

# Invariant Mass Distribution of Jet Pairs Produced in Association with a $W/Z$ boson in $p\bar{p}$ Collisions at CDF

The CDF Collaboration<sup>1</sup>

<sup>1</sup><http://www-cdf.fnal.gov>

## Abstract

We present a study of the invariant mass spectra of jets in events with one identified lepton and missing transverse energy at the CDF II experiment at the Fermilab Tevatron. Events of this signature indicate the interface between electroweak and strong processes, and are critical to studies of vector boson production, top-quark physics, Higgs boson physics, and searches for beyond-the-standard-model particles. We present detailed analyses of several control regions which are used to calibrate detector response and models of instrumental background. We present a search for high-mass resonances decaying into jets, and find no significant excess above the standard-model-background prediction.

PACS numbers:

## I. INTRODUCTION

At hadron colliders, the production of jet pairs in association with vector bosons offers measurements of fundamental standard model (SM) parameters and tests of theoretical predictions. Many beyond-the-standard-model (BSM) scenarios also predict significant deviations from the SM in these signatures [1, 2, 3]. In a previous publication, the CDF collaboration reported a disagreement between data and the SM prediction using a data sample corresponding to  $4.3 \text{ fb}^{-1}$  of data [4]. Assuming an excess of events over the background prediction appearing as a narrow Gaussian distribution, the statistical significance of the reported disagreement was 3.2 standard deviations. In this current document, we report an update to the previous analysis using the full CDF Run II data set, more than doubling the candidate event sample size. In addition to the larger data set, we present the investigation of a series of systematic uncertainties which were not conclusive using the prior, smaller, data sample, or were previously unconsidered.

As a result of these studies, new calibrations of the detector response and instrumental backgrounds is performed, yielding far better agreement with the standard model prediction from Monte Carlo (MC) event generators. Using this improved model of the detector and the background processes, we perform measurements of diboson production cross sections in the semileptonic decay modes ( $WW \rightarrow l\nu jj$ ,  $WZ \rightarrow l\nu jj$ ,  $WZ/ZZ \rightarrow l^+l^-jj$ , and  $WZ/ZZ \rightarrow \nu\nu jj$ .) Good agreement with the predictions of the SM via next-to-leading-order (NLO) predictions is found. We also present searches for the excess of events over the background prediction described in Ref. [4]. No significant excess is observed, and a detailed study of the affect of the new calibrations on the dijet mass spectra is presented.

We will start with a description of the CDF apparatus and the event selection (Sec. II). The following two sections show two sets of corrections that greatly improve the agreement between data and predictions (Sec. III and Sec. IV). The fitting method used in the analysis is described next in Sec. V. The systematic uncertainties considered are discussed in Sec. VI. The results, before and after applying the new corrections are shown in Sec. VII and Sec. VIII. A quick summary of studies performed in orthogonal samples are presented in Sec. IX and we conclude in Sec. X.

## II. SELECTION

The CDF II detector is described in detail elsewhere [5]. The detector is cylindrically symmetric around the proton beam axis which is oriented in the positive  $z$  direction. The polar angle,  $\theta$ , is measured from the origin of the coordinate system at the center of the detector with respect to the  $z$  axis. Pseudorapidity, transverse energy, and transverse momentum are defined as  $\eta = -\ln \tan(\theta/2)$ ,  $E_T = E \sin \theta$ , and  $p_T = p \sin \theta$ , respectively. The central and plug calorimeters, which respectively cover the pseudorapidity regions of  $|\eta| < 1.1$  and  $1.1 < |\eta| < 3.6$ , surround the tracking system with a projective tower geometry. The detector has a charged particle tracking system immersed in a 1.4 T magnetic field, aligned coaxially with the  $p\bar{p}$  beams. A silicon microstrip detector provides tracking over the radial range 1.5 to 28 cm. A 3.1 m long open-cell drift chamber, the central outer tracker (COT), covers the radial range from 40 to 137 cm and provides up to 96 measurements with alternating axial and  $\pm 2^\circ$  stereo superlayers. The fiducial region of the silicon detector extends to  $|\eta| \sim 2$ , while the COT provides coverage for  $|\eta| \lesssim 1$ . Muons are detected up to  $|\eta| < 1.0$  by drift chambers located outside the hadronic calorimeters.

The transverse momentum resolution is measured to be  $\delta p_T/p_T \approx 0.07\% \cdot p_T$  (GeV/c) for the combined tracking system. The transverse energy is measured in each calorimeter tower where the polar angle ( $\theta$ ) is calculated using the measured  $z$  position of the event vertex and the tower location. Contiguous groups of calorimeter towers with signals are identified and summed together into an energy cluster. Electron candidates are identified in the central electromagnetic calorimeter (TCE) as isolated, mostly electromagnetic, clusters that match a reconstructed silicon track in the pseudorapidity range  $|\eta| < 1.1$ . The electron transverse energy is reconstructed from the electromagnetic cluster with a precision  $\sigma(E_T)/E_T \approx 13.5\%/\sqrt{E(\text{GeV})} \oplus 2\%$ . Jets are identified as a group of electromagnetic calorimeter energy and hadronic calorimeter energy clusters populating a cone of radius  $\Delta R \approx \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} \leq 0.4$  units around a high- $E_T$  seed cluster. Jet energies are corrected for calorimeter nonlinearity, losses in the gaps between towers, and multiple primary interactions. The jet energy resolution is approximately  $\sigma(E_T) \approx [0.1E_T + 1.0\text{GeV}]$ .

Muon candidates are detected in three separate subdetectors. After at least five interaction lengths in the calorimeter, central muons first encounter four layers of planar drift chambers (CMU), capable of detecting muons with  $p_T > 1.4\text{GeV}/c$ . Four additional

layers of planar drift chambers (CMP) behind another 60 cm of steel detect muons with  $p_T > 2.8 \text{ GeV}/c$ . These two systems cover the same central pseudorapidity region with  $|\eta| \lesssim 0.6$ . A track that is linked to both CMU and CMP stubs is called a CMUP muon. Muons that exit the calorimeters at  $0.6 \lesssim |\eta| \lesssim 1.0$  are detected by the CMX system of four drift layers. Muon candidates are then identified as isolated tracks that extrapolate to line segments or “stubs” in the muon subdetectors.

Missing transverse energy ( $\cancel{E}_T$ ) is defined as the opposite of the vector sum of all calorimeter tower energy depositions projected on the transverse plane. It is used as a measure of the sum of the transverse momenta of the particles that escape detection, most notably neutrinos. The corrected energies are used for jets in the vector sum defining  $\cancel{E}_T$ . The muon momentum is also added for any minimum ionizing high- $p_T$  muon found in the event.

Muon and electron candidates used in this analysis are identified during data taking with the CDF trigger system, a three-level filter with tracking information available at the first level. The first stage of the central electron trigger (TCE) requires a track with  $p_T > 8 \text{ GeV}/c$  pointing to a tower with  $E_T > 8 \text{ GeV}$  and  $E_{HAD}/E_{EM} < 0.125$ . The first stage of the muon trigger requires a track with  $p_T > 4 \text{ GeV}/c$  (CMUP) or  $8 \text{ GeV}/c$  (CMX) pointing to a muon stub. A complete lepton reconstruction is performed online in the final trigger stage, where we require  $E_T > 18 \text{ GeV}$  for central electrons (TCE) and  $p_T > 18 \text{ GeV}/c$  for muons (CMUP, CMX).

We select events with one and only one electron  $E_T > 20 \text{ GeV}$  or muon with  $p_T > 20 \text{ GeV}/c$  large transverse missing energy ( $\cancel{E}_T > 25 \text{ GeV}$ ) and exactly two jets with  $E_T > 30 \text{ GeV}$  and  $|\eta| < 2.4$ . In order to reject multijet backgrounds we impose the following cuts: transverse mass cut  $m_T > 30 \text{ GeV}$ , azimuthal angle between the most energetic jet and  $\cancel{E}_T$ ,  $\Delta\phi(\cancel{E}_T, j_1) > 0.4$ , difference in pseudorapidity between the two jets,  $|\Delta\eta(j_1, j_2)| < 2.5$  and the transverse momentum of the dijet system  $p_T^{jj} > 40 \text{ GeV}/c$ . The event’s primary vertex is calculated by fitting a subset of well-measured tracks coming from the beam line and is required to be within 60 cm of the center of the CDF II detector. The longitudinal coordinate  $z_0$  of the lepton track at point of closest approach to the beam line must be within 5 cm of the primary vertex to ensure that the lepton and the jets come from the same hard interaction.

### III. JES MODELING

The jets used in this analysis have their energies, as measured by the calorimeter, corrected for a number of effects that distort the true jet energy. These effects include consistency across  $|\eta|$  and time, contributions from multiple  $p\bar{p}$  interactions per beam crossing (pileup) and the underlying event, the non-linear response of the calorimeter, and energy radiated outside of the jet cone. The jet energy scale (JES) corrections applied are described in detail in [6].

These energy corrections, however, do not distinguish between the response of gluon and quark jets. The largest energy corrections, which correct the energy scale of calorimeter jets to better match that of particle jets and the initial parton energies, are derived using PYTHIA [7] dijet Monte Carlo simulations. Differences in the response of gluon and quark jets between MC and data may lead to differences in the measured energies of these objects that are not covered by the previously assigned systematic uncertainties on the JES.

In what follows, we derive a correction for the response of quark and gluon jets in simulated events using two independent samples of jets with different quark fractions, balanced against objects of known momentum. We use events where a jet balances with a high- $E_T$  photon, which are rich in quark jets, and utilize the significant number of  $Z \rightarrow \ell^+\ell^- + \text{jet}$  events available in the full CDF dataset, which are more rich in gluon jets. We construct the balance of the jet with these better measured reference objects:

$$K_{Z/\gamma} = (E_T^{\text{jet}}/p_T^{Z/\gamma}) - 1 . \quad (1)$$

For well-measured jets,  $K_{z/\gamma} = 0$ . Rather than derive full and separate JES corrections for quark and gluon jets in data and simulation, we compare the balance in data and simulation and derive an additional correction to be applied to simulated jets, based upon whether these jets are matched to quarks or gluons.

#### A. Dataset and Event Selection

The dataset that forms our  $Z$ -jet balancing sample comes from a suite of high- $E_T$  electron and high- $p_T$  muon triggers with an integrated luminosity of  $8.9 \text{ pb}^{-1}$ . We require two leptons consistent with resulting from the decay of a  $Z$  boson, where those leptons have a more

relaxed selection criteria than that described in Sec. II. We also require events have exactly one measured jet with (uncorrected)  $E_T > 3$  GeV within  $|\eta| \leq 2.4$ . Additionally, we ensure that the  $Z$  boson and jet are back-to-back by requiring the  $\Delta\phi$  between the two objects be  $> 2.8$  radians, and require  $p_T^Z > 10$  GeV/ $c$ .

For the  $\gamma$ -jet balancing sample, we closely mirror the selection requirements described in Ref. [6]. We use events collected with an isolated central photon trigger covering the same data period as that of the high- $p_T$  lepton samples. We compare these data to PYTHIA simulations of both  $\gamma$ + jet production as well as dijet production which contaminates our  $\gamma$ -jet balancing sample.

In order to avoid trigger biases, we require  $E_T^\gamma > 27$  GeV and  $0.2 \leq |\eta_\gamma| \leq 0.6$  in both data and MC. To decrease the contribution from dijet production, where a jet mimics our photon selection, we require the energy in the calorimeter and momentum in the tracking system contained within a cone of  $R = 0.4$  around the photon to be less than 1 GeV and 2 GeV/ $c$ , respectively. As in the  $Z$ -jet balancing sample, we require events have exactly one measured jet with (uncorrected)  $E_T > 3$  GeV within  $|\eta| \leq 2.4$ , and demand the  $\Delta\phi$  between the jet and photon be  $> 3.0$  radians. We further reduce contamination of our sample by vetoing events with large pileup (more than 1 reconstructed interaction point), and by removing events with  $\cancel{E}_T/E_T^\gamma > 0.8$ , which likely contain activity from cosmic rays.

## B. Determination of Correction

We derive separate corrections for the quark and gluon jet energy scales in data and simulation using our  $Z$ -jet and  $\gamma$ -jet balancing samples in the following way. Suppose  $K_Z$  and  $K_\gamma$  are the necessary corrections on jet energies to balance the reconstructed  $Z$  or  $\gamma$  in the  $Z$ -jet and  $\gamma$ -jet balancing samples. Each correction should be a weighted average of separate corrections for quark and gluon jets:  $K_q$  and  $K_g$ , respectively. If  $F_X^{q/g}$  is the quark/gluon fraction in sample  $X$ , then we can write:

$$K_Z = F_Z^q K_q + F_Z^g K_g = F_Z^q K_q + (1 - F_Z^q) K_g \quad (2)$$

$$K_\gamma = F_\gamma^q K_q + F_\gamma^g K_g = F_\gamma^q K_q + (1 - F_\gamma^q) K_g, \quad (3)$$

or, solving for  $K_q$  and  $K_g$ :

$$K_q = \frac{1}{F_\gamma^q - F_Z^q} [(1 - F_Z^q)K_\gamma - (1 - F_\gamma^q)K_Z] \quad (4)$$

$$K_g = \frac{1}{F_\gamma^q - F_Z^q} [F_\gamma^q K_Z - F_Z^q K_\gamma] . \quad (5)$$

These equations may be written separately for data and MC simulation (thus with distinct  $K_X^{\text{data}}$  and  $K_X^{\text{MC}}$ ), and may include a dependence on the energy of the jet ( $F_X^q \rightarrow F_X^q(E_T^{\text{jet}})$  and  $K_X \rightarrow K_X(E_T^{\text{jet}})$ ).

In order to solve for  $K_q$  and  $K_g$ , we need  $K_{Z/\gamma}$  and  $F_{Z/\gamma}^q$ . For the former, we extract a singular value as a function of  $E_T^{\text{jet}}$  by constructing the balancing distribution, as defined in Eq. 1, in bins of  $E_T^{\text{jet}}$ , and fit the distribution around its peak with a Gaussian. We perform these fits separately in data and simulation, and use the mean and uncertainty on the mean of the fitted Gaussian as the value of  $K_{Z/\gamma}(E_T^{\text{jet}})$  and its uncertainty. This provides us an estimate of the most probable value of the balancing distribution, which we use for  $K_{Z/\gamma}(E_T^{\text{jet}})$  in order to avoid effects from a small number of highly mismeasured jets, which will more dramatically alter the mean and median of the distribution.

The distributions of the balancing variables in data and simulation are shown in Fig. 1. We not only see that the jets are not well-measured (the balancing variable is non-zero), but that in the  $Z$ -jet balancing, a sample largely dominated by gluon jets, we see significant disagreement between data and simulated jets. We do not see such disagreement in the  $\gamma$ -jet balancing, indicating our MC simulation models the behavior of the jets in this sample (dominated by quark jets) well.

We determine  $F_{Z/\gamma}^q$  in simulation by matching jets to their originating partons. In our  $\gamma$ -jet balancing sample, we find that the quark fraction is about 85% at  $E_T^{\text{jet}} \sim 30$  GeV, and drops to about 71% at  $E_T^{\text{jet}} \sim 70$  GeV. In the  $Z$ -jet balancing sample, these fractions are  $\sim 38\%$  and  $\sim 49\%$  in the same  $E_T^{\text{jet}}$  regions. In data, it is not possible to directly match jets to their originating parton, and we must rely on simulation to extract a value of  $F_{Z/\gamma}^q(E_T^{\text{jet}})$ . Because we are trying to correct for discrepancies in the reconstruction of quark and gluon jets between data and simulation, we cannot simply use the simulation-derived  $F_{Z/\gamma}^q$  values from each jet  $E_T$  bin. Instead, we can use the reconstructed  $p_T^{Z/\gamma}$ , which should match in data and simulation, and determine the quark/gluon fraction based on that value. Therefore, we parameterize the  $F_{Z/\gamma}^q$  from simulation as a function of  $p_T^{Z/\gamma}$  and determine the  $F_{Z/\gamma}^q$  data in each jet  $E_T$  of the data based on  $p_T^{Z/\gamma}$  distribution in the data.

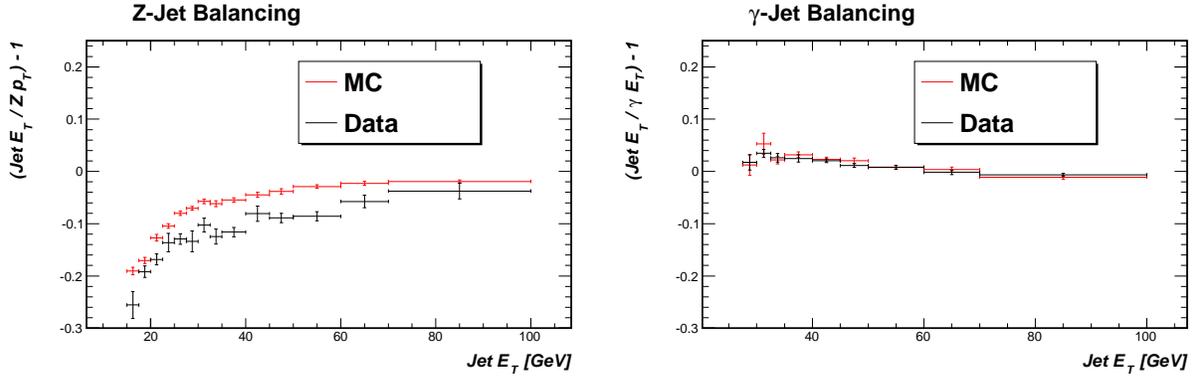


FIG. 1: The balancing distributions,  $K_Z$  (*left*) and  $K_\gamma$  (*right*), in data (*black*) and MC simulation (*red*) as a function of  $E_T^{\text{jet}}$ . The uncertainties are solely the uncertainty on the mean of a Gaussian fit to the balancing distributions in bins of  $E_T^{\text{jet}}$ .

Using Eqs. 4-5, we construct distributions of  $K_q$  and  $K_g$  as a function of the jet  $E_T$ , which are shown in Fig. 2. We see good agreement between data and MC in  $K_q$ , but worse agreement in  $K_g$ , where data appears consistently lower than MC. This suggests that MC is systematically overestimating gluon jet energies, relative to the data.

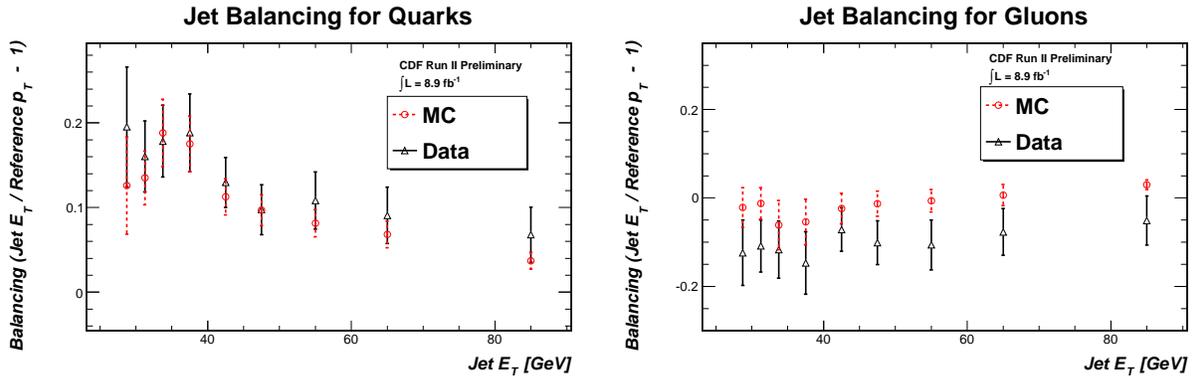


FIG. 2: The derived balancing variable for quark jets,  $K_q$ , (*left*) and gluon jets,  $K_g$ , (*right*) in data (*black*) and MC simulation (*red*) as a function of  $E_T^{\text{jet}}$ . The uncertainties on each point are from the uncertainties from the mean of the Gaussian fit and the uncertainties on the quark fractions, added in quadrature. We see better agreement between data and simulation in the energy scale of quark jets than that of gluon jets, following from the behavior seen in Fig. 1.

Using the distributions of  $K_q$  and  $K_g$ , we determine the corrections that need to be applied

to MC jets in order to best match the energy scale of the data. These MC corrections are defined as  $(K_q^{\text{Data}} + 1)/(K_q^{\text{MC}} + 1)$  for quark jets, and  $(K_g^{\text{Data}} + 1)/(K_g^{\text{MC}} + 1)$  for gluon jets, and are shown in Fig. 3. Due to the photon trigger used to select the  $\gamma$ -jet balancing sample, we do not have reliable balancing information for jets below 27.5 GeV in that sample, limiting the full range over which we may derive corrections. Since we are interested in jets down to energies around 20 GeV, we extrapolate to lower jet energies the quark jet energy correction derived for jets with  $E_T \geq 27.5$  GeV, and use the  $Z$ -jet balancing sample to extract a gluon correction assuming this extrapolated quark correction.

As both the quark and gluon corrections appear flat in jet energy for jets with  $E_T \geq 15$  GeV, we fit them to a constant. We find that to better match the data, quark jet energies in MC should be increased by 1.4%, while gluon jet energies should be decreased by 7.9%.

### C. Uncertainties on Simulated Jet Energy Corrections

We consider the following sources of uncertainty on the corrections presented in Sec. III B.

- *Fit/Statistical Uncertainty:* We use the standard deviation of the necessary corrections on simulation of each jet  $E_T$  bin to capture the spread of the MC corrections around the assumed flat correction function. This is an uncertainty of  $\pm 2.0\%$  for quark jet energies, and  $\mp 2.5\%$  for gluon jet energies.
- $F_Z^q$ : We compare the distribution of the jet QG value ([8]) in data and simulation, and fit the data distribution using quark and gluon templates from simulation. We take the average deviation of this jet QG extracted value from the nominal MC value as a systematic on  $F^q$ , constant across jet  $E_T$ . This uncertainty is  $\sim 10\%$  (absolute) in the  $Z$ -jet sample, and so we vary the quark fraction of the  $Z$ -jet sample by  $\pm 10\%$  and recalculate the corrections for quark and gluon jets. This translates to an uncertainty of  $\pm 0.6\%$  for quark jet energies, and  $\mp 2.1\%$  for gluon jet energies.
- $F_\gamma^q$ : We follow a similar procedure of fitting the jet QG value in the  $\gamma$ -jet sample, and obtain a similar uncertainty of  $\pm 10\%$  (absolute) on the quark fraction. This translates to an uncertainty of  $\pm 1.8\%$  for quark jet energies, and  $\mp 2.7\%$  for gluon jet energies.

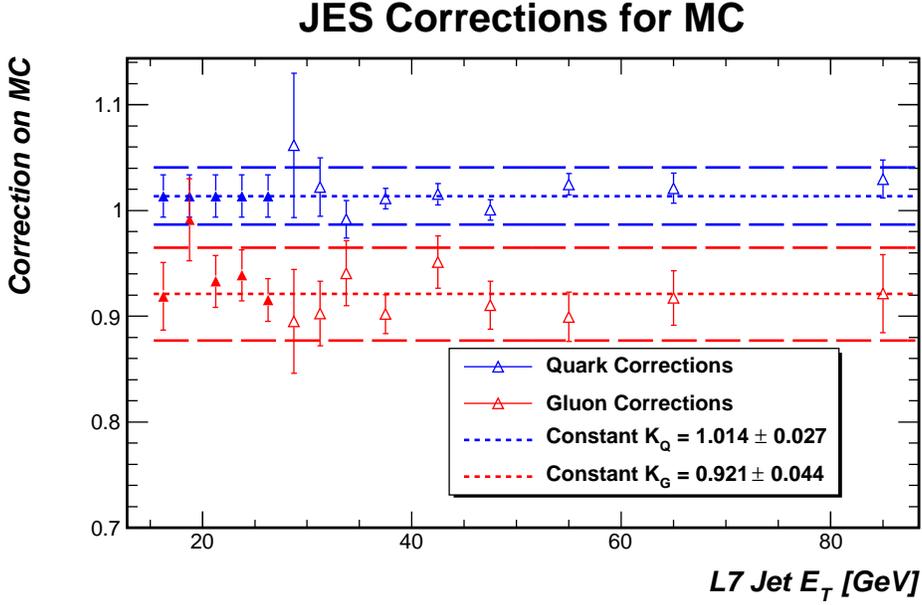


FIG. 3: The derived correction for simulated quark jets (*blue*) and gluon jets (*red*) as a function of  $E_T^{\text{jet}}$ . The open triangles represent corrections derived using both  $\gamma$ -jet and  $Z$ -jet balancing samples, while the filled triangles represent the assumed flat correction for quarks and the corresponding correction for gluons calculated from the  $Z$ -jet balancing sample alone. The error bars shown are statistical uncertainties only. The short dashed lines are the fits of the correction to a constant across jet  $E_T$ , and the long dashed lines represent the total systematic uncertainty bands on that constant correction, further described in Sec. III C.

- *Low  $E_T$  Extrapolation*: We check the dependence of the gluon jet energy corrections on the assumed quark jet corrections for low  $E_T$  jets by moving the quark jet  $E_T$  for these jets by  $\pm 2\%$ . We see a small change in the fit gluon energy corrections:  $\mp 0.4\%$  of the jet energy.
- *Number of Interaction Vertices Dependence*: The  $\gamma$ -jet sample as a cut on the number of reconstructed interaction vertices to reduce contamination from pileup. The  $Z$ -jet sample does not place such a cut, in order to retain as many events as possible. We check the effect this cut has by checking for any shift in the corrections when the cut is placed on the  $Z$ -jet sample. We see a change to the quark jet energies of  $\pm 0.2\%$ , and the gluon jet energies of  $\mp 1.2\%$ .

The uncertainties are summarized in Tab. I. Because the corrections shift the energy response in simulation to better match data, the quark jet and gluon jet energy correction uncertainties are anti-correlated: if the quark jet energy correction goes up, the gluon jet energy correction must go down in order to further compensate for that shift, and *vice versa*. The uncertainties are similar in magnitude to the default CDF jet energy scale uncertainties [6].

		Quark jets	Gluon jets
Jet Energy Correction		1.014	0.921
Uncertainty	Fit/Statistics	0.020	0.025
	$F_Q^{Z\text{-jet}}$	0.006	0.021
	$F_Q^{\gamma\text{-jet}}$	0.018	0.027
	Low $E_T$ Extrapolation		0.004
	$N_{vert}$ difference	0.002	0.012
Total Uncertainty		$\pm 0.027$	$\mp 0.044$

TABLE I: Summary of the additional jet energy corrections applied to MC jets, and the uncertainty on those corrections. The uncertainties for the quark jet and gluon jet energy corrections are anti-correlated, as they must work in concert to match the balancing distributions in data.

#### IV. MULTIJET BACKGROUND MODELING

Multijet events can be wrongly identified as signal when in 3-jet event, one of the jets fakes the lepton. Usually, this mismeasurement can also result in relatively large missing transverse energy. In muon events, the multijet background is negligible ( $< 0.5\%$ ) but is close to 10% in electron events. In this section we will concentrate on describing the multijet estimation and the systematic analysis for events with electrons. Same methods are applied to muon events, however, since the multijet contribution to muon events is much smaller, the effect of this background to the final analysis is much less important.

To model the multijet background we use exactly the same selection as described in Sec. II except for some of the electron identification cuts which are inverted. The objects identified

with those inverted cuts we call nonelectrons. This ensures that the sample we use for the multijet model is orthogonal to our signal sample and at the same time, kinematically, as close as possible. Nevertheless, several corrections have to be applied in order to obtain a good multijet model. First, although the lepton identification cuts have been inverted, there is a contribution of real leptons to this sample, that is due to the fact that our cuts are not 100% efficient. To account for this, we subtract from the nonelectron sample all processes with real lepton contributors (EWK, top, W+jets). This is done bin by bin for any variable of interest by taking the data content of a bin and subtracting the Monte Carlo prediction for that bin.

The second correction to the nonelectron sample is needed because of the trigger bias. In order for this sample to be a good multijet background model, the kinematics of the 3-jet process which contains the misidentified lepton candidate has to be identical to the process that contains the nonelectron. Since in reality this was a 3-jet event, neither the misidentified electron candidate, nor the nonelectron represent a good measurement of the original object, but rather the jet that contains them is what needs to be considered. To find this jet we search for a jet object in a cone of  $\Delta R < 0.4$  around the electron/nonelectron. The energy distribution of this matched jet in the misidentified electron and nonelectron samples should be the same. In order to test this, we look in a control region enriched in multijet events. The selection criteria for the control region are the same as for the signal region, except that we require events to have either  $\cancel{E}_T < 20$  GeV or  $m_T < 30$  GeV. The estimated fraction of multijet events in this region is 84%. When comparing the energy distribution of the jets matched to identified electrons and jets matched to the nonelectrons in this control region we find large discrepancies as seen in Fig. 4. This discrepancy is due to the fact that the electron trigger cuts on electromagnetic energy only. The jets matched to identified electrons are richer in electromagnetic energy than the jets matched to nonelectrons so in order to have an nonelectron at the same energy as the identified electron, the energy of the jet matched to the nonelectron has to be higher. To remove the trigger bias we reweight events in the nonelectron sample such that the energy spectrum of the jets matched to identified electrons is the same as the energy spectrum of jets matched to nonelectrons. This reweighting is done in the control region and the same weights are used in the signal region.

The third correction to the nonelectron sample addresses the question of how well the above described matched jet represents a true measurement of the initial parton. In par-

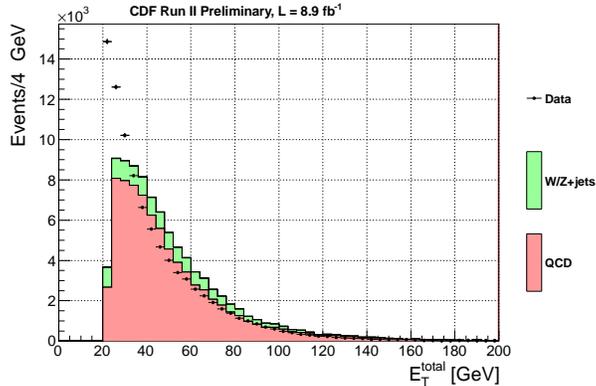


FIG. 4: The energy distribution of the jets matched to identified electrons in the control region as observed in the data and as predicted by the nonelectron based model

particular we are interested in whether the difference between the parton energy and the jet energy is the same in the identified electron sample and the nonelectron sample. We investigate this difference in Monte Carlo 2-jet QCD events. It turns out that the energy of the jets matched to nonelectrons is systematically lower than the energy of the jets matched to electrons given the same parton energy. We derive an energy dependent correction based on this Monte carlo study.

In order to test how well these corrections work we go back to the control region described before. An important kinematic distribution for this analysis is the  $P_T$  of the 2-jet system used to calculate the invariant mass for the obvious reason that this  $P_T$  is very correlated with the invariant mass itself. Fig. 5 shows this quantity before and after applying our correction in the control region defined above. Same level of significant improvement can be seen in other kinematical variables as well.

Aside from the control region, we also need to investigate how the multijet correction affects our final selected sample, as described in Sec. II. To increase the statistical accuracy of the sample we make a looser selection by removing the 2-jet system  $P_T$  cut and lower the jet  $E_T$  cuts to 25 GeV. The same variable as above,  $P_T$  of the 2-jet system, is shown in Fig. 6.

To estimate the amount of multijet background in our selected sample we perform a three component fit to the  $\cancel{E}_T$  distribution in the data. The three components are the multijet background, the W+jets and the rest of EWK processes. The latter component is constrained to the theoretical prediction whereas the W+jets and the multijet background

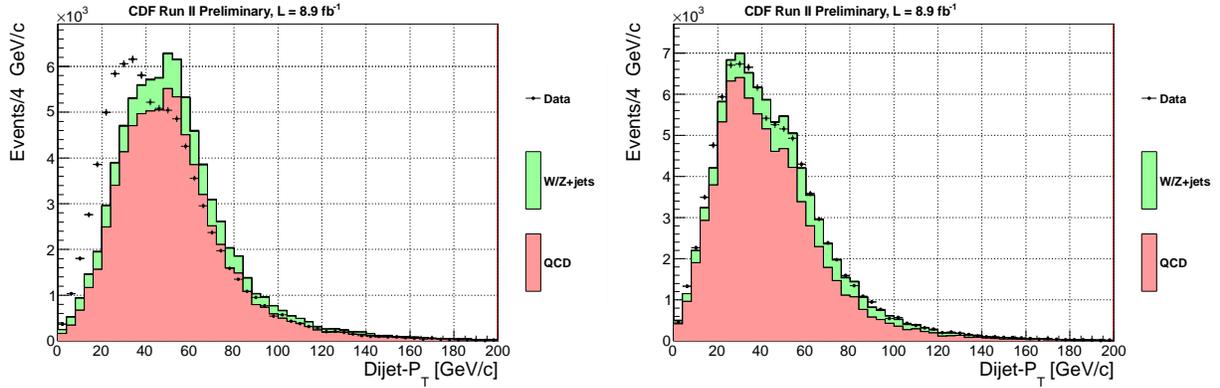


FIG. 5:  $P_T$  of the 2-jet system before (left) and after (right) applying the corrections described in text for the multijet enriched selection

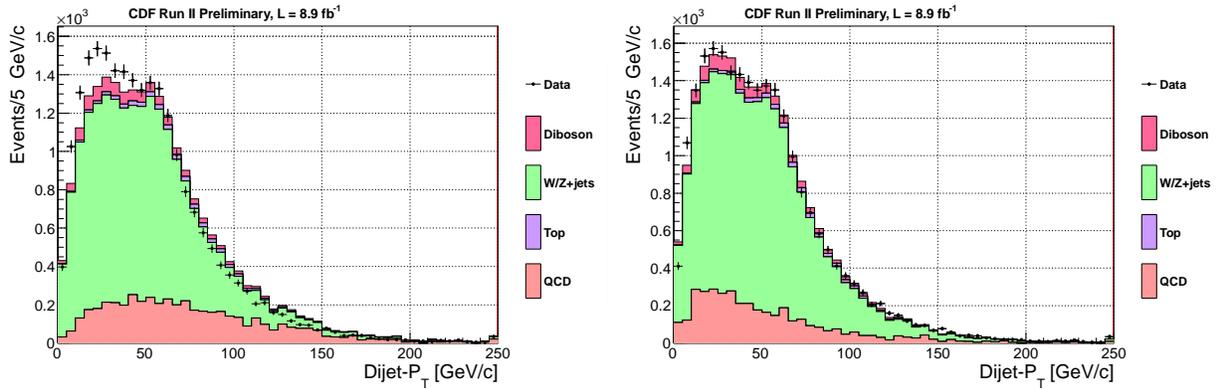


FIG. 6:  $P_T$  of the 2-jet system before (left) and after (right) applying the corrections described in text for the loose selection

are allowed to freely float in the fit. The  $W+ jets$  shape is taken from MonteCarlo and the multijet background shape is derived from the nonelectron sample using the corrections described before. The fit is shown in Fig. 7 from which we estimate the amount of multijet background in the signal sample to be  $7.8\% \pm 0.2\%$ , where the uncertainty is statistical only. A similar procedure is used for muons and the result is shown in the same Fig. 7, although as mentioned before, the multijet contribution to the muons sample is a negligible  $0.27\% \pm 0.01\%$ . There are several systematic uncertainties considered in this multijet background estimation: jet energy scale, choice of the variable to fit in, theoretical uncertainties on the EWK cross sections. etc. We sum in quadrature all those systematic uncertainties for a total systematic uncertainty on the multijet background estimation of 14.1%.

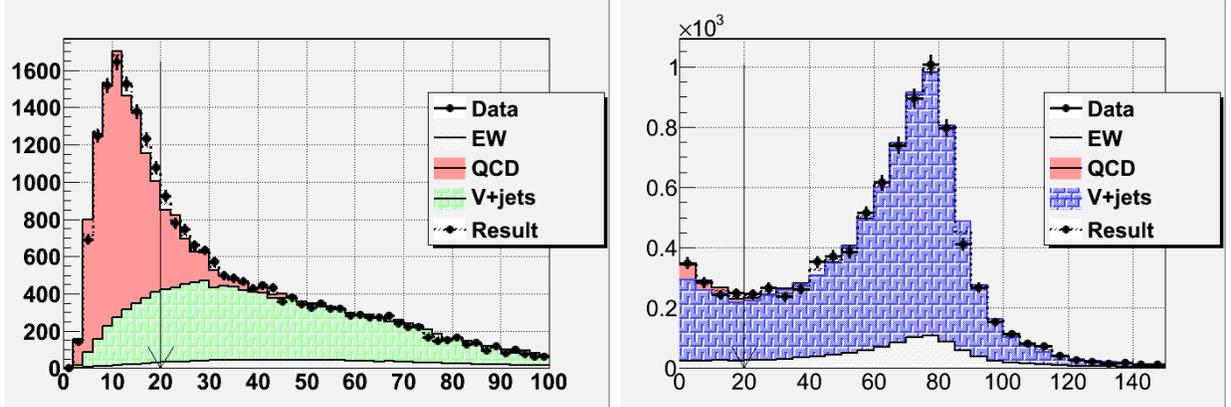


FIG. 7:  $\cancel{E}_T$  distribution for electrons (left) and  $m_T$  distribution for muons (right). The amount of multijet background is determined from fits to those distributions

## V. FIT TECHNIQUES

In order to extract information about the strength of the diboson signal and to test the hypothesis of the presence of other signals, fits must be performed, comparing the data with sums of signal and background predictions. These predictions have associated systematic uncertainties which are expected to enlarge the uncertainties on the fitted cross sections and to weaken the sensitivity to test for new particles. These uncertainties are parameterized with nuisance parameters, such as the  $W$ +jets cross section, the integrated luminosity, and the jet energy scales. The data are used in the fits to constrain both the parameters of interest and also the nuisance parameters. Some nuisance parameters, such as the  $W$ +jets background normalization, are highly constrained by “sideband” regions of the  $m_{jj}$  distributions, away from the expected location of the signal. Other parameters, such as those describing the uncertainties in the PDF’s, are less well constrained by the data.

Two approaches are followed here to extract cross sections and to set upper limits on the production rates of new particles. The first approach is to maximize the binned likelihood function  $L(\text{data}|\vec{\theta}, \vec{\nu})\pi(\vec{\nu})$ , which expresses the probability of observing the data given the model parameters  $\vec{\theta}$  and the nuisance parameters  $\vec{\nu}$ . The likelihood is a product of Poisson probabilities to observe the data in each bin, and the function  $\pi(\vec{\nu})$  is a product of Gaussian constraints, one for each nuisance parameter, which incorporates external information about each parameter, as measured in control samples or other measurements or predictions. The nuisance parameters describe systematic uncertainties in three forms – bin-by-bin uncertain-

ties which are considered uncorrelated for each bin of each template; shape uncertainties, which are coherent distortions of the contents of each bin of a template, parameterized by a single nuisance parameter, and rate uncertainties, which affect the overall normalization of one or more templates. Rate and shape uncertainties may be correlated with each other. For example, the jet energy scale shifts the location of the signal peak to the left or to the right in the  $m_{jj}$  distribution, and it affects the acceptance of the signal due to the requirement that jets pass the minimum jet  $E_T$  cut. Correlations are taken into account by allowing each source of systematic uncertainty parameterized by a single nuisance parameter affect the rates and shapes of multiple templates. A detailed description of the construction of the likelihood function is given in Ref. [9]. Restrictions are placed on the allowed ranges of the nuisance parameters to ensure that all predictions are non-negative. The program MINUIT [10, 11] is used to perform the maximization of the likelihood function. The associated uncertainties are computed using MINOS, part of MINUIT.

The second approach used is Bayesian in nature. The Gaussian constraints on the nuisance parameters are re-interpreted as prior distributions for each nuisance parameter. The prior function for a nuisance parameter vanishes if that value of the parameter results in a negative prediction for a signal or background component. A uniform prior is assumed for the signal cross section, which is also assumed to be non-negative. The marginalized likelihood [12] is given by the integral of the likelihood multiplied by the prior functions for the nuisance parameters

$$L_m(\text{data}|\vec{\theta}) = \int L(\text{data}|\vec{\theta}, \vec{\nu}) \times \pi(\vec{\nu}) d\vec{\nu}. \quad (6)$$

Since the prior for the signal rate is uniform,  $L$ , after normalizing to unit area, is the posterior density in the parameters of interest  $\vec{\theta}$ . The best-fit cross section then corresponds to that value which maximizes  $L_m$ , and the  $\pm 1$  s.d. uncertainties are given by the shortest interval which contains 68% of the integral of the posterior density. The 95% C.L. upper limit on the cross section is that for which 95% of the area of the posterior corresponds to smaller production cross sections. The integrals over the nuisance parameters and also over the parameters of interest are performed using the Metropolis-Hastings algorithm [13, 14].

## VI. SYSTEMATICS

As described in Sec. V, there are two categories of systematic uncertainties considered. They enter the fit as nuisance parameters that affect the normalization and the shape of the invariant mass templates that are supplied for each physical process that contributes.

The templates, and the uncertainties on their normalizations, are listed below.

1. EWK ( $W/Z$ +jets): Normalizations are allowed to float in the fit, unconstrained.
2.  $t\bar{t}$  and single top: The normalization of the  $t\bar{t}$  template is constrained to the measured cross section as reported in [15]. The single top normalization is constrained to the theoretical cross section with an uncertainty of 11% [16, 17].
3. Multijet background: We use our data-driven estimate with an uncertainty of 15%. There are multiple sources that contribute to this uncertainty but the dominant one is from using different variables to do the fit for multijet background described in Sec. IV.
4.  $WW, WZ, ZZ$ : We use the NLO cross section and apply a Gaussian constraint to the number of events centered on this value with a width equal to the theoretical uncertainty of 6% [18].

Other sources of systematic uncertainties that affect the template normalizations are: luminosity (6%), initial and final state radiation (2.5%), PDFs (2.2%), jet energy scale (2.7% for quark jets and 4.8% for gluon jets with a 100% anti-correlation), jet energy resolution (0.7%) and trigger efficiency (2.2%).

In addition to uncertainties on the normalizations of each template, we consider other systematic uncertainties that may affect the shape of templates. The most important are the jet energy scale and the renormalization and the factorization scale  $Q^2$  in  $W/Z$ +jets Monte Carlo. For the former, two alternative templates are obtained by varying the jet energy scale within its expected  $\pm 1\sigma$  uncertainty. For the latter, the  $Q^2$  parameter used in the Monte Carlo generation is doubled and halved in order to obtain two alternative templates.

## VII. INVARIANT MASS SPECTRUM BEFORE CORRECTIONS

We first perform a fit in dijet invariant mass without considering the corrections described in the previous sections. Here, there is no correction to the multijet background

invariant mass shape (Sec. IV), no jet energy scale correction (Sec. III). Similar to [4], aside from Standard Model contributions we allow in the fit a gaussian component centered at  $144 \text{ GeV}/c^2$  with a width of  $14.3 \text{ GeV}/c^2$ . Assuming an acceptance identical to the one for a  $140 \text{ GeV}cc$  Higgs produced in association with a  $W$  boson, the cross section determined would be  $2.4 \pm 0.6 \text{ pb}$ . The probability to observe an excess larger than what is seen in Fig. 8 assuming only backgrounds and systematic uncertainties is  $2.6 \times 10^{-5}$  corresponding to  $4.2\sigma$  in terms of gaussian standard deviations. A similar discrepancy between data and prediction can be observed in both electrons and muons as shown in Fig. 9. Nonetheless, it is clear that a simple gaussian does not describe well the excess observed in data.

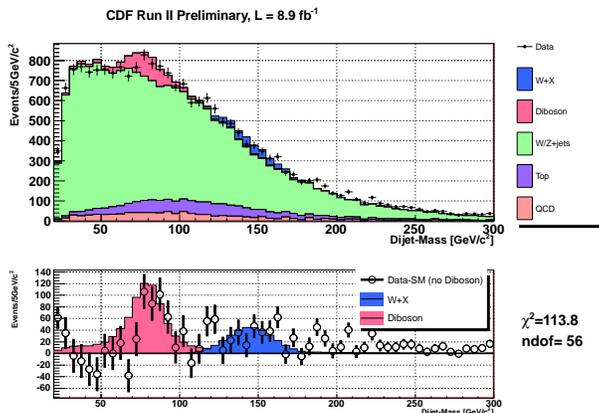


FIG. 8: Fit to the dijet invariant mass distribution similar to [4]. The corrections described in text have not been applied here. Bottom figure shows data with all backgrounds (except the diboson contribution) subtracted.

## VIII. INVARIANT MASS SPECTRUM AFTER CORRECTIONS

We now perform the same fit as in Sec. VII after including the corrections described in (Sec. IV) and (Sec. III). We will first look at the effect of the correction due to different response to quarks and gluons as described in Sec. III. The two samples (electrons and muons) are shown in Fig. 10. Good agreement between data and the prediction is seen in the muon sample while the agreement is better, but still poor in the electron sample. However, after applying the correction to the multijet background the electron sample shows a good agreement too as seen in Fig. 11.

The result is shown in Fig. 12. In this case, the data is consistent with no excess around

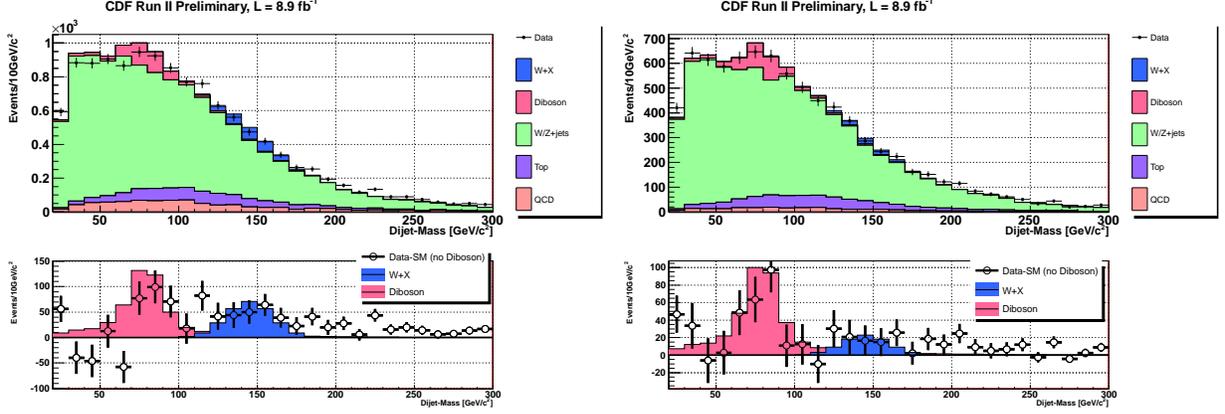


FIG. 9: Fit to the dijet invariant mass distribution similar to [4]. The corrections described in text have not been applied here. Bottom figures shows data with all backgrounds (except the diboson contribution) subtracted. On the left we show the invariant mass distribution in electron events and on the right the invariant mass distribution in muon events.

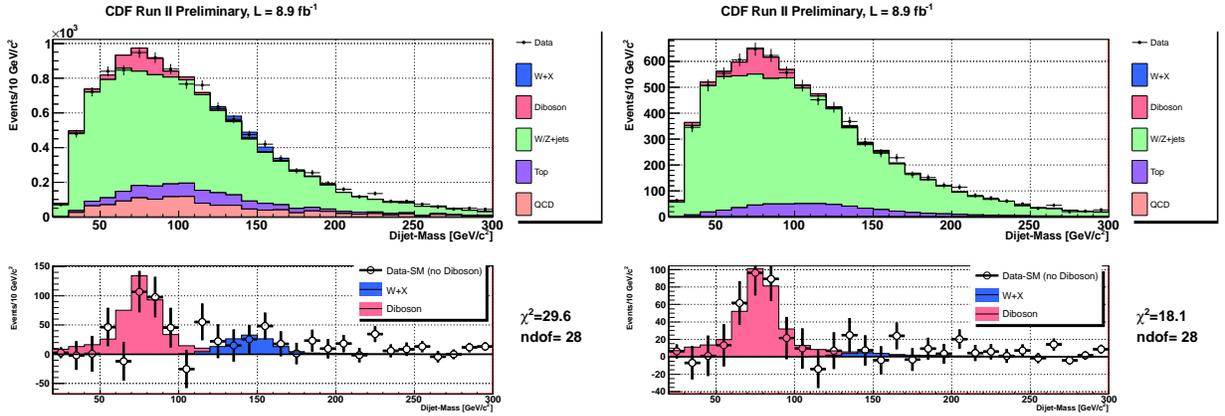


FIG. 10: Fit to the dijet invariant mass distribution for electrons (left) and muons (right). The corrections described in text have not been applied here. Bottom figure shows data with all backgrounds (except the diboson contribution) subtracted.

$144 \text{ GeV}/c^2$ . The probability to observe an excess larger than what is seen in Fig. 12 assuming only backgrounds and systematic uncertainties is 72% corresponding to  $0.35\sigma$  in terms of gaussian standard deviations.

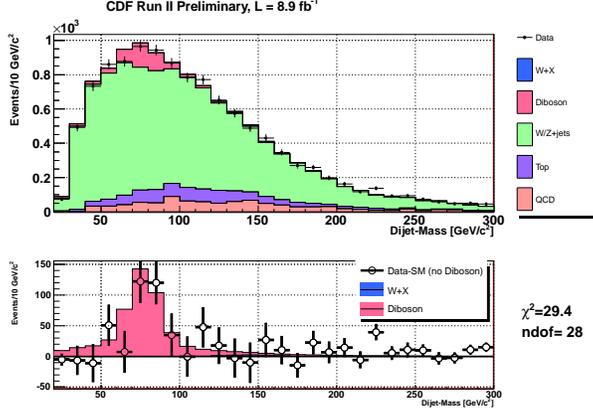


FIG. 11: Fit to the dijet invariant mass distribution in the electron sample. All corrections described in text have been applied here. Bottom figure shows data with all backgrounds (except the diboson contribution) subtracted.

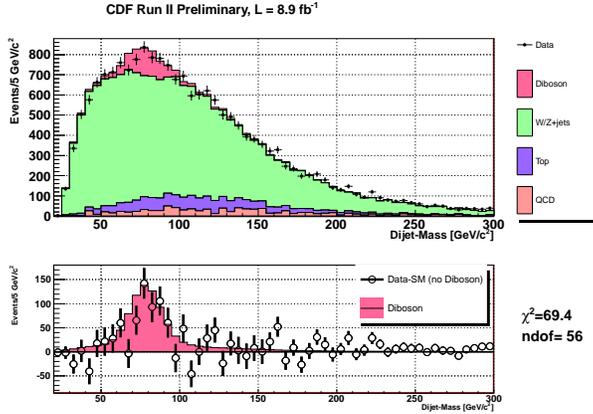


FIG. 12: Fit to the dijet invariant mass distribution similar to [4]. The corrections described in text have been applied here. Bottom figure shows data with all backgrounds (except the diboson contribution) subtracted.

## IX. ORTHOGONAL SAMPLES

Similar analyses are performed in two orthogonal samples. Same fit in invariant mass is done as described in Sec. V and same kinematic cuts are applied as in Sec. II. The only differences are in the objects selected.

### A. $lljj$ sample

We select here two charged leptons (electrons or muons) above 20 GeV and two jets with  $E_T$  above 25 GeV. The invariant mass of the two leptons is required to be within 15 GeV of the Z mass. The dominant contributors to this sample are the Z+2jets processes. The JES correction described in Sec. III is applied to the jets selected in Monte Carlo. Given the negligible multijet contribution to this sample the multijet corrections are ignored here. Similar to Sec. VIII we show in Fig. 13 the invariant mass distribution of the two jets after all backgrounds (except the dibosons) are subtracted. We observe good agreement between data and predictions.

### B. $\cancel{E}_T jj$ sample

We select here two jets with  $E_T$  above 35 GeV and 25 GeV respectively and no other jet above 15 GeV together with a calorimeter transverse momentum imbalance ( $\cancel{E}_T$ ) of more than 50 GeV. The selected jets are required to be further apart than 1.0 in  $\Delta R$ . Multijet events in this sample are predicted this contribution using a data driven model. However, since these multijet events do not enter the sample as fake leptons but rather as jet mis-measurements that results in large  $\cancel{E}_T$  and the data is collected with a completely different trigger, the corrections described in Sec. IV do not apply. We show in Fig. 14 the invariant mass distribution of the two jets after all backgrounds (except the dibosons) are subtracted. We observe good agreement between data and predictions.

## X. CONCLUSION

We presented in this note the invariant mass spectrum of two jets produced together with a heavy boson at CDF. Since the last CDF publication on this topic ([4]) a number of systematic effects we investigated. The most important, the Monte Carlo response to quarks and gluons and the multijet background resulted in corrections that were applied to our model of the dominant backgrounds. Using all data collected by CDF, there is good agreement between data and the Standard Model prediction.

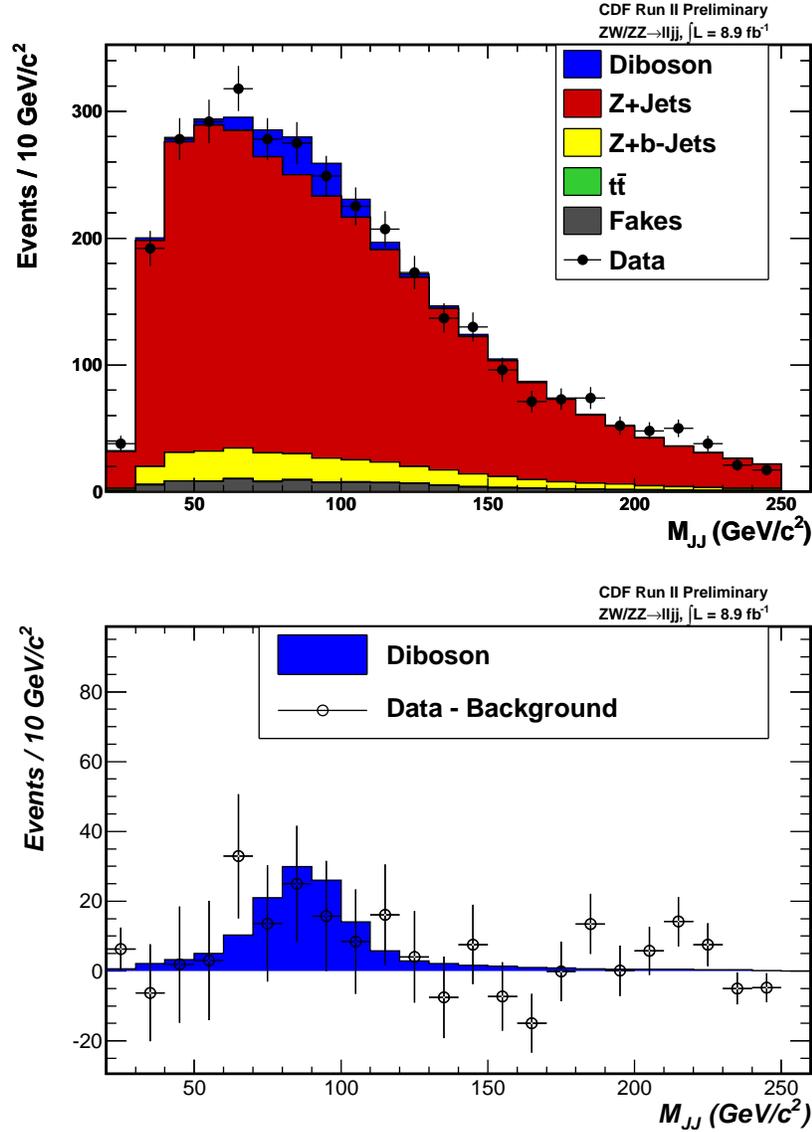


FIG. 13: Invariant mass distribution of two jets produced together with two charged leptons (electrons or muons).

### 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

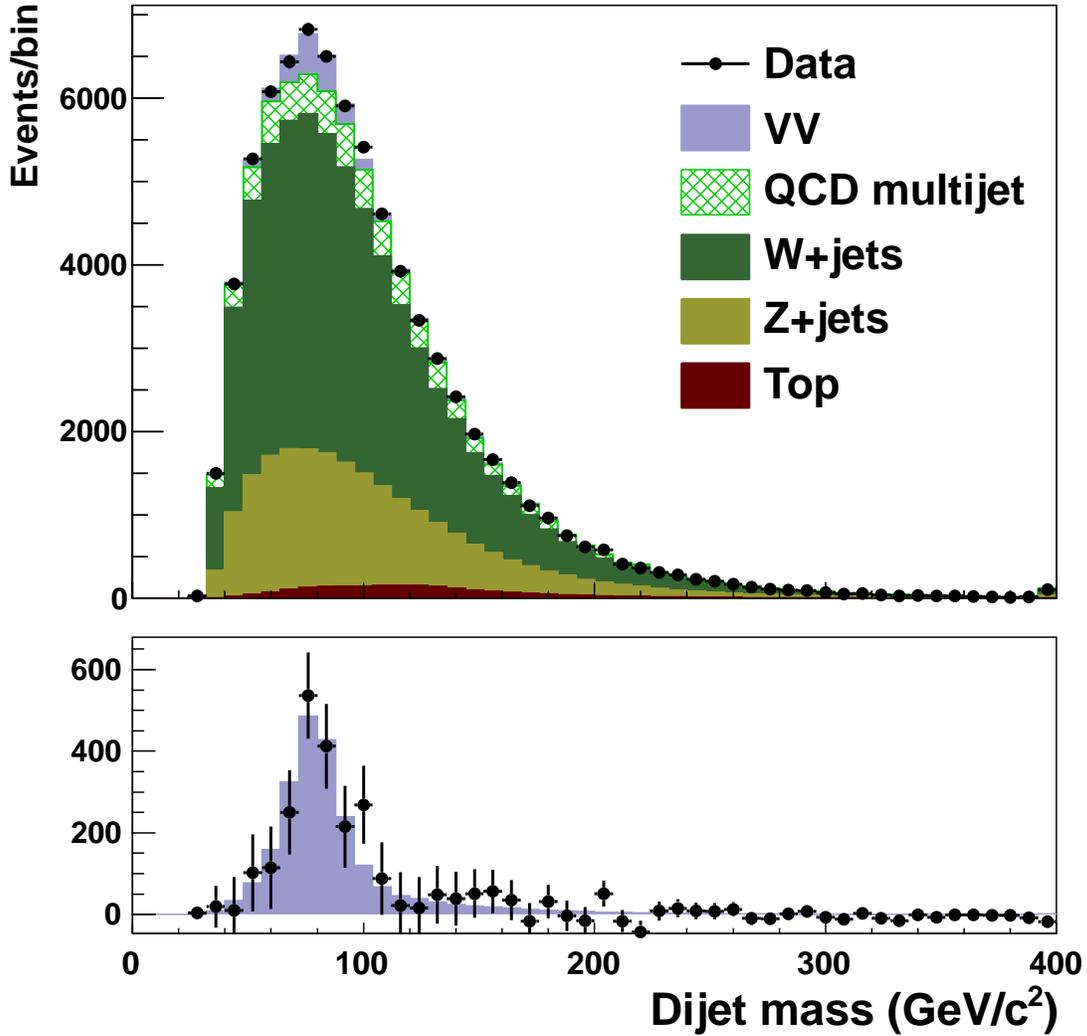


FIG. 14: Invariant mass distribution of two jets produced together with large  $\cancel{E}_T$ .

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 Bildung und Forschung, Germany; the Korean World Class University Program, the National Research Foundation of Korea; the Science and Technology Facilities Council and the Royal Society, UK; the Russian Foundation for Basic Research; the Ministerio de Ciencia e Innovación, and Programa Consolider-Ingenio 2010, Spain; the Slovak R&D Agency; the Academy of Finland; the Australian Research Council (ARC); and the EU community Marie Curie Fellowship contract 302103.

- 
- [1] K. Hagiwara et al., Nucl. Phys. **B 282**, 253 (1987).
- [2] M. Kober, B. Koch, and M. Bleicher, Phys. Rev. D **76**, 125001 (2007).
- [3] E. J. Eichten, K. Lane, and A. Martin, Phys. Rev. Lett. **106**, 251803 (2011), URL <http://link.aps.org/doi/10.1103/PhysRevLett.106.251803>.
- [4] T. Aaltonen et al. (CDF Collaboration), Phys. Rev. Lett. **106**, 171801 (2011), URL <http://link.aps.org/doi/10.1103/PhysRevLett.106.171801>.
- [5] A. Abulencia et al. (CDF Collaboration), J. Phys. G **34**, 2457 (2007).
- [6] A. Bhatti et al., Nucl. Instrum. Methods A **566**, 375 (2006).
- [7] T. Sjöstrand et al., J. High Energy Phys. **05**, 026 (2006).
- [8] W. R. Ketchum (????).
- [9] T. Aaltonen et al. (CDF Collaboration), Phys.Rev. **D82**, 112005 (2010), 1004.1181.
- [10] F. James and M. Roos, Comput.Phys.Commun. **10**, 343 (1975).
- [11] F. James (2006).
- [12] J. Beringer et al. (Particle Data Group), Phys.Rev. **D86**, 010001 (2012).
- [13] N. Metropolis et al., Journal of Chemical Physics **21**, 1087 (1953).
- [14] W. Hastings, Biometrika **57**, 97 (1970).
- [15] T. Aaltonen et al. (CDF Collaboration), *Combination of top pair production cross section results*, URL [http://www-cdf.fnal.gov/physics/new/top/2009/xsection/ttbar\\_combined\\_46invfb](http://www-cdf.fnal.gov/physics/new/top/2009/xsection/ttbar_combined_46invfb).
- [16] N. Kidonakis, Phys. Rev. D **74**, 114012 (2006).
- [17] N. Kidonakis, PoS(DIS2010) **196** (2010), arXiv:1005.3330v1.
- [18] J. M. Campbell and R. K. Ellis, Phys. Rev. D **60**, 113006 (1999).