Search for light vector resonances decaying to quarks at $\sqrt{s} = 13$ TeV

The CMS Collaboration

Abstract

We perform a search for light, narrow vector resonances decaying to quarks in the mass range from 100-300 GeV produced in association with a high transverse momentum jet using 2.7 fb$^{-1}$ of 2015 $\sqrt{s} = 13$ TeV proton-proton collision data collected by CMS. Novel jet substructure methods and background estimation techniques are employed to search for a resonance in the jet mass distribution originating from a new particle in whose decay the quarks are merged into a single jet. We demonstrate the CMS sensitivity to vector resonances in the mass-coupling phase space not yet explored at the LHC or by any past experiments.
1 Introduction

Searches for narrow resonances in the dijet invariant mass spectrum are a central feature of the physics program of every collider experiment of the past half-century. Theoretically, this is well-motivated due to the many classes of models of new physics predicting resonances with significant couplings to quarks such as string resonances [1, 2]; scalar diquarks [3]; excited quarks [4, 5]; axigluons [6, 7]; color-octet colorons [8]; color octet scalars [9]; new gauge bosons ($W'$ and $Z'$) [10]; Randall-Sundrum gravitons [11]. There has been a long history of searches for dijet resonances. The first results by UA1 [12] and UA2 [13, 14] with $\sqrt{s} = 630$ GeV collisions at the Sp̄pS have been extended to larger values of resonance masses by CDF [15–19] and DØ [20] with the $\sqrt{s} = 1.96$ TeV collisions at the Tevatron, and by ATLAS [21–27] and CMS [28–34] with $\sqrt{s} = 7$ TeV, $\sqrt{s} = 8$ TeV and $\sqrt{s} = 13$ TeV pp collisions at the Large Hadron Collider (LHC). A review of the available results can be found in Ref [35].

This analysis focuses on the search for leptophobic vector ($Z'$) resonances decaying to quarks in the mass region from 100-300 GeV. This mass range has not yet been fully probed, neither at the LHC where current searches probe down to 250 GeV [36], nor at past experiments where searches have been made down to 140 GeV [13, 14]. Results obtained at different collider energies are interpreted in the framework of a specific leptophobic resonance model described in [37]. This model is described with the following Lagrangian:

$$L \sim \frac{g_B}{6} Z'_{\mu} \overline{q} \gamma^\mu q$$

(1)

where $Z'_{\mu}$, $\overline{q}$, and $q$ are the $Z'$ and quark fields, respectively, and $g_B$ is a dimensionless coupling constant and is related to a $U(1)_B$ gauge symmetry of baryon number. We examine this new gauge boson coupling in the perturbative regime where values of total coupling are less than unity and therefore $g_B < 6$. This flavor-independent $Z'$ vector resonance model can be then interpreted as generic searches for other vector-like resonances to quarks as in [37].

With the increase of collision energy and beam intensity at hadron colliders, there has been a loss of sensitivity for light resonances with couplings to quarks and gluons. The main experimental difficulties originate from the large cross section of background multijet events at low dijet mass. The limited resources for recording and storing these data forced all hadron-collider experiments to use tighter trigger requirements to reduce the output rate. In this analysis, in order to overcome these difficulties, we require a high transverse momentum ($p_T$) signal jet recoiling against another high $p_T$ ISR jet. Here, the $Z' \rightarrow q\overline{q}$ system is reconstructed as a single boosted merged jet. We then require the leading $p_T$ jet to have a mass ($m_{Z'}$) and 2-prong substructure consistent with an object decaying into a quark-antiquark pair. Our selection variables are designed to be decorrelated between the jet mass, $p_T$, and jet substructure in order to perform a robust search over a smoothly-falling background jet mass distribution. In Section 4 the background is estimated using a data-driven

Analysis Strategy

We outline here the general strategy of the analysis and the structure of this note. In Section 2 we give a brief overview of the CMS detector. In Section 3 we describe the data and simulated samples and event selection. In order to isolate the $Z'$ signal and overcome trigger restrictions we require a high transverse momentum ($p_T$) signal jet recoiling against another high $p_T$ ISR jet. Here, the $Z' \rightarrow q\overline{q}$ system is reconstructed as a single boosted merged jet. We then require the leading $p_T$ jet to have a mass ($m_{Z'}$) and 2-prong substructure consistent with an object decaying into a quark-antiquark pair. Our selection variables are designed to be decorrelated between the jet mass, $p_T$, and jet substructure in order to perform a robust search over a smoothly-falling background jet mass distribution. In Section 4 the background is estimated using a data-driven
QCD scaling sideband method where the events failing the mass and substructure requirements are used to predict the jet mass distribution from QCD in the signal region. Standard model (SM) candles from the W and Z inclusive processes, also produced in association with a high transverse momentum ISR jet, have a very similar topology to the Z' signal. They are used to validate the analysis method as a signal proxy and further constrain systematic effects related to a potential signal. Section 5 describes the systematic uncertainties for the background and signal contributions. This includes a validation of the Z' tagging techniques using merged jets from W bosons in \( t\bar{t} \) events. Finally, in Section 6, limits are set in the \( g_{B} \) coupling-mass plane in the 100-300 GeV mass range.

\[ \text{Figure 1: An example Feynman diagram of a } Z' \rightarrow q\bar{q} \text{ resonance production with an initial-state radiation gluon.} \]

2 CMS detector

The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the superconducting solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. Forward calorimeters extend the pseudorapidity [38] coverage provided by the barrel and endcap detectors. Muons are measured in gas-ionization detectors embedded in the steel flux-return yoke outside the solenoid. A more detailed description of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, can be found in Ref. [38].

3 Event reconstruction, simulation and selection

This study uses proton-proton collision events from the 2015 Run 2 dataset corresponding to 2.7 fb\(^{-1}\) at \( \sqrt{s} = 13 \text{ TeV} \). Events are selected using a two-tier trigger system. Events satisfying loose jet requirements at the first level (L1) are examined by the high-level trigger (HLT). We use a logical “OR” of the following HLT trigger requirements which make a selection on the total hadronic transverse energy in the event (\( H_T \)) and, in some cases, in conjunction with a selection on the mass of the jet after cleaning it of soft radiation with the jet trimming technique [39] (\( m_{\text{trimmed}} \)): 
- $H_T > 650\,\text{GeV}$ and $m_{\text{trimmed}} > 50\,\text{GeV}$
- $p_T > 350\,\text{GeV}$ and $m_{\text{trimmed}} > 30\,\text{GeV}$
- $H_T > 800\,\text{GeV}$

Jets at the trigger level are reconstructed using the anti-$k_T$ algorithm [40] with a distance parameter of 0.8, which we denote as AK8. The variable $H_T$ is defined as the scalar sum of the $p_T$ of AK8 jets in the event. We determine that this trigger strategy is fully efficient at selecting events offline with a AK8 jet with $p_T > 500\,\text{GeV}$. More details on jet reconstruction will be given later in this Section.

Background samples are generated using the MadGraph [41] generator and hadronized using Pythia8 [42, 43]. Samples considered are multijet QCD, $W \rightarrow q' \bar{q}$, and $Z \rightarrow q \bar{q}$. They are produced in bins of $H_T$ in order to be efficient at generating events passing the trigger threshold. For the $W$ and $Z$ events, we apply higher order QCD and EWK corrections to those processes to improve modeling of the high $p_T$ W and Z events [44–46].

Signal samples are also generated using the MadGraph generator and are also hadronized with Pythia8. The same higher order NLO QCD corrections that are applied to the $W$ and $Z$ simulation are also applied to the signal simulation. The NLO EWK corrections are not applied since we are searching for $Z'$ in the context of a leptophobic model.

Jets used in the analysis are reconstructed from clustering particle flow candidates in the event using the anti-$k_T$ algorithm with a distance parameter of 0.8. To mitigate the effect of pileup, the pileup per particle identification (PUPPI) algorithm [47] is used to weight the particle flow candidates prior to jet clustering based on the likelihood of coming from the primary vertex. Further corrections are applied as a function of jet pseudorapidity ($\eta$) and transverse momentum to account for detector non-linearities. Noise filtering is also applied to the jets. The AK8 PUPPI jets are required to have $p_T > 500\,\text{GeV}$ to be fully efficient with respect to the trigger and must be within $|\eta| < 2.5$. We veto events containing identified and isolated electrons and muons with $p_T > 10\,\text{GeV}$.

Substructure observables are used to identify a $Z'$ merged into a single AK8 jet. The most important variable is the groomed jet mass which peaks at the $Z'$ mass for signal and pushes the mass of background quark- and gluon-initiated jets towards lower values. For the mass of the jet we choose to use the soft drop mass groomer [48, 49] with parameters $z = 0.1$ and $\beta = 0$. Since the $Z' \rightarrow q \bar{q}$ jet is 2-prong we use the substructure tools typically applied for W jet tagging. We use a variable that is based on the $n$-subjettiness ratio $\tau_2/\tau_1$, also referred to as $\tau_{21}$, which is widely used in other searches [50–52] to determine how consistent a jet is with having a 2-prong structure. There are some challenges with the $\tau_{21}$ variable related to its correlation with the groomed jet mass. In particular, a selection based on a given $\tau_{21}$ value sculpts the jet mass distributions depending on the $p_T$ of the jet, causing a non-trivial shape of the jet mass distribution. This makes searching for a generic jet mass peak over a large range in $p_T$ particularly challenging. Therefore, we propose to employ the DDT transformation of $\tau_{21}$ where DDT stands for designed decorrelated tagger.

The DDT transformation is described in [53] and its purpose is to remove the correlation between $\tau_{21}$ and a QCD-invariant type variable [54] such as $\rho$, where $\rho = \log(m^2/p_T^2)$, by applying a linear transformation. Then selecting on $\tau_{21}^{\text{DDT}}$ keeps the background QCD mass distribution smoothly falling. In addition, this behavior will be predictable as a function of $p_T$. As in [53], we evaluate the correlation between $\tau_{21}$ and mass and $p_T$ to perform the DDT transformation. In Fig. 2 (left), we show the profile plot in QCD simulation of $\tau_{21}$ versus the QCD scaling variable $\rho$ in different bins of $p_T$. A linear behavior exists in different bins of $p_T$ al-
though each $p_T$ bin has a different offset. The region of the profile plot which is non-linear on the left side of the distribution corresponds to very small jet mass values we do not probe in this analysis, where $\rho = -8$ corresponds to a jet mass of $\sim 10$ GeV for a jet of $p_T = 500$ GeV. The non-linear region near $\rho \sim 0$ corresponds to a cutoff in the distribution from having a fixed-R jet algorithm where there are only a small fraction of the events. We remove this $p_T$ dependence by defining $\rho_{DDT} = \log(m^2 / (p_T \times \mu))$ where $\mu$ is a scale to keep the value of $\rho_{DDT}$ dimensionless and is set to 1 GeV. Now the relevant region for the analysis is above $\rho_{DDT} = 0$ which corresponds to a jet mass of $\sim 22$ GeV for a jet of $p_T = 500$ GeV. Then, in Fig. 2 (right), we perform a transformation to make this correlation distribution flat for the mean value of $\tau_{21}^{DDT}$. This change of variable is a simple linear transformation:

$$\tau_{21}^{DDT} = \tau_{21} - k \times \rho_{DDT}$$

The slope $k = -0.063$ is fit from the distributions. Based on optimization of the analysis sensitivity given a Z' with $g_B = 1$, we choose to set a tagging requirement on the Z' candidate of $\tau_{21}^{DDT} < 0.38$. This corresponds to approximately 20% signal efficiency for a Z' signal over the mass range in this search.

Figure 2: Profiled distribution of $\tau_{21}$ versus $\rho$ (left) and $\tau_{21}^{DDT}$ versus $\rho_{DDT}$ (right) for multijet QCD MC, for different leading jet $p_T$ values, which demonstrates the (de-)correlation of the variables using the DDT transformation. The region of the right distribution where $\rho_{DDT} < 0$, corresponds to very small jet mass values we do not probe in this analysis, while the non-linear region where $\rho_{DDT} > 4.7$ corresponds to high mass jet values where $\rho_{DDT}$ acquires non-physical values for a given $p_T$. Error bars denote the uncertainty on the mean for the profiled distributions.

A comparison of the data and simulation of the leading $p_T$ jet soft drop mass and $\tau_{21}^{DDT}$ variables are shown in Fig. 3. There exists some residual disagreement between the data and simulation for the two variables. Therefore we perform an estimation of the QCD background using data-driven techniques. Corrections to the resonant SM backgrounds (W and Z), as well as the signal shapes, are derived by using the W jet peak in semi-leptonic $t\bar{t}$ events. This will be discussed in greater detail in Section 5.

4 Background estimation

The aim of the background estimation technique is to predict the background jet mass shape from data control regions with as little reliance on simulation as possible. We predict the background in the jet mass region from 30-330 GeV to include the SM processes, W and Z. Including
Figure 3: Data-to-simulation comparison of the leading $p_T$ jet soft drop mass (left) and $\tau_{21}^{DDT}$ (right) after kinematic selections on the leading $p_T$ jet which is dominated by QCD multijet processes (pink) with subdominant contributions from inclusive SM $W$, $Z$, and $t\bar{t}$ processes. Residual differences in data and simulation demonstrate the need for a data-driven background estimation method.

this region allows us to constrain the $Z'$ shape and normalization by determining in situ the $W$ and $Z$ contribution. We then set limits on $Z'$ masses in the range from 100-300 GeV.

We make use of both mass sidebands and $\tau_{21}^{DDT}$ sidebands to extrapolate the QCD shape in the signal region. We define a transfer factor (TF) that translates from the fail - $\tau_{21}^{DDT} > 0.38$ - to the pass - $\tau_{21}^{DDT} < 0.38$ - region as a function of $\rho_{DDT}(\tau_{21}^{DDT}, p_T)$. We choose this parameterization of TF because the $\tau_{21}^{DDT}$ variable is designed to be decorrelated with respect to $\rho_{DDT}$ in bins of jet $p_T$, as shown in Fig. 2. The method uses the near-flat correlations between $\tau_{21}^{DDT}$ and $\rho_{DDT}$ as a function of $p_T$ to build an estimate of the QCD events in the signal region. We scale events from the fail region of the $\tau_{21}^{DDT}$ requirement defines the passing signal region and the failing sideband region. The goal of the method is to predict the $\rho_{DDT}$ distribution in the signal region and then translate it back into a jet mass distribution.

To determine this transfer factor plane, we fit in $\{\rho_{DDT}, p_T\}$ space with polynomial functions using only the data. Before doing the fit, we subtract the SM $W$/$Z$ contributions expected from simulation, corrected at NLO, in order to estimate only non-resonant backgrounds such as the QCD multijet background. For a given $Z'$ mass, we exclude from the fit the window defined as 10% around the $Z'$ mass, approximately the jet mass resolution, which could otherwise bias the TF measurement. This is the missing strip in the transfer factor plane in Fig. 4 (left). A unique mass window is required for each $Z'$ mass hypothesis and therefore a unique fit and background estimation is made in each case.

The resulting transfer factor plane, as an example for a $Z'$ mass of 110 GeV, is shown in Fig. 4 where the fit now interpolates into the signal region. This transfer factor plane can be slightly different for each mass hypothesis as a different window in mass is excluded for each $Z'$ mass in the search. With this transfer factor plane, we can translate the fail sample events in data directly into an expectation for the QCD background in the signal region. Uncertainties in the
transfer factors come from shape uncertainties derived from the fit parameters, the finite statistics in the fail region, and simulation closure from performing the same fit in QCD simulation. A detailed discussion of these uncertainties is given in Section 5. We note that because the input to the fit is the pass-to-fail transfer factor and not the overall normalization, this fit also takes advantage of information below the trigger threshold to provide more data to further constrain it. Therefore we begin this fit at jet $p_T$ of 350 GeV. Using this transfer factor plane, we can then predict the number of events in the pass signal region given the event yields in the fail region. Results from the QCD background prediction are presented below.

Figure 4: Left: Diagram illustrating the background estimation technique which estimates the signal region ("pass") QCD background distribution from the sideband region ("fail") in the $\{\rho_{DDT}, p_T\}$ phase space. The empty strip in the pass-to-fail plane is the signal region in $\rho_{DDT}$ (or jet mass) for a given $Z'$ mass. Right: The transfer factor $TF(\rho_{DDT}, p_T)$ space, now interpolated over the signal region, for a $Z'$ with a mass of 110 GeV as measured in data which translates "fail" region events into the signal region.

An alternative analysis method consists of searching for a resonant peak above the smoothly falling background shape by performing a simultaneous signal plus background fit. In addition, the W and Z processes are included in the fit and have the same modeling and systematic uncertainties as the primary background estimation technique. Comparable results for upper limits on the $Z'$ cross section, to within 1σ, are found between the sideband extrapolation background estimation technique and the alternative background estimation technique. A preference is given to the sideband method as it uses more information from the data to determine the background and is less susceptible to biases.

5 Systematic Uncertainties

The background from QCD is estimated using the sideband fit method which fits for the shape of the $n_{\text{pass}}/n_{\text{fail}}$ in the $\rho_{DDT}$ sidebands. The uncertainties in this method come from 3 main sources:

- Uncorrelated per bin in the mass distribution: (1) statistical uncertainties from the "fail" sample and (2) closure statistical uncertainties estimated using the data fit
- Uncertainty on the transfer factor shape from the fit parameter errors correlated across all jet mass bins
- MC closure uncertainty of the method demonstrated with QCD simulation correlated across all jet mass bins
The systematic effects for the W, Z, and signal shapes and normalization are tied closely together since they are all affected by similar systematic uncertainties. We constrain jet mass scale, jet mass resolution, and $\tau_{21}$ selection efficiency using a sample of merged W jets in semi-leptonic $t\bar{t}$ events in data. This region is defined by selecting a high energy muon (electron) with $p_T > 53$ (120) GeV, $E_T^{\text{miss}} > 80$ GeV, a high $p_T$ AK8 jet with $p_T > 200$ GeV, and a b-tagged jet with $\Delta R > 0.8$ from the AK8 jet. Using the same $\tau_{21}$ selection as in the analysis, we look at the passing and failing regions for real, merged W jets in data and simulation. This is shown in Fig. 5 where the fail sample is shown on the left and the pass sample is shown on the right. A simultaneous fit to the two samples is performed in order to extract the tagging efficiency of a merged W jet in simulation and data. A scale factor between data and simulation is then computed.

From a study of this semi-leptonic $t\bar{t}$ event sample we measure the data-to-simulation scale factor on the $\tau_{21}$ selection to be $0.95 \pm 0.20$. The mass scale between data and simulation is set to be $-0.59 \pm 0.87$ GeV. The jet mass resolution data-to-simulation scale factor is measured to be $1.10 \pm 0.12$. These uncertainties determine the initial values for the W, Z, and signal shapes and they are then further constrained in the final fit because of the presence of the W and Z peaks in our background estimation. The W and Z rates and shapes are determined in situ with the fit for a possible signal.

Finally, additional systematic uncertainties are applied to the W, Z, and $Z'$ signal yields from higher order corrections to the boson $p_T$ distributions, jet energy scale uncertainties, pileup and the luminosity uncertainty.

The systematic effects for the various background and signal components are summarized in Table 1. Given the background estimation method and associated systematic uncertainties, studies are performed in the simulation injecting signal events and determining the bias in the measured signal cross section. No significant bias is observed in these studies.
6 Results

Next, we combine the various SM contributions and search for a potential signal from a Z’ resonance. We set limits in the Z’ mass range from 100-300 GeV and the SM backgrounds are estimated starting from 30 GeV. This has the benefit of using the W and Z SM processes as standard candles to which we may correlate systematics uncertainties with the Z’ signal. The QCD background is evaluated as described in Section 4. The SM W and Z contributions are estimated from simulation with additional corrections. These corrections are primarily from higher order NLO differential cross section calculations and data-to-simulation scale factors related to tagging efficiency and jet mass scale and resolution differences.

The number of observed events in the data was found to be consistent with the predicted background from SM processes. The results are interpreted in terms of an upper limit on the production cross section. Upper limits are computed using the modified frequentist approach for confidence levels (CLs), taking the profile likelihood as the test statistic [55, 56] in the asymptotic approximation. Results are obtained from combined signal and background binned likelihood fits to the shape of the jet mass distribution as shown in Fig. 6. A unique estimation of the QCD background is made for each Z’ mass and so the final distributions are shown for a Z’ with a mass of 135 GeV (top) and 200 GeV (bottom). The upper limits on the cross section are given in Fig. 7 (top) and are compared to cross sections for coupling values of $g_B = 0.5, 1.0$ that are close to our current sensitivity.

Finally, these limits are translated into the $g_B$ versus $m_{Z'}$ plane. In Fig. 7 (bottom), we show previous results from UA2, CDF, and both ATLAS and CMS. Recent ATLAS results from Run 2 in [26, 36] are scaled to the coupling $g_B$ using a factor of 6 which relates $g_B$ to the $g_q$ coupling presented in [57]. Our limits are the first results in the region below 140 GeV and are the most sensitive in the mass regime less than 300 GeV.

7 Summary

We present the first limits on a search for a light Z’ boson decaying to a quark-antiquark pair in the mass range from 100-300 GeV with the CMS detector. We do not observe any excess above the SM prediction and set limits on the Z’ coupling to quarks, $g_B$, as a function of the Z’

<table>
<thead>
<tr>
<th>Systematic Effect</th>
<th>QCD bkg.</th>
<th>SM W/Z and Z’ signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>QCD fit parameters</td>
<td>shape</td>
<td>-</td>
</tr>
<tr>
<td>QCD bin-by-bin statistics</td>
<td>per bin</td>
<td>-</td>
</tr>
<tr>
<td>QCD fit closure</td>
<td>shape</td>
<td>-</td>
</tr>
<tr>
<td>W tag scale factor</td>
<td>-</td>
<td>20%</td>
</tr>
<tr>
<td>NLO QCD pT corrections$^\Delta$</td>
<td>-</td>
<td>8-12%</td>
</tr>
<tr>
<td>jet energy/mass resolution$^\dagger$</td>
<td>-</td>
<td>10%</td>
</tr>
<tr>
<td>MC statistics$^\dagger$</td>
<td>-</td>
<td>per bin</td>
</tr>
<tr>
<td>luminosity</td>
<td>-</td>
<td>2.7%</td>
</tr>
<tr>
<td>pileup</td>
<td>-</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>jet energy/mass scale$^\dagger$</td>
<td>-</td>
<td>1%</td>
</tr>
</tbody>
</table>

Table 1: Summary of the systematic uncertainties for signal and background. Symbol $^\dagger$ denotes this is only applied to Z’ signal. Symbol $^\dagger$ denotes a shape uncertainty on the peaking SM W/Z and Z’ signal shape. Symbol $^\Delta$ denotes EWK uncertainties are also included for the SM W/Z processes.
Figure 6: Final inputs to the statistical interpretation for a $Z'$ mass of 135 GeV (top) and 200 GeV (bottom). The QCD background prediction, including uncertainties, is shown in the blue boxes while the sum of the SM processes is shown in the blue line. Contributions from the $W$, $Z$, and $Z'$ are given as well. In the bottom panel, the ratio of the data to the background prediction, including uncertainties, is shown.
Figure 7: 95% confidence level upper limit on the $Z'$ production cross section compared to the theoretical cross section for a $Z'$ with $g_B = 0.5, 1$ (top) and the translation of that upper limit to a limit on $g_B$ (bottom). Limits from other relevant searches are also shown. Recent ATLAS results from Run 2 in [26, 36] are scaled to the coupling $g_B$. 
mass. Our limits are the most stringent in the mass range less than 300 GeV. For masses below 140 GeV, they are the only limits set.
References


