The Search for Single Production of First and Second Generation Leptoquarks in pp collisions at the LHC

by David Nash
B.S. in Physics, Brandeis 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

February 24, 2016

Dissertation directed by

Darien Wood
Professor of Physics

Acknowledgements

I would like to first express thanks to my advisor, Professor Darien Wood, for his support, insight, and genuine interest in my success over the course of my work. His input and guidance made the work for this dissertation possible. I would also like to especially thank Professor Emanuela Barberis, whose close collaboration was also integral to my productiveness and effectiveness as a scientist throughout my tenure as a PhD student. Gratitude is also due to the rest of my Thesis Committee, Professors Brent Nelson and Toyoko Orimoto, for their input and encouragement.

I would also like to thank my collaborators from both Northeastern University and CMS. In particular I am grateful to Professor Joseph Haley, Doctor Daniele Trocino, Doctor David Morse, Doctor Darin Baumgartel, Doctor Matthew Chasco, and Doctor Ren-Jie Wang for their guidance, friendship, and collaboration while at Northeastern University. I also thank Guido Magazzu, Doctor Piet Verwilligen, Doctor Archana Sharma, Doctor Cesare Calabria, David Sheffield, Professor Ketino Kaadze, Doctor Francesco Santanastasio, Doctor Jim Hirschauer, Professor John Paul Chou, Doctor Daniele del Re, and Doctor Juan Alcarez whose support and collaboration was invaluable to me as I worked on a variety of projects at CMS.

I additionally gratefully acknowledge support of the US National Science Foundation, grants PHY-1066578 and PHY-1401984.

Finally, I would like to thank my parents for all their support, who imbued me with the healthy sense of curiosity that made this work possible.

Abstract of Dissertation

Leptoquarks are theoretical particles proposed by various extensions to the Standard Model of particle physics. They couple to both quarks and leptons, explaining the correspondence between the three generations of quarks and leptons in the Standard Model that are essential to ensure its renormalizability. Experimental constraints indicate that leptoquarks would only couple to a single generation, and this thesis describes searches performed with the CMS detector for leptoquarks of the first or second generation, produced singly and decaying to final states containing either two electrons and one jet or two muons and one jet.

The search is based on a data sample of proton-proton collisions at the LHC with center-of-mass energy $\sqrt{s} = 8\,\text{TeV}$ recorded with the CMS detector and corresponding to an integrated luminosity of 19.6 fb^{-1}. No signal-like excess is observed. Upper limits are set on leptoquark cross sections at the 95% confidence level via the CL_S modified frequentist method. Single production of first-generation leptoquarks with a coupling $\lambda = 1$ and branching fraction $\text{B}(\text{LQ} \to eq) = 1$ is excluded for masses below 1755 GeV, and second-generation leptoquarks with a coupling $\lambda = 1$ and branching fraction $\text{B}(\text{LQ} \to \mu q) = 1$ is excluded for masses below 660 GeV. These lower limits on leptoquark mass are the most stringent limits for single production of leptoquarks to date.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Standard Model of Particle Physics

1.1.1 Properties of the Standard Model

The Standard Model (SM) of particle physics consists of a mathematical framework and physical description of all subatomic processes observed to this point. Since its inception it has shown remarkable predictive power, although it has also been adapted over time to assimilate new measured phenomena. The current state of the SM dates to the initial proposal that electroweak interactions could be described by an $SU(2) \times U(1)$ symmetry followed by the inception of the concept that these interactions could be described via spontaneous gauge symmetry breaking and the Higgs mechanism [1–5]. The strong force was later added, including quantum chromodynamics (QCD), quark mixing angles, and the full 3-generation formulation [1, 6–8]. The current state of the SM is a non abelian gauge theory with the symmetry group

$$SU(3)_C \times SU(2)_L \times U(1)_Y \tag{1.1}$$

where $SU(3)_C$ denotes the strong QCD interaction, $SU(2)_L$ the weak interactions, and $U(1)_Y$ the electromagnetic interaction. The generator of the $U(1)_Y$ group, Y, is called the hypercharge and the generator of the $SU(2)_L$, T, is called the weak isospin. Right handed fermions have a weak isospin of zero, and thus do

not interact via the weak force. Before electroweak symmetry breaking there are 3 weak force bosons ($W^{+,-,3}$), one electromagnetic boson (B), and two complex scalar fields ($\phi^{+,0}$). Electroweak symmetry breaking causes the mixing of the third weak boson and the B boson to create a neutral weak force mediator (Z) and a neutral electromagnetic force mediator (γ), and only the neutral scalar field is the only survivor of the pair becoming the Higgs boson (H). This symmetry breaking enforces the necessary gauge invariance of $SU(2)_L \times U(1)_Y$ and gives the $W^{\pm}$ and Z bosons their mass.

Color charge is the generator of the strong QCD $SU(3)_C$ group, and results in eight gluon fields that are the strong force mediators. There are three possible values of color (r,g,b) and only the quarks and the gluons have color charge, with quarks possessing one color charge and gluons possessing two. Unlike photons, gluons can interact with themselves, but single quarks and gluons can not be isolated and thus bare color charge is not an observable quantity.

There are 12 fermions total in the SM, six leptons and six quarks. The leptons are divided in to three generations: Electrons and electron neutrinos (e^-, ν_e, muons and muon neutrinos (μ^-, ν_μ), and taus and tau neutrinos (τ, ν_τ). Every charged lepton has an oppositely charged anti particle, and possesses a lepton number (L) of one. The quarks are likewise organized in a three generation scheme, each possessing pairs of quarks with electromagnetic charges of $2/3$ and $-1/3$: the up and down quarks, charm and strange, and top and bottom. The quarks have a baryon number of $1/3$, and due to the triple-valued color charge they possess and the necessary "colorlessness" of the strong force they exist in either combinations of three quarks (hadrons) or quark anti-quark pairs (mesons). Both quarks and leptons have electromagnetic charges (Q_{em}) dependent on their hypercharge and the third component of their weak isospin, according to the relation in 1.2.

$$Q_{em} = T_3 + \frac{1}{2}Y \tag{1.2}$$

The fermions are summarized in Table 1.1.

The bosons, which are the force mediators described above, also follow the same relationship between charge, weak isospin, and hypercharge as the fermions. They, however, do not possess lepton number or baryon number. The bosons are summarized in Table 1.2.

Fermion type	Q_{em}	T_3	T	Y	L	B	Spin	Color charge
ℓ_L	-1	$-1/2$	$1/2$	-1	1	0	$1/2$	0
$\nu_{\ell,L}$	0	$1/2$						
ℓ_R	-1	0	0	-2	1	0	$1/2$	0
u_L	$2/3$	$1/2$	$1/2$	$1/3$	0	$1/3$	$1/2$	1
d_L	$-1/3$	$-1/2$						
u_R	$2/3$	0	0	$4/3$	0	$1/3$	$1/2$	1
d_R	$-1/3$			$-1/3$				

TABLE 1.1: The standard model fermions, listed with their quantum numbers. Note that leptons and quarks are listed independent of generation, i.e.: $\ell \in \{e, \mu, \tau\}$, $u \in$ {u,c,t}, and $d \in$ {d,s,b}. [9, 10]

Boson type	Q_{em}	T_3	T	Y	L	B	Spin	Color charge
W^+	1	1						
Z	0	0	1	0	0	0	1	0
W^-	-1	1						
γ	0	0	0	0	0	0	1	0
g	0	0	0	0	0	0	1	2
H	0	$-1/2$	$1/2$	1	0	0	1	0

TABLE 1.2: The standard model bosons, listed with their quantum numbers [9, 10].

The allowed interactions of the SM are depicted in Figure 1.1. Note that leptons and quarks cannot interact with each other directly, without the "help" of the force mediators.

1.1.2 Beyond the Standard Model

While the Standard Model has been heavily probed by experiments and proves to be a very powerful predictive tool, it does have numerous limitations. It possesses no description of the force of gravity whatsoever, despite its considerations of every other known fundamental force. Explanations have been attempted to account for this via the combination of principles of supersymmetry and gravity known as "supergravity" theories, or SUGRA. Of those fundamental forces, the Standard Model accounts for the relative strengths of the electromagnetic, weak, and strong forces but cannot do so for the difference between the electroweak force strength and the gravitational. This difference is known as the "hierarchy problem" and is also closely tied to the problem of large radiative corrections to the Higgs mass

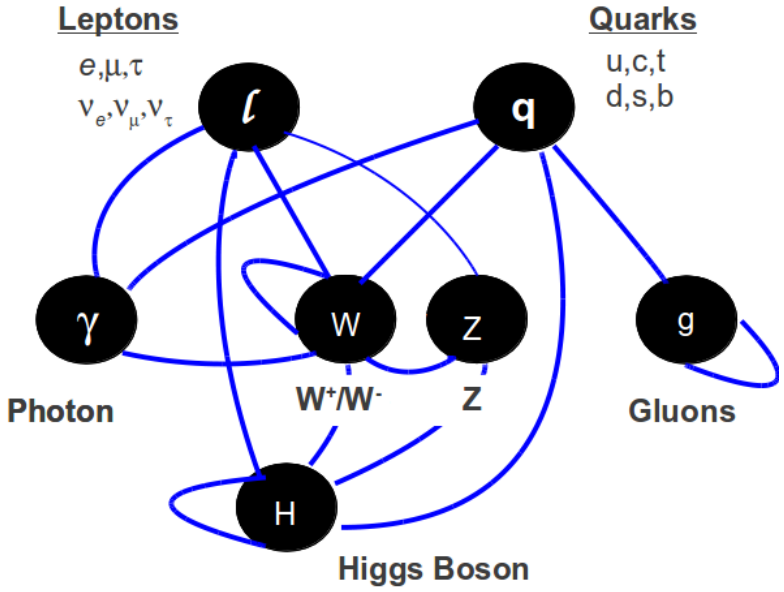


FIGURE 1.1: The particles of the SM depicted with all tree level permitted interactions. The top row contains the fermions, the center row contains the force mediating bosons, and the bottom row contains the Higgs boson.

(a "fine-tuning" problem). Additionally, the SM and the Higgs mechanism specifically predict massive leptons with massless neutrinos, which conflicts with the experimentally confirmed model of neutrino oscillation.

In addition to these problems, the SM leaves open a few compelling questions. One question, first revealed by the field of astrophysics, is that of dark matter. Observations of rotation curves of galaxies indicated that nonluminous mass must be present. An alternative theory of modified Newtonian dynamics was proposed, but then disproven initially by observations of the Bullet cluster and later by other similar phenomena. The Bullet cluster data showed gravitational lensing occurring in nonluminous regions of two colliding galaxies, which made the issue of actual nonluminous matter concrete. Measurements of the portion of our universe that is made of dark matter indicate that it is actually the majority of all matter in our universe, and there is no mechanism in the SM that could explain that fact. While neutrinos qualify for the qualitative properties of dark matter they cannot explain the quantity of mass that must be nonluminous, and thus there is no dark matter "candidate" particle present in the SM. Numerous models beyond the SM (BSM), such as supersymmetry and string theory contain such candidates. Additional questions raised in astrophysics that the SM cannot explain are the observed rate of expansion and the necessary abundance of "dark energy" that contributes to this expansion.

Outside of the field of astrophysics the SM offers an additional question: it lacks any explanation for the symmetry between the three generations of quark pairs and the three generations of lepton pairs. The existence of this symmetry implies a presence of a mechanism to account for it. Numerous extensions to the SM contain such a mechanism: a hypothetical new particle that possesses both lepton and quark number. This particle is called a leptoquark and a search for its decay signatures at the LHC is the main subject of this thesis.

1.2 Leptoquarks

1.2.1 Theoretical considerations

Various extensions to the SM include predictions of new bosons that carry both lepton and baryon number, motivated by the essential symmetry of leptons and quarks in it. Triangle anomalies that otherwise threaten the renormalizability of the SM are canceled by the requirement

$$\sum_n Q_{en}^2 (Q_L - Q_R)_n = 0 \tag{1.3}$$

for each of the three fermionic families, where Q_{en}, Q_{Ln}, and Q_{Rn} denote the elecromagnetic, left-, and right-handed neutral current charges [11].

These new bosons, called leptoquarks (LQ) are hypothetical color-triplet bosons with spin 0 (scalar LQ) or 1 (vector LQ) that are predicted by many extensions of the standard model (SM) of particle physics, such as Grand Unified Theories [12–19], technicolor schemes [20–22], and composite models [23]. They carry fractional electric charge ($\pm\frac{1}{3}$ for LQs considered in this paper) and both baryon and lepton numbers and thus couple to a lepton and a quark. Existing experimental limits on flavor changing neutral currents and other rare processes disfavor leptoquarks that couple to a quark and lepton of a different SM generation or of more than one SM generation [24, 25].

The description of the LQ phenomenology considered in this thesis begins with an effective Lagrangian put forth by W. Buchmuller, R. Ruckl, and D. Wyler in 1986 in the lead up to the turning on of the HERA collider. [26, 27] with the

general dimensionless SU(3) × SU(2) × U(1) invariant couplings of scalar and vector leptoquarks to leptons and quarks, satisfying baryon and lepton number conservation:

$$\mathcal{L} = \mathcal{L}^{f}_{|\mathrm{F}|=0} + \mathcal{L}^{f}_{|\mathrm{F}|=2} \tag{1.4}$$

The $\mathcal{L}^{f}_{|\mathrm{F}|=0,2}$ Lagrangians describe the Yukawa interactions of LQs with leptons and quarks, changing the fermion number F by 0 or 2, where $F = 3B + L$, B is the baryon number and L is the leptons number. $\mathcal{L}^{f}_{|\mathrm{F}|=0,2}$ are diagonal in flavor:

$$\begin{aligned}
\mathcal{L}^{f}_{|\mathrm{F}|=0} &= (h_{2L}\bar{u}_R\ell_L + h_{2R}\bar{q}_L i\sigma_2 e_R)S_{\frac{1}{2}} + \tilde{h}_{2L}\bar{d}_R\ell_L\tilde{S}_{\frac{1}{2}} \\
&\quad +(h_{1L}\bar{q}_L\ell_L + h_{1R}\bar{d}_R\gamma^\mu e_R)V_{0\mu} + \tilde{h}_{1R}\bar{u}_R\gamma^\mu e_R\tilde{V}_{0\mu} \\
&\quad +h_{3L}\bar{q}_L\vec{\sigma}\gamma^\mu\ell_L\vec{V}_{1\mu} + h.c.,
\end{aligned} \tag{1.5}$$

$$\begin{aligned}
\mathcal{L}^{f}_{|\mathrm{F}|=2} &= (g_{2L}\bar{d}^c_R\gamma^\mu\ell_L + g_{2R}\bar{q}^c_L\gamma^\mu e_R)V_{\frac{1}{2}\mu} + \tilde{g}_{2L}\bar{u}^c_R\gamma^\mu\ell_L\tilde{V}_{\frac{1}{2}\mu} \\
&\quad +(g_{1L}\bar{q}^c_L i\sigma_2\ell_L + g_{1R}\bar{u}^c_R e_R)S_0 + \tilde{g}_{1R}\bar{d}^c_R e_R\tilde{S}_0 \\
&\quad +g_{3L}\bar{q}^c_L i\sigma_2\vec{\sigma}\ell_L\vec{S}_1 + h.c.,
\end{aligned} \tag{1.6}$$

where S ($\tilde{S}$) and V ($\tilde{V}$) are scalar and vector leptoquark fields, respectively, q_L and ℓ_L are the left handed quark and lepton SU(2) doublets, and e_R, d_R, u_R are the right handed charged leptons, down-, and up-quarks, respectively. Charge conjugated fields are denoted by $\psi_c = \bar{\psi}^T$, σ_i are the Pauli spin matrices. The subscripts L, R, of the coupling constants g_i and h_i denote the lepton chirality. The LQ indices give the weak isospin.

1.2.2 Signatures at the LHC

In pp collisions LQs can be produced either singly or in pairs, with very similar final state decay products. In pair production, LQs are produced predominantly via gluon-gluon fusion with additional contributions from quark-quark annihilation, as indicated in Eq. 1.7 and Eq. 1.8

LQ type	Q_{em}	T_3	T	Y	F=3B+L	$\lambda_L(\ell q)$	$\lambda_L(\nu q)$	$\lambda_R(\ell q)$	Final States
$S_{0,L}$	1/3	0	0	2/3	−2	g_{1L}	$-g_{1L}$	0	$\ell_L^+\bar{u}_L, \bar{\nu}_L\bar{d}_L$
$S_{0,R}$	1/3	0	0	2/3	−2	0	0	g_{1R}	$\ell_R^+\bar{u}_R$
$\tilde{S}_{0,R}$	4/3	0	0	8/3	−2	0	0	$\tilde{g}_{1R}$	$\ell_R^+\bar{d}_R$
$S_{\frac{1}{2},L}$	5/3	1/2	1/2	7/3	−2	h_{2L}	0	0	$\ell_L^+u_L$
$S'_{\frac{1}{2},L}$	2/3	−1/2				0	h_{2L}	0	$\bar{\nu}_L u_L$
$S_{\frac{1}{2},R}$	5/3	1/2	1/2	7/3	−2	0	0	h_{2R}	$\ell_R^+u_R$
$S'_{\frac{1}{2},R}$	2/3	−1/2				0	0	$-h_{2R}$	$\ell_R^+d_R$
$\tilde{S}_{\frac{1}{2},L}$	2/3	1/2	1/2	1/3	−2	$\tilde{h}_{2L}$	0	0	$\ell_L^+d_L$
$\tilde{S}'_{\frac{1}{2},L}$	−1/3	−1/2				0	$\tilde{h}_{2L}$	0	$\bar{\nu}_L d_L$
$S_{1,L}$	4/3	1				$-\sqrt{2}g_{3L}$	0	0	$\ell_L^+\bar{d}_L$
$S_{1,L}$	1/3	0	1	2/3	−2	$-g_{3L}$	0	$-g_{3L}$	$\ell_L^+\bar{u}_L, \bar{\nu}_L\bar{d}_L$
$S_{1,L}$	−2/3	−1					0	$\sqrt{2}g_{3L}$	$\bar{\nu}_L\bar{u}_L$
$V_{0,L}$	2/3	0	0	4/3	−2	h_{1L}	h_{1L}	0	$\ell_L^+d_R, \bar{\nu}_L u_R$
$V_{0,R}$	2/3	0	0	4/3	−2	0	0	h_{1R}	$\ell_R^+d_L$
$\tilde{V}_{0,R}$	5/3	0	0	10/3	−2	0	0	$\tilde{h}_{1R}$	$\ell_R^+u_L$
$V_{\frac{1}{2},L}$	4/3	1/2	1/2	5/3	−2	g_{2L}	0	0	$\ell_L^+d_R$
$V'_{\frac{1}{2},L}$	1/3	−1/2				0	g_{2L}	0	$\bar{\nu}_L\bar{d}_R$
$V_{\frac{1}{2},R}$	4/3	1/2	1/2	5/3	−2	0	0	g_{2R}	$\ell_R^+\bar{d}_L$
$V'_{\frac{1}{2},R}$	1/3	−1/2				0	0	g_{2R}	$\ell_R^+\bar{u}_L$
$\tilde{V}_{\frac{1}{2},L}$	1/3	1/2	1/2	−1/3	−2	$\tilde{g}_{2L}$	0	0	$\ell_L^+\bar{u}_R$
$\tilde{V}'_{\frac{1}{2},L}$	−2/3	−1/2				0	$\tilde{g}_{2L}$	0	$\nu_L\bar{u}_R$
$V_{1,L}$	5/3	1				$\sqrt{2}h_{3L}$	0	0	$\ell_L^+u_R$
$V_{1,L}$	2/3	0	1	4/3	−2	$-h_{3L}$	0	h_{3L}	$\ell_L^+\bar{d}_R, \bar{\nu}_L u_R$
$V_{1,L}$	−1/3	−1					0	$\sqrt{2}h_{3L}$	$\bar{\nu}_L d_R$

TABLE 1.3: Scalar (S and $\tilde{S}$) and vector (V and $\tilde{V}$) type LQs, with their quantum numbers, coupling strengths, and final state decay modes. Quantum numbers given are fermion number (F), hypercharge (Y), weak isospin (T), the third component of weak isospin (T_3), and electric charge (Q_{em}). Note that the tilde indicates a difference in hypercharge value between two LQ types.

$$g + g \rightarrow \mathrm{LQ} + \overline{\mathrm{LQ}} \tag{1.7}$$

$$q + \overline{q} \rightarrow \mathrm{LQ} + \overline{\mathrm{LQ}} \tag{1.8}$$

These decay modes result in a final state of two leptons and two jets. Single production of LQs occurs in association with a lepton via quark-gluon fusion, as indicated in Eq. 1.9, and thus results in a final state of two leptons and one jet.

$$q + g \rightarrow \mathrm{LQ} + \ell \tag{1.9}$$

Of all the diagrams contributing to pair production, only one posseses any λ-dependence at all and the overall dependence on λ in pair production is negligible [28]. In contrast to this single production diagrams, depicted in Figures 1.2 through 1.4, do scale with the coupling. Single production cross sections decrease more slowly with the coupling, exceeding pair production for leptoquark masses on the order of 1 TeV for $\lambda = 0.6$.

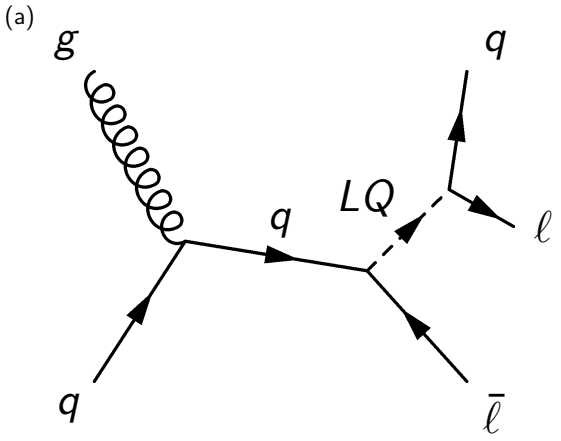


FIGURE 1.2: The s-channel resonant LQ production diagram.

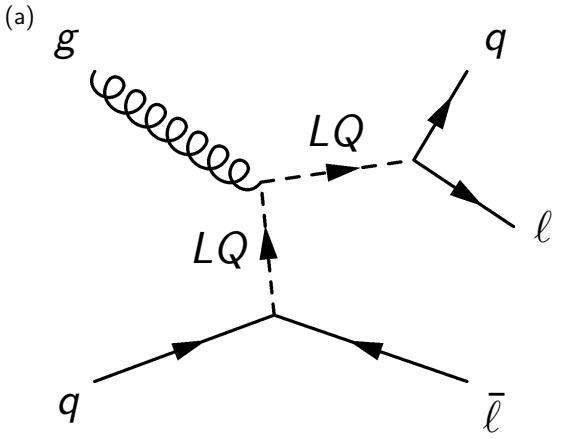


FIGURE 1.3: The t-channel LQ production diagram with a resonant production component.

The main single leptoquark production mode at the LHC is the resonant diagram shown in Figure 1.2. However, significant contributions are made by the diagrams with non-resonant components shown in Figures 1.3 and 1.4. These contributions

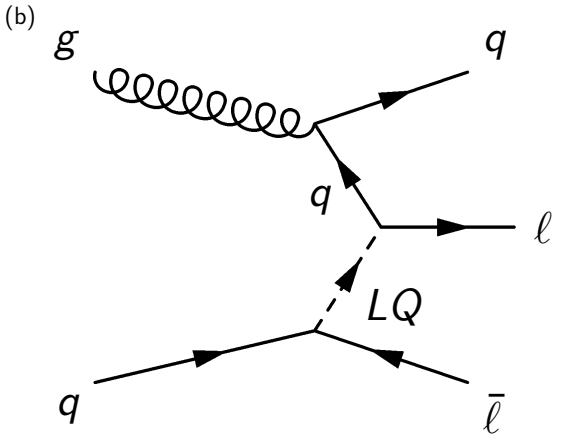


FIGURE 1.4: The t-channel nonresonant LQ production diagram.

increase with both the LQ mass and coupling; the invariant mass distribution of a first generation LQ, of LQ mass $M_{\mathrm{LQ}} = 1$ TeV and coupling $\lambda = 1.0$, possesses a tail that extends to very low masses that is comparable to the peak in magnitude. This phenomenon is discussed in further detail in Section 4.1.

Due to the parton distribution function (PDF) dependence of single LQ production and the suppression of 2nd generation quarks in the proton PDF, the cross sections for second generation LQs is significantly suppressed with respect to first generation LQs.

The signatures for LQ single production include three-object final states with leptons (ℓ), neutrinos (ν), and a single jet (j). There are two distinct final state signatures, $\ell^+\ell^- j$ and $\nu\nu j$. This thesis considers single LQ production with $\beta = 1.0$, thus only final state signatures with two leptons and a jet. The major SM backgrounds that can mimic this signature are Z boson production in association with one or more jets and $t\bar{t}$ production. In the case of Z boson production, the Z decays to two leptons yielding the same final state particles as single LQ production. There are three possible decay modes for $t\bar{t}$ production: fully leptonic, partially leptonic, and hadronic. It is the fully leptonic mode, $t\bar{t} \to bW(\to \ell\nu)bW(\to \ell\nu)$ that can mimic single LQ production with two charged leptons, although it does also possess additional missing transverse energy from the neutrinos. Other SM processes that are additional small contributions to the total background that of events with two charged leptons and at least one jet are diboson (WW, WZ,ZZ) production, single t-quark production, and QCD multijet production in which two jets are misidentified as leptons.

1.2.3 Previous searches

In addition to the indirect limits that restrict LQs to coupling to same-generation leptons discussed in Section 1.7, various collaborations have performed direct searches for LQs.

Early searches for scalar leptoquarks were conducted at the HERA ep collider by the H1 [29–31] and ZEUS [32] collaborations and found no evidence for leptoquarks. Limits were set at 95% confidence level of $M_{\mathrm{LQ}}(\text{1st gen.}) < 250$ GeV for $|F| = 0$ with the assumption that $\lambda > \lambda_{em}$. Later results in 1997 from the HERA collider had intriguing implications for leptoquarks and increased interest in searches for them. Both the H1 and ZEUS collaborations reported excesses in neutral current deep inelastic scattering events ($e^+p \to e^+X$) at high values of Q^2, where Q^2 is the four-momentum transfer [33, 34]. Figure 1.6 depicts the excess measured by the H1 collaboration, showing both the total number of events measured and the ratio to the SM expectation versus Q^2. The reconstructed invariant mass of the final state electron and jet is compared to the SM expectation in Figure 1.6 with an excess visible at high values of Q^2, concentrated near values of M_{e^+j} of 200 GeV.

Searches for LQs following the observation of this excess have ruled out LQs at various detectors. The D0 collaboration produced limits on singly first generation LQs of $M_{\mathrm{LQ}} > 274$ GeV ($\lambda = 1.0$ and $\beta = 1.0$) [35]. In 2011, the H1 collaboration published comprehensive limits on singly produced LQs of various types as a function of λ [31], placing a limit on the S_0^R type LQ that this thesis primarily considers of $M_{\mathrm{LQ}} > 500$ GeV ($\lambda = 1.0$ and $\beta = 1.0$). These limits are shown in Figure 1.7.

More recently, in results published in concert with the single LQ results presented in this thesis, the CMS collaboration produced limits on pair production of LQs of 1010 (1030) GeV in the first and second generations, assuming $\beta = 1.0$ [36]. Data being collected at this very moment at a center of mass energy of 13 TeV comprise data collected on collisions at the highest energies performed to this date. This increase in energy roughly doubles the single LQ production cross sections at the LHC, allowing for greater reach for this statistically limited search.

The 20 fb^{-1} of data collected by CMS with a center-of-mass energy of $\sqrt{s} =$ 8 TeV was unprecedented for hadron colliders, both with respect to total integrated

luminosity and collision energy. For a statistically limited search such as the search for single LQ production, this provides significantly greater reach than achieved in the past at other colliders. Additionally, advances in detector technology, described in Chapter 2, and in reconstruction, described in Chapter 3, provide the ability to even more precisely measure particle kinematics. These techniques along with the increase in integrated luminosity also create a reduction of the systematic uncertainties associated with measurements made with the CMS detector.

In this thesis, a search for single production of first- and second-generation scalar LQs with the CMS detector is presented, as a function of LQ mass and coupling λ, using all the 8 TeV collision data collected. The SM backgrounds are studied in detail, and data-driven methods for estimating the most significant backgrounds are utilized.

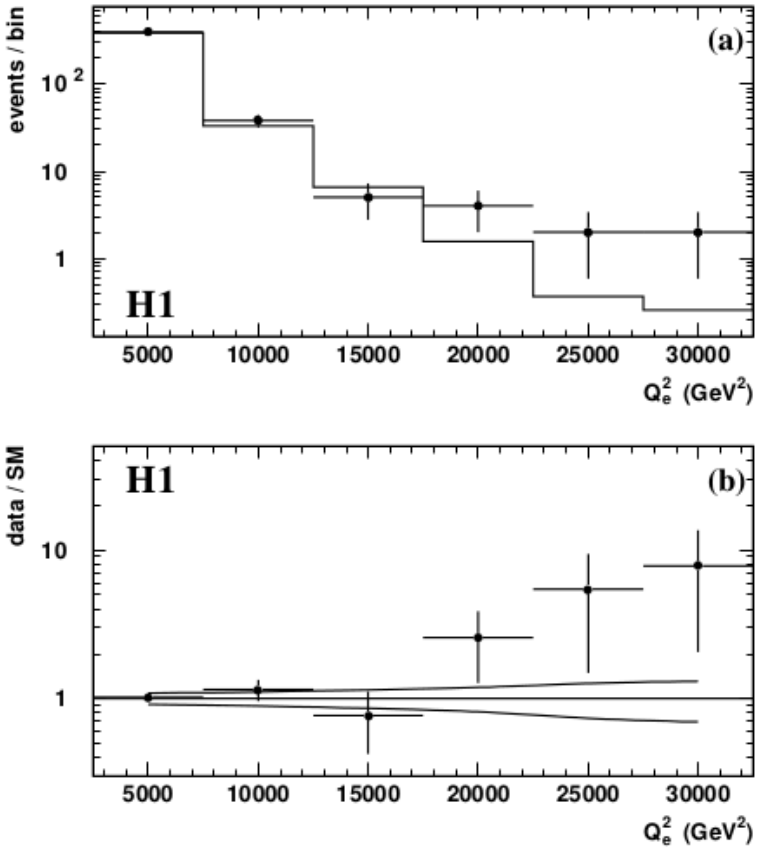


FIGURE 1.5: Observation of an excess of high Q_e^2 events measured by the H1 collaboration, where the subscript e represents the "e-method" used to compute Q^2, using only information from the scattered electron [33]. The number of events is shown in (a), with the solid black circles representing data and the solid black line representing the neutral current deep inelastic scattering expectation. The ratio of data over expectation is shown in (b), with the lines above and below unity representing the $\pm 1\sigma$ levels calculated including both statistical and systematic errors of the expected values.

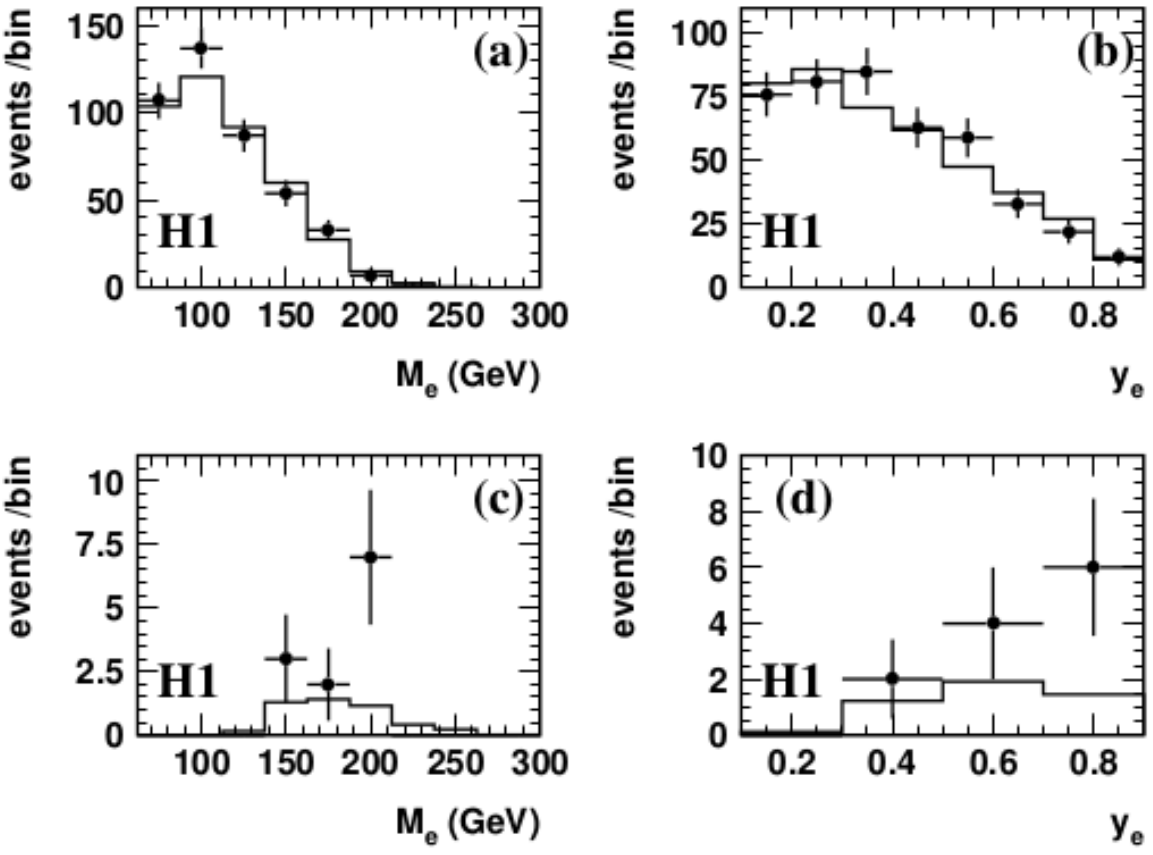


FIGURE 1.6: Observation of an excess in the M_e distribution at high values of Q_e^2. The subscript e represents the "e-method" used to compute Q^2, using only information from the scattered electron [33]. Distributions of M_e and y_e (where y_e is the the ratio of the electron four momentum over the jet four-momentum) of events with $2500\,\text{GeV} < Q_e^2 < 15000\text{GeV}$ are shown in (a) and (b), respectively. The same distributions for $Q_e^2 > 15000\,\text{GeV}$ are shown in (c) and (d).

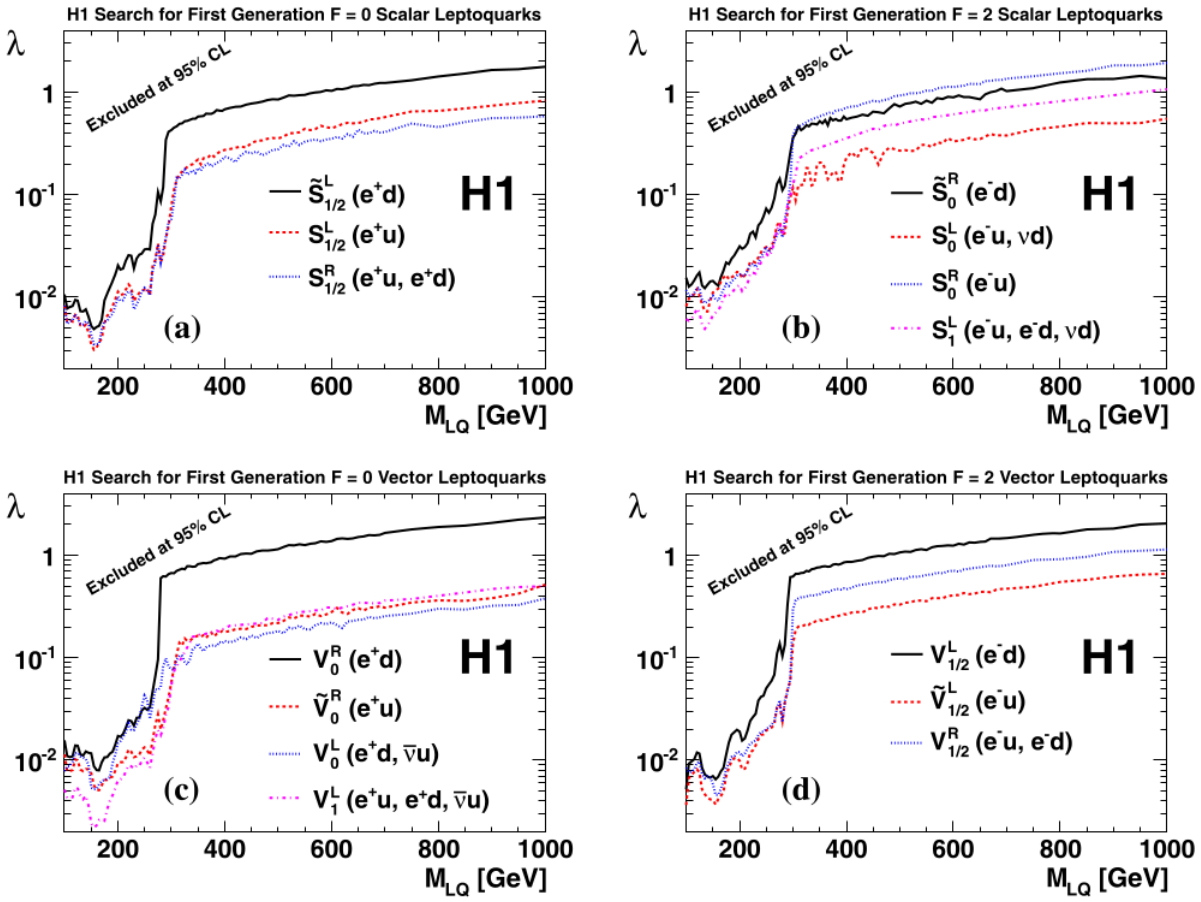


FIGURE 1.7: Exclusion limits produced by the H1 collaboration, for all 14 LQ types as described by the LQ model in Section . The limits are expressed on λ as a function of LQ mass, for scalar LQs with F = 0 (a) and F = 2 (b) as well as for vector LQs with F = 0 (c) and F = 2 (d). The domains above the curves are excluded at 95% confidence level.

Chapter 2

The Experimental Apparatus

This chapter describes the Large Hadron Collider (LHC) [37] and the Compact Muon Solenoid (CMS) [38] detector which records data from pp and heavy ion collisions and is located at one of the main interaction points along the LHC. The CMS detector is a general-purpose detector that recorded data from 2010 to 2012 corresponding to 36 pb^{-1} and 5 fb^{-1} at $\sqrt{7}$ TeV and 19.6 fb^{-1} at $\sqrt{8}$ TeV. Section 2.1 contains details on the design and specifications of the LHC and section 2.2 describes the CMS detector, its sub-systems, and a number of upgrades (both completed and planned) that it has undergone.

2.1 The Large Hadron Collider

The LHC is a two-ring-superconducting-hadron accelerator and collider that sits in the 26.7 km tunnel that was constructed for and housed the Large Electron Positron collider (LEP). It is capable of delivering pp, lead-lead, and proton-lead collisions. This section will focus on pp collisions.

2.1.1 Design

In 1994 the proposal to the CERN council to construct the LHC in the existing LEP tunnel was approved. With the Superconducting Super Collider (SSC) canceled just the year before due to concerns related to rising costs, the ability to use a preexisting tunnel was a significant motivation for construction of the LHC to

be located there. As a result the layout of the LHC was strongly influenced by geometry of the LEP tunnel.

Before reaching the full collision energies, particles in the LHC must go through a series of accelerators in which they are successively brought to higher and higher energies. This system, composed of an initial linear accelerator and a number of synchrotrons, is called the injection chain and consists of the following systems:

- The linear accelerators (LINACs) are the first step in the chain. The LINAC2 accelerator generates the protons at an energy of 50 MeV for injection in to the smallest synchrotron.
- The Proton Synchrotron Booster (PSB) prepares the protons for injection in to the next step, the Proton Synchrotron (PS). In it the protons are accelerated to 1.4 GeV.
- The PS then raises the proton energy to 26 GeV.
- The last step before injection in to the LHC itself, the Super Proton Synchrotron (SPS) accelerates the proton beams to 450 GeV.
- Finally, once in the LHC, the proton beams are accelerated to the energies at which they will collide. In 2010 and 2011 the operating energy here was 3.5 TeV, in 2012 the energy per beam was increased to 4 TeV. Currently, in 2015, the proton beams circulating in the LHC each possess an energy of 6.5 TeV.

These accelerators are depicted in Figure 2.1 with their approximate relative sizes and locations.

Once accelerated to these energies the beams are made to collide at four points along the LHC ring. The LHC has eight arcs and eight straight sections. Each straight section is 528 m long and can serve as an experimental or insertion point. Point 1, in the center of the first "octant", is the location of the ATLAS (A Toroidal LHC Apparatus) experiment, while CMS is located across the LHC ring at Point 5. Points 2 and 8 are the locations of the ALICE (A Large Ion Collider Experiment) and LHCb (LHC beauty) experiments, as well as the injection points for Beam 1 and Beam 2 (the main colliding beams used in the LHC). These 4 points are where the beams cross at interaction points (IPs), where the β function is low

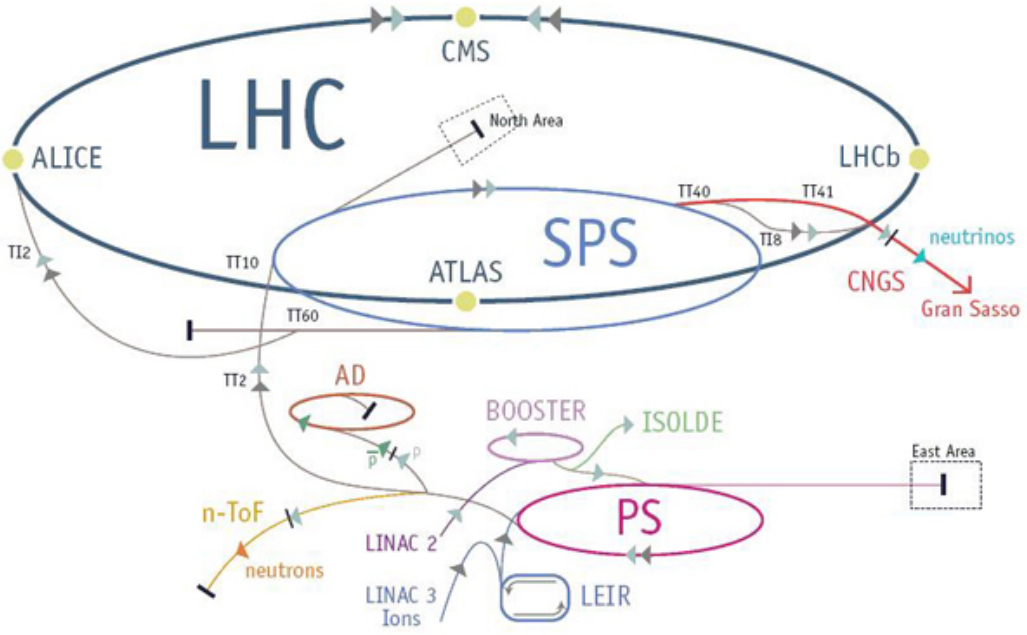


FIGURE 2.1: A schematic of the injection chains of the successive accelerators in to the LHC. The LINACs at the bottom of the diagram inject the protons or the lead ions which enter the LHC via the Booster, PS, and SPS. The approximate relative size of each system is shown, as well as relative locations.

and the beams are squeezed. The remaining 4 octants do not contain IPs, rather Points 3 and 7 contain collimations systems, Point 4 contains an RF system, and Point 6 is the location of the beam dump system. This consists of a combination of horizontally deflecting fast-pulsed magnets and vertically deflecting double steel septum magnets that serve to vertically extract both beams [37]. This layout is depicted in Figure 2.2.

Each of the eight arcs that compose the majority of the LHC's 26.7 km is composed of 23 cells, each 106.9 m long, making each arc approximately 2.5 km in length. Each arc cell itself is composed of two half cells each containing a cold mass (6.63 m long cryostat), a short straight section (SSS), and three dipole magnets 14.3 m in length. This comes to a total of 1104 main dipoles in use in the LHC ring straight sections and with 128 more in use in the straight regions the LHC contains 1232 dipole magnets in total. Each dipole is held to a temperature of 1.9 K in operation and provides a magnetic field (at the 7 TeV beam energy) of 8.33 T. At this operating level the current through the dipole magnets is 11.85 kA. The layout of the dipole magnets is given in Figure 2.3. The SSSs contain the main quadrupole magnets and a variety of magnets such as skew quadrupole correctors,

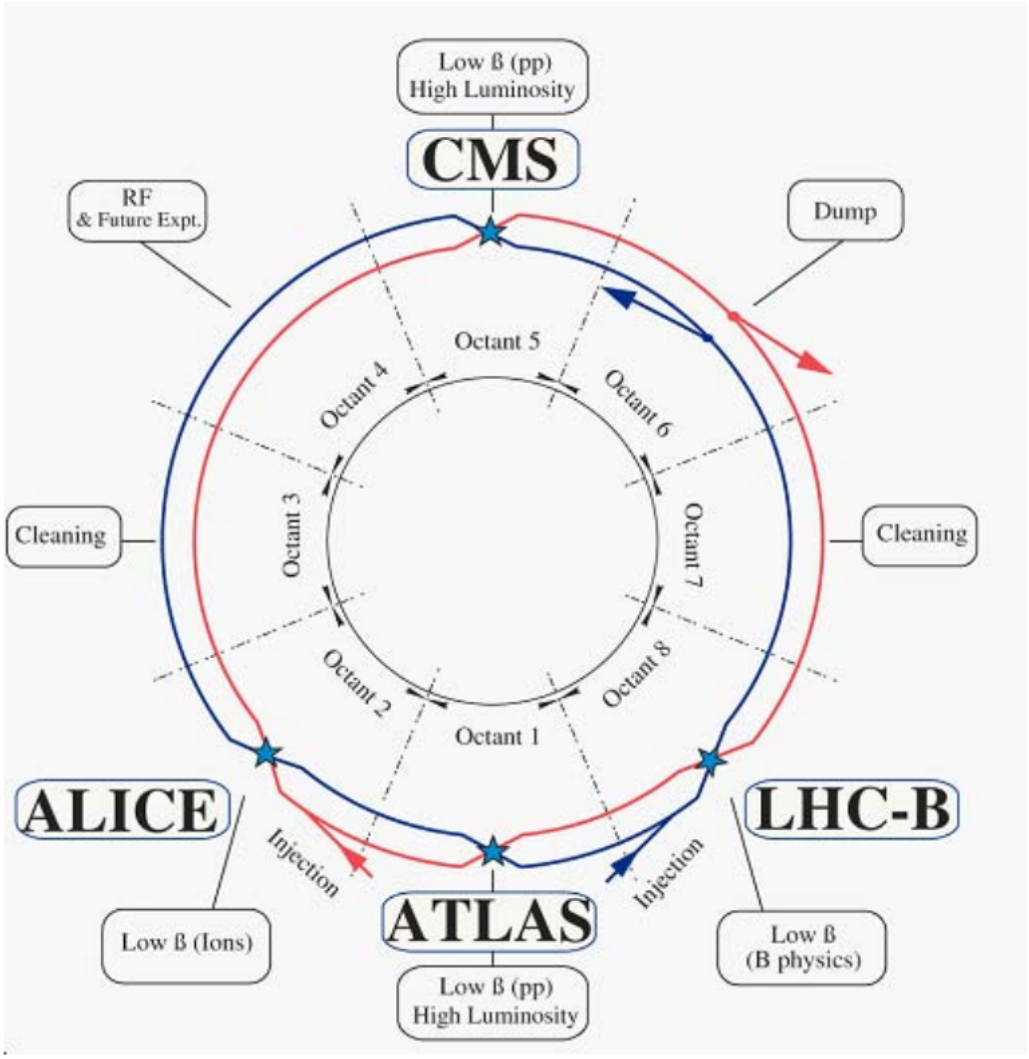


FIGURE 2.2: A schematic of the layout of the LHC. Points 1, 2, 5, and 8 are the locations of the 4 experiments conducted at the LHC [37].

sextupole-dipole correctors, tuning quadrupoles, and octupoles.

2.1.2 Specifications and parameters

Because the aim of the LHC is to reveal physics beyond the standard model, not only must the beam energy be high to reveal processes suppressed in nature due to kinematics, but also integrated luminosity is a limiting variable for statistically limited searches. The integrated luminosity over a period of time, T, is given by:

$$\mathcal{L} = \int \mathrm{L}dt \tag{2.1}$$

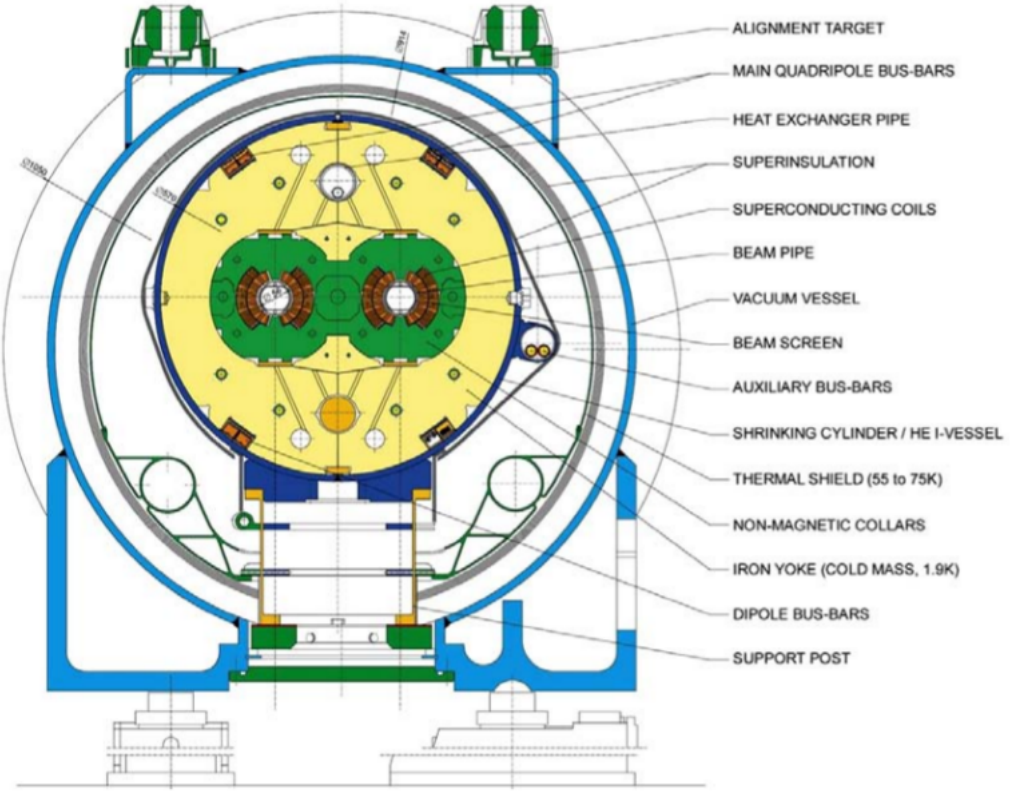


FIGURE 2.3: Diagram of a cross section of an LHC "cryodipole": the combined assembly housing the superconducting dipole magnets and the cold mass [37].

where L is the luminosity and is measured in number of collisions the LHC can produce per unit area per unit time (machine luminosity). Given a particular physics process of interest, P, the number of expected events for such a process over a period of time T is:

$$N_{event}(P) = \int_T \mathrm{L}\sigma_{event}(P)dt \tag{2.2}$$

where $\sigma_{event}(P)$ is the cross section for the event of the process in question and depends on the the energy in the event and the model in question. The instantaneous luminosity, L, depends only on the beam parameters and is given by the expression:

$$L = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi\epsilon_n \beta^*} F \tag{2.3}$$

where N_b is the number of particles per bunch, n_b is the number of bunches per beam, f_{rev} is the revolution frequency, γ_r is the relativistic gamma factor, ϵ_n is the normalized transverse beam emittance, β^* is the beam beta function at the collision point, and F is the geometric luminosity reduction factor due to the crossing angle at the IP. The value for F is given by the following expression:

$$F = \left(1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*}\right)^2\right)^{-1/2} \tag{2.4}$$

where θ_c is the full crossing angle at the IP, σ_z is the RMS bunch length, and σ^* is the transverse RMS beam size at the IP. This assumes round beams, with $\sigma_z \ll \beta$, as well as assuming equal beam parameters for both beams.

The discovery of rare processes at the LHC relies upon high energies, high instantaneous luminosities, and significant run times. The maximum particle density per bunch is limited by nonlinear beam-beam interactions that occur when bunches from the two beams collide with each other. This beam-beam interaction is measured by the linear tune shift which is given by:

$$\xi = \frac{N_b r_p}{4\pi\epsilon_n} \tag{2.5}$$

where r_p is the classical proton radius, $r_p = e^2/(4\pi\epsilon_0 m_p c^2)$. Previous experience with hadron colliders has indicated that the total linear tune shift including all IPs should not exceed 0.015 [37]. The LHC runs pp collisions in 3 experiments simultaneously, so parameters at the LHC must satisfy $\xi < 0.005$.

The LHC beam energy was 3.5 TeV per beam in 2010 and 2011, increasing to 4 TeV in 2012. The peak instantaneous luminosity during these run periods was a fraction of the nominal value, starting at 2% in 2010 and increasing to 35% in 2011 and 77% in 2012. The running parameters for 2012 and nominal design conditions are given in Table 2.1. Achieving 77% of nominal with half the nominal number of bunch was achieved partially due to the excellent beam quality delivered by the injectors, yielding an above-nominal number of protons per bunch. Challenges have arisen, unique to operating complex systems and electronics in the LHC environment. Occasional beam dumps can be caused by Unidentified Falling Objects (UFOs), and Single Event Effects (SEEs) caused by beam induced radiation to tunnel electronics was a significant source of inefficiency at the onset of the 2011 run [39].

As the 2015 run is ongoing, with the higher beam energy of 6.5 TeV, UFOs pose an increased challenge to LHC operation. Looking forward, a series of long shutdowns will prepare the LHC for the High Luminosity LHC program, for which CMS is preparing upgrades to handle and take advantage of the increased rate.

	Parameter	LHC Design Value	Achieved in 2012 run
N_b	Number of particles per bunch	1.15×10^{11}	$1.6 - .17 \times 10^{11}$
n_b	Number of bunches per beam	2808	1374
f_{rev}	Revolution frequency	11.25 kHz	11.25 kHz
γ_r	Relativistic gamma factor	7461	4263
ϵ_n	Transverse beam emittance	3.75μm	2.5μm
β^*	Beta function at IP	0.55	0.6
θ_c	Crossing angle at IP	285μrad	290μrad
σ_z	RMS bunch length	7.55 cm	9 cm
σ^*	Transverse RMS beam size at IP	16.6μm	19μm
L	Peak instantaneous luminosity	1×10^{34}	7.7×10^{33}
	Bunch spacing	25 ns	50 ns
	Stored beam energy	362 MJ	140 MJ
	Beam energy	7 TeV	4 TeV

TABLE 2.1: The targeted design values for the running parameters of the LHC and the values achieved in the 2012 run [39].

2.2 The Compact Muon Solenoid

The IP at Point 5 is the location of the Compact Muon Solenoid (CMS), a general purpose detector used for a variety of higgs and exotic searches and SM measurements. In 2012, an integrated luminosity of 21.79 fb^{-1} was recorded by the CMS detector, the integrated luminosity and peak daily luminosities and shown in Figure 2.4. As of the end of 2015 it has recorded a luminosity of 3.81 fb^{-1}, shown in Figure 2.5.This section describes the CMS detector and its subsystems. Section 2.2.7.3 goes in to greater detail regarding the electronics used in the endcap muon subsystem as the author of this thesis was responsible for work on upgrading a component of those systems.

2.2.1 Overview of the Detector

The CMS detector is located at Point 5 on the LHC ring, approximately 100 m underground near the French village of Cessy, between Lake Geneva and the Jura

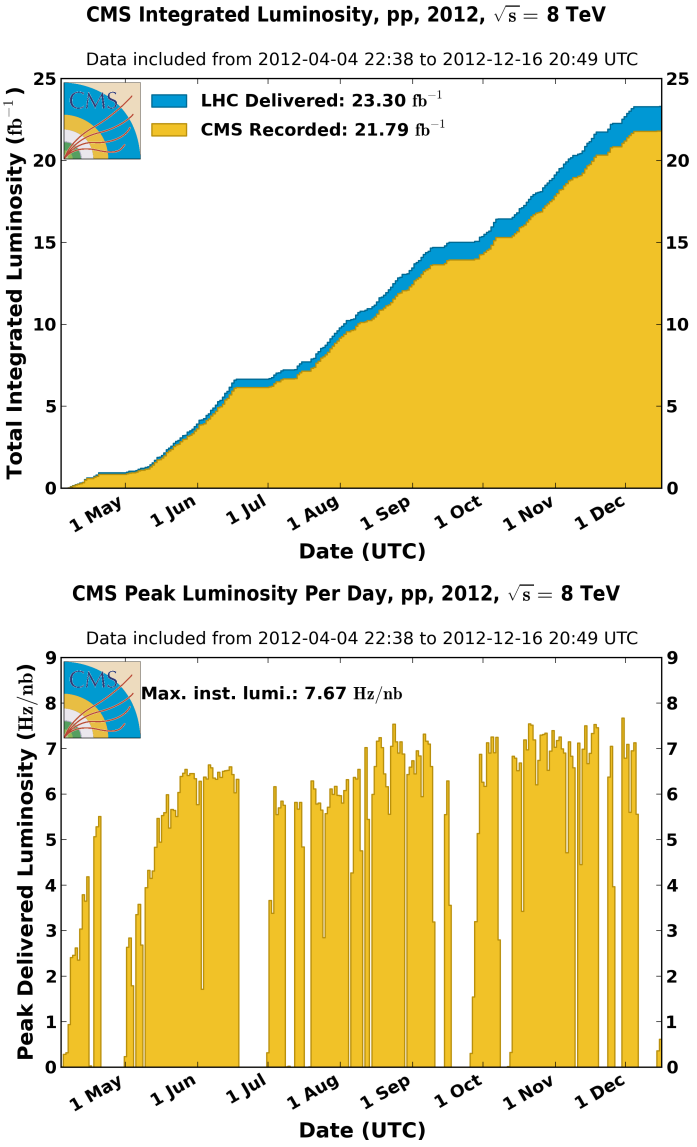


FIGURE 2.4: The integrated luminosity (top) delivered to and recorded by the CMS detector and the peak instantaneous luminosity (bottom) in 2012.

mountain range. It is designed to operate with pp collisions at $\sqrt{s} = 14$ TeV and a design luminosity of 10^{34} cm^{-2}s^{-1}. With an expected proton-proton cross section of 100 mb at that design energy the event rate will be approximately 10^9 inelastic events/s. Computing limitations allow for approximately 100 events/s for storage and subsequent analysis so the online selection process (called the "trigger", discussed in Section 2.2.8) must reduce this very large rate.

The high density of protons (within one bunch) and the short time between bunches, 25ns in 2015 and 50ns in 2012, result in the phenomeon called pileup in which multiple inelastic collisions are present in a single event. The average number of collisions per bunch crossing measured in 2012 was 21, and in 2015 this

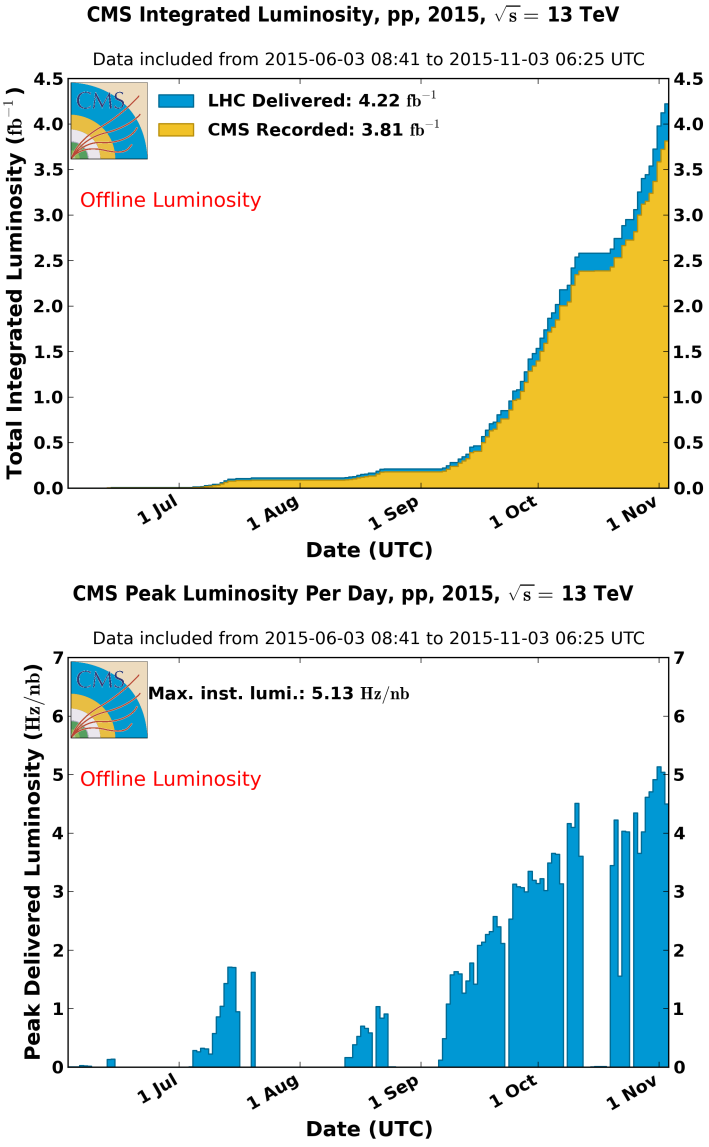


FIGURE 2.5: The integrated luminosity (top) delivered to and recorded by the CMS detector and the peak instantaneous luminosity (bottom) in 2015 by December 16th.

value dropped to approximately 15. This so called pileup-distributions for 2012 is shown in Figure 2.6.

This results in a significant challenge to avoid confusing products from different interactions in the same bunch crossing, especially when response times from sub-detectors can exceed the time between bunch crossings. This effect is reduced in the CMS detector through the use of high-granulaty subdetectors with good time resolution that can achieve low occupancy. This necessitates a very large number of detector channels, resulting in millions of total detector electronic channels that require good synchronization. This provides the ability to determine which

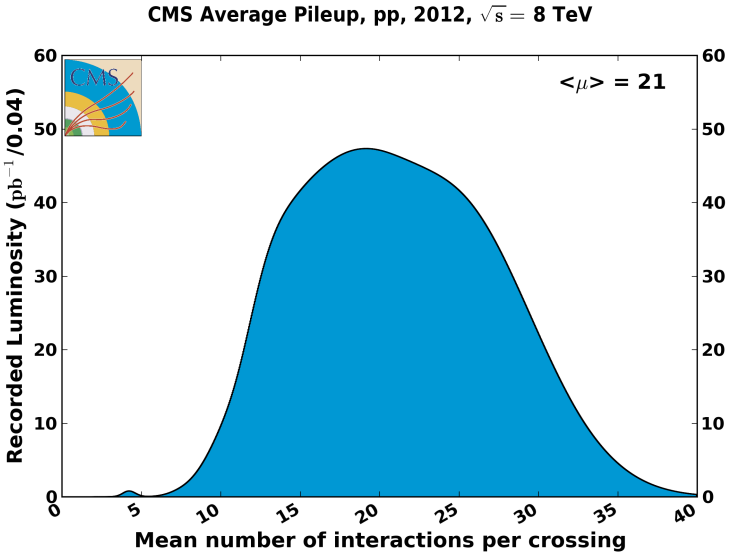


FIGURE 2.6: The distribution of the number of interactions per event measured in 2012 (top).

particles correspond to which event and which collision vertex within the event, and thus correspondingly the ability to accurately reconstruct the charges and momenta of the particles.

The overall detector requirements for CMS are as follows:

- Muon systems:
  - Good muon identification and momentum resolution for a wide range of momenta and angles
  - Dimuon mass resolution of approximately one percent at 100 GeV (near the mass of the Z boson)
  - Near perfect charge determination for muons with momenta less than 1 TeV
- Electromagnetic calorimeter:
  - Good electromagnetic energy resolution
  - Diphoton and Dielectron mass resolution of approximately one percent at 100 GeV
  - Rejection of π^0 decays
  - Efficient photon and lepton isolation at high luminosities

  - Wide geometric coverage

- Inner Tracker:
  - Good charged-particle momentum resolution and reconstruction efficiency
  - Efficient triggering and offline tagging for τ leptons and jets from b-quark decays

- Hadronic calorimeter:
  - Good missing-transverse-energy and dijet-mass resolution
  - Large and hermetic geometric coverage and fine lateral segmentation (enabling the above point)

The layout of the subsystems of CMS is shown in Figure 2.7. Each is described in detail in sections 2.2.3 through 2.2.7.

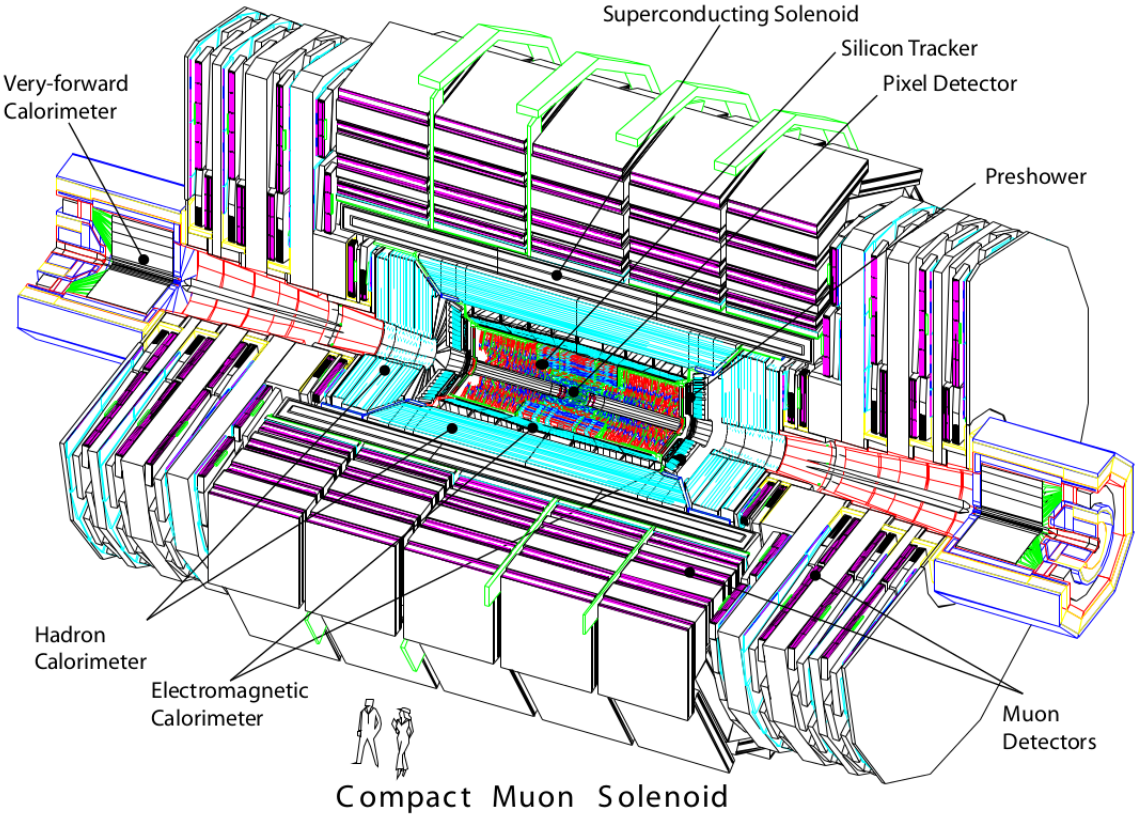


FIGURE 2.7: A cutout diagram of the subsystems of the CMS detector, to scale. The locations of the Solenoid magnet, Calorimeters, Trackers, and Muon Detectors are shown. Figures of two people standing by the detector are given for a sense of size.

2.2.2 Conventions and Coordinates

The convention for describing the geometry of the CMS detector is a right-handed coordinate system, defined with the origin at the nominal IP. The z-axis points along the beam axis towards the Jura mountains from Point 5. The y-axis points vertically upward and the x-axis points radially inward to the center of the LHC circle. Thus the polar angle θ is taken from the z-axis and the azimuthal angle ϕ is measured from the x-axis in the x-y plane.

Traditionally in collider physics the polar angle θ is not used, rather the psuedorapidity, η, is used. Pseudorapidity is an approximation of the quantity called rapidity, y. Unlike speeds, at relativistic velocities rapidity is an additive quantity and the difference in rapidity between two particles ejected by a collision is invariant under Lorentz transfomations along the z-axis. Also, production of particles tends to be close to uniform as a function of y. The rapidity, y, is related to the Lorentz γ-factor as well as energy and momentum of a particle and is given by the expression:

$$y = \cosh^{-1}(\gamma) = \tanh^{-1}\left(\frac{|\mathbf{p}|c}{E}\right) = \frac{1}{2}\ln\left(\frac{E+|\mathbf{p}|c}{E-|\mathbf{p}|c}\right). \tag{2.6}$$

Typically in colliders a version of the rapidity, y, is used that is defined only on the z-axis with respect to longitudinal momentum:

$$y \equiv \frac{1}{2}\ln\left(\frac{E+p_L c}{E-p_L c}\right). \tag{2.7}$$

In the limit where the mass of the particle is negligible or where the particle is travelling close to the speed of light (applicable in the case of particles coming from collisions in the LHC) the rapidity is approximated by a simple function of the polar angle, θ. This value is called the psuedorapidity, η, and is given by the expression:

$$\eta \equiv -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]. \tag{2.8}$$

Both precision measurements and searches performed at the CMS experiment use kinematics defined in the transverse plane. This is due to the fact that momentum in that plane can be determined to a high degree of accuracy and momentum conservation can be applied. Also, like η, the transverse momentum (p_T) is invariant under longitudinal Lorentz transformations. Thus p_T and η are the main kinematic variables used by most searches, including the search discussed in this thesis.

2.2.3 The Superconducting Solenoid

The superconducting magnet is designed to reach a field of 4 T in a free bore diameter of 6 m and length of 12.5 m, with a stored energy of 2.6 GJ at the full current. The flux is returned via a 10,000 ton iron yoke which is comprised of five wheels and two endcaps composed of three disks each. The solenoid itself has a 6.3 m cold bore and weighs 220 tons. To achieve a 4 T field a very large number of ampere-turns is required (41.7 MA-turns) and the winding of the solenoid is composed of four layers. This winding is composed of a stabilised and reinforced niobium-titanium (NbTi) conductor.

The ratio between the stored energy and cold mass of 220 t is very high, 11.6 KJ/kg, which results in a very significant mechanical deformation of 0.15% during energizing. These values are much higher than other solenoidal detector magnets, a comparison of the stored energy and energy-over-mass ratios of the CMS magnet and other detector magnets is given in Figure 2.8. The main parameters of the CMS magnet are listed in Table 2.2.

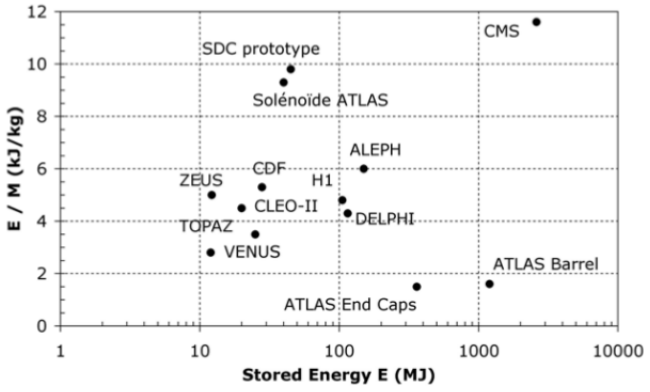


FIGURE 2.8: A comparison of the stored energy and energy-over-mass ratio E/M for CMS and other detector magnets [38].

General parameters	
Magnetic length	12.5 m
Cold bore diameter	6.3 m
Central magnetic induction	4 T
Total Ampere-turns	41.7 MA-turns
Nominal current	19.14 kA
Inductance	14.2 H
Stored energy	2.6 GJ
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 2.2: The design parameters of the CMS magnet [38].

The purpose of the magnet is to provide a field that bends muon paths, allowing for precise momentum measurements. In order to increase the longevity of the magnet, the decision to run it with a field of 3.8 T was made, which degraded the achievable muon momentum resolution by approximately 5% [40]. The levels of the magnetic field and structure of the field lines can be seen in Figure 2.9, the return of the field the iron yoke is clearly visible.

2.2.4 The Silicon Tracker

The innermost layer of the CMS detector is the inner tracking system, which is composed of an inner pixel tracker with 1440 modules and an outer strip tracker containing 15,148 strip modules. With a total of approximately 200 m^2 of active silicon area the CMS inner tracker is the largest silicon tracker ever built [38]. At the LHC design luminosity there is an average of approximately 1000 particles

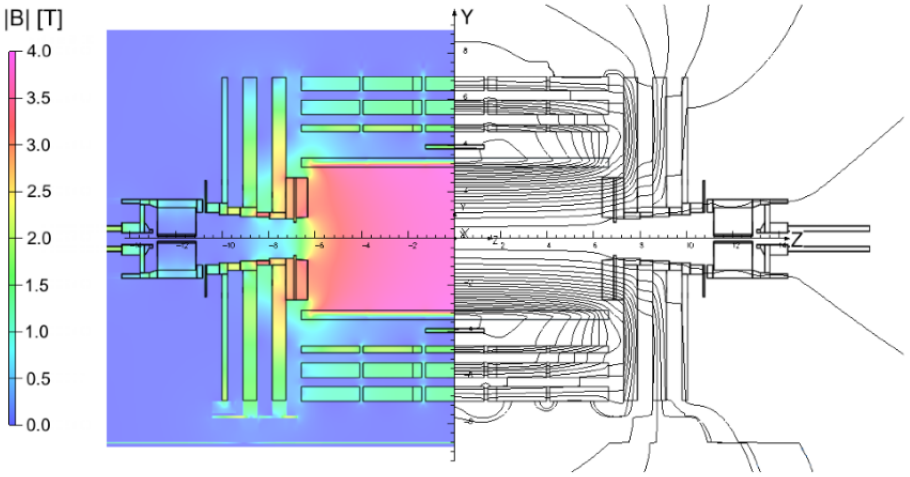


FIGURE 2.9: A detailed map of the $|B|$ field (left) and the field lines (right) for a longitudinal section of CMS at a magnetic flux density of 3.8 T. Each field line is an increment in magnetic flux of 6 Wb. [40].

expected per bunch crossing traversing the tracking system, from approximately 20 overlapping proton-proton interactions. At a radius of 4 cm, the location of the innermost layer of the pixel tracker, this leads to a hit rate density of 1 MHz/mm^2. In order to meet the resolution requirements the pixel size is 100×150 μm^2 in $r - \phi$ and z, respectively, which results in an occupancy of approximately 10^{-4} per pixel per LHC bunch crossing. At the intermediate radii where larger strips (10cm $\times$ 80μm) are located the occupancy is 2-3% per strip and at the outer radii the cell sizes increase to 25cm$\times$180μm and maintain an occupancy of apprixmately 1%.

The layout of the CMS tracker is given in Figure 2.10 and is composed of 4 major component systems. The innermost system is the pixel detector (PIXEL), composed of three cylindrical layers in the barrel (BPix) located at radii of 4.4, 7.3, and 10.2 cm and two endcap disks (FPix) located at $z = \pm 34.5$ and $z = \pm 46.5$ cm. In total the PIXEL system covers an area of approximately 1 m^2 and is composed of 66 million pixels. The pixel detector covers a pseudorapidity range of $|\eta| < 2.5$, matching the acceptance of the central tracker. The layout of the pixel detector and its hit coverage is given in Figure 2.11.

The other three systems compose the strip tracker and are located at radii from 20 cm to 116 cm. The Tracker Inner Barrel and Inner Disks (TIB/TID) are composed of 4 barrel layers supplemented by 3 encdcap disks at each end, covering a radius up to 55 cm. Thus the TIB/TID delivers up to 4 $r - \phi$ measurements per trajectory. It uses 320 μm thick silicon micro-strip sensors oriented parallel to the

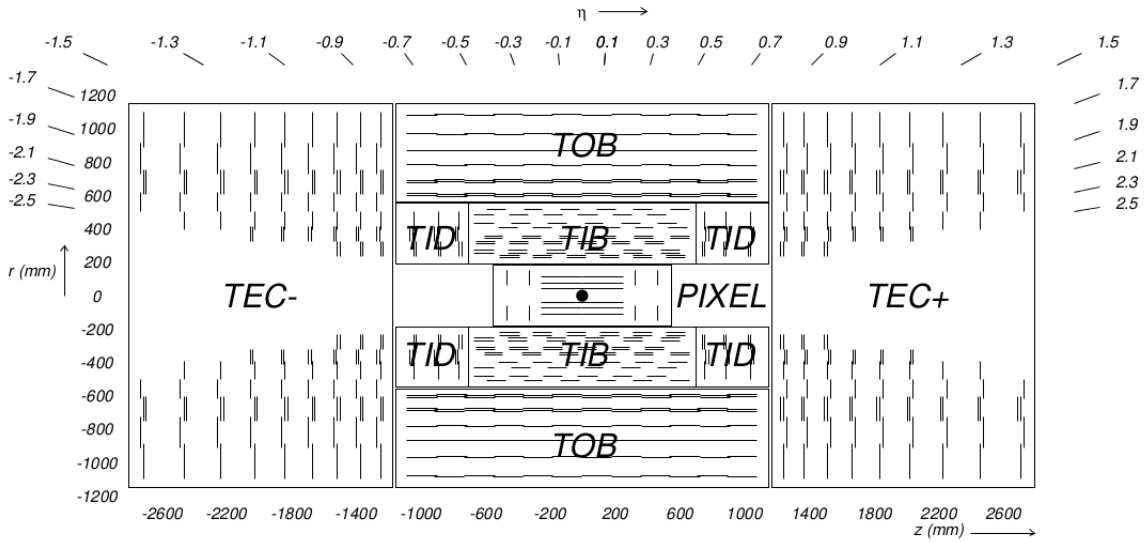


FIGURE 2.10: The longitudinal cross section of the layout of the CMS tracker [38]. Each line represents a module, with double lines indicating back-to-back modules that deliver stereo hits.

beam axis in the barrel and perpendicular to it in the endcaps. The strip pitch (inter-strip distance) is 80 μm on layers 1 and 2 and 120 μm on layers 3 and 4 of the TIB, varying from 100 μm to 141 μm in the TID. The strip pitch, width, and length are chosen to minimize inter-strip capacitance which in turn optimizes the resolution and occupancy as well as ensuring high voltage operational stability [41]. The TIB/TID is in turn surrounded by the Tracker Outer Barrel (TOB) which consists of 6 barrel layers of 500 μm thick sensors extending to a radius of 116 cm. The strip pitches are 183 μm on the first 4 layers and 122 μm on the last 2 layers. The TOB extends in z between $\pm$118 cm and beyond it lie the Tracker EndCaps (TEC). Each TEC consists of 9 disks covering 124cm $< |z| <$ 282cm and 22.5cm $< |z| <$ 113.5cm. There are 4 inner rings with 320 μm thick strips and 5 outer rings with 500 μm thick strips, averaging pitches ranging from 97 μm to 184 μm.

In addition to the modules listed above, the first two layers and rings of the TIB, TID, and TOB as well as rings 1, 2, and 5 of the TECs contain a second micro-strip module mounted back-to-back with a stereo angle of 100 mrad which provides a measurement in the second coordinate (z in the case of the barrel, r in the case of the disks). The strip tracker layout in total assures at least approximately 9 hits in the $|\eta| < 2.4$ range, 4 of which are two-dimensional measurements. The complete coverage of the strip tracker ends at $|\eta| = 2.5$, the overall number of measurement points in the strip tracker is given as a function of η in Figure 2.12.

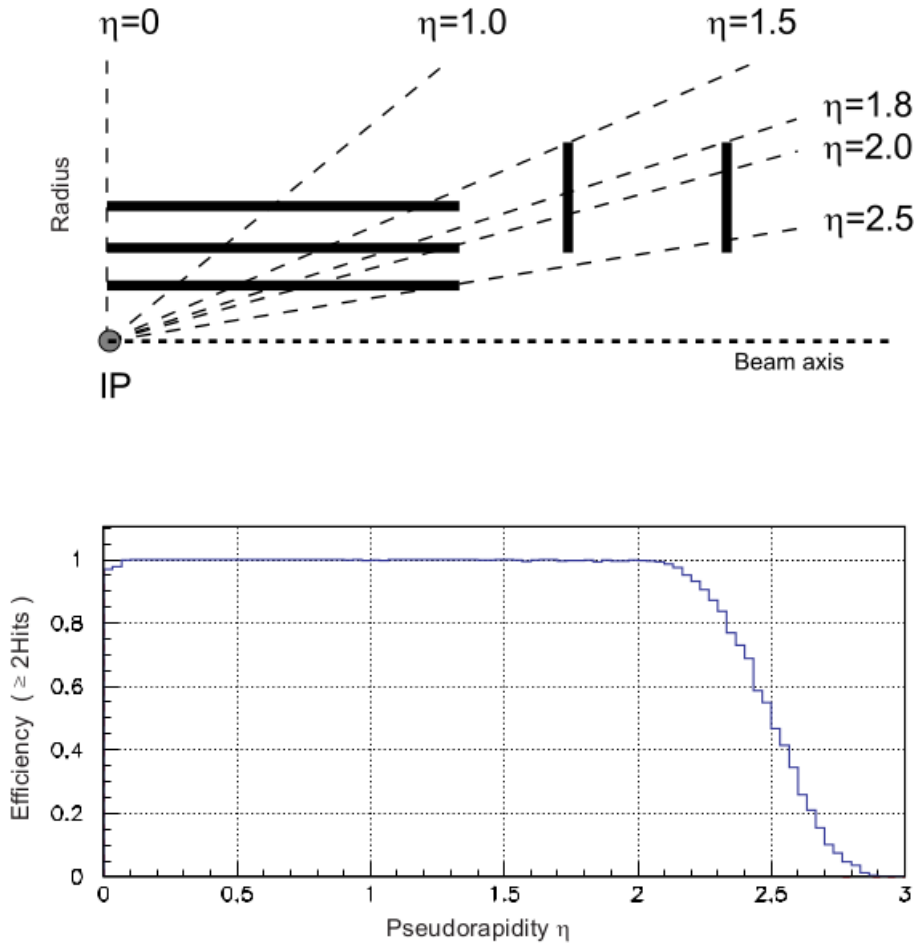


FIGURE 2.11: The geometrical layout of the pixel detector (top) and the hit coverage with respect to pseudorapidity (bottom) [38].

The pixel and strip detector modules performed very well in Run I (in 2011 and 2012); the hit efficiencies are shown in Figure 2.13. The tracker allows for accurate reconstruction of charged particle tracks as well as high-resolution measurements of the interaction vertices, which are both described in detail in Section 3.1.

2.2.5 The Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) immediately surrounds the silicon tracker. The ECAL is a hermetic and homogeneous calorimeter which is composed of 75,848 lead tungstate ($PbWO_4$) crystals in total, with 61,200 located in the central barrel section and 7,324 in each endcap. A preshower detector is located in front of the endcap crystals, which serves to identify neutral pions in the endcaps and help electron identification against minimum ionizing particles [42].

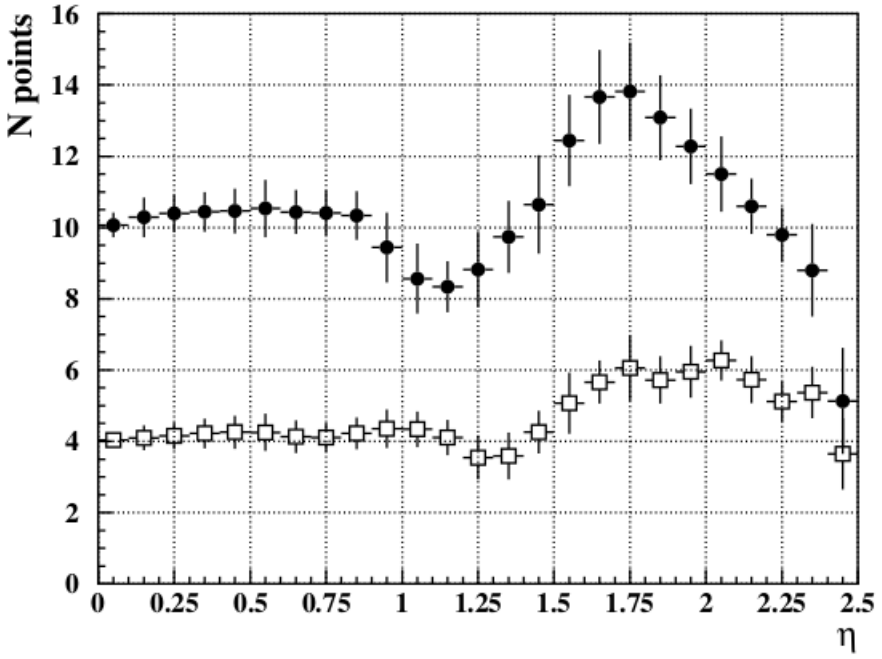


FIGURE 2.12: The average number of track hits in the strip tracker with respect to η, black circles represent the total number while white squares represent the hits in stereo layers only [38].

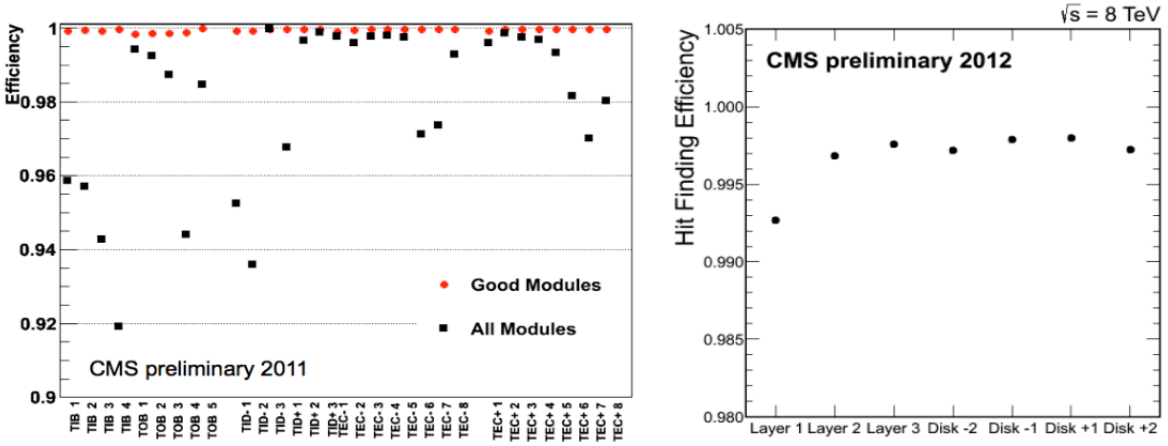


FIGURE 2.13: The single hit efficiency of the strip (left) and pixel (right) trackers [41]. Only the good (fully operational) modules were considered for the pixels.

The barrel portion of the ECAL (EB) covers a pseudorapidity range of $|\eta| < 1.479$ with an inner radius of 1.29 m. The granularity is 360-fold in the ϕ-direction, and (2×85)-fold in η, resulting in the total of 61,200 individual crystals. The crystals themselves are tapered in shape, slightly varying with position in η. The endcap portion (EE) covers a pseudorapidity range of $1.479 < |\eta| < 3.0$, with a longitudinal distance from the inner surface to the interaction point of 315.4

cm. The preshower detector (ES) sits directly in front of each EB, covering a pseudorapidity range of $1.653 < |\eta| < 2.6$. The overall layout of the ECAL is given in Figure 2.14.

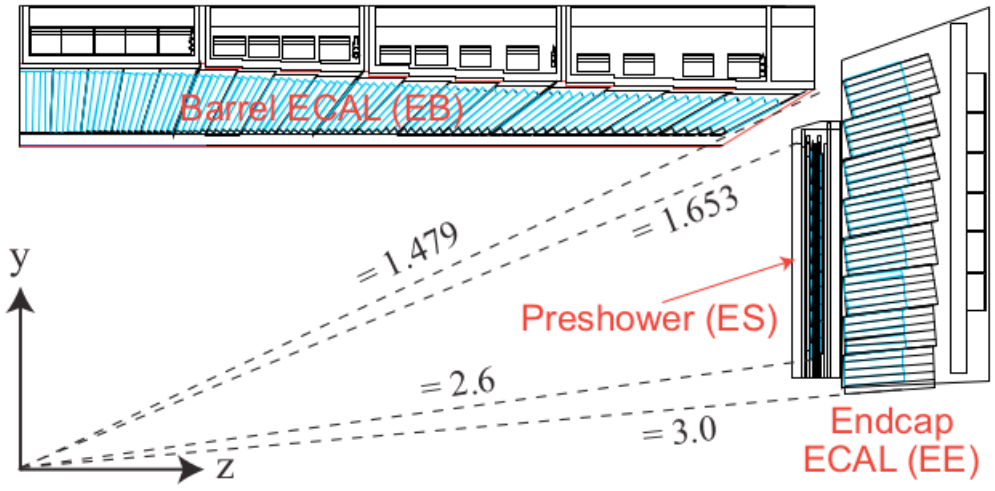


FIGURE 2.14: The layout of the ECAL, showing the Barrel ECAL (EB), Endcap ECAL (EE), and Preshower ECAL (ES) [42]. The dashed lines are labeled with pseudorapidity values for reference.

A driving criterion in the design of the ECAL was the ability to detect decays to two photons, a main decay mode of the Higgs boson. The characteristics of the $PbWO_4$ crystals made them an appropriate choice to help meet this criterion. The density (8.28 g/cm^3) and short radiation length of 0.89 cm (defined as the mean path length over which a relativistic particle loses energy by a factor of $1/e$) allows for a compact design. The small Molière radius of 2.2 cm (the radius of a cylinder transverse to a charged particle's direction of flight in which on average at least 90% of the energy of the particle is deposited) provides a fine granularity. Additionally, in the years leading up to the construction of the LHC, $PbWO_4$ crystal production improved, becoming capable of producing optically clear and radiation-hard crystals. The scintillation decay time of the $PbWO_4$ crystals in the ECAL is on the same order as the LHC design bunch crossing time. Approximately 80% of the light is emitted in 25 ns. At the design operation temperature of 18° C the light output is approximately 4.5 photoelectrons per MeV. The scintillated light is blue-green in color, with a broad maximum wavelength of 420-430 nm [38].

The scintillated light is collected by photodetectors located at the end of each crystal. In the barrel, avalanche photodiodes (APDs) that are specially produced for the CMS ECAL are used. Two are placed on the backs of each crystal module. In the endcap, the photodetectors used are vacuum phototriodes (VPTs), also

specially produced for the ECAL, with one per crystal module. ECAL modules of both the endcap and the barrel are shown with their attached photodetectors in Figure 2.15.

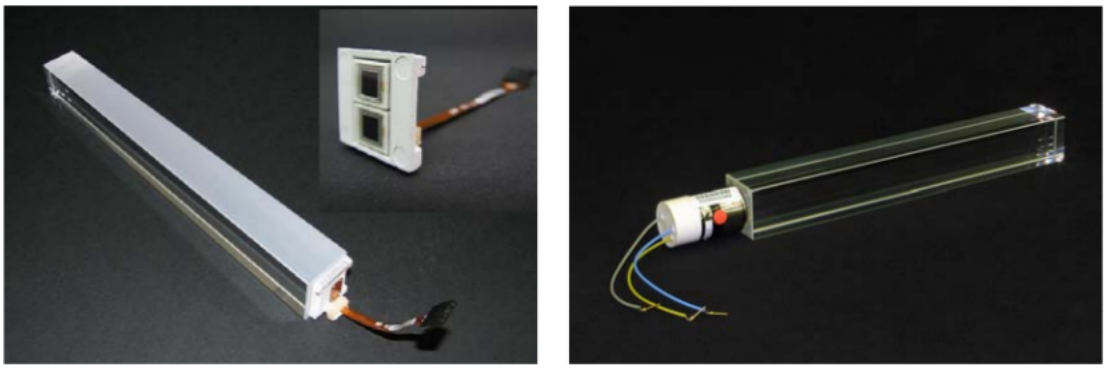


FIGURE 2.15: $PbWO_4$ crystals with their photodetectors attached [42]. A barrel crystal with attached APD (left), detached pair of APDs (left insert), and endcap crystal with VPT attached (right).

The APDs and VPTs are fast and radiation hard and are able to operate in a 4-T magnetic field. The $PbWO_4$ crystals output a relatively small amount of light, so both photodetector types must have strong amplification. The APDs operate at a gain of 50 and cover an active area of 25 mm^2. The VPTs, of which there are only one per crystal, operate at a mean gain of 10.2 in a zero field and cover an active area of approximately 280 mm^2. When the VPTs are placed in a strong axial magnetic field the response is slightly reduced, and there is a variation of response with the angle of the VPT axis with respect to the field. The mean response in a 4-T field is typically at least 90% of that in a zero magnetic field. Both the APDs and VPTs were tested and screened to ensure reliable operation for 10 years under the high luminosity LHC conditions.

2.2.6 The Hadronic Calorimeter

The hadronic calorimeter (HCAL) is located directly outside of the ECAL and its role is to measure the energies of hadron jets and to help infer the missing transverse energy in an event that may result from neutrinos and/or exotic particles. The HCAL is composed of 4 main components: the barrel (HB), the endcap (HE), the outer (HO), and the forward (HF) calorimeters. Each subcomponent is a sampling calorimeter using the well known calorimetry strategy of tile and wavelength shifting fibers [38]. The layout of these subsystems are given in Figure 2.16.

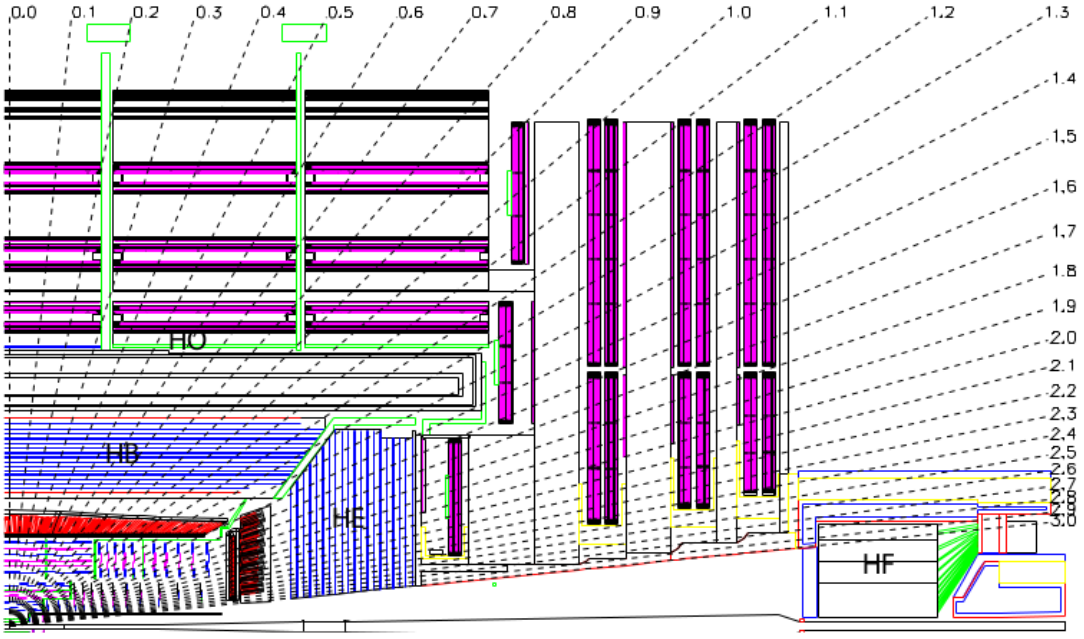


FIGURE 2.16: A longitudinal cross section of the layout of the CMS detector featuring the locations of the hadronic calorimeter: the barrel (HB), the endcap (HE), the outer HCAL (HO), and the forward HCAL (HF) [38].

The HB inner radius is restricted by the outer edge of the ECAL, $R = 1.77$ m, and the inner edge of the magnet coil, $R = 2.95$ m. It is composed of 36 identical azimuthal wedges which form each of two half barrels (HB+ and HB-), together covering the pseudorapidity range of $|\eta| < 1.3$. Each wedge is is segmented azimuthally is segmented in to 4 sectors and is composed of brass absorber plates that are bolted together in a staggered fashion as to produce a configuration which contains no projective dead material for the full extent of the wedge. For structural integrity the brass absorbers are reinforced by steel plates on the front and back of the wedges. The scintillator, which is made of plastic, is divided into 16 η sectors producing a segmentation of $(\Delta\eta, \Delta\phi) = (0.087, 0.087)$. The light from the scintillators is brought out by wavelength shifting fibers and after exiting the scintillator the fibers are spliced to clear fibers. These clear fibers bring the light to the hybrid photodiode for eventual data collecting.

The HE cover a large portion of the pseudorapidity range, $1.3 < |\eta| < 3$, which is a region containing approximately 34% of final state particles. Because the HE is inserted into the end of a high T solenoidal magnet, the absorber had to be composed of a non-magnetic material. In order to simultaneously satisfy that condition and others such as a maximum number of interaction lengths to contain

hadronic showers, C26000 cartridge brass was used [38]. As with the HB, the HE absorbers are staggered to avoid creation of any projective dead material. In addition, they are shaped to minimize any cracks between HB and HE, rather than being shaped to optimize single-particle energy resolution. This is due to the fact that jet resolution in the HE is limited by pileup, magnetic effects, and parton fragmentation in any case. Also as with the HB, the HE contain scintillators whose light is brought out via wavelenth shifting fibers to photodetectors and readout electronics.

Between the electronic and hadronic barrel calorimeters (EB and HB) there is not enough stopping power in the central pseudorapidity region to provide adequate containment for hadronic showers. Thus, to ensure that there is sufficent sampling depth for tha tregion, the HCAL is extended outside the solenoid with the HO, also called a tail catcher. The HO uses the solenoid coil as an extra absorber and is used to identify showers that start late and to measure any additional shower energy deposited after the HB. The HO is placed as the first sensitive layer in each of the five rings of the iron yoke (2.536 m wide along the z-axis). It is segmented in to 12 identical azimuthal sectors, closely matching the geometry of the barrel muon system. Together with the HE and the HB the HO is segmented in the r, z plane according to Figure 2.17.

The HF is located in a region that experiences huge particle fluxes. The front face is located 11.2 m from the interaction point, with an inner radius of 12.5 cm and an outer radius of 130 cm. It receives on average 760 GeV per proton-proton interaction as opposed to 100 GeV for the rest of the detector. To cope with this environment the HF is housed in a hermetic radiation shield which consists of layers of 40 cm thick steel, 40 cm of concrete, and 5 cm of polyethylene. Additionally, the active elements of the HF are radiation-hard quartz fibers. The calorimeter itself is subdivided into two segments longitudinally, one which runs the full depth of the detector and one which only starts at a depth of 22 cm from the front, which are read out separately. This allows the HF to distinguish between showers generated by electrons and photons, which deposit a majority of their energy in the first 22 cm, from hadronic showers which deposit their energy roughly equally throughout the calorimeter.

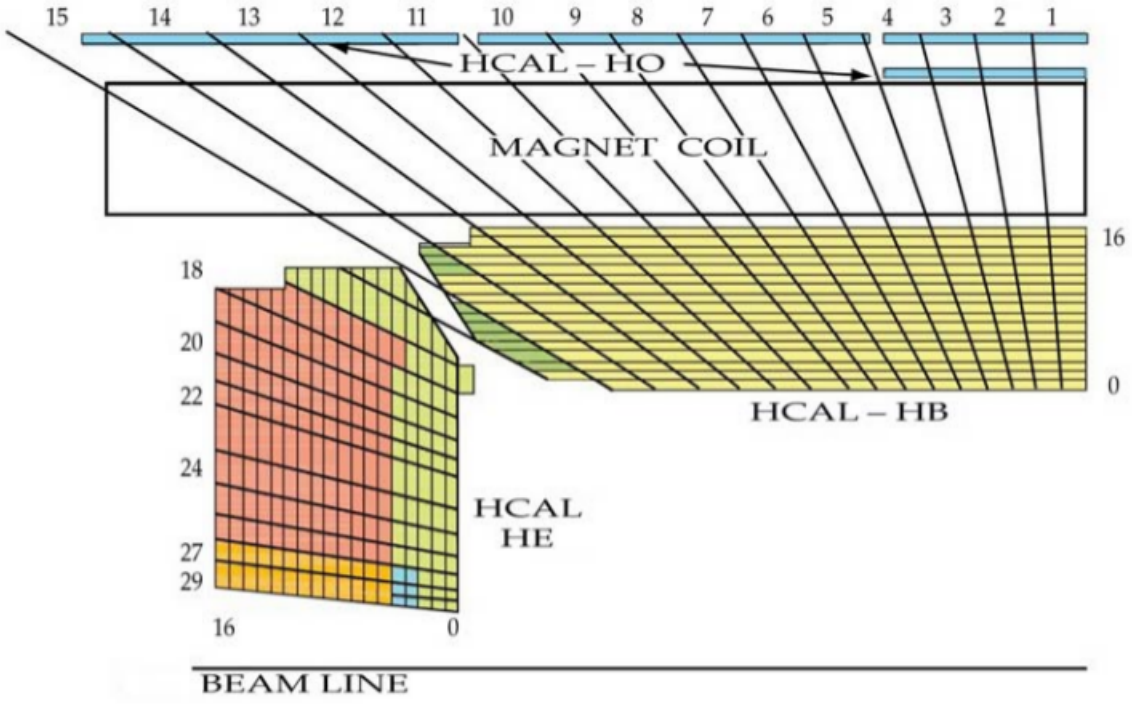


FIGURE 2.17: A cross section in the r, z plane of the segmentation of the HB, HE, and HO [38].

2.2.7 The Muon system

Precision measurement of muons is of critical importance, as many final state decays of exotic processes and signatures of supersymmetry include one or more muon muons, and the "gold plated" SM Higgs decay mode of $h \to ZZ \to 4\mu$ (as well as a few other Higgs decay modes) produces many muons in its decay states. Muon final states possess great discovery potential, as final states in which all leptons are muons are less affected by radiative losses in the tracker system. Thus, as indicated by the experiment's middle initial, muon detection is a main role that the CMS detector must perform well. The combined muon system has three functions: muon identification, momentum measurement, and triggering. Good muon momentum resolutions and trigger capability are facilitated by the high-field solenoid magnet and the flux-return yoke. The yoke also serves to absorb hadrons to assist with muon identification.

The muon system has the capability to reconstruct muon momentum and charge over the entire kinematic range of the LHC, and is composed of three separate types of gaseous particle detectors. As with the calorimeters, the muon system

is composed of a cylindrical barrel section and two planar endcap sections. The eventual background rate the muon system would experience was uncertain during construction, so a dedicated trigger system was added in both the barrel and endcap sections. This system, consisting of Resistive Plate Chambers (RPC) provides a fast and fully independent trigger and capable of tagging the time of ionizing events in less than 25ns, extends over a large pseudorapidity range, $|\eta| < 1.6$. In the barrel, where the overall muon rate and neutron-induced background is low, and the magnetic field is low as well, so drift chambers with rectangular drift cells are used. This system, the Drift Tubes (DT), covers a pseudorapidity range of $|\eta| < 1.2$. In the endcap regions the muon and background rates are high, and the magnetic field is large and non-uniform, so Cathode Strip Chambers (CSC) are used because of their fast response time, fine segmentation, and radation-hardness. The CSCs cover a pseudorapidity range of $0.9 < |\eta| < 2.4$. The layout of these three systems is shown in a quadrant cutout of the CMS detector in Figure 2.18.

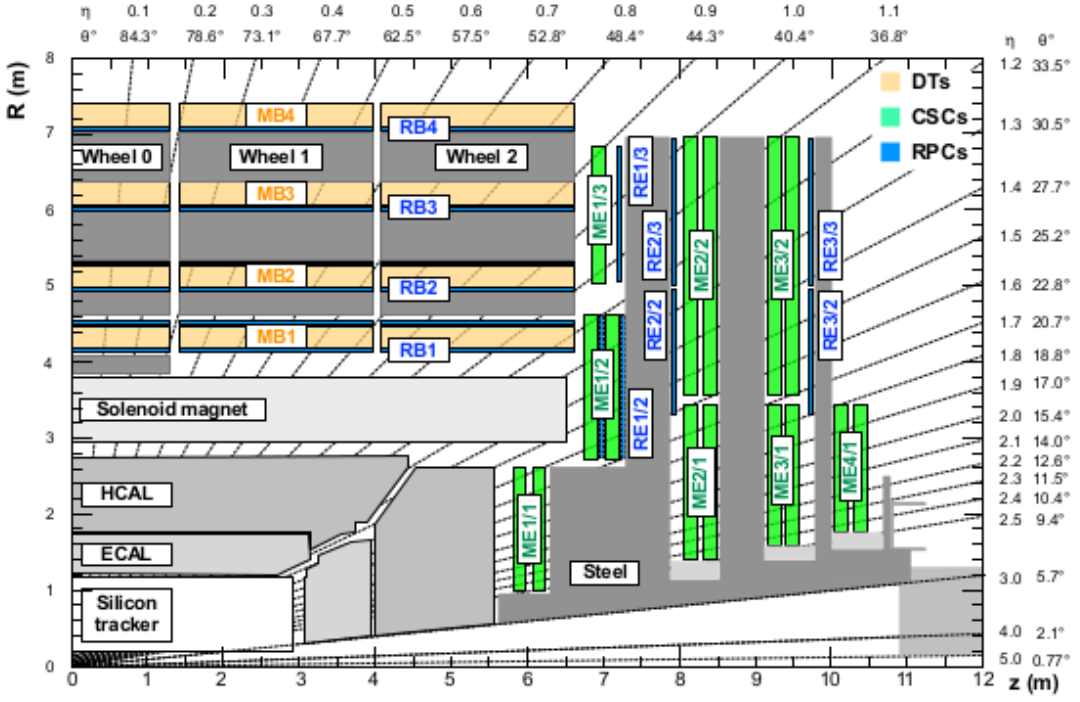


FIGURE 2.18: The cross section of a quadrant of the CMS detector, in the $R-z$ plane [40]. The muon systems are featured, the four DT stations (light orange) are labeled MB, the CSCs (green) are labeled ME, and the RPCs (blue) which are in both barrel and endcap are labeled both MB and ME, respectively. MB stands for "muon barrel" and ME for "muon endcap". The dark gray regions represent the steel disks.

2.2.7.1 The Resistive Plate Chamber system

The RPCs are gaseous parallel-plate detectors that provide a time resolution comparable to that of scintillators. An RPC is capable of identifying the time of an ionizing event in a significantly shorter time than the 25ns between two consecutive LHC bunch crossings. A muon trigger based on RPCs can identify the bunch crossing to which a muon track is associated unambiguously, and thus the RPCs are a detector system dedicated to triggering. The RPCs cover a pseudorapidity of $|\eta| < 1.6$ which covers the entire range of the DTs and a portion of the range of the CSCs.

The RPCs are double-gap (called "up" and "down" gaps) gaseous detectors, operated in avalanch mode with common pick-up read-out strips located between the two gaps. Each gap consists of a pair of 2 mm thick bakelite plates coated in a thin graphite layer encapsulating the gap which is filled by gas mixture of 95.2% Freon, 4.5% isobutane, and 0.3% sulphur hexaflouride. The plates are held to a voltage of 9.6 kV, and muons that traverse the gas ionize an atom in the gas, causing an avalanch which induces a charge that is readout by the on-board electronics.

The RPCs are partioned in the η-direction, in two and three partitions (called rolls) in the barrel and endcap, respectively. The layout of a typical barrel RPC is shown in Figure 2.19. The RPCs are arranged in stations following a similar sequence to the DTs and RPCs. The RPC barrel (RB) has 4 stations, RB1-4, while the RPC endcap (RE) has 3 stations, RE1-3, totalling 480 chambers overall.

2.2.7.2 The Drift Tube system

The DTs are located in the barrel, covering a pseudorapidity range of $|\eta < 1.2|$. In this region the muon rate is low, the neutron background is small (except for at the outermost layer of the DTs), and the magnetic field is predominantly uniform with a strength of 0.4 T and lower [40]. Thus drift chambers are used with rectangular cells and electrical field shaping implemented. There are four stations in the barrel, labeled MB1-4, which are divided into 12 ϕ-segments.

The DTs contain basic elements called drift cells, with a transverse area of $42 \times 13\text{mm}^2$ and a 50 μm diameter gold-plated anode wire at the center. A voltage of 3600 V is applied to the wire, and 4 electrodes are used (including 2 cathode

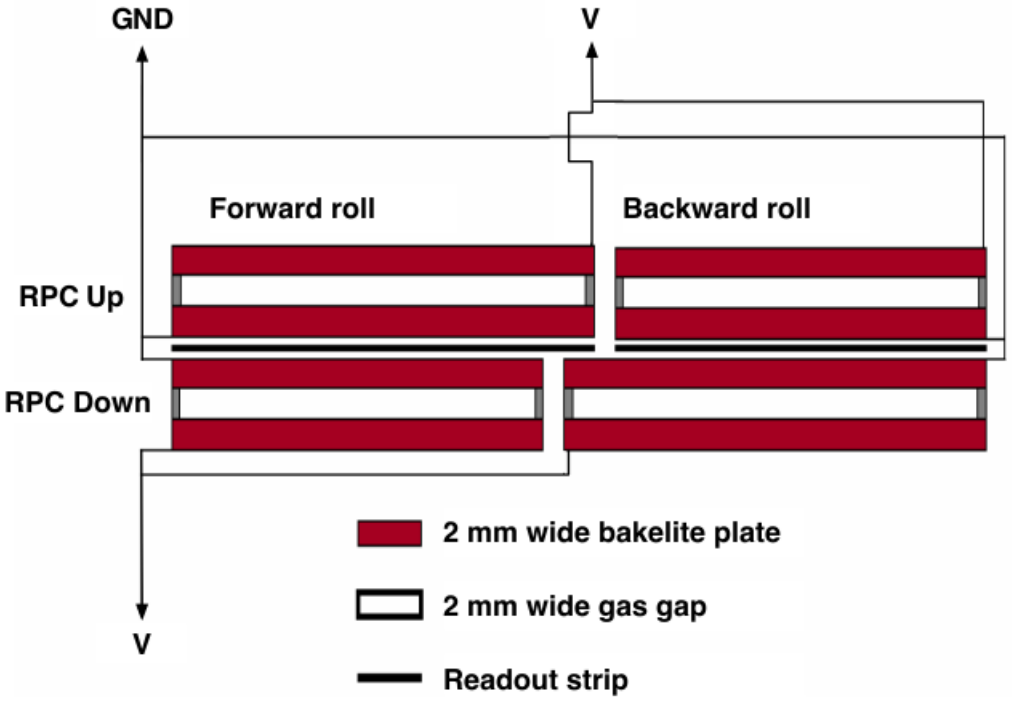


FIGURE 2.19: A schematic of a barrel RPC with two partitions (rolls) [40].

strips) to shape the drift field: 2 on the ground planes between layers and 2 on the side walls of the tube. These electrodes are held at 1800 and -1200 V, respectively. The layout of the drift cell with the drift paths and an example incident muon path is depicted in Figure 2.20.

The gas mixture which is ionized by incident muons is an 85%/15% blend of argon (Ar) and carbon dioxide (CO_2). This provides good quenching properties, as well as a saturated drift velocity of approximately 55 μm/ns and a maximum drift time of 400 ns. These drift cells are staggered, with four layers of parallel cells forming a superlayer (SL). Each DT chamber consists of 2 SLs that measure the $r-\phi$ coordinates and one orthogonal SL that measures the $r-z$ coordinate (except for the outermost layer of DTs, MB4, which only has an $r-\phi$ layer).

2.2.7.3 The Cathode Strip Chamber system

The CSCs cover a pseudorapidity range of $1.2 < |\eta| < 2.4$, and consist of a total of 473 chambers in the first LHC run. There are 108 in the first station (ME1), 54 in the second and third (ME2 and ME3), and 18 in the inner ring of the outermost

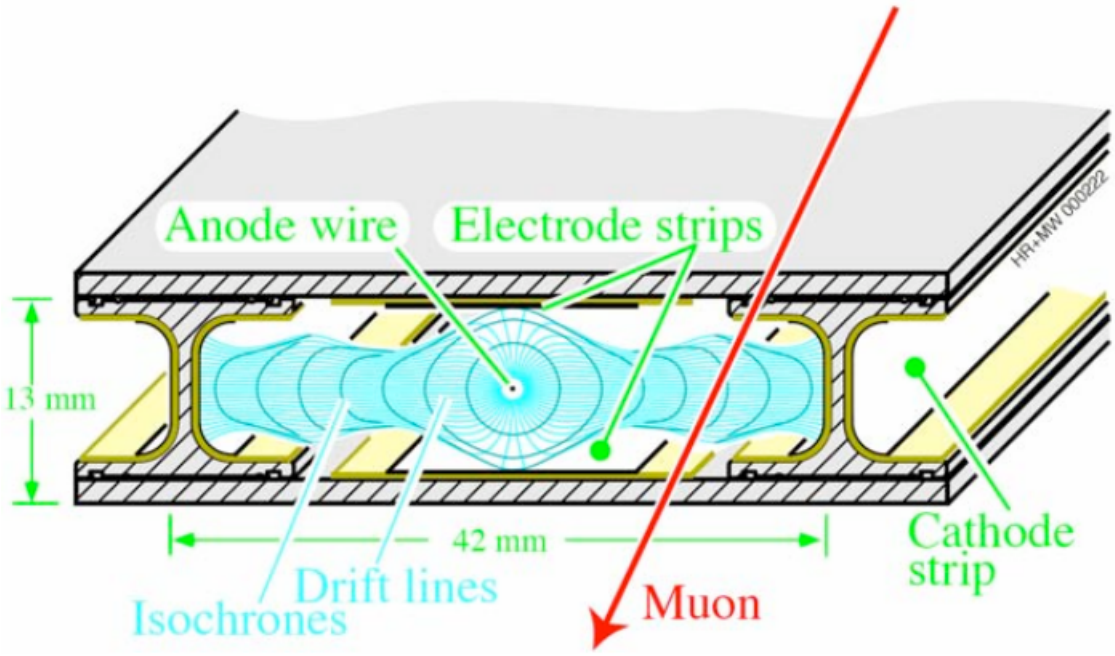


FIGURE 2.20: A schematic of a drift cell, shown with the drift lines and an example incident muon path [38].

station (ME4), as depicted in Figure 2.18. On one side of the CMS detector, 5 CSCs were placed in an outer ring on the 4th station (ME4/2), and these were expanded to the full 18 per endcap in the second LHC run.

CSCs are trapezoidal in shape and cover either 10° or 20° in ϕ. All except for ME1/3 overlap and provide continuous ϕ-coverage. They are comprised of 6 anode wire planes interspersed with 7 cathode panels. The wires run azimuthally and define the radial coordinate of a track. The strips are milled directly on to the cathode panels and run lengthwise at a constant width in $\Delta\phi$. While there are varying total sizes of chambers, the largest ones (ME2/2, ME3/2, and ME4/2) are approximately 3.4×1.5 m^2 in size. A sample schematic layout of a typical CSC is given in Figure 2.21.

The nominal gas mixture used in the CSCs for ionization is 40% Ar, 50% CO_2, and 10% CF_4. The primary role of the CO_2 is as a non-flammable quencher for achieving large gas gains, and the role of the CF_4 is to prevent polymerization on the anode wires and increase drift speed. The nominal operating voltage was chosen to be 3.6kV which corresponds to a gas gain on the order of 7×10^4, except

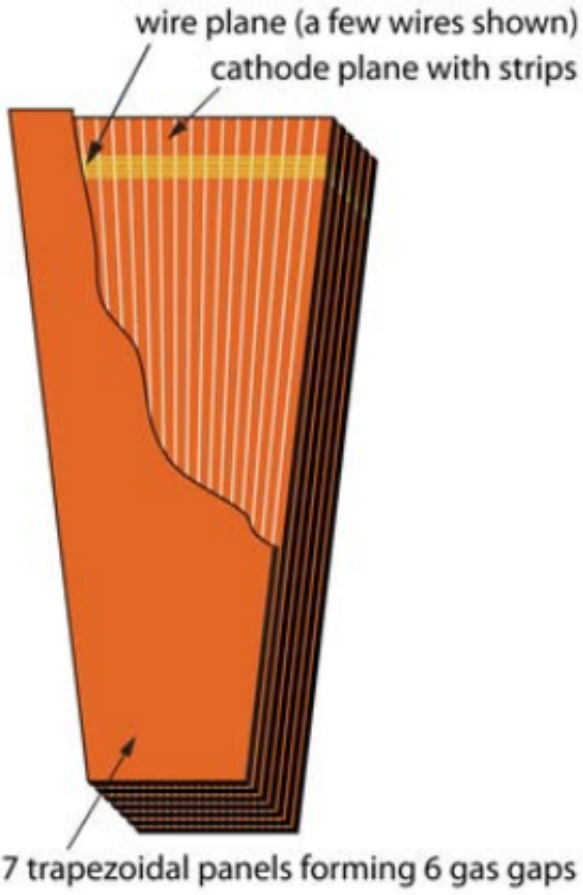


FIGURE 2.21: A schematic of a CSC, shown with all six layers [38].

for the inner stations (ME1), for which 3.3kV was chosen. These voltages provide very high efficiencies with an adequate signal-to-noise ratio.

Each chamber has a set of anode front end boards (AFEBs) that are amplifier-discriminators and serve to initially shape and read out charge avalanches from incident muons ionizing the gas. All the AFEBs send their output to an FPGA-based anode local charged track (ALCT) board, of which there is one per chamber. The ALCT checks every bunch crossing for patterns in the six planes that are consisent with a muon track originating from the interaction point. Any pattern that is found is called an ALCT, and serves as a trigger primitive that is transferred downstream for further decisions on whether the event should be recorded as data. A depiction of this avalanche and ALCT formation is given in Figure 2.22.

Each CSC has 5 cathode front end boards (CFEBs) that serve to manage the charge deposited on the cathode strips. The CFEBs contain more logic than

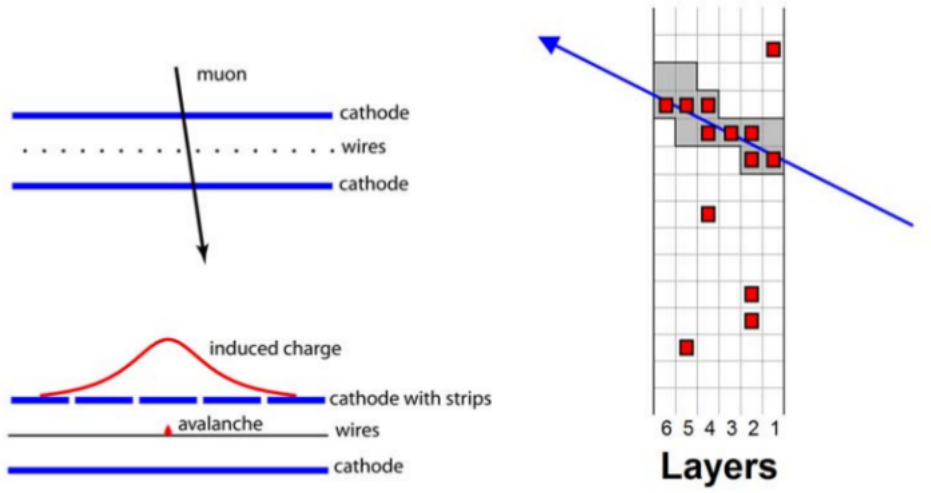


FIGURE 2.22: A schematic view of one gap, depicting the principle of CSC operation (right). The charge induced on the cathode strips is interpolatd, and a localization of the avalanche along the direction of the wire is found. A pattern of these hits is quickly matched to form an anode local charged track (ALCT, right) [38].

the AFEBs, and use a comparator network that uses on board amplifier shaper outputs to achieve a resolution corresponding to each half strip. This is achived via a comparison for every group of 3 adjacent strips, determining the amplitude of the central strip signal and comparing it to the central-to-left and central-to-right signal. Thus if the charge on the central strip is above the threshold, and the charge on the right strip is larger than on the left, the hit position must be in the right half of the central strip. This pattern is called a cathode local charged track (CLCT), and is depicted in Figure 2.23.

The half-strip cathode local charged track hits and the anode local charged track hits are both sent to a piece of off-chamber electronics called the trigger motherboard (TMB). There is one TMB per chamber, and each can create two 2-dimensional LCTs from the ALCT and CLCT that it receives from the chamber. These 2D LCTs are sent on to muon port cards (MPCs), each of which collects hits from 9 chambers. The MPC collects the 2D LCTs, sorts them, and finds the 3 highest quality candidates to send further upstream to the Level-1 muon trigger electronics. This path is shown in Figure 2.24.

The raw data are collected by the data acquisition, or DAQ, motherboards (DMBs). They are also located off-chamber, in peripheral crates, with the TMBs and MPC, and there is one per chamber. The data passed to the DMBs consist of anode and

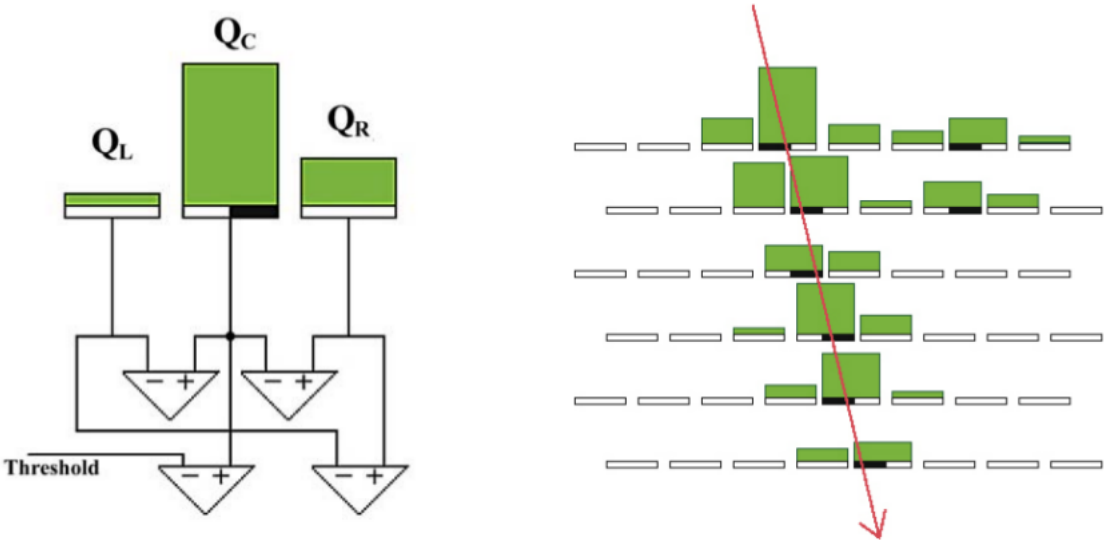


FIGURE 2.23: A schematic depiction of the CFEB comparator network (left) and how half-strip resolution is acheived. A muon track is depicted in red (right) and the white rectangles represent cathode strips. Each half-strip where a hit is found is colored black, and the green rectangles represent the amplitude of the signal, showing how the comparator network makes the half-strip decisions [38].

cathode comparator hits in a time window that is up to 32 bunch crossings long. These data that are collected by the DMB are in turn passed on to the detector dependent unit (DDU) and then to a data concentration card (DCC, not used after Run 1 ended in early 2013) and then finally on to the CMS filter farm in order to be processed by the CMS high level trigger (HLT) software. The approximate event size per chamber is 4-5 kBytes. These electronics are also shown in Figure 2.24.

If the Level-1 trigger accepts the data from a collision (this decision being made by Level-1 electronics from all subsystems), a "Level 1 accept", or L1A, is sent back out. This signal, as well as the LHC clock and all other control signals, is distributed to all the CSC electronics by the clock control board (CCB). This is also shown in Figure 2.24. The raw anode and cathode local charged track data is only sent upstream in coincidence with an L1A signal, so the CSC read-out system is intrinsically zero-suppressed.

2.2.7.4 Upgrades during the first long shutdown and the optical DMB

In order to handle the increased rate resulting from higher instantaneous luminosities moving forward in the lifetime of the LHC and of CMS, one of the many

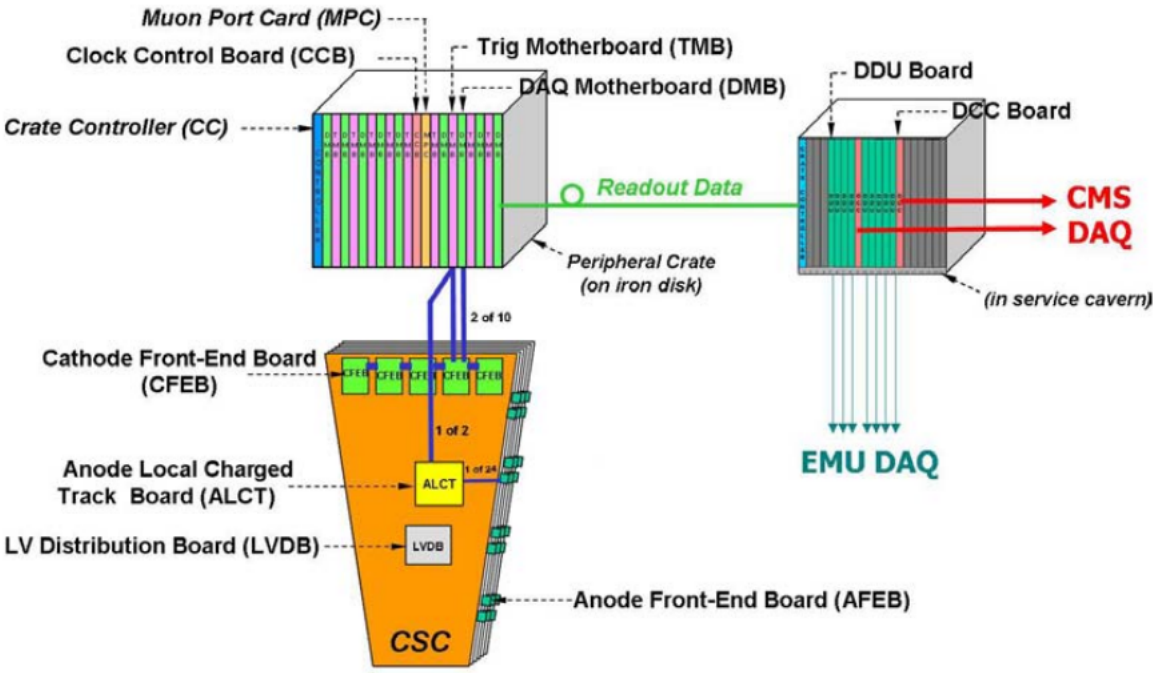


FIGURE 2.24: The schematic layout of the cathode strip chamber trigger and read-out electronics [38].

upgrades performed during the long shutdown of 2013 and 2014 was a new version of the DMB of the CSC subystem with an optical data link. Specifically, the CSCs closest to the IP (the ME1/1 chambers) needed DMB upgrades moving in to 2015 for LHC Run 2, and for LHC Run 4 the ME2/1, ME3/1, and ME4/1 chambers will need them as well. This was planned in concert with changes to the onboard electronics of the ME1/1 chambers, increasing the number of CFEBs from 5 to 7, as well as upgrading them to be digital (DCFEBs), no longer relying on analog switched capacitor arrays for data storage and ordering. All told, this upgrade for the CSCs resulted in a readout system for the ME1/1s shown in Figure 2.25.

The author of this thesis worked on the VME protocol module for the ODMB, which handled communication between the ODMB and PCs, in order to perform testing and communicate with other board in turn connected to the ODMB. The overall layout of the firmware is shown in Figure 2.26, which was converted to VHDL from the orignal schematics with with the DMB firmware was defined. This and many other upgrades performed allowed CMS to operate at full performance moving in to 2015.

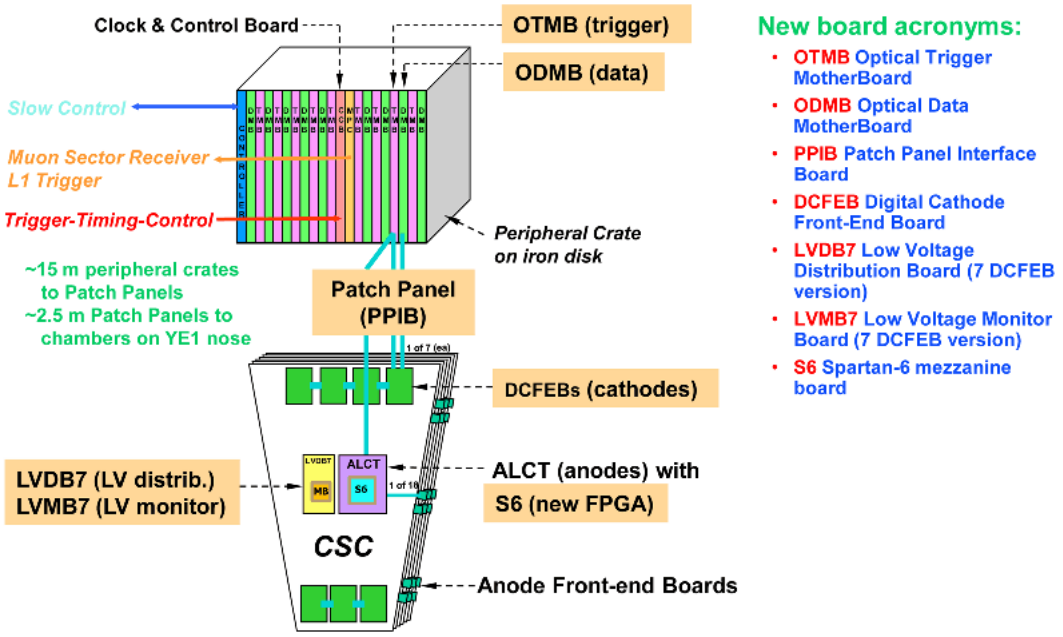


FIGURE 2.25: The schematic layout of the cathode strip chamber trigger and read-out electronics after the upgrades performed during the first long shutdown [38].

2.2.8 The Trigger system

The LHC delivers collisions to the CMS detector at a beam crossing interval of 25ns (or 50ns in 2011 and 2012), corresponding to a crossing frequency of 40 MHz at design parameters. At this design frequency, and at nominal instantaneous luminosity, an average of 20 collisions per crossing will occur. As it is impossible to store and process data for some many events, a drastic reduction in the rate of event storage is necessary. This necessary reduction is performed by the trigger system, which is the the beginning of the physics event selection process.

The procedure of performing this reduction in rate is separated in to two steps, and thus two systems. The first is the Level-1 Trigger system (L1 Trigger) and it is composed of custom-designed, mostly programmable electronics. The second is the High-Level Trigger (HLT), which is a software-based system implemented in a filter farm of approximately one thousand commercial processors. Combined, the L1 Trigger and HLT effect a rate reduction of at least a factor of 10^6.

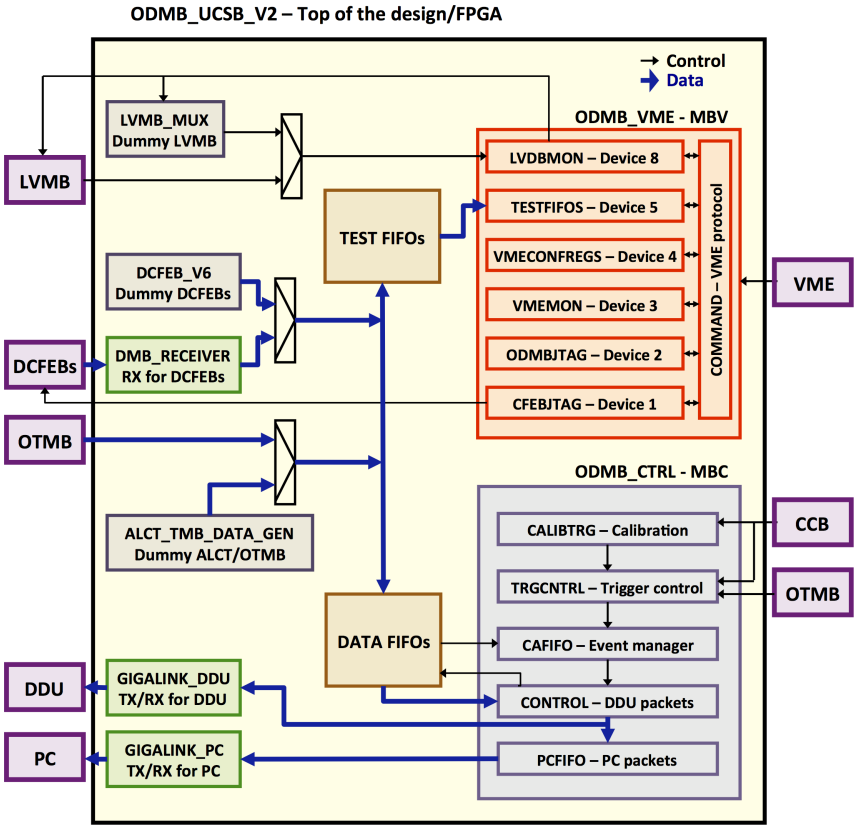


FIGURE 2.26: A schematic layout of the ODMB firmware. The red block in the upper right represents the VME protocol, for sending command messages to the ODMB from a PC via the peripheral crate and through it to other on-chamber electronics. The blue block in the lower right represents the module that is responsible for receiving CMS control signals from the CCB, as well as passing on data packets to the DDU.

2.2.8.1 The L1 Trigger

The design output rate of the L1 Trigger is 100 kHz, which in practice is limited to 30 kHz, with an approximate safety factor of three. The L1 Trigger uses relatively coarsely segmented data from the calorimeters and the muon system, while keeping the high-resolution data in pipelines in the front-end subdetector electronics. In order to ensure flexibility, the hardware for the L1 Trigger is implemented in FPGA (Field Programmable Gate Array) technology in most places, but some ASICs and memory lookup tables (LUTs) are also used where speed and radiation resistance is important.

The L1 Trigger is split in to local, regional, and global components. At the bottom end, "closest" to the subdetectors themselves, the local component called Trigger Primitive Generators (TPGs) are based on energy deposits in the calorimeters and track segments or hit patterns in the muon chambers. These TPGs are passed

to the Regional Triggers which combine their information and use logic to rank and sort trigger objects such as electron or muon candidates in regions limited spatially. The ranking is determined as a function of energy or momentum and quality, which reflects the confidence in the L1 measurements which is in turn based on the subdetector properties, electronics, and the information available. The Global Calorimeter and Muon triggers determine the best-quality calorimeter and muon objects across the entire CMS detector and transfer those candidates to the Global Trigger. The Global Trigger, finally, takes the decision to reject an event or to accept it for further decision making by the HLT. This decision is based both on algorithmic calculations with the candidate objects and on the readiness of the subdetectors and the data acquisition system (DAQ), and is sent back to the subdetectors via the L1 Accept signal (L1A). The overall architecture of the L1 Trigger system is shown in Figure 2.27.

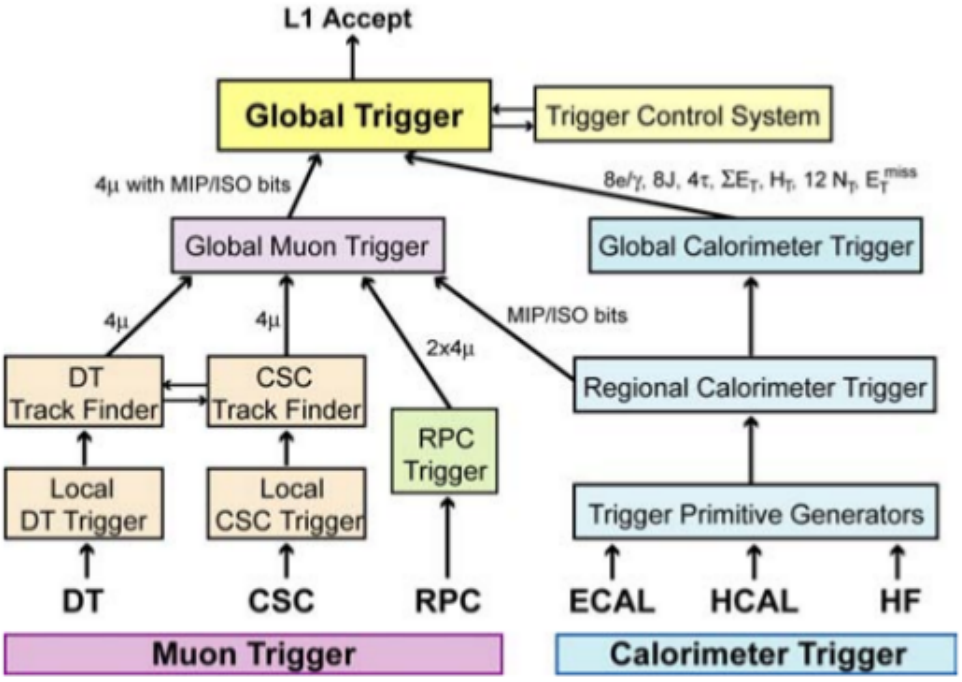


FIGURE 2.27: The architecture of the L1 Trigger [38]. The trigger layers are shown with what information is sent to the Global Trigger: Muons (μ), minimum-ionizing particle and isolation bits (MIP/ISO), electrons and photons (e/γ), jets (J), taus (τ), total transverse energy (ΣE_T), total hadronic transverse energy(H_T), number of jets passing various transverse energy requirements (N_T), and missing transverse energy (E_T^{miss}).

The TPGs in the Calorimeter Trigger subdivide both calorimeters in to trigger towers. The TPGs compute the sums of transverse energies measured in the ECAL

crystals or HCAL read-out towers in order to obtain the entire trigger tower E_T and determine the correct bunch crossing. The TPGs then are transmitted to the Regional Calorimeter Trigger (RCT) which is responsible for multiple tasks. The RCT determines regional candidates for electrons and photons via determining the tower with the largest energy deposit and then applying two shower profile requirements: a fine-grained crystal energy profile that reflects teh lateral shape of a shower, and a requirement based on the ratio of deposited energies in the hadronic and electromagnetic portions. It also requires there be at least one quiet corner in one of the course sections surrounding the hit. This strategy is depicted in Figure 2.28. The RCT also sums transverse energies in the calorimeters, determines τ-veto bits based on the narrower shape of τ decays, and produces information relevant for muons in the form of minimum-ionizing particle (MIP) and isolation (ISO) bits. Finally, the Global Calorimeter Trigger (GCT) determines the jets, total transverse energy and missing transverse energy, jet counts, and the jet-only transverse energy sum given a programmable threshold (H_T).

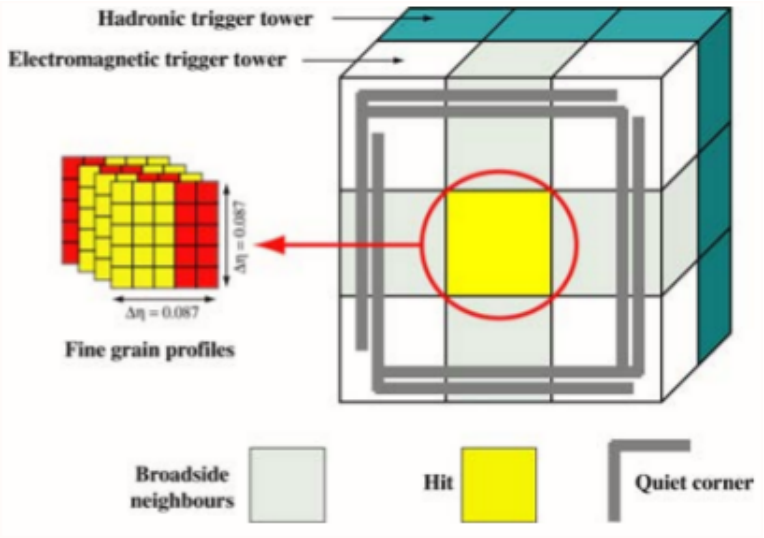


FIGURE 2.28: The algorithm the RCT uses for electron/photon triggering [38].

The TPGs in the muon trigger are sourced from all three muons subsystems. The CSCs in the endcap deliver three-dimensional track segments (straight line segments called local charged tracks, or LCTs). The DTs likewise contribute track segments, which are track segments in the ϕ-projection and hit pattterns in the η-projection. All chambers also identify the bunch crossing from which the event originated. The regional muon trigger is based on the DT and CSC Track Finders (DTTF, CSCTF), which serve to create tracks out of the individual segments they receive and thus identify muon candidates along with their transverse momenta,

locations, and quality. The functionality of both the DTTF and CSCTF fits into high-density FPGAs: DTTF segmentation in the central wheel is 2×12 half-width sectors and in the four outer wheels it is 12 full-width sectors each, while the CSCTF segmentation is by 2×6 60°-sectors. The CSCTF synchronizes the data and uses extrapolation to form an overall track. Then the p_T is assigned via lookup tables based on the ϕ-information from three stations. The overall strategy for the DTTF is the same, which also uses some pattern matching in the η coordinate. This algorithm is depicted, for the DTTF, in Figure 2.29. The Global Muon Trigger (GMT) is fed the muon candidates from the DTTF, CSTF, and RPC Trigger and uses them to reduce repeat muon measurements via merging kinematic parameters and canceling duplicates. Muons are also back extrapolated to the vertex using the MIP/ISO bits from the calorimeters which are in turn also added to the GMT output. Finally, the muons are sorted by transverse momentum and quality, to create four candidates to be given to the Global Trigger.

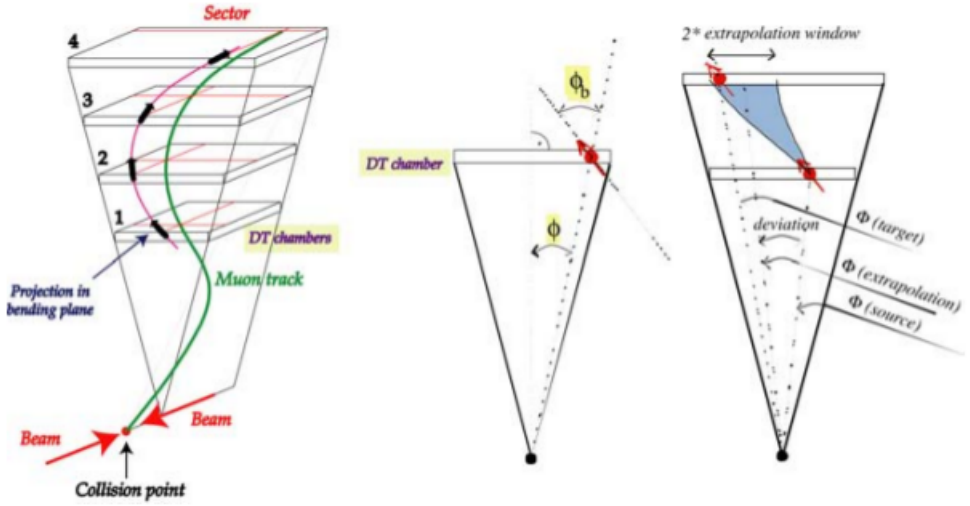


FIGURE 2.29: The algorithm the DTTF uses for muon candidate track extrapolation [38]. The CSCTF uses a very similar scheme, with 3-dimensional segments.

The Global Trigger (GT) takes the decision to accept or reject an event at Level-1 based on the trigger objects that are delivered to it from the GCT and GMT. The algorithms vary in complexity, with the most basic ones consisting of applying p_T or E_T thresholds to a single object, or requiring minimum jet multiplicities. Location and quality information is available as well, so more complex algorithms taking into account event topology cand be programmed. Up to 128 algorithms can be executed in paralle, and for normal physics data this single trigger mask is applied, with the L1A decision being taken accordingly.

2.2.8.2 The HLT

Unlike the L1 Trigger, the HLT operates on software. It serves to further reduce the output rate of the L1 Trigger of approximately 30 kHz to 100 Hz. It consists of a variety of trigger paths based on object kinematics and quality criteria. However, since it has access to the complete data read-out, it is capable of performing more complex calculations similar to those used by off-line analysis software. Its goal is to both reduce rate and ensure that events saved match the interest of physics studies, so events are partially reconstructed even in the cases of simple kinematics-based criteria. HLT algorithms evolve with experience, optimizing the usefullness for off-line analysis. Additionally, in 2011 and 2012, the instantaneous luminosity delivered by the LHC increased continuosly, so some HLT trigger paths evolved with time to compensate for this and maintain a relatively constant rate of 100 Hz throughout the run periods. Some HLT triggers also "prescaled" events, only saving a sampling of events (saving a $1/N$ fraction of events, where N is called a prescale) that would otherwise increase the rate by large quantities, these prescales were also varied throughout run periods.

Chapter 3

Event Reconstruction

Event reconstruction is the process of computing quantities useful for physics analyses and constructing software objects called "physics" objects which are associated quantities that describe the kinematic properties of particles in an event. It is a software operation which is fundamentally a data reduction procedure whose primary client is the data analysis. Reconstruction can be divided in to 3 fundamental steps: local reconstruction that occurs within a single subdetector module, global reconstruction that occurs within a whole subdetector system, and the combination of the reconstructed objects to produce higher-level objects.

Local reconstruction uses as input either real data acquired during run time that was triggered on by the HLT and stored, or simulated data produced to represent the real data. The data collected by the DAQ system is in the form of the digitized response of the various subsystems' modules to incident particles (charge avalanches, energy deposits in crystals, etc), so these input data objects are called "digis" - a reference to the fact that they are either digitized detector responses or simulations thereof. The output of local reconstruction software is a reconstruction of the physical hit that occurred in the subdetector module. Called a "rechit", these data objects are typically position measurements taken from times or clusters of strips or pixels in tracking-type subdetector systems (Muon and Tracker), and locations of energy cluster deposits in the calorimeters (ECAL and HCAL). These rechits are used as input in the following reconstruction step.

In the global reconstruction step rechits from multiple subdector modules are combined, and then data from multiple subdetectors are combined to form physics objects. For instance, rechits from muon detectors are used to produce reconstructed

tracks representing candidate muon tracks and rechits in the tracker are used to produce reconstructed charged particle tracks, which are then combined to form global muon tracks.

Finally, the last step combines the reconstructed data objects from various individual subdetectors to produce the higher-level objects that are ultimately used in high-level triggering and/or for physics analyses. As an example, a given track in the Tracker subsystem and a track in either the DT or CSC subsystem (Muon subdetectors) may be combined to form final muon candidates, and likewise electron candidates from the calorimeters are matched to tracks in the Tracker.

Figure 3.1 depicts the path that data from a muon takes through this reconstruction algorithm to eventually become a set of kinematics quantities identified as a "muon". In depth descriptions of the reconstruction of all physics objects follow in sections 3.1 through 3.4.

3.1 Track and Vertex Reconstruction

Tracks formed by charged particles in the tracking system are reconstructed via an iterative trajectory finding method called the Combinatorial Track Finder (CTF). This method makes use of both pixel hits and strip hit in in local reconstruction, and pixel hits created via a fast algorithm are used for initial track seeding.

For "first-pass" hit reconstruction, the transverse and longitudinal coordinates are determined via a simple computation. In the case of a single pixel hit the center of the pixel is taken, whereas if the hit contains multiple pixels the hit position is determined using the relative charge of the two pixels at either end. The position is then corrected for the Lorentz drift of the collected charge in the magnetic field. These simple hit coordinates are used for track seeding, the first step of the global track reconstruction.

Over the lifetime of the detector the radiation exposure can significantly impact a pixel module's charge collection efficiency, and thus degrade the performance of the above standard hit reconstruction method. The hit position can become biased by up to 50μm, so a template-based algorithm is used for the steps after track seeding in global track reconstruction [43]. In this method the distribution of charge in the pixel module is compared to expected projected distributions in

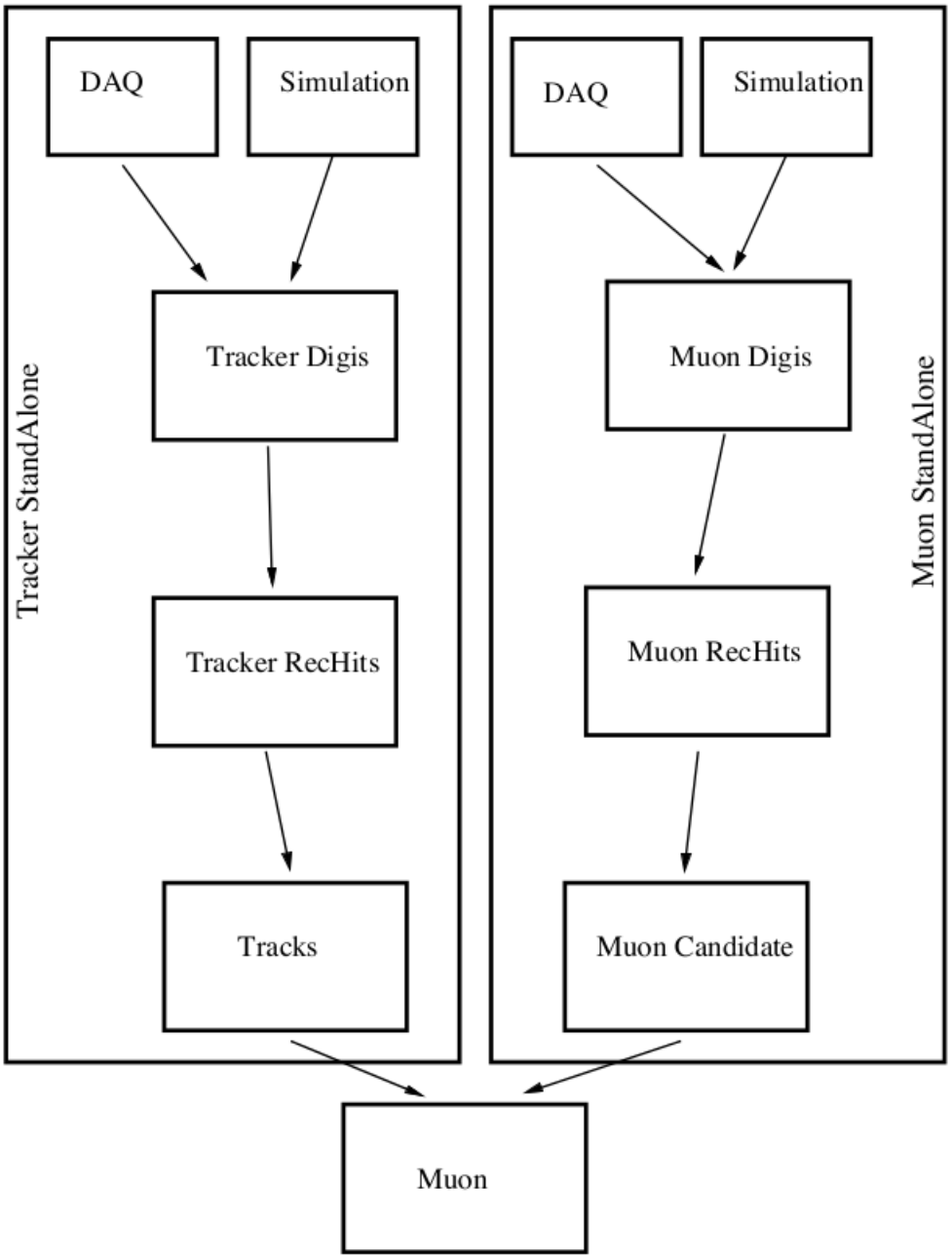


FIGURE 3.1: An example of a path that data from a muon takes through reconstruction, in the Tracker and Muon subsystems [42].

order to estimate hit positions. These expected distributions, called templates, are produced in a simulation called PIXELAV which can include descriptions of behavior of irradiated pixel sensors. Thus, with updates over the course of the pixel tracker's lifetime, template-based algorithms can maintain a more realistic picture of sensor behavior and thus reduce position bias and improve resolution of the hit reconstruction.

In the case of hit reconstruction in the strip detector, a seeding method is used.

Clusters are seeded by any channel that is saved offline that has a charge of at least three times the corresponding channel noise. Neighboring strips are added to this seed if their charge it twice their strip noise, and a cluster overall is kept if the total charge is a factor of five larger than the total cluster noise. The position of the hit for each cluster is then determined from the charge-weighted average of strip positions, corrected for Lorentz drift as well as for inefficient charge collection near the backplane of the sensitive strip volume. These strip hits are used in the successive steps of global track reconstruction after seeding.

The global reconstruction of tracks takes the hits in the tracker subdetectors and estimates the momentum and positions of the charged particles responsible for those hits. The momentum is computed by the CTF from the bending of the trajectory of the particle in the magnetic field and thus the key role of the global reconstruction is to create these trajectories. The CTF is an adaptation of the combinatorial Kalman filter [44–46] which is a basic Kalman filter [47] that includes pattern recognition with track fitting in the same framework. The CTF process is applied 6 times, iteratively, with the strategy of searching for the tracks that are easiest to find in the first iterations and the tracks that are hardest in the last iterations. Thus the hits associated with tracks are removed in each iteration, reducing the combinatorial multiplicity in subsequent steps that search for more difficult track classes such as low-p_T or displaced tracks. Each iteration consists of four steps:

- Seed generation provides initial track candidates based on only two or three hits. The seed contains the initial estimate of trajectory parameters with their uncertainties.
- A Kalman filter finds the tracks, extrapolating the seed trajectories along the expected path and searching for additional hits to be assigned to the track candidate.
- A Kalman filter and smoother fit the track, to provide an improved estimate of the parameters of each trajectory.
- Finally, track selection applies a set of quality criteria, discarding tracks that fail, and sets various quality flags on passing tracks

The six successive iterations vary in how the initial seed is configured and the criteria applied in the final selection, with the same Kalman filter-based method applied to track finding and fitting.

The resolution of the reconstructed parameters of the tracks is studied using simulated events, using the differences between the generated values and the reconstructed values. Figure 3.2 shows the p_T resolution, as a percentage of p_T, for singly produced and isolated muons.

Reconstruction of the primary vertices aims to measure the location and the associated uncertainty of all the proton-proton interaction vertices in each event, including those from the hard scattering vertex (the "main" event vertex) and any vertices from pileup collisions, using the reconstructed tracks. It is performed in three steps:

- Track selection, in which tracks likely to have been produced promptly in the primary interaction region are chosen by imposing requirements on the maximum value of significance of the transverse impact parameter relative to the beam spot, the number of strip and pixel hits associated with the track, and the normalized χ^2 from the trajectory fit [43].
- Clustering of tracks that appear to originate from the same interaction vertex is performed by an association system called "deterministic annealing" [48] which included "soft" assignments, allowing tracks to be associated with more than one cluster.
- Finally each candidate vertex is fitted to find its position and covariance matrix as well as parameters such as the number of degrees of freedom and weights of tracks used to indicate likelihood of success of the fit. Each track is assigned a weight between 0 and 1 that reflects the likelihood it belongs to the vertex.

The resolution of the reconstructed primary-vertex position is greatly dependent on the number of tracks used in the fit, as well as the p_T of those tracks. This dependence is shown in Figure 3.3.

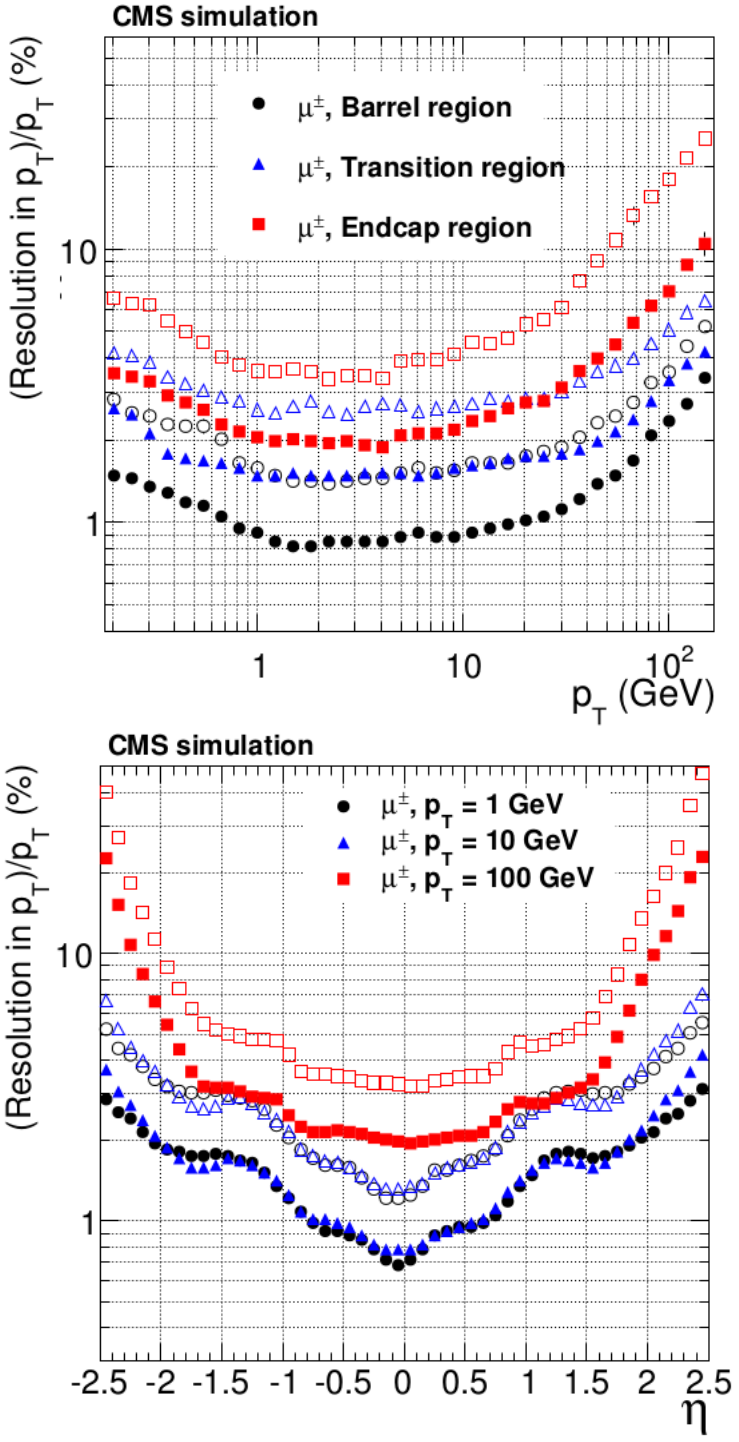


FIGURE 3.2: Resolution of the track p_T, for singly produced and isolated muons with p_T=1,10, and 100 GeV. The solid symbols represent the half-width for 68% intervals, while open represents 90% [43].

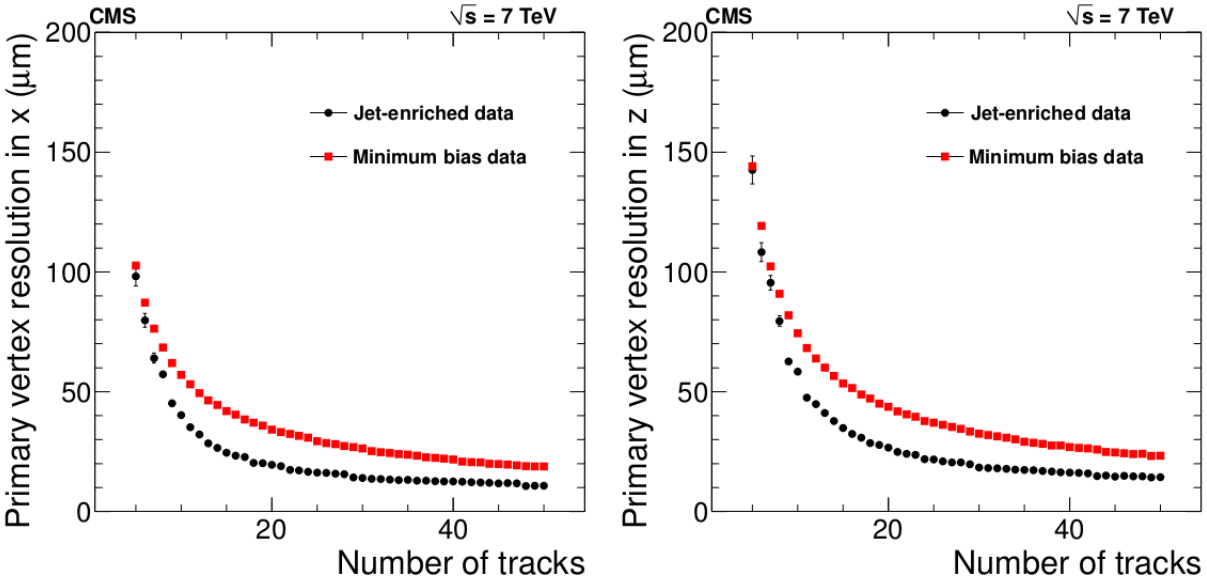


FIGURE 3.3: Primary vertex resolution as a function of the number of tracks in the vertex. The jet-enriched data contains tracks with a significantly higher p_T values [43].

3.2 Muon Reconstruction and Identification

3.2.1 Muon Reconstruction

Aside from neutrinos and other particles that only weakly interact and cannot be detected in any subdetector in CMS, muons are the only particles that pass through all the detectors inside the magnet solenoid and form tracks in the gas based muon system. An example of a muon's path through CMS is shown in Figure 3.4. The CSCs, DTs, and RPCs can form their own muon tracks as well as identify associated tracks from the tracker, and as a result muons are globally reconstructed in both the muon systems and in the tracker system.

As in other subsystems, the muon systems begin their reconstruction with local reconstruction - building reconstructed hits either out of simulated data or actual collected data. The main objects that the DTs use for this are hits in the volume of the drift cells, for which the drift times are converted to drift distances. Separate $r-\phi$ and $r-z$ projections are combined in to 3-D "rechits" and a line segment is formed from aligned hits within the chamber. The best segment candidate is chosen from any segments sharing hits, and in turn the hit information and segment fit is updated.

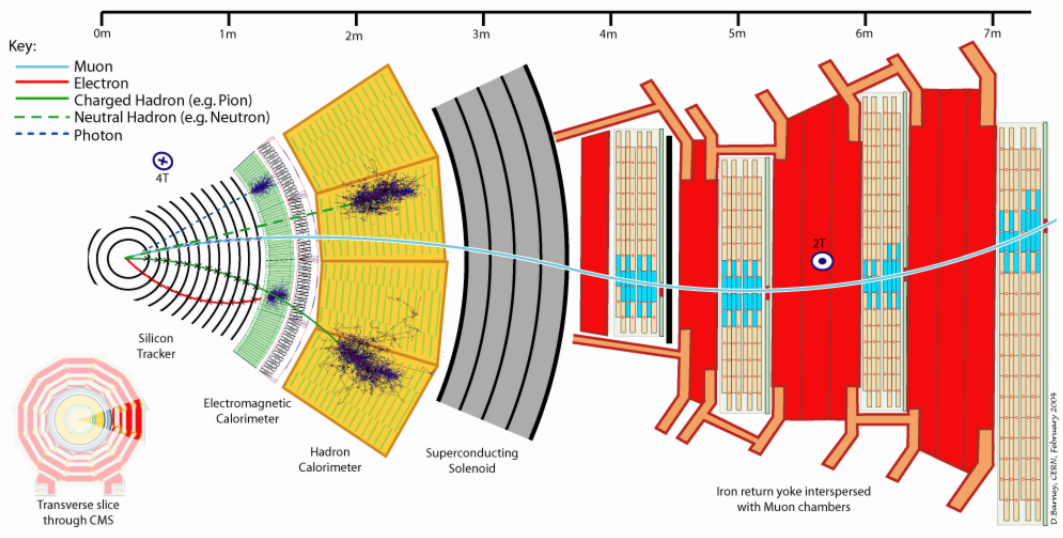


FIGURE 3.4: A slice of a the CMS detector, on the transverse plane, depicting various particles interacting with each specific subsystem. A muon is shown in blue, traversing the tracker, calorimeters, and magnetic solenoid to leave a track in a muon system. Its track is bent in two directions, in the inner 4 T field and the outer 2 T field.

A very similar method is used for the CSCs. In each of the six layers of a CSC chamber the pulse height is measured to determine the probable hit position, and a hit is produced based on the intersection of the locations on the strip and on the wire. Then a segment is produced by connecting the first and last hits in a chamber, and including any hits within a clustering window along the line connecting those two hits. This segment must pass a requirement on the resulting χ^2/(number of degrees of freedom) of the fit of those hits that are included. A minimum of three hits is required for each segment, although segments tend to include closer to the maximum possible 6 hits per segment (due to the 6 layers of the CSCs), as shown in Figure 3.5.

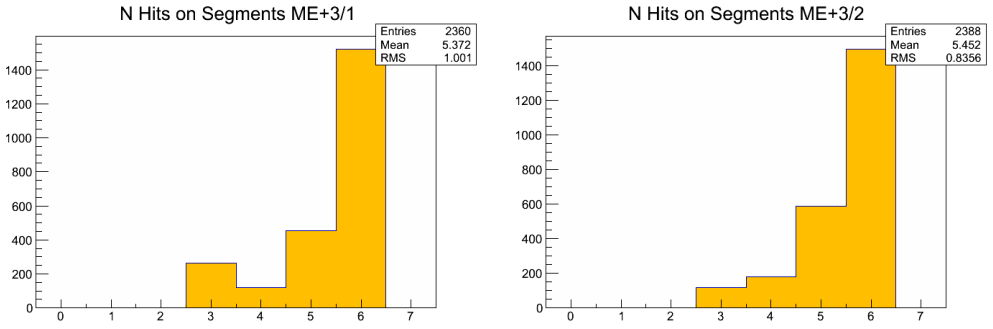


FIGURE 3.5: The average number of hits per segment for a given run in 2012. The inside disk of the third station of the CSCs (left, ME+3/1) has slightly more "showered" muons than the outside disk of the same station (right, ME+3/2).

Finally, the RPCs are used in local reconstruction to obtain neighboring clusters with average positions that match the reconstructed hits in either the DTs or the CSCs.

Global reconstruction of muons begins with "standalone" reconstruction of tracks only in the muon chambers themselves. In standalone reconstruction, the track segments that are built during local reconstruction are used as seeds for muon trajectories. The track segments from the innermost detectors are taken as state vectors (vectors containing their positions, direction, and momentum, along with covariance matrices) to start an inside-out Kalman fitting procedure [42, 46]. A predicted state vector is created at the next chamber by using the parameters of the initial state vector and propagating the trajectory outwards. This projection takes the muon energy loss in the material, the effects of multiple scattering, and the non uniformity of the magnetic field into account. At that chamber, assuming that the actual track segment there passes a certain χ^2 requirement, the predicted state is updated with the new information, adjusting parameters of the state vector for the next iteration. Once the outermost chamber is reached, the "Kalman Update" step is complete, and an outside-in fit is performed in the exact same manner, in order to smooth the resulting trajectory. Together with the "Kalman Update", this "Kalman Smoothing" step composes the entire Kalman Filter that is applied to the muon track segments. In the DTs, since each hit is only 2-D, the entire segment is used for this procedure, but the CSC hits are 3-D so each hit is used individually.

The final step of muon reconstruction, called global reconstruction, includes information from the tracker. There are two approaches to this, one which produces objects called tracker muons, and one which produces objects referred to as global muons.

In tracker muon reconstruction, all tracks in the tracker with $p_T > 0.5\,\text{GeV}$ and total momentum $p > 2.5\,\text{GeV}$ are considered to be possible muon candidates and their trajectories are extrapolated to the muon system taking the magnetic field, average expected energy losses, and multiple coulomb scattering as done in the Kalman Filtering into account. There, if at least one muon segment (the short track stubs created during local muon reconstruction) matches the track position, the track is tagged as muon, and a track muon data object is created. This matching is performed in the local coordinates of the chamber, and a match is declared if the distance between track and segment coordinates in local x is less

than 3 cm or if the pull for local x is less than 4 [49]. The pull is defined as the difference between the positions of the segment and track, divided by the combined uncertainties.

If tracker muon reconstruction can be thought of as "inside-out", then global muon reconstruction is "outside-in". To create global muons, the standalone muon track is propagated back in to the tracker. The same Kalman Filter technique is applied here, using the tracker hits as well, and any tracker hits passing the χ^2 requirements are added to the trajectory, which iterates through the tracker. An example event display showing global muons in both the CSCs and the DTs is shown in Figure 3.6. This global muon fit can significantly improve the momentum resolution compared to the tracker-only fit for muons with large transverse momenta, $p_T \gtrsim 200\,\text{GeV}$.

The global muon fit is used for the muons considered in the physics analysis presented in this thesis. While it is sufficient for the needs of the analysis within this thesis, additional refinements can be made to muon momentum calculation. The "sigma-switch" method uses the global-muon fit if the results of the tracker-only and global fits for the ratio of the muon charge q to the muon moment p, q/p, are within $2\sigma_{q/p}$ of each other, and if the p_T values for both fits are at least 200 GeV; and it uses the tracker muon fit otherwise [50]. The "Tune P" algorithm chooses between the tracker muon fits and a global fit using only the first muon station also using a "picky" fit which imposes tight cuts on the compatibility of hits in the muon chambers which appear to contain electromagnetic showers (large numbers of hits). The momentum resolutions of the global muon fit method as well as tracker-only and these two refinements are shown in Figure 3.7.

3.2.2 Muon Identification

At low transverse momenta ($p_T \lesssim 30\,\text{GeV}$), the source of the majority of muons is semileptonic decays of heavy-flavor hadrons. Light-flavor hadron decays and hadron showers that are not fully contained in the calorimeters also contribute significantly in this regime. The details of selection criteria applied to muons can very significantly impact the relative weights of these various background contributions. However, at higher transverse momenta ($p_T \gtrsim 30\,\text{GeV}$), muon decays from W and Z bosons dominate. Muon decays from hypothetical high mass resonance particles would likewise possess high transverse momentum. To reduce

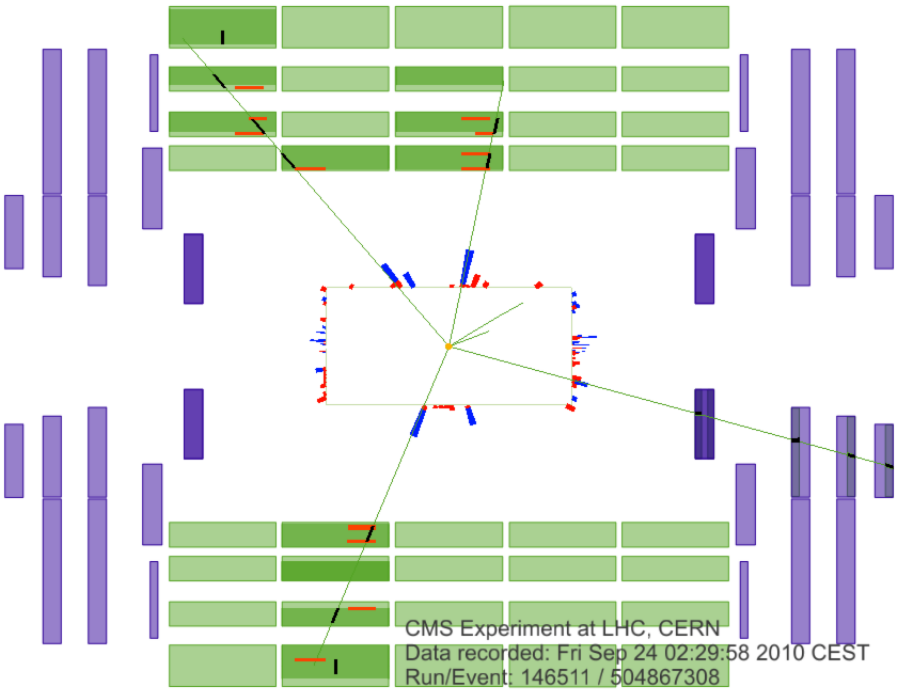


FIGURE 3.6: A longitudinal view of a collision event in which four muons were reconstructed. The thin green curves in the inner cylinder represent the tracks of charged particles which were reconstructed in the inner tracker with transverse momentum $p_T > 1\,\mathrm{GeV}/c$. Those that extend to the muon system represent the tracks of muons that were reconstructed with both the inner tracker and the muon system. Three muons were detected by the DTs and RPCs, and the fourth by the CSCs. The short black lines in the muon system show segments that were included in the muon track, and the horizontal red lines are depicted the positions of RPC hits. The energy depositions are shown in red (for ECAL) and blue (for HCAL) on the outer edge of the inner cylinder. [49]

various sources of backgrounds, in general a "tight" muon selection is applied, with the follow requirements:

- The muon must be a global muon, and the global muon track fit must possess a $\chi/d.o.f.$ less than 10.
- At least one hit from a muon chamber must be included in the global muon track fit.

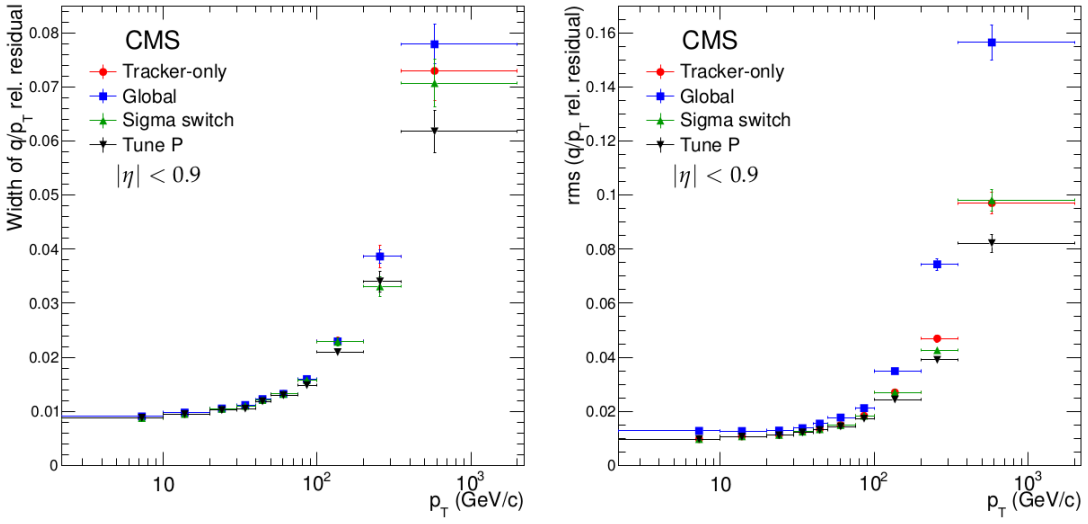


FIGURE 3.7: The resolutions of q/p_T for various muon momentum calculation algorithms, as measured using cosmic muons. The muons are detected traversing CMS vertically, and are split into top and bottom legs (detected in the DT chambers at the top of CMS, and at the bottom). The relative residuals are then defined as $R(q/p_T) = [q/p_T(top) - (q/p_T(bottom)]/[\sqrt{2}q/p_T(bottom)]$. The Gaussian widths of the relative residuals are are shown in (a) and sample RMS values, truncated at ± 1, are shown in (b). In each case, results from the tracker muon fit are in red, global fit in blue, and the more advanced algorithms of the Sigma switch and Tune P are shown in green and black, respectively. [49].

- The track from the tracker that is associated with the muon must be matched to muon segments in at least two separate muon stations, and must contain at least 10 tracker hits in the trajectory, with at least one from the pixel tracker.

- The transverse impact parameter, the distance of the closest approach in the $x - y$ plane of the track with respect to the primary vertex, $|d_{xy}|$ must be less than 2 mm.

However, a selection is optimized for high p_T muons, not those with at least p_T 30 GeV but significantly more ($p_T \gtrsim 200\,\text{GeV}$). This selection is as follows:

- The muon must be a global muon.

- At least one hit from a muon chamber must be included in the global muon track fit.

- The track from the tracker that is associated with the muon must be matched to muon segments in at least two separate muon stations, with hits in at least 5 tracker layers, with at least one from the pixel tracker.
- The transverse impact parameter, the distance of the closest approach in the $x-y$ plane of the track with respect to the primary vertex, $|d_{xy}|$ must be less than 2 mm. Additional the longitudinal distance, $|d_z|$ must be less than 5 mm
- The track used for the final momentum calculation must have a relative p_T error of less than 30%

These criteria listed above are applied to the muon fits themselves, but additional quality criteria are added regarding activity from other particles in the vicinity of the muon. This can be used to discriminate muons from decays of W and Z bosons (or heavy resonances) from those that are produced by heavy-flavor decays and hadron decays in flight. Various isolation algorithms are optimized for different analysis requirements. The best isolation algorithm for W and Z decays combines the tracker and the calorimeters:

- Combined relative isolation: The scalar sum of the p_T of all tracker tracks and sum of energies measured in ECAL and HCAL towers in a cone of radius $\Delta R \equiv \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} < 0.3$ centered on the muon track direction is computed. The energy deposits in the muon track itself are not part of this sum. The relative isolation, I^{rel}_{comb}, is defined as the ratio of that scalar sum to the muon track p_T.

For muons decays from heavy resonances that decay to muon pairs, the following algorithm is optimum:

- Tracker relative isolation: The scalar sum of the p_T of all tracker tracks only in a cone of radius $\Delta R \equiv \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} < 0.3$ centered on the muon track direction is computed. Like the combined relative isolation, the energy deposits in the muon track itself are not part of this sum. The relative isolation, I^{rel}_{trk}, is defined as the ratio of that scalar sum to the muon track p_T.

In each case, for the muon to be considered isolated the relative isolation is required to be below a certain threshold, which is in turn optimized for various analyses.

3.2.3 Future upgrades for muon reconstruction

By 2023, the quadrupoles that focus the beams at the CMS collision regions are expected to near the end of their lifetimes due to radiation exposure. The third planned long shutdown for the LHC will occur at that point, with the plan to replace these with low-β quadrupole triplet magnets. Also, crab-cavities will be added to optimize the overlap of the bunches of protons at the interaction region. These two changes will significantly increase the LHC luminosity, and thus the pileup will become an even greater challenge for the experiment. This era of operation will be called "Phase-II" and the LHC's new configuration will be named the HL-LHC (High Luminosity LHC).

Many subsystems will receive sweeping upgrades to perform well in this environment. One such upgrade will be implemented on the the muon subsystems located in the endcaps. Improved RPCs will be installed, and chambers of a new type, Gas Electron Multiplier (GEM), will be implemented with good position resolution in order to improve momentum resolution in the standalone muon trigger and matching with tracks in the global muon trigger. Finally, to take advantage of space that will become free due to endcap calorimeter upgrades, and extended pixel tracker coverage in η, an additional station of muon detectors will be added using a triple layer of GEMs, to be called ME0, will be installed. The position of these planned chambers are shown in Figure 3.8, and their planned layout is shown in Figure 3.9.

The position of these planned chambers will increase the acceptance range for tracker muons, allowing identification of muons up to the planned extent of the ME0 chamber, $|\eta| < 2.8$, shown in Figure 3.10. This region will experience a large amount of muons from pileup and as a result selection of signal muons from W, Z, and Higgs decays may be challenging. The author of this thesis worked on preparation of early studies of basic selection algorithms for optimization of signal muon selection and preparation of code to perform reconstruction of muons in the planned chambers in simulation. In the selection algorithm studies, a requirement on the difference in the local ϕ between the track state and the segment is applied: a "loose" selection requiring this to be less than 0.5, and a "tight" selection requiring it to be less than 0.15. Early studies of the efficiency of these selections and the resulting background yield, performed in simulation, are shown

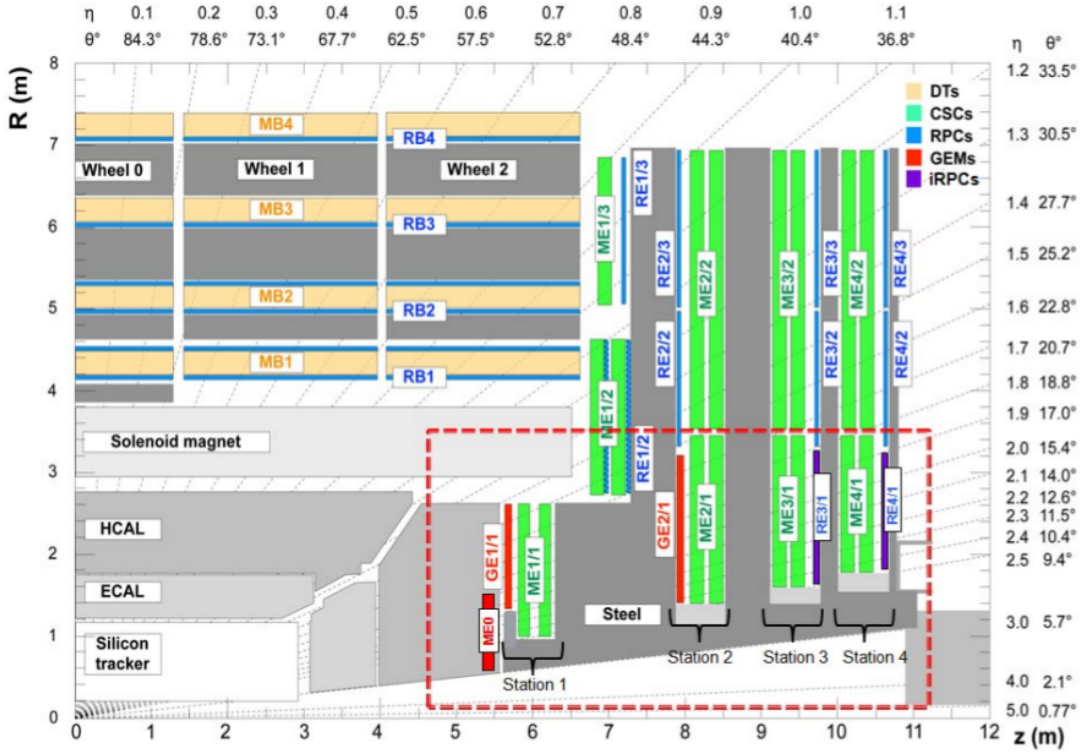


FIGURE 3.8: A quadrant of the muon system, showing the standard DT chambers (yellow), RPC (blue), and CSC (green). The locations for the new muon chambers planned for Phase-II are indicated by the dashed box, including the GEM stations (ME0, GE1/1, GE2/1, in red) and the improved RPC stations (RE3/1 and RE4/1, in purple) [51].

in Figure 3.11. They show a clear increase in acceptance in the higher η regions where the chambers will be installed.

3.3 Electron Reconstruction and Identification

3.3.1 Electron Reconstruction

The reconstruction of electrons in CMS, as for other physics objects, begins with local reconstruction. While the primary units of muon local reconstruction are track stubs, electrons (and photons) leave a majority of their energy in a small number of ECAL crystals via showers. Approximately 97% of electron incident energy is deposited in a 5x5 grid of crystals. This simple crystal window (of a fixed size) is useful for measuring unconverted photons and simplifying measurements of "low-radiating" electrons for calibration of the crystals [52]. However, typical

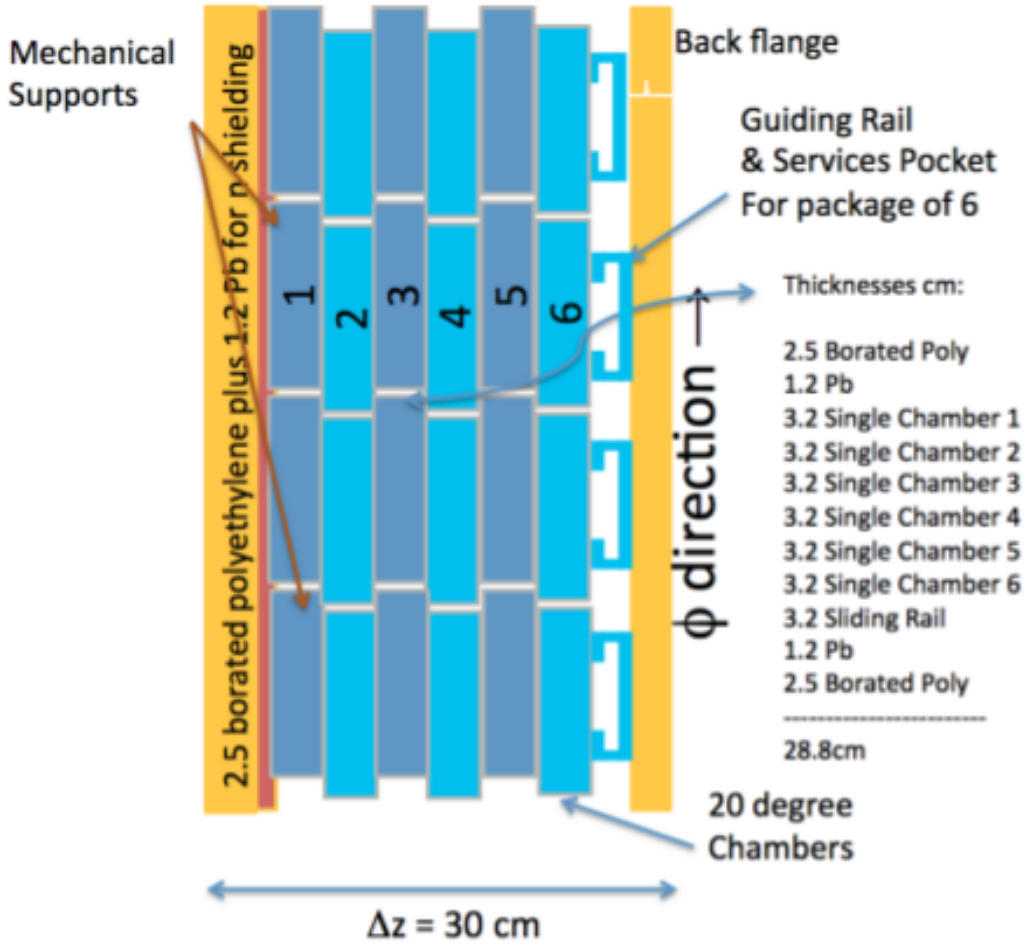


FIGURE 3.9: A preliminary planned layout of a 6-layer ME0 layout in the $\phi - z$ plane, using triple-GEM chambers. [51].

electrons are more complex to handle. Electrons traversing the silicon tracker material give off energy via radiation of bremsstrahlung photons which, due to the 4 T solenoidal magnetic field, reaches the ECAL spread in ϕ.

Thus, in order to correctly measure the electron energy at the primary vertex and to minimize the cluster containment variations, it is essential to collect these bremsstrahlung photons. This is done via "super-clustering" algorithms, so-called because they create superclusters composed of clusters spread in the ϕ-direction. One such algorithm is called the "Hybrid" algorithm, which attempts to take advantage of the simple geometry of the ECAL barrel and exploits the properties of the lateral showers in the transverse direction while searching for the separated bremsstrahlung energy in ϕ. In Hybrid super-clustering a "seed" cluster is a collection over ϕ of contiguous 3x5 crystal clusters in $\phi \times \eta$, which is separated

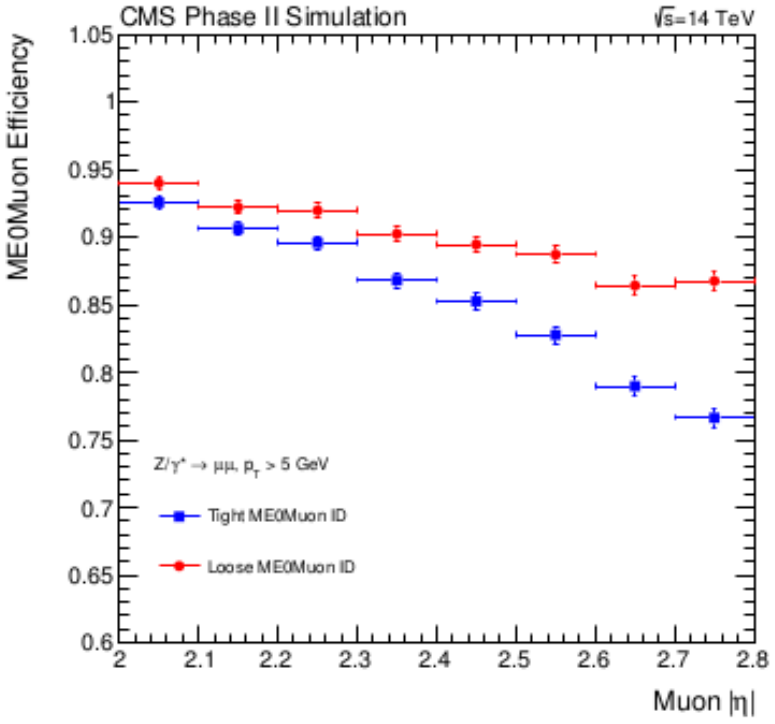


FIGURE 3.10: The efficiency of the loose ($\Delta\phi(track, segment) < 0.5$) and tight ($\Delta\phi(track, segment)) < 0.15$) selection for muons in the ME0 detector as simulated in Drell Yan events. The increase in acceptance can be seen in $|\eta| > 2.4$ [51].

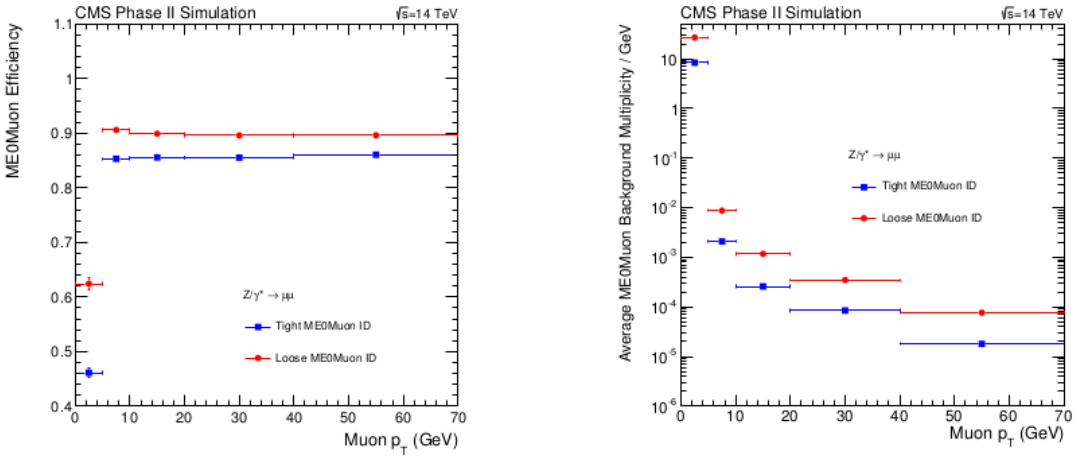


FIGURE 3.11: The efficiency and background yields of the loose ($\Delta\phi(track, segment) < 0.5$) and tight ($\Delta\phi(track, segment)) < 0.15$) selection, vs. p_T for muons in the ME0 detector, as simulated in Drell Yan events [51]. The background simulation included hadronic punchthrough and multiple scattering effects, but does not include neutron backgrounds.

from other collections by a valley in which less than 100 MeV is observed in one 3x5 cluster. The Island algorithm builds clusters by connecting rows of crystals with monotonically decreasing energies while moving away from a seed crystal.

Then superclusters are build by collecting other Island clusters in the ϕ-direction (in both directions) around each Island cluster. In either case, it is the resulting supercluster that forms the basic locally reconstructed unit.

As with the muon reconstruction, once the local reconstruction is complete tracks from the tracker are used to describe the complete trajectory of the electron, associating them with nearby "local hits" which are the calorimeter superclusters in the case of electrons. This trajectory building requires a seed, for which two different types are used. Tracker-driven seed-based reconstruction uses hits in the tracker much as the tracker muons do, and search for superclusters to associate with the track trajectory. Provided that a tight χ^2 requirement is used in the trajectory building, this method is shown to be very effective for high p_T electrons [42]. In the case of low p_T electrons and in general for detailed electron analysis, to ensure a high purity the superclusters themselves are used as trajectory seeds and are back-propagated through the tracker.

The standard treatment of the Kalman Filter for material effects is valid for the high energy electrons that do not release as much bremsstrahlung radiation in the tracker material, but the Gaussian modeling of the material effects fails for the low energy electrons that do radiate significantly. A nonlinear filter approach is used in this case, called the Gaussian Sum Filter (GSF) which better describes the propagation of electrons. In the GSF method all the state vectors and errors are Gaussian mixtures (weighted sums of Gaussians), and the weight of the components of the mixture depends on the measurements.

3.3.2 Electron Identification

As in the case of muons, various selection criteria are established, using properties of the electron candidates. These criteria are used to reduce the contamination from a variety of backgrounds, such as electrons produced within jets or from photon conversion as well as other physics objects that are misreconstructed as electrons (jets, for example). One such selection is referred to as the "High Energy Electron Pair" ID (HEEP ID), a set of selection criteria optimized for high dielectron invariant mass [53]. The HEEP ID uses GSF electrons as candidates, but applies a variety of cuts to them, and uses the energy from the superclusters alone, not a weighted average between supercluster and GSF track, as GSF electrons use. Over time this selection evolved, with small tunes made to various

parameters, version 4.1 is shown is Table 3.1. It has two separate sets of criteria, one for electrons located in the barrel, and another for electrons located in the endcap.

Variable	Barrel	Endcap
E_T	> 35	> 35
$\lvert\eta_{SC}\rvert$	$\lvert\eta\rvert < 1.442$	$1.56 < \lvert\eta\rvert < 2.5$
seed	ECAL seeded	ECAL seeded
Missing hits	≤ 1	≤ 1
$\lvert d_{xy}\rvert$[cm]	< 0.02	< 0.05
$\Delta\eta_{in}$	< 0.005	< 0.007
$\Delta\phi_{in}$	< 0.06	< 0.06
H/E	< 0.05	< 0.05
$\sigma_{i\eta i\eta}$	-	< 0.03
Shape	$E^{2\times5}/E^{5\times5} > 0.94$ OR $E^{1\times5}/E^{5\times5} > 0.83$	-
Calo iso. [GeV]	$< 2 + 0.3 \times E_T + 0.28 \times \rho$	$< 2.5 + 0.28 \times \rho$ if $E_T < 50$ $< 2.5 + 0.28 \times \rho + 0.03 \times (E_T - 50.)$ if $E_T > 50$
Tracker iso. [GeV]	< 5.0	< 5.0

TABLE 3.1: HEEP ID v4.1 selection criteria, for electron identification and isolation.

These parameters range from kinematic requirements, such as E_T, to shower shape requirements, such as $E^{2\times5}/E^{5\times5}$. In depth, these variables are defined as:

- E_T**:** The transverse energy of the electron. In this case it is the calibrated energy measurement from the ECAL supercluster, multiplied by $sin(\theta_{track})$, where θ_{track} is the polar angle of the electron GSF track measured in the innermost layer of the tracker, and extrapolated to the interaction vertex. While the weighted average between GSF and supercluster energy provides better performance for low energy electrons, at high energies the weighted average is nearly identical to the supercluster alone. In rare situations the weighted average discards the supercluster entirely, and uses the GSF track measurement alone, which can lead to misassigning very high energy to low energy electrons. Thus, the HEEP selection uses only the supercluster energy.

- $|\eta_{SC}|$**:** The pseudorapidity of the supercluster used in the reconstruction of the electron candidate. Note that this not the pseudorapidity of the track, and electrons located near the gap between the barrel and endcap (where the reconstruction efficiency and energy resolution are poor) are not considered.

- **Seed:** As discussed in Section 3.3.1, two electron seeding algorithms are used in CMS. The HEEP ID v4.1 uses only electrons that are ECAL-driven.

- **Missing hits:** The number of expected hits in the tracker that are actually missing from the GSF track of the electron candidate. Electrons produced via photon conversion can often possess missing hits within the tracker.

- $|d_{xy}|$**:** The transverse impact parameter, as in the muon criteria, the distance of the closest approach in the $x - y$ plane of the track with respect to the primary vertex.

- $\Delta\eta_{in}$**:** The difference between the pseudorapidity, as measured from the GSF track at the innermost layer of the tracker and extrapolated to the interaction vertex, and measured from the electron's supercluster (η_{SC}). This is required to be close to zero.

- $\Delta\phi_{in}$**:** The difference between the ϕ value of the electron, as measured from the GSF track at the innermost layer of the tracker and extrapolated to the interaction vertex, and measured from the electron's supercluster (η_{SC}). As in the case of $\Delta\eta_{in}$, this is required to be close to zero.

- H/E**:** The ratio of the energy measured by the HCAL inside a cone of radius $\Delta R \equiv \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.15$ centered on the supercluster position over the energy of the electron's supercluster. This is required to be small.

- $\sigma_{i\eta i\eta}$**:** The measure of the spread in pseudorapidity of the electron energy inside a 5×5 block of ECAL crystals, centered on the seed crystal. The spread of electron energy is expected to be narrow in η.

- $E^{2\times5}/E^{5\times5}$**:** The fraction of the total electron energy within a block of 2 crystals in the η-direction by 5 crystals in the /*phi*-direction. To reduce background contamination from electrons that are contained in gets, this is required to be large.

- **Calo iso:** This is the isolation of both the ECAL and the HCAL, the sum of the energy in the ECAL and in the "HCAL depth 1". The ECAL isolation sum is the sum of the transverse energy of all the barrel hits with at least $E > 0.08\,\text{GeV}$ and all endcap hits with at least $E > 0.1\,\text{GeV}$, in a cone of radius $\Delta R \equiv \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.3$. An inner of three by three crystals is excluded from the sum. The HCAL depth 1 energy is the sum

of the depth 1 hits in the same radius size, with an inner cone of radius $\Delta R = 0.15$ excluded. Depth 1 hits are hits with specific lengths in the HCAL towers, not necessarily penetrating the entire tower. The total sum of both the ECAL and HCAL energies in their respective cones comprise the calorimeter isolation value, which is required to be small.

- **Tracker iso:** The tracker isolation is the sum of the p_T of the tracks within a cone of radius $\Delta R = 0.3$ centered around the GSF track with a minimum p_T of 0.7 GeV and minimum longitudinal distance from the interaction vertex of 0.2 cm. Tracks within an inner cone of $\Delta R = 0.04$ of the GSF track are not considered. As with the calorimeter isolation, this is required to be small.

3.4 Particle Flow Reconstruction and jets

The analysis described in this thesis, as well as many others performed at CMS, makes use of the particle flow (PF) event reconstruction method for for jets and missing transverse energy (E_T^{miss}). This algorithm combines information from all CMS subdetectors in order to identify all stable particles in the event, *i.e.*, electrons, muons, photons, and both charged and neutral hadrons. This list of individual particles is then used to build jets, to determine the E_T^{miss}, and to reconstruct and identify taus from their decays products. It is also used to quantify charged lepton isolation with respect to other particles and to tag identify jet flavor. The PF algorithm begins via identification of "fundamental elements", charged particle tracks and calorimetric energy clusters. These are then topologically linked to form "blocks". Finally, the PF algorithm interprets these blocks in terms of particles, and completes the jet building and E_T^{miss} calculation [54].

3.4.1 Fundamental elements

To create the charged particle fundamental element, the charged particle trajectories created via iterative tracking discussed in Section 3.1 are used. The momentum of these charged hadrons is measured by the tracker with a far superior resolution to that of the calorimeters for p_T values up to at least several hundreds of GeV. The tracker also provides a precise measurement of the charged particle's direction at the production vertex. The tracking fake rate must be kept small

since fake tracks, with randomly distributed momentum, could lead to large energy excesses. Thus, the iterative tracking method is essential, with the tracks both seeded and reconstructed with very tight tracking criteria. Successive iterations that use softer criteria improve the efficiency to avoid charged hadrons that become solely detected by the calorimeters, thus leading to a degraded energy resolution.

The calorimetric energy fundamental elements are clusters that are formed from energy deposits in both the ECAL and HCAL. The purpose of the clustering algorithm is to detect and measure the energy and direction of neutral particles, to separate neutral particles from charged particles, to reconstruct and identify electrons along with their accompanying Bremsstrahulung photons, and to help the energy measurement for the charged hadrons for which the track parameters were not computed accurately (for low-quality or high-p_T tracks). The algorithm consists of three steps, starting with cluster seeds which are simply local calorimeter cells with energy maxima above a given threshold. Topological clusters are grown from the seeds by aggregating cells with at least one side in common with a cell already in the cluster. Then these clusters are used as seeds are used as seeds to the PF flow algorithm.

3.4.2 Link algorithm

Any given particle can potentially yield several fundamental elements: one charged-particle track, several calorimeter clusters, and one muon track. Thus these fundamental elements must be linked via a linking algorithm to both fully reconstruct each particle, and to remove any possible double counting. The link algorithm is performed for each pair of elements in the event and defines a distance between each one to quantify the quality of the link, and produces blocks of elements linked either directly or indirectly.

A link between a charged-particle track and calorimeter cluster is created by extrapolating the track from its outermost hit position in the tracker to the ECAL and the HCAL. Its extrapolated in the ECAL to a depth that corresponds to the typical expected maximum of a longitudinal electron shower profile, and in the HCAL to a depth corresponding to one interaction length typical of a hadron shower. If the position in either corresponding calorimeter is within the cluster boundaries, then a link is created. The envelope is typically enlarged by up to

the size of a single cell in each direction in order to account for gaps between cells, cracks in the intervening space between calorimeter modules, and for the effect of multiple scattering on low-momentum charged particles and uncertainty on the position of the shower maximum. The link distance in this case is defined as the separation between the extrapolated track position and cluster position in the (η, ϕ) plane.

A link is created between two calorimeter clusters when the cluster position in the more granular detector (ECAL) is located within the cluster envelope of the less granular detector (HCAL). As in the case of the track-cluster link, the envelope can be slightly enlarged, and the link distance is defined as the distance between the two cluster positions in the (η, ϕ) plane.

Finally, a link is created between a charged-particle track in the tracker and a muon track in the muon system when the global fit between the two tracks passes a requirement on the value of χ^2. In the case of multiple possible fits, only the pairing with the smallest χ^2 is linked. For this link the χ^2 value itself is used as the link distance.

3.4.3 PF algorithm

Once the link algorithm creates each block, the particle flow algorithm then examines each block and steps through particle types successively to describe all the elements. First, each global muon results in a PF muon if the combined momentum is compatible within three standard deviations of the tracker-only momentum. This track is then removed from the block. Then electrons are initially identified first by their short pre-shower tracks in the tracker, this is called the pre-identification stage. These pre-identified electrons are in turn refit with the GSF method in order to track the trajectories throughout the entirety of the ECAL. Final identification of electrons is performed with a combination of parameters in the tracker and ECAL, yielding PF electrons which are also removed from the block.

After muons and electrons are removed from the block, tighter quality criteria are applied to the remaining tracks in the block, requiring that the relative uncertainty of the measured p_T be less than the relative calorimetric energy resolution expected

for charged hadrons [54]. These elements then give rise to either charged hadrons, photons, or neutral hadrons (or, rarely, additional muons).

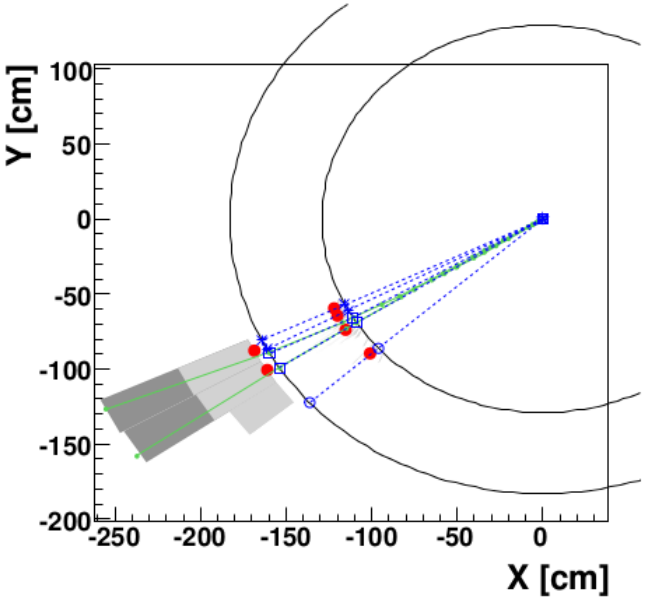
Neutral particles (both neutral hadrons and photons) are identified via a comparison between track momentum and energy detected in the calorimeters - which makes use of a calibrated sum of the ECAL and HCAL energies [54].

In this comparison, several tracks could be linked to the same HCAL cluster, in which the sum of their momentum is used for the comparison to the HCAL cluster. Conversely, if a single track is linked to multiple HCAL clusters, only the closest cluster is used. The same is true for multiple ECAL clusters, although then the possibility and then the decision whether to keep the links to other ECAL clusters that may be from hadronic shower fluctuations or to reject those that may be overlapping photons is taken. This decision is taken via ordering the linked clusters by distance to the closest track and sequentially keeping each cluster as long as the total calibrated calorimetric energy remains smaller than the total charged-particle momentum.

In the case where the total calibrated calorimetric energy is still smaller than the total track momentum by more than three standard deviations, a search for muons and for fake tracks is performed, with relaxed criteria. First, any global muon not already selected by the algorithm for which the momentum estimate possesses a precision of at least 25% is treated as a PF muon. Tracks are sequentially removed from the block, ordered according to their p_T uncertainty, until either all tracks with p_T uncertainties over 1 GeV are examined or when the removal of a track would cause the total track momentum to be smaller than the calibrated calorimetric energy. This procedure is needed in only very rare cases, affecting less than 0.3 per mil of the tracks used in the PF reconstruction method.

Each of the remaining tracks in the block result in PF charged hadrons, in which the momentum and energy of the charged particles is taken directly from the track momentum, under the charged pion mass hypothesis. If the calibrated calorimetric energy matches the track momentum within uncertainties, the momentum is redefined via a fit of the measurements in both the track and the calorimeters. If the closest ECAL and HCAL clusters linked to the track have calibrated energies significantly larger than the associated track, a PF photon or PF neutral hadron is created.

Finally, remaining ECAL and HCAL clusters, either not linked to any track or for which the link was removed, yield PF photons or PF neutral hadrons, respectively. Also, E_T^{miss} is expressed as the magnitude of the negative vector sum of the p_T of each PF particle. An example event display for a jet built with the PF algorithm is given in Figure 3.12. The PF particles created using the PF algorithm are also used in clustering to create jets, described in Section 3.4.4.



(a) The (x, y) view

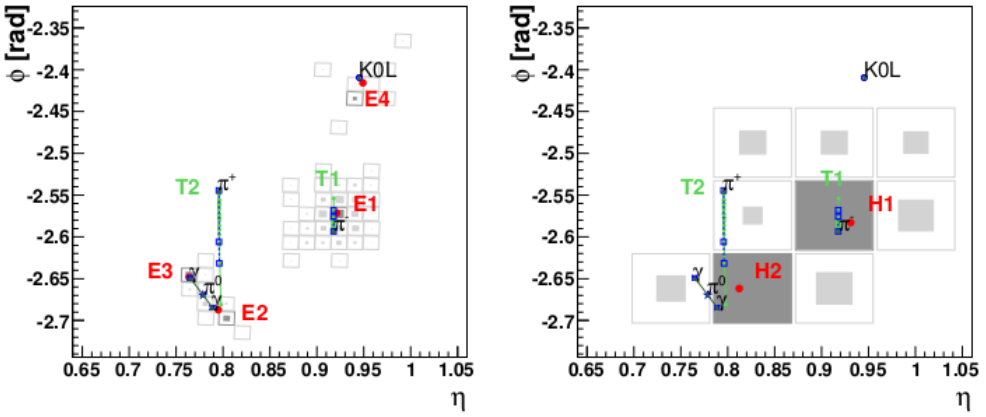


FIGURE 3.12: An event display for a simple hadronic jet, shown in the (x, y) view (a) and the (η, ϕ) view on both the ECAL surface (b) and the HCAL surface (c). There are four constituent particles detected, a K_L^0, a π^-, and two photons from a π^+ decay. The two photons are visible in the ECAL (b) but not the HCAL (c). The two charged pions are reconstructed as charged-particle tracks, and are shown as vertical solid lines in both (η, ϕ) views and as circular arcs in the (x, y) view. These tracks clearly point towards two HCAL clusters (c). The cluster positions in all three views are represented by dots and the simulated particles by dashed lines. [54].

3.4.4 Jet clustering, energy correction, and identification

Once all the PF particles have been defined the charged hadrons, neutral hadrons, and other possible jet-constituents must be combined into jets. At CMS this can be done with calorimeter deposits alone (the resulting particles are called "Calo jets") or from particles reconstructed via the PF algorithm, resulting in PF jets. In the following discussion of how these PF particles are grouped in to jets, they will be referred to collectively as "objects".

In general there are two methods of grouping objects into jets: a cone based method which seeks to search for stable cones and can include a split-merge step to disentangle overlapping cones (SISCone is one such algorithm [55, 56]), and an iterative pair-wise clustering method. The main method used by many analyses at CMS, and the one presented in this thesis, is called the anti-k_T algorithm. As with any clustering algorithm, it introduces a distance between objects i and j, d_{ij}, and between object i and the beam, d_{iB}. It iteratively steps through the list of objects in an event, and if d_{ij} is less than d_{iB}, it combines the two objects i and j into a pseudojet object and continues the clustering with the next object, or if d_{iB} is less than d_{ij} the algorithm declares object i a jet and removes it from the list. These distances are recalculated and the entire procedure is repeated until no objects remain. In general, the distinctions between methods lie in the definitions of the distance measures:

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2}, \tag{3.1}$$

$$d_{iB} = k_{ti}^{2p}, \tag{3.2}$$

where $\Delta_{ij}^2 = (y_i - j_j)^2 + (\phi_i - \phi_j)^2$ and k_{ti}, y_i, and ϕ_i are the transverse momentum, rapidity, and azimuthal angle of particle i, respectively. The radius parameter R provides the angular size of the jets, and the power p governs the relative power of the energy versus the geometrical (Δ_{ij}^2) scale. The standard inclusive k_T algorithm is the case in which $p = 1$, and the inclusive Cambridge/Aachen algorithm is the special case in which $p = 0$. While negative values of p might seem counter intuitive, the case where $p = -1$ is actually the definition of the anti-k_T algorithm, and it performs well comparatively to the other methods, as shown in Figure 3.13.

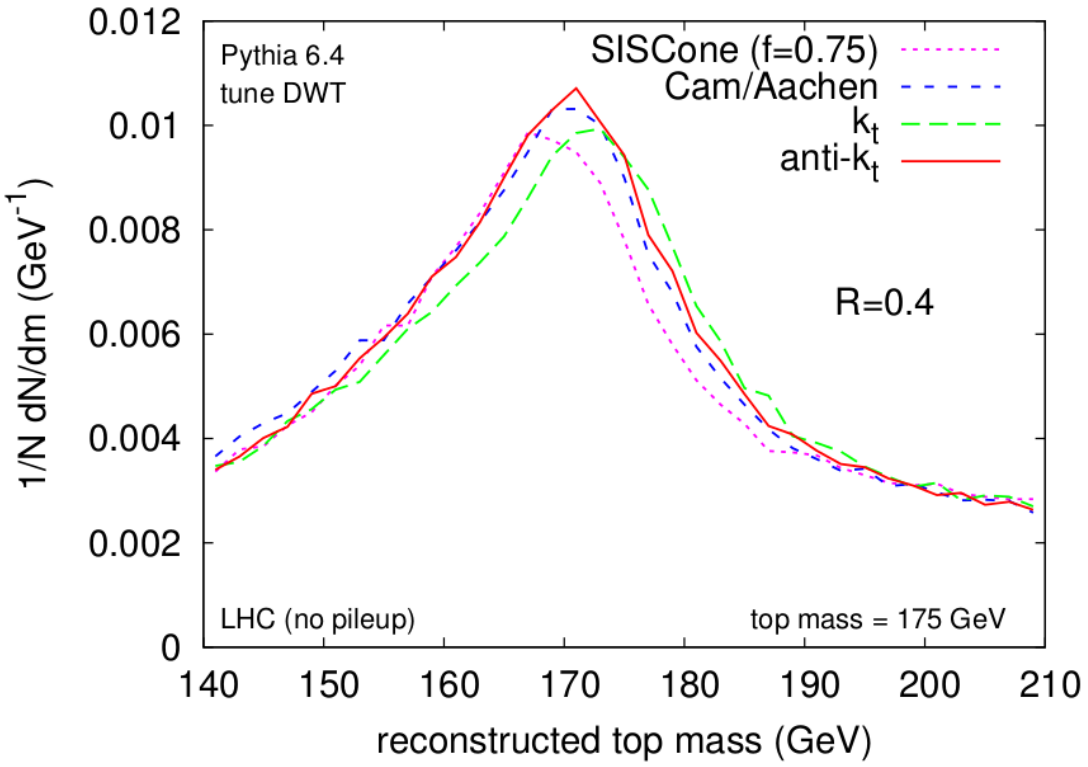


FIGURE 3.13: Reconstructed top mass in simulated LHC $t\bar{t}$ events. Both the t and the $\bar{t}$ decay to a b quark and W boson which in turn decays hadronically resulting in a final state of with two b quarks and 4 other quarks. The four non-b jets are paired according to the combination that best reproduces the W masses and then those pairs are paired with b-jets to minimize the mass-difference between the two t-jets. Four algorithms are compared, the SISCone cone-based algorithm, the Cambridge/Aachen $p = 0$ algorithm, the simple k_T algorithm with $p = 1$, and the anti-k_T algorithm with $p = 1$. All algorithms used a value of $R = 0.4$, and the marginal improved performance of the anti-k_T algorithm is shown. [55].

The energies of the reconstructed jets are corrected in order to improve the correspondence between the reconstructed energy value and the true particle level energy value. The jet energy is calibrated using the p_T balance of dijet and γ+jet events to account for a variety of factors possibly affecting the measurement [57]. In total, the complete correction is applied as a multiplicative factor C to each component of the raw four-momentum vector p_μ^{raw} of the jet:

$$p_\mu^{cor} = C \cdot p_\mu^{raw}. \tag{3.3}$$

The correction factor C is composed of various factors intended to correct for different phenomena:

- The offset correction C_{offset} removes extra energy from either noise or pile-up interactions.
- The simulation correction C_{sim} accounts for the bulk of the non-uniformity in η and non-linearity in p_T of the calorimeter response.
- Relative and absolute residual corrections C_{rel} and C_{abs} account for the small differences between data and simulation, such as resolution biases introduced by the clustering algorithm.

All the components are applied in sequence as described in Equation 3.4:

$$C = C_{offset}(p_T^{raw}) \cdot C_{sim}(p'_T, \eta) \cdot C_{rel}(\eta) \cdot C_{abs}(p''_T), \tag{3.4}$$

where p'_T is the transverse momentum of the jet after applying only the offset correction and p''_T is the transverse momentum after applying all previous corrections (other than the residual absolute energy scale resolution). The overall corrections for PF jets and a couple other jet types are shown in Figure 3.14.

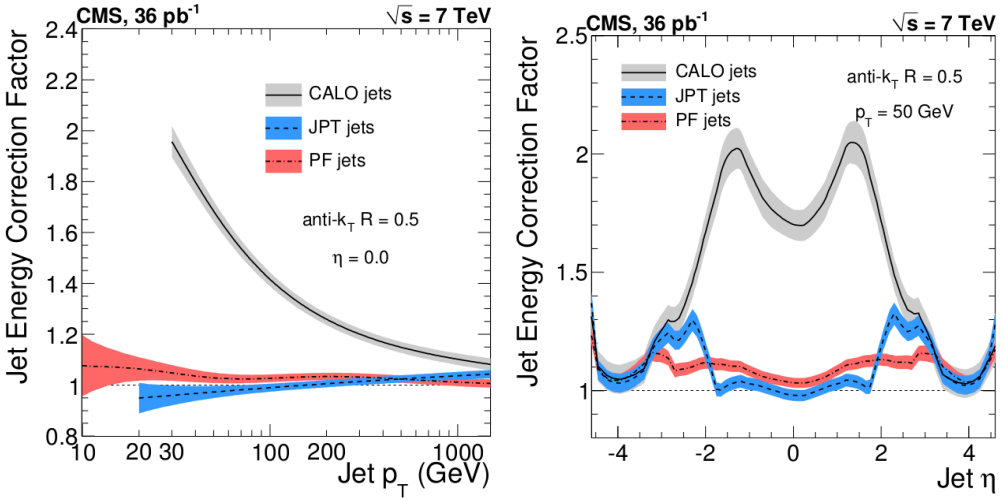


FIGURE 3.14: Total jet-energy corrections factors, as a function of p_T for $\eta = 0.0$ on the left and as a function of η for $p_T = 50\,\text{GeV}$ on the right. It is shown for Calo Jets (jets reconstructed with only calorimeter deposits), JPT jets (jets reconstructed using track information in addition to calorimeter information), and PF jets. The bands show the corresponding uncertainties of the measurements, which do not exceed 15%, for jets with p_T values of at least 10 GeV. [57]

As with muons and with electrons, after reconstruction various selection criteria can be applied to PF jets for identification and use in physics analyses. One such set of selection criteria is called the "loose" jet ID, and is defined below:

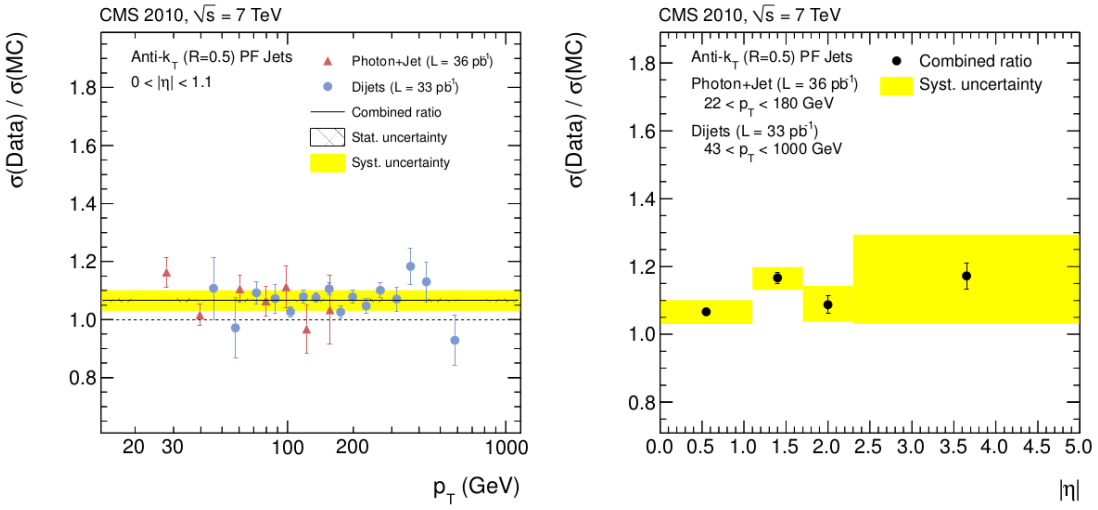


FIGURE 3.15: The ratio of jet p_T resolutions in data and simulation samples versus p_T with $|\eta| < 1.1$ (left) as measured in the dijet and γ+jet samples, with a combined fit to both datasets. The result from the combined fits versus p_T is shown in various η ranges (right). [57]

- The fraction of the jet energy corresponding to neutral hadrons must be less than 0.99.
- Likewise, the fraction of jet energy corresponding to neutral electromagnetic energy deposits must also be less than 0.99.
- There must be at least two constituents clustered to create the jet.

For jets with $|\eta| < 2.4$ the following additional criteria are applied:

- The fraction of the jet energy corresponding to charged hadrons must be greater than zero.
- The fraction of the jet energy corresponding to charged electromagnetic energy deposits must be less than 0.99.
- There must be at least one charged constituent in the clustering to create the jet.

The loose jet ID is fully efficient ($> 99.9\%$) for real high p_T jets, its background rejection is also high, and it is used in most physics analyses at CMS as well as the analysis presented in this thesis. Some analysis use the "tight" jet ID, for which rather than requiring that the fractions of jet energy corresponding to neutral hadrons and neutral electromagnetic energy deposits be less than 0.99, they are both requered to be less than 0.90.

3.5 Event simulation and detector response

Simulation of events is an essential component of measurements and searches for exotic processes at CMS. For searches, the signal process must be simulated as well as any backgrounds for which data-driven measurements are not possible. For measurements and studies such as those for future upgrades, or for detector performance, likewise specific processes must be simulated. There are two significant portions of event simulation: generation of the actual physics processes in the event, and simulation of the various detector subsystems' response to the final state particles produced by those processes. For the first step, programs called "event generators" are used.

Commonly used event generators in CMS include MadGraph [58], powheg [59–62], CalcHEP [63], and pythia [64]. All event generators use specific parton distribution functions (PDFs) to describe the distribution of the various partons within the protons involved in the collisions. pythia and MadGraph use the leading order CTEQ6L1 [65] PDF set, CalcHEP uses the CTEQ6L PDF set, and powheg uses the CTEQ6m PDF set. The powheg, MadGraph, and CalcHEP [63] event generators are effectively simply matrix element calculators, and thus only produce the undecayed final state particles. They are interfaced with pythia to handle parton showering and hadronization for the quarks. Additionally, in pythia matching between jets and partons is performed using the k_T-MLM scheme [66].

For any given simulation once the final state particles have been determined, a full simulation of the CMS detector's response to those particles must be produced in order to perform the standard reconstruction for the events. The simulation is performed with the Geant4 package [67] and the CMS object-oriented framework, for all subsystems within CMS. The Geant4 package includes both the simulation of all hadronic and electromagnetic interactions as well as the geometry of the CMS detector. Once the physical response has been simulated, the individual electronic response for each subsystem's readout electronics is performed in a step called digitization. This produces effectively the same signals those subsystems would see during data-taking, referred to as "digis". It is these data objects that then in turn are fed in to the standard reconstruction algorithms.

It is also necessary to simulate the effect of pileup, additional pp interactions taking place in the same bunch crossing as the interaction of interest (in-time pileup) and

interactions in the detector from neighboring bunches (out-of-time pileup). Due to the fact that conditions in the LHC are continuously changing from run to run, it is not possible to know the mean number and profile of pileup interactions in advance. As a result, pileup events are added to the events according to a minimum-bias sample regarding the expected profile the running period - and after that all analyses at CMS additionally reweight the pileup profile to match the actual one measured in data.

Chapter 4

The Search for Leptoquarks

4.1 Theoretical issues

As discussed in Section 1.2.2, while the main mode of single leptoquark production, significant contributions are made by diagrams with non-resonant components. These contributions increase with both the LQ mass and coupling; the invariant mass distribution of a first generation LQ, of mass $M_{\mathrm{LQ}} = 1$ TeV and coupling $\lambda = 1.0$, possesses a tail extending to very low masses that is comparable to the peak in magnitude. The reconstructed shape of the resonance peak itself is not strongly affected by λ. Even for very low LQ mass, these off-shell decays are visible in the $M_{\ell q}$ spectrum, as shown for a first generation LQ hypothesis of $M_{LQ} = 300\,\mathrm{GeV}$, for two couplings in Figure 4.1. Contrasting that with the same spectrum for $M_{LQ} = 1000\,\mathrm{GeV}$ in Figure 4.2, one can see that at higher mass and coupling values the off-shell tail can become significant with respect to the peak.

Additionally, interference with the $qg \to qZ/\gamma^* \to q\ell^+\ell^-$ SM process can occur at dilepton masses in the vicinity of the Z boson mass peak and at lower energies. Treatments for this interference region and the above-described low-mass off-shell tail of the lepton-jet mass distribution are necessary. In order to study the single LQ mass peak alone, a selection is performed at the generator-level to remove the bulk of the off-shell tail and the interference region. This selection is described in further detail in Section 4.4.1.

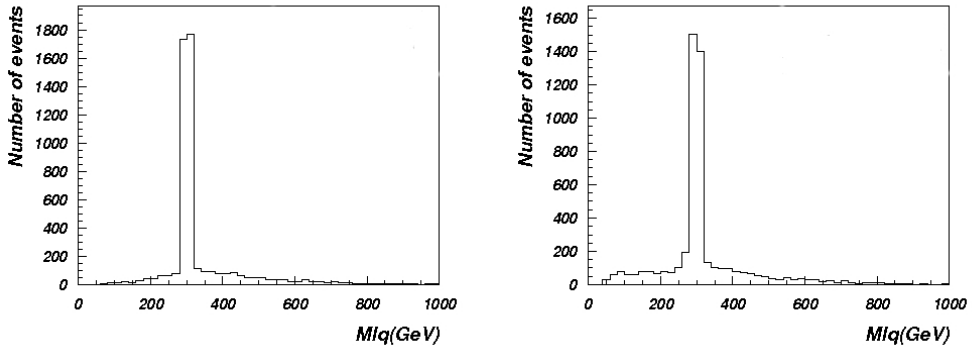


FIGURE 4.1:
M_{eq} as simulated in CALCHEP for a first generation LQ, with $M_{eq} = 300\,\text{GeV}$ and $\lambda = 0.2$ (left), and with $M_{eq} = 300\,\text{GeV}$ and $\lambda = 1.0$.

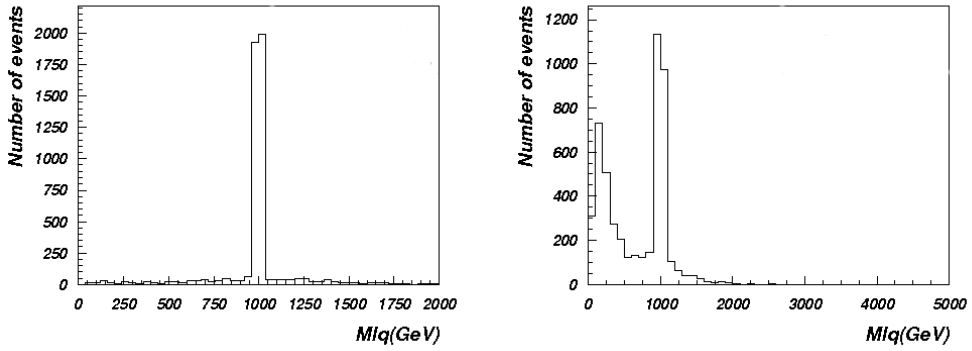


FIGURE 4.2:
M_{eq} as simulated in CALCHEP for a first generation LQ, with $M_{eq} = 1000\,\text{GeV}$ and $\lambda = 0.2$ (left), and with $M_{eq} = 1000\,\text{GeV}$ and $\lambda = 1.0$.

4.2 Analysis Strategy

Single LQ production presents an attractive target for searches at the LHC for various reasons. As discussed in Section 1.1.2, LQs are components to many models of physics beyond the SM. Probing single LQ production also provides a means to determine coupling-dependent constraints on LQ models. Finally, single LQ production has a very significant production cross section at a pp collider such as the LHC, even becoming larger than pair production for certain values of coupling. Single LQ production also provides a final state with a pair of well-isolated and energetic electrons or muons (assuming nonzero values of β).

The analysis presented herein takes β to be 1.0, and studies multiple values of λ. It considers two channels, one for first- and one for second-generation production.

Thus the final states considered are eej and $\mu\mu j$. Due to the suppression of the second-generation LQ production (Section 1.2.2), only one, relatively high, value of λ is considered.

The basic strategy for each channel is fundamentally the same, differing mainly in the trigger used to select events. The eej channel uses a double-electron trigger, whereas the $\mu\mu j$ channel uses the standard single-muon trigger that was used by most analyses for studying high p_T muons at the LHC. Thereafter, kinematic requirements are essentially the same between the channels, requiring at least two leptons and at least one jet, and additionally applying thresholds on p_T values high enough to guarantee good quality physics objects and low enough to define a region dominated by SM backgrounds. Once the backgrounds, primarily those described in Section 1.2.2 that can mimic the single LQ production signal, are well understood in this "preselection" region, a search for the optimal final selection is performed.

The final selection is optimized for each channel separately by maximizing $S/\sqrt{S+B}$, where S is the number of signal events in the simulation passing a given selection and B is the number of background events in the simulation passing the same selection. We optimize for each LQ mass hypothesis by varying the requirements on $M_{\ell j}$, defined as the higher of the two possible lepton-jet mass combinations, and S_{T}, the scalar sum of the p_T of the two highest p_T leptons and the p_T of the highest p_T jet.

4.3 Data and simulation samples

The data were collected during all four 8 TeV pp run periods in 2012 at the CERN LHC and correspond to a total integrated luminosity of $19.6\,\mathrm{fb}^{-1}$. In the ee$j$ channel, events are selected using a trigger that requires two electrons with $p_{\mathrm{T}} > 33\,\mathrm{GeV}$ and $|\eta| < 2.4$ and in the $\mu\mu j$ channel, events are selected using a trigger that requires one muon with $p_{\mathrm{T}} > 40\,\mathrm{GeV}$ and $|\eta| < 2.1$.

Simulated samples for the signal processes are generated for a range of leptoquark mass hypotheses between 300 and 3300 GeV and coupling hypotheses between 0.4 and 1.0 in the ee$j$ channel, and a range of leptoquark mass hypotheses between 300 and 1800 GeV and a coupling hypothesis of 1.0 in the $\mu\mu j$ channel. Production

of LQs in the $\mu\mu j$ channel is suppressed because of the proton PDF as discussed in Section 1.2.2.

The main sources of background are $t\bar{t}$, Z/γ^*+jets, W+jets, diboson (ZZ, ZW, WW)+ jets, single top quark, and QCD multijet production. The shape of the kinematic distributions of the $t\bar{t}$ + jets background is estimated from a study based on data described in Section 4.5; the simulation sample for the normalization of the $t\bar{t}$+jets background as well as the samples for the Z/γ^*+jets and W+jets backgrounds are generated with MadGraph 5.1 [58]. Single top quark samples (s-, t-channels, and W boson associated production) are generated with powheg 1.0 [59–62] and diboson samples are generated with pythia (version 6.422) [64] using the Z2 tune [68]. The QCD multijet background is estimated from data.

For the simulation of signal samples, the CalcHEP [63] generator is used for calculation of the matrix elements. The signal cross sections are computed at leading order (LO) with CalcHEP and are listed in Table A.1 in the appendix. Blank entries were not considered because of the small size of the cross section. The resonant cross sections σ_{res} are shown in Fig. 4.3 and are defined by the kinematics selections given in Section 4.4.

The pythia and MadGraph simulations use the CTEQ6L1 [65] PDF sets, those produced with CalcHEP use the CTEQ6L PDFs, and the powheg simulation uses the CTEQ6m set. All of the simulations use pythia for the treatment of parton showering, hadronization, and underlying event effects. For both signal and background simulated samples, the simulation of the CMS detector is based on the Geant4 package [67]. All simulated samples include the effects of extra collisions in a single bunch crossing as well as collisions from nearby bunch crossings (out-of-time pileup and in-time pileup, respectively).

In the ee$j$ channel, the background and signal are rescaled by a uniform trigger efficiency scale factor of 0.996, which is measured in [69]. In the $\mu\mu j$ channel, the background and signal are rescaled by muon η-dependent efficiency factors of 0.94 ($|\eta| \leq 0.9$), 0.84 ($0.9 < |\eta| \leq 1.2$), and 0.82 ($1.2 < |\eta| \leq 2.1$). An uncertainty of 1% is assigned to these factors to account for variations during data-taking periods and statistical uncertainties.

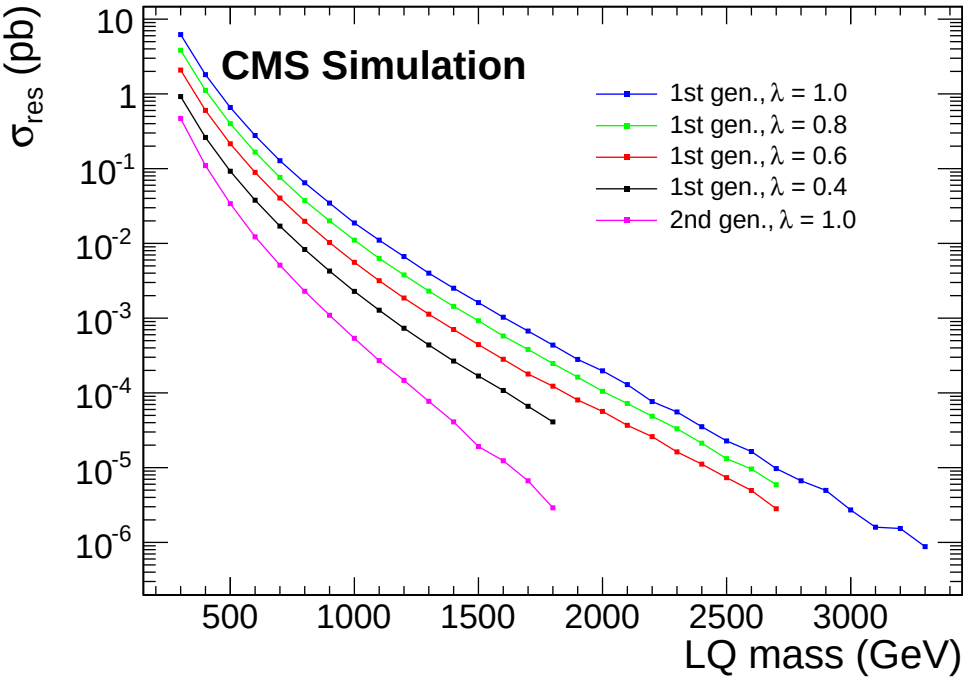


FIGURE 4.3: Cross sections for single LQ production, calculated at LO in CALCHEP and scaled by the acceptance of the requirements described in Section 4.4.1, as a function of the LQ mass in GeV.

4.4 Event selection

The process of selecting events for analysis can be broken down in to three fundamental selections. First, a resonant region is created that isolates the resonant part of the LQ signal from the offshell and SM interference regions. This selection is applied to the simulated LQ signal events only. Then, the preselection region is defined by applying quality criteria to the physics objects in the events and defining the kinematic region of interest. Finally, a search for the optimal final selection is performed for each LQ mass.

4.4.1 Resonant selection

As discussed in Section 1.2.2, owing to the unique aspects of single LQ decays, two generator-level requirements are applied to the simulated signal samples. The first is $M_{\ell\ell} > 110\,\text{GeV}$, to remove LQ decays that are in the Z boson interference region. The second is a requirement on $M_{\ell j}$, defined as the higher of the two possible lepton-jet mass combinations, chosen to remove the t-channel diagram

contributions in the low-mass off-shell region, while preserving most of the resonant signal. This requirement is set at $M_{\ell j} > 0.67\, M_{\mathrm{LQ}}$ for the first-generation studies and $M_{\ell j} > 0.75\, M_{\mathrm{LQ}}$ for the second-generation studies. The thresholds for $M_{\ell j}$ were chosen separately for each channel, because of the differences in the distribution shape. Selection regions for sampled LQ masses are given for first generation in Figure 4.4 and for second generation in Figure 4.5. The dilepton invariant mass requirement at the generator level precisely matches the reconstruction level requirement at the preselection. Together, these two requirements define the resonant region. Cross sections at the generator level before and after these requirements are provided in Table A.1, in the appendix.

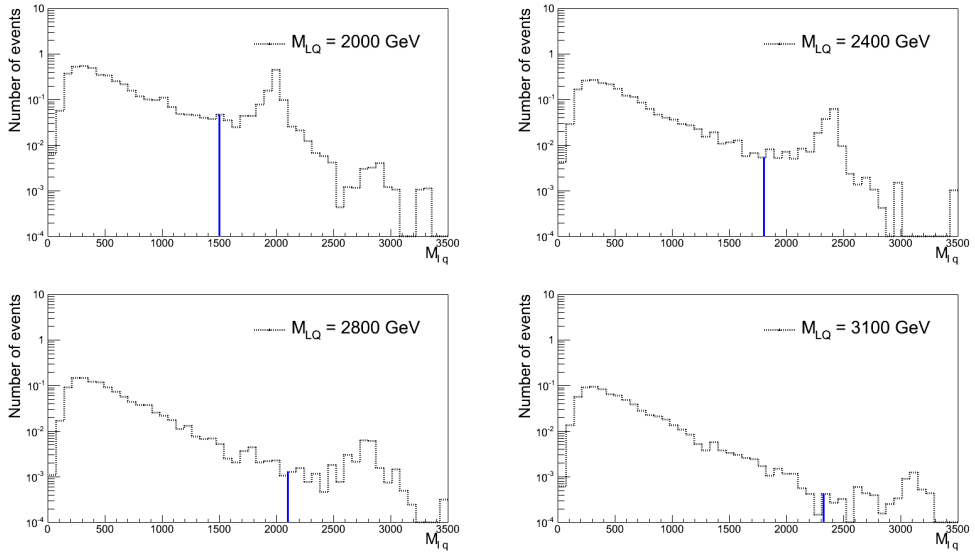


FIGURE 4.4: First generation LQ invariant mass distributions with the cut location of $M_{\mu j} > 0.67\, M_{\mathrm{LQ}}$ indicated by the thick vertical blue lines. Distrubutions for $M_{\mathrm{LQ}} = 2000\,\mathrm{GeV}$ (upper left), $M_{\mathrm{LQ}} = 2400\,\mathrm{GeV}$ (upper right), $M_{\mathrm{LQ}} = 2800\,\mathrm{GeV}$ (lower left), and $M_{\mathrm{LQ}} = 3100\,\mathrm{GeV}$ (lower right) are shown.

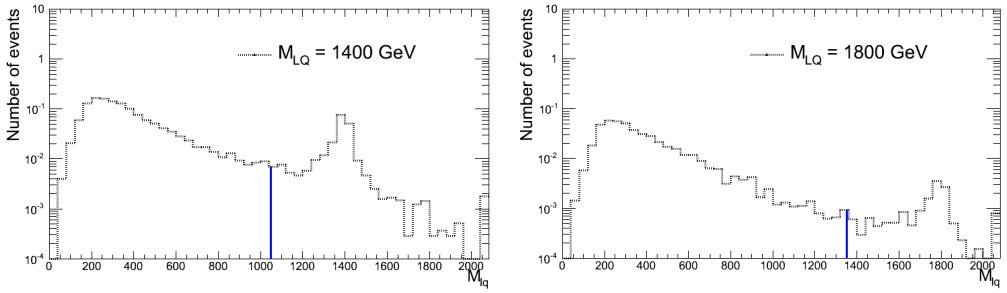


FIGURE 4.5: Second generation LQ invariant mass distributions with the cut location of $M_{\mu j} > 0.75\, M_{\mathrm{LQ}}$ indicated by the vertical blue lines. The distribution on the left is for an LQ with $M_{\mathrm{LQ}} = 1400\,\mathrm{GeV}$ and the right is for an LQ with $M_{\mathrm{LQ}} = 1800\,\mathrm{GeV}$.

4.4.2 Preselection

In all events with muons in the final state, the muons are required to be high p_T muons as described in Section 3.2.2. Electrons are required to satisfy the HEEP v4.1 criteria described in Section 3.3.2. Jets are reconstructed with the PF algorithm described in Section 3.4, are clustered via the anti-k_T algorithm with a distance parameter of 0.5, and must pass the loose jet ID described in Section 3.4.4. Additionally, jets involved in the analysis must pass these additional criteria:

- $p_T > 45\,\text{GeV}$
- $|\eta| < 2.4$
- Spatial separation from leptons of $\Delta R > 0.3$

The first two requirements ensure high quality jets in the region of interest, and the third removes any possible double counting of leptons as jets. Finally, all the jet energy corrections described in Section 3.4.4 are applied.

Once the high quality physics objects (electrons and jets or muons and jets) are selected, the final kinematic selection defining the preselection region is applied:

- $p_{T\ell 1} > 45\,\text{GeV}$
- $p_{T\ell 2} > 45\,\text{GeV}$
- $|\eta_{\ell 1}| < 2.1$
- $|\eta_{\ell 2}| < 2.1$
- $p_{Tjet} > 45\,\text{GeV}$
- $|\eta_{jet}| < 2.4$
- $M_{\ell\ell} > 50\,\text{GeV}$
- $\Delta R_{\ell\ell} > 0.3$
- $S_T > 250\,\text{GeV}$

Two additional requirements are added for the sake of background studies. In order to reduce the background from QCD multijets production, and in turn to facilitate its study an opposite sign requirement is added:

- $Charge_{e1} \times Charge_{e2} = -1$

Also, for all background studies other than the Drell–Yan plus jets background normalization study, the Z boson mass peak is removed from the preselection region with the following criterion:

- $M_{\ell\ell} > 110\,\text{GeV}$

All told, these criteria define the preselection region.

4.4.3 Final Selection

After the preselection is defined, additional selection criteria are applied to maximize statstical significance. This optimization is performed with the simulated signal and backgrounds only, to reduce the possibility of introduction of bias. For every mass hypothesis, thresholds on $M_{\ell j}$ and S_T are varied, and a complete scan is performed to find the pair that maximizes $S/\sqrt{S+B}$. Throughout the course of the analysis, other parameters were included in the optimization scan, such as p_T, but resulted in negligible improvements to the significance.

Once the individual optimized selections were found in the scan, the resulting cuts were smoothed with a linear fit and kept constant for LQ masses above the point where low statistics resulted in fluctuations of the cuts. This is done since optimization can fall victim to overtraining, especially in the regions of high LQ mass where numbers of background and signal events become small. The fit and the threshold reduce the likelihood that the procedure chooses S_T and $M_{\ell j}$ thresholds based on fluctuations. This is shown in Figure 4.6 for the first generation selection and in Figure 4.7 for the second generation selection.

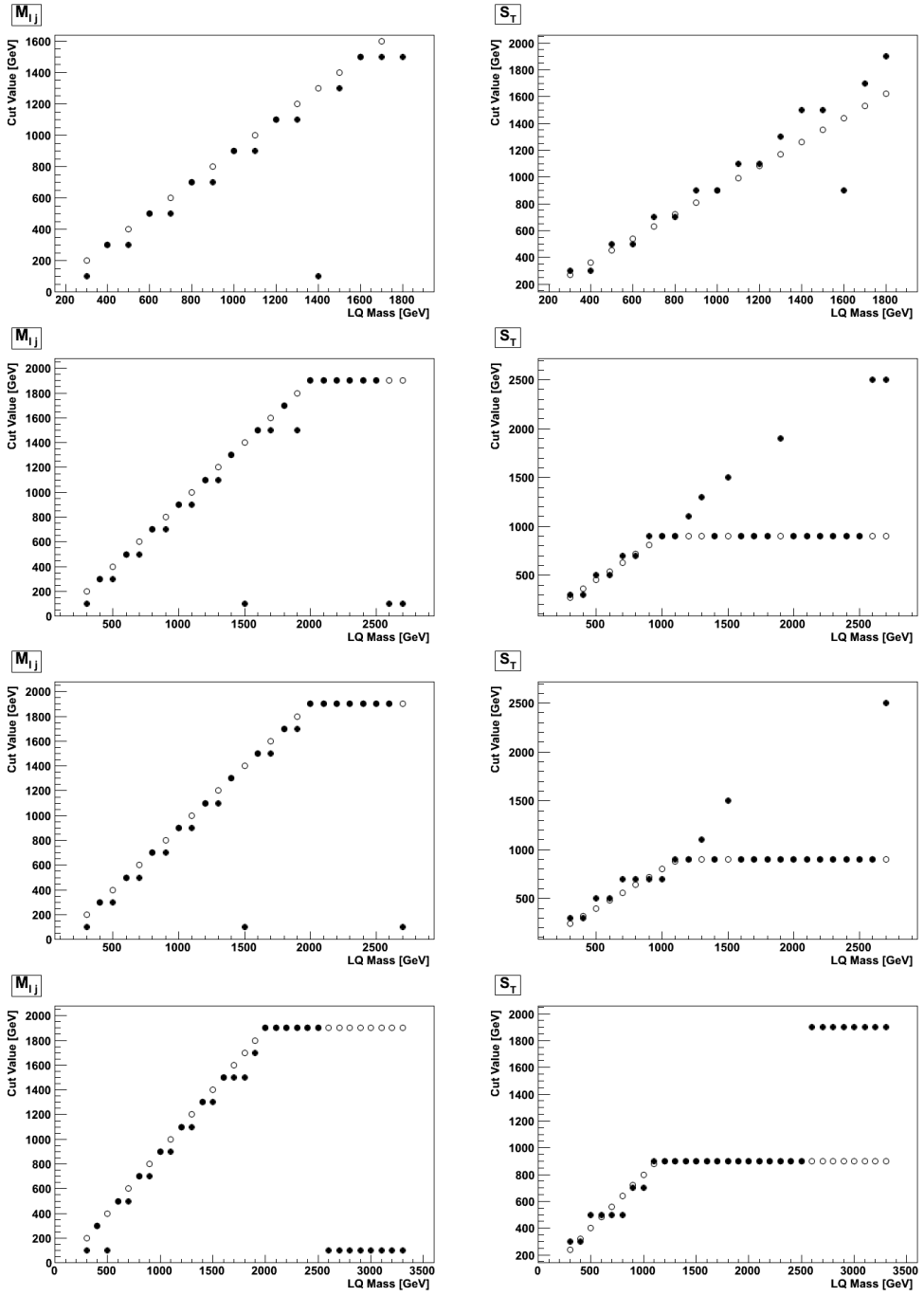


FIGURE 4.6: Final selection thresholds on M_{ej} (left) and S_T (right) vs LQ mass in GeV, for $\lambda = 0.4$ (first row), $\lambda = 0.6$ (second row), $\lambda = 0.8$ (third row), and $\lambda = 1.0$ (fourth row). The black dots represent the original scan results and the white dots are the adjusted values.

4.5 Background studies

The SM processes that mimic the signal signature are $Z/\gamma^* + \text{jets}$, $t\bar{t}$, single top quark, diboson +jets, W + jets, and QCD multijets events where the jets are

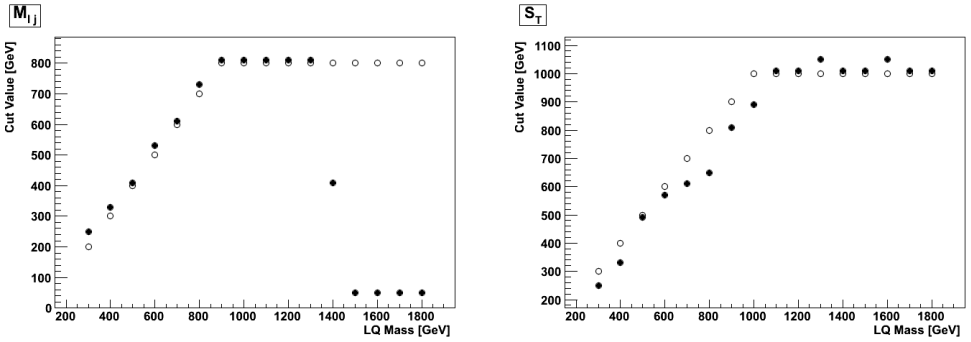


FIGURE 4.7: Final selection thresholds on $M_{\mu j}$ (left) and S_T (right) vs LQ mass in GeV, for $\lambda = 1.0$. The black dots represent the original scan results and the white dots are the adjusted values.

misidentified as leptons. The dominant contributions come from the first two processes, whereas the other processes provide minor contributions to the total number of background events.

4.5.1 Pileup interactions

Because it is impossible to know the mean number and profile of pileup interactions in CMS in advance, the pileup profile in simulated samples must be rescaled to the one measured in data. The main heuristic to determine if this procedure is performed properly is the comparison of the number of vertices in data and simulation after this rescaling is performed. Both the overall shape and the scale should agree, and this is shown for the eej channel in Figure 4.8 and for the $\mu\mu j$ channel in Figure 4.9.

4.5.2 Drell–Yan plus jets

4.5.2.1 Normalization

The contribution from the $Z/\gamma^* + \text{jets}$ background is estimated with a simulated sample that is normalized to agree with data at preselection in the Z-enriched region of $80 < M_{\ell\ell} < 100\,\text{GeV}$, where $M_{\ell\ell}$ is the dilepton invariant mass. With this selection the data sample (with non-$Z/\gamma^*+\text{jets}$ simulated samples subtracted) is compared to $Z/\gamma^* + \text{jets}$ in simulation. The resulting scale factor, representing the ratio of the measured yield to the predicted yield, is $R_Z = 0.98 \pm 0.01\,(\text{stat})$ in

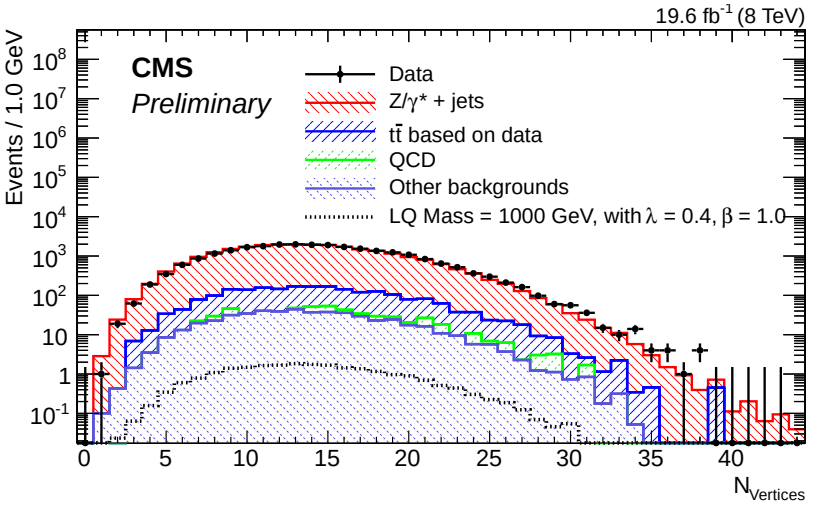


FIGURE 4.8: The distribution of the number of vertices in events passing the preselection criteria in the eej channel, after rescaling the simulation to data.

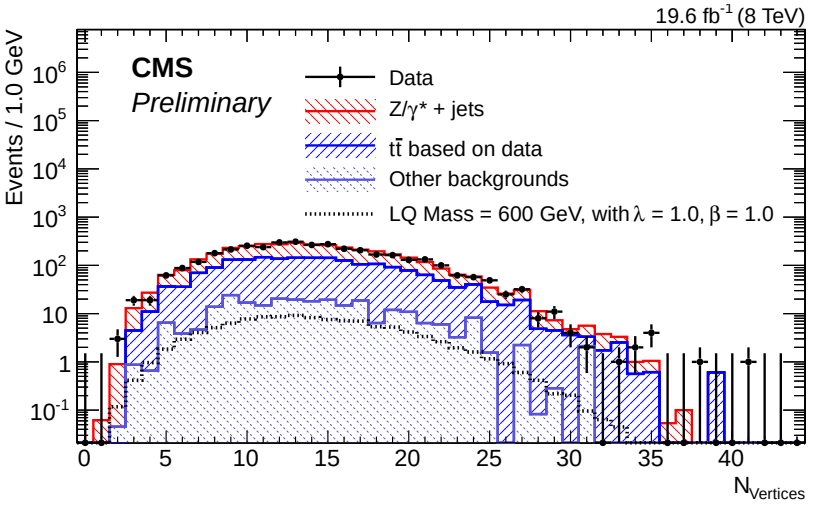


FIGURE 4.9: The distribution of the number of vertices in events passing the preselection criteria in the $\mu\mu j$ channel, after rescaling the simulation to data.

both the ee$j$ and $\mu\mu j$ channels. This scale factor is then applied to the simulated $Z/\gamma^* +$ jets sample in the signal region of $M_{\ell\ell} > 110\,\text{GeV}$.

4.5.2.2 Dijet transverse momentum

In order to account for possible mismodeling of the $p_{\text{T}}(\ell\ell)$ spectrum of the $Z/\gamma^* +$ jets background sample, where $p_{\text{T}}(\ell\ell)$ is the scalar sum of the two highest p_{T} leptons in the event, a bin-by-bin rescaling of yields at preselection and full selection is performed, with scale factors measured in an inverted $M_{\ell\ell}$ selection ($M_{\ell\ell} < 110\,\text{GeV}$), shown at preselection in Figures 4.12 and 4.13 for the eej and

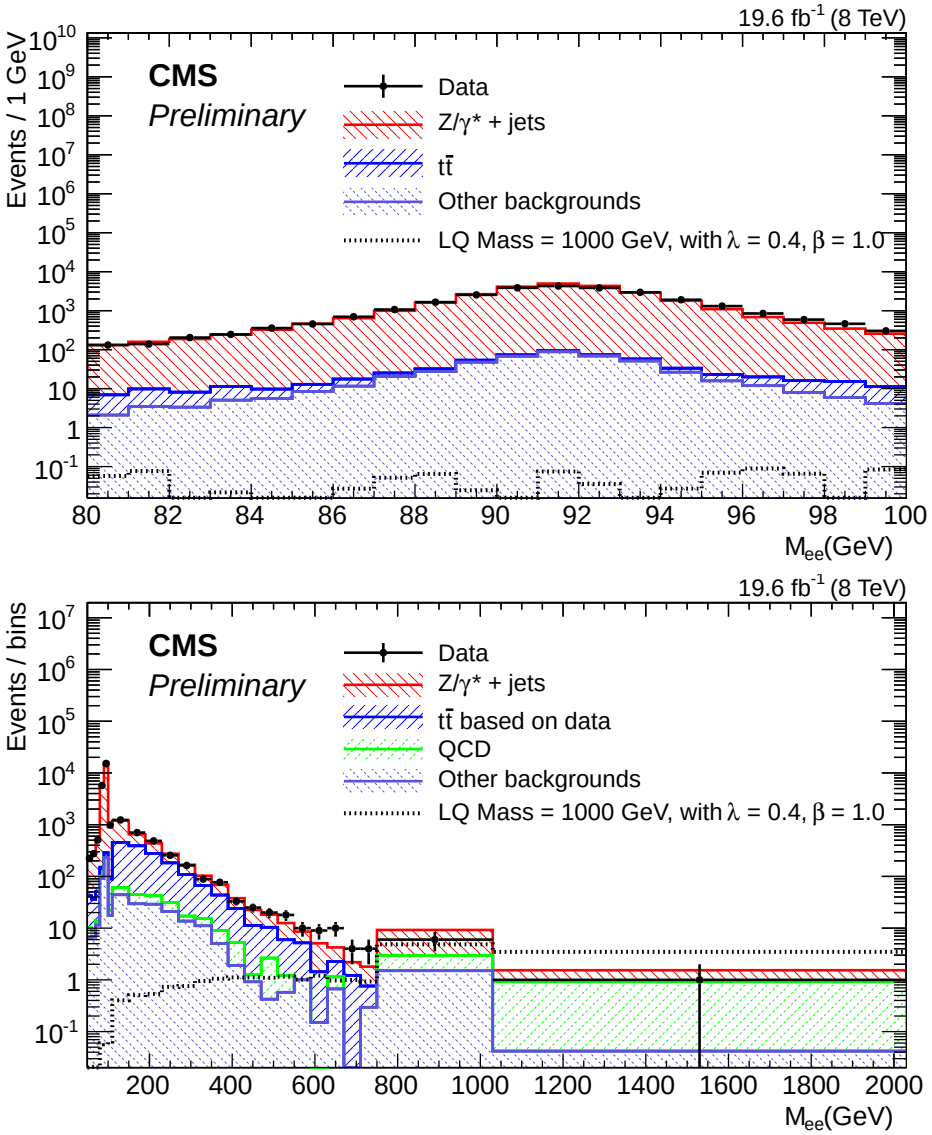


FIGURE 4.10: The dielectron invariant mass distribution of events passing the preselection criteria. The Z+jets contribution from simulation is normalized to data in the Z-enriched mass peak region shown in the top plot, 80 GeV $< M_{ee} <$ 100 GeV. The bottom plot shows the full range of the dielectron invariant mass distribution before applying the $M_{ee} > 110$ GeV requirement.

$\mu\mu j$ channels, respectively. These scale factors differ from unity by 1% to 10% at full selection, depending on the $p_{\mathrm{T}}(\ell\ell)$ bin, and are applied to the $Z/\gamma^* + \text{jets}$ sample in the signal region of $M_{\ell\ell} > 110$ GeV.

4.5.3 Top quark pair production

We estimate the $t\bar{t}$ background with a $t\bar{t}$-enriched eμ sample in data, selected using the single muon trigger. We use a selection that is identical to our signal selection

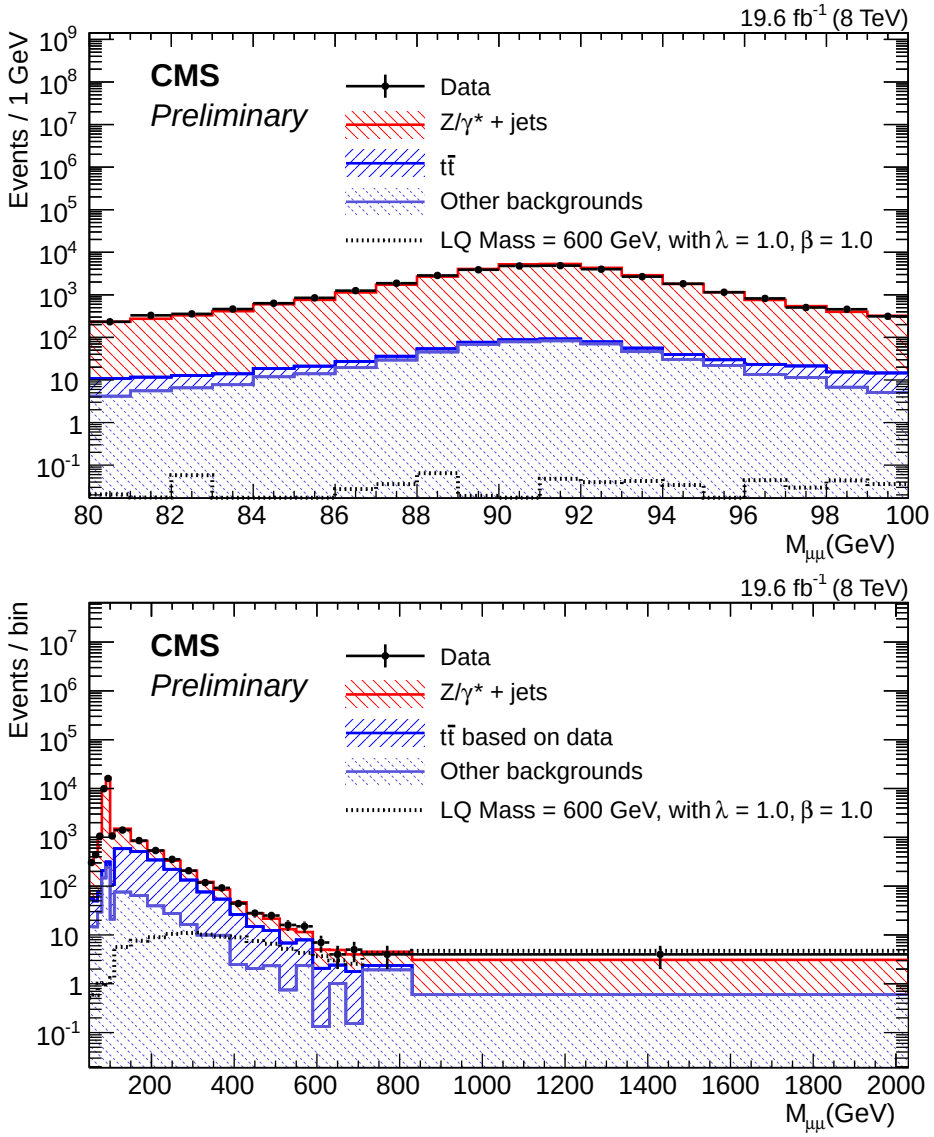


FIGURE 4.11: The dimuon invariant mass distribution of events passing the preselection criteria. The Z+jets contribution from simulation is normalized to data in the Z-enriched mass peak region shown in the top plot, 80 GeV $< M_{\mu\mu j} <$ 100 GeV. The bottom plot shows the full range of the dielectron invariant mass distribution before applying the $M_{\mu\mu} > 110$ GeV requirement.

in terms of kinematics requirements, except that we require at least a single muon and a single electron rather than requiring two same-flavor leptons. The eμ sample is considered to be signal-free, because limits on flavor changing neutral currents imply that LQ processes do not present a different-flavor decay topology [24, 25]. The $t\bar{t}$ background is largely dominant in the eμ sample with respect to the other backgrounds. The sample is normalized to account for the different branching fractions of the eμ and ee or $\mu\mu$ final states (B(eμ) = 2 B(ee or $\mu\mu$)) and the difference in electron and muon identification and isolation efficiencies, collectively

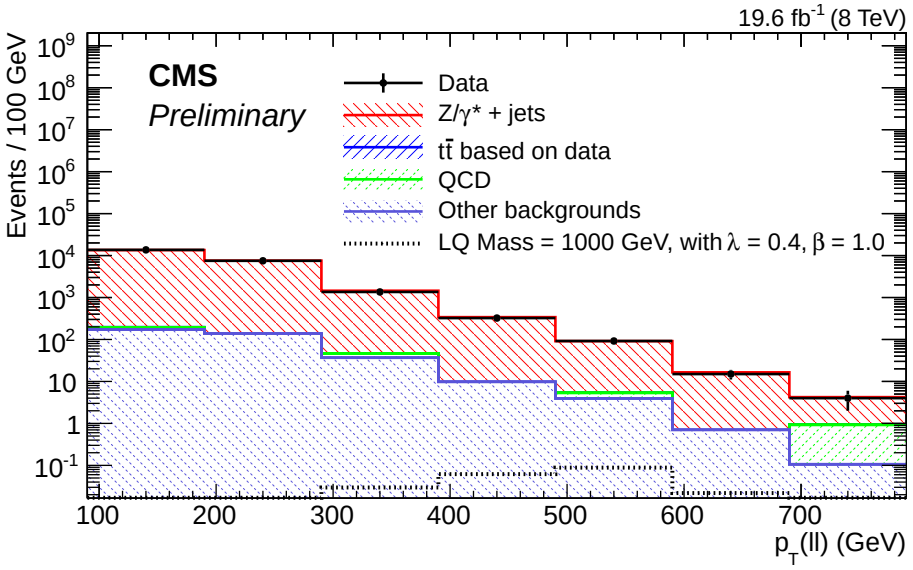


FIGURE 4.12: Inverted dielectron invariant mass selection for rescaling the Drell–Yan plus jets yield. For each bin, the measured value in the $M_{\ell\ell} > 110\,\text{GeV}$ region is scaled for the value measured here. The same procedure is performed for each set of final selection criteria.

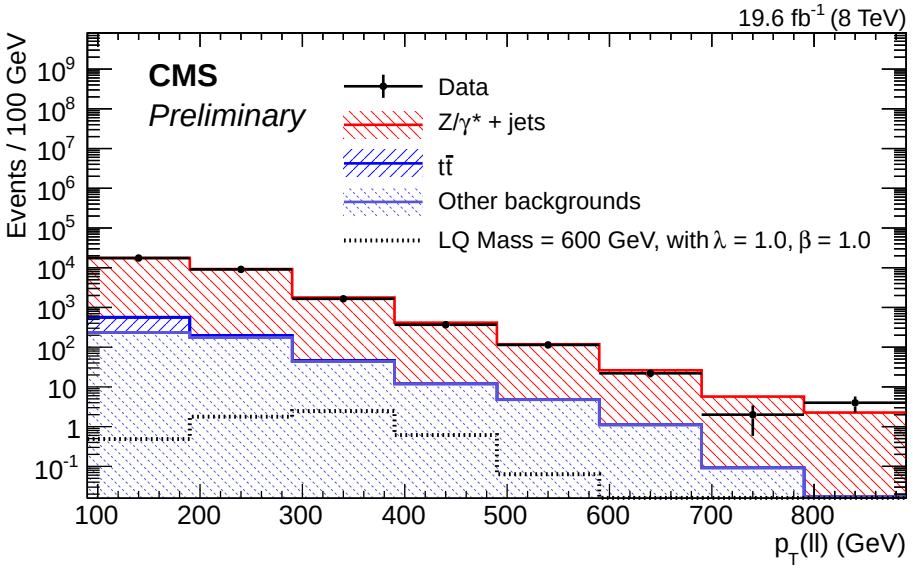


FIGURE 4.13: Inverted dimuon invariant mass selection for rescaling the Drell–Yan plus jets yield. For each bin, the measured value in the $M_{\ell\ell} > 110\,\text{GeV}$ region is scaled for the value measured here. The same procedure is performed for each set of final selection criteria.

taken to be $R_{(\text{ee or }\mu\mu)/\text{e}\mu}$, for the two channels. The sample is also normalized by the ratio of the single-muon trigger efficiency and either the double-electron trigger efficiency or single-muon trigger efficiency in two muon final states ($R_{\text{trig},(\text{ee or }\mu\mu)}$). The overall normalization is taken from simulation and the eμ sample is taken from data. The eμ sample is then used to estimate the contribution from $t\bar{t}$ at

preselection and final selection. The number of estimated $t\bar{t}$ events, $N^{t\bar{t},\text{est}}$, is

$$\begin{aligned} N^{t\bar{t},\text{est}} &= (N^{e\mu,\text{data}} - N^{e\mu,\text{non-}t\bar{t}\text{ sim}}) \times \\ &\quad R_{(\text{ee or }\mu\mu)/e\mu}\, R_{\text{trig},(\text{ee or }\mu\mu)}, \end{aligned} \tag{4.1}$$

with

$$R_{\text{trig,ee}} = \frac{\epsilon_{\text{ee}}}{\epsilon_\mu}, \tag{4.2}$$

$$R_{\text{trig},\mu\mu} = \frac{1-(1-\epsilon_\mu)^2}{\epsilon_\mu} = 2 - \epsilon_\mu, \tag{4.3}$$

where ϵ_μ and ϵ_{ee} are the single-muon trigger and double-electron trigger efficiencies, respectively, and $N^{e\mu,\text{data}}$ and $N^{e\mu,\text{non-}t\bar{t}\text{ sim}}$ are the number of events observed in data and in other backgrounds in the eμ sample, respectively. $R_{\text{trig},\mu\mu}$ is the ratio of the efficiency of a single muon trigger on a dimuon sample over the efficiency on a single muon sample (the numerator is the likelihood of failure on two muons).

As a check that this method works, the measured scale factor is applied to an $e\mu$ selection in simulation, not data, and is compared to the $\ell\ell$ distributions in simulation.

As a form of closure test, plots comparing the rescaled simulated $e\mu$ sample to the simulated eej or $\mu\mu j$ samples are examined. Scales and overall shapes should agree, or the rescaling method does not work. These closure tests are shown in Figure 4.14 for the eej channel and Figure 4.15 for the $\mu\mu j$ channel. Plots comparing the estimate that is used for the $t\bar{t}$ background to the originally simulated sample are shown in Figures 4.16 and 4.17 for the eej and $\mu\mu j$ channels, respectively.

4.5.4 QCD multijets

The contribution from QCD multijet processes is determined by a method that makes use of the fact that neither signal events nor events from other backgrounds produce final states with same-charge leptons at a significant level. We create four selections, with both opposite-sign (OS) and same-sign (SS) charge requirements, as well as isolated and non-isolated requirements. Electrons in isolated events must

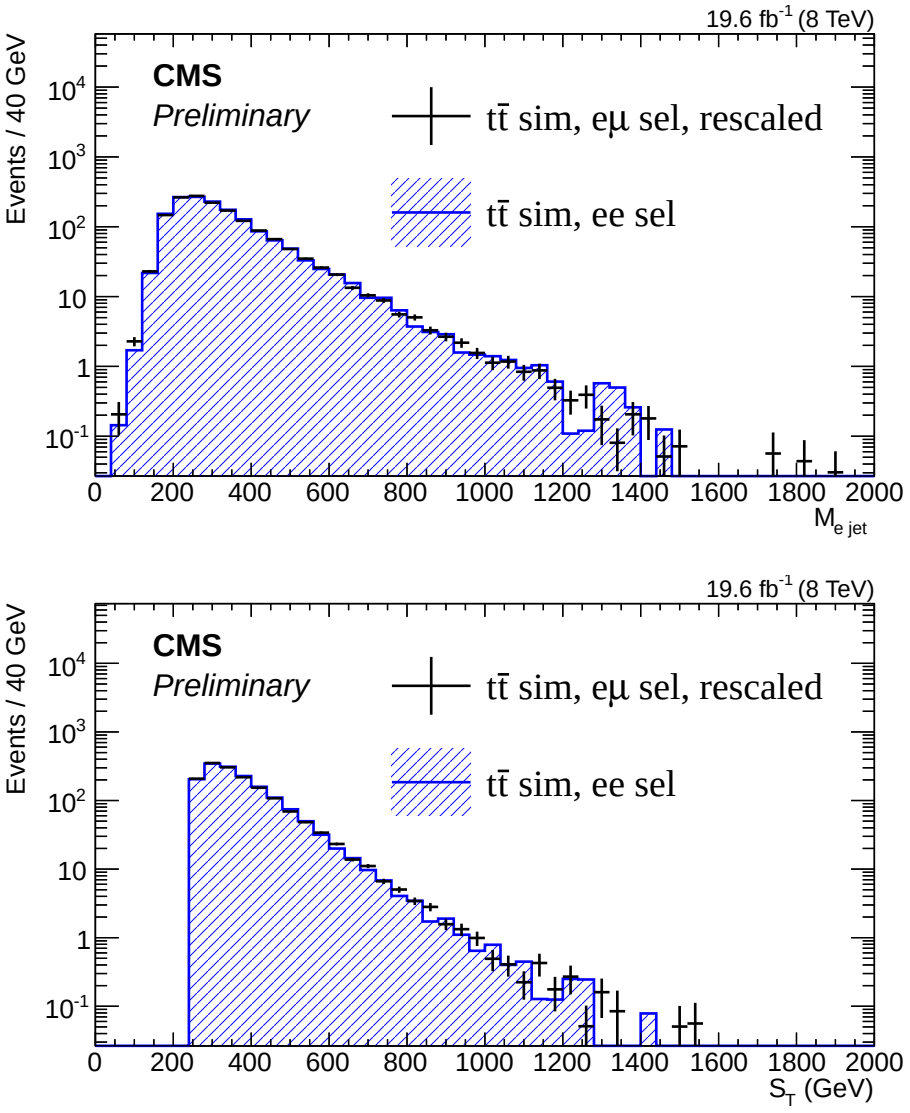


FIGURE 4.14: Lepton jet invariant mass (top) and S_T (bottom) in the closure test performed in the ee$j$ channel. The black crosses represent the simulated $t\bar{t}$ sample with the $e\mu j$ selection after rescaling and the blue histogram represents the simulated $t\bar{t}$ sample with the eej selection. The agreement in the distributions indicates that the scale factor was correctly measured.

pass the isolation criteria optimized for high-energy electrons [69] and muons are required to have a relative tracker isolation less than 0.1, as discussed in Section 3.2.1. Non-isolated events are those with leptons failing these criteria. The four selections are as follows,

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} \text{OS+isolated} & \text{OS+non-isolated} \\ \text{SS+isolated} & \text{SS+non-isolated} \end{pmatrix}. \tag{4.4}$$

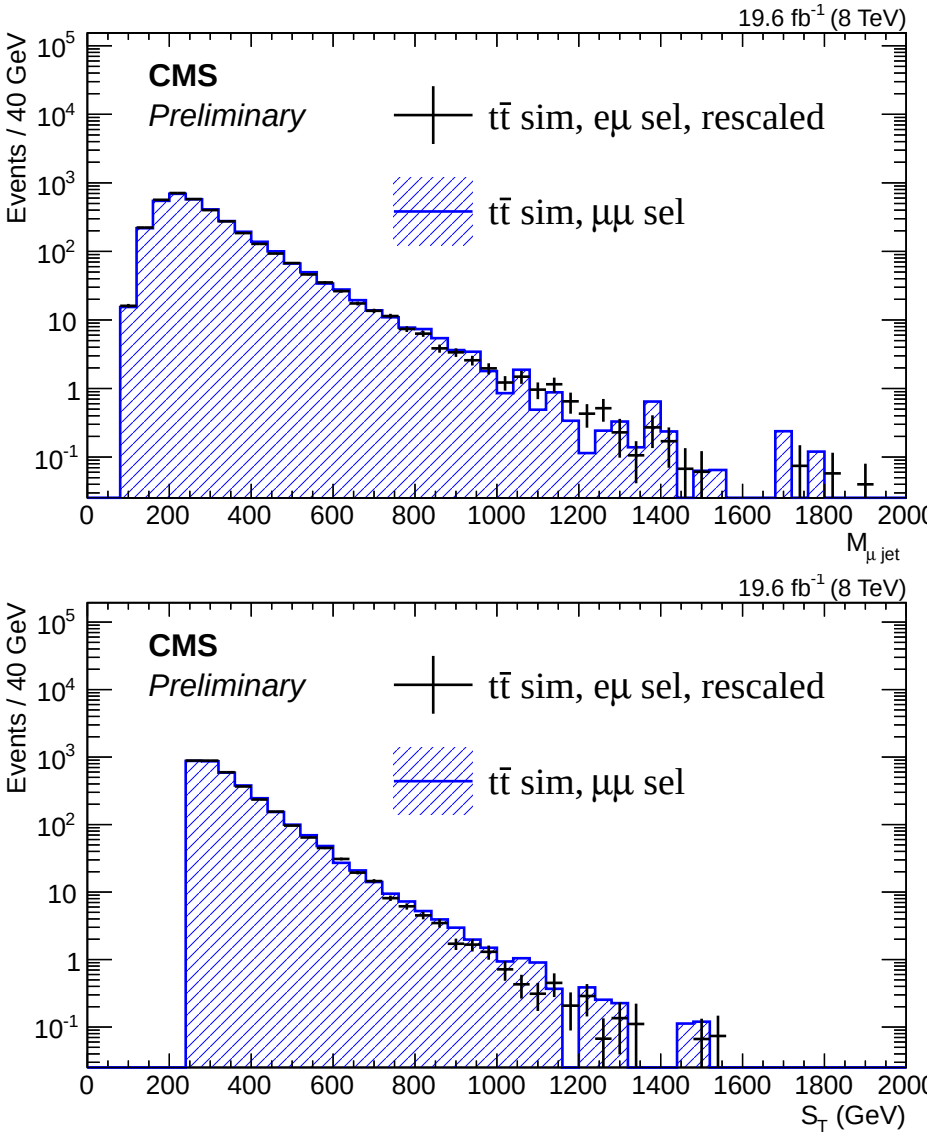


FIGURE 4.15: Lepton jet invariant mass (top) and S_T (bottom) in the closure test performed in the $\mu\mu j$ channel. The black crosses represent the simulated $t\bar{t}$ sample with the $e\mu j$ selection after rescaling and the blue histogram represents the simulated $t\bar{t}$ sample with the $\mu\mu j$ selection.The agreement in the distributions indicates that the scale factor was correctly measured.

The shape of the background is taken from the SS region with isolation requirements, and the normalization is obtained from the ratio between the number of OS events and the number of SS events in the non-isolated selection. Thus, the number of events, $N^{\text{QCD, est}}$, is estimated by

$$N^{\text{QCD, est}} = r_{B/D}\, N_C^{(\text{data} - \text{non-QCD sim})}, \tag{4.5}$$

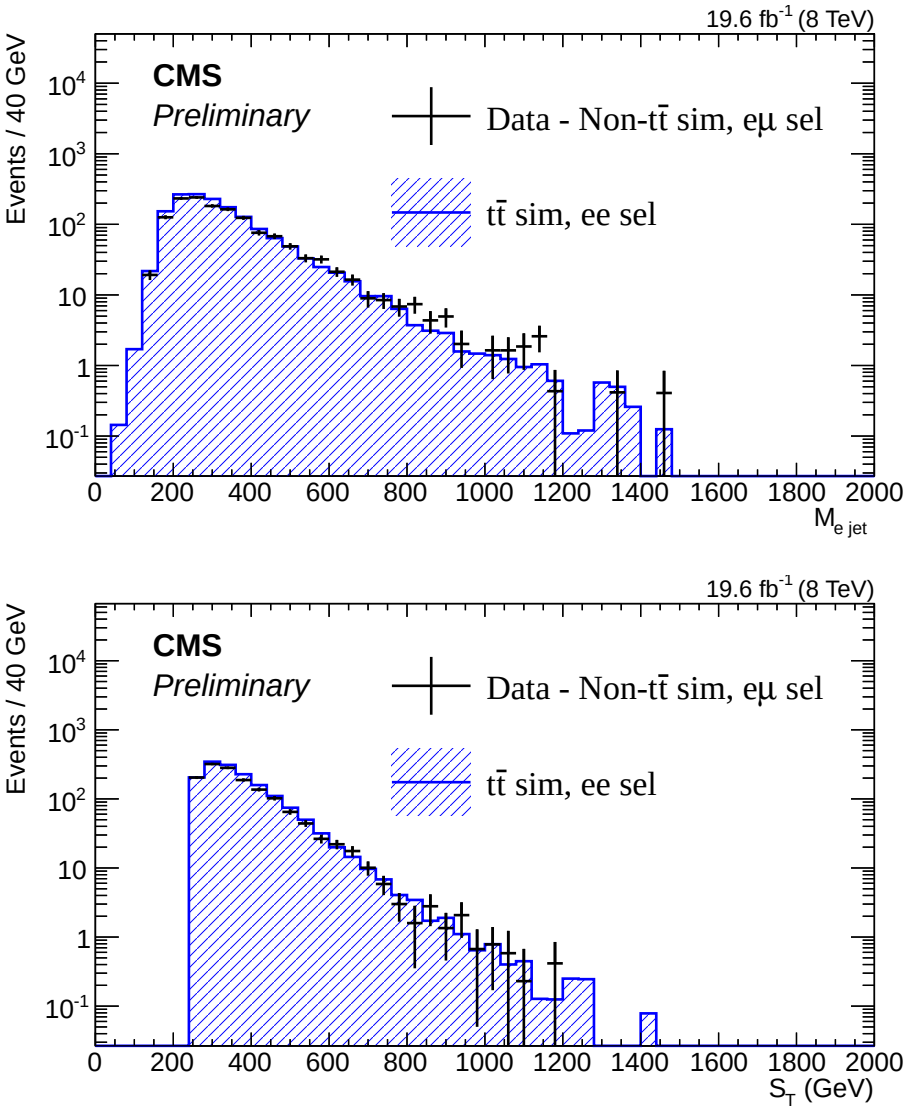


FIGURE 4.16: Comparison of data driven estimate and simulated sample for the $t\bar{t}$ background in the eej channel. The black crosses represent the data-driven estimate that is used for the $t\bar{t}$ background estimate (after rescaling) and the blue histogram represents the simulated estimate.

where $N_C^{(\text{data} - \text{non-QCD sim})}$ is the number of events in region C of Eq. (4.4) and $r_{B/D}$ is the ratio of the number of events (measured in data with simulated non-QCD backgrounds subtracted) in regions B and D.

This strategy assumes that shape of the same sign and opposite sign regions are the same, and that overall shape is not affected by charge. To check this, regions A and B are added and compared to regions C and D added (after normalizing the two distribuations to each other), as shown in Figure 4.18.

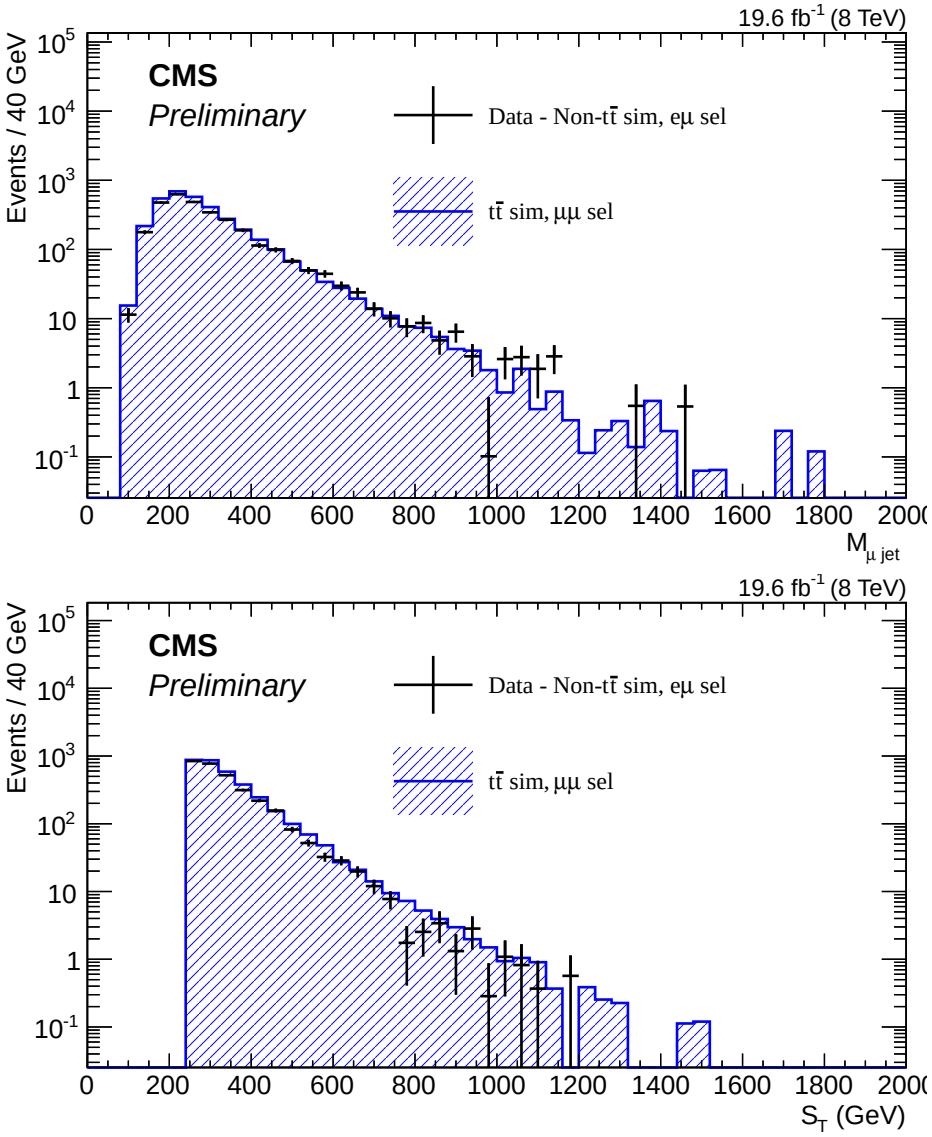


FIGURE 4.17: Comparison of data driven estimate and simulated sample for the $t\bar{t}$ background in the $\mu\mu j$ channel. The black crosses represent the data-driven estimate that is used for the $t\bar{t}$ background estimate (after rescaling) and the blue histogram represents the simulated estimate.

The result is that QCD multijet processes account for 2% (1%) of the total SM background in the ee$j$ ($\mu\mu j$) channel.

4.6 Background checks at preselection

Other than the backgrounds described in the previous Section, the contributions of the remaining backgrounds (diboson+jets, W+jets, single top quark) are small and are determined entirely from simulation.

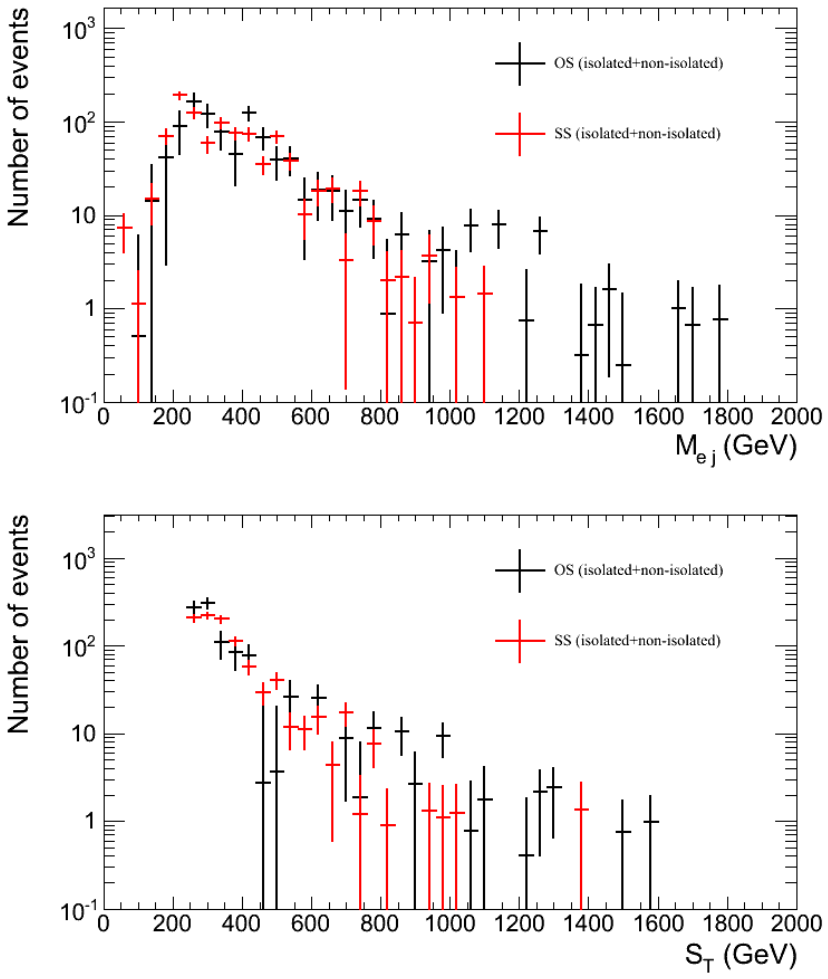


FIGURE 4.18: A comparison of same sign and opposite sign distributions in the QCD multijets background study. The black markers represent the SS selection and the red represent the OS selection.

After combining all the backgrounds, both those estimated via data-driven studies and from simulation, they are compared at preselection to check for good agreement. This preselection check is performed to ensure good background modeling, and the $M_{\ell\ell} > 110\,\text{GeV}$ cut is included.

The preselection level distributions for single object quantities are shown in Figures 4.19 through 4.22, while the distributions for M_{ee}, S_{T}, and $M_{\ell j}$ are shown in Figs. 4.23 and 4.24. In all cases the observed data and estimated backgrounds are also compared with a signal LQ mass of 1000 GeV in the ee$j$ channel, and with a signal LQ of mass 600 GeV, in the $\mu\mu j$ channel. Data and background are found to be in agreement.

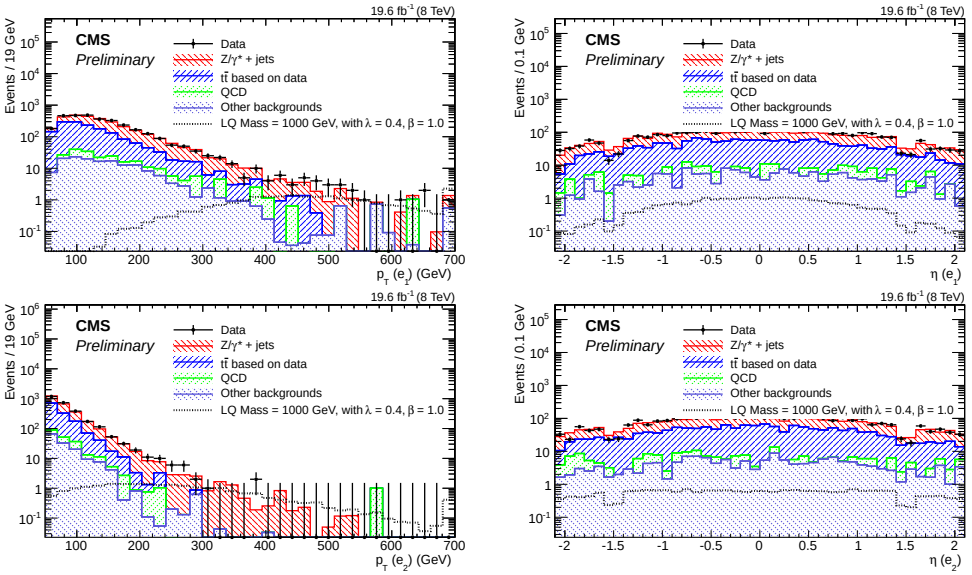


FIGURE 4.19: Distributions of leading electron p_T (top left) and η (top right) and the second leading electron p_T (bottom left) and η (bottom right) at preselection in the ee$j$ channel. The points represent the data and the stacked histograms show the expected background contributions.

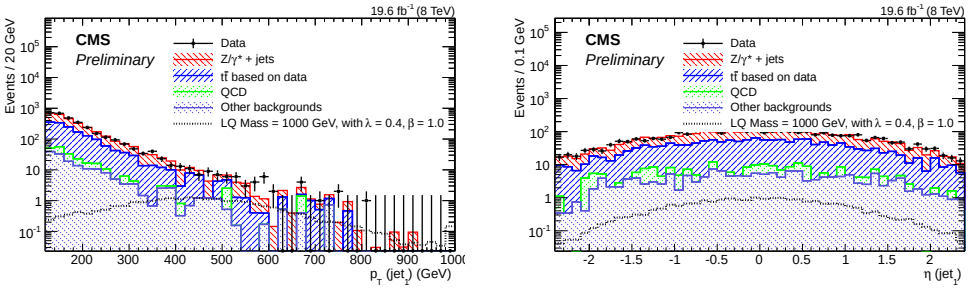


FIGURE 4.20: Distributions of leading jet p_T (left) and η (right) at preselection in the ee$j$ channel. The points represent the data and the stacked histograms show the expected background contributions.

4.7 Final selection

As described in Section 4.7, final selection varies from mass hypothesis to mass hypothesis. A sampling of thresholds is shown in Figures 4.1 and 4.2 for the eej and $\mu\mu j$ channels, respectively. Sampled event yields for the eej channel, its signal, and the $\mu\mu j$ are shown in Figures 4.3 through 4.5, respectively. More complete tables can be found in Appendix A.

The observed data and background predictions are compared after final selection for $\lambda = 0.4$ and a signal LQ mass of 1000 GeV in the ee$j$ channel and a signal LQ mass of 600 GeV in the $\mu\mu j$ channel and are shown in Figs. 4.25 and 4.26.

TABLE 4.1: The ee$j$ channel threshold values for S_{T}, M_{ej}, and $M_{ej,\mathrm{gen}}$ *vs.* LQ mass (for all couplings), and the corresponding observed limits.

M_{LQ} (GeV)	S_{T} threshold (GeV)	M_{ej} threshold (GeV)	$M_{ej,\mathrm{gen}}$ threshold (GeV)
300	250	200	200
600	480	500	400
900	720	800	600
1800	900	1700	1200
3300	900	1900	2200

TABLE 4.2: The $\mu\mu j$ channel threshold values for S_{T}, $M_{\mu j}$, and $M_{\mu j,\mathrm{gen}}$ *vs.* LQ mass, and the corresponding observed limits.

M_{LQ} (GeV)	S_{T} threshold (GeV)	$M_{\mu j}$ threshold (GeV)	$M_{\mu j,\mathrm{gen}}$ threshold (GeV)
300	300	200	225
600	600	500	450
900	900	800	675
1200	1000	800	900
1800	1000	800	1350

TABLE 4.3: Data and background yields after final selection for the ee$j$ channel for first-generation LQs, shown with statistical and systematic uncertainties. "Other backgrounds" refers to diboson+jets, W+ jets, single-top quark, and QCD. The values do not change above 2000 GeV.

M_{LQ}(GeV)	Data	Total background	Z/γ^* + jets	$t\bar{t}$	Other backgrounds
300	3007	$2830 \pm 40 \pm 170$	1362 ± 19	1238 ± 27	230 ± 15
600	370	$329 \pm 12 \pm 24$	192.9 ± 6.3	102.7 ± 7.9	33.3 ± 5.8
900	46	$36.9 \pm 3.4 \pm 6.6$	23.9 ± 1.9	7.6 ± 2.1	5.5 ± 1.9
1200	7	$5.2 \pm 1.6 \pm 1.8$	3.17 ± 0.61	$0.39^{+0.53}_{-0.39}$	1.6 ± 1.3
1800	0	$0.0^{+1.3}_{-0.0} \pm 0.0$	$0.0^{+0.22}_{-0.0}$	$0.0^{+0.41}_{-0.0}$	$0.0^{+1.2}_{-0.0}$

TABLE 4.4: Signal yields in eventsafter final selection in the ee$j$ channel for first-generation LQs shown with statistical and systematic uncertainties, for different values of λ and for $\beta = 1.0$. The first uncertainty listed is statistical and the second is systematic.

M_{LQ} (GeV)	$\lambda = 0.4$	$\lambda = 0.6$	$\lambda = 0.8$	$\lambda = 1.0$
300	$3540 \pm 60 \pm 200$	$7880 \pm 130 \pm 420$	$14390 \pm 240 \pm 820$	$22600 \pm 400 \pm 1200$
600	$289 \pm 3 \pm 18$	$666 \pm 8 \pm 33$	$1188 \pm 14 \pm 76$	$1920 \pm 30 \pm 100$
900	$35.9 \pm 0.4 \pm 1.4$	$82.5 \pm 0.9 \pm 3.3$	$145.7 \pm 1.8 \pm 5.6$	$231 \pm 3 \pm 11$
1800	$0.29 \pm 0.0 \pm 0.03$	$0.76 \pm 0.01 \pm 0.08$	$1.31 \pm 0.03 \pm 0.13$	$2.02 \pm 0.06 \pm 0.24$
3300				$0.004 \pm 0.001 \pm 0.001$

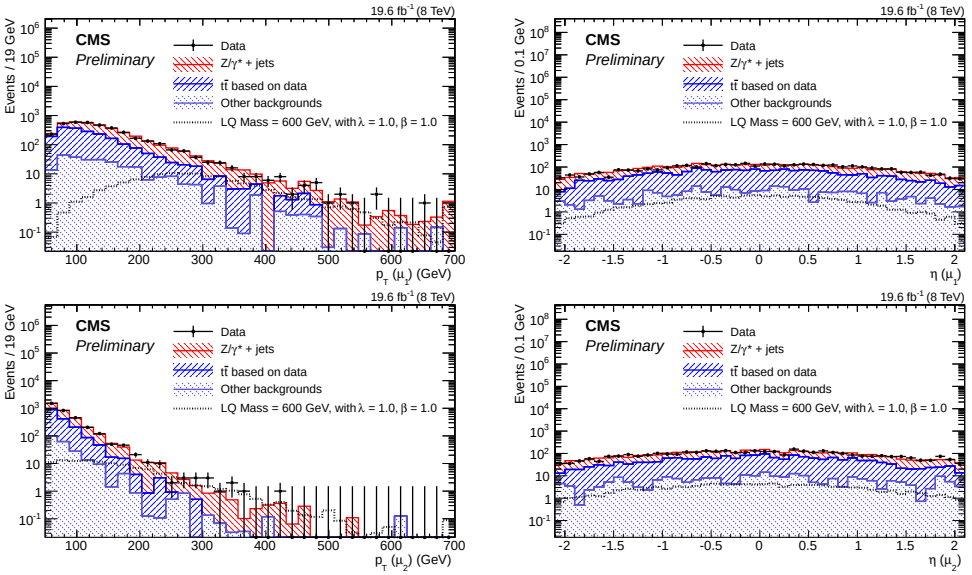


FIGURE 4.21: Distributions of leading muon p_T (top left) and η (top right) and the second leading muon p_T (bottom left) and η (bottom right) at preselection in the $\mu\mu j$ channel. The points represent the data and the stacked histograms show the expected background contributions.

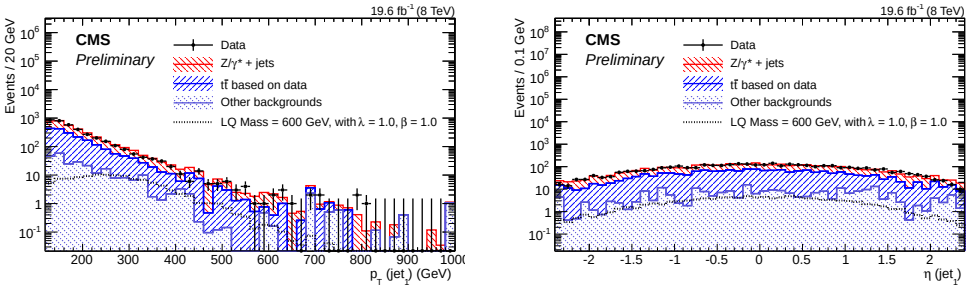


FIGURE 4.22: Distributions of leading jet p_T (left) and η (right) at preselection in the $\mu\mu j$ channel. The points represent the data and the stacked histograms show the expected background contributions.

TABLE 4.5: Data, signal, and background yields in events after final selection in the $\mu\mu j$ channel shown with statistical and total systematic uncertainties, for $\lambda = 1.0$ and $\beta = 1.0$. "Other backgrounds" refers to diboson+jets, W+ jets, single-top quark, and QCD.

M_{LQ}(GeV)	Signal	Data	Total background	$Z/\gamma^* + \mathrm{jets}$	$t\bar{t}$	Other backgrounds
300	$2130 \pm 30 \pm 290$	3036	$3120 \pm 40 \pm 370$	1541 ± 20	1362 ± 32	214 ± 15
600	$77.1 \pm 1.1 \pm 9.5$	238	$246 \pm 10 \pm 32$	155.6 ± 5.6	73.8 ± 7.7	16.4 ± 4.3
900	$4.67 \pm 0.1 \pm 0.84$	27	$25.7 \pm 3.5 \pm 4.6$	14.9 ± 1.5	4.8 ± 2.0	5.9 ± 2.5
1800	$0.0135 \pm 0.0012 \pm 0.0066$	17	$15.5 \pm 3.0 \pm 3.3$	7.6 ± 1.1	2.6 ± 1.5	5.3 ± 2.4

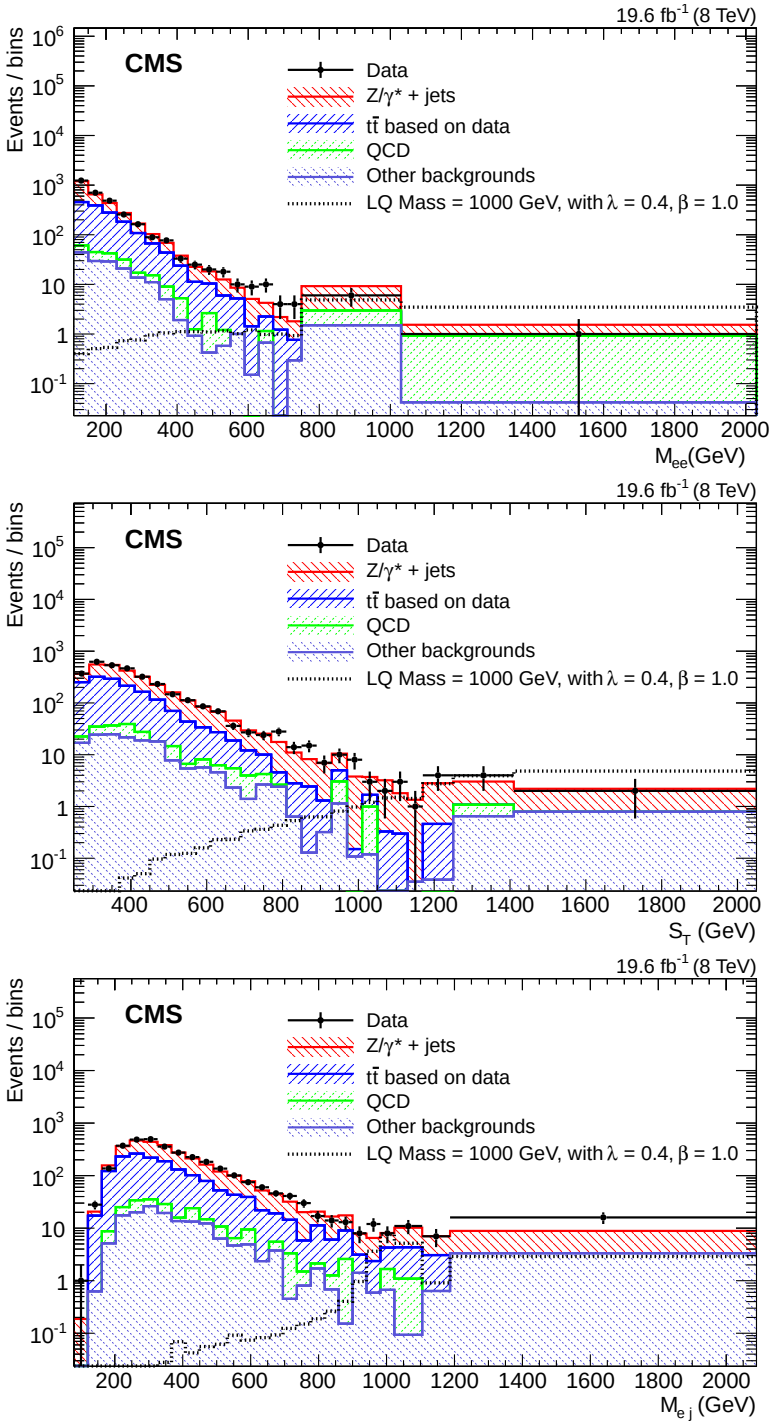


FIGURE 4.23: Distributions of M_{ee} (top), S_{T} (middle), and M_{ej} (bottom) at preselection in the ee$j$ channel. "Other backgrounds" include diboson, W+ jets, and single top quark contributions. The points represent the data and the stacked histograms show the expected background contributions. The open histogram shows the prediction for an LQ signal for $M_{\mathrm{LQ}} = 1000\,\mathrm{GeV}$ and $\lambda = 0.4$. The horizontal error bars on the data points represent the bin width. The last bin includes overflow.

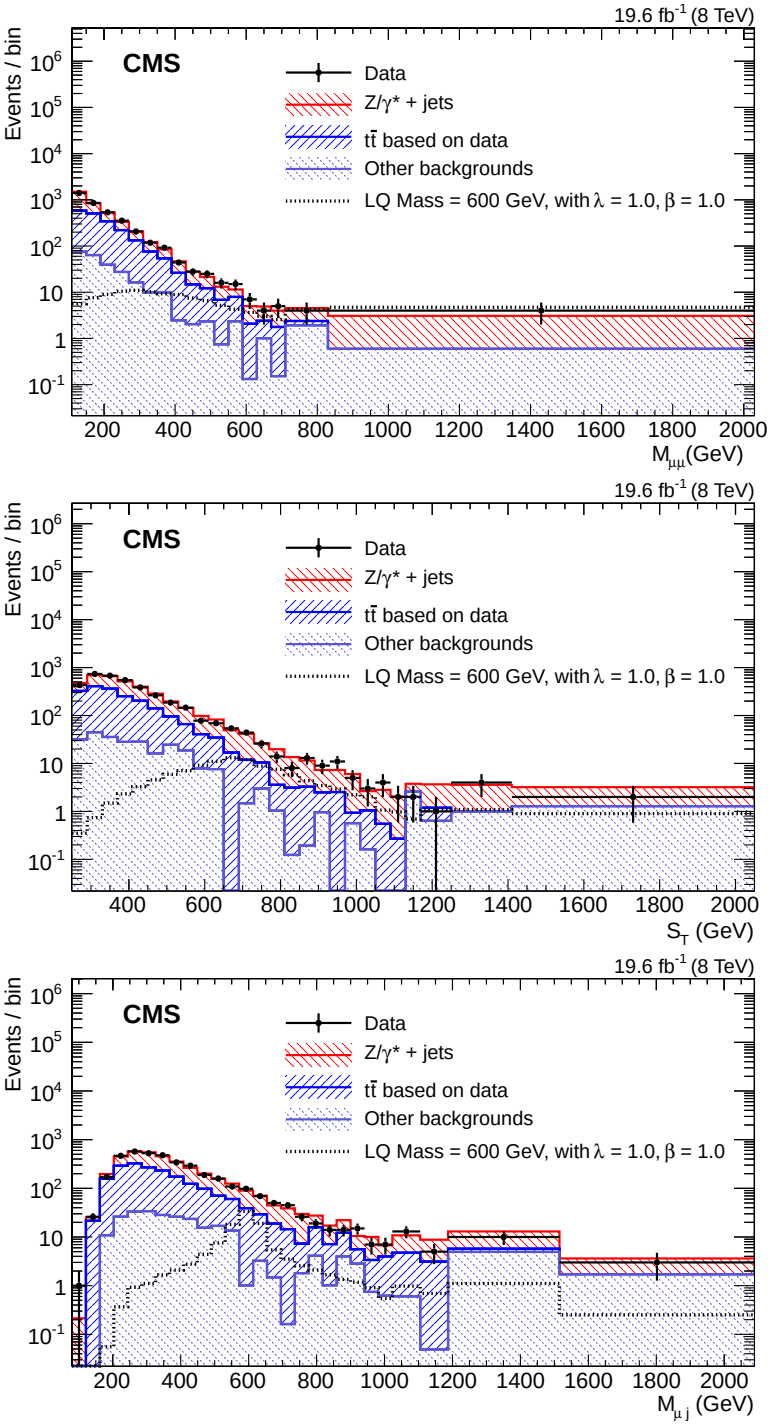


FIGURE 4.24: Distributions of $M_{\mu\mu}$ (top), S_{T} (middle), and $M_{\mu j}$ (bottom) at preselection in the $\mu\mu j$ channel. "Other backgrounds" include diboson, W+ jets, single top quark, and QCD multijet contributions. The points represent the data and the stacked histograms show the expected background contributions. The open histogram shows the prediction for an LQ signal for $M_{\mathrm{LQ}} = 600\,\mathrm{GeV}$ and $\lambda = 1.0$. The horizontal error bars on the data points represent the bin width. The last bin includes overflow.

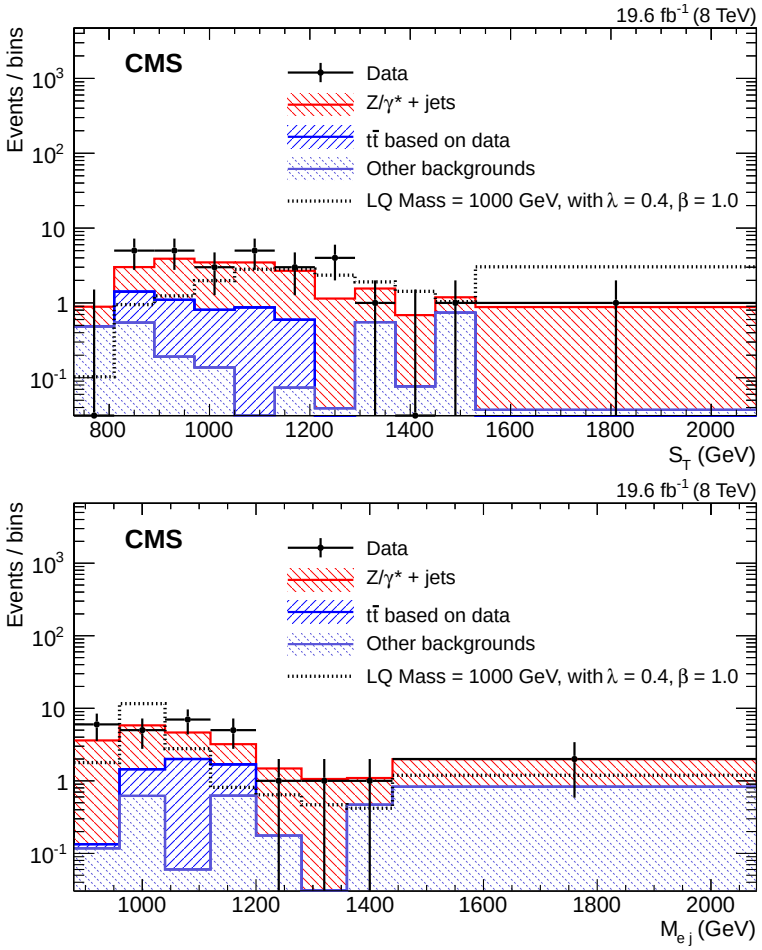


FIGURE 4.25: Distributions of S_{T} and $M_{\mathrm{e}j}$ at final selection, in the ee$j$ channel. The points represent the data and the stacked histograms show the expected background contributions. The open histogram shows the prediction for an LQ signal for $M_{\mathrm{LQ}} = 1000\,\mathrm{GeV}$ and $\lambda = 0.4$. The horizontal error bars on the data points represent the bin width. The last bin includes overflow.

4.8 Systematic uncertainties

Systematic uncertainties are generally assessed by varying quantities in the simulated backgrounds and determining the variation in the event yields at final selection. Uncertainties are included on the kinematic quantities as well as overall event normalization variations due to uncertainties on the pileup modeling, integrated luminosity, and data-driven background rescaling factors. Additionally, uncertainties in the choice of factoriation and renormalization scale and jet-parton matching thresholds that can affect background shapes are evaluated, as well as PDF uncertainties on both the signal and the background cross sections. This chapter outlines how these uncertainties are computed and provides examples of

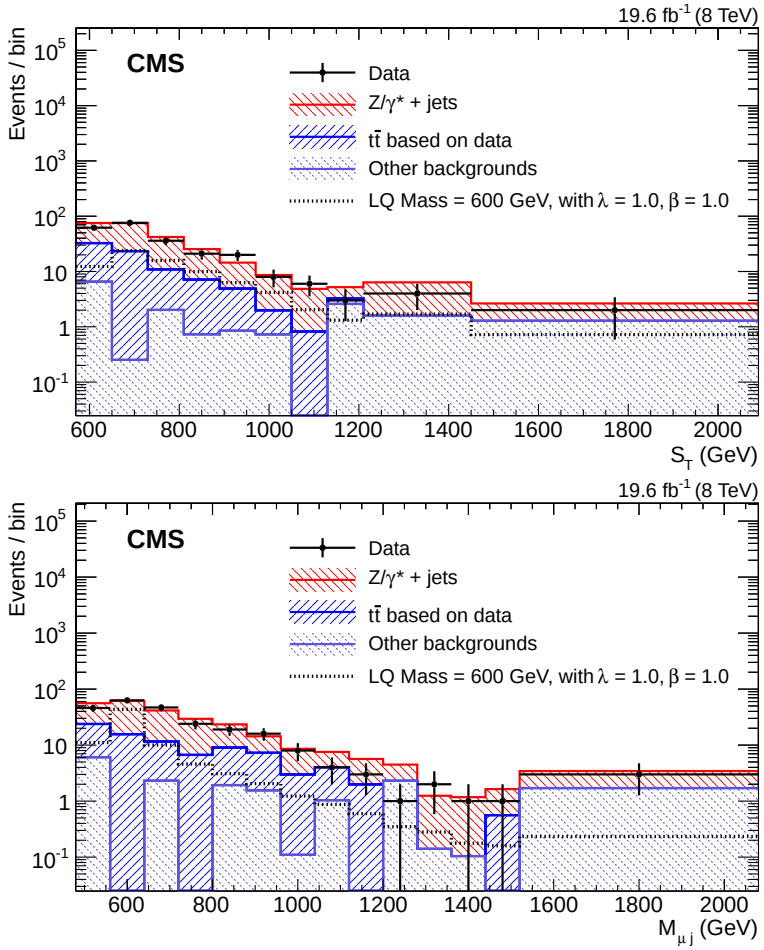


FIGURE 4.26: Distributions of S_{T} and $M_{\mu j}$ at final selection, in the $\mu\mu j$ channel. The points represent the data and the stacked histograms show the expected background contributions. The open histogram shows the prediction for an LQ signal for $M_{\mathrm{LQ}} = 600\,\mathrm{GeV}$ and $\lambda = 1.0$. The horizontal error bars on the data points represent the bin width. The last bin includes overflow.

uncertainties for both channels evaluated at final selection for $M_{\mathrm{LQ}} = 600$ GeV, shown in Table 4.6.

4.8.1 Background normalization

The uncertainties on the normalization of the main backgrounds and the QCD multijet background are discussed in Section 4.5 and are summarized here:

- An uncertainty of 0.6% on R_Z, the scale factor for the Drell–Yan plus jets simulated sample, in both channels

- An uncertainty of 0.5% on $R_{t\bar{t}}$, the overall scale factor for the data-driven study for the $t\bar{t}$ background estimate, in both channels
- An uncertainty of 15% on the overall QCD background measured from data, in both channels

Additionally, in the data-driven study to estimate the $t\bar{t}$ background, there exists a possiblity of contamination from first generation LQs in the control sample for the $\mu\mu j$ channel measurement. This is due to the fact that the single generation diagrams produce one additional jet, thus there could be a non-negligible chance that an initial state radiation (ISR) or final state radiation (FSR) jet could be measured as a muon with a higher p_T than one of the electrons, results in an $e\mu j$ final state. While both the first and second generation diagrams could both contain ISR/FSR jets, the second generation LQ production is so suppressed that this contamination is vanishingly small for the eej channel measurement. This contamination was estimated by measuring the number of first generation signal LQ events in the $e\mu j$ selection, and adding them to the $e\mu j$ meausured event yields to determine the potential variation due to the contamination. Thus, the following uncertainty was evaluated:

- In the mumuj channel only, an uncertainty of 4.0% on the $t\bar{t}$ background normalization to account for potential signal contamination

4.8.2 Background shape

The simulated modeling of the major backgrounds can introduce uncertainties on the shape due to factorization/renormalization scales and jet-parton matching thresholds. Because the $t\bar{t}$ background is estimated from data, these simulation-dependent uncertainties do not affect the measurement. While they do affect the W+jets estimate, that background is very minor and thus the impact is negligible. However, the Z/γ^*+ jets background shape is taken from simulation, and thus these uncertainties must be assessed for that background. This is evaluated using samples simulated with MadGraph with both factorization/renormalization scales and jet-parton matching thresholds varied by a factor of two both up and down. The procedure to measure this uncertainty is as follows:

1. All four simulated samples are fully reconstructed via all the same methods and with the same criteria applied as the main background sample. These four samples include those with both the factorization/renormalization scale and jet-parton matching threshold varied up by a factor of two, both varied down by a factor of two, as well as with the scale varied up and threshold varied down, and vice versa.

2. The normalization of the Drell–Yan plus jets background is performed, with the standard initial event selection applied. Thus only changes in event yields due to the scale and threshold variations are considered.

3. Final event criteria are applied to each of the four samples. Given sufficient population in the samples, the nominal method would be to apply the final event criteria for each LQ mass hypothesis. However, the systematically varied samples produced for the 8 TeV LHC run contained a significantly smaller number of events than the central, unvaried, sample. Thus, in cases where statistical variation in the sample dominated the total variation, the overall value is instead taken at preselection.

4. Once the variation from the central value is measured for each of the four systematically varied samples, the maximal variation is taken as the size of the uncertainty.

Additionally, as described in Section 4.5, the Z/γ^*+ jets background is scaled in a bin-by-bin fashion from the $p_{\mathrm{T}}(\ell\ell)$ distribution with an inverted $M_{\ell\ell} < 110\,\mathrm{GeV}$ selection. The uncertainty for this rescaling is assessed by taking the weighted average of the uncertainties from each $p_{\mathrm{T}}(\ell\ell)$ bin used in the rescaling.

4.8.3 Energy and momentum scales

For the energy and momentum scales of each of the objects in the events (electron, muon, jet), the quantity in question is varied in the simulated signal and background samples and the overall event yield is measured again, with the final uncertainty being taken to be the relative uncertainty of the new yield over the central yield.

Jet energy scale uncertainties are estimated by assigning p_{T}- and η-dependent uncertainties in jet energy corrections as discussed in Ref. [70], and varying the jet p_{T} according to the magnitude of that uncertainty.

For electrons, an energy scale uncertainty of 0.6% in the ECAL barrel and 1.5% in the ECAL endcap is assigned to electrons [71]. For muons, a p_{T}-dependent scale uncertainty of $5\%\,(p_{\mathrm{T}}/1\,\mathrm{TeV})$ is applied [49], in order to precisely quantify the uncertainty for high-momentum muons. In both cases the jet p_{T} is varied according to the corresponding uncertainty, as with jets.

4.8.4 Energy and momentum resolutions

For the energy and momentum resolutions, the same fundamental method is used as for scale uncertainties. The final uncertainty is taken as the relative variation of the systematically varied samples.

The uncertainty in the jet energy resolution is assessed by modifying the p_{T} difference between the generator-level and reconstructed jets by an η-dependent value [70]. In order to compute this uncertainty, the generator-level jet is matched to each reconstructed jet within ΔR, taking the closest of any multiple candidate matchees. The resulting uncertainty varies with jet p_T and η, and is between 5% and 30% for most jets.

For both electrons and muons, the momentum value of each lepton is smeared by random values according to a gaussian distribution centered on the measured value and with a width corresponding to the uncertainty. This resolution uncertainty is taken to be 10% for electrons in both the ECAL barrel and endcap [71], and a p_T dependent value is used for muons, in the range of 1–4% [49].

4.8.5 Parton Distribution Functions

PDF uncertainties in the background and signal acceptance, as well as the background and signal cross sections, are calculated using the PDF4LHC recommendations described in Refs. [72, 73]. This method is applied to the PDF and α_s variations of the CTEQ6.6 [74], MSTW2008 [75], and NNPDF2.0 [76] PDF sets. In this method each variation is applied and the event yields at each final selection

are measured, with the final uncertainty being taken to be the envelope defined by the maximal values, in a very similar technique to that described for the factorization/renormalization scale and jet-parton matching threshold described in Section 4.8.2. The PDF uncertainties are not applied for the background sources measured from data. The overall measured uncertainties are 4–10% for signal and 3–9% for background. The PDF uncertainty is larger in the $\mu\mu j$ channel because of the large uncertainty associated with the s-quark PDF.

4.8.6 Integrated Luminosity and Event and Object Acceptances

The normalization of the simulated background and signal processes are determined according to the total measured integrated luminosity of the 8 TeV data, corresponding to 19.6 fb^{-1}. The uncertainty of that measured luminosity is taken to be 2.6% [77].

Additional uncertainties are applied on the event acceptance of the trigger used to select events, and on the identification of the electrons and muons. A difference observed in the data-to-simulation comparison of the muon trigger efficiency for a single muon yields a 1% uncertainty, and an additional uncertainty on the muon identification and isolation efficiency is evaluated to be 2%.

For electrons, the uncertainty on the trigger efficiency is negligible, and a 0.6% uncertainty in reconstruction, identification, and isolation requirements is applied.

4.8.7 Pileup

The uncertainty from modeling the pileup interactions in simulation is measured by varying the mean of the pileup interactions up and down by 6%. This estimate takes in to account the uncertainty in the luminosity as well as the total inelastic cross section [78]. This uncertainty is assessed in the same way as the object uncertainties, varying the pileup interactions and then measuring the variation in the final event yields.

TABLE 4.6: Systematic uncertainties (in %) and their effects on total signal (S) and background (B) in both channels for $M_{\mathrm{LQ}} = 600$ GeV final selection.

Systematic	eej		$\mu\mu j$	
uncertainty	S(%)	B(%)	S(%)	B(%)
Jet energy scale	0.3	1.0	0.7	1.4
Jet energy resolution	0.1	0.3	0.3	0.4
Electron energy scale	0.2	2.1	-	-
Electron energy resolution	0.1	0.6	-	-
Muon energy scale	-	-	2.4	3.7
Muon energy resolution	-	-	0.2	1.1
Electron reco/ID/iso	1.2	0.1	-	-
Muon reco/ID/iso	-	-	2.0	0.1
Trigger	-	-	1.0	0.1
QCD normalization	-	0.0	-	0.1
$t\bar{t}$ normalization	-	0.2	-	1.1
$Z/\gamma^* +$ jets normalization	-	0.3	-	0.3
$Z/\gamma^* +$ jets shape	-	5.2	-	5.6
$Z/\gamma^* +$ jets $p_{\mathrm{T}}(\ell\ell)$ scale factor	-	2.6	-	3.0
PDF	3.5	3.0	3.0	2.8
Pileup	2.5	0.6	2.8	1.9
Integrated luminosity	2.6	0.3	2.6	0.2
Statistical uncertainty	1.3	3.5	1.4	4.3
Total	5.3	8.1	6.05	8.1

4.8.8 Statistical uncertainty

Finally, a statistical uncertainty associated with the size of the simulated sample is included for both background and signal.

The systematic uncertainties are listed in Table 4.6, together with their effects on signal and background yields, corresponding to the final selection values optimized for $M_{\mathrm{LQ}} = 600$ GeV.

4.9 Limit calculation

The observed data in both eej and $\mu\mu j$ channels are consistent with the standard model background prediction. An upper limit on the leptoquark cross section is set by using the CL_S modified frequentist method [79, 80] with the final event yields. A log-normal probability function is used to model the systematic uncertainties, whereas statistical uncertainties are described with gamma distributions

with widths determined according to the number of events simulated or measured in data control regions.

To isolate the limits for resonant LQ production, we apply the resonant requirements at the generator level on both the lepton+jet mass, $M(\ell, j) > (0.67 \text{ or } 0.75)\, M_{\text{LQ}}$ (for the first- or second- generation LQs, respectively), and on the dilepton mass, $M_{\ell\ell} > 110\,\text{GeV}$. These requirements make the limits extracted from data more conservative and are discussed in Section 4.4. A resonant cross section σ_{res} is computed with respect to those requirements. Limits are then computed with the reduced sample of simulated signal events and compared to σ_{res}.

The 95% confidence level (C.L.) upper limits on $\sigma_{\text{res}}\,\beta$ as a function of leptoquark mass are shown in Fig. 4.27 together with the resonant cross section predictions for the scalar leptoquark single production cross section. The uncertainty band on the theoretical cross section prediction corresponds to uncertainties in the total cross section due to PDF variations with an additional +70% uncertainty, due to the k factor [81]. The observed limits are listed in Tables A.2 and A.3 in the appendix.

By comparing the observed upper limit with the theoretical production cross-section times branching fraction, we exclude single leptoquark production at 95% C.L. for LQ masses below the values given in Table 4.7.

TABLE 4.7: 95% C.L. lower limits on scalar LQ masses ($\beta = 1.0$).

LQ generation, coupling	Excluded mass (GeV)
First gen., $\lambda = 0.4$	860
First gen., $\lambda = 0.6$	1175
First gen., $\lambda = 0.8$	1355
First gen., $\lambda = 1.0$	1755
Second gen., $\lambda = 1.0$	660

Limits on single production of the S_0^R type LQ from the H1 collaboration exclude LQ production up to 500 GeV ($\lambda = 1.0$) and up to 350 GeV ($\lambda = 0.6$) [31]. As discussed in Section 1.2.3, these limits can also be compared with the contemporaneously produced limits of 1010 (1030) GeV in the first (second) generation, assuming $\beta = 1.0$, by the LQ pair production searches performed at CMS [36]. The limits on pair production are β-dependent, falling to 850 (890) GeV in the first (second) generation for $\beta = 0.5$. They are, however, independent of λ. The limits on single production, listed above, are conversely λ-dependent and are for $\beta = 1.0$.

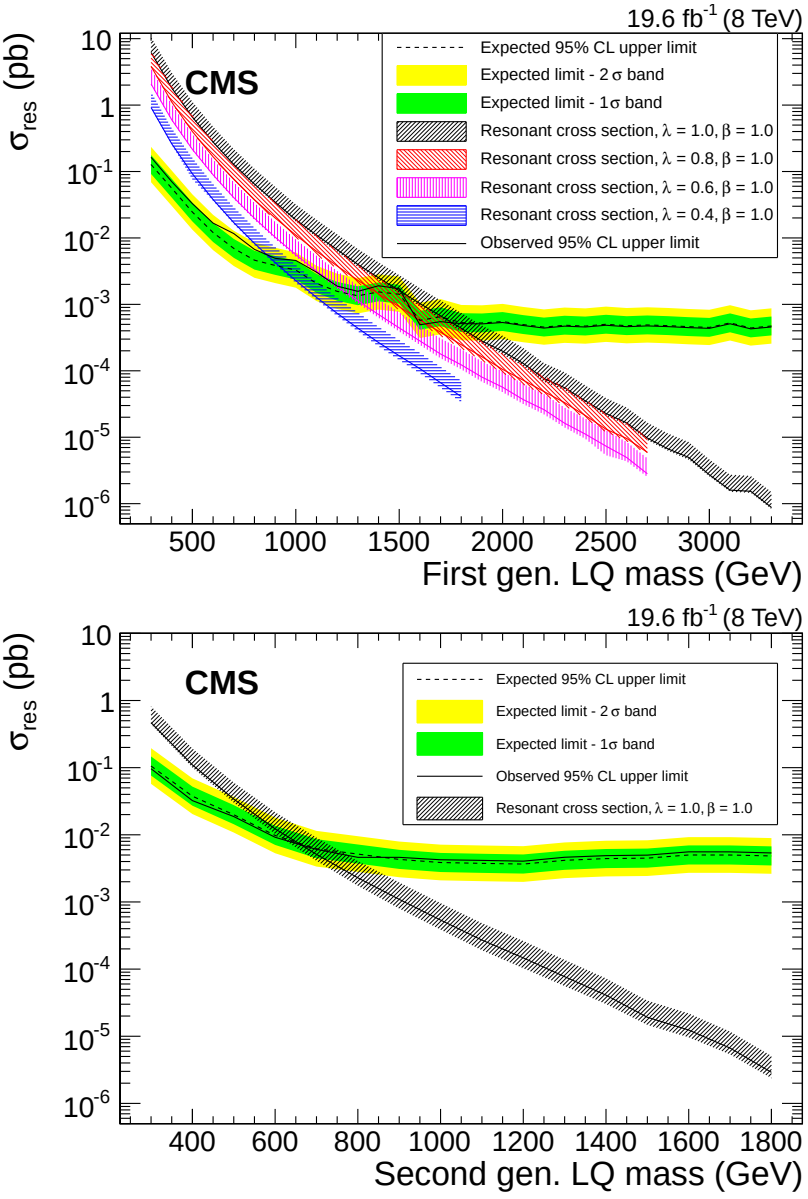


FIGURE 4.27: Expected and observed upper limits at 95% C.L. on first and second generation leptoquark single production resonant cross section as a function of the leptoquark mass. First generation limits are shown on the top plot with a resonant region of $M_{\ell j} > 0.66\, M_{\mathrm{LQ}}$, $M_{\ell\ell} > 110\,\mathrm{GeV}$ and second generation limits are shown on the bottom plot with a resonant region of $M_{\ell j} > 0.75\, M_{\mathrm{LQ}}$, $M_{\ell\ell} > 110\,\mathrm{GeV}$. The uncertainty bands on the observed limit represent the 68% and 95% confidence intervals. The uncertainty band on the theoretical cross section includes uncertainties due to PDF variation and the k factor.

Chapter 5

Conclusion and prospects

This thesis has described a search for the single production of first and second generation leptoquarks in final states with two electrons and a jet or two muons and a jet using 8 TeV pp collisions data corresponding to an integrated luminosity of 19.6 fb^{-1}. For each final state, mass, and coupling under consideration, the number of observed events is in good agreement with the Standard Model prediction. Thus, a 95% confidence level lower limit is set on the leptoquark mass, excluding single production of first generation leptoquarks with coupling $\lambda = 1.0$ with masses below 1755 GeV and second generation leptoquarks with masses below 660 GeV. These are the most stringent limits to date for single production. The first generation limits for couplings greater than 0.6 are stronger than those from pair production and are the most stringent overall limits on leptoquark production in the first generation to date. Additionally, the search for single production of leptoquarks adds a λ-dependence to the leptoquark studies performed at CMS that could not be studied in the β-dependent pair production searches.

This analysis is in the process of being updated for the 2015-2016 CMS datasets, which will consist of 13 TeV pp collisions. This very significant increase in collision energy will greatly extend the sensitivity to leptoquark masses, beyond 2 TeV with $\lambda = 1.0$ in the first generation channel.

Bibliography

[1] C. Burgess and G. Moore. The standard model: A primer. 2007.

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

[3] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964. doi: 10.1103/PhysRevLett.13.508. URL http://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[4] Steven Weinberg. A Model of Leptons. *Phys.Rev.Lett.*, 19:1264–1266, 1967. doi: 10.1103/PhysRevLett.19.1264.

[5] A. Salam and J. C. Ward. Electromagnetic and weak interactions. *Physics Letters*, 13:168–171, November 1964. doi: 10.1016/0031-9163(64)90711-5.

[6] S. L. Glashow, J. Iliopoulos, and L. Maiani. Weak interactions with lepton-hadron symmetry. *Phys. Rev. D*, 2:1285–1292, Oct 1970. doi: 10.1103/PhysRevD.2.1285. URL http://link.aps.org/doi/10.1103/PhysRevD.2.1285.

[7] Steven Weinberg. Physical processes in a convergent theory of the weak and electromagnetic interactions. *Phys. Rev. Lett.*, 27:1688–1691, Dec 1971. doi: 10.1103/PhysRevLett.27.1688. URL http://link.aps.org/doi/10.1103/PhysRevLett.27.1688.

[8] Nicola Cabibbo. Unitary symmetry and leptonic decays. *Phys. Rev. Lett.*, 10:531–533, Jun 1963. doi: 10.1103/PhysRevLett.10.531. URL http://link.aps.org/doi/10.1103/PhysRevLett.10.531.

[9] J. Beringer and et al. Arguin. Review of particle physics. *Phys. Rev. D*, 86:010001, Jul 2012. doi: 10.1103/PhysRevD.86.010001. URL http://link.aps.org/doi/10.1103/PhysRevD.86.010001.

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

[11] Johannes Blumlein, Edward Boos, and Alexander Kryukov. Leptoquark pair production in hadronic interactions. *Z.Phys.*, C76:137–153, 1997. doi: 10.1007/s002880050538.

[12] Howard Georgi and S. L. Glashow. Unity of all elementary-particle forces. *Phys. Rev. Lett.*, 32:438, 1974.

[13] Jogesh C. Pati and Abdus Salam. Unified lepton-hadron symmetry and a gauge theory of the basic interactions. *Phys. Rev. D*, 8:1240, 1973. doi: 10.1103/PhysRevD.8.1240.

[14] Jogesh C. Pati and Abdus Salam. Lepton number as the fourth color. *Phys. Rev. D*, 10:275, 1974. doi: 10.1103/PhysRevD.11.703.2.

[15] H. Murayama and T. Yanagida. A viable SU(5) GUT with light leptoquark bosons. *Mod. Phys. Lett. A*, 7:147, 1992. doi: 10.1142/S0217732392000070.

[16] Harald Fritzsch and Peter Minkowski. Unified interactions of leptons and hadrons. *Annals Phys.*, 93:193, 1975. doi: 10.1016/0003-4916(75)90211-0.

[17] Goran Senjanovic and Aleksandar Sokorac. Light leptoquarks in SO(10). *Z. Phys. C*, 20:255, 1983. doi: 10.1007/BF01574858.

[18] Paul H. Frampton and Bum Hoon Lee. Su (15) grand unification. *Phys. Rev. Lett.*, 65:2209, Oct 1990. doi: 10.1103/PhysRevLett.65.2209. URL http://link.aps.org/doi/10.1103/PhysRevLett.65.2209.

[19] Paul H. Frampton and Thomas W. Kephart. Higgs sector and proton decay in SU(15) grand unification. *Phys. Rev. D*, 42:3892, 1990. doi: 10.1103/PhysRevD.42.3892.

[20] Savas Dimopoulos and Leonard Susskind. Mass without scalars. *Nucl. Phys. B*, 155:237, 1979. doi: 10.1016/0550-3213(79)90364-X.

[21] Savas Dimopoulos. Technicolored signatures. *Nucl. Phys. B*, 168:69, 1980. doi: 10.1016/0550-3213(80)90277-1.

[22] Edward Farhi and Leonard Susskind. Technicolor. *Phys. Rept.*, 74:277, 1981. doi: 10.1016/0370-1573(81)90173-3.

[23] Barbara Schrempp and Fridger Schrempp. Light leptoquarks. *Phys. Lett. B*, 153:101, 1985. doi: 10.1016/0370-2693(85)91450-9.

[24] W. Buchmuller and D. Wyler. Constraints on SU(5) type leptoquarks. *Phys. Lett. B*, 177:377, 1986. doi: 10.1016/0370-2693(86)90771-9.

[25] O. Shanker. πl2, kl3, and k^0-$\bar{K}^0$ constraints on leptoquarks and supersymmetric particles. *Nucl. Phys. B*, 204:375, 1982. doi: 10.1016/0550-3213(82)90196-1.

[26] Alexander Belyaev, Claude Leroy, Rashid Mehdiyev, and Alexander Pukhov. Leptoquark single and pair production at LHC with CalcHEP/CompHEP in the complete model. *JHEP*, 09:005, 2005. doi: 10.1088/1126-6708/2005/09/005.

[27] W. Buchmuller, R. Ruckl, and D. Wyler. Leptoquarks in lepton - quark collisions. *Phys. Lett. B*, 191:442, 1987. doi: 10.1016/0370-2693(87)90637-X.

[28] JoAnne L. Hewett and Thomas G. Rizzo. Much ado about leptoquarks: A Comprehensive analysis. *Phys. Rev.*, D56:5709–5724, 1997. doi: 10.1103/PhysRevD.56.5709.

[29] S. Aid et al. A Search for leptoquarks at HERA. *Phys. Lett.*, B369:173–185, 1996. doi: 10.1016/0370-2693(95)01601-5.

[30] I. Abt et al. A Search for leptoquarks, leptogluons and excited leptons in H1 at HERA. *Nucl. Phys.*, B396:3–26, 1993. doi: 10.1016/0550-3213(93)90255-N.

[31] F.D. Aaron, C. Alexa, V. Andreev, S. Backovic, A. Baghdasaryan, et al. Search for first generation leptoquarks in ep collisions at HERA. *Phys. Lett. B*, 704:388, 2011. doi: 10.1016/j.physletb.2011.09.017.

[32] M. Derrick et al. Search for leptoquarks with the ZEUS detector. *Phys. Lett.*, B306:173–186, 1993. doi: 10.1016/0370-2693(93)91155-G.

[33] Observation of events at very high q2 in ep collisions at hera. *Zeitschrift für Physik C Particles and Fields*, 74(2):191–205, 2014. ISSN 1431-5858. doi: 10.1007/s002880050383. URL http://dx.doi.org/10.1007/s002880050383.

[34] J. Breitweg and et al Derrick. Comparison of zeus data with standard model predictions for e+p to e+x scattering at high x and q2. *Zeitschrift für Physik*

C Particles and Fields, 74(2):207–220, 1997. ISSN 0170-9739. doi: 10.1007/s002880050384. URL http://dx.doi.org/10.1007/s002880050384.

[35] V.M. Abazov et al. Search for single production of scalar leptoquarks in p anti-p collisions decaying into muons and quarks with the D0 detector. *Phys. Lett. B*, 647:74, 2007. doi: 10.1016/j.physletb.2007.01.064.

[36] Vardan Khachatryan et al. Search for Pair Production of First and Second Generation Leptoquarks in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV. 2015.

[37] Lyndon Evans and Philip Bryant. LHC Machine. *JINST*, 3:S08001, 2008. doi: 10.1088/1748-0221/3/08/S08001.

[38] S. Chatrchyan et al. The CMS experiment at the CERN LHC. *JINST*, 3: S08004, 2008. doi: 10.1088/1748-0221/3/08/S08004.

[39] Mike Lamont. Status of the lhc. *Journal of Physics: Conference Series*, 455(1):012001, 2013. URL http://stacks.iop.org/1742-6596/455/i=1/a=012001.

[40] The CMS collaboration. The performance of the cms muon detector in proton-proton collisions at s = 7 tev at the lhc. *Journal of Instrumentation*, 8 (11):P11002, 2013. URL http://stacks.iop.org/1748-0221/8/i=11/a=P11002.

[41] Paolo Azzurri. The cms silicon strip tracker. *Journal of Physics: Conference Series*, 41(1):127, 2006. URL http://stacks.iop.org/1742-6596/41/i=1/a=011.

[42] G. L. Bayatian et al. CMS physics: Technical design report. 2006.

[43] Serguei Chatrchyan et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014. doi: 10.1088/1748-0221/9/10/P10009.

[44] Pierre Billoir and S. Qian. Simultaneous pattern recognition and track fitting by the Kalman filtering method. *Nucl. Instrum. Meth.*, A294:219–228, 1990. doi: 10.1016/0168-9002(90)91835-Y.

[45] R. Mankel. A Concurrent track evolution algorithm for pattern recognition in the HERA-B main tracking system. *Nucl. Instrum. Meth.*, A395:169–184, 1997. doi: 10.1016/S0168-9002(97)00705-5.

[46] R. Fruhwirth. Application of Kalman filtering to track and vertex fitting. *Nucl. Instrum. Meth.*, A262:444–450, 1987. doi: 10.1016/0168-9002(87)90887-4.

[47] A. Strandlie and W. Wittek. Propagation of covariance matrices of track parameters in homogeneous magnetic fields in CMS. 2006.

[48] Wolfgang Waltenberger, Rudolf Frühwirth, and Pascal Vanlaer. Adaptive vertex fitting. *Journal of Physics G: Nuclear and Particle Physics*, 34(12): N343, 2007. URL http://stacks.iop.org/0954-3899/34/i=12/a=N01.

[49] Serguei Chatrchyan et al. Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV. *JINST*, 7:P10002, 2012. doi: 10.1088/1748-0221/7/10/P10002.

[50] CMS Collaboration. Performance of cms muon reconstruction in cosmic-ray events. *Journal of Instrumentation*, 5(03):T03022, 2010. URL http://stacks.iop.org/1748-0221/5/i=03/a=T03022.

[51] J Butler, D Contardo, M Klute, J Mans, and L Silvestris. Technical Proposal for the Phase-II Upgrade of the CMS Detector. Technical Report CERN-LHCC-2015-010. LHCC-P-008, CERN, Geneva. Geneva, Jun 2015. URL https://cds.cern.ch/record/2020886. Upgrade Project Leader Deputies: Lucia Silvestris (INFN-Bari), Jeremy Mans (University of Minnesota) Additional contacts: Lucia.Silvestris@cern.ch, Jeremy.Mans@cern.ch.

[52] S. Baffioni, C. Charlot, F. Ferri, D. Futyan, P. Meridiani, I. Puljak, C. Rovelli, R. Salerno, and Y. Sirois. Electron reconstruction in CMS. *Eur. Phys. J.*, C49:1099–1116, 2007. doi: 10.1140/epjc/s10052-006-0175-5.

[53] S. Chatrchyan and et al. V. Khachatryan. Search for narrow resonances in dilepton mass spectra in pp collisions at. *Physics Letters B*, 714 (2–5):158 – 179, 2012. ISSN 0370-2693. doi: http://dx.doi.org/10.1016/j.physletb.2012.06.051. URL http://www.sciencedirect.com/science/article/pii/S0370269312006909.

[54] 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. URL http://cdsweb.cern.ch/record/1194487.

[55] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k(t) jet clustering algorithm. *JHEP*, 04:063, 2008. doi: 10.1088/1126-6708/2008/04/063.

[56] Gavin P. Salam and Gregory Soyez. A Practical Seedless Infrared-Safe Cone jet algorithm. *JHEP*, 05:086, 2007. doi: 10.1088/1126-6708/2007/05/086.

[57] CMS Collaboration. Determination of the jet energy scale using $Z \to e^+e^-$ + jet p_T balance and a procedure for combining data driven corrections. CMS Physics Analysis Summary CMS-PAS-JME-09-005, 2009. URL http://cdsweb.cern.ch/record/1194492.

[58] J. Alwall and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014. doi: 10.1007/JHEP07(2014)079.

[59] S. Alioli et al. A general framework for implementing nlo calculations in shower monte carlo programs: the powheg box. *JHEP*, 06:043, 2010. doi: 10.1007/JHEP06(2010)043.

[60] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. NLO single-top production matched with shower in POWHEG: s- and t-channel contributions. *JHEP*, 09:111, 2009. doi: 10.1088/1126-6708/2009/09/111.

[61] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11:070, 2007. doi: 10.1088/1126-6708/2007/11/070.

[62] Paolo Nason. A New method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004. doi: 10.1088/1126-6708/2004/11/040.

[63] M. Bondarenko, A. Belyaev, L. Basso, E. Boos, V. Bunichev, et al. High Energy Physics Model Database : Towards decoding of the underlying theory (within Les Houches 2011: Physics at TeV Colliders New Physics Working Group Report). 2012. URL https://hepmdb.soton.ac.uk.

[64] Torbjorn Sjöstrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006. doi: 10.1088/1126-6708/2006/05/026.

[65] J. Pumplin et al. New generation of parton distributions with uncertainties from global qcd analysis. *JHEP*, 07:012, 2002. doi: 10.1088/1126-6708/2002/07/012.

[66] J. Alwall, S. Höche, F. Krauss, N. Lavesson, L. Lönnblad, F. Maltoni, M.L. Mangano, M. Moretti, C.G. Papadopoulos, F. Piccinini, S. Schumann, M. Treccani, J. Winter, and M. Worek. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *The European Physical Journal C*, 53(3):473–500, 2008. ISSN 1434-6044. doi: 10.1140/epjc/s10052-007-0490-5. URL http://dx.doi.org/10.1140/epjc/s10052-007-0490-5.

[67] S. Agostinelli et al. GEANT4 — a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250, 2003. doi: 10.1016/S0168-9002(03)01368-8.

[68] Vardan Khachatryan et al. Charged particle multiplicities in pp interactions at $\sqrt{s} = 0.9$, 2.36, and 7 TeV. *JHEP*, 01:079, 2010. doi: 10.1007/JHEP01(2011)079.

[69] Vardan Khachatryan et al. Search for physics beyond the standard model in dilepton mass spectra in proton-proton collisions at $\sqrt{s} = 8$ TeV. Technical report, 2014.

[70] Serguei Chatrchyan et al. Determination of jet energy calibration and transverse momentum resolution in CMS. *JINST*, 6:P11002, 2011. doi: 10.1088/1748-0221/6/11/P11002.

[71] Vardan Khachatryan et al. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at sqrt(s)=8 TeV. 2015.

[72] Michiel Botje, Jon Butterworth, Amanda Cooper-Sarkar, Albert de Roeck, Joel Feltesse, Stefano Forte, Alexander Glazov, Joey Huston, Ronan McNulty, Torbjörn Sjöstrand, and Robert S. Thorne. The PDF4LHC working group interim recommendations. 2011.

[73] Sergey Alekhin, Simone Alioli, Richard D. Ball, Valerio Bertone, Johannes Blumlein, et al. The PDF4LHC working group interim report. 2011.

[74] Pavel M. Nadolsky, Hung-Liang Lai, Qing-Hong Cao, Joey Huston, Jon Pumplin, Daniel Stump, Wu-Ki Tung, and C.-P. Yuan. Implications of cteq global analysis for collider observables. *Phys. Rev. D*, 78:013004, Jul 2008. doi: 10.1103/PhysRevD.78.013004. URL http://link.aps.org/doi/10.1103/PhysRevD.78.013004.

[75] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur. Phys. J.*, C63:189–285, 2009. doi: 10.1140/epjc/s10052-009-1072-5.

[76] Richard D. Ball, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, José I. Latorre, Juan Rojo, and Maria Ubiali. A first unbiased global {NLO} determination of parton distributions and their uncertainties. *Nuclear Physics B*, 838(1–2):136 – 206, 2010. ISSN 0550-3213. doi: http://dx.doi.org/10.1016/j.nuclphysb.2010.05.008. URL http://www.sciencedirect.com/science/article/pii/S0550321310002853.

[77] CMS Collaboration. Cms luminosity based on pixel cluster counting - summer 2013 update. CMS Physics Analysis Summary CMS-PAS-LUM-13-001, 2013. URL http://cdsweb.cern.ch/record/1598864.

[78] Serguei Chatrchyan et al. Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 7\,\text{TeV}$. *Phys. Lett. B*, 722:5, 2013. doi: 10.1016/j.physletb.2013.03.024.

[79] A L Read. Presentation of search results: the CL_S technique. *J. Phys. G*, 28:2693, 2002. doi: 10.1088/0954-3899/28/10/313.

[80] Thomas Junk. Confidence level computation for combining searches with small statistics. *Nucl. Instrum. Meth. A*, 434:435, 1999. doi: 10.1016/S0168-9002(99)00498-2.

[81] J.B. Hammett and D.A. Ross. NLO Leptoquark Production and Decay: The Narrow-Width Approximation and Beyond. 2015.

Appendix A

Appendix

A.0.1 Signal cross sections

This section contains a table of first- and second-generation LQ cross sections, computed at LO in CalcHEP and scaled for the resonant selection (Table A.1).

A.0.2 Final selection

This section contains the reference tables for the final selection criteria and the corresponding observed limits for the ee$j$ and the $\mu\mu j$ channels.

A.0.3 Event yields

This section contains tables of data, background, and signal yields after the final selection. Event counts vary between the two channels due to differences in the optimized thresholds for S_{T} and $M_{\ell j}$ as well as differences in the electron and muon efficiencies. The first listed uncertainty is statistical, the second is systematic; in cases where only one uncertainty is listed it is statistical.

TABLE A.1: Signal cross sections calculated at LO in CALCHEP. Resonant cross sections scaled by the acceptance of the selections described in Section 4.4 are listed under each corresponding LO cross section.

M_{LQ} (GeV)	First gen., $\lambda = 0.4$ (pb)	First gen., $\lambda = 0.6$ (pb)	First gen., $\lambda = 0.8$ (pb)	First gen., $\lambda = 1.0$ (pb)	Second gen., $\lambda = 1.0$ (pb)
300	1.04	2.39	4.38	7.12	0.579
	0.921	2.08	3.83	6.21	0.468
400	0.291	0.675	1.25	2.06	0.139
	0.261	0.601	1.11	1.81	0.11
500	0.102	0.239	0.451	0.755	0.0446
	0.0924	0.215	0.4	0.658	0.034
600	0.0413	0.0984	0.189	0.322	0.0176
	0.0378	0.0891	0.166	0.278	0.0122
700	0.0186	0.0451	0.088	0.154	0.00807
	0.017	0.0404	0.0763	0.128	0.00511
800	0.00904	0.0223	0.0446	0.0797	0.00418
	0.00829	0.0198	0.0374	0.0647	0.00229
900	0.00467	0.0118	0.0242	0.0443	0.00237
	0.00427	0.0103	0.02	0.0346	0.00109
1000	0.00254	0.00657	0.0139	0.0261	0.00145
	0.00228	0.00559	0.0111	0.0188	0.000537
1200	0.00084	0.00234	0.00526	0.0104	0.00064
	0.000733	0.00186	0.00378	0.00667	0.000147
1400	0.00032	0.00097	0.00233	0.00485	0.00033
	0.000267	0.000705	0.00144	0.00252	4.09e-05
1600	0.00014	0.00045	0.00117	0.00255	0.00019
	0.000108	0.000282	0.000577	0.00103	1.24e-05
1800	6e-05	0.00024	0.00065	0.00147	0.00011
	4.1e-05	0.000123	0.000247	0.000436	2.9e-06
2000		0.00014	0.00039	0.00092	
		5.66e-05	0.000105	0.000197	
2500		5e-05	0.00014	0.00035	
		7.35e-06	1.32e-05	2.28e-05	
3000				0.00016	
				2.72e-06	
3300				0.00011	
				8.79e-07	

TABLE A.2: The ee$j$ channel threshold values for S_{T}, $M_{\mathrm{e}j}$, and $M_{\mathrm{e}j,\mathrm{gen}}$ *vs.* LQ mass (for all couplings), and the corresponding observed limits.

M_{LQ} (GeV)	S_{T} threshold (GeV)	$M_{\mathrm{e}j}$ threshold (GeV)	$M_{\mathrm{e}j,\mathrm{gen}}$ threshold (GeV)	Observed limit on σ_{res} (pb)
300	250	200	200	0.16
400	320	300	266	0.07
500	400	400	333	0.033
600	480	500	400	0.017
700	560	600	466	0.012
800	640	700	533	0.0067
900	720	800	600	0.0049
1000	800	900	666	0.0046
1200	900	1100	800	0.0019
1400	900	1300	933	0.0019
1600	900	1500	1066	0.00049
1800	900	1700	1200	0.00051
2000	900	1900	1333	0.00053
2500	900	1900	1666	0.00048
3000	900	1900	2000	0.00044
3300	900	1900	2200	0.00046

TABLE A.3: The $\mu\mu j$ channel threshold values for S_{T}, $M_{\mu j}$, and $M_{\mu j,\mathrm{gen}}$ *vs.* LQ mass, and the corresponding observed limits.

M_{LQ} (GeV)	S_{T} threshold (GeV)	$M_{\mu j}$ threshold (GeV)	$M_{\mu j,\mathrm{gen}}$ threshold (GeV)	Observed limit on σ_{res} (pb)
300	300	200	225	0.096
400	400	300	300	0.032
500	500	400	375	0.019
600	600	500	450	0.0092
700	700	600	525	0.0061
800	800	700	600	0.0046
900	900	800	675	0.0046
1000	1000	800	750	0.0042
1200	1000	800	900	0.004
1400	1000	800	1050	0.0049
1600	1000	800	1200	0.0056
1800	1000	800	1350	0.0054

TABLE A.4: Data and background yields after final selection for the ee$j$ channel for first-generation LQs, shown with statistical and systematic uncertainties. "Other backgrounds" refers to diboson+jets, W+ jets, single-top quark, and QCD. The values do not change above 2000 GeV.

M_{LQ}(GeV)	Data	Total background	Z/γ^* + jets	$t\bar{t}$	Other backgrounds
300	3007	$2830 \pm 40 \pm 170$	1362 ± 19	1238 ± 27	230 ± 15
400	1766	$1660 \pm 30 \pm 110$	873 ± 15	637 ± 19	151 ± 12
500	807	$736 \pm 18 \pm 49$	409.8 ± 9.6	251 ± 12	75.6 ± 8.6
600	370	$329 \pm 12 \pm 24$	192.9 ± 6.3	102.7 ± 7.9	33.3 ± 5.8
700	186	$149 \pm 8 \pm 12$	91.6 ± 4.1	40.9 ± 4.9	16.7 ± 4.2
800	91	$73.7 \pm 5.6 \pm 7.0$	46.3 ± 2.8	21.1 ± 3.5	6.3 ± 3.3
900	46	$36.9 \pm 3.4 \pm 6.6$	23.9 ± 1.9	7.6 ± 2.1	5.5 ± 1.9
1000	28	$18.3 \pm 2.5 \pm 4.8$	11.7 ± 1.3	3.7 ± 1.5	2.9 ± 1.5
1200	7	$5.2 \pm 1.6 \pm 1.8$	3.17 ± 0.61	$0.39^{+0.53}_{-0.39}$	1.6 ± 1.3
1400	4	$1.8 \pm 1.3 \pm 1.5$	1.0 ± 0.31	$0.0^{+0.41}_{-0.0}$	$0.8^{+1.2}_{-0.8}$
1600	0	$0.2^{+1.2+0.4}_{-0.2-0.2}$	0.17 ± 0.12	$0.0^{+0.41}_{-0.0}$	$0.1^{+1.2}_{-0.1}$
1800	0	$0.0^{+1.3}_{-0.0} \pm 0.0$	$0.0^{+0.22}_{-0.0}$	$0.0^{+0.41}_{-0.0}$	$0.0^{+1.2}_{-0.0}$
2000	0	$0.0^{+1.3}_{-0.0} \pm 0.0$	$0.0^{+0.22}_{-0.0}$	$0.0^{+0.41}_{-0.0}$	$0.0^{+1.2}_{-0.0}$

TABLE A.5: Signal yields after final selection in the ee$j$ channel for first-generation LQs shown with statistical and systematic uncertainties, for different values of λ and for $\beta = 1.0$.

M_{LQ} (GeV)	$\lambda = 0.4$	$\lambda = 0.6$	$\lambda = 0.8$	$\lambda = 1.0$
300	$3540 \pm 60 \pm 200$	$7880 \pm 130 \pm 420$	$14390 \pm 240 \pm 820$	$22600 \pm 400 \pm 1200$
400	$1577 \pm 22 \pm 85$	$3600 \pm 50 \pm 190$	$6330 \pm 80 \pm 340$	$9990 \pm 150 \pm 530$
500	$670 \pm 10 \pm 160$	$1504 \pm 18 \pm 85$	$2670 \pm 30 \pm 140$	$4270 \pm 60 \pm 210$
600	$289 \pm 3 \pm 18$	$666 \pm 8 \pm 33$	$1188 \pm 14 \pm 76$	$1920 \pm 30 \pm 100$
700	$138.1 \pm 1.6 \pm 6.2$	$320 \pm 4 \pm 15$	$559 \pm 7 \pm 27$	$885 \pm 12 \pm 41$
800	$67.8 \pm 0.8 \pm 3.3$	$158.2 \pm 1.8 \pm 6.5$	$275 \pm 3 \pm 12$	$446 \pm 6 \pm 19$
900	$35.9 \pm 0.4 \pm 1.4$	$82.5 \pm 0.9 \pm 3.3$	$145.7 \pm 1.8 \pm 5.6$	$231 \pm 3 \pm 11$
1000	$19.26 \pm 0.22 \pm 0.88$	$43.6 \pm 0.5 \pm 1.8$	$77.9 \pm 1.0 \pm 3.1$	$118.3 \pm 1.8 \pm 4.6$
1200	$6.14 \pm 0.07 \pm 0.25$	$13.8 \pm 0.2 \pm 1.2$	$25.44 \pm 0.35 \pm 0.98$	$39.7 \pm 0.6 \pm 1.9$
1400	$2.2 \pm 0.0 \pm 0.2$	$5.07 \pm 0.07 \pm 0.28$	$9.13 \pm 0.14 \pm 0.58$	$13.78 \pm 0.26 \pm 0.88$
1600	$0.8 \pm 0.0 \pm 0.1$	$1.89 \pm 0.03 \pm 0.15$	$3.3 \pm 0.06 \pm 0.26$	$5.24 \pm 0.12 \pm 0.46$
1800	$0.29 \pm 0.0 \pm 0.03$	$0.76 \pm 0.01 \pm 0.08$	$1.31 \pm 0.03 \pm 0.13$	$2.02 \pm 0.06 \pm 0.24$
2000		$0.31 \pm 0.01 \pm 0.04$	$0.497 \pm 0.014 \pm 0.071$	$0.81 \pm 0.03 \pm 0.12$
2500		$0.039 \pm 0.001 \pm 0.032$	$0.064 \pm 0.003 \pm 0.016$	$0.102 \pm 0.006 \pm 0.023$
3000				$0.0134 \pm 0.0015 \pm 0.0029$
3300				$0.004 \pm 0.001 \pm 0.001$

TABLE A.6: Data, signal, and background yields after final selection in the $\mu\mu j$ channel shown with statistical and total systematic uncertainties, for $\lambda = 1.0$ and $\beta = 1.0$. "Other backgrounds" refers to diboson+jets, W+ jets, single-top quark, and QCD.

M_{LQ}(GeV)	Signal	Data	Total background	Z/γ^* + jets	$t\bar{t}$	Other backgrounds
300	$2130 \pm 30 \pm 290$	3036	$3120 \pm 40 \pm 370$	1541 ± 20	1362 ± 32	214 ± 15
400	$721 \pm 9 \pm 91$	1371	$1440 \pm 30 \pm 170$	774 ± 14	548 ± 21	118 ± 11
500	$228 \pm 3 \pm 27$	558	$577 \pm 17 \pm 75$	340.7 ± 8.6	182 ± 12	54.3 ± 8.1
600	$77.1 \pm 1.1 \pm 9.5$	238	$246 \pm 10 \pm 32$	155.6 ± 5.6	73.8 ± 7.7	16.4 ± 4.3
700	$28.0 \pm 0.5 \pm 3.7$	100	$102 \pm 6 \pm 14$	70.1 ± 3.5	22.3 ± 4.3	9.5 ± 2.7
800	$10.7 \pm 0.2 \pm 1.6$	48	$52.3 \pm 4.7 \pm 7.6$	32.3 ± 2.3	12.3 ± 3.2	7.7 ± 2.6
900	$4.67 \pm 0.1 \pm 0.84$	27	$25.7 \pm 3.5 \pm 4.6$	14.9 ± 1.5	4.8 ± 2.0	5.9 ± 2.5
1000	$2.1 \pm 0.05 \pm 0.46$	17	$15.5 \pm 3.0 \pm 3.3$	7.6 ± 1.1	2.6 ± 1.5	5.3 ± 2.4
1200	$0.7 \pm 0.02 \pm 0.22$	17	$15.5 \pm 3.0 \pm 3.3$	7.6 ± 1.1	2.6 ± 1.5	5.3 ± 2.4
1400	$0.195 \pm 0.008 \pm 0.088$	17	$15.5 \pm 3.0 \pm 3.3$	7.6 ± 1.1	2.6 ± 1.5	5.3 ± 2.4
1600	$0.06 \pm 0.003 \pm 0.032$	17	$15.5 \pm 3.0 \pm 3.3$	7.6 ± 1.1	2.6 ± 1.5	5.3 ± 2.4
1800	$0.0135 \pm 0.0012 \pm 0.0066$	17	$15.5 \pm 3.0 \pm 3.3$	7.6 ± 1.1	2.6 ± 1.5	5.3 ± 2.4