Search for charged Higgs bosons decaying into a top quark and a bottom quark in the fully hadronic final state at 13 TeV

The CMS Collaboration

Abstract

A search for charged Higgs bosons decaying into a top and a bottom quark-antiquark pair in the fully hadronic final state is presented. The analysis uses LHC proton-proton collision data recorded with the CMS detector in 2016 at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 35.9 fb$^{-1}$. No significant deviation above the expected background is observed. Model-independent upper limits at 95% confidence level are set on the charged Higgs boson production cross section times branching fraction in two scenarios. For production in association with a top quark, limits of 21.3 to 0.007 pb are calculated for charged boson masses in the range 200 GeV to 3 TeV. Combination with data from a search in the leptonic final states results in improved limits of 9.25 to 0.005 pb. The complementary s-channel production of a charged Higgs boson is investigated in the mass range 800 GeV to 3 TeV and the corresponding upper limits are 4.5 to 0.023 pb. These results are interpreted in different minimal supersymmetric extensions of the standard model.
1 Introduction

The observation [1–3] of a Higgs boson [4, 5] has initiated a new field of study at the CERN LHC to explore its properties and to better understand its role in the standard model (SM). Many models beyond the SM predict an extended Higgs sector. Minimal extensions often include a second complex Higgs doublet and are known as two-Higgs-doublet models (2HDMs) [6–9]. The two doublets lead to five physical particles: two charged Higgs bosons, $H^\pm$, two neutral scalars, h and H, and one neutral pseudoscalar, A. The 2HDMs are further classified according to the couplings of the doublets to fermions. One of the popular 2HDMs is the minimal supersymmetric standard model (MSSM) [10, 11] where one doublet couples to the up quarks and the other to the down quarks and the charged leptons (Type-II 2HDMs). The production and decay of the $H^\pm$ depend on its mass and on the parameters of the model. No charged fundamental scalar boson exists in the SM and the discovery of such a particle would uniquely point to physics beyond the SM.

We describe a search for charged Higgs bosons decaying to a top and bottom quark-antiquark pair ($H^+ \rightarrow t\bar{b}$). The boson is produced in association with a top and a bottom quark in the so-called four-flavor scheme (4FS), in association with a top quark in the five-flavor scheme (5FS) [12, 13], or via the s-channel production process. The corresponding leading order (LO) Feynman diagrams are shown in Fig. 1. Charge conjugate processes are implied throughout this note.

![Figure 1: LO diagrams for the production of a heavy charged Higgs boson, showing the top quark associated production process in the 4FS (left), the 5FS (middle), and the s-channel process (right).](image)

Various searches for the signature $H^+ \rightarrow t\bar{b}$ have been performed by the ATLAS and CMS Collaborations at a center-of-mass energy of $\sqrt{s} = 8$ TeV [14, 15] and $\sqrt{s} = 13$ TeV [16, 17] and have been interpreted in the context of 2HDMs. Results of searches for a light charged Higgs boson produced in the decay of a top quark and subsequently decaying into c$\bar{s}$ or into c$\bar{b}$ are presented in Refs. [18, 19]. Limits on production of a charged Higgs boson in the $\tau^+\nu_\tau$ decay channel have been obtained at center-of-mass energies of $\sqrt{s} = 8$ and 13 TeV [15, 20–22]. Charged current processes from low-energy precision flavor observables like tauonic B meson decays and $b \rightarrow s\gamma$ are affected by the presence of the charged Higgs boson. These currently provide the best lower indirect limit on the charged Higgs boson mass ($m_{H^\pm}$) in the Type-II 2HDM [23, 24]. Complementary searches for additional neutral heavy Higgs bosons decaying to a pair of third generation fermions have been performed by the ATLAS and CMS Collaborations at $\sqrt{s} = 8$ and 13 TeV in $t\bar{t}$, $b\bar{b}$, and $\tau\tau$ decay channels [25–30]. The production of charged Higgs bosons via vector boson fusion with decays via W and Z bosons is predicted by models containing Higgs triplets [31]. These searches are discussed in Refs. [32, 33].

The results presented here are based on proton-proton collision data collected at $\sqrt{s} = 13$ TeV by the CMS experiment in 2016, corresponding to an integrated luminosity of 35.9 $fb^{-1}$. The
search investigates fully hadronic events, targeting signal events with hadronic W boson decays and vetoing events with isolated charged leptons. The fully hadronic $t\bar{b}$ decay of the charged Higgs boson offers the advantage of the largest accessible branching fractions, $\sim45\%$ ($\sim67\%$) for the top associated (s-channel) production. In addition, all the final state objects are visible, enabling the reconstruction of the invariant mass of the charged Higgs boson candidate. This analysis is the first to report results using the fully hadronic $t\bar{b}$ final state to examine top associated and s-channel production of a charged Higgs boson.

This search targets two distinct event topologies. The charged Higgs boson is reconstructed from either four resolved jets, two b-tagged jets and two jets from the hadronic decay of a W boson, or as a single top-flavored or W boson jet paired with one or two b-tagged jets. The second topology is expected for highly boosted final states.

Model-independent upper limits on the product of the charged Higgs boson production cross section and branching fraction into a top and bottom quark-antiquark pair ($\sigma B$), as a function of $m_{H^\pm}$ are presented in this note. These limits can be recast into model dependent limits. The search is sensitive to any narrow resonant charged state decaying to a top quark and a bottom quark. Beyond the 2HDM interpretations, the top-bottom decay mode is relevant in the more general context of exotic resonance searches, motivated by $W'$ boson models [34, 35]. Results are also interpreted in scenario-specific limits, where the underlying free parameters (e.g. $m_{H^\pm}$, $\tan\beta$, defined as the ratio of the vacuum expectation values of the two Higgs doublets, and branching fractions) are fixed by the specific scenario.

2 The 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 solenoid volume are 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 endcap sections. Forward calorimeters extend the pseudorapidity ($\eta$) coverage beyond these barrel and endcap detectors. Muons are measured in gas-ionization detectors embedded in the steel flux-return yoke outside the solenoid. Events of interest are selected using a two-tiered trigger system [36]. The first level, composed of specialized hardware processors, uses information from the calorimeters and muon detectors, while the second level consists of a farm of processors running a version of the full event reconstruction software optimized for fast processing. 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. [37].

3 Event samples, object selections, and event simulation

The main SM backgrounds in this analysis are multijets events produced exclusively through the strong interaction and top quark-antiquark pair production. Other sources of background include Drell–Yan and W+jets (referred to as V+jets) processes, dibosons (WZ, ZZ, WW, VH), tribosons, single top production, and $t\bar{t}+X$ processes with $X = (W, Z, \gamma, H, \text{or } t\bar{t})$.

Simulated samples are produced using various Monte Carlo (MC) event generators. These are used to study the characteristics of the signal, to calculate detector acceptance, and also to estimate part of the SM background. Signal samples are generated using the 4FS at next-to-leading-order (NLO) precision in perturbative QCD with the MADGRAPH5_aMC@NLO v2.3.3 [38] generator for a range of $m_{H^\pm}$ hypotheses from 200 GeV to 3 TeV.
The s-channel signal processes are simulated using LO COMPHEP 4.5.2 [39] following a $W_R'$ model in the small width approximation, in the mass range from 800 GeV up to 3 TeV [35]. The QCD multijet background is simulated using the MadGraph5_aMC@NLO v2.2.2 event generator at LO. The $t\bar{t}$ sample is generated using the Powheg v2.0 [40–42] at NLO in QCD [43] assuming a top quark mass of 172.5 GeV. The V+jets background samples are generated at LO precision with the MadGraph [38] generator while single top quark samples are generated at NLO precision with the MadGraph5_aMC@NLO v2.2.2 and Powheg v2.0 generators [44, 45]. The production of $t\bar{t}$ in association with $W$, $Z$ or $\gamma$ is simulated at NLO using MadGraph5_aMC@NLO v2.2.2. The production of $t\bar{t}$ in association with $H$ where $H$ decays to a $b\bar{b}$ pair is generated using Powheg v2.0 at next-to-next-to-leading order (NNLO). Other decay channels are generated at NNLO in QCD and NLO in electroweak corrections. Parton distribution functions (PDFs) are modeled using the NNPDF3.0 [46] parametrization. Parton showering and fragmentation are performed using the Pythia v8.212 [47] generator. The CUETP8M2T4 [48] tune is used to characterize the underlying event in the $t\bar{t}$ background, while the CUETP8M1 [49] tune is used for all other background processes. The samples are normalized to the most precise available cross section calculations, corresponding most often to NLO or NNLO [50–62] accuracy. The production cross sections for the heavy charged Higgs boson signals are computed in the 4FS and 5FS schemes, which differ at finite order in perturbation theory. These are combined to obtain the total production cross section using the Santander matching scheme [13]. Typical values are of the order of 1 pb for a mass of 200 GeV, down to about $10^{-4}$ pb for a mass of 3 TeV [12, 63–67]. Branching fractions $B(H^+ \to t\bar{b})$ are computed with the HDECAY package [68] for different values of $\tan \beta$. The response of the CMS detector for all generated samples is simulated using the Santander matching scheme [13]. Typical values are of the order of 1 pb for a mass of 200 GeV, down to about $10^{-4}$ pb for a mass of 3 TeV [12, 63–67]. Branching fractions $B(H^+ \to t\bar{b})$ are computed with the HDECAY package [68] for different values of $\tan \beta$. The response of the CMS detector for all generated samples is simulated using the Santander matching scheme [13]. Typical values are of the order of 1 pb for a mass of 200 GeV, down to about $10^{-4}$ pb for a mass of 3 TeV [12, 63–67]. Branching fractions $B(H^+ \to t\bar{b})$ are computed with the HDECAY package [68] for different values of $\tan \beta$. The response of the CMS detector for all generated samples is simulated using the Santander matching scheme [13]. Typical values are of the order of 1 pb for a mass of 200 GeV, down to about $10^{-4}$ pb for a mass of 3 TeV [12, 63–67]. Branching fractions $B(H^+ \to t\bar{b})$ are computed with the HDECAY package [68] for different values of $\tan \beta$. The response of the CMS detector for all generated samples is simulated using the Santander matching scheme [13]. Typical values are of the order of 1 pb for a mass of 200 GeV, down to about $10^{-4}$ pb for a mass of 3 TeV [12, 63–67]. Branching fractions $B(H^+ \to t\bar{b})$ are computed with the HDECAY package [68] for different values of $\tan \beta$. The response of the CMS detector for all generated samples is simulated using the Santander matching scheme [13]. Typical values are of the order of 1 pb for a mass of 200 GeV, down to about $10^{-4}$ pb for a mass of 3 TeV [12, 63–67]. Branching fractions $B(H^+ \to t\bar{b})$ are computed with the HDECAY package [68] for different values of $\tan \beta$. The response of the CMS detector for all generated samples is simulated using the Santander matching scheme [13]. Typical values are of the order of 1 pb for a mass of 200 GeV, down to about $10^{-4}$ pb for a mass of 3 TeV [12, 63–67]. Branching fractions $B(H^+ \to t\bar{b})$ are computed with the HDECAY package [68] for different values of $\tan \beta$. The response of the CMS detector for all generated samples is simulated using the Santander matching scheme [13]. Typical values are of the order of 1 pb for a mass of 200 GeV, down to about $10^{-4}$ pb for a mass of 3 TeV [12, 63–67]. Branching fractions $B(H^+ \to t\bar{b})$ are computed with the HDECAY package [68] for different values of $\tan \beta$. The response of the CMS detector for all generated samples is simulated using the Santander matching scheme [13]. Typical values are of the order of 1 pb for a mass of 200 GeV, down to about $10^{-4}$ pb for a mass of 3 TeV [12, 63–67]. Branching fractions $B(H^+ \to t\bar{b})$ are computed with the HDECAY package [68] for different values of $\tan \beta$. The response of the CMS detector for all generated samples is simulated using the Santander matching scheme [13]. Typical values are of the order of 1 pb for a mass of 200 GeV, down to about $10^{-4}$ pb for a mass of 3 TeV [12, 63–67]. Branching fractions $B(H^+ \to t\bar{b})$ are computed with the HDECAY package [68] for different values of $\tan \beta$. The response of the CMS detector for all generated samples is simulated using the Santander matching scheme [13]. Typical values are of the order of 1 pb for a mass of 200 GeV, down to about $10^{-4}$ pb for a mass of 3 TeV [12, 63–67]. Branching fractions $B(H^+ \to t\bar{b})$ are computed with the HDECAY package [68] for different values of $\tan \beta$. The response of the CMS detector for all generated samples is simulated using the Santander matching scheme [13]. Typical values are of the order of 1 pb for a mass of 200 GeV, down to about $10^{-4}$ pb for a mass of 3 TeV [12, 63–67]. Branching fractions $B(H^+ \to t\bar{b})$ are computed with the HDECAY package [68] for different values of $\tan \beta$. The response of the CMS detector for all generated samples is simulated using the Santander matching scheme [13]. Typical values are of the order of 1 pb for a mass of 200 GeV, down to about $10^{-4}$ pb for a mass of 3 TeV [12, 63–67]. Branching fractions $B(H^+ \to t\bar{b})$ are computed with the HDECAY package [68] for different values of $\tan \beta$. The response of the CMS detector for all generated samples is simulated using the Santander matching scheme [13]. Typical values are of the order of 1 pb for a mass of 200 GeV, down to about $10^{-4}$ pb for a mass of 3 TeV [12, 63–67]. Branching fractions $B(H^+ \to t\bar{b})$ are computed with the HDECAY package [68] for different values of $\tan \beta$. The response of the CMS detector for all generated samples is simulated using the Santander matching scheme [13]. Typical values are of the order of 1 pb for a mass of 200 GeV, down to about $10^{-4}$ pb for a mass of 3 TeV [12, 63–67]. Branching fractions $B(H^+ \to t\bar{b})$ are computed with the HDECAY package [68] for different values of $\tan \beta$. The response of the CMS detector for all generated samples is simulated using the Santander matching scheme [13]. Typical values are of the order of 1 pb for a mass of 200 GeV, down to about $10^{-4}$ pb for a mass of 3 TeV [12, 63–67]. Branching fractions $B(H^+ \to t\bar{b})$ are computed with the HDECAY package [68] for different values of $\tan \beta$. The response of the CMS detector for all generated samples is simulated using the Santander matching scheme [13]. Typical values are of the order of 1 pb for a mass of 200 GeV, down to about $10^{-4}$ pb for a mass of 3 TeV [12, 63–67]. Branching fractions $B(H^+ \to t\bar{b})$ are computed with the HDECAY package [68] for different values of $\tan \beta$. The response of the CMS detector for all generated samples is simulated using the Santander matching scheme [13]. Typical values are of the order of 1 pb for a mass of 200 GeV, down to about $10^{-4}$ pb for a mass of 3 TeV [12, 63–67]. Branching fractions $B(H^+ \to t\bar{b})$ are computed with the HDECAY package [68] for different values of $\tan \beta$. The response of the CMS detector for all generated samples is simulated using the Santander matching scheme [13]. Typical values are of the order of 1 pb for a mass of 200 GeV, down to about $10^{-4}$ pb for a mass of 3 TeV [12, 63–67]. Branching fractions $B(H^+ \to t\bar{b})$ are computed with the HDECAY package [68] for different values of $\tan \beta$. The response of the CMS detector for all generated samples is simulated using the Santander matching scheme [13]. Typical values are of the order of 1 pb for a mass of 200 GeV, down to about $10^{-4}$ pb for a mass of 3 TeV [12, 63–67].
Jets consistent with originating from a heavy-flavor hadron are identified using the combined secondary vertex (CSV) b tagging algorithm [79], at the medium or loose working points. These are defined such that the efficiency to select light-flavor quarks (u, d, or s) or gluons as b jets is about 1% or 10%, and the corresponding efficiency for tagging jets from a b quark decay is 65% or 80%.

The scalar $p_T$ sum of all selected jets in an event is denoted as $H_T$, while the missing transverse momentum vector $\vec{p}_T^{\text{miss}}$ is defined as the projection onto the plane perpendicular to the beam axis of the negative vector sum of the momenta of all reconstructed PF candidates [80]. Its magnitude is referred to as $p_T^{\text{miss}}$. Quality requirements are applied to remove a small fraction of events in which detector effects, such as the electronic noise, can affect the $p_T^{\text{miss}}$ reconstruction. The energy scale corrections applied to jets are propagated to the calculation of $H_T$ and $p_T^{\text{miss}}$.

A collection of large radius jets is used to reconstruct and identify the boosted objects from the W boson and top quark decays by means of the anti-$k_T$ clustering algorithm with a distance parameter of 0.8. In order to discriminate against multijet backgrounds, we exploit both the reconstructed jet mass, which is required to be close to the W boson or top quark mass, and the two or three prong jet substructure produced by the particle cascades corresponding to the two quark jets of a W boson decay and to three jets in the case of top quark candidates. The soft-drop algorithm [81] is used to remove soft and wide-angle radiation. The use of soft-drop grooming reduces the jet mass $m_{SD}$ for background QCD multijet events when large jet masses arise from soft-gluon radiation. The W (top) candidates are required to have $65 < m_{SD} < 105$ GeV (135 < $m_{SD} < 220$ GeV), $p_T > 200 (400)$ GeV, and $|\eta| < 2.4$. The hard substructures are identified using the N-subjettiness [82] ratios ($\tau_3/\tau_2 < 0.67$ for the top jet and $\tau_2/\tau_1 < 0.6$ for the W jet). Finally, because top-quark jets contain a b quark and W jets do not, additional discrimination power is achieved by using a subjet b tagging criterion [79].

### 4 Search strategy

The analysis aims to reconstruct the full event in order to search for a local enhancement in the top and bottom quark-antiquark invariant-mass spectrum.

Because of the large cross section for the multijet background, restrictive trigger requirements are needed to reduce the data-recording rate. The data used for this search are collected with an inclusive online selection of $H_T^{\text{miss}} > 900$ GeV, with $H_T^{\text{miss}}$ being defined as the scalar sum of small radius jets with $p_T > 30$ GeV. Events are also acquired with a dedicated large radius jet trigger requiring $p_T > 360$ GeV and a mass after jet trimming [83] of at least 30 GeV. Furthermore, events satisfying trigger requirements of $H_T^{\text{miss}} > 540(400)$ GeV and six jets with $p_T > 40(30)$ GeV, are selected if at least one (two) of them satisfies b tagging criteria. In the low $m_{H^\pm}$ regions, the sensitivity of the hadronic final state is limited by the relatively high trigger thresholds.

Two analyses are performed, targeting different regions of signal parameter space. The first analysis targets charged Higgs bosons with high mass. It uses collimated hadronically decaying top quark or W boson candidates to distinguish signal events, while strongly suppressing background processes. It is expected that the decay products of resonances with mass of $\mathcal{O}(\text{TeV})$ have average transverse momenta of several hundred GeV. As a consequence, the particles emerging from subsequent decays are very collimated and cannot be resolved by the standard clustering algorithm, but are instead reconstructed as a single large radius jet. In less boosted signal scenarios, the hadronic decay products of the W boson may be reconstructed in
4. Search strategy

Dedicated techniques are applied to exploit the substructure of these objects. We refer to this as the boosted analysis. The events are categorized according to the number of jets containing a b hadron decay to further separate the SM background and capture signals with both high and low number of b quarks.

The second analysis is optimized for charged Higgs bosons with lower masses that decay to moderately boosted top quarks, often resolved as three separate small cone jets, one of which is b-tagged and the other two jets resulting from the W boson decay. The resolved top quark candidates (t\textsubscript{res}) are identified using a multivariate boosted decision tree with gradient boost (BDTG) classifier. This is trained on simulated top quark-antiquark pair events using the TMVA package [84]. The classifier exploits properties of the top quark and its decay products such as masses, angular separations, and other kinematic properties. Additional input variables are quark-gluon, charm-light and b tagging discriminator values for each of the three jets. The signal enriched region is defined by requiring events with at least seven jets, three or more of which are b-tagged, and containing two identified resolved top quark candidates. This analysis is referred to as the resolved analysis.

In order to reduce the contribution of W and Z boson decays, both analyses veto the presence of an isolated charged lepton (electron or muon) with $p_T \geq 10$ GeV, or an isolated hadronically decaying tau lepton with $p_T \geq 20$ GeV. To further reduce semileptonic final states, the boosted analysis requires events to have $p_T^{miss} \leq 200$ GeV. The lepton veto ensures that leptonic final states of $H^\pm$ boson decays are not considered. These are covered by dedicated analyses [17].

4.1 Event categories in the boosted analysis

Events with at least one b jet are considered. The highest $p_T$ b jet is chosen among the small radius jets that satisfy the loose working point of the b jet tagger. A top quark candidate is identified as top jet or the combination of a W jet and subleading in $p_T$ b jet with the closest invariant mass to the top-quark mass. The charged Higgs boson candidate is then reconstructed as the sum of the four-vectors of the top quark candidate and the highest $p_T$ b jet. We introduce four mutually exclusive categories. The labels “t1b” and “t0b” refer to events containing a large radius jet identified as top jet where at least one, or none, of the subjets satisfies the medium working point of the b tagging algorithm. The “wbb” and “wbj” categories require the presence of a W jet paired with two or one additional medium-tagged b jets, respectively. The wbj category required a second additional jet satisfying the loose working point. If multiple candidates are available, the jet having the largest CSV discriminator value is used.

The signal is characterized as a peak in the invariant-mass spectrum of top and bottom quark candidates, which mostly consist of SM multijet processes. The shape of the charged Higgs boson candidate mass distribution is dominated by the detector resolution and the wrong-pairing background, where jets are not correctly matched to the decay products of the charged Higgs boson. The full width at half maximum of the reconstructed mass distribution for correct jet assignments is used to describe the mass resolution and events falling outside this window are used to constrain the background. The typical mass resolution is of the order of 150 GeV for a charged Higgs boson with a mass of 1 TeV. The distribution of the invariant mass is shown in Fig. 2. We search for an excess of events in the $H_T$ data distribution in a window around different charged Higgs boson candidate masses in each of the categories listed above.

To better separate signal from background the event categories are subdivided to exploit differences in jet and b-jet multiplicities. For signal events produced in association with a top quark, we expect at least three b quarks in the final state and a large number of extra jets not participating in the charged Higgs boson reconstruction. Signal produced in the s-channel contains
Figure 2: The SM background for the event sample with one top jet as function of the charged Higgs boson candidate mass. The category t1b is shown. The mass distribution for the model with \( m_{H^\pm} = 1 \text{ TeV} \) is displayed on top of the backgrounds and normalized with a cross section times branching fraction of 1 pb. The signal mass window “in” is shown together with the sidebands “below” and “above” for the mass hypothesis of 1 TeV.

Two b quarks and fewer extra jets. We therefore consider different requirements on the number of b jets: exactly one b jet, exactly two b jets and at least three b jets. We also distinguish two categories based on the number of additional small radius jets, either less than three or at least three such jets.

The signal-rich regions are analyzed together with signal-depleted regions using a binned maximum likelihood fit to the data that simultaneously determines the contributions from signal and the major background sources.

4.2 Event selection in the resolved analysis

A multivariate analysis is employed to select top quark candidates in events containing seven or more jets. We employ a BDTG classifier that is trained using a simulated t\( T \) sample. The signal objects are considered to be three small radius jet combinations in which each individual jet is matched to the decay product of a top quark at generator level. Similarly, background objects are defined as three-jet combinations in which at least one jet is not matched to a top quark decay product. The input variables used for the BDTG training (19 in total), calculated from these jet combinations, are described in detail in Ref. [85]. In the BDTG response distribution values close to -1 are mainly populated by top quark candidates from QCD multijet processes, while values close to +1 are dominated by top quark candidates from t\( T \) or signal events. In this analysis we require resolved top candidates to have a BDTG score > 0.4, yielding a signal object efficiency of 92% and a background object efficiency of 6%.

Events with at least three b jets passing the CSV medium working point and at least four additional jets are selected. The first top quark candidate is identified by pairing each b-tagged jet with all two-jet combinations and retaining the combination having the highest BDTG value.
The same procedure is applied for the second candidate using only the remaining jets as inputs. To reduce the combinatorial background, we require the combined three-jet system to have invariant mass less than 400 GeV.

The efficiency of the BDTG requirement as a function of the $p_T$ of the generated top quark in $t\bar{t}$ events is shown in Fig. 3 (left) along with the misidentification rate observed in a QCD multijet sample. At the plateau the tagging efficiency reaches 50%. The observed decrease in efficiency in the high-$p_T$ region is due to top quark decay products becoming increasingly collimated, resulting in a jet-to-parton matching inefficiency. The misidentification rate is less than 8% for the entire $p_T$ range considered.

To reconstruct the invariant mass of the charged Higgs boson candidate, we use the resolved top quark candidate with the highest $p_T$, shown in Fig. 3 (right), and the leading in $p_T$ b jet that is not used in the reconstruction of the two selected top candidates. The invariant mass of the charged Higgs boson candidate is used in a binned maximum likelihood fit to extract the signal in the presence of the SM background.

5 Backgrounds

The dominant backgrounds arise from QCD multijet and top-quark production associated with additional light flavor, charm or bottom quarks. Contributions from more rare processes such as single top, $V$+jets, dibosons, tribosons, and four top-quark production, are found to be small.
5.1 Background estimations in events with boosted W boson and top quark candidates

We estimate the multijet and top-quark backgrounds using a data-driven method that exploits a number of the background enriched regions. These control regions are included in a simultaneous fit with the signal enriched regions to determine the normalization and the shape of the background distributions.

Because the cross section for multijet production is large, this background can produce many events satisfying our signal selection requirements. The distribution of $m_{SD}$ for signal peaks around the W-boson or the top-quark mass, while the multijets background spectrum is peaked at lower soft-drop masses. This background is estimated from simulation and corrected to data using a control region enriched with jets arising from the hadronization of single quarks or gluons. The control region is defined by reverting the N-subjettiness requirements used to identify the top and W jets. The normalization is determined for each event category using sideband regions around the signal mass windows in the invariant mass spectrum of the top and bottom pair. We validate this correction by applying the technique in define an orthogonal control region, defined by requiring that no b-tagged jets are identified. The shapes of the N-subjettiness distributions and kinematics of jets having $m_{SD}$ consistent with either a top jet or W jet are found to be consistent with events passing the signal selection.

The contribution from the $t\bar{t}$ process arises from fully hadronic final states or states with a leptonic decay of a W boson where the charged lepton is outside the kinematic acceptance of the CMS detector or evades identification by the dedicated lepton vetoes. Such events contain a pair of genuine b quarks and boosted W jets and top jets. In order to normalize the $t\bar{t}$ background, a lepton enriched set of events is used to describe the kinematics for the top-quark pair production and the normalization is left floating in the final fit. The control region is defined by requiring a lepton (e, $\mu$) with $10 < p_T < 35$ GeV, $p_T^{miss} > 100$ GeV, at least one b jet. This ensures orthogonality with the searches for charged Higgs bosons in the leptonic channels [17].

5.2 Background estimations in events with resolved top quarks

The main backgrounds for the resolved analysis can be decomposed into events containing either genuine b jets or events with at least one light quark or gluon jet erroneously tagged as a b jet. We refer to the latter as fake b jets. This is measured with a data-driven technique through the use of control regions that are defined by inverting the BDTG requirement, the b jet selection, or both. Background containing genuine b jets is modeled using simulation.

The shape of the charged Higgs boson candidate mass distribution in the background is obtained from events orthogonal to the signal region (SR) by requiring that only two (of at least three) b jets pass the CSV medium working point, and the remaining jets only pass the loose CSV working point. This region is referred to as the application region (AR). In order to compensate for the different selection efficiencies between these two regions, transfer factors are used to normalize the AR to the SR. These transfer factors are determined by taking the ratio of events in two additional control regions that are orthogonal to each other and to both the AR and the SR. The first control region is obtained by inverting the BDTG requirement for the second top quark candidate, and the second control region is obtained by also altering the b jet selection as described above for the AR. In order to minimize the effect of kinematic differences between the loose and medium working points, the background from fake b-tagged jets is evaluated separately in $p_T$ and $|\eta|$ bins of the b-tagged jet used in the reconstruction of the invariant mass of the charged Higgs boson candidate.
6. Systematic uncertainties

Because the SR and associated control regions are mutually orthogonal, the expected yield of fake b events passing the signal selections can be predicted as:

$$N_i^{SR} = \sum_i N_i^{AR} \cdot \left( \frac{N_i^{CR1}}{N_i^{CR2}} \right),$$  \hspace{1cm} (1)

where CR1(2) refers to the first (second) control region and the index $i$ runs over all $p_T$ and $|\eta|$ bins of the aforementioned leading b jet.

6 Systematic uncertainties

The systematic uncertainties are divided into two categories: those that affect the estimation of the background from SM processes, and those that affect the expected signal yields.

The events used in this search are largely collected where the trigger efficiency is close to 100%. The trigger efficiency is extracted from data and the uncertainties in trigger correction factors applied to the simulation are less than 5%.

The uncertainty from pileup modeling is estimated by varying the total inelastic cross section of 69.2 mb by 5% [86]. The uncertainty on the integrated luminosity is estimated to be 2.5% [87].

Uncertainties on the background prediction that also affect the signal arise from the jet energy scale [88], from the scale factors correcting the efficiency and misidentification rate for b tagging [79], and from the reconstruction and identification efficiencies of the leptons. In addition, uncertainties arising from the simulation-to-data corrections for boosted top and W tagging and the BDTG response are applied in the boosted and resolved analyses, respectively. The variations in the jet selection and jet energy scale are propagated to the $H_T$, $p_T^{miss}$ and $H^\pm$ candidate yields and invariant mass.

For the boosted analysis a normalization uncertainty of 50% is applied for the QCD multijet background. This uncertainty is treated to be uncorrelated among the boosted top- and W-tagged event categories and it is 100% correlated within each category and across the signal regions and the side bands. An additional uncertainty to account for shape variations in modeling the $H_T$ observable is parametrized linearly as a function of $H_T$ and reaches 30% for an $H_T$ of 1 TeV. These uncertainties are then constrained by studying the control region used to correct the simulation and the resulting variation in expected QCD multijet background yield is approximately 23%.

The systematic uncertainties affecting the fake b background measurement in the resolved analysis can be divided into three components. The first component consists of events containing jets from b quark decays that fail the b tagging requirement and is subtracted from the control regions used in the measurement. The uncertainty on the normalization of this component is estimated by propagating all the uncertainties related to the simulation of electroweak and top quark processes. The other two components account for the statistical and systematic uncertainties in determining the transfer factors applied in the normalization of the AR. Statistical fluctuations in the value of the transfer factors can result in rate and shape differences in the predicted invariant mass distribution. Similarly, the definition of the control region affects the individual transfer factors and subsequently the invariant mass shape in the AR. The aforementioned contributions affect the expected event yield by approximately 5%.

For $t\bar{t}$ and single top quark processes, the effect of the top quark mass on the cross sections is
estimated by varying the top quark mass by 1.0 GeV around the nominal value of 172.5 GeV. Uncertainties from factorization and renormalization scales in the inclusive cross sections are estimated for each simulated process by varying the matrix element scales independently from each other by factors of 0.5 and 2 with respect to the default value. The PDF uncertainties are treated as fully correlated for all processes that share the same dominant partons in the initial state of the matrix element (i.e. gg, gq, or qq) [89].

Finally, the limited number of simulated background and signal events leads to statistical fluctuations in the nominal predictions. The effects are considered in the limit calculations using a Barlow–Beeston lite approach [90, 91], which assigns the combined statistical uncertainty in each bin to the process dominating the background yield in that bin.

Tables 1 and 2 summarize the various sources of systematic uncertainty and their impact on signal yield and the expected background in data, for the boosted and resolved analyses respectively.

Table 1: The systematic uncertainties affecting signal and background for the boosted analysis, evaluated prior to fitting to data, summed over all final states and categories. The numbers are given in percentage and describe the effect of each nuisance parameter on the overall background normalization. Nuisance parameters with a check mark also affect the shape of the $H_T$ spectrum. Sources that do not apply in a given category are marked with long-dashed lines. For the $H^\pm$ signal, the values for $m_{H^\pm} = 1$ TeV are shown.

<table>
<thead>
<tr>
<th>Source</th>
<th>Shape</th>
<th>$H^\pm$</th>
<th>QCD multijet</th>
<th>$t\bar{t}$</th>
<th>$tW, t\bar{t} + X$</th>
<th>Electroweak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trigger efficiency</td>
<td></td>
<td>5.0</td>
<td>4.5</td>
<td>0.39</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>Pileup</td>
<td>✓</td>
<td>0.42</td>
<td>1.4</td>
<td>0.05</td>
<td>$&lt; 0.01$</td>
<td>0.03</td>
</tr>
<tr>
<td>Integrated luminosity</td>
<td></td>
<td>2.5</td>
<td>—</td>
<td>0.2</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>Lepton efficiency</td>
<td></td>
<td>5</td>
<td>—</td>
<td>0.39</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>Jet energy scale and resolution</td>
<td>✓</td>
<td>3.0</td>
<td>5.8</td>
<td>0.4</td>
<td>0.04</td>
<td>0.12</td>
</tr>
<tr>
<td>b-jet identification</td>
<td>✓</td>
<td>2.4</td>
<td>12</td>
<td>0.24</td>
<td>0.03</td>
<td>0.12</td>
</tr>
<tr>
<td>Unclustered $p_T^{miss}$ energy scale</td>
<td>✓</td>
<td>0.23</td>
<td>—</td>
<td>0.02</td>
<td>$&lt; 0.01$</td>
<td>0.01</td>
</tr>
<tr>
<td>Jet $m_{SD}$ scale</td>
<td>✓</td>
<td>1.3</td>
<td>2.5</td>
<td>0.07</td>
<td>0.02</td>
<td>—</td>
</tr>
<tr>
<td>N-subjettiness scale</td>
<td></td>
<td>2.0</td>
<td>—</td>
<td>0.17</td>
<td>0.02</td>
<td>—</td>
</tr>
<tr>
<td>QCD normalization</td>
<td></td>
<td>—</td>
<td>28</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>QCD shape</td>
<td>✓</td>
<td>—</td>
<td>$&lt; 0.01$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Top quark mass</td>
<td></td>
<td>—</td>
<td>—</td>
<td>0.21</td>
<td>0.02</td>
<td>—</td>
</tr>
</tbody>
</table>

Theory source

<table>
<thead>
<tr>
<th>Source</th>
<th>Shape</th>
<th>$H^\pm$</th>
<th>QCD multijet</th>
<th>$t\bar{t}$</th>
<th>$tW, t\bar{t} + X$</th>
<th>Electroweak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale (acceptance)</td>
<td>✓</td>
<td>2.1</td>
<td>—</td>
<td>0.53</td>
<td>—</td>
<td>0.04</td>
</tr>
<tr>
<td>Scale, PDF (cross section)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.43</td>
<td>0.04</td>
<td>0.05</td>
</tr>
</tbody>
</table>

7 Results and interpretation

The expected SM event yields from a background-only fit to the data are shown in Fig. 4 and Table 3 for the boosted and resolved analyses, respectively. For the boosted analysis, the background predictions are broken down into various categories of signal and background enriched regions and in total 98 distributions are fitted. The shape of the $H_T$ distribution in the boosted analysis and the invariant mass of the charged Higgs boson in the resolved analysis are used to assess the agreement with the background-only hypothesis or the presence of the signal and
Table 2: The systematic uncertainties of the backgrounds and the signal for the resolved analysis, evaluated prior to fitting to data, summed over all final states and categories. The numbers are given in percentage and describe the effect of each nuisance parameter on the overall background normalization. Nuisance parameters with a check mark also affect the shape of the $H^\pm$ candidate mass spectrum. Sources that do not apply in a given category are marked with long-dashed lines. For the $H^\pm$ signal, the values for $m_{H^\pm} = 500$ GeV are shown.

<table>
<thead>
<tr>
<th>Source</th>
<th>Shape</th>
<th>$H^\pm$ fake b</th>
<th>$t\bar{t}$</th>
<th>$t_\ell W, t_\ell +X$</th>
<th>Electroweak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trigger efficiency</td>
<td>5.0</td>
<td>0.09</td>
<td>0.69</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td>Pileup</td>
<td>✓</td>
<td>&lt; 0.01</td>
<td>—</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Integrated luminosity</td>
<td>2.5</td>
<td>0.09</td>
<td>0.35</td>
<td>0.02</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Lepton efficiency</td>
<td>0.32</td>
<td>—</td>
<td>0.04</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Jet energy scale and resolution</td>
<td>✓</td>
<td>8.5</td>
<td>0.24</td>
<td>1.6</td>
<td>0.09</td>
</tr>
<tr>
<td>b-jet identification</td>
<td>✓</td>
<td>5.0</td>
<td>—</td>
<td>0.64</td>
<td>0.04</td>
</tr>
<tr>
<td>$t_\text{res}$ tagging</td>
<td>✓</td>
<td>8.9</td>
<td>0.24</td>
<td>1.8</td>
<td>0.04</td>
</tr>
<tr>
<td>Transfer factors</td>
<td>✓</td>
<td>—</td>
<td>4.0</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Top quark mass</td>
<td>—</td>
<td>0.09</td>
<td>0.39</td>
<td>0.02</td>
<td>—</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Theory source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale (acceptance)</td>
</tr>
<tr>
<td>Scale, PDF (cross section)</td>
</tr>
</tbody>
</table>

are shown in Fig. 5. The contribution of a hypothetical charged Higgs boson with a mass of 1 TeV or 800 GeV and $\sigma B = 1$ pb is also displayed.

Table 3: Number of expected and observed events for the resolved analysis after all selections. For background processes, the event yields and their corresponding uncertainties are prior to the background-only fit to the data. For the $H^\pm$ mass hypotheses of 500, 650, and 800 GeV, the signal yields are normalized to a $\sigma B = 1$ pb and the total systematic uncertainties prior to the fit are shown.

<table>
<thead>
<tr>
<th>Process</th>
<th>Events ± (stat) ⊕ (syst)</th>
</tr>
</thead>
<tbody>
<tr>
<td>fake b</td>
<td>6152 ± 292</td>
</tr>
<tr>
<td>Genuine b</td>
<td>1067 $^{+185}_{-187}$</td>
</tr>
<tr>
<td>Total expected from the SM</td>
<td>7220 ± 336</td>
</tr>
<tr>
<td>Observed</td>
<td>7124</td>
</tr>
<tr>
<td>$H^\pm$ signal, $m_{H^\pm} = 500$ GeV</td>
<td>183 ± 26</td>
</tr>
<tr>
<td>$H^\pm$ signal, $m_{H^\pm} = 650$ GeV</td>
<td>218 $^{+30}_{-31}$</td>
</tr>
<tr>
<td>$H^\pm$ signal, $m_{H^\pm} = 800$ GeV</td>
<td>234 ± 33</td>
</tr>
</tbody>
</table>

The observed data agree with the predicted SM background processes. The results of the search are interpreted to set upper limits on the product of the charged Higgs boson production cross section and the branching fraction into a top and bottom quark-antiquark pair. The upper limits are calculated at the 95% confidence level (CL) using the LHC-style CLs method [92, 93] and the asymptotic approximation for the test statistics [94, 95]. These results are shown in Fig. 6 for the associated and s-channel productions, respectively.

Exclusion limits are placed on the production cross section of a charged Higgs boson associated
with a top quark, $\sigma_{H^{\pm}}B(H^{\pm} \to tb) = \sigma(pp \to H^{\pm}t\bar{b})B(H^{\pm} \to t\bar{b}) + \sigma(pp \to H^{-}t\bar{b})B(H^- \to t\bar{b})$, for masses from 200 GeV to 3 TeV in the range 21.3 to 0.007 pb. The boosted analysis has the best sensitivity for $m_{H^{\pm}}$ larger than 800 GeV while the resolved analysis limits are most stringent at lower masses. The boosted analysis sets upper limits from 4.5 to 0.023 pb on the charged Higgs boson production cross section in the s-channel, $\sigma(pp \to H^{\pm})B(H^{\pm} \to tb)$, for masses from 800 GeV to 3 TeV extending previous results [14].

Model-dependent upper limits are obtained by comparing the observed limit with theoretical predictions. The MSSM $m_{h}^{mod-}$ benchmark scenario [11] is designed to give a mass of approximately 125 GeV for the light CP-even Higgs boson over a wide region of the parameter space. The $M_{125}^{\tilde{\chi}}$ scenario [96] is characterized by significant mixing between higgsinos and gauginos, a compressed EW-ino mass spectrum, and with a phenomenology that resembles the Type-II 2HDM with MSSM-inspired Higgs couplings compatible with $m_{h} \approx 125$ GeV for large masses of the $m_{A}$ boson. Figure 7 shows the excluded parameter space in the MSSM scenarios. In the $m_{h}^{mod-}$ scenario tan $\beta$ values from 0.25 to 0.86 are excluded for $m_{H^{\pm}}$ values between 200 and 1.2 TeV, while in the $M_{125}^{\tilde{\chi}}$ scenario tan $\beta$ values from 0.45 to 0.86 are excluded for $m_{H^{\pm}}$ between 200 and 650 GeV.
Figure 5: Variables used in the limit extraction. The $H_T$ distribution for the boosted analysis (left), summed over all associated production channels, with the expected signal shown for $m_{H^\pm} = 1$ TeV. The invariant mass of the charged Higgs boson candidates for the resolved analysis (right), with the expected signal shown for $m_{H^\pm} = 800$ GeV. The distributions are binned according to the statistical precision of the samples, leading to wider bins in the tail of the distributions.
Figure 6: Upper limits at 95% CL on the cross section times branching fraction as function of $m_{H^±}$ for the process $\sigma_{H^±} B(H^± \to t b)$ (left) and for $\sigma(pp \to H^±) B(H^± \to t b)$ (right). The observed upper limits are shown by the solid black markers. The median expected limit (dashed line), 68% (inner green band), and 95% (outer yellow band) confidence-interval expected limits are also shown.

Figure 7: Excluded parameter space region in the $m_{h^{\text{mod}}}$ scenario (left) and $M_{125}(\tilde{\chi})$ (right). The observed upper limits are shown by the solid black markers. The median expected limit (dashed line), 68% (inner green band), and 95% (outer yellow band) confidence-interval expected limits are also shown. The region below the red line is excluded assuming that the observed neutral Higgs boson is the light CP-even 2HDM Higgs boson with a mass of $125 \pm 3$ GeV, where the uncertainty is the theoretical uncertainty in the mass calculation.
8 Combination with the leptonic final states

In Ref. [17] a search is presented for a charged Higgs boson with mass greater than the top quark and decaying into a top and bottom quark-antiquark pair in the complementary leptonic final states. Events are selected by the presence of a single isolated charged lepton (electron or muon) or an opposite sign dilepton pair ($ee, \mu\mu, e\mu$). These are categorized according to the jet multiplicity and number of b-tagged jets and multivariate techniques are used to enhance the signal and background discrimination in each category. The search is based on the same proton-proton collision data collected by the CMS experiment at a center-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 35.9 fb$^{-1}$.

These results are combined with the fully hadronic channel analyses to calculate 95% CL combined upper limits on the product of the cross section and the branching fraction as function of the $m_{H^\pm}$ for the process $\sigma_{H^\pm} B(H^\pm \rightarrow tb)$. The limits are shown in Fig. 8 and Table 4. The common experimental and theoretical nuisance parameters that share the same mechanism between the final states are correlated while the uncertainties from different sources described in Section 6 are assumed to be uncorrelated. The single lepton final state has the best sensitivity in the whole $m_{H^\pm}$ range from 200 GeV to 3 TeV, while the dilepton channel contributes in the low $m_{H^\pm}$ regime, i.e. $\leq 1.5$ TeV, and the fully hadronic channel improves the overall sensitivity at larger values of $m_{H^\pm}$.

![Figure 8](image_url)

Figure 8: Upper limits at 95% CL on the cross section times branching fraction as function of $m_{H^\pm}$ for the process $\sigma_{H^\pm} B(H^\pm \rightarrow tb)$. The median expected limit (dashed line), 68% (inner green band), and 95% (outer yellow band) confidence-interval expected limits are also shown (left). The relative expected contribution of each channel to the overall combination is shown (right). The black dashed corresponds to the combined expected limits while the red, magenta, and blue represent the contributing channels.

9 Summary

Results are presented from a search for charged Higgs bosons decaying into a top and a bottom quark in the fully hadronic final state. The search targets two distinct event topologies. The charged Higgs boson is reconstructed from either four resolved jets, two b-tagged jets and
Table 4: The upper limit at 95% CL on the $\sigma_{H^\pm} B(H^\pm \to tb)$ with the combined fully hadronic, single-lepton, and dilepton final states.

<table>
<thead>
<tr>
<th>$m_{H^\pm}$ (GeV)</th>
<th>Expected limits (pb)</th>
<th>Observed limits (pb)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-2 s.d.</td>
<td>-1 s.d.</td>
</tr>
<tr>
<td>200</td>
<td>2.02</td>
<td>2.71</td>
</tr>
<tr>
<td>220</td>
<td>1.25</td>
<td>1.69</td>
</tr>
<tr>
<td>250</td>
<td>0.86</td>
<td>1.15</td>
</tr>
<tr>
<td>300</td>
<td>0.66</td>
<td>0.89</td>
</tr>
<tr>
<td>350</td>
<td>0.45</td>
<td>0.61</td>
</tr>
<tr>
<td>400</td>
<td>0.36</td>
<td>0.48</td>
</tr>
<tr>
<td>500</td>
<td>0.19</td>
<td>0.26</td>
</tr>
<tr>
<td>650</td>
<td>0.12</td>
<td>0.16</td>
</tr>
<tr>
<td>800</td>
<td>0.078</td>
<td>0.11</td>
</tr>
<tr>
<td>1000</td>
<td>0.051</td>
<td>0.069</td>
</tr>
<tr>
<td>1500</td>
<td>0.022</td>
<td>0.030</td>
</tr>
<tr>
<td>2000</td>
<td>0.012</td>
<td>0.017</td>
</tr>
<tr>
<td>2500</td>
<td>0.008</td>
<td>0.011</td>
</tr>
<tr>
<td>3000</td>
<td>0.005</td>
<td>0.007</td>
</tr>
</tbody>
</table>

two jets from the hadronic decay of a W boson, or as a single top-flavored or W boson jet paired with one or two b-tagged jets. The data are collected with the CMS detector in 2016 at a center-of-mass energy $\sqrt{s} = 13$ TeV with an integrated luminosity of 35.9 fb$^{-1}$. No significant deviation above the expected background is observed. Model-independent upper limits at 95% confidence level are set on the charged Higgs boson production cross section times branching fraction in two scenarios. For production in association with a top quark, limits of 21.3 to 0.007 pb are calculated for charged boson masses in the range 200 GeV to 3 TeV. Combination with data from a search in the leptonic final states results in improved limits of 9.25 to 0.005 pb. The complementary s-channel production of a charged Higgs boson is investigated in the mass range 800 GeV to 3 TeV and the corresponding upper limits are 4.5 to 0.023 pb. Exclusion regions in the parameter space of the minimal supersymmetric standard model $m_{h}^{mod}$ and $M_{h}^{125}(\tilde{\chi})$ benchmark scenarios are presented.

References


[22] CMS Collaboration, “Search for charged Higgs bosons in the $H^\pm \rightarrow \tau^\pm \nu_\tau$ decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV”, arXiv:1903.04560. Submitted to *JHEP*.


