SEARCH FOR Z' PRODUCTION IN 4 B-TAGGED JET FINAL STATES IN PROTON-PROTON COLLISIONS

A Dissertation

by

ANDREA DELGADO

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPY

Chair of Committee,	Ricardo Eusebi
Committee Members,	Bhaskar Dutta
	Teruki Kamon
	Charles M. Folden III
Head of Department,	Grigory Rogachev

May 2020

Major Subject: Physics

ABSTRACT

The LHCb experiment has reported discrepancies at a level of 3 sigma in the ratio in which B mesons decay to muons and electrons. Some theories attempt to explain these anomalies by theorizing the existence of new particles beyond the standard model, such as a new heavy neutrl gauge boson (Z') coupling mostly to third-generation fermions, specifically b-quarks. This study performs a search for Z' boson at the LHC with the CMS experiment. The main production channel is b-quark fusion, and since the b-quark parton distribution functions (PDFs) are at least ten times lower when compared to the gluon PDFs, a substantial contribution of bottom quarks coming from gluon-splitting to Z' production is expected. In short, a study for $Z' \rightarrow b\bar{b}$ decays is presented. The final state consists of 4 b jets, with the two extra jets coming from the initial gluon splitting. Results correspond to 35.9 fb^{-1} of proton-proton collision data recorded by the CMS detector at the LHC with a center-of-mass-energy of 13 TeV during 2016.

DEDICATION

To Elvis, Simon, and Alice

ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my committee for their time spent reviewing my work and the final document. I'm extremely grateful to Dr. Ricardo Eusebi for the invaluable teachings during my time at Texas A&M. Many thanks to Dr. Denis Rathjens, my success and completion of my dissertation would not have been possible without his support. I would also like to extend my deepest gratitude to Dr. Alexx Perloff, who taught me a lot of the technical skills needed for this project at the very beginning of the program.

I cannot begin to express my thanks to my husband, Elvis, who provided me with encouragement and patience throughout my academic career. I'm deeply indebted to my parents, my dad, Javier, who taught me the value of working in what you love and to my mom, who taught me that you could have both a family and a career if you work hard. Finally, I would like to thank my children, Alice and Simon, who made me stronger, better, and more fulfilled than I could have ever imagined.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professors Ricardo Eusebi, Bhaskar Dutta, and Teruki Kamon of the Department of Physics and Astronomy and Professor Charles M. Folden III of the Department of Chemistry.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by the Texas A&M University System Louis Stokes Alliance for Minority Participation (TAMUS LSAMP) Bridge to the Doctorate (BTD) Cohort IX (2013-2015) Program by the National Science Foundation Award No. HRD-1301877. This material is also based upon work supported by the National Science Foundation Graduate Research Fellowship.

NOMENCLATURE

ATLAS	A Toroidal LHC ApparatuS
BEH	Brout-Englert-Higgs
BFF	Bottom Fermion Fusion
CERN	European Center for Nuclear Research
CMS	Compact Muon Solenoid
CSV	Combined Secondary Vertex
CTF	Combinatorial Track Finder
DAQ	Data Acquisition
EMCAL	Electromagnetic Calorimeter
EW	Electroweak
EWSB	Electroweak Symmetry Breaking
FCNC	Flavor-Changing Neutral-Current
GR	General Relativity
HCAL	Hadron Calorimeter
HLT	High Level Trigger
LHC	Large Hadron Collider
η	Pseudorapidity
PF	Particle Flow
p_T	Transverse Momentum
PU	Pile-up
PV	Primary Vertex
QCD	Quantum Chromodynamics

QED	Quantum Electrodynamics
QFT	Quantum Field Theory
SM	Standard Model of particle physics
VBF	Vector Boson Fusion

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

Particle physics is a branch of physics that studies the nature of the constituents of matter and radiation. Particle physicists investigate the irreducibly smallest detectable particles and the fundamental interactions necessary to explain their behavior.

By our current understanding, these elementary particles are excitations of the quantum fields that also govern their interactions. The currently dominant theory explaining these fundamental particles and fields, along with their dynamics, is called the Standard Model (SM).

The SM is amazingly successful, yet it leaves many basic questions unanswered. For example, parameters such as the cosmological constant, Higgs mass, or neutron electric dipole moment are unexplained. Also, phenomena such as dark matter and baryogenesis are still unexplained. Therefore, there is strong evidence that the SM is still incomplete. Thus, modern particle physics generally investigates the SM and its various possible extensions, e.g., to the newest "known" particle, the Higgs boson, or even the oldest known force field, gravity.

A possible approach in the search for new particles involves performing precise measurements of the properties of known decays of hadrons that are accurately described by the SM. For example, processes that occur via the weak force, such as the decay of a kaon (a hadron containing a strange quark) or of a b-hadron (which includes a bottom quark), are particularly interesting. As a consequence of Quantum Field Theory (QFT), such decays can occur through transient particles that have a physical mass greater than the amount of mass-energy available from the decaying particle. These particles are referred to as "virtual." Massive new particles can cause significant deviations from the SM predictions of the decay rate and of the dynamics of the decay products.

Of particular interest are the anomalies reported by the LHCb and Belle collaborations in the decay of B-mesons. These anomalies might present a challenge to the current standard model assumption of lepton universality. Also, a possible explanation of such deviations from the SM expectation could imply the existence of flavor-violating couplings, which are not described by the SM. Therefore, a massive Z' with a flavor changing b-s quark coupling, and a non-universal

coupling to leptons could easily accommodate such anomalies.

According to Einstein's famous equation, E = mc^2, energy, and mass are interchangeable. Therefore, to produce massive particles, like the Z', a large amount of energy is required. The LHC has been designed and constructed to produce highly energetic proton-proton collisions in which a variety of elementary particles may be generated. High energy collisions also enable the study of tiny distance scales. Massive particles produced in the proton-proton collision typically decay very rapidly. Highly advanced detectors, such as CMS, are needed to observe and measure the properties of their decay products.

This dissertation will present a search for a Z' boson in the $Z \rightarrow b\bar{b}$ channel using 13 TeV proton-proton data collected by the CMS detector. In this scenario, bottom quarks arising from gluon splitting can fuse into Z' allowing the LHC to probe it.

The document will be organized in the following way. Section 2 will present an overview of the SM, the Higgs mechanism, and a brief overview of the current experimental results on the B-meson anomalies. Finally, the theoretical framework that proposes a flavor violating coupling mediated by the Z' will also be introduced. The LHC and CMS will be described in Section 3. Section 4 describes the reconstruction of an event at CMS as well as the physics objects used in the analysis. Section 5 discusses the analysis workflow from data samples used to signal extraction techniques, while the results are presented in Section 6. In Section 7, the concluding remarks are presented.

2. THEORETICAL FRAMEWORK

2.1 The Standard Model

Particle physics is the study of the fundamental constituents of matter and the forces between them. For more than 40 years these have been described by the so-called standard model of particle physics (SM), which aims to provide, at least in principle, a basis for understanding all known particle interactions. The SM currently fails to include gravity due to the difficult task of combining the quantum theory used to describe the microscopic world and the general theory of relativity. Furthermore, its theorized force carrier, the graviton has not been found experimentally.

The SM can be understood as arising from an underlying symmetry of the universe, which combines the theory of electroweak (EW) interactions and that of quantum chromodynamics (QCD). In mathematical terms, the SM is formed from the gauge groups $SU(3)_C \times SU(2)_L \times U(1)_{EM}$.

2.2 Structure and Particle Content

All the phenomena described by particle physics can be explained in terms of the properties and interactions of a small number of particles of four distinct types: two spin-1/2 families of fermions called leptons and quarks; one family of spin-1 bosons (called gauge bosons) which act as "force carriers", and a spin-0 particle, called the Higgs boson [22, 23]. We should note that all particles in the SM are assumed to be elementary, i.e. they do not have internal structure or excited states. See Figure 2.1 for reference.

In this section, the particle content of the SM will be introduced, along with the various force carriers. In the following sections, the specifics of particle-particle interactions will be explained in detail.

2.2.1 Fermions

Fermions are elementary particles with half-integer spin. They constitute the matter content of the SM, which accounts for 12 named fermions that interact via the weak and electromagnetic force (with the exception of neutrinos). Also, they obey Fermi-Dirac statistics and the Pauli exclusion

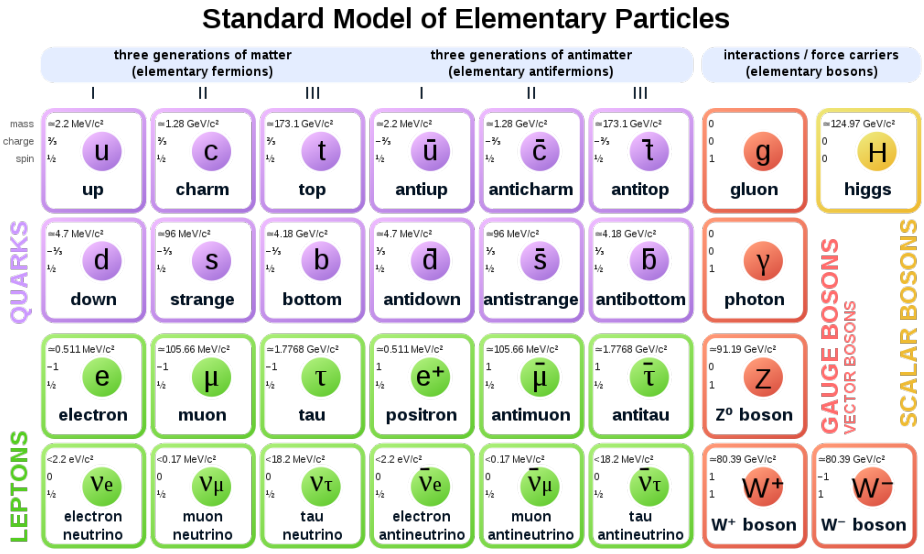

Figure 2.1: Particles of the Standard Model of particle physics. Reprinted from [1]

principle, meaning that no two fermions can occupy the same quantum state within a quantum system simultaneously.

2.2.1.1 Leptons

There are six known leptons, and they occur in pairs called generations, which we can write as doublets:

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}, \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}, \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix} \tag{2.1}$$

The three charged leptons (e^-, μ^-, τ^-) all have electric charge $Q = -q$. Associated with them are three neutral leptons, or neutrinos, called the electron neutrino, mu neutrino, and tau neutrino, respectively, all of which have very small masses. The six distinct types of leptons are also referred to as having different "flavors".

The charged leptons interact via both the electromagnetic and weak forces, whereas for neutral leptons only weak interactions have been observed. We should note that each generation of leptons has an associated quantum number. The electron number, which is defined for any state by

$$L_e \equiv N(e^-) - N(e^+) + N(\nu_e) - N(\bar{\nu_e}), \tag{2.2}$$

where $N(e^-)$ is the number of electrons present, and so on. For single-particle states, $L_e = 1$ for e^- and ν_e, $L_e = -1$ for e^+ and $\bar{\nu_e}$, and $L_e = 0$ for all other particles.

The form of Equation 2.2 also applies to the heavier lepton generations. Finally, in the SM, lepton numbers are individually conserved in all known interactions.

2.2.1.2 Quarks

Currently, there are six known quarks in the SM. Like the leptons, these six distinct types, or flavors, occur in pairs, or generations, denoted

$$\begin{pmatrix} u \\ d \end{pmatrix}, \begin{pmatrix} c \\ s \end{pmatrix}, \begin{pmatrix} t \\ b. \end{pmatrix} \tag{2.3}$$

Each generation consists of a quark with electromagnetic charge +2/3 (u,c, or t) together with a quark of charge -1/3 (d,s,b), in units of q. They are called the down(d), up(u), strange(s), charmed(c), bottom(b) and top(t) quarks. Each of these particles has an anti-particle version, with the same quantum numbers, but opposite charge. Furthermore, each quark also carries a *color charge* which can be red, green, or blue. This is a result of the strong force interaction of the quarks and will be explained in more detail in Section 2.3.3.

Quarks are known to bound to other quarks in states that we call hadrons. Hadrons can be bound states of two or three quarks called mesons and baryons, respectively. Recently, the LHCb collaboration reported the observation of a new type of hadron, a so-called *pentaquark*, which is a bound state composed of four quarks and one anti-quark[24].

2.2.2 Bosons

The SM bosons are the mediators of the interaction between the matter content of the SM, but also within themselves. They have integer spin quantum numbers and follow Bose-Einstein statistics, which means that they are not limited to single occupancy of the same quantum state.

There are 6 named bosons: the gluon, the photon, the $W^{\pm}$, and the Z, which have spin 1; and the Higgs boson, which corresponds to a scalar field and therefore has spin 0.

2.3 Particle Interactions

The interactions of the particles described in the previous section can be described in the mathematical framework of gauge field theory. Three of the four fundamental forces of nature are described in the SM (electromagnetism, the strong and the weak force). To each of these forces belongs a physical theory, its corresponding charge, (i.e. electric charge, color or flavor) and an associated boson as mediator.

Charges correspond to the time-invariant generators of a symmetry group, and specifically, to the generators that commute with the Hamiltonian. The invariance of the charge corresponds to the vanishing commutator

$$[Q, H] = 0 \tag{2.4}$$

for a given charge Q and Hamiltonian H. Thus charges are associated with conserved quantum numbers; which are the eigenvalues q of the generator Q [25].

Modern theories describe these forces in terms of quantum fields, namely quantum electrodynamics (QED), QCD and the unified electroweak quantum field theory. One feature all these theories have in common is that they are all gauge invariant. This is important because it is a fundamental requirement from which the detailed properties of the interaction are deduced, as we shall see later in this section.

To describe each of the three SM interactions or forces, we will start with a Lagrangian that describes the general dynamics of a given system of particles. Then we will study its invariance(variance) under a global(local) gauge transformation. We will see that in order to maintain gauge invariance after a local transformation, we will need to introduce additional gauge fields and their corresponding covariant derivatives. Finally, we will take a look at the conservation laws arising from the symmetry of the gauge invariance [26].

2.3.1 Quantum Electrodynamics

QED describes the dynamics of the electromagnetic interaction between fermions and the boson mediating the interaction, the photon. QED corresponds to the U_{EM} group and it was the first discovered example of gauge symmetry.

In QFT, particles are represented by fields[27], which are in turn represented mathematically by Lagrangian densities $\mathcal{L}$. If we start with the Lagrangian density for the Dirac spin-1/2 fermion field[28]

$$\mathcal{L} = \bar{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi \tag{2.5}$$

where γ^{μ} are the gamma matrices[29], ψ is a four-component column vector representing the wave function of a spin 1/2 particle (or Dirac spinor), $\bar{\psi} = \psi^{\dagger}\gamma^{0}$, and m is the mass of the particle. The Lagrangian is invariant under a global U(1) transformation of the form

$$\psi \rightarrow \psi' = e^{-i\alpha}\psi \tag{2.6}$$

while the parameter α is kept a constant. If instead, α is allowed to vary as a function of space-time, then Equation 2.6 becomes a local U(1) transformation and the Lagrangian density becomes

$$\mathcal{L} \rightarrow \mathcal{L}' = \mathcal{L} + \bar{\psi}\gamma^{\mu}(\partial_{\mu}\alpha(x))\psi \tag{2.7}$$

which is not invariant under the local transformation.

In order to restore local gauge invariance, a gauge field A_{μ} representing the photon and the covariant derivative

$$D_{\mu} = \partial_{\mu} + iqA_{\mu}, \tag{2.8}$$

where q (electric charge) are introduced. The new gauge field transforms as

$$A_\mu \to A'_\mu = A_\mu + \partial_\mu \chi(x), \tag{2.9}$$

where $\chi(x)$ is an arbitrary function of space-time. The covariant derivative has the same transformation properties as ψ and is chosen to replace ∂_μ.

After introducing these modifications, the Lagrangian takes the form:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{2.10}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the electromagnetic field strength tensor.

By looking at the resulting Lagrangian after the introduction of new gauge fields we can see that it does not include a mass term for the photon field A (i.e. no term proportional to $m^2 A_\mu A^\mu$). At this point, the theory posits an infinite range for the interaction (which is experimentally verified).

The final form of the Lagrangian includes lepton-photon interactions, as well as those in the form of $l^+ l^- \gamma$ and a quadratic term in the field strength tensor which is the photon kinetic energy. It can be generalized to include all the electromagnetically-charged fermions in the SM by taking the form

$$\mathcal{L} = \sum_i \bar{\psi}_i(i\gamma^\mu D_\mu - m_i)\psi_i - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{2.11}$$

where $i = e, \mu, \tau, u, d, c, s, t, b$.

2.3.2 Electroweak Interaction

The story of weak interactions starts with Henri Becquerel's discovery of radioactivity in 1896 and its subsequent classification into alpha, beta and gamma decays of the nucleus by Ernest Rutherford and others. But the real understanding of beta-decay in the sense we now know it came only after Enrico Fermi formulated a physical mechanism for such process in 1934.

The main ingredient for Fermi's theory had been provided by Wolfgang Pauli. To solve the puzzle of the continuous energy spectrum of the electrons emitted in the beta-decay of the nuclei,

Pauli had suggested that along with the electron, an almost massless neutral particle was also emitted. Fermi succeeded in incorporating Pauli's suggestion and thus was born the theory of weak interactions [30].

In the 1960s Glashow, Salam, and Weinberg had developed a theory[31, 32, 33] that unified electromagnetic and weak interactions in a way that is often compared to the unification of electric and magnetic interactions by Faraday and Maxwell a century earlier. This new theory made several remarkable predictions, including the existence of a neutral vector boson Z^0 and of weak reactions arising from its exchange.

The EW interaction is based on a local $SU(2)_L \times U(1)_Y$ gauge symmetry where L and Y are the generators of the symmetry. Here, electromagnetic and weak interactions are unified into a single non-abelian gauge theory. In order to understand this unification, we will start with a fermionic doublet representing an $SU(2)$ symmetry

$$\psi = \begin{pmatrix} \psi_1(x) \\ \psi_2(x) \end{pmatrix}, u_R, d_R \tag{2.12}$$

which transforms under the three dimensional rotation

$$\psi \rightarrow exp < i\alpha^i \frac{\sigma_i}{2} > \psi \tag{2.13}$$

which is the three dimensional version of Equation 2.6 and σ^i are the Pauli sigma matrices: the three non-commuting generators of the SU(2) transformations.

Just like in Section 2.3.1, we allow the parameter α to vary as a function of space-time so that

$$\psi(x) \rightarrow V(x)\psi(x), \tag{2.14}$$

where $V(x) = exp(i\alpha^i(x)\frac{\sigma^i}{2})$.

In order to keep the Lagrangian invariant under this transformation, we introduce additional fields. Since SU(2) has three generators there are also three gauge fields $A^i_\mu(x)$. The covariant

derivative for a SU(2) gauge invariant Lagrangian is

$$D_\mu = \partial_\mu - igA^i_\mu \frac{\sigma^i}{2} \tag{2.15}$$

and therefore

$$A^i_\mu(x)\frac{\sigma^i}{2} \to V(x)(A^i_\mu(x)\frac{\sigma^i}{2} + \frac{i}{g}\partial_\mu)V^\dagger(x) \tag{2.16}$$

To simplify this calculation, we can expand $V(x)$ to first order in α

$$A^i_\mu \frac{\sigma^i}{2} \to A^i_\mu \frac{\sigma^i}{2} + \frac{1}{g}(\partial_\mu \alpha^i)\frac{\sigma^i}{2} + i[\alpha^i \frac{\sigma^i}{2}, A^i_\mu \frac{\sigma^i}{2}] + ... \tag{2.17}$$

The covariant derivative will have the form

$$D_\mu \psi \to (1 + i\alpha^i \frac{\sigma^i}{2})D_\mu \psi \tag{2.18}$$

Due to the non-commutativity of the generators of this symmetry, the field strength tensor has an extra term

$$F^i_{\mu\nu} = \partial_\mu A^i_\nu - \partial_\nu A^i_\mu + g\epsilon^{ijk}A^j_\mu A^k_\nu \tag{2.19}$$

We can the construct the Yang-Mills Lagrangian

$$\mathcal{L} = -\frac{1}{4}(F^i_{\mu\nu})^2 + \bar{\psi}(i\gamma^\mu \partial_\mu - igA^i_\mu \frac{\sigma^i}{2})\psi \tag{2.20}$$

Now we introduce the local gauge invariance requirement for the Lagrangian and introduce new gauge fields with their associated covariant derivatives.

But first, we should note that the SM fermions possess a fundamental property called chirality, which describes how a given particle's wave function behaves under rotation. In the SM, the left-handed components of the electron neutrino and electron are grouped into an SU(2) doublet. Since

the right-handed component of the electron is invariant under SU(2), it is placed in a singlet, i.e.:

$$L_e = \begin{pmatrix} \nu_e \\ e_L \end{pmatrix}, e_R \tag{2.21}$$

And so on for the heavier generations of leptons.

Within the SM framework, neutrinos are weakly-interacting massless particles. As such, neutrinos wouldn't be able to change their handedness, but with mass, they can. Until now there is no experimental evidence for right handed neutrinos.

The kinetic energy term of the electroweak Lagrangian for first generation leptons can be represented by:

$$\mathcal{L}^e_{KE} = L^\dagger_e \tilde{\sigma}^\mu i\partial_\mu L_e + e^\dagger_R \sigma^\mu i\partial_\mu e_R \tag{2.22}$$

where $\sigma = (\sigma^0, \sigma^1, \sigma^2, \sigma^3)$, $\tilde{\sigma} = (\sigma^0, -\sigma^1, -\sigma^2, -\sigma^3)$, σ^0 is an identity matrix, and the σ^i are again the Pauli matrices. This Lagrangian is invariant under the global $SU(2)_L \times U(1)_Y$ transformation:

$$L \to L' = e^{i\theta} U L \tag{2.23}$$

$$e_R \to e'_R = e^{2i\theta} e_R \tag{2.24}$$

where

$$U = e^{-ia^k \sigma^k} \tag{2.25}$$

and θ and a^k are real numbers parameterizing the transformation.

Again, the Lagrangian is not invariant under a transformation where these parameters are allowed to vary as a function of space-time, i.e. a local transformation.

To restore invariance, we can introduce additional gauge fields and replace the space-time

derivatives with an appropriately chosen covariant derivative. This time, we introduce a U(1) gauge field $B_\mu(x)$ and three SU(2) gauge fields $W_\mu(x) = W_\mu^k(x)\sigma_k$. Such fields must transform as

$$B_\mu(x) \to B'_\mu(x) = B_\mu(x) + \frac{2}{g_1}\partial_\mu\theta(x) \tag{2.26}$$

$$W_\mu(x) \to W'_\mu(x) = U(x)W_\mu(x)U^\dagger(x) + \frac{2i}{g_2}(\partial_\mu U(x))U^\dagger(x) \tag{2.27}$$

where g_1 and g_2 are dimensionless parameters of the theory, the coupling strengths of the interactions. The necessary covariant derivatives are given by

$$D_\mu L_e = (\partial_\mu + i\frac{g_1}{2}YB_\mu + i\frac{g_2}{2}W_\mu)L_e \tag{2.28}$$

$$D_\mu e_R = (\partial_\mu + i\frac{g_1}{2}YB_\mu)e_R \tag{2.29}$$

where Y is the weak hypercharge operator, whose eigenvalues are listed in Table 2.1. The weak hypercharge values can be calculated as $Y = 2(Q - T_3)$, where T_3 is the third component of the weak isospin quantum number T.

	Particle	Q	T_3	Y	B	L
Quarks	$q_L = \begin{pmatrix} u \\ d \end{pmatrix}_L$	$\begin{pmatrix} 2/3 \\ -1/3 \end{pmatrix}$	$\begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix}$	1/3	1/3	0
	u_R	2/3	0	4/3	1/3	0
	d_R	-1/3	0	-2/3	1/3	0
Leptons	$l_L = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L$	$\begin{pmatrix} 0 \\ -1 \end{pmatrix}$	$\begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix}$	-1	0	1
	e_R	-1	0	-2	0	1

Table 2.1: Quantum numbers of the SM fermions

The covariant derivatives transform according to the same rule as the fields themselves. Combining the kinetic and gauge interaction terms of the Lagrangian yields

$$\mathcal{L} = \mathcal{L}_{KE} + \mathcal{L}_{gauge} = L_e^\dagger \tilde{\sigma}^\mu i D_\mu L_e + e_R^\dagger \sigma^\mu i D_\mu e_R - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \sum_{i=1}^{3} \frac{1}{4} W_{\mu\nu}^i W^{i\mu\nu} \tag{2.30}$$

where $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$ and $W_{\mu\nu} = [\partial_\mu + (\frac{ig_2}{2})W_\mu]W_\nu - [\partial_\nu + (\frac{ig_2}{2})W_\nu]W_\mu$ are the field strength tensors. This Lagrangian is now locally invariant.

The mediators of the electroweak force are the physical bosons $W^\pm$, the Z and the photon. All of them result from the combination of the newly introduced gauge fields as in the following way

- The $W^\pm$ are linear combinations of the W_1 and W_2, which are electrically charged and given by

$$W_\mu^\pm = \frac{W_\mu^1 \mp i W_\mu^2}{\sqrt{2}} \tag{2.31}$$

- The W_3 and B gauge fields are electrically neutral. The physical Z and photon are linear combinations of these fields, given by

$$Z_\mu = W_\mu^3 cos\theta_W - B_\mu sin\theta_W \tag{2.32}$$

$$A_\mu = W_\mu^3 sin\theta_W - B_\mu cos\theta_W \tag{2.33}$$

  where the Weinberg angle θ_W is defined by $sin\theta_W = g_1/\sqrt{g_1^2 + g_2^2}$.

The interactions contained in the Lagrangian couple the $W^\pm$ bosons to the left-handed lepton components only, unlike the photon and Z bosons which couple to both the left- and right-handed components.

Furthermore, we can now assign a value to the electromagnetic charge e proportional to the interaction strength

$$g_2 sin\theta_W = g_1 cos\theta_W = e. \tag{2.34}$$

Finally, in order to include second and third generation leptons, the leptonic portion of the Lagrangian generalizes to

$$\mathcal{L}^l = \sum_{leptons} (L_e^\dagger \tilde{\sigma}^\mu i D_\mu L_e + e_R^\dagger \sigma^\mu i D_\mu e_R) - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \sigma_{i=1}^3 \frac{1}{4} W_{\mu\nu}^i W^{i\mu\nu} \tag{2.35}$$

Quarks are included in the EW sector in a similar manner. The left-handed components of the u and d quark are placed in SU(2) doublets, and the right-handed components in singlets.

$$Q_u = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, u_R, d_R \tag{2.36}$$

The second and third generation quarks can be represented in the same way. The covariant derivatives acting on the quark fields have the same form as those which act on the lepton fields. Therefore, the dynamic portion of the u and d quark Lagrangian is given by

$$\mathcal{L}_{KE}^q = \sum_{quarks} Q_u^\dagger \tilde{\sigma}^\mu i D_\mu Q_\mu + u_R^\dagger \sigma^\mu i D_\mu u_R + d_R^\dagger \sigma^\mu i D_\mu d_R \tag{2.37}$$

Again, the W bosons couple only to the left-handed quark components, while the Z and photon couple to the right-handed components as well.

The full electroweak Lagrangian is a result of the addition of the lepton and quark kinetic components, as well as the gauge interaction component.

$$\mathcal{L}^{EW} = \mathcal{L}_{KE}^l + \mathcal{L}_{KE}^q + \mathcal{L}_{gauge} \tag{2.38}$$

Note that a U(1) transformation of the form $L_{e,\mu,\tau} \rightarrow e^{i\alpha} L_{e,\mu,\tau}$, $e, \mu, \tau_R \rightarrow e, \mu, \tau^{i\alpha} e, \mu, \tau_R$ leaves the EW Lagrangian invariant, which leads to conservation of lepton number. Additionally, a U(1) transformation multiplying all negatively (positively) charged fields by $e^{i\alpha}(e^{-i\alpha})$ leaves the Lagrangian invariant, and implies conservation of electric charge.

On the other hand, the EW Lagrangian is not invariant under charge conjugation C or a parity transformation P. Charge conjugation is the operation of exchanging all particles with antiparticles

and vice-versa. A parity transformation is the inversion of spatial coordinates, $r \to -r$. The neutral current interactions, mediated by the Z and photon, preserve combined CP invariance. However, combined CP symmetry is violated by weak current interactions, mediated by the $W^{\pm}$, in the quark sector. A third important potential symmetry is time reversal T, where $t \to -t$. Combined CPT invariance is required to maintain Lorentz invariance. Therefore, the breaking of CP also implied the breaking of T symmetry.

Although CP is not conserved, there is good reason to believe that all interactions are invariant under the combined operation of CPT, taken in any order. This result is called the CPT theorem[34] and can be shown to hold in any relativistic quantum theory in which signals cannot propagate faster than the speed of light.

2.3.3 Strong Interaction

QCD is the theory that describes the interaction between quarks via the strong force. It is represented by a local $SU(3)_C$ gauge symmetry and the interaction mediator is the gluon.

Associated with the $SU(3)_C$ symmetry are several conserved quantum numbers, called color charges, which play a similar role in strong interactions to that played by e in electromagnetic interactions. Color charges can be green, red, and blue but only color neutral (or colorless) states have been observed in nature. Baryons contain equal parts of each color and mesons contain color-anticolor pairs.

In QCD, quarks are represented in this theory as color triplets

$$q_u = \begin{pmatrix} u_r \\ u_g \\ u_b \end{pmatrix} \tag{2.39}$$

and gluons contain two color charges. The eight known combinations of color charges for the gluon are represented by eight gauge fields that are a direct consequence of the 8 non-abelian generators of SU(3), the Gell-Mann matrices[35].

As in the previous sections, we start building the interaction from an $SU(3)$ Lagrangian that is

globally invariant in the form

$$\mathcal{L}^q_{QCD} = \sum_{i=1}^{6} \bar{q}_i i\gamma^\mu \partial_\mu q_i \tag{2.40}$$

This Lagrangian is invariant under a transformation of the form $q_i \rightarrow q_i' = Uq_i$ where U is a member is a member of $SU(3)$. If we allow for a transformation of the form $U(x)$, the Lagrangian is no longer invariant. To return invariance, we introduce 8 gauge fields ($G_\mu(x)$) which represent the gluon fields and an appropriate covariant derivative. They will transform as

$$G_\mu \rightarrow G'_\mu = UG_\mu U^\dagger + \frac{i}{g_s}(\partial_\mu U)U^\dagger \tag{2.41}$$

$$D_\mu q_i = (\partial_\mu + ig_s G_\mu)q_i \tag{2.42}$$

where g_s is the dimensionless coupling strength of the color interaction.

The field strength tensor for QCD is:

$$G_{\mu\nu} = \partial_\mu G_\nu - \partial_v G_\mu + ig_s(G_\mu G_\nu - G_\nu G_\mu) \tag{2.43}$$

and the locally $SU(3)$ gauge invariant QCD Lagrangian is then given as:

$$\mathcal{L}^q_{QCD} = \sum_{i=1}^{6}(\bar{q}_i i\gamma^\mu D_\mu q_i) - \frac{1}{4}\sum_{i=1}^{8} G^i_{\mu\nu} G^{i\mu\nu} \tag{2.44}$$

In contrast to the EW interaction, C,P, and T are all conserved. The range of the strong force interaction is about 10^{-15} m, which is enough to act on nucleons, i.e. protons and neutrons to form atomic nuclei.

Finally, QCD is a strongly coupled theory at low energies and large distance scales, but weakly interacting at high energies and small distance scales. This fact is responsible for the hadronic bound states of quarks. Moreover, unlike QED, its mediator, the gluon, interacts with itself. At low energy scales, i.e. the non-perturbative regime, QCD calculations are extremely difficult and

techniques such as lattice gauge theory must be exploited. On the other hand, at a high energy scale, or equivalently small distance scales, the strong interaction becomes weakly interacting and quarks are effectively free. In this regime the usual techniques of perturbation theory can be used, allowing high-precision calculations.

2.3.4 Brout-Englert-Higgs Mechanism and the Higgs Boson

As we have seen from the previous section, the EW and QCD Lagrangians do not contain any mass terms. Gauge invariance seems to imply that the spin-1 gauge bosons have zero masses. This is acceptable for QED and QCD, where the gauge bosons are the photons and the gluons, which do indeed have zero mass. However, the $W\pm$ and Z^0 bosons are very heavy, and therefore, not massless as they would if gauge invariance was exact.

This problem is overcome by introducing the Brout-Englert-Higgs(BEH) mechanism which postulates that the various particles in the SM interact with a new type of scalar field, the Higgs field(s). This field differs from others in its behavior in the so-called vacuum state by having a non-zero value, unlike the other fields introduced previously. The value v is not invariant under a gauge transformation, and will spontaneously break the symmetry of the Lagrangian in a process we will refer to as spontaneous electroweak symmetry breaking (EWSB).

The Goldstone theorem postulates that for every spontaneously broken continuous symmetry there will be a new massive scalar "Goldstone" boson. The number of new bosons will be equal to the number of broken generators of the symmetry group. The massless SM bosons then acquire mas by absorbing these Goldstone bosons.

The BEH mechanism is also used to generate mass for the quarks and electrically charged leptons. The neutrinos, photon, and gluons remain massless, as observed experimentally.

Remember from previous section that there are four massless electroweak gauge bosons, W^1, W^2, W^3, and B^0. The experimentally observed bosons, however, are the massless photon, and three massive bosons (the $W^\pm$ and Z). We also know that electric charge Q is conserved in EW interactions. This means that the $SU(2)_L \times U(1)_Y$ EW theory is broken such that a new $U(1)_{EM}$ symmetry group is formed which corresponds to electromagnetism.

In order for three gauge bosons to acquire mass they must absorb three Goldstone bosons. The simplest method to accomplish this is to introduce a complex, scalar $SU(2)$ doublet Φ with hypercharge $Y = 1$.

$$\Phi = \begin{pmatrix} \Phi_A \\ \Phi_B \end{pmatrix} = \begin{pmatrix} \phi_1 \\ i\phi_2 \\ \phi_3 \\ i\phi_4 \end{pmatrix}, \tag{2.45}$$

The part of the SM Lagrangian which includes the EW gauge bosons and the leptons can be written as

$$\mathcal{L}_{SM} = -\frac{1}{4} W^a_{\mu\nu} W^{\mu\nu}_a - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + \bar{L}_i (iD_\mu \gamma^\mu) L_i + \bar{e}_{R,i} (iD_\mu \gamma^\mu) e_{R,i} \tag{2.46}$$

where i runs over the three generations, μ and ν are Lorentz indices, and a runs over the generators in the gauge group. The field strengths are given by

$$W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g_2 \epsilon^{abc} W^b_\mu W^c_\nu \tag{2.47}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{2.48}$$

and the covariant derivatives for the left- and right-handed leptons are

$$D_\mu L_L = (\partial_\mu - i g_2 T_a W^a_\mu - i g_1 Y B_\mu) L_L \tag{2.49}$$

$$D_\mu e_R = (\partial_\mu - i g_1 Y B_\mu) e_R \tag{2.50}$$

where T_a are the generators of the $SU(2)_L$ gauge group and g_1, g_2 are the coupling constants for the EW interaction.

The scalar part of the Lagrangian required by the addition of a scalar field is then

$$\mathcal{L}_S = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi^\dagger \Phi) \tag{2.51}$$

where the first term is the kinetic term and the second term is the scalar potential. While the form of the scalar potential is not known from first principles, we can make the assumption that it takes the form

$$V(\Phi^\dagger\Phi) = \mu^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2 \tag{2.52}$$

The value of λ must be positive in order for the vacuum to be stable. The sign of μ^2 specifies one of two cases for the potential.

- When $\mu^2 > 0$, the potential $V(\Phi)$ is always positive and has a minimum at

$$< 0|\Phi|0 > \equiv \Phi_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{2.53}$$

  where no spontaneous symmetry breaking can occur.

- When $\mu^2 < 0$ the potential has a minimum value not located at the origin. In this case, the neutral component of the scalar field will acquire a vacuum expectation value v

$$< 0|\Phi|0 > = \Phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{2.54}$$

  where $v = \sqrt{\frac{-\mu^2}{\lambda}}$

By only adding a vev to the neutral component of the scalar field, electromagnetism is unbroken and the $U(1)_{EM}$ symmetry keeps a conserved electric charge $Q = T_3 + \frac{Y}{2}$.

We can then expand the scalar field Φ around the minimum Φ_0 to get

$$\Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{2.55}$$

where $h(x)$ is a new scalar field.

Next we insert this field into the kinetic part of the Lagrangian and redefine the gauge fields as

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) \tag{2.56}$$

$$Z_\mu = \frac{1}{\sqrt{g_1^2 + g_2^2}}(g_2 W_\mu^3 - g_1 B_\mu) \tag{2.57}$$

$$A_\mu = \frac{1}{\sqrt{g_1^2 + g_2^2}}(g_2 W_\mu^3 + g_1 B_\mu) \tag{2.58}$$

which correspond to the observed gauge bosons.

After this the covariant derivative becomes

$$|D_\mu \Phi|^2 = \frac{1}{2}(\partial_\mu H)^2 + \frac{1}{2}g_2^2(v+H)^2 W_\mu^+ W^{\mu-} + \frac{1}{8}(v+H)^2(g_1^2+g_2^2)Z_\mu Z^\mu \tag{2.59}$$

From here we can see that the photon A_μ remains massless, but that the mass terms for the W and Z bosons take the general forms $M_W^2 W_\mu W^\mu$ and $M_Z^2 Z_\mu Z^\mu/2$ respectively.

Thus the masses of the electroweak gauge bosons are

$$M_W = \frac{1}{2}vg_2 \tag{2.60}$$

$$M_Z = \frac{1}{2}v\sqrt{g_1^2 + g_2^2} \tag{2.61}$$

$$M_A = 0 \tag{2.62}$$

Three of the degrees of freedom from the scalar field, which would have been two charged and one neutral Goldstone boson, have been absorbed by the gauge bosons in order to give them mass. There is one remaining degree of freedom, an oscillation in the radial direction of the scalar potential, which corresponds to the neutral Higgs boson.

Finally, fermions acquire mass by adding couplings between the fermion fields and the scalar

field to the SM Lagrangian. The part of the Lagrangian that corresponds to the first generation fermions is given by

$$\mathcal{L}_F = -G_e \bar{L} \Phi e_R - G_d \bar{Q} \Phi d_R - G_u \bar{Q} \tilde{\Phi} u_R + h.c. \tag{2.63}$$

where $\tilde{\Phi} = i\tau_2 \Phi^*$ is the conjugate of Φ with negative hypercharge.

There are additional terms added to the full Lagrangian which correspond to the second and third generations which are not shown here.

By substituting Φ into the previous Lagrangian we find

$$\mathcal{L}_F = -\frac{1}{\sqrt{2}}(v + H)(G_e \bar{e}_L e_R + G_d \bar{d}_L d_R + G_u \bar{u}_L u_R) + h.c. \tag{2.64}$$

where $h.c.$ is a placeholder for the hermitian conjugate terms.

The fermion masses take the form $m\bar{f}_L f_R + h.c.$, which means that the fermion masses for the first generation are

$$m_e = \frac{G_e v}{\sqrt{2}}, m_u = \frac{G_u v}{\sqrt{2}}, m_d = \frac{G_d v}{\sqrt{2}} \tag{2.65}$$

The second and third generations have similar mass terms. For the case of the neutrinos, since there is no right handed neutrino in the SM the neutrinos that do exist remain massless.

Finally, the coupling constants, G, and the fermion masses are not predicted by the SM, so they must be measured and added to the model.

2.4 Beyond the Standard Model

The SM evolved in response to a series of experimental discoveries over a period of several decades, and it turned out to be a remarkably successful theory. At the present time, provided non-zero neutrino masses are incorporated, all experimental observations in particle physics are consistent with the SM, but there is no reason to suppose that there will not be more surprises in

the future, as higher energy regions are explored.

Also, there are a few experimental facts which suggest that the SM may not be a complete theory of nature. For example, there is strong evidence that the particles of the SM can only account for a small fraction of the matter in the Universe, and the observed predominance of matter over antimatter cannot be understood in the framework of the SM.

Moreover, there is still an incompatibility between general relativity (GR) (which can be thought of as the theory of gravitation) and the SM (the theory that describes the other three fundamental forces) because space-time is not quantized in GR.

Finally, the SM itself embodies many assumptions and more than twenty free parameters, giving rise to many questions like

- Can the number of parameters be reduced?
- Why are there three generations of quarks and leptons, rather than just the one that is required to describe "ordinary matter", i.e. the neutrons and protons?
- Are the quarks really point-like particles, or will they turn out to be composite when we are able to explore a higher energy regime?
- Why does the weak interaction violate CP invariance, but not the strong interaction?

Many theories have been proposed to try to answer these and other questions, and a few experimental programs have been set up to test them.

2.5 Lepton universality

All known experimental data are consistent with the assumption that the interactions of the electron and its neutrino are identical with those of the muon and its associated neutrino and the tau and its neutrino, provided the mass differences are taken into account. This fundamental assumption is called the universality of lepton interactions.

We will illustrate universality of this rule by looking at the leptonic decays [36]

$$\mu^+ \to e^+ + \nu_e + \bar{\nu}_\mu, \tag{2.66}$$

$$\mu^- \to e^- + \bar{\nu}_e + \nu_\mu, \tag{2.67}$$

$$\tau^- \to \mu^- + \bar{\nu}_\mu + \nu_\tau, and \tag{2.68}$$

$$\tau^- \to e^- + \bar{\nu}_e + \nu_\tau \tag{2.69}$$

of the muon and tau leptons at rest.

To simplify the calculation, we will work to lowest order only and we will use the zero-range approximation (a zero-range point interaction with strength equal to the Fermi constant $G_F = 1.66 \times 10^{-5} GeV^{-2}$), since the masses of the leptons are very small compared with the rest energy of the W bosons mediating the weak interaction.

We start by considering the muon decay whose rate has the form (in the zero-range approximation)

$$\Gamma(\mu^- \to e^- + \bar{\nu}_e + \nu_\mu) = K G_F^2 m_\mu^5 \tag{2.70}$$

since we are assuming the electron and neutrino masses are zero. Here, K is a dimensionless constant whose value will depend on the precise form of the interaction. If we assume this is the same for muon and tau leptons the same argument gives

$$\Gamma(\tau^- \to e^- + \bar{\nu}_e + \nu_\tau) = K G_F^2 m_\tau^5 \tag{2.71}$$

Likewise, $e - \mu$ universality gives

$$\Gamma(\tau^- \to e^- + \bar{\nu}_e + \nu_\tau) = \Gamma(\tau^- \to \mu^- + \bar{\nu}_\mu + \nu_\tau) \tag{2.72}$$

This explains why the experimental branching ratios for the two leptonic decay modes of the tau lepton are, to a good approximation, equal. A full calculation, taking into account final state

masses, gives the ratio $\Gamma(\tau^- \to \mu^- + \bar{\nu}_\mu + \nu_\tau)/\Gamma(\tau^- \to e^- + \bar{\nu}_e + \nu_\tau) = 0.973$, whereas the experimental value is 0.976 ± 0.003.

It also gives a relation between the μ and τ lifetimes[1]

$$\tau_l = \frac{1}{\Gamma_{tot}} = \frac{B(l^- \to e^- \bar{\nu}_e \nu_l)}{\Gamma(l^- \to e^- \bar{\nu}_e \nu_l)} \tag{2.73}$$

where l can be the μ or τ lepton and Γ_{tot} is the total decay width and therefore

$$B(l^- \to e^- \bar{\nu}_e \nu_l) = \frac{\Gamma(l^- \to e^- \bar{\nu}_e \nu_l)}{\Gamma_{tot}} \tag{2.74}$$

is the branching ratio. Experimentally, $B = 1$ and 0.1783 ± 0.0004 for $l = \mu$ and τ, respectively[37]. Thus from 2.66and 2.70 we have

$$\frac{\tau_\tau}{\tau_\mu} = \frac{B(\tau^- \to e^- \bar{\nu}_e \nu_\tau)}{B(\mu^- \to e^- \bar{\nu}_e \nu_\mu)} (\frac{m_\nu}{m_\tau})^5 = (1.326 \pm 0.003) \times 10^{-7} \tag{2.75}$$

This agreement, involving lifetimes that differ by seven orders of magnitude, is impressive evidence of the universality of lepton interactions.

2.6 B-hadron anomalies

So far, no definite violation of lepton universality has been observed. However, the wealth of data on rare leptonic and semi-leptonic b hadron decays that have been accumulated at the LHC so far seem to challenge the rule.

In particular, current data on rare $b \to sll$ decays show an intriguing pattern of deviations from the SM predictions both for branching ratios [38], [39],[40],[41],[42] and angular distributions[43], [44], [45].

The latest global fits find that data consistently points with high significance to a non-standard effect that can be described by a four-fermion contact interaction[46]

[1]Note the use of natural units ($\hbar$ = c = 1)

$$C_9(\bar{s}\gamma^\nu P_L b)(\bar{\mu}\gamma_\nu \mu). \tag{2.76}$$

Nonetheless, the main obstacle towards conclusively establishing a beyond-SM effect is the inability to exclude large hadronic effects as the origin of the apparent discrepancies. In this respect, observables in $b \to sll$ transitions that are practically free of hadronic uncertainties are of particular interest. Among them are lepton flavor universality ratios such as the branching ratios $R_K = \frac{BR(B^+ \to K^+ \mu^+ \mu^-)}{BR(B^+ \to K^+ e^+ e^-)}$ and $R_{K^*} = \frac{BR(B^+ \to K^* \mu^+ \mu^-)}{BR(B^+ \to K^* e^+ e^-)}$.

In the SM, the only sources of lepton flavor universality violation are the leptonic Yukawa couplings, which are responsible for both the charged lepton masses and their interactions with the Higgs. However, Higgs interactions do not lead to any observable effects in rare b decays and lepton mass effects become relevant only for a very small dilepton invariant mass squared (q^2) close to the kinematic limit $q^2 \sim 4m_l^2$.

Over a broad range of q^2 the SM accurately predicts $R_k = R_{k^*} = 1$, with theoretical uncertainties on the order of 1%[47] which contradict experimental results which show a deviation from the expected SM value in the 2.4-2.6 σ range (Figure 2.2. A more recent study[48], [49] combined the results for R_K and R_{K^*}, resulting in a 4σ deviation from the SM.

2.6.1 $b \to s$ quark transitions

As we have seen from the previous section, over the last few years, many observables related to the flavour-changing neutral-current (FCNC) transitions $b \to l^+ l^-$ have exhibited important deviations from SM expectations. Therefore, in this section, we will take a closer look at these transitions and their sensitivity to potential new physics.

A $b \to s$ quark transition is an example of a FCNC process[50]. In such process, the s and b quark interact via a quantum-loop transition involving predominantly a W boson and either and up, charm, or top quark as shown in Figure 2.3.

Within the SM, the lowest order processes that could mediate the $b \to s$ quark transitions are at least of third order and are suppressed by angular momentum conservation and by the chiral nature

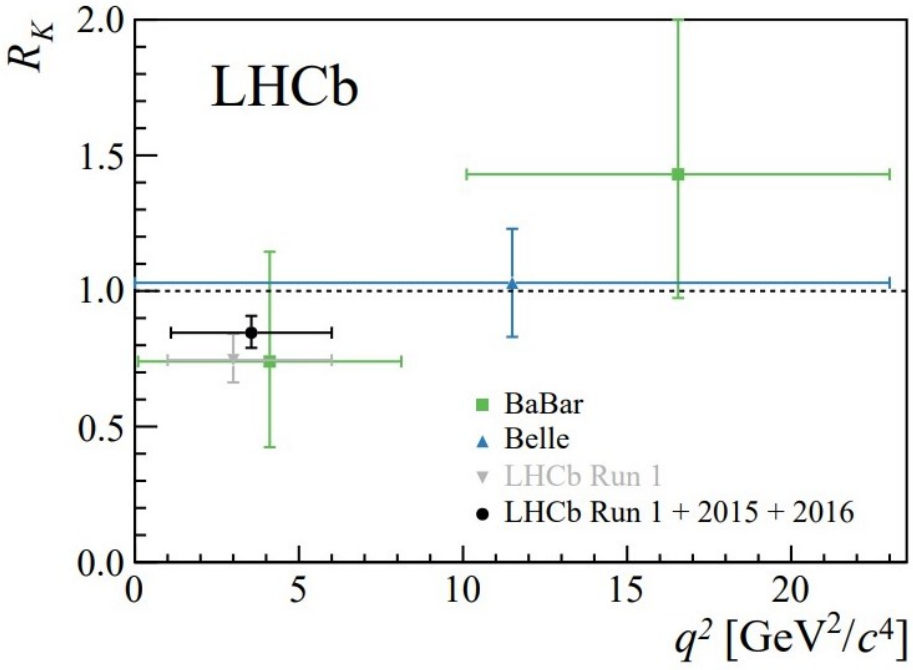

Figure 2.2: Summary of the $R_{K^{(*)}}$ measurements performed at the B-factories and by the LHCb experiment. Results are presented using different colored markers. Reprinted from [2]

of the weak force. This suppression is not necessarily present for new-physics particles and that is what makes the study of this decay of particular interest in probing for physics beyond the SM.

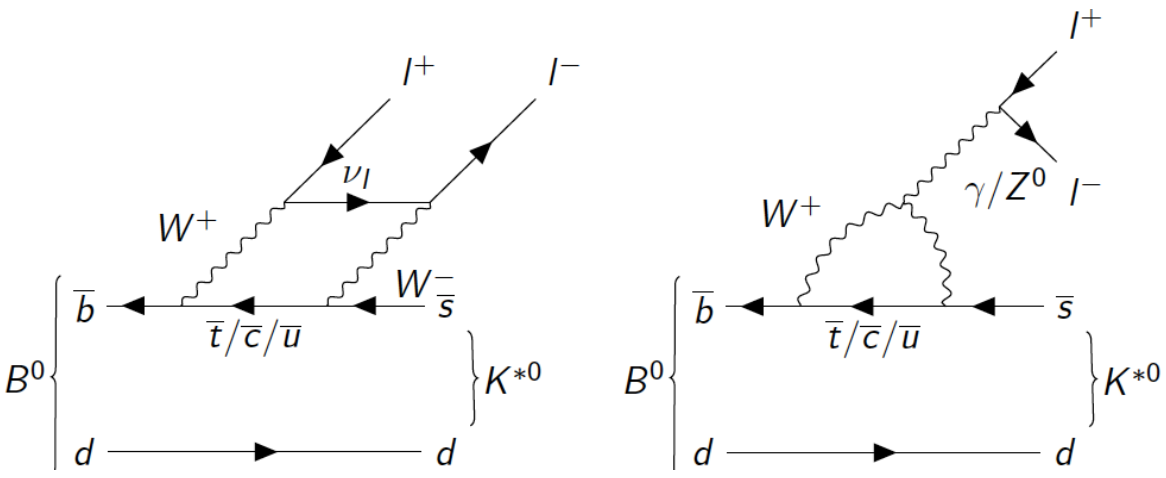

Figure 2.3: Lowest order Feynmann diagrams for $b \to s$ quark transition. Reprinted from [3]

2.7 The Z'

There is a multitude of new-physics models that could explain some or all of the b decay anomalies. One example of an extension to the SM that could explain these deviations involves a heavy version of the Z boson, denoted Z'[3]. Such an extension to the model must satisfy direct searches for such particles at the CMS[51], [52] and ATLAS [53] experiments, which in practice means either that the Z' candidate must be at least 30 times heavier than the SM Z boson, or that it must have small couplings to the up and down quarks. If the Z' is very heavy, it would not have a sizable impact on the decay compared with the SM contribution, unless it could change the flavor of quarks directly without going through a quantum-loop transition.

A generic framework of a minimal extension to the SM which explains B anomalies has been introduced in [54] and we collect here only those formulae of that paper that are essential for our study. The new physics contribution to rare B decays can be described by the following effective Lagrangian

$$\mathcal{L} \supset \frac{4G_F}{\sqrt{2}} V_{tb} V_{ts}^* \frac{e^2}{16\Pi^2} C_9 O_9 + h.c. \tag{2.77}$$

where C_9 is a Wilson coefficient and the effective operator O_9,

$$O_9 = (\bar{s}\gamma_\mu P_L b)(\bar{\mu}\gamma^\mu \mu) \tag{2.78}$$

describes a four-fermion interaction, with a left-handed $b - s$ current and a vector current for μ. To fit the current data [55], the new physics contribution to C_9 needs to be $-1.59^{+0.46}_{-0.56}$.

In this model, an extra $U(1)$ gauge group has been introduced, resulting in a new gauge boson, the Z'. This newly introduced particle would have a flavor changing quark coupling δ_{bs} and a nonuniversal lepton coupling.

Including the contribution from Equation 2.77, the dominant terms in the Lagrangian that are allowed by all the existing constraints in order to address the anomalies are then

$$\mathcal{L} \supset Z'_\mu[g_\mu \bar{\mu}\gamma^\mu \mu + g_\mu \bar{\nu}_\mu \gamma^\mu P_L \nu_\mu + g_b \sum_{q=t,b} \bar{q}\gamma^\mu P_L q + (g_b \delta_{bs} \bar{s}\gamma^\mu P_L b + h.c.)] \tag{2.79}$$

The Z' mass is constrained to be less than 5.5(10) TeV in the 1(2)σ range to explain the B anomalies[56]. As the mass gap between the Z' and the SM Z becomes smaller, interference problems start to arise and becomes harder to probe at the LHC. Therefore, for this analysis the lower bound in the search is 250 GeV.

2.7.1 4b Bottom Fermion Fusion

As we can conclude from Equation 2.79, the Z' does not significantly couple to first or second generation quarks, which could explain why it has not been observed experimentally yet. However, the Z' can be produced through its couplings to b quarks originating either from sea quarks, or gluon-splitting.

Therefore, the Z' is associated either with two,one, or no b-jets depending on the number of quarks from gluon splitting. The Z' can decay into pairs of b quarks, muons, muon neutrinos, and, if kinematically allowed, top quarks.

The relevant final states at the LHC are dimuon or di-b resonances. The cross sections behave as follows:

$$\sigma(pp \to Z' \to \mu\mu) \propto 2g_b^2(1 + k\delta_{bs}^2)g_\mu^2 \tag{2.80}$$

$$\sigma(pp \to Z' \to b\bar{b}) \propto 3g_b^4(1 + k\delta_{bs}^2) \tag{2.81}$$

where δ_{bs} regulates the possible production of Z' through b-s quark fusion, and where k contains the s-quark PDF contributions.

For this analysis, the special case when the Z' is associated with two b-jets coming from gluon splitting (a process we will refer to as Bottom Fermion fusion (BFF) (see Figure 2.4)) and has a

di-b jet final state is considered.

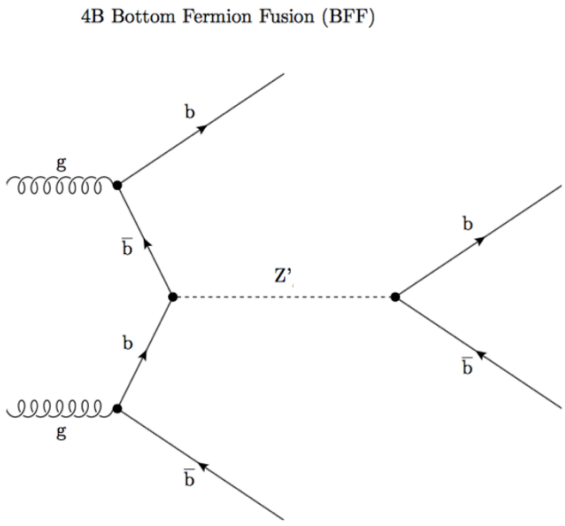

Figure 2.4: Feynman diagram for bottom fermion fusion (BFF).

2.7.2 Flavour-violating coupling δ_{bs}

In order to provide an explanation for B-decay anomalies, we need to consider the flavour-violating coupling δ_{bs}. Allowing the Z' boson to couple to s quarks in addition to b quarks results in two times more ways to produce the Z' and two times more ways for it to decay. A non-zero δ_{bs} will allow the Z's to be produced by b and $\bar{s}$ quarks (in addition to $b\bar{b}$ ones) and this significantly enhances the production cross section since the PDF for the s quark is significantly higher than that for the b quark at the LHC, as we can see from Figure 2.5.

Also, we can see from Equation 2.81 and 2.80 that when δ_{bs} goes to zero, the flavor conserving contribution dominates the production of Z'. Likewise, when δ_{bs} is large but still satisfies the B anomalies (so smaller g_b) the flavour violating contribution dominates.

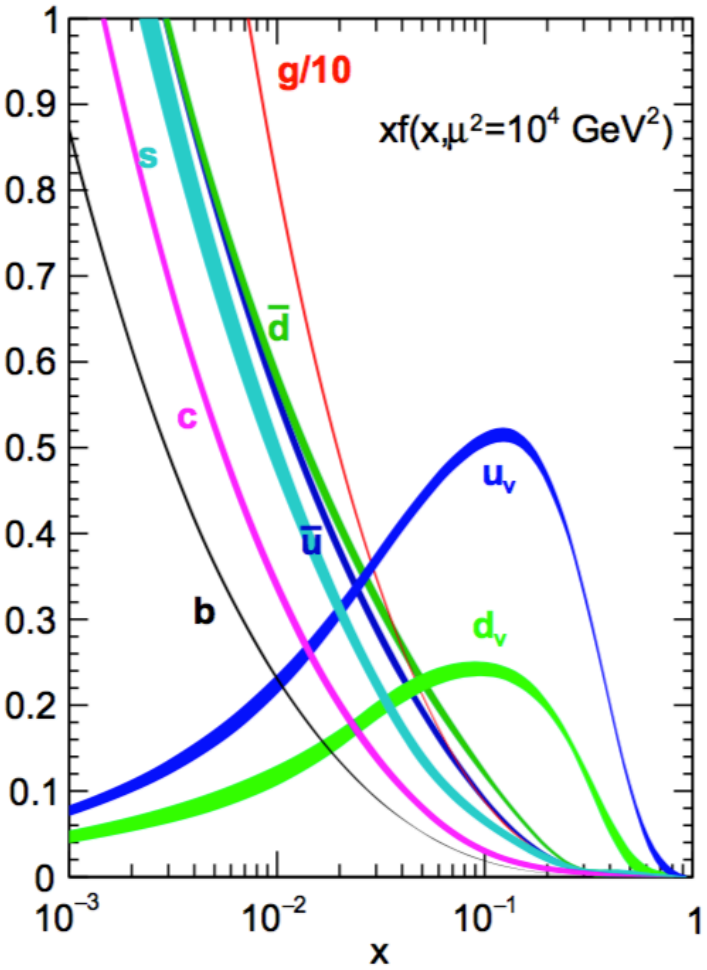

Figure 2.5: Parton distribution functions at the $\mu^2 = 10^4 GeV^2$ mass scale. Here, the vertical axis is the number density of the parton. Effectively it's a probability density, but normalized to the expected number of a given parton in the proton. The width in each line is the uncertainty in the PDF for that particular parton. Caption and figure reprinted from [4].

3. THE LHC AND CMS DETECTOR

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [57] experiment at the European Organization for Nuclear Research (CERN) is the world's largest and most powerful particle accelerator in operation today. Located at the border between Switzerland and France, it consists of a 27-km circumference ring of superconducting magnets and accelerating structures. See Figure 3.1 for reference.

Within the ring, protons are accelerated to a speed close to that of light and made to collide at 4 points:

- CMS (Compact Muon Solenoid) [58],
- ATLAS (A Toroidal LHC ApparatuS) [59],
- ALICE (A Large Ion Collider Experiment) [60],
- and LHCb (Large Hadron Collider beauty) [61].

ATLAS and CMS are two general-purpose particle detectors located at opposite sides on the LHC ring. These are "onion-type" detectors in the sense that their general layout surrounds the interaction point with sub-detector systems aimed to measure a specific property of the particle to be detected.

The other two detectors, ALICE and LHCb are designed for specific purposes, like studying heavy-ion collisions and performing precision measurements of CP-violation and the physics of B-mesons, respectively. For this study, data collected by the CMS experiment is used.

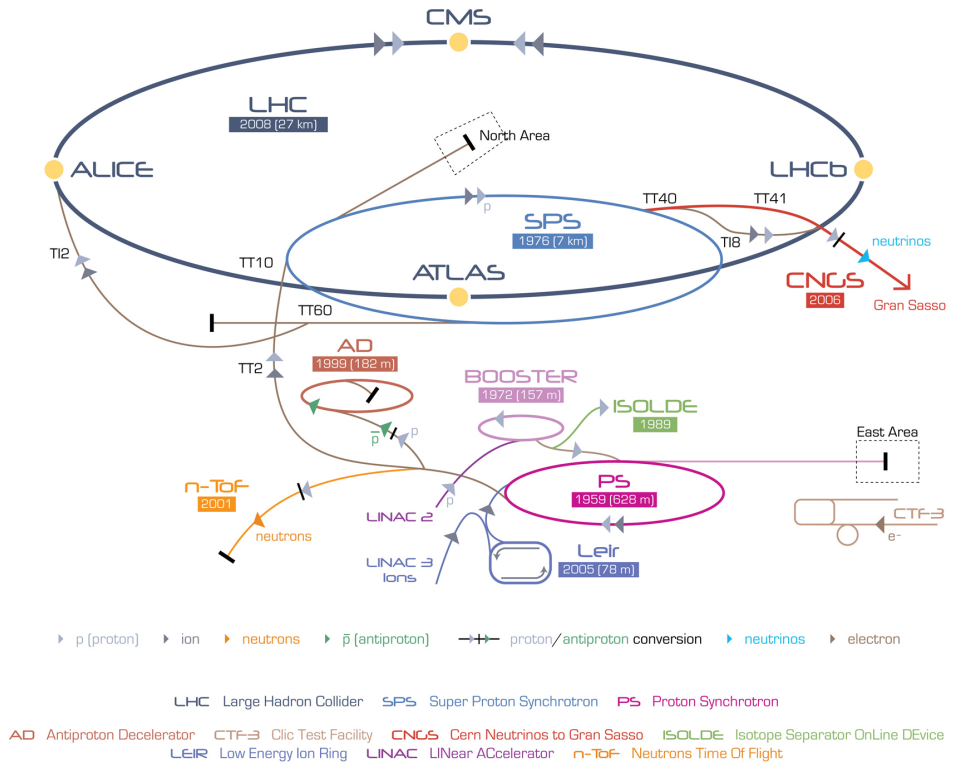

Figure 3.1: Schematic diagram for the LHC experiment at CERN. Reprinted from [5].

According to Einstein's famous equation $E = mc^2$, energy and mass are interchangeable. Therefore in order to produce heavy particles, a large amount of energy is required. The LHC was designed to produce highly energetic proton-proton, lead-proton or lead-lead beam collisions in which a variety of elementary particles can be produced.

The beams circulating the LHC ring are not continuous streams of particles, but rather trains of regularly spaced proton bunches. The experiment was designed to operate with 2,808 bunches of protons per beam, containing about 1.5×10^{11} protons per bunch separated by 25 ns, corresponding to a collision frequency of 40 MHz.

The LHC started operation on November 23, 2009. Through its 2010-2011 run, the LHC operated at a center-of-mass energy of 7 TeV. Then in 2012, the energy was increased to 8 TeV, and again to 13 TeV in 2015, after a shutdown in 2013 that lasted two years.

3.2 The CMS Detector

The CMS detectors was designed with the goal of identifying the particles coming out of the proton-proton collisions as well as to characterize their momentum, position, and trajectory at the moment of the collision. The goals of the CMS physics program range from studying the SM (including the Higgs boson) to searching for extra dimensions and dark matter. It even has a very successful heavy ion program.

In particular, the central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. The solenoidal volume contains a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (EMCAL) and a brass and scintillator hadron calorimeter (HCAL). Each layer of the detector exploits a property of the particle to be detected to measure its energy, momentum, position, and direction.

The CMS detector was not built on site like other giant detectors of the LHC experiment, but it was constructed in 15 sections at ground level before being lowered into an underground cavern near Cessy in France and then reassembled. The complete detector is 21 m long, 12 m wide and 15 m high.

The layout of the detector can be seen in Figure 3.2. The following sections will describe each of the sub-detectors and its properties.

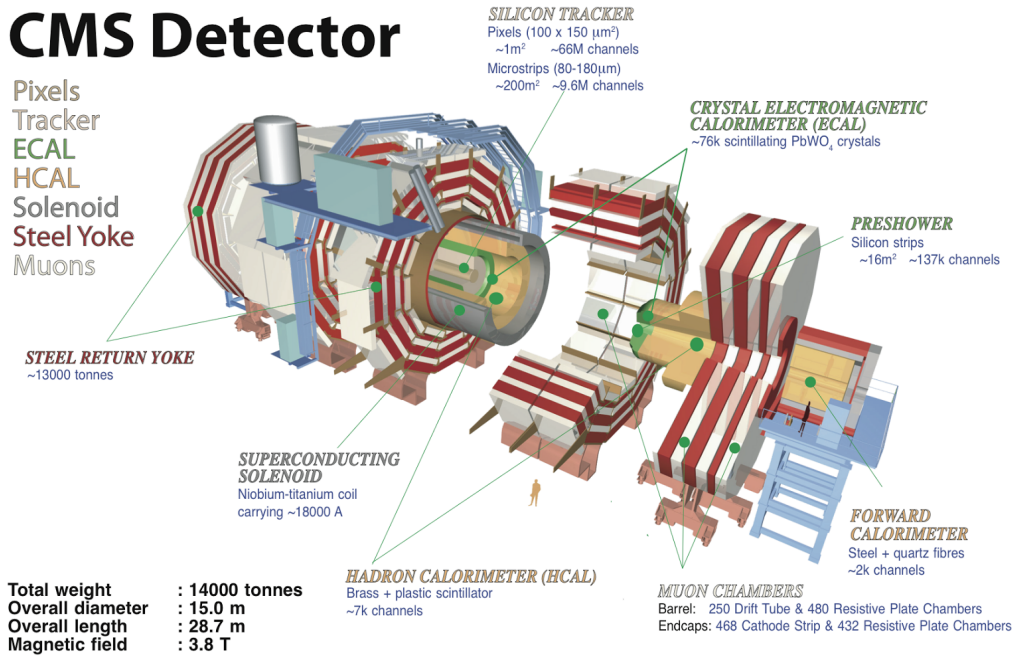

Figure 3.2: Schematic diagram for the CMS experiment with its sub-detector systems and a person for scale. Reprinted from [6].

3.2.1 Coordinate System

The CMS experiment uses a right-handed coordinate system, with the origin at the nominal collision point, the x-axis pointing to the center of the LHC ring, the y-axis pointing up (perpendicular to the LHC plane), and the z-axis along the anticlockwise beam direction. The polar angle θ is measured from the positive z-axis and the azimuthal angle ϕ is measured from the positive x-axis in the $x-y$ plane. The radius r denotes the distance from the z-axis and the pseudorapidity η is defined as $\eta = -log[tan(\theta/2)]$. η is preferently used by CMS particle physicists to measure forward-ness of relativistic particles in the detector since any differences in this coordinate are invariant under boosts in the $z-$direction and particle production is roughly uniform in η. See Figure 3.3 for reference.

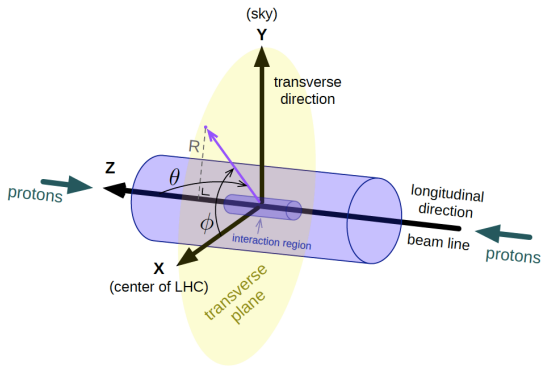

Figure 3.3: Diagram for the CMS detector coordinate system. Reprinted from [7]

3.2.2 Solenoid

The CMS magnet[62] is one of the main features of the experiment. It delivers a 4T magnetic field, which is 100,000 times stronger than that of Earth over a length of 12.5 m and a free-bore radius of 3.15 m.

Its job is to bend the paths of charged particles emerging from high-energy collisions in the LHC. The higher momentum particles get their path curved less than the lighter ones, and as a result, curvature is an important tool for momentum measurements.

The strong magnetic field, combined with the high-precision position measurements in the tracker and muon detectors, allows for accurate measurement of the momentum of high-energy particles.

The CMS solenoid magnet is made of coils of wire that produce a uniform magnetic field when electricity flows through them.It is the largest superconducting magnet ever built, weighting about 12,000 tonnes. In order for it to be superconducting, it needs to be cooled down to -268.5 C, which is a degree warmer than outer space.

The tracker and calorimeter detectors fit inside the magnet while the muon detectors are interleaved with a 12-sided iron structure that surrounds the magnet coils and contains and guides the field. Made up of three layers, this "return yoke" reaches out 14 meters in diameter and also

acts as a filter, allowing through only muons and weakly interacting particles such as neutrinos. The enormous magnet also provides most of the experiment's structural support, and must be very strong itself to withstand the forces of its own magnetic field.

3.2.3 Tracker and Pixel Detector

The main purpose of the tracker sub-detector system is to reconstruct the trajectory of charged particles. Charged particles move in a helicoidal way which can be parameterized as a function of p_T, η, ϕ, z_0, and d_0. Here, z_0 and d_0 are the maximum longitudinal and transverse impact parameters relative to the centre of the beam spot, respectively. The beam spot is the luminous region produced by the collisions of proton beams.

As particles traverse the detector, they leave a ionization trail or *hits*. Without further analysis, it is not known which particle triggered which hits. Particle tracking consists in reconstructing the trajectories of the charged particles from the tracker measurements.

Momentum analysis in CMS makes use of the magnetic field provided by its super-conducting solenoid. The tracker sub-detector is not only able to measure the momentum of charged particles but also determines their direction at their production vertex.

A description of the hardware used for tracking is described in the following pharagraphs and the algorithms used for reconstructing and analyzing particle tracks can be found in Section 4.2.

The full sillicon inner tracking system[63] is a cylinder-shaped detector with an outer radius of 1.2 m and a length of 5.6 m. The barrel(each of the two endcaps) includes three(two) layers of pixel detectors, surrounded by ten(twelve) layers of micro-strip detectors. The 16,588 silicon sensor modules are finely segmented into 66 million 150×100 μm pixels and 9.6 million 80-to180 μm-wide strips. Figure 3.4 shows the layout of the tracker and its subsystems.

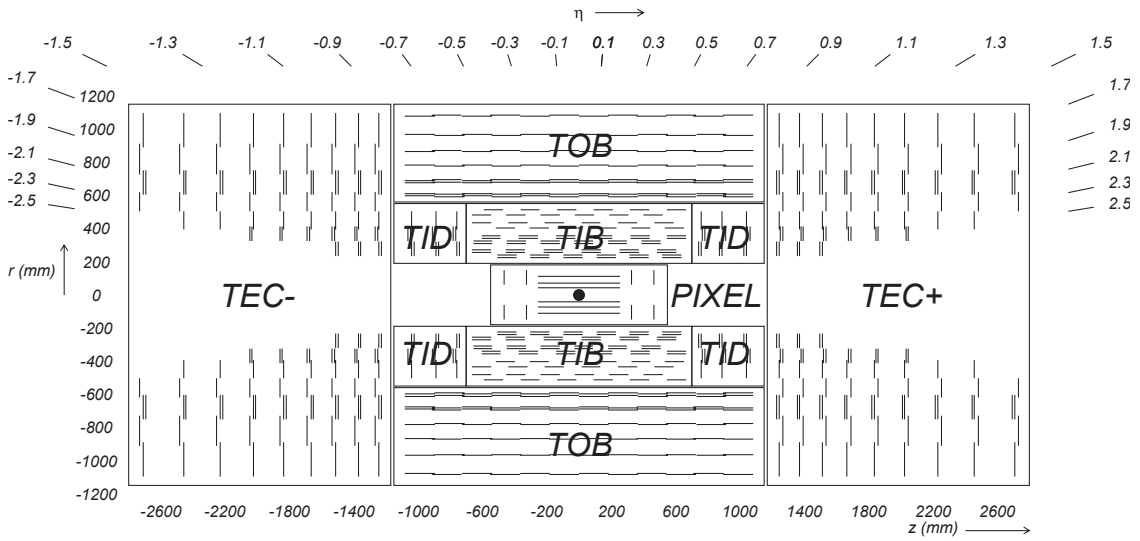

Figure 3.4: Layout of the CMS detector tracker with subsystems labeled. Reprinted from [8].

The pixel detector is made up of three barrel layers, called BPIX, and two endcap layers called the FPIX. The BPIX contains 48 million pixels and the FPIX contains another 18 million pixels. In total it consists of 1440 hybrid silicon detector modules, each with a dimension of $100 \times 150 \mu m^2$. The small pixel size enables track resolutions of $10 \mu m$ in the transverse plane and $20 \mu m$ in the z-direction. The pixel detector is what gives CMS its excellent secondary vertex tagging ability in addition to producing seed tracks for the strip tracker and the high level trigger (HLT).

Likewise, the silicon strip detector is made up of four subsystems. The Tracker Inner Barrel (TIB) has four layers of 320 μm strips. At each end of the TIB is a three-layer Tracker Inner Disks (TID), which contains strips of the same thickness. The Tracker Outer Barrel (TOB) is the six layer system which surrounds the TIB/TID. The first four layers of the TOB use $500 \mu m$ thick strips, and the last two layers use 122 μm thick strips. The Tracker EndCaps (TEC) are on either side of the previous setup and contain nine disks with up to seven layers of strips. These strips are 320 μm thick in the inner four rings and 500 μm in the outer three rings. In total, the strip detector contains 9.3 million silicon strips.

The tracker measures the p_T of charged hadrons at normal incidence with a resolution of 1% for

$p_T < 20$ GeV[64]. The relative resolution then degrades with increasing p_T to reach the calorimeter energy resolution for track momenta of several hundred GeV.

3.2.4 Calorimeters

Calorimeters are an important class of detector used for measuring the energy and direction of a particle (or collection of particles) by its total absorption. They differ from most other detectors in that the nature of the particle is changed by the detector. Moreover, calorimeters can detecto both neutral and charged particles.

During the interaction with this type of detector, the particle will generate a cascade(s) of secondary particles by a process we will refer to as *showering*.

The absorption of a particle in a calorimeter is a statistical process governed by the Poisson distribution and therefore, the relative precision of energy measurements $\Delta E/E$ varies like $E^{-1/2}$ for large E.

Since the characteristics of electromagnetic and hadronic showers are somewhat different it is convenient to describe each subsystem separately.

3.2.4.1 Electromagnetic Calorimeter

Electromagnetic calorimeters are designed to measure the energy of electrons and photons. The CMS ECAL[65] is a homogeneous calorimeter made out of lead tungstate ($PbWO_4$) crystals totaling 75,848 units. The detector is divided up into two sections which provide a coverage of $|\eta|< 1.479$ in the barrel region (EB) and $1.479 < |\theta|< 3.0$ in two endcap regions (EE). There are also preshower detectors (PS) in each of the endcaps, in fornt of the EE, which cover a pseudorapidity range of $1.653 < |\eta|< 2.6$. Figure 3.5 shows the structure of the ECAL.

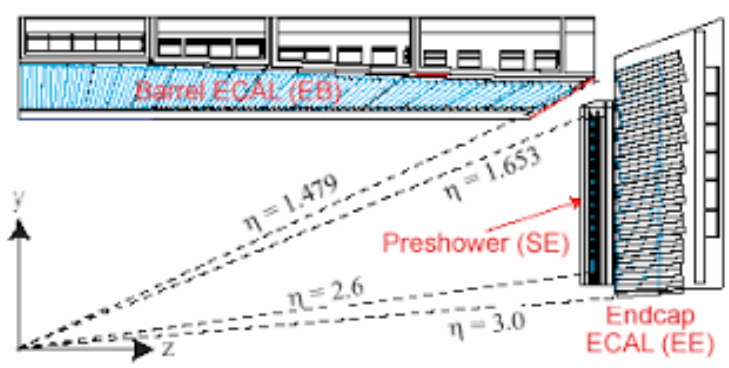

Figure 3.5: A schematic of the CMS ECAL detector with its subsystems labeled. Reprinted from [9].

Each calorimeter cristal has a depth of 230 mm, which corresponds to 25.8 radiation lengths (X_0) for $PbWO_4$, sufficient to contain more than 98% of the energy of electrons and photont up to 1 TeV. The scintillation light produced in the crystals is read out by avalanche photodiodes (APDs), which produce approximately 4.5 photoelectrons per MeV at room temperature.

The crystal transverse size matches the small Moliére radius of $PbWO_4$, 2.2 cm. This fine transverse granularity makes it possible to fully resolve hadron and photon energy deposits as close as 5 cm from one another.

The intrinsic energy resolution (σ) of the ECAL barrel was measured with an ECAL supermodule directly exposed to an electron beam[66]. The relative energy resolution is typically parameterized as a function of the electron energy as

$$\frac{\sigma_E}{E} = \frac{2.8\%}{E[GeV]} \oplus \frac{12\%}{\sqrt{E[GeV]}} \oplus 0.3\% \tag{3.1}$$

3.2.4.2 Hadron Calorimeter

Hadronic calorimeters measure the energy of hadrons, as their name suggests. The CMS HCAL[67] is a sampling calorimeter, meaning it finds a particle's position, energy and arrival time using alternating layers of "absorber" and fluorescent scintillator or "active" materials that

produce a rapid light pulse when the particle passes through.The produced light is then collected by optic fibers that feed it into readout boxes where photodetectors amplify the signal. The amount of light in a given region is summed up over many layers of tiles in depth, called a "tower".

The HCAL is organized into barrel (HB and HO), endcap (HE), and forward (HF). There are 36 barrel "wedges", each weighting 26 tonnes. These form the last layer of detector inside the magnetic coil. A few additional layers, the outer barrel (HO), sit outside the coil, ensuring no energy leaks out the back of the HB undetected. Similarly, 36 endcap wedges measure particle energies as they emerge through the ends of the solenoid magnet. In the barrel, the HCAL absorber thickness amounts to almost six interaction lengths at normal incidence, and increases to over ten interaction lengths at larger pseudorapidities. The HO material corresponds to 1.4 interaction lengths at normal incidence.

Lastly, the two hadronic forward calorimeters (HF) are positioned at either end of CMS, to detect particles coming out of the collision region at shallow angles relative to the beam line. These receive the bulk of the particle energy contained in the collision so must be very radiation resistant. Figure 3.6 shows the structure and position of the HCAL subsystems.

Combined, the ECAL and HCAL can measure the energy deposited by a charged pion with a resolution of $\sigma/E \approx 100\%/\sqrt{E[GeV]} \oplus 5\%$ [68], assuming an average jet particle composition.

3.2.5 Muon System

Since muons can penetrate several meters of iron without interacting, gas-ionization detector chambers were placed at the very edge of the experiment embedded in the steel flux-return yoke in order to detect them. This allows for a pseudorapidity coverage of $|\eta| < 2.4$.

The muon system consists of 1400 muon chambers which can be classified into three categories, according to the technology used: 250 drift tubes (DTs), 540 cathode strip chambers(CSCs), and 610 resistive plate chambers(RPCs).

The barrel region of the detector contains DTs and RPCs, while the endcap region contains CSCs and RPCs. The layout of the muon system can be seen in Figure 3.7.

The four muon DT stations sitting outside the magnet coil are interleaved with the iron "return

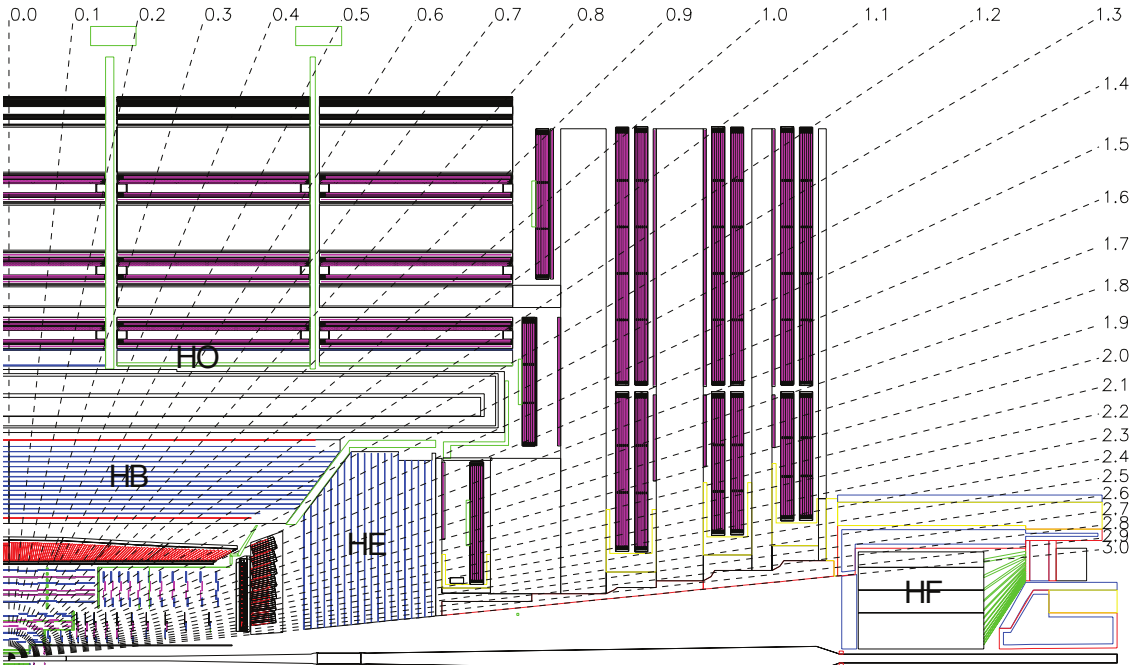

Figure 3.6: Structure and position of the CMS HCAL sub-detector systems. Reprinted from [10].

yoke" (shown in red in Figure 3.8, for the barrel region), which not only returns the flux from the solenoid, but also shields the muon chambers from hadrons.

The CSCs track the particle's position and allow for triggering, while the RPCs form a redundant trigger system, which quickly decides to keep the acquired muon data or not. Because of the many layers of detector and different specialties of each type, the system is naturally robust and able to filter out background noise.

The muon system on its own has a resolution of 15-40$\%$ depending on η. Matching muons to tracks measured in the silicon tracker results in a relative transverse momentum resolution for muons with 20$<p_T<$ 100 GeV of 1.3-2.0$\%$ in the barrel and better than 6$\%$ in the endcaps. The p_T resolution in the barrel is better than 10$\%$ for muons up to 1 TeV [69].

3.2.6 Luminosity Measurement

Two important features of a particle accelerator are its center of mass energy and its instantaneous luminosity ($\mathcal{L}$). The largest the center of mass energy the more massive particles can be created and found, therefore, the larger the luminosity, the higher the chances of creating the

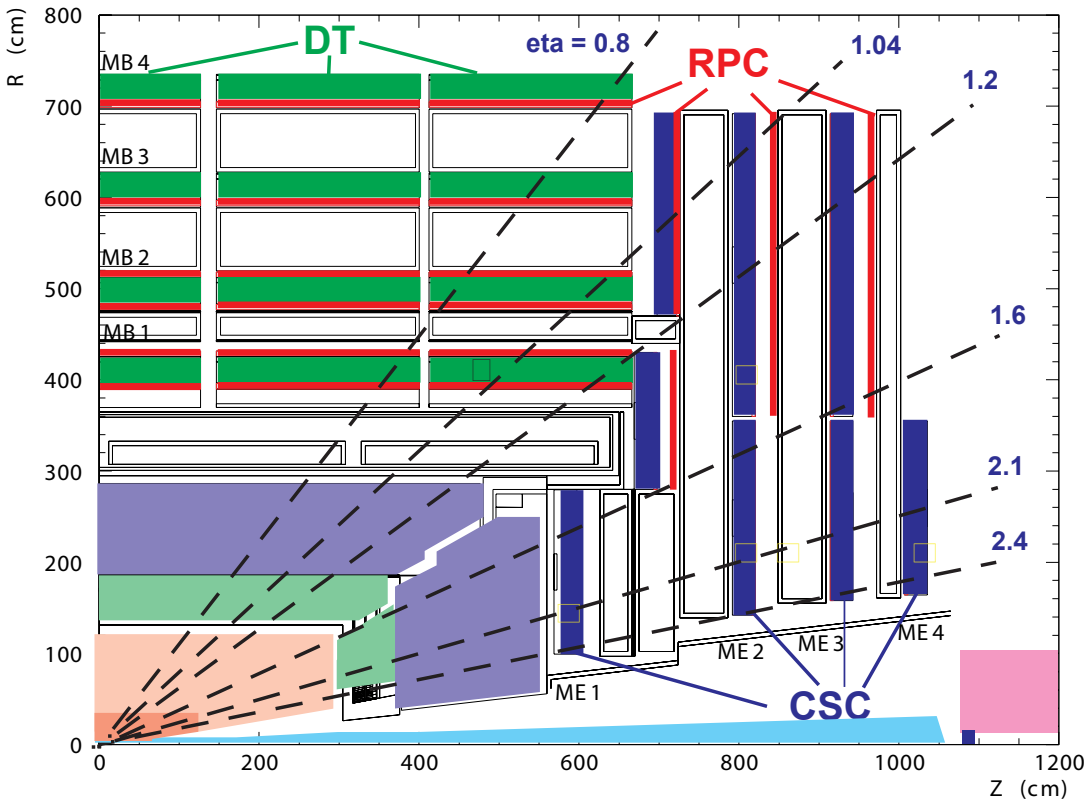

Figure 3.7: Layout of the CMS muon system. Reprinted from [11].

particle.

Besides measuring the kinematics of each of the particles traversing the detector, CMS must also measure the instantaneous luminosity delivered by the LHC. Both the pixel detector, and the HF are able to measure the luminosity to varying degrees of accuracy.

For a given process, the number of interactions (N) is the product of $\mathcal{L}$ integrated over the data taking time period and the cross section for the process in question (σ_{ref}):

$$N = \sigma_{ref} \int \mathcal{L}(t)dt \tag{3.2}$$

The Van der Meer (VdM) scan method measures the size and shape of the interaction region of the colliding beams. This is achieved by displacing the beams in the x and y- (transverse) planes and measuring the relative interaction rates as a function of the transverse beam separation.

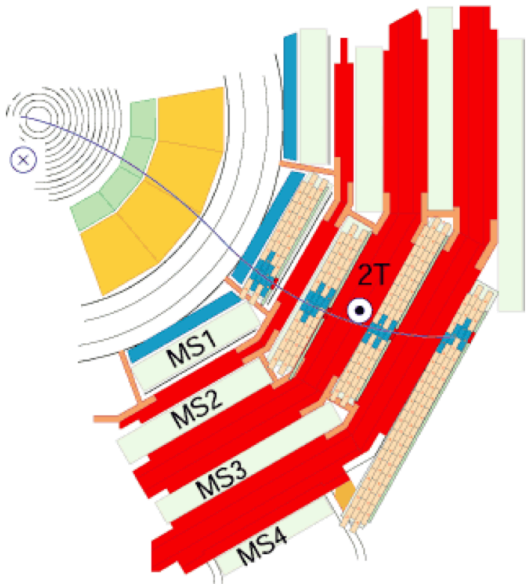

Figure 3.8: A muon, in the plane perpendicular to the LHC beams, leaves a curved trajectory in four layers of muon detectors or stations. Reprinted from [11].

For Gaussian beams, the luminosity as a function of the transverse displacement (δu) can then be expressed as:

$$\mathcal{L}(\delta u) = \mathcal{L}_0 exp[-\frac{\delta u^2}{2\sigma_u^2}] \tag{3.3}$$

where

$$\mathcal{L}_0 = \frac{N_1 N_2 f N_b}{2\pi\sqrt{(\sigma_{1x}^2 + \sigma_{2x}^2)(\sigma_{1y}^2 + \sigma_{2y}^2)}} \tag{3.4}$$

and $\sigma_u = \sqrt{\sigma_{1u}^2 + \sigma_{2x}^2}$ with $u = x, y$ for each separation plane, N_b the number of colliding bunches and f the revolution frequency.

A fit of the measured interaction rates as a function of the reparation will allow to determine the effective beam size as well as the maximum achievable collision rate ($\dot{N}$)

$$\dot{N} = \mathcal{L}\sigma \tag{3.5}$$

In practice, the scans are performed by moving the beams step-wise across each other in the two transverse planes.

4. EVENT RECONSTRUCTION

As explained in the previous section, the LHC experiment consists of two beams of hadrons circulating in opposite directions. These beams cross at the interaction points where collisions take place. This small, longitudinal region is usually referred to as the luminous or interaction region. Then, the origin of one or more new particles is called a vertex. Particularly, the set of detector measurements associated to a single beam-beam crossing will be referred to as an event. This section explains how a reconstruction software is used to process the raw information and identify physics objects for a given event.

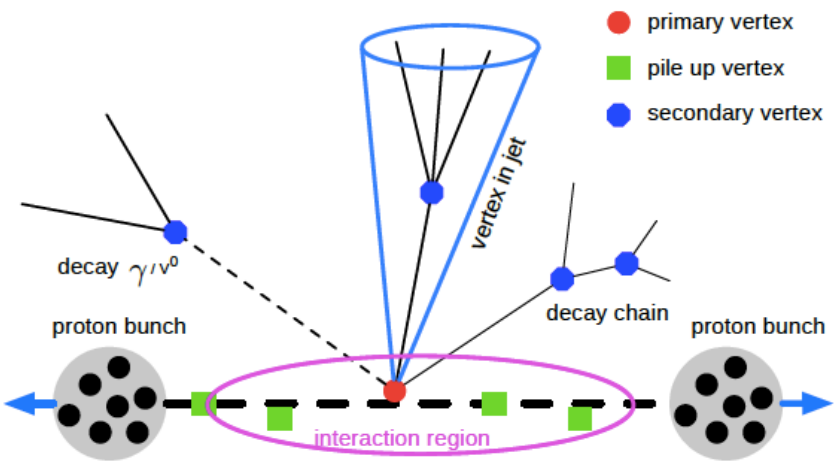

Figure 4.1: Schematic diagram for a reconstructed event at the LHC. Reprinted from [12].

4.1 Data Acquisition

A beam in the LHC is not a continuous string of particles, but it is divided into proton bunches a few centimeters long squeezed down to the size of a human hair at the collision point. Elsewhere in the ring, the beam size is normally less than a millimeter. The LHC is designed to run with 2808

bunches per beam, separated by a gap of 25 ns, which corresponds to a collision rate of 40 MHz. At this rate, it would be impossible to keep or even transmit all the collected data.

The CMS *Data Acquisition (DAQ) and trigger system* was specifically designed to cope with the large amount of data acquired by the detector. The so-called trigger systems perform an online event filtering in an effort to reduce the amount of data before writing it to disk.

The interaction point of a particular proton-proton collision is referred to as the *primary vertex* (PV), all other proton-proton interactions that took place during the same bunch-crossing are called *pile-up* (PU). Secondary vertices refer to those production points (other than the PV) where particles are created either from the decay or hard-scattering of the particles associated with a particular PV. This is shown in Figure 4.1. The higher the collision energy, the more interesting the collision, since higher energy collisions are more likely to produce particles or interactions that have not yet been observed. For this reason, triggers are configured to prefer higher momentum objects.

4.1.1 L1 Trigger and HLT

Whenever the LHC is performing at its peak, about one billion proton-proton interactions take place every second inside the CMS detector. To select events of potential physics interest, the CMS trigger[70] utilizes a two-level system including an L1 hardware trigger and an High Level Trigger (HLT) array of commercially available computers running high-level physics algorithms.

The L1 of the CMS trigger is an extremely fast process that selects events containing candidate objects, e.g. ionization deposits consistent with a muon, or energy clusters consistent with a physics object such as an electron or photon. During this process, only coarsely segmented data from calorimeter and muon detectors is used, while all the high-resolution data is held in pipeline memories in the front-end electronics. At the end of this process, no more than 100 kHz of the stored events are forwarded to the next trigger system, the HLT.

The data processing of the HLT is structured around the concept of an *HLT path*. This path is a set of algorithmic processing steps run in a predefined order that rudimentarily reconstruct physics objects based on specific HLT objects. See Figure 4.2 for reference. The reconstruction modules and selection filters of the HLT use the software framework that is also used for offline

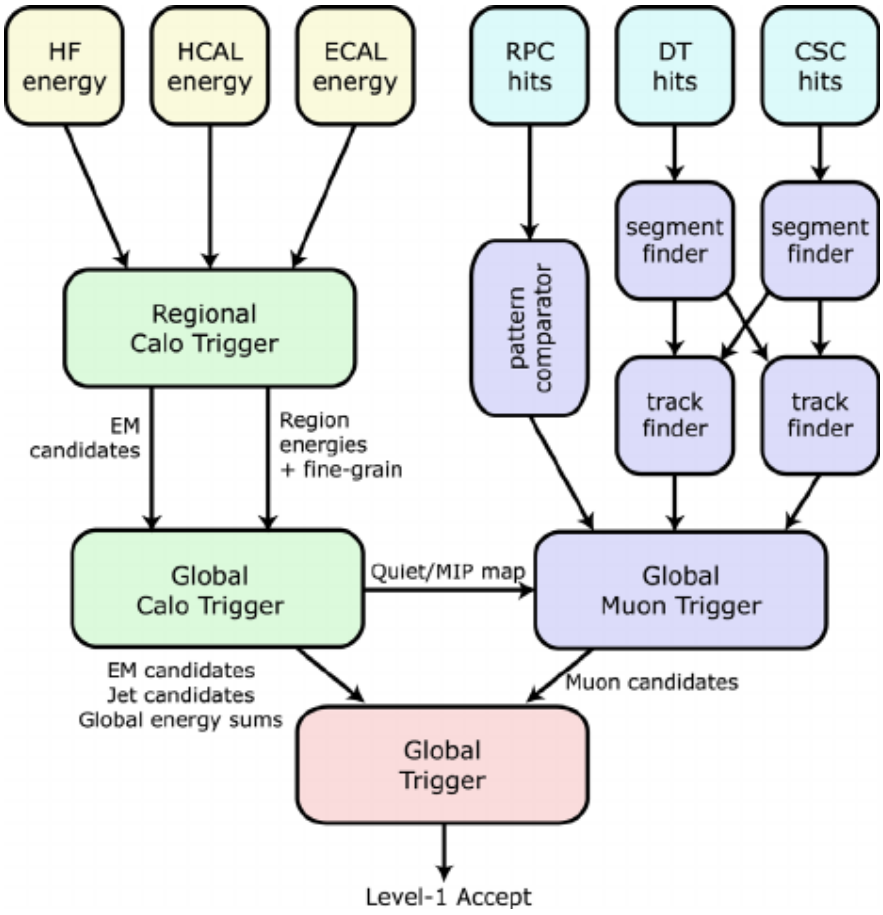

Figure 4.2: The CMS Level-1 Trigger. Reprinted from [13]

reconstruction and analyses.

As a result of this procedure, the average rate of accepted events is reduced to about 400 Hz. Event data is stored locally on disk and eventually transferred to the CMS Tier-0 computing center for offline processing and permanent storage.

4.1.2 T1 sites and data storage

The CMS computing system operates on a tiered structure. A Tier-0 computing center is located at CERN where the data is transferred to the HLT and a first set of reconstruction occurs. From there, it is transferred to one of seven Tier-1 computing centers located around the world. At the Tier-1 centers, a full reconstruction of the data is performed. Furthermore, there are 55 Tier-2 centers which can be accessed by the collaboration members for data processing and storage. See

Figure 4.3 for reference.

The analysis presented here was performed at one of the Tier-3 centers, the Texas A&M University Brazos HPC cluster[71].

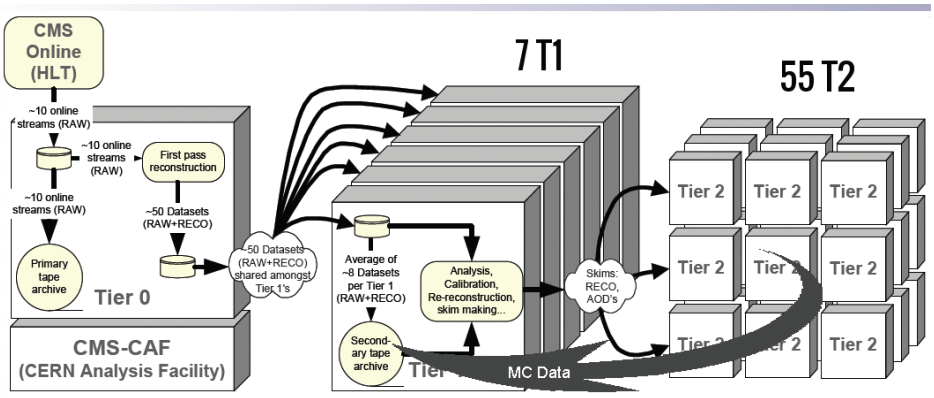

Figure 4.3: Flow of CMS detector data through the tiers. Reprinted from [14]

The data itself is also processed in three data tiers. The first layer of this system is the RAW data format, which is created by unpacking detector streams passed on from the HLT trigger. It includes measurements from the different sub-detectors, as well as some information provided by the HLT and L1 triggers. RAW data is then reconstructed into PF objects, as explained in Section 4.2. This step is called RECO, which is short for reconstruction and contains both the detector and physics object information.

After the RECO step, *analysis object data* (AOD) is generated from a subset of the RECO information. AOD objects are typically comprised of only high-level physics objects, and therefore, of reduced size.

4.2 Particle Flow Event Reconstruction

In the previous section we described how data was managed and stored during the acquisition process. This section will focus on how raw detector information is interpreted.

Raw detector data is measured in the form of hits in the tracker or the muon system, as well as energy depositions in the calorimeters. Then, the trajectories of charged particles, or tracks, are reconstructed from the position hits in the tracker. From the collection of tracks in an event, interaction vertices are reconstructed.

An optimal event description can be achieved by correlating the basic elements from all sub-detectors (tracks and clusters) to identify each final-state particle, and by combining the corresponding measurements to reconstruct the particle properties on the basis of this identification. At CMS, this approached is called *particle-flow (PF) reconstruction*.

The reconstructed and identified individual particle list includes muons, electrons, photons, as well as charged and neutral hadrons. This list of individual particles is then used to build jets (from which the quark and gluon four-momentum in inferred), to determine the E_T^{miss} (defined as the modulus of the vector sum of the p_T of all reconstructed particles and therefore provides an estimate of the direction and energy of the neutrinos and other invisible particles), to reconstruct and identify taus from their decay products, etc.

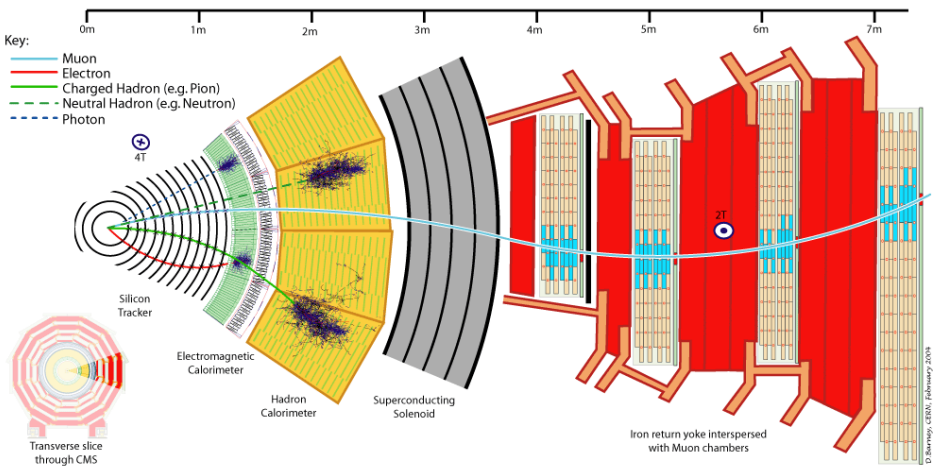

Figure 4.4: Cross-sectional view of the CMS detector with all of the sub-detectors labeled. The colored lines correspond to different particle types. Each particle interacts with different pieces of the detector and may or may not be bent by the magnetic field. Reprinted from [15]

During PF reconstruction, photons and neutral hadrons are identified by ECAL and HCAL clusters with no associated tracks, respectively. Electrons can be identified by associating a track to an ECAL cluster and from possible Bremsstrahlung photons radiated by the electron in the tracker on its way to the ECAL; with a momentum-to-energy ratio compatible with unity, and not connected to an HCAL cluster. Finally, muons and neutrinos would traverse the calorimeters with little or no interactions. While neutrinos would escape undetected, muons would be identified by a track in the inner tracker connected to a track in the muon detectors. Muons are reconstructed in isolation as well as in jets. Finally, the presence of neutrinos can be detected by the E_T^{miss} in the event. See Figure 4.4.

The PF concept was developed and used for the first time by the ALEPH experiment at LEP[72]. In particular, CMS is very well suited for PF reconstruction due to its highly-segmented tracker, a fine-grained ECAL, and an hermetic HCAL. Also, the CMS magnet is large enough to accommodate the tracker and both the ECAL and HCAL, thereby minimizing the amount of material in front of the calorimeters.

In the following sections, the general PF reconstruction work-flow will be described, starting by the reconstruction of its fundamental elements, the charged particle tracks and the calorimeter clusters. These elements are then grouped and interpreted in terms of particles.

4.2.1 Iterative Tracking

The first step of the PF reconstruction process consists of the reconstruction of hits in the pixel and strip tracker[73] and is referred to as local reconstruction.

The next step is track reconstruction, which refers to the process of using the reconstructed hits to obtain estimates for the momentum and position parameters of the charged particles responsible for the tracker hits. The momentum of charged hadrons is measured in the tracker with a resolution vastly superior to that in the calorimeter. Furthermore, the tracker provides a precise measurement of the charged particle direction at the production vertex.

The tracking software at CMS[73] is commonly referred to as the *Combinatorial Track Finder* (CTF), which is an adaptation of the Combinatorial Kalman Filter [74, 75, 76], which in turn is an

extension of the Kalman filter[77] to allow pattern recognition and track fitting to occur in the same framework. The collection of reconstructed tracks is produced by multiple passes or iterations of the same CTF track reconstruction sequence, in a process called iterative tracking.

The basic idea of iterative tracking is that tracks of relatively large p_T and those produced near the interaction region are searched for during the initial iteration. During successive iterations, hits unambiguously assigned to candidates found in previous iterations tracks are removed. By doing so, the combinatorial complexity is reduced, and subsequent iterations searching for more difficult types of tracks (e.g., low p_T, or greatly displaced tracks) is simplified.

Each iteration proceeds in four steps:

- Seed generation, which provides track candidates consisting of a few (2 or 3) hits. Seeds are generated in the innermost layers of the tracker and are commonly referred to as *proto-tracks*.
- Track finding, which is based on a Kalman filter. It extrapolates the seed trajectories along the expected flight path of a charged particle, searching for additional hits that can be assigned to the track candidate.
- Track fitting. A module that is used to provide the best possible estimate of the parameters of each trajectory by means of a Kalman filter.
- Track selection. This step sets the quality flags and discards tracks that fail certain specified criteria.

A total of six iterations are used, each with different seed generation, p_T, and impact parameter requirements. Once the hits that are associated with so-called *fake tracks* are removed, the seeding criteria is loosened, and therefore, tracking efficiency is increased. From iteration 4 and on, the constraints on the tracks closer to the interaction point are slowly relaxed. This allows for reconstruction of secondary charged particles created from photon conversions and nuclear interactions in the tracker volume.

4.2.2 Calorimeter Clustering

Clustering in the calorimeters is the process of grouping detector cells that register hits together with the purpose of (i) detecting and measuring the energy and direction of stable neutral particles, (ii) being able to separate these neutral particles from energy deposits associated with charged hadrons, (iii) reconstructing and identifying electrons and all possible Bremsstrahlung photons, and (iv) helping the energy measurement of charged hadrons for which the track parameters were not determined accurately, which is sometimes the case for high-p_T tracks.

The clustering algorithm is performed separately in each of the following sub-detectors: ECAL barrel and endcap, HCAL barrel and endcap, as well as in the pre-shower. It proceeds via three steps[78]:

1. Identify 'cluster seeds'. These are defined as the cell in a calorimeter with a local maximum of energy (above some set threshold).

2. Expand from the seed to grow 'topological clusters'. This is done by aggregating calorimeter cells that have at least one side in common with the seed cell, and also have an energy over a particular threshold.

3. Repeat the process of cluster growing, now using new cells that are part of the cluster.

In this sense, a "seed" gives rise to a "particle-flow cluster". If a cell is identified by two clusters, the energy is shared between the clusters according to the distance from the cell to the center of each cluster. The cluster energies and positions are iteratively determined as new cells are added to the cluster.

4.2.3 Linking Tracks and Clusters

Once the basic PF elements are available, the next step in reconstructing a particle is the so-called *link algorithm*. This algorithm can test any pair of elements in the event. In order to prevent the computing time of the link algorithm from growing quadratically with the number of particles,

the pairs of elements considered by the link procedure are restricted to the nearest neighbors in the (η, ϕ) plane, as obtained with a k-dimensional tree[79].

If two elements are found to be linked, the algorithm defines a metric between these two elements, aimed at quantifying the quality of the link. The link algorithm then produces *PF blocks* of elements associated either by a direct link or by an indirect link through common elements.

The link between tracks and calorimeter clusters proceeds by extrapolating the last measured hit in the tracker to one of the three detectors[78]:

- The two layers of the pre-shower detector,
- the ECAL, at a depth corresponding to the expected maximum of the electron shower profile,
- the HCAL, to a depth corresponding to one interaction length.

The track is then linked to a cluster in these detectors if the extrapolated position is within the cluster boundaries. Additionally, to link Bremsstrahlung photons to their associated electron, tangents to the track are extrapolated to the ECAL and any cluster found within those boundaries is also linked.

Similarly, links between the calorimeters are formed when a cluster from the more granular calorimeter (pre-shower or ECAL) is within the cluster envelope of the less granular calorimeter (ECAL or HCAL). Finally, muon tracks are linked to charged particle tracks by a global fit between the two sets of tracks.

4.3 Physics Object Reconstruction

With the tracks identified, calorimeter clusters formed, and the linking of clusters to tracks; particles can then be reconstructed. The PF process begins by reconstructing muons, then electrons and photons, and finally charged hadrons. As each particle is reconstructed, the tracks and clusters associated with it are removed from the collection of blocks used to form candidate particles, which ensures that energy deposits attributed to one particle are not used twice. The hadrons are then clustered together to form *jets*, and these jets can additionally be identified as coming from

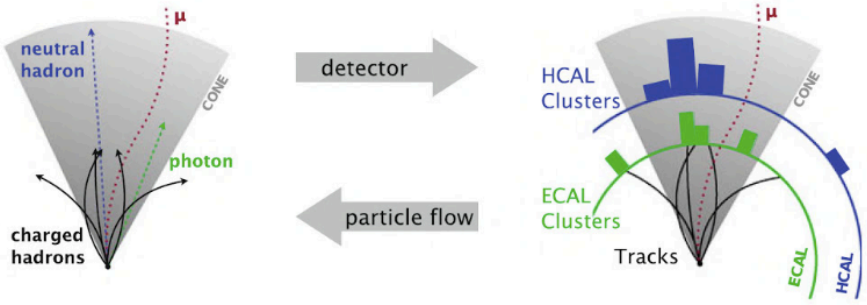

Figure 4.5: CMS Particle Flow algorithm. The diagram shows how collisions lead to particle decays and final state particles. On the right side of the diagram the tracks and deposits in the CMS detector are shown. The left side shows that PF candidates are derived from detector information and then become input for the PF algorithm that uses them to construct high-level physics objects like electrons, which are then used by analysts to reconstruct the collision event. Reprinted from [16]

tau leptons or b quarks[64, 80] (Figure 4.5). There is also Pileup Jet Identification (PU Jet ID)[81], a c-quark jet identification probability, as well as composite jet identification, e.g. a top-jet or a W-jet or Z-jet but these higher level objects are not relevant for this analysis.

4.4 Jets

During proton-proton collisions, the confined state of quarks and gluons is broken. Shortly after the collision, partons hadronize and a bunch of particles is generated by this process. These particles are usually collimated in a given direction due to the boosted nature of the parton, and thereby produce a jet or spray of particles around it. See Figure 4.6 for reference.

In practice, jets are the result of clustering groups of charged hadrons, photons, and neutral hadrons with the occasional muon or electron. The energy fraction in jets is divided amongst them with a breakdown of roughly 65%, 25%, and 10% respectively. This is illustrated in Figure 4.7. For this study, jets were reconstructed from PF candidates clusters using the anti-k_T algorithm[18] as defined in the FASTJET package[82].

Jet clustering algorithms work by defining a distance parameter d_{ij} between PF candidates i and j and the distance between such cluster and the beam d_{iB}. These are defined as

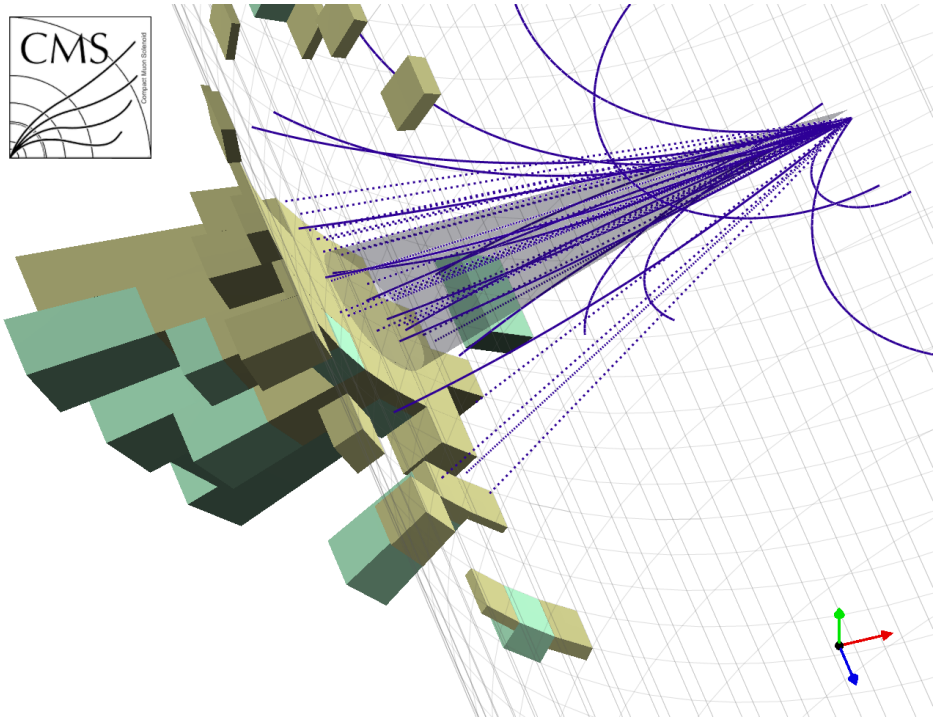

Figure 4.6: Schematic view of a jet with tracks and calorimeter deposits at CMS. Reprinted from [17]

$$d_{ij} = min(k_{ti}^{2p}, k_{tj}^{2p})\frac{\Delta_{ij}^2}{R^2} \tag{4.1}$$

$$d_{iB} = k_{ti}^{2p} \tag{4.2}$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, and k_{ti}, y_i, and ϕ_i are the transverse momentum, rapidity, and azimuth of particle i, respectively. R is a user-defined radius parameter, and p is a measure of the relative power of energy vs geometric scales. Particularly, for the anti-k_T algorithm, $p = -1$, and Equation 4.1 reduces to

$$d_{ij} = min(\frac{1}{p_{ti}^2}, \frac{1}{p_{tj}^2})\frac{\Delta_{ij}^2}{R^2} \tag{4.3}$$

The algorithm[82] loops over all PF candidate objects, calculating d_{ij} for each pair of objects.

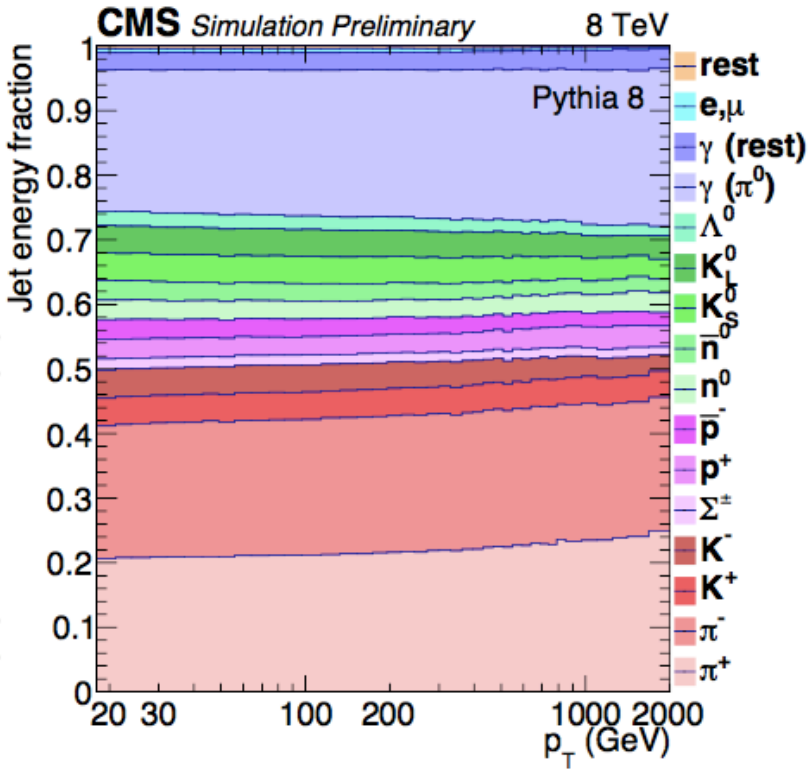

Figure 4.7: Particle composition for a jet. The energy fraction is relatively constant as a function of p_T^{jet} and corresponds to roughly 65%, 25%, and 10% charged hadrons, photons, and neutral hadrons, respectively. Reprinted from [18]

Once it does this, it selects the two objects with the lowest value of d_{ij} and combines them. Subsequently, it calculates the distance of any merged clusters to their nearest neighbors. This process is repeated until the smallest value of d_{ij} satisfies the condition $d_{ij} > d_{iB}$.

As a result, the cutoff limit of $1/p_T^2$ defines a maximum size that the algorithm will look to cluster particles inside. The construction of d_{ij} using the inverse p_T^2 has a result of producing values of d_{ij} that are smaller for objects with a higher p_T, given equal separation. As a result, softer particles will tend to cluster to higher p_T particles long before they would cluster amongst themselves. If no hard particles are present, the jet object will simply cluster soft p_T particles in a circle in an $\eta - \phi$ space of radius R.

The clustering of the anti-k_T algorithm leads to jets with a large p_T being reconstructed as perfect circles. Figure 4.8 shows a display of the anti-k_T algorithm for a distance parameter R=1.

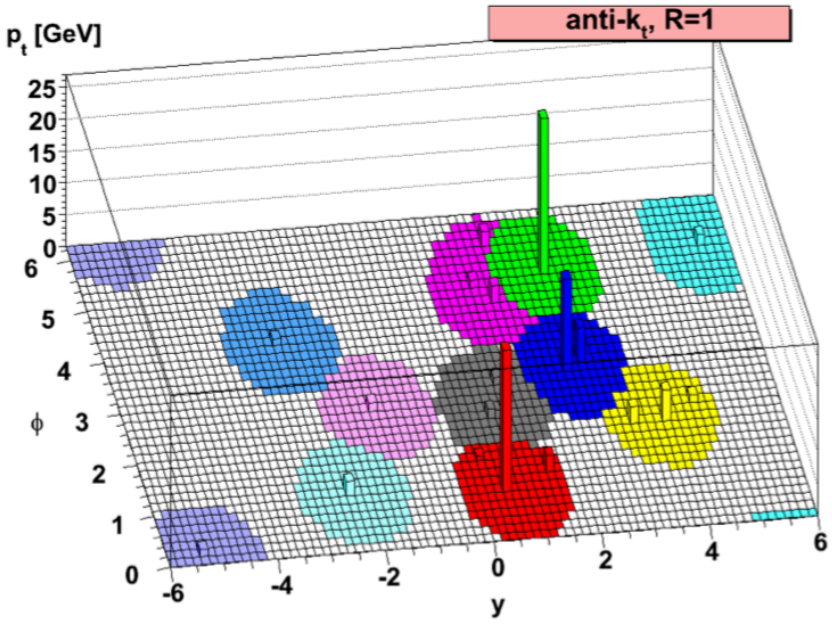

Figure 4.8: A sample parton-level event clustered with the anti-k_T algorithm. Reprinted from [18]
.

Notice that the green jet around $y = 2$ and $\phi = 5$ has a circular shape, while the smaller jets loose some of their clusters to their higher momentum neighbors.

Finally, the anti-k_T algorithm is both infrared and collinear safe. Infrared safety implies that the jet clustering algorithm is insensitive to the emission of soft, wide angle particles. Under this condition, two jets would not be merged due to one of them producing a soft-momentum particle between them. Collinear safety means that if there is a splitting which results in two parallel high-p_T particles, a single jet is produced and the jet properties will not be different from a jet where this splitting did not occur. If the algorithm follows these two properties, it is referred to as being IRC safe.

After the clustering procedure, the momentum and energy of the reconstructed jets still might not be the same as those from the initial parton. This could be due to out of cone showering, the presence of additional pileup energy produced during the same bunch crossing as the primary vertex, or detector effects. To correct for this, CMS adopted a factorized approach[83] with three

levels of correction, each one targeting a specific effect. Correction factors obtained at each level are applied sequentially in the order in which they were obtained. The goal is to make sure each jet has a relative response

$$\mathcal{R}_{rel} = \frac{p_T^{reco}}{p_T^{ref}} \tag{4.4}$$

as close as possible to unity. Here p_T^{reco} is the reconstructed jet p_T and p_T^{ref} is the true reference p_T of the jet at generator level.

The process of correcting the jet 4-momentum by means of a scale or weight obtained from matching the reconstructed jet information to that of the reference jet in Monte Carlo is referred to as jet energy correction (JEC).

The first level of correction, commonly referred to as the L1FastJet[84] corrections, starts by removing pileup or electronic noise energy that may have made it into the jet reconstruction. This multiplicative correction will only remove energy from within the jet and will take the form in Equation 4.5, where ρ is the median energy density of the event and A is the jet area. f is an estimate of the average amount of energy added to an event due to pileup (offset) inside the jet per unit of jet area [83, 84].

$$p_T^{L1Corrected} = p_T^{uncorrected}(1 - A\frac{f(\eta, \rho, A)}{p_T^{uncorrected}}) \tag{4.5}$$

While the L1 corrections attempt to remove pileup and electronic noise from jet energy measurements, the L2Relative and L3Absolute MC corrections aim to correct the jet energy response so that it matches that of the particle level jet. The L2Relative correction compensates for the nonlinearity in the jet response as a function of η while the L3Absolute correction does so as a function of p_T. All three corrections are applied to both data and simulation. An additional level of correction, called L2L3Residual, is applied to data only, as a function of η, in order to correct for the difference in scale between the data and simulation.

A final level of modification to the reconstructed objects is an η dependent smearing factor

applied to the jet 4-momenta coming from the MC samples. The distribution of jet energies within the MC simulation tends to be more sharply peaked and less broad than the same distribution in data, resulting in a smaller jet energy resolution (JER) than we can realistically measure using the CMS detector. The deterministic "smearing" method recommended by CMS[85] matches the MC jet energy resolution to the one measured in data.

The reconstructed jet p_T is scaled by a correction factor C_{JER} as determined in Equation 4.6, where C_η is a correction factor derived as a function of η. The multiplicative JER correction factor is then used to modify the jet 4-momentum as in Equation 4.7.

$$C_{JER} = max\left(0.0, \frac{p_T^{GEN}}{p_T^{RECO}} + C_\eta \cdot \left(1 - \frac{p_T^{GEN}}{p_T^{RECO}}\right)\right) \tag{4.6}$$

$$\mathbf{X}_{Jet}^{corrected} = C_{JER} \cdot \mathbf{X}_{Jet}^{RECO} \tag{4.7}$$

A set of quality cuts, collectively called PF jet identification, are applied to the resulting collection of jets to ensure that only real, hard scatter PF jets are used during the analysis[86]. Several working points are defined at varying levels of efficiency and purity, but this analysis makes use of the tight criteria shown in Table 4.1[81].

Table 4.1: Cut based PF jet identification requirements for the tight working point.

Cut Variable	Cut Value		
	Tight		
η	$\lvert\eta\rvert \leq 2.7$	$2.7 < \lvert\eta\rvert \leq 3.0$	$\lvert\eta\rvert > 3.0$
Neutral Hadron Fraction	<0.90	<0.98	-
Neutral EM Fraction	<0.90	> 0.01	<0.90
$n_{constituents}$	>1	-	-
Muon fraction	<0.8	-	-
Number of Neutral Particles	-	> 2	>10
and for $\lvert\eta\rvert \leq 2.4$ in addition apply			
Charged Hadron Fraction	> 0		
Charged Multiplicity	> 0		
Charged EM Fraction	< 0.90		

All cuts on the jet energy fractions are made on the raw jets, before any energy correction is applied. In addition to the PF jet quality cuts, this analysis requires that all jets be within $|\eta| < 2.6$ and to have a $p_T > 30$ GeV.

4.5 b-tagging

Some jets are produced from a b-quark that after enters a bound state with another quark becoming part of a B meson that has a long lifetime which subsequently decays after it has traveled some distance inside the inner tracker. B-tagging is the identification of jets at some confidence level as having contained a B meson. See Figure 4.9 for reference.

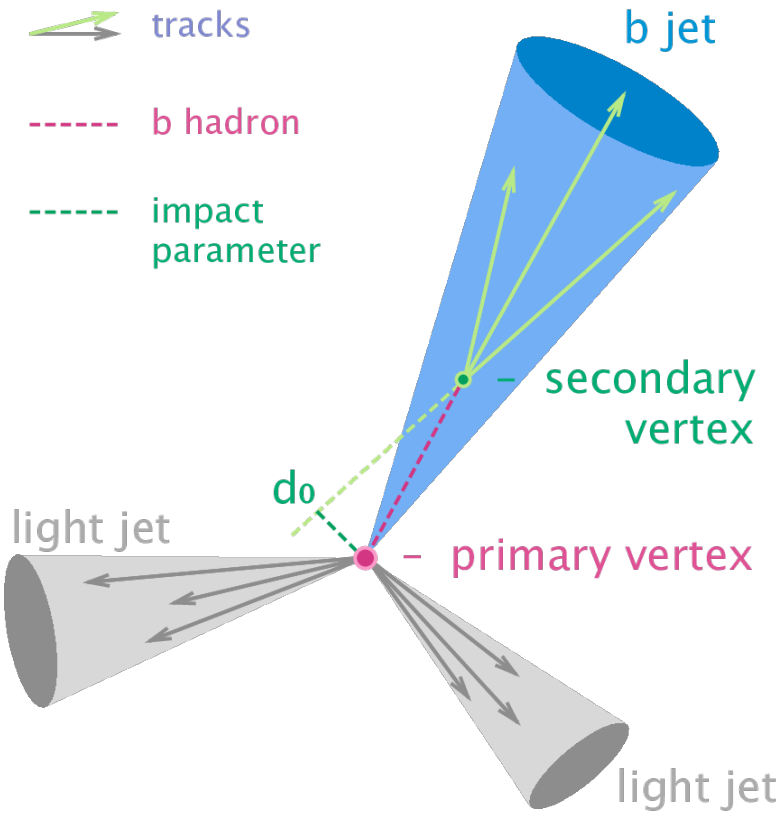

Figure 4.9: Diagram showing the common principle of identification of jets initiated by B hadron decays. Reprinted from [19]

A variety of b-tagging algorithms have been developed by CMS to select b-quark jets[87]

based on variables such as the impact parameters of the charged-particle tracks, the properties of reconstructed decay vertices, and the presence or absence of a lepton, or combinations thereof. These algorithms heavily rely on machine learning tools like deep neural networks. Currently, CMS makes use of the DeepCSV[20, 80] algorithm, which uses a deep neural network trained by using about 50 million simulated jets.

Table 4.2: Input variables used for the CSVv2 algorithm.

Input variable
Secondary vertex 2D flight distance significance
Number of secondary vertices
Track η_{rel}
Corrected secondary vertex mass
Number of track from secondary vertex
Secondary vertex energy ratio
$\Delta R(Secondaryvertex, jet)$
3D interaction point significance of the first four tracks
Track $p_{T,rel}$
$\Delta R(track, jet)$
Track $p_{T,rel}$ ratio
Track distance
Track decay length
Summed tracks E_T ratio
ΔR(summed tracks, jet)
First track 2D interaction point significance above c threshold
Number of selected tracks
Jet p_T
Jet η

The DeepCVS algorithm uses the reconstructed tracks and secondary vertices found by using the *inclusive vertex finding* (IVF) algorithm [88]. The same input variables used for the CSV(Combined Secondary Vertex)v2 tagger (Table 4.2 from [20]) are used, with the difference that the track-based variables use up to six tracks in the training of the DeepCSV. Jets are randomly selected in such a way that similar jet p_T and η distributions are obtained for all jet flavors. These distributions are also used as input variables in the training to take into account the corre-

lation between the jet kinematics and the other variables. The distribution of all input variables is preprocessed to center the mean of each distribution around zero and to obtain a root-mean-square value of unity. All of the variables are presented to the multi-variable analysis (MVA) in the same way because of the preprocessing.

The training is performed using jets with p_T between 20 GeV and 1 TeV, and within the tracker acceptance. The relative ratio of jets of each flavor is set to 2:1:4 for b:c:udsg jets. A mixture of $t\bar{t}$ and multi-jet events is used to reduce the possible dependency of the training on the heavy-flavor quark production process.

The training of the deep neural network is performed using the KERAS[89] deep learning library, interfaced with the TENSORFLOW[90] library that is used for low-level operations such as convolutions. The neural network uses four hidden layers that are fully connected, each with 100 nodes. For the nodes in the last layer, a normalized exponential function is used for the activation to be able to interpret the output value as a probability for a certain jet flavor category, $P(f)$. The output layer contains five nodes corresponding to five jet flavor categories used in the training. These categories are defined according to whether the jet contains exactly one b hadron, at least two b hadrons, exactly one c hadron and no b hadrons, at least two c hadrons and no b hadrons, or none of the aforementioned categories. Each of these categories is completely independent of the others, and the reasoning behind the chosen categorization has to do with the ability of identifying jets containing two b or c hadrons.

The tagger can categorize individual jets in so-called "Tight" (DeepCSV T), "Medium" (Deep CSV M), and "Loose" (DeepCSV L) categories or working points. These working points correspond to 0.1, 1, and 10 $\%$ misidentification rates, respectively.

Figure 4.10 shows the b-jet efficiency as a function of the jet p_T for the DeepCSV algorithm at different working points. These efficiencies are obtained on simulated $t\bar{t}$ events using jets within tracker acceptance with p_T > 20 GeV.

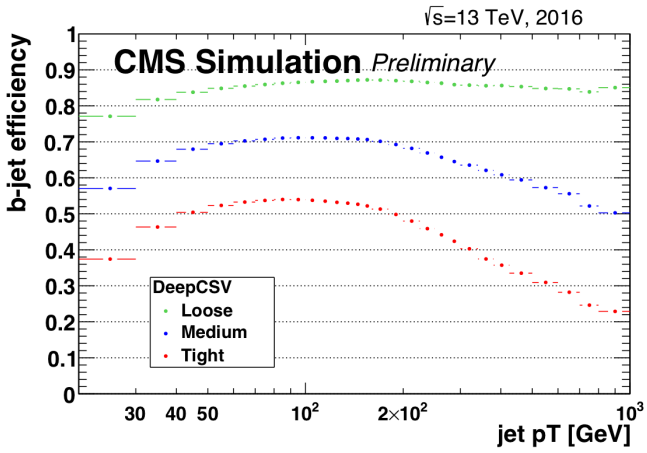

Figure 4.10: b-jet efficiency as a function of jet p_T for the DeepCSV algorithm for different working points. Reprinted from [20].

4.6 Event Generation

Searching for new physics can basically be reduced to a search for deviations from the SM. For this reason, extremely accurate signal modeling and SM background predictions are required. For a given analysis, these predictions take the form of samples of events representing various physics scenarios, as they would be seen in the CMS detector. Monte Carlo (MC) event generators are used to simulate proton-proton collisions resulting in a variety of final states. The passage of these final states through the CMS detector is then simulated, and the resulting detector level data is analyzed. The reconstruction algorithms described in the previous section are then run on the simulated data, allowing for a direct comparison with real data.

4.6.1 Event Generators

Particle physics event generators aim to give a complete description of particle collisions. This involves a combination of perturbative physics at large energy scales, and non-perturbative physics at small energy scales, which is described by phenomenological models.

In practice, event generators are software packages which take a specific initial state as input and simulate a selected subset of outcomes, from an interaction between the specified initial state particles. Bare partons produced in the hard scattering undergo gluon radiation or splitting, and

subsequently undergo hadronization to form colorless hadrons. Unstable particles produced in the hard interaction are made to decay to stable particles according to their known, or imposed, branching fractions and lifetimes.

During the event generation process, strongly-interacting particles which take part in hard scattering interactions with large momentum transfer can be considered free due to asymptotic freedom. The large energy scale involved allows treatment of the interaction using perturbative methods. Under these assumptions, the proton-proton cross section at the LHC for a given N particle final state is given by

$$\sigma_N = \sum_{a,b} \int_0^1 dx_1 \int_0^1 dx_2 f_a(x_1, \mu^2) f_b(x_2, \mu^2) \hat{\sigma}_N^{ab} \tag{4.8}$$

where the sum is over all parton species a and b within protons 1 and 2. $f_i(x_j, \mu^2)$ is the probability (calculated at renormalization scale μ^2) of finding parton species i carrying a momentum fraction x_j of the parent proton j, and $\hat{\sigma}_N^{ab}$ is the partonic cross section for initial state $a + b$. The function $f_i(x_j, \mu^2)$ is referred to as a parton distribution function (PDF), and usually refers to the probability density for finding a particle with a certain longitudinal momentum fraction x at a resolution scale μ^2 (Figure 2.5 is an example of such distribution at the 10^4 GeV^2 mass scale).

The main task involved in simulation of the hard interaction is the evaluation of the integral in Equation 4.8. The partonic cross section is itself given by

$$\begin{aligned} \hat{\sigma}_N^{ab} = \int_{cuts} d\hat{\sigma}_N^{ab} = & \frac{(2\pi)^4 S}{4\sqrt{(p_1 \cdot p_2)^2 - m_1^2 m_2^2}} \times \\ & \int_{cuts} \left[\prod_{i=1}^{N} \frac{d^3 q_i}{(2\pi)^3 2E_i} \right] \delta^4 \left(p_1 + p_2 - \sum_i^N q_i \right) |\mathcal{M}_{p_1 p_2 \to \{\vec{q}\}}^{ab}|^2 \end{aligned} \tag{4.9}$$

where p_i are the incoming particle four-momenta, $q_i(E_i)$ are the outgoing particle four-momenta, S is a product of factors $1/j!$ for each set of j identical particles in the final state, and $\mathcal{M}_{p_1 p_2 \to \{\vec{q}\}}^{ab}$ is the parton level matrix element (ME) for the process. The event generator must build and evaluate

all Feynman diagrams associated with the given process to determine the parton level ME, or these must be hard-coded by the package authors.

The number of diagrams is directly proportional to the final state multiplicity and therefore becomes a complex problem very quickly. Next-to-leading-order (NLO) generators have recently been developed which include loop diagrams. This inclusion complicates the ME calculation enormously as divergences arise in real and virtual contributions which must cancel. Once the MEs have been evaluated, the evaluation of the multidimensional phase space integration required for the random sampling is performed using MC integration techniques[91].

The underlying event (UE) refers to all the semi-hard interactions that the spectator partons that did not take place in the initial hard interaction undergo with each other. Because these spectator interactions are typically soft, they are not calculable by perturbative methods and empirical models are used to describe them.

Bare partons may be produced as a result of the hard interaction. These strongly-interacting particles are perturbatively evolved from the scale of the hard interaction through successive branchings down to a lower energy scale at which they combine to form colorless hadrons, the hadronization scale. These successive branchings are the origin of hadronic jets, whereby individual quarks and gluons lead to a cascade of particles moving in the general direction of the original parton as they inherit its momentum. The probability for a quark or gluon to branch into two partons as it evolves from scale t to $t' < t$ as well as the kinematics of such a branching can be calculated from first principles, accurate to fixed order in the strong coupling; the results of which are known as the DGLAP evolution equations[92]. Thus, partons are recursively evolved down to the hadronization scale through successive branchings. After showering of an $N-1$ particle final state, an additional hard parton can be radiated, thus producing overlap with an N particle interaction hard state. The colored proton remnants which did not take part in the hard interaction can produce showers as well.

At the hadronization scale, the showering ceases and the colored partons group to form colorless hadrons. This regime is not amenable to perturbative calculations, and no first-principle

theory is viable. Various phenomenological methods have been developed to model hadronization, including the *Lund-string-model*[93]. In the Lund model, quarks q and anti-quarks $\bar{q}$ are the end points of Lund strings, in a similar way to that of the positive and negative charges of electric dipoles. Unlike the electrodynamic field, the chromodynamic field is self-interacting, and the field lines can be thought of as forming a bundle or a string. Gluons g are excitations or force carriers on these string lines. Once strings are formed, they dissolve or fragment into hadrons based on a probability distribution derived from data. The number, types, and kinematics of particles produced from the string fragmentation depend upon the number, types and kinematics of partons that form them.

In this analysis the generation of physical events proceeds in three steps, the first one being the ME calculation using MADGRAPH[94]. Then PYTHIA[95] is used to simulate parton shower and hadronization.

PYTHIA is a general purpose, tree level partonic matrix element generator capable of performing parton showering, hadronization, and UE simulation. A variety of $2 \rightarrow 1, 2, 3$ processes are included. Full spin correlations are included in the decays of unstable resonances. Shower evolution proceeds in terms of decreasing time-like virtuality, and imposes angular ordering by veto. Shower evolution is accurate to the LL level. The Lund string model is used for hadronization. UE interactions are described perturbatively as multiple nearly-independent $2 \rightarrow 2$ scatterings. For this study, the values for $xqcut = 30$ and $qcut = 60$ are used to simulate signal samples.

4.6.2 Detector Simulation

The next step in the simulation of events chain is the simulation of how particles will interact with the detector and its constituent materials and how the readout electronics will behave. To simulate the response of the CMS detector, the generators are interfaced with a sophisticated detector simulation based on the GEANT4[96] software package, which takes into account the exact detector geometry as well as all materials used.

The alignment, calibration, and other conditions which may change over time are periodically checked and stored in a database. These conditions are used for both offline simulation and re-

construction as well as for online activities. A snapshot of the conditions at some point in time is called a global tag. For reference, this analysis uses the $80X\_dataRun2\_2016LegacyRepro\_v4$ and $80X\_mcRun2\_asymptotic\_2016\_TrancheIV\_v8$ global tags for data and simulation, respectively.

The final state particles from the event generator are sent to the detector simulation, which tracks the particles as they move through the detector depositing energy into what are called simulated hits. While the models of electromagnetic interactions are extremely precise, the hadronic interactions have a greater uncertainty associated with them. The simulation goes through the data acquisition process, simulating the responses of the avalanche photodiodes and readout electronics. The resulting information is then analyzed by the same reconstruction process that the real data undergoes and is stored using the ROOT software library.

5. ANALYSIS

In this study, a search for a Z' coupling primarily to bottom quarks is presented. Specifically, the bottom-fermion fusion channel with two initial state gluons splitting to bottom pairs, two bottoms of which generate the Z' that subsequently decays to bottoms again, is aimed at.

The anomalies in B-meson decays reported by the Belle and LHCb experiments are the primary motivation for the search for a Z' coupling to bottom quarks and also able to embody a flavor-chainging neutral current between s- and b-quarks. Searches for a new heavy neutral boson have been performed previously[3, 97, 98, 99, 100, 101], though the bottom-fermion fusion channel itself has so far only been explored in a Delphes study with the Z' decaying to muons[102].

This search is performed using a data sample of proton-proton collisions at a center-of-mass energy of 13 TeV collected by the CMS experiment in 2016 and corresponding to an integrated luminosity of $35.9 fb^{-1}$. Data for this analysis were collected using an HLT path that uses multi-jet and b-tagging requirements as described in Section 5.4. The Z' mass is reconstructed offline from its decay products out of the two leading jets in p_T in the event. Summarily, a bump in the falling invariant mass spectrum of background events would indicate the presence of a new background.

As one could expect from an all-hadronic final state search, such a signal would be buried under an immense background of multi-jet events produced by QCD interactions. For this purpose, a signal region (SR) is identified by means of a specific b-tagging selection on the four leading jets on p_T and an offline trigger as described in Sections 5.5 and 5.4. The background control regions (CR) are orthogonal to the SR in terms of their respective trigger and b-tagging selection criteria. Then, a data-driven method is used to estimate the background contribution in the SR due to the poorly-simulated multi-jet QCD backgrounds.

5.1 Data and Monte Carlo Samples

5.1.1 Data

The analysis makes use of the full 2016 CMS dataset of 13 TeV data. Table 5.1 shows the data samples used for this analysis, which correspond to a total of $35.9 fb^{-1}$ of data.

Dataset	Run Range	Integrated Luminosity
/BTagCSV/Run2016B-07Aug17-v*/AOD	272007-275376	5.8
/BTagCSV/Run2016C-07Aug17-v1/AOD	275657-276283	2.5
/BTagCSV/Run2016D-07Aug17-v1/AOD	276315-276811	4.3
/BTagCSV/Run2016E-07Aug17-v1/AOD	276831-277420	4.1
/BTagCSV/Run2016F-07Aug17-v1/AOD	277772-278808	3.1
/BTagCSV/Run2016G-07Aug17-v1/AOD	278820-280385	7.5
/BTagCSV/Run2016H-07Aug17-v1/AOD	280919-284044	8.5
Total BTagCSV	**272007–284044**	**35.9**
/SingleMu/Run2016B-07Aug17-v*/AOD	272007-275376	5.8
/SingleMu/Run2016C-07Aug17-v1/AOD	275657-276283	2.5
/SingleMu/Run2016D-07Aug17-v1/AOD	276315-276811	4.3
/SingleMu/Run2016E-07Aug17-v1/AOD	276831-277420	4.1
/SingleMu/Run2016F-07Aug17-v1/AOD	277772-278808	3.1
/SingleMu/Run2016G-07Aug17-v1/AOD	278820-280385	7.5
/SingleMu/Run2016H-07Aug17-v1/AOD	280919-284044	8.5
Total SingleMu	**272007–284044**	**35.9**

Table 5.1: The datasets analyzed for this analysis.

Data is processed in the CMSSW_9_4_9 framework to yield *n-tuples*, which include a subset of the information contained in the official CMS datasets. These smaller datasets were analyzed using privately created analysis software.

The other years (2017/2018) at the same center-of-mass energy of 13 TeV were not considered, due to changes in the trigger menu, necessitating a different analysis strategy than presented in this study.

5.1.2 Monte Carlo

This analysis makes use of MC simulation to give a rough idea of the hadronic background behavior, as well as an estimate on how a potential Z' signal would look like in data. Table 5.2 lists

all of the MC samples for the considered SM background processes, which are centrally produced by the CMS collaboration. Every dataset name is followed by /RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM.

The signal sample kinematics and yields are taken from MC. The datasets for this purpose were created privately and specifically for this analysis. The MC signal events are simulated to leading order (LO) in QCD precision. Different samples are generated for $m_{Z'}$ ranging from 350 to 950 GeV in 150 GeV steps. Also, the model attempts to exclude a phase space spanned by g_b and δ_{bs} and therefore, different values for δ_{bs} were considered ranging from 0.0 to 1.0 in steps of 0.25. The coupling parameter g_b is set to 1.0. The cross-section for the signal samples can be found in Table 5.3. The MADGRAPH5 v4.2.4[94] event generator was used with the NNPDF3.0[103] leading order PDFs taken from the LHAPDF 6.2.1 PDF set. The showering and hadronization of partons were simulated with PYTHIA 8[104] with the CUETP8M1 NNPDF23LO tune. Finally, the hadronization and reconstruction were performed by using the CMSSW release 8_0_26.

MC background events were generated using the leading-order matrix element generator MADGRAPH5 v4.2.4. Parton shower and hadronization are included using PYTHIA 8, and the matrix element is matched to the parton shower using the MLM scheme. The Z2* tune is used to describe the underlying event. This tune is identical to the Z1 tune but uses the CTEQ6L PDFs.

All generated events are processed through a GEANT4 simulation of the CMS apparatus. Additional proton-proton interactions within a bunch crossing (pileup) are added to the simulation, with a frequency distribution chosen to match that observed in data. During this data-taking period, the mean number of interactions per bunch crossing is 25.

5.1.2.1 Multi-jet QCD Background

It is well-known that the QCD process is challenging to model to the desired level of accuracy. Additionally, the event selection in the analysis requires four jets. This requirement reduces vastly the number of QCD MC events that pass the selection criteria. Therefore, when using the MC samples, we are left with a statistically limited sample that is almost useless for describing this background.

Background Processes		
Process	Dataset Name	Cross Section [pb]
QCD_HT50to100	QCD_HT50to100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	246300000.0
QCD_HT100to200	QCD_HT100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	28060000.0
QCD_HT200to300	QCD_HT200to300_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1710000.0
QCD_HT300to500	QCD_HT300to500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	347500.0
QCD_HT500to700	QCD_HT500to700_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	32060.0
QCD_HT700to1000	QCD_HT700to1000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	6829.0
QCD_HT1000to1500	QCD_HT1000to1500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1207.0
QCD_HT1500to2000	QCD_HT1500to2000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	120.0
QCD_HT2000toInf	QCD_HT2000toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	25.25
TTJets	TTJets_TuneCUETP8M2T4_13TeV-amcatnloFXFX-pythia8	831.76

Table 5.2: List of background MC datasets and cross sections used in the analysis.

Signal samples			
$m_{Z'}$[GeV]	δ_{bs}	g_b	Cross Section [pb]
350	0.0	1.0	232.71
	0.25	1.0	296.8
	0.5	1.0	493.1
	0.75	1.0	821.9
	1.0	0.85	928.1
500	0.0	1.0	38.86
	0.25	1.0	51.59
	0.5	1.0	92.38
	0.75	0.95	150.3
	1.0	0.8	176.3
650	0.0	1.0	11.66
	0.25	1.0	15.635
	0.5	1.0	28.96
	0.75	0.9	43.01
	1.0	0.8	57.49
800	0.0	1.0	4.388
	0.25	1.0	5.947
	0.5	1.0	11.2
	0.75	0.9	16.9
	1.0	0.8	22.69
950	0.0	1.0	1.902
	0.25	1.0	2.604
	0.5	1.0	4.979
	0.75	0.9	7.582
	1.0	0.8	10.24

Table 5.3: List of signal MC datasets used in the analysis and their corresponding cross-sections.

Rather than relying on MC for the QCD background sample, a data-driven method to estimate the contribution from background in the signal region will be used. This method is explained in detail in Section 5.6.

5.2 Object Selection

As described in Chapter 4, CMS provides every user a list of reconstructed objects (i.e., jets, electrons, etc.) to be used in the analysis. However, the reconstruction algorithms are intentionally generic so that the objects they return apply to a wide array of physics analyses. Specific groups within CMS called physics object groups (POGs) are responsible for developing object quality criteria which must be implemented by each study to prevent fake or poorly reconstructed objects. This section will discuss the object selection criteria to identify jets, which all meet or exceed the object requirements as set by the relevant POGs. Only events which contain a sufficient number of objects of the right quality will be used in this analysis.

Jets must meet the Tight Jet ID criteria, as described in Section 4.4. Moreover, selected jets must have a $p_T > 30$ GeV and $|\eta| < 2.4$ to always have b-tag discriminator information available.

5.3 Data and MC Corrections

Despite the efforts to make sure the MC simulation models data properly, there can still exist discrepancies between the observed data and simulation. Often this occurs because the exact data-taking conditions are not known in advance, like the pileup conditions during the actual run of the experiment. In other cases, these discrepancies arise from the precision of the physics generators, which might not be enough to model data accurately. Detector conditions may also change over time due to e.g. radiation damage. Data, as a measurement cannot be wrong, only wrongly interpreted.

In Section 4, I have already discussed some object-specific corrections like jet energy corrections and resolution. For other discrepancies, it is often necessary to reweight the full event rather than a specific object, and they are particular to the analysis. In the following, the corrections used will be explained.

5.3.1 JEC and Residuals

The details of the factorized approach adopted by CMS for jet energy calibration were discussed in Section 4.4. The final step in this chain is aiming to provide the same jet energy scale to all events. The residual corrections applied to data, account for data to MC differences after simulated corrections are applied.

5.3.2 JER Smearing

The jet p_T resolution is relatively broad compared to that of many other physics objects, and the biases caused by jet resolution smearing can be crucial for finding the signature of a particle like the Z' on top of the steeply falling QCD spectrum. At CMS, the particle-level JER is determined from MC simulation, and then a data/MC scale factor is obtained from data-based methods. Note that JECs are applied as described in Section 4.4 before deriving JER.

A detailed description of the methods can be found at [105]. Still, strictly speaking, one could say that the measurement of JER is an extension of the methods used for measuring JES, but instead of looking at the mean of the response distribution, we are interested in its width.

5.3.3 Pileup Re-weighting

Up to this point, pileup has been described as additional proton-proton interactions within an event, besides the process that produced the physics objects we are interested in studying. Still, there are several properties of pileup worth noting. First, pileup is related to additional objects (tracks or energy), which might be found in the same bunch crossing as the event under study. In reality, there are two different kinds of pileup. There is indeed the kind that comes from additional proton-proton interactions within the same bunch crossing, known as "in-time" pileup. Still, there is also energy from pileup added to objects because it was left in the sub-detectors from bunch crossings before or after the current one. This effect is known as the "out-of-time" pileup and is mostly because the integration window of the sub-detectors can be larger than 25 ns.

Second, the exact number of proton-proton interactions within an event, μ is related to the instantaneous luminosity. The latter can vary within any given data-taking period and even within

a luminosity section (LS). As a benchmark, the average number of proton-proton interactions per bunch crossing in 2016 was 23[21] as it can be seen from 5.1.

Finally, pileup can affect reconstruction efficiency and even the observed kinematics of all the objects used in the analysis. Therefore it is of vital importance that the pileup distribution in MC follows that of data as closely as possible.

Since the MC samples used in CMS are usually produced before the data was taken, the actual pileup conditions were unknown during the generation process. Thus, MC was generated assuming specific situations that might not match data that well. A broad distribution of μ values, the number of mini-bias pileup events overlaid on the hard scatter event, is generally chosen so as to cover all pileup conditions which might be experienced over the course of a data-taking period. Unsurprisingly, the anticipated μ distribution rarely matches the observed one in data and thus the MC must be reweighted such that the μ distributions match [106].

To generate a histogram for the average number of interactions per bunch crossing coming from data, we make use of the approved pileupCalc tool provided by CMS. This tool takes as input the measured total inelastic cross-section $\sigma_{inelastic}$ = 69.2 mb, a "pileup file" in JSON format, and the JSON file used in the analysis to select "good events." CMS provides the JSON files for every data-taking period. These files include every run number and luminosity section, which are matched to a given average instantaneous luminosity.

The per-event weights as a function of μ are created by dividing the normalized distribution from data by the normalized MC based distribution. The weights are then applied to each MC event by looking up the weight for the true number of pileup interactions used to generate that specific event[107]. The distributions of pileup interactions in MC and data can be seen in Figure 5.2 for MC generated signal for a Z' with a mass of 500 GeV.

5.3.4 Deep CSV Reweighting

In Section 4.5, a criterion for tagging a jet as being produced by a b quark by using the DeepCSV discriminant was introduced. The derivation of this discriminant is described in [80].

This analysis relies heavily on the ability to identify b jets, so the discriminant, or at least

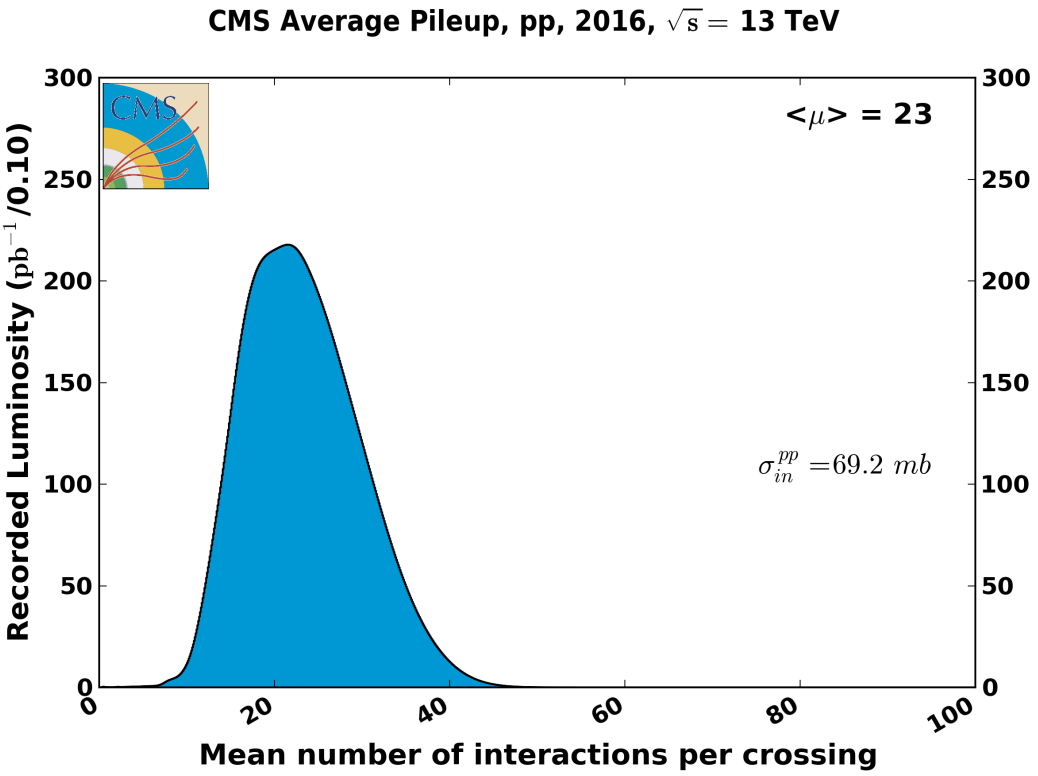

Figure 5.1: Mean number of interactions per bunch crossing for the 2016 proton-proton run at $\sqrt{s}$=13 TeV. The two plots shown here use the same data, but different values for the minimum bias cross-section were assumed. The left plot uses the "CMS recommended" value of 69.2 mb, which is determined by finding the best agreement with data and is recommended for CMS analyses. Reprinted from [21].

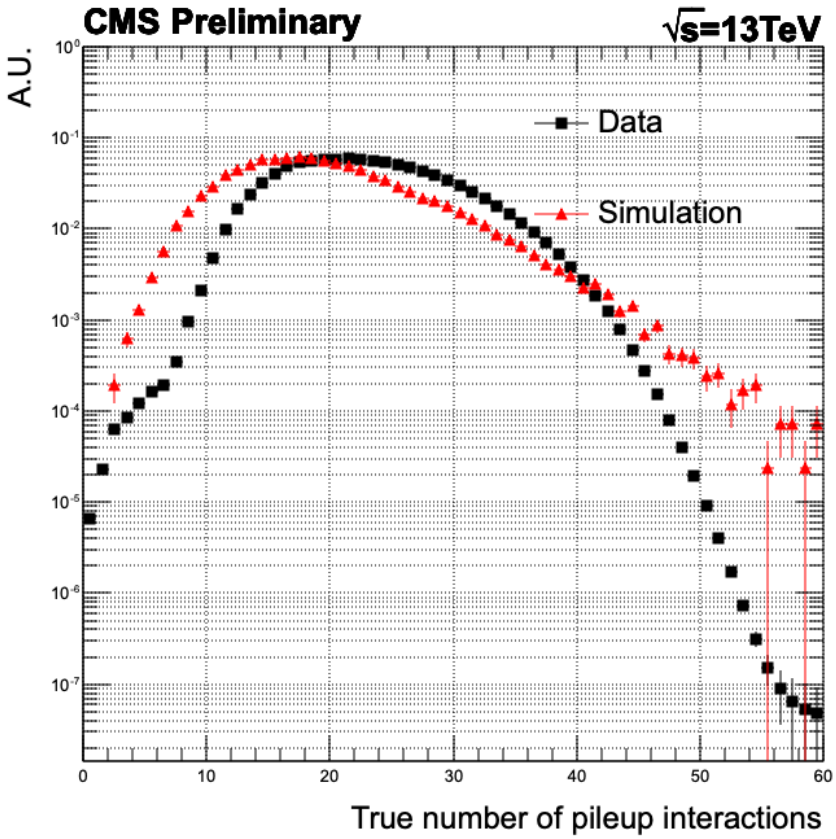

Figure 5.2: Distributions of the true number of pileup interactions in data and simulation. The MC signal corresponds to a 500 GeV Z' with a δ_{bs} = 0 and g_b = 1. Distribution from data corresponds to the full 2016 dataset (35.9 fb^{-1}).

the working point used should behave the same way in both data and MC to accurately describe the rate of observing a b jet. The tagging efficiency in data is usually not the same as that in MC, so a correction to the DeepCSV discriminant or tagging probability must be made. For this analysis, an event reweighting is done using scale factors provided by the b-tagging POG[108] and MC b-tagging efficiencies computed for each signal MC sample. The goal of this method[109] is to predict the correct event yield in data by giving each event a weight respective to differences pertaining to the working point combination used in that event and each working point's difference in (mis-)tagging probability with respect to data, in the form of scale factors (SF).

The probability of a given configuration of jets in MC simulation and data is defined as:

$$P(MC) = \prod_{i=tagged} \epsilon_i \prod_{j=nottagged} (1 - \epsilon_j) \tag{5.1}$$

$$P(Data) = \prod_{i=tagged} SF_i\epsilon_i \prod_{j=nottagged} (1 - SF_j\epsilon_j) \tag{5.2}$$

where ϵ_i is the MC b-tagging efficiency for a given exclusive b-tagging working point.

The B-tag POG measures the b-tagging efficiency scale factors for b and light flavor jets. These scale factors, as well as the MC efficiencies, depending on the jet flavor, jet p_T, and jet η. The b jet scale factors are divided into five p_T bins of p_T < 40 GeV, 40 GeV < p_T < 60 GeV, 60 GeV < p_T 100 GeV, 100 GeV < p_T 160 GeV, p_T > 160 GeV. Also, only jets with $|\eta| \geq$ 2.4 are considered.

The event weight is then calculated as

$$w = \frac{P(Data)}{P(MC)} \tag{5.3}$$

Since multiple operating points are used in the analysis, the "tagged" or "not tagged" categories become more complex. The expressions above then take the form:

$$P(MC) = \prod_{i=taggedT} \epsilon_i^T \prod_{j=taggedM,notT} (\epsilon_j^M - \epsilon_j^T) \prod_{k=taggedL,notM} (\epsilon_k^L - \epsilon_k^M) \prod_{l=nottagged} (1 - \epsilon_l^L) \tag{5.4}$$

$$P(Data) = \prod_{i=taggedT} SF_i^T \epsilon_i^T \prod_{j=taggedM,notT} (SF_j^M \epsilon_j^M - SF_j^T \epsilon_j^T) \prod_{k=taggedL,notM} (SF_k^L \epsilon_k^L - SF_k^M \epsilon_k^M) \prod_{l=nottagged} (1 - SF_l^L \epsilon_l^L) \quad (5.5)$$

5.4 Trigger

Events used in this analysis are selected by trigger algorithms that save events with a large amount of hadronic activity. The L1 seeds for these triggers either select events with a large scalar sum of transverse jet momentum (H_T) or that contain a jet with a high p_T. The HLT algorithm then places requirements on either the event H_T, jet p_T, or jet tagging.

The final state, characterized by four b-jets, is particularly challenging to trigger on with the usual multi-trigger paths due to the immense rate of QCD multi-jet events. A trigger exploiting the b-tag information at the HLT level was chosen to allow low thresholds for jet p_T appropriate for the analysis and an acceptable signal/background rate.

The use of b-tagging at the trigger level also reduces the multi-jet contribution to the selected data sample. The paths used are DoubleJet90_Double30_TripleCSV087 and Quad45_TripleCSV087 and they are described below.

Schematically, these HLT paths seed on events that contain basically high L1 HT. Other multi-jet seeds have been used in logic OR to increase the L1 trigger efficiency. Then, the HLT paths select events with four HLT anti-kT4 jets above four optimized p_T thresholds, and then require a minimum HLT b-tagging value for three jets with full regional HLT tracking information.

The HLT path structure can be summarized as:

DoubleJet90_Double30_TripleBTagCSV_p087:

- L1_TripleJet_84_68_48_VBF OR L1_TripleJet_88_72_56_VBF OR
  L1_TripleJet_92_76_64_VBF OR
  L1_HTT280 OR L1_HTT300 OR L1_HTT320 OR
  L1_SingleJet170 OR L1_SingleJet180 OR
  L1_SingleJet200 OR L1_DoubleJetC100 OR

L1_DoubleJetC112 OR L1_DoubleJetC120

- Reconstruct anti-k_T 0.4 L1FastJet corrected Calo-jets
  - 2 jets with $|\eta|$ < 2.6 and p_T > 90 GeV
  - 4 jets with $|\eta|$ < 2.6 and p_T > 30 GeV
- Fast Primary Vertex Reconstruction
- Online CSV computation
  - 3 Calo-jets with CSV>0.87
- PF reconstruction sequence
  - 2 PF jets with $|\eta|$ < 2.6 and p_T > 90 GeV
  - 4 PF jets with $|\eta|$ < 2.6 and p_T > 30 GeV

QuadJet45_TripleBTagCSV_p087:

- L1_QuadJetC50 OR L1_QuadJetC60 OR L1_HTT280 OR L1_HTT300 OR L1_HTT320 OR L1_TripleJet_84_68_48_VBF OR L1_TripleJet_88_72_56_VBF OR L1_TripleJet_92_76_64_VBF
- Reconstruct anti-k_T 0.4 L1FastJet corrected Calo-jets
  - 4 jets with $|\eta|$ < 2.6 and p_T > 45 GeV
- Fast Primary Vertex Reconstruction
- Online CSV computation
  - 3 Calo-jets with CSV>0.87
- PF reconstruction sequence

– 4 PF jets with $|\eta| < 2.6$ and $p_T > 45$ GeV

The triggers described here employ an online version of the CSV algorithm which is implemented at Level 3 of the HLT and therefore does not have access to the full collections available to its offline version counterpart as explained below.

- **Primary Vertex** Only information from the pixel detector is used to reconstruct the primary vertex position. Pixel tracks that are compatible with vertices reconstructed through the Fast Primary Vertex algorithm are used.

- **Tracks** While full tracker information is available at L3, not all tracks are reconstructed due to time limitations. The resulting track collection consists of tracks compatible with the "Pixel Primary Vertex" and the eight leading jets in p_T in the event.

- **Jets** Only Calo-jets are available to the HLT CSV algorithm, while PF jets are available to the offline-CSV algorithm.

5.4.1 Trigger Efficiency Estimation

We follow a data-driven approach to measure the trigger efficiency, which can be measured in the following way:

$$P(T) = \frac{N(T)}{N_{tot}} \tag{5.6}$$

where $P(T)$ is the probability of the trigger T to pass an event, $N(T)$, and N_{tot} are the number of triggered and total events. For those passing a particular offline selection, this expression becomes:

$$P(T|S) = \frac{N(T\&S)}{N_{tot}} \tag{5.7}$$

where $P(T|S)$ is the probability of the trigger T to accept an event after the selection S, $N(S)$ is the number of events that satisfy the selection S, and $N_{T\&S}$ is the number of events that meet both the selection and the trigger T requirements.

The turn-on curve can be interpreted as a set of $P(T|S_i)$, where each S_i corresponds to a bin of the efficiency distribution. Let us assume the trigger is composed by two requirements on some properties of trigger objects, C_1 and C_2. We can write the trigger efficiecy as the conditional probability of:

$$P(T|S) = P(C_1\&C_2|S) \tag{5.8}$$
$$= \frac{N(C_1\&C_2\&S)}{N(S)} \tag{5.9}$$
$$= \frac{N(C_1\&C_2\&S)}{N(C_1\&S)} \frac{N(C_1\&S)}{N(S)} \tag{5.10}$$
$$= P(C_2|S, C_1) \cdot P(C_1|S) \tag{5.11}$$

which can be generalized to:

$$P(T|S) = P(C_1\&C_2\&C_3\&...|S) \tag{5.12}$$
$$= P(C_2\&C_3\&...|S, C_1) \cdot P(C_1|S) \tag{5.13}$$
$$= P(C_3\&...|S, C_1, C_2) \cdot P(C_2|S, C_1) \cdot P(C_1|S) \tag{5.14}$$
$$= \prod_{i=1}^{n} P(C_i|S, C_1, ..., C_{i-1}) \tag{5.15}$$

for an n number of C_n cuts in the selection. This means that the efficiency of the trigger T can be evaluated as a product of the efficiency of a single cut given the previous cuts.

Now, we want to estimate the efficiency of a given trigger as a function of the offline variables to obtain a set of weights whose product would yield the same effect as requiring the trigger. For example, if we take the specific case of a trigger which has the following three cuts:

- C_1: four calo-jets with p_T > 45 GeV
- C_2: three calo-CSV CSV > 0.3

- C_3: four PF-jets with p_T > 45 GeV

and we want to measure efficiency as a function of the offline variables. Let's assume that the cuts depend only on the offline variables CSV_3 and p_{T4} (CSV value of the third jet in decreasing CSV and the p_T of the fourth jet in decreasing order of p_T). We can estimate the trigger efficiency as follows:

$$P(T|S) = P(C_1|S) \cdot P(C_2|S, C_1) \cdot P(C_3|S, C_1, C_2) \tag{5.16}$$

The first term can be evaluated using the events collected by the ZeroBias trigger (or any trigger that is independent of T), and it can be parameterized with a turn-on function:

$$f_1(p_{T4}) = P(C_1|p_{T4}) \tag{5.17}$$

The turn-on of the second term must be evaluated on the events that pass the first cut. Then, we can obtain the efficiency of the third term as:

$$f_3(p_{T4}) = P(C_3|C_1, C_2, p_{T4}) \tag{5.18}$$

Once all turn-on curves are fitted we can obtain the trigger efficiency as a function of p_{T4} and CSV_3 in the following way:

$$P(T|CSV_3, p_{T4}) = P(C_3|S, C_1, C_2) \cdot P(C_2|S, C_1) \cdot P(C_1|S) = f_3(p_{T4}) \cdot f_2(CSV_3) \cdot f_1(p_{T4}) \tag{5.19}$$

In other words, if we collected data with the trigger T, we get data/MC agreement applying to the simulated events the following weights:

$$w(CSV_3, p_{T4}) = f_3(p_{T4}) \cdot f_2(CSV_3) \cdot f_1(p_{T4}) \tag{5.20}$$

For the particular case of the triggers selected for this analysis, the efficiency is measured in data using the SingleMuon dataset. We start by selecting events that pass the *HLT_BIT_HLT_IsoMu24* and an additional cut on quality of the event muon: loose muon POG ID and $p_T > 40$ GeV. We also, require that there are at least four jets with p_T > 30 GeV. Then, we obtain weights for each trigger according to specific selection requirements, which will be explained in detail in the following sections.

5.4.1.1 QuadJet45_TripleBTagCSV_p087 efficiency

The weights for the QuadJet45_TripleBTagCSV_p087 trigger can be parameterized as follows:

$$\begin{aligned} w_{Quad45}(p_{T1}, p_{T2}, p_{T3}, p_{T4}, CSV_3) = & \, TurnOnL1Pt1Pt2Pt3Pt4(p_{T1} + p_{T2} + p_{T3} + p_{T4}) \\ & \cdot TurnOnCaloPt4(p_{T4}) \\ & \cdot Eff(CSV_3) \cdot TurnOnPFPt4(p_{T4}) \end{aligned} \tag{5.21}$$

where:

- C_1 = TurnOnL1Pt1Pt2Pt3Pt4($p_{T1} + p_{T2} + p_{T3} + p_{T4}$): turn-on curve for events passing the L1 trigger requirements, as a function of the p_T sum of the four leading calo jets.
- C_2 = TurnOnCaloPt4(p_{T4}): turn on curve for events passing the hltQuadCentralJet45 trigger and C_1, as a function of the fourth leading calo jet in p_T.
- C_3 = Eff(CSV_3) is the efficiency for events passing C_1, C_2, and the hltBTagCaloCSVp087Triple trigger, as a function of the third leading jet in offline CSV value.
- C_4 = TurnOnPFPt4(p_{T4}): turn on curve for events passing the hltQuadPFCentralJetLooseID45 trigger, C_1, C_2, and C_3 as a function of the fourth leading PF jet in p_T.

These turn on curves are shown in Figure 5.3. The trigger efficiency is parameterized by using a fit to the individual turn-on curves as a function of the p_T, sum of p_T, or CSV value of calo or PF jets.

The overall trigger efficiency is validated using the same preselection used for the estimate. We compare distributions derived by applying the weights as computed in 5.21 with what we obtain if

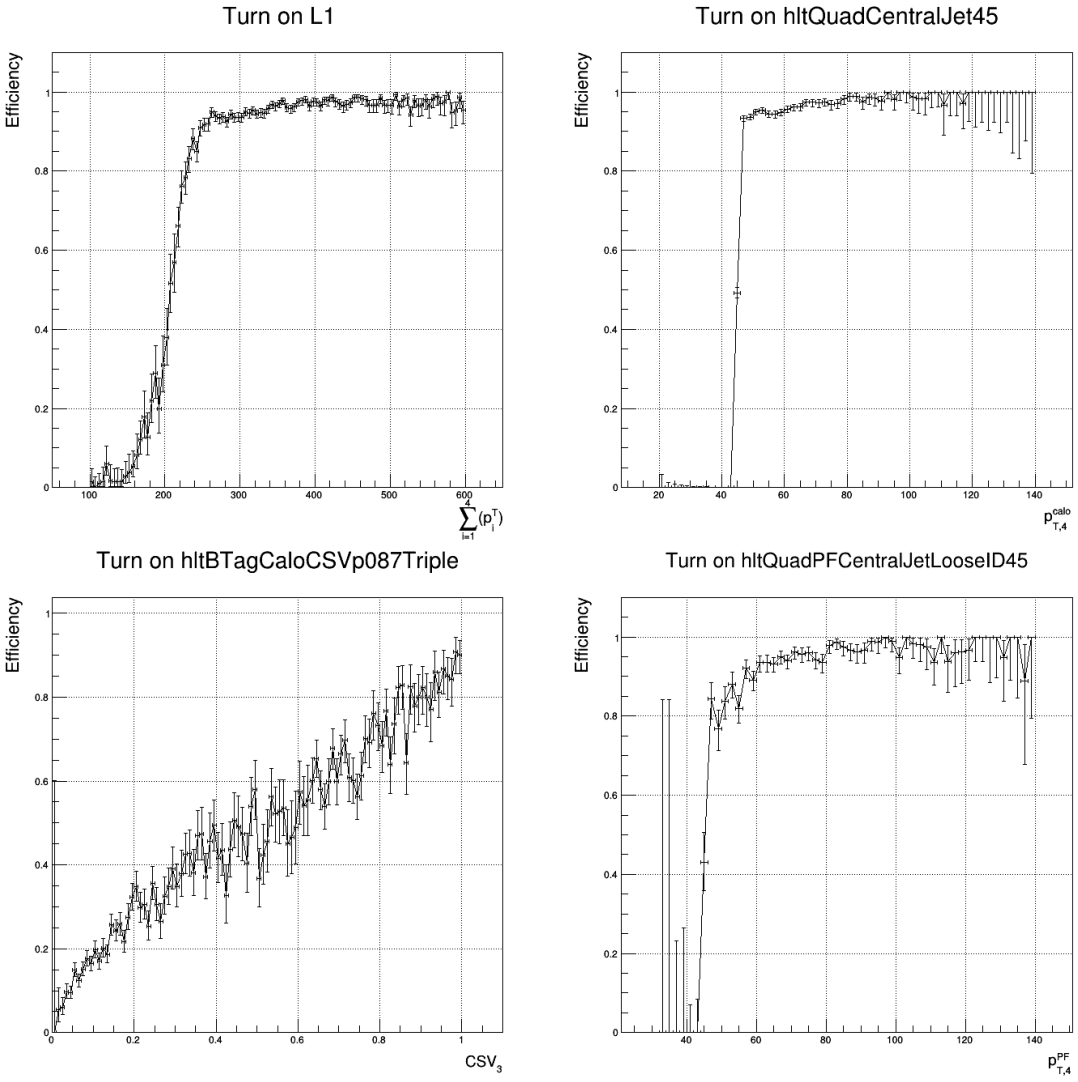

Figure 5.3: Turn on plots for QuadJet45_TripleBTagCSV_p087 trigger.

we apply the trigger bit, reporting a good agreement that validates the method. The results can be found in Figure 5.4.

5.4.1.2 DoubleJet90_Double30_TripleBTagCSV_p087 efficiency

The weights for the DoubleJet90_Double30_TripleBTagCSV_p087 trigger can be parameterized as follows:

$$\begin{aligned} w_{Di90Di30}(p_{T1}, p_{T2}, p_{T3}, p_{T4}, CSV_3) = & \, TurnOnL1Pt1Pt2Pt3Pt4(p_{T1} + p_{T2} + p_{T3} + p_{T4}) \\ & \cdot TurnOnCaloPt4(p_{T4}) \\ & \cdot TurnOnCaloPt2(p_{T2}) \cdot Eff(CSV_3) \\ & \cdot TurnOnPFPt4(p_{T4}) \cdot TurnOnPFPt2(p_T) \end{aligned} \tag{5.22}$$

where the measured efficiencies for the L1 trigger requirements, calo jets, and PF jets selections are shown in Figure 5.5.

As in the case for the Quad45 trigger, a closure test was performed to validate the method, as can be seen in Figure5.6.

In both cases, there is good agreement between weighted/trigger events, particularly in the region of interest.

5.5 Event Selection

As described previously, the signal consists of fully hadronic final states with four jets coming from b-hadrons. Therefore, we have designed our event selection to identify events containing at least four central jets with a $p_T > 30$ GeV and passing the Tight Jet ID criteria. Two of these jets (the most energetic ones) are expected to be associated with the resonance, and therefore their reconstructed invariant mass is used as the discriminant variable.

Furthermore, since b jet identification plays a crucial role in identifying the signal, we identified a set of tagging requirements on the four leading jets in p_T that yield the highest significance in MC. These were obtained by first calculating the probability for each jet in the event to be exclusively categorized (or tagged) as T, M, L, or X (not tagged). Categorization according to the DeepCSV algorithm. See Table 5.4 for selection values.

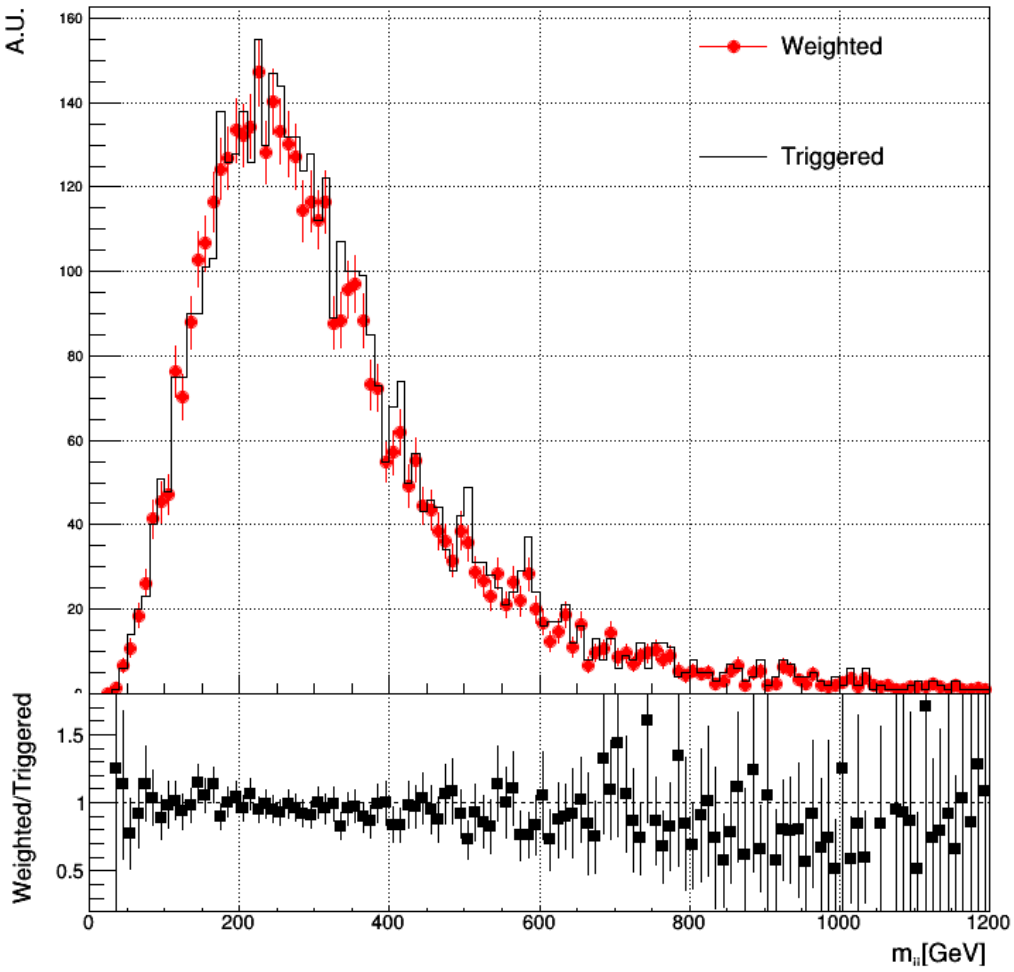

Figure 5.4: Invariant mass distribution for weighted (red) and triggered (black) events for QuadJet45_TripleBTagCSV_p087 trigger bit. The distribution is shown for SingleMu dataset for events passing preselection and 2T b-tagging selection.

DeepCSV working point	Value (prob b + prob bb)
T	>0.8958
M	>0.6324
L	>0.2219

Table 5.4: Working points minimum values for categorization.

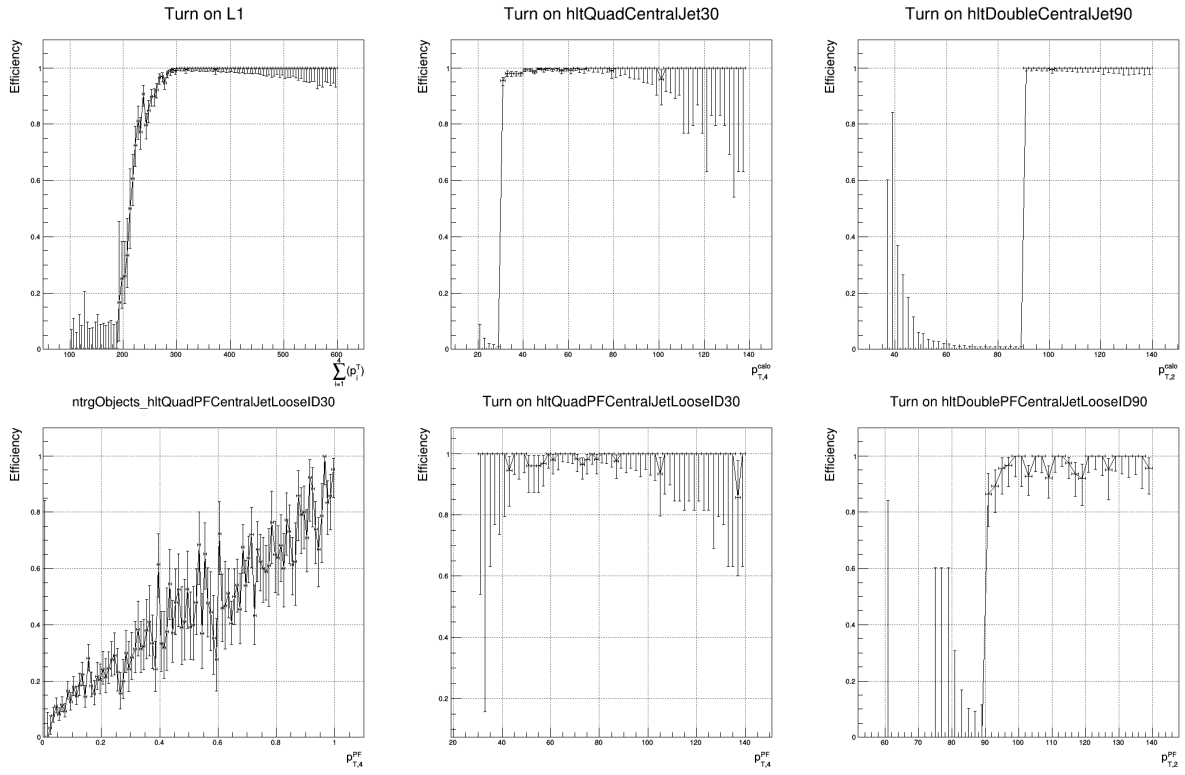

Figure 5.5: Turn on plots for DoubleJet90_Double30_TripleBTagCSV_p087.

For each permutation of T, M, L, or X (i.e., TTTT, TMTT, etc.) the individual jet tagging probability is multiplied to obtain a total bin probability, i.e.:

$$P(TTTT) = P(T, j_1) \cdot P(T, j_2) \cdot P(T, j_3) \cdot P(T, j_4) \tag{5.23}$$

Then we would calculate the significance for every permutation by using the bin probability 5.23 and the number of expected events assuming 100% acceptance (no tagging requirements), in both signal and background. Later, we started adding as many bins as possible and calculated their combined significance from the most significant bin downwards. The set of bins that provide the highest combined significance are listed in Table 5.5.

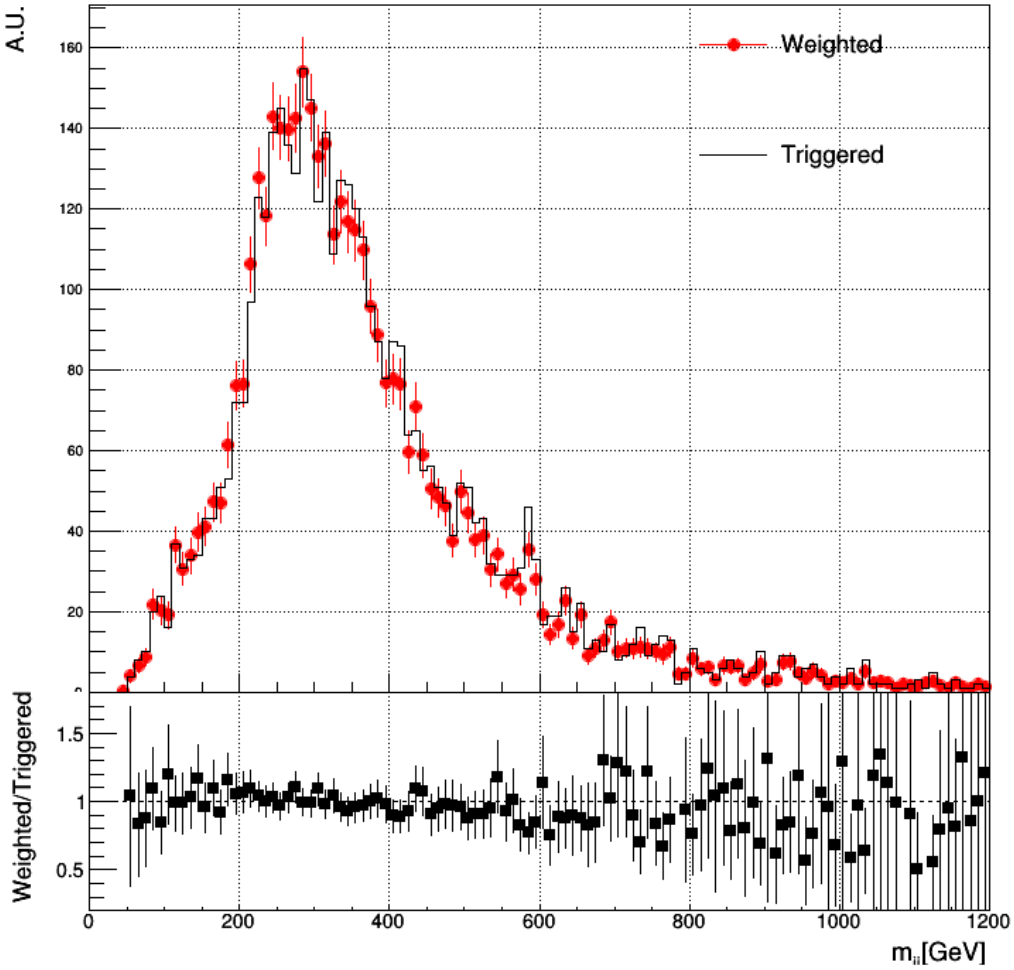

Figure 5.6: Invariant mass distribution for weighted (red) and triggered (black) events for DoubleJet90_Double30_TripleBTagCSV_p087 trigger bit. The distribution is shown for SingleMu dataset for events passing preselection and 2T b-tagging selection.

TTTT	TTTM	TTMT
TMTT	MTTT	TTTL
TTLT	TLTT	TTTX
TTMM	TMTM	MTTM
TMMT	MMTT	MTMT

Table 5.5: List of b-tagging requirements for the four leading jets in p_T in the event (T4 tag list).

5.6 Data-driven Background Estimation

Given that MC cannot be used for background estimation, a data-driven method will be used to minimize systematic uncertainties arising from poorly understood multi-jet QCD backgrounds. The background contribution can be estimated from data by finding appropriate signal/control regions. An optimal SR would have a good signal/background ratio, and a good CR would have a high-statistics background shape, low signal contamination, and a sensible way to extrapolate the background events in the CR to the background events in the SR.

We define three CR's as in Table. CR2 inversion to SR is the QuadJet45_TripleBTagCSV_p087 and not DoubleJet90_Double30_TripleBtagCSV_p087 trigger to avoid overlap.

Trigger/b-tag requirement	T4 tags	T2 tags & !T4 tags
DoubleJet90Double30	SR	CR1
QuadJet45 &!DoubleJet90Double30	CR2	CR3

Table 5.6: Control region definition.

In CR1 and 3, the tagging requirements on the 3rd and 4th leading jet on p_T have been relaxed to gain statistics while keeping those on the first two leading jets, i.e., events pass selection if the two leading jets pass any of the requirements on T2 tag list (Table 5.7).

TT	MM	TL
TM	MT	

Table 5.7: List of b-tagging requirements for the two leading jets in p_T in the event (T2 tag list).

In Figure 5.7, we compare the m_{jj} distributions in MC and data for a given CR. From here, we can see that MC reproduces the shape of data approximately, but fails to provide enough statistics or accurate normalization. For this reason, we will obtain the distribution of m_{jj} in both SR and

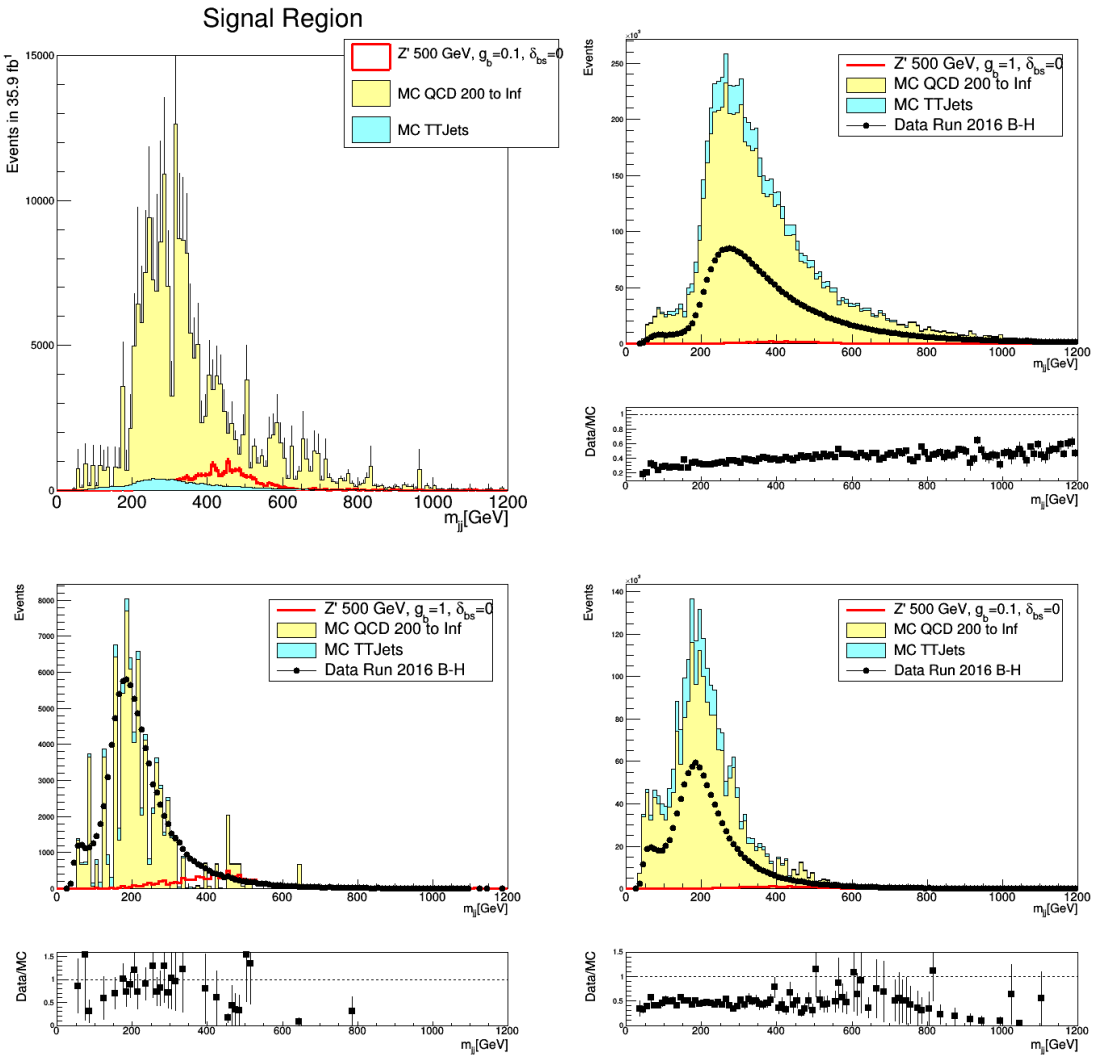

Figure 5.7: Invariant mass distribution in data and MC for the three control and signal regions defined in the text. Signal MC corresponds to $m_{Z'}$ = 500 GeV with δ_{bs}=0 and g_b=1.

CR from data. The background distribution in the SR is expected to follow the bin by bin products of the distribution in CR1*CR2/CR3.

5.6.1 Method Validation

To validate the data-driven method without looking at the SR, we have defined four additional CR's. These are a split of CR1 and CR3 into a triple, not quadruple b-tagged and double, not triple b-tagged events (Table 5.8). The T3 tag list was obtained by ommiting the tagging requirements

on the fourth leading jet on the T4 tag list (Table 5.9).

Trigger/b-tag requirement	T3 tags & !T4 tags	T2 tags & !(T4 tags ‖ T3 tags)
DoubleJet90Double30	CR1A	CR1B
QuadJet45 &!DoubleJet90Double30	CR3A	CR3B

Table 5.8: Control region definition for in-data validation.

TTT	TMM	MTT
TMT	MTM	TLT
TTL	TTM	MMT

Table 5.9: List of b-tagging requirements for the three leading jets in p_T in the event (T3 tag list).

This allows for a cross-check in data that it is not significantly signal-contaminated, as we can see in Figure 5.8.

Figure 5.9 shows the m_{jj} distribution in CR1A and prediction (CR1B*CR3A/CR3B) for the data sample. Since the ratio of these two distributions is flat in our region of interest, we can conclude that this method works as expected.

5.7 Yields

After applying all of the object and event selections, corrections, and event weights, we can now look at the expected yields for the simulated signals and background. Table 5.10 shows the event yields for the signal region. Also, Table 5.10 includes the acceptance in percentage. Here, event yields have been normalized to the total number of expected events in the sample for a given cross-section.

5.8 Systematic Uncertainties

The input to the statistical analysis is a set of histograms of the Z' m_{jj} distributions and their associated systematic uncertainties. Given that this is a shape analysis, it is essential to consider

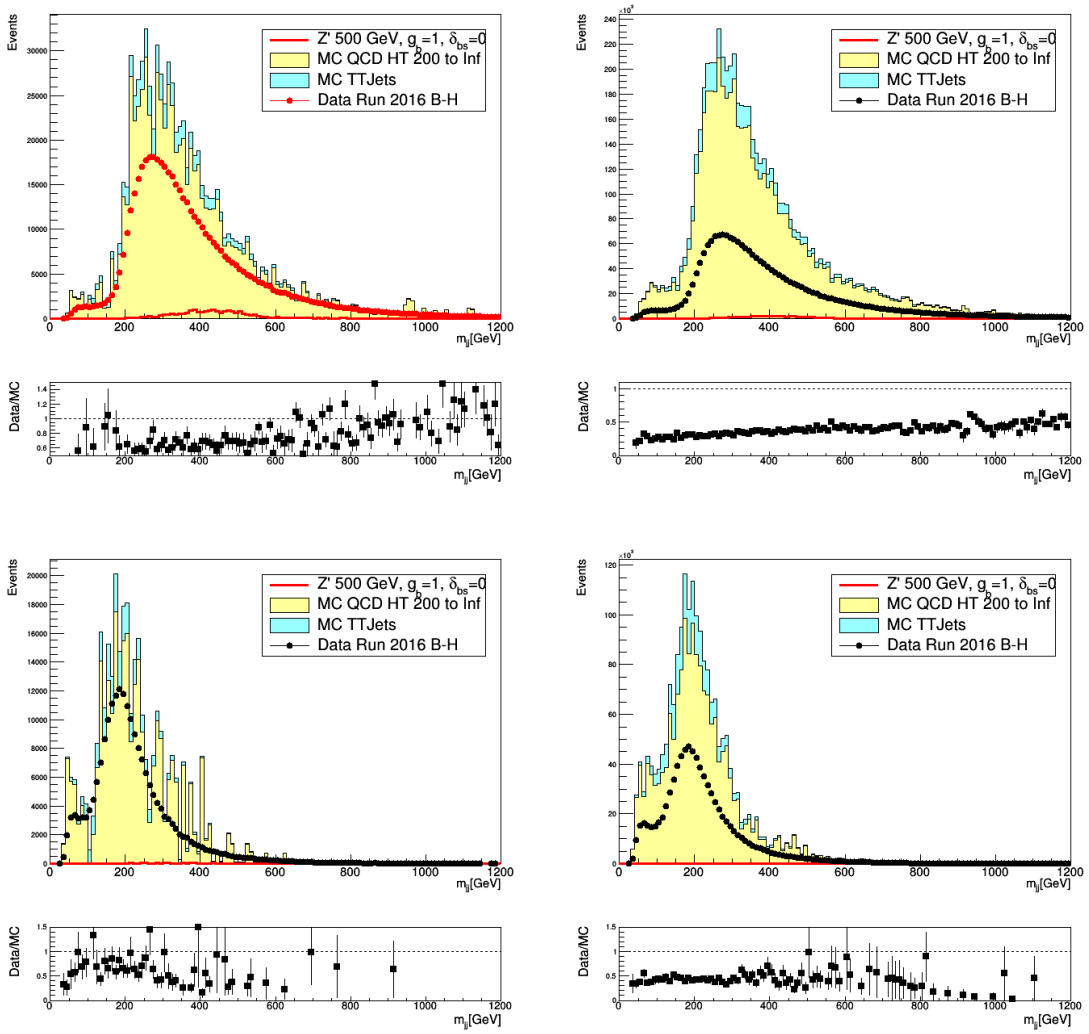

Figure 5.8: Invariant mass distribution in data and MC for the four additional control regions designed to validate the method. Signal MC corresponds to $m_{Z'}$ = 500 GeV with δ_{bs}=0 and g_b=1.

	δ_{bs}	Event Yields	Acceptance (%)
$m_{Z'}$ = 350 GeV	0.0	79315.6 $\pm$ 27.4	0.95 $\pm$ 0.003
$m_{Z'}$ = 500 GeV	0.0	16493.7 $\pm$ 13.9	1.18 $\pm$ 0.009
	0.25	17035.8 $\pm$ 12.5	0.92 $\pm$ 0.007
	0.5	18649.6 $\pm$ 10.2	0.56 $\pm$ 0.004
	0.75	19174.9 $\pm$ 8.3	0.36 $\pm$ 0.002
	1.0	12264.1 $\pm$ 4.9	0.19 $\pm$ 0.001
$m_{Z'}$ = 650 GeV	0.0	4792.4 $\pm$ 7.4	1.15 $\pm$ 0.016
$m_{Z'}$ = 800 GeV	0.0	1924.8 $\pm$ 4.8	1.22 $\pm$ 0.027
$m_{Z'}$ = 950 GeV	0.0	790.9 $\pm$ .02	1.16 $\pm$ 0.041

Table 5.10: Event yields and acceptances after event selection, object corrections, and event weights for simulated signals.

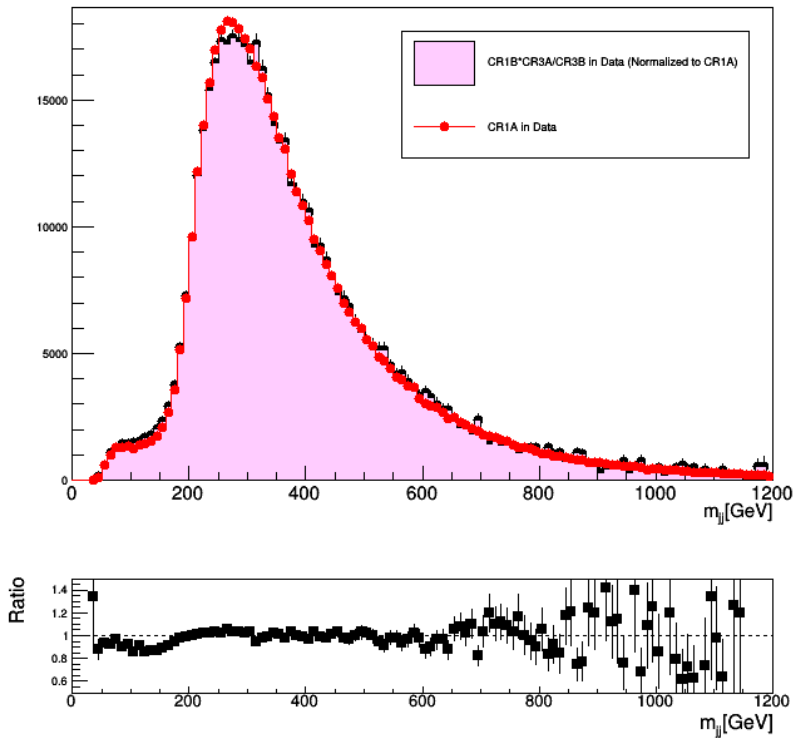

Figure 5.9: Invariant mass distribution for CR1A and the product of CR1B*CR3A/CR3B. This crosscheck allows us to conclude that we can extrapolate the background contribution in a given region by using the shape and yields from the other 3.

systematic uncertainties that may change the expected yields (rate changes), the shape of the discriminant variable, or both.

Table 5.11 summarizes all of the systematic uncertainties considered for this analysis, with one systematic per line. Each source of uncertainty is assigned either a shape or "log-normal" (lnN) type depending on how a deviation from nominal affects the m_{jj} distribution.

5.8.1 PDF

To estimate the uncertainty associated with the PDF prediction used in the generation of the samples, we follow the recommendations from the PDF4LHC group[103]. To compute the PDF uncertainty, one has to evaluate the cross-section σ N_{mem} +1 times, where N_{mem} is the number of error sets (symmetric eigenvectors) of the PDF set,

$$\sigma^{(k)}, k = 0, ..., N_{mem}, \tag{5.24}$$

In particular, N_{mem} = 100 in $PDF4LHC15\_100$. Then, the PDF uncertainty is given by:

$$\delta^{pdf}\sigma = \sqrt{\sum_{k=1}^{N_{mem}} (\sigma^{(k)} - \sigma^{(0)})^2} \tag{5.25}$$

The uncertainty is to be understood as a 68% confidence level. Figure 5.10 shows the nominal m_{jj} distribution for a Z' with m = 500 GeV, δ_{bs} = 0, and g_b = 1 in the signal region, as well as the distribution after PDF weights. Figure 5.11 shows the acceptance (in percentage) in the signal region for the generated signal as a function of δ_{bs} (left) and $m_{Z'}$(right).

Source	**Type**	**Rate Uncertainty [%]**	**Notes**
PDF choice	lnN	1.8-5.9	PDF uncertainty for BFF initiated processes
α_s	lnN	3.0-4.2	α_s uncertainty for BFF initiated processes
Luminosity 13 TeV	lnN	2.5	Signal
Jet Energy Scale	shape	0-0.23	Signal
Jet Energy Resolution	shape	0-0.57	Signal
Pileup Weight	lnN	6.5-9.3	Signal
DeepCSV Weight	lnN	5.6-8.5	Signal
Trigger Weight	lnN	3.2-4.9	Signal
background normalization	shape	5.0	Scale uncertainty for data-driven background prediction

Table 5.11: Summary of the systematic uncertainties used in this analysis.

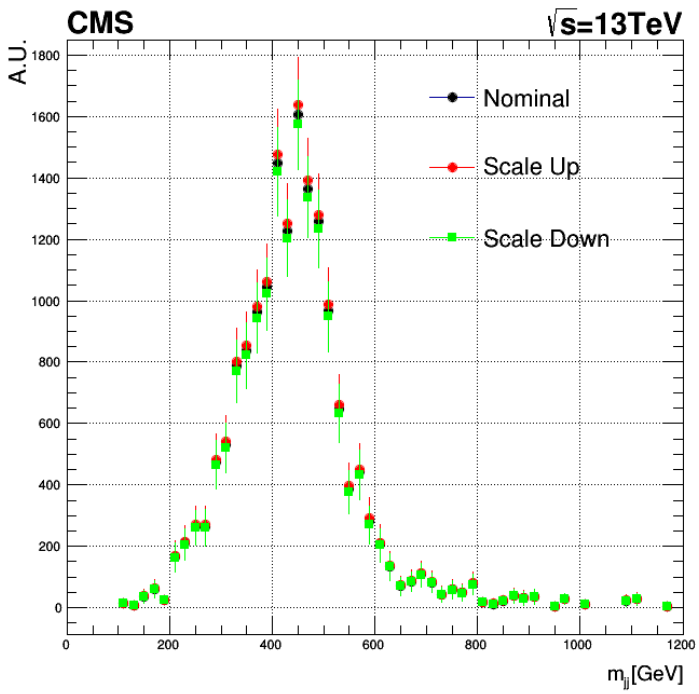

Figure 5.10: Nominal invariant mass distribution for a Z' with m = 500 GeV, δ_{bs} = 0 and g_b = in the signal region (black); after PDF weight scaled up (red)/ down(green).

.

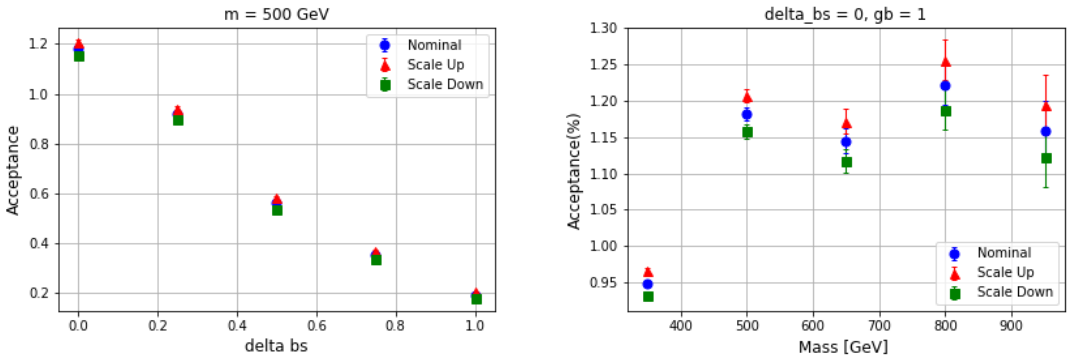

Figure 5.11: Nominal acceptance (blue), acceptance after PDF systematics scaled up (red) / down (green) in signal region as a function of δ_{bs} (left) and $m_{Z'}$ (right).

The effect of the PDF uncertainties strongly depends on the Q^2 that governs the scale of the

hard interaction. The higher the mass Q^2, the higher the x of the partons involved in the interaction, and so the uncertainty.

5.8.2 ISR/FSR

The PDF4LHC15 sets are based on the following value of the strong coupling constant $\alpha_s(m_Z^2)$ and of its associated uncertainty,

$$\alpha_s(m_Z^2 = 0.1180 \pm 0.0015), \tag{5.26}$$

at the 68% confidence level. The α_s uncertainty can be computed as

$$\delta^{\alpha_s}\sigma = \frac{\sigma(\alpha_s = 0.1195) - \sigma(\alpha_s = 0.1165)}{2} \tag{5.27}$$

corresponding to an uncertainty $\delta\alpha_s$ = 0.0015 at the 68% confidence level. Figure 5.12 shows the nominal m_{jj} distribution for a Z' with m = 500 GeV, δ_{bs} = 0, and g_b = 1 in the signal region, as well as the distribution after α_s weights. Figure 5.13 shows the acceptance (in percentage) in the signal region for the generated signal as a function of δ_{bs} (left) and $m_{Z'}$(right).

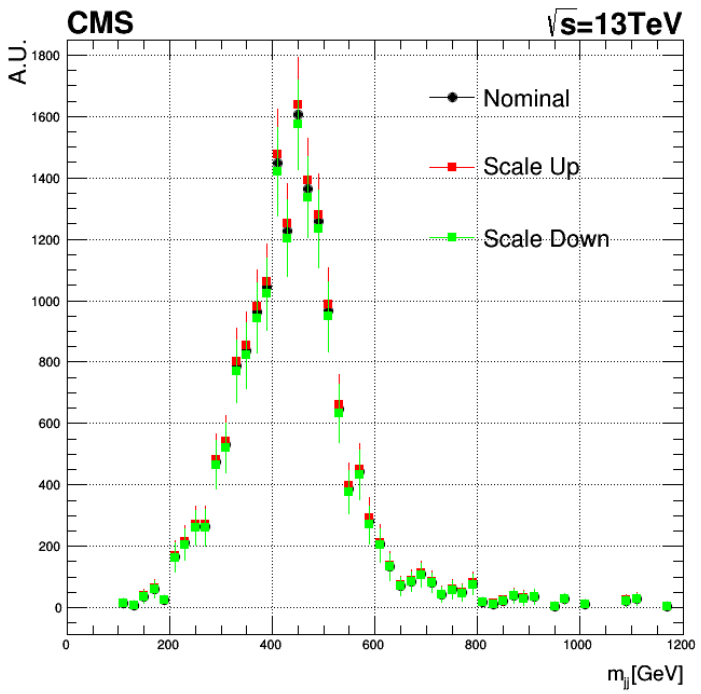

Figure 5.12: Nominal invariant mass distribution for a Z' with m = 500 GeV, δ_{bs} = 0 and g_b = in the signal region (black); after α_s weight scaled up (red)/ down(green).

.

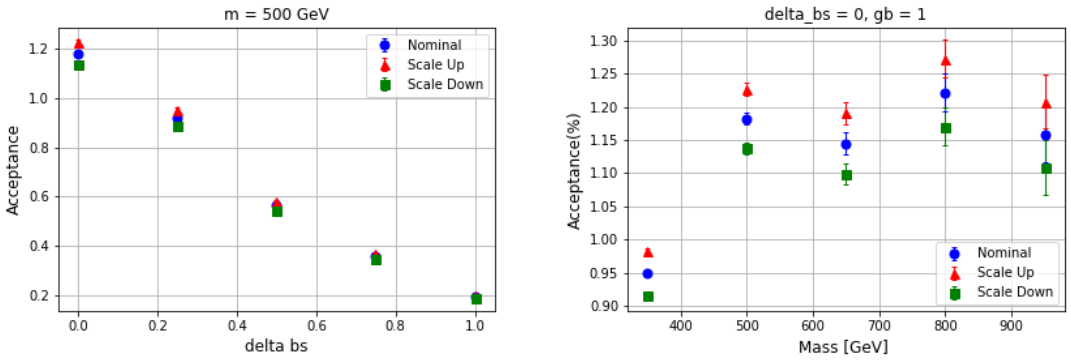

Figure 5.13: Nominal acceptance (blue), acceptance after α_s-related systematics scaled up (red) / down (green) in signal region as a function of δ_{bs} (left) and $m_{Z'}$ (right).

5.8.3 LHC Luminosity

A flat rate uncertainty of 2.5% is applied to all of the simulated samples accounting for the uncertainty on the LHC luminosity and thus, the simulation normalizations[110].

5.8.4 Pileup Weights

The necessity of the pileup weights was discussed in Section 5.3.3. The number of pileup interactions in a single bunch crossing is given by

$$N_i = \frac{\mathcal{L} \cdot \sigma_{minimumbias}}{v_{orbit}} \tag{5.28}$$

where $\mathcal{L}$ is the instantaneous luminosity, $\sigma_{minimumbias}$ is the total minimum bias cross-section for an event at the LHC, and v_{orbit} is the LHC orbit frequency (11246 Hz). In this calculation, the minimum bias cross-section is used, but its actual value is unknown.

To asses the effect of a systematic uncertainty due to the choice of $\sigma_{minimumbias}$ = 69.3 mb, a $\pm 4.6\%$ variation was used, and the pileup weights were recalculated.

As it turns out, the shape changes were negligible, but the rate changes due to this shift can be seen in Figure 5.14.

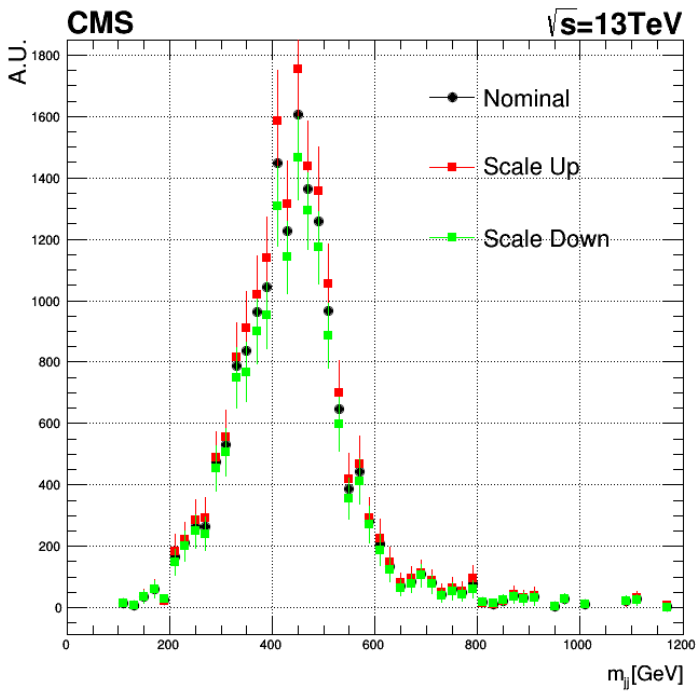

Figure 5.14: Nominal invariant mass distribution for a Z' with m = 500 GeV, δ_{bs} = 0 and g_b = in the signal region (black); after PU-related systematics scaled up (red)/ down(green).

Figure 5.15 shows the acceptance (in percentage) in the signal region for the generated signal as a function of δ_{bs} (left) and $m_{Z'}$(right). Here, we see that the impact of the PU weight variation is greater for smaller values of δ_{bs}. This behavior is expected, as pileup will directly affect our b-tagging efficiency and preselection.

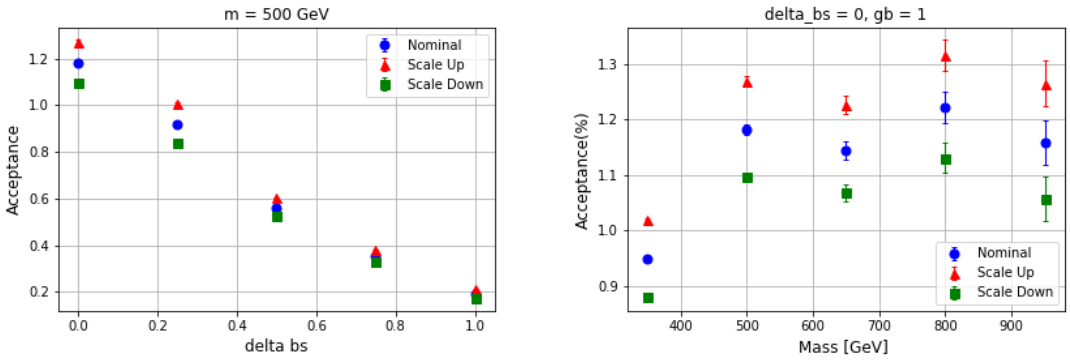

Figure 5.15: Nominal acceptance (blue), acceptance after PU-related systematics scaled up (red) / down (green) in signal region as a function of δ_{bs} (left) and $m_{Z'}$ (right).

5.8.5 Jet energy scale

The jet energy corrections used to correct the energy scale back to the particle level were discussed in Section 4.4. The uncertainty on this correction originates from several uncorrelated sources, but for simplicity, we use the total combined uncertainty. For *M* uncorrelated sources, the total uncertainty $S(p_T, \eta)$ is given by:

$$S(p_T, \eta) = \sqrt{\sum_i^M s_i^2(p_T, \eta)} \tag{5.29}$$

where $s_i(p_T, \eta)$ is the uncertainty for a single source i. The JES uncertainty varies as a function of p_T and η and is $\leq 4\%$ in all regions of phase space[111].

To evaluate the effect this uncertainty has on the m_{jj} distribution, the jet energies used to reconstruct the mass were shifted by $\pm 1\sigma$ using the procedures given in [112] and [113]. Figure 5.16 shows the type of variations expected for the signal sample.

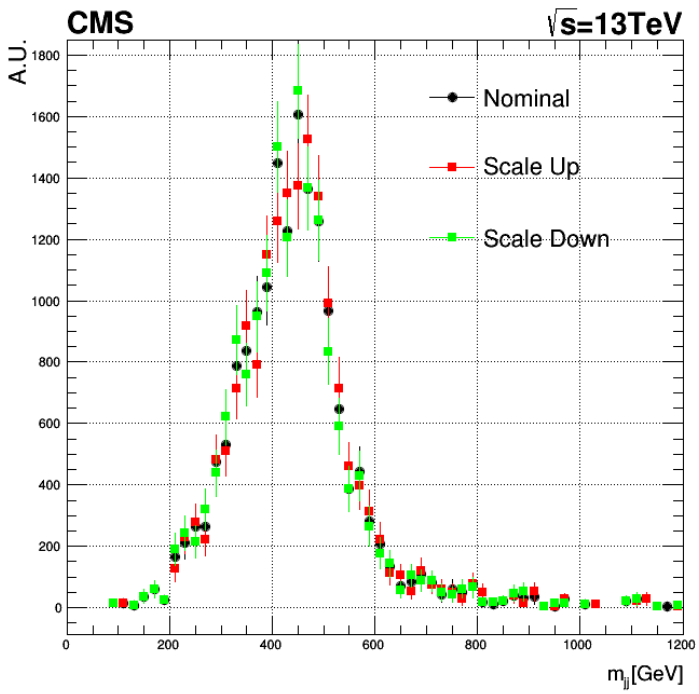

Figure 5.16: Nominal invariant mass distribution for a Z' with m = 500 GeV, δ_{bs} = 0 and g_b = in the signal region (black); after JES-related systematics scaled up (red)/ down(green).

The variation in yields due to JES shifts was negligible for all mass points and δ_{bs} considered on the analysis. The JES systematic uncertainties only affect the shape of the m_{jj} distribution.

5.8.6 Jet energy resolution

The jet energy resolution is smeared, by default, according to CMS standard jet energy corrections. To estimate the systematic effects related to such smearing, an additional $\pm 1\sigma$ on the uncertainty of the JES is applied[114]. Then, the m_{jj} is recalculated.

Figure 5.17 shows the nominal m_{jj} distribution for a Z' with m = 500 GeV, δ_{bs} = 0 and g_b = 1 in the signal region, as well as the distribution after additional JER smearing. Figure 5.18 shows the acceptance (in percentage) in the signal region for generated signal as a function of δ_{bs} (left) and $m_{Z'}$(right).

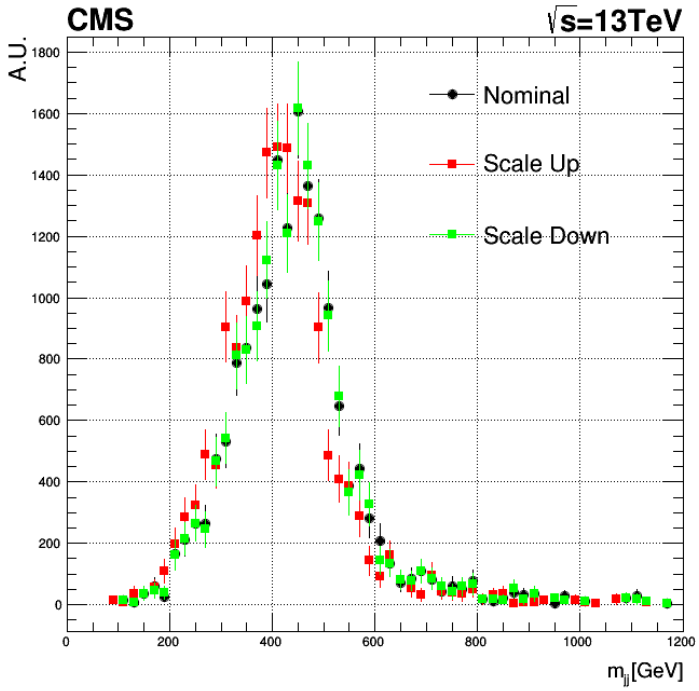

Figure 5.17: Nominal invariant mass distribution for a Z' with m = 500 GeV, δ_{bs} = 0 and g_b = 1 in the signal region (black) after $+1\sigma$(red)/ -1σ(green) JER smearing.

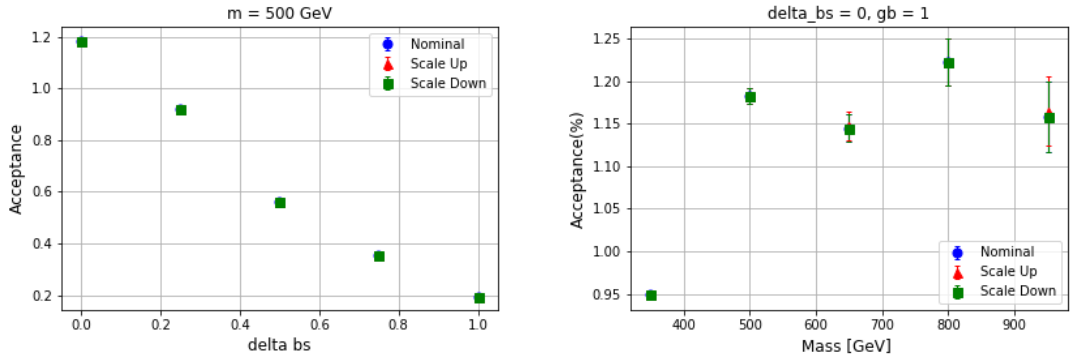

Figure 5.18: Nominal acceptance (blue), acceptance after $+1\sigma$(red)/ -1σ(green) JER smearing in signal region as a function of δ_{bs} (left) and $m_{Z'}$ (right).

Just like for the case of JES uncertainties, the additional JER smearing does not seem to have a direct impact on the acceptance, but it does modify the shape of the m_{jj} distribution.

5.8.7 Trigger Weights

As described in Section 5.4, the efficiency of the trigger HLT paths used for the analysis is measured as a function of the event observables such as the p_T of the calo and PF jets, as well as the CSV value for the third leading jet in CSV in the event. Fits on the measured turn-on components in Figure 5.3 (b) are then used as a weight. The uncertainty of the trigger fits is propagated to the simulated signal samples.

Figure 5.19 shows the nominal m_{jj} distribution for a Z' with m = 500 GeV, δ_{bs} = 0 and g_b = 1 in the signal region, as well as the distribution after trigger efficiency uncertainties, are scaled up/down. Figure 5.20 shows the acceptance (in percentage) in the signal region for the generated signal as a function of δ_{bs} (left) and $m_{Z'}$(right).

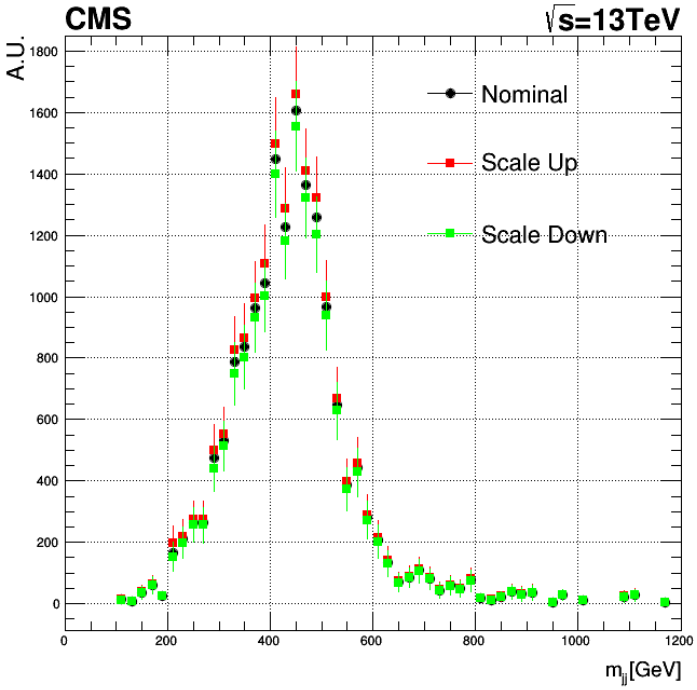

Figure 5.19: Nominal invariant mass distribution for a Z' with m = 500 GeV, δ_{bs} = 0 and g_b = in the signal region (black); after trigger efficiency statistical uncertainties are scaled up(red)/ down(green).

.

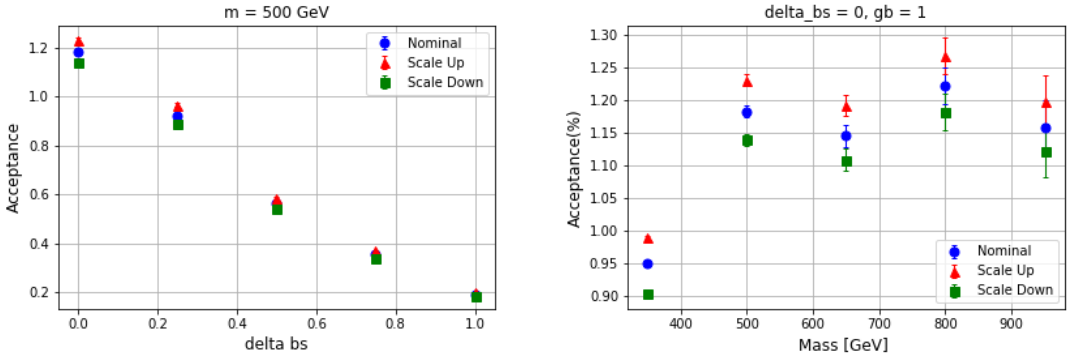

Figure 5.20: Nominal acceptance (blue), acceptance after trigger efficiency statistical uncertainties are scaled up(red)/ down(green) in signal region as a function of δ_{bs} (left) and $m_{Z'}$ (right).

5.8.8 Background estimation

As described in Section 5.6, background estimation in the signal region is estimated from the signal-depleted control regions. In order to estimate the uncertainties derived from the extrapolation, we propagate the errors on a bin-by-bin basis. In this case, the uncertainty for the prediction of A is

$$\sigma_S R = SR\sqrt{(\frac{\sigma_{CR1}}{CR1})^2 + (\frac{\sigma_{CR2}}{CR2})^2 + (\frac{\sigma_{CR3}}{CR3})^2} \tag{5.30}$$

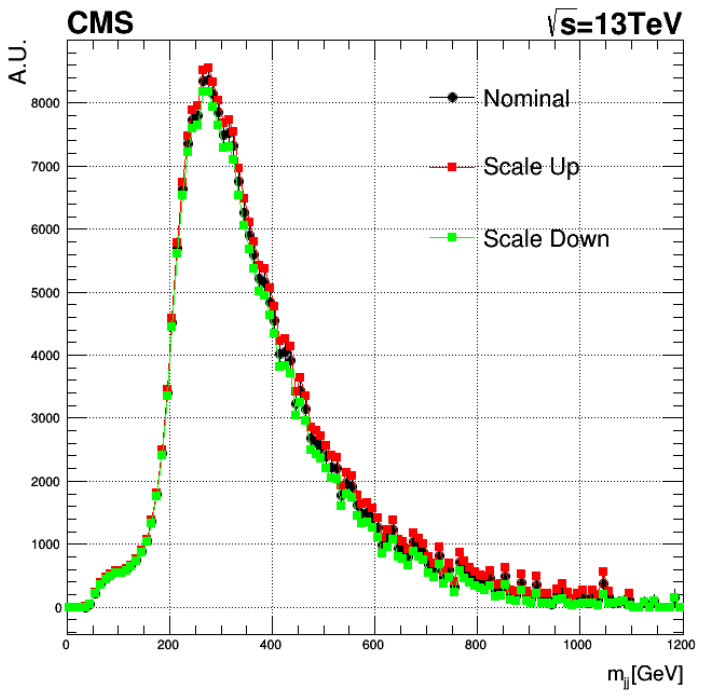

Figure 5.21: Nominal invariant mass distribution for background estimation from data (black) and after propagation of uncertainties up(red) and down (green).

6. RESULTS

What follows is an overview of the statistical tool used to set an upper limit to the potential Z' production cross-section. The results include upper limits in $Z' \rightarrow b\bar{b}/b\bar{s}$ production at $\sqrt{s}$ = 13 TeV as a function of the Z' mass, δ_{bs}, and g_b.

Such quantities can be interpreted directly from the Lagrangian of the new physics model proposed in this analysis and therefore constrain the potential BSM theories that describe data best, even if no evidence for new physics is found in the analysis.

6.1 Statistical model

A statistical evaluation is performed on the observed and expected bin-by-bin counts in the m_{jj} distribution when comparing to the background prediction with and without the presence of several test Z' signal hypotheses. These hypothetical signals differing by their mass and δ_{bs} coupling value (Figure 6.1).

The Combine Higgs Tool[115] is used for this purpose. Its output is called the signal strength, μ, and is defined as $\mu = \sigma/\sigma_{Z'}$. Here, σ is the excludable cross-section, and $\sigma_{Z'}$ is the input cross-section provided by the signal model. Combine is a CMS-maintained software package for statistical analysis based on RooStats. The model was provided to this tool using plain text datacard files for each signal mass and the m_{jj} shape via histograms produced with ROOT. With these definitions, the number of expected events in a given bin is $\lambda_i = \mu s_i + b_i$, where s_i and b_i are the number of expected signal and background events, respectively.

The statistical model is described by the likelihood, $L(n|\lambda(\mu))$, of observing n events given the model $\lambda(\mu)$, and it is built by the product of the Poisson probability evaluated at each of the bins in the m_{jj} distribution. To complete this model, we include the sources of systematic uncertainties that might affect the number of expected events. Each source is defined by a nuisance parameter, θ_u so that

$$\lambda_i(\mu, \theta_1...\theta_{N_u}) = \mu s_i(\theta_1...\theta_{N_u}) + \sum_{b=1}^{N_b} \lambda_{i,b}(\theta_1...\theta_{N_u}) \tag{6.1}$$

This way, the proper statistical model corresponds to the following binned likelihood

$$L(n|\lambda(\mu, \theta)) = \prod_{i=1}^{N_{bin}} \frac{\lambda_i(\mu, \theta)}{n_i!} \dot{e}^{-\lambda_i(\mu,\theta)} \tag{6.2}$$

6.2 The upper limit

In the statistical model language, the desired probability, $P(SM+Z'|Data)$, corresponds to the posterior probability $P(\lambda(\mu)|n)$. The latter is computed using Bayes theorem from the likelihood of the statistical model in Equation 6.2, and integrating over all the nuisance parameters and with the proper normalization

$$P(\lambda(\mu)|n) = \frac{\int d\theta' L(n|\lambda_\mu, \theta')\pi(\mu, \theta')}{\int d\mu' d\theta' L(n|\lambda_{\mu'}, \theta')\pi(\mu', \theta')} \tag{6.3}$$

Where π is called the prior probability density function. With this choice, all positive values for the signal cross-section are assumed to be equally likely. Integrating the posterior probability from zero to μ

$$p(\mu) = \int_0^\mu P(n|\lambda_{\mu'}) d\mu' \tag{6.4}$$

This expression provides a way to estimate how likely it is for data to be consistent with the SM + Z' hypothesis for a given Z' mass, and signal strength μ, given a particular measurement. This analysis uses a 95% confidence level (CL) definition to set an exclusion on μ. This means that if $p(\mu)$ < 0.05, the particular choice of signal strength for a given Z' hypothesis is disfavored by the observations and is therefore excluded. The value of μ for which $p(\mu)$ = 0.05 is called the excluded signal strength, μ_{exc}, at 95% CL. Consequently, signal strengths for which $p(\mu)$>0.05 are not excluded by the measurement. It is important to note that, if for some mass hypothesis, $p(\mu)$ >0.05, it does not mean that the SM+Z' hypothesis is correct. The latter only means that the

considered hypothesis is not consistent with the measurement, probably because μ is too weak for the sensitivity of the analysis.

For every Z' mass-δ_{bs} point considered, two different μ_{exc} are computed, the so-called expected and observed μ_{exc} at 95% CL. If the observed μ_{exc} at 95% CL is less than unity, that region in phase-space is said to be excluded in the model. The relationship between the excluded signal strength and cross-section can be obtained from the definition of signal strength, $\sigma_{exc} = \mu_{exc}\sigma_{Z'}$.

6.3 Constraints to the Z' production at $\sqrt{s}$ = 13 TeV

The m_{jj} distributions showing the background predictions and the observed data are shown in Figure 6.1. We report an almost perfect agreement between the SR and background estimation in data, at least in the region of interest. The m_{jj} distribution for five Z' mass points is also shown in solid lines for comparison. The low acceptance of the simulated signal events leads to a non-uniform behavior of the m_{jj} distributions. These shapes are used as input to the statistical analysis.

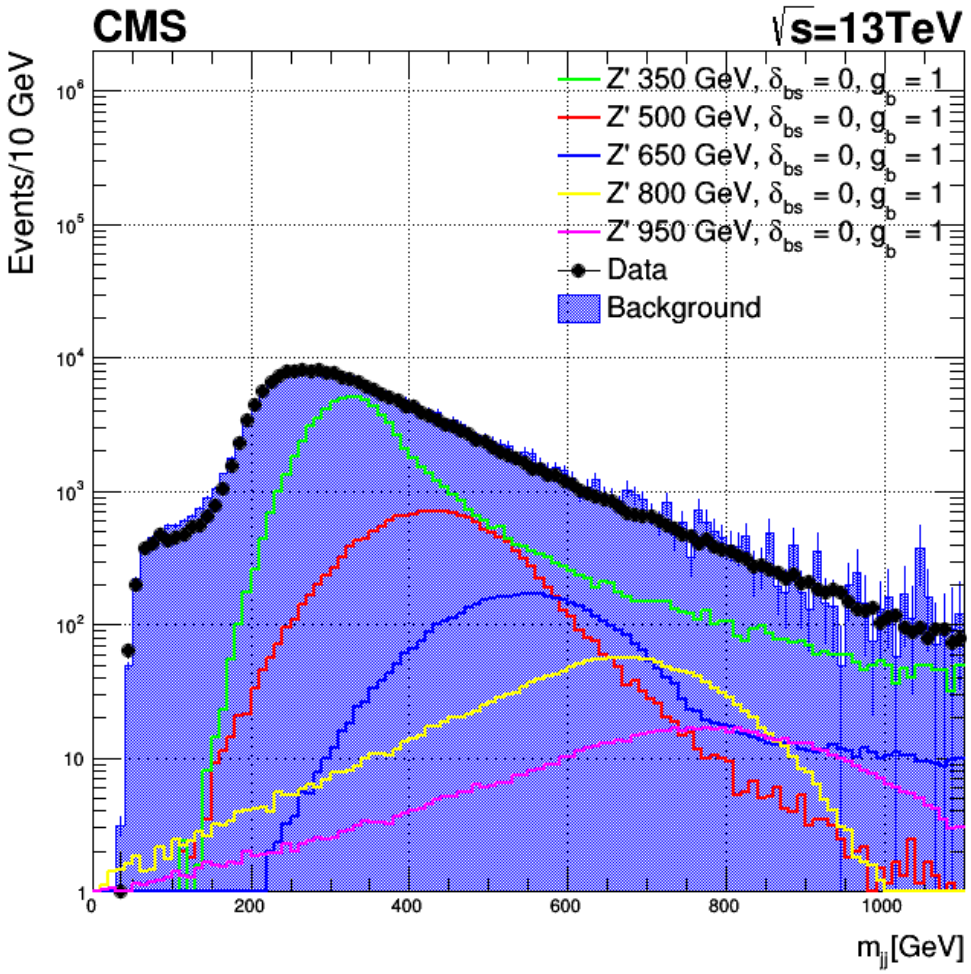

Figure 6.1: The m_{jj} distribution in signal MC (colored solid lines) and data (black markers). Also shown here, is the background prediction from data in the signal region (filled purple histogram).

Figure 6.2 shows the observed and expected 95% CL cross-section, $\sigma_{Z'} \times B(Z' \to b\bar{b})$ at the 95% CL. The green and yellow bands correspond to one and two standard deviations from the 95% CL expected cross-section. The theoretical value is also plotted for different g_b values. We can exclude the theoretical model where its expected cross-section exceeds the 95% CL observed cross-section.

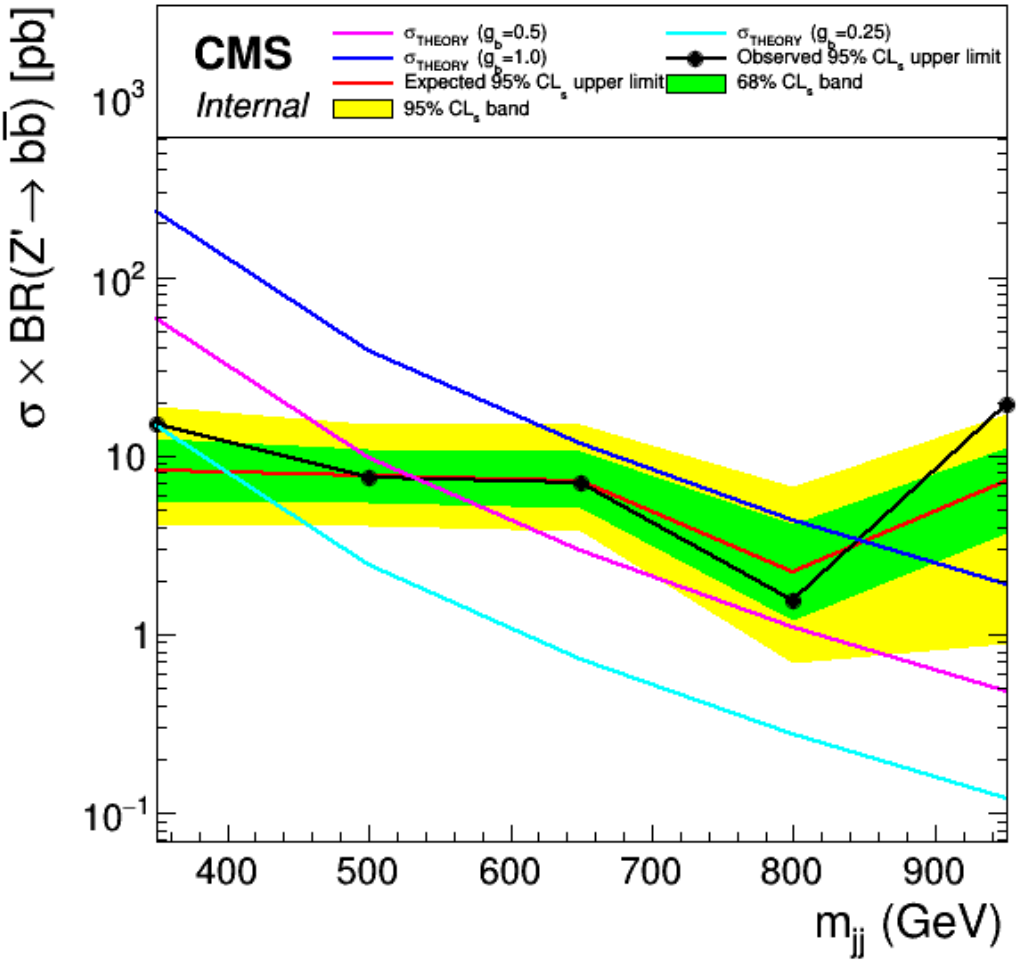

Figure 6.2: Expected (red line) and observed (black line) 95% CL limits to $\sigma_{Z'} \times B(Z' \to b\bar{b})$ for Z' bosons at $\sqrt{s}$ = 13 TeV, as a function of Z' mass for δ_{bs} = 0. The shaded green and yellow bands represent the one and two sigma uncertainty bands. Theoretical values also displayed for g_b = 0.25, 0.5, and 1.0

The Z' production can be excluded for almost every mass point considered in the analysis, except for the case of 950 GeV, when g_b = 1.0. For g_b = 0.5, only the 350, and 500 GeV mass points can be excluded, while none can be excluded when g_b = 0.25.

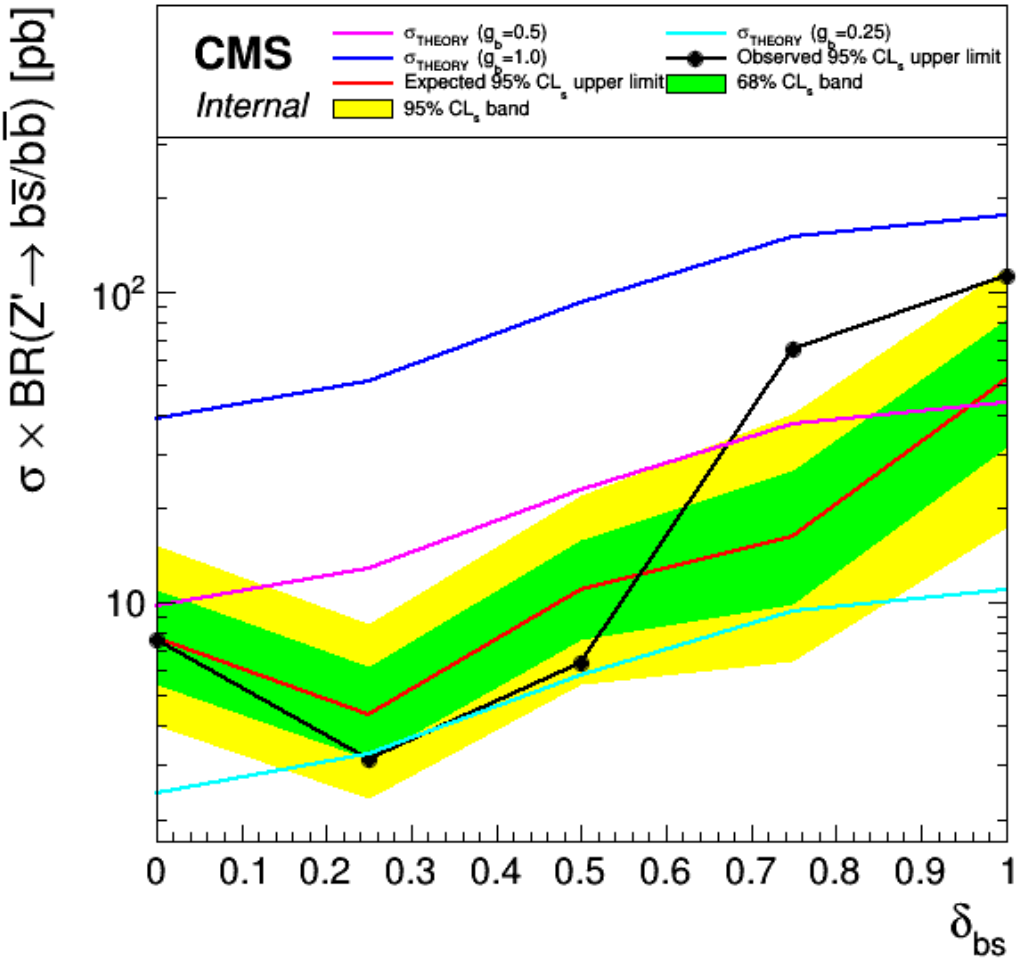

Figure 6.3: Expected (red line) and observed (black line) 95% CL limits to $\sigma_{Z'} \times B(Z' \to b\bar{b}/b\bar{s})$ for 500 GeV Z' bosons at $\sqrt{s}$ = 13 TeV, as a function of δ_{bs}. The shaded green and yellow bands represent the one and two sigma uncertainty bands. Theoretical values also displayed for g_b = 0.25, 0.5, and 1.0

Figure 6.3 shows the observed and expected cross-section, $\sigma_{Z'} \times B(Z' \to b\bar{b}/b\bar{s})$ at the 95% CL, as a function of the coupling parameter δ_{bs} for $m_{Z'}$ = 500 GeV. When δ_{bs} increases, the expected yield of events in the signal region rises marginally due to a small but existing acceptance for diagrams including s-quarks, but the production cross-section rises disproportionately, degrading the sensitivity. As an example, the ratio of expected events at δ_{bs} = 0.75 to δ_{bs} = 0 is 1.3 while the cross-sections for the same differ by a factor of 4.3. The observed better sensitivity for low but

non-zero δ_{bs} is due to variation in fit shape quality. With this in mind, the region in phase space we can exclude includes all δ_{bs} points for g_b = 1.0. When g_b = 0.5, all points below δ_{bs} < 0.75 can be excluded. Finally, for g_b = 0.25, only the δ_{bs} = 0.25 point can be excluded at 95$\%$ CL.

7. CONCLUSIONS

This dissertation presented a search for a new heavy gauge boson decaying to a pair of b quarks in a 4 b-jet final state in pp collisions at $\sqrt{s}$ = 13 TeV in 35.9 fb^{-1} of data collected by the CMS experiment during the 2016 run of the LHC. To do so, a novel analysis technique was employed using an optimized b-tagging selection and two HLT paths triggering on events with four jets, three of them b-tagged. Furthermore, this analysis constitutes the first purely hadronic search for a BFF Z', as previous efforts used leptonic final states.

No significant excess is found, and therefore, a 95% CL_s upper limit on Z' production cross-section is set. The 350 GeV mass point can almost be excluded for g_b = 0.25, we further exclude g_b = 0.5 for masses up to 520 GeV, and g_b = 1 up to 840 GeV. We also explore the δ_{bs}-dependence of the 500 GeV mass point limit. In that case, we exclude g_b = 0.5 for δ_{bs} values below 0.65 and $g_b = 1$ up to $\delta_{bs} = 1$.

With the amount of data analyzed, we can exclude the Z' model in some regions of the parameter space based in δ_{bs} and g_b. The Z' as an explanation of B-anomalies cannot be ruled out for the other regions of phase space.

REFERENCES

[1] "The standard model of particle physics." https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg, 2016. Online; accessed June 7, 2019.

[2] R. Aaij, C. Abellán Beteta, B. Adeva, M. Adinolfi, C. Aidala, Z. Ajaltouni, S. Akar, P. Albicocco, J. Albrecht, F. Alessio, and et al., "Search for lepton-universality violation in $b^+ \to k^+l^+l$ decays," *Physical Review Letters*, vol. 122, May 2019.

[3] A. J. Buras and J. Girrbach, "Left-handed Z' and Z FCNC quark couplings facing new $b \to s\mu^+\mu^-$ data," *J. High Energ. Phys.*, vol. 9, 2013.

[4] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, "Parton distributions for the LHC," *Eur. Phys. J.*, vol. C63, pp. 189–285, 2009.

[5] K. Andersen and T. Eberle, "The genesis 2.0 project." http://media.vanityfair.com/photos/54cbf6ad1ca1cf0a23ac6c85/master/w_690,c_limit/image.jpg, 2010. Online; accessed May 29, 2019.

[6] T. Sakuma, "A cutaway diagram of the cms detector." https://en.wikipedia.org/wiki/Compact_Muon_Solenoid, 2013. Online; accessed May 29, 2019.

[7] "Diagram for the CMS detector coordinate system." https://hypatia.iasa.gr/applet/en/help.html, 2011.

[8] C. Collaboration, C. Warsaw, K. Pozniak, R. Romaniuk, and W. Zabolotny, "Performance of cms silicon tracker," *Journal of Instrumentation*, vol. 5, p. 2010, 03 2010.

[9] A. Benaglia, "The CMS ECAL performance with examples," *Journal of Instrumentation*, vol. 9, pp. C02008–C02008, feb 2014.

[10] G. Dolinska, *Measurement of the Cross Sections of $t\bar{t}$ Production in the Boosted Regime with the CMS Detector at $\sqrt{s} = 8$ TeV*. PhD thesis, Hamburg U., 2016.

[11] F. Bechtel and P. Schleper, "The underlying event in proton-proton collisions," 04 2020.

[12] "Reconstructed Event at the LHC." https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.lhc_p_collisions, 2011.

[13] J. Brooke, D. Cussans, R. Frazier, G. Heath, D. Machin, D. Newbold, S. Galagadera, S. Madani, and A. Shah, "Hardware and firmware for the cms global calorimeter trigger," 07 2019.

[14] "Detector data flow through Hardware tiers." https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookComputingModel, July 2018.

[15] "The CMS Detector and the Token Bit Manager." https://www.phys.ksu.edu/reu2014/wabehn/, Jan 2017.

[16] "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET," Tech. Rep. CMS-PAS-PFT-09-001, CERN, Geneva, Apr 2009.

[17] Q. D. Blog, "Jets at cms." https://www.quantumdiaries.org/2011/06/01/anatomy-of-a-jet-in-cms/, 2010. [Online; accessed June 27, 2019.

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

[19] Wikipedia contributors, "B-tagging — Wikipedia, the free encyclopedia," 2019. [Online; accessed 19-June-2019].

[20] T. C. Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV," *Journal of Instrumentation*, vol. 13, pp. P05011–P05011, may 2018.

[21] "CMS Luminosity Public Results." https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults, 2016.

[22] 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," *Physics Letters B*, vol. 716, no. 1, pp. 1 – 29, 2012.

[23] S. Chatrchyan *et al.*, "Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc," *Physics Letters B*, vol. 716, no. 1, pp. 30 – 61, 2012.

[24] R. Aaij *et al.*, "Observation of $j\psi$ resonances consistent with pentaquark states in $\Lambda_b^0 \to j\psi K^- p$ decays," *Phys. Rev. Lett.*, vol. 115, p. 072001, Aug 2015.

[25] Wikipedia contributors, "Charge (physics) — Wikipedia, the free encyclopedia," 2019. [Online; accessed 7-June-2019].

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

[27] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*. Reading, USA: Addison-Wesley, 1995.

[28] Wikipedia contributors, "Fermionic field — Wikipedia, the free encyclopedia," 2018. [Online; accessed 7-June-2019].

[29] Wikipedia contributors, "Gamma matrices — Wikipedia, the free encyclopedia," 2019. [Online; accessed 7-June-2019].

[30] G. Rajasekaran, "Fermi and the Theory of Weak Interactions," *Resonance J. Sci. Educ.*, vol. 19, no. 1, pp. 18–44, 2014.

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

[32] S. Weinberg, "A model of leptons," *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, Nov 1967.

[33] A. Salam, "Weak and electromagnetic interactions," in *Elementary particle theory* (N. Svartholm, ed.), pp. 367–377, Almquist & Wiksell.

[34] J. Schwinger, "The theory of quantized fields. i," *Phys. Rev.*, vol. 82, pp. 914–927, Jun 1951.

[35] T. P. Cheng and L. F. Li, *GAUGE THEORY OF ELEMENTARY PARTICLE PHYSICS*. 1984.

[36] B. Martin and G. Shaw, *Particle Physics*. Wiley, 2017.

[37] C. Patrignani *et al.*, "Review of Particle Physics," *Chin. Phys.*, vol. C40, no. 10, p. 100001, 2016.

[38] The LHCb collaboration, "Differential branching fractions and isospin asymmetries of $b \to k^* \mu^+ \mu$ decays," *Journal of High Energy Physics*, vol. 2014, p. 133, Jun 2014.

[39] The LHCb collaboration, "Measurements of the s-wave fraction in $b^0 \to k^+ \pi \mu^+ \mu$ decays and the $b^0 \to k^{*0}(892) \mu^+ \mu$ differential branching fraction," *Journal of High Energy Physics*, vol. 2016, p. 47, Nov 2016.

[40] R. Aaij *et al.*, "Test of lepton universality using $B^+ \to K^+ \ell^+ \ell^-$ decays," *Phys. Rev. Lett.*, vol. 113, p. 151601, Oct 2014.

[41] The LHCb collaboration, "Test of lepton universality with $b_0 \to k^{*0} l^+ l$ decays," *Journal of High Energy Physics*, vol. 2017, p. 55, Aug 2017.

[42] R. Aaij *et al.*, "Search for lepton-universality violation in $B^+ \to K^+ e^+ e^-$ decays," *Phys. Rev. Lett.*, vol. 122, no. 19, p. 191801, 2019.

[43] R. Aaij *et al.*, "Measurement of form-factor-independent observables in the decay $B^0 \to K^{*0} \mu^+ \mu^-$," *Phys. Rev. Lett.*, vol. 111, p. 191801, Nov 2013.

[44] T. L. collaboration, "Angular analysis of the $B^0 \to K^{*0} \mu^+ \mu^-$ decay using 3 fb1 of integrated luminosity," *Journal of High Energy Physics*, vol. 2016, p. 104, Feb 2016.

[45] A. Abdesselam *et al.*, "Angular analysis of $B^0 \to K^*(892)^0 \ell^+ \ell^-$," in *Proceedings, LHCSki 2016 - A First Discussion of 13 TeV Results: Obergurgl, Austria, April 10-15, 2016*, 2016.

[46] W. Altmannshofer, C. Niehoff, P. Stangl, and D. M. Straub, "Status of the $b \to k^* mu^+ mu^-$ anomaly after moriond 2017," *The European Physical Journal C*, vol. 77, p. 377, Jun 2017.

[47] M. Bordone, G. Isidori, and A. Pattori, "On the standard model predictions for r_K and r_{K^*}," *The European Physical Journal C*, vol. 76, p. 440, Aug 2016.

[48] G. D'Amico, M. Nardecchia, P. Panci, F. Sannino, A. Strumia, R. Torre, and A. Urbano, "Flavour anomalies after the rk measurement," *Journal of High Energy Physics*, vol. 2017, p. 10, Sep 2017.

[49] B. Capdevila, A. Crivellin, S. Descotes-Genon, J. Matias, and J. Virto, "Patterns of new physics in $b \to sll$ transitions in the light of recent data," *Journal of High Energy Physics*, vol. 2018, p. 93, Jan 2018.

[50] F. Archilli, M. O. Bettler, P. Owen, and K. A. Petridis, "Flavour-changing neutral currents making and breaking the standard model," *Nature*, vol. 546, pp. 221 EP –, 06 2017.

[51] T. C. Collaboration, "Search for narrow resonances in dilepton mass spectra in proton-proton collisions at $\sqrt{s}$ = 13ătev and combination with 8 tev data," *Physics Letters B*, vol. 768, pp. 57 – 80, 2017.

[52] A. M. Sirunyan *et al.*, "Search for high-mass resonances in dilepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV," *JHEP*, vol. 06, p. 120, 2018.

[53] T. A. Collaboration", "Search for new high-mass resonances in the dilepton final state using proton-proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector," Tech. Rep. ATLAS-CONF-2016-045, CERN, Geneva, Aug 2016.

[54] M. Abdullah, M. Dalchenko, B. Dutta, R. Eusebi, P. Huang, T. Kamon, D. Rathjens, and A. Thompson, "Bottom-quark fusion processes at the LHC for probing Z models and B -meson decay anomalies," *Phys. Rev.*, vol. D97, no. 7, p. 075035, 2018.

[55] W. Altmannshofer, P. Stangl, and D. M. Straub, "Interpreting hints for lepton flavor universality violation," *Phys. Rev. D*, vol. 96, p. 055008, Sep 2017.

[56] W. Altmannshofer, S. Gori, M. Pospelov, and I. Yavin, "Quark flavor transitions in $L_\mu - L_\tau$ models," *Phys. Rev. D*, vol. 89, p. 095033, May 2014.

[57] A. Breskin and R. Voss, *The CERN Large Hadron Collider: Accelerator and Experiments*. Geneva: CERN, 2009.

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

[59] G. Aad *et al.*, "The ATLAS Experiment at the CERN Large Hadron Collider," *Journal of Instrumentation*, vol. 3, no. 08, p. S08003, 2008.

[60] K. Aamodt *et al.*, "The ALICE experiment at the CERN LHC," *JINST*, vol. 3, p. S08002, 2008.

[61] A. A. Alves, Jr. *et al.*, "The LHCb Detector at the LHC," *JINST*, vol. 3, p. S08005, 2008.

[62] G. Acquistapace *et al.*, "CMS, the magnet project: Technical design report," 1997.

[63] C. collaboration, "The cms tracker: addendum to the technical design report," *CERN/LHCC*, vol. 16, 01 2000.

[64] T. C. Collaboration, "Particle-flow reconstruction and global event description with the CMS detector," *Journal of Instrumentation*, vol. 12, pp. P10003–P10003, oct 2017.

[65] *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical Design Report CMS, Geneva: CERN, 1997.

[66] Q. Ingram, "Energy resolution of the barrel of the CMS electromagnetic calorimeter," *Journal of Instrumentation*, vol. 2, pp. P04004–P04004, apr 2007.

[67] "CMS: The hadron calorimeter technical design report," 1997.

[68] E. Yazgan and the CMS ECAL/HCAL Collaborations, "The CMS barrel calorimeter response to particle beams from 2 to 350 GeV/c," *Journal of Physics: Conference Series*, vol. 160, p. 012056, apr 2009.

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

[70] T. C. Collaboration, "The CMS trigger system," *Journal of Instrumentation*, vol. 12, pp. P01020–P01020, jan 2017.

[71] "Brazos Computational Resource." http://brazos.tamu.edu/acknowledge.html, 2019.

[72] T. A. Collaboration, "Performance of the aleph detector at lep," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 360, no. 3, pp. 481 – 506, 1995.

[73] S. Chatrchyan *et al.*, "Description and performance of track and primary-vertex reconstruction with the CMS tracker," *JINST*, vol. 9, p. P10009, 2014.

[74] P. Billoir, "Progressive track recognition with a Kalman like fitting procedure," *Comput. Phys. Commun.*, vol. 57, pp. 390–394, 1989.

[75] P. Billoir and S. Qian, "Simultaneous pattern recognition and track fitting by the kalman filtering method," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 294, no. 1, pp. 219 – 228, 1990.

[76] R. Mankel, "A Concurrent track evolution algorithm for pattern recognition in the HERA-B main tracking system," *Nucl. Instrum. Meth.*, vol. A395, pp. 169–184, 1997.

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

[78] "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET," 2009.

[79] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Commun. ACM*, vol. 18, pp. 509–517, 1975.

[80] D. Guest, J. Collado, P. Baldi, S.-C. Hsu, G. Urban, and D. Whiteson, "Jet flavor classification in high-energy physics with deep neural networks," *Phys. Rev. D*, vol. 94, p. 112002, Dec 2016.

[81] "Jet identification." https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetID#Recommendations_for_13_TeV_2016, 2018.

[82] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual," *Eur. Phys. J. C*, vol. 72, p. 1896, 2012.

[83] The CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in cms," *Journal of Instrumentation*, 2011. http://iopscience.iop.org/1748-0221/6/11/P11002/.

[84] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas," *Physics Letters B*, 2008.

[85] "Jet Energy Resolution." https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetResolution, 2017.

[86] N. Saoulidou, "Particle flow jet identification criteria," CMS Analysis Note CMS-AN-2010-003, CERN, Geneva, Jun 2010.

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

[88] W. Adam, "Track and vertex reconstruction in cms," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 582, no. 3, pp. 781 – 784, 2007. VERTEX 2006.

[89] F. Chollet *et al.*, "Keras." https://keras.io, 2015.

[90] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. Software available from tensorflow.org.

[91] K. Binder, "Applications of monte carlo methods to statistical physics," *Reports on Progress in Physics*, vol. 60, pp. 487–559, may 1997.

[92] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language," *Nucl. Phys.*, vol. B126, pp. 298–318, 1977.

[93] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, "Parton fragmentation and string dynamics," *Physics Reports*, vol. 97, no. 2, pp. 31 – 145, 1983.

[94] J. Alwall, C. Duhr, B. Fuks, O. Mattelaer, D. G. Öztürk, and C.-H. Shen, "Computing decay rates for new physics theories with FeynRules and MadGraph 5," *Comput. Phys. Commun.*, vol. 197, pp. 312–323. 29 p, Feb 2014. Comments: 29 pages, 2 figures.

[95] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An Introduction to PYTHIA 8.2," *Comput. Phys. Commun.*, vol. 191, pp. 159–177, 2015.

[96] S. Agostinelli *et al.*, "Geant4 a simulation toolkit," *Nucl. Instrum. Meth. A*, vol. 506, no. 3, pp. 250 – 303, 2003.

[97] R. Foot, X. G. He, H. Lew, and R. R. Volkas, "Model for a light Z-prime boson," *Phys. Rev.*, vol. D50, pp. 4571–4580, 1994.

[98] S. Chatrchyan *et al.*, "Search for Anomalous $t\bar{t}$ Production in the Highly-Boosted All-Hadronic Final State," *JHEP*, vol. 09, p. 029, 2012. [Erratum: JHEP03,132(2014)].

[99] S. Chatrchyan *et al.*, "Search for Z ' Resonances Decaying to $t\bar{t}$ in Dilepton + Jets Final States in pp Collisions at $\sqrt{s} = 7$ TeV," *Phys. Rev.*, vol. D87, no. 7, p. 072002, 2013.

[100] M. Carena, A. Daleo, B. A. Dobrescu, and T. M. P. Tait, "Z' gauge bosons at the Tevatron," *Phys. Rev.*, vol. D70, p. 093009, 2004.

[101] K. R. Lynch, E. H. Simmons, M. Narain, and S. Mrenna, "Finding Z' bosons coupled preferentially to the third family at LEP and the Tevatron," *Phys. Rev.*, vol. D63, p. 035006, 2001.

[102] M. Abdullah, M. Dalchenko, B. Dutta, R. Eusebi, P. Huang, T. Kamon, D. Rathjens, and A. Thompson, "Bottom-quark fusion processes at the lhc for probing z' models and b-meson decay anomalies," *Phys. Rev. D*, vol. 97, p. 075035, Apr 2018.

[103] Jon Butterworth and Stefano Carrazza and Amanda Cooper-Sarkar and Albert De Roeck and Joël Feltesse and Stefano Forte and Jun Gao and Sasha Glazov and Joey Huston and Zahari Kassabov and Ronan McNulty and Andreas Morsch and Pavel Nadolsky and Voica Radescu and Juan Rojo and Robert Thorne, "PDF4LHC recommendations for LHC Run II," *Journal of Physics G: Nuclear and Particle Physics*, vol. 43, no. 2, p. 023001, 2016.

[104] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An introduction to pythia 8.2," *Computer Physics Communications*, vol. 191, pp. 159 – 177, 2015.

[105] T. C. collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV," *Journal of Instrumentation*, vol. 12, pp. P02014–P02014, feb 2017.

[106] "Pileup Studies." https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupInformation, 2017.

[107] "Utilities for Accessing Pileup Information for Data." https://twiki.cern.ch/twiki/bin/view/CMS/PileupJSONFileforData, 2017.

[108] "Heavy flavour tagging for 13 tev 2016 legacy data." https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation#Recommendation_for_13_TeV_Data, 2019.

[109] "Methods to apply b-tagging efficiency scale factors." https://twiki.cern.ch/twiki/bin/viewauth/CMS/BTagSFMethods, 2020.

[110] "Cms luminosity measurements for the 2016 data taking period," CMS Physics Analysis Summary CMS-PAS-LUM-17-001, CERN, Geneva, 2017.

[111] V. Khachatryan *et al.*, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV," *JINST*, vol. 12, p. P02014. 92 p, Jul 2016. Replaced with the published

version. Added the journal reference and DOI. All the figures and tables can be found at http://cms-results.web.cern.ch/cms-results/public-results/publications/JME-13-004/.

[112] "Jet Energy Corrections: Official Software Tools for applying JEC Corrections and Uncertainties." https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookJetEnergyCorrections, 2017.

[113] "Jet energy scale uncertainty sources." https://twiki.cern.ch/twiki/bin/viewauth/CMS/JEC UncertaintySources, 2017.

[114] "Jet energy scale and resolution in the cms experiment," CMS Physics Analysis Summary Jet Energy Scale and Resolution in the CMS Experiment, CERN, Geneva, 2015.

[115] "Documentation of the RooStats-based statistics tools for Higgs PAG." https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideHiggsAnalysisCombinedLimit, 2017.