The search for the pair production of second-generation scalar leptoquarks and measurements of the differential cross sections of the W boson produced in association with jets with the CMS detector at the LHC

by Darin C. Baumgartel

B.A. in Physics and Economics, Rutgers University
M.S. in Physics, Northeastern University

A dissertation submitted to

The Faculty of
the College of Science of
Northeastern University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

September 15, 2014

Dissertation directed by

Emanuela Barberis
Professor of Physics

Acknowledgements

I would like to express my profound gratitude to my advisor, Professor Emanuela Barberis, for her endless support, guidance, and willingness to thoughtfully (and frequently) answer my questions. Her active involvement in my work and genuine motivation to see me succeed have made the past few years of research a more productive and enjoyable experience than I could have imagined. I would also like to thank my Thesis Committee members, Professor George Alverson, Professor Haim Goldberg, and Professor Darien Wood, whom I have had the privilege of learning from during my studies, and whose support and guidance has made this work possible.

I am grateful to my current and former collaborators at Northeastern University and CMS, whom I have had the joy of working with over the past six years. In particular, I would like to thank Professor Joseph Haley, Professor Toyoko Orimoto, Professor Oana Boeriu, Dr. Tim Cox, Dr. David Morse, Dr. Daniele Trocino, Dr. Jinzhong Zhang, Bhawan Uppal, Matthew Chasco, Apichart Hortiangtham, and David Nash, for their help, support, and collaboration.

I am immensely thankful to my mother, Kathie Bryan, my late father, David C. Baumgartel, my stepfather, Dave Bryan, and my brother, David K. Baumgartel, for shaping me into the person I am today. While my family may insist that they have no idea how they raised a physicist, it was because of them that I had the courage to pursue my goals. Finally, I would like to thank my fiancée, Elizabeth Bury, for

her unwavering support and her ability to make me laugh when life seemed otherwise daunting.

Abstract of Dissertation

Since the formulation of the Standard Model of particle physics, numerous experiments have sought to observe the signatures of the subatomic particles by examining the outcomes of charged particle collisions. Over time, advances in detector technology and scientific computing have allowed for unprecedented precision measurements of Standard Model phenomena and particle properties. Although the Standard Model has displayed remarkable predictive power, extensions to the Standard Model have been formulated to account for unexplained phenomena, and these extensions often infer the existence of additional subatomic particles. Consequently, experiments at particle colliders often endeavor to search for signatures of physics beyond the Standard Model. These searches and measurements are often complementary pursuits, as searches are often limited by the precision of estimations of the Standard Model backgrounds.

At the forefront of present-day collider experiments is the Large Hadron Collider at CERN, which delivers proton-proton collisions with unprecedented energy and luminosity. Collisions are recorded with detectors located at interaction points along the ring of the Large Hadron Collider. The CMS detector is one of two general-purpose detectors at the Large Hadron Collider, and the high-precision detection of particles from collision events in the CMS detector make the CMS detector a powerful tool for both Standard-Model measurements and searches for new physics.

The Standard Model is characterized by three generation of quarks and leptons.

This correspondence between the generations of quarks and leptons is necessary to allow for the renormalizability of the Standard Model, but it is not an inherent property of the Standard Model. Motivated by this compelling symmetry, many theories and models propose the existence of leptoquark bosons which mediate transitions between quarks and leptons. Experimental constraints indicate that leptoquarks would couple to a single generation, and this thesis describes searches for leptoquarks produced in pairs and decaying to final states containing either two muons and two jets, or one muon, one muon-neutrino, and two jets. Searches are conducted with collision data at center-of-mass energies of both 7 TeV and 8 TeV. No compelling evidence for the existence of leptoquarks is found, and upper limits on the leptoquark mass and cross section are placed at the 95% confidence level. These limits are the most stringent to date, and are several times larger than limits placed previously at hadron collider experiments.

While the pair production of massive leptoquark bosons yields final states which have strong kinematic differences from the Standard Model processes, the ability to exploit these differences is limited by the ability to accurately model the backgrounds. The most notable of these backgrounds is the production of a W boson in association with one or more jets. Since the W+jets process has a very large cross section and a final state containing missing energy, its contribution to the total Standard Model background is both nominally large and more difficult to discriminate against than backgrounds with only visible final state objects. Furthermore, estimates of this background are not easily improved by comparisons with data in control regions, and simulations of the background are often limited to leading-order predictions. To improve the understanding and modeling of this background for future endeavors, this thesis also presents measurements of the W+jets process differentially as a function of several variables, including the jet multiplicity, the individual jet transverse momenta and

pseudorapidities, the angular separation between the jets and the muon, and the scalar sum of the transverse momenta of all jets. The agreement of these measurements with respect to predictions from event leading-order generators and next-to-leading-order calculations is assessed.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Standard Model of Particle Physics

The Standard Model (SM) of particle physics provides the best known description of the subatomic particles and the strong and electroweak forces which govern their interactions. Since the early 1970's, the SM has shown a remarkable predictive power which has been verified experimentally to high precision, and has been used to predict the existence of a number of experimentally confirmed subatomic particles, including the W and Z bosons [1–4], the top quark [5,6], the tau neutrino [7], and most recently the Higgs boson [8, 9]. This Section gives a brief overview of the SM, and a more thorough description can be found in Refs [10–12].

The fundamental particles incorporated into the SM include 12 fermions of spin-$\frac{1}{2}$, including three generations of two fractionally charged quarks and three generations of unit-charged leptons and neutrally charged lepton-neutrinos. The fundamental forces between the fermions are mediated by spin-1 gauge bosons. These include the massless and neutral photon (γ), which mediates the electromagnetic force between charged particles, the W^+, W^-, and Z^0 bosons, which mediate the weak nuclear interaction,

and the eight gluons, which mediate the strong interaction.

The theory of quantum electrodynamics (QED) fully describes the mediation of electromagnetic interaction via photons. The range of the electromagnetic force is infinite due to the zero rest mass of the photon. The electroweak theory further formulates the weak interaction of electrically charged particles in terms of the photon and the W and Z bosons. These vector bosons are massive and have a short lifetime, which gives the weak force a small interaction strength compared to that of the electromagnetic force, for which the weak force is named. The strong force, described by quantum chromodynamics (QCD), mediates the interaction between particles carrying color charge. It is the strongest of the fundamental forces and acts on the shortest distance scale. It is responsible both for the binding of nucleons in an atomic nucleus, and for the binding of quarks in an individual nucleon.

There are six flavors of leptons, divided among three generations, each consisting of one charged lepton and one neutrino. The first generation consists of the electron (e^-) and the electron neutrino (ν_e), while the second and third generations consist of the muon and muon neutrino (μ^- and ν_μ), and the tau and tau neutrino (τ^- and ν_τ), respectively. Each charged lepton has an oppositely-charged antiparticle. As spin-$\frac{1}{2}$ fermions, the charged leptons carry an intrinsic magnetic dipole moment inversely proportional to their mass, and proportional to a g-factor which has higher-order corrections referred to as the anomalous dipole moment. Predictions of the anomalous dipole moment have often served as a fundamental test of the SM. Leptons also interact via the weak interaction, and quantum numbers in addition to charge and spin are needed to describe this interaction. The helicity is defined by the direction of the projection of the lepton's spin onto its' momentum, and the chirality is defined by the representation of the Poincaré group under which the lepton transforms: left-handed (L) or right-handed (R). In the SM, the weak interactions involving the W boson affect

only left-handed leptons, while the Z boson may interact regardless of chirality. Each generation of leptons can be represented by a doublet in the weak-isospin $T = 1/2$ representation of the SU(2) gauge group. In this representation, the left-handed particles are eigenstates of the third component (T_3) of the weak-isospin. A fourth and final quantum-number, the weak hypercharge (Y), relates the electric charge and the weak isospin via the equality $Y = 2(Q - T_3)$. Thereby, the electromagnetic and weak forces are combined into in a unified electroweak description.

The quarks are fractionally-charged fermions which possess an additional property known as color (or color-charge), which comes in three values (r, g, and b), and determines the possible interactions via the strong force. The three generations of quarks each consist of one quark of electric charge $+\frac{2}{3}$ (including the up, charm, and top quarks) and one quark of charge $-\frac{1}{3}$ (including the down, strange, and bottom quarks). Hadrons, or particles which are composites of multiple quarks, exist in colorless combinations including baryons with exactly one r, g, and b quark, and mesons consisting of one quark and one anti-quark of the same color. Familiar examples of baryons include protons and neutrons, while a variety of pseudoscalar and vector mesons exist. Quarks interact electromagnetically, weakly, and strongly. Quarks may absorb or emit a W boson, causing a flavor transformation and allowing for common processes such as beta decay and inverse beta decay. The strong interactions are mediated by eight gluons, represented by a linearly independent color octet of superpositions of quark-antiquark states. The interactions of quarks and gluons together are described by QCD, which is defined by an SU(3) gauge symmetry. The spinor fields of the fundamental representation represent the quarks, while the vector fields of the adjoint representation represent the gluons.

The origin of the mass of the fermions and gauge bosons in the SM is described by the Higgs mechanism. As described so far, the SM is rooted in a gauge theory in the

$\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ symmetry group, with the $\mathrm{SU}(3)_C$ component describing the strong interaction and the $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ describing the electroweak interaction. The local gauge invariance of this theory requires the particles to be massless. The Higgs mechanism introduces an additional scalar (spin-0) field which is a four-component complex doublet of the SU(2) symmetry. The Higgs field interacts with the W and Z bosons under the $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ symmetry. The ground state of the Higgs field can be chosen such that the Higgs field has a non-zero vacuum expectation value, v, and the SU(2) and $\mathrm{U}(1)_Y$ symmetries are broken. This non-zero v gives mass to the W and Z bosons. Since the $\mathrm{U}(1)_{\mathrm{em}}$ symmetry remains unbroken, the photon remains massless. The fermion masses originate from interactions with the Higgs field via Yukawa couplings. The three generations of quarks and leptons, along with their mass and quantum numbers are detailed in Table 1.1.

While leptons are often observed separately, single quarks have never been observed, due to a phenomena known as "quark confinement" or "color confinement". Any mechanism which would separate quarks in a color-neutral bound state would add an amount of energy to the system such that it would become favorable to create a quark-antiquark pair from the vacuum, which would then form color-neutral bound states with the quarks previously undergoing separation. The separation of quarks is a common occurrence in collider experiments, and the principle of color confinement causes such interactions to result in narrow jets of color-neutral mesons and baryons.

Many searches for beyond-the-standard-model (BSM) physics are motivated by some inherent limitations of the SM. The description of the fundamental forces in the SM lacks a description of gravity altogether. In doing so, while the SM is capable of explaining the relative strength of the electromagnetic, weak, and strong forces, it offers no explanation for the relative difference in the strength of the electroweak and gravitational forces. This observed difference, known as the "hierarchy problem", is

also fundamentally related to the problem of fine-tuning of the radiative corrections of the Higgs mass, which has been observed to be much lighter than the Planck mass. Furthermore, the SM and the Higgs mechanism predict massive charged leptons and massless neutrinos, while experimental confirmations of neutrino oscillation indicate that neutrinos should be massive.

The SM also leaves many open questions in the context of astrophysics. While several astrophysical observations have indicated the existence of dark matter which accounts for the majority of the known mass of the universe, the SM offers no viable candidates for this dark matter. Furthermore, the SM does not provide a mechanism for the observed expansion of the universe and the large amount of so-called dark energy which contributes to this expansion.

Lastly, the SM offers no explanation for the compelling symmetry between the three generations of leptons and the three generations of quarks. Numerous extensions to the SM attempt to address some or all of these shortcomings, and this symmetry is addressed in several theories and models with the addition of a new boson, the leptoquark, which couples to both leptons and quarks. The search for this particle, and precision measurements of one of the most crucial and challenging backgrounds to this search, comprise the body of this thesis.

	Generation	Fermion	Mass (GeV/c^2)	T	T_3	Q	Y
Leptons	1	e_L^-	5.11×10^{-4}	1/2	−1/2	−1	−1
		e_R^-		0	0		−2
		ν_e	$< 2 \times 10^{-9}$	1/2	1/2	0	−1
	2	μ_L^-	1.06×10^{-1}	1/2	−1/2	−1	−1
		μ_R^-		0	0		−2
		ν_μ	$< 2 \times 10^{-4}$	1/2	1/2	0	−1
	3	τ_L^-	1.78	1/2	−1/2	−1	−1
		τ_R^-		0	0		−2
		ν_τ	$< 2 \times 10^{-2}$	1/2	1/2	0	−1
Quarks	1	u_L	$\approx 2.3 \times 10^{-3}$	1/2	1/2	2/3	1/3
		u_R		0	0		4/3
		d_L	$\approx 4.8 \times 10^{-3}$	1/2	−1/2	−1/3	1/3
		d_R		0	0		−2/3
	2	c_L	≈ 1.3	1/2	1/2	2/3	1/3
		c_R		0	0		4/3
		s_L	$\approx 1 \times 10^{-1}$	1/2	−1/2	−1/3	1/3
		s_R		0	0		−2/3
	3	t_L	$\approx 1.73 \times 10^2$	1/2	1/2	2/3	1/3
		t_R		0	0		4/3
		b_L	≈ 4.2	1/2	−1/2	−1/3	1/3
		b_R		0	0		−2/3

Table 1.1: The quarks and leptons, along with their masses and quantum numbers, including the weak isospin (T), the third component of the weak isospin (T_3), the weak hypercharge (Y), and the electric charge (Q) [13].

1.2 Leptoquarks

Inherent to the SM is a compelling symmetry between quarks and leptons. Both consist of three generations and are grouped into weak isospin doublets and singlets. The renormalizability of the SM depends explicitly on the cancellation of the lepton and quark contributions to the triangle anomalies. Motivated by this symmetry, several theories beyond the SM propose a fundamental relationship between the quarks and leptons via a new boson called the leptoquark (LQ), which mediates quark-lepton transitions. Such theories include SU(5) grand unified theories [14] (GUTs), Pati-Salam SU(4) [15], composite models [16], technicolor [17–19], and superstring-inspired E_6 models [20].

The leptoquark would be a fractionally charged, colored boson which decays into a charged lepton and a quark with an unknown branching fraction β and an unknown lepton and quark coupling strength λ, and may be either spin-0 (scalar) or spin-1 (vector). It is generally considered that LQs couple to quarks and leptons of the same generation because of constraints from experiments sensitive to flavor changing neutral currents, lepton family-number violation, and other rare processes [21]. Thus, LQs are classified by the generation of quarks and leptons to which they decay. First-generation LQs may decay to an e or a ν_e and a u or a d quark. Second-generation LQs may decay to a μ or a ν_μ and a c or an s quark. Third-generation LQs may decay to an τ or a ν_τ and a t or a b quark. Leptoquarks were first proposed in SU(5) GUTs, which allowed for proton decay via leptoquark exchanges. Consequently, leptoquarks in GUT models were restricted to very large masses ($\approx 10^{15}$ GeV) to avoid rapid proton decay. Other theories in which baryon number and lepton number are separately conserved allow for leptoquarks in mass ranges accessible to modern colliders.

As with most searches for LQs with colliders, this thesis considers LQ production

according to a general effective model, described in Section 1.2.1. The signatures of LQ production at the LHC are discussed in Section 1.2.2, and Section 1.2.3 describes previous searches for LQs at colliders, as well as the searches presented in this thesis.

1.2.1 Buchmüller-Rückl-Wyler Effective Model

This thesis describes a search for leptoquarks described by the effective Lagrangian put forth by Buchmüller, Rückl, and Wyler [22] in 1986, in advance of data-taking at the HERA collider. This Lagrangian is constructed with dimensionless couplings for the scalar and vector LQs under SU(3)×SU(2)×U(1) invariance, and the leptoquark interactions are assumed to conserve both baryon number (B) and lepton number (L). For quarks, $B = +1/3$, and for leptons, $L = 1$. To avoid rapid proton decay, the leptoquarks are assumed to conserve B and L separately. The leptoquark interactions can be characterized by the fermion number $F = 3B + L$, and the Lagrangian takes the form of Eqn. 1.1,

$$\mathcal{L} = \mathcal{L}_{|F|=0} + \mathcal{L}_{|F|=2} \tag{1.1}$$

with $\mathcal{L}_{|F|=0}$ $\mathcal{L}_{|F|=2}$ defined in Eqn. 1.2 and Eqn. 1.3.

$$\begin{aligned}
\mathcal{L}_{|F|=0} &= (h_{2L}\bar{u}_R\ell_L + h_{2R}\bar{q}_L i\tau_2 e_R)S_{\frac{1}{2}} + \tilde{h}_{2L}\bar{d}_R\ell_L\tilde{S}_{\frac{1}{2}} \\
&+ (h_{1L}\bar{q}_L\gamma^\mu\ell_L + h_{1R}\bar{d}_R\gamma^\mu e_R)V_{0\mu} \\
&+ \tilde{h}_{1R}\bar{u}_R\gamma^\mu e_R\tilde{V}_{0\mu} + h_{3L}\bar{q}_L\vec{\tau}\gamma^\mu\ell_L V_{1\mu} + h.c.,
\end{aligned} \tag{1.2}$$

$$\begin{aligned}
\mathcal{L}_{|F|=2} &= (g_{1L}\bar{q}^c_L i\tau_2\ell_L + g_{1R}\bar{u}^c_R e_R)S_0 \\
&+ \tilde{g}_{1R}\bar{d}^c_R e_R\tilde{S}_0 + g_{3L}\bar{q}^c_L i\tau_2\vec{\tau}\ell_L\vec{S}_1 \\
&+ (g_{2L}\bar{d}^c_R\gamma^\mu\ell_L + g_{2R}\bar{q}^c_L\gamma^\mu e_R)V_{\frac{1}{2}\mu} \\
&+ \tilde{g}_2\bar{u}^c_R\gamma^\mu\ell_L\tilde{V}_{\frac{1}{2}\mu} + h.c.
\end{aligned} \tag{1.3}$$

Here, the SM fermion content is assumed. The scalar and vector leptoquarks are represented by S and V respectively, following the conventions of Ref. [23]. The Pauli matrices are denoted τ_i and the Dirac matrices are represented by γ^μ. The SU(2) quark and lepton doublets are represented by q_L and ℓ_L, respectively. The color and generational indices are omitted for simplicity. The singlet fields are denoted by u_R, d_R, and e_R, and the charge-conjugate fields are denoted with the superscript c, i.e. $\psi^c = C\overline{\psi}^T$. The subscript of the LQ (0, $\frac{1}{2}$, or 1) denotes the isospin under $\mathrm{SU}(2)_L$. The subscripts L and R refer to the chirality of the charged lepton, and the Yukawa coupling ($\lambda_{L,R}$) is denoted as $g_{iL,R}$ in $\mathcal{L}_{|F|=2}$ and $h_{iL,R}$ in $\mathcal{L}_{|F|=0}$, where i represents the generation of the LQ coupling.

The LQs with $|F| = 2$ are restricted to couple to ℓq, whereas the LQs with $|F| = 0$ couple to $\ell\overline{q}$. Two distinct leptoquarks may share the same fermion number, spin, and isospin, but differ in hypercharge and electric charge, and in this case are distinguished by a tilde. A detailed overview of the distinct leptoquark types, along with their quantum numbers, couplings, and decay modes, is given in Table 1.2. In total, the BRW model allows for seven distinct scalar LQs and seven distinct vector LQs. While each LQ has one or more well-defined decay modes, the searches are conducted assuming an unknown branching fraction (β) of the LQ to a charged lepton and quark, in essence relaxing the assumption that LQs couple only to SM fermions and gauge bosons, and retaining generality.

Several experimental observations place constraints on LQ interactions. In order to satisfy constraints from bounds on flavor-changing neutral currents and from rare pion and kaon decays [22, 24], it is assumed that leptoquarks couple to quarks and leptons of a single generation. One simple example of this is the $K \to \mu e$ process, which would have large LQ-induced contributions if LQs coupled to more than one generation. Accordingly, leptoquark searches are conducted in decay modes specific to

	LQ	F	Y	T	T_3	Q	$\lambda_L(\ell q)$	$\lambda_R(\ell q)$	$\lambda_L(\nu q)$	Decay Modes
Scalar	$S_{0,L}$	-2	2/3	0	0	1/3	g_{1L}	0	$-g_{1L}$	$e_L^+\overline{u}_L$, $\overline{\nu}_L\overline{d}_L$
	$S_{0,R}$	-2	2/3	0	0	1/3	0	g_{1R}	0	$e_R^+\overline{u}_R$
	$\tilde{S}_{0,R}$	-2	8/3	0	0	4/3	0	$\tilde{g}_{1R}$	0	$e_R^+\overline{d}_R$
	$S_{\frac{1}{2},L}$	0	7/3	1/2	1/2	5/3	h_{2L}	0	0	$e_L^+u_L$
					$-1/2$	2/3	0	0	h_{2L}	$\overline{\nu}_L u_L$
	$S_{\frac{1}{2},R}$	0	7/3	1/2	1/2	5/3	0	h_{2R}	0	$e_R^+u_R$
					$-1/2$	2/3	0	$-h_{2R}$	0	$e_R^+d_R$
	$\tilde{S}_{\frac{1}{2},L}$	0	1/3	1/2	1/2	5/3	$\tilde{h}_{2L}$	0	0	$e_L^+d_L$
					$-1/2$	2/3	0	0	$\tilde{h}_{2L}$	$\overline{\nu}_L d_L$
	$S_{1,L}$	-2	2/3	1	1	4/3	$-\sqrt{2}g_{3L}$	0	0	$e_L^+\overline{d}_L$
					0	1/3	$-g_{3L}$	0	$-g_{3L}$	$e_L^+\overline{u}_L$, $\overline{\nu}_L\overline{d}_L$
					-1	$-2/3$	0	0	$-\sqrt{2}g_{3L}$	$\overline{\nu}_L\overline{u}_L$
Vector	$V_{0,L}$	0	4/3	0	0	2/3	h_{1L}	0	h_{1L}	$e_L^+d_R$, $\overline{\nu}_L u_L$
	$V_{0,R}$	0	4/3	0	0	2/3	0	h_{1R}	0	$e_R^+d_L$
	$\tilde{V}_{0,R}$	0	10/3	0	0	5/3	0	$\tilde{h}_{1R}$	0	$e_R^+u_L$
	$V_{\frac{1}{2},L}$	-2	5/3	1/2	1/2	4/3	g_{2L}	0	0	$e_L^+\overline{d}_R$
					$-1/2$	1/3	0	0	g_{2L}	$\overline{\nu}_L\overline{d}_R$
	$V_{\frac{1}{2},R}$	-2	5/3	1/2	1/2	4/3	0	g_{2R}	0	$e_R^+\overline{d}_L$
					$-1/2$	1/3	0	g_{2R}	0	$e_R^+\overline{u}_L$
	$\tilde{V}_{\frac{1}{2},L}$	-2	-1/3	1/2	1/2	1/3	$\tilde{g}_{2L}$	0	0	$e_L^+\overline{u}_R$
					$-1/2$	$-2/3$	0	0	$\tilde{g}_{2L}$	$\overline{\nu}_L\overline{u}_R$
	$V_{1,L}$	0	4/3	1	1	5/3	$\sqrt{2}h_{3L}$	0	0	$e_L^+u_R$
					0	2/3	$-h_{3L}$	0	$-h_{3L}$	$e_L^+\overline{d}_R$, $\overline{\nu}_L u_R$
					-1	$-1/3$	0	0	$\sqrt{2}h_{3L}$	$\overline{\nu}_L d_R$

Table 1.2: The types of scalar and vector lepqtoquarks, their decay modes, and their quantum numbers including the weak isospin (T), the third component of the weak isospin (T_3), the weak hypercharge (Y), the electric charge (Q), and the fermion number (F). Also shown are the couplings.

each generation. Searches for second generation LQs are presented in this thesis.

An additional constraint is given by low energy data. For instance, a first generation LQ coupled to $u\overline{e}_R$ with a coupling constant λ_L, and to $d\nu_L$ with a coupling constant λ_L, results in the interaction Lagrangian given in Eqn 1.4,

$$\mathcal{L} = \frac{\lambda_R \lambda_L}{2M_{\mathrm{LQ}}^2} \overline{u}_R d_L \overline{e}_R \nu_L \tag{1.4}$$

which would have large contributions to the $\pi^+ \to e^+ \overline{\nu}_e$ process. Comparison with experimental data can constrain the LQ mass as a function of the couplings as $M_{\mathrm{LQ}} > 2 \times 10^5 \times |\lambda_L \lambda_R|$ GeV, indicating one or both Yukawa couplings must be negligible.

1.2.2 LQ Pair Production and Final-State Signatures in pp Collisions at the LHC

In pp collisions, LQs can be produced singly or in pairs. Single production occurs in association with a lepton through quark-gluon fusion, and may yield a final state similar to that of pair production, but with one fewer jet. For LQ masses below 1 TeV, pair-production processes are dominant. When compared to single LQ production, the relatively higher cross section of pair-production, combined with the presence of four high-momentum final-state particles, allows for more powerful LQ searches in the TeV mass range. The production of LQ-$\overline{\mathrm{LQ}}$ pairs at the LHC is dominated by gluon-gluon fusion with additional contributions from quark-antiquark annihilation, as indicated in Eqn. 1.5 and Eqn. 1.6.

$$g + g \to \mathrm{LQ} + \overline{\mathrm{LQ}} \tag{1.5}$$

$$q + \overline{q} \to \mathrm{LQ} + \overline{\mathrm{LQ}} \tag{1.6}$$

As color triplet particles, the dominant diagrams for LQ pair production do not depend on the Yukawa coupling (λ). The dominant diagrams contributing to the pair production of LQs are given in Figure 1.1. Of these diagrams, only one exhibits λ-dependence, which is of order λ^2, and this contribution is effectively negligible [25].

Given the lack of coupling dependence of the dominant LQ pair production mechanisms, searches for LQ pairs are often performed just as a function of LQ mass and branching fraction to a charged lepton and a quark.

The cross section for the pair production of scalar LQs in hadron collisions has been calculated at next-to-leading order (NLO), using the method described in Ref [26]. This method considers higher-order QCD radiative corrections, including gluon bremsstrahlung and initial-state (IS) parton splittings, and significantly reduces the dependence on the renormalization and factorization scales. The cross-section for the LQ pair production process for center-of-mass energies of $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV, for masses in the range of 200 – 1200 GeV, are presented in Table 1.3.

The signatures for LQ pair production include four-object final states with leptons (ℓ), neutrinos (ν), and jets (j), with three distinct types of final state signatures, $\ell^+\ell^- jj$, $\ell^\pm \nu jj$, and $\nu\nu jj$. In final states with a charged lepton, the major SM backgrounds which can mimic the final-state signatures of LQ pair production include the production of a vector boson (W or Z) in association with 2 or more jets, and $\mathrm{t\bar{t}}$ production. The $\mathrm{t\bar{t}}$ background may originate from multiple decay modes. The fully leptonic decay mode, $\mathrm{t\bar{t}} \to bb\mathrm{W}(\to \ell\nu)\mathrm{W}(\to \ell\nu)$, contains the necessary final-state objects to mimic the LQ pair production in the $\ell\ell jj$ final state, albeit with extra significant missing transverse energy from the neutrinos in the event. Similarly, the semi-leptonic decay mode, $\mathrm{t\bar{t}} \to bb\mathrm{W}(\to \ell\nu)\mathrm{W}(\to jj)$, contains the necessary final-state objects to mimic the LQ pair production in the $\ell\nu jj$ final state, but with two extra (heavy-flavor) jets in the final state. Additional small contributions to the SM background prediction

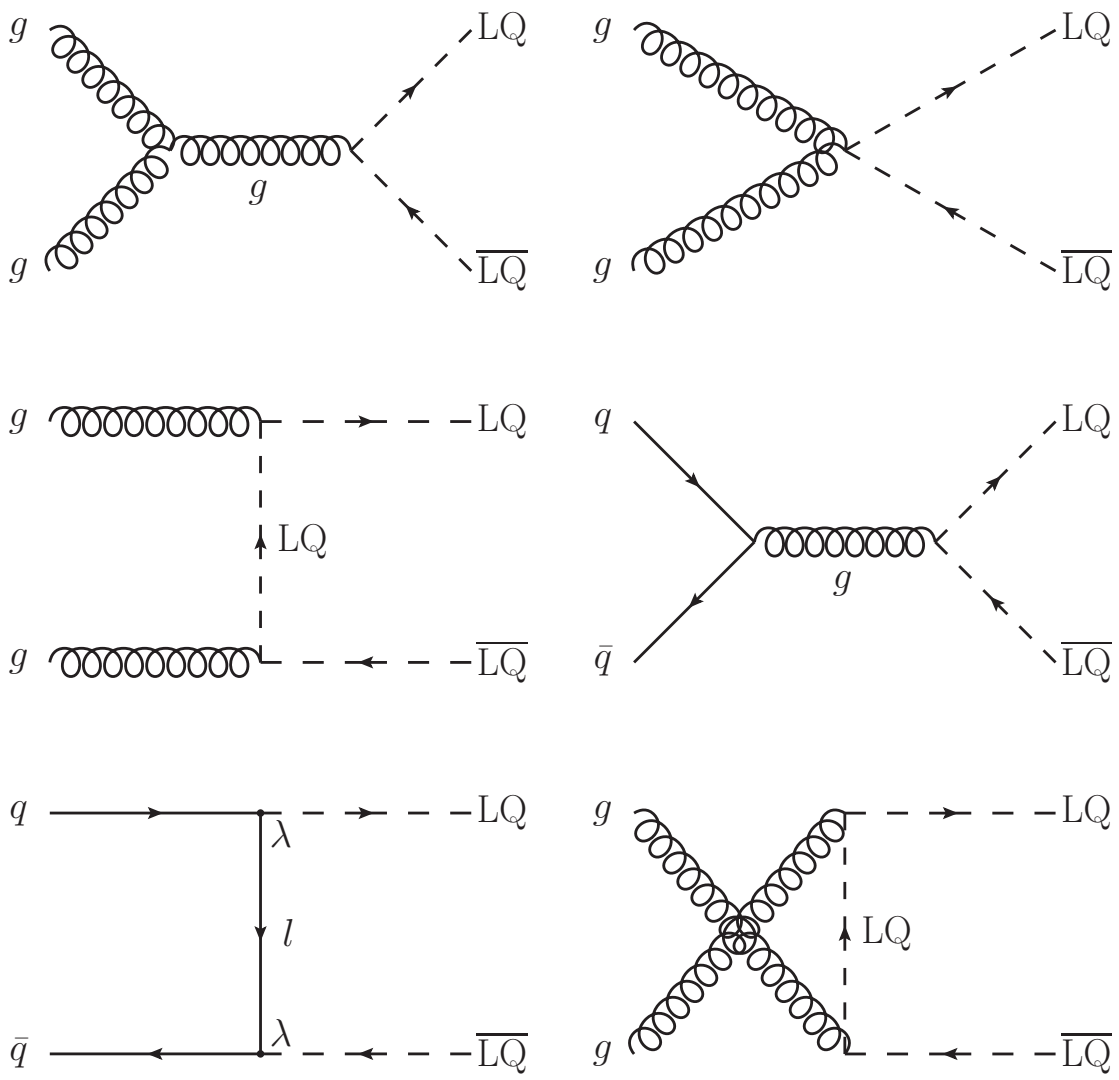

Figure 1.1: Dominant leading order diagrams for the pair production of LQs in pp collisions. The production via gluon-gluon fusion is dominant. All diagrams, except for the t-channel lepton exchange diagram (bottom-left) are independent of the Yukawa coupling (λ). The bottom-left diagram, of order λ^2, can be considered as negligible.

	$\sqrt{s} = 7$ TeV		$\sqrt{s} = 8$ TeV	
M_{LQ} [GeV]	σ_{NLO} [pb]	CTEQ6 PDF Unc. [pb]	σ_{NLO} [pb]	CTEQ6 PDF Unc. [pb]
200	1.19×10^{1}	9.86×10^{-1}	1.74×10^{1}	1.24×10^{0}
250	3.47×10^{0}	3.72×10^{-1}	5.26×10^{0}	4.87×10^{-1}
300	1.21×10^{0}	1.58×10^{-1}	1.89×10^{0}	2.14×10^{-1}
350	4.77×10^{-1}	7.28×10^{-2}	7.70×10^{-1}	1.02×10^{-1}
400	2.05×10^{-1}	3.57×10^{-2}	3.42×10^{-1}	5.20×10^{-2}
450	9.48×10^{-2}	1.85×10^{-2}	1.63×10^{-1}	2.78×10^{-2}
500	4.63×10^{-2}	9.96×10^{-3}	8.20×10^{-2}	1.15×10^{-2}
550	2.36×10^{-2}	5.58×10^{-3}	4.31×10^{-2}	8.93×10^{-3}
600	1.24×10^{-2}	3.21×10^{-3}	2.35×10^{-2}	5.30×10^{-3}
650	6.76×10^{-3}	1.90×10^{-3}	1.32×10^{-2}	3.22×10^{-3}
700	3.77×10^{-3}	1.14×10^{-3}	7.61×10^{-3}	2.00×10^{-3}
750	2.14×10^{-3}	7.00×10^{-4}	4.48×10^{-3}	1.26×10^{-3}
800	1.24×10^{-3}	4.37×10^{-4}	2.69×10^{-3}	8.10×10^{-4}
850	7.32×10^{-4}	2.76×10^{-4}	1.64×10^{-3}	5.27×10^{-4}
900	4.36×10^{-4}	1.76×10^{-4}	1.01×10^{-3}	3.47×10^{-4}
950	2.63×10^{-4}	1.13×10^{-4}	6.34×10^{-4}	2.31×10^{-4}
1000	1.60×10^{-4}	7.37×10^{-5}	4.01×10^{-4}	1.55×10^{-4}
1050	9.82×10^{-5}	4.83×10^{-5}	2.56×10^{-4}	1.05×10^{-4}
1100	6.06×10^{-5}	3.18×10^{-5}	1.65×10^{-4}	7.18×10^{-5}
1150	3.77×10^{-5}	2.10×10^{-5}	1.07×10^{-4}	4.92×10^{-5}
1200	2.35×10^{-5}	1.40×10^{-5}	6.96×10^{-5}	3.40×10^{-5}

Table 1.3: The NLO calculation of the cross-section and CTEQ6 PDF uncertainty for scalar LQs with masses ranging from 200 GeV to 1200 GeV, at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV. These quantities were calculated with the methods described in Ref [26].

originate from diboson (WW, WZ, and ZZ) production, single t-quark production, and QCD multijet production with one or two jets misidentified as leptons.

As noted previously, this thesis considers LQ production and decays with the branching fraction β of an LQ to a charged lepton and a quark as a free parameter. For the case of pair production, this results in an effective branching fraction of β^2 in the $\ell\ell jj$ channel and $2\beta(1-\beta)$ in the $\ell\nu jj$ channel.

1.2.3 Previous Search Results and Prospects

In addition to the indirect limits which restrict the values of LQ couplings discussed in Section 1.2.1, several direct experimental limits have already been placed on the LQ mass and production cross section.

The earliest direct searches for scalar leptoquarks were conducted in s-channel production modes at the HERA ep collider by the H1 [27,28] and ZEUS [29] collaborations, neither of which found evidence for leptoquarks. Limits were set at the 95% confidence level, which excluded $|F| = 0$ first-generation LQs with $M_{\mathrm{LQ}} < 250$ GeV and $|F| = 2$ first-generation LQs with $M_{\mathrm{LQ}} < 215$ GeV, assuming a $\lambda > \lambda_{\mathrm{EM}}$. The ZEUS limit on LQ left-handed and right-handed coupling as a function of LQ mass, for different LQ decay modes, is shown in Figure 1.2.

Later, in 1997, the H1 and ZEUS collaborations reported an excess [30, 31] of neutral-current deep-inelastic scattering ($e^+p \rightarrow e^+X$) events at high Q^2, where Q is the four-momentum transfer. This excess was concentrated near M_{e^+j} (positron-jet invariant mass) values of 200 GeV. Figure 1.3 shows the excess of data events over the prediction of neutral-current deep-inelastic scatter events observed by the H1 collaboration. This observation increased interest in LQ searches [25].

Subsequent searches with the D0 and CDF detectors at the Tevatron were able to rule out LQs with $\beta = 1$ and $M_{\mathrm{LQ}} < 242$ GeV using a combination of the $eejj$ channel

searches performed with both detectors [32]. Since then, for $\beta = 1(0.5)$, searches at D0 have ruled out first-generation scalar LQs with a mass of 299 (284) GeV [33] and second-generation scalar LQs with a mass of 316 (270) GeV [34]. The D0 limits on the branching fraction β, as a function of the LQ mass, are given in Figure 1.4.

To date, the CMS and ATLAS detectors at the LHC have recorded 5 fb^{-1} of data with a center-of-mass energy $\sqrt{s} = 7$ TeV, and 20 fb^{-1} of data with a center-of-mass energy $\sqrt{s} = 8$ TeV. This is an unprecedented feat for a hadron collider, both in terms of integrated luminosity and collision energy. The increase in energy from that of the Tevatron to the 8 TeV running of the LHC gives LQ cross section increases of nearly a factor of 10. The increase in integrated luminosity also allows for further reach in LQ searches, which are statistically limited. Furthermore, advances in detector technology, described in Chapter 2, and advances in reconstruction, described in Chapter 3, allow for precise measurements of kinematic quantities. The larger integrated luminosities and advanced reconstruction techniques also result in a reduction of the systematic uncertainties associated with the measurements with the CMS detector.

In this thesis, searches for second-generation scalar LQs with the CMS detector are presented, as a function of LQ mass and β, using all of the pp collision data recorded to date with the CMS detector. As part of this search, the SM backgrounds to the LQ signal are studied in detail, and data-driven approaches to estimating these backgrounds are utilized. One particularly challenging background is the W+jets background, which gives a significant contribution to the total SM background in searches in the $\ell\nu jj$ channel. This background, and the unique challenges which it presents, are introduced in Section 1.3, and measurements of the differential cross-section of the W+n jets process are presented in Chapter 6.

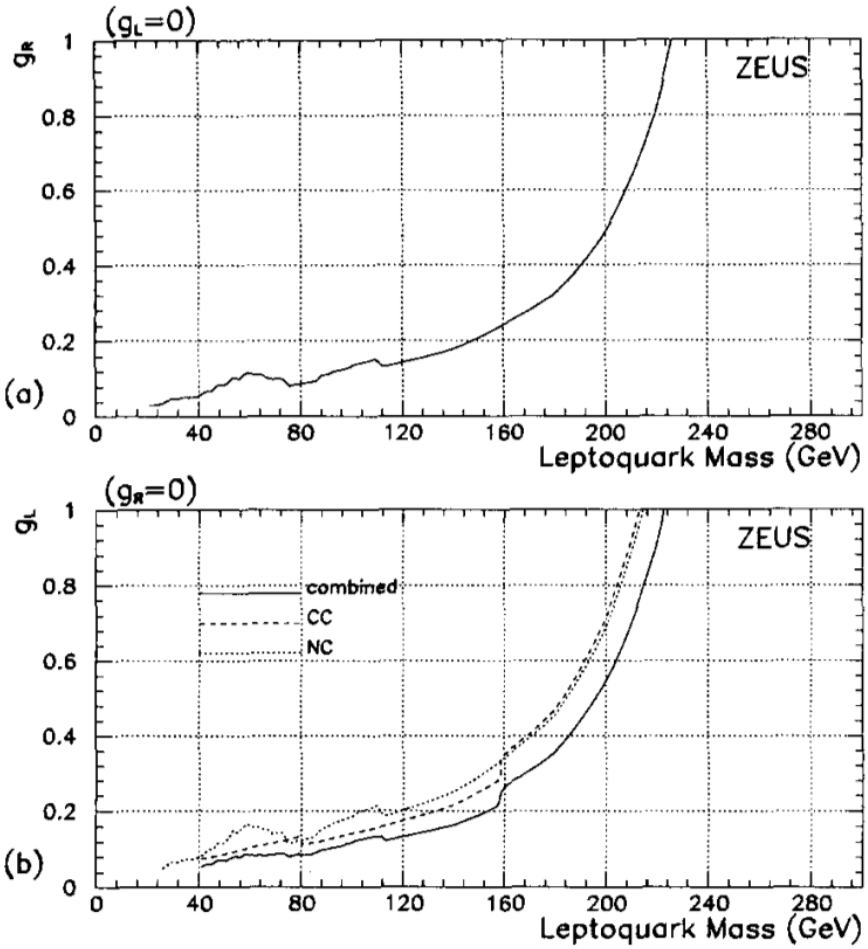

Figure 1.2: The ZEUS limits on the LQ mass and couplings, including (a) the limit on the right-handed coupling from the LQ $\rightarrow ej$ decay mode, assuming $\beta = 1$. (b) the limit on the left-handed coupling for the LQ $\rightarrow ej$ decay mode decay mode (dotted), for the LQ $\rightarrow \nu j$ decay mode (dashed), and for the combination (solid), assuming $\beta = 1/2$. [29].

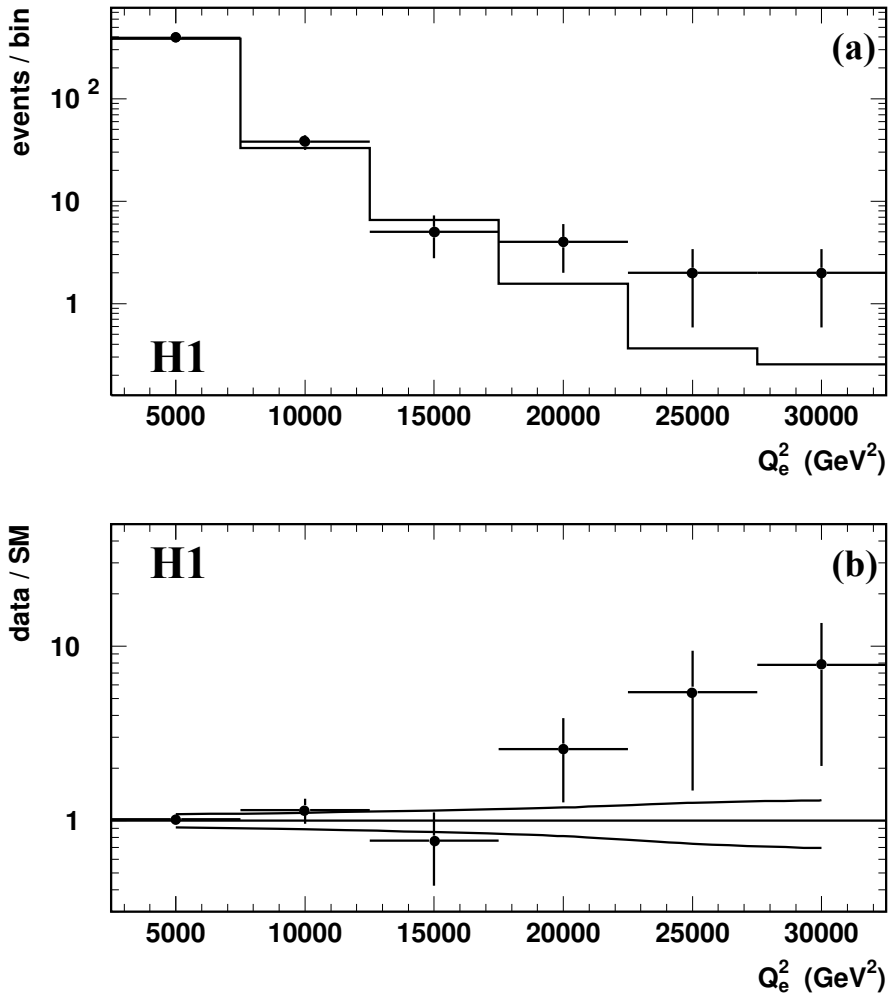

Figure 1.3: The H1 observation of an excess of high Q_e^2 events. (a) Black circular markers indicate the observation in data, and the solid line indicates the SM prediction. (b) The ratio of the data to the SM prediction is given, with the lines above and below unity indicating the $\pm 1\sigma$ uncertainty range. [30].

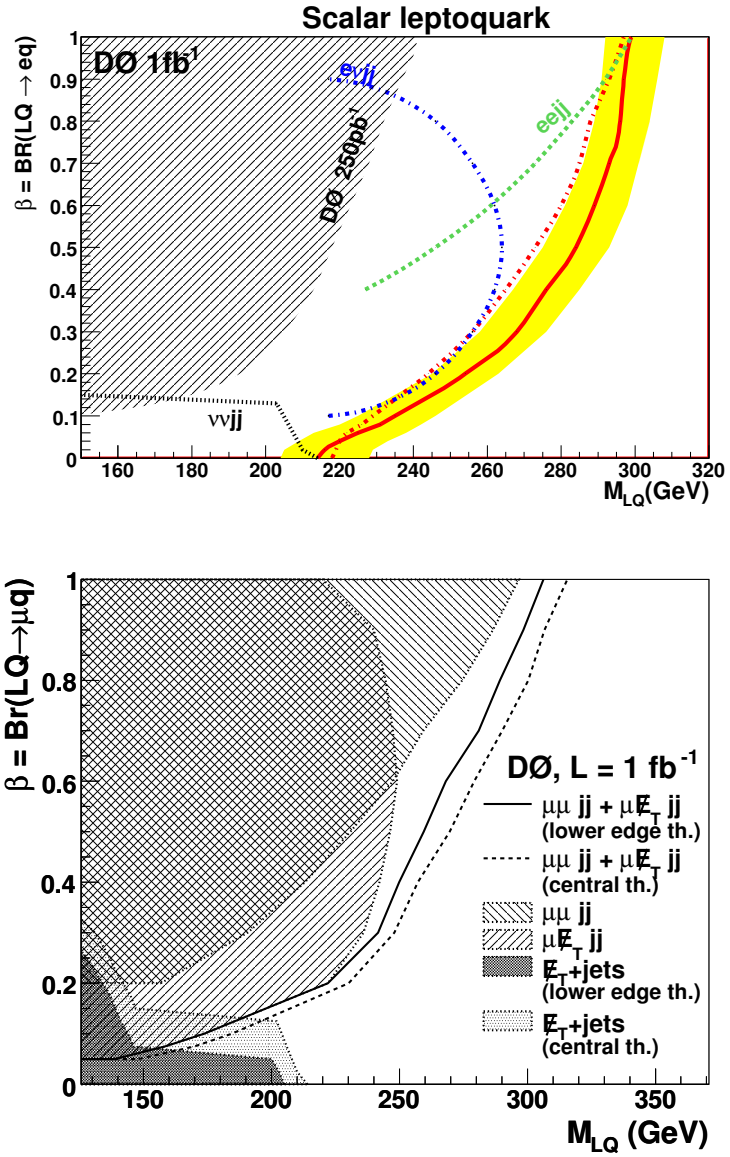

Figure 1.4: The limits on the LQ branching fraction (β) to a charged lepton and a quark, as a function of the LQ mass, for scalar leptoquarks, using 1 fb^{-1} of data collected with the D0 detector. The plot on the top displays limits in on first-generation leptoquarks [33], and the plot on the bottom displays limits on second-generation leptoquarks [34].

1.3 The W+jets Process

The W+jets process is a large contribution to the background for many searches and measurements. The cross section for the W+jets process is several times that of the Z+jets process, and the presence of a neutrino in the final state makes it inherently more difficult to model and to discriminate against using kinematic thresholds. Conversely, some of these properties also make the production of the W boson in association with jets an ideal process for SM measurements as a test of perturbative QCD. In Section 1.3.1, the motivations for studying the W+jets process are introduced. In Section 1.3.2, the current state-of-the-art Monte-Carlo methods and NLO predictions of the W+jets process are discussed. In Section 1.3.3, previous measurements of this process in hadron colliders are summarized.

1.3.1 Motivations for Studying

The W+jets process, as a background to LQ searches, presents unique challenges. In comparison to other SM backgrounds, the W+jets process is more difficult to determine in a fully data-driven manner, and is more difficult to discriminate against by setting kinematic thresholds. For comparison, the $t\bar{t}$ background can be estimated in a fully data-driven manner in the decay modes with two leptons by using a properly reweighted $t\bar{t}$-enriched sample with an electron and a muon in the final state. In final states with just one lepton, the $t\bar{t}$ background still contains two heavy-flavor jets which can be tagged in reconstruction, and their presence can be used to suppress the $t\bar{t}$ contamination, if needed. The Z+jets background is also more easily estimated and removed from the event selection procedure, since the dilepton final state has a sharp peak in the dilepton invariant mass. A simulated Z+jets background sample can be normalized in the Z invariant mass peak, and thresholds can be placed on the invariant

mass to greatly reduce the Z+jets background.

By contrast, the W+jets background is larger and more difficult to reduce than either the Z+jets or $t\bar{t}$ background. The W+jets production cross section is several times larger than that of the Z+jets background. The neutrino in the final state reduces the ability to exploit a sharp mass peak for reduction. The missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) is the reconstructed observable associated with presence of a neutrino in the final state, and the $E_{\mathrm{T}}^{\mathrm{miss}}$ is measured with much lower resolution than that of a lepton. The inability to measure the longitudinal component of the missing energy in pp collisions also means that a transverse substitute for the invariant mass should be used. Compared to an invariant mass, the transverse mass distribution exhibits a broader peak, and thresholds placed above the W mass will remove a smaller fraction of the background than similar cuts placed on the Z invariant mass. Furthermore, the large majority of jets produced in association with the W boson are light-flavor jets, and hence no additional jet properties can be exploited to reduce the W+jets background. Accordingly, the only assessment based on data which can be made on the W+jets background is a normalization in a control region in which the LQ signal is small. The determination of the W+jets background shape must rely on simulation.

The challenges posed by the W+jets background are, however, not unique to exotic searches at high momentum scales and large jet multiplicities. The W+jets processes also provide the main background to rare standard model processes, such as $t\bar{t}$ [35] and single top-quark production [36], and to Higgs boson production. The substantial presence of the W+jets process as a background to searches and measurements provides a very compelling reason to measure the differential cross-sections of the W+jets process at high values of the momentum scales and jet multiplicities.

Furthermore, measurements of the W+jets process are of substantial importance on their own accord. Measurements of the production of vector bosons in association

with jets are fundamental tests of perturbative QCD (pQCD). The perturbative regime is defined by high Q^2, where quarks and gluons interact weakly and behave similarly to free particles. In events where a vector boson is produced in association with jets, the presence of the heavy vector boson ensures a high Q^2, thus allowing for direct and reliable comparisons of W+jets in data with that of predictions from pQCD.

Direct measurements of the W+jets process, differentially as a function of the jet multiplicity and kinematic variables, also allow for an assessment of the reliability of the Monte-Carlo (MC) simulations commonly used for SM predictions. The choice of kinematic variables for which the differential cross sections are measured often include energy and momentum variables, such as the transverse momenta of the jets, and the H_{T}, which is defined as the scalar sum of the p_{T} of all jets, and sometimes also the lepton. It is important to study the distribution of the jet p_{T} and the observable H_{T} because they are sensitive to higher order corrections, and are often used to discriminate against background in searches for signatures of physics beyond the SM. Additionally, H_{T} is commonly used as a component of the factorization and renormalization scale in event generators simulating W+jets production. It is also common to measure angular variables, such as jet rapidities and pseudorapidities, as well as angular separations among the jets, and between the jets and the leptons. It is important to study the agreement between the predicted and measured distributions of these angular variables, because they are sensitive to the modeling of parton emission.

1.3.2 Theoretical Predictions

The production of a W boson in association with n jets in hadron collisions involves one electroweak interaction vertex for the W production and n strong interaction vertices. To the first order, the jet production rate decreases with the number of strong interaction vertices like α_S^n. To date, calculations of the W+jets process have been made at

NLO in QCD for the associated production with jets. These NLO predictions greatly reduce the sensitivity to the factorization and renormalization scale choices, which often result in large uncertainties in leading order calculations. Existing tools for the NLO prediction of the W+jets process include MCFM [37] and BlackHat+Sherpa [38]. The measurements presented in this thesis are compared to BlackHat+Sherpa, for which predictions are available at NLO for parton multiplicities up to four [39].

These NLO calculations represent the current state of the art for predictions of the W+jets processes at hadron colliders. Leading order predictions of the W+jets process are commonly extracted from matrix element (ME) plus parton shower (PS) Monte-Carlo simulations. Such simulations involve a leading-order ME calculation of the underlying hard process with a given final state parton multiplicity, which is interfaced to a PS calculation. Complications arise when considering the ways in which a final state of a given parton multiplicity can come about. For an event with $n+1$ partons in the final state, all of the partons may have resulted from the soft radiation evolution, or they may have evolved from an initial n-parton configuration where a hard, large-angle emission resulted in the presence of an extra parton. To avoid double counting in regions where the ME and PS overlap, a so-called matching scheme is utilized. Common matching schemes are the CKKW and MLM matching schemes, and for a full description of these schemes, the reader is referred to Ref [40].

Common ME+PS MC event generators used at CMS include MadGraph [41] interfaced with Pythia [42], and Sherpa [43–46]. Implementations of MadGraph +Pythia use the MLM matching scheme, whereas Sherpa uses the CKKW matching scheme. It should also be noted that these ME+PS event generators use different underlying models of hadronization. Pythia uses (by default) the string (a.k.a. Lund) method of hadronization [47], whereas Sherpa uses a cluster fragmentation method implemented in the AHADIC++ package, based on Ref [48].

1.3.3 Previous Measurements

Previous measurements of the cross-section of the W+jets process, in leptonic decay modes, have been performed at hadron colliders including the Tevatron and the LHC. Measurements have been performed with $p\overline{p}$ collisions with the D0 detector [49, 50] as a function of jet multiplicity, jet p_T, and a variety of angular variables. Figure 1.5 shows the D0 measurements of the $W \to e\nu$+jets cross section, as a function of jet multiplicity, for up to four jets, using data corresponding to 4.2 fb^{-1}. Results are compared to predictions from ROCKET+MCFM [51, 52]. Measurements have also been made with the CDF detector [53], albeit with a smaller integrated luminosity of 320 pb^{-1}. Figure 1.6 compares measured results from CDF and D0, as a function of the transverse momentum or energy of the three leading jets. CDF measurements are compared to theoretical predictions from pQCD and NLO predictions from MCFM [37]. Measurements from D0 are also made for the scalar sum of the jet transverse momentum and lepton transverse momentum (H_T), for jet multiplicities up to four [50], using data corresponding to 3.7 fb^{-1}, as shown in Figure 1.7.

At the LHC, measurements have been made with pp collisions with an integrated luminosity of 36 pb^{-1} at $\sqrt{s} = 7$ TeV with the ATLAS [56] and CMS [57] detectors. The ATLAS measurements have been produced as a function of several variables, including jet multiplicity and p_T, H_T, and several angular variables. In Figure 1.8, the ATLAS measurements of differential cross section as a function of the leading jet p_T and the H_T are shown for inclusive jet multiplicities up to four jets. These measurements are among the only measurements of the W+jets process measured with LHC data as a function of kinematic variables pertaining to the jets, and are done with only a fraction of the total integrated luminosity delivered by the LHC to date. Measurements have also been made with data from the CMS detector, for the W+n jets cross-section, for $n = 1$–4, where the W boson decays to either an electron or a muon, as a fraction of the

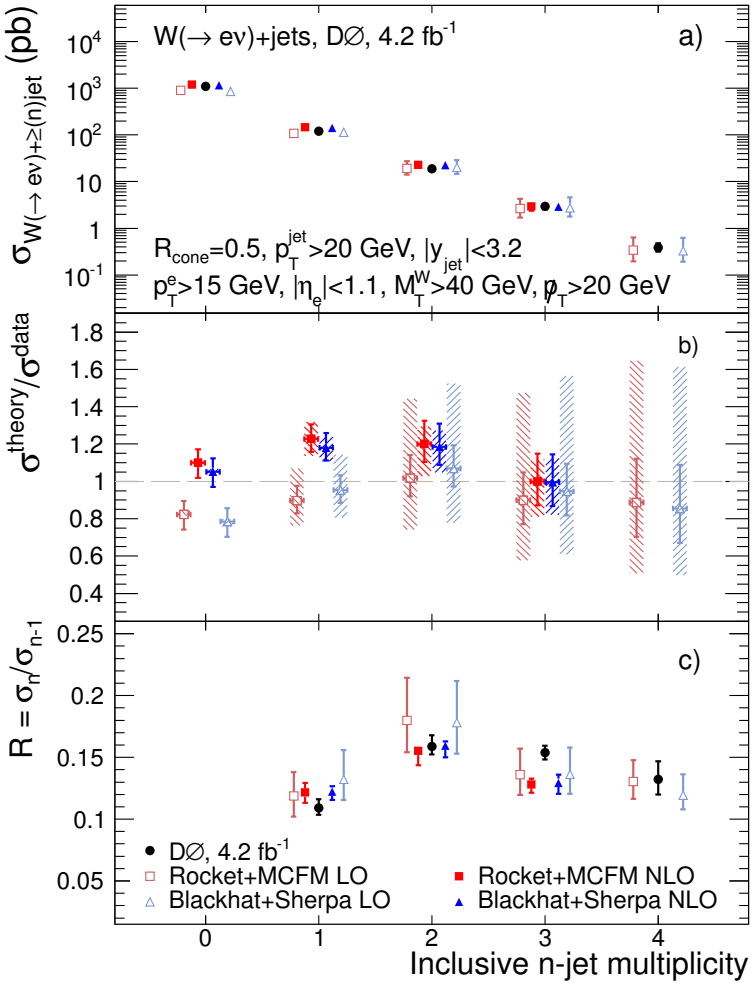

Figure 1.5: The W+jets cross section measurement with 4.2 fb^{-1} of p$\overline{\text{p}}$ collision data with the D0 detector [49], as a function of the inclusive jet multiplicity for jet multiplicities up to four. The measurements are compared to leading order and next-to-leading order predictions from ROCKET+MCFM and BLACKHAT+SHERPA.

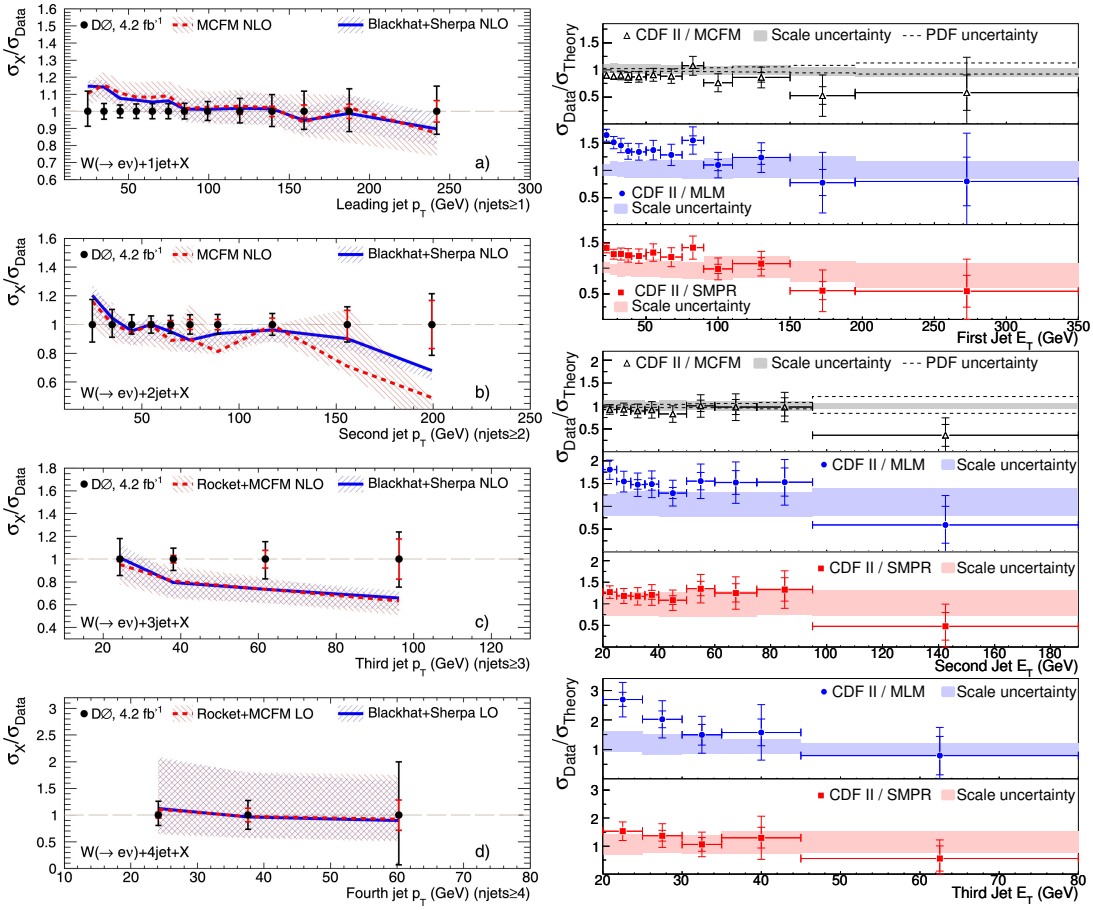

Figure 1.6: Comparisons of measured cross sections of the W+jets processes and theoretical predictions, as a function of the transverse momentum (energy) of the jets, with $p\overline{p}$ collision data from the Tevatron experiment. The left plot shows the comparison between measurements with the D0 detector [49] using 4.2 fb^{-1} of $p\overline{p}$ collision data, with leading order and next-to-leading order predictions from ROCKET+MCFM and BLACKHAT+SHERPA. The right plot shows the comparison between measurements with the CDF detector [53] using 320 pb^{-1} of $p\overline{p}$ collision data, with predictions from pQCD using the MLM [54] and SMPR (CKKW) [55] matching schemes, and from NLO predictions from MCFM [37].

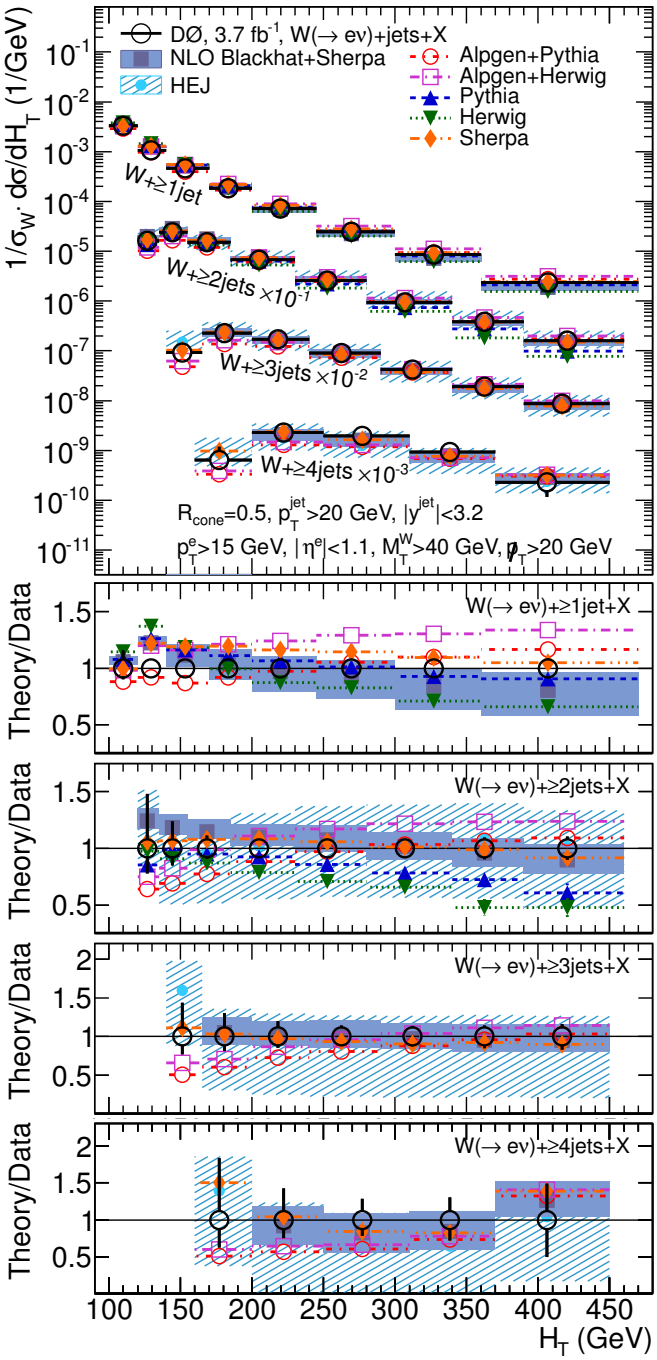

Figure 1.7: The W+jets cross-section measurement with 3.7 fb^{-1} of p$\overline{\text{p}}$ collision data with the D0 detector [49], as a function of the scalar sum of the jet transverse momentum and lepton transverse momentum (H_{T}), for jet multiplicities up to four. The measurements are compared to leading order event generators, and to NLO predictions from BlackHat+Sherpa.

inclusive W boson production cross section. Results are compared to MC generators, including MadGraph and Pythia . These results are shown in Figure 1.9.

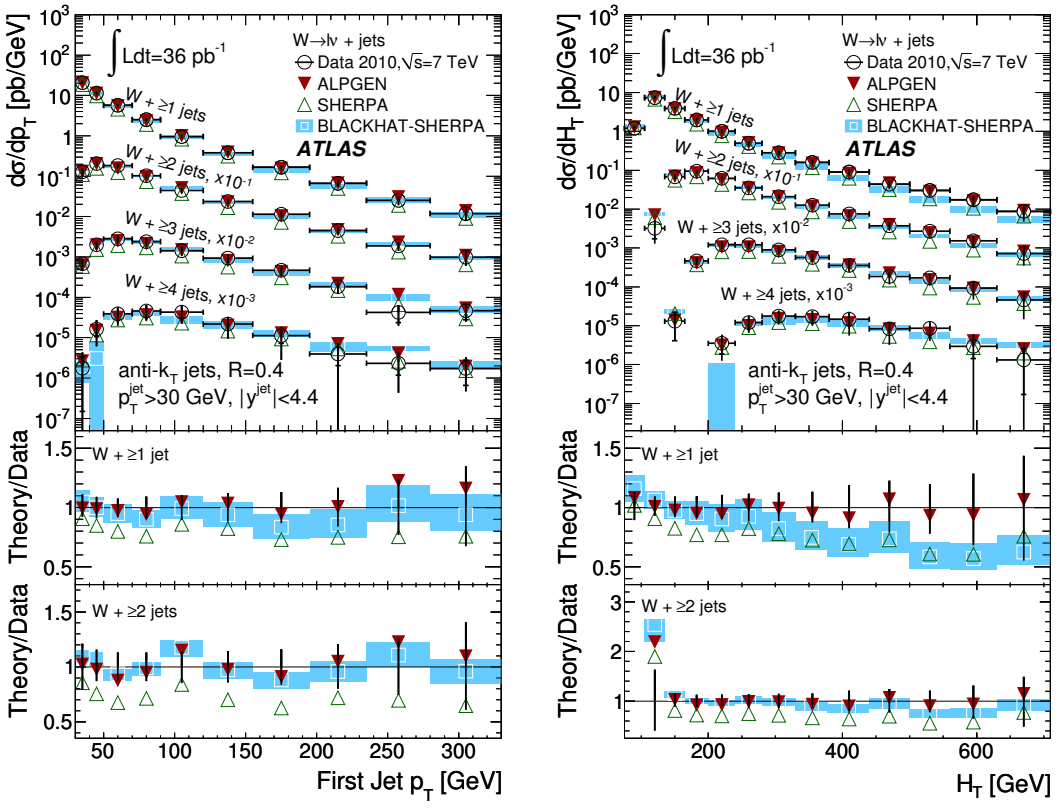

Figure 1.8: Selected ATLAS results for the W+n jets cross-section. The plot on the left shows the cross-section as a function of the p_{T} of the leading jet, for inclusive jet multiplicities up to four. The plot on the right shows the H_{T} for inclusive jet multiplicities up to four, where H_{T} is the scalar sum of the p_{T} of all partons and of the lepton. Results are compared to predictions from MC generators, including Alpgen and Sherpa, and to NLO predictions from BlackHat+Sherpa [56].

Since the CMS and ATLAS data-taking in 2010, which recorded 36 pb^{-1} of collision data, an additional 5 fb^{-1} of data at $\sqrt{s} = 7$ TeV data has been recorded. This data allows for unprecedented precision measurements of the W+jets cross section. With sufficient reduction of the backgrounds, the W+jets cross section can be measured for jet multiplicities up to six jets, and the distributions of the individual jet p_{T} may

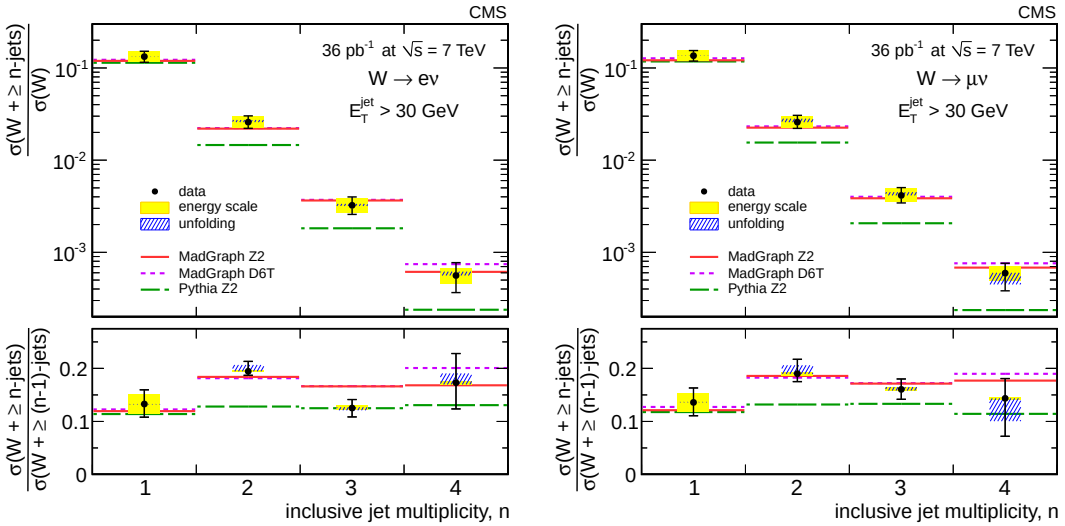

Figure 1.9: CMS measurements of the W+n jets cross-section, for $n = 1$–4, as a fraction of the inclusive W boson production cross section. Measurements were determined from 36 pb^{-1} of pp collision data, and are compared to predictions from MC generators, including MadGraph and Pythia [57].

be measured up to 800 (200) GeV for the first (fourth) leading jet. Additionally, distributions of the H_{T} may be measured up to 1.6 TeV for jet multiplicities up to four jets. This can be compared with previous measurements, which measured jet multiplicities up to four, and p_{T} (H_{T}) distributions up to 350 (700) GeV.

Thus, in addition to the searches for leptoquark signatures to be presented in this thesis, measurements of the differential cross sections of the W boson in association with jets are presented. These measurements are performed in the W $\rightarrow \mu\nu$ decay mode, taking advantage of the high detection and reconstruction efficiency of muons with the CMS detector. Results include measurements of the cross section of the W+n jets process for $n = 1$–6, and measurements of the differential cross section as a function of jet p_{T} and H_{T}, as well as angular variables such as the jet pseudorapidity and the difference in azimuthal angle between each jet and the muon, for the first four leading jets.

Chapter 2

The Experimental Apparatus

This chapter details the Large Hadron Collider (LHC), and the Compact Muon Solenoid (CMS) detector which records data from pp collisions at one of the LHC interaction points. From 2010-2012, the CMS detector recorded pp collision data from LHC running at center-of-mass energies of $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV, with data samples corresponding to 36 pb^{-1} and 5 fb^{-1} with $\sqrt{s} = 7$ TeV, and 20 fb^{-1} with $\sqrt{s} = 8$ TeV. The sections proceed as follows: Section 2.1 describes the LHC apparatus; and Section 2.2 describes the CMS Detector and its sub-systems.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is 26.7 km hadron accelerator and collider producing proton-proton (pp) and lead-ion collisions at multiple interaction points along the accelerator ring. This work will focus solely on pp collisions.

2.1.1 Construction and Layout

The LHC project was approved by the the CERN Council in 1994, and construction of the LHC as a 14 TeV collider was approved in 1996. Plans to construct the LHC at CERN were motivated by the ability to re-use the tunnel constructed for the LEP collider [58]. Protons are accelerated along the 26.7 km main accelerator ring buried beneath parts of France and Switzerland at depths varying from 45 to 170 m.

The layout of the LHC was determined by the geometry of the pre-existing LEP tunnel. Protons produced at the CERN accelerator complex are injected via two 2.5 km transfer tunnels, and undergo acceleration to incrementally higher energies via a series of proton synchrotrons as follows:

- A linear accelerator (LINAC2) generates protons at an energy of 50 MeV.
- The Proton Synchrotron Booster (PSB) accelerates protons to energies of 1.4 GeV.
- The Proton Synchrotron (PS) further accelerates protons to energies of 26 GeV.
- Protons are transferred to the Super Proton Synchrotron (SPS), where they undergo acceleration to 450 GeV.
- In the main ring, protons are accelerated to the energy to be used for collision. Collisions in 2010 and 2011 were performed with 3.5 TeV protons, while collisions in 2012 have been performed with 4 TeV protons.

The LHC injection system, along with the preexisting LEP structure, are illustrated in Fig. 2.1.

Unlike proton-antiproton ($p\overline{p}$) colliders such as the Tevatron, the LHC collides two protons and thereby requires separate rings for the acceleration of the protons

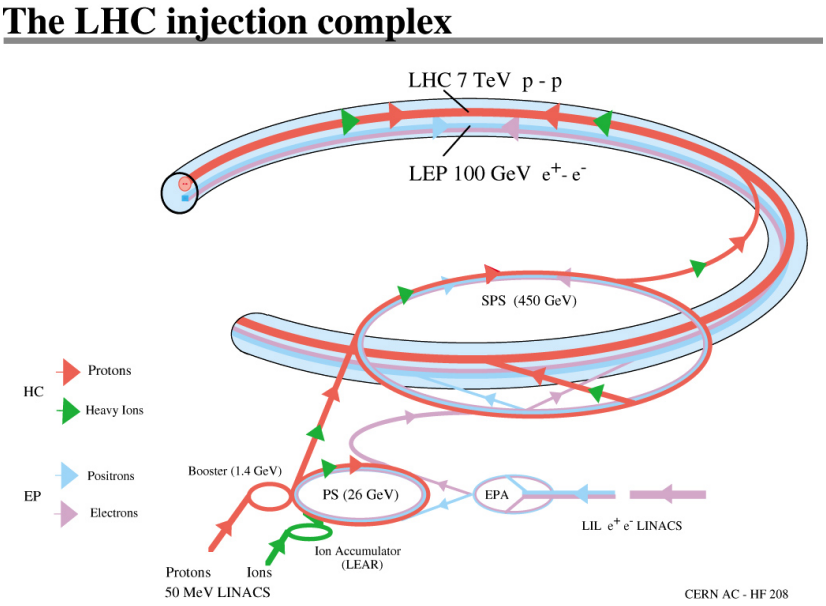

Figure 2.1: A schematic of the LHC injection system, indicating the layout of the linear accelerators (LINACs) injecting protons or ions into an initial accelerating Proton Synchrotron (PS) and Super Proton Synchrotron (SPS), before injecting into the main LHC ring [59].

in opposite directions. To satisfy space constraints in the tunnel and reduce cost, the superconducting magnets which accelerate the protons along the rings were constructed with a twin-bore design, illustrated in Fig 2.2. The 1232 main dipole magnets in the ring accelerate the beam up to an intended design energy of 7 TeV under a magnetic field strength of 8.33 T using a 11.85 kA current. During operation, magnets are maintained at a temperature of 1.9 K using super-fluid pressurized helium. Interspersed with the dipole magnets are 392 quadrupole which serve to focus the beam.

Along the LHC ring are four interaction points housing detectors. The locations of the interaction points are illustrated in Fig 2.3. Two general-purpose detectors are in place: Point 1 houses A Toroidal LHC ApparatuS (ATLAS) [60] and Point 5 houses the Compact Muon Solenoid (CMS) [61] detector. Points 2 and 8 house A Large Ion Collider (ALICE) [62] and the LHC beauty (LHCb) [63] experiments,

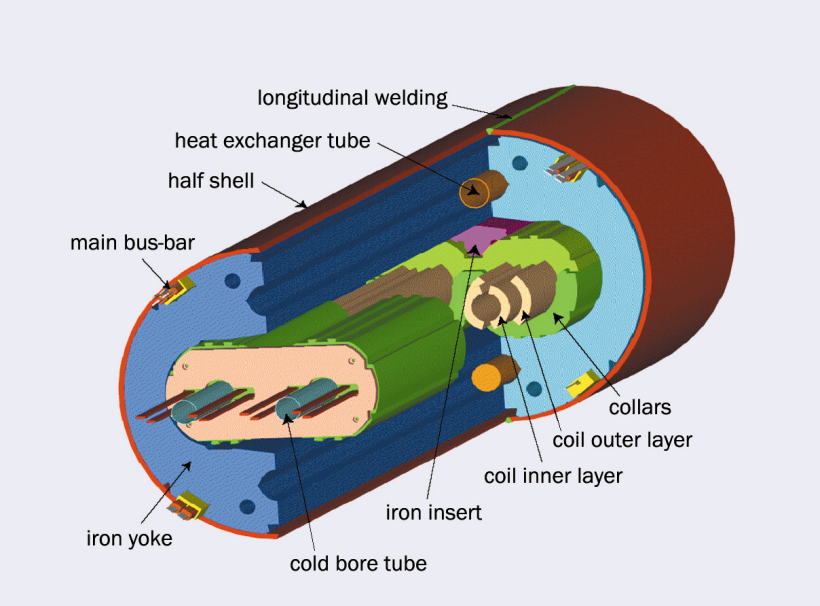

Figure 2.2: A disected view of an LHC diple magnet, displaying the two cold bore tubes use to accelerate protons in opposite directions [58].

respectively. ALICE is optimized to study lead ion collisions at energies of 2.76 TeV. LHCb is optimized for measurements involving the bottom quark.

2.1.2 Design Specifications and Running Parameters

The LHC is designed to deliver pp collisions with unprecedented energies and luminosities. The luminosity ($\mathcal{L}$) defined in Equation 2.1, gives a measure of the number of collisions a collider is able to produce per unit time per unit area. Here, dN/dt is the number of collisions per unit time, and σ represents the process cross-section.

$$\mathcal{L} = \frac{dN/dt}{\sigma} \tag{2.1}$$

For collisions of two beams, $\mathcal{L}$ can is expressed as a function of the beam parameters

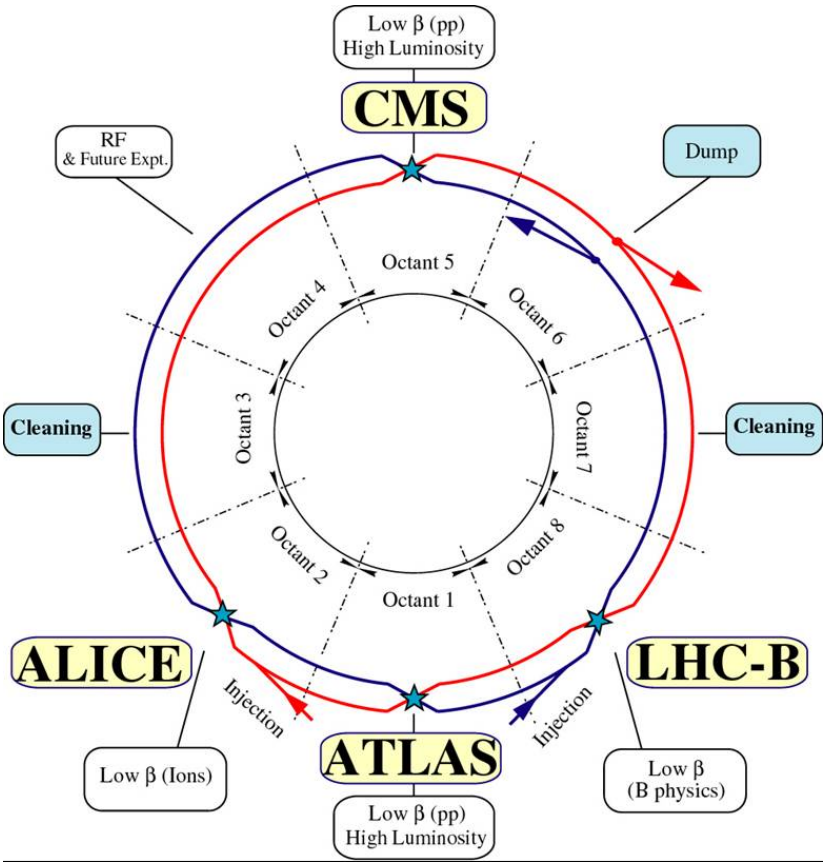

Figure 2.3: A schematic view of the LHC ring, divided into eight octant arcs with four crossings housing the ALICE, ATLAS, LHC-B, and CMS detectors [64].

as in Equation 2.2

$$\mathcal{L} = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi \epsilon_n \beta^*} \left(1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*}\right)^2\right)^{-1/2} \tag{2.2}$$

The parameters in Equation 2.2 are defined in Table 2.1, along with the values corresponding to the design specifications of the LHC. The quantity $\left(1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*}\right)^2\right)^{-1/2}$ can also be generalized as the geometric luminosity reduction factor (F), which accounts for the presence of a crossing angle at the interaction point (IP).

The maximum luminosity achieved over the 2010-2012 running period of the LHC has been 7.7×10^{33} cm^{-2}s^{-1} . This can be compared with the design specification for

	Definition	LHC Design Value	Achieved in 2012
N_b	Number of particles per bunch	1.15×10^{11}	1.7×10^{11}
n_b	number of bunches per beam	2808	1374
f_{rev}	Revolution frequency	11.25 kHz	11.25 kHz
γ_r	The Lorentz factor	7461	4264
ϵ_n	Norm. transverse beam emittance	3.75×10^{-6} m	2.4×10^{-6} m
β^*	Envelope function at IP	0.55 m	0.6 m
θ_c	Full crossing angle at IP	285 μrad	290 μrad
σ_z	RMS bunch length	7.55×10^{-2} m	9×10^{-2} m
σ^*	Transverse RMS beam size at IP	16.7×10^{-6} m	18.8×10^{-6} m
$\mathcal{L}$	Total instantaneous luminosity	10^{34} cm$^2 s$	7.7×10^{33} cm$^2 s$

Table 2.1: The parameters which determine the luminosity ($\mathcal{L}$) in Equation 2.2.

the peak luminosity at $\sqrt{s} = 8$ TeV of $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1} . The total integrated luminosities delivered to, and recorded by the CMS detector, and the peak instantaneous luminosities by day in 2011 and 2012, are shown in Figure 2.4.

2.2 The Compact Muon Solenoid Detector

One collision interaction point of the LHC houses the Compact Muon Solenoid (CMS) detector, a general purpose detector used for a variety of searches and measurements. This Section describes the CMS detector and the CMS subsystems which are optimized for the detection of different types of particles. An expanded description of the endcap muon cathode strip chamber (CSC) calibrations is given in Section 2.2.7, since the author of this thesis was responsible for this work.

2.2.1 Coordinates and Conventions

As a convention, the geometry of the CMS experiment can be described with a right-handed coordinate system, defined by an origin at the nominal interaction point. The x and y axes are taken as perpendicular to the beam direction, with the x axis pointing

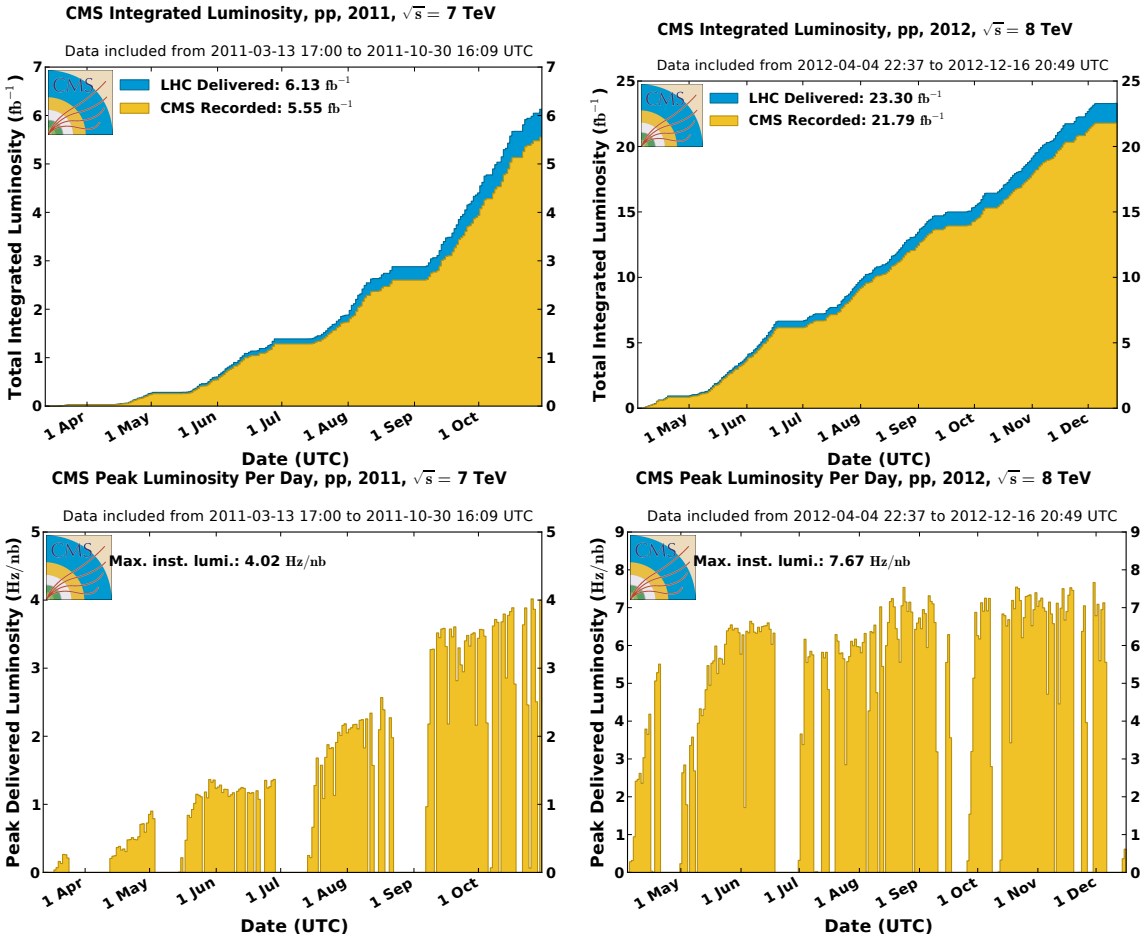

Figure 2.4: The total integrated luminosities delivered to and recorded by the CMS detector in 2011 (top left) and 2012 (top right), and the peak instantaneous luminosities by day in 2011 (bottom left) and 2012 (bottom right).

to the center of the LHC ring, and the y axis pointing upward. The z axis points in the direction of the counter-clockwise beam. Hence, a polar angle θ is taken with respect to the positive z axis, and the azimuthal angle ϕ is measured in the x-y plane perpendicular to the beam direction.

In collider physics, instead of the polar angle θ, the rapidity (φ) and pseudorapidity (η) are often used. These quantities are defined in Equation 2.3 and Equation 2.4. The rapidity is a useful quantity because the production of particles is nearly uniform as a function of φ, and the difference in rapidity between two particles ejected from a col-

lision is invariant under Lorentz transformations along the z-axis. The psuedorapidity is an approximation to the rapidity which relies only the polar angle θ. In practice, it can be easier to determine the pseudorapidity than the rapidity, because detectors are often incapable of making precise measurements of particles close to the beam - i.e. particles with large p_z.

$$\varphi = \frac{1}{2} \ln \left(\frac{E + cp_z}{E - cp_z} \right) \tag{2.3}$$

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right) \tag{2.4}$$

Searches and measurements in particle physics often make use of kinematics in the plane transverse to the beam direction, in which momentum conservation can be applied and the momentum can be most reliably determined. Like η, transverse momenta (p_{T}) are also invariant under longitudinal relativistic transformations.

2.2.2 Overview of the Detector

The CMS detector is located at Point 5 along the LHC ring, between Lake Geneva and the Jura Mountains near Cessy, France. It is designed to operate with pp collisions at $\sqrt{s} = 14\,\mathrm{TeV}$ and luminosities of $10^{34}\,\mathrm{cm^2 s^{-1}}$. While the total inelastic event rate under design conditions is approximately 1 billion events per second, limitations on computational and storage capabilities necessitate a reduction in event rate to about 100 events per second, which is accomplished with a so-called "trigger" system, discussed in Section 2.2.8.

The bunching of protons and small time-distances between bunches (25–50 ns) results in a phenomenon called pileup (PU), or the presence of multiple inelastic collisions in a given event. This results in the presence of thousands of charged particles in each event. During the 2012 running of the LHC, the CMS detector recorded an average of

21 collision events per bunch crossing. The distribution of the number of interactions per bunch crossing is shown in Figure 2.5.

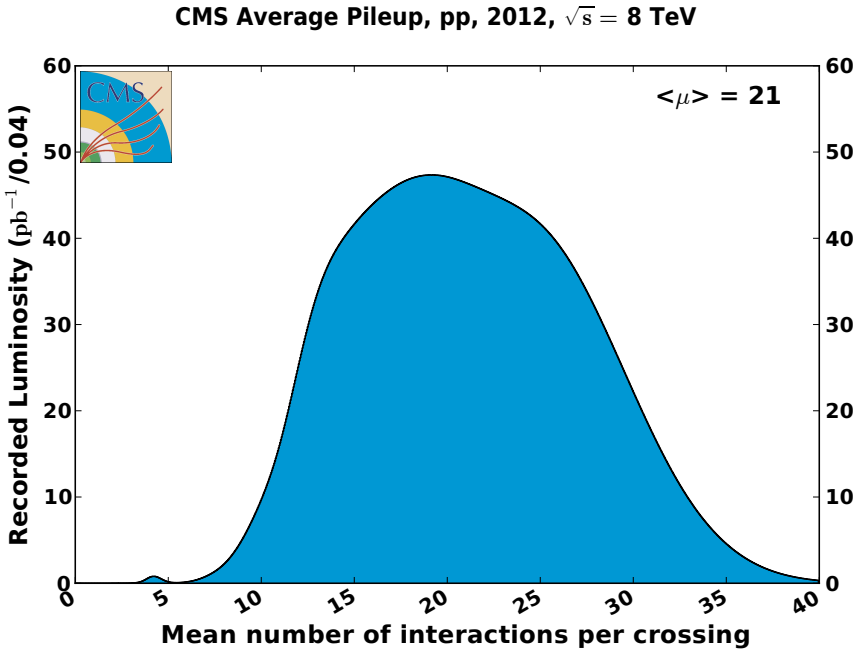

Figure 2.5: A histogram of the number of interactions per bunch crossing, recorded with the CMS detector over the course of the full 2012 running period.

Additional consideration must be given to the response time of the detector components, which can be larger than time between bunches. The goal is to accurately determine which particles correspond to which event and which collision vertex within the event, and to accurately reconstruct the charge and momenta of the particles. Namely, the particle reconstruction goals include: reliable muon, electron, and photon identification and momentum resolution, resulting in dilepton mass resolutions of approximately 1% near the mass of the Z boson; near-perfect muon charge determination for muons with $p_{\mathrm{T}} < 1\,\mathrm{TeV}$; good resolution for charged particles in the inner tracker; efficient tagging and triggering for τ leptons and jets resulting from b-quark decays; efficient photon and lepton isolation; and good estimation of missing transverse energy.

The CMS detector is described in terms of its main components (subsystems) in Sections 2.2.3 through 2.2.8. The basic layout of the CMS detector subsystems is illustrated in Figure 2.6.

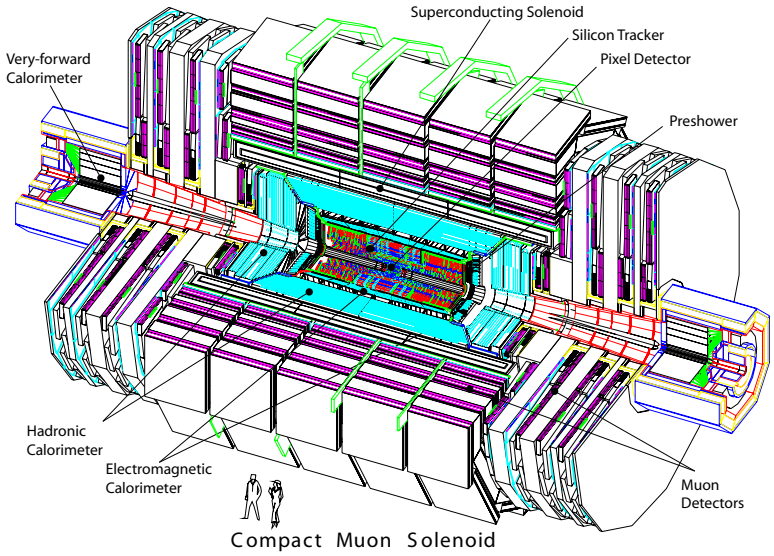

Figure 2.6: A cutout perspective of the CMS detector, illustrating the locations of the superconducting solenoid, calorimeters, pixel detectors, and muon system [65].

2.2.3 The Superconducting Magnet

The superconducting magnet is built around a coil that is 6 m in diameter and 12.5 m long, and was designed to produce a 4.0 T magnetic field and store 2.6 GJ of energy. The magnet produces a remarkably uniform field for its volume. The coil, constructed from niobium-titanium (NbTi), consists of 41.7 MA-turns and is wound in four layers. The winding itself took five years to complete, and was completed in 2005, and the magnet was successfully tested at an operating temperature of 4.5 K with cosmic-ray observations in 2006.

The solenoid consists of five separate modules, and the cold mass is 220 tons. This

results in a stored energy-to-mass (E/M) ratio of 11.6 kJ/kg, which gives mechanical deformations of approximately 0.15%. This is considerably larger than magnets used in previous experiments, as depicted in Figure 2.7. The flux is returned through an in iron return yoke weighing approximately 10000 tons. Five barrel wheels and 6 endcap disks comprise the yoke, and this design allows for easier assembly of the subdetectors.

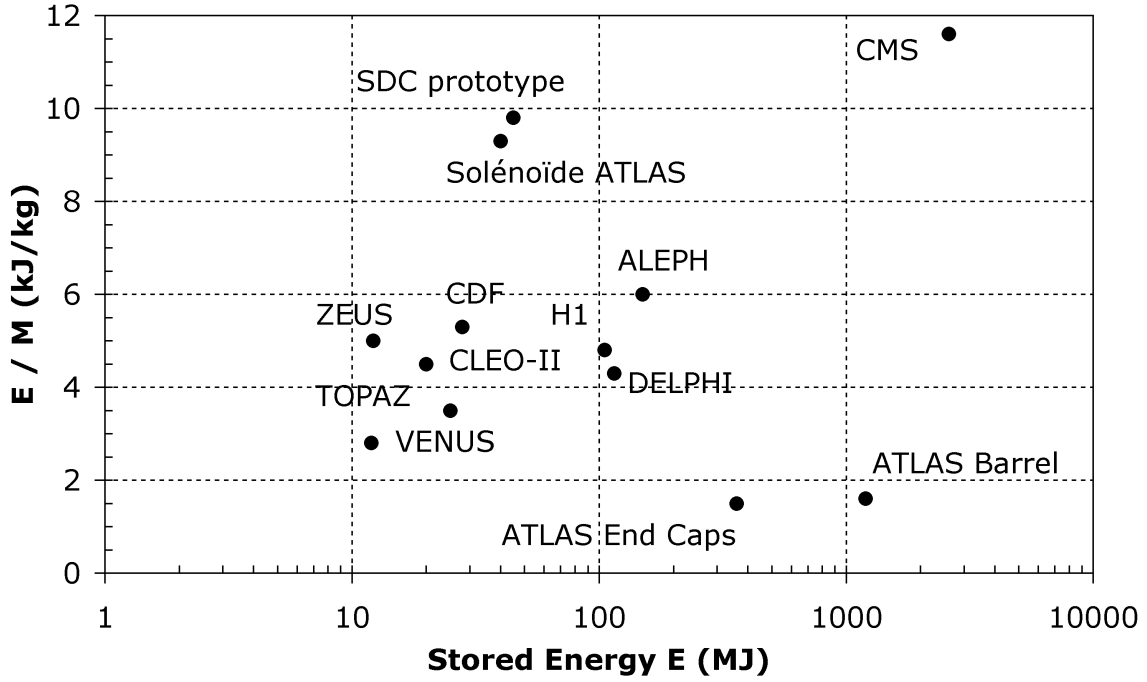

Figure 2.7: The energy-to-mass ratio for the CMS superconducting solenoid, compared to that of other experiments [61].

While the magnet was designed to operate at 4 T, in practice it has operated at 3.8 T to increase longevity. The operating parameters, including field strength, stored energy, and current, are compared to the design specifications in Table 2.2. Table 2.3 details the properties of the magnet, including the design parameters and dimensions of the cold mass and the iron yoke. A schematic of the magnetic field magnitude and direction is given in Figure 2.8 for the 2010-2012 operational value of 3.8 T.

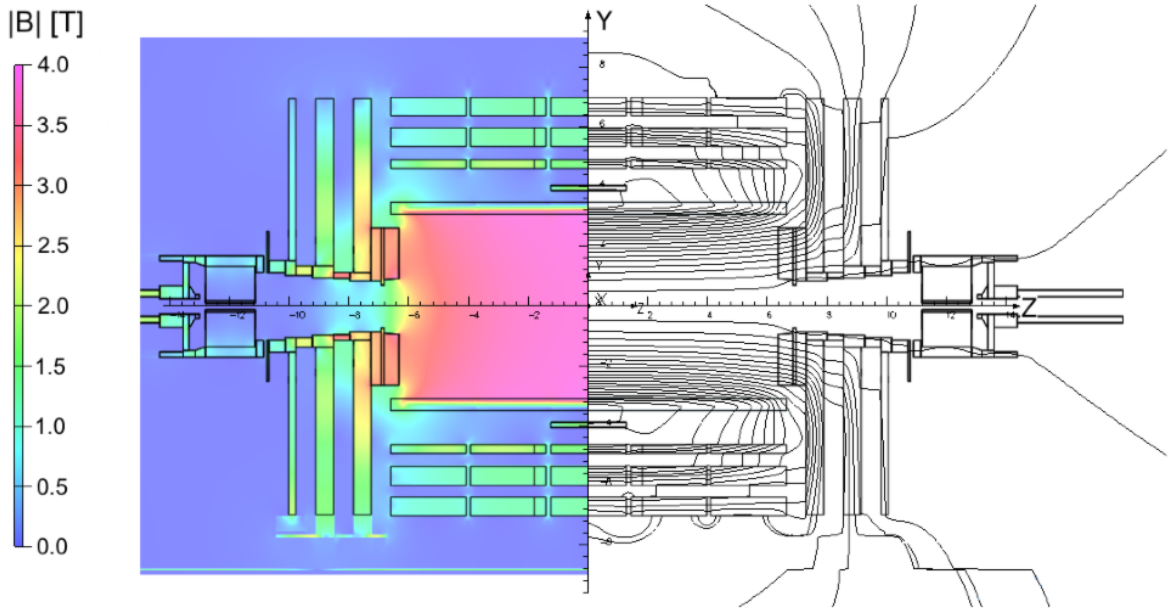

Figure 2.8: The absolute value of the magnetic field $|B|$ (left side) , and the field lines depicting the B-field direction (right side). [66].

Value	Design Specification	Used in Operation
Current (kA)	19.50	18.16
Energy storage (GJ)	2.6	2.3
Field strength (T)	4.0	3.8

Table 2.2: The design specifications of the magnet, compared to the operational values used in running from 2010-2012 [61].

2.2.4 The Silicon Tracker

The innermost layer of the CMS detector consists of a silicon tracker capable of determining the locations of charged particles with 10 μm precision in three dimensions, allowing for accurate reconstruction of charged particle tracks. It is the largest silicon detector ever constructed, with over 205 m^2 of sensors. The inner pixel detector consists of 66 million pixels, and the outer layer consists of 9.6 million strip channels. The tracker was designed to work under operating conditions with approximately 1000 particles originating from 20 overlapping collisions in bunches spaced 25 ns apart, which

General parameters	
Magnetic length	12.5 m
Cold bore diameter	6.3 m
Central magnetic induction	4T
Total Ampere-turns	41.7 MA-turns
Nominal current	19.14 kA
Inductance	14.2 H
Stored energy	2GJ
Cold mass	
Layout	Five modules mechanically and electrically coupled
Radial thickness of cold mass	312 mm
Radiation thickness of cold mass	3.9 X_0
Weight of cold mass	220 t
Maximum induction on conductor	4.6 T
Temperature margin w.r.t. operating temperature	1.8 K
Stored energy/unit cold mass	11.6 kJ/kg
Iron yoke	
Outer diameter of the iron flats	14 m
Length of barrel	13 m
Thickness of the iron layers in barrel	300, 630 and 630 mm
Mass of iron in barrel	6000 t
Thickness of iron disks in endcaps	250, 600 and 600 mm
Mass of iron in each endcap	2000 t
Total mass of iron in return yoke	10000 t

Table 2.3: The properties of the magnet, including the design parameters and dimensions of the cold mass and the iron yoke [61].

requires remarkable speed and granularity.

The tracker consists of four major parts, illustrated in Figure 2.9. The innermost pixel detector (PIXEL) covers approximately $1\,\mathrm{m}^2$, and is composed of three cylindrical barrel layers (BPix) at radii of 4.4, 7.3, and 10.2 cm and two endcap discs (FPix) at $z = \pm 34.5\,\mathrm{cm}$ and $z = \pm 46.5\,\mathrm{cm}$. The FPix cover the radial dimension from approximately 6 to 15 cm. The positions of the BPix and FPix and their η coverage are illustrated in Figure 2.10. Matching the acceptance of the central tracker, the pixel

detectors provide $|\eta|$ coverage up to 2.5.

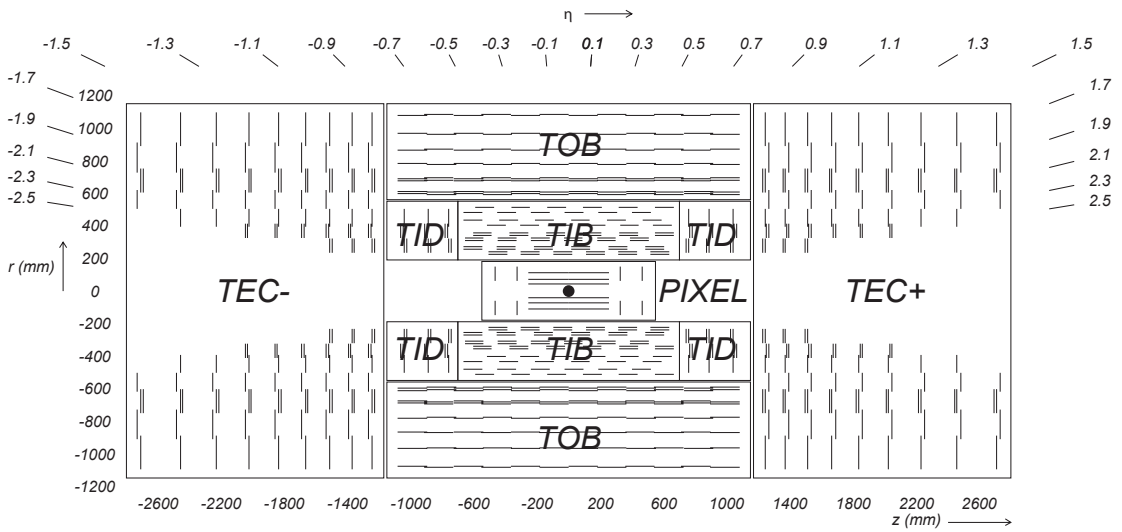

Figure 2.9: A longitudinal view of the CMS tracker, with single lines indicating detector modules and double-lines indicating back-to-back detector modules [61].

The silicon strip tracker surrounds the pixel detector, and is composed of three subsystems. Directly surrounding the pixel detectors is the Tracker Inner Barrel (TIB), which consists of four concentric cylinders at inner radii of 25.50, 33.90, 41.85, and 49.80 cm, each of which extends to ± 70 cm in the z direction. The Tracker Inner Disk (TID) extends the coverage of the TIB, and consists of three disks on each end. Together, the TIB and TID extend to a outer radius of 55 cm and to ± 118 cm in the z direction. They deliver up to four measurements of (r,ϕ) position for a single particle direction. The micro-strip sensors which comprise the TIB and TID are 320 μm thick, and extend parallel to the z axis on the TIB and in the radial direction on the TID. In the TIB, the pitch of the micro-strips is 80 μm (120 μm) in the inner (outer) two layers. The TID strips have variable pitch, with a minimum of 100 μm and a maximum of 141 μm.

The Tracker Outer Barrel (TOB) surrounds the TIB and TID, and extends to a

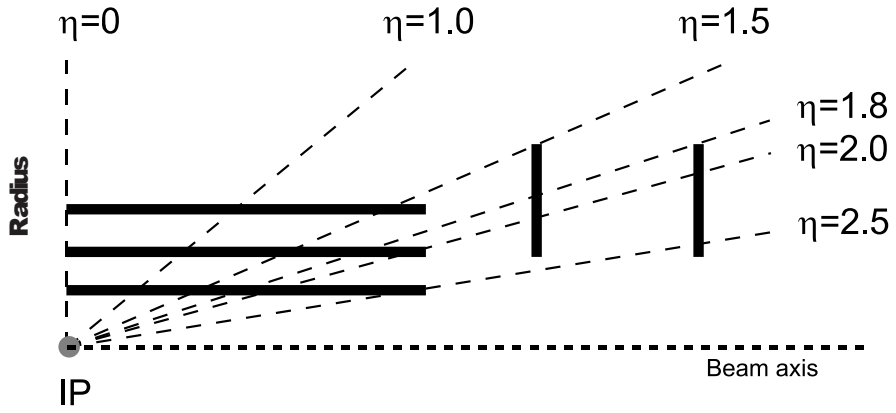

Figure 2.10: The pseudorapidity coverage of the barrel pixel layers (BPix), displayed as solid black horizontal bars, and of the endcap discs (FPix) displayed as solid black vertical bars [61].

radius of 116 cm and to ±118 cm in the z direction. The 6 layers of the TOB provide an additional 6 measurements of the (r,ϕ) position of a particle. The barrel layers of the TOB have micro-strip sensors with a thickness of 500 μm and a pitch of 183 μm (122 μm) in the first four (outer two) layers.

The outermost subsystem in the z direction is the Tracker EndCaps (TECs), which can be referred to separately for the endcap on the $+z$ side (TEC+) and on the $-z$ side (TEC−). The TEC+ and TEC− are symmetrically constructed, with each consisting of nine disks with an inner radius decreasing with the $|z|$ position, providing up to 9 measurements of the (r,ϕ) position of a particle. The TEC disks consists of up to 7 rings of radial micro-strip sensors with widths of 230 μm to 500 μm and a pitch of 97 μm to 184 μm. The TEC disks extend from 124 cm to 282 cm in the $|z|$ direction, with a radial coverage beginning at 22.5 cm for the inner TEC disks, and an outer radius of 113.5 cm.

The strip detector subsystems are capable of measuring the z coordinate in the barrel and the r coordinate in the disks. This is done with an additional layer of silicon sensors with a small stereo angle rotation of 1000 mrad, and is present on a

subset of the rings and layers of the detectors. These rings and layers are as follows: layers 1 and 2 of the TIB and TOB; rings 1 and 2 of the TIDs; rings 1, 2, and 5 of the TECs. For most pseudorapidities, this guarantees at least 9 total hits in the tracker, and at least four two-dimensional measurements from the stereo layers, as illustrated in Figure 2.11. The tracker construction allows for accurate reconstruction of charged particle tracks and high-resolution measurements of interaction vertices, which are further detailed in Section 3.1.

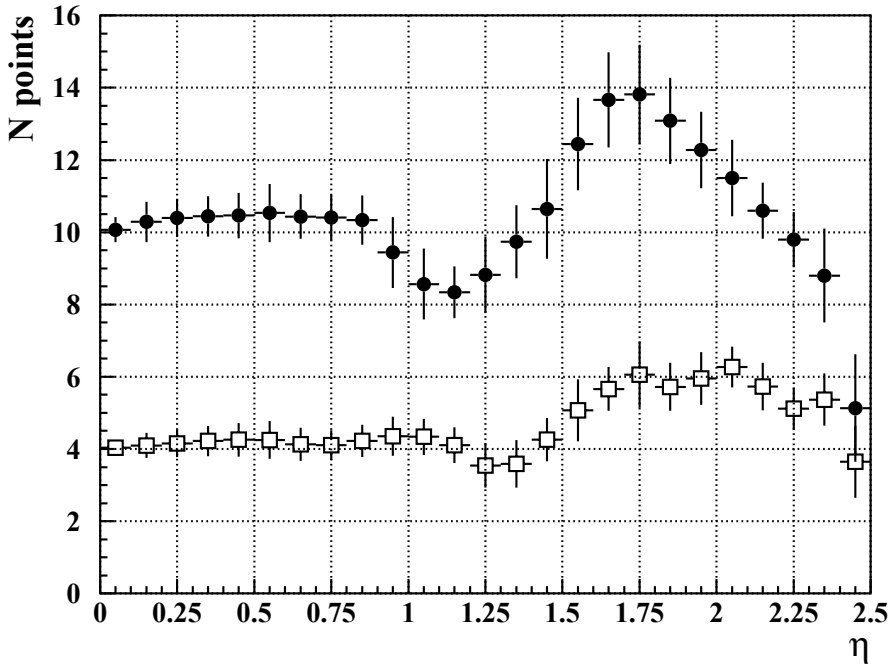

Figure 2.11: The number of measured hits in the strip tracker versus η, with solid markers representing the total, and open markers representing two-dimensional measurements in stereo layers. [61].

2.2.5 The Electromagnetic Calorimeter

Immediately surrounding the silicon tracker is the electromagnetic calorimeter (ECAL), which is designed for high-precision measurements of electron and photon energies. The ECAL consists of 75,848 $PbWO_4$ (lead tungstate) crystals, with 61,200 crystals in the

barrel (EB) region, and 7324 in each endcap (EE). The $PbWO_4$ is extremely dense and able to stop high energy particles. The crystalline structure also has the capability of scintillating in fast bursts which is suitably quick for operation at the LHC, where time between colliding bunches is 25-50 ns depending on running conditions. In addition to the barrel and encaps, a "preshower" detector (ES) also surrounds the ECAL endcaps, and is designed to be able to distinguish between single high-energy photons and pairs of spatially close low-energy photons.

The choice of $PbWO_4$ was motivated by the operating conditions and goals of the detector. Lead tungstate was chosen because:

- The density, $8.25\,\mathrm{g/cm^3}$, is higher than that of stainless steal, making it ideal for stopping high-energy particles and allowing for the construction of a compact detector.

- The radiation length, defined as the mean path length required to reduce the energy (via bremsstrahlung) of a relativistic charged particle by a factor of $1/e$, is short (0.89 cm). This allows for a compact design.

- The Molière radius, defined as the radius of a cylinder transverse to a charged particle's direction of motion in which 90% of the particle's shower energy is deposited, is small (2.2 cm). This allows for high granularity and spatial resolution, which is especially important for detecting Higgs boson decays to two photons.

- The scintillation decay time is short, with 80% of light emitting in less than 25 ns, the design bunch-spacing of the LHC.

The EB consists of crystals beginning at an inner radius of 1.29 m, and has $|\eta|$ coverage up to 1.479. The EE extend the ECAL η coverage up to $|\eta| = 3.0$, and are in the plane transverse to the beam axis at 315.4 cm from the interaction point.

A schematic of the ECAL, including the layout of the EB, EE, and ES, is given in Figure 2.12.

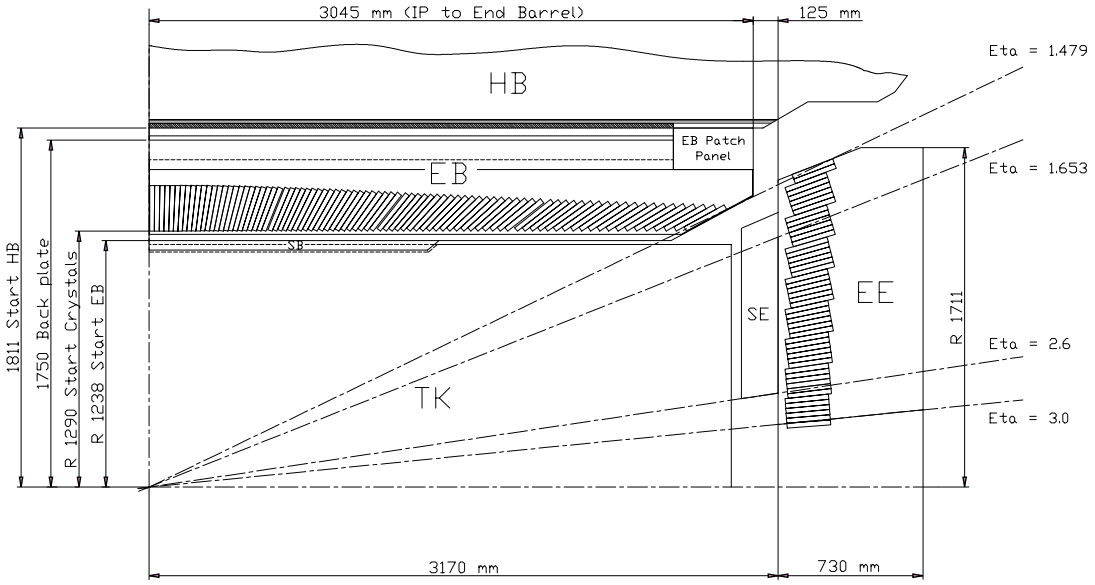

Figure 2.12: A schematic of the main components of the ECAL, including the barrel (EB), endcap (EE), and the preshower (SE, or ES) [67].

In the EB, Avalanche photodiodes (APD) are used as photodetectors, whereas vacuum phototriodes (VPT) are used in the EE. The APDs and VPTs are fast and tolerant to radiation, and able to operate in the high (4 T) magnetic field of the CMS detector. Since the $PbWO_4$ crystals output relatively small amounts of light, strong amplification is needed in the APD and VPTs. The APDs cover $5 \times 5\,\mathrm{mm}^2$. A pair are mounted on each EB crystal. The APDs were screened to ensure operation for 10 years under high luminosity. The VPTs are single-gain photomultipliers covered in a copper mesh for operation in the magnetic field. One VPT is affixed to each EE crystal. The VPTs were tested to ensure functionality at a wide range of angles with respect to the CMS magnetic field.

2.2.6 The Hadron Calorimeter

The CMS hadron calorimeter (HCAL) is designed to measure the energies of hadron jets and to infer the apparent missing transverse energy in an event. The HCAL subsystems include the barrel calorimeter (HB), the endcap calorimeter (HE), the outer calorimeter (HO), and the forward calorimeter (HF). The layout of the HCAL subsystems is illustrated in Figure 2.13.

The HB design is restricted to the region between the ECAL and the magnet, $1.77\,\mathrm{m} < r < 2.95\,\mathrm{m}$. The HB is a sampling calorimeter which consists of 36 wedges divided into two half-barrels, indicated by their z position relative to the interaction point as HB+ and HB−. Each wedge is identical and spans the full range $0 < \eta < 1.3$. The wedges are made from flat brass absorber plates parallel to the beam axis but staggered to improve coverage losses due to dead material. To increase structural strength, the inner and outer plates are made from stainless steel. The plastic scintillator is divided into 16 sectors which cover an area defined by an η and ϕ width of 0.087. The active scintillating medium uses a tile and wavelength shifting (WLS) fiber method. Light from tiles are collected with WLS fibers situated in grooves of the scintillator, which are spliced to clear fibers connected to optical cables taking the light to an optical decoding unit.

The HE extend the coverage provided by the HB, including the range of $1.3 < |\eta| < 3.0$, and must be able to endure the CMS operating conditions, including heavy radiation doses and large magnetic field. The HE brass absorbers have a staggered structure to avoid areas of dead material. Most of the HE towers have 16 layers, with the exception of tower 18 which has an additional inner layer which improves coverage in the η gap region where photodetectors and front-end electronics are installed. As with the HB, the light is collected with WLS fibers in a tray structure.

The HB and HE are segmented in η for readout in a tower structure, illustrated in

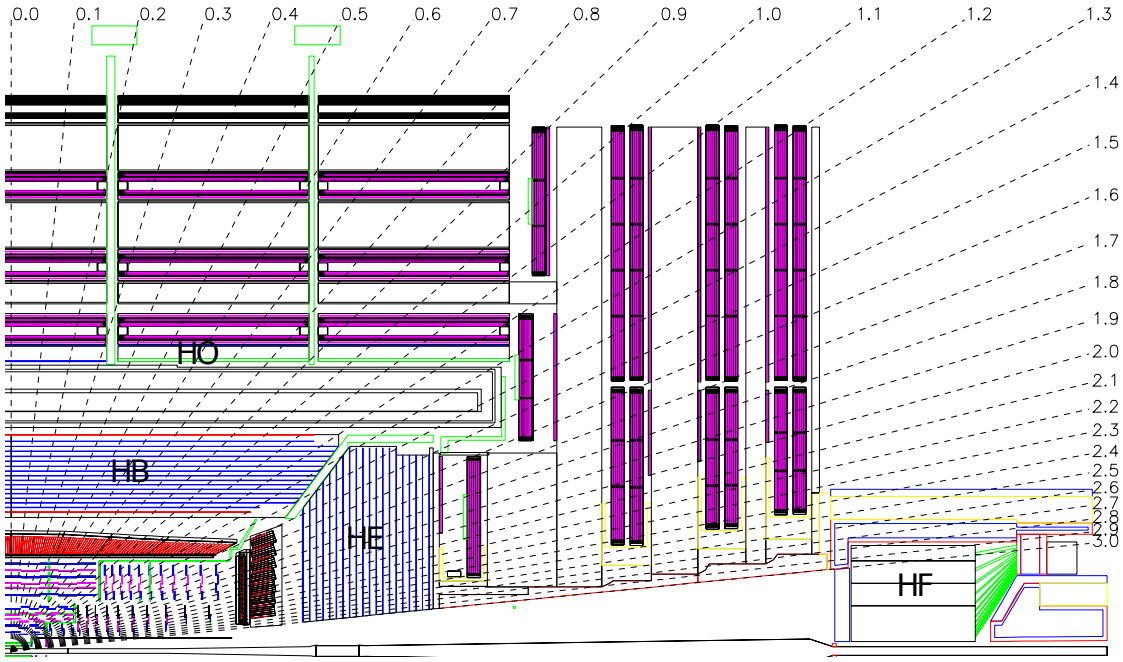

Figure 2.13: A quarter section schematic of the CMS detector indicating the positions of the HCAL subsystems, including the barrel (HB), the endcap (HE), the outer (HO), and the forward (HF) calorimeters [61].

Figure 2.14.

The HCAL Outer calorimeter (HO) captures the majority of hadron showers not contained by the EB and HB. The HO extends outside the solenoid, and also acts as an absorber. The HO is placed as the first sensitive layer of each of the five rings of the iron yoke (see Section 2.2.3). The HO layers are segmented according to the geometry of the muon system. Each ring has 12 ϕ divisions (sectors), separated by stainless steel beams composing the infrastructure supporting the muon system and magnet return yoke.

The Forward calorimeter (HF) is a cylindrical steel structure located 11.2 m from the nominal interaction point of the CMS detector, and spans the radial dimensions from $12.5 < r < 130$ cm. The HF is divided into 36 wedges azimuthally (18 on the $+z$ side and 18 on the $-z$ side). The calorimeter is divided into two longitudinal segments

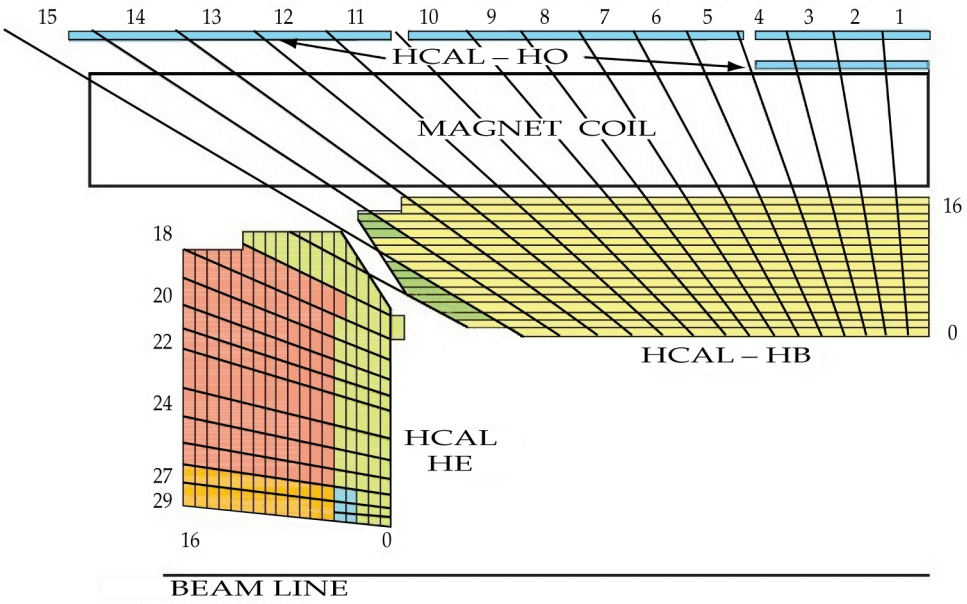

Figure 2.14: The segmented tower structure of the HB, HE, and HO of the ECAL. Coloring indicates optical grouping for readout. [61].

which read out separately, allowing the discrimination between photon showers and electron showers. The high radiation environment of the forward region, with deposits over 700 GeV per pp interaction and an expected 10-year radiation dose of 10 MGy, demands exceptional shielding of the calorimeter. The radiation shielding consists of 40 cm of steel, 40 cm of concrete, and 5 cm of polyethylene. The detector is housed in this shield structure. Bundled fibers from the absorber deliver signals to read-out boxes housing photomultiplier tubes with steel/polystyrene shielding.

2.2.7 The Muon System

Precise and robust detection of muons is critically important for CMS, as several signatures of Higgs decays (e.g. to to ZZ$\rightarrow$ 4μ) need to be discriminated from large backgrounds, and many signatures of supersymmetry and exotic processes involve one or more leptons. Good muon resolution also aids in precision measurements, espe-

cially those involving differential cross sections in which final-state kinematics are to be measured.

The CMS muon system is divided into three subdetectors for optimal muon detection in the barrel, in the endcap, and for triggering. The muon system consists of drift tubes with rectangular drift cells in the barrel region ($|\eta| < 1.2$), cathode strip chambers in the endcap region ($0.9 < |\eta| < 2.4$), and resistive plate chambers in the $|\eta| < 1.6$ region with fast, high-resolution determination of muon p_{T} for triggering purposes. The design of the muon system takes into consideration the large background from neutrons generated in collisions and captured in detector material. Such interactions between neutrons and nuclei lead to the emission of photons, which generate electrons that are detected as hits in muons chambers. This process, which is dominant in the endcap region, is mitigated by the use of the multi-layer cathode strip chambers. The layout of the muon subsystems in a quadrant of the CMS detector is illustrated in Figure 2.15.

The Drift Tube Chambers

The drift tube (DT) chambers provide coverage for muon detection in the $|\eta| < 1.2$ region, where the magnetic field is approximately uniform at 0.4 T and the neutron background is small [66]. The chambers consist of rectangular drift cells, and they are divided into 4 radial stations (labeled MB1–4) in between layers of the magnetic flux return plates. Of these four stations, the three inner stations contain 60 chambers each, while the final station contains 70 chambers. The DTs are divided into 5 "wheels", labeled by their relative position in the z direction as W–2, W–1, W0, W+1, and W+2. The layout of the DT system is illustrated in Figure 2.16.

The DT chambers consist of rectangular drift cells. The DTs contain approximately 172000 wires. The drift cells are are 21 mm in length in the transverse direction where

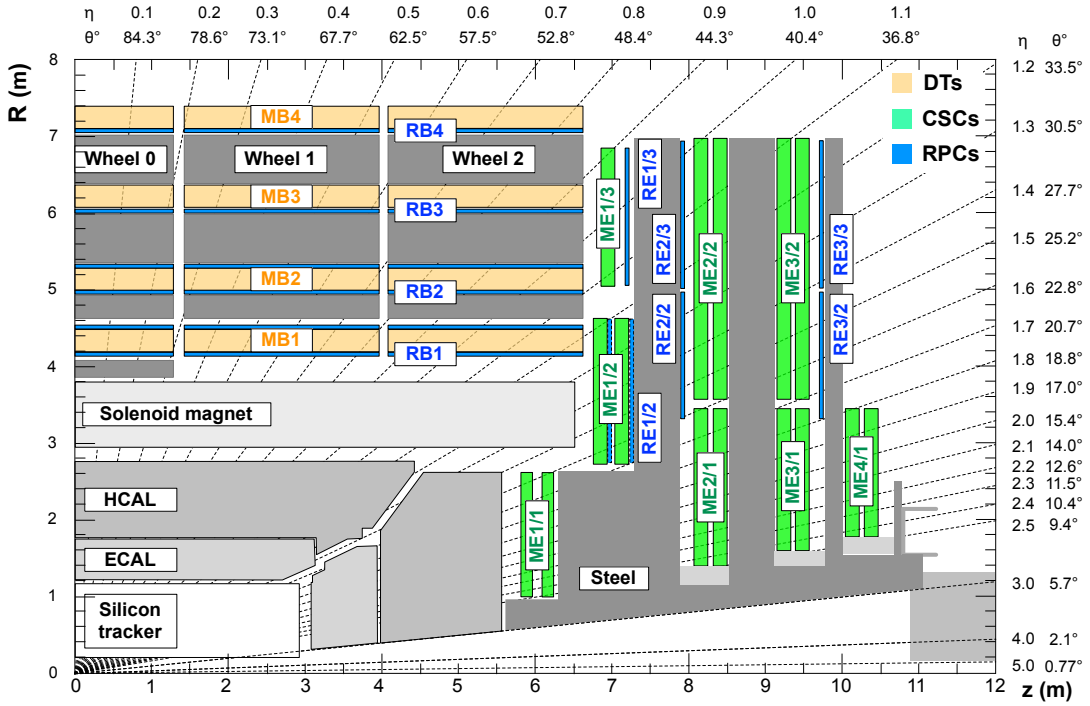

Figure 2.15: A quadrant of the CMS detector, indicating the positions of the CSCs, the RPCs, and the DTs [66].

drift time would be maximal. This design choice was motivated by the expected drift time of a muon in the DT gas mixture, which is 85% argon and 15% carbon dioxide. The drift cells are characterized by a single centrally-positioned gold-plated stainless-steel anode wire to which a +3600 V potential is applied. Four electrodes serve to shape the drift field: two on the wide walls at −1200 V and two above and below the wire at +1800 V.

DT chambers consist of so-called super-layers (SL), each containing four layers of drift cells staggered by half a cell length in each layer. The wires in the 2 outer SLs are oriented parallel to the beam direction and use to measure bening in the r-ϕ plane. The inner SL wires are perpendicular to the beam line and are used to measure z positions.

The Cathode Strip Chambers

The cathode-strip chambers (CSCs) are the central feature of the endcap muon system. The CSCs are are designed with a short drift path resulting in a fast response, and to tolerate the non-uniform magnetic field in the endcap region, as illustrated in Figure 2.8. The CSCs also have fine segmentation and are radiation resistant.

Each CSC is a trapezoidal chamber consisting of 6 layers of radial cathode strips interleaved with 7 anode wires panels, with wires roughly perpendicular to the strips. This allows for two coordinates of muon position measurement, with precise measurement of the ϕ value from the cathode strips and coarser measurement of the r position from the anode wires, which have readout channels ganged into groups. Schematics of the CSCs and the strip and wire orientations are given in Figure 2.19.

The CSCs cover a range of $0.9 < |\eta| < 2.4$, and are divided into 4 stations in both the $+z$ and $-z$ endcap, with rings of chambers, and are conventionally labelled by the endcap (+ or −), followed by the station, ring, and chamber number. In total, there are 72 chambers in each of ME1/1, ME1/2, ME1/3, ME2/2, and ME3/2. There are 36 chamber in each of ME4/1, ME2/1, and ME3/1. The ME4/2 chambers were mostly not present in 2010-2012 running, however 5 ME+4/2 chambers were in place. Currently, all ME4/2 chambers are installed for future data-taking. The size of the chambers varies with station and ring, with the largest being the ME2/2 and ME3/2 chambers, each occupying an area of approximately $3.4 \times 1.5\,\mathrm{m}^2$. In total, the CSCs have sensitive planes covering about $5000\,\mathrm{m}^2$, with a gas volume over $50\,\mathrm{m}^3$.

The operating mechanism of the CSCs is the detection of an avalanche of positive ions resulting from a traversing muon knocking electrons from atoms in the gas mixture (50%CO_2, 40%Ar, 10%CF_4). The avalanching ions move from anode wires in to the cathode strips, inducing a charge distribution perpendicular to the wire direction. The layered system of the CSC allows for redundancy of this phenomena in close succes-

sion. Predefined patterns of hits involving four or more layers (out of six possible) per chamber are used to reconstruct trajectories. The operating principle in a single layer and in an overall chamber is illustrated in Figure 2.18.

Cathode strips are read out in 200k channels, and wires are read out in about 180k channels. The readout electronics of the CSC system is calibrated and the functioning of its components (including chambers, strips, and wires) is constantly monitored. Electronics calibrations include adjustments for counting rates, chamber noise levels, and channel connectivity. To improve reconstruction, it is essential to calibrate the detector response according to the strip-to-strip crosstalk, the strip channel noise, the strip pedestals, and the strip channel electronic gains.

The cathode strips are connected to 16-channel amplifier-shaper application-specific integrated circuit (ASIC) chips, with outputs sampled every 50 ns with voltages stored in a switched capacitor array (SCA). Each ASIC contains two internal capacitors for each cathode amplifier channel. Each cathode front-end board contains a precision external capacitor to service the amplifier shaper. The CSC calibration constants are calculated according to the results of test pulses delivered in parallel by these capacitors. The test pulse amplitude is varied to determine the linearity, offset, and saturation of the amplifier, as well as the electronic gain used to normalize the pulse height measurements for reconstruction of muon hit positions. A short pulse is injected into each amplifier channel to determine cross-talk between neighboring strips. The cross-talk is determined for each of the two neighboring strips for a given strip, in terms of both magnitude and slope as a function of time. The crosstalk calibrations are used to model the crosstalk in simulation, and to deconvolve the effect of crosstalk in data. In practice, the reconstruction is relatively insensitive to the crosstalk. The pedestals for each strip are determined by sampling the the amplifier output at 20MHz with no input signal. The charge measured in each SCA time bin is the pedestal. The

covariance between the pedestals in each SCA time bin defines a "noise matrix" as $C_{ij} = \langle Q_i \cdot Q_j \rangle - \langle Q_i \rangle \langle Q_j \rangle$, where Q_i is defined as the charge in SCA time bin i, and the expectation values indicate the average over a large number of events.

Calibration constants are monitored continually over time through dedicated calibration runs taken between periods of data taking. Calibration constants are updated for use in reconstruction when reconstruction tests indicate notable improvements in the strip position resolution. The reconstruction is most sensitive to changes in the gains and pedestals, and relatively insensitive to changes in the crosstalk calibration constants. The differences in the values of the gains and pedestals for each of the 220,000 strip channels, before and after operation in 2010, is given in Figure 2.19. Correlation between the values on a given front-end board are visible [66].

The Resistive Plate Chambers

The resistive plate chambers (RPCs) are interspersed with the DTs and CSCs as illustrated in Figures 2.15, following a similar station-and-wheel design. The barrel wheel layout is identical to that of the DTs. The RPCs are a dedicated triggering detector system. They are characterized by very fast time resolution which improves determination of the beam crossing time at high luminosity (small bunch spacings of 25 ns). The RPCs measure muons in the range of $|\eta| < 1.6$, inclusive of the entire DT coverage and a portion of the CSC coverage.

The RPCs are double-gap chambers with 2-mm thick resistive plates and 2-mm thick gas gaps. The gas mixture is 95.2% Freon ($C_2H_2F_4$), 4.5% isobutane (i-C_4H_{10}), and 0.3% sulfur hexafluoride (SF_6). A relative humidity of 40–50% is maintained through the addition of water vapor to the gas mixture. The resistive plates are coated with conductive graphite and a 9.6 kV potential is applied. Muons traversing the gas will ionize atoms in the gas and avalanches induce a charge read out by readout strips

aligned in η between the two gaps.

The barrel RPCs (RB) consist of four stations RB1-4, and the endcap (RE) consists of three stations RE1-3. The innermost RBs (RB1 and RB2) consist of two layers surrounding the two innermost DT layers, while the outer RBs consist of only 1 layer on the inner-facing side of the DTs.

2.2.8 The Trigger System

In pp collisions with a bunch separation of 25 ns, the crossing frequency is approximately 40 MHz, and many simultaneous collisions can occur at each crossing due to pileup. Data processing time and storage capabilities limit the number of collisions which can be recorded. To reduce the rate of data processing and recording, a triggering system is implemented with the goal of quickly filtering out events which are not of importance to physics analyses.

Triggering is a two-step process, consisting of a Level-1 (L1) Trigger using programmable electronics, and a High-Level Trigger (HLT) implemented in software and executed on a computing farm with approximately one thousand processors. Combining the rate reduction capabilities of the L1 and HLT, less than one in one million events are retained for physics analysis.

The Level-1 Trigger

The L1 trigger is a hardware trigger designed to reduce the output rate to 100 kHz, a factor of several hundred below that of the design collision rate of the LHC. The L1 trigger decisions are based on coarsely segmented data from the ECAL, HCAL, and muon system. The corresponding full-resolution data is stored in memory on the front-end electronics, to be utilized if the event passes the L1 trigger. The hardware components of the L1 trigger consist of field-programmable gate-array (FPGA) technology, as well

as ASICs and programmable lookup tables (LUTs).

The L1 trigger is divided into three components: the Local Trigger, the Regional Trigger, and the Global Trigger. The Local Trigger is based on energy deposits in the calorimeter trigger towers and track segments or hit patterns in the muon chambers. In the calorimeters, these objects are often referred to as Trigger Primitive Generators (TPGs). For TPGs, the calorimeters are divided into trigger towers, and the TPGs sum the E_{T} of the ECAL crystals and HCAL read-out towers to attain the trigger E_{T}. The local muon trigger utilizes the DTs and CSCs to calculate local trigger information. Barrel DTs provide track segments in the ϕ-projection and hit patterns in the η-projection, while the CSCs provide track segments in three dimensions.

The Regional Triggers combine trigger primitive information and use pattern logic to sort trigger objects and rank them according to quality and kinematic properties, such as energy and momentum. Quality criteria are based on the uncertainty in the parameters measured at L1, given limitations on information and imperfect knowledge of the detectors and electronics. The Regional Muon Trigger uses track-finders in the DT and CSC to complete tracks by joining segments determined in the local trigger. The RPCs also provide track candidates with excellent timing resolution. The Regional Calorimeter Trigger determines electron and photon transverse energy based on the TPGs, as well as information necessary for calculations of muon isolation utilized in muon triggers.

Objects of the highest rank, including calorimeter deposits and muons, are transmitted to the Global Trigger which is responsible for combining the information from the subsystem trigger primitives and making the final determination of the event retention at L1. The workflow of the L1 trigger is illustrated in Figure 2.20.

The High-Level Trigger

The HLT is a software trigger which serves to reduce the rate of stored events to approximately 100 Hz. The HLT consists of a series of trigger paths corresponding to requirements on the presence of reconstructed particles with kinematic thresholds and quality criteria imposed. The implementation of the HLT has the overall goal of minimizing CPU while reliably retaining events needed to meet the goals of physics analyses. Events are reconstructed partially, with objects that can be used for event selection. A first round of event rejection is based on information from the calorimeter and muon detectors. Following this, CPU-intensive tracker track reconstruction is performed, and events can be rejected based on the status of more well-reconstructed objects, such as electron candidates.

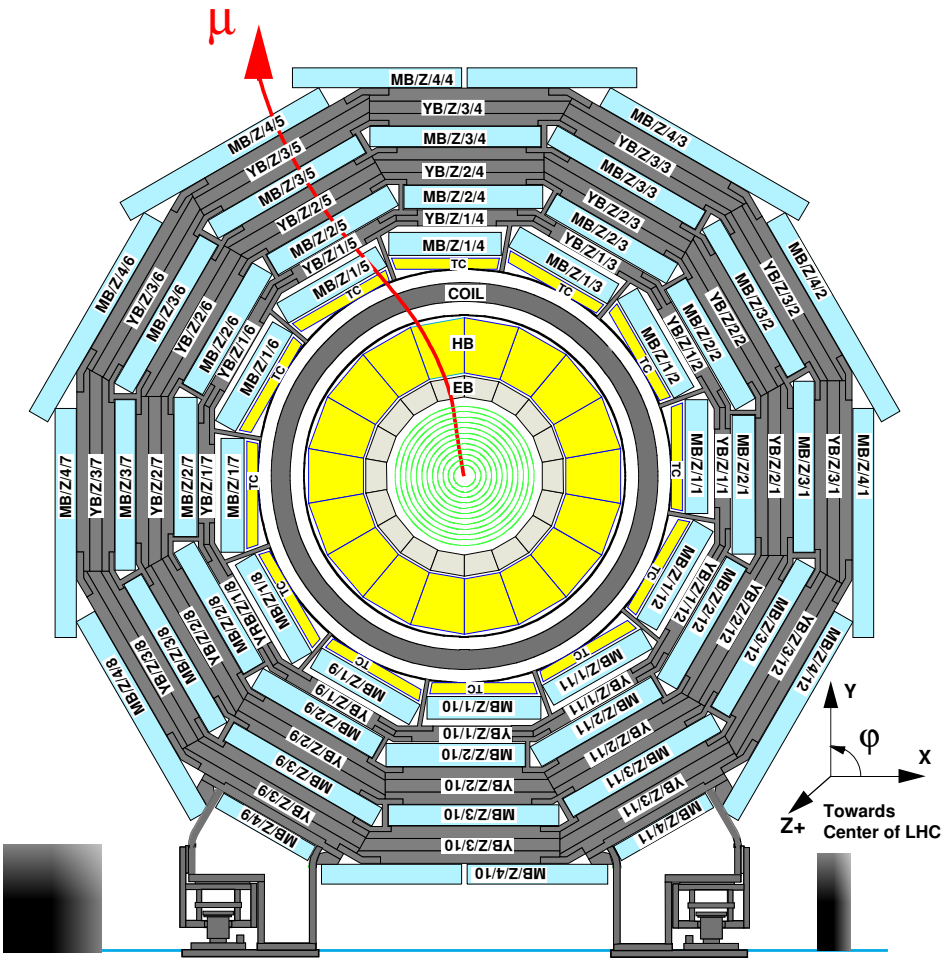

Figure 2.16: A cross-sectional view of the DT system, detailing the four radial layers of the DTs (MB1–4), and the chamber segmentation in the azimuthal angle [61]. A muon originating from the interaction point traverses the DT chambers with a curved trajectory due to the magnetic field.

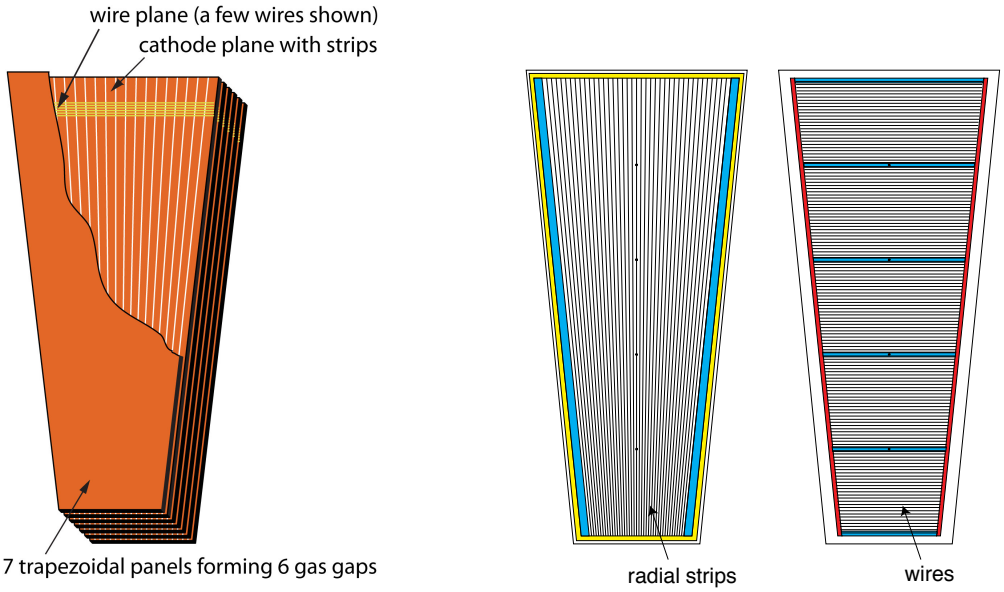

Figure 2.17: An illustration of a cathode strip chamber indicating orientation of the strips (radial) and wires (perpendicular ot central strip), as well as the 6 layers of the chamber stacked in the direction perpendicular to the strip-wire plane (left), and finer detail of the strip-wire plane in a single layer of a CSC (right) [68].

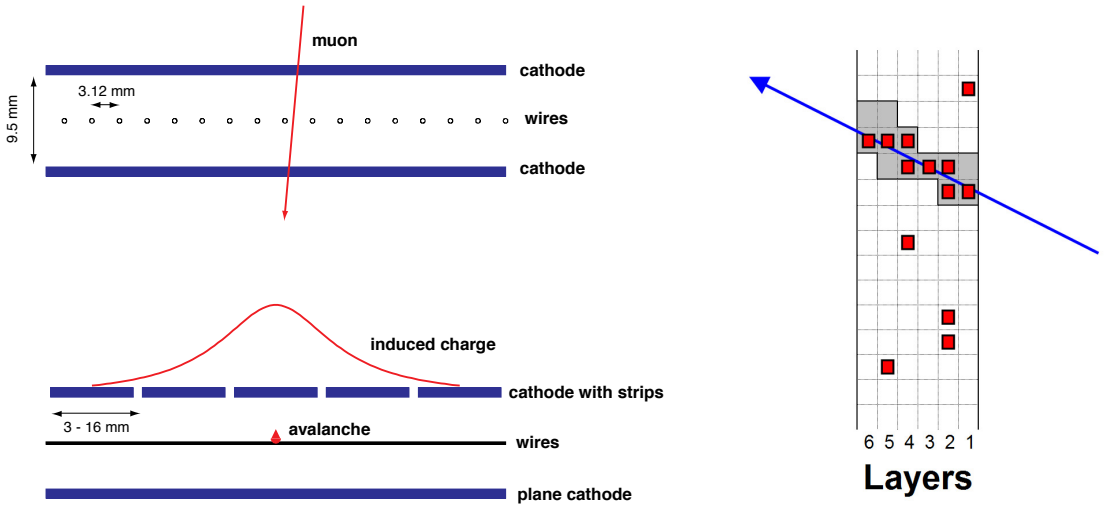

Figure 2.18: A diagram illustrating the operating principle of the CSCs. A muon passing through the causes an avalanche of positive ions near a wire, inducing a charge on the cathode strips (left). Hits in the CSC layers are matched to predefined patterns requiring hits in at least four planes (right) [68].

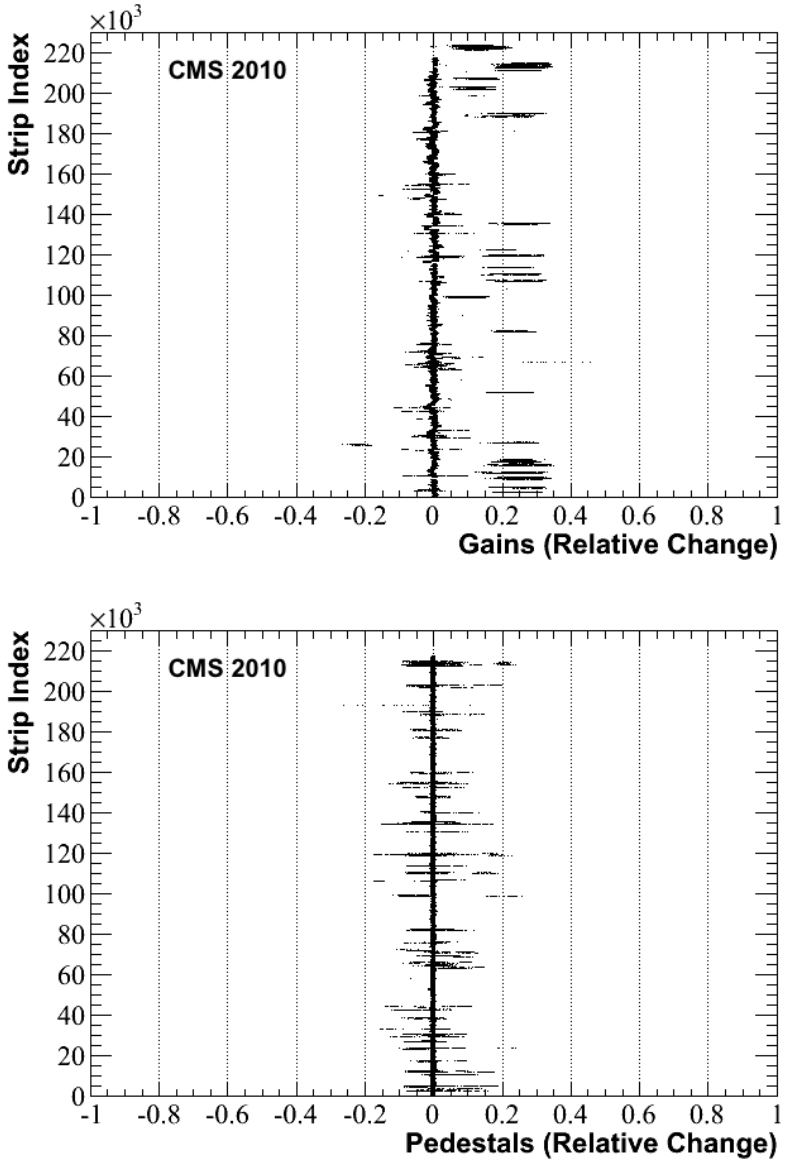

Figure 2.19: Changes in the values of the CSC calibration constants in the CMS running in 2010 for gains (top) and pedestals (bottom). The vertical axis indicates the "strip index" which increases sequentially with endcap, station, ring, and chamber. Clusters of values localized around a given strip index value having similar values of relative change indicate correlation between changes on a given front-end board [66].

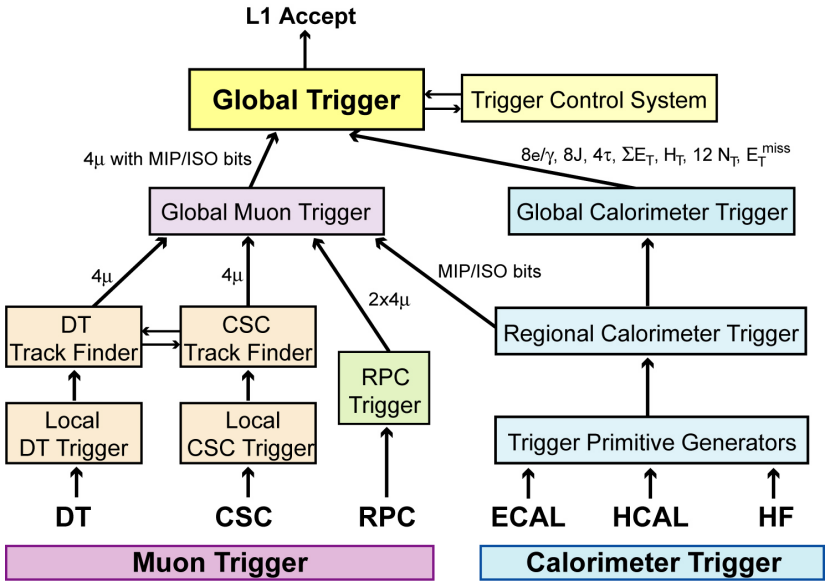

Figure 2.20: A diagram of the architecture of the L1 trigger. Local triggers using the CMS subsystems evaluate trigger primitives, which are combined in regional triggers and used by a global trigger to determine which events are accepted at L1 to be passed on to the HLT [61].

Chapter 3

Event Reconstruction

3.1 Track and Vertex Reconstruction

The fast and reliable determination of particle tracks and interaction vertices is of fundamental importance to the physics and performance goals of the LHC. Thousands of tracks are produced from up to 20 simultaneous collision events, and hits from these tracks are determined from the tracker's silicon pixel and microstrips sensor signals with high resolution. The extremely high granularity of the tracker allows for low average occupancy and accurate trajectory determination. After the reconstruction of a set of tracks corresponding to an event with multiple collisions, a set of vertices can be determined by considering the geometrical properties of the tracks. This section will first describe the reconstruction of tracks using a set of tracker hits, followed by the reconstruction of the vertices using the set of reconstructed tracks.

Track reconstruction is performed with the combinatorial track finder [69, 70], and begins with the process of track "seeding". Track seeds consist of a set of trajectories and errors which are candidates for full track reconstruction. Seeds are based on triplets or pairs of hits in the pixel layers which are compatible with the transverse

position of the interaction region of the detector (the "beamspot"), as determined with a beamspot fitter [71], and which satisfy a transverse momentum threshold. For each seed, a track is determined using a Kalman filter to extrapolate the trajectory outward to successive layers. Hits with a measured position that is consistent with the predicted position are included in the track. The track hit collection is cleaned of poor-quality hits according to criteria on the χ^2 between the hit measurement and predicted position, and ranked according to quality. Because the Kalman updating occurs in parallel for each track, the procedure is resistant to bias. Conditions on the number of shared hits in a track are used to remove double-counting of tracks, and truncation of the number of candidate tracks at each layer limits the computational expense of the algorithm. After the tracks are determined, the final track parameters are improved with a least-squares fit of the Kalman filter, giving a better estimate at the border of the tracker.

Interaction vertices are determined from the track collection. An initial set of "prompt" vertices is determined from a set of tracks with conditions on the transverse distance with respect to the beamspot (the "impact parameter"), the number of tracker hits, and the normalized track χ^2. The prompt tracks are clustered with an adaptive vertex fit [72] in the beamline (z) direction, and tracks in each vertex are weighted according to their compatibility with the vertex. Primary vertex candidates are filtered according to their compatibility with the beam line and ranked according to the sum in quadrature of the transverse momenta of the tracks associated with the vertex. Vertex position resolution depends on the number of charged tracks originating at the vertex, as displayed in Figure 3.1.

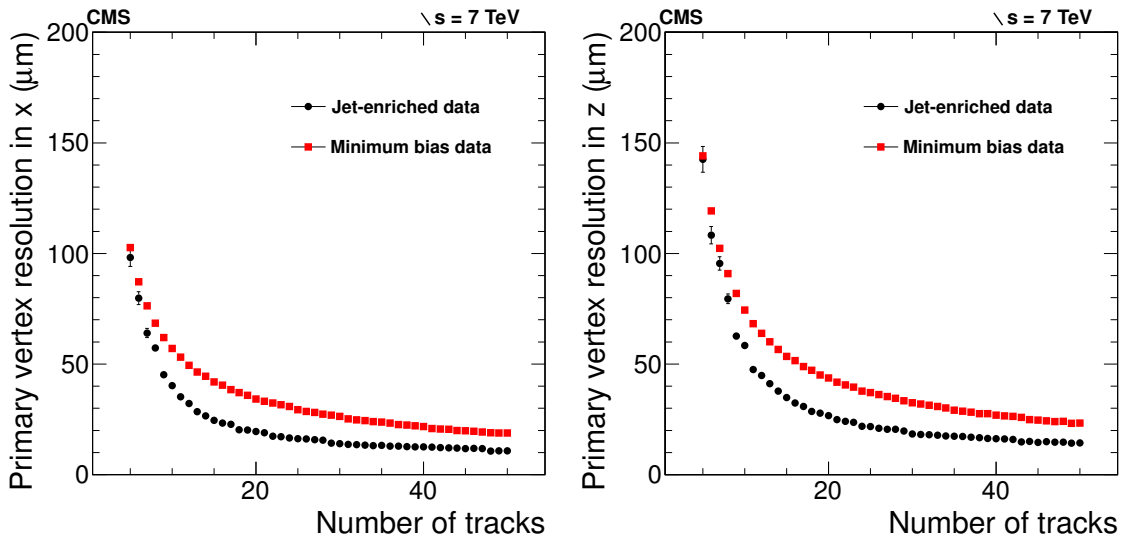

Figure 3.1: The resolution of the primary vertex position as a function of the number of charged tracks originating from the vertex, with data collected from pp collisions at $\sqrt{s} = 7\,\text{TeV}$ [73].

3.2 Muon Reconstruction

Muons are reconstructed as global objects after reconstruction of local muons in the muon subsystem, including the DTs, the CSCs, and the RPCs, and in the tracker. The detection of a muon in the tracker and muon chambers is depicted in a transverse view of the CMS detector in Figure 3.2. Muon reconstruction begins with seeds which provide an initial estimate of the trajectory parameters. For the reconstruction used in the HLT, these seeds are based on the L1 trigger objects. For offline reconstruction, segment patterns in the muon subsystems determined in local reconstruction are used as seeds. This description will focus on the offline muon reconstruction which relies on the local reconstruction.

The reconstruction of muons as global objects begins with the local reconstruction, which consists of a set of pattern-recognition algorithms, and makes use of local objects such as reconstructed hits and track segments. In the barrel DTs, hits are

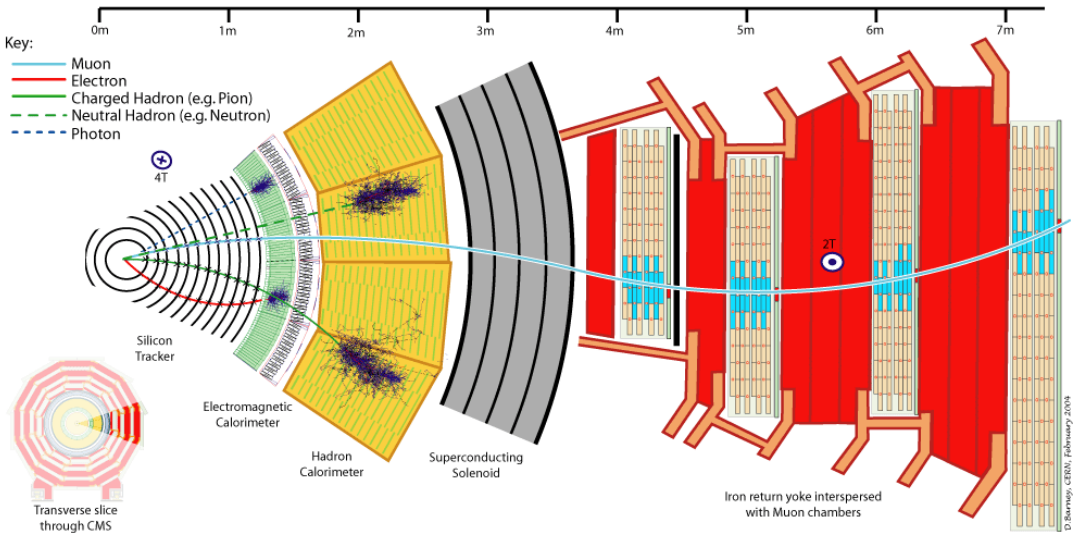

Figure 3.2: A slice of the CMS detector, depicting the subsystems which detect different types of particles. A muon is depicted being detected in the tracker and the muon chambers, and bending in the inner 4T magnetic field and outer 2T magnetic return field.

reconstructed from a series of pairs of points equidistant from the tube's central wire. Track segments are determined from a linear fit on the hit positions in the θ and ϕ projections. Identified segments are required to have three or more reconstructed hits from different layers. In the CSCs, the hits are obtained from strip-based and wire-based information, which allows for the determination of the local x, y, and ϕ coordinates. A typical muon registers hits in 3-6 strips. Muons induce a charge distribution on the strips consistent with the shape of a Gatti function [61], and a parameterization of the Gatti shower shape determines the center of gravity of the induced shower for the measurement of the local x coordinate. The strip-to-strip crosstalk is deconvolved following the procedure described in Section 2.2.7. Wire readouts occur in groups, with widths ranging from 2 to 5 cm. The two-dimensional hits are built from overlaps of a strip cluster and wire group. Segments are built from hits in the 6 layers of each CSC, using a spanning tree algorithm with one hit per layer and at least 3 layers. In the

RPCs, strips with readout are used in the local reconstruction to attain neighboring-strip clusters with average positions corresponding to reconstructed hit positions. The efficiency of local segment reconstruction in the DTs and CSCs is detailed in Table 3.1.

Muon Station	DT Efficiency (%)		CSC Efficiency (%)	
	Data	MC	Data	MC
Station 1	99.2 ± 0.4	98.05 ± 0.03	98.9 ± 0.9	97.8 ± 0.1
Station 2	99.0 ± 0.4	98.98 ± 0.03	96.8 ± 0.9	95.5 ± 0.1
Station 3	99.1 ± 0.4	99.08 ± 0.04	96.8 ± 0.9	94.1 ± 0.1
Station 4	98.9 ± 0.6	99.00 ± 0.04	94.9 ± 1.6	91.7 ± 0.2

Table 3.1: The efficiency of the local reconstruction of segments in the drift tubes and cathode strip chambers, in 2011 operations of the CMS detector, by muon station [74].

The second level of muon reconstruction, often referred to as Level-2 or standalone reconstruction, reconstructs muons as objects by combining information from the muon detectors and omitting the tracker. The segments obtained in local reconstruction are treated as seeds to the standalone track reconstruction, with trajectories determined from a Kalman-filter fitting performed radially outward. The homogeneity of the magnetic field in the DT barrel allows for the use of segments as measurements in the Kalman-filter procedure, while the varying magnetic field in the CSC requires the use of 3-D reconstructed hits as measurements. A requirement on the χ^2 of each measurement reduces contamination from spurious hits due to showering, delta rays, and pair production. A reverse Kalman-filter fitting is then applied (radially inward) to attain the track parameters at the innermost muon station, and the track is extrapolated to the interaction point.

The final stage of muon reconstruction, termed Level-3 or global reconstruction, combines the information of the standalone muons with the tracker information. For each standalone muon, the track parameters at the innermost muon station are used to estimate the track parameters at the outer edge of the tracker. Tracker layers and

regions of interest are used to perform a regional track reconstruction, and the definition of the region of interest (ROI) is tuned to balance optimal values of reconstruction efficiency, misidentification rate, and CPU time. In the ROI, two hits from two different tracker layers define a seed, and a Kalman-filter approach is used to reconstruct the muon track in the ROI. This reconstruction also filters out incompatible hits based on a χ^2 of the track fit, and smooths the final set of hits including the hits in the muon chamber. An additional cleaning of incompatible hits is then applied, including cleaning of hits in the muon stations with incompatible values of χ^2. For accurate reconstruction of high p_{T} muons approaching 1 TeV, a compatibility is evaluated between the tracker-only reconstruction and a reconstruction using the tracker and the first muon station, in order to detect muon bremsstrahlung. The muon q/p_{T} resolution as a function of p_{T} is given in Figure 3.3.

3.3 Electron Reconstruction

Electrons are reconstructed with tracker and ECAL information. As with other objects, the reconstruction begins with a seed. Electron seeds come in two varieties: ECAL-driven seeds which are best-suited for high-p_{T} isolated electrons, and tracker-driven seeds which are optimal for low p_T electrons and electrons in jets. The fundamental unit of ECAL activity relevant to the reconstruction is the "supercluster" - a group of one or more clusters of ECAL energy deposits with a shape characteristic of energy deposits from an electron trajectory bending in a magnetic field. Superclusters are determined with a "Hybrid" algorithm which considers a fixed bar of 3 or 5 crystals in η and a dynamic range in the ϕ direction. Smaller clusters can be determined with an "Island" algorithm which considers a single crystal satisfying an energy deposit threshold, and adds neighboring crystals iteratively in ϕ and then in η.

Figure 3.3: The q/p_{T} resolution of muons determined with cosmic rays, as a function of the muon p_{T}, using a tracker-only muon reconstruction and a reconstruction combining the tracker and muon system information. The cosmic rays are detected as traversing the CMS detector from top to bottom, and are split into the top and bottom legs. The Gaussian width of the $[q/p_{\mathrm{T}}(\mathrm{top}) - q/p_{\mathrm{T}}(\mathrm{bottom})] \,/\, \left[\sqrt{2}q/p_{\mathrm{T}}(\mathrm{bottom})\right]$ is plotted as a function of $q/p_{\mathrm{T}}(\mathrm{bottom})$ as measured with the tracker.

A reconstructed electron consists of a single track matched to an interaction vertex and to an electromagnetic supercluster, with $p_{\mathrm{T}} > 5\,\mathrm{GeV}$ and $|\eta| < 2.5$. ECAL superclusters motivate the determination of pixel seeds, consisting of two pixel hits, used to determine the primary electron track. The energy-weighted mean position of the supercluster is extrapolated backward through the magnetic field toward the pixel detector, considering both possible electron charges, and trajectories are determined using a pixel matching algorithm detailed in Ref [75].

Due to bremsstrahlung emission, electrons require a dedicated track reconstruction strategy. For high p_{T}, it is suitable to use a standard Kalman-filter algorithm with χ^2-

based cleaning of the hits used in track evolution. For low-p_{T} electrons, the multiple scattering can be suitably modeled with a Gaussian Sum Filter (GSF) fit which is a nonlinear generalization of the Kalman-filter.

Different classes of electrons are used depending on analysis-specific needs at CMS, and can vary according the the range of p_{T} of the electrons being measured, as well as the expected hadronic activity in the event. A description of the classes of electrons and identification criteria are given in Ref [76].

3.4 Particle Flow Reconstruction, Jets, and $E_{\mathrm{T}}^{\mathrm{miss}}$

The analyses presented in this dissertation make use of the particle-flow (PF) [77] algorithm of reconstruction for jets and missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) which combines information from all CMS sub-detectors to identify stable particles and measure kinematic quantities. Hadronic jets, composed of charged hadrons, neutral hadrons, and photons, must be reliably reconstructed for an accurate determination of the $E_{\mathrm{T}}^{\mathrm{miss}}$ in the event. The particle-flow method improves the jet and $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction by first determining a set of "fundamental elements" in the event — consisting of charged particle tracks, calorimeter clusters, and muon tracks. The granularity of the tracker and strength of the magnetic field allow for charged hadron momentum measurement that is superior to that of the calorimeters for most charged hadrons, while those with very high p_{T} benefit more from calorimeter measurements. To reliably measure charged hadron tracks with high efficiency and a low contamination from fake tracks, an iterative tracking strategy is used, in which the tight criteria used for track seeding is progressively loosened in several iterations, and each iteration reconstructs tracks with a method that cleans the track of poorly matched hits.

Calorimeter clusters are determined separately in the barrel and endcaps of the

ECAL, and of the HCAL. Cells corresponding to energy deposits greater than a threshold value (based on the standard deviation of calorimeter noise measurements) are treated as cluster seeds, and cells with energy deposits neighboring the seed are grouped into a larger topological cluster. Topological clusters serve as seeds to PF clusters.

Since a given particle can be expected to register hits in multiple sub-detectors, the PF elements based on these hits must be connected by a "link algorithm" to achieve a particle-based reconstruction. This algorithm pairs elements based on a distance parameter which quantifies the quality of the link, and differs for different types of elements. Charged tracks and calorimeter clusters are linked according to a distance defined by the physical distance in (η, ϕ) between the extrapolated track and the cluster center, for clusters in which the extrapolated track falls within the cluster boundary. Clusters of two different calorimeters (ECAL, HCAL, or PS) can be linked when the cluster position of the more granular calorimeter (PS or ECAL) is within the cluster envelope of the less granular calorimeter (ECAL or HCAL), and the physical distance in the (η, ϕ) plane again defines the link distance. A link between a charged track and a muon can also be determined when a global fit between the two tracks returns a χ^2 below a certain threshold, and in this case the link distance is simply defined as the χ^2 value.

Linked elements, referred to as "blocks", are used to reconstruct a set of particles describing the entire event. For each block, reconstruction begins with muons, which can be reconstructed reliably with a low fake rate. Global muons are considered PF muons if the global muon momentum is consistent with the tracker-muon momentum within three standard deviations, and the track is removed from the block. Electrons in the block are identified by associated short pre-shower tracks, and these electrons are refit with a GSF and extrapolated to the ECAL so that final identification can make use of both the ECAL and the tracker. The PF electron track and ECAL clusters are

then removed from the block.

Remaining tracks are filtered according to a tight quality requirement to reduce fake tracks, and tracks surviving these criteria can be connected to ECAL and HCAL clusters. Determination of neutral particles is performed by comparison of the measured track momentum with the hadronic energy deposits, accounting for the possibility of multiple tracks associated with he same HCAL cluster. Tracks linked to multiple clusters are associated to their closest cluster. Overlapping photons in ECAL clusters are treated by ignoring the links and ordering ECAL clusters for each track according to track-cluster distance, and retaining the link if the total calorimeter energy is smaller than the charged particle momentum. When track momentum exceed calorimeter energy deposits by more than three standard deviations, conditions applied to muons and fake tracks are loosened and corresponding block modifications are repeated until the entire track collection with $p_T > 1$ GeV is considered or the track momentum is smaller than the calorimeter energy.

Remaining tracks are considered PF charged hadrons, and tracks with momenta consistent with calorimeter energies are refit to include the calorimeter clusters. Calorimeter clusters linked to tracks with significantly less momentum define PF photons and PF neutral hadrons, where PF neutral hadrons are defined with the excess of the ECAL energy minus the track momentum, should any excess exist. Remaining ECAL and HCAL clusters are considered PF photons and PF neutral hadrons. Jets are clustered according to standard algorithms described in Section 3.5 using the PF particle set. The performance of the jet reconstruction using the particle-flow algorithm is compared to that of simple calorimeter jets by examining the differences in jet energy resolution as a function of jet p_{T}, in Figure 3.4. The jet p_{T} resolutions are assessed with asymmetry methods using γ+jet and dijet samples [78]. The particle-flow method is observed to result in better jet p_{T} resolution at all p_{T} values. The comparison between

the jet energy resolution in data and simulation is shown in Figure 3.5, and differences of less than 10% are generally observed.

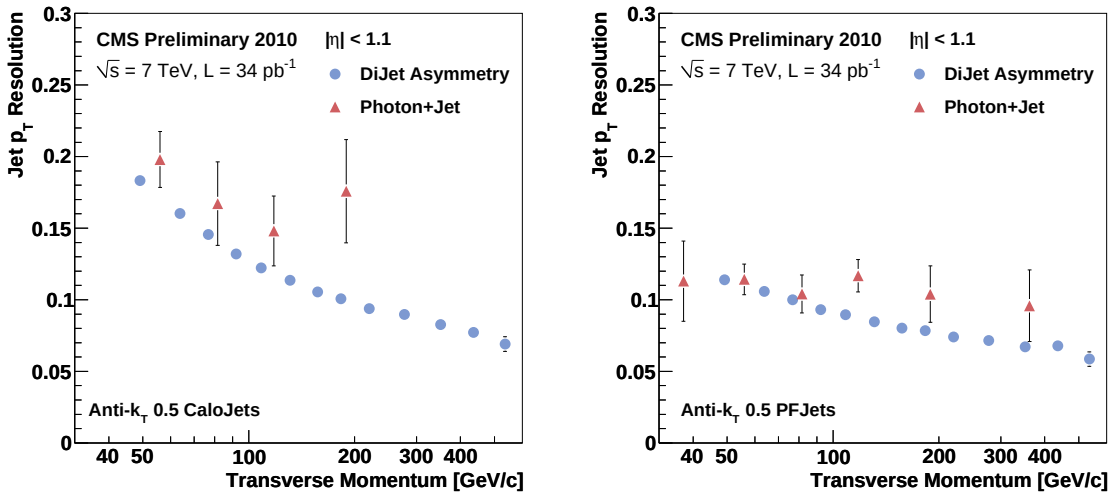

Figure 3.4: The jet energy resolution as a function of jet transverse momentum, for jets with $|\eta| < 1.1$. Particle-flow jets (right) are compared with jets clustered from calorimeter deposits (left), and a dramatic improvement in resolution is evident [78].

Finally, the $E_{\mathrm{T}}^{\mathrm{miss}}$ is simply the magnitude of the negative vector sum in the transverse plane of all reconstructed PF particles. Comparisons between particle-flow $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution and calorimeter based $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution are shown in Figure 3.6 [79].

3.5 Particle Identification and Quality Criteria

3.5.1 Muon Selection

Reconstructed muons are identified with quality criteria which are in place to reject fake and cosmic muons and guarantee isolation from jets. The quality criteria are based on subdetector measurements associated with the muon track. Some common muon selections used at CMS are as follows:

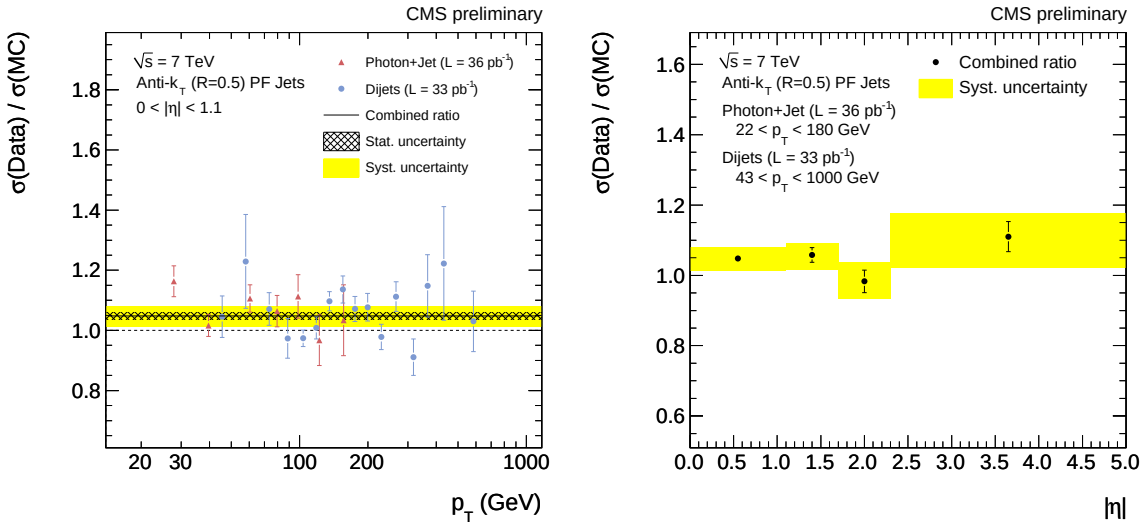

Figure 3.5: (Left) The ratio of jet p_{T} resolutions in data and simulation, as a function of jet p_{T}. (Right) The ratio of jet p_{T} resolution in data and simulation, in different ranges of $|\eta|$. [78].

- *Tight muon selection*
  - The muon is a global muon with a χ^2/d.o.f. of the global muon track fit less than 10.
  - At least one hit is present in a muon chamber included in the global muon track fit.
  - The tracker track associated with the muon is required to be matched to muon segments in at least two muon stations, and to use more than 10 tracker hits to determine the track parameters, with at least one hit in the pixel detector.
  - The transverse impact parameter (distance of closest approach in the x-y plane) of the track with respect to the primary vertex of $|d_{xy}| < 2\,\mathrm{mm}$.
- *High-p_T muon selection*

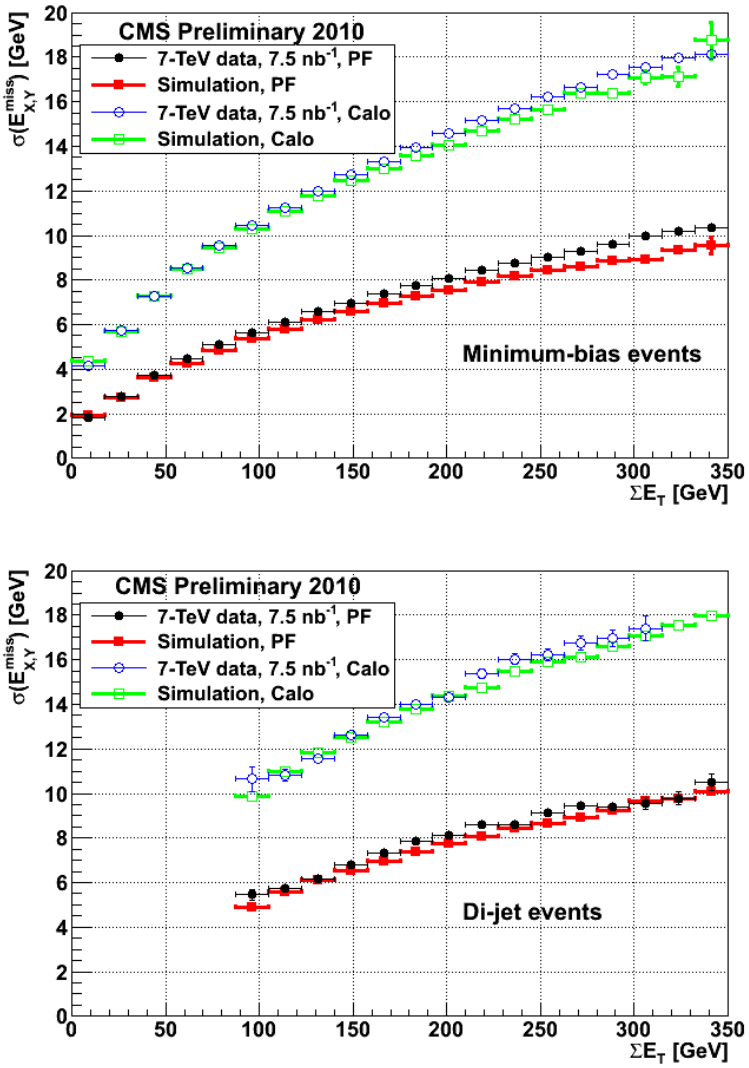

Figure 3.6: The $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution in data and simulation, using 7 TeV data collected in 2010, comparing particle-flow $E_{\mathrm{T}}^{\mathrm{miss}}$ to $E_{\mathrm{T}}^{\mathrm{miss}}$ calculated with calorimeter deposits. (Top) The $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution in minimum-bias events. (Bottom) The $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution in dijet events [79].

- The muon is a global muon.
- At least one hit is present in a muon chamber included in the global muon track fit.
- The tracker track associated to the muon is required to be matched to muon segments in at least two muon stations, with at least 5 tracker layers containing hits, and at least one hit in the pixel detector.
- The transverse impact parameter of the track with respect to the primary vertex is required to be $|d_{xy}| < 2\,\text{mm}$, and the longitudinal distance of the track with respect to the primary vertex is required to be $|d_z| < 5\,\text{mm}$.
- The track used for momentum determination is required to have a relative error in p_T of less than 30%.

Muon quality criteria are combined with an isolation criterion that selects muons with relatively small or no activity from other particles in the vicinity of the muon, such that accurate reconstruction of the muon is more likely and muons originating from jets and jets incorrectly identified as muons can be removed. Muon isolation is a subdetector-based quantity which can consider one or all of the ECAL, HCAL, and tracker. For high-p_T muons with Bremsstrahlung, the simplest and most reliable estimation of activity surrounding the muon can be taken from the tracker tracks surrounding the muon. For muons of low p_T, it is most useful to consider the ECAL and HCAL energy deposits surrounding the muon, especially in events with one or more jets. In general, we can define two types of subdetector-based isolation:

- *Tracker-only isolation* is defined as the sum of the p_T of tracks surrounding the muon, excluding the muon track itself, divided by the muon p_T. The typical cone size for this quantity is $\Delta R = 0.3$, and a tight (loose) cut on the relative isolation (Iso_Rel) is $\text{Iso}_\text{Rel} < 0.05(0.10)$.

- *Combined isolation* is defined as the sum of the p_{T} of tracks surrounding the muon, excluding the muon track itself, added to the the energy deposits in the ECAL and HCAL surrounding the muon, divided by the muon p_{T}. The typical cone size for this quantity is $\Delta R = 0.4$, where $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$, and a tight (loose) cut on the relative isolation is $\mathrm{Iso}_{\mathrm{Rel}} < 0.12(0.20)$.

3.5.2 Jet Clustering

The analyses presented in this thesis rely on the anti-k_{T} algorithm [80] for the clustering of final-state particles into jets. In simulations at the particle level, jet clustering is performed using the set of generated particles after decays and fragmentation. At the reconstructed level, in either data or simulation, jets can be created from calorimeter deposits (Calo jets) or from particle-flow particles (PF jets). In the following description of the anti-k_{T} algorithm, for generality, the entities being combined into jets will be referred to as "objects".

The anti-k_{T} algorithm is a "clustering" algorithm. Clustering algorithms use iterative pair-wise clustering, beginning with a single object and building a successively larger collection of objects to define the final jet. This can be contrasted with "cone" algorithms which aim to define spatial regions wherein the summed objects have maximal momentum. Clustering algorithms, including anti-k_{T}, rely on a distance between to objects, i and j being considered for clustering (d_{ij}), and the distance between a given object and the beam direction, d_{iB}.

The anti-k_{T} begins with the set of objects, including the beam direction B. The distance parameters are defined as:

$$d_{ij} = \min\left(k_{\mathrm{T}i}^{2p}, k_{\mathrm{T}j}^{2p}\right)\frac{\Delta_{ij}^2}{R^2} \tag{3.1}$$

$$d_{iB} = k_{\mathrm{T}i}^{2p} \tag{3.2}$$

where $\Delta_{ij} \equiv (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, $k_{\mathrm{T}i}$ is the transverse momentum of the i^{th} object, y_i is the rapidity of the i^{th} object, and ϕ_i is the azimuthal angle of the i^{th} object. The variable R is referred to as the radius parameter, and it governs the angular size scale of the resulting jets. The power parameter p governs the relative power of the energy. A value of $p = 1$ defines the classic k_{T} clustering algorithm, while a value of $p = -1$ is referred to as the anti-k_{T} algorithm, which has been shown to be infrared and collinear safe and yield sensible results [80]. A comparison among jet algorithms in the reconstruction of the top quark mass in $\mathrm{t\bar{t}}$ events is illustrated in Figure 3.7.

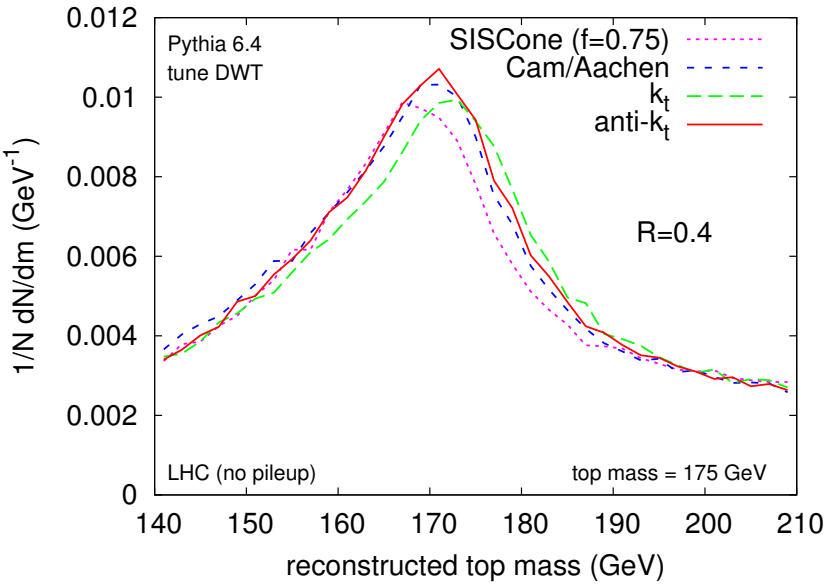

Figure 3.7: The reconstruction of the top quark mass in $\mathrm{t\bar{t}}$ events simulated with Pythia. Both quarks decay hadronically to a b quark and W boson, and both W bosons decay hadronically. The four non-b jets are paired in the combination that yields the closest invariant mass to the of the W boson, and the resulting light-jet pairs are paired with the b-jets according to the pairing which yields the smallest difference in invariant mass. Several jet algorithms are considered, and the marginally better performance of the anti-k_{T} algorithm is observed [80].

3.5.3 Jet Energy Corrections

The reconstructed jet energies are corrected to improve the correspondence between the reconstructed energy and the true particle-level energy. The jet energy is calibrated using the p_{T} balance of dijet and γ+jet events [81] to account for the following effects: nonuniformity and nonlinearity of the ECAL and HCAL energy response to neutral hadrons, the presence of extra particles from pileup interactions, the thresholds used in jet constituent selection, reconstruction inefficiencies, and possible biases introduced by the clustering algorithm.

Figure 3.8 displays the size of the jet energy residual correction for differences in data and simulation after corrections for known detector effects and running conditions, as well as the total uncertainty on the jet energy corrections as a function of the jet p_{T}. For values of $p_{\mathrm{T}} > 100$ GeV, the jet energy correction uncertainty is at the 1% level.

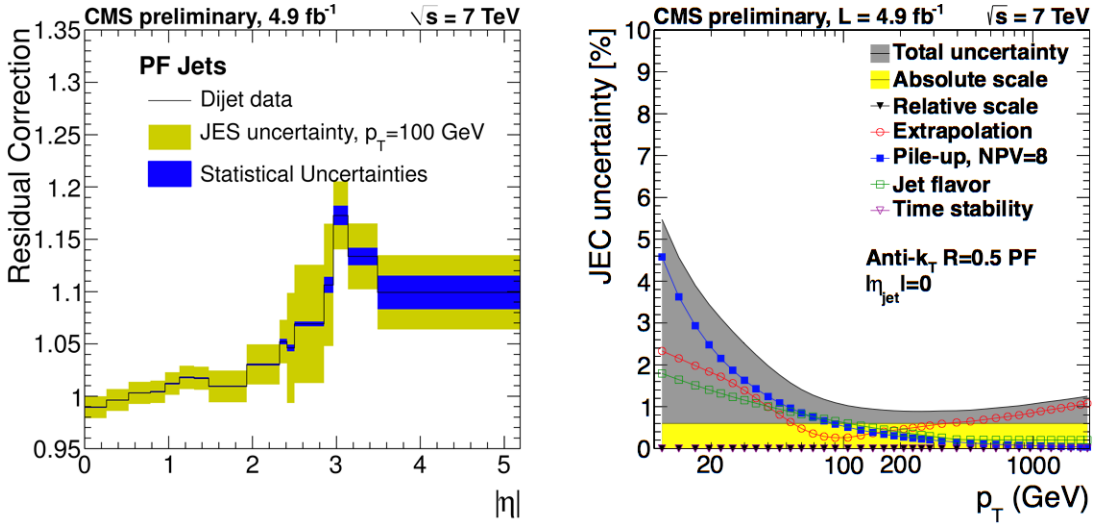

Figure 3.8: (Left): The residual correction factor determined from di-jet data and jet energy correction uncertainties, for a jet with $E_{\mathrm{T}} = 100$ GeV. (Right): The jet energy scale uncertainties as a function of jet p_{T} for a central jet with $\eta = 0$. [82]

3.5.4 Reconstructed Jet Identification

A set of jet identification quality criteria is defined for PF jets. The quality criteria used in results presented in this work, corresponding to the CMS "loose" jet identification, are as follows:

- The fraction of the jet energy corresponding to neutral hadrons (those not linked to tracks in the PF algorithm) must be less than 0.99.
- The fraction of the jet energy corresponding to neutral electromagnetic energy deposits must less than 0.99.
- The number of constituents clustered to create the jet must be more than 1.

Additional quality criteria can be applied for a cleaner sample of jets in the forward region ($\eta > 2.4$), however these jets are not considered in the work presented here.

3.5.5 Heavy-Flavor Jet Tagging

The accurate determination of jets originating from b-quarks is crucial to measurements of SM processes and searches for new physics. The tagging of jets as b-jets can be based on several aspects of the tracks and vertices associated with the jets and overall event reconstruction, such as track impact parameters (IPs), the properties of reconstructed vertices, and the leptonic content of the jet. One simple yet effective method of tagging a b-jet relies on the IP of the tracks associated with the jet, with respect to the primary vertex. The high resolution of the CMS tracker allows for an accurate determination of the IP, defined as the distance of closest approach between a track and a vertex in all three dimensions. Tracks originating from the primary vertex will tend to have very small IPs, while tracks originating from particles within the jet (i.e. from a secondary vertex) will have larger IPs. Since the measurement of each IP has an associated

uncertainty, the impact parameter significance is defined simply as the ratio of the IP to its estimated uncertainty.

The significance of the track IPs for a given jet can be used to discriminate between heavy flavor and light jets. The "track counting" (TC) algorithm sorts the tracks associated with a jet according to IP significance. For a jet with n tracks, this allows for n possible discriminating variables corresponding to the significance of the n^{th} IP. Two variants of this algorithm are commonly used. A "track-counting high-effiency" (TCHE) algorithm uses $n = 2$, while the "track-counting high-purity" (TCHP) uses $n = 3$. The TCHE discriminator in data and simulation, for a center-of-mass energy of 7 TeV, is displayed in Figure 3.9.

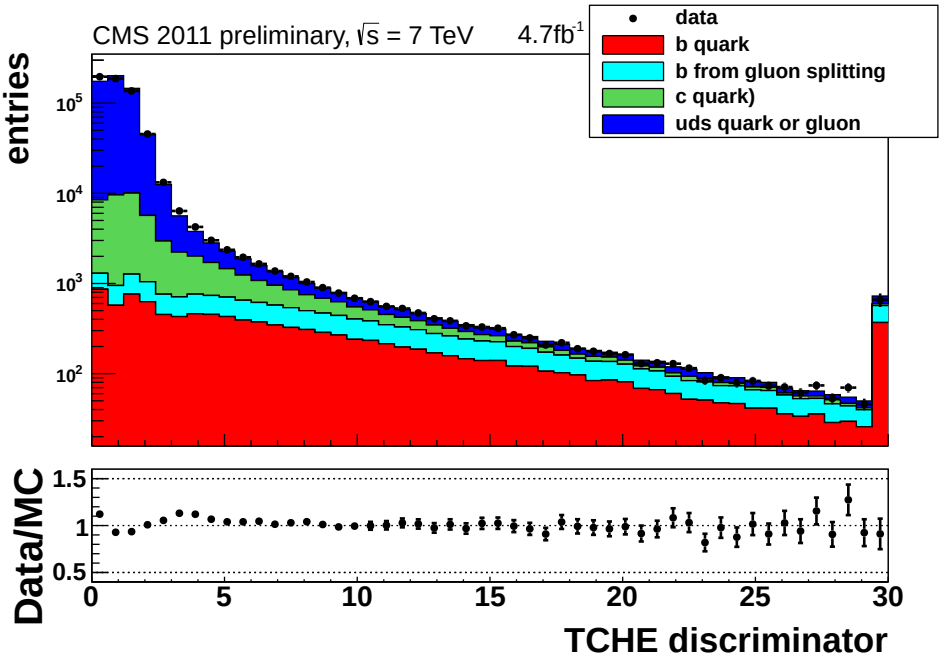

Figure 3.9: The track-counting high-efficiency discriminator, corresponding to the IP significance of the track in the jet with the second-highest IP significance. Good agreement between data and simulation is observed. [83]

For a given b-tag discriminator, a threshold on the discriminator can be applied to define a jet as b-tagged. A limited set of thresholds for any given discriminator are

chosen as standard "working points", and have been studied in detail at CMS [83]. Some of the analyses presented in this thesis make use of the medium working point of the TCHE tagger (TCHEM), with a threshold at 3.3, to define a b-tagged jet.

3.6 Simulation of Events and Detector Response

Event simulation is a fundamental component of searches and measurements with the CMS detector. For the estimation of signal processes, and for the estimation of backgrounds for which data-driven estimation is not feasible, simulation is relied upon. The two major aspects of simulation are the generation of the underlying event, and the simulation of the detector response to the final-state particles of the generated event.

Commonly-used event generators in CMS include Pythia [84], Sherpa [43–46], Madgraph [41], and Powheg [85–88]. Simulations using leading-order (LO) generators are produced with the LO CTEQ6L1 parton distribution functions (PDF) [89]. The simulation with Sherpa uses the CTEQ6.6m PDF, and the simulations with Powheg use the CTEQ6m PDF. Samples of MadGraph and Powheg rely on an interface to Pythia for parton showering and hadronization. In Pythia, matching between jets and partons is performed with the kT-MLM matching scheme, while in Sherpa, matching is performed with the CKKW matching scheme [40].

For a given simulation, after the set of final-state particles is determined, a full simulation of the detector response and a reconstruction of that response is required for direct comparison of simulation with reconstructed data. The detector simulation is based on the Geant 4 [90] simulation toolkit and the CMS object-oriented framework and event model, and it is implemented for all CMS subsystems. The Geant 4 toolkit provides a robust description of the hadronic and electromagnetic interactions, as well

as a modeling of the CMS detector geometry. For each simulated detector hit, the corresponding electronic readouts are simulated in process referred to as "digitization".

In a high luminosity environment, a critical part of the simulation is the addition of multiple pileup interactions to the simulated event. Since it not possible to know the luminosity profile for a given running period in advance, and since analyses may use a subset of all available data, pileup events are added to the events according to a generic profile which can be reweighted according to the targeted running conditions. In addition to the injection of events in the central proton bunch (in-time pileup), pileup from neighboring bunches (out-of-time pileup) is also modeled. Pileup events are simulated as additional minimum-bias events, and these events are never re-used for a given MC sample to avoid correlations between subsets of events. The number of pileup interactions added to the simulation is chosen according to the pileup multiplicity distribution observed in the data during the data-taking periods of interest.

Chapter 4

The Search for Leptoquarks with the Early CMS Data

Searches for second generation scalar leptoquarks have been conducted with all of the the pp collision data recorded with the CMS detector. While the later searches used large datasets corresponding to $5\,\text{fb}^{-1}$ at $\sqrt{s} = 7$ TeV and $20\,\text{fb}^{-1}$ at $\sqrt{s} = 8$ TeV, in both the $\mu\mu jj$ and $\mu\nu jj$ final states, a simpler search with the first 34 pb^{-1} of $\sqrt{s} = 7$ TeV collision data was able to make significant improvements over previous searches at hadron colliders. This section presents an overview of the first search at CMS with the $\mu\mu jj$ final state. Greater detail is given for the searches with the full CMS datasets, which are presented in Section 5.

For the search presented here [91], the signature of the decay of pair-produced second-generation scalar leptoquarks consists of two muons and two jets with high transverse momentum. Because the dataset being used is relatively small, and because the event selection yields relatively low background levels, this search is statistically limited. As there are few viable control regions for background studies, and statistical uncertainties are dominant, a conservative assessment of the systematic uncertainties

is performed, and background estimation relies on simulation.

This section proceeds as follows: Section 4.1 describes the data and simulated samples used to conduct the search. Section 4.2 describes the selection applied to events. Section 4.3 describes the normalization of the background contributions. Section 4.4 describes the systematic uncertainties. Section 4.5 gives the results in terms of event yields, and limits on the LQ cross section and branching ratio β.

4.1 Data and Simulation Samples

The Monte Carlo (MC) signal events are generated in the LQ mass range 250-500 GeV, using the PYTHIA event generator (version 6.422) and tune D6T [92, 93]. The main background processes that can mimic the signature of the LQ signal are $Z/\gamma^* + \text{jets}$, $t\bar{t}$, VV (WW, ZZ, WZ), W+jets, and multijet events. The $t\bar{t}$ and VV background are generated with MADGRAPH [94,95]; Z/γ^*+jets and W+jets events are generated with ALPGEN [96]; and muon-enriched QCD multi-jet events are generated with PYTHIA . In MADGRAPH and ALPGEN samples, parton showering and hadronization is performed with PYTHIA.

4.2 Event Selection

Events are selected by a single-muon trigger without isolation requirements and with lower p_{T} thresholds which vary during the data-taking period. Additional requirements on the muon p_{T} in selected events are applied during the analysis which are well-beyond the trigger requirement. The combined HLT and first-level trigger efficiency in these events is approximately 92% per muon, and simulated samples are adjusted for this acceptance rate.

Muons are selected according to the "Tight" identification criteria described in Section 3.5.1. Muons are required to have $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.4$, with at least one muon in the pseudorapidity region $|\eta| < 2.1$. To reduce contamination from jets that are miscreconstructed as muons, a relative isolation requirement condition is placed on the muons. The combined relative isolation, described in Section 3.5.1, is applied, with a threshold at 0.05. In addition, the two muon candidates are required to be separated from each other by at least $\Delta R = 0.3$. The efficiency of selecting dimuon events is 61-70% for the LQ mass range of 200 GeV to 500 GeV.

Jets are reconstructed from calorimeter deposits using the anti-$\mathrm{k_T}$ [80] algorithm with a distance parameter $R = 0.5$ and are required to have $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 3.0$. Jet energy-scale corrections derived from MC simulated events are applied to establish a relative uniform response in η and an absolute uniform response in p_{T}. A residual jet energy correction is derived from data according to the p_{T} balance in dijet events, and this correction is applied to jets in data.

Additional selection requirements are placed on two variables, which are effective at discriminating the LQ signal from the major sources of background. The first is the dimuon invariant mass, $M_{\mu\mu}$. The second variable, S_{T}, is defined as the sum of the magnitudes of the p_{T} of the two highest p_{T} muons and the two highest p_{T} jets. The two muons in the signal events come from the decays of two high-mass particles, and they tend to form a large invariant mass. Thus, events are selected if $M_{\mu\mu} > 115$ GeV. This helps to reduce the contribution from $\mathrm{Z}/\gamma^* + \mathrm{jets}$ processes, which is one of the largest backgrounds. In addition, the LQ pair is expected to have a large S_{T}. The lower threshold on S_{T} is optimized for different LQ mass hypotheses by using a Bayesian approach [97, 98] to minimize the expected upper limit on the LQ cross section in the absence of an observed signal. The S_{T} cut helps to further reduce background sources, most noticeably $\mathrm{t\bar{t}}$. The optimal S_{T} threshold values for each mass hypothesis are given

in Table 4.1. While the LQ signal is expected to peak in the mass distribution of the μ-jet pairs, the S_{T} variable gives sufficient power of discrimination in the range of LQ masses considered. However, since the μ-jet mass distribution would nevertheless be important to establish the signal in case an excess is observed, it would be examined in case evidence of an excess is observed. The μ-jet mass distribution is also used in later searches with higher LQ mass reach to improve discriminating power.

M_{LQ}	$S_{\mathrm{T}} >$ (GeV)
200	310
225	350
250	400
280	440
300	440
320	490
340	530
400	560
450	620
500	700

Table 4.1: The S_{T} thresholds for different signal mass hypotheses.

4.3 Background Estimation

In the LQ mass range probed by this search, the largest background is from the $\mathrm{Z}/\gamma^{*}+$ jets process. The initial estimate of this background is taken from simulation, and a data-driven normalization is determined by comparing $\mathrm{Z}/\gamma^{*}+$jets events from data and MC samples in the Z boson peak, $80 < M_{\mu\mu} < 100$ GeV. After selecting two muons and two jets with $p_{\mathrm{T}} > 30$ GeV and applying a preliminary requirement of $S_{\mathrm{T}} > 250$ GeV, the ratio of data to MC events (R_Z) is calculated in the $80 < M_{\mu\mu} < 100$ GeV region. In this region, it is found that $R_Z = 1.28 \pm 0.14$, where the uncertainty is based on the statistical uncertainty in the data and simulation samples. The factor R_Z is applied to

the simulated Z+jets sample for all subsequent stages of the analysis. A comparison of the data and simulation in the $M_{\mu\mu}$ is displayed in Figure 4.1.

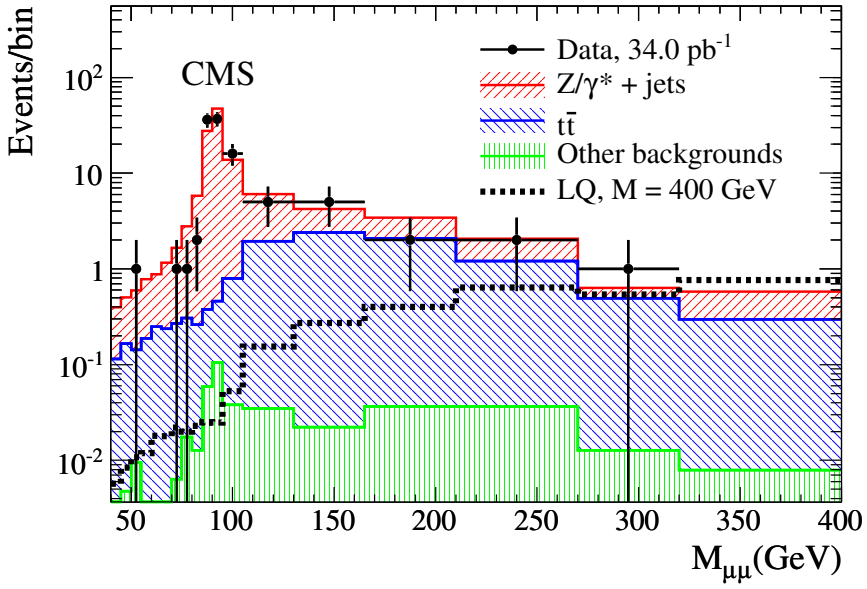

Figure 4.1: The distribution of $M_{\mu\mu}$ after requiring at least two muons and at least two jets with $p_{\mathrm{T}} > 30$ GeV and $S_{\mathrm{T}} > 250$ GeV, The $\mathrm{Z}/\gamma^* \to \mu\mu$ + jets is rescaled by $R_Z = 1.28 \pm 0.14$, as determined by comparing the data and simulation in the $80 < M_{\mu\mu} < 100$ GeV region. Other backgrounds correspond to VV and W+jets. Uncertainties are statistical.

The contribution from $\mathrm{t\bar{t}}$ is estimated with the MC sample, using normalization and uncertainties determined from data [99]. The contribution from W+jets is estimated with MC, and is negligible once the full event selection is applied. The small contribution from VV is estimated from MC. The multijet background is estimated with a multijet-dominated control sample of data containing same-sign dimuon events. Using this method, the multijet contribution is found to be negligible.

Using these background estimations, the data can be compared to the predicted background and LQ signal in the distributions of several variables. Figure 4.2 and Figure 4.3 display the p_{T} and η of the two leading muons and jets, respectively. Fig-

ure 4.4 displays the S_{T} and the invariant masses of the jet-muon pairs ($M_{\mu\mathrm{j}}$) for both pairs in the events. To pair muons to jets, the combination is chosen which minimizes the difference in invariant mass between the two μ-jet pairs. The $M_{\mu\mathrm{j}}$ distribution is particularly important, because a sharp peak would be evident in this distribution should an LQ signal be present. The $M_{\mu\mathrm{j}}$ distribution, after applying the optimized S_{T} for an LQ mass of 300 GeV, is shown in Figure 4.5. Good agreement is found in all distributions, and no excess of events beyond the SM prediction is observed.

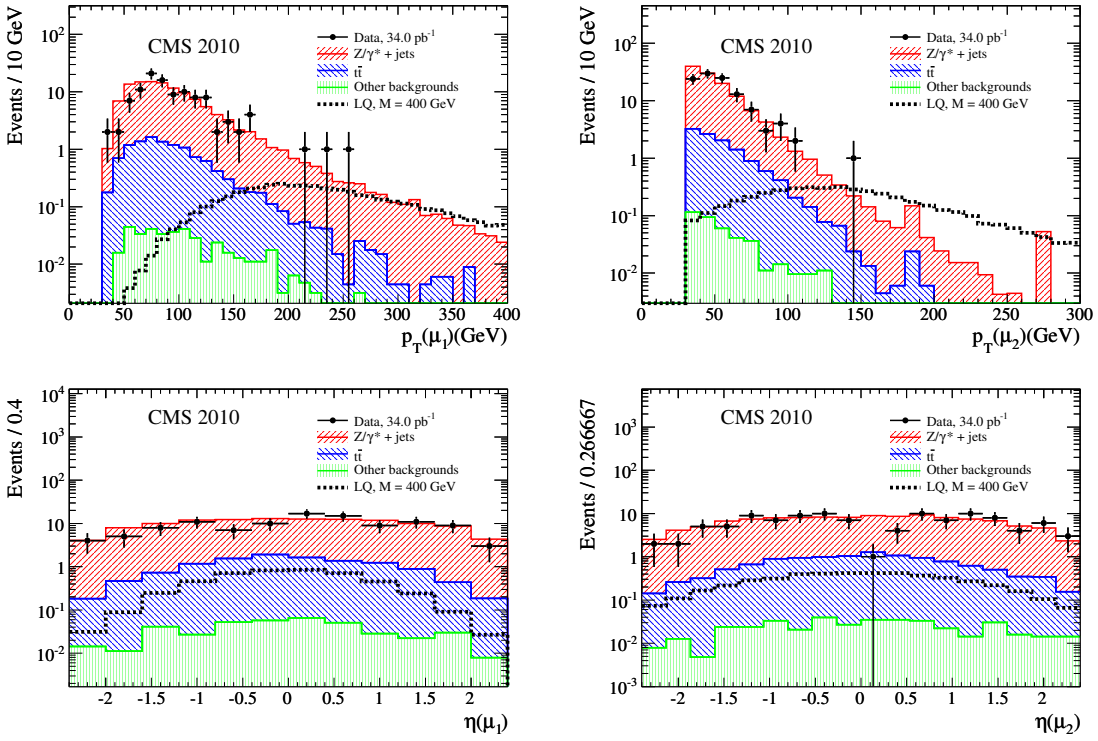

Figure 4.2: The distributions of the p_{T} and η of the two leading muons, after requiring at least two muons and at least two jets with $p_{\mathrm{T}} > 30$ GeV and $S_{\mathrm{T}} > 250$ GeV. Good agreement is observed in all distributions. Other backgrounds correspond to VV and W+jets. Uncertainties are statistical. Overlaid is the signal contribution for an LQ hypothesis mass of 400 GeV.

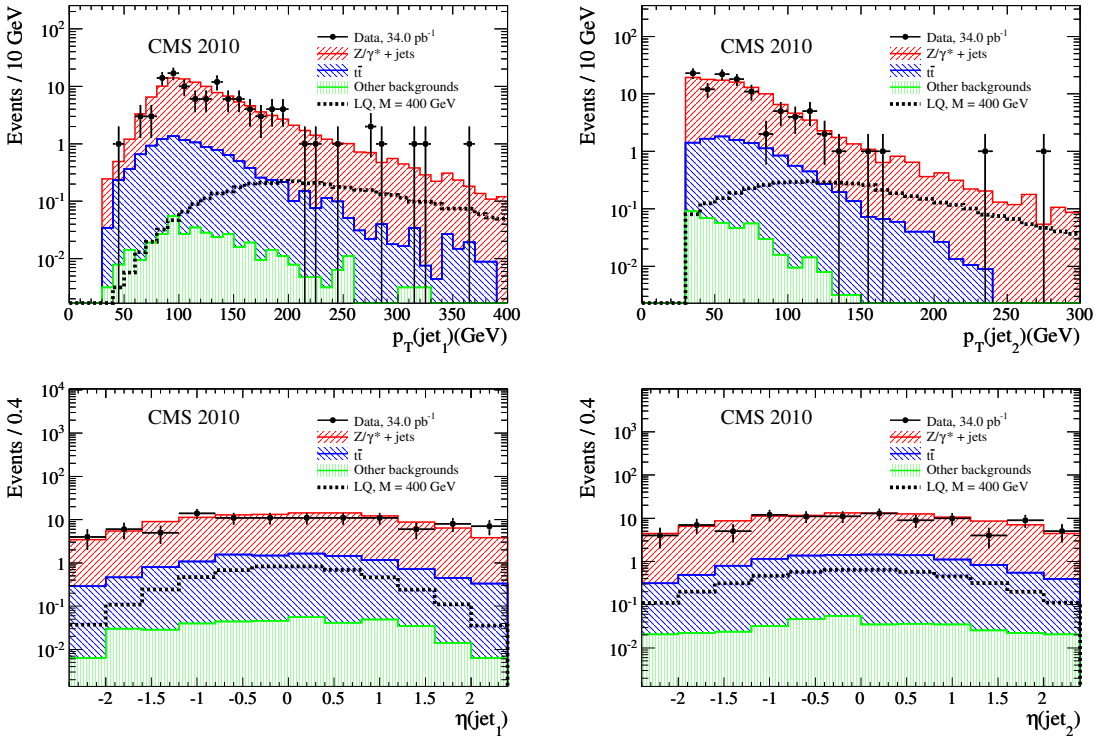

Figure 4.3: The distributions of the p_{T} and η of the two leading jets, after requiring at least two muons and at least two jets with $p_{\mathrm{T}} > 30$ GeV and $S_{\mathrm{T}} > 250$ GeV. Good agreement is observed in all distributions. Other backgrounds correspond to VV and W+jets. Uncertainties are statistical. Overlaid is the signal contribution for an LQ hypothesis mass of 400 GeV.

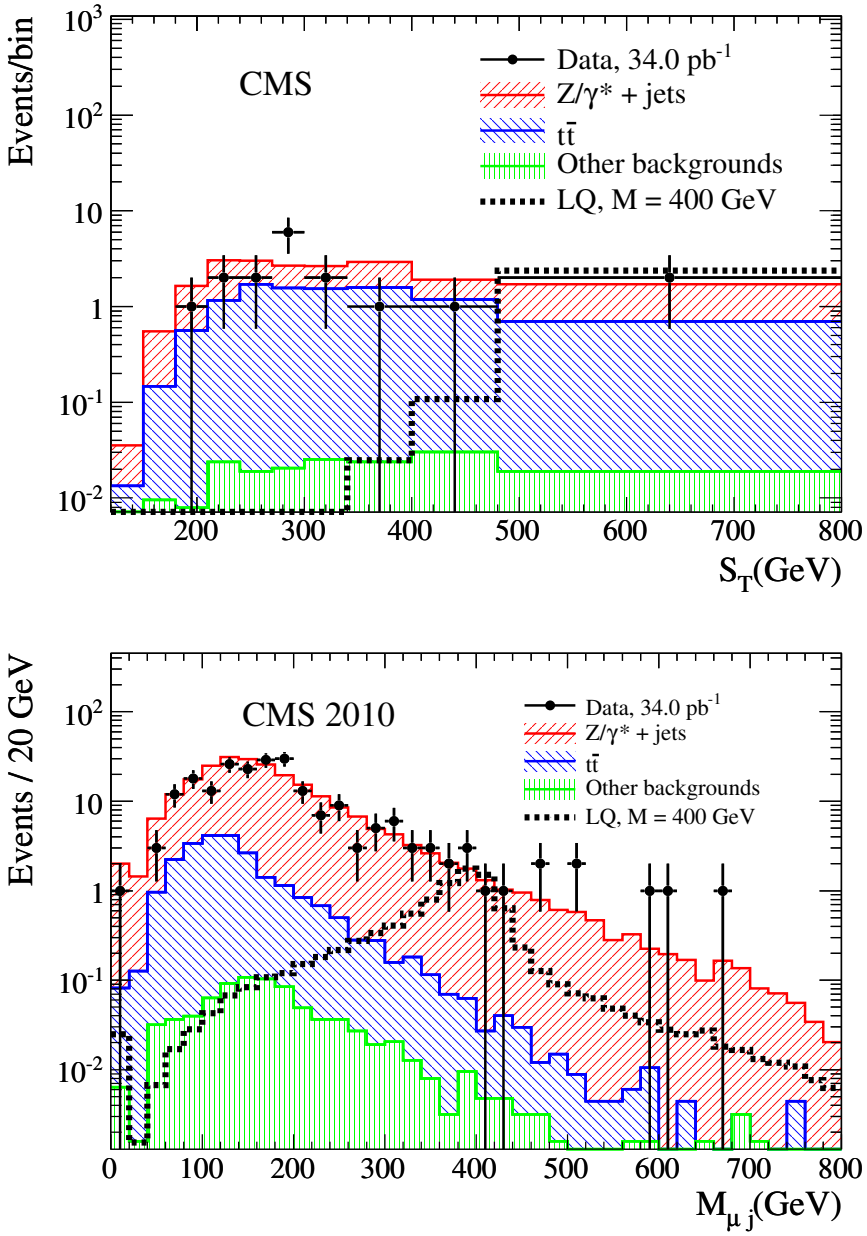

Figure 4.4: The S_{T} and the invariant masses of the jet-muon pairs for both pairs in the events. To pair muons to jets, the combination is chosen which minimizes the difference in invariant mass between the two μ-jet pairs. Other backgrounds correspond to VV and W+jets. Uncertainties are statistical. Overlaid is the signal contribution for an LQ hypothesis mass of 400 GeV.

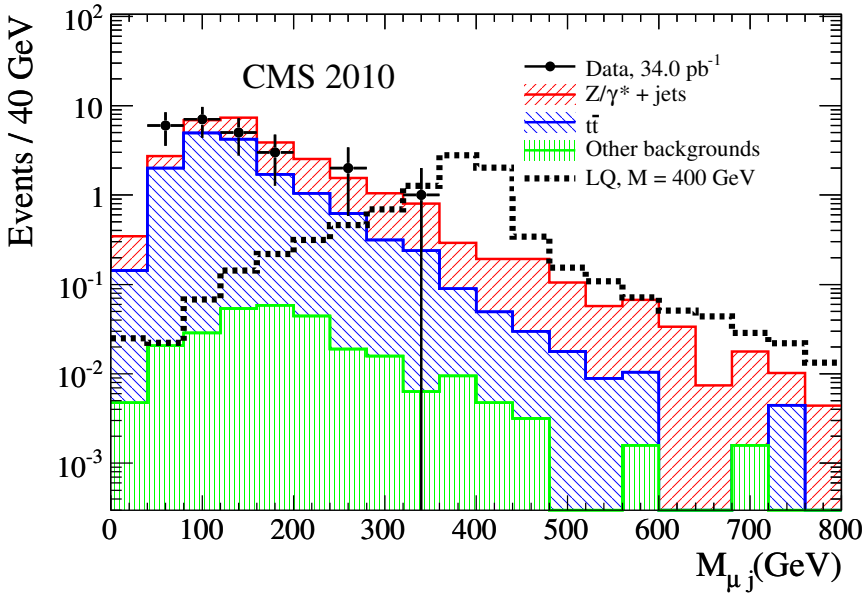

Figure 4.5: The $M_{\mu \mathrm{j}}$ after the final selection S_{T} cut is applied for an LQ mass of 300 GeV. No apparent signal is observed. Other backgrounds correspond to VV and W+jets. Uncertainties are statistical.

4.4 Systematic Uncertainties

Several sources of systematic uncertainties are considered in this analysis. Since the total uncertainty is dominated by statistical uncertainty, conservative uncertainties are assigned to the systematic variations. The systematic uncertainties, their magnitude, and the relative impact on the number of signal and background events are summarized in Table 4.2.

Uncertainties due to the energy and momentum scales of the jets and leptons are assessed by varying the jet energies and muon momentum by nominal values. A 5% systematic uncertainty is assigned to the jet energy scale (JES) [100] of each jet. A smaller, $\sim$ 1% systematic uncertainty comes from the muon momentum scale. The effect of these variations on the expected signal are small. The 300 GeV LQ signal efficiency changes by 2% and 1% due to JES and muon momentum scale uncertain-

ties, respectively. The effect of the muon momentum scale uncertainty on the total background is estimated to be $< 0.5\%$. The JES contributes 2% to the estimate of the Z/γ^*+jets background described above and 15% to the estimate of the VV background from MC. The effect of jet energy and muon momentum resolution on expected signal and backgrounds is found to be negligible.

Additional uncertainties pertain to the normalizations of the simulated samples. The uncertainty on the integrated luminosity is taken as 11% [101]. The uncertainty on the value of R_Z is calculated by considering the statistical uncertainty on the data and simulated samples in the Z mass window. This uncertainty is found to be 11%, and is used as an uncertainty on the estimated $Z/\gamma^* +$ jets background. Additionally, an uncertainty of 16% is assigned on the shape of the Z/γ^*+jets background by comparing the number of $Z/\gamma^* +$ jets events surviving final S_T cut selections in MadGraph samples with factorization/renormalization scales and jet-parton matching thresholds varied by a factor of two. A 41% systematic uncertainty is taken from the CMS measurement of the $t\bar{t}$ production cross section [99] and assigned to the estimate of the $t\bar{t}$ background; it includes the effect of JES on the estimate of the $t\bar{t}$ background.

A 5% systematic uncertainty per muon is assigned due to differences in reconstruction, identification, trigger, and isolation efficiencies between data and MC [102], resulting in a 10% uncertainty on the efficiency of selecting events with two muons both for the signal and background processes. A theoretical uncertainty on the LQ signal production cross sections due to the choice of renormalization/factorization scales is calculated by varying the scales between half and twice the LQ mass, and is found to be 14-15% for LQ masses between 200 and 500 GeV. The effect on the signal acceptance of additional jets generated via initial and final state radiation is found to be less than 1%.

The 90% C.L. PDF uncertainties on LQ cross section is obtained using the

CTEQ6.6 [103] PDF error set following a standard prescription and have been found to vary from 8 to 22% for leptoquarks in the mass range of 200 - 500 GeV [26]. The effect of PDF uncertainties is less than 0.5% on signal acceptance. The PDF uncertainties are not considered for background sources with uncertainties determined from data.

Systematic Uncertainty	Magnitude	Effect on Signal	Effect on Background
JES	5%	2%	–
JES & Data Backgr. Est.	–	–	26%
Muon Momentum Scale	1%	1%	$< 0.5\%$
Muon Pair Reco/ID/Iso	10%	10%	$< 0.05\%$
Integrated Luminosity	11%	11%	–
Total		15%	26%

Table 4.2: Systematic uncertainties and their effects on number of signal and background events.

4.5 Results

The event yields from data, expected LQ signal (for several mass hypotheses), signal selection efficiency times acceptance, and expected SM backgrounds are summarized in Table 4.3.

One candidate event survives the full selection criteria corresponding to a leptoquark mass hypothesis of 400 GeV, and no candidates survive for criteria corresponding to masses greater than 450 GeV. An upper limit on the LQ cross section is set using a Bayesian method [97, 98] with a flat signal prior probability distribution. A log-normal probability density function is used to integrate over the systematic uncertainties. Using Poisson statistics, a 95% confidence level (C.L.) upper limit is obtained on $\sigma \times \beta^2$. This is shown in Fig. 4.6 together with the NLO predictions for the scalar

M_{LQ} (GeV)	MC Signal Samples		Monte Carlo Background Samples				Events in
	Selected	Acceptance	Selected Events in				
	Events	× Efficiency	$t\bar{t}$ + jets	Z+jets	Others	All	Data
200	160±20	0.388±0.003	4.6±0.1	4.08±0.07	0.1±0.01	8.8±0.2	5
225	89±9	0.421±0.003	3.1±0.1	2.99±0.05	0.07±0.01	6.2±0.1	3
250	51±5	0.437±0.003	1.88±0.09	1.92±0.04	0.051±0.009	3.9±0.1	3
280	28±3	0.467±0.003	1.15±0.07	1.53±0.03	0.038±0.008	2.72±0.08	3
300	21±2	0.518±0.004	1.15±0.07	1.53±0.03	0.038±0.008	2.72±0.08	3
320	14±1	0.509±0.004	0.64±0.05	1.12±0.02	0.019±0.005	1.78±0.06	2
340	9±1	0.508±0.003	0.4±0.04	0.79±0.01	0.01±0.004	1.20±0.04	1
400	4.0±0.4	0.578±0.004	0.31±0.04	0.67±0.01	0.01±0.004	0.99±0.04	1
450	1.9±0.2	0.600±0.004	0.19±0.03	0.49±0.01	0.006±0.003	0.69±0.03	0
500	0.9±0.1	0.602±0.004	0.09±0.02	0.277±0.006	0.003±0.002	0.37±0.02	0

Table 4.3: The data event yields in 34 pb^{-1} for different leptoquark mass hypotheses, together with the optimized S_T threshold values (in GeV) for each mass, background predictions, number of expected LQ signal events (S), and signal selection efficiency times acceptance (ϵ_S). M_{LQ} and S_T values are listed in GeV. The $Z/\gamma^* \to \mu\mu$ + jets and $t\bar{t}$ contributions are rescaled by the normalization factors determined from data. Other backgrounds correspond to VV, W+jets, and multijet processes. Uncertainties are from MC statistics.

LQ pair production cross section. The 95% C.L. exclusion on β as a function of LQ mass is also shown in Fig. 4.6. The systematic uncertainties reported in Table 4.2 are included in the calculation as nuisance parameters. With the assumption that $\beta = 1$, second-generation scalar leptoquarks with masses less than 394 GeV are excluded at 95% C.L., 78 GeV higher than the limit set at the D0 Experiment at the Tevatron [34]. This is in agreement with the expected limit of 394 GeV. The corresponding observed limit on cross section is 0.223 pb. If the lower edge of the theoretical $\sigma \times \beta^2$ curve is used, the observed (expected) limit on LQ mass is 379 (377) GeV and the observed limit on cross section is 0.224 pb.

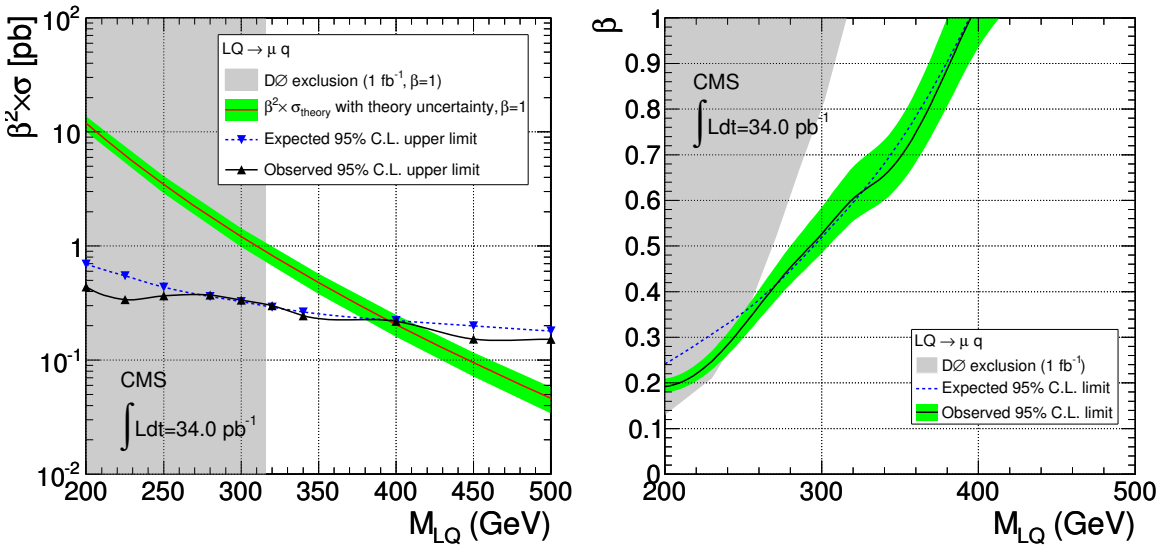

Figure 4.6: (Left) The expected and observed 95% C.L. upper limit on the scalar leptoquark pair production cross section multiplied by β^2 as a function of the LQ mass, together with the NLO theoretical cross section curve. The shaded band on the theoretical values includes PDF uncertainties and the error on the leptoquark production cross section due to renormalization and factorization scale variation by a factor of two. The shaded region is excluded by the current D0 limits [34]. (Right) The minimum β for 95% C.L. exclusion of the leptoquark hypothesis as a function of leptoquark mass. The observed limit and corresponding uncertainty band is obtained by considering the observed upper limit and theoretical branching ratio and its uncertainty in the left-hand figure. Note: The shaded area excluded by the D0 experiment was determined with combined information from the decay channel with two muons and two jets and the decay channel with one muon, missing transverse energy, and two jets.

Chapter 5

Leptoquark Searches with the Full CMS 7 TeV and 8 TeV Datasets

The pair production of second-generation scalar LQs at the LHC would be observable in three possible final states. These final states consist of two jets and either two muons ($\mu\mu jj$), one muon and one neutrino ($\mu\nu jj$), or two neutrinos ($\nu\nu jj$). This Chapter will present searches in the $\mu\mu jj$ and $\mu\nu jj$ final states. The searches presented here [104, 105] are performed with pp collision data from the CMS detector taken in 2011 and 2012, corresponding to $5\,\text{fb}^{-1}$ at $\sqrt{s} = 7$ TeV and $20\,\text{fb}^{-1}$ at $\sqrt{s} = 8$ TeV.

This Chapter proceeds as follows: Section 5.1 describes the selection of muons and jets, as well as the initial selection of events containing final-state objects characteristic of LQ pair production in the $\mu\mu jj$ and $\mu\nu jj$ channels. Section 5.2 provides details on the simulated samples used to estimate signal and background, and the methods used to improve or fully model the different backgrounds using real data. Section 5.3 contains event distributions of the data, the SM background, and the simulated LQ signal at the intial selection levels. Section 5.4 describes optimization of event selection for each LQ mass hypothesis, and Section 5.5 describes the systematic uncertainties on the signal

and background predictions. Lastly, Section 5.6 gives the distributions and final event counts at the optimized selection levels, as we as limits on the LQ pair production cross section as a function of mass and of the branching fraction β.

5.1 Event Selections

5.1.1 Object Identification

Muons are selected according to the "Tight" and "High-p_{T}" identification criteria described in Section 3.5.1. Muons are required to pass a minimum p_{T} threshold, and to be reconstructed in the fiducial region used for the high-level trigger, i.e. with $|\eta| < 2.1$. To reduce contamination from jets that are misreconstructed as muons, a relative isolation requirement is placed on the muons. The tracker-only relative isolation, described in Section 3.5.1, is applied, with a threshold at 0.1. This isolation is robust and well-modeled in simulation, even for very high-p_{T} muons, since it is unaffected by bremsstrahlung in the outer layers of the detector.

Jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ are reconstructed using a particle-flow algorithm described in Section 3.5.1 with the "Loose" jet identification requirements applied. The distance parameter used for the jet clustering is $R = 0.5$. The jet energy is calibrated using p_T balance of dijet and γ+jet events [81], and the corrections to the momentum of each jet are propagated into the measurement of the $E_{\mathrm{T}}^{\mathrm{miss}}$ to improve the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution. Jets are required to have $|\eta| < 2.4$, and a spatial separation from muon candidates of $\Delta R > 0.3$.

5.1.2 Initial Event Selection

Before determining and applying optimized event selection for each LQ hypothesis mass, events are initially selected to contain the appropriate final-state objects, while retaining sufficient amounts of the major backgrounds to conduct detailed studies of the background estimation. The initial event selection, or "preselection", is a set of kinematic thresholds on the identified muons, jets, and $E_{\mathrm{T}}^{\mathrm{miss}}$, specific to the $\mu\mu jj$ and $\mu\nu jj$ final states. The 2011 and 2012 CMS datasets differ in both energy and integrated luminosity, and it is therefore reasonable to utilize more stringent preselection requirements in the 2012 dataset.

The preselection thresholds can be summarized as follows:

1. Lower limits on the transverse momenta of the final-state objects characteristic of LQ pair production: In the $\mu\mu jj$ channel, this includes the two leading muons and two leading jets. In the $\mu\nu jj$ channel, this includes the leading muon, the missing transverse energy, and the two jets.

2. Maximum values for the magnitude of the muon and jet pseudorapidities, ensuring that the muons and jets are in the fiducial region of the detector necessary to meet identification and reconstruction quality definitions.

3. A minimum value of the ΔR between the jets and the muon(s). Individual jets failing this condition are disregarded, but the the event itself may be retained.

4. A minimum value of the $\Delta\phi$ between the $E_{\mathrm{T}}^{\mathrm{miss}}$ and the leading jet, reducing the possible effects of artificially large $E_{\mathrm{T}}^{\mathrm{miss}}$ caused by jet p_{T} mismeasurement. This only pertains to the $\mu\nu jj$ channel.

5. A minimum value of the $\Delta\phi$ between the $E_{\mathrm{T}}^{\mathrm{miss}}$ and the muon, reducing the W+jets and QCD multijet backgrounds, and reducing the effects of muon p_{T}

mismeasurement on the $E_{\mathrm{T}}^{\mathrm{miss}}$. This only pertains to the $\mu\nu jj$ channel.

6. Minimum thresholds on $M_{\mu\mu}$ in the $\mu\mu jj$ channel, and on $M_{\mathrm{T}}^{\mu\nu}$ in the $\mu\nu jj$ channel.

7. Minimum threshold on the S_{T}, defined as scalar sum of the p_{T} of the four final state objects. In the $\mu\mu jj$ channel $S_{\mathrm{T}}^{\mu\mu} \equiv p_{\mathrm{T}\mu 1}+p_{\mathrm{T}\mu 2}+p_{\mathrm{T}j1}+p_{\mathrm{T}j2}$, and in the $\mu\nu jj$ channel, $S_{\mathrm{T}}^{\mu\nu} \equiv p_{\mathrm{T}\mu 1} + E_{\mathrm{T}}^{\mathrm{miss}} + p_{\mathrm{T}j1} + p_{\mathrm{T}j2}$. The S_{T} variable is a powerful discriminating variable used in LQ pair production searches, and a minimum threshold is placed to give a set of preselected events with high p_{T} objects, but that would still be heavily dominated by background process and useful for studies of the SM backgrounds.

Table 5.1 details the preselection thresholds in both the 7 TeV and 8 TeV searches, for the $\mu\mu jj$ and $\mu\nu jj$ channels.

5.2 Signal and Background Estimation

5.2.1 Overview of Simulated Samples and Normalizations

The estimation of the scalar LQ signal is performed with Pythia for LQ mass hypotheses from 250 GeV to beyond 1 TeV. The LQ signal samples are normalized to an NLO cross-section, calculated with the methods described in Section 1.2.2. Estimations of the SM backgrounds are determined with simulation. The list of MC generators and cross-sections used for background MC simulation are given in Table 5.2. The cross sections used to normalize the simulations of the background processes are the state-of-the-art calculations at the time the analysis was conducted. With few exceptions, all calculations are performed at next-to-leading order or better. For the vector boson production in association with one or more jets, the contributions are modeled with

	$5\,\text{fb}^{-1}$ at $\sqrt{s} = 7$ TeV		$20\,\text{fb}^{-1}$ at $\sqrt{s} = 8$ TeV	
Quantity	$\mu\mu jj$ channel	$\mu\nu jj$ channel	$\mu\mu jj$ channel	$\mu\nu jj$ channel
$p_{\text{T}}(\mu_1)$ [GeV]	> 40	> 40	> 45	> 45
$p_{\text{T}}(\mu_2)$ [GeV]	> 40	—	> 45	—
$\lvert\eta(\mu_1)\rvert$	< 2.1	< 2.1	< 2.1	< 2.1
$\lvert\eta(\mu_2)\rvert$	< 2.1	—	< 2.1	—
$p_{\text{T}}(j_1)$ [GeV]	> 30	> 40	> 125	> 125
$p_{\text{T}}(j_2)$ [GeV]	> 30	> 40	> 45	> 45
$\lvert\eta(j_1)\rvert$	< 2.4	< 2.4	< 2.4	< 2.4
$\lvert\eta(j_2)\rvert$	< 2.4	< 2.4	< 2.4	< 2.4
$\Delta R(\mu, j)$	> 0.3	> 0.3	> 0.3	> 0.3
$E_{\text{T}}^{\text{miss}}$ [GeV]	—	> 55	—	> 55
$\lvert\Delta\phi(E_{\text{T}}^{\text{miss}}, j_1)\rvert$	—	> 0.5	—	> 0.5
$\lvert\Delta\phi(E_{\text{T}}^{\text{miss}}, \mu)\rvert$	—	> 0.8	—	> 0.8
$M_{\mu\mu}$ [GeV]	> 50	—	> 50	—
$M_{\text{T}}^{\mu\nu}$ [GeV]	—	> 50	—	> 50
$S_{\text{T}}^{\mu\mu}$ [GeV]	> 250	—	> 300	—
$S_{\text{T}}^{\mu\nu}$ [GeV]	—	> 250	—	> 300

Table 5.1: The preselection requirements applied to events in the searches performed with the 2011 and 2012 CMS pp collision data, corresponding to $5\,\text{fb}^{-1}$ at $\sqrt{s} = 7$ TeV and $20\,\text{fb}^{-1}$ at $\sqrt{s} = 8$ TeV.

MadGraph and Sherpa , including up to four real parton emissions, and are normalized to NNLO cross-sections calculated with FEWZ [106]. The $\text{t}\bar{\text{t}}$ contribution is modeled with MadGraph , and normalized to cross sections determined in Ref [107] (7 TeV) and Ref [108] (8 TeV). In the 8 TeV $\text{t}\bar{\text{t}}$ simulation, the fully-leptonic, semi-leptonic, and fully-hadronic processes are simulated separately. Diboson samples are modeled with Pythia , and single top-quark samples are modeled with Powheg. All diboson samples and single top-quark samples use NLO cross section calculations from MCFM [109–112], with the exception of the 8 TeV single-top quark cross sections,

taken from Ref [107]. Samples of the Z+jets, W+jets, and $t\bar{t}$ processes are normalized to data in control regions. The QCD multijet sample is enriched to contain final states with identified muons, and normalized to an LO cross section from PYTHIA .

	$5\,\text{fb}^{-1}$ at $\sqrt{s} = 7$ TeV		$20\,\text{fb}^{-1}$ at $\sqrt{s} = 8$ TeV	
Process	MC Generator	σ[pb]	MC Generator	σ[pb]
W+jets	SHERPA	31314	MADGRAPH	36257
Z+jets	SHERPA	3048	MADGRAPH	3504
$t\bar{t}$ (inclusive)	MADGRAPH	165	—	—
$t\bar{t}$ (leptonic decays)	—	—	MADGRAPH	26.2
$t\bar{t}$ (semileptonic decays)	—	—	MADGRAPH	103.7
$t\bar{t}$ (full hadronic deays)	—	—	MADGRAPH	104.1
single t-quark (t-channel)	POWHEG	42.6	POWHEG	56.4
single t-quark (s-channel)	POWHEG	2.97	POWHEG	3.79
single t-quark (tW-channel)	POWHEG	5.3	POWHEG	11.1
single $\bar{t}$-quark (t-channel)	POWHEG	22.0	POWHEG	30.7
single $\bar{t}$-quark (s-channel)	POWHEG	1.63	POWHEG	1.76
single $\bar{t}$-quark (tW-channel)	POWHEG	5.3	POWHEG	11.1
WW	PYTHIA	43.0	PYTHIA	54.8
WZ	PYTHIA	18.2	PYTHIA	33.2
ZZ	PYTHIA	5.9	PYTHIA	17.7
QCD, μ-enriched $p_{\text{T}} > 15$ GeV	PYTHIA	3782730	PYTHIA	5636940

Table 5.2: The MC generators used for W+jets signal and SM background simulation, and the corresponding cross sections. The cross sections of the W+jets and Z+jets processes are calculated with FEWZ [106] at NNLO. The 7 TeV $t\bar{t}$ cross section is determined in Ref [107], and the 8 TeV cross section is determined in Ref [108]. All diboson samples and single top-quark samples use NLO cross section calculations from MCFM [109–112], with the exception of the 8 TeV single-top quark cross sections, taken from Ref [107]. Samples of the Z+jets, W+jets, and $t\bar{t}$ processes will be normalized to data in control regions. The QCD multijet sample is normalized to an LO cross section from PYTHIA .

The major background contributions from Z+jets, W+jets, and $t\bar{t}$ processes are normalized in data control regions. In the special case of the $t\bar{t}$ process in the $\mu\mu jj$ channel, the process can be fully estimated with a data sample with an $e\mu jj$ signature. In all analysess, the QCD multijet contamination is studied in control regions and found to be at negligible levels. Section 5.2.2 and Section 5.2.3 detail the methods

used to normalize and model the major backgrounds.

5.2.2 Background Estimation in the $\mu\mu jj$ Channel

The main processes that can mimic the signature of the signal in the $\mu\mu jj$ channel are $Z/\gamma^* + \text{jets}$, $t\bar{t}$, VV+jets, W+jets, and QCD multijets. The contributions from the latter three processes are small once a fully-optimized event selection is applied. This section describes the estimation of the Z/γ^*+jets, $t\bar{t}$, and QCD multijet backgrounds.

Estimation of Z/γ^*+jet Background

The contribution from the principal background, Z+jets, is estimated with MC simulation normalized to data at the initial selection level in the Z-enriched region 80 GeV $< M_{\mu\mu} <$ 100 GeV. In this region, the rescaling factor R_Z is calculated as the sum of all events in data, minus the contributions non-Z backgrounds, divided by the number of Z+jets simulated events, as described in Eqn. 5.1. In the 7 TeV data, where the Z+jets contribution is modeled with SHERPA, a scale factor of $R_Z = 1.27 \pm 0.02$ is calculated. In the 8 TeV data, where the Z+jets contribution is modeled with MADGRAPH, the resulting scale factor is $R_Z = 0.92 \pm 0.01$. The uncertainties are based on the statistical uncertainty in the data and simulation samples used to calculate the ratio. Figure 5.1 displays the distributions of $M_{\mu\mu}$ in data and simulation, after the R_Z rescaling factor has been applied to the Z+jets contribution, in the 8 TeV analysis.

$$R_Z = \left(N_{\text{data}} - N_{t\bar{t}} - N_{VV} - N_{\text{W+jets}} - N_{\text{single-top}}\right) / N_{\text{Z+jets}} \tag{5.1}$$

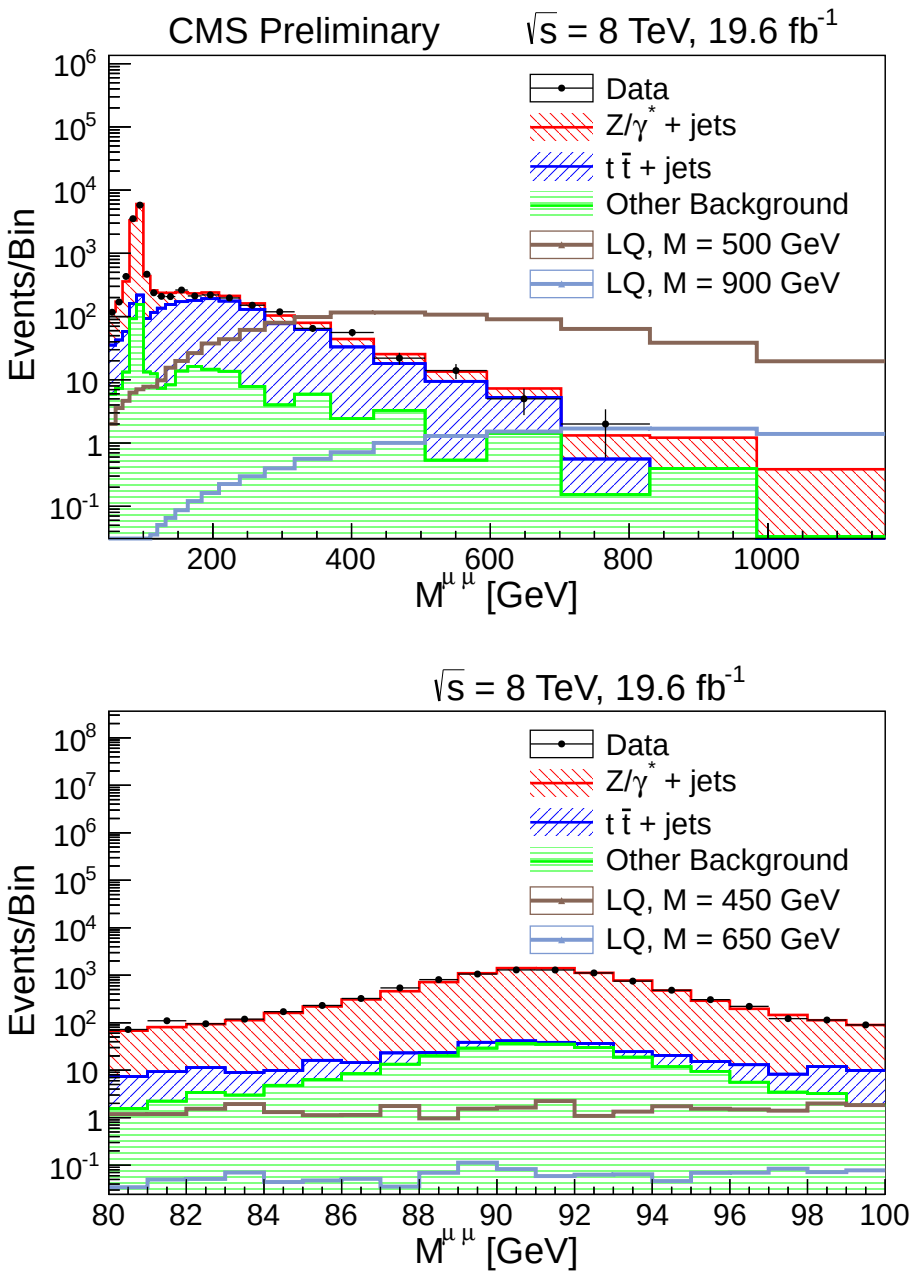

Figure 5.1: The dimuon invariant mass, in events satisfying the 8 TeV preselection criteria in the $\mu\mu jj$ channel. The Z+jets contribution from MadGraph is normalized to data in the Z-enriched mass peak region, 80 GeV $< M_{\mu\mu} <$ 100 GeV. The top depicts a wide range of $M_{\mu\mu}$, and the bottom depicts the $M_{\mu\mu}$ range in which the Z+jets rescaling factor is calculated.

Estimation of the $t\bar{t}$ Background

The $t\bar{t}$ background in this channel can be estimated completely with data, using a sample of events with an $e\mu jj$ final state. Since additional final-selection cuts are applied to several variables, this method does not yet determine any specific distribution, but instead determines a collection of events and appropriate weights which can be treated as a $t\bar{t}$ sample for all subsequent steps of the analysis. The $t\bar{t}$ estimation proceeds as follows:

1. The set of all $e\mu$ events in data and in non-$t\bar{t}$ simulation are gathered, and referred to as $\{\text{Data}^{e\mu}\}$ and $\left\{\text{MC}^{e\mu}_{\text{non-}t\bar{t}}\right\}$.

2. The data events, minus the non-$t\bar{t}$ MC events determined in step [1.], provide a sample which can be used to estimate the $t\bar{t}$ contribution in the $e\mu$ final state: $\left\{t\bar{t}^{\,e\mu}\right\} = \left\{\{\text{Data}^{e\mu}\}, \left\{-\text{MC}^{e\mu}_{\text{non-}t\bar{t}}\right\}\right\}$.

3. Since the efficiency of the muon and electron selection identification and isolation criteria are nonidentical, and since $t\bar{t}$ processes with an $e\mu$ final state are twice as likely as processes with a $\mu\mu$ final state, a factor $R_{\mu\mu,e\mu}$ is needed which accounts for all of of these effects. The factor $R_{\mu\mu,e\mu}$ can be easily calculated in $t\bar{t}$ MC, as the ratio of $\mu\mu$ events surviving preselection to that of $e\mu$ events surviving preselection.

4. Since a single-muon trigger is used when collecting both the $\mu\mu$ and $e\mu$ events, a factor R_{trigger} must be used to account for this effect. If the efficiency of the single-muon trigger for an individual muon is denoted ϵ_μ, then the corresponding efficiency for an event with two muon is $1-(1-\epsilon_\mu)^2 = 2\epsilon_\mu - \epsilon_\mu^2$. Thus, $R_{\text{trigger}} = \left(2\epsilon_\mu - \epsilon_\mu^2\right)/\epsilon_\mu = 2 - \epsilon_\mu$.

5. Lastly, the correction factors from steps [3.] and [4.] can be applied to the

$\left\{t\bar{t}^{e\mu}\right\}$ sample determined in Step [2.], yielding a properly normalized sample representing the $t\bar{t}$ contribution to the $\mu\mu$ final state:
$\left\{t\bar{t}^{\mu\mu}\right\} = \left\{t\bar{t}^{e\mu}\right\} \times R_{\mu\mu,e\mu} \times R_{\text{trigger}}$.

In practice, the factor $R_{\mu\mu,e\mu}$ is found to be slightly larger than 0.5 due to the better identification efficiency of muons, compared to that of electrons. The factor R_{trigger} depends on measurements of the single-muon trigger efficiency, which vary with the muon p_{T} an η.

The comparison between the S_T and $M_{\ell j}$ distributions at preselection for the $e\mu jj$ channel can be seen in the Figure 5.2. It is observed that the estimation of the $t\bar{t}$ $e\mu jj$ background with MC is compatible with the $e\mu jj$ data. While this would indicate that the $t\bar{t}$ contribution is well-modeled by simulation, the data-driven method ensures the best-possible modeling of the $t\bar{t}$ background, and allows for a reduction in the total systematic uncertainties.

Estimation of QCD Multijet Background

An initial estimate of the QCD prediction is performed with the QCD multijet MC, and it is found that no QCD multijet MC events pass the selection criteria applied to optimize the signal selection. This outcome is expected, since the isolation critieria applied to the muons efficiently reject contamination from jets misreconstructed as muons. Thus, it is very unlikely to have two jets miscreconstructed as muons in a single event.

To validate the assumption that contribution to the background from the QCD multijet process is negligible, a data-driven estimation of the QCD multijet background is performed using a QCD-enriched control sample of data events. While most SM processes with a dimuon final state are expected to have opposite-sign muons, jets which are misreconstructed as muons may be reconstructed as same-sign muons in

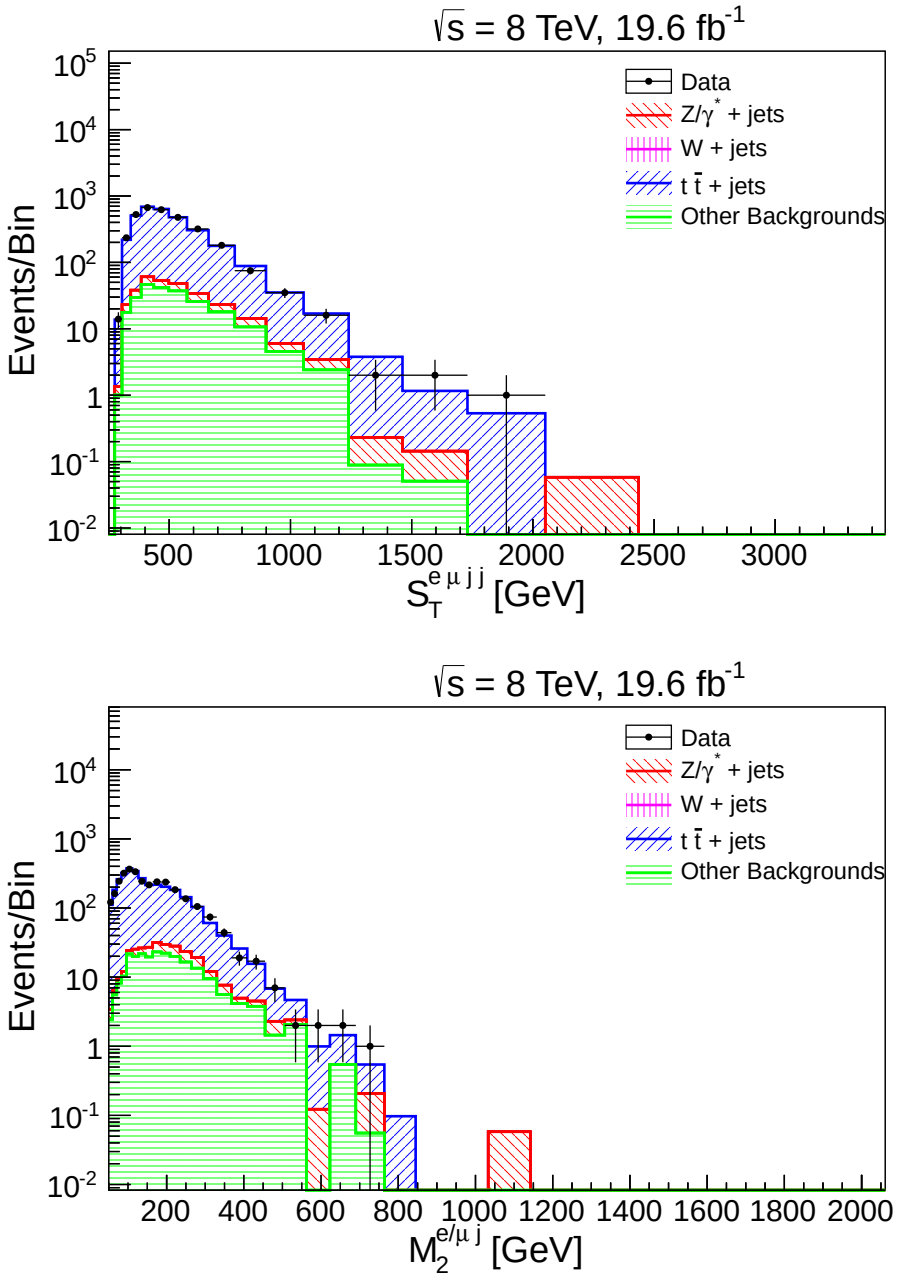

Figure 5.2: Distributions of the S_T and $M_{\ell j}$ for the eμjj channel, for data and MC predictions, with events passing preselection in the $e\mu jj$ channel.

a substantial fraction of events. Thus, a QCD-enriched control sample in data can be obtained by selecting events which have two same-sign muons. To increase the number of events in the QCD-enriched sample, and to greatly increase the purity of this sample, the isolation condition is removed from the muon selection process. Then, the normalization of the same-sign non-isolated sample needs only to be adjusted for the muon isolation efficiency and the same-sign to opposite-sign ratio.

The ratio of same-sign to opposite-sign muons in QCD, as well as the efficiency of the muon isolation, are determined from the QCD multijet MC sample. To do so, it is necessary to rely on the modeling of the muon isolation in the simulation. The relative isolation for the leading and sub-leading muon is depicted in Figure 5.3, and the agreement in the QCD-dominated region is found to be good. For the optimized event selection criteria, to be described in Section 5.4, the number of QCD events (N_{QCD}) can determined in terms of the same sign (N_{SS}^{MC}) and opposite sign (N_{OS}^{MC}) events in the control region in the MC sample, the isolation efficiency in MC (ϵ_{ISO}), and the number of same-sign non-isolated events in data ($N_{SS,Non-Iso}^{D}$) as:

$$N_{QCD} = \frac{N_{SS}^{MC} + N_{OS}^{MC}}{N_{SS}^{MC}} \times \epsilon_{ISO} \times N_{SS,Non-Iso}^{D}. \tag{5.2}$$

This method is applied to searches with both the 7 TeV and 8 TeV data. The estimates of the parameters calculated for this method are given in Table 5.3 for the analysis with the 8 TeV dataset. For all optimized selections in the $\mu\mu jj$ channel at 7 TeV and 8 TeV, the QCD multijets process is less than 1% of the total background prediction. Thus, it is considered negligible for all subsequent parts of the analysis.

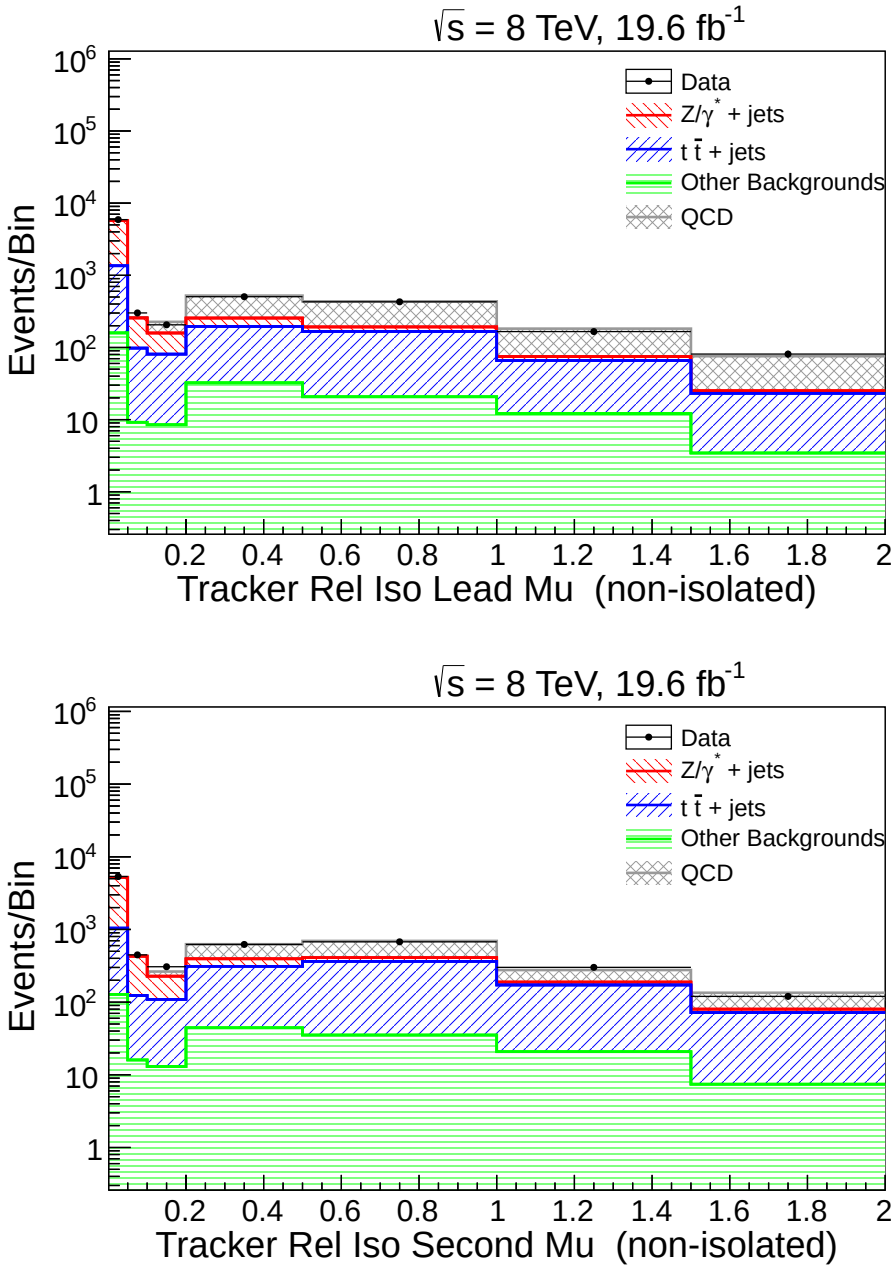

Figure 5.3: The relative tracker isolation for the two leading muons in the non-isolated $\mu\mu jj$ channel.

Parameter	Value
$\frac{N_{SS}^{MC}+N_{OS}^{MC}}{N_{SS}^{MC}}$	2.80 ± 0.33
ϵ_{ISO}	0.203 ± 0.091
ϵ_{ISO}^2	0.041 ± 0.037
$\frac{N_{SS}^{MC}+N_{OS}^{MC}}{N_{SS}^{MC}} \times \epsilon_{ISO}^2$	0.115 ± 0.104

Table 5.3: The parameters used to estimate the QCD multijet contribution to the $\mu\mu jj$ channel in the 8 TeV search, corresponding to Eqn 5.2.

5.2.3 Background Estimation in the $\mu\nu jj$ Channel

The main processes that can mimic the signature of the signal are the production of a W boson in association with jets and $t\bar{t}$ production. Additional small contributions come from single top-quark production, diboson processes (WW, WZ, and ZZ), and instrumental background, caused by the misidentification of jets as muons in QCD multijet processes. For second-generation LQ searches in the $\mu\nu jj$ channel, all background processes are modeled with simulation. The normalizations of the W+jets and $t\bar{t}$ simulations are adjusted after comparing data to simulation in control regions. Contributions from smaller backgrounds are estimated directly with simulation, and the QCD multijet contribution is shown to be negligible. This section describes these background estimations.

Estimation of W+jets and $t\bar{t}$ Backgrounds

The background contributions from W+jets and $t\bar{t}$ processes can be estimated with simulation. Because the $\mu\nu jj$ preselection is background-dominated, control regions in this selection can be used to correct the normalizations of the simulations to data, as long as these control regions are removed from the final event selection. The transverse

mass between the muon and $E_{\mathrm{T}}^{\mathrm{miss}}$ is defined as $M_{\mathrm{T}} = \sqrt{2p_{\mathrm{T}\mu}E_{\mathrm{T}}^{\mathrm{miss}}(1-\cos(\Delta\phi))}$, where $\Delta\phi$ is the difference in azimuthal angle between the muon momentum direction and the $E_{\mathrm{T}}^{\mathrm{miss}}$ direction. The transverse mass window $70 < M_{\mathrm{T}} < 110$ GeV is used to determine both the W+jets and the $\mathrm{t\bar{t}}$ normalization factors, using two mutually exclusive selections.

To accomplish the simultaneous normalization of the W+jets and $\mathrm{t\bar{t}}$ samples, the contributions from data and simulation in the M_{T} window are separated into a sub-sample with three or fewer reconstructed jets, and a sub-sample with four or more reconstructed jets. This method is viable because the contribution from W+jets decreases rapidly with jet multiplicity, as discussed in Section 1.3.2. However, the $\mathrm{t\bar{t}}$ contribution is expected to have higher jet multiplicities, because the semileptonic $\mathrm{t\bar{t}}$ decays contain two b-quark jets, and two jets from a hadronic decay of a W boson.

In the two control regions in the W transverse mass window, the data and simulation event yields can be used to form a system of equations which determine the W rescaling (R_{W}) and $\mathrm{t\bar{t}}$ rescaling ($R_{\mathrm{t\bar{t}}}$) :

$$\begin{aligned} N_1 &= R_{\mathrm{t\bar{t}}}N_{1,\mathrm{t\bar{t}}} + R_{\mathrm{W}}N_{1,\mathrm{W}} + N_{1,O}; \\ N_2 &= R_{\mathrm{t\bar{t}}}N_{2,\mathrm{t\bar{t}}} + R_{\mathrm{W}}N_{2,\mathrm{W}} + N_{2,O}. \end{aligned} \tag{5.3}$$

Here, N_i, $N_{i,\mathrm{W}}$, $N_{i,\mathrm{t\bar{t}}}$, and $N_{i,O}$ are the number of events in data, W+jets, $\mathrm{t\bar{t}}$, and other MC backgrounds passing selection i. Details of these parameters in the 8 TeV search are given in Table 5.4.

The M_{T} peak region, in which the $\mathrm{t\bar{t}}$ and W+jets rescaling factors are calculated, is shown in Figure 5.4, with the $\mathrm{t\bar{t}}$-enriched and W+jets-enriched regions depicted separately.

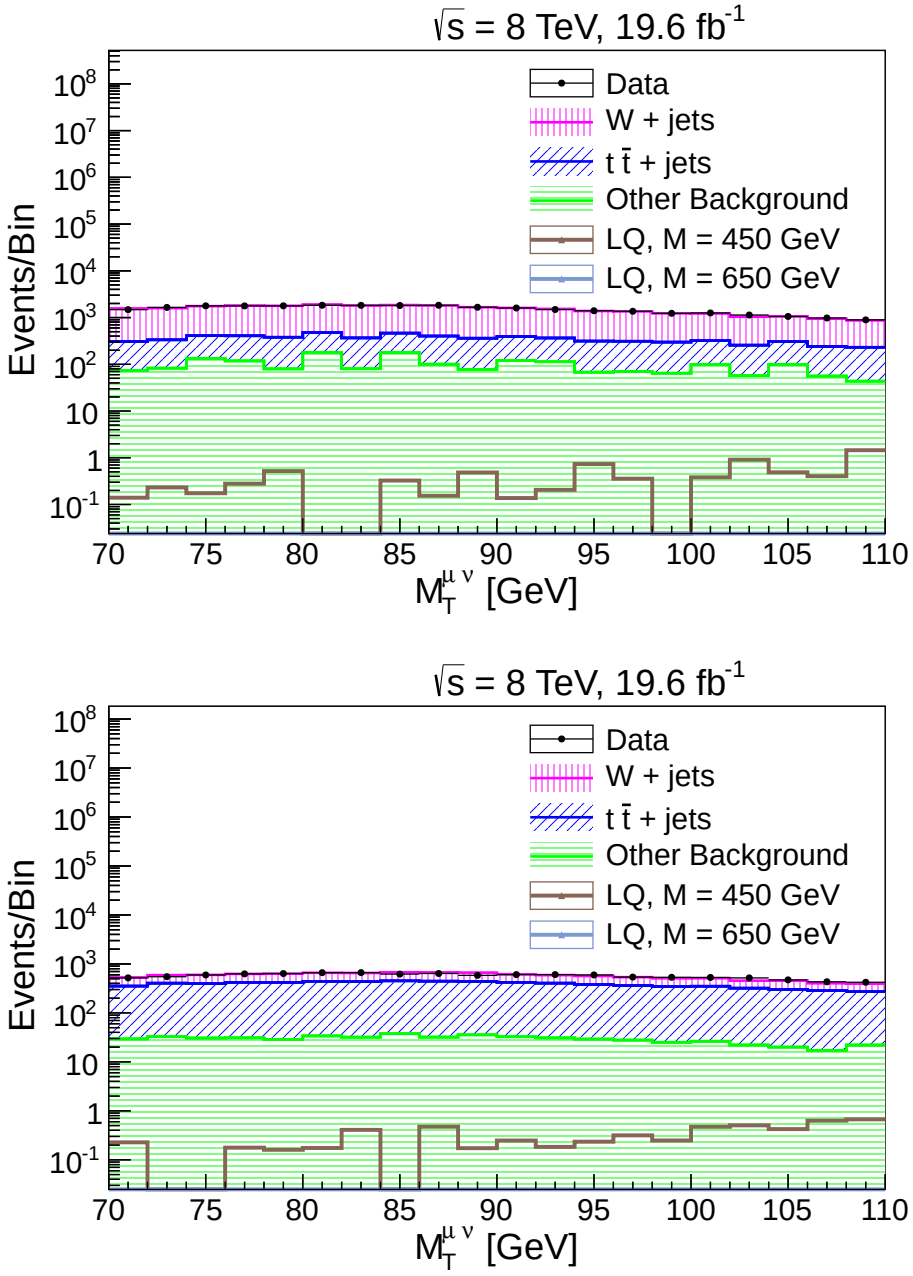

Figure 5.4: The transverse mass peak region, $70 < M_{\mathrm{T}} < 110$ GeV, used for the determination of the rescaling factors for the W+jets and $\mathrm{t\bar{t}}$ simulation. Rescaling factors have been applied. (Top) Events with three reconstructed jets or fewer, dominated by W+jets processes. (Bottom) Events with four or more jets, dominated by $\mathrm{t\bar{t}}$ production.

Parameter	Value
N_1	29807
N_2	11346
$N_{1,t\bar{t}}$	5122.8 ± 29.6
$N_{2,t\bar{t}}$	7166.3 ± 35.8
$N_{1,W}$	23947.0 ± 141.9
$N_{2,W}$	3871.1 ± 41.2
$N_{1,0}$	1934.5 ± 149.3
$N_{2,0}$	588.4 ± 13.6
$R_{t\bar{t}}$	0.99 ± 0.02
R_{W}	0.95 ± 0.01

Table 5.4: The parameters used to estimate the W+jets and $t\bar{t}$ rescaling factors in the $\mu\nu jj$ channel in the 8 TeV search.

Multijet Background

The contribution from QCD multijet processes after all selection criteria are applied is expected to be negligible. No events in the simulated QCD multijet sample survive the full selection criteria optimized for any leptoquark mass hypothesis. However, as multijet processes are difficult to accurately model by MC simulation, the QCD multijet contribution is further studied to ensure that the QCD multijet background is negligible for this search.

A QCD multijet-enriched data sample is obtained by requiring events with $E_{\mathrm{T}}^{\mathrm{miss}}$ <10 GeV and exactly one muon that passes all muon identification requirements, but imposing no requirement on the muon isolation. In this QCD-dominated

region, we first derive a global rescaling factor for the QCD MC. The agreement between data and MC in this region is assessed to be good using a comparison before and after the global rescaling, as displayed in Figure 5.5. Using the data in this region, a ratio of multijet muons that satisfy the tracker-relative isolation requirement to the total number of multijet muons (the "fake-rate") is calculated. Then, the global scale factor and the data-driven fake-rate are applied to the non-isolated QCD MC sample at each final selection.

An example of this calculation in the 8 TeV data is as follows: In the non-isolated low-$E_{\mathrm{T}}^{\mathrm{miss}}$ region, the global rescaling factor applied to the QCD MC is calculated to be $R_{\mathrm{QCD}} = 1.22 \pm 0.03$. The muon isolation fake rate is calculated to be: $f_{\mathrm{Iso}} = 0.086 \pm 0.004$. Applying the product of these factors, $f_{\mathrm{Iso}} \times R_{\mathrm{QCD}}$, to the non-isolated QCD MC at the optimized LQ final selections, the contribution from QCD is estimated to be always less than 2%, and to decrease with LQ mass. Beyond LQ masses of 600 GeV, the QCD contamination is less than 1%.

According to these studies, the contribution to the total background prediction from QCD multijet processes is considered negligible in all subsequent part of the search.

5.3 Distributions of Data and Background at Preselection

5.3.1 Preselection Distributions of the 7 TeV Data

The distributions of the p_{T} of the two muons and two jets selected by the $\mu\mu jj$ preselection in the 7 TeV search are shown in Figure 5.6. Figure 5.7 displays the S_{T} and muon-jet invariant mass distributions in the $\mu\mu jj$ channel. Good agreement between the data and the SM background prediction is observed. The distribution in scalar LQ

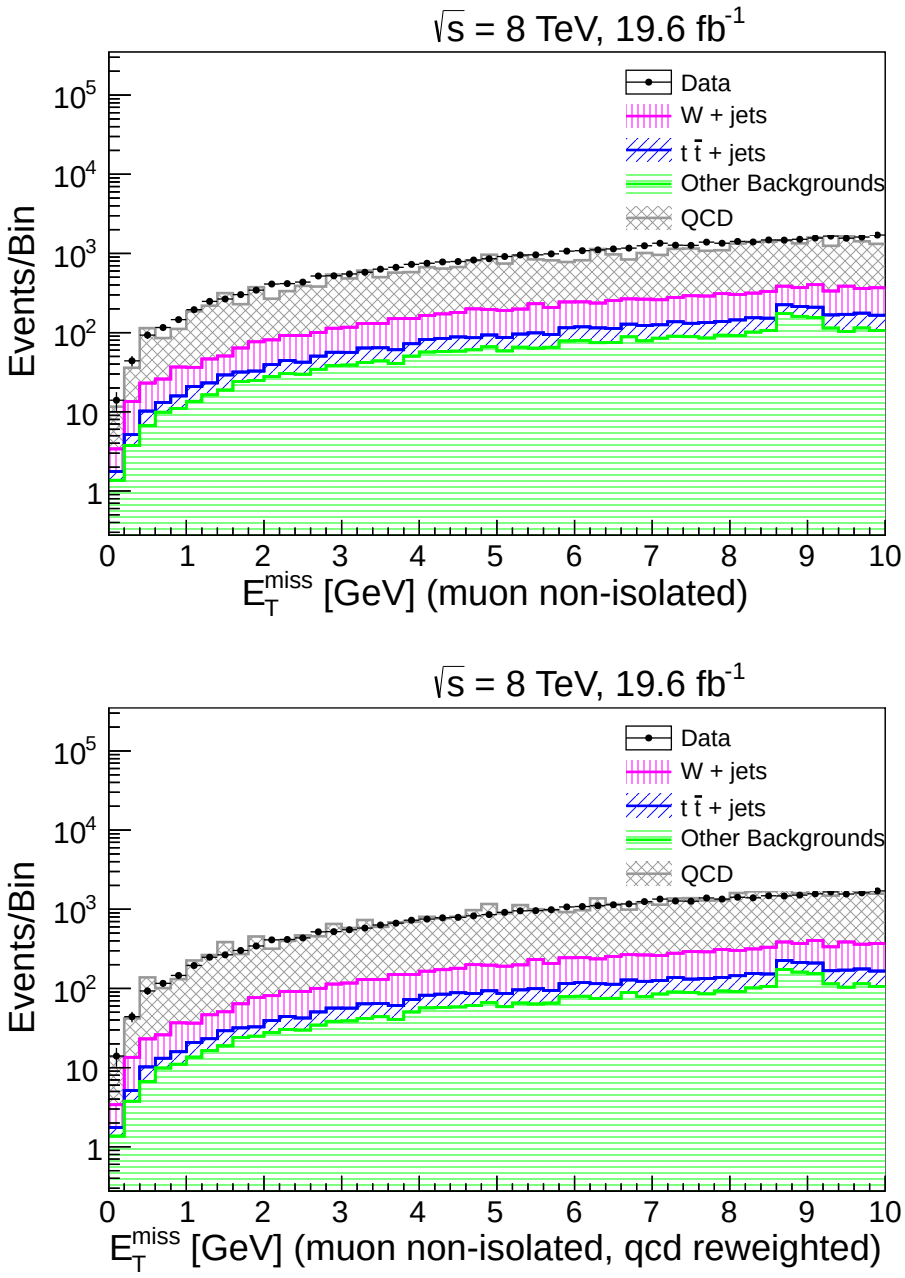

Figure 5.5: The QCD-enriched low $E_{\mathrm{T}}^{\mathrm{miss}}$ region, with non-isolated muons. (Top) The $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions before the global normalization is applied to the QCD simulation, and (Bottom) a after the global normalization correction is applied to the QCD MC.

simulation is overlaid, for an LQ mass hypothesis of 400 GeV.

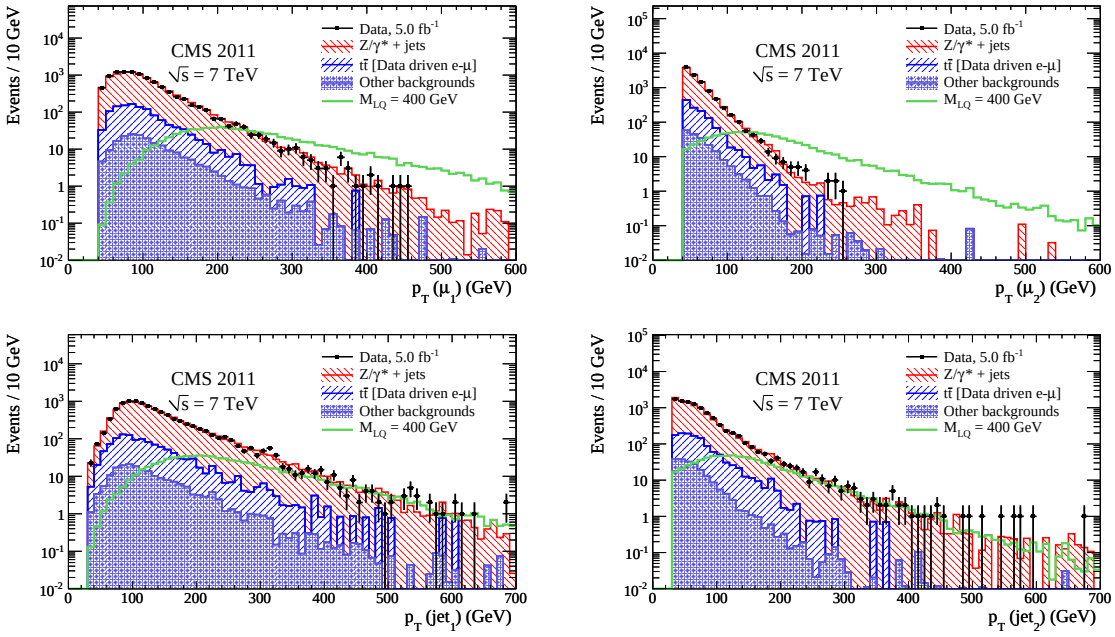

Figure 5.6: Transverse momentum distributions in $\mu\mu jj$ channel in the 7 TeV search. The distributions of p_{T} of the four final-state objects are shown, including the leading muon (top left), the sub-leading muon (top right), the leading jet (bottom left), and the sub-leading jet (bottom right). Distributions contain events that pass the initial selection level. The data are indicated by the points, and the SM backgrounds are given as cumulative histograms. The expected contribution from a leptoquark signal with $M_{\mathrm{LQ}} = 400$ GeV is also shown.

Similarly, in the $\mu\nu jj$ channel of the 7 TeV search, the distributions of the $E_{\mathrm{T}}^{\mathrm{miss}}$ and the p_{T} of the muon and the jets are displayed in Figure 5.8. Events are selected by the $\mu\mu jj$ preselection in the 7 TeV search are shown in Figure 5.6. Figure 5.9 displays the S_{T} and muon-jet invariant mass distributions. Again, good agreement between the data and the SM background prediction is observed, and the signal distribution for $M_{\mathrm{LQ}} = 400$ GeV is overlaid.

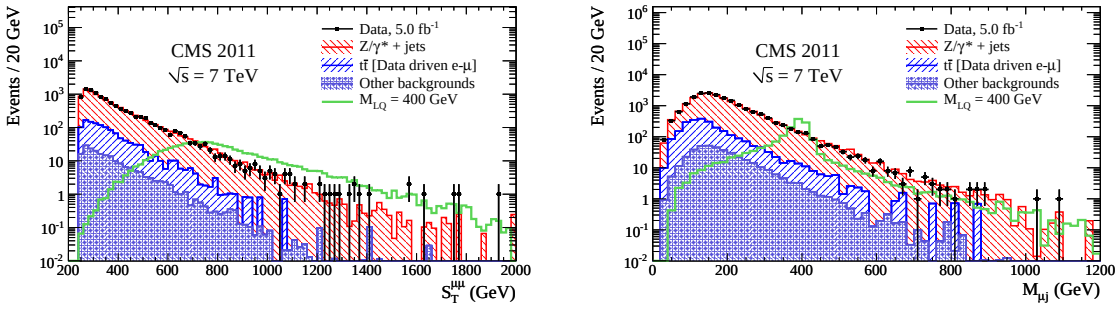

Figure 5.7: For the $\mu\mu jj$ channel in the 7 TeV search, the distributions of S_{T} and of $M_{\mu j}$ for each of the two muon-jet pairs for events that pass the initial selection level. The data are indicated by the points, and the SM backgrounds are given as cumulative histograms. The expected contribution from a leptoquark signal with $M_{\mathrm{LQ}} = 400$ GeV is also shown.

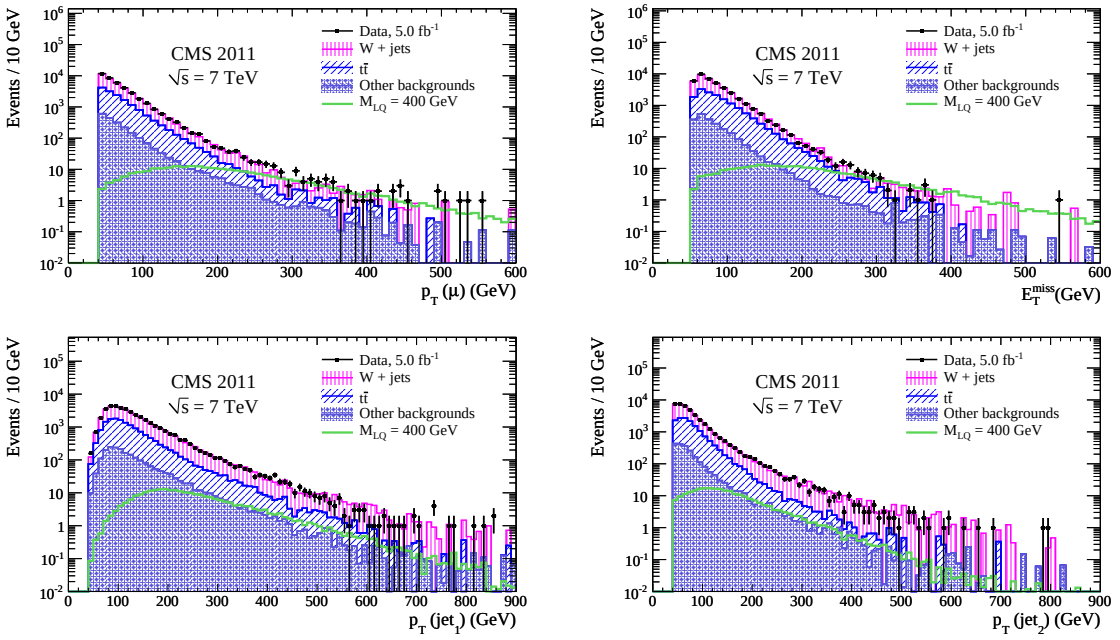

Figure 5.8: Transverse momentum distributions in $\mu\nu jj$ channel in the 7 TeV search. The distributions of p_{T} of the four final-state objects, including the leading muon (top left), the $E_{\mathrm{T}}^{\mathrm{miss}}$ (top right), the leading jet (bottom left), and the sub-leading jet (bottom right). Distributions contain events that pass the initial selection level. The data are indicated by the points, and the SM backgrounds are given as cumulative histograms. The expected contribution from a leptoquark signal with $M_{\mathrm{LQ}} = 400$ GeV is also shown.

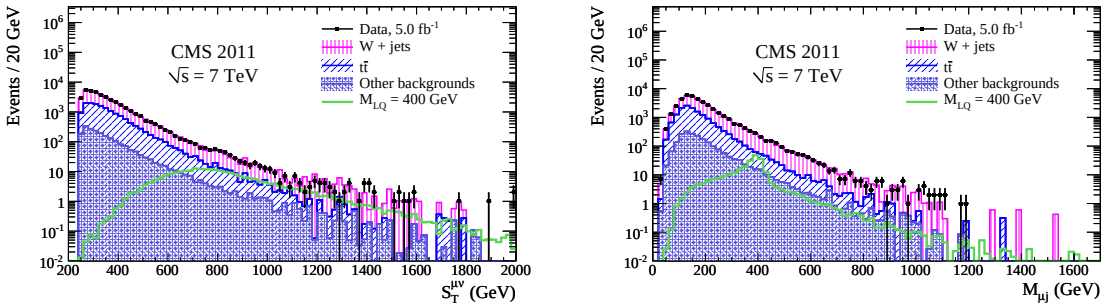

Figure 5.9: For the $\mu\nu jj$ channel in the 7 TeV search, the distributions of S_{T} and of $M_{\mu j}$ for events that pass the initial selection level. The data are indicated by the points, and the SM backgrounds are given as cumulative histograms. The expected contribution from a leptoquark signal with $M_{\mathrm{LQ}} = 400$ GeV is also shown.

5.3.2 Preselection Distributions of the 8 TeV Data

In the 8 TeV search, good agreement is found between the data and the SM background prediction for all distributions at preselection level in the $\mu\mu jj$ channel. Figure 5.10 displays the distributions of the p_{T} of the leading and sub-leading muon, and the p_{T} of the leading and sub-leading jets. Figure 5.11 displays the distributions on which cuts will be placed for the optimized LQ signal selection, including the S_{T}, the $M_{\mu\mu}$, and $M_{\mathrm{min}}(\mu, \mathrm{jet})$, the smallest of the muon-jet mass combinations. For all plots the Z/γ^*+jets and $\mathrm{t\bar{t}}$ MC predictions are normalized to data as described in Section 5.2.2.

Similarly, good agreement is found between the data and the SM background prediction is found for all final state distributions at preselection level in the $\mu\nu jj$ channel. Figures 5.12 displays the muon p_{T}, the $E_{\mathrm{T}}^{\mathrm{miss}}$, and the p_{T} of the leading and the sub-leading jets. Figure 5.13 displays the distributions of quantities to be used for the optimized LQ signal selection, including S_{T}, M_{T}, and $M(\mu, \mathrm{jet})$. For all plots the W+jets and $\mathrm{t\bar{t}}$ predictions are normalized to data as described in Sections 5.2.3.

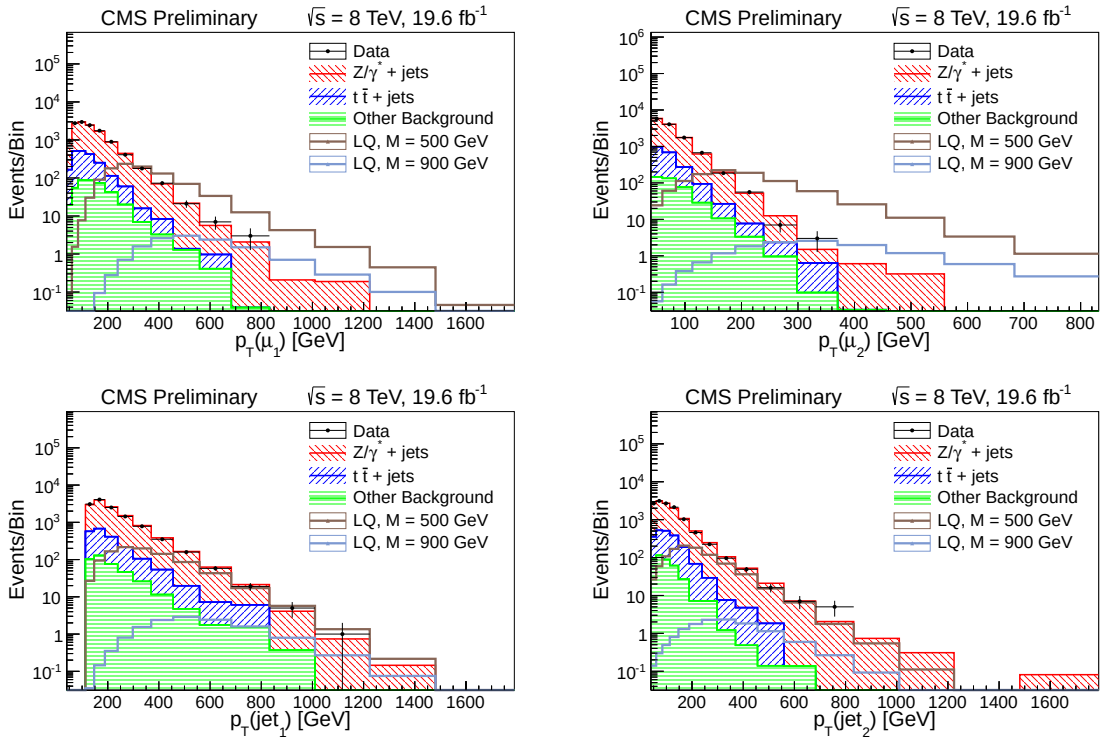

Figure 5.10: Distributions of the muon and the jet p_{T}s at preselection level in the $\mu\mu jj$ channel of the 8 TeV search. "Other Background" include diboson, W+jets, and single-top contributions. The data are indicated by the black markers, and the predictions from LQ signal for LQ hypothesis masses of 500 GeV and 900 GeV are overlaid.

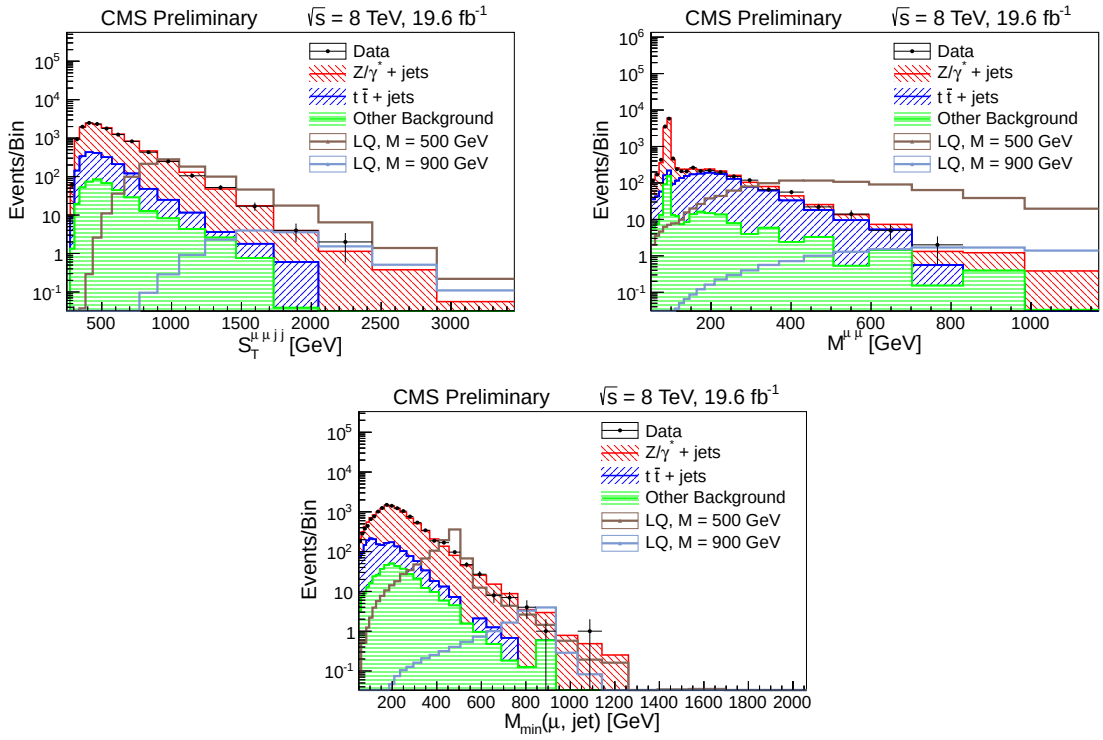

Figure 5.11: Distributions of S_{T}, $M_{\mu\mu}$, and of $M_{\mathrm{min}}(\mu, \mathrm{jet})$ at preselection level in the $\mu\mu jj$ channel of the 8 TeV search. "Other Background" include diboson, W+jets, and single-top contributions. The data are indicated by the black markers, and the predictions from LQ signal for LQ hypothesis masses of 500 GeV and 900 GeV are overlaid.

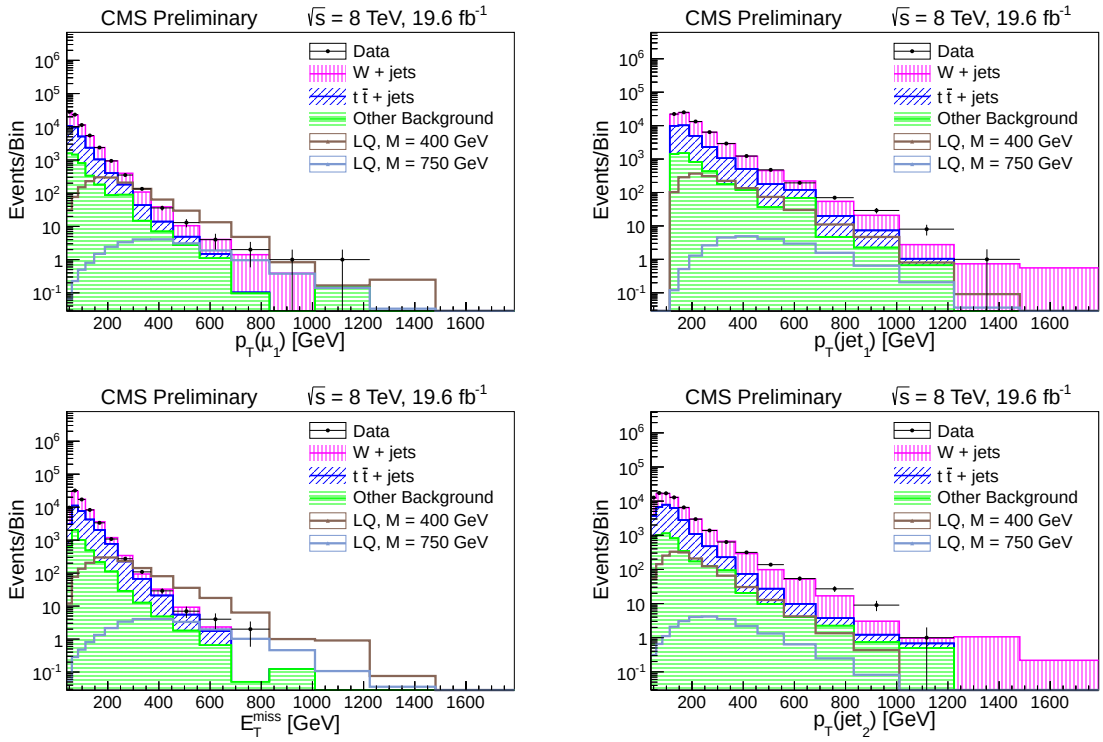

Figure 5.12: Distributions of the muon, $E_{\mathrm{T}}^{\mathrm{miss}}$, and the jet p_{T}s at preselection level in the $\mu\nu jj$ channel of the 8 TeV search. "Other Background" include diboson, Z+jets, and single-top contributions. The data are indicated by the black markers, and the predictions from LQ signal for LQ hypothesis masses of 400 GeV and 750 GeV are overlaid.

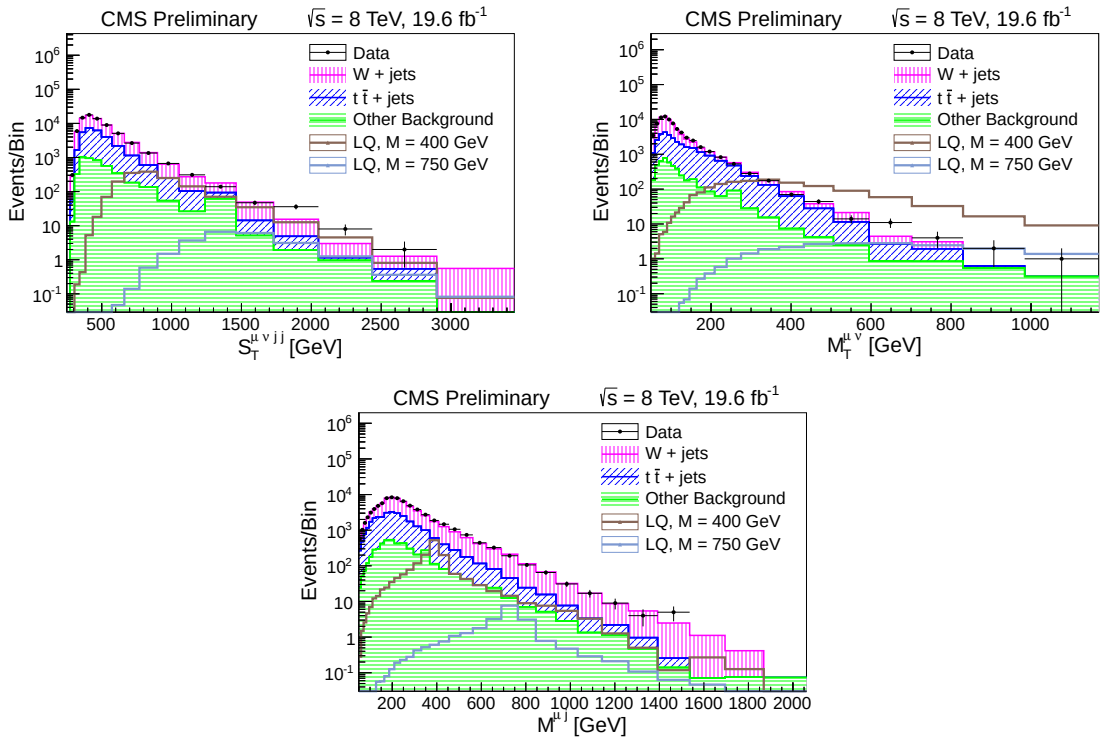

Figure 5.13: Distributions of S_{T}, M_{T}, and of $M(\mu, \mathrm{jet})$ at preselection level in the $\mu\nu jj$ channel of the 8 TeV search. "Other Background" include diboson, Z+jets, and single-top contributions. The data are indicated by the black markers, and the predictions from LQ signal for LQ hypothesis masses of 400 GeV and 750 GeV are overlaid.

5.4 Optimization of the Signal-Background Separation

For each LQ mass hypothesis, an optimized event selection is applied to the data and simulated signal and background. The selected events are examined for evidence of an excess or discrepancy compatible with that of an LQ signal. In the absence of any excess, the total data and simulation event yields are used to set 95% confidence level upper limits on the pair-production cross section of scalar LQs. This section describes the optimization of the selection applied for each LQ mass hypothesis.

5.4.1 Optimized $\mu\mu jj$ Channel Selection

As described in Section 5.1, the $\mu\mu jj$ channel initial event selection consists of events containing at least two muons and at least two jets. The two highest transverse momentum muons and the two highest transverse momentum jets are selected as the decay products from a pair of leptoquarks. After the initial selection criteria are applied to the events, the signal-background separation is optimized by maximizing the figure of merit, $S/\sqrt{S+B}$, where S is the total number of expected signal events and B is the total number of expected background events. This figure of merit is chosen because it represents the total expected signal in terms of the Poisson uncertainty on the total event count in the presence of signal, $\sqrt{S+B}$. It also has the property of being robust against statistical fluctuations leading to zero expected background, which is possible with MC samples containing a limited number of events. In such a case, a figure of merit $S/\sqrt{B}$ would be automatically maximized for zero-background scenarios, regardless of signal content, whereas maximization of $S/\sqrt{S+B}$ would still seek a maximal signal expectation.

To maximize the figure of merit, optimized thresholds are placed on several kine-

matic quantities which are known to have powerful signal-background separation in LQ pair-production searches. The variables optimized for each leptoquark mass hypothesis in the $\mu\mu jj$ channel are:

- $M_{\mu\mu}$, the invariant mass of the dimuon pair, used to remove the majority of the contribution from the $\mathrm{Z}/\gamma^* + \mathrm{jets}$ background.

- S_{T} defined as the scalar sum of the transverse momenta of final state objects in the event: $S_T = p_T(\mu_1) + p_T(\mu_2) + p_T(jet_1) + p_T(jet_2)$. The lower threshold on S_T is optimized for different leptoquark mass hypotheses.

- $M_{\mu j}^{\mathrm{min}}$ defined as the smallest of the two muon-jet invariant masses. To calculate this variable, a pairing between the muons and jets is performed by calculating the muon-jet invariant masses for both muon-jet pairings, i.e. (μ_1:jet$_1$, μ_2:jet$_2$) and (μ_1:jet$_2$, μ_2:jet$_1$). The muon-jet pairing which is chosen is the one which minimizes the invariant mass difference between the two muon-jet combinations. Considering the two muon-jet invariant masses, $M_{\mu j}^{\mathrm{min}}$ is the smaller of the two.

The optimization is performed by scanning a 3-D space for the cuts giving the maximum $S/\sqrt{S+B}$, with cuts tested at 5 GeV increments. For each LQ mass hypothesis, the optimized cuts on each quantity are plotted. A quadratic function is fit to the cuts as a function of LQ mass, and the values of the fit function are taken as the values used in the final optimized selection. The reasons for this approach aref twofold:

1. The quadratic function fit is smooth and increasing as a function of LQ mass. Therefore, the events passing each subsequent optimized selection will be a subset of the events selected for lower masses.

2. Optimization on a large feature set can be prone to overtraining, especially in the case where the total number of signal or background simulated events becomes

small and subject to relatively large statistical fluctuations. At high LQ mass, where MC event yields become small, overtraining becomes more likely. Taking the result from the fit function mitigates the ability of the optimization procedure to choose thresholds based on fluctuations in statistically limited samples.

Due to low background event yields after the stringent high-LQ-mass selections are applied, optimization of the thresholds becomes impossible beyond LQ masses of 1 TeV. So, beyond this selection, the optimized criteria are held constant.

The results of the optimization for the 8 TeV search in the $\mu\mu jj$ channel are displayed in Figure 5.14, along with the quadratic fits. The values of the final selection thresholds, taken from the quadratic fits and rounded to the nearest 5 GeV, are summarized in Table 5.5.

M_{LQ}	300	350	400	450	500	550	600	650	700	750	800	850	900	950	$\geq$1000
$S_T >$	380	460	540	615	685	755	820	880	935	990	1040	1090	1135	1175	1210
$M_{\mu\mu} >$	100	115	125	140	150	165	175	185	195	205	215	220	230	235	245
$M_{\mu j}^{\mathrm{min}} >$	115	115	120	135	155	180	210	250	295	345	400	465	535	610	690

Table 5.5: Optimized thresholds for different mass hypothesis of the $\mu\mu jj$ signal, for the 8 TeV analysis. Values are taken from the quadratic fit in Figure 5.14, and rounded to the nearest 5 GeV. All values are expressed in GeV.

A similar procedure is applied in the 7 TeV search, however cuts are optimized in common with a first-generation LQ search [104]. The resulting optimized thresholds are summarized in Table 5.6 for the $\ell\ell jj$ channel.

M_{LQ}	250	350	400	450	500	550	600	650	750	850	900
$S_{\mathrm{T}} >$	330	450	530	610	690	720	770	810	880	900	920
$M_{\mu\mu} >$ (GeV)	100	110	120	130	130	130	130	130	140	150	150
$M_{\ell j}^{\mathrm{min}} >$	60	160	200	250	300	340	370	400	470	500	520

Table 5.6: Optimized thresholds for different mass hypothesis of the $\mu\mu jj$ signal, for the 7 TeV analysis. The optimized thresholds are determined in common with those used for a first-generation LQ search [104]. All values are expressed in GeV

Data and background predictions with the optimized LQ event selections in the 8

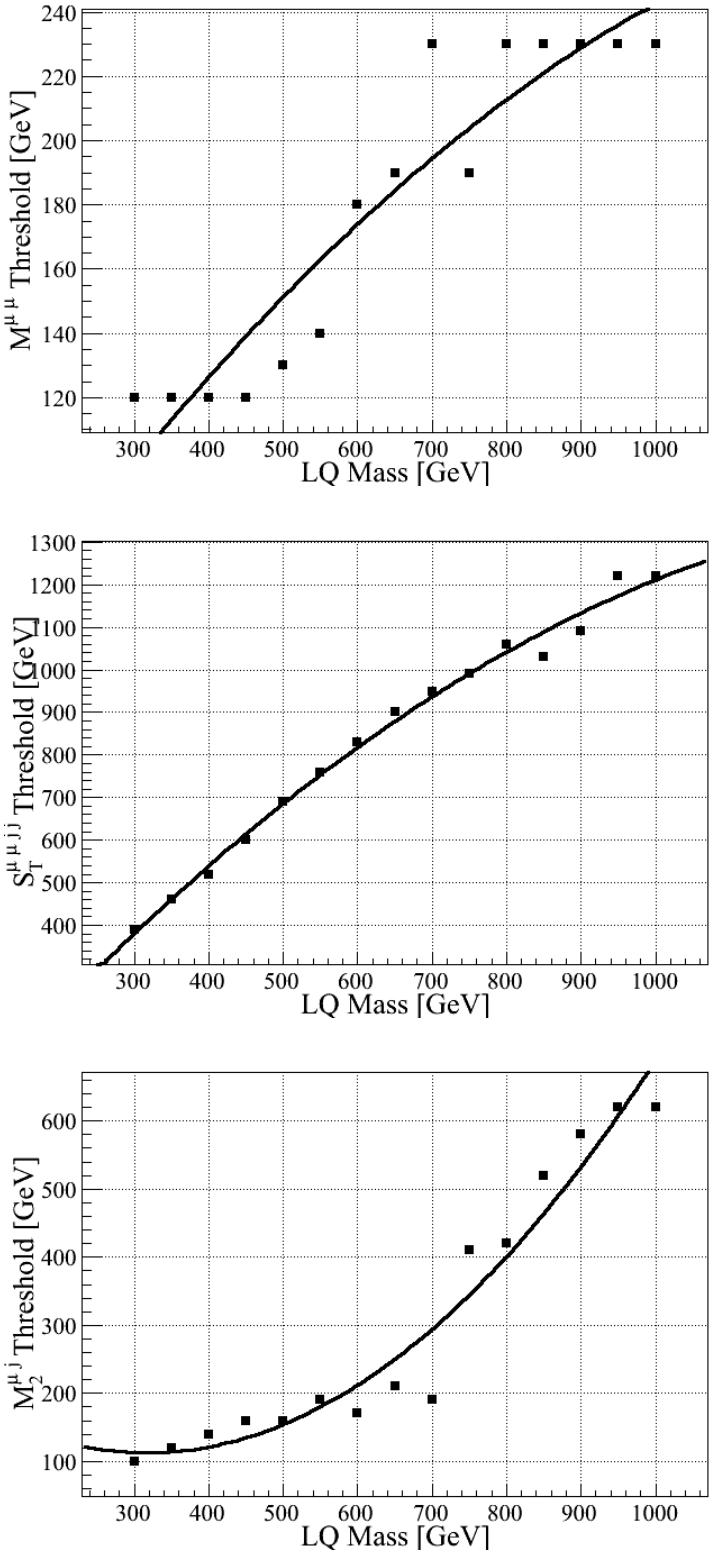

Figure 5.14: Optimized quantity threhsods, fitted to quadratic functions, for the $\mu\mu jj$ optimization in the 8 TeV search.

TeV search are shown in Figures 5.15 and 5.16, which compare S_{T} and $M_{\mu j}^{\mathrm{min}}$ for a signal LQ mass of 500 GeV and 900 GeV respectively. The gray shaded bands give the uncertainty on the background prediction, determined as the sum in quadrature of the statistical uncertainty on background and systematic uncertainty on background, and detailed in Section 5.5.

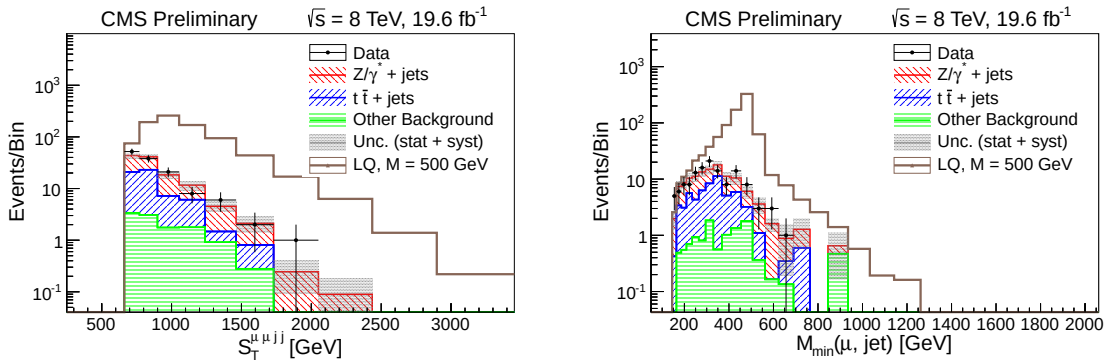

Figure 5.15: Distributions of S_{T} and $M_{\mu j}^{\mathrm{min}}$ reconstruction at final selection level for a LQ mass of 500 GeV in the $\mu\mu jj$ channel. The gray shaded region indicates the statistical and systematic uncertainty on the background prediction. "Other Background" include diboson, W+jets, and single-top contributions.

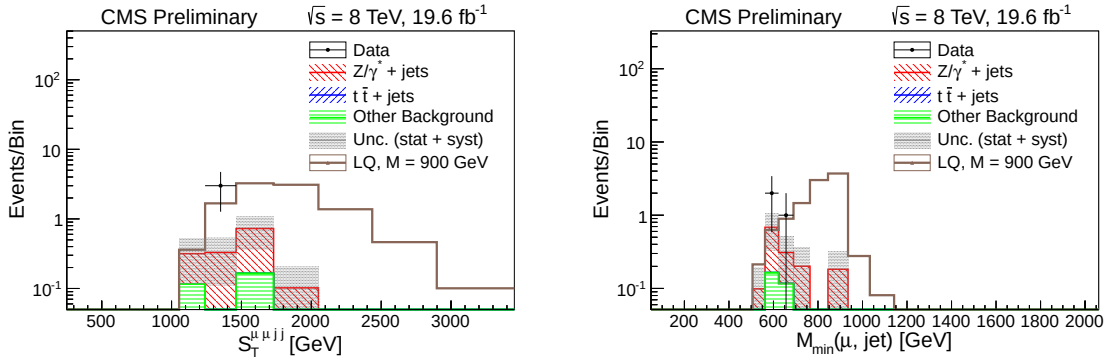

Figure 5.16: Distributions of S_{T} and $M_{\mu j}^{\mathrm{min}}$ reconstruction at final selection level for a LQ mass of 900 GeV in the $\mu\mu jj$ channel. The gray shaded region indicates the statistical and systematic uncertainty on the background prediction. "Other Background" include diboson, W+jets, and single-top contributions.

Similarly, for the 7 TeV search, Figure 5.17 displays the the distributions of S_{T} (left) and of $M_{\mu j}$ with the $\mu\mu jj$ signal selection optimized for $M_{\mathrm{LQ}} = 600$ GeV.

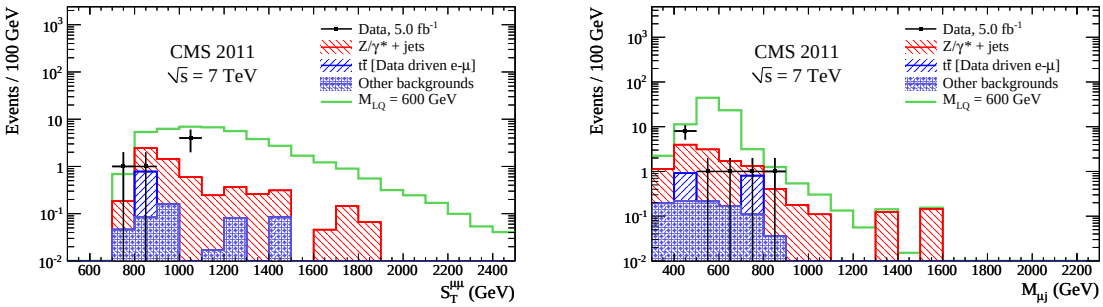

Figure 5.17: The distributions of S_{T} (left) and of $M_{\mu j}$ for each of the two muon-jet pairs (right) for events that pass the final selection criteria optimized for a signal leptoquark mass of 600 GeV in the $\mu\mu jj$ channel for the 7 TeV search. The expected contribution from a leptoquark signal with $M_{\mathrm{LQ}} = 600$ GeV is also shown.

5.4.2 Optimized $\mu\nu jj$ Channel Selection

Event selection in the $\mu\nu jj$ channel proceeds in a similar manner to that of the $\mu\mu jj$ channel. Beginning with events passing the $\mu\nu jj$ preselection requirements, optimized thresholds are placed on a set of kinematic quantities to optimize the signal significance $S/\sqrt{S+B}$.

The following variables are optimized for each leptoquark mass hypothesis:

- M_{T}, the transverse mass of the muon and $E_{\mathrm{T}}^{\mathrm{miss}}$, to veto the W transverse mass peak and greatly reduce the contribution from the W+jets background.

- S_{T} defined as the scalar sum of the transverse momenta of final state objects in the event: $S_T = p_T(\mu_1) + E_{\mathrm{T}}^{\mathrm{miss}} + p_T(jet_1) + p_T(jet_2)$.

- $M_{\mu j}$ defined as the invariant mass of the muon jet combination which minimizes the $\mathrm{LQ} - \overline{\mathrm{LQ}}$ transverse mass difference. To calculate this variable, a pairing between the muons and jets is performed by calculating the muon-jet and $E_{\mathrm{T}}^{\mathrm{miss}}$-jet transverse masses for both muon-jet and $E_{\mathrm{T}}^{\mathrm{miss}}$-jet pairings, i.e. ($\mu$:jet$_1$, $E_{\mathrm{T}}^{\mathrm{miss}}$:jet$_2$) and ($\mu$:jet$_2$, $E_{\mathrm{T}}^{\mathrm{miss}}$:jet$_1$). The pairing which is chosen is the one which minimize

the difference between the muon-jet and the $E_{\mathrm{T}}^{\mathrm{miss}}$-jet transverse masses. After the optimal pairing is determined, $M_{\mu j}$ is simply the invariant mass of the muon and its matched jet.

The results of the optimization for 19.6 fb^{-1} of integrated luminosity in the $\mu\nu jj$ channel are displayed in Figure 5.18.

The optimized $\mu\nu jj$ thresholds are summarized in Table 5.7.

M_{LQ}	300	350	400	450	500	550	600	650	700	750	800	850	900	950	$\geq$1000
$S_T >$	455	540	625	715	800	890	980	1070	1160	1250	1345	1435	1530	1625	1720
$M_{\mathrm{T}} >$	155	180	205	225	245	260	275	290	300	310	315	320	320	325	320
$M_{\mu j} >$	125	150	175	200	225	250	280	305	330	355	380	410	435	465	490

Table 5.7: Optimized thresholds for different mass hypothesis of the $\mu\nu jj$ signal, for the 8 TeV analysis. Values are taken from the quadratic fit in Figure 5.18, and rounded to the nearest 5 GeV. All values are expressed in GeV.

As in the $\mu\mu jj$ channel, the optimized cuts applied in the 7 TeV $\mu\nu jj$ channel are optimized in common with a first-generation LQ search [104]. The resulting optimized thresholds are summarized in Table 5.8 for the $\ell\nu jj$ channel.

M_{LQ}	250	350	400	450	500	550	600	650	750	850
$S_{\mathrm{T}} >$	450	570	650	700	800	850	890	970	1000	1000
$E_{\mathrm{T}}^{\mathrm{miss}} >$	100	120	120	140	160	160	180	180	220	240
$M_{\mu j} >$	150	300	360	360	360	480	480	540	540	540

Table 5.8: Optimized thresholds for different mass hypothesis of the $\ell\nu jj$ signal, for the 7 TeV analysis. The optimized thresholds are determined in common with those used for a first-generation LQ search [104]. All values are expressed in GeV.

Data and background predictions at final selection level for the 8 TeV search are also shown in Figure 5.19, which compare S_{T} and $M_{\mu j}$ for a signal LQ mass of 500 GeV. The gray shaded bands give the uncertainty on the background prediction, determined as the sum in quadrature of the statistical uncertainty on background and systematic uncertainty on background, and detailed in Section 5.5.

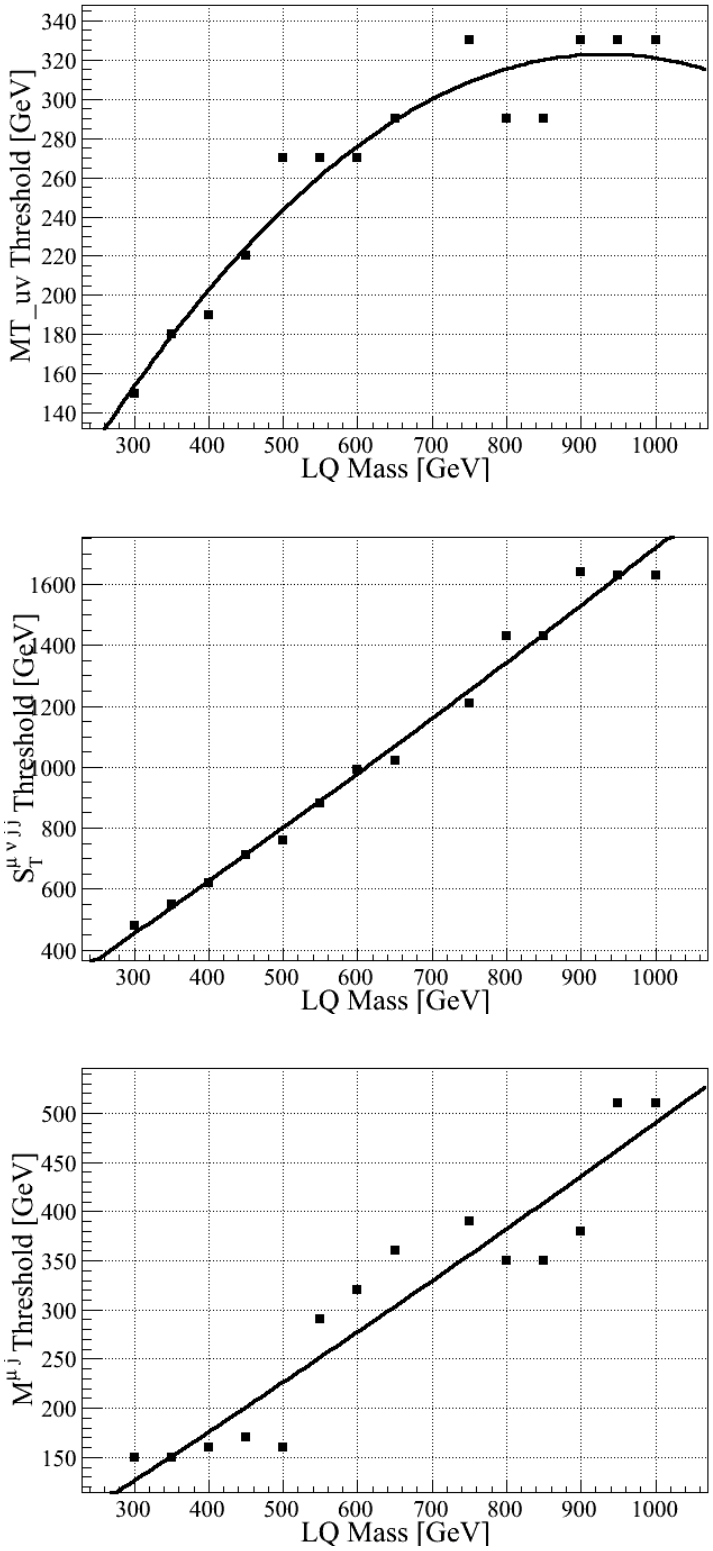

Figure 5.18: Optimization scan fitted to a quadratic function for the $\mu\mu jj$ optimization variables.

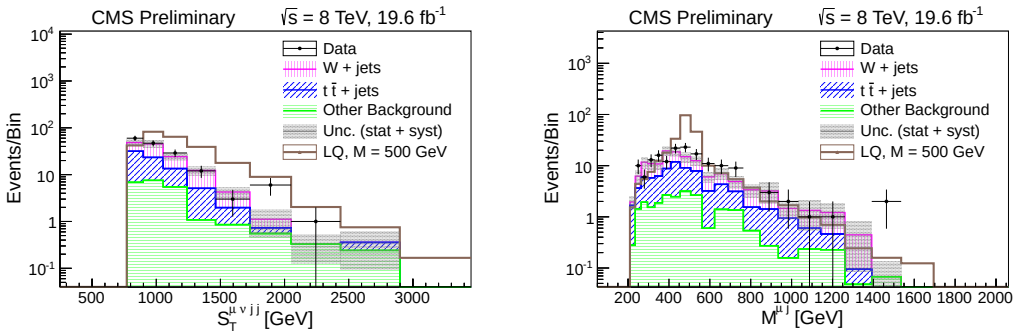

Figure 5.19: Distributions of S_{T} and $M_{\mu j}$ reconstruction at final selection level for a LQ mass of 500 GeV in the $\mu\nu jj$ channel. The gray shaded region indicates the statistical and systematic uncertainty on the background prediction. "Other Background" include diboson, Z+jets, and single-top contributions.

Similarly, for the 7 TeV search, Figure 5.20 displays the the distributions of S_{T} (left) and of $M_{\mu j}$ with the $\mu\nu jj$ signal selection optimized for $M_{\mathrm{LQ}} = 600$ GeV.

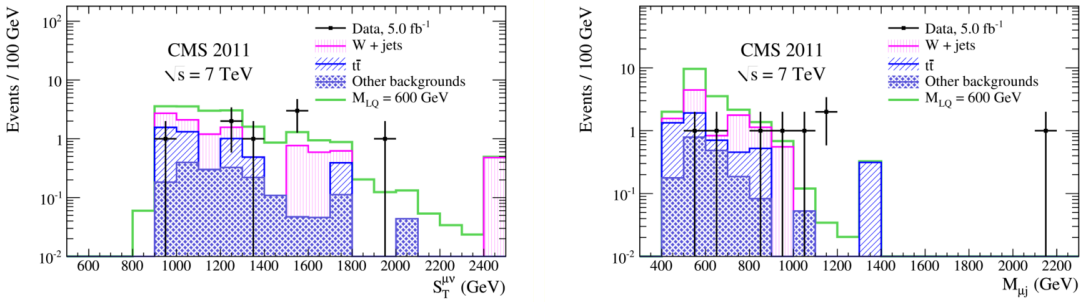

Figure 5.20: $\mu\nu jj$ channel: the distributions of S_{T} (left) and of $M_{\mu j}$ (right) for events that pass the final selection criteria optimized for a signal leptoquark mass of 600 GeV. The expected contribution from a leptoquark signal with $M_{\mathrm{LQ}} = 600$ GeV is also shown.

5.5 Systematic Uncertainties

Systematic uncertainties are assessed by varying quantities in the simulated background and determining the variation in the final event yields after applying the optimized selections. Uncertainties on kinematic quantities, such as the jet and muon energy and

momentum scales and resolutions, are considered, as well as global event normalization variations due to uncertainty on the pileup modeling, integrated luminosity, and data-driven background rescaling factors. PDF uncertainties on the signal and background cross sections and event selections are evaluated, as well as uncertainties in the choice of factorization and renormalization scale and jet-parton matching thresholds.

5.5.1 Energy and Momentum Scales and Resolutions

The uncertainty on the muon momentum scale is assessed by varying the muon momentum up and down by a conservative value. For the 7 TeV analysis, the uncertainty is taken to be 1% [113], and for the 8 TeV analysis, the uncertainty is parameterized as a function of the muon p_{T} as $5\% \times p_{\mathrm{T}}/\mathrm{TeV}$, to more precisely quantify the uncertainty for high-momentum muons. The uncertainty on the muon momentum resolution is assessed by modifying the values of the muon momentum by random values according to a Gaussian distribution. For the 7 TeV analysis, the width of this Gaussian smearing is taken conservatively to be 4% of the muon momentum [113], and in the 8 TeV analysis, a p_{T}-dependent value of 1-4% is used.

The jet energy scale uncertainty is assessed by varying the jet energy up and down according to the uncertainty in the jet-energy correction factors [114] applied to the simulation. This uncertainty is taken to be a conservative value of 4% in the 7 TeV search, while in the 8 TeV search this uncertainty is reduced by considering a more accurate uncertainty as a function of the jet p_{T} and η. The jet energy resolution uncertainties are assessed by varying the p_{T} difference between each reconstructed jet and its corresponding generator-level jet according to the resolution uncertainties [114]. This uncertainty varies with jet p_{T} and η, and is usually between 5% and 14%. To calculate this uncertainty, the generator-level jet matched to each reconstructed jet is that which is closest in ΔR, considering generator-level jets with $\Delta R < 0.5$, if any.

The effect of energy scale and resolution uncertainties are calculated for the (minor) background sources for which no data rescaling is applied. For the background sources for which data rescaling is applied, residual uncertainties are calculated (relative to the initial selection used to derive the rescaling factor).

5.5.2 Background Normalization and Shape

The data-driven estimation of the normalizations of the Z+jets, W+jets, and $t\bar{t}$ background contributions have associated uncertainties based on the statistical uncertainties of the control samples used to determine the scale factors. Table 5.9 details the values of these uncertainties, which range from 1% to 5%.

	$5\,\text{fb}^{-1}$ at $\sqrt{s} = 7$ TeV		$20\,\text{fb}^{-1}$ at $\sqrt{s} = 8$ TeV	
Sample	$\mu\mu jj$	$\mu\nu jj$	$\mu\mu jj$	$\mu\nu jj$
Z+jets Simulation	2%	—	1%	—
W+jets Simulation	—	2%	—	1%
$t\bar{t}$ Simulation	—	4%	—	2%
$t\bar{t}$ $e\mu$ Data	5%	—	2%	—

Table 5.9: The uncertainties on the data-driven reweighting factors for the simulation and data samples used to estimate the major backgrounds from W+jets, Z+jets, and $t\bar{t}$ processes.

Shape uncertainties on the Z+jets, W+jets and $t\bar{t}$ backgrounds are determined using MadGraph MC samples with factorization/renormalization scales and jet-parton matching thresholds varied by a factor of 2. In the $\mu\mu jj$ channel, the uncertainty on the Z+jets background is assessed, but no uncertainty on the $t\bar{t}$ background is needed because the $t\bar{t}$ estimate uses the $e\mu jj$ data sample. In the $\mu\nu jj$ channel, the uncertainties on the W+jets and $t\bar{t}$ backgrounds are assessed. The uncertainties are determined in the following way:

- Four alternative fully-reconstructed MC samples are examined for each back-

ground process. The samples considered include (1) a sample with the factorization and renormalization scale both varied up by a factor of 2; (2) a sample with the factorization and renormalization scale both varied down by a factor of 2; (3) a sample with the jet-parton matching threshold varied up by a factor of 2; and (4) a sample with with the jet-parton matching threshold varied down by a factor of 2.

- Since the major backgrounds are normalized to data in control regions at the initial event selection, variations in the scales and matching threshold are considered after this selection level, i.e. only changes in event yields with respect to the initial selection are considered.

- The variation in the final event yield is considered with each of the four samples. Since the systematically varied samples contain a limited number of events compared to that of the central sample, this must be approximated with an intermediate selection which retains enough events such that the variation is not dominated by the statistical variation in the samples. For the 7 TeV analysis, this variation was assessed considering all of the optimized selection cuts for an LQ mass of 250 GeV applied separately, and taking the maximal variation in event yield with respect to the central sample. For the 8 TeV analysis, the variation is calculated considering the optimized selection cuts for an LQ mass of 300 GeV, with the cuts on $M_{\mu\mu}$ and M_{T} removed when assessing the Z+jets and W+jets variations.

- After assessing the variations in each of the four systematically varied samples with respect to the central sample, the maximal variation is taken as the size of the systematic uncertainty.

The magnitudes of the systematic uncertainties on the background shapes are given in

Table 5.10

Background	channel	7 TeV Search	8 TeV Search
Z+jets	$\mu\mu jj$	10%	3%
W+jets	$\mu\nu jj$	11%	9%
$t\bar{t}$	$\mu\nu jj$	10%	20%

Table 5.10: The magnitudes of the uncertainties on the background shapes, for the W+jets, Z+jets, and $t\bar{t}$ processes. The uncertainties consider variations of a factor of 2 in the jet-parton matching threshold and factorization/renormalization scale.

5.5.3 Parton Distribution Functions

PDF uncertainties on the theoretical cross section of leptoquark production and on the final selection acceptance have been calculated using the PDF4LHC [115] prescriptions, with PDF and α_s variations of the MSTW2008 [116], CTEQ6.6 [103] and NNPDF2.0 [117] PDF sets taken into account. Uncertainties on the cross section vary from 10 to 30 % for leptoquarks in the mass range of 200–900 GeV, while the effect of the PDF uncertainties on signal acceptance varies from 1 to 3%. The PDF uncertainties are not considered for background sources with uncertainties determined from data.

5.5.4 Integrated Luminosity, Event and Object Acceptances

The initial normalizations of all simulated background and signal processes are determined according to a total measured integrated luminosity for the data-taking period of interest. The uncertainty on the integrated luminosity is taken as 2.2% [118] in the 7 TeV data, and 4.4% in the 8 TeV data [119].

Additional uncertainties are applied on the event acceptance of the single-muon trigger used to select events and on the identification of the muons. Recent measure-

ments of the muon reconstruction, identification, trigger, and isolation efficiencies using Z$\rightarrow \mu\mu$ events show very good agreement between data and MC events [120]. A $\sim 1\%$ discrepancy is observed in the data-to-MC comparison of the muon trigger efficiency for a single muon passing the other identification requirements. An additional uncertainty on the muon identification and isolation efficiency is evaluated to be 2%. These uncertainties are applied to both signal and background estimates, for each muon in the final state.

Lastly, an uncertainty on the modeling of pileup interactions in the MC simulation is determined by varying the mean of the distribution of pileup interactions by $6-8\%$, depending on the data-taking period. This conservative value accounts for uncertainty in the luminosity and the uncertainty in the total inelastic cross section [121].

The systematic uncertainties and their effects on signal and background are summarized in Table 5.11 for both the $\mu\mu jj$ and $\mu\nu jj$ channels in the 7 TeV search. The representative figures in this table correspond to the final selection optimized for $M_{\mathrm{LQ}} = 600$ GeV. Similar values are presented in Table 5.12, for the 8 TeV search using the optimized selection for an LQ mass of 700 GeV.

Systematic	Magnitude	$\mu\mu jj$		$\mu\nu jj$	
Uncertainties	(%)	S(%)	B(%)	S(%)	B(%)
Jet Energy Scale	4	1	1	3	7
Background Modeling	—	–	9	–	10
Muon Momentum Scale	1	0.5	4	1	2
Muon Reco/ID/Iso	1	2	–	1	–
Jet Resolution	5-14	< 0.5	< 0.5	< 0.5	2.5
Muon Resolution	4	< 0.5	5	< 0.5	1
Pileup	8	0.5	< 0.5	1	4
Integrated Luminosity	2.2	2.2	–	2.2	–
Total		3	11	4	13

Table 5.11: For the 7 TeV search, systematic uncertainties and their effects on signal (S) and background (B) in all channels for the $M_{LQ} = 600$ GeV final selection. All uncertainties are symmetric.

Systematic Uncertainties	Magnitude (%)	$\mu\mu jj$ $S(\%)$	$\mu\mu jj$ $B(\%)$	$\mu\nu jj$ $S(\%)$	$\mu\nu jj$ $B(\%)$
Jet Energy Scale	Variable	0.5	7	< 0.5	4.5
Background Modeling	–	–	5.5	–	13
Muon Momentum Scale	$5 \times p_{\mathrm{T}}$/TeV	1	3	< 0.5	16
Muon Reco/ID/Iso	2	4	1	2	0.5
Jet Resolution	5-14	< 0.5	1	< 0.5	2.5
Muon Resolution	1 - 4	< 0.5	< 0.5	0.5	7.5
Pileup	6	0.5	1.5	< 0.5	5
Integrated Luminosity	4.4	4.4	1	4.4	0.6
Total		6	10	5.5	24

Table 5.12: For the 8 TeV search, systematic uncertainties and their effects on signal (S) and background (B) in all channels for the $M_{LQ} = 600$ GeV final selection. All uncertainties are symmetric.

5.6 Results

5.6.1 Event Yields with the Optimized Signal Selections

The event yields for the optimized selections corresponding to each LQ hypothesis mass, in the $\mu\mu jj$ and $\mu\nu jj$ channels for the 7 TeV analysis, are displayed in Table 5.13 and Table 5.14. Total background and signal counts are displayed along with their corresponding statistical uncertainties. The data agree well with the SM background predictions.

The number of events selected in data, estimated signal, and estimated backgrounds, at the optimized final selections in the 8 TeV searches for the $\mu\mu jj$ and $\mu\nu jj$ channels are shown Table 5.13 and Table 5.14, respectively.

In both the $\mu\mu jj$ and $\mu\nu jj$ channels, in the searches with 7 TeV data and with 8 TeV data, no significant excess beyond the predicted SM backgrounds is observed. With no compelling evidence for LQ-like signatures observed, limits on the second-

generation LQ pair production cross sections can be places as a function of LQ mass and a function of the branching fraction β.

M_{LQ}	Z+jets	$\mathrm{t\bar{t}}$	Other BG	Signal	Data	Total BG
400	46.9 ± 2.4	30.1 ± 4.2	$3.58^{+0.70}_{-0.40}$	629.3 ± 4.0	68	$80.6^{+4.9}_{-4.9}$
500	10.4 ± 1.1	4.1 ± 1.6	$0.89^{+0.62}_{-0.21}$	136.73 ± 0.86	14	$15.4^{+2.0}_{-1.9}$
550	7.29 ± 0.94	2.4 ± 1.2	$0.53^{+0.60}_{-0.17}$	70.49 ± 0.40	9	$10.2^{+1.6}_{-1.5}$
600	5.03 ± 0.75	0.59 ± 0.59	$0.47^{+0.60}_{-0.16}$	37.39 ± 0.22	6	$6.1^{+1.1}_{-1.0}$
650	3.82 ± 0.65	0.59 ± 0.59	$0.24^{+0.59}_{-0.12}$	20.56 ± 0.13	5	$4.7^{+1.1}_{-0.9}$
750	2.03 ± 0.47	$0.00^{+0.67}_{-0.00}$	$0.09^{+0.59}_{-0.08}$	6.529 ± 0.038	1	$2.1^{+1.0}_{-0.5}$
850	1.56 ± 0.42	$0.00^{+0.67}_{-0.00}$	$0.08^{+0.59}_{-0.08}$	2.327 ± 0.014	0	$1.6^{+1.0}_{-0.4}$

Table 5.13: Individual background (BG) sources, expected signal, data and total background event yields after the initial (first row) and final selections for the $\mu\mu jj$-channel search with $\sqrt{s} = 7$ TeV data. Other BG includes single top, W+jets, and VV+jets. Only statistical uncertainties are reported.

M_{LQ}	Z+jets	$\mathrm{t\bar{t}}$	Other BG	Signal	Data	Total BG
400	31.8 ± 4.4	23.9 ± 2.6	5.68 ± 0.64	118.3 ± 1.2	60	61.5 ± 5.2
500	18.0 ± 3.3	9.6 ± 1.7	3.38 ± 0.51	33.94 ± 0.28	26	31.0 ± 3.7
550	5.7 ± 1.8	4.2 ± 1.1	2.21 ± 0.43	15.51 ± 0.14	12	12.1 ± 2.1
600	5.5 ± 1.8	3.2 ± 1.0	1.80 ± 0.39	9.23 ± 0.08	8	10.5 ± 2.1
650	2.9 ± 1.2	1.14 ± 0.59	1.13 ± 0.31	4.964 ± 0.043	7	5.1 ± 1.3
750	2.9 ± 1.2	0.51 ± 0.36	0.99 ± 0.29	1.840 ± 0.015	6	4.4 ± 1.2
850	2.7 ± 1.1	0.51 ± 0.36	0.76 ± 0.23	0.6906 ± 0.0051	6	4.0 ± 1.2

Table 5.14: Individual background (BG) sources, expected signal, data, and total background event yields after the initial (first row) and final selections for the $\mu\nu jj$-channel search with $\sqrt{s} = 7$ TeV data. Other BG includes single top, Z+jets, and VV+jets. Only statistical uncertainties are reported.

M_{LQ}	Signal	Z+Jets	$t\bar{t}$	Other BG	Total BG	Data
300	14980 ± 110	716.2 ± 8.4	612 ± 18	86.7 ± 5.0	1415 ± 20 ± 45	1461
350	6975 ± 46	307.2 ± 5.5	368 ± 14	54.2 ± 4.1	730 ± 15 ± 16	714
400	3369 ± 22	176.5 ± 4.1	178.7 ± 9.4	29.6 ± 3.0	384.8 ± 10.7 ± 9.3	394
450	1664 ± 10	97 ± 3.0	89.3 ± 6.6	18.9 ± 2.4	205.3 ± 7.6 ± 5.5	210
500	859.4 ± 5.2	61.9 ± 2.4	48.5 ± 4.8	11.2 ± 1.9	121.6 ± 5.7 ± 4.8	128
550	459.3 ± 2.8	35.1 ± 1.8	25.5 ± 3.4	7.5 ± 1.6	68.1 ± 4.2 ± 2.7	75
600	252.3 ± 1.5	23 ± 1.4	15.84 ± 2.76	5.85 ± 1.41	44.7 ± 3.4 ± 2.0	44
650	143.87 ± 0.86	15.1 ± 1.13	8.86 ± 1.98	$4.08\ ^{+1.32}_{-1.25}$	28 ± 2.6 ± 1.3	24
700	82.02 ± 0.49	9.66 ± 0.91	5.97 ± 1.72	$2.99\ ^{+1.12}_{-1.04}$	18.6 ± 2.2 ± 1.3	15
750	48.06 ± 0.29	6.37 ± 0.74	1.41 ± 0.7	$1.54\ ^{+0.78}_{-0.67}$	$9.32\ ^{+1.29}_{-1.22}$ ± 0.87	11
800	28.73 ± 0.17	3.85 ± 0.58	1.55 ± 0.77	$1.13\ ^{+0.71}_{-0.59}$	$6.53\ ^{+1.2}_{-1.13}$ ± 0.85	9
850	17.43 ± 0.11	2.2 ± 0.42	0.56 ± 0.56	$1.12\ ^{+0.72}_{-0.59}$	$3.88\ ^{+1.0}_{-0.92}$ ± 0.67	5
900	10.337 ± 0.064	1.19 ± 0.31	$0.0\ ^{+0.59}_{-0.0}$	$0.28\ ^{+0.45}_{-0.2}$	$1.47\ ^{+0.81}_{-0.37}$ ± 0.43	3
950	6.333 ± 0.04	0.71 ± 0.24	$0.0\ ^{+0.59}_{-0.0}$	$0.117\ ^{+0.658}_{-0.117}$	$0.83\ ^{+0.91}_{-0.26}$ ± 0.29	1
1000	3.845 ± 0.025	0.38 ± 0.17	$0.0\ ^{+0.59}_{-0.0}$	$0.0\ ^{+0.65}_{-0.0}$	$0.383\ ^{+0.894}_{-0.171}$ ± 0.031	0
1050	2.557 ± 0.016	0.38 ± 0.17	$0.0\ ^{+0.59}_{-0.0}$	$0.0\ ^{+0.65}_{-0.0}$	$0.383\ ^{+0.894}_{-0.171}$ ± 0.031	0
1100	1.714 ± 0.01	0.38 ± 0.17	$0.0\ ^{+0.59}_{-0.0}$	$0.0\ ^{+0.65}_{-0.0}$	$0.383\ ^{+0.894}_{-0.171}$ ± 0.031	0
1150	1.1465 ± 0.0069	0.38 ± 0.17	$0.0\ ^{+0.59}_{-0.0}$	$0.0\ ^{+0.65}_{-0.0}$	$0.383\ ^{+0.894}_{-0.171}$ ± 0.031	0
1200	0.7554 ± 0.0045	0.38 ± 0.17	$0.0\ ^{+0.59}_{-0.0}$	$0.0\ ^{+0.65}_{-0.0}$	$0.383\ ^{+0.894}_{-0.171}$ ± 0.031	0

Table 5.15: Event yields at final selection level for the $\mu\mu jj$-channel search with $\sqrt{s} = 8$ TeV data. Other BG includes single top, W+jets, and VV+jets. Uncertainties are Poisson uncertainties on the MC background, except for the second uncertainty for "Total BG", which gives the total sysematic uncertainty as detailed in Section 5.5. Systematic uncertainties are dominated by energy scale and shape uncertainties.

M_{LQ}	Signal	W+Jets	$t\bar{t}$	Other BG	Total BG	Data
300	5032 ± 69	990 ± 21	1741 ± 14	362 ± 11	3093 ± 27 ± 383	3276
350	2322 ± 28	418 ± 14	604.5 ± 8.1	201.8 ± 9.5	1224 ± 18 ± 137	1315
400	1032 ± 11	195.8 ± 9.1	243.6 ± 5.1	75.8 ± 4.1	515 ± 11 ± 60	594
450	512.8 ± 8.6	101.4 ± 6.6	110.4 ± 3.5	41.6 ± 2.9	253.3 ± 8.0 ± 28	289
500	257.6 ± 2.7	59.3 ± 5.0	53.9 ± 2.4	23.6 ± 2.1	136.8 ± 5.9 ± 15	158
550	139.2 ± 1.6	37.1 ± 3.9	24.5 ± 1.6	14.1 ± 1.7	75.8 ± 4.6 ± 8.7	87
600	75.77 ± 0.8	19.2 ± 2.7	13.7 ± 1.2	7.4 ± 1.1	40.3 ± 3.2 ± 4.8	53
650	43.18 ± 0.45	12.1 ± 2.2	7.48 ± 0.89	3.98 ± 0.71	23.6 ± 2.5 ± 3.6	32
700	24.51 ± 0.26	7.2 ± 1.7	4.82 ± 0.71	$2.37\ ^{+0.77}_{-0.45}$	$14.4\ ^{+2.0}_{-1.9}$ ± 2.6	22
750	14.63 ± 0.15	5.3 ± 1.5	2.87 ± 0.55	$1.87\ ^{+0.76}_{-0.42}$	$10\ ^{+1.7}_{-1.6}$ ± 2.3	16
800	8.879 ± 0.097	3.8 ± 1.4	1.41 ± 0.39	$1.6\ ^{+0.74}_{-0.4}$	$6.9\ ^{+1.6}_{-1.5}$ ± 1.9	12
850	5.346 ± 0.056	0.92 ± 0.53	0.75 ± 0.28	$1.16\ ^{+0.72}_{-0.36}$	$2.83\ ^{+0.94}_{-0.7}$ ± 0.98	6
900	3.265 ± 0.036	0.6 ± 0.43	0.63 ± 0.26	$0.86\ ^{+0.7}_{-0.32}$	$2.09\ ^{+0.86}_{-0.59}$ ± 0.92	4
950	2.056 ± 0.022	0.39 ± 0.39	0.42 ± 0.21	$0.73\ ^{+0.7}_{-0.3}$	$1.54\ ^{+0.83}_{-0.54}$ ± 0.54	4
1000	1.287 ± 0.014	0.39 ± 0.39	0.252 ± 0.145	$0.61\ ^{+0.69}_{-0.28}$	$1.25\ ^{+0.8}_{-0.5}$ ± 0.38	4
1050	0.9091 ± 0.0091	0.39 ± 0.39	0.252 ± 0.145	$0.61\ ^{+0.69}_{-0.28}$	$1.25\ ^{+0.8}_{-0.5}$ ± 0.38	4
1100	0.6274 ± 0.0061	0.39 ± 0.39	0.252 ± 0.145	$0.61\ ^{+0.69}_{-0.28}$	$1.25\ ^{+0.8}_{-0.5}$ ± 0.38	4
1150	0.4292 ± 0.0043	0.39 ± 0.39	0.252 ± 0.145	$0.61\ ^{+0.69}_{-0.28}$	$1.25\ ^{+0.8}_{-0.5}$ ± 0.38	4
1200	0.2989 ± 0.0027	0.39 ± 0.39	0.252 ± 0.145	$0.61\ ^{+0.69}_{-0.28}$	$1.25\ ^{+0.8}_{-0.5}$ ± 0.38	4

Table 5.16: Event yields at final selection level for the $\mu\mu jj$-channel search with $\sqrt{s} = 8$ TeV data. Other BG includes single top, Z+jets, and VV+jets. Uncertainties are Poisson uncertainties on the MC background, except for the second uncertainty for "Total BG", which gives the total systematic uncertainty as detailed in Section 5.5. Systematic uncertainties are dominated by energy scale and shape uncertainties.

5.6.2 Results with 5 fb^{-1}of pp Collisions at $\sqrt{s} = 7$ TeV

The number of observed events in data passing the full selection criteria is consistent with the SM background prediction in all decay channels. An upper limit on the leptoquark pair-production cross section is therefore set using the CL_S modified frequentist approach [122, 123]. A log-normal probability function is used to integrate over the systematic uncertainties. Uncertainties of statistical nature are described with Γ distributions with widths determined by the number of events simulated in MC samples or observed in data control regions.

The 95% CL upper limits on $\sigma \times \beta^2$ or $\sigma \times 2\beta(1-\beta)$ as a function of leptoquark mass are shown together with the NLO predictions for the scalar leptoquark pair-production cross section in Fig 5.23. The theoretical cross sections are represented for different values of the renormalization and factorization scale, μ, varied between half and twice the leptoquark mass (blue shaded region). The PDF uncertainties are taken into account in the theoretical cross section values.

By comparing the observed upper limit with the theoretical cross section values, second-generation scalar leptoquarks with masses less than 840 (620) GeV are excluded with the assumption that $\beta = 1$ (0.5). This is to be compared with median expected limit of 800 (610) GeV.

The observed and expected limits on the branching fraction β as a function of leptoquark mass can be further improved using the combination of the $\ell\ell jj$ and $\ell\nu jj$ channels, as shown in Figure 5.24. The combination is calculated for a given β value by reweighting the $\mu\mu jj$ and $\mu\nu jj$ signal contributions by the appropriate branching fraction, β^2 and $2\beta(-\beta)$, respectively. The combined limit is expressed as a function of β by considering a finely binned array of β values in the range of $0 < \beta \leq 1$. For a given β, the cross section limit at each LQ mass hypothesis is calculated. With the combined limit and the theoretical LQ cross section expressed as a function of

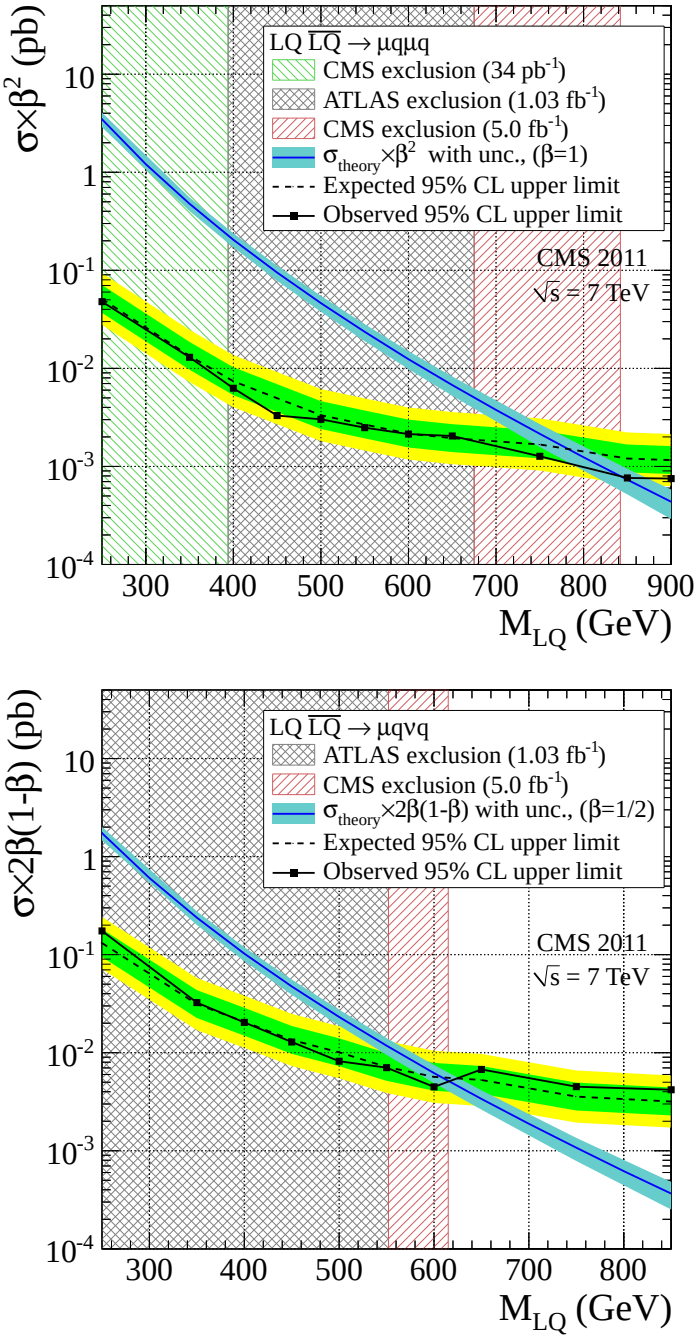

Figure 5.21: The top (bottom) plot gives the expected and observed upper limits at 95% CL on the LQ pair-production cross section times β^2 ($2\beta(1-\beta)$) as a function of the second-generation LQ mass. The expected limits and uncertainty bands represent the median expected limits and the 68% and 95% confidence intervals. The green and gray hatched areas represent the regions excluded by the previous CMS limit [124] and ATLAS limit [125], respectively. The red (rightmost) region is excluded by this analysis. The σ_{theory} curves and their bands represent, respectively, the theoretical scalar LQ pair-production cross section.

LQ mass, the intersection of these two curves, with a log-linear interpolation between points, is taken limit on LQ mass. The limit on LQ mass as a function of β is given in Fig. 5.24. This combination of channels results in an exclusion of second-generation scalar leptoquarks with masses less than 650 GeV for $\beta = 0.5$, compared with median expected limits of 670 GeV. This corresponds to an increase of 20 GeV (30 GeV) beyond that of the single $\mu\mu jj$ ($\mu\nu jj$) channel.

5.6.3 Results with 20 $\mathbf{fb^{-1}}$ of pp Collisions at $\sqrt{s} = 8$ TeV

In the same manner as in the 7 TeV search, data in the 8 TeV search are consistent with no-signal hypothesis. Upper 95% CL limits on the leptoquark production cross section are determined using the asymptotic CL_S modified frequentist approach [122,123], and a log-normal probability function is used to integrate over the systematic uncertainties.

The 95% CL upper limits on $\sigma \times \beta^2$ or $\sigma \times 2\beta(1-\beta)$ as a function of leptoquark mass are shown together with the NLO predictions for the scalar leptoquark pair production cross section in Fig. 5.23. The theoretical cross sections are represented for different values of the renormalization and factorization scale, μ, varied between half and twice the LQ mass (blue shaded region). The PDF uncertainties are taken into account in the theoretical cross section values. A slightly higher-than-expected cross-section upper limit in is visible in the $\mu\nu jj$ channel for the entire range of masses, due to a small excess in data which is well within the calculated systematic and statistical uncertainties. This excess is visible at all mass hypothesis because each mass hypothesis has a set of final selection events which is a subset of the those in the final selection of preceding mass hypothesis.

By comparing the observed upper limit with the theoretical cross section values, second-generation scalar leptoquarks with masses less than 1070 (740) GeV are excluded with the assumption that $\beta = 1$ (0.5). This is to be compared with median

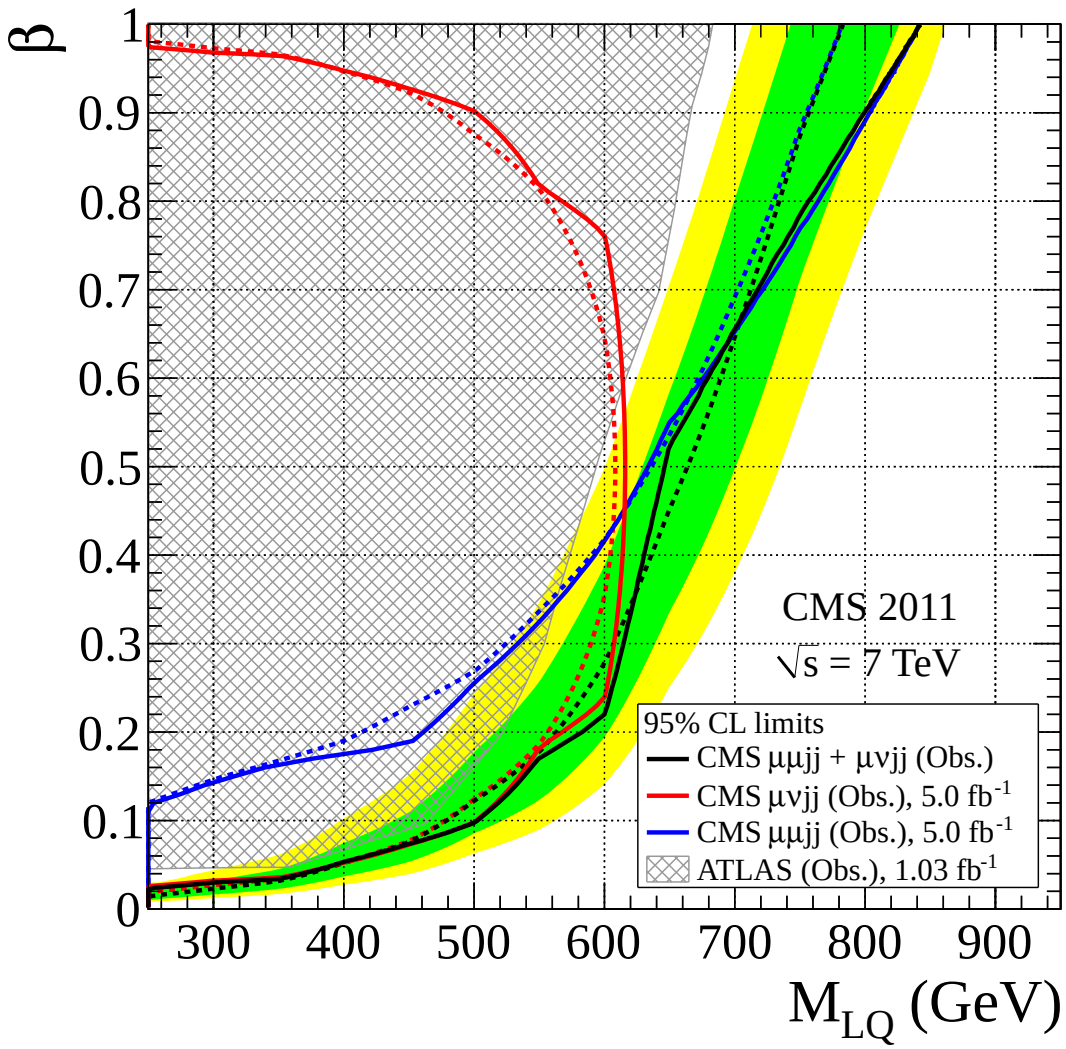

Figure 5.22: The expected and observed exclusion limits at 95% CL on the second-generation leptoquark hypothesis in the β versus mass plane using the central value of signal cross section for the individual $\mu\mu jj$ and $\mu\nu jj$ channels and their combination. The dark green and light yellow expected limit uncertainty bands represent the 68% and 95% confidence intervals. Solid lines represent the observed limits in each channel, and dashed lines represent the expected limits. The systematic uncertainties reported in Table 5.11 are included in the calculation. The shaded region is excluded by the current ATLAS limits [125].

expected limits of 1045 (790) GeV.

The combination of the $\mu\mu jj$ and $\mu\nu jj$ channels, shown in Figure 5.24, excludes LQ masses as a function of β using the intersection of the central theory cross-section and the excluded cross-section. The method of determining the mass limit as a function of β is the same as that described in Section 5.6.2. The combination improves the mass exclusion for values of $0 < \beta < 1$. Using the combined channels, second-generation scalar leptoquarks with masses less than 785 GeV can be excluded for $\beta = 0.5$, compared with an expected limit of 870 GeV. This represents an increase in the mass limit of 15 GeV (45 GeV) compared to the limit computed with the $\mu\mu jj$ ($\mu\nu jj$) channel only.

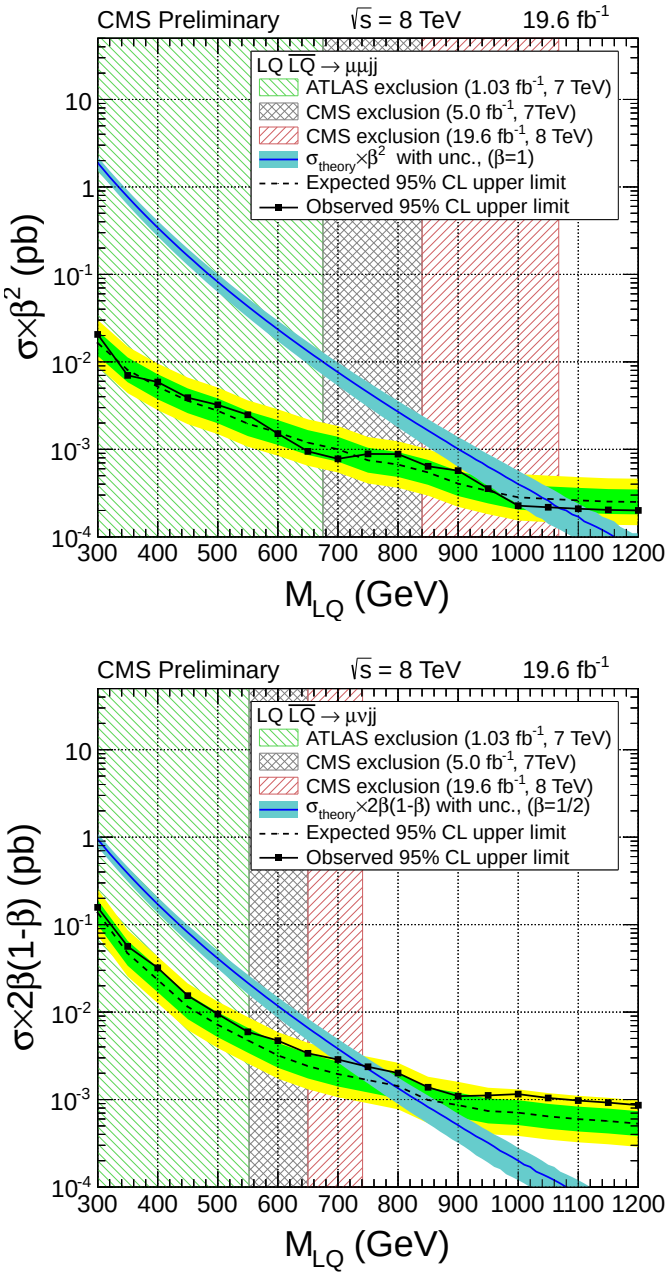

Figure 5.23: The top (bottom) plots gives the expected and observed upper limits at 95% CL on the LQ pair production cross section times β^2 ($2\beta(1-\beta)$) as a function of the second-generation LQ mass obtained with the $\mu\mu jj$ ($\mu\nu jj$) analysis. The expected limits and uncertainty bands represent the median expected limits and the 68% and 95% confidence intervals. The green hatched region is excluded by the ATLAS limit [125] determined with 7 TeV data. The gray hatched region is excluded by the CMS limit detailed in Section 5.6.2. The red rightmost hatched region is newly excluded. The σ_{theory} curves and their bands represent, respectively, the theoretical scalar LQ pair production cross section and the uncertainties due to the choice of PDF and renormalization/factorization scales.

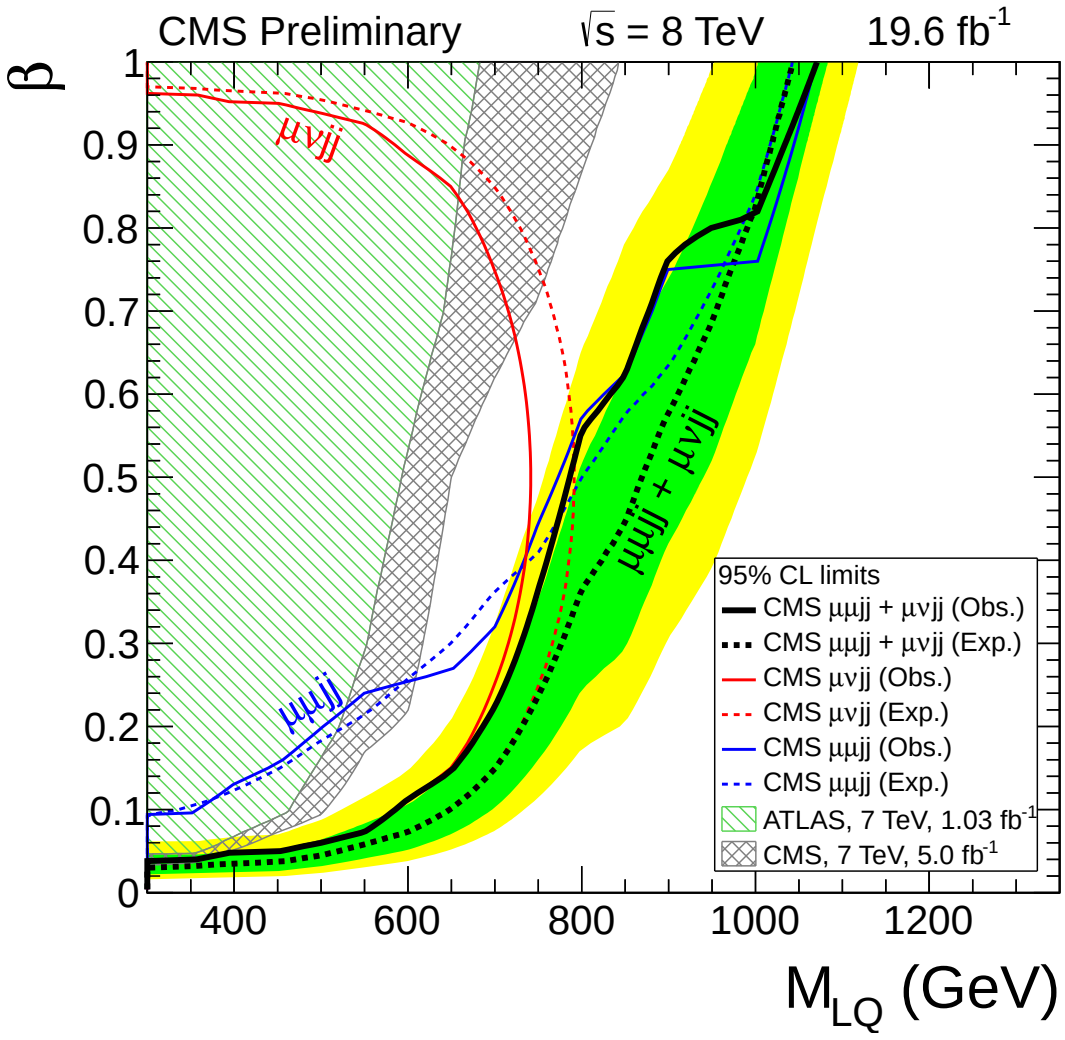

Figure 5.24: The expected and observed exclusion limits at 95% CL second-generation leptoquark mass as a function of the branch fraction β. Limits are expressed for the central value of signal cross section given in Table 1.3. The dark green and light yellow expected limit uncertainty bands represent the 68% and 95% confidence intervals on the combination. Limits for the individual $\mu\mu jj$ and $\mu\nu jj$ channels are also given as in the mass versus beta plane. Solid lines represent the observed limits in each channel, and dashed lines represent the expected limits. The leftmost shaded region is excluded by the most recent ATLAS 7TeV result [125], and the rightmost shaded region is excluded by the CMS 7 TeV Result detailed in Section 5.6.2.

Chapter 6

W+jets Differential Cross Section Measurements

Differential cross sections are measured for the W+jets process, using data corresponding to 5 fb^{-1} of pp collisions at $\sqrt{s} = 7\,\text{TeV}$ [126]. Cross section measurements are performed as a function of the inclusive and exclusive jet multiplicities, the transverse momentum and the pseudorapidity of the jets, the difference in azimuthal angle between the direction of each jet and that of the muon, and H_{T}, which is defined as the scalar sum of the p_{T} of all jets with $p_{\text{T}} > 30$ GeV and $\eta < 2.4$. It is important to study the distributions of the jet p_{T} and the observable H_{T} because they are sensitive to higher order corrections, and are often used to discriminate against background in searches for signatures of physics beyond the SM. Additionally, H_{T} is commonly used as a component of the factorization and renormalization scale in event generators simulating W+jets production at the LHC. The η distributions of jets and the azimuthal separations between the jets and the muon are also important, because they are sensitive to the modeling of parton emission.

The measurements are conducted in the following manner:

1. Simulations of the W+jets signal and all relevant backgrounds are performed, with sufficient numbers of events generated to fully model the data-taking period.

2. The simulated signal and background are fully reconstructed to be compared directly with the real data.

3. An event selection is used to obtain a high-purity W+jets signal sample.

4. The background predictions are checked in background-enriched control regions, and the QCD multijet background, which can not be simulated, is determined with a sample of QCD enriched data in a region orthogonal to the signal region.

5. For each distribution to be measured, the distribution of the backgrounds to the W+jets signal is subtracted from the distribution of the data, to estimate the W+jets contribution to the reconstructed data.

6. This distribution is then deconvolved (unfolded) for detector effects and corrected for efficiencies and acceptances, using the reconstructed W+jets signal MC to quantify the detector effects.

7. This entire procedure is repeated for a multitude of systematic uncertainties to assess variations in the final unfolded distributions, and the unfolded distributions are presented with these uncertainties.

8. Comparisons of the unfolded data to predictions from different MC simulations and NLO predictions are presented.

This Chapter proceeds as follows: Section 6.1 describes the samples of real and simulated data, and Section 6.2 details the event selection applied to the samples. Section 6.3 describes the methods used to estimate the backgrounds to the W+jets process, using simulated samples and control samples in data. Section 6.4 describes

the particle-level event selection (the fiducial region) for which the cross section will be calculated. Section 6.5 describes the deconvolution of the detector effects via regularized unfolding. Section 6.6 describes the sources of systematic uncertainty, and how they affect the measured distributions. Section 6.7 gives details of the theoretical predictions from event generators and NLO predictions, to which the measured distributions will be compared. Lastly, Section 6.8 presents the measured distributions, and describes the agreement with the different predictions.

6.1 Data and Simulation Samples

These measurements use proton-proton (pp) collision data at a center-of-mass energy of $\sqrt{s} = 7$ TeV recorded with the CMS detector at the LHC in 2011 and corresponding to an integrated luminosity of $5.0 \pm 0.1\text{fb}^{-1}$ [118].

All fully-reconstructed MC samples in this analysis were produced by the CMS collaboration. The reconstruction of simulated events in the CMS detector based on GEANT4 [90], as is further detailed in Section 3. MC Samples and corresponding cross-sections and uncertainties are detailed in Table 6.1. The W+jets and Z/γ^*+jets samples are normalized to next-to-next-to-leading order (NNLO) inclusive cross sections calculated with FEWZ [106]. Single top-quark and VV samples are normalized to NLO inclusive cross sections calculated with MCFM [109–112]. The $t\bar{t}$ contribution is normalized to the NNLO + next-to-next-leading logarithm (NNLL) predicted cross section from Ref. [127].

The primary W$\rightarrow \ell\nu$+jets signal sample is generated with MadGraph 5.1.1 [41], and is used to determine the detector response in the unfolding procedure described in Section 6.5. Parton showering and hadronization of the MadGraph samples is performed with Pythia 6.424 using the Z2 tune [128]. The detector response is also

Process	MC Generator	σ [pb]	Unc(σ) [%]
W+jets	MadGraph	31314	–
W+jets	Sherpa	31314	–
$t\bar{t}$	MadGraph	172	8 %
Z/γ^*+jets	MadGraph	3048	4.3 %
t (t-chan)	Powheg	42.6	6 %
t (s-chan)	Powheg	2.97	6 %
t (tW-chan)	Powheg	5.3	6 %
$\bar{t}$ (t-chan)	Powheg	22	6 %
$\bar{t}$ (s-chan)	Powheg	1.63	6 %
$\bar{t}$ (tW-chan)	Powheg	5.3	6 %
WW	Pythia	43	4 %
ZZ	Pythia	5.9	4 %
WZ	Pythia	18.2	4 %

Table 6.1: The MC generators used for W+jets signal and SM background simulation, and the corresponding cross sections.

determined using a different W+jets event sample generated with Sherpa 1.3.0 [43–46], and is used in the evaluation of systematic uncertainties due to the unfolding of the data.

The main sources of background are the production of $t\bar{t}$, single top-quark, Z/γ^*+jets, dibosons (ZZ/WZ/WW), and multijets. With the exception of multijet production, all backgrounds are estimated from simulation. The simulated samples of $t\bar{t}$ and Z/γ^*+jets are generated with MadGraph 5.1.1; single top-quark samples (s-, t-, and tW- channels) are generated with Powheg [85–88]; VV samples, where V represents either a W boson or a Z boson, are generated with version 6.424 using the Z2 tune [128]. The simulations with MadGraph and Pythia use the CTEQ6L1 parton distribution functions (PDF) [89]. The simulation with Sherpa uses the CTEQ6.6m PDF, and the simulations with use the CTEQ6m PDF.

These simulations include multiple collisions in a single bunch crossing (pileup) by matching the number of pileup events delivered by the LHC during the data-taking

periods.

6.2 Event Selection

Events are retained if they pass a trigger requiring one isolated muon with $p_{\mathrm{T}} > 24$ GeV and $|\eta| < 2.1$. Simulated events are required to pass an emulation of the trigger requirements applied to the data.

Muon candidates are reconstructed as tracks in the muon system that are matched to tracks reconstructed in the inner tracking system [113]. Muon candidates are required to have $p_{\mathrm{T}} > 25$ GeV, and to be reconstructed within the fiducial volume used for the high-level trigger muon selection, i.e. within $|\eta| < 2.1$. This ensures that the offline event selection requirements are as stringent as the trigger. To ensure a precise measurement of the transverse impact parameter of the muon track relative to the interaction point, only muon candidates with tracks containing more than 10 hits in the silicon tracker and at least one hit in the pixel detector are considered. To reject muons from cosmic rays, the transverse impact parameter of the muon candidate with respect to the primary vertex is required to be less than 2 mm. These identification criteria correspond to the "Tight muon" criteria described in Section 3.5.1. In addition, an isolation requirement is applied to the muon candidates by demanding that the relative isolation is less than 0.15, where the relative isolation is defined as the sum of the transverse energy deposited in the calorimeters (ECAL and HCAL) and of the p_{T} of charged particles measured with the tracker in a cone of $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.3$ around the muon candidate track (excluding this track), divided by the muon candidate p_{T}.

Jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ are reconstructed using the CMS particle-flow algorithm [77, 129], describe in Section 3.4. Jets are reconstructed using the anti-k_{T} [80, 130] algorithm with a distance parameter of 0.5. The jet energy is calibrated using the p_{T} balance

of dijet and γ+jet events [81]. Only jets with $p_{\mathrm{T}} > 30$ GeV, $|\eta| < 2.4$, and a spatial separation of $\Delta R > 0.5$ from the muon are considered. To reduce the contamination from pileup jets, jets are required to be associated to the same primary vertex as the muon. The vertex associated to each jet is the one that has the largest number of p_{T}-weighted tracks in common with the jet. The contamination from pileup jets is estimated with the signal simulation, with pileup events simulated with PYTHIA , and found to be less than 1%.

The missing momentum vector, $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, is defined as the negative of the vectorial sum of the transverse momenta of the particles reconstructed with the particle-flow algorithm, and the $E_{\mathrm{T}}^{\mathrm{miss}}$ is defined as the magnitude of the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ vector. The measurement of the $E_{\mathrm{T}}^{\mathrm{miss}}$ in simulation is sensitive to the modeling of the calorimeter response and resolution and to the description of the underlying event. To account for these effects, the $E_{\mathrm{T}}^{\mathrm{miss}}$ in W+jets simulation is corrected for the differences in the detector response between data and simulation, using a method detailed in Ref. [131]. A recoil energy correction is applied to the W+jets simulation on an event-by-event basis, using a sample of Z$\rightarrow \mu\mu$ events in data and simulation. The transverse recoil vector, defined as the negative vector sum of the missing transverse energy and the transverse momenta of the lepton(s), is divided into components parallel and perpendicular to the boson direction. The mean and the width of the transverse recoil vector components are parameterized as a function of the Z-boson p_{T} in data and simulation. The ratio of the data and simulation parameterizations is used to adjust the transverse recoil vector components in each simulated event, and a new $E_{\mathrm{T}}^{\mathrm{miss}}$ is computed using the corrected recoil components.

Events are required to contain exactly one muon satisfying the conditions described above and one or more jets with $p_{\mathrm{T}} > 30$ GeV. Events are required to have $M_{\mathrm{T}} > 50$ GeV, where M_{T}, the transverse mass of the muon and missing transverse energy,

is defined as $M_{\mathrm{T}} \equiv \sqrt{2p_{\mathrm{T}}^{\mu} E_{\mathrm{T}}^{\mathrm{miss}} (1 - \cos\Delta\phi)}$, where p_{T}^{μ} is the muon p_{T} and $\Delta\phi$ is the difference in azimuthal angle between the muon momentum direction and the $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ vector.

6.3 Estimation of the Backgrounds and Selection Efficiencies

6.3.1 The Normalization of Simulation Samples

All background sources except for the multijet production are modeled with simulation. The simulated event samples are corrected for differences between data and simulation in muon identification efficiencies and event trigger efficiency. A "tag-and-probe" method [113] is used to determine the differences between simulation and data for the efficiency of the trigger and for the muon identification and isolation criteria. This method uses $\mathrm{Z} \rightarrow \mu\mu$ events from both data and simulated samples where the "tag" muon is required to pass the most stringent muon identification and isolation criteria. The efficiency measurements use the "probe" muon, which is required to pass minimal quality criteria. Trigger efficiency corrections are determined as a function of the muon η, and are in general less than 5%. Muon isolation and identification efficiency corrections are determined as a function of the muon p_{T} and η, and are generally less than 2%. Corrections to the simulation are applied on an event-by-event basis in the form of event weights.

6.3.2 Data-Driven Normalization of Major Backgrounds

The shape and normalizations of the Z/γ^{*}+jets and $\mathrm{t\bar{t}}$ predictions are cross-checked in selected data samples. The Z+jets background is compared to data in a Z-boson-

dominated data sample that requires two well-identified, isolated muons. The $t\bar{t}$ background is compared to data in a control region requiring at least two b-tagged jets. Background estimations from simulation and from data control samples agree within the uncertainties described in Section 6.6.

6.3.3 The QCD Multijet Background Estimation

The QCD multijet background is estimated using a data control sample with an inverted muon isolation requirement. The shape and normalization of the QCD background is estimated entirely from data, using a control sample of data events with inverted isolation. The process of QCD normalization is repeated for every distribution being measured, since different distributions require different jet multiplicities, which can affect the efficiency of the isolation in QCD. For the case of examining jet multiplicity itself, the QCD isolation efficiency is determined in bins of jet multiplicity. For all other distributions it is determined as a single value as described below. Also, since the QCD measurement is affected by the subtraction of other backgrounds, namely W+jets, the W+jets contribution is normalized to data in the $M_{\mathrm{T}} > 50$ GeV region before the QCD studies are conducted.

Control regions are defined as described in Table 6.2, and the process for determining the QCD background distributions is as follows:

	$M_{\mathrm{T}} > 50$ GeV	$M_{\mathrm{T}} < 50$ GeV
$\mathrm{ISO_{Rel}} < 0.15$	**A.** Signal region	**B.** Numerator of fake rate calculation
$\mathrm{ISO_{Rel}} > 0.15$	**C.** QCD shape	**D.** Denominator of fake rate calculation

Table 6.2: Definition of control regions used in QCD study.

0. In region A, a global scale factor for the W+jets simulation is calculated as

$$f_{\mathrm{W}} = \frac{N^{A}_{\mathrm{Data}} - N^{A}_{\mathrm{Background}}}{N^{A}_{\mathrm{W}}}. \tag{6.1}$$

Here, N^{A}_{Data} denotes the total number of data events in region A, N^{A}_{W} denotes the total number W+jets MC events in region A, and $N^{A}_{\mathrm{Background}}$ denotes the total number of background events in region A, except for W+jets, i.e. Z+jets, $\mathrm{t\bar{t}}$, single-top, and di-boson. The scale factor f_{W} is applied to all W+jets contributions in regions A,B,C and D for calculations below.

1. In region C, a histogram (QCD^{C}_{i}) is created, with i bins in the same binning to be used for unfolding in the signal region. The QCD^{C}_{i} histogram determines the QCD shape, and is given by:

$$\mathrm{QCD}^{C}_{i} = \mathrm{Data}^{C}_{i} - \mathrm{MC}^{C}_{i}. \tag{6.2}$$

Here, MC^{C}_{i} consists of all MC, including W+jets.

2. Regions B and D are used to calculate the isolation fake rate:

$$f_{B/D} = \frac{N^{B}_{\mathrm{Data}} - N^{B}_{\mathrm{MC}}}{N^{D}_{\mathrm{Data}} - N^{D}_{\mathrm{MC}}}. \tag{6.3}$$

Here, N^{B}_{Data} and N^{D}_{Data} are the number of data events in data in regions B and D respectively, and N^{B}_{MC} and N^{D}_{MC} are the number of MC events, including W+jets, in regions B and D respectively.

5. The isolation fake-rate ($f_{B/D}$) is applied to the histogram from step [1.] to give

a final QCD histogram to be used in the signal region:

$$\mathrm{QCD}_i^A = f_{B/D} \times \mathrm{QCD}_i^C. \tag{6.4}$$

6 Since f_{W} should consider QCD contamination, which had not yet been estimated in Step (0.), Steps (0. - 5.) are repeated several times until the values of f_{W} and $f_{B/D}$ stabilize (See Table 6.3).

The QCD multijet contribution to the total W+jets background estimate is 32.7% (1.9%) for selections requiring at exactly 1(4) jets, or 2.6%(0.3%) of the total SM prediction from MC. The total uncertainty on the QCD prediction ranges between 1.3% and 31% for jet multiplicities of 1–4.

6.3.4 Suppression of the $t\bar{t}$ Background

The dominant background to W+jets production is $t\bar{t}$ production, which has a larger total contribution than that of the W+jets signal in events with four or more jets. In order to reduce the level of $t\bar{t}$ contamination, a veto is applied to events with one or more b-tagged jets. Heavy-flavor tagging is based on a tag algorithm [132] that exploits the long lifetime of b-quark hadrons. This algorithm calculates the signed impact parameter significance of all tracks in the jet that satisfy high-quality reconstruction and purity criteria, and orders the tracks by decreasing significance. The impact parameter significance is defined as the ratio of the impact parameter to its estimated uncertainty. For jets with two or more significant tracks, a high-efficiency b-quark discriminator is defined as the significance of the second-most significant track. The size of the $t\bar{t}$ background is illustrated in Fig. 6.1, before and after the implementation of the b-jet veto, using the event selection described in Section 6.2. The expected contributions for the different processes in Fig. 6.1 are shown as a function of the jet multiplicity, along with

	≥ 1 jet			≥ 2 jet		
It.	f_{W}	$f_{B/D}$	$\delta_{B/D}$ (%)	f_{W}	$f_{B/D}$	$\delta_{B/D}$ (%)
0	0.956	0.04121	0.9	0.960	0.02043	1.4
1	0.981	0.03984	0.9	0.986	0.01955	1.4
2	0.981	0.03989	0.9	0.985	0.01959	1.4
3	0.981	0.03989	0.9	0.985	0.01959	1.4
4	0.981	0.03989	0.9	0.985	0.01959	1.4
	≥ 3 jet			≥ 4 jet		
It.	f_{W}	$f_{B/D}$	$\delta_{B/D}$ (%)	f_{W}	$f_{B/D}$	$\delta_{B/D}$ (%)
0	0.889	0.01067	3.9	0.868	0.00745	9.0
1	0.908	0.00998	3.9	0.883	0.00692	9.0
2	0.907	0.01003	3.9	0.882	0.00695	9.0
3	0.907	0.01002	3.9	0.882	0.00695	9.0
4	0.907	0.01002	3.9	0.882	0.00695	9.0

Table 6.3: The iterative determination of f_{W} and $f_{B/D}$ for QCD multijet background estimation, providing a few percent correction to each factor. The resulting QCD estimation is less-dependent on the MC modeling of the W+jets contribution which must be subtracted in the QCD control regions. The factor f_{W} plays no additional role in analysis.

the observed data. Differences in the tagging and mistagging rates between data and simulation are measured as a function of the jet p_{T} in multijet and $\mathrm{t\bar{t}}$ events [132], and are used to correct the tagging rates of the jets in simulation. For jet multiplicities of 1 to 6, the b-jet veto eliminates 44 – 84% of the predicted $\mathrm{t\bar{t}}$ background, while eliminating 3 – 26% of the predicted W+jets signal.

6.4 The Particle Level Selection

For the measurement of cross sections, the particle level is defined by a W boson, which decays into a muon and a muon neutrino, produced in association with one or more

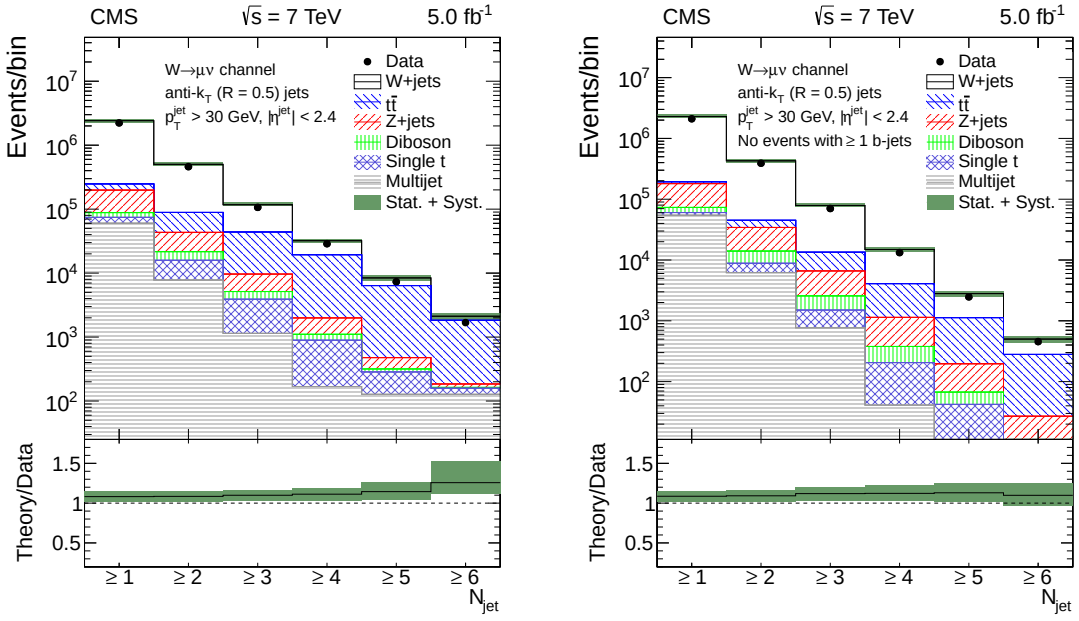

Figure 6.1: The jet multiplicity in data and simulation before (left) and after (right) the b-jet veto. The W+jets contribution is modeled with MadGraph 5.1.1+Pythia 6.424. The solid band indicates the total statistical and systematic uncertainty in the W+jets signal and background predictions, as detailed in Section 6.6. This includes uncertainties in the jet energy scale and resolution, the muon momentum scale and resolution, the pileup modeling, the b-tagging correction factors, the normalizations of the simulations, and the efficiencies of reconstruction, identification, and trigger acceptance. A substantial reduction in the expected $t\bar{t}$ background is observed in the right plot.

jets. Decays of the W boson to a tau and a tau neutrino, resulting in a final state with a muon, are not included in the particle level definition. Kinematic thresholds on the particle-level muon, M_{T}, and jets are identical to those applied to the reconstructed objects. The particle-level selection also includes the requirement of exactly one muon with $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.1$, and $M_{\mathrm{T}} > 50$ GeV. The particle-level $E_{\mathrm{T}}^{\mathrm{miss}}$ is defined as the negative of the vectorial sum of the transverse momenta of all visible final state particles. To account for final-state radiation, the momenta of all photons in a cone of

$\Delta R < 0.1$ around the muon are added to that of the muon. Jets are clustered using the anti-k_{T} [80] algorithm with a distance parameter of 0.5. Clustering is performed using all particles after decay and fragmentation, excluding neutrinos and the muon from the W-boson decay. Additionally, jets are required to have $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.4$, and to be separated from the muon by $\Delta R > 0.5$.

6.5 Unfolding

The reconstructed distributions are corrected to the particle level with the method of regularized singular value decomposition (SVD) [133] unfolding, using the RooUnfold toolkit [134]. For each distribution, the total background, including the multijet estimate from data and all simulated processes except the W boson signal, is subtracted from the data before unfolding. A response matrix, defining the migration probability between the particle-level and reconstructed quantities, as well as the overall reconstruction efficiency, is computed using W+jets events simulated with MadGraph +Pythia .

The SVD unfolding procedure begins with a response (or migration) matrix $\hat{A}$ which describes the relationship between a true distribution x^{ini} and a measured distribution b^{ini} as given by an MC simulation and reconstruction,

$$\hat{A}x^{\mathrm{ini}} = b^{\mathrm{ini}}. \tag{6.5}$$

The exact solution to Equation 6.5 would entail calculating the exact inverse of $\hat{A}$, but this would yield an unstable result with rapidly oscillating bin content. Instead, the regularized SVD method seeks to remove non-significant terms from the solution via regularization, to produce a stable result.

To do so, $\hat{A}$ is factorized into the product of an orthogonal matrix ($\hat{U}$), a diagonal

matrix ($\hat{S}$), and a transposed orthogonal matrix ($\hat{V}^T$),

$$\hat{A} = \hat{U}\hat{S}\hat{V}^T. \tag{6.6}$$

The diagonal elements of $\hat{S}$ in Equation 6.6 can be expressed as $S_{ii} \equiv s_i$ and are referred to as the singular values of $\hat{A}$. An ordering of the s_i is performed by interchanging rows of $\hat{U}$ and $\hat{V}$, which retains the same set of equations.

The goal is to solve for a data distribution x which is observed after reconstruction as b, and characterized with the response matrix as $\hat{A}x = b$. An exact solution for x in terms of the SVD decomposition is given in Equation 6.7.

$$x = \hat{V}\hat{S}^{-1}\hat{U}^T b \tag{6.7}$$

Typically, the last two terms of the r.h.s. of Equation 6.7 is expressed as $d \equiv \hat{U}^T b$. When the values of d are not statistically significant, the exact solution from Equation 6.7 will not yield sensible results.

Assuming that for some value of i, values of d_i will be come insignificant, the SVD solution can be regularized according to the singular values s_i and and regularization parameter τ, such that:

$$d_i'(\tau) = d_i \frac{s_i^2}{s_i^2 + \tau} \tag{6.8}$$

The recommended choice of τ is $\tau = s_k^2$, where d_k is the first statistically insignificant component of d. In practice, this allows for i different choices of τ, and most implementations including RooUnfoldSvd allow instead for the corresponding choice of k, often referred to as k_{reg}.

A more through explanation of the recommended procedure is given in Reference [133].

6.5.1 The Unfolding Response

The unfolding is trained with a response object which accounts for:

1. Bin-to-bin migration between the generated and reconstructed quantities,

2. The efficiency of reconstruction. Events with a generated quantity and no corresponding reconstructed quantity are considered to have missed the reconstruction, and are termed "misses")

3. The occurrence of reconstructed quantities with no real corresponding generated quantity. Events in this category are considered "fakes".

To accomplish this task, the response is built from three objects:

1. *HGenVsReco2D*, a two-dimensional histogram containing events in the particle-level phase space which also pass the reconstructed level selection.

2. *HGen1D*, a one-dimensional histogram of the particle-level quantity filled with events in the particle-level phase space.

3. *HReco1D*, a one-dimensional histogram of the reconstructed quantity filled with events passing the selection applied to the real data.

Additionally, some kinematic quantities benefit from an underflow and overflow region of the reconstructed quantity to account for the migration of events with a generated quantity in the phase space and a reconstructed quantity outside the phase space. This can be made clearer with a couple examples.

For Jet multiplicity, the response components are as follows:

1. *HGenVsReco2D* is filled with events with one or more reconstructed jets and one or more particle-level jets.

2. *HGen1D* is filled with events with one or more particle-level jets.

3. *HReco1D* is filled with events with one or more reconstructed jets.

For the lead jet transverse momentum, the reconstructed p_{T} range is defined as 10-910 GeV, and the particle-level p_{T} range is defined as 30-850 GeV. The differences in reconstructed and particle-level ranges provide a migration underflow/overflow for a more accurate measurement. The response components are as follows.

1. *HGenVsReco2D* is filled with events with one or more reconstructed jets (in the reconstructed range) and one or more particle-level jets (in the particle-level range). The reconstructed quantity is the p_{T} of the highest-p_{T} identified and reconstructed jet. The particle-level quantity is the p_{T} of the highest-p_{T} particle-level jet meeting the particle-level jet selection criteria. No matching between the particle-level and reconstructed jet is applied.

2. *HGen1D* is filled with events with one or more particle-level jets in the particle-level range. The particle-level quantity is the p_{T} of the highest-p_{T} particle-level jet meeting the particle-level jet selection criteria. No matching between the particle-level and reconstructed jet is applied.

3. *HReco1D* is filled with events with one or more reconstructed jets in the reconstructed range. The reconstructed quantity is the p_{T} of the highest p_{T} identified and reconstructed jet.

A similar procedure is adopted for jet η, where the reconstructed jet range is $-2.5 < \eta < 2.5$ and the particle-level jet range is $-2.4 < \eta < 2.4$. Quantities such as $\Delta\phi$ have no overflow or underflow because all quantities must occupy the range $0 - \pi$. Unfolding of the H_T also has no underflow or overflow, because the H_T is defined as the scalar

sum of jet transverse momenta, where a jet is defined to have $p_T > 30$ GeV, so no $H_T < 30$ may exist.

6.5.2 The Unfolding Procedure

The unfolding procedure can be summarized as follows:

1. Choose an appropriate range and binning for desired distribution for the measurement. During the unfolding, the reconstructed variable will be restricted to this range.

2. Choose an appropriate larger range at the particle-level for the variable. This underflow/overflow region will account for events in which the particle-level value differs enough from the reconstruction-level value such that it would be outside the reconstruction-level range specified in step (1).

3. Choose a binning for the unfolding distributions and response matrix in which the bin widths are less than or equal to the smallest bin width to be used for the final presentation distributions. Unfolding binning is chosen such that every x-value for every bin-edge in the presentation distribution is also a bin-edge for the unfolding binning (e.g. if the presentation binning is [0,10,20,40], the unfolding binning could be [0,5,10,15,20,25,30,35,40,50], but could not be [0,15,30,45,60]).

4. The response matrix is created using W+jets MADGRAPH MC and the binning specified in step (3). The response matrix is constructed for events in the particle-level phase space passing reconstructed event selection, including cuts on M_{T}, muon p_{T}, jet p_{T}, and the b-tagged jet event veto.

5. The regularization parameter k_{reg} is chosen as the value of i for which the remaining d_i values become insignificant, falling significantly below unity or becoming

flat as a function of i. All values of i between 2 and half the number of bins are considered.

6. The distributions for unfolded data and reconstruction-level data are compared. A histogram rebinning is applied to convert from the unfolding binning used for the SVD derived in step (3) to the final presentation-style binning, initially chosen in step (1).

For a given particle-level quantity Q with a corresponding reconstructed quantity $Q^{'}$, the migration probability from an interval $a < Q < b$ to an interval $c < Q^{'} < d$ is defined as the fraction of events with $a < Q < b$ that have $c < Q^{'} < d$. The unfolding of the jet multiplicity is performed with a response defined by the number of particle-level jets versus the number of reconstructed jets. For particle-level jet multiplicities of 1 to 6, 4% to 51% of simulated events exhibit migration to different values of reconstructed jet multiplicity. The unfolding of the kinematic distributions of the n^{th} jet is performed with a response defined by the kinematic quantity of the n^{th}-highest-p_{T} particle-level jet versus that of the n^{th}-highest-p_{T} reconstructed jet. To achieve a full migration from the selection of reconstructed events to the particle-level phase space, no matching between reconstructed and particle-level jets is applied. Since a reconstructed event may contain a final state muon from the $\text{W} \to \tau\nu$ process, the contamination from the $\text{W} \to \tau\nu$ process is estimated with simulation to be at the 1% level, and is removed during the unfolding procedure.

The b-jet veto is treated as an overall event selection condition in the same manner as the basic kinematic conditions on the M_{T} and muon p_{T}. Events failing this condition are treated as nonreconstructed in the unfolding response, so that the cross section obtained after unfolding is valid for W-boson decays with associated jets of any flavor.

Systematics uncertainties, discussed further in Section 6.6, are assessed by performing variations and repeating the entirety of the unfolding procedure to estimate the effect on the final unfolded data distribution.

6.5.3 Unfolded Distributions

For each measure distribution, the plots detailing the different components of the unfolding procedure are provided in Appendix A. The format of these plots is as follows:

- Top Left: A histogram of reconstructed data vs reconstructed simulation with the binning and axis range of the final measurement distribution.
- Bottom Left: The particle-level-vs-reconstructed quantity two-dimensional histogram representing the response matrix to be used in the unfolding. Grid lines are superimposed to indicate two regions. The central region is the region in which the unfolded result will be derived. The periphery is an underflow/overflow region. The reconstruction is confined to the central region, whereas the particle-level quantity spans the central region and periphery.
- Top right: The distribution of the simulated W+jets and the reconstructed and particle-levels, along with reconstructed and unfolded data. The underflow/overflow periphery is marked with dashed lines. Binning is chosen according to the procedure described in Section 6.5.2. Also drawn on this plot is the closure test of the unfolding procedure conducted on the reconstruction-level MC.
- Bottom right: The distributions, rebinned with a variable binning, for the unfolded data divided by reconstructed data, superimposed with the particle-level quantity divided by the reconstructed quantity in simulation.

6.6 Systematic Uncertainties

The sources of systematic uncertainties considered in this analysis are described below. The entirety of the unfolding procedure is repeated for each systematic variation, and the unfolded data results with these variations are compared with the central (unvaried) results to extract the uncertainties in the unfolded data distributions.

In most distributions, the dominant sources of systematic uncertainty include the jet energy scale and resolution uncertainties, which affect the shape of all reconstructed distributions, as well as the overall event acceptance. The jet energy scale uncertainties are estimated by assigning a p_{T}- and η-dependent uncertainty in jet energy corrections as discussed in Ref. [81], and by varying the jet p_{T} by the magnitude of the uncertainty. The uncertainties in jet energy resolution are assessed by increasing the p_{T} difference between the reconstructed and particle-level jets by an η-dependent value [81]. The jet energy uncertainties are determined by varying the p_{T} of the jets in data rather than in simulation.

Muon momentum scale and resolution uncertainties also introduce uncertainties in the overall event acceptance. A muon momentum scale uncertainty of 0.2% and a muon momentum resolution uncertainty of 0.6% are assumed [113]. The effects of these uncertainties are assessed by directly varying the momentum scale and randomly fluctuating the muon momentum in the simulation.

Variations for uncertainties in the energy and momentum scales and resolutions affect the size and shape of the background distribution to be subtracted from the data distribution, as well as the acceptance of W+jets simulated events, which define the response matrix used for unfolding. The variations are also propagated to the measurement of $E_{\mathrm{T}}^{\mathrm{miss}}$, which affects the acceptance of the $M_{\mathrm{T}} > 50$ GeV requirement.

Another important source of systematic uncertainty is the choice of the generator

used in the unfolding procedure. The size of this uncertainty is assessed by repeating the unfolding procedure with a response trained on a separate simulated sample generated with SHERPA 1.3.0. The absolute value of the difference between the data unfolded with a response matrix trained on SHERPA and with a response matrix trained on MADGRAPH is treated as a symmetric uncertainty on the measurement.

Other minor sources of systematic uncertainty include the uncertainties in the background normalization, the b-tagging efficiency, the modeling of the Wb contribution in the signal simulation, integrated luminosity, the pileup modeling, the trigger and object identification efficiencies, and the finite number of simulated events used to construct the response. Background normalization uncertainties are determined by varying the cross sections of the backgrounds within their theoretical uncertainties. For the Z+jets process, a normalization uncertainty of 4.3% is calculated as the sum in quadrature of the factorization/renormalization scale and PDF uncertainties calculated in FEWZ [106]. For the diboson and single top-quark processes, uncertainties are calculated with MCFM [109–112] to be 4% and 6%, respectively. The uncertainty on the $t\bar{t}$ modeling is assessed by taking the difference between data and simulation in a control region with two or more b-tagged jets, and is estimated to be 5 to 12% for jet multiplicities of 2 to 6. The estimate of the multijet background has an uncertainty based on the limited number of events in the multijet sample and in the control regions where the multijet sample normalization is calculated, and other systematic variations affecting the backgrounds in the multijet control regions introduce variations in the multijet normalization and template shape. For the b-tagging algorithm used to veto events containing b jets, uncertainties in the data/simulation ratio of the b-tagging efficiencies are applied. For jets with $p_{\mathrm{T}} > 30$ GeV, these uncertainties range from 3.1% to 10.5%. An additional uncertainty is ascribed to the normalization of the Wb content in the simulation by examining the agreement between data and simulation

as a function of jet multiplicity in a control region defined by requiring exactly one b-tagged jet. An increase in the normalization of the Wb process of 120% is considered, yielding an uncertainty in the measurement of 0.5 to 11% for jet multiplicities of 1 to 6. The uncertainty in the integrated luminosity is 2.2% [118]. An uncertainty in the modeling of pileup in simulation is determined by varying the number of simulated pileup interactions by 5% to account for the uncertainty in the luminosity and the uncertainty in the total inelastic cross section [121], as determined by a comparison of the number of reconstructed vertices in $Z \to \mu\mu$ events in data and simulation. Uncertainties in the differences between data and simulation efficiencies of the trigger, muon isolation, and muon identification criteria are generally less than 1%. An additional uncertainty due to the finite number of simulated events used to construct the response matrix is calculated by randomly varying the content of the response matrix according to a Poisson uncertainty in each bin.

The effect of the systematic variations on the measured cross section as a function of the exclusive jet multiplicity is illustrated in Fig. 6.2. The uncertainties given in Fig. 6.2 are the total uncertainty for each jet multiplicity. The corresponding ranges of systematic uncertainty across bins of jet p_{T} are given in Table 6.4.

n	p_{T} [GeV]	Statistical	JES, JER	Generator	MC stat.	Other	Total
1	30–850	0.1–3.2	3.4–24	0.9–9.6	0.2–11	2.3–6.6	4.5–29
2	30–550	0.4–2.4	4.6–12	1.6–13	0.8–11	2.9–5.9	6.9–21
3	30–450	0.6–16	6.0–23	2.7–48	1.0–45	4.5–11	9.4–73
4	30–210	1.6–10	11–15	6.4–21	2.4–23	7.2–26	16–43

Table 6.4: Ranges of percent uncertainties for the measurement of $d\sigma/dp_{\mathrm{T}}$ of the n^{th} jet in events with n or more jets. The uncertainties displayed include the statistical uncertainty propagated through the unfolding procedure (Statistical), the jet energy scale and resolution (JES/JER), the choice of generator used in the unfolding procedure (Generator), the uncertainty due to a finite number of simulated events used to construct the response (MC Stat.), and all other systematic uncertainties (Other) detailed in Section 6.6.

6.7 Comparisons with Theoretical Predictions

6.7.1 Comparisons with ME+PS Event Generators

The measured W+jets cross sections are compared to the predictions from several generators with an overall normalization determined by the NNLO inclusive cross sections calculated with FEWZ [106]. We consider W+jets signal processes generated with MadGraph 5.1.1 using the CTEQ6L1 PDF set, with Sherpa 1.4.0 using the CT10 [135,136] PDF set. The Sherpa sample is a separate sample from that used for the evaluation of uncertainties in Section 6.6. The MadGraph and Sherpa predictions provide leading-order (LO) matrix element (ME) calculations at each jet multiplicity, which are then combined into inclusive samples by matching the ME partons to particle jets. Parton showering (PS) and hadronization of the MadGraph sample is performed with Pythia 6.426 using the Z2 tune. The MadGraph+Pythia calculation includes the production of up to four partons. The jet matching is performed following the k_{T}-MLM prescription [40], where partons are clustered using the algorithm with a distance parameter of $D = 1$. The k_{T} clustering thresholds are chosen to be 10 GeV and 20 GeV at the matrix-element and parton-shower level, respectively. The factorization scale for each event is chosen to be the transverse mass computed after k_{T}-clustering of the event down to a $2\rightarrow2$ topology. The renormalization scale for the event is the k_{T} computed at each vertex splitting. The factorization scale for each event is chosen to be the transverse mass computed after k_{T}-clustering of the event down to a $2\rightarrow2$ topology. The renormalization scale for the event is the k_{T} computed at each vertex splitting. The predictions from Sherpa include the production of up to four partons. The matching between jets and partons is performed with the CKKW matching scheme [40], and the default factorization and renormalization scales are used.

The predictions from MadGraph+Pythia and Sherpa are shown with statistical uncertainties only. These MadGraph+Pythia and Sherpa samples are processed through the Rivet toolkit [137] in order to create particle level distributions, which can be compared with the unfolded data.

6.7.2 Comparisons with NLO Predictions from BlackHat+Sherpa

Results are also compared with predictions from BlackHat+Sherpa [38] using the CT10 PDF set. The BlackHat+Sherpa samples represent fixed-order predictions at the level of ME partons of W + n jets at NLO accuracy, for n = 1, 2, 3, 4, and 5 jets. Each measured distribution for a given inclusive jet multiplicity is compared with the corresponding fixed-order prediction from BlackHat+Sherpa. The choice of renormalization and factorization scales for BlackHat+Sherpa is $\hat{H}'_{\mathrm{T}}/2$, where $\hat{H}'_{\mathrm{T}} \equiv \sum_m p_{\mathrm{T}}^m + E_{\mathrm{T}}^{\mathrm{W}}$, m represents the final state partons, and $E_{\mathrm{T}}^{\mathrm{W}}$ is the transverse energy of the W boson.

The BlackHat+Sherpa samples were produced with the decay $W^{\pm} \to e^{\pm}\nu$, so we simply treat the electron as if it were a muon from $W^{\pm} \to \mu^{\pm}\nu$. Furthermore, the events contain matrix-element partons without showering/hadronization or underlying event; therefore, a non-perturbative correction must be applied before comparing with unfolded data. Following the recommendation of the BlackHat authors [138], the non-perturbative corrections were determined from a set of Madgraph(+Pythia) events with non-perturbative effects enabled and disabled, as listed in Table 6.5. The correction is then applied to each histogram on a bin-by-bin basis and it is typically less than 5%. The model-dependence of the non-perturbative is typically negligible [139], and so no uncertainty on the model-dependence is considered.

Pythia	Non-perturbative Effects		
parameter	Enabled	Disabled	
MSTP(81)	21	20	Disable multiple parton interaction, but do not change (p_T-ordered) parton showering.
MSTP(82)	4	0	Multiple parton model without hard interactions (note: this is actually not necessary as MPI was already disabled by MSTP(81)=20)
MSTJ(1)	1	0	Disable hadronization.

Table 6.5: Pythia settings for turning on/off non-perturbative effects.

The BlackHat+Sherpa prediction also includes uncertainties due to the PDF and variations of the factorization and renormalization scales. The nominal prediction is given by the central value of the CT10 PDF set, and the PDF uncertainty considers the envelope of the error sets of CT10, MSTW2008nlo68cl [140], and NNPDF2.1 [141] according to the PDF4LHC prescription [115, 142]. The factorization and renormalization scale uncertainty is determined by varying the scales simultaneously by a factor 0.5 or 2.0.

6.8 Results

The cross sections for exclusive and inclusive jet multiplicities are given in Fig. 6.3. In Figs. 6.4– 6.7 the differential cross sections are presented.

The unfolded exclusive and inclusive jet multiplicity distributions, shown in Fig. 6.3, are found to be in agreement, within uncertainties, with the predictions of the generators and with the NLO calculation of BlackHat+Sherpa. Table 6.6 details the measured cross sections as a function of the inclusive and exclusive jet multiplicity.

The jet p_{T} unfolded distributions for inclusive jet multiplicities from 1 to 4 are shown in Fig. 6.4. The predictions of BlackHat+Sherpa are in agreement with the mea-

sured distributions within the systematic uncertainties, while MadGraph+Pythia is observed to overestimate the yields up to 50% (45%) for the first (second) leading jet p_{T} distributions at high-p_{T} values. The predictions from Sherpa are found to agree well for the second-, third-, and fourth-leading jet p_{T} distributions, while an excess of slightly more than one standard deviation can be seen at high-p_{T} values for the leading jet p_{T} distribution. Similar observations hold for MadGraph+Pythia and Sherpa predictions in the H_{T} distributions for inclusive jet multiplicities of 1–4, as shown in Fig. 6.5. Since the BlackHat+Sherpa NLO prediction for $H_{\mathrm{T}}(\geq 1$ jet) is a fixed-order prediction with up to two real partons, contributions from higher jet multiplicities are missing, which results in an underestimation in the tail of the distribution [143]. Similar observations have been made with W+jets measurements at D0 [50] and ATLAS [56]. In general, Sherpa models the H_{T} distributions better than other generators.

The distributions of the jet η and of the difference in azimuthal angle between each jet and the muon are shown in Figs. 6.6 and 6.7, respectively. The measurements of the jet η agree with predictions from all generators, with MadGraph+Pythia and BlackHat+Sherpa performing best. The measurements of the $\Delta\phi$ between the leading jet and the muon are underestimated by as much as 38% by BlackHat+Sherpa, with similar, but smaller, underestimations in predictions from MadGraph+Pythia and Sherpa .

Examples of the variation in the BlackHat+Sherpa prediction due to the choice of PDF are given in Fig. 6.3–6.7, in which the predictions with the MSTW2008nlo68cl and NNPDF2.1 PDF sets are compared to the prediction from the CT10 PDF set, and to the measurements from data. The distributions determined with the different PDF sets are consistent with one another.

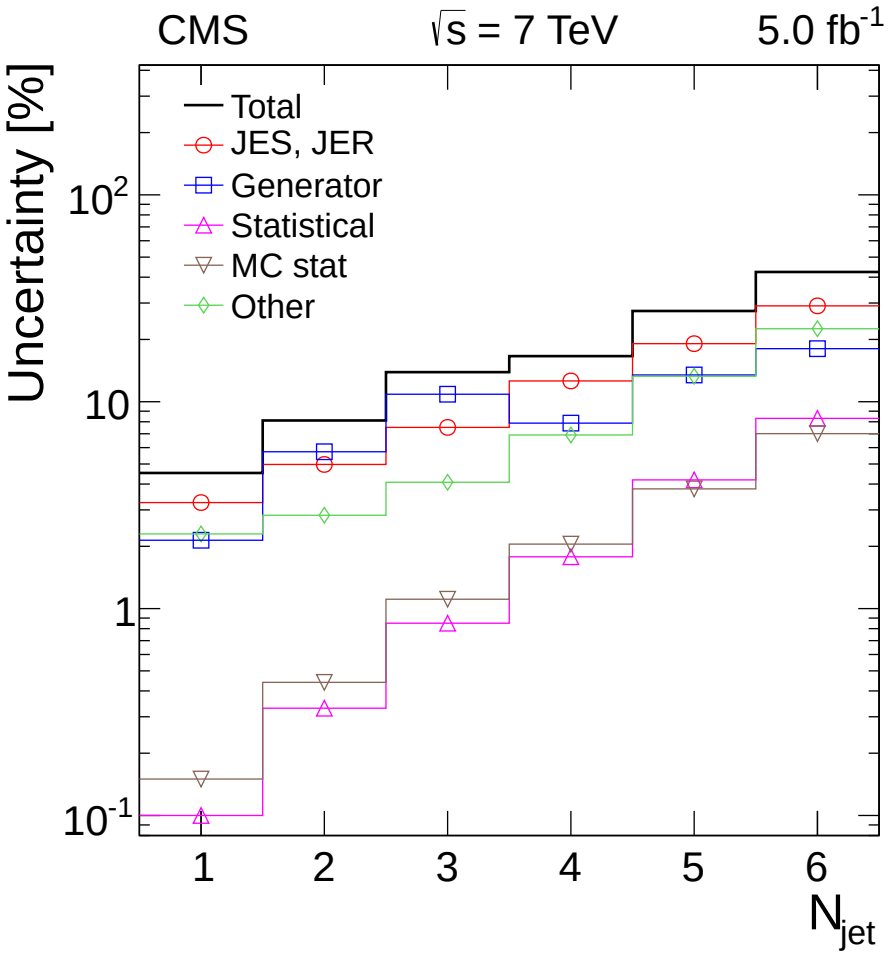

Figure 6.2: The dominant systematic uncertainties in the measurement of the W+jets cross section as a function of the exclusive jet multiplicity. The systematic uncertainties displayed include the jet energy scale and resolution (JES, JER), the choice of generator used in the unfolding procedure (Generator), the statistical uncertainty propagated through the unfolding procedure (Statistical), the uncertainty due to a finite number of simulated events used to construct the response (MC stat.), and all other systematic uncertainties (Other) detailed in Section 6.6. The uncertainties presented here correspond to the weighted average of the values shown in Table 6.4.

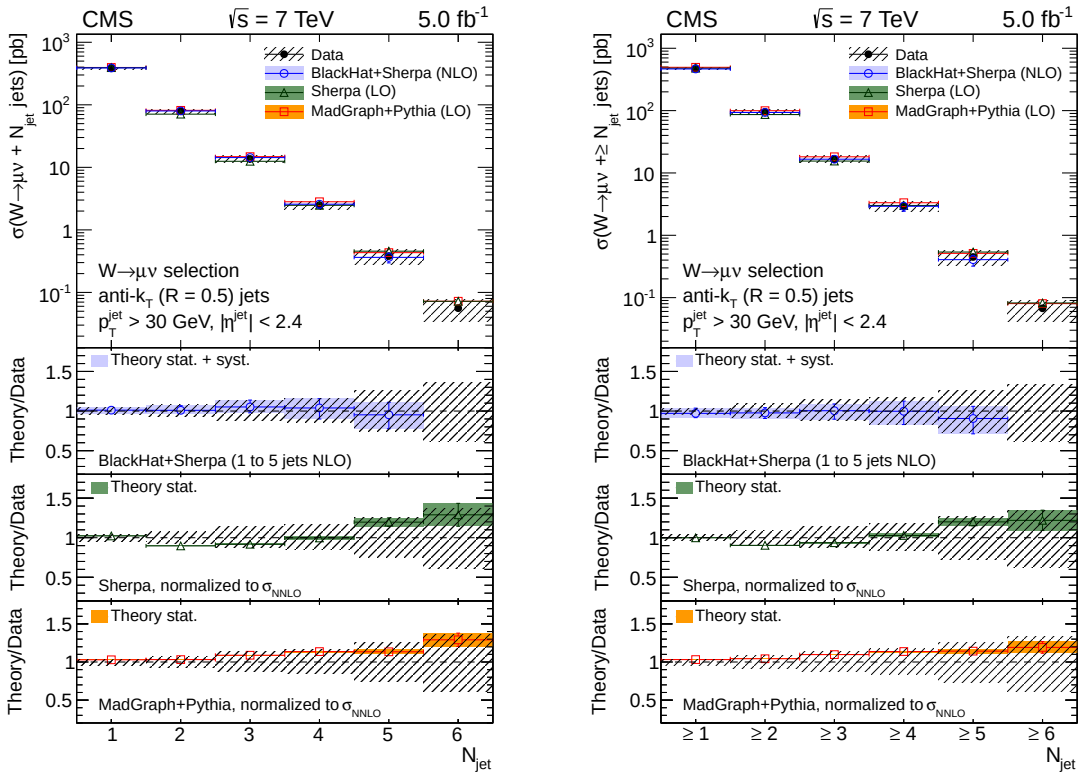

Figure 6.3: The cross section measurement for the exclusive and inclusive jet multiplicities, compared to the predictions of MadGraph 5.1.1 + Pythia 6.426, Sherpa 1.4.0, and BlackHat+Sherpa (corrected for hadronization and multiple-parton interactions). Black circular markers with the gray hatched band represent the unfolded data measurement and its uncertainty. Overlaid are the predictions together with their statistical uncertainties (Theory stat.). The BlackHat+Sherpa uncertainty also contains theoretical systematic uncertainties (Theory syst.) described in Section 6.8. The lower plots show the ratio of each prediction to the unfolded data.

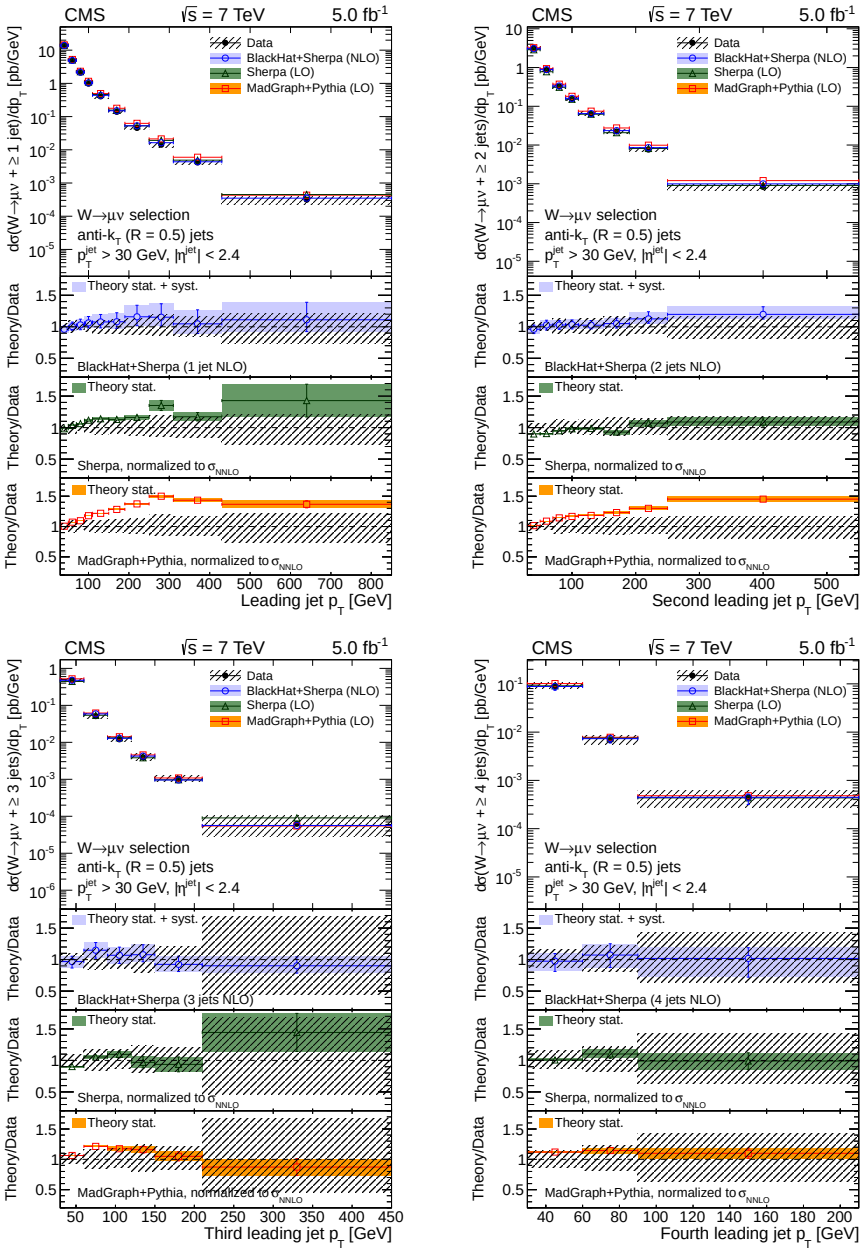

Figure 6.4: The differential cross section measurement for the leading four jets' transverse momenta, compared to the predictions of MadGraph 5.1.1 + Pythia 6.426, Sherpa 1.4.0, and BlackHat+Sherpa (corrected for hadronization and multiple-parton interactions). Black circular markers with the gray hatched band represent the unfolded data measurement and its uncertainty. Overlaid are the predictions together with their statistical uncertainties (Theory stat.). The BlackHat+Sherpa uncertainty also contains theoretical systematic uncertainties (Theory syst.) described in Section 6.8. The lower plots show the ratio of each prediction to the unfolded data.

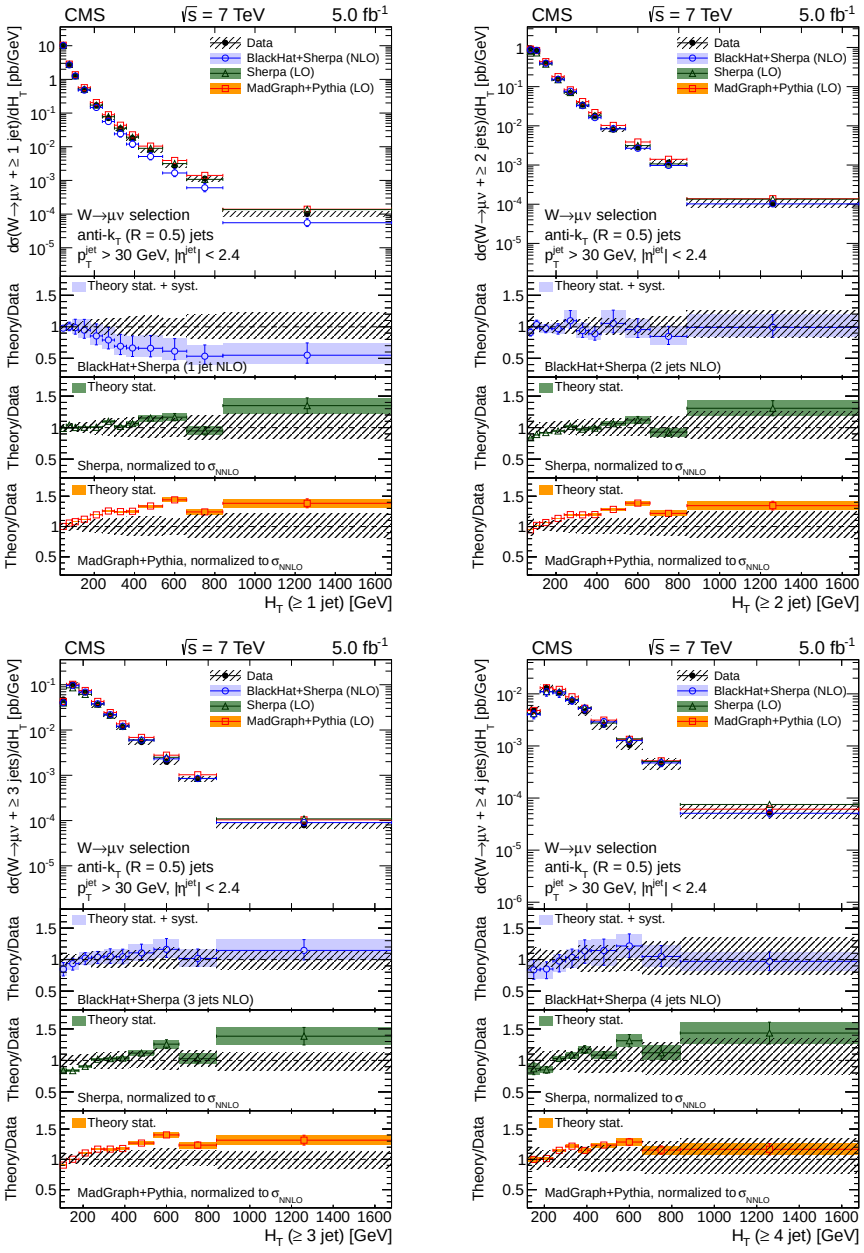

Figure 6.5: The differential cross section measurement for H_{T} for inclusive jet multiplicities 1–4, compared to the predictions of MadGraph 5.1.1 + Pythia 6.426, Sherpa 1.4.0, and BlackHat+Sherpa (corrected for hadronization and multiple-parton interactions). Black circular markers with the gray hatched band represent the unfolded data measurement and its uncertainty. Overlaid are the predictions together with their statistical uncertainties (Theory stat.). The BlackHat+Sherpa uncertainty also contains theoretical systematic uncertainties (Theory syst.) described in Section 6.8. The lower plots show the ratio of each prediction to the unfolded data.

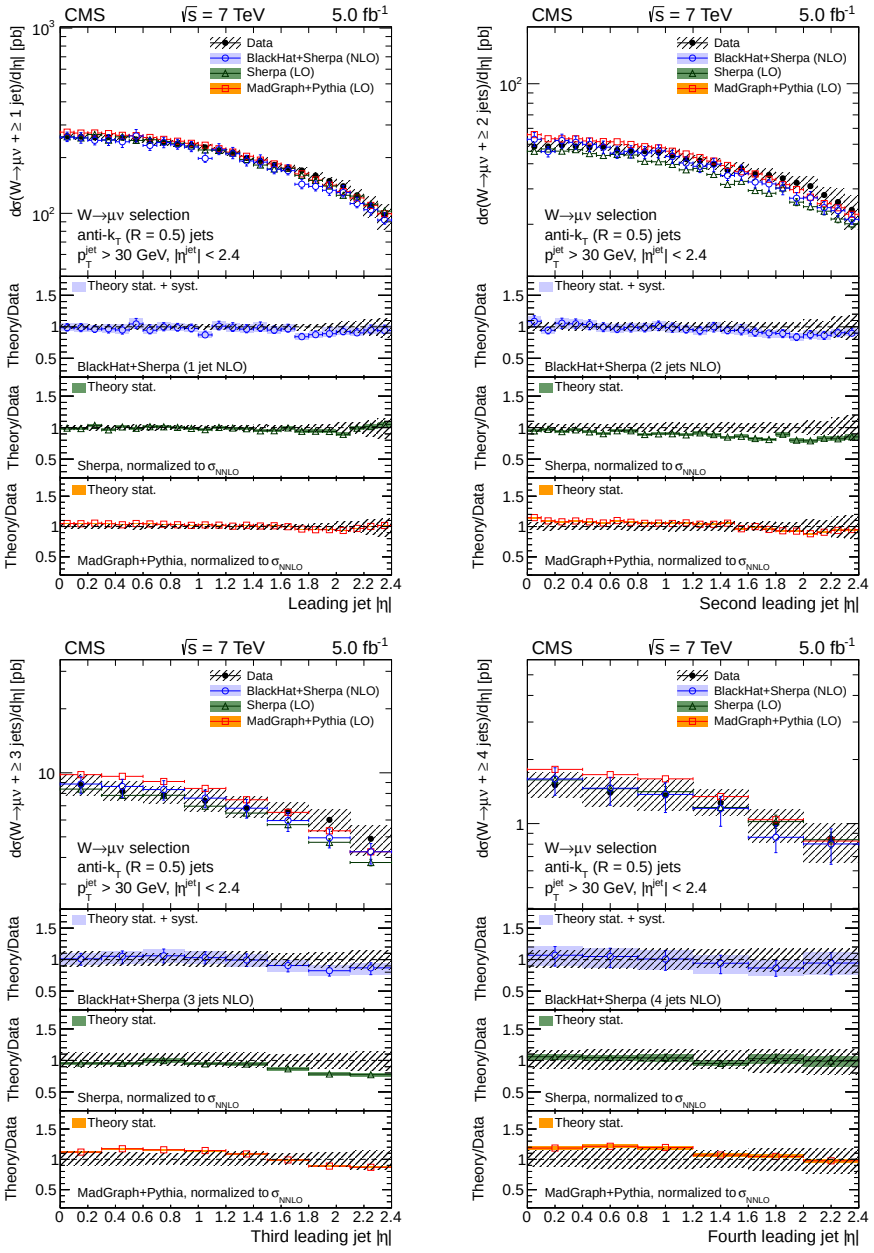

Figure 6.6: The differential cross section measurement for the pseudorapidity of the four leading jets, compared to the predictions of MadGraph 5.1.1 + Pythia 6.426, Sherpa 1.4.0, and BlackHat+Sherpa (corrected for hadronization and multiple-parton interactions). Black circular markers with the gray hatched band represent the unfolded data measurement and its uncertainty. Overlaid are the predictions together with their statistical uncertainties (Theory stat.). The BlackHat+Sherpa uncertainty also contains theoretical systematic uncertainties (Theory syst.) described in Section 6.8. The lower plots show the ratio of each prediction to the unfolded data.

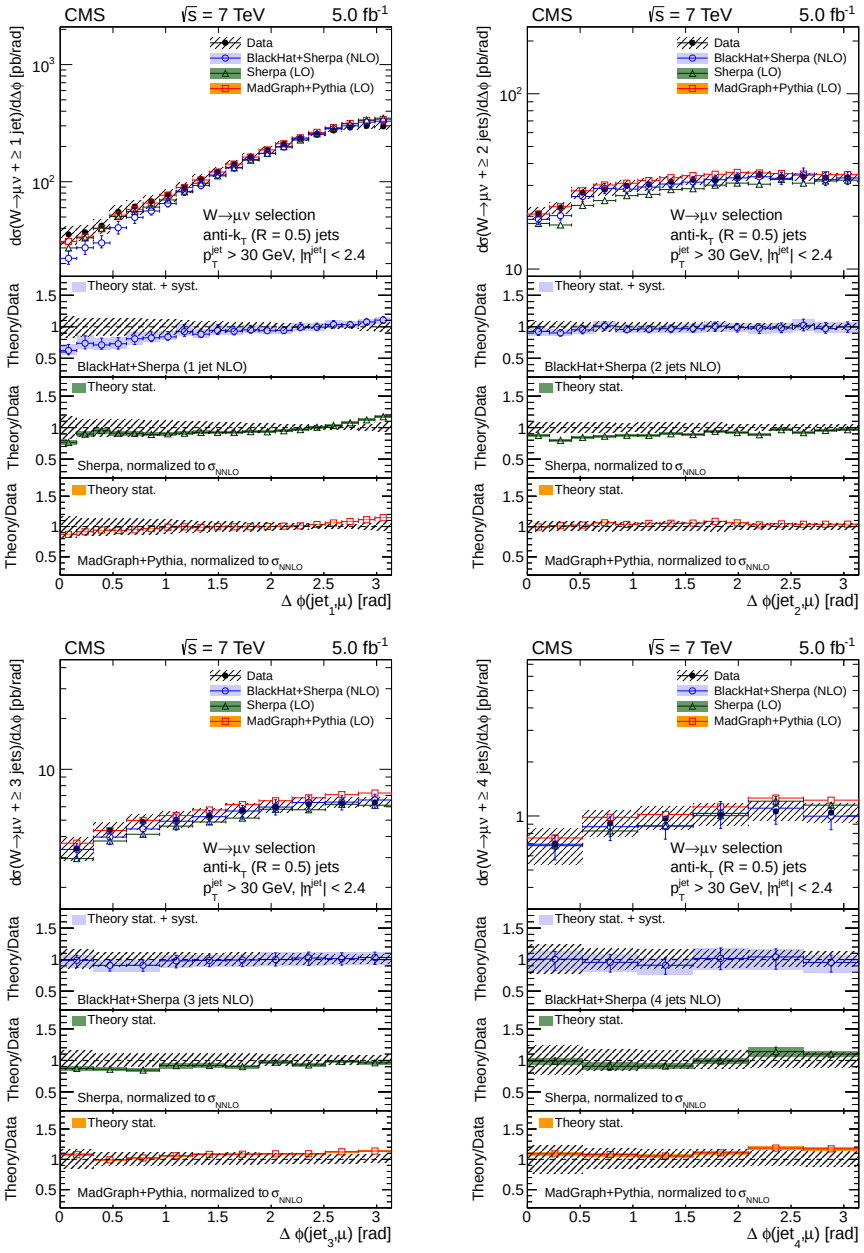

Figure 6.7: The differential cross section measurement in $\Delta\phi(\text{jet}_n, \mu)$, for $n = 1$ - 4, compared to the predictions of MadGraph 5.1.1 + Pythia 6.426, Sherpa 1.4.0, and BlackHat+Sherpa (corrected for hadronization and multiple-parton interactions). Black circular markers with the gray hatched band represent the unfolded data measurement and its uncertainty. Overlaid are the predictions together with their statistical uncertainties (Theory stat.). The BlackHat+Sherpa uncertainty also contains theoretical systematic uncertainties (Theory syst.) described in Section 6.8. The lower plots show the ratio of each prediction to the unfolded data.

Jet multiplicity	Exclusive σ [pb]	Inclusive σ [pb]
1	384^{+15}_{-17}	480^{+18}_{-20}
2	$79.1^{+6.2}_{-5.9}$	$95.6^{+8.5}_{-8.0}$
3	$13.6^{+1.9}_{-1.6}$	$16.6^{+2.3}_{-2.0}$
4	$2.48^{+0.40}_{-0.36}$	$2.93^{+0.52}_{-0.48}$
5	$0.382^{+0.097}_{-0.097}$	$0.45^{+0.12}_{-0.12}$
6	$0.056^{+0.020}_{-0.022}$	$0.067^{+0.023}_{-0.026}$

Table 6.6: Cross section measurements with statistical and systematic uncertainties for inclusive and exclusive jet multiplicities up to 6 jets.

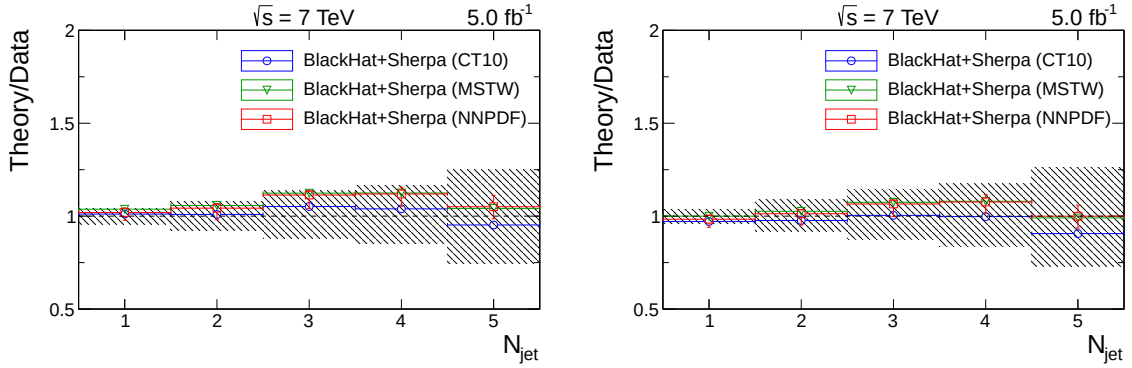

Figure 6.8: The ratio of the predictions of the jet multiplicity distributions from BlackHat+Sherpa, to the unfolded data measurement, using the CT10, MSTW2008nlo68cl and NNPDF PDF sets. The circular, triangular, and square markers indicate the predictions using the CT10, MSTW2008nlo68cl, and NNPDF PDF sets, respectively. The gray hatched band indicates the total uncertainty in the unfolded data measurement.

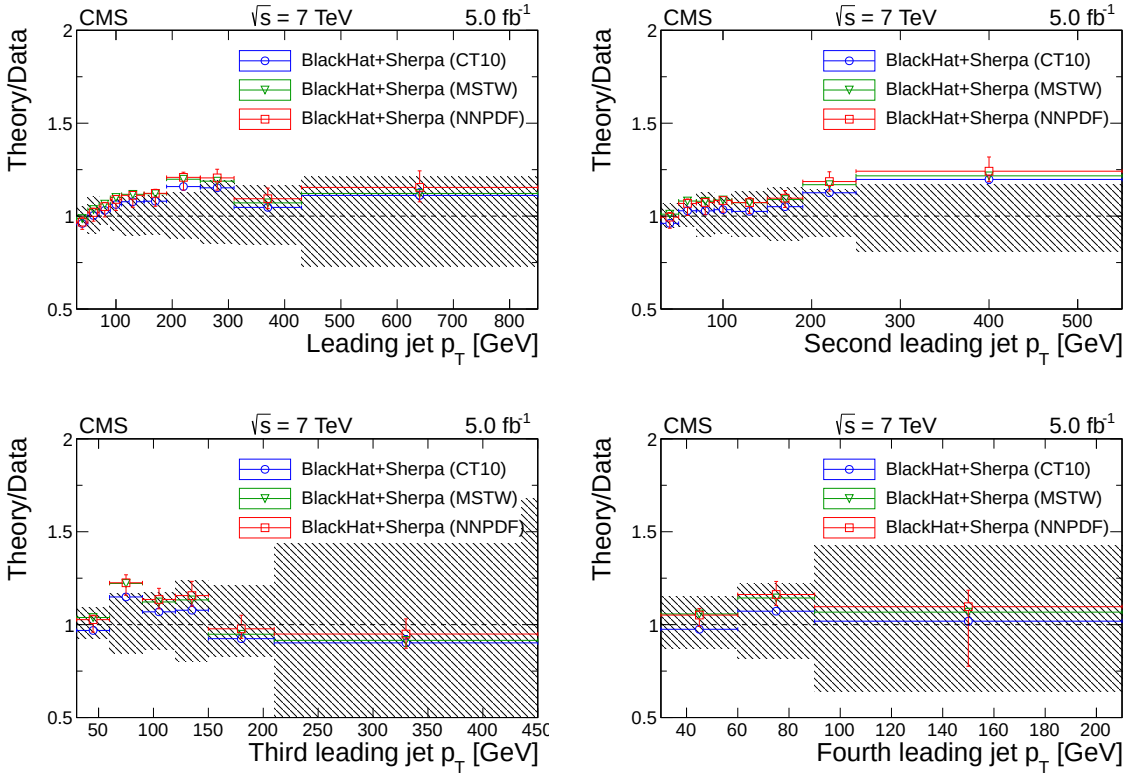

Figure 6.9: The ratio of the predictions of the p_{T} of the four leading jets from BlackHat+Sherpa, to the unfolded data measurement, using the CT10, MSTW2008nlo68cl and NNPDF PDF sets. The circular, triangular, and square markers indicate the predictions using the CT10, MSTW2008nlo68cl, and NNPDF PDF sets, respectively. The gray hatched band indicates the total uncertainty in the unfolded data measurement.

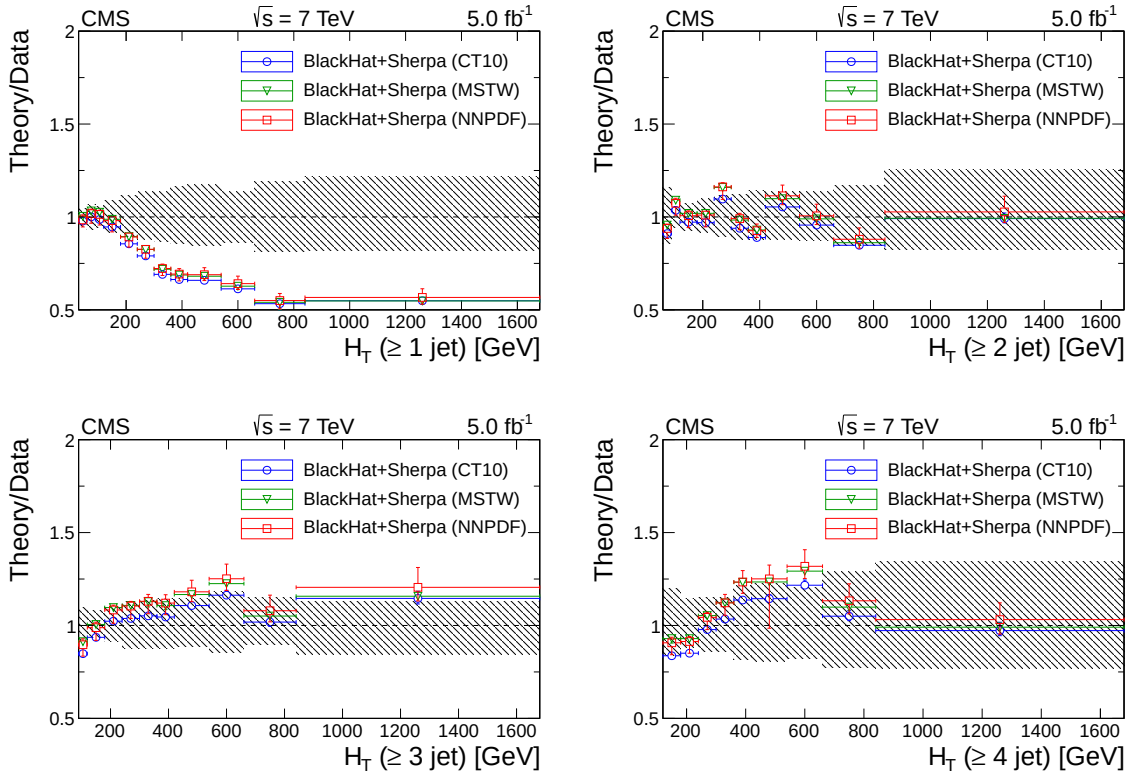

Figure 6.10: The ratio of the predictions of the H_{T} for jet multiplicities of 1–4 from BlackHat+Sherpa, to the unfolded data measurement, using the CT10, MSTW2008nlo68cl and NNPDF PDF sets. The circular, triangular, and square markers indicate the predictions using the CT10, MSTW2008nlo68cl, and NNPDF PDF sets, respectively. The gray hatched band indicates the total uncertainty in the unfolded data measurement.

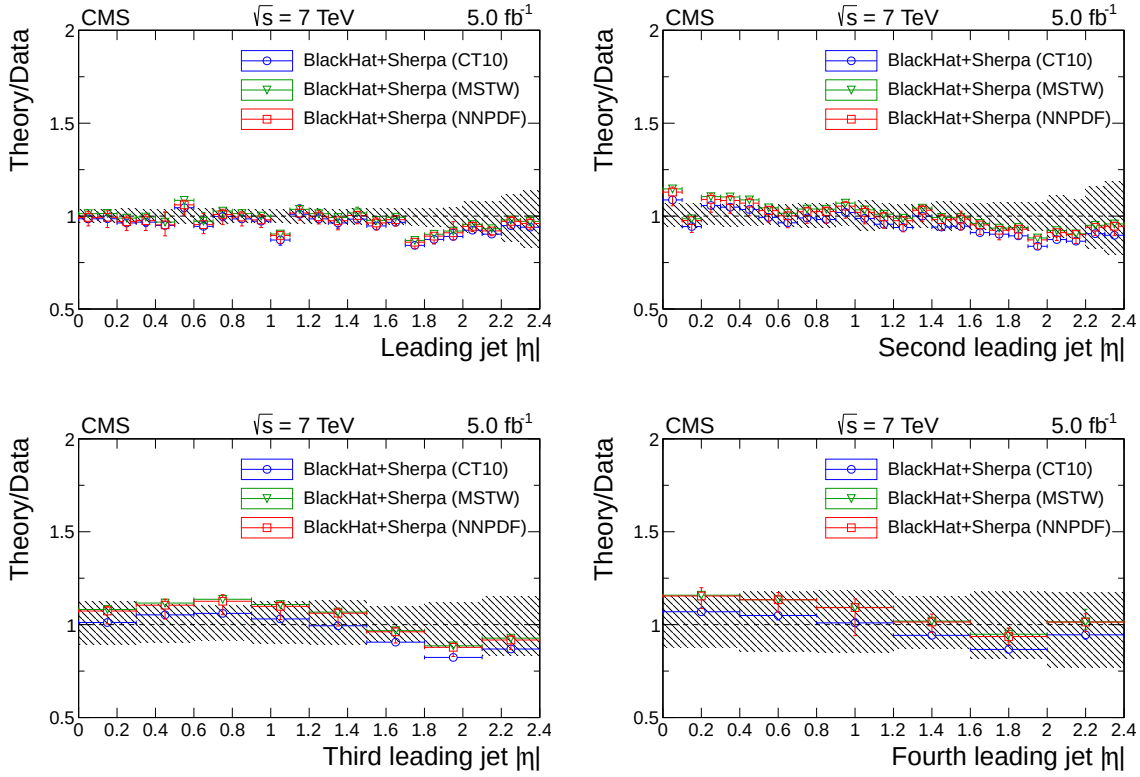

Figure 6.11: The ratio of the predictions of the η of the four leading jets from BlackHat+Sherpa to the unfolded data measurement, using the CT10, MSTW2008nlo68cl and NNPDF PDF sets. The circular, triangular, and square markers indicate the predictions using the CT10, MSTW2008nlo68cl, and NNPDF PDF sets, respectively. The gray hatched band indicates the total uncertainty in the unfolded data measurement.

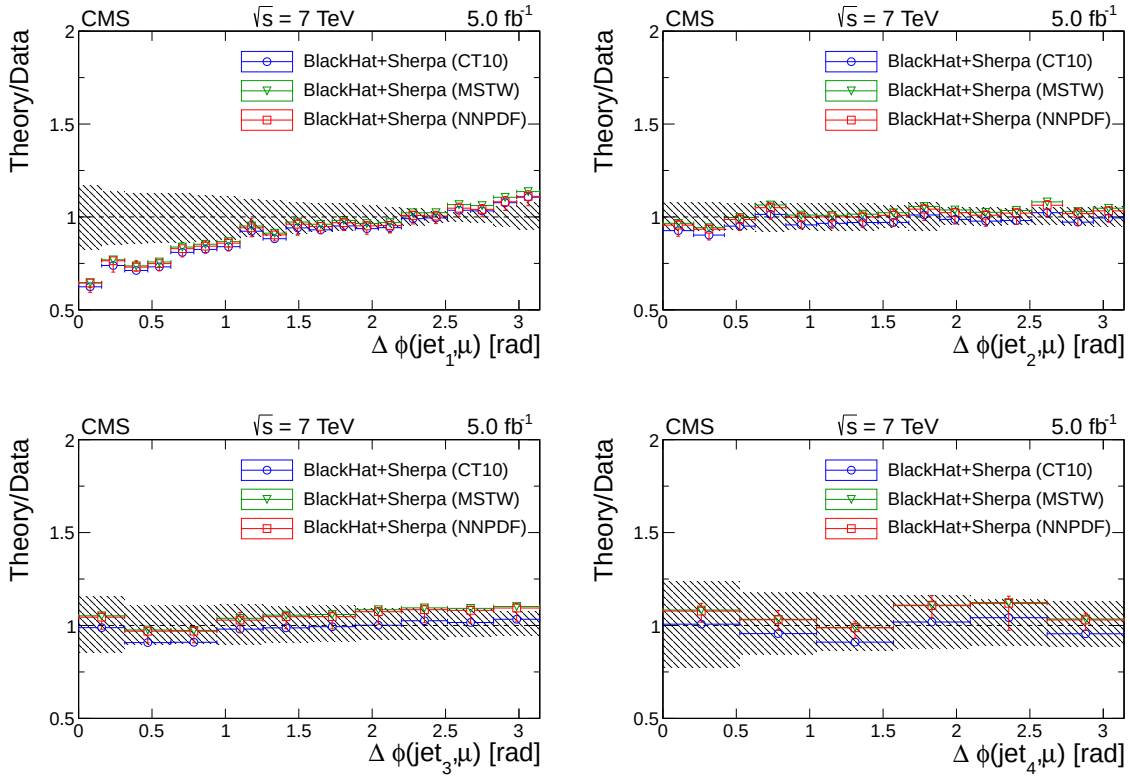

Figure 6.12: The ratio of the predictions of the $\Delta\phi(\text{jet}, \mu)$ of the four leading jets from BlackHat+Sherpa to the unfolded data measurement, using the CT10, MSTW2008nlo68cl and NNPDF PDF sets. The circular, triangular, and square markers indicate the predictions using the CT10, MSTW2008nlo68cl, and NNPDF PDF sets, respectively. The gray hatched band indicates the total uncertainty in the unfolded data measurement.

Chapter 7

Conclusion

This thesis has presented searches for the pair production of second generation scalar leptoquarks in the $\mu\mu jj$ and $\mu\nu jj$ final states, using data collected with the CMS detector at the LHC. Multiple searches were performed, using 5 fb^{-1}of pp collision data collected at $\sqrt{s}$ = 7 TeV, and 20 fb^{-1}of pp collision data collected at $\sqrt{s}$ = 8 TeV. Since no evidence for the existence of second generation leptoquarks has been found, upper limits on the mass and cross section have been determined at the 95% confidence level. In the searches conducted at 8 TeV, second-generation scalar leptoquarks with masses less than 1070 (785) GeV can be excluded for $\beta = 1.0$ (0.5). These limits are the most stringent to date, and are several times larger than limits set by previous collider experiments.

Motivated by the large background contribution to the leptoquark searches from the W+jets process, measurements of this process have also been presented, using CMS data corresponding to 5 fb^{-1} of pp collisions at $\sqrt{s} = 7$ TeV. The measured cross sections, corrected for detector effects, include the jet multiplicity, the individual jet transverse momenta and pseudorapidities, the angular separation between the jets and the muon, and the scalar sum of the transverse momenta of all jets. The cross sections

have been compared with predictions from event generators, MadGraph+Pythia and NLO predictions from BlackHat+Sherpa. While predictions have been found to describe the jet multiplicity within the uncertainties, some key differences with respect to the predictions have been observed. These differences include some overestimation of the leading and subleading jet p_{T} at high p_{T} values, and in the distribution of H_{T} at high H_{T} values, by MadGraph+Pythia and Sherpa . The predictions from BlackHat+Sherpa underestimate the measurement of the cross section as a function of H_{T} for $N_{\mathrm{jet}} \geq 1$, since the contribution from W+$\geq$3 jets is missing from an NLO prediction of W+$\geq$1 jet. The distributions of $\Delta\phi$ between the leading jet and the muon are underestimated by all predictions for $\Delta\phi$ values near zero, with the largest disagreement visible in BlackHat+Sherpa.

Bibliography

[1] M. Banner et al., "Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider", *Physics Letters B* **122** (1983), no. 56, 476 – 485, doi: http://dx.doi.org/10.1016/0370-2693(83)91605-2.

[2] G. Arnison et al., "Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540$ GeV", *Physics Letters B* **122** (1983), no. 1, 103 – 116, doi: http://dx.doi.org/10.1016/0370-2693(83)91177-2.

[3] G. Arnison et al., "Experimental observation of lepton pairs of invariant mass around 95 GeV/c^2 at the CERN SPS collider", *Physics Letters B* **126** (1983), no. 5, 398 – 410, doi: http://dx.doi.org/10.1016/0370-2693(83)90188-0.

[4] P. Bagnaia et al., "Evidence for $Z^0 \to e^+e^-$ at the CERN pp collider", *Physics Letters B* **129** (1983), no. 12, 130 – 140, doi: http://dx.doi.org/10.1016/0370-2693(83)90744-X.

[5] CDF Collaboration, "Observation of top quark production in $p\overline{p}$ collisions with the Collider Detector at Fermilab", *Phys. Rev. Lett.* **74** (Apr, 1995) 2626–2631, doi: 10.1103/PhysRevLett.74.2626.

[6] D0 Collaboration, "Search for high mass top quark production in $p\overline{p}$ collisions at $\sqrt{s} = 1.8$ TeV", *Phys. Rev. Lett.* **74** (Mar, 1995) 2422–2426, doi: 10.1103/PhysRevLett.74.2422.

[7] K. Kodama et al., "Observation of tau neutrino interactions", *Physics Letters B* **504** (2001), no. 3, 218 – 224, ihttp://dx.doi.org/10.1016/S0370-2693(01)00307-0.

[8] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *J. High Energy Phys.* **06** (2013) 081, doi: 10.1007/JHEP06(2013)081.

[9] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC",

Phys.Lett. **B716** (2012) 1–29, doi: 10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[10] F. Halzen and A. D. Martin, "Quarks and leptons: An introductory course in modern particle physics". Wiley, 1984.

[11] D. Griffiths, "Introduction to elementary particles". Wiley, 2008.

[12] M. Srednicki, "Quantum field theory". Cambridge University Press, 2007.

[13] Particle Data Group Collaboration, "Review of particle physics (RPP)", *Phys.Rev.* **D86** (2012) 010001, doi: 10.1103/PhysRevD.86.010001.

[14] H. Georgi and S. L. Glashow, "Unity of all elementary-particle forces", *Phys. Rev. Lett.* **32** (Feb, 1974) 438–441, doi: 10.1103/PhysRevLett.32.438.

[15] J. C. Pati and A. Salam, "Lepton number as the fourth color", *Phys. Rev.* **D10** (1974) 275–289, doi: 10.1103/PhysRevD.10.275.

[16] B. Schrempp and F. Schrempp, "Light leptoquarks", *Phys. Lett.* **B153** (1985) 101, doi: 10.1016/0370-2693(85)91450-9.

[17] S. Dimopoulos and L. Susskind, "Mass without scalars", *Nucl. Phys.* **B155** (1979) 237–252, doi: 10.1016/0550-3213(79)90364-X.

[18] S. Dimopoulos, "Technicolored signatures", *Nucl. Phys.* **B168** (1980) 69–92, doi: 10.1016/0550-3213(80)90277-1.

[19] E. Farhi and L. Susskind, "Technicolor", *Phys. Rept.* **74** (1981) 277, doi: 10.1016/0370-1573(81)90173-3.

[20] J. L. Hewett and T. G. Rizzo, "Low-energy phenomenology of superstring inspired E(6) models", *Phys. Rept.* **183** (1989) 193, doi: 10.1016/0370-1573(89)90071-9.

[21] W. Buchmuller and D. Wyler, "Constraints on SU(5) type leptoquarks", *Phys. Lett.* **B177** (1986) 377, doi: 10.1016/0370-2693(86)90771-9.

[22] W. Buchmuller, R. Ruckl, and D. Wyler, "Leptoquarks in lepton quark collisions", *Phys. Lett.* **B191** (1987) 442–448, doi: 10.1016/0370-2693(87)90637-X.

[23] M. S. J. T. A. Djouadi, T. Kohler, "(eb), (et) type letoquarks at ep colliders", volume 46. Z. Phys. C, 1990.

[24] O. Shanker, "$\pi\ell 2$, K$\ell 3$, and K^0-$\overline{K}^0$ constraints on leptoquarks and supersymmetric particles", *Nucl. Phys. B* **204** (1982) 375, doi: 10.1016/0550-3213(82)90196-1.

[25] J. L. Hewett and T. G. Rizzo, "Much ado about leptoquarks: A comprehensive analysis", *Phys. Rev.* **D56** (1997) 5709–5724, arXiv:hep-ph/9703337.

[26] Krämer, M. and Plehn, T. and Spira, M. and Zerwas, P. M., "Pair production of scalar leptoquarks at the CERN LHC", *Phys. Rev. D* **71** (2005) 057503, arXiv:hep-ph/0411038.

[27] H1 Collaboration, "A search for leptoquarks, leptogluons and excited leptons in H1 at HERA", *Nuclear Physics B* **396** (1993), no. 1, 3 – 23, doi: 10.1016/0550-3213(93)90255-N.

[28] H1 Collaboration, "A search for leptoquarks at HERA", *Physics Letters B* **369** (1996), no. 2, 173 – 185, doi: 10.1016/0370-2693(95)01601-5.

[29] ZEUS Collaboration, "Search for leptoquarks with the ZEUS detector", *Physics Letters B* **306** (1993), no. 12, 173 – 186, doi: 10.1016/0370-2693(93)91155-G.

[30] H1 Collaboration, "Observation of events at very high Q^2 in ep collisions at HERA", *Z.Phys.* **C74** (1997) 191–206, doi: 10.1007/s002880050383, arXiv:hep-ex/9702012.

[31] ZEUS Collaboration, "Comparison of ZEUS data with standard model predictions for $e^+p \to e^+X$ scattering at high x and Q^2", *Z.Phys.* **C74** (1997) 207–220, doi: 10.1007/s002880050384, arXiv:hep-ex/9702015.

[32] CDF Collaboration, D0 Collaboration, "Combined limits on first generation leptoquarks from the CDF and D0 experiments", arXiv:hep-ex/9810015.

[33] D0 Collaboration, "Search for pair production of first-generation leptoquarks in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV", *Phys. Lett.* **B681** (2009) 224, arXiv:0907.1048.

[34] D0 Collaboration, "Search for pair production of second generation scalar leptoquarks", *Phys. Lett.* **B671** (2009) 224, arXiv:0808.4023.

[35] CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section in pp collisions at $\sqrt{s} = 7$ TeV with lepton + jets final states", *Phys. Lett. B* **720** (2012) 83, doi: 10.1016/j.physletb.2013.02.021, arXiv:1212.6682.

[36] CMS Collaboration, "Measurement of the single-top-quark t-channel cross section in pp collisions at $\sqrt{s} = 7$TeV", *JHEP* **12** (2012) 035, doi: 10.1007/JHEP12(2012)035, arXiv:1209.4533.

[37] J. M. Campbell and R. K. Ellis, "Next-to-leading order corrections to W^+ 2 jet and Z^+ 2 jet production at hadron colliders", *Phys.Rev.* **D65** (2002) 113007, doi: 10.1103/PhysRevD.65.113007, arXiv:hep-ph/0202176.

[38] C. F. Berger et al., "Next-to-leading order QCD predictions for W+3-jet distributions at hadron colliders", *Phys. Rev. D* **80** (2009) 074036, doi: 10.1103/PhysRevD.80.074036, arXiv:0907.1984.

[39] Z. Bern et al., "Ntuples for NLO events at hadron colliders", *Comput. Phys. Commun.* **185** (2014) 1443, doi: 10.1016/j.cpc.2014.01.011, arXiv:1310.7439.

[40] J. Alwall et al., "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions", *Eur. Phys. J C* **53** (2008), no. 3, 473–500, doi: 10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.

[41] J. Alwall et al., "MadGraph 5 : Going beyond", *JHEP* **06** (2011) 128, doi: 10.1007/JHEP06(2011)128, arXiv:1106.0522.

[42] T. Sjöstrand et al., "High energy physics event generation with Pythia 6.1", *Comput. Phys. Commun.* **135** (2001) 238, doi: 10.1016/S0010-4655(00)00236-8, arXiv:hep-ph/0010017.

[43] T. Gleisberg and S. Höeche, "Comix, a new matrix element generator", *JHEP* **0812** (2008) 039, doi: 10.1088/1126-6708/2008/12/039, arXiv:0808.3674.

[44] S. Schumann and F. Krauss, "A parton shower algorithm based on Catani-Seymour dipole factorisation", *JHEP* **0803** (2008) 038, doi: 10.1088/1126-6708/2008/03/038, arXiv:0709.1027.

[45] T. Gleisberg et al., "Event generation with Sherpa 1.1", *JHEP* **0902** (2009) 007, doi: 10.1088/1126-6708/2009/02/007, arXiv:0811.4622.

[46] S. Höeche et al., "QCD matrix elements and truncated showers", *JHEP* **0905** (2009) 053, doi: 10.1088/1126-6708/2009/05/053, arXiv:0903.1219.

[47] B. Andersson et al., "Parton fragmentation and string dynamics", *Phys.Rept.* **97** (1983) 31–145, doi: 10.1016/0370-1573(83)90080-7.

[48] J.-C. Winter, F. Krauss, and G. Soff, "A modified cluster hadronization model", *Eur.Phys.J.* **C36** (2004) 381–395, doi: 10.1140/epjc/s2004-01960-8, arXiv:hep-ph/0311085.

[49] D0 Collaboration, "Measurements of inclusive W+jets production rates as a function of jet transverse momentum in $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV", *Phys. Lett. B* **705** (2011) 200, doi: 10.1016/j.physletb.2011.10.011, arXiv:1106.1457.

[50] D0 Collaboration, "Studies of W boson plus jets production in $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV", *Phys. Rev. D* **88** (2013) 092001, doi: 10.1103/PhysRevD.88.092001, arXiv:1302.6508.

[51] R. K. Ellis et al., "One-loop amplitudes for W^+ 3 jet production in hadron collisions", *JHEP* **0901** (2009) 012, doi: 10.1088/1126-6708/2009/01/012, arXiv:0810.2762.

[52] W. Giele and G. Zanderighi, "On the numerical evaluation of one-loop amplitudes: The gluonic case", *JHEP* **0806** (2008) 038, doi: 10.1088/1126-6708/2008/06/038, arXiv:0805.2152.

[53] CDF Collaboration, "Measurement of the cross section for W-boson production in association with jets in $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV", *Phys. Rev. D* **77** (2008) 011108, doi: 10.1103/PhysRevD.77.011108, arXiv:0711.4044.

[54] M. L. Mangano et al., "Matching matrix elements and shower evolution for top-quark production in hadronic collisions", *JHEP* **0701** (2007) 013, doi: 10.1088/1126-6708/2007/01/013, arXiv:hep-ph/0611129.

[55] S. Mrenna and P. Richardson, "Matching matrix elements and parton showers with HERWIG and PYTHIA", *JHEP* **0405** (2004) 040, doi: 10.1088/1126-6708/2004/05/040, arXiv:hep-ph/0312274.

[56] ATLAS Collaboration, "Study of jets produced in association with a W boson in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector", *Phys. Rev. D* **85** (2012) 092002, doi: 10.1103/PhysRevD.85.092002, arXiv:1201.1276.

[57] CMS Collaboration, "Jet production rates in association with W and Z bosons in pp collisions at $\sqrt{s} = 7$ TeV", *JHEP* **01** (2012) 010, doi: 10.1007/JHEP01(2012)010, arXiv:1110.3226.

[58] G. Bachy et al., "The LEP collider: construction, project status and outlook", *Part. Accel.* **26** (Aug, 1989) 19–32. 15 p.

[59] J.-L. Caron, "The LHC injection complex", (May, 1993). AC Collection. Legacy of AC. Pictures from 1992 to 2002.

[60] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *Journal of Instrumentation* **3** (2008), no. 08, S08003.

[61] CMS Collaboration, "The CMS experiment at the CERN LHC", *Journal of Instrumentation* **3** (2008), no. 08, S08004.

[62] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *Journal of Instrumentation* **3** (2008), no. 08, S08002.

[63] LHCb Collaboration, "The LHCb Detector at the LHC", *Journal of Instrumentation* **3** (2008), no. 08, S08005.

[64] L. Evans and P. Bryant, "LHC machine", *Journal of Instrumentation* **3** (2008), no. 08, S08001.

[65] CMS Collaboration, "Detector Drawings", (Mar, 2012). CMS Collection.

[66] T. CMS, "The performance of the CMS muon detector in proton-proton collisions at s = 7 TeV at the LHC", *Journal of Instrumentation* **8** (2013), no. 11, P11002.

[67] C. Collaboration, "ECAL Technical Design Report (TDR) Figures from Chapter 1", (Dec, 1997). CMS Collection.

[68] CMS Collaboration, "The CMS muon project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997.

[69] CMS Collaboration, "Tracking and Vertexing Results from First Collisions", CMS Physics Analysis Summary CMS-PAS-TRK-10-001, (2010).

[70] W. Adam et al., "Track reconstruction in the CMS tracker", Technical Report CMS-NOTE-2006-041, CERN, Geneva, (Dec, 2006).

[71] T. Miao et al., "Beam position determination using tracks", Technical Report CMS-NOTE-2007-021, CERN, Geneva, (Aug, 2007).

[72] R. Frhwirth, W. Waltenberger, and P. Vanlaer, "Adaptive vertex fitting", Technical Report CMS-NOTE-2007-008, CERN, Geneva, (Mar, 2007).

[73] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", `arXiv:1405.6569`.

[74] CMS Collaboration, "The performance of the CMS muon detector in proton–proton collisions at $\sqrt{s} = 7$ TeV at the LHC", *J. Instrum.* **8** (2013) P11002, doi: 10.1088/1748-0221/8/11/P11002.

[75] CMS Collaboration, "CMS physics: Technical design report volume 1: Detector performance and software". Technical Design Report CMS. CERN, Geneva, 2006.

[76] CMS Collaboration, "Electron reconstruction and identification at $\sqrt{s} = 7$ TeV", CMS Physics Analysis Summary CMS-PAS-EGM-10-004, (2010).

[77] CMS Collaboration, "Particle-flow event reconstruction in CMS and performance for jets, taus, and $E_{\mathrm{T}}^{\mathrm{miss}}$", CMS Physics Analysis Summary CMS-PAS-PFT-09-001, (2009).

[78] CMS Collaboration, "Jet energy resolution in CMS at $\sqrt{s} = 7$ TeV", CMS Physics Analysis Summary CMS-PAS-JME-10-014, (2010).

[79] 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).

[80] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-kt jet clustering algorithm", *JHEP* **04** (2008) 063, arXiv:0802.1189.

[81] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *JINST* **6** (2011) 11002, doi: 10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.

[82] CMS Collaboration, "Jet energy scale performance in 2011",.

[83] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, doi: 10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.

[84] T. Sjostrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual", *JHEP* **0605** (2006) 026, doi: 10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

[85] P. Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **11** (2004) 040, doi: 10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.

[86] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, doi: 10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[87] S. Alioli et al., "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043, doi: 10.1007/JHEP06(2010)043, arXiv:1002.2581.

[88] S. Alioli et al., "NLO single-top production matched with shower in POWHEG: s- and t-channel contributions", *JHEP* **09** (2009) 111, doi: 10.1007/JHEP02(2010)011, 10.1088/1126-6708/2009/09/111, arXiv:0907.4076.

[89] J. Pumplin et al., "New generation of parton distributions with uncertainties from global QCD analysis", *JHEP* **07** (2002) 012, doi: 10.1088/1126-6708/2002/07/012, arXiv:hep-ph/0201195.

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

[91] CMS Collaboration, "Search for pair production of second-generation scalar leptoquarks in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Rev. Lett.* **106** (2010) 201803, doi: 10.1103/PhysRevLett.106.201803.

[92] R. Field, "Physics at the Tevatron", *Acta Phys. Polon.* **B39** (2008) 2611.

[93] R. Field, "Studying the underlying event at CDF and the LHC", in *Proceedings of the First International Workshop on Multiple Partonic Interactions at the LHC MPI'08, October 27-31, 2008*, P. Bartalini and L. Fanó, eds. Perugia, Italy, October, 2009. arXiv:1003.4220.

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

[95] J. Alwall et al., "MadGraph/MadEvent v4: The new web generation", *JHEP* **09** (2007) 028, doi: 10.1088/1126-6708/2007/09/028.

[96] M. L. Mangano et al., "ALPGEN, a generator for hard multiparton processes in hadronic collisions", *JHEP* **07** (2003) 001, arXiv:hep-ph/0206293.

[97] Particle Data Group Collaboration, "Review of particle physics", *Phys. Lett.* **B667** (2008) 1–1340, doi: 10.1016/j.physletb.2008.07.018.

[98] I. Bertram et al., "A Recipe for the construction of confidence limits", technical report, (2000). FERMILAB-TM-2104.

[99] CMS Collaboration, "First measurement of the cross section for top-quark pair production in proton-proton collisions at $\sqrt{s} = 7$ TeV", (2010). arXiv:1010.5994.

[100] CMS Collaboration, "Determination of the jet energy scale in CMS with pp collisions at $\sqrt{s} = 7$ TeV", CMS Physics Analysis Summary CMS-PAS-JME-10-010, (2010).

[101] CMS Collaboration, "Measurement of CMS Luminosity", CMS Physics Analysis Summary CMS-PAS-EWK-10-004, (2010).

[102] CMS Collaboration, "Measurements of inclusive W and Z cross sections in pp collisions at $\sqrt{s} = 7$ TeV", *J. High Energy Phys.* **01** (2010) 080, doi: 10.1007/JHEP01(2011)080.

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

[104] CMS Collaboration, "Search for pair production of first- and second-generation scalar leptoquarks in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Rev. D* **86** (2012) 052013, doi: 10.1103/PhysRevD.86.052013.

[105] CMS Collaboration, "Search for pair production of second-generation scalar leptoquarks in pp collisions at $\sqrt{s} = 8$ TeV with the CMS Detector", CMS Physics Analysis Summary CMS-PAS-EXO-12-042, (2012).

[106] K. Melnikov and F. Petriello, "Electroweak gauge boson production at hadron colliders through $O(\alpha_s^2)$", *Phys. Rev. D* **74** (2006) 114017, doi: 10.1103/PhysRevD.74.114017, arXiv:hep-ph/0609070.

[107] N. Kidonakis, "Higher-order corrections to top-antitop pair and single top quark production", *Proceedings of the DPF-2009 Conference* (2009) arXiv:0909.0037.

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

[109] J. Campbell, R. Ellis, and F. Tramontano, "Single top production and decay at next-to-leading order", *Phys. Rev. D* **70** (2004) 094012, doi: 10.1103/PhysRevD.70.094012, arXiv:hep-ph/0408158.

[110] J. Campbell and F. Tramontano, "Next-to-leading order corrections to Wt production and decay", *Nucl. Phys. B* **726** (2005) 109, doi: 10.1016/j.nuclphysb.2005.08.015, arXiv:hep-ph/0506289.

[111] J. M. Campbell et al., "Next-to-leading-order predictions for t-Channel single-top production at hadron colliders", *Phys. Rev. Lett.* **102** (2009) 182003, doi: 10.1103/PhysRevLett.102.182003, arXiv:0903.0005.

[112] J. M. Campbell, R. K. Ellis, and C. Williams, "Vector boson pair production at the LHC", *JHEP* **07** (2011) 018, doi: 10.1007/JHEP07(2011)018, arXiv:1105.0020.

[113] CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$TeV", *J. Instrum.* **7** (2012) P10002, doi: 10.1088/1748-0221/7/10/P10002.

[114] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *J. Instrum.* **6** (2011) P11002, doi: 10.1088/1748-0221/6/11/P11002.

[115] M. Botje et al., "The PDF4LHC working group interim recommendations", (2011). arXiv:1101.0538.

[116] A. D. Martin et al., "Parton distributions for the LHC", *Eur. Phys. J. C* **63** (2009) 189, doi: 10.1140/epjc/s10052-009-1072-5, arXiv:0901.0002.

[117] R. D. Ball et al., "A first unbiased global NLO determination of parton distributions and their uncertainties", *Nucl. Phys. B* **838** (2010) 136, doi: 10.1016/j.nuclphysb.2010.05.008, arXiv:1002.4407.

[118] CMS, "Absolute calibration of the luminosity measurement at CMS: Winter 2012 update", CMS Physics Analysis Summary CMS-PAS-SMP-12-008, (2012).

[119] CMS Collaboration, "CMS luminosity based on pixel cluster counting - Summer 2013 update", CMS Physics Analysis Summary CMS-PAS-LUM-13-001, (2013).

[120] CMS Collaboration, "Search for narrow resonances in dilepton mass spectra in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Lett. B* (2012) doi: 10.1016/j.physletb.2012.06.051, arXiv:1206.1849.

[121] CMS Collaboration, "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 7$TeV", *Phys. Lett. B* **722** (2013) 5, doi: 10.1016/j.physletb.2013.03.024.

[122] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instrum. Meth.* **A434** (1999) 435–443, doi: 10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.

[123] A. L. Read, "Modified frequentist analysis of search results (The CL(s) method)",. Prepared for Workshop on Confidence Limits, Geneva, Switzerland, 17-18 Jan 2000.

[124] CMS Collaboration, "Search for pair production of second-generation scalar leptoquarks in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Rev. Lett.* **106** (2011) 201803, doi: 10.1103/PhysRevLett.106.201803, arXiv:1012.4033.

[125] ATLAS Collaboration, "Search for pair production of first or second generation leptoquarks in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC", *Phys.Rev.* **D83** (2011) 112006, doi: 10.1103/PhysRevD.83.112006, arXiv:1104.4481.

[126] CMS Collaboration, "Measurement of the differential cross section of W bosons produced in association with jets in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Lett. B* **741** (2015) 12, doi: 10.1016/j.physletb.2014.12.003.

[127] M. Czakon, P. Fiedler, and A. Mitov, "The total top quark pair production cross-section at hadron colliders through $O(\alpha_S^4)$", *Phys. Rev. Lett.* **110** (2013) 252004, doi: 10.1103/PhysRevLett.110.252004, arXiv:1303.6254.

[128] CMS Collaboration, "Charged particle multiplicities in pp interactions at $\sqrt{s} = 0.9$, 2.36, and 7TeV", *JHEP* **01** (2010) 079, doi: 10.1007/JHEP01(2011)079, arXiv:1011.5531.

[129] CMS Collaboration, "Commissioning of the particle-flow event reconstruction with the first LHC collisions recorded in the CMS detector", CMS Physics Analysis Summary CMS-PAS-PFT-10-001, (2010).

[130] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual", *Eur. Phys. J. C* **72** (2012) 1896, doi: 10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[131] CMS Collaboration, "Measurement of the inclusive W and Z production cross sections in pp collisions at $\sqrt{s} = 7$ TeV with the CMS experiment", *J. High Energy Phys.* **10** (2011) 132, doi: 10.1007/JHEP10(2011)132.

[132] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) 04013, doi: 10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.

[133] A. Hocker and V. Kartvelishvili, "SVD approach to data unfolding", *Nucl. Instrum. Meth. A* **372** (1996) 469, doi: 10.1016/0168-9002(95)01478-0, arXiv:hep-ph/9509307.

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

[135] H.-L. Lai et al., "New parton distributions for collider physics", *Phys. Rev. D* **82** (2010) 074024, doi: 10.1103/PhysRevD.82.074024, arXiv:1007.2241.

[136] J. Gao et al., "The CT10 NNLO global analysis of QCD", *Phys. Rev. D* **89** (2014) 033009, doi: 10.1103/PhysRevD.89.033009, arXiv:1302.6246.

[137] A. Buckley et al., "Rivet user manual", *Comput. Phys. Commun.* **184** (2013) 2803, doi: 10.1016/j.cpc.2013.05.021, arXiv:1003.0694.

[138] Z. Bern et al., "Four-Jet Production at the Large Hadron Collider at next-to-leading order in QCD", *Phys. Rev. Lett.* **109** (2012) 042001, doi: 10.1103/PhysRevLett.109.042001, arXiv:1112.3940.

[139] Z. Bern et al., "Four-Jet Production at the Large Hadron Collider at next-to-leading order in QCD", *Phys. Rev. Lett.* **109** (2012) 042001, doi: 10.1103/PhysRevLett.109.042001, arXiv:1112.3940.

[140] A. D. Martin et al., "Parton distributions for the LHC", *Eur. Phys. J. C* **63** (2009) 189, doi: 10.1140/epjc/s10052-009-1072-5, arXiv:0901.0002.

[141] R. D. Ball et al., "A first unbiased global NLO determination of parton distributions and their uncertainties", *Nucl. Phys. B* **838** (2010) 136, doi: 10.1016/j.nuclphysb.2010.05.008, arXiv:1002.4407.

[142] S. Alekhin et al., "The PDF4LHC working group interim report", (2011). arXiv:1101.0536.

[143] D. Maître and S. Sapeta, "Simulated NNLO for high-p_{T} observables in vector boson + jets production at the LHC", *Eur. Phys. J. C* **73** (2013) 2663, doi: 10.1140/epjc/s10052-013-2663-8, arXiv:1307.2252.

Appendix A

Additional W+jets unfolding validation plots

This appendix provides figures pertaining to the unfolding of the distributions measured for the W+jets process. The format and content of these plots are detailed in Section 6.5.3.

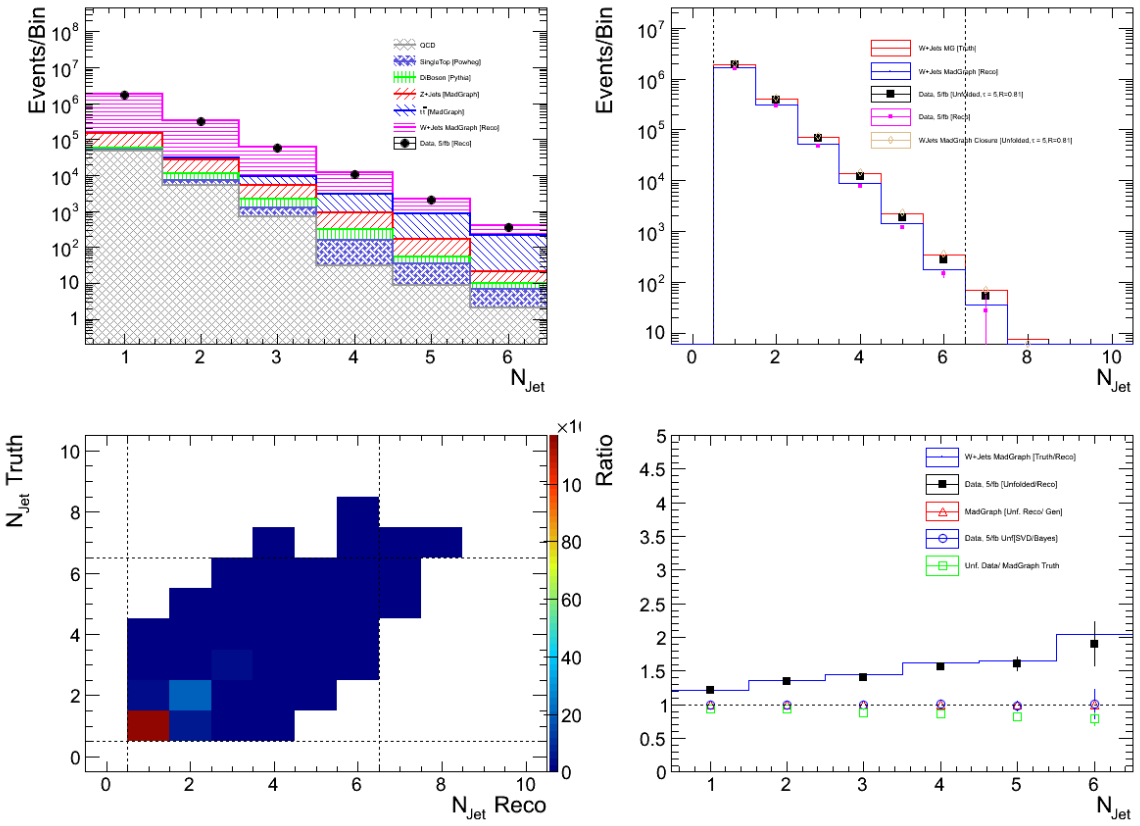

Figure A.1: The unfolding of exclusive jet multiplicity distribution.

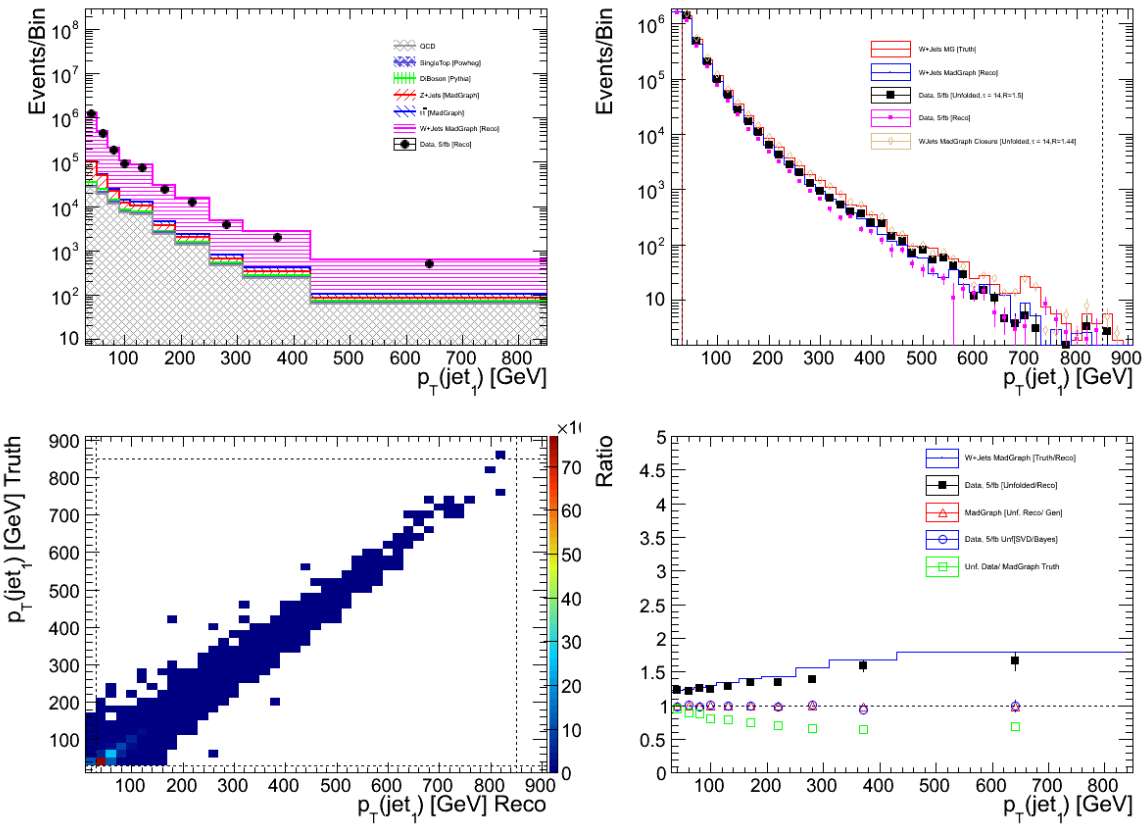

Figure A.2: The unfolding of the $p_{\mathrm{T}}(\mathrm{jet}_1)$ distribution.

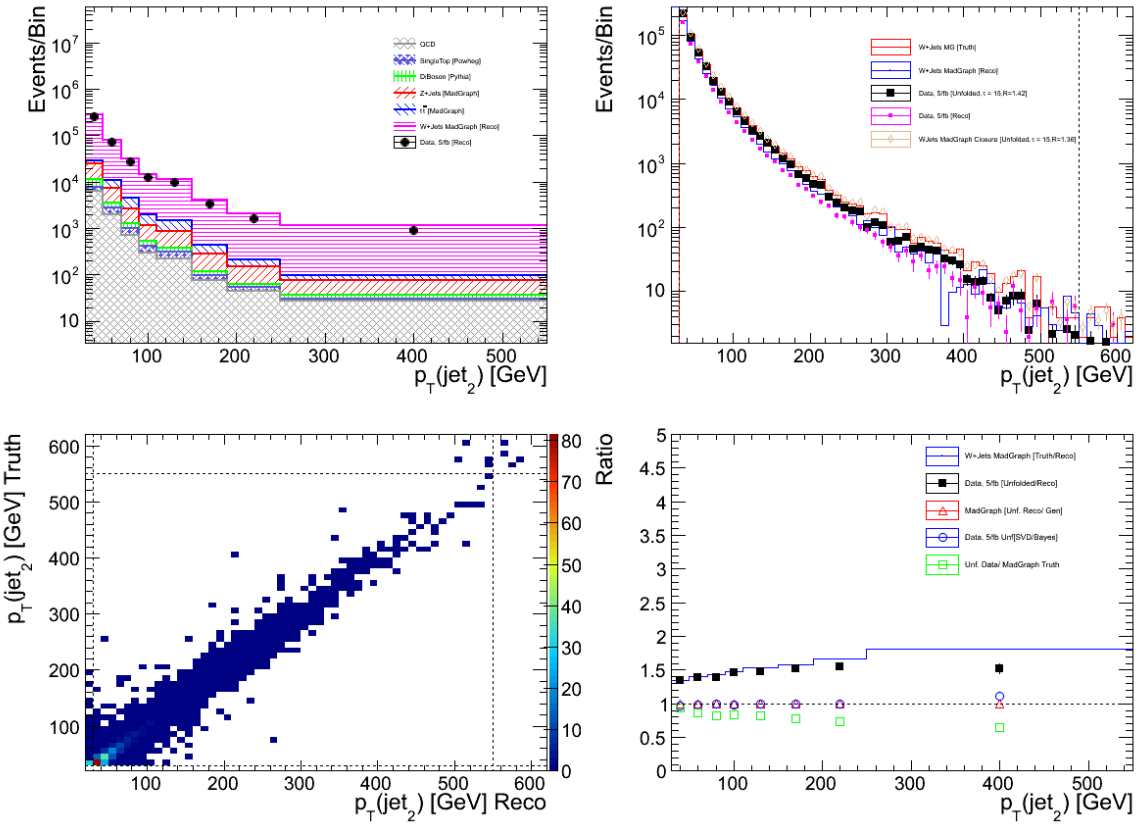

Figure A.3: The unfolding of the $p_{\mathrm{T}}(\mathrm{jet}_2)$ distribution.

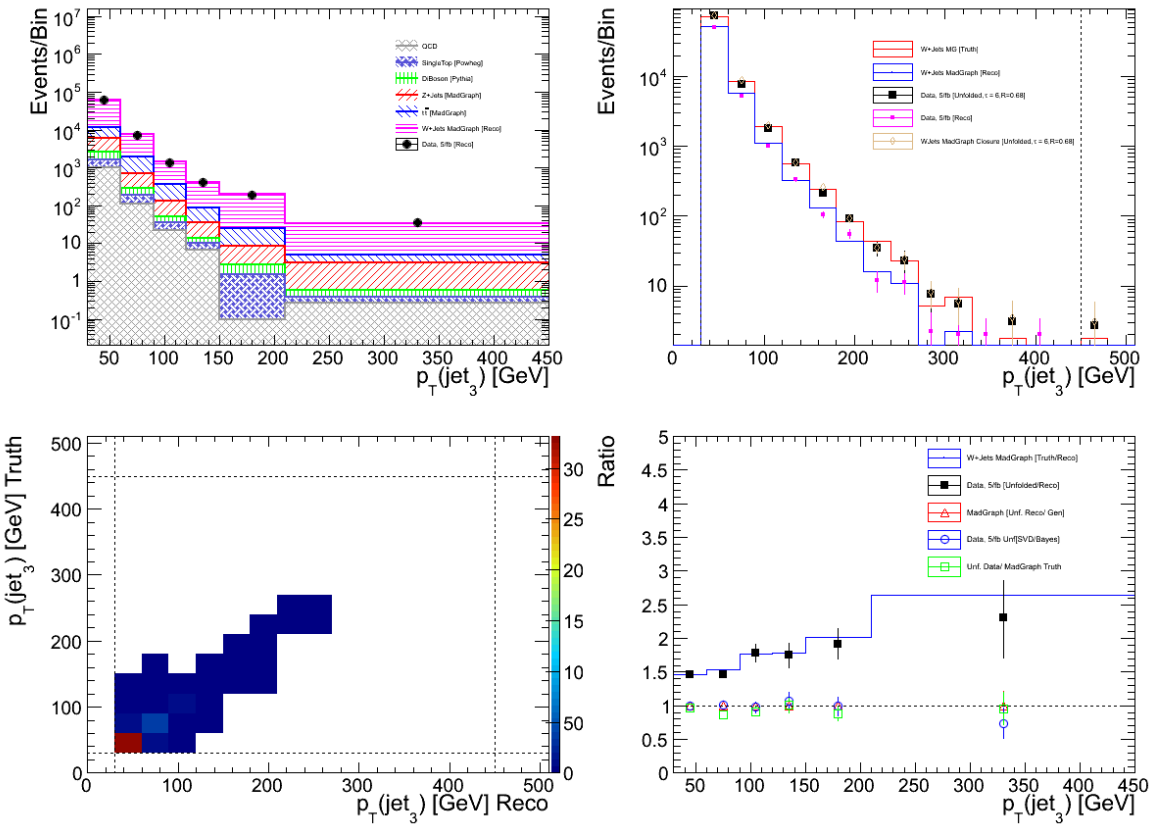

Figure A.4: The unfolding of the $p_{\mathrm{T}}(\mathrm{jet}_1)$ distribution.

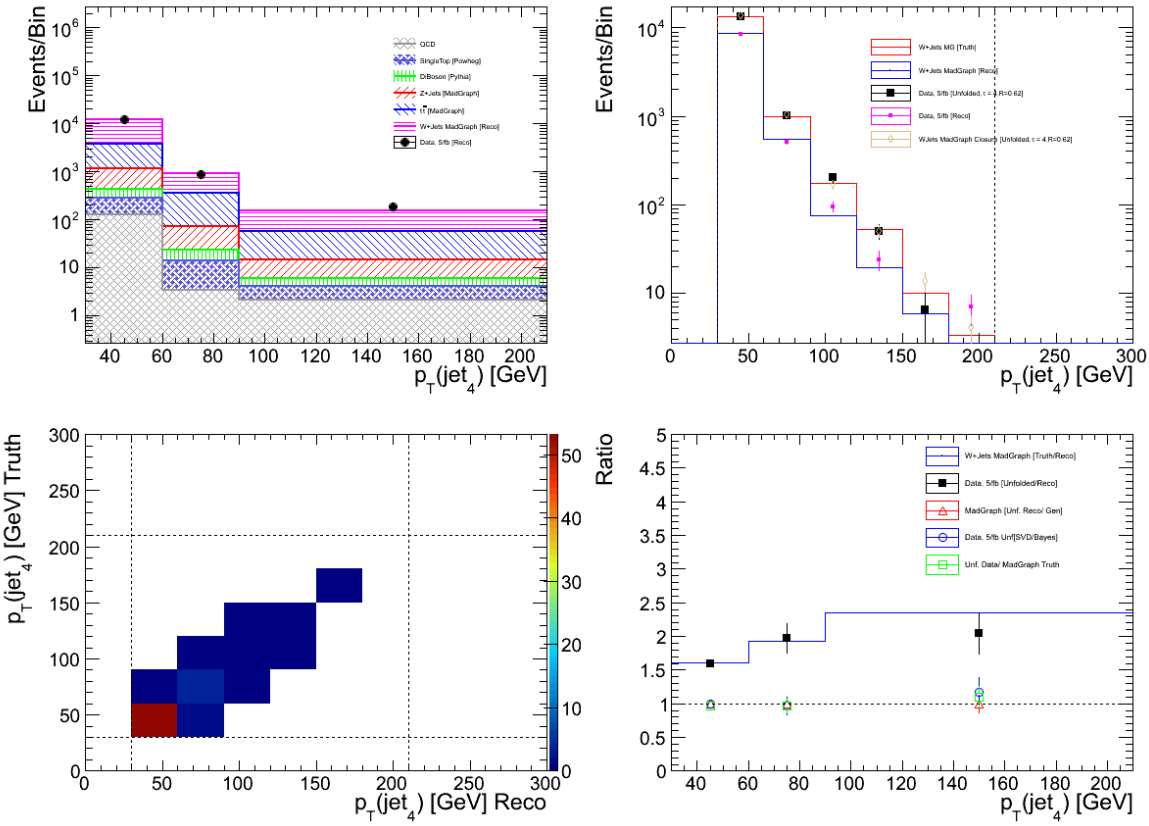

Figure A.5: The unfolding of the $p_{\mathrm{T}}(\mathrm{jet}_4)$ distribution.

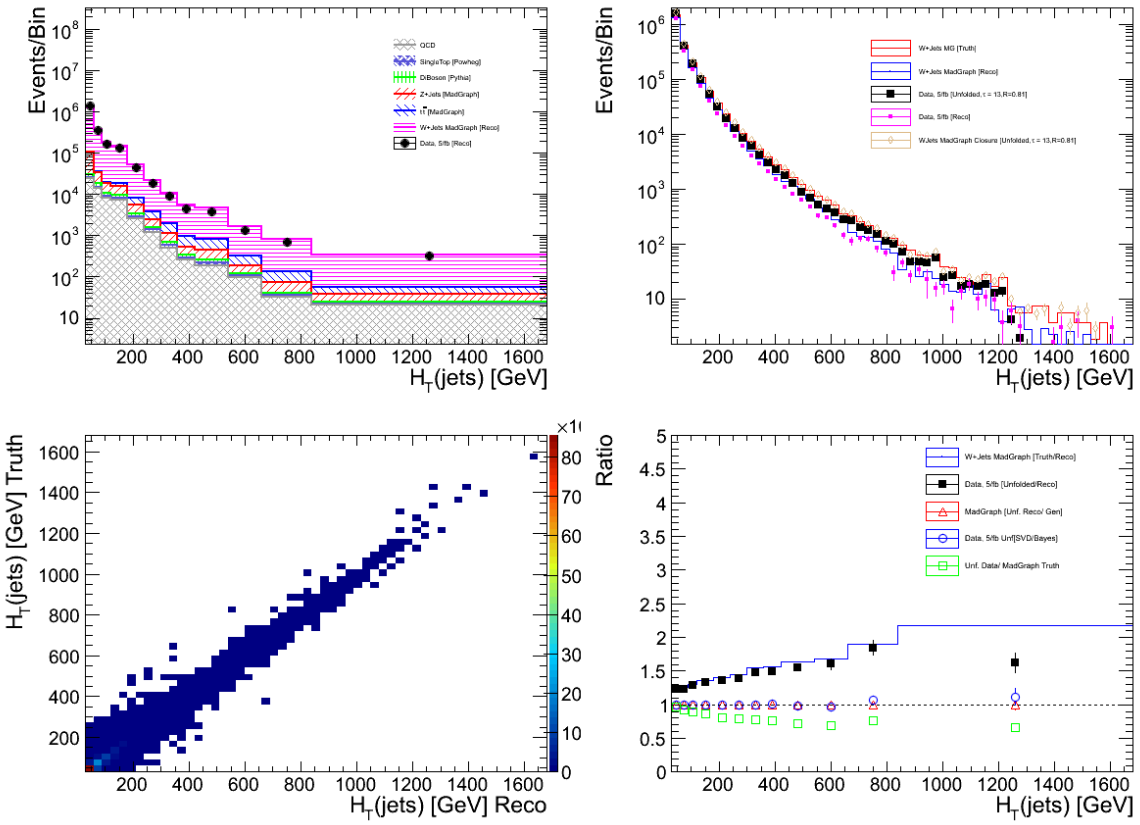

Figure A.6: The unfolding of the H_{T} distribution for events with one or more jet.

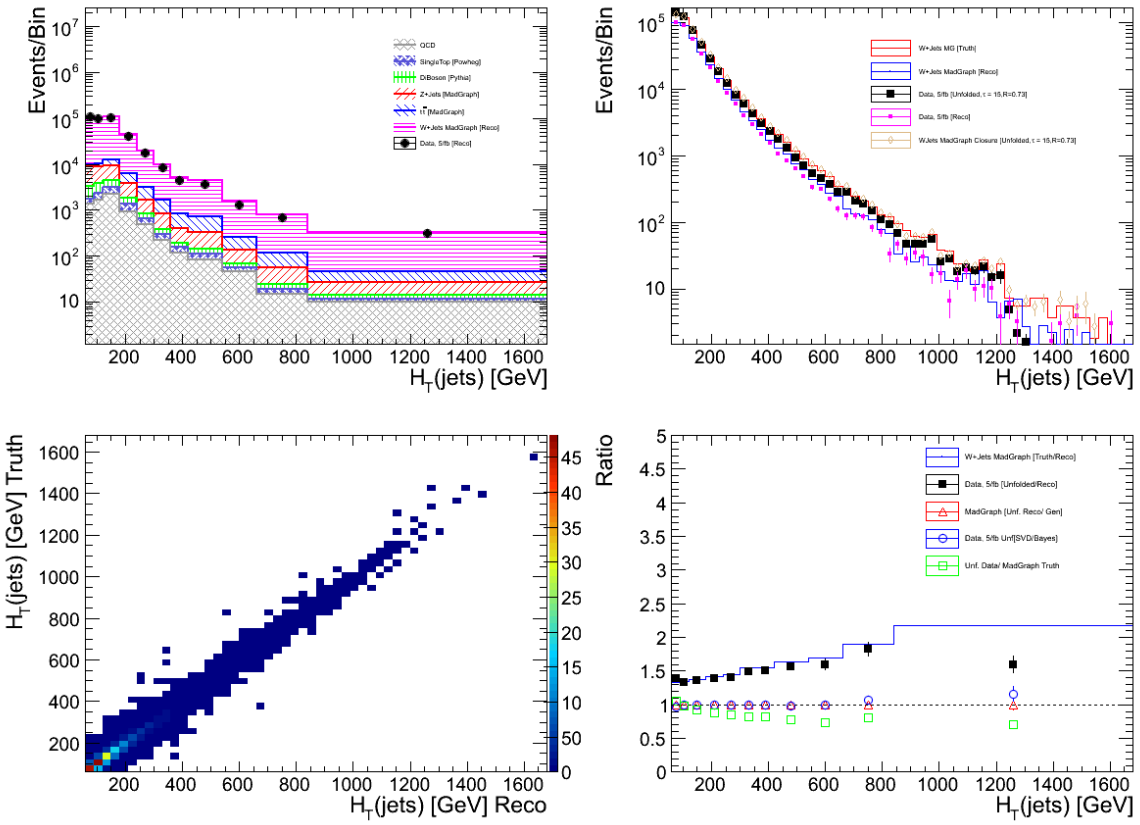

Figure A.7: The unfolding of the H_{T} distribution for events with two or more jets.

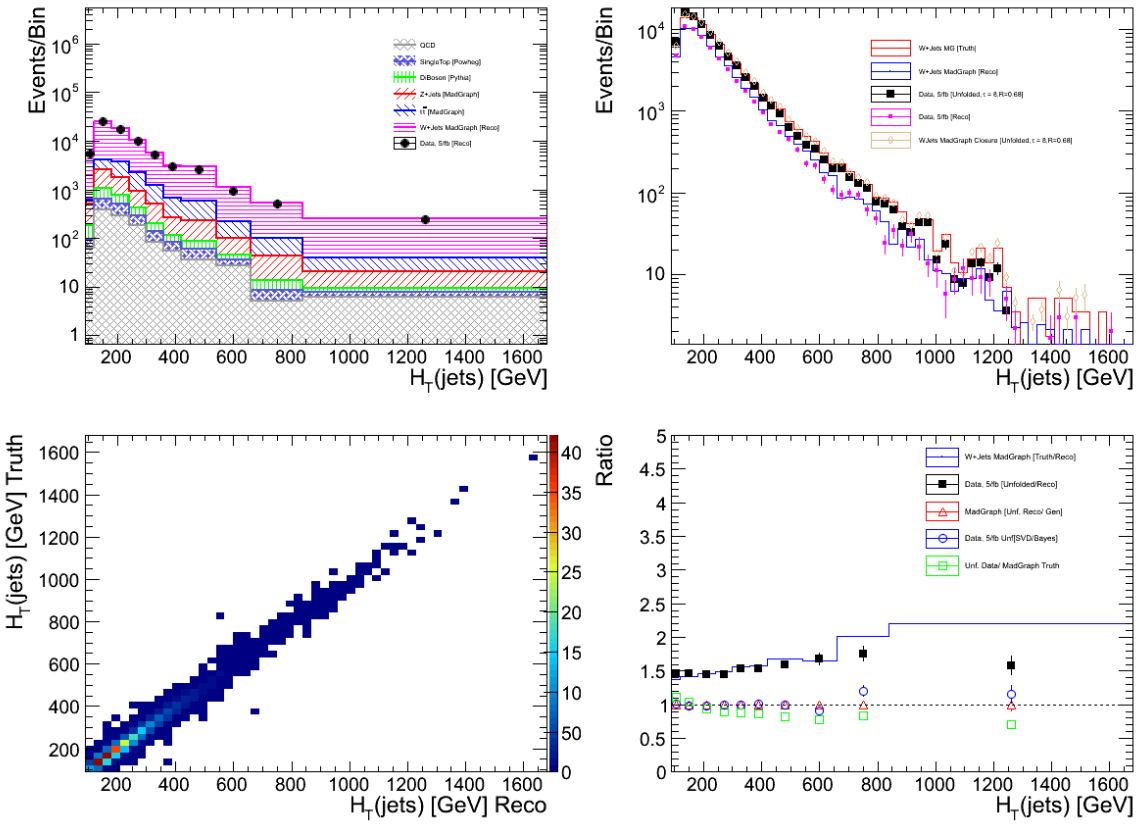

Figure A.8: The unfolding of the H_{T} distribution for events with three or more jets.

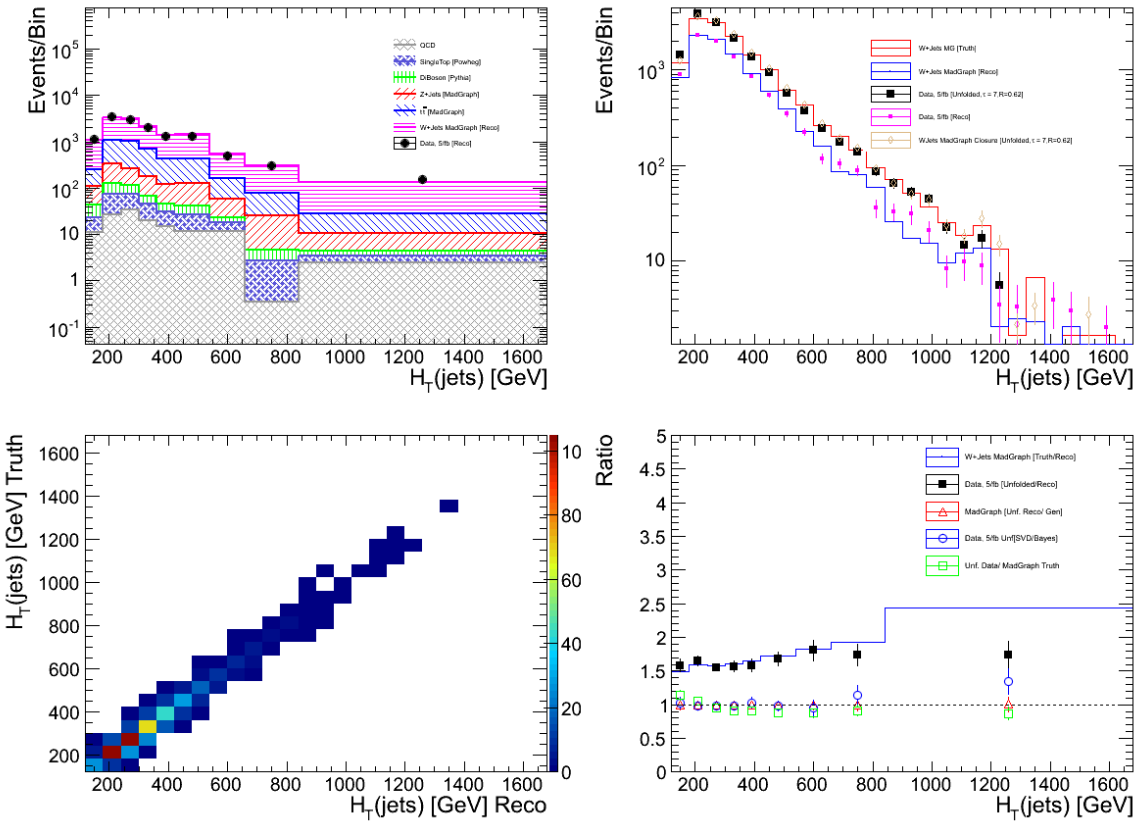

Figure A.9: The unfolding of the H_{T} distribution for events with four or more jets.

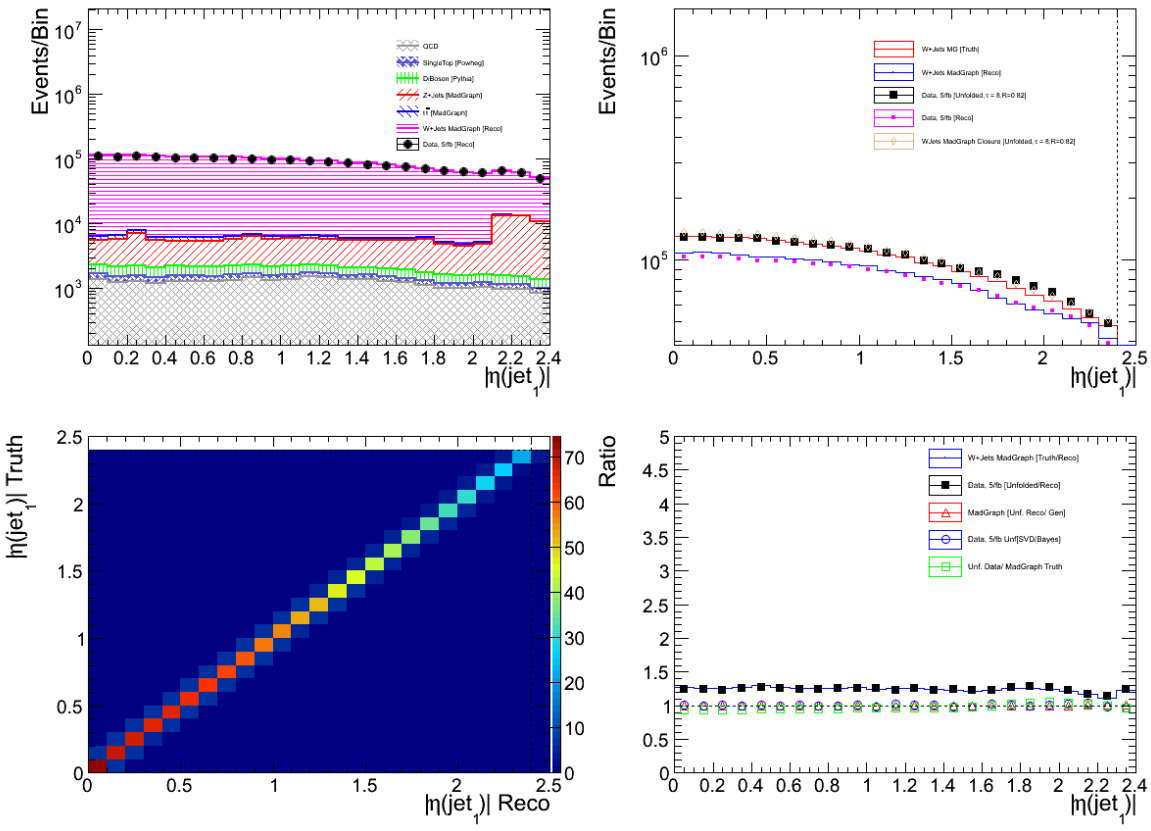

Figure A.10: The unfolding of the $\eta(\text{jet}_1)$ distribution.

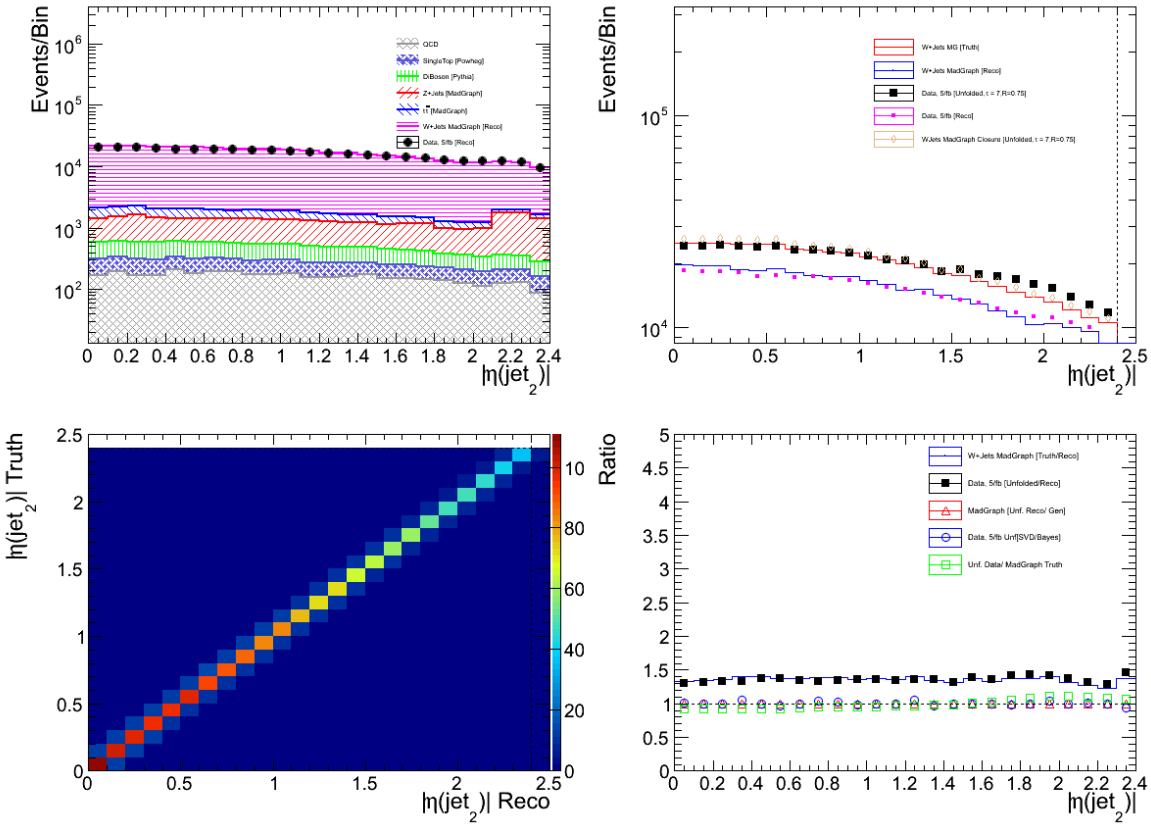

Figure A.11: The unfolding of the $\eta(\text{jet}_2)$ distribution.

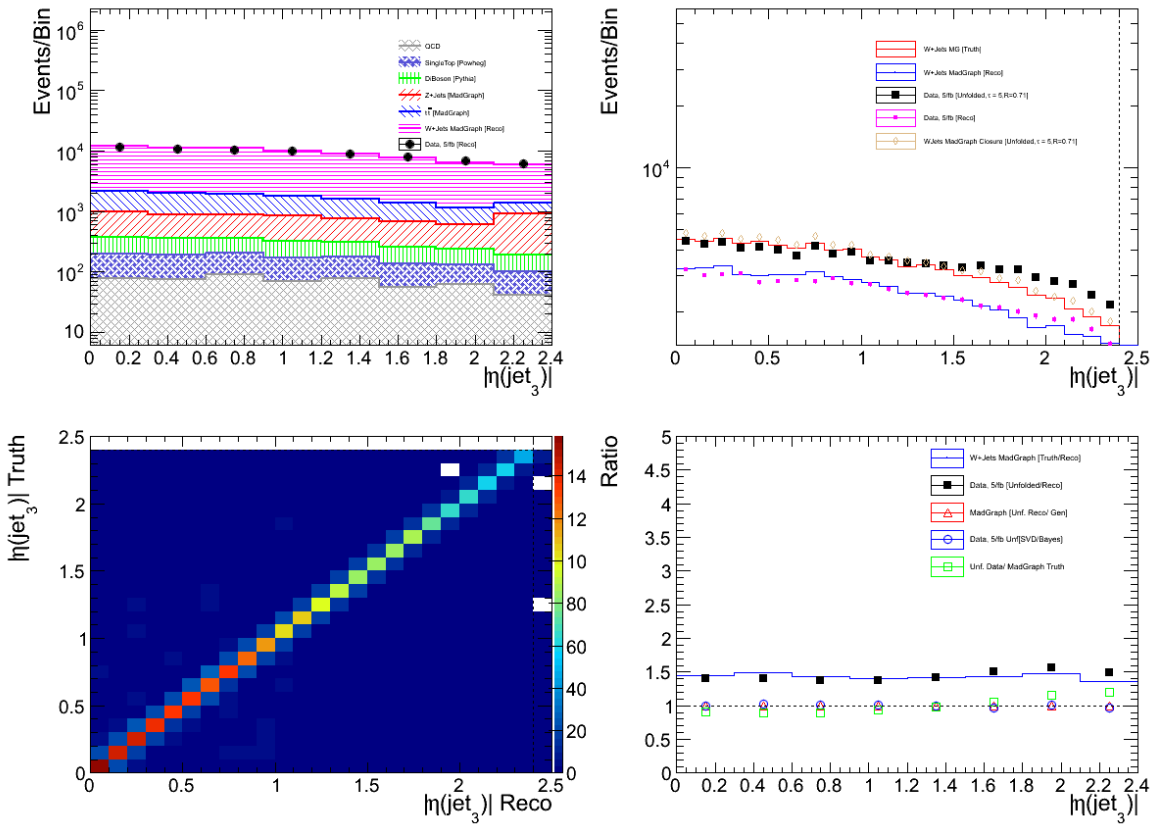

Figure A.12: The unfolding of the $\eta(\text{jet}_3)$ distribution.

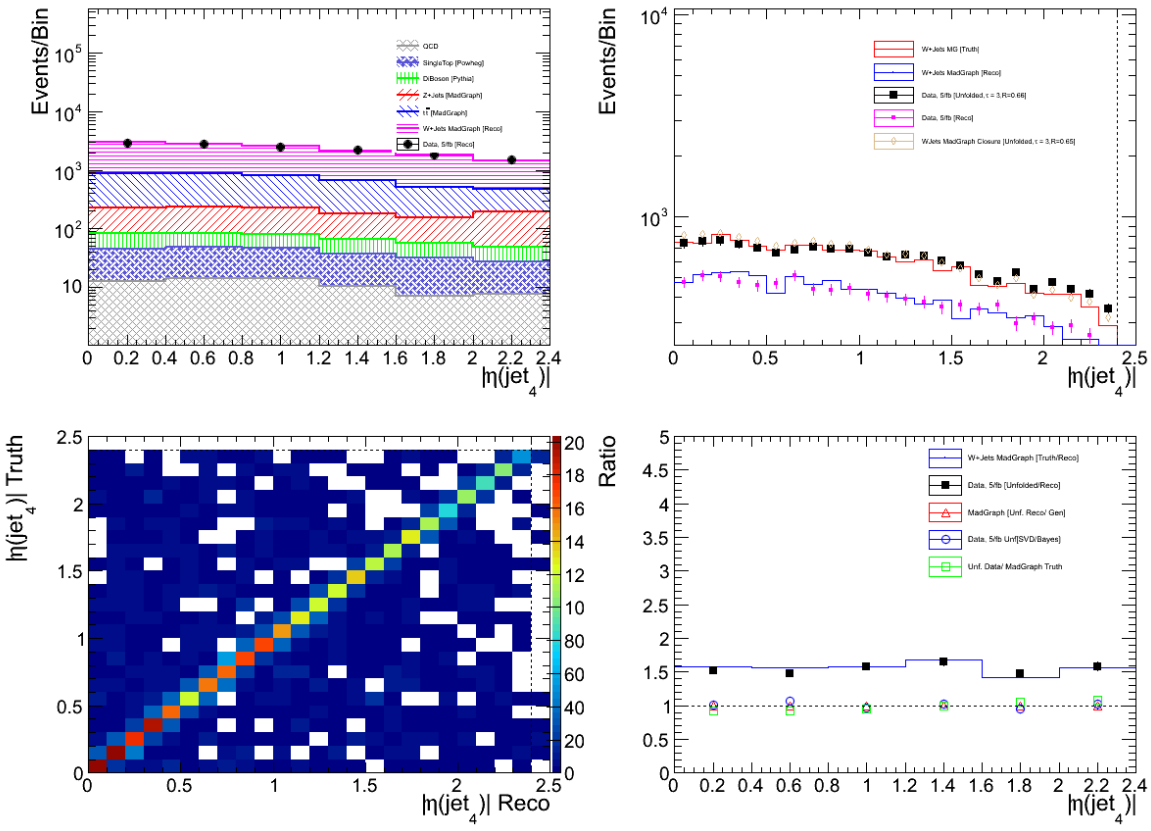

Figure A.13: The unfolding of the $\eta(\text{jet}_4)$ distribution.

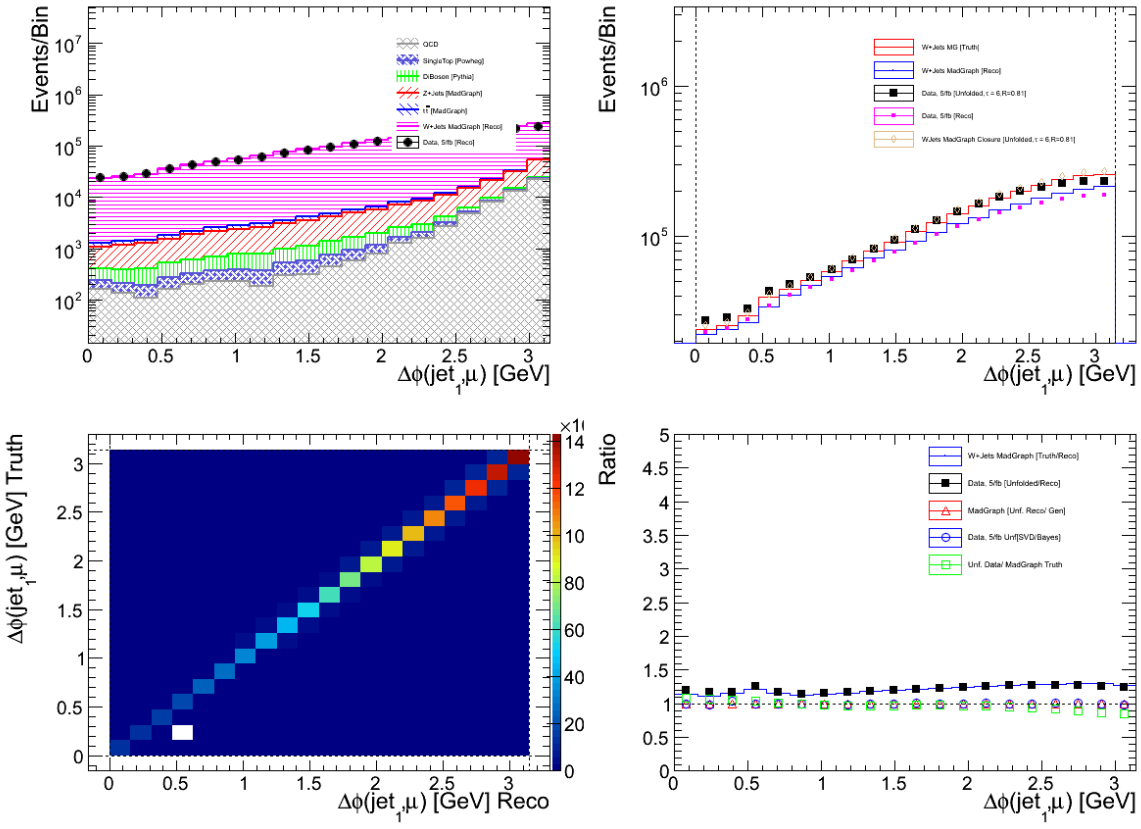

Figure A.14: The unfolding of the $\Delta\phi(\text{jet}_1, \mu)$ distribution.

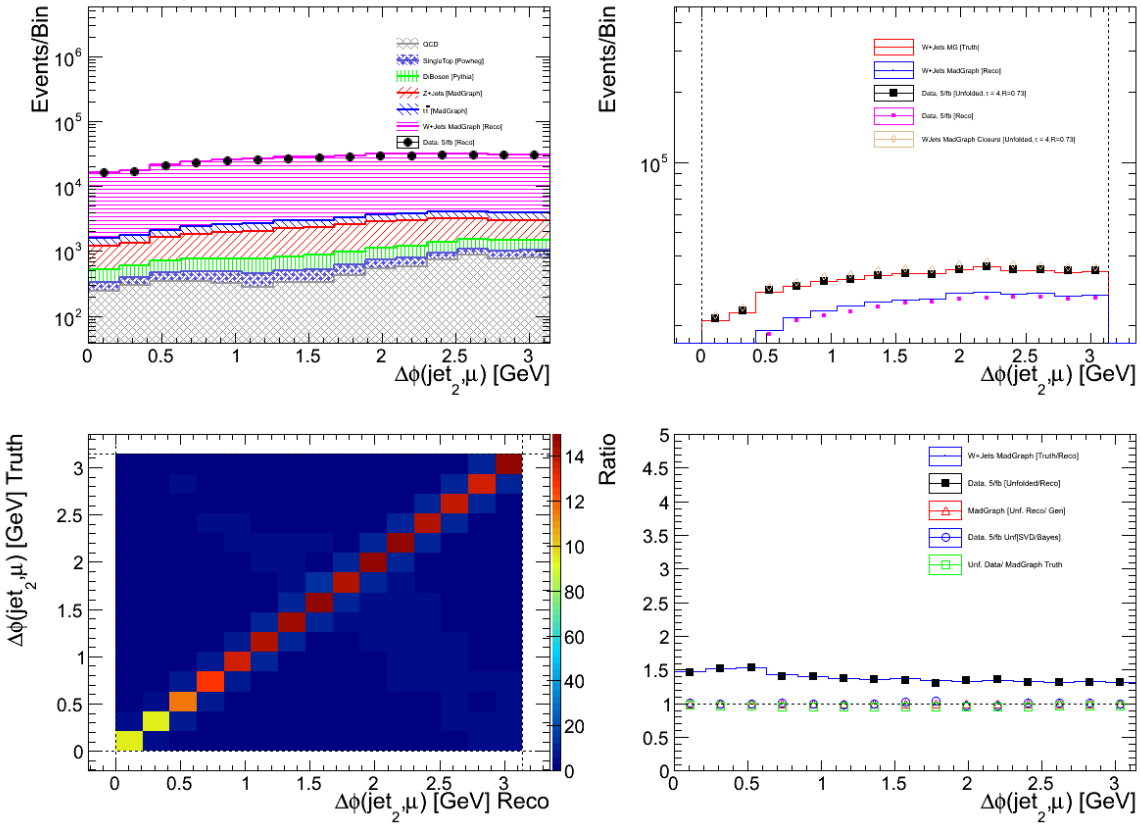

Figure A.15: The unfolding of the $\Delta\phi(\text{jet}_2, \mu)$ distribution.

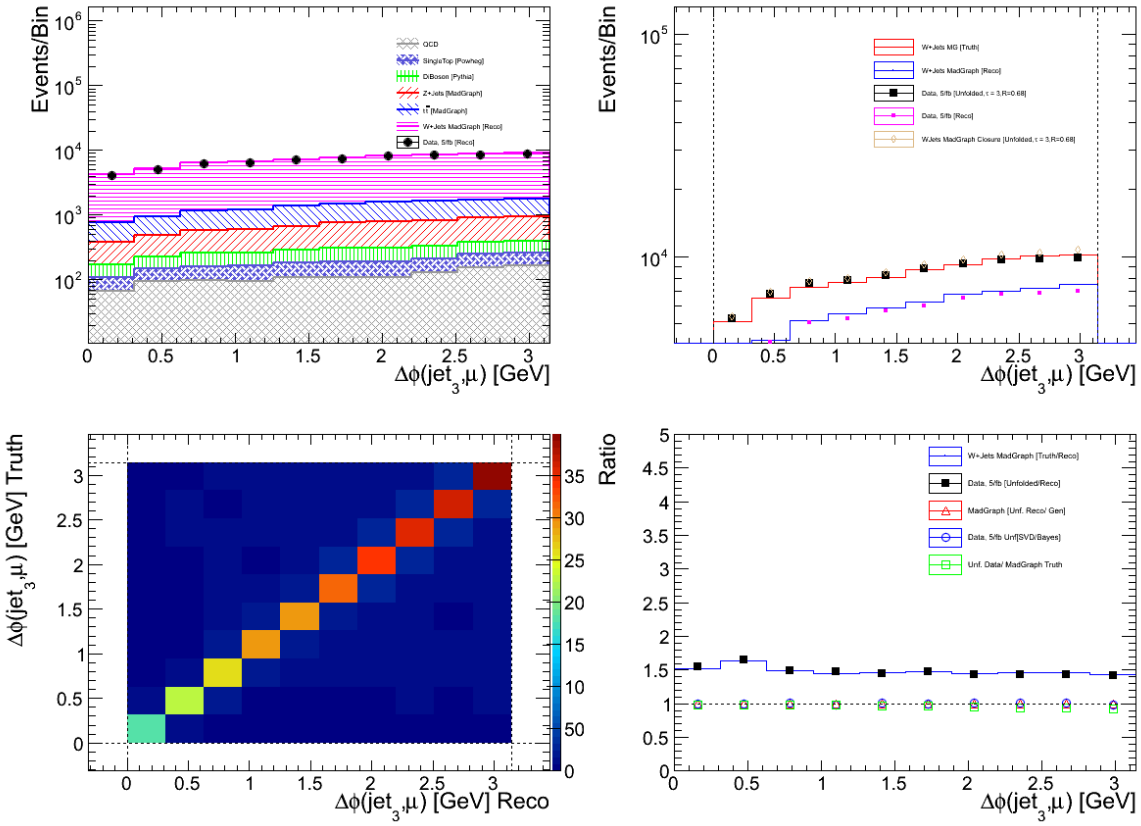

Figure A.16: The unfolding of the $\Delta\phi(\text{jet}_3, \mu)$ distribution.

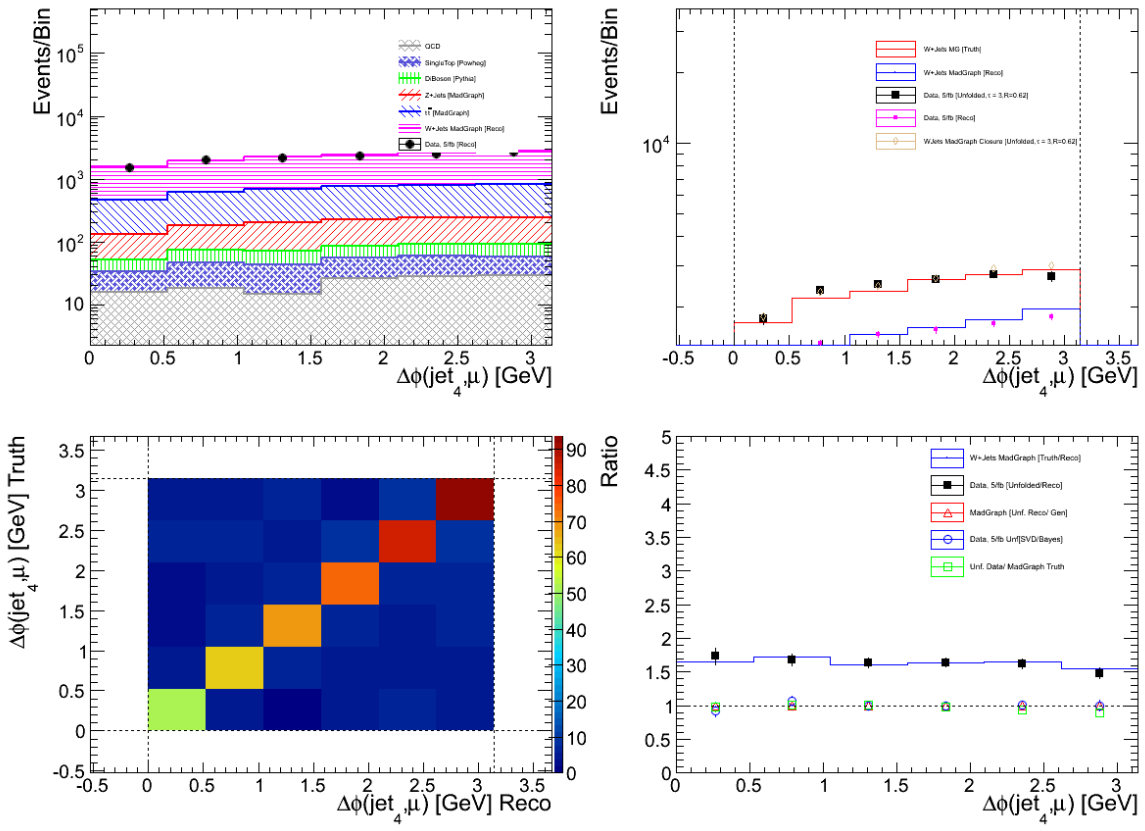

Figure A.17: The unfolding of the $\Delta\phi(\text{jet}_4, \mu)$ distribution.