August 20, 2007

Search for the Standard Model Higgs boson in the $\not\!\!E_T$ and b-jet Signature in $p\bar{p}$ collisions at $\sqrt{s} = 1.96 TeV$

The CDF Collaboration

Abstract

We search for the Higgs boson produced in association with a Z or a W boson. We consider a scenario where the Z decays into neutrinos or the lepton originating from the W-decay escapes detection and the Higgs decays into a $b\bar{b}$ pair. Therefore the expected signature is missing transverse energy, no leptons, and two b-jets. We process 1.7 fb^{-1} of data collected in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96 TeV in Run II of the Tevatron. No significant excess above backgrounds is observed, so we set limits on associated Higgs boson production cross-section for Higgs masses in range from 110 to 150 GeV/c^2.

Preliminary results for Summer 2007 Conferences

1 Introduction

In the Higgs mechanism of the Standard Model, the fermions and weak gauge bosons acquire mass via interaction with the Higgs field. The Higgs field is described by a complex doublet. Three of the four real fields of the doublet couple to the SU(2) gauge bosons. The observable quantum of the fourth Higgs field is called the Higgs boson. The existence of this undiscovered particle is the cornerstone of the Standard Model[1].

Direct searches performed with the LEP experiments have constrained the Higgs mass to be larger than 114.4 GeV at 95% C.L [2]. As the Tevatron is now taking data, it might be possible to set higher limits on the Higgs mass, or even to find evidence for its existence. We have developed an analysis searching for the associate production of the Higgs with a Z/W boson in a mass range above the LEP limit.

At the Tevatron, the most probable production mode of the Higgs is by gluon fusion through a virtual top loop. Around 70% of the Higgs would decay into two b-quarks yielding two b-jets in the final state. Since the QCD b-quark production is an irreducible background, this analysis would have a low sensitivity. The second most frequent production mode is when a virtual W or Z decays into a W or Z and a Higgs. In this case, it is possible to trigger on the decay products of the W/Z boson and significantly reduce the QCD background.

We are analyzing Z-Higgs and W-Higgs associated productions when the Z decays into two neutrinos, or the W decays leptonically but the charged lepton escapes the detection. Because the neutrinos will not be detected in the calorimeter, either, they lead to an unbalanced transverse energy sum in the transverse plain ($\not\!\!E_T$). We do not expect isolated leptons in these events. To remain mutually exclusive with the dedicated WH analysis, a set of loose lepton selection requirements are used to veto events with isolated leptons. Lepton veto also helps to reduce the W/Z + jets background. We are performing a unbiased analysis in which all the considerable Standard Model background processes, such as QCD heavy flavor multi-jet production, top production, and electro-weak processes, are simulated. After obtaining a good agreement between data and background prediction in control regions, we perform an optimization of the signal selection cuts on the Monte Carlo simulation. This way, the final selection in the signal region remains completely unbiased.

2 Data Sample and Event Selection

We use data collected through March 2007, which corresponds to 1.7 fb^{-1} integrated luminosity. The events are collected by CDF II detector with a trigger that selects events with $\not{E}_T > 25$ GeV at Level 1 at least two Level 2 clusters with $E_T > 10$ GeV and $\not{E}_T > 35$ GeV at Level 3 .

In the first step of the analysis, both the Monte Carlo and real data events are to pass a set of quality cuts to ensure that the possible beam and detector effects are removed from the data sample making it compatible with the simulation. The standard CDF jet clustering algorithm is used [3] with a jet cone of radius 0.4. Jet energies are corrected for calorimeter non-uniformity, non-linearity and energy loss in the un-instrumented regions of calorimeter and energy coming from different $p\overline{p}$ interactions during the same bunch crossing. The $\not{E}_T$ of the event is then corrected with new jet energies with the following formula:

$$\not{E}_x{}^{Corr} = \not{E}_x{}^{Raw} - \sum_i (E_x^{Corr,i} - E_x^{Raw,i}) \tag{1}$$

$$\not{E}_y{}^{Corr} = \not{E}_y{}^{Raw} - \sum_i (E_y^{Corr,i} - E_y^{Raw,i}) \tag{2}$$

The azimuth of $\not{E}_T$ in the transverse plane is calculated as:

$$\varphi_{\not{E}_T} = \arctan\left(\frac{\not{E}_y{}^{Corr}}{\not{E}_x{}^{Corr}}\right) \tag{3}$$

The trigger efficiency is obtained from data and is used to scale the signal and Monte-Carlo backgrounds to correct for event loss during data taking. The efficiency of the two-jet requirement is 100% if the offline transverse energy of the most energetic jet is above 35 GeV, the second most energetic jet is above 25 GeV, and at least one of the jets has $|\eta| < 1.0$. The overall efficiency of the online event selection is then parametrized by the offline corrected $\not{E}_T$ and applied on the Monte Carlo samples providing a proper scaling for the simulated events.

The final requirement imposed on the data and simulation before comparing them is the b-tag requirement. We use two categories of SECVTX b-tagging algorithm [4], tight and loose. The main difference between the loose and tight tagging algorithms is that the loose tagger has more efficient track selection. The b-tagging efficiency for tight (loose) tagger is $\sim$ 40% (50%) and mistag rate is $\sim$ 2% (4%).

In this analysis, we first look for 2 jets tagged by loose algorithm. If this fails we look for at least one jet tagged by tight algorithm.

Below is the summary of the basic selection cuts used in analysis:

- Exactly two jets with $E_T^{Corr} > 20$ GeV, $|\eta| < 2.4$, and 0.1<EM Fract.< 0.9
- 1^{st} jet $E_T^{Corr} > 35$ GeV, and $|\eta| < 2.0$

- 2^{nd} jet $E_T^{Corr} > 25$ GeV, and $|\eta| < 2.0$
- 2 loose b-tags or at least 1 tight b-tag
- no isolated leptons (electron or muon)

3 Background Estimation

In this analysis, we simulate processes which yield real taggable objects, that is, when a b- or a c-quark pair is created. Events with light flavor jets, after requiring a positive tag in the event during the basic selections, are considered to be mistags and are estimated from the data. Consequently, a tagged jet in the simulation is considered to have a real tag only if there is a heavy flavor quark inside the R=0.4 radius cone of the jet to avoid double-counting the mistags. All Monte-Carlo samples were generated with Pythia.

The QCD sample is scaled by a factor such that the total number of simulated and mistag events becomes equal to the number of data events passing the basic selection cuts (single and double-tag QCD events are scaled separately). This scale factor is measured in control regions.

Two classes of top-production are considered in this analysis. The pair-production and the single top-production in the t- and s-channels. They both yield a significant contribution to the background in the signal region. In addition, the presence of the single top background becomes more pronounced after requiring only two hard jets in the event. Due to the large mass and the semi-leptonic decay of the top, these events are energetic, bear large $\not\!\!E_T$ and high jet multiplicity. The top quark pair production is normalized to luminosity with CDF latest combined cross-section computed at top mass 171 GeV [5] ($\sigma_{t\bar{t}} = 7.8 \pm 0.9$ pb). The single top samples are normalized to luminosity with theoretical values for cross-section [6] ($\sigma_{W^\star \to tb} = 1.12 \pm 0.06$ pb, $\sigma_{gW \to tb} = 2.34 \pm 0.12$ pb).

In the di-boson samples, the bosons' decays are inclusive. In the W/Z + jets samples, the bosons are forced to decay into leptons, or b-quarks. The electroweak backgrounds thus include the following processes: W to leptons + h.f., Z to leptons + h.f., WW/WZ/ZZ inclusive decays.

The mistagged light flavor multi-jet processes are estimated from the data [4]. The method is based on the assumption that positive mistags are due to resolution effects through the same mechanism as the negative tags; therefore, the positive mistag rate is proportional to the negative one. The number of negative tags are estimated from the jets in the data, the tag-rate of which is parametrized by their E_T, $|\eta|$ and SecVTX track multiplicity, in an event with a given number of Z vertexes, primary vertex Z position and $\sum E_T$ (scalar summed E_T of all jets in the event).

After defining two control regions in the events passing the basic selection criteria mentioned above, the simulated SM background is compared to the data. In the first control region all the leptons are vetoed from the events, and the azimuthal angular separation between the second leading jet and the $\not\!\!E_T$, $\varphi(2^{nd}\, jet, \not\!\!E_T)$, is less than 0.4. This control region is dominated by QCD multi-jet events, and is used to corroborate the shapes and normalization of the b/c-filtered Pythia Monte Carlo sample with respect to the data in the control region. The second control region contains at least one lepton or isolated track and $\varphi(2^{nd}\, jet, \not\!\!E_T) > 0.4$. This region is sensitive to Electroweak processes, and is used to check the overall shapes and normalizations of the Monte Carlo. The double-tag events in this control region are dominated by the top processes, yielding an extra handle in the cross-check.

Process	1 Tight tag	2 Loose Tags
QCD h.f.	$24337.1 \pm 111.4 \pm 5445.4$	$3768.5 \pm 45.8 \pm 688.2$
Top	$7.1 \pm 0.4 \pm 0.8$	$2.3 \pm 0.2 \pm 0.4$
Di-boson	$1.1 \pm 0.2 \pm 0.2$	$0.1 \pm 0.1 \pm 0.1$
W + h.f.	$26.2 \pm 2.7 \pm 11.1$	$1.1 \pm 0.5 \pm 0.4$
Z + h.f.	$8.7 \pm 1.2 \pm 3.6$	$0.9 \pm 0.4 \pm 0.4$
Mistag	$6181.0 \pm 63.6 \pm 498.5$	$415.2 \pm 10.0 \pm 71.1$
Total bckg	30561.2 ± 5469.7	4188.1 ± 693.4
Observed	29431	4190

Table 1: Number of expected background events in the QCD control region

Process	1 Tight tag	2 Loose Tags
QCD h.f.	$50.7 \pm 5.1 \pm 12.6$	$7.0 \pm 2.0 \pm 1.9$
Top	$134.8 \pm 1.6 \pm 16.4$	$55.9 \pm 1.0 \pm 9.0$
Di-boson	$14.7 \pm 0.8 \pm 2.5$	$1.9 \pm 0.2 \pm 0.4$
W + h.f.	$80.5 \pm 4.1 \pm 34.9$	$8.2 \pm 1.3 \pm 3.6$
Z + h.f.	$17.5 \pm 1.8 \pm 7.9$	$1.3 \pm 0.5 \pm 0.7$
Mistag	$86.5 \pm 4.3 \pm 6.3$	$3.7 \pm 1.2 \pm 0.8$
Total bckg	384.7 ± 42.6	77.9 ± 10.3
Observed	373	79

Table 2: Number of expected background events in the EWK control region

After achieving a good agreement between the simulation and the data in the control regions (Tables 1, 2, Figures 4, 5, 6, 7), a set of cuts are selected optimizing the signal Monte Carlo against the background prediction. The optimization yields the following selection requirements:

- $\varphi(1^{st} jet, \not\!\!E_T) > 0.8$
- $\frac{\not\!\!H_T}{H_T} > 0.45$, where H_T is the scalar sum and $\not\!\!H_T$ is the magnitude of the vectorial sum of the p_T's of the two leading jets.

- $1^{st} jet\ E_T > 60$ GeV
- $\not{E}_T > 70$ GeV

Process	Single Tag	Double Tags
QCD h.f.	$157.4 \pm 9.0 \pm 49.1$	$10.6 \pm 2.4 \pm 3.9$
Top	$48.2 \pm 1.0 \pm 4.1$	$14.0 \pm 0.5 \pm 2.1$
Di-boson	$11.5 \pm 0.6 \pm 2.4$	$1.9 \pm 0.2 \pm 0.4$
W + h.f.	$59.9 \pm 4.1 \pm 26.6$	$4.6 \pm 1.1 \pm 2.0$
Z + h.f.	$28.3 \pm 1.9 \pm 12.5$	$4.1 \pm 0.7 \pm 1.9$
Mistag	$98.2 \pm 7.3 \pm 12.8$	$4.7 \pm 1.0 \pm 1.1$
Total bcgk	403.5 ± 60.1	39.9 ± 6.1
Observed	443	51
($VH\ m_H = 115\,\mathrm{GeV/c^2}$)	(1.9)	(1.2)

Table 3: Number of expected background and signal events in the signal region after applying the final cuts.

4 Cross-Section Limits

The observed excess in double-loose tagged sample (Table 3) has been extensively studied. No systematic source for disagreement has been found. The probability of having such an excess as a result of background only fluctuation was estimated to be ~3%.

The systematic uncertainties are classified as correlated and uncorrelated errors considering the relations between the signal and the background processes. The correlated errors are taken into account separately for each processes in the limit calculation. The uncorrelated systematic uncertainties are: statistical error in negative tag estimate, negative-positive tag rate asymmetry factor, QCD multi-jet Monte Carlo normalization (14% in single tagged, 6.3% in double tagged sample), MC statistical fluctuations. The correlated systematics are: luminosity (6.0%), b-tagging efficiency scale factor between data and Monte Carlo (4.3% for single and 10.2% for double tags), trigger efficiency (3%), lepton veto efficiency (2%), PDF uncertainty (2%) and Jet Energy Scale. ISR/FSR systematic uncertainties (between 1% and 5%) are applied on the signal.

Considering the systematic uncertainties listed above, we computed the expected limit for the Higgs cross-section when the Higgs is produced with a Z/W boson and decays to two b-quarks where Z decays to neutrinos and W to leptons. We use Bayesian method for deriving the limits[8]. Table 4 shows the final result. All the cross-sections are ratios with respect to the Standard Model cross-section.

Higgs mass	VH limit, 1 Tight Tag		VH limit, 2 Loose Tags		VH limit, Combined	
(GeV)	Predicted	Observed	Predicted	Observed	Predicted	Observed
110	$19.7^{+9.7}_{-6.0}$	36.6	$10.4^{+4.4}_{-2.9}$	18.7	$9.3^{+4.4}_{-2.9}$	18.5
115	$22.7^{+9.5}_{-7.2}$	37.2	$11.1^{+4.4}_{-3.3}$	20.8	$9.7^{+5.0}_{-2.8}$	19.7
120	$27.5^{+11.4}_{-7.7}$	40.8	$13.0^{+6.5}_{-3.9}$	25.2	$11.5^{+5.5}_{-3.7}$	22.6
125	$31.2^{+14.8}_{-9.3}$	46.6	$15.9^{+6.6}_{-4.8}$	30.1	$13.4^{+6.1}_{-4.1}$	26.6
130	$40.6^{+16.7}_{-12.6}$	58.7	$19.5^{+10.6}_{-5.5}$	39.3	$16.6^{+7.3}_{-5.3}$	33.4
135	$52.0^{+22.4}_{-16.7}$	74.6	$24.7^{+10.7}_{-7.5}$	48.3	$21.0^{+9.7}_{-6.3}$	43.0
140	$71.6^{+31.5}_{-23.7}$	110.0	$35.3^{+17.5}_{-10.9}$	64.3	$31.5^{+16.4}_{-7.2}$	61.5
150	$172.3^{+71.7}_{-61.4}$	238.6	$77.1^{+37.1}_{-22.4}$	133.6	$72.1^{+30.9}_{-23.4}$	127.0

Table 4: The predicted and observed cross-section limits of the ZH/WH processes combined when $H \to b\bar{b}$ divided by the SM cross-section

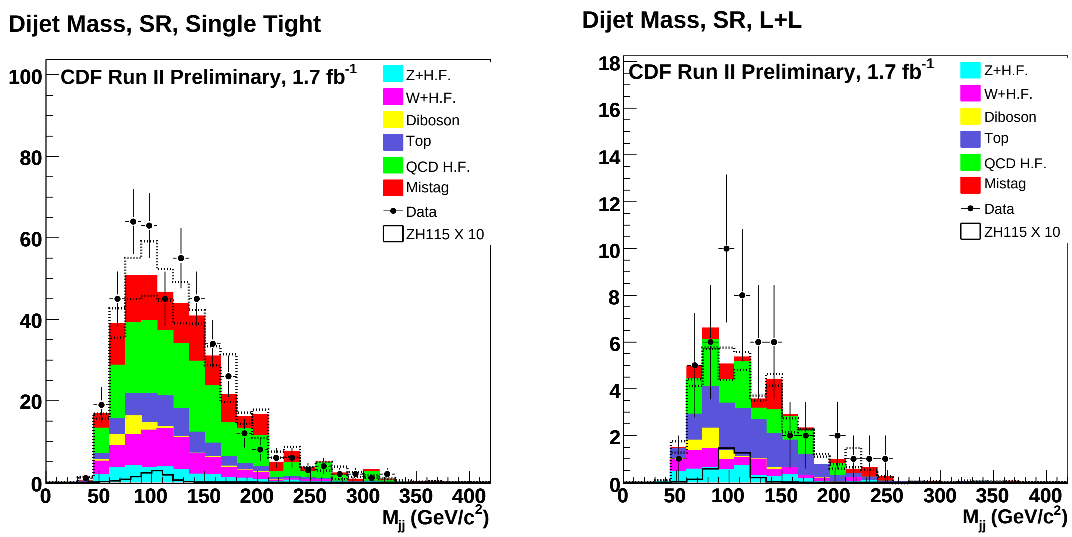

Figure 1: Dijet invariant mass in the Signal Region, single- and double-tagged events

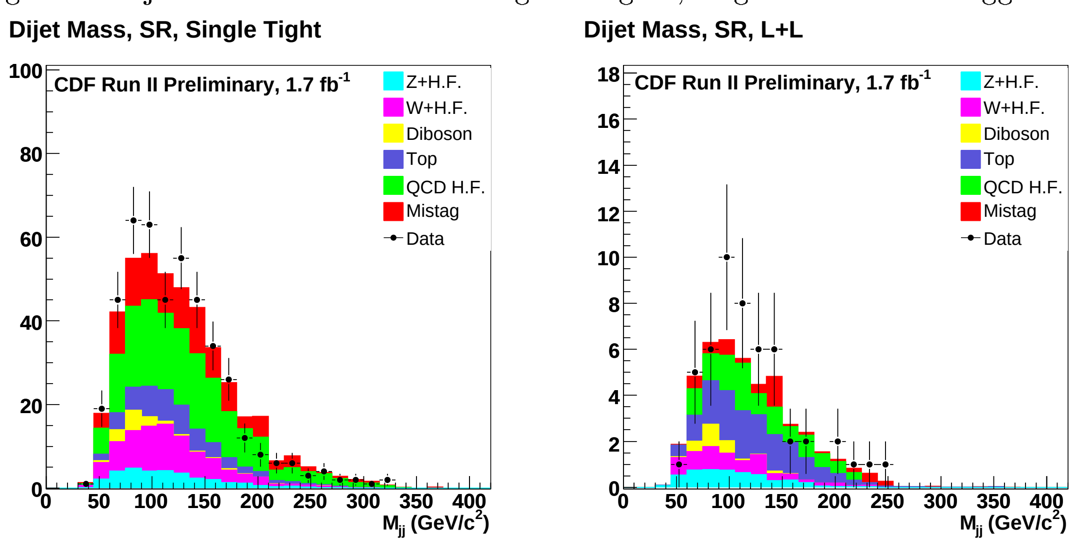

Figure 2: Dijet invariant mass in the Signal Region, single- and double-tagged events. Predicted backgrounds are varied witin systematic uncertainties to produce the best chi-squared fit to data.

We have performed a direct search for the Standard Model Higgs boson decaying into b-jet pairs in 1.7 fb^{-1} data accumulated in Run II of the CDF detector. We do not observe any significant excess over the background predicted by the Standard Model, thus we set a 95 % C.L. upper limit for the Higgs boson at varous masses. The observed cross-section limits agree with the expected ones within 2σ.

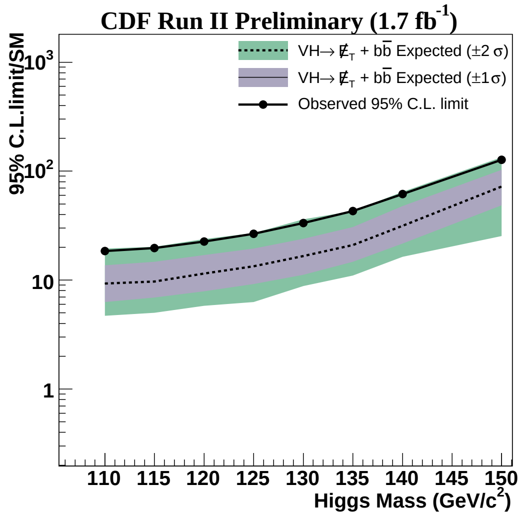

Figure 3: The 95 % C.L.Cross-Section upper limits

References

[1] J. Gunion et al., The Higgs Hunter's Guide (Addison-Wesley; New York, 1990).

[2] A. Heister et al., (The LEP Higgs Working Group), Phys. Lett. B 565, 61 (2003).

[3] F. Abe et al., Phys. Rev. D 45, 1448 (1992).

[4] D. Acosta et al., Phys. Rev. D. 71, 052003 (2005)

[5] CDF collaboration, Combination of top quark pair production cross section measurements, CDF/PHYS/TOP/PUBLIC/8148

[6] N. Kidonakis, "Single top production at the Tevatron: Threshold resummation and finite-order soft gluon corrections," Phys. Rev. D **74** (2006) 114012 [arXiv:hep-ph/0609287].

[7] D. Acosta et al., Phys. Rev. Lett. 93, 142001 (2004)

[8] T. Junk, Sensitivity, Exclusion and Discovery with Small Signals, Large Backgrounds, and Large Systematic Uncertainties, CDF/DOC/STATISTICS/PUBLIC/8128

5 Control Regions Plots

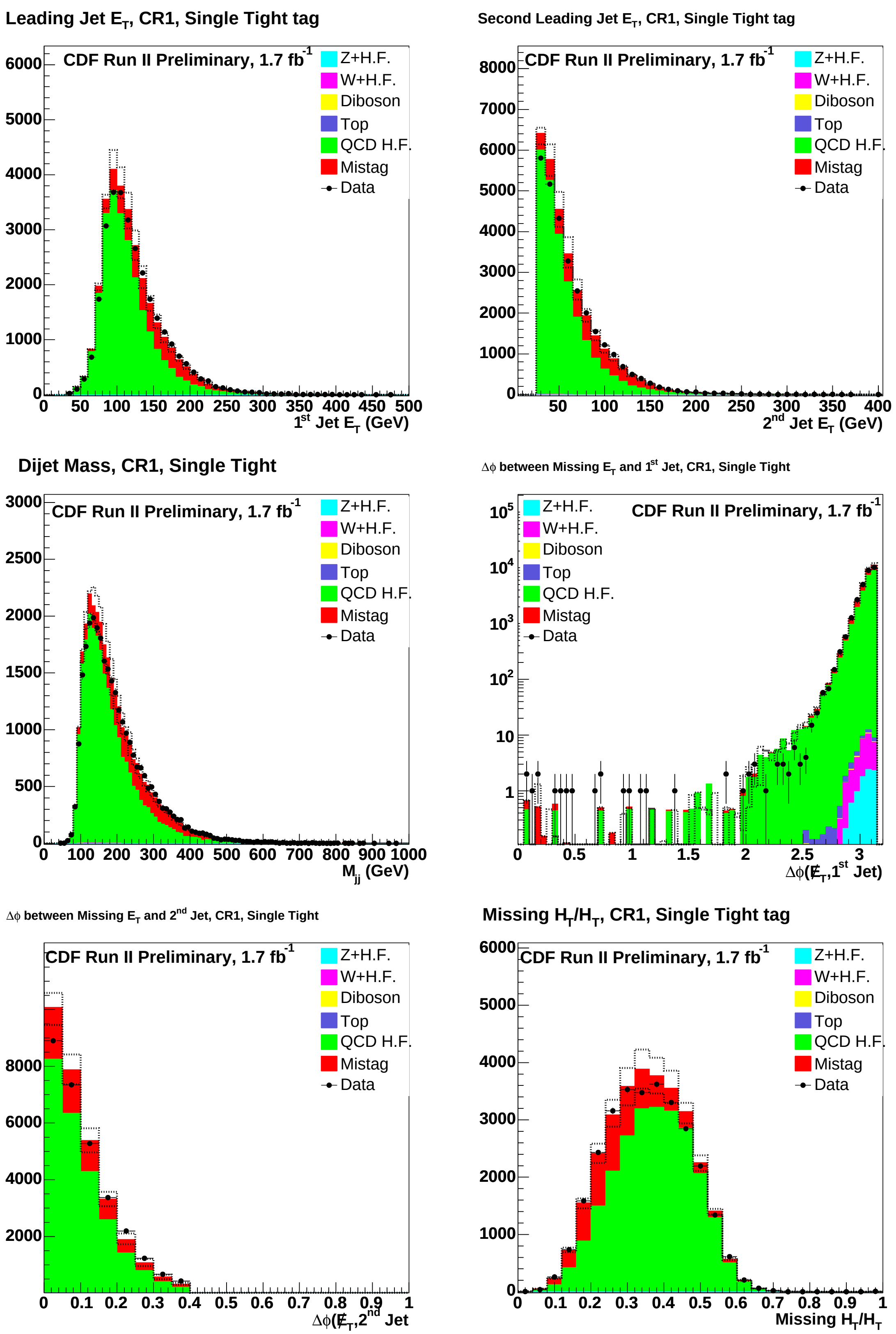

Figure 4: Control Region 1, Single Tag plots, from the top: 1^{st} jet E_T, 2^{nd} jet E_T, dijet mass, $\Delta\phi(\not{E}_T, 1^{st}jet)$, $\Delta\phi(\not{E}_T, 2^{nd}jet)$, $missing\ H_T/H_T$

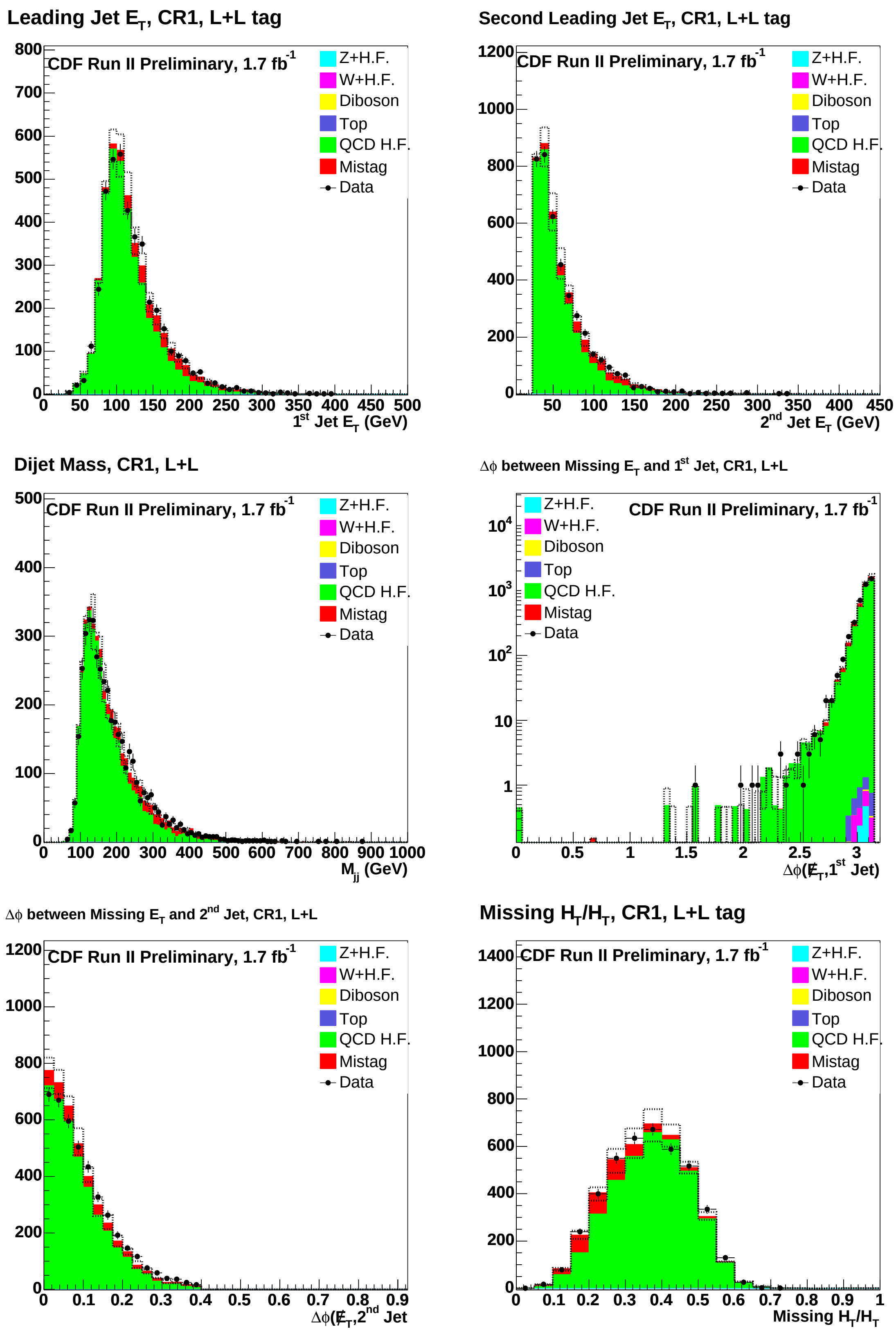

Figure 5: Control Region 1, Double Tag plots, from the top: 1^{st} jet E_T, 2^{nd} jet E_T, dijet mass, $\Delta\phi(\not{E}_T, 1^{st}jet)$, $\Delta\phi(\not{E}_T, 2^{nd}jet)$, $missing\ H_T/H_T$

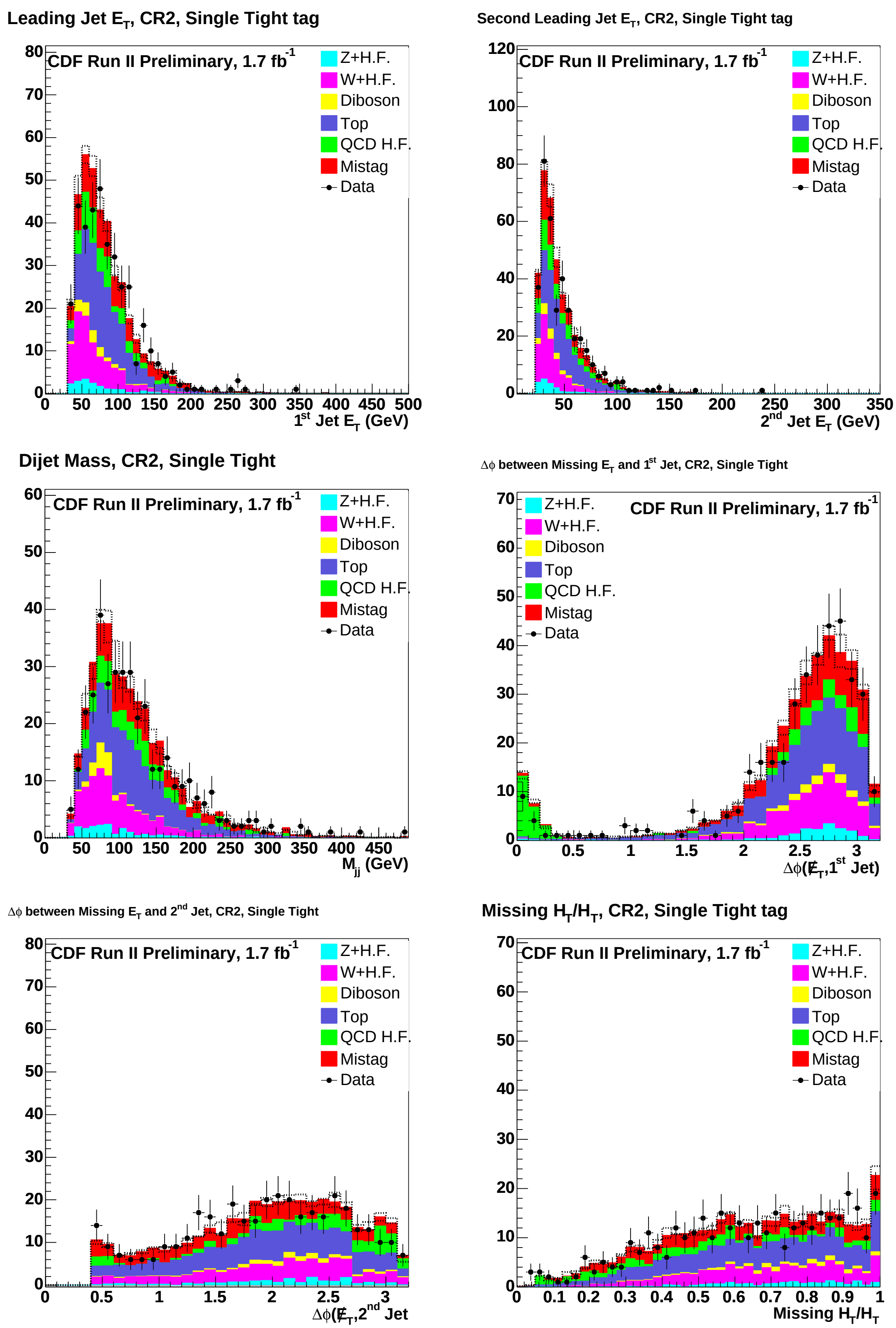

Figure 6: Control Region 2, Single Tag plots, from the top: 1^{st} jet E_T, 2^{nd} jet E_T, dijet mass, $\Delta\phi(\not{E}_T, 1^{st} jet)$, $\Delta\phi(\not{E}_T, 2^{nd} jet)$, $missing\ H_T/H_T$

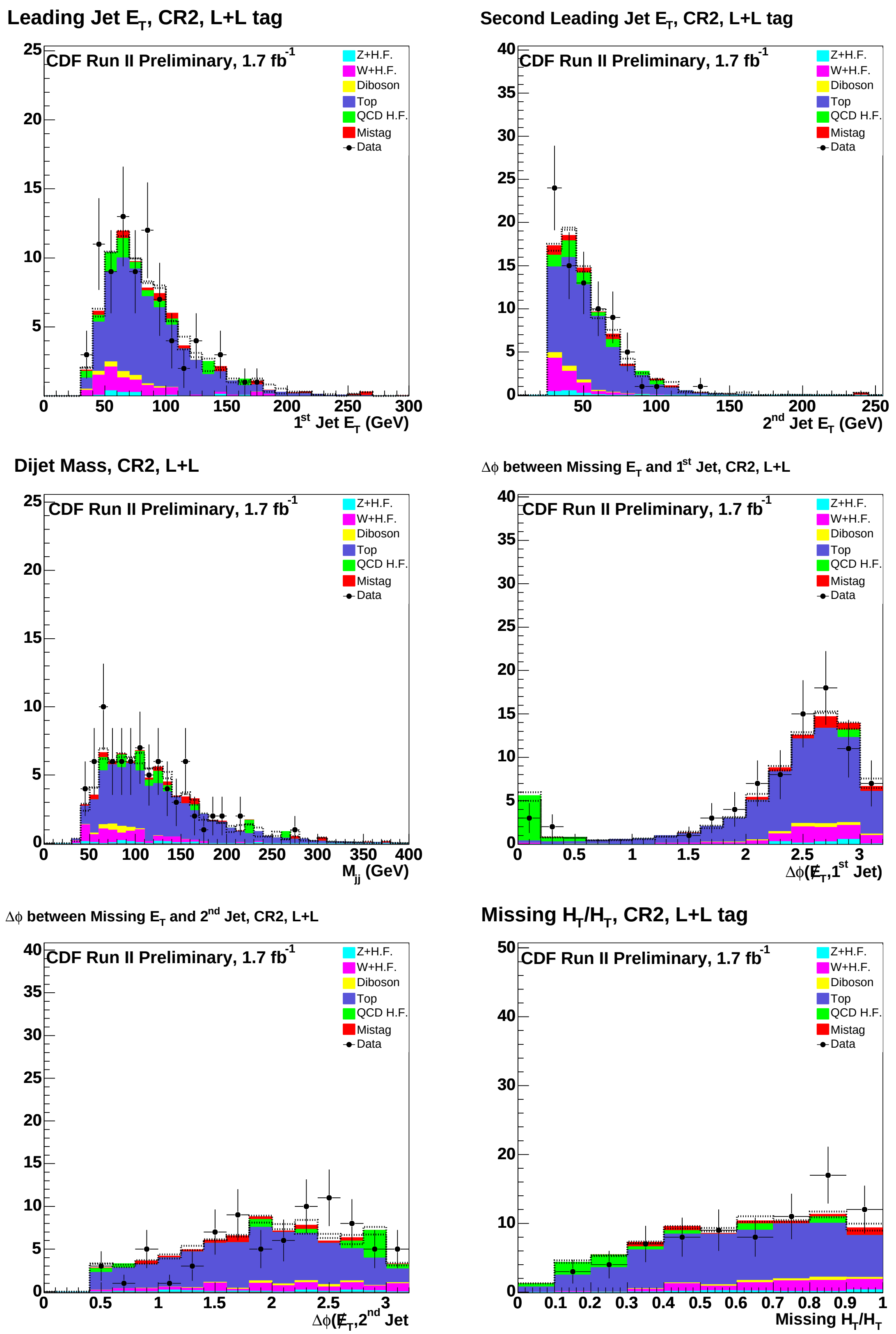

Figure 7: Control Region 2, Double Tag plots, from the top: 1^{st} jet E_T, 2^{nd} jet E_T, dijet mass, $\Delta\phi(\not{E}_T, 1^{st} jet)$, $\Delta\phi(\not{E}_T, 2^{nd} jet)$, $missing\ H_T/H_T$