Invariant Mass Distribution of Jet Pairs Produced in Association with a W boson in pp Collisions at √s = 1.96 TeV

We report a study of the invariant mass distribution of jet pairs produced in association with a $W$ boson using data collected with the CDF detector which correspond to an integrated luminosity of 4.3 fb$^{-1}$. The observed distribution has an excess in the 120-160 GeV/c$^2$ mass range which is not described by current theoretical predictions within the statistical and systematic uncertainties.

In this letter we report studies of the properties of this excess.

PACS numbers: 12.15.Ji, 12.38.Qk, 14.80.-j

Measurements of associated production of a $W$ boson and jets are fundamental probes of the electroweak sector of the standard model (SM) and are an essential starting point for searches for physics beyond the SM. Several important processes share this signature, such as diboson production, associated production of a $W$ and a light Higgs boson and searches for new phenomena [1, 2]. The CDF collaboration at the Fermilab Tevatron collider recently measured the cross section for the production of either an additional $W$ or a $Z$ boson in association to...
a $W$ ($WW$ or $WZ$ diboson production) from a lepton plus jets final state described in Ref. [3]. One of the two methods described in that work uses the invariant mass of the two-jet system ($M_{jj}$) to extract a $WW + WZ$ signal from data. Here we perform a statistical comparison of that spectrum with expectations by including additional data and further studying the $M_{jj}$ distribution for masses higher than 100 GeV/c$^2$, with minimal changes to the event selection with respect to the previous analysis. We find a statistically significant disagreement with current theoretical predictions.

The parts of the CDF II detector [4] relevant to this analysis are briefly described here. The tracking system is composed of silicon microstrip detectors and an open-cell drift chamber inside a 1.4 T solenoid. Electromagnetic lead-scintillator and hadronic iron-scintillator sampling calorimeters segmented in a projective tower geometry surround the tracking detectors. A central calorimeter covers a pseudorapidity range $|\eta| < 1.1$, while "plug" calorimeters extend the acceptance into the region $1.1 < |\eta| < 3.6$ [5]. Outside the calorimeters are muon detectors composed of scintillators and drift chambers. Cherenkov counters around the beam pipe provide the collider luminosity measurement [6].

The trigger selection used to collect the data sample required a central and high $p_T$ electron (mmuon). Further event selection requirements are applied offline to reject backgrounds and reduce the sensitivity to systematic uncertainties. We require the presence of one electron (muon) candidate with $E_T$ ($p_T$) $> 20$ GeV (GeV/c) and $|\eta| < 1.0$ plus missing transverse energy $E_T > 25$ GeV. Both electrons and muons are required to be isolated ($Iso < 0.1$) [7] to reject leptons from semileptonic decays of heavy flavor hadrons and hadrons misidentified as leptons. Jets are clustered using a fixed-cone algorithm with radius $\Delta R = \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2} = 0.4$, and their energies are corrected for detector effects that are of the order of 25% for jet $E_T = 30$ GeV [8]. Jets with an electron or muon in a cone $\Delta R = 0.52$ around the jet axis are removed. Cosmic rays and photon-conversion candidates are removed. We require events to have exactly two jets each with $E_T > 30$ GeV and $|\eta| < 2.4$, and the dijet system to have $p_T > 40$ GeV/c.

The transverse mass $M_T(W)$ [5] of the lepton + $E_T$ system must be greater than 30 GeV/c$^2$; the two jets must be separated by $|\Delta \eta| < 2.5$. To suppress multijet background, we further require that the direction of $E_T$ and of the most energetic jet are separated azimuthally by $|\Delta \phi| > 0.4$. These requirements do not sculp the dijet invariant mass distribution at masses above 100 GeV/c$^2$.

To remove contamination from $Z$ production, we reject events where an additional lepton is found using looser criteria and the invariant mass of the two leptons is in the range 76-106 GeV/c$^2$. We further reject events with two identified leptons, where the $E_T$ ($p_T$) threshold for the second lepton is decreased to 10 GeV (GeV/c), to suppress other sources of real dileptons such as leptonic decays of both final state $W$'s in $t\bar{t}$ and dibosons with jets. The main difference with respect to the selection criteria used in Ref. [3] is that the jet $E_T$ threshold is increased from 20 GeV to 30 GeV, motivated by the interest in a higher invariant mass range. The resulting sample is dominated by events where a $W$ boson, which decays leptonically, is produced in association with jets ($W$+jets). Minor contributions to the selected sample come from $WW+WZ$, $tt$, $Z$+jets, single top production and multijet QCD sources. Predictions for these processes, with the exception of the multijet QCD component, are obtained using event generators and a GEANT-based CDF II detector simulation [9]. The diboson, $t\bar{t}$, and single top components are simulated using the PYTHIA event generator [10]. The $W$+jets and $Z$+jets processes are simulated using a matrix element Leading Order event generator ALPGEN [11] with an interface to PYTHIA providing parton showering and hadronization [12][13]. Multijet QCD events, where one jet is misidentified as a lepton, are modeled with data containing anti-isolated muons ($Iso > 0.2$) or candidate electrons failing quality cuts [13]. The normalization of the $Z$+jets component is based on the measured cross section [14], while for $t\bar{t}$, single top, and diboson production the NLO predicted cross sections are used [15]. The detection efficiencies for $Z$+jets, $t\bar{t}$, single top, and diboson contributions are determined from simulation. The normalization of the multijet QCD component and a preliminary estimation of the $W$+jets component are obtained by fitting the $E_T$ spectrum in data to the sum of all contributing processes.

We perform a combined binned $\chi^2$ fit, for electron and muon events, to the dijet invariant mass ($M_{jj}$) spectrum using predictions for the multijet QCD, $WW$, $WZ$, $Z$+jets, $W$+jets, $t\bar{t}$, and single top processes. The final $W$+jets normalization is determined by minimizing this $\chi^2$ and all other contributions are constrained to be within the variance of their expected normalization.

We fit the dijet mass distribution in the range 28-200 GeV/c$^2$ defined a priori in the measurement of the $WW/WZ$ cross section [3]. Figs. 1(a) and (b) show the extrapolation of this fit in the extended range of mass up to 300 GeV/c$^2$. The fit is stable with respect to changes in the fit range and histogram binning. Our model describes the data within uncertainties, except in the mass region $\sim 120-160$ GeV/c$^2$, where an excess over the simulation is seen. The fit $\chi^2$/ndf is 77.1/84, where ndf is the number of degrees of freedom. The $\chi^2$/ndf computed only in the region 120-160 GeV/c$^2$ is 26.1/20. However the Kolmogorov-Smirnov (KS) test, which is more sensitive to a localized excess, yields a probability of $6 \times 10^{-5}$ [16].

We try to model the excess with an additional Gaussian peak and perform a $\Delta \chi^2$ test of this hypothesis. The Gaussian is chosen as the simplest hypothesis compatible with the assumption of a two jet decay of a narrow
FIG. 1: The dijet invariant mass distribution. The sum of electron and muon events is plotted. In the left plots we show the fits for known processes only (a) and with the addition of a hypothetical Gaussian component (c). On the right plots we show, by subtraction, only the resonant contribution to $M_{jj}$ including $WW$ and $WZ$ production (b) and the hypothesized narrow Gaussian contribution (d). In plot (b) and (d) data points differ because the normalization of the background changes between the two fits. The band in the subtracted plots represents the sum of all background shape systematic uncertainties described in the text. The distributions are shown with a 8 GeV/c^2 binning while the actual fit is performed using a 4 GeV/c^2 bin size.

Resonance with definite mass. The width of the Gaussian is fixed to the expected dijet mass resolution by scaling the width of the $W$ peak in the same spectrum:

$$\sigma_{\text{resolution}} = \sigma_W \sqrt{\frac{M_{jj}}{M_W}} = 14.3 \text{ GeV}/c^2,$$

where $\sigma_W$ and $M_W$ are the resolution and the average dijet invariant mass for the hadronic $W$ in the $WW$ simulations respectively, and $M_{jj}$ is the dijet mass where the Gaussian template is centered.

In the combined fit, the normalization of the Gaussian is free to vary independently for the electron and muon samples, while the mean is constrained to be the same. The result of this alternative fit is shown in Figs. (c) and (d). The inclusion of this additional component brings the fit into good agreement with the data. The fit $\chi^2/ndf$ is 56.7/81 and the Kolmogorov-Smirnov test returns a probability of 0.05, accounting only for statistical uncertainties. The $W$+jets normalization returned by the fit including the additional Gaussian component is compatible with the preliminary estimation from the $E_T$ fit. The $\chi^2/ndf$ in the region 120-160 GeV/c^2 is 10.9/20.
TABLE I: Results of the combined fit. The ratios of the number of events in the excess to the number of expected diboson events in the electron and muon samples are statistically compatible with each other.

<table>
<thead>
<tr>
<th></th>
<th>Electrons</th>
<th>Muons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excess events</td>
<td>156 ± 42</td>
<td>97 ± 38</td>
</tr>
<tr>
<td>Excess events / expected diboson</td>
<td>0.60 ± 0.18</td>
<td>0.44 ± 0.18</td>
</tr>
<tr>
<td>Mean of the Gaussian component</td>
<td>144 ± 5 GeV/c²</td>
<td></td>
</tr>
</tbody>
</table>

The values of parameters returned by the combined fit are shown in Table I, where the mean of the Gaussian peak represents the experimentally measured value i.e. it is not corrected back to the parton-level.

We take the difference between the $\chi^2$ of the two fits ($\Delta \chi^2$), with and without the additional Gaussian structure to assess the significance of the excess. The expected distribution of $\Delta \chi^2$ is computed numerically from simulated background-only experiments and used to derive the p-value corresponding to the $\Delta \chi^2$ actually observed. In order to account for the trial factor within our search window, 120-200 GeV/c², in each pseudoexperiment we calculate the $\Delta \chi^2$ varying the position of the Gaussian component in steps of 4 GeV/c². The largest $\Delta \chi^2$ for each pseudoexperiment is used to define the p-value distribution.

In deriving the p-value we account for systematic uncertainties that affect the background shapes and the normalization of constrained components. Normalization uncertainties of unconstrained components are considered as part of the statistical uncertainty. The largest systematic uncertainties arise from the modeling of the $W$+jets and multijet QCD shapes. For $W$+jets we consider, as an alternative, the $M_{jj}$ distributions obtained by halving or doubling the renormalization scale ($Q^2$) in Alpgen. For multijet QCD, we change our model using different lepton isolation ranges. The systematic uncertainty due to uncertainties in the jet energy scale (±3%) affects all components with the exception of multijet QCD, which is derived from data. For each systematic effect we consider the two extreme cases. For each of the possible combinations of systematic effects we calculate a different $\Delta \chi^2$ distribution and take the conservative approach of using the distribution that returns the highest p-value. The total systematic effect on the extracted number of excess events, defined as the number of events fitted by the Gaussian component, in the electron and muon samples is found to be 10% and 9%, respectively. The dominant systematic effects arise from the $W$+jets renormalization scale (6.7%), the jet energy scale (6.1%) and QCD shape (1.9%). Assuming only background contributions, and systematic errors, the probability to observe an excess larger than in the data is $7.6 \times 10^{-4}$ corresponding to a significance of 3.2 standard deviations for a Gaussian distribution. For comparison, the p-value without taking into account systematic uncertainties is $9.9 \times 10^{-5}$.

To investigate possible mismodeling of the $W$+jets background we consider various configurations of our systematic uncertainties. The combination of systematic uncertainties that fits the data best is shown in Fig. (a) where $Q^2$ is doubled and the QCD shape is varied. The KS probability for this fit is 0.28. The fit $\chi^2$/ndf outside the 120-160 GeV/c² region is 50.3/66, indicating that the dijet mass distribution is well modeled within our systematic uncertainties. This choice of systematic uncertainties returns a p-value intermediate between the central configuration and the most conservative combination. In order to test “Next to Leading Order” contributions to the $W$+2 partons prediction, we compare a sample of $W$+2 partons simulated with Alpgen and interfaced to Pythia for showering to a sample of $W$+2 partons simulated using the MCFM generator. We extract a correction as a function of $M_{jj}$ that is applied to the Alpgen + Pythia sample used in our background model. The statistical significance obtained with the MCFM reweighted $W$+jets background model is 3.4σ.

Details of a large set of additional checks can be found in Ref. [13]. In particular we verified that the background model describes the data in several independent control regions and satisfactorily reproduces the kinematic distributions of jets, lepton, and $E_T$. The excess is stable against 5 GeV variations of the thresholds used for all of the kinematic selection variables, including variations of the jet $E_T > 30$ GeV threshold. This analysis employs requirements on jets of $E_T > 30$ GeV and $p_T > 40$ GeV/c for the dijet system, which improves the overall modeling of many kinematic distributions. We also test a selection only requiring jet $E_T > 20$ GeV as in Ref. [19]. This selection, which increases the background by a factor of 4, reduces the statistical significance of the excess to about 1σ.

We study the $\Delta R_{jj}$ distribution to investigate possible effects that could result in a mismodeling of the dijet invariant mass distribution. We consider two control regions, the first defined by events with $M_{jj} < 115$ and $M_{jj} > 175$ GeV/c² and the second defined by events with $p_T < 40$ GeV/c. We use these regions to derive a correction as a function of $\Delta R_{jj}$ to reweight the events in the excess region. We find that the reweightings change the statistical significance of the result by plus or minus one sigma. However, the $\Delta R_{jj}$ distribution is strongly correlated to $M_{jj}$ and the control regions both have significantly different distributions of $\Delta R_{jj}$. Reweighting our $W$+jets sample to correct for the differences observed in $\Delta R_{jj}$ in the control samples may be indicative of the effect of correcting $\Delta R_{jj}$ mismodeling or may introduce bias in the $M_{jj}$ distribution. In addition, the $\Delta R_{jj}$ distribution is consistent within the
one sigma variation of the systematic uncertainties for events outside the excess mass region as shown in Fig. 2 (b). The data-background comparison of the $\Delta R_{jj}$ distribution has $\chi^2$/ndf of 26.7/18 and a KS probability of 0.022 when compared with best-fit systematic model. For these reasons, we present these studies as cross checks and quote the significance in the unweighted sample as our primary result.

We look for evidence in favor or against the hypothesis that the excess in the 120-160 GeV/$c^2$ mass range is from a new (non-SM) physics source. Since non-SM particles may in general couple to both massive electroweak gauge bosons we have investigated the shape of the dijet mass distribution in $Z+\text{jets}$ events. No statistically significant deviation from the SM expectations in this sample is observed. We increase the jet $E_T$ threshold in steps of 5 GeV and check the fraction of excess events that are selected as a function of the jet $E_T$. The result is compatible with expectation from a Monte Carlo simulation of a $W$ boson plus a particle with a mass of 150 GeV/$c^2$ and decaying into two jets [19]. In this model, we estimate a cross section times the particle branching ratio into dijets of the order of 4 pb. The cross section of the observed excess is not compatible with SM $WH$ production whose $\sigma \cdot BR(H \rightarrow b\bar{b})$ is about 12 fb for $m_H = 150$ GeV/$c^2$ [20]. To check the flavor content with this selection, we identify jets originating from a $b$-quark by requesting a displaced secondary vertex for tracks within the jet cone. We compare the fraction of events with at least one $b$-jet in the excess region (120-160 GeV/$c^2$) to that in the sideband regions (100-120 and 160-180 GeV/$c^2$), and find them to be compatible with each other. Dedicated CDF searches for $WH \rightarrow l\nu b\bar{b}$ using events with reconstructed displaced vertices from $b$ hadron decay, and looser selection criteria, have not found any significant excesses using final analysis discriminants trained to identify Higgs bosons in the mass range 100-150 GeV/$c^2$ [19].

Finally, to investigate the possibilities of a parent resonance or other quasi-resonant behavior, we consider the $M_{\text{lepton},\nu, jj}$ and the $M_{\text{lepton},\nu, jj} - M_{jj}$ distributions for events with $M_{jj}$ in the range 120-160 GeV/$c^2$ and, to investigate the Dalitz structure of the excess events, the distribution of $M_{\text{lepton},\nu, jj} - M_{jj}$, in bins of $M_{jj}$. The distributions are compatible in shape with the background-only hypothesis in all cases.

In conclusion, we study the invariant mass distribution of jet pairs produced in association with a $W$ boson. The best fit to the observed dijet mass distribution using known components, and modeling the dominant $W+\text{jets}$ background using ALPGEN+PYTHIA Monte Carlo, shows a statistically significant disagreement. One possible way to interpret this disagreement is as an excess in the 120-160 GeV/$c^2$ mass range. If we model the excess as a Gaussian component with a width compatible with the dijet invariant mass resolution, and perform a $\Delta \chi^2$ test for the presence of this additional component, we obtain a p-value of $7.6 \times 10^{-3}$, corresponding to a significance of 3.2 standard deviations, after accounting for all statistical and systematic uncertainties.

We thank the Fermilab theory group for helpful suggestions, particularly J. Campbell, E. Eichten, R. K. Ellis, C. Hill, and A. Martin. We are grateful to K. Lane.
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 Research Council and the Royal Society, UK; the Institut National de Physique Nucleaire et Physique des Particules/CNRS; the Russian Foundation for Basic Research; the Korean World Class University Program, the National Research Foundation of Korea; the Science and Technology Facilities Council and the Royal Society, UK; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Science and Technology of China; the Swiss National Science Foundation of Canada; the National Science Council of Japan; the Natural Sciences and Engineering Research Council of Canada; the National Research Council and the Royal Society, UK; the Institut National de Physique Nucleaire et Physique des Particules/CNRS; the Russian Foundation for Basic Research; the Korean World Class University Program, the National Research Foundation of Korea; the Science and Technology Facilities Council and the Royal Society, UK; the Institut National de Physique Nucleaire et Physique des Particules/CNRS; the Russian Foundation for Basic Research; the Korean World Class University Program, the National Research Foundation of Korea; the Science and Technology Facilities Council and the Royal Society, UK; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Science and Technology of China; the Swiss National Science Foundation; the KfW (German Federal Ministry of Education and Research); the Spanish Ministry of Science and Innovation; the Slovak R&D Agency; the Ministerio de Ciencia e Innovación, and Programa Consolider-Ingenio 2010, Spain; the Slovak R&D Agency; and the Academy of Finland.

[5] We use a cylindrical coordinate system with its origin in the center of the detector, where $\theta$ and $\phi$ are the polar and azimuthal angles, respectively, and pseudo-rapidity is $\eta = -\ln \tan(\theta/2)$. The transverse energy $E_T$ (momentum $p_T$) is defined as $E \sin \theta$ ($p \sin \theta$). The missing $E_T$ ($\vec{E}_T$) is defined by $\vec{E}_T = -\sum_i E_{T,i} \hat{n}_i$, where $\hat{n}_i$ is a unit vector perpendicular to the beam axis and pointing at the $i^{th}$ calorimeter tower. $\vec{E}_T$ is corrected for high-energy muons and also jet energy corrections. We define $E_T = |\vec{E}_T|$. The transverse mass of the $W$ is defined as $M_T(W) = \sqrt{2p_T(E_T(1 - \cos(\Delta \phi^\nu)))}$.
[7] Lepton isolation (ISO) is defined as $\sum_{E_T}^E_{\mu}p_T$ for electrons and muons respectively, where $\sum E_T$ is the calorimetric energy in a cone 0.4 around the lepton.
[15] The reported KS probability corresponds to the KS test between data and background distributions. It does not account for the fact that the background distributions are constrained by fits to data. The reported values are thus an upper limit on KS probability.
[20] $M_{[\text{lepton}, v, jj]}$ denotes the total invariant mass of the lepton, neutrino and dijet system. We reconstruct the longitudinal component of the neutrino momentum by imposing a $W$ mass of 80.398 GeV/$c^2$. We consider both real solutions for the $p_x$ of the neutrino and we discard complex solutions of the $W$ mass equation.