

# **Machine Learning-enhanced feasibility studies on the production of new particles with preferential couplings to third generation fermions at the LHC**

Thesis submitted in partial fulfillment  
of the requirements for the degree of  
**Doctor in Sciences - Physics**

**Presented by:**  
Cristian Fernando Rodríguez Cruz

**Directed by:**  
Prof. Andrés Florez (advisor)  
Universidad de los Andes  
Prof. Joel Jones-Perez (co-advisor)  
Pontificia Universidad Católica del Perú

High Energy Physics Research Group  
Phenomenology of Particle Physics

**Universidad de los Andes**  
Faculty of Sciences  
Department of Physics  
Bogotá D.C., Colombia  
September 21, 2025. Pre-alpha version



### **Dedication**

to be added



## Acknowledgements

to be added



## Abstract

The Standard Model (SM) of particle physics is the most successful framework for describing the subatomic world. It is continuously tested in experiments worldwide, with the Large Hadron Collider (LHC) being the flagship project in this endeavor. One of the primary goals of the LHC is to precisely measure SM parameters and search for deviations that could signal new physics.

In recent years, reported anomalies, such as those in B-meson decays from LHCb, BaBar, and Belle experiments, along with the potential discrepancy in the muon's magnetic moment ( $g - 2$ ) from Fermilab, suggest a violation of lepton flavor universality (LFU). These observations provide a compelling window into physics beyond the SM. Among the proposed SM extensions to explain LFU violation, many introduce new particles with preferential couplings to third and second-generation fermions. Popular candidates include heavy states such as  $Z'$  bosons,  $\phi'$  scalars, and leptoquarks (LQs), among others.

This work presents two phenomenological studies proposing different strategies to probe new models, such as the 4321 [1],  $U(1)_{T_R^3}$  [2], that extend the SM particle content to explain clues on LFU violation. The studies use benchmark scenarios where the new particle fields have preferential couplings to second and third generation SM-fermions, depending on the model. The hypothetical signal and background samples are generated using Monte Carlo simulations, emulating the current running conditions of the LHC and the performance of the CMS detector. The expected sensitivity for the different signal models under study is obtained by performing a detailed analysis of the available (non-excluded) experimental phase-space, boosted by machine learning (ML) techniques to optimize the discovery potential for these exotic states.



---

## CONTENTS

<b>Contents</b>	<b>ix</b>
<b>INTRODUCTION</b>	<b>1</b>
<b>1 PHENOMENOLOGICAL FRAMEWORK</b>	<b>5</b>
<b>1.1 Detectors and Subsystems</b> . . . . .	6
<b>1.1.1 Collision Parameters</b> . . . . .	11
<b>1.2 Jets Reconstruction</b> . . . . .	12
<b>1.2.1 Jet algorithms</b> . . . . .	13
<b>1.2.2 <math>\tau</math> Tagging at Multipurpose Detectors</b> . . . . .	15
<b>1.2.3 B Tagging at Multipurpose Detectors</b> . . . . .	16
<b>1.3 The CMS Detector</b> . . . . .	17
<b>BIBLIOGRAPHY</b>	<b>21</b>
<b>PUBLICATIONS</b>	<b>27</b>
On the sensitivity reach of vectorial leptoquark production with preferential couplings to third generation fermions at the LHC . . . . .	29
Probing light scalars and vector-like quarks at the high-luminosity LHC . . . . .	47



---

## INTRODUCTION

---

The pursuit of a fundamental description of nature’s building blocks and their interactions is a central endeavor of modern physics. This quest has led to the development of the Standard Model (SM) of particle physics, a quantum field theory that encapsulates our current understanding of the subatomic world. With breathtaking precision, the SM describes the electromagnetic, weak, and strong nuclear forces and classifies all known elementary particles. Its triumphs are undeniable, crowned by the landmark discovery of the Higgs boson at the Large Hadron Collider (LHC) in 2012, which confirmed the mechanism for generating mass and represented the final piece of the SM puzzle.

Yet, for all its success, the Standard Model is universally acknowledged to be an incomplete theory. It offers no candidate for dark matter, cannot account for the asymmetry between matter and antimatter in the universe, does not incorporate gravity, and leaves the mass of the Higgs boson itself unnaturally unstable under quantum corrections—a problem known as the hierarchy problem. These profound theoretical shortcomings provide a clear motivation for physics beyond the Standard Model (BSM). However, the most compelling guide for this search has always come from experimental data itself.

The primary mission of the LHC is not only to consolidate the SM but to probe its boundaries and discover new physics. While no direct evidence of new particles has been found so far, a series of subtle but persistent discrepancies—termed “anomalies”—have emerged from experiments worldwide, suggesting a potential crack in the SM’s foundation.

A particularly intriguing set of these anomalies points towards a violation of Lepton Flavor Universality (LFU). The SM predicts that the electroweak force couples with identical strength to the three charged leptons (electrons, muons, and taus), a fundamental principle known as LFU. The most significant and long-standing hints of LFU violation come from measurements of semileptonic B-meson decays. The ratios  $R(D^{(*)}) = \mathcal{B}(B \rightarrow D^{(*)}\tau\nu_\tau)/\mathcal{B}(B \rightarrow D^{(*)}\ell\nu_\ell)$ , where  $\ell$  is a muon or electron, have been measured by the BaBar, Belle, and LHCb collaborations to consistently exceed the SM predictions by a combined significance of approximately  $3\sigma$ - $4\sigma$ . This deviation suggests that B mesons are more likely to decay to a final state containing a tau lepton than the SM allows, providing a compelling hint of new physics that couples preferentially to the third generation. Furthermore, the longstanding discrepancy in the muon’s anomalous magnetic moment ( $g - 2$ ), recently confirmed with increased precision by the Fermilab experiment, adds another layer of

intrigue, as it also hints at new physics potentially coupled preferentially to the second generation.

While each anomaly individually requires careful scrutiny, their collective persistence has generated significant excitement, as they seem to point towards new physics that breaks lepton flavor universality, potentially involving enhanced couplings to heavier fermions.

The pattern of these LFU-violating anomalies has inspired a vast landscape of theoretical models extending the SM. A common thread among the most promising explanations is the introduction of new heavy particles that mediate interactions with non-universal couplings to the different generations of fermions. This generational hierarchy is crucial to evade tight constraints from precision measurements on electrons (first generation) while affecting processes involving muons and taus.

In this thesis, we contextualize and present two of our phenomenological studies that propose different strategies to probe new physics models, such as the 4321 [1] and  $U(1)_{T_R^3}$  [2] models, which extend the SM particle content to explain the observed LFU violation. These models introduce new particles with preferential couplings to second and third-generation fermions, making them prime candidates for explaining the experimental anomalies.

The experimental challenge lies in probing these models at the LHC. The proposed new particles are often heavy, leading to low production rates, and their decay signatures are complex and overwhelmed by enormous Standard Model backgrounds. Given the immense number of theoretical possibilities and the finite resources available to experimental collaborations, it is impossible to pursue every potential signature with equal vigor. This is where **phenomenological feasibility studies** become critical. They provide a vital bridge between theory and experiment by performing a detailed *a priori* assessment of the discovery potential for a given signal model. By using Monte Carlo simulations to emulate the detector response and analysis chain, these studies can identify the most promising signatures, optimize event selection criteria, and estimate the sensitivity achievable with the available data. This process is essential for prioritizing the experimental program, justifying the dedication of significant computing and human resources to a particular search, and ultimately guiding the LHC experiments towards the most well-motivated and detectable signals of new physics.

This thesis contributes to this effort by presenting two dedicated phenomenological studies that propose and develop novel strategies to probe the 4321 and  $U(1)_{T_R^3}$  models at the LHC. The work is situated at the intersection of theoretical model-building and experimental high-energy physics, with the explicit goal of assessing the feasibility of these searches.

The core methodology of this research involves:

1. Defining **benchmark scenarios** within each model, selecting specific mass points and coupling structures that explain the LFU anomalies while remaining experimentally viable.

2. Using **Monte Carlo simulation** to accurately generate the hypothetical signal processes alongside the dominant SM background processes, emulating the run conditions of the LHC and the performance of the CMS detector.
3. Performing a detailed analysis of the available experimental phase-space. Given the high-dimensionality of the final states (e.g., involving multiple jets, leptons, and missing energy) and the complex, overlapping kinematical distributions of signal and background, traditional “cut-and-count” analyses are often sub-optimal. To overcome this, we employ advanced **Machine Learning (ML) techniques**, specifically supervised learning algorithms such as Boosted Decision Trees (BDTs) or Deep Neural Networks (DNNs). These algorithms are trained to learn the complex, non-linear correlations between many kinematic variables (e.g., invariant masses, angular separations, transverse momenta) to construct powerful discriminators that optimally separate the rare signal events from the large and diverse SM backgrounds. This ML-enhanced approach significantly boosts the analysis sensitivity, allowing for the detection of weaker signals or the setting of more stringent limits than would otherwise be possible.
4. Deriving the **expected sensitivity** for each model, establishing the exclusion limits or discovery potential that the LHC experiments could achieve with the current dataset. This final step is the ultimate quantitative measure of the search’s feasibility.

The structure of this thesis is as follows. We begin by establishing the theoretical foundation with a review of the Standard Model in Chapter 2. Chapter 3 then details the experimental context, describing the LHC and the CMS detector, and introduces the general analysis techniques employed, including a discussion on the application of Machine Learning in high-energy physics. The original phenomenological work of this thesis is presented in the subsequent chapters: Chapter 4 details a search for new physics in the process  $pp \rightarrow t\bar{t}\mu^+\mu^-$ , while Chapter 5 presents a search for vector leptoquarks in the process  $pp \rightarrow \tau^+\tau^- + b\text{-jets}$ . Finally, Chapter 6 concludes by summarizing our findings and discussing their implications for the field, along with an outlook on future prospects.



---

## PHENOMENOLOGICAL FRAMEWORK FOR LHC SEARCHES AND ANALYSIS

---

Since its formulation, the Standard Model (SM) has proven remarkably successful in describing the fundamental particles and interactions, and its parameters have been measured with increasing precision over several decades [3, 4]. However, as discussed in the previous chapter, various theoretical and experimental observations suggest that the SM is incomplete [5, 6]. As outlined previously, this is motivated by theoretical shortcomings such as the hierarchy problem [7, 8], the absence of a dark matter candidate [9], and non-zero neutrino masses [10] as well as by experimental anomalies [11–13]. These limitations motivate the search for new physics (NP) beyond the SM (BSM) [6, 14].

The search for BSM physics proceeds along two main axes: the construction of theoretical extensions to the SM [6, 8, 14], and the development of experimental methods to probe them [15–17]. A necessary condition for any viable BSM model is consistency with existing experimental data, which places strong constraints on its parameter space [3, 18, 19]. These constraints include lower limits on the masses of new particles from direct searches at high-energy colliders [18, 19], and upper bounds on couplings and mixing angles from precision measurements at both high and low energies [20, 21], which are sensitive to virtual corrections [22].

Phenomenology connects theoretical models to experimental observables by calculating cross sections, decay rates, and other signatures for given model parameters [15, 23, 24]. A critical function of this field is to assess the experimental feasibility of BSM scenarios—evaluating whether predicted signals would be observable above background processes given the capabilities of current and future detectors [15–17]. This involves estimating production rates [15, 23, 25], modeling detector acceptance and efficiency [17, 26–28], and developing discrimination variables to maximize signal-to-background ratios [29–32]. This feasibility assessment is essential for designing analysis strategies, particularly at the Large Hadron Collider (LHC), where signals of new physics must be discriminated from large Standard Model backgrounds [15, 17].

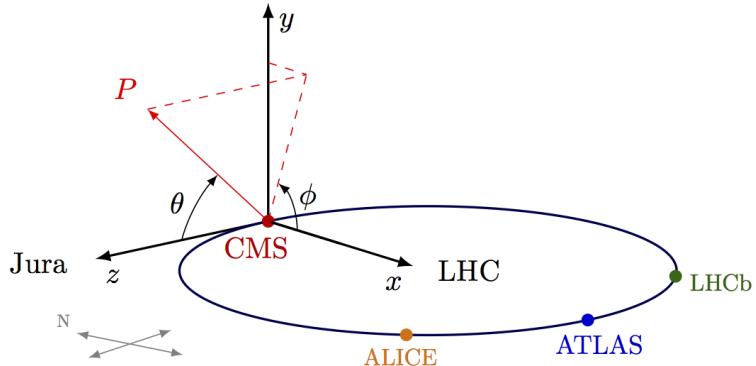
This feasibility assessment is essential for designing analysis strategies at the Large Hadron Collider (LHC), a proton-proton (pp) collider operating since 2009 [26, 27, 33]. The LHC has provided data at center-of-mass energies from 7 TeV to 13.6 TeV [19, 28]. During Run I (2010–2013), operations at 7–8 TeV led to the discovery of the Higgs boson using a

dataset corresponding to an integrated luminosity of roughly  $30 \text{ fb}^{-1}$  [34, 35]. Run II (2015–2018) significantly expanded this dataset, collecting approximately  $140 \text{ fb}^{-1}$  at 13 TeV [19, 28]. Run III (2022–2025) is currently underway at 13.6 TeV and is more than doubling the available data, with a target of over  $300 \text{ fb}^{-1}$  [28]. Future operations will be dominated by the High-Luminosity LHC (HL-LHC), starting around 2029, which is designed to accumulate an unprecedented integrated luminosity of  $3000 \text{ fb}^{-1}$  [36, 37]. This vast increase in data volume enables searches for exceedingly rare processes but also requires discriminating potential signals of new physics from correspondingly large and complex SM backgrounds, making sophisticated phenomenological tools increasingly important [15, 17].

### 1.1 DETECTORS AND SUBSYSTEMS

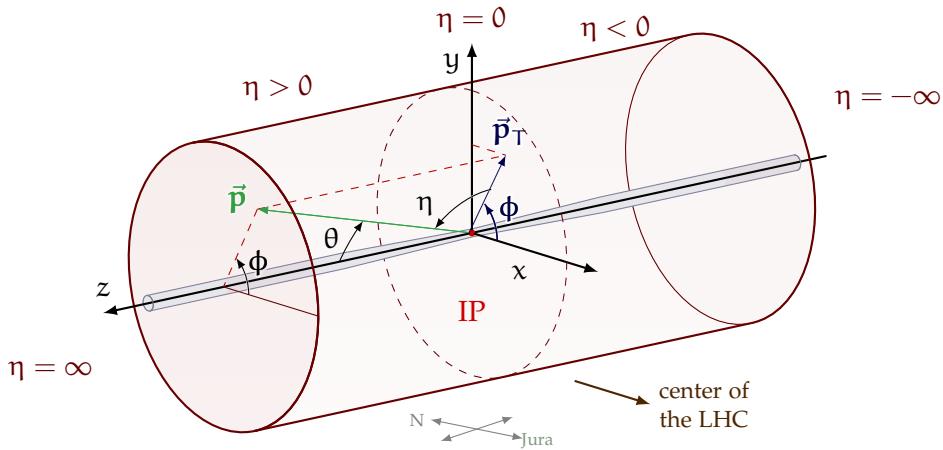
When two particle bunches from colliding beams cross each other, they generate individual interactions known as events [26, 27]. At the LHC, the beam intensity is so high that multiple interactions can take place in a single event; this phenomenon is referred to as pile-up [37, 38]. In other words, the probability that several proton-proton interactions occur within the same bunch crossing is non-negligible, leading to multiple overlapping events in a single detector readout [24, 39]. These collisions occur at four main interaction points, each hosting a large particle detector designed to record and analyze the outcomes [26, 27].

Among them, the Compact Muon Solenoid (CMS) and ATLAS are the largest and most comprehensive experiments [26, 27]. Both are multipurpose detectors with broad physics programs, capable of exploring a wide range of phenomena [26, 27]. They perform precision measurements within the electroweak sector of the Standard Model [3], probe the dynamics of quarks and gluons (including through heavy-ion collisions) [17], and conduct extensive searches for physics beyond the Standard Model using pp collision data [34, 35]. While CMS and ATLAS differ in their detector designs and reconstruction strategies, their physics goals are largely overlapping, and their results are complementary [26, 27].



**Figure 1.1:** Coordinate system employed by the CMS experiment (retrieved from [40]).

Throughout this work, phenomenological studies and comparisons are primarily developed in the context of CMS, although several results from ATLAS are also referenced, given the close alignment in sensitivity and scope [26, 27]. Measurements performed at CMS adopt a right-handed coordinate system with its origin at the nominal collision point [27]. The z-axis is defined along the beam direction, the x-axis points radially inward toward the center of the LHC ring, and the y-axis points vertically upward [27]. The azimuthal angle  $\phi$  is measured in the transverse (xy) plane from the x-axis, while the polar angle  $\theta$  is measured from the z-axis, as shown in Fig. 1.1 [27]. Moreover, for kinematic analysis at hadron colliders, the Cartesian coordinate system is often reparameterized into quantities that are more physically meaningful and experimentally convenient as shown in Fig. 1.2 [41]:



**Figure 1.2:** Detailed reparametrization of the coordinate system employed by the CMS experiment (retrieved from [40])

**PSEUDO-RAPIDITY** ( $\eta$ ) Instead of using the polar angle, CMS measurements involve the pseudo-rapidity [3, 27], defined by

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

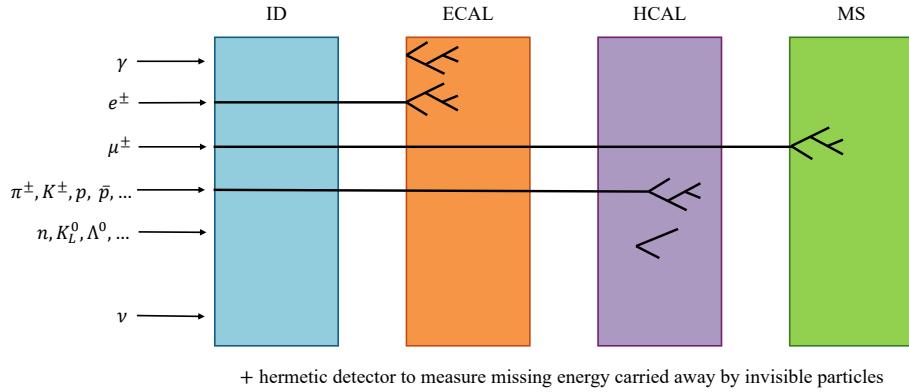
The main advantage of using the pseudo-rapidity is that distributions over it tend to be closer to a uniform distribution than those over the polar angle [41], see Fig. 1.2. Furthermore, the difference in pseudo-rapidity is invariant under Lorentz boosts along the beam direction [3].

**TRANSVERSE MOMENTUM** ( $p_T$ ) Refers to the component of momentum which is perpendicular to the beam line [41]. It is usually preferred over full momentum because momentum along the beamline may just be left over from the beam particles, while the transverse momentum is always associated with whatever physics happened at the vertex [41], see Fig. 1.2.

**AZIMUTHAL ANGLE** ( $\phi$ ) Measures the angle in the transverse plane relative to the  $x$ -axis, providing the directional component perpendicular to the beam line [27].

Together, the triplet  $(p_T, \phi, \eta)$  forms a natural coordinate system that fully describes a particle's three-momentum vector at a hadron collider [3, 41]. The full four-momentum  $(E, p_x, p_y, p_z)$  can be reconstructed from these quantities, typically supplemented by either the particle's mass hypothesis (for identified particles like electrons or muons) or the energy deposited in the calorimeters (for neutral objects like photons or jets) [24, 42, 43]. This  $(p_T, \phi, \eta)$  system serves as the fundamental framework for defining physical objects, calculating event variables, and performing analyses at the LHC, providing both experimental convenience and physical insight into the collision dynamics [24, 41].

A key challenge is isolating the primary hard interaction from the additional concurrent pile-up interactions [37, 38]. This is accomplished by reconstructing distinct interaction vertices along the beam direction and associating charged particles to their point of origin using the CMS tracking and vertexing algorithms [41, 44]. The ultimate aim of the reconstruction chain is to identify all stable particles produced in the collision and measure their four-momenta, thereby enabling the identification of the underlying fundamental process [41].

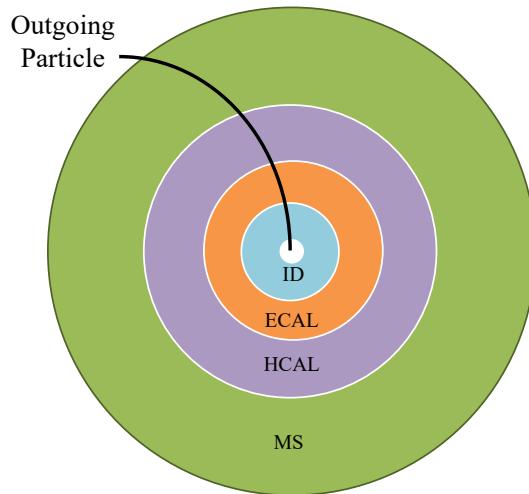


**Figure 1.3:** Illustration of high-energy particles being identified by consecutive types of subdetectors in a typical collider experiment. The curvature of the tracks in the magnetic field is not shown for simplicity. Representation of which particles and kinds of detectors are used in a multipurpose detector such as CMS or ATLAS.

However, the reconstruction is complicated by several factors [27, 41]. The initial state of the colliding protons is not fully known, as they are composite particles made up of quarks and gluons (collectively referred to as partons) [25, 45]. The fraction of the proton's momentum carried by each parton is described by parton distribution functions (PDFs), which are determined experimentally and recommended for LHC Run II/III use by PDF4LHC efforts [25, 46]. Consequently, the total momentum along the beam axis ( $z$ ) is not balanced on an event-by-event basis [45].

Furthermore, not all particles are stable enough to reach the detector; some decay before being detected, and only their decay products are observed [3]. The design of a collider experiment, illustrated in Fig. 1.3, is optimized for the identification and energy measurement of the particles produced in high-energy collisions [17, 27].

Finally, some hypothetical particles, such as those comprising dark matter, along with known neutrinos, interact very weakly with matter and escape direct detection [3, 9]. Therefore, a hermetic detector design is crucial to infer their presence by accurately measuring the imbalance of energy and momentum in the transverse plane, referred to as missing transverse momentum [41, 47].



**Figure 1.4:** Schematic representation of a transverse section of a generic multipurpose detector. The inner detector (ID) is used to measure the trajectories of charged particles, the electromagnetic calorimeter (ECAL) measures the energy of photons and electrons, the hadronic calorimeter (HCAL) measures the energy of hadrons, and the muon system (MS) identifies and measures muons. The missing transverse momentum (MET) is inferred from the momentum imbalance in the transverse plane.

In this way, a typical collider experiment comprises several main detector subsystems that are used jointly to detect and measure the properties of particles produced in the collision [17, 26, 27, 41]. A *schematic representation* of such a generic multipurpose detector is shown in Fig. 1.4 [17, 27]. The detector features an "onion-like" design of several concentric layers, each optimized to identify different types of particles and measure their properties [27, 41].

The innermost subsystem, the inner detector (ID) or tracker, is immersed in a strong axial magnetic field (typically 1–4 T) [27, 44]. It is designed to reconstruct the trajectories of charged particles, which are bent by the magnetic field [41, 44]. The direction and curvature of these trajectories,

called **tracks**, yield the particle's momentum vector and electric charge [3, 44]. The most common long-lived charged particles from the SM are leptons (electrons  $e$  and muons  $\mu$ ) and hadrons (pions  $\pi$ , kaons  $K$ , and protons  $p$ ) [3]. In some detectors, the ID is complemented by a Cherenkov light detector (RICH) to measure particle velocity and aid particle identification [3, 48]. Combined with the momentum measurement, this velocity helps determine the particle mass, allowing for differentiation between pions, kaons, and protons [3, 48].

After the tracker, particles enter the electromagnetic calorimeter (ECAL), which is designed to fully absorb photons, electrons, and positrons [27, 42]. These particles deposit all their energy in the ECAL by initiating an electromagnetic shower via bremsstrahlung and  $e^+e^-$  pair production [42]. Electrons are identified as charged tracks that point to a compact, high-energy deposit in the ECAL [42].

The hadronic calorimeter (HCAL) surrounds the ECAL and is built to absorb hadrons and measure their energy through hadronic interactions [17, 27]. High-energy quarks and gluons hadronize into collimated sprays of hadrons known as **jets**; a jet's energy is measured by combining calorimeter deposits with track momenta, an approach formalized in particle-flow reconstruction [24, 39, 41].

Muons are unique as they can penetrate the calorimeters; a dedicated muon system outside the calorimeters identifies and measures muons, and muon tracks in the ID are matched to tracks in the muon chambers [27, 43]. Since the detector is nearly hermetic, momentum conservation in the transverse plane is a powerful tool: any significant imbalance (missing transverse momentum, MET) signals undetected neutral particles such as neutrinos or potential dark-matter candidates [41, 47].

Since the detector is nearly hermetic (covering almost the full solid angle), momentum conservation in the plane transverse to the beam line ( $x$ - $y$  plane) is a powerful tool. The vector sum of the momenta in the transverse plane ( $\vec{p}_T$ ) of all detected particles should be zero. Any significant imbalance indicates the presence of undetected, neutral particles that did not interact with the detector, such as neutrinos or hypothetical new particles. This imbalance is called missing transverse momentum (MET) and is formally defined as:

$$\vec{p}_T^{\text{miss}} \equiv - \sum_i \vec{p}_{T,i}$$

where the sum runs over all reconstructed particles (e.g., leptons, photons, jets) or calorimeter deposits in the event.

This detector design, optimized for identifying and measuring Standard Model particles, also makes it a powerful instrument for searching for new physics beyond the SM through unusual signatures or an excess of events with large MET.

### 1.1.1 COLLISION PARAMETERS

A aim of particle physics experiments is to quantify how frequently different processes occur and to characterize the properties of the particles involved. These quantities are expressed in terms of **cross-sections**, normalized to the **luminosity** delivered by the accelerator.

The cross-section ( $\sigma$ ) quantifies the probability for a specific process to occur. Formally, it represents the effective area of a target particle presented to an incoming beam particle for an interaction to happen. It has units of area, typically barn (b), where  $1 \text{ b} = 10^{-28} \text{ m}^2$ .

In the context of proton-proton (pp) collisions at the LHC, the concept is generalized. Since both colliding particles are composite, the cross-section for a specific process is calculated by considering the interactions between their constituent partons (quarks and gluons). The total cross-section for a process  $\text{pp} \rightarrow X$  is given by the convolution of the parton distribution functions (PDFs) and the partonic cross-section  $\hat{\sigma}_{ij \rightarrow X}$ :

$$\sigma(\text{pp} \rightarrow X) = \sum_{i,j} \int_0^1 dx_1 dx_2 f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \hat{\sigma}_{ij \rightarrow X}(\hat{s}, \mu_F^2, \mu_R^2), \quad (1.1)$$

where:

- The sum runs over all possible parton types  $i, j$  (e.g.,  $u, d, g$ ) in the two protons.
- $f_i(x, \mu_F^2)$  is the PDF, representing the probability density to find a parton of type  $i$  carrying a fraction  $x$  of the proton's momentum at a factorization scale  $\mu_F$ .
- $\hat{s} = x_1 x_2 s$  is the square of the center-of-mass energy for the colliding partons, with  $s$  being the square of the pp center-of-mass energy (e.g., 13.6 TeV).
- $\mu_R$  is the renormalization scale.
- $\hat{\sigma}_{ij \rightarrow X}$  is the partonic cross-section for the hard scattering process  $ij \rightarrow X$ .

Then, in one hand, the cross-section  $\sigma$  is a theoretical quantity that encapsulates the fundamental physics of the interaction, independent of the accelerator's performance. On the other hand, the **luminosity** ( $\mathcal{L}$ ) is a property of the particle accelerator and beams. It measures the density of particles in the colliding beams and thus the rate at which interactions can occur. The instantaneous luminosity is defined by:

$$\mathcal{L} = \frac{fn_1 n_2}{4\pi\sigma_x\sigma_y}, \quad (1.2)$$

where  $f$  is the revolution frequency of the bunches,  $n_1$  and  $n_2$  are the numbers of particles in each bunch, and  $\sigma_x$  and  $\sigma_y$  are the transverse

dimensions of the beams at the interaction point. The integrated luminosity is the integral of the instantaneous luminosity over time:

$$L = \int \mathcal{L} dt. \quad (1.3)$$

The primary unit of integrated luminosity is the inverse barn ( $b^{-1}$ ), commonly  $fb^{-1}$ .

The expected number of events ( $N$ ) for a given process in a collider experiment connects the cross-section and integrated luminosity through the relation

$$N = \sigma \cdot L \cdot \epsilon. \quad (1.4)$$

Here,  $\epsilon$  is the product efficiency  $\times$  acceptance of the detector for the process in question [17, 41]. It estimates the fraction of produced events that are actually detected and reconstructed, accounting for detector geometry, resolution, and analysis selection criteria [41].

On one side, if a process is known, we can measure the number of events  $N$ , know the integrated luminosity  $L$  from accelerator measurements [28], and calculate the efficiency  $\epsilon$  from simulation (event generation, parton shower and detector simulation) [15, 17, 23]. Equation 1.4 is then solved for  $\sigma$  to provide a measurement of the production rate, with statistical interpretation based on likelihood methods [49]. On the other side, if we are searching for a new process not predicted by the Standard Model, we calculate the expected background  $N_{bkg}$  from known SM processes using Monte Carlo and data-driven techniques [15, 24]. We then observe a number of events  $N_{obs}$ . Any significant excess,  $N_{obs} - N_{bkg}$ , can be interpreted as a potential signal, and Equation 1.4 can be used to set an upper limit on the cross-section  $\sigma_{BSM}$  of the hypothetical new process using standard limit-setting procedures (CLs, profile-likelihood, treatments of nuisances) [50–52].

## 1.2 JETS RECONSTRUCTION

---

Quarks and gluons are never observed as free particles because of colour confinement [53, 54]. Nevertheless, perturbative QCD treats them as the relevant short-distance degrees of freedom: factorization theorems and asymptotic freedom justify computing hard-scattering matrix elements for incoming and outgoing partons even though QCD becomes non-perturbative at low scales [45]. The strong coupling  $\alpha_s$  grows large and effectively “blows up” around the confinement scale  $\Lambda_{QCD}$  [3]; consequently, something must happen to quarks and gluons before they reach the detector [23]. In practice, the gluon and all quarks except the top hadronize, producing cascades of baryons and mesons that themselves undergo further decays; hadronization is modelled e.g. with the Lund string or cluster models [23, 53, 54]. At the LHC, these hadrons typically carry energies comparable to the electroweak scale, and relativistic boosts tend to collimate their decay products into

narrow bunches [55]. Those collimated collections of hadrons are the jets we measure at hadron colliders and the objects we use to infer the partons produced in the hard interaction [39, 55].

Each high-energy parton produced in a collision, such as a quark from the process  $gg \rightarrow q\bar{q}$ , undergoes hadronization over a distance scale of  $\sim 10^{-15}$  m, producing a jet of hadrons [23, 53]. The energy composition of these jets is phenomenologically well established and is the basis of particle-flow reconstruction: on average roughly  $\sim 60\%$  of the jet energy is carried by charged particles (mostly  $\pi^\pm, K^\pm$ ),  $\sim 30\%$  by photons (from  $\pi^0 \rightarrow \gamma\gamma$ ) and  $\sim 10\%$  by neutral hadrons [41]. In high-energy jets the particles can be too collimated to be resolved individually in coarse calorimeter segmentation; nevertheless, the jet four-momentum is reconstructed from clustered PF candidates or calorimeter deposits and then corrected using jet energy corrections derived from simulation and in-situ data [17, 24, 41].

Phenomenologically one usually assumes that each high-energy parton yields a jet and that the measured jet four-momentum can, to useful accuracy, be related to the original parton four-momentum [56, 57]. Jets are therefore defined operationally using recombination (clustering) algorithms such as Cambridge–Aachen [58] or the (anti-)k<sub>T</sub> family [39]. Experimentally this means grouping a large number of energy depositions (or particle-flow candidates) observed in the calorimeters and tracker into a much smaller set of jets or sub-jets [41]. Nothing in the raw detector data, however, indicates a priori how many jets there should be: the clustering procedure and the choice of a resolution scale fix the outcome [55]. In practice one must either specify the desired number of final jets or choose a resolution/stop criterion (for example a distance parameter R, a clustering distance cut, or a jet-mass/sub-jet-resolution threshold) that determines the smallest substructure to be considered a separate parton-like object [59].

Modern reconstruction at the LHC typically uses particle-flow (PF) candidates as input together with infrared- and collinear-safe clustering algorithms to define jet four-momenta [24, 41]. The anti-k<sub>T</sub> algorithm [39], implemented in FastJet [24], is widely used in ATLAS and CMS; it groups candidates by proximity in the rapidity–azimuth ( $y, \phi$ ) plane with a typical distance parameter  $R \sim 0.4\text{--}0.6$  and is relatively insensitive to soft radiation and pileup when combined with area-based subtraction techniques [60]. After clustering, jet energy corrections (JEC) derived from simulation and in-situ calibrations compensate for detector response, pileup, and underlying-event effects [61], while jet-substructure and tagging algorithms (mass-drop, N-subjettiness, SoftDrop, etc.) help infer the flavour and origin of the initiating parton [59, 62, 63].

### 1.2.1 JET ALGORITHMS

Recombination (or sequential clustering) algorithms formalise the intuitive idea that parton showering produces collinear and soft splittings [55, 56, 58]: two nearby and kinematically compatible sub-jets are merged if they

are more likely to have originated from a single parton [56, 58]. A practical implementation requires a measure of “distance” between objects [24, 56]; common choices combine an angular separation in the rapidity–azimuth plane,  $\Delta R_{ij}$ , with a transverse-momentum weighting [39, 56]. Typical distance measures are [39, 56, 58]

$$\begin{aligned} k_T : \quad y_{ij} &= \frac{\Delta R_{ij}}{R} \min(p_{T,i}, p_{T,j}), \quad y_{iB} = p_{T,i}, \\ C/A : \quad y_{ij} &= \frac{\Delta R_{ij}}{R}, \quad y_{iB} = 1, \\ \text{anti-}k_T : \quad y_{ij} &= \frac{\Delta R_{ij}}{R} \min(p_{T,i}^{-1}, p_{T,j}^{-1}), \quad y_{iB} = p_{T,i}^{-1}. \end{aligned} \quad (1.5)$$

The parameter  $R$  balances jet–jet and jet–beam criteria and sets the geometric size of jets; in LHC analyses, typical values are  $R \sim 0.4\text{--}0.7$  depending on the physics target [24].

Two operational modes are useful to distinguish. In an exclusive algorithm, one supplies a resolution scale  $y_{cut}$  and proceeds iteratively:

1. compute  $y^{\min} = \min_{i,j}\{y_{ij}, y_{iB}\}$ ;
2. if  $y^{\min} = y_{ij} < y_{cut}$  merge  $i$  and  $j$  and repeat;
3. if  $y^{\min} = y_{iB} < y_{cut}$  remove  $i$  as beam radiation and repeat;
4. stop when  $y^{\min} > y_{cut}$  and keep remaining sub-jets as jets.

An inclusive algorithm omits  $y_{cut}$  and instead declares a sub-jet a final-state jet when its jet–beam distance is the smallest quantity; iteration continues until no inputs remain. Inclusive algorithms therefore produce a variable number of jets, while exclusive algorithms deliver a scale-dependent fixed set.

A practical question is how to combine the kinematics of merged objects. The most common choice in modern experiments is the E-scheme: four-vectors are added, which preserves energy–momentum and yields a physical jet mass useful for substructure and boosted-object tagging. An alternative is to sum three-momenta and rescale the energy to enforce a massless jet; this can be appropriate when the analysis targets massless parton kinematics, but it discards potentially useful jet-mass information.

From a theoretical and experimental viewpoint, important properties are infrared and collinear safety: a jet algorithm should give stable results under the emission of soft particles or collinear splittings. The  $k_T$ , C/A and anti- $k_T$  families are constructed to satisfy these requirements. Their practical behavior differs:  $k_T$  naturally follows the physical shower history soft-first clustering, C/A is purely geometric useful for declustering and substructure studies, while anti- $k_T$  produces regular, cone-like jets that are robust and convenient experimentally.

Corrections for pileup and the underlying event are necessary at the LHC. These corrections depend on the jet area and are typically performed by estimating an event-wide transverse-momentum density and subtracting the corresponding contribution proportional to the jet area. Finally,

because inclusive algorithms can produce jets arbitrarily close to the beam, a minimum jet  $p_T$  threshold, commonly 20–100 GeV depending on the analysis, is imposed to ensure experimental observability and theoretical control.

### 1.2.2 $\tau$ TAGGING AT MULTIPURPOSE DETECTORS

The  $\tau$  lepton decays hadronically with a probability of  $\sim 65\%$ , producing a narrow “ $\tau$ -jet” that contains only a few charged and neutral hadrons [3, 64]. Hadronic decays are dominated by one- and three-prong topologies and often include neutral pions that promptly convert to photons, giving a sizable electromagnetic fraction in the calorimeters [3, 65]. When the  $\tau$  momentum is large compared to its mass the decay products are highly collimated [32, 64]: for  $p_T > 50$  GeV roughly 90% of the visible energy is contained within a cone of radius  $R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.2$  [64]. These properties motivate the use of small signal cones and narrow isolation annuli in reconstruction [32, 64].

Identification exploits three complementary classes of observables [32, 41, 64, 65]:

- Calorimetric isolation and shower-shape variables [64, 65]: hadronic  $\tau$  decays deposit localized energy in ECAL+HCAL [64]. Experiments use isolation sums and shape ratios to quantify peripheral activity [64, 65]. Example variables are

$$\Delta E_T^{12} = \frac{\sum_{0.1 < \Delta R < 0.2} E_{T,j}}{\sum_{\Delta R < 0.4} E_{T,i}}, \quad P_{ISOL} = \sum_{\Delta R < 0.40} E_T - \sum_{\Delta R < 0.13} E_T, \quad (1.6)$$

which suppress QCD jets that populate the isolation ring [64].

- Charged-track isolation and prong topology [41, 64]: the few, collimated charged tracks of a  $\tau$  allow powerful selections. A common procedure defines a matching cone of radius  $R_m$  around the calorimeter jet axis to select candidate tracks above a  $p_T^{\min}$  threshold; the leading track ( $tr_1$ ) defines a narrow signal cone  $R_S$  (1- or 3-prong hypotheses) and a larger isolation cone  $R_I$  is scanned for additional tracks [41, 64].
- Lifetime and vertexing observables [3, 44]: the finite  $\tau$  lifetime ( $c\tau \approx 87$   $\mu\text{m}$ ) produces displaced tracks and, for multi-prong decays, a reconstructible secondary vertex; impact-parameter significances and secondary-vertex properties are exploited to separate genuine  $\tau_h$  from prompt jets or leptons [3, 44].

Additional discriminants include the invariant mass of the visible decay products computed from tracks and calorimeter clusters, electromagnetic energy fractions (sensitive to  $\pi^0 \rightarrow \gamma\gamma$ ), and dedicated shower-strip grouping for nearby photons. For example, invariant-mass reconstruction

commonly uses a jet cone  $\Delta R_{\text{jet}} \lesssim 0.4$  while excluding calorimeter clusters matched to tracks by a minimum separation  $\Delta R_{\text{track}} \gtrsim 0.08$  to reduce double counting.

Reconstruction algorithms combine these inputs. CMS’s Hadron-Plus-Strips (HPS) and modern DeepTau methods explicitly build decay-mode hypotheses and use strip-clustering of photons plus multivariate or deep-learning discriminators to reject jets, electrons, and muons [32, 66]. ATLAS employs analogous calorimeter+track based MVAs and BDTs [67]. Typical working points trade efficiency versus background: medium points often give  $\tau_h$  efficiencies of order 50–70% with light-jet misidentification rates in the per-mille to percent range, depending on kinematics and pileup.

Practical implementations tune cone sizes, isolation thresholds, and MVA inputs to the kinematic region and analysis goals; the choice of working point is driven by the signal-to-background optimization for the search or measurement at hand.

### 1.2.3 B TAGGING AT MULTIPURPOSE DETECTORS

Jets originating from bottom quarks ( $b$ -jets) exhibit several distinctive properties that enable their identification. The relatively long lifetime of  $b$  hadrons (order 1.5 ps) produces displaced charged tracks and often reconstructible secondary vertices a few millimetres from the primary interaction point. The large  $b$ -hadron mass yields decay products with sizable transverse momentum relative to the jet axis, and semileptonic branching fractions produce soft electrons or muons inside the jet. These features form the basis for  $b$ -tagging [68].

Practical algorithms exploit individual signatures or combine them:

- **Track-counting:** counts tracks with large impact-parameter significance to identify a  $b$ -like topology [68].
- **Jet-probability:** evaluates the compatibility of the jet’s track impact-parameter distribution with the primary vertex hypothesis [68].
- **Secondary-vertex:** explicitly reconstructs displaced vertices and uses their kinematic properties (decay length significance, vertex mass) [68].
- **Soft-lepton taggers:** identify low- $p_T$  leptons inside jets from semileptonic  $b$  decays [68].

Modern taggers combine many observables in multivariate or deep-learning classifiers to maximize discrimination power. Contemporary approaches exploit rich, low-level inputs (track-by-track and PF-candidate information, vertex features and kinematics) and advanced network architectures (DeepCSV/DeepJet, RNN/sequence, graph/set networks) [31, 67, 68]. These developments yield measurable performance gains: modern deep classifiers typically improve  $b$  efficiency at fixed mistag rate relative to classical taggers, and allow continuous discriminants with tunable operating

points. Calibration with data-driven scale factors (from  $t\bar{t}$ , multijet or dilepton control samples) and propagation of associated systematic uncertainties remain essential for physics results [68].

- Deep feed-forward networks (e.g. DeepCSV/DeepJet) ingest a large set of high-level and per-track inputs to produce powerful binary or multi-class discriminants that separate b, c and light-flavour jets.
- Sequence models and recurrent networks (RNN-based taggers) process an arbitrary ordered list of track-level variables, improving sensitivity by directly exploiting per-track correlations and order-dependent information (impact-parameter sequences, track kinematics).
- Graph- and set-based architectures and combined particle+vertex networks (sometimes referred to as “DeepFlavour”-style models) aggregate heterogeneous inputs and return per-flavour probabilities, enabling natural multi-classification and calibrated operating points.

These developments yield measurable performance gains: modern deep classifiers typically improve b efficiency at fixed mistag rate (or reduce mistag rates at fixed efficiency) relative to classical taggers. The continuous output of such networks permits analyses to choose operating points (loose/medium/tight) corresponding to desired efficiencies or mistag targets. Calibration remains essential: data-driven scale factors derived from control samples (e.g.  $t\bar{t}$ , multijet, dilepton) are applied to correct simulation, and systematic uncertainties from the calibration, flavour composition, and kinematic extrapolation are propagated to physics results.

Examples in use are CMS DeepCSV / DeepJet and ATLAS MV2 / DL1 [32, 67], which illustrate the transition from expert-designed high-level variables to large-scale machine learning leveraging low-level detector information. Typical medium working points yield b-tag efficiencies of order 60–80% with light-jet misidentification rates at or below the percent level; the precise choice of working point is tuned per analysis to optimise sensitivity while accounting for calibration and systematic uncertainties.

### 1.3 THE CMS DETECTOR

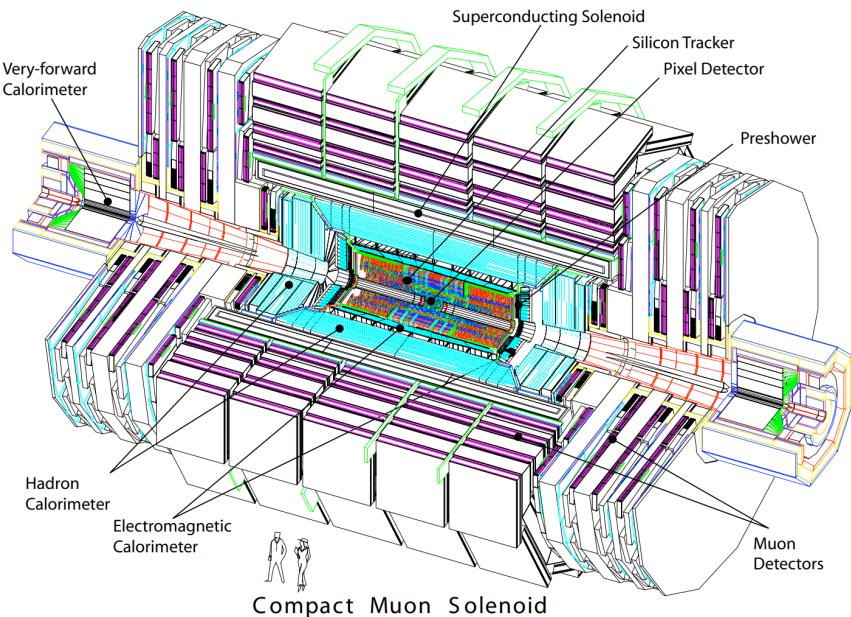
---

CMS is a general-purpose detector at the LHC [69]. With a length of 21.6 m, a diameter of 14.6 m, and a weight of 14,000 tonnes, its cylindrical geometry is divided into a central barrel section and two endcaps. This design provides hermetic coverage to accurately measure momentum and energy balance, which is crucial for identifying non-interacting particles like neutrinos through missing transverse energy.

The detector is constructed from concentric layers of sub-detectors, as illustrated in Figure 1.5. The innermost component is the silicon tracker, comprising a pixel detector and silicon strip tracker. It reconstructs the

trajectories of charged particles and measures their transverse momenta ( $p_T$ ) with a resolution of  $\approx 0.7\%$  for 10 GeV particles within a pseudorapidity range of  $|\eta| < 2.5$ .

Surrounding the tracker is the calorimetric system. The electromagnetic calorimeter (ECAL) is made of lead-tungstate crystals. It is designed to measure electrons and photons with a high resolution of  $\approx 0.6\%$  for 50 GeV electrons. The hadronic calorimeter (HCAL), located outside the ECAL, is a brass-scintillator sampling calorimeter that measures hadrons (e.g., charged pions, kaons, protons) with an energy resolution of  $\approx 18\%$  for 50 GeV pions. Together, the ECAL and HCAL cover  $|\eta| < 3$ . The coverage is extended to  $|\eta| < 5$  with steel and quartz-fiber hadron calorimeters in the forward regions.



**Figure 1.5:** Layout of the CMS experiment at the CERN LHC. (retrieved from [69]).

A key feature of CMS is its large superconducting solenoid, which encloses the tracker and calorimeters. The solenoid is constructed from a niobium-titanium alloy and cooled to 4.2 K with liquid helium. It generates a uniform magnetic field of 3.8 T throughout the tracking volume, enabling precise momentum measurement from the curvature of charged particle tracks.

The outermost system is dedicated to muon identification and measurement. Gas-ionization detectors are embedded in the steel flux-return yoke that surrounds the solenoid. This system provides triggering and tracking capabilities for muons up to  $|\eta| < 2.4$ . The combination of the inner tracker and the muon system allows for a robust identification and momentum measurement of muons across a wide kinematic range.

The geometrical segmentation of the barrel and endcaps defines the detector's acceptance in terms of pseudorapidity. The central barrel provides optimal coverage for  $|\eta| \lesssim 1.5$ , while the endcaps extend the acceptance

to  $|\eta| \lesssim 2.5$  for the tracker and calorimeters, and to  $|\eta| \lesssim 2.4$  for the muon system.

This segmentation impacts the detection efficiency. The silicon trackers are highly efficient in the barrel, where particles cross the layers perpendicularly. In the endcaps, the reduced hit multiplicity from shallow-angle traversals leads to a slight decrease in tracking efficiency and resolution. The calorimeters are also optimized to maintain performance across  $\eta$ , though the material budget and granularity vary.

Muon reconstruction performance exhibits regional differences. In the barrel, drift tubes (DTs) provide high spatial resolution, while in the endcaps, cathode strip chambers (CSCs) and resistive plate chambers (RPCs) are used to handle higher background rates and non-uniform magnetic fields. The assumed identification efficiency for muons (electrons) is 95% (85%), with a mis-identification rate of 0.3% (0.6%) [36, 42, 43].

For the identification of heavy-flavor jets, we adopt the DeepCSV algorithm [68]. We use its “medium” working point, which provides a b-tagging efficiency of 70% with a light-flavor jet misidentification rate of approximately 1% across the entire  $p_T$  spectrum. The “loose” (85% efficiency, 10% mis-id) and “tight” (45% efficiency, 0.1% mis-id) working points were also explored during the analysis optimization.

For hadronically decaying  $\tau$  leptons ( $\tau_h$ ), we use the DeepTau algorithm [32], which employs a deep neural network combining isolation and lifetime information to identify  $\tau_h$  decay modes. The “medium” working point is chosen for this analysis, providing a  $\tau_h$  identification efficiency of 70% and a misidentification rate of 0.5% for jets originating from light quarks and gluons. This working point was selected through an optimization process that maximized the discovery reach of the analysis.



---

## BIBLIOGRAPHY

---

- [1] A. Flórez et al. “On the sensitivity reach of LQ production with preferential couplings to third generation fermions at the LHC”. In: *The European Physical Journal C* 83.11 (Nov. 2023), p. 1023. ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-023-12177-4](https://doi.org/10.1140/epjc/s10052-023-12177-4).
- [2] Umar Sohail Qureshi et al. “Probing light scalars and vector-like quarks at the high-luminosity LHC”. In: *Eur. Phys. J. C* 85.4 (2025), p. 379. DOI: [10.1140/epjc/s10052-025-14085-1](https://doi.org/10.1140/epjc/s10052-025-14085-1). arXiv: [2410.17854 \[hep-ph\]](https://arxiv.org/abs/2410.17854).
- [3] P.A. Zyla et al. “Review of Particle Physics”. In: *PTEP* 2020.8 (2020), p. 083C01. DOI: [10.1093/ptep/ptaa104](https://doi.org/10.1093/ptep/ptaa104).
- [4] Steven Weinberg. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19 (21 1967), pp. 1264–1266. DOI: [10.1103/PhysRevLett.19.1264](https://doi.org/10.1103/PhysRevLett.19.1264).
- [5] H.P. P Nilles. “Supersymmetry, supergravity and particle physics”. In: *Physics Reports* 110.1-2 (1984), pp. 1–162. ISSN: 03701573. DOI: [10.1016/0370-1573\(84\)90008-5](https://doi.org/10.1016/0370-1573(84)90008-5).
- [6] I. Doršner et al. “Physics of leptoquarks in precision experiments and at particle colliders”. In: *Phys. Rept.* 641 (2016), pp. 1–68. DOI: [10.1016/j.physrep.2016.06.001](https://doi.org/10.1016/j.physrep.2016.06.001). arXiv: [1603.04993 \[hep-ph\]](https://arxiv.org/abs/1603.04993).
- [7] Lisa Randall and Raman Sundrum. “A Large mass hierarchy from a small extra dimension”. In: *Phys. Rev. Lett.* 83 (1999), pp. 3370–3373. DOI: [10.1103/PhysRevLett.83.3370](https://doi.org/10.1103/PhysRevLett.83.3370). arXiv: [hep-ph/9905221](https://arxiv.org/abs/hep-ph/9905221).
- [8] G. F. Giudice. “Naturalness after LHC”. In: *arXiv e-prints* (2013). arXiv: [1307.7879 \[hep-ph\]](https://arxiv.org/abs/1307.7879).
- [9] Gianfranco Bertone, Dan Hooper, and Joseph Silk. “Particle dark matter: evidence, candidates and constraints”. In: *Physics Reports* 405 (2005), pp. 279–390. DOI: [10.1016/j.physrep.2004.08.031](https://doi.org/10.1016/j.physrep.2004.08.031). arXiv: [hep-ph/0404175 \[hep-ph\]](https://arxiv.org/abs/hep-ph/0404175).
- [10] P. K. Mohapatra, R. N. Mohapatra, and P. B. Pal. “Implications of E6 Grand Unification”. In: *Phys. Rev. D* 33 (1986), p. 2010. DOI: [10.1103/PhysRevD.33.2010](https://doi.org/10.1103/PhysRevD.33.2010).
- [11] LHCb Collaboration. “Test of lepton universality using  $B^+ \rightarrow K^+ \ell^+ \ell^-$  decays”. In: *Phys. Rev. Lett.* 113 (2014), p. 151601. DOI: [10.1103/PhysRevLett.113.151601](https://doi.org/10.1103/PhysRevLett.113.151601). arXiv: [1406.6482 \[hep-ex\]](https://arxiv.org/abs/1406.6482).
- [12] LHCb Collaboration. “Test of lepton universality with  $B^0 \rightarrow K^{*0} \ell^+ \ell^-$  decays”. In: *JHEP* 08 (2017), p. 055. DOI: [10.1007/JHEP08\(2017\)055](https://doi.org/10.1007/JHEP08(2017)055). arXiv: [1705.05802 \[hep-ex\]](https://arxiv.org/abs/1705.05802).

- [13] BaBar Collaboration. “Evidence for an excess of  $\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$  decays”. In: *Phys. Rev. Lett.* 109 (2012), p. 101802. doi: 10.1103/PhysRevLett.109.101802. arXiv: 1205.5442 [hep-ex].
- [14] D. Buttazzo et al. “B-physics anomalies: a guide to combined explanations”. In: *JHEP* 11 (2017), p. 044. doi: 10.1007/JHEP11(2017)044. arXiv: 1706.07808 [hep-ph].
- [15] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *JHEP* 07 (2014), p. 079. doi: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].
- [16] A. Alloul et al. “FeynRules 2.0 - A complete toolbox for tree-level phenomenology”. In: *Comput. Phys. Commun.* 185 (2014), pp. 2250–2300. doi: 10.1016/j.cpc.2014.04.012. arXiv: 1310.1921 [hep-ph].
- [17] DELPHES 3 Collaboration. “DELPHES 3, A modular framework for fast simulation of a generic collider experiment”. In: *JHEP* 02 (2014), p. 057. doi: 10.1007/JHEP02(2014)057. arXiv: 1307.6346 [hep-ex].
- [18] Georges Aad et al. “Search for high-mass dilepton resonances using  $139 \text{ fb}^{-1}$  of  $pp$  collision data collected at  $\sqrt{s} = 13 \text{ TeV}$  with the ATLAS detector”. In: *Phys. Lett. B* 796 (2019), pp. 68–87. doi: 10.1016/j.physletb.2019.07.016. arXiv: 1903.06248 [hep-ex].
- [19] Albert M Sirunyan et al. “Search for resonant and nonresonant new phenomena in high-mass dilepton final states at  $\sqrt{s} = 13 \text{ TeV}$ ”. In: *JHEP* 07 (2021), p. 208. doi: 10.1007/JHEP07(2021)208. arXiv: 2103.02708 [hep-ex].
- [20] S. Schael et al. “Measurement of the  $W$  boson mass and width in  $e+e-$  collisions at LEP”. In: *The European Physical Journal C* 47.2 (2006), pp. 309–335. doi: 10.1140/epjc/s2006-02576-8.
- [21] Wolfgang Altmannshofer and David M. Straub. “New physics in  $b \rightarrow s$  transitions after LHC run 1”. In: *The European Physical Journal C* 75.8 (2015). doi: 10.1140/epjc/s10052-015-3602-7.
- [22] M. Ciuchini et al. “Constraints on lepton universality violation from rare  $B$  decays”. In: *Phys. Rev. D* 107 (5 2023), p. 055036. doi: 10.1103/PhysRevD.107.055036.
- [23] T. Sjöstrand et al. “An introduction to PYTHIA 8.2”. In: *Comput. Phys. Commun.* 191 (2015), pp. 159–177. doi: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].
- [24] M. Cacciari, G. P. Salam, and G. Soyez. “FastJet User Manual”. In: *Eur. Phys. J. C* 72 (2012), p. 1896. doi: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].
- [25] NNPDF Collaboration. “Parton distributions for the LHC Run II”. In: *JHEP* 04 (2015), p. 040. doi: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].

- [26] G. Aad et al. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08003. doi: [10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003).
- [27] S. Chatrchyan et al. "The CMS Experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. doi: [10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- [28] Paul Lujan et al. *The Pixel Luminosity Telescope: A detector for luminosity measurement at CMS using silicon pixel sensors*. Tech. rep. CMS-DN-21-008. Accepted by Eur. Phys. J. C. 2022. arXiv: 2206.08870 [physics.ins-det]. URL: <https://arxiv.org/abs/2206.08870>.
- [29] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *J. Mach. Learn. Res.* 12 (2011), pp. 2825–2830. ISSN: 1532-4435. doi: [10.48550/arXiv.1201.0490](https://doi.org/10.48550/arXiv.1201.0490).
- [30] T. Chen and C. Guestrin. "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. San Francisco, California, USA: ACM, 2016, pp. 785–794. ISBN: 978-1-4503-4232-2. doi: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785).
- [31] E. Bols et al. "Jet flavour classification using DeepJet". In: *Journal of Instrumentation* 15.12 (2020), P12012. doi: [10.1088/1748-0221/15/12/P12012](https://doi.org/10.1088/1748-0221/15/12/P12012).
- [32] CMS Collaboration. "Identification of hadronic tau lepton decays using a deep neural network". In: *JINST* 17.07 (2022), Po7023. doi: [10.1088/1748-0221/17/07/P07023](https://doi.org/10.1088/1748-0221/17/07/P07023). arXiv: 2201.08458 [hep-ex].
- [33] Lyndon Evans and Philip Bryant. "LHC Machine". In: *Journal of Instrumentation* 3.08 (2008), S08001–S08001. doi: [10.1088/1748-0221/3/08/s08001](https://doi.org/10.1088/1748-0221/3/08/s08001).
- [34] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716 (2012). ATLAS Higgs-discovery paper; verify metadata if needed, pp. 1–29. doi: [10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020).
- [35] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716 (2012). CMS Higgs-discovery paper; verify metadata if needed, pp. 30–61. doi: [10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021).
- [36] CMS Collaboration. *Study of the Discovery Reach in Searches for Supersymmetry at CMS with 3000/fb*. CMS Physics Analysis Summary CMS-PAS-FTR-13-014. Accessed on January 07, 2023. CERN, 2013. URL: <https://cds.cern.ch/record/1607141>.
- [37] G. Apollinari et al. *High-Luminosity Large Hadron Collider (HL-LHC): Technical and performance projections*. Tech. rep. CERN Yellow Report / TBD. HL-LHC technical design / projection report; please verify exact report number/title and replace placeholder fields as desired. CERN, 2017.

- [38] D. Bertolini et al. "Pileup Per Particle Identification". In: *JHEP* 10 (2014). Introduces the PUPPI algorithm for per-particle pileup mitigation; ampliamente usado en ATLAS/CMS analyses., p. 59. arXiv: 1407.6013 [hep-ph].
- [39] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-kt jet clustering algorithm". In: *Journal of High Energy Physics* 2008.04 (2008), p. 063. doi: 10.1088/1126-6708/2008/04/063.
- [40] Izaak Neutelings. CMS coordinate system. URL: [https://tikz.net/axis3d\\_cms](https://tikz.net/axis3d_cms).
- [41] CMS Collaboration. "Particle-flow reconstruction and global event description with the CMS detector". In: *JINST* 12 (2017). Resumen del algoritmo de reconstrucción particle-flow utilizado en CMS; útil para la discusión de PF en el texto., P10003.
- [42] A.M. Sirunyan et al. "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC". In: *JINST* 16.05 (2021), P05014. doi: 10.1088/1748-0221/16/05/P05014.
- [43] A.M. Sirunyan et al. "Performance of the reconstruction and identification of high-momentum muons in proton-proton collisions at  $\sqrt{s} = 13$  TeV". In: *JINST* 15.02 (2020), Po2027. doi: 10.1088/1748-0221/15/02/P02027.
- [44] The CMS Collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *Journal of Instrumentation* 9.10 (Oct. 2014), P10009–P10009. ISSN: 1748-0221. doi: 10.1088/1748-0221/9/10/p10009.
- [45] John C. Collins, Davison E. Soper, and George F. Sterman. "Factorization of Hard Processes in QCD". In: *Adv. Ser. Direct. High Energy Phys.* 5 (1989), pp. 1–91. doi: 10.1142/9789814503266\_0001. arXiv: hep-ph/0409313.
- [46] J. Butterworth, S. Carrazza, A. Cooper-Sarkar, A. De Roeck, J. Feltesse, S. Forte, J. Gao, S. Glazov, J. Huston, Z. Kassabov, R. McNulty, A. Morsch, P. Nadolsky, V. Radescu, J. Rojo and R. Thorne. "PDF4LHC recommendations for LHC Run II". In: *J. Phys. G* 43 (2016), p. 023001. doi: 10.1088/0954-3899/43/2/023001. arXiv: 1510.03865 [hep-ph].
- [47] Albert M Sirunyan et al. "Performance of missing transverse momentum reconstruction in proton-proton collisions at  $\sqrt{s} = 13$  TeV using the CMS detector". In: *JINST* 14.07 (2019), Po7004. doi: 10.1088/1748-0221/14/07/P07004. arXiv: 1903.06078 [hep-ex].
- [48] William R. Leo. *Techniques for Nuclear and Particle Physics Experiments*. Springer, 1994. ISBN: 978-3-540-57280-0, 978-3-642-57920-2. doi: 10.1007/978-3-642-57920-2.
- [49] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur. Phys. J. C* 71 (2011). [Erratum: Eur.Phys.J.C 73, 2501 (2013)], p. 1554. doi: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].

- [50] Alexander L. Read. "Presentation of search results: The  $CL_s$  technique". In: *J. Phys. G* 28 (2002). Ed. by M. R. Whalley and L. Lyons, pp. 2693–2704. doi: [10.1088/0954-3899/28/10/313](https://doi.org/10.1088/0954-3899/28/10/313).
- [51] Wolfgang A. Rolke, Angel M. López, and Jan Conrad. "Limits and confidence intervals in the presence of nuisance parameters". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 551.2–3 (Oct. 2005), 493–503. ISSN: 0168-9002. doi: [10.1016/j.nima.2005.05.068](https://doi.org/10.1016/j.nima.2005.05.068).
- [52] Gary J. Feldman and Robert D. Cousins. "Unified approach to the classical statistical analysis of small signals". In: *Physical Review D* 57.7 (Apr. 1998), 3873–3889. ISSN: 1089-4918. doi: [10.1103/physrevd.57.3873](https://doi.org/10.1103/physrevd.57.3873).
- [53] Bo Andersson et al. "Parton Fragmentation and String Dynamics". In: *Phys. Rept.* 97 (1983), pp. 31–145. doi: [10.1016/0370-1573\(83\)90080-7](https://doi.org/10.1016/0370-1573(83)90080-7).
- [54] B. R. Webber. "A QCD Model for Jet Fragmentation Including Soft Gluon Interference". In: *Nucl. Phys. B* 238 (1984), pp. 492–528. doi: [10.1016/0550-3213\(84\)90333-X](https://doi.org/10.1016/0550-3213(84)90333-X).
- [55] Gavin P. Salam. "Towards Jetography". In: *Eur. Phys. J. C* 67 (2010), pp. 637–686. doi: [10.1140/epjc/s10052-010-1314-6](https://doi.org/10.1140/epjc/s10052-010-1314-6). arXiv: [0906.1833 \[hep-ph\]](https://arxiv.org/abs/0906.1833).
- [56] S. Catani et al. "Longitudinally invariant  $K_t$  clustering algorithms for hadron hadron collisions". In: *Nucl. Phys. B* 406 (1993), pp. 187–224. doi: [10.1016/0550-3213\(93\)90166-M](https://doi.org/10.1016/0550-3213(93)90166-M).
- [57] Stephen D. Ellis and Davison E. Soper. "Successive combination jet algorithm for hadron collisions". In: *Phys. Rev. D* 48 (1993), pp. 3160–3166. doi: [10.1103/PhysRevD.48.3160](https://doi.org/10.1103/PhysRevD.48.3160). arXiv: [hep-ph/9305266](https://arxiv.org/abs/hep-ph/9305266).
- [58] Yuri L. Dokshitzer et al. "Better jet clustering algorithms". In: *JHEP* 08 (1997), p. 001. doi: [10.1088/1126-6708/1997/08/001](https://doi.org/10.1088/1126-6708/1997/08/001). arXiv: [hep-ph/9707323](https://arxiv.org/abs/hep-ph/9707323).
- [59] Jesse Thaler and Ken Van Tilburg. "Identifying Boosted Objects with N-subjettiness". In: *JHEP* 03 (2011), p. 015. doi: [10.1007/JHEP03\(2011\)015](https://doi.org/10.1007/JHEP03(2011)015). arXiv: [1011.2268 \[hep-ph\]](https://arxiv.org/abs/1011.2268).
- [60] M. Cacciari, G. P. Salam, and G. Soyez. "The catchment area of jets". In: *JHEP* 04 (2008). Detalla conceptos de área de jets y técnicas relacionadas con la resta de pile-up., p. 005. doi: [10.1088/1126-6708/2008/04/005](https://doi.org/10.1088/1126-6708/2008/04/005).
- [61] Garvita Agarwal. "Jet Energy Scale and Resolution Measurements in CMS". In: *PoS ICHEP2022* (2022), p. 652. doi: [10.22323/1.414.0652](https://doi.org/10.22323/1.414.0652). arXiv: [2301.02175 \[hep-ex\]](https://arxiv.org/abs/2301.02175).
- [62] Jonathan M. Butterworth et al. "Jet substructure as a new Higgs search channel at the LHC". In: *Phys. Rev. Lett.* 100 (2008), p. 242001. doi: [10.1103/PhysRevLett.100.242001](https://doi.org/10.1103/PhysRevLett.100.242001). arXiv: [0802.2470 \[hep-ph\]](https://arxiv.org/abs/0802.2470).

- [63] Andrew J. Larkoski et al. “Soft Drop”. In: *JHEP* 05 (2014), p. 146. doi: [10.1007/JHEP05\(2014\)146](https://doi.org/10.1007/JHEP05(2014)146). arXiv: [1402.2657](https://arxiv.org/abs/1402.2657) [hep-ph].
- [64] A. M. Sirunyan et al. “Performance of reconstruction and identification of  $\tau$  leptons decaying to hadrons and  $\nu_\tau$  in pp collisions at  $\sqrt{s} = 13$  TeV”. In: *JINST* 13.10 (2018), P10005. doi: [10.1088/1748-0221/13/10/P10005](https://doi.org/10.1088/1748-0221/13/10/P10005). arXiv: [1809.02816](https://arxiv.org/abs/1809.02816) [hep-ex].
- [65] Georges Aad et al. “Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at  $\sqrt{s}=8$  TeV”. In: *Eur. Phys. J. C* 75.7 (2015), p. 303. doi: [10.1140/epjc/s10052-015-3500-z](https://doi.org/10.1140/epjc/s10052-015-3500-z). arXiv: [1412.7086](https://arxiv.org/abs/1412.7086) [hep-ex].
- [66] CMS Collaboration. “Performance of the CMS Level-1 trigger in proton-proton collisions at  $\sqrt{s} = 13$  TeV”. In: *J. Inst.* 17.05 (2022), Po5024. doi: [10.31526/lhep.2022.195](https://doi.org/10.31526/lhep.2022.195). arXiv: [2202.08678](https://arxiv.org/abs/2202.08678) [hep-ex].
- [67] ATLAS Collaboration. “Measurement of the energy response of the ATLAS calorimeter to charged pions from  $W^\pm \rightarrow \tau^\pm (\rightarrow \pi^\pm \nu_\tau) \nu_\tau$  events in Run 2 data”. In: *Eur. Phys. J. C* 82.3 (2022), p. 223. doi: [10.1140/epjc/s10052-022-10933-6](https://doi.org/10.1140/epjc/s10052-022-10933-6). arXiv: [2202.13948](https://arxiv.org/abs/2202.13948) [hep-ex].
- [68] CMS Collaboration. “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”. In: *JINST* 13.04 (2018), P05011. doi: [10.1088/1748-0221/13/05/P05011](https://doi.org/10.1088/1748-0221/13/05/P05011). arXiv: [1712.07158](https://arxiv.org/abs/1712.07158) [hep-ex].
- [69] The CMS Collaboration. “The CMS experiment at the CERN LHC”. In: *Journal of Instrumentation* 3.08 (2008), S08004–S08004. doi: [10.1088/1748-0221/3/08/s08004](https://doi.org/10.1088/1748-0221/3/08/s08004).

---

## PUBLICATIONS

---

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis

lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.



# On the sensitivity reach of LQ production with preferential couplings to third generation fermions at the LHC

A. Flórez<sup>1,a</sup>, J. Jones-Pérez<sup>2,b</sup>, A. Gurrola<sup>3,c</sup>, C. Rodriguez<sup>1,d</sup>, J. Peñuela-Parra<sup>1,e</sup>

<sup>1</sup> Departamento de Física, Universidad de Los Andes, Cra. 1 # 18a-12, Bogotá, Colombia

<sup>2</sup> Sección Física, Departamento de Ciencias, Pontificia Universidad Católica del Perú, Apartado 1761, Lima, Peru

<sup>3</sup> Department of Physics and Astronomy, Vanderbilt University, Nashville, TN 37235, USA

Received: 31 July 2023 / Accepted: 22 October 2023 / Published online: 11 November 2023  
© The Author(s) 2023

**Abstract** Leptoquarks (LQs) are hypothetical particles that appear in various extensions of the Standard Model (SM), that can explain observed differences between SM theory predictions and experimental results. The production of these particles has been widely studied at various experiments, most recently at the Large Hadron Collider (LHC), and stringent bounds have been placed on their masses and couplings, assuming the simplest beyond-SM (BSM) hypotheses. However, the limits are significantly weaker for LQ models with family non-universal couplings containing enhanced couplings to third-generation fermions. We present a new study on the production of a LQ at the LHC, with preferential couplings to third-generation fermions, considering proton-proton collisions at  $\sqrt{s} = 13$  TeV and  $\sqrt{s} = 13.6$  TeV. Such a hypothesis is well motivated theoretically and it can explain the recent anomalies in the precision measurements of B-meson decay rates, specifically the  $R_{D^{(*)}}$  ratios. Under a simplified model where the LQ masses and couplings are free parameters, we focus on cases where the LQ decays to a  $\tau$  lepton and a b quark, and study how the results are affected by different assumptions about chiral currents and interference effects with other BSM processes with the same final states, such as diagrams with a heavy vector boson,  $Z'$ . The analysis is performed using machine learning techniques, resulting in an increased discovery reach at the LHC, allowing us to probe new physics phase space which addresses the B-meson anomalies, for LQ masses up to 5.00 TeV, for the high luminosity LHC scenario.

## 1 Introduction

After more than ten years collecting data, the LHC has confirmed that the Standard Model (SM) is indeed the correct theory describing particle physics for energies below the TeV scale. Nevertheless, there exist reasons to expect the SM to be a low-energy effective realization of a more complete theory. On the theoretical side, we do not know if gravity should be quantized, or if the gauge interactions should be unified, and if so, we do not know how to solve the associated hierarchy problems on the Higgs mass. Moreover, we have no explanation for fermion family replication, nor for the lack of CP violation in the strong sector. This expectation for physics beyond the SM (BSM) is reinforced experimentally, where the observation of neutrino masses, dark matter, and the baryon asymmetry in the Universe, cannot be explained by the SM.

Leptoquarks (LQs) are hypothetical bosons carrying both baryon and lepton number, thus interacting jointly with a lepton and a quark. They are a common ingredient in SM extensions where quarks and leptons share the same multiplet. Typical examples of these can be found in the Pati-Salam [1] and  $SU(5)$  GUT [2] models. In addition, they can also be found in theories with strong interactions, such as compositeness [3]. Due to their exotic coupling which allows quark-lepton transitions, they have a diverse phenomenology, which naturally leads to several constraints. An important one comes from proton decay, which forces the LQ mass to values close to the Planck scale, unless baryon and lepton numbers are not violated. Furthermore, in models where the latter are conserved, the LQ can still be subject to a wide variety of bounds [4–9]. Examples of these come from meson mixing, electric and magnetic dipole moments, atomic parity violation tests, rare decays, and direct searches. Nevertheless, the significance of each bound is a model dependent question.

<sup>a</sup> e-mail: ca.florez@uniandes.edu.co (corresponding author)

<sup>b</sup> e-mail: jones.j@pucp.edu.pe

<sup>c</sup> e-mail: alfredo.gurrola@vanderbilt.edu

<sup>d</sup> e-mail: c.rodriguez45@uniandes.edu.co

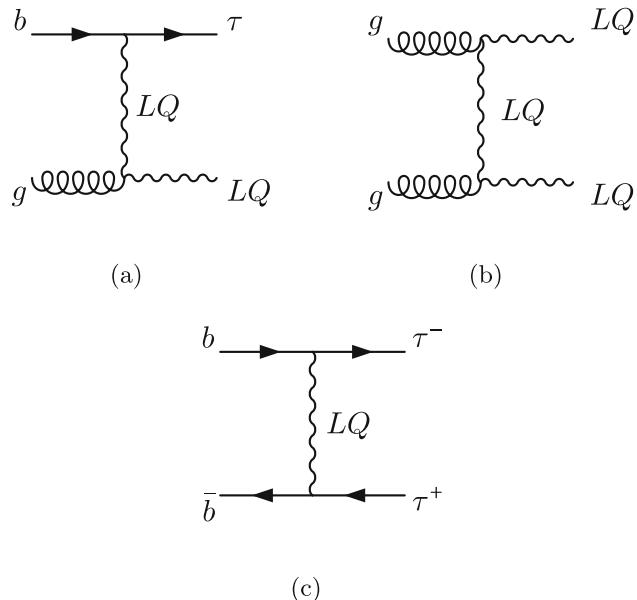
<sup>e</sup> e-mail: j.penuela@uniandes.edu.co

In the last years, an increased interest in low scale LQs has emerged due to the anomalies in the precision measurements of the B-meson decay rates. As it is well known, these corresponded mainly to deviations in the  $R_{K^{(*)}}$  [10–13] and  $R_{D^{(*)}}$  [14–25] ratios, which measure the violation of lepton flavour universality (LFU). What followed was a very intense theoretical development, aiming to explain the anomalies by TeV scale LQ exchange at tree level [26–41]. Before the end of 2022, it was generally agreed that, within proposed single LQ solutions, the only candidate capable of addressing all B-meson anomalies simultaneously and surviving all other constraints was a vector LQ ( $U_1$ ), transforming as  $(\mathbf{3}, \mathbf{1}, 2/3)$ , and coupling mainly to third-generation fermions via  $b \tau$  and  $t \nu_\tau$  vertices [36, 39]. In spite of a recent re-analysis of  $R_{K^{(*)}}$  data showing this ratio to be compatible with the SM prediction [42–45], the solution to the  $R_{D^{(*)}}$  anomaly is still an open question and remains a valid motivation for the study of scenarios where new particles have preferential couplings to third-generation fermions. Thus, it is still of interest to continue exploring the possibility of observing the  $U_1$  LQ at the LHC [41].

As expected, the theoretical community has extensively participated in probing LQ models by scrutinizing search strategies, recasting LHC results, and predicting the reach in the parameter space via different searches involving third-generation fermions (see for instance [46–55]). In addition, several 13 TeV searches for LQs decaying into  $t/b$  and  $\tau/\nu$  final states have been performed by the CMS [56–64] and ATLAS [65–71] collaborations.

Of the searches above, we find [62] particularly interesting. Here, the CMS collaboration explores signals corresponding to  $t \nu b \tau$  and  $t \nu \tau$  final states, with  $137 \text{ fb}^{-1}$  of proton-proton (pp) collision data. The former is motivated by LQ pair production, with one LQ decaying into  $t \nu$  and the other into  $b \tau$ , while the latter arises from a single LQ produced in association with a  $\tau$ , with a subsequent LQ decay into  $t \nu$  (see Fig. 1 for the corresponding diagrams). From the combination of both production channels, the search excludes  $U_1$  masses under  $1.3 - 1.7 \text{ TeV}$ , with this range depending on the  $U_1$  coupling to gluons and on its coupling  $g_U$  in the  $b_L \tau_L$  vertex.

What makes this search particularly attractive is that, for the first time, an LHC collaboration directly places (mass dependent) bounds on  $g_U$ . This is important, since having information on this parameter is crucial in order to understand if the  $U_1$  is really responsible for the  $R_{D^{(*)}}$  anomaly. The inclusion of the single-LQ production mode is important, since its cross-section is directly proportional to  $g_U^2$ . However, as can be seen in Fig. 6 of [62], the current constraints are dominated by pair production, with single-LQ production playing a subleading role. While this is expected [49], it still leads us to ponder the possibility of improving the sensitivity of LHC searches to single-LQ production, and thus



**Fig. 1** Representative Feynman diagrams of single (a), pair (b), and non-resonant (c) leptoquark production in proton-proton collision experiments. In single and pair production, the diagrams shown involve t-channel LQ exchange, dominant for lower LQ mass. However, for larger mass there exist s-channel diagrams featuring a virtual bottom quark and gluon, respectively

on achieving better constraints on  $g_U$ . Other complementary and similar searches to [62] were carried out by both ATLAS [70] and CMS [64].

It is also well known, though, that searches for an excess in the high- $p_T$  tails of  $\tau$  lepton distributions can strongly probe  $g_U$ , up to very large LQ masses. Indeed, as shown in [41, 72], the new physics effective operators contributing to  $R_{D^{(*)}}$  also contribute to an enhancement in the  $\text{p p} \rightarrow \tau\tau$  production rates. This has motivated a large number of recasts [39, 41, 49, 51, 73–77], as well as a CMS search explicitly providing constraints in terms of  $U_1$  [63]. Nevertheless, it is important to note that for these  $\text{p p} \rightarrow \tau\tau$  processes, the LQ participates non-resonantly, so contributions to the  $\text{p p} \rightarrow \tau\tau$  rates and kinematic distributions from non-LQ BSM diagrams containing possible virtual particles, such as a heavy neutral vector boson  $Z'$ , could spoil a straightforward interpretation of any possible excess [51]. Thus, it is also necessary to understand how the presence of other virtual particles can affect the sensitivity of an analysis probing  $g_U$ .

In this work we study the projected LQ sensitivity at the LHC, considering already available pp data as well as the expected amount of data to be acquired during the High-Luminosity LHC (HL-LHC) runs. We explore a proposed analysis strategy which utilizes a combination of single-, double-, and non-resonant-LQ production, targeting final states with varying  $\tau$ -lepton and b-jet multiplicities. The

studies are performed considering various benchmark scenarios for different LQ masses and couplings, also taking into account distinct chiralities for the third-generation fermions in the LQ vertex. We also assess the impact of a companion  $Z'$ , which is typical of gauge models, in non-resonant LQ probes, and find that interference effects can have a significant effect on the discovery reach. We consider this effect to be of high interest, given that non-resonant LQ production can have the largest cross-section, and thus could be an important channel in terms of discovery potential.

An important aspect of this work is that the analysis strategy is developed using a machine learning (ML) algorithm based on Boosted Decision Trees (BDT)[78]. The output of the event classifier is used to perform a profile-binned likelihood test to extract the overall signal significance for each model considered in the analysis. The advantage of using BDTs and other ML algorithms has been demonstrated in several experimental and phenomenological studies [50, 79–84]. In our studies, we find that the BDT algorithm gives sizeable improvement in signal significance.

This paper is organized as follows. In Sect. 2 we present our simplified model and review the model parameters which are relevant for solving the B-meson anomalies. Section 3 describes the details associated with the analysis strategy and the simulation of signal and background samples. Section 4 contains the results of the study, including the projected sensitivity for different benchmark scenarios considered. Finally, in Sect. 5 we discuss the implication of our results and prospects for future studies.

## 2 A simplified model for the $U_1$ leptoquark

Extending the SM with a massive  $U_1$  vector LQ is not straightforward, as one has to ensure the renormalizability of the model. Most of the theoretical community has focused on extensions of the Pati-Salam (PS) models which avoid proton decay, such as the scenario found in [85]. Other examples include PS models with vector-like fermions [86–88], the so-called 4321 models [89–91], the twin PS<sup>2</sup> model [92, 93], the three-site PS<sup>3</sup> model [94–96], as well as composite PS models [97–99].

In what follows, we shall restrict ourselves to a simplified non-renormalizable lagrangian, understood to be embedded into a more complete model. The SM is thus extended by adding the following terms featuring the  $U_1$  LQ:

$$\begin{aligned} \mathcal{L}_{U_1} = & -\frac{1}{2} U_{\mu\nu}^\dagger U^{\mu\nu} + M_U^2 U_{1\mu}^\dagger U_1^\mu \\ & -ig_s U_{1\mu}^\dagger T^a U_{1\nu} G^{a\mu\nu} - i\frac{2}{3}g' U_{1\mu}^\dagger U_{1\nu} B^{\mu\nu} \\ & + \frac{g_U}{\sqrt{2}} [U_{1\mu} (\bar{Q}_3 \gamma^\mu L_3 + \beta_L^{st} \bar{Q}_2 \gamma^\mu L_3 \\ & + \beta_R \bar{b}_R \gamma^\mu \tau_R) + h.c.] \end{aligned} \quad (1)$$

where  $U_{\mu\nu} \equiv \mathcal{D}_\mu U_{1\nu} - \mathcal{D}_\nu U_{1\mu}$ , and  $\mathcal{D}_\mu \equiv \partial_\mu + ig_s T^a G_\mu^a + i\frac{2}{3}g' B_\mu$ . As evidenced by the second line above, we assume that the LQ has a gauge origin.<sup>1</sup>

The third and fourth lines in Eq. (1) shows the LQ interactions with SM fermions, with coupling  $g_U$ , which we have chosen as preferring the third generation.<sup>2</sup> These are particularly relevant for the LQ decay probabilities, as well as for the single-LQ production cross-section. The  $\beta_L^{st}$  parameter, which is the  $LQ \rightarrow s\tau$  coupling in the  $\beta_L$  matrix (see footnote), is chosen to be equal to 0.2, following the fit done in [75], in order to simultaneously solve the  $R_{D^{(*)}}$  anomaly and satisfy the  $p p \rightarrow \tau^+ \tau^-$  constraints. Although  $\beta_L^{st}$  technically alters the single-LQ production cross-section and LQ branching fractions, we have confirmed that a value of  $\beta_L^{st} = 0.2$  results in negligible impact on our collider results, and thus is ignored in our subsequent studies.

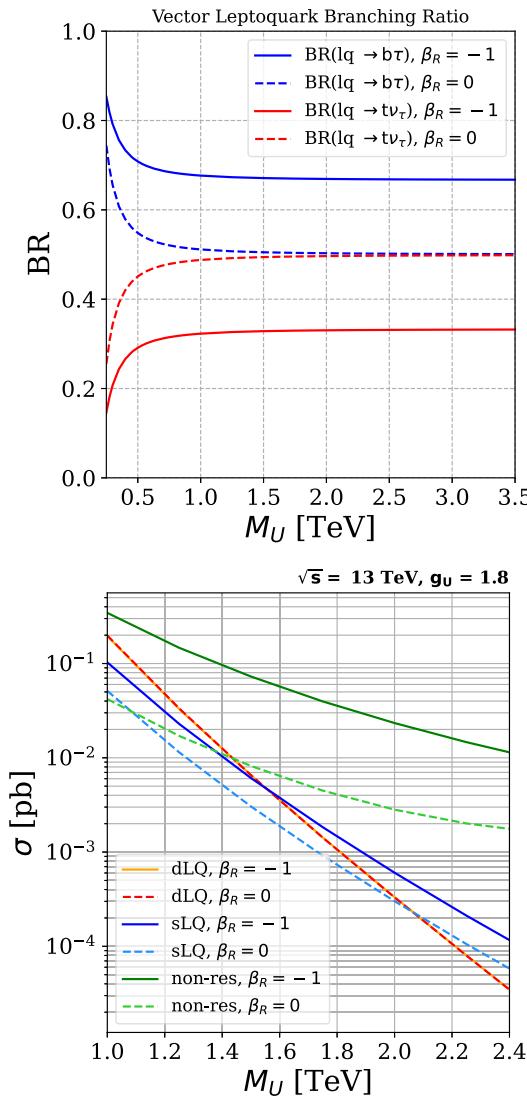
The LQ right-handed coupling is modulated with respect to the left-handed one by the  $\beta_R$  parameter. The choice of  $\beta_R$  is important phenomenologically, as it affects the LQ branching ratios,<sup>3</sup> as well as the single-LQ production cross-section. To illustrate the former, Fig. 2 (top) shows the LQ  $\rightarrow b\tau$  and LQ  $\rightarrow t\nu$  branching ratios as functions of the LQ mass, for two values of  $\beta_R$ . For large LQ masses, we confirm that with  $\beta_R = 0$  then  $BR(LQ \rightarrow b\tau) \approx BR(LQ \rightarrow t\nu) \approx \frac{1}{2}$ . However, for  $\beta_R = -1$ , as was chosen in [38], the additional coupling adds a new term to the total amplitude, leading to  $BR(LQ \rightarrow b\tau) \approx \frac{2}{3}$ . The increase in this branching ratio can thus weaken bounds from LQ searches targeting decays into  $t\nu$  final states, which motivates exploring the sensitivity in  $b\tau$  final states exclusively. Note that although a  $BR(LQ \rightarrow b\tau) \approx 1$  scenario is possible by having the LQ couple exclusively to right-handed currents (i.e.,  $g_U \rightarrow 0$ , but  $g_U \beta_R \neq 0$ ), it does not solve the observed anomalies in the  $R_{D^{(*)}}$  ratios. Therefore, although some LHC searches assume  $BR(LQ \rightarrow b\tau) = 1$ , we stress that in our studies we assume values of the model parameters and branching ratios that solve the  $R_{D^{(*)}}$  ratios.

To further understand the role of  $\beta_R$  at colliders, Fig. 2 (bottom) shows the cross-section for single-LQ (sLQ), double-LQ (dLQ), and non-resonant (non-res) production, as a function of mass and for a fixed coupling  $g_U = 1.8$ , assuming  $p p$  collisions at  $\sqrt{s} = 13$  TeV. We note that this benchmark scenario with  $g_U = 1.8$  results in a LQ  $\rightarrow b\tau$  decay width that is <5% of the LQ mass, for mass values

<sup>1</sup> The couplings in the second line of Eq. (1) can be found in the literature as  $g_s \rightarrow g_s(1 - \kappa_U)$  and  $g' \rightarrow g'(1 - \tilde{\kappa}_U)$ , in order to take into account the possibility of an underlying strong interaction.

<sup>2</sup> Before the demise of the  $R_{K^{(*)}}$  anomaly [42–45], a  $3 \times 3$   $\beta_L$  matrix would be used instead, with values fitted to solve all B meson anomalies.

<sup>3</sup> Having  $\beta_L^{st}$  different from zero also opens new decay channels. These, however, are either suppressed by  $\beta_L^{st}$  and powers of  $\lambda_{CKM}$ . In any case, this effect would decrease  $BR(LQ \rightarrow b\tau)$  and  $BR(LQ \rightarrow t\nu)$  by less than 3%.



**Fig. 2** Top: The LQ  $\rightarrow$   $b\tau$  and LQ  $\rightarrow$   $t\nu$  branching ratios for  $\beta_R = 0$  (solid lines) and  $\beta_R = -1$  (dashed lines). Bottom: Signal cross-section as a function of the LQ mass, for  $\sqrt{s} = 13$  TeV, with  $g_U = 1.8$ . We show single, pair, and non-resonant production, for  $\beta_R = -1, 0$  in solid and dashed lines, respectively

from 250 GeV to 2.5 TeV. In the figure, we observe that, since dLQ production is mainly mediated by events from quantum chromodynamic processes, the choice of  $\beta_R$  does not affect the cross-section. However, for sLQ production, a non-zero value for  $\beta_R$  increases the cross-section by about a factor of 2 and by almost one order of magnitude in the case of non-res production. These results shown in Fig. 2 are easily understood by considering the diagrams shown in Fig. 1. The LQ mass value where the sLQ production cross-section exceeds the dLQ cross-section depends on the choice of  $g_U$ .

We also note that to solve the  $R_{D^{(*)}}$  anomaly, the authors of [75] point out that the wilson coefficient  $C_U \equiv g_U^2 v_{SM}^2 / (4 M_U^2)$  is constrained to a specific range of values,

and this range depends on the value of the  $\beta_R$  parameter. Therefore, the allowed values of the coupling  $g_U$  depend on  $M_U$  and  $\beta_R$ , and thus our studies are performed in this multi-dimensional phase space.

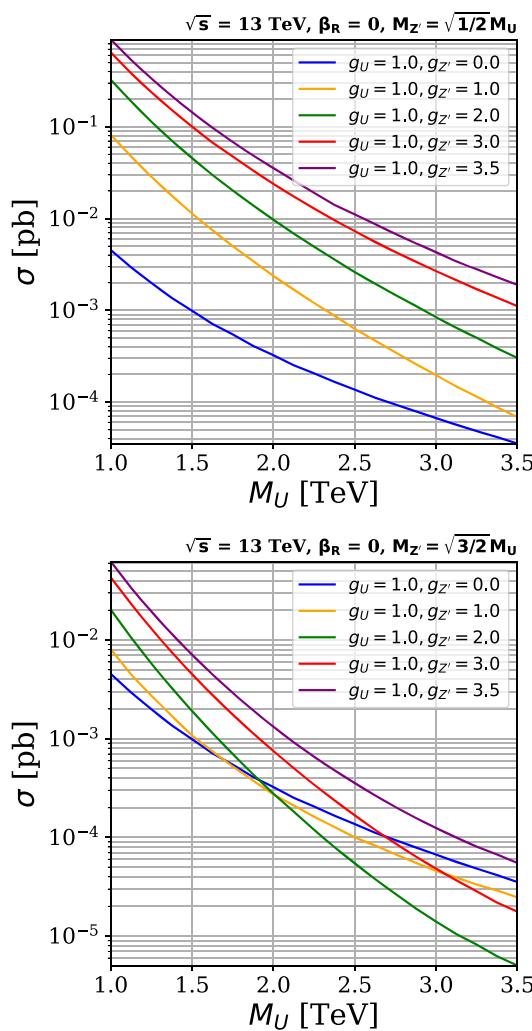
As noted in Sect. 1, we study the role of a  $Z'$  boson in  $p p \rightarrow \tau\tau$  production. The presence of a  $Z'$  boson in LQ models has been justified in various papers, for example, in [51]. The argument is that minimal extensions of the SM which include a massive gauge  $U_1$  LQ, uses the gauge group  $SU(4) \times SU(3)' \times SU(2)_L \times U(1)_{T_R^3}$ . Such an extension implies the presence of an additional massive boson,  $Z'$ , and a color-octet vector,  $G'$ , arising from the spontaneous symmetry breaking into the SM.<sup>4</sup> The  $Z'$  in particular can play an important role in the projected LQ discovery reach, as it can participate in  $p p \rightarrow \tau\tau$  production by s-channel exchange, both resonantly and as a virtual mediator. To study the effect of a  $Z'$  on the  $p p \rightarrow \tau\tau$  production cross-sections and kinematics, we extend our benchmark Lagrangian in Eq. (1) with further non-renormalizable terms involving the  $Z'$ . Accordingly, we assume the  $Z'$  only couples to third-generation fermions. Our simplified model is thus extended by:

$$\begin{aligned} \mathcal{L}_{Z'} = & -\frac{1}{4} Z'_{\mu\nu} Z'^{\mu\nu} + \frac{1}{2} M_{Z'}^2 Z'_\mu Z'^\mu \\ & + \frac{8Z'}{2\sqrt{6}} Z'^\mu (\zeta_q \bar{Q}_3 \gamma_\mu Q_3 + \zeta_t \bar{t}_R \gamma_\mu t_R \\ & + \zeta_b \bar{b}_R \gamma_\mu b_R - 3\zeta_\ell \bar{L}_3 \gamma_\mu L_3 - 3\zeta_\tau \bar{\tau}_R \gamma_\mu \tau_R) \end{aligned} \quad (2)$$

where the constants  $M_{Z'}$ ,  $g_{Z'}$ ,  $\zeta_q$ ,  $\zeta_t$ ,  $\zeta_b$ ,  $\zeta_\ell$ ,  $\zeta_\tau$ , are model dependent.

We study two extreme cases for the  $Z'$  mass, following [100], namely  $M_{Z'} = \sqrt{\frac{1}{2}} M_U < M_U$  and  $M_{Z'} = \sqrt{\frac{3}{2}} M_U > M_U$ . We also assume the LQ and  $Z'$  are uniquely coupled to left-handed currents, i.e.  $\zeta_q = \zeta_\ell = 1$  and  $\zeta_t = \zeta_b = \zeta_\tau = 0$ . With these definitions, Fig. 3 shows the effect of the  $Z'$  on the  $\tau\tau$  production cross-section, considering  $g_U = 1$ ,  $\beta_R = 0$ , and different  $g_{Z'}$  couplings. On the top panel, the cross-sections corresponding to the cases where  $M_{Z'} = \sqrt{\frac{1}{2}} M_U$  are shown. As expected, the  $\tau\tau$  production cross-section for the inclusive case (i.e.,  $g_{Z'} \neq 0$ ) is larger than that for the LQ-only non-res process ( $g_{Z'} = 0$ , depicted in blue). This effect increases with  $g_{Z'}$  and, within the evaluated values, can exceed the LQ-only cross-section by up to two orders of magnitude. In contrast, a more intricate behaviour can be seen in the bottom panel of Fig. 3, which corresponds to  $M_{Z'} = \sqrt{\frac{3}{2}} M_U$ . Here, for low values of  $M_U$ , a similar

<sup>4</sup> Naively, the LQs are associated to the breaking of  $SU(4) \rightarrow SU(3)_{[4]} \times U(1)_{B-L}$ , the  $G'$  arises from  $SU(3)_{[4]} \times SU(3)' \rightarrow SU(3)_c$ , and the  $Z'$  comes from the breaking of  $U(1)_{B-L} \times U(1)_{T_R^3} \rightarrow U(1)_Y$ . Notice that the specific pattern of breaking, and the relations between the masses and couplings, are connected to the specific scalar potential used.



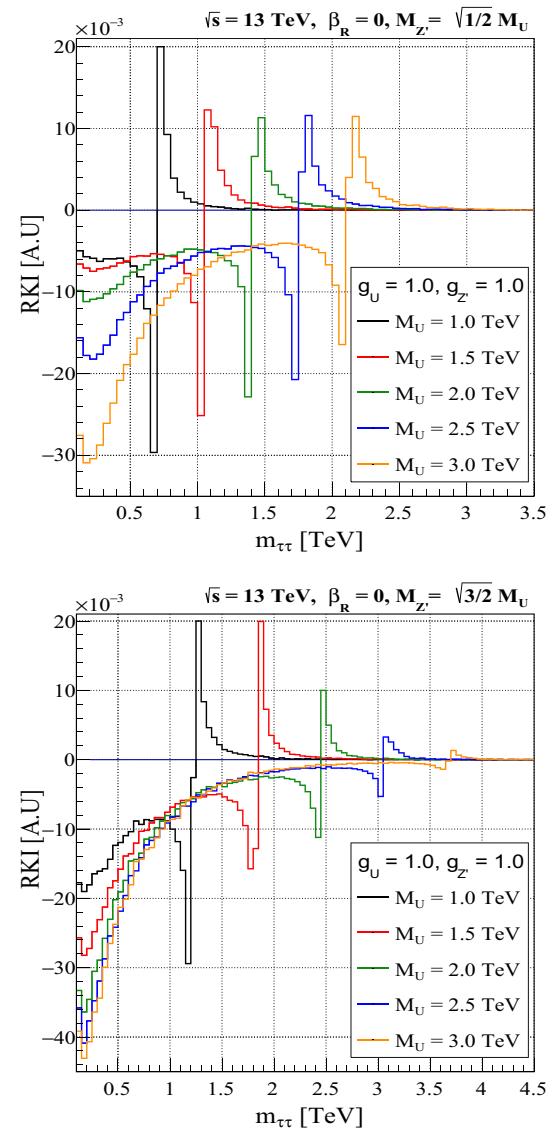
**Fig. 3**  $\tau\tau$  cross-section as a function of the LQ mass for different values of  $g_U$  and  $g_{Z'}$ . The estimates are performed at  $\sqrt{s} = 13$  TeV,  $\beta_R = 0$ ,  $M_{Z'} = \sqrt{1/2}M_U$  (top), and  $M_{Z'} = \sqrt{3/2}M_U$  (bottom)

increase in the cross-section is observed. However, for higher values of  $M_U$ , the inclusive  $p p \rightarrow \tau\tau$  cross-section is smaller than the LQ-only  $\tau\tau$  cross-section. This behaviour suggests the presence of a dominant destructive interference at high masses, leaving its imprint on the results.

In order to further illustrate the effect, Fig. 4 shows the relative kinematic interference (RKI) as a function of the reconstructed invariant mass  $m_{\tau\tau}$ , for  $g_{Z'} = 1$  and varying values of  $M_U$ . The RKI parameter is defined as

$$\text{RKI}(m_{\tau\tau}) = \frac{1}{\sigma_{\text{LQ}+Z'}} \left[ \frac{d\sigma_{\text{LQ}+Z'}}{dm_{\tau\tau}} - \left( \frac{d\sigma_{\text{LQ}}}{dm_{\tau\tau}} + \frac{d\sigma_{Z'}}{dm_{\tau\tau}} \right) \right], \quad (3)$$

where  $\sigma_X$  is the production cross-section arising due to contributions from  $X$  particles. For example,  $\sigma_{\text{LQ}+Z'}$  represents the inclusive cross-section where both virtual LQ and s-channel  $Z'$  exchange contribute. For both cases, we can



**Fig. 4** The relative kinematic interference (RKI), as a function of the reconstructed mass of two taus, for different LQ masses. The studies are performed assuming  $\sqrt{s} = 13$  TeV,  $\beta_R = 0$ ,  $g_U = 1.0$ ,  $g_{Z'} = 1.0$ ,  $M_{Z'} = \sqrt{1/2}M_U$  (top), and  $M_{Z'} = \sqrt{3/2}M_U$  (bottom)

observe the presence of deep valleys in the RKI curves when  $m_{\tau\tau} \rightarrow 0$ , indicating destructive interference between the LQ and the  $Z'$  contributions. This interference generates a suppression of the differential cross-section for lower values of  $m_{\tau\tau}$  and, therefore, in the integrated cross-section.

The observed interference effects are consistent with detailed studies on resonant and non-resonant  $p p \rightarrow t\bar{t}$  production, performed in reference [101].

### 3 LQ search strategy and simulation

Our proposed analysis strategy utilizes single-LQ (i.e.  $p p \rightarrow \tau\tau$  LQ), double-LQ (i.e.  $p p \rightarrow \text{LQ LQ}$ ), and non-resonant LQ

production (i.e.  $p p \rightarrow \tau\tau$ ) as shown in Fig. 1. At leading order in  $\alpha_s$ , since we focus on  $U_1 \rightarrow b\tau$  decays, the sLQ process results in the  $b\tau\tau$  mode, the dLQ process results in the  $bb\tau\tau$  mode, and the non-res process results in the  $\tau\tau$  mode. Therefore, in all cases we obtain two  $\tau$  leptons, with either 0, 1, or 2 b jets. The  $\tau$  leptons decay to hadrons ( $\tau_h$ ) or semi-leptonically to electrons or muons ( $\tau_\ell$ ,  $\ell = e$  or  $\mu$ ). To this end, we study six final states:  $\tau_h\tau_{h/\ell}$ ,  $b\tau_h\tau_{h/\ell}$ , and  $bb\tau_h\tau_{h/\ell}$ , which can be naively associated to non-res, sLQ and dLQ production, respectively. Nevertheless, experimentally it is possible for b jets to not be properly identified or reconstructed, leading, for instance, to a fraction of dLQ signal events falling into the  $b\tau_h\tau_{h/\ell}$  and  $\tau_h\tau_{h/\ell}$  categories. Similarly, soft jets can fake b jets, such that non-res processes can contribute to the  $b\tau_h\tau_{h/\ell}$  and  $bb\tau_h\tau_{h/\ell}$  final states. This kind of signal loss and mixing is taken into account in our analysis.<sup>5</sup>

The contributions of signal and background events are estimated using Monte Carlo (MC) simulations. We implemented the  $U_1$  model from [51], adjusted to describe the lagrangian in Eqs. (1) and (2), using FeynRules (v2.3.43) [102, 103]. The branching ratios and cross-sections have been calculated using MadGraph5\_aMC (v3.1.0) [104, 105], the latter at leading order in  $\alpha_s$ . The corresponding samples are generated considering  $p p$  collisions at  $\sqrt{s} = 13$  TeV and  $\sqrt{s} = 13.6$  TeV. All samples are generated using the NNPDF3.0 NLO [106] set for parton distribution functions (PDFs) and using the full amplitude square SDE strategy for the phase-space optimization due to strong interference effects with the  $Z'$  boson. Parton level events are then interfaced with the PYTHIA (v8.2.44) [107] package to include parton fragmentation and hadronization processes, while DELPHES (v3.4.2) [108] is used to simulate detector effects, using the input card for the CMS detector geometric configurations, and for the performance of particle reconstruction and identification.

At parton level, jets and leptons are required to have a minimum transverse momentum ( $p_T$ ) of 20 GeV, while b jets are required to have a minimum  $p_T$  of 30 GeV. Additionally, we constrain the pseudorapidity ( $\eta$ ) to  $|\eta| < 2.5$  for b jets and leptons, and  $|\eta| < 5.0$  for jets. The production cross-sections shown in the bottom panel of Figs. 2 and 3 are obtained with the aforementioned selection criteria.

Table 1 shows the preliminary event selection criteria for each channel at analysis level. The channels are divided based on the multiplicity of b jets,  $N(b)$ , number of light leptons,  $N(\ell)$ , number of hadronic tau leptons,  $N(\tau_h)$ , and kinematic

<sup>5</sup> Note that further signal mixing can also occur at the event generation level by including terms at larger order in  $\alpha_s$ . For example, in the non-res diagram in Fig. 1, one of the initial b could come from a  $g \rightarrow b\bar{b}$  splitting, leading to non resonant production of  $b\tau_h\tau_{h/\ell}$ . Simulating and studying the role of such NLO contributions is outside the scope of this work.

**Table 1** Preliminary event selection criteria used to filter events before feeding them to the BDT algorithm. A  $\Delta R(p_i, p_j) > 0.3$  requirement is imposed between all pairs of reconstructed particle candidates  $p_i, p_j$

Variable	Threshold					
	$\tau_h\tau_h$	$b\tau_h\tau_h$	$bb\tau_h\tau_h$	$\tau_h\tau_\ell$	$b\tau_h\tau_\ell$	$bb\tau_h\tau_\ell$
$N(b)$	= 0	= 1	$\geq 2$	= 0	= 1	$\geq 2$
$p_T(b)$	–	$\geq 30$ GeV	–	$\geq 30$ GeV	–	$\leq 2.4$
$ \eta(b) $	–	$\leq 2.4$	–	$\leq 2.4$	–	$\leq 2.4$
$N(\ell)$	= 0	–	–	= 1	–	–
$p_T(e)$	–	–	$\geq 35$ GeV	–	–	–
$p_T(\mu)$	–	–	$\geq 30$ GeV	–	–	–
$ \eta(\ell) $	–	–	$\leq 2.4$	–	–	–
$N(\tau_h)$	= 2	–	–	= 1	–	–
$p_T(\tau_h)$	$\geq 50$ GeV	–	–	–	–	–
$ \eta(\tau_h) $	$\leq 2.3$	–	–	–	–	–
$\Delta R(p_i, p_j)$	$\geq 0.3$	–	–	–	–	–

criteria based on  $\eta$ ,  $p_T$  and spatial separation of particles in the detector volume ( $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ ). The minimum  $p_T$  thresholds for leptons are chosen following references [62, 63, 67], based on experimental constraints associated to trigger performance. Following reference [109], we use a flat identification efficiency for b jets of 70% across the entire  $p_T$  spectrum with misidentification rate of 1%. These values correspond with the “medium working point” of the CMS algorithm to identify b jets, known as DeepCSV. We also explored the “Loose” (“Tight”) working point using an efficiency of 85% (45%) and mis-identification rate of 10% (0.1%). The “medium working point” was selected as it gives the best signal significance for the analysis.

For the performance of  $\tau_h$  identification in DELPHES, we consider the latest technique described in [110], which is based on a deep neural network (i.e. DeepTau) that combines variables related to isolation and  $\tau$ -lepton lifetime as input to identify different  $\tau_h$  decay modes. Following [110], we consider three possible DeepTau “working points”: (i) the “Medium” working point of the algorithm, which gives a 70%  $\tau_h$ -tagging efficiency and 0.5% light-quark and gluon jet mis-identification rate; (ii) the “Tight” working point, which gives a 60%  $\tau_h$ -tagging efficiency and 0.2% light-quark and gluon jet mis-identification rate; and (iii) the “VTight” working point, which gives a 50%  $\tau_h$ -tagging efficiency and 0.1% light-quark and gluon jet mis-identification rate. Similar to the choice of b-tagging working point, the choice of  $\tau_h$ -tagging working point is determined through an optimization process which maximizes discovery reach. The “Medium” working point was ultimately shown to provide the best sensitivity and therefore chosen for this study. For muons (electrons), the assumed identification efficiency is 95% (85%), with a 0.3% (0.6%) mis-identification rate [111–113].

**Table 2** The number of simulated events for the signal and background samples

Sample	$t\bar{t}$	single t	$VV$	$V+\text{jets}$	signals
$N_{\text{events}} \times 10^{-6}$	24.31	11.50	32.35	39.45	0.60

After applying the preliminary selection criteria, the primary sources of background are production of top quark pairs ( $t\bar{t}$ ), and single-top quark processes (single t), followed by production of vector bosons with associated jets from initial or final state radiation ( $V+\text{jets}$ ), and pair production of vector bosons ( $VV$ ). The number of simulated MC events used for each sample is shown in Table 2.

We use two different sets of signal samples. The first set includes various  $\{M_U, g_U\}$  scenarios, for two different values of  $\beta_R \in \{0, -1\}$ . We generate signal samples for  $M_U$  values between 250 and 5000 GeV, in steps of 250 GeV. The considered  $g_U$  coupling values are between 0.25 and 3.5, in steps of 0.25. Although the signal cross-sections depend on both  $M_U$  and  $g_U$ , the efficiencies of our selections only depend on  $M_U$  (for all practical purposes) since the decay widths are relatively small compared to the mass of  $M_U$  ( $\frac{\Gamma_U}{M_U} < 5\%$ ), and thus more sensitive to experimental resolution. In total there are 280  $\{M_U, g_U, \beta_R\}$  scenarios simulated for this first set of signal samples, and for each of these scenarios two subsets of samples are generated, which are used separately for the training and testing of the machine learning algorithm. The second set of signal samples is used to evaluate interference effects between LQs and the  $Z'$  bosons in non-res production. Using benchmark values  $g_U = 1.8$  and  $\beta_R = 0$ , we consider various  $\{M_U, g_{Z'}\}$  scenarios for two different  $Z'$  mass hypotheses,  $(M_{Z'}/M_U)^2 \in \{\frac{1}{2}, \frac{3}{2}\}$ . The  $M_U$  values vary between 500 and 5000 GeV, in steps of 250 GeV. The  $g_{Z'}$  coupling values are between 0.25 and 3.5, in steps of 0.25. Therefore, in total there are 280  $\{M_U, g_{Z'}, (M_{Z'}/M_U)^2\}$  scenarios simulated for this second set of signal samples, and for each of these scenarios a total of  $6.0 \times 10^5$  MC events are generated.

As noted previously, the simulated signal and background events are initially filtered using selections which are motivated by experimental constraints, such as the geometric constraints of the CMS detector, the typical kinematic thresholds for reconstruction of particle objects, and the available triggers. The remaining events after the preliminary event selection criteria are used to train and execute a BDT algorithm for each signal point in the  $\{M_U, g_U\}$  space, in order to maximize the probability to detect signal amongst background events. The BDT algorithm is implemented using the scikit-learn [114] and xgboost (XGB) [115] python libraries. We use the XGBClassifier class from the xgboost library, a 10-fold cross validation using

the scikit-learn method (GridCV<sup>6</sup>) for a grid in a hyperparameter space with 75, 125, 250, and 500 estimators, maximum depth in 3, 5, 7, 9, as well as learning rates of 0.01, 0.1, 1, and 10. For the cost function, we utilize the default mean square error (MSE). Additionally, we use the tree method based on the approximate greedy algorithm (histogram-optimized), referred to as hist, with a uniform sample method. These choices allow us to maximize the detection capability of the BDT algorithm by carefully tuning the hyperparameters, selecting an appropriate cost function, and utilizing an optimized tree construction method.

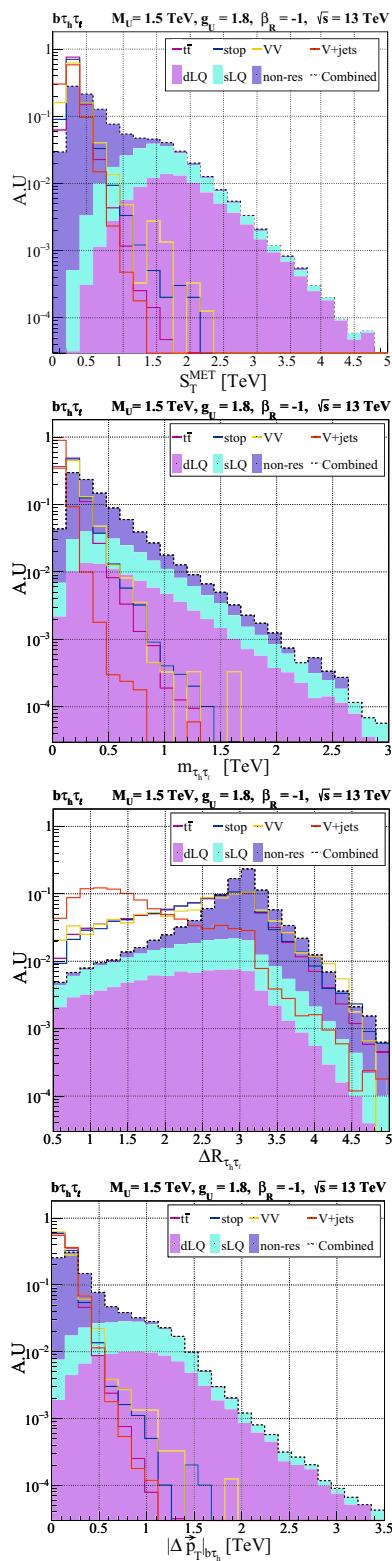
For each of the six analysis channels and  $\{M_U, g_U\}$  signal point, the binary XGB classifier was trained (tested) with 20% (80%) of the simulated events, for each signal and background MC sample. Over forty kinematic and topological variables were studied as input for the XGB. These included the momenta of b jets and  $\tau_{h,\ell}$  candidates; both invariant and transverse masses of pairs of  $\tau$  objects and of b  $\tau$  combinations; angular differences between b jets, between  $\tau$  objects, and between the  $\tau_{h,\ell}$  and b jets; and additional variables derived from the missing momentum in the events. After studying correlations between variables and their impact on the performance of the BDT, we found that only eight variables were necessary and responsible for the majority of the sensitivity of the analysis. The variable that provides the best signal to background separation is the scalar sum of the  $p_T$  of the final state objects ( $\tau_h$ ,  $\tau_{h/\ell}$ , and b jets) and the missing transverse momentum, referred to as  $S_T^{MET}$ :

$$S_T^{MET} = |\vec{p}_T^{miss}| + \sum_{\tau_h, \tau_{h/\ell}, b} |\vec{p}_T| \quad (4)$$

The  $S_T^{MET}$  variable has been successfully used in LQ searches at the LHC, since it probes the mass scale of resonant particles involved in the production processes. Other relevant variables include the magnitude of the vectorial difference in  $p_T$  between the two lepton candidates ( $|\Delta \vec{p}_T|_{\tau_h \tau_{h/\ell}}$ ), the  $\Delta R_{\tau_h \tau_{h/\ell}}$  separation between them, the reconstructed dilepton mass  $m_{\tau_h \tau_{h/\ell}}$ , and the product of their electric charges ( $Q_{\tau_h} \times Q_{\tau_{h/\ell}}$ ). We also use the  $|\Delta \vec{p}_T|$  between the  $\tau_h$  candidate and  $\vec{p}_T^{miss}$ , and (if applicable) the  $|\Delta \vec{p}_T|$  between the  $\tau_h$  candidate and the leading b jet. For the final states including two  $\tau_h$  candidates, the one with the highest  $p_T$  is used.

Figure 5 shows some relevant topological distributions, including  $S_T^{MET}$  on the top, for the b  $\tau_h \tau_\ell$  category. In the Figure we include all signal production modes to this channel, with each component weighted with respect to their total contribution to the combined signal. The combined signal distribution is normalised to unity. We also show all background processes contributing to this channel, each of them

<sup>6</sup> GridCV is a method that allows to find the best combination of hyperparameter values for the model, as this choice is crucial to achieve an optimal performance.



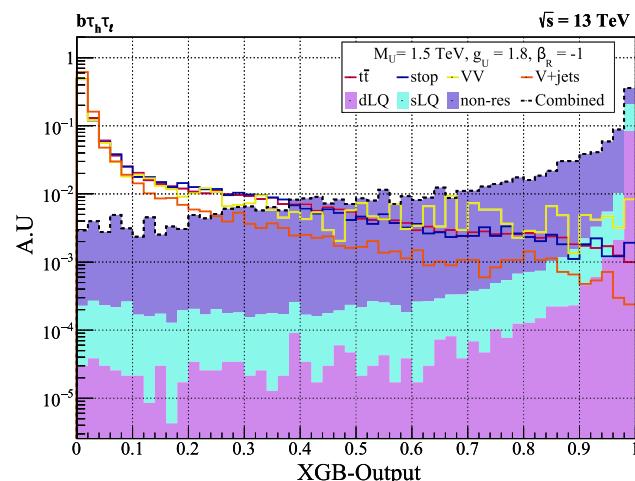
**Fig. 5**  $S_T^{MET}$ ,  $m_{\tau_h \tau_\ell}$ ,  $\Delta R_{\tau_h \tau_\ell}$ ,  $|\Delta \vec{p}_T|_{b\tau_h}$  signal and background distributions for the  $b\tau_h \tau_\ell$  channel. The signal distributions are generated for a benchmark sample with LQ mass of 1.5 TeV maximally coupled to right-handed currents. The combined distribution (shown as a stacked histogram) is the sum of the distributions, correctly weighted according to their respective cross-sections, assuming a coupling  $g_U = 1.8$

individually normalised to unity. We find that the combined signal is dominated by sLQ production for large values of  $S_T^{MET}$ , while non-res production dominates for small  $S_T^{MET}$ . Interestingly, the backgrounds also sit at low  $S_T^{MET}$  values, since  $S_T^{MET}$  is driven by the mass scale of the SM particles being produced, in this case top quarks and Z/W bosons. This suggest that the sLQ and dLQ signals can indeed be separated from the SM background. As expected, the  $S_T^{MET}$  sLQ and dLQ signal distributions have a mean near  $M_U$ , representative of resonant production, and a broad width as expected for large mass  $M_U$  hypotheses when information about the  $z$ -components of the momenta of objects is not utilised in the  $S_T^{MET}$  calculation.

Figure 5 (second from the top) shows the reconstructed mass of the ditau system, for the  $b\tau_h \tau_\ell$  search channel. Since the two  $\tau$  candidates in signal events arise from different production vertices (e.g., each  $\tau$  candidate in dLQ production comes from a different LQ decay chain), the ditau mass distribution for signal scales as  $m_{\tau_h \tau_\ell} \sim p_T(\tau_h) + p_T(\tau_\ell)$ , and thus has a tail which depends on  $M_U$  and sits above the expected SM spectrum. On the other hand, the SM  $m_{\tau_h \tau_\ell}$  distributions sit near  $m_{Z/W}$  since the  $\tau$  candidates in SM events arise from Z/W decays.

Figure 5 (third from the top) shows the  $\Delta R_{\tau_h \tau_\ell}$  distribution for the  $b\tau_h \tau_\ell$  channel. In the case of the  $p p \rightarrow \tau \tau$  non-res signal distribution, the two  $\tau$  leptons must be back-to-back to preserve conservation of momentum. Therefore, the visible  $\tau$  candidates,  $\tau_h$  and  $\tau_\ell$ , give rise to a  $\Delta R_{\tau_h \tau_\ell}$  distribution that peaks near  $\pi$  radians. In the case of sLQ production, although the LQ and associated  $\tau$  candidate must be back-to-back, the second  $\tau$  candidate arising directly from the decay of the LQ does not necessarily move along the direction of the LQ (since the LQ also decays to a b quark). As a result, the  $\Delta R_{\tau_h \tau_\ell}$  distribution for the sLQ signal process is smeared out, is broader, and has a mean below  $\pi$  radians. On the other hand, the  $\tau_h$  candidate in  $t\bar{t}$  events is often a jet being misidentified as a genuine  $\tau_h$ . When this occurs, the fake  $\tau_h$  candidate can arise from the same top quark decay chain as the  $\tau_\ell$  candidate, thus giving rise to small  $\Delta R_{\tau_h \tau_\ell}$  values. This difference in the signal and background distributions provides a nice way for the ML algorithm to help decipher signal and background processes.

As noted above, the  $|\Delta \vec{p}_T|$  distribution between the visible  $\tau$  candidates and the b-quark jets is an important variable to help the BDT distinguish between signal and background processes. The discriminating power can be seen in Fig. 5 (bottom), which shows the  $|\Delta \vec{p}_T|$  between the  $\tau_h$  and b-jet candidate of the  $b\tau_h \tau_\ell$  channel. In the case of dLQ production, the b quarks and  $\tau$  leptons from the LQ  $\rightarrow b\tau$  decay acquire transverse momentum of  $p_T \sim M_U/2$ . However, when the  $\tau$  lepton decays hadronically (i.e.  $\tau \rightarrow \tau_h \nu$ ), a large fraction of the momentum is lost to the neutrino. Therefore, the  $|\Delta \vec{p}_T|_{b\tau_h}$  distribution for the dLQ (and sLQ)



**Fig. 6** Postfit XGB-output normalised distribution in the  $b\tau_h\tau_\ell$  channel, for LQ mass of 1.5 TeV, constant coupling  $g_U = 1.8$ , and maximally coupled to right-handed currents

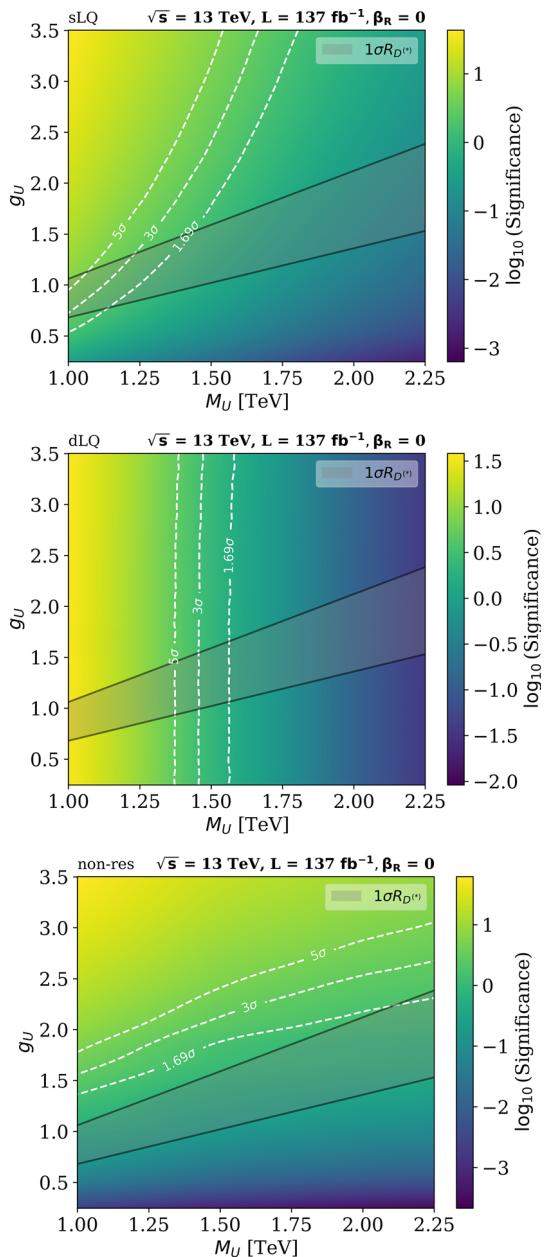
process peaks below  $M_U/2$ . On the other hand, for a background process such as  $V+jets$ , the  $b$  jet arises due to initial state radiation, and thus must balance the momentum of the associated vector boson (i.e.  $p_T(b) \sim p_T(V) \sim m_V$ ). Since the visible  $\tau$  candidate is typically produced from the  $V$  boson decay chain, its momentum (on average) is approximately  $p_T(\tau_h) \sim p_T(V)/4 \sim m_V/4$ . Therefore, to first order, the  $|\Delta\vec{p}_T|$  distribution for the  $V+jets$  background is expected to peak below the  $m_V$  mass.

Lets us turn to the results of the  $b\tau_h\tau_\ell$  BDT classifier, which is shown in Fig. 6 for the different signal production modes and backgrounds. Similar to Fig. 5, the distribution for each individual signal production mode is weighted with respect to their total contribution to the combined signal. The background distributions and combined signal distribution are normalized to an area under the curve of unity. Figure 6 shows the XGB distributions for a signal benchmark point with  $M_U = 1.5$  TeV,  $g_U = 1.8$ , and  $\beta_R = -1$ . The XGB output is a value between 0 and 1, which quantifies the likelihood that an event is either signal-like (XGB output near 1) or background-like (XGB output near 0). We see that the presence of the sLQ and dLQ production modes is observed as an enhancement near a XGB output of unity, while the backgrounds dominate over the low end of the XGB output spectrum, especially near zero. In fact, over eighty percent of the sLQ and dLQ distributions reside in the last two bins, XGB output greater than 0.96, while more than sixty percent of the backgrounds fall in the first two bins, XGB output less than 0.04. It is also interesting to note that in comparison to the sLQ and dLQ distributions in Fig. 6, non-res is broader and not as narrowly peaked near XGB output of 1, which is expected due to the differences in kinematics described

above. Overall, if we focus on the last bin in this distribution, we find approximately 0.2% of the background, in contrast to 22% of the non-res, 78% of the sLQ, and 91% of the dLQ signal distributions. These numbers highlight the effectiveness of the XGB output in reducing the background in the region where the signal is expected.

The output signal and background distributions of the XGB classifier, normalised to their cross section times pre-selection efficiency times luminosity, are used to perform a profile binned likelihood statistical test in order to determine the expected signal significance. The estimation is performed using the `RooFit` [116] package, following the same methodology as in Refs. [117–132]. The value of the significance ( $Z_{sig}$ ) is measured using the probability to obtain the same outcome from the test statistic in the background-only hypothesis, with respect to the signal plus background hypothesis. This allows for the determination of the local p-value and thus the calculation of the signal significance, which corresponds to the point where the integral of a Gaussian distribution between  $Z_{sig}$  and  $\infty$  results in a value equal to the local p-value.

Systematic uncertainties are incorporated as nuisance parameters, considering log-priors for normalization and Gaussian priors for shape uncertainties. Our consideration of systematic uncertainties includes both experimental and theoretical effects, focusing on the dominant sources of uncertainty. Following [133], we consider a 3% systematic uncertainty on the measurement of the integrated luminosity at the LHC. A 5% uncertainty arises due to the choice of the parton distribution function used for the MC production, following the PDF4LHC prescription [134]. The chosen PDF set only has an effect on the overall expected signal and background yields, but the effect on the shape of the XGB output distribution is negligible. Reference [110] reports a systematic uncertainty of 2–5%, depending on the  $p_T$  and  $\eta$  of the  $\tau_h$  candidate. Therefore, we utilize a conservative 5% uncertainty per  $\tau_h$  candidate, independent of  $p_T$  and  $\eta$ , which is correlated between signal and background processes with genuine  $\tau_h$  candidates, and correlated across XGB bins for each process. We assumed a 5%  $\tau_h$  energy scale uncertainty, independent of  $p_T$  and  $\eta$ , following the CMS measurements described in [110]. Finally, we assume a conservative 3% uncertainty per  $b$ -jet candidate, following reference [135], and an additional 10% uncertainty in all the background predictions to account for possible mismodeling by the simulated samples. The uncertainties on the background estimates are typically derived from collision data using dedicated control samples that have negligible signal contamination and are enriched with events from the specific targeted background. The systematic uncertainties on the background estimates are treated as uncorrelated between background processes.



**Fig. 7** Signal significance for different coupling scenarios and LQ masses, without right-handed currents, using the combination of all search channels. The results pertaining to sLQ, dLQ and non-res production are displayed respectively from the top. These results are for  $\sqrt{s} = 13 \text{ TeV}$  and  $137 \text{ fb}^{-1}$

#### 4 Results

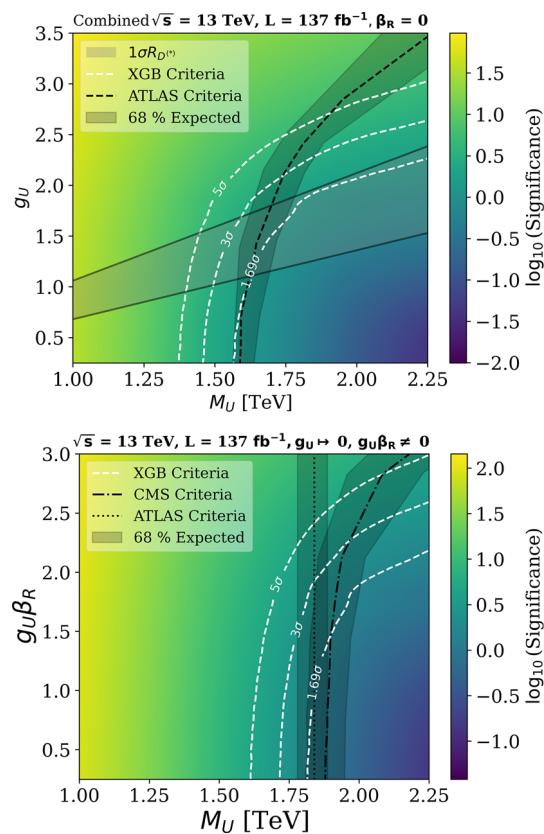
The expected signal significance for sLQ, dLQ and non-res production, and their combination, is presented in Fig. 7. Here, the significance is shown as a heat map in a two dimensional plane of  $g_U$  versus  $M_U$ , considering exclusive couplings to left-handed currents, i.e.  $\text{BR}(\text{LQ} \rightarrow b\tau) = \frac{1}{2}$ . The dashed lines show the contours of constant signal significance. The  $1.69\sigma$  contour represents exclusion at 95%

confidence level, and the  $3-5\sigma$  contours represent potential discovery. The grey band defines the set of  $\{M_U, g_U\}$  values that can explain the B-meson anomalies,  $C_U \sim 0.01$  for this scenario. The estimates are performed under the conditions for the second run, RUN-II, of the LHC ( $\sqrt{s} = 13 \text{ TeV}$  and  $L = 137 \text{ fb}^{-1}$ ). We find that the dLQ interpretation plot (Fig. 7 second from the top) does not depend on  $g_U$ , which is expected due to dLQ production arising exclusively from interactions with gluons. For this reason, the dLQ production process provides the best mode for discovery when  $g_U$  is small. On the other hand, the non-res channel is more sensitive to changes in the coupling parameter  $g_U$ , as its production cross-section depends on  $g_U^4$ . Therefore, the non-res production process provides the best mode for discovery when  $g_U$  is large. These results confirm the expectations from previous analyses (see for instance [49]), in the sense that the dLQ and non-res processes complement each other nicely at low and high  $g_U$  scenarios. The sLQ channel combines features from both the dLQ and non-res channels, in principle making it an interesting option to explore different scenarios and gain a better understanding of LQ properties, but the evolution of the signal significance in the full phase space is more complicated as it involves resonant LQ production with a cross-section that depends non-trivially on  $M_U$ ,  $g_U$ , and the LQ coupling to gluons. However, Fig. 7 shows that the sLQ production process can provide complementary and competitive sensitivity to the non-res and dLQ processes, in certain parts of the phase space.

The top panel of Fig. 8 presents the sensitivity of all signal production processes combined, and compares our expected exclusion region with the latest one from the ATLAS Collaboration [70]. The comparison suggests that our proposed analysis strategy provides better sensitivity than current methods being carried out at ATLAS, especially at large values of  $g_U$ . In particular, we find that with the pp data already available from RUN-II, our expected exclusion curves begin to probe solutions to the B-anomalies for LQ masses up to  $2.25 \text{ TeV}$ .

Figure 8 shows the expected signal significance considering  $\text{BR}(\text{LQ} \rightarrow b\tau) = 1$ , in order to compare our analysis with the corresponding results from the CMS [64] and ATLAS [71] Collaborations. Let us emphasize again that  $\text{BR}(\text{LQ} \rightarrow b\tau)$  depends on  $\beta_R$ , as illustrated on the top panel of Fig. 2. Thus, although the  $\text{BR}(\text{LQ} \rightarrow b\tau) = 1$  scenario is a possible physical case, it does not solve the observed anomalies in the  $R_{D^{(*)}}$  ratios, as it corresponds to the case where LQs couple exclusively to right-handed currents.

With this in mind, the scenario studied by CMS in [64] considers couplings only to left-handed currents, setting artificially the condition  $\text{BR}(\text{LQ} \rightarrow b\tau) = 1$ . In order to compare, we scale the efficiency  $\times$  acceptance of our selection criteria for  $\beta_R = 0$ , by a factor of 2.0 for sLQ and 4.0 for dLQ. According to Fig. 8, the ML approach that we have followed

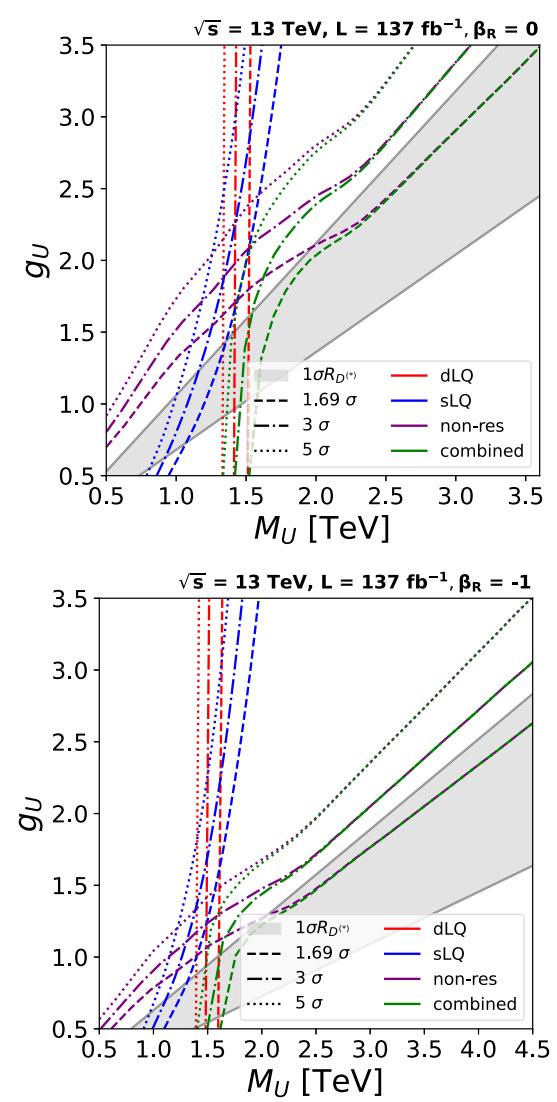


**Fig. 8** The top (bottom) panel shows signal significance comparison with ATLAS [70] (CMS and ATLAS [64, 71]) background only hypothesis, for the combination of all channels, with uniquely coupling to left-handed (right-handed) currents. The estimates are performed at  $\sqrt{s} = 13$  TeV and  $137 \text{ fb}^{-1}$

again suggests an optimisation of the signal and background separation, having the potential of improving the regions of exclusion ( $1.69\sigma$ ) with respect to that of CMS. In the bottom panel of the figure we have also included a similar exclusion by ATLAS [71]. However, since ATLAS only considers dLQ production in the analysis, the results are not entirely comparable, so are included only as a reference.

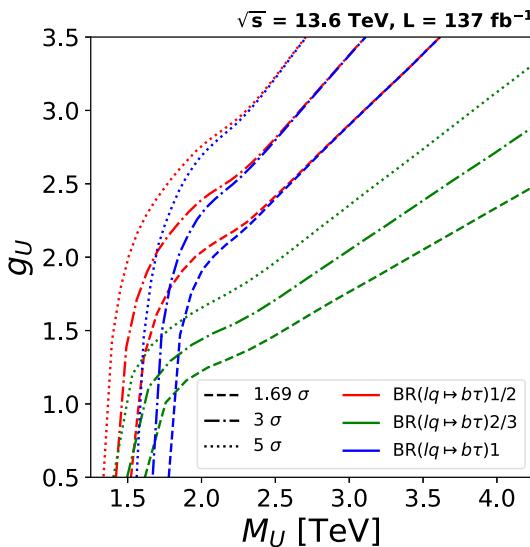
We now turn to the role of  $\beta_R$ , and our capacity of probing the regions solving the B-meson anomalies. Figure 9 shows the maximum significant contours, under LHC RUN-II conditions, for the different LQ production mechanisms and their combination, considering scenarios with only left-handed currents ( $\beta_R = 0$ , top) and with maximal right-handed currents ( $\beta_R = -1$ , bottom). We find a noticeable improvement in signal significance in all channels when taking  $\beta_R = -1$ , as is expected from the increase in  $\text{BR}(\text{LQ} \rightarrow b\tau)$  branching ratio and production cross-sections (see Fig. 2). However, the region solving the B-meson anomalies also changes, preferring lower values of  $g_U$ , such that in both cases we find ourselves just starting to probe this region at large  $M_U$ .

The combined significance contours for the different BR scenarios that have been considered is presented in Fig. 10.



**Fig. 9** Signal significance for different coupling scenarios and LQ masses for all channels. This plot summarizes our results with  $\beta_R = 0$  (without right-handed currents) and  $\beta_R = -1$  (maximally coupled to right-handed currents). The estimates are performed at  $\sqrt{s} = 13$  TeV and  $137 \text{ fb}^{-1}$

These contours illustrate the regions of exclusion for the three cases of interest, namely exclusive left-handed currents ( $\text{BR}(\text{LQ} \rightarrow b\tau) = \frac{1}{2}, \beta_R = 0$ ), maximal left and right couplings ( $\text{BR}(\text{LQ} \rightarrow b\tau) = \frac{2}{3}, \beta_R = -1$ ), and exclusive right-handed currents ( $\text{BR}(\text{LQ} \rightarrow b\tau) = 1, g_U \rightarrow 0, g_U \beta_R = 1$ ). For small  $g_U$ , we find that the exclusive right-handed scenario is most sensitive, while the exclusive left-handed case is the worst. The reason for this is that this region is excluded principally by dLQ production, such that having the largest branching ratio is crucial in order to have a large number of events. For larger couplings, both exclusive scenarios end up having similar exclusion regions, with the  $\beta_R = -1$  case being significantly more sensitive. The reason in this case is that the exclusion is dominated by

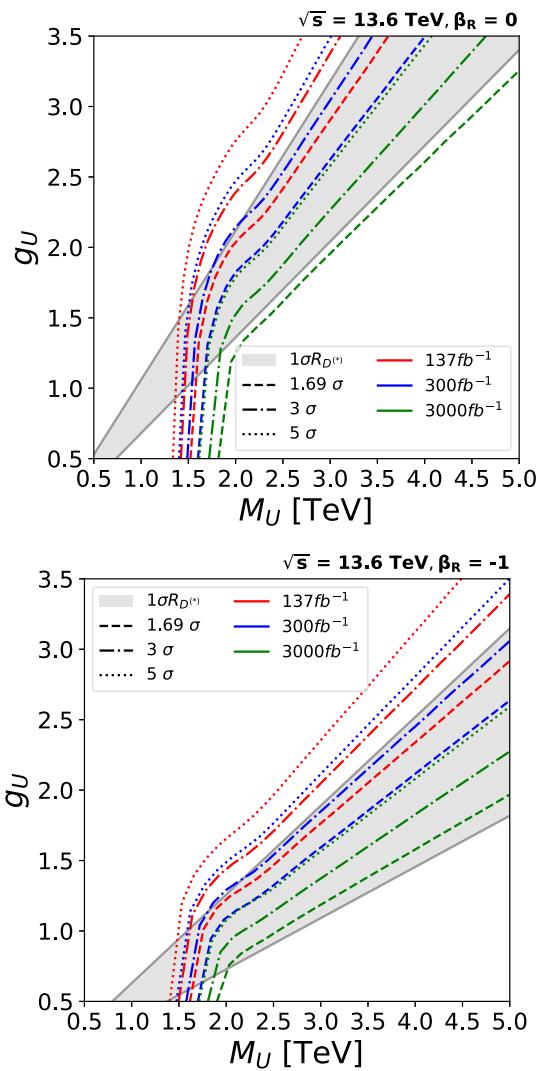


**Fig. 10** Signal significance for different coupling scenarios and LQ masses, considering the case without coupling to right-handed currents  $\text{BR}(\text{LQ} \rightarrow b\tau) = \frac{1}{2}$ , the case maximally coupled to right- and left-handed currents  $\text{BR}(\text{LQ} \rightarrow b\tau) = \frac{2}{3}$ , and the case uniquely coupled to right-handed currents  $\text{BR}(\text{LQ} \rightarrow b\tau) = 1$ . The estimates are performed at  $\sqrt{s} = 13 \text{ TeV}$  and  $137 \text{ fb}^{-1}$

non-res, which has a much larger production cross-section if both currents are turned on.

In order to finalise our analysis of the LQ-only model, we show in Fig. 11 the expected combined significance in the relatively near future. For this, considering  $\sqrt{s} = 13.6 \text{ TeV}$ , we show contours for the sensitivity corresponding to integrated luminosities of  $137 \text{ fb}^{-1}$ ,  $300 \text{ fb}^{-1}$ , and  $3000 \text{ fb}^{-1}$ , for scenarios with only left-handed currents (top) and with maximal coupling to right-handed currents (bottom). Note that for  $\beta_R = 0$  ( $\beta_R = -1$ ), couplings  $g_U$  close to  $3.18$  ( $1.85$ ) and  $M_U = 5.0 \text{ TeV}$  can be excluded with  $1.69\sigma$  significance for the high luminosity LHC era, allowing us to probe the practically the entirety of the B-meson anomaly favored region. Note that the background yields for the high luminosity LHC might be larger due to pileup effects. Nevertheless, as it was mentioned in Sect. 3, we have included a conservative 10% systematic uncertainty associated with possible fluctuations on the background estimations. Although effects from larger pileup might be significant, they can be mitigated by improvements in the algorithms for particle reconstruction and identification, and also on the data-analysis techniques.

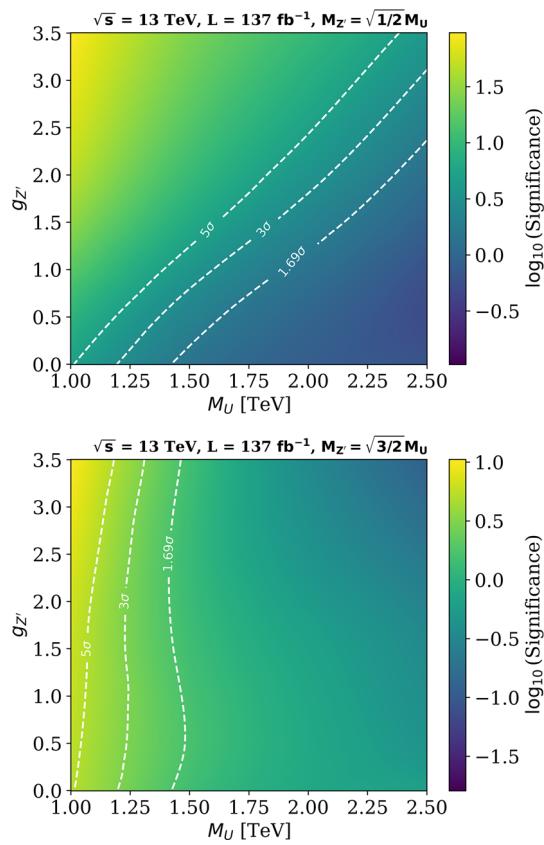
As commented on the Introduction, non-res production can be significantly affected by the presence of a companion  $Z'$ , which provides additional s-channel diagrams that add to the total cross-section and can interfere destructively with the LQ t-channel process (see Figs. 3 and 4). From our previous results, we see that non-res always is of high importance in determining the exclusion region, particularly at large  $M_U$ .



**Fig. 11** Projected signal significance for different coupling scenarios and LQ masses maximally coupled to right-handed currents. The estimates are performed at  $\sqrt{s} = 13.6 \text{ TeV}$ ,  $137 \text{ fb}^{-1}$ ,  $300 \text{ fb}^{-1}$  and  $3000 \text{ fb}^{-1}$

and  $g_U$ , meaning it is crucial to understand how this role is affected in front of a  $Z'$  with similar mass.

The change in sensitivity on the non-res signal significance due this interference effect with the  $Z'$  boson is shown in Fig. 12. We consider two opposite cases for the  $Z'$  mass:  $M_{Z'}^2 = M_U^2/2$  (top) and  $M_{Z'}^2 = 3M_U^2/2$  (bottom). Our results are shown on the  $g_{Z'} - M_U$  plane, for a fixed  $g_U = 1.8$  and  $\beta_R = 0$ . For the  $M_{Z'}^2 = M_U^2/2$  scenario, there is an overall increase in the total cross-section, with a larger  $g_{Z'}$  implying a larger sensitivity. This means that our ability to probe smaller values of  $g_U$  could be enhanced, as a given observation would be reproduced with both a specific  $g_U$  and vanishing  $g_{Z'}$ , or a smaller  $g_U$  with large  $g_{Z'}$ . Thus, for a large enough  $g_{Z'}$ , it could be possible to enhance non-res to the point that the entire region favoured by B-anomalies could

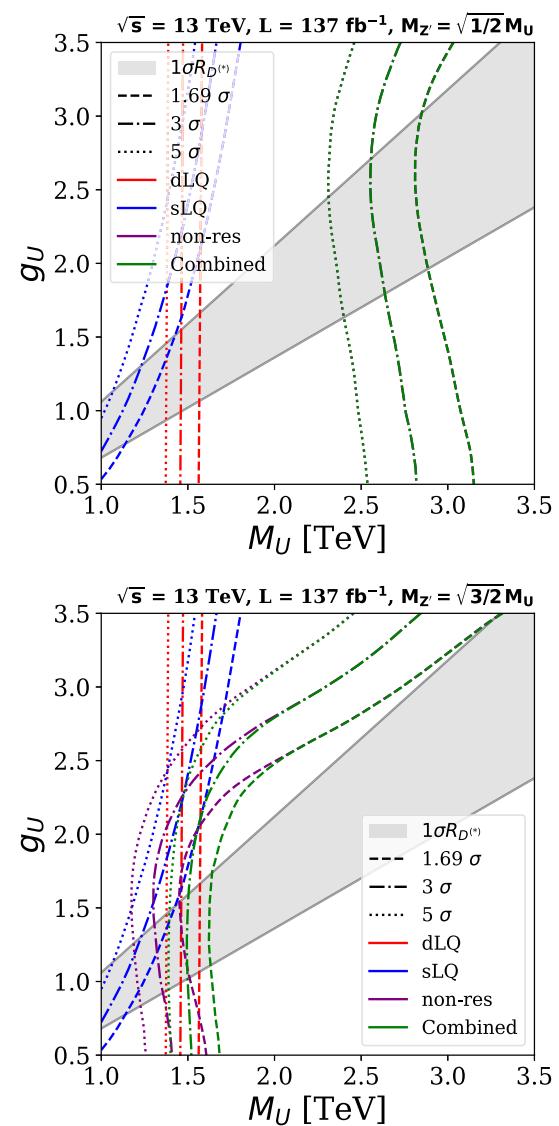


**Fig. 12** Change on the non-res signal significance for different  $Z'$  coupling scenarios and LQ masses. The estimates are performed at  $\sqrt{s} = 13.0$  TeV,  $\beta_R = 0$ ,  $g_U = 1.8$ ,  $M_{Z'} = \sqrt{1/2}M_U$  (top), and  $M_{Z'} = \sqrt{3/2}M_U$  (bottom)

be ruled out. In contrast, for  $M_{Z'}^2 = 3M_U^2/2$  the cross-section is strongly affected by the large destructive interference, such that a larger  $g_{Z'}$  does not necessarily imply an increase in sensitivity. In fact, as can be seen in the bottom panel, for large  $M_U$  the significance is reduced as  $g_{Z'}$  increases, leading to the opposite conclusion than above, namely, that a large  $g_{Z'}$  could reduce the effectiveness of non-res.

The impact of the above can be seen in Fig. 13, which shows our previous sensitivity curves on the  $M_U - g_U$  plane, but this time with a  $Z'$  contribution to non-res. We use the same values of  $M_{Z'}$  as before, but fix  $g_{Z'} = 3.5$ . For smaller  $M_{Z'}$  (top), the non-res contribution is enhanced so much, that both sLQ and dLQ play no role whatsoever in determining the exclusion region. We find that, for small  $g_U$ , the sensitivity is dominated by  $Z'$  production such that, since  $M_U$  is related to  $M_{Z'}$ , LQ masses up to  $\sim 3$  TeV are excluded. This bound is slightly relaxed for larger values of  $g_U$ , which is attributed to destructive interference effects due to an increased LQ contribution.

The bottom panel of Fig. 13 shows that case where  $M_{Z'}$  is larger than  $M_U$ . As expected from our previous discussion, the behaviour and impact of non-res is modified. For small



**Fig. 13** Signal significance for different coupling scenarios and LQ masses, for all channels, with an additional  $Z'$  contribution to non-res production. We set  $\beta_R = 0$  and  $g_{Z'} = 3.5$ , taking  $M_{Z'}^2$  equal to  $M_U^2/2$  (3 $M_U^2/2$ ) on the top (bottom) panel

$g_U$ , we again have the pure  $Z'$  production dominating the non-res cross-section, leading to a null sensitivity on  $g_U$ , similar to what happens in dLQ. In contrast, for very large  $g_U$ , we find that the pure LQ non-res production is the one that dominates, and we recover sensitivity regions with a slope similar to those shown in Figs. 7, 8, 9, 10 and 11, shifted towards larger values of  $g_U$ . For intermediate values of this coupling, the destructive interference have an important effect again, twisting the exclusion region slightly towards the left. Still, even in this case, we find that sLQ plays a marginal role in defining the combined exclusion region, and that the final result again depends primarily on dLQ and non-res production.

## 5 Discussion and conclusions

Experimental searches for LQs with preferential couplings to third generation fermions are currently of great interest due to their potential to explain observed tensions in the  $R_{(D)}$  and  $R_{(D^*)}$  decay ratios of B mesons with respect to the SM predictions. Although the LHC has a broad physics program on searches for LQs, it is very important to consider the impact of each search within wide range of different theoretical assumptions within a specific model. In addition, in order to improve the sensitivity to detect possible signs of physics beyond the SM, it is also important to strongly consider new computational techniques based on machine learning (ML). Therefore, we have studied the production of  $U_1$  LQs with preferential couplings to third generation fermions, considering different couplings, masses and chiral currents. These studies have been performed considering p p collisions at  $\sqrt{s} = 13$  TeV and 13.6 TeV and different luminosity scenarios, including projections for the high luminosity LHC. A ML algorithm based on boosted decision trees is used to maximize the signal significance. The signal to background discrimination output of the algorithm is taken as input to perform a profile binned-likelihood test statistic to extract the expected signal significance.

The expected signal significance for sLQ, dLQ and non-res production, and their combination, is presented as contours on a two dimensional plane of  $g_U$  versus  $M_U$ . We present results for the case of exclusive couplings to left-handed, mixed, and exclusive right-handed currents. For the first two, the region of the phase space that could explain the B meson anomalies is also presented. We confirm the findings of previous works that the largest production cross-section and best overall significance comes from the combination of dLQ and non-res production channels. We also find that the sensitivity to probe the parameter space of the model is highly dependent on the chirality of the couplings. Nevertheless, the region solving the B-meson anomalies also changes with each choice, such that in all evaluated cases we find ourselves just starting to probe this region at large  $M_U$ .

Our studies compare our exclusion regions with respect to the latest reported results from the ATLAS and CMS Collaborations. The comparison suggests that our ML approach has a better sensitivity than the standard cut-based analyses, especially at large values of  $g_U$ . In addition, our projections for the HL-LHC cover the whole region solving the B-anomalies, for masses up to 5.00 TeV.

Finally, we consider the effects of a companion Z' boson on non-res production. We find that such a contribution can have a considerable impact on the LQ sensitivity regions, depending on the specific masses and couplings. In spite of this, we still consider non-res production as an essential channel for probing LQs in the future.

**Acknowledgements** The authors would like to thank Gino Isidori for fruitful discussions. A.F, J.P, and C.R thank the constant and enduring financial support received for this project from the faculty of science at Universidad de Los Andes (Bogotá, Colombia) and the Colombian Science Ministry MinCiencias, with the grant program 70141, contract number 164-2021. J.J.P. acknowledges funding by the *Dirección de Gestión de la Investigación* at PUCP, through grant No. DGI-2021-C-0020. A.G acknowledges the funding received from the Physics & Astronomy department at Vanderbilt University and the US National Science Foundation. This work is supported in part by NSF Award PHY-1945366 and a Vanderbilt Seeding Success Grant.

**Data availability statement** This manuscript has no associated data or the data will not be deposited. [Authors' comment: This is a theoretical study and no experimental data].

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Funded by SCOAP<sup>3</sup>. SCOAP<sup>3</sup> supports the goals of the International Year of Basic Sciences for Sustainable Development.

## References

1. J.C. Pati, A. Salam, Lepton number as the fourth color. *Phys. Rev. D* **10**, 275–289 (1974). (**Erratum:** *Phys. Rev. D* **11**, 703–703 (1975))
2. H. Georgi, S.L. Glashow, Unity of all elementary particle forces. *Phys. Rev. Lett.* **32**, 438–441 (1974)
3. B. Schrempp, F. Schrempp, Light leptoquarks. *Phys. Lett. B* **153**, 101–107 (1985)
4. M. Leurer, A comprehensive study of leptoquark bounds. *Phys. Rev. D* **49**, 333–342 (1994)
5. S. Davidson, D.C. Bailey, B.A. Campbell, Model independent constraints on leptoquarks from rare processes. *Z. Phys. C* **61**, 613–644 (1994)
6. M. Leurer, Bounds on vector leptoquarks. *Phys. Rev. D* **50**, 536–541 (1994)
7. J.L. Hewett, T.G. Rizzo, Much ado about leptoquarks: a comprehensive analysis. *Phys. Rev. D* **56**, 5709–5724 (1997)
8. F.S. Queiroz, K. Sinha, A. Strumia, Leptoquarks, dark matter, and anomalous LHC events. *Phys. Rev. D* **91**(3), 035006 (2015)
9. I. Doršner, S. Fajfer, A. Greljo, J.F. Kamenik, N. Košnik, Physics of leptoquarks in precision experiments and at particle colliders. *Phys. Rep.* **641**, 1–68 (2016)
10. LHCb Collaboration, Test of lepton universality using  $B^+ \rightarrow K^+ \ell^+ \ell^-$  decays. *Phys. Rev. Lett.* **113**(2014)
11. LHCb Collaboration, Test of lepton universality with  $B^0 \rightarrow K^{*0} \ell^+ \ell^-$  decays. *JHEP* **08**, 055 (2017)
12. LHCb Collaboration, Search for lepton-universality violation in  $B^+ \rightarrow K^+ \ell^+ \ell^-$  decays. *Phys. Rev. Lett.* **122**(19), 191801 (2019)
13. LHCb Collaboration, Test of lepton universality in beauty-quark decays. *Nat. Phys.* **18**(3), 277–282 (2022)

14. B.B. Collaboration, Evidence for an excess of  $\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$  decays. *Phys. Rev. Lett.* **109**, 101802 (2012)
15. B.B. Collaboration, Measurement of an excess of  $\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$  decays and implications for charged Higgs bosons. *Phys. Rev. D* **88**(7), 072012 (2013)
16. Belle Collaboration, Measurement of  $\mathcal{R}(D)$  and  $\mathcal{R}(D^*)$  with a semileptonic tagging method (2019)
17. B. Collaboration, Measurement of the  $\tau$  lepton polarization and  $R(D^*)$  in the decay  $\bar{B} \rightarrow D^*\tau^-\bar{\nu}_\tau$  with one-prong hadronic  $\tau$  decays at Belle. *Phys. Rev. D* **97**, 012004 (2018)
18. B. Collaboration, Measurement of the branching ratio of  $\bar{B}^0 \rightarrow D^{*+}\tau^-\bar{\nu}_\tau$  relative to  $\bar{B}^0 \rightarrow D^{*+}\ell^-\bar{\nu}_\ell$  decays with a semileptonic tagging method. *Phys. Rev. D* **94**, 072007 (2016)
19. B. Collaboration, Measurement of the  $\tau$  lepton polarization and  $R(D^*)$  in the decay  $\bar{B} \rightarrow D^*\tau^-\bar{\nu}_\tau$ . *Phys. Rev. Lett.* **118**, 211801 (2017)
20. B. Collaboration, Measurement of the branching ratio of  $\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$  relative to  $\bar{B} \rightarrow D^{(*)}\ell^-\bar{\nu}_\ell$  decays with hadronic tagging at Belle. *Phys. Rev. D* **92**, 072014 (2015)
21. LHCb Collaboration, Measurement of the ratio of branching fractions  $\mathcal{B}(\bar{B}^0 \rightarrow D^{*+}\tau^-\bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \rightarrow D^{*+}\mu^-\bar{\nu}_\mu)$ . *Phys. Rev. Lett.* **115**(11), 111803 (2015). [Erratum: *Phys. Rev. Lett.* **115**, 159901 (2015)]
22. LHCb Collaboration, Measurement of the ratio of branching fractions  $\mathcal{B}(\bar{B}^0 \rightarrow D^{*+}\tau^-\bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \rightarrow D^{*+}\mu^-\bar{\nu}_\mu)$ . *Phys. Rev. Lett.* **115**, 111803 (2015). [Erratum: *Phys. Rev. Lett.* **115**, 159901 (2015)]
23. LHCb Collaboration, Measurement of the ratio of the  $B^0 \rightarrow D^{*-}\tau^+\nu_\tau$  and  $B^0 \rightarrow D^{*-}\mu^+\nu_\mu$  branching fractions using three-prong  $\tau$ -lepton decays. *Phys. Rev. Lett.* **120**, 171802 (2018)
24. LHCb Collaboration, Test of lepton flavor universality by the measurement of the  $B^0 \rightarrow D^{*-}\tau^+\nu_\tau$  branching fraction using three-prong  $\tau$  decays. *Phys. Rev. D* **97**(7), 072013 (2018)
25. LHCb Collaboration, Measurement of the ratios of branching fractions  $\mathcal{R}(D^*)$  and  $\mathcal{R}(D^0)$  (2023)
26. G. Hiller, M. Schmaltz,  $R_K$  and future  $b \rightarrow s\ell\ell$  physics beyond the standard model opportunities. *Phys. Rev. D* **90**, 054014 (2014)
27. B. Gripaios, M. Nardecchia, S.A. Renner, Composite leptoquarks and anomalies in  $B$ -meson decays. *JHEP* **05**, 006 (2015)
28. R. Alonso, B. Grinstein, J.M. Camalich, Lepton universality violation and lepton flavor conservation in  $B$ -meson decays. *JHEP* **10**, 184 (2015)
29. L. Calibbi, A. Crivellin, T. Ota, Effective field theory approach to  $b \rightarrow s\ell\ell^{(\prime)}$ ,  $B \rightarrow K^{(*)}\nu\bar{\nu}$  and  $B \rightarrow D^{(*)}\tau\nu$  with third generation couplings. *Phys. Rev. Lett.* **115**, 181801 (2015)
30. S. Fajfer, N. Košnik, Vector leptoquark resolution of  $R_K$  and  $R_{D^{(*)}}$  puzzles. *Phys. Lett. B* **755**, 270–274 (2016)
31. M. Bauer, M. Neubert, Minimal leptoquark explanation for the  $R_{D^{(*)}}$ ,  $R_K$ , and  $(g-2)_\mu$  anomalies. *Phys. Rev. Lett.* **116**(14), 141802 (2016)
32. D. Bećirević, N. Košnik, O. Sumensari, R. Zukanovich-Funchal, Palatable leptoquark scenarios for lepton flavor violation in exclusive  $b \rightarrow s\ell_1\ell_2$  modes. *JHEP* **11**, 035 (2016)
33. A. Crivellin, D. Müller, T. Ota, Simultaneous explanation of  $R(D^*)$  and  $b \rightarrow \mu^+\mu^-$ : the last scalar leptoquark standing. *JHEP* **09**, 040 (2017)
34. G. D'Amico, M. Nardecchia, P. Panci, F. Sannino, A. Strumia, R. Torre, A. Urbano, Flavour anomalies after the  $R_{K^*}$  measurement. *JHEP* **09**, 010 (2017)
35. G. Hiller, I. Nisandžić,  $R_K$  and  $R_{K^*}$  beyond the standard model. *Phys. Rev. D* **96**(3), 035003 (2017)
36. D. Buttazzo, A. Greljo, G. Isidori, D. Marzocca, B-physics anomalies: a guide to combined explanations. *JHEP* **11**, 044 (2017)
37. D. Bećirević, I. Doršner, S. Fajfer, N. Košnik, D.A. Faroughy, O. Sumensari, Scalar leptoquarks from grand unified theories to accommodate the  $B$ -physics anomalies. *Phys. Rev. D* **98**(5), 055003 (2018)
38. C. Cornellà, J. Fuentes-Martín, G. Isidori, Revisiting the vector leptoquark explanation of the  $B$ -physics anomalies. *JHEP* **07**, 168 (2019)
39. A. Angelescu, D. Bećirević, D.A. Faroughy, F. Jaffredo, O. Sumensari, Single leptoquark solutions to the  $B$ -physics anomalies. *Phys. Rev. D* **104**(5), 055017 (2021)
40. G. Belanger et al., Leptoquark manœuvres in the dark: a simultaneous solution of the dark matter problem and the  $R_{D^{(*)}}$  anomalies. *JHEP* **02**, 042 (2022)
41. J. Aebrischer, G. Isidori, M. Pesut, B.A. Stefanek, F. Wilsch, Confronting the vector leptoquark hypothesis with new low- and high-energy data. *Nature* **83**(2), 153 (2023)
42. LHCb Collaboration, Test of lepton universality in  $b \rightarrow \ell^+\ell^-$  decays (2022)
43. LHCb Collaboration, Measurement of lepton universality parameters in  $B^+ \rightarrow K^+\ell^+\ell^-$  and  $B^0 \rightarrow K^{*0}\ell^+\ell^-$  decays (2022)
44. A. Greljo, J. Salko, A. Smolkočić, P. Stangl, Rare  $b$  decays meet high-mass Drell-Yan. *JHEP* **5**, 2023 (2023)
45. M. Ciuchini, M. Fedele, E. Franco, A. Paul, L. Silvestrini, M. Valli, Constraints on lepton universality violation from rare  $b$  decays. *Phys. Rev. D* **107**, 055036 (2023)
46. B. Diaz, M. Schmaltz, Y. Zhong, The leptoquark Hunter's guide: Pair production. *JHEP* **10**, 097 (2017)
47. I. Doršner, A. Greljo, Leptoquark toolbox for precision collider studies. *JHEP* **05**, 126 (2018)
48. N. Vignaroli, Seeking leptoquarks in the  $t\bar{t}$  plus missing energy channel at the high-luminosity lhc. *Phys. Rev. D* **99**, 035021 (2019)
49. M. Schmaltz, Y. Zhong, The leptoquark Hunter's guide: large coupling. *JHEP* **01**, 132 (2019)
50. A. Biswas, D. Kumar-Ghosh, N. Ghosh, A. Shaw, A.K. Swain, Collider signature of  $U_1$  leptoquark and constraints from  $b \rightarrow c$  observables. *J. Phys. G* **47**(4), 045005 (2020)
51. M.J. Baker, J. Fuentes-Martín, G. Isidori, M. König, High- $p_T$  signatures in vector-leptoquark models. *Eur. Phys. J. C* **79**(4), 334 (2019)
52. U. Haisch, G. Polesello, Resonant third-generation leptoquark signatures at the Large Hadron Collider. *JHEP* **05**, 057 (2021)
53. A. Bhaskar, T. Mandal, S. Mitra, M. Sharma, Improving third-generation leptoquark searches with combined signals and boosted top quarks. *Phys. Rev. D* **104**(7), 075037 (2021)
54. J. Bernigaud, M. Blanke, I.M. Varzielas, J. Talbert, J. Zurita, LHC signatures of  $\tau$ -flavoured vector leptoquarks. *JHEP* **08**, 127 (2022)
55. R. Leonardi, O. Panella, F. Romeo, A. Gurrola, H. Sun, S. Xue, Phenomenology at the LHC of composite particles from strongly interacting Standard Model fermions via four-fermion operators of NJL type. *EPJC* **80**(309) (2020)
56. C.M.S. Collaboration, Search for heavy neutrinos or third-generation leptoquarks in final states with two hadronically decaying  $\tau$  leptons and two jets in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *JHEP* **03**, 077 (2017)
57. C.M.S. Collaboration, Search for third-generation scalar leptoquarks and heavy right-handed neutrinos in final states with two tau leptons and two jets in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *JHEP* **07**, 121 (2017)
58. C.M.S. Collaboration, Search for third-generation scalar leptoquarks decaying to a top quark and a  $\tau$  lepton at  $\sqrt{s} = 13$  TeV. *Eur. Phys. J. C* **78**, 707 (2018)
59. C.M.S. Collaboration, Constraints on models of scalar and vector leptoquarks decaying to a quark and a neutrino at  $\sqrt{s} = 13$  TeV. *Phys. Rev. D* **98**(3), 032005 (2018)
60. C.M.S. Collaboration, Search for a singly produced third-generation scalar leptoquark decaying to a  $\tau$  lepton and a bottom

- quark in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *JHEP* **07**, 115 (2018)
61. C.M.S. Collaboration, Search for heavy neutrinos and third-generation leptoquarks in hadronic states of two  $\tau$  leptons and two jets in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *JHEP* **03**, 170 (2019)
  62. C.M.S. Collaboration, Search for singly and pair-produced leptoquarks coupling to third-generation fermions in proton-proton collisions at  $s=13$  TeV. *Phys. Lett. B* **819**, 136446 (2021)
  63. CMS Collaboration, Searches for additional Higgs bosons and for vector leptoquarks in  $\tau\tau$  final states in proton-proton collisions at  $\sqrt{s} = 13$  TeV (2022)
  64. CMS Collaboration, The search for a third-generation leptoquark coupling to a  $\tau$  lepton and a b quark through single, pair and nonresonant production at  $\sqrt{s} = 13$  TeV (2022)
  65. ATLAS Collaboration, Searches for third-generation scalar leptoquarks in  $\sqrt{s} = 13$  TeV pp collisions with the ATLAS detector. *JHEP* **06**, 144 (2019)
  66. ATLAS Collaboration, Search for a scalar partner of the top quark in the all-hadronic  $t\bar{t}$  plus missing transverse momentum final state at  $\sqrt{s} = 13$  TeV with the ATLAS detector. *Eur. Phys. J. C* **80**(8), 737 (2020)
  67. ATLAS Collaboration, Search for pair production of third-generation scalar leptoquarks decaying into a top quark and a  $\tau$ -lepton in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector. *JHEP* **06**, 179 (2021)
  68. ATLAS Collaboration, Search for new phenomena in final states with  $b$ -jets and missing transverse momentum in  $\sqrt{s} = 13$  TeV  $pp$  collisions with the ATLAS detector. *JHEP* **05**, 093 (2021)
  69. 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**(11), 112005 (2021)
  70. ATLAS Collaboration, Search for leptoquarks decaying into the  $b\tau$  final state in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector (2023)
  71. ATLAS Collaboration, Search for pair production of third-generation leptoquarks decaying into a bottom quark and a  $\tau$ -lepton with the atlas detector (2023)
  72. D.A. Faroughy, A. Greljo, J.F. Kamenik, Confronting lepton flavor universality violation in B decays with high- $p_T$  tau lepton searches at LHC. *Phys. Lett. B* **764**, 126–134 (2017)
  73. A. Angelescu, D. Bećirević, D.A. Faroughy, O. Sumensari, Closing the window on single leptoquark solutions to the  $B$ -physics anomalies. *JHEP* **10**, 183 (2018)
  74. A. Bhaskar, D. Das, T. Mandal, S. Mitra, C. Neeraj, Precise limits on the charge-2/3 U1 vector leptoquark. *Phys. Rev. D* **104**(3), 035016 (2021)
  75. C. Cornellà, D.A. Faroughy, J. Fuentes-Martín, G. Isidori, M. Neubert, Reading the footprints of the B-meson flavor anomalies. *JHEP* **08**, 050 (2021)
  76. L. Allwicher, D.A. Faroughy, F. Jaffredo, O. Sumensari, F. Wilsch, Drell-Yan tails beyond the standard model. *JHEP* **3**, 2023 (2023)
  77. U. Haisch, L. Schnell, S. Schulte, Drell-Yan production in third-generation gauge vector leptoquark models at NLO+PS in QCD. *JHEP* **2**, 2023 (2023)
  78. J. H. Friedman, Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **29**(5), 1189–1232 (2001). Publisher: Institute of Mathematical Statistics
  79. X. Ai, S.C. Hsu, K. Li, C.T. Lu, Probing highly collimated photon-jets with deep learning. *JPCS* **2438**(1), 012114 (2023)
  80. ATLAS Collaboration, Search for the standard model Higgs boson produced in association with top quarks and decaying into a  $b\bar{b}$  pair in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector. *Phys. Rev. D* **97**(7), 072016 (2018)
  81. S. Chigusa, S. Li, Y. Nakai, W. Zhang, Y. Zhang, J. Zheng, Deeply learned preselection of Higgs dijet decays at future lepton colliders. *Phys. Lett. B* **833**, 137301 (2022)
  82. Y.L. Chung, S.C. Hsu, B. Nachman, Disentangling boosted Higgs boson production modes with machine learning. *JINST* **16**, P07002 (2021)
  83. J. Feng, M. Li, Q.S. Yan, Y.P. Zeng, H.H. Zhang, Y. Zhang, Z. Zhao, Improving heavy Dirac neutrino prospects at future hadron colliders using machine learning. *JHEP* **9**, 2022 (2022)
  84. D. Barbosa, F. Díaz, L. Quintero, A. Flórez, M. Sanchez, A. Gurrola, E. Sheridan, and F. Romeo, Probing a  $Z'$  with non-universal fermion couplings through top quark fusion, decays to bottom quarks, and machine learning techniques. *EPJC* **83**(413) (2023)
  85. N. Assad, B. Fornal, B. Grinstein, Baryon Number and Lepton Universality Violation in Leptoquark and Diquark Models. *Phys. Lett. B* **777**, 324–331 (2018)
  86. L. Calibbi, A. Crivellin, T. Li, Model of vector leptoquarks in view of the  $B$ -physics anomalies. *Phys. Rev. D* **98**(11), 115002 (2018)
  87. M. Blanke, A. Crivellin,  $B$  Meson anomalies in a Pati-Salam model within the Randall-Sundrum background. *Phys. Rev. Lett.* **121**(1), 011801 (2018)
  88. S. Iguro, J. Kawamura, S. Okawa, Y. Omura, TeV-scale vector leptoquark from Pati-Salam unification with vectorlike families. *Phys. Rev. D* **104**(7), 075008 (2021)
  89. L. Di Luzio, A. Greljo, M. Nardecchia, Gauge leptoquark as the origin of  $B$ -physics anomalies. *Phys. Rev. D* **96**(11), 115011 (2017)
  90. A. Greljo, B.A. Stefanek, Third family quark-lepton unification at the TeV scale. *Phys. Lett. B* **782**, 131–138 (2018)
  91. L. Di Luzio, J. Fuentes-Martín, A. Greljo, M. Nardecchia, S. Renner, Maximal flavour violation: a Cabibbo mechanism for leptoquarks. *JHEP* **11**, 081 (2018)
  92. S.F. King, Twin Pati-Salam theory of flavour with a TeV scale vector leptoquark. *JHEP* **11**, 161 (2021)
  93. M. Fernández-Navarro, S.F. King, B-anomalies in a twin Pati-Salam theory of flavour including the 2022 LHCb  $R_{K^{(*)}}$  analysis. *JHEP* **2**, 2023 (2023)
  94. M. Bordone, C. Cornellà, J. Fuentes-Martín, G. Isidori, A three-site gauge model for flavor hierarchies and flavor anomalies. *Phys. Lett. B* **779**, 317–323 (2018)
  95. M. Bordone, C. Cornellà, J. Fuentes-Martín, G. Isidori, Low-energy signatures of the  $PS^3$  model: from  $B$ -physics anomalies to LFV. *JHEP* **10**, 148 (2018)
  96. J. Fuentes-Martín, G. Isidori, J.M. Lizana, N. Selimovic, B.A. Stefanek, Flavor hierarchies, flavor anomalies, and Higgs mass from a warped extra dimension. *Phys. Lett. B* **834**, 137382 (2022)
  97. B. Gripaios, Composite leptoquarks at the LHC. *JHEP* **02**, 045 (2010)
  98. R. Barbieri, C.W. Murphy, F. Senia,  $B$ -decay anomalies in a composite leptoquark model. *Eur. Phys. J. C* **77**(1), 8 (2017)
  99. R. Barbieri, A. Tesi,  $B$ -decay anomalies in Pati-Salam SU(4). *Eur. Phys. J. C* **78**(3), 193 (2018)
  100. J. Fuentes-Martín, G. Isidori, M. König, N. Selimović, Vector leptoquarks beyond tree level. III. Vectorlike fermions and flavor-changing transitions. *Phys. Rev. D* **102**, 115015 (2020)
  101. A. Djouadi, J. Ellis, A. Popov, J. Quevillon, Interference effects in  $t\bar{t}$  production at the LHC as a window on new physics. *JHEP* **03**, 119 (2019)
  102. N.D. Christensen, C. Duhr, FeynRules - Feynman rules made easy. *Comput. Phys. Commun.* **180**, 1614–1641 (2009)
  103. A. Alloul, N.D. Christensen, C. Degrande, C. Duhr, B. Fuks, FeynRules 2.0 - A complete toolbox for tree-level phenomenology. *Comput. Phys. Commun.* **185**, 2250–2300 (2014)
  104. J. Alwall, C. Duhr, B. Fuks, O. Mattelaer, D.G. Öztürk, C.H. Shen, Computing decay rates for new physics theories with FeynRules

- and MadGraph 5\_aMC@NLO. *Comput. Phys. Commun.* **197**, 312–323 (2015)
105. J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.S. Shao, T. Stelzer, P. Torrielli, M. Zaro, The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP* **07**, 079 (2014)
  106. NNPDF Collaboration. Parton distributions for the LHC Run II. *JHEP* **04**, 040 (2015)
  107. T. Sjöstrand, S. Ask, J.R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C.O. Rasmussen, P.Z. Skands, An introduction to PYTHIA 8.2. *Comput. Phys. Commun.* **191**, 159–177 (2015)
  108. DELPHES 3 Collaboration. DELPHES 3, A modular framework for fast simulation of a generic collider experiment. *JHEP* **02**, 057 (2014)
  109. C.M.S. Collaboration, Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *JINST* **13**(04), P05011 (2018)
  110. C.M.S. Collaboration, Identification of hadronic tau lepton decays using a deep neural network. *JINST* **17**(07), P07023 (2022)
  111. CMS Collaboration. Study of the Discovery Reach in Searches for Supersymmetry at CMS with 3000/fb. CMS Physics Analysis Summary CMS-PAS-FTR-13-014, CERN, 2013. Accessed on January 07, (2023)
  112. A.M. Sirunyan et al., Performance of the reconstruction and identification of high-momentum muons in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *JINST* **15**(02), P02027 (2020)
  113. A.M. Sirunyan et al., Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. *JINST* **16**(05), P05014 (2021)
  114. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011)
  115. T. Chen and C. Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’16, pages 785–794, New York, NY, USA, 2016. Association for Computing Machinery
  116. The ROOT team. Root - an object oriented data analysis framework. *Nucl. Inst. & Meth. in Phys. Res. A*, 389:81–86, 1997. Proceedings AIHENP’96 Workshop, Lausanne, Sep. 1996. See also “ROOT”[software], Release v6.18/02, and <https://root.cern/manual/root/>
  117. D. Barbosa, F. Díaz, L. Quintero, A. Flórez, M. Sanchez, A. Gurrola, E. Sheridan, F. Romeo, Probing a  $Z'$  with non-universal fermion couplings through top quark fusion, decays to bottom quarks, and machine learning techniques. *Eur. Phys. J. C* **83**(5), 413 (2023)
  118. A. Flórez, A. Gurrola, W. Johns, P. Sheldon, E. Sheridan, K. Sinha, B. Soubasis, Probing axionlike particles with  $\gamma\gamma$  final states from vector boson fusion processes at the LHC. *Phys. Rev. D* **103**(9), 095001 (2021)
  119. A. Flórez, A. Gurrola, W. Johns, J. Maruri, P. Sheldon, K. Sinha, S.R. Starko, Anapole dark matter via vector boson fusion processes at the LHC. *Phys. Rev. D* **100**(1), 016017 (2019)
  120. A. Flórez, Y. Guo, A. Gurrola, W. Johns, O. Ray, P. Sheldon, S. Starko, Probing heavy spin-2 bosons with  $\gamma\gamma$  final states from vector boson fusion processes at the LHC. *Phys. Rev. D* **99**(3), 035034 (2019)
  121. A. Flórez, K. Gui, A. Gurrola, C. Patiño, D. Restrepo, Expanding the reach of heavy neutrino searches at the LHC. *Phys. Lett. B* **778**, 94–100 (2018)
  122. A. Flórez, A. Gurrola, W. Johns, Y. Do Oh, P. Sheldon, D. Teague, T. Weiler, Searching for new heavy neutral gauge bosons using vector boson fusion processes at the LHC. *Phys. Lett. B* **767**, 126–132 (2017)
  123. A. Flórez, L. Bravo, A. Gurrola, C. Ávila, M. Segura, P. Sheldon, W. Johns, Probing the stau-neutralino coannihilation region at the LHC with a soft tau lepton and a jet from initial state radiation. *Phys. Rev. D* **94**(7), 073007 (2016)
  124. B. Dutta, S. Ghosh, A. Gurrola, D. Julson, T. Kamon, J. Kumar, Probing an MeV-Scale Scalar Boson in Association with a TeV-Scale Top-Quark Partner at the LHC. *JHEP* **03**(164) (2023)
  125. R. Arnowitt, B. Dutta, A. Gurrola, T. Kamon, A. Krislock, D. Toback, Determining the Dark Matter Relic Density in the Minimal Supergravity Stau-Neutralino Coannihilation Region at the Large Hadron Collider. *Phys. Rev. Lett.* **100**, 231802 (2008)
  126. B. Dutta, A. Gurrola, T. Kamon, A. Krislock, A.B. Lahanas, N.E. Mavromatos, D. V. Nanopoulos, Supersymmetry Signals of Super-critical String Cosmology at the Large Hadron Collider. *Phys. Rev. D* **79**, 055002 (2009)
  127. C. Avila, A. Flórez, A. Gurrola, D. Julson, S. Starko, Connecting Particle Physics and Cosmology: Measuring the Dark Matter Relic Density in Compressed Supersymmetry at the LHC. *Phys. Dark Univ.* **27**, 100430 (2020)
  128. B. Dutta, A. Gurrola, W. Johns, T. Kamon, P. Sheldon, K. Sinha, Vector Boson Fusion Processes as a Probe of Supersymmetric Electroweak Sectors at the LHC. *Phys. Rev. D* **87**, 035029 (2013)
  129. A. Delannoy, B. Dutta, A. Gurrola, W. Johns, T. Kamon, E. Luiggi, A. Melo, P. Sheldon, K. Sinha, K. Wang, S. Wu, Probing Dark Matter at the LHC using Vector Boson Fusion Processes. *Phys. Rev. Lett.* **111**, 061801 (2013)
  130. B. Dutta, T. Ghosh, A. Gurrola, W. Johns, T. Kamon, P. Sheldon, K. Sinha, S. Wu, Probing Compressed Sleptons at the LHC using Vector Boson Fusion Processes. *Phys. Rev. D* **91**, 055025 (2015)
  131. B. Dutta, W. Flanagan, A. Gurrola, W. Johns, T. Kamon, P. Sheldon, K. Sinha, K. Wang, S. Wu, Probing Compressed Top Squarks at the LHC at 14 TeV. *Phys. Rev. D* **90**, 095022 (2014)
  132. B. Dutta, W. Flanagan, A. Gurrola, W. Johns, T. Kamon, P. Sheldon, K. Sinha, K. Wang, S. Wu, Probing Compressed Bottom Squarks with Boosted Jets and Shape Analysis. *Phys. Rev. D* **92**, 095009 (2015)
  133. Paul Lujan The Pixel Luminosity Telescope: A detector for luminosity measurement at CMS using silicon pixel sensors. Technical Report CMS-DN-21-008, et al., Accepted by Eur (Phys. J. C., 2022)
  134. J. Butterworth, S. Carrazza, A. Cooper-Sarkar, A. De Roeck, J. Feltesse, S. Forte, J. Gao, S. Glazov, J. Huston, Z. Kassabov, R. McNulty, A. Morsch, P. Nadolsky, V. Radescu, J. Rojo, R. Thorne, PDF4LHC recommendations for LHC Run II. *J. Phys. G* **43**, 023001 (2016)
  135. A.M. Sirunyan et al., Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *JINST* **13**, P05011 (2018)





# Probing light scalars and vector-like quarks at the high-luminosity LHC

U. S. Qureshi<sup>1,a</sup>, A. Gurrola<sup>1,b</sup>, A. Flórez<sup>2,c</sup>, C. Rodriguez<sup>2,d</sup>

<sup>1</sup> Department of Physics and Astronomy, Vanderbilt University, Nashville, TN, USA

<sup>2</sup> Physics Department, Universidad de los Andes, Bogotá, Colombia

Received: 30 October 2024 / Accepted: 15 March 2025

© The Author(s) 2025

**Abstract** A model based on a  $U(1)_{T_R^3}$  extension of the Standard Model can address the mass hierarchy between generations of fermions, explain thermal dark matter abundance, and the muon  $g - 2$ ,  $R_{(D)}$ , and  $R_{(D^*)}$  anomalies. The model contains a light scalar boson  $\phi'$  and a heavy vector-like quark  $\chi_u$  that can be probed at CERN's large hadron collider (LHC). We perform a phenomenology study on the production of  $\phi'$  and  $\chi_u$  particles from proton–proton (pp) collisions at the LHC at  $\sqrt{s} = 13.6$  TeV, primarily through  $g-g$  and  $t-\chi_u$  fusion. We work under a simplified model approach and directly take the  $\chi_u$  and  $\phi'$  masses as free parameters. We perform a phenomenological analysis considering  $\chi_u$  final states to b-quarks, muons, and neutrinos, and  $\phi'$  decays to  $\mu^+\mu^-$ . A machine learning algorithm is used to maximize the signal sensitivity, considering an integrated luminosity of  $3000 \text{ fb}^{-1}$ . The proposed methodology can be a key mode for discovery over a large mass range, including low masses, traditionally considered difficult due to experimental constraints.

## 1 Introduction

The Standard Model (SM) of particle physics, despite its successful account of numerous experimental findings involving strong, electromagnetic, and weak interactions, confirmed by CERN's large hadron collider (LHC) is regarded as a lower-energy manifestation of a more comprehensive theory. This perspective arises from unresolved questions regarding the origins of dark matter, electroweak symmetry breaking scales, lepton flavor universality, the anomalous muon mag-

netic moment [1–8], discrepancies in the  $R_{(D)}$  and  $R_{(D^*)}$  ratios from b-meson decays [9–21], as well as theoretical conundrums about whether gravity should be quantized, how gauge interactions can be unified, and the fine-tuning problem associated with the Higgs boson mass. Furthermore, the SM offers no explanation for fermion family replication nor for the lack of CP violation in the strong sector. These theoretical gaps, coupled with the experimental observation of phenomena such as neutrino masses, dark matter, and the baryon asymmetry in the universe, which cannot be explained by the SM, reinforce the expectation for physics beyond the SM (BSM).

As a result, several theoretical models have been put forth to address the limitations of the SM over the past decade. Despite differing theoretical motivations and resulting implications, a common thread among these ideas is the introduction of new particles, that, depending on the model, might be probed via proton–proton (pp) collisions at the LHC. A myriad of ideas have been suggested to investigate BSM physics, driving a substantial amount of exploration at the LHC. Said research has significantly limited the scope of theories and established exclusion bounds, extending to multi-TeV ranges for the masses of newly predicted particles within certain models [22–28]. Possible reasons for the absence of evidence could be attributed to new particle masses being at the scale where they are too large to be produced at the LHC energies and likely with exceptionally low production rates. In the scenario where the masses of the new particles might be probed at the LHC, a vast amount of data might be needed, together with advanced analysis techniques, to enhance the probability of detection. Alternatively, it is conceivable that new physics diverges from the conventional assumptions made in many BSM theories and the associated explorations. As a result, these new physics phenomena could remain hidden in processes that have not yet been thoroughly examined.

<sup>a</sup> e-mail: umar.sohail.qureshi@vanderbilt.edu

<sup>b</sup> e-mail: alfredo.gurrola@vanderbilt.edu

<sup>c</sup> e-mail: ca.florez@uniandes.edu.co (corresponding author)

<sup>d</sup> e-mail: c.rodriguez45@uniandes.edu.co

Minimal extensions to the SM, considering new  $U(1)_\chi$  symmetry groups, are among the most studied BSM scenarios. For example, the  $U(1)_{T_R^3}$  symmetry, where families of right-handed fermions of the SM and possible extensions, such as right-handed neutrinos, are charged, was originally studied in the context of left-right symmetry models [29–31]. In these studies,  $U(1)_{T_R^3}$  is identified as the subgroup of  $SU(2)_R$  defined by its diagonal (electric-charge neutral) generator,  $T_R^3$ . In addition, it is often suggested that  $U(1)_{T_R^3}$  is a subspecies of a  $U(1)_{B-L}$  symmetry since the breaking of the  $U(1)_{B-L} \times U(1)_{T_R^3}$  leads to the  $U(1)_Y$  symmetry. This naturally motivates the presence of a massive and electrically neutral  $Z'$  gauge boson [32–36]. However, in the breaking of  $U(1)_{B-L} \times U(1)_{T_R^3} \rightarrow U(1)_Y$ , it follows that the Higgs doublet  $H$ , since it is a singlet of  $U(1)_{B-L}$ , acquires its hypercharge by inheritance from a charge under  $U(1)_{T_R^3}$ . Consequently, the vacuum expectation value (VEV) of  $H$  couples both symmetry-breaking scales for  $U(1)_Y$  and  $U(1)_{T_R^3}$ . Alternatively, these symmetry-breaking scales can be decoupled by adding an additional  $U(1)_G$  group where fermions of the SM are singlets and  $H$  is not. Therefore, the hypercharge comes from  $U(1)_G$  for the  $H$  and from  $U(1)_{T_R^3}$  for fermions, i.e.  $Y = Q_{T_R^3} + \frac{1}{2}Q_{B-L} + Q_G$  [37]. Moreover, one can ask for scenarios where the hypercharge is not related to the  $U(1)_{T_R^3}$  charge.

Recently, theoretical and phenomenological efforts have emerged around scenarios where the low-energy gauge symmetry of the SM is extended by appending the Abelian gauge group  $U(1)_{T_R^3}$ , whose spontaneous symmetry-breaking is not linked to the electroweak one [38–43]. In these scenarios, the gauge boson of  $U(1)_{T_R^3}$  is associated with a massive dark photon  $A'$  whose longitudinal mode arises from a Higgs-like mechanism involving a complex scalar field,  $\phi$ . This field is a singlet under the SM group, with its CP-odd component associated with the  $A'$  mass and the CP-even giving rise to a dark Higgs,  $\phi'$ . To cancel gauge anomalies, a right-handed  $\nu_R$  neutrino must be included for each generation of the SM that couples to  $U(1)_{T_R^3}$ . Furthermore, to correctly explain the origin of fermion masses in a UV-complete theory, a set of new vector-like quarks ( $\chi_u, \chi_d, \chi_\ell, \chi_\nu$ ) must be included. These new particles are singlets under  $U(1)_{T_R^3}$  and charged like SM right-handed fermions, as in the universal see-saw mechanism [44–49].

In this phenomenology study, we devise a LHC search strategy for the light GeV-scale scalar boson  $\phi'$  produced in association with a heavy TeV-scale  $\chi_u$ , the partner particle of the top quark, through a previously unexplored production and final state channel. Particularly, we explore the production of  $pp \rightarrow t\chi_u\phi'$ , in contrast to  $pp \rightarrow TT \rightarrow t\phi't\phi'$  with hadronic [50–52] di-photon  $\phi'$  [53, 54] decays. Due to the non-trivial  $\chi - t - \phi'$  coupling, processes where the final state includes  $t\chi_u\phi'$  are allowed in pp colliders through the  $\chi_u - t$

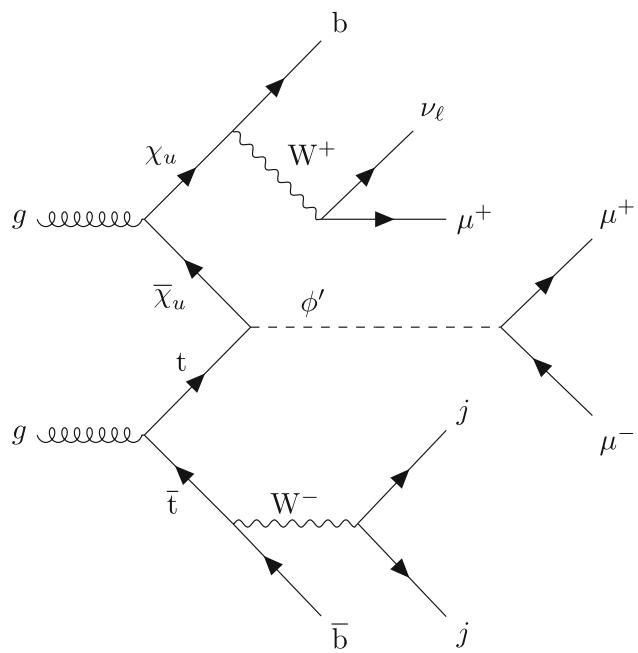
fusion, see Fig. 1. Since the  $\chi_u$  couples to SM quarks and gluons, it can be produced in large quantities. Furthermore, its energetic decay products can be detected alongside the  $\phi'$  mediator particle that has significant transverse momentum. Therefore, if the  $\phi'$  decays into SM particles that are observable in the detector's central region, this strategy can be very effective at reducing the SM background, and thus improve the long-term LHC discovery reach for heavy top partners and GeV-scale mediators, which are typically hard to detect using conventional methods at hadron colliders. Moreover, since it is possible to have  $\chi_u \rightarrow t\phi'$  decays (and  $\bar{\chi}_u \rightarrow \bar{t}\phi'$ ), the same  $pp \rightarrow t\chi_u\phi'$  state may arise from  $\chi_u\bar{\chi}_u$  production diagrams with quantum chromodynamic (QCD) vertices, where one  $\chi_u$  decays to  $t\phi'$ , as shown in Fig. 2. As a consequence, the energetic products from  $\chi_u\bar{\chi}_u$  decays can be readily detected, particularly when they occur alongside a mediator particle that carries substantial transverse momentum, providing greater sensitivity than that of searches where either  $\chi_u$  or  $\phi'$  are considered in isolation.

We probe the scenario where the scalar  $\phi'$  has family non-universal fermion couplings, as was suggested in [39], and thus can address several issues with the SM. We focus on the  $\phi'$  decay to a pair of muons since, at the experimental level, muons generally have high reconstruction and identification efficiencies, which allow for the development of relatively low  $p_T(\mu)$  triggers, and provide clean signatures to remove the copious QCD multijet SM background. A key component of this study is the development of an analysis strategy utilizing a machine learning (ML) algorithm based on Boosted Decision Trees (BDT) [55]. The event classifier's output is employed to conduct a profile-binned likelihood test, which is used to determine the overall signal significance for each model examined in the analysis. The effectiveness of BDTs and other ML algorithms has been validated in numerous experimental and phenomenological studies [36, 56–65]. Our findings indicate that the BDT algorithm significantly enhances signal significance.

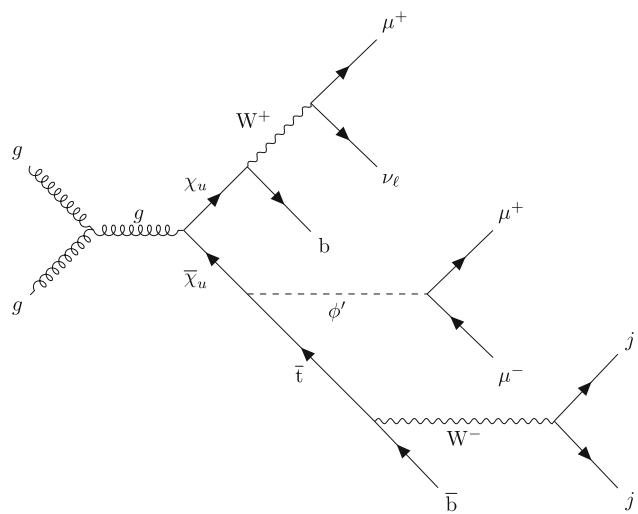
The rest of this paper is structured as follows. Section 3 discusses details of the minimal  $U(1)_{T_R^3}$  model. Section 2 provides an overview of current relevant results at the LHC. Section 4 explains how the Monte Carlo (MC) simulation samples are produced for this study. In Sect. 5 we discuss the motivation and details of our machine learning workflow, and in Sect. 6, the main results are presented. We conclude with a short discussion in Sect. 7.

## 2 Experimental considerations

The ATLAS and CMS collaborations at CERN have conducted various searches for heavy vector-like quarks (T). These searches utilized pp collisions at center-of-mass energies of  $\sqrt{s} = 8$  and 13 TeV. The studies primarily focused



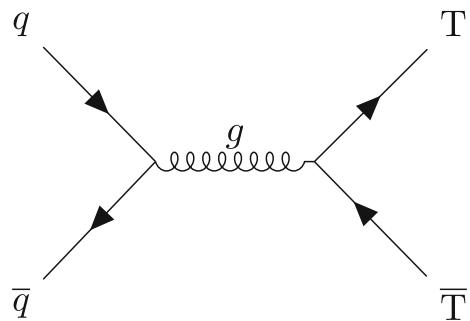
**Fig. 1** Representative Feynman diagram for the production of a  $\phi'$  boson in association with a  $\chi_u$  vector-like quark through the fusion of a top quark and  $\chi_u$  vector-like quark. Once again, the  $\phi'$  decays to a pair of muons, the top quark decays fully hadronically, and the  $\chi_u$  decays semi-leptonically to muons, neutrinos and  $b$ -jets



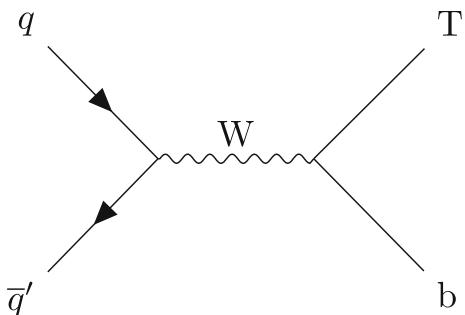
**Fig. 2** Representative Feynman diagram for the production of a  $\phi'$  boson in association with a  $\chi_u$  vector-like quark through the fusion of a gluon pair from incoming protons. The  $\phi'$  decays to a pair of muons, the top quark that decays fully hadronically, and the  $\chi_u$  decay semi-leptonically to muons, neutrinos and jets

on T production through gluon-mediated QCD processes, either in pair production from quark-antiquark annihilation (Fig. 3) or in single-T production from electroweak processes involving associated quarks (Fig. 4).

In those studies, T decays into  $bW$ ,  $tZ$ , or  $tH$  have been considered. In the context of T pair production,  $T\bar{T}$ , via



**Fig. 3** Representative Feynman diagram for T pair production via gluon-mediated QCD processes



**Fig. 4** Representative Feynman diagram for single T production via electroweak processes

QCD processes, the cross sections are well-known and solely depend on the mass of the vector-like quark. Assuming a narrow T decay width ( $\Gamma/m(T) < 0.05$  or 0.1) and a 100% branching fraction to  $bW$ ,  $tZ$ , or  $tH$ , these searches have set stringent bounds on  $m(T)$ , excluding masses below almost 1.5 TeV at 95% confidence level [66–73]. The most recent analysis from the CMS collaboration probes T-quark production via  $pp \rightarrow Tqb$ , in final states with  $T \rightarrow tZ$  or  $T \rightarrow tH$ , considering scenarios with preferential couplings to third-generation fermions. The analysis sets 95% confidence level upper limits of 68–1260 fb on the production cross section, for T masses ranging from 600–1200 GeV [67]. The latest studies from ATLAS probe vector-like quarks using the single-T production mode with the  $T \rightarrow tH$  decay channel leading to a fully hadronic final state [68], the single-T production mode with the  $T \rightarrow tZ$  decay channel leading to a multileptonic final state [69], the  $TT$  pair production mode with various T decay channels leading to multileptonic final states [70], and the  $TT$  pair production mode with various T decay channels leading to a single lepton plus missing momentum final state [71, 72]. The multilepton search offers the greatest sensitivity in most of the phase space, but the missing transverse energy based search has better sensitivity for low branching fraction  $\mathcal{B}(T \rightarrow Wb)$  and high  $\mathcal{B}(T \rightarrow Ht)$ . These searches have similar sensitivities for the singlet and doublet models, resulting in exclusion bounds for masses below about 1.25 TeV and 1.41 TeV, respectively.

A key consideration in the model interpretations summarized above is that the T branching fractions depend on the chosen model. The excluded mass range is less restrictive for specific branching fraction scenarios, such as  $\{\mathcal{B}(T \rightarrow tZ), \mathcal{B}(T \rightarrow bW), \mathcal{B}(T \rightarrow tH)\} = \{0.2, 0.6, 0.2\}$ , excluding masses below about 0.95 TeV. Moreover, if the  $T \rightarrow \phi't$  decay is allowed, or if the branching fractions  $\mathcal{B}(T \rightarrow tH/bW)$  are lower, the limits previously quoted must be re-evaluated. The authors of Ref. [74] emphasize that bounds on  $m(T)$  can be around 500 GeV when  $T \rightarrow t\phi'$  decays are permitted. Therefore, to facilitate a comprehensive study, benchmark scenarios in this paper are considered down to  $m(\chi_u) = 500$  GeV.

### 3 The minimal $U(1)_{T_R^3}$ model

#### 3.1 Scalar potential

In this model, the SM is extended by the Abelian gauge symmetry  $U(1)_{T_R^3}$ , where only right-handed fermions are charged. We assume two independent Higgs mechanisms, one with a Higgs doublet  $H$  for electroweak symmetry breaking and the other with a Higgs singlet  $\phi$  for the  $U(1)_{T_R^3}$  symmetry breaking. Both scalars have independent vacuum expectation values (VEVs),  $\langle H \rangle = v_h/\sqrt{2}$  and  $\langle \phi \rangle = v_\phi/\sqrt{2}$ , allowing us to express the doublet and singlet Higgs fields, following a Kibble parametrization, as

$$H = \begin{pmatrix} G_+ \\ \frac{1}{\sqrt{2}}(v_h + \rho_0 + iG_0) \end{pmatrix} \quad (1)$$

$$\phi = \frac{1}{\sqrt{2}}(v_\phi + \rho_\phi + iG_\phi). \quad (2)$$

In Eqs. (1) and (2),  $G_\pm$ ,  $G_0$ , and  $G_\phi$  are the Goldstone bosons that allow the SM  $W^\pm$  and Z bosons and the dark photon  $A'$ , associated with the  $U(1)_{T_R^3}$  symmetry, to acquire mass. The  $\rho_h$  and  $\rho_\phi$  are an orthogonal mixture of the SM Higgs boson and the dark Higgs

$$\begin{pmatrix} h \\ \phi' \end{pmatrix} = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} \rho_0 \\ \rho_\phi \end{pmatrix}, \quad (3)$$

that results from the diagonalization of the mass matrices arising from the gauge invariant potential

$$\begin{aligned} \mathcal{V}(\phi, H) = & \mu_H^2 H^\dagger H + \mu_\phi^2 \phi^* \phi \\ & + \lambda (H^\dagger H)(\phi^* \phi) + \lambda_H (H^\dagger H)^2 + \lambda_\phi (\phi^* \phi)^2. \end{aligned} \quad (4)$$

The tadpole equations are given from the minimization of the potential as

$$\frac{\partial \mathcal{V}}{\partial H} = \frac{v_h}{\sqrt{2}} \left( \mu_H^2 + \lambda_H v_h^2 + \frac{1}{2} \lambda v_\phi^2 \right) = 0, \quad (5)$$

$$\frac{\partial \mathcal{V}}{\partial \phi} = \frac{v_\phi}{\sqrt{2}} \left( \mu_\phi^2 + \lambda_\phi v_\phi^2 + \frac{1}{2} \lambda v_h^2 \right) = 0. \quad (6)$$

The masses of the scalar bosons can be written as

$$\begin{aligned} m_{h,\phi'}^2 = & \frac{1}{2} \left( \lambda_H v_h^2 + \lambda_\phi v_\phi^2 \right) \\ & \pm \sqrt{\lambda^2 v_h^2 v_\phi^2 + \left( \lambda_H v_h^2 - \lambda_\phi v_\phi^2 \right)^2}, \end{aligned} \quad (7)$$

and the mixing angle  $\alpha$  as

$$\tan \alpha = \frac{-\lambda v_h v_\phi}{\lambda_H v_h^2 - \lambda_\phi v_\phi^2 - \sqrt{\lambda^2 v_h^2 v_\phi^2 + \left( \lambda_H v_h^2 - \lambda_\phi v_\phi^2 \right)^2}}. \quad (8)$$

#### 3.2 The universal seesaw mechanism

In the model, each electrically charged SM fermion  $f$  has a mass protected by both VEVs. In turn, they acquire mass from the mixture with a vector-like fermion  $\chi_f$ , which is charged as the right-handed component of the respective SM fermion, in a UV complete theory. The terms in the Lagrangian density that contribute to the mass of physical fermions are,

$$\begin{aligned} -\mathcal{L} \supset & Y_{f_L} \bar{f}'_L \chi'_R H + Y_{f_R} \bar{\chi}'_{fL} f'_R \phi^* + m_{\chi'_f} \bar{\chi}'_{fL} \chi'_{fR} \\ & + \text{h.c.} \end{aligned} \quad (9)$$

Therefore, in the vacuum, the mass matrix is

$$M_f = \begin{pmatrix} 0 & Y_{f_L} v_h / \sqrt{2} \\ Y_{f_R} v_\phi / \sqrt{2} & m_{\chi'_f} \end{pmatrix}. \quad (10)$$

The left- and right-handed components of the physical fermions ( $f$ ,  $\chi_f$ ) are given by two rotations  $\mathcal{R}(\theta_{f_{L,R}})$  as,

$$\begin{pmatrix} f_{L,R} \\ \chi_{f_{L,R}} \end{pmatrix} = \begin{pmatrix} \pm \cos \theta_{f_{L,R}} & \mp \sin \theta_{f_{L,R}} \\ \sin \theta_{f_{L,R}} & \cos \theta_{f_{L,R}} \end{pmatrix} \begin{pmatrix} f'_{L,R} \\ \chi'_{f_{L,R}} \end{pmatrix}, \quad (11)$$

in a way that  $\mathcal{R}(\theta_{f_L}) M_f \mathcal{R}^{-1}(\theta_{f_R}) = \text{diag}(m_f, m_{\chi_f})$  up to a phase. Assuming real parameters, the physical masses and the mixing angles are given by

$$m_f m_{\chi_f} = \frac{(Y_{f_L} v_h)(Y_{f_R} v_\phi)}{2}, \quad (12)$$

$$m_f^2 + m_{\chi_f}^2 = m_{\chi'_f}^2 + \frac{1}{2} \left( Y_{f_L}^2 v_h^2 + Y_{f_R}^2 v_\phi^2 \right), \quad (13)$$

$$\tan \theta_{f_{L,R}} = \frac{\sqrt{2}}{m_{\chi'_f}} \left( \frac{Y_{f_{L,R}} v_{h,\phi}}{2} - \frac{m_f^2}{Y_{f_{L,R}} v_{h,\phi}} \right). \quad (14)$$

The Yukawa interactions of the physical fermions with the scalar bosons have the form

$$-\mathcal{L}_{\text{yuk}} = h \bar{\psi}_{f_L} \mathcal{Y}_h \psi_{f_R} + \phi' \bar{\psi}_{f_L} \mathcal{Y}_\phi \psi_{f_R}, \quad (15)$$

with  $\psi_f = (f, \chi_f)^T$ , and the matrices  $\mathcal{Y}_{f_{L,R}}$  given by

$$\mathcal{Y}_h = \frac{1}{\sqrt{2}} \mathcal{R}(\theta_{f_L}) (Y_{f_L} \sigma_+ \cos \alpha - Y_{f_R} \sigma_- \sin \alpha) \mathcal{R}^{-1}(\theta_{f_R}) \quad (16)$$

$$\mathcal{Y}_\phi = \frac{1}{\sqrt{2}} \mathcal{R}(\theta_{f_L}) (Y_{f_L} \sigma_+ \sin \alpha + Y_{f_R} \sigma_- \cos \alpha) \mathcal{R}^{-1}(\theta_{f_R}), \quad (17)$$

where  $\sigma_\pm = (\sigma_1 \pm i \sigma_2)/2$  are the ladder Pauli matrices.

### 3.3 Minimal UV-complete theory

The model must provide non-zero masses for all the SM fermions and be free of gauge anomalies. So, we must have at least one full generation of vector-like fermions  $\{\chi_u, \chi_d, \chi_\ell, \chi_v\}$  and the right-handed component of the SM neutrinos,  $\nu_R$ , charged as shown in Table 1. Therefore, the Yukawa interactions in the UV-complete theory must be of the form

$$\begin{aligned} -\mathcal{L} \supset & Y_{Lu}^i \bar{q}_L^i \chi'_{uR} \tilde{H} + Y_{Ru}^i \bar{\chi}'_{uL} u_R^i \phi^* + m_{\chi_u} \bar{\chi}'_{uL} \chi'_{uR} \\ & + Y_{Ld}^i \bar{q}_L^i \chi'_{dR} H + Y_{Rd}^i \bar{\chi}'_{dL} d_R^i \phi + m_{\chi_d} \bar{\chi}'_{dL} \chi'_{dR} \\ & + Y_{L\ell}^i \bar{\ell}_L^i \chi'_{\ell R} H + Y_{R\ell}^i \bar{\chi}'_{\ell L} \ell_R^i \phi + m_{\chi_\ell} \bar{\chi}'_{\ell L} \chi'_{\ell R} \\ & + Y_{Lv}^i \bar{\ell}_L^i \chi'_{vR} \tilde{H} + Y_{Rv}^i \bar{\chi}'_{vL} \nu_R^i \phi^* + m_{\chi_v} \bar{\chi}'_{vL} \chi'_{vR} \\ & + \text{h.c.}, \end{aligned} \quad (18)$$

where the  $i$  index runs over the three generations of fermions. The simultaneous diagonalization of the mass matrices of each fermion sector will have a similar structure to the one presented in Eqs. (12) and (13) and the Yukawa matrices will have a similar structure of Eqs. (16) and (17) but codifying the  $CKM$  matrix. For the neutrino sector, the structure of the mass matrix will be more complex due to the presence of the additional Majorana mass term for the vector-like neutrino  $\chi'_v$ .

**Table 1** Minimal field content of the model and their representations under the SM and  $U(1)_{T_R^3}$  gauge groups

Field	$SU(3)_C$	$SU(2)_L$	$U(1)_Y$	$U(1)_{T_R^3}$
$q'_L$	<b>3</b>	<b>2</b>	1/6	0
$\ell'_L$	<b>1</b>	<b>2</b>	-1/2	0
$H$	<b>1</b>	<b>2</b>	1/2	0
$u'^c_R$	<b>3</b>	<b>1</b>	-2/3	-2
$d'^c_R$	<b>3</b>	<b>1</b>	1/3	2
$\ell'^c_R$	<b>1</b>	<b>1</b>	1	2
$\nu'^c_R$	<b>1</b>	<b>1</b>	0	-2
$\phi$	<b>1</b>	<b>1</b>	0	2
$\chi'_{uL}$	<b>3</b>	<b>1</b>	2/3	0
$\chi'^c_{uR}$	<b>3</b>	<b>1</b>	-2/3	0
$\chi'_{dL}$	<b>3</b>	<b>1</b>	-1/3	0
$\chi'^c_{dR}$	<b>3</b>	<b>1</b>	1/3	0
$\chi'_{\ell L}$	<b>1</b>	<b>1</b>	-1	0
$\chi'^c_{\ell R}$	<b>1</b>	<b>1</b>	1	0
$\chi'_{vL}$	<b>1</b>	<b>1</b>	0	0
$\chi'^c_{vR}$	<b>1</b>	<b>1</b>	0	0

The numbers in bold indicate the dimensional representation under  $SU(2)_L$  and  $SU(3)_C$  in which the field transforms

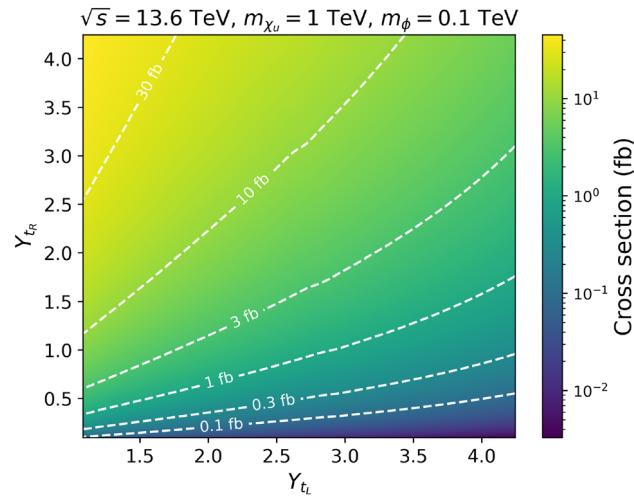
## 4 Samples and simulation

The minimal  $U(1)_{T_R^3}$  model described in Sect. 3 is implemented into the `FeynRules` package [75], which generates the Feynman rules and exports them into a Universal `FeynRules Output` (`UFO`) [76]. The resulting `UFO` is utilized as input for a generator to produce the MC samples. Both signal and background events are generated with the `MadGraph5_aMC@NLO` v3.2.0 program [77, 78] at leading order (LO) in QCD, considering pp beams colliding with a center-of-mass energy of  $\sqrt{s} = 13.6$  TeV. Each signal and background sample is generated separately, with no interference effects between the signal and background considered. The impact of these interference effects has been evaluated, and for all values of  $\chi_u$  and  $\phi'$  masses considered, the effect on the signal plus background cross section is found to be less than  $< 0.5\%$ . Additionally, the effect on the shape of the b-jet  $p_T$  distribution is less than 6% for  $p_T < 300$  GeV and less than 2% for b-jet  $p_T > 300$  GeV. We use the `NNPDF3.0 NLO` [79] set for parton distribution functions (PDFs) for all event generation. Parton-level events are then interfaced with `PYTHIA` (v8.2.44) [80] to account for parton showering and hadronization processes. Finally, we use `DELPHES` (v3.4.2) [81] to simulate smearing and other detector effects using the CMS detector geometric configurations and parameters for particle identification and reconstruction, using the CMS input card with 140 average pileup interactions. All signal cross sections used

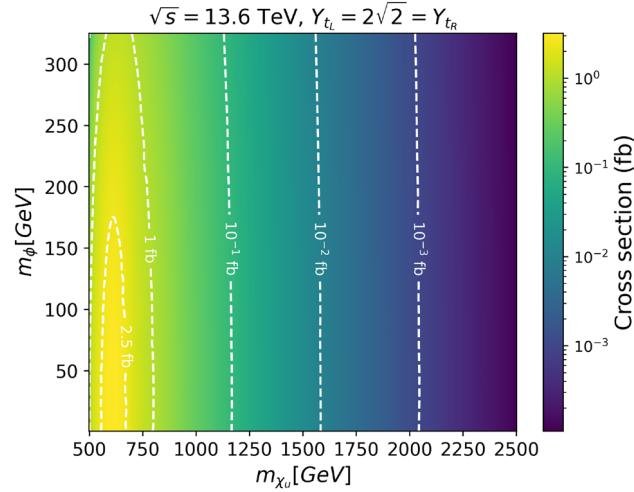
in this analysis are obtained requiring the following kinematic criteria on leptons ( $\ell$ ), b quarks, and light-quark/gluon jets ( $j$ ) at parton level in MadGraph:  $p_T(\ell) > 35$  GeV,  $|\eta(b)| < 2.5$ ,  $|\eta(\ell)| < 2.3$ ,  $p_T(j) > 20$  GeV, and  $|\eta(j)| < 5$ . These parton-level selections were applied exclusively to the signal processes to restrict event generation to the relevant phase space regions. For background processes, these default parton level requirements in MadGraph were imposed:  $p_T(\ell) > 10$  GeV,  $|\eta(\ell)| < 2.5$ ,  $p_T(j) > 20$  GeV,  $|\eta(j)| < 5$ , and  $|\eta(b)| < 5$ . This ensures that the phase space regions for the background near the analysis-level selection criteria are adequately described after parton showering since the pre-selections at the analysis level are more stringent than the parton-level requirements. Furthermore, we use the MLM algorithm for jet matching and jet merging. The parameters  $x_{q\bar{q}cut}$  and  $qcut$  of the MLM algorithm are set to 30 and 45 respectively to ensure continuity of the differential jet rate as a function of jet multiplicity. Each simulated signal and background sample is produced separately at LO, with one million events at the generation level, neglecting potential interference effects between the signal and background due to the suppression caused by the different orders of magnitude in the coupling constants of the signal and background.

Signal samples are generated considering the production of a  $\phi'$  boson, an associated  $\chi_u$  vector-like quark, and a top quark ( $pp \rightarrow \chi_u t \phi'$ ), inclusive in both  $\alpha$  and  $\alpha_s$  (see Figs. 1 and 2). We have used the implementation of the  $U(1)_{T_R^3}$  model in Ref. [43]. Signal samples were created considering coupling values of  $Y_{t_R} = Y_L = 2\sqrt{2}$  in the range of masses  $m(\phi') \in \{5, 10, 50, 100, 325\}$  GeV for the dark higgs and  $m(\chi_u) \in \{0.50, 0.75, 1.0, 1.5, 2.0, 2.5\}$  TeV for the vector-like quark  $\chi_u$  [82]. The production cross section for  $pp \rightarrow \chi_u t \phi'$  is highly dependent on the choice of the Yukawa couplings in the Lagrangian. The  $\chi_u-t$  fusion process shown in Fig. 1 is dominated by the  $Y_{t_R}$  coupling. However, the decay  $\chi_u \rightarrow t \phi'$  shown in Fig. 2 is inversely proportional to the  $Y_{t_L}$  coupling. This effect is shown in Fig. 5, which displays the total signal cross section, as a function of  $Y_{t_R}$  and  $Y_{t_L}$ , for a benchmark point with  $m(\phi') = 100$  GeV and  $m(\chi_u) = 1.0$  TeV.

We target signal events where the top quark decays hadronically into a bottom quark and two jets ( $t \rightarrow bW \rightarrow bqq\bar{q}'$ ), the  $\chi_u$  decays semileptonically into a  $b$  quark, lepton, and neutrino (via  $\chi_u \rightarrow bW$  and  $W \rightarrow \mu\nu_\mu$ ), and the  $\phi'$  produces two muons. We note that the scalar  $\phi'$  particle could result from the mixture of the SM Higgs boson and additional scalar fields, and the Yukawas of the fermions could additionally arise from the mixing of the SM fermions with additional copies of the associated vector-like fermions. Therefore, the  $\phi'$  branching ratios are dependent on the chosen mechanism and model by which this mixture occurs, see for example, Refs. [83–86]. For the purpose of this work, and similar to Refs. [39, 43], the considered bench-



**Fig. 5** Signal production cross section,  $pp \rightarrow \chi_u t \phi'$ , in the  $Y_{t_R}$  versus  $Y_{t_L}$  plane, for a benchmark point with  $m(\phi') = 100$  GeV and  $m(\chi_u) = 1.00$  TeV. The white-dashed contours show specific cross section values in the two dimensional plane



**Fig. 6** Projected cross section (fb) plot for  $pp \rightarrow t \chi_u \phi'$  and subsequent decay as a function of  $m(\chi_u)$  and  $m(\phi')$

mark signal scenarios have  $\mathcal{B}(\chi_u \rightarrow b W)$  of about 0.5 and  $\mathcal{B}(\phi' \rightarrow \mu^+ \mu^-) = 0.98$ . Figure 6 shows the production cross section in fb, as a function of  $m(\phi')$  and  $m(\chi_u)$  masses, assuming the aforementioned decays, branching ratios, and couplings.

We note that for the parameter space of focus in this paper, the total mass of the  $t-\chi_u$  system is larger than  $m(\phi')$ , thus the large rest energy of the  $t-\chi_u$  system is converted into potentially large momentum values for the  $\phi'$ . Similarly, the  $t$ -quark produced through the  $\chi_u-t$  fusion interaction can also have large momentum values, and thus in some cases the hadronic  $t$  decay products cannot be fully reconstructed independently of each other. This results in three possible  $t$  reconstruction scenarios: a fully merged scenario where the

$W \rightarrow jj$  system and the b quarks are very collimated and reconstructed as a single “fat jet” (henceforth referred to as a FatJet, FJ); a partially merged scenario, where the decay products of the W boson form a single FatJet but the b quark can still be separately identified; and an un-merged scenario where all decay products can be independently identified. Jets are clustered using the anti- $k_t$  algorithm [87] using the FastJet (v3.4.2) [88] package with a distance parameter of  $R = 0.4$  for standard jets and  $R = 0.8$  for fat jet objects. Each scenario has an associated identification efficiency and misidentification rate, which depends on the choice of the boosted  $t/W$  algorithm (our choice of efficiency and misidentification rates is described later).

Based on the above details, the final state of interest in this paper consists of three muons (two from the  $\phi'$  decay and one from the  $\chi_u$  decay), a (possibly boosted) top-tagged system, at least one  $b$ -tagged jet, and large missing transverse momentum ( $\vec{p}_T^{\text{miss}}$ ). For the partially merged and un-merged scenarios, there will be two  $b$  quarks present in the final state (one of which is part of the top tagged system).

We consider background sources from SM processes which can give similar objects in the final state as those expected for signal. Several background sources were considered and studied, such as QCD multijet events, production of vector boson pairs (VV : WW, ZZ, WZ), vector boson triplets (VVV : WWZ, WZZ, ZZZ, WWW), top-quark pairs in association with weak bosons ( $t\bar{t}X$ ), and  $t\bar{t}\bar{t}\bar{t}$  processes. The dominant sources of SM background events are from the  $t\bar{t}X$ , ZZW, and  $t\bar{t}\bar{t}\bar{t}$  processes. The  $t\bar{t}X$  background is primarily associated production of a  $Z/\gamma^*$  from  $t\bar{t}$  fusion processes. The ZZW process becomes a background when one Z decays  $b\bar{b}$ , another Z decays to a pair of muons, and the W decays to a muon and a neutrino. Events from ZZW and  $t\bar{t}\bar{t}\bar{t}$  have been combined, after being weighted by their corresponding production cross section. The combination is presented as the “ $b\bar{b}\mu\mu\nu\nu$ ” background in the remainder of this paper. The  $t\bar{t}X$  process is presented as part of the “ $t\bar{t}\mu^+\mu^-$ ” background. Table 2 shows the production cross sections for the dominant background sources. The rest of the aforementioned background processes do not contribute meaningfully in our context, accounting for  $\ll 1\%$  of the total expected background yield.

The identification of leptons, boosted top quarks, and bottom quarks plays an important role in the ability to identify signal events, the ability to minimize the rate of SM backgrounds, and thus also the discovery reach in the high-luminosity environment of the LHC. It is worth noting that the reconstruction and identification of leptons and the decay products of the top/bottom quarks may be non-trivial at the High-Luminosity LHC (HL-LHC) due to the presence of a potentially large number of secondary pp interactions (pileup). The impact of pileup on the new physics discovery reach, and the importance of pileup mitigation at CMS

**Table 2** A summary of dominant SM backgrounds produced by pp collisions and their cross sections in pb, as computed by MadGraph with  $n = 10^6$  events

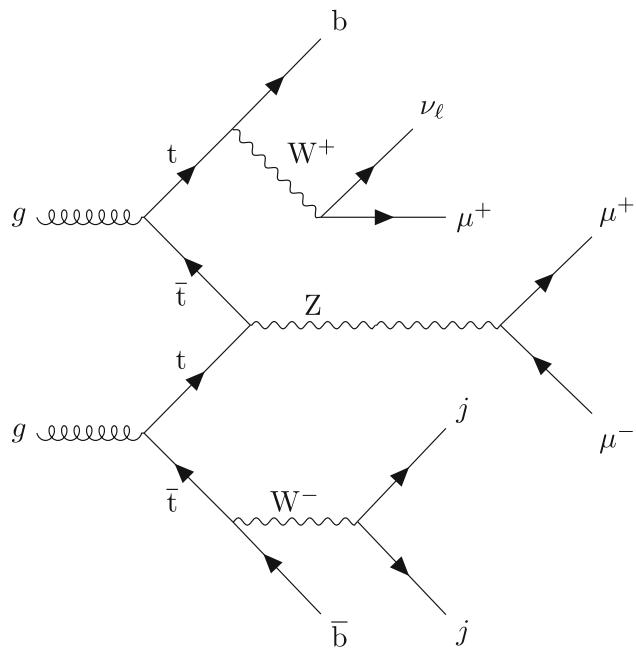
Background process	Cross-section $\sigma$ [pb]
$pp \rightarrow t\bar{t} \mu^+ \mu^-$	$2.574 \times 10^{-3}$
$pp \rightarrow b\bar{b} \mu\mu\nu\nu$	$4.692 \times 10^{-4}$

and ATLAS has been outlined in many papers, for example in Ref. [89]. We note the expected performance of the upgraded ATLAS and CMS detectors for the HL-LHC is beyond the scope of this work; however, the studies presented here do attempt to provide reasonable expectations by conservatively assuming some degradation in lepton and hadron identification efficiencies, using Ref. [89] as a benchmark, and considering the case of 140 average pileup interactions.

For muons with  $|\eta| < 1.5$ , the assumed identification efficiency is 95% with a 0.3% misidentification rate [89, 90]. The performance degrades linearly with  $\eta$  for  $1.5 < |\eta| < 2.5$ , and we assume an identification efficiency of 65% with a 0.5% misidentification rate at  $|\eta| = 2.5$ . Similarly, the charged hadron tracking efficiency, which contributes to the jet clustering algorithm and missing transverse momentum ( $\vec{p}_T^{\text{miss}}$ ) calculation, is 97% for  $1.5 < |\eta| < 2.5$ , and degrades to about 85% at  $|\eta| = 2.5$ . These potential inefficiencies due to the presence of secondary pp interactions contribute to how well the lepton and top kinematics can be reconstructed. Following Refs. [91, 92], we consider the “Loose” working point for the identification of the fully merged (partially merged) t decays, which results in 80–85% top (W) identification efficiency and 11–25% misidentification rate, depending on the FatJet transverse momentum ( $p_T^{FJ}$ ). Following Ref. [93], we consider the “Loose” working point of the DeepCSV algorithm [94], which gives a 70–80% b-tagging efficiency and 10% light quark mis-identification rate. The choice of boosted  $t/W$  and b-tagging working points is determined through an optimization process that maximizes discovery reach. It is noted the contribution from SM backgrounds with a misidentified boosted  $t/W$  is negligible, and thus our discovery projections are not sensitive to uncertainties related to the boosted  $t/W$  misidentification rates (Fig. 7).

## 5 Data analysis using machine learning

The analysis of signal and background events is performed utilizing machine learning techniques. A machine learning-based approach offers sizeable advantages when compared to traditional event classification techniques. Unlike conventional methods, machine learning models have the capability to simultaneously consider all kinematic variables, allowing them to efficiently navigate the complex and



**Fig. 7** Representative Feynman diagram for a background event. A  $Z$  boson is produced in association with a top quark through the fusion of a top, anti top pair from incoming protons. The  $Z$  boson subsequently decays to a pair of muons and the two spectator top quarks decay semi-leptonically and purely hadronically to muons, neutrinos and jets, resulting in the same final states as the signal event

high-dimensional space of event kinematics. Consequently, machine learning models can effectively enact sophisticated selection criteria that take into account the entirety of this high-dimensional space. This makes them ideal for high-energy physics applications.

The BDT method is a powerful machine learning technique that has proven its effectiveness in various applications, particularly in the field of collider physics. In this method, decision trees are trained greedily in a sequential manner, with each tree focusing on learning the discrepancies or residuals between its predictions and the expected values obtained from the previously trained tree. This iterative process aims to progressively minimize errors, making BDTs a particularly effective approach for enhancing model performance.

In the context of collider physics, BDTs have demonstrated their utility in addressing classification problems. In particular, BDTs can effectively discriminate between signal and background events, enabling accurate and efficient event classification. Their ability to handle subtle non-linear relationships within the data with high interpretability makes BDTs a valuable tool to handle large amounts of data with a large number of parameters for each event.

The first step in our workflow involves the use of a specialized *MadAnalysis Expert Mode C++* script [95]. This script extracts essential kinematic and topological information from the simulated samples. The script will process the aforementioned variables contained within these

files and transform them into a structured and informative CSV (Comma-Separated Values) format that can be used to train our machine learning models. These kinematic variables include crucial details about the events, such as particle momenta, energies, and topologies, providing the fundamental building blocks for our machine learning analysis. Figure 13 shows the features that are used for training the machine learning models and their importance for a benchmark point.

To account for the differential significance of various events, we apply cross-section weighting. This ensures that the relative importance of signal and background events is appropriately balanced in the dataset. This weighting is crucial for addressing the varying likelihood of observing different types of events in high-energy physics experiments. The prepared and weighted datasets are then passed to our *MadAnalysis Expert Mode C++* script, where the simulated signal and background events are initially filtered, before being passed to the CSV file for use by the machine learning algorithm. The filtering process requires at least one well-reconstructed and identified  $b$ -jet candidate, at least one jet (regular or FJ) not tagged as a  $b$  jet, and exactly three identified muons. The filtering selections are motivated by experimental constraints, such as the geometric constraints of the CMS/ATLAS detectors, the typical kinematic thresholds for the reconstruction of particle objects, and the available lepton triggers which also drive the minimal kinematic thresholds. Selected jets must have  $p_T > 30 \text{ GeV}$  and  $|\eta(j)| < 5.0$ , while  $b$ -jet candidates with  $p_T > 20 \text{ GeV}$  and  $|\eta(b)| < 2.5$  are chosen. The  $\mu$  object must pass a  $p_T > 35 \text{ GeV}$  threshold and be within a  $|\eta(\ell)| < 2.3$ . We will refer to this filtering criteria as pre-selections. The efficiency of the pre-selections depends on  $m(\phi')$  and  $m(\chi_u)$ , but is typically about 25–30% for the signal samples. Events passing this pre-selection are used as input for the machine learning algorithm, which classifies them as signal or background, using a probability factor.

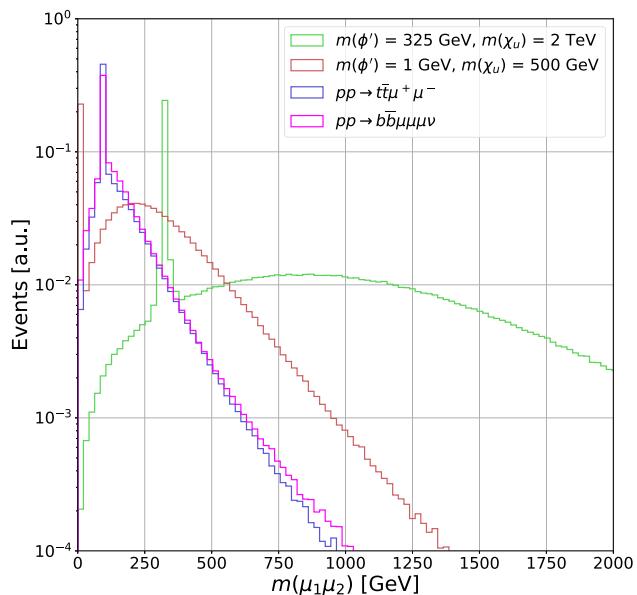
We explore the performance of a diverse set of machine learning models, specifically three neural networks of differing architectures and a BDT algorithm. To ensure robust model assessment, we employed a standard 90–10 train-test split of the dataset, partitioning it into a 90% portion for training and a 10% portion for testing. This division allows us to gauge the generalization capabilities of our models on unseen data.

The training and evaluation of the BDT were carried out in a high-performance computing environment. Specifically, an Nvidia A100 GPU was used. The canonical PyTorch [96] deep learning framework was employed for configuring, training, and evaluating the neural networks. PyTorch is well-regarded for its flexibility and performance in deep learning applications.

For the BDT algorithm, we used hyperparameters  $\eta = 0.3$ ,  $\gamma = 0$ , and  $\text{max\_depth} = 6$ . The XGBoost [97]

**Table 3** Train/test results for the ML models

Model	Train/test acc.	Training time
BDT	N.A./0.9993	6s
Neural network 1	0.9999/0.9997	1 h 58 min
Neural network 2	0.9999/0.9998	2 h 12 min
Neural network 3	0.9999/0.9998	2 h 32 min

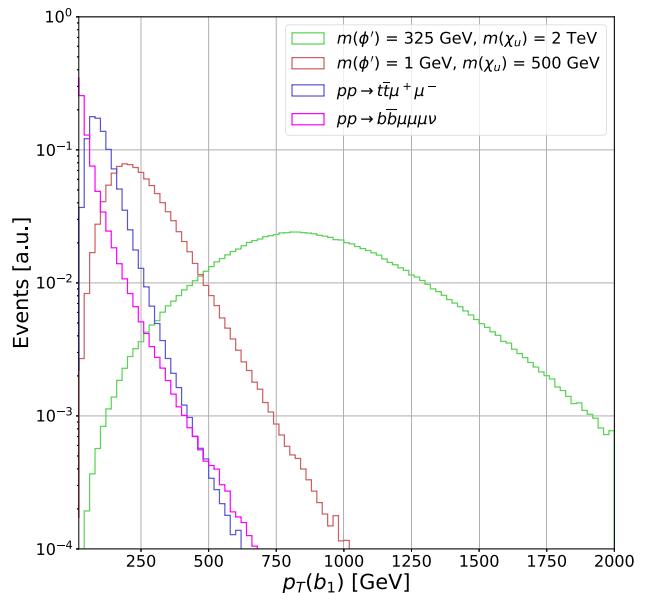
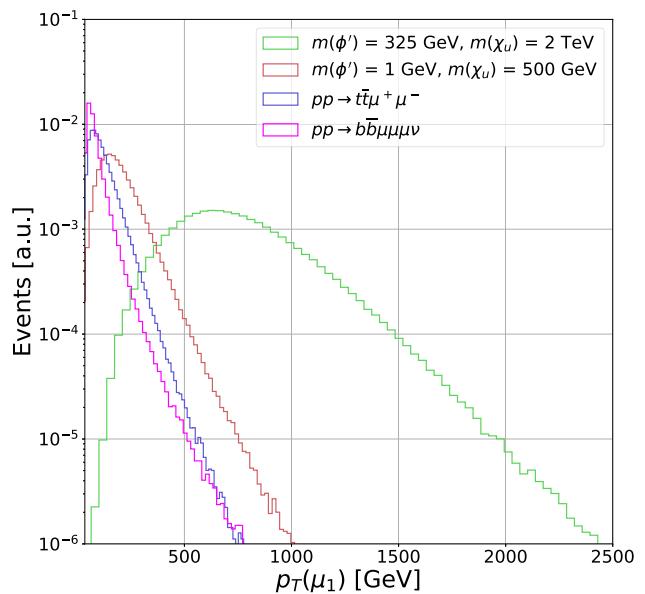
**Fig. 8** Invariant mass distribution of the muon pair with the highest and second highest transverse momentum. The distributions are shown for the two main SM background processes and two signal benchmark points

library was used for the implementation of the Boosted Decision Tree algorithm. It offers high efficiency, optimization, and interpretability, making it a suitable choice for this particular task (Table 3).

It is worth mentioning that we experimented with deep neural networks of various architectures. Although we found that they yield similar signal sensitivity to the BDT, the complex nature of the studies in this work (particle objects considered, experimental constraints in a high luminosity LHC, etc.) motivates the use of a BDT over a deep neural network because of its usefulness, efficiency, and simplicity in understanding the machine learning output in addition to significantly shorter training times. Therefore, we perform our proceeding analysis using the BDT. The outcomes of our model training and evaluation are presented in Table 3.

## 6 Results

Figures 8, 9, and 10 show relevant kinematic distributions for two benchmark signal points and the dominant SM back-

**Fig. 9** Transverse momentum distribution of the leading b-quark jet candidate. The distributions are shown for the two main SM background processes and two signal benchmark points**Fig. 10** Transverse momentum distribution of the leading muon candidate. The distributions are shown for the two main SM background processes and two signal benchmark points

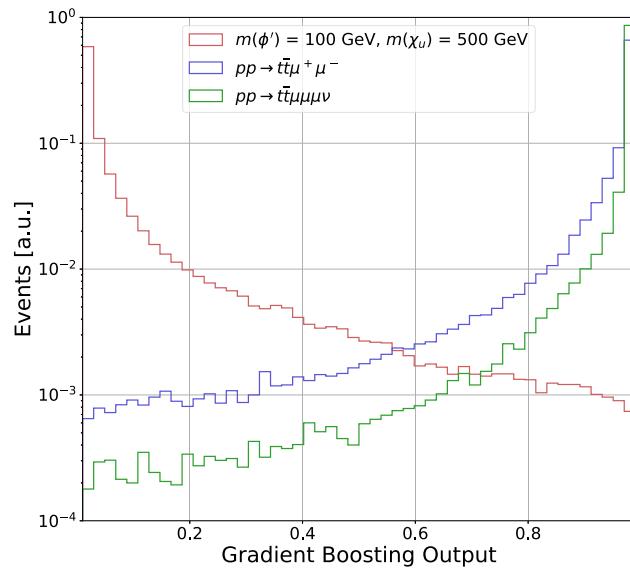
grounds, using the subset of events passing the pre-selections defined above. The signal benchmark points in these figures are  $m(\phi') = 325 \text{ GeV}, m(\chi_u) = 2 \text{ TeV}$ , and  $m(\phi') = 1 \text{ GeV}, m(\chi_u) = 500 \text{ GeV}$ . The distributions are normalized such that the area under the curve is unity. These distributions correspond to the reconstructed mass,  $m(\mu_1, \mu_2)$ , between the two muon candidates with the highest transverse momentum ( $\mu_1$  and  $\mu_2$ ), the transverse momentum of the b-jet can-

dicate with the highest transverse momentum  $p_T$  ( $b_1$ ), and the muon candidate with the highest transverse momentum  $p_T$  ( $\mu_1$ ), respectively. These distributions are among the variables identified by the BDT algorithm with the highest signal to background discrimination power (see Fig. 13).

As can be seen from Fig. 8, the  $\phi'$  mass can be reconstructed through its associated muon decay pair, which is observed as a peak in the  $m(\mu_1, \mu_2)$  distribution around the expected  $m(\phi')$  value, and has low- and high-mass tails which are a consequence of cases where the leading and/or sub-leading muon is not from the  $\phi'$  decay, but rather from the associated W boson from the  $\chi_u$  decay. For the backgrounds, muons come from Z (W) decays. Therefore, the  $m(\mu_1, \mu_2)$  background distributions show a peak near  $m_{W/Z}$ , combined with a broad distribution indicative of the combination of two muon candidates from different decay vertices. We note that the  $\phi' \rightarrow \mu^+ \mu^-$  decay width depends on the square of the  $\phi' \rightarrow \mu^+ \mu^-$  coupling and  $\frac{m_\mu^2}{m(\phi')^2}$  and is thus suppressed by the relatively small muon mass. For the new physics phase space considered in this paper, the  $\phi'$  decay width is less than 1% of the  $\phi'$  resonant mass. Furthermore, as indicated previously, the signal/background interference effects are small and negligible compared to effects from experimental resolution. Therefore, the width of the  $m(\mu_1, \mu_2)$  signal distributions is driven by the experimental resolution in the reconstruction of the muon momenta, as well as the probability that the two leading muons are the correct pair from the  $\phi'$  decay. Since the probability that the two highest- $p_T$  muons are the correct pair from the  $\phi' \rightarrow \mu^+ \mu^-$  decay depends on  $m(\phi')$  and  $m(\chi_u)$ , it is important to include all possible combinations of dimuon pairs (i.e.,  $m(\mu_1, \mu_3)$  and  $m(\mu_2, \mu_3)$ ) in the training of the BDT.

Figure 9 shows the distribution for the b-jet candidate with the highest  $p_T$ ,  $p_T(b_1)$ , for the same simulated samples shown in Fig. 8. Based on the signal topology and our choice of parameter space (i.e.,  $m(\chi_u) > m_t$ ), it is expected that the leading b-jet candidate comes from the  $\chi_u$  decay, with an average  $p_T$  close to  $\frac{m(\chi_u) - m_W}{2}$ , as observed in Fig. 9. For the  $t\bar{t}\mu^+ \mu^-$  background, the b-jet candidates come from top-quark decays. Therefore, their average transverse momentum is expected to be  $\frac{m_t - m_W}{2} \approx 45$  GeV, as observed in Fig. 9. On the other hand, the b-jet candidates for the  $bb\bar{\mu}\mu\mu\nu$  background can come from off-mass-shell  $Z^*/\gamma^*$ , and thus typically have an even softer spectrum in comparison to the  $t\bar{t}\mu^+ \mu^-$  background.

Figure 10 shows the distribution for the muon candidate with the highest  $p_T$ ,  $p_T(\mu_1)$ . Similar to Fig. 9, when  $m(\chi_u) > m_t$  it is expected that the leading muon candidate comes from the  $\chi_u$  decay, with an average  $p_T$  of approximately  $\frac{m(\chi_u) - m_W}{4}$ , as observed in Fig. 10. For the major SM backgrounds, the muon candidates come from Z/W/ $\gamma^*$  decays. Therefore, their average transverse momentum is expected to be much lower,



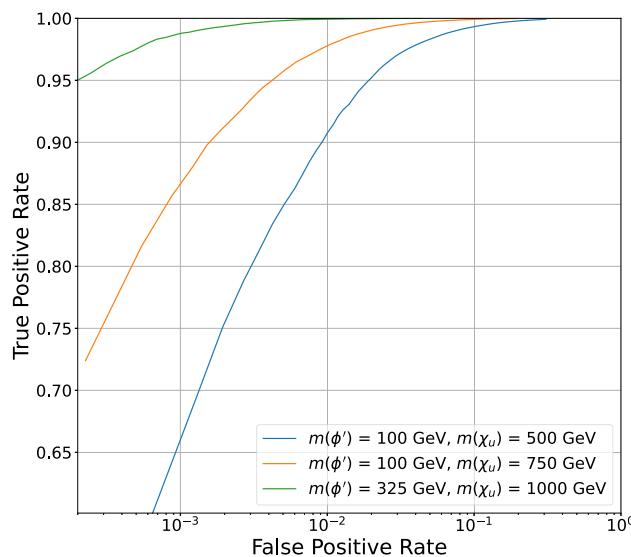
**Fig. 11** Output of the gradient boosting algorithm for a benchmark  $m(\phi') = 100$  GeV and  $m(\chi_u) = 500$  GeV signal, and dominant backgrounds. The distributions are normalized to unity

$\frac{m_{Z/W}}{4} \approx 40 - 45$  GeV. This kinematic feature provides a nice handle to discriminate high  $m(\chi_u)$  signal events amongst the large SM backgrounds, which have lower average  $p_T(\mu)$  constrained by the SM weak boson masses.

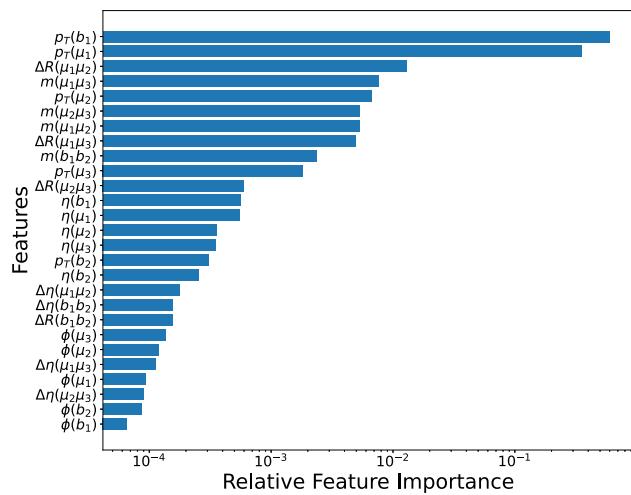
In addition to these aforementioned variables in Figs. 8, 9 and 10, several other kinematic variables were included as inputs to the BDT algorithm. In particular, 27 such variables were used in total, and these included the momenta of b and muon candidates; invariant masses of pairs of muons; angular differences between b jets and between the muons.

As mentioned above, the variables  $m(\mu_i, \mu_j)$  for  $i, j \neq 1$  provide some additional discrimination between signal and background when the leading muons are not a  $\phi'$  decay candidate. The angular separation variables, such as  $\Delta R(\mu_i, \mu_j)$ , are designed to be sensitive to lower mass  $\phi'$ , since the low rest mass of those particles means they acquire more boost, and thus smaller angular separation  $\Delta R$  between the muon candidates. The trained BDT returns the discriminating power of each of its inputs, and the feature importance for each variable is shown in Fig. 13 for a signal benchmark point with  $m(\phi') = 325$  GeV and  $m(\chi_u) = 2000$  GeV.

Figure 11 shows the distributions for the output of the BDT algorithm, normalized to unity, for the representative signal benchmark point of  $m(\phi') = 1$  GeV,  $m(\chi_u) = 0.5$  TeV and the two dominant backgrounds. The output of the BDT algorithm is a value between 0 and 1, which quantifies the likelihood that an event is either background-like (BDT output near 1) or signal-like (BDT output near 0). Figure 12 illustrates the true positive rate (TPR), defined as the probability of correctly selecting signal events using the BDT output, plotted against the false positive rate (FPR), defined as the



**Fig. 12** Receiver operating characteristic curve of the BDT algorithm for three different signal benchmark scenarios



**Fig. 13** Relative importance of features in training for a benchmark signal scenario with  $m(\phi') = 325 \text{ GeV}$  and  $m(\chi_u) = 2000 \text{ GeV}$

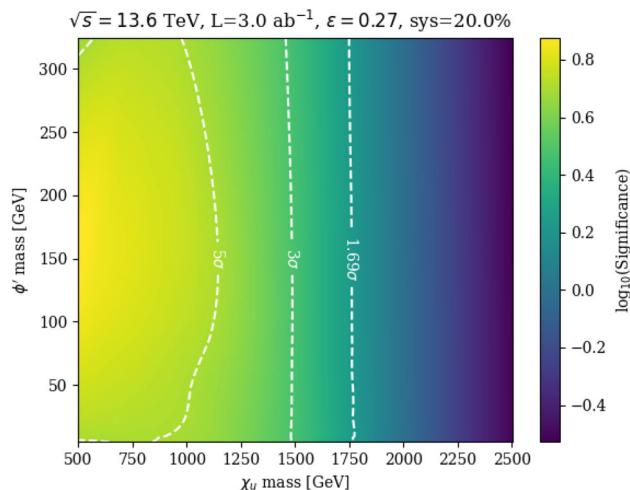
probability of incorrectly selecting background events. For example, for  $m(\phi') = 100 \text{ GeV}$  and  $m(\chi_u) = 500 \text{ GeV}$ , when signal events are selected at 65% probability, the background is selected at about  $10^{-3}$  probability. We note that the primary discriminating feature between the signal and background is the boosted b-jet  $p_T$  coming from the  $\chi_u$  vector-like quark. The  $p_T$  of said b jet increases with  $m(\chi_u)$ , peaking at around  $[m(\chi_u) - m(W)]/2$ . This enhanced boost increases the separation between signal and background, improving the performance of the BDT algorithm as  $m(\chi_u)$  increases (Fig. 13).

The outputs from the BDT machine learning algorithm are used to perform a profile-bin likelihood analysis to estimate the signal significance for a luminosity of  $3000 \text{ fb}^{-1}$ ,

corresponding to the expected amount of collected data by the end of the LHC era. For this purpose, the BDT distributions are normalized to cross section times pre-selection efficiency times luminosity for the different signal models. The significance is then calculated using the expected bin-by-bin yields of the BDT output distribution in a profile likelihood fit, using the ROOTFit [98] package developed by CERN. The expected signal significance  $Z_{\text{sig}}$  is calculated using the probability of obtaining the same test statistic for the signal plus background and the signal-null hypotheses, defined as the local  $p$ -value. Similar to Refs. [99–105], the significance corresponds to the point where the integral of a Gaussian distribution between  $Z_{\text{sig}}$  and  $\infty$  results in a value equal to the local  $p$ -value. The estimation of  $Z_{\text{sig}}$  incorporates systematic uncertainties. The uncertainty values have been included as nuisance parameters, considering lognormal priors for normalization and Gaussian priors for uncertainties associated with the modeling of the shapes similar to Refs. [106, 107].

The systematic uncertainties that have been included result from experimental and theoretical constraints. A 1–5% systematic uncertainty, depending on the simulated MC sample, has been included to account for the choice of Parton Distribution Function (PDF) set. The systematic uncertainty effect was incorporated following the PDF4LHC [98] recommendations. This systematic uncertainty has a small impact on the expected event yields for signal and background, but it does not affect the shape of the BDT output distribution. We additionally considered theoretical uncertainties related to the absence of higher-order contributions to the signal cross sections, which can change the pre-selection efficiencies and the shapes of kinematic variables used as inputs to the BDT algorithm. This uncertainty was calculated by varying the renormalization and factorization scales by  $\times 2$ , and studying the resulting change in the bin-by-bin yields of the BDT distributions. They are found to be at most 2% in a given bin.

Regarding experimental uncertainties, following experimental measurements from CMS on the estimation of the integrated luminosity, a conservative 3% effect has been included [108]. A 5% systematic uncertainty associated with the reconstruction and identification of b-quark jets has been included, independent of  $p_T$  and  $\eta$  of the b-jet candidates. According to Ref. [93], this uncertainty is correlated between signal and background processes with genuine b-jets and is also correlated across BDT bins for each process. For muons, we include a 2% uncertainty associated with the reconstruction, identification, and isolation requirements, and a 3% systematic uncertainty to account for scale and resolution effects on the momentum and energy measurement. We consider jet energy scale uncertainties ranging from 2–5%, contingent on  $\eta$  and  $p_T$ , resulting in shape-based uncertainties on the BDT output distribution. Jet energy scale uncertainties were assumed to range from 1–5%, contingent on  $\eta$  and  $p_T$ . These assumptions lead to shape-based uncertainties on the



**Fig. 14** Signal significance for the high luminosity LHC era, considering with  $3000 \text{ fb}^{-1}$  of collected data

BDT output distribution, varying from 1–2%. Additionally, we include a 10% systematic uncertainty to account for errors in the signal and background predictions. Considering all the various sources of systematic uncertainties, our conservative estimate yields a total effect of about 20%.

Figure 14 shows the expected signal significance considering an integrated luminosity of  $3000 \text{ fb}^{-1}$ . The significance is shown as a heat map in a two-dimensional plane for different  $\phi'$  and  $\chi_u$  masses. The x-axis corresponds to  $m(\chi_u)$ , the y-axis to  $m(\phi')$ , and the heat map to  $\log_{10}(Z_{\text{sig}})$ . The white dashed lines are contours of constant signal significances of  $1.69\sigma$ ,  $3\sigma$  and  $5\sigma$  to represent regions of possible exclusion, evidence of new physics, and discovery, respectively. Under these conditions,  $\phi'(\chi_u)$  masses ranging from 1 to 325 GeV (500 to 1800 GeV) can be probed. The range for a discovery with  $5\sigma$  signal significance varies from  $\chi_u$  masses from  $m(\chi_u) = 770$ – $1100$  GeV, depending  $m(\phi')$ . For large  $m(\chi_u)$ , the significance is almost independent of  $m(\phi')$  because the primary discriminating feature—the boosted  $b$ -quark originating from  $\phi'$ —is driven predominantly by the large  $m(\chi_u)$ , with the kinematic impact of  $m(\phi')$  being relatively negligible.

## 7 Discussion

The LHC will continue to run with pp collisions at  $\sqrt{s} = 13.6$  TeV for the next decade. Given the increase in the integrated luminosity expected from the high-luminosity program, it is important to consider unexplored new physics phase space that diverges from the conventional assumptions made in many BSM theories, and which could have remained hidden in processes that have not yet been thoroughly examined. It is additionally crucial to explore advanced analysis techniques, in particular the use of artificial intelligence

algorithms, to enhance the probability of detecting these rare corners where production cross sections are lower and discrimination from SM backgrounds is difficult.

In this work, we examine a model based on a  $U(1)_{T_R^3}$  extension of the SM, which can address various conceptual and experimental issues with the SM, including the mass hierarchy between generations of fermions, the thermal dark matter abundance, and the muon  $g - 2$ ,  $R_{(D)}$ , and  $R_{(D^*)}$  anomalies. This model contains a light scalar boson  $\phi'$ , with potential masses below the electroweak scale, and TeV-scale vector-like quarks  $\chi_u$ . We consider the scenario where the scalar  $\phi'$  has family non-universal fermion couplings and  $m(\phi') \geq 1$  GeV, as was suggested in Ref. [39], and thus the  $\phi'$  can primarily decay to a pair of muons. Previous works in Refs. [43, 53] considered scenarios motivating a search methodology with a merged diphoton system from  $\phi' \rightarrow \gamma\gamma$  decays. The authors of Ref [43], in which  $m(\phi') < 1$  GeV, indeed pointed out that if the  $\phi'$  is heavier than about 1 GeV, then decays to  $\mu^+\mu^-$  can become the preferable mode for discovery, which is the basis for the work presented in this paper. We further note that the final state topology studied in this paper would represent the most important mode for discovery at  $m(\phi') < 2m_t$  where the  $\phi' \rightarrow t\bar{t}$  decay is kinematically forbidden.

The main result of this paper is that we have shown that the LHC can probe the visible decays of new bosons with masses below the electroweak scale, down to the GeV-scale, by considering the simultaneous production of heavy QCD-coupled particles, which then decay to the SM particles that contain large momentum values and can be observed in the central regions of the CMS and ATLAS detectors. The boosted system combined with innovative machine learning algorithms allows for the signal extraction above the lower-energy SM background. The LHC search strategy described here can be used to discover the prompt decay of new light particles. An important conclusion from this paper is that the detection prospects for low-mass particles are enhanced when it is kinematically possible to simultaneously access the heavy degrees of freedom which arise in the UV completion of the low-energy model. This specific scenario in which the couplings of the light scalars are generationally dependent, with important coupling values to the top quark, is an ideal example which would be difficult to directly probe at low energy beam experiments.

The proposed data analysis represents a competitive alternative to complement searches already being conducted at the LHC, allowing us to probe  $\phi'$  masses from 1 to 325 GeV, for  $m(\chi_u)$  values up to almost 2 TeV, at the HL-LHC. Therefore, we strongly encourage the ATLAS and CMS Collaborations to consider the proposed analysis strategy in future new physics searches.

**Acknowledgements** The authors would like to thank Prof. Joel Jones-Pérez for fruitful discussions. A.F. and C.R. thank the constant and enduring financial support received for this project from the Faculty of Science at Universidad de Los Andes (Bogotá, Colombia) through the projects INV-2023-178-2999 and INV-2023-175-2957. A.G. and U.Q. acknowledge the funding received from the Physics & Astronomy Department at Vanderbilt University and the US National Science Foundation. This work is supported in part by NSF Award PHY-1945366 and a Vanderbilt Seeding Success Grant.

**Data Availability Statement** Data will be made available on reasonable request. [Authors' comment: The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.]

**Code Availability Statement** Code/software will be made available on reasonable request. [Authors' comment: The code/software generated during and/or analysed during the current study is available from the corresponding author on reasonable request.]

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Funded by SCOAP<sup>3</sup>.

## References

1. G.W. Bennett et al., Phys. Rev. D **73**, 072003 (2006). <https://doi.org/10.1103/PhysRevD.73.072003>
2. M. Tanabashi et al., Phys. Rev. D **98**, 030001 (2018). <https://doi.org/10.1103/PhysRevD.98.030001>
3. M. Davier, A. Hoecker, B. Malaescu, Z. Zhang, Eur. Phys. J. C **77**(12), 827 (2017). <https://doi.org/10.1140/epjc/s10052-017-5161-6>
4. M. Davier, A. Hoecker, B. Malaescu, Z. Zhang, Eur. Phys. J. C **80**(3), 241 (2020). <https://doi.org/10.1140/epjc/s10052-020-7792-2>
5. T. Blum, P.A. Boyle, V. Gülpers, T. Izubuchi, L. Jin, C. Jung, A. Jüttner, C. Lehner, A. Portelli, J.T. Tsang, Phys. Rev. Lett. **121**, 022003 (2018). <https://doi.org/10.1103/PhysRevLett.121.022003>
6. A. Keshavarzi, D. Nomura, T. Teubner, Phys. Rev. D **97**, 114025 (2018). <https://doi.org/10.1103/PhysRevD.97.114025>
7. T. Blum, N. Christ, M. Hayakawa, T. Izubuchi, L. Jin, C. Jung, C. Lehner, Phys. Rev. Lett. **124**, 132002 (2020). <https://doi.org/10.1103/PhysRevLett.124.132002>
8. F. Campanario, H. Czyż, J. Gluza, T. Jeliński, G. Rodrigo, S. Tracz, D. Zhuridov, Phys. Rev. D **100**, 076004 (2019). <https://doi.org/10.1103/PhysRevD.100.076004>
9. B. Collaboration, Phys. Rev. Lett. **109**, 101802 (2012). <https://doi.org/10.1103/PhysRevLett.109.101802>
10. B. Collaboration, Phys. Rev. D **88**(7), 072012 (2013). <https://doi.org/10.1103/PhysRevD.88.072012>
11. Belle Collaboration, Phys. Rev. D **92**, 072014 (2015). <https://doi.org/10.1103/PhysRevD.92.072014>
12. L. Collaboration, Phys. Rev. Lett. **115**(11), 111803 (2015). <https://doi.org/10.1103/PhysRevLett.115.111803>. [Erratum: Phys. Rev. Lett. **115**, 159901 (2015)]
13. LHCb Collaboration, Phys. Rev. Lett. **115**, 111803 (2015). <https://doi.org/10.1103/PhysRevLett.115.111803>. [Erratum: Phys. Rev. Lett. **115**, 159901 (2015)]
14. Belle Collaboration, Phys. Rev. D **94**, 072007 (2016). <https://doi.org/10.1103/PhysRevD.94.072007>
15. Belle Collaboration, Phys. Rev. Lett. **118**, 211801 (2017). <https://doi.org/10.1103/PhysRevLett.118.211801>
16. LHCb Collaboration, Phys. Rev. Lett. **120**, 171802 (2018). <https://doi.org/10.1103/PhysRevLett.120.171802>
17. Belle Collaboration, Phys. Rev. D **97**, 012004 (2018). <https://doi.org/10.1103/PhysRevD.97.012004>
18. L. Collaboration, Phys. Rev. D **97**(7), 072013 (2018). <https://doi.org/10.1103/PhysRevD.97.072013>
19. Belle Collaboration, (2019). <https://doi.org/10.48550/arXiv.1904.08794>
20. Belle Collaboration, Phys. Rev. Lett. **124**, 161803 (2020). <https://doi.org/10.1103/PhysRevLett.124.161803>
21. L. Collaboration, (2023). <https://doi.org/10.48550/arXiv.2302.02886>
22. S. Navas et al., Phys. Rev. D **110**(3), 030001 (2024). <https://doi.org/10.1103/PhysRevD.110.030001>
23. A.M. Sirunyan et al., JHEP **03**, 170 (2019). [https://doi.org/10.1007/JHEP03\(2019\)170](https://doi.org/10.1007/JHEP03(2019)170)
24. V. Khachatryan et al., Phys. Rev. Lett. **118**(2), 021802 (2017). <https://doi.org/10.1103/PhysRevLett.118.021802>
25. V. Khachatryan et al., JHEP **02**, 048 (2017). [https://doi.org/10.1007/JHEP02\(2017\)048](https://doi.org/10.1007/JHEP02(2017)048)
26. V. Khachatryan et al., JHEP **03**, 077 (2017). [https://doi.org/10.1007/JHEP03\(2017\)077](https://doi.org/10.1007/JHEP03(2017)077)
27. A.M. Sirunyan et al., JHEP **07**, 121 (2017). [https://doi.org/10.1007/JHEP07\(2017\)121](https://doi.org/10.1007/JHEP07(2017)121)
28. V. Khachatryan et al., JHEP **11**, 189 (2015). [https://doi.org/10.1007/JHEP11\(2015\)189](https://doi.org/10.1007/JHEP11(2015)189)
29. J.C. Pati, A. Salam, Phys. Rev. D **10**(1), 275 (1974). <https://doi.org/10.1103/PhysRevD.10.275>
30. R.N. Mohapatra, J.C. Pati, Phys. Rev. D **11**(9), 2558 (1975). <https://doi.org/10.1103/PhysRevD.11.2558>
31. G. Senjanovic, R.N. Mohapatra, Phys. Rev. D **12**(5), 1502 (1975). <https://doi.org/10.1103/PhysRevD.12.1502>
32. L. Di Luzio, J. Fuentes-Martin, A. Greljo, M. Nardecchia, S. Renner, J. High Energy Phys. **2018**(11), 81 (2018). [https://doi.org/10.1007/JHEP11\(2018\)081](https://doi.org/10.1007/JHEP11(2018)081)
33. M.J. Baker, J. Fuentes-Martín, G. Isidori, M. König, Eur. Phys. J. C **79**(4), 334 (2019). <https://doi.org/10.1140/epjc/s10052-019-6853-x>
34. L. Michaels, F. Yu, JHEP **03**, 120 (2021). [https://doi.org/10.1007/JHEP03\(2021\)120](https://doi.org/10.1007/JHEP03(2021)120)
35. P.S.B. Dev, W. Rodejohann, X.J. Xu, Y. Zhang, JHEP **06**, 039 (2021). [https://doi.org/10.1007/JHEP06\(2021\)039](https://doi.org/10.1007/JHEP06(2021)039)
36. A. Flórez, J. Jones-Pérez, A. Gurrola, C. Rodríguez, J. Peñuela-Parra, Eur. Phys. J. C **83**(11), 1023 (2023). <https://doi.org/10.1140/epjc/s10052-023-12177-4>
37. B. Dutta, S. Ghosh, J. Kumar, in *Snowmass 2021* (2022)
38. B. Dutta, S. Ghosh, J. Kumar, Phys. Rev. D **100**(7), 075028 (2019). <https://doi.org/10.1103/PhysRevD.100.075028>
39. B. Dutta, S. Ghosh, J. Kumar, Phys. Rev. D **102**(7), 075041 (2020). <https://doi.org/10.1103/PhysRevD.102.075041>
40. B. Dutta, S. Ghosh, J. Kumar, Phys. Rev. D **102**(1), 015013 (2020). <https://doi.org/10.1103/PhysRevD.102.015013>
41. B. Dutta, S. Ghosh, P. Huang, J. Kumar, Phys. Rev. D **105**(1), 015011 (2022). <https://doi.org/10.1103/PhysRevD.105.015011>

42. V. De Romeri, J. Nava, M. Puerta, A. Vicente, Phys. Rev. D **107**, 095019 (2023). <https://doi.org/10.1103/PhysRevD.107.095019>
43. B. Dutta, S. Ghosh, A. Gurrola, D. Julson, T. Kamon, J. Kumar, J. High Energy Phys. **2023**(3), 164 (2023). [https://doi.org/10.1007/JHEP03\(2023\)164](https://doi.org/10.1007/JHEP03(2023)164)
44. Z. Berezhiani, Phys. Lett. B **129**(1), 99 (1983). [https://doi.org/10.1016/0370-2693\(83\)90737-2](https://doi.org/10.1016/0370-2693(83)90737-2)
45. D. Chang, R.N. Mohapatra, Phys. Rev. Lett. **58**, 1600 (1987). <https://doi.org/10.1103/PhysRevLett.58.1600>
46. A. Davidson, K.C. Wali, Phys. Rev. Lett. **59**, 393 (1987). <https://doi.org/10.1103/PhysRevLett.59.393>
47. S. Rajpoot, Mod. Phys. Lett. A **02**(05), 307 (1987). <https://doi.org/10.1142/S0217732387000422>
48. K.S. Babu, R.N. Mohapatra, Phys. Rev. Lett. **62**, 1079 (1989). <https://doi.org/10.1103/PhysRevLett.62.1079>
49. K.S. Babu, R.N. Mohapatra, Phys. Rev. D **41**, 1286 (1990). <https://doi.org/10.1103/PhysRevD.41.1286>
50. A. Bhardwaj, T. Mandal, S. Mitra, C. Neeraj, Phys. Rev. D (2022). <https://doi.org/10.1103/physrevd.106.095014>
51. A. Bhardwaj, K. Bhide, T. Mandal, S. Mitra, C. Neeraj, Phys. Rev. D (2022). <https://doi.org/10.1103/physrevd.106.075024>
52. J. Bardhan, T. Mandal, S. Mitra, C. Neeraj, Phys. Rev. D (2023). <https://doi.org/10.1103/physrevd.107.115001>
53. S. Banerjee, D. Barducci, G. Bélanger, C. Delaunay, J. High Energy Phys. (2016). [https://doi.org/10.1007/jhep11\(2016\)154](https://doi.org/10.1007/jhep11(2016)154)
54. J.M. Alves, G. Branco, A. Cherchiglia, C. Nishi, J. Penedo, P.M. Pereira, M. Rebelo, J. Silva-Marcos, Phys. Rep. **1057**, 1–69 (2024). <https://doi.org/10.1016/j.physrep.2023.12.004>
55. J.H. Friedman, Ann. Stat. **29**(5), 1189 (2001). <https://doi.org/10.1214/aos/1013203451>. (Publisher: Institute of Mathematical Statistics)
56. X. Ai, S.C. Hsu, K. Li, C.T. Lu, JPCS **2438**(1), 012114 (2023). <https://doi.org/10.1088/1742-6596/2438/1/012114>
57. A. Collaboration, Phys. Rev. D **97**(7), 072016 (2018). <https://doi.org/10.1103/PhysRevD.97.072016>
58. A. Biswas, D. Kumar-Ghosh, N. Ghosh, A. Shaw, A.K. Swain, J. Phys. G **47**(4), 045005 (2020). <https://doi.org/10.1088/1361-6471/ab6948>
59. Y.L. Chung, S.C. Hsu, B. Nachman, JINST **16**, P07002 (2021). <https://doi.org/10.1088/1748-0221/16/07/P07002>
60. J. Feng, M. Li, Q.S. Yan, Y.P. Zeng, H.H. Zhang, Y. Zhang, Z. Zhao, JHEP (2022). [https://doi.org/10.1007/jhep09\(2022\)141](https://doi.org/10.1007/jhep09(2022)141)
61. D. Barbosa, F. Díaz, L. Quintero, A. Flórez, M. Sanchez, A. Gurrola, E. Sheridan, F. Romeo, EPJC (2023). <https://doi.org/10.1140/epjc/s10052-023-11506-x>
62. S. Chigusa, S. Li, Y. Nakai, W. Zhang, Y. Zhang, J. Zheng, Phys. Lett. B **833**, 137301 (2022). <https://doi.org/10.1016/j.physletb.2022.137301>
63. E. Arganda, D.A. Díaz, A.D. Perez, R.M. Sandá Seoane, A. Szynkman, Phys. Rev. D **109**, 055032 (2024). <https://doi.org/10.1103/PhysRevD.109.055032>
64. S. Ajmal, J.T. Gaglione, A. Gurrola, O. Panella, M. Presilla, F. Romeo, H. Sun, S.S. Xue, J. High Energy Phys. (2024). [https://doi.org/10.1007/jhep08\(2024\)176](https://doi.org/10.1007/jhep08(2024)176)
65. B. Dutta, A. Gurrola, K. Hatakeyama, W. Johns, T. Kamon, P. Sheldon, K. Sinha, S. Wu, Z. Wu, Phys. Rev. D (2015). <https://doi.org/10.1103/physrevd.92.095009>
66. C. Collaboration, Review of searches for vector-like quarks, vector-like leptons, and heavy neutral leptons in proton-proton collisions at  $\sqrt{s} = 13$  TeV at the CMS experiment (2024)
67. C. Collaboration, Search for production of a single vector-like quark decaying to Th or Tz in the all-hadronic final state in pp collisions at  $\sqrt{s} = 13$  TeV (2024)
68. G. Aad et al., Phys. Rev. D **105**(9), 092012 (2022). <https://doi.org/10.1103/PhysRevD.105.092012>
69. G. Aad et al., Phys. Rev. D **109**(11), 112012 (2024). <https://doi.org/10.1103/PhysRevD.109.112012>
70. G. Aad et al., Phys. Lett. B **843**, 138019 (2023). <https://doi.org/10.1016/j.physletb.2023.138019>
71. G. Aad et al., Eur. Phys. J. C **83**(8), 719 (2023). <https://doi.org/10.1140/epjc/s10052-023-11790-7>
72. G. Aad et al., JHEP **08**, 153 (2023). [https://doi.org/10.1007/JHEP08\(2023\)153](https://doi.org/10.1007/JHEP08(2023)153)
73. G. Aad et al., arXiv:2403.09292
74. G. Cacciapaglia, T. Flacke, M. Park, M. Zhang, Phys. Lett. B **798**, 135015 (2019). <https://doi.org/10.1016/j.physletb.2019.135015>
75. A. Alloul, N.D. Christensen, C. Degrande, C. Duhr, B. Fuks, Comput. Phys. Commun. **185**, 2250 (2014). <https://doi.org/10.1016/j.cpc.2014.04.012>
76. C. Degrande, C. Duhr, B. Fuks, D. Grellscheid, O. Mattelaer, T. Reiter, Comput. Phys. Commun. **183**, 1201 (2012). <https://doi.org/10.1016/j.cpc.2012.01.022>
77. J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.S. Shao, T. Stelzer, P. Torrielli, M. Zaro, JHEP **07**, 079 (2014). [https://doi.org/10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079)
78. J. Alwall, C. Duhr, B. Fuks, O. Mattelaer, D.G. Öztürk, C.H. Shen, Comput. Phys. Commun. **197**, 312 (2015). <https://doi.org/10.1016/j.cpc.2015.08.031>
79. N. Collaboration, JHEP **04**, 040 (2015). [https://doi.org/10.1007/JHEP04\(2015\)040](https://doi.org/10.1007/JHEP04(2015)040)
80. T. Sjöstrand, S. Ask, J.R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C.O. Rasmussen, P.Z. Skands, Comput. Phys. Commun. **191**, 159 (2015). <https://doi.org/10.1016/j.cpc.2015.01.024>
81. D. Collaboration, JHEP **02**, 057 (2014). [https://doi.org/10.1007/JHEP02\(2014\)057](https://doi.org/10.1007/JHEP02(2014)057)
82. A.C. Canbay, O. Cakir, Phys. Rev. D **108**, 095006 (2023). <https://doi.org/10.1103/PhysRevD.108.095006>
83. G. Cacciapaglia, A. Deandrea, S. Vatani, Phys. Rev. D (2023). <https://doi.org/10.1103/physrevd.108.016010>
84. G. Blankenburg, G. Isidori, J. Jones-Perez, Eur. Phys. J. C **72**, 2126 (2012). <https://doi.org/10.1140/epjc/s10052-012-2126-7>
85. J. Jones-Perez, J. Phys. Conf. Ser. **447**, 012060 (2013). <https://doi.org/10.1088/1742-6596/447/1/012060>
86. L. Calibbi, J. Jones-Perez, A. Masiero, J.h. Park, W. Porod, O. Vives, PoS EPS-HEP2009, 167 (2009). <https://doi.org/10.22323/1.084.0167>
87. M. Cacciari, G.P. Salam, G. Soyez, J. High Energy Phys. **2008**(04), 063 (2008). <https://doi.org/10.1088/1126-6708/2008/04/063>
88. M. Cacciari, G.P. Salam, G. Soyez, Eur. Phys. J. C (2012). <https://doi.org/10.1140/epjc/s10052-012-1896-2>
89. Study of the discovery reach in searches for supersymmetry at CMS with 3000/fb. Geneva (2013). <https://cds.cern.ch/record/1607141>. CMS-PAS-FTR-13-014
90. A. Sirunyan et al., JINST **15**(02), P02027 (2020). <https://doi.org/10.1088/1748-0221/15/02/P02027>
91. A.M. Sirunyan et al., JINST **15**(06), P06005 (2020). <https://doi.org/10.1088/1748-0221/15/06/P06005>
92. M. Aaboud et al., Eur. Phys. J. C **79**(5), 375 (2019). <https://doi.org/10.1140/epjc/s10052-019-6847-8>
93. A. Sirunyan et al., JINST **13**, P05011 (2018). <https://doi.org/10.1088/1748-0221/13/05/P05011>
94. E. Bols, J. Kieseler, M. Verzetti, M. Stoye, A. Stakia, J. Instrum. **15**(12), P12012 (2020). <https://doi.org/10.1088/1748-0221/15/12/P12012>
95. E. Conte, B. Fuks, G. Serret, Comput. Phys. Commun. **184**(1), 222 (2013). <https://doi.org/10.1016/j.cpc.2012.09.009>
96. A. Paszke et al., Pytorch: an imperative style, high-performance deep learning library. (2019)
97. T. Chen, C. Guestrin, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data*

- Mining (Association for Computing Machinery, New York, 2016), KDD '16, pp. 785–794. <https://doi.org/10.1145/2939672.2939785>
98. J. Butterworth, S. Carrazza, A. Cooper-Sarkar, A. De Roeck, J. Feltesse, S. Forte, J. Gao, S. Glazov, J. Huston, Z. Kassabov, R. McNulty, A. Morsch, P. Nadolsky, V. Radescu, J. Rojo, R. Thorne, J. Phys. G **43**, 023001 (2016). <https://doi.org/10.1088/0954-3899/43/2/023001>
99. A. Flórez, A. Gurrola, W. Johns, P. Sheldon, E. Sheridan, K. Sinha, B. Soubasis, Phys. Rev. D **103**(9), 095001 (2021). <https://doi.org/10.1103/PhysRevD.103.095001>
100. A. Flórez, A. Gurrola, W. Johns, J. Maruri, P. Sheldon, K. Sinha, S.R. Starko, Phys. Rev. D **100**(1), 016017 (2019). <https://doi.org/10.1103/PhysRevD.100.016017>
101. A. Flórez, Y. Guo, A. Gurrola, W. Johns, O. Ray, P. Sheldon, S. Starko, Phys. Rev. D **99**(3), 035034 (2019). <https://doi.org/10.1103/PhysRevD.99.035034>
102. A. Flórez, K. Gui, A. Gurrola, C. Patiño, D. Restrepo, Phys. Lett. B **778**, 94 (2018). <https://doi.org/10.1016/j.physletb.2018.01.009>
103. A. Florez, A. Gurrola, W. Johns, Y. Do Oh, P. Shendon, D. Teague, T. Weiler, Phys. Lett. B **767**, 126 (2017). <https://doi.org/10.1016/j.physletb.2017.01.062>
104. A. Flórez, L. Bravo, A. Gurrola, C. Ávila, M. Segura, P. Sheldon, W. Johns, Phys. Rev. D **94**(7), 073007 (2016). <https://doi.org/10.1103/PhysRevD.94.073007>
105. R. Leonardi, O. Panella, F. Romeo, A. Gurrola, H. Sun, S.S. Xue, Eur. Phys. J. C (2020). <https://doi.org/10.1140/epjc/s10052-020-7822-0>
106. C. Natalia, F. Andrés, G. Alfredo, J. Will, S. Paul, T. Cheng, Long-term LHC discovery reach for compressed higgsino-like models using VBF processes. (2021)
107. A. Flórez, A. Gurrola, W. Johns, P. Sheldon, E. Sheridan, K. Sinha, B. Soubasis, Phys. Rev. D **103**, 095001 (2021). <https://doi.org/10.1103/PhysRevD.103.095001>
108. P. Lujan et al., The Pixel Luminosity Telescope: a detector for luminosity measurement at CMS using silicon pixel sensors. Tech. Rep. CMS-DN-21-008 (2022). Accepted by Eur. Phys. J. C