CDF/PHYS/TOP/PUBLIC/8410

Measurement of the b-jet Cross-Section for $W^{\pm}+ b\bar{b}$ Production

The CDF Collaboration
URL http://www-cdf.fnal.gov
(Dated: July 24, 2006)

We present the results of a measurement of the $W^{\pm}+ b\bar{b}$ b-jet cross-section at the CDF experiment. The cross-section is defined to be proportional to the number of b-jets from $W^{\pm}+ b\bar{b}$ events with one or two jets, and a leptonically decaying $W^{\pm}$ with decay products passing kinematics cuts ($p_T(\ell^{\pm})$ $\geq$20.0 GeV, $|\eta(\ell^{\pm})|$ $\leq$1.1, $p_T(\nu)$ $\geq$25.0 GeV). The cross-section is measured to be 0.90$\pm$0.20(stat.)$\pm$0.26(syst.)pb.

I. INTRODUCTION

This note describes a measurement of the b-jet cross section for $W^{\pm}$+$b\bar{b}$ events produced in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV with the CDF detector at the Fermilab Tevatron. $W^{\pm}$+$b\bar{b}$ refers to the production of a $W^{\pm}$ boson simultaneously with the production of a pair of $b\bar{b}$ quarks that predominantly come from gluon splitting, as shown in Figure 1. The measurement uses secondary vertex tagging for b-jet identification in order to enrich the data sample in events where $W^{\pm}$+$b\bar{b}$ was produced. This measurement provides a test of the QCD calculations of the $W^{\pm}$+$b\bar{b}$ cross section [1], and also allows for a new determination of this process which is of much importance as a background source in Standard Model Higgs and Single-Top analyses.

The CDF detector is described in detail in [2].

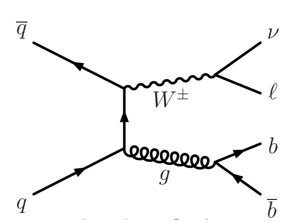

FIG. 1: Leading-Order (LO) Feynman diagram for $W^{\pm}+ b\bar{b}$ production.

II. DATA SAMPLE & EVENT SELECTION

This analysis is based on an integrated luminosity of 695 pb^{-1} collected with the CDFII detector between March 2002 and May 2005. The data are collected with an inclusive lepton trigger that requires an electron with $E_T > 18$ GeV or a muon with $p_T > 18$ GeV/c. From this inclusive lepton dataset we select events offline with a reconstructed isolated electron E_T (muon P_T) greater than 20 GeV, missing transverse energy ($\not\!\!E_T$) ≥ 25 GeV and 1 or 2 calorimeter jets with $E_T > 20$ GeV and within $|\eta| \leq 2.0$ [3]. The calorimeter jets (*i.e.*- formed from calorimeter towers using a cone-algorithm with radius=0.4) have had their energies corrected for several effects such that their energy should be equal, on average, to the energy of the underlying hadrons that formed them[4].

The dataset selected above, called "lepton+jets", is dominated by the production of W bosons accompanied by multiple jets that typically originate from light-flavor (u, d, s) quarks and gluons. To improve the signal to background we identify events with one or more b-jets (*i.e.* - jets originating from the production of b quarks) by searching for decays of B hadrons, identifiable due to their significant lifetimes, inside jets. The technique we use to identify these long-liftetime decays is called secondary vertex tagging, or SECVTX.

The identification of jets containing the decays of B hadrons (b-decays) relies on the fact that these hadrons have lifetimes long enough to allow them to travel several millimeters before decaying. Charged particle tracks that have an impact parameter, d_0, that is significantly removed from the primary $p\bar{p}$ interaction point are fit to a common secondary interaction point. In order for a jet to be identified as coming from a b-decay, or tagged, by this algorithm it must contain at least two tracks that are significantly displaced from the primary interaction point.

III. METHOD

The main idea in this analysis is to measure the cross-section for $W^{\pm}+b\bar{b}$ in a manner that is minimally dependent on theoretical models of this process. The leading-order (LO) predictions for $W^{\pm}$+heavy flavor processes ($W^{\pm}+b\bar{b}$, $W^{\pm}+c\bar{c}$, $W^{\pm}+c$) are thought to suffer from large uncertainties due to next-to-leading-order (NLO) effects and also because of the non-negligible masses of the quarks involved in the calculations[5]. Since the Monte Carlo (MC) predictions for these $W^{\pm}$+heavy flavor processes could have significant uncertainty in their distribution of events among the different jet multiplicity bins, our method avoids reliance on this variable to isolate the signal and instead focuses on the properties of jets that are tagged by the SECVTX algorithm.

A. SECVTX Mass Fit

Not all of the jets tagged by the SECVTX algorithm are actually due to true b-decays, due to material interactions, long-lived light-flavor decays, and resolution of the detector. The number of tagged jets in the selected data sample that are due to true b-decays is determined by fitting the SECVTX mass, which is the invariant mass of all the tracks that form the secondary vertex, of the jets with templates for different quark flavors. The SECVTX mass has excellent discriminating power between jets originating from bottom, charm, and light-flavor decays as shown in Figure 2. A negative-log likelihood fit is performed to find the value of the fraction of tagged jets due to b-jets, as shown in Figure 3.

The SECVTX mass fit indicates that 32.8±3.4%, or 415.7±43.2 jets (statistical uncertainty only), of the tagged jets in the selected data sample are due to true b-decays. To isolate the portion of these 415.7 tagged b-jets that are due to $W^{\pm}+b\bar{b}$ production, we estimate the number of tagged b-jets from a wide variety of background processes and subtract them from the total. After subtracting backgrounds the number of tagged b-jets from $W^{\pm}+b\bar{b}$ should be determined, and correcting this for efficiency and acceptance factors gives us a cross-section that can be compared to theoretical predictions.

B. $W^{\pm}+b\bar{b}$ Acceptance

The total acceptance, which is an estimate of the efficiency of our selection criteria on true $W^{\pm}$+$b\bar{b}$ events, is obtained using LO $W^{\pm}+b\bar{b}+N$ parton (N =0,1,2 partons beyond the $b\bar{b}$ pair) MC made using the Alpgen generator[6]. The acceptance is defined as the number of calorimeter b-jets from $W^{\pm}+b\bar{b}$ events passing all selection (except tagging) with 1 or 2 calorimeter jets divided by the number of hadron b-jets from $W^{\pm}+b\bar{b}$ events with 1 or 2 hadron jets before applying any selection. Hadron jets, which are included in this analysis to test the earlier statement that our corrected calorimeter jets should be equal in energy to the hadrons that produced them, are formed in MC samples by running the jet algorithm over final-state hadrons from the shower evolution programs.

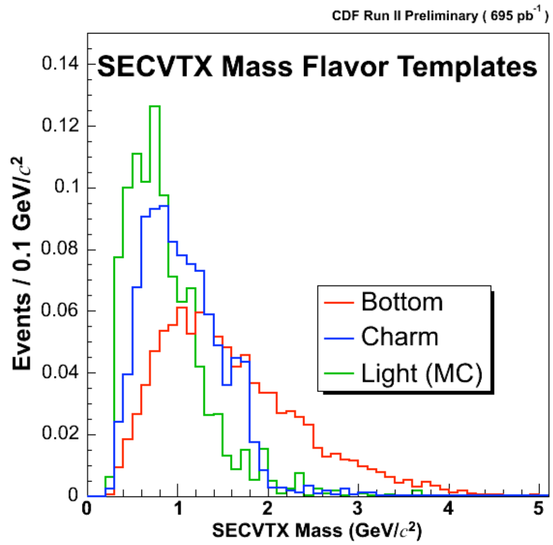

FIG. 2: Templates of SECVTX mass for tagged jets matched to different flavors. Using MC, tagged jets are matched to different flavors of hadrons using a cone of $\Delta R \leq 0.4$. If a jet is matched to both a bottom and charm hadron, it is included in the bottom jet template.

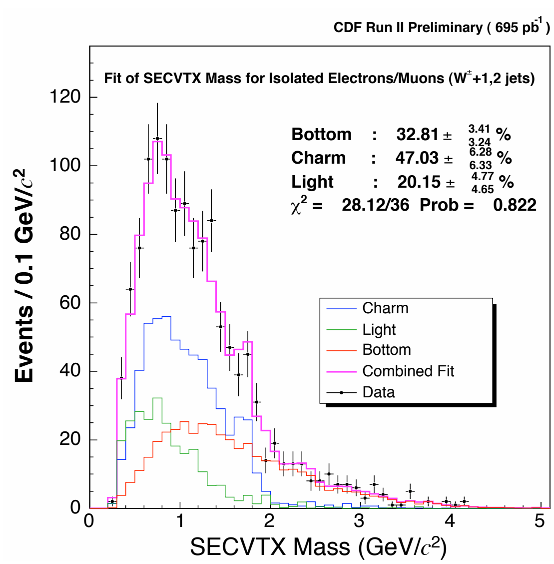

FIG. 3: Results of a negative-log likelihood fit of the templates of Figure 2 to the tagged jets in the data that passes all selection criteria. The likelihood fit indicates that 32.8±3.4% of the tagged jets (415.7±43.2 tagged jets) originate from b-decays.

We factorize the acceptance into two terms, $\mathcal{A}_{\rm jet}$ and $\mathcal{A}_{\rm selection}$, to separately examine the effects of selection and to test the claim about the equivalence of hadron jets and corrected calorimeter jets. $\mathcal{A}_{\rm jet}$ is the ratio of the number of calorimeter b-jets to the number of hadron b-jets before any selection (for events with 1 or 2 hadron/calorimeter jets), and it should be very close to unity if the claim about corrected calorimeter jets is true. $\mathcal{A}_{\rm selection}$ is the ratio of the number of calorimeter b-jets after selection to the number of calorimeter b-jets before selection, and it tells us the efficiency of the cuts we apply.

Finally, the MC used in this analysis has had a parton-jet matching scheme applied to eliminate the possibility of over-representing regions of $W^{\pm}+b\bar{b}$ phase-space. The final acceptance values for each trigger channel are derived from a weighted sum over the individual $W^{\pm}+b\bar{b}$+N parton MC samples, with the weight provided by the contribution of each sample to the overall $W^{\pm}+b\bar{b}$ generator-level cross-section. Table I summarizes the calculation of the $W^{\pm}+b\bar{b}$

acceptance.

Process	Trigger	$\mathcal{A}_{\text{jet}}$	$\mathcal{A}_{\text{selection}}$	$\mathcal{A} = \mathcal{A}_{\text{jet}} \cdot \mathcal{A}_{\text{sel.}}$	$\mathcal{A}_{\text{trig}}$
$W^{\pm}(\rightarrow e^{\pm}\nu) + b\bar{b}$+0p	CEM	0.952±0.006	0.538±0.004	0.512±0.005	0.505±0.007
$W^{\pm}(\rightarrow e^{\pm}\nu) + b\bar{b}$+1p	CEM	0.951±0.005	0.517±0.004	0.492±0.005	
$W^{\pm}(\rightarrow e^{\pm}\nu) + b\bar{b}$+2p	CEM	0.971±0.008	0.485±0.005	0.471±0.006	
$W^{\pm}(\rightarrow \mu^{\pm}\nu) + b\bar{b}$+0p	CMUP	0.953±0.006	0.320±0.003	0.305±0.003	0.301±0.004
$W^{\pm}(\rightarrow \mu^{\pm}\nu) + b\bar{b}$+1p	CMUP	0.955±0.006	0.307±0.003	0.293±0.003	
$W^{\pm}(\rightarrow \mu^{\pm}\nu) + b\bar{b}$+2p	CMUP	0.976±0.008	0.286±0.003	0.279±0.004	
$W^{\pm}(\rightarrow \mu^{\pm}\nu) + b\bar{b}$+0p	CMX	0.953±0.006	0.134±0.002	0.128±0.002	0.127±0.002
$W^{\pm}(\rightarrow \mu^{\pm}\nu) + b\bar{b}$+1p	CMX	0.954±0.006	0.130±0.002	0.124±0.002	
$W^{\pm}(\rightarrow \mu^{\pm}\nu) + b\bar{b}$+2p	CMX	0.976±0.008	0.126±0.002	0.123±0.002	

TABLE I: Acceptance values for main $W^{\pm} + b\bar{b}$ MC samples. The last column contains the final acceptance, derived from a weighted sum over the individual $W^{\pm} + b\bar{b}$+N parton samples. $\mathcal{A}_{\text{jet}}$ is very close to unity for all samples, which indicates that the corrected calorimeter jets are doing a good job of describing the energy of the underlying hadrons that were produced.

C. Efficiencies

The measured number of b-jets from $W^{\pm} + b\bar{b}$ in the selected data will be corrected by several efficiency factors that are estimated from a combination of data and MC. The trigger used to identify events with high E_T electrons has an efficiency of 0.9755±0.0055, while the trigger used to identify events with high p_T muons has an efficiency of 0.9157±0.0031 for CMUP muons and 0.9623±0.0028 for CMX muons. An efficiency of 0.9555±0.003 is applied to correct for the fraction of $p\bar{p}$ collisions that occur outside of the fiducial ($|z_{vtx}| \leq$60cm) region of the Central Outer Tracker. Possible differences in the efficiency to identify electrons and muons in MC versus in data, which would alter our estimates of $W^{\pm} + b\bar{b}$ acceptance, are accounted for by applying a factor of 0.9810±0.0030 for electrons, 0.9472±0.0043 for CMUP muons, and 1.0014±0.0178 for CMX muons.

The efficiency to tag a b-jet in $W^{\pm} + b\bar{b}$ events with 1 or 2 jets, $\epsilon_{b-\text{tag}}$, is determined in MC by calculating the fraction of b-jets that pass all selection criteria that are also tagged by the SECVTX algorithm. As in the acceptance calculation, $\epsilon_{b-\text{tag}}$ receives weighted contributions from all available $W^{\pm} + b\bar{b} + N$parton MC, yielding a final value of 0.37±0.029.

IV. BACKGROUNDS

A combination of data and MC is used to estimate the total number of tagged b-jets from sources other than $W^{\pm}+b\bar{b}$ that can pass all selection criteria.

A. QCD Background

The dominant background process is that from QCD multijet production, in which there is no real $W^{\pm}$ production in the initial hard scatter but due to a combination of jets faking leptons, mismeasured jet energy, or semileptonic b-decays these events can pass all selection criteria. A four-sector method is used to estimate the number of QCD events that pass all selection criteria and end up in our chosen data sample. The four-sector method assumes that the $\not\!\!E_T$ and lepton isolation in QCD events are uncorrelated, and divides the two-dimensional plane of these two variables into four sectors (one of which is the signal region). The number of QCD events in the three non-signal sectors are then used to estimate the number inside the signal region using the uncorrelated assumption. This procedure yields an estimate of 161.5±40.5 QCD events that pass all selection and have 1 or 2 calorimeter jets, at least one of which is tagged by the SECVTX algorithm.

To estimate the number of tagged b-jets from QCD, we fit the SECVTX mass distribution of tagged data events with $\not\!\!E_T \leq$15.0 GeV and high lepton isolation ($\geq$0.2), which should be dominated by QCD events. The b-fraction obtained for these events, 62±6% for electrons and 71±8% for muons, is applied to the respective electron and muon pieces of the tagged QCD event estimate from the four-sector method. This procedure yields 102.6±27.6 tagged b-jets from QCD events in our signal region.

B. MC-Derived Backgrounds

MC samples are used to estimate the remaining non-$W^{\pm}+b\bar{b}$ sources of tagged b-jets in the selected data sample. The MC samples have been simulated using CDF software, and the predicted number of tagged b-jets is estimated assuming the same luminosity as the data sample. Table II summarizes the non-$W^{\pm}+b\bar{b}$ processes considered in these calculations, their theoretical cross-sections, and the final estimated number of tagged b-jets each contributes to the signal region.

Process	Cross-Section (pb)	Tagged b-jets (1+2 jet events)
$W^{\pm}(\to\tau\nu)+b\bar{b}$	2.998±0.083	5.84±0.32
$W^{\pm}W^{\pm}$	12.4±0.25	0.74 ± 0.17
$W^{\pm}Z$	3.96±0.06	6.88 ± 0.32
ZZ	1.58±0.02	0.23 ± 0.04
Single-top (s-channel)	0.88±0.05	18.69 ± 0.62
Single-top (t-channel)	1.98±0.08	28.01 ± 0.67
$Z(\to\ell^{+}\ell^{-})+b\bar{b}+$	0.5376±0.0008	2.43 ± 0.16
$Z(\to\tau\tau)$	252.0±9.0	2.68 ± 1.03
$t\bar{t}$	6.7±0.7	57.22 ± 3.55
Total	-	122.72 ± 3.84

TABLE II: Summary of tagged b-jets from MC sources. Uncertainties shown are those due to MC statistics and generator-level cross-section.

V. SYSTEMATIC UNCERTAINTIES

Many sources of systematic uncertainty were investigated, the results of which are summarized in Table III.

Variable	Source	$\Delta\sigma_{W^{\pm}+b\bar{b}}(\%)$
n_{fit}	Dijet Mass Study	13.9
	Light-Flavor Model	11.1
	Double HF	7.4
	MC Statistics	2.2
n_{QCD}	MET vs. Isol.	13.5
	QCD Model	5.3
	Dijet Mass Study	3.7
	Light-Flavor Model	0.5
n_{MC}	b-tag Scale Factor	6.0
	Luminosity	3.9
	MC Statistics	2.0
$\epsilon_{b-\mathrm{tag}}$	b-tag Scale Factor	8.7
Denominator	Luminosity	6.0
	Jet Energy	5.9
	Herwig/Pythia	3.8
	PDF	3.5
	Collision Q^2	1.6
	Trig. Efficiencies	0.33
	Lepton ID	0.33
	z_0 Efficiency	0.31
Total		28.8

TABLE III: Final systematic uncertainties in the calculation of the $W^{\pm}+b\bar{b}$ cross-section.

The leading source of uncertainty is due to possible differences in the MC modeling of the SECVTX mass of b-jets with respect to the same shape for b-jets in the data. The SECVTX mass shapes of MC b-jets and jets from a double-tagged dijet data sample, which is estimated to be ≈92% pure in its b-jet content, are compared after subtracting the estimated shape of c-jets and light-flavor jets from this data. Any shape difference in this comparison is used to smear the b-jet templates used in the fit of the lepton+jets data sample, and the subsequent change in the measured b-fraction results in a 13.9% systematic uncertainty on the $W^{\pm}+b\bar{b}$ b-jet cross-section. The effect also alters the estimation of the number of tagged b-jets from QCD events, which also relies on a SECVTX mass fit, leading to an additional 3.7% systematic uncertainty on the cross-section.

The estimation of the number of tagged b-jets from the QCD process suffers from large uncertainties. The four-sector method of QCD event estimation contributes an uncertainty of 13.5%, derived by varying the boundaries used in the four-sector calculation and recording the shifts in event estimates, to the $W^{\pm}+b\bar{b}$ cross-section. The determination of the b-fraction of tagged QCD events leads to a 5.3% uncertainty, derived by fitting the SECVTX mass of different regions of the $\not{E}_T$ versus lepton isolation plane and recording the shifts in b-fraction, in the $W^{\pm}+b\bar{b}$ cross-section.

The modeling of the SECVTX mass of light-flavor jets contributes an 11.1% uncertainty which is estimated by noting the differences obtained in the result when using light-flavor jets from MC samples versus using negatively SECVTX tagged jets from the data (which are dominated by light-flavor jets). Fluctuating the fraction of our SECVTX mass templates that are due to jets containing one versus two heavy-flavor hadrons by 10%, which causes small differences in the shape, leads to a 7.4% systematic uncertainty.

Differences between the expected SECVTX tagging efficiency in Monte Carlo versus that in data leads to a 8.7% systematic uncertainty on the cross-section from the $W^{\pm}+b\bar{b}$ b-jet tagging efficiency estimate and 6.0% uncertainty on the cross-section from the estimate of the number of tagged b-jets from background processes. The luminosity recorded at CDF has a 6% uncertainty which is present in the final $W^{\pm}+b\bar{b}$ cross-section, and this effect on the estimate of background b-jet processes also produces an additional 3.9% uncertainty in the final cross-section. The cross-section denominator also contributes additional systematics due to jet-energy uncertainty (5.9%), differences obtained when using Herwig versus Pythia to shower Monte Carlo samples (3.8%), choice of generator parton distribution function (3.5%), Q^2 form in the evaluation of the initial hard-scattering process (1.6%), trigger efficiencies (0.33%), lepton identification scale factors (0.33%), and z-vertex efficiency (0.31%)[7, 8].

VI. RESULTS

We calculate the cross section as

$$\sigma_{W^{\pm}+b\bar{b}} \times \mathrm{BR}(W^{\pm} \to \ell^{\pm}\nu) = \frac{n_{b-\mathrm{jets}}(\mathrm{fit}) - n_{b-\mathrm{jets}}(\mathrm{bkgd})}{\epsilon \times \mathcal{A}_{W^{\pm}+b\bar{b}} \times \int \mathcal{L}dt} \quad (1)$$

where the denominator is summed over the different trigger leptons (CEM, CMUP, CMX) that are used in the analysis. $n_{b-\mathrm{jets}}(\mathrm{fit})$ is the number of tagged b-jets returned in the fit of the data (415.7±43.2), and $n_{b-\mathrm{jets}}(\mathrm{bkgd})$ (which receives contributions from QCD, n_{QCD}, as well as MC-derived sources, n_{MC}) is the total number of expected background b-jets (225.3±32.2). The cross-section defined in Eq. 1 is proportional to the number of b-jets in $W^{\pm}+b\bar{b}$ events with 1 or 2 jets. The final $W^{\pm}+b\bar{b}$ b-jet cross-section is found to be:

$$0.90 \pm 0.20 \pm 0.26 \text{ pb},$$

where the first uncertainty is statistical and the second is systematic. This result is found to be in good agreement with LO Alpgen predictions, which yield a value of 0.74±0.18pb.

Acknowledgments

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural Sciences and Engineering Research Council of Canada; the National Science Council of the Republic of China; the Swiss National Science Foundation; the A.P. Sloan Foundation; the Bundesministerium fuer Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Particle Physics and Astronomy Research Council and the Royal Society, UK; the Russian Foundation for Basic Research; the Comision Interministerial de Ciencia y Tecnologia, Spain; and in part by the European Community's

Human Potential Programme under contract HPRN-CT-20002, Probe for New Physics.

---

[1] M.L. Mangano, et al., Multijet Matrix Elemetns And Shower Evolution In Hadronic Collisions: $W^{\pm}+b\bar{b}+n$ Jets As A Case Study, Nucl. Phys.B **632**, 343 (2002);

[2] F. Abe, et al., Nucl. Instrum. Methods Phys. Res. A **271**, 387 (1988); D. Amidei, et al., Nucl. Instum. Methods Phys. Res. A **350**, 73 (1994); F. Abe, et al., Phys. Rev. D **52**, 4784 (1995); P. Azzi, et al., Nucl. Instrum. Methods Phys. Res. A **360**, 137 (1995); The CDFII Detector Technical Design Report, Fermilab-Pub-96/390-E

[3] An isolated lepton is one for which no more than 10% extra energy is measured in a cone of ΔR$\equiv \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} \leq$0.4 around the lepton.

[4] A. Bhatti, et al., submitted to Nucl. Instrum. Methods Phys. Res. A, hep-ex/0510047.

[5] J. Campbell and J. Huston, Phys. Rev. D **70**, 094021 (2004);

[6] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau, A. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP 0307:001,2003,hep-ph/0206293

[7] T. Sjostrand et al., High-Energy-Physics Event Generation with PYTHIA 6.1, Comput. Phys. Commun. **135**, 238 (2001).

[8] G. Corcella et al., HERWIG 6: An Event Generator for Hadron Emission Reactions with Interfering Gluons (including supersymmetric processes), JHEP **01**, 10 (2001).