CDF/PHYS/ELECTROWEAK/PUB/11219

Search for Higgs-like particle produced in association with b-quark jets at CDF II

The CDF Collaboration

(Dated: July 21, 2017)

Abstract

This note describes the search for a Higgs-like particle decaying into two b-quark jets produced in association with b-quark jets. We search for an enhancement in the invariant mass of the two b-quarks leading jets in triply b tagged events. A limit on the cross section times the branching ratio into $b\bar{b}$ is set using a sample of 5.4 fb^{-1} of $p\bar{p}$ events collected with a dedicated on-line trigger path. The invariant mass shape of the dominant QCD background is parametrized using a data-driven technique, validated with the measurement of the $Z \to b\bar{b}$ cross section, reducing in this way the dependence of the analysis on simulation.

CONTENTS

I. INTRODUCTION

The discovery of Higgs boson [1, 2] with measured properties in agreement with the expectations of the Standard Model (SM) does not exclude the existence of new neutral scalar particles ϕ, direct indication of new physics. Many SM extensions predict new particles that strongly couple to the b quark and in general the $b\bar{b}$ decay mode is relevant in any context of exotic resonance searches. Higgs like particles decaying into b-quark jets are foreseen for example in the minimal supersymmetric extension of the SM (MSSM) [3], in two-Higgs-doublet models (2HDM) [4] and dark-matter models involving mediator particles with a large coupling to b quarks [5, 6].

In this note, a search for a narrow neutral scalar particle ϕ decaying into b-quark jets in multi b-quark jets final states is described. Since an inclusive search is difficult due to large multijet backgrounds, the analysis relies on the case where the ϕ boson is produced in association with one or more b-quarks. Final state with at least three b-quark jets represents a powerful search channel, with the third b-quark jet providing additional suppression of the large multijet background.

Searches for heavy resonances decaying into $b\bar{b}$ jets, and produced in association with b-quark jets have already been performed by the CDF [7] and D0 [8] experiments at the Tevatron collider and by the CMS [9] experiment at LHC. The two Tevatron experiments have reported a deviation, at the level of 2σ, from the SM expectations in the two b-quark jets invariant mass around $100 - 150$ GeV/c^2 [10]. The CMS collaboration has excluded a resonance compatible with a MSSM model particle and has set an upper limit in the $(\mathrm{M}_A, \tan\beta)$ parameter space.

The analysis presented here has been performed with data collected by the CDF II detector at Tevatron, using a data sample corresponding to 5.4 fb^{-1}. Events are selected by requiring online at least one b-quark jet [11] in order to suppress the large rate of multijet background production. The different initial production state and the lower center of mass energy with respect to LHC makes this search competitive with the CMS experiment measurement especially in the low b-quark jets invariant mass region, where the 2σ deviation has been observed. Because of the various possible theoretical frameworks, the analysis is kept model independent, i.e. no particular theoretical model is tested and the upper limit is set on the production cross section $\sigma(p\bar{p} \to \phi b) \times \mathcal{B}r(\phi \to b\bar{b})$.

In Sec. II data and Monte Carlo samples are described, while in Sec. III the evaluation of the background is explained. The fitting procedure used to measure signal and background events is reported in Sec. IV with the systematic uncertainties in Sec. V. The results and conclusions are summarized in Sec. VI and Sec. VII.

II. DATA SELECTION

A data sample corresponding to 5.4 fb^{-1} of integrated luminosity was collected with the DIJET_BTAG trigger [11]. It was optimized for $H \to b\bar{b}$ events selection and it was used efficiently with any final state with b-quark jets. The trigger algorithm performed an on-line b-quark jet tagging exploiting the long b hadron lifetime by searching for tracks coming from a secondary vertex displaced from the primary one. It also looked for two calorimeter jets, whose energy thresholds were kept as low as possible, 15 GeV, allowing to search for dijet resonances at low invariant mass. For more details about the trigger selection see [11, 12].

The b tagging algorithm used in the analysis is the so called SecVtx [13], which searches

for a secondary vertex inside the jet formed by the decay of a b hadron. The offline event selection requires at least three jets, each of them with a tight SecVtx tag and $E_{\mathrm{T}} > 22$ GeV.

To summarize:

- b **tag trigger** jet is a jet that fired the DIJET_BTAG trigger and has a tight SecVtx tag;
- b **tagged** jet is a jet that has a tight SecVtx tag;

A. Simulated data

Monte Carlo samples are used to evaluate the efficiency and the acceptance of the ϕb signal process at different ϕ mass points and to extract the response of the SecVtx-tagging algorithm to the different jet flavor.

Signal samples for a variety of ϕ masses are generated using Pythia 6.216 [14], MSUB=121 which correspond to the $gg \to b\bar{b}h^0$ process with a p_{T} cut of 15 GeV/c on a quark which can be either the b or the $\bar{b}$.

The signal events selection efficiencies vary from 3.7‰ to 8.7‰ as a function of the invariant mass of the neutral scalar and are shown in Figure 1.

As described in [12], simulated data samples have been scaled to reproduce the data by applying scale factors for the trigger and the SecVtx b tagging.

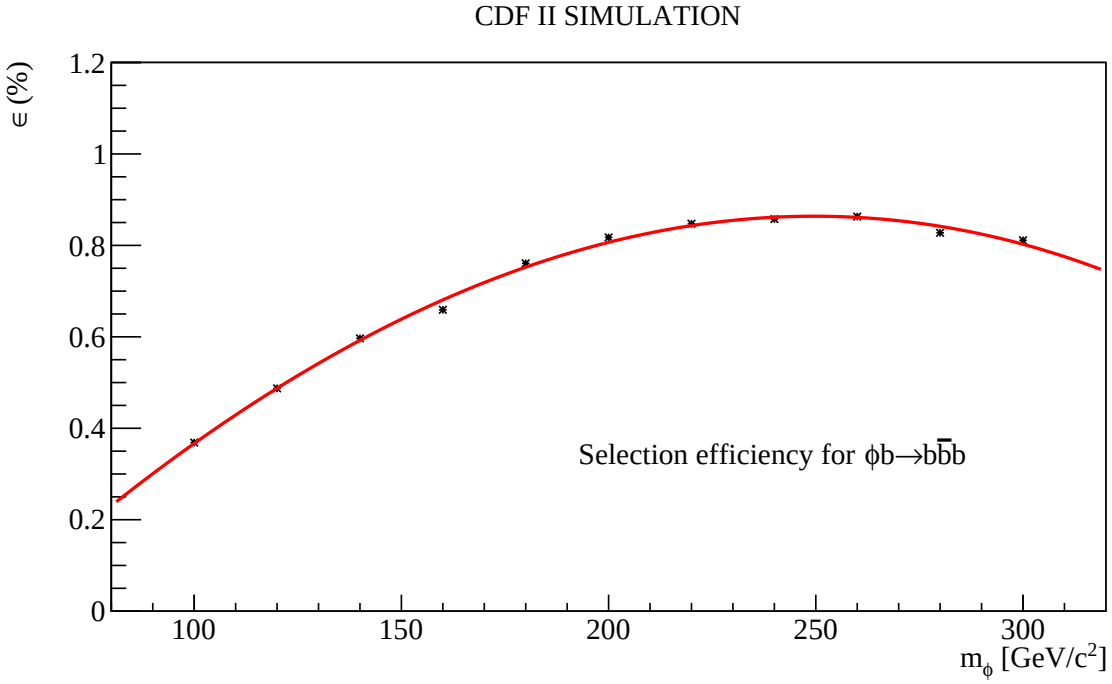


FIG. 1. Signal acceptance as function of the mass of the neutral scalar particle.

III. BACKGROUND ESTIMATE

The data sample is mainly composed by heavy flavour multijets, originating by a large number of production mechanisms [15] for which the rates suffer of large uncertainties.

Heavy quark production can be categorized into three types of processes: Flavor Creation, Flavor Excitation and Gluon Splitting. Flavor Creation, $q\bar{q} \to b\bar{b}+X$, refers to the lowest-order QCD $b\bar{b}$ production diagrams. This process includes $b\bar{b}$ production through $q\bar{q}$ annihilation and gluon fusion, plus higher-order corrections to these processes. Because this production is dominated by two-body final states, it tends to yield $b\bar{b}$ pairs that are back-to-back in $\Delta\phi$ and balanced in p_{T}.

Flavor Excitation, $bq \to bq + X$, refers to diagrams in which a $b\bar{b}$ pair from the quark sea of the proton or antiproton is excited into the final state because one of the quarks from the $b\bar{b}$ pair undergoes a hard QCD interaction with a parton from the other beam particle. Because only one of the quarks in the $b\bar{b}$ pair undergoes the hard scatter, this production mechanism tends to produce b quarks with asymmetric p_{T}. Often, one of the b quarks will be produced with high rapidity and not be detected in the central region of the detector.

Gluon Splitting, $qg \to qg + X$ followed by $g \to b\bar{b}$, refers to diagrams where the $b\bar{b}$ pair arises from a gluon splitting in the initial or final state. Neither of the quarks from the $b\bar{b}$ pair participates in the hard QCD scatter. Depending on the experimental range of b quark p_{T} sensitivity, gluon splitting production can yield a $b\bar{b}$ distribution with a peak at small $\Delta\phi$.

It is possible to obtain more than two heavy quarks in the final state by combining these processes in a single event. Given all the possible final states with heavy quarks, it is not possible to rely on direct calculation of the multijet production.

Other processes that can contribute to the heavy flavour multijet are the $Z + jet$ and the $t\bar{t}$ productions. We expect a contribution of these processes to be less that 1% of the total events and it is already included in the double tagged events sample used to build the background templates.

A. Data driven background templates

The measurement of the $Z \to b\bar{b}$ production cross section [12] has been performed using the same data sample and a similar analysis technique. The signal events are obtained by fitting the double tagged sample using background templates built starting from the single tagged sample.

As already shown in the MSSM Higgs analysis [7], the fact that the triple-tagged jets sample predominantly contains at least two b-quark jets is of major importance. The double tagged sample is then the natural starting point to build the different heavy flavour multijet background templates used to describe the triple tagged data. The effect of requiring a third tag, whose efficiency depend upon the flavor of the jet, can be simulated using a parametrization of the SecVtx response evaluated using MC samples. The flavor composition of the triple tagged jets sample will ultimately be determined by fitting the data.

The tagging probabilities, used to parametrize the response of the SecVtx algorithm, represent the efficiency to tag a b-, c- and light quark initiated jet as a b-quark jet as function of its E_{T} and η and therefore are referred as tagging matrices. They are constructed per jet, under the assumption that the probability to tag a jet depends only on its kinematics and not on the event topology, therefore the tagging matrices determined for the $Z \to b\bar{b}$ analysis [12] are used.

The events in the double-btag sample, with an additional third untagged jet, are organized in two categories, bbx and xbb. These categories depend on the $E_{\rm T}$ rank of the untagged jet which is represented by the lower-case letter x, with the caveat that no distinction is made between the two leading jets. The ranking in descending $E_{\rm T}$ of the three jets is incorporated in the nomenclature adopted here, e.g. xbb means a sample of events where the third leading jet and either one of the two leading jets is tagged. From these categories six background templates are constructed by weighting the untagged jet with the tagging matrices for the different flavour hypotheses: light quark ($Q = udsg$), charm (C) or beauty (B).

B. The $x_{\rm tags}$ variable definition

To better discriminate among the multijet production mechanisms a second variable derived from the TagMass is introduced in the fit templates, alongside m_{12} the invariant mass of the two highest momentum b-quark jets. The TagMass is sensitive to the flavor of the parton initiating the jet. Light quarks and gluons, which can generate a secondary vertex tag only due to tracks mis-measurements, as well as c quarks initiated jets have Tagmass distribution peaking at lower values with respect to b quark initiated jets. Figure 2 shows the Tagmass distribution for MC jets coming from b, c and light quarks.

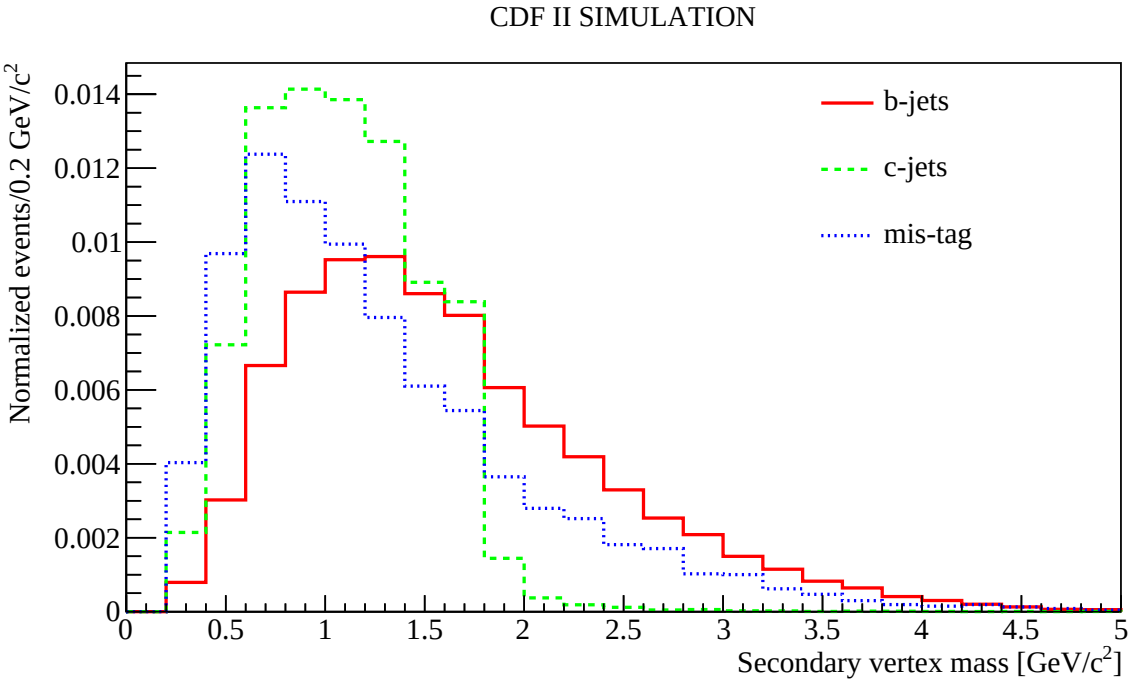


FIG. 2. Tagmass distribution for b-, c- and light-jets as obtained with Monte Carlo simulation.

The variable $x_{\rm tags}$ allows a better separation between backgrounds with high and low TagMass values. We built $x_{\rm tags}$ following the recipe described in Ref [7]. Because no distinction is made between the two leading jets in the flavor classification scheme, $x_{\rm tags}$ is constructed to be symmetric under their interchange, as it is m_{12}.

We define the $x_{\rm tags}$ variable as:

$$x_{\rm tags} = \begin{cases} min(m_{3,\rm tag}, 3) & : \quad m_{1,\rm tag} + m_{2,\rm tag} < 2 \\ min(m_{3,\rm tag}, 3) + 3 & : \; 2 < m_{1,\rm tag} + m_{2,\rm tag} < 4 \\ min(m_{3,\rm tag}, 3) + 6 & : \quad m_{1,\rm tag} + m_{2,\rm tag} > 4 \end{cases},$$

where $m_{1,\rm tag}$ is the TagMass of the leading jet, $m_{2,\rm tag}$ is the TagMass of the second leading jet and $m_{3,\rm tag}$ is the TagMass of the third leading jet.

The $m_{1,\rm tag} + m_{2,\rm tag}$ provides the sensitivity to *bcb* and *bqb* components versus the other components, the $m_{3,\rm tag}$ separates out *bbc* and *bbq*.

In order to compute $x_{\rm tags}$ for the background components we need to simulate not only the bias on the invariant mass m_{12} due to the requirement of the third tag, but also its expected value of TagMas. Therefore the tagging matrices are parametrized as function of jet $E_{\rm T}$, η and TagMass.

Because the untagged jet in double-tagged events does not have any Secondary Vertex to compute $x_{\rm tags}$, we assign all possible values of the TagMass for that jet, with the caveat of weighting it with the proper weight taken from the tagging matrices. For each weighted TagMass value we then compute $x_{\rm tags}$ and fill the background template histogram.

By construction then, each event has multiple entries in the background template, each with the same value of m_{12} but with varying values for $x_{\rm tags}$.

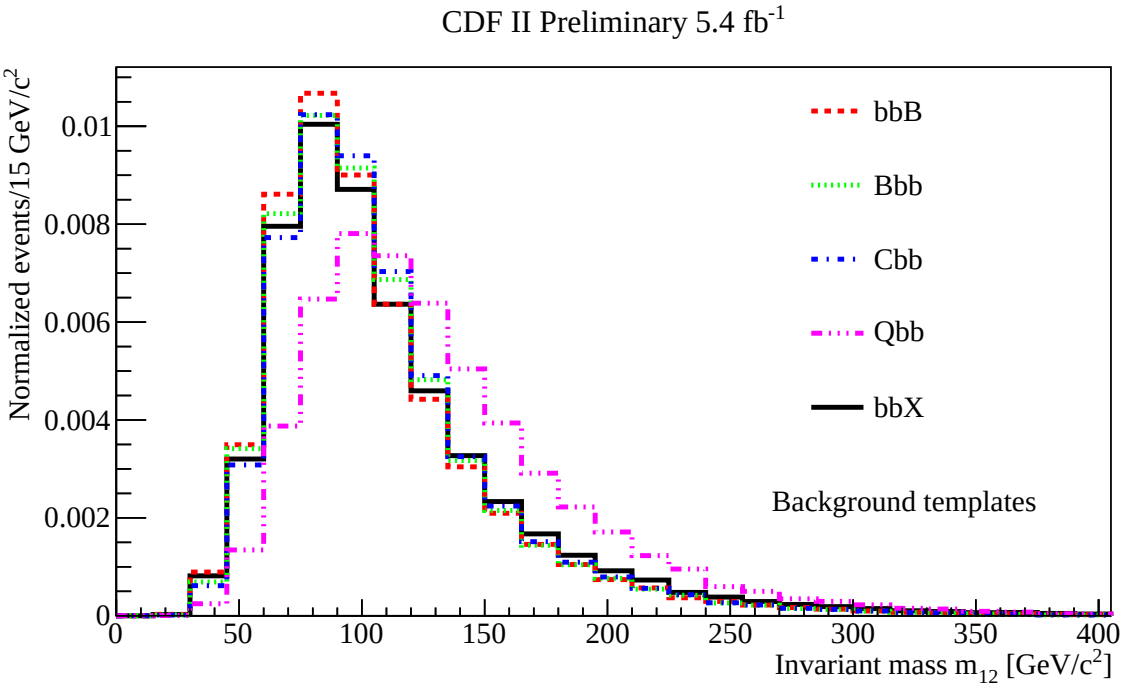
Figure 3 shows the m_{12} and $x_{\rm tags}$ distributions of heavy flavor multijet background components. The average of the *bbC* and *bbQ* templates (*bbX*) is used in the fit because they are too similar to discriminate. The *bbx* double tagged sample is composed of about 130k events, the *xbb* double tagged sample is composed of about 140k events.

C. Background model validation

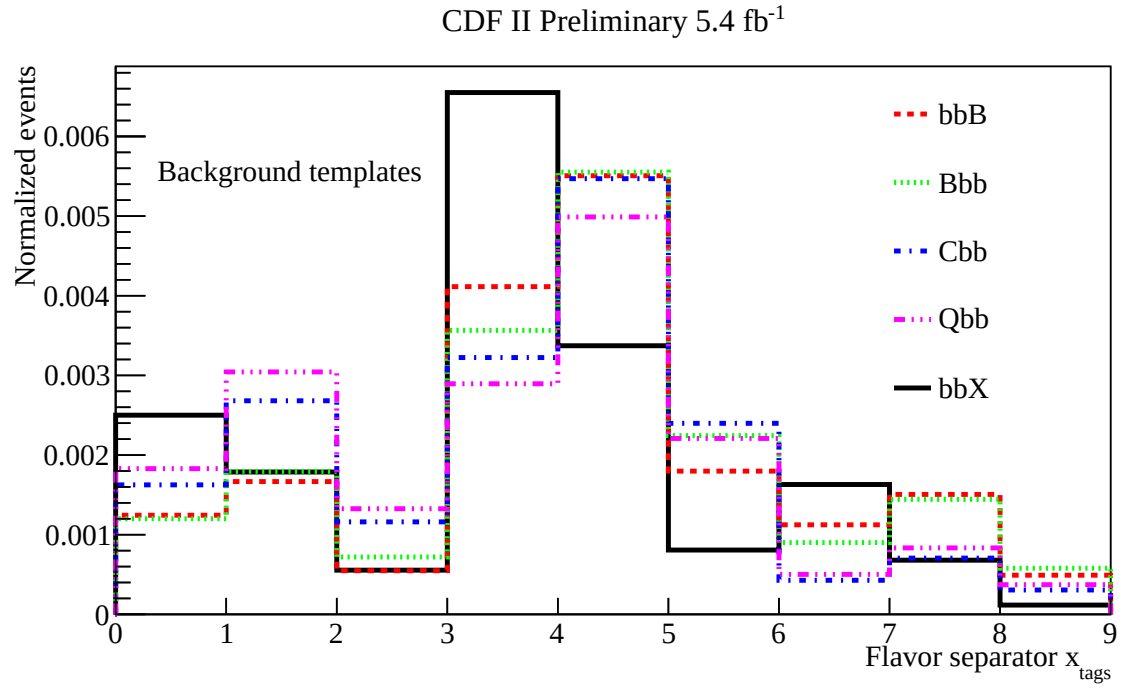
The model used to parameterize the background contributions is extracted from simulated data and it is tested using the collected Tevatron data. The events with two positive and one negative tagged jets, where the negative tag can be on any of the three jets, are selected from the initial data sample. A jet has a negative tag if the secondary vertex is found on the opposite side of the primary vertex respect to the jet direction. Jets negatively tagged are predominantly constituted by light-flavor tagged jets because of the finite resolution on the position of the tracking system. This sample is expected to be almost a pure sample of *bbq* and *qbb* events. In order to verify the background parametrization, the selected data is fit by a two dimensional binned maximum-likelihood with m_{12} and $x_{\rm tags}$ as variables . Figure 4 shows the result of the fit of the data projected on the variables m_{12} and $x_{\rm tags}$. Data is described by a combination of the *bbQ* and *Qbb* templates only, as expected.

IV. FIT DESCRIPTION AND RESULTS

A binned maximum-likelihood method is used to fit data, 5616 events with triple tagged jets. The likelihood function is a joint probability of the Poisson likelihood distribution of each bin $\nu_{ij}^{n_{ij}} e^{-\nu_{ij}}/n_{ij}!$, where n_{ij} is the number of observed events in the i-th bin of m_{12} and



(a)



(b)

FIG. 3. Invariant mass distribution of the two leading jets, m_{12} (a), and the x_{tags} (b) background templates variables.

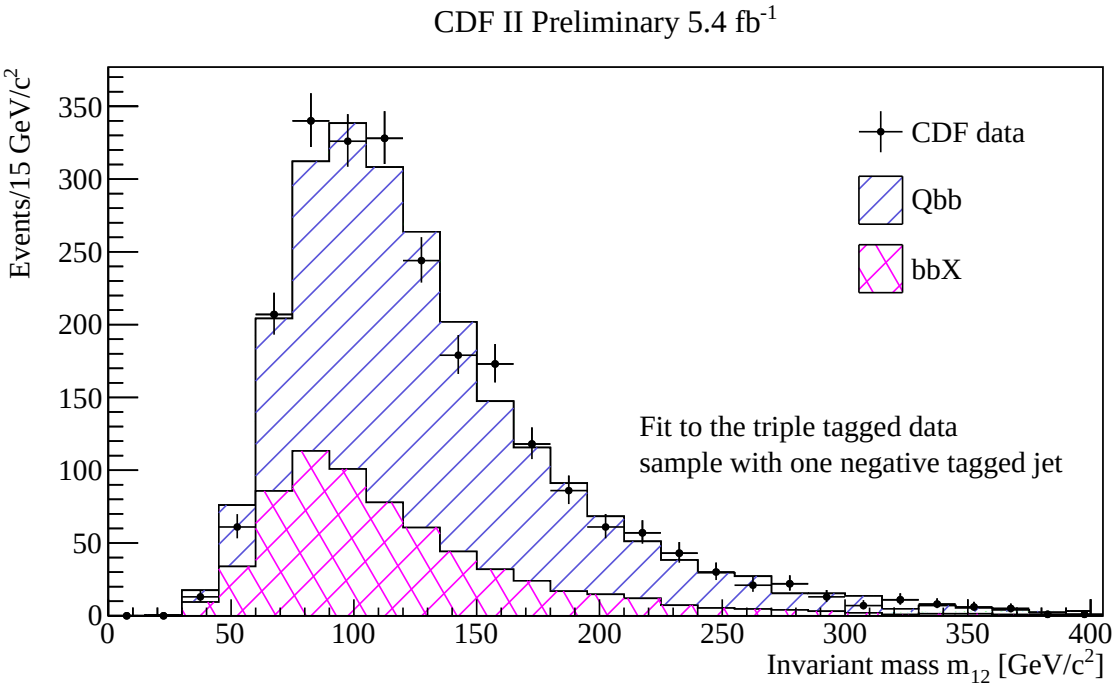


FIG. 4. Invariant mass distribution of the two leading jets in the two positive and one negative tagged jets with the result of the fit projected into the m_{12} variable.

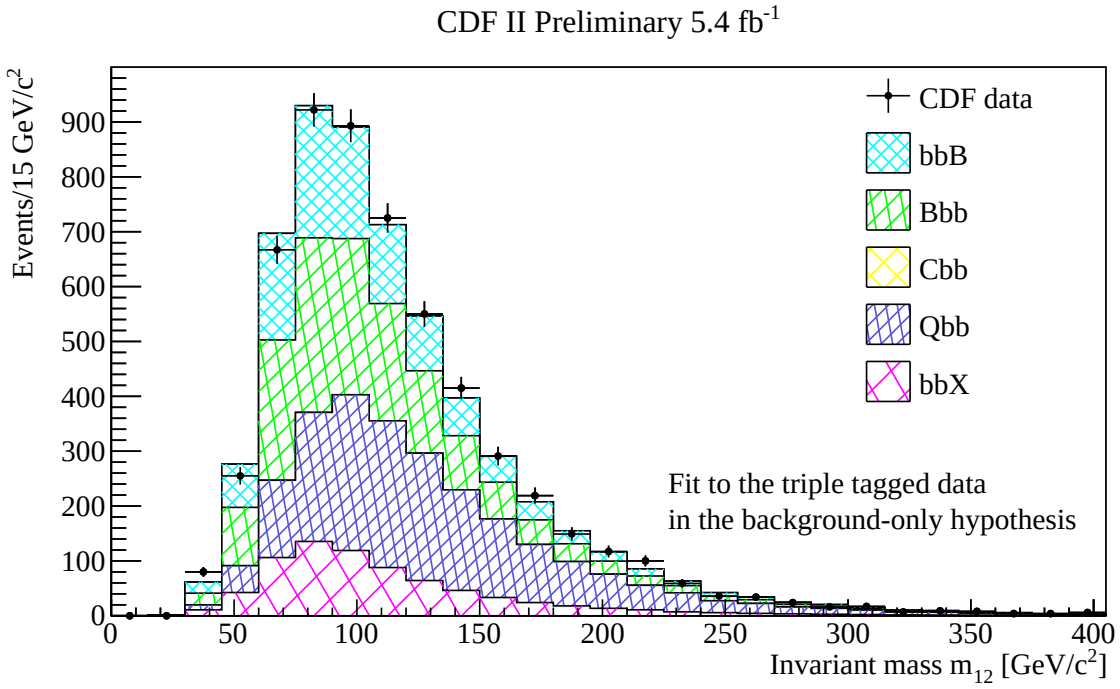
the j-th bin of $x_{\rm tags}$, the expected events in that bin ν_{ij} is given by:

$$\nu_{ij} = \sum_b N_b f_{b,ij} + N_s f_{s,ij};$$

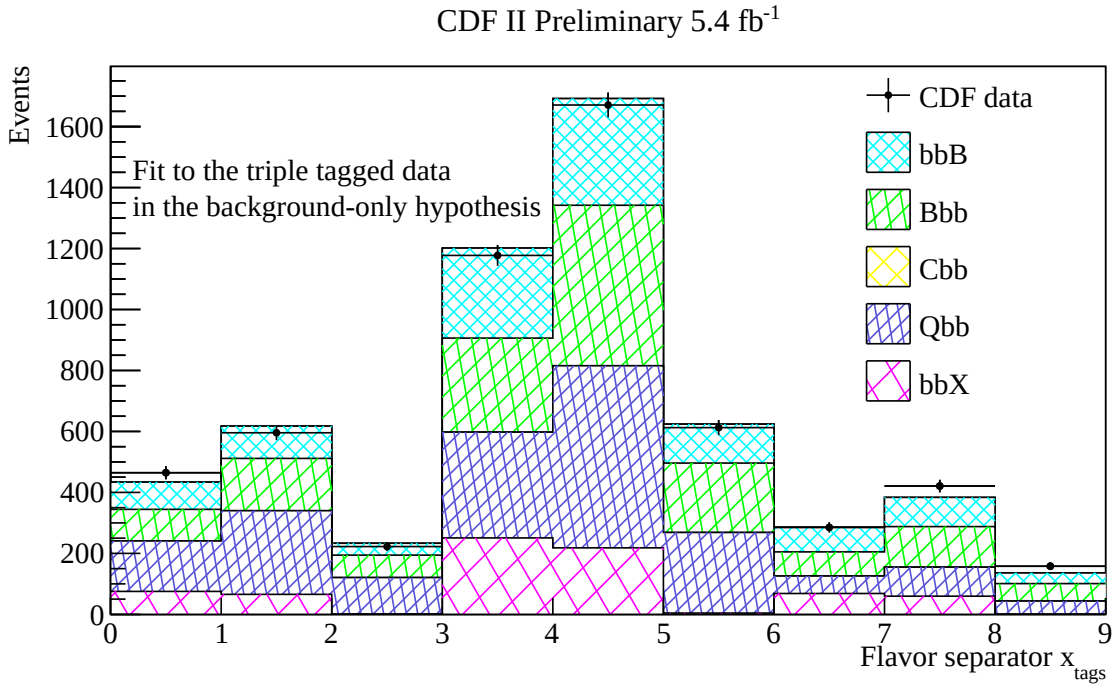
where b represents the five background templates, $f_{b,ij}$ and $f_{s,ij}$ are the bin contents in terms of various background components and of neutral scalar signal. The five N_b and N_s are the free parameters of the fit which represent the normalized number of events of each component. Figure 5 shows the result of the fit projected into m_{12} and $x_{\rm tags}$, when only background templates are considered. No systematic uncertainties are included in the fit, which has a goodness-of-fit $\chi^2/d.o.f.$ of 0.8. Table I summarizes the number of events returned by the fit for each component compared to the expected value obtained as explained in the following.

An a priori estimate of the background components is performed starting from the double tagged sample. The study at MC generator-level in [7], where the relevant analysis conditions are equal to the ones used here, predicts that in events with two b-quark jets, the third jet is from b-quark in the 2% of the time and from c-quark in the 4% of the time. The Monte Carlo study demonstrated that these numbers do not depend on the jets energy ordering. The background templates derived from the double-tagged events are normalized to $N_{b\bar{b}}\epsilon_f$, where $N_{b\bar{b}}$ is the number of double-tagged events and ϵ_f is the predicted tag efficiency for a jet under the flavor hypothesis f, extract from the tagging matrices. Therefore, the normalization of each component in the triple tagged sample is predicted by scaling the number $N_{b\bar{b}}$ by the fraction of jets expected to be of flavor f. The Qbb and the bbQ components of the bbX template, are verified by using the fraction of negative tagged jet in the triple tagged sample.

The predictions match the results of the fit except for the Cbb, where the expected 550 events seem to be included in the Bbb and bbB fitted components. One of the results of the



(a)



(b)

FIG. 5. Triple tagged events fit results projected into m_{12} (a) and x_{tags} (b), under the background only hypothesis.

CDF II Preliminary 5.4 fb^{-1}

Backgrounds in the $\phi b \to b\bar{b}b$ search

Background component	Best fit in the background only hypothesis result	Expected events normalizing the double tagged sample
bbB	1227 ± 891	950
Bbb	1672 ± 738	1280
Cbb	< 90 (1σ)	550
Qbb	1964 ± 169	1820
bbX	742 ± 293	1080

TABLE I. Events yields as returned by the fit to the triple tagged sample in the background only hypothesis, compared to the Standard Model expectations calculated as explained in the text.

fit is that the anti-correlation between the Bbb and bbB components is -0.973. This leads to large fit uncertainties for these two backgrounds.

In order to evaluate the quality of the background only fit to data, the systematic uncertainties have to be considered. The fitting procedure explained above is applied to a set of pseudo-experiments generated including the systematic uncertainties to test the background only hypothesis and the background plus signal one.

V. SYSTEMATICS UNCERTAINTIES

Systematic uncertainties affect both the signal and the background description. They can modify the normalization of the fit results, denoted as rate, and also the m_{12} and x_{tags} distributions of the components templates, denoted as shape. Table II summarizes the systematics uncertainties considered.

CDF II Preliminary 5.4 fb^{-1}

Systematic uncertainties on the $\phi b \to b\bar{b}b$ search

Source	Variation	Applies to	Type
Luminosity	5.9%	Signal	Rate
Offline b-tag	5% per jet	Signal	Rate
Online and offline b-tag combined	4%	Signal	Rate
JES	$7-4\%$	Signal	Rate/Shape
x_{tags}	3%	Signal	Shape
PDFs	2%	Signal	Rate
Template stat. uncertainty	-	Background	Shape
Heavy flavor normalization	5%	Background	Rate

TABLE II. Summary of Systematic Uncertainties.

Online and the offline b tagging systematic uncertainties are taken from Ref. [16] where they have been evaluated for this specific online tagging algorithm. The systematic uncertainty on the signal efficiency due to the CDF jet energy correction is estimated by shifting the energy of the MC jets by $\pm 1\sigma_c$ of the standard jet energy correction. In this way both, the acceptance and the shape of the signal are modified. The acceptance changes from 7% to 4%, depending on the mass of the ϕ. The MC signal samples have been generated using

the default set of PDFs at CDF, the CTEQ5L set. The uncertainty due to this choice has been evaluated, as for Ref [12], by generating MC samples using the CTEQ6L set and taking the difference in acceptance as error. The uncertainty due to the limited statistics of the components templates is taken into account in the limit calculator, following a Poisson statistics for each bin as explained in Sec. VI. The mass of the SecVtx tags, used to build the $x_{\rm tags}$ variable, has been varied by $\pm 3\%$ around the chosen values and the impact on the measurement evaluated. The normalization used for the heavy flavour templates depends on the offline tagging algorithm, and a 5% uncertainty is applied.

VI. CROSS SECTION LIMIT

The search for a neutral scalar Higgs-like resonance, ϕ, is performed in the mass range of $100 - 300$ GeV/c^2 by fitting the m_{12} and the $x_{\rm tags}$ distributions using the procedure described in the previous sections. Figure 6 shows the result of the fit performed including a signal template with the mass of ϕ of 160 GeV/c^2.

To compute the sensitivity and set 95% confidence level upper limits on the production cross section times the branching ratio of a narrow scalar as a function of mass we use a modified frequentist CL_S method [17]. The limit calculator is based on the MCLIMIT package [18]. Simulated experiments are generated based on the background predictions in Table I. The predictions for the numbers of each background type and for the signal are randomly varied for each simulated experiment according to the systematics in Table II. These generated pseudo-experiments are then fitted under the the background only hypothesis and in the background plus signal hypothesis. The expected limit is then computed by using as test statistic the difference in χ^2 of these fits under the different hypotheses. The observed limit is computed following the same procedure fitting the data instead of the pseudo-experiments.

The expected number of events are translated into $\sigma \times \mathcal{B}r$ using the acceptance showed in Figure 1, the integrated luminosity and the data/MC scale factors for online and offline b tagging algorithm.

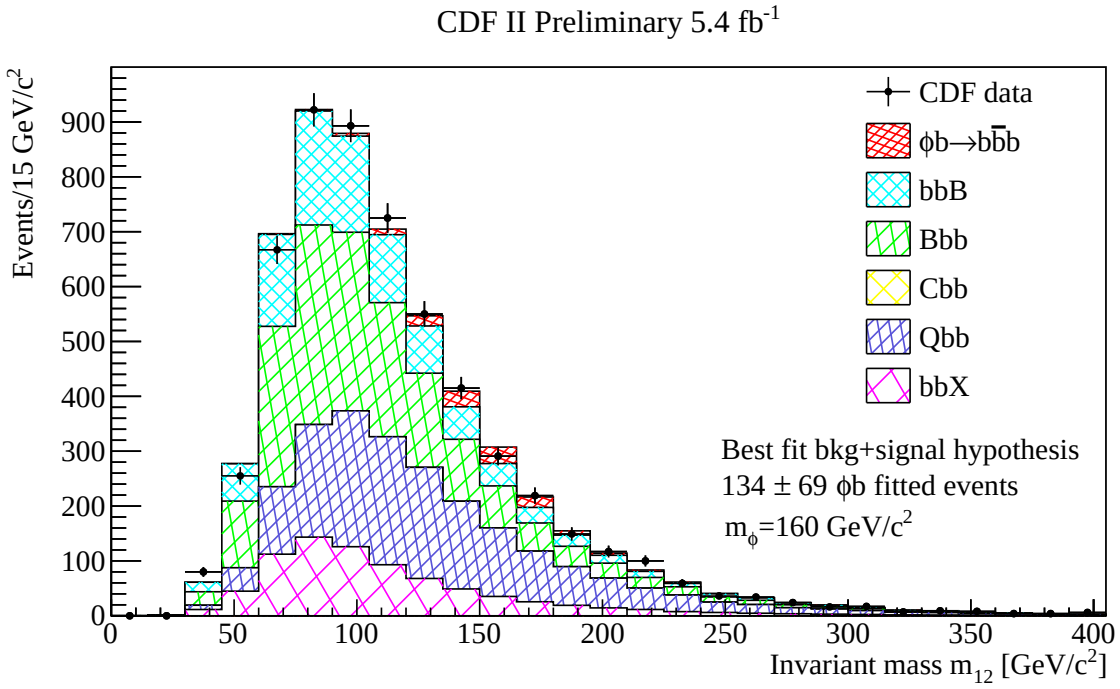
The observed limits and the median expected 95% C.L. limits as a function of the mass of the scalar particle are shown in Table III and in Figure 7. Figure 7 also shows the $\pm 1\sigma$ and $\pm 2\sigma$ bands of the expected limits.

All points of the observed limit are within 2σ band of the expected limit, indicating that there is none statistically significant excess.

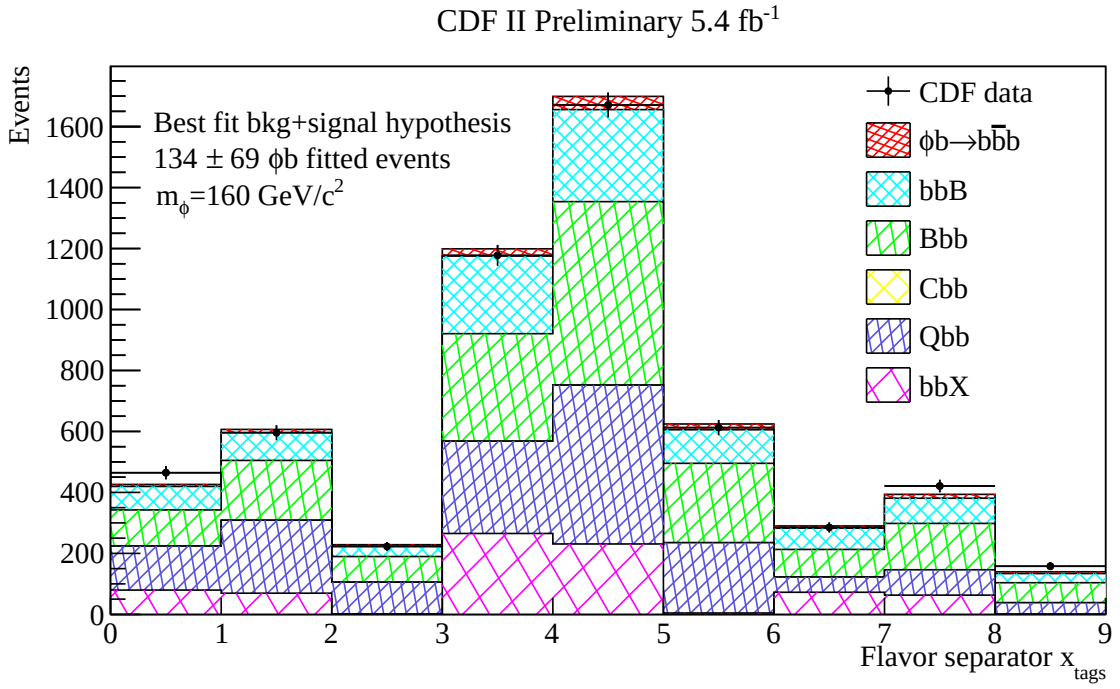
VII. CONCLUSION

A search for a Higgs-like particle decaying into a pair of b-quark jets and produced in association with at least one additional b-quark jet at CDF has been described.

No hint of deviation from the SM background expectations has been observed upper limits on the cross section times branching ratio the $100 - 300$ GeV/c^2 mass range are calculated. The result improves the previous combined limit of CDF and D0 and classifies the 2σ exces



(a)



(b)

FIG. 6. Result of the fit to the triple tagged data projected into m_{12} (a) and $x_{\rm tags}$ (b). A signal component with a mass of the ϕ of 160 GeV/c^2 is added to the background templates.

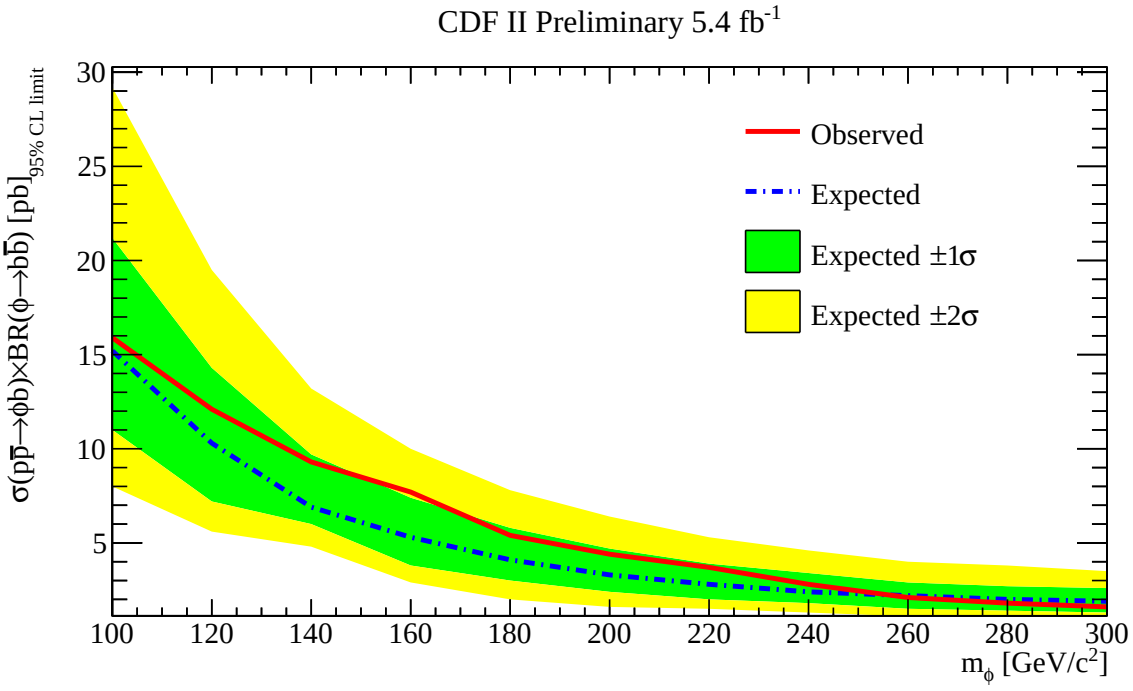
CDF II Preliminary 5.4 fb^{-1}

	95% C.L. upper limit on $\sigma \times \mathcal{B}r$ [pb]	
m_ϕ [GeV/c^2]	Expected	Observed
100	15.2	15.9
120	10.3	12.1
140	6.9	9.3
160	5.3	7.7
180	4.1	5.4
200	3.3	4.4
220	2.8	3.7
240	2.4	2.8
260	2.2	2.1
280	2.0	1.8
300	1.9	1.6

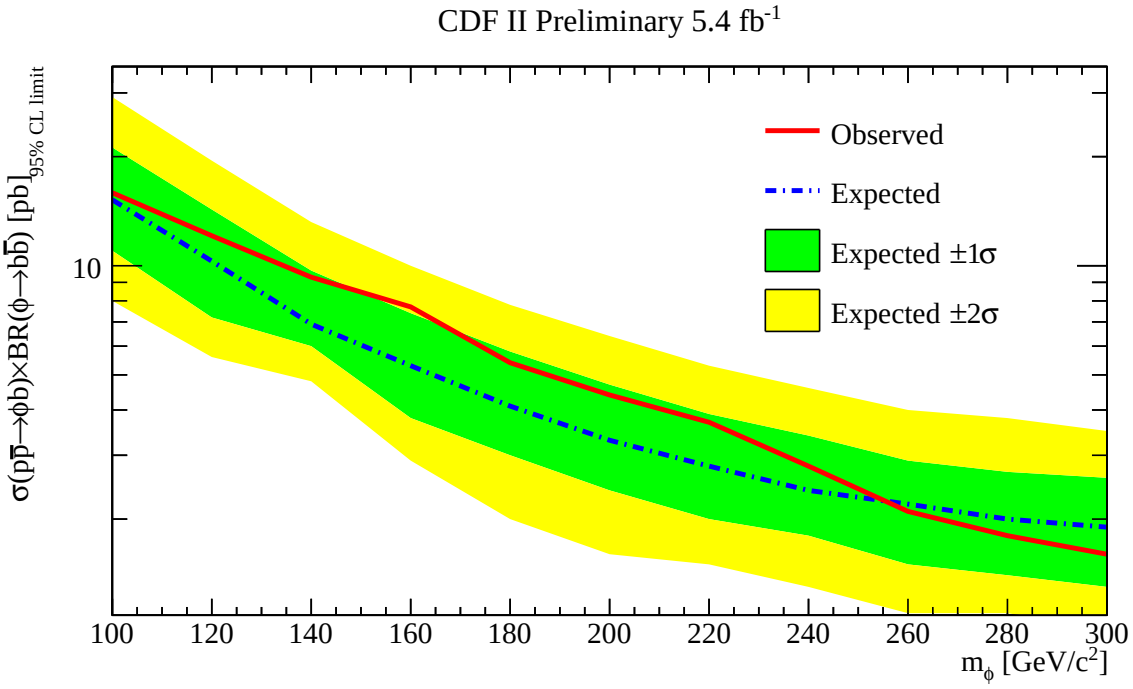
TABLE III. Median expected and observed limits on $\sigma \times \mathcal{B}r$, in pb.

in the 100-160 GeV mass range as statistical fluctuation.

[1] S. Chatrchyan *et al.* (CMS), Phys. Lett. **B716**, 30 (2012), arXiv:1207.7235 [hep-ex].
[2] G. Aad *et al.* (ATLAS), Phys. Lett. **B716**, 1 (2012), arXiv:1207.7214 [hep-ex].
[3] H. P. Nilles, Phys. Rept. **110**, 1 (1984).
[4] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, M. Sher, and J. P. Silva, Phys. Rept. **516**, 1 (2012), arXiv:1106.0034 [hep-ph].
[5] E. Izaguirre, G. Krnjaic, and B. Shuve, Phys. Rev. **D90**, 055002 (2014), arXiv:1404.2018 [hep-ph].
[6] A. Berlin, D. Hooper, and S. D. McDermott, Phys. Rev. **D89**, 115022 (2014), arXiv:1404.0022 [hep-ph].
[7] T. Aaltonen *et al.* (CDF), Phys. Rev. **D85**, 032005 (2012), arXiv:1106.4782 [hep-ex].
[8] V. M. Abazov *et al.* (D0), Phys. Lett. **B698**, 97 (2011), arXiv:1011.1931 [hep-ex].
[9] V. Khachatryan *et al.* (CMS), JHEP **11**, 071 (2015), arXiv:1506.08329 [hep-ex].
[10] T. Aaltonen *et al.* (CDF, D0), Phys. Rev. **D86**, 091101 (2012), arXiv:1207.2757 [hep-ex].
[11] S. Amerio, M. Casarsa, G. Cortiana, J. Donini, D. Lucchesi, and S. Pagan Griso, IEEE Trans. Nucl. Sci. **56**, 1690 (2009).
[12] D. Acosta *et al.* (CDF Collaboration), CDF-PUB-11219.
[13] D. Acosta *et al.* (CDF Collaboration), Phys. Rev. D **71**, 052003 (2005).
[14] T. Sjöstrand, S. Mrenna, and P. Skands, J. High Energy Phys. **05**, 026 (2006), we use PYTHIA version 6.216.
[15] D. Acosta *et al.* (CDF), Phys. Rev. **D71**, 092001 (2005), arXiv:hep-ex/0412006 [hep-ex].
[16] S. Amerio *et al.*, CDF Note 11195.
[17] A. L. Read, *Advanced Statistical Techniques in Particle Physics. Proceedings, Conference, Durham, UK, March 18-22, 2002*, J. Phys. **G28**, 2693 (2002), [,11(2002)].
[18] T. Junk, Nucl. Instrum. Meth. **A434**, 435 (1999), arXiv:hep-ex/9902006 [hep-ex].



(a)



(b)

FIG. 7. The observed 95% C.L. upper limits on the cross section times the branching ratio. Linear scale on top (a), log scale on bottom (b).