CDF/ANAL/TOP/CDFR/8698
January 17, 2008
Version 1.0

Measurement of the b Jet Production Cross Section in Events with a $W^{\pm}$ Boson

Christopher Neu, Evelyn Thomson, Joel Heinrich

University of Pennsylvania

Abstract

The b jet production cross section is measured for events with a $W^{\pm}$ boson in 1.9/fb of CDF Run II data. $W^{\pm} + b$ jet events are selected by requiring a high p_T central electron or muon, large $\not\!\!E_T$ and one or two high E_T central jets, among which jets are further required to be b-tagged. The increased purity of Ultratight SECVTX b-tagging is exploited in an effort to better understand the impact of b and non-b sources in the tagged sample. Assuming Standard Model cross sections for top and diboson processes, the b jet production cross section for events with a leptonically decaying $W^{\pm}$ boson is measured to be 2.74 $\pm$ 0.27 (stat) $\pm$ 0.42(syst) pb. This measured cross section provides a data-driven measurement of a difficult-to-predict background for single top and Higgs searches.

1 Introduction

$W^{\pm}$ production in association with b jets is an important process in Run II at CDF. $W^{\pm} + b$ jets production is an obstacle to understanding two prominent Run II targets: the search for single top production and the search for a low-mass Higgs in the WH channel. The final state in each of these processes has a $W^{\pm}$ boson with one or two b jets; hence $W^{\pm} + b$ jets poses a challenge since it is an irreducible background of the two more interesting processes.

The importance of understanding $W^{\pm} + b$ jets is amplified by the large predicted rate of such events. From the theory side, tree-level predictions are available for $W^{\pm}$ + up to 2 b jet production from $\mathcal{O}(\alpha_s^2)$ to $\mathcal{O}(\alpha_s^5)$ [1]. An example of one of these LO diagrams is shown in Figure 1. These predictions, which include effects from the nonzero b mass, indicate that the inclusive event cross section for $W^{\pm} + b$ jets production, with $W \to \ell\nu$, is in the range 2-3pb. Calculations at NLO have demonstrated a factor of $\sim$2 enhancement over LO results [2], [3]. These predictions can be compared to the SM predictions for single top production ($\sim$1pb for

s- and t-channel inclusive [4] with a leptonically decaying $W^{\pm}$) and WH (∼0.01-0.03pb for the mass range 100 GeV/c^2 < M_H < 140 GeV/c^2 [5] with a leptonically decaying $W^{\pm}$). Hence $W^{\pm}$ + b jets is a significant source of background events for these two searches.

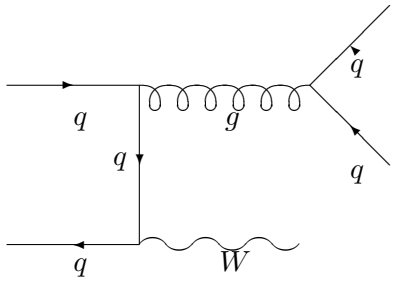

Figure 1: An example of tree level production of $W^{\pm}$ + b-jets.

Theory predictions for the rate of $W^{\pm}$ + b jet production are widely distrusted. At CDF, a technique was developed in Run 1 [6] to derive the $W^{\pm}$ + LF and $W^{\pm}$ + HF backgrounds in the $W^{\pm}$ + jets sample in the context of the top search. This technique is widely known as "Method 2". The recipe to arrive at a prediction for $W^{\pm}$ + b jets events in Method 2 relies on measuring the ratio of the rate of Monte Carlo (MC) $W^{\pm}$ + b jets to the rate of inclusive MC $W^{\pm}$ + jets events. It is assumed that while the absolute rates of these events is poorly modeled in MC, the relative rate of heavy flavor to inclusive flavor is usable. The MC-based prediction for the HF contribution is then scaled by a data-to-MC fudge factor, k, which is measured in a control sample (either inclusive jets [7] or, more recently in the Run 2 single top search [8] in the $W^{\pm}$ + 1 jet bin). The HF fractions from the $W^{\pm}$ + jets MCs underestimates the HF content in the data; the k factor scales the HF prediction up by 40% . This treatment succeeds in providing an estimate for $W^{\pm}$ + b jets events; however from Table 1 one can see that the technique has very limited precision, only being able to predict the $W^{\pm}$ + b jets contribution to within 30-40% . Most of this loss of precision is due to imperfect knowledge of the HF k factor. While this level of accuracy was tolerable for extracting a relatively large signal like $t\bar{t}$ production, small statistics searches such as single top and Higgs would benefit from a more precise knowledge of the rate of these events.

The goal then is to precisely measure the $W^{\pm}$ + b jets cross section in the data. With such a result one could provide a new measurement of the k factor with improved uncertainty. More desirable and more ambitious would be to use the direct measurement of $W^{\pm}$ + b jets production as a feedback loop into the improvement of MC tools, to improve the rate predictions and verify the kinematic distributions for such events.

Process	1 jet	2 jets	3 jets	4 jets	≥5 jets
Pretag Data	196160.0 ± 0.0	32242.0 ± 0.0	5496.0 ± 0.0	1222.0 ± 0.0	272.0 ± 0.0
WW	12.7 ± 1.4	31.7 ± 3.5	10.1 ± 1.1	2.5 ± 0.3	0.7 ± 0.1
WZ	7.0 ± 0.6	14.8 ± 1.2	4.0 ± 0.3	0.9 ± 0.1	0.2 ± 0.0
ZZ	0.2 ± 0.0	0.6 ± 0.0	0.3 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
ttbar	13.5 ± 1.9	108.9 ± 15.2	242.6 ± 33.7	252.0 ± 34.9	85.3 ± 11.8
s-chan	8.2 ± 0.8	30.4 ± 3.0	9.4 ± 0.9	2.0 ± 0.2	0.4 ± 0.0
t-chan	21.7 ± 1.9	31.4 ± 2.8	9.4 ± 0.8	1.9 ± 0.2	0.3 ± 0.0
Z+jets	35.9 ± 4.6	26.8 ± 3.3	9.8 ± 1.2	2.4 ± 0.3	0.5 ± 0.1
W+b-jets	387.8 ± 149.2	241.1 ± 94.6	65.9 ± 26.5	16.5 ± 6.7	3.5 ± 1.5
W+c-jets	692.6 ± 271.3	231.7 ± 92.1	54.0 ± 22.0	12.9 ± 5.2	2.6 ± 1.1
W+LF jets	693.6 ± 95.7	235.5 ± 40.0	66.4 ± 14.2	16.7 ± 3.6	3.3 ± 1.1
Fake$W$	345.0 ± 138.0	249.6 ± 99.9	74.3 ± 29.7	13.6 ± 11.7	4.6 ± 4.6
Total Prediction	2218.2 ± 451.2	1202.6 ± 214.3	546.3 ± 67.7	321.6 ± 39.1	101.4 ± 13.0
Observed	2872.0 ± 0.0	1416.0 ± 0.0	586.0 ± 0.0	371.0 ± 0.0	113.0 ± 0.0

Table 1: Method 2 table for events passing standard l+jets selection with ≥ 1 Tight SECVTX tag in 1.9/fb of CDF data assuming ttbar cross section of 6.7+-0.7pb.

1.1 Previous Result

The $W + b\bar{b}$ cross section was measured in the CDF 700/pb data sample [9]. The b jet cross section from $W + b\bar{b}$ production was measured to be

$$\sigma_{b-\text{jets}}(W + b-\text{jets}) \times BR(W \to \ell\nu) = 0.90 \pm 0.20(stat) \pm 0.30(syst)\text{pb}$$

which can be compared to the theory cross section provided by ALPGEN v2.1 [10]:

$$\sigma_{b-\text{jets}}(W + b-\text{jets}) \times BR(W \to \ell\nu) = 0.74 \pm 0.18(syst)\text{pb}$$

The analysis strategy used in this previous result is similar in spirit to the one described herein.

The 700/pb result was systematics limited. The largest source of systematic error came from potential inadequacies of the b jet model which is used to extract the b content of the selected data sample. Additional significant amount of uncertainty was incurred from ambiguities that arise in the handling of one specific type of background, that from b jets in fake W events.

In an attempt to improve the measured $W^{\pm} + b$ jet cross section, in the current analysis an effort was made to reduce these sources of error significantly. The strategy employed here was two-fold:

- use high purity Ultratight SECVTX in order to better understand b jet behavior in the data
- use a highly effective veto of fake W events to minimize the impact of mismodeling those fake W events that remain

The effect of these two analysis improvements are described below.

2 Event Samples: Data and Monte Carlo

The events considered in this analysis are required to satisfy the high p_T central lepton triggers. Details of the L1/L2/L3 criteria can be found elsewhere (for example [11]). The trigger path names used in the analysis:

- ELECTRON_CENTRAL_18
- MUON_CMUP18
- MUON_CMX18

For the analysis we only consider CEM, CMUP and CMX trigger leptons. The events are organized into two analysis datasets, *bhel*XX and *bhmu*XX. This data was assembled using Production v6.1.1 and accessed via high p_T lepton skims provided by the Top group.

The analysis uses good run list v18, which accommodates data through Period 12 [**?**]. We consider only runs in which the subsystems responsible for detecting electrons and muons were functional [1]. We additionally require that the runs we consider have a fully operational silicon detector. Considering these requirements, this constitutes approximately 1.9/fb of data.

The MC samples used in this analysis are those that are necessary to describe the physics of the $W^{\pm}$ + jets sample. ALPGEN v2.10′ [10] is used to $Z/W^{\pm}$ + b jets, $Z/W^{\pm}$ + c jets, and $Z/W^{\pm}$ + light flavor jets samples. The parton showering in these ALPGEN samples is accomplished via Pythia v6.325 [**?**]. Pythia v6.216 is also used in constructing samples of $t\bar{t}$ and diboson events. The complete list of MC samples including production and ntuple DSIDs is available elsewhere [13]. These MC samples include run-dependent effects through the 1.2/fb run range, including the effect of multiple interactions as a function of instantaneous luminosity.

Calibrations of the tagged b jet sample in data relies on jets in the 8 GeV muon sample, from the trigger paths MUON_CMUP8 and MUON_CMUP8_DPS (production dataset *bmcl*XX). A corresponding HF-enriched Pythia dijet MC sample is used for consistency checks (MC DSID

[1] For runs less than 150145 the CMX system status is ignored since the CMX status information is unreliable for that era of CDF.

btopla). Inclusive jet data on trigger paths JET_20 and JET_50 (datasets *gjt1*XX and *gjt2*XX) were used to probe the behavior of tagged light flavor jets in the data.

TopNtuples were used exclusively throughout the analysis due to the more complete SECVTX tagging information available in these files over other alternatives.

3 Event Selection

Among the events firing the central high p_T lepton triggers, we seek to identify events that have a leptonically decaying $W^{\pm}$ boson and 1 or 2 jets, among which at least one is from b quark production. To this end we apply the following event selection criteria.

Beyond the lepton trigger requirements, events are further required to have a $p_T > 20$ GeV/c electron or muon. Attention is restricted to CEM electrons and CMUP and CMX muons only. The standard "tight" electron and muon identification criteria from the Joint Physics recommendations are used universally [14]. The selected tight lepton is required to be isolated; the ratio, I, of the E_T in the R=0.4 cone around the candidate to the candidate track p_T is required to be small, $I < 0.1$. This requirement preferentially selects leptons from $W^{\pm}$ decay instead of those from semileptonic hadron decay or sources of fake leptons. Electrons are required to not be consistent with photon conversion. Muons are subject to a cut on the track χ^2 to avoid background from decays-in-flight. Finally, muons consistent with cosmic rays traversing the CDF detector are vetoed.

Events are required to have large missing transverse energy, $\not\!\!E_T$. Raw $\not\!\!E_T$ is corrected to take into account the primary interaction point of the event, the presence of high p_T muons, and for corrected jet energies. The cut used in this analysis ($\not\!\!E_T > 25$ GeV) is larger than the standard one used in other analyses examining the same final state in the Top and Higgs groups. The higher $\not\!\!E_T$ cut is motivated by the desire to reduce fake $W^{\pm}$ background as much as possible(see below for more). In a measurement of $W^{\pm}$ + b-jet production, signal statistics are not a problem so such a cut is reasonable. The high p_T lepton and $\not\!\!E_T$ requirements constitute the $W^{\pm}$ selection.

In this analysis we seek only events that have a single tight lepton; events with additional tight leptons (or additional objects satisfying the "loose" or tight-but-non-isolated lepton definition) are rejected to protect against dilepton $t\bar{t}$ background. Events consistent with $Z \to \ell\ell$ production are vetoed by rejecting events that have a tight lepton and an isolated track that have a reconstructed mass near the Z.

Events are finally subject to a veto that targets fake $W^{\pm}$ events. These events arise mostly form QCD multijet production in which a jet fakes a lepton signature and the event contains spurious missing transverse energy. A strong fake $W^{\pm}$ veto was developed in the context of the single top search [**?**] and was employed here as well.

Jets are identified using the JetClu algorithm and have a cone size of 0.4. The towers from the leptons qualifying for the dilepton $t\bar{t}$ veto above are removed before clustering of the jets. The jets considered in this analysis are corrected to L5 with v08 of the CDF jet corrections [15] and are required to have E_T>20 GeV and $\mid \eta \mid$<2.0. Events are required to have 1 or 2 such reconstructed jets; most $W^{\pm}$ + b jet signal has 1 or 2 jets, and considering higher jet multiplicities would introduce $t\bar{t}$ background at a significant rate.

Jets are further required to possess an Ultratight SECVTX b tag. Ultratight SECVTX provides a sample of increased b purity over the more widely used Tight and Loose SECVTX operating points. With Ultratight SECVTX the rate of tagging LF jets is reduced by x10 and the rate of tagging charm jets is reduced by x4 at the expense of a reduction by 50% of the tags of real b jets [16]. Ultratight SECVTX was designed specifically for this $W^{\pm}$ + b jet analysis where one can afford such a loss in signal efficiency for the sake of significantly reducing the tags from non-$b$jets and better understanding of the tagged sample.

Within this selected sample one can measure the production cross section for $W^{\pm}$ + b jet production. Herein the pretag selected sample is referred to with the shorthand notation "W12j".

4 The $W^{\pm}$ + b-jet Cross Section: Definition

The b-jet cross section for $W^{\pm}$ + b jets production is defined as

$$\sigma_{b-\mathrm{jets}}(W + b-\mathrm{jets}) \times BR(W \to \ell\nu) = \frac{n^{b\ \mathrm{jets}}_{W+Nb}}{\mathcal{L} \cdot \mathcal{A}^{b\ \mathrm{jets}}_{W+Nb} \cdot \epsilon^{b}{}_{\mathrm{tag}}} \tag{1}$$

We seek here the production cross section of b $j$ets from $W^{\pm}$ + b jets as opposed to the cross section of $e$vents from $W^{\pm}$ + b jets. This choice is motivated by the desire to insulate the result, wherever possible, from the possibly flawed details of how our $W^{\pm}$ + b jets model (here ALPGEN) handles the physics outside our acceptance. In order to measure an event cross section, one would measure the rate of events that satisfy our event selection, and then extrapolate those results to an inclusive event cross section. This is not useful for experimentalists, since our CDF result would be influenced by uncertainty in our $W^{\pm}$ + b jets model. This is not useful for theorists since the result would be intimately tied to one specific model of these events. Hence we avoid this problem by measuring the b jet production cross section in events with a $W^{\pm}$ boson.

We further insulate the result from model dependency by restricting the phase space of events we consider. MC events are required to possess:

- a HEPG electron or muon with $p_T > 20$ GeV/c, $\mid \eta \mid < 1.1$

- a HEPG neutrino with $p_T > 25$ GeV/c
- 1 or 2 total MC-level jets with $E_T > 20$ GeV/c^2, $\mid \eta \mid < 2.0$

These restrictions are chosen in such a way as to match the corresponding analysis level requirements placed on reconstructed objects. The construction of jets in the MC will occur in Section 8.

With these clear definitions one can calculate the hadron level b jet cross section for events with a leptonically decaying $W^{\pm}$ boson. For the hadron level cross section calculation, the b jets are chosen via a $\Delta R < 0.4$ match to an outgoing b parton from the matrix element calculation. One can translate from event production cross section to the b jet cross section via

$$\sigma_{b-\text{jets}} \times BR = \frac{\sigma_{\text{event}} \times BR}{N_{\text{evt}}} \cdot n^{\text{1or2}}_{\text{bhadjets}} \tag{2}$$

where $\sigma_{\text{evt}} \times BR$ is the event cross section times leptonic $W^{\pm}$ branching ratio in ALPGEN for a given sample, N_{evt} is the number of generated events, and $n^{\text{1or2}}_{\text{bhadjets}}$ is the number of b-matched MC-level jets in the events having 1 or 2 total qualifying MC-level jets.

Sample	DSID	$\sigma_{\text{evt}} \times BR$ (pb)	N_{evt}	$n_{b-\text{jets}}^{\text{1or2}}$	$\sigma_{b-\text{jets}} \times BR$ (pb)	w
Wevbb0p	btop0w	2.98	1542539	2.915e+05	0.5631	0.722
Wevbb1p	btop1w	0.89	1545970	2.76e+05	0.1589	0.204
Wevbb2p	btop2w	0.29	1498550	1.196e+05	0.02314	0.030
Wevcc0p	ctop0w	5.00	2005399	49	0.0001222	0.000
Wevcc1p	ctop1w	1.79	1968365	68	6.184e-05	0.000
Wevcc2p	ctop2w	0.628	1885915	55	1.831e-05	0.000
Wevc0p	stopw0	17.1	1943317	44	0.0003872	0.000
Wevc1p	stopw1	3.39	1896728	72	0.0001287	0.000
Wevc2p	stopw2	0.507	1837070	60	1.656e-05	0.000
Wevc3p	stopw3	0.083	1745440	28	1.331e-06	0.000
Wev0p	ptopw0	1800	4868357	65	0.02403	0.031
Wev1p	ptopw1	225	4563248	168	0.008284	0.011
Wev2p	ptop2w	35.3	872814	43	0.001739	0.002
Wev3p	ptop3w	5.59	831222	33	0.0002219	0.000
Wev4p	ptop4w	1.03	775589	8	1.062e-05	0.000
Total					0.780	
Wmvbb0p	btop5w	2.98	1524880	2.897e+05	0.5661	0.721
Wmvbb1p	btop6w	0.89	1508029	2.716e+05	0.1603	0.204
Wmvbb2p	btop7w	0.29	1506613	1.209e+05	0.02328	0.030
Wmvcc0p	ctop5w	5.00	1982424	49	0.0001236	0.000
Wmvcc1p	ctop6w	1.79	1961120	77	7.028e-05	0.000
Wmvcc2p	ctop7w	0.628	1949189	72	2.32e-05	0.000
Wmvc0p	stopw5	17.1	1975397	56	0.0004848	0.001
Wmvc1p	stopw6	3.39	1911713	78	0.0001383	0.000
Wmvc2p	stopw7	0.507	1840847	73	2.011e-05	0.000
Wmvc3p	stopw8	0.507	1754673	36	1.04e-05	0.000
Wmv0p	ptopw5	1800	4955756	72	0.02615	0.033
Wmv1p	ptopw6	225	4648605	135	0.006534	0.008
Wmv2p	ptop7w	35.3	872511	46	0.001861	0.002
Wmv3p	ptop8w	5.59	839645	26	0.0001731	0.000
Wmv4p	ptop9w	5.59	774744	12	7.989e-05	0.000
Total					0.785	

Table 2: Predicted b-jet cross section from ALPGEN. The events considered are required to pass the MC level restrictions on the charged lepton, neutrino and MC-level jets. ALPGEN samples are separated in electronic and muonic $W^{\pm}$ decays; the predicted ALPGEN cross section is similar for both types of samples.

A summary of the ALPGEN samples used in this calculation and the result are contained in Table 2. ALPGEN provides $W^{\pm}$ + b jet events from diagrams with N = 0, 1, 2, etc., additional outgoing light flavor partons. In order to construct a sample that is representative of $W^{\pm}$+b-jet production, one needs to include contributions from higher order diagrams. At CDF we construct a $W^{\pm}$ + b jet sample by generating samples with up to 2 additional outgoing light flavor partons. In the generation of these samples, filtering was put in place to increase the efficiency for generating useful events. Generated events were simulated if and only if

- they contain at least one b quark with $p_T > 8$ GeV/c and $\mid \eta \mid < 3.0$, OR
- if $\Delta R(b, \overline{b}) < 0.4$ and $p_T < 8$ GeV/c for both $b$and $\overline{b}$, then $p_T > 8$ GeV/c and $\mid \eta \mid < 3.0$ for the $b\overline{b}$ system.

This choice of filtering means that $W^{\pm}$ events with parton-level low p_T b jets are not included in the $W^{\pm}$ + b jet samples generated at CDF. Despite the MC-level phase space restrictions, this amounts to some small amount of additional contributions that can be recovered through the remaining $W^{\pm}$ + jets MCs. All contributors are listed in Table 2. The contribution from non-$W^{\pm}$+b jet samples is $\sim$5% .

Because jets can be produced both from the hard scatter process and the parton shower, a matching scheme is in place in the ALPGEN samples used in this analysis that guards against double counting events into specific jet multiplicities. Specifically, the N=0,1 parton samples use exclusive matching (where events are produced if and only if each outgoing parton from the hard scatter is matched to a corresponding MC-level jet) and the N=2 parton sample uses inclusive matching (where events are produced if and only if the number of MC-level jets meets or exceeds the number of outgoing partons). In this way the true representation of the $W^{\pm}$+b-jets process is the sum of the events in the N=0,1,2 parton samples. To get the ALPGEN prediction for the b jet cross section in $W^{\pm}$+b production one must sum the values from each of the three samples. One can see from Table 2 that this prediction is $\sim$0.78 pb, and this is consistent for both the $W \rightarrow e\nu$ and $W \rightarrow \mu\nu$ decay samples.

5 Measuring the $W^{\pm}$ + b-jet Cross Section

In Equation 1, the number of b jets from $W^{\pm}$ + b jets production is not trivially extracted from the tagged sample in the data, the total yield of which we refer to as $n_{\rm UT}$. There are three sources contributing to the b-tagged sample: bottom, charm and light flavor (LF) jets. In this analysis the fraction of actual b-jets in the tagged data, $f^{b\ \rm jets}$, is determined by fitting the distribution of vertex mass, M_{vert}, of tagged jets to templates for b, c and LF species. The contribution from the background b sources ($n^{b\ \rm jets}_{\rm bkgd}$) is estimated from MC or determined in the data, and then subtracted:

$$\sigma_{b-\text{jets}}(W+b\text{–jets}) \times BR(W \to \ell\nu) = \frac{n_{\text{UT}} \cdot f^{b\ \text{jets}} - n^{b\ \text{jets}}_{\text{bkgd}}}{\mathcal{L} \cdot \mathcal{A}^{b\ \text{jets}}_{W+Nb} \cdot \epsilon^{b}_{\ \text{tag}}} \tag{3}$$

The denominator is comprised of the luminosity $\mathcal{L}$, the acceptance $\mathcal{A}^{b\ \text{jets}}_{W+Nb}$ and tag efficiency $\epsilon^{b}_{\ \text{tag}}$. The acceptance and tag efficiency are measured in the $W^{\pm} + b$ jets MC samples.

With these pieces in place one can calculate the jet-level cross section in the data, folding in the systematic errors appropriately. Vertex mass fits, background subtraction and acceptance are discussed in the following sections.

6 Determining the Composition of the b-Tagged Sample

Tagged jets contain b, c, and light flavor. The invariant mass of the tracks that comprise the found secondary vertex can be used to discriminate between the different species within the tagged sample. The mass of the tracks comprising the secondary vertex is correlated to the mass of the parent particle decaying at that point; thus the vertex mass for b jets is larger than that for c and LF since the mass of B hadron is larger than that of charm and LF states. From Figure 2 one can see the discriminating power of the vertex mass.

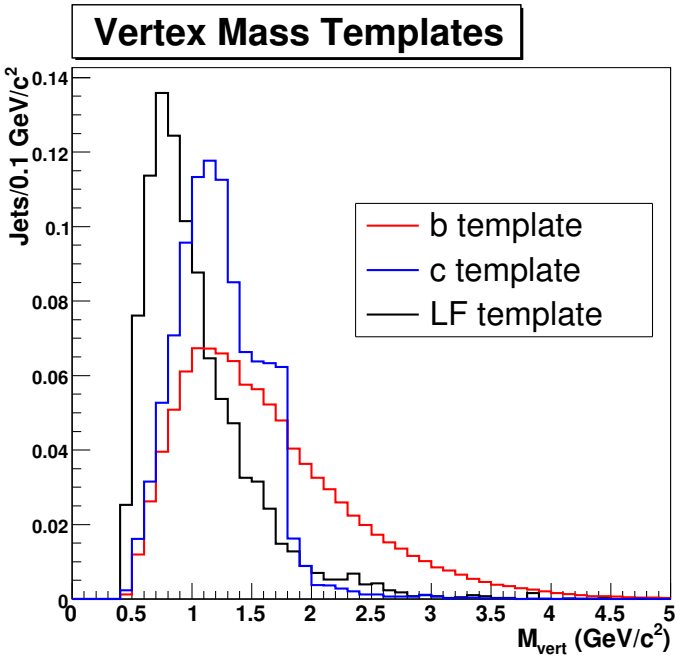

Figure 2: Vertex mass templates for Ultratight SECVTX tagged b, c, and LF jets.

Templates of the vertex mass for b and c jets were made in the MC by ΔR<0.4 matching tagged jets to B and C hadrons in the CDF OBSP bank. Cases in which the decay $B \rightarrow C + X$ occur are assigned to the b species. The HF templates come from a mixture of the main sources of these events in the $W^{\pm}$ + jets sample. The contribution of each is weighted in according to their predicted contribution to the $W^{\pm}$ + 1,2 tagged jet sample. The weight for each sample w_{template} is calculated via

$$w_{\text{template}} = \sigma_{\text{generated}} \cdot \frac{N_{\text{pass}}}{N_{\text{generated}}} \tag{4}$$

where $\sigma_{\text{generated}}$ is the generated cross section and $\frac{N_{\text{pass}}}{N_{\text{generated}}}$ is the ratio of total number of events satisfying the complete event selection with one or more tagged jets to the total number of generated events.

For b jets the main contributors to the tagged W12j sample are $W^{\pm}$+ b-jets, $t\bar{t}$, and single top. The main contributors to c tags in the W12j sample are $W^{\pm}$+ c-jets, from $W + c\bar{c}$ and $W + c$ processes. The weighted constructed b and c templates are shown in Figure 3.

The feature in the c shape near M_{vert}= 1.8 GeV/c^2 is attributable to D^0 (mass = 1.865 GeV/c^2) and $D^{\pm}$ (mass = 1.869 GeV/c^2) vertices in which the invariant mass of the constituent tracks very nearly reproduces that of the parent. This feature is so prominent because D^0 and $D^{\pm}$ are the most common c hadron state, either through direct production or through photon emission from an excited c state. The feature can be seen also in the b shape; these are instances in which the decay $B \rightarrow D^0/D^{\pm}X$ occurs, and the found secondary vertex truly contains tracks from the tertiary charm state decay.

An attempt was made to construct the light flavor model in the same fashion as that of b and c. LF jets are defined as those tagged jets that are not matched with ΔR<0.4 to any B or C hadron. The main contributor of tags of LF jets to the W12j sample comes from $W^{\pm}$ + LF jets for which MC samples exist. However the statistics for Ultratight tagged jets in events satisfying the W12j selection are very low; this necessitated using the high statistics inclusive jet MCs (one with p_T >18 GeV/c, one with p_T >40 GeV/c) for the LF shape. A comparison of the various LF template options is shown in Figure 4. A data-driven LF shape was also constructed from negative tagged jets in the inclusive jet data. This option was abandoned though because of the non-trivial nature of extracting the true HF content of the negatively tagged sample; the problem of determining a pure LF sample is avoided in the MC where one knows the species within each jet.

The fitted contributions from each species proceeds via a binned maximum likelihood fit.

(a) (b)

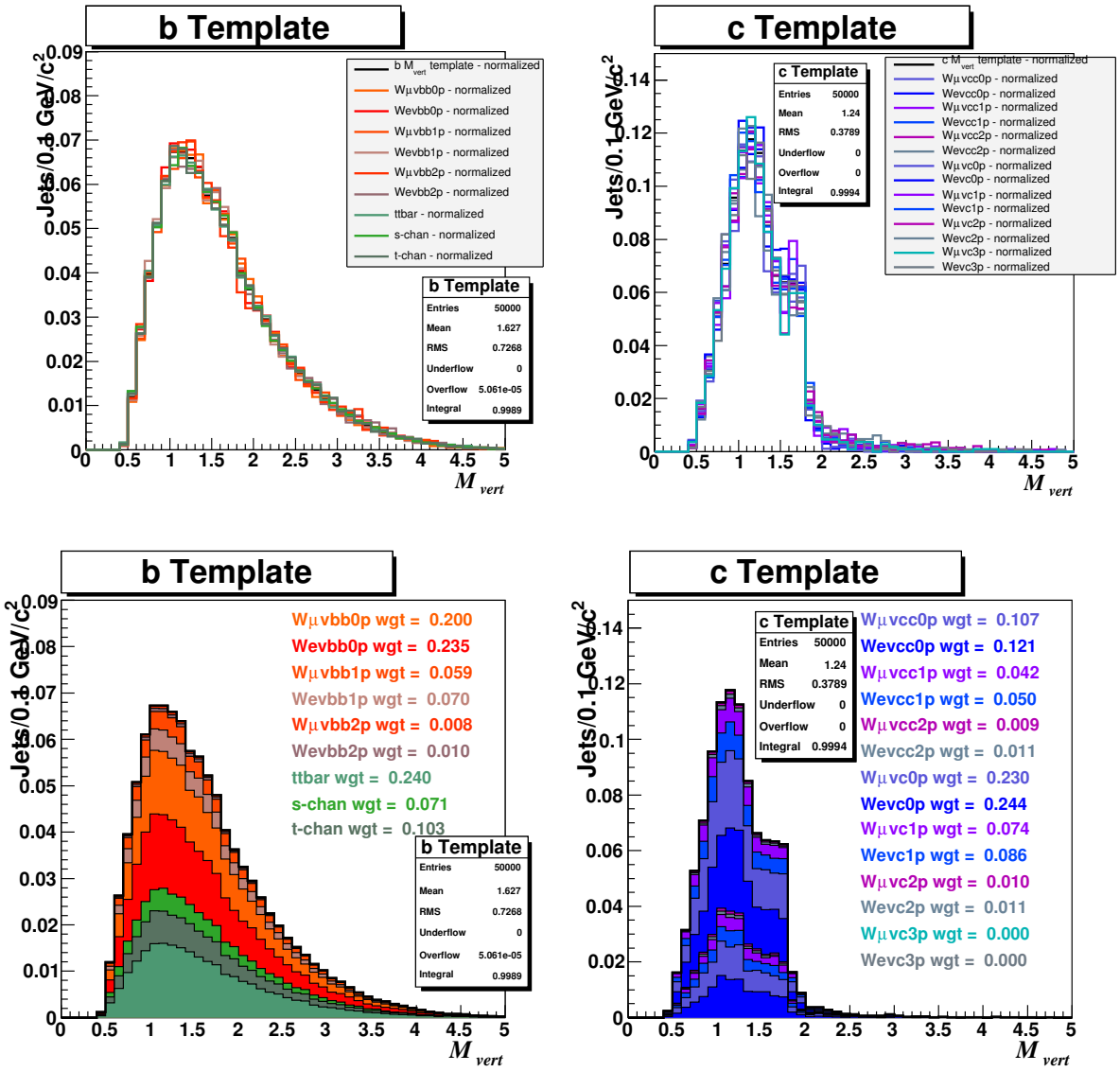

Figure 3: Contributions to the b (a) and c (b) vertex mass templates. Top row are the normalized distributions from each source overlayed; bottom row are the stacked contribution, weighted according to their contribution to the selected tagged sample from MC studies.

The Poisson probability $P(n_i|\mu_i)$ of observing n_i tagged jets in bin i of the vertex mass distribution given μ_i, the expected number of total tagged jets in bin i, is given by

$$P(n_i|\mu_i) = \frac{e^{-\mu_i}\mu_i^{n_i}}{n_i!} \tag{5}$$

μ_i is given by

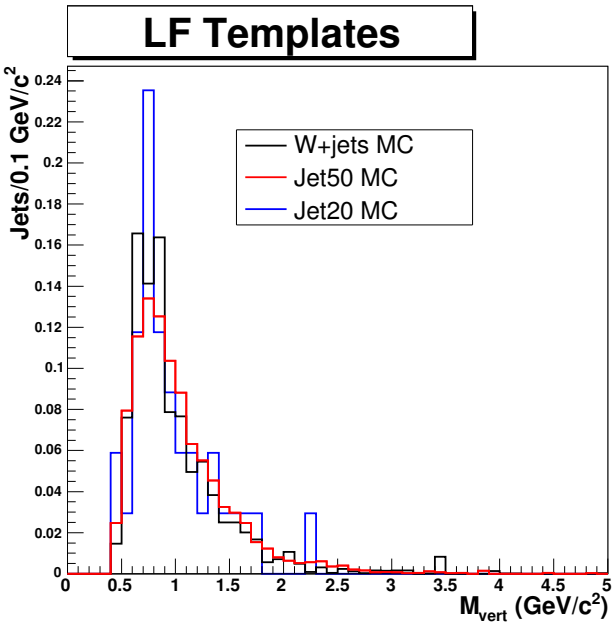

Figure 4: Comparison of different LF template sources. The Jet50 MC shape, because of its healthy statistics, is used by default in vertex mass fits. The vertex mass was shown to be consistent for all samples regardless of kinematic cuts.

$$\mu_i = N_{\rm tot}[f_b^{\rm fit} \cdot N_b^i + f_c^{\rm fit} \cdot N_c^i + (1.0 - f_b^{\rm fit} - f_c^{\rm fit}) \cdot N_{LF}^i] \tag{6}$$

Here N_x^i is the normalized contribution in bin i of the vertex mass template for species x, and $f_x^{\rm fit}$ is the fit fraction for species x. The fit for the contributions from the three species b, c, and LF is reduced to a two-component problem by the restriction $f_b^{\rm fit} + f_c^{\rm fit} + f_{LF}^{\rm fit} = 1.0$.

The likelihood, L, can be constructed across the $N_{\rm bins}$ of the vertex mass distribution:

$$L = \prod_{i=1}^{N_{\rm bins}} P(n_i|\mu_i) \tag{7}$$

$$= \prod_{i=1}^{N_{\rm bins}} \frac{e^{-\mu_i}\mu_i^{n_i}}{n_i!} \tag{8}$$

and once the logs are taken

$$lnL = ln[\prod_{i=1}^{N_{\rm bins}} \frac{e^{-\mu_i}\mu_i^{n_i}}{n_i!}] \quad (9)$$

$$= \sum_{i=1}^{N_{\rm bins}} [-\mu_i + n_i ln\mu_i + const] \quad (10)$$

one can maximize the likelihood with respect to the adjustable species fractions, $f_x^{\rm fit}$. The MINUIT [18] package is used for this purpose.

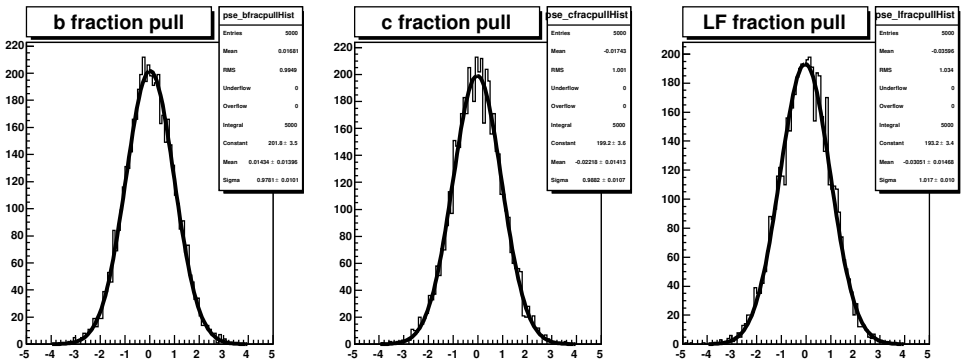

Figure 5: Pull distributions for 5000 pseudoexperiments. Pseudodata for these experiments were constructed in the b/c/LF ratio 70/15/15.

Pseudoexperiment studies demonstrated that the vertex mass shapes were reliable for the purpose of fitting and that the fits were unbiased. Figure 5 shows the pull distributions from a set of 5000 pseudoexperiments performed on pseudodata constructed under the scenario b/c/LF = 70/15/15. The pseudodata assumed the statistics of the selected Ultratight tagged sample in 1.9/fb. From the pulls one can see that the fits find the correct answer (pulls means ∼0.0) and the fit errors are well-behaved (pull RMS values ∼1.0). Pseudoexperiments were performed at a variety of b/c/LF scenarios. The vertex mass shapes successfully found the true b/c/LF fractions in all cases.

Figure 6 shows the absolute error on the fitted b fraction from the 5000 pseudoexperiments in the b/c/LF = 70/15/15 scenario. The mean of the distribution is ∼0.048, which corresponds to a relative error on the b fraction of ∼6.9% given $f_b^{true} = 0.70$. An expected relative error of 5-10% is expected on $f_b^{\rm fit}$ under all f_b^{true} scenarios. This error is purely from the fit procedure and is strongly influenced by the statistics of the data distribution.

An alternative method for constructing vertex mass templates was developed in which the b template contained just b jets from $W^{\pm}$ + b-jet events. In this case a shape and normalization

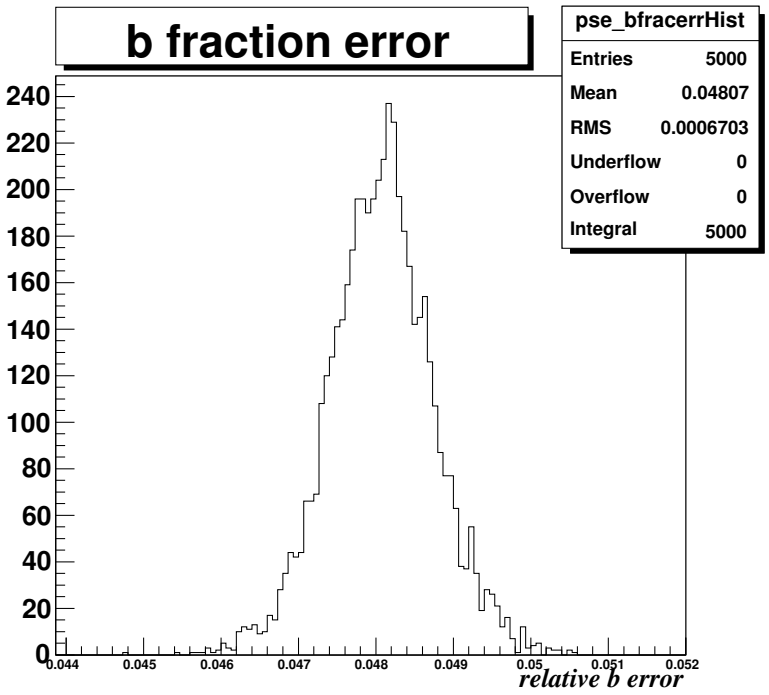

Figure 6: Absolute error on the fitted b fraction from 5000 pseudoexperiments under the b/c/LF = 70/15/15 scenario. With a mean of ∼0.048, this corresponds to a error/true value = 0.048/0.70 = 6.9% relative error.

was assumed for the top backgrounds, and the vertex mass fits in the data attempted to fist subtract these backgrounds from the shape and fit the residual distribution. This alternative was abandoned due to the significant influence on the result from the systematic error on the $t\bar{t}$ and single top cross sections.

6.1 Vertex Mass Shape Systematics

The vertex mass of b jets in MC might not be an accurate representation of b jets in the data. A calibration sample of data b jets was constructed to probe this. Events are collected within the 8 GeVmuon sample, path MUON_CMUP8. A high p_T non-isolated muon is identified via:

- $p_T > 9$ GeV/c
- CMU $|\Delta x| < 3.0$
- CMP $|\Delta x| < 5.0$
- $I > 0.1$

The tracks of selected muons are further required to originate within 5cm in z of the primary vertex and to be fiducial to the CDF silicon detector. The non-isolation requirement preferen-

tially selects muons buried within jets. The muon-jet, j^{μ} then is a L5-corrected E_T>9 GeVjet matched (ΔR<0.4) to the identified muon. This muon-jet is paired with a back-to-back ($\Delta\phi$> 2 radians) partner jet, referred to as the away-jet, j^A. The j^A is required to have $E_T > 15$ GeV.

By requiring that the dijet system be back-to-back one enhances the fraction of jet pairs in which the flavor on each side of the system is equivalent; by demanding one of the jets have a high p_T muon one enriches the system in heavy flavor. The b calibration sample selection is completed by applying the following tagging requirements:

- the j^A is required to possess an Ultratight SECVTX tag
- the j^{μ} is required to possess an Ultratight SECVTX tag and that tagged vertex must have $M_{vert} > 1.7$ GeV/c^2

In this double-tagged dijet system, the purity of b jets in the j^A sample is very high. Studies of Pythia dijet MC indicate this sample is essentially 100% pure in b jets. Because of this one can use this sample of away-jets as a b calibration in the data. Figure 7 contains a comparison of the b vertex mass template attained from MC and that of the b data calibration.

The difference in shape between the MC template and the data calibration was measured in pseudoexperiments. The b-jet pseudodata was drawn from the b calibration shape, along with the default pseudodata sources for charm and light flavor, and then the species contributions were fitted for using the default templates. The data-to-MC b shape difference manifests itself as a $\sim$8% relative effect on the fitted b fraction f_b, see Figure 8. This systematic covers all data-to-MC vertex mass differences for b jets, including fragmentation effects, B hadron decay branching ratios, rates of single/multiple B hadrons inside the b jet and tracking simulation deficiencies.

Uncertainties in the charm and light flavor shapes also influence the fitted b fraction. Unlike the calibration of the b shape, a pure sample of tagged jets in the data from charm and light flavor is not available. So alternative techniques using MC events must be constructed for probing these effects.

From studies of the b shape discrepancy between data and MC, it was found that the uncertainty on the rate of single/multiple B hadrons inside jets had the greatest effect on the fitted b fraction. Here we assume the rate of single/multiple C hadrons will have the greatest effect on $f_c^{\rm fit}$, and that this in turn will be the largest charm shape systematic on the fitted b fraction. Existing $W^{\pm}$ + c-jet MC samples were used; the single and multiple c jet samples were separated and reweighted together to provide a c template with varied single-to-multiple C hadron ratio. The test templates were then used to construct pseudodata which were then

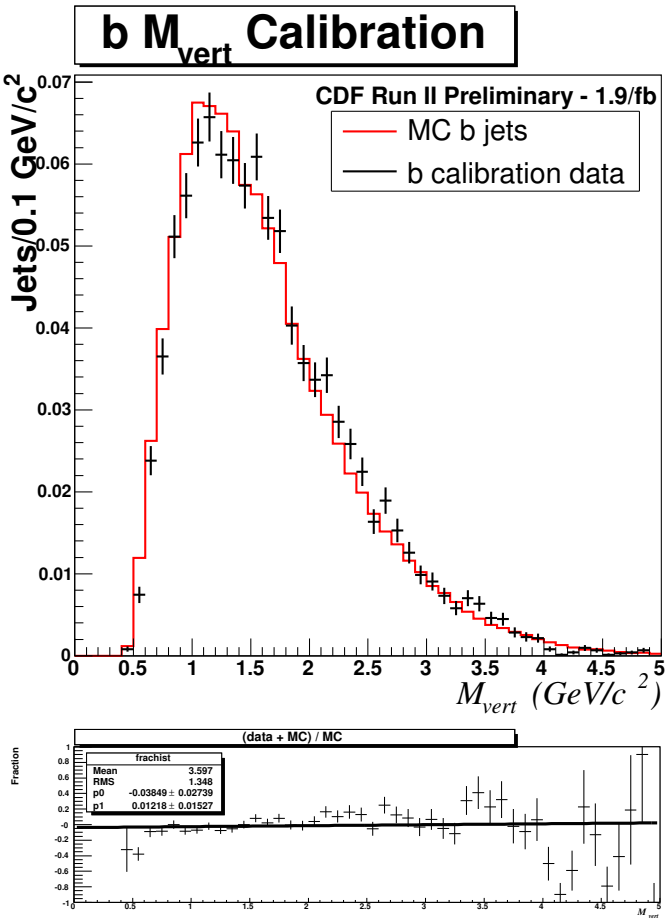

Figure 7: Comparison of default vertex mass shape for b jets versus a calibration shape from data.

fit with the original templates. The single/ multi C hadron multiplicity has little effect on the fitted b-fraction; the relative error on the fitted b fraction f_b was found to be ~ 1 %.

The default light flavor template is made from positively tagged jets matched to strictly light flavor hadrons in Jet50 jet MC. This may not be an accurate model for LF; in order to probe what kind of systematic error this choice incurs, other LF models were considered: positive LF tags in Jet20 MC, positive LF tags in $W^{\pm}$+jets MC, and negative tags in Jet20 data. The MC based LF templates have the virtue of being LF pure; the data driven LF template has the virtue of containing a large amount of statistics. Negative tags in the data are not a pure model for LF, though, since real heavy flavor can contribute to the negative tags. Regardless we use the results from these alternative templates, however flawed, to gauge the effect of systematic

(a) (b)

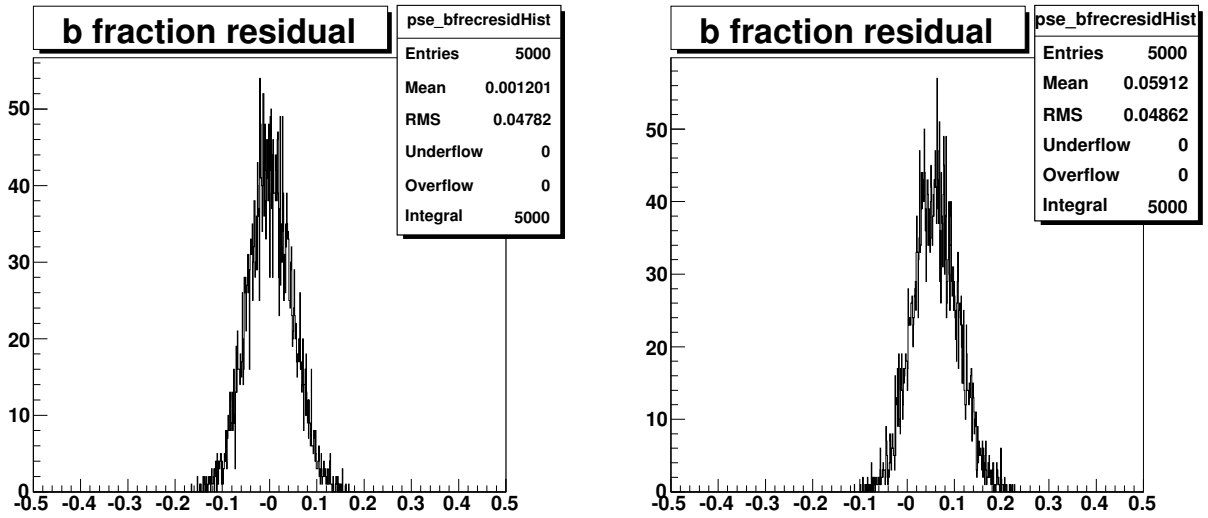

Figure 8: Fitted b fraction residual for 5000 pseudoexperiments under the b/c/LF = 70/15/15 scenario. (a) contains the residual for b pseudodata made from the b template; under these conditions the fitter generally gets the correct answer (residual = $f_b^{fit} - f_b^{true} \sim 0.0$). (b) contains the residual for pseudoexperiments with b pseudodata made from the b calibration sample from the data; one can see that the residual in this case is non-zero ($f_b^{fit} - f_b^{true} \sim 0.06$), which is taken as a 8% systematic on the ultimate fitted b fraction.

error in LF shape on the fitted b-fraction. The relative error on the fitted b fraction was found to be ~ 3 %.

7 Background Sources of b Jets

Not all tagged b jets in the W12j sample are from $W^{\pm}+$ b jet production. Background sources of b jets need to be subtracted from the overall yield of tagged b jets to get the contribution from signal. The background assessment is split into two categories: top and electroweak backgrounds and fake $W^{\pm}$ backgrounds.

7.1 Top and Electroweak Backgrounds

Paired and single top production will contribute to the tagged W12j sample since the decay $t \to Wb$ occurs nearly 100% of the time. Electroweak processes such as diboson production will also contribute to the sample. To estimate the contributions from these backgrounds we measure the rate by which these events satisfy the complete event selection in MC; the samples used for

study are listed in Table 3. Small backgrounds, such as $W^{\pm}$ + jets, where $W \to \tau\nu$, and Z+jets production are also predicted from MC. For these purposes, SM NLO cross sections and $m_t = 175$ GeV/c^2 are universally assumed. Tag rates in the MC are multiplied by the data-to-MC efficiency scale factor. An integrated luminosity of 1.9/fb is assumed for the total background yields.

Process	DSID	N_{gen}	σ (pb)
$t\bar{t}$	ttop75	4719380	6.7 ± 0.7
s-channel	stop00	918173	0.28 ± 0.12
t-channel	stopm0	1012500	0.636 ± 0.28
WZ	itopwz	2340140	3.96 ± 0.06
ZZ	itopzz	2323810	1.58 ± 0.02
WW	itopww	2284860	12.4 ± 0.25
$W + bb + 0$p, $W \to \tau\nu$	dtop0w	1503020	2.98 ± ∼0.3
$W + bb + 1$p, $W \to \tau\nu$	dtop1w	1523790	0.888 ± ∼0.1
$W + bb + \geq 2$p, $W \to \tau\nu$	dtop2w	1478880	0.286 ± ∼0.03
$Z + bb + 0$p, $Z \to e^+e^-$	ztopb0	532205	0.511 ± ∼0.05
$Z + bb + 1$p, $Z \to e^+e^-$	ztopb1	525955	0.134 ± ∼0.01
$Z + bb + \geq 2$p, $Z \to e^+e^-$	ztopb2	519500	0.0385 ± ∼0.004
$Z + bb + 0$p, $Z \to \mu^+\mu^-$	ztopb5	530793	0.511 ± ∼0.05
$Z + bb + 1$p, $Z \to \mu^+\mu^-$	ztopb6	525695	0.134 ± ∼0.01
$Z + bb + \geq 2$p, $Z \to \mu^+\mu^-$	ztopb7	536159	0.0385 ± ∼0.004
$Z + bb + \geq 0$p, $Z \to \tau^+\tau^-$	ztopbt	1456740	0.625 ± ∼0.06

Table 3: MC samples used for tagged b jet background estimation.

Simply counting the number of Ultratight tagged jets in events passing the W12j selection is insufficient; one needs to know the fraction of these tagged jets that are actually matched to b's. These fractions ${f_b}^{\text{tag}}$ for each MC sample are listed in Table 4. Most processes have ${f_b}^{\text{tag}} > 0.95$; this is because most of these processes are sources of real b jets at LO. Distinct among these is the WW process, in which b jets can only occur in splitting of initial or final state radiated gluons. Hence ${f_b}^{\text{tag}}$ is small for WW production.

Table 4 contains the predicted background tagged b jet yields for each sample. The yields are separated into the W1j and W2j samples for later convenience.

7.2 Fake $W^{\pm}$ Background from QCD Jets

The second class of background one needs to consider in the $W^{\pm}$ + b-jet measurement is that coming from spurious $W^{\pm}$ signals in QCD multijet events. The $W^{\pm}$ signature can occur in

Process	f_b^{tag}	
	W+1j	W+2j
$t\bar{t}$	0.997 ± 0.001	0.992 ± 0.002
s-channel	0.9990 ± 0.0003	0.9990 ± 0.0004
t-channel	0.9991 ± 0.0004	0.9952 ± 0.0006
WZ	0.917 ± 0.009	0.919 ± 0.009
ZZ	0.90 ± 0.04	0.92 ± 0.03
WW	0.07 ± 0.02	0.10 ± 0.02
$W + bb + 0\text{p}$, $W \to \tau\nu$	1.0 ± 0.0	1.0 ± 0.0
$W + bb + 1\text{p}$, $W \to \tau\nu$	0.973 ± 0.008	0.991 ± 0.005
$W + bb + \geq 2\text{p}$, $W \to \tau\nu$	0.93 ± 0.02	0.96 ± 0.02
$Z + bb + 0\text{p}$, $Z \to e^+e^-$	1.0 ± 0.0	1.0 ± 0.0
$Z + bb + 1\text{p}$, $Z \to e^+e^-$	0.96 ± 0.02	0.993 ± 0.010
$Z + bb + \geq 2\text{p}$, $Z \to e^+e^-$	1.0 ± 0.0	0.97 ± 0.05
$Z + bb + 0\text{p}$, $Z \to \mu^+\mu^-$	1.0 ± 0.0	1.0 ± 0.0
$Z + bb + 1\text{p}$, $Z \to \mu^+\mu^-$	0.987 ± 0.005	0.998 ± 0.002
$Z + bb + \geq 2\text{p}$, $Z \to \mu^+\mu^-$	0.93 ± 0.02	0.97 ± 0.01
$Z + bb + \geq 0\text{p}$, $Z \to \tau^+\tau^-$	0.987 ± 0.004	0.992 ± 0.003

Table 4: Fraction of Ultratight SECVTX tagged jets that are matched to B hadrons.

these events for several reasons:

- narrow jets that mimic electrons
- semileptonic hadron decay within a jet
- jet punch-through to the muon system
- mismeasured jet energies that give rise to false missing energy

This background is particularly burdensome because it is difficult to model and hence the overall contribution is hard to predict with satisfactory precision. In the case of tagging analyses, there is the additional problem of knowing the heavy flavor content of the tagged sample. To minimize the impact of the uncertainty on this background process, a strict fake $W^{\pm}$ veto (see Section 3) was designed to remove as much of these events as possible.

The remaining fake $W^{\pm}$ background is estimated from the data. The procedure is as follows. The $\not\!\!E_T$ requirement is relaxed in the W12j selected data sample. Templates of $\not\!\!E_T$ are taken from different models for the contributing processes: $W^{\pm}$+LF, $W^{\pm}$+HF, $t\bar{t}$, single top, diboson and fake $W^{\pm}$. Then the entire $\not\!\!E_T$ shape is fit; this provides one with the fraction of events in

Process	n^b_{W+1j}	n^b_{W+2j}	n^b_{W+12j}
$t\bar{t}$	7.1 ± 1.0	66.0 ± 9.2	73.1 ± 10.1
s-channel	4.0 ± 1.7	18.2 ± 7.9	22.2 ± 9.6
t-channel	13.4 ± 6.1	19.9 ± 9.0	33.4 ± 15.0
WZ	2.6 ± 0.2	6.5 ± 0.6	9.1 ± 0.9
ZZ	0.07 ± 0.008	0.21 ± 0.02	0.28 ± 0.03
WW	0.19 ± 0.04	0.64 ± 0.10	0.83 ± 0.12
$W + bb + 0$p, $W \to \tau\nu$	3.2 ± 0.4	2.1 ± 0.3	5.2 ± 0.7
$W + bb + 1$p, $W \to \tau\nu$	0.40 ± 0.06	1.4 ± 0.2	1.8 ± 0.2
$W + bb + \geq 2$p, $W \to \tau\nu$	0.04 ± 0.005	0.25 ± 0.04	0.29 ± 0.04
$Z + bb + 0$p, $Z \to e^+e^-$	0.18 ± 0.03	0.31 ± 0.04	0.49 ± 0.07
$Z + bb + 1$p, $Z \to e^+e^-$	0.03 ± 0.004	0.13 ± 0.02	0.16 ± 0.02
$Z + bb + \geq 2$p, $Z \to e^+e^-$	0.001 ± 0.0002	0.023 ± 0.003	0.024 ± 0.004
$Z + bb + 0$p, $Z \to \mu^+\mu^-$	2.0 ± 0.3	0.9 ± 0.1	2.9 ± 0.4
$Z + bb + 1$p, $Z \to \mu^+\mu^-$	0.26 ± 0.04	0.75 ± 0.10	1.01 ± 0.14
$Z + bb + \geq 2$p, $Z \to \mu^+\mu^-$	0.021 ± 0.003	0.13 ± 0.02	0.15 ± 0.02
$Z + bb + \geq 0$p, $Z \to \tau^+\tau^-$	0.57 ± 0.08	0.91 ± 0.13	1.48 ± 0.20

Table 5: Top, EWK and diboson b jet backgrounds for ultratight SECVTX for 1900/pb.

the sample that come from each process. From this one can re-apply the $\not\!\!E_T$ cut and determine the contribution from the fake $W^\pm$ background in the final selected sample.

The important detail of this technique is the model used for the fake $W^\pm$ events. Here we use as the fake $W^\pm$ model *antielectron* objects. An antielectron is defined as an object that satisfies the kinematical requirements of a candidate electron but fails at least 2 of the remaining requirements. The antielectrons are collected on the high p_T electron trigger but are orthogonal to the sample ultimately used in the measurement.

Figure 9 shows the $\not\!\!E_T$ fits used to determine the fake $W^\pm$ contribution to the selected event sample for events with one or more Ultratight tag. Individual fits are performed in each jet multiplicity bin separately and for each trigger lepton category. This is done because of the definition for the fake $W^\pm$ veto is different in each of these categories, and hence the residual fake $W^\pm$ contribution will be different. Antielectrons are used as a model for fake $W^\pm$ events from both electron and muon triggers. From Figure 9 one can see that the fake $W^\pm$ contribution is largest for electron triggers, but on average contributes just a few percent to the selected sample across categories.

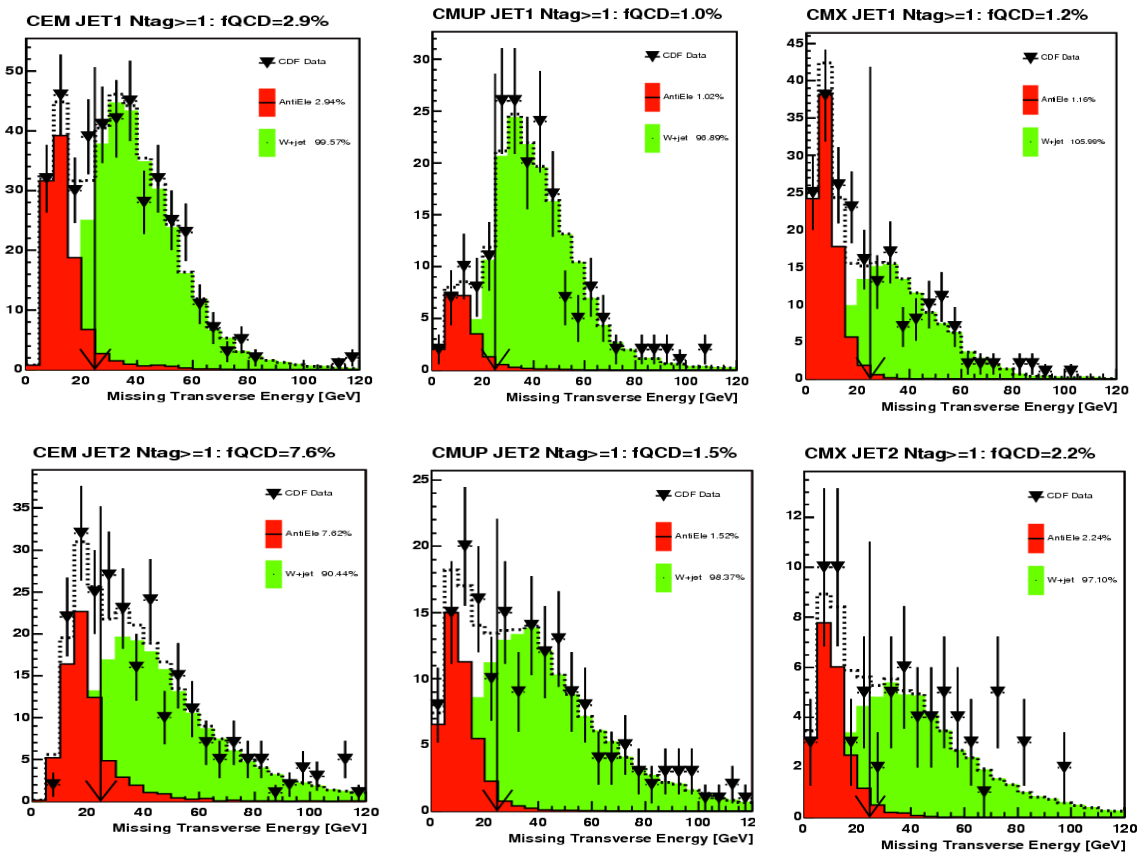

Figure 9: The $\not\!\!E_T$ fits that are used to determine the contribution from fake $W^{\pm}$ events to the W12j sample.

Once one has the overall prediction for the tagged jets in fake $W^{\pm}$ events in the W12j sample, one needs to determine the HF content of that tagged sample. Ideally one could continue to use tagged jets in the antielectron events that pass the full selection including tagging; this is not possible however because of a lack of statistics. It was determined that the antielectron event sample is severely reduced via the $\not\!\!E_T > 25$ GeV requirement.

So in an effort to recoup lost statistics and make a statement on the HF content of the tagged sample from fake $W^{\pm}$ events, the $\not\!\!E_T$ requirement was relaxed completely ($\not\!\!E_T > 0$ GeV), the tagged jets in these antielectron events were collected and the vertex mass of the newfound statistics was fit (see Figure 10(a)). To probe the $\not\!\!E_T$ influence on HF content of the fake $W^{\pm}$ sample, three different minimum $\not\!\!E_T$ cuts were considered. From Figure 10 (a)-(c) one can see that a large HF content is favored ($f_b^{QCD} >\sim 0.70$) in these events. From the studies of these various $\not\!\!E_T$ cuts, the fake $W^{\pm}$ b fraction $f_b^{QCD} = 0.8 \pm 0.2$ was determined to be a reasonable choice.

(a) (b) (c)

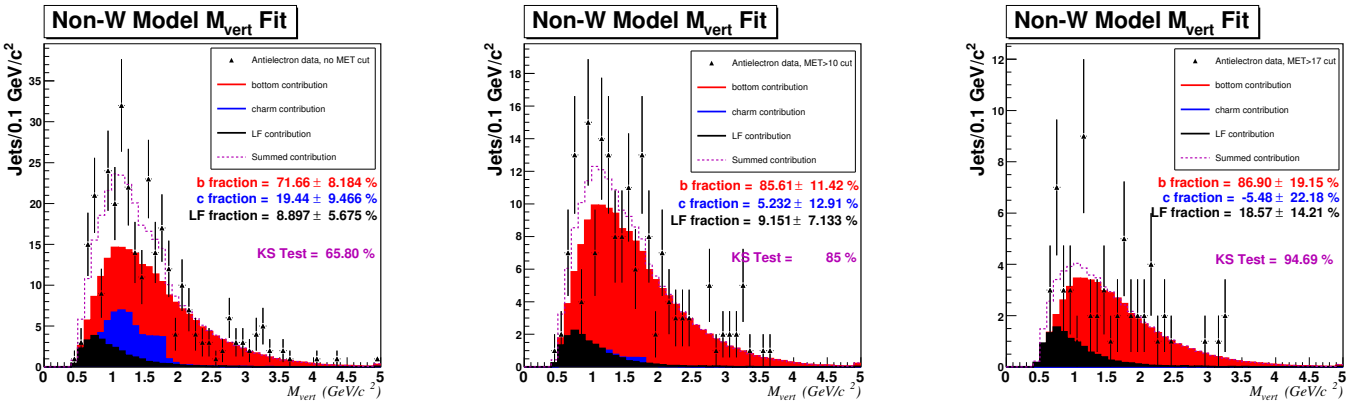

Figure 10: Vertex mass fit to tagged jets in anti-electron data events. The MET requirement in these events was relaxed in order to collect sufficient tagged jet statistics. Three MET cuts were investigated: $\not\!\!E_T >$ 0 GeV- no cut (a); $\not\!\!E_T >$ 10 GeV(b); and $\not\!\!E_T >$ 17 GeV. From these results we set $f_b^{QCD} = 0.8 \pm 0.2$.

7.3 Summary of Backgrounds

With the background studies completed one can construct a complete summary table, Table 6, for all the tagged b jet background contributions to the W12j sample. Notable are the main contributors: $t\bar{t}$ contributes 41% of the total background, single top 32% and fake $W^{\pm}$ 15% .

Process	n^b_{W+1j}	n^b_{W+2j}	n^b_{W+12j}
$t\bar{t}$	7.1 ± 1.0	66.0 ± 9.2	73.1 ± 10.1
s-channel	4.0 ± 1.7	18.2 ± 7.9	22.2 ± 9.6
t-channel	13.4 ± 6.1	19.9 ± 9.0	33.4 ± 15.0
WZ	2.6 ± 0.2	6.5 ± 0.6	9.1 ± 0.9
ZZ	0.07 ± 0.008	0.21 ± 0.02	0.28 ± 0.03
WW	0.19 ± 0.04	0.64 ± 0.10	0.83 ± 0.12
$W + bb$+Np, $W \to \tau\nu$	3.6 ± 0.4	3.7 ± 0.3	7.3 ± 0.8
$Z + bb$+Np, $Z \to e^+e^-$	0.21 ± 0.03	0.46 ± 0.05	0.67 ± 0.08
$Z + bb$+Np, $Z \to \mu^+\mu^-$	2.3 ± 0.3	1.8 ± 0.2	4.1 ± 0.4
$Z + bb+ \geq$Np, $Z \to \tau^+\tau^-$	0.57 ± 0.08	0.91 ± 0.13	1.48 ± 0.20
Non-W	9.4 ± 3.7	15.1 ± 6.3	24.5 ± 8.4
Total	43.4 ± 7.5	133.4 ± 21.0	176.8 ± 22.3

Table 6: Summary of b jet backgrounds for ultratight SECVTX for 1900/pb.

8 Acceptance

The acceptance to detect the b jets in $W^{\pm}$+b-jet production is calculated in the ALPGEN samples listed in Table 2. The events considered must pass the MC-level requirements listed in Section 4. By restricting attention to a specific region of the possible phase space, the result is protected from the impact of inadequacies of the modeling of $W^{\pm}$ and b jet kinematics in regimes where we have no experimental sensitivity in this analysis.

The acceptance comes in the denominator of the b-jet cross section expression from Section 4:

$$\sigma_{b-\text{jets}}(W+b\text{−jets}) \times BR(W \to \ell\nu) = \frac{n_{\text{UT}} \cdot f^{b \text{ jets}} - n^{b \text{ jets}}_{\text{bkgd}}}{\mathcal{L} \cdot \mathcal{A}^{b \text{ jets}}_{W+Nb} \cdot \epsilon^{b}_{\text{tag}}} \quad (11)$$

$$= \frac{n_{\text{UT}} \cdot f^{b \text{ jets}} - n^{b \text{ jets}}_{\text{bkgd}}}{\sum_t [\mathcal{L} \cdot \mathcal{A}^{b \text{ jets}}_{W+Nb} \cdot \epsilon^{b}_{\text{tag}}]_t} \quad (12)$$

$$= \frac{n_{\text{UT}} \cdot f^{b \text{ jets}} - n^{b \text{ jets}}_{\text{bkgd}}}{\sum_t [\mathcal{L}_t \cdot \epsilon_{z_0,t} \cdot \epsilon_{\text{ trig},t} \cdot \epsilon_{\text{ lepID},t} \cdot (\mathcal{A}_t \cdot \epsilon^{b}_{\text{tag},t})]} \quad (13)$$

where $t \in CEM, CMUP, CMX$. In this Section the focus is $(\mathcal{A}_t \cdot \epsilon^{b}_{\text{tag},t})$, which is the acceptance times efficiency for identifying and selecting the b jets from the signal process. This quantity has contributions from each of the ALPGEN signal samples for each $W^{\pm}$ decay mode:

$$(\mathcal{A}_t \cdot \epsilon^{b}_{\text{tag},t}) = \sum_{i \in \text{samples}} [w_i \cdot (\mathcal{A}_t \cdot \epsilon^{b}_{\text{tag},t})_i] \quad (14)$$

This term is a sum over the ALPGEN $W^{\pm}$+jets MC samples ($W \to e\nu$ samples contribute to CEM, $W \to \mu\nu$ samples contribute to CMUP and CMX). The weights w_i are given by

$$w_i = \frac{(\sigma_{\text{bhadjet}} \times BR)_i}{\sum_{i \in \text{samples}} (\sigma_{\text{bhadjet}} \times BR)_i} \quad (15)$$

and are determined by the contribution from each ALPGEN $W^{\pm}$ + jets sample to the overall predicted ALPGEN cross section. These weights are given in Table 2.

The quantity $(\mathcal{A}_t \cdot \epsilon^{b}_{\text{tag},t})_i$ relates reconstructed-level quantities to MC-level quantities. It can be expanded into a product of three constituent factors:

$$(\mathcal{A}_t \cdot \epsilon^{b}_{\text{tag},t})_i = \mathcal{A}_{\text{jet}_i} \times \mathcal{A}_{\text{sel}_i} \times \epsilon_{\text{UT}_i} \quad (16)$$

where $\mathcal{A}_{\text{jet}}$ is given by

$$\mathcal{A}_{\text{jet}} = \frac{\#\ b\text{-matched CDF jets in MC signal events satisfying phase space}}{\#\ b\text{-matched MC-level jets in MC signal events satisfying phase space}}$$

and encodes the effect of smearing jets at MC level and how they manifest themselves at reconstructed level.

$\mathcal{A}_{\rm sel}$ is given by

$$\mathcal{A}_{\rm sel} = \frac{\#\ b\text{-matched CDF jets in events passing W12j pretag sel and phase space req's}}{\#\ b\text{-matched CDF jets in MC signal events satisfying phase space}}$$

and encodes the effect of the applying the pretag event selection.

Finally $\epsilon_{\rm UT}$ is given by

$$\epsilon_{\rm UT} = \frac{\#\ b\text{-matched UT-tagged CDF jets in events passing pretag W12j, phase space reqs}}{\#\ b\text{-matched CDF jets in events passing W12j pretag sel and phase space req's}} \times f_{UT}$$

and is the per-b-jet Ultratight tag efficiency in data for b jets in events passing the full pretag event selection. The factor f_{UT} is the data-to-MC Ultratight SECVTX tag efficiency scale factor, which is necessary to translate a MC-based tag rate into the actual tag efficiency in the data.

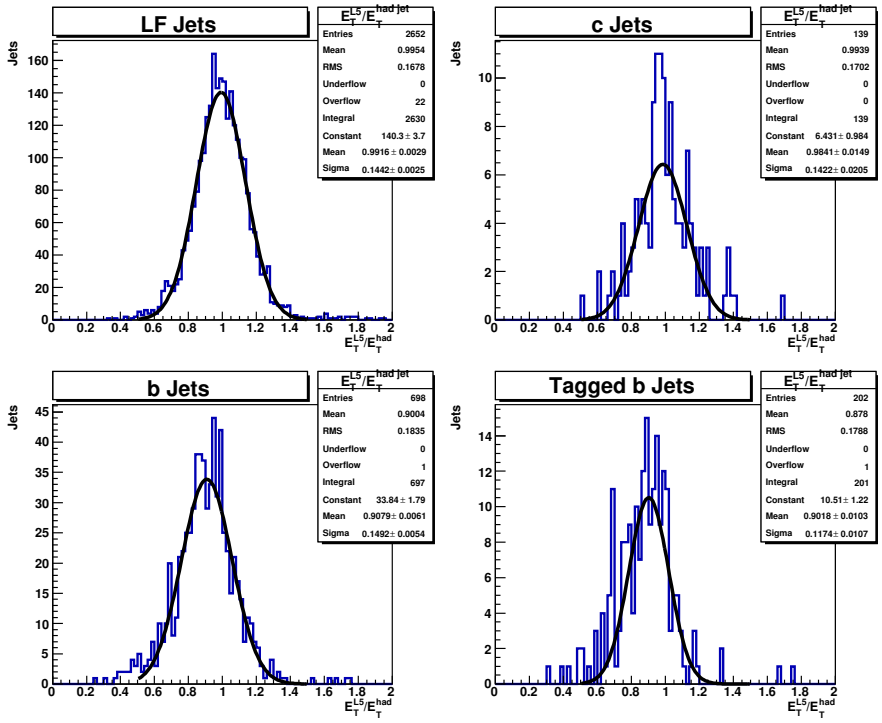

Figure 11: E_T studies of MC-level and reconstructed level jets. MC-level and reconstructed level jets are required to match with $\Delta R < 0.4$.

Important for the acceptance denominator is the definition of a jet at MC level. This is accomplished here via the SpartyJet package [17] which allows one to make JetClu0.4 jets out

of final state particles in MC events. Per CDF convention, all final state particles are considered for jet construction except those from $W^{\pm}$ decay. Reconstructed jets will not be exact replicas of the jets at MC level; effects from fragmentation, hadronization, shower, and the response of the CDF detector smear the true jet energies. The convention for MC-level jet construction will create another source of mismatch: particles that interact minimally with the CDF calorimetry will be included in MC-level jets, and their presence will clearly not be accounted for in the reconstructed CDF jets. This effect should be largest for b jets, where semileptonic decays of B hadrons will play a significant role. This effect is clear from Figure 11, where one can see the deficit in reconstructed jet E_T for b jets, both pretag and tagged, with respect to the corresponding MC-level jets. No additional correction is applied to account for this mismatch; instead the effect of the E_T mismatch is encoded in the jet smearing term of the acceptance, $\mathcal{A}_{\rm jet}$.

	A_{jet}	A_{sel}	$\epsilon_{\rm UT}$	w	$(A \times \epsilon_{tag})_{CEM}$
Wevbb0p	0.7871 ± 0.0008	0.4815 ± 0.001	0.1556 ± 0.001	0.7218	0.04256 ± 0.0003
Wevbb1p	0.6798 ± 0.0009	0.5629 ± 0.001	0.1606 ± 0.001	0.2037	0.01252 ± 9e-05
Wevbb2p	0.6811 ± 0.001	0.576 ± 0.002	0.1592 ± 0.002	0.02966	0.001853 ± 2e-05
Wevcc0p	1.306 ± 0.01	0.3438 ± 0.06	0 ± 0	0.0001566	0 ± 0
Wevcc1p	1.132 ± 0.01	0.3636 ± 0.05	0.03143 ± 0.03	7.927e-05	0 ± 0
Wevcc2p	0.9455 ± 0.03	0.4615 ± 0.07	0.03667 ± 0.04	2.348e-05	0 ± 0
Wevc0p	1.295 ± 0.01	0.1754 ± 0.05	0 ± 0	0.0004963	0 ± 0
Wevc1p	1.069 ± 0.01	0.3377 ± 0.05	0 ± 0	0.000165	0 ± 0
Wevc2p	1.2 ± 0.01	0.625 ± 0.06	0.01956 ± 0.02	2.123e-05	0 ± 0
Wevc3p	0.6786 ± 0.09	0.6316 ± 0.1	0 ± 0	1.707e-06	0 ± 0
Wev0p	1.538 ± 0.01	0.22 ± 0.04	0 ± 0	0.03081	0 ± 0
Wev1p	1.125 ± 0.01	0.3069 ± 0.03	0 ± 0	0.01062	0 ± 0
Wev2p	0.7209 ± 0.07	0.6452 ± 0.09	0.044 ± 0.05	0.002229	0 ± 0
Wev3p	0.7576 ± 0.07	0.56 ± 0.1	0 ± 0	0.0002845	0 ± 0
Wev4p	1.125 ± 0.01	0.4444 ± 0.2	0 ± 0	1.362e-05	0 ± 0
Total					0.0569 ± 0.0003

Table 7: Acceptance for CEM trigger in ALPGEN $W^{\pm}$+jets samples.

Tables 7-9 contain the acceptance values for each of the three trigger samples we consider in this analysis. One can see that the A_{jet} values are different for $W \to e\nu$ and $W \to \mu\nu$ samples; this is a consequence of the choice in how we have factorized $(A \times \epsilon_{tag})$. Since the jets in the numerator of A_{jet} are not required to be in events that have satisfied the lepton identification in the $W^{\pm}$ selection, these jets will have instances in which the reconstructed lepton object

	A_{jet}	A_{sel}	ϵ_{UT}	w	$(A \times \epsilon_{tag})_{CMUP}$
Wmvbb0p	0.8494 ± 0.0007	0.2559 ± 0.0009	0.1557 ± 0.001	0.7208	0.02439 ± 0.0002
Wmvbb1p	0.8867 ± 0.0006	0.2466 ± 0.0009	0.1579 ± 0.001	0.2041	0.007043 ± 7e-05
Wmvbb2p	0.9532 ± 0.0006	0.239 ± 0.001	0.1556 ± 0.002	0.02964	0.001051 ± 2e-05
Wmvcc0p	0.898 ± 0.04	0.1364 ± 0.05	0 ± 0	0.0001574	0 ± 0
Wmvcc1p	1.052 ± nan	0.2469 ± 0.05	0 ± 0	8.95e-05	0 ± 0
Wmvcc2p	0.9861 ± 0.01	0.09859 ± 0.04	0 ± 0	2.954e-05	0 ± 0
Wmvc0p	0.75 ± 0.06	0.2619 ± 0.07	0 ± 0	0.0006173	0 ± 0
Wmvc1p	1.231 ± nan	0.3438 ± 0.05	0 ± 0	0.0001761	0 ± 0
Wmvc2p	0.9726 ± 0.02	0.2676 ± 0.05	0 ± 0	2.56e-05	0 ± 0
Wmvc3p	1.139 ± nan	0.1707 ± 0.06	0 ± 0	1.325e-05	0 ± 0
Wmv0p	0.6528 ± 0.06	0.1915 ± 0.06	0 ± 0	0.0333	0 ± 0
Wmv1p	0.9926 ± 0.007	0.2985 ± 0.04	0 ± 0	0.008321	0 ± 0
Wmv2p	0.8261 ± 0.06	0.2105 ± 0.07	0 ± 0	0.00237	0 ± 0
Wmv3p	1 ± 0	0.1923 ± 0.08	0 ± 0	0.0002204	0 ± 0
Wmv4p	1.25 ± nan	0.3333 ± 0.1	0 ± 0	0.0001017	0 ± 0
Total					0.0325 ± 0.0002

Table 8: Acceptance for CMUP trigger in ALPGEN $W^{\pm}$+jets samples.

actually fails the lepton ID. In such cases these objects do not disappear from the event record — they appear in the jet list. If one examines A_{jet} but further requires the jets in the numerator occur in events with a reconstructed lepton satisfying our event selection, then A_{jet} is more consistent between the $W \to e\nu$ and $W \to \mu\nu$ samples.

One can also see that the per-b-jet tag efficiency is consistent for all samples considered. This is to be expected, since the tagger behavior should not depend on the trigger lepton type.

Generally the acceptance is very different from 100% . Certainly one will lose b jets when using the pure Ultratight tagging operating point. But upstream of the tagging it is interesting to note that, even in these MC events with a high p_T central charged lepton, high p_T neutrino and 1 or 2 central high E_T jets, we lose a significant portion of these b jets. A small study was performed to examine this loss in acceptance. The main losses were found to be in the 1,2 reconstructed jet requirement and the lepton identification requirement. Events are lost through the jet multiplicity requirement when MC-level $E_T > 20$ GeV jets are reconstructed with E_T that is too low to meet our reconstructed jet definition. Jets can also be lost through this requirement when at MC level there are 1 or 2 tight MC-level jets but at reconstructed level there are 3 or more. Such events would be outside our acceptance, and the reconstructed b jets would not be counted in the acceptance calculation. The lepton identification criteria is

	A_{jet}	A_{sel}	$\epsilon_{\rm UT}$	w	$(A \times \epsilon_{tag})_{CMX}$
Wmvbb0p	0.8494 ± 0.0007	0.1345 ± 0.0007	0.1543 ± 0.002	0.7208	0.01271 ± 0.0002
Wmvbb1p	0.8867 ± 0.0006	0.1339 ± 0.0007	0.1624 ± 0.002	0.2041	0.003935 ± 5e-05
Wmvbb2p	0.9532 ± 0.0006	0.1311 ± 0.001	0.1594 ± 0.003	0.02964	0.0005902 ± 1e-05
Wmvcc0p	0.898 ± 0.04	0.1591 ± 0.06	0 ± 0	0.0001574	0 ± 0
Wmvcc1p	1.052 ± nan	0.08642 ± 0.03	0 ± 0	8.95e-05	0 ± 0
Wmvcc2p	0.9861 ± 0.01	0.2113 ± 0.05	0 ± 0	2.954e-05	0 ± 0
Wmvc0p	0.75 ± 0.06	0.2143 ± 0.06	0 ± 0	0.0006173	0 ± 0
Wmvc1p	1.231 ± nan	0.1042 ± 0.03	0 ± 0	0.0001761	0 ± 0
Wmvc2p	0.9726 ± 0.02	0.1549 ± 0.04	0 ± 0	2.56e-05	0 ± 0
Wmvc3p	1.139 ± nan	0.122 ± 0.05	0 ± 0	1.325e-05	0 ± 0
Wmv0p	0.6528 ± 0.06	0.06383 ± 0.04	0 ± 0	0.0333	0 ± 0
Wmv1p	0.9926 ± 0.007	0.1119 ± 0.03	0 ± 0	0.008321	0 ± 0
Wmv2p	0.8261 ± 0.06	0.1316 ± 0.05	0 ± 0	0.00237	0 ± 0
Wmv3p	1 ± 0	0.1923 ± 0.08	0 ± 0	0.0002204	0 ± 0
Wmv4p	1.25 ± nan	0.1333 ± 0.09	0 ± 0	0.0001017	0 ± 0
Total					0.0172 ± 0.0002

Table 9: Acceptance for CMX trigger in ALPGEN $W^{\pm}$+jets samples.

somewhat stringent; one could consider investigating loosening this criteria in the context of the WH and single top searches. But that optimization study will be pursued elsewhere.

8.1 Acceptance Systematics

Several sources of systematic error on the acceptance times efficiency are considered. The first source is imperfect jet energy measurements at CDF. The measured energies of jets at CDF are corrected for various effects, including imperfect calorimeter response, eta-dependent influence on the jet energy measurement, contributions to the jet cone from non-primary $p\overline{p}$ interactions and accounting for uninstrumented regions of the detector. These corrections have some uncertainty on them, and this uncertainty affects the per-jet acceptance in this analysis. Systematically measuring jet energies low would artificially reduce the per-jet acceptance; too few events would contain the 1 or 2 $E_T > 20$ GeV jets we demand as part of the event selection. Systematically measuring jet energies high would likewise artificially inflate the per-jet acceptance. Table 10 demonstrates the effect of adjusting the jet energy corrections within their ± 1 σ uncertainties on $(A \times \epsilon_{tag})$; this effect manifests itself as a relative 3% error on this quantity.

The second source of systematic error on the acceptance comes from the choice in renor-

malization and factorization scale used in the ALPGEN signal MC samples. This choice could have an impact on the jet E_T and $|\eta|$ distributions, and since in the event selection we demand events posses exactly 1 or 2 $E_T > 20$ GeV, $|\eta| < 2.0$ jets, pushing these distributions around could affect the acceptance. Table 11 demonstrates the effect the choice of scale Q^2 has on the acceptance. The default scale is $Q^2 = k \cdot (M_W^2 + \sum_p (m_p^2 + p_{T,p}^2))$ with $k = 1$; to test this effect we considered alternative ALPGEN samples with $k = 0.5$ and $k = 2$. The choice of Q^2 manifests itself as a $\sim$3% effect on the acceptance.

The third source of systematic error on the acceptance is the choice in parton distribution function used in the ALPGEN signal MC samples. This systematic was evaluated in the previous $W^{\pm}$+b-jet analysis and was found to have a small effect on the cross section result; its effect on the cross section is dwarfed by that of the systematic incurred from the imperfections in the b model, and in the acceptance itself this source is far smaller than that from uncertainty in the tag efficiency in data. We do not repeat this evaluation here and instead assume that the PDF choice affects the acceptance at the $\sim$2% level, as motivated by the findings in the previous analysis.

The fourth source of systematic error in $(A \times \epsilon_{tag})$ comes from imprecise knowledge of the Ultratight tag efficiency in data b jets. The tag efficiency for b jets is calibrated in data; however this calibration lacks statistics in b jets in the E_T range that we probe in this analysis. Hence we must account for this possible error in the tag efficiency measurement; this manifests itself as a relative error of 6% on the tag efficiency, ϵ_{tag}.

		$(A \times \epsilon_{tag})_i$	$(A \times \epsilon_{tag})_{CEM}$
	Wevbb0p	0.0466 ± 0.0004	
JES $+1\sigma$	Wevbb1p	0.0132 ± 0.0001	0.0616 ± 0.0004
	Wevbb2p	0.0018 ± 3e-05	
	Wevbb0p	0.0447 ± 0.0003	
Default	Wevbb1p	0.0133 ± 0.0001	0.0600 ± 0.0003
	Wevbb2p	0.0020 ± 3e-05	
	Wevbb0p	0.0434 ± 0.0004	
JES -1σ	Wevbb1p	0.0126 ± 0.0001	0.0581 ± 0.0004
	Wevbb2p	0.0021 ± 3e-05	

		$(A \times \epsilon_{tag})_i$	$(A \times \epsilon_{tag})_{CMUP}$
	Wmvbb0p	0.0271 ± 0.0003	
JES $+1\sigma$	Wmvbb1p	0.0072 ± 7e-05	0.035 ± 0.0003
	Wmvbb2p	0.0011 ± 3e-05	
	Wmvbb0p	0.0257 ± 0.0002	
Default	Wmvbb1p	0.0076 ± 6e-05	0.034 ± 0.0002
	Wmvbb2p	0.0011 ± 2e-05	
	Wmvbb0p	0.0241 ± 0.0003	
JES -1σ	Wmvbb1p	0.0075 ± 7e-05	0.033 ± 0.0003
	Wmvbb2p	0.0012 ± 3e-05	

		$(A \times \epsilon_{tag})_i$	$(A \times \epsilon_{tag})_{CMX}$
	Wmvbb0p	0.0133± 0.0003	
JES $+1\sigma$	Wmvbb1p	0.0038± 7e-05	0.018 ± 0.0003
	Wmvbb2p	0.0006 ± 2e-05	
	Wmvbb0p	0.0133± 0.0002	
Default	Wmvbb1p	0.0038± 6e-05	0.018 ± 0.0002
	Wmvbb2p	0.0006 ± 1e-05	
	Wmvbb0p	0.0123± 0.0003	
JES -1σ	Wmvbb1p	0.0038± 7e-05	0.017 ± 0.0003
	Wmvbb2p	0.0007 ± 2e-05	

Table 10: Effect on acceptance in CMX trigger samples due to ± 1 σ variations in the jet energy corrections.

		$(A \times \epsilon_{tag})_i$	$(A \times \epsilon_{tag})_{CEM}$
	Wevbb0p	0.0467 ± 0.0006	
k =0.5	Wevbb1p	0.0128 ± 0.0002	0.0615 ± 0.0006
	Wevbb2p	0.0020 ± 4e-05	
	Wevbb0p	0.0447 ± 0.0003	
Default $k = 1$	Wevbb1p	0.0133 ± 0.0001	0.0600 ± 0.0003
	Wevbb2p	0.0020 ± 3e-05	
	Wevbb0p	0.0456 ± 0.0006	
k =2.0	Wevbb1p	0.0135 ± 0.0002	0.0611 ± 0.0006
	Wevbb2p	0.0019 ± 4e-05	

		$(A \times \epsilon_{tag})_i$	$(A \times \epsilon_{tag})_{CMUP}$
	Wmvbb0p	0.0257± 0.0003	
k =0.5	Wmvbb1p	0.0076± 7e-05	0.034 ± 0.0003
	Wmvbb2p	0.0011 ± 2e-05	
	Wmvbb0p	0.0257 ± 0.0002	
Default $k = 1$	Wmvbb1p	0.0076 ± 6e-05	0.034 ± 0.0002
	Wmvbb2p	0.0011 ± 2e-05	
	Wmvbb0p	0.0266± 0.0003	
k =2.0	Wmvbb1p	0.0076± 7e-05	0.035 ± 0.0003
	Wmvbb2p	0.0011 ± 2e-05	

		$(A \times \epsilon_{tag})_i$	$(A \times \epsilon_{tag})_{CMX}$
	Wmvbb0p	0.0133± 0.0003	
k =0.5	Wmvbb1p	0.0038± 7e-05	0.018 ± 0.0003
	Wmvbb2p	0.0007 ± 2e-05	
	Wmvbb0p	0.0133± 0.0002	
Default $k = 1$	Wmvbb1p	0.0038± 6e-05	0.018 ± 0.0002
	Wmvbb2p	0.0006 ± 1e-05	
	Wmvbb0p	0.0133± 0.0003	
k =2.0	Wmvbb1p	0.0038± 7e-05	0.018 ± 0.0003
	Wmvbb2p	0.0006 ± 2e-05	

Table 11: Effect on acceptance in CMX trigger samples due to Q^2 variations.

9 Impact of Systematic Errors

The various sources of systematic error have been described in the preceding sections. A summary is contained in Table 12, where the relative effect on the ultimate result, $\sigma_{b-\text{jets}}(W + b-\text{jets}) \times BR(W \to \ell\nu)$, is listed for each source of error.

Source	$\frac{\delta_{\sigma_{b-\text{jets}} \times BR}}{\sigma_{b-\text{jets}} \times BR}$ (%)
b shape modeling	8
c shape modeling	1
LF shape modeling	3
UT tag efficiency	6
Luminosity	6
Top Cross Sections	2
Fake $W^{\pm}$ $\not\!\!E_T$ fits	1
Tagged Fake $W^{\pm}$ b fraction	1
Jet Energy Scale	3
Q^2	3
PDF	2
$\lvert z_0 \rvert$ efficiency	<1
Trigger efficiency	<1
Lepton ID efficiency	<1

Table 12: Summary of the sources of systematic error and their impacts on the measurement.

The dominant uncertainty is from systematic error in our model for the b vertex mass. The 8% relative uncertainty comes from exchanging the MC-based b template for one from a calibration b sample in the data. This was also the leading source of systematic error in the previous $W^{\pm}$+b-jets analysis. Here however the result benefits from the utilization of Ultratight SECVTX, with which we can construct a pure sample of b jets in the data. A similar exercise was done in the previous result for the standard Tight SECVTX; however uncertainty in the purity of the calibration sample for the Tight operating point translated into a larger incurred uncertainty in the fitted b fraction, and from that a larger effect in $\sigma_{b-\text{jets}} \times BR$ (∼22% in the previous result).

Additional uncertainty is incurred through our inability to precisely measure tag efficiency for b jets in the data. This systematic error is incurred because the standard calibration technique for tagging algorithms exploits a sample of jets that are predominantly at meager jet E_T, 20 GeV $< E_T <$ 40 GeV. However the typical signal b jets in top measurements and Higgs

searches have $E_T > 40$ GeV. Because of this mismatch a large systematic is applied to the tag efficiency measurement to cover the imprecisely known behavior at large jet E_T (a 6% relative effect). This additional systematic is not exactly fair in the case of signal b jets from $W^{\pm}$+b-jet production, since these b jets are not as energetic as those from top or Higgs decay. So in the future a parameterized tag efficiency with a parametrized systematic uncertainty would offer the potential to reduce the effect of this systematic error.

10 Result

With the various pieces of the analysis in hand, one can turn to the data and complete the result. In this analysis we consider essentially 1.9/fb of CDF Run 2 data. This corresponds to 175712 total events passing the pretag W12j selection, among which there are finally 943 Ultratight tagged jets in total. The breakdown of these yields as a function of trigger source are shown in Table 13.

Trigger Sample	W12j Pretag Events	W12j Pretag Jets	UT tags
CEM	98004	111226	504
CMUP	47243	54030	294
CMX	30465	34414	145
Total	175712	199670	943

Table 13: Yields of Ultratight tagged jets in data events passing the W+1,2 jet selection.

The vertex mass fit in these 943 Ultratight tagged jets in the data is shown in Figure 12. The fitted b fraction is found to be $f_b^{\rm fit} = 0.71 \pm 0.05({\rm stat}) \pm 0.06({\rm syst})$. Recall the prediction from pseudoexperiments was that the statistical error would correspond to a ∼7% effect on the fitted b fraction, the validity of which is borne out here. Recall also that the systematic error amounts to a ∼9% effect on the fitted b fraction and is driven by the b shape calibration. This fitted b fraction corresponds to 670 ± 48(stat) ± 57(syst) tagged b jets; backgrounds contribute 177 ± 22(syst) to this sample, meaning that 493 ± 48(stat) ± 61(syst) are from $W^{\pm}$+b-jet production.

The fit values for the three species can be checked for consistency in kinematic and tagging variables. Figures 13- 15 demonstrate that the species fractions found from the vertex mass do a good job of representing the data in these other variables.

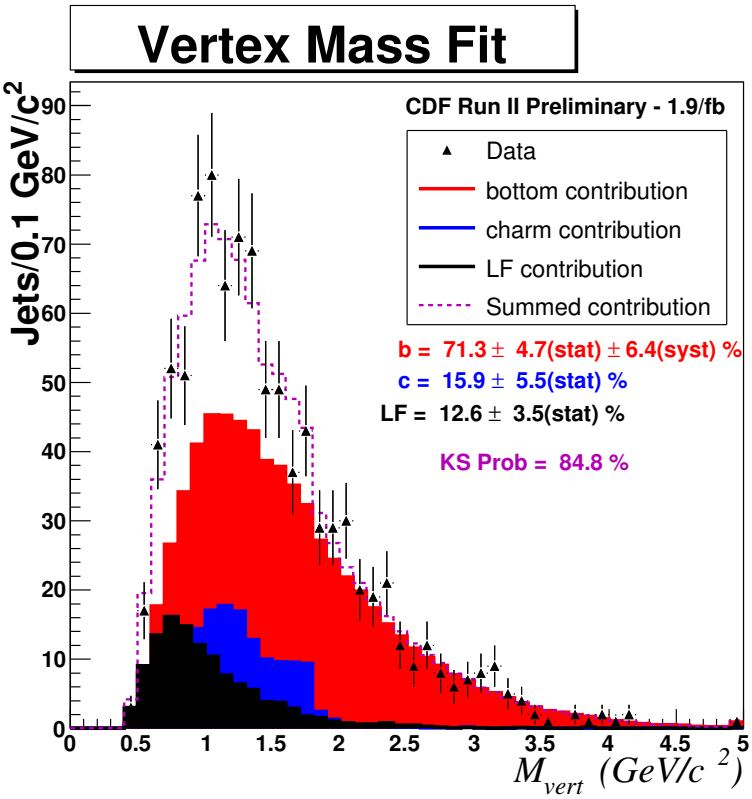

Figure 12: Vertex mass fit for Ultratight tagged jets in selected sample in 1.9/fb.

The denominator from the cross section expression Equation 13 includes contributions from the luminosity, acceptance and efficiency. A summary of these values for each trigger sample can be found in Table 14. The quantities ϵ_{z_0}, $\epsilon_{\rm trig}$ and $\epsilon_{\ell \rm ID}$ were taken from the CDF Joint Physics recommendations for 1.9/fb analyses [19]. Statistical errors on the acceptance and efficiency are considered to be negligible; only systematic errors are assigned for these quantities. The largest source of systematic error in the denominator comes from the Ultratight tag efficiency measurement and the uncertainty on the integrated luminosity measurement.

	CEM	CMUP	CMX
$\epsilon_{z_0} \times \epsilon_{\rm trig} \times \epsilon_{\ell \rm ID}$	0.915 ± 0.005	0.815 ± 0.001	0.915 ± 0.007
$(A \times \epsilon_{tag})$	0.057 ± 0.005	0.033 ± 0.003	0.017 ± 0.001
Luminosity (1/pb)	1916 ± 115	1916 ± 115	1883 ± 113

Table 14: Denominator information for the three trigger sources.

(a) (b)

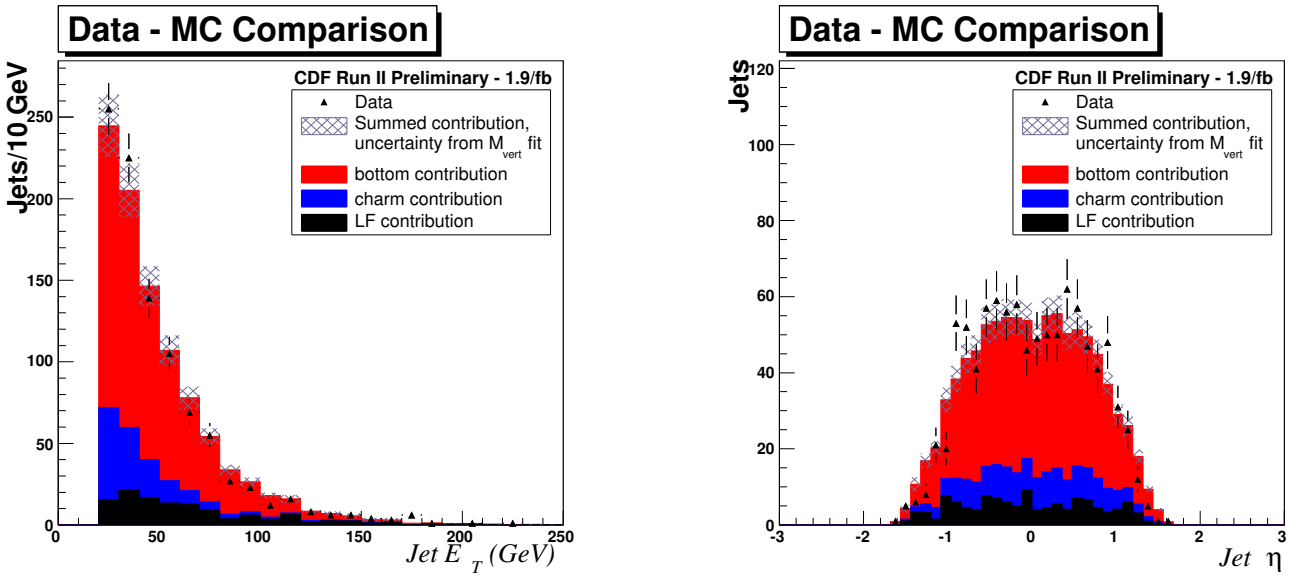

Figure 13: Comparison of corrected jet E_T (a) and detector η (b) for data and MC. The species fractions are those found in the vertex mass fit. The MC shapes are made from the same samples from which the vertex mass templates are constructed.

Folding in the pieces into Equation 13 yields the measured b jet cross section in $W^{\pm}$+b-jet events:

$$\sigma_{b-\text{jets}}(W + b-\text{jets}) \times BR(W \to \ell\nu) = 2.74 \pm 0.27(stat) \pm 0.42(syst)\text{pb} \qquad (17)$$

This measured value can be compared to the ALPGEN prediction of 0.78 pb; the mismatch indicates that the ALPGEN prediction is 3-4 times lower than measured in the data. A prediction from MCFM [20] is being pursued, and one for Pythia is being planned.

A mismatch between data and the ALPGEN prediction is not wholly unexpected. Distrust of the production cross section of $W^{\pm}$+b-jet events (which is dependent on the b jet production cross section in such events) motivated the invention of Method 2. Agreement between the MC prediction and the data for $W^{\pm}$+b-jets would be further surprising given that ALPGEN has a deficit with respect to data in its prediction for inclusive $W^{\pm}$ production [22].

This result has a relative statistical precision of ∼10% , and a systematic precision of ∼16% . This is a significant increase in precision over the previous result (statistical ∼22% , systematic ∼33%). This result is also a significant increase in precision over the current predictions for $W^{\pm}$+b-jet events from Method 2. It is hoped that this increase in precision can be exploited in the construction of the $W^{\pm}$+ jets background prediction for the single top and Higgs analyses. Exploration of these possibilities is the focus of current work.

(a) (b)

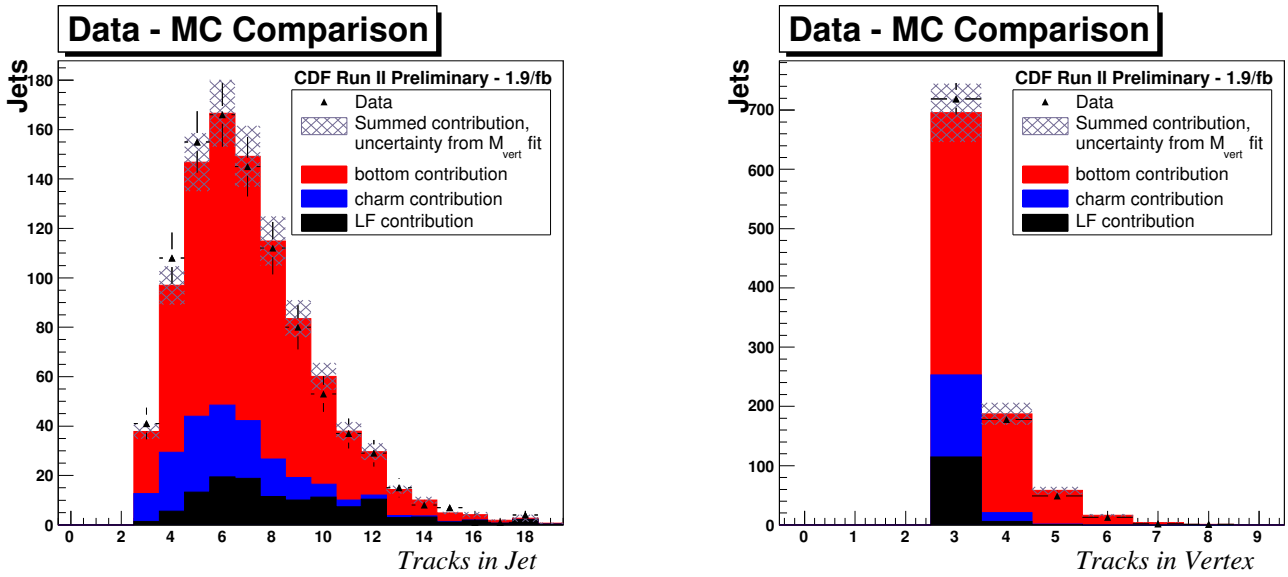

Figure 14: Comparison of number of total tracks inside the jet (a) and number of tracks inside the tagged vertex (b) for data and MC. The species fractions are those found in the vertex mass fit. The MC shapes are made from the same samples from which the vertex mass templates are constructed.

This result is in contrast to the 700/pb result, which saw good agreement with the ALPGEN prediction (measured: 0.90 ± 0.20(stat) ± 0.30(syst) pb; prediction: 0.74 ± 0.18 (syst) pb). An investigation as to how the two versions of the analysis differ was performed and the details are included here [21]; this comparison will be documented thoroughly in a forthcoming version of this note.

11 Summary and Conclusions

We have measured the production cross section for b jets in events with a $W^{\pm}$ boson in 1.9/fb of CDF Run 2 data. Care was taken to insulate the result for influence of the model used for the signal events. The measured jet cross section is 2.74 ± 0.27 (stat) ± 0.42 (syst) pb; this jet cross section applies specifically to events that possess a high p_T central charged lepton, a high p_T neutrino and exactly 1 or 2 high E_T central jets. This result is 3-4 times higher than the prediction from ALPGEN. This result is an improvement over the previous $W^{\pm}$+b jet analysis, which had precision only to within ∼30%. This more precise measurement will be incorporated into $W^{\pm}$ + b-jet predictions for Higgs and single top searches.

(a) (b)

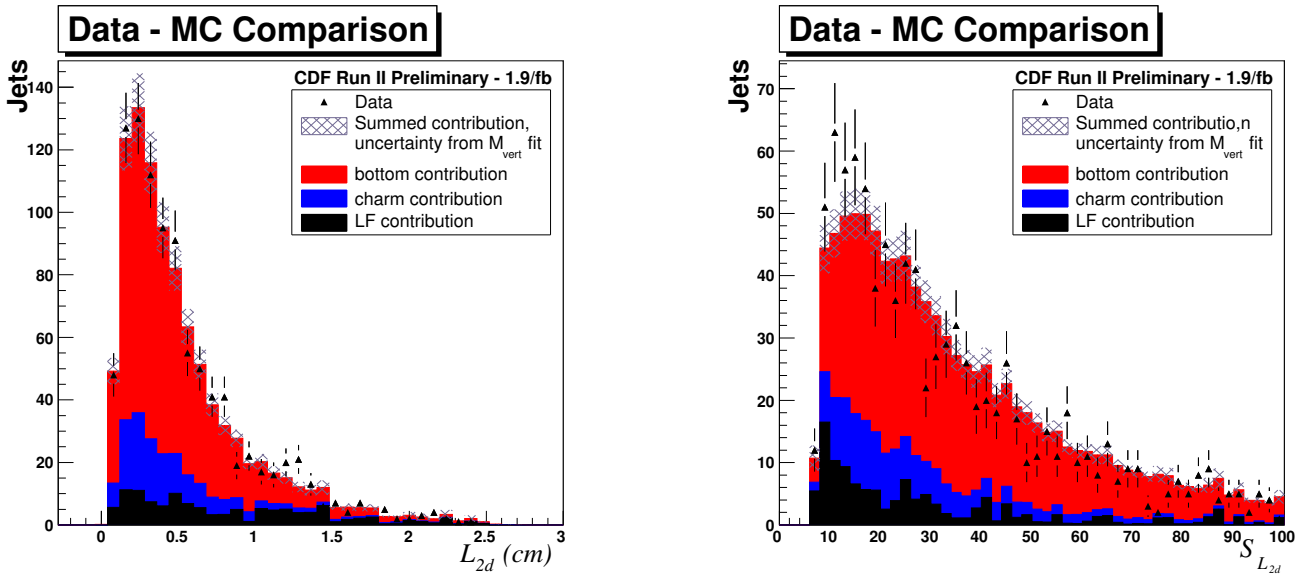

Figure 15: Comparison of L_{2d} (a) and its significance (b) for tagged vertices in data and MC. The species fractions are those found in the vertex mass fit. The MC shapes are made from the same samples from which the vertex mass templates are constructed.

References

[1] M. L. Mangano, M. Moretti and R. Pittau, Nucl. Phys. B **632**, 343 (2002) [arXiv:hep-ph/0108069].

[2] J. Campbell, R. K. Ellis, F. Maltoni and S. Willenbrock, Phys. Rev. D **75**, 054015 (2007) [arXiv:hep-ph/0611348].

[3] F. Febres Cordero, L. Reina and D. Wackeroth, Phys. Rev. D **74**, 034007 (2006) [arXiv:hep-ph/0606102].

[4] B. W. Harris, E. Laenen, L. Phaf, Z. Sullivan, S. Weinzierl, Phys.Rev. D66 (2002) 054024,arXiv:hep-ph/0207055.

[5] T. Hahn, S. Heinemeyer, F. Maltoni, G. Weiglein and S. Willenbrock, arXiv:hep-ph/0607308.

[6] CDF 3389, "Wbb and Wcc Backgrounds in Top SVX Channel".G. Unal, 1995.

[7] CDF 8768, "Calibration of Heavy-Flavor Production in QCD Data".S. Grinstein, *et al.*,2007.

[8] CDF 9125, "W′' -like Resonances in the tb Decay Channel with 1.9/fb". C. Cully. See Sec 3.1.

[9] CDF 8334, "Updated Measurement of the $W + b\bar{b}$ Cross Section".

[10] "ALPGEN, A Generator for Hard Multiparton Processes in Hadronic Collisions", M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau, A. Polosa, JHEP 0307:001,2003, hep-ph/0206293.

[11] CDF 8767, "Measurement of the Top Pair Cross Section in Lepton+Jets in 1.12/fb", Sherman, et al.

[12] Information on CDF's good run lists can be found at http://www-cdf.fnal.gov/internal/dqm/goodrun/good.html

[13] Information on the top group's Gen6 MC samples can be found at http://www-cdf.fnal.gov/internal/physics/top/RunIIMC/topmc6/index.shtml

[14] Information on the Joint Physics recommendations on lepton selection can be found at http://ncdf70.fnal.gov:8001/PerfIDia/PerfIDia.html.

[15] See the jet energy corrections website: http://www-cdf.fnal.gov/internal/physics/top/jets/corrections.html.

[16] CDF 7932, "Ultratight SECVTX Tagging: Efficiency Studies"

[17] SpartyJet home page: http://www.pa.msu.edu/ huston/SpartyJet/SpartyJet.html.

[18] See http://wwwasdoc.web.cern.ch/wwwasdoc/minuit/minmain.html.

[19] See http://www-cdf.fnal.gov/internal/physics/joint_physics/instructions/JPScaleFactor/.

[20] "Production of a W boson and two jets with one b-quark tag", J. Campbell, R.K. Ellis, F. Maltoni, S. Willenbrock, Phys. Rev. D75:054015 (2007), hep-ph/0611348.

[21] See pages 34-35 of my talk at the 12/7/2007 Higgs Discovery Group meeting.

[22] Internal communication. Improved source needed. Official source exists, just need to cite it exactly.