Search for resonances in the mass distribution of jet pairs with one or two jets identified as b-jets in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

The ATLAS Collaboration

DOI
10.1016/j.physletb.2016.05.064

Publication date
2016

Document Version
Final published version

Published in
Physics Letters B

License
CC BY

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Download date: 11 Aug 2023
Search for resonances in the mass distribution of jet pairs with one or two jets identified as $b$-jets in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

The ATLAS Collaboration*

1. Introduction

Many extensions to the Standard Model (SM) predict the existence of new massive particles that couple to quarks or gluons. If produced in proton–proton ($pp$) collisions at the Large Hadron Collider (LHC), these new beyond-the-SM (BSM) particles could decay into quarks ($q$) or gluons ($g$), creating resonant excesses in the two-jet (dijet) invariant mass distributions [1–6]. If the new particle couples to the $b$-quark and decays into $bb$, $bq$ or $b\bar{q}$ pairs, a dedicated search for dijet resonances with one or both jets identified as originating from a $b$-quark (“$b$-jet”) could greatly increase the signal sensitivity.

Prior resonance searches in dijet events containing $b$-jets were performed by the CDF [7] and CMS [8,9] experiments, probing the mass ranges 200–750 GeV and 1–4 TeV respectively. Excited heavy-flavour quarks have been investigated in alternative decay modes as well [10]. No BSM phenomena have been observed yet. The increase in centre-of-mass energy of the $pp$ collisions at the LHC from $\sqrt{s} = 7$ and 8 TeV to 13 TeV provides a new energy regime in which to search for such a heavy resonance. This is particularly true for heavy states coupling to $b$-quarks from the proton sea, when compared to states produced by valence quarks. The parton luminosity to create a 2 TeV object increases by an additional factor of 2–3 for $bb$ and $b\bar{q}$ over $q\bar{q}$ and $qq$ pairs, when increasing the centre-of-mass energy from 8 TeV to 13 TeV. The total production rate for dijet BSM signals can become large enough to allow a good signal sensitivity even with a relatively small data sample. In this paper the search for a new narrow resonance decaying to $b$-quarks with the ATLAS detector, using 3.2 $fb^{-1}$ integrated luminosity of proton–proton collisions at $\sqrt{s} = 13$ TeV, is reported. The mass range 1.1–5.0 TeV is probed.

The results are interpreted in the context of two benchmark processes shown in Fig. 1: an excited heavy-flavour quark $b^*$ and a new gauge boson $Z'$. Excited quarks are a consequence of quark compositeness models that were proposed to explain the general structure and mass hierarchy of quarks [11,12]. The $Z'$ boson arises in many extensions to the SM with an additional $U(1)$ group. Two $Z'$ models are considered, one with SM-like fermion couplings in the Sequential Standard Model (SSM) and a leptophobic $Z'$ model [13,14]. All benchmark model decays are expected to result in a narrow resonance superimposed on a smoothly falling dijet invariant mass distribution. This search divides the events into samples with one or two jets identified as $b$-jets to enhance the signal sensitivity to the benchmark models $b^* \rightarrow bg$ and $Z' \rightarrow bb$. In addition, the results are interpreted in the context of possible Gaussian-shaped signal contributions to the dijet invariant mass spectra where one or both jets are identified as $b$-jets. The results, presented in terms of the cross section times acceptance times branching ratio $(\sigma \times A \times BR)$, are quoted for contributions with widths of up to 15% of the resonance mass.

2. The ATLAS detector

The ATLAS experiment [15] at the LHC is a multi-purpose particle detector with a forward–backward symmetric cylindrical...
3. Data and simulated event samples

The data used in this analysis were collected by the ATLAS detector in pp collisions at the LHC with a centre-of-mass energy of 13 TeV during 2015. Events were recorded using a jet-based trigger requiring at least one jet with a transverse momentum \( p_T \) of at least 360 GeV. The full dataset corresponds to an integrated luminosity of 3.2 fb\(^{-1}\) with an associated uncertainty of 5% after applying quality criteria to the data. The measurement of the integrated luminosity is derived, following a methodology similar to that detailed in Ref. [18], from a calibration of the luminosity scale using a pair of \( x-y \) beam-separation scans.

Monte Carlo (MC) simulated event samples are used to model the expected signals and study the composition of SM background processes. The QCD dijet process is simulated with Pythia8 [19] using the A14 tuned parameter set [20] for the modelling of the parton shower, hadronization and underlying event. The leading-order (LO) parton distribution function (PDF) set NNPDF2.3 [21] is used for the generation of events. The renormalization and factorization scales are set to the average transverse momentum \( p_T \) of the two leading jets. The EvtGen decay package [22] is used for bottom and charm hadron decays.

The three signal samples are generated with Pythia8 using the A14 set of tuned parameters and the NNPDF2.3 PDF set. For the \( b^* \) model, the compositeness scale is set to the excited-quark mass and 85% of decays are to \( b\bar{g} \). The remaining decay modes are to a SM gauge boson (\( Z \) boson, \( W \) boson or photon) and a \( b\)-quark. In the SSM \( Z' \) model, the \( Z' \) boson has the same couplings to SM fermions as the SM \( Z \) boson and the bottom quark decay branching ratio \( BR(Z' \rightarrow b\bar{b}) = 13.8\% \). The leptophobic \( Z' \) model differs by having vanishing couplings to leptons. The corresponding value of \( BR(Z' \rightarrow b\bar{b}) = 18.9\% \). For both, only decays to \( b\)-quark pairs are simulated. The intrinsic decay width is \( \Gamma \sim 0.6\% \) of the resonance mass for the \( b^* \) model and \( \Gamma \sim 3\% \) of the mass for the SSM \( Z' \) boson.

The generated samples are processed with the ATLAS detector simulation [23], which is based on the GEANT4 package [24]. To account for additional pp interactions from the same or close-by bunch crossings, a number of minimum-bias interactions generated using Pythia8 and the MSTW2008LO PDF [25] set are superimposed onto the hard scattering events. The MC samples are re-weighted to match the collisions per bunch crossing observed in the data.

4. Event reconstruction and selection

Jets are reconstructed from noise-suppressed topological clusters [26] of energy deposited in the calorimeters using the anti-\( k_t \) algorithm [27] with a radius parameter of 0.4. Jet energies and directions are corrected by the jet calibrations derived from \( \sqrt{s} = 13 \) TeV simulation, and \( pp \) collision data taken at \( \sqrt{s} = 8 \) TeV and \( \sqrt{s} = 13 \) TeV, as described in Ref. [28]. Jets are required to have \( p_T > 50 \) GeV. Events where any of the three leading jets with \( p_T > 50 \) GeV is compatible with non-collision background or calorimeter noise are removed. Events are preselected in the same way as in the dijet analysis of Ref. [5], requiring that the \( p_T \) of the leading jet is greater than 440 GeV to ensure full trigger efficiency. An additional requirement is placed on the jet pseudorapidity, \( |\eta| < 2.4 \), to ensure tracker coverage for \( b\)-jet identification. The analysis is performed in an unbiased dijet mass range of \( m_{bb} > 1.1 \) TeV. To reduce the background from QCD multijet processes and enhance \( s \)-channel processes, the rapidity difference \( y^* = (y_1 - y_2)/2 \) between the two leading jets is required to be \( |y^*| < 0.6 \). Here \( y_1 \) and \( y_2 \) are the rapidities of the leading and sub-leading jet respectively.

To identify jets originating from \( b \)-hadrons (\( b\)-tagging) a multivariate algorithm that combines information about the impact parameters of inner detector tracks associated with the jet, the presence of displaced secondary vertices, and the reconstructed flight paths of \( b \) and \( c \)-hadrons associated with the jet [29,30] is employed. The \( b\)-tagging working point with 85% efficiency, as determined when integrating over all jets in a simulated sample of \( tt \) events, is chosen because it gives the highest signal sensitivity. As the average jet energies in this analysis are larger than in \( tt \) events and the \( b\)-tagging efficiency drops with jet \( p_T \), the per-jet efficiencies are below 85% and are roughly 50% for jets with a \( p_T \) of 1 TeV.

The \( b \)-jet identification algorithm is applied to the two leading jets, and events are categorized as inclusive, single \( b\)-tagged “1b” or double \( b\)-tagged “2b”, in order to enhance the sensitivity of dif-

---

1 ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upwards. Cylindrical coordinates \((r, \phi)\) are used in the transverse plane, \( \phi \) being the azimuthal angle around the z-axis. The pseudorapidity is defined in terms of the polar angle \( \theta \) as \( \eta = -\ln \tan(\theta/2) \). Angular distance is measured in units of \( \Delta R = \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2} \).
different signal compositions. The "1b" category is defined inclusively, including events from the "2b" category.

The per-event b-tagging efficiencies as functions of the reconstructed invariant mass are shown in Fig. 2. Efficiencies are for benchmark models with different b and Z' resonance masses, after the event selection is applied. The tagging efficiency for Z' events in the inclusive "1b" category is higher than for b events because this process has more b-quarks in the final state. At high mass, the gluon from the decay of the b* has a higher probability to produce a bb-pair, which causes the event tagging efficiency to be comparable for the Z' and b*. The tagging efficiency in the "2b" category is about 2.5 times lower at low mass and a factor 10 lower at high mass compared to the inclusive "1b" category for the same Z' events. The average light-flavour jet rejection factor for jets passing the kinematic selection is approximately 30 for jet transverse momenta up to ∼ 1 TeV.

Correction factors are applied to the simulated event samples to compensate for differences between data and simulation in b-tagging efficiencies and mis-identification rates. These corrections were derived from comparisons of samples of b-quark-enriched events in data and simulation [31]. The average combined signal acceptance and efficiency is around 20% for the b* benchmark in the "1b" category and drops with increasing mass from 9% at 1.5 TeV to 2% at 5.0 TeV for the Z' signals for the "2b" category.

5. Dijet mass spectrum

The dijet mass spectrum is predominantly composed of jets arising from QCD interactions. Fig. 3 shows the comparison between data and Pythia8 multijet MC simulation. The simulated distributions are normalized to the number of events observed in the data in each category separately. The bin widths are chosen to approximate the mll resolution as derived from simulated QCD processes, which range from 3% at 1.0 TeV to 2% at 5.0 TeV. Good agreement between the shapes of the Pythia8 multijet predictions and the data is found. The inclusive distribution, not restricted in the inner tracking detector acceptance, was analysed in Ref. [5].

The dijet background estimation does not rely on the simulation as it is obtained directly from a fit to the mll distribution. The following parameterization ansatz is adopted to fit the distribution in the mll range from 1.1 TeV up to the last data point of the inclusive, "1b" and "2b" mass distributions separately.

\[
f(z) = p_1(1 - z)^p_2 z^{p_3},
\]

where \( p_i \) are free parameters and \( z = m_{ll}/\sqrt{s} \). This ansatz was used in previous searches [5] and is found to provide a satisfactory fit to leading-order Pythia8 multijet MC simulation at \( \sqrt{s} = 13 \) TeV. Employing Wilks’ theorem [32], a log-likelihood statistic is used to confirm that no additional parameters are needed to model these distributions for a data set as large as the one used for this analysis.

The results of the fits are shown in Fig. 4. The fits of this ansatz to the data without considering systematic uncertainties return \( p \)-values of 0.73, 0.90 and 0.66 for the inclusive, “1b” and “2b” categories respectively. The \( p \)-value was calculated as a goodness-of-fit measure using a \( \chi^2 \) test statistic determined from pseudo-experiments.

The lower panels of Fig. 4 show the significances of bin-by-bin differences between the data and the fit. These equivalent Gaussian significances are calculated from the Poisson probability, considering only statistical uncertainties.

The statistical significance of any localized excess in the dijet mass distribution is quantified using the BUMPHunter algorithm [33]. The algorithm compares the binned \( m_{ll} \) distribution of the data to the fitted background estimate, considering contiguous mass intervals in all possible locations, from a width of two bins to one-half of the distribution. For each interval in the scan, it computes the significance of any excess found. The algorithm identifies the intervals 1493–1614 GeV in the “1b” and 3596–3827 GeV in the “2b” sample, indicated by the two vertical lines in Fig. 4, as the most discrepant intervals. The statistical significance of these.
outcomes is evaluated using the ensemble of Poisson outcomes across all intervals scanned, by applying the algorithm to many pseudo-data samples drawn randomly from the background fit. Without including systematic uncertainties, the probability that fluctuations of the background model would produce excesses at least as significant as those observed in the data, anywhere in the distribution, is greater than 60% in the “1b” and “2b” categories. Thus, there is no evidence of localized contributions to the mass distribution from BSM phenomena.

6. Systematic uncertainties

Uncertainties in the parameters of the fitted background function Eq. (1) are evaluated by fitting the ansatz to pseudo-data drawn via Poisson fluctuations around the fitted background model. The uncertainty in the prediction in each $m_b$ bin is taken to be the root mean square of the function value for 10000 generated pseudo-experiments. To estimate an uncertainty due to the choice of background parameterization, one additional degree of freedom, $2\Delta \log(\sigma)$, is appended as a multiplicative factor to the nominal ansatz (Eq. (1)), and the difference between the estimated parameters from the two fits is taken as an uncertainty.

The uncertainty in the jet energy scale is estimated using various methods in 8 TeV data, corrected to the new centre-of-mass energy by taking the difference between the 8 TeV and 13 TeV runs into account using MC simulation [28]. The jet energy scale uncertainty used in this analysis relies on a set of three nuisance parameters [34]. For untagged jets, it is within the range 1–5% for jet transverse momenta greater than 200 GeV.

The relative additional uncertainty in the energy scale of b-tagged jets is estimated using the MC samples and verified with data following the method described in Ref. [35]. The ratio $r_{\text{h}}$ is the sum of track transverse momenta inside the jet to the total jet transverse momentum measured in the calorimeter is used for this estimate. The double ratio of $r_{\text{h}}$ from data and simulation is formed and compared for inclusive jets and b-jets.

relative additional uncertainty for jets with $200 < p_T < 800$ GeV is found to be less than 2.6%, and this value is subsequently used in the higher $p_T$ regions. This relative uncertainty is applied in addition to the nominal jet energy scale uncertainty. The maximum uncertainty for b-tagged jets is estimated to be 6% and is conservatively applied to all $p_T$ regions.

The uncertainty in the jet energy resolution is estimated using the same method as the untagged jet energy scale uncertainty and relies on an additional Gaussian smearing of the reconstructed jet energies in MC simulation. For jets with $p_T > 50$ GeV, the uncertainty is less than 2%.

The uncertainty introduced by the application of the b-tagging algorithm is the largest systematic uncertainty in the analysis. The uncertainty in the measured tagging efficiency of b-jets is estimated by studying $t\bar{t}$ events in 13 TeV data for jet $p_T$ up to 200 GeV [31]. The uncertainties in the measured rate of mistagging c-jets and light-flavour jets are estimated in 8 TeV data. The uncertainties are extrapolated to 13 TeV, taking into account the addition of the new IBL system as well as reconstruction and tagging improvements. An additional term is included to extrapolate the measured uncertainties to the high-$p_T$ region of interest. This term is calculated from simulated events by considering variations on the quantities affecting the b-tagging performance such as the impact parameter resolution, percentage of poorly measured tracks, description of the detector material, and track multiplicity per jet. The dominant effect on the uncertainty when extrapolating at high-$p_T$ is related to the different tagging efficiency when smearing the tracks impact parameters based on the resolution measured in data and simulation. The difference in the impact parameter resolution is due to effects from alignment, dead modules and additional material not properly modelled in the simulation. The impact of the b-tagging efficiency uncertainty increases with jet $p_T$ and reaches 50% above 2 TeV.
7. Results

Due to the absence of a signal, 95% credibility-level upper limits are set on the cross section for new processes that would produce a contribution to the dijet mass distribution with $b$-tagging. The signal shapes are taken as provided by $b^* \to bg$ and $Z' \to bb$ production processes.

The limits on $b^*$ and $Z'$ cross sections are shown in Figs. 5 and 6. The limits were obtained using a Bayesian method [36]. The Bayesian credible intervals were calculated using a posterior probability density from the likelihood function for the observed mass spectrum obtained by a fit to the background (Eq. (1)), while the signal shape was derived from MC simulations. The limit is interpolated between discrete values of the mass to create a continuous curve. The systematic uncertainties associated with the uncertainty in the integrated luminosity, jet energy scale, jet energy resolution, $b$-tagging and alternative fit functions are all included in the limit-setting.

Fig. 5 shows that the $b^*$ model, with the decay to $g + b(\bar{b})$, is excluded for $b^*$ masses from 1.1 TeV up to 2.1 TeV at leading-order in QCD. Fig. 6 shows that the leptophobic $Z' \to bb$ model with SM-like couplings to quarks is excluded up to 1.5 TeV at leading-order in QCD. The present data are not sufficient to provide an exclusion limit for the SSM $Z'$ model.

As shown in Fig. 7, narrow resonance contributions of various widths with visible cross sections $\sigma \times A \times \text{BR}$ ranging from approximately 0.4 to 0.001 pb are excluded in the mass range 1.5–5.0 TeV. These limits should be used when low-mass off-shell tails from PDFs and non-perturbative effects on the narrow resonance signal shape can be safely truncated or neglected and, after applying the selection described in Section 4, the reconstructed mass distribution approximates a Gaussian distribution. For a detailed description of how to use these limits, see the instructions in Ref. [37]. To estimate the $b$-tagging efficiency, invariant-mass-dependent correction factors as given in Fig. 2 can be used.

8. Summary

A search for new resonances decaying to jets with a single or double $b$-tag in $p p$ collisions with the ATLAS detector at the LHC is presented. The dataset corresponds to an integrated luminosity of 3.2 fb$^{-1}$ collected at $\sqrt{s} = 13$ TeV in 2015. The studies use the dijet invariant mass $m_{jj}$ in the range of 1.1–5.0 TeV with $b$-tagging applied to the leading and sub-leading jets and categorize the events according to their $b$-jet multiplicity.

The background from jets initiated by $b$-quarks is well described by the leading-order parton-shower models. The dijet background is estimated using a template-matching method with MC simulation.
ground is also well described by the analytic fit function with three parameters which is used in the light-flavour dijet analysis [5].

No evidence of a significant excess of events is found compared to the expectations of the Standard Model. The largest observed local excess is less than 2σ for both the single and double b-tag channels.

The expected contribution from the b* model is excluded in the mass range 1.1–2.1 TeV at leading-order in QCD using the single b-jet channel. The results can not exclude contributions from the SSM Z′ → bb model in the mass range 1.1–5.0 TeV in the double b-jet channel. For the leptophobic Z′ model with SM-like couplings to quarks, the mass range 1.1–1.5 TeV is excluded at leading-order in QCD in this channel.

This analysis excludes generic high-mass particles decaying to two jets, where one or two jets originate from b-quarks, with visible cross sections ranging from 0.4 to 0.001 pb in the mass range 1.1–5.0 TeV. The exclusion limits are applicable for resonances exhibiting a Gaussian shape and width similar to the b∗ or Z′ models. The limits were calculated assuming that the width of the Gaussian signal is 15%, 10% or 7% of its mass.

Acknowledgements

We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions without whom ATLAS could not be operated efficiently.

We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWF and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN, CONICYT, Chile; CAS, MOST and NSFC, China; COLCIENCIAS, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS, CEA-DSM/IRFU, France; DFG and AvH Foundation, Germany; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; FOM and NWO, Netherlands; RCN, Norway; MNiSW and NCN, Poland; FCT, Portugal; MINEFA, Romania; MES of Russia and NRC KI, Russian Federation; JINR; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DST/NRF, South Africa; MINECO, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNS and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TAEK, Turkey; STFC, United Kingdom; DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, the Canada Council, Canarie, CRC, Compute Canada, FQRNT, and the Ontario Innovation Trust, Canada; EPLANET, ERC, FP7, Horizon 2020 and Marie Skłodowska-Curie Actions, European Union; Investissements d’Avenir Labex and Idex, ANR, Région Auvergne and Fondation Partager le Savoir, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF; BRF, GIF and Minerva, Israel; BRF, Norway; Generalitat de Catalunya, Generalitat Valenciana, Spain; the Royal Society and Leverhulme Trust, United Kingdom.

The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN and the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK) and BNL (USA) and in the Tier-2 facilities worldwide.

References


The ATLAS Collaboration
