ATLAS CONF Note

ATLAS-CONF-2021-018

6th April 2021

Search for single vector-like B quark production and decay via $B \to bH(b\bar{b})$ in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

The ATLAS Collaboration

A search is presented for single production of a vector-like B quark decaying to a Standard Model b-quark and the Standard Model Higgs boson, which decays into $b\bar{b}$. The search is carried out on 139 fb^{-1} of proton–proton collision data at $\sqrt{s}$ = 13 TeV collected by the ATLAS detector at the LHC between 2015 and 2018. No significant deviation from the Standard Model background prediction is observed. Mass-dependent exclusion limits at the 95% confidence level are drawn on the single production cross-section of a vector-like B quark occurring as part of a (B, Y) weak isospin doublet and various hypotheses on the coupling strength κ. Within such hypotheses the search excludes the presence of a vector-like B quark resonance with 1.0 TeV $< m_B <$ 1.28 TeV and 1.46 TeV $< m_B <$ 2.0 TeV for $\kappa = 0.25$, and the 1.0 TeV $< m_B <$ 2.0 TeV resonance mass range for $\kappa = 0.3$. These intervals exclude resonances of relative width larger than 5% for masses 1.0 TeV $< m_B <$ 1.76 TeV.

ATLAS-CONF-2021-018
06 April 2021

1 Introduction

The observation of a particle compatible with the Higgs boson at the Large Hadron Collider (LHC) [1, 2] completed the full picture of the fundamental particles predicted to exist according to the Standard Model (SM). Nevertheless, the comparatively low observed Higgs boson mass conflicts with the expected effect of the higher-order quantum loop mass corrections, which would push the physical Higgs boson mass towards the Planck scale. Such observations and naturalness arguments [3] suggest the existence of an as-of-yet-undiscovered mechanism beyond the SM preventing the aforementioned divergent contributions [4]. Theoretical extensions of the SM attempt to provide a natural solution to this issue by postulating the Higgs boson to be either a composite particle [5, 6] or a pseudo-Nambu-Goldstone boson, such as in the Little Higgs model [7]. In these models, an additional symmetry corresponds to a new strong interaction, whose bound states include vector-like Quarks (VLQs).

VLQs are predicted to be spin-1/2 particles that transform as a triplet under colour gauge symmetry and whose left and right-handed components both have the same electroweak quantum numbers [8]. They couple to the SM fermions via Yukawa couplings [9] and therefore interact principally with the third-generation SM quarks, the two most massive fundamental fermions. Theoretical constraints on the renormalisability of the coupling constants restrict the occurrence of VLQs to seven gauge covariant multiplets. Vector-like T and B quarks, the vector-like equivalent of the third-generation SM quarks, can exist as singlets, doublets and triplets, whereas the X and Y VLQs, respectively with exotic charges $Q_X = 5/3$ and $Q_Y = -4/3$, can exist as part of either of gauge doublets and triplets with either or both the T and B quarks.

VLQs are expected to be produced at the LHC either in pairs, mediated by the strong interaction, or singly, via the exchange of an intermediate electroweak gauge boson. VLQ pair production is a pure QCD process with a cross-section that, at leading order, depends solely on the VLQ mass, whereas VLQ single production cross-sections are additionally and strongly affected by both the coupling strength to the SM quarks and the multiplet considered [8]. Furthermore, single VLQ production may overtake pair production as the principal VLQ production mechanism above the TeV threshold depending on the values of the coupling strength with the SM quarks.

The theoretical framework for VLQs sets a common bare-mass term in the Yukawa Lagrangian for all VLQs, resulting in a mass split of the order of 1-10 GeV among the various weak eigenstates. The main consequence of this feature is the heavy suppression of all cascade decays of one VLQ into another, which in turn causes the bulk of the total decay width to be determined by the coupling to the SM third-generation quarks. For T and B VLQs, the allowed decays are neutral-current conversion into the SM equivalent ($T \to tH, tZ$ and $B \to bH, bZ$) or charged-current final states via the emission of a W boson ($T \to bW$ and $B \to tW$). Likewise, the production of a single final-state vector-like B occurs by virtue of the $tW \to B$ and $bZ \to B$ vertices, and similarly for the top quark partner T.

The kinematic properties of the signal are inferred from a theory-agnostic phenomenological model of VLQ single production [10–12] where both the left and right-handed components of VLQ mix with third-generation SM quarks via Yukawa couplings, giving rise to the interaction vertices mentioned above. Under this picture, the coupling constants between the vector-like B quark and the W, Z and H bosons, which regulate both the production cross-section and decay width, are given by:

$$c_W = \kappa\sqrt{\frac{2\xi_W}{\rho_W}}, \qquad c_Z = \frac{m_Z}{m_W} \times \kappa\sqrt{\frac{2\xi_Z}{\rho_Z}}, \qquad c_H = \frac{1}{2}\frac{g_W m_B}{m_W} \times \kappa\sqrt{\frac{2\xi_H}{\rho_H}}, \tag{1}$$

in which $\rho_{W,Z,H}$ are dimensionless kinematic factors approximately equal to 1 for $m_B > 1$ TeV and $\xi_{W,Z,H}$ dimensionless constants which determine the coupling hierarchy and summing to unity. Furthermore, all three coupling constants scale with the universal coupling strength κ, which also uniquely determines the resonance width for a VLB of a given mass m_B:

$$\Gamma_B = \frac{g^2}{128\pi} \times \frac{m_B^3}{m_W^2} \times \kappa^2 \tag{2}$$

with g being the weak coupling constant. The various theoretically motivated multiplet scenarios are reflected in the choice of values for the ξ constants: for a B singlet, $\xi_W = 0.5$ and $\xi_Z = \xi_H = 0.25$, for a (T, B) doublet $\xi_W = \xi_Z = 0.5$ and $\xi_H = 0$ and for a (B, Y) doublet $\xi_W = 0$ and $\xi_Z = \xi_H = 0.5$. In the asymptotic, high m_B limit, which holds to a good approximation down to $m_B \sim 1$ TeV, the values of ξ in each multiplet state correspond to the branching fractions of the B in the respective decay mode.

This article details the search for a single-produced vector-like B quark (VLB) occurring as part of a (B,Y) weak isospin doublet and decaying to the $B \to bH$, $H \to b\bar{b}$ final state, as shown in Figure 1. Under this theoretical scenario, in which $\xi_W = 0$ results in a vanishing coupling to the W boson, the vector-like B can only be generated as a result of the electroweak interaction of an initial state b quark and a Z boson.

A vector-like B quark can also be produced non-resonantly, via t-channel exchange. As the resonant production dominates for the narrow width B states considered in this note, these non-resonant contributions are neglected.

This search establishes 95% confidence level limits on the phenomenological coupling constant c_Z appearing in the simplified VLQ Lagrangian introduced in [10, 12], significantly expanding the regions of the wider VLQ phenomenological phase space probed at collider experiments.

Several searches carried out on Run 1 and Run 2 ATLAS data [13–15] have targeted VLQ-compatible signatures. Vector-like B quark single production has been probed using the full Run 2 dataset at CMS in the $bZ(b\bar{b})$ and $bH(b\bar{b})$ modes [16, 17] , and ATLAS in the $bH(\gamma\gamma)$ final state, using $\sqrt{s} = 13$ TeV data collected in the 2015, 2016 and 2017 data-taking periods, corresponding to a total integrated luminosity of 80 fb^{-1}. No evidence of single-produced vector-like B quarks was found and a 95% confidence level exclusion limit on both the singlet (BR($B \to bH$) ≈ 0.5) and (B, Y) doublet (BR($B \to bH$) ≈ 0.25) hypotheses, ruling out masses below ≈ 1200 GeV.

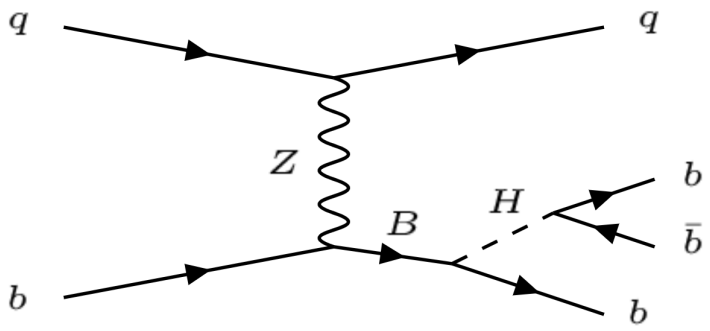


Figure 1: Feynman diagrams of the main leading-order single production mode of a vector-like B quark, as mediated by Z boson. The diagram displays the VLB decay resulting in the final state targetted by this search.

2 ATLAS detector

The ATLAS detector [18] is a general-purpose particle detector used to investigate a broad range of physics processes. It includes inner tracking devices surrounded by a 2.3 m diameter superconducting solenoid, electromagnetic (EM) and hadronic calorimeters and a muon spectrometer with a toroidal magnetic field. The Inner Detector (ID) consists of a high-granularity silicon pixel detector, including the insertable B-layer [19, 20] installed after Run 1 of the LHC, a silicon strip detector, and a straw-tube tracker. It is immersed in a 2 T axial magnetic field and provides precision tracking of charged particles with pseudorapidity $|\eta| < 2.5$.[1] The calorimeter system consists of finely segmented sampling calorimeters using lead/liquid-argon for the detection of EM showers up to $|\eta| < 3.2$, and copper or tungsten/liquid-argon for EM and hadronic showers for $1.5 < |\eta| < 4.9$. In the central region ($|\eta| < 1.7$), a steel/scintillator hadronic calorimeter is used. Outside the calorimeters, the muon system incorporates multiple layers of trigger and tracking chambers within a magnetic field produced by a system of superconducting toroids, enabling an independent precise measurement of muon track momenta for $|\eta| < 2.7$. A dedicated trigger system is used to select events [21]. The first-level trigger is implemented in hardware and uses the calorimeter and muon detectors to reduce the accepted event rate to below 100 kHz. This is followed by a software-based high-level trigger, which reduces the rate further, to 1 kHz on average.

3 Data and simulated samples

This analysis uses 139 fb^{-1} of LHC proton-proton (pp) collisions at $\sqrt{s} = 13$ TeV collected by ATLAS during LHC's Run 2 from 2015 to 2018. The data used were collected during stable beam conditions with all relevant detector systems functional and producing good quality data [22]. Events were selected online through a trigger signature that requires a single anti-k_t jet [23] with radius parameter $R = 1.0$ (large-R jet) with transverse energy (E_{T}) thresholds of 420 GeV and 460 GeV in the 2015-16 and 2017-18 data-taking periods, respectively. This trigger requirement is $> 99\%$ efficient for events passing the analysis selection of a large-R jet with a transverse momentum (p_{T}) over 480 GeV.

The VLB signal is modelled by means of a Monte Carlo (MC) simulation based on the phenomenological Lagrangian outlined in the previous section. Signal samples were generated with MadGraph5_aMC@NLO [24] under the four-flavour scheme, using the leading order (LO) NNPDF2.3 PDF sets [25]. Parton showering and hadronization was handled by Pythia [26] v8.212 and a set of tuned parameters called the ATLAS A14 tune [26] for the underlying event. The EvtGen program v1.2.0 [27] was used to model the properties of b-hadron decays. The detector response was simulated with Geant4 4 [28] and the events were processed with the same reconstruction software as that used for data. All simulated samples include the effects due to multiple pp interactions per bunch-crossing (pile-up), as well as the effect on the detector response due to interactions from bunch crossings before or after the one containing the hard interaction.

Samples are generated with resonance pole mass ranging from 1 TeV to 2 TeV, in steps of 200 GeV. To facilitate the interpretation of the search result in terms of exclusion limits on the coupling constant c_Z, an event-by-event reweighting mechanism was introduced at the matrix-element level to simulate the effects

[1] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector. The positive x-axis is defined by the direction from the IP to the centre of the LHC ring, with the positive y-axis pointing upwards, while the beam direction defines the z-axis. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudorapidity η is defined in terms of the polar angle θ by $\eta = -\ln\tan(\theta/2)$. The transverse momentum (p_{T}) is defined relative to the beam axis and is calculated as $p_{\mathrm{T}} = p\sin(\theta)$.

on the signal properties of different values of the coupling strength κ introduced in Eq. 1, ranging from $\kappa = 0.1$ to $\kappa = 1.6$. All samples were generated with a nominal coupling constant $\kappa = 1.0$, inclusively in terms of the Higgs boson decay modes. Signal samples are normalized to values of the cross-section calculated at next-to-leading order (NLO) assuming the narrow-width approximation and a five-flavour PDF scheme.

As the final state searched for is fully hadronic, multijet production is overwhelmingly the dominant source of background and a fully data-derived background estimation is performed, as described in Section 6. As a cross-check, simulated $t\bar{t}$ samples were also studied to confirm that the small (few percent) contribution from this process can be fully accounted for by the data-derived background estimation technique. The contribution from top quark pair production is modelled at NLO in the strong coupling by the MC generator POWHEG BOXV2 [29–31] fitted with the NNPDF3.0 NLO [32] parton distribution function (PDF) for matrix-element calculations. The top pair-production cross-section is scaled to a next-to-next-to-leading-order (NNLO) calculation in QCD including resummation of next-to-next-to-leading logarithmic soft gluon terms with Top++2.0 [33–37]. Parton showering, hadronization and the underlying event were simulated using PYTHIAV8.230 [38] with the LO NNPDF2.3 [25] PDF set and the A14 set of tuned parameters [26].

The contribution to the total SM background from other processes, such as Z+jets production, is estimated to be negligible on the grounds of their cross-section and the acceptance of the event selection with respect to such processes.

4 Object reconstruction

The production and decay of a VLB with a mass in the range of 1-2 TeV targeted in this search results in a final state composed of a high-p_{T} Higgs boson decaying to $b\bar{b}$, an energetic jet from the b-quark originating directly from the VLB decay, and an additional, softer, forward jet from the spectator quark involved in the hard scatter (as shown in Figure 1). The boosted Higgs boson decay system is reconstructed as a single large-radius (large-R) jet displaying a two-pronged energy profile which originates from the hadronization of the b- and $\bar{b}$-quarks. Conversely, the b-quark from the initial VLB decay and the spectator quark, both expected features of a signal-like event, are reconstructed as standard, small-radius (small-R) jets. The identification of three b-hadrons in this event topology is key to the background suppression.

Events are checked to remove those with noise bursts or coherent noise in the calorimeters, as well as those containing large energy deposits from non-collision or cosmic sources of background. Collision vertices are reconstructed from inner detector tracks with $p_{\mathrm{T}} > 0.5$ GeV. The primary vertex in each event is chosen to be the one with the largest sum of the squared transverse momenta of all associated tracks. Events without a reconstructed primary vertex are rejected.

Events containing isolated, charged leptons (electrons or muons) are removed in this analysis, since no leptons are expected in the final state under study. This applies to events containing electron candidates with $E_{\mathrm{T}} > 25$ GeV that satisfy the 'loose' identification criteria defined in Ref. [39] or muons with $p_{\mathrm{T}} > 25$ GeV satisfying the 'medium' quality requirements [40]. Also, to preserve orthogonality with VLB searches targeting the $H \to \gamma\gamma$ channel, events with isolated photons that pass the 'tight' identification criteria [41] are removed if a pair of photons has an invariant mass in the range $105 - 160$ GeV.

Small-R jets are reconstructed using the anti-k_t algorithm [23], with a radius parameter $R = 0.4$, out of inner detector tracks associated with the primary vertex and calorimeter clusters selected through a particle flow reconstruction algorithm [42]. An appropriate energy calibration is applied to both the input

calorimeter clusters [43] and the final reconstructed jet [44]. Additionally, a pileup subtraction procedure [45] is applied along with a global sequential calibration to account for flavour dependencies. To suppress jets arising from pile-up, a jet-vertex-tagging (JVT) technique using a multivariate likelihood [46] is applied to jets with $p_{\mathrm{T}} < 60$ GeV, ensuring that selected jets are matched to the primary vertex. Jets with $|\eta| > 2.4$, falling outside the inner detector acceptance, undergo a tighter selection via a specifically designed and trained forward-JVT algorithm [47].

Large-R jets are built from three-dimensional topological clusters of energy deposits in the calorimeter, calibrated to the hadronic energy scale with the local cluster weighting (LCW) procedure [48], on which the anti-k_t algorithm is applied with radius parameter $R = 1.0$. The reconstructed jets are 'trimmed' [49] to reduce contributions from pile-up and soft interactions. This is performed by reclustering the jet constituents into subjets using the k_t algorithm [50, 51] with a radius parameter $R = 0.2$ and discarding subjets with p_{T} less than 5% of the parent jet p_{T} [52]. The large-R jet four-momentum is then recomputed from the four-momenta of the remaining subjets and corrected using simulation [43, 53].

Small-R jets in the range $|\eta_j| < 2.5$ are recognized as containing a b hadron (henceforth b-tagged) using the "DL1r" algorithm [54]. This algorithm is based on a multivariate classification technique with an artificial deep neural network combining information from the impact parameters of tracks, reconstructed muons in jets and topological properties of secondary and tertiary decay vertices reconstructed from the tracks associated to the jet. The b-jets are selected in the analysis using the working point corresponding to an efficiency of 70% for identifying true b-jets in simulated SM $t\bar{t}$ events. The corresponding mis-tagging efficiencies for jets containing c-hadrons and light-flavour jets are estimated to be, respectively, 10% and 0.2%.

In order to explore the presence of one or multiple b-hadrons within the large-R jet radius, as expected in boosted $H \to b\bar{b}$ decays, variable-radius (VR) track jets are matched to the large-R jet via "ghost association" [55–57], and subsequently inspected for b-tagging. Track jets are built from inner detector tracks using the anti-k_t algorithm with a radius parameter R inversely proportional to the jet p_{T} [58]:

$$R \to R_{\mathrm{eff}}(p_{\mathrm{T}}) = \rho / p_{\mathrm{T}} \,.$$

The ρ-parameter that controls the effective radius, R_{eff} is set to $\rho = 30$ GeV. There are two additional parameters, R_{min} and R_{max}, used to set the minimum and maximum bounds on R_{eff}, and these are set to 0.02 and 0.4, respectively [59]. The values of these parameters were determined by examining the efficiency of identifying two b-jets within a large-radius jet associated with a high-p_{T} Higgs boson decaying into a b-quark pair [60]. As for small-R jets, b-tagging is performed by means of a version of the "DL1" algorithm specifically re-trained for tagging VR track jets. For a 70% efficiency in identifying true b-jets as measured in simulated SM $t\bar{t}$ events, the mis-tag efficiency for c and light jets is approximately around 10% and 0.25%, respectively. For both small-R jets and track jets, the efficiencies of identifying b-jets, c-jets, and light-flavour jets are corrected in the simulation to account for deviations from the efficiencies observed in data [54].

A veto is enforced on events where any of the track jets inspected for b-tagging are found to be collinear with any other track jet in the event with $p_T > 10$ GeV. For the purpose of this cleaning procedure, track jet collinearity is defined by the geometrical separation ΔR between the two examined track jets being smaller than both of the jet effective radii. The collinear veto prevents events with ambiguous track to track-jet matching, and therefore uncalibrated flavour tagging performance, from entering the analysis, at the cost of an observed 6% signal efficiency loss across the available resonance mass spectrum.

5 Event selection and categorization

As mentioned above, Higgs candidates (HC) are reconstructed as single large-R jets. Events are selected if they have at least one eligible HC reconstructed out of a large-R jet with $p_{\mathrm{T}}^{\mathrm{HC}} > 480$ GeV, $|\eta| < 2.0$, invariant mass (m_{HC}) in the range 105-135 GeV, and at least two matched b-tagged track-jets with $p_{\mathrm{T}} > 50$ GeV. HCs are then classified on the grounds of the b-tagged track jet multiplicity, allowing candidates with a higher number of b-tagged jets to be prioritized in case of multiple eligible HCs being present within a single event. The two HC categories are labelled "`H2T2B`", for candidates with two matched b-tagged track-jets, and "`H2T1B`" otherwise. The presence of at least one b-tagged track-jet matched to the large-R jet is required for Higgs candidate eligibility.

VLB candidates are formed by combining a HC with a b-tagged small-R jet required to have $p_{\mathrm{T}} > 480$ GeV, $|\eta| < 2.5$ and an angular distance $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} > 2.0$ from the HC. Two further selection criteria are applied to exploit the subjet structure of the HC in the signal and the correlation between $p_{\mathrm{T}}^{\mathrm{HC}}$ and the VLB candidate mass, m_{B}. The first is captured by the quantity $\log\Delta R^*$, defined as:

$$\log\Delta R^* \;=\; \log\left[\frac{\Delta R\Big(\mathrm{tj0},\,\mathrm{tj1}\Big)}{\min\left[R_{\mathrm{eff}}^{\mathrm{tj0}},\,R_{\mathrm{eff}}^{\mathrm{tj1}}\right]}\right]$$

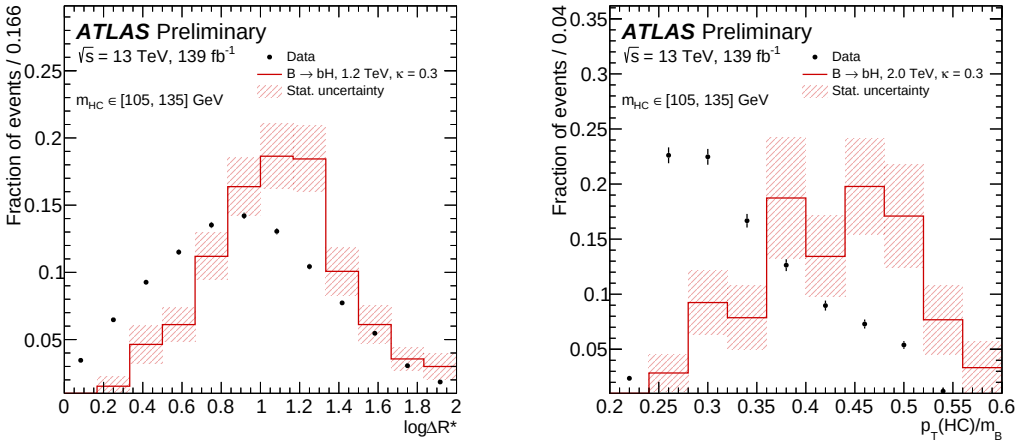
where tj0, tj1 are the two highest-p_{T} track-jets associated to the HC, and $R_{\mathrm{eff}}^{\mathrm{tj0}}, R_{\mathrm{eff}}^{\mathrm{tj1}}$ their p_{T}-dependent effective radii. The second quantity is the ratio between the HC p_T and the reconstructed VLB invariant mass, $p_{\mathrm{T}}^{\mathrm{HC}}/m_{\mathrm{B}}$, which is sensitive to whether the QCD background event is produced as a result of an s- or t-channel process. The distributions of $\log\Delta R^*$ and $p_{\mathrm{T}}^{\mathrm{HC}}/m_{\mathrm{B}}$ are shown in Figure 2. Events are selected if they satisfy $\log\Delta R^* > 0.67$ and $p_{\mathrm{T}}^{\mathrm{HC}}/m_{\mathrm{B}} > 0.4$. If more than one VLB candidate fulfills the event selection requirements in an event at this point, the one with the lowest $p_{\mathrm{T}}^{\mathrm{B}}/m_{\mathrm{B}}$ is chosen, as the search targets a low-p_{T}, high-mass decaying particle. Finally, since the signal event topology involves a forward spectator quark, events are required to have at least one small-R jet with $p_{\mathrm{T}} > 40$ GeV and $|\eta| > 2.5$.

The search is restricted to a data sub-sample of maximal signal purity by defining a signal region (SR) on the grounds of the HC mass, required to range from 105 GeV and 135 GeV, and its b-tagging category, required to be `H2T2B`. The lower-purity sample of `H2T1B` events passing the full event selection and the HC mass requirement is preserved for the purpose of validating the background modelling procedure. A comprehensive breakdown of all pre-selection, event reconstruction and kinematic selection criteria is available in Table 1.

The full event selection efficiency on simulated signal events varies as a function of the coupling strength κ regulating the resonance width. For a $m_B = 1200$ GeV, $\kappa = 1.0$ benchmark signal, approximately 9.5% of all simulated $B \rightarrow bH$ events (inclusive with respect to the Higgs decay) have one eligible VLB candidate, but only 0.9% eventually pass the kinematic selection outlined above to enter the signal region. The main factors affecting the reconstruction and selection efficiency are the large-R jet p_{T} threshold, set at 480 GeV to ensure 100% trigger efficiency, the triple b-tagging efficiency and the requirement of signal region events to feature at least one jet in the forward region of the detector.

Table 1: Summary of all pre-selection, reconstruction and kinematic selection steps leading up to the full definition of the signal region and a number of orthogonal control data samples that will be used for validation purposes.

Pre-Selection							
$\geqslant$ 1 large-R Jet, $p_{\mathrm{T}} > 480$ GeV							
No Leptons & no $\gamma\gamma$ pairs with $m_{\gamma\gamma} \in [105,\ 160]$ GeV							
$\geqslant$ 2 associated track jets to large-R jet, $\geqslant$ 1 b-tagged track jet							
$\geqslant$ 1 small-R jet with $p_{\mathrm{T}} > 300$ GeV							
ΔR(small-R jet, large-R jet) > 2.0							
HC Reconstruction							
Any large-R jet with $p_{\mathrm{T}} > 480$ GeV							
$\geqslant$ 2 ghost-matched track jets with $p_{\mathrm{T}} > 50$ GeV							
Pass collinear veto							
Highest b-tag multiplicity: 2 track jets				Highest b-tag multiplicity: 1 track jet			
Select candidate with largest m_{HC}							
VLB Candidate Reconstruction							
HC + small-R jet, p_{T}(small-R jet) > 480 GeV							
ΔR(small-R, large-R) > 2.5							
Kinematic Selection							
$\log \Delta R^{*} > 0.67$							
$p_{\mathrm{T}}^{HC}/m_{B} > 0.4$							
$m_{\mathrm{HC}} \in [105,\ 135]$ GeV							
$\geqslant$ 1 Forward Jets		= 0 Forward Jets		$\geqslant$ 1 Forward Jets		= 0 Forward Jets	
Small-R jet b-tagging status							
Tag	No Tag	Tag	No Tag	Tag	No Tag	Tag	No Tag
SR	Control Samples						



(a) (b)

Figure 2: (a) $\log\Delta R^*$ distributions for the 1.2 TeV VLB signal and background; and (b) $p_T\,(HC)/m_B$ distributions for the 2 TeV VLB signal and background. The background is approximated here by data events that pass the above event selection, except the criteria related to the two displayed variables, and additionally the small-R jet is required to not be b-tagged, leading to a very signal-depleted data sample functioning as proxy for background of the high-purity data sample. Furthermore, the displayed events belong to a 150 GeV-wide window centered on the signal resonance pole mass. All histograms are normalised to the same area for an easier comparison of the shapes. The shaded areas on the data points refer to the statistical uncertainties only, while the shaded error bars about the signal distribution refer to the statistical uncertainty on the MC simulation.

6 Data-driven background modelling

As mentioned in Section 3, the background in this analysis is largely dominated (over 90%) by events featuring QCD production of multiple jets (multijet) in the final state, with the majority of the remainder traceable to t-quark pair production. The well-established ABCD background modelling method is used to carry out a fully data-derived determination of the expected background. The shape of the m_{B} distribution, which is used as the discriminating variable in the statistical analysis of the data, is likewise estimated through a modified ABCD-like procedure.

The two event properties used to define the ABCD partitioning, which are required to be uncorrelated to produce an unbiased estimation of the background, are (a) the b-tagging classification of the small-R jet in the VLB candidate; and (b) the presence of forward jets in the event (see Figure 3). Within the context of this analysis, events passing the full selection described in Section 5 fall in region A, while the remaining regions contain events with no forward jets (B and D) and/or a non-b-tagged, small-R jet in the VLB candidate (C and D). Given the present ABCD partitioning, the A-region data sample where $m_{HC} \in [105,\ 135]$ GeV is predicted to have maximal signal purity. For this reason, A-region data events were temporarily removed from the analysis (henceforth "*blinded*") to avoid a bias in the modelling of the SM background originating from the possible presence of signal.

Provided that the two properties used to split the events in the ABCD regions are uncorrelated, the ratio of background events in regions A and B would be equal to that in regions C and D. From this identity it follows that the number of expected background events in region A, $\tilde{N}_{\mathrm{A}}$, can be estimated to be $\tilde{N}_{\mathrm{A}} = k_{\mathrm{fwd}} N_{\mathrm{B}}$, where $k_{\mathrm{fwd}} = N_{\mathrm{C}}/N_{\mathrm{D}}$.

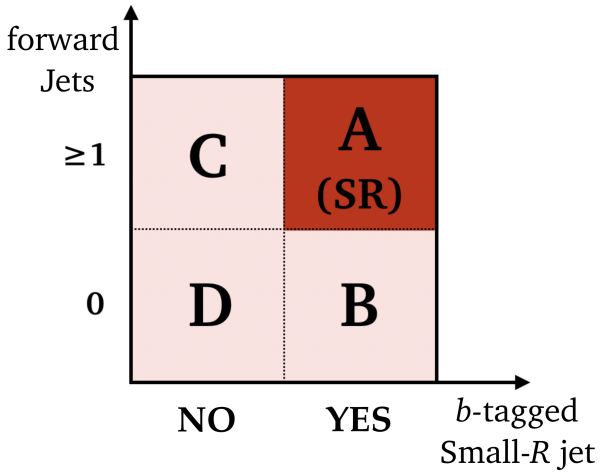


Figure 3: Schematic representation of the ABCD partition employed in this search.

While the two event variables used to define the ABCD partitioning are required to be as uncorrelated as possible, a level of residual correlation is expected and accounted for by computing the value of the correlation-sensitive R_{corr} estimator, defined as:

$$R_{\text{corr}} = \frac{N_{\text{A}}}{N_{\text{B}}} \cdot \frac{N_{\text{D}}}{N_{\text{C}}} \,.$$

The value of R_{corr} can only be computed in support data samples, where data is not blinded in any ABCD region. Three such samples are defined by the present analysis. The first sample contains events selected with the same selection criteria as in Section 5, except that the HC invariant mass is required to be in the range $m_{HC} \in [135,\ 200]$ GeV, which defines the Validation Sideband (VS) mass window. This event sample is labelled "`H2T2B_VS`", to distinguish it from the main analysis sample, where the HC falls in the Higgs mass window (signal region), labelled "`H2T2B`". The other two control samples comprise events that satisfy the same selection criteria as `H2T2B` events, except that only one of the track jets associated with the HC is b-tagged. These events are labelled "`H2T1B`" with the `SR` and `VS` labels when falling respectively in the Higgs mass window or in the Higgs mass sideband. All these control samples have significantly lower sensitivity than the `H2T2B` sample, owing to their smaller signal content and larger expected background, and are estimated only to be sensitive to VLB production cross-sections that are already excluded by previous searches.

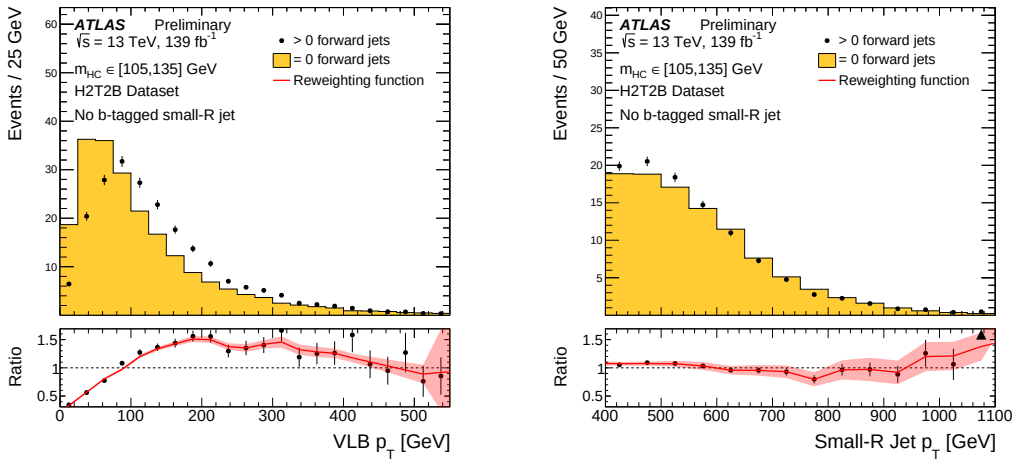
The values for R_{corr} computed in the three aforementioned control sample are consistent within their statistical uncertainties:

$$R_{\text{corr}}^{\text{H2T2B\_VS}} = 1.11 \pm 0.05 \qquad R_{\text{corr}}^{\text{H2T1B\_VS}} = 1.12 \pm 0.02 \qquad R_{\text{corr}}^{\text{H2T1B\_SR}} = 1.11 \pm 0.03 \,.$$

These values, consistently greater than one, imply a slight underestimation of the background yield by the uncorrected ABCD method as a result of the residual correlation between the ABCD axes. Consequently, a correction to the background prediction in the signal region (region A of `H2T2B`) is implemented by scaling the transfer factor k_{fwd} by R_{corr} itself:

$$\tilde{N}_{\text{A}} = R_{\text{corr}}^{\text{H2T2B\_VS}} \cdot k_{\text{fwd}} \cdot N_{\text{B}} \,.$$

Additionally, the event sample in region B is reweighted using per-event weights derived by comparing the distributions of two kinematic variables in regions C and D, to compensate for differences between events with 0 (or no) forward jets and at least one forward jet. The two variables elected to serve for this purpose are the VLB p_{T} and the p_{T} of the small-R jet participating in the VLB reconstruction, which display the largest discrepancies between data distributions in regions C and D, as shown in Figure 4. The weights in each of the two variables are calculated from the bin-by-bin ratios of the normalised distributions, with an additional non-parametric smoothing applied using Gaussian kernel regression [61] to smooth out the effects of any statistical fluctuations. The template ratio regression method used to extract smooth, continuous event weight functions comes naturally with an associated $\pm 1\sigma$ uncertainty band, as displayed in Figure 4. The overall weights are the product of the weights extracted from the two kinematic variables. After this reweighting procedure, all kinematic distributions in region D are found to be in good agreement with those in region C, indicating a satisfactory level of closure for the kinematic reweighting method. The uncertainties on the event weights applied to each B-region event are propagatated to the final background model for signal region data and assumed as systematic uncertainties on the shape and the yield of the SM background prediction for signal region data.



(a) (b)

Figure 4: Comparison of the distributions in ABCD regions C and D for the two variables used to calculate the correction weights: (a) the VLB p_T and (b) the small-R jet p_T. All events shown belong to the Higgs mass window $m_{HC} \in [105,\ 135]$ GeV, but only C-region events have one or more forward jets. The normalization scale factor k_{fwd} is already applied to D-region events. The non-parametric corrections and uncertainties are shown with the red line and pink shaded area, respectively, in each of the ratio plots. The product of corrections in (a) and (b) are applied to B-region events. The solid black triangles indicate points that fall outside the vertical range of the figure. The last bin of every distribution incorporates the overflow.

6.1 Effect of signal contamination on the background model

In an ideal scenario, the ABCD partitioning is tuned in such a way as to contain the large majority of signal events within the blinded target region, with only minimal spillage in the remaining three regions. A perfectly clean signal separation is however impossible in practice due to inefficiencies in flavour tagging and in the identification of the spectator quark as a forward jet. This results in a non-negligible fraction of the predicted signal yield to fall into regions B, C and D, potentially affecting the modelling of the multijet

background. A breakdown of the distribution of data and signal events across the ABCD plane is displayed in Table 2. The signal spillage in regions C and D is respectively of the order of 2% and 0.2%, and results

Table 2: Event yields in the four ABCD regions in the Higgs mass window ($m_{HC} \in [105,\ 135]$ GeV) in the reference $m_B = 1.2$ TeV $\kappa = 0.3$ (B, Y) isospin doublet signal hypothesis and in data. The expected signal region data yield (ABCD-region A), corresponding to the model for the SM background contribution in the signal region, is displayed in bold face with its associated uncertainty.

$m_B = 1.2$ TeV, $\kappa = 0.3$

fwd jets \ b-tags	2	3
≥ 1	96	125
$= 0$	58	78

Data

fwd jets \ b-tags	2	3
≥ 1	5310	**257± 25**
$= 0$	23800	972

in an effect on the value of the scaling factor k_{fwd} or the event consistently smaller than their established statistical uncertainties. The number of signal events in region B is however non-negligible, and would translate in an overestimation of the signal region background of approximately 15% of the predicted A-region signal yield.[2] An MC-derived correction of the background model is implemented to compensate for the potential loss in sensitivity caused by this effect, as described in greater detail in Section 8.

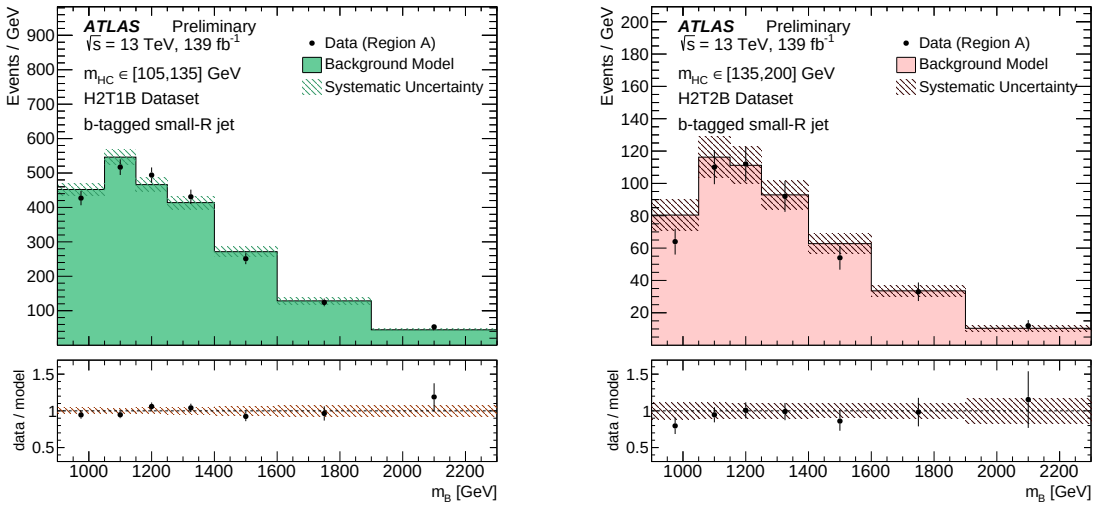
6.2 Background model validation

The full background modelling procedure is applied to the previously defined control samples, the `H2T2B_VS` and `H2T1B` samples, and the resulting predictions compared with the never-blinded A-region data for such samples, thereby serving as validation samples for the background estimation procedure before it is applied to the high-sensitivity signal region data. In the context the `H2T1B` validation sample, the looser requirement on the number of b-tagged track jets associated to the Higgs candidate results in a number of predicted QCD events approximately one order of magnitude larger than the estimated `H2T2B` signal region background yield. The larger available statistics, and the consequentially smaller statistical fluctuations on both the background prediction and the data, renders the `H2T1B` sample a primary tool for evaluating the performance of the background modelling procedure before it can be applied on the signal region.

Figure 5 shows the agreement between the predicted SM background and the data in the validation samples. The shaded areas about the background prediction represent the total systematic uncertainty on the background prediction in sample, defined as the quadrature sum of the individual contributions from: kinematic reweighting on VLB p_{T} and small-R p_{T}, R_{corr} scaling and the statistical fluctuations of the B-region data sample propagated through the modelling procedure (as explained in greater detail in Section 7).

In each validation sample the agreement between the SM background prediction and the data is evaluated via the calculated χ^2 score of the data with respect to the predicted background. No statistically significant or consistently observed discrepancy between data and prediction is detected in any of the validation samples.

[2] The background overestimation caused by signal spillage in region B is estimated as $N_{\text{cont}} \simeq N_B^{\text{sig}} \times k_{\text{fwd}} \times R_{\text{corr}}$.



(a) (b)

Figure 5: Comparison of the VLB reconstructed invariant mass distribution between ABCD-region A data (points) and predicted background (solid histogram) in (a) the `H2T1B` data sample and (b) in the validation sideband of the `H2T2B` sample. The shaded areas about the background prediction account for the pre-fit total systematic uncertainty on either background prediction. The last bin of each distribution incorporates the overflow.

7 Systematic uncertainties

The prediction of the expected signal from the simulated signal samples is affected by uncertainties in the modelling of the reconstruction and calibration techniques used in this analysis. These uncertainties do not affect the background prediction, which is entirely data-derived. The systematic uncertainties in the background modelling stem from the uncertainty in the residual correlation correction, ΔR_{corr} and the uncertainty in the procedure used to derive the event weights for the kinematic shape corrections, described in Section 6, which are applied to B-region events in order to generate the background prediction for the signal region.

7.1 Signal uncertainties

The uncertainties in the p_{T} scale of small and large-R jets (JES), as well as the mass scale for large-R jets (JMS), are evaluated by combining information from the detector reconstruction performance in MC-derived studies as well as in LHC data (*in situ*) [44, 62]. The uncertainty in the jet p_{T} resolution [44] (JER) is obtained by smearing the jet p_{T} in simulated signal events using a p_{T}- and η-dependent parametrization of the jet energy resolution derived in momentum imbalance measurements. The smeared jet p_{T} distribution is then propagated through the event reconstruction and selection, and its effect on the final observable is ultimately evaluated. This procedure gives rise to a one-sided systematic uncertainty, which is then symmetrized before input to the statistical fit. The JES and JER uncertainties for both small-R and large-R jets are modelled in the fit as respectively 30 and 13 individual nuisance parameters describing the individual sub-sources of uncertainty for the scale and resolution of the reconstructed jet energy. Nuisance parameters regulating the same physical source of uncertainty for both small-R and large-R jets are treated as fully correlated.

The large-R jet mass resolution uncertainty [63] (large-R JMR) is obtained by smearing the jet mass using a Gaussian function with a width determined in MC-derived studies on the reconstructed t-quark, W and Higgs boson mass peaks. The combined impact of all these uncertainties on the signal yield varies from 12% to 18%, with the large-R jet mass resolution dominating both at the low (10%) and high (17%) end of the investigated VLB resonance mass range.

The efficiency of flavour tagging in MC events is corrected to match the observed efficiency in data by means of data-derived correction factors calculated centrally in bins of the jet p_{T} and η and applied to MC events. The uncertainties associated with this correction in the signal Monte Carlo simulation are estimated by varying the η- and p_{T}-dependent correction factors applied to each jet within a range that reflects the systematic uncertainty in the measured b-tagging efficiency in data [54]. The impact of the b-tagging uncertainties on the expected signal yield is found to be between 2.8% and 3.1% across the VLB mass range of the search. Uncertainties on the b-tagging efficiency sharing the same physical source are assumed to be fully correlated across small-R and large-R jets.

The uncertainty in the modelling of the $\log\Delta R^*$ variable is assessed by propagating a conservative uncertainty on the track-jet p_{T}, assuming the worst case of a single-track track-jet and using the ATLAS track p_{T} resolution [64], through the event selection step that involves $\log\Delta R^*$. This leads to an asymmetric $+3\%/-6\%$ uncertainty on the expected signal yield. A possible uncertainty originating from mismodelling in the showering and hadronization processes is investigated by comparing the $\log\Delta R^*$ distribution shape in simulated Pythia8 and Herwig7 multijet events in the kinematic region targeted by the analysis. No substantial discrepancy is observed, leading to a negligible systematic uncertainty.

The impact of a number of theoretical uncertainties on the modelling of signal events is investigated. The uncertainty introduced by neglecting higher-order contributions to the renormalization group equation, as well as the uncertainty on the factorization scale, are propagated through the full reconstruction and event selection procedure to assess the impact on the predicted signal yield and shape of the discriminating variable. This procedure was applied to a number of signal resonance mass and width hypotheses covering the theoretical phase space targeted by the search, for a more comprehensive picture. As conservative 4% uncertainty is introduced to account for the renormalization scale, while the impact of the factorization scale uncertainty is found to be 5% across the mass range.

Due to the limited size of the simulated signal samples, the statistical uncertainty on the MC signal simulations is accounted for in the fit by means of as many uncorrelated nuisance parameters as are the bins of the final observable. For each bin, the $\pm1\sigma$ variation is taken to be the well-known Poisson uncertainty on a weighted event count $\sum_{i,\ i\in bin} w_i^2$, with w_i being the weight assigned to every event belonging to the bin under consideration.

Finally, the uncertainty in the combined 2015–2018 integrated luminosity is 1.7% [65], obtained using the LUCID-2 detector [66] for the primary luminosity measurements.

7.2 Background uncertainties

The systematic uncertainty on the background prediction comes entirely from the sources of uncertainty introduced within the data-derived background estimation procedure outlined in Section 6.

The per-event kinematic reweighting factors come with an associated uncertainty in the form of $\pm1\sigma$ weight variations derived from the Kernel regression confidence bands. The reweighting based on the VLB p_{T} introduces a relatively flat uncertainty on the background m_{B} distribution, ranging from 3.5% at

the low-mass end of the reconstructed VLB candidate invariant mass range ($m_{\mathrm{B}} = 950$ GeV) to 4.5% at the high-mass range ($m_{\mathrm{B}} = 2250$ GeV). The second reweighting, derived from the small-R jet p_{T} spectrum, likewise leads to an uncertainty ranging from ±5% in the low-mass region to ±12% at the high-mass end.

A second source of uncertainty, amounting to ±5% on the predicted background yield, is associated with the statistical uncertainty on the cross-ratio R_{corr} by which the background model is rescaled to compensate for the residual correlation between the two event variables chosen to form the ABCD axes.

A third source of uncertainty on the background accounts for a possible dependence of both the per-event kinematic reweighting factors and k_{fwd} on the HC mass, which may result in a biased prediction as such quantities are computed globally within the Higgs mass window $m_{\mathrm{HC}} \in [105,\ 135]$ GeV. An alternative set of weights and k_{fwd} is extracted from the regions C and D of the validation sideband (`H2T2B_VS`) sample and applied to the region B events in the Higgs mass window (signal region) sample to produce an alternative estimate for the background. The difference between the VS-based prediction and the nominal one is taken as a systematic uncertainty on the reweighting model stability with respect to the HC mass. The predicted impact on the predicted SM background yield ranges from below 1% for the lowest values of the probed mass range to about 4% above 2 TeV.

Finally, the shape of the background prediction is affected by the bin-wise Poisson fluctuations in the region B data, out of which the model is constructed. The uncertainty on each m_{B} bin on the final model originating from this source is computed by propagating the Poisson uncertainty of the relevant bin through the whole modelling procedure, and accounts for an uncertainty on the predicted yields ranging from $7 - 8$% in the low mass region to $10 - 20$% in the high mass region, where it is the dominant source of uncertainty on the background prediction.

Table 3 displays a comprehensive summary of the impact of each category of systematic uncertainties on the yield of each of the affected event samples. To highlight the shifting impact and hierarchy of the systematic sources throughout the investigated resonance mass range, relative systematic uncertainties are displayed for both a relatively low-mass 1.2 TeV resonance and for the highest available simulated mass sample (2.0 TeV).

8 Statistical analysis

The statistical interpretation of the data is carried out using a binned maximum-likelihood fit to the invariant mass distribution, m_{B}, of the reconstructed VLB candidate, based on the expected signal and background yields. The likelihood model is defined as:

$$\mathcal{L} = \prod_i P_{\mathrm{pois}}(n_i|\lambda_i) \times \mathcal{N}(\boldsymbol{\theta})$$

where $P_{\mathrm{pois}}(n_i|\lambda_i)$ is the Poisson probability to observe n_i events when λ_i events are expected in bin i of the m_{B} distribution, and $\mathcal{N}(\boldsymbol{\theta})$ is a series of Gaussian or log-normal distributions for the nuisance parameters, $\boldsymbol{\theta}$, corresponding to the systematic uncertainties related to the signal and background yields in each bin. The λ_i are expressed as $\lambda_i = \mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta})$, with μ being the signal strength, defined as a signal cross-section in units of the theoretical prediction, left as the free-floating parameter of interest to be determined by the fit, and $s_i(\boldsymbol{\theta})$ and $b_i(\boldsymbol{\theta})$ being the expected numbers of signal and background events, respectively. Nuisance parameters are allowed to float in the fit within their constraints $\mathcal{N}(\boldsymbol{\theta})$, and thus

Table 3: Relative effect of the pre-fit systematic uncertainties of each group of sources on the yields of the predicted background and two simulated signals with a VLB mass respectively equal to 1.2 or 2.0 TeV, and $\kappa = 0.3$. The size of the systematic uncertainties affecting the background is similarly provided for reconstructed VLB candidate invariant of 1.2 TeV and 2.0 TeV.

	VLB Mass = 1.2 TeV	VLB Mass = 2.0 TeV	Background	
Systematic			At 1.2 TeV	At 2.0 TeV
b-tagging	3%	3%	⟋	
JER	1%	2%	⟋	
JES	4%	4%	⟋	
Large-R JES	3%	1%	⟋	
Large-R JMR	13%	21%	⟋	
Large-R JMS	2%	4%	⟋	
Luminosity	2%	2%	⟋	
MC Statistics	5%	6%	⟋	
Renorm. scale	4%		–	
Factor. scale	5%		–	
VLB p_T weights	–	–	4%	5%
$R = 0.4$ jet p_T weights	–	–	5%	12%
Model Stability	–	–	1%	4%
R_{corr}	–	–	5%	
B-region statistics	–	–	8%	18%
TOTAL	**15%**	**23%**	**11%**	**23%**

alter the normalization and shape of the signal m_{B} distribution as well as the shape of the background m_{B} distribution.

The potential effect on the background model of the signal spillage outside the signal region (as introduced in Section 6) is compensated by subtracting the signal contribution in the B-region from the background at the fitting stage. Before subtraction, B-region signal events are scaled by $k_{\text{fwd}} \times R_{\text{corr}}$ and kinematically reweighted using the same weight functions employed in the background modelling phase, in order to reproduce their actual contribution to the signal region background model. The sample subtraction is effectively accomplished by assigning a negative signal strength parameter μ_{cont} to the contamination sample and constraining it to the fitted value of the parameter of interest μ by imposing $\mu_{\text{cont}} = -\mu$. This operation ensures that both the signal yield and the magnitude of the resulting background contamination are simultanously determined in the statistical fit.

As with the principal signal MC simulations, the subtracted contamination sample is affected by the systematic uncertainties relating to the theoretical and experimental understanding of VLB single production and its detection at ATLAS. As a consequence, the nuisance parameters regulating such uncertainties are

set to be fully correlated across all the MC-derived signal samples. In addition, the contamination sample is also affected by the custom uncertainties introduced by the kinematic reweighting procedure applied to the ABCD-Region B signal simulation to generate the estimate for the effect of the signal spillage onto the background model. The nuisance parameters regulating the impact of such uncertainties in the fit are set to be fully correlated with those of the same source in the QCD background model.

The information about μ is extracted from a likelihood fit to the data under the signal-plus-background hypothesis, using a test statistic based on the profile likelihood ratio. The distributions of the test statistic under the signal-plus-background and background-only hypotheses are obtained using asymptotic formulae [67]. The systematic uncertainties with the largest post-fit impact on μ at $m_{\mathrm{B}} = 1.2$ TeV and $m_{\mathrm{B}} = 2.0$ TeV are the uncertainties on the R_{corr} correction at the low-mass end and the uncertainty on the reweighting on the small-R jet p_{T} in the higher mass range. The level of agreement between the observed data and the background prediction is assessed by computing the local p-value p_0 for the observed value of the *profile likelihood* test statistics given its asymptotic distribution in the background-only hypothesis. The p-value p_0 is defined as the probability to observe an excess at least as large as the one observed in data, under the background-only hypothesis. Expected and observed upper limits are set at 95% confidence level (CL) on the cross-section for single production of a VLB into the decay channel under investigation fraction ($\sigma(pp \rightarrow B \rightarrow bH)$) using the $\mathrm{CL_s}$ prescription [68].

9 Results

Following the statistical interpretation scheme outlined in Section 8, a binned maximum-likelihood fit is performed on the signal region data. The data yield in the signal region, as well as the background yields and their relative uncertainties after the statistical fit are referenced in Table 4. The post-fit comparison between the m_B spectrum in data and in the background model is displayed in Figure 6. The yield displayed corresponds to the expected production for the (B, Y) isospin doublet scenario, used as the benchmark model for a VLB signal in the present search.Owing to the generally good agreement between the data and its prediction, the maximum-likelihood fit does not produce noticeable shifts of the nuisance parameters regulating the background model from their nominal values. The largest discrepancy between the data and the predicted SM background contribution is observed at a resonance mass $m_B = 1.3$ TeV, corresponding to a local p-value p_0 equal to 0.03 and a local significance of 1.9σ for the $m_B = 1.3$ TeV, $\kappa = 0.3$ hypothesis.

As no significant excess is observed in data over the prediction for the SM background, 95% CL exclusion limits on the production cross-section are drawn by means of the *profile likelihood* technique introduced in Section 8 for the two theoretical benchmark scenarios determined by the coupling strength κ. Figure 7 displays mass-dependent 95% exclusion limits on the production cross-section for two differing values of the coupling strength κ. The red solid line overlaid to the exclusion plots represents the theoretical single-production cross-section for a VLB occurring as a (B, Y) weak isospin doublet as calculated at NLO.

The ranges of the resonance mass excluded at 95% CL for the two values of κ considered are inferred from Figure 7 as the mass intervals for which the expected or observed excluded cross-section is smaller than the relevant theoretical prediction. VLB resonances with coupling strength $\kappa = 0.3$ are excluded across the full mass range investigated, whereas for a $\kappa = 0.25$, the exclusion is limited to the 1.0 TeV $< m_B <$ 1.28 TeV and 1.46 TeV $< m_B <$ 2.0 TeV resonance mass ranges.

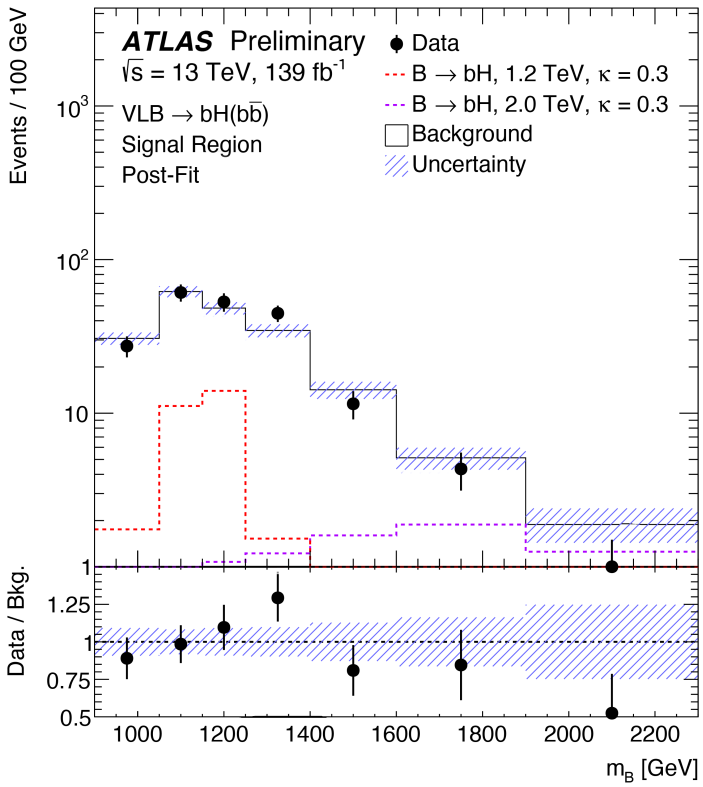


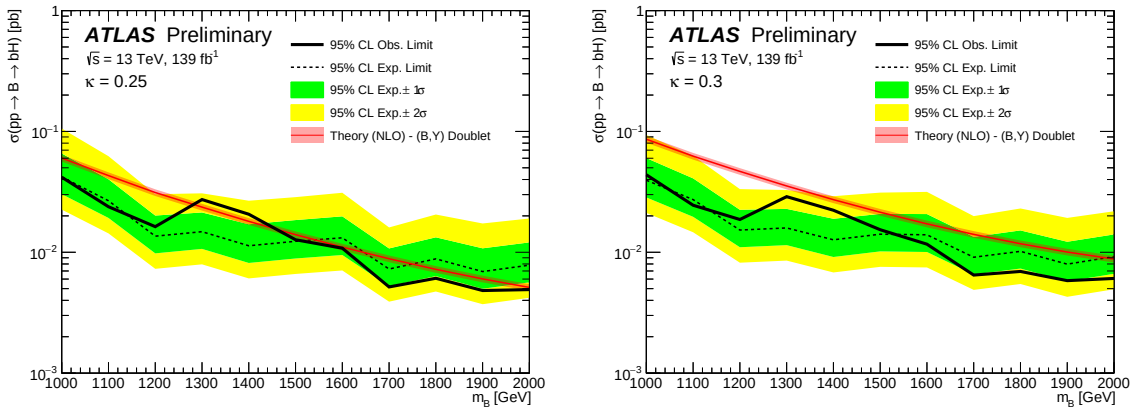
Figure 6: Comparison between data (points) and SM data-derived background model (solid white) after a binned maximum-likelihood background-only fit in the signal region. The shaded area represent the total systematic uncertainty on the background model. The expected contributions of a VLB signal of mass equal to 1.2 TeV and 2.0 TeV, occurring as part of a (B, Y) doublet with $\kappa = 0.3$, are overlaid to the data distribution as the red and violet dashed lines respectively. The lower panel shows the ratio between the bin-wise data and background yields. The last bin of both distributions incorporeate the overflow.

The cross-section limits presented above are then interpreted in the form of mass-dependent 95% exclusion limits on the value of the phenomenological Lagrangian parameter c_Z, representing the coupling constant between the VLB and the SM Z boson. As the (B, Y) doublet state implies $\xi_W = 0$ and $\xi_Z = \xi_H = 0.5$, it follows from Eq. 1 that in such scenario $c_W = 0$ and the values of c_Z and c_H scale with each other, justifying the choice to intepret the search results in terms of limits on the value of a single phenomenological parameter.

Figure 8 displays the expected and observed mass-dependent limits on the value of c_Z for a VLB quark occurring as part of (B, Y) doublet. Such limits are computed by examining the excluded production cross section observed for each available signal benchmark, as arranged on a $(m_B,\ \kappa)$ grid. For each available resonance mass, the lowest value of κ for which the signal is excluded is converted into a value of c_Z through the $c_Z = m_Z/m_W \times \kappa$ identity for a (B, Y) doublet. The dashed gray lines overlaid to Figure 8 offer a visual connection between the values of the phenomenological coupling c_Z and the relative width of the VLB resonance, from which it can be inferred that the search is able to exclude a VLB doublet resonance with relative width larger than 5% in the 1.0 TeV < m_B < 1.75 TeV mass range.

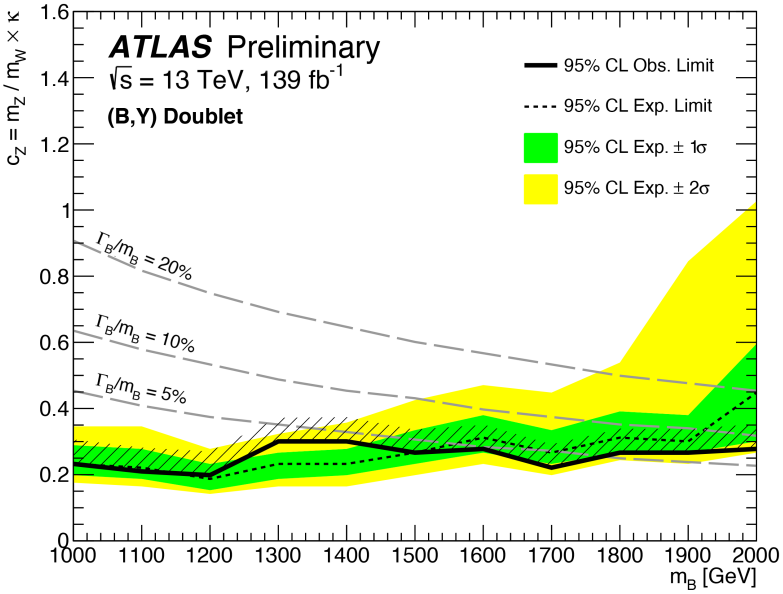
Sample		Pre-Fit	Post-Fit
VLB, 1.2 TeV	$\kappa = 0.15$	35 ± 4	–
	$\kappa = 0.25$	90 ± 7	–
	$\kappa = 0.3$	125 ± 21	–
VLB, 1.6 TeV	$\kappa = 0.15$	10 ± 1	–
	$\kappa = 0.25$	33 ± 3	–
	$\kappa = 0.3$	49 ± 3	–
VLB, 2.0 TeV	$\kappa = 0.15$	5 ± 1	–
	$\kappa = 0.25$	11 ± 1	–
	$\kappa = 0.3$	18 ± 2	–
Background estimation		257 ± 25	260 ± 17
Data		**262**	

Table 4: Comparison of the yields of the SM background and of a VLB of varying mass and κ before and after a background-only fit. The post-fit signal yield is accordingly zero by definition (conditional fit with $\mu = 0$). The observed data yield in the signal region is also displayed.



(a) (b)

Figure 7: Expected (dashed line) and observed (solid black line) 95% CL exclusion limits on the production cross-section of a single VLB into the Higgs decay mode. The green and yellow bands about the expected limits represent respectively the $\pm 1\sigma$ and $\pm 2\sigma$ confidence intervals for the observed limit to fall assuming a background-only scenario. The red solid line and shaded area respectively trace the evolution of doublet production cross-section as a function of the resonance mass and its relative theoretical uncertainty account for uncertainties in the renormalization and factorization scales. Limits are presented for two values of the coupling strength κ: 0.25 (a) and 0.3 (b).



(a)

Figure 8: Mass-dependent expected and observed 95% CL exclusion limits on the c_Z phenomenological coupling in the (B, Y) doublet scenarios. The range of excluded values of the c_Z coupling can be inferred as the area above the observed exclusion limit line. The gray dashed lines represent the loci of $(m_B,\ \kappa)$ configurations resulting in constant values of the relative resonance width Γ_B/m_B.

10 Conclusions

A search is presented for single-production of a vector-like B quark decaying to a SM b-quark and a Standard Model Higgs boson, itself decaying to a pair of bottom quarks. The search is carried out on proton–proton collision data with center-of-mass energy $\sqrt{s}$ = 13 TeV collected at the ATLAS experiment at the LHC between 2015 and 2018, corresponding to a total integrated luminosity of 139 fb^{-1}.

Evidence of a signal is sought in the form of an excess in the spectrum of the reconstructed invariant mass of the selected vector-like B quark candidate, defined as the combination a large-R and a small-R jet satisfying the criteria outlined in Section 4. No statistically significant evidence of a signal is found, with the selected data being in a good agreement with the expected Standard Model contribution as modelled using a fully data-derived procedure.

The search results are presented in the form of mass-dependent limits at the 95% CL on the B production cross-section under the assumption of the VLB arising as part of a (B, Y) weak isospin doublet.

With no statistically significant deviation from the expected SM background contribution, and with previously-published searches for VLB pair-production at the LHC having established the lower bound for the existence of a VLB in the (B,Y) scenario at 1.2 TeV, this search was able to substantially expand the region of the VLB parameter space probed at collider experiments, excluding the presence of a VLB resonance relative width greater than for masses between 1.0 TeV and 2.0 TeV in the hypothesis where $\kappa = 0.3$, and in the 1.0 TeV $< m_B <$ 1.28 TeV and 1.46 TeV $< m_B <$ 2.0 TeV mass ranges for $\kappa = 0.25$. The aforementioned limits translate into a resonance originating from single-production of a VLB (B, Y) doublet with relative width larger than 5% being excluded in the 1.0 TeV $< m_B <$ 1.76 TeV mass range.

References

[1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1, arXiv: 1207.7214 [hep-ex] (cit. on p. 2).

[2] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30, arXiv: 1207.7235 [hep-ex] (cit. on p. 2).

[3] L. Susskind, *Dynamics of Spontaneous Symmetry Breaking in the Weinberg-Salam Theory*, Phys. Rev. **D20** (1979) 2619 (cit. on p. 2).

[4] A. de Gouvea, D. Hernandez and T. M. P. Tait, *Criteria for Natural Hierarchies*, Phys. Rev. **D89** (2014) 115005, arXiv: 1402.2658 [hep-ph] (cit. on p. 2).

[5] M. J. Dugan, H. Georgi and D. B. Kaplan, *Anatomy of a Composite Higgs Model*, Nucl. Phys. **B254** (1985) 299 (cit. on p. 2).

[6] K. Agashe, R. Contino and A. Pomarol, *The Minimal composite Higgs model*, Nucl. Phys. **B719** (2005) 165, arXiv: hep-ph/0412089 [hep-ph] (cit. on p. 2).

[7] M. Schmaltz and D. Tucker-Smith, *Little Higgs review*, Ann. Rev. Nucl. Part. Sci. **55** (2005) 229, arXiv: hep-ph/0502182 [hep-ph] (cit. on p. 2).

[8] J. A. Aguilar-Saavedra, R. Benbrik, S. Heinemeyer and M. Pérez-Victoria, *Handbook of vectorlike quarks: Mixing and single production*, Phys. Rev. **D88** (2013) 094010, arXiv: 1306.0572 [hep-ph] (cit. on p. 2).

[9] G. Cacciapaglia et al., *Heavy Vector-like Top Partners at the LHC and flavour constraints*, JHEP **03** (2012) 070, arXiv: 1108.6329 [hep-ph] (cit. on p. 2).

[10] O. Matsedonskyi, G. Panico and A. Wulzer, *On the Interpretation of Top Partners Searches*, JHEP **12** (2014) 097, arXiv: 1409.0100 [hep-ph] (cit. on pp. 2, 3).

[11] A. Roy, N. Nikiforou, N. Castro and T. Andeen, *Novel interpretation strategy for searches of singly produced vectorlike quarks at the LHC*, Phys. Rev. D **101** (11 2020) 115027, URL: https://link.aps.org/doi/10.1103/PhysRevD.101.115027 (cit. on p. 2).

[12] M. Buchkremer, G. Cacciapaglia, A. Deandrea and L. Panizzi, *Model Independent Framework for Searches of Top Partners*, Nucl. Phys. **B876** (2013) 376, arXiv: 1305.4172 [hep-ph] (cit. on pp. 2, 3).

[13] ATLAS Collaboration, *Search for single production of vector-like quarks decaying into Wb in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **05** (2019) 164, arXiv: 1812.07343 [hep-ex] (cit. on p. 3).

[14] ATLAS Collaboration, *Combination of the searches for pair-produced vector-like partners of the third-generation quarks at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. Lett. **121** (2018) 211801, arXiv: 1808.02343 [hep-ex] (cit. on p. 3).

[15] ATLAS Collaboration, *Search for pair- and single-production of vector-like quarks in final states with at least one Z boson decaying into a pair of electrons or muons in pp collision data collected with the ATLAS detector*, Phys. Rev. **D98** (2018) 112010, arXiv: 1806.10555 [hep-ex] (cit. on p. 3).

[16] CMS Collaboration, *Search for single production of vector-like quarks decaying to a Z boson and a top or a bottom quark in proton-proton collisions at* $\sqrt{s}$ = 13 *TeV*, JHEP **05** (2017) 029, arXiv: 1701.07409 [hep-ex] (cit. on p. 3).

[17] CMS Collaboration, *Search for single production of vector-like quarks decaying to a b quark and a Higgs boson*, JHEP **06** (2018) 031, arXiv: 1802.01486 [hep-ex] (cit. on p. 3).

[18] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003 (cit. on p. 4).

[19] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, ATLAS-TDR-19; CERN-LHCC-2010-013, 2010, URL: https://cds.cern.ch/record/1291633 (cit. on p. 4).

[20] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, JINST **13** (2018) T05008, arXiv: 1803.00844 [physics.ins-det] (cit. on p. 4).

[21] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, Eur. Phys. J. C **77** (2017) 317, arXiv: 1611.09661 [hep-ex] (cit. on p. 4).

[22] ATLAS Collaboration, *ATLAS data quality operations and performance for 2015–2018 data-taking*, JINST **15** (2020) P04003, arXiv: 1911.04632 [physics.ins-det] (cit. on p. 4).

[23] M. Cacciari, G. P. Salam and G. Soyez, *The anti-*k_t *jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph] (cit. on pp. 4, 5).

[24] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer and T. Stelzer, *MadGraph 5 : Going Beyond*, JHEP **06** (2011) 128, arXiv: 1106.0522 [hep-ph] (cit. on p. 4).

[25] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. **B867** (2013) 244, arXiv: 1207.1303 [hep-ph] (cit. on pp. 4, 5).

[26] ATLAS Collaboration, *ATLAS Pythia 8 tunes to* 7 *TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: https://cds.cern.ch/record/1966419 (cit. on pp. 4, 5).

[27] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462** (2001) 152 (cit. on p. 4).

[28] S. Agostinelli et al., *GEANT4 – a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250 (cit. on p. 4).

[29] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv: hep-ph/0409146 (cit. on p. 5).

[30] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv: 0709.2092 [hep-ph] (cit. on p. 5).

[31] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv: 1002.2581 [hep-ph] (cit. on p. 5).

[32] R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040, arXiv: 1410.8849 [hep-ph] (cit. on p. 5).

[33] M. Cacciari, M. Czakon, M. Mangano, A. Mitov and P. Nason, *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Phys. Lett. **B710** (2012) 612, arXiv: 1111.5869 [hep-ph] (cit. on p. 5).

[34] M. Beneke, P. Falgari, S. Klein and C. Schwinn,
Hadronic top-quark pair production with NNLL threshold resummation,
Nucl. Phys. **B855** (2012) 695, arXiv: 1109.1536 [hep-ph] (cit. on p. 5).

[35] P. Bärnreuther, M. Czakon and A. Mitov, *Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to $q\bar{q} \to t\bar{t} + X$*, Phys. Rev. Lett. **109** (2012) 132001,
arXiv: 1204.5201 [hep-ph] (cit. on p. 5).

[36] M. Czakon and A. Mitov,
NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels,
JHEP **12** (2012) 054, arXiv: 1207.0236 [hep-ph] (cit. on p. 5).

[37] M. Czakon and A. Mitov,
Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders,
Comput. Phys. Commun. **185** (2014) 2930, arXiv: 1112.5675 [hep-ph] (cit. on p. 5).

[38] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159,
arXiv: 1410.3012 [hep-ph] (cit. on p. 5).

[39] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton–proton collision data*, ATLAS-CONF-2016-024, 2016,
URL: https://cds.cern.ch/record/2157687 (cit. on p. 5).

[40] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **76** (2016) 292, arXiv: 1603.05598 [hep-ex]
(cit. on p. 5).

[41] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data*, JINST **14** (2019) P12006,
arXiv: 1908.00005 [hep-ex] (cit. on p. 5).

[42] ATLAS Collaboration,
Jet reconstruction and performance using particle flow with the ATLAS Detector,
Eur. Phys. J. C **77** (2017) 466, arXiv: 1703.10485 [hep-ex] (cit. on p. 5).

[43] ATLAS Collaboration,
Jet energy measurement with the ATLAS detector in proton–proton collisions at $\sqrt{s} = 7$ TeV,
Eur. Phys. J. C **73** (2013) 2304, arXiv: 1112.6426 [hep-ex] (cit. on p. 6).

[44] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$, TeV with the ATLAS detector*, (2020), arXiv: 2007.02645 [hep-ex] (cit. on pp. 6, 13).

[45] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D **96** (2017) 072002,
arXiv: 1703.09665 [hep-ex] (cit. on p. 6).

[46] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, Eur. Phys. J. C **76** (2016) 581,
arXiv: 1510.03823 [hep-ex] (cit. on p. 6).

[47] ATLAS Collaboration,
Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector,
Eur. Phys. J. C **77** (2017) 580, arXiv: 1705.02211 [hep-ex] (cit. on p. 6),
Erratum: Eur. Phys. J. C **77** (2017) 712.

[48] ATLAS Collaboration,
Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1,
Eur. Phys. J. C **77** (2017) 490, arXiv: 1603.02934 [hep-ex] (cit. on p. 6).

[49] D. Krohn, J. Thaler and L.-T. Wang, *Jet Trimming*, JHEP **02** (2010) 084,
arXiv: 0912.1342 [hep-ph] (cit. on p. 6).

[50] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*,
Phys. Rev. **D48** (1993) 3160, arXiv: hep-ph/9305266 (cit. on p. 6).

[51] S. Catani, Y. Dokshitzer, M. Seymour and B. Webber,
Longitudinally-invariant k_T-clustering algorithms for hadron-hadron collisions,
Nucl. Phys. **B406** (1993) 187, ISSN: 0550-3213,
URL: https://www.sciencedirect.com/science/article/pii/055032139390166M
(cit. on p. 6).

[52] ATLAS Collaboration, *Performance of jet substructure techniques in early $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector*, ATLAS-CONF-2015-035, 2015,
URL: https://cds.cern.ch/record/2041462 (cit. on p. 6).

[53] ATLAS Collaboration, *Jet energy measurement and its systematic uncertainty in proton–proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Eur. Phys. J. C **75** (2015) 17,
arXiv: 1406.0076 [hep-ex] (cit. on p. 6).

[54] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **79** (2019) 970,
arXiv: 1907.05120 [hep-ex] (cit. on pp. 6, 14).

[55] ATLAS Collaboration, *Performance of jet substructure techniques for large-R jets in proton–proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, JHEP **09** (2013) 076,
arXiv: 1306.4945 [hep-ex] (cit. on p. 6).

[56] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, Phys. Lett. **B659** (2008) 119,
arXiv: 0707.1378 [hep-ph] (cit. on p. 6).

[57] M. Cacciari, G. P. Salam and G. Soyez, *The Catchment Area of Jets*, JHEP **04** (2008) 005,
arXiv: 0802.1188 [hep-ph] (cit. on p. 6).

[58] D. Krohn, J. Thaler and L.-T. Wang, *Jets with Variable R*, JHEP **06** (2009) 059,
arXiv: 0903.0392 [hep-ph] (cit. on p. 6).

[59] ATLAS Collaboration, *Variable Radius, Exclusive-k_T, and Center-of-Mass Subjet Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS*, ATL-PHYS-PUB-2017-010, 2017,
URL: https://cds.cern.ch/record/2268678 (cit. on p. 6).

[60] ATLAS Collaboration, *Identification of boosted Higgs bosons decaying into b-quark pairs with the ATLAS detector at 13 TeV*, Eur. Phys. J. C **79** (2019) 836, arXiv: 1906.11005 [hep-ex]
(cit. on p. 6).

[61] E. A. Nadaraya, *On estimating regression*, Theory of Probability and its Applications **9** (1964) 141
(cit. on p. 11).

[62] ATLAS Collaboration, *Identification of high transverse momentum top quarks in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, JHEP **06** (2016) 093, arXiv: 1603.03127 [hep-ex]
(cit. on p. 13).

[63] ATLAS Collaboration, *Jet mass reconstruction with the ATLAS Detector in early Run 2 data*, ATLAS-CONF-2016-035, 2016, URL: https://cds.cern.ch/record/2200211 (cit. on p. 14).

[64] ATLAS Collaboration, *Track Reconstruction Performance of the ATLAS Inner Detector at $\sqrt{s}$ = 13 TeV*, ATL-PHYS-PUB-2015-018, 2015, URL: https://cds.cern.ch/record/2037683 (cit. on p. 14).

[65] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s}$ = 13 TeV using the ATLAS detector at the LHC*, ATLAS-CONF-2019-021, 2019, URL: https://cds.cern.ch/record/2677054 (cit. on p. 14).

[66] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13** (2018) P07017, URL: https://doi.org/10.1088/1748-0221/13/07/p07017 (cit. on p. 14).

[67] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011) 1554, Erratum: Eur. Phys. J. C **73** (2013) 2501, arXiv: 1007.1727 [physics.data-an] (cit. on p. 17), Eur. Phys. J. C **73** (2013) 2501.

[68] A.L. Read, *Presentation of search results: the CL_s technique*, J. Phys. G **28** (2002) 2693 (cit. on p. 17).