

## Search for single production of vector-like quarks decaying to a b quark and a Higgs boson

G. RAUCO on behalf of the CMS COLLABORATION

*Universität Zürich - Zürich, Switzerland*

received 6 September 2018

**Summary.** — A search is presented for the singly produced heavy vector-like quark  $B$ , in its decay to a Higgs boson and a  $b$  quark. The Higgs boson decays to a highly boosted  $b\bar{b}$  quark pair and is reconstructed as a single collimated jet. The analysis exploits data collected by the CMS experiment in proton-proton collisions at  $\sqrt{s} = 13$  TeV at the CERN LHC. The corresponding integrated luminosity is  $35.9 \text{ fb}^{-1}$ . The data are in agreement with the background expectations, and upper limits at 95% confidence level on the  $B$  quark cross section times the branching ratio are set. For a narrow  $B$  quark with a mass between 700 and 1800 GeV, values in the range 1.28–0.07 pb are obtained. Similar sensitivities are achieved also in the case of more realistic scenarios where the production of  $B$  quarks with widths of 10, 20 and 30% of the resonance mass are considered. With this search, the current knowledge on vector-like quarks was extended, being this analysis the first at the CERN LHC for the single production of a  $B$  quark through its fully hadronic decay channel, and also the first study investigating finite resonance widths of the  $B$  quark.

### 1. – Introduction

On July 4, 2012, the ATLAS and CMS experiments announced the discovery of the Higgs (H) boson and with it the standard model (SM) of particle physics was confirmed. However, several problems related to the nature of the electroweak symmetry breaking and to the hierarchy between the electroweak and the Planck mass scales remain unsolved. To cope with this issue several models beyond the SM have been elaborated, that include the introduction of new particles allowing the cancellation of loop corrections to the H mass. Vector-like quarks (VLQ) are one of the possible new particles predicted by the theory.

The first search at the LHC for the electroweak production of a VLQ  $B$ , with electrical charge  $-1/3 e$ , with  $e$  the proton charge, decaying to a bottom ( $b$ ) quark and a H boson is presented here. The H boson is considered in its decay to a pair of  $b$  quarks. The final state of this analysis is therefore fully hadronic. The typical signature of the events

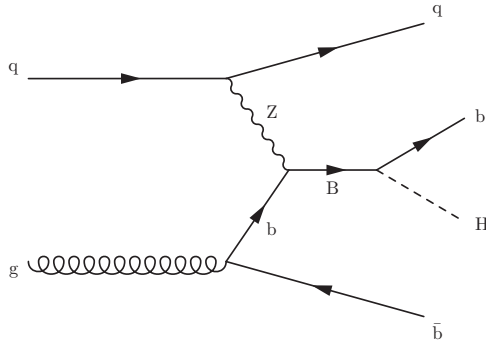


Fig. 1. – Representative Feynman diagram [1].

here investigated consists of a boosted H boson, reconstructed as a large cone jet, and multiple narrower jets originated from the hadronisation of light-flavored and b quarks. In fig. 1, a representative Feynman diagram illustrates the process under study.

In addition to the decay channel considered in this analysis, the  $B \rightarrow Hb$ , the B quark can also decay into  $Zb$ ,  $Wt$  and in additional lighter states considering models with model-dependent branching fractions. Here the B quark is considered both belonging to a singlet or a doublet representation. The branching ratios in  $Hb$ ,  $Zb$ , and  $Wt$  depend on the VLQ mass and are approximately 25, 25 and 50%, 50, 50, and 0%, respectively for the singlet and doublet state.

As opposite to previous searches for the B quark, that just assumed the resonance width to be narrow compared to the experimental resolution, in this search we explore also more realistic scenarios of non-negligible width, with values up to 30% of the resonance mass.

The presented search is performed using proton-proton (pp) events collected by the CMS experiment at a centre-of-mass energy of 13 TeV, corresponding to an integrated luminosity of  $35.9 \text{ fb}^{-1}$ .

In sect. 2, a brief description of the CMS experiment is presented. Signal modeling and background simulations are then described, and the theoretical framework used for the results interpretation is introduced in sect. 3. Section 4 focuses on the analysis workflow, describing the event selection and the main tools used in the search. The multijet background estimation strategy is also described, followed by the systematics uncertainties accounted in this study. Finally, the results and their theoretical interpretation are discussed in sect. 5.

## 2. – The CMS detector

The CMS detector [2] is one of the two multipurpose experiments situated at the Large Hadron Collider. It is a cylindrical-shaped machine, with length 21.6 m and diameter of approximately 14.6 m. The main feature of the detector is a superconducting solenoid generating a magnetic field of 3.8 T. Many sub-detectors are installed inside the solenoid: a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter, and a brass and scintillator hadron calorimeter, each composed of a barrel and two end sections. Outside the solenoid is located a steel flux-return yoke, and the system for muon detection consisting of gas-ionization chambers. A schematic view of the machine is pictured in fig. 2.

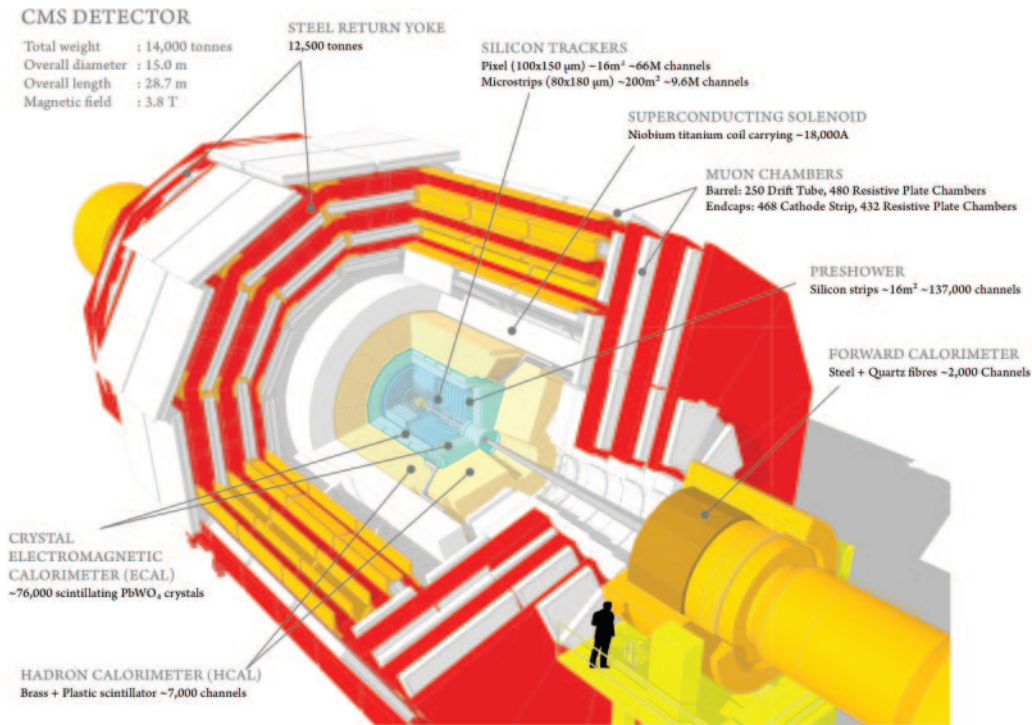


Fig. 2. – An exploded view of the CMS detector, showing the outer muon chambers (white) embedded in iron (red). Internally can be seen the calorimeter system (ECAL, in green, and HCAL, in orange). The tracker is located in the core of the central barrel.

### 3. – Modeling and interpretation

**3.1. Signal and backgrounds simulation.** – The signature of the events studied in this analysis is characterized by the presence of multiple jets rich in heavy-flavor content and a highly boosted H boson. Therefore, the most important background is the production of jets through the strong quantum chromodynamic (QCD) interaction, labelled as multijet events. Minor contributions arise from  $t\bar{t}$  events and electro-weak production of W/Z bosons in association with jets. The background from multijet events is estimated from data in control regions, while all the others mechanisms are estimated using Monte Carlo (MC) simulation.

Multijet events and electroweak backgrounds are simulated at leading order (LO) using the MADGRAPH5 aMC@NLO 2.2.2 generator. It is further interfaced to PYTHIA 8.2 with the CUETP8M1 underlying-event tune to include the simulation of the parton-shower. The pair-production of top quark is generated using POWHEG v2 at next-to-leading order (NLO), also interfaced to PYTHIA.

Signal event samples have been generated at LO with the generator MADGRAPH5\_aMC@NLO, modeled using the universal FEYNRULES output. The generator is then matched to PYTHIA 8 for the parton-shower simulation. Several benchmark models have been considered depending on the hypothesis made on the B quark mass, that is ranging from 700 up to 1800 GeV in steps of 100 GeV. In the

narrow-width regime, the intrinsic decay width of the B quark set fixed to 1 GeV. More realistic scenarios of B quarks with large widths (10, 20, or 30% of the mass hypothesis) have also been studied and the corresponding signal event samples generated.

All simulated events are finally processed with GEANT4 to include the full simulation of the CMS detector.

**3.2. Theoretical framework.** – The process of the single production of a VLQ B, decaying in a final state X, has a total cross section described as

$$(1) \quad \sigma(C_1, C_2, m_B, \Gamma_B, X) = C_1^2 C_2^2 \hat{\sigma}_{AW}(m_B, \Gamma_B),$$

where  $C_1$  and  $C_2$  are the production and decay couplings corresponding to the interactions through which a B quark is produced and decays, and  $\hat{\sigma}_{AW}$  is the reduced cross section for a resonance of arbitrary width (AW). In the approximation of  $\Gamma_B/m_B$  approaching zero, production and decay can be factorized, resulting in the following expression for the cross section:

$$(2) \quad \sigma(C_1, C_2, m_B, \Gamma_B) = \sigma_{\text{prod}}(C_1, m_{\text{prod}}) \mathcal{B}_{B \rightarrow X} = C_1^2 \hat{\sigma}_{NWA}(m_B) \mathcal{B}_{B \rightarrow X},$$

where  $C_1$  is the B production coupling, and information for the parameters  $C_2$  and  $\Gamma_B$  are included in the branching fraction for the specific B quark decay  $\mathcal{B}_{B \rightarrow X}$ , while  $\hat{\sigma}_{NWA}$  is the reduced cross section in the narrow-width approximation (NWA).

In this search, we rely on the assumption that the B quark can decay into Hb, Zb, and Wt and that their corresponding branching fractions depend on the weak mixing angle and on the coupling strength, that can be fixed to tune the desired resonance width. More details on the model and the setting of the parameters can be found in ref. [1].

## 4. – Analysis strategy

**4.1. Event selection.** – The topology of this search is boosted, given the presence of an energetic H boson arising from the decay of the B quark. The H boson is expected to have a very large transverse momentum, and the two b quarks originating from its decay emerge therefore very collimated and close to each other in the  $\eta$ - $\phi$  space and are reconstructed as a single large jet.

The first step of the selection is defined online (trigger level) and is based on the jet activity  $H_T$ , an hadronic variables defined as the scalar  $p_T$  sum of all the narrow jets with  $p_T > 30$  GeV and  $|\eta| < 3$ . The events, to be selected, must ensure a  $H_T$  in excess of 900 GeV. To increase the trigger efficiency, collisions where at least one jet with  $p_T > 450$  GeV is identified are saved.

To further identify signal-like events, a request on the narrow jets multiplicity is also applied, asking for at least 3 jets with  $p_T > 30$  GeV and  $|\eta| < 4$ , among which there must be at least one coming from the hadronization of a b-quark. Selected events should also include at least one large jet, reconstructed using a radius cone parameter 0.8 instead of the usual 0.4 size employed for the narrow jets reconstruction. Higgs-tagging requirements are applied to these large jets to identify the Higgs boson candidates, and are based on a dedicated pruning algorithm, that with a iterative procedure discard soft and wide angle radiations inside the jet. With this procedure a pruned-jet mass is defined and it is requested to fall within the window of 105–135 GeV. Furthermore, two subjets

are obtained using the soft-drop algorithm and both of them should satisfy b-tagging requirements.

The main focus of this analysis is the reconstruction of the B quark candidate, that is achieved considering the selected Higgs jet candidate and the  $p_T$ -leading jet b-tagged jet, leading to a significant reduction of the combinatorial background. Furthermore, the selected b jet is requested not to overlap with the Higgs candidate, and this request is achieved by applying a distance in the  $\eta$ - $\phi$  space between the two objects in excess of 1.2.

To further reduce the overwhelming multijet background, the hadronic activity is required to be greater of 950 GeV for B mass hypothesis between 700 and 1500, and of 1250 for the higher part of the mass spectrum. The search is therefore split in two parts: the “low-mass” and the “high-mass” analysis.

Finally, a light flavor quark produced is associated with the VLQ B, and it is typically emitted in the forward region of the detector. This distinctive feature strongly enhances the signal to background discrimination, and is exploited by separating events in two categories: a high-purity category is defined by requiring at least one forward jet, while a second category, containing a large fraction of events, is characterized by the absence of jets with  $|\eta| > 2.4$ .

**4.2. Background estimation.** – The core strategy of this analysis consists of a “bump-hunt” search over the expected background in the reconstructed mass  $m_{bH}$  distribution and the signal is extracted by means of a binned maximum likelihood fit. To reduce the dependence of the fit strategy on the modeling of the dominant multijet background, this contribution is extracted from data. The procedure used is called “ABCD method”. A two dimensional phase space is defined by two variables: the b-tagged subjet multiplicity of the Higgs jet candidate and its reconstructed pruned mass,  $m_J$ . Therefore, in addition to the nominal signal region (here referred to region A), three multijet enriched regions are used to define the ABCD method: region B, with two b-tagged subjets, and  $75 < m_J < 105$  GeV or  $m_J > 135$  GeV, region C, with one b-tagged subjet, and  $105 < m_J < 135$  GeV, and region D, with one b-tagged subjet, and  $75 < m_J < 105$  GeV or  $> 135$  GeV.

Assuming the b-tagged subjet multiplicity and the Higgs boson pruned mass as uncorrelated, and labelling with  $N_A$ ,  $N_B$ ,  $N_C$ , and  $N_D$  the yields in the four regions, the multijet background yield in the signal region A is given by the relationship  $N_A = N_B N_C / N_D$ , after the  $t\bar{t}$  from the MC simulation is subtracted from data. The electroweak contributions are not subtracted as they are negligible. The shape, on the other hand, is estimated from the  $m_{bH}$  distribution in the control region C, given that the correlation between mass spectrum and the b jet multiplicity is expected to be negligible.

Several checks are performed to validate the method, on simulation only and cross-checked on data. The procedure is applied to signal-depleted regions from high-mass sidebands and the agreement between the background estimated in the modified signal region and the observed data is good.

**4.3. Systematic uncertainties.** – Several sources of systematic uncertainties are considered in this search and taken into account as nuisance parameters in the statistical interpretation and results extraction. Their effect is evaluated by propagating the uncertainty to the  $m_{bH}$  distribution and to the event yields.

Corrections due to misalignment between simulation and data are applied. An uncertainty of 2.5% on the integrated luminosity is considered. Differences between the b tagging efficiency in data and MC are also accounted for. The transverse momenta of narrow and fat jets are corrected to consider the jet energy scale and resolution, in

addition to the pruned mass scale and resolution. Scale factors are applied to account for discrepancies between the trigger efficiency in data and in simulations, as well as to reweight simulated events to match the pileup distributions in data.

Theoretical uncertainties, such as the choice of factorization and normalization scales, are also evaluated. The uncertainty from the choice of parton distribution functions is also estimated and considered.

Specifically referred to this search, systematic uncertainties accounting for the mis-modelling of the forward jets multiplicity, are applied. The magnitude of this correction is evaluated by taking the normalization difference in data and in MC, and results affecting the two forward jet based categories of 0.5% and 2.0%, respectively when 0 and at least 1 forward jets are identified.

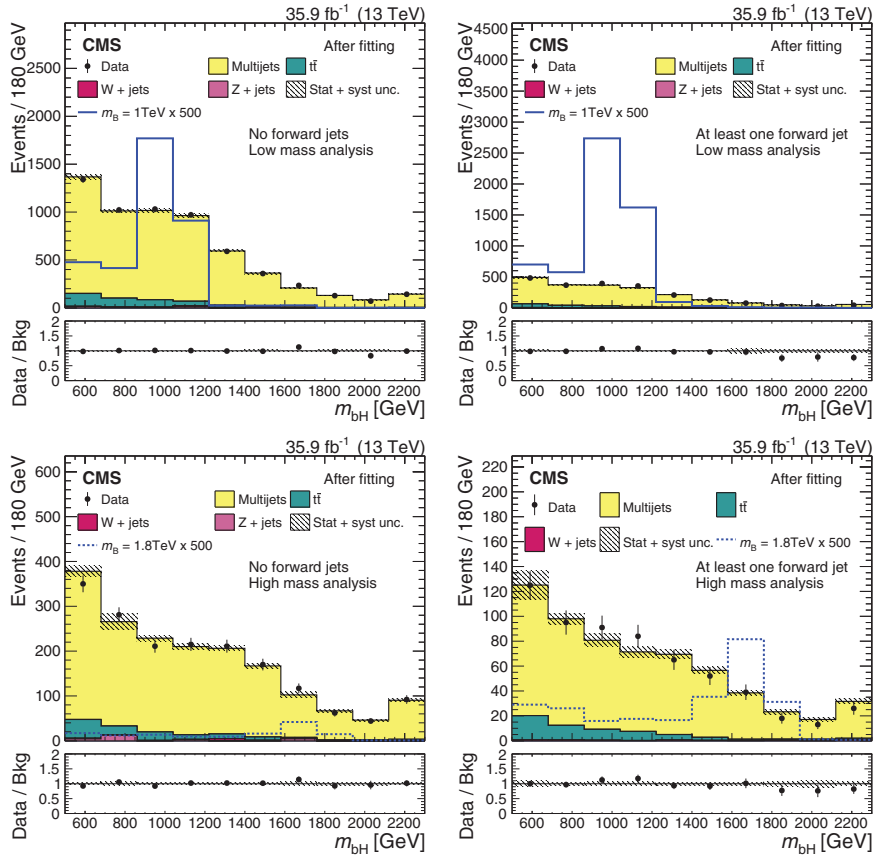


Fig. 3. – Distribution in the reconstructed B quark mass, after applying all selections to events with no forward jets (left) and to events with at least one forward jet (right), compared to the background distributions estimated after fitting [1]. The upper and lower plots refer to the low and high-mass  $m_B$  analyses, respectively. The expectations for signal MC events are given by the blue histogram lines. Different contributions to background are indicated by the colour filled histograms. The grey-hatched error band shows total uncertainties in the background expectation. The ratios of observations to background expectations are given in the lower panels, together with the total uncertainties prior to fitting, indicated by the grey-hatched band.

TABLE I. – Observed and expected fitted event yields in the signal region in both low mass and high mass ranges [1]. The multijet background expectation yields are derived from data, while the yields for the other sources of background are derived from MC simulation.

Category	Source	Low-mass yields	High-mass yields
Zero forward jets category	Multijet	$5416 \pm 60$	$1612 \pm 24$
	$t\bar{t}$	$394 \pm 46$	$117 \pm 18$
	W + jets	$29 \pm 13$	$10.5 \pm 4.3$
	Z + jets	$43 \pm 15$	$23 \pm 23$
	Total backgrounds	$5882 \pm 42$	$1762 \pm 26$
	Observed in data	$5886 \pm 77$	$1753 \pm 42$
	Expected signal	$7.3 \pm 0.3$	$0.27 \pm 0.01$
At least one forward jet category	Multijet	$1938 \pm 23$	$549 \pm 10$
	$t\bar{t}$	$163 \pm 20$	$58 \pm 17$
	W + jets	$11.5 \pm 4.2$	$4.3 \pm 1.4$
	Z + jets	$2 \pm 10$	$0.0 \pm 0.0$
	Total backgrounds	$2115 \pm 21$	$612 \pm 15$
	Observed in data	$2107 \pm 46$	$608 \pm 25$
	Expected signal	$11.5 \pm 0.3$	$0.51 \pm 0.01$

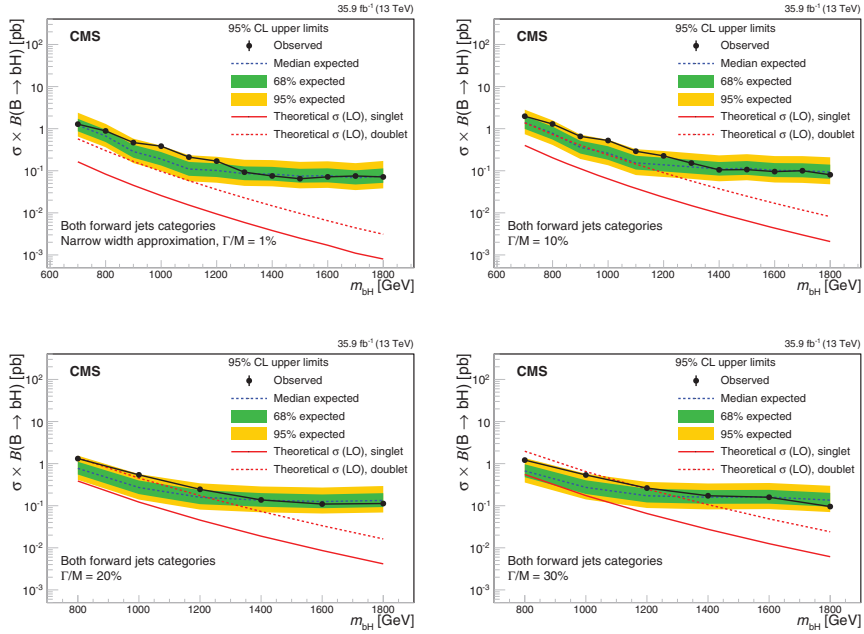


Fig. 4. – The median observed and expected 95% CL upper limits on the product of the B quark production cross section and branching fraction as a function of the signal mass, assuming narrow-width resonances (upper-left) and widths of 10% (upper-right), 20% (lower-left), and 30% (lower-right) of the resonance mass for the B quark [1]. The results are shown for the combination of 0 and  $> 0$  forward-jet categories. The continuous red curves correspond to the theoretical expectations for singlet and doublet models.

Finally, uncertainties on the estimated multijet background are applied. Statistical uncertainties are obtained from control samples in data and propagated to  $m_{bH}$  in the signal region by shifting the observed event yield in regions B and D of  $\pm 1\sigma$ , and deriving again the expected distribution in the signal region. The statistical uncertainty in region C, being the region from which the shape of the  $m_{bH}$  spectrum is derived, is considered in the signal extraction. The bin content of each bin was therefore allowed to fluctuate independently according to Poisson statistics and with this procedure a systematic uncertainty in the shape of the multijet background estimated distribution was obtained. In addition, the difference between the observed and predicted yields obtained in the method validation test performed in the large-mass sidebands, was accounted for normalization uncertainties and corresponds to  $\approx 5\text{--}10\%$ .

## 5. – Results and interpretation

A binned maximum likelihood fit is performed on the  $m_{bH}$  spectrum. The fitted distributions are presented in fig. 3, and the corresponding yields are listed in table I. In all categories, an agreement is obtained between the observed distributions and the background-only hypothesis, therefore the obtained results are consistent with the SM predictions. Upper limits are set on the product of the cross section and branching fraction of a B quark decaying to Hb, produced in association with another b quark and a light-flavour quark, as a function of  $m_{bH}$  and exclusion limits at 95% confidence level (CL) are calculated. The observed and expected combined upper limits from the two forward jet based categories are given in fig. 4. In the narrow width regime, values of  $\sigma\mathcal{B}(bH)$  between 1.28–0.07 are excluded at the 95% confidence level, for masses ranging in 700–1800 GeV. Those values are compared to theoretical predictions computed at NLO, considering both the singlet and the doublet models and assuming a branching ratio of the VLQ B in a b quark and a H boson of  $\approx 25\%$ . The same approach is followed to extract the results assuming more realistic scenarios of resonance width over masses fixed to 10, 20, and 30%. Similar sensitivities are obtained, and exclusions limits lying between 0.08 and 1.97, 0.11 and 1.32, and 0.10 and 1.22 pb, respectively.

## REFERENCES

- [1] CMS COLLABORATION, *Search for single production of vector-like quarks decaying to a b quark and a Higgs boson*, to be published in *JHEP*, arXiv:1802.01486 [hep-ex].
- [2] CMS COLLABORATION, *JINST*, **3** (2008) S08004.