Sapienza
Università di Roma

Search for leptoquarks coupling to muons in lepton-quark collisions at LHC

Scuola di Dottorato in Scienze Astronomiche, Chimiche, Fisiche e Matematiche
'Vito Volterra'
Physics (XXXVI cycle)

Mattia Campana
ID number 1654846

Advisors
Prof. Francesco Santanastasio
Prof. Paolo Meridiani

Academic Year February 2024

Search for leptoquarks coupling to muons in lepton-quark collisions at LHC
PhD thesis. Sapienza University of Rome

This thesis has been typeset by LaTeX and the Sapthesis class.

Author's email: mattia.campana0395@gmail.com

Per me si va ne la città dolente,
per me si va ne l'etterno dolore,
per me si va tra la perduta gente.

Contents

Introduction

The Standard Model (SM) is heralded as a monumental achievement in scientific theories. It offers profound insights into the fundamental constituents of matter and the forces governing their interactions, all encapsulated within a quantum field theoretical framework. The fidelity of the Standard Model has been rigorously validated through many experiments, yielding measurements of exquisite precision, especially over the past century. This has established the SM as a paramount framework for understanding the interactions that govern strong and electroweak forces.

Notwithstanding its remarkable accomplishments, the SM does not serve as a comprehensive depiction of the universe. It lacks integration of gravitational interactions and fails to clarify the disproportionate weakness of gravity compared to the electroweak and strong forces. The model faces theoretical challenges, notably its dependency on finely adjusted parameters to reconcile phenomena at vastly different energy scales. Finally, it has not identified a viable particle candidate for dark matter nor elucidates the observed preponderance of matter over antimatter in the universe. It does not account for neutrino flavor oscillations and masses and the reason for an apparent symmetry between quarks and leptons generations within the SM.

Several extensions to the SM have been developed to provide solutions to these open questions. These theoretical models of new physics often predict the existence of new particles that could be observed, for example, in collisions at accelerators, such as the LHC. Leptoquarks (LQs) are particles that emerge naturally in some extensions of the SM. They couple to a lepton and a quark through a coupling λ and can explain the symmetry between quarks and leptons. These particles were first introduced in the Pati-Salam model in the context of Grand Unification Theories (GUTs) over forty years ago and have been studied and searched for in many different experiments, such as *ep* collisions at HERA and $p\bar{p}$ at Tevatron and pp collisions at LHC.

This thesis will exploit a novel LQ production mechanism using LHC as a lepton-quark collider. Due to quantum fluctuations, protons also contain charged leptons, making it possible to study lepton-induced processes. By picking a lepton from one beam and a quark from the other beam, it becomes possible to study the resonant single LQ production ($l + q \rightarrow LQ$). The LQ then decays into a lepton and a quark and the experimental signature is a narrow bump in the lepton-jet invariant mass distribution corresponding to the LQ resonance mass. The search is performed with the Compact Muon Solenoid (CMS) detector at CERN, analyzing data from proton-proton collisions provided by the Large Hadron Collider (LHC), the world's largest and most powerful particle accelerator providing proton-proton collisions at a center of mass energy up to 13.6 TeV. The data used for the analysis were collected during Run 2 (2016-2018), corresponding to an integrated luminosity of 138 fb^{-1}. This thesis studies the final state in which the LQ decays into a muon and a quark, the latter being either light (i.e. u quark) or heavy (b quark), resulting in a final state with a muon, a jet, and in some cases, a second muon, which comes from the initial state of this production mechanism. Final states with top quarks are not considered in this thesis.

Chapter 1 offers a brief review of the theoretical foundation of the SM, presenting the open questions in the SM. LQs are presented as particles that arise naturally from the SM framework and can explain some of its open questions. LHC is described, and elements of the physics of pp interactions at hadron colliders are also introduced. A description of the existing LQ searches and the current lower limits on LQ mass are presented. The last section is dedicated to the novel production mechanism, which is the focus of this thesis, describing also the strategy of the analysis.

Chapter 2 details the experimental apparatus: the CMS detector is described in its various subdetectors.

Chapter 3 describes the LQ sample generation process and the kinematic distributions to understand what the signal looks like. It also presents the cross sections of this process and how they change under model assumptions.

Chapter 4 describes the reconstruction of the physics objects used for this analysis, focusing on jets and muons which are the key ingredients for this search.

Chapter 5 introduces the data and simulation samples and the criteria adopted for the event preselection. This chapter also shows the distributions of the interesting kinematic variables in data, compared to the simulated ones.

Chapter 6 describes the analysis strategy. The first part is dedicated to the details of the signal selection optimization performed with a Boosted Decision Tree. The second part presents the mass spectrum of the muon-jet system, detailing the background estimation by fitting data with empirical functions. The tests performed to validate the method used are presented at the end of the chapter.

Chapter 7 presents the analysis results, where they are interpreted as limits on the LQ mass and coupling. A comparison with previous LQ searches in different channels is also presented.

Chapter 8 gives a summary of the analysis results.

Chapter 1

Leptoquarks at the Large Hadron Collider

This chapter begins with a concise overview of the Standard Model of particle physics, laying the groundwork for the subsequent discourse. It then introduces leptoquarks as hypothetical particles predicted by various theories extending beyond the Standard Model. The research into leptoquarks, mainly focusing on the mechanisms for generating scalar leptoquarks, is scrutinized. It then proceeds to present a novel channel of lepton-induced leptoquark production, the subject of this thesis. The discussion encompasses its potential impact and significance in the ongoing research landscape, offering insights into how this novel approach may enhance our understanding of particle physics and potentially lead to groundbreaking discoveries.

1.1 The Standard Model of Particle Physics

The Standard Model (SM) is the quintessential framework in particle physics, offering a comprehensive description of the interactions between the elementary components of matter [120]. It encapsulates the interactions mediated by three foundational forces of nature: the strong, electromagnetic, and weak forces, and explains the presence of mass in elementary particles. Initiated by Sheldon Glashow's efforts in 1961 to unify electromagnetic and weak interactions, the SM underwent significant evolution with the inclusion of the Higgs mechanism[128][110], as developed by Steven Weinberg [171] and Abdus Salam [163]. This addition was pivotal, giving the model its contemporary form. Through extensive experimental verification, the SM has proven to be a well-tested theory[126]. The sub-atomic interactions within the SM are framed within a consistent quantum field theory, characterized by a renormalizable gauge theory based on the non-Abelian symmetry group:

$$SU(3) \times SU(2) \times U(1). \tag{1.1}$$

This group's non-Abelian nature signifies that the commutation relations between the group's generators are non-trivial. Moreover, the theory is gauge invariant: the interaction terms are invariant under group transformations. In the SM, fermions — spin-1/2 particles — are described in terms of symmetric interactions under the group mentioned above. The fermionic interactions in the SM's Lagrangian are constructed

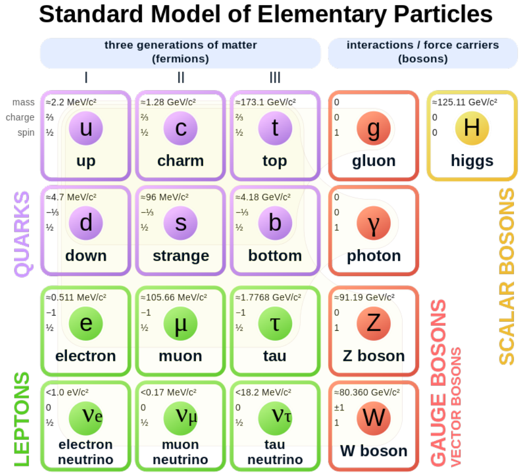


Figure 1.1. The Standard Model of elementary particles [20].

to respect this group symmetry, a foundational aspect of the theory. According to Noether's theorem [148], each symmetry within a system corresponds to a conserved quantity. The SM employs the electroweak symmetry group $SU(2) \times U(1)$, which governs the weak isospin and hypercharge, and $SU(3)$ for the color charge associated with quantum chromodynamics. For $SU(2)$, the subscript L designates the chirality of interactions. In quantum field theory, fermion fields represent the Lorentz group and are distinguished by their chirality, denoted by the operator γ^5, which defines particles' helicity for massless states. Chirality is aligned with the spin direction relative to a particle's momentum, with left-handed particles participating exclusively in weak interactions within the SM. The fermions of the SM are categorized into two distinct types based on their interaction strength. Quarks, which exhibit color charge, engage in strong interactions, while leptons, lacking color charge, are involved only in electroweak interactions. The fermion fields of the SM, which encompass a diverse array of particles, are illustrated as follows:

$$L_{i,L} = \begin{pmatrix} \nu_i \\ e_i \end{pmatrix}_L, \quad e^{\alpha}_{L,i} = \begin{pmatrix} u^i_{\alpha} \\ d^i_{\alpha} \end{pmatrix}_L, \quad u^{\alpha}_{R,i}, \quad d^{\alpha}_{R,i}. \tag{1.2}$$

Here, i represents the flavor index, grouping quark and lepton flavors into three families. Each family comprises left-handed doublets and right-handed singlets, adhering to the transformation properties of the electroweak symmetry. The chiral nature of these states is explicit in their transformation under $SU(2)_L$ as doublets or singlets, influencing their participation in the electroweak interactions. The SM is fundamentally divided into two branches: Quantum Chromodynamics (QCD), the theory describing the strong force interactions, and the Electroweak (EW) theory, which unifies electromagnetic and weak forces. The SM Lagrangian can be therefore written as:

$$\mathcal{L} = \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{EW}}. \tag{1.3}$$

To maintain gauge invariance across both QCD and EW theories, the model introduces spin-1 bosons that correspond to the symmetry group's generators, which are

the force carriers within the SM.
Because of the symmetries of the model, the Lagrangian cannot contain explicit mass terms for the particles involved. Contrary to observations, this would imply that all fundamental particles are massless. This problem has been resolved with the Brout-Englert-Higgs mechanism. The seminal works detail a mechanism known as spontaneous symmetry breaking, through which massive fermions and bosons gain mass via interactions with the newly postulated scalar, termed the Brout-Englert-Higgs (or simply Higgs) boson [128][110].

1.1.1 Open questions in the Standard Model

The SM, rooted in rigorous mathematics, has successfully elucidated aspects of the electroweak and strong forces. The ATLAS [22] and CMS [64] experiments at CERN in 2012 confirmed its predictions by identifying a Higgs-like particle. Yet, the SM does not claim to be an all-encompassing theory of the universe, leaving several phenomena unaddressed.
The SM (with massless neutrinos) has 19 free parameters whose numerical values are established by experiments, which span over a vast range of energy scales. This feature does not align with the naturalness principle, which demands that the free parameters of a theory fall in the same order of magnitude [32]. Disparities are evident within the parameters themselves, such as the quark masses, which exhibit marked hierarchies and lack a theoretical basis within the model and similarly for charged leptons. Further, the SM necessitates fine-tuning to reconcile the Higgs mass with loop-level contributions. They come in two, competitive contributions and appear to be precisely $fine\,tuned$ to allow for the Higgs mass to be extremely small with respect to the value of the correction terms themselves.
Several pieces of evidence hint at the non-completeness of the SM theory: a brief summary of phenomena non explained by the SM is reported below.

- Experimental observations of neutrino flavor oscillation prove that SM neutrinos are massive particles [53, 11], while no mass term can be accounted for neutral leptons in the SM [18, 146].
- It accounts for only a fraction of the universe's energy content, leaving dark matter and dark energy largely unexplained, constituting most of the universe's mass-energy balance [151, 150].
- While the SM's interactions do not differentiate between matter and antimatter, the observable universe is dominated by matter, a disparity not yet clarified by the SM [37].
- The SM is compatible with special relativity but does not integrate general relativity, failing to unify all four fundamental forces [116, 118, 159].
- The SM does not explain why there are three generations of fermions and what is the reason for an apparent symmetry between quarks and leptons [112].

1.2 Leptoquarks in theories beyond the Standard Model

One of the prominent attributes of the Standard Model is the evident symmetry between quarks and leptons. This symmetry is reflected in the organization of quarks and leptons within weak isospin multiplets because these groupings are duplicated across three different generations of fermions. This symmetry is crucial for the precise negation of chiral (triangle) anomalies across each fermion generation [155], ensuring the theory's renormalizability. A necessary condition for this cancellation is the existence of exactly three color states for quarks. Consequently, the observed symmetry between quarks and leptons in the Standard Model may suggest the presence of a more foundational theory where quarks and leptons are interrelated. Within such a theoretical framework, one might anticipate the emergence of new particles known as leptoquarks (LQs), which couple to lepton-quark pairs and act as intermediaries for lepton-quark transitions.
Grand Unified Theories (GUTs), such as the Georgi-Glashow and Pati-Salam models [154], posit that quarks and leptons could be unified under a broken $SU(5)$ or $SU(4)$ symmetry respectively, wherein LQs manifest as gauge bosons. Further discussions on LQs regarding technicolor theories [111, 136, 105, 117] and composite models [164, 5, 173] appear in the literature.
The leptoquarks are color triplets under $SU(3)_C$, appearing as scalar (spin-0) or vector (spin-1) bosons carrying both baryon (B) and lepton (L) number and fractional electric charge. Their exact properties, such as spin, weak isospin, electric charge, chirality of the fermion couplings, and fermion number (F), depend on the structure of each specific model. Because different models contain leptoquarks of many differing properties, direct searches for leptoquarks at collider experiments are typically carried out in the context of effective leptoquark models. With limited quark and lepton field combinations in the SM, the classification of possible LQs is feasible. According to the Buchmüller-Rückl-Wyler classification, ten potential LQ multiplets, comprising scalars and vectors, have been identified [43]. The detailing of their couplings to fermions, alongside their fermion number (F), can be represented as:

$$F = 3B + L. \tag{1.4}$$

LQs with $|F| = 0$ couple to fermion-antifermion pairs, while those with $|F| = 2$ couple exclusively to either two fermions or two antifermions, constrained by the electric charge limits that LQs can possess.
Depending on the type, leptoquarks can decay either to a charged lepton and a quark or to a neutrino and a quark. It is customary to express the branching ratio for a leptoquark decaying to a charged lepton and a quark, $\mathrm{BR}(LQ \to l^{\pm}q)$, as β. In the minimal Buchmüller-Rückl-Wyler (mBRW) model, β is constrained to values of 0, 1/2, or 1. In broader theoretical frameworks often utilized in collider experiments, β is considered a free parameter. This flexibility is permissible if certain mBRW model assumptions are relaxed [33, 15, 127] and leptoquarks may couple to with fields beyond the Standard Model (SM) gauge bosons and fermions. However, for collider searches, it is generally presumed that the branching ratio to a neutrino and a quark, $\mathrm{BR}(LQ \to \nu q)$, is $1 - \beta$. In the analysis proposed in this thesis, we will focus on LQ decays with $\beta = 1$; therefore the LQ always decays into a charged

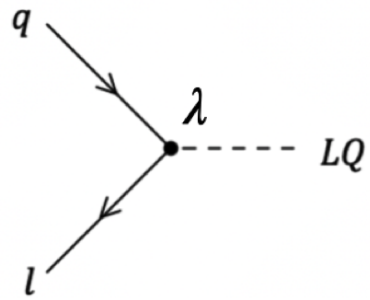


Figure 1.2. LQ coupling to both a quark and a lepton.

lepton and a quark.

Historically, leptoquark searches have only considered models that couple to a lepton and quark of the same generation, motivated by limits placed on flavor-changing neutral currents [166], proton decay [149, 141], and other rare processes. Observational evidence supports inter-generational mixing in leptoquark decays. Assuming non-chiral LQs with couplings to left-handed and right-handed fermions, cross-generational LQs can also be an explanation for the deviations observed in the anomalous magnetic moment of the muon because of loop-level modifications of EM interactions of the muon [50, 38]. These hypotheses are supported by recent analyses suggesting that LQs might be the underlying cause for the observed discrepancies in neutrino masses, with the loop corrections providing a plausible explanation for their observed minuteness [52, 139]. Furthermore, new theoretical frameworks have been proposed to account for the discrepancies noted in the W boson mass measurements reported by the CDF experiment [125, 51]. The model presented in this thesis assumes a cross-generational coupling only between quark families, while there is no mixing between the lepton families.

1.3 The Large Hadron Collider

The Large Hadron Collider (LHC), the largest particle collider to date, operates as a proton-proton collider within the European Organization for Nuclear Research (CERN), positioned near Geneva. This complex houses a circular ring 27 kilometers long where superconducting magnets bend the particles to follow its paths, and radio frequency cavities increase their velocities. Depicted in Figure 1.3, the CERN accelerator complex is illustrated, highlighting the experiments and beamlines.

Initially, protons are sourced from hydrogen gas and are then grouped into clusters known as "bunches" within the Linac2, a linear accelerator, where they reach energies of 50 MeV. These protons are then propelled into the Proton Synchrotron Booster (PSB), which encompasses four acceleration rings and boosts their energies up to 1.4 GeV. From there, they are funneled into the Proton Synchrotron (PS), a larger acceleration ring of 628 meters in circumference, which increases the proton energy to 25 GeV, organizing the bunches to maintain a 25 ns interval between each bunch. The proton bunches proceed to the Super Proton Synchrotron (SPS), where acceleration continues until they reach 450 GeV. Subsequently, these proton beams are channeled

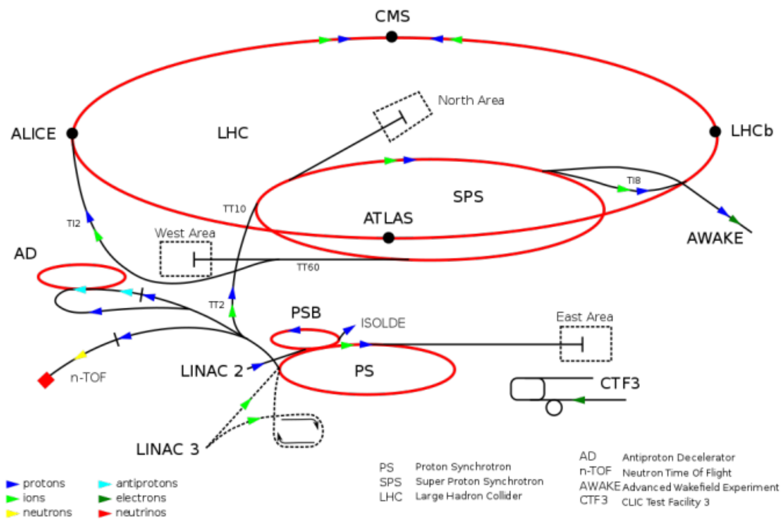


Figure 1.3. Structure of the CERN accelerating complex. Linac2, Booster, SPS, and LHC represent the acceleration chain; interaction points are shown along the LHC ring, together with heavy ion experiments [119].

into the main collider ring in opposing trajectories within two ultra-high vacuum tubes. Here, superconducting magnets and radio frequency cavities are instrumental in maintaining the protons on course and increasing their energies. Upon achieving peak energy levels, the protons collide at four interaction points, aligning with the four principal experiments stationed there: ATLAS (*A Toroidal LHC Apparatus*) [99], ALICE (*A Large Ion Collider Experiment*) [98], CMS (*Compact Muon Solenoid*) [100], and LHCb (*LHC Beauty*) [101].

The LHC is classified as a hadronic collider, leveraging the complex structure of hadrons made of smaller components: quarks, bounded by gluons, the carriers of the strong nuclear force. At substantial energies surpassing their rest mass, hadrons lose their structural coherence, leading to collisions among their constituent quarks that carry varying proportions of the proton momentum. This leads to fluctuating effective center-of-mass energies for each collision event, thus positioning the LHC as a "discovery machine", capable of probing a vast energy spectrum.

Since its inception in 2009, the LHC is projected to continue operations until at least ~ 2040. An outline is shown in Figure 1.4. Its operational life is segmented into Runs for collision and data acquisition, interspersed with shutdown periods dedicated to maintenance and enhancements. Runs are marked by consistent readings of two key collider metrics: instantaneous peak luminosity and center of mass energy ($\sqrt{s}$). Significant upgrades and adjustments to the collider mechanisms and accelerator infrastructure are performed during the long shutdowns, which correspond with the experiment upgrade phases, allowing for the refinement and stabilization of these parameters for subsequent Runs.

The instantaneous peak luminosity of the LHC is given by:

$$\mathcal{L} = \frac{\gamma n_b N^2 f_{\text{rev}} R}{4\pi \beta^* \epsilon_n}, \tag{1.5}$$

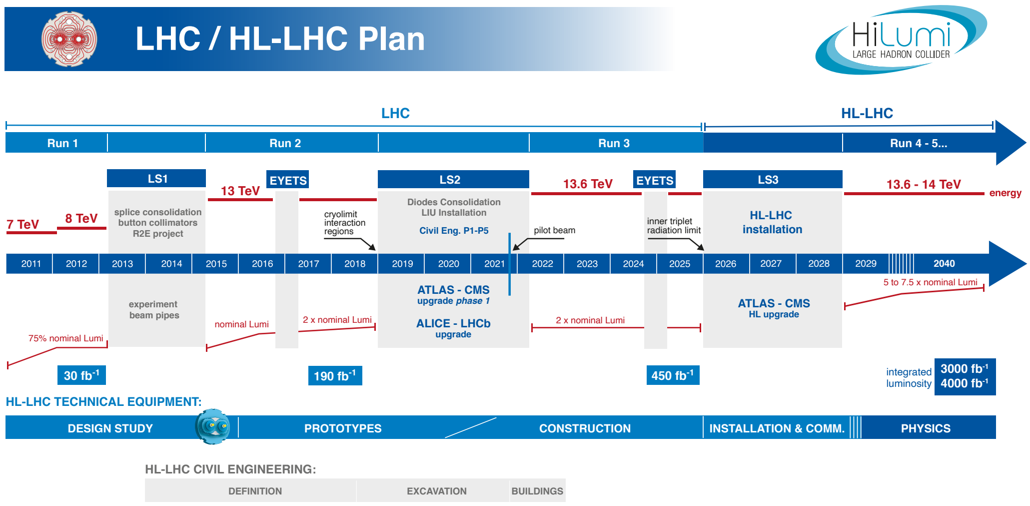


Figure 1.4. Operational chronology for the LHC Extending to 2037. Progressions in the center of mass energies are marked in red, while luminous enhancements are charted in green. The LHC baseline luminosity level is set at $L = 1 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$. Cumulative luminosity tallies for each run phase are delineated [1].

where γ is the Lorentz factor, n_b represents the number of proton bunches, N is the proton count per bunch, f_{rev} is the revolution frequency, β^* the beta function at the collision point, ϵ_n the normalized transverse emittance, and R denotes the luminosity reduction factor due to the geometric layout of the crossing beams [170]. The center of mass energy (in red in Figure 1.4), denoted as $\sqrt{s}$, initially set to 7 TeV, was elevated to 13 TeV in the transition from Run 1 to Run 2, owing to advancements in the accelerator system and the integration of stronger superconducting magnets for beam deflection. The record collision energy of 13.6 TeV has been reached in 2022 Run 3 collisions. The peak luminosity gradually rose, as seen in the red line in Figure 1.4, through LHC operating years by tuning the beam's focus in the proximity of the interaction regions. Another important parameter is the integrated luminosity, a measure integral to estimating collision counts, which is represented by:

$$L = \int L(t)dt, \tag{1.6}$$

where $L(t)$ is the time-dependent luminosity. Fluctuations in peak luminosity occur over time, due to ongoing beam focusing and the incremental attenuation of beam quality with each successive interaction. The number of collisions, N, is derived from the total proton-proton cross-section, σ_{pp}, and the integrated luminosity:

$$N = \sigma_{pp} \int L(t)dt, \tag{1.7}$$

For individual final states i, the number of events, N_i, is computed as:

$$N_i = \sigma_i \int L(t)dt, \tag{1.8}$$

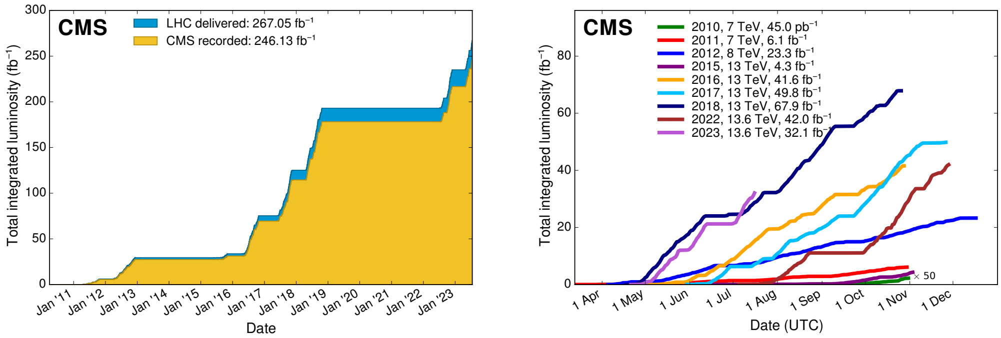


Figure 1.5. Left: integrated luminosity recorded for CMS from May 2011 to November 2023 [57]. Right: CMS integrated luminosity data per year from 2010 to 2023 [57].

with σ_i representing the cross-section for the specific final state.

Figure 1.5 illustrates the aggregate integrated luminosity provided by the Large Hadron Collider (LHC) over its years of operation, highlighting the annual luminosity data. The integrated luminosity has steadily increased, nearly doubling between 2016 and 2018. The commencement of the LHC's first operational run, termed Run 1 and spanning from 2009 to 2011, featured beam energies that surpassed the Tevatron 0.98 TeV per beam, thereby establishing the LHC as the world's most potent particle accelerator. The beams' energy was subsequently increased, reaching 3.5 TeV in 2011 and 4 TeV in 2012. The month of July 2012 was significant due to the discovery of a new particle with a mass of 125 GeV, later identified as the Higgs boson, aligning with one of the cornerstone goals of the LHC project. Post reaching 30 fb^{-1} of integrated luminosity, the LHC underwent its first Long Shutdown (LS1), during which the collider peak luminosity neared 75% of the planned maximum as outlined in the design. This period was utilized for upgrading the collider, preparing it to handle beams at 7 TeV, while the magnets were conditioned for 6.5 TeV beams. Run 2 commenced post-LS1 in 2015, pushing the center of mass energy to 13 TeV and achieving and surpassing the design luminosity of the collider. Run 2 provided the experiments with an integrated luminosity five times larger than that of Run 1, delivering a peak luminosity of 2.14×10^{34} $\text{cm}^{-2}\text{s}^{-1}$ at $\sqrt{s} = 13$ TeV, exceeding the LHC design luminosity of $L = 1 \times 10^{34}$ $\text{cm}^{-2}\text{s}^{-1}$. The 2018 integrated luminosity is shown in the left plot of Figure 1.5. Throughout their activity, LHC experiments have conducted precision measurements that substantially agree with the SM predictions or have set constraints on BSM phenomena. Data collection activities for Run 3 resumed in July 2022. The LHC will have for another substantial shutdown, LS3, to escalate peak luminosities in preparation for the High Luminosity LHC (HL-LHC) program [6], which aims to increase peak luminosity by a factor of 5 to 7.5 relative to the original LHC design. By 2025, the beam-focusing quadrupoles at the ATLAS and CMS interaction points are expected to reach the end of their operational life due to sustained radiation exposure. These will be replaced with modern quadrupole triplets, and supplementary crab cavities [6] will be installed to enhance bunch overlap at collision points. Additionally, a novel scheme for bright bunch trains [6], expected to be initiated during LS2, will facilitate the production of more intense

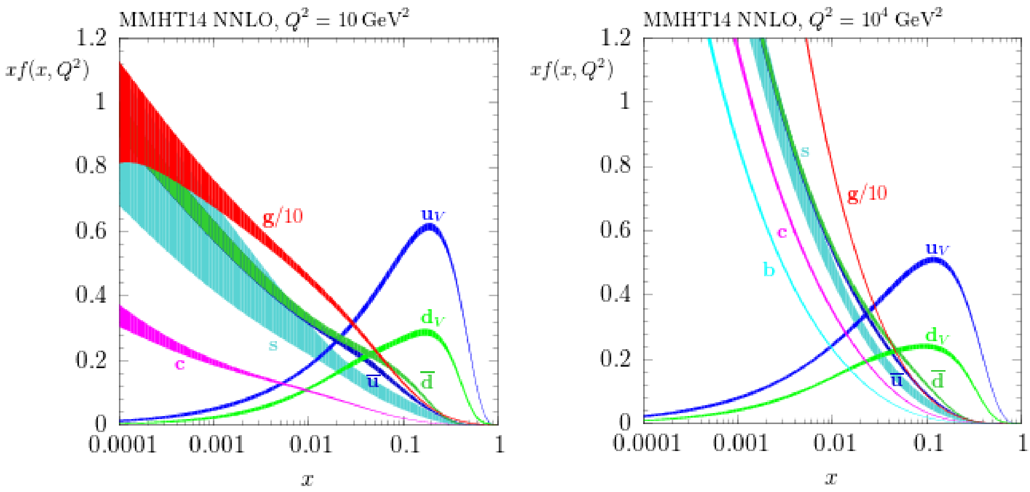


Figure 1.6. The momentum distributions (PDFs) of all flavours of the partons at two scales $Q^2 = 10\,GeV^2$ (left) and $Q^2 = 10^4\,GeV^2$ (right). These curves are deduced from NNLO QCD fits to global hard-scattering data, with uncertainties arising from experimental errors and model assumptions [123].

beams and finer beam cross-sections at collision points, substantially augmenting LHC luminosity.

1.3.1 The proton substructure

Efforts to decipher the binding of partons within protons have led to progressive insights into the characteristics of the proton. The Quark-Parton Model (QPM) [54], conceived in the 1960s, initially sought to clarify outcomes from deep inelastic scattering experiments at the Stanford Linear Accelerator Center (SLAC) involving colliding protons with electrons. Subsequent studies, benefiting from data from experiments like EMC [3], H1 [7], and ZEUS [129], expanded the QPM to encompass scattering kinematics using perturbative Quantum Chromodynamics (QCD) [108] adjustments. The advancement of this model, known as the Enhanced Quark Parton Model [104], delineates the proton as a conglomeration of three primary quarks, termed valence quarks, encircled by a dynamic sea of gluon-generated quark-antiquark pairs. These sea partons are in a perpetual exchange governed by the strong force.

Parton distribution functions (PDFs) are essential tools for describing the chance of finding a parton inside the proton with a specific momentum portion. These functions, which vary with the parton's momentum fraction x, are based on perturbative quantum chromodynamics (pQCD) calculations and validated by experimental data. Collaborative efforts have led to the extraction of PDFs from a wide range of experimental data, providing insights into the proton's structure. Figure 1.6 shows PDFs for two different energy scales. The definition of the PDFs implies the normalisation condition that the sum over the momenta of all partons equals the proton momentum, or equivalently:

$$\sum_i \int_0^1 x f_i(x, \mu^2) dx = 1. \tag{1.9}$$

Additionally, the fact that a proton has the valence quark structure uud can be expressed by the following two equations:

$$\int_0^1 \left[f_u(x, \mu^2) - f_{\bar{u}}(x, \mu^2)\right] dx = 2, \tag{1.10}$$

$$\int_0^1 \left[f_d(x, \mu^2) - f_{\bar{d}}(x, \mu^2)\right] dx = 1. \tag{1.11}$$

With the foundational concepts in place for representing the proton's internal structure, we can formulate the principal equation for determining the cross sections of high-energy proton-proton interactions. This computation is grounded in the QCD factorization theorem and can be articulated as follows:

$$\sigma_{pp \to X+Y}(\mu_F, \mu_R) = \sum_{i,j} \int_0^1 \int_0^1 dx_1 dx_2 f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \hat{\sigma}_{ij \to X}(x_1, x_2, \mu_R). \tag{1.12}$$

Here, the summation encompasses all types of partons, i.e., $i, j = g, u, \bar{u}, d, \bar{d}, \ldots$, and Y represents the resultant hadrons from non-interacting partons. The parton level cross-section $\hat{\sigma}_{ij \to X}$ is calculable in perturbative QCD and depends on the chosen renormalization scale μ_R. The important statement of the factorization theorem is that the complete calculation for hadron collisions can be separated into a partonic cross section which is convoluted with the parton distribution functions of the hadron. Hereby, the former can be calculated perturbatively by making use of the asymptotic freedom of QCD in the high-energy regime whereas the latter encodes the non-perturbative effects at small energies. The separation of both energy regimes is dictated by the choice of the factorization scale μ_F. While the renormalization group equations, which govern the running of the PDFs, are known as DGLAP equations [16, 107, 122], they need to be anchored with some starting value extracted from experimental data. Nowadays, several groups perform fits to experimental data and provide PDF sets like the examples shown in Figure 1.6. Even though the renormalization scale μ_R and the factorization scale μ_F describe conceptually different aspects of the calculation, they are often chosen to be the same $\mu = \mu_R = \mu_F$. Another important consequence of the proton structure for pp collisions at the LHC is that the longitudinal momentum of the interacting parton-parton system is unknown on a per-event basis as it not only depends on the known four-momenta of the colliding protons but also on the, experimentally unknown, momentum fractions x_1 and x_2 of the two colliding partons.

1.3.2 Proton-proton collisions at LHC

Figure 1.7 visually represents a typical proton-proton (pp) collision.
The rate of events, which is directly proportional to the cross-section σ at a given instantaneous luminosity, is given by

$$R = \mathcal{L} \cdot \sigma. \tag{1.13}$$

The anticipated total inelastic cross-section σ_{pp} for proton-proton collisions has been measured to be $\sim$ 100 mb at an energy of 13 TeV.
The inelastic cross-section encompasses two interaction categories. The first, known as *soft collisions*, occurs when the colliding protons exchange minimal momentum. In this case particle scattering at large angle is suppressed and most of the final state particles escape down the beam pipe. The second category, *hard collisions*, involves interactions between quark and gluon (i.e. the proton constituents), leading to significant momentum transfers perpendicular to the beam's direction (p_T). Hard collisions are the primary source of significant physics events. Despite being less frequent than soft collisions, the high-p_T remnants of soft events can compete with hard collisions regarding event rate, contributing to the background in high-p_T signal events.
In hard interactions, the effective center-of-mass energy $\sqrt{s}$ is determined by the energy shared between the two partons and is proportional to the energy fraction they carry. The process is calculated based on matrix elements at some order in perturbation theory (usually leading order (LO) or next-to-leading order (NLO)) integrated over a given phase space.

Quarks and gluons are the fundamental constituents of protons and carry color

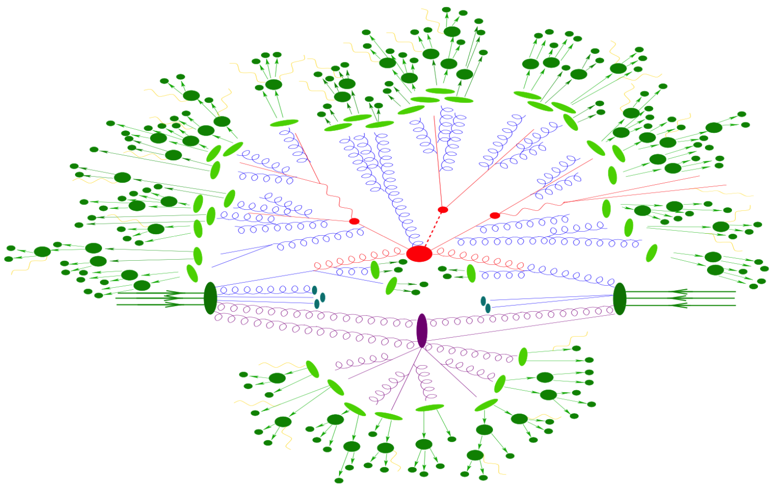


Figure 1.7. Schematic representation of a proton-proton collision, incoming protons are visualized as three green parallel lines approaching from both the left and right sides of the diagram. The interaction region where hard scattering occurs is highlighted in red. The fragmentation process, which follows the scattering, is illustrated with blue lines. Subsequently, hadron formation is depicted using light green ovals, transitioning into stable hadrons represented by dark green shapes. The photon emission process, a consequence of quantum electrodynamics radiation, is depicted with wavy yellow lines. In addition to the primary interaction, the underlying event, which may include secondary hard interactions, is encapsulated by a violet oval. Proton remnants, which are not involved in the primary interaction, are indicated by small cyan circles [172].

charge and, hence, they are not directly observable in experiments. They rather manifest themselves as collimated jets of particles, consisting mostly of hadrons and photons, but they may also contain leptons from decays of unstable hadrons. For the rest of this document, jets are considered to be the observable embodiment

of quarks and gluon. Further details on the detection and reconstruction of these objects is given in Section 4.2.
Parton showering [145] models the conversion from quarks and gluons to observable jets. These models simulate the development of quarks and gluons after a collision, known as final state radiation, and also trace back the interactions that occurred before the collision, known as initial state radiation. In general, parton shower models can be broken down into the following parts:

- **Fragmentation** (also called parton shower) models summarise the effect of repeated gluon splitting $g \to gg$ or $g \to q\bar{q}$ and gluon radiation $g \to qg$ which results in a cascade of generated partons. These models' calculations are approximate and valid up to the energy level where the strong force interactions are not mathematically perturbative. For the evolution of the incoming partons backward in time, a matching with the factorization scale of the PDF has to be performed as the PDF already catches effects from gluon splitting and radiation at lower energies. Furthermore, if the hard process is calculated at a higher order in QCD (usually at NLO), a matching between real emissions of partons already included in the calculation of the hard process and additional splittings simulated by the fragmentation model has to be done to avoid double-counting of certain contributions [144].
- **Hadronization** is the process where quarks and gluons come together to form hadrons, occurring at energy scales where the strong force is not perturbative. The models that describe hadronization are based on empirical data [138].
- **Decay** of hadrons, although not directly a part of parton shower models, is well-understood through electroweak interactions and QCD principles. Therefore, these decay processes are often included in the simulations of parton showers [145].

1.4 Leptoquarks searches at colliders

Various approaches have been explored for the detection of leptoquarks. At lepton colliders such as e^+e^- and $\mu^+\mu^-$, leptoquarks can be generated in pairs through s-channel photon and Z boson exchange, or via t-channel quark exchange. Leptoquarks can also be singly produced. The cross-sections for such events are influenced by the Yukawa coupling (λ) - i.e. the coupling of a quark q and a lepton ℓ to the leptoquark LQ - which remains undetermined.
In pp collisions, leptoquarks can be produced either singly or in pairs. Moreover, they can be exchanged in a non-resonant production of two leptons.

1.4.1 Pair Production

During proton-proton collisions at the LHC, LQs are predominantly produced by pair production. There are mainly two processes:

1. Gluon-gluon fusion:
$$g + g \to LQ + \overline{LQ}, \tag{1.14}$$

where LQ stands for a leptoquark and $\overline{LQ}$ for its antiparticle.

2. Quark-antiquark annihilation:

$$q + \overline{q} \to LQ + \overline{LQ}, \tag{1.15}$$

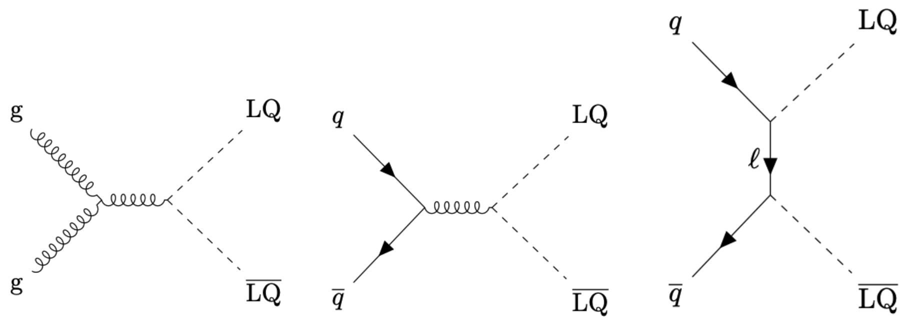


Figure 1.8. Feynman diagram for gluon-gluon fusion, quark-antiquark annihilation, t-channel production respectively on the left, center, right.

While gluon-gluon fusion is the dominant production mode for small leptoquark masses, quark-antiquark annihilation processes become more important with higher leptoquark masses. There is also a possibility for production through a t-channel lepton exchange. This process is notably less significant and involves the Yukawa coupling λ.
The production cross-section is almost independent of the coupling λ. Therefore, the majority of previous searches have concentrated on pair production. However, at higher masses, single production of LQs becomes a more favorable approach for LQ searches. The cross-section for pair production decreases rapidly with increasing scalar leptoquark mass due to the phase-space suppression for producing two particles.

1.4.2 Single Production (gluon initiated)

Single production of LQs occurs in association with a lepton via quark-gluon fusion:

$$g + q \to LQ + l, \tag{1.16}$$

resulting in the production of a single leptoquark LQ and a lepton l. Feynman diagrams for the single leptoquark production processes are presented in Figure 1.9. In both s-channel and t-channel processes, the Yukawa coupling λ plays a central role in the interaction between a quark and a lepton with the leptoquark. The cross-section for this process is proportional to λ^2; therefore, for high LQ masses and larger couplings, this production mechanism becomes dominant over pair production ($M_{LQ} \approx 1$ TeV and $\lambda \approx 0.6$).

1.4.3 Non-resonant dilepton production

The non-resonant production of two leptons through the t-channel exchange of a LQ is an additional leptoquark mechanism production. Such a process involves

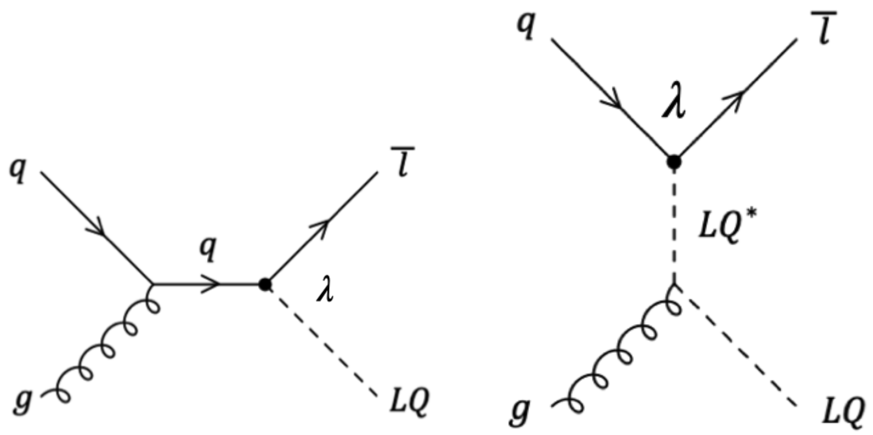


Figure 1.9. Feynman diagram for the s-channel (t-channel) single leptoquark production on the left (right).

quark-antiquark annihilation, represented by the following reaction:

$$q + \bar{q} \to l^+ l^-, \tag{1.17}$$

where l^+ and l^- are leptons of opposite charges in the final state. The Feynman diagram corresponding to the dilepton production mechanism is depicted in Figure 1.10. The cross-section scales with λ^4, adding signal sensitivity at high values of

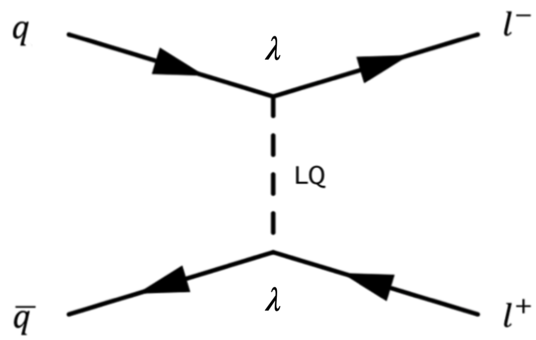


Figure 1.10. Feynman diagram relevant for the dilepton production mediated by a leptoquark.

the coupling strength λ.

1.4.4 Current limits on leptoquarks

Figure 1.11 shows the sensitivity of different physics channels in the λ vs. mass plane in terms of exclusion of a scalar LQ model. Due to phase space constraints, single production is anticipated to overshadow pair production for masses around 1 TeV, enhancing sensitivity for such processes. This sensitivity notably depends on the leptoquark-lepton-quark coupling, λ, since single production is proportional to λ^2. Enhanced coupling strengths allow for amplified sensitivity by analyzing non-resonant dileptonic states that emerge from the t-channel exchange of leptoquarks,

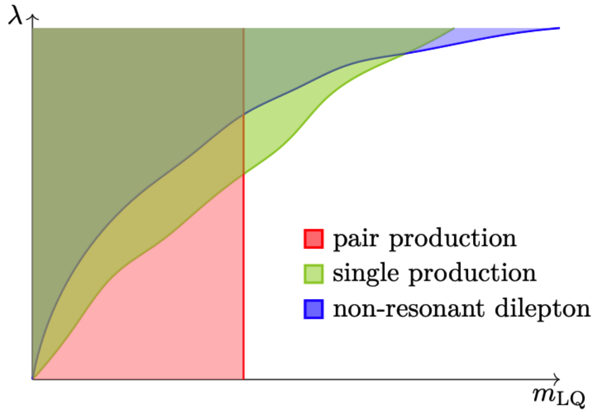


Figure 1.11. Schematic illustration of the exclusion reach of searches for LQ pair production, LQ single production, and non-resonant dilepton production as a function of the LQ mass (M_{LQ}) and the coupling strength (λ) [31]. The intersection between the pair production and the single production is for $M_{LQ} \sim 1$ TeV and $\lambda \sim 0.6$.

scaling with λ^4.

The most stringent limits LQs are reported in Table 1.1. The constraints on mass values are contingent on the specific model and the intrinsic nature of the leptoquarks (LQs). As explained before, the pair production process of LQs does not depend on the coupling constant λ, whereas single production exhibits a dependence. The limit under consideration pertains to a coupling constant $\lambda = 1$, as it represents the magnitude of coupling of primary interest in the context of lepton-induced searches, which is the central focus of this dissertation.

As explained in the following sections, this dissertation will focus on LQs decaying into muons in the case of $\beta = 1$. Under the precedent assumptions, the previous research endeavors have effectively ruled out LQs with mass scales below 1.5 TeV.

1.5 Lepton-induced single LQ production

The study of LHC processes has predominantly focused on those initiated by quarks and gluons within proton-proton collisions. Recent literature [46] introduces the prospect of examining lepton-initiated processes at the LHC. Consequently, this suggests an additional mechanism for leptoquark resonant production at LHC, which expands the scope of LHC's investigation capabilities.

1.5.1 Lepton Parton Distribution Functions

Quantum fluctuations allow the presence of leptons and photons within a proton, albeit in significantly lesser quantities compared to quarks and gluons [46]. When leptons initiate a large momentum transfer scattering, the process becomes perturbative, the lepton densities also obey an evolution equation, and a partonic calculation of the process, including higher-order corrections, becomes possible. Figure 1.12 shows the lepton PDF for an electron inside a proton, computed at the same energy scale as Figure 1.6 (right).

Process	Channel	Mass Limit	Luminosity	Experiment	$\sqrt{s}$
Single production (gluon)	LQ $\to qe$	M_{LQ} < 1755 GeV	19.6 fb^{-1}	CMS [66]	8 TeV
Single production (gluon)	LQ $\to te$	M_{LQ} < 1580 GeV	138 fb^{-1}	ATLAS [28]	13 TeV
Pair production	LQ $\to qe$	M_{LQ} < 1435 GeV	35.9 fb^{-1}	CMS [78]	13 TeV
Pair production	LQ $\to qe$	M_{LQ} < 1800 GeV	138 fb^{-1}	ATLAS [24]	13 TeV
Pair production	LQ $\to te$	M_{LQ} < 1120 GeV	138 fb^{-1}	CMS [92]	13 TeV
Pair production	LQ $\to te$	M_{LQ} < 1610 GeV	138 fb^{-1}	ATLAS [27]	13 TeV
Single production (gluon)	LQ $\to q\mu$	M_{LQ} < 660 GeV	19.6 fb^{-1}	CMS [66]	8 TeV
Single production (gluon)	LQ $\to t\mu$	M_{LQ} < 1590 GeV	138 fb^{-1}	ATLAS [28]	13 TeV
Pair production	LQ $\to q\mu$	M_{LQ} < 1530 GeV	35.9 fb^{-1}	CMS [79]	13 TeV
Pair production	LQ $\to b\mu$	M_{LQ} < 1810 GeV	138 fb^{-1}	CMS [93]	13 TeV
Pair production	LQ $\to q\mu$	M_{LQ} < 1700 GeV	138 fb^{-1}	ATLAS [24]	13 TeV
Pair production	LQ $\to t\mu$	M_{LQ} < 1420 GeV	138 fb^{-1}	CMS [92]	13 TeV
Pair production	LQ $\to t\mu$	M_{LQ} < 1640 GeV	138 fb^{-1}	ATLAS [27]	13 TeV
Single production (gluon)	LQ $\to b\tau$	M_{LQ} < 800 GeV	138 fb^{-1}	CMS [97]	13 TeV
Single production (lepton)	LQ $\to u\tau$	M_{LQ} < 2070 GeV	138 fb^{-1}	CMS [124]	13 TeV
Single production (lepton)	LQ $\to d\tau$	M_{LQ} < 1800 GeV	138 fb^{-1}	CMS [124]	13 TeV
Pair production	LQ $\to b\tau$	M_{LQ} < 1460 GeV	138 fb^{-1}	ATLAS [30]	13 TeV
Pair production	LQ $\to t\tau$	M_{LQ} < 1430 GeV	138 fb^{-1}	ATLAS [26]	13 TeV
Pair production	LQ $\to b\tau$	M_{LQ} < 1220 GeV	138 fb^{-1}	CMS [97]	13 TeV
Pair production	LQ $\to u\tau$	M_{LQ} < 1250 GeV	138 fb^{-1}	ATLAS [25]	13 TeV
Single,non-resonant and pair production	LQ $\to b\tau$	M_{LQ} < 1280 GeV	138 fb^{-1}	ATLAS [29]	13 TeV

Table 1.1. Limits on LQ masses. For pair production, the cross-section is independent of the coupling λ. The limit is reported for single production for $\lambda = 1$. All limits are reported under the assumption of LQs decaying in charged leptons 100% of the times (β = 1).

A precise knowledge of the lepton densities can be used to study lepton-initiated processes involving leptons in the initial state, even at hadron colliders. These PDFs are extremely small, so their contribution to the SM processes is generally nonrelevant with respect to processes initiated by coloured partons. However, in processes like the single leptoquark production, the contribution of a lepton-initiated process can become comparable, if not more significant, than the one initiated by quarks and gluon only, as explained in Section 3.2.1.

1.5.2 LHC as a lepton-quark collider

The process, shown in Figure 1.13 consists of a lepton from a proton colliding with a quark from the other proton, producing a LQ, which then decays into a high energy lepton and quark:

$$\ell + q \to LQ \to \ell + q. \tag{1.18}$$

The new lepton PDF allowed for precise computation of the cross-section, calculated at NLO, of the resonant LQ production from a lepton-induced process. Figure 1.14 shows the cross-section NLO/LO ratio and its uncertainty for a LQ coupling to a muon and an up quark. Coupling to other quarks have similar uncertainties on the NLO/LO ratio [121]. We can see that the process is known with great precision, with small uncertainties on cross-section ($\leq$ 4%) [46] and k factor (few %), i.e. the ratio of NLO to LO total cross-section.

For this particular process, assuming a scalar LQ as in [45] and in the limit of

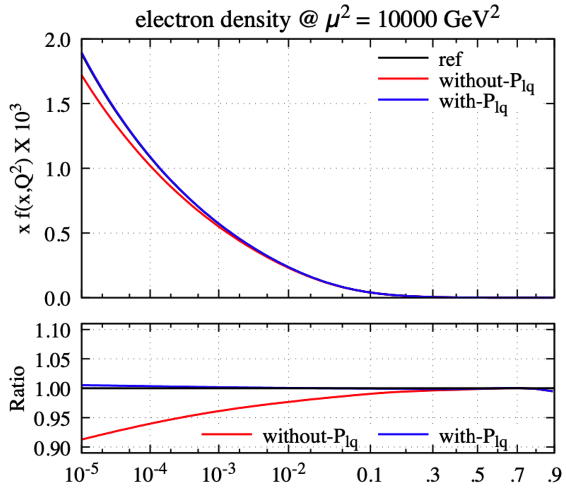


Figure 1.12. The electron PDF is shown in black, while the blue and red curves are two different approaches to higher-order correction of the lepton PDF [46].

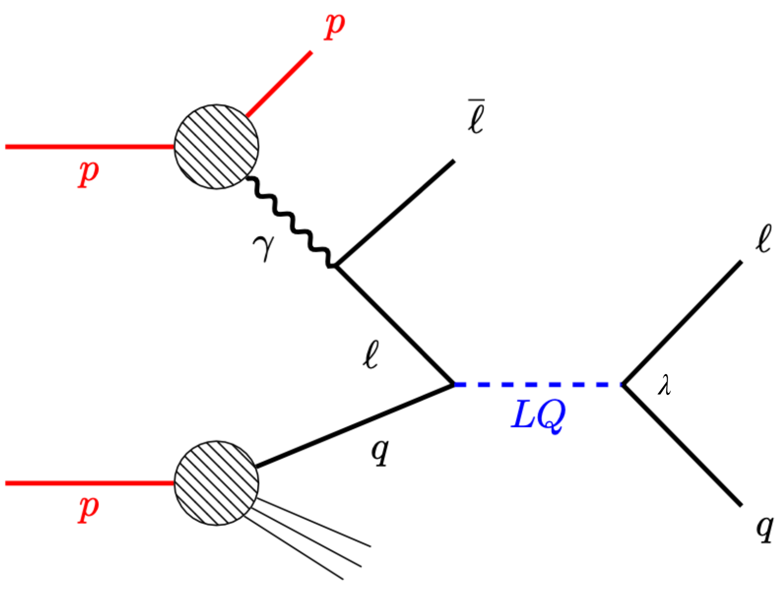


Figure 1.13. Feynman diagram for single LQ resonant lepton induced process.

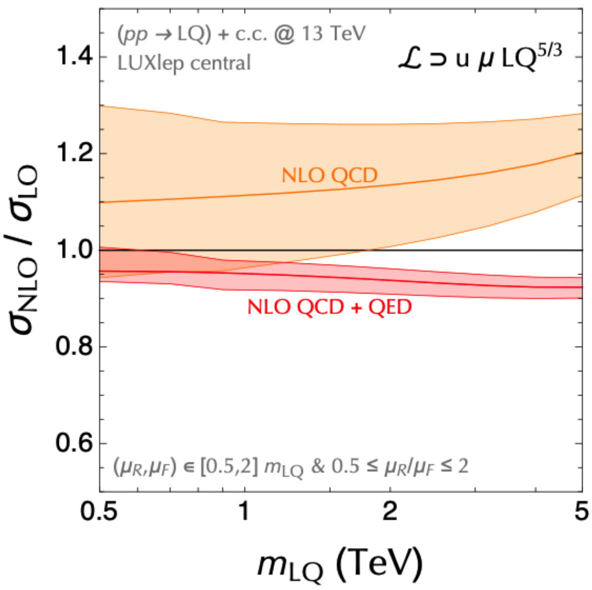


Figure 1.14. $\frac{\sigma_{NLO}}{\sigma_{LO}}$ for resonant scalar production at $\sqrt{s} = 13$ TeV in red. The colored bands are the uncertainties in the prediction [121].

$M_{LQ} >> m_\ell, m_q$, the width of the resonance can be computed as

$$\Gamma = \frac{|\lambda|^2}{16\pi} M, \tag{1.19}$$

where λ is the LQ coupling constant.
The LQ can couple to all families of leptons and a previous CMS analysis already covered the LQ decay into a tau lepton and a quark [124]. To further explore this new production mechanism, in this analysis we will study an LQ decaying into a muon and a quark. The final state is therefore composed of a high-p_T muon and a jet, the experimental signature of a quark as explained in Section 4.2. Both these objects can be reconstructed using the CMS detector (Chapter 2); therefore, it is possible to fully reconstruct the new particle mass from the invariant mass of the muon-jet system ($M_{\mu jet}$). Since the process is resonant, the presence of a LQ would produce a striking signature, represented by a bump on the smoothly falling $M_{\mu jet}$ spectrum that arises from multiple SM processes (Figure 1.15).

We are considering LQ with masses at the TeV scale, so it is produced almost at rest. We can then consider a two-body decay in the transverse plane, with the muon and quark coming from the LQ with almost the same transverse momentum ($p_T^\mu \sim p_T^q$) and back-to-back in the transverse plane. Further kinematic considerations are reported in Section 3.2. It is also possible that the lepton coming from the photon (Figure 1.13), which does not interact with the quark, is reconstructed by the CMS detector (Section 3.2). Therefore, we will study two orthogonal signal regions depending on the number of reconstructed muons (1 or 2) in the final state. After a preselection with loose cuts to reduce the background, the signal is optimized with a Boosted Decision Tree (BDT) algorithm (Section 6.1). We decided to have two categories based on the BDT output for each signal region: a high-purity category

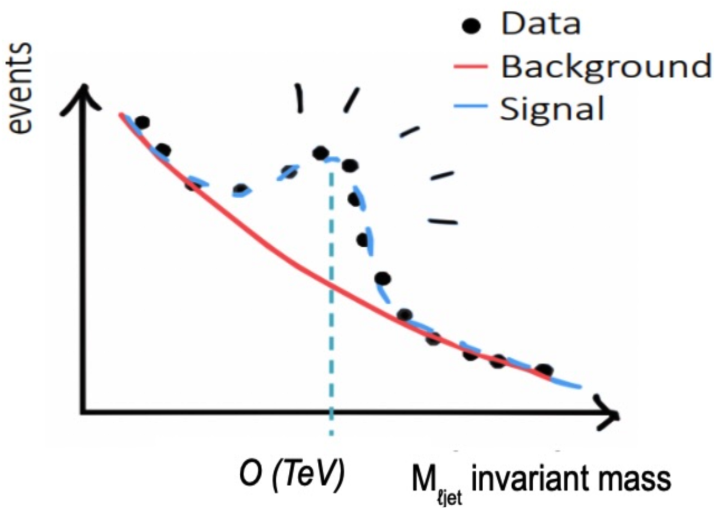


Figure 1.15. Sketch of a signature from a LQ resonance.

with low background contamination and a low-purity category to recover signal efficiency. Furthermore, every BDT category is split according to whether the jet with highest p_T is originated from a b-quark, resulting in a total of 8 independent event categories that are used simultaneously to search for LQs. To perform this analysis, we use the data recorded by CMS during Run 2 as described in Chapter 6. The results for this search are then presented in Chapter 7.

Chapter 2

The CMS experiment at LHC

The Compact Muon Solenoid (CMS) detector is built around a huge solenoid magnet, from which it takes its name. The CMS detector was constructed in fifteen sections at ground level before being lowered into an underground cavern near Cessy in France and reassembled.
The detector has a cylindrical shape with a diameter of 15 meters and is 22 meters long with a ~ 14500 tons weight.
CMS is designed to detect particles derived from the collision of hadrons and measures particle masses, momenta, energies, and charges. It is formed by different detector layers, concentric and roughly cylindrical, in a typical "onion" shape structure. The layers are specialized in different measurements or sensitive to a specific category of particles. It is divided into three parts: one central, called *barrel*, and two lateral, called *endcap*.
Working at high luminosity, the detector is designed to operate in a high radiation environment, maintaining good performances over several years of data taking and distinguishing processes of interest from backgrounds.

2.1 Coordinate system

The standard reference framework for the CMS detector is a right-handed coordinate system situated at the nominal interaction point. It is oriented such that the z axis aligns with the direction of the beam, the y axis points vertically upward, and the x axis is orthogonal, pointing radially towards the center of the LHC. Angular measurements are described relative to the beam direction: ϕ is the azimuthal angle in the x-y plane, and θ is the polar angle in the y-z plane. Pseudorapidity, denoted as η, is defined by the expression

$$\eta = -\ln\tan\left(\frac{\theta}{2}\right), \tag{2.1}$$

and is favored over θ because, for particles considered massless in good approximation, which is a valid assumption for many particles produced in high-energy interactions at the LHC, η corresponds directly to rapidity y, given by

$$y = \frac{1}{2}\ln\left(\frac{E-p_z}{E+p_z}\right), \tag{2.2}$$

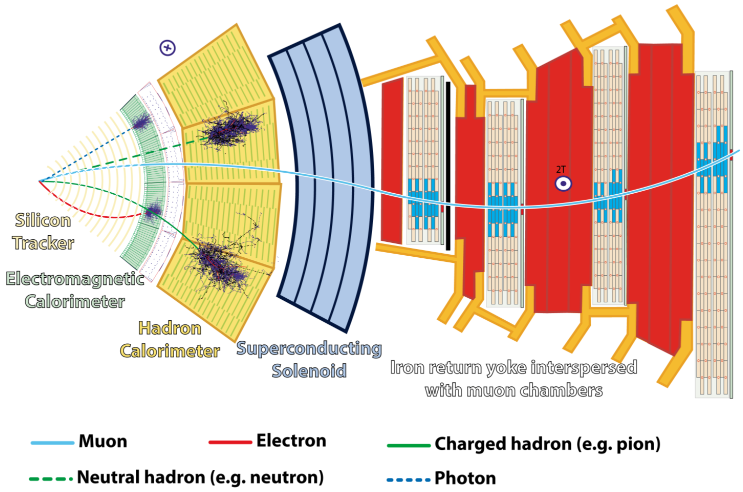


Figure 2.1. Section of the CMS detector. The lines are the particles passing through the various detectors [39].

assuming the speed of light $c = 1$. The variable y approaches zero for particles with no longitudinal momentum component and diverges negatively for particles with a longitudinal momentum component significantly larger than their transverse momentum. Differences in rapidity Δy, and correspondingly $\Delta\eta$ for massless particles, remain invariant under Lorentz boosts along the z axis. This invariant nature makes $\Delta\eta$ a crucial quantity in the analysis of collisions, capturing the essence of the partonic content within the incoming hadrons, which possess varying longitudinal momentum fractions, resulting in collisions at varying center-of-mass energies.

2.2 Detectors and subsystems

Detectors typically consist of three primary components [40]: a central barrel region, characterized by its cylindrical shape encompassing the beamline, and two endcap sections that extend the detector's coverage to areas adjacent to the beamline at greater distances from the point of interaction. Figure 2.1 presents an illustrative representation of the CMS detector.

2.2.1 Solenoid Magnet

The CMS experiment is designed to be a state-of-the-art, compact detector system emphasizing the muon subsystem. A significant magnetic bending power is essential to achieve high precision in muon momentum measurement. The CMS magnet facilitates this. Comprising a superconducting Niobium-Titanium solenoid, it generates a uniform magnetic field of 3.8 T at its core and conducts an electrical current of 18 kA. The energy stored in the magnetic field amounts to 2.4 GJ. The solenoid's length of 13 m and diameter of 5.9 m accommodate the tracking system and the calorimeters. The magnetic field outside the solenoid is guided back by an iron yoke, providing structural support for the detector.

2.2.2 Tracker

Encircling the beamline, the innermost detector has a radius of 130 cm, handling the higher flux of particles [133]. Its crucial function is to record the paths of charged particles with outstanding resolution while causing minimal interference with their natural trajectories and kinematic details. Composed of silicon-based sensors, the tracker is sectioned into the Pixel and Silicon Strips detectors. The Pixel detector consists of four barrel layers in a cylindrical arrangement and three endcap disks, each constructed from hybrid pixel detectors measuring 100 x 150 mm^2; its furthest transverse reach is 16 cm, and the endcaps are located within 50 cm of the interaction zone. The Silicon Strips detector extends to a transverse radius of 1.2 m, within a pseudorapidity range of $|\eta| < 2.5$. Strips within this detector have a thickness between 320 µm and 500 µm, and the spacing between them varies from 122 µm to 205 µm. The silicon's high segmentation and sensitivity enable position measurements within the layers to a precision of about 10 µm, essential for reconstructing accurate particle tracks through the tracker's various layers.

2.2.3 Electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) [58] is a scintillating homogeneous calorimeter with a radius between 1.2 and 1.8 meters, and it is used to measure the energy released by electrons and photons. It is composed of lead tungstate ($PbWO_4$), and it is divided into two geometrical regions: the barrel, the central region, composed of 61200 crystals, and the endcaps, two sections that delimit the barrel laterally, composed by 7324 crystals each. Figure 2.2 shows a schematic view of its structure.

The choice of this material is due to the requirement of high granularity, good

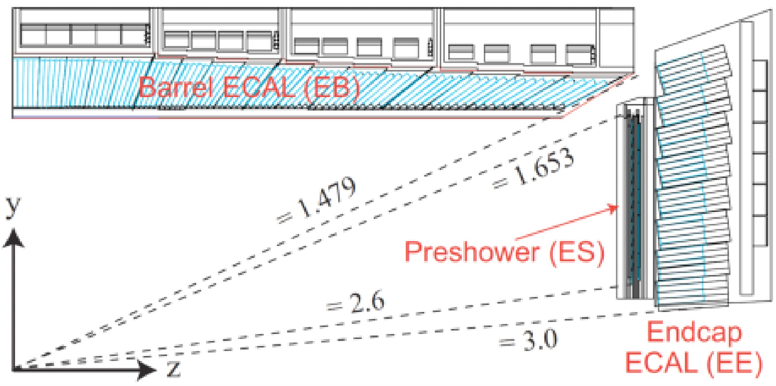


Figure 2.2. Schematic view of a quarter of the ECAL system in the (r, z) plane [60].

energy resolution, and high radiation resistance, all characteristic features of $PbWO_4$ crystals. In fact the $PbWO_4$ has

- interaction length (X_0) [135] of 0.89 cm, thus the electromagnetic shower is completely contained longitudinally in the 23 cm crystal;
- low Moliere radius (R_m) [135] of 2.2 cm, ensures lateral shower containment and, therefore, allows for the realization of a detector with high granularity;

- is a fast scintillator: the scintillation decay time is of the same order of magnitude as the LHC bunch crossing time (25 ns), enabling ECAL to measure the particles energy with a little overlap between signals from different bunch crossing;
- good radiation resistance.

The crystals are read out by avalanche photodiodes (APDs) in the barrel region and vacuum phototriods (VPTs) in the endcaps. A preshower system is placed before the endcap sections of ECAL to separate photons from the primary vertex from those coming from a π_0 in-flight decay, which could be misreconstructed as a single photon.
The energy resolution of a homogeneous electromagnetic calorimeter is [61]:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{a}{\sqrt{E}}\right)^2 + \left(\frac{b}{E}\right)^2 + c^2 \tag{2.3}$$

Where:

- The first term is the stochastic contribution, which considers fluctuations in the number of collected photoelectrons. These fluctuations include number of photons emitted, the number of photons collected, and the quantum efficiency of the photodetector;
- the second term is due to electronic noise in the readout system. The contribution depends on the instantaneous luminosity and on η;
- The third term, dominant at high energy, considers several effects, including calorimeter calibration and other systematic uncertainties.

The energy resolution in the barrel, for electrons, is measured to be [9]:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{2.8\%\,\mathrm{GeV}^{\frac{1}{2}}}{\sqrt{E}}\right)^2 + \left(\frac{12\%\,\mathrm{GeV}}{E}\right)^2 + (0.3\%)^2. \tag{2.4}$$

The relative resolution with respect to the particle energy goes from 1.5% to less than 0.4% [9]. In the ECAL, energy deposits from electromagnetic showers are analyzed to ascertain information about the centroid position. Since these showers yield varying energies across the crystals, the centroid position is computed as the energy-weighted average position of the crystal.

2.2.4 Hadronic calorimeter

The hadronic calorimeter (HCAL) [59] measures the energy of hadrons produced in pp collisions and, in combination with ECAL, also the missing transverse energy flow. The design of HCAL originates from the necessity of containing the hadronic shower inside the volume of the detector, which occupies the limited space between the ECAL and the solenoid. Contrasting with ECAL, HCAL employs a sampling methodology, integrating passive absorbing layers with active detection layers. Although this design

does not measure the energy deposited in the absorbing material, it is beneficial for its compactness and capacity to absorb hadronic shower radiation.
The HCAL's absorption layers are made of brass for their optimal interaction length, mechanical robustness, and non-magnetic nature, preventing interference with the CMS magnet's field. The active layers consist of plastic scintillators emitting prompt photonic signals upon particle interaction. These photons are then captured by fibers that shift the wavelength. Structurally, HCAL is divided into six segments: the central barrel (HB), a pair of endcaps (HE), two forward sections (HF), and an extra outer layer surrounding the barrel next to the magnet coil (HO). Figure 2.3 shows the detailed HCAL longitudinal section layout.

The barrel part, HB, spans a pseudorapidity range of $|\eta| < 1.4$ and is arranged

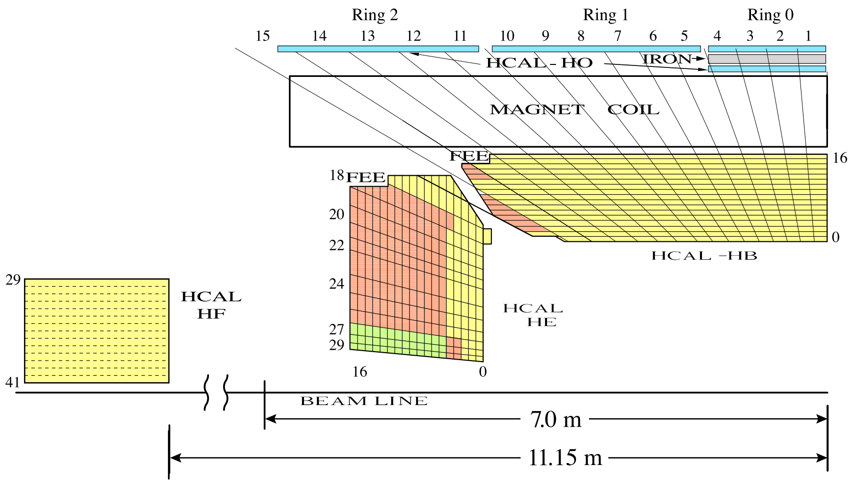


Figure 2.3. Longitudinal view of HCAL [96].

into towers, aiding in pinpointing the incident particles' entry and trajectory from the observed showers. The endcaps, HE, extend the coverage to $1.3 < |\eta| < 3.0$, matching HB's granularity to facilitate corrections for escaping hadronic showers from high-energy particles. These corrections are crucial, especially for energies surpassing 500 GeV, to avoid affecting the muon detectors. The HO layer is added to the HB, enhancing one radiation length to the calorimeter. HF, composed of steel/quartz fibers, covers $3.0 < |\eta| < 5.0$. It is crafted from distinct materials due to its distant placement from the CMS magnet's center and high exposure to particles from shallow-angle pp collisions.
The HCAL's collective energy resolution for single pions can be expressed as:

$$\left(\frac{\sigma}{E}\right) = \left(\frac{52.9\%}{\sqrt{E(\mathrm{GeV})}}\right)^2 + (5.7\%)^2. \tag{2.5}$$

Here, the noise component is considered negligible. HCAL's energy resolution substantially exceeds ECAL's, so it predominantly influences the combined resolution in the analyses involving both HCAL and ECAL measurements.

2.2.5 Muon system

The muon system [137] is the outermost of the CMS subdetectors. It has the aim of detecting muons, the only charged particles which are able to pass through the calorimeters without being absorbed. It is placed outside the magnet coil, and it has a pseudorapidity reach of $|\eta| < 2.4$. It is subdivided in a barrel and two endcaps:the barrel covers the region of $0 < |\eta| < 1.2$, and the endcaps the region $1.2 < |\eta| < 2.4$. Both regions are made of four layers of measuring stations, embedded in the iron of the magnet return yoke, where the return field of the solenoid is about 1.5 T. A sketch of the muon system is shown in Fig. 2.4.

Different experimental techniques in different regions of the detector are used:

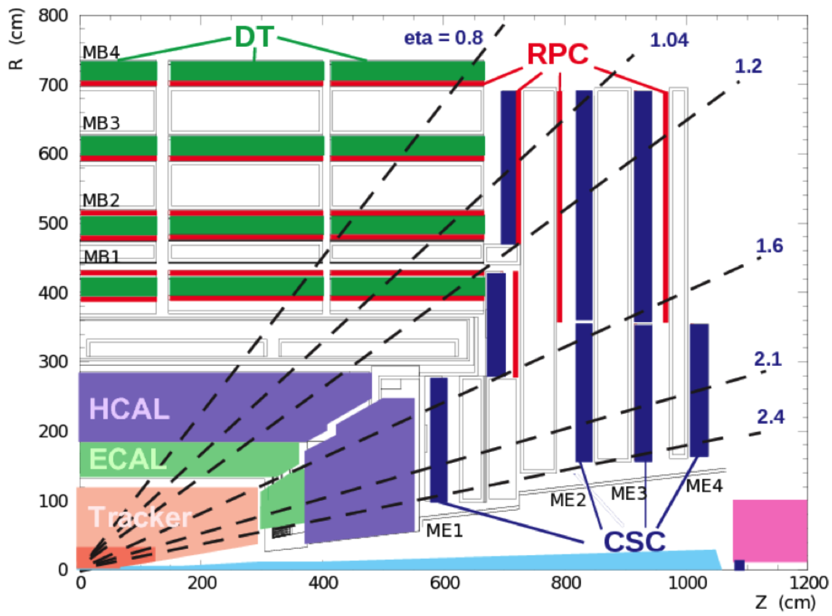


Figure 2.4. Layout of one quarter of section the CMS muon system [137].

- Drift tubes (DT): This kind of detectors is used in the central part of the muon system. Each chamber is made of twelve 4-cm-wide tubes containing a stretched wire within a gas volume.
- Cathode strip chambers (CSC): In the endcap region, where particle multiplicity is higher, arrays of anode wires, crossed with cathode strips, within a gas volume are used for muon detection.
- Resistive Plate Chambers (RPC): Both barrel and endcaps are equipped with this fast gaseous detectors. They consist of two parallel plates separated by a gas volume. Their excellent time resolution (3 ns) makes them suitable to be used also as fast high-efficiency triggers.

2.2.6 Trigger system

The LHC generates bunch crossings at a rate of 40 MHz, corresponding to a 25 ns interval. The CMS, however, has a data storage capacity for proton-proton collision information at around 2 GB/s, which translates to an event rate of approximately 1 kHz, assuming an average event occupies about 2 MB. Recording every collision event produced is beyond the CMS's capability, as most low-energy collisions do not contribute significantly to the CMS physics objectives. Therefore, a trigger mechanism [67] is imperative to isolate the noteworthy events from high-energy parton-parton collisions. The CMS trigger system employs a dual-tiered approach to decrease the event frequency from 40 MHz to roughly 1 kHz. This system comprises:

- A Level-1 (L1) trigger [82], constructed from hardware processors that execute rapid logical operations on signals from the sub-detectors.
- A High-Level Trigger (HLT) [165], a software-based system hosted on a multi-processor computing farm, analyzes reconstructed data for decision-making.

The L1 trigger mitigates the initial event rate to near 100 kHz within a latency period of 3.2 microseconds, leveraging high-resolution data from calorimeters and muon detectors. It retains essential event characteristics in pipeline memory, employing various algorithms to summarize the event data. Events that pass the L1 criteria are passed to the HLT for detailed analysis. The HLT further reduces the output rate to about 1 kHz. It is an array of computers running high-level physics algorithms designed to identify particular event structures, with each algorithmic path dissecting and making decisions based on complex data attributes. The modularity of the HLT allows for dynamic adaptation, ensuring the utmost adaptability of the software. These HLT paths, constructed with the objective of swift regional reconstruction and prompt rejection of non-essential data, prioritize rapid processing and storage of significant events in distinct Primary Datasets (PD) for subsequent offline analysis. For instance, events that include high transverse momentum muons are cataloged in the SingleMuon dataset, contingent on meeting the HLT's predefined selection criteria.

Chapter 3

Leptoquark signal generation and kinematics

This chapter describes the signal generation of a scalar LQ from a lepton induced process and the kinematics of the particles associated with the LQ decay process.

3.1 Event generation

Scalar leptoquark signal datasets are generated at the NLO level through POWHEG [130] and undergo parton showering (as described in Sec. 1.3.2) via HERWIG [41]. The initial generation phase utilizes POWHEG, which manages various leptoquark flavor combinations characterized by distinct masses and coupling strengths. The simulation output is formatted as Les Houches Event (LHE) files, where a number uniquely identifies each particle by the Particle Data Group (PDG) particle numbering convention. The models used and lepton PDFs (LUXlep-NNPDF31 nlo_as_0118_luxqed) are referenced in [46, 44]. At present, there is no established framework for vector leptoquark simulations for this lepton-induced process, so no vector leptoquark model can be probe from a single production lepton-induced process. Each particle is described by a four-vector representing its physical properties and an associated mass value. After LHE file generation, HERWIG takes over for parton showering, furnishing a comprehensive depiction of the particle process. HERWIG's parton shower capabilities extend to having leptons in the initial stage of the process, such as the ones under investigation in this study. These output files are subsequently integrated into the full CMS simulation and reconstruction pipeline via Geant4 [115] to be utilized as a signal model for a CMS analysis.
We study two different signal hypotheses: a LQ coupling to u quark and b quark. Datasets are created for b ($b\mu$) and u ($u\mu$) couplings at different mass and λ values. Production cross-sections with solely b couplings are roughly an order of magnitude lower than those for u couplings; this discrepancy is attributed to the production side involving sea quarks rather than valence quarks, due to a higher probability for valence quark to carry a higher fraction of the proton momentum as shown in Figure 1.6. Mass values in the light and heavy quark scenarios are 700, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, and 5000 GeV. The couplings generated for each mass are 0.1,0.5,1.0,1.5,2.0. In order to probe intermediate masses and couplings an

interpolation procedure is used, as described later in Section 6.3.1.
Since the experimental signature and efficiency are similar for all light quarks (u,d,s), the datasets are generated with only u couplings. Therefore, with the u datasets and rescaling to the different cross sections, it is, in principle, possible to interpret the result with different quark hypotheses. Due to the entirely different experimental signature, the top quark case is not studied in this dissertation.

3.1.1 Leptoquark mass distributions

The LQ mass distribution is a convolution of the PDF (manifesting itself as a tail in the distributions) of the initial quark and lepton (Equation 1.12) with the cross-section of a hard-scattering process (manifesting itself as a peak) described with a Breit-Wigner, which is used to parameterize the resonant peak [153]. The Breit-Wigner is parametrized as:

$$A(s) \approx \frac{\alpha}{M_{BW}^2 - s - iM_{BW}\Gamma_{BW}}, \tag{3.1}$$

where α contains the information on the coupling of the initial and final state, M_{BW} is the mass of the resonance, Γ_{BW} is the width of the resonance and s is the center of mass energy.
Narrow resonances with different widths characterize the LQ mass distribution depending on the coupling and mass hypotheses. The absolute width (Γ) for a scalar LQ is described by Equation 1.19, from which we can compute the relative intrinsic width:

$$\frac{\Gamma}{M_{LQ}} = \frac{\lambda^2}{16\pi}, \tag{3.2}$$

The relative intrinsic width for different coupling hypotheses is reported in Table 3.1. Figure 3.1 shows the LQ mass distribution for several mass hypotheses, and, as expected, the resonance becomes wider with the increase of the LQ mass hypotheses. The tails at low mass are due to the convolution of the PDF with a Breit-Wigner (parametrizing the cross-section of the hard scattering process), as explained before. Fig 3.2 shows the invariant mass distribution at the generator level for three coupling hypotheses and $M_{LQ} = 4$ TeV. The tails at low mass for higher couplings are again due to the convolution of the PDF with a Breit-Wigner (parametrizing the cross-section of the hard scattering process), as explained before. This becomes more evident for the heavy quark hypothesis where, due to a higher PDF with a smaller proton momentum fraction for the b quark with respect to the u quark in the proton, the tails are more enhanced than the light quark.

3.2 Kinematic distributions

This section will describe the kinematics of the particles associated with the LQ decay process. Due to its unique experimental signature, the LQ decays into a lepton and a quark (which will be reconstructed as a jet, as explained in Sec 4.2). We denote the momenta of the decay products, the muon and the quark, as p_μ and p_q, respectively. Since M_{LQ} is at the TeV scale, the final state will be composed of a high-p_T quark

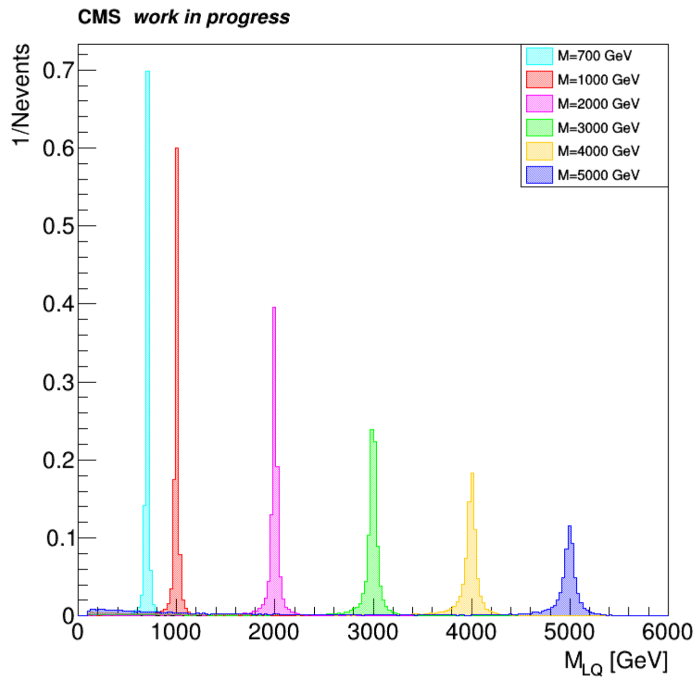


Figure 3.1. LQ mass distribution for different M_{LQ} and $\lambda_{u\mu} = 1$.

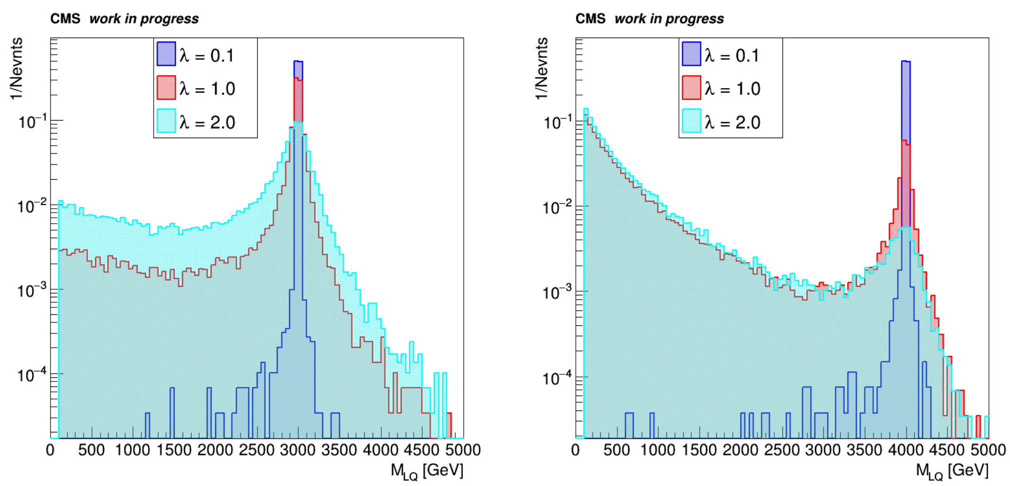


Figure 3.2. Invariant mass distribution of the muon-jet system at generator level for different couplings hypotheses and $M_{\mu jet} = 4$TeV for a LQ produced by a light quark (heavy quark) on the left (right).

λ	Γ/M
0.1	0.2%
0.5	0.5%
1.0	2%
1.5	4.5%
2.0	8%

Table 3.1. The ratio of width to mass as a function of the coupling constant λ.

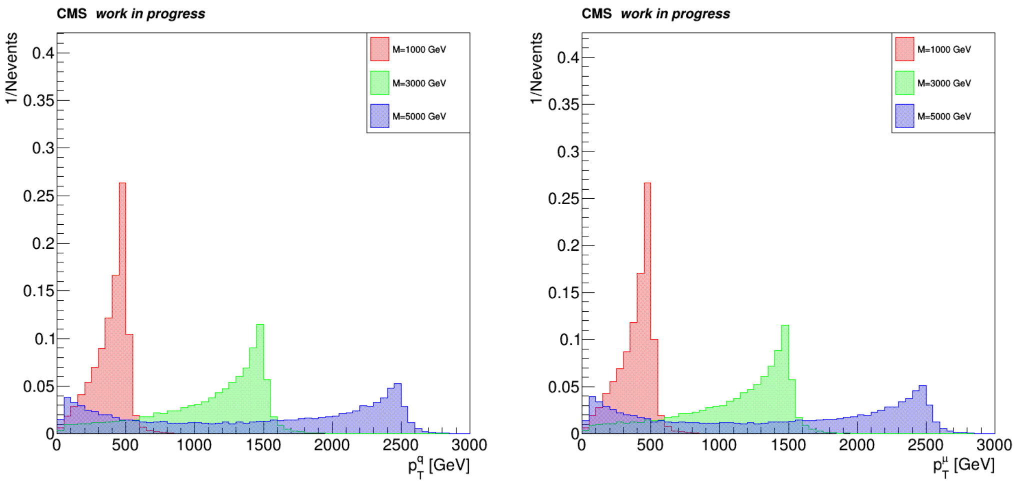


Figure 3.3. Transverse momentum distribution of leading jet (muon) for different M_{LQ} and $\lambda_{u\mu} = 1$ on the left (right). The tails at low p_T are due to the convolution of the PDF with a Breit-Wigner (parametrizing the cross-section of the hard scattering process) as for the invariant muon+jet mass distribution.

and muon. Figure 3.3 shows the distribution of the quark and muon coming from the LQ. The distribution peaks at $\frac{M_{LQ}}{2}$, with some tails at lower p_T. Given the LQ's substantial mass, it is typically produced almost at rest and with small transverse momentum. The lepton and the jet are expected to have substantial azimuthal angular separation. Figure 3.4 displays the $\Delta\phi_{\mu q}$ distributions for various mass hypotheses. The distribution has a peak at around π, which indicates a back-to-back production of the two objects. Due to the kinematic of a two-body decay, the module of the transverse momentum of the quark is approximately the same as the muon. The ratio between the muon p_T^μ and the jet p_T^{jet} has to peaks around at one, with some lower tails, as can be seen in Fig 3.5.

In the experimental signature, we do not expect neutrinos in the final state (since we are assuming $\beta = 1$); therefore, if we could reconstruct perfectly the event, no missing transverse energy (MET), defined in the next chapter, would be expected. It is also interesting to study the properties of the second muon produced from the initial state photon (see Figure 1.13 with Feynman diagram), which usually has a lower p_T than the one coming from the LQ, as illustrated in Figure 3.6 (left). Moreover, this low-p_T muon usually has a greater η. Figure 3.6 (right) shows the distributions of η for both muons. Due to the large η, the second muon may often

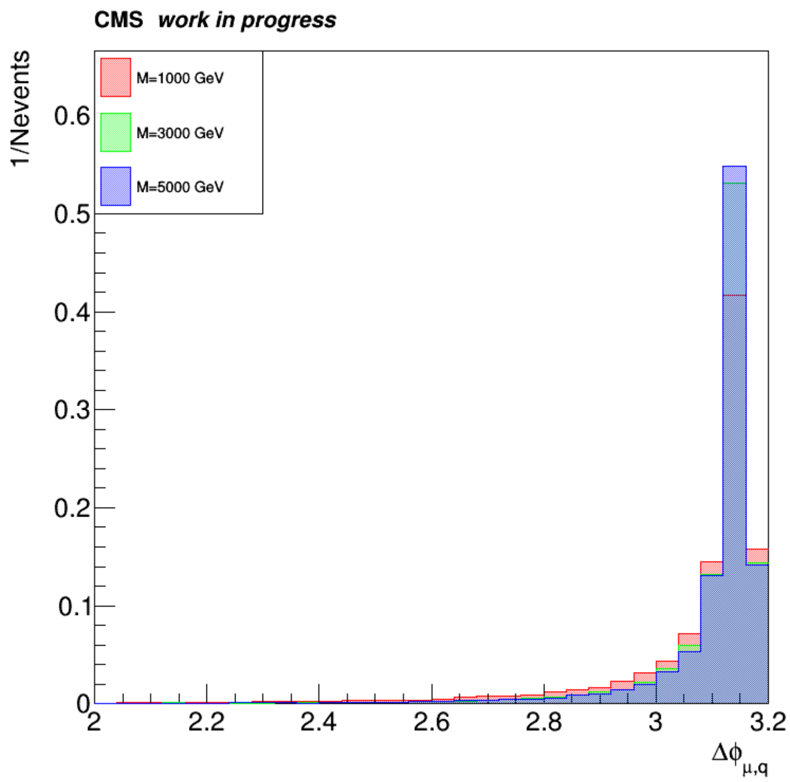


Figure 3.4. Distribution of the angle $\Delta\phi_{\mu q}$ between the leading muon and the leading jet for different M_{LQ} and $\lambda_{u\mu} = 1$.

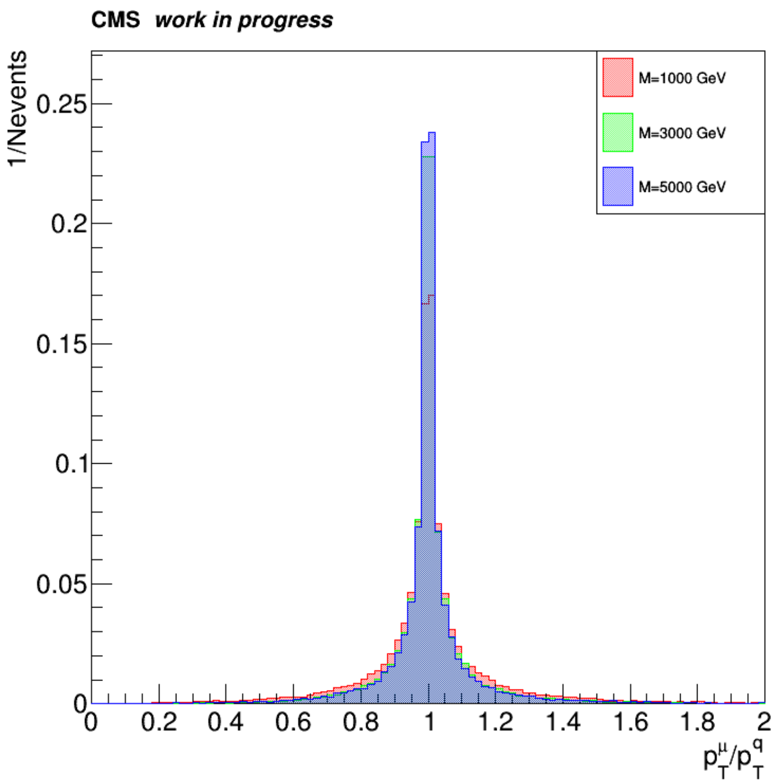


Figure 3.5. Distribution of the ratio of the transverse momentum of the leading muon and leading jet for different M_{LQ} and $\lambda_{u\mu} = 1$.

be out of the tracker acceptance ($|\eta| < 2.4$) and thus it is not detected.

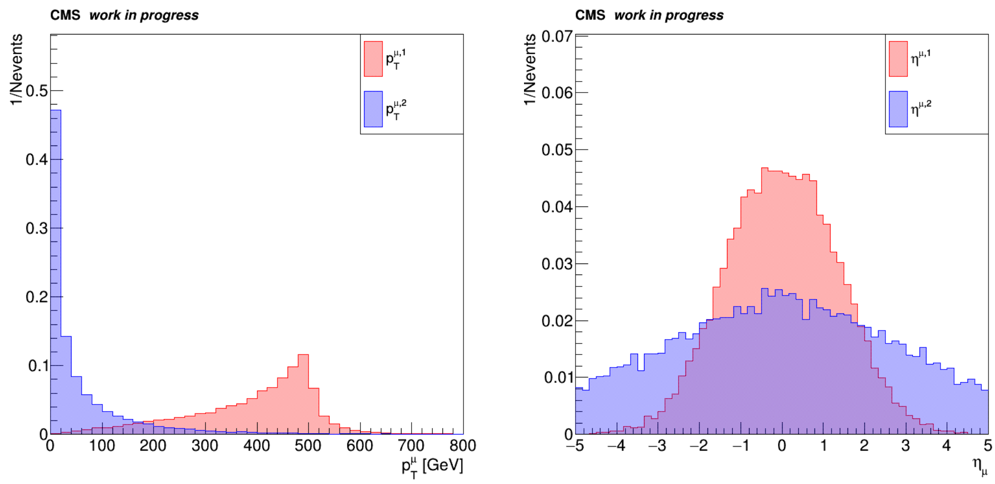


Figure 3.6. Transverse momentum distributions (η) for leading muon and second muon for $M_{LQ} = 1$ TeV and $\lambda_{u\mu} = 1$ on the left (right).

3.2.1 Signal cross sections

LQ can, in principle, couple to both valence or sea quarks depending on the specific LQ model considered. The cross-section for valence quarks, i.e., the quarks contributing to the quantum numbers of a hadron, is greater than the one for sea quarks, as reported in Table 3.2. For the sea quarks, the cross-section decreases increasing the mass of the quark ($m_s > m_c > m_b$). Both these cross-section properties are due

$m_{LQ} = 3$ TeV, $\lambda = 1$	σ $[fb]$ $\pm 5\%$ [44]
$u\mu$	0.572
$d\mu$	0.242
$s\mu$	0.0419
$c\mu$	0.0251
$b\mu$	0.0123

Table 3.2. Cross section depending on LQ-quark coupling.

to the PDF of the quarks inside the proton (as can be seen in Figure 1.6).
Fig 3.7 left (right) shows the value of the cross-section as a function of the mass (coupling) of the resonance, and, as expected, it decreases with the mass and increases with the coupling.

The total cross-section for the lepton initiated process for $M_{LQ} \sim 1$ TeV is ~ 2 times greater than the one for the gluon-initiated process. Even if the PDF for a gluon in the initial state (Figure 1.6) is approximately 1000 to 20000 times larger than the lepton PDF (Figure 1.12), the enhancement due to the resonant production of a LQ with respect to an off-shell quark is such that the lepton-induced process is more probable. For higher LQ mass hypotheses, as can be seen in Table 3.3,

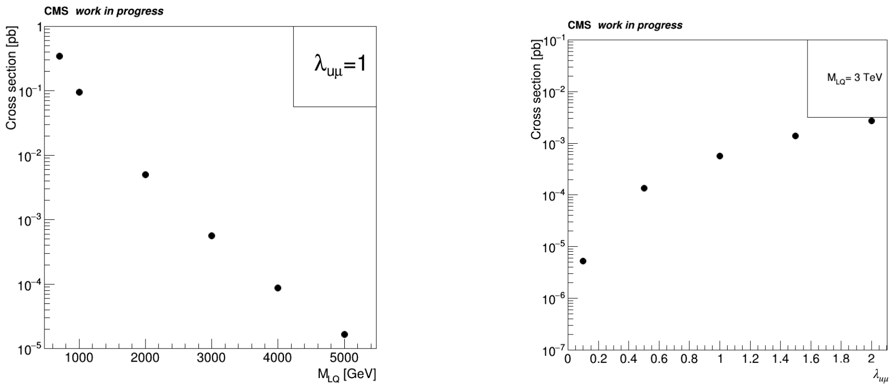


Figure 3.7. Cross sections (in pb) of for u-μ LQ couplings as function of M_{LQ} ($\lambda_{u\mu}$) on the left (right).

the enhancement due to the resonant production becomes greater, such that for $M_{LQ} = 3$ TeV, the lepton-induced production cross-section is $\sim$ 25 time bigger than the gluon induced one. The results are the same considering different couplings or leptons in the initial state [44].

	$\frac{\sigma_{\text{lepton}}}{\sigma_{\text{gluon}}}$		
m_{LQ}[GeV]	$\lambda = 0.1$	$\lambda = 0.5$	$\lambda = 1.0$
1000	1.7	1.7	1.7
2000	7.6	7.6	7.6
3000	26.0	26.3	25.6

Table 3.3. Cross section ratio of lepton-initiated and gluon-initiated single-leptoquark production (uμ) production for different couplings and masses.

Chapter 4

Event reconstruction

This chapter describes the reconstruction of physics objects, focusing on the reconstruction of high-p_T muons and jets.

4.1 The Particle Flow algorithm

The CMS detector employs a comprehensive particle identification strategy using the Particle Flow (PF) [70] algorithm to reconstruct and identify all the stable particles produced in a proton-proton collision. It runs on an event basis to match groups of tracks and deposits and identify the particles that have caused them. To do so, it mainly relies on an efficient and pure track reconstruction, clustering algorithms that are able to distinguish overlapping showers, and a linking algorithm to build relations among different deposits.

At first, the tracks of charged particles are reconstructed from the points of their trajectory recorded by the silicon tracker. The tracker also combines the tracks information to reconstruct the points (vertices) where the pp interactions took place. The vertex with the most high-energetic tracks associated is called the primary (or leading) vertex. A 3.8 T magnetic field generated by the CMS solenoidal magnet bends the particle trajectories. The bending of the tracks allows precise measurement

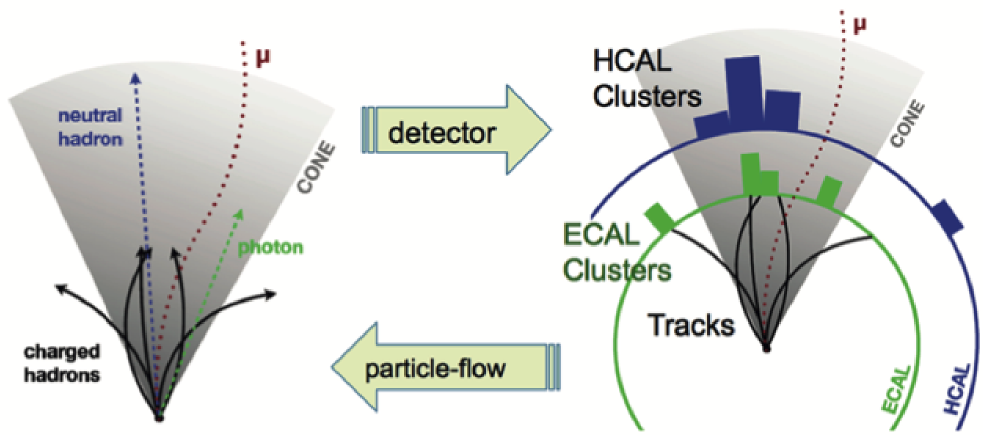


Figure 4.1. Schematic view of how Particle Flow reconstructs the various candidates [157].

of the p_T of charged particles. The energies of photons and electrons are measured by ECAL (as described in Section 2.2.3), with excellent resolution, while the energy of hadrons is measured by the CMS hadronic calorimeter HCAL (as described in Section 2.2.4), with a contribution of ECAL for charged hadrons. Muons, besides the neutrinos, are the only particles that escape the calorimeters and their track is reconstructed in the muon chambers (as described in Section 2.2.5).
For each event, the algorithm builds the list of reconstructed objects (blocks) with their relations and kinematic parameters. This list represents a global event description, and it allows the rejection of soft contributions and the identification of relations among particles.
The reconstruction process of the PF algorithm incorporates several types of blocks, each corresponding to different particle candidates:

- Electron candidates are extrapolated from correlations between charged tracks and one or more ECAL clusters.
- Charged hadron candidates emerge from charged tracks aligned with calorimeter clusters, either ECAL or HCAL, that do not meet the electron candidate criteria.
- Photon candidates are discerned from ECAL energy deposits that lack an associated charged track.
- Candidates for hadronic particles in the forward regions are deduced from energy deposits within the hadronic forward calorimeters. The classification as hadronic or electromagnetic is based on the depth profile of the energy dispersion.
- Muon candidates are deduced with high precision by integrating information from the tracking system and the muon chambers.

The compilation of the Particle Flow (PF) Candidate list is designed to reflect the Particle Flow algorithm's analysis of a proton-proton collision event within the CMS framework. This list is an effort to reproduce the actual particle composition of the event as accurately as possible. Comprehensive data on the momentum and energy of the PF Candidates are meticulously recorded and made available for analysis. Jet reconstruction within CMS is achieved by combining PF Candidates through a variety of clustering methodologies, discussed in the next section.

4.2 Jet reconstruction

Jets observed by the detectors are not physical objects but are the experimental signatures of a quark and a gluon undergone through a hadronization process, as described before. Therefore, the process of defining jets introduces a degree of subjectivity as different algorithms employed to cluster PF Candidates into jets can influence the inferred jet attributes, such as momentum, energy, or spatial extent.
The prototypical objective is for the jet's four-momentum to replicate that of the originating parton. Although this is only true to some approximation, contingent

upon theoretical and empirical guidelines that shape algorithm development, aiming to align jet characteristics with those of the progenitor parton as closely as feasible. Crucial for any robust jet clustering algorithm are two characteristics: infrared and collinear (IRC) insensitivity. An algorithm is infrared insensitive if the incidental emission of soft gluons during hadronization does not perturb the jet clustering outcome. Conversely, an algorithm exhibits collinear insensitivity if it remains unaltered when a parton bifurcates into two (e.g., a gluon branching into a quark pair).
Sequential recombination algorithms constitute a comprehensive suite of IRC-safe methodologies for jet clustering. This family encompasses the k_t [109], Cambridge/Aachen [106], and anti-k_t algorithms [47]. These algorithms typically agglomerate pairs of particles when a specifically defined metric, related to the transverse momentum of the particles and denoted as $k_{T,i}$, falls below a predetermined threshold. Distances are measured in two fashions: one between pairs of particles i and j and another between a particle and the beamline.

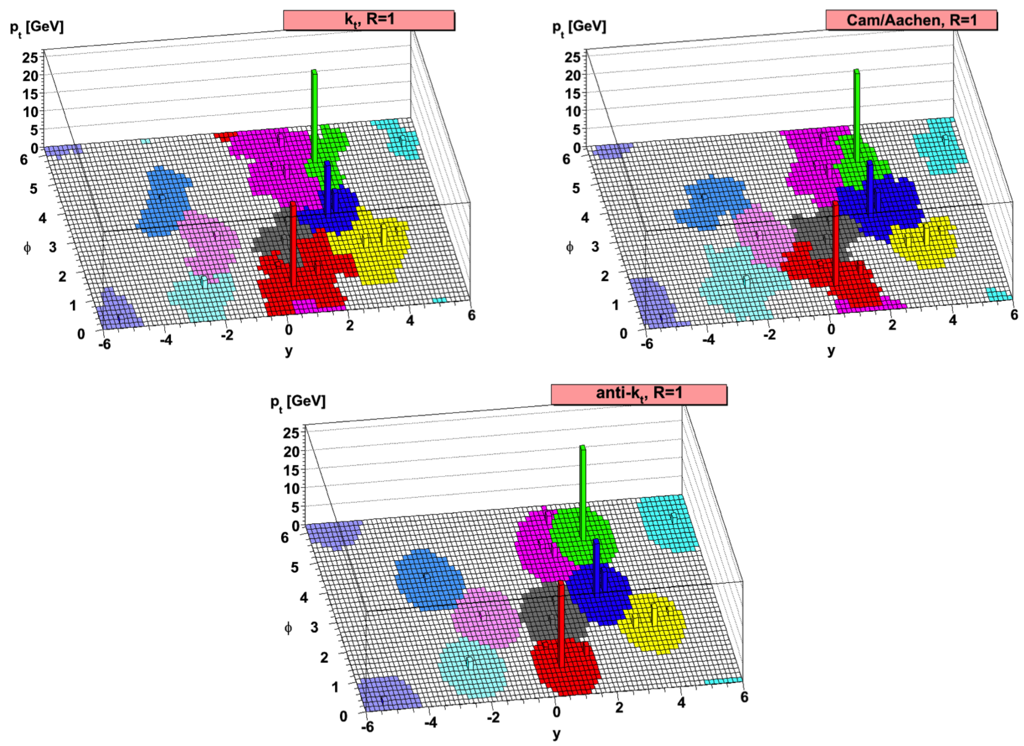


Figure 4.2. Illustration of different jet clustering algorithms [47].

The distances are articulated as follows:

$$d_{ij} = \min(k_{T,i}^{2p}, k_{T,j}^{2p}) \frac{\Delta R_{ij}^2}{R^2}, \quad d_{iB} = k_{T,i}^{2p}, \tag{4.1}$$

where $\Delta R_{ij} = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, with $k_{T,i}$, y_i, and ϕ_i representing the transverse momentum, rapidity, and azimuth of particle i, respectively. Parameters R and p

are algorithm-specific constants.
The operational steps for sequential recombination algorithms are:

- Compute the distances d_{iB} and d_{ij} for each particle and pair of particles.
- The minimum of all d_{iB} and d_{ij} is designated as $d_{\min}$.
- If $d_{\min}$ is a d_{ij}, the particles i, j are merged into a single protojet by summing their four-momenta.
- If $d_{\min}$ is a d_{iB}, the particle is considered non-mergeable and is excised from the list, with the remaining particles reevaluated.
- Iterate the algorithm, using the resulting protojets as new inputs until only non-mergeable entities remain.

The output from jet clustering algorithms is an ensemble of jets characterized by their four-momenta, as determined through the sequential recombination process. The parameter R known as the distance parameter, correlates with the jet's spatial expansion. A larger R value decreases the distance d_{ij} relative to d_{iB}, leading to the amalgamation of more particles into a jet. The radial extent of the jet in the $\eta\phi$-space is analogous to the R parameter utilized in the clustering process.
The parameter p dictates the clustering algorithm's nature: $p = 1$ yields the k_T algorithm; $p = 0$ aligns with the Cambridge/Aachen algorithm; and $p = -1$ results in the anti-k_T algorithm. In the k_T algorithm, low-energy particles are clustered initially, followed by higher-energy ones. This order is inverted in the anti-k_T algorithm. The Cambridge-Aachen algorithm, with $p = 0$, operates purely geometrically. Figure 4.2 demonstrates the distinctive behavior of these algorithms, with the anti-k_T algorithm typically producing jets with more defined boundaries.
CMS analyses usually utilize the anti-k_T algorithm with parameters $R = 0.4$ (AK4) or $R = 0.8$ (AK8). The PF algorithm retains all jet constituent properties, permitting jet reclustering with different algorithms or distance parameters. The FASTJET package is commonly employed for k_T algorithm-based clustering [48].
For this analysis, we expect a high-p_T jet in the final state, which is the experimental signature for the quark decaying from the LQ. We use standard AK4 jets, with $R = 0.4$ and p_T>100 GeV. So, any jets that spatially overlap with isolated and fully identified electron or muon candidates within $\Delta R < 0.4$ are excluded from consideration [140, 134]. The typical energy resolution for these objects is 5% for jet with $p_T > 100$ GeV. Jet energy scales have uncertainties of the order of 1 - 2% [68].

4.2.1 Jet Energy Calibration

In this section, we will discuss the jet energy corrections (JEC), which are a set of tools to correct the jet energy for the many effects that modify it, first above all the non-linearity of the response of CMS calorimeters.
The CMS collaboration adopted a factorized solution to the problem of JECs, where each level of correction takes care of a different effect. Each level consists of a scaling of the jet four-momentum with a scale factor (correction) which depends on jets p_T and η. The levels of correction are applied sequentially (the output of each step is

the input to the next) and with a fixed order, following the scheme shown in Figure 4.3:

- The first step in jet calibration is to estimate and subtract the energy not associated with the hard scattering interaction, i.e. coming from pileup (PU) and noise. Section 4.5 will describe the PUPPI method, adopted for the mitigation of pileup effects;
- The second step, the simulated response corrections, accounts for detector non-uniformity and is derived from PU corrected jets.
- The last step, the residual corrections for data, corrects the residual difference between data and simulation after the application of the other corrections. Figure 4.4 shows the ratio between residual correction evaluated on data/MC for collision data as a function of the p_T of the jets.

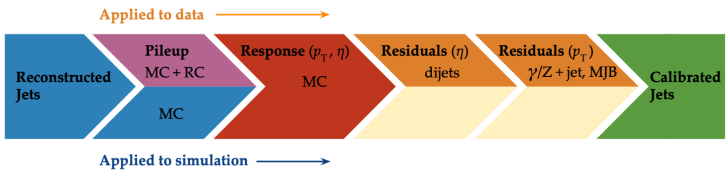


Figure 4.3. Consecutive stages of jet energy corrections, for data (upper row) and simulation (lower row) [88].

For further details on the JEC evaluation methods and on the performances on 13 TeV data the reader can refer to [80, 69].

4.3 Jet b-tagging

To identify the jets from a b quark, CMS uses the DeepJet algorithm [72]. DeepJet uses a deep machine learning algorithm, and the discriminating variables exploit the fact that long living particles, such as B-hadrons, travel a considerable distance from the primary vertex before their decay happens. The Impact Parameter (IP), Figure 4.5, is the variable used to define the distance between the primary and secondary vertices. The typical value for the B-hadrons corresponds to $c\tau \sim 450$ μm that, in CMS, can be measured with precision between 30 μm and hundreds μm depending on the p_T of the hadron.

The AK4 jets are considered to be coming from b-quark if they pass a given threshold on the value of DeepJet [95], which is defined based on the selection efficiency of b-originated jets and the mistagging efficiency of light quark-originated jets. The chosen working point is "Medium", meaning it provides a misidentification probability of less than 1% [95].

4.4 Muon reconstruction

In the CMS detector, the local reconstruction of muon tracks utilizes data from individual muon chambers — such as RPC, CSC, or DT — to detect the trajectory

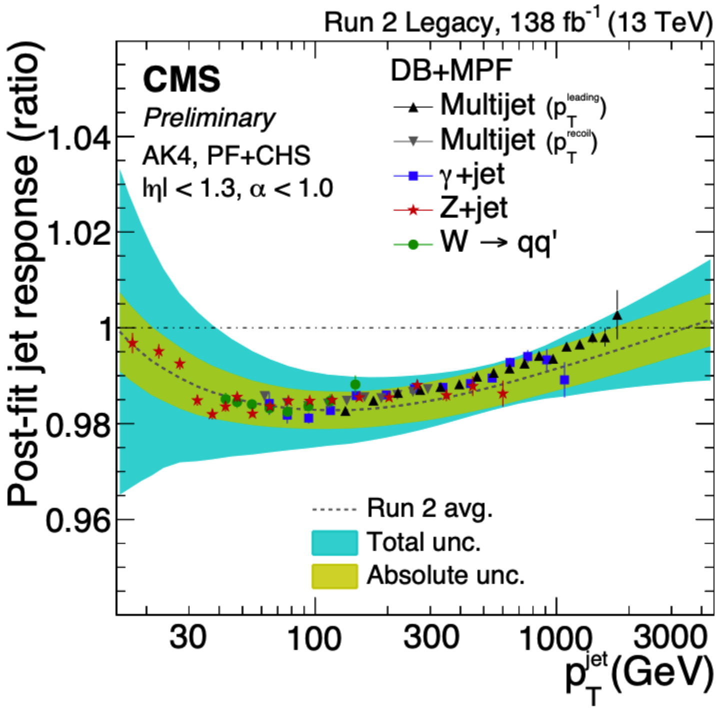


Figure 4.4. Global fit to the ratio of the jet response (line with yellow band) using the contribution of different SM process for full Run2 data [10].

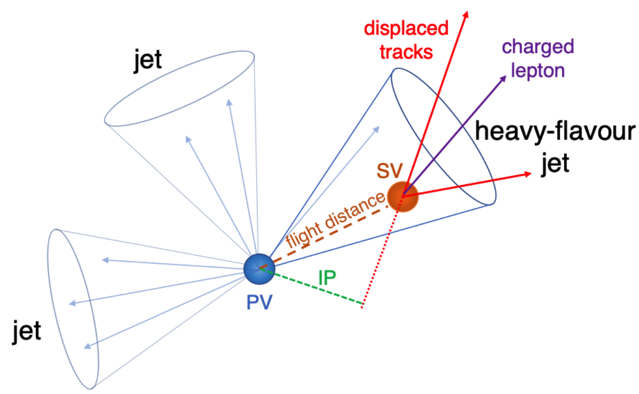


Figure 4.5. Illustration of a heavy-flavour jet with a secondary vertex (SV) from the decay of a b or c hadron resulting in charged-particle tracks (including possibly a soft lepton) that are displaced with respect to the primary interaction vertex (PV), and hence with a large impact parameter (IP) value [72].

of muons. When traversed by muons or other charged particles, the chambers experience ionization of the contained gas, generating electrical signals along the detector's wires and strips [135]. These electrical pulses, routinely referred to as "hits", are spatially precise and are interpreted using an array of algorithms tailored to the specific technology of the detector. Each hit's exact position is deciphered from these signals, enabling the reconstruction of particle paths.

In CMS's established protocol for reconstructing events from proton-proton interactions, tracks are initially determined separately in the inner tracker and the muon detectors. These are referred to as tracker tracks and standalone muon tracks, respectively, and subsequently merged to enhance muon track reconstruction [167]. Tracker tracks are built using an iterative approach, running a sequence of tracking algorithms, each with slightly different logic. After each iteration step, hits that have been associated with reconstructed tracks are removed from the set of input hits to be used in the following step. This approach maintains high performance and reduces processing time[167]. Standalone-muon trajectories are deduced by harnessing data from various muon subdetectors, utilizing a Kalman filter[132] approach to assimilate all relevant CSC, DT, and RPC data along the path of a muon. The initiation of this process relies on seed clusters composed of DT or CSC segment groups.

Conversely, tracker muon paths are generated in an "inside-out" fashion, starting from the tracker tracks and extending towards the muon detectors, seeking alignment with DT or CSC segments. A tracker track is deemed a muon track if it aligns with at least one muon segment after extrapolation, contingent on the track having sufficient transverse momentum ($p_T > 2.5$ GeV). This matching utilizes local coordinates, with specific criteria for the minimum allowable distance or uncertainty ratio between the track and segment [167].

Global muon tracks are the product of an "outside-in" strategy, which matches standalone muon tracks to tracker tracks. This amalgamation relies on the Kalman filter to assess the compatibility of the standalone muon and tracker tracks based on the kinematics of the muon trajectory extrapolated from the muon system to the tracker and vice versa.

Tracker muons have high efficiency in regions of the CMS detector with less instrumentation (for routing of detector services), especially for muons with lower transverse momentum (p_T). Tracker muons generally correspond to segments only in the detector's innermost station, enhancing the likelihood of accurate identification by mitigating the risk of misidentifying hadronic shower remnants as muons. The global muon reconstruction framework, utilizing standalone muon track data, achieves substantial efficiency for muons traversing multiple muon stations, thereby reducing misidentification rates relative to tracker muons. When the data from both the inner tracker and muon systems are integrated, the measurement accuracy of global muons' p_T is notably enhanced. This improvement is most apparent for p_T values exceeding 200 GeV. Standalone-muon tracks, in contrast, typically exhibit poorer momentum resolution and greater inclusion of cosmic muons than their global or tracker counterparts [167].

Experimentally, one can delineate the principal variances between high-transverse momentum (high-p_T) and low-p_T muons. The p_T resolution of the reconstructed trajectory degrades as the muon momentum escalates. Within a section of orbit where the magnetic field B is nearly uniform, the p_T quantification is dependent on

B and the track's curvature radius R:

$$p_T[\text{GeV}] = |0.3B[\text{T}]R[\text{m}]|. \tag{4.2}$$

The magnetic field is precisely monitored and maintained at a uniform value of 3.8 T within the solenoid's tracking volume. The curvature radius R is computationally derived from the track's arc length L and the sagitta s of the track through the relationship:

$$R[\text{m}] \approx \frac{L[\text{m}]^2}{8s[\text{m}]}, \tag{4.3}$$

an approximation that holds when $L/R \ll 1$ [83]. Consistent assignment of arithmetic signs to R, s, and the charge q (in proton charge units) yields:

$$s[\text{m}] \approx \frac{(0.3B[\text{T}]L[\text{m}]^2)(q)}{8p_T[\text{GeV}]} = \frac{(0.3BL^2)}{8}\kappa, \tag{4.4}$$

where $\kappa = q/p_T$ is referred to as the (signed) curvature of the muon track.
As the p_T increases and the sagitta in the tracker decreases, the enhancement of p_T measurements is feasible by using the large BL^2 within both the tracker and the muon system. This remains valid as long as the transverse momentum is sufficiently high to prevent multiple Coulomb scattering in the calorimeters and steel flux-return yoke of the solenoid from impairing the precision of the measurement [83]. High-p_T muon trajectory reconstruction and momentum quantification necessitate the correlation of tracks reconstructed within the inner tracker and those within the muon system, which are spaced over three meters apart, forming a global track. Nonetheless, at TeV energy scales, the muon p_T resolution may be influenced by the alignment of the hits used for track reconstruction due to the diminished sagitta.
For a muon with substantial momentum traversing the magnet's steel flux-return yoke, the radiative energy losses, which encompass bremsstrahlung and the production of electron-positron pairs, become significant relative to ionization energy losses. The threshold energy for a muon in iron, designated as E_c^{iron}, where ionization energy losses are equal with radiative losses, is approximately 300 GeV [152]. Beyond this critical energy, as muons propagate through the steel interconnecting the muon subsystems, radiative losses predominate. This process leads to the generation of particle cascades, namely electromagnetic showers, potentially resulting in additional hits within the muon detection apparatus. The occurrence of such showers exerts a profound influence on muon detection performance metrics, including trigger efficiency, track reconstruction fidelity, and the precision of p_T measurements. The muons showering is predominantly a function of their total momentum rather than the transverse component, which is more frequently employed in physics analyses. For muons with longitudinal momentum components, possessing a transverse momentum (p_T) exceeding 200 GeV, it is possible for their energies to surpass the critical energy for iron, denoted as E_c^{iron} [83].
This presence of additional particles resulting from electromagnetic showers can lead to spurious signals within the muon detectors, resulting in extra reconstructed hits and segments. These may be mistakenly incorporated by the path reconstruction algorithm in lieu of the legitimate muon track fragments, or they could even render the muon track reconstruction unfeasible. The high-p_T requires careful treatment of

the information from the muon system. A set of specially developed TeV-muon track refits has been developed to address this issue: the “tracker-plus-first-muon-station” (TPFMS) fit [83], the “Picky” fit [83], and the “dynamic truncation” (DYT) fit [83]. The momentum assignment is finally performed by the “TuneP” algorithm, which chooses the best muon reconstruction among the tracker-only track, TPFMS, DYT, and Picky fits [83].

Muon Identification

An array of metrics and selection parameters has been established to calibrate the trade-off between reconstruction efficacy and signal purity in muon analysis. Variables include the goodness of track fit expressed as a chi-squared value (χ^2), the count of hits associated with each track within the inner tracker or muon detectors, and the alignment precision between the tracker and standalone muon tracks, particularly for global muons. The compatibility of muon segments is quantified by extending the tracker track into the muon detector system, considering the number of aligned segments across stations and the accuracy of this alignment in both position and orientation. The compatibility score spans from 0 to 1, where 1 denotes maximum compatibility.
A specialized algorithm designed to detect discontinuities dissects the tracker track at multiple points along its path. This kink-finding algorithm evaluates the two resultant tracks from each division point, with a high χ^2 suggesting that the tracks represent separate trajectories. Additionally, other factors derived from outside the reconstructed muon track are utilized, such as the congruence with the primary vertex, identified by the vertex with the maximal sum of p_T^2 of its constituents.
Employing these determinants, muons are classified into principal identification categories used in CMS physics analyses.
For this analysis, muon candidates must fulfill high transverse momentum (high-pT) identification criteria [73]. Specifically, these high-pT ID muons need to be categorized as global muons. The global muon track fitting should incorporate at least one hit from the muon chambers, and segments must be detected in at least two different chambers. The relative uncertainty on the best-fit track’s transverse momentum (p_T) should be below 30%. Additional requirements include constraints on the impact parameter close to the primary vertex: it must be less than 2 mm in the transverse plane and less than 5 mm in the longitudinal direction. The track should also include a minimum of one-pixel hits and traverse more than five layers of the tracker with hits. These stringent conditions enhance the precision of momentum measurements while minimizing potential contaminants like hadronic punch-through, cosmic ray muons, and in-flight decays from mesons. Reconstructed muons are additionally subjected to a condition on tracker-based relative isolation: $\frac{\sum_{p_T}(\text{tracker tracks from PV})}{p_T(muon)}$. This condition is evaluated within a cone centered around the muon track, defined by $\Delta R < \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.05$.
These criteria meets the requirements described described in Table 4.1.

Criterion	Specification
Global Muon Identification	≥ 1 hit in a chamber, segments in ≥ 2 different chambers
Transverse Momentum Uncertainty	p_T uncertainty $< 30\%$
Impact Parameter Constraints	Transverse < 2 mm, Longitudinal < 5 mm
Minimum Pixel Hits	≥ 1 pixel hit
Tracker Layer Traversal	> 5 layers with hits
Isolation Condition	$\frac{\sum p_T(\text{tracker tracks from PV})}{p_T(\text{muon})} < 0.05$

Table 4.1. Criteria for High-pT ID Muons Classification and Isolation.

4.5 Missing transverse energy reconstruction

According to momentum conservation, the net transverse momentum of all final-state particles from pp collisions should balance to zero. The missing transverse momentum (MET), denoted as p_T^{miss}, represents the discrepancy in momentum within the plane perpendicular to the colliding beams. The magnitude of this vector, p_T^{miss}, serves as a crucial observable for inferring the presence of non-interacting particles such as neutrinos originating from Standard Model (SM) processes involving W and Z boson decays. Additionally, MET is instrumental in the search for hypothetical dark matter candidates posited by theories extending the SM. The accurate reconstruction and interpretation of p_T^{miss} are contingent upon the resolution of the experimental apparatus, as it is affected by factors such as measurement errors, misidentification of particle types, detector imperfections and PU.
The MET is reconstructed using the PUPPI (PileUp Per Particle Identification) algorithm [77], an algorithm that removes charged particles with tracks not originating at the primary vertex and downweights neutral particles based on the probability that they originate from pileup. This probability is calculated considering the global pile-up density as well as locally adjacent charged particles of the hard process [42]. The PUPPI algorithm exploits tracking information, local particle distribution, and event pileup properties in order to assign an individual weight for each article (in our case, to each PF candidate). The weights are in the range of 0 to 1 and represent the degree to which PF candidates are likely to be produced from the leading vertex (LV), i.e. the primary collision of interest. When applying the PUPPI algorithm in object reconstructions, the momentum of each PF candidate is rescaled with its weight accordingly.
The PUPPI algorithm starts with charged PF Candidates, allocating a binary weight (either 1 or 0) based on the association of the candidate's track to the vertex. Charged PF candidates linked with the LV receive a weight of 1. If a charged PF candidate is unaffiliated with any vertex and its longitudinal impact parameter d_z is less than 0.3 cm, it is also granted a weight of 1. Otherwise, a weight of 0 is assigned.
For neutral PF candidates, the local shape descriptor α is defined according to the surrounding PF candidates and it is defined as:

$$\alpha_i = \log\left(\sum_{\substack{j\neq i,\\ \Delta R_{ij}<0.4}} \left(\frac{p_{T_j}}{\Delta R_{ij}}\right)^2\right) \quad \begin{cases} \text{for } |\eta_i| < 2.5, & j \text{ are charged PF candidates from LV} \\ \text{for } |\eta_i| \geq 2.5, & j \text{ are all kinds of reconstructed PF candidates} \end{cases} \tag{4.5}$$

where ΔR_{ij} is the angular distance in the η-ϕ space between the candidate i and other candidates j and is computed as $\Delta R_{ij} = \sqrt{(\Delta\eta_{ij})^2 + (\Delta\phi_{ij})^2}$. Figure 4.6 shows the α distribution for charged particles from the LV and the PU vertices and for neutral particles with $|\eta| < 2.5$. Then Charged PF candidates from pileup and

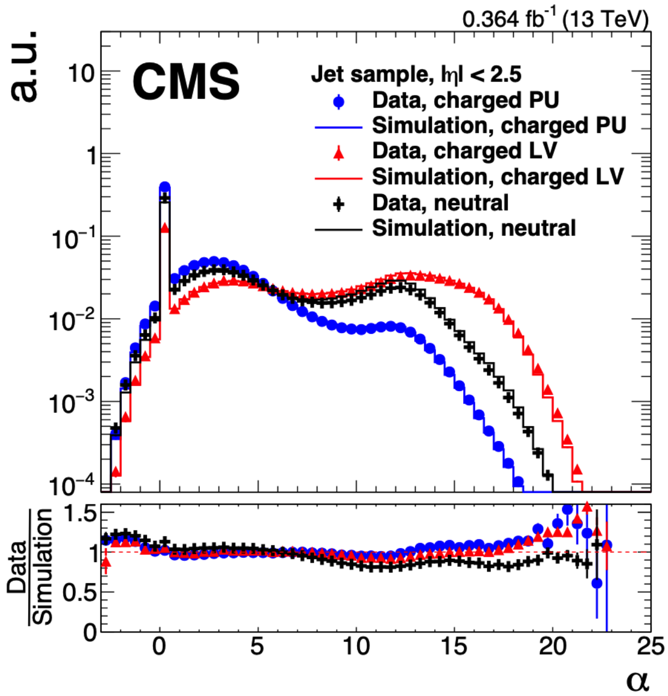


Figure 4.6. Data-to-simulation comparison for the α distribution in the jet sample for charged particles associated with the LV (red triangles), charged particles associated with PU vertices (blue circles), and neutral particles (black crosses) for $|\eta| < 2.5$ [56].

all neutral candidates are evaluated using a χ^2 like metric that measures their α value's consistency with the pileup profile of the event, enabling more effective pileup mitigation. For the i candidate a large χ_i^2 suggests a significant α_i, implying it is likely coming from the LV. In the final step, the signed χ^2 value is converted into a weight using the following relation:

$$w_i = F^{-1}_{\chi^2,\mathrm{NDF}=1}(\chi_i^2), \tag{4.6}$$

where $F_{\chi^2,\mathrm{NDF}=1}$ represents the cumulative distribution function (CDF) of a χ^2 distribution with one degree of freedom.
The PUPPI algorithm incorporates the individual PUPPI weights w_i for the PF candidates. It computes the p_T^{miss} by summing the weighted transverse momentum vectors of all PF candidates, resulting in the following expression:

$$\vec{p}_T^{\,\mathrm{miss}} = -\sum_i w_i \vec{p}_{T,i}, \tag{4.7}$$

where $\vec{p}_{T,i}$ is the transverse momentum vector of the i-th PF candidate, and w_i is the corresponding PUPPI weight. Figure 4.7 shows the PUPPI weight distribution for neutral particles for both jet and the PU samples.

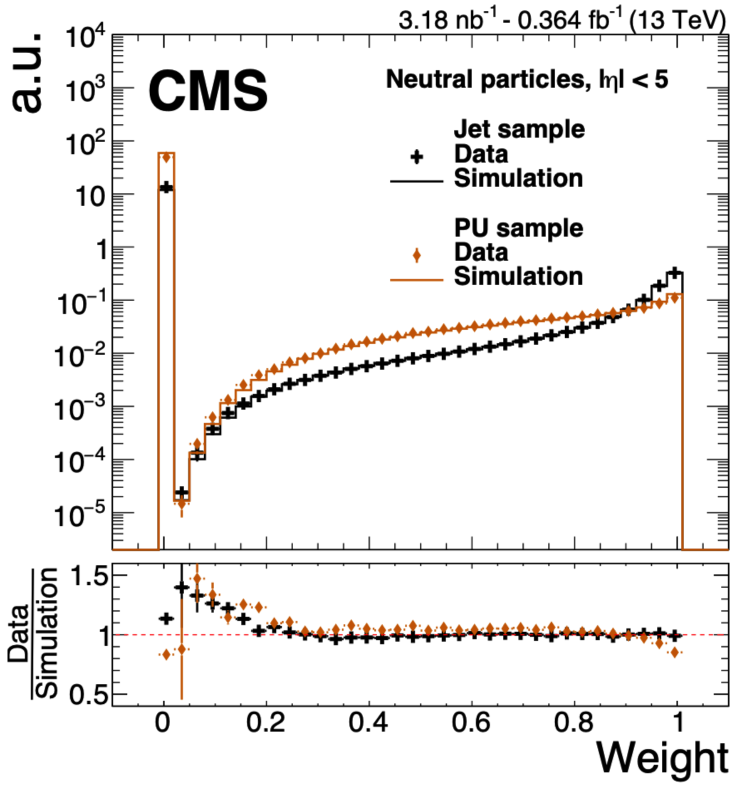


Figure 4.7. Data-to-simulation comparison for the PUPPI weight distribution for neutral particles in the jet sample (black crosses) and the PU sample (orange diamonds) [56].

4.6 Electron reconstruction

Electrons generate signals within the tracker layers in the CMS detector through ionization processes. They are then deviated by the solenoidal magnetic field arranged orthogonally to the plane, leading to the generation of electromagnetic showers upon entering the ECAL, where these showers are ultimately absorbed. Consequently, two subsystems of the CMS detector are essential for reconstructing electron events: the tracker, which measures transverse momentum, and the ECAL, which is responsible for energy measurements.
As an electron moves through the CMS tracker and is influenced by the magnetic field, its trajectory, governed by the Lorentz force, allows for determining the particle's transverse momentum. Subsequently, the electron enters the electromagnetic calorimeter. It is completely absorbed via the development of an electromagnetic shower within ECAL lead-tungstate crystals, providing an energy measurement for the impinging particle. The total electron reconstruction efficiency is reported in η bins and for different p_T ranges for simulated Drell-Yan processes samples and 2017 data in Figure 4.8.

Several algorithms are used sequentially to reconstruct electron candidates from the

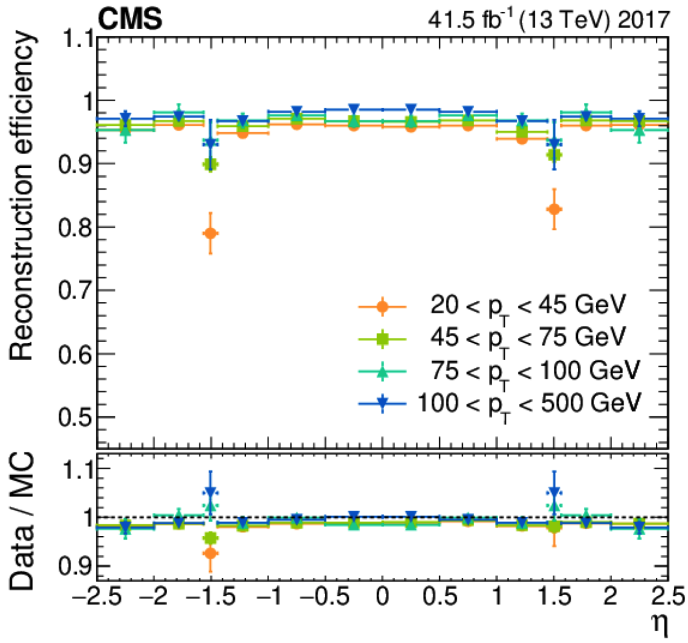


Figure 4.8. Electron reconstruction efficiency versus η in data (upper panel) and data-to-simulation efficiency ratios (lower panel) for the 2017 data-taking period. The vertical bars on the markers represent the combined statistical and systematic uncertainties. The region $1.44 < |\eta| < 1.57$ corresponds to the transition between the barrel and endcap regions of ECAL and is not considered in physics analyses [86].

measurement of scintillation light in each single crystal in the ECAL detector [87]. The first step builds the ECAL *Rechits*, the measurement of the amount of energy deposited in each crystal at each bunch crossing. Following this, the PFClustering algorithm develops rudimentary energy clusters by identifying and combining the

crystals that exhibit peak energy levels, referred to as "seeds", with their adjoining crystals. An electron, while passing through the Pixel and Tracker detectors, emits bremsstrahlung photons that will leave a trace of small energy clusters in the ECAL detector near the main impact point. To incorporate the energy contributions from secondary clusters, thus enhancing the electron reconstruction, an additional algorithm, named SuperClustering, has been developed.
The SuperCluster (SC) candidate is exclusively derived from local ECAL information. In this framework, electrons are delineated by primary charged-particle tracks converging to ECAL SCs. The Gaussian Sum Filter (GSF) algorithm, a refined tracking approach, is applied to electrons to accurately parameterize their tracks by considering the alterations induced by bremsstrahlung photon emissions [8]. Additional ECAL clusters can be included in the SC if they are compatible with secondary emissions and photon conversions from the main electron track. The final object is called refined SC and is the baseline for the formation of electron candidates. Following this foundational step, supplementary criteria are employed to loosely identify isolated electrons and photons. An energy regression algorithm, which operates on a semi-parametric Boosted Decision Tree (BDT) framework, is utilized to integrate characteristics of the tracks and the ECAL energy deposits. For an exhaustive exposition of electron reconstruction processes, one is directed to consult the specified reference [87].

Electron Identification

In the CMS experiment, robust strategies for electron identification are paramount to achieve high selection purity, particularly imperative for mitigating background interferences.
Criteria tailored for electron identification within CMS target candidates presenting a p_T greater than 20 GeV, bifurcated into cut-based and multivariate methodological frameworks. The former encompasses a sequence of defined thresholds over an array of identification variables, embracing supercluster-to-track matching parameters, the energy deposition ratio across the hadronic and electromagnetic calorimeters, and isolation measures.
Particle isolation is evaluated by delineating a cone around the particle's trajectory, characterized by a radius ΔR, generally set to 0.3 or 0.4, within which the total particle energy is aggregated. The isolation ratio I/E_T quantifies how isolated the particle is within the detector. These variables are particularly productive in electron identification. They are designed to differentiate authentic electron events from those stemming from alternative sources, such as pion misidentification. In high-energy particle interactions, particularly those involving hard QCD processes, pions are predominantly generated within dense jets and are consequently seldom found in isolation. The abundance of jet events in such processes implies that the occurrence of an isolated pion capable of depositing energy in the ECAL is notably rare. Despite the rarity, it's crucial to account for the fact that after implementing selective criteria, a significant count of these isolated pion events may persist, warranting consideration in the analysis.
In this analysis, we do not expect electrons in the final state since we are considering LQs couplings with muons. Therefore, we need to identify electrons to veto events

containing them properly. The Tight working point [87] for electron identification was chosen to identify electrons and veto their presence in this analysis. The requirements for the tight working point in the barrel region are shown in Table 4.2.

Variable	Barrel
$\sigma_{i\eta i\eta}$	< 0.010
$\lvert\Delta\eta_{\text{seed}}\rvert$	< 0.025
$\lvert\Delta\phi_{\text{in}}\rvert$	< 0.022 rad
H/E	$0.026 + \frac{1.15}{E_{\text{SC}}} + \frac{0.0324\rho}{E_{\text{SC}}}$
I_{combined}/E_T	$< 0.029 + \frac{0.51\ \text{GeV}}{E_T}$
$\lvert\frac{1}{E} - \frac{1}{p}\rvert$	$< 0.016\ \text{GeV}^{-1}$
Number of missing hits	≤ 1
Pass conversion veto	Yes

Table 4.2. Selection criteria for the Barrel category.

Chapter 5

Data samples and event selection

5.1 Dataset and trigger selection

The analysis presented in this thesis uses pp collision data at a center-of-mass energy of 13 TeV collected with the CMS detector at the LHC in 2016, 2017, and 2018, corresponding to an integrated luminosity of 138 fb^{-1}.
We search for a signal in a final state with at least a high-p_T muon and a high-p_T jet almost back-to-back in the transverse plane. These events are collected in the SingleMuon dataset, which is defined as the OR of different triggers, requiring that at least a muon in the final state with a $p_T > 50$ GeV is present. The muon with the highest energy in the event is required to have a p_T greater than 55 GeV, ensuring a 100% trigger efficiency.
The SingleMuon dataset is divided into different datasets, corresponding to 2 datasets for the first year of data-taking (2016), 1 for 2017, and 1 for 2018. Operational parameters at the LHC exhibited variations across the 2016, 2017, and 2018 dataset years. The data taken during 2016 are split into two different datasets (called preVFP and postVFP from now on) based on the different algorithms used to reconstruct the tracks. Additionally, CMS subdetectors were calibrated periodically to address, for example, alterations due to radiation exposure over time. Corrections of physics objects [10, 76, 143, 142, 94] yield consistent distributions during the various datasets. This uniformity allows for the consolidation and collective analysis of the datasets as a singular entity, as done for the final results of this analysis (and explained later). The trigger paths used in this analysis are listed in Table 5.1

5.2 Monte Carlo simulation

5.2.1 Background samples

The relevant background for the search here presented arises from SM processes with at least one high-p_T jet and muon in the final state: W+Jets, DY+Jets, top production, and QCD multijet production. Another process, negligible with respect to the others also at selection, is the lepton-induced lepton+jet SM scattering. However, we generate signal samples using the same lepton PDF [46, 44] used for the signal generation, with a model provided by the author of [45].

Year	Trigger
2016	HLT_Mu50 HLT_TkMu50
2017-2018	HLT_Mu50 HLT_OldMu100 HLT_TkMu100

Table 5.1. List of triggers used in this analysis. The triggers with "Tk" in the name are using the tracker information only, while the others are using the global muon information (so muon chamber + tracker). From 2017-2018 the trigger has been improved with a third trigger path and a higher threshold for the p_T^{μ} for the tracker-only trigger to increase the trigger efficiency at high-p_T^{μ}.

In this analysis, the background is estimated from data as described later in Section 6.2. The MC simulation is only used to optimize the analysis selection and to describe the LQ signal.
The `MADGRAPH5 aMC@NLO` 2.6.5 event generator [17] is used for DY+jets and W+jets processes. These are simulated at the NLO with the FxFx jet matching and merging [113]. To increase the number of simulated events, samples binned in vector boson p_T are used.
The `MADGRAPH5 aMC@NLO` generator is also used for diboson production, while `POWHEG` 2.0 [147, 13, 14, 12, 114] is employed for $t\bar{t}$ and single top quark productions. For the QCD samples, the generator used is PYTHIA.
These generators are interfaced with `PYTHIA` 8.240 for parton showering and hadronization and for the leptons decay. The `PYTHIA` parameters that affect the description of the underlying event are set according to the CP5 tune [84]. The parton density function (PDF) set used is NNPDF 3.1 [36, 35, 34]. Simulated events include additional proton-proton interactions per bunch crossing modeled to match the profile observed in data. The simulations are processed through a `GEANT4` [115] simulation of the CMS detector. All the MC samples used are described in Table 5.2.

5.2.2 Leptoquark signal samples

The signal samples utilized are the ones described in Sec. 3.1. The samples generated are listed in Table 5.3 and 5.4 with the relative cross sections. We have generated 2 model benchmarks (light and heavy quarks) for several LQ masses (from 0.7 TeV to 5 TeV) and couplings (from 0.1 to 2.0).

5.3 Preselection

Events are selected in the final state with at least a muon and a jet. Then, the events are divided into two independent signal regions, depending on the number of reconstructed ($p_T^{\mu} > 7$ GeV and $|\eta| < 2.4$) muons (1 or 2). From now on, the leading muon (jet) is the reconstructed muon (jet) with the highest p_T in the event. The second muon will be called sub-leading muon and will be indicated with a 2 at

Sample	bin p_T [GeV]	**cross section [pb]**
QCD	80<p_T<120	2336000
QCD	120<p_T<170	407300
QCD	170<p_T<300	103500
QCD	300<p_T<470	6830
QCD	470<p_T<600	551.2
QCD	600<p_T<800	156.7
QCD	800<p_T<1000	26.25
QCD	1000<p_T<1400	7.465
QCD	1400<p_T<1800	0.6487
QCD	1800<p_T<2400	0.08734
QCD	2400<p_T<3200	0.005237
QCD	p_T>3200	0.0001352
W+Jets	inclusive	61526
W+Jets	100<p_T<250	697.7
W+Jets	250<p_T<400	25.18
W+Jets	400<p_T<600	3.177
W+Jets	p_T>600	0.4946
$t\bar{t}$ + jets - full leptonic	inclusive	88.29
$t\bar{t}$ + jets - hadronic	inclusive	377.96
$t\bar{t}$ + jets - semi leptonic	inclusive	365.34
Single top - Single t, t channel	inclusive	136.02
Single top - Single $\bar{t}$, t channel	inclusive	80.95
Single top - t + W^+	inclusive	35.85
Single top - $\bar{t}$ + W^-	inclusive	35.85
Single top - s channel	inclusive	3.376
Diboson - ZZ	inclusive	16.52
Diboson - WZ	inclusive	47.13
Diboson - WW	inclusive	113.9
DY+Jets	0<p_T<50	1402
DY+Jets	50<p_T<100	375.3
DY+Jets	100<p_T<250	91.79
DY+Jets	250<p_T<400	3.495
DY+Jets	400<p_T<650	0.4803
DY+Jets	p_T>650	0.04465

Table 5.2. The MC simulated samples used in the analysis with their cross-section.

M_{LQ} [TeV]	Cross section [pb] $\lambda_{u\mu}$				
	$\lambda = 0.1$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$
0.7	$3.5 * 10^{-3}$	$8.7 * 10^{-2}$	0.35	0.78	1.4
1.0	$9.5 * 10^{-4}$	$2.4 * 10^{-2}$	$9.5 * 10^{-2}$	0.21	0.38
1.5	$1.8 * 10^{-4}$	$4.6 * 10^{-3}$	$1.8 * 10^{-2}$	$4.2 * 10^{-2}$	$7.6 * 10^{-2}$
2.0	$4.9 * 10^{-5}$	$1.2 * 10^{-3}$	$5.0 * 10^{-3}$	$1.1 * 10^{-2}$	$2.1 * 10^{-2}$
2.5	$1.5 * 10^{-5}$	$3.9 * 10^{-4}$	$1.6 * 10^{-3}$	$3.7 * 10^{-3}$	$7.1 * 10^{-3}$
3.0	$5.3 * 10^{-6}$	$1.3 * 10^{-4}$	$5.7 * 10^{-4}$	$1.4 * 10^{-3}$	$2.7 * 10^{-3}$
3.5	$1.9 * 10^{-6}$	$4.9 * 10^{-5}$	$2.1 * 10^{-4}$	$5.5 * 10^{-4}$	$1.1 * 10^{-3}$
4.0	$7.3 * 10^{-7}$	$1.9 * 10^{-5}$	$8.7 * 10^{-5}$	$2.3 * 10^{-4}$	$5.1 * 10^{-4}$
4.5	$2.8 * 10^{-7}$	$7.6 * 10^{-6}$	$3.7 * 10^{-5}$	$1.0 * 10^{-4}$	$2.5 * 10^{-4}$
5.0	$1.1 * 10^{-7}$	$3.1 * 10^{-6}$	$1.7 * 10^{-5}$	$5.3 * 10^{-5}$	$1.3 * 10^{-4}$

Table 5.3. Cross-section values for different M_{LQ} mass points and coupling $\lambda_{u\mu}$. Uncertainties on cross-section are of the order of $\sim 5\%$ for this mass range.

M_{LQ} [TeV]	Cross section [pb] $\lambda_{b\mu}$				
	$\lambda = 0.1$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$
0.7	$2.7 * 10^{-4}$	$6.9 * 10^{-3}$	$2.8 * 10^{-2}$	$6.3 * 10^{-2}$	0.11
1.0	$5.3 * 10^{-5}$	$1.3 * 10^{-3}$	$5.4 * 10^{-3}$	$1.3 * 10^{-2}$	0.023
1.5	$6.3 * 10^{-6}$	$1.6 * 10^{-4}$	$6.7 * 10^{-4}$	$1.6 * 10^{-3}$	$3.1 * 10^{-3}$
2.0	$1.1 * 10^{-6}$	$2.8 * 10^{-5}$	$1.2 * 10^{-4}$	$3.2 * 10^{-4}$	$6.7 * 10^{-4}$
2.5	$2.4 * 10^{-7}$	$6.4 * 10^{-6}$	$3.0 * 10^{-5}$	$8.5 * 10^{-5}$	$1.9 * 10^{-4}$
3.0	$5.9 * 10^{-8}$	$1.6 * 10^{-6}$	$8.9 * 10^{-6}$	$2.8 * 10^{-5}$	$7.0 * 10^{-5}$
3.5	$1.6 * 10^{-8}$	$4.9 * 10^{-7}$	$3.1 * 10^{-6}$	$1.1 * 10^{-5}$	$3.1 * 10^{-5}$
4.0	$4.2 * 10^{-9}$	$1.6 * 10^{-7}$	$1.3 * 10^{-6}$	$5.5 * 10^{-6}$	$1.6 * 10^{-5}$
4.5	$1.1 * 10^{-9}$	$6.4 * 10^{-8}$	$6.6 * 10^{-7}$	$3.0 * 10^{-6}$	$9.1 * 10^{-6}$
5.0	$2.7 * 10^{-10}$	$3.0 * 10^{-8}$	$3.8 * 10^{-7}$	$1.8 * 10^{-6}$	$5.7 * 10^{-6}$

Table 5.4. Cross-section values for various M_{LQ} at and couplings $\lambda_{b\mu}$. Uncertainties on cross-section are of the order of $\sim 5\%$ for this mass range.

the top in the labels. A first preselection is applied to all events in the SingleMuon dataset and MC simulation matching trigger requirements. This preselection aims to reduce the overall background while maintaining most of the signal events. To pass the preselection, events have to satisfy the following criteria:

- We require 1 (or 2) muon in the final state, depending on the signal region - N_{μ}=1 (2);
- p_T of the leading muon (p_T^{μ}) has to be greater than 55 GeV to avoid the smaller trigger efficiency characterizing muon with p_T near the trigger threshold, thus ensuring a $\sim$ 100% trigger efficiency;
- p_T of the leading jet (p_T^{jet}) has to be greater than 100 GeV;
- The invariant mass of the μ+jet system ($M_{\mu jet}$) has to be greater than 300 GeV to remove events in the not interesting low mass region;

- A veto on the presence of reconstructed electrons in the final state $N_{ele} = 0$, to remove the spurious top and diboson production events and avoid or reduce potential overlap with other LQ analyses involving also electrons in the final state;

- The ratio between the p_T of the leading muon and jet has to be included between 0.6 and 1.6 ($0.6 < \frac{p_T^{\mu}}{p_T^{jet}} < 1.6$) and the azimuth angle between the leading muon and the leading jet has to be greater than 2.5 ($\Delta\phi_{\mu jet} > 2.5$). These criteria are motivated by Figures 3.5 and 3.4, showing, respectively, the distribution of the ratio betweenp_T^{μ} and p_T^{jet} and the distribution of $\Delta\phi_{\mu q}$ for various LQ hypotheses;

- Met significance ($\frac{MET}{\sqrt{\sum E_T}}$) has to be less than 10 $GeV^{1/2}$, to reduce the W+jets and top production background which mainly contain neutrinos in the final state. As stated above, our signal has no neutrinos, so the event's missing energy must be low. In an ideal detector, a non-zero value of the MET indicates the presence of weakly interacting particles. Still, experimental effects such as object misreconstruction, finite detector resolution, or detector noise can produce a large MET in an event with no neutrinos in the final state. Since the amount of MET in an event is approximately proportional to the square root of the total transverse energy [19], it is interesting to look at MET significance, which is defined as:
$$\text{MET significance} = \frac{\text{MET}}{\sqrt{\sum E_T}}, \tag{5.1}$$
where the $\sum E_T$ is the sum of the total transverse energy of the event. For signal events, this quantity is almost independent of the LQ mass. Therefore, it is more suitable to make a single cut independent from the LQ mass hypothesis and not introduce a dependence on the mass in the BDT training. The distribution of both MET and MET significance are shown in Fig 5.1 for different signal hypotheses, showing the latter one to be less dependent on the LQ mass.

- We require at least one jet in the final state (N_{jet} >=1);

- For the two muons signal region, we further require the invariant mass of the two muons ($M_{\mu\mu}$) to be greater than 110 GeV to reduce the DY+Jets events around the peak of the Z.

The list with all the preselection criteria for both categories is presented in Table 5.5.
The signal efficiency for all the kinematic requirements is between 40 and 60% for all signal hypotheses considered.

5.4 Background description

The relevant background for the search presented here arises from SM processes with at least one jet and at least one muon in the final state.

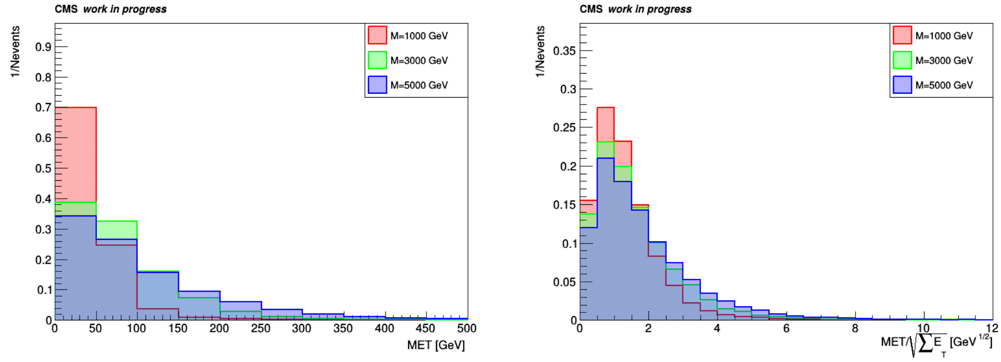


Figure 5.1. Distribution of MET (MET significance) for different M_{LQ} and $\lambda_{u\mu} = 1$ on the left (right).

1 Muon signal region		2 Muons signal region	
Preselection Criteria	**Requirement**	**Preselection Criteria**	**Requirement**
N_μ	$= 1$	N_μ	$= 2$
Leading muon p_T	> 55 GeV	Leading muon p_T	> 55 GeV
Leading Jet p_T	> 100 GeV	Leading Jet p_T	> 100 GeV
$M_{\mu jet}$	> 300 GeV	$M_{\mu jet}$	> 300 GeV
N_{ele}	$= 0$	N_{ele}	$= 0$
$p_T^\mu / p_T^{\text{jet}}$	$0.6 < \cdot < 1.6$	$p_T^\mu / p_T^{\text{jet}}$	$0.6 < \cdot < 1.6$
$\Delta\phi_{\mu\text{jet}}$	> 2.5	$\Delta\phi_{\mu\text{jet}}$	> 2.5
$\frac{\text{MET}}{\sqrt{\text{SumET}}}$	$< 10 GeV^{1/2}$	$\frac{\text{MET}}{\sqrt{\text{SumET}}}$	$< 10 GeV^{1/2}$
N_{jet}	≥ 1	N_{jet}	≥ 1
		$M_{\mu\mu}$	> 110 GeV

Table 5.5. Preselection criteria for 1 and 2 muon signal regions.

The most significant background contribution for the 1 Muon signal region comes from W+jets production, where W decays leptonically. Figure 5.2 shows an example of this process. This scenario typically involves one lepton with high transverse missing momentum ($p_{\text{T}}^{\text{miss}}$) and includes at least one jet that recoils against the W boson. Since this process includes a neutrino in the final state, we can reduce it for example by requiring a cut on the MET significance.

Other background contributions also arise from semi-leptonic $t\bar{t}$ events, as well as from single top (shown in Figure 5.3) and diboson productions. The top can decay into a W boson in these processes, containing a non-negligible number of high-p_T jet and lepton in the final state.

Another background process is the QCD multijet production, where, in some cases, a jet could be misidentified as a muon. Since the cross-section for QCD multijet production is significant, its contribution is not negligible (although small), even if the probability of identifying a jet as a muon is small.

For the two muon signal region, the main sources of backgrounds are processes with two leptons, like fully leptonic $t\bar{t}$ and DY+Jets, the latter shown in in Figure 5.4. These processes can also contribute to the 1 Muon signal region when the second

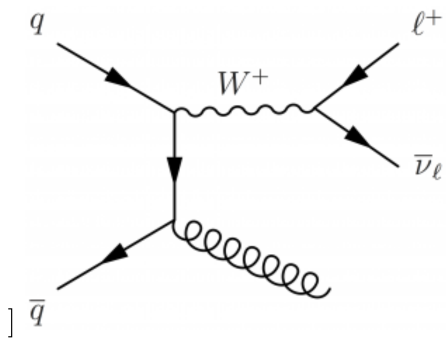


Figure 5.2. Feynman diagrams for W+jets production.

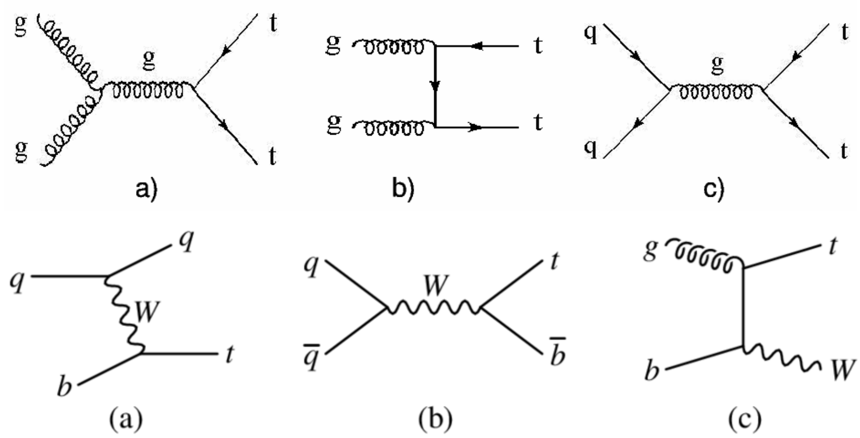


Figure 5.3. LO $t\bar{t}$ production (single top) on the top (bottom).

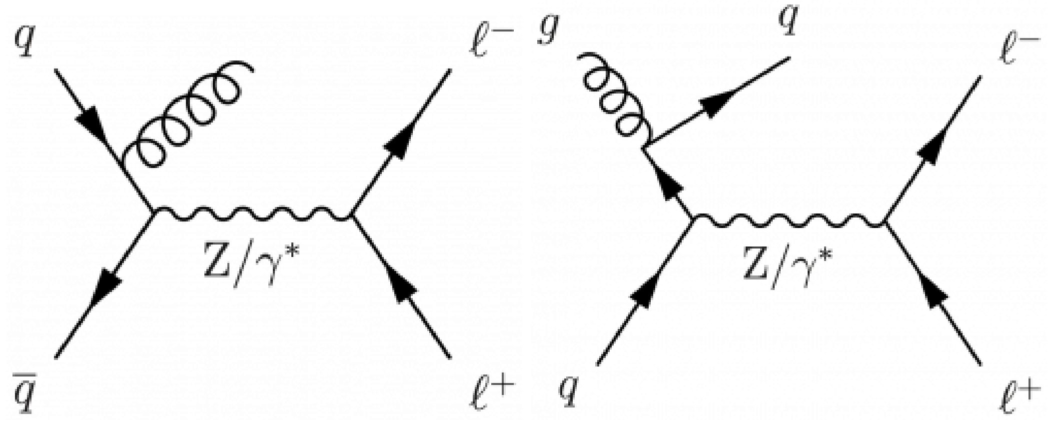


Figure 5.4. DY+Jets process at LO.

electron (muon) is not reconstructed according to the conditions established for the electron veto (muon reconstruction).
Even if we estimate the background in a data-driven way, we use the MC to underline the agreement between data and MC in some interesting kinematic distributions, shown later in this chapter. This check is important to verify the quality of the reconstruction and check the capability of the MC to reconstruct the data in the phase space of this analysis.

5.4.1 Control regions

A control region is the phase space selected to isolate a specific physics process from the signal, usually by requiring cuts orthogonal to the region of interest. Two different control regions are defined: one for top production and one for Drell-Yan processes. We defined these two control regions to compute a normalization factor for the MC simulation. These scale factors are calculated to account for discrepancies and align the simulated events with the observed data. They are computed by normalizing the yield of the MC to the observed yield in each specific control region. Then, they are applied to the MC simulation in the signal regions where the search is performed. The scale factors are computed starting from the background processes with less contamination from events of other background processes and signal ones. The following control regions are defined:

- For top production, a muon and an electron in the final state are required ($N_e = 1$ and $N_\mu = 1$) on top of the other preselection cuts.

- For the Drell-Yan, on top of the other preselection cuts, in the final state, the presence of two muons is required ($N_\mu = 2$). Moreover, the invariant mass of the two muons has to be included between 60 GeV and 110 GeV (60 GeV $< M_{\mu\mu} <$ 110 GeV), corresponding to the peak of the Z boson.

Some distributions, after the scale factors are applied, are shown in Fig. (5.5-5.6) for 2018 data. We can see a good agreement between data and MC in both control regions for the distributions shown. In particular, in the $M_{\mu\mu}$ distribution for the

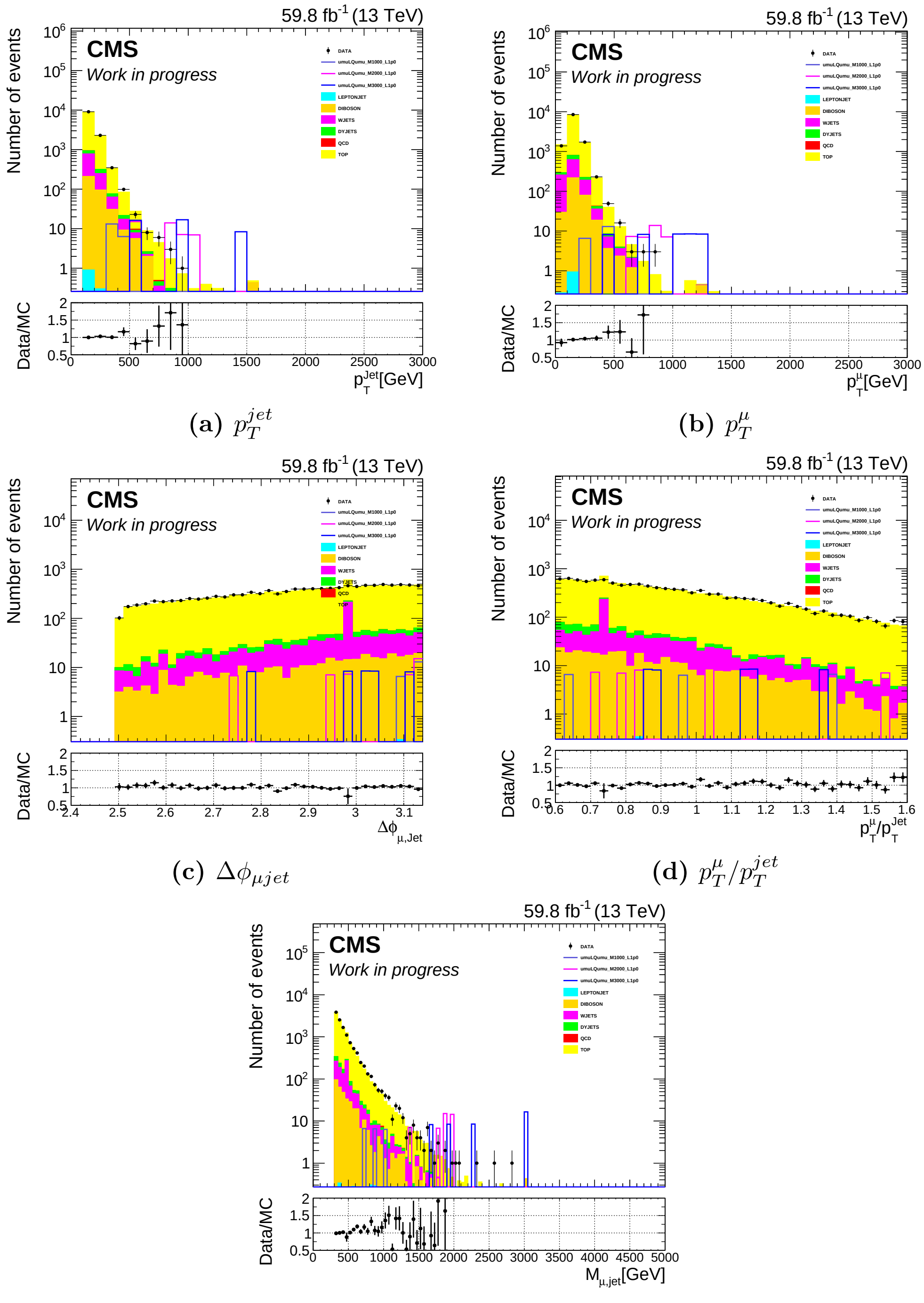
DY+Jets control region, we see low signal contamination around the peak of the Z. For the W+jets, events are taken at the preselection level (still fully dominated by background), and data and MC in that region are compared. The scale factor is computed with the preselection cut applied.
These scale factors are applied for the various datasets and are reported in the table 5.6 for all the main background processes considered. The values are roughly coherent with different years. The scale factors for W+Jets and DY+Jets are similar, as we expect, given that they are similar processes and have a value $\sim 90\%$ for all years, while the $t\bar{t}$ is compatible with 1. Only statistical uncertainties are reported.

Process	**2016 (preVFP)**	**2016 (postVFP)**	**2017**	**2018**
TTBAR	0.96 ± 0.06	1.02±0.06	0.97±0.07	1.00±0.04
DY+Jets	0.886±0.009	0.91±0.01	0.959±0.007	0.918±0.006
W+Jets	0.89±0.01	0.90±0.01	0.94±0.01	0.91±0.01

Table 5.6. Data/MC scale factors for various background processes.

5.5 Comparison between data and simulation

We conducted a thorough data check of the kinematic distribution for events that met the selection criteria. Since we used the MC to optimize the signal, we verified the agreement between the data and the simulations. The outcomes of this evaluation are detailed in this section.
Figure 5.7 - 5.8 shows the distributions of several kinematic variables for the 2018 dataset. The distributions for other datasets are shown in Appendix A. Each data distribution is compared to the corresponding one from the MC simulated samples summed. The distributions for two examples of simulated signal samples are also shown. Each simulated sample is generated with the 2018 conditions and scaled to match the 2018 luminosity. The event distributions in data show an acceptable level of agreement with those from MC simulated samples, ensuring that the data sample is not affected by pathologies. Some distributions (as p_T^{μ} or p_T^{jet}) are blinded above 500 GeV at the preselection level in order not to look at the tails which are sensitive to signal. The other distributions are not blinded since signal distribution is relatively broad, and there is no possibility of introducing any bias by looking at them at preselection stage.
The distribution of both p_T^{μ} and p_T^{jet}, Figure 5.7a, 5.7b, 5.8a and 5.8b, show good agreement between data and simulation. The azimuthal angle between the leading muon and the leading jet (Figure 5.7d and 5.8d) shows a good agreement for both 1 and 2 muons case, with the signal having a peak at π. The distribution of p_T^{μ}/p_T^{jet}, Figure 5.7e shows the different distribution between signal and background; in fact, the signal has a peak around 1, while the background processes show a broad distribution. Figure 5.8e shows the transverse momentum of the second muon $p_T^{\mu,2}$ distribution, which is significantly lower than the leading muon, allowing for a univocal choice of muon to use to compute the invariant mass of the leptoquark. Figure 5.8f shows the difference in the pseudorapidity of the two muons, where

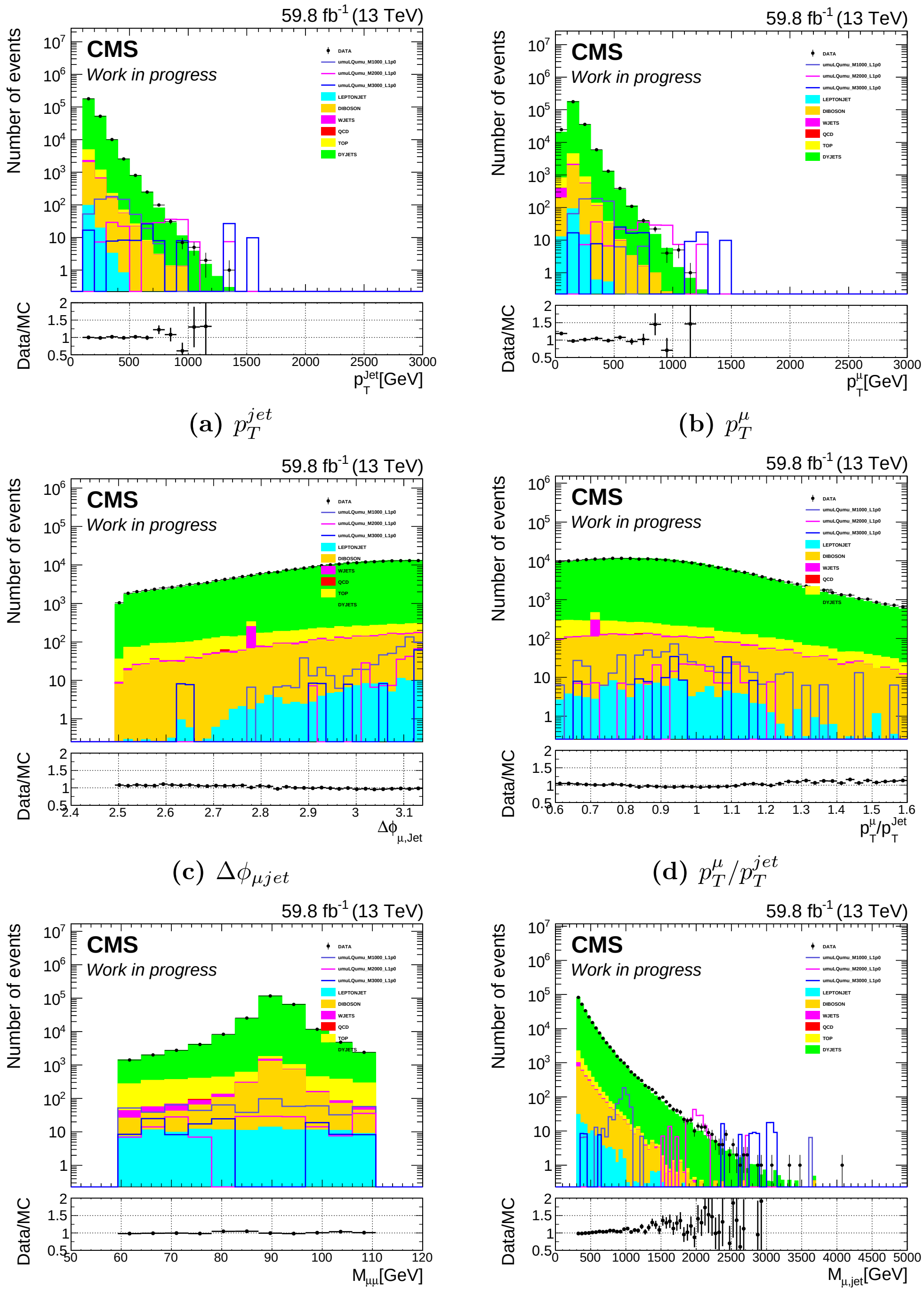


(a) p_T^{jet} **(b)** p_T^{μ}

(c) $\Delta\phi_{\mu jet}$ **(d)** p_T^{μ}/p_T^{jet}

(e) $M_{\mu jet}$

Figure 5.5. Kinematic distributions for the top control region for the 2018 dataset.



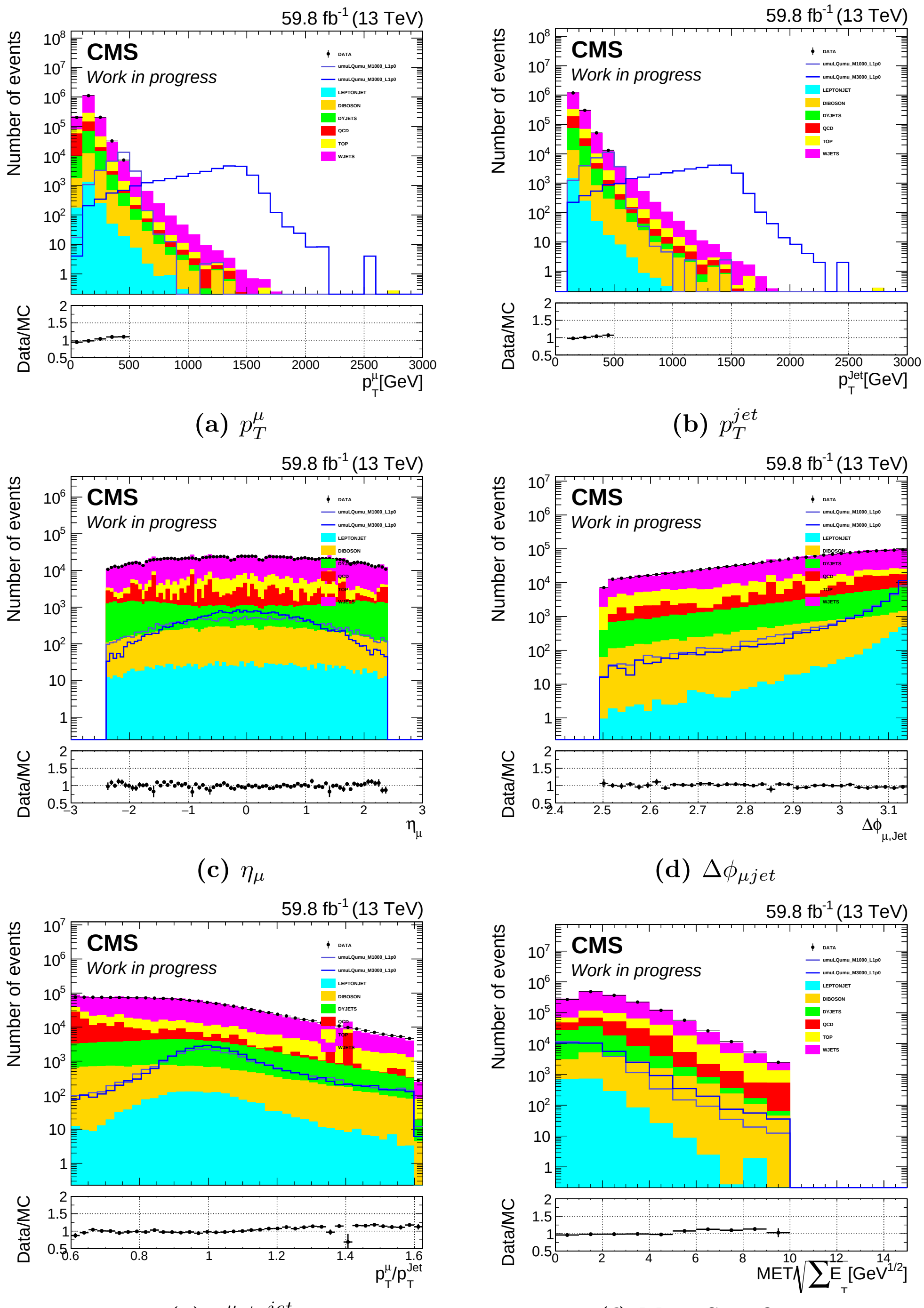
(a) p_T^{jet} **(b)** p_T^{μ}

(c) $\Delta\phi_{\mu jet}$ **(d)** p_T^{μ}/p_T^{jet}

(e) $M_{\mu\mu}$ **(f)** $M_{\mu jet}$

Figure 5.6. Kinematic distributions for the DY+Jets control region for the 2018 dataset.

background processes peak around 0, while the signal has two broad peaks for higher values of $\Delta\eta_{\mu\mu}$.
The $M_{\mu jet}$ distributions for both 1 and 2 muon signal region - Figure 5.9a - 5.9b - show the expected smoothly decreasing trend, which is in acceptable agreement with the background prediction. The two signal samples shown correspond to a cross-section of 1pb, about 100–1000 times the expected limit. The data are blinded above 1 TeV to avoid looking at the tails which are sensitive to signal.

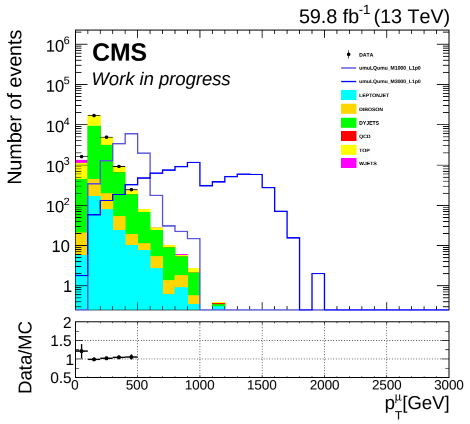


(a) p_T^{μ} **(b)** p_T^{jet}

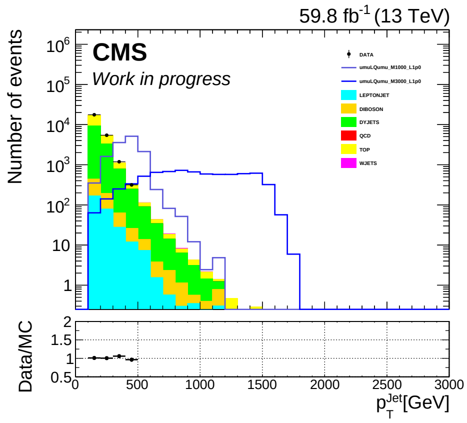
(c) η_{μ} **(d)** $\Delta\phi_{\mu jet}$

(e) p_T^{μ}/p_T^{jet} **(f)** MET Significance

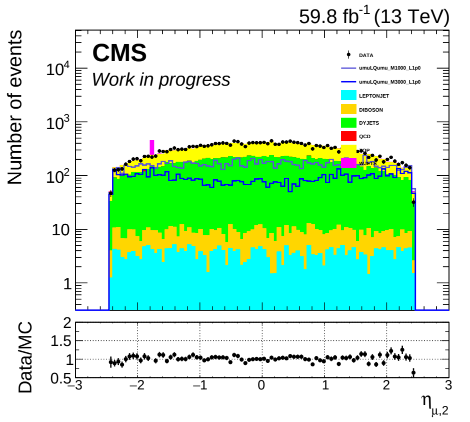
Figure 5.7. Comparison between 2018 data and simulated samples for some interesting distributions in the 1 Muon preselection signal region. The distributions of two simulated signal samples are also shown. They correspond to signal hypotheses with M_{LQ}= 1 TeV $\lambda = 1.0$ and M_{LQ}= 3 TeV $\lambda = 1.0$. The signal cross-section is set to a default value of 1 pb (to make signal histograms clearly visible in the plot). This value is two to three orders of magnitude larger than the expected cross-section limit of this search. Both p_T^{μ} and p_T^{jet} distributions are blinded above 500 GeV.



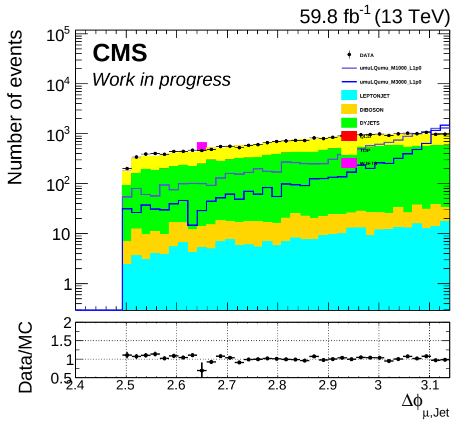
(a) p_T^{μ}



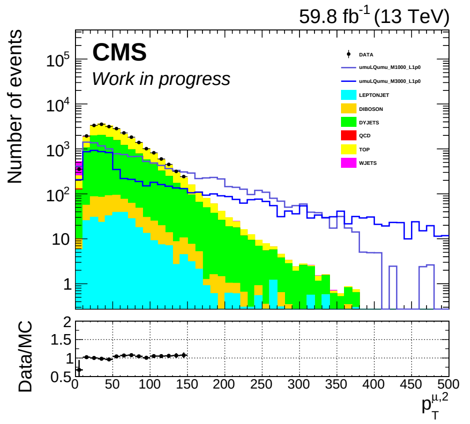
(b) p_T^{jet}



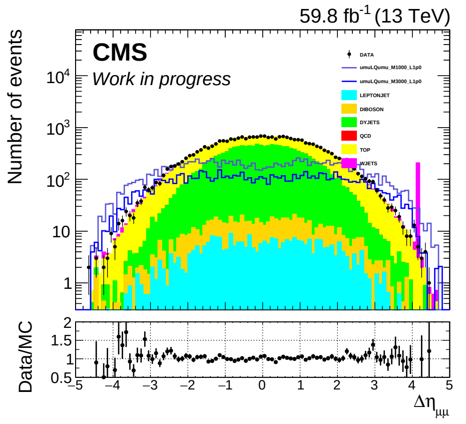
(c) $\eta_{\mu,2}$



(d) $\Delta\phi_{\mu jet}$

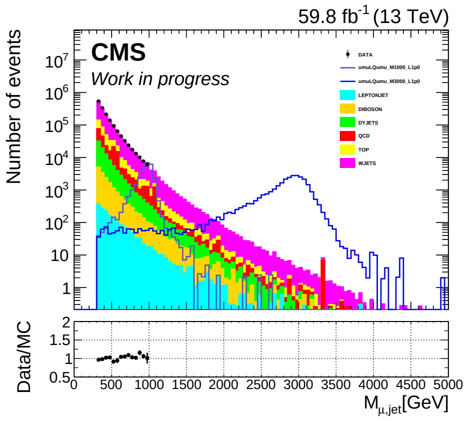


(e) $p_T^{\mu,2}$

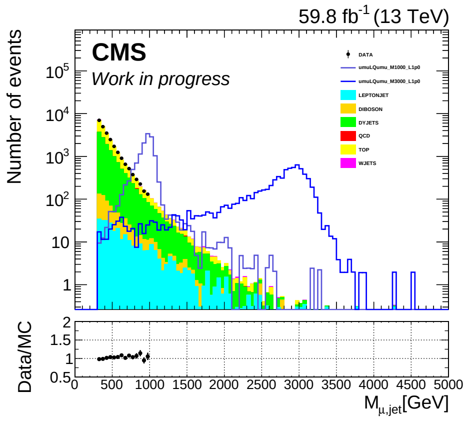


(f) $\Delta\eta_{\mu\mu}$

Figure 5.8. Comparison between 2018 data and simulated samples for interesting distributions in the 2 Muons preselection signal region. The distributions of two simulated signal samples are also shown. They correspond to signal hypotheses with M_{LQ}= 1 TeV $\lambda = 1.0$ and M_{LQ}= 3 TeV $\lambda = 1.0$. The signal cross-section is set to a default value of 1 pb (to make signal histograms clearly visible in the plot). This value is two to three orders of magnitude larger than the expected cross-section limit of this search. Both p_T^{μ} and p_T^{jet} distributions are blinded above 500 GeV, while $p_T^{\mu,2}$ is blinded above 150 GeV. The peak for the W+Jets samples present in some of the distribution are due to the large weight for inclusive samples, due to the low statistic of these samples.



(a) 1-muon signal region



(b) 2-muon signal region

Figure 5.9. Invariant mass of leading muon and leading jet system ($M_{\mu jet}$) in the 1 (2) muon signal region at preselection on the left (right). The distributions of two simulated signal samples are also shown. They correspond to signal hypotheses with M_{LQ}= 1 TeV $\lambda = 1.0$ and M_{LQ}= 3 TeV $\lambda = 1.0$. The signal cross-section is set to a default value of 1 pb (to make signal histograms clearly visible in the plot). This value is two to three orders of magnitude larger than the expected cross-section limit of this search. The $M_{\mu jet}$ distribution for both signal regions is blinded above 1 TeV.

Chapter 6

Analysis strategy

This Chapter describes the analysis strategy for this LQ search. The final state topology we are interested in is characterized by at least a high-p_T jet and 1 or 2 muons. The next sections will describe the selection optimization, the categories definition, the signal model, and the method used for the fit to the $M_{\mu jet}$ spectra, reporting the fit quality checks performed.

6.1 Selection optimisation

This analysis categorizes events according to the number of muons in the final state (1 or 2). A Boosted Decision Tree (BDT) approach is designed to maximize the significance of a potential LQ signal in data using the ROOT TMVA software [168]. This selection can also be done independently for each variable (a method commonly called "rectangular cut"), with the advantage of clear physical meaning and simple systematic uncertainties determination, but with the drawback of neglecting all the information in the variable correlations. Multivariate analysis methods, such as BDT, are used instead to perform event selection in a multidimensional approach to exploit variable correlations and maximize signal sensitivity.
The BDT is the predominant multivariate method employed in this thesis for classifying events into signal and background groups. It operates by constructing a decision tree where each node executes a decision based on the value of a single variable, typically resulting in a rectangular cut. However, the tree allows for diverse selection paths, enabling multiple hyper-rectangular sub-selections in the multidimensional parameter space. This flexibility permits the BDT to partition the parameter space into several adaptive hyper-boxes that conform to the signal sample distribution, unlike traditional rectangular cuts that define only a single hyper-box. The construction of a BDT involves a critical process known as *training*. During training, the BDT algorithm receives two distinct datasets representing pure signal and pure background. The algorithm then optimizes all selection paths and cuts with the inherent knowledge of the events' categories. A subsequent *testing* phase assesses the BDT's effectiveness using independent datasets that are statistically separate from those used during training. The BDT then classifies this mix of data, and the classifier's performance is evaluated post-implementation.
One challenge in BDT training is the potential for *over-training*, where the algorithm

adjusts to the training data too closely due to an excess number of parameters with respect to the number of events. To mitigate over-training, it is advisable to moderate the BDT's complexity by reducing its degrees of freedom, such as limiting the number or depth of each tree.
It is noteworthy that over-training can lead to sub-optimal performance when the BDT is applied to independent datasets, as the classifier may perform poorly due to its excessive specialization. The BDT is particularly susceptible to this issue compared to other multivariate techniques, given its highly adaptable decision-making process. Monitoring for over-training involves comparing outcomes during the testing phase with the training performance.

6.1.1 BDT training

The selected distributions for the optimization process must exhibit distinct characteristics between signal and background shapes; hence, they are named *discriminating variables*. To train the BDT, we used discriminating variables minimally influenced by the LQ mass since the reconstructed LQ mass will be the final observable for signal extraction. Subsequent sections will elaborate on the specific attributes of the BDT for different final states and the criteria for determining the final categories. The events for BDT training are split in half: one is used for training a BDT, while the other is used for the testing process. The signal used in the BDT training is a composite of various signal samples, incorporating all different LQ masses and coupling values at the same time, and only the uμ samples. The input variables are selected to minimize mass dependence, facilitating the mixing of samples with different masses. The background for the training comprises a weighted mix of all the different MC background samples described before except for the QCD, which is excluded due to its small contribution after full selection and low statistics of the samples. We train both signal and background with the total event weight, which takes into account both the correction to the physics objects due to reconstruction, ID, trigger efficiency and the generator weight due to NLO generation.
The input variables for the BDT are arranged by their importance, with the most important ones listed at the top, and are presented in Table 6.1 for 1 muon and 2 muons signal regions. Figure 6.1 and 6.2 show the distributions of the discriminating variables used as input for the BDT training for the 1 muon and 2 muon signal region respectively.
Figure 6.3 left (right) shows the BDT output for both training and test datasets for 1 muon (2 muon) signal region. A good separation is visible between the background and the signal. The BDT is not overtrained, as the training and test datasets are compatible.

6.1.2 Optimisation of signal significance

Even though the input variables for the BDT were selected to have limited dependence on the mass, the BDT distribution still has some variations for signal samples with

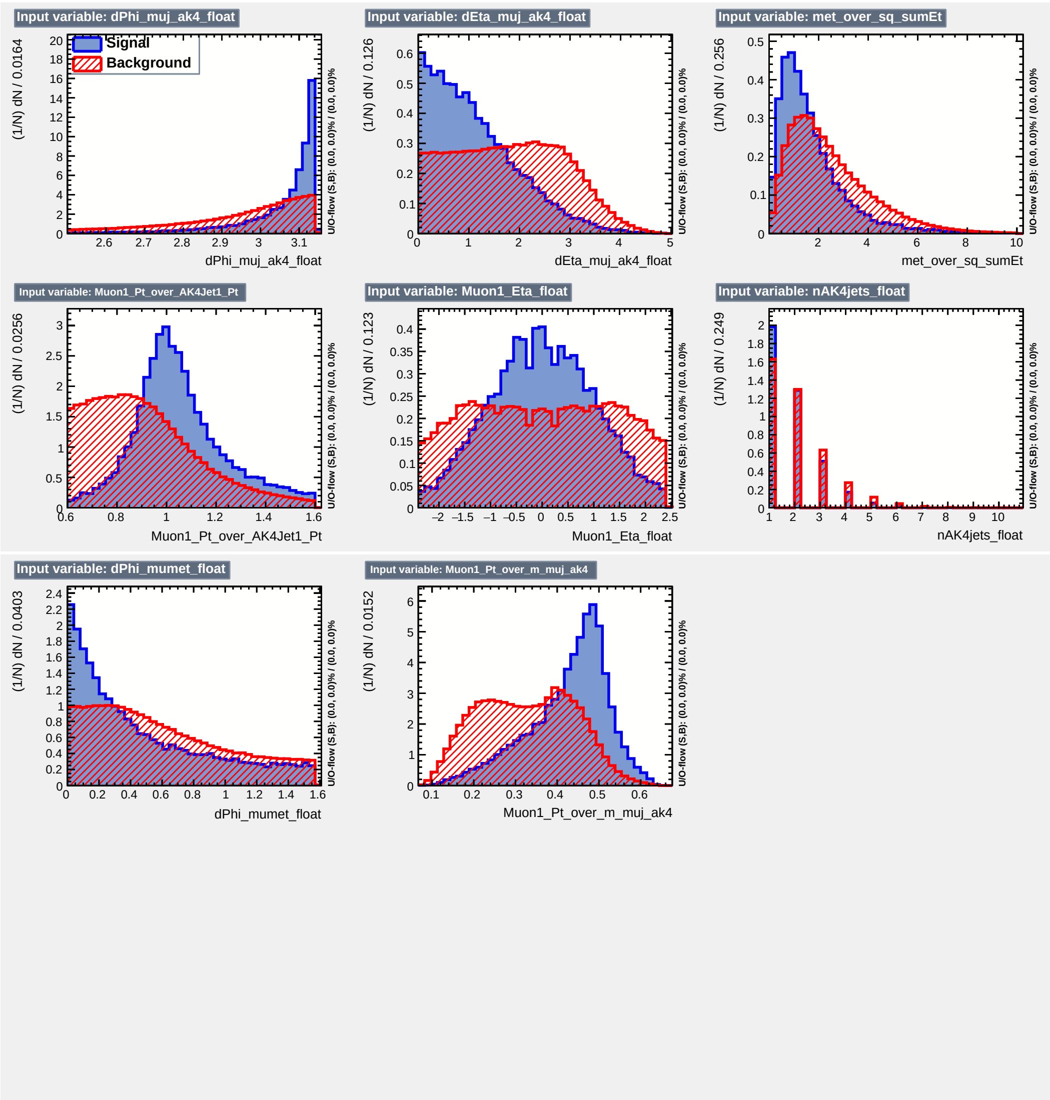


Figure 6.1. Discriminating variables used as input of the BDT for the 1-muon signal region.

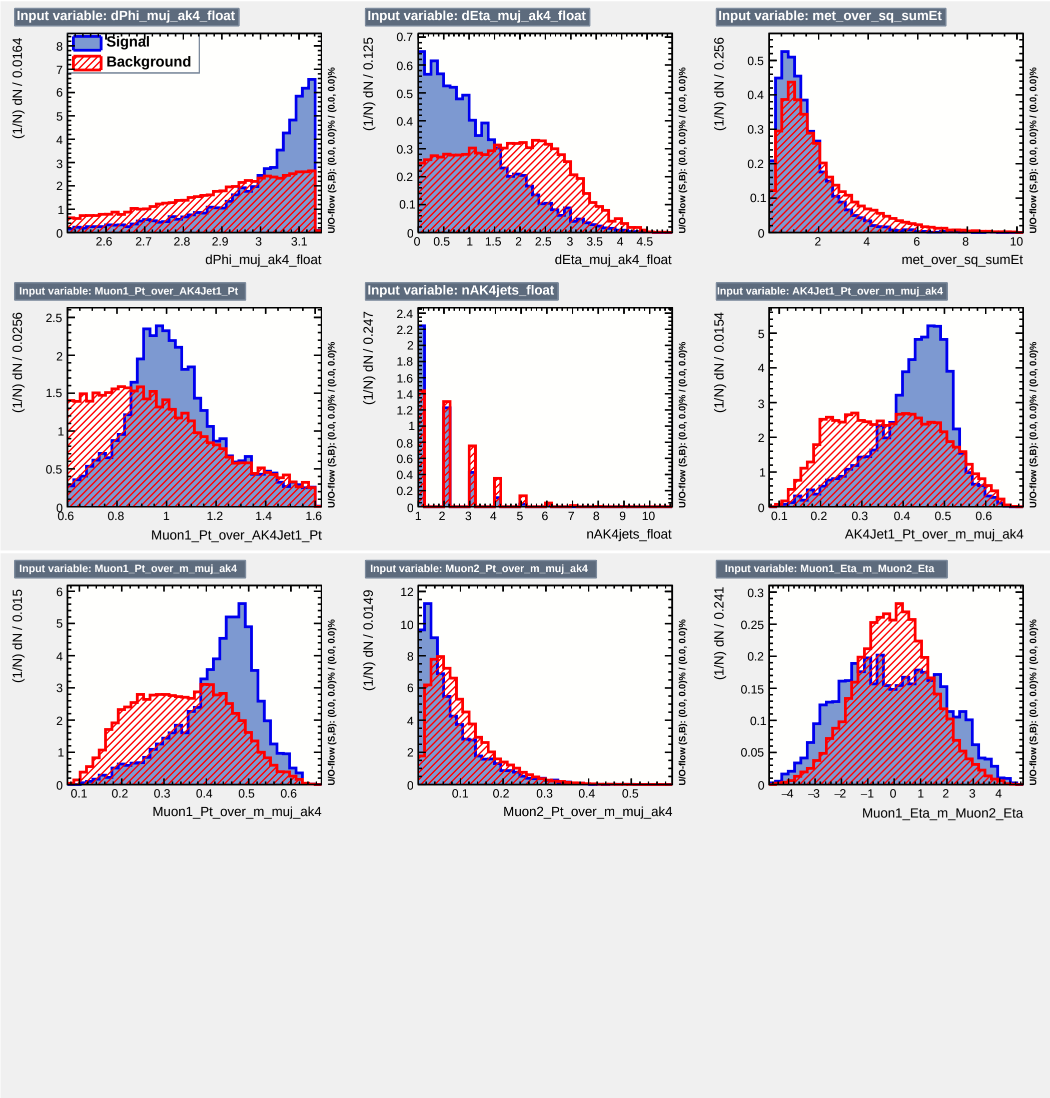


Figure 6.2. Discriminating variables used as input of the BDT for the 2-muon signal region.

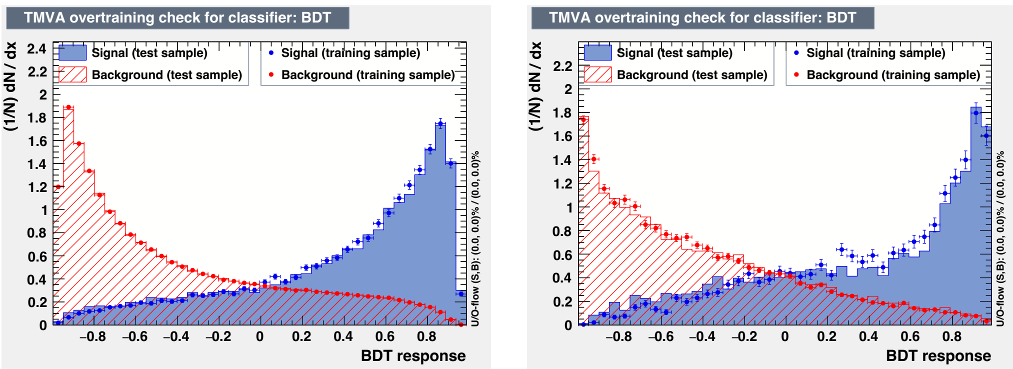


Figure 6.3. BDT output from the training and test for 1-muon signal region (2-muon signal region) on the left (right).

1 Muon	2 Muon
$\frac{p_T^\mu}{p_T^{jet}}$	$\Delta\phi_{\mu jet}$
$\frac{p_T^\mu}{M_{\mu jet}}$	$\frac{p_T^\mu}{p_T^{jet}}$
$\Delta\phi_{\mu jet}$	$\frac{p_T^{\mu,2}}{M_{\mu jet}}$
$\Delta\phi_{\mu MET}$	$\Delta\eta_{\mu\mu}$
$\frac{MET}{\sqrt{SumE_T}}$	$\frac{p_T^\mu}{M_{\mu jet}}$
η_μ	$\frac{p_T^{jet}}{M_{\mu jet}}$
$\Delta\eta_{\mu jet}$	$\frac{MET}{\sqrt{SumE_T}}$
N_{jets}	$\Delta\eta_{\mu jet}$
	N_{jets}

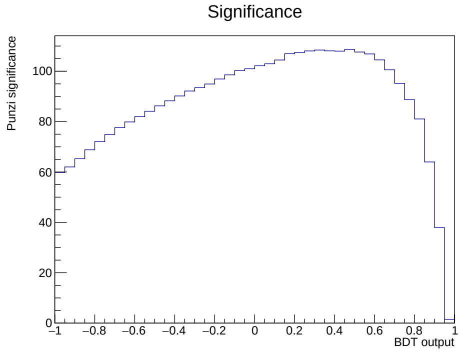
Table 6.1. Discriminating variables used as input for the BDT training. The most important ones are listed at the top.

different LQ masses. The BDT score cut selection must be optimized to maximize the signal significance. Therefore, the signal significance as a function of the BDT score cut was studied. We define the significance using the Punzi significance [160]:

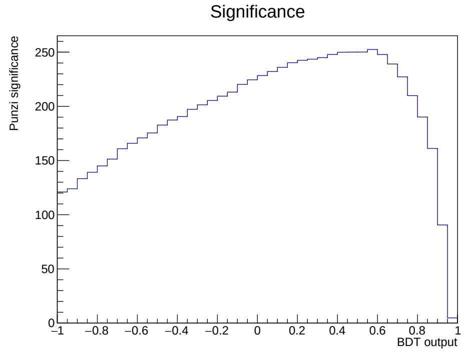
$$\text{Punzi Significance} = \frac{S}{\frac{a}{2} + \sqrt{B}}, \tag{6.1}$$

where S is the number of signal events, a is the number of standard deviations corresponding to a one-sided Gaussian test at a certain significance value, in this case 3, and B is the number of background events. We use this significance because it works well also when the number of background events is small ($B \sim 0$) and the standard significance (defined as $S = \frac{S}{\sqrt{B}}$) would go to infinity.
The number of signal and background events is computed in a $\pm 10\%$ window from the nominal LQ mass of the sample used to compute the significance, corresponding to ≈ 2 standard deviations of the 5% LQ mass resolution. The signal yield is computed as the sum of the number of events of the signal events that pass the BDT score cut. The background yield is calculated as the sum of the number of events of the background events that pass the BDT score cut.
For the 1 muon signal regions, the signal significance as a function of the BDT score cut is shown in Figure 6.4. The plots represent different LQ masses hypotheses. The significance shows a maximum BDT output of ~ 0.4 for all the signal samples. However, this value is not optimal for all the samples, especially in the hypothesis of larger LQ masses ($M_{LQ} > 3\,\text{TeV}$), where this signal region is expected to have more

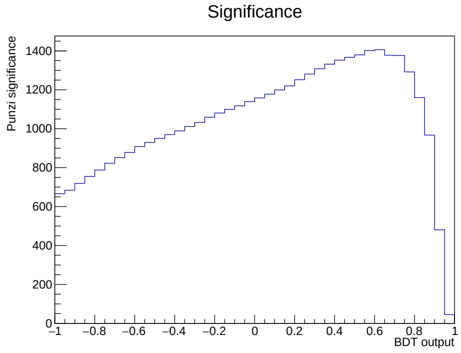
sensitivity. Therefore, we decided to have two categories: one with a BDT score in



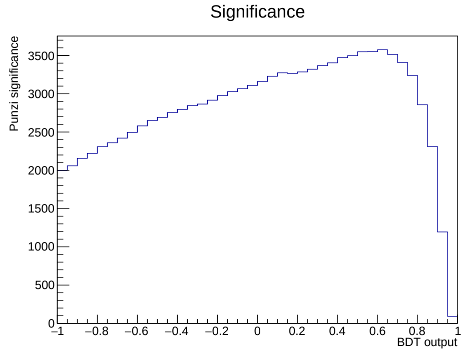
(a) M_{LQ} = 0.7 TeV and $\lambda_{u\mu}$=1



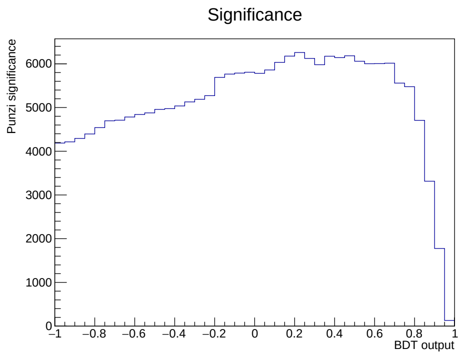
(b) M_{LQ} = 1.0 TeV and $\lambda_{u\mu}$=1



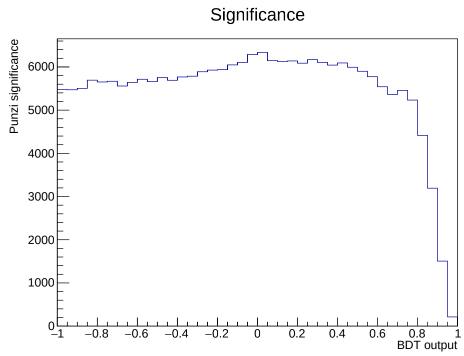
(c) M_{LQ} = 2.0 TeV and $\lambda_{u\mu}$=1



(d) M_{LQ} = 3.0 TeV and $\lambda_{u\mu}$=1



(e) M_{LQ} = 4.0 TeV and $\lambda_{u\mu}$=1



(f) M_{LQ} = 5.0 TeV and $\lambda_{u\mu}$=1

Figure 6.4. Punzi significance (arbitrary units) for the 1 Muon signal region for several M_{LQ} hypotheses and λ=1.

the range of 0-0.4 and one with a BDT score greater than 0.4. The tighter cut is the one that optimizes the signal significance, allowing for an optimal signal selection for all signal hypotheses, while the looser cut is used to have a second category with low purity to recover the efficiency. It has been observed that the optimal BDT boundaries do not vary with the coupling parameter λ.
After the BDT categorization, events are categorized according to whether the leading p_T jet, essential for constructing the LQ candidate, is b-tagged. The medium DeepJet Working Point (WP) is selected for optimal analysis sensitivity.

There are 4 signal categories in the 1 Muon signal region. They are defined as follows:

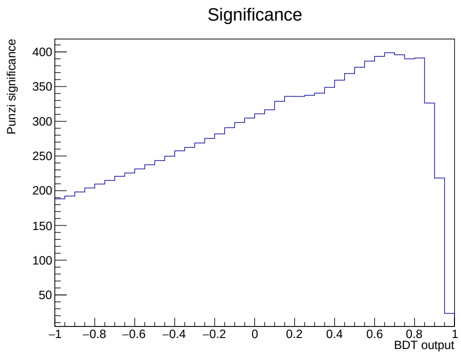
- **Loose+btag**: the BDT score is in the range of 0-0.4, and the leading jet is b-tagged;
- **Loose+no-btag**: the BDT score is in the range of 0-0.4, and the leading jet is not b-tagged;
- **Tight+btag**: the BDT score is greater than 0.4, and the leading jet is b-tagged;
- **Tight+no-btag**: the BDT score is greater than 0.4, and the leading jet is not b-tagged;

The same study has been repeated for the 2 muon signal region, with the result shown in Figure 6.5. In this case, there is a significant dependence on mass. This is related to the fact that the background is essentially zero for the 2-muon channel at high mass. In this signal region, we have background events in the low mass region ($M_{LQ} < 3\,\text{TeV}$), which is also where we expect a higher sensitivity with respect to the 1 muon signal region. To avoid introducing unnecessary complications, since the 1 muon signal region dominates the sensitivity in the high mass region, we optimize the BDT cut for the low LQ mass region, keeping the same selection for all mass hypotheses. The maximum of the BDT score is at ~ 0.6, for $M_{LQ} < 3$ TeV, but is not optimal for all samples. Therefore, in analogy with the 1 muon signal region, we added a second category to recover the signal efficiency (the cut region is BDT = [-0.2-0.6] see Table 6.2). After the BDT categorization, we used the same WP of the 1 Muon case for the b-tagging of the leading jet.

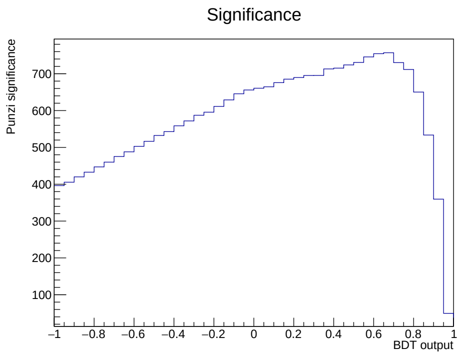
There are 4 signal categories in the 2 Muon final state. They are defined as follows:

- **Loose+btag**: the BDT score is in the range of -0.2-0.6, and the leading jet is b-tagged;
- **Loose+no-btag**: the BDT score is in the range of -0.2-0.6, and the leading jet is not b-tagged;
- **Tight+btag**: the BDT score is greater than 0.6, and the leading jet is b-tagged;
- **Tight+no-btag**: the BDT score is greater than 0.6, and the leading jet is not b-tagged;

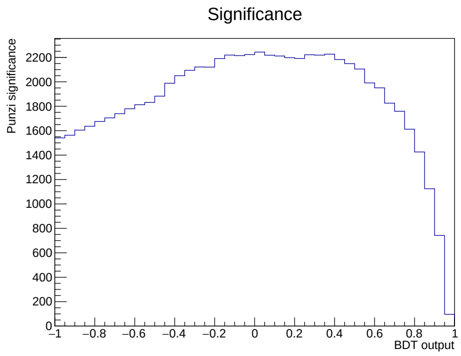
Table 6.2 summarizes the final selection for both signal regions, which have 4 categories each.



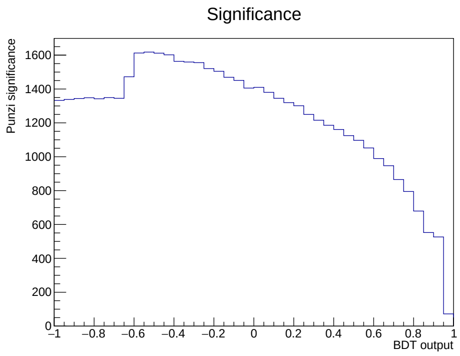
(a) $M_{LQ} = 0.7$ TeV and $\lambda_{u\mu}$=1



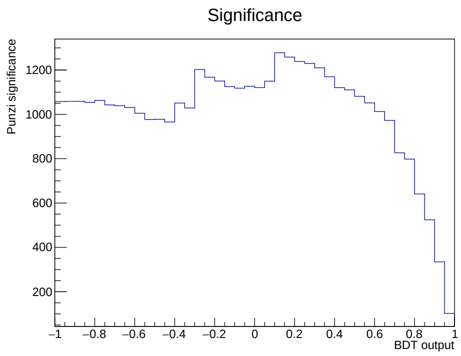
(b) $M_{LQ} = 1.0$ TeV and $\lambda_{u\mu}$=1



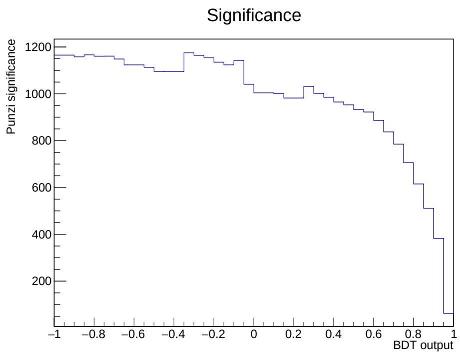
(c) $M_{LQ} = 2.0$ TeV and $\lambda_{u\mu}$=1



(d) $M_{LQ} = 3.0$ TeV and $\lambda_{u\mu}$=1



(e) $M_{LQ} = 4.0$ TeV and $\lambda_{u\mu}$=1



(f) $M_{LQ} = 5.0$ TeV and $\lambda_{u\mu}$=1

Figure 6.5. Punzi significance (arbitrary units) for the 1 Muon signal region for several M_{LQ} hypotheses and λ=1.

	1 Muon signal region				2 Muons signal region			
Preselection (tab. [5.5])	applied				applied			
N_μ	1				2			
$M_{\mu\mu}$	-				$M_{\mu\mu} > 110$ GeV			
BDT score	$0 < . < 0.4$ (loose)		≥ 0.4 (tight)		$-0.2 < . < 0.6$ (loose)		≥ 0.6 (tight)	
leading jet b-tag	no	yes	no	yes	no	yes	no	yes
final category	loose + no b-tag	loose + b-tag	tight + no b-tag	tight + b-tag	loose + no b-tag	loose + b-tag	tight + no b-tag	tight + b-tag

Table 6.2. Final category selection for the two signal regions.

6.2 Fit method

The analysis technique employed utilizes the $M_{\mu jet}$ mass spectra across various event categories to identify potential signals indicative of a narrow resonance on top of the steeply falling background distribution. The signal strength is extracted from a maximum likelihood fit using all eight event categories considered simultaneously. The $M_{\mu jet}$ mass is binned with variable sizes approximately corresponding to the mass resolution (see Appendix B). The fitting function for each category c, denoted by $m_{\mu j}$, comprises two elements:

- A smoothly falling function $B_c(M_{\mu jet})$ to model the background.
- A signal histogram of the $M_{\mu jet}$ spectra, characterizing the resonance.

A likelihood binned function L, which is the product of Poissonian function, is used. The likelihood is represented as:

$$L = \prod_{c=1}^{n_c} \prod_{i=1}^{n_b} \text{Poisson}(x_{ic}|\lambda_{ic}) = \prod_{c=1}^{n_c} \prod_{i=1}^{n_b} \frac{\lambda_{ic}^{x_{ic}} e^{-\lambda_{ic}}}{x_{ic}!}. \tag{6.2}$$

Here, i indexes the bins n_b within the $M_{\mu jet}$ spectrum and c indexes the categories n_c. The variables x_{ic} and λ_{ic} represent the observed and expected event counts in the i^{th} bin of the c^{th} category, respectively. Specifically, λ_{ic} is defined by:

$$\lambda_{ic} = \mu s_{ic} + b_{ic}, \tag{6.3}$$

where μ is the signal strength modifier, s_{ic} is the expected signal, and b_{ic} is the expected background. These components are defined as follows:

- $b_{ic} = N_{ic}(B) = \int_{m_{ic,\text{low}}}^{m_{ic,\text{high}}} B_c(M_{\mu jet})\, dM_{\mu jet}$, where $B_c(M_{\mu jet})$ is the background function integrated over the $M_{\mu jet}$ mass bin range to yield the expected background event count.
- $s_{ic} = N_{ic}(S)$ is the expected number of signal events in a given bin of a $M_{\mu jet}$ histogram assuming an equal cross-section for all signals equal to 1 pb.
- The signal strength modifier μ is a multiplicative factor that scales the signal cross-section.

6.2.1 Background estimation

The first thing to do is then to determine the background model to be used in the fit. In this particular case, it's an empirical function that describes the background spectrum and depends on a number of parameters to be determined.
To simplify the analysis, we merged the different years, thus having only 8 event categories on which do the fit. To validate this choice, we compare the shape of the invariant mass distribution of the muon + jet system for all 8 categories. The invariant mass distribution of each dataset was first normalized to the same luminosity and then divided by the 2018 data. Thus, we compare the shape of 4 datasets. The result is shown in Fig. 6.6, and the yellow band is the statistical uncertainty on the 2018 data, where the denominator is always the 2018 data. The data are compatible within the statistical uncertainty.

MC simulations within each category reveal that the $M_{\mu jet}$ mass spectra from background processes diminish smoothly. The strategy for this analysis involves searching for a narrow resonance over an exponentially decaying background.
We choose to use two empiric families of function to describe the background that better fits the data and were used previously in similar searches by CMS and ATLAS:

- The so-called "Standard dijet" function, which has already been utilized by the previous dijet searches [84, 85, 74, 65, 62]. This function will be called f_1 in the following sections.
- The so-called "UA2" function, utilized in ATLAS and CDF similar searches [156, 23, 21, 49]. This function will be called f_2 in the following sections.

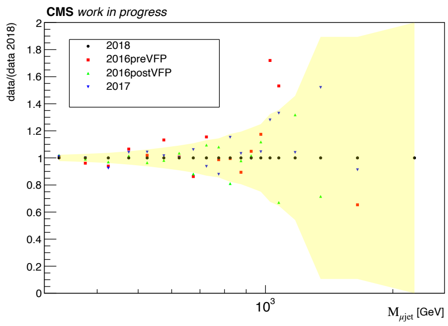
Both families of functions with different numbers of parameters are listed in Table 6.3, where $x = \frac{M_{\mu jet}}{\sqrt{s}}$.

A Fisher F-test, described in Sec 6.2.2, is used to determine the optimal number of

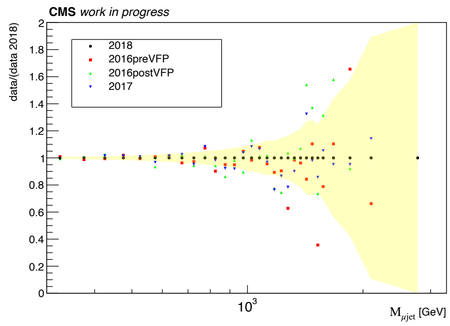
Number of Parameters	dijet function	UA2 function
2	$f_1^{2\text{par}} : p_0(1-x)^{p_1}$	$f_2^{2\text{par}} : \frac{1}{\left(\frac{p_0}{x}\right)^{p_1}}$
3	$f_1^{3\text{par}} : \frac{p_0(1-x)^{p_1}}{x^{p_2}}$	$f_2^{3\text{par}} : \frac{exp^{-p_2*(\frac{p_0}{x})}}{\left(\frac{p_0}{x}\right)^{p_1}}$
4	$f_1^{4\text{par}} : \frac{p_0(1-x)^{p_1}}{x^{p_2+p_3*\log x}}$	$f_2^{4\text{par}} : \frac{exp^{-p_2*(\frac{p_0}{x})-p_3*\left(\frac{p_0}{x}\right)^2}}{\left(\frac{p_0}{x}\right)^{p_1}}$
5	$f_1^{5\text{par}} : \frac{p_0(1-x)^{p_1}}{x^{p_2+p_3*\log x+p_4\log^2 x}}$	

Table 6.3. Standard dijet (UA2) functions tested in the left (right) column.

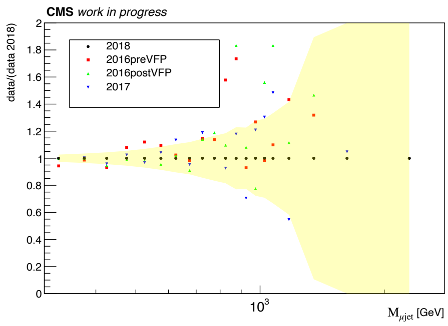
the parameters for each of the two function families. Each of the parameters of f_1 and f_2 is free to vary during the fit, with flat priors assumed. The background is modeled alongside the signal, modified by the signal strength μ. This procedure, inclusive of signal-plus-background, mitigates bias in signal extraction. Tests to confirm the absence of bias are detailed in Sec 6.4.1,6.4.2.
A blinding policy was strictly followed during the analysis, with preliminary tests validating the background-only fits. The blinding policy was to perform the only background fit in the whole invariant mass range, but showing the data only up to 1



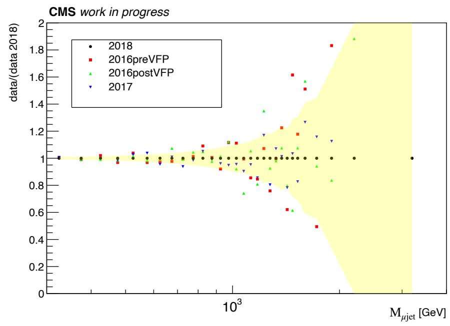
(a) 1-muon, loose, btag category



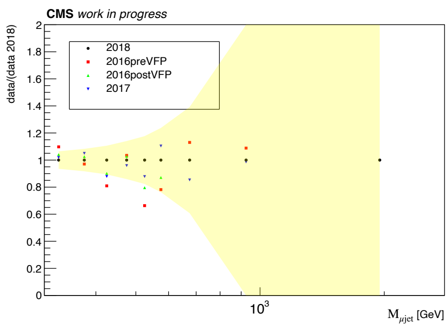
(b) 1-muon, loose, nobtag category



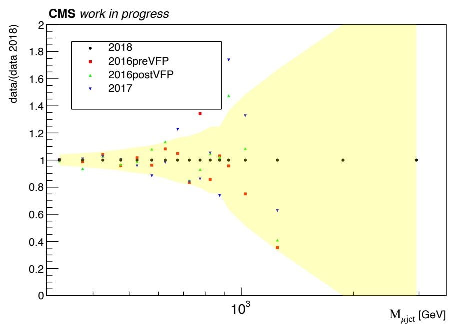
(c) 1-muon, tight, btag category



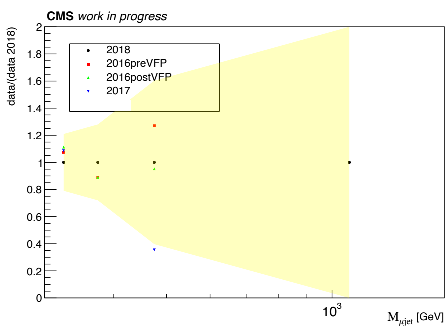
(d) 1-muon, tight, nobtag category



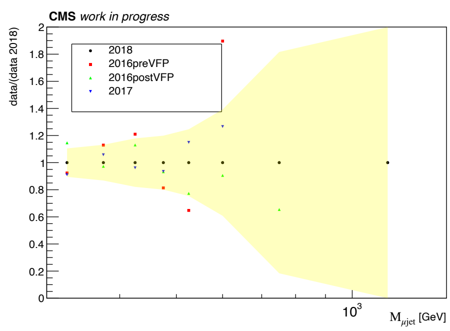
(e) 2-muon, loose, btag category



(f) 2-muon, loose, btag category



(g) 2-muon, loose, btag category



(h) 2-muon, loose, btag category

Figure 6.6. Comparison of the invariant mass shape of the 4 different datasets by normalizing the data to the same luminosity. The yellow band is the statistical uncertainty on the 2018 data, where the denominator is always the 2018 data.

TeV. The background fit procedure is validated by the statistical test performed in Section 6.4.
We choose the starting range of the fit for both functions to be 450 GeV to have the same range for all the categories while maintaining a good fit quality. Since the resonance searched has a minimum invariant mass at the TeV scale - since lower masses were excluded by previous searches reported in Table 1.1 - the precise starting point of the fit is not critical for this analysis.

6.2.2 Choice of fit function

Since we estimate the background from empirical functions, it is important to determine the number of parameters sufficient to describe the background distribution. In fact, when the number of parameters is sufficient, adding more parameters does not significantly improve the fit.
This section outlines the analysis, to choose the number of parameters for the background fitting function. Data were used to fit the $M_{\mu jet}$ mass distribution across all event categories. This approach enabled independent testing of the fitting function across various categories with distinct event counts. The testing begins with the hypothesis that there are two fitting models, M_0 with n parameters and M_1 with $n+1$ parameters:

$$M_0 : (\theta_1, \ldots, \theta_n) \quad \text{and} \quad M_1 : (\theta_1, \ldots, \theta_{n+1}). \tag{6.4}$$

Given observations y, the likelihoods for these models post-fitting are:

$$L(y|M_0, \theta_1, \ldots, \theta_n) \quad \text{and} \quad L(y|M_1, \theta_1, \ldots, \theta_{n+1}). \tag{6.5}$$

According to the Neyman-Pearson lemma, the likelihood ratio is the most effective discriminator between the two hypotheses:

$$LR(\theta_{n+1}) = \frac{L(y|M_0, \hat{\theta}_1, \ldots, \hat{\theta}_n)}{L(y|M_1, \hat{\theta}_1, \ldots, \hat{\theta}_{n+1})}. \tag{6.6}$$

Assuming Gaussian errors for each observation y_i, the likelihoods are products of Gaussian distributions, leading to:

$$LR(\theta_{n+1}) = \prod_i \frac{\exp\left(-\frac{(y_i - \mu_i(M_0))^2}{2\sigma_i^2}\right)}{\exp\left(-\frac{(y_i - \mu_i(M_1))^2}{2\sigma_i^2}\right)}. \tag{6.7}$$

Here, y_i represents the data, i.e., the number of entries for each bin, and σ_i the Gaussian standard deviations, and $\mu_i(M_0)$ and $\mu_i(M_1)$ the model predictions. By taking the logarithm of the likelihood ratio and multiplying it by -2, we can express the result as:

$$LLR = -2 \log LR(\theta_{n+1}) = \sum_i \left(\frac{(y_i - \mu_i(M_1 0)^2}{\sigma_i^2} \right) - \sum_i \left(\frac{(y_i - \mu_i(M_1))^2}{\sigma_i^2} \right). \tag{6.8}$$

This can be expressed as the difference between two chi-squared distributions:

$$LLR = \chi^2(\nu_0 = n) - \chi^2(\nu_0 = n+1) = \chi^2(\nu = 1). \tag{6.9}$$

Here, ν_0 represents the number of fitting parameters, and $\nu = 1$ represents the degrees of freedom. The LLR statistic is then the difference in χ^2 values from fits of the two models, where the resulting LLR follows a chi-squared distribution with one degree of freedom. Interpretation of the right-tail p-value from the LLR distribution is as follows:

- For a p-value ≥ 0.05, the difference in chi-squared values is not statistically significant, suggesting a preference for the model with fewer parameters.
- For a p-value < 0.05, model M_1 is statistically favored over model M_0.

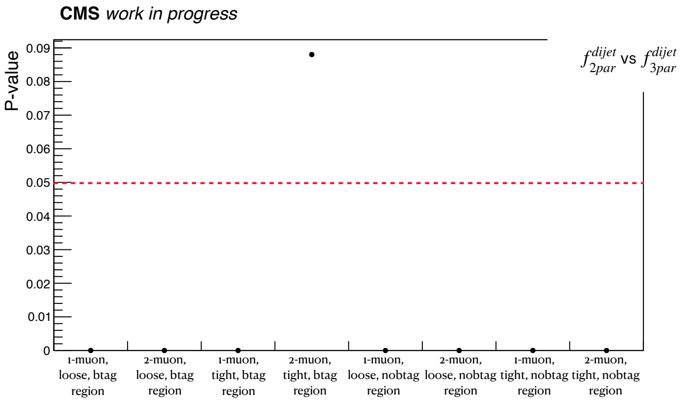
The analysis was conducted using both families functions from the dijet and UA2 family functions, all possessing varying parameter counts as detailed in Table 6.3. The investigative process unfolded in the following manner:

- Background events from data were categorized as described in Table 6.2.
- For every category, $M_{\mu jet}$ mass spectra were fitted utilizing the approach described in the previous section, employing the functions referenced in Table 6.3.
- The chi-squared value for each fit was computed for every category under the assumption of Gaussian uncertainties for the $M_{\mu jet}$ mass bin contents, denoted by $\sigma_i = \sqrt{N_i}$, with N_i representing the count of events in the i^{th} bin.
- Subsequently, the LLR was calculated to evaluate the comparative fit quality between the pairs f_{2par} and f_{3par}, as well as between f_{3par} and f_{4par} and f_{4par} and f_{5par}.
- The LLR's p-values were calculated for all categories and functions.

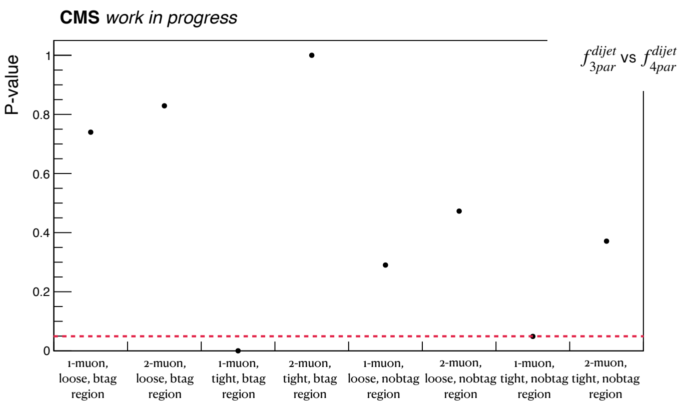
The results, shown in Figure 6.7 and based on the analysis of the standard dijet functions, indicated that f_{3par} generally outperformed f_{2par}. Additionally, the inclusion of a fourth parameter in f_{4par} did not substantially enhance the model's fit to the data, except for one category: 1 Muon, tight, btag category. The results for the UA2 function are shown in Appendix C. The final numbers of parameters chosen for each category are summarized in table 6.4.

6.2.3 Muon-jet invariant mass distributions

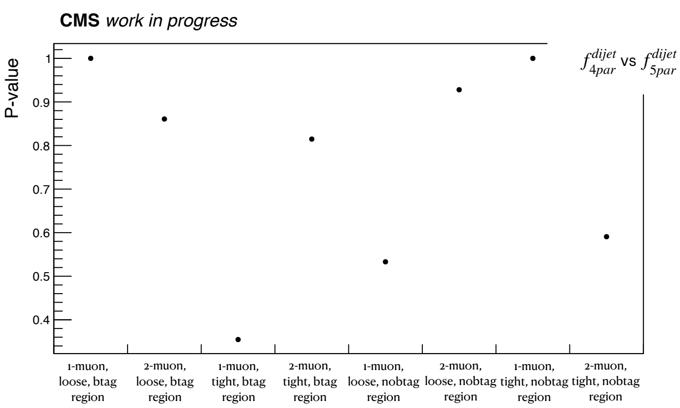
In this section, we show the final fit for both background functions. The data are blinded after 1 TeV while the fit is performed on all ranges. The data are blinded after the 1 TeV to avoid introducing a human bias in the analysis by looking at the data and optimizing the analysis should a deviation from the background be found. The result is shown in Figures 6.8 -6.9. The top part in each frame shows the data



(a) f^{dijet}_{2par} vs f^{dijet}_{3par}



(b) f^{dijet}_{3par} vs f^{dijet}_{4par}



(c) f^{dijet}_{4par} vs f^{dijet}_{5par}

Figure 6.7. Numbers of parameters for all categories for standard dijet function.

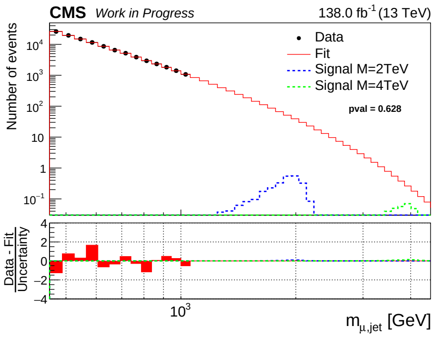
Category	**standard dijet function**	**UA2 fucntion**
1 Muon BDT tight + btag	4 parameters	4 parameters
1 Muon BDT tight + no-btag	3 parameters	3 parameters
1 Muon BDT loose + btag	3 parameters	3 parameters
1 Muon BDT loose + no-btag	3 parameters	3 parameters
2 Muon BDT tight + btag	3 parameters	3 parameters
2 Muon BDT tight + no-btag	3 parameters	3 parameters
2 Muon BDT loose + btag	3 parameters	3 parameters
2 Muon BDT loose + no-btag	3 parameters	3 parameters

Table 6.4. Number of parameters chosen for each category.

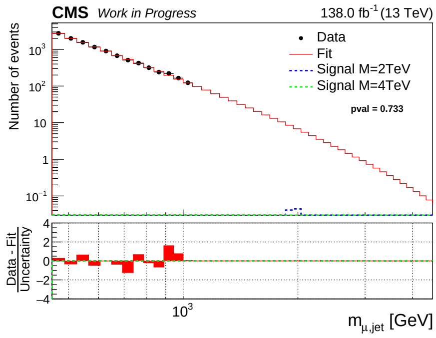
distribution, the background component of the fit. The bottom part shows, for each bin, the pulls, defined as:

$$\text{pull} = \frac{N_{\text{data}} - N_{\text{fit}}}{\sqrt{N_{\text{data}}}}. \tag{6.10}$$

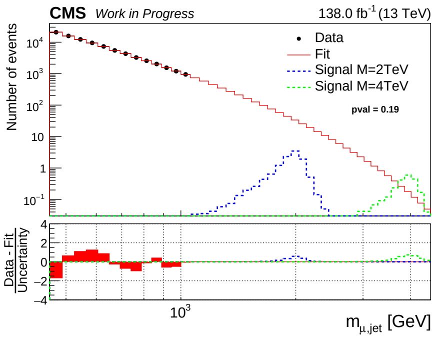
The p-value shown in the plots is the p-value of the fit for all bins. The p-values obtained suggest a good-quality fit for all categories.



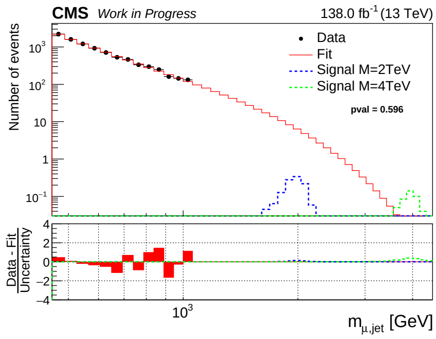
(a) 1-muon, loose, nobtag category



(b) 1-muon, loose, btag category

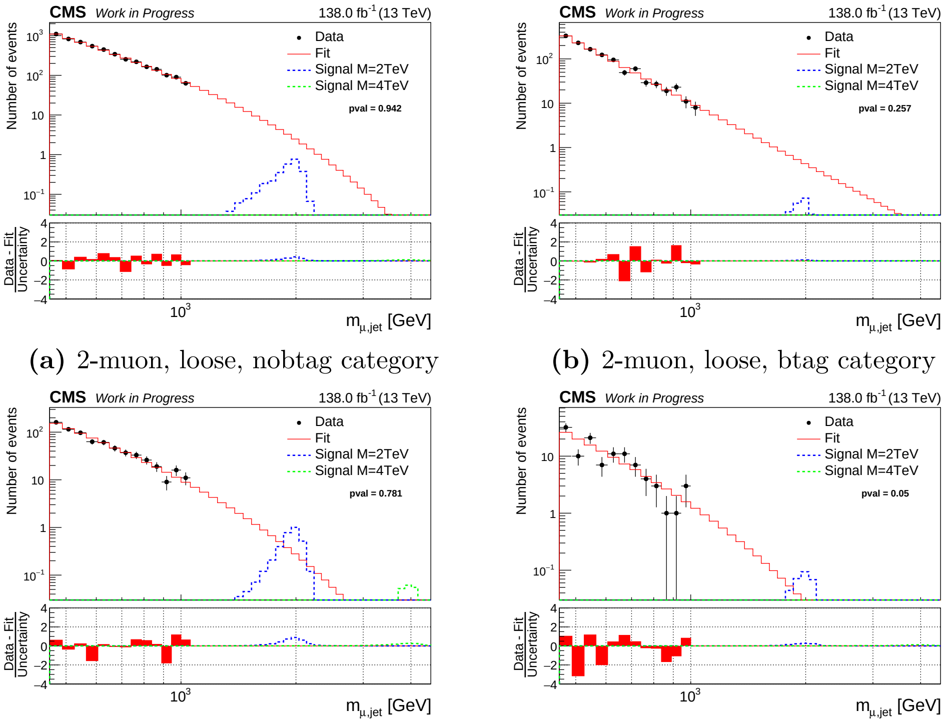


(c) 1-muon, tight, nobtag category



(d) 1-muon, tight, btag category

Figure 6.8. Background fits for the standard dijet function for all the categories for the 1 muon signal region. The black points are shown up to a reconstructed mass of 1 TeV. The red line is the background estimation obtained using the standard dijet function. The blue (green) line represents a signal hypothesis with M_{LQ} = 2 (4) TeV and $\lambda_{u\mu}$ =1, which is normalized to the expected upper-limit cross-section. The horizontal axis shows the value of the $M_{\mu jet}$ spectra, while the vertical axis shows the number of events in each bin. The bottom part shows the difference between the data and the background fit function divided by the statistical uncertainty. We do not see any trend in the residual distributions. The background fit procedure is validated by the statistical test performed in Section 6.4.



(a) 2-muon, loose, nobtag category

(b) 2-muon, loose, btag category

(c) 2-muon, tight, nobtag category

(d) 2-muon, tight, btag category

Figure 6.9. Background fits for the standard dijet function for all the categories for the 2 muon signal region. The black points are shown up to a reconstructed mass of 1 TeV. The red line is the background estimation obtained using the standard dijet function. The blue (green) line represents a signal hypothesis with M_{LQ} = 2 (4) TeV and $\lambda_{u\mu}$ =1, which is normalized to the expected upper-limit cross-section. The horizontal axis shows the value of the $M_{\mu jet}$ spectra, while the vertical axis shows the number of events in each bin. The bottom part shows the difference between the data and the background fit function divided by the statistical uncertainty. We do not see any trend in the residual distributions. The 3σ deviation from the background found in the 2-muon, tight, btag category is considered to be compatible with a statistical fluctuations since the excess is compensated in the adjacent bin and no visible trend is present. The background fit procedure is validate by the statistical test performed in Section 6.4.

6.3 Leptoquark signal model

Signal shapes are generated using MC samples of signal processes, as detailed in Chapter 3, and are listed in Table 5.3 and 5.4. The complete selection described in Table 6.2 is also applied to signal samples, which are thus divided into different event categories based on the number of muons, BDT selection, and presence or absence of a btagged leading jet.
The $m_{\mu j}$ signal distributions derived from these categories describe the signal in the signal+background fit.

6.3.1 Muon-jet mass shape interpolation

The shapes for each simulated signal sample are interpolated to obtain the signal shapes for the intermediate M_{LQ} and λ values. The interpolation is performed using the linear interpolation algorithm described in Ref [161].
We take a single, normalized (to 1) shape for each signal at specific M_{LQ} and λ values to interpolate. At first, shapes are interpolated between LQ mass points every 100 GeV at fixed λ, as described in Fig 6.10 (left) for a specific coupling. The second step is to interpolate between λ values with a step of 0.25 at fixed mass, as shown in Fig 6.10(right).

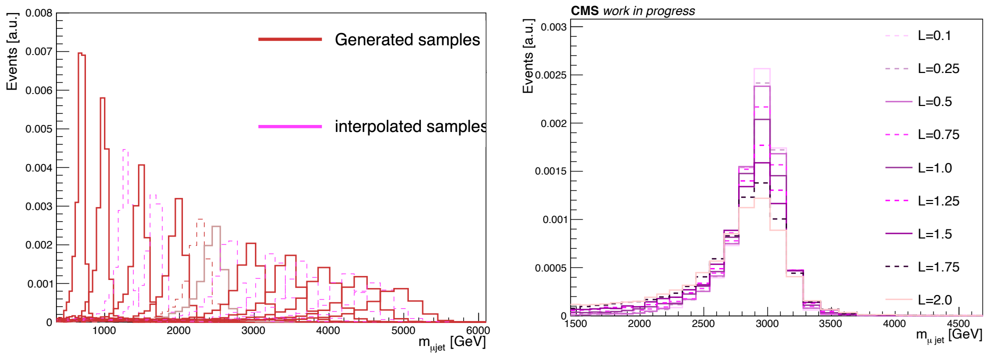


Figure 6.10. Linear signal interpolation interpolation of between mass at fixed $\lambda_{u\mu} = 1$ (coupling at fixed $M_{LQ} = 3$ TeV) points on the left (right) for a light quark produced LQ for the 1-muon, tight, nobtag category. The full lines are the generated samples, while the dotted lines are the interpolated ones.

6.3.2 Efficiency interpolation

In conjunction with signal shapes, the signal efficiencies across each category have been interpolated for the median values of M_{LQ} and λ. The signal efficiency for each category is defined as follows:

$$\epsilon_i = \frac{N_i}{N_{gen}}, \tag{6.11}$$

where N_i is the number of events passing the final selection for the category i, defined as the integral of the histogram, and N_{gen} is the total number of generated events for the considered signal sample.

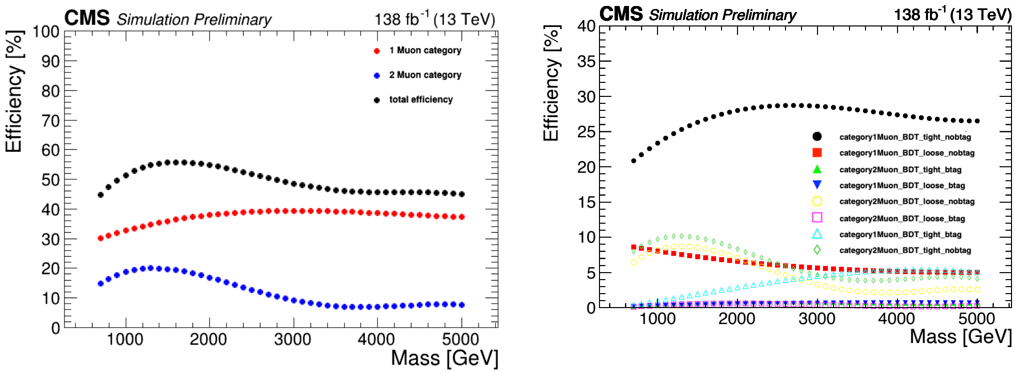


Figure 6.11. Total signal efficiency (single category efficiency) as a function of the LQ mass on the left (right) for $\lambda = 1$ and u quark.

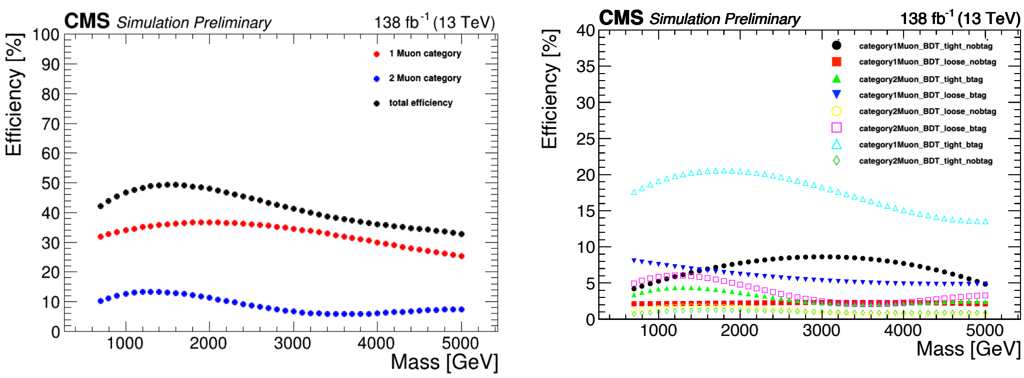


Figure 6.12. Total signal efficiency (single category efficiency) as a function of the LQ mass on the left (right) for $\lambda = 1$ and b quark.

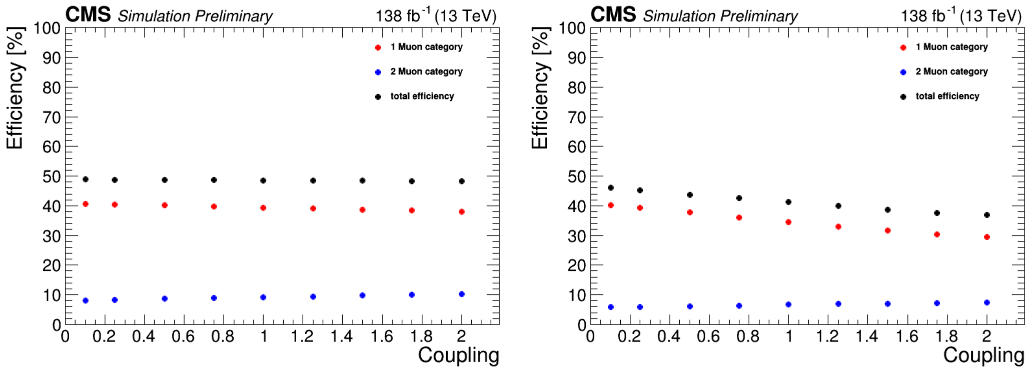


Figure 6.13. Total signal efficiency as a function of the coupling on the left (right) for $M_{\mu j}$=3 TeV and u (b) quark.

We studied the variation of the efficiency as a function of the LQ mass, as shown in Figures 6.11 and 6.12 and the coupling λ, as shown in Fig. 6.13. The interpolation is performed using a polynomial function of the form:

$$\epsilon(m_{\mu j}, \lambda) = \sum_{i=0}^{n} a_i x^i, \tag{6.12}$$

where x is the LQ mass or the coupling λ. We performed a fit on each single category with a polynomial that can vary for each category in order to achieve a good quality fit and a good description of the efficiency variation vs mass or coupling.
As shown in Figure 6.11 left, the efficiency for the light quark hypothesis (LQ->$u\mu$) increases with low mass (up to 2 TeV) due to a higher preselection acceptance for the 2 muon signal regions due to the process kinematic. Figure 6.11 right shows the efficiency for the eight categories separately. The 1-muon, tight, and nobtag category has higher efficiency, as expected since the leading jet is not coming from a b-quark.
Figure 6.12 left shows the total efficiency for LQ($b\mu$) hypothesis. The efficiency then decreases at higher mass in the bμ case due to the higher tails in the $M_{\mu j}$ distribution as discussed in Section 3.1.1. The efficiency for the single categories is reported in Figure 6.12 right, where the category with the highest efficiency is the 1-muon, tight and btag as expected.
Figure 6.13 shows the efficiency as a function of the coupling for a LQ($u\mu$) (LQ($b\mu$)) on the left (right). The efficiency for both models is similar for small coupling ($\lambda < 0.5$), while it decreases with the increase of the coupling for the LQ($b\mu$) model due to the higher tails in the mass distribution.

6.4 Validation of fit method

Two tests were performed to check the fit procedure described: goodness of fit and bias tests. These verify that the background fit function is suitable to describe the background in data and does not introduce a significant bias in the signal extraction procedure.

6.4.1 Goodness-of-fit test

We perform a goodness of fit test for all event categories using the function f_1, under the assumption that solely background processes contribute to the data, thus implementing a background-centric fit. Pseudodata sets are then generated from f_1 functions, assuming Poissonian fluctuation within each histogram bin. These pseudodata sets are subsequently refitted under the background-only assumption. This procedure (generation and fit) was replicated 1000 times. During each iteration, the test statistic q was calculated. This statistic, q, extends the chi-squared test to scenarios where bin contents are not necessarily normally distributed, particularly applicable in cases of low-count bins.
The test statistic is conceptualized as the likelihood ratio contrasting the empirical fit to the data and the saturated model [102], the latter being a model in which the predicted event count for each histogram bin is precisely the observed count:

$$q = -2\log\left(\prod_{c=1}^{n_c}\prod_{i=1}^{n_b}\frac{\text{Poisson}(x_{ic}|\lambda_{ic})}{\text{Poisson}(x_{ic}|x_{ic})}\right) = 2\sum_{c=1}^{n_c}\sum_{i=1}^{n_b}\left[\lambda_{ic} - x_{ic} + x_{ic}\log\left(\frac{x_{ic}}{\lambda_{ic}}\right)\right]. \tag{6.13}$$

The test statistic q thus serves as a gauge for fit adequacy, with its value reflecting the degree of concordance between the fitted model and the observed data distribution. The distribution of q obtained from the fit to the pseudodata distributions is roughly Gaussian. The value of q is then evaluated also for the fit to the data (q_{data}), and it is compared with the mean value of q from the pseudodata (toy experiments). To quantify this comparison, we introduce a p-value defined as the proportion of toy experiments where q exceeds q_{data}, relative to the total number of toy experiments (n_{toys}):

$$p\text{-value} = \frac{n_{q>q_{\text{data}}}}{n_{\text{toys}}}. \tag{6.14}$$

A p-value approaching zero implies a stark divergence between the model and the data, indicating that the chosen fitting function may be inadequate, as it fits the data poorly compared to the pseudodata. Conversely, a p-value nearing unity suggests potential overfitting if the fitting model possesses more degrees of freedom than the data can reliably constrain. Figure 6.14 illustrates the q distributions, the specific q_{data}, and the corresponding p-value. The p-value of 0.30 indicates a good fit quality. Thus, the different dijet functions (with their different parameters for each category) provide a good quality fit overall.

6.4.2 Bias study

The background estimation is acquired through fitting with an empirical function, f_1. Given that the exact form of the $M_{\mu jet}$ distribution for background processes is not fully established, f_1 represents one of several potential fitting functions. For example f_2 can fit the data with the same performance as f_1. Moreover, the presence of a signal can influence the fit, potentially causing the fit function to adapt to signal features, hence introducing bias. Consequently, bias was evaluated, and the potential for bias in both f_1 and f_2 was assessed. Our procedure was as follows:

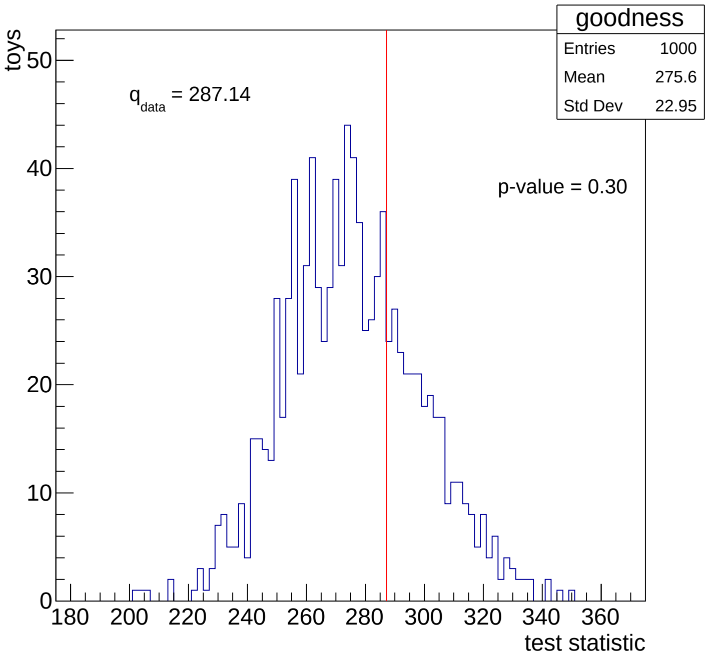


Figure 6.14. Distributions of the test statistic (q) obtained from 1000 toy experiments for the background estimation obtained using the standard dijet function. The red vertical lines mark the test statistic values evaluated from data (qdata).

- A sample of 250 pseudodata distributions is generated for all signal hypotheses considered using one of the two background functions. These functions were parameterized based on the optimal fit to actual collision data. A signal commensurate with the expected upper limit from the analysis was incorporated into this pseudodata.
- The distributions are fitted with the same generation function or with the other one, following the same fit method used for the analysis of actual data. The resulting signal cross-sections were determined alongside their standard deviations.
- The discrepancy between the signal cross-section extracted from the injected pseudodata and the actual injected value provided a measure of the signal bias.

The procedure described above is repeated also without the injection of a signal. The outlined method is reiterated for each signal hypothesis to assess the potential bias, which we quantify by calculating the "pull" of the signal cross-section. This pull is the difference between fitted (σ_{fit}) and the injected (σ_{inj}) signal cross sections, normalized by the standard deviation error (σ_{err}) derived from the fit:

$$\text{pull} = \frac{\sigma_{\text{fit}} - \sigma_{\text{inj}}}{\sigma_{\text{err}}}. \tag{6.15}$$

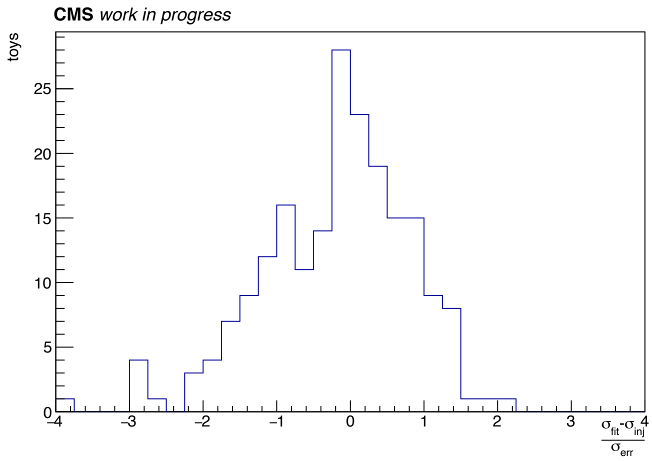
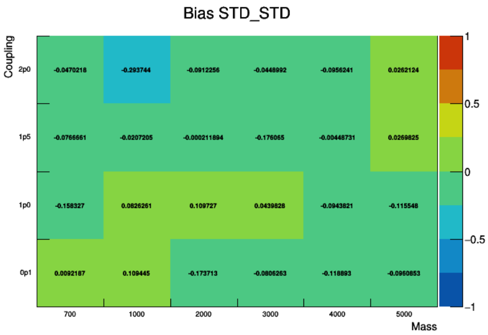


Figure 6.15. Pull distribution plots of fitted signal strength parameter (r) from 250 toys both generated and fitted with function f1. In this plot, the injected signal cross-section is equal to 3 times the expected upper limit.

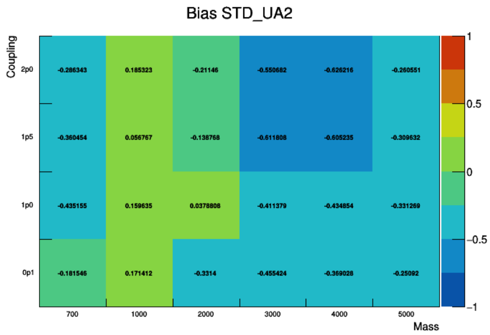
Figure 6.15 shows the pull distributions obtained for a signal hypothesis for the test performed both generating and fitting with f1. The mean of the histograms gives our estimation of the bias in units of σ_{err}. The signal injected (σ_{inj}) is equal to 3 times the expected upper limit.
In general, an absolute value of the bias below 0.5 is considered negligible. In this case, in fact, the total uncertainty on the signal cross-section from the combination of the bias and the fit uncertainty would be:

$$\sigma_{\mathrm{err}}^{\mathrm{tot}} = \sqrt{\sigma_{\mathrm{err}}^2 + (0.5\sigma_{\mathrm{err}})^2} \approx 1.1\sigma_{\mathrm{err}} \tag{6.16}$$

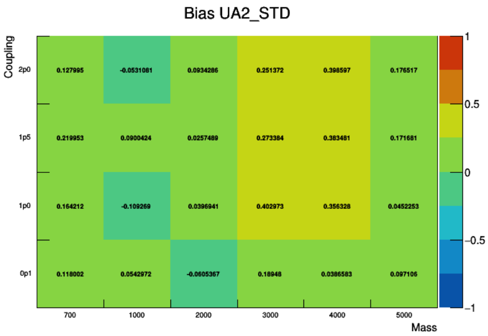
The increase due to the bias on the total uncertainty would be about 10% of the fit uncertainty -i.e. the error on the final limit due to the bias would increase of 10% - therefore it can be considered negligible.
The bias study has been performed by generating and fitting with the same function and generating with a function and fitting with the other one. The bias for a signal injected with a cross-section equal to 3*expected limit is shown in Figures 6.16 and 6.17 respectively for LQ($u\mu$) and LQ($b\mu$). The bias is generally lower than 0.5, and thus it is negligible. Since fitting with the f_1 function gives less bias, we can conclude that f_1 is the function that introduces the smaller bias between the two, so we choose it for the fit. An analog study performed in the case of $\sigma_{inj} = 0$ is shown in Appendix D, to verify the bias introduced by the background fitting function in case of no signal.



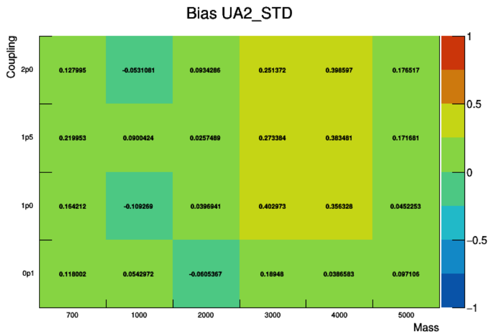
(a) Generated and fitted with f_1



(b) Generated with f_1 and fitted with f_2

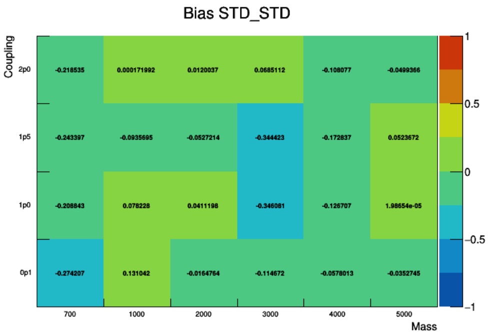


(c) Generated with f_2 and fitted with f_1

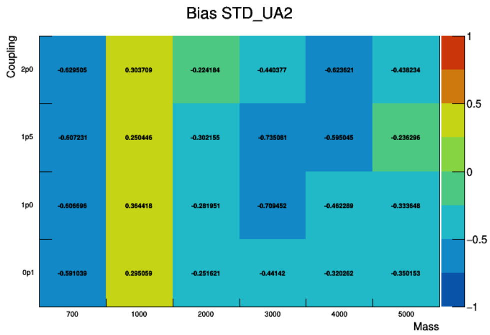


(d) Generated and fitted with f_2

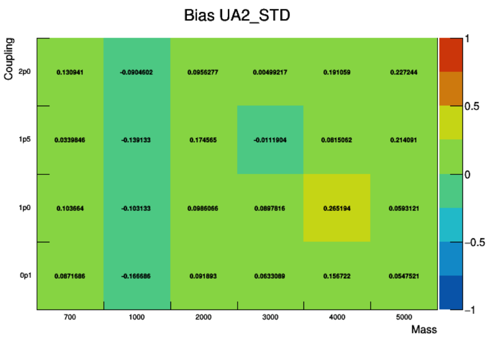
Figure 6.16. Mean of pull distribution for the bias study versus $\lambda_{u\mu}$ and M_{LQ} from pseudodata. A signal is injected with a cross-section equal to the 3*expected limit. In the z-axes of the plots, the bias is shown.



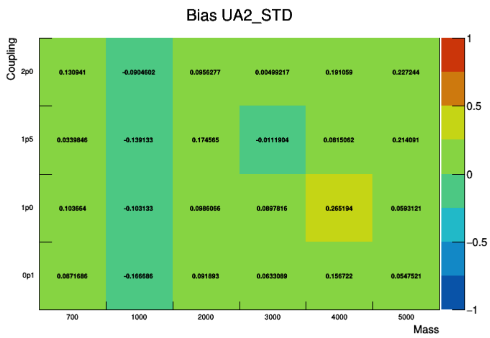
(a) Generated and fitted with f_1



(b) Generated with f_1 and fitted with f_2



(c) Generated with f_2 and fitted with f_1



(d) Generated and fitted with f_2

Figure 6.17. Mean of pull distribution for the bias study versus $\lambda_{b\mu}$ and M_{LQ} from pseudodata. A signal is injected with a cross-section equal to the 3*expected limit. In the z-axes of the plots, the bias is shown.

Chapter 7

Results and cross-section limits

In this chapter, we present the sensitivity from the search for signal processes leading to the production of a LQ resonance. The results are the expected upper limits on the cross-section of the signal, utilizing the complete dataset of pp collisions from the LHC Run 2, at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 138 fb^{-1}, and adheres to the strategy detailed in Chapter 6. Since the analysis has not been approved yet, it is not possible to show the observed limits but only the expected ones. Finally, we looked at 1% of the statistics also in the signal region.

7.1 Limits evaluation

The absence of a statistically significant excess in event counts during the analysis dictates the establishment of upper bounds for the cross-section on producing a LQ decaying into muons. To calculate these bounds, we utilize a variant of the frequentist approach, known as the CL$_s$ technique [131, 162]. The corresponding test statistic within the LHC CL$_s$ framework [4] is presented:

$$\hat{q}_\mu = -2 \ln \frac{L(x|\mu, \hat{\theta}_\mu)}{L(x|\hat{\mu}, \hat{\theta})}, \tag{7.1}$$

where $\hat{\theta_\mu}$ refers to the conditional maximum likelihood estimators of θ, given the signal strength parameter μ and "data" refer to the pseudo-data (toys). The pair of parameter estimators $\hat{\mu}$ and $\hat{\theta}$ correspond to the global maximum of the likelihood. The lower constraint $0 \leq \hat{\mu}$ is dictated by physics (the signal rate is positive), while the upper constraint $\hat{\mu} \leq \mu$ is imposed by hand in order to guarantee a one-sided (not detached from zero) confidence interval. Physics-wise, this means that upward fluctuations of the data such that $\hat{\mu} > \mu$ are not considered as evidence against the signal hypothesis, namely a signal with strength μ.
In the evaluation of the CL_s limit, the probability density functions (pdfs) $f(\hat{q}_\mu|\mu, \hat{\theta}_\mu)$ and $f(\hat{q}_\mu|0, \hat{\theta}_0)$ are used. The first is the pdf of $\hat{q}_\mu$ assuming a signal with strength μ in the signal+background hypothesis, while the second is for the background-only hypothesis ($\mu = 0$). The two pdfs are not known a priori. Still, the Asymptotic Formulae [103], valid in the limit of a large event sample, provide a useful approximation that avoids the estimation of them through the use of toy datasets.

We define a pair of p-values corresponding to the observed data under the signal-plus-background and the background-only hypotheses, denoted as p_μ and p_b, respectively:

$$p_\mu = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{obs}}|\text{signal+background}) = \int_{\tilde{q}_\mu^{\text{obs}}}^{\infty} f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu) d\tilde{q}_\mu, \quad (7.2)$$

$$1 - p_b = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{obs}}|\text{background-only}) = \int_{\tilde{q}_0^{\text{obs}}}^{\infty} f(\tilde{q}_\mu|0, \hat{\theta}_0) d\tilde{q}_\mu. \quad (7.3)$$

From these, $CL_s(\mu)$ is derived as:

$$CL_s(\mu) = \frac{p_\mu}{1 - p_b}. \quad (7.4)$$

The critical value of μ at which CL_s reaches 0.05 sets the exclusion threshold in our analysis, corresponding to a 95% confidence level exclusion. The $CL_s(\mu)$ metric decreases with increasing μ, thus excluding higher values of μ with greater confidence.
The expected CL_s limit on μ is evaluated by assuming $1 - p_b = 0.5$, i.e., the cumulative of the pdf for the background-only fit $f(\hat{q}_\mu|0, \hat{\theta}_0^{obs})$ crosses the quantile of 50%, which corresponds to the median of $f(\hat{q}_\mu|0, \hat{\theta}_0^{obs})$. The $\pm 1\sigma$ (68%) band is defined by the crossings of the 16% and 84% quantiles, while crossings at 2.5% and 97.5% define the $\pm 2\sigma$ (95%) band.

7.1.1 Systematic uncertainties

Systematic uncertainties are incorporated as nuisance factors within the framework of setting limits. The primary contributors to systematic variance include:

- Jet Energy Scale (JES);
- Jet Energy Resolution (JER);
- Muon Momentum Scale (MMS);
- Muon Momentum Resolution (MMR);
- Luminosity measurements.

Given that the model for the background is empirically derived, the uncertainties above are only factored in for the resonance signal assessment.
The JES and JER uncertainties (σ_{JES} and σ_{JER}) translate, respectively, into uncertainties in the position of the signal peak and its width in the muon-jet invariant mass distribution. Since we use histograms as input for our signal models, we implemented the systematic by shifting the p_T^{jet} by $\pm 2\%$ for the JES. After shifting the p_T, we recompute the invariant mass of the muon-jet system and re-apply the selection for all the categories as described in Table 6.2. Figure 7.1 shows the effect of the JES systematic on the signal for the 1muon, tight and nobtag category. The impact on the efficiency is less than 1%.
The effect of these uncertainties is propagated to the limits by fitting a morphed histogram. The JES and JER uncertainties are assumed flat in all the p_T range and

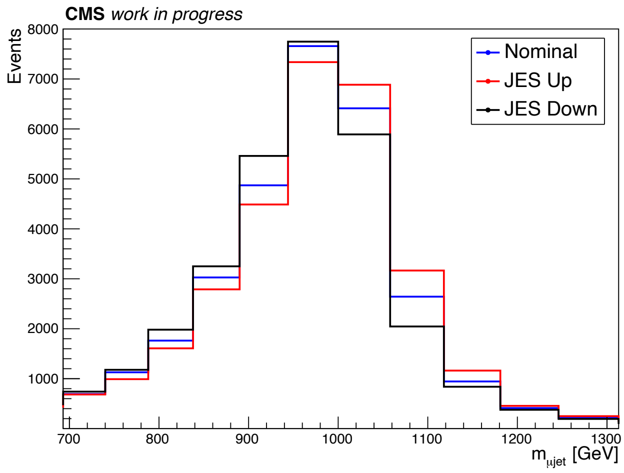


Figure 7.1. Example of JES systematic effect on the $M_{\mu jet}$. The blue curve is the histogram without any systematic applied, while the red (black) line is the effect of the up (down) shift of the p_T^{jet} on the $M_{\mu jet}$ distribution for the 1 muon, tight and nobtag category. The JES effect on the efficiency is less than 1%.

are respectively a shift of ±2% on the peak position and a smearing of the p_T of ±10%. These values have been evaluated by previous CMS analyses (as [81]) for anti-k_T jets with a distance parameter R = 0.4.

The MMS and MMR resolution affect the muon-jet invariant mass distribution similarly to the jet case discussed above. So, respectively, a shift in the peak and the widening of the signal distribution both depend on the muon's energy. The maximum effect is a shift of the mass shape of ±2% and a variation on the width of ±8%.

The uncertainty on the integrated luminosity is 1.6% [89, 71, 75], and it is propagated to the normalization of the signal, in addition to the other uncertainties on the normalization.

The combined effect of all the uncertainties above on the limit is estimated to be below $\sim$ 10% for all the signal hypotheses considered.

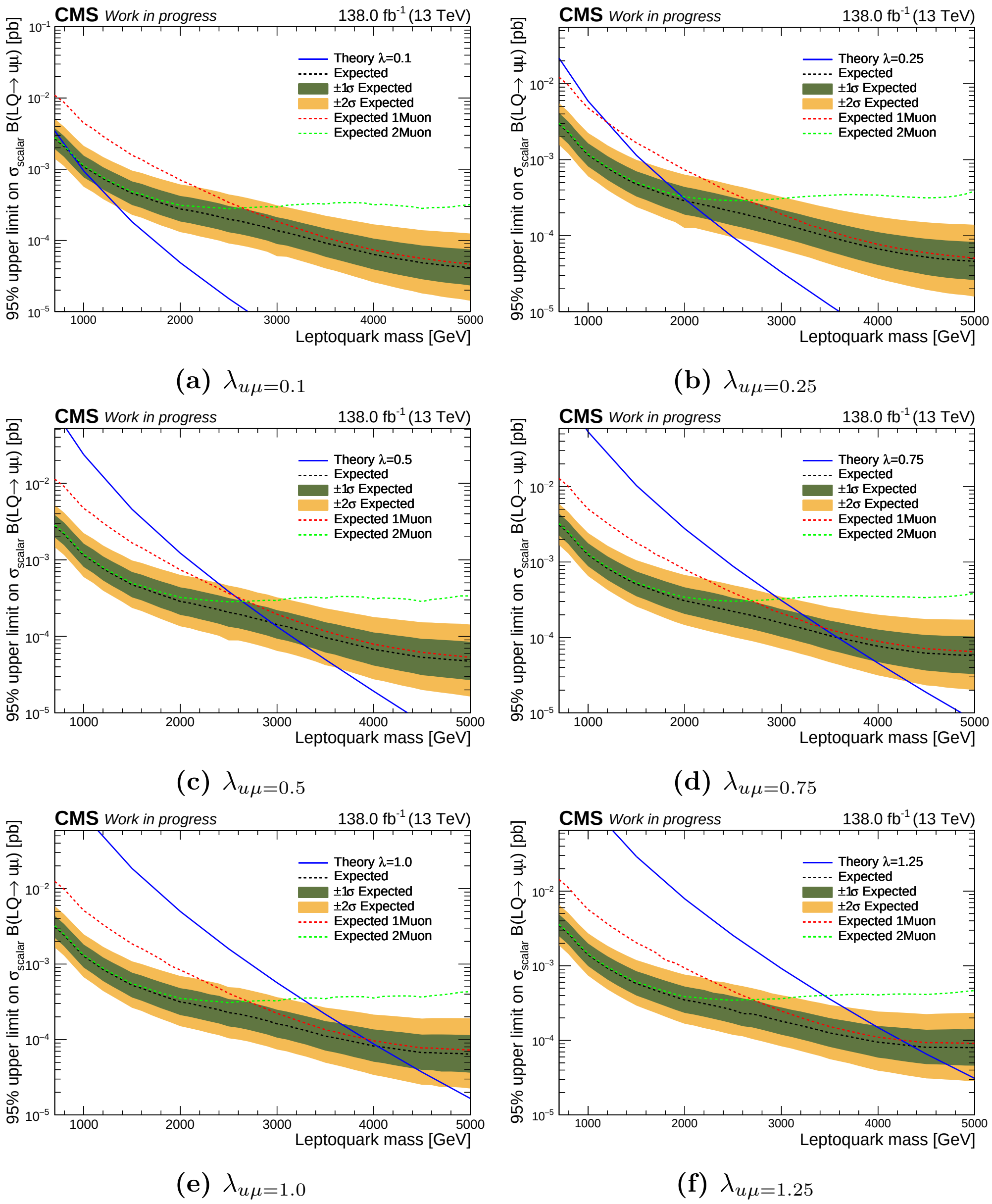
We do not consider additional uncertainty on the background shape parameters because they are all considered nuisance parameters distributed with a flat prior around the best-fit values in a sufficiently large range, for which the limit is found to be stable. Moreover the correction to the physics object used to compute the invariant mass spectrum, i.e. jets and muon, show a linear trend, as visible in Figures 4.4. Therefore there are no uncertainties due to the physics objects correction that could affect the falling spectrum of the muon-jet invariant mass distribution, so no uncertainty from the data is considered.

A source of systematic uncertainty on the signal can arise from biases due to the choice of a specific functional form for the background modeling. In Section 6.4.2, we show that the bias introduced by the chosen fit function is well below 50% of the statistic uncertainty; therefore, it is considered negligible for the reasons described in

Section 6.4.2. Thus, we do not need to introduce systematic uncertainties to account for the bias.

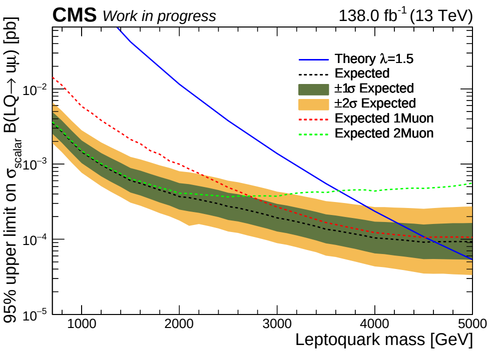
7.1.2 Upper limits on LQ signal cross-section

In this section, we report the final results and the exclusion limits that our analysis can set on the signal cross section for LQ($u\mu$) ed LQ($b\mu$) models introduced in Section 3.1.
Figures 7.2 and 7.3 show the expected limits on the LQ production cross-section for a LQ($u\mu$) model. Each Figure shows the expected limit and its uncertainty bands as a function of the LQ mass for the different λ values tested. The low mass region ($M_{LQ} < 2.5$ TeV) is dominated by the 2 muon signal region, while above the 1 muon signal region dominates. This is due to the almost zero background; thus, the greater signal efficiency of the 1-muon signal region results in a better sensitivity with respect to the 2-muon signal region. Due to the smaller cross-section, despite lower expected limits, the LQ mass we are able to exclude for low couplings is smaller than the one for high couplings.
Figures 7.4 and 7.5 show the expected limits on the LQ production cross-section for a LQ($b\mu$) model. We do not show the limits for $M_{LQ} > 3.5$ TeV since above that mass, for some categories, it is not possible to observe a peak in the signal shape due to the high tails at low mass, especially at high coupling ($\lambda \geq 0.5$). In the observed mass range, the same considerations made for the LQ($u\mu$) model are valid.

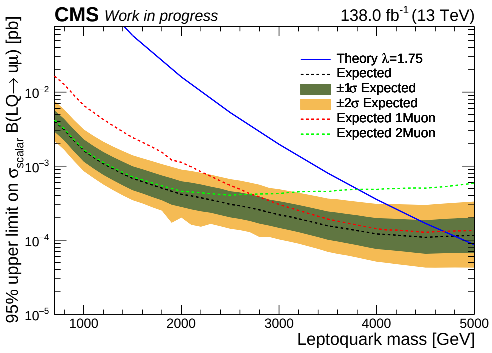


(a) $\lambda_{u\mu}$=0.1 (b) $\lambda_{u\mu}$=0.25 (c) $\lambda_{u\mu}$=0.5 (d) $\lambda_{u\mu}$=0.75 (e) $\lambda_{u\mu}$=1.0 (f) $\lambda_{u\mu}$=1.25

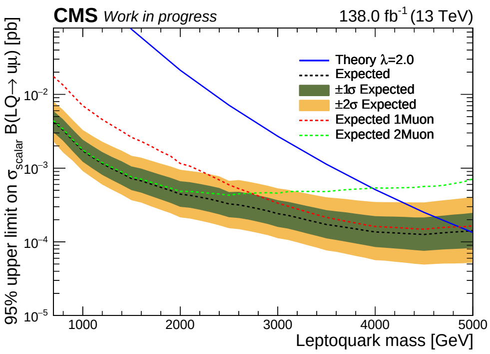
Figure 7.2. Expected (dashed line) 95% CLs limits on $\sigma(LQ)$ as a function of M_{LQ}, for fixed λ. Uncertainty bands ($\pm 1\sigma, \pm 2\sigma$) on expected limits are also shown. The dashed red (green) line shows the expected limits for the 1 muon (2 muons) signal region.



(a) $\lambda_{u\mu=1.5}$

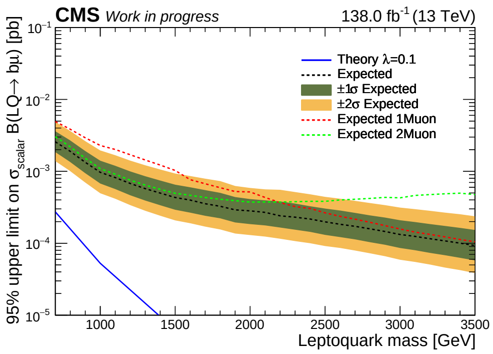


(b) $\lambda_{u\mu=1.75}$

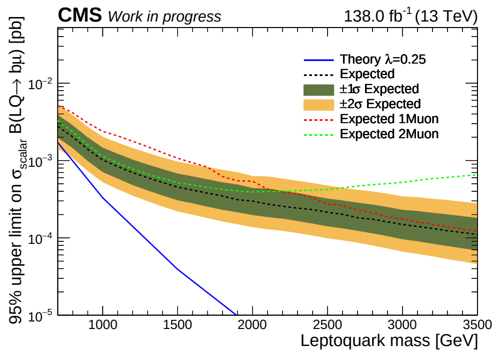


(c) $\lambda_{u\mu=2.0}$

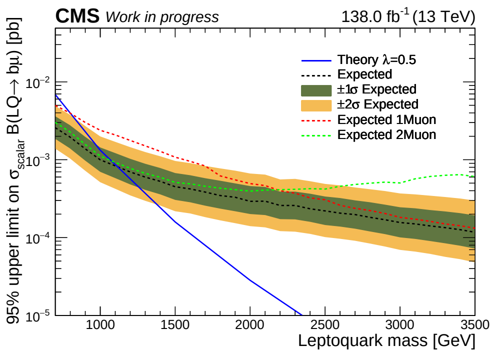
Figure 7.3. Expected (dashed line) 95% CLs limits on $\sigma(LQ)$ as a function of M_{LQ}, for fixed λ. Uncertainty bands ($\pm1\sigma, \pm2\sigma$) on expected limits are also shown. The dashed red (green) line shows the expected limits for the 1 muon (2 muons) signal region.



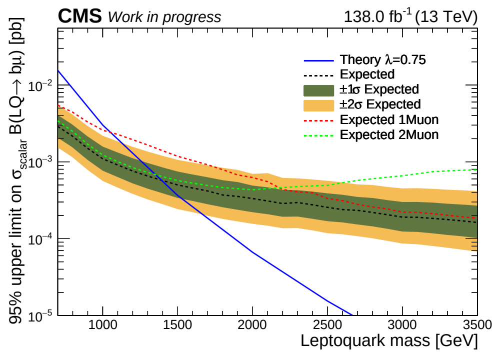
(a) $\lambda_{b\mu=0.1}$



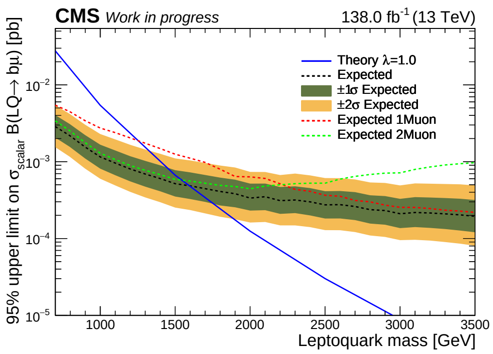
(b) $\lambda_{b\mu=0.25}$



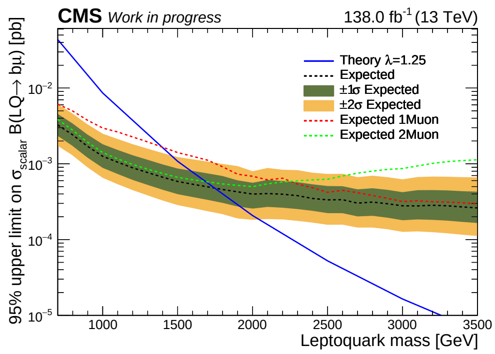
(c) $\lambda_{b\mu=0.5}$



(d) $\lambda_{b\mu=0.75}$

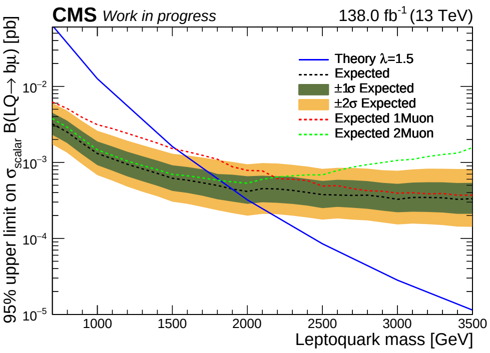


(e) $\lambda_{b\mu=1.0}$

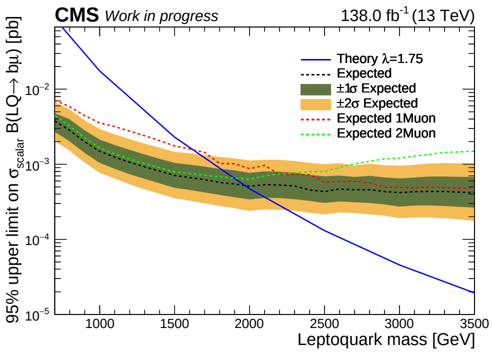


(f) $\lambda_{b\mu=1.25}$

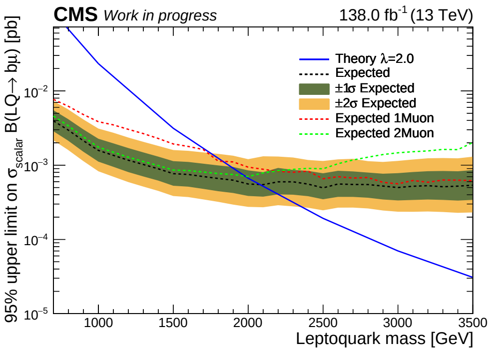
Figure 7.4. Expected (dashed line) 95% CLs limits on $\sigma(LQ)$ as a function of M_{LQ}, for fixed λ. Uncertainty bands ($\pm1\sigma, \pm2\sigma$) on expected limits are also shown. The dashed red (green) line shows the expected limits for the 1 muon (2 muons) signal region.



(a) $\lambda_{b\mu=1.5}$



(b) $\lambda_{b\mu=1.75}$



(c) $\lambda_{b\mu=2.0}$

Figure 7.5. Expected (dashed line) 95% CLs limits on $\sigma(LQ)$ as a function of M_{LQ}, for fixed λ. Uncertainty bands ($\pm1\sigma, \pm2\sigma$) on expected limits are also shown. The dashed red (green) line shows the expected limits for the 1 muon (2 muons) signal region.

The expected limits can also be seen as a function of the M_{LQ} and λ in Figure 7.6 for a LQ($u\mu$) hypotheses. The black area shown in the figure marks the contour of the region of the λ-M_{LQ}-plane corresponding to the signal hypotheses that our search can exclude. The exclusion is obtained by comparing the expected limit with the theoretical prediction of $\sigma(LQ)$, evaluated for a specific choice of the couplings of the model. The signal hypotheses with the expected limits below the theoretical predictions are excluded.
Figure 7.6 also shows the current exclusion due to a CMS LQ pair production analysis [79]. The exclusion area is a rectangle since the LQ pair-production cross-section is independent of the coupling. The comparison between the black and the blue line clearly shows that the analysis presented here is much more sensitive to the model under study with respect to the CMS pair production search, especially for high values LQ mass and coupling (M_{LQ}>1.6 TeV and $\lambda \geq$0.2).
Figure 7.7 shows the expected limits as a function of M_{LQ} and λ for a heavy quark

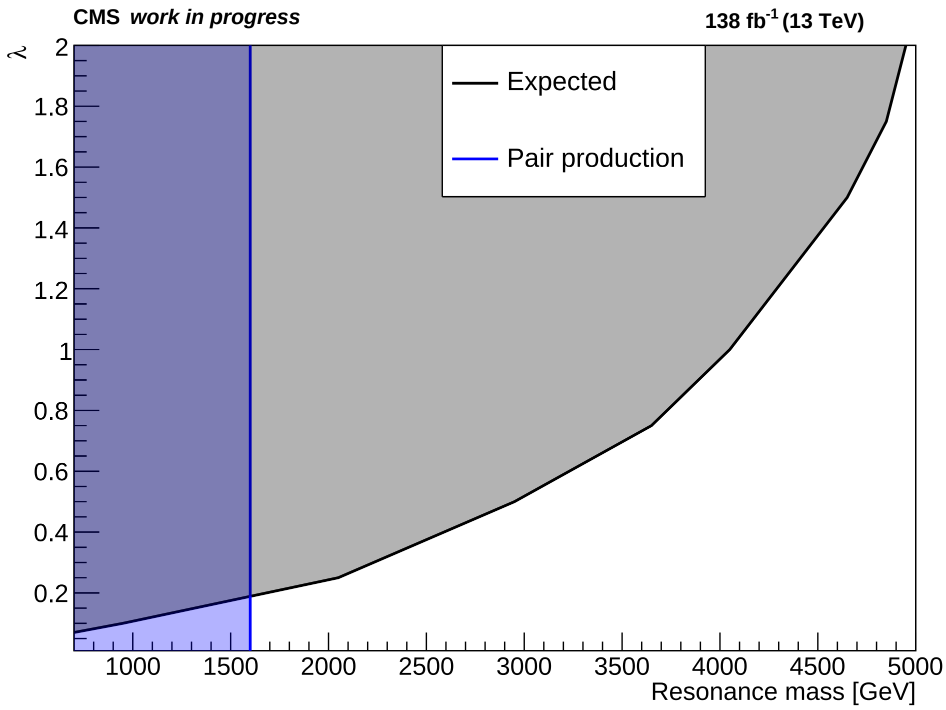


Figure 7.6. Expected upper limits on $\sigma(LQ)$, as a function of $\lambda_{u\mu}$ vs. M_{LQ}, for a LQ decaying into muons and light quark. The excluded regions are compared with those obtained from previous CMS searches [79]. The presented analysis can improve the CMS sensitivity in the high mass ($M_{LQ} > 1.5$ TeV) and high coupling ($\lambda \geq 0.2$).

scenario. The comparison between the black and the blue line clearly shows that the analysis presented here is more sensitive to the model under study with respect to the CMS pair production search, especially for high values LQ mass and coupling (M_{LQ}>1.8 TeV and λ >1).
We can conclude that the analysis presented here could enhance the sensitivity to new LQ particles decaying into muons and both light and heavy quarks with

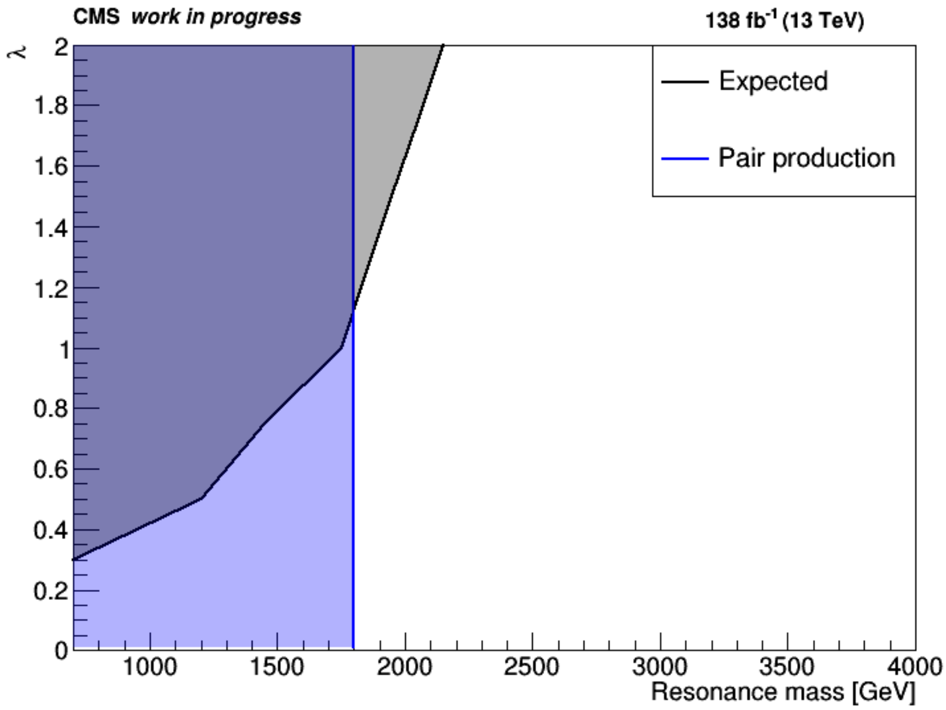


Figure 7.7. Expected upper limits on $\sigma(LQ)$, as a function of $\lambda_{b\mu}$ vs. M_{LQ}, for a LQ decaying into muons concerning. The excluded regions are compared with those obtained from previous CMS searches [93]. The presented analysis can improve the CMS sensitivity in the high mass ($M_{LQ} > 1.8$) TeV and high coupling ($\lambda > 1$).

respect to previous CMS searches. In the benchmark model studied, the excluded limit on the LQ mass is extended by up to 2 TeV for the light quark scenario concerning previous resonant searches depending on the coupling hypotheses. The analysis presented here in the heavy quark scenario can enhance the sensitivity to LQ particles in the high mass and coupling regime.

7.2 Unblinding of 1% of the dataset

Since the analysis has not yet been approved by the CMS Collaboration, it is not possible, according to CMS standard procedures, to look at the data in the signal region. Therefore, we have unblinded only a small fraction of the data (1% the complete Run 2 statistics) since we do not expect to be more sensitive than previous analyses with these statistics. However, this allows us to show all steps of the analysis from beginning to end using the data. A test selecting randomly 1% of all statistics of the 3 years of data-taking and looking at the data in the whole mass range was performed. It was possible to see the data in the whole mass range instead of up to 1 TeV. However, due to the limited statistics available, all the 2 muon signal region categories are summed together in a single category since they did not have enough events. The function used to estimate the background is the standard dijet function (with 3 parameters) since it is used to evaluate the expected limits.

To evaluate the fitting robustness, as for the full dataset, we perform a goodness of

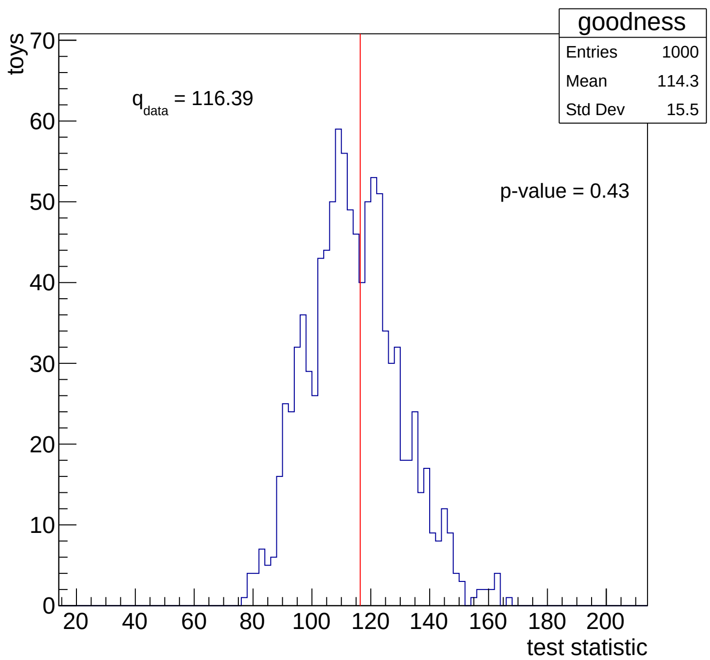


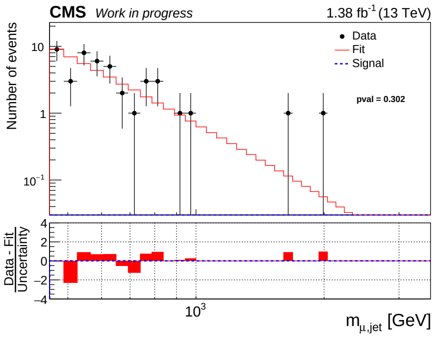
Figure 7.8. Distributions of the test statistic (q) obtained from 1000 toy experiments for the background estimation obtained using the standard dijet function. The red vertical lines mark the test statistic values evaluated from data (qdata).

fit test for all event categories using the function f_1 in the hypotheses of background only. Fig 7.8 illustrates the q distribution and the corresponding p-value, confirming a good fit quality.

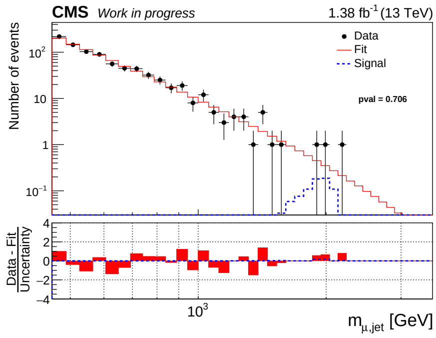
Fig 7.9 shows the fit to the data unblinded in the whole mass range using background-only fit. A LQ($u\mu$) signal with a mass of 2 TeV is superimposed on the plot. The bottom part shows the difference between the background estimated by the fit function and the data divided by the statistical uncertainty for each bin. It is possible to see that the agreement between the data and the background estimation is good overall, but a slight excess in the data can be seen around 2 TeV. The excess is present in a region not yet excluded by previous searches in for all the signal hypotheses shown.

Figures 7.10 and 7.11 show the obtained limits on a LQ particle decaying into a muon and a light quark. Each figure shows the observed limit, the expected limit, and its uncertainty bands as a function of the M_{LQ} for the different λ tested. A slight excess at 2σ level (local significance) is found for a M_{LQ} of $\sim$ 2 TeV for all the coupling considered. The events contributing to the excess are visible in Figure 7.9. The excess found is in a mass region not excluded by previous analysis.

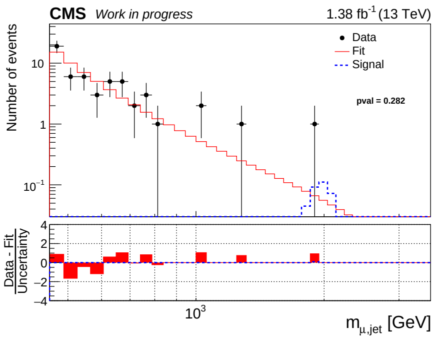
Figures 7.12 and 7.13 show the obtained limits for a LQ($b\mu$) model. Each Figure shows the observed limit, the expected limit, and its uncertainty bands as a function of the M_{LQ} for the different λ tested. We obtained the same results of the LQ($u\mu$) model.



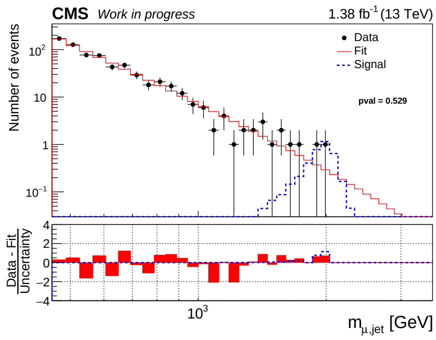
(a) 1 Muon signal region: loose BDT cut and btag on leading jet



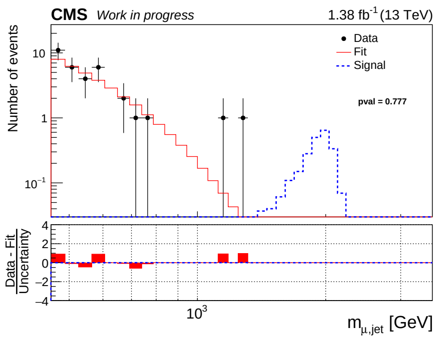
(b) 1 Muon signal region: loose BDT cut and no-btag on leading jet



(c) 1 Muon signal region: tight BDT cut and btag on leading jet

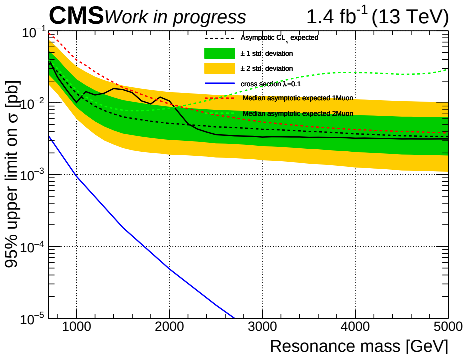


(d) 1 Muon signal region: tight BDT cut and no-btag on leading jet

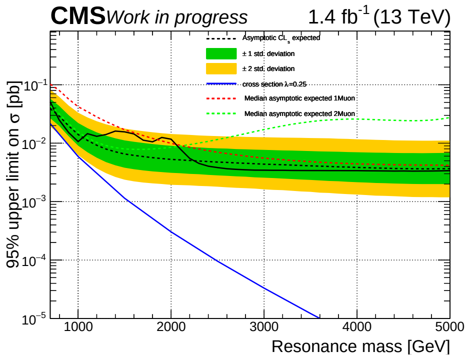


(e) 2 Muon signal region: all categories summed together

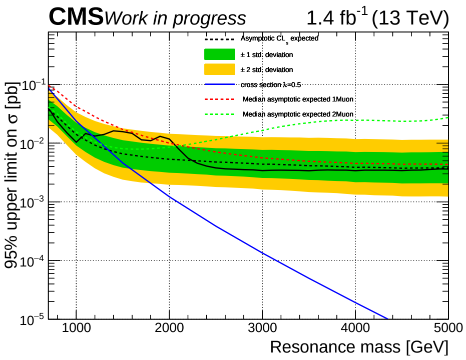
Figure 7.9. Background fit for the standard dijet function for all the categories. The black points are the unblinded CMS data since we are using only 1% of the full Run 2 statistics. The red line is the background estimation obtained using the standard dijet function. The blue line represents a signal with a $M_{LQ} = 2$ TeV and $\lambda = 1$, which is normalized to the observed limit cross-section. The horizontal axis shows the value of the $M_{\mu jet}$ spectra, while the vertical axis shows the number of events in each bin. The bottom part shows the difference between the data and the background fit function divided by the statistical uncertainty.



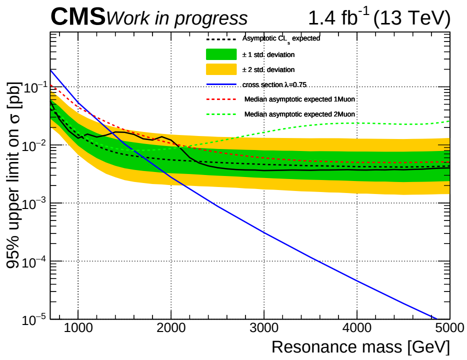
(a) $\lambda_{u\mu=0.1}$



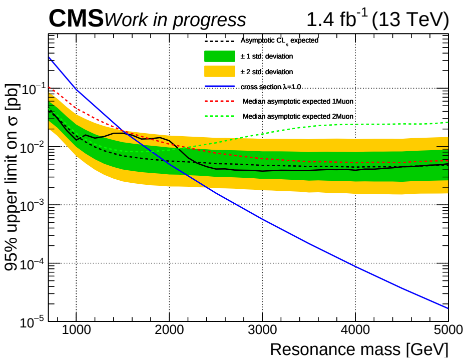
(b) $\lambda_{u\mu=0.25}$



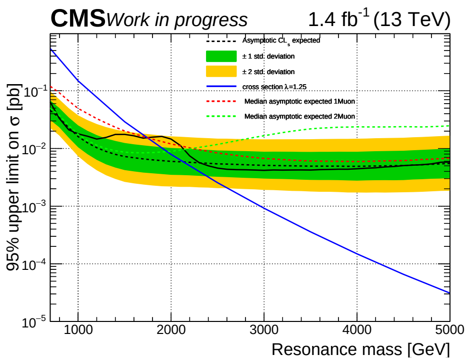
(c) $\lambda_{u\mu=0.5}$



(d) $\lambda_{u\mu=0.75}$

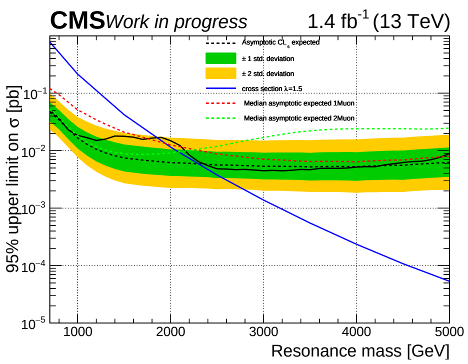


(e) $\lambda_{u\mu=1.0}$

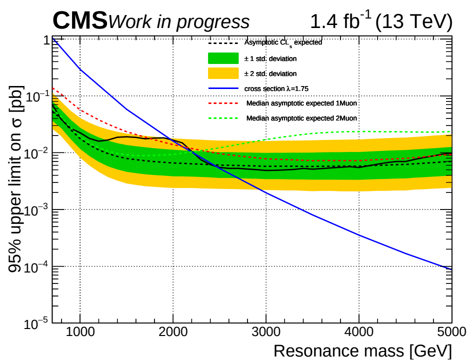


(f) $\lambda_{u\mu=1.25}$

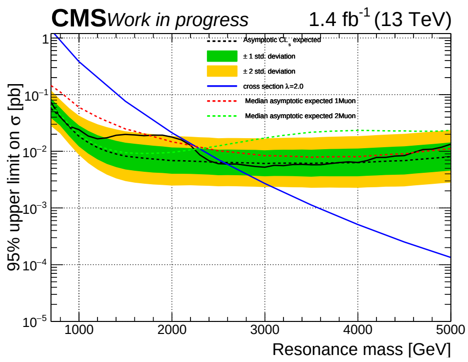
Figure 7.10. Observed (full line), expected (dashed line) 95% CLs limits on $\sigma(LQ)$ as a function of M_{LQ}, for fixed λ. Uncertainty bands ($\pm 1\sigma, \pm 2\sigma$) on expected limits are also shown. The dashed red (green) line shows the expected limits for the 1 muon (2 muons) signal region.



(a) $\lambda_{u\mu=1.5}$

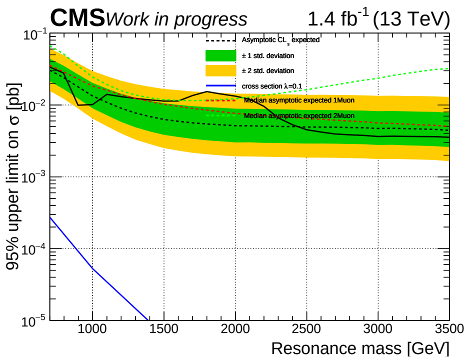


(b) $\lambda_{u\mu=1.75}$

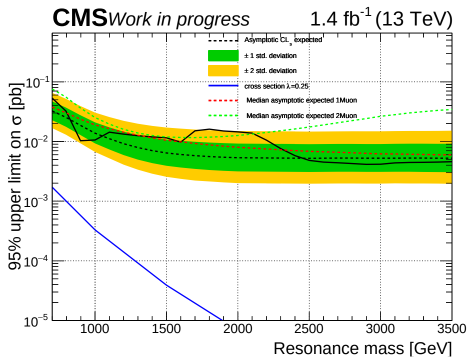


(c) $\lambda_{u\mu=2.0}$

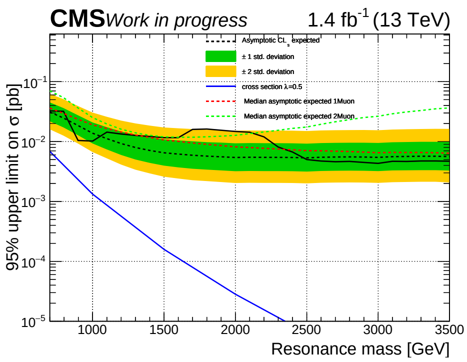
Figure 7.11. Observed (full line), expected (dashed line) 95% CLs limits on $\sigma(LQ)$ as a function of M_{LQ}, for fixed λ. Uncertainty bands ($\pm 1\sigma, \pm 2\sigma$) on expected limits are also shown. The dashed red (green) line shows the expected limits for the 1 muon (2 muons) signal region.



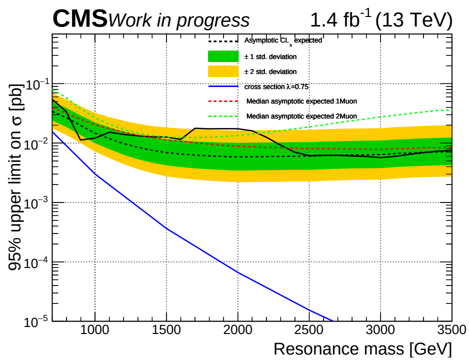
(a) $\lambda_{b\mu=0.1}$



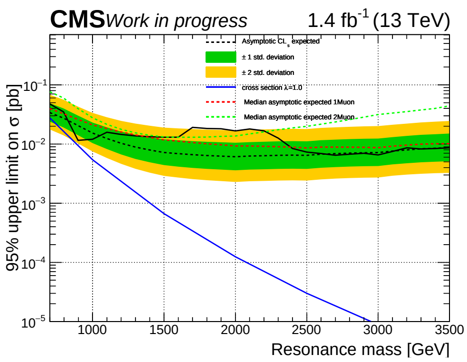
(b) $\lambda_{b\mu=0.25}$



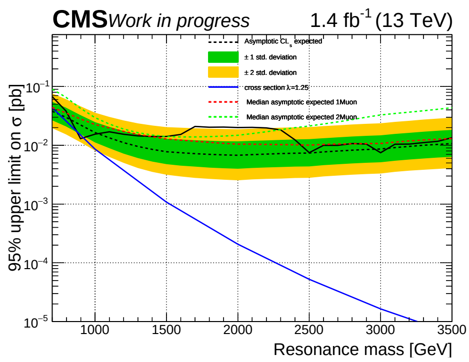
(c) $\lambda_{b\mu=0.5}$



(d) $\lambda_{b\mu=0.75}$

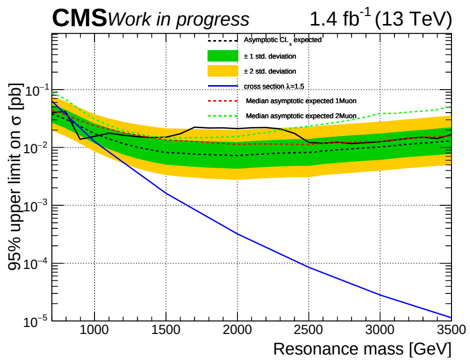


(e) $\lambda_{b\mu=1.0}$

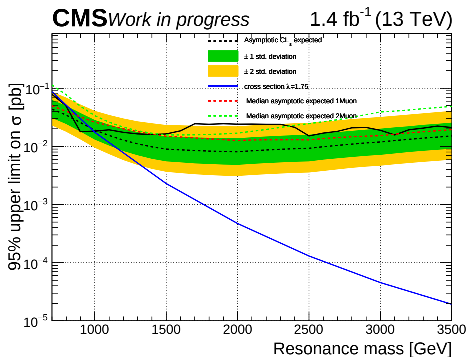


(f) $\lambda_{b\mu=1.25}$

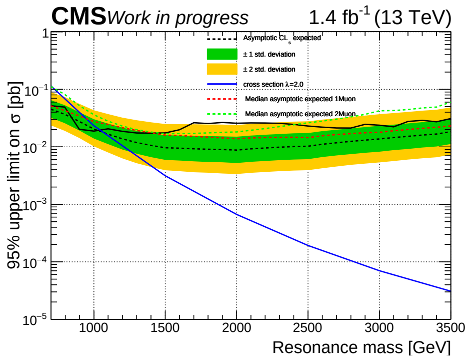
Figure 7.12. Observed (full line), expected (dashed line) 95% CLs limits on $\sigma(LQ)$ as a function of M_{LQ}, for fixed λ. Uncertainty bands ($\pm 1\sigma, \pm 2\sigma$) on expected limits are also shown. The dashed red (green) line shows the expected limits for the 1 muon (2 muons) signal region.



(a) $\lambda_{b\mu=1.5}$



(b) $\lambda_{b\mu=1.75}$



(c) $\lambda_{b\mu=2.0}$

Figure 7.13. Observed (full line), expected (dashed line) 95% CLs limits on $\sigma(LQ)$ as a function of M_{LQ}, for fixed λ. Uncertainty bands ($\pm 1\sigma, \pm 2\sigma$) on expected limits are also shown. The dashed red (green) line shows the expected limits for the 1 muon (2 muons) signal region.

Chapter 8

Conclusions

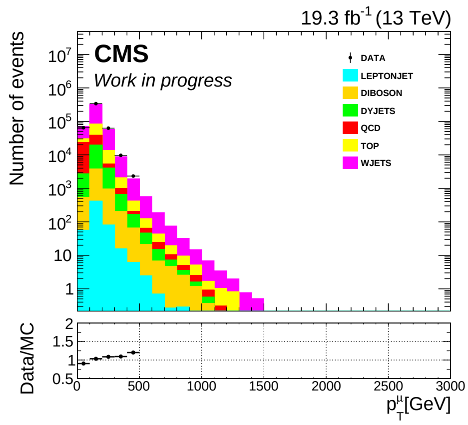
This dissertation presents a novel search for new physics beyond the Standard Model using the data collected by the CMS experiment at CERN during the LHC Run 2. This work extends the search for leptoquarks (LQs), hypothetical particles foreseen by some extensions of the Standard Model, exploiting a novel production mechanism in proton-proton collision at LHC. Due to quantum fluctuations, protons also contain charged leptons, making it possible to study lepton-induced processes. By picking a lepton from one beam and a quark from the other beam, it becomes possible to study the resonant single LQ production and the corresponding decay ($l+q \rightarrow LQ \rightarrow l+q$). The analysis uses data uses data from proton-proton collisions produced at the CERN LHC at a center-of-mass energy of 13 TeV with an integrated luminosity of 138 fb^{-1}.
The analysis targets LQs with masses at the TeV scale, focusing on the muon channel. The final state is composed of a high-p_T jet and muon coming from the LQ decays. An additional signal region has been added if a second muon, which is usually soft (low p_T) and produced typically at higher η compared to the first one, is inside the detector acceptance and reconstructed. The kinematic properties of the signal final state were used to establish preselection criteria for both signal regions (1 and 2 muons) to select LQ events. The analysis sensitivity is optimized through a BDT to preserve the signal efficiency while rejecting the background for both signal regions. For each signal region, two categories were defined based on the output of the BDT score to maximize the signal significance. All categories are then split based on the b-tagging signature of the leading jet, resulting in a total 8 independent event categories, which are used simultaneously to search for LQs.
A signal+background fit is used to search in the invariant mass of the muon+jet system for excesses above the exponentially decaying background that has been estimated via an empirical function, compatible with the presence of a signal peak. The fit quality is found to be good, and it was verified that no bias was introduced with the signal+background fit.
Since the LQ mass and lepton-quark-LQ (λ) coupling are unknown, we scanned a wide range of values. We searched from 1 TeV (compatible with current limits on LQs) up to 5 TeV for LQ couplings ranging from 0.1 to 2.0. Expected exclusion limits are derived through a fit on the invariant mass of the jet and muon system in the hypothesis of light (u) or heavy (b) quarks. The analysis is currently in the

approval phase within the CMS Collaboration. For this reason, the final results, including the data spectra in the signal regions, cannot be shown in public yet. Expected upper limits on signal cross-section limits are determined to provide an estimate of the physics reach of this new search. These results show that is possible to enhance the the lower limits on LQ mass from 1.6 TeV up to 2 TeV (5 TeV) for $\lambda = 0.2$ (2.0) in the light quark scenario. In the hypothesis of an LQ decaying into a b quark, the search presented is expected to exclude at 95% confidence level (CL) LQs up to 2 TeV for $\lambda > 1$ compared to 1.8 TeV from a previous limit from an LQ pair production analysis.
Finally, since it was not possible to show the result with the complete dataset, a study looking at data in the whole mass range with only 1% of the 138 fb^{-1} of the full Run 2 was performed. Beyond confirming the analysis method, this study shows a small excess at 2σ level for $M_{LQ} \sim 2$ TeV both in the $M_{\mu jet}$ mass spectrum and in the observed limit.
This novel production mechanism improves the sensitivity of LQ searches by extending the discovery power for LQs, particularly in regions of lower couplings where previous analyses were less sensitive. Consequently, this work could significantly impact the CMS LQ search strategy (especially at high mass and coupling), setting a precedent for future analyses in the field of high-energy particle physics.

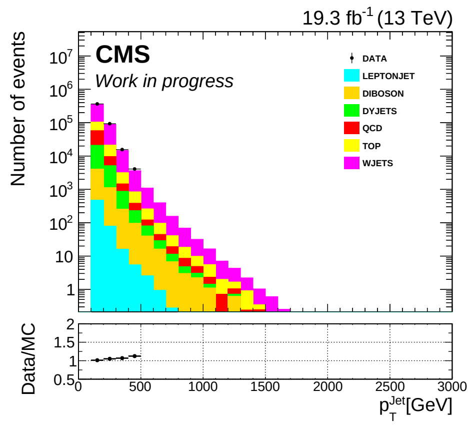
Appendix A

Kinematic distributions for 2016 and 2017 datasets

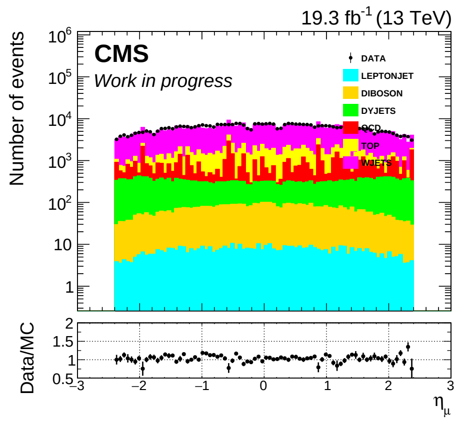
In this section, the same kinematic distributions shown for the 2018 dataset in Section 5.5 are reported for the 2016 and 2017 datasets. All the datasets show good agreement between data and MC, and all the kinematic distributions have the same behavior for all datasets.



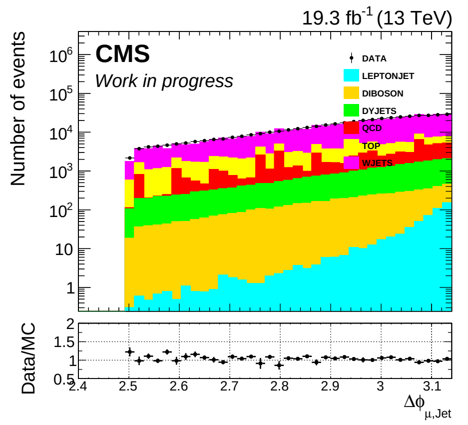
(a) p_T^{μ}



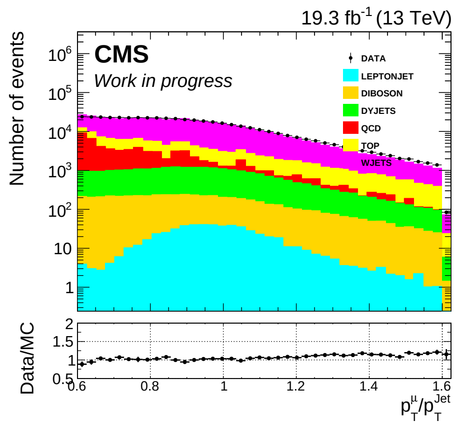
(b) p_T^{jet}



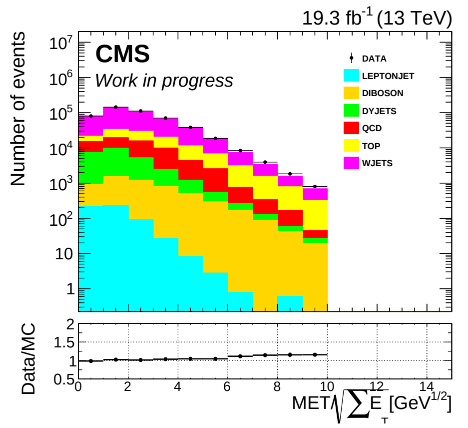
(c) η_μ



(d) $\Delta\phi_{\mu jet}$

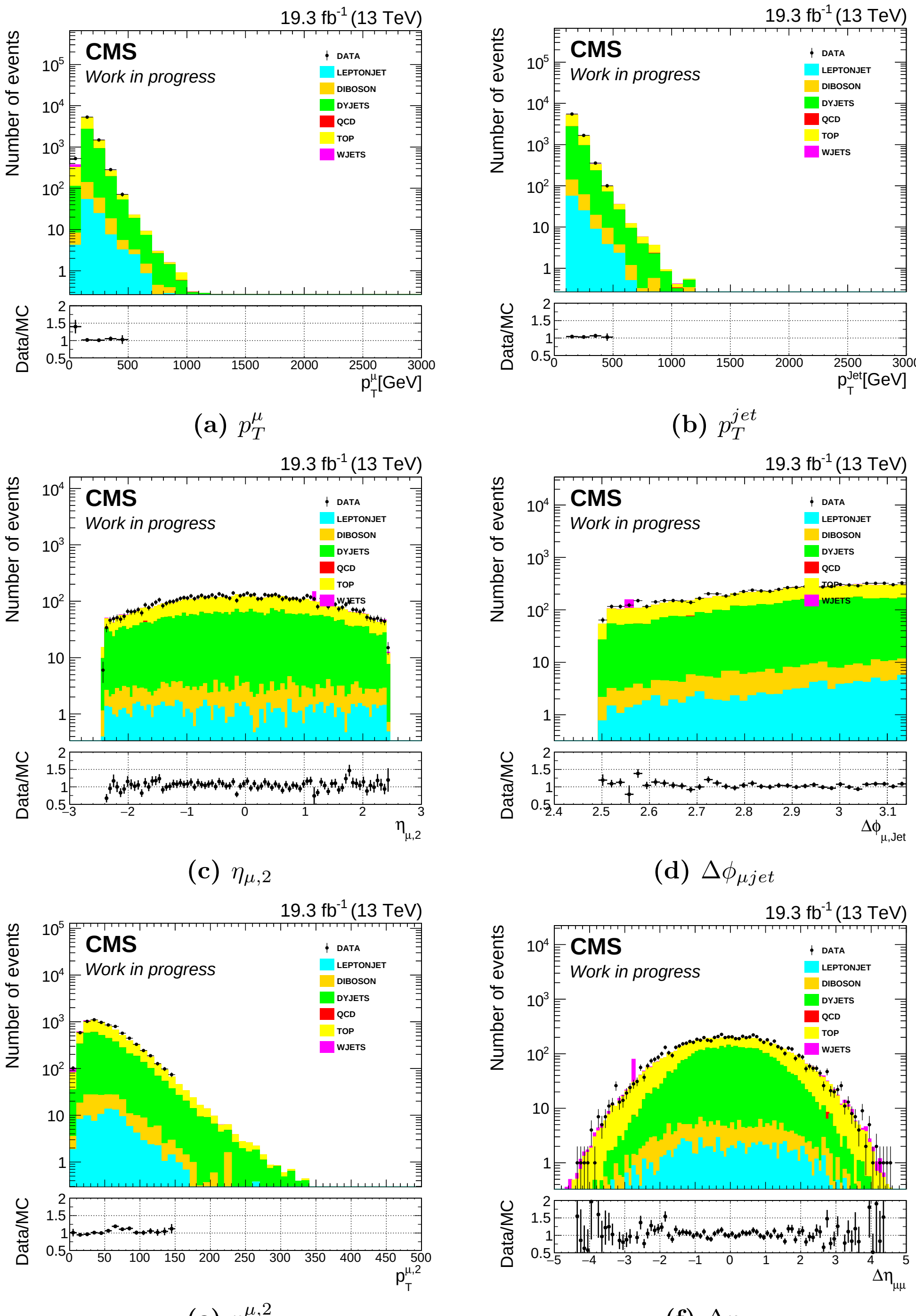


(e) p_T^{μ}/p_T^{jet}



(f) MET Significance

Figure A.1. Comparison between 2016preVFP data and simulated samples for some interesting distributions in the 1 Muon signal region at preselection.

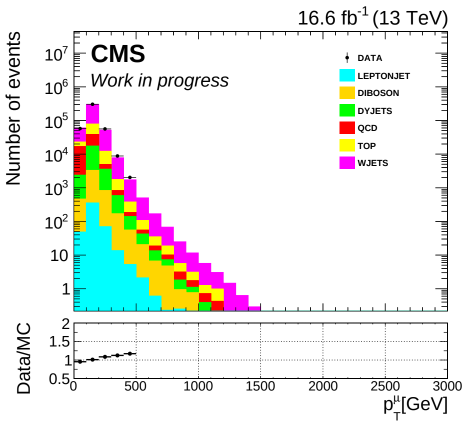


(a) p_T^{μ} **(b)** p_T^{jet}

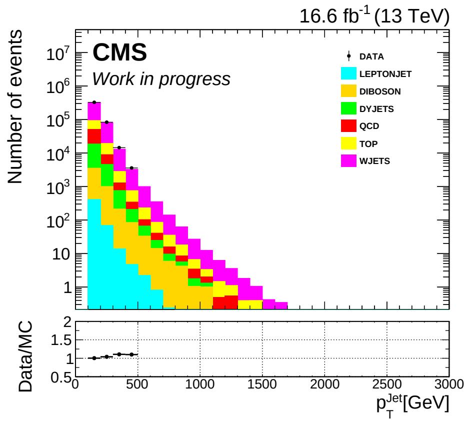
(c) $\eta_{\mu,2}$ **(d)** $\Delta\phi_{\mu jet}$

(e) $p_T^{\mu,2}$ **(f)** $\Delta\eta_{\mu\mu}$

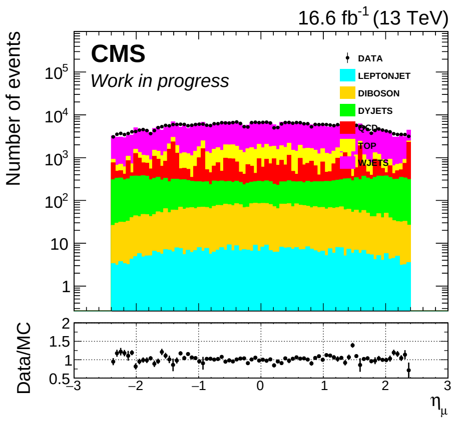
Figure A.2. Comparison between 2016preVFP data and simulated samples for interesting distributions in the 2 Muons signal region at preselection.



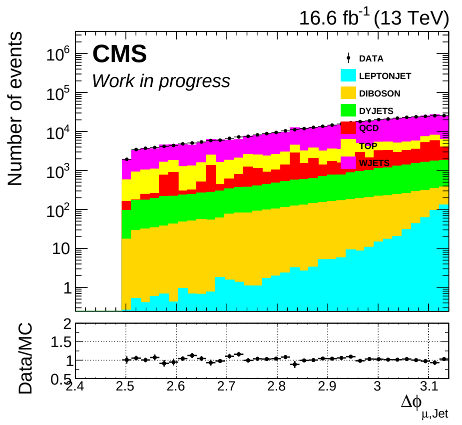
(a) p_T^{μ}



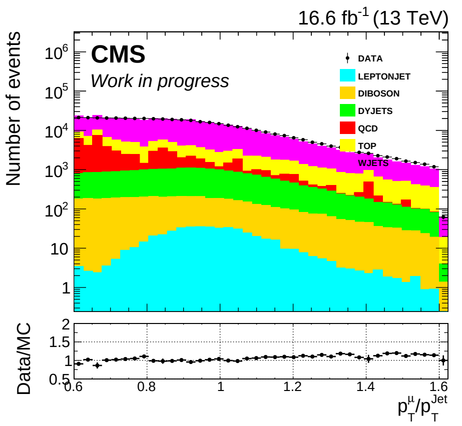
(b) p_T^{jet}



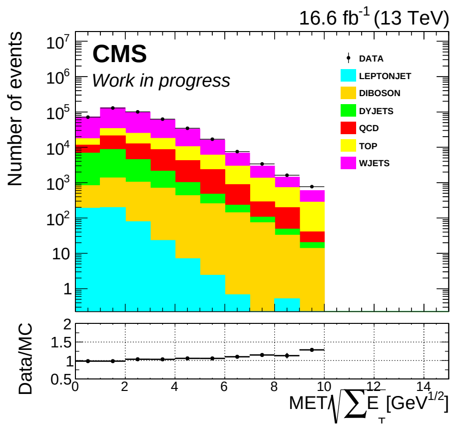
(c) η_{μ}



(d) $\Delta\phi_{\mu jet}$

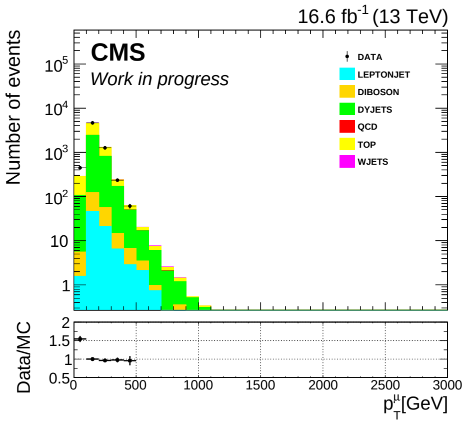


(e) p_T^{μ}/p_T^{jet}

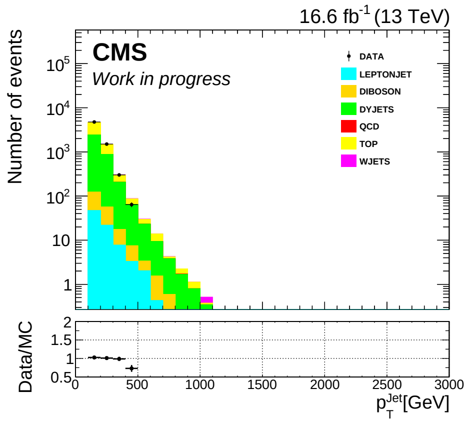


(f) MET Significance

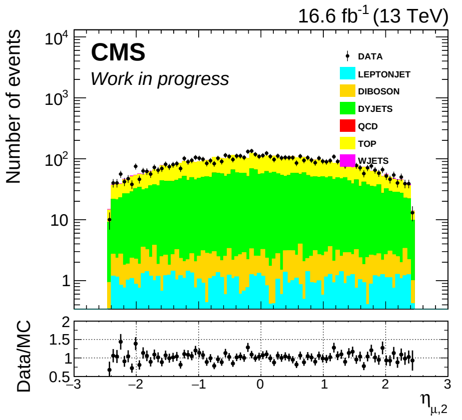
Figure A.3. Comparison between 2016postVFP data and simulated samples for some interesting distributions in the 1 Muon signal region at preselection.



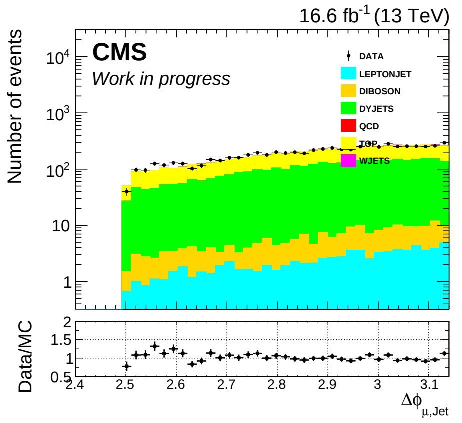
(a) p_T^{μ}



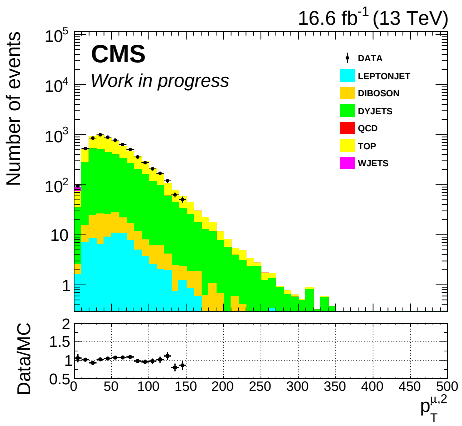
(b) p_T^{jet}



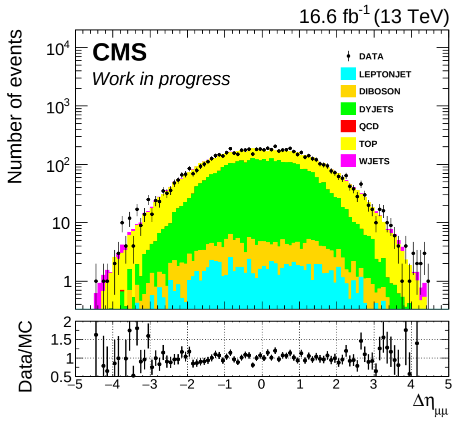
(c) $\eta_{\mu,2}$



(d) $\Delta\phi_{\mu jet}$

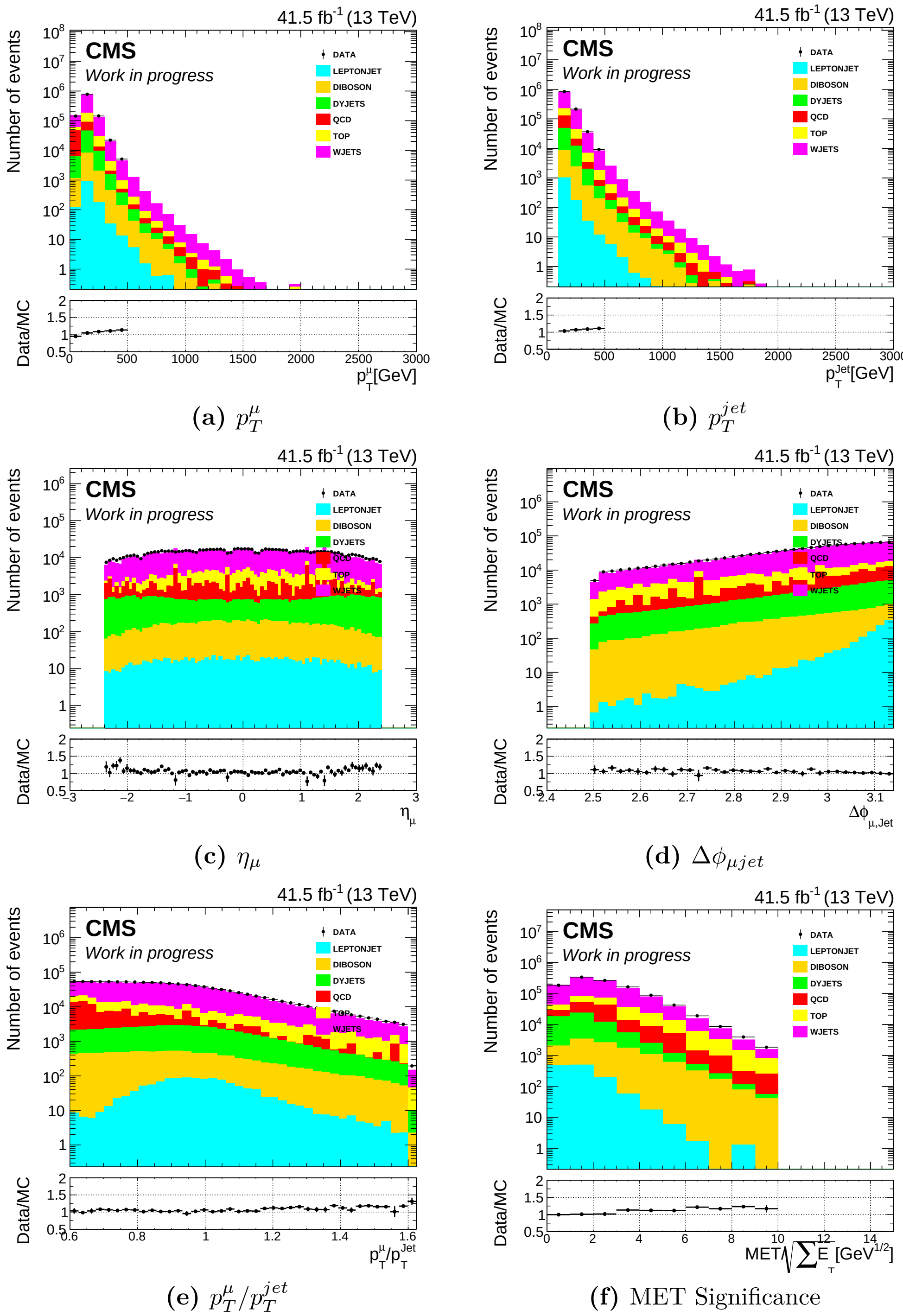


(e) $p_T^{\mu,2}$



(f) $\Delta\eta_{\mu\mu}$

Figure A.4. Comparison between 2016postVFP data and simulated samples for interesting distributions in the 2 Muons signal region at preselection.

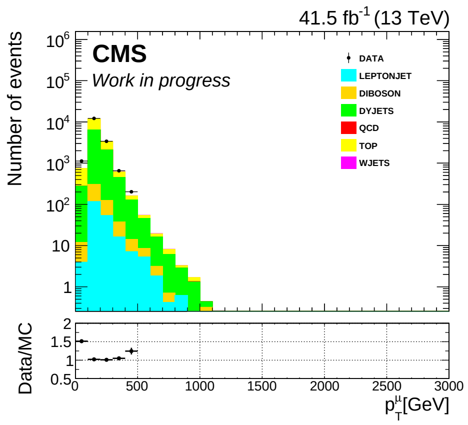


(a) p_T^{μ} **(b)** p_T^{jet}

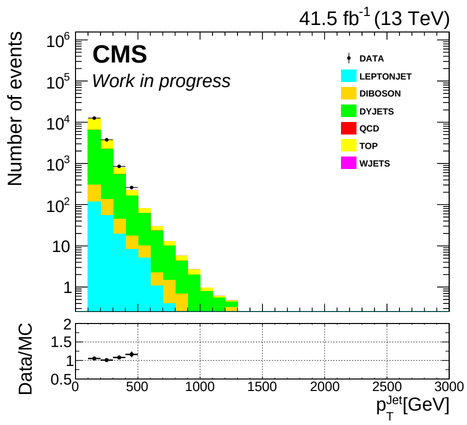
(c) η_{μ} **(d)** $\Delta\phi_{\mu jet}$

(e) p_T^{μ}/p_T^{jet} **(f)** MET Significance

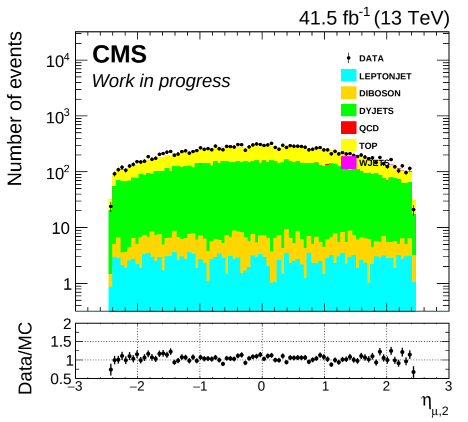
Figure A.5. Comparison between 2017 data and simulated samples for some interesting distributions in the 1 Muon signal region at preselection.



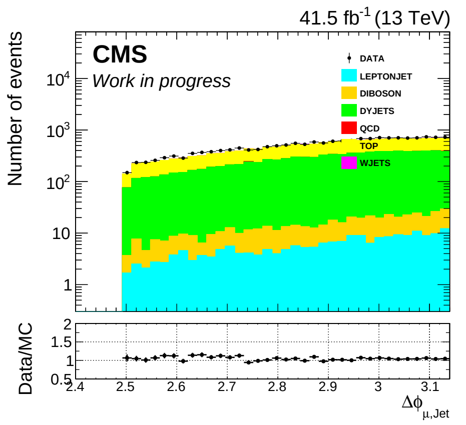
(a) p_T^{μ}



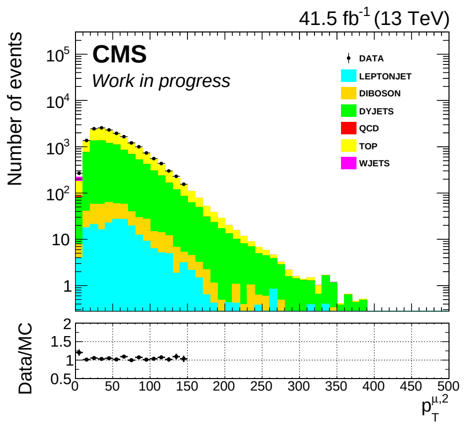
(b) p_T^{jet}



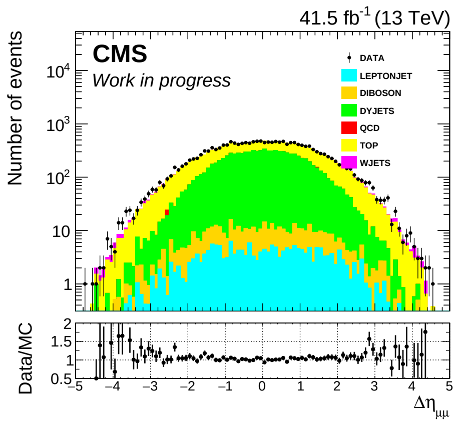
(c) $\eta_{\mu,2}$



(d) $\Delta\phi_{\mu jet}$



(e) $p_T^{\mu,2}$



(f) $\Delta\eta_{\mu\mu}$

Figure A.6. Comparison between 2017 data and simulated samples for interesting distributions in the 2 Muons signal region at preselection.

Appendix B

Binning for $M_{\mu jet}$ spectra

For the $M_{\mu jet}$ spectra shown in this thesis, we chose the following binning:

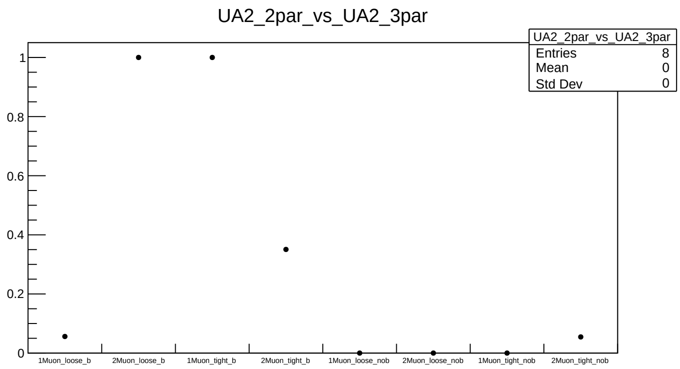
$$\begin{aligned}\text{bin edges} = [&1, 3, 6, 10, 16, 23, 31, 40, 50, 61, 74, 88, 103, 119, 137, 156, 176, 197, 220, 244,\\ &270, 296, 325, 354, 386, 419, 453, 489, 526, 565, 606, 649, 693, 740, 788, 838, 890, 944, 1000,\\ &1058, 1118, 1181, 1246, 1313, 1383, 1455, 1530, 1607, 1687, 1770, 1856, 1945, 2037, 2132, 2231,\\ &2332, 2438, 2546, 2659, 2775, 2895, 3019, 3147, 3279, 3416, 3558, 3704, 3854, 4010, 4171, 4337,\\ &4509, 4686, 4869, 5058, 5253, 5455, 5663, 5877, 6099, 6328, 6564, 6808, 7150]\end{aligned} \tag{B.1}$$

The binning has been fixed in the previous dijet analysis in CMS, and the bin size corresponds roughly to the experimental resolution at that mass.

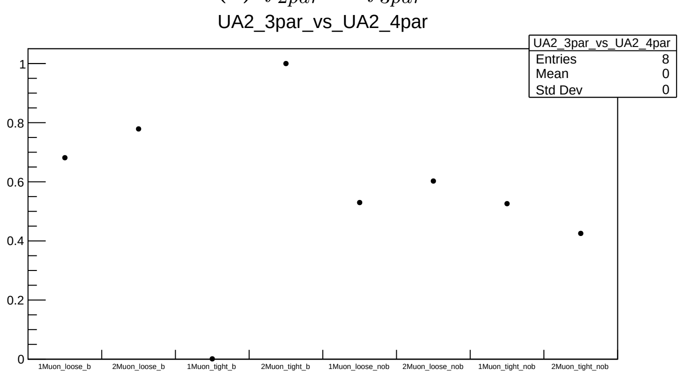
Appendix C

UA2 function distributions

In this section, the distributions that are not shown in Chapter 6 for the UA2 functions are reported.

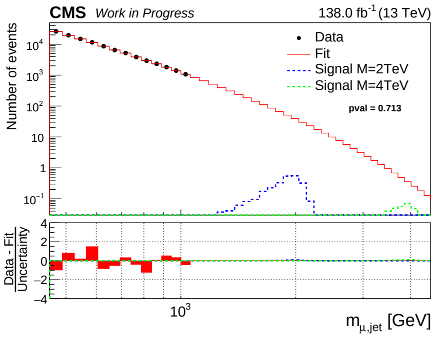


(a) f_{2par}^{UA2} vs f_{3par}^{UA2}

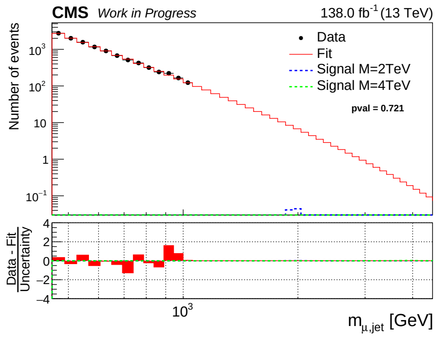


(b) f_{3par}^{UA2} vs f_{4par}^{UA2}

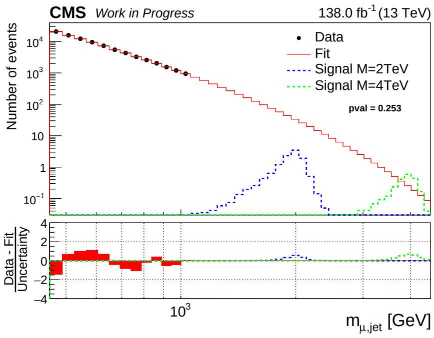
Figure C.1. UA2 function fit test for all the categories.



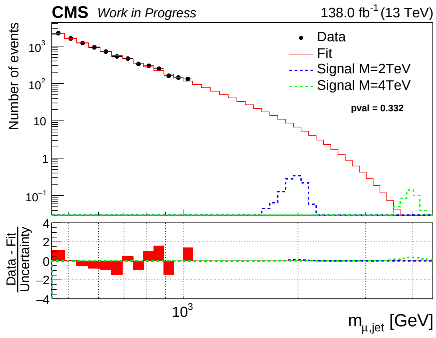
(a) 1-muon, loose, nobtag category



(b) 1-muon, loose, btag category

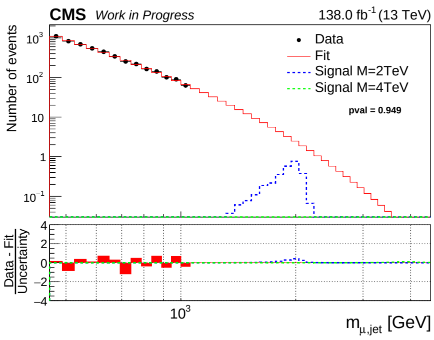


(c) 1-muon, tight, nobtag category

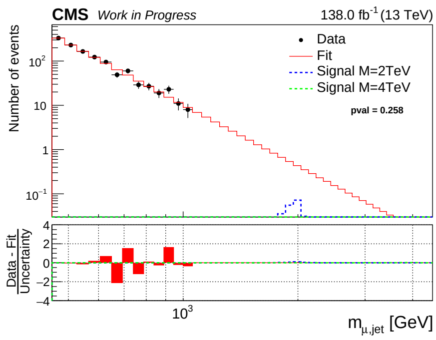


(d) 1-muon, tight, btag category

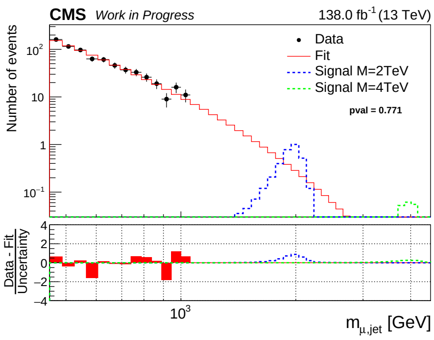
Figure C.2. Background fits for the standard dijet function for all the categories for the 1 muon signal region. The black points are shown up to a reconstructed mass of 1 TeV. The red line is the background estimation obtained using the standard dijet function. The blue (green) line represents a signal hypothesis with M_{LQ} = 2 (4) TeV and $\lambda_{u\mu}$ =1, which is normalized to the expected upper-limit cross-section. The horizontal axis shows the value of the $M_{\mu jet}$ spectra, while the vertical axis shows the number of events in each bin. The bottom part shows the difference between the data and the background fit function divided by the statistical uncertainty.



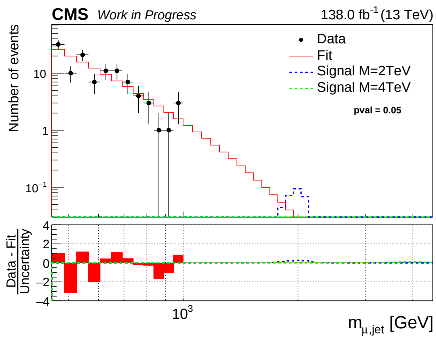
(a) 2-muon, loose, nobtag category



(b) 2-muon, loose, btag category



(c) 2-muon, tight, nobtag category



(d) 2-muon, tight, btag category

Figure C.3. Background fits for the standard dijet function for all the categories for the 2 muon signal region. The black points are shown up to a reconstructed mass of 1 TeV. The red line is the background estimation obtained using the standard dijet function. The blue (green) line represents a signal hypothesis with M_{LQ} = 2 (4) TeV and $\lambda_{u\mu}$ =1, which is normalized to the expected upper-limit cross-section. The horizontal axis shows the value of the $M_{\mu jet}$ spectra, while the vertical axis shows the number of events in each bin. The bottom part shows the difference between the data and the background fit function divided by the statistical uncertainty.

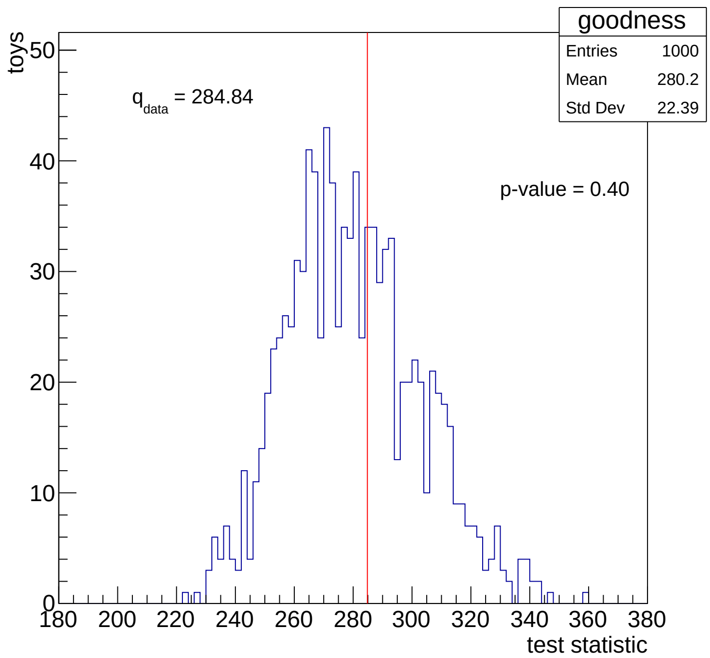


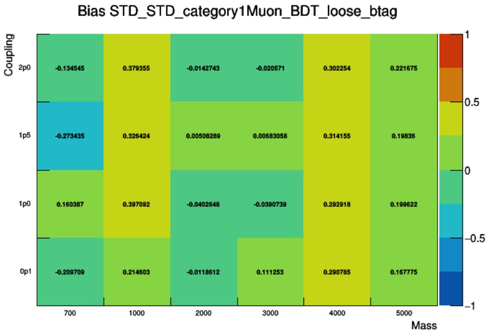
Figure C.4. Distributions of the test statistic (q) obtained from 1000 toy experiments for the background estimation obtained using the UA2 function. The red vertical lines mark the test statistic values evaluated from data (qdata).

Appendix D

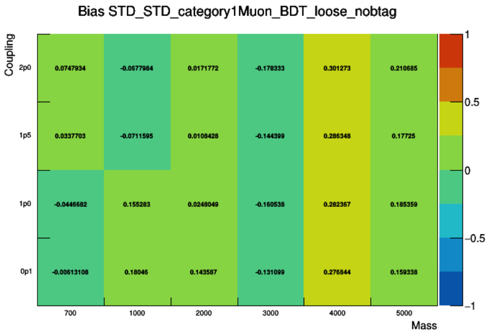
Bias study with no signal injection

In this section, the results of the bias study in the case of no signal injection for both LQ($u\mu$) and LQ($b\mu$) are reported. The results are shown for all the combinations of functions not shown in Section 6.4.2.

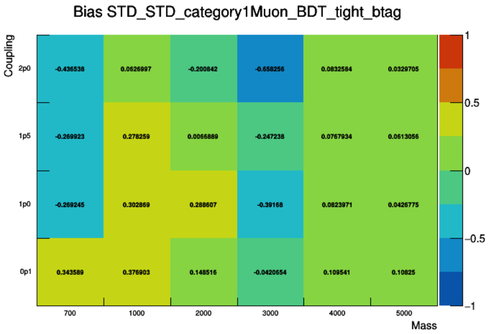
The region for $M_{LQ} \geq 4$ (2) TeV for the 1 (2) muon signal region shows only positive values for the bias. This is because σ_{fit} is forced to be positive in that region. At such high values of M_{LQ}, in fact, the $M_{\mu jet}$ distribution is poorly populated, and a negative fluctuation of σ_{fit} would cause the signal+background distribution to be negative, which is unrealistic. As a consequence, the bias in that region can only be positive. Moreover, in some cases, the PDF of the signal can become negative; thus, it is not possible to perform the fit. For these cases, the bias value in the graph is 99. This does not happen in Figures 6.16 and 6.17 because the injection of a signal populates the spectra at high $M_{\mu jet}$ values, and the total distribution is always positive. In this case, a negative value of the bias arises when $\sigma_{fit} < \sigma_{inj}$ on average.



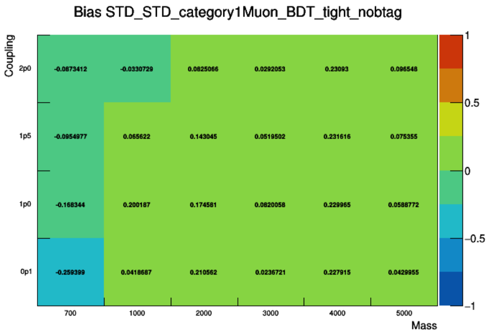
(a) 1-muon, loose, btag category



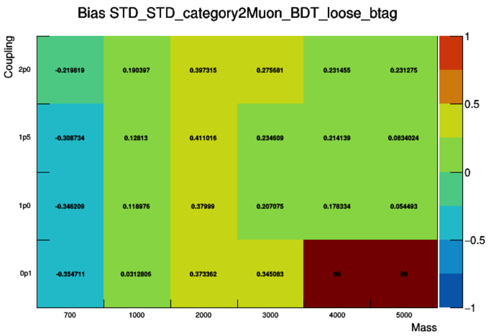
(b) 1-muon, loose, nobtag category



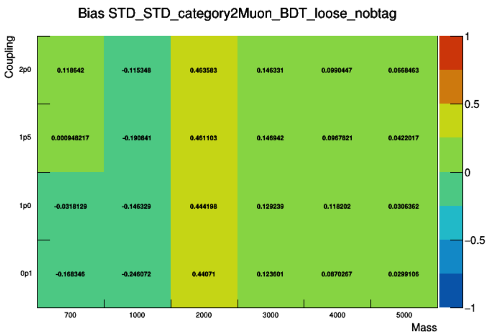
(c) 1-muon, tight, btag category



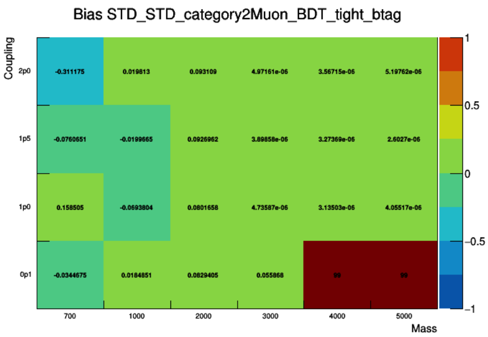
(d) 1-muon, tight, nobtag category



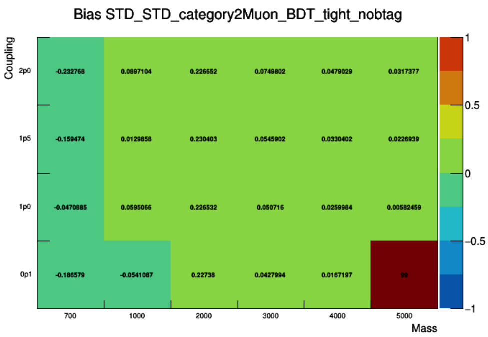
(e) 2-muon, loose, btag category



(f) 2-muon, loose, nobtag category

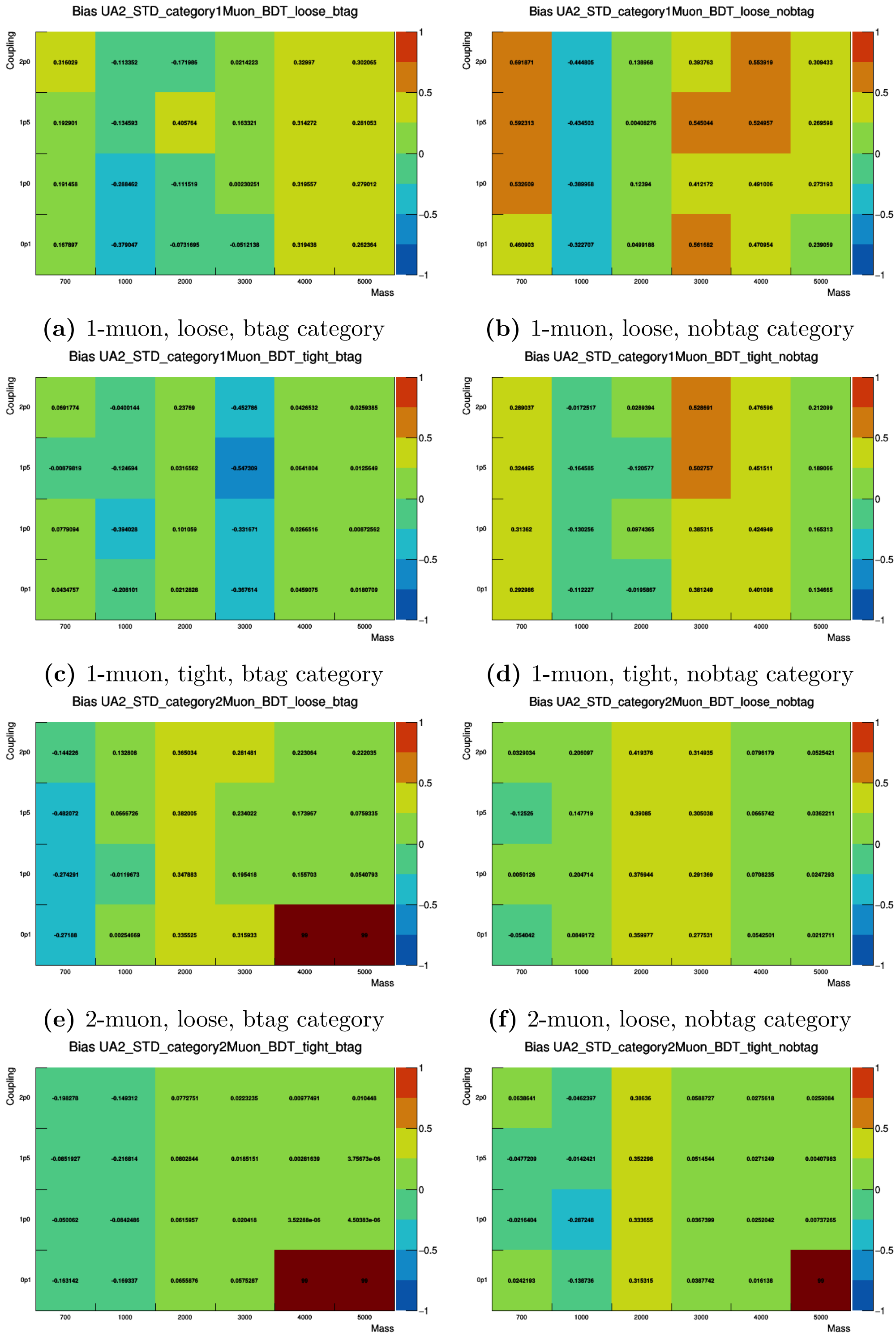


(g) 2-muon, tight, btag category



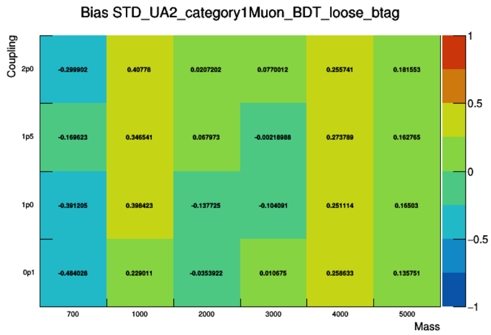
(h) 2-muon, tight, nobtag category

Figure D.1. Mean of pull distribution for the bias study versus $\lambda_{u\mu}$ and M_{LQ} from pseudodata both generated and fitted with function f_1. No signal is injected. In the z-axes of the plots, the bias is shown.

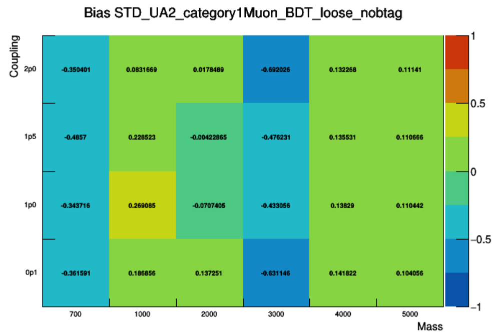


(a) 1-muon, loose, btag category
(b) 1-muon, loose, nobtag category
(c) 1-muon, tight, btag category
(d) 1-muon, tight, nobtag category
(e) 2-muon, loose, btag category
(f) 2-muon, loose, nobtag category
(g) 2-muon, tight, btag category
(h) 2-muon, tight, nobtag category

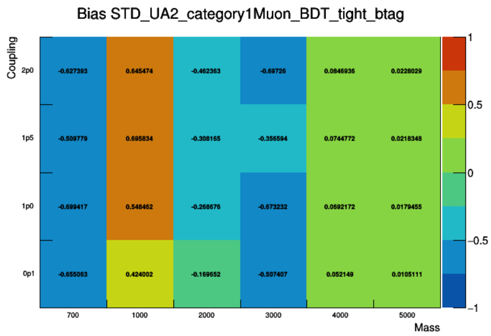
Figure D.2. Mean of pull distribution for the bias study versus $\lambda_{u\mu}$ and M_{LQ} from pseudodata both generated with f_2 and fitted with f_1. No signal is injected. In the z-axes of the plots, the bias is shown.



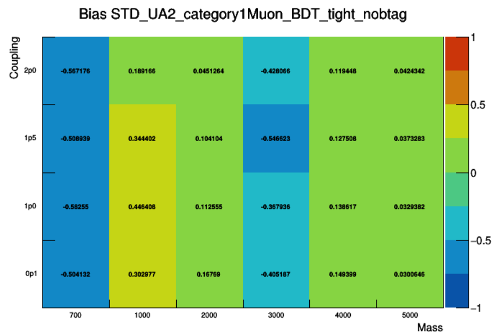
(a) 1-muon, loose, btag category



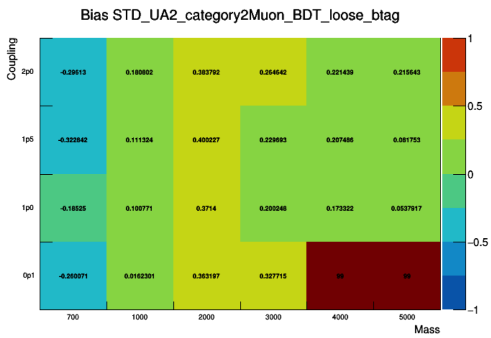
(b) 1-muon, loose, nobtag category



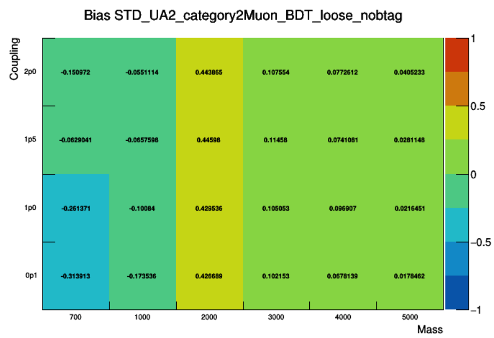
(c) 1-muon, tight, btag category



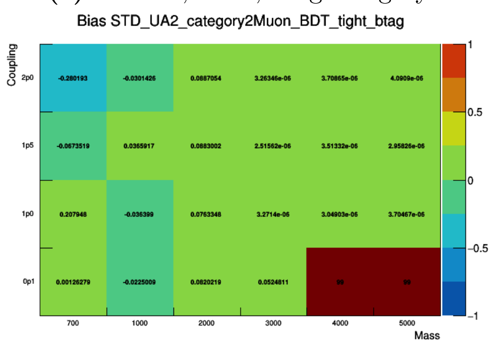
(d) 1-muon, tight, nobtag category



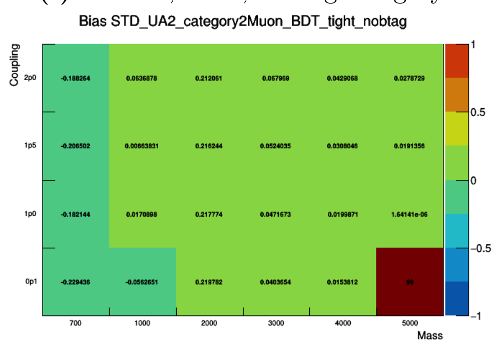
(e) 2-muon, loose, btag category



(f) 2-muon, loose, nobtag category

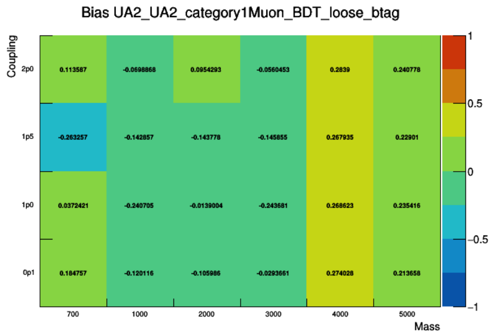


(g) 2-muon, tight, btag category

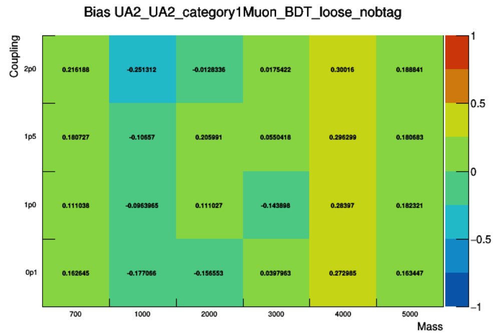


(h) 2-muon, tight, nobtag category

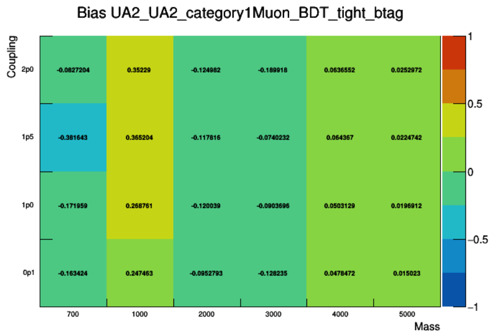
Figure D.3. Mean of pull distribution for the bias study versus $\lambda_{u\mu}$ and M_{LQ} from pseudodata generated with f_1 and fitted with f_2. No signal is injected. In the z-axes of the plots, the bias is shown.



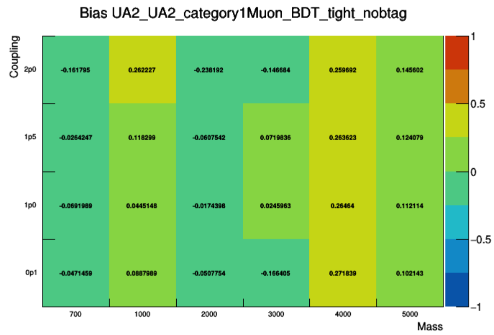
(a) 1-muon, loose, btag category



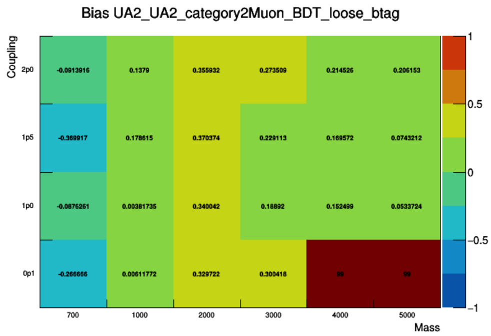
(b) 1-muon, loose, nobtag category



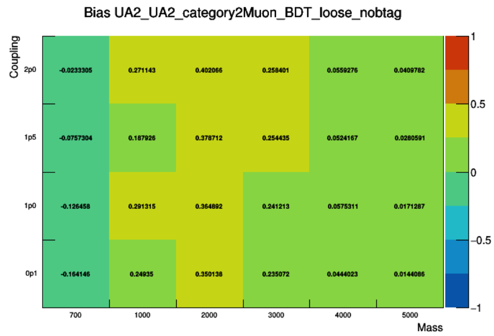
(c) 1-muon, tight, btag category



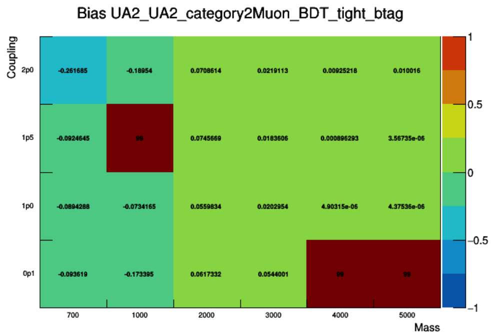
(d) 1-muon, tight, nobtag category



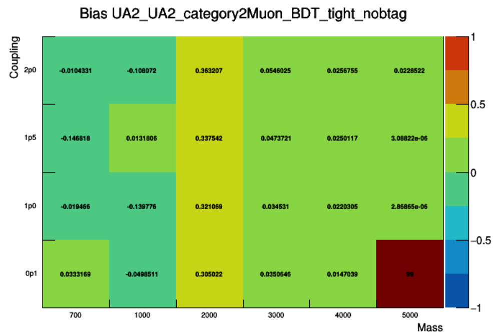
(e) 2-muon, loose, btag category



(f) 2-muon, loose, nobtag category

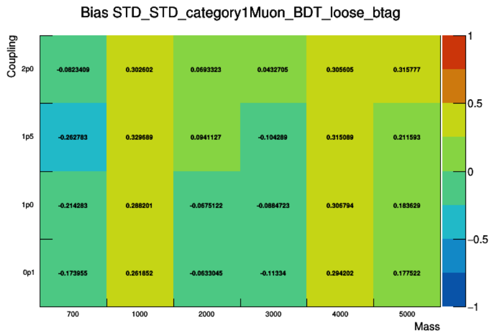


(g) 2-muon, tight, btag category

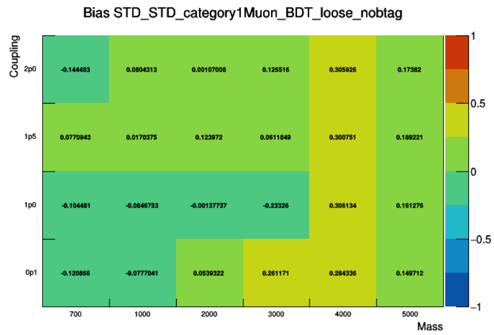


(h) 2-muon, tight, nobtag category

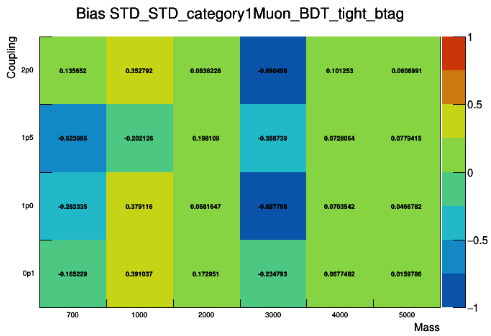
Figure D.4. Mean of pull distribution for the bias study versus $\lambda_{u\mu}$ and M_{LQ} both generated and fitted with f_2. No signal is injected. In the z-axes of the plots, the bias is shown.



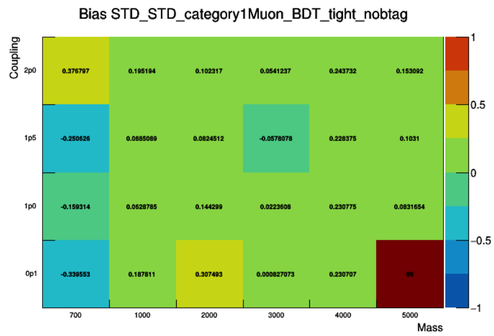
(a) 1-muon, loose, btag category



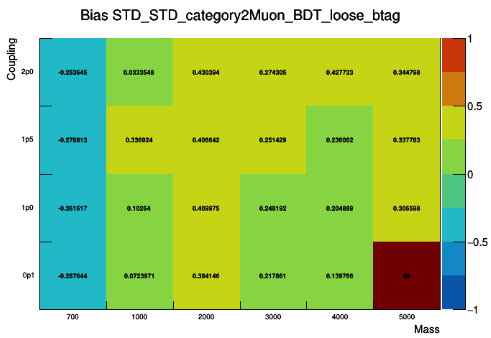
(b) 1-muon, loose, nobtag category



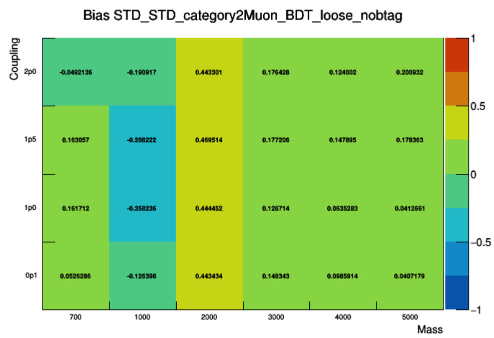
(c) 1-muon, tight, btag category



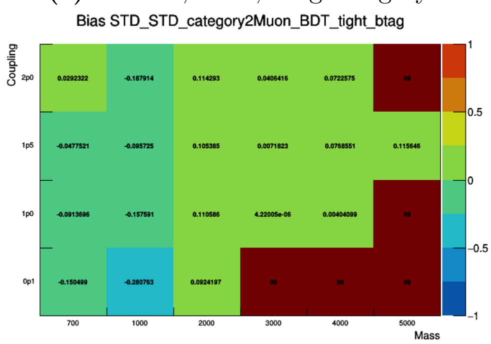
(d) 1-muon, tight, nobtag category



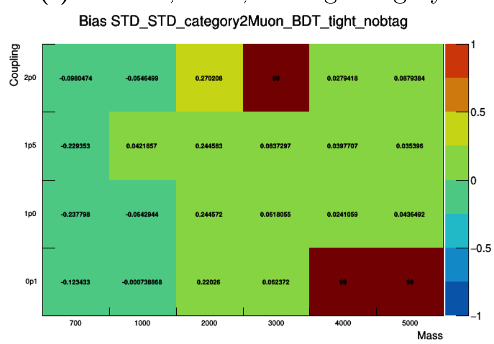
(e) 2-muon, loose, btag category



(f) 2-muon, loose, nobtag category

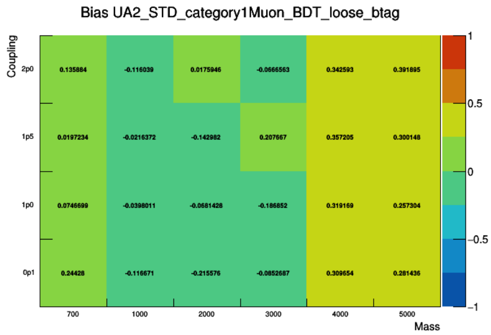


(g) 2-muon, tight, btag category

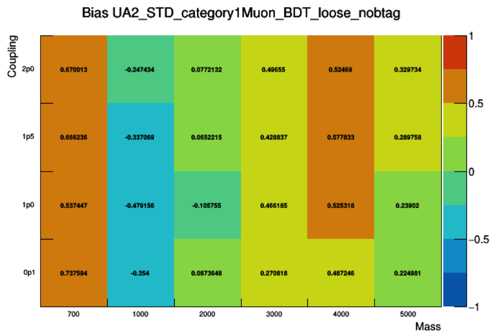


(h) 2-muon, tight, nobtag category

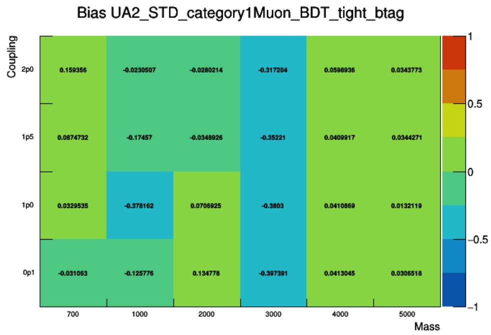
Figure D.5. Mean of pull distribution for the bias study versus $\lambda_{b\mu}$ and M_{LQ} from pseudodata both generated and fitted with function f_1. No signal is injected. In the z-axes of the plots, the bias is shown.



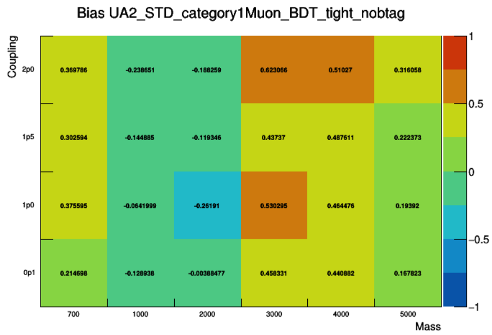
(a) 1-muon, loose, btag category



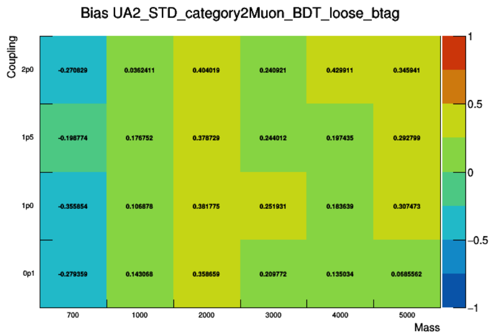
(b) 1-muon, loose, nobtag category



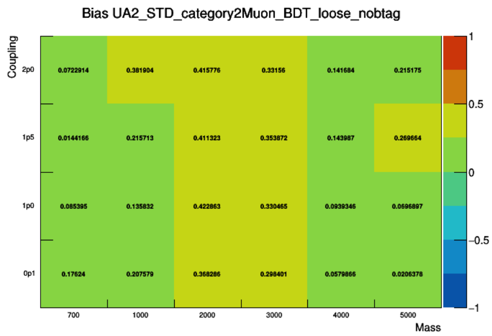
(c) 1-muon, tight, btag category



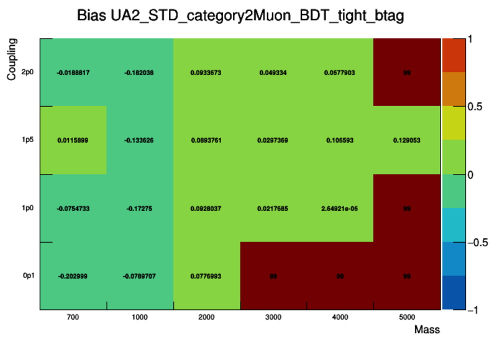
(d) 1-muon, tight, nobtag category



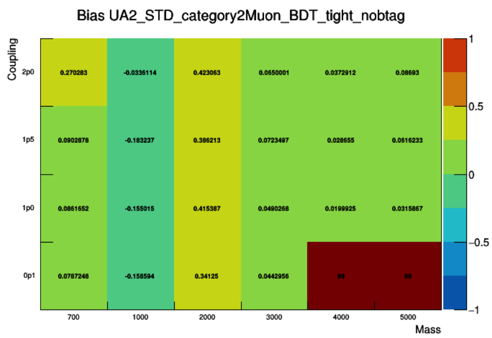
(e) 2-muon, loose, btag category



(f) 2-muon, loose, nobtag category

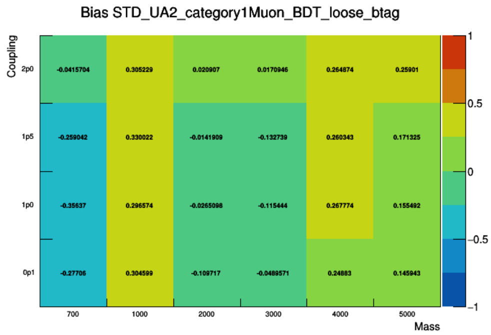


(g) 2-muon, tight, btag category

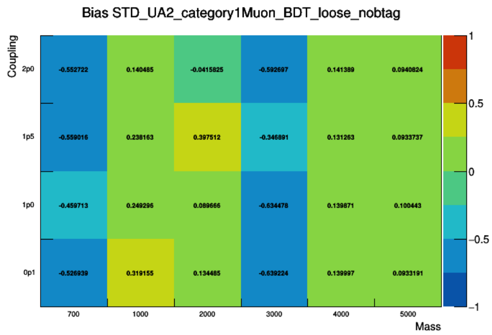


(h) 2-muon, tight, nobtag category

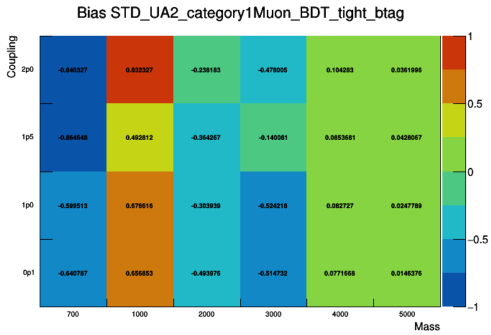
Figure D.6. Mean of pull distribution for the bias study versus $\lambda_{b\mu}$ and M_{LQ} from pseudodata both generated with f_2 and fitted with f_1. No signal is injected. In the z-axes of the plots, the bias is shown.



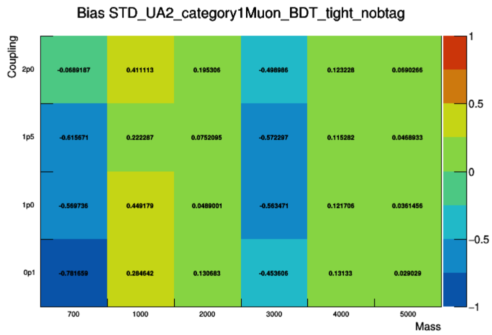
(a) 1-muon, loose, btag category



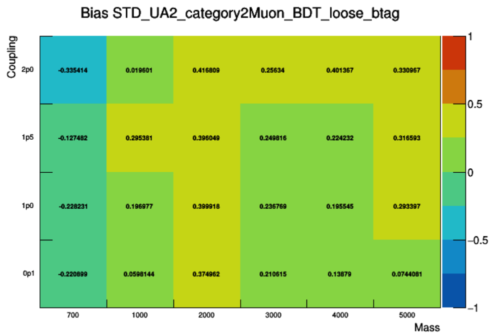
(b) 1-muon, loose, nobtag category



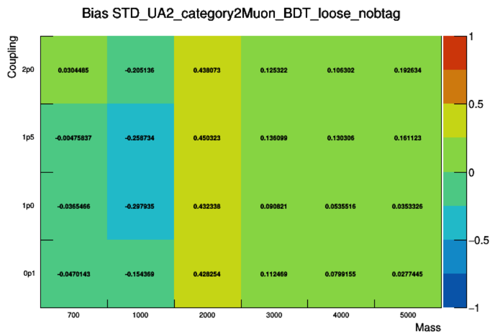
(c) 1-muon, tight, btag category



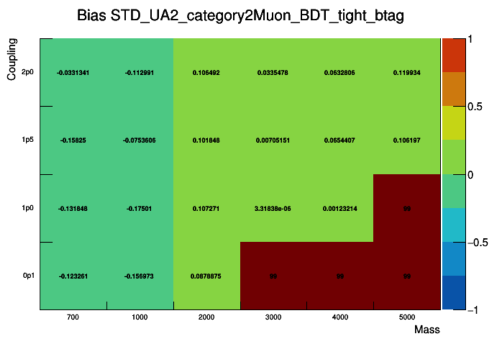
(d) 1-muon, tight, nobtag category



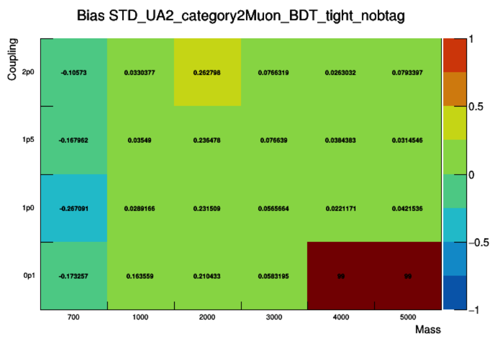
(e) 2-muon, loose, btag category



(f) 2-muon, loose, nobtag category

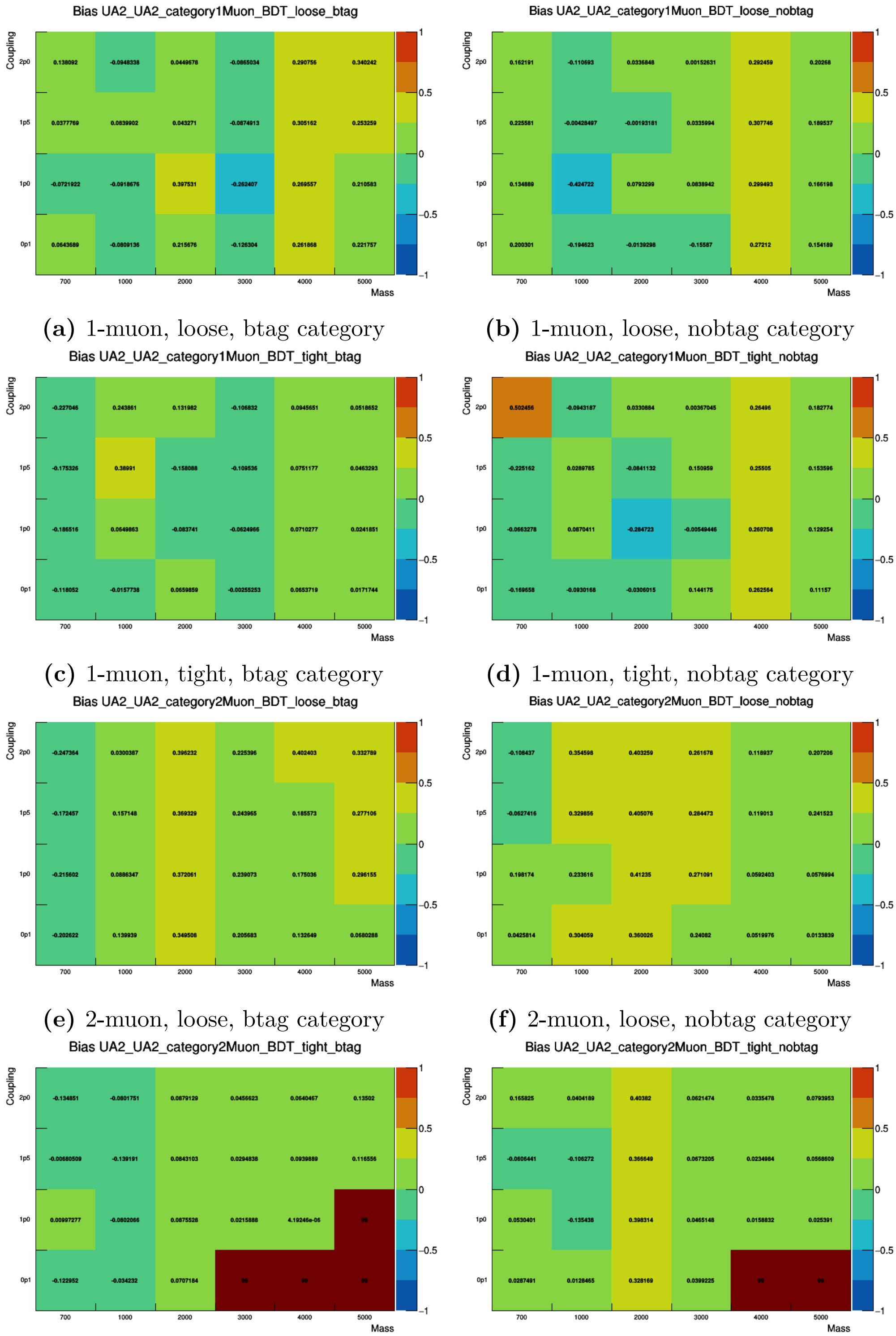


(g) 2-muon, tight, btag category



(h) 2-muon, tight, nobtag category

Figure D.7. Mean of pull distribution for the bias study versus $\lambda_{b\mu}$ and M_{LQ} from pseudodata generated with f_1 and fitted with f_2. No signal is injected. In the z-axes of the plots, the bias is shown.



(a) 1-muon, loose, btag category

(b) 1-muon, loose, nobtag category

(c) 1-muon, tight, btag category

(d) 1-muon, tight, nobtag category

(e) 2-muon, loose, btag category

(f) 2-muon, loose, nobtag category

(g) 2-muon, tight, btag category

(h) 2-muon, tight, nobtag category

Figure D.8. Mean of pull distribution for the bias study versus $\lambda_{b\mu}$ and M_{LQ} both generated and fitted with f_2. No signal is injected. In the z-axes of the plots, the bias is shown.

Appendix E

ECAL and double weights

ECAL Trigger Primitives (TPs) are integral to the Level 1 Calorimeter Trigger inputs. The ECAL TPs during the LHC Phase-I electronics deployment are based on *strips*, each consisting of 5 crystals in the EB and 1-5 crystals in the EE. At a frequency of every 25 ns, a digitized sample of the ECAL signal from each crystal is recorded, as depicted in Figure E.1. This process results in a consistent output of ADC counts during data acquisition. These readings are linearized to normalize gains across amplifiers and combined for each strip. Subsequently, 10 linearized samples from a strip are processed through a digital filter composed of 10 FIR (Finite impulse response) weights. Each weight multiplies its corresponding sample value, and the products are summed to derive the transverse energy (E_T), as presented in Equation E.1. Here, S_i signifies the digitized sample "i," with "i" ranging from 0 to 10, corresponding to the 10 digitized samples of an ECAL pulse within 0-225 ns. The w_i symbol represents the FIR weight allocated to the i-th sample in the ECAL

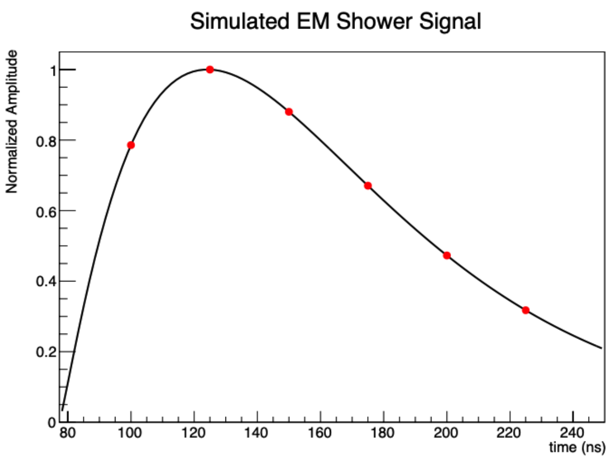


Figure E.1. ECAL analog pulse shape example, with digitized samples taken every 25 ns [169].

strip, predetermined. Hence, their sum equals zero, which is essential for dynamic pedestal subtraction:

$$E_T = \sum_{i=1}^{10} S_i \times w_i, \quad \sum_{i=1}^{10} w_i = 0. \tag{E.1}$$

This process is reiterated for each successive 10-sample set, with the sample window shifting by 25 ns each time, producing a steady stream of E_T values. Furthermore, a peak-finding algorithm is applied to identify which BX, within a certain range, has the maximal E_T. The energies from the strips are summed to form trigger towers—5 strips in the EB and 1-5 in the EE—to create TPs. An ECAL TP represents the E_T of a trigger tower for a given BX and includes up to two distinguishing feature bits. One bit is used to discern electromagnetic signals from jets in the EB, and another is allocated to negate anomalous signals known as "spikes". If a TP is established for a BX, subsequent BXs in the same window are precluded from TP generation. An ECAL TP represents the E_T of a trigger tower for a specific BX, with provisions for up to two feature bits. A fine-grain bit is employed to distinguish electromagnetic (EM) signals from those of jets. In the EB, an additional bit is designated for the rejection of anomalous signals, referred to as "spikes." When a TP is instantiated for a BX, subsequent BXs in the same interval are ineligible for TP generation. TPs that are non-zero are relayed to the Trigger Concentrator Card (TCC) for further processing and temporal alignment before being dispatched to the L1 trigger. This process and the computation of strip E_T are illustrated in Figure E.2.

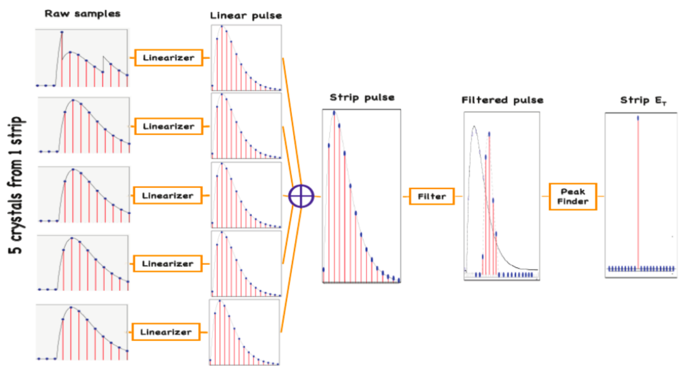


Figure E.2. ECAL strip E_T formation [169].

Each ECAL TP in the EB and EE is composed of an E_T value, which is the sum of its strip E_T values, supplemented with information bits and a BX assignment. ECAL TPs are generated on the detector and transmitted to the Level-1 trigger in sync with the LHC collision rate of 40 MHz. Given that the transverse momentum of colliding LHC proton bunches is zero, the detection of hits with high E_T or p_T is indicative of a potentially significant hard interaction between protons, making it a critical parameter for L1 trigger decisions.

E.1 Double Weights

During LHC Runs 1 and 2, the on-detector ECAL FENIX chip, a custom ASIC, was employed for reconstructing energy to compute E_T sums for ECAL TPs, applying a singular set of weights to the digitized signals. Upon the second Long Shutdown (LS2), it was discovered that the ECAL FENIX chip has the facility to store and utilize dual sets of weights. This enhancement essentially duplicates the ECAL FENIX data stream, bifurcating it into two paths that are electronically equivalent, each corresponding to a FIR filter, termed the "EVEN" and "ODD" filters, as depicted in Figure E.3. This feature was implemented in the ECAL FENIX chip for potential further use but was never used during Runs 1 and 2.

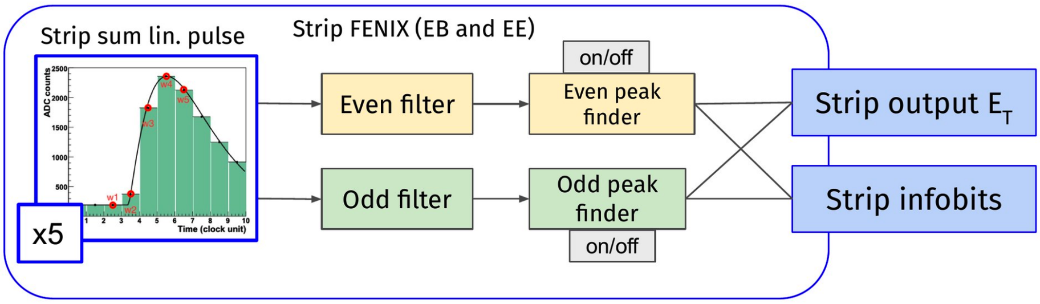


Figure E.3. ECAL double weights mechanism [169].

E.1.1 Spikes

At the ECAL, an oft-observed phenomenon is a direct ionization within the EB APDs, which gives rise to anomalous signals termed "spikes". Figure E.4 depicts the possible spike formation processes. Unlike electromagnetic showers that originate from the interactions of LHC collisions, these spikes are not interesting for the physics CMS program. They must be efficiently excluded to regulate trigger rates

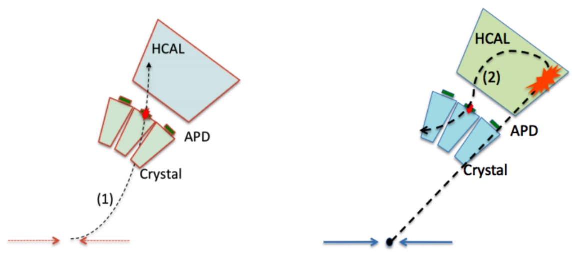


Figure E.4. Origin of the particles producing spikes in the ECAL EB APD sensors [2].

and safeguard the accuracy of offline reconstructions of electrons, photons, and jets. Moreover, the precursors to these spikes may traverse the CMS detector for some duration before directly ionizing the EB APDs, thus potentially rendering these signals out-of-sync with electromagnetic signals. A strategy employed at L1 to remove spikes is known as the "spike killer" [158]. It functions by implementing a topological cut, leveraging the observation that spikes usually deposit all their energy into a solitary ECAL crystal as a consequence of the direct ionization of APDs. In contrast, EM showers are anticipated to disperse energy over several crystals. The operational principle of the spike killer is illustrated in Figure E.5.

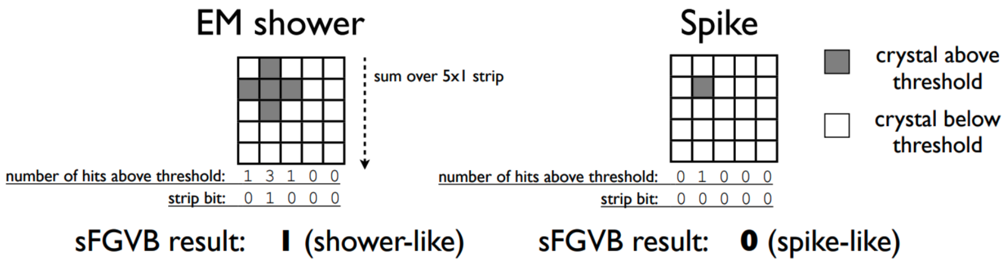


Figure E.5. Operation of the strip Fine-Grained Veto Bit (sFGVB) on an electromagnetic shower (left) and a spike-like energy deposit (right) [158].

The spike killer utilizes a per-strip bit, the Fine-Grained Veto Bit (sFGVB), which is activated to 1 if at least two crystals in a strip are above a predetermined energy threshold. If a trigger tower, comprising 25 crystals organized into 5 strips, has at least one strip with a sFGVB set to 1, it is conserved as it resembles an EM shower, characterized by its dispersed energy. However, if a trigger tower lacks any strip with a sFGVB of 1, the TP energy is deemed zero if its energy does not exceed the spike killer's killing threshold of 16 GeV.
To refine the spike killer for Run 3, where increased noise and PU were expected, the energy threshold per crystal was heightened. This modification is a proactive measure against the expected rise in noise and PU for all crystals. The prevalence of spike contamination in Trigger Primitives (TPs), as observed in Run 2 and the proposed operating point for Run 3, is depicted in Figure E.6. Notably, the contamination

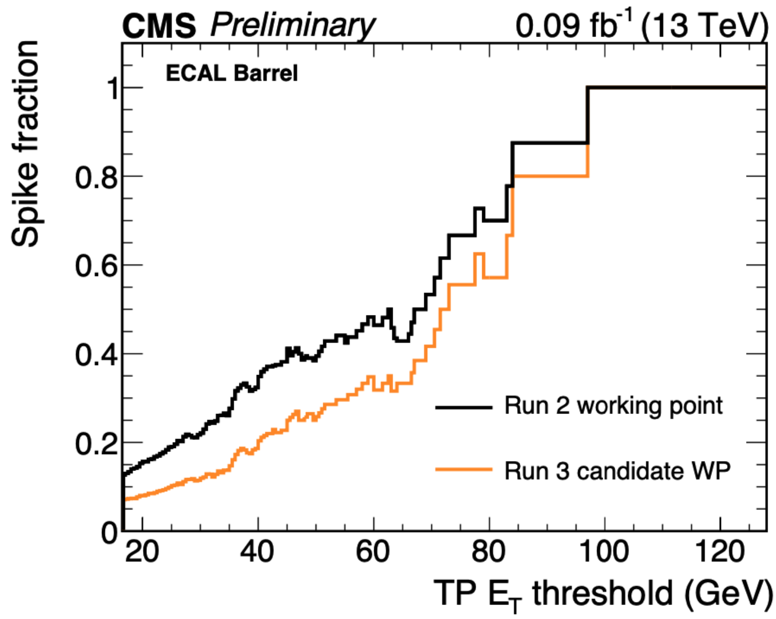


Figure E.6. Spike fraction vs. TP E_T threshold with a Run 2 and Run 3 candidate working point of the existing ECAL L1 spike killer. The data comes from a ZeroBias dataset recorded in July 2018 with a peak pileup of 50 [90].

graph, which utilizes data from a ZeroBias dataset, indicates a significant incidence of spikes at elevated energies. This is because spikes are more apt to result in a high-energy signal due to the direct ionization of APDs, in contrast to high-energy EM showers which are the result of actual high-energy particles from proton-proton collisions.

This demonstrates that while updating the settings of the existing spike killer to a prospective Run 3 working point does mitigate additional spikes at Level-1, ample opportunity for refinement remains, especially in the high-energy domain. Furthermore, at L1, there is presently no spike killer mechanism for the low-energy range of 0-16 GeV, given that this is beneath the spike killer's threshold.

E.1.2 Optimization of double weights

The initial strategy for refining ECAL's dual weight application was to retain the original amplitude weights within the EVEN filter and implement a secondary set of weights in the ODD filter dedicated to timing for the computation of on-detector timing metrics for trigger primitives. Studies revealed the timing weights' capability to discern between concurrent and spike-induced out-of-time signals. Since a timing cut was not possible in the ECAL FENIX chip, the idea is to calculate dual amplitudes utilizing two distinct sets of weights and employ a comparator in the electronics to assert a boolean flag if one amplitude surpasses the other. When an ODD set of weights is refined for the detection of out-of-time signals, it is anticipated to produce a larger amplitude for such signals than the Run 2 weights, which are designed for in-time signals. Hence, the methodology adopted is to optimize an ODD set of weights for the identification of out-of-time signals.

Selecting an odd set of weights for tagging out-of-time signals entails a multifaceted, multivariate problem, necessitating an assessment of the realistic signal energy spectrum, the spike energy spectrum, the spike timing PDF, and the influence of pileup on signal waveform alteration. A numerical optimization was established to deduce the ideal weight sets to supplant the second FIR filter weights to extract ODD weight sets optimized for maximizing signal efficiency alongside spike rejection. This optimization harnessed simulated signal waveforms characterized by an analytic representation and the simulated pileup, in addition to the simulation of spike waveforms from an independent simulation. The optimization was structured as a problem of minimizing loss, utilizing gradient descent and reverse loss propagation to augment the rejection of spikes while curtailing the rejection of signals. This methodology was integrated into a loss definition, as illustrated in Figure E.7.

One of the pivotal parameters in the optimization process is the minimum separation between amplitude values calculated by the EVEN (default) and ODD (out-of-time sensitive) weight sets, symbolized as δ_{min}. Altering this parameter yields various operational points. Figure 5.15 illustrates the segmentation of a simulated spike timing PDF, identified as out-of-time at discrete operational points, referenced in [91, 55].

In the spike timing PDF, the majority of spikes register nearly concurrently with EM signals, yet some display a delayed tail indicative of out-of-time activity. This is attributed to the fact that spike progenitors traverse the CMS detector before

$$
\begin{aligned}
L = {} & (\lambda_{Signal} \times L_{SigEff}) + \\
& (\lambda_{Spike} \times L_{SpikeRej}) + \\
& (\lambda_{Norm} \times W2LossNorm) + \\
& W2LossLimit
\end{aligned}
$$

(a) Loss function

$$
L_{SigEff} = \begin{cases} if((A_{w2,d1} - A_{w1,d1}) \geq \delta_{\min}): & (A_{w2,d1} - A_{w1,d1}) \\ if((A_{w2,d1} - A_{w1,d1}) < \delta_{\min}): & 0 \end{cases}
$$

$$
L_{SpikeRej} = \begin{cases} if((A_{w1,d2} - A_{w2,d2}) \geq \delta_{\min}): & (A_{w1,d2} - A_{w2,d2}) \\ if((A_{w1,d2} - A_{w2,d2}) < \delta_{\min}): & 0 \end{cases}
$$

(b) Use of δ_{min} in loss

$$
W2LossLimit = \begin{cases} if\left(\sum_{i=1}^{5} |w_{2,i}| < 1\right): & 0 \\ if\left(\sum_{i=1}^{5} |w_{2,i}| \geq 1\right): & \sum_{i=1}^{5} |w_{2,i}| \cdot (-100) \end{cases}
$$

(c) Definition of ODD weights loss limit.

Figure E.7. Loss definition used in the optimization of ODD set of amplitude weights [169].

ionizing the EB APDs. It is observed that elevating the $\delta_{\min}$ value aids in tagging these delayed spikes. Intuitively, an increased $\delta_{\min}$ value is employed to select spike examples that exhibit significant differences in EVEN and ODD amplitudes during optimization, which are typically associated with temporally shifted spikes. It is also imperative to quantify the anticipated improvement in spike rejection for various $\delta_{\min}$ operational points and to assess their influence on EM shower-like signals. In order to identify a reasonable trade-off between signal efficiency and spike rejection, only signals with ET leq 3 GeV are considered as simulated signals with ET $>$ 3 GeV have an efficiency near 100%. Additionally, only spikes with a timing greater than 10 ns are considered, as the working points chosen are not effective at tagging in-time spikes. Table E.1 shows the signal selection efficiency and the spike rejection

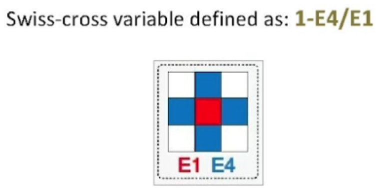
$\delta_{\min}$ (GeV)	Signal efficiency (%)	Spike rejection (%)
0.5	78.2	77.6
2.5	95.6	62.5
5.0	95.7	19.2

Table E.1. Signal efficiency and spike rejection for different values of $\delta_{\min}$ [55].

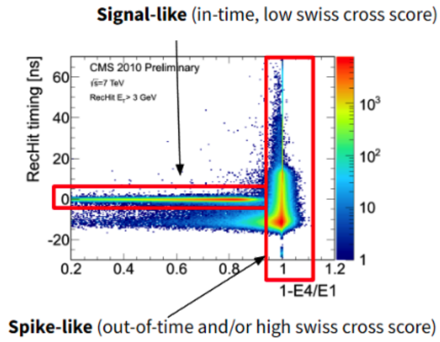
as a function of the WP chosen. It is shown that moving from the $\delta_{\min}$ of 2.5 GeV to a 5.0 GeV working point results in a minimal gain in signal efficiency (0.1%), while a significant portion of spike rejection is lost (43.3%). This implies that the $\delta_{\min} = 2.5$ GeV working point offers a favorable compromise between signal efficiency at low E_T and overall spike rejection.

E.1.3 Commissioning for Run3

An offline Severity assignment is employed to categorize ECAL TPs in the dataset as akin to signal or spike. Each ECAL TT is comprised of 25 crystals, and an offline energy assessment is performed for each crystal within high-energy event regions at L1A, denoted as a reconstructed hit or rec hit. In parallel, each rec hit is assigned a timing value and a severity level. A severity level of zero signifies a rec hit without issues in the data, whereas a level of three indicates a rec hit is out of time based on its timing. A severity level of four implies the rec hit satisfies one or more criteria: it is flagged as out-of-time or does not pass a topological cut, the swiss cross cut. Since spikes originate from isolated APD hits and not from dispersed EM showers, they usually concentrate their energy in a single crystal, identifiable by a swiss cross



(a) Swiss cross definition, illustrated by the energy in a 3x3 ECAL crystal portion. E1 = energy of the central crystal, E4 = sum of the energies of the central crystal's four surrounding neighbors [158].



(b) Reconstructed time vs. swiss cross score [63].

Figure E.8. Swiss cross definition, and reconstructed hit timing vs. swiss cross score from a 2010 CMS data sample.

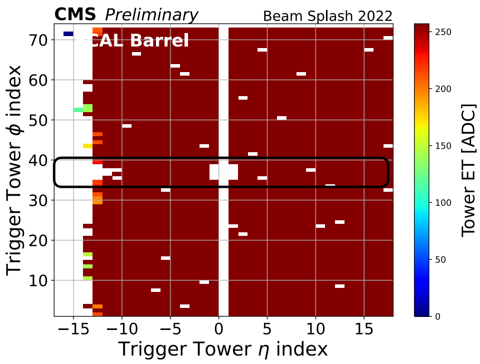
variable, depicted in Figure E.8. Consequently, rec hits with a severity level of zero (or four) correspond typically to signal-like (or spike-like) events, mapped in the respective region of reconstructed time versus swiss-cross score in Figure E.8. To allocate a severity level and reconstructed time to an EB TP, the highest energy reconstructed hit within a TP is identified. This hit then imparts its severity level and timing to the TP.

During the commissioning of CMS and LHC for Run 3, the accelerator complex orchestrated beam splashes to the experiments. This entailed the closure of an LHC collimator upstream from CMS, leading to a proton bunch interaction and the ensuing production of a particle shower. These particles, principally muons, pervade the entire CMS detector.

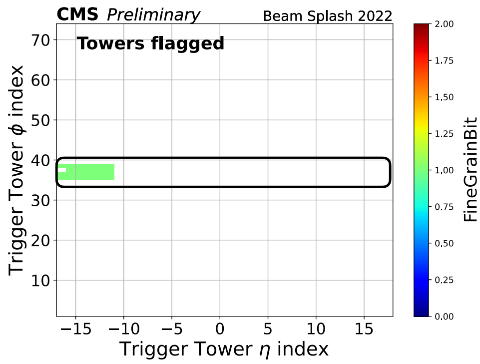
An extensive spectrum of ECAL reconstructed hit timings is registered during beam splash occurrences. This arises from the particle shower reaching the CMS detector unidirectionally. The CMS system is calibrated to trigger events based on ECAL activity in the vicinity of $\eta = 0$, which synchronizes in-time hits with the anticipated timing of proton-proton bunch collisions. As the particle cascade interacts with the ECAL over time, these hits are chronologically categorized as positively out-of-time. In contrast, hits resulting from the particle shower impacting the ECAL prior to the triggering of the event are denoted as negatively out-of-time in relation to $\eta = 0$.

In this configuration, the majority of the ECAL barrel operated with its standard Run 2 setup. However, a "killing + tagging" mode was established for two central supermodules to address the expected negative out-of-time signals. In this mode, ECAL strips that exhibit a larger odd amplitude compared to the even amplitude have their energy set to zero, a process referred to as "killing." Concurrently, if a considerable extent of energy zeroing is detected within an ECAL Trigger Tower, a flag is raised, denoting the "tagging" process. This method's functionality was scrutinized and is considered a potentially valuable technique for future applications. For example, it could provide insights into which ECAL regions have undergone energy reduction due to the double weights mechanism. Observations indicated that energy readings in these regions were lower than in other supermodules with similar TP timings due to the particle splash propagating from negative to positive pseudorapidity (η) values. TPs in these regions were appropriately tagged, marking the first instance of ECAL running with double weights in the "killing + tagging" mode. This demonstrates the efficacy of this previously untested configuration. The resulting ECAL TP timings, along with the flagged TPs, are depicted in Figure E.9.

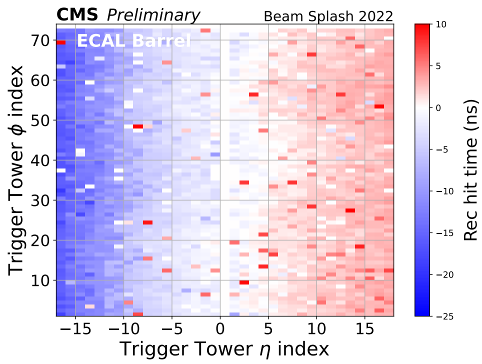
Observations indicate that the TP timings, derived by assigning the timing of



(a) TP energy distribution.



(b) TPs tagged as out of time.



(c) TP timing distribution.

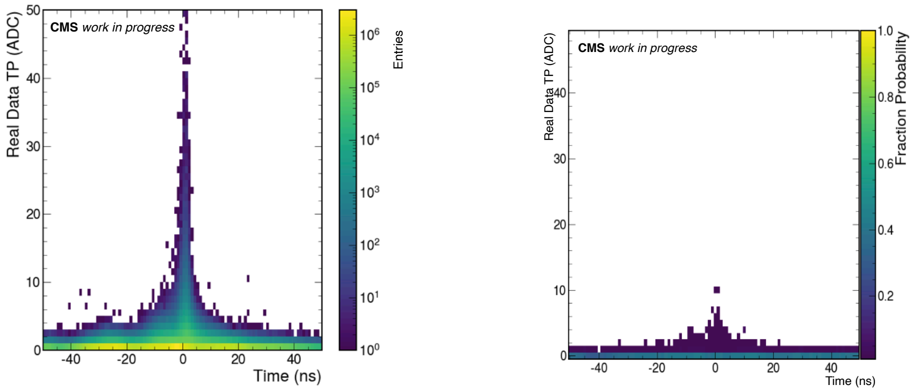
Figure E.9. ECAL TP energies, TPs tagged as out-of-time by double weights in killing + tagging mode and TP timing distribution during a 2022 CMS beam splash.

the highest energy reconstructed hit in the TT, vary from about −25 ns to 10 ns. TPs identified by the double weights mechanism as out-of-time possess timings in the negative domain, ranging from −10 ns to −15 ns, and in-time signals are not erroneously tagged. This constitutes the initial instance of out-of-time tagging at the ECAL TP and validates the functionality of the double weights system for tagging out-of-time signals in the data.

In addition, during data-taking in 2022, during a 2-hour period, for the first time, ECAL took data with pp collisions at 13.6 TeV in full readout mode running with double weights in tagging+killing mode $\delta_{min} = 2.5$ GeV working point. The reason a full-readout run was included in this study is that, in a full-readout, information from all ECAL crystals is saved. In non-full-readout runs, there is a selective readout procedure in which low-interest regions are not readout. It is desirable to study the full readout runs when comparing low-energy TP energies to ensure that all ECAL information is available in the events and be able to study the effects of double weights properly.

Two different categories are studied: signals (TPs assigned to severity 0 reconstructed hits), which are in time (matched reconstructed crystal hit time $|t| < 3$ns), and spikes (TPs assigned to severity 4 reconstructed hits), which are out-of-time (matched reconstructed crystal hit time $t > 10$ns). The data TP energies vs. offline matched times and the fraction of TPs tagged over the total are shown in Figure E.10 for severity zero matched signal-like TPs, and Figure E.11 for severity 4 matched spike-like TPs. We checked the timings of TPs that have some energy subtracted by double weights in the full timing range, not just restricted to in time and very later. From these distributions, it is observed that Double weights are able to tag TPs, which are positively out-of-time (spikes-like signals). For signals-like events, there is a negligible amount of tagging seen for in-time signals. This shows that with ECAL double weights, there is some potential to lower the spike contamination rate, including in the high energy regime.

From these distributions, it is observed that Double weights are able to tag TPs,



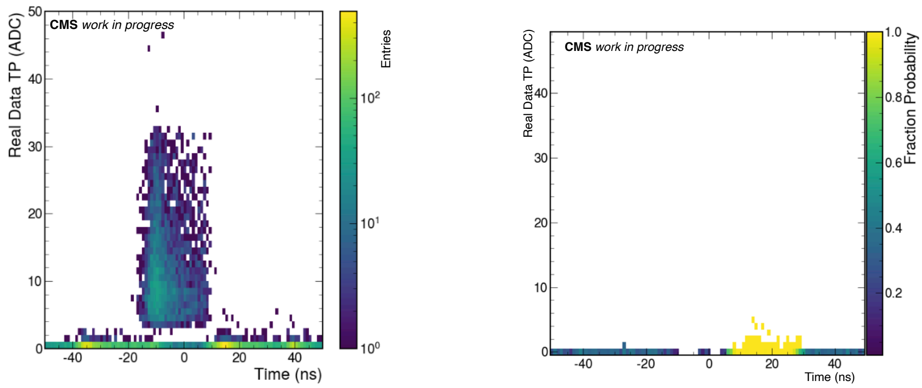
(a) All severity 0 signal TPs: Data energy vs. time

(b) Signal TPs tagged by double weights.

Figure E.10. 2022 LHC full-readout severity 0 signal TPs, all and tagged.

which are positively out-of-time, as well as some spikes, which are negative out-of-time. For signals, there is a non-negligible amount of tagging seen for in-time signals. To properly quantify the effects of the double weights for both signals and spikes, the resulting tagging probability (average energy fractions subtracted) from in-time signals and out-of-time spikes are shown in Figures E.12 left (right) and E.13 left (right). The outcome of this study is that low energy signals have a non-zero probability of having energy subtracted, which decreases as the TP energy increases, and that out-of-time spikes have a large percentage of energy subtracted, which increases as spike energy increases.

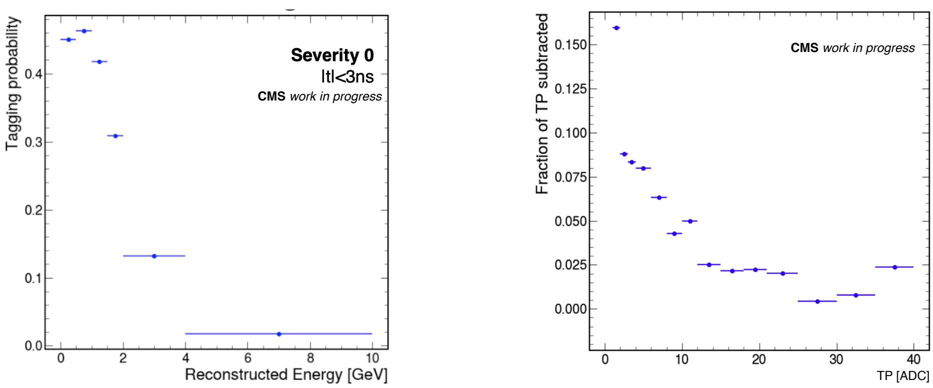
The results of these studies with double weights demonstrate a potential benefit at the ECAL TP level, as high-energy spikes are identified and eliminated with minimal impact on in-time low-energy signals. A subsequent critical evaluation involves the impact of double weights on CMS L1 metrics. This encompasses the effects on L1



(a) All severity 4 signal TPs: Data energy vs. time.

(b) Signal TPs tagged by double weights.

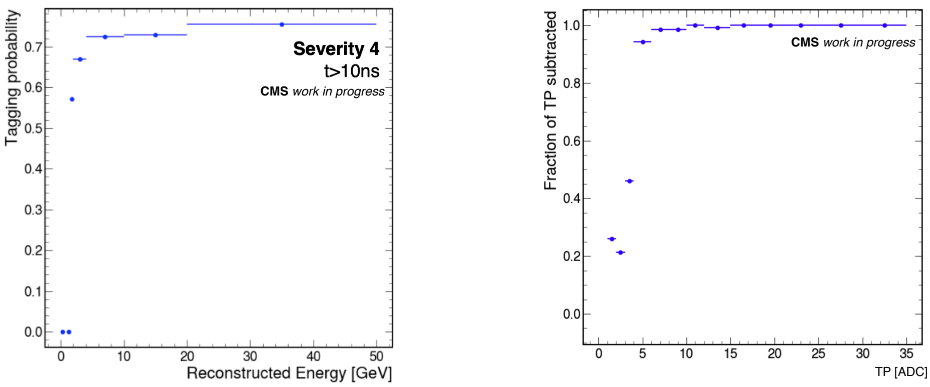
Figure E.11. 2022 LHC full-readout severity 4 signal TPs, all and tagged.



(a) Tagging probability for in-time signals as a function a function of TP transverse energy.

(b) Average energy fractions subtracted from in time signals.

Figure E.12. Tagging probability in data and average energy subtracted for Severity 0 matched TPs with offline matched reconstructed times < 3 ns.



(a) Tagging probability for out-of-time signals as a function a function of TP transverse energy.

(b) Average energy fractions subtracted from out-of-time signals.

Figure E.13. Tagging probability in data and average energy subtracted for Severity 4 matched TPs with offline matched reconstructed times $>$ 10 ns.

rates and the L1 turn-on curves. A conceivable advantage would be a reduction in the L1 rates owing to the exclusion of spikes, which might enable a reduction of the L1 seed energy thresholds. Once these studies are completed in the next months, the commissioning of the double weights will start for the last year of data-taking.

Bibliography

[1] LS3 schedule change. Available from: `https://hilumilhc.web.cern.ch/article/ls3-schedule-change`.

[2] Origin of spikes. Available from: `https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/RetrieveFile?docid=2941&filename=spike-1.png&version=1`.

[3] Proposed experiments and equipment for a programme of muon physics at the SPS. Tech. rep., CERN, Geneva (1974). Available from: `https://cds.cern.ch/record/594962`.

[4] Procedure for the LHC Higgs boson search combination in Summer 2011. Tech. rep., CERN, Geneva (2011). Available from: `https://cds.cern.ch/record/1379837`.

[5] Abbott, L. and Farhi, E. Are the weak interactions strong? *Physics Letters B*, **101** (1981), 69. Available from: `https://www.sciencedirect.com/science/article/pii/0370269381904925`, `doi:https://doi.org/10.1016/0370-2693(81)90492-5`.

[6] Aberle, O. e. a. *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*. CERN Yellow Reports: Monographs. CERN, Geneva (2020). Available from: `https://cds.cern.ch/record/2749422`, `doi:10.23731/CYRM-2020-0010`.

[7] Abt, I., et al. The h1 detector at hera. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **386** (1997), 310. Available from: `https://www.sciencedirect.com/science/article/pii/S0168900296008935`, `doi:https://doi.org/10.1016/S0168-9002(96)00893-5`.

[8] Adam, W., Frühwirth, R., Strandlie, A., and Todorov, T. Reconstruction of electrons with the gaussian-sum filter in the cms tracker at the lhc. *Journal of Physics G: Nuclear and Particle Physics*, **31** (2005), N9. Available from: `https://dx.doi.org/10.1088/0954-3899/31/9/N01`, `doi:10.1088/0954-3899/31/9/N01`.

[9] Adzic, P., et al. Energy Resolution Performance of the CMS Electromagnetic Calorimeter. Tech. rep., CERN, Geneva (2006). Available from: `https://cds.cern.ch/record/1000388`.

[10] Agarwal, G. Jet Energy Scale and Resolution Measurements in CMS. *PoS*, **ICHEP2022** (2022), 652. arXiv:2301.02175, doi:10.22323/1.414.0652.

[11] Ahmad, Q. R. et al. Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions produced by ^{8}B solar neutrinos at the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, **87** (2001), 071301. arXiv:nucl-ex/0106015, doi:10.1103/PhysRevLett.87.071301.

[12] Alioli, S., Hamilton, K., Nason, P., Oleari, C., and Re, E. Jet pair production in powheg. *Journal of High Energy Physics*, **2011** (2011). Available from: http://dx.doi.org/10.1007/JHEP04(2011)081, doi:10.1007/jhep04(2011)081.

[13] Alioli, S., Nason, P., Oleari, C., and Re, E. Nlo higgs boson production via gluon fusion matched with shower in powheg. *Journal of High Energy Physics*, **2009** (2009), 002. Available from: https://dx.doi.org/10.1088/1126-6708/2009/04/002, doi:10.1088/1126-6708/2009/04/002.

[14] Alioli, S., Nason, P., Oleari, C., and Re, E. A general framework for implementing nlo calculations in shower monte carlo programs: the powheg box. *Journal of High Energy Physics*, **2010** (2010). Available from: http://dx.doi.org/10.1007/JHEP06(2010)043, doi:10.1007/jhep06(2010)043.

[15] Altarelli, G., Giudice, G., and Mangano, M. What if charged current events at large q2 are observed at hera? *Nuclear Physics B*, **506** (1997), 29. Available from: https://www.sciencedirect.com/science/article/pii/S0550321397005348, doi:https://doi.org/10.1016/S0550-3213(97)00534-8.

[16] Altarelli, G. and Parisi, G. Asymptotic freedom in parton language. *Nuclear Physics B*, **126** (1977), 298. Available from: https://www.sciencedirect.com/science/article/pii/0550321377903844, doi:https://doi.org/10.1016/0550-3213(77)90384-4.

[17] Alwall, J., et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, **2014** (2014). Available from: http://dx.doi.org/10.1007/JHEP07(2014)079, doi:10.1007/jhep07(2014)079.

[18] An, F. P., et al. Improved measurement of electron antineutrino disappearance at daya bay. *Chinese Physics C*, **37** (2013), 011001. Available from: http://dx.doi.org/10.1088/1674-1137/37/1/011001, doi:10.1088/1674-1137/37/1/011001.

[19] Apresyan, A. and behalf of the CMS Collaboration), O. Performance of met reconstruction in cms. *Journal of Physics: Conference Series*, **587** (2015), 012006. Available from: https://dx.doi.org/10.1088/1742-6596/587/1/012006, doi:10.1088/1742-6596/587/1/012006.

[20] Arbuzov, A. Quantum Field Theory and the Electroweak Standard Model. (2018), 1. 35 pages. Available from: `https://cds.cern.ch/record/2315477`, `arXiv:1801.05670`, `doi:10.23730/CYRSP-2017-004.1`.

[21] ATLAS Collaboration. Search for new particles in two-jet final states in 7 tev proton-proton collisions with the atlas detector at the lhc. *Phys. Rev. Lett.*, **105** (2010), 161801. Available from: `https://link.aps.org/doi/10.1103/PhysRevLett.105.161801`, `doi:10.1103/PhysRevLett.105.161801`.

[22] ATLAS Collaboration. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, **716** (2012), 1–29. Available from: `http://dx.doi.org/10.1016/j.physletb.2012.08.020`, `doi:10.1016/j.physletb.2012.08.020`.

[23] ATLAS Collaboration. Search for new phenomena in dijet mass and angular distributions from pp collisions with the atlas detector. *Physics Letters B*, **754** (2016), 302–322. Available from: `http://dx.doi.org/10.1016/j.physletb.2016.01.032`, `doi:10.1016/j.physletb.2016.01.032`.

[24] Atlas Collaboration. Search for pairs of scalar leptoquarks decaying into quarks and electrons or muons in $\sqrt{s} = 13$ tev pp collisions with the atlas detector. *Journal of High Energy Physics*, **2020** (2020), 112. Available from: `https://api.semanticscholar.org/CorpusID:260745858`.

[25] ATLAS Collaboration. Search for new phenomena in *pp* collisions in final states with tau leptons, *b*-jets, and missing transverse momentum with the atlas detector. *Phys. Rev. D*, **104** (2021), 112005. Available from: `https://link.aps.org/doi/10.1103/PhysRevD.104.112005`, `doi:10.1103/PhysRevD.104.112005`.

[26] ATLAS Collaboration. Search for pair production of third-generation scalar leptoquarks decaying into a top quark and a τ-lepton in pp collisions at $\sqrt{s} = 13$ tev with the atlas detector. *Journal of High Energy Physics*, **2021** (2021). Available from: `http://dx.doi.org/10.1007/JHEP06(2021)179`, `doi:10.1007/jhep06(2021)179`.

[27] ATLAS Collaboration. Search for leptoquark pair production decaying to $t\ell^-\bar{t}\ell^+$ in multilepton final states in pp collisions at 13 tev with the atlas detector. Tech. rep., CERN, Geneva (2022). Available from: `https://cds.cern.ch/record/2816335`.

[28] ATLAS Collaboration. Search for leptoquark pair production decaying into $te^-\bar{t}e^+$ or $t\mu^-\bar{t}\mu^+$ in multi-lepton final states in *pp* collisions at 13 tev with the atlas detector (2023). `arXiv:2306.17642`.

[29] ATLAS Collaboration. Search for leptoquarks decaying into the bτ final state in pp collisions at $\sqrt{\mathrm{s}} = 13$ TeV with the ATLAS detector. *Journal of High Energy Physics*, **2023** (2023). Available from: `http://dx.doi.org/10.1007/JHEP10(2023)001`, `doi:10.1007/jhep10(2023)001`.

[30] ATLAS COLLABORATION. Search for pair production of third-generation leptoquarks decaying into a bottom quark and a τ-lepton with the ATLAS detector. *Eur. Phys. J. C*, **83** (2023), 1075. Available from: `https://cds.cern.ch/record/2851425`, `arXiv:2303.01294`, `doi:10.1140/epjc/s10052-023-12104-7`.

[31] AUSTRUP, V. A. Search for scalar and vector leptoquarks decaying into quarks and leptons of different generations (2022). Presented 01 Dec 2022. Available from: `https://cds.cern.ch/record/2847970`.

[32] BAAK, M., GOEBEL, M., HALLER, J., HOECKER, A., KENNEDY, D., KOGLER, R., MÖNIG, K., SCHOTT, M., AND STELZER, J. The electroweak fit of the standard model after the discovery of a new boson at the lhc. *The European Physical Journal C*, **72** (2012). Available from: `http://dx.doi.org/10.1140/epjc/s10052-012-2205-9`, `doi:10.1140/epjc/s10052-012-2205-9`.

[33] BABU, K., KOLDA, C., AND MARCH-RUSSELL, J. Implications of a charged-current anomaly at hera. *Physics Letters B*, **408** (1997), 261. Available from: `https://www.sciencedirect.com/science/article/pii/S0370269397007661`, `doi:https://doi.org/10.1016/S0370-2693(97)00766-1`.

[34] BALL, R. D., BERTONE, V., CARRAZZA, S., DEL DEBBIO, L., FORTE, S., GUFFANTI, A., HARTLAND, N. P., AND ROJO, J. Parton distributions with qed corrections. *Nuclear Physics B*, **877** (2013), 290. Available from: `https://www.sciencedirect.com/science/article/pii/S0550321313005233`, `doi:https://doi.org/10.1016/j.nuclphysb.2013.10.010`.

[35] BALL, R. D., BERTONE, V., CERUTTI, F., DEL DEBBIO, L., FORTE, S., GUFFANTI, A., LATORRE, J. I., ROJO, J., AND UBIALI, M. Unbiased global determination of parton distributions and their uncertainties at nnlo and at lo. *Nuclear Physics B*, **855** (2012), 153. Available from: `https://www.sciencedirect.com/science/article/pii/S0550321311005463`, `doi:https://doi.org/10.1016/j.nuclphysb.2011.09.024`.

[36] BALL, R. D., ET AL. Parton distributions from high-precision collider data: Nnpdf collaboration. *The European Physical Journal C*, **77** (2017). Available from: `http://dx.doi.org/10.1140/epjc/s10052-017-5199-5`, `doi:10.1140/epjc/s10052-017-5199-5`.

[37] BARR, S. M., SEGRE, G., AND WELDON, H. A. The Magnitude of the Cosmological Baryon Asymmetry. *Phys. Rev. D*, **20** (1979), 2494. `doi:10.1103/PhysRevD.20.2494`.

[38] BAUER, M. AND NEUBERT, M. Minimal leptoquark explanation for the $R_{D^{(*)}}$, R_K, and $(g-2)_\mu$ anomalies. *Phys. Rev. Lett.*, **116** (2016), 141802. Available from: `https://link.aps.org/doi/10.1103/PhysRevLett.116.141802`, `doi:10.1103/PhysRevLett.116.141802`.

[39] BAYATIAN, G., ET AL. Cms physics technical design report: Addendum on high density qcd with heavy ions. (2007).

[40] Bayatian, G. L. et al. Cms physics: Technical design report volume 1: Detector performance and software. (2006).

[41] Bellm, J. et al. Herwig 7.0/Herwig++ 3.0 release note. *Eur. Phys. J. C*, **76** (2016), 196. arXiv:1512.01178, doi:10.1140/epjc/s10052-016-4018-8.

[42] Bertolini, D., Harris, P., Low, M., and Tran, N. Pileup per particle identification. *Journal of High Energy Physics*, **2014** (2014). Available from: http://dx.doi.org/10.1007/JHEP10(2014)059, doi:10.1007/jhep10(2014)059.

[43] Buchmüller, W., Rückl, R., and Wyler, D. Leptoquarks in lepton-quark collisions. *Physics Letters B*, **191** (1987), 442. Available from: https://www.sciencedirect.com/science/article/pii/037026938790637X, doi:https://doi.org/10.1016/0370-2693(87)90637-X.

[44] Buonocore, L., Greljo, A., Krack, P., Nason, P., Selimović, N., Tramontano, F., and Zanderighi, G. Resonant leptoquark at nlo with powheg. *Journal of High Energy Physics*, **2022** (2022). Available from: http://dx.doi.org/10.1007/JHEP11(2022)129, doi:10.1007/jhep11(2022)129.

[45] Buonocore, L., Haisch, U., Nason, P., Tramontano, F., and Zanderighi, G. Lepton-quark collisions at the large hadron collider. *Phys. Rev. Lett.*, **125** (2020), 231804. Available from: https://link.aps.org/doi/10.1103/PhysRevLett.125.231804, doi:10.1103/PhysRevLett.125.231804.

[46] Buonocore, L., Nason, P., Tramontano, F., and Zanderighi, G. Leptons in the proton. *Journal of High Energy Physics*, **2020** (2020). Available from: http://dx.doi.org/10.1007/JHEP08(2020)019, doi:10.1007/jhep08(2020)019.

[47] Cacciari, M., Salam, G. P., and Soyez, G. The anti-ktjet clustering algorithm. *Journal of High Energy Physics*, **2008** (2008), 063–063. Available from: http://dx.doi.org/10.1088/1126-6708/2008/04/063, doi:10.1088/1126-6708/2008/04/063.

[48] Cacciari, M., Salam, G. P., and Soyez, G. Fastjet user manual: (for version 3.0.2). *The European Physical Journal C*, **72** (2012). Available from: http://dx.doi.org/10.1140/epjc/s10052-012-1896-2, doi:10.1140/epjc/s10052-012-1896-2.

[49] CDF Collaboration. Search for new particles decaying into dijets in proton-antiproton collisions at $\sqrt{s} = 1.96$ TeV. *Phys. Rev. D*, **79** (2009), 112002. Available from: https://link.aps.org/doi/10.1103/PhysRevD.79.112002, doi:10.1103/PhysRevD.79.112002.

[50] Chakraverty, D., Choudhury, D., and Datta, A. A non-supersymmetric resolution of the anomalous muon magnetic moment. *Physics Letters B*, **506** (2001), 103. Available from: https://www.sciencedirect.com/science/article/pii/S0370269301004191, doi:https://doi.org/10.1016/S0370-2693(01)00419-1.

[51] Chowdhury, T. A. and Saad, S. Leptoquark-vectorlike quark model for the cdf anomalies, and neutrino masses. *Physical Review D*, **106** (2022). Available from: `http://dx.doi.org/10.1103/PhysRevD.106.055017`, `doi:10.1103/physrevd.106.055017`.

[52] Chua, C.-K., He, X.-G., and Hwang, W.-Y. Neutrino mass induced radiatively by supersymmetric leptoquarks. *Physics Letters B*, **479** (2000), 224. Available from: `https://www.sciencedirect.com/science/article/pii/S0370269300003257`, `doi:https://doi.org/10.1016/S0370-2693(00)00325-7`.

[53] Cleveland, B. T., Daily, T., Davis, R., Jr., Distel, J. R., Lande, K., Lee, C. K., Wildenhain, P. S., and Ullman, J. Measurement of the solar electron neutrino flux with the Homestake chlorine detector. *Astrophys. J.*, **496** (1998), 505. `doi:10.1086/305343`.

[54] Close, F. The quark parton model. *Reports on Progress in Physics*, **42** (1979), 1285. Available from: `https://dx.doi.org/10.1088/0034-4885/42/8/001`, `doi:10.1088/0034-4885/42/8/001`.

[55] CMS Collaboration. CMS ECAL public plots: Cms-dp-2022-007. April, 2022. Available from: `https://cmsdpgplots.web.cern.ch/#/ecalapproved/CMS-DP-2022_007`.

[56] CMS Collaboration. Pileup mitigation at cms in 13 tev data. *Journal of Instrumentation*, **15**. Available from: `http://dx.doi.org/10.1088/1748-0221/15/09/P09018`, `doi:10.1088/1748-0221/15/09/p09018`.

[57] CMS Collaboration. Public CMS Luminosity Information. Available from: `https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults`.

[58] CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report*. CERN, Geneva (1997). Available from: `https://cds.cern.ch/record/349375`.

[59] CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report*. CERN, Geneva (1997). Available from: `https://cds.cern.ch/record/357153`.

[60] CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical design report. CMS. CERN, Geneva (2006). Available from: `https://cds.cern.ch/record/922757`.

[61] CMS Collaboration. Determination of jet energy calibration and transverse momentum resolution in cms. *Journal of Instrumentation*, **6** (2011), P11002–P11002. Available from: `http://dx.doi.org/10.1088/1748-0221/6/11/P11002`, `doi:10.1088/1748-0221/6/11/p11002`.

[62] CMS Collaboration. Search for resonances in the dijet mass spectrum from 7 tev pp collisions at cms. *Physics Letters B*, **704** (2011), 123–142. Available from: `http://dx.doi.org/10.1016/j.physletb.2011.09.015`, `doi:10.1016/j.physletb.2011.09.015`.

[63] CMS Collaboration. ECAL Detector Performance Plots: more on 2011 data. (2012). Available from: `https://cds.cern.ch/record/1457923`.

[64] CMS Collaboration. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, **716** (2012), 30. Available from: `https://www.sciencedirect.com/science/article/pii/S0370269312008581`, `doi:https://doi.org/10.1016/j.physletb.2012.08.021`.

[65] CMS Collaboration. Search for Narrow Resonances using the Dijet Mass Spectrum in pp Collisions at sqrt s of 7 TeV. Tech. rep., CERN, Geneva (2012). Available from: `https://cds.cern.ch/record/1461223`.

[66] CMS Collaboration. Search for single production of scalar leptoquarks in proton-proton collisions at $\sqrt{s}$= 8 tev. *Physical Review D*, **93** (2016). Available from: `http://dx.doi.org/10.1103/PhysRevD.93.032005`, `doi:10.1103/physrevd.93.032005`.

[67] CMS Collaboration. The cms trigger system. *Journal of Instrumentation*, **12** (2017), P01020–P01020. Available from: `http://dx.doi.org/10.1088/1748-0221/12/01/P01020`, `doi:10.1088/1748-0221/12/01/p01020`.

[68] CMS Collaboration. Jet energy scale and resolution in the cms experiment in pp collisions at 8 tev. *Journal of Instrumentation*, **12** (2017), P02014–P02014. Available from: `http://dx.doi.org/10.1088/1748-0221/12/02/P02014`, `doi:10.1088/1748-0221/12/02/p02014`.

[69] CMS Collaboration. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *JINST*, **12** (2017), P02014. Available from: `https://cds.cern.ch/record/2198719`, `arXiv:1607.03663`, `doi:10.1088/1748-0221/12/02/P02014`.

[70] CMS Collaboration. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, **12** (2017), P10003. Available from: `https://cds.cern.ch/record/2270046`, `arXiv:1706.04965`, `doi:10.1088/1748-0221/12/10/P10003`.

[71] CMS Collaboration. CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV. Tech. rep., CERN, Geneva (2018). Available from: `https://cds.cern.ch/record/2621960`.

[72] CMS Collaboration. Identification of heavy-flavour jets with the cms detector in pp collisions at 13 tev. *Journal of Instrumentation*, **13** (2018), P05011–P05011. Available from: `http://dx.doi.org/10.1088/1748-0221/13/05/P05011`, `doi:10.1088/1748-0221/13/05/p05011`.

[73] CMS Collaboration. Performance of the cms muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$=13 tev. *Journal of Instrumentation*, **13** (2018), P06015–P06015. Available from: http://dx.doi.org/10.1088/1748-0221/13/06/P06015, doi:10.1088/1748-0221/13/06/p06015.

[74] CMS Collaboration. Search for narrow and broad dijet resonances in proton-proton collisions at $\sqrt{s} = 13$ tev and constraints on dark matter mediators and other new particles. *Journal of High Energy Physics*, **2018** (2018). Available from: http://dx.doi.org/10.1007/JHEP08(2018)130, doi:10.1007/jhep08(2018)130.

[75] CMS Collaboration. CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV. Tech. rep., CERN, Geneva (2019). Available from: https://cds.cern.ch/record/2676164.

[76] CMS Collaboration. Muon reconstruction performance during Run II. (2019). Available from: https://cds.cern.ch/record/2682902.

[77] CMS Collaboration. Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ tev using the cms detector. *Journal of Instrumentation*, **14** (2019), P07004–P07004. Available from: http://dx.doi.org/10.1088/1748-0221/14/07/P07004, doi:10.1088/1748-0221/14/07/p07004.

[78] CMS Collaboration. Search for pair production of first-generation scalar leptoquarks at $\sqrt{s} = 13$ tev. *Physical Review D*, **99** (2019). Available from: http://dx.doi.org/10.1103/PhysRevD.99.052002, doi:10.1103/physrevd.99.052002.

[79] CMS Collaboration. Search for pair production of second-generation leptoquarks at $\sqrt{s}$ =13 tev. *Physical Review D*, **99** (2019). Available from: http://dx.doi.org/10.1103/PhysRevD.99.032014, doi:10.1103/physrevd.99.032014.

[80] CMS Collaboration. Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016-2018. (2020). Available from: https://cds.cern.ch/record/2715872.

[81] CMS Collaboration. A multi-dimensional search for new heavy resonances decaying to boosted WW, WZ, or ZZ boson pairs in the dijet final state at 13 TeV. *The European Physical Journal C*, **80** (2020). Available from: http://dx.doi.org/10.1140/epjc/s10052-020-7773-5, doi:10.1140/epjc/s10052-020-7773-5.

[82] CMS Collaboration. Performance of the cms level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ tev. *Journal of Instrumentation*, **15** (2020), P10017–P10017. Available from: http://dx.doi.org/10.1088/1748-0221/15/10/P10017, doi:10.1088/1748-0221/15/10/p10017.

[83] CMS Collaboration. Performance of the reconstruction and identification of high-momentum muons in proton-proton collisions at $\sqrt{s} = 13$ tev. *Journal of Instrumentation*, **15** (2020), P02027–P02027. Available from: http://dx.doi.org/10.1088/1748-0221/15/02/P02027, doi:10.1088/1748-0221/15/02/p02027.

[84] CMS Collaboration. Search for high mass dijet resonances with a new background prediction method in proton-proton collisions at $\sqrt{s} = 13$ tev. *Journal of High Energy Physics*, **2020** (2020). Available from: http://dx.doi.org/10.1007/JHEP05(2020)033, doi:10.1007/jhep05(2020)033.

[85] CMS Collaboration. Search for high mass dijet resonances with a new background prediction method in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, **2005** (2020), 033. Available from: https://cds.cern.ch/record/2699883, arXiv:1911.03947, doi:10.1007/JHEP05(2020)033.

[86] CMS Collaboration. Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. *JINST*, **16** (2021), P05014. Available from: https://cds.cern.ch/record/2747266, arXiv:2012.06888, doi:10.1088/1748-0221/16/05/P05014.

[87] CMS Collaboration. Electron and photon reconstruction and identification with the cms experiment at the cern lhc. *Journal of Instrumentation*, **16** (2021), P05014. Available from: http://dx.doi.org/10.1088/1748-0221/16/05/P05014, doi:10.1088/1748-0221/16/05/p05014.

[88] CMS Collaboration. Jet energy scale and resolution measurement with Run 2 Legacy Data Collected by CMS at 13 TeV. (2021). Available from: https://cds.cern.ch/record/2792322.

[89] CMS Collaboration. Precision luminosity measurement in proton-proton collisions at s = 13 tev in 2015 and 2016 at cms. *The European physical journal. C, Particles and fields*, **81** (2021), 800. doi:10.1140/epjc/s10052-021-09538-2.

[90] CMS Collaboration. ECAL trigger for Run 3. (2022). Available from: https://cds.cern.ch/record/2812785.

[91] CMS Collaboration. ECAL trigger for Run 3. (2022). Available from: https://cds.cern.ch/record/2808229.

[92] CMS Collaboration. Inclusive nonresonant multilepton probes o new phenomena at $\sqrt{s} = 13$ tev. *Phys. Rev. D*, **105** (2022), 112007. Available from: https://cds.cern.ch/record/2801793, arXiv:2202.08676, doi:10.1103/PhysRevD.105.112007.

[93] CMS Collaboration. A search for pair production of leptoquarks decaying to muons and bottom quarks at $\sqrt{s} = 13$ TeV. Tech. rep., CERN, Geneva (2023). Available from: https://cds.cern.ch/record/2868023.

[94] CMS COLLABORATION. Performance of the local reconstruction algorithms for the CMS hadron calorimeter with Run 2 data. *JINST*, **18** (2023), P11017. Available from: https://cds.cern.ch/record/2862211, arXiv:2306.10355, doi:10.1088/1748-0221/18/11/P11017.

[95] CMS COLLABORATION. Performance summary of AK4 jet b tagging with data from proton-proton collisions at 13 TeV with the CMS detector. (2023). Available from: https://cds.cern.ch/record/2854609.

[96] COLLABORATION, C. Performance of the cms hadron calorimeter with cosmic ray muons and lhc beam data. *Journal of Instrumentation*, **5** (2010), T03012. Available from: https://dx.doi.org/10.1088/1748-0221/5/03/T03012, doi:10.1088/1748-0221/5/03/T03012.

[97] COLLABORATION, C. Search for a third-generation leptoquark coupled to a τ lepton and a b quark through single, pair, and nonresonant production in proton-proton collisions at $\sqrt{s} = 13$ tev (2023). arXiv:2308.07826.

[98] COLLABORATION, T. A., ET AL. The alice experiment at the cern lhc. *Journal of Instrumentation*, **3** (2008), S08002. Available from: https://dx.doi.org/10.1088/1748-0221/3/08/S08002, doi:10.1088/1748-0221/3/08/S08002.

[99] COLLABORATION, T. A., ET AL. The atlas experiment at the cern large hadron collider. *Journal of Instrumentation*, **3** (2008), S08003. Available from: https://dx.doi.org/10.1088/1748-0221/3/08/S08003, doi:10.1088/1748-0221/3/08/S08003.

[100] COLLABORATION, T. C., ET AL. The cms experiment at the cern lhc. *Journal of Instrumentation*, **3** (2008), S08004. Available from: https://dx.doi.org/10.1088/1748-0221/3/08/S08004, doi:10.1088/1748-0221/3/08/S08004.

[101] COLLABORATION, T. L., ET AL. The lhcb detector at the lhc. *Journal of Instrumentation*, **3** (2008), S08005. Available from: https://dx.doi.org/10.1088/1748-0221/3/08/S08005, doi:10.1088/1748-0221/3/08/S08005.

[102] COUSINS, R. D. Generalization of chisquare goodness-of-fit test for binned data using saturated models, with application to histograms. Available from: http://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf.

[103] COWAN, G., CRANMER, K., GROSS, E., AND VITELLS, O. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, **71** (2011). Available from: http://dx.doi.org/10.1140/epjc/s10052-011-1554-0, doi:10.1140/epjc/s10052-011-1554-0.

[104] DEVENISH, R. AND COOPER-SARKAR, A. 69QCD improved parton model. In *Deep Inelastic Scattering*. Oxford University Press (2003). ISBN 9780198506713. Available from: https://doi.org/10.1093/acprof:oso/9780198506713.003.0004, arXiv:https://academic.oup.com/book/0/chapter/156624170/chapter-ag-pdf/44947560/book_9589_section_156624170.ag.pdf, doi:10.1093/acprof:oso/9780198506713.003.0004.

[105] Dimopoulos, S. Technicoloured signatures. *Nuclear Physics B*, **168** (1980), 69. Available from: `https://www.sciencedirect.com/science/article/pii/0550321380902771`, `doi:https://doi.org/10.1016/0550-3213(80)90277-1`.

[106] Dokshitzer, Y., Leder, G., Moretti, S., and Webber, B. Better jet clustering algorithms. *Journal of High Energy Physics*, **1997** (1997), 001–001. Available from: `http://dx.doi.org/10.1088/1126-6708/1997/08/001`, `doi:10.1088/1126-6708/1997/08/001`.

[107] Dokshitzer, Y. L. Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics. *Sov. Phys. JETP*, **46** (1977), 641.

[108] Ecker, G. Quantum Chromodynamics. (2006). Available from: `https://cds.cern.ch/record/943008`, `doi:10.5170/CERN-2006-014.55`.

[109] Ellis, S. D. and Soper, D. E. Successive combination jet algorithm for hadron collisions. *Physical Review D*, **48** (1993), 3160–3166. Available from: `http://dx.doi.org/10.1103/PhysRevD.48.3160`, `doi:10.1103/physrevd.48.3160`.

[110] Englert, F. and Brout, R. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, **13** (1964), 321. Available from: `https://link.aps.org/doi/10.1103/PhysRevLett.13.321`, `doi:10.1103/PhysRevLett.13.321`.

[111] Farhi, E. and Susskind, L. Technicolour. *Physics Reports*, **74** (1981), 277. Available from: `https://www.sciencedirect.com/science/article/pii/0370157381901733`, `doi:https://doi.org/10.1016/0370-1573(81)90173-3`.

[112] Foot, R., Lew, H., and Volkas, R. R. Quark-lepton symmetry (1992). `arXiv:hep-ph/9211287`.

[113] Frederix, R. and Frixione, S. Merging meets matching in mc@nlo. *Journal of High Energy Physics*, **2012** (2012). Available from: `http://dx.doi.org/10.1007/JHEP12(2012)061`, `doi:10.1007/jhep12(2012)061`.

[114] Frixione, S., Nason, P., and Oleari, C. Matching nlo qcd computations with parton shower simulations: the powheg method. *Journal of High Energy Physics*, **2007** (2007), 070. Available from: `https://dx.doi.org/10.1088/1126-6708/2007/11/070`, `doi:10.1088/1126-6708/2007/11/070`.

[115] GEANT4 Collaboration. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **506** (2003), 250. Available from: `https://www.sciencedirect.com/science/article/pii/S0168900203013688`, `doi:https://doi.org/10.1016/S0168-9002(03)01368-8`.

[116] Gell-Mann, M., Ramond, P., and Slansky, R. Complex spinors and unified theories (2013). arXiv:1306.4669.

[117] Georgi, H. and Glashow, S. L. Unextended hypercolor and unification. *Phys. Rev. Lett.*, **47** (1981), 1511. Available from: https://link.aps.org/doi/10.1103/PhysRevLett.47.1511, doi:10.1103/PhysRevLett.47.1511.

[118] Georgi, H., Quinn, H. R., and Weinberg, S. Hierarchy of interactions in unified gauge theories. *Phys. Rev. Lett.*, **33** (1974), 451. Available from: https://link.aps.org/doi/10.1103/PhysRevLett.33.451, doi:10.1103/PhysRevLett.33.451.

[119] Gerber, C. E. LHC Highlights and Prospects. (2021). Available from: https://cds.cern.ch/record/2771843, arXiv:1909.10919, doi:10.23730/CYRSP-2021-002.161.

[120] Glashow, S. L. The renormalizability of vector meson interactions. *Nuclear Physics*, **10** (1959), 107. Available from: https://www.sciencedirect.com/science/article/pii/0029558259901968, doi:https://doi.org/10.1016/0029-5582(59)90196-8.

[121] Greljo, A. and Selimovic, N. Lepton-quark fusion at hadron colliders, precisely (2020). arXiv:2012.02092.

[122] Gribov, V. N. and Lipatov, L. N. Deep inelastic e p scattering in perturbation theory. *Sov. J. Nucl. Phys.*, **15** (1972), 438.

[123] Harland-Lang, L. A., Martin, A. D., Motylinski, P., and Thorne, R. S. Parton distributions in the lhc era: Mmht 2014 pdfs. *The European Physical Journal C*, **75** (2015). Available from: http://dx.doi.org/10.1140/epjc/s10052-015-3397-6.

[124] Hayrapetyan, A., et al. Search for scalar leptoquarks produced via τ-lepton–quark scattering in pp collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, **132** (2024), 061801. Available from: https://link.aps.org/doi/10.1103/PhysRevLett.132.061801, doi:10.1103/PhysRevLett.132.061801.

[125] He, S.-P. Leptoquark and vector-like quark extended model for simultaneous-explanation of w boson mass and muon g–2 anomalies*. *Chinese Physics C*, **47** (2023), 043102. Available from: http://dx.doi.org/10.1088/1674-1137/ac9e4c, doi:10.1088/1674-1137/ac9e4c.

[126] Herrero, M. J. The standard model (1998). arXiv:hep-ph/9812242.

[127] Hewett, J. L. and Rizzo, T. G. Don't stop thinking about leptoquarks: Constructing new models. *Phys. Rev. D*, **58** (1998), 055005. Available from: https://link.aps.org/doi/10.1103/PhysRevD.58.055005, doi:10.1103/PhysRevD.58.055005.

[128] Higgs, P. W. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, **13** (1964), 508. Available from: https://link.aps.org/doi/10.1103/PhysRevLett.13.508, doi:10.1103/PhysRevLett.13.508.

[129] Holm, U. The ZEUS detector: Status report 1993. (1993). doi:10.3204/PUBDB-2017-12635.

[130] Ježo, T. and Nason, P. On the treatment of resonances in next-to-leading order calculations matched to a parton shower. *Journal of High Energy Physics*, **2015** (2015), 1–47. Available from: http://dx.doi.org/10.1007/JHEP12(2015)065, doi:10.1007/jhep12(2015)065.

[131] Junk, T. Confidence level computation for combining searches with small statistics. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **434** (1999), 435–443. Available from: http://dx.doi.org/10.1016/S0168-9002(99)00498-2, doi:10.1016/s0168-9002(99)00498-2.

[132] Kalman, R. E. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, **82** (1960), 35.

[133] Karimäki, V., et al. *The CMS tracker system project: Technical Design Report*. Technical design report. CMS. CERN, Geneva (1997). Available from: https://cds.cern.ch/record/368412.

[134] Kirschenmann, H. and behalf of the CMS collaboration), O. Determination of the jet energy scale in cms. *Journal of Physics: Conference Series*, **404** (2012), 012013. Available from: https://dx.doi.org/10.1088/1742-6596/404/1/012013, doi:10.1088/1742-6596/404/1/012013.

[135] Knoll, G. F. *Radiation Detection and Measurement, 3rd ed.* John Wiley and Sons, New York, 3rd edition edn. (2000). ISBN 978-0-471-07338-3, 978-0-471-07338-3.

[136] Lane, K. and Ramana, M. V. Walking technicolor signatures at hadron colliders. *Phys. Rev. D*, **44** (1991), 2678. Available from: https://link.aps.org/doi/10.1103/PhysRevD.44.2678, doi:10.1103/PhysRevD.44.2678.

[137] Layter, J. G. *The CMS muon project: Technical Design Report*. Technical design report. CMS. CERN, Geneva (1997). Available from: https://cds.cern.ch/record/343814.

[138] Li, H. T. and Skands, P. A framework for second-order parton showers. *Physics Letters B*, **771** (2017), 59. Available from: https://www.sciencedirect.com/science/article/pii/S0370269317303684, doi:https://doi.org/10.1016/j.physletb.2017.05.011.

[139] Mahanta, U. Neutrino masses and mixing angles from leptoquark interactions. *Phys. Rev. D*, **62** (2000), 073009. Available from: https://link.aps.org/doi/10.1103/PhysRevD.62.073009, doi:10.1103/PhysRevD.62.073009.

[140] Malara, A. Reconstruction of jets and missing transverse momentum at the cms experiment: Run 2 and perspective for run 3 (2021). arXiv:2012.06271.

[141] Matsumoto, R., et al. Search for proton decay via $p \to \mu^+ K^0$ in 0.37 megaton-years exposure of super-kamiokande. *Phys. Rev. D*, **106** (2022), 072003. Available from: `https://link.aps.org/doi/10.1103/PhysRevD.106.072003`, `doi:10.1103/PhysRevD.106.072003`.

[142] Monti, F. Performance of the CMS Electromagnetic Calorimeter in LHC Run2. Tech. rep., CERN, Geneva (2020). Available from: `https://cds.cern.ch/record/2797453`.

[143] Mukherjee, S. Electron and Photon Performance in CMS in Run-2 and Prospects for Run-3. *Phys. At. Nucl.*, **84** (2021), 571. Available from: `https://cds.cern.ch/record/2800516`, `doi:10.1134/S1063778821040207`.

[144] Nagy, Z. and Soper, D. E. Parton showers with quantum interference. *Journal of High Energy Physics*, **2007** (2007), 114. Available from: `https://dx.doi.org/10.1088/1126-6708/2007/09/114`, `doi:10.1088/1126-6708/2007/09/114`.

[145] Nagy, Z. and Soper, D. E. What is a parton shower? *Physical Review D*, **98** (2018). Available from: `http://dx.doi.org/10.1103/PhysRevD.98.014034`, `doi:10.1103/physrevd.98.014034`.

[146] Nakamura, K. and Group), P. D. Review of particle physics. *Journal of Physics G: Nuclear and Particle Physics*, **37** (2010), 075021. Available from: `https://dx.doi.org/10.1088/0954-3899/37/7A/075021`, `doi:10.1088/0954-3899/37/7A/075021`.

[147] Nason, P. A new method for combining nlo qcd with shower monte carlo algorithms. *Journal of High Energy Physics*, **2004** (2004), 040. Available from: `http://dx.doi.org/10.1088/1126-6708/2004/11/040`, `doi:10.1088/1126-6708/2004/11/040`.

[148] Noether, E. Invariant variation problems. *Transport Theory and Statistical Physics*, **1** (1971), 186–207. Available from: `http://dx.doi.org/10.1080/00411457108231446`, `doi:10.1080/00411457108231446`.

[149] Ohlsson, T. Proton decay. *Nuclear Physics B*, **993** (2023), 116268. Available from: `https://www.sciencedirect.com/science/article/pii/S0550321323001979`, `doi:https://doi.org/10.1016/j.nuclphysb.2023.116268`.

[150] Oort, J. H. The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems. *Bull. Astron. Inst. Netherlands*, **6** (1932), 249. Available from: `https://cds.cern.ch/record/436532`.

[151] P. Ade et al. Planck2013 results. i. overview of products and scientific results. *Astronomy amp; Astrophysics*, **571** (2014), A1. Available from: `http://dx.doi.org/10.1051/0004-6361/201321529`, `doi:10.1051/0004-6361/201321529`.

[152] Particle Data Group. Review of particle physics. *Phys. Rev. D*, **98** (2018), 030001. Available from: `https://link.aps.org/doi/10.1103/PhysRevD.98.030001`, `doi:10.1103/PhysRevD.98.030001`.

[153] Particle Data Group. Review of Particle Physics, 2018-2019. *Phys. Rev. D*, **98** (2018), 030001. Available from: `https://cds.cern.ch/record/2636832`, `doi:10.1103/PhysRevD.98.030001`.

[154] Pati, J. C. and Salam, A. Unified lepton-hadron symmetry and a gauge theory of the basic interactions. *Phys. Rev. D*, **8** (1973), 1240. Available from: `https://link.aps.org/doi/10.1103/PhysRevD.8.1240`.

[155] Peskin, M. E. and Schroeder, D. V. *An Introduction To Quantum Field Theory*. Westview Press (1995).

[156] Petridou, C. Experimental tests of the standard model using the recently collected ua2 data. *Nuclear Physics B - Proceedings Supplements*, **22** (1991), 8. Available from: `https://www.sciencedirect.com/science/article/pii/092056329190346G`, `doi:https://doi.org/10.1016/0920-5632(91)90346-G`.

[157] Petrucciani, G., Rizzi, A., and Vuosalo, C. Mini-AOD: A New Analysis Data Format for CMS. A new analysis data format for CMS. Tech. Rep. 7, CERN, Geneva (2015). Available from: `https://cds.cern.ch/record/2029414`, `arXiv:1702.04685`, `doi:10.1088/1742-6596/664/7/072052`.

[158] Petyt, D. A. Mitigation of anomalous APD signals in the CMS electromagnetic calorimeter. *J. Phys. Conf. Ser.*, **404** (2012), 012043. `doi:10.1088/1742-6596/404/1/012043`.

[159] Polchinski, J. *String Theory*. Cambridge Monographs on Mathematical Physics. Cambridge University Press (1998).

[160] Punzi, G. Sensitivity of searches for new signals and its optimization. *eConf*, **C030908** (2003), MODT002. `arXiv:physics/0308063`.

[161] Read, A. Linear interpolation of histograms. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **425** (1999), 357. Available from: `https://www.sciencedirect.com/science/article/pii/S0168900298013473`, `doi:https://doi.org/10.1016/S0168-9002(98)01347-3`.

[162] Read, A. L. Presentation of search results: the cls technique. **28** (2002), 2693. Available from: `https://dx.doi.org/10.1088/0954-3899/28/10/313`, `doi:10.1088/0954-3899/28/10/313`.

[163] Salam, A. Weak and Electromagnetic Interactions. *Conf. Proc. C*, **680519** (1968), 367. `doi:10.1142/9789812795915\_0034`.

[164] Schrempp, B. and Schrempp, F. Light leptoquarks. *Physics Letters B*, **153** (1985), 101. Available from: `https://www.sciencedirect.`

com/science/article/pii/0370269385914509, doi:https://doi.org/10.1016/0370-2693(85)91450-9.

[165] SERT, H. CMS Run 2 High Level Trigger Performance. CMS High Level Trigger performance in Run 2. Tech. rep., CERN, Geneva (2020). Available from: https://cds.cern.ch/record/2797787, doi:10.22323/1.364.0165.

[166] SHER, M. Flavor-changing neutral currents in the higgs sector. *Modern Physics Letters A*, **37** (2022). Available from: http://dx.doi.org/10.1142/S0217732322300117, doi:10.1142/s0217732322300117.

[167] SIRUNYAN, A. M. ET AL. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV. *JINST*, **13** (2018), P06015. arXiv:1804.04528, doi:10.1088/1748-0221/13/06/P06015.

[168] THERHAAG, J. TMVA Toolkit for multivariate data analysis in ROOT. *PoS*, **ICHEP2010** (2010), 510. doi:10.22323/1.120.0510.

[169] VALSECCHI, D. First evidence of VBS in semileptonic decays with $WVjj \to l\nu qqjj$ final state and optimization of the CMS electromagnetic calorimeter for Run III. Available from: https://boa.unimib.it/handle/10281/364125.

[170] VAN DER MEER, S. Calibration of the effective beam height in the ISR. Tech. rep., CERN, Geneva (1968). Available from: https://cds.cern.ch/record/296752.

[171] WEINBERG, S. A model of leptons. *Phys. Rev. Lett.*, **19** (1967), 1264. Available from: https://link.aps.org/doi/10.1103/PhysRevLett.19.1264, doi:10.1103/PhysRevLett.19.1264.

[172] WIEDEMANN, U. A. HIP and HEP. HIP and HEP. *PoS*, **ICHEP2020** (2021), 046. Available from: https://cds.cern.ch/record/2748813, arXiv:2101.01971, doi:10.22323/1.390.0046.

[173] WUDKA, J. Composite leptoquarks. *Physics Letters B*, **167** (1986), 337. Available from: https://www.sciencedirect.com/science/article/pii/0370269386903564, doi:https://doi.org/10.1016/0370-2693(86)90356-4.