Search for resonances below 1.2 TeV from the mass distribution of $b$-jet pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

The ATLAS Collaboration

Abstract

Searches for resonances using the dijet invariant mass spectrum with two jets identified as $b$-jets are performed with the ATLAS detector at the Large Hadron Collider. The dataset consists of an integrated luminosity of 3.2 fb$^{-1}$ of proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. The dijet mass distribution from 0.57 TeV to 1.2 TeV is studied. No significant deviations from the Standard Model expectation have been observed in the data. These results are used to exclude $Z'$ bosons at the 95% credibility level in the inspected mass range. Contributions of a Gaussian shaped signal with visible cross sections ranging from approximately 0.3 to 0.02 pb are also excluded.
1. Introduction

Many extensions to the Standard Model (SM) predict the existence of new massive particles that couple to quarks. 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), creating resonant excesses in the two-jet (dijet) invariant mass distributions [1–6]. If the new particle strongly couples to the b-quark, as favoured by many BSM models, a dedicated search for dijet resonances with two jets containing b-hadrons (“b-jet”) could greatly increase the sensitivity to signals such as the one arising from the Z’ model discussed below.

Resonance searches using two b-jet events have been performed by the CDF experiment at the Tevatron [7] and the CMS experiment at √s = 7 and 8 TeV. The increase in centre-of-mass energy of the pp collisions at the LHC to 13 TeV provides a new energy regime and heavy resonances in the mass range 1.1–5.0 TeV have been probed by the ATLAS experiment [8]. For resonances below 1.1 TeV, due to the limitations of the trigger bandwidth, an alternate trigger strategy is needed. Two recent searches from ATLAS and CMS [9, 10] employ trigger level objects to search for new resonances in the inclusive jet distribution and reach down to 425 GeV and 500 GeV in dijet mass respectively. An alternative strategy recently reported by the ATLAS collaboration [11] evades the trigger limitations by examining data where the light resonance is boosted in the transverse direction via recoil from a high-pT initial-state radiation photon to probe dijet masses as low as 250 GeV.

In this paper the mass range 0.57–1.2 TeV is probed with the ATLAS detector, using a b-jet trigger and looking at proton-proton collisions at √s = 13 TeV with an integrated luminosity of 3.2 fb−1. The results are interpreted in the context of a benchmark process, a new gauge boson Z’ decaying to b-quarks as shown in Figure 1. The Z’ boson arises in many extensions to the SM with an additional U(1) group. This report considers a Z’ model with SM-like fermion couplings in the Sequential Standard Model (SSM) [12]. This results in a narrow resonance superimposed on a smoothly falling invariant mass distribution originated from multijet backgrounds. In addition, the results are interpreted in the context of possible Gaussian-shaped signal contributions to the b-tagged dijet invariant mass spectra. The results, presented in terms of the cross section times kinematic acceptance times b-tagging efficiency times branching ratio (σ × A × ϵ × BR), are quoted for different hypotheses of signal widths.

![Figure 1: Leading-order Feynman diagram for the process considered: q̅q → Z’ → b̅b.](image-url)
2. The ATLAS detector

The ATLAS experiment [13] at the LHC is a multi-purpose particle detector with a forward-backward symmetric cylindrical geometry and a near $4\pi$ coverage in solid angle. It consists of an inner tracking detector surrounded by a thin superconducting solenoid providing a 2 T axial magnetic field, electromagnetic and hadronic calorimeters, and a muon spectrometer. The inner tracking detector covers the pseudorapidity range $|\eta| < 2.5$. It consists of, in ascending order of radius from the beam-line, silicon pixel, silicon microstrip, and transition radiation tracking detectors. The pixel detectors are crucial for $b$-jet identification. For the second LHC data-taking period at $\sqrt{s} = 13$ TeV, a new inner pixel layer, the Insertable B-Layer (IBL) [14, 15], was added at a mean sensor radius of 3.2 cm from the beam-line. Lead/liquid-argon (LAr) sampling calorimeters provide electromagnetic (EM) energy measurements with high granularity. A hadronic (steel/scintillator-tile) calorimeter covers the central pseudorapidity range ($|\eta| < 1.7$). The end-cap and forward regions are instrumented with LAr calorimeters for EM and hadronic energy measurements up to $|\eta| = 4.9$. A two-level trigger system, using custom hardware followed by a software-based level, is used to reduce the event rate to a maximum of around 1 kHz for offline storage.

3. Data and simulated event samples

The high-quality data were recorded in stable beam conditions and with all relevant sub-detectors. Events are selected by a trigger requiring at least two $b$-tagged jets [16], one with a transverse momentum $p_T > 175$ GeV and the other one with $p_T > 60$ GeV.

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 [17] using the A14 parameter set [18] for the modelling of the parton shower, hadronization and underlying event. The leading-order parton distribution function (PDF) set NNPDF2.3 [19] is used for the generation of events. The renormalisation and factorisation scales are set to the average transverse momentum $p_T$ of the two leading jets. The EvtGen decay package [20] is used for bottom and charm hadron decays. $Z'$ boson signal events are simulated for masses of 0.6, 0.8, 1 and 1.25 TeV and are generated with the same setup of the QCD dijet samples. The intermediate mass points are created from parameter interpolations of a template fit to the generated signal points. $Z'$ bosons are only allowed to decay into a pair of $b$-quarks. The $Z'$ signal is normalised to the next-to-leading order cross section in the strong coupling [21–23] using a neutral vector boson model [24] with SM-like quark couplings. The $Z'$ width is set to 3%.

The generated samples are processed with the ATLAS detector simulation [25], which is based on the GEANT4 package [26]. 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 [27] set are superimposed onto the hard scattering events.

---

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 $\Delta R \equiv \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2}$.
4. Object reconstruction and event selection

Groups of contiguous calorimeter cells (topological clusters) are formed based on the significance of the cell energy deposit over calorimeter noise [28]. Topological clusters are grouped into jets using the anti-$k_t$ algorithm [29] with radius parameter $R = 0.4$. Jet four-momenta are computed by summing over the topological clusters that constitute each jet, treating each cluster as a four-momentum vector with zero mass. Jet calibrations, derived from $\sqrt{s} = 13$ TeV simulation and collision data taken at the same energy, are used to correct the jet energies and directions to those of the particles from the hard-scatter interaction. This calibration procedure, described in Ref. [30], is improved by a data-derived correction to the relative calibration of jets in forward regions compared to central jets.

Charged-particle tracks are reconstructed with a $p_T$ threshold of 400 MeV. Event vertices are formed from these tracks and are required to have at least two tracks. The primary vertex is chosen as the vertex with the largest $\Sigma p_T^2$ of the associated tracks. The $b$-jets are identified ($b$-tagged) by exploiting the relatively long lifetime and large mass of $b$-hadrons. The $b$-tagging methods are based on the presence of tracks with a large impact parameter with respect to the primary vertex, the presence of displaced secondary vertices, and the reconstructed flight paths of $b$- and $c$-hadrons associated with the jet [16, 31]. This information is combined into a single discriminant by employing a multivariate algorithm. This analysis uses a $b$-tagging criterion that, in simulated $t\bar{t}$ events, provides an average efficiency of 70% for $b$-jets and a $c$-jet (light-jet) mis-tag rate less than 20% (1%). At the trigger level, the impact parameter and displaced secondary vertex information is combined in a likelihood. With differences in charged particle reconstruction and the discriminants themselves, $b$-tagging in the trigger (online) is not fully correlated to $b$-tagging in reconstructed events (offline). The $b$-jet trigger efficiencies are measured using data and simulated $t\bar{t}$ events in the dilepton channel, and the simulation efficiencies are corrected to match the data. For the $b$-tagged offline jets, the $b$-jet triggers efficiencies range from 87% at $p_T \sim 100$ GeV to 60% at $p_T \sim 600$ GeV.

Events containing at least two $b$-jets with $|\eta| < 2.4$ are selected for this analysis. The $p_T$ of the leading (highest $p_T$) and sub-leading (second highest $p_T$) $b$-jets should be greater than 230 GeV and 90 GeV, respectively. These criteria ensure the adopted trigger to reach a kinematic plateau. In addition, the leading $b$-jet pair should satisfy $|y^*| < 0.6$ where $y^* = 0.5(y_1 - y_2)$, and $y_1$ ($y_2$) is the rapidity of the leading (sub-leading) $b$-jet. This requirement favours $s$-channel production and reduces the large background contribution from $t$-channel multijet processes.

The per-event $b$-tagging efficiencies as a function of the reconstructed invariant mass of the leading $b$-jet pair ($m_{jj}$) are shown in Figure 2. The efficiencies are for benchmark models with different $Z'$ resonance masses, after the event selection is applied.

The tagging efficiency is dependent on $Z'$ mass, varying between approximately 45% and 15%. The effect of the $b$-tagging selection at the trigger level is also shown in Figure 2, reflecting the differences between the trigger and offline $b$-tagging algorithms. The analysis acceptance varies between 30% and 40% across the inspected mass range and is mainly driven by the $y^*$ and $\eta$ requirements. An additional term in the acceptance as high as 50% is present for low reconstructed invariant masses due to the $p_T$ and $m_{jj}$ requirements. The effect is due to the low-mass off-shell tail in the $Z'$ production mechanism.
5. Invariant mass spectrum of two \( b \)-jets and statistical techniques

The \( m_{jj} \) spectrum is predominantly composed of jets arising from QCD interactions. The bin widths are chosen to approximate the \( m_{jj} \) resolution. The multijet background estimation does not rely on the simulation as it is obtained directly from a fit to the \( m_{jj} \) distribution in data. The following parametrisation ansatz is adopted to fit the distribution in the \( m_{jj} \) range from 0.57 TeV to 1.2 TeV:

\[
f(x) = p_1 (1 - x)^{p_2} x^{p_3 + p_4 (\ln x) + p_5 (\ln x)^2},
\]

where \( p_i \) are free parameters, and \( x = m_{jj} / \sqrt{s} \). The \( x^0 \) term is motivated by the leading-order QCD matrix element, and the \((1 - x)^p\) term is a common parametrisation for the behaviour of parton distributions with the property of vanishing as \( x \) approaches unity. This ansatz was used in previous searches [5, 8].

Employing Wilk’s theorem [32], a log-likelihood statistic is used to determine if the background estimation would be significantly improved by an additional degree of freedom. A significant improvement is delineated as a Wilk’s \( p \)-value less than 0.05. It is found that adding the fourth and fifth parameters does not improve the likelihood significantly, therefore the first three parameters are used for the nominal fit. Additional checks have been performed on leading-order QCD simulated events and confirmed the three-parameter function was able to provide a satisfactory fit.

6. Systematic uncertainties

Systematic uncertainties considered in this analysis can be separated into those affecting the data-driven background estimate and the signal determined from simulation. Neither is the dominant limiting factor on the search sensitivity.
The systematic uncertainties on the background estimate take into account the uncertainty in the choice of a fit function Eq. (1) and the uncertainty in the parameters of the fit. The former is estimated by increasing the number of degrees of freedom for the fit function. The latter is estimated by fitting pseudo-experiments generated with a Poisson statistics around the predicted background.

The uncertainty in the jet energy scale is estimated using untagged jets in 13 TeV data and simulation by following the method described in Ref. [30]. In the jet $p_T$ range considered in this analysis, an additional uncertainty in the energy scale of $b$-jets was proven to be negligible. The total scale uncertainty is approximately 1% across the full mass range considered. The uncertainty in the jet energy resolution is estimated using the same method as for the untagged jet energy scale uncertainty and relies on an additional Gaussian smearing of the reconstructed jet energies in MC simulation. The resolution uncertainty is less than 1.5%.

The uncertainty introduced by the application of the $b$-tagging algorithm is derived from comparisons of samples of $b$-quark-enriched events in data and simulation [16, 33]. 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. The uncertainties are estimated to be approximately 30% and represent the largest systematic uncertainty for jets with $p_T$ below 1 TeV. The uncertainty in the measured tagging efficiency of $b$-jets is estimated by studying $t\bar{t}$ events in 13 TeV data. 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 variation 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 uncertainty on the $b$-jet trigger efficiency is estimated from the relative trigger response between data and simulation in a $b$-jet enriched sample of $t\bar{t}$ events in the dilepton channel. The events are selected with a dedicated lepton trigger also running the jet reconstruction and the $b$-tagging algorithm but without using the $b$-tagging discriminant for the trigger selection. Various sources of uncertainties are considered related to the data statistics, the modelling of the $b$-tagging discriminant and the purity in truth $b$-jets of the sample. The total systematic uncertainty is found to be $p_T$ dependent and to range between approximately 10% and 30% in the mass interval of interest and dominated by the limited data available to determine the high jet $p_T$ efficiency.

The uncertainty on the integrated luminosity is estimated to be ±5%, following a methodology similar to that detailed in Ref. [34].

7. Results

The statistical significance of any localised excess in the $m_{jj}$ distribution is quantified using the BUMP Hunter algorithm [35]. The algorithm compares the binned $m_{jj}$ 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 statistical significance 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. The result of the analysis is shown in Figure 3. The most discrepant region has a $p$-value of 0.70. Thus, no evidence for statistically significant deviations from the background hypothesis is observed.
In the absence of a signal, 95% credibility-level upper limits are set on the cross sections for new processes that would produce a contribution to the dijet mass distribution with $b$-tagging. The limits are obtained using a Bayesian method [36]. The Bayesian credibility intervals are calculated using a posterior probability density from the likelihood function for the observed mass spectrum obtained by the fit to the background in Eq. (1), while the signal shape is taken as provided by the $Z'$ simulated samples. The limits are interpolated between discrete values of the mass to create a continuous curve. The systematic uncertainties associated with the integrated luminosity, jet energy scale, jet energy resolution, $b$-tagging, trigger and alternative fit functions are included in the limit setting. Figure 4 shows the 95% credibility-level upper limits for the SSM $Z'$ model.

Narrow resonance contributions of various widths are also considered and, as shown in Figure 5, visible cross sections ranging from approximately 0.3 to 0.02 pb are excluded in the mass range 0.65–1.15 TeV at the 95% credibility level.

8. Conclusions

A search for new resonances decaying to two $b$-jets in $pp$ 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 of two $b$-jets in the range of 0.57–1.2 TeV.
Figure 4: Observed (filled circles) and expected (dotted line) 95% credibility-level upper limits on the cross section ($\sigma$) times kinematic acceptance ($A$) times $b$-tagging efficiency ($\epsilon$) times branching ratio (BR) for the SSM $Z'$ model. The dashed line shows the signal prediction with the $Z'$ production cross section computed at the next-to-leading order (NLO).

Figure 5: The 95% credibility-level upper limits on the cross section ($\sigma$) times kinematic acceptance ($A$) times $b$-tagging efficiency ($\epsilon$) times branching ratio (BR) for resonances exhibiting a generic Gaussian shape. The circles, squares and triangles correspond to the cases where the width of the Gaussian signal is 10%, 6% and 4% of the signal mass, respectively. The figure also shows as a red line the case where the width is given by the dijet mass resolution.
No evidence of a significant excess of events compared to the expectation of the SM is found. The largest observed local excess is less than $1.5\sigma$. These results are used to exclude SSM $Z'$ bosons in the mass range between 0.65 TeV and 1.0 TeV. Contributions of a Gaussian shaped signal with visible cross sections ranging from approximately 0.3 to 0.02 pb are also excluded.
References


A. Mass spectrum without a signal overlaid

Figure 6 shows the result of the analysis as shown in Figure 3 but without a signal overlaid.

Figure 6: The $m_{jj}$ spectrum overlaid with the fit to the background function together with the results from BumpHunter. The lower panel shows the significance per bin of the data with respect to the background fit, in terms of the number of standard deviations, considering only the statistical fluctuations. The region between the two vertical lines presents the most discrepant region with a $p$-value of 0.70.