Measurement of the cross section for W-boson production in association with jets in pp collisions at \( \sqrt{s} = 1.96 \) TeV


(CDF Collaboration*)

*Institute of Physics, Academia Sinica, Taipei, Taiwan 11529, Republic of China
*Argonne National Laboratory, Argonne, Illinois 60439
*Institut de Fisica d’Altes Energies, Universitat Autonoma de Barcelona, E-08193, Bellaterra (Barcelona), Spain
*Baylor University, Waco, Texas 76798
*Istituto Nazionale di Fisica Nucleare, University of Bologna, I-40127 Bologna, Italy
*Brandeis University, Waltham, Massachusetts 02254
*University of California, Davis, Davis, California 95616
*University of California, Los Angeles, Los Angeles, California 90024
*University of California, San Diego, La Jolla, California 92039
*University of California, Santa Barbara, Santa Barbara, California 93106
*Instituto de Fisica de Cantabria, CSIC-University of Cantabria, 39005 Santander, Spain
*University of Cambridge, Cambridge, United Kingdom 5213
*Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637
*Comenius University, 842 48 Bratislava, Slovakia; Institute of Experimental Physics, 040 01 Kosice, Slovakia
*Joint Institute for Nuclear Research, RU-141980 Dubna, Russia
*Duke University, Durham, North Carolina 27708
*Fermi National Accelerator Laboratory, Batavia, Illinois 60510
*University of Florida, Gainesville, Florida 32611
*Laboratori Nazionali di Frascati, Istituto Nazionale di Fisica Nucleare, I-00044 Frascati, Italy
*University of Geneva, CH-1211 Geneva 4, Switzerland
*University of Glasgow, Glasgow G12 8QQ, United Kingdom
*Harvard University, Cambridge, Massachusetts 02138
*Division of High Energy Physics, Department of Physics, University of Helsinki and Helsinki Institute of Physics, FIN-00014, Helsinki, Finland
*University of Illinois, Urbana, Illinois 61801
We present a measurement of the cross section for W-boson production in association with jets in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. The analysis uses a data sample corresponding to an integrated luminosity of 320 pb$^{-1}$ collected with the CDF II detector. W bosons are identified in their electron decay channel and jets are reconstructed using a cone algorithm. For each $W + \geq n$-jet sample ($n = 1 - 4$) we measure $\sigma(pp \rightarrow W + \geq n$-jet) $\times B(W \rightarrow e\nu)$ with respect to the transverse energy $E_T$ of the $n^{th}$-highest $E_T$ jet above 20 GeV, for a restricted $W \rightarrow e\nu$ decay phase space. The cross sections, corrected for all detector effects, can be directly compared to particle level $W +$ jet(s) predictions. We present here comparisons to leading order and next-to-leading order predictions.

PACS numbers: 14.70.Fm, 13.87.Ce, 12.38.Qk, 13.85.Ni
Final states containing a vector boson \( V \) (\( V = W, Z \)) and multiple jets (\( V + \text{jet(s)} \)) are a key signal channel for important standard model (SM) processes such as \( t \bar{t} \) or single top production, as well as a search channel for the Higgs boson and for physics beyond the SM. The production of \( V + \text{jet(s)} \) via quantum chromodynamics (QCD) presents a very large background to these processes. The ability to describe it accurately is therefore crucial, as well as being a stringent test of the power of perturbative QCD predictions. Consequently, a precise measurement of the cross section for QCD \( V \) production as a function of the transverse energy \( E_T^{\text{jet}} \) [2] of the \( n \)-th-leading jet (the highest \( E_T \) jet for \( W + \geq 1 \)-jet, the second highest \( E_T \) jet for \( W + \geq 2 \)-jet, etc.), for \( n \leq 4 \) and \( E_T^{\text{jet}} > 20 \text{ GeV} \). In order to minimize the dependence of the measurement on the modeling of the \( W \) boson production and decay kinematics, we quote cross sections defined within a limited \( W \) decay phase space: \( E_T > 20 \text{ GeV} \), \( |\eta| < 1.1 \), \( E_T^{\text{jet}} > 30 \text{ GeV} \), and \( m_W^W > 20 \text{ GeV}/c^2 \) [2]. The range of \( E_T^{\text{jet}} \) extends up to 350 GeV in the \( W + 1 \)-jet sample, a significant increase in the measured phase space compared to previous \( V + \text{jet(s)} \) measurements [3, 4]. Furthermore, the differential spectra presented here are for the first time corrected for all detector effects and represent absolute particle level cross sections [5] free, within systematic uncertainties, of any experimental bias. As such, they provide a benchmark which can be directly used for background estimates and for the validation and tuning of QCD phenomenological models. At the end of this paper we present, as examples, comparisons of our results with some of the available predictions. It is important to note that the cross section is not corrected for effects resulting from the interaction between the proton and anti-proton remnants (the "underlying event").

This analysis uses \( 320 \pm 18 \text{ pb}^{-1} \) of data collected using the CDF II [6] detector during the Tevatron Run II period. The CDF II detector is a general-purpose detector designed to study \( pp \) collisions at the Fermilab Tevatron. Inside a 1.4 T solenoidal magnetic field, a large open-cell drift chamber and an eight-layer silicon system provide precise charged-particle tracking information. Outside the solenoid electromagnetic and hadronic sampling calorimeters surround the tracking volume, allowing for the measurement of particle energies over the range \( |\eta| < 3.6 \). Finely segmented detectors located at electromagnetic shower maximum are used for electron identification. Forward gas Cerenkov detectors measure the fraction of bunch crossings that result in an inelastic \( pp \) collision and thereby determine the instantaneous luminosity delivered to the experiment.

The selection of \( W \to e\nu \) events proceeds as follows. An online trigger system selects events containing an electromagnetic calorimeter cluster with \( E_T > 18 \text{ GeV} \) associated with a high \( p_T \) track. Offline, electron candidates are required to pass standard identification cuts [7] and to have \( E_T > 20 \text{ GeV} \). The sample is enriched with events containing a neutrino by requiring that the missing transverse energy \( \not{E}_T \) [2], corrected for the calorimeter jet energy scale and the potential presence of muons [8], satisfies \( \not{E}_T > 30 \text{ GeV} \). To reduce background contamination further, the \( W \) transverse mass [2] \( m_W^W \) is required to satisfy \( m_W^W > 20 \text{ GeV}/c^2 \). In addition, \( Z \to e^+e^- \) events are rejected by a veto algorithm [7].

The jets in each \( W \to e\nu \) event are reconstructed using the \textsc{jetclu} cone algorithm [9] with cone radius \( R = \sqrt{\Delta \phi^2 + \Delta \eta^2} = 0.4 \). Starting from seed locations corresponding to calorimeter towers with \( E_T > 1 \text{ GeV} \), all nearby towers with \( E_T > 0.1 \text{ GeV} \) are used to search for stable cones. To resolve ambiguities with overlapping cones, cones sharing an energy fraction greater than 0.75 are merged into a single jet; otherwise the shared towers are assigned to the closest jet. We apply a jet energy scale (JES) correction [10] such that the measured \( E_T^{\text{jet}} \) is on average equal to the summed \( E_T \) of the particles within the jet cone that are the result of the \( p\bar{p} \to W + X \) interaction. Jets are required to have \( E_T^{\text{jet}} > 20 \text{ GeV} \) and \( |\eta| < 2.0 \).

To ensure negligible overlap of electron and jet energy deposits, events are rejected if any jet lies within \( \Delta R = 0.52 \) of the \( W \) decay electron [3]. Using this jet definition, we divide the inclusive \( W \to e\nu \) candidate events into \( W + \geq n \)-jet samples, and in each sample form the \( E_T^{\text{jet}} \) spectrum of the \( n \)-th \( E_T \)-ordered jet.

The processes which contribute background events to our \( W \) candidate sample can be divided into two categories: "leptonic" and "multi-jet". The leptonic background contains real electrons and/or neutrinos from boson decay and includes \( W \to \tau\nu \), \( Z \to e^+e^- \), \( WW \), \( W\gamma \), and top pair production. The multi-jet background arises from QCD interaction events in which one or more jets are incorrectly reconstructed in the detector as electrons and have mis-measured energy, resulting in large event \( \not{E}_T \). Background estimation proceeds as follows. For each \( W + \geq n \)-jet sample, a background-enriched event sample is constructed by removing the \( \not{E}_T > 30 \text{ GeV} \) requirement in the \( W \) selection. Multi-jet, leptonic, and signal \( \not{E}_T \) histograms are then fit to the data in the range [0,100] GeV. The measured \( \not{E}_T \) spectrum, and the result of the fit, are shown in Fig. 1 for the \( W + \geq 1 \)-jet sample. For this fit the leptonic background and signal processes are modeled by applying the \( W \) event selection minus the \( \not{E}_T > 30 \text{ GeV} \) requirement to detector simul-
lated Monte Carlo event samples of these processes. The multi-jet background is modeled using an event sample selected from the same 320 pb$^{-1}$ analysis dataset by requiring that at least two of the electron identification criteria fail. Kinematic cuts are unchanged, resulting in a background-dominated sample that accurately reflects the kinematic distributions of the multi-jet background events in the signal sample. It is necessary to correct this multi-jet sample for $\sim 5\%$ contamination from signal and leptonic background events, estimated by applying the multi-jet selection criteria to the Monte Carlo simulations of these processes. In the fit, only the normalizations of the multi-jet and signal $E_T$ histograms are allowed to float. The normalization of the $W \rightarrow \tau \nu$ and $Z \rightarrow e^+ e^-$ histograms relative to the signal is fixed by the well-established relationships between these cross sections [11]. The normalizations of the $WW$, $W \gamma$, and top pair production histograms are determined using the recently measured cross sections for these processes [12]. Once the fit is performed, the background fractions in each $W+ \geq n$-jet sample are obtained by integrating the respective histograms, with their fitted normalization, above the $E_T$ cut of 30 GeV. These fractions are then used to normalize the jet $E_T$ distributions of each background model relative to the candidates to give the background correction as a function of $E_T^{jet}$. This method offers a more accurate description of the kinematics of the multi-jet background when compared with previous approaches [3]. Fig. 1 demonstrates a successful modeling of the $E_T$ spectrum. Similarly good agreement was found in the electron $E_T$ and $m_T^W$ distributions across all jet multiplicities.

Rarely, reconstructed jets may originate from separate $p\bar{p}$ interactions in the same bunch crossing. To account for this effect, corrections extracted from the data are applied to the measured cross sections. The corrections are less than 2% in the lowest $E_T^{jet}$ bins and decrease rapidly with increasing $E_T^{jet}$.

The total background fraction increases with increasing jet multiplicity and transverse energy. At low $E_T^{jet}$ it is 10% (40%) in the 1-jet (4-jet) sample, rising to 90% at the highest $E_T^{jet}$ for all jet multiplicities. Multi-jet events contribute $\sim 70\%$ of the overall background in the 1-jet sample. At high jet multiplicities and high $E_T^{jet}$ the contribution from top pair production becomes increasingly important, climbing to 50% (80%) of the total background in the 2-jet (3, 4-jet) sample. The systematic uncertainty on the background estimate is 15% at low $E_T^{jet}$ independent of the jet multiplicity, rising to 50% (20%) at the highest $E_T^{jet}$ in the 1-jet (4-jet) sample. At low jet multiplicities, this is dominated by the limited statistics of the multi-jet background sample. At high jet multiplicities, the 12% uncertainty on the measured top pair production cross section dominates the total systematic.

We have used fully simulated signal Monte Carlo samples to correct the event yield for the efficiency of the $W \rightarrow e\nu$ selection criteria. Samples for each jet multiplicity $n$ were obtained using the ALPGEN v1.3 [13] event generator for the $W+n$-parton final state, and the PYTHIA v6.3 [14] Monte Carlo program for the parton shower and hadronization. PYTHIA includes an underlying event model, hereafter referred to as Tune A, which has been tuned to describe Tevatron data [15]. The efficiency correction factor is defined as the ratio of two subsets of the generated signal events: in the numerator, the number of reconstructed events which pass the $W \rightarrow e\nu$ selection criteria, and in the denominator, the number of generator-level events which pass the electron, neutrino and transverse-mass cuts corresponding to our cross section definition. This is found to be $(60 \pm 3)\%$, independent of event jet multiplicity and kinematics within the quoted uncertainty. By comparing $Z \rightarrow e^+ e^-$ measured and simulated event samples, we confirm that electron identification efficiencies are well reproduced by our Monte Carlo across all measured phase space to within 5%. This uncertainty also covers the observed variation in efficiency obtained by changing the number of final state partons in the ALPGEN matrix element (ME) calculation and the parton showering program from PYTHIA to HERWIG v6.5 [16]. Additionally, the $W$ candidate event yield must be corrected to account for the efficiency of the online trigger to accept high $E_T$ electrons. This is independent of jet kinematics and found to be 96.2$\pm$0.6% [7].

A further correction to the event yield is required to form particle level $W \rightarrow e\nu+\geq n$-jet cross sections as a function of $E_T^{jet}$. This correction factor accounts for the effect of calorimeter jet energy resolution on the measured cross section and is determined as follows. Using the ALPGEN-PYTHIA simulated signal samples, two cross sections in each $E_T^{jet}$ bin are determined: one defined by clustering generator-level particles into jets, the other by clustering after detector reconstruction, using the same calorimeter level jet definition (including JES correction) as the one used in the data. The correction factor for each bin is then defined as the ratio of the particle to calorimeter level cross section in that bin. To avoid dependence of such a correction on the assumed particle level $E_T^{jet}$ distribution, an iterative procedure is used to reweight the events at the particle level until the particle level $E_T^{jet}$ distribution agrees with the corresponding

![FIG. 1: The results of fitting the signal and background $E_T$ template distributions to the data in the $W+\geq 1$-jet sample before the final $E_T$ cut is applied.](image-url)
data-unfolded distribution to within the systematic uncertainties of the measurement. The correction factors vary between 0.95 and 1.2 over the measured range of \( E_T \).

The total systematic uncertainty on the cross section introduced by the jet energy measurement ranges between 5% and 20%, increasing with increasing \( E_T \). This is dominated by the approximately 3% [10] uncertainty on the JES correction. The effect of this uncertainty on the cross section is estimated by applying this variation to an ALPGEN-PYTHIA simulated signal sample reweighted to match the data. The sensitivity of the measurement to jet energy resolution uncertainties, estimated by varying the calorimeter resolution in the simulation, is much smaller by comparison.

The measured differential cross sections \( d\sigma(W \rightarrow e\nu + \sum n - \text{jet})/dE_T^{\text{th}-\text{jet}} \) are listed in Table I. For each inclusive jet multiplicity sample the cross sections are given with respect to the \( E_T \) of the \( n^{\text{th}}\)-leading jet in the sample, \( E_T^{\text{th}-\text{jet}} \). The quoted statistical uncertainties are on the event yield in each bin. The systematic uncertainties are the sum in quadrature of the effects introduced by the uncertainty on the background estimation, acceptance correction, and jet energy measurement. A 5.8% uncertainty in the integrated luminosity is not included. In summary, the total systematic uncertainty on the measured cross sections is < 20% at low \( E_T^{\text{jet}} \) increasing to 50% – 80% at high \( E_T^{\text{jet}} \). At low \( E_T^{\text{jet}} \) this is dominated by the uncertainty on the jet energy scale, whereas at high \( E_T^{\text{jet}} \) it is dominated by the background uncertainty.

We proceed to compare the measured cross sections to some of the available theoretical predictions. Leading order (LO) perturbative QCD calculations exist for the matrix element of \( V+n \) partons, with \( n \leq 6 \). They are included in Monte Carlo event generators [13, 17, 18] where the initial and final state partons are evolved through a perturbative parton shower (PS) and eventually hadronized. Additionally, the generator may include a model of the underlying event. In this LO plus PS approach, ambiguities may arise as a result of the hard emission of gluon radiation during the parton shower evolution. For example, a \( V+n \)-parton event may be reconstructed, after the shower, with a jet multiplicity \( n \neq n \), and the question naturally arises as to whether the event should or should not be counted in the estimate of the \( n\)-jet cross section. This problem has been studied extensively in the literature, leading to the development of three merging algorithms, usually known as CKKW [19], Lonnblad’s [20], and MLM [21]. A merging algorithm ensures that a given configuration in the multi-jet phase space enters into the calculation once and only once.

We present here the first comparisons of \( W + n \)-jet(s) implementations of the CKKW and MLM schemes to data. We use an implementation of the CKKW scheme hereafter referred to as SMPR [22]. Details of the generation parameters and systematic uncertainties are given in [22] for the SMPR model and in [23] for the MLM model. The former uses MADGRAPH v4 [17] for the ME generation, PYTHIA v6.3 for the PS, and CTEQ6L1 [24] parton distribution functions (PDFs), the latter uses ALPGEN v2.12, HERWIG v6.5, and CTEQ5L [25] PDFs respectively. Following generation, the SMPR and MLM predictions are formed by clustering the final state particles into jets using the JETCLU algorithm. The uncertainties on these predictions cover variation of the renormalization scale by a factor 0.5–2. This dominates the overall uncertainty in the absolute rates [23].

In addition, we present comparisons of our data with next-to-leading order (NLO) predictions for \( W + 1 \) and \( W + 2 \) jets obtained with the MCFM [26] program. These were generated using the CTEQ6.1M PDFs [24] and a renormalization and factorization scale \( \mu = \sqrt{m_W^2 + (p_T^W)^2} \). We define an uncertainty due to the choice of \( \mu \) by generating with a lower scale, \( \mu = p_T^{\text{jet}} \), and a higher scale, \( \mu = 2 \times \sqrt{m_W^2 + (p_T^W)^2} \). Additionally, the variation due to the uncertainty on the PDFs has been computed using the Hessian method [24]. This PDF uncertainty is also broadly applicable to the SMPR and MLM predictions.

In the case of the NLO predictions, the final states are not evolved through a parton shower nor hadronized. Jets are reconstructed with a cone algorithm \( R = 0.4 \), such that two partons are merged if they are within 1.3 \( \times \) \( R \) of each other and within \( R \) of the resulting jet centroid [27]. This is still considered to be sufficient to give a reasonable description of the perturbative structure of the jet [28]. However, before comparing with data, the non-perturbative effects of hadronization and the underlying event have to be considered. These two effects are of opposite sign, leading to some cancellation. We have observed, using PYTHIA TUNE A, that below 50 GeV the hadronization effect dominates, leading to an overall decrease of the cross section with respect to the parton level that is within 10%. At higher \( E_T^{\text{jet}} \), the correction is driven by the underlying event leading to an increase of at most 5%. A detailed study of this correction is outside the scope of this paper, and we do not apply any such corrections to the MCFM predictions.

The upper plot of Fig. 2 shows, as a function of the jet multiplicity \( n \), the ratio of data to theory for the cumulative cross sections \( \sigma_n = \sigma(W \rightarrow e\nu + \sum n - \text{jet}, E_T^{\text{th}-\text{jet}} > 25 \text{ GeV}) \). The lower plot shows the ratio \( \sigma_n/\sigma_{n-1} \). In Fig. 3 the ratios of the measured differential cross sections \( d\sigma(W \rightarrow e\nu + \sum n - \text{jets})/dE_T^{\text{th}-\text{jet}} \) to the predictions are shown for \( n = 1-3 \). The difference observed in Fig. 2 between the measured cross sections and SMPR or MLM predictions reflects the LO nature of these calculations. All the predictions show good agreement with the data in the cross section ratios \( \sigma_n/\sigma_{n-1} \). Fig. 3 shows that the variation in the \( W + n \)-jet cross section as a function of \( E_T^{\text{jet}} \) is better reproduced by the SMPR prediction than by the MLM. A possible explanation is the absence of a tuned underlying event model.
FIG. 2: Top: the ratio of data to theory for the cumulative cross sections $\sigma_n = \sigma(W \rightarrow e\nu/\bar{e}\bar{\nu} \geq n$-jet; $E_T^{n=1-st-jet} > 25 \text{ GeV}$) as a function of the jet multiplicity $n$. Bottom: $\sigma_n/\sigma_{n-1}$ for data, MLM, SMPR and MCFM. Inner (outer) error bars denote the statistical (total) uncertainties on the measured cross sections.

FIG. 3: Ratio of the measured cross sections $d\sigma(W \rightarrow e\nu/\bar{e}\bar{\nu} \geq n$-jet)/$dE_{T,m=1-st-jet}$ to the MLM, SMPR and MCFM predictions. Inner (outer) error bars denote the statistical (total) uncertainties on the measured cross sections.

to measure the differential $\sigma(W \rightarrow e\nu/\bar{e}\bar{\nu} \geq n$-jet) as a function of the transverse energy of the $n$-th leading jet for $n = 1 - 4$. The cross sections, corrected for all detector effects, can be directly compared to the particle level predictions of $W + 1-4$ Monte Carlo generators.

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. We also thank J. Campbell, M. Mangano, and S. Mrenna. 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 für Bil-

TABLE 1: The measured cross section $d\sigma(W \rightarrow e\nu/\bar{e}\bar{\nu} \geq n$-jet)/$dE_{T,m=1-st-jet}$ for each $E_{T,m=1-st-jet}$-bin $(n = 1 - 4)$, with statistical (first) and systematic (second) uncertainties. An overall 5.8% uncertainty on the integrated luminosity is not included.

<table>
<thead>
<tr>
<th>$E_{T,m=1-st-jet}$ (GeV)</th>
<th>$d\sigma/dE_{T,m=1-st-jet}$ (pb/GeV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20–25</td>
<td>0.46 ± 0.07 ± 0.29</td>
</tr>
<tr>
<td>25–30</td>
<td>2.80 ± 0.06 ± 0.21</td>
</tr>
<tr>
<td>30–35</td>
<td>1.92 ± 0.05 ± 0.17</td>
</tr>
<tr>
<td>35–40</td>
<td>1.31 ± 0.04 ± 0.14</td>
</tr>
<tr>
<td>40–50</td>
<td>0.839 ± 0.023 ± 0.093</td>
</tr>
<tr>
<td>50–60</td>
<td>0.498 ± 0.018 ± 0.063</td>
</tr>
<tr>
<td>60–75</td>
<td>0.259 ± 0.011 ± 0.038</td>
</tr>
<tr>
<td>75–90</td>
<td>0.158 ± 0.008 ± 0.024</td>
</tr>
<tr>
<td>90–110</td>
<td>0.056 ± 0.005 ± 0.011</td>
</tr>
<tr>
<td>110–150</td>
<td>0.0225 ± 0.0023 ± 0.0044</td>
</tr>
<tr>
<td>150–195</td>
<td>0.0035 ± 0.0011 ± 0.0023</td>
</tr>
<tr>
<td>195–350</td>
<td>0.00039 ± 0.00022 ± 0.00038</td>
</tr>
</tbody>
</table>

in the HERWIG component of the MLM prediction. We observe good agreement between the MCFM predictions and data in both cumulative and differential cross section comparisons.

In summary, we have used 320 pb$^{-1}$ of CDF II data
dung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Science and Technology Facilities Council and the Royal Society, UK; the Institut National de Physique Nucléaire et Physique des Particules/CNRS; the Russian Foundation for Basic Research; the Comisión Interministerial de Ciencia y Tecnología, Spain; the European Community’s Human Potential Programme; the Slovak R&D Agency; and the Academy of Finland.


[2] We use a cylindrical coordinate system along the proton direction in which $\theta$ ($\phi$) is the polar (azimuthal) angle. We define $\eta = -\ln(\tan(\theta/2))$, $p_T = p \sin \theta$, $E_T = E \sin \theta$, $p_T = -\sum E_i^T \hat{n}_i$ where $\hat{n}_i$ is a unit vector in the azimuthal plane that points from the beamline to the $i^{th}$ calorimeter tower, and $m_W^T = \sqrt{2E_T^\text{ele}} E^T(1 - \cos \Delta \phi (E^T, E^T))$.


