SEARCH FOR VECTOR-LIKE QUARK PAIR PRODUCTION WITH MULTILEPTON FINAL STATES USING 19.5 FB^{-1} OF PP COLLISIONS AT $\sqrt{\text{S}}$=8 TeV

By

CHRISTIAN CONTRERAS-CAMPANA

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Physics and Astronomy

written under the direction of

Prof. Stephen Schnetzer

and approved by

____

____

____

____

____

New Brunswick, New Jersey

January, 2015

ABSTRACT OF THE DISSERTATION

Search for Vector-Like Quark Pair Production with Multilepton Final States using 19.5 fb^{-1} of pp Collisions at $\sqrt{\mathrm{s}}$=8 TeV

By CHRISTIAN CONTRERAS-CAMPANA

Dissertation Director:

Prof. Stephen Schnetzer

A search for pair-production of vector-like partners of the b quark, b', using 19.5 fb^{-1} of integrated luminosity in pp collisions at $\sqrt{s} = 8$ TeV collected by the CMS experiment at the LHC, is carried out in events with at least three leptons. Observed multilepton events are categorized into exclusive channels according to the amount of expected Standard Model background in order to increase the search sensitivity. The observations are consistent with Standard Model predictions. The search is interpreted in the context of a vector-like b' quark for different b' masses and for varying branching fractions to the bZ, tW, and bH final states. b' quarks with masses less than values in the range of $520-785$ GeV, depending on the values of the branching fraction, are excluded at the 95% confidence level.

Acknowledgments

First and foremost, I would like to thank my advisor Prof. Steve Schnetzer for giving me the opportunity to work with the high-energy particle physics group at Rutgers. During my PhD years, I had the chance to work along side great colleagues, all of who played a role in helping me develop as a physicist. I would like to express my appreciation and give many thanks to Matt and Sanjay not only for their help throughout my time with the multilepton group, but also for their influence in making me a better and more efficient programmer. Also, many thanks go out to Richard for laying the foundations of our group's work. Furthermore, I also want to thank Dean Hidas for all our great conversations over coffee during my stay at CERN. Moreover, during my PhD program, I greatly benefited from the support and collaboration of the conveners of the CMS B2G group. Their support, feedback, and direction lead to the results discussed in this thesis. Additionally, I thank Prof. Sunil Somalwar, from whom I learned a great deal, for serving as an unofficial co-advisor and helping me navigate through the CMS analysis publication process. I would also like to extend recognition to my thesis committee members, namely, Yuri Gershtein, Sergei Lukyanov, Girsh Blumberg, and Michael Sokoloff.

Over the course of my academic career, both undergraduate and graduate, countless people have guided me, and while it is impossible to thank everyone individually, there are a few I would be remiss not mention. From Mount San Antonio College, I would like to acknowledge my physics Professor Martin Mason for his encouragement and influencing my interest in particle physics, without which I never would have pursued this field. I have to thank friends Preema, Rishi, and Anthony, both in the US and at CERN, who have helped make my graduate school experience memorable.

I am truly grateful to my parents and family members who have given me all the freedom and opportunities that one could ever hope to have. Para mis queridos padres, Victor y Jeannette, muchos agradecimientos por haberme dado las oportunidades de seguir mis sueños. No hubiera sido posible sin ustedes. I would like to also thank my dear friend Ignacio and the Rios Caviedes family for their heartfelt influence on me while living in Chile.

I reserve special thanks to my wife Claudia and brother Emmanuel who have accompanied, encouraged, and inspired me in ways that made me the person I am today. The countless hours they spent with me working in the office or at home was a big part of what made the work described in this thesis possible. Emmanuel it brings me great joy to have been able to work with you all these years, and that we have been able to take this journey together, for which I will forever be grateful.

I would like to express my sincere appreciation and thanks to Megan for not only being a great sister-in-law but also being a wonderful friend. Last but not least, I would like to give many thanks to Sue, Richard, Karen, Scott, and Ann, as well as the rest of the Parker family, for all the amazing Friday night dinners and holiday celebrations. All of you made New Jersey feel like a home.

Dedication

To my beautiful wife Claudia and my twin brother Emmanuel

for taking this journey of searching for more than just particles with me...

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this dissertation, we present a search for a vector-like b' quark, through their pair production and subsequent decay to a bottom quark and either a Z, W, or H boson. We conduct a multichannel counting experiment for events with three or more leptons and at least one b-jet in the final state.

The thesis is arranged as follows. Chapter 1 is a brief outline. Chapter 2 is an introduction to the Standard Model (SM) and an overview of a possible extension of the SM involving vector-like quarks. Chapter 3 describes the experimental setup, focusing on the Large Hadron Collider accelerator and the CMS detector. Chapter 4 discusses collision data and Monte Carlo simulation. Chapter 5 describes event and object reconstruction. Chapter 6 discusses object selection and efficiency. Chapter 7 details the analysis strategy, and background estimations. Chapter 8 describes the sources of systematic uncertainties associated with the analysis. Chapter 9 provides an overview of the statistical and limit setting procedures. Chapter 10 summarizes the results and interpretation.

Chapter 2

Theoretical Overview

In this chapter we give a theoretical overview of fundamental particles and their interactions in the context of the Standard Model of particle physics. The second part of this chapter focuses on a possible extension of the Standard Model, introducing a scenario with new massive vector-like quarks.

2.1 The Standard Model

The Standard Model (SM) of particle physics is a theoretical description of elementary particles and their interactions. This chapter provides a short overview of the SM, following closely the description in Ref. [1].

The theoretical and experimental discoveries made since the 1930's have resulted in a incredible insight into the structure of matter. Everything found in the known universe is built from fundamental constituents, with perhaps the exception of dark energy and dark matter, referred to as elementary particles, which are governed by four fundamental interactions. The SM provides our best understanding of how fundamental particles interact with one another. The SM successfully explains all experimental results and precisely predicts a variety of physics phenomena. The excellent agreement with experimental results, establishes the SM as a well tested theory. To briefly review, particle interactions are described by locally gauge invariant quantum field theories (QFT), such as the electromagnetic, weak, and strong interactions. A fourth fundamental force, gravity, is not described by the SM, due to the weakness of gravity at scales where quantum effects become apparent. Particle interactions are mediated by the exchange of spin-1 gauge fields, which includes the $W^{\pm}$ and Z^0 bosons for the weak interaction, photons for the electromagnetic interaction, and gluons for the strong interaction. The local gauge symmetry group of the SM is described by,

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

The symmetry group $\mathrm{SU(3)_C}$, where C denotes color charge, represents the the strong interaction describing the fundamental interaction among quarks and gluons. The symmetry group $\mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$, represents the electroweak force, which describes the interaction of leptons, quarks, and gauge bosons. The subscript L means that the SU(2) part of the weak theory only acts on left-handed fields of spinors and $Y = 2Q - T_3$ refers to the weak hypercharge of the unified electroweak theory, where Q corresponds to the electromagnetic charge and $T_3 = \dfrac{\sigma_3}{2}$ refers to the third component of weak isospin.

In the SM, all elementary particles are either *fermions* with half-integer spin, that correspond to the building blocks of matter, or *bosons* with integer spin, referred to as the force mediators. Fermions obey Fermi-Dirac statistics while bosons obey Bose-Einstein statistics. Fermions are further categorized into leptons and quarks. Leptons are arranged into three *generations*. Each generations in the SM forms an isospin doublet of left-handed states $\binom{\nu_\ell}{\ell}_L$ (ℓ is an e, μ, or τ), with non-zero weak isospin and a singlet of right-handed state ℓ_R, with zero weak isospin. One of the particles in the lepton pair carries an integer charge (-1), while the other doublet partner (ν_e, ν_μ, ν_τ) is electrically neutral. The leptons besides having charge, also are assigned a lepton quantum number, for particles (L = +1) and for antiparticles (L = −1), and (L = 0) for non-lepton particles. The difference between the two categories of fermions is that quarks participate in both the electroweak and strong interaction, while leptons do not experience the strong interactions. Quarks have fractional charges of $-\frac{1}{3}$ or $+\frac{2}{3}$ and are assigned a baryon quantum number B, for quarks ($\mathrm{B} = +\frac{1}{3}$) and for anti-quarks ($\mathrm{B} = -\frac{1}{3}$), while (B = 0) for non-quark particles. The quark flavors known to exist are up, down, charm, strange, bottom, and top, and are represented by isospin doublets $\binom{U}{D}_L$. Quarks carry an additional form of charge, referred to as the color charge of the strong force, with three values r, b, and g. A more detailed discussion is found in Section 2.1.4. The particles of each generations differ only by their masses, where particles in the first generation are lighter than the corresponding particles in the next generation. The Higgs boson with spin-0 is responsible for electroweak symmetry breaking, is discussed later in Section 2.1.3. Table 2.1 summarizes the properties of these elementary fermions and bosons shown in Figure 2.1.

The SM can be derived from a Lagrangian density $\mathcal{L}_{SM}$, following the Lagrangian formalism and using Noether's theorem [2], which relates continuous symmetries of a system to physically conserved quantities. The SM Lagrangian density depends on 19 free parameters, whose values must be determined by experimental measurements. These free parameters of the SM include the masses of the fermions, generated by the Yukawa couplings of the fermion fields to the Higgs field, the gauge couplings, various mixing angles, a CP-violating phase, and the Higgs self-interaction strength and quadratic coupling.

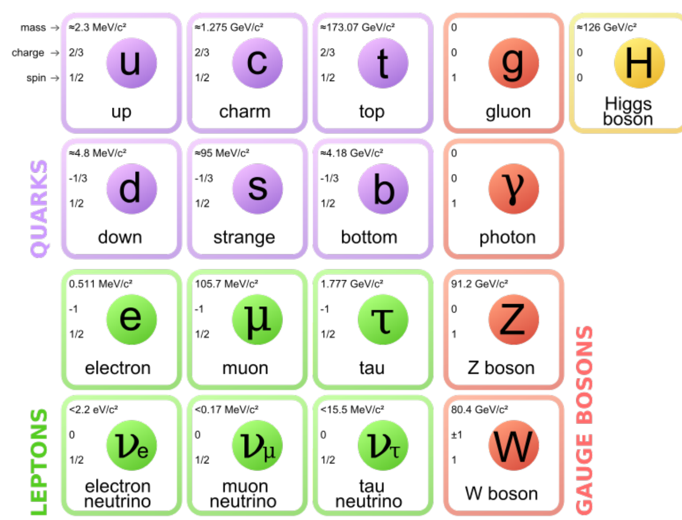


Figure 2.1: Standard Model of elementary particles physics [3].

Particle	Spin S	Charge Q	Lepton number L	Baryon number B	Mass (MeV)
Leptons					
electron (e)	$\frac{1}{2}$	-1	1	0	0.511
electron neutrino (ν_e)	$\frac{1}{2}$	0	-1	0	$< 2.2 \times 10^{-6}$
muon (μ)	$\frac{1}{2}$	-1	1	0	105.7
muon neutrion (ν_μ)	$\frac{1}{2}$	0	-1	0	< 0.17
tau (τ)	$\frac{1}{2}$	-1	1	0	1.77×10^3
tau neutrion (ν_τ)	$\frac{1}{2}$	0	-1	0	< 15.5
Quarks					
up (u)	$\frac{1}{2}$	$\frac{2}{3}$	0	1	2.4
down (d)	$\frac{1}{2}$	$-\frac{1}{3}$	0	1	4.8
charm (c)	$\frac{1}{2}$	$\frac{2}{3}$	0	1	1.27×10^3
strange (s)	$\frac{1}{2}$	$-\frac{1}{3}$	0	1	104
top (t)	$\frac{1}{2}$	$\frac{2}{3}$	0	1	171.2×10^3
botom (b)	$\frac{1}{2}$	$-\frac{1}{3}$	0	1	4.2×10^3
Gauge boson					
photon (γ)	1	0	0	0	0
W boson ($W^{\pm}$)	1	$\pm$ 1	0	0	809.4×10^3
Z boson (Z^0)	1	0	0	0	91.2×10^3
gluon (g)	1	0	0	0	0
Higgs boson (H)	0	0	0	0	126×10^3

Table 2.1: The elementary particles of the SM with their spin, charge, lepton number, baryon number, and mass. The electric charge is given in units of the elementary charge e and spin in units of $\hbar$.

2.1.1 Quantum Electrodynamics

Quantum electrodynamics (QED) is an Abelian gauge theory, that describes the electromagnetic interactions between electrically charged fermions mediated by a massless spin-1 gauge boson, the *photon*. Requiring the QED Lagrangian density be invariant under the local gauge invariance symmetry group $U(1)_Q$, introduces the photon field, in which the gauge transformation is a function of the space-time point. The free QED Lagrangian density for a Dirac fermion field ψ with mass m is described by,

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi, \tag{2.2}$$

where $\bar{\psi}$ is the conjugate of the two dimensional Dirac Spinor ψ and γ^{μ} are the Dirac matrices. The algebra of these matrices is defined by the anti-commutation relation,

$$\{\gamma^{\mu}, \gamma^{\nu}\} = 2g^{\mu\nu}, \tag{2.3}$$

where $g^{\mu\nu}$ is the metric tensor.

Elements of the symmetry group $U(1)_Q$ are phase rotations of the fermion field ψ. The field theory must have local gauge invariance in order for the theory to be renormalizeable. Applying a local U(1) transformation, the terms in the QED Lagrangian transform according to,

$$\psi \rightarrow e^{iQ\theta(x)}, \tag{2.4a}$$

$$\partial^{\mu}\psi \rightarrow e^{iQ\theta(x)}\partial_{\mu}\psi + iQ\partial_{\mu}(\theta(x))e^{iQ\theta(x)}\psi, \tag{2.4b}$$

Since, the local transformation phase θ depends on local space-time coordinates, $\theta = \theta(x)$, where x is the space-time position, the Lagrangian gains an extra term due to the derivative of $\theta(x)$.

In order to maintain a gauge invariant Lagrangian, this extra term must be canceled, which is accomplished by introducing a covariant derivative, D_μ, that contains an additional spin-1 field A_μ,

$$D_\mu = \partial_\mu + ieA_\mu, \tag{2.5}$$

by replacing the normal derivative with the covariant derivative, the terms in the Lagrangian remain invariant under local gauge transformation, according to,

$$D_\mu\psi \to (\partial_\mu + iQA_\mu^{'})e^{iQ\theta(x)}\psi = e^{iQ\theta(x)}(\partial_\mu + iQ(A_\mu^{'} + \partial_\mu\theta(x)))\psi, \tag{2.6}$$

where the gauge field A_μ transforms as,

$$A_\mu \to A_\mu^{'} = A_\mu - \partial_\mu\theta(x), \tag{2.7}$$

and the Lagrangian is now invariant under local U(1) transformations. The principle of local gauge invariance introduces an interaction term between the fermion field ψ and the vector field A_μ into the Lagrangian, that describes the interaction between photons and fermions. The QED Lagrangian density, after including a kinematic term for the new vector field A_μ, is expressed as,

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}, \tag{2.8}$$

where $F_{\mu\nu}$ is the electromagnetic field strength tensor, given by,

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{2.9}$$

A mass term for the gauge field is forbidden, since a term of the form $\frac{1}{2}m_A^2 A_\mu A^\mu$ would not persevere local gauge invariance. Therefore, gauge bosons mediating the electromagnetic interaction, are massless photons. The electromagnetic coupling constant is given by α, also known as the fine-structure constant of QED, and is given by,

$$\alpha = \frac{e^2}{4\pi} = \frac{1}{137}. \tag{2.10}$$

Although, the charge is modified by the vacuum polarization loops, for all practical purposes the variation in α is extremely small, increasing from 1/137 very slowly with decreasing distance.

2.1.2 Electroweak Interactions

The Glashow-Weinberg-Salam model [4, 5] that unifies the weak interaction and quantum electrodynamics, is called the electroweak interaction described by the gauge group SU(2)$_L$⊗U(1)$_Y$. It is a chiral theory because the right-handed and left-handed fermion states transform differently under the SU(2)$_L$ group. The electroweak Lagrangian density is given by,

$$\mathcal{L}_{EWK} = -\frac{1}{4}\sum_{3}^{A} F^{A}_{\mu\nu}F^{A\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + i\bar{\psi}_L \not{D}_L \psi_L + i\bar{\psi}_R \not{D}_R \psi_R, \tag{2.11}$$

where the sum is taken over all fermion fields ψ. The right and left handed components are chiral projections of the fields ψ and are defined by $\psi_{R,L} = \frac{1}{2}(1 \pm \gamma^5)\psi$, with $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. The spinors ψ_L are doublets under SU(2)$_L$, whereas the right-handed spinors ψ_R are singlets. The first two terms are kinematic terms of the gauge bosons of SU(2)$_L$⊗U(1)$_Y$ with,

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{2.12a}$$

$$F^{A}_{\mu\nu} = \partial_\mu W^{A}_{\nu} - \partial_\nu W^{A}_{\mu} - g\epsilon_{ABC} W^{B}_{\mu} W^{C}_{\nu}, \tag{2.12b}$$

where B_μ is the gauge field of the group U(1)$_Y$, W$^{A}_{\mu}$ for $A = 1, 2, 3$ are the gauge fields of the of the group $\mathrm{SU}(2)_L$, and ϵ_{ABC} are the structure constants of $\mathrm{SU}(2)_L$.

The covariant derivates are given by,

$$D_\mu = \partial_\mu + ig_1\frac{1}{2}W_\mu \cdot \mathbf{T} + ig_2\frac{1}{2}YB_\mu, \tag{2.13}$$

where the weak-isospin operator $\mathbf{T}$ and weak hypercharge Y, represent the generators of the groups $\mathrm{SU}(2)_L$ and $\mathrm{U}(1)_Y$, respectively. The covariant derivative D_μ results in three gauge bosons, with a gauge coupling g_1 for B_μ and a weak gauge coupling g_2 for the W_μ^a boson.

The physical gauge fields are a linear combination of the W_μ and B_μ gauge fields, given by,

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), \tag{2.14a}$$

$$Z_\mu^0 = \frac{1}{\sqrt{g_1^2 + g_2^2}}(g_2 W_\mu^3 - g_1 B_\mu) = -B_\mu \mathrm{sin}\theta_W + W_\mu^3 \mathrm{cos}\theta_W, \tag{2.14b}$$

$$A_\mu = \frac{1}{\sqrt{g_1^2 + g_2^2}}(g_1 W_\mu^3 + g_2 B_\mu) = B_\mu \mathrm{cos}\theta_W + W_\mu^3 \mathrm{sin}\theta_W, \tag{2.14c}$$

and $\tan\theta_W = g_2/g_1$, where θ_W is known as the Weinberg angle or weak angle. A list of the chiral fermion doublets can be found in Table 2.2

Particles	1^{st} **Gen.**	2^{nd} **Gen.**	3^{rd} **Gen.**	**T**	$\mathbf{T_3}$	**Y**
Quarks	$\binom{u}{d}_L$	$\binom{c}{s}_L$	$\binom{t}{b}_L$	1/2	$\binom{+1/2}{-1/2}_L$	+1/3
Leptons	$\binom{\nu_e}{e}_L$	$\binom{\nu_\mu}{\mu}_L$	$\binom{\nu_\tau}{\tau}_L$	1/2	$\binom{+1/2}{-1/2}_L$	−1

Table 2.2: The quantum numbers of electroweak chiral doubles and singles, where Y is the weak hypercharge, T is the weak isospin, and T_3 the third component of T.

Because of the chirality of the theory, local gauge invariance requires all fermions to be massless, which is in disagreement with observed experimental results. A mechanism through which fermions and gauge bosons acquire mass while preserving gauge invariance is called “electroweak symmetry breaking” [6].

2.1.3 Electroweak symmetry breaking

The unified electroweak field theory provides a relationship between the $W^{\pm}$ and Z boson, but it does not provide a mechanism by which bosons and fermions acquire their masses. Fermion and boson masses are not allowed to be introduce directly into the Lagrangian, since this would break the local gauge invariance symmetry. Spontaneous symmetry breaking (SSB) provides a way for fermions and boson to acquire mass, while persevering the local gauge invariance of the the Lagrangian, but not the vacuum state of the system.

The Higgs mechanism induces SSB, through the interaction specified by its potential. This mechanism introduces a complex SU(2) doublet scalar field Φ with the Lagrangian density given by,

$$\mathcal{L}_{Higgs} = |D_\mu \Phi|^2 - V(\Phi), \tag{2.15a}$$

$$V(\Phi) = \mu^2 |\Phi|^2 + \lambda |\Phi|^4, \tag{2.15b}$$

$$D_\mu = \partial_\mu + i g_1 W_\mu \cdot \mathbf{T}/2 - i g_2 B_\mu, \tag{2.15c}$$

where W_μ^a and B_μ are the Yang-Mills fields corresponding to the $\mathrm{SU(2)}_L$ and $\mathrm{U(1)}_Y$ gauge group. The generators T^a of the group are represented by Pauli matrices. The general gauge invariant potential $\mathrm{V}(\Phi)$ represents the Higgs potential. For the Higgs potential to respect a lower bound, the parameter λ has to be positive, resulting in a potential with a parabolic shape around its minimum, when $\mu^2 < 0$. A non-zero vacuum expectation value results in a spontaneous breaking of the $\mathrm{SU(2)}_L \otimes \mathrm{U(1)}_Y$ symmetry to the $\mathrm{U(1)}_Q$ symmetry, by introducing the field Φ, called the “Higgs fields”, described by,

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \quad Y_\Phi = +1. \tag{2.16}$$

To generate gauge boson masses, the Higgs field is given a non-zero vacuum expectation value, v (determined experimentally),

$$\langle\Phi\rangle_0 = \langle 0|\Phi|0\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix}, \text{ with } v = \sqrt{-\frac{\mu^2}{\lambda}}, \tag{2.17}$$

where the charge component ϕ^+, has a vacuum expectation value of zero to respect the electromagnetic symmetry group $\mathrm{U}(1)_Q$. A simple choice for the ground state, $\phi_1 = \phi_2 = \phi_4 = 0$ and $\phi_3 = v$, can be used to generate gauge boson masses by using ϕ for the vacuum expectation value $\langle\Phi\rangle_0 = \langle 0|\Phi|0\rangle$ in the Lagrangian of Equation 2.15a, following the derivation found in [7],

$$\mathcal{L}_{Higgs} = (\frac{1}{2}vg)^2 W_\mu^+ W^{-\mu} + \frac{1}{8}v^2(W_\mu^3, B_\mu)\begin{pmatrix} g_1^2 & -g_1g_2 \\ -g_1g_2 & g_2^2 \end{pmatrix}. \tag{2.18}$$

By comparing the first term with the mass term expected for the charged boson $W^\pm$, $M_W^2 W^+W^-$, we have,

$$M_W = \frac{1}{2}vg_2. \tag{2.19}$$

Similarly, for the neutral vector bosons, we have,

$$M_Z = \frac{1}{2}v\sqrt{g_1^2 + g_2^2}, \tag{2.20a}$$

$$M_A = 0. \tag{2.20b}$$

The Higgs mechanism generates fermion masses though Yukawa-type interaction between the fermions and the Higgs field, with a coupling constant $g_f = \sqrt{2}m_f/v$ proportional to the fermion mass. Fermions acquire mass via Yukawa couplings to the Higgs doublet scalar field $\Phi(x)$, by introducing the following terms to the Lagrangian:

$$\mathcal{L}_f = -\lambda_e L\Phi e_R - \lambda\bar{Q}\Phi U_R - \lambda\bar{Q}\Phi D_R + h.c., \tag{2.21}$$

where L and Q represent the lepton and quark doublets and the anti-fermions acquire their masses by including the Hermitian conjugate (h.c.) in the Lagrangian. Following spontaneous symmetry breaking through the Higgs mechanism,

$$m_\ell = \frac{\lambda_\ell v}{\sqrt{2}},\ m_U = \frac{\lambda_U v}{\sqrt{2}},\ m_D = \frac{\lambda_D v}{\sqrt{2}}, \tag{2.22}$$

for the masses of leptons, the up-type and down-type quarks, respectively.

Besides generating masses for fermions and gauge bosons the mechanism also predicts the existence of an additional scalar particle, referred to as the Higgs boson. The search for the Higgs boson has been one of the main goals of experiments at the LHC. A new particle, compatible with the Higgs boson prediction, was discovered in 2012 by two experiments at the LHC simultaneously, CMS [8] and ATLAS [9], with a mass $\mathrm{m_H} = 125.03^{+0.26}_{-0.27}(\text{stat.})^{+0.13}_{-0.15}(\text{syst.})$ GeV [10], thus completing the SM picture.

2.1.4 Quantum Chromodynamics

The theory of the strong interaction, Quantum Chromodynamics (QCD), is described by a non-Abelian symmetry group $\mathrm{SU}(3)_C$. QCD interactions bind quarks and anti-quarks into the observed "hadrons", such as mesons and baryons. The generators of the $\mathrm{SU}(3)_C$ group, give rise to eight massless gauge bosons, called *gluons*. The theory of QCD perseveres the principle of gauge invariance, however, instead of a single electric charge, there exist three color charges r (red), g (green), and b (blue). Quark spinors are assigned a three-component color vector. The mediators of the strong interaction are colored gluons, where each gluon carries both a color and an anti-color. The gluon-gluon, quark-gluon interaction vertices are shown in Figure 2.2. These color anti-color combinations yield nine different gluon states, a color octet and one color singlet [6], given by the representations,

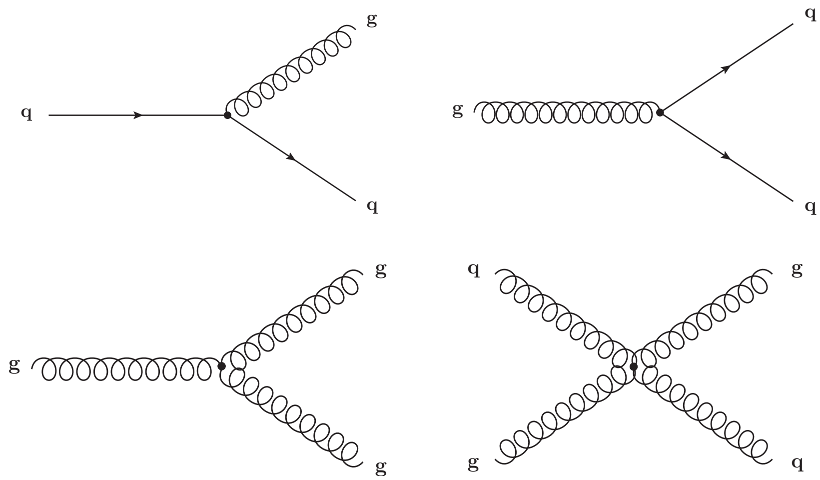


Figure 2.2: Feynman diagrams for interactions in QCD. Top: gluon emission and gluon absorption. Bottom: gluon-gluon interaction.

$$\begin{pmatrix} (rb+b\bar{r})/\sqrt{2} & -i(r\bar{g}-g\bar{r})/\sqrt{2} \\ (b\bar{g}+g\bar{b})/\sqrt{2} & -i(r\bar{b}+b\bar{r})/\sqrt{2} \\ (r\bar{r}+b\bar{b})/\sqrt{2} & -i(b\bar{g}+g\bar{b})/\sqrt{2} \\ (r\bar{g}+g\bar{r})/\sqrt{2} & (r\bar{r}+b\bar{b}-2g\bar{b})/\sqrt{2} \end{pmatrix}, \left((r\bar{r}+b\bar{b}+g\bar{g})\sqrt{3}\right). \tag{2.23}$$

The color singlet state with a colorless combination $(r\bar{r}+b\bar{b}+g\bar{b})$ is invariant under rotation in color space, and does not participate in the QCD interactions. From the field theoretic description these eight spin-1 gluon fields correspond to the 3×3 Gell-Mann matrices λ_a, corresponding to the generators of the SU$(3)_C$ group, where a $= \{1, 8\}$,

$$[\lambda^a, \lambda^b] = i f^{abc} \frac{\lambda^c}{2}, \tag{2.24}$$

and f^{abc} are the structure constants of the SU(3) symmetry group.

The locally gauge invariant QCD Lagrangian is then described by,

$$\mathcal{L}_{QCD} = \bar{\Psi} \not{D}_\mu \Psi - \frac{1}{2g_s^2} \mathrm{Tr}\{G_{\mu\nu} G^{\mu\nu}\}, \tag{2.25}$$

where Ψ are the quark fields, in the fundamental representation of the SU(3) symmetry group. The covariant derivatives are specified by,

$$\not{D}_\mu = i\gamma^\mu(\partial_\mu - ig_s\frac{\lambda_a}{2}G^a_\mu). \tag{2.26}$$

The gluon field strength tensor is given by,

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f_{abc} G^b_\mu G^c_\nu, \tag{2.27}$$

where $G^a_{\mu\nu}$ correspond to the gluon fields, in the adjoint representation of the SU(3) gauge group. The gauge coupling g_s is related to the strong coupling constant $\alpha_s = g_s^2/4\pi$.

2.1.5 Asymptotic freedom and confinement

The structure of the $SU(3)_C$ group implies interactions between gluons, which leads to *asymptotic freedom* and *confinement*. Calculating physical quantities such as cross sections and decay rates to leading order (i.e. tree-level Feynman diagrams), is straight forward. Evaluating higher order diagrams entails integrating over arbitrarily large momentum, referred to as ultraviolet (UV) divergences, since in the theory there is no intrinsic momentum cutoff. The procedure for isolating the UV divergences, knowns as “renormalization”, removes them from the physically measurable quantities [11]. The renormalization procedure redefines the bare parameters (not physically relevant), which are part of the QCD Lagrangian, including such quantities as the fermion masses, and the constants (e.g. α_s,). In addition, the renormalization introduces a energy scale λ_{QCD}.

The renormalization couplings depend on Q^2, the square of the momentum transfer of the process, required by the procedure for removing divergences. Due to the dependence of the strong coupling constant α_s on Q^2, the coupling is running,

$$\alpha_s(Q^2) = \frac{4\pi}{11 - \frac{2}{3}n_f \ln(Q^2/\Lambda_{QCD})}, \tag{2.28}$$

where n_f corresponds to the number of flavors active at the energy scale of the calculation and Λ_{QCD} is the scale at which the coupling diverges. The running coupling constant $\alpha_s(Q^2)$, is small at large momentum transfer, Q^2, so-called “asymptotic freedom”, and large at low momentum transfer, referred to as “confinement” of quarks and gluons within hadrons. At small distances, where the strong coupling is small, perturbative calculations still apply. While confinement is a non-perturbative effect, leading to “hadronization” of quarks and gluons, occurring when pairs of quarks and anti-quarks are created from the vacuum, and that then combine into colorless hadrons states.

2.2 Beyond the Standard Model

The SM successfully describes the interactions of fundamental particles, and with the discovery of a Higgs-like boson, the theory is completed. However, the SM still leaves many long-standing questions unanswered. The SM theory encounters several difficulties, if the SM is valid up to an energy scale Λ, the Higgs boson mass should receive radiative quantum corrections from vacuum polarization of the order Λ, that are quadratically divergent, leading to the “hierarchy problem”, if not properly cancelled. Possible solutions require fine-tuning at every order in the perturbative expansion. Supersymmetry (SUSY) provides a natural solution to the hierarchy problem, whereby fermion contributions to the loop correction of the Higgs boson mass are cancelled by a boson counterpart arising from additional superpartners of the SM particles. These superpartners have opposite contributions, leading to a cancelation of the Higgs mass divergences. Therefore, it is not necessary to fine-tune the parameters of the SM. Results from Run I of the LHC have ruled out many SUSY scenarios allowing models such as those involving vector-like quarks to become an appealing alternative solution to the hierarchy problem. The SM also lacks a theoretical explanation for why the number of generations of leptons and quarks is exactly three, as indicated by the Large Electron-Positron (LEP) experiments [12] at CERN. There are, in addition, other theoretical challenges. For example, fermion masses introduces a “naturalness problem”. In a natural theory, masses are expected to have the same order of magnitude, which is not the case in the SM, where the top quark mass is much larger than the rest of the quarks. Cosmological observation challenges the SM as well. The baryon anti-baryon asymmetry in the universe is not entirely understood. In addition, astronomical observations indicate that the energy density in the universe is composed of only about $4-5\%$ of ordinary baryonic matter, while $20-25\%$ is composed of dark matter, and the remaining $70-76\%$ of dark energy. Dark matter is non-baryonic matter that interacts gravitationally and possibly with a new type of weak interaction. Another feature not addressed by the SM is further unification of the interactions. The SM does not incorporate gravity, nor is there an accepted quantum theory of gravity. At the electroweak scale, the strength of gravity is negligible, but its quantum effects become relevant at the Planck scale Λ_{planck}. For these reasons, searches Beyond the Standard Model (BSM) are theoretically and experimentally of fundamental importance.

2.2.1 Vector-like quark phenomenology

The discovery of the SM-like Higgs boson and measurements of its production rate [13, 14], have evidently ruled out the possibility of extra quarks with chiral couplings, which receive their mass through Yukawa coupling with the Higgs doublet, such as fourth generation chiral quarks [15, 16]. Also, electroweak precision measurements severely constrain the existence of additional chiral quarks. Introduction of additional quarks would enhance the production rate of the Higgs boson [17], which is in direct conflict with the observed data at the Large Hadron Collider (LHC). Alternatively, vector-like quarks (VLQ) [18, 19] are hypothetical spin-1/2 particles that do not acquire their masses from Yukawa coupling to a Higgs doublet and can mix with SM quarks. Unlike chiral quarks, tree-level Flavor-Changing Neutral Currents (FCNC) [20, 21] are not suppressed for vector-like quarks, that therefore, do not obey the Glashow-Iliopoulos-Maiani (GIM) mechanism [22]. Vector-like fermions can introduce new sources of CP violation [23, 24, 25, 26]. Additional heavy quarks may result in enough intrinsic matter and anti-matter asymmetry to explain the observed baryon asymmetry in the universe [27]. More importantly, for the purpose of this analysis, vector-like quarks can be analyzed in a model-independent approach in terms of a few free parameters. They decay into a variety of final states with branching fractions treated as free parameters in the searches.

A fermion is "vector-like", if its left- and right-handed chiral states transform in the same way under the gauge symmetry group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, resulting in only the vector term remaining in the weak charged current. For this reason gauge invariant mass terms, $\Phi_L \Phi_R$, are not forbidden by the local gauge symmetry and are independent of the Higgs coupling. These quarks can be represented as a weak isospin singlet, doublet or triplet. In the same way that supersymmetry has the potential to stabilize the mass of the Higgs boson, partner fermions to the third generation can serve a similar purpose. Heavy vector-like quarks may help solve the hierarchy problem by reducing the size of the loop correction to the Higgs mass by adding extra radiative corrections to the Higgs mass [28]. In order for such cancellations to be effective, the masses of the partner quarks must be at or below the 1 TeV scale.

Despite representing a break from the established pattern of quark generations in the SM, vector-like quarks have been introduced in many extensions of the SM. A brief description of scenarios that predict the presence of the vector-like quarks include the following:

- In the so-called Little Higgs scenario, VLQs arise as partners of the SM fermions represented in larger multiplets [29, 30], which ensure cancelations of the top-loop quadratic divergence of the Higgs mass.

- In composite Higgs models, the VLQs are excited resonances of the bound states which form SM fermions [31].

- In extra-dimensional models with SM quarks in the bulk, the VLQs are general Kaluza-Klein (KK) excitations of those bulk fields [32].

- In non-minimal supersymmetric (SUSY) extension of the SM, VLQs are introduced to increase corrections to the Higgs mass without affecting the electroweak precision measurements [33].

At the LHC, the main production channel for vector-like quarks is gluon fusion, as shown in Figure 2.3. QCD production through gluon fusion of b'$\bar{\text{b}}$' pairs, followed by b' decay, lead to various final states.

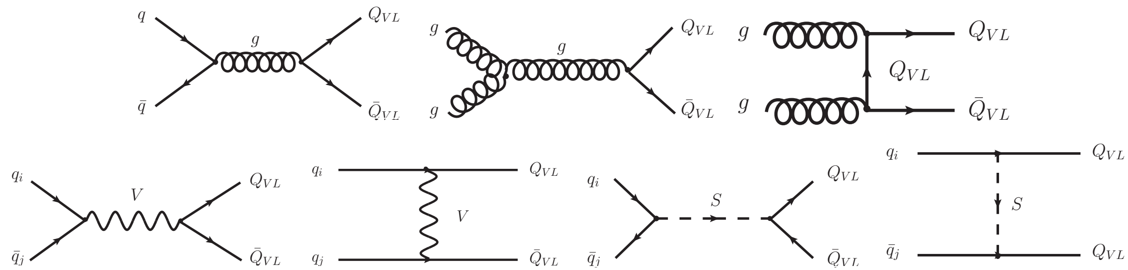


Figure 2.3: Top: shows the dominant and model-independent QCD contributions, the most important of which is the gluon-gluon fusion at the LHC. Bottom: shows the subdominant and model-dependent Electroweak contributions.

The motivation for pair-produced vector-like quarks, is that such searches depends only minimally on the strength of the coupling to weak bosons, making it essentially model-independent. The coupling only needs to be large enough to ensure prompt decays.

In this dissertation, a search for a heavy b' partner quark that is pair-produced in *pp* collisions is presented. Final sates considered include three or more leptons. We assume that the b' quark mass, $M_{\text{b'}}$, is larger than the sum of the W and top quark mass. Vector-like fermions decay by exchange of electroweak gauge and Higgs bosons, $\text{W}^{\pm}$, Z, and H. Moreover, we keep the search for new physics general by also considering flavor-changing neutral current decays to a bottom quark with a Z boson (i.e. b' $\rightarrow$ bZ) or to a bottom quark with a SM Higgs bosons (i.e. b' $\rightarrow$ bH). The latter could potentially be the more significant of the two FCNC decay channels [34]. We consider the branching fraction to the three modes to be free, but subject to the constraint that the branching fractions add to unity. We do not consider other decay modes of the b'. A b' quark can potentially decay in three different states, allowing for six distinct event topologies for pair-produced b' quarks: bZbZ, tWtW, bHbH, bZtW, bZbH, and tWbH. Figure 2.4 illustrates an example of vector-like quark pair production with subsequent decay to the bWbW channel. Taking into consideration the top decay to bW, the possible decay modes are bbZZ, bbWWWW, bbHH, bbWWZ, bbZH and bbWWH. These in turn can have multilepton final states from leptonic decay of the Higgs and vector bosons. We assume the SM Higgs boson to have a mass of 125 GeV with branching ratios obtained from Reference [35].

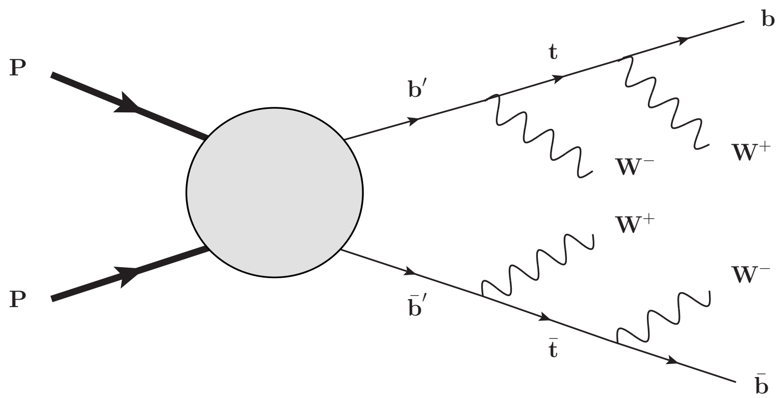


Figure 2.4: Feynman diagram of proton-proton collision to b'b' production to bWbW decay.

Chapter 3

Experimental Apparatus

In this chapter, we give a description of Large Hadron Collider (LHC) and the Compact Muon Solenoid (CMS) detector with a focus on the features relevant for this analysis. More details can be found in the references provided.

3.1 The Large Hadron Collider

The LHC [36] is a particle accelerator located along the Swiss-French border at the European Organization for Nuclear Research (CERN). The LHC was designed to accelerate and collide protons (or heavy ions) at a center-of-mass energy of $\sqrt{s} = 14$ TeV with an instantaneous luminosity of $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1}. Figure 3.1 shows the 27 km circumference ring with its two multi-purpose experiments: the CMS and the ATLAS detectors. Additionally, three more specialized experiments are also placed around the ring: LHCb to explore B-physics in detail, ALICE for heavy ion collisions studies, and TOTEM to measure the total proton-proton (pp) cross section to high precision.

The LHC has not reached yet its design energy of 14 TeV. During Run I from 2010−2013, the two proton beams were brought to collisions with 3.5 TeV and 4 TeV per beam, reaching a center-of-mass energy, $\sqrt{s}$, of 7 and 8 TeV, respectively. Protons for the LHC beams are produced from a tank of hydrogen gas and injected from a Duoplasmatron source into a linear accelerator (LINAC), that accelerates these protons to 50 MeV. Afterwards, the protons are transferred to the Proton Synchrotron Booster (PSB), which further increases their energy to 1.4 GeV. This step is followed by the Proton Synchrotron (PS), which

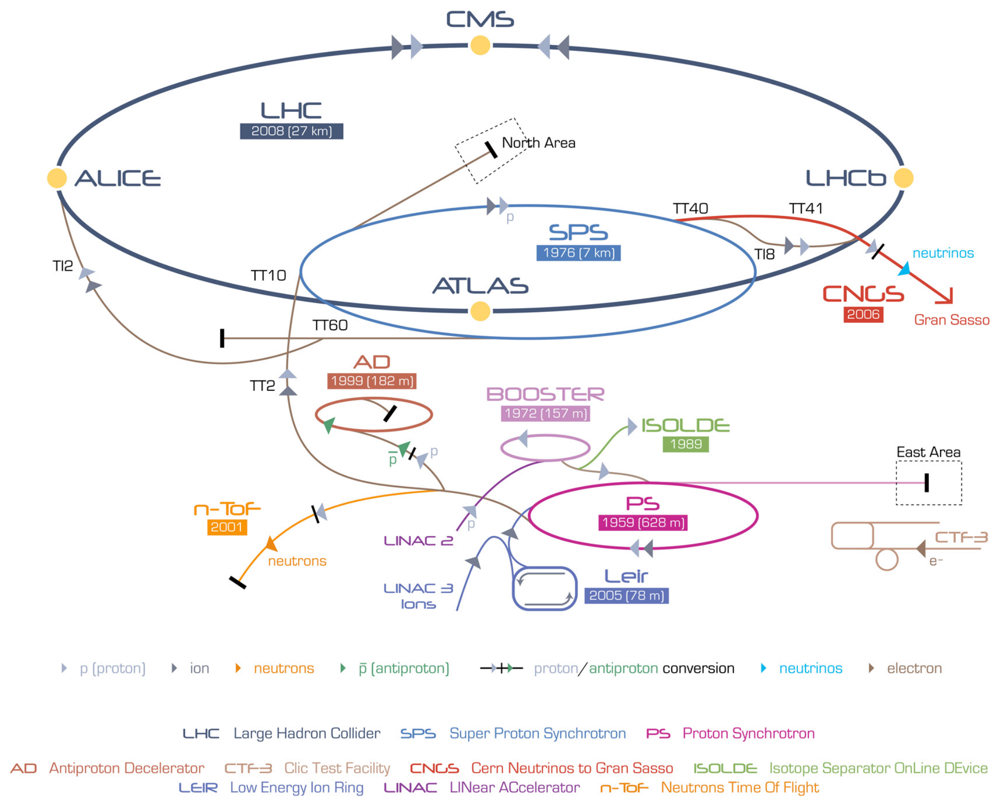


Figure 3.1: The CERN accelerator complex [37]. Protons are boosted as they travel through the LINAC, boosted into the main LHC ring.

accelerates them to 26 GeV. Finally, the Super Proton Synchrotron (SPS) provides proton bunches with energy of up to 450 GeV. The injection sequence provides the LHC with up to 2808 bunches of protons per beam, with a bunch length of about 53 mm and a transverse width of 15 μm, with more than 10^{11} protons per bunch and a 25 ns beam crossing interval. A more detailed list of the parameter values for the LHC can be found in Table 3.1. During Run I the nominal operating conditions were a bunch spacing of 50 ns with a maximum number of 1287 bunches per beam [38].

The LHC ring consists of superconducting dipole magnets (NbTi) that are designed to provide a magnetic field up to 8.3 Tesla. This magnetic field is needed to maintain the circular path of the protons around the main LHC ring. The dipole magnets are designed to operate at temperature well below 2 Kelvin by using superfluid helium for cooling. Radio Frequency (RF) cavities provide the acceleration of the proton bunches inside the LHC.

There are in total eight RF cavities, which operate at 400 MHz. The proton beams which circulate in both directions in separate rings are brought together for collisions within the various experiments distributed around the LHC, such as CMS, and ATLAS. Quadrupole magnets are used to squeeze the beams at the collision point.

Besides the beam energy, another important parameter is the luminosity, which is proportional to the rate of collisions. The total instantaneous luminosity [$\mathrm{cm}^{-2}\mathrm{s}^{-1}$] can be expressed as,

$$\mathcal{L} = \frac{N_b^2 n_b^2 f_{rev} \gamma}{4\pi \epsilon_n \beta^*} F, \tag{3.1}$$

where N_b is the number of particles per bunch crossing, n_b is the number of bunches per beam, and f_{rev} is the revolution frequency. The emittance ϵ_n corresponds to the average normalized phase space occupied by the beam and describes the spread in momentum and position of the protons. The measure of the transverse beam width at the collision point, is referred to as β^*. The cross sectional area $A = 4\pi\epsilon_n\beta^*/F\gamma$, where F is the geometric factor describing the crossing angle at the interaction point, and γ is the Lorentz factor. The total integrated luminosity delivered in 2012 by the LHC experiment reached $L = \int \mathcal{L} \cdot dt = 23.3\ \mathrm{fb}^{-1}$. A detailed discussion about luminosity measurement is given in Section 3.2.7.

Parameters	**Values for *pp* collisions**
center-of-mass energy	8 TeV
Number of protons per bunch	1.1×10^{11}
Number of bunches	2808
Designed luminosity	10^{34} $\mathrm{cm}^{-2}s^{-1}$
Luminosity duration	10 Hours
Bunch Length	53 mm
Beam radius at interaction point (IP)	15 μm
Time between collisions	25 ns
Bunch crossing rate	40 MHz
Circumference	27 Km
Dipole field	8.3 Tesla

Table 3.1: LHC design parameters for *pp* collisions in 2012 [39].

3.2 The Compact Muon Solenoid detector

The CMS experiment [40] is situated near the village of Cessy in France. It is a general purpose particle detector designed to study *pp* collisions at the LHC. A main feature of CMS, is the compact superconducting solenoid designed to precisely measure muons. The solenoid produces a magnetic field of 3.8 Tesla parallel to the beam axis allowing the momentum measurements of charged particles in the inner tracker system from the bending of the trajectories of these particles. The CMS detector has an onion-like structure covering a 2π azimuthal angle around the beamline. As one moves radially outward from the beam, there is the tracker system, followed by the electromagnetic and hadronic calorimeter detectors. The outermost part is composed of the muon detectors. The separate parts are described in more detail in the following sections, while further details can be found in Ref. [40]. Figure 3.2 shows a schematic of the CMS detector.

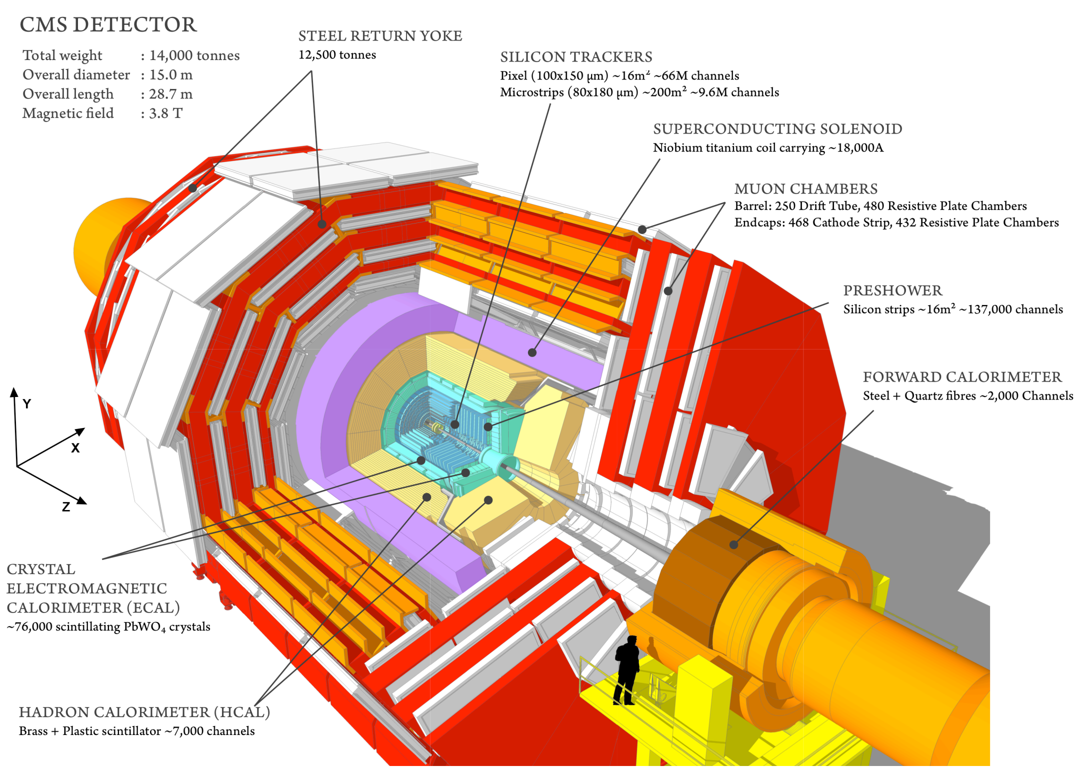


Figure 3.2: CMS detector [41].

3.2.1 Detector coordinate system

The CMS detector is described using a right-handed coordinate system, with the origin centered at the nominal collision point of the detector. The x-axis of the coordinate points radially inward towards the center of the LHC ring, the y-axis points vertically upward, and the z-axis points parallel along the counterclockwise beam direction. The detector is cylindrically symmetric along the *pp* beam direction and is therefore described using a cylindrical coordinate system. The polar angle θ is measured from the z-axis (longitudinal direction), the azimuthal angle ϕ is measured with respect to the x-y plane and the radial direction is given by r. We need to consider relativistic invariant coordinates, at a center-of-mass energy of 8 TeV, when referring to a particle's trajectory. The rapidity y is used instead of θ, since the difference between rapidities of two particles is invariant with respect to Lorentz boosts along the z-direction (beam axis), while θ is not invariant.

The rapidity of a particle is given by,

$$y = \frac{1}{2} \ln \frac{\mathrm{E} + \mathrm{p_z}}{\mathrm{E} - \mathrm{p_z}}, \tag{3.2}$$

where E represents the energy of a particle and p_z is the particle's momentum vector in the z-direction. In hadron collisions, Lorentz-invariance is importance, since it allows the definition of observables independent of the p_z of the initial state. For relativistic particles, when $E \gg m$, rapidity can be approximated by the "pseudo-rapidity", η, and can be expressed in terms of the polar angle, θ, and defined according to,

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

The direction perpendicular to the beamline corresponds to $\eta = 0$ ($\theta = 90^o$), while for $\eta = 4$ ($\theta = 2.1^o$) the particle points almost parallel to the beamline. We denote a particle's energy and momentum in the transverse direction by E_T and p_T, which can be calculated from their x- and y-components, for example,

$$p_{\mathrm{y}} = p_{\mathrm{T}} \sin\theta, \tag{3.4a}$$

$$p_{\mathrm{x}} = p_{\mathrm{T}} \cos\theta. \tag{3.4b}$$

Particles that escape detection by the CMS detector produce an imbalance of the total momentum measured in the transverse plane, demoted by $\vec{\not{E}}_{\mathrm{T}}$, and its magnitude is called "missing transverse energy",

$$E_{\mathrm{T}}^{\mathrm{miss}} = -\sum_{i} p_{\mathrm{T}}^{i}, \tag{3.5}$$

where the sum is over all detected particles i in the collision event.

3.2.2 Magnet

The CMS detector uses a strong 3.8 Tesla magnetic field provided by a superconducting coil [42], which bends charged particle trajectories allowing for precise measurements of the particle momenta. The CMS magnet design aim was to achieve a momentum resolution ($\Delta p/p$) of about 10% for muons with a momentum of 1 TeV over a pseudo-rapidity region up to $|\eta| = 2.4$. With this level of precision, it is possible to unambiguously identify the sign of the muon charge. The magnetic system consists of two main components, the superconducting solenoid and the iron return yoke in the barrel and endcap. The solenoid has a length of 12.9 m with an inner diameter of 5.9 m. The iron yoke weighs 7000 tons, more than half of the total weight of the CMS detector. The solenoid is comprised of five separate sections placed within a cryostat and cooled to 1.9 K to maintain superconductivity. The iron return yoke is interlaced by the muon detectors which are located outside the solenoid.

3.2.3 Tracking system

The purpose of the inner tracker system [42] of the CMS detector is to provide precise measurements of the trajectories of charged particles originating from the *pp* collisions produced by the LHC, including precise reconstruction of secondary vertices. The tracker system consists of the high precision pixel detector and an outer strip tracker, designed to reconstruct particles trajectories known as "tracks". These tracks are later used to reconstruct muons, electrons, and hadrons with a momentum resolution accuracy of about 1.5% for charged particles with p_T of 100 GeV, and high efficiency in the pseudo-rapidity region $|\eta| < 2.5$.

The tracker volume utilizes silicon sensor technology, which can tolerate large radiation doses without major deterioration in its performance. Charged particles transversing the silicon sensors produce electron-hole pairs. A signal is measured as a result of the positive and negative charge carries drifting towards the surface electrodes.

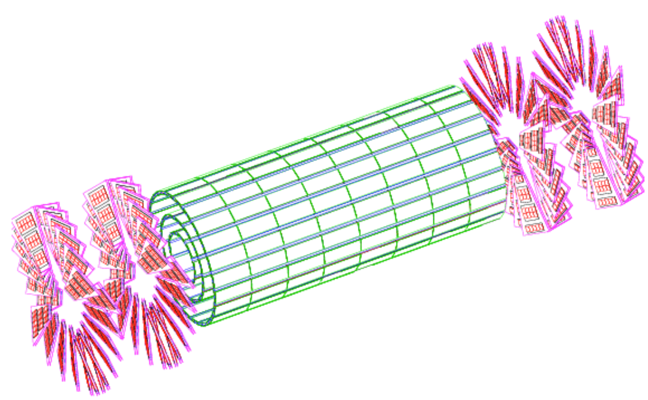


Figure 3.3: Schematic view of the CMS pixel detector [43]. The three barrel pixel (BPIX) layers are shown in green and the four endcap disks (FPIX) are shown in pink.

Pixel detector

The silicon pixel detector, referred to as the Barrel Pixel (BPIX) system is comprised of three barrel layers located at radial distances of 4.4, 7.3, and 10.2 cm each with a length of 53 cm. Each layer is further divided into ladders covering the full cylindrical surface, 20 ladders in the first layer, 32 ladders for the following layer, and 44 ladders for the last layer. Every ladder is composed of 8 modules, each of which has a 4×2 array of 8 mm by 8 mm Read-Out Chips (ROC) with a pixel size of $100 \times 150\ \mu\text{m}^2$. A total of 768 pixel modules comprise the pixel barrel detector.

A Forward Pixel (FPIX) system supplements the barrel pixel detector, organized into four disks, two at each end of the BPIX detector, consisting of blades arranged in a fan-like structure. There are two disks per side at $|z| = 34.5$ cm and $|z| = 46.5$ cm and extending from 6 to 15 cm in the radial direction. Figure 3.3 shows a schematic of the pixel system, with the BPIX in green and FPIX in pink. The pixel detector covers a pseudo-rapidity region of $|\eta| < 2.5$.

Silicon tracker

The silicon strip detector is also divided into barrel and disks components that cover the central and forward regions, respectively. The two parts situated in the barrel region are called the Tracker Inner Barrel (TIB) and the Tracker Outer Barrel (TOB) and are arranged concentrically about the beamline. The TIB extends up to $|z| = 65$ cm, comprised of four layers, based on silicon sensors with a strip pitch between 80 and 100 μm. The inner two layers contain "stereo" modules, which provide a measurement of ϕ. The TOB extends to $|z| = 110$ cm and is comprised of six layers with strip pitches between 120 and 180 μm. Figure 3.4 shows the tracker system, where each line-element represents a silicon strip module and double line-elements are the stereo modules, which are mounted back to back.

In order to cover the forward region, two endcap tracker systems are installed on both sides of the strip barrel detector, the first is called the Track Inner Disk (TID) and the second the Tracker End Cap (TEC). The TIC consists of three disks, filling the empty region between the TIB and TEC. The TEC component is comprised of nine disks covering the region between 120 cm $< |z| <$ 280 cm with a pseudo-rapidity coverage of $0 < |\eta| < 2.5$.

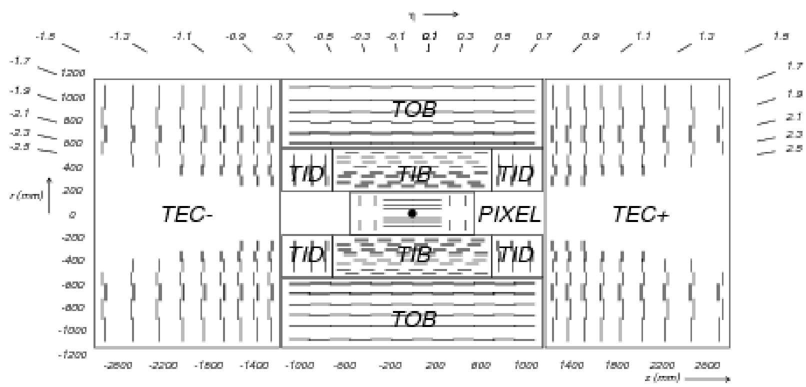


Figure 3.4: Schematic cross section through the CMS tracker in the $r-z$ plane [44]. The line-elements correspond to a detector module and double line-element indicates stereo strip modules.

3.2.4 Calorimetry

In addition to measuring precisely the tracks and momenta of charged particles, it is also important to measure their energy as precisely as possible, especially for neutral particles that do not leave a signature in the tracker system. There are two distinct calorimeter systems in CMS, the Electromagnetic Calorimeter (ECAL) and the Hadronic Calorimeter (HCAL), with the aim of measuring the energies of all particles (except neutrinos and muons) produced in the pp collisions. Calorimeters are designed such that incident particles deposits all their energy while traversing the calorimeter.

Electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) [42] is designed to measure the energy of electrons and photons. Electrons primarily loose their energy by ionization due to interactions with the electric field of the atomic electrons of the material and emission of Bremsstrahlung radiation arising from interactions with the charge atomic nucleus. Photons interact with the atomic nuclei of the material and convert to electron-positron pairs. This electromagnetic shower in the material leads to further production of secondary photons and electron-positron pairs depending on the radiation length of the material. Eventually the electrons and positrons dissipate their energy through the ionization and excitation processes. The radiation length X_0 is defined as the characteristic length, after which the energy of a highly-energetic electron is reduced by a factor of $1/e$, while high-energy photons yield e^+e^- pairs with a mean free path of $7/9X_0$.

The ECAL, shown schematically in Figure 3.5, consists of about 80,000 lead-tungstate scintillating crystal ($PbWO_4$), chosen because of the material's short radiation length of $X_0 = 0.89$ cm and its fast scintillation, roughly 80% of the light is emitted within 25 ns. The $PbWO_4$ crystal are arranged cylindrically around the tracking system and covers a pseudo-rapidity region $|\eta| < 1.48$. These crystals are highly transparent and scintillate as electrons transverse them. Photodiodes are used to measure the scintillator light emission. The photodiodes work on the basis of the photoelectric effect, whereby photons free an electron from its bound state in the material, converting the light signal into a measurable current. CMS has installed specially designed silicon Avalanche Photodiodes (APD) and Vacuum Phototriodes (VPT) for use in the endcap and barrel region, respectively.

The ECAL relative energy resolution σ_{rel} for electrons hitting the center of a crystal is parametrized by,

$$\frac{\sigma_{rel}(E)}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c, \tag{3.6}$$

where the quadratic sum is denoted by $\oplus$, defined as $x \oplus y = \sqrt{x^2 + y^2}$. The parameter $a = 2.8\%$ represents the intrinsic stochastic fluctuations of the number of particles produced in the shower process [45]. The parameter $b = 12\%$, describes the contribution from electronic noise, and the parameter $c = 0.3\%$, is a constant term that incorporates the different systematic uncertainties due to detector non-uniformity and calibration uncertainties.

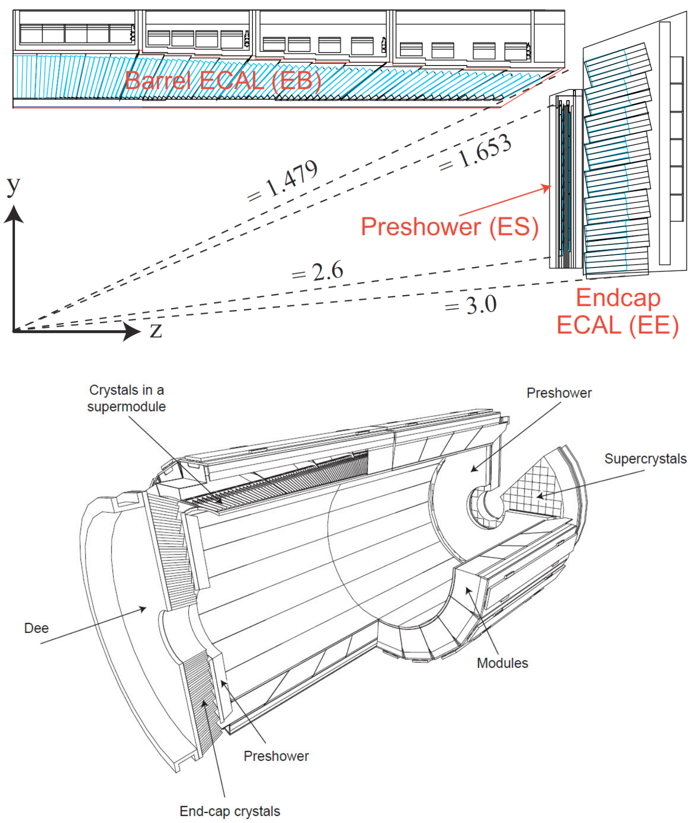


Figure 3.5: Top: Qauter view of the CMS electromagnetic calorimeter system. Bottom: ECAL presenting the arrangement of crystal modules, super modules, encamps, and preshower for front and bottom [44].

To ensure a precise energy resolution measurement, and to reduce *punch-through* the ECAL detector material thickness is greater than $22X_0$ in the barrel and greater than $24X_0$ in the endcaps.

Hadronic calorimeter

The CMS hadronic calorimeter (HCAL) [42] is a sampling calorimeter detector situated outside the electromagnetic calorimeter and inside the superconducting coil. The HCAL is designed to measure the energy of hadrons, such as pions, kaons, and protons, allowing the reconstruction of hadronic jets and indirectly the missing energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) with high precision, discussed in more details in Section 5.4. Hadrons contain quarks and undergo strong interactions when transversing a material thereby producing a hadronic shower of particles. The nuclear interaction length λ_I is characterized by the mean distance travelled by hadrons, before loosing $1/e$ of their energy through inelastic nuclear interactions.

The hadronic calorimeter is located radially between the outer radius of the ECAL, at $r = 1.77$ m, and the superconducting solenoid, at $r = 2.95$ m. The HCAL consists of brass layers as absorbers, interleaved by plastic scintillator layers as the active material. The brass layers have 60 mm thickness in the barrel and 80 mm thickness in the endcaps, while the scintillator tiles are 4 mm thick. As particles travel through the absorbers, they interact with the detector material producing a shower of secondary particles. The charged particles in the shower produce scintillation light are collected by wavelength shifting fibers. The readout electronics, consists of hybrid photodiodes (HPD). Photons striking the photocathode of the HPD free the electrons in the material, which are accelerated towards a pixelated silicon diode. The signal is amplified through secondary electrons produced in the processes.

Similar to the ECAL, the HCAL is divided into sub-detectors comprised of a barrel region (HB) and two endcap regions (HE), and hadron forward (HF) calorimeter, as shown schematically in Figure 3.6. These sub-detectors cover a pseudo-rapidity range of $|\eta| < 3.0$. The HCAL extends beyond the solenoid magnet in both the barrel and endcap sections to reduce the amount of energy that can escape detection with an additional layer, called the Outer Hadron calorimeter (HO), situated outside the solenoid magnet in the barrel region in order to catch the tail of the shower. The HO uses the iron return yoke as a passive absorber material and scintillator plates as the active material. In addition, the HF calorimeter made of radiation-hard quartz fiber and steel absorbers, further extends the η range up to the region of $|\eta| < 5.0$. The HF is located 11.2 m away, in the z-direction, from the interaction point. The HF measures signals from Cherenkov radiation produced in the quartz fiber from light emitted when charged particles travel faster than the speed of light in the material.

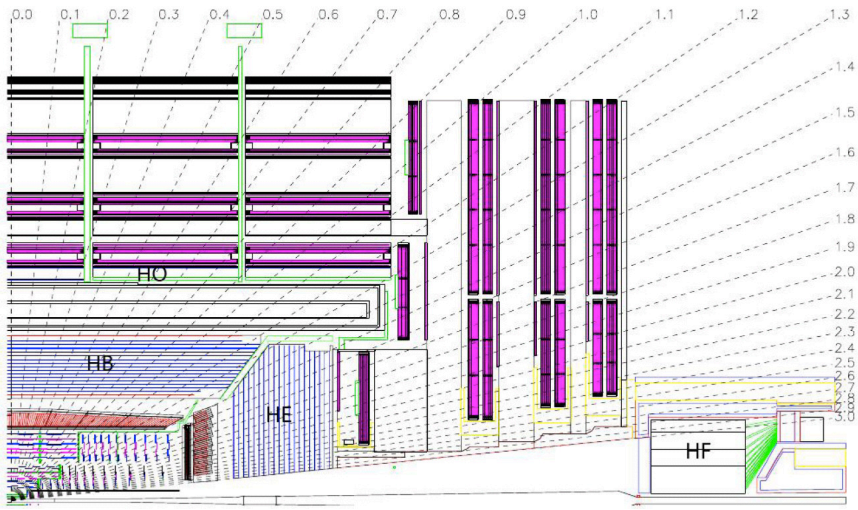


Figure 3.6: Longitudinal view of one quarter of the detector, showing HCAL components: hadron barrel (HB), hadron outer (HO), hadron endcap (HE) and hadron forward (HF) [44].

The HCAL energy resolution, σ/E, for a given energy E is parameterized by,

$$\frac{\sigma(E)}{E} = \frac{s}{\sqrt{E}} \oplus c, \tag{3.7}$$

where $s = 65\%$ is the stochastic term and $c = 5\%$ is a constant term. The HCAL is designed to detect signals ranging from a single minimum ionizing muon up to an energy deposit of 3 TeV.

3.2.5 Muon system

The purpose of the muon system [42] is to identify and reconstruct the trajectory of muons by accurately measuring their momenta. Muons are minimum ionizing particles (MIP) that do not interact strongly with the inner regions of the CMS detector, although they leave track signatures from ionization. For this reason, the muon system is located outside of the tracking and calorimeter systems. The muon system is located outside of the solenoid and covers a pseudo-rapidity region of $|\eta| < 2.4$. Three types of gaseous detectors are employed to measure muons, the Drift Tubes chambers (DT), Resistive Plate Chambers (RPC), and

Cathode Strip Chambers (CSC). As muons transverse the gas chambers they ionize the atoms along their trajectory. The positively charged ions and negatively charged electrons produce a signal, generated by their drift in the applied electric field between the positive voltage anode and the negative voltage cathode.

The muon barrel region is divided into four stations and is separated by layers of the iron return yoke. Figure 3.7 shows the layout of the CMS muon system. The stations labeled MB1-4 are located at radii of approximately, 4.0, 4.9, 5.9, and 7.0 m from the interaction point (IP), where stations MB1 is closest to the IP. The DT (green) chambers are installed between two RPCs, covering a pseudo-rapidity range of $|\eta| < 1.2$ with four concentric cylinders. The endcap regions uses CSCs (violet), consisting of four stations with a total of 468 cambers, covering an additional pseudo-rapidity region of $0.9 < \eta < 2.4$. Every cathode strip chamber consists of six layers of cathode plates with radial strips and six anode wire planes, which are arranged to maximize coverage of muons. The main feature of the CMS muon system is its ability to "trigger" on the p_{T} of muons with high efficiency. A discussion on triggers within CMS is given in Section 3.2.6. Additionally, the muon system has dedicated RPCs (red) installed in the barrel and endcap regions, which cover a pseudo-rapidity region $|\eta| < 1.6$. A total of 36 RPCs are mounted on the two outer rings of each of the endcap stations. The RPCs are gaseous parallel-plate detector cambers, where the plates function as the anode and cathode component. Muons traveling through the chamber ionize the gas atoms, where the free electrons in turn strike other atoms resulting in an avalanche of electrons and photons. The endcap muon system allows muon tracks to be reconstructed with high precision in ϕ.

3.2.6 Trigger System

A primary challenge for data taking is the 25 ns beam crossing interval, corresponding to a crossing frequency of 40 MHz. The peak luminosity delivered during 2012 by the LHC, 8×10^{33} cm^{-2} s^{-1}, times the total inelastic cross section of about 70 mb, leads to

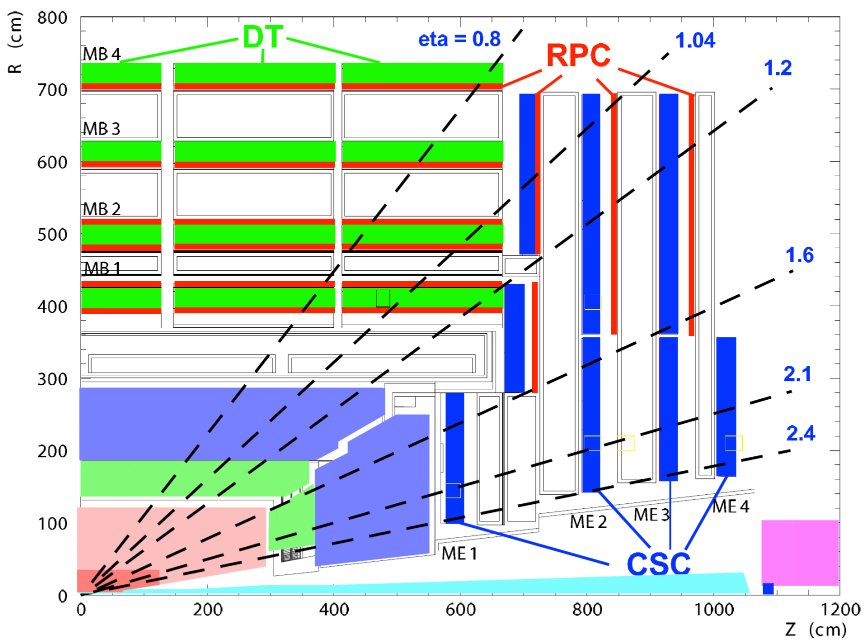


Figure 3.7: Longitudinal view of the CMS muon system [44]. The MB1-4 stations in the barrel and endcap, as well as the three sub-system: the drift tube chambers (DTs), resistive plate chambers (RPCs), and cathode strip chambers (CSCs) are shown.

a maximum event rate of 5.6×10^8 s^{-1}. To keep the event rate at a manageable level, for production, storage, and processing, this rate needs to be significantly reduced. This reduction is achieved by looking for event characteristics that indicate events worth keeping for further analyses, such as searches for new physics or precision measurements. CMS uses a "trigger system", to record events, for further processing. The CMS trigger system [42] consists of two levels: the Level-1 (L1) trigger reduces the data rate to under 100 kHz and a second, higher level trigger, HLT further reduces the trigger event rate to under 100 Hz. The trigger systems are described in more details in the following sections.

Level-1 trigger system

The Level-1 trigger is implemented using custom designed programable electronic hardware to reduce the data rate of events down to 100 kHz, achieving a trigger decision, on whether to accept or reject an event, within 3.2 μs. The decision made by the trigger is based on primitive trigger objects, provided by the calorimeter and muon sub-systems. Figure 3.8

shows the L1 trigger system logic scheme.

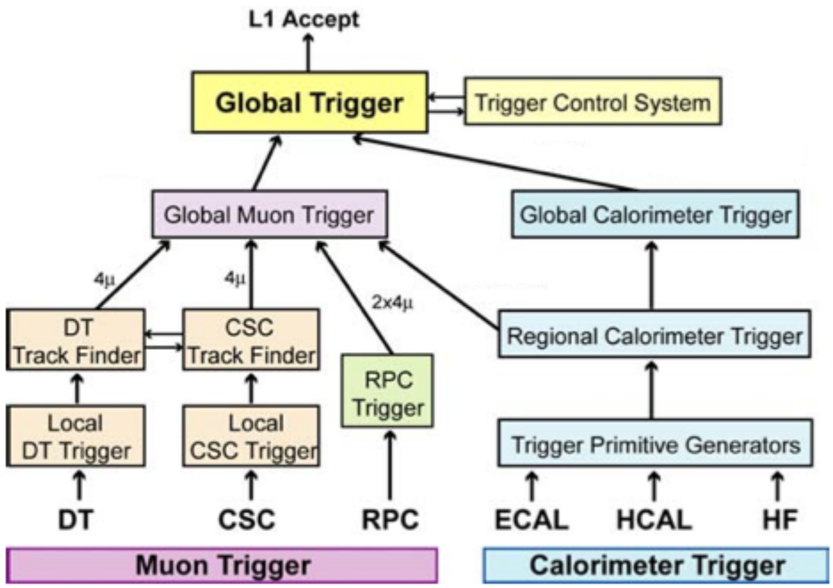


Figure 3.8: Schematic Level-1 trigger system [42]. L1 trigger decision flow of CMS before data transfer to the DAQ.

Level-1 calorimeter trigger

The calorimeter component of the L1 trigger consist of the Trigger Primitive Generators (TPG) that collect energy deposit information from the calorimeters. The TPG defines a list of trigger towers, where a trigger tower in the barrel region maps single HCAL cells on to 5×5 crystal arrays of the ECAL and sums up the respective energies for a corresponding (η, ϕ)-element of 0.087×0.087 [46]. The trigger primitives from both the ECAL and HCAL are subsequently transmitted to the Regional Calorimeter Trigger (RCT), which combines information from the 4×4 trigger towers. The RCT finds the electron/photon candidates and determines the energy deposited. These electrons/photons are identified based on the trigger towers with largest energy deposit as well as being required to satisfying quality selections depending on the ratio of HCAL to ECAL, H/E, energy deposits. Four isolated and four non-isolated electron/photon candidates with the highest E_{T} are sent to the Global Calorimeter Trigger (GCT). The global component is comprised of the GCT and Global

Muon Trigger (GMT). Information collected by the GCT and GMT, further ranks the highest calorimeter and assign muon objects, and is transferred to the Global Trigger (GT), where the second level of trigger decision of whether to accept or reject an event is performed by the High-Level Trigger (HLT).

Level-1 muon trigger

The muon trigger sub-systems are implemented using the DT, CSC, and RPC components. The local level of reconstruction begins with receiving electronic information from the DT system in the barrel region and the CSC system in the endcap region. This information is collected by the DT Track Finder (DTTF). The DTTF links the hit segment information from the different chambers, reconstructing tracks consistent with the trajectory of a muon, assigns a transverse momentum, and a charge to the candidate particle. The DT provides the four highest p_{T} candidates, found in the barrel and endcap, to the GMT. The GMT matches the muon from the DT, CSC, and RPC sub-systems, each of which provide muon candidates independently, and combines their information. The calorimeter information from the RCT are used to determine whether a muon candidate is isolated. Finally, the GMT sends the best four muon candidates, including their isolation criteria, and correlations of the different muon sub-system, to the GT, responsible for the the accept-reject decision.

High-Level Trigger system

The High-Level Trigger (HLT) [42] system is a software trigger system implemented on a computing farm with over 13,000 CPU cores. A special CMS reconstruction software is used, referred to as “online reconstruction”. The HLT system has access to the full detector readout allowing it to make complex computations, designed to further reduce the output event rate of the L1 trigger, $\mathcal{O}(100\ \mathrm{kHz})$, down to a more manageable rate of $\mathcal{O}(100\ \mathrm{Hz})$. Events accepted by the HLT trigger are sorted in primary datasets (PD), based on the trigger decisions and written to storage. Event reconstruction performed further downstream is referred to as "offline reconstruction", a more detailed description is found in Chapter 5. The HLT selects events using a list of triggers, selecting an event based on objects, above certain energy or momentum thresholds. The trigger objects can be single objects, such as muons, electrons, taus, jets, and photons, or can be composite objects, for example, $E_{\mathrm{T}}^{\mathrm{miss}}$. The full list of triggers is called the HLT "trigger menu". The trigger name indicates the HLT selection, for instance, HLT_El20_eta2p1 corresponds to an event that contains an electron above 20 GeV threshold and $|\eta| < 2.1$.

Triggers can be configured to accept every event that pass selection requirements, referred to as "un-prescaled" triggers. Triggers that accept one out of every N events that passes the criteria, are known as "pre-scaled" triggers, where N indicates the "trigger prescale" value. A pre-scale value of 1,000 corresponds to recording one event for every 1,000 events passing the trigger selection requirements. This allows the LHC to maintain a manageable output trigger rate. Pre-scale triggers are typically used for studies on trigger efficiencies where event selection is not important while un-prescaled triggers are used in cases where higher event acceptance is important, such as searches for rare processes.

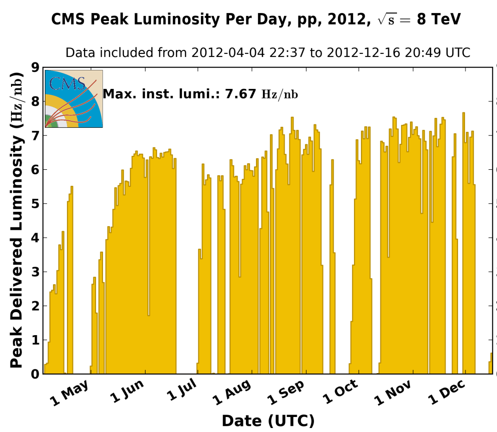
3.2.7 Luminosity measurement

In collider physics, processes are characterized by their cross section. An accurate measurement of the relative luminosity delivered and recorded by CMS is necessary for the cross section determination. Data is recorded in separate runs, with a granularity referred to as a "luminosity section" (LS) , corresponding to about a 23 second time interval. The luminosity $\mathcal{L}$ is measured by both the HF calorimeter and the pixel detector, with the pixel method preferred since it has smaller dependencies on multiple interactions and other beam conditions. The luminosity of pp collisions is determined from the average number of pixel clusters $\langle n \rangle$ in zero-bias events (zero-bias triggers require only bunch crossing occurrences):

$$\mathcal{L} = \frac{\langle n \rangle f_{rev}}{\sigma_{visible}}, \tag{3.8}$$

where f_{rev} is the proton beam revolution frequency with 11246 Hz and $\sigma_{visible} = A\sigma_{inelastic}$ is the total inelastic pp cross section within the detector acceptance A. The calibration of $\sigma_{visible}$ is performed using Van der Meer (VdM) scans. The VdM technique involves scanning the proton beams through one another in the horizontal and vertical direction to determine the width of the beams at their point of collision.

The luminosity measurement is calculated from the number of pixel clusters per event for the specified LS for a given dataset, which is then multiplied by the luminosity lifetime. The total integrated luminosity L is then the sum of all luminosity sections relevant to the analysis with an associated uncertainty is about 2.5% [47]. The delivered (blue) and recorded (yellow) integrated luminosities for the 2012 data taking period are shown in Figure 3.9.



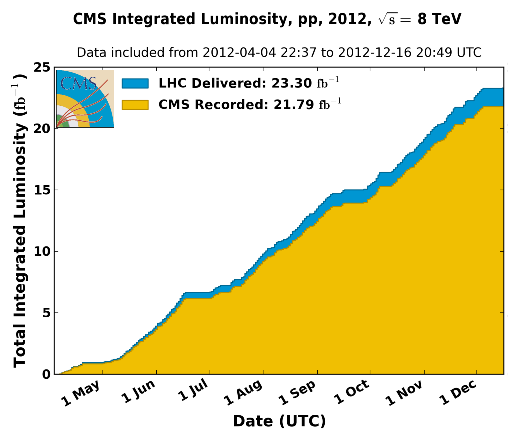


Figure 3.9: Peak instantaneous luminosity (left), and cumulative integrated luminosity (right) by CMS [38], delivered (blue) and recorded (yellow), as function of time. This analysis uses a certified subset of the recorded data, which corresponds to 19.5 fb^{-1}.

Chapter 4

Collision and simulated data

In the following sections we will briefly discuss the data used in this analysis. Additionally, an overview is given about the simulation of signal and background events.

4.1 Collision data samples

The analysis is performed with a total integrated luminosity of 19.5 fb^{-1} of proton-proton collisions collected by the CMS experiment during the 2012 run of the LHC at $\sqrt{s} = 8$ TeV. We used only data that were certified as good for analysis, meaning all sub-detectors, triggers, and physics objects such as leptons, tracks, photons, jets, and $E_{\mathrm{T}}^{\mathrm{miss}}$, have the expected performance. Events are stored in primary datasets following a positive trigger decision, as described in Section 3.2.6. Afterwards, the collected events are reconstructed following the algorithms described in Chapter 5. Multiple primary datasets are used for this analysis such as the DoubleMu, DoubleElectron, and MuEG datasets. The data taking period is split up into different runs, more details about these different runs and corresponding integrated luminosity for the aforementioned datasets can be found in Table 4.1 with their corresponding integrated luminosities. Following the CMS naming convention, the DoubleMu dataset is the collection of events which pass trigger criteria requiring at least two muons, while the DoubleElectron dataset contains events with at least two electrons. The MuEG dataset is comprised of events with at least one muon and one electron or photon. The H_{T}, a measure of jet activity, dataset is used for efficiency measurements as discussed in Section 6.3. In addition, we use the SingleMu dataset for muon identification and isolation efficiency measurements as discussed in Section 6.2. For each of the datasets we select events where specific trigger paths were satisfied. The list of trigger paths are listed in Appendix B. For example, the HLT_Mu17_Mu8 trigger path selects events with two muons, the first muon with $p_{\mathrm{T}} > 17$ GeV and the other muon with $p_{\mathrm{T}} > 8$ GeV. More stringent selections are applied at the analysis level. Identification and isolation requirements can also be applied at the trigger level.

During event reconstruction the data is monitored to select data with good detector status and data quality. CMS developed procedures for this Data Quality Monitoring (DQM) in order to provide quality flags that are later used in the analysis of the data. Collision data is split into different runs and luminosity sections (LS). Recorded events with erroneous noise in the hadron calorimeter are rejected, including as well beam-scraping events with large tracker occupancy leading to a high fraction of low quality tracks. The "good" runs and luminosity blocks are selected based on the CMS DQM and physics validation. Events reconstructed immediately after being recorded are referred to as "prompt reconstruction", while reprocessed data, including for example, updated detector alignment conditions, are called "re-reconstructed data". The trigger menu evolves throughout the runs. Therefore, trigger requirements for the data adjust to trigger conditions in order to adapt to the increasing instantaneous luminosity.

4.2 Simulated samples

Monte Carlo simulations are used in high energy particle physics to predict and to model processes that occur in particle collisions. Modeling expected background processes properly is especially important for any physics analysis searching for new physics. A numerical approach is needed because of the complex final states emerging from collision processes with a high multiplicity of particles, that cannot be calculated analytically. The Monte Carlo (MC) method is a numerical integration approach based on pseudo random number generators. With the increasing number of extra partons in the final state the dimension of the phase space becomes too large, beyond the level of simple numerical integration. Even

Primary Dataset	**Reco details**	**Luminosity** (fb^{-1})
MuEG	Run2012A-recover-06Aug2012-v1	0.082
MuEG	Run2012A-13Jul2012-v1	0.809
MuEG	Run2012B-13Jul2012-v1	4.403
MuEG	Run2012C-24Aug2012-v1	0.495
MuEG	Run2012C-PromptReco-v2	6.584
MuEG	Run2012D-PromptReco-v1	7.718
DoubleMu	Run2012A-recover-06Aug2012-v1	0.082
DoubleMu	Run2012A-13Jul2012-v1	0.809
DoubleMu	Run2012B-13Jul2012-v4	4.403
DoubleMu	Run2012C-24Aug2012-v1	0.495
DoubleMu	Run2012C-PromptReco-v2	6.557
DoubleMu	Run2012D-PromptReco-v1	7.719
DoubleElectron	Run2012A-recover-06Aug2012-v1	0.082
DoubleElectron	Run2012A-13Jul2012-v1	0.809
DoubleElectron	Run2012B-13Jul2012-v1	4.403
DoubleElectron	Run2012C-24Aug2012-v1	0.495
DoubleElectron	Run2012C-PromptReco-v2	6.575
DoubleElectron	Run2012D-PromptReco-v1	7.727

Table 4.1: Datasets used in this analysis.

to calculate a parton level cross section with many partons will include integration over all intermediate states and their final state decay modes, with spin, color state and parton density function (PDF).

The generation of simulated collision events begins with the hard process. For example, the production of a $t\bar{t}$ pair from gluon fusion. The matrix element (ME) of the hard process is calculated first with the momenta of the ingoing particles being randomly chosen, based on input PDFs, and the momentum of the outgoing particle being randomly distributed in the available kinematic phase space. After the hard interaction is generated, higher order QCD effects are incorporated using parton shower (PS) models. Partons can emit gluons called both initial-state radiation (ISR) before and final-state radiation (FSR) after the hard interaction. Next follows the hadronization process of the partons at an energy scale where perturbative QCD does not hold. Many hadronization models exist and depend on the generator used.

MadGraph [48, 49] generates tree-level events with extra partons in the final state on the basis of ME calculations. MadGraph is interfaced with the PYTHIA [50] event generator to simulate PS and hadronization. MadGraph is better suited to describe the hard interaction, while PYTHIA can handle the hadronization process and low energy physics. After the PS and hadronization are performed, jets are formed based on the Particle-Flow event reconstruction algorithm [51] and matched to the original partons generated at the matrix element stage in order to avoid double counting of the partons produced in the final state when interfacing with the ME and PS generators. This is accomplished at the MadGraph level by requiring a minimum p_{T} between patrons at the matrix-element level and a matching scale parameter Q_{cut}, the maximum distance between a jet and a parton to be matched with each other. Furthermore, this determines the scale of the transition between the perturbative and non-perturbative regime, mainly between the ME and PS generator transition region.

The main SM backgrounds for this analysis that includes $\mathrm{t\bar{t}}$ pairs, and VV (double vector) boson + jets processes are generated using MadGraph with PDF CTEQ6L11 [52]. In the MC samples multiple proton-proton interactions in the same or adjacent bunch crossings, known as pileup, are simulated using PYTHIA and superimposed on the hard collision. Simulated events are reweighted to match the pileup distributions in data. Pileup does not form part of the hard interaction, but can contribute additional low p_{T} objects to the event.

Simulated events are scaled such that the number of events corresponds to the expected yield based on the integrated luminosity and the cross section of the process. These expected yields are determined by multiplying the number of MC events passing certain selection criteria by a factor f_{MC},

$$f_{MC} = \frac{\mathcal{L} \cdot \sigma}{N_{simulated}}, \tag{4.1}$$

where $\mathcal{L}$ is the integrated luminosity of the corresponding dataset, σ is the cross section of the simulated process, and $N_{\text{simulated}}$ is the number of simulated events for the respective process. All SM processes are normalized to cross section calculations at next-to-leading order (NLO) or next-to-next-to-leading order (NNLO) when available and otherwise to leading order (LO). All simulated events are passed through a Geant4-based model [53] of the CMS detector. In order to improve both the signal and backgrounds modeling, lepton reconstruction and identification efficiencies, trigger efficiencies, jet energy scales, b-jet scale factors, and resolutions in the MC simulation are corrected according to values measured in data.

4.2.1 Signal samples

For the signal samples in this analysis, the CMS detector response is simulated using the full simulation package [53]. The simulated events are reconstructed and analyzed with the same software used to process collision data. MC simulations of signal and SM processes are used to tune the analysis, to estimate some of the backgrounds, and to calculate the signal acceptance in the search regions.

For the signal events, specified by the vector-like-quark signal model, as explained in the theory Section 2.2, parameters are generated according to the fourth generation Les Houches accord standards [54]. The production of vector-like-quark pairs is modeled with MadGraph 5.1.5.4, including up to two additional partons at the matrix-element level. The pair produced b' quarks are then decayed and hadronized using PYTHIA6. The b' quark can potentially decay into three different states bZ, tW, and bH, as previously mentioned. The cross sections (see Table 4.2) for the signal MC samples are calculated at full NNLO+NNLL resummation using the Top++ software package [55]. Samples were produced for b' masses of 300-1000 GeV in 50 GeV increments. We use simulated events for the purpose of estimating signal acceptance and to determine the expected signal yields in each of the different search channels.

Quark mass (GeV)	**cross section** (pb)	**Scale errors** (%)	**PDF errors** (%)
300	12.90	+2.33 −2.68	+3.36 −3.30
350	5.297	+2.28 −2.50	+3.57 −3.53
400	2.386	+2.24 −2.34	+3.73 −3.68
450	1.153	+2.21 −2.22	+3.88 −3.81
500	0.590	+2.18 −2.11	+4.04 −3.94
550	0.315	+2.15 −2.02	+4.21 −4.06
600	0.174	+2.13 −1.94	+4.42 −4.20
650	0.0999	+2.11 −1.88	+4.67 −4.34
700	0.0585	+2.06 −1.82	+4.97 −4.50
725	0.0452	+2.09 −1.79	+5.14 −4.59
750	0.0350	+2.06 −1.77	+5.31 −4.69
775	0.0273	+2.06 −1.75	+5.50 −4.80
800	0.0213	+2.03 −1.75	+5.70 −4.92
850	0.0132	+1.99 −1.71	+6.11 −5.21
900	0.00828	+2.03 −1.71	+6.57 −5.54
950	0.00525	+1.97 −1.68	+7.09 −5.89
1000	0.00336	+1.92 −1.72	+7.66 −6.32

Table 4.2: Approximate NNLO b' pair production cross sections computed with HATHOR [56].

4.2.2 Background samples

Simulated events are used to determine background contributions that can not be estimated using data based methods. These SM samples are produced with either PYTHIA, MadGraph v1.4.4, or the powheg [57] MC event generator with CTEQL16.6 or CTEQ6M PDF [58]. We normalize the SM processes to the their cross sections calculated at NLO or NNLO when available [59, 60, 61, 62, 63, 64, 65, 66] and to LO otherwise.

The simulated samples used for this analysis are listed in Table 4.3. These samples are used to compare with data and to estimate background contributions. The first column gives the MC sample name, the second column corresponds to the number of events generated, and the third column represents the theoretical cross section, used to determine the expected number of events of a given sample in the data.

MC sample (AODSIM)	**N events**	**cross section** (pb)
DYJetsToLL_M-10To50filter_8TeV-madgraph	7,131,530	11050.0
DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball	30,459,503	3532.8
TTJets_FullLeptMGDecays_8TeV-madgraph	12,119,013	26.5895
TTJets_SemiLeptMGDecays_8TeV-madgraph	25,423,514	108.51
TTGJets_8TeV-madgraph	71,598	2.166
TTWJets_8TeV-madgraph	196,046	0.2057
TTZJets_8TeV-madgraph_v2	209,677	0.232
TTWWJets_8TeV-madgraph	217,213	0.002
TBZToLL_4F_TuneZ2star_8TeV-madgraph-tauola	148504	0.0114
ZZZNoGstarJets_8TeV-madgraph	224,902	0.0192
WWWJets_8TeV-madgraph	220,170	0.08217
ZZJetsTo4L_TuneZ2star_8TeV-madgraph-tauola	4,804,781	0.1769
WZJetsTo3LNu_TuneZ2_8TeV-madgraph-tauola	2,016,678	1.0575
WWJetsTo2L2Nu_TuneZ2star_8TeV-madgraph-tauola	1,932,249	5.8123
WJetsToLNu_TuneZ2Star_8TeV-madgraph-tarball	18,393,090	37509
WWGJets_8TeV-madgraph	215,121	1.44
WWZNoGstarJets_8TeV-madgraph	222,234	0.0633
GluGluToHToTauTau_M-125_8TeV-powheg-pythia6	967566	1.2466
GluGluToHToWWTo2LAndTau2Nu_M-125_8TeV-powheg-pythia6	299975	0.4437
GluGluToHToZZTo4L_M-125_8TeV-powheg-pythia6	995117	0.0053
VBF_HToTauTau_M-125_8TeV-powheg-pythia6	1000000	0.0992
VBF_HToWWTo2LAndTau2Nu_M-125_8TeV-powheg-pythia6	299687	0.0282
VBF_HToZZTo4L_M-125_8TeV-powheg-pythia6	49876	0.000423
WH_ZH_TTH_HToTauTau_M-125_8TeV-pythia6-tauola	200000	0.0778
WH_ZH_TTH_HToWW_M-125_8TeV-pythia6	200408	0.254

Table 4.3: Monte-Carlo samples used for background estimations. The Summer12_DR53X-PU_S10_START53_V7A or tags correspond to the pileup scenario and the alignment and calibration conditions centrally determined by CMS. TTJets fully-leptonic used version v2, TTJets semi-leptonic uses ext-v1 and all other MC used v2. TBZToLL used pileup scenario Summer12_DR53XPU_S10_START53_V7Cv1.

Chapter 5

Event reconstruction and selection

In this chapter, we describe the algorithms used by CMS to reconstruct objects such as leptons, tracks, vertices, and jets, that are used in the search for new physics.

5.1 Particle-Flow reconstruction algorithm

The CMS Particle-Flow (PF) algorithm [51] performs the reconstruction and identification of all stable particles such as electrons, muons, photons, and charged and neutral hadrons in an event by using the full information from all available sub-detectors. As particles transverse the detector they leave signals behind, where charged particles produce tracks, neutral and charged particles deposit energy in calorimeters, or in the case of muons, leave hits in the muon chambers. Particle candidates in an event are reconstructed based on tracks and calorimeter clusters that are linked together using a linking algorithm, resulting in a list of PF candidates found in the event. For example, in muon reconstruction, tracks reconstructed in the inner tracker system are matched to tracks reconstructed with the muon system and are combined if the global track fit has an acceptable χ^2 (a measure of goodness-of-fit) to provide final muon candidates. Charged hadrons are reconstructed and identified as tracks in the inner tracker and linked to the HCAL if the particle p_{T} is sufficiently large to reach the calorimeter.

5.2 Tracks from charged particles

In the uniform 4 Tesla field of the solenoidal magnet, charged particles follow a helical path parallel to the direction of the magnetic field. These particles leave energy deposits in the tracker detector sensors along their trajectories that are reconstructed as hits in the detector volume. Tracks are used in CMS in the reconstruction of electrons, muons, taus, hadrons, jets, and in the determination of the primary interaction vertices. Track reconstruction is performed by the Combinatorial Tracker Finder (CFT) algorithm [67]. Seeding and p_{T} requirements are changed for each iteration of the algorithm. The initial track estimate including its uncertainty is called the "seed" and is based on a triplet of hits in the tracker system or on pairs of hits combined with an additional constraint from the beamspot. The seed tracks are propagated outward by a Kalman filter [51] algorithm that searches for compatible hits based on predicted trajectories. The Kalman filter relies on information of the current sate of the trajectory, its uncertainty, statistical noise, and the underlying physics process of the particle interaction. With each iteration, hits are associated to tracks and removed from the collections of tracker hits, resulting in a smaller hit collection to be used in the subsequent iteration. The reconstructed tracks are filtered after each iteration to remove tracks that are likely to be fake tracks.

The CTF algorithm performs six iterations as follows:

- Iteration 0: Looks at tracks with $p_{\mathrm{T}} > 0.9$ GeV originating close to the interaction point and that have at least three hits in the pixel detector.
- Iteration 1-2: Tries to find tracks with exactly two hits in the pixel detector or with lower p_{T} then the previous iterations.
- Iteration 3-5: Tries to reconstruct non-prompt tracks originating further away from the primary interaction point.

Tracks that satisfy selection criteria after each iteration, such as the χ^2 per degree of freedom of fitted tracks, their distance from the primary vertex, and the number of layers that have hits, are labelled as “high purity”.

5.3 Vertex reconstruction

For the purpose of this analysis, we are interested in particles coming from the hard collision of two protons and therefore it is important to identify the primary vertex. The reconstruction of vertices in the event becomes increasingly difficult to handle as effects of multiple interactions of protons in the same bunch crossing resulting in pileup increases with higher center-of-mass energy and instantaneous luminosity. The vertex reconstruction starts from prompt tracks, selected based on the quality criteria mentioned in the previous section. Then these selected tracks are clustered using the Deterministic Annealing (DA) clustering algorithm [68] in the z-direction.

For every track, the closest point of approach in the z-coordinate to the beamline, is denoted by z_i, with an associated uncertainty σ_i. The cluster algorithm proceeds by finding a possible set of vertex candidates, denoted by z_j, which are assigned to a track. The χ^2 quantity can be used as a measure of performance, which is defined as,

$$\chi^2 = \sum_{ij} p_{ij} \cdot \frac{(z_i - z_j)^2}{\sigma_i^2}, \tag{5.1}$$

where p_{ij} is interpreted as a probability.

Instead of calculating the p_{ij} and z_j pair that minimizes the total χ^2 directly, the DA algorithm determines the most likely distribution for p_{ij}, given $< \chi^2 >$, and proceeds to decrease $< \chi^2 >$ until it finds a local minimum [68]. After tracks are assigned to various vertices, a vertex fit is performed with the Adaptive Vertex Fitter (AVF) [69], where each track receives a weight depending on its χ^2 contribution to the vertex. This weight w_i ranges between 0 (outliers) and 1 (good tracks), based on compatibility with the common vertex. Later, the weights are summed, giving effectively the total number of accepted tracks by the AFV. Each vertex, based on the sum of track weights, is assigned a number of degrees of freedom, $n_{\mathrm{dof}} = 2 \cdot \Sigma w_i - 3$ [68]. Reconstructed vertices are sorted according to the $\sum p_{\mathrm{T}}^2$ of the tracks found in the track cluster, where the vertex corresponding to the highest sum is considered the primary vertex.

5.4 Jets

Quarks and gluons cannot be observed as isolated particles, rather the hadronization process turns individual quarks and gluons into many hadrons. Experimentally, a “jet” is defined a cluster of reconstructed objects such as tracks, calorimeter towers, and particle candidates. Jets are observed as collimated sprays of particles originating from partons of the underlying hard interaction processes. Charged hadrons produce tracks in the pixel and silicon strip tracker and deposit energy in the electromagnetic and hadronic calorimeters. In the CMS collaboration, several jet algorithms have been developed. Jet reconstruction can be based either on standalone calorimeter information, or tracking and calorimeter information, so-called “jets-plus-tracks”, or PF candidates which takes advantage of information from all sub-detectors with improved jet energy resolution and smaller uncertainties.

5.4.1 Jet algorithm

In this section, we briefly describe the jet algorithm relevant to this particular analysis. In order to handle processes involving quarks and gluons in the final sate, we need to establish a jet algorithm that clusters particle together to form a jet.

Sequential recombination algorithms

Recombination algorithms are based on hierarchical clustering. Typically these types of algorithms work by calculating a ‘distance’ between particles, referred to as jet candidates, and then recombining them pairwise according to a given prescription, eventually yielding the final set of jets. Sequential algorithms are specified by their recombination metrics, relying on the clustering of jet candidates based on a minimum measured distance between two objects (d_{ij}) as well as between a candidate jet and the beamline (d_{iB}) [70],

$$d_{ij} \equiv min(k_{Ti}^{2p}, k_{Tj}^{2p}) \cdot \frac{\Delta R_{ij}^2}{R^2}, \tag{5.2}$$

$$d_{iB} \equiv k_{Ti}^{2p} \tag{5.3}$$

here k_{Ti} is the transverse moment of particles or jet candidates and $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_j - \phi_j)^2$, where y is the rapidity and ϕ is the azimuthal angle around the beam direction. The value of p specifies the type of jet clustering algorithm. For p = 1, the algorithm is called the k_t algorithm [71], p = 0 results in the Cambridge-Aachen algorithm [72], and p = -1 is called the anti-k_t algorithm [73]. These jet algorithms require the specification of a cutoff angular parameter R, which controls the size of the jets and discerns when jet candidates should be promoted to a jet. In this analysis we look at jets clustered with the anti-k_t algorithm ($p = -1$) and a distance parameter $R = 0.5$.

For each pair of particles in the event, the minimal d_{ij} and d_{iB} are determined. The anti-k_t algorithm tends to clusters particles out to a distance R from the center of a jet beginning with the hardest jet candidates first, resulting in a more regular jet shape, while a softer jet exhibit more of an irregular shape [73]. Therefore, anti-k_t algorithm jets are easier to calibrate.

The recombination algorithm proceeds as follows:

- Compute d_{ij} and d_{iB} for every particle in the final state, and find the one with the minimum value.
- If the minimum is d_{iB}, then declare particle i a jet, remove it from the list of jet candidates, and return to the previous step.
- If the minimum is d_{ij}, combine particles i and j, and return to the first step.

- Iterate until all the particles in the event have been classified as jets.

Two important features of these sequential algorithms are that they are manifestly infrared (IR) and collinear safe to ensure that the final set of jets are unchanged when soft emission of radiation or collinear splitting of hard particles occurs. In general, these jet algorithms should remain stable with respect to higher-order perturbative QCD corrections and not change the results of the jet clustering.

5.5 b-jet identification

The identification of jets originating from light- or heavy-flavor quarks is an important tool to select events which contain b quarks in the final state while suppressing non-b quark backgrounds. During the hadronization process of b quarks, B mesons are produced. These hadrons travel typically a distance of a few millimeters in the detector before decaying. The charged particles from this decay are reconstructed as tracks with large impact parameters that can be used to identify the position of the B meson decay, leading to a so-called "secondary vertex". This can then be used as a discriminating variable in order to "tag" a b-jet. The so-called "Combined Secondary Vertex" (CSV) algorithm is used for this analysis, described in further detail in Ref. [74]. The CVS tagger algorithm uses a neural network approach combing information from track based lifetime and secondary vertex reconstruction.

Working points are classified based on the average mistag rate for light-flavor jets, that is defined as the probability of misidentifying a jet as a b jet. For this analysis we chose the "medium" working point defined by an average mistag rate of 1% and a b-tagging efficiency of approximately 70%. Figure 5.1 shows how the efficiency in simulation and data do not match. Therefore, differences of the b-tagging efficiency and mistag rate between simulated events and data are corrected by reweighting the simulated events.

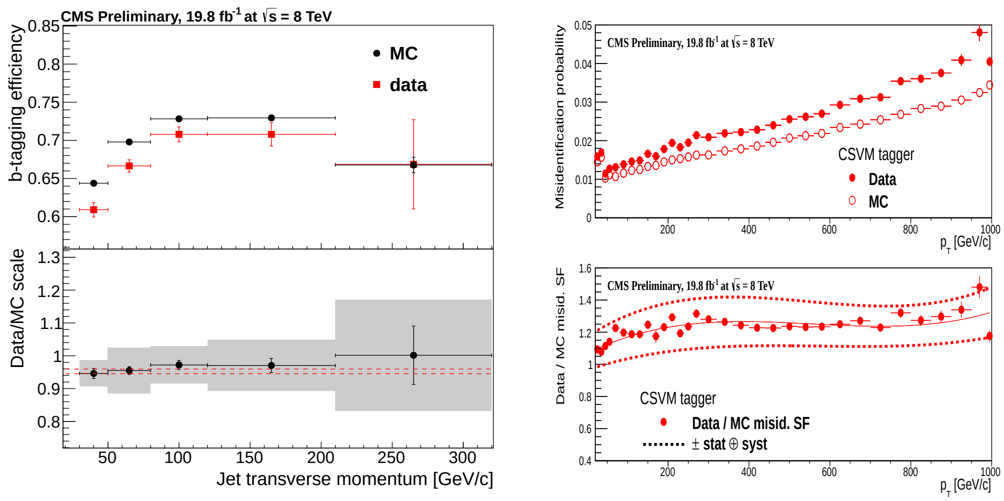


Figure 5.1: Identification efficiency (left) or b jets with CSV tagger and medium operation point for data and simulation. The misidentification rate (right) for the CSVM tagger [75].

5.6 Leptons

In this section, we will discuss other important objects used in this analysis such as muons, electrons, and taus. The focus here lies on the reconstruction algorithms. More details about identification criteria applied in the analysis are given in Chapter 6.

5.6.1 Electrons

Electrons are the lightest charged lepton and while transversing the dense material of the ECAL deposit a large fraction of their energy in the calorimeter. In CMS, electrons are reconstructed using two sub-detectors the inner tracker, which reconstructs electron tracks, and the ECAL, where electrons deposits the majority of their energy. The major challenge in electron reconstruction is due to the fact that electrons tend to emit large portions of their energy from Bremsstrahlung radiation in the tracker material, which significantly alter their trajectory. Therefore, both track and energy reconstructions algorithms must account for these deviations. Electrons deposit their energy in the ECAL in a narrow region in η. In order to reconstruct as much of the electron's energy as possible the reconstruction algorithm makes use of a "supercluster" (SC) pattern, where a supercluster is a collection of one or more associated clusters of energy deposited in the ECAL system. The aim of the SC is to recover energy lost due to Bremsstrahlung photons and conversion pairs. The tracker-driven reconstruction is better suited for low p_{T} electron, while the ECAL-driven algorithm is optimized for isolated electrons in the p_{T} range for Z and W decay.

5.6.2 Muons

Muons produced in the *pp* collisions are measured by reconstructing the muon's trajectories. A more detailed description on muon reconstruction in CMS can be found in Ref. [76]. Muons are minimum ionizing particles (MIP) particles which leave hits in the outer part of the detector (i.e. muon system), and signatures in the tracker system.

In general, muon reconstruction is performed in three stages: Tracker reconstruction, Standalone reconstruction, and Global reconstruction, which are described below.

- *Tracker reconstruction*: The starting point for the reconstruction of tracker muons are tracks in the silicon tracker system. These tracks are matched with the calorimeter and muon system for compatible signatures.

- *Standalone reconstruction*: A Kalman-filter [77] technique is applied to reconstruct the track trajectory, using standalone information from the muon system (DT, RPC, and CSC). After reconstruction the track is linked to the interaction vertex.

- *Global muons reconstruction*: Track positions from the standalone reconstruction are combined with information from the inner tracker. If these are compatible a global muon track is obtained using the hit information from both the muon system and silicon tracker.

Muons candidates are selected among the reconstructed muon track candidates by applying minimal requirements on the track segments in the muon and inner tracker systems as well as taking into account small energy deposits in the calorimeter.

5.6.3 Taus

Tau leptons are primarily identified via their decay products. In 35% of the cases, taus decay semi-leptonically to a neutrino and lepton (electron or muon) and 65% of the time they decay hadronically typically into one or three charged mesons such as π^+ and π^-, which are often accompanied by neutral pions, π^0. Experimentally taus produce signatures of narrow jets with low particle multiplicity. The CMS collaboration developed several algorithms for identifying hadronic taus [78], based on the various tau decay modes, through the reconstruction of daughter pions.

Hadronic Plus Strips algorithm

This section will briefly discuss hadronic tau reconstruction using the Hadronic Plus Strip (HPS) [78] algorithm, which is based on PF jets. The magnetic field of the CMS solenoid bends electrons and positrons, originating from photon conversions from π^0 decay, broadening the calorimeter energy deposit in the azimuthal direction. The algorithm takes this into account by reconstructing photons in so-called "strips", which are objects built out of EM particles, PF photons and electrons. These strips allow for greater coverage in the η and ϕ direction. The strip reconstruction begins by centering a strip in the core of the most energetic EM particle inside the PF jet. Next, the algorithm searches for EM particles within the region $\Delta\eta = 0.05$ and $\Delta\phi = 0.20$ around the strip center. This is repeated until no other particle can be associated with the strip. The strips satisfying a minimum p_{T} requirement ($p_{\mathrm{T}}^{\mathrm{strip}} > 1\mathrm{GeV}$) are combined with the charged hadron in order to reconstruct individual hadronic tau decay modes.

Decay topologies considered by the HPS tau identification algorithm are:

1. *Single hadron*: corresponds to the decay modes $\pi^-\nu_\tau$ and $\pi^-\nu_\tau\pi^0$, where the neutral pions have too low of an energy to be reconstructed as strips.

2. 1 *hadron* + 1 *strip*: corresponds to the decay mode $\pi^-\nu_\tau\pi^0$ in events where the

photons coming from a π^0 decay are very close together in the calorimeter.

3. 1 *hadron* + 2 *strips*: corresponds to the decay mode $\pi^-\nu_\tau\pi^0$ in events where the photons originating from π^0 decay are well separated.

4. 3 *hadrons*: corresponds to the decay mode $\pi^-\pi^+\pi^-\nu_\tau$, where the three hadrons are required to come from the same secondary vertex in the event.

The hadronic tau constituents are required to be isolated, with cone of size $\Delta R = 0.5$ around the direction of the tau candidate, from other particles in the event. Therefore, three working points (WP) are defined as either "loose", "medium", and "tight", by adjusting the p_{T} thresholds for particles considered in the isolation cone. For example, 'a 'loose' WP has a probability of about 1% for jets to be misidentified as a hadronic tau.

5.7 Photons

Photons, similar to electrons, are reconstructed from superclusters and their momentum is assigned based on the position of the supercluster as well as that of the primary reconstructed vertex [79]. Calorimeter energy expected for charged pions with momenta given by reconstructed track is subtracted from the cluster. The remaining clusters without a linked track are classified as photons. Selection requirements are applied to distinguish photons from electrons, such as requiring the photons not match any pixel hits, consistent with tracks from the interaction region. Isolation criteria are imposed, that use different isolation variables calculated based on HCAL and ECAL energy deposits, including a $\sum p_{\mathrm{T}}^{\mathrm{Track}}$ in a $\Delta R = 0.4$ cone around the photon candidate.

5.8 Missing transverse energy

Missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$, corresponds to the imbalance of the measured particle momentum in the plane transverse to the beam direction. From momentum conservation, the vector sum of all particles in the collision event must be equal to the initial patron's momentum. Particles such as the neutrino escape detection producing an imbalance of the transverse momentum.

$E_{\mathrm{T}}^{\mathrm{miss}}$ is calculated using the PF algorithm and is estimated in terms of this momentum imbalance. The missing transverse momentum vector is defined as the negative vector sum of all the visible PF candidates in the event and given by,

$$\vec{\not{E}}_{\mathrm{T}} = -\sum_{i} \vec{p}_{\mathrm{T,i}}, \tag{5.4}$$

where the magnitude of this vector is the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$.

Chapter 6

Particle identification and efficiency

The previous chapter presented general reconstruction algorithms used by the CMS collaboration in order to identify leptons, $E_{\mathrm{T}}^{\mathrm{miss}}$, jets, and photons. In this chapter we will discuss selection criteria used to reduce misidentified objects and provide an overview of the final objects used in this analysis.

6.1 Object selection

This analysis requires the presence of at last three reconstructed lepton candidates. The allowed candidates include electrons, muons, and hadronically-decaying taus; taus decaying leptonically are included in the electron and muon categories. The matching candidate tracks must satisfy quality requirements and spatially match with the energy deposits in the electromagnetic calorimeter or the tracks in the muon detectors, as appropriate. Details of the reconstruction and identification can be found in [80] for muons and in [81] for electrons and have been summarized in Section 5.6. The selections requirements applied to these objects are designed to maximize the number of good reconstructed objects, while minimizing the background coming from fake objects.

6.1.1 Electrons

Electrons are reconstructed from the Particle-Flow algorithm by a combination of track and energy deposit in the ECAL system. After the isolation requirement on the leptons is imposed, the most significant background sources are residual non-prompt leptons from heavy quark decays. These leptons have a higher probability of being isolated because of their larger momentum with respect to the jet axis, and can be misidentified as prompt leptons. This background is reduced by requiring that the leptons originate from within 0.5 cm of the primary vertex in the z-direction, denoted by d_z, and that the impact parameter, denoted by d_0, between the track and the event vertex in the plane transverse to the beam axis be small, $|d_0| \leq 0.02$ cm. This ensures that electrons are consistent with being produced directly at the primary vertex. We require electrons to have $p_\mathrm{T} > 10\,\mathrm{GeV}$ and $|\eta| < 2.4$. Selection requirements based on several types of variables are discussed below which ensure the selection of quality objects.

H/E: Electrons deposit most of their energy in the ECAL and only a small fraction of it in the HCAL. Therefore, the ratio of energy deposited by the electron in the HCAL to ECAL is small. The non-zero value is due to the tail of the electron shower that punches-through to the HCAL. This quantity can be used to distinguish electrons from pions as the pions have a greater HCAL energy deposit fraction.

$\sigma_{i\eta i\eta}$: Measures the electromagnetic clustering width in the ECAL, where iη corresponds to the ith detector element, in the η-direction, as discussed in section 3.2.1.

$\Delta\eta_{In}$ and $\Delta\phi_{In}$: These quantities are used to match the electron to the energy deposit in the ECAL in both η and ϕ, respectively. This ensures the track matches an electron rather then to a charged pions.

$1/E - 1/p$: This quantity for the electron differs from pions in that they are nearly equal to each other. Therefore, imposing a requirement on this value to be less than 5% removes hadronic tracks.

$Relative\ Isolation$: We measure the relative isolation of the electron from other activity in the event, denoted as I_{rel}. Relative isolation is defined as the ratio of the scalar sum of the transverse track momenta and the transverse calorimeter energy deposits within a $\Delta R < 0.3$ cone around the electron candidate direction at the origin, to the transverse momentum of the electron candidate. This quantity allows background contributions with softer p_T spectrum than the electron to be effectively suppress.

A complete list of selection criteria and their cut values are shown in Table 6.1 and additional details of the electron reconstruction at CMS is found in [82].

Observable	**Value or Range** Barrel	Endcap
$\|\eta\|$	< 2.4	< 2.4
$\Delta\eta_{in}$	< 0.007	< 0.009
$\Delta\phi_{in}$	< 0.15	< 0.10
$\sigma_{i\eta i\eta}$	< 0.01	< 0.03
H/E	< 0.12	< 0.10
$d_0(vertex)$	< 0.02	< 0.02
$d_Z(vertex)$	< 0.1	< 0.2
$\|1/E - 1/p\|$	< 0.05	< 0.05
Relative PF isolation ($RelIso$)	< 0.15	< 0.15
Conversion rejection cut	0	0
Number of expected inner hits	< 2	< 2
ΔR to nearest selected muon	> 0.1	> 0.1

Table 6.1: Electron selection criteria, several electron ID requirements are different for the barrel ($|\eta| < 1.44$) and endcap ($1.56 < |\eta| < 2.4$) regions [85].

Further selection criteria are applied to all electrons:

- Electrons in the barrel-endcap transition region $1.44 < |\eta| < 1.57$ are rejected.
- An electron should not be within ΔR of 0.1 of a selected muon.
- d_0 and d_z are calculated with respect to the first good vertex.
- Effective area, A_{eff}, corrections for electrons.
  - A_{eff} are taken from official CMS numbers. This is the geometric area of the isolation cone scaled by a factor which accounts for the residual dependence of the average pileup on the electron.

6.1.2 Muons

Muons as previously discussed in section 5.6.2 are reconstructed in three stages. The first is the 'tracker muon' reconstruction based on tracker information alone. The second is the 'stand alone muon' reconstruction based on the information from the muon chambers only. In the last stage, the combined information from the tracker and muon chambers are linked together to get a combined fit for the so-called 'global muon'. We select muons with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and $|\eta| < 2.4$. Several other selection criteria described below must be satisfied by the muons to be selected for the purpose of the analysis.

Particle-Flow muon: Particles identified by the PF event reconstruction as PF muons are accepted in to the event selection. The matching particle candidates must satisfy quality requirements and spatial matching with the energy deposit in the ECAL and the tracks in the muon detector.

Impact Parameter (IP): Muons are required to originate from within 0.5 cm of the primer vertex in the z-direction. The impact parameter d_o between the track and event vertex in the plane transverse to the beam axis must be less than 0.02 cm, which reduces the background contribution from muons originating from jets and from pileup vertices.

Global muon fit: Global muon are reconstructed from the muon trajectory based on hits in the silicon tracker and the muon chambers [83]. This information serves as an input for a fit of the muon trajectory whose quality is determined in terms of a χ^2 per degree of freedom (ndof). Good quality muon tracks are required to have $\chi^2/ndof < 10$.

Number of hits in the muon chambers: Fitted global-muon tracks are required to have hits in the muon chambers, which ensures muons do not come from hadronic punch-throughs or in flight decays.

Number of hits in the tracker: We require a minimum of five hits in the tracker along with hits in the muon chamber, which ensures a good p_{T} measurement.

Relative Isolation : Semi-leptonic decay of hadrons containing heavy flavor quarks, can give rise to muons produced during the hadronization process of a b quark. The muons are constituents of the jet and should not be considered as "good" muons for this analysis. In order to reject these muons we compute an isolation variable based on the p_{T} of any particles surrounding the muon candidate with a $\Delta R = 0.3$ cone. This isolation divided by the p_{T} of the muon has to be less than 0.15.

A comprehensive list on all analysis level muon selection requirements along with their values are found in Table 6.2 and details about CMS muon reconstruction can be found in [80].

Observable	**Value or Range**
$\lvert\eta\rvert$	< 2.4
Global muon	1
PF muon	1
χ^2/d.o.f.	< 10
$\lvert dz\rvert$from vertex	< 0.5 cm
$\lvert d0\rvert$ from vertex	< 0.02 cm
Number of valid pixel hits	> 0
Number of tracker LayersWM*	> 5
Number of valid hits in muon chamber	> 0
Number of muon stations with muon segments	> 1
Relative isolation within $\Delta R < 0.3$, with β corrections for PU	< 0.15

Table 6.2: Muon selection criteria [85].

6.1.3 Taus

We consider hadronic tau decays that yield either a single charged track (one-prong) or three charged tracks (three-prong) with or without additional electromagnetic energy from neutral pion decays. The hadronic tau candidates are reconstructed using the Hadron Plus Strips (HPS) PF algorithm [84], that considers the various hadronic decay modes and rejects candidates that appear to be poorly reconstructed electrons or muons. We require tau candidates to have $p_{\mathrm{T}} \geq 20$ GeV, $|\eta| < 2.5$ and $\Delta R > 0.1$ from selected leptons.

The selection criteria for hadronic taus are as described below.

Hadronic Plus Strip (*HPS*) *taus*: Hadronically decaying taus are as reconstructed by the HPS algorithm [78], that uses charged hadrons and photons to construct the main decay modes of the hadronic tau (1 changed hadron, 1 charged hadron + photons, and 3 charged hadrons).

ByDecayModelFinding: Is a discriminant boolean variable calculated by the HPS algorithm, that can either be 1 if the algorithm is able to reconstruct a valid hadronic decay of the tau and 0 otherwise.

AgainstElectronMVA: Discriminant used to rejects tau candidates which are already selected as electron candidates by the PF algorithm.

AgainstMuonTight: Discriminant used to reject fake hadronic taus from muons.

ByLooseCombinedIsolationDBSumPtCorr: Discriminator calculates $\sum p_{\mathrm{T}} \geq 20$ GeV of charged and neutral candidates with $p_{\mathrm{T}} > 0.5$ GeV and $\Delta R < 0.5$, where it is a assigned a value of 1 if the isolation is less than 2 GeV.

6.1.4 Photons

We consider photons with $p_{\mathrm{T}} \geq 20$ GeV and $|\eta| < 2.5$. The cone size for all isolation variables and photon selection criteria used are listed in Table 6.3.

Observable	**Value or Range**	
	Barrel	Endcap
Conversion safe electron veto	1	1
Single tower H/E	< 0.06	< 0.05
$\sigma_{i\eta i\eta}$	< 0.011	< 0.034
ρ corrected relative PF charged hadron isolation	< 0.06	< 0.05
ρ corrected relative PF neutral hadron isolation	< 0.16	< 0.10
ρ corrected PF photon isolation	< 0.08	< 0.12

Table 6.3: Selection criteria for barrel and endcap photons [85].

6.1.5 Jets

Jets are reconstructed from PF objects using the anti-k_{T} algorithm [86] with a distance parameter of 0.5. Jet energy scale corrections obtained from data and MC simulation are applied to account for the nonlinear response of the calorimeter and pileup effects [87, 88], corresponding to multiple interactions in a bunch crossing coming from different proton-proton collisions.

The selection criteria for jets [85] are given by:

- PFJet: Apply L1FastL2L3 corrections to the MC simulation, and L1FastL2L3residual corrections to data. Jet energy corrections are based on MC simulation. Since the

MC does not properly reproduce real detector response, additional corrections have to be applied to correctly describe the data. These residual corrections are about 2% in the barrel and up to 10% for the endcap.

- Require all jets to have $p_{\mathrm{T}} > 30$ GeV.
- Neutral hadron fraction of total jet energy < 0.99, corresponding to the fraction of the total jet energy associated with hadronic energy deposits, not linked to tracks.
- Neutral EM fraction of total jet energy < 0.99, related to the fraction of the total jet energy associated with electromagnetic energy deposits, not linked to the tracks.
- Number of constituents in jet > 1.
- Apply additional requirements for jets with $|\eta| < 2.4$:
  - Charged hadron fraction > 0, corresponding to the fraction of the total jet energy associated with hadronic energy deposit, linked to tracks
  - Number of charged/neutral constituents > 1.
  - Charged EM fraction < 0.99, related to the fraction of total jet energy associated with the electromagnet energy deposits linked to tracks.

6.1.6 B-tagged jets

We categorize events according to the absence or presence of one or more b-jets. The CMS Combined Secondary Vertex algorithm [74] is used to identify jets that are consistent with having originated from the hadronization of b quarks. The working point is chosen such that we obtain a b-tagging efficiency of 70%, a c-tagging misidentification of $10 - 20\%$, and a misidentification rate for light flavor jets of 1%. Systematic uncertainties related to b-tag/mis-tag scale factors are determined by varying corresponding p_{T} dependent scale factors by $\pm 1\sigma$ of their uncertainty.

6.1.7 Missing transverse energy

The missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$, is defined as the magnitude of the transverse component of the vectorial sum of the momenta of all PF candidates in the event. Comparisons between data and simulation show good modeling of $E_{\mathrm{T}}^{\mathrm{miss}}$ for processes with genuine $E_{\mathrm{T}}^{\mathrm{miss}}$ from neutrinos [89, 90].

The selection criteria for $E_{\mathrm{T}}^{\mathrm{miss}}$ are given by:

- Select PF $E_{\mathrm{T}}^{\mathrm{miss}}$.
- Apply the CMS collaboration recommended list of filters:
  - *CSC tight beam halo filter*: Beam halo noise arises from detector induced secondary particles that result in showers from beam-gas collisions inside the vacuum chamber or charged particles deflected by the magnetic field of the beam-line optics. The CSC beam halo filter is used to identify events with large beam backgrounds.
  - *HBHE noise filter*: Designed to reject isolated noise from the HCAL barrel and HCAL endcap readout electronics, that can be incorrectly reconstructed as hadronic energy deposits [91].
  - *Primary vertex filter*: Require at least one "good" primary vertex to be reconstructed in each collision event. A well identified primary vertex has least four degrees of freedom and a position with in $|z| < 24$ cm and $r < 2$ cm in order to ensure good collision candidates and reject noisy events due to pileup.
  - *ECAL dead cell trigger primitive (TP) filter*: High energy particles depositing their energy in noisy crystals cells in the ECAL that are masked and not included in the event reconstruction, can lead to fake E_{T}^{miss}. These crystals still posses their TP information, which can be used to filter out events where high energy is measured in the masked cells by the TP system.

– *Tracking failure filter*: In the case where there are too many clusters in relation to the number of tracks, the tracking algorithm can fail. To reject these events, a cut on the $\sum p_{\mathrm{T}}$ of all tracks belonging to the primary vertex divided by the $\mathrm{H_T}$ of all jets in the events is applied.

– *Bad EE Supercrystal filter*: Rejects events with anomalously high energy in superclusters in the ECAL.

6.2 Lepton efficiency and scale factors

The previous section focused on the identification of different objects that are important for this analysis. In this section, we will discuss measurements of the efficiency of the lepton identification and isolation requirements. The efficiency criteria refers to the ability of a certain set of selection criteria to properly identify an object, in this case electrons, muons, and taus. A so-called "tag-and-probe" method based on $\mathrm{Z} \rightarrow \ell^{+}\ell^{-}$ events is used for the measurements. The tag-and-probe technique involves selecting tight leptons, referred to as the 'tag' object, that pass analysis level selection requirements and 'probe' objects that satisfy looser selection requirements. The efficiency of a certain selection under study corresponds to the fraction of events where the probe lepton satisfies these selection requirements. The efficiency is parametrized as a function of the p_{T} and η of the probe lepton.

A tag muon is required to pass the same analysis level selections as isolated muons, except that the p_{T} requirement is increased to $p_{\mathrm{T}} > 20$ GeV. Probe muons are required to be global muons with $p_{\mathrm{T}} > 5$ GeV and $|\eta| < 2.1$. In the case of probe electrons, the selection requirements are loosened on the values of $\sigma_{I\eta I\eta}$, $\Delta\phi$, $\Delta\eta$, and H/E from the analysis criteria. The identification and isolation efficiency for leptons are measured in data and in simulation and compared with each other. However, the MC simulated efficiencies do not necessarily match to the efficiency measured in data and to account for the differences a data-to-simulated "scale factor" must be applied. These scale factors can be parameterized based on certain variables, e.g. p_{T} and η, to be used to correct the MC modeling. The identification efficiency scale factor is the ratio between the efficiency of the identification selection when measured in data and when measured in MC simulation. We fit this ratio of the two efficiencies and use the parameterized function to correct the MC simulations for each lepton in an event based on its p_{T} and η.

The dimuon and dielectron invariant mass distributions are shown in Figures 6.1 and 6.2 for different probe lepton p_{T} ranges. The MC simulation is normalized to have the same number of events under the Z boson mass region ($80 - 100$ GeV) as in data. For the various p_{T} ranges, we count the number of events within the Z boson mass region before and after applying the selection requirements. The invariant mass distribution in the range of $55 - 125$ GeV is fitted with a polynomial function to remove the background contribution inside the Z mass region. We estimate the identification and isolation efficiency separately.

The lepton identification efficiencies correspond to the probe selection efficiency after the identification criteria for probes are applied. The lepton isolation efficiency is estimated by applying the isolation requirement to probe leptons which pass the identification selection, referred to as 'good probe' leptons. We define the the lepton isolation efficiency as,

$$\epsilon_{\text{Isolation Efficiency}} = \frac{\text{N(good probe leptons which satisfy isolation requirements)}}{\text{N(total number of good probe leptons)}} \tag{6.1}$$

The lepton identification efficiencies are shown in Figures 6.3 and 6.4 with their corresponding data to MC ratio as a function of the probe lepton p_{T} for both electrons and muons, respectively. The identification efficiency for muons is modeled to within a percent by the MC simulation. In the case of $p_{\mathrm{T}} < 25$ GeV, electron and muon isolation efficiency measurements in data are lower than that measured in MC. Figures 6.5 and 6.6 show both the isolation efficiency and the data to MC ratio as a function of probe p_{T} for muons and electrons, respectively. We apply efficiency scale factors to the MC simulation to account for the differences in lepton efficiency found in data. The scale factors are determined by

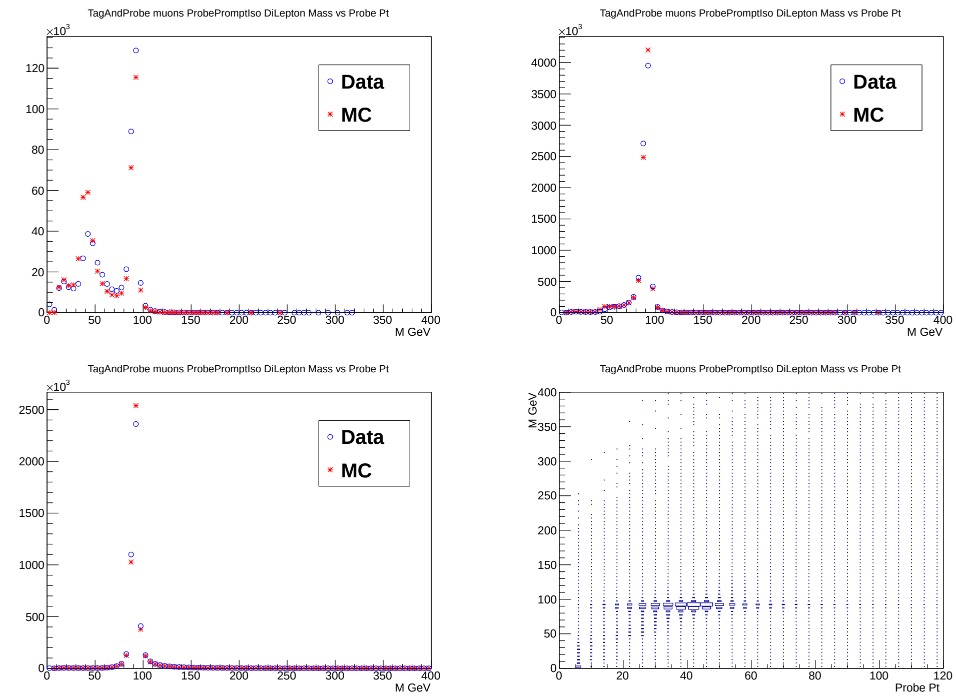


Figure 6.1: Dimuon invariant mass between a tag and probe muon. Shown is the mass for probe p_{T} from 12-24 GeV (top left), 24-48 GeV (top right), and > 48 GeV (bottom left). The mass versus probe p_{T} is shown bottom right [92].

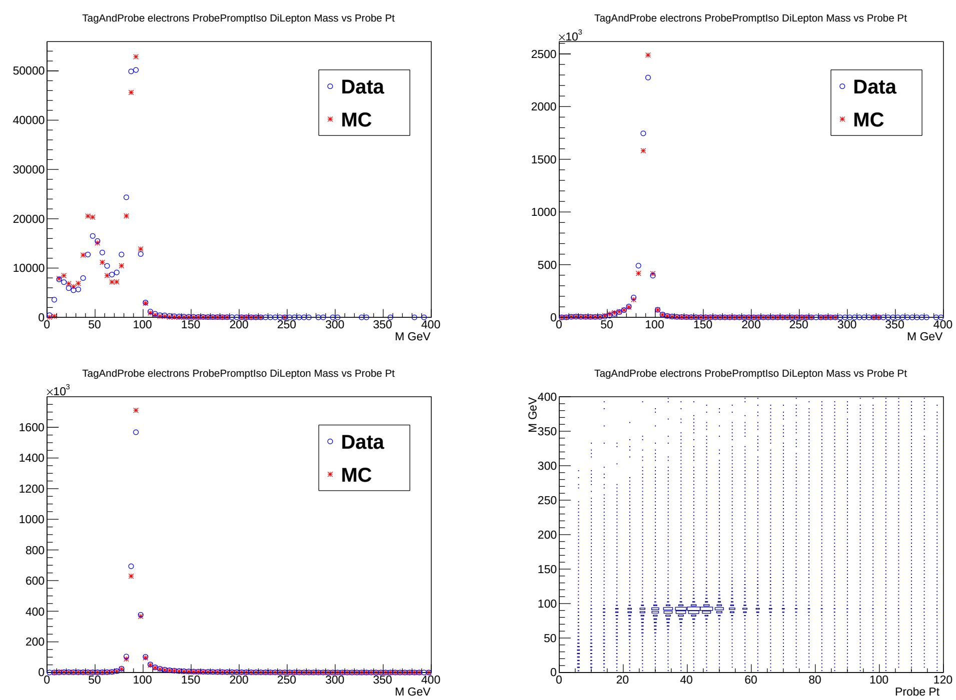


Figure 6.2: Dielectron invariant mass between a tag and probe electron. Shown is the mass for probe p_{T} from 12-24 GeV (top left), 24-48 GeV (top right), and > 48 GeV (bottom left). The mass versus probe p_{T} is shown bottom right [92].

fitting the ratio between data and MC isolation efficiencies to Equation 6.2, proposed in Ref. [93],

$$\epsilon(p_{\mathrm{T}}) = \epsilon_{\infty} \times Erf(\frac{p_{\mathrm{T}} - C}{\sigma}) + \epsilon_{C} \times (1 - Erf(\frac{p_{\mathrm{T}} - C}{\sigma})). \tag{6.2}$$

- ϵ_{∞} = Efficiency in the active (plateau) region at high momenta value.
- C = Specific p_{T} selection value for leptons.
- ϵ_C = Efficiency value at p_{T} = C.
- σ = Determines the rate of change in value as p_{T} decreases.

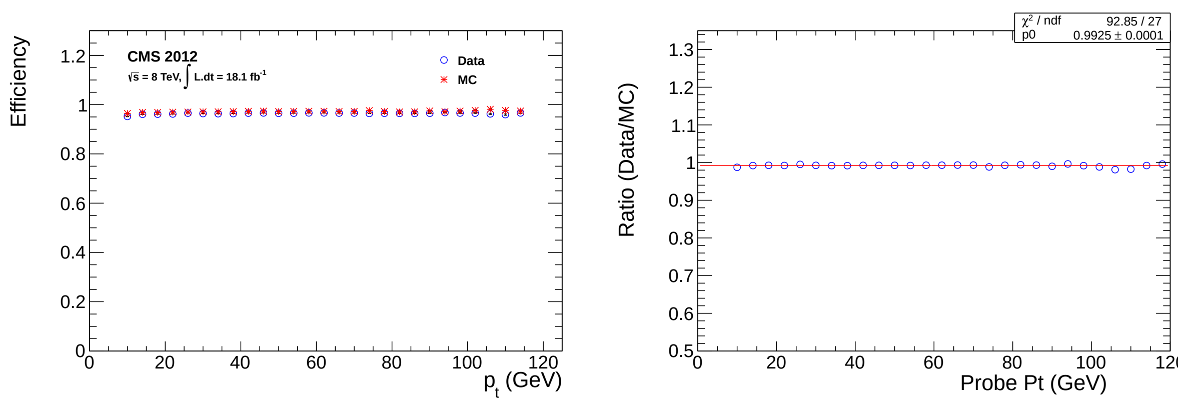


Figure 6.3: Muon identification efficiency as a function of probe p_{T} (left) and ratio of data and MC (right) [92].

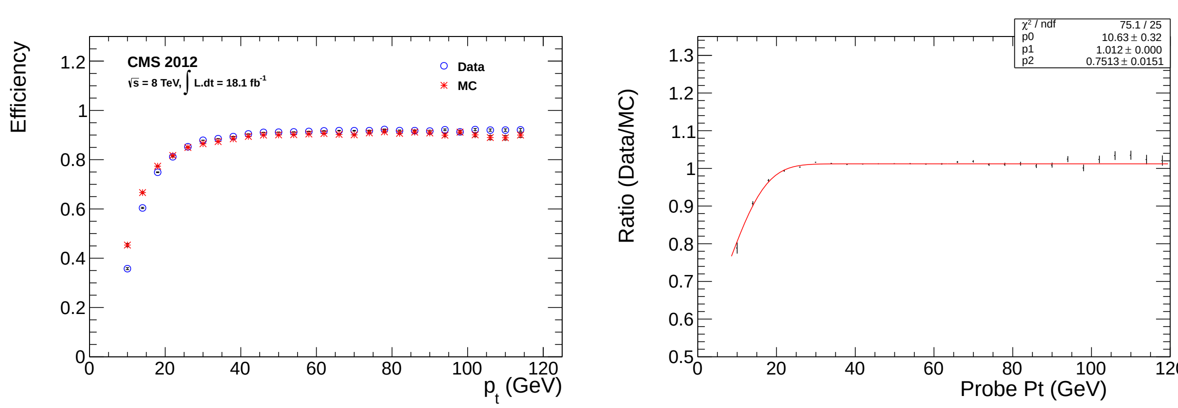


Figure 6.4: Electron identification efficiency as a function of probe p_{T} (left) and ratio of data and MC (right) [92].

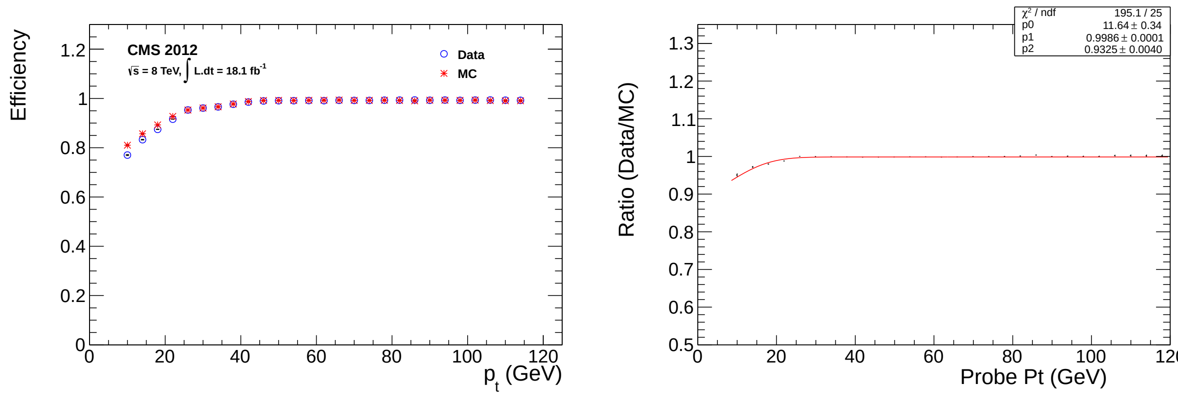


Figure 6.5: Muon isolation efficiency as a function of probe p_{T} (left) and ratio of data and MC (right) [92].

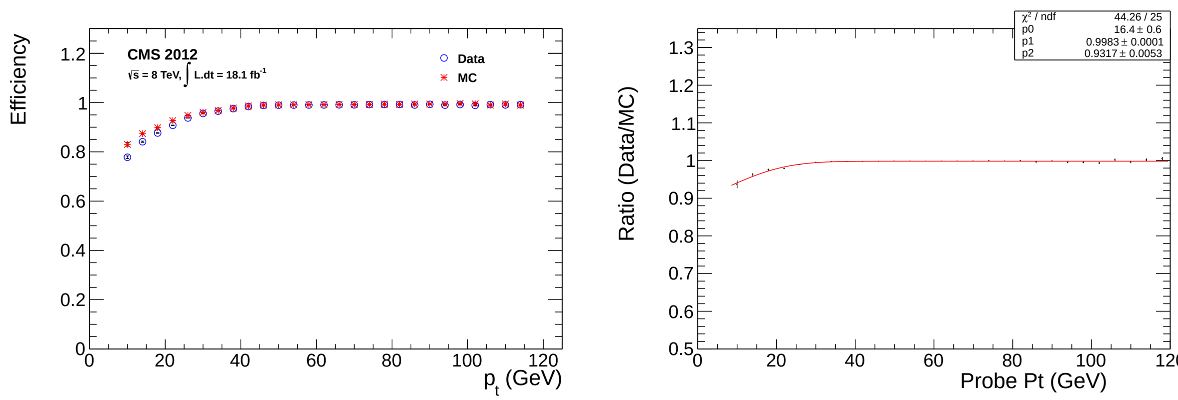


Figure 6.6: Electron isolation efficiency as a function of probe p_{T} (left) and ratio of data and MC (right) [92].

We can also use Equation 6.2 to estimate the electron and muon identification efficiency scale factors, where the efficiencies and scale factors depend on event properties. These quantities decrease with increasing jet activity and isolation efficiency is inversely proportional to pileup. In order to ensure the efficiencies and scale factors are appropriate for the kinematic properties of the events of interest, they are calculated as a function of η ($|\eta| < 1.5$ for barrel and $1.5 < \eta < 2.1$ for endcap), number of pileup vertices, and number of jets. Changes with the number of vertices and number of jets, serve as systematic uncertainties on the lepton efficiencies.

The resulting fit parameters for both the muon and electron isolation efficiency are as follows [92]:

- $\sigma_\mu = 11.6361 \pm 0.3416$ (stat) ± 2.3697 ($\mathrm{syst_{BE}}$) ± 1.8662 ($\mathrm{syst_{jet}}$) ± 1.7979 ($\mathrm{syst_{vert.}}$)
- $(\epsilon_\infty)_\mu = 0.9985 \pm 0$ (stat) ± 0.002 ($\mathrm{syst_{BE}}$) ± 0.0009 ($\mathrm{syst_{jet}}$) ± 0.0002 ($\mathrm{syst_{vert.}}$)
- $(\epsilon_C)_\mu = 0.9324 \pm 0.0039$ (stat) ± 0.0371 ($\mathrm{syst_{BE}}$) ± 0.1041 ($\mathrm{syst_{jet}}$) ± 0.0166 ($\mathrm{syst_{vert.}}$)
- $\sigma_e = 16.4017 \pm 0.5597$ (stat) ± 0.5075 ($\mathrm{syst_{BE}}$) ± 1.9723 ($\mathrm{syst_{jet}}$) ± 2.839 ($\mathrm{syst_{vert.}}$)
- $(\epsilon_\infty)_e = 0.9982 \pm 0.0001$ (stat) ± 0.001 ($\mathrm{syst_{BE}}$) ± 0.0004 ($\mathrm{syst_{je}t}$) ± 0.0001 ($\mathrm{syst_{vert.}}$)
- $(\epsilon_C)_e = 0.9316 \pm 0.0052$ (stat) ± 0.0054 ($\mathrm{syst_{BE}}$) ± 0.015 ($\mathrm{syst_{jet}}$) ± 0.005 ($\mathrm{sys_{vert.}}$)

6.3 Trigger efficiency

The data used in this analysis was collected by several different triggers. The triggers are not modeled in the MC simulation, therefore the trigger efficiency is measured in data and applied to the simulation. The triggers selected depend on the types of physics objects being considered in the search. In a multilepton search we want events which contain leptons, meaning events recorded that contain at least one lepton. We use un-prescaled lepton triggers that have the lowest possible p_T threshold for these objects. As the p_T threshold decreases, the frequency of firing such triggers increases and due to the limitation of the rate on processing, prescale factors are applied to limit this rate [1]. A comprehensive list of triggers is found in Appendix B.

The data used in this multilepton search is from double-lepton triggers. The various dilepton trigger efficiencies are obtained from a comparison with independent jet energy triggers. The trigger efficiencies are measured from $\mathrm{H_T}$ triggered data samples that have isolated leptons passing the lepton selection criteria. This method ensures an unbiased selection due to the use of independent trigger paths. If the efficiency of the i^{th} trigger is ϵ_i and the j^{th} trigger is ϵ_j, and the triggers are uncorrelated, the efficiency for an event to satisfy both triggers is $\epsilon_{ij} = \epsilon_i \times \epsilon_j$. For each dilepton-trigger of interest we select events

[1] For example, a pre-scale of 10 means the trigger will fire at every 10^{th} event.

that contain at least two leptons with p_{T} above 10 GeV and measure the fraction of events that fired the aforementioned trigger. This fraction represents the trigger efficiency, which can be formally defined as,

$$\epsilon_{\text{Trigger Efficiency}} = \frac{\text{N(events with 2 reconstructed leptons \& dilepton trigger fired)}}{\text{N(total number of events with 2 reconstructed leptons)}}. \quad (6.3)$$

In the case of the double-electron trigger, the efficiency is given by the ratio of the number of events passing the two good isolated electron criteria and that fired at least one of the dielectron analysis triggers to the number of events which have two good isolated electrons. $\mathrm{H_T}$ triggers are used as 'tag' triggers. In order to remove the effect of the turn-on curve of these $\mathrm{H_T}$ triggers we estimate the dilepton trigger efficiencies in the active region, where the tag trigger efficiency reaches a plateau. These trigger efficiencies as a function of the probe lepton p_{T} have smooth turn-on curves. In addition, we require $E_{\mathrm{T}}^{\mathrm{miss}} > 180$ GeV (or $H_{\mathrm{T}} > 300$ GeV) and $E_{\mathrm{T}}^{\mathrm{miss}} > 70$ GeV, which helps remove trigger biases and correlation between lepton and $\mathrm{H_T}$ triggers.

Due to the large number of double-muon triggers we determine the trigger efficiency for the logical "OR" of all triggers in this category. The efficiency for the double-electron and electron-muon triggers are calculated in the same manner. We measure the double-muon trigger efficiency to be 90% with no significant p_{T} dependence. The double-electron trigger efficiency is 95% when the sub-leading electron p_{T} is greater than 20 GeV, while that of the electron-muon trigger is 93% when the sub-leading lepton p_{T} is also greater than 20 GeV. Lastly, the double-electron trigger efficiency is 82% when the sub-leading electron p_{T} is less than 20 GeV,and for the electron-muon trigger the efficiency is 86% for the same p_{T} range. The trigger efficiencies for the double-electron and double-muon triggers as function of the sub-leading lepton p_{T} are shown in Figure 5.7. The simulated events are weighted by the probability for an event to satisfy the double-lepton triggers. The uncertainty in the corrections of the MC simulations translates into a systematic uncertainty in the irreducible backgrounds, as described in Section 6.7.

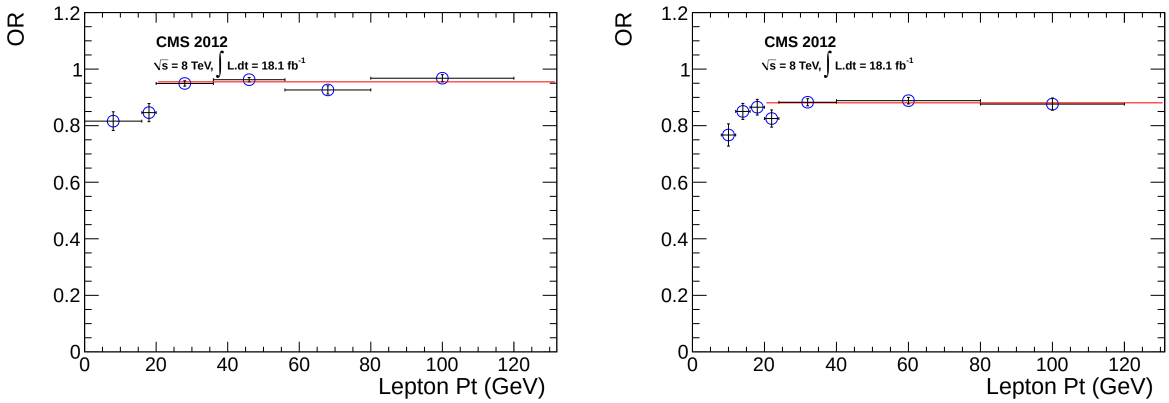


Figure 6.7: Dimuon "OR" trigger efficiency (left) and dielectron trigger "OR" efficiency (right) [85].

Chapter 7

Analysis strategy

In this chapter we, describe the search strategy based on a multichannel counting experimental approach. The general philosophy is to use exclusive search channels classified by quantities such as lepton flavor, invariant mass of dileptons representing Z-boson candidates, number of b-jets, and a kinematic quantity called S_T. We will discuss both the background pruning method to minimize the amount of SM contributions and background estimations for the different channels containing signal events.

7.1 Search strategy

Events are categorized on the basis of the number of leptons, lepton and jet flavor, charge and flavor combinations, and various kinematic quantities. To maintain high sensitivity, the search channels with hadronic tau candidates are separated from electron and muon channels due to the larger backgrounds arising from a higher tau misidentification rate.

We categorize each event in terms of the number of opposite-sign same flavor (OSSF) dilepton pairs. Each identified electron or muon is used only once for this assignment. As an example, both $\mu^+\mu^-\mu^-$ and $\mu^+\mu^-e^-$ have one OSSF pairs, $\mu^+\mu^+e^-$ has no OSSF pairs, and $\mu^+\mu^-e^+e^-$ has two OSSF pairs. The amount of SM background across the various channels varies considerably. All lepton charge combinations are considered as different channels, since the ones containing OSSF pairs or hadronic tau candidates suffer from larger backgrounds as compared to those channels with no OSSF pair. Jets are prone to be misidentified as hadronic taus, and Drell-Yan (DY) processes can contribute to OSSF pair channels.

We further classify events based on the presence of a leptonically decaying Z, if at least one OSSF pair has a reconstructed invariant mass, $m(\ell^+\ell^-)$, inside the Z-mass region (75-105 GeV), referred to as "on-Z". In this context, ℓ represents either an electron or muon. Events with $m(\ell^+\ell^-)$ outside the Z-mass region are referred to as "off-Z". We reject events with $m(\ell^+\ell^-) < 12$ GeV to avoid backgrounds from particles such as J/ψ, Υ mesons, and low mass DY processes. In order to remove contributions from leptons that arise from conversion of final state radiation in Z boson decays, we reject events with $|m(\ell^+\ell^-) - m_Z| > 15$ GeV and $|m(\ell^+\ell^-\ell'^{\pm}) - m_Z| < 15$ GeV or with $|m(\ell^+\ell^-\ell^{\pm}) - m_Z| < 15$ GeV, where $\ell\ell$ represents same-flavor lepton pair, and $\ell\ell'$ represents opposite-flavor lepton pair.

The SM background can be further reduced by requirements on the $\mathrm{S_T}$ of the event, where we define $\mathrm{S_T}$ as the scalar sum of $E_{\mathrm{T}}^{\mathrm{miss}}$ and the p_{T} of all isolated leptons and jets (see Equation 7.1). $\mathrm{S_T}$ is a useful quantity in this search because its distribution peaks near the sum of the parent particle masses. Therefore, events containing the production and decay of heavy particles, such as the signal events in this analysis, are expected to have much larger values of $\mathrm{S_T}$ than SM backgrounds. We divide the $\mathrm{S_T}$ distribution into several ranges: 0-0.3, 0.3-0.6, 0.6-1.0, 1.0-1.5, 1.5-2.0, and > 2.0 TeV. The different $\mathrm{S_T}$ ranges are designed such that the backgrounds from ZZ and WZ processes are in the lowest range and $\mathrm{t\bar{t}}$ background are in the lowest subsequent range, thus leaving the signal to occupy the highest part of the $\mathrm{S_T}$ spectrum with relatively small background.

$$\mathrm{S_T} = \sum \mathrm{p_T^{jet}} + \sum \mathrm{p_T^{lepton}} + \mathrm{E_T^{miss}}. \tag{7.1}$$

7.2 Background pruning

In order to enhance sensitivity to the new physics signal, it is important to reduce background contributions from the SM by applying basic object and event selection criteria. SM processes that are similar to the 'signal' of interest contribute to the 'background' of the search. Understanding and estimating SM background sources precisely in data improves the potential for discovery since any excess events can be a signal for new physics. In this analysis the predominant SM background arises from dilepton processes, such as Z+jets when accompanied by a third non-prompt lepton that passes the selection criteria, WZ production leading to three leptons, and $\mathrm{t\bar{t}}$ production followed by leptonic decays of the W bosons. Another contribution to the three lepton category comes from asymmetric internal photon conversion. The background estimation techniques employ data-based as well as simulation-based methods, both of which are described in the following sections.

The steps taken to minimize SM backgrounds are as follows:

- To reduce background contributions from low mass resonances such as J/ψ and Υ mesons we require events with an OSSF lepton pair to have the pair mass greater than 12 GeV.

- Light-leptons originating from jets (appearing to be prompt and isolated) are treated as fake leptons for the purpose of the analysis. We want to look at leptons that come from the hard interaction of the proton-proton collision, that are associated with the primary vertex. In the case of jets originating from b quarks, leptons can be produced through heavy-flavor hadron decays. These leptons tend to be produced further away from the hard interaction. Backgrounds of this type can be suppressed by isolation and vertex requirements on the leptons.

- We categorize events in a variety of exclusive channels according to lepton flavor, number of b-jets, number of OSSF pairs, and the invariant mass of these pairs. This allows the construction of search channels with greater signal to background ratios, which improves sensitivity.

MC simulation and data-based methods are used to obtain background estimations. Both of these methods were originally developed for the “Background and Efficiency Determination Methods for Multilepton Analyses”, a more detailed description can be found in [92].

7.3 Background estimation using a data-based approach

In certain cases, the MC simulation does not properly describe events due to either mis-modeling of the tail of a particular kinematic distribution, or in the case of asymmetric photon conversion, due to a generator level minimum cut off on p_{T}, as discussed in Section 7.4.3. For these reasons we employ data-based techniques to account for the background contribution from Z + jets and WW + jets events, where the jets can produce a fake lepton. To estimate SM background contributions of these types for the channels of interest, we use the data to determine the conversion factors. The driving principle behind a data-based technique is to find a proxy object which resembles the object of interest, in this case the fake object, in the event. Proxy objects are chosen such that they posses similar features and dependence on event kinematic properties as the fake object, but occur more frequently in the event than the fake object. This has the benefit of reducing statistical uncertainties and making it easier to handle systematics in the background estimates. We calculate the ratio of the rate of production of the fake object to that of the proxy object in a control region, verify it in another control region, and apply it to events in the signal-like search regions in data. A more detailed description for estimating the fake lepton backgrounds for different lepton flavors and sources is given in the the following sections.

7.3.1 Background from the production of light-lepton from jets

Light-leptons, electrons and muons, that originate from jets are considered fake leptons in this analysis. The rate for jets to produce fake light leptons depends on several factors, such as jet shape, jet $\mathrm{p_T}$ spectra, and the probability for jets to posses heavy flavor content, that might not be properly modeled in the MC simulation.

A data-based approach is used to estimate the probability for jets to produce light lepton candidates, that appear to be prompt and isolated. In order to understand these light lepton candidates originating from jets, we use $t\bar{t}$ control regions in data requiring $E_{\mathrm{T}}^{\mathrm{miss}} < 50$ GeV and $\mathrm{H_T} < 200$ GeV, where $\mathrm{H_T}$ is the $\sum p_{\mathrm{T}}^{jet}$ of all the jets in the events. The obtained predictions for fake light leptons are then validated in the multilepton control regions prior to applying them to the signal regions. We accomplish this by finding the relationship for the rate of jets that produce isolated light-lepton candidates to the rate of jets that produce isolated tracks (see Equation 7.2), such as those coming from pions and kaons. We then use $2\ell + 1$ *isolated track* in data to predict the number of $2\ell + 1$ *fake lepton* events, in the signal regions to obtain the three lepton background contribution. The background for the four lepton signal regions is determined in a similar way.

We relate the number of isolated lepton N_{ℓ}^{Iso} candidates coming from jets to the number of isolated track N_{Track}^{Iso} candidates from jets by the conversion factor (i.e. fake rate) f_{ℓ}, given by,

$$N_{\ell}^{\mathrm{Iso}} = f_{\ell} \times N_{\mathrm{Track}}^{\mathrm{Iso}}, \text{ where } \ell \ = \ \mathrm{e} \text{ or } \mu. \tag{7.2}$$

The conversion factor between isolated tracks and isolated lepton candidates originating from jets is sensitive to the heavy flavor jet content in the dataset used. The relative composition of light and heavy flavor jets in the data can differ between the control region and the signal region, leading to under or over estimation of the background yields. To understand the variation in the conversion factor that might appear across various datasets we use other objects to gauge this change. An isolated track can serve as a proxy object and other objects such as non-isolated leptons, or non-isolated prompt and non-prompt tracks can be used to test changes in the conversion factors. In equation 7.3, the conversion factor is expressed in terms of objects that can be measured in data for both signal and control regions [85]. The number of non-isolated leptons and tracks are measured directly from data where the conversion factors are applied. Also, the parametrization of the isolation efficiencies is obtained from a control sample.

The conversion factor f_ℓ is then giving by,

$$f_\ell = \frac{N_\ell^{\text{non}-\text{Iso}}}{N_{\text{Track}}^{\text{non}-\text{Iso}}} \times \frac{\epsilon_\ell^{\text{Iso}}}{\epsilon_{\text{Track}}^{\text{Iso}}}, \text{ where } \ell = \text{e or } \mu. \tag{7.3}$$

- N_ℓ: Number of non-isolated lepton candidates in the dataset of interest
- $\mathrm{N}_{\text{Track}}$: Number of non-isolated tracks
- $\epsilon_\ell^{\text{Iso}}/\epsilon_{\text{Track}}^{\text{Iso}}$: Isolation efficiency ratio of leptons to tracks

Jet flavor composition has an affect on the conversion factor since jets originating from heavy flavor quarks are more likely to produce fake leptons candidates than lighter flavor quarks. Since tracks from heavy flavor jets have a broader impact parameter distribution compared to tracks from light flavor jets this can be used to estimate the changes in the flavor composition of jets in the data.

In order to parametrize the isolation efficiency ratio we define another parameter called R_{dxy}, which is sensitive to the heavy flavor content in the various datasets of interest. R_{dxy} is defined as the ratio of the number of non-isolated tracks with large impact parameter ($|d_0| > 0.02$ cm), where the impact parameter is the perpendicular distance between the primary vertex and the point of closest approach of the track, to the number of non-isolated tracks with small impact parameter ($|d_0| < 0.02$ cm) in a giving data sample. R_{dxy} is determined from tracks within a p_{T} range of 8-24 GeV and $|\eta| < 2.4$. In the case where the dataset of interest predominantly contains b-jets, R_{dxy} will have a range of $0.20-0.30$, while datasets with very little heavy flavor jet content will have R_{dxy} in the range of $0.03-0.04$. The conversion factor between isolated tracks and electron (muon) candidates is measured to be 0.7%±0.2% (0.6%±0.2%) in a dilepton data sample. The parameter R_{dxy} is calculated after subtracting the expected contribution from simulated $\mathrm{t\bar{t}}$ processes. The relationship between the isolation efficiency ratio, ϵ_{Ratio}, and R_{dxy} is determined by using the analytic relationship between the two quantities from the curves shown in Figure 7.1. A more detailed derivation of this relationship can be found in Appendix C .

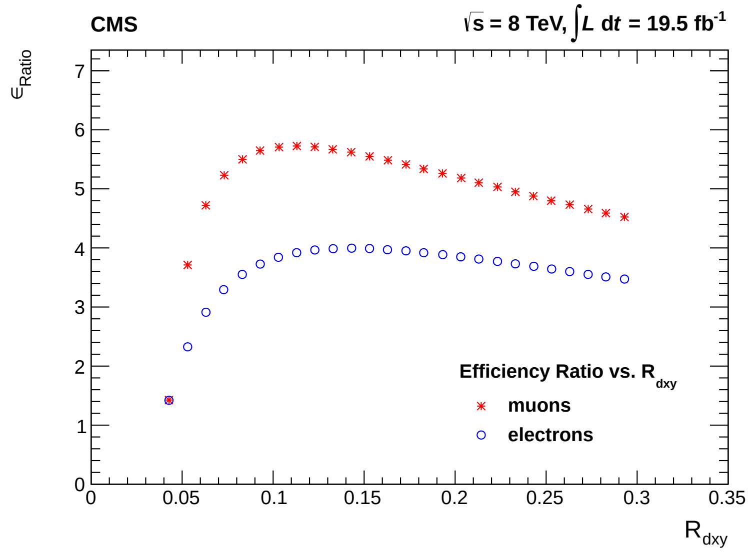


Figure 7.1: Ratio of isolation efficiency of leptons to tracks vs R_{dxy} for both electrons and muons [95].

7.3.2 Background from the production of hadronic taus from jets

The primary source of fake hadronic taus are jets. The background contribution for hadronically decaying taus, as opposed to electrons and muons, are much more challenging to determine based on isolation requirements alone. A data based approach is used to determine this type of background using an isolation sideband method [92]. Absolute isolation I_{abs} is defined as the isolation energy deposit in a cone of $\Delta R < 0.3$ around the tau candidate. An isolation requirement strongly reduces the SM background from misidentified leptons. We can determine a tau fake rate (i.e. conversion factor), f_τ, as the ratio of the number of tau candidates in the signal region, having absolute isolation of $I_{abs} < 2\,\mathrm{GeV}$, to the number of hadronic tau candidates with $6\ \mathrm{GeV} < I_{abs} < 15\ \mathrm{GeV}$. Figure 7.2 shows the isolation distribution of hadronic tau candidates for a soft jet p_{T} spectra (low-p_{T}) in red and for a hard jet p_{T} spectra (high-p_{T}) in blue. The tau fake rate is sensitive to both jet multiplicity and jet flavor. Therefore, we define another parameter, f_{SB}, as the ratio of the number of tau candidates in the sideband region ($6\ \mathrm{GeV} < I_{abs} < 15\ \mathrm{GeV}$) to the total number of non-isolated ($I_{abs} > 15\ \mathrm{GeV}$) tau candidates for the given dataset. The variable f_{SB} gives a handle in understanding the kind of isolation distribution a sample can have depending on its jet multiplicity and jet p_{T} spectra.

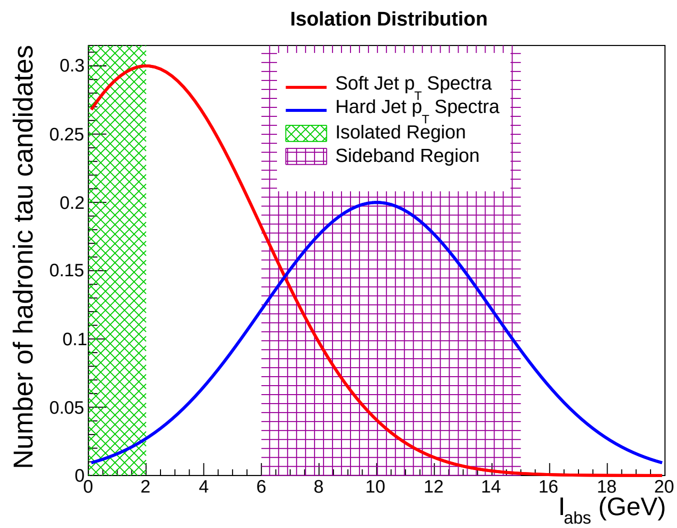


Figure 7.2: Model for isolation distribution showing various isolation regions used in the estimation of fake hadronic tau background. The red curve represents the isolation distribution for a soft jet p_{T} spectra while the blue curve for the hard jet p_{T} spectra. The isolation region $I_{abs} < 2\,\mathrm{GeV}$ in green, sideband region $6\ \mathrm{GeV} < I_{abs} < 15\ \mathrm{GeV}$ in magenta, and others $I_{abs} > 15\ \mathrm{GeV}$ in white.

To understand the contribution from reconstructed fake hadronic taus, we loosen the isolation requirements to get a conversion factor between loose ($I_{abs} > 15\ \mathrm{GeV}$) taus and tight ($I_{abs} < 2\ \mathrm{GeV}$) isolated taus. This is accomplished by determining the correlation between f_τ and f_{SB} by dividing data into different ranges of $\sum p_T^{\mathrm{Track}}$ of all the tracks in the event associated with the primary vertex. We calculate f_τ and f_{SB} for each range and find the functional dependence. Figure 7.3 shows f_τ vs f_{SB} for dilepton events with $E_{\mathrm{T}}^{\mathrm{miss}} < 100\ \mathrm{GeV}$ and $H_{\mathrm{T}} < 200\ \mathrm{GeV}$, in the p_T ranges $20 - 40\ \mathrm{GeV}$ (right plot) and $40 - 60\ \mathrm{GeV}$ (left plot), where we bin data in $\sum p_{\mathrm{T}}^{\mathrm{Track}}$ associated with the primary vertex. The isolation distribution for the hadronic tau candidates vary based on these $\sum p_{\mathrm{T}}^{\mathrm{Track}}$ bin ranges, as shown in Figure 7.4.

We measure the conversion factor to be $20\% \pm 6\%$. We find that this ratio is the same for dilepton data and jet-triggered data within 30% of itself, which is assigned as a systematic

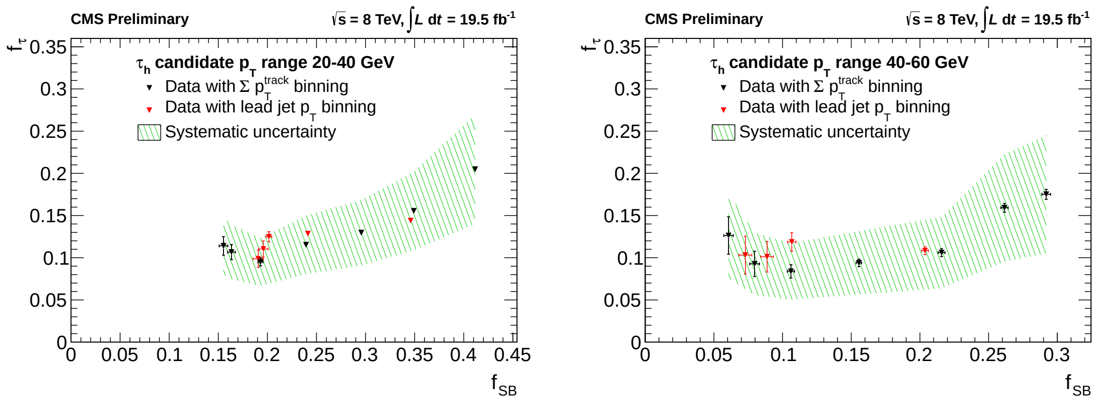


Figure 7.3: f_τ vs f_{SB} for hadronic taus with visible tau p_T between 20-40 GeV (left) and 40-60 GeV (right) in dilepton data [95].

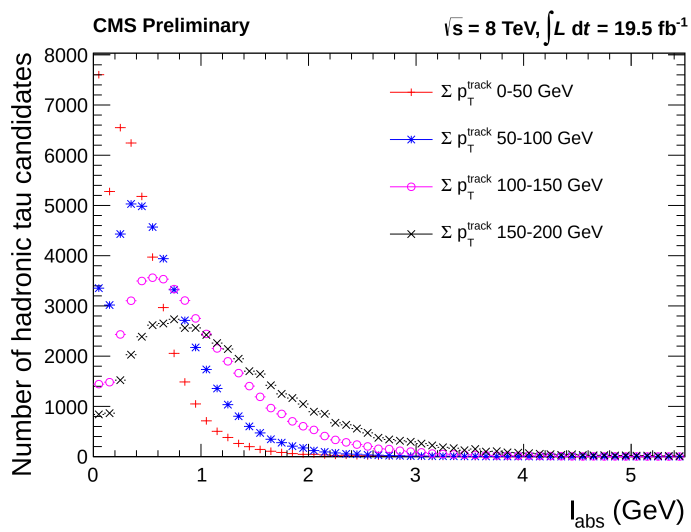


Figure 7.4: Isolation distribution of hadronic tau candidates for different jet activities ($\sum p_{\mathrm{T}}^{\mathrm{Track}}$) [92].

uncertainty. The tau fake rate is applied to the $2\ell + 1$ *sideband tau* candidate event sample to estimate the fake hadronic tau contribution.

7.4 Background estimation using MC simulation

We consider multiple background processes in this multilepton search: electrons and muons originating from jets, jets faking taus, $\mathrm{t\bar{t}}$ production, VV + jets (V = W or Z) processes, and asymmetric photon conversions. Each of these types of background is estimated separately. The types of background that cannot be distinguished from the signal scenario are referred to as "irreducible", since there are no selection criteria that would improve the signal to background ratio significantly. Simulated events are used to determine these irreducible backgrounds and a validation is performed using control regions in data. These control regions are selected in such a way that only one particular background contribution under study dominates. Scale factors are determined for each background contribution in the control regions, which are then applied to model those background in the signal regions.

7.4.1 Background from $\mathrm{t\bar{t}}$ production

Simulated events are used for SM background prediction involving $\mathrm{t\bar{t}}$ processes, since their jet flavor composition and p_T spectra have been fully studied and are well understood. However, we verify that the MC generator accurately simulates the $\mathrm{t\bar{t}}$ background contributions in a control region. We correct and validate $\mathrm{t\bar{t}}$ simulations in single lepton and dilepton control regions, respectively. The opposite sign eμ dilepton $\mathrm{t\bar{t}}$ control region is dominated by the fully-leptonic decay of $\mathrm{t\bar{t}}$, since these decays guarantee two real leptons. We observe good agreement in the kinematic properties such as $\mathrm{S_T}$, $\mathrm{H_T}$, and $E_\mathrm{T}^\mathrm{miss}$ between simulation and data, as can be seen in Figure 7.5.

In the case of the single muon control region we require events to have a prompt and isolated muon with $p_\mathrm{T} > 30$ GeV, at least three jets with $p_\mathrm{T} > 40$ GeV, where one of them must be tagged as a b-jet, and $\mathrm{S_T} > 300$ GeV. We can study the isolation distribution, as

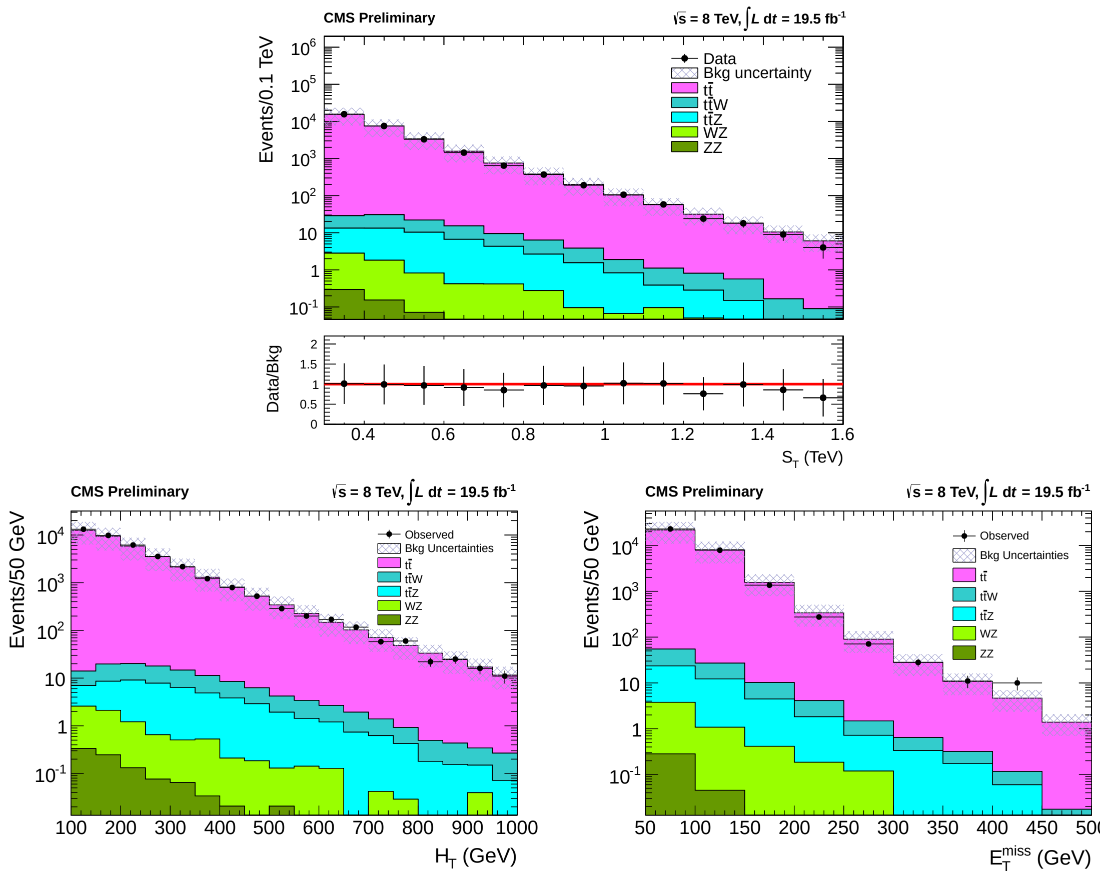


Figure 7.5: The S_T distribution of datasets dominated by $t\bar{t}$ in the opposite sign $e\mu$ dilepton control region. Also included are the H_T and E_T^{miss} distribution [95].

shown in Figure 7.6, for non-prompt muons originating from b-jets by considering additional muons found far from the leading b-jet, which have a large impact parameter. We assume the isolation distribution of muons coming from b-jets does not depend on the impact parameter. From the non-prompt single muon isolation distribution (left plot), in the $I_{abs} < 0.2$ GeV region, we extract a scale factor of 1.5 and validate it in the control region with the non-prompt single electron isolation distribution (right plot).

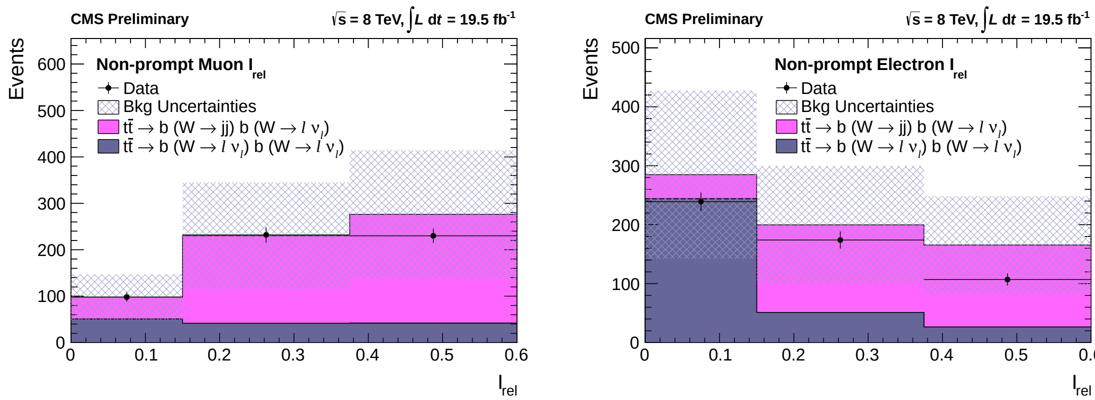


Figure 7.6: Left: Relative isolation distribution of the non-prompt μ in the single lepton $\mathrm{t\bar{t}}$ control region. Right: Relative isolation distribution of the non-prompt e in the single lepton $\mathrm{t\bar{t}}$ control region. The 1.5 scale factor has been applied to these distribution [95].

We measure a scale factor of 1.5 from the isolation distribution of light leptons coming from jets, in a single lepton control region, to correct the simulation of $\mathrm{t\bar{t}}$ events. Moreover, the scale factor is applied to events selection where the simulated $\mathrm{t\bar{t}}$ processes can contribute a fake lepton. For example, in a trilepton search region, where one of the light lepton comes from a heavy jet.

7.4.2 Diboson + jets production and rare SM processes backgrounds

SM processes that can produce three prompt and isolated leptons events are diboson process, such as WZ + jets and ZZ + jets, where both bosons decay leptonically. Simulation is corrected to match the measured lepton efficiency, trigger efficiency, b-jet identification efficiency, and $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution in data. The $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution is unfavorably affected by both pileup and jet activity where stochastic contributions result in poorer $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution. Therefore, a smearing factor for the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution is determined as a function of the number of vertices, to account for pileup, and of $\mathrm{H_T}$ for the jet activity in the event. A large number of vertices in an event indicates a large extraneous energy in reconstructed objects due to pileup affecting the Gaussian width of the distributions. $E_{\mathrm{T}}^{\mathrm{miss}}$ smearing is added to the MC simulation to match data, where the amount of smearing is determined on an event by event basis depending on the number of vertices and the $\mathrm{H_T}$ of the event. A larger $\mathrm{H_T}$ indicates higher jet activity, leading to systematically larger tails in the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution due to misreconstruction. This affect is considered in the systematic uncertainties, which are estimated by varying the weighted factors and determined by the level of migration in the number of events observed in different $E_{\mathrm{T}}^{\mathrm{miss}}$ and $\mathrm{H_T}$ ranges.

We verify the simulation by comparing with a data sample enriched in WZ production, which represents the dominant contribution to trilepton signatures from diboson + jets. WZ events can be selected by requiring three leptons, 50 GeV $< E_{\mathrm{T}}^{\mathrm{miss}} <$ 100 GeV, an on-shell Z, and $\mathrm{H_T} < 200\,\mathrm{GeV}$.

The transverse mass is defined as,

$$M_{\mathrm{T}}^2 = 2 \cdot p_{\mathrm{T}}^{\ell} \cdot E_{\mathrm{T}}^{\mathrm{miss}}(1 - cos\Delta\phi), \tag{7.4}$$

where $\Delta\phi$ is the angle between the lepton (ℓ = e or μ) and the $E_{\mathrm{T}}^{\mathrm{miss}}$. This variable is representative of the W boson having an edge around 80 GeV. Figure 7.7 shows the transverse mass distribution of the W boson. The normalization factor for the simulated WZ events is determined by normalizing the M_{T} distribution to data in the region 50 GeV $< M_{\mathrm{T}} <$ 120 GeV. We validate this normalization of the WZ simulation in the control region requiring $\ell^+\ell^-$ pairs to have an invariant mass of 75 GeV $< m_{\ell^+\ell^-} <$ 105 GeV and $\mathrm{H_T}$ < 200 GeV, with the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution, as shown in Figure 7.8.

For the four or more lepton search channels, the main source of irreducible background

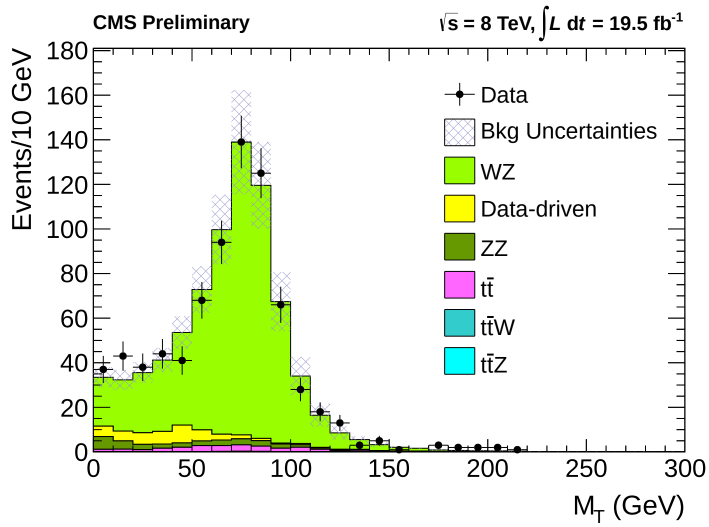


Figure 7.7: The transverse mass M_T distribution of events in a data sample enriched in WZ requiring an OSSF pair with $m_{\ell\ell}$ in the Z mass range and $100\,\text{GeV} < E_{\mathrm{T}}^{\text{miss}} < 150\,\text{GeV}$ [85].

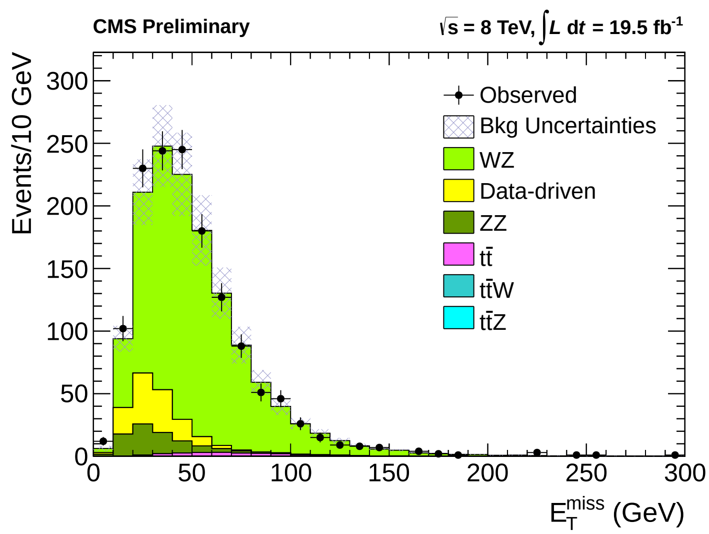


Figure 7.8: Distributions for $E_{\mathrm{T}}^{\text{miss}}$ in the WZ and opposite sign $e\mu$ dilepton control regions [95].

arises from ZZ + jet processes. The ZZ cross section is normalized in the ZZ event dominated control region, that require four lepton events with two OSSF pairs which are on-Z, with $\mathrm{H_T} < 200$ GeV, and $E_{\mathrm{T}}^{\text{miss}} < 50$ GeV. We normalize the ZZ simulation to data in the four lepton invariant mass distribution. We validate this normalization in a control region which requires at least one on-Z OSSF pair, as shown in Figure 7.9.

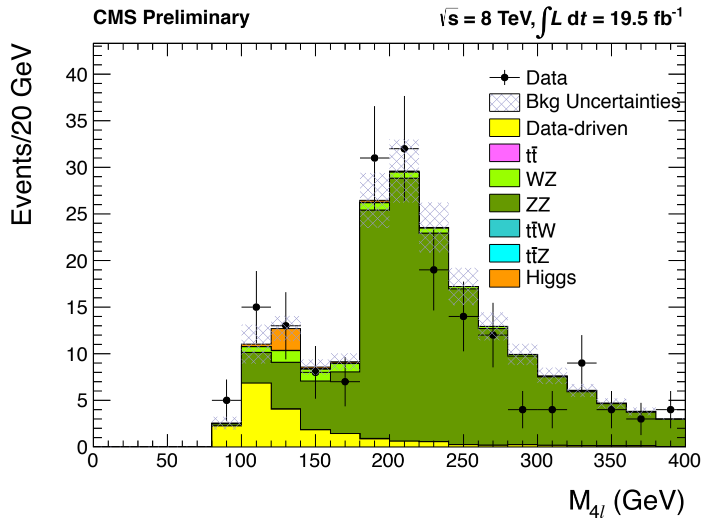


Figure 7.9: Distributions for $M_{4\ell}$ in the ZZ control regions with at least one $\ell^{\pm}\ell^{\mp}$ on-Z [95].

We consider other backgrounds arising from rare irreducible SM processes. These rare processes are top pair production in association with a W, Z, or Higgs boson, e.g. $\mathrm{t\bar{t}W}$, $\mathrm{t\bar{t}Z}$, and $\mathrm{t\bar{b}Z}$. The prediction from these background is derived from MC simulation, with cross sections of 0.2057 pb, 0.232 pb, and 0.0114 pb [62, 96] for $\mathrm{t\bar{t}W}$, $\mathrm{t\bar{t}Z}$, $\mathrm{t\bar{b}Z}$, respectively, with next-to-leading order (NLO) or next-to-next-to-leading order (NNLO) precision. We assign a 50% systematic uncertainty to account for the uncertainty in the NLO cross section calculation and for the limited experimental cross-check measurements for these processes. We include as well background from SM Higgs processes such as gluon-gluon fusion, vector boson fusion, and associated production with a W boson, Z boson, and top quark pair.

7.4.3 Backgrounds from asymmetric internal photon conversions

Photons converting to a pair of $\ell^{+}\ell^{-}$ leptons are a source of background for multilepton searches. In particular, there are two types of photon conversions, “external” and “internal”. External photon conversion involves on-shell photons interacting with the material of the detector producing an $\ell^{+}\ell^{-}$ pair. Conversions of this type predominately result in $\mathrm{e^{+}e^{-}}$ pairs as compared to $\mu^{+}\mu^{-}$, due to the much higher mass of muons, and are accounted for in the electron selection of the analysis. In the case of internal photon conversion, the photon is off-shell and can produce muons nearly as often as electrons. Therefore, additional care is needed for this type of conversion. The process of asymmetric internal photon conversion, is one in which the p_{T} of the leptons is small and either does not pass the selection criteria or the lepton is not reconstructed. These types of events are not accurately accounted for in the MC simulation, because a minimum lepton p_{T} is implemented at the generator-level. Drell-Yan processes with such photon conversions lead to a significant background in the case of three lepton final state signatures. This motivates a data-based method for estimating background contributions from asymmetric internal photon conversion, in the sense that one of the leptons carries most if not all the photon p_{T} while the other soft lepton (≤ 1 GeV) either does not pass the lepton selection or is not reconstructed.

The most important source of this type of background in multilepton analyses involves a Z boson decaying to a pair of leptons, where an asymmetric internal conversion of a final state radiation off-shell photon results in additional leptons. This final state radiation has the affect of causing the invariant mass of the lepton pair decaying from the Z boson to not reconstruct the Z boson mass while the internal conversion adds one additional lepton to the event. A diagram of this process is shown in Figure 7.10, where the Z boson decays to a pair of muons and one of the muons emits an off-shell photon, which in turn produces a $\mu^+\mu^-$ pair. Events of this type can appear as a three-lepton event if one of the muons carries most of the off-shell photon's momentum.

In order to measure this background we assume that the rate for producing on-shell photons from SM processes is proportional to the rate of producing off-shell photons that yield asymmetric conversions to a pair of leptons. We define a conversion factor C_ℓ, as the ratio of the probability for an off-shell photon to produce asymmetrically a pair of leptons, which pass the selection criteria, to the rate for an on-shell photon, that passes the full analysis-level selection requirements. We measure the conversion factor using final state photon conversions on the Z boson for both muon and electron candidates in a control

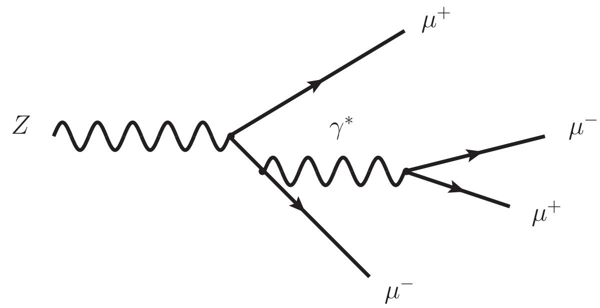


Figure 7.10: Feynman diagram showing a Z boson decaying to a pair of electrons and an asymmetric FSR decay to a pair of muons from one of the electrons.

region in the data. We look at events in the low $E_{\mathrm{T}}^{\mathrm{miss}}$, low $\mathrm{S_T}$, low $\mathrm{H_T}$, and on-Z region. As previously mentioned, if one of the decay products of the Z boson emits final state radiation, the dilepton mass will be off the Z boson mass peak, but the three-body mass reconstruction of the $\ell^+\ell^-\gamma^*$ will be on the Z mass region. The internal photon conversion factor can be expressed as,

$$C_\ell = \frac{N_{\ell^+\ell^-\ell^\pm} \text{ events which make a Z boson}}{N_{\ell^+\ell^-\gamma} \text{ events with on-Z boson}}. \tag{7.5}$$

To calculate the conversion factor we divide the number of events with $|m(\ell^+\ell^-\ell'^\pm) - m_\mathrm{Z}| < 15\,\mathrm{GeV}$ or $|m(\ell^+\ell^-\ell^\pm) - m_\mathrm{Z}| < 15\,\mathrm{GeV}$ by the number of events with $|m(\ell^+\ell^-\gamma) - m_\mathrm{Z}| < 15\,\mathrm{GeV}$. Figure 7.11 shows the three-body $\ell^+\ell^-\gamma$ invariant mass distribution where the $\ell^+\ell^-$ pair is not on the Z peak. In principle, this distribution has no contribution from external photon conversions and the peak is entirely from internal photon conversions.

The measured muon conversion factor C_μ is $0.7\% \pm 0.1\%$ and the electron conversion factor C_e is $2.1\% \pm 0.3\%$. The uncertainties are statistical only and we assign systematic uncertainties of 50% to these conversion factors. This is due to the underlying assumption that the production of on-shell photons is proportional to the production off-shell photons that result in asymmetric conversion. The measured conversion factors are then used to estimate the background from photon conversions in the signal regions from the observed

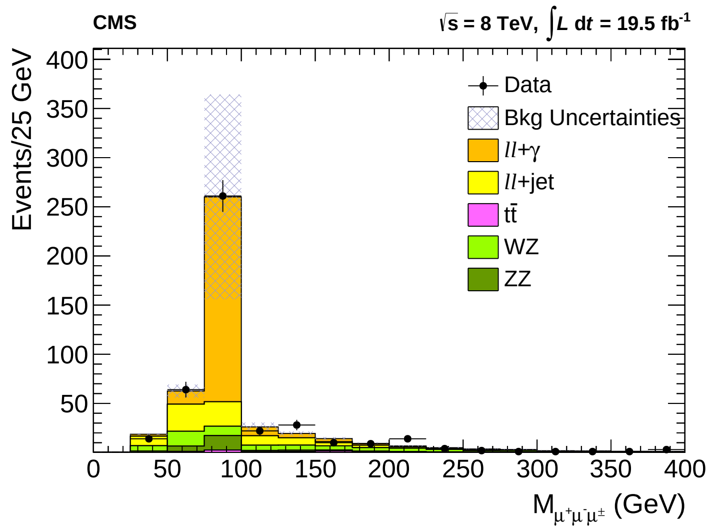


Figure 7.11: $m_{\mu^+\mu^-\mu^\pm}$ where both reconstructed $m_{\mu+\mu-}$ are either below (< 75 GeV) or above (> 105 GeV) Z mass region [95].

number of $\ell^+\ell^-\gamma$ events in the signal regions. The background contribution from these converted photons is small after final selections requirements.

Chapter 8

Uncertainties

An overview of the uncertainties associated with this analysis arising from different sources will be presented in this chapter, followed by a description of how these systematic uncertainties are evaluated for both signal and background.

8.1 The nature of uncertainties

Experimental results depend on the accuracy of the observed measurements. Statistical uncertainties are especially important in high energy particle physics due to the nature of the processes involved. Occurrences of decays follow a Poisson distribution with mean λ and a standard deviation σ defined by,

$$\sigma = \sqrt{N}, \tag{8.1}$$

where N is the number of events in a counting experiment. Additionally, there are also systematic sources of inaccuracy that need to be considered in order to accurately estimate uncertainties on the measurement. In general, systematic errors arise due to imperfect modeling of the observables and unavoidable biases in the measurements. In principle, a way to estimate the influence of an uncertainty on an analysis is to vary a quantity within its uncertainty and then observe the effect on the final result.

Sources of uncertainties affecting background and signal models that are considered include: uncertainty on the measured luminosity, jet energy scale uncertainty, b-tagging scale factor uncertainty, $\mathrm{t\bar{t}}$ and hadronic tau fake rate uncertainty, trigger and lepton efficiency uncertainties, and $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution uncertainty. The effect of the relevant systematic uncertainties for this analysis are evaluated and discussed in more detailed in the following sections.

8.2 Sources of systematic uncertainties

In this section, we describe the systematic uncertainties associated with this analysis. In principle, we determine systematic uncertainties for each channel by shifting events up and down, based on smearing or scaling of events and then taking the difference between the nominal values for a given channel and the values which result from shifting a particular quantity. In this multichannel counting experiment, bin migration is an important point to consider. As an example, the jet energy scale uncertainty will cause events to migrate from one $\mathrm{S_T}$ bin to another. Correlation or anti-correlation between the various channels are taken into account by keeping track of the relative sign of channel migration of events.

8.2.1 Integrated luminosity uncertainty measurement

The estimated uncertainty on the integrated luminosity measurement is 2.5% (sys.) $\pm$ 0.5% (stat.) [47]. The systematic uncertainties on the CMS calibration scan measurements are determined using the Van der Meer technique [97, 98]. Table 8.1 summarizes the sources and contributions of the individual systematic uncertainties associated with the integrated luminosity, which include stability across pixel detector regions, pixel gains and pedestals, dynamic inefficiencies, length-scale corrections, beam width evolution, beam intensity, scan-to-scan variations, and afterglow.

The luminosity measurement is based on a pixel cluster counting method. In this technique, an effective pixel cluster cross section is determines using a VdM scan, where the cross section is used to determine the luminosity for each luminosity section. The luminosity section can be estimated by counting the number of pixel cluster per zero-bias trigger [99]. A more detailed description of the luminosity measurement is found in Section 3.2.7.

All sources of uncertainties are summed in quadrature to obtain the total systematic uncertainty of 2.5% on the luminosity measurement, where the dominant contributions are attributed to stability across pixel detector regions a 1% uncertainty, as well as 2% from the model fit. The *Stability* across pixel detector regions systematic accounts for the small variations in the fractional cluster counts observed in the different barrel and endcap layers, while the *fit model* uncertainty is associated with the choice of fit model for $\sigma_{visible}$ depending the fit function (e.g. single Gaussian plus constant or double Gaussian plus constant) used to model the inelastic cross section [99].

Type	Source of Uncertainty	Correction (%)	Uncertainty (%)
Integration	Stability	-	1
	Dynamic inefficiencies	-	0.5
	Afterglow	2	0.5
Normalization	Fit model	-	2
	Beam current calibration	-	0.3
	Ghosts and satellites	-0.4	0.2
	Length scale	-0.9	0.5
	Emittance growth	-0.1	0.2
	Orbit Drift	0.2	0.1
	Beam-beam	1.5	0.5
	Dynamic-β	-	0.5
	Total		2.5

Table 8.1: Summary of the systematic uncertainties. If applicable percentage for normalization and on the total luminosity for afterglow effects is also reported [99].

8.2.2 Jet energy scale uncertainty

Jet energy scale uncertainties for signal and background events are calculated by shifting the p_{T} of each jet in the event up and down by a factor that varies as a function of p_{T} and η. This affects the $\mathrm{S_T}$ value as well as the number of b-tagged jets in the events, due to event migration between channels. We take as the systematic uncertainty the change which causes the largest variation in the expected number of events among all channels. This procedure results in an uncertainty of about 0.5% for the jet energy scale correction.

8.2.3 Lepton identification and isolation efficiency uncertainty

The results for the tag-and-probe method for the lepton identification and isolation efficiencies has been previously presented in Section 6.2. The isolation and identification efficiency for data and MC simulation do not exactly match, as seen in Figures 6.3 - 6.6. The disagreement between data and simulation is large for low p_T values. Therefore, we parameterize the ratio of lepton efficiency of data to MC. We correct MC simulation by scaling the events by this ratio for each reconstructed lepton in the event.

We fit the ratio of the lepton isolation efficiency between data and MC as a function of the probe lepton p_{T}. The fit parameters are calculated for two event selections that require either one or three jets in a Z + jets sample where the Z boson decays leptonically. We compare the values of the fit parameter and assign an uncertainty based on their differences. The total systematic uncertainty is the sum in quadratures of all the individual uncertainties of the fit parameters.

8.2.4 B-tagging scale factor uncertainties

During the hadronization process B hadrons are produced out of b quarks and gluons, forming jets that can be tagged as b-jets, as discussed in Section 5.5. Data and MC samples have different b-tagging efficiencies. Therefore, scale factors (SF) are derived and applied to MC simulated events to match the measured efficiency in data. In general, the SF is based on the ratio of the efficiency measured in data, ϵ_{data}, to the measured efficiency in simulation, ϵ_{MC}, and is parametrized as a function of the jet p_{T} and η. The SF are used to correct for residual discrepancies between data and MC and are subject to effects of systematic uncertainties. Sources of systematic uncertainty include the mis-modeling of the light jet p_{T} spectra. We follow the prescription given in Ref. [100] in order to apply the scale factors. In particular, the fit functions for the efficiency for generator level b-jet, c-jet and light jet to be b-tagged, are parametrized based on the jet p_{T}. Generator level information for each MC simulated event is used, such as the jet flavor and whether or not a jet is b-tagged. We correct the simulation on an event by event basis in order to correct the overall MC b-tagging efficiency of the simulation sample to match that of data. We can proceed to construct a SF for each event by multiplying weights of each jet in the event as follows,

$$SF = \prod_{i=1}^{\mathrm{N_{jet}}} w_i, \tag{8.2}$$

where the weight for each jet in an event corresponds to either w_i^{tagged} or $w_i^{\text{un-tagged}}$ depending on wether or not the jet was reconstructed as a b-jet:

- In the case the jet is tagged as a b-jet, w_i is defined as,

$$w_i^{\text{tagged}} = \frac{\epsilon_{Data}}{\epsilon_{MC}}. \tag{8.3}$$

  The numerator corresponds to the probability for a b-jet to be tagged in data, while the denominator is the probability for the b-jet to be tagged in MC.

- In the case the jet is not tagged as a b-jet, w_i is then defined as,

$$w_i^{\text{un-tagged}} = \frac{(1-\epsilon_{Data})}{(1-\epsilon_{MC})} = \frac{(1-w_i^{\text{tagged}}\cdot\epsilon_{MC})}{(1-\epsilon_{MC})}. \tag{8.4}$$

Here the ratio corresponds to the probability for the b-jet to not be tagged in data divided by the probability for the b-jet to not be tagged in MC. The effect of the measured SF are accounted for by reweighting the events. The efficiency in data is given by $\epsilon_{\mathrm{Data}} = w \cdot \epsilon_{\mathrm{MC}}$. The b-jet scale efficiency mainly depends on the p_T and flavor of the jet, at generator level, whether it is a light or heavier flavor jet (c- or b-jet). The product of weights for all jets in the event then gives an overall scale factor for the event.

We determine the systematic uncertainties of these weights by varying all the light jet scale factors up and down within their uncertainty (σ), while keeping the b/c scale factors fixed. Similarly, we vary the b/c-jet scale factors by $\pm\ 1\sigma$, as we keep the light jet scale factor fixed. Afterwards, we add in quadratures the deviation from the total weight w. Typically this amounts to a 6% uncertainty.

8.2.5 $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution uncertainties

A detailed look at the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution modeling and associated uncertainties can be found in Ref. [101]. We give a brief overview of the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution method, and how it is used to determine a systematic uncertainty.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution is parameterized based on the number of reconstructed vertices in the event, $\mathrm{N_{vertex}}$, to account for pileup and on the $\mathrm{H_T}$ in the event to account for jet activity. An event with a large $\mathrm{H_T}$ value can degrade the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution, as a consequence of a possible mismeasurement of the jet energy. The $E_{\mathrm{T}}^{\mathrm{miss}}$ x- and y-components are found to be approximately Gaussian. Therefore, the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution is models by a sum of Rayleigh distributions, given by,

$$p(E_{\mathrm{T}}^{\mathrm{miss}}) = \sum_{ij} w_{ij} \frac{E_{\mathrm{T}}^{\mathrm{miss}}}{\sigma_{ij}^2} e^{-E_{\mathrm{T}}^{\mathrm{miss}\,2}/2\sigma_{ij}^2}, \tag{8.5}$$

where "i" corresponds to the number of vertices and "j" represents a bin of $\mathrm{H_T}$ of width 40 GeV. The weight w_{ij} represents the fraction of events in the channel, which have i vertices and $\mathrm{H_T}$ given by j times 40 GeV. The width parameters σ_{ij}, which characterize the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution, are fitted for in dilepton events. The $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution width is parameterization with respect to $\mathrm{N_{vertex}}$ and $\mathrm{H_T}$. This coefficient in simulation is adjusted to match that found in data.

To correct the MC simulation to match data, smearing factors are applied to the MC sample based on the amount of $E_{\mathrm{T}}^{\mathrm{miss}}$ in the event, determined on an event-by-event basis depending on $\mathrm{N_{vertex}}$ and $\mathrm{H_T}$. This smearing factor applied to the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution has an associated systematic uncertainty. In order to evaluate the systematic uncertainty on the

Source of Uncertainty	**Uncertainty** (%)
Luminosity	2.5
$E_{\mathrm{T}}^{\mathrm{miss}}$ Resolution $(0-50\ \mathrm{GeV}, 50-100\ \mathrm{GeV}, >100\ \mathrm{GeV})$	(−3, +4, +4)
Jet energy scale WZ	0.5
b-tagging scale factor	0.1 (WZ), 6 ($\mathrm{t\bar{t}}$)
Muon ID/Isolation at 10 (100) GeV	11 (0.2)
Electron ID/Isolation at 10 (100) GeV	14 (0.6)
Tau ID/isolation at 10 (100) GeV	2 (1.1)
Dilepton trigger efficiency	5
$\mathrm{t\bar{t}}$ cross-section/fake contribution	50
WZ normalization	6
ZZ normalization	12
Asymmetric internal conversion fake rate	50
Fake muons (electrons) contribution	30 (30)
Fake tau contribution	30

Table 8.2: The systematic uncertainties associated with this analysis. The $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution systematic is given for WZ background on Z for different selection requirements on $E_{\mathrm{T}}^{\mathrm{miss}}$ and for different selection on $\mathrm{M_T}$ given a requirement of $E_{\mathrm{T}}^{\mathrm{miss}} > 50\,\mathrm{GeV}$ [85].

$E_{\mathrm{T}}^{\mathrm{miss}}$ resolution, we vary the amount of $E_{\mathrm{T}}^{\mathrm{miss}}$ smearing, based on the Rayleigh distribution, and determine the migration in the number of events among the different $E_{\mathrm{T}}^{\mathrm{miss}}$ bins. The systematic uncertainty can be either correlated or anti-correlated across the different $E_{\mathrm{T}}^{\mathrm{miss}}$ channels, since the number of events must be conserved between the channels. An increase in the number of events in the higher $E_{\mathrm{T}}^{\mathrm{miss}}$ region, results in a deficit in the number of events in the lower $E_{\mathrm{T}}^{\mathrm{miss}}$ region, corresponding to an anti-correlation in the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution systematic uncertainty due to event migration in the lower and higher $E_{\mathrm{T}}^{\mathrm{miss}}$ region.

8.3 Effects due to systematic uncertainties

In this section, we discuss the effects of various systematic uncertainties on signal and background estimations. Table 8.2 summarizes the sources of uncertainties relevant for this analysis. The total systematic uncertainty varies between 3% and 30% depending on the specific search channel.

8.3.1 Simulated signal and background uncertainties

We consider several sources of systematic uncertainty that affect the simulated SM background and signal, e.g. simulation scale, b-tagging scale factors, $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution.

The number of $\mathrm{t\bar{t}}$ events is in principal well known in the single lepton and dilepton control regions. Background simulations are effected by theoretical uncertainties on the cross section calculations, which come from parton distribution function uncertainties and the renormalization/factorization scale uncertainties. Furthermore, the background contributions found for this analysis depend heavily on the number of fake leptons arising from b-jets. A conservative systematic uncertainty of 50%, shown in Table 8.2, is assigned to account for the cross section uncertainty and for the estimate of the background due to misidentified leptons from $\mathrm{t\bar{t}}$ processes to cover the level of disagreement between data and simulation in the $I_{rel} < 0.15$ GeV region of the relative isolation distribution for non-prompt muons (see Figure 7.6).

We assign systematic uncertainties of 6% and 12% to the WZ and ZZ cross sections, respectively, in order to cover the level of disagreement between data and simulation found in two separate control regions, as previously described in Section 7.4.2 and shown in Figure 7.7. Systematic uncertainties associated with b-tagging scale factors, jet energy scale, luminosity, and $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution affect both signal and SM background estimates. The effect of the b-tagging scale factor uncertainty for WZ and $\mathrm{t\bar{t}}$ are evaluated to be 0.1% and 6%, respectively. The jet energy scale procedure for WZ results in a systematic uncertainty of 0.5%. The luminosity systematic uncertainty of 2.5% does not have an effect on data-based background estimates and only effects irreducible MC backgrounds and signals.

Signal and background simulated samples are subject to systematic uncertainties from trigger, lepton identification, and lepton isolation efficiencies. The lepton trigger systematic uncertainty for the MC simulation is based on the number and p_{T} spectra of isolated electrons and muons. We require each event to fire at least one of several triggers: double-muon, double-electron, and muon-electron, as discussed in Section 6.3. Trigger efficiencies are measured from data using an unbiased H_T triggered sample. We calculate the probability for at least one of the double-lepton triggers to be fired, which gives a measure of the trigger efficiency. We determine an overall 5% systematic uncertainty for the lepton trigger efficiency.

Sources of uncertainty related to lepton identification and isolation efficiency scale factors depend on the lepton p_T spectra. This systematic uncertainty is determined based on the deviation of the parameters used to fit the ratio of efficiencies between data and MC simulation, which include the number of pileup vertices, and the number of jets in the event. The difference of the lepton efficiency between data and MC is taken into account by scaling the MC simulation by appropriate scale factors, as previously discussed in Section 6.2. Muon and electron identification and isolation efficiencies agree between data and simulation to better than 1% for lepton $p_T > 20$ GeV. The electron identification scale factor uncertainties are 14% for a p_T of 10 GeV and 0.6% at 100 GeV. For muons the identification scale factor uncertainties are 11% for p_T of 10 GeV and 0.2% at 100 GeV. Tau identification scale factor uncertainties are 2% for a p_T of 10 GeV and 1.1% for a p_T of 100 GeV.

8.3.2 Uncertainties on data-based methods

Uncertainties associated with using a data-based approach for background predictions are derived from the accuracy of the proxy object technique used for light lepton fake rate, hadronic tau fake rate, and internal photon conversion used to estimate the backgrounds.

The dominant source of systematic uncertainties arise from the lepton fake rates. The light-lepton fake rate uncertainty is determine by taking the difference between the light-lepton fake rates, as measured in a Z + jet sample, based on two different event selections. For example, we estimate the electron fake rate by selecting $\mu^+\mu^-e^+$ events, where the opposite-sign muons reconstruct a Z boson, while the electron comes from the jet, and compare this lepton fake rate when selecting instead $e^+e^-e^+$ events. The measured fake rate on the third lepton should not depend on the decay of the Z boson. Therefore, we take this difference, of about 30%, and assign it as the uncertainty on the light-lepton fake rate technique. A systematic uncertainty of 30% on the muon fake rate is determined in a similar manner.

The hadronic tau fake rate f_τ, as described in Section 7.3.2, has an associated systematic uncertainty which is based on the amount of uncertainty necessary to cover the difference between the f_τ measured using $\sum p_T^{\mathrm{Track}}$ and $p_T^{\mathrm{Lead\ Jet}}$ bin types (Figure 7.3). The difference in hadronic tau fake rate, of about 30%, measured based on these bin types is assigned as the systematic uncertainty.

We assign a moderately conservative 50% systematic uncertainty on the asymmetric internal photon conversion fake rate due to the theoretical assumption of proportionality between off-shell and on-shell photons, as previously mentioned. The predominant uncertainties arise from internal photon conversion, $\mathrm{t\bar{t}}$ fake contribution, and hadronic tau fake rate uncertainties.

Chapter 9

Statistical analysis

In this chapter, we introduce the methods and tools needed to set upper limits on cross sections, that can then be used to set lower limits on the mass of hypothetical particles. The application of these techniques to the searches for massive vector-like quarks are discussed in Chapter 10. We begin this chapter by giving an overview of a statistical model for a multichannel approach, then we briefly discuss the theory behind hypothesis testing, followed by a description of the limit setting procedure.

9.1 Statistical method

A "statistical model" specifies the probability of observing a given dataset as a function of a set of underlying parameters, including the parameter about which we aim to make a statistical statement, such as a signal across section (or signal rate). Several statistical methods for limit setting can be applied depending on the statistical model.

We introduce a simple statistical model, that we subsequently extend to adapt to the multichannel analysis used in this study. This model involves a simple counting experiment in one channel, where the expected number of background events, b, is known and the signal cross section, σ_{signal}, is the parameter of interest. The number of observed events, n, completely specifies the data set. We assume that the probability to observe n follows a Poisson probability distribution,

$$\mathcal{P}(n|\lambda) = \frac{\lambda^n e^{-\lambda}}{n!}, \tag{9.1}$$

where λ, the mean of the distribution, is a function of the model parameters, σ_{signal}. We introduce a *signal strength modifier* μ, which scales the number of predicted signal events, s, calculated for a expected signal cross section $\sigma^{signal}_{predicted}$, where λ is now a function of μ,

$$\lambda(\mu) = b + \mu \cdot s. \tag{9.2}$$

The probability to observe n for a data set, given μ, is completely specified by Equations 9.1 an 9.2.

We extend this model from a simple single channel to a multichannel counting experiment, which includes all the search channels. In this extension of the model, the probability to observe n for a given data set is given by the product of Poisson probability distributions over all channels,

$$P(n|\lambda) = \prod_{i}^{N_{channel}} \mathcal{P}_i^{poiss}(n_i|\lambda_i) = \prod_{i}^{N_{channel}} \frac{\lambda_i^{n_i} e^{-\lambda_i}}{n_i!}, \tag{9.3}$$

where for each channel i, λ_i is the sum of the expected number of background events plus the scaled predicted number of signal events. Another modification to the model is to express the expected number of background events, b, for each channel as the sum of the different contributing background process considered, which are determined separately.

Additionally, we also introduce systematic uncertainties to the model. For each source of systematic uncertainty, an associated model parameter θ, so-called nuisance parameter, is introduced. For each background process the number of expected events is written as a function of these nuisance parameters, which can affect the overall rate of a process.

A more generalized model can now be written as,

$$P(n|\lambda, \theta) = \prod_{i}^{N_{channel}} \mathcal{P}_i^{poiss}(n_i|\lambda_i(\mu, \theta)), \tag{9.4}$$

$$\lambda_i(\mu,\theta) = \sum_{i}^{N_{background}} b_i(\theta) + \mu \cdot s_i(\theta), \tag{9.5}$$

where both the Poisson probability distribution and mean are expressed in terms of a set of nuisance parameter θ. The probability of observing a data set, given the parameters μ and θ is now defined by Equations 9.4 and 9.5. In the following section we describe how this expression can be read, for a fixed data set n, as a function of μ and θ, known as the "likelihood function" $\mathcal{L}(n|\mu,\theta)$.

9.2 Limit setting procedure

In this section, we describe the theory behind hypothesis testing and the statistical limit setting procedure [102]. In order to derive exclusion limits we need to make a comparison between two different hypotheses. The first is the background-only hypothesis (H_b) and the other is the signal plus background (H_{s+b}) hypothesis. In order to place limits on a potentially new physics signal we need to quantify how incompatible the data is with the H_{s+b} hypothesis scenario. There are in general two schools of thought regarding statistical inference, the frequentist and Bayesian approaches. In frequentist statistics, probability is interpreted as the frequency of an outcome when the measurement is repeated in an experiment. While in the Bayesian approach the term probability can be interpreted as "degree of belief" for a parameter of interest [102]. In Bayesian statistics, one begins by introducing a prior probability distribution function (*pdf*), reflecting the "degree of belief" about the parameter of interest, corresponding to those being constrained in the analysis. This 'prior *pdf*' describes the a-priori knowledge of where the parameter of interest should lie and Baye's theorem allows an update of it in light of the new data.

9.2.1 Modified frequentist method-CL$_\mathrm{s}$

A modified frequentist CL$_s$ method [103, 104, 105], as introduced by CMS and ATLAS, combines frequentist and Bayesian features. The CL$_s$ technique modifies certain features, which appear in searches with a small signal on top of a large background, where overestimating the background can yield small upper limits on the signal cross section. Moreover, it avoids false exclusions when the experiment has little sensitivity to signal. This effect is mitigated in the CL$_s$ construction by considering the compatibility of observation with the background-only hypothesis. Additionally, the CL$_s$ approach provides a means of setting upper limits on cross sections and masses derived from theoretical models where the possible range of the model parameters is constrained.

Estimates for both signal and background yields are subject to several uncertainties that are handled by introducing nuisance parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_N)$, where the signal and background expectations become functions of the nuisance parameters, i.e. s(θ) and b(θ), respectively. The nuisance parameters are modeled with probability density functions, typically using log-normal, flat, or Gaussian distributions, where $\bar{\theta}$ corresponds to the best estimate of the nuisance parameter, i.e. background/signal error values. We can define the binned likelihood function $\mathcal{L}(n|\mu, \theta)$, where n (i.e. *data*) represents either the experimental observation or pseudo-data used to construct sampling distributions, as follows,

$$\mathcal{L}(n|\mu, \theta) = \mathcal{P}_{poiss}(n|\mu \cdot s(\theta) + b(\theta)) \cdot p(\bar{\theta}|\theta). \tag{9.6}$$

The likelihood function includes the *pdf* terms, denoted by $p(\bar{\theta}|\theta)$, constraining the nuisance parameter θ associated with systematic uncertainties. Conventionally, a log-normal probability distribution is chosen for $p(\bar{\theta}|\theta)$, referred to as the auxiliary measurement. In the case of signal plus background, the Poisson distribution corresponds to the product of Poisson probabilities over all N channels, with observed events n_i, and expected events

$\mu \cdot s_i \ + \ b_i$,

$$\mathcal{P}_{poiss}(n|s+b) = \prod_{i}^{N} \frac{(\mu \cdot s_i + b_i)^{n_i} e^{-(\mu \cdot s_i + b_i)}}{n_i!}. \tag{9.7}$$

The signal estimation, s_i, depends on the expected signal cross section, branching fractions, detection efficiency for the signal, and integrated luminosity. The number of observed events for a given channel i is n_i.

For the case of the background-only hypothesis, the probability distributions are given by,

$$\mathcal{P}_{poiss}(n|b) = \prod_{i}^{N} \frac{b_i^{n_i} e^{-b_i}}{n_i!}. \tag{9.8}$$

The background estimates b_i depend on the SM background cross sections, the integrated luminosity, and selection efficiencies. For a multichannel counting experiment the full likelihood function is given by the product of the individual likelihoods for each channel, i.e. $\mathcal{L}(\mu,\theta) = \prod_i \mathcal{L}_i(\mu,\theta_i)$.

In order to determine the compatibility of data with the H_{s+b} and the H_b hypotheses, we construct a test statistics q_μ based on the ratio of the two likelihood functions. The test statistics is defined as,

$$q_\mu = -2\ln\frac{\mathcal{L}(\mathrm{n}|\mu,\hat{\theta}_\mu)}{\mathcal{L}(\mathrm{n}|\hat{\mu},\hat{\theta})}, \ \text{ with } 0 \le \hat{\mu} \le \mu. \tag{9.9}$$

where θ represents the nuisance parameters, $\hat{\theta}_\mu$ correspond to the maximal likelihood estimator (MLE), for the parameter θ for a specified value of the signal strength modifier μ given the data n, which is assumed to be Poisson distributed. The parameter estimator $\hat{\mu}$ and $\hat{\theta}$ represent the estimators of the parameters μ and θ, which maximize the likelihood, by fitting both θ and μ, given the observed data n. Two physics motivated conditions are imposed, the first is a lower bound constraint, $0 \le \hat{\mu}$, to ensure the signal rate is positive. The second is a upper bound constraint, $\hat{\mu} \le \mu$, which ensures that upward fluctuations of the data, such that $\hat{\mu} > \mu$ are not considered as evidence against the signal hypothesis, namely a signal with strength μ [106].

Following the "LHC-style" prescription, profiling is used, where the best estimator for the nuisance parameters, $\hat{\theta}$, are obtained by a fit to data. With the profiling method the nuisance parameters are calculated by performing a constrained maximum-likelihood fit [107]. An uncertainty on a nuisance parameter, e.g. luminosity, efficiency, background rate, cross section, can be in general described in the form of a probability density function. Gaussian pdf's are disfavored, since they are not well suited for positively defined observables (e.g. efficiency, cross section, and luminosity). The use of log-normal functions for modeling systematic uncertainties of non-statistical nature, is preferred.

The test statistic q_μ, can be used to distinguish between background-like and signal-like scenarios, and is found by numerical minimization of the negative log-likelihood using a general purpose maximizer, such as the Markov Chain Monte-Carlo method, in LandS, as discussed in Section 9.3.

We next summarize the steps to calculate the observed and expected limits, following closely [103].

9.2.2 Observed limit

The principal method for deriving exclusion limits is based on the $\mathrm{CL_s}$ method. For the purpose of this analysis we use an LHC-style prescription, which uses the profile likelihood test statistic, q_μ. The following steps are taken to calculate an observed limit on the signal strength μ.

1. Determine the observed value of the test statistics q_μ^{obs}, fort the fixed signal strength μ, that is under test.

2. Calculate estimator values for the nuisance parameters $\hat{\theta}_0^{obs}$ $(\mu = 0)$, and $\hat{\theta}_\mu^{obs}$, which

maximizes the likelihood for the H_b (background-only) and H_{s+b} (signal and background) hypothesis, respectively.

3. Following the frequentist approach, Monte Carlo pseudo-data, are generated in order to construct *pdf*s, i.e. $f(q_\mu|\mu,\hat{\theta}^{obs}_\mu)$ and $f(q_0|0,\hat{\theta}^{obs}_0)$, for the test statistics of the H_{s+b} ($\mu = 1$) and H_b ($\mu = 0$) hypotheses. A sampling distribution of the test statistics q_μ is shown in Figure 9.1 where,

    - q^{obs}_μ (black line), is calculated in step one.
    - $f(q_\mu|\mu,\hat{\theta}^{obs}_\mu)$ (in red), assuming fixed signal strength μ and corresponding best fit nuisance parameter $\hat{\theta}^{obs}_\mu$, given the observed data.
    - $f(q_\mu|\mu = 0,\hat{\theta}^{obs}_{\mu=0})$ (in blue), for background only hypothesis H_b and corresponding best fit nuisance parameter $\hat{\theta}^{obs}_0$, given the observed data.

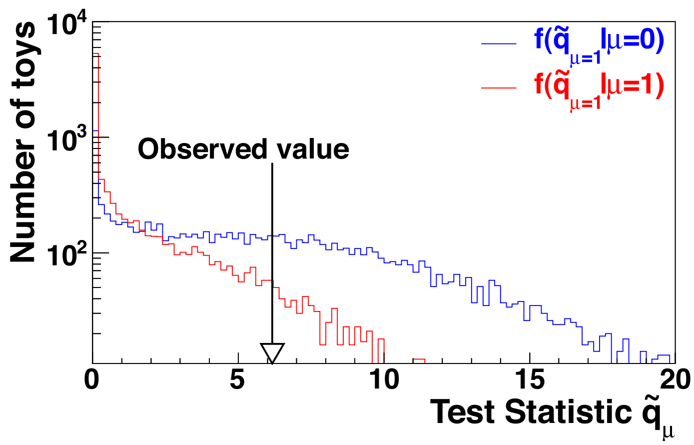


Figure 9.1: Test statistic distribution for a set of generated pseudo-data for *signal + background* (red) and *background-only* (blue) hypothesis [103].

4. From these *pdf*s, $f(q_\mu|\mu,\theta^{obs}_\mu)$ and $f(q_\mu|0,\theta^{obs}_0)$, we define p-values as the integrals,

$$p_\mu = \mathcal{P}(q_\mu \geq q^{\text{obs}}_\mu|H_{s+b}) = \int_{q^{obs}_\mu}^{\infty} \mathrm{f}(\mathrm{q}_\mu|\mu,\hat{\theta}^{\text{obs}}_\mu)\,\mathrm{d}q_\mu. \tag{9.10}$$

$$1 - p_b = \mathcal{P}(q_\mu \geq q_\mu^{\rm obs}|H_b) = \int_{q_0^{obs}}^{\infty} \mathrm{f}(\mathrm{q}_\mu|0, \hat{\theta}_0^{\rm obs})\,\mathrm{d}q_\mu. \tag{9.11}$$

where p_μ corresponds to the probability to get a results as or less compatible with the signal-plus-background hypothesis, while $p_{\rm b}$ is the probability to obtain a result as or less compatible with the background-only hypothesis, than the observed data.

5. The CL$_s$ upper limit, defined as the ratio between the two previous probabilities, at the (1-α)% confidence level is the value of μ for which,

$$\mathrm{CL_s}(\mu, \mathrm{q}_\mu^{\rm obs}) = \frac{\mathrm{p}_\mu}{1 - \mathrm{p_b}} = \frac{\mathrm{CL_{s+b}}(\mu, \mathrm{q}_\mu^{\rm obs})}{\mathrm{CL_b}(\mathrm{q}_\mu^{\rm obs})} \leq \alpha, \tag{9.12}$$

The CL$_{s+b}$ value corresponds to the probability to observe a data set with true signal μ, with a test statistic value equal to or larger than q_μ, and CL$_b$ is the corresponding probability without signal ($\mu = 0$). The denominator CL$_b$ servers to prevent exclusion if there is low sensitivity to signal. The pseudo datasets are generated according to the statistical model in equations 9.1 and 9.2, where the test statistic value $q_\mu(n)$ is calculated for each pseudo data set, n, to get the empirical probability distribution function for the underlying test statistic distribution, such as those shown in Figure 9.1. Each search channel is treated as statistically independent when the combined exclusion limits are calculated.

6. The signal model is excluded at the (1-α)% confidence level if $\mu = 1$ and CL$_s \leq \alpha$, where α is typically chosen to have a value of 0.05. In this analysis we quote the 95% confidence level limit on the theory cross section multiplied by the branching fraction, which corresponds to $\mu^{95\%}$ (i.e. $\mu^{95\%} = \sigma^{95\%}/\sigma_{predicted}$) times $\sigma_{predicted} \times BF$[1].

When the signal strength μ equals zero, it is expected that CL$_{s+b} \leq 0.05$, meaning that 5% of all the searches will result in excluding a signal strength of zero. The CL$_s$ method

[1] BF refers to the product of branching fractions for each of the two b' quarks

accounts for the fact that what is observed is a downward fluctuation in the background that causes a deficit in the observed number of events, that is inconsistent with the expected background. This can lead to the H_{s+b} hypothesis being excluded event when the expected signal is quite small, such as when there is no real experimental sensitivity [108]. The CL_s confidence level is designed to regulate against this behavior for CL_{s+b}, which is handled by the denominator value CL_b.

9.2.3 Expected limit

The background-only hypothesis can be used to determine the expected sensitivity for a new physics model. This is accomplished by generating pseudo-data based on expected background yields, which are then treated as if they were the real data. Following the prescription given in Section 9.2.1, we can calculate CL_s values for each test statistic. The $\mu^{95\%}$ value is calculated for each of the background-only pseudo-data experiments and a cumulative probability distribution of these $\mu^{95\%}$ values is generated. The point for which the cumulative probability distribution crosses the 50%-quantile corresponds to the median expected value, as illustrated in Figure 9.2. We can obtain with the same procedure the $\pm 1\sigma$ (68%) and $\pm 2\sigma$ (95%) uncertainty bands on the expected limit as defined by the crossing of the 16%/84% and 2.5%/97.5% quantiles of the cumulative distribution [103].

9.3 The LandS framework

If no significant deviation from the predicted SM background is found, we can proceed to set limits on the cross section for new physics processes. This is accomplished by calculating the maximal number of signal events, known as the "upper limit", for which the observed number of events is statistically consistent.

For this analysis we compute the exclusion limit based on the modified frequentist CL_s method as previously described in Section 9.2.1. We use the LandS [109] software to compute the 95% confidence level limits using the LHC-style CLs prescription. This computation

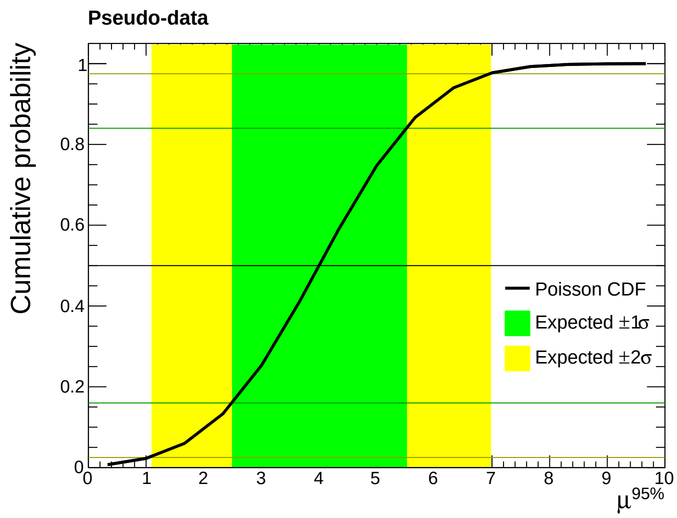


Figure 9.2: Cumulative probability distribution with 2.5%, 16%, 50%, 84%, and 97.5% quantiles (horizontal lines), which defines the median expected limit including the $\pm 1\sigma$ (68%) and $\pm 2\sigma$ (95%) bands for expected value of μ (*background-only hypothesis*).

yields the observed limit as well as the expected limit with one- and two-sigma uncertainty bands. As input LandS takes the combination of observed events, background estimation, and expected signal yield for each selected channel, including the uncertainties associated with the analysis. For each channel, nuisance parameters are defined to describe the effect of systematic uncertainties on the signal and background yields, as well as the statistical uncertainties on both yields. We use log-normal constraints on the nuisance parameters for the signal and background statistical and systematic uncertainties. While systematic uncertainties in many cases are correlated across channels, statistical uncertainties are not. All sources of systematic uncertainties are either fully correlated (either positive or negative) or uncorrelated (independent). Examples of nuisance parameters are the luminosity uncertainty, trigger efficiency uncertainty and others that were discussed in Section 8.1.

Given the large number of channels, the various search regions will not contribute equally for every particular signal scenario in the signal parameter space. In order to reduce the computational resources required, a combined limit is calculated using only the channels expected to have the highest sensitivity. Therefore, channels with no signal expectation are removed from the calculation. The channels are added in decreasing order of sensitivity until 90% of the expected signal yield is included in the limit calculation. The sensitivity of a given channel is determined by looking at its expected signal strength given by $\frac{\sigma_{95\%}}{\sigma_{predicted}}$. The discarded channels contain 10% of the signal, but with large SM backgrounds, hence, there is a large computational gain for a minimal gain on sensitivity. The expected exclusion limits obtained are consistent whether the fraction of expected signal yield included is chosen to be 90% or 95%. Even though about $30 - 40$ channels end up being used in the limit computation, roughly only the top 15 most sensitive channels contribute significantly to the results, the other channels are kept to conservatively make sure the calculation remains stable.

Chapter 10

Results

A detailed discussion of the statistical methods used in this analysis was presented in Section 9.1. We will discuss in this chapter the experimental results and the exclusion limits placed on the pair production of new vector-like b' quarks.

10.1 Multilepton results and interpretation

In this section, we interpret the results of the searches using multilepton final states. Events with at least three leptons are selected, including up to one hadronic tau candidate. These events are categorized into multiple exclusive signal regions based on the number and flavor of the leptons, the presence or absence of an opposite-sign, same-flavor lepton pair and its invariant mass, the presence or absence of b-tagged jets, and the S_T range. Tables 10.1 and 10.2 show the number of observed and total expected SM background events for the three and four lepton exclusive search channels, respectively. There are 66 signal channels (combinations of b' decaying to a b and either a Z, W, or H boson) and over one hundred search channels. The results in this chapter are based on 19.5 fb^{-1} of CMS data at $\sqrt{s}=$ 8 TeV.

Backgrounds in this analysis, such as dilepton $t\bar{t}$ events with fake leptons are, as previously mentioned, estimated from simulation, while additional sources of fake leptons are estimated using data-based methods. Backgrounds from WZ and ZZ diboson processes are estimated from simulation, with a correction to the E_T^{miss} resolution based on comparisons to data in control regions. Backgrounds estimated using MC simulation are corrected for trigger efficiency, lepton efficiency, pileup, and b-jet efficiency. We also observe contributions from rare SM processes like $t\bar{t}W$ and $t\bar{t}Z$, which are estimated from simulation as well.

The experimental results for a few selected multilepton channels are presented in Figures 10.1 and 10.2. For a complete list of results, see Appendix A. The three plots in Figure 10.1 show the S_T distribution of data compared with the background prediction for events that contain three light leptons (no taus), one OSSF pair not consistent with the Z boson mass, and one b-tagged jet. Each bin in the S_T distribution is considered to be a separate channel. Different decay modes of pair produced b' with a mass of 750 GeV are overlaid in each of the plots. From left to right b'b' $\rightarrow$ bZbZ, b'b' $\rightarrow$ tWtW, b'b' $\rightarrow$ bHbH are shown. Based on this example, we can clearly see that different search regions are more sensitive to different decay modes. Although, the b' $\rightarrow$ bH decay mode has the same signature as the tW decay, it is suppressed by a factor of 0.044, which comes from the squared branching fraction of 21% of the Higgs boson decaying to WW, which drastically decreases the expected signal yield. The signal yield for b' to bZ is low as well, since off-Z channels are not sensitive to this type of decay mode. The data yields in the signal regions are found to be broadly consistent with the expected SM backgrounds within uncertainties.

Figure 10.2 shows a different search region with four leptons (no taus), two OSSF pairs consistent with the Z mass, and one b-tagged jet. Again the distributions for pair produced b' with a mass of 750 GeV are overlaid for the three different decay modes. With the requirements placed on the reconstruction of two Z boson candidates this search channel has a good sensitivity to the b'b' $\rightarrow$ bZbZ decay mode.

N_{OSSF}	on- or off-Z	S_T (TeV)	$N_{\tau_h}=0$, $N_{b\text{-jets}}=0$ Obs.	Exp.	$N_{\tau_h}\geq 1$, $N_{b\text{-jets}}=0$ Obs.	Exp.	$N_{\tau_h}=0$, $N_{b\text{-jets}}\geq 1$ Obs.	Exp.	$N_{\tau_h}\geq 1$, $N_{b\text{-jets}}\geq 1$ Obs.	Exp.
0	-	>2.0	0	<0.02	0	0 ± 0.02	0	0 ± 0.02	0	0 ± 0.02
0	-	$1.5-2.0$	0	<0.02	0	0 ± 0.02	0	0 ± 0.02	0	0 ± 0.02
0	-	$1.0-1.5$	0	<0.02	0	0 ± 0.02	0	0 ± 0.02	0	0.007 ± 0.02
0	-	$0.6-1.0$	0	<0.02	0	0.12 ± 0.11	0	0.05 ± 0.05	†0	0.12 ± 0.1
0	-	$0.3-0.6$	0	0.09 ± 0.06	1	0.5 ± 0.19	0	0.001 ± 0.02	0	0.28 ± 0.12
0	-	$0-0.3$	0	0.05 ± 0.05	2	1.1 ± 0.45	0	0.0003 ± 0.02	0	0.25 ± 0.16
1	offZ	>2.0	0	<0.02	0	0 ± 0.02	0	0 ± 0.02	0	0 ± 0.02
1	onZ	>2.0	0	<0.02	0	0 ± 0.02	0	0 ± 0.02	0	0 ± 0.02
1	offZ	$1.5-2.0$	0	<0.02	0	0.007 ± 0.02	0	0 ± 0.02	0	0 ± 0.02
1	onZ	$1.5-2.0$	0	<0.02	0	0.02 ± 0.03	0	0.01 ± 0.03	0	0.007 ± 0.02
1	offZ	$1.0-1.5$	0	0.002 ± 0.02	0	0.12 ± 0.07	†0	0.03 ± 0.04	0	0.02 ± 0.02
1	onZ	$1.0-1.5$	1	0.06 ± 0.06	0	0.1 ± 0.07	0	0.11 ± 0.08	0	0.04 ± 0.04
1	offZ	$0.6-1.0$	0	0.06 ± 0.04	2	0.48 ± 0.17	0	0.06 ± 0.07	†0	0.3 ± 0.13
1	onZ	$0.6-1.0$	0	0.43 ± 0.15	0	1.7 ± 0.6	0	0.5 ± 0.29	†0	0.7 ± 0.33
1	offZ	$0.3-0.6$	0	0.27 ± 0.11	4	2.1 ± 0.5	0	0.33 ± 0.17	0	1.2 ± 0.43
1	onZ	$0.3-0.6$	5	1.8 ± 0.47	10	12 ± 3	2	1 ± 0.5	2	1.6 ± 0.5
1	offZ	$0-0.3$	2	0.48 ± 0.18	18	8.3 ± 2.1	0	0.04 ± 0.04	1	0.6 ± 0.3
1	onZ	$0-0.3$	2	3 ± 0.9	43	41 ± 10	2	0.07 ± 0.04	2	1 ± 0.4
2	offZ	>2.0	0	1e-05 $\pm$ 0.02	-	-	0	0 ± 0.02	-	-
2	onZ	>2.0	0	0.002 ± 0.02	-	-	0	0.02 ± 0.03	-	-
2	offZ	$1.5-2.0$	0	0.0002 ± 0.02	-	-	0	0 ± 0.02	-	-
2	onZ	$1.5-2.0$	0	0.05 ± 0.03	-	-	0	0.01 ± 0.02	-	-
2	offZ	$1.0-1.5$	0	0.01 ± 0.02	-	-	0	0 ± 0.02	-	-
2	onZ	$1.0-1.5$	1	0.6 ± 0.26	-	-	†0	0.1 ± 0.05	-	-
2	offZ	$0.6-1.0$	0	0.11 ± 0.04	-	-	0	0.14 ± 0.08	-	-
2	onZ	$0.6-1.0$	4	5.9 ± 2.0	-	-	1	1 ± 0.39	-	-
2	offZ	$0.3-0.6$	3	1 ± 0.3	-	-	1	0.22 ± 0.1	-	-
2	onZ	$0.3-0.6$	26	42 ± 10	-	-	4	3.2 ± 1	-	-
2	offZ	$0-0.3$	7	8.2 ± 2.3	-	-	0	0.18 ± 0.07	-	-
2	onZ	$0-0.3$	*135	122 ± 29	-	-	1	1 ± 0.26	-	-
Total4	All	All	186	187 ± 39	80	68 ± 15	11	8.3 ± 2.7	5	6.3 ± 1.6

Table 10.1: Observed (Obs.) yields for four lepton events from 19.5 fb^{-1} recorded in 2012. The channels are broken down by the number of and mass of any opposite-sign same-flavor pairs (whether on or off Z), whether the leptons include taus, whether there are any b jets present and the S_T. Expected (Exp.) yields are the sum of simulation and data-driven estimates of backgrounds in each channel. The channels are exclusive. Channels marked with an asterisk are used as control regions and are excluded from the limit calculations. Also, those channels with a dagger mark are used in the limit setting procedure and are representative of the top most sensitive channels for the b' decay with mass of 500 GeV where $\mathcal{B}$(b' $\rightarrow$ bH) = 100% [85].

N_{OSSF}	$m(\ell^+\ell^-)$	S_T	$N_{\tau_h}=0$, $N_{b\text{-jets}}=0$		$N_{\tau_h}=1$, $N_{b\text{-jets}}=0$		$N_{\tau_h}=0$, $N_{b\text{-jets}}\geq 1$		$N_{\tau_h}=1$, $N_{b\text{-jets}}\geq 1$	
	(GeV)	(TeV)	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
0	-	>2.0	0	<0.02	0	0.04 ± 0.05	0	0 ± 0.02	0	0 ± 0.22
0	-	$1.5-2.0$	0	0.07 ± 0.06	0	0.18 ± 0.19	0	0.05 ± 0.06	0	0.46 ± 0.28
0	-	$1.0-1.5$	0	0.21 ± 0.18	2	2.6 ± 1.2	0	0.36 ± 0.14	2	3.9 ± 2
0	-	$0.6-1.0$	†3	3.1 ± 1	†26	28 ± 12	2	4.9 ± 1.9	†46	58 ± 28
0	-	$0.3-0.6$	32	27 ± 10	289	290 ± 129	42	39 ± 17	410	480 ± 241
0	-	$0-0.3$	72	79 ± 22	1194	1324 ± 330	37	32 ± 15	316	331 ± 160
1	>105	>2.0	0	0.001 ± 0.02	0	0 ± 0.21	0	0 ± 0.03	0	0 ± 0.21
1	<75	>2.0	0	0.004 ± 0.02	0	0 ± 0.21	0	0.01 ± 0.04	0	0 ± 0.21
1	onZ	>2.0	0	0.2 ± 0.12	0	0.009 ± 0.21	0	0.04 ± 0.06	0	0.04 ± 0.05
1	>105	$1.5-2.0$	0	0.15 ± 0.09	0	0.22 ± 0.22	0	0.08 ± 0.05	0	0.2 ± 0.18
1	<75	$1.5-2.0$	1	0.11 ± 0.08	0	0.03 ± 0.05	0	0.07 ± 0.05	0	0.06 ± 0.07
1	onZ	$1.5-2.0$	3	1.1 ± 0.6	0	0.31 ± 0.17	1	0.28 ± 0.18	0	0.25 ± 0.12
1	>105	$1.0-1.5$	2	1 ± 0.4	1	1.3 ± 0.6	0	0.5 ± 0.22	1	2.1 ± 1.2
1	<75	$1.0-1.5$	0	1.1 ± 0.38	1	0.9 ± 0.44	†1	0.6 ± 0.27	0	1 ± 0.7
1	onZ	$1.0-1.5$	11	15 ± 6.9	9	5.9 ± 1.6	2	3.3 ± 1.2	1	1.7 ± 0.6
1	>105	$0.6-1.0$	13	10 ± 2.4	21	23 ± 7.2	†7	7.4 ± 2.4	23	28 ± 14
1	<75	$0.6-1.0$	14	10 ± 3.6	21	11 ± 3.4	†4	8.3 ± 2.6	†14	12 ± 6
1	onZ	$0.6-1.0$	106	111 ± 40	108	70 ± 17	†16	24 ± 7	17	17 ± 4.7
1	>105	$0.3-0.6$	63	65 ± 12	285	372 ± 96	36	35 ± 13	169	187 ± 94
1	<75	$0.3-0.6$	84	86 ± 21	290	279 ± 71	52	56 ± 22	167	171 ± 87
1	onZ	$0.3-0.6$	*669	735 ± 166	*2099	2705 ± 772	122	108 ± 24	325	284 ± 73
1	>105	$0-0.3$	180	195 ± 33	1620	1712 ± 482	17	17 ± 6.4	97	79 ± 35
1	<75	$0-0.3$	617	644 ± 102	10173	9211 ± 2694	62	74 ± 28	297	288 ± 97
1	onZ	$0-0.3$	*4255	4439 ± 691	*49916	49192 ± 14670	*140	149 ± 24	795	826 ± 229
Total3	All	All	6125	6430 ± 916	66055	65233 ± 19038	541	564 ± 150	2680	2774 ± 903

Table 10.2: Observed (Obs.) yields for three lepton events. The channels are broken down by the number of and mass of any opposite-sign, same-flavor pairs (whether on or off Z), whether the leptons include taus, whether there are any b jets present and the S_T. Expected (Exp.) yields are the sum of simulation and data-driven estimates of backgrounds in each channel. The channels are exclusive. Channels marked with an asterisk are used as control regions and are excluded from the limit calculations. Also, those channels marked with a dagger are a representative subset of the top most sensitive channels for the b' decay, with a mass of 500 GeV and $\mathcal{B}(\text{b'} \to \text{bH}) = 100\%$, which are used in the limit setting procedure [85].

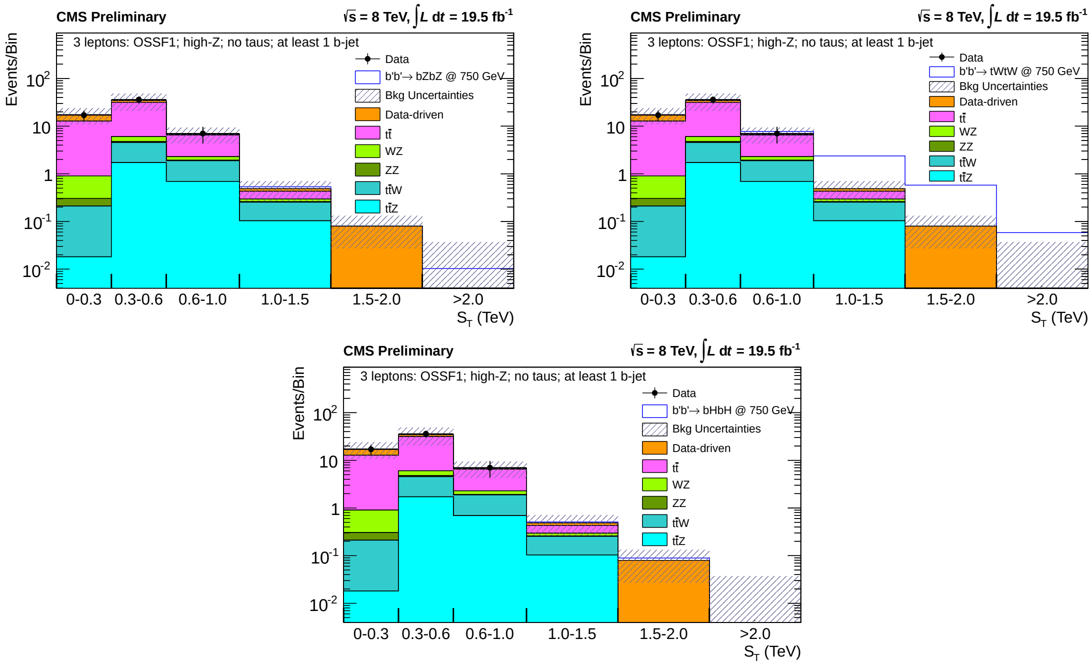


Figure 10.1: Channel with 3 leptons with none of them being a tau, such that they form 1 opposite-sign same-flavor pair having invariant mass above the Z-window, and at least 1 b jet [85].

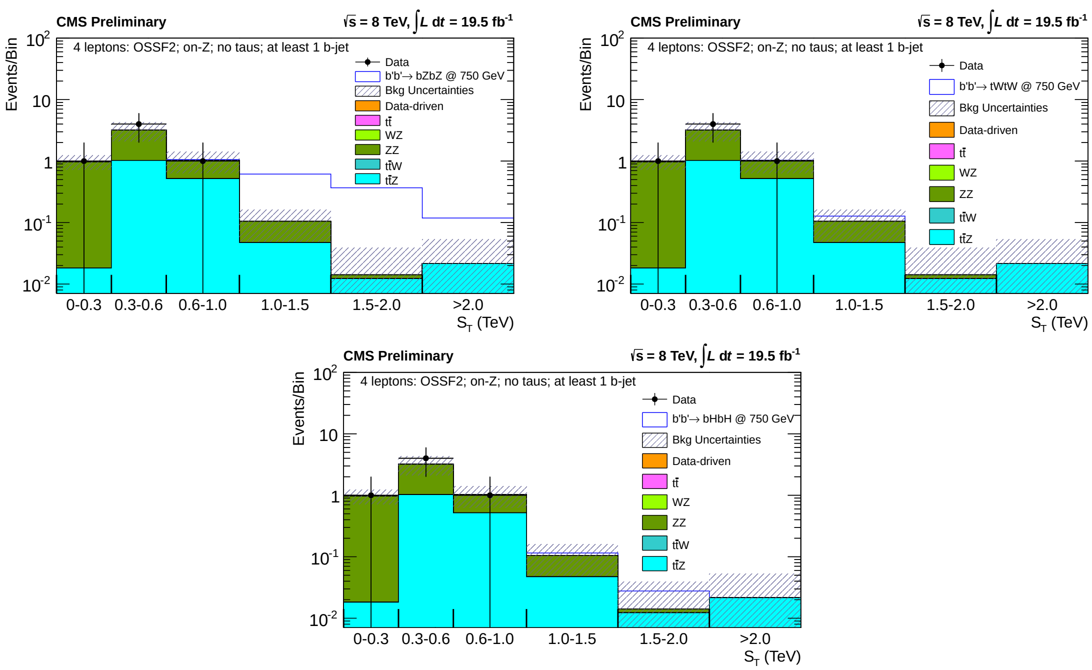


Figure 10.2: Channel with 4 leptons with none of them being a tau, such that they form 2 opposite-sign same-flavor pairs where at least one of them is on-Z and with at least 1 b jet [85].

10.2 Exclusion limits on exotic b' quarks

No significant excesses above the SM predictions are observed. We interpret the result in the context of a model involving the vector-like b' quark decaying to three different modes (b' $\rightarrow$ bZ, b' $\rightarrow$ tW, and b' $\rightarrow$ bH) as a function of the branching fractions. In order to determine the sensitivity for various branching fraction scenarios, we perform a simultaneous fit across all exclusive channels to compute the likelihood of observing a signal. We exclude b' with masses below where the theory line intersects with the observe limit curve. The theory curve corresponds to the theoretical cross section as a function of b' mass. We exclude the production of b' quarks at the 95% CL for b' masses as a function of branching fraction.

Figure 10.3 shows the expected and observed upper limits on the cross section times branching fraction ($\sigma \times B$) for the cases $\mathcal{B}$(b' $\rightarrow$ bZ) = 100%, $\mathcal{B}$(b'$\rightarrow$ tW) = 100%, and $\mathcal{B}$(b' $\rightarrow$ bH) = 100% with corresponding b' masses excluded up to 685, 785, and 520 GeV, respectively. The green and yellow shaded regions correspond to the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty bands on the expected limit. This is one of the first result showing exclusions in the $\mathcal{B}$(b' $\rightarrow$ bH) channel using multilepton final states.

Figure 10.4 shows the expected and observed mass limits for a varying branching fraction of $\mathcal{B}$(b' $\rightarrow$ tW) and $\mathcal{B}$(b' $\rightarrow$ bZ) assuming that $\mathcal{B}$(b' $\rightarrow$ bH) = 0%. The x-axis represents the b' mass and the y-axis shows $\mathcal{B}$(b' $\rightarrow$ bZ). All points to the left of the curve are excluded.

The expected exclusion curve in Figure 10.4 can be estimated from the branching fractions for a b' pair to decay into three and four leptons, as a function of $\mathcal{B}$(b'$\rightarrow$ bZ). Defining $\alpha \equiv \mathcal{B}$(b'$\rightarrow$ bZ) and assuming $\mathcal{B}$(b'$\rightarrow$ bH) to be zero, the the different branching fraction of a b' pair decay can be written as bZbZ = α^2, bZtW = 2α(1-α), and tWtW = $(1-\alpha)^2$, respectively. The branching fraction of $\mathcal{B}$(b'b' $\rightarrow \geq$ 3 leptons) is then given by,

$$f(\alpha) = (0.36\%)\alpha^2 + (2.65\%)2\alpha(1-\alpha) + (5.1\%)(1-\alpha)^2 \tag{10.1}$$

The first coefficient, as seen in table 10.3, is 0.36% and corresponds to the probability of bZbZ to decay into four leptons, while 2.65%, and 5.1% correspond to the branching fraction of bZtW and tWtW to decay into three or more leptons. The mass limit is expected to improve as α decreases, as seen in Equation 10.1. As bZ decay modes become dominant, the 3 lepton channels are suppressed and the limit worsens.

Decy mode	**4-lepton BR (%)**	**3-lepton BR (%)**	**Total (%)**
bZbZ	$(6\%)(6\%) = 0.36$	0	0.36
tWtW	$(25\%)^4 = 0.4$	$4(25\%)^3(75\%) = 4.7$	5.1
bHbH	$(21\%)^2(30\%)^3(70\%)$ $+2(2.5\%)(21\%)(10\%)^2(90\%)(30\%)^2 = 0.084$	$4(2.5\%)(10\%)^2(90\%)$ $+(21\%)^2(30\%)^4 = 0.13$	0.131

Table 10.3: Percentage yield of 3 and 4 lepton final states for various b'b' decay modes.

Figure 10.5 shows the exclusion curves for the benchmark branching fractions $\mathcal{B}$(b' $\rightarrow$ tW) = 50%, $\mathcal{B}$(b' $\rightarrow$ bH) = 25%, and $\mathcal{B}$(b' $\rightarrow$ bZ) = 25%. This benchmark is referred to as "democratic" decay modes, for which we calculate observed and expected exclusion limit of 694 GeV. The benchmark branching fractions relate to the asymptotic limit where the mass of the heavy vector-like quark goes to infinity (high mass limit), which is in agreement with what is expected from the Goldstone equivalence theorem [110].

Figure 10.6 shows the expected and observed curves respectively as a function of the branching fractions to b' $\rightarrow$ bZ, b' $\rightarrow$ tW, and b' $\rightarrow$ bH. The x-axis is the b' mass and the y-axis is $\mathcal{B}$(b' $\rightarrow$ bZ) similar to Figure 10.4. The different contours are curves of fixed $\mathcal{B}$(b' $\rightarrow$ bH). All points to the left of a given curve are excluded at the 95% CL. As $\mathcal{B}$(b' $\rightarrow$ bH) increases, the total acceptance into three- and four-leptons decreases which results in less sensitivity.

The full interpretation of the results taking into account all combinations of the various branching fractions is shown in Figure 10.7 for the observed limits (on the left) and expected limits (on the right). Table 10.4 shows the observed and expected limits at the 95% CL for several of these branching fraction combinations.

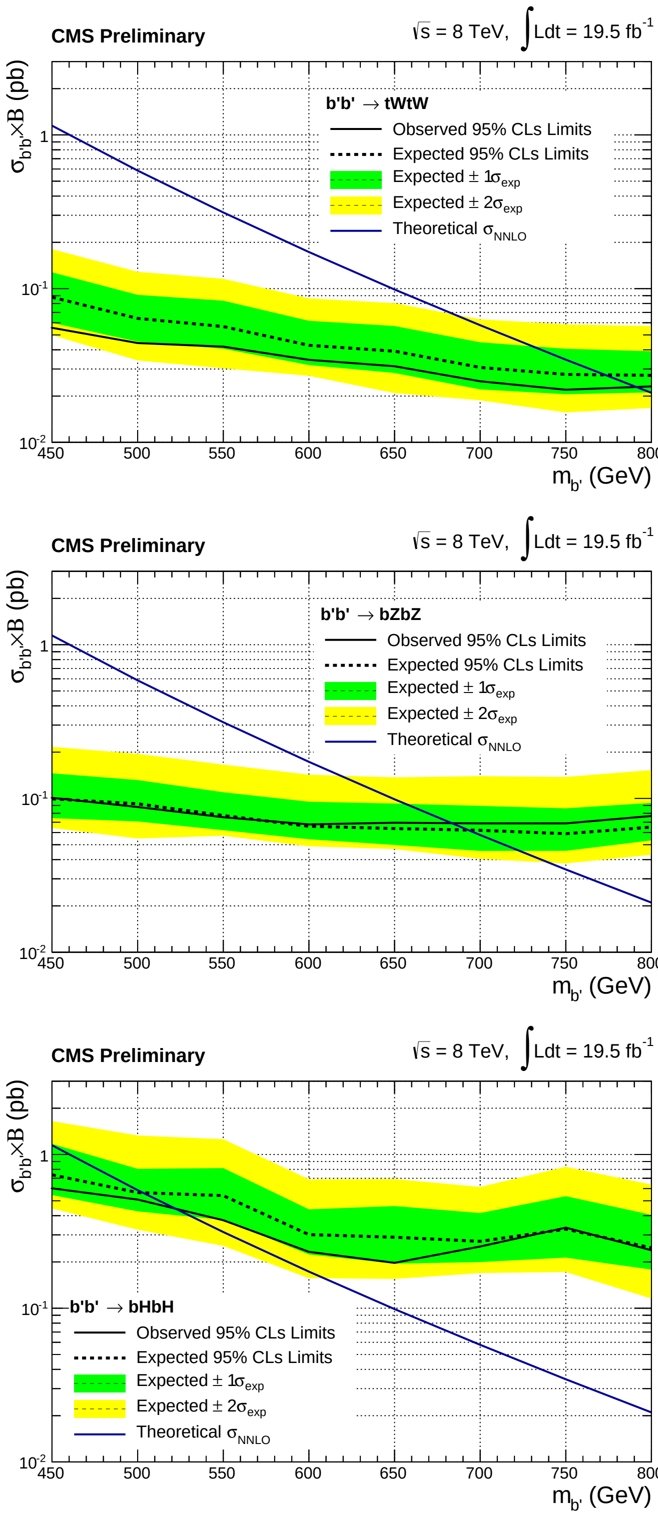


Figure 10.3: Cross section times branching fraction exclusion curves for a b' as a function of its mass for the decay modes b'b' $\rightarrow$ tWtW (top), b'b' $\rightarrow$ bZbZ (middle), and b'b' $\rightarrow$ bHbH (bottom). The figures show expected (dashed), observed (solid) exclusions, and theory (blue). The green and yellow bands correspond to the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties on the expected limit [85].

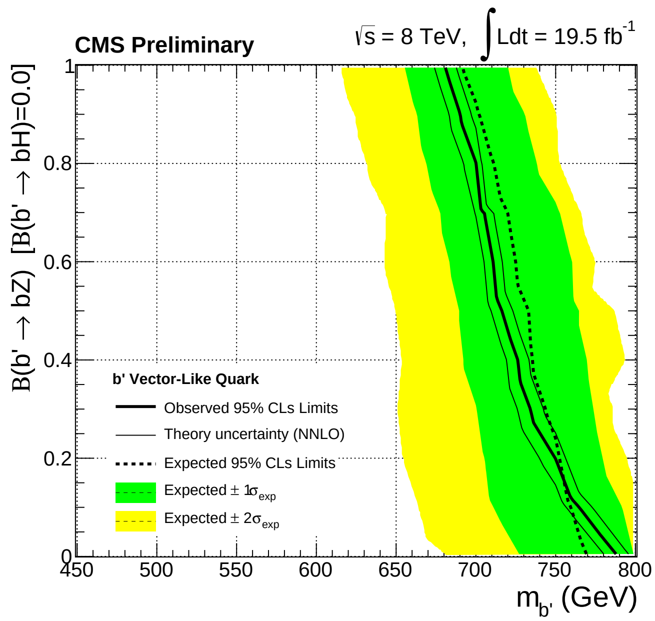


Figure 10.4: Exclusion limits for pair-produced b' quarks decaying into multilepton final states in the two-dimensional plane of the branching fraction of b' → bZ vs. b' mass.The branching fraction for b' → bH is set to zero. Points to the left of the curve are excluded. The y = 0 axis corresponds to b'b' → tWtW and the $y = 1$ axis to b'b' → bZbZ [85]. We apply a conservative 10% theory uncertainty.

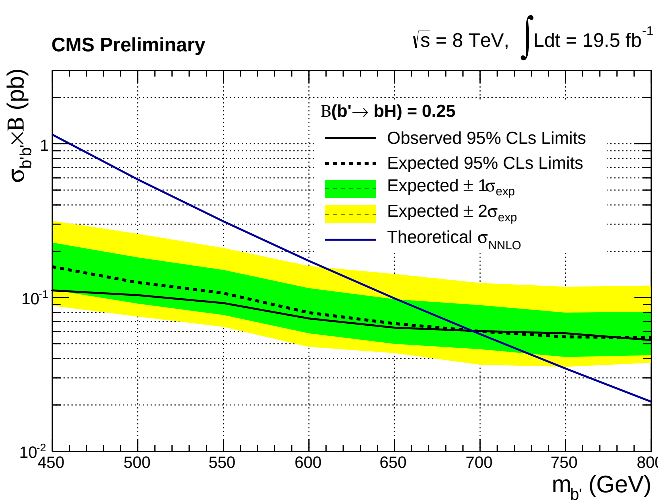


Figure 10.5: Observed and expected 95% CL upper limits for b' quark production cross section for branching fraction to tW, bH, and bZ of 50%, 25%, and 25%, respectively [85].

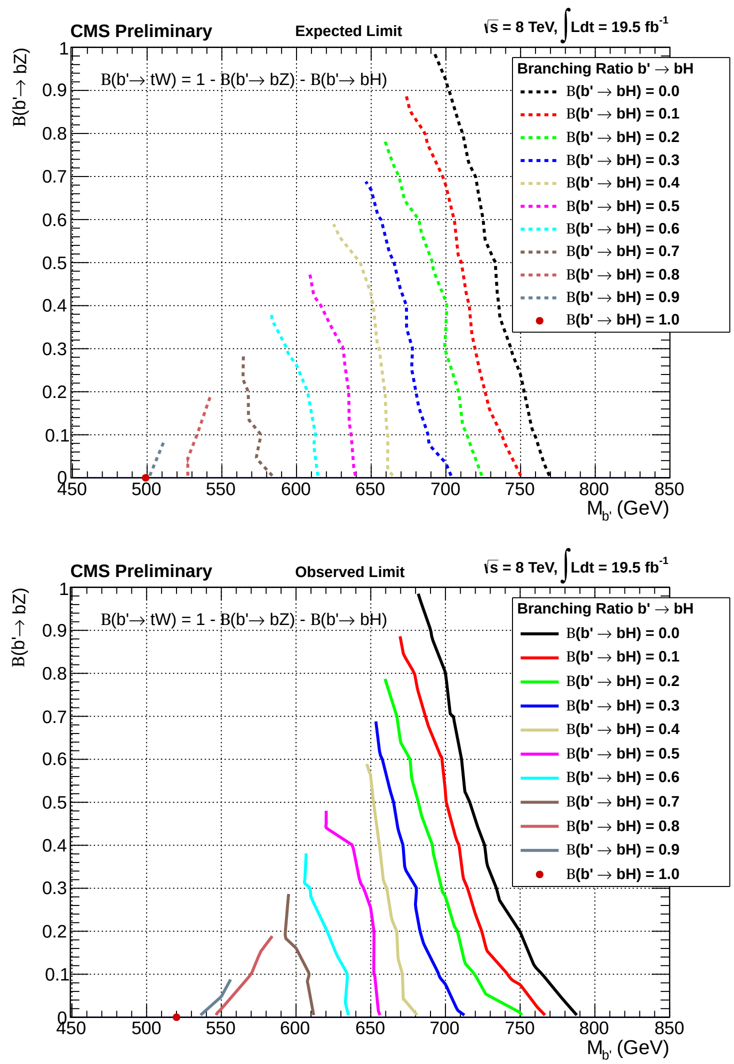


Figure 10.6: Expected (top) and observed (bottom) exclusion curves as a function of branching fractions. The $\mathcal{B}$(b' $\rightarrow$ bZ) is plotted as a function of the b' mass and the various curves represent fixed $\mathcal{B}$(b' $\rightarrow$ bH) from 0.0 (right most) to 1.0 (left most) [85].

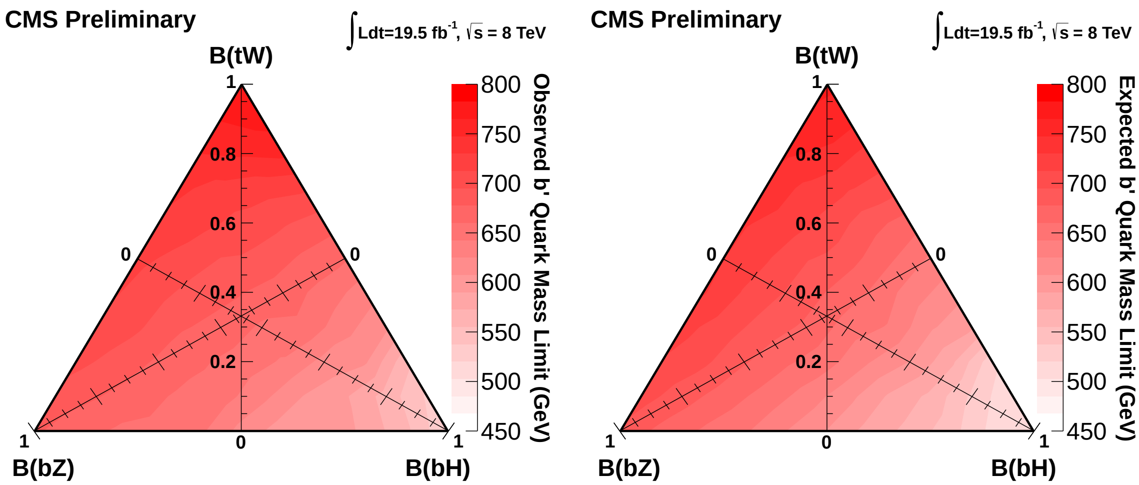


Figure 10.7: Observed (left) and expected (right) limits with varied branching fraction of tW, bH, and bZ in steps of 0.1. Each point on the triangle corresponds to a unique combination of the three branching fractions and the vertices represent a simplified models with 100% branching fraction into the three final states [85].

Comb #	$\mathcal{B}$(H→tW)	$\mathcal{B}$(H→bH)	$\mathcal{B}$(H→bZ)	Obs. (GeV)	Exp. (GeV)	$\pm 1\sigma$ (GeV)	$\pm 2\sigma$ (GeV)
(0)	0.5	0.3	0.3	694	692	[651,723]	[605,759]
(1)	0.0	0.0	1.0	680	691	[655,719]	[617,750]
(2)	0.0	0.1	0.9	666	672	[637,707]	[600,711]
(3)	0.0	0.2	0.8	657	657	[616,685]	[578,708]
(4)	0.0	0.3	0.7	654	645	[596,668]	[561,672]
(5)	0.0	0.4	0.6	646	625	[578,657]	[542,672]
(6)	0.0	0.5	0.5	618	606	[564,643]	[517,648]
(7)	0.0	0.6	0.4	598	583	[547,620]	[501,639]
(8)	0.0	0.7	0.3	605	566	[519,613]	[474,631]
(9)	0.0	0.8	0.2	584	543	[501,593]	[457,628]
(10)	0.0	0.9	0.1	557	514	[466,546]	[438,627]
(11)	0.0	1.0	0.0	520	502	[447,528]	[423,607]
(12)	0.1	0.0	0.9	688	703	[661,724]	[622,742]
(13)	0.1	0.1	0.8	678	686	[649,716]	[607,717]
(14)	0.1	0.2	0.7	667	669	[629,701]	[589,725]
(15)	0.1	0.3	0.6	657	653	[610,682]	[567,709]
(16)	0.1	0.4	0.5	650	635	[589,665]	[547,700]
(17)	0.1	0.5	0.4	629	613	[567,652]	[525,665]
(18)	0.1	0.6	0.3	606	587	[546,645]	[499,659]
(19)	0.1	0.7	0.2	592	568	[516,615]	[474,651]
(20)	0.1	0.8	0.1	565	533	[484,595]	[451,632]
(21)	0.1	0.9	0.0	534	503	[458,543]	[436,587]

Table 10.4: Sets of branching fraction values and the observed and expected 95% CL upper limits for the combined electron, muon, and tau channels [85].

10.3 Conclusion

In this dissertation, we carried out a search for physics beyond the SM which manifests itself in a variety of possible multilepton final states. It has been performed using a total integrated luminosity of 19.5 fb^{-1} of LHC data at a center-of-mass energy of $\sqrt{s} = 8$ TeV collected during 2012. We have estimated the background from SM processes using both MC simulations and data-based methods. We have performed a search for pair production of vector-like, b' quarks, in multilepton final states. We binned the data based on multiple exclusive channels arranged according to the amount of expected Standard Model background observed, in order to increase the search sensitivity for new signal. We see good agreement between observations and expectations. The search is interpreted as a function of b' mass depending on the branching fractions to bZ, tW, and bH states. We exclude b' quarks at the 95% confidence level with masses less than values in the range $520 - 785$ GeV, depending on the values of the branching fraction.

Appendix A

Additional results plots

The following plots show the distribution of observation, expected SM background, and signal yield as a function of S_T for various 3 and 4 lepton channels. These plots represent the case where the b' mass is 750 GeV and correspond to the results found in Ref. [85].

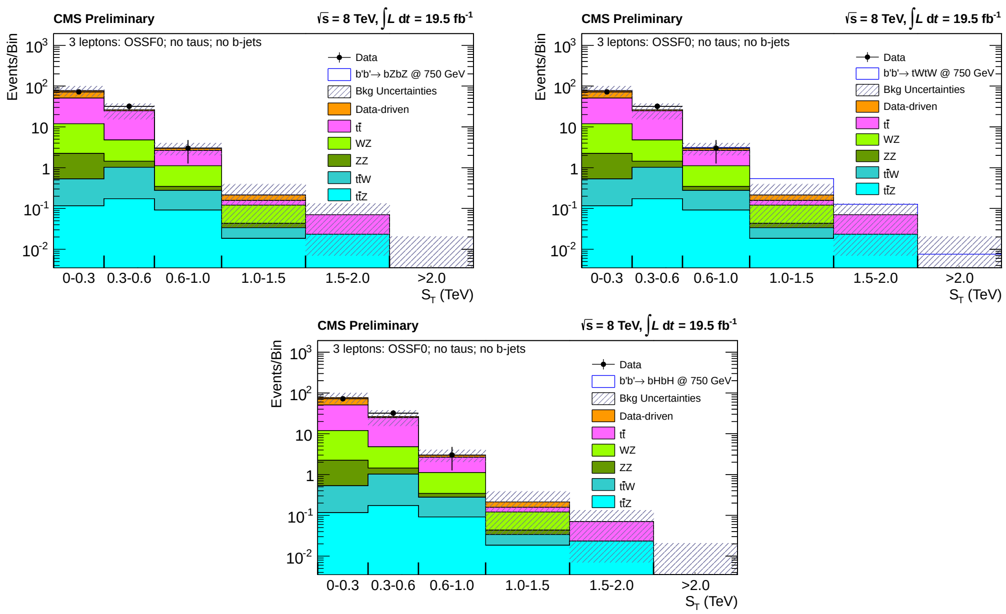


Figure A.1: 3-lepton + OSSF0 + on-Z + Tau0 + b0.

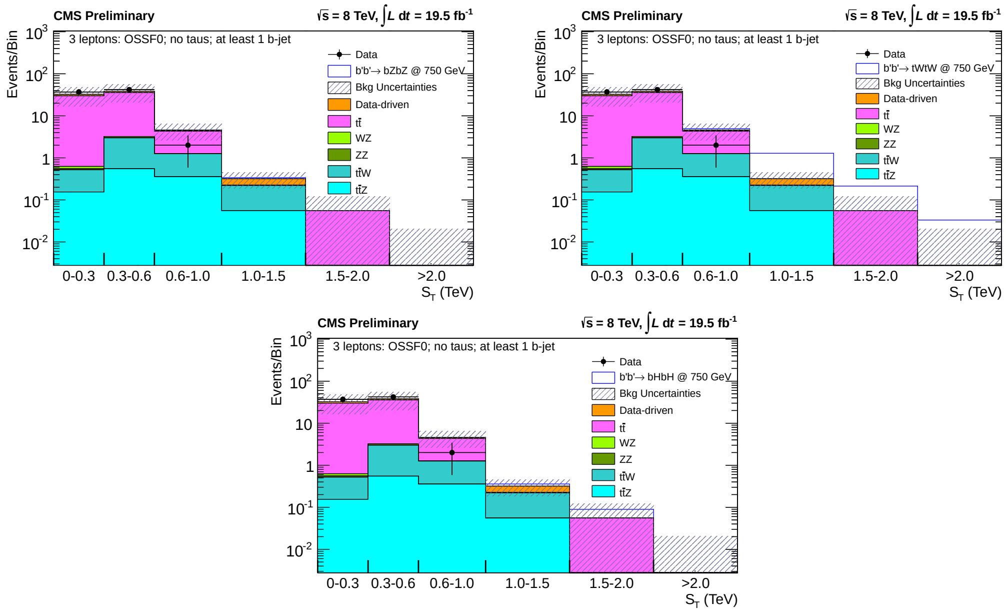


Figure A.2: 3-lepton + OSSF0 + on-Z + Tau0 + b1.

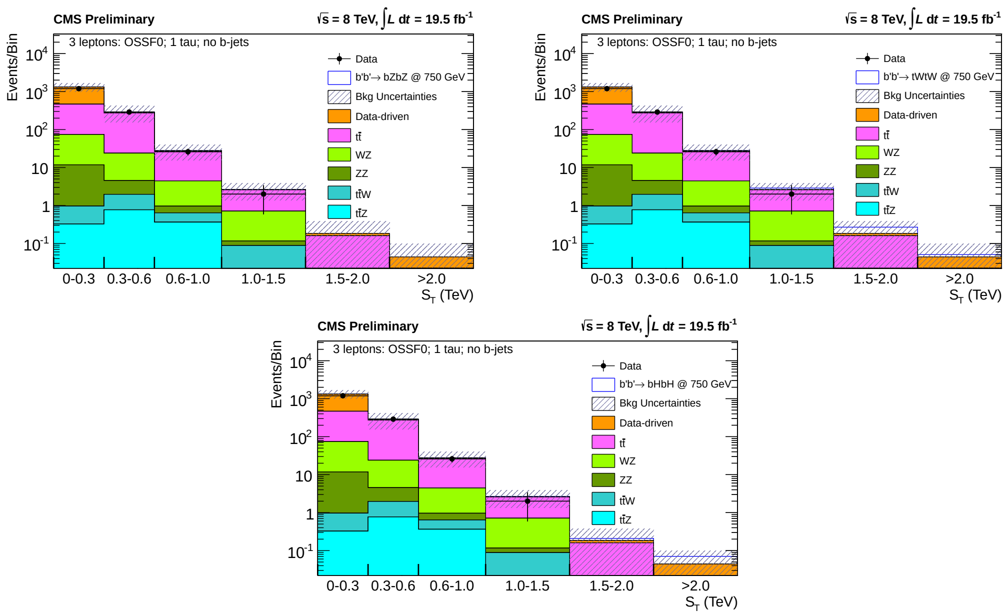


Figure A.3: 3-lepton + OSSF0 + on-Z + Tau1 + b0.

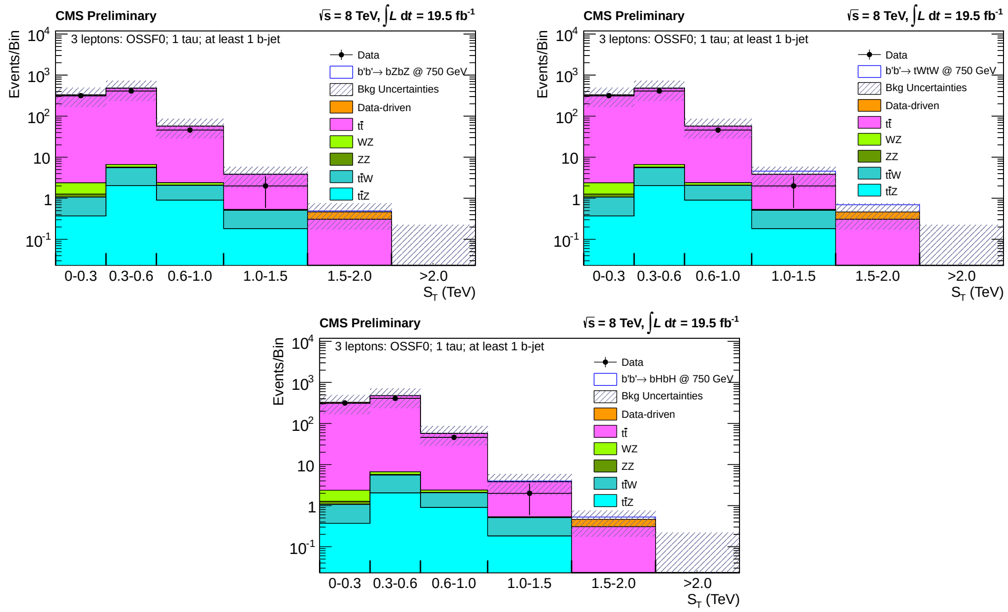


Figure A.4: 3-lepton + OSSF0 + on-Z + Tau1 + b1.

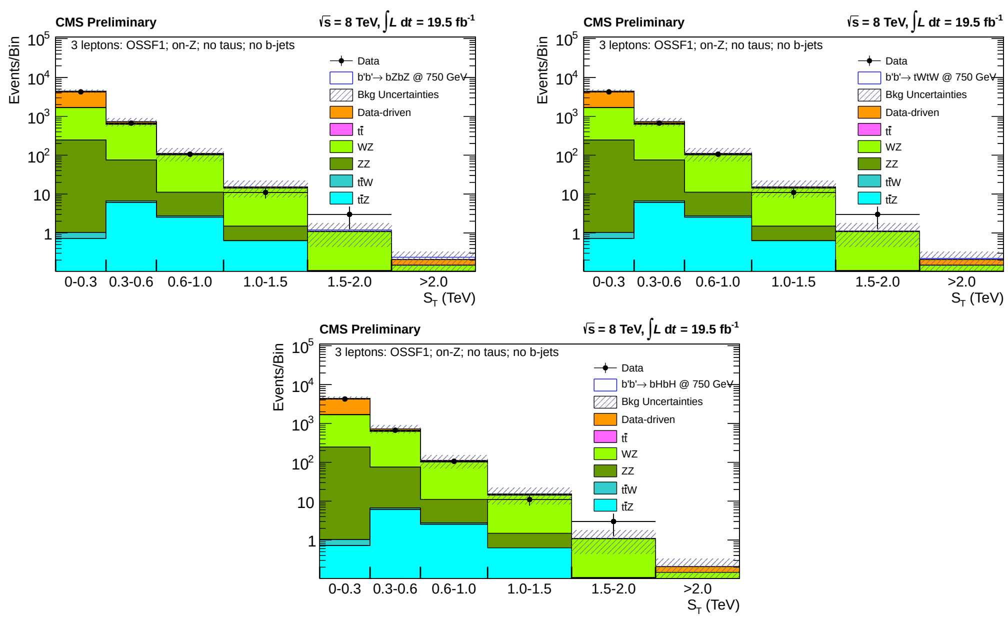


Figure A.5: 3-lepton + OSSF1 + on-Z + Tau0 + b0.

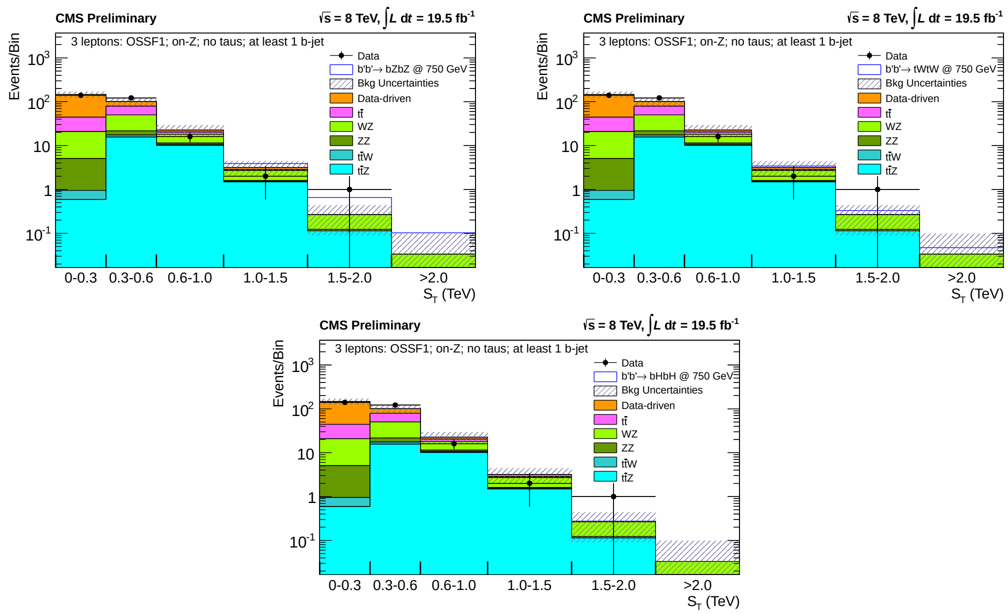


Figure A.6: 3-lepton + OSSF1 + on-Z + Tau0 + b1.

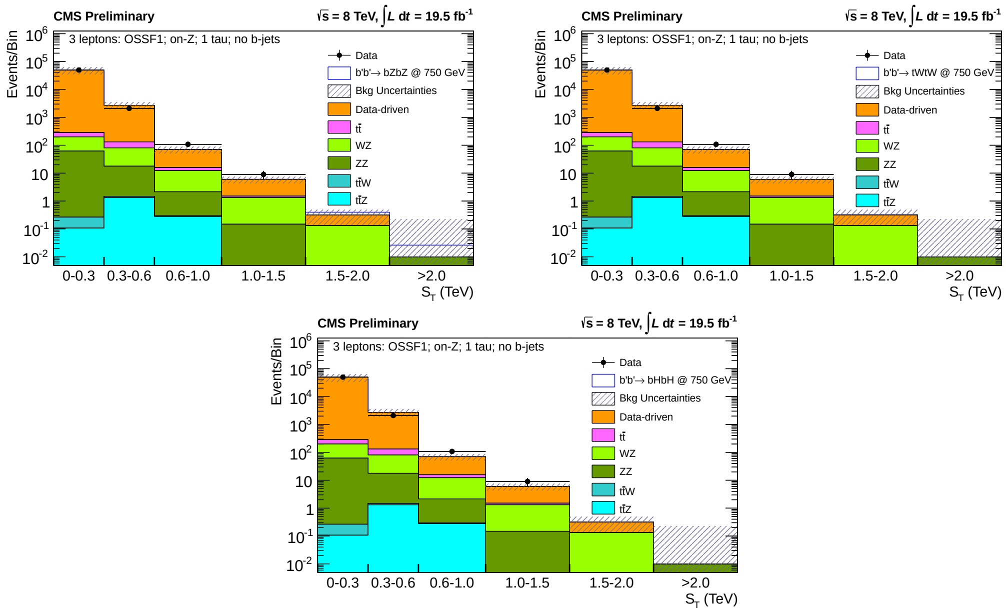


Figure A.7: 3-lepton + OSSF1 + on-Z + Tau1 + b0.

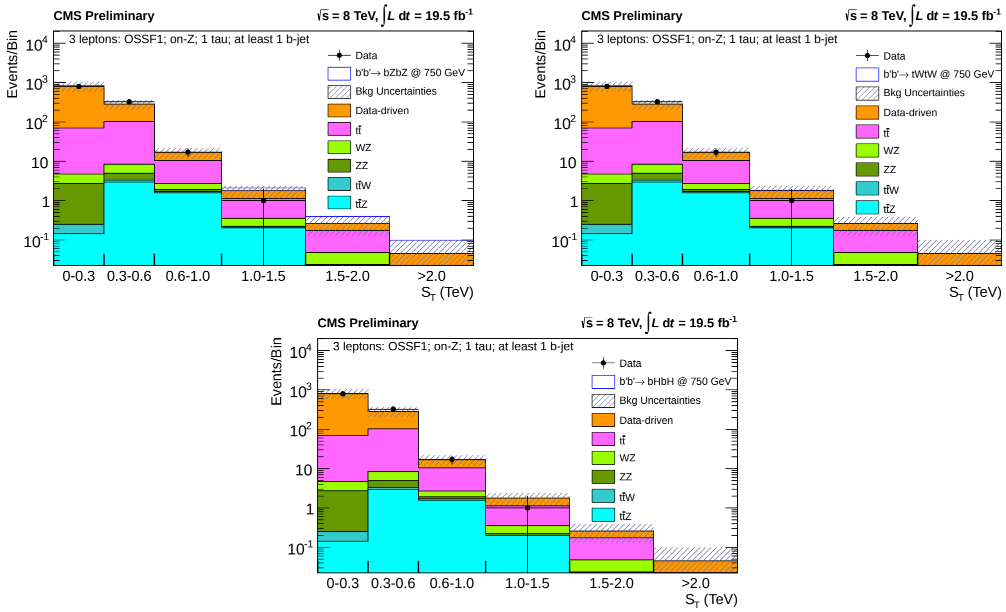


Figure A.8: 3-lepton + OSSF1 + on-Z + Tau1 + b1.

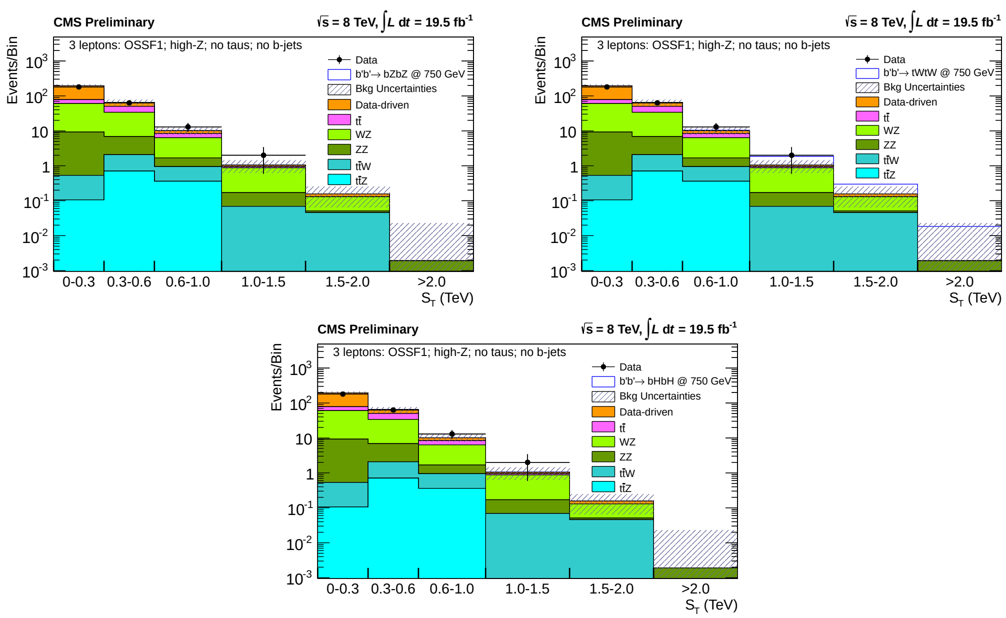


Figure A.9: 3-lepton + OSSF1 + above-Z + Tau0 + b0.

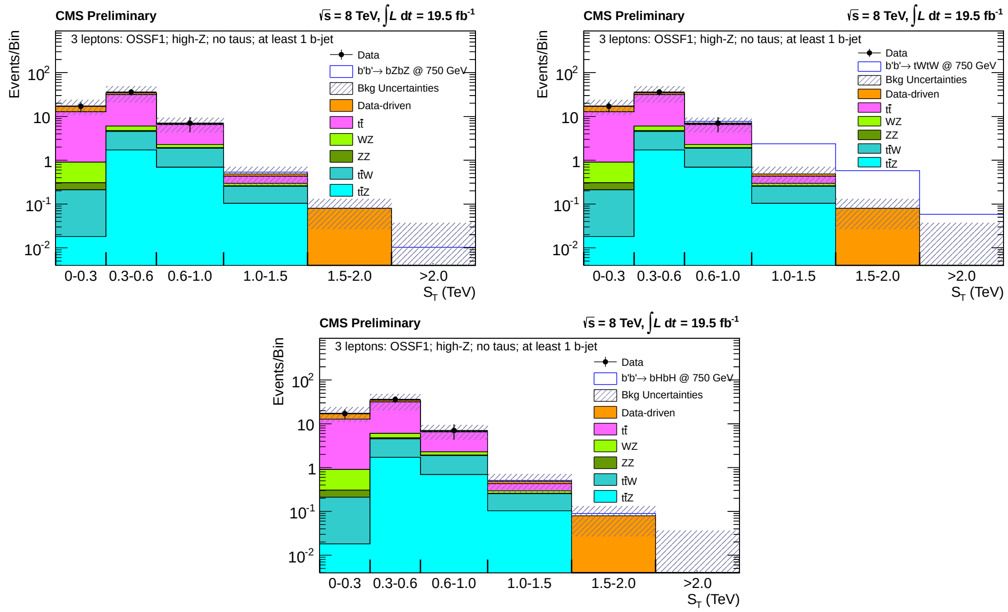


Figure A.10: 3-lepton + OSSF1 + above-Z + Tau0 + b1.

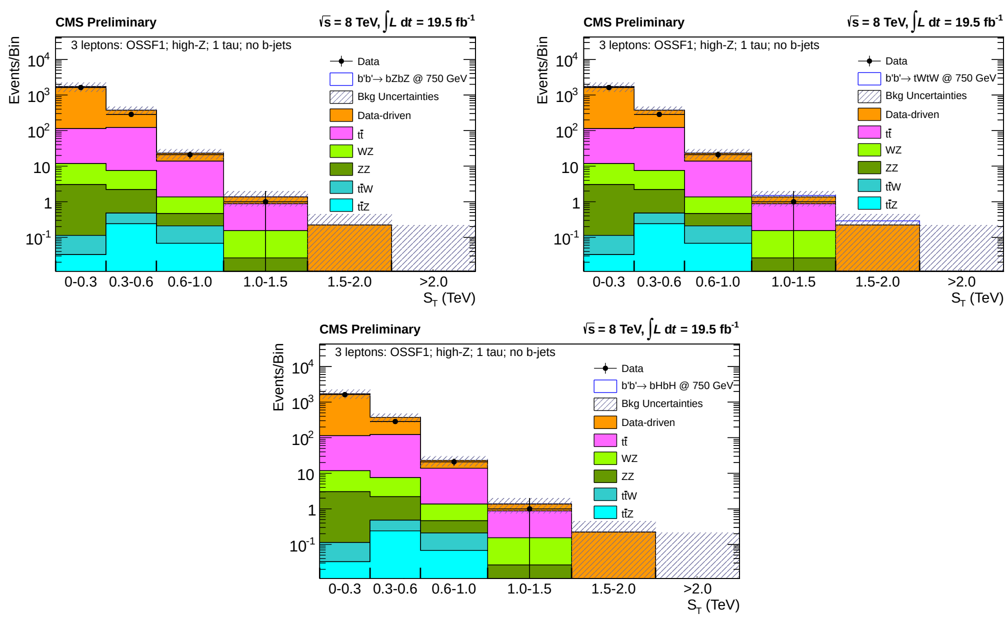


Figure A.11: 3-lepton + OSSF1 + above-Z + Tau1 + b0.

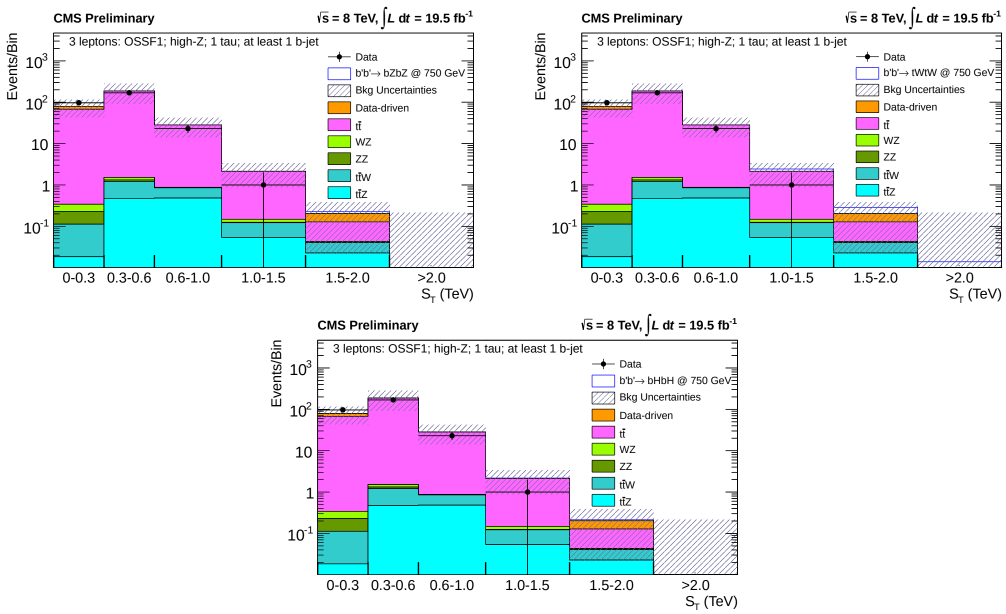


Figure A.12: 3-lepton + OSSF1 + above-Z + Tau1 + b1.

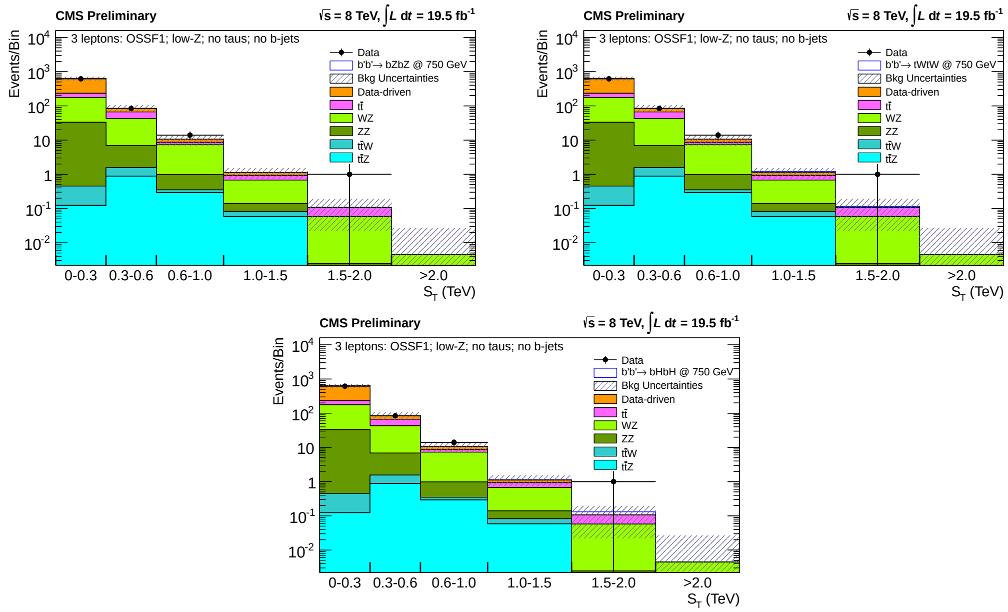


Figure A.13: 3-lepton + OSSF1 + below-Z + Tau0 + b0.

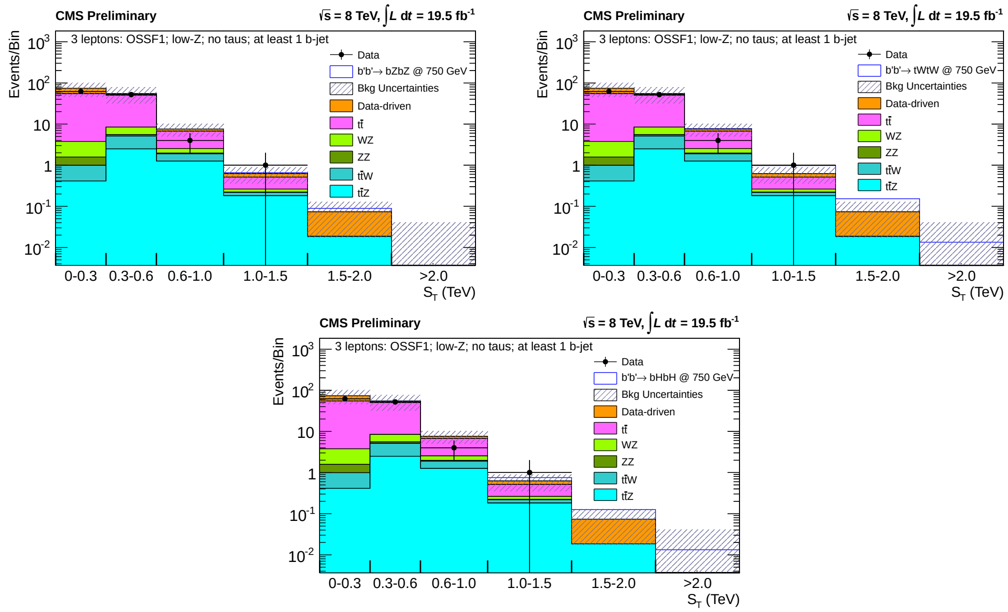


Figure A.14: 3-lepton + OSSF1 + below-Z + Tau0 + b1.

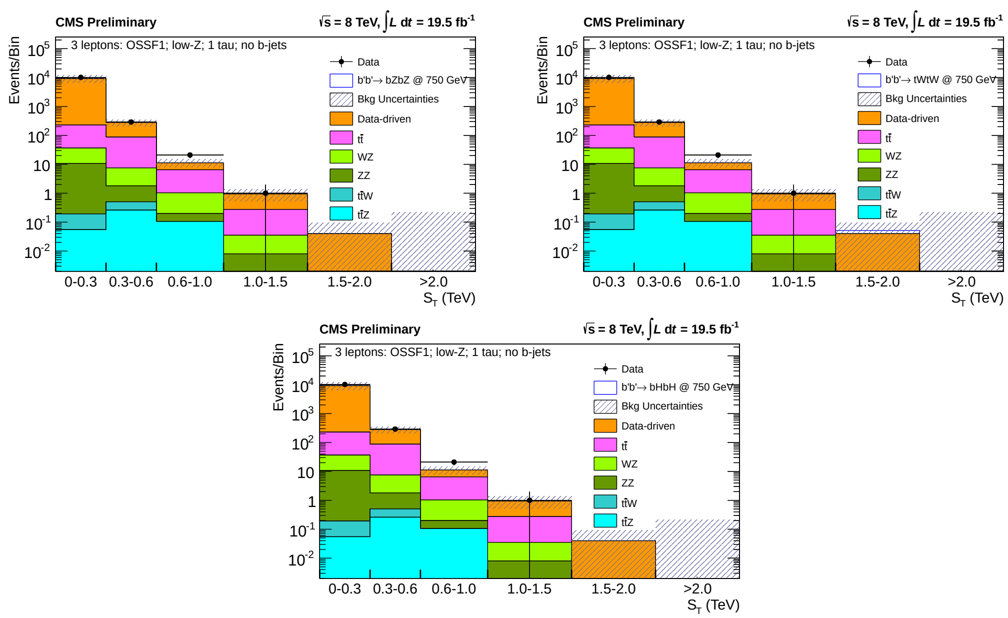


Figure A.15: 3-lepton + OSSF1 + below-Z + Tau1 + b0.

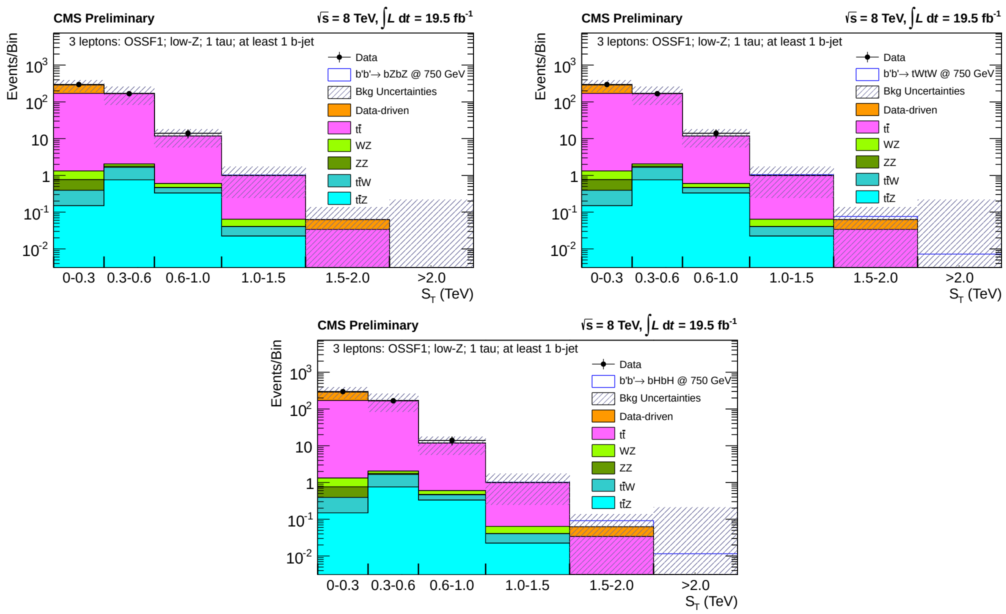


Figure A.16: 3-lepton + OSSF1 + below-Z + Tau1 + b1.

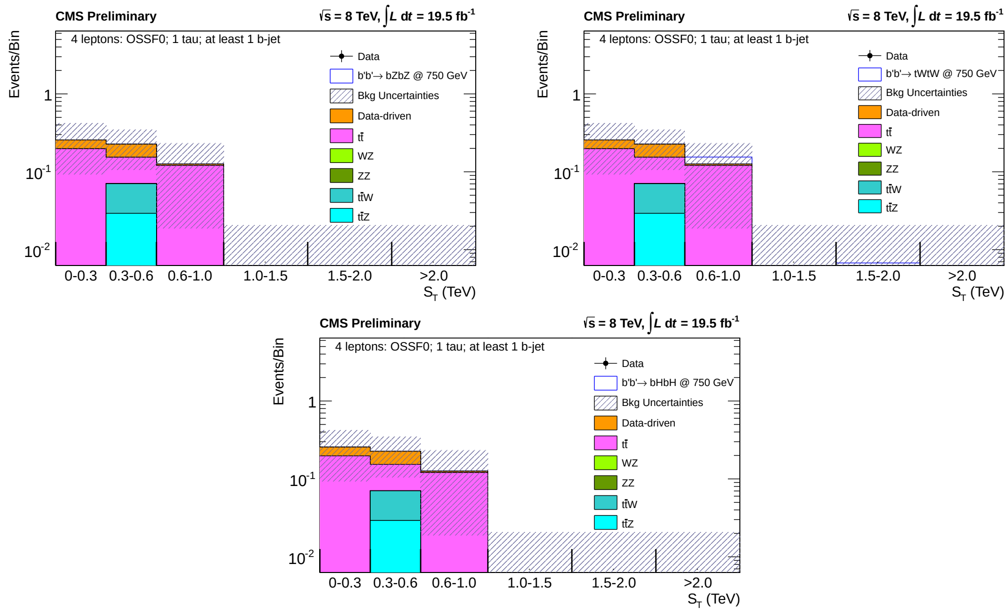


Figure A.17: 4-lepton + OSSF0 + on-Z + Tau1 + b1.

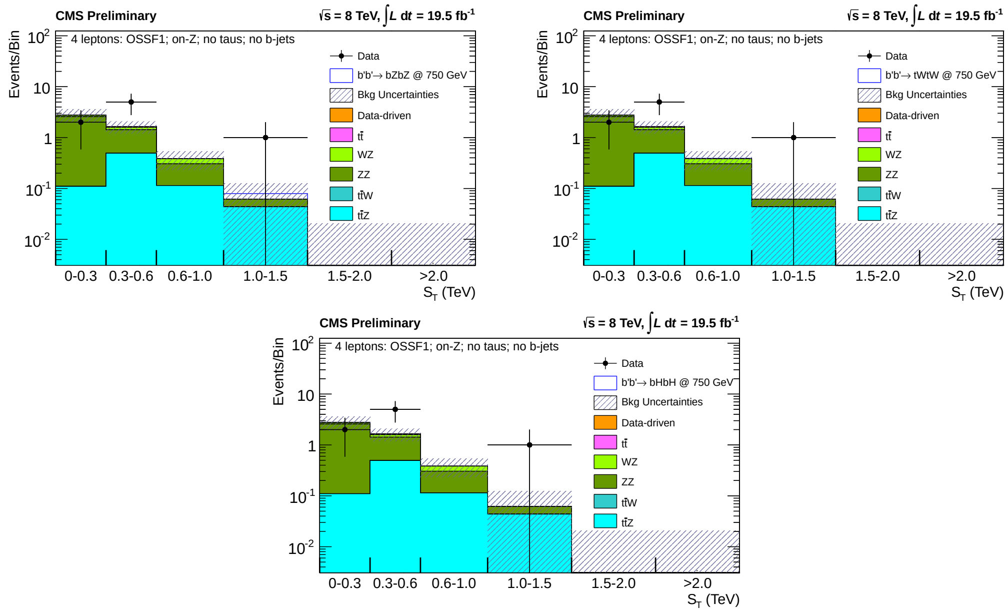


Figure A.18: 4-lepton + OSSF1 + on-Z + Tau0 + b0.

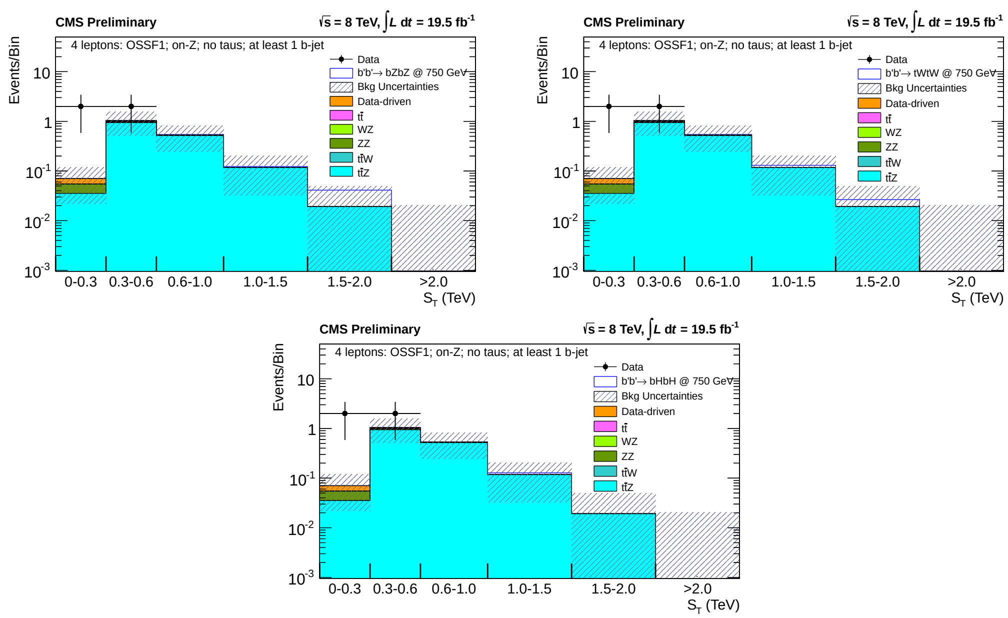


Figure A.19: 4-lepton + OSSF1 + on-Z + Tau0 + b1.

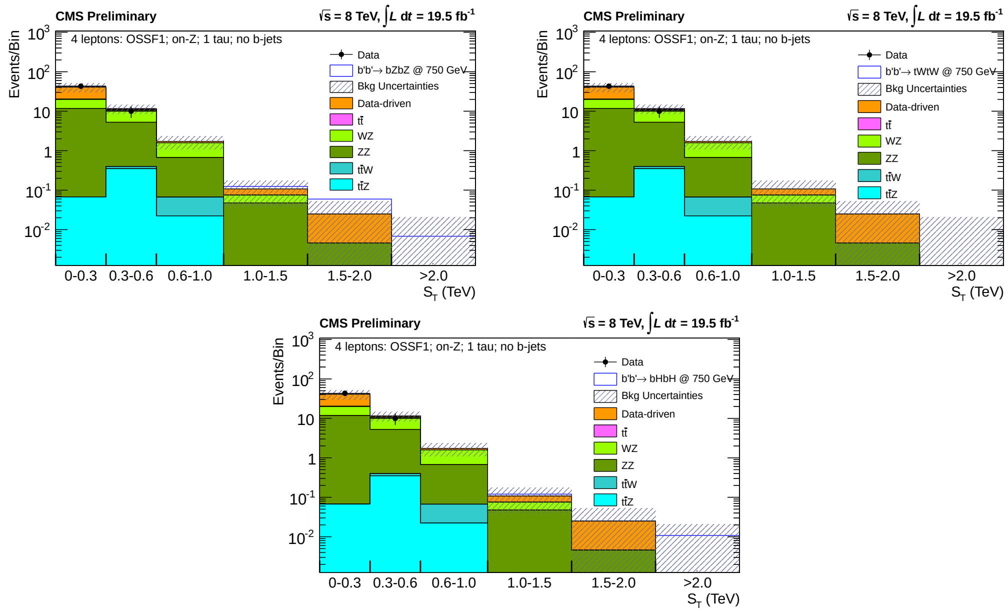


Figure A.20: 4-lepton + OSSF1 + on-Z + Tau1 + b0.

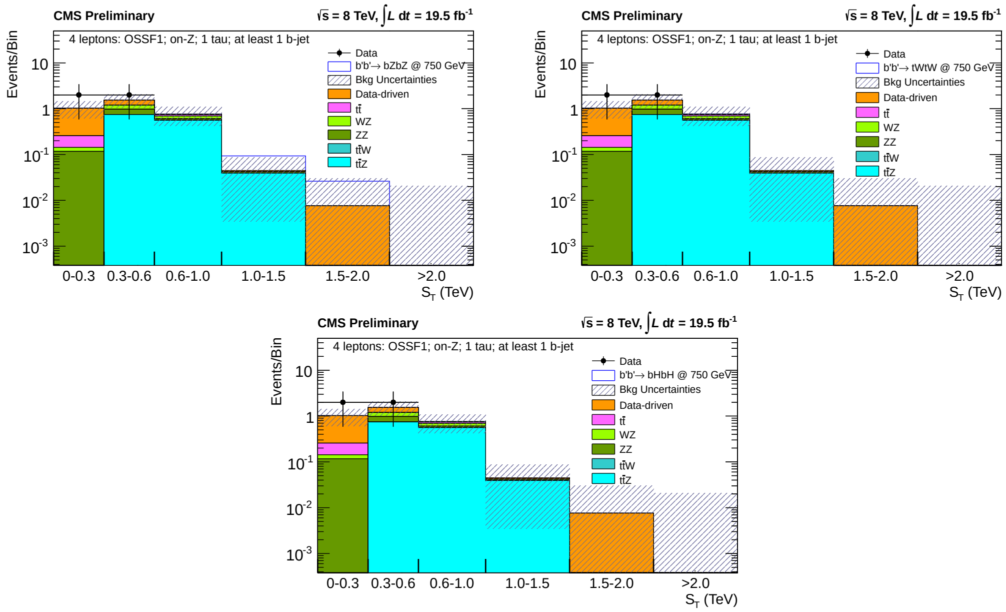


Figure A.21: 4-lepton + OSSF1 + on-Z + Tau1 + b1.

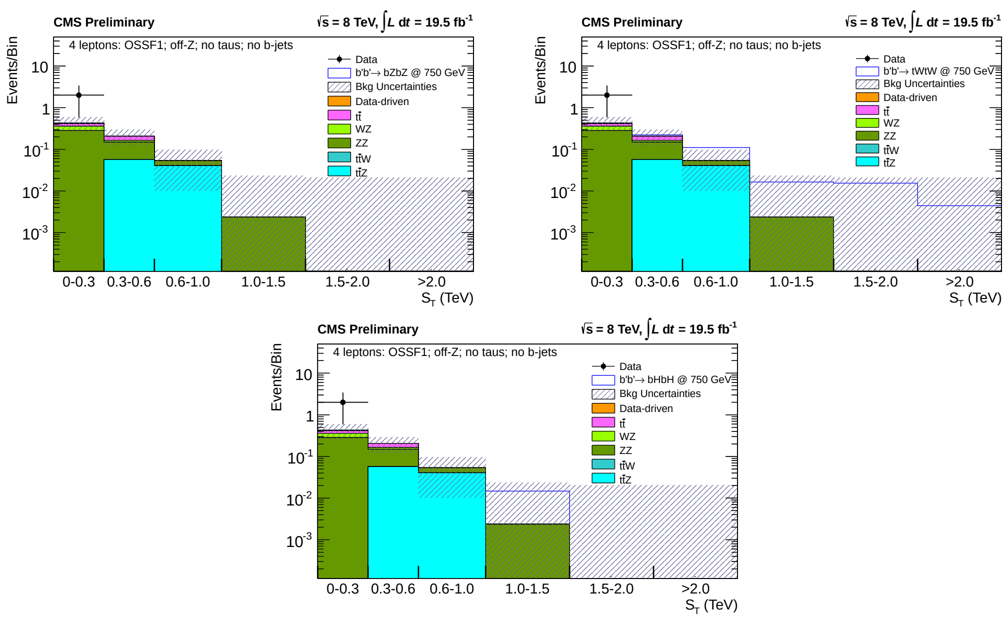


Figure A.22: 4-lepton + OSSF1 + off-Z + Tau0 + b0.

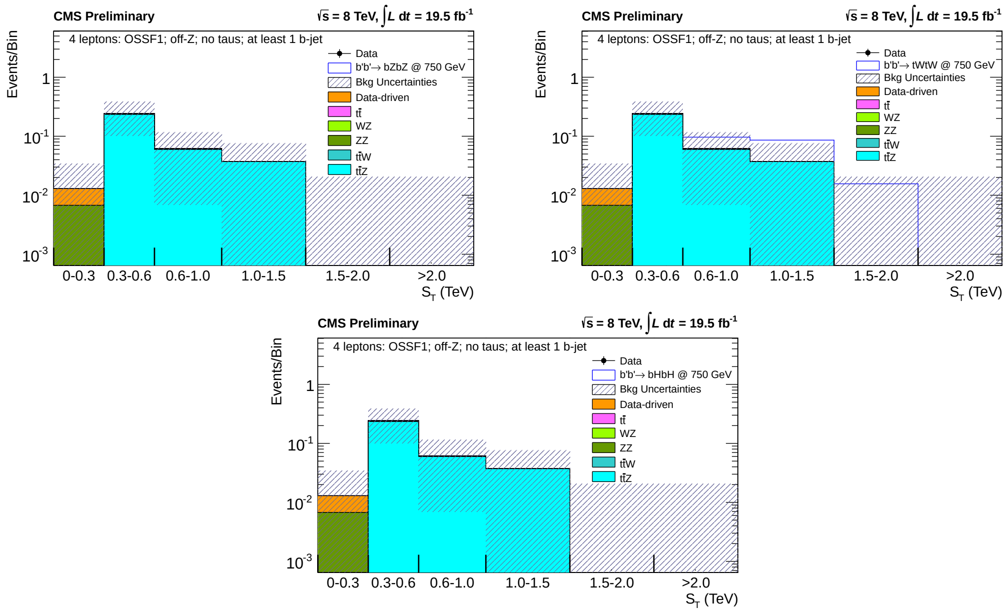


Figure A.23: 4-lepton + OSSF1 + off-Z + Tau0 + b1.

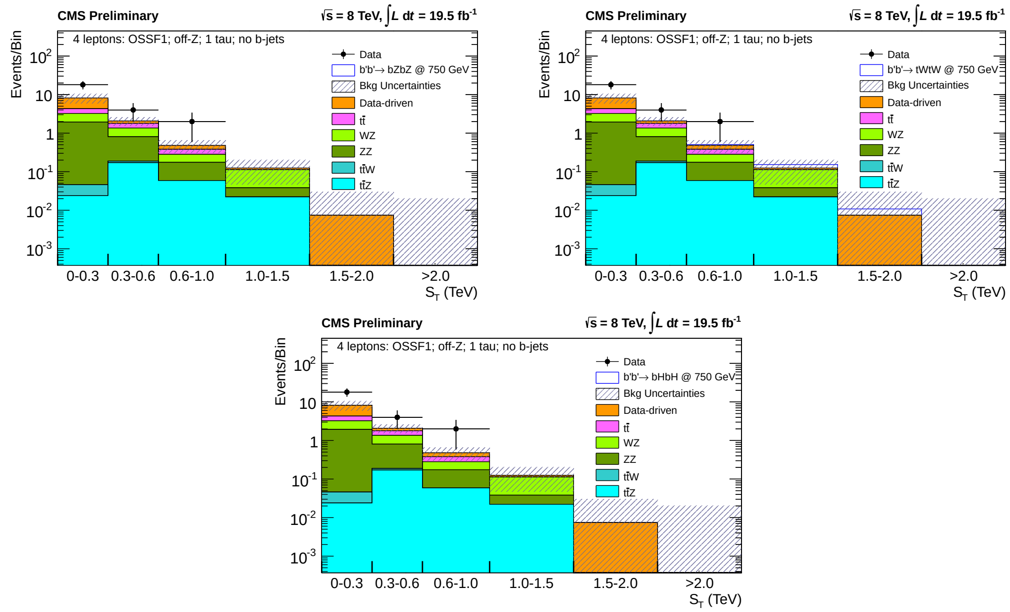


Figure A.24: 4-lepton + OSSF1 + off-Z + Tau1 + b0.

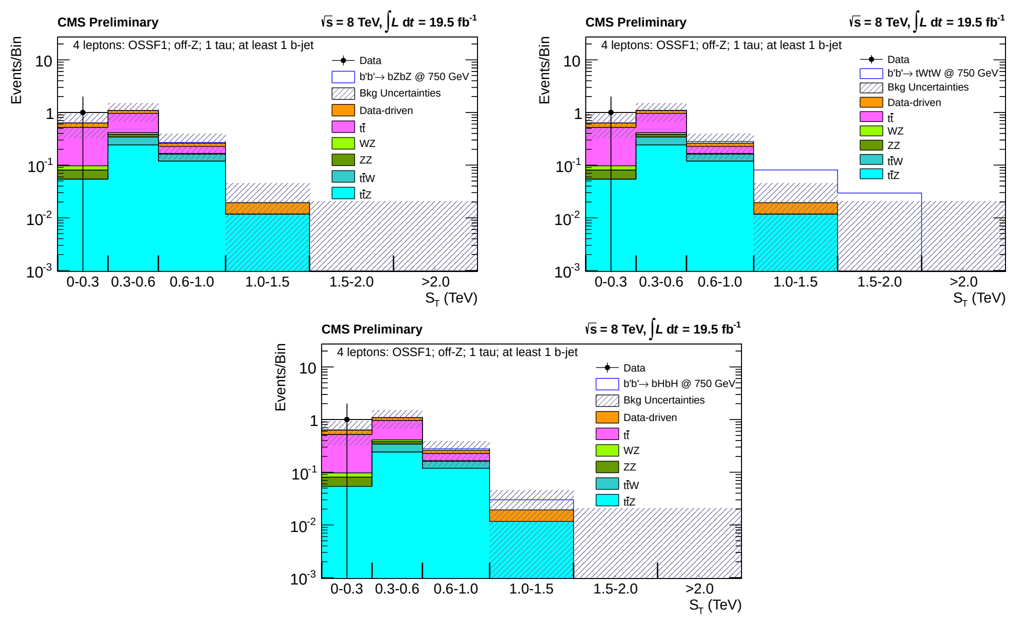


Figure A.25: 4-lepton + OSSF1 + off-Z + Tau1 + b1.

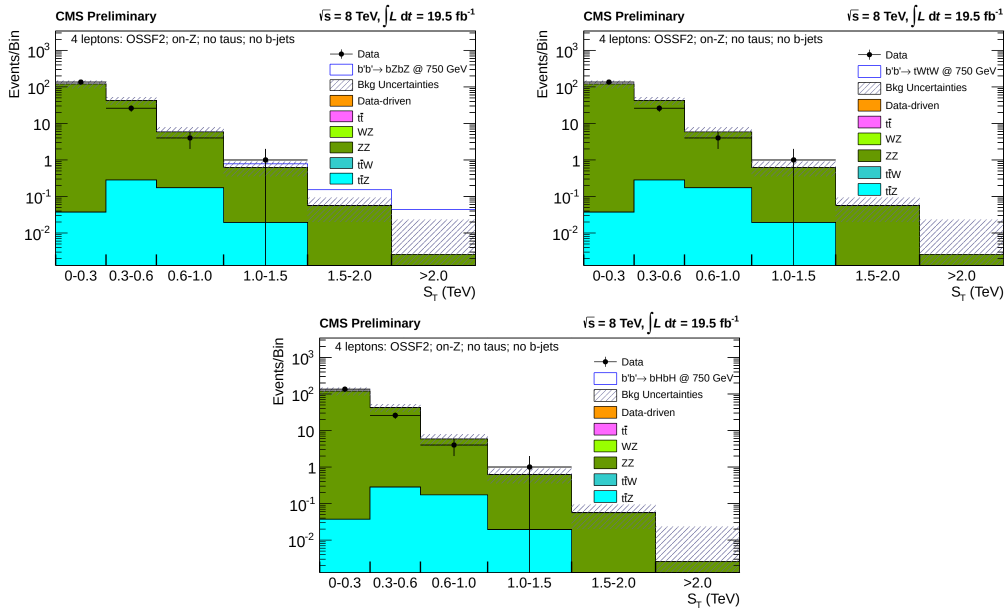


Figure A.26: 4-lepton + OSSF2 + on-Z + Tau0 + b0.

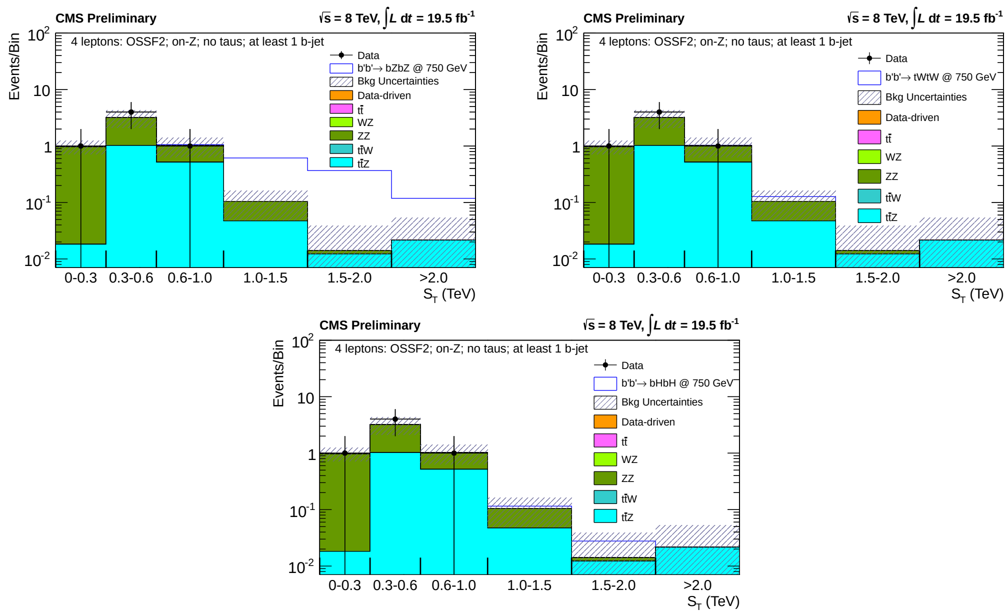


Figure A.27: 4-lepton + OSSF2 + on-Z + Tau0 + b1.

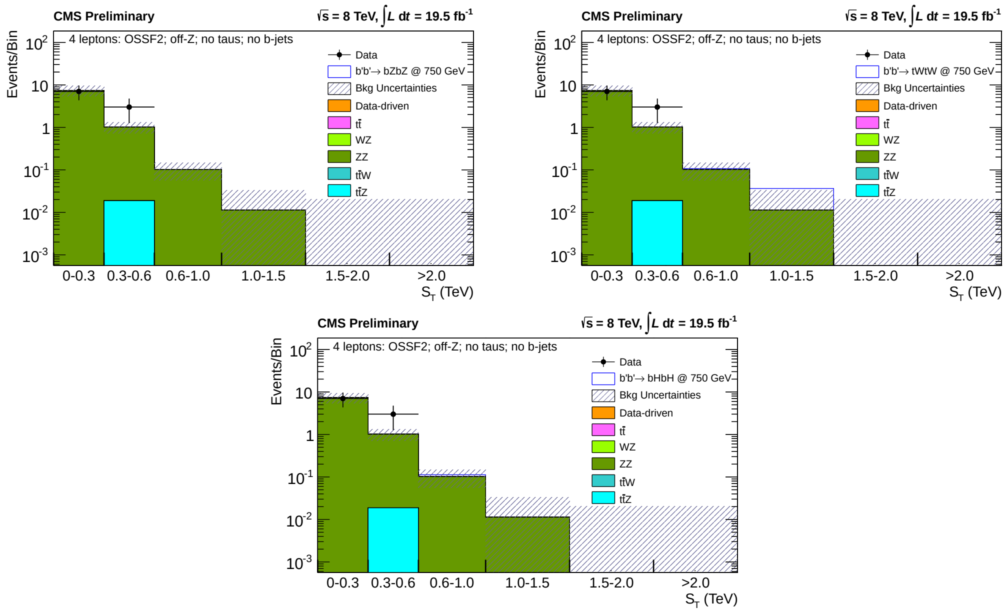


Figure A.28: 4-lepton + OSSF2 + off-Z + Tau0 + b0.

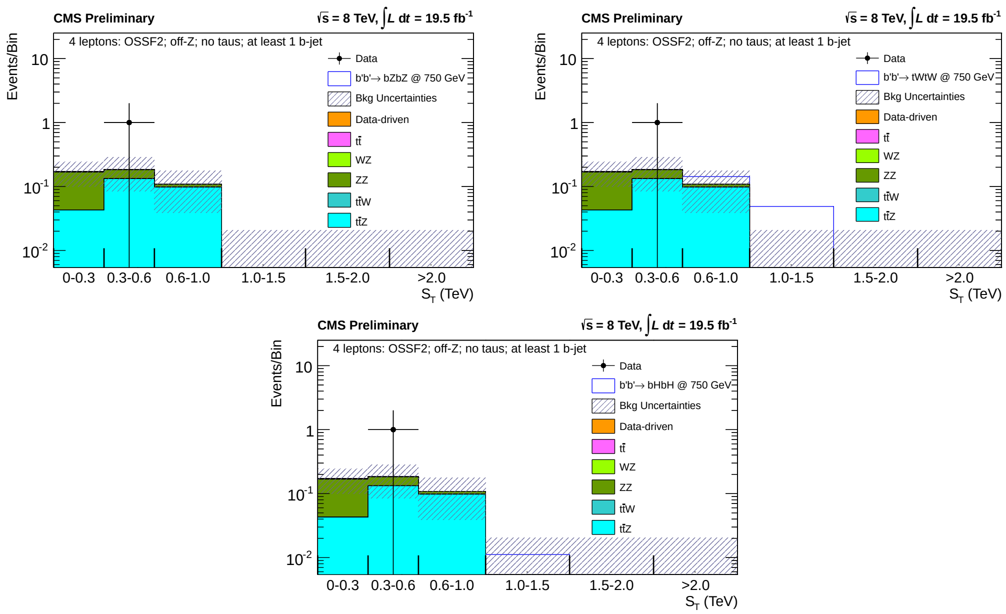


Figure A.29: 4-lepton + OSSF2 + off-Z + Tau0 + b1.

Appendix B

List of Triggers

Below is a list of all the un-pre-scaled triggers used in this multilepton analysis [92]. In order to keep track and no event missed in the course of the analysis we used the logical OR of all the various triggers and their corresponding version. The trigger efficiency is calculated for the logical OR. During the course of data taking these efficiencies are monitored for any significant deviation which then could affect the trigger scale factors for MC simulation.

HLT trigger paths for each datasets. Asterisk (*) in the path names are wildcard to match every versions:

DoubleMuon

- HLT_Mu17_Mu8_v*

Double Electron

- HLT_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_v*

Muon-Electron

- HLT_Mu8_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_v*
- HLT_Mu17_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_v*

Single Electron

- HLT_Ele80_CaloIdVT_TrkIdT_v*
- HLT_Ele100_CaloIdVT_TrkIdT_v*
- HLT_Ele90_CaloIdVT_GsfTrkIdT_v*

Single Muon

- HLT_Mu40_eta2p1_v*
- HLT_Mu50_eta2p1_v*
- HLT_IsoMu24_eta2p1_v*
- HLT_IsoMu30_eta2p1_v*
- HLT_IsoMu34_eta2p1_v*
- HLT_IsoMu40_eta2p1_v*

Appendix C

Derivation of $\mathrm{R}_{\mathrm{dxy}}$

We parametrize the isolation efficiency ϵ_{ratio} ratio and R_{dxy} in terms of free parameters α, which provided the contribution from the nominal sample (Z+jets) and a purely dominant *b*-jet sample. Derivation of the analytical relationship between ϵ_{ratio} and R_{dxy} is found in Ref. [92].

We define the following quantities:

$$N_{\mathrm{NP}} \equiv \text{Number of non-prompt tracks } (d_{xy} > 0.02\,cm). \tag{C.1}$$

$$N_{\mathrm{P}} \equiv \text{Number of prompt tracks } (d_{xy} < 0.02\,cm). \tag{C.2}$$

$$R_{\mathrm{dxy}} \equiv \text{Ratio between the number of non-prompt to prompt tracks.} \tag{C.3}$$

$$N^{\mathrm{Track}}_{\mathrm{non-Iso}} \equiv \text{Number of non-isolated tracks.} \tag{C.4}$$

$$N^{\mathrm{Track}}_{\mathrm{Iso}} \equiv \text{Number of isolated tracks.} \tag{C.5}$$

$$N^{\ell}_{\mathrm{non-Iso}} \equiv \text{Number of isolated leptons.} \tag{C.6}$$

$$\epsilon^{\mathrm{Iso}}_{\ell} \equiv \frac{N^{\ell}_{Iso}}{N^{\ell}_{non-Iso}}. \tag{C.7}$$

$$\epsilon^{\mathrm{Iso}}_{\mathrm{Track}} \equiv \frac{N^{\mathrm{Track}}_{Iso}}{N^{\mathrm{Track}}_{non-Iso}}. \tag{C.8}$$

R_{dxy} as a function of α with range [0,1]. Notation: N^0 corresponding to 0 *b*-jets in the sample and N^1 for the sample with maximum b-jets in the sample.

$$R_{dxy}(\alpha) = \frac{(1-\alpha)\cdot\frac{N^0_{\mathrm{NP}}}{N^0_{\mathrm{NP}}+N^0_{\mathrm{P}}} + \alpha\cdot\frac{N^1_{\mathrm{NP}}}{N^1_{\mathrm{NP}}+N^1_{\mathrm{P}}}}{(1-\alpha)\cdot\frac{N^0_{\mathrm{P}}}{N^0_{\mathrm{NP}}+N^0_{\mathrm{P}}} + \alpha\cdot\frac{N^1_{\mathrm{NP}}}{N^1_{\mathrm{NP}}+N^1_{\mathrm{P}}}} \tag{C.9}$$

$$\alpha = \frac{1}{1+\frac{\frac{R_{dxy}N^1_{\mathrm{P}}+N^1_{\mathrm{NP}}}{N^1_{\mathrm{P}}+N^1_{\mathrm{NP}}}}{\frac{N^0_{\mathrm{NP}}-R_{dxy}N^0_{\mathrm{P}}}{N^1_{\mathrm{NP}}+N^1_{\mathrm{P}}}}} \tag{C.10}$$

$$\alpha = \frac{1}{1+\frac{R_{dxy}-R^1_{dxy}}{R^0_{dxy}-R_{dxy}}\cdot\frac{1+R^0_{dxy}}{1+R^1_{dxy}}} \tag{C.11}$$

$$\tag{C.12}$$

The ratio of isolation efficiencies, ϵ_{ratio}, is the ratio between isolation efficiency of leptons to the isolation efficiency of tracks:

$$\epsilon_{ratio} = \frac{\epsilon^{\mathrm{Iso}}_{\ell}}{\epsilon^{\mathrm{Iso}}_{\mathrm{Track}}} \tag{C.13}$$

$\epsilon_{\mathrm{ratio}}$ as a function of α:

$$\epsilon^{\mathrm{Iso}}_{\ell}(\alpha) = \frac{(1-\alpha)\cdot\frac{N^{\ell,0}_{\mathrm{Iso}}}{N^{\ell,0}_{\mathrm{Iso}}+N^{\ell,0}_{\mathrm{non-Iso}}} + \alpha\cdot\frac{N^{\ell,0}_{\mathrm{Iso}}}{N^{\ell,1}_{\mathrm{Iso}}+N^{\ell,1}_{\mathrm{non-Iso}}}}{(1-\alpha)\cdot\frac{N^{\ell,0}_{\mathrm{non-Iso}}}{N^{\ell,0}_{\mathrm{Iso}}+N^{\ell,0}_{\mathrm{non-Iso}}} + \alpha\cdot\frac{N^{\ell,1}_{\mathrm{non-Iso}}}{N^{\ell,1}_{\mathrm{Iso}}+N^{\ell,1}_{\mathrm{non-Iso}}}} \tag{C.14}$$

$$= \frac{\epsilon^{\mathrm{Iso},0}_{\ell}\cdot\epsilon^{\mathrm{Iso},1}_{\ell} + \epsilon^{\mathrm{Iso},0}_{\ell} + \alpha\cdot(\epsilon^{\mathrm{Iso},1}_{\ell} - \epsilon^{\mathrm{Iso},0}_{\ell})}{1+\epsilon^{\mathrm{Iso},1}_{\ell} + \alpha\cdot(\epsilon^{\mathrm{Iso},1}_{\ell} - \epsilon^{\mathrm{Iso},0}_{\ell})} \tag{C.15}$$

$$\epsilon^{\mathrm{Iso}}_{\mathrm{Track}}(\alpha) = \frac{\epsilon^{\mathrm{Iso},0}_{\mathrm{Track}}\cdot\epsilon^{\mathrm{Iso},1}_{\mathrm{Track}} + \epsilon^{\mathrm{Iso},0}_{\mathrm{Track}} + \alpha\cdot(\epsilon^{\mathrm{Iso},1}_{\mathrm{Track}} - \epsilon^{\mathrm{Iso},0}_{\mathrm{Track}})}{1+\epsilon^{\mathrm{Iso},1}_{\mathrm{Track}} + \alpha\cdot(\epsilon^{\mathrm{Iso},1}_{\mathrm{Track}} - \epsilon^{\mathrm{Iso},0}_{\mathrm{Track}})} \tag{C.16}$$

$$\epsilon_{\mathrm{ratio}}(\alpha) = \frac{\frac{\frac{1}{\alpha}\cdot(\epsilon^{\mathrm{Iso},0}_{\ell}\cdot\epsilon^{\mathrm{Iso},1}_{\ell}+\epsilon^{\mathrm{Iso},0}_{\ell})+(\epsilon^{\mathrm{Iso},1}_{\ell}-\epsilon^{\mathrm{Iso},0}_{\ell})}{\frac{1}{\alpha}\cdot(1+\epsilon^{\mathrm{Iso},1}_{\ell})+(\epsilon^{\mathrm{Iso},1}_{\ell}-\epsilon^{\mathrm{Iso},0}_{\ell})}}{\frac{\frac{1}{\alpha}\cdot(\epsilon^{\mathrm{Iso},0}_{\mathrm{Track}}\cdot\epsilon^{\mathrm{Iso},1}_{\mathrm{Track}}+\epsilon^{\mathrm{Iso},0}_{\mathrm{Track}})+(\epsilon^{\mathrm{Iso},1}_{\mathrm{Track}}-\epsilon^{\mathrm{Iso},0}_{\mathrm{Trac}})}{\frac{1}{\alpha}(1+\epsilon^{\mathrm{Iso},1}_{\mathrm{Track}})+(\epsilon^{\mathrm{Iso},1}_{\mathrm{Track}}-\epsilon^{\mathrm{Iso},0}_{\mathrm{Track}})}} \tag{C.17}$$

Re-express in terms of α:

$$\epsilon_{\text{ratio}}(R_{dxy}) = \frac{\frac{(1+\frac{R_{dx}-R^1_{dx}}{R^0_{dx}-R_{dx}}\cdot\frac{1+R^0_{dx}}{1+R^1_{dx}})\cdot(\epsilon_\ell^{\text{Iso},0}\cdot\epsilon_\ell^{\text{Iso},1}+\epsilon_\ell^{\text{Iso},0})+(\epsilon_\ell^{\text{Iso},1}-\epsilon_\ell^{\text{Iso},0})}{(1+\frac{R_{dxy}-R^1_{dxy}}{R^0_{dxy}-R_{dxy}}\cdot\frac{1+R^0_{dxy}}{1+R^1_{dxy}})\cdot(1+\epsilon_\ell^{\text{Iso},1})+(\epsilon_\ell^{\text{Iso},1}-\epsilon_\ell^{\text{Iso},0})}}{\frac{(1+\frac{R_{dxy}-R^1_{dxy}}{R^0_{dxy}-R_{dxy}}\cdot\frac{1+R^0_{dxy}}{1+R^1_{dxy}})\cdot(\epsilon_{\text{Track}}^{\text{Iso},0}\cdot\epsilon_{\text{Track}}^{\text{Iso},1}+\epsilon_{\text{Track}}^{\text{Iso},0})+(\epsilon_{\text{Track}}^{\text{Iso},1}-\epsilon_{\text{Track}}^{\text{Iso},0})}{(1+\frac{R_{dxy}-R^1_{dxy}}{R^0_{dxy}-R_{dxy}}\cdot\frac{1+R^0_{dxy}}{1+R^1_{dxy}})\cdot(1+\epsilon_{\text{Track}}^{\text{Iso},1})+(\epsilon_{\text{Track}}^{\text{Iso},1}-\epsilon_{\text{Track}}^{\text{Iso},0})}} \tag{C.18}$$

$$\tag{C.19}$$

Bibliography

[1] M. Peskin and D. Schroeder, *An Introduction to Quantum Field Theory.* Addison-Wesley Publishing Company, 1995.

[2] E. Noether, "Invariante variationsprobleme," *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, vol. 1918, pp. 235–257, 1918.

[3] Wikipedia, "Standard model of particle physics." `http://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg`.

[4] S. Glashow, "Partial Symmetries of Weak Interactions," *Nucl.Phys.*, vol. 22, pp. 579–588, 1961.

[5] S. Weinberg, "A Model of Leptons," *Phys.Rev.Lett.*, vol. 19, pp. 1264–1266, 1967.

[6] D. Griffiths, *Introduction to Elementary Particles.* WILEY-VCH, 2011.

[7] F. Halzen and A. D. Martin, *QUARKS AND LEPTONS: AN INTRODUCTORY COURSE IN MODERN PARTICLE PHYSICS.* New York, USA: Wiley, 1984.

[8] S. Chatrchyan *et al.*, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Phys.Lett.*, vol. B716, pp. 30–61, 2012.

[9] G. Aad *et al.*, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Phys.Lett.*, vol. B716, pp. 1–29, 2012.

[10] "Precise determination of the mass of the Higgs boson and studies of the compatibility of its couplings with the standard model," Tech. Rep. CMS-PAS-HIG-14-009, CERN, Geneva, 2014.

[11] R. Ellis, W. Stirling, and B. Webber, *QCD and Collider Physics.* Cambridge University Press, 1996.

[12] S. Schael *et al.*, "Precision electroweak measurements on the Z resonance," *Phys.Rept.*, vol. 427, pp. 257–454, 2006.

[13] "Combination of standard model Higgs boson searches and measurements of the properties of the new boson with a mass near 125 GeV," Tech. Rep. CMS-PAS-HIG-13-005, CERN, Geneva, 2013.

[14] "Combined coupling measurements of the Higgs-like boson with the ATLAS detector using up to 25 fb^{-1} of proton-proton collision data," Tech. Rep. ATLAS-CONF-2013-034, CERN, Geneva, Mar 2013.

[15] P. H. Frampton, P. Hung, and M. Sher, "Quarks and leptons beyond the third generation," *Phys. Rept.*, vol. 330, p. 263, 2000.

[16] B. Holdom, W. Hou, T. Hurth, M. Mangano, S. Sultansoy, *et al.*, "Four Statements about the Fourth Generation," *PMC Phys.*, vol. A3, p. 4, 2009.

[17] S. Chatrchyan *et al.*, "Combined search for the quarks of a sequential fourth generation," *Phys.Rev. D*, vol. 86, p. 112003, 2012.

[18] J. Aguilar-Saavedra, R. Benbrik, S. Heinemeyer, and M. Pérez-Victoria, “Handbook of vectorlike quarks: Mixing and single production,” *Phys.Rev.*, vol. D88, no. 9, p. 094010, 2013.

[19] B. A. Dobrescu, K. Kong, and R. Mahbubani, “Prospects for top-prime quark discovery at the Tevatron,” *JHEP*, vol. 0906, p. 001, 2009.

[20] G. Branco and L. Lavoura, “On the Addition of Vector Like Quarks to the Standard Model,” *Nucl.Phys.*, vol. B278, p. 738, 1986.

[21] F. del Aguila and M. J. Bowick, “The Possibility of New Fermions With Δ I = 0 Mass,” *Nucl.Phys.*, vol. B224, p. 107, 1983.

[22] S. L. Glashow, J. Iliopoulos, and L. Maiani, “Weak interactions with lepton-hadron symmetry,” *Phys. Rev. D*, vol. 2, pp. 1285–1292, Oct 1970.

[23] G. C. Branco, T. Morozumi, P. A. Parada, and M. N. Rebelo, “CP,” *Phys. Rev. D*, vol. 48, pp. 1167–1175, Aug 1993.

[24] Y. Nir and D. Silverman, “z,” *Phys. Rev. D*, vol. 42, pp. 1477–1484, Sep 1990.

[25] F. del Aguila, J. Aguilar-Saavedra, and G. Branco, “CP violation from new quarks in the chiral limit,” *Nucl.Phys.*, vol. B510, pp. 39–60, 1998.

[26] G. Barenboim, F. Botella, G. Branco, and O. Vives, “How sensitive to FCNC can B0 CP asymmetries be?,” *Phys.Lett.*, vol. B422, pp. 277–286, 1998.

[27] W.-S. Hou, “Source of CP Violation for the Baryon Asymmetry of the Universe,” *Chin.J.Phys.*, vol. 47, p. 134, 2009.

[28] K. Babu, I. Gogoladze, M. U. Rehman, and Q. Shafi, “Higgs Boson Mass, Sparticle Spectrum and Little Hierarchy Problem in Extended MSSM,” *Phys.Rev.*, vol. D78, p. 055017, 2008.

[29] M. Schmaltz and D. Tucker-Smith, “Little Higgs review,” *Ann. Rev. Nucl. Part. Sci.*, vol. 55, pp. 229–270, 2005.

[30] N. Arkani-Hamed, A. G. Cohen, and H. Georgi, “Electroweak symmetry breaking from dimensional deconstruction,” *Phys. Lett. B*, vol. 513, pp. 232–240, 2001.

[31] P. Lodone, “Vector-like quarks in a ’composite’ Higgs model,” *JHEP*, vol. 0812, p. 029, 2008.

[32] S. Gopalakrishna, T. Mandal, S. Mitra, and G. Moreau, “LHC Signatures of Warped-space Vectorlike Quarks,” 2013.

[33] J. Kang, P. Langacker, and B. D. Nelson, “Theory and Phenomenology of Exotic Isosinglet Quarks and Squarks,” *Phys.Rev.*, vol. D77, p. 035003, 2008.

[34] M. Sher, “Fourth generation b' decays into b + higgs,” 1999.

[35] S. Heinemeyer *et al.*, “Handbook of LHC Higgs Cross Sections: 3. Higgs Properties,” 2013.

[36] L. Evans and P. Bryant, “LHC Machine,” *JINST*, vol. 3, p. S08001, 2008.

[37] CERN, “The CERN accelerator complex.” `http://te-epc-lpc.web.cern.ch/te-epc-lpc/machines/pagesources/Cern-Accelerator-Complex.jpg`.

[38] Public, “CMS luminosity information.” `https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults`.

[39] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report.* Geneva: CERN, 2004.

[40] S. Chatrchyan *et al.*, "The CMS experiment at the CERN LHC," *JINST*, vol. 3, p. S08004, 2008.

[41] T. Lenzi, "Development and Study of Different Muon Track Reconstruction Algorithms for the Level-1 Trigger for the CMS Muon Upgrade with GEM Detectors," 2013.

[42] S. Chatrchyan *et al.*, "The CMS experiment at the CERN LHC," *JINST*, vol. 3, p. S08004, 2008.

[43] B. Isildak, "Measurement of the differential dijet production cross section in proton-proton collisions at $\sqrt{s} = 7$ tev," 2013.

[44] C. Collaboration], "Description and performance of track and primary-vertex reconstruction with the cms tracker." `https://inspirehep.net/record/1298029/files/figs_2011_cmsTracker_TrackerLayout.png`.

[45] S. Chatrchyan *et al.*, "Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in *pp* Collisions at $\sqrt{s} = 7$ TeV," *JINST*, vol. 8, p. P09009, 2013.

[46] J. Brooke, "Performance of the CMS Level-1 Trigger," *PoS*, vol. ICHEP2012, p. 508, 2013.

[47] "CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update," Tech. Rep. CMS-PAS-LUM-13-001, CERN, Geneva, 2013.

[48] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, "Madgraph 5: Going beyond," *JHEP*, vol. 1106, p. 128, 2011.

[49] F. Maltoni and T. Stelzer, "MadEvent: Automatic event generation with MadGraph," *JHEP*, vol. 02, p. 027, 2003.

[50] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual," *JHEP*, vol. 0605, p. 026, 2006.

[51] "Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector," 2010.

[52] J. Pumplin, D. Stump, J. Huston, H. Lai, P. M. Nadolsky, *et al.*, "New generation of parton distributions with uncertainties from global QCD analysis," *JHEP*, vol. 0207, p. 012, 2002.

[53] S. Agostinelli *et al.*, "GEANT4: A simulation toolkit," *Nucl. Instrum. Meth. A*, vol. 506, p. 250, 2003.

[54] P. Z. Skands, B. Allanach, H. Baer, C. Balazs, G. Belanger, *et al.*, "SUSY Les Houches accord: Interfacing SUSY spectrum calculators, decay packages, and event generators," *JHEP*, vol. 0407, p. 036, 2004.

[55] M. Czakon, P. Fiedler, and A. Mitov, "The total top quark pair production cross-section at hadron colliders through $O(\alpha_S^4)$," *Phys.Rev.Lett.*, vol. 110, p. 252004, 2013.

[56] CMS, "Heavy exotic quarks pair production cross sections for CMS at 8 TeV." `https://twiki.cern.ch/twiki/bin/viewauth/CMS/B2GMonteCarlo`, 2012.

[57] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method," *JHEP*, vol. 0711, p. 070, 2007.

[58] P. M. Nadolsky *et al.*, "Implications of CTEQ global analysis for collider observables," *Phys. Rev. D*, vol. 78, p. 013004, 2008.

[59] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations," *JHEP*, vol. 06, p. 029, 2002.

[60] S. Frixione, P. Nason, and B. R. Webber, “Matching NLO QCD and parton showers in heavy flavor production,” *JHEP*, vol. 08, p. 007, 2003.

[61] N. Kidonakis, “Differential and total cross sections for top pair and single top production,” pp. 831–834, 2012.

[62] J. M. Campbell and R. K. Ellis, “$t\bar{t}W^{\pm}$ production and decay at NLO,” *JHEP*, vol. 1207, p. 052, 2012.

[63] M. Garzelli, A. Kardos, C. Papadopoulos, and Z. Trocsanyi, “$t\bar{t}W^{\pm}$ and $t\bar{t}Z$ Hadroproduction at NLO Accuracy in QCD with Parton Shower and Hadronization Effects,” *JHEP*, vol. 11, p. 056, 2012.

[64] J. M. Campbell and R. Ellis, “MCFM for the Tevatron and the LHC,” *Nucl. Phys. Proc. Suppl.*, vol. 205-206, pp. 10–15, 2010.

[65] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush, “W Physics at the LHC with FEWZ 2.1,” *Comput. Phys. Commun.*, vol. 184, pp. 208–214, 2013.

[66] R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, R. Pittau, *et al.*, “Four-lepton production at hadron colliders: aMC@NLO predictions with theoretical uncertainties,” *JHEP*, vol. 02, p. 099, 2012.

[67] V. Khachatryan *et al.*, “CMS tracking performance results from early LHC operation,” 2010.

[68] C. Collaboration], ““Offline Primary Vertex Reconstruction with Deterministic Annealing Clustering”,” *CMS Analysis Note*, vol. no. AN-2011/014, 2011.

[69] C. Collaboration], ““Adaptive Vertex Fitting”,” *CMS Analysis Note*, vol. no. AN-2007/008, 2007.

[70] G. P. Salam, “Towards Jetography,” *Eur.Phys.J.*, vol. C67, pp. 637–686, 2010.

[71] S. D. Ellis and D. E. Soper, “Successive combination jet algorithm for hadron collisions,” *Phys.Rev.*, vol. D48, pp. 3160–3166, 1993.

[72] Y. L. Dokshitzer, G. Leder, S. Moretti, and B. Webber, “Better jet clustering algorithms,” *JHEP*, vol. 9708, p. 001, 1997.

[73] M. Cacciari, G. P. Salam, and G. Soyez, “The Anti-k(t) jet clustering algorithm,” *JHEP*, vol. 0804, p. 063, 2008.

[74] S. Chatrchyan *et al.*, “Identification of b-quark jets with the CMS experiment,” *JINST*, vol. 8, p. P04013, 2013.

[75] C. Collaboration, “Performance of b tagging at sqrt(s)=8 TeV in multijet, ttbar and boosted topology events,” 2013.

[76] CMS , “Performance of cms muon reconstruction in pp collision events at $\sqrt{s} = 7$TeV,” 2012. Submitted to *J. Inst.*

[77] R. Fruhwirth, “Application of Kalman filtering to track and vertex fitting,” *Nucl.Instrum.Meth.*, vol. A262, pp. 444–450, 1987.

[78] S. Chatrchyan *et al.*, “Performance of tau-lepton reconstruction and identification in CMS,” *JINST*, vol. 7, p. P01001, 2012.

[79] “Photon reconstruction and identification at sqrt(s) = 7 TeV,” Tech. Rep. CMS-PAS-EGM-10-005, CERN, 2010. Geneva, 2010.

[80] CMS, "Performance of muon identification in pp collisions at $\sqrt{s} = 7$ TeV," *CMS Physics Analysis Summary*, vol. CMS-PAS-MUO-10-002, 2010.

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

[82] "Electron reconstruction and identification at sqrt(s) = 7 TeV," Tech. Rep. CMS-PAS-EGM-10-004, CERN, Geneva, 2010.

[83] "Search for Vector-Like b' Pair Production with Multilepton Final States in pp collisions at $\sqrt{s} = 8$ TeV," Tech. Rep. CMS-PAS-B2G-13-003, CERN, Geneva, 2013.

[84] S. Chatrchyan *et al.*, "Performance of CMS muon reconstruction in *pp* collision events at $\sqrt{s} = 7$ TeV," *JINST*, vol. 7, p. P10002, 2012.

[85] CMS, "Performance of τ-lepton reconstruction and identification in cms," *JINST*, vol. 7, no. 01, p. P01001, 2012.

[86] M. Cacciari, G. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *JHEP*, vol. 04, p. 063, 2008.

[87] G. Soyez, G. P. Salam, J. Kim, S. Dutta, and M. Cacciari, "Pileup subtraction for jet shapes," *Phys.Rev.Lett.*, vol. 110, p. 162001, 2013.

[88] A. Perloff, "Pileup measurement and mitigation techniques in CMS," *J.Phys.Conf.Ser.*, vol. 404, p. 012045, 2012.

[89] CMS, "Missing transverse energy performance in minimum-bias and jet events from proton-proton collisions at $\sqrt{s} = 7$ TeV," *CMS Physics Analysis Summary*, vol. CMS-PAS-JME-10-004, 2010.

[90] CMS, "CMS MET performance in events containing electroweak bosons from pp collisions at $\sqrt{s} = 7$ TeV," *CMS Physics Analysis Summary*, vol. CMS-PAS-JME-10-005, 2010.

[91] CMS Collaboration, "Identification and filtering of uncharacteristic noise in the CMS hadron calorimeter," *Journal of Instrumentation*, vol. 5, p. 3014, Mar. 2010.

[92] CMS, "Background and efficiency determination methods for multilepton analyses," *CMS Analysis Note*, vol. CMS-AN-2012/257, 2012.

[93] C. Collaboration, "Search for new physics with same-sign dileptons using the 2011 dataset of cms," *CMS Analysis Note*, vol. CMS-AN-2011/258, 2011.

[94] Public, "Search for vector-like b' pair production with multilepton final states in pp collisions at $\sqrt{s} = 8$ tev." `https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsB2G13003`.

[95] M. Garzelli, A. Kardos, C. Papadopoulos, and Z. Trocsanyi, "t$\bar{\text{t}}$ $W^{\pm}$ and t$\bar{\text{t}}$ Z Hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effects," *JHEP*, vol. 1211, p. 056, 2012.

[96] S. van der Meer, "Calibration of the Effective Beam Height in the ISR," 1968.

[97] S. M. White, H. Burkhardt, and P. Puzo, *Determination of the Absolute Luminosity at the LHC*. PhD thesis, Orsay, Université Paris-Sud 11, Orsay, 2010. Presented on 11 Oct 2010.

[98] C. Collaboration, "CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update," 2013.

[99] C. I. twiki], "Apply SF for b/c/light tags to directly tagged jets in MC," https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFUtil.

[100] C. Collaboration], "$E_{\mathrm{T}}^{\mathrm{miss}}$ Resolution Dependence on Pileup," *CMS Analysis Note*, vol. no. AN-2012/342, 2012.

[101] K. O. et al., "Particle data group statistics," no. 38, 090001, 2014.

[102] "Procedure for the LHC Higgs boson search combination in summer 2011," Tech. Rep. ATL-PHYS-PUB-2011-011, CERN, Geneva, Aug 2011.

[103] T. Junk, "Confidence Level Computation for Combining Searches with Small Statistics," *Nucl. Instrum. Meth. A*, vol. 434, pp. 435–443, 1999.

[104] A. L. Read, "Presentation of search results: The CL(s) technique," *J. Phys. G*, vol. 28, pp. 2693–2704, 2002.

[105] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *European Physical Journal C*, vol. 71, p. 1554, Feb. 2011.

[106] J. C. Street and E. C. Stevenson, "New evidence for the existence of a particle of mass intermediate between the proton and electron," *Phys. Rev.*, vol. 52, pp. 1003–1004, Nov 1937.

[107] e. a. V. Buescher, "Recommendation of the ad-hoc committee on limit-setting procedures to be used by d$\emptyset$ in run ii," no. 4686, 2004.

[108] Mingshui Chen, et al., "Lands software package," 2013.

[109] G. Cacciapaglia, A. Deandrea, D. Harada, and Y. Okada, "Bounds and Decays of New Heavy Vector-like Top Partners," *JHEP*, vol. 1011, p. 159, 2010.