Measurement of $Z \to b\bar{b}$ cross section and search for a Higgs-like particle produced in association with b quarks at CDF

Luigi Marchese[1,a] on behalf of the CDF Collaboration

[1] *University of Oxford*

Abstract. We present a measurement of the $Z \to b\bar{b}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV . We use a data set of 5.46 fb^{-1} collected by the CDF experiment at the Tevatron collider during Run II using a dedicated trigger path which required a displaced vertex compatible with a b-hadron decay. A data-driven procedure is applied to estimate the dijet mass spectrum of the non-resonant multijet background. Using a similar strategy we set one of the most stringent upper limits on the production of a Higgs-like particle in association with b quarks. We also set a limit on the inclusive SM $H \to b\bar{b}$.

1 Introduction

With the Higgs boson discovery at the ATLAS and CMS [1][2] experiments, high-energy physics is in transition: after a long period of search and exploration, an era of consolidation and precise measurements has just started, complementing the direct search for new physics beyond the Standard Model (SM). At the moment, several Higgs properties are still measured with low precision. The decay $H \to b\bar{b}$ is still not observed, despite its large branching ratio, 58%, [3] due to the overwhelming $b\bar{b}$ QCD background. Compared to the LHC searches, the search for $H \to b\bar{b}$ at CDF benefits from a lower background, but it suffers of low production cross section. We used a data-driven techinque to evaluate the $b\bar{b}$ QCD background. We validate this technique by reconstructing $Z \to b\bar{b}$. The measurement of $p\bar{p} \to Z \to b\bar{b}$ cross section at CDF helps in understanding the production of electroweak bosons at center-of-mass energy of 1.96 TeV with initial conditions complementary to the LHC experiments [4].
Furthermore, we report a search for a narrow neutral scalar particle ϕ decaying into b-quark jets in events with an additional b-quark jet. Several SM extensions predict the existence of new particles decaying into b-quark jets, such as the minimal supersymmetric extension of the SM (MSSM) [5] or the two-Higgs-doublet models (2HDM) [6] and dark-matter models involving mediator particles with a large coupling to b quarks [7][8]. In this note I describe the $Z \to b\bar{b}$ cross section and the b-jet energy-scale measurement by using events with at least two b-tag jets. Using the same dataset and analysis strategy an upper limit to the inclusive $p\bar{p} \to H \to b\bar{b}$ process is set. Finally, a search for a narrow neutral scalar particle ϕ decaying into b-quark jets in events with an additional b-quark jet is described. The analysis is limited to the case where the ϕ boson is produced in association with one or more additional b-quark jets in order to suppress the large multijet background.
This note is structured as follows:

[a] e-mail: marchese@fnal.gov

- Section 2 describes the measurement in the $Z \to b\bar{b}$ channel and the search for a SM $H \to b\bar{b}$. Within this section, Sec. 2.1 describes the data and Monte Carlo (MC) simulation samples used for this analysis. We use a sample of 5.4 fb^{-1} of data collected by the CDF detector requiring at least one displaced secondary vertex in the event compatible with a b-hadron decay. The full dataset collected by the CDF experiment of 10 fb^{-1} is reduced to 5.4 fb^{-1} bacuse of the trigger path applied. The evaluation of the $b\bar{b}$ QCD background using a data-driven technique is presented is Section 2.2. The fitting procedure used to measure signal and background is discussed in Section 2.3. Finally, the measurement of the $Z \to b\bar{b}$ cross section is described in Section 2.4 and the inclusive limit on the $p\bar{p} \to H \to b\bar{b}$ is described in Section 2.5;
- Section 3 reports the results of the search for a narrow neutral scalar particle ϕ decaying into b-quark jets in a data sample of events with at least three b-quark jets;
- Section 4 reports the final conclusions.

2 $Z \to b\bar{b}$ and SM $H \to b\bar{b}$

2.1 Data and MC samples

Data used in this measurement are collected with the DIJET BTAG trigger. This trigger algorithm exploits the long b-hadron lifetime searching for tracks coming from a secondary vertex displaced from the primary one and keeps the jet energies as low as possible to reconstruct events with low dijet invariant mass allowing the exploration of invariant-mass regions not covered by the LHC experiments. This trigger path is designed and optimized for $H \to b\bar{b}$ event collection, but can be used for any final state with b jets since it is aimed at high efficiency on any final state with b jets. The trigger algorithm combines the information of the CDF II eXtremelyFastTracker (XFT), improved on line tracking, and the information of the Secondary Vertex (SecVxt) from the Silicon Vertex Trigger (SVT) to perform an efficient track-jet matching.
Selected events are required to have:

- two jets with $E_T > 22$ GeV;
- at least one jet with a secondary vertex found by using the tight SecVxt [9] b-tagging algorithm. This sample, referred as single tag, is used for the background determination;
- at least two jets with a secondary vertex identified by the tight SecVxt b-tagging algorithm. This sample, referred as double tag, is fitted to extract the signal yield.

A jet is defined as:

- b-tag trigger jet if it fires the DIJET BTAG trigger and has a tight SecVtx tag;
- b-tag jet if has tight SecVtx tag.

Monte Carlo simulation is used to evaluate the efficiencies and the acceptance of signals, $Z \to b\bar{b}$ and $H \to b\bar{b}$, and to determine the shape of the two b-jet invariant-mass distribution for signals and background. Simulated QCD $b\bar{b}, c\bar{c}$ and generic di-jet events are produced using Pythia v. 6.216 [10] with the underlying event modeled using tune "A". The CTEQ5L parton distribution functions (PDF) are used. $H \to b\bar{b}$ with mass of 125 GeV/c^2 and $Z \to b\bar{b}$ events are also generated with the same version of Pythia with tune "A". The standard CDF procedure is used for MC generation: events are generated and simulated through the detector using the same luminosity as in the periods corresponding to the data taking. The efficiency of the trigger on the $Z \to b\bar{b}$ is about 5%, while on the $H \to b\bar{b}$ process is about 10%.

Data selected as described above are expected to be dominated by b-hadrons. The TagMass, defined as the invariant mass of all particles with tracks originating from the secondary vertex assumed to be charged pions (139 MeV/c^2) is used to determine the heavy flavor content of the sample. The TagMass is sensitive to the flavor of the parton initiating the jet. Since light quarks and gluons can generate a secondary vertex tag only because of track mis-measurements, they are characterized by low invariant-mass distributions. Hadrons originating in b quarks have larger invariant mass with respect to those originating in light- and also c-quarks, so they are distinguishable on a statistical basis.

In Fig.1 we report the TagMass distribution of the b-tag trigger jets, which is fitted with a binned maximum likelihood as the sum of three contributions: b quark, c quark and light quarks. Templates of the TagMass distribution of the b-tag trigger jets in the b, c and light-quark jets assumption for each of the three components are obtained from MC samples. Individual contributions in the b-tag trigger jets are the following: the fraction of b-quark jets is (75 ± 2)%, (7 ± 1)% for c-quark jets and (18 ± 2)% for light quarks jets, where the uncertainty is the quadratic sum of the statistical and the systematic uncertainties in the MC templates. As Fig.1 shows, high values of TagMass correspond to an almost pure sample of b-quark jets.

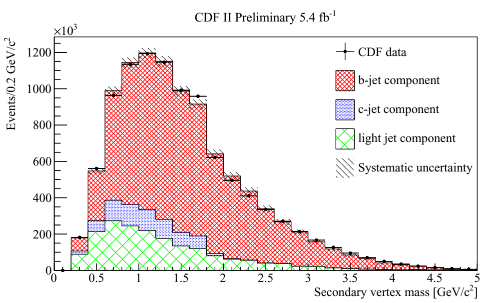

Figure 1. Invariant-mass distribution of the charged tracks of the secondary vertex for a sample of events with at least one b-tag-trigger jet.

2.2 Background estimate

The invariant-mass distributions for $b\bar{b}$ and bc and bq from multijet production is necessary to describe the components of the double-tagged sample together with the signal. The contribution of multiple non-b-tag events is negligible and it is not considered. The sample used to search for the signal largely consists of heavy-quark jets with at least two SecVtx tags. Since these events are produced by mechanisms not precisely predicted by theoretical models, a data-driven strategy is used to evaluate the background contribution, as in similar CDF measurements [11]. The method proceeds as follow:

- We first estimate the efficiency to tag b, c and light-quark-initiated jets as a b jet. This per-jet probability is evaluated as a function of the jet (E_T, η) in MC samples. In this note we refer to these probabilities as tagging matrices;
- Starting from the single b-tag jet data sample, we infer the flavor of the non-tagged jet by weighting it with the tagging matrices for b, c and light-quark hypotheses;
- We calculate the invariant mass of the b-tag trigger and the second "flavour determined" jets assuming the predicted shape of the background invariant mass in the three jet flavor hypotheses.

Following this method, the expected background is determined using data which include the searched signal but with a contribution less than 1% and in a specific invariant-mass region not biasing the background description which spans a large region. It has to be noted that this strategy holds only

if the b-tag trigger jet is a jet originating from a b quark. Therefore, we ask for a TagMass of the b-tag trigger jet greater than 1.8 GeV/c^2 thus rejecting c and light quarks initiated jets.
This strategy allows only for the shape extraction of the background contributions. Normalizations are left as free parameters of the data fit.

2.3 Fit description

The double-tagged sample is made of 925338 dijet events. Data are fitted using a signal template with background contributions obtained as described in Section 2.2. We look for an enhancement riding atop the continuum background of the invariant-mass distribution of the two leading (highest E_T) tagged jets, m_{12}. The signal template is obtained from fully simulated Monte Carlo $Z \to b\bar{b}$ events. Since the invariant-mass distribution templates in the b and c flavour hypotheses are very similar and the fit is not able to distinguish them, they are merged considering a fixed contribution. We assign a systematic uncertainty for this assumption.
We perform a binned maximum-likelihood fit, where the likelihood is defined as:

$$\mathcal{L} = \prod_{i}^{N} \frac{n_s^i P_S(m_{jj}^i) + \sum_b n_b^i P_b(m_{jj}^i)}{n_s^i + \sum_b n_b^i} \tag{1}$$

where $\mathcal{L}$ is the product, over all the bins, of the probability that the events in the i.th bin with invariant mass m_{jj} are described by the 4 background p.d.f. $P_b(m_{jj})$ and the signal p.d.f. $P_S(m_{jj})$; n_s and n_b are the free parameters which are constrained to be greater than or equal to zero.

2.4 $Z \to b\bar{b}$ results

In Fig.2 we show data with the result of the fit superimposed. The fit returns a sizable signal component with the light-quark component compatible with zero, indicating that the sample is constituted by $b\bar{b}$ jets. The goodness of the fit is estimated by calculating the χ^2/NDF, which is found to be 0.7. We evaluate the significance of the signal as obtained from the fit by computing the p-value, i.e. the probability that the observed signal is a background fluctuation. The observed $Z \to b\bar{b}$ signal has an observed significance greater than 5σ. In Fig.3 we compare the double-tagged-events invariant-mass distribution after the background subtraction to the Monte Carlo $Z \to b\bar{b}$ signal template. The selected Z sample can be used to measure the cross section for Z-boson production multiplied by the branching ratio of the decay to b-quark pairs, after we take into account efficiency and the luminosity of the sample. The calculated cross section value is then $\sigma_Z \cdot B(Z \to b\bar{b}) = 1.11 \pm 0.08$ nb, where the uncertainty is statistical only. Systematic uncertainties are related to data/MC differences, fitting procedure and background modeling. The total systematic-uncertainty is 11.4%.

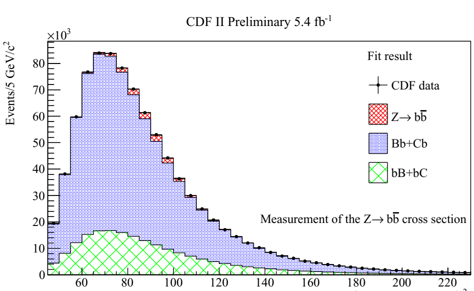

Figure 2. Double-tagged events invariant-mass distribution with the result of the fit. In red the fitted $Z \to b\bar{b}$, in blue the Bb+Cb and in green the bB+bC background. Capital letter indicates the b-tag trigger jet.

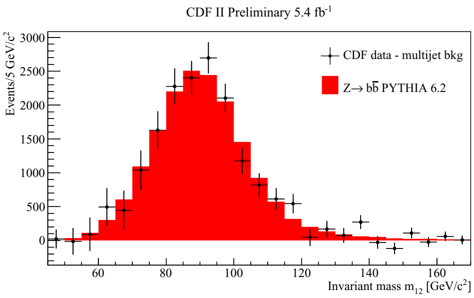

Figure 3. Double-tagged events invariant-mass distribution after the background subtraction. The Z peak is in good agreement with the MC signal template, in red.

2.5 Search for inclusive production of Standard Model $H \to b\bar{b}$

The predicted total SM-Higgs production cross section for the Tevatron is 1.23 ± 0.22 pb [3], while the branching ratio into a pair of b-quarks is (58.4 ± 3.3)% [3]. The selection efficiency can be estimated using MC samples, as described in Sec. 2.1. The signal template shape which has a resolution σ of 19 GeV/c^2 is also MC estimated. Given the integrated luminosity of 5.4 fb^{-1}, we expect about 36 signal events. The background under the signal region, defined as the 2σ region around the nominal Higgs mass, is made of about 670 k events. In Fig.4 we report the result of the fit to the double-tagged sample with the Higgs component scaled by a factor 10^3 with respect to the SM expectations. The selection of the events and the fitting strategy are the same as used for the Z channel with the $H \to b\bar{b}$ template added. Normalizations of all the components are the free unconstrained parameters of the fit. The fit returns 0 ± 91 Higgs events.

We set a 95% confidence level (C.L.) limit on the inclusive production of the SM Higgs using a modified frequentist CL_S method [12]. Pseudo-experiments are generated in the background only hypothesis. The CL_S for the expected limit is then evaluated using as test statistic the difference in χ^2 of the fits to the pseudo-experiments using only the background templates and the fits using both background and signal templates. Applying the same procedure to data instead of the pseudo-experiments, it is possible to estimate the observed limit. Systematic uncertainties regarding both the normalization and the shape are introduced in the limit calculator as nuisance parameters. In Fig.5 we report the expected and the observed CL_S limits obtained as a function of the ratio between the cross-section upper limit and the Standard Model cross section. The observed(expected) upper limit at 95% C.L. on the $p\bar{p} \to H \to b\bar{b}$ is found to be 33(46) times the SM cross section.

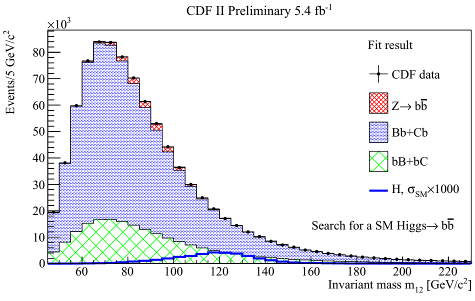

Figure 4. Invariant-mass distribution of the double-tagged data sample with the result of the fit that includes the $H \to b\bar{b}$ decay. The normalization of the Higgs signal is set to x10^3 the expected SM cross section for illustrative purposes.

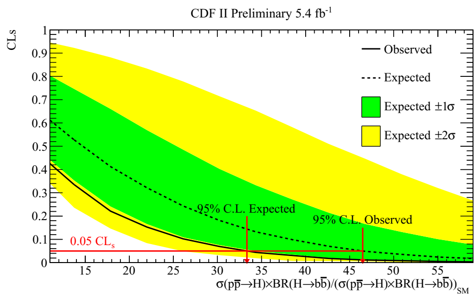

Figure 5. Observed (black solid line) and expected (black dashed line) CL_S as function of the cross section times the branching ratio normalized to SM $H \to b\bar{b}$. Green and yellow regions are respectively the 1σ and 2σ bands on the expected CL_S. The 95% C.L. limits are indicated by the red vertical arrows under the red horizontal line.

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

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 [13] and D0 [14] experiments at the Tevatron collider and by the CMS [15] 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 [16]. The Tevatron initial state and center-of-mass energy provides a complementary search environment to the LHC experiments, especially in the low b-quark-dijet invariant-mass region, where the 2σ deviation has been observed.
This analysis is based on the same dataset and analysis strategy used in the Z channel described in the previous section. Signal events are obtained by fitting the double-tagged sample using background templates built starting from the single tagged sample. Results are model-independent, i.e. no particular theoretical model is tested and the upper limit is set on the production cross section $\sigma(p\bar{p} \to \phi b)$ x $B(\phi \to b\bar{b})$, because of the various theoretical frameworks able to interpret them.
MC samples are generated for a variety of ϕ masses using Pythia 6.216 corresponding 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 $\bar{b}$. As for the Z channel, MC 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 flavour. Signal-event-selection efficiencies are in the range [3.7- 8.7]‰, depending on the invariant mass of the neutral scalar.
The multijet background templates used to describe the triple-jet-tagged data are built starting from the double -agged sample used in the Z analysis. The inclusion of a third tag, whose efficiency depends on the flavor of the jet, can be simulated using a parametrization of the SecVtx response. The flavor composition of the triple-jet-tagged data is determined by fitting data. The flavour of the third jet is inferred using the same tagging matrices described in the previous section. Events in the double-btag sample, with an additional third untagged jet, are organized in two categories, bbx and xbb where x can be a c-, b or light-quark-initiated jet. The order is as a function of the E_T. 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). Fig.6 shows the invariant mass of the two highest momentum b-quark jets, m_{12}, for the various components of the heavy-flavour multijet background. An additional variable, x_{tag}, is studied in order to have a better separation between backgrounds with high and low TagMass values. The variable is defined as reported in [11].

A sample of 5616 events with three tagged jets are fitted using a binned maximum-likelihood method. We search for a neutral scalar Higgs-like resonance, ϕ, in the mass range of 100 - 300 GeV/c^2 by fitting the m_{12} and the x_{tags} distributions using the procedure described in the previous

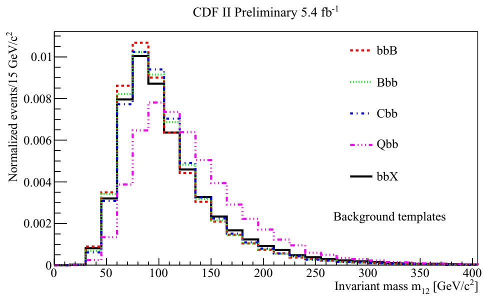

Figure 6. Invariant-mass distribution of the two leading jets m_{12} in the background templates.

section. In Fig.7 we show the result of the fit performed when we include a signal template with m_ϕ of 160 GeV/c^2.

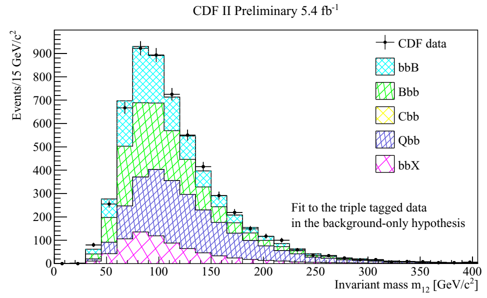

Figure 7. Result of the fit to the triple tagged data projected into m_{12} when a signal component with a mass of ϕ = 160 GeVc^2 is added to the background templates.

As done for the SM Higgs search, a 95% confidence-level upper limits on the production cross section times the branching ratio of a narrow scalar as a function of mass is set using a modified frequentist CL_S method. The procedure is the same as described in Sec.2.5 but with the appropriate pseudo-experiments. Using as a test statistic the difference in χ^2 of these fits under the different hypotheses, background only and background plus signal, we derive an expected limit. Fitting data with the same procedure we derive an observed limit. In Fig.8 we report the observed limits and the median expected 95% C.L. limits as a function of the mass of the scalar particle. We also show the $\pm 1\sigma$ and $\pm 2\sigma$ bands of the expected limits. All points of the observed limit are within the 2σ band of the expected limit, indicating no excess of significant signal events.

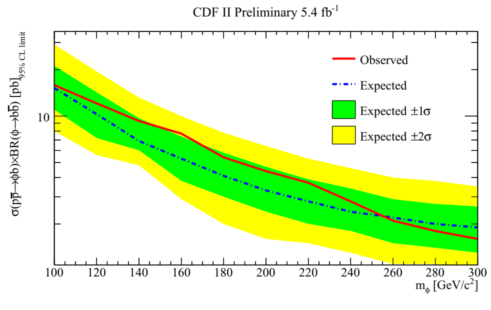

Figure 8. The observed 95% C.L. upper limits on the cross section times the branching ratio. The plot is shown in log scale. See Fig. 5 for a complete description of the plot format.

4 Conclusions

Observation of the $Z \rightarrow b\bar{b}$ decay at CDF has been reported and the production cross section times the branching ratio measured with a final value of 1.11 ± 0.08(stat) ± 0.13(sys) nb. Results are consistent with NLO theoretical calculations combined with the measured $Z \rightarrow b\bar{b}$ branching ratio, which predicts σ_Z x $B(Z \rightarrow b\bar{b})$ = 1.13 ± 0.02 nb. Using the same dataset and a similar strategy, a search for SM $H \rightarrow b\bar{b}$ is performed. The observed upper limit at 95% C.L. on the $p\bar{p} \rightarrow H \rightarrow b\bar{b}$ with an invariant mass of 125 GeV/c^2 is 40.6 pb which corresponds to 33 times the SM expectations. This is the first inclusive limit on the $p\bar{p} \rightarrow H \rightarrow b\bar{b}$. In addition, 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 performed. These results complement the previous combined limit of CDF and D0 using the full statistics collected by the CDF experiment in Run II and suggest that the 2σ excess in the 100-160 GeV mass range in the $b\bar{b}$ dijet-mass distribution from events with an additional b-quark jet is a statistical fluctuation [16]. A limit on the cross section times branching ratio in the 100 - 300 GeV/c^2 mass range has been set.

References

[1] ATLAS Collaboration, , Phys. Lett. **B 716**, 01 (2012).
[2] CMS Collaboration, Phys. Lett. **B 716**, 30 (2012).
[3] C. Patrignani et al. (Particle Data Group), Chin. Phys. **C 40**, 100001 (2016).
[4] ATLAS Collaboration, Phys. Lett. **B 738**, 25 (2014), arXiv:1404.7042 [hep-ex].
[5] H. P. Nilles, Phys. Rept. **110**, 1 (1984).
[6] 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].
[7] E. Izaguirre, G. Krnjaic, and B. Shuve, Phys. Rev. **D 90**, 055002 (2014), arXiv:1404.2018 [hep-ph].
[8] A. Berlin, D. Hooper, and S. D. McDermott, Phys. Rev. **D 89**, 115022 (2014), arXiv:1404.0022 [hep-ph].
[9] D. Acosta et al. (CDF Collaboration), Phys. Rev. **D 71**, 052003 (2005).
[10] T. Sjostrand, S. Mrenna, and P. Skands, J. High Energy Phys. **05**, 026 (2006), we use *pythia* version 6.216.
[11] T. Aaltonen et al. (CDF), Phys. Rev. **D 85**, 032005 (2012), arXiv:1106.4782 [hep-ex].
[12] A. L. Read, Advanced Statistical Techniques in Particle Physics. Proceedings, Conference, Durham, UK, March 18-22, 2002, J. Phys. **G 28**, 2693 (2002).
[13] T. Aaltonen et al. (CDF), Phys. Rev. **D 85**, 032005 (2012), arXiv:1106.4782 [hep-ex].
[14] V. M. Abazov et al. (D0), Phys. Lett. **B 698**, 97 (2011), arXiv:1011.1931 [hep-ex].
[15] V. Khachatryan et al. (CMS), JHEP **11**, 071 (2015), arXiv:1506.08329 [hep-ex].
[16] T. Aaltonen et al. (CDF and D0 Collaborations), Phys. Rev. **D 86**, 091101 (2012), arXiv:1207.2757 [hep-ex].