Study of Dijet Invariant Mass Distribution in lνjj Final States

Arán García-Bellido
University of Rochester
On behalf of the D0 Collaboration

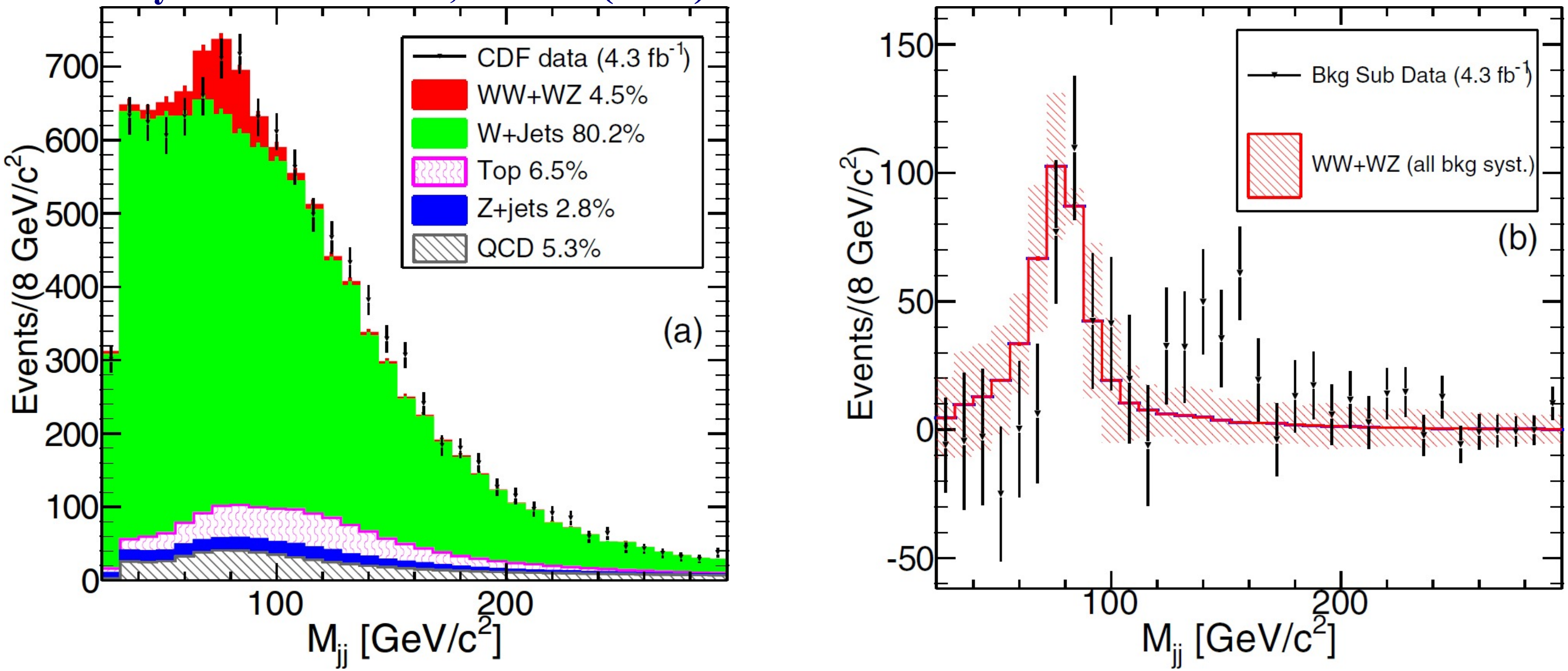
Physics at the LHC 2011 - Perugia - June 11, 2011

Slides based on Wine & Cheese seminar on June 10 by Joe Haley

What's so hot about lνjj?

- CDF has reported an excess of events in the dijet mass spectrum above the expected Standard Model contributions
- Does DØ confirm this excess?

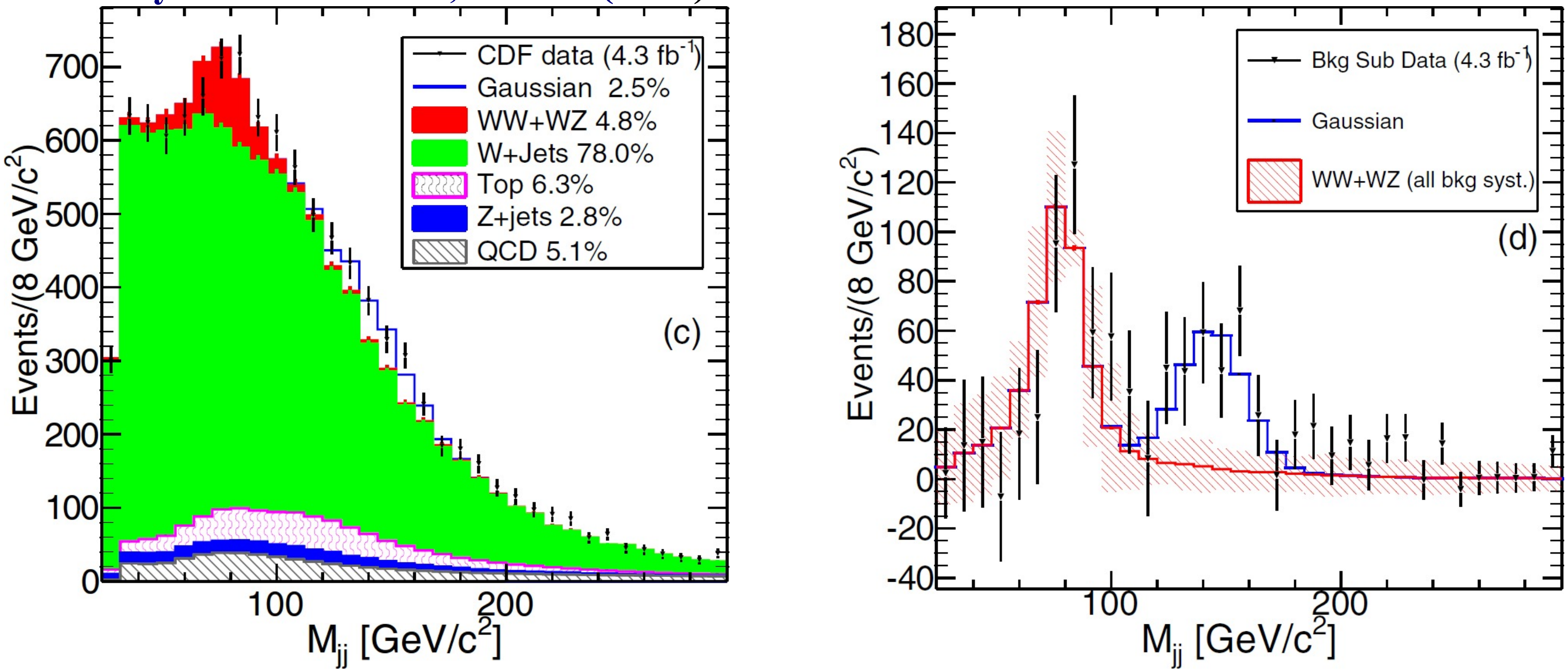
Phys. Rev. Lett. 106, 171801 (2011)



The CDF Excess

- Using 4.3 fb^{-1} integrated luminosity the CDF data show an excess of 3.2 standard deviations around a dijet mass ~145 GeV
  - Modeled by a Gaussian with width expected from jet resolution
  - If this is a resonance from some new particle, X, then $\sigma(p\bar{p} \to WX) \approx 4$ pb
    - Assumes BR(X→jj) = 1.0 and the same efficiency as WH→lvbb with m_H=150 GeV

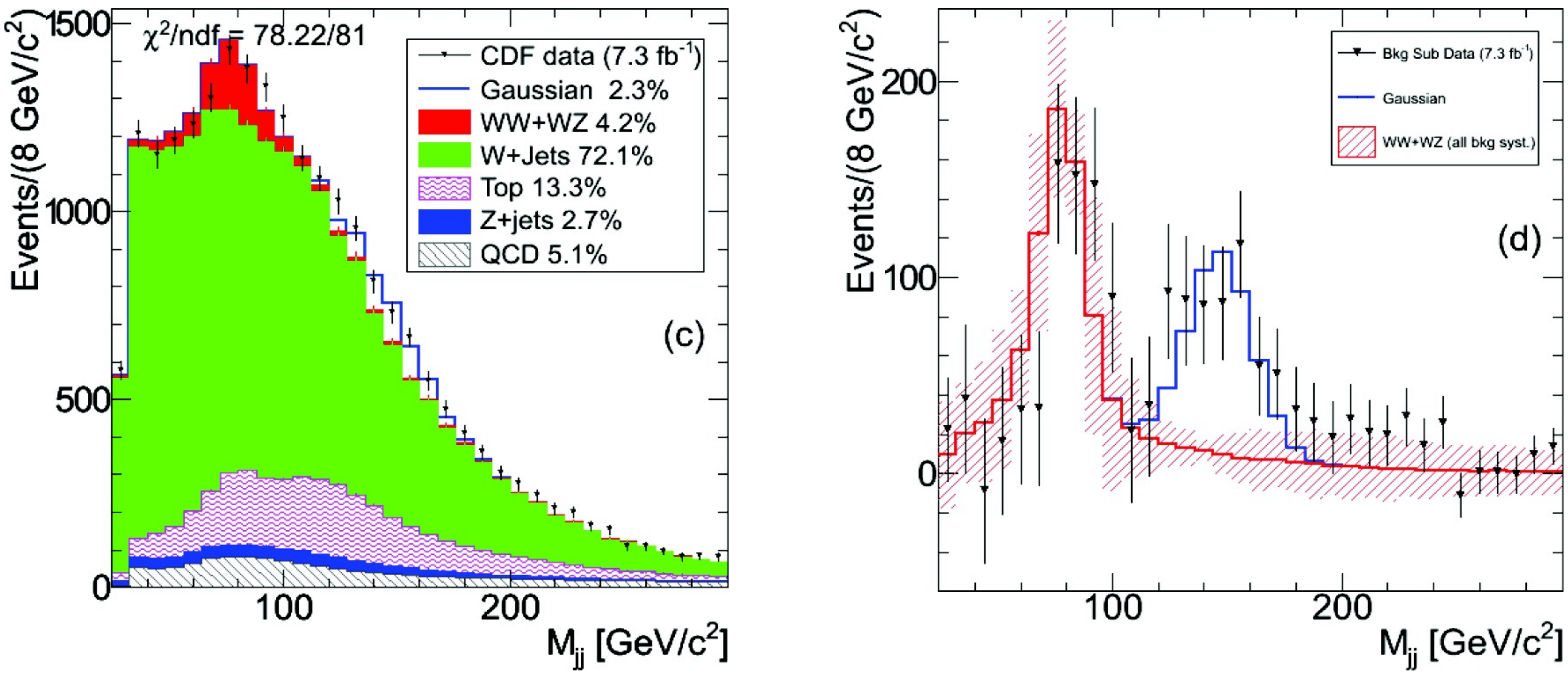
Phys. Rev. Lett. 106, 171801 (2011)



The CDF Bump

- CDF has updated results using an integrated luminosity of 7.3 fb^{-1}
  - Significance of excess now exceeds 4 σ

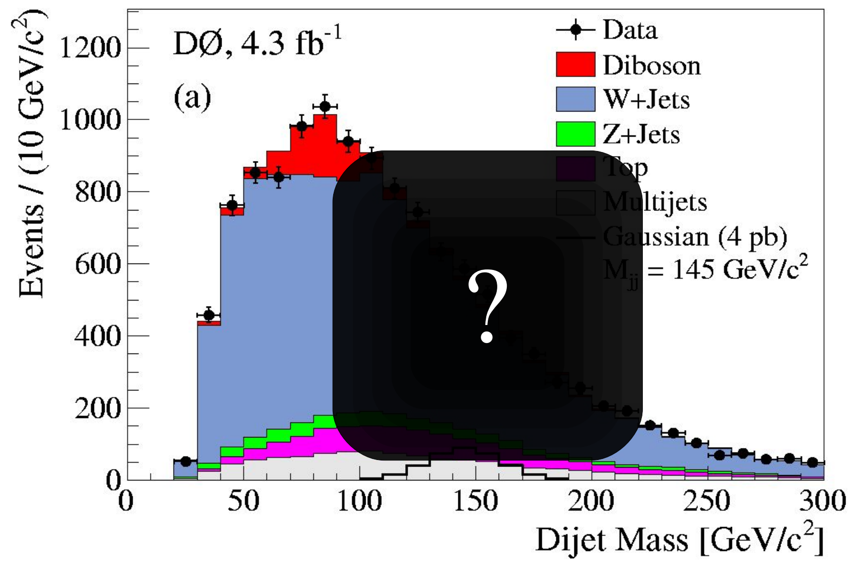
www-cdf.fnal.gov/physics/ewk/2011/wjj/7_3.html



Analysis Outline

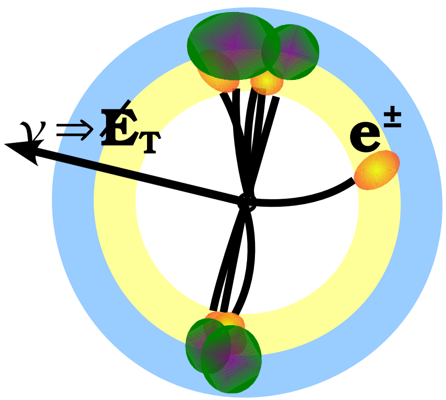
- Try to mirror what was done in the CDF analysis
  - Started from ongoing DØ diboson analysis and modified the kinematic selection to replicate the CDF publication
  - Make similar assumptions on modeling an excess

- Study the dijet mass distribution in the DØ data
  - Fit SM contributions to the data
    - Do we have an excess of events around M_{jj} = 145 GeV?
  - Include a model for WX→lνjj
    - How large of an excess do the DØ data support?

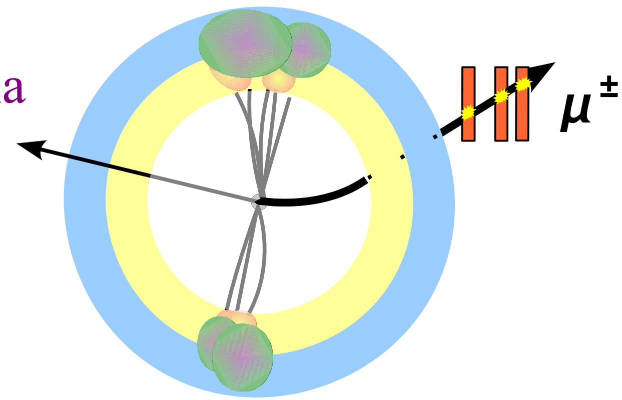


Event Selection

- 4.3 fb^{-1} of integrated luminosity collected by the DØ detector
  - Events containing W(→lν) and 2 jets
  - Require isolated lepton (e or μ): $p_T \geq 20$ GeV, $|\eta| < 1.0$
  - Neutrino: MET > 25 GeV
  - Lepton+Neutrino system: 30 GeV < $m_T(W)$ < 200 GeV
  - Veto events with more than one charged lepton



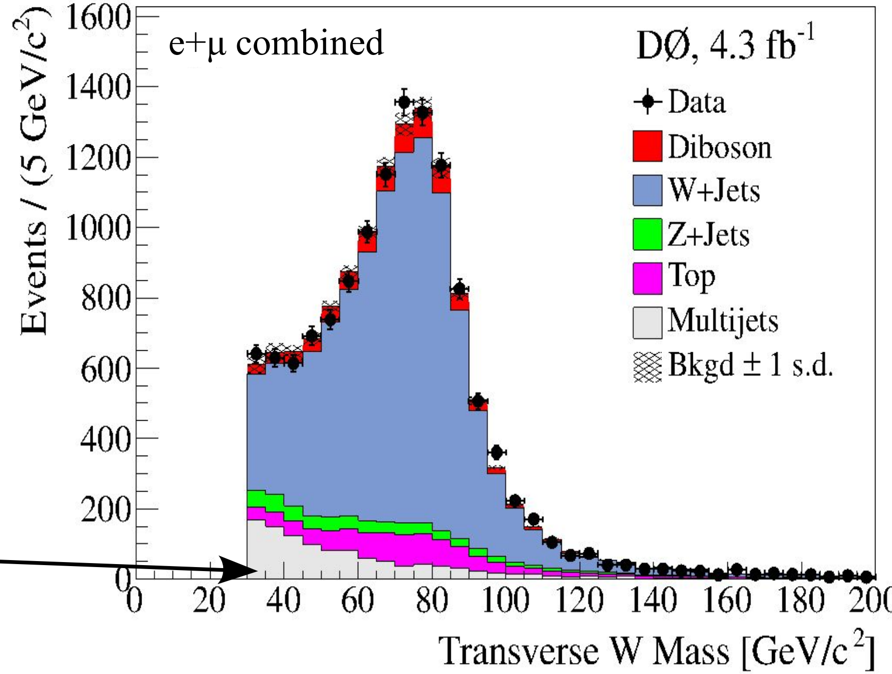
  - Two Jets:
    - $p_T(j_1)$ and $p_T(j_2) \geq 30$ GeV, $|\eta_{detector}| \leq 2.5$
    - Veto events with additional jets meeting these criteria
  - Dijet System
    - $p_T(j_1 j_2) \geq 40$ GeV
    - $|\Delta\eta(j_1,j_2)| \leq 2.5$



  - Reduce mis-measured E_T
    - $\Delta\varphi(j_1,\mathrm{MET}) \geq 0.4$

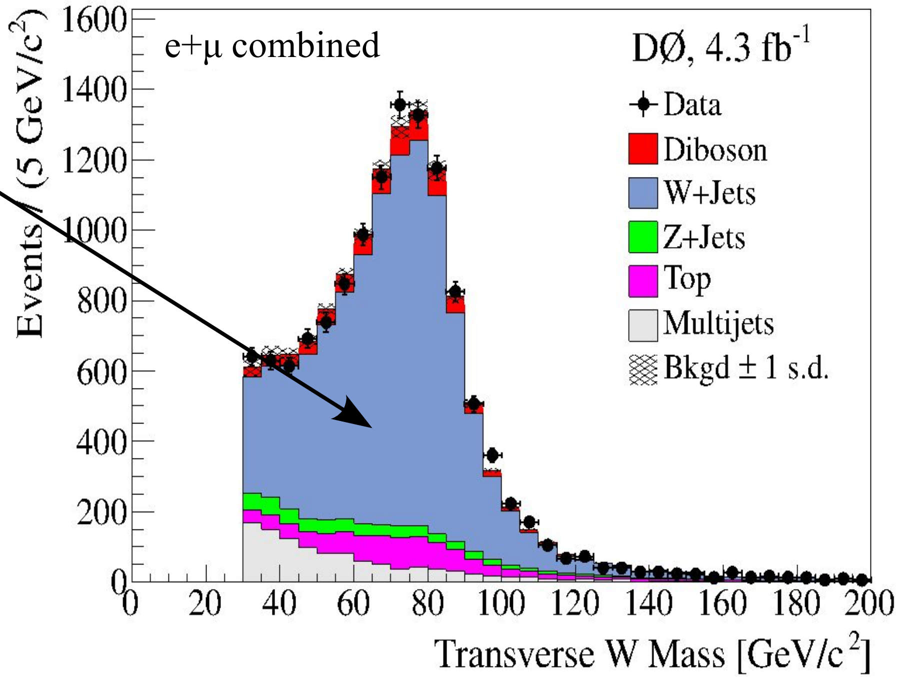
SM Predictions

- Processes with a high p_T lepton modeled via Monte Carlo generators
  - Pythia: WW, WZ, ZZ
  - CompHep+Pythia: single top
  - Alpgen+Pythia: W+jets, Z+jets, $t\bar{t}$
  - With Geant-based detector simulation
- Multijet background
  - A jet is mis-identified as a lepton
  - Estimated from multijet enriched data
    - Muon channel: Reverse isolation cuts
    - Electron channel: Release EM quality criteria
  - Corrected for contributions already accounted for by MC
  - Normalization determined by fitting the $m_T(W)$ distribution



SM Predictions

- Dominated by W+jets (~75%)
  - Vital to understand this background when looking differences of a few percent
  - MC generators are meant to reproduce the SM, but are only approximations
    - Make assumptions and simplifications
    - Many "knobs" to tune
    - Different generators give different results
- No corrections applied to Alpgen prediction
  - Systematic effects allow for variation of W+jets background shape

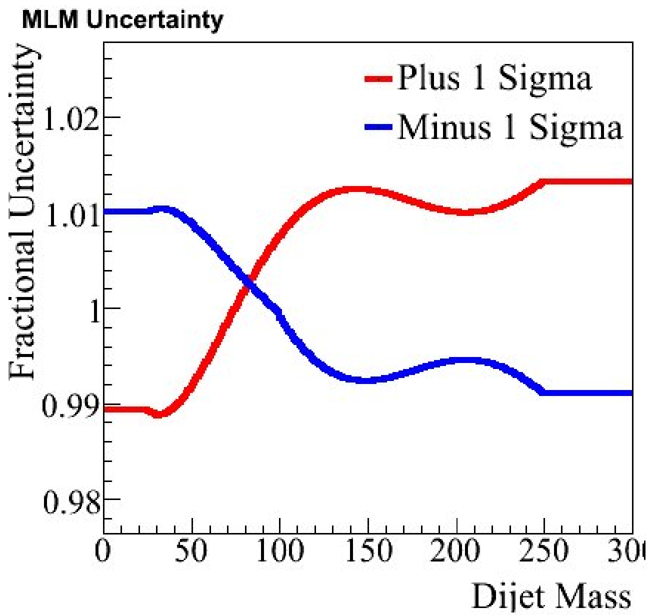


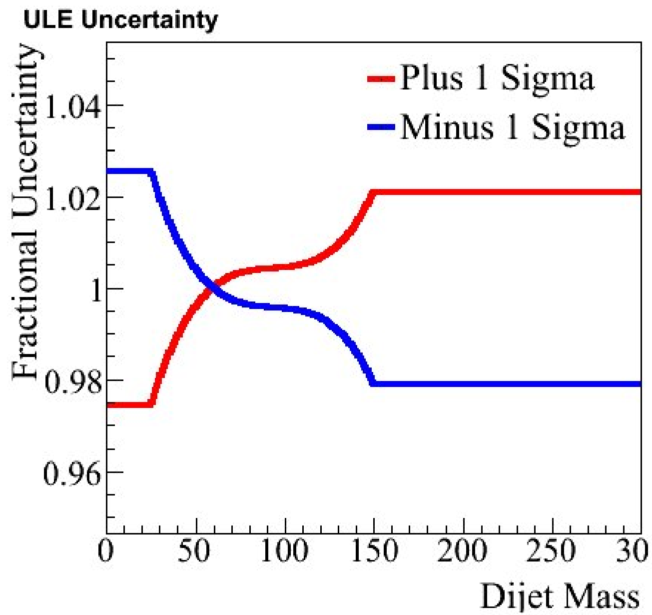
Sample composition: fit result

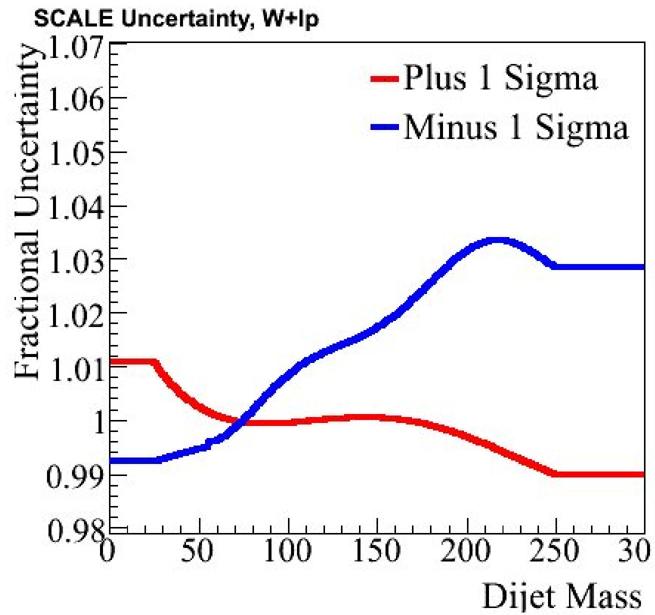
	Electron channel	Muon channel
Dibosons	434 ± 38	304 ± 25
W+jets	5620 ± 500	3850 ± 290
Z+jets	180 ± 42	350 ± 60
$t\bar{t}$ + single top	600 ± 69	363 ± 39
Multijet	932 ± 230	151 ± 69
Total predicted	7770 ± 170	5020 ± 130
Data	7763	5026

W+jets Modeling

- Large NLO/LO k-factor
- Uncertainties due to:
  - P_T threshold for Alpgen MLM matching prescription
  - Parton shower model (Pythia vs. Herwig) and underlying event model (tunes)
  - Renormalization/Factorization scale choice







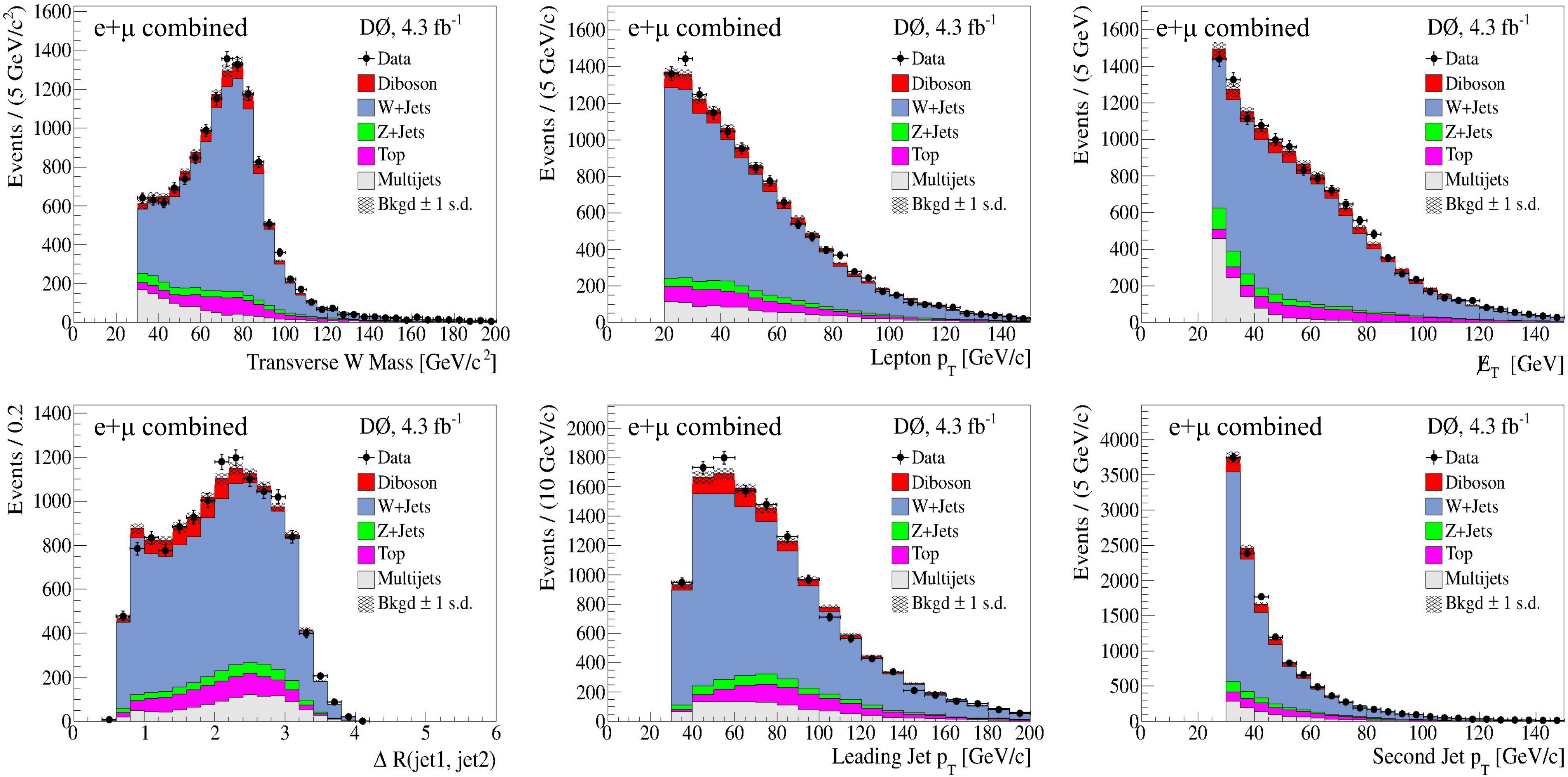
Fit of SM to Data

- Fit dijet mass distributions for all SM processes to the data
  - Construct a χ^2 function from the ratio of Poisson likelihoods and include prior information on the systematic uncertainties
  - Float cross sections for Diboson and W+jets contributions

Source of systematic uncertainty	Diboson signal	W+jets	Z+jets	Top	Multijet	Nature
Trigger/Lepton ID efficiency	±5	±5	±5	±5		N
Trigger correction, muon channel	±5	±5	±5	±5		D
Jet identification	±1	±1	±2	±1		D
Jet energy scale	±10	±5	±7	±5		D
Jet energy resolution	±6	±1	±3	±6		D
Jet vertex confirmation	±3	±3	±4	±1		D
Luminosity	±6.1	±6.1	±6.1	±6.1		N
Cross section		±6.3	±6.3	±10		N
V+hf cross section		±20	±20			N
V+2 jets/V+3 jets cross section		±10	±10			N
Multijet normalization					±20	N
Multijet shape, electron channel					±1	D
Multijet shape, muon channel					±10	D
Diboson modeling	±8					D
Parton distribution function	±1	±5	±4	±3		D
Unclustered Energy correction	± <1	±3	±3	± <1		D
ALPGEN η and $\Delta R(jet1, jet2)$ corrections		$\pm < 1$	$\pm < 1$			D
ALPGEN W p_T correction		$\pm < 1$				D
ALPGEN correction Diboson bias	±1	±1	±1	±1		D
Renormalization and factorization scales		±1	±1			D
ALPGEN parton-jet matching parameters		±1	±1			D
Parton shower and Underlying Event		±2	±2			D

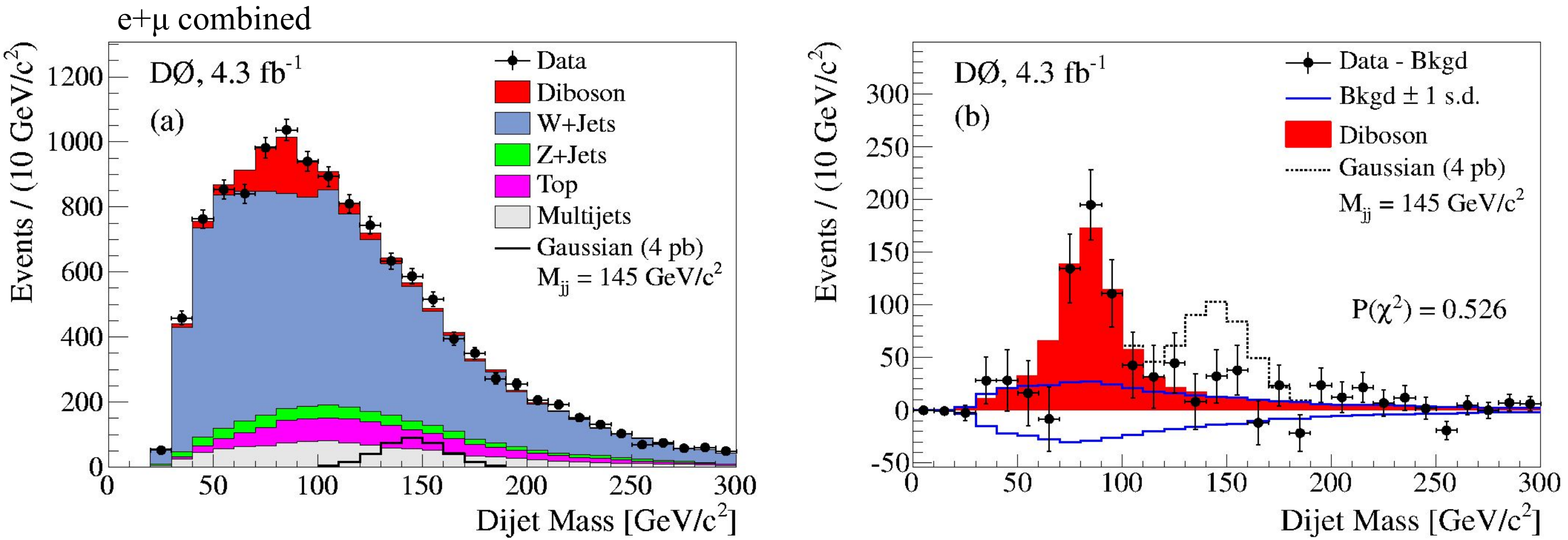
Fit of SM to Data

- Kinematic distributions after the fit in the dijet mass



Fit of SM to Data

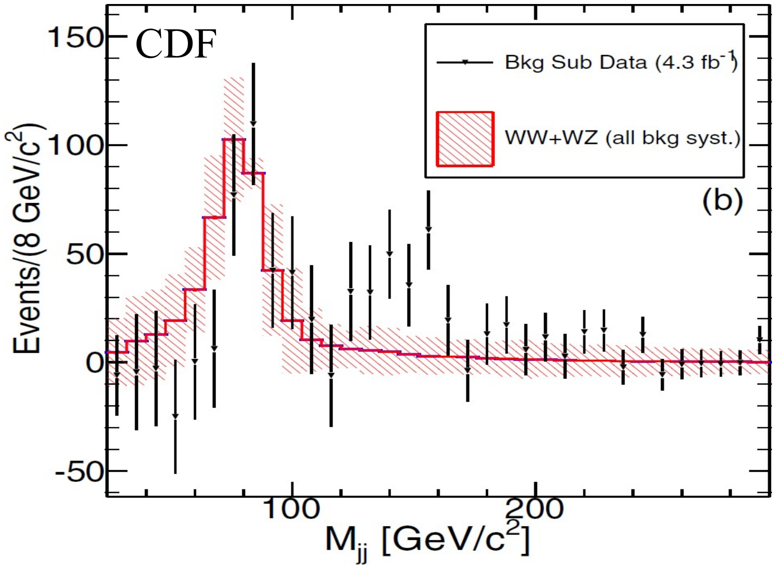
- The dijet mass distributions after fitting the SM processes to the data

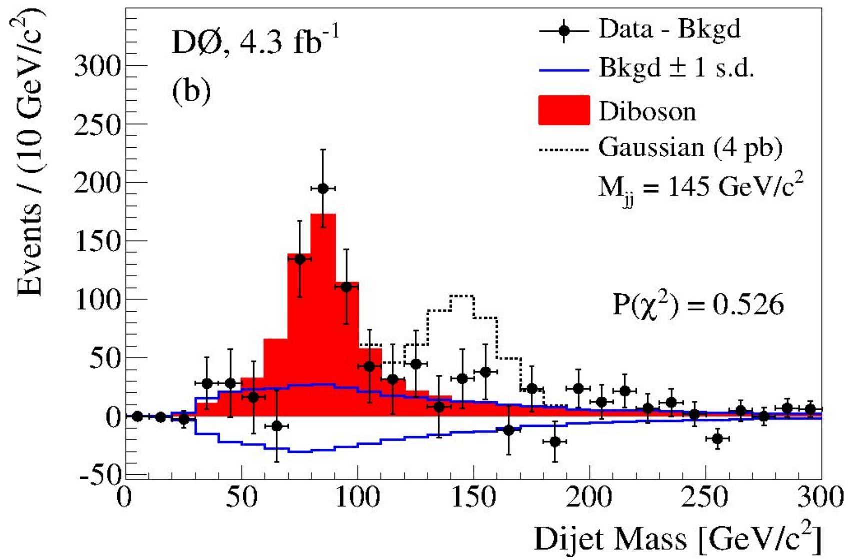


- The DØ data are consistent with the SM prediction

Fit of SM to Data

- The dijet mass distributions after fitting the SM processes to the data

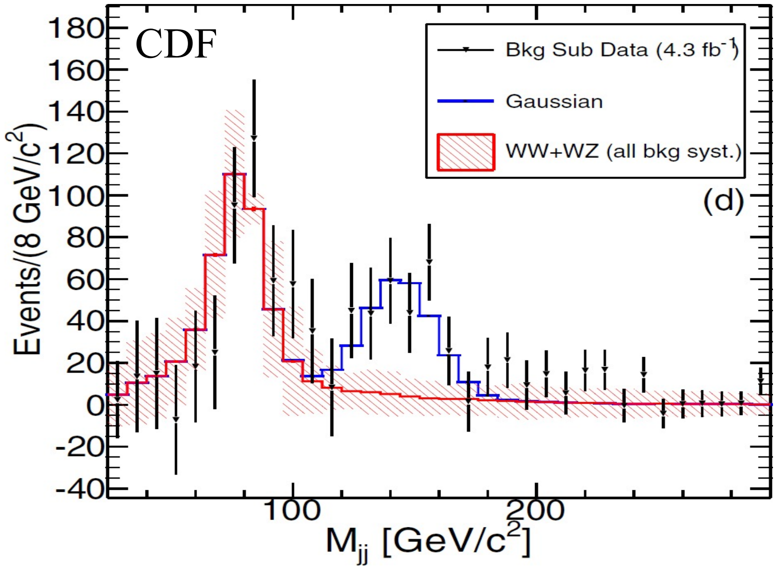


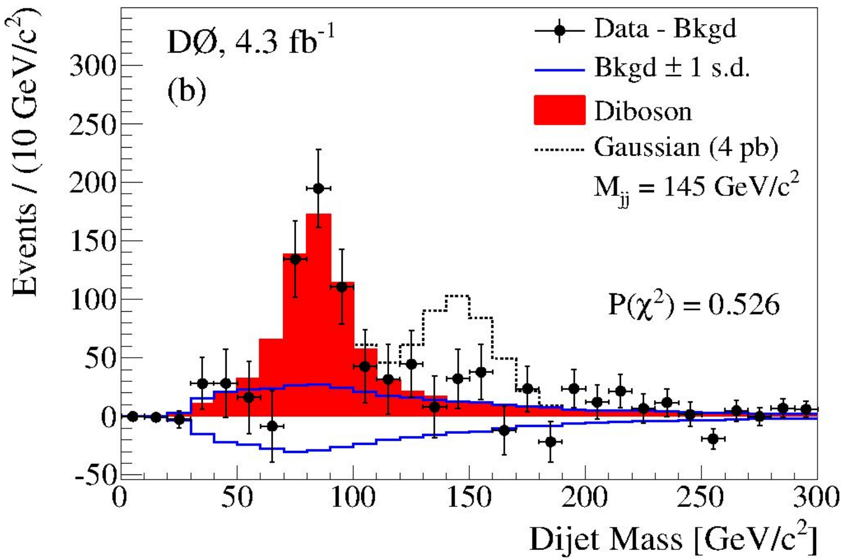


- The DØ data are consistent with the SM prediction

Fit of SM to Data

- The dijet mass distributions after fitting the SM processes to the data

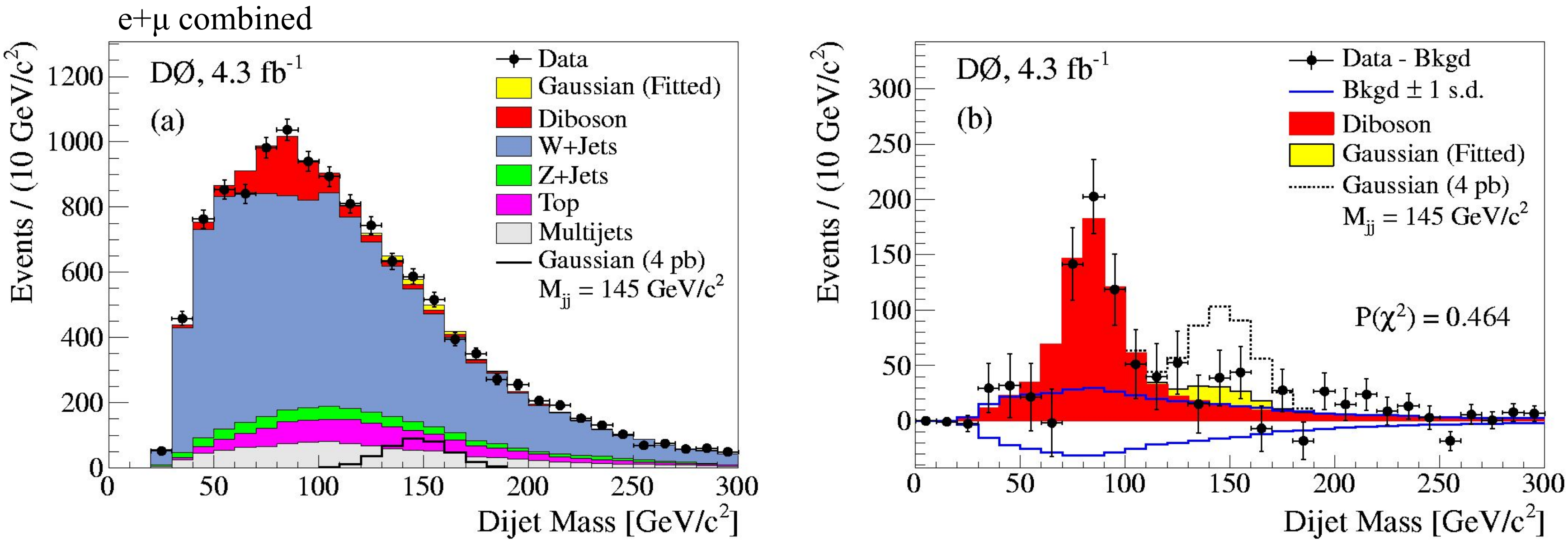




- The DØ data are consistent with the SM prediction
- What if we fit to a resonance like the excess seen by CDF?
  - Quantify whether the DØ data are consistent with such an excess
  - Assume Gaussian distribution, width determined by experimental resolution

Fitting WX

- Fit WX→lνjj template to the data along with SM processes
  - Floating normalizations of WX, diboson, and W+jets



- Fitting a signal is also consistent with no excess
  - How large of an excess is allowed by the DØ data?

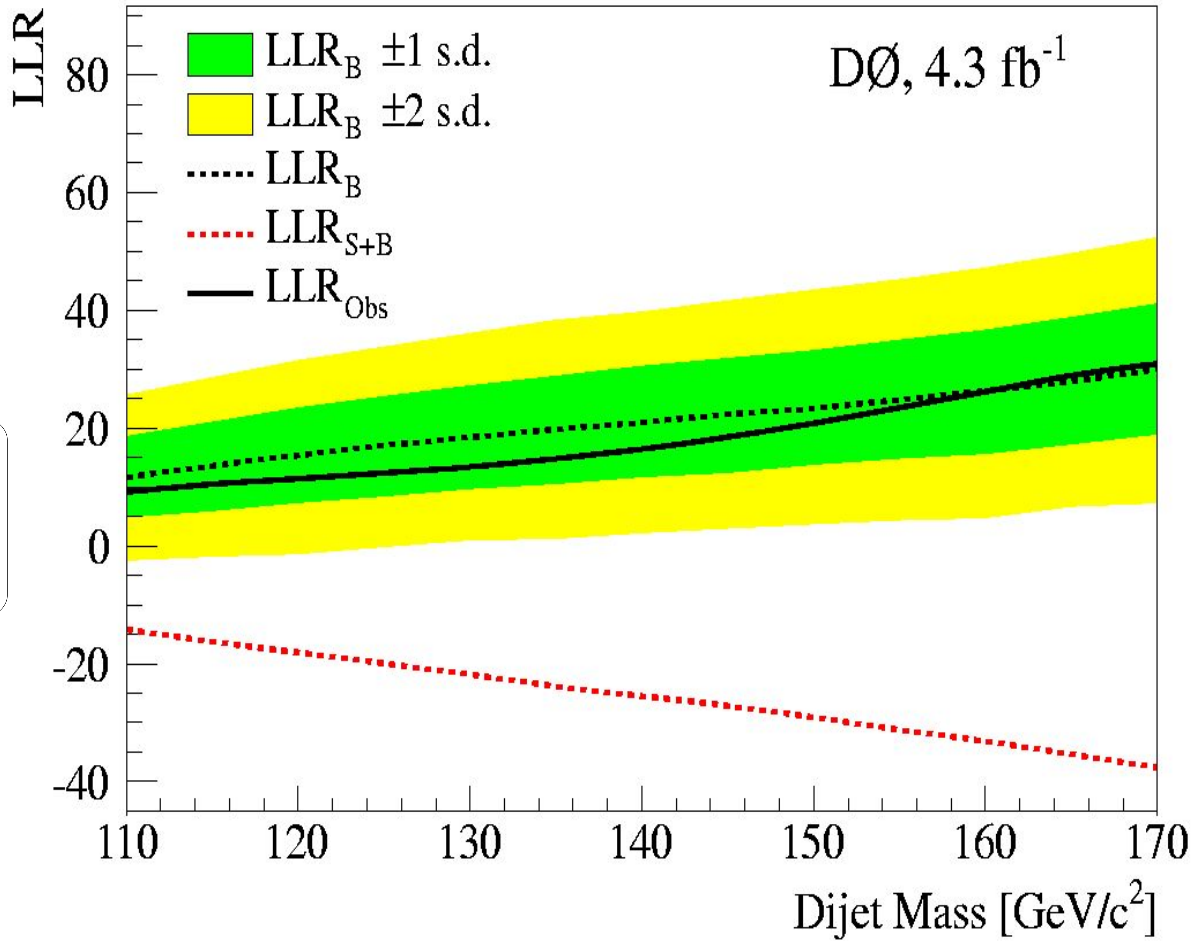
Limit Setting

- Frequentist approach
  - Test statistic: Ratio between S+B fit and B-only fit

$$LLR = -2\log\left(\frac{P(D;S+B)}{P(D;B)}\right) =$$
$$= \chi^2(D|S+B) - \chi^2(D|B)$$

D = observed number of events
S = predicted number of signal events
B = predicted number of background events

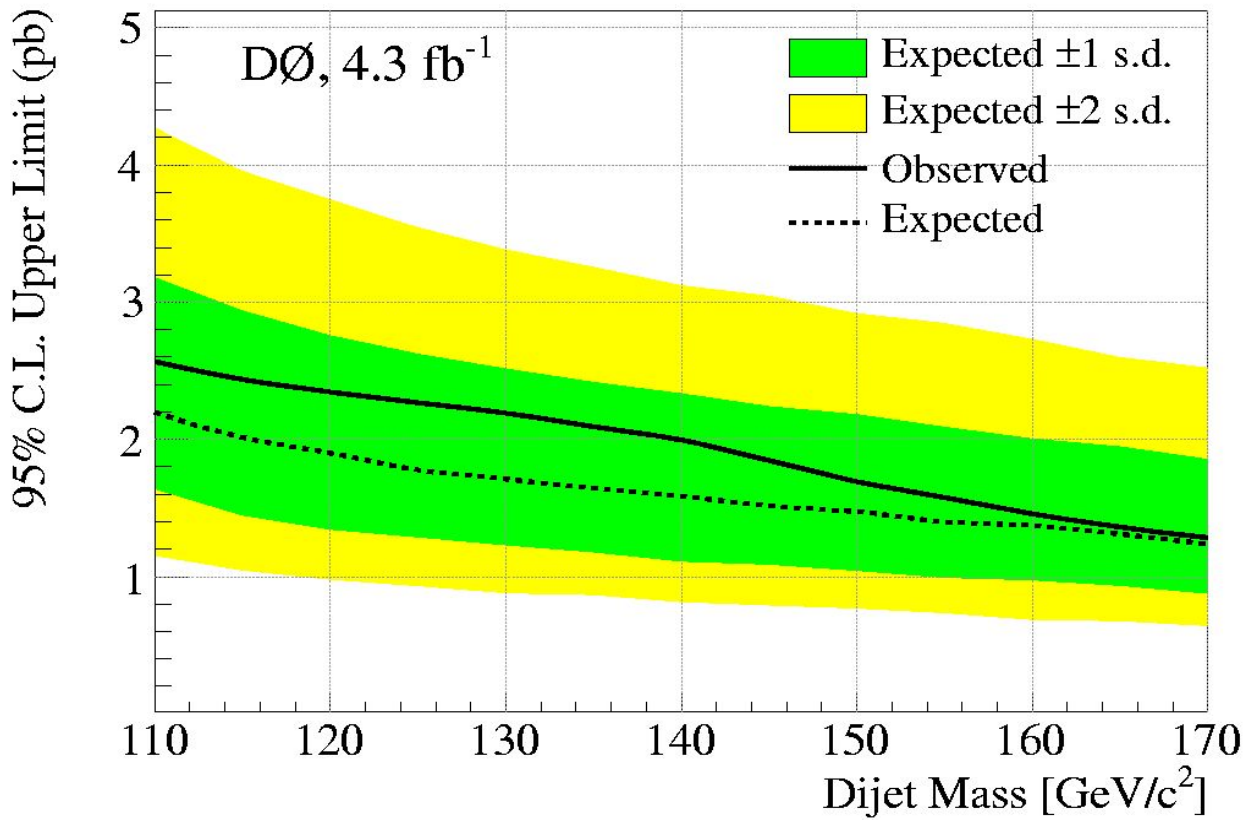
- Compare observed LLR to the predicted LLR distributions over the range of dijet mass



Limits on WX

- 95% CL upper limits on WX→lνjj as a function of reconstructed M_{jj}

  - For Mjj = 145 GeV
    - 95% CL exclusion for cross sections greater than 1.9 pb

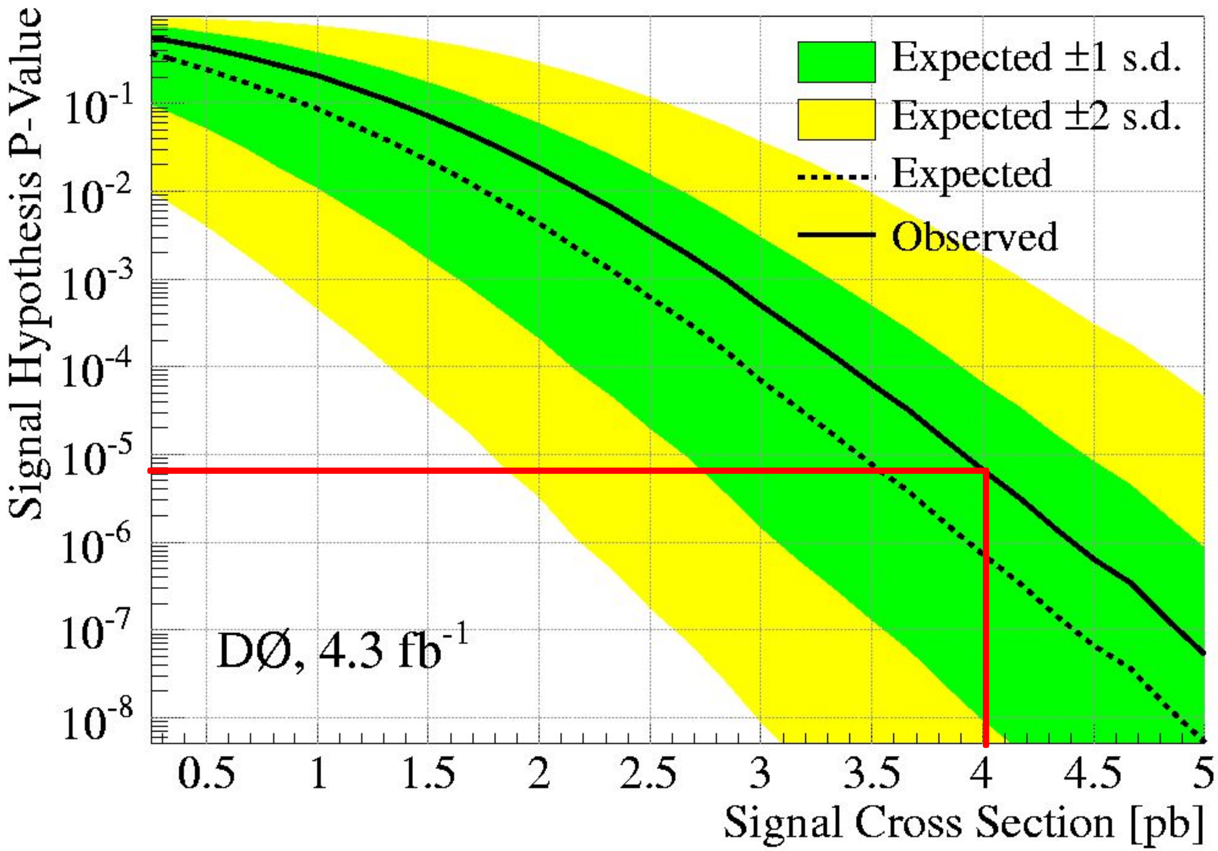


- Can also ask: How strongly is an excess at 145 GeV ruled out?

Limits on WX

- How strongly do the DØ data rule out an excess at 145 GeV?

  - For a cross section of 4 pb as reported by CDF
    - Exclude at 99.999% CL
    - 4 standard deviations



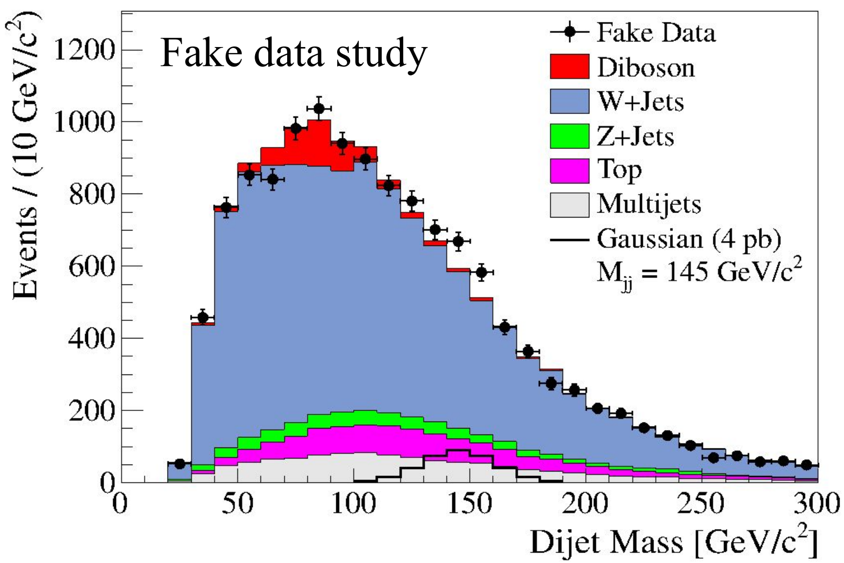
⇒ The DØ data are not consistent with the excess seen by CDF

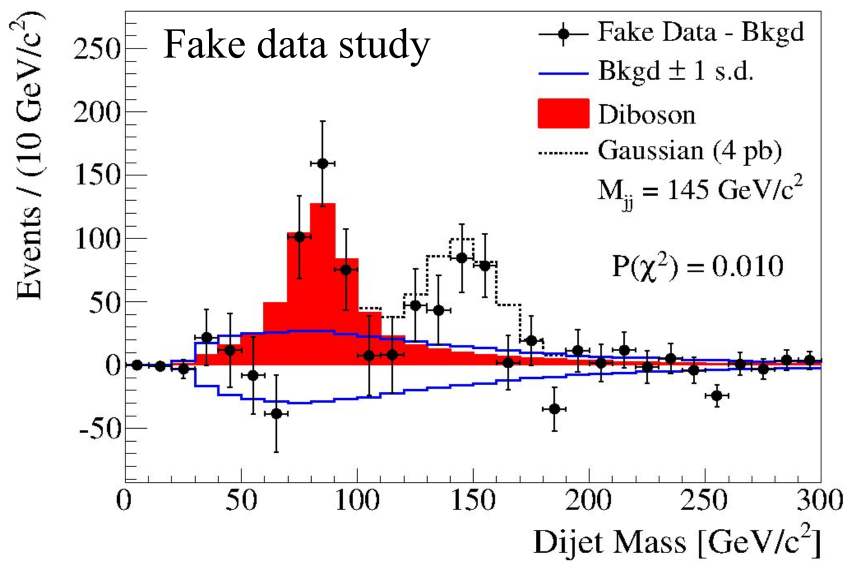
Tests with Signal Injection

- What if there really were a 4 pb dijet mass resonance at 145 GeV?
  - What would it look like?
- Make a signal-injected mock "data" sample
  - Composed of data + WX template @ 145 GeV
  - Confirm that our studies would find that signal

- Fitting the SM processes to the signal-injected data:

e+μ combined

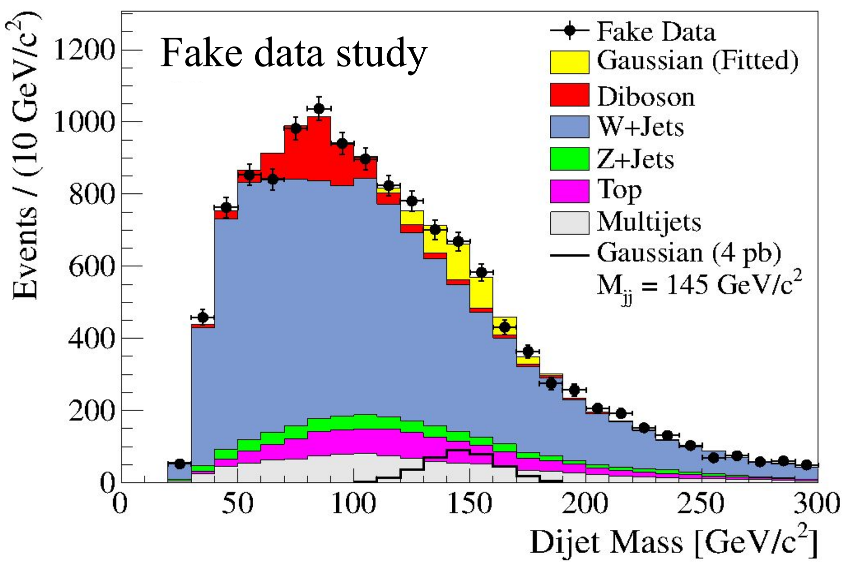


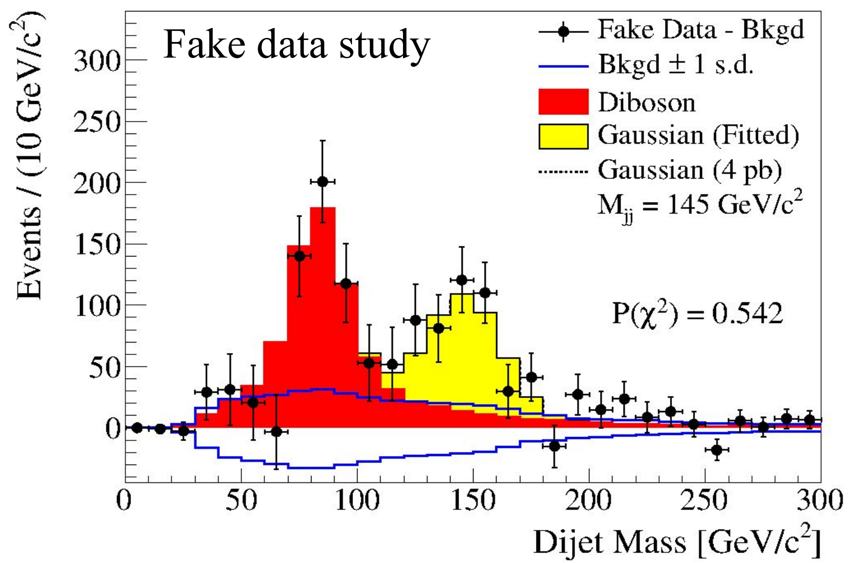


Tests with Signal Injection

- What if there really were a 4 pb dijet mass resonance at 145 GeV?
  - What would it look like?
- Make a signal-injected mock "data" sample
  - Composed of data + WX template @ 145 GeV
  - Confirm that our studies would find that signal
- Fitting the SM + WX template to the signal-injected data:

e+μ combined





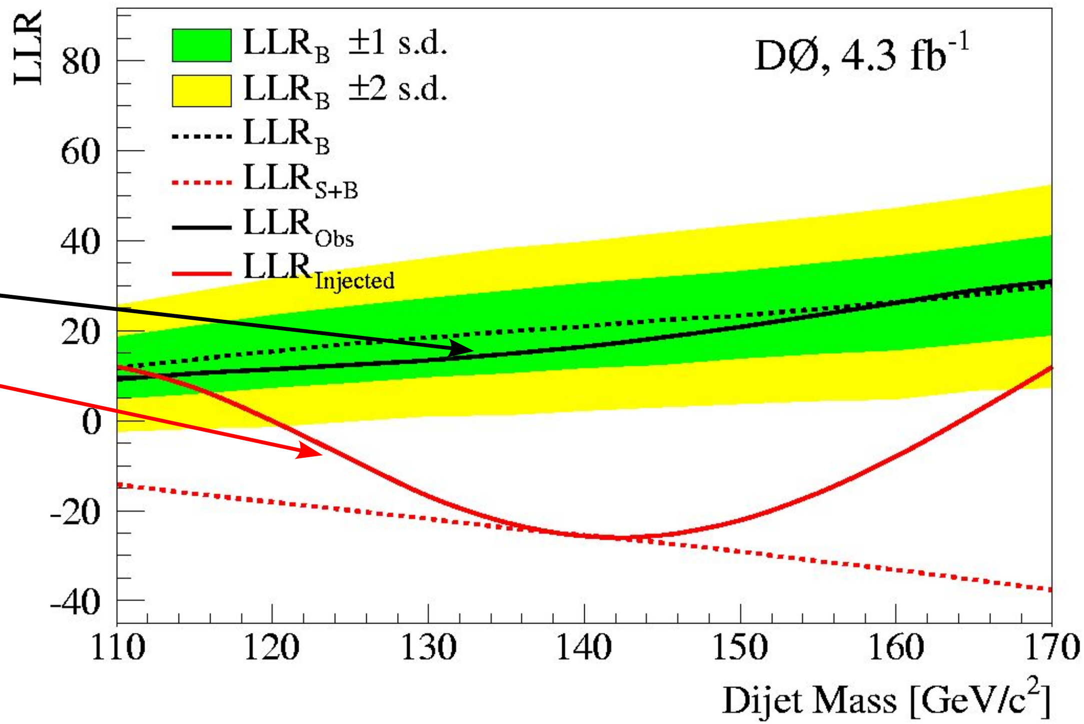
Tests with Signal Injection

- What if there really were a 4 pb dijet mass resonance at 145 GeV?
  - What would it look like?
- Make a signal-injected mock "data" sample
  - Composed of data + WX template @ 145 GeV
  - Confirm that our studies would find that signal

- LLR distribution

Observed Data LLR

Signal-Injected LLR



⇒ We clearly would have seen a 4 pb excess if it were there

Differences Between CDF and DØ

- Detector and Object Reconstruction
  - Different jet cone algorithms

- Monte Carlo generators

	CDF:	DØ:
PDF set:	CTEQ5L	CTEQ6L1
Pythia version:	v6.326	v6.409
Pythia tune:	tune A	"DØ tune A" (like tune A, but for CTEQ6L1)
Alpgen version:	v2.1	v2.11_wcfix

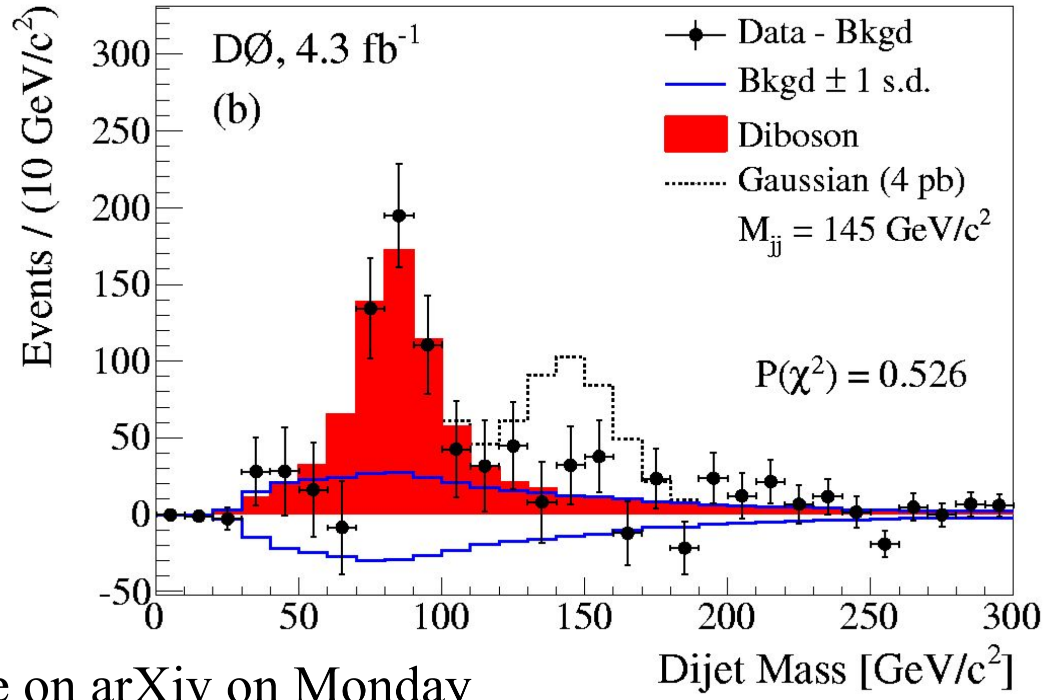
- Alpgen parameters and uncertainties
  - DØ assigns uncertainties on kinematic modeling, parton-jet matching, parton shower model, renormalization/factorization scale, PDF

Summary

- Used the same selection as the CDF analysis
- Studied the dijet mass spectrum in the range 110 – 170 GeV

  ⇒ DØ data are consistent with the SM predictions

  - We verified that:
    - We would see a 4 pb excess if it were in the DØ data
    - We get consistent results with or without kinematic Alpgen corrections
  - For a resonance at 145 GeV
    ⇒ Exclude 1.9 pb at 95% CL
    ⇒ Exclude 4 pb at 99.999% CL



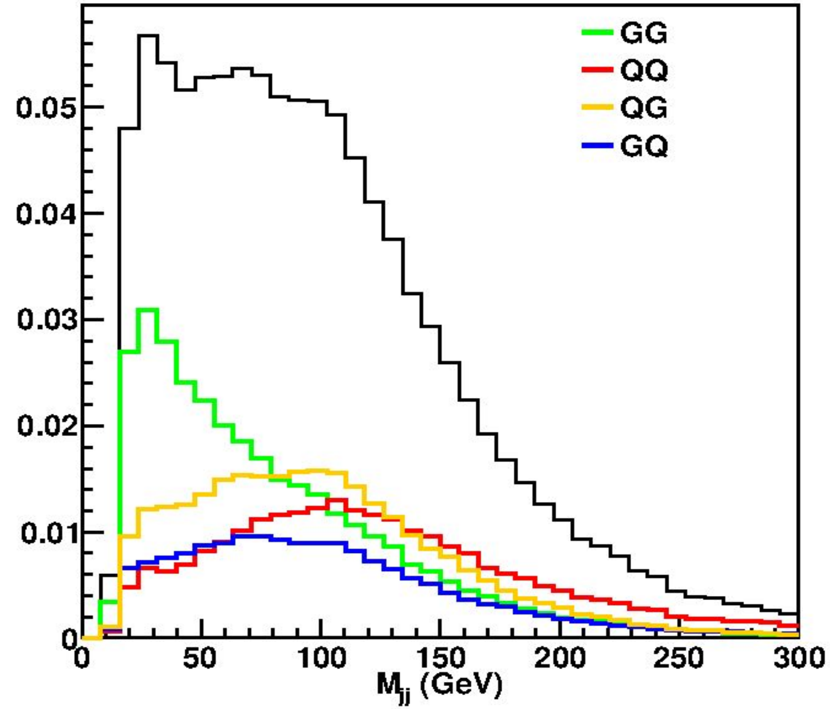
Submitted to Phys. Rev. Lett., available on arXiv on Monday
http://www-d0.fnal.gov/Run2Physics/WWW/results/final/HIGGS/H11B/

Thank you

Event Selection

- Jets:
  - Reconstruction:
    - DØ iterative mid-point cone algorithm with radius R=0.5
    - Must be a hadronic shower and not contain noisy calorimeter cells
    - At least two tracks originating from the primary interaction point
  - Jet Energy Scale
    - Measured in γ+jet and dijet events
    - Correct energy to particle-level
      - Correct for detector response, out of cone showering, overlap with pileup energy
  - Relative Data/MC Correction
    - Measured in Z+jet events
    - Different correction depending on quark vs. gluon content

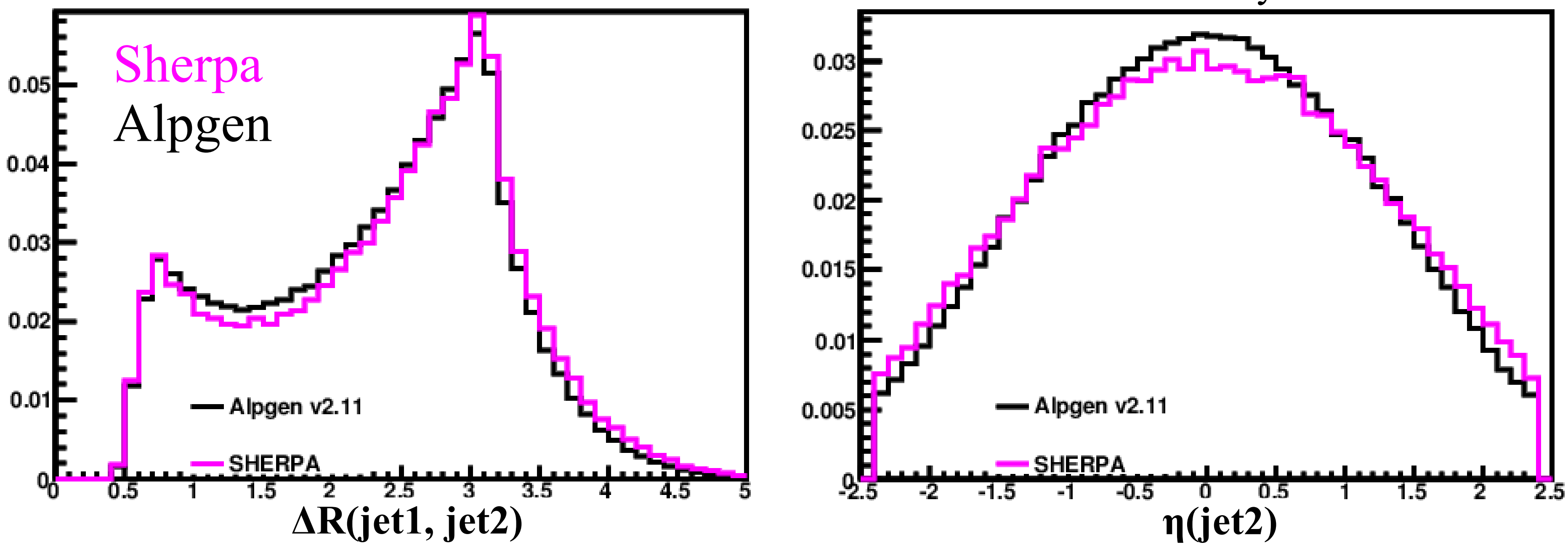
Plot courtesy of Adam Martin



W+jets Modeling

- We know that Alpgen is not the final answer in modeling W+jets
  - Different generators have different predictions

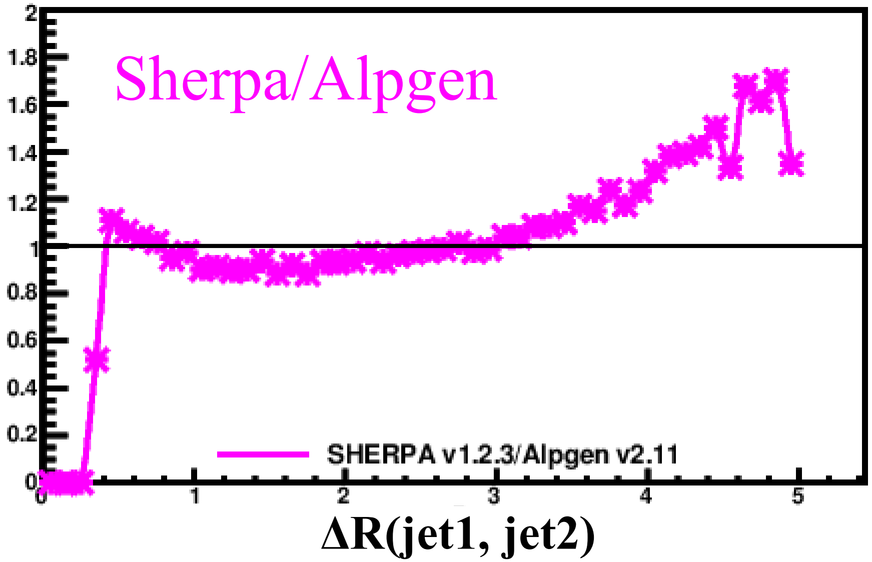
Plots courtesy of Adam Martin

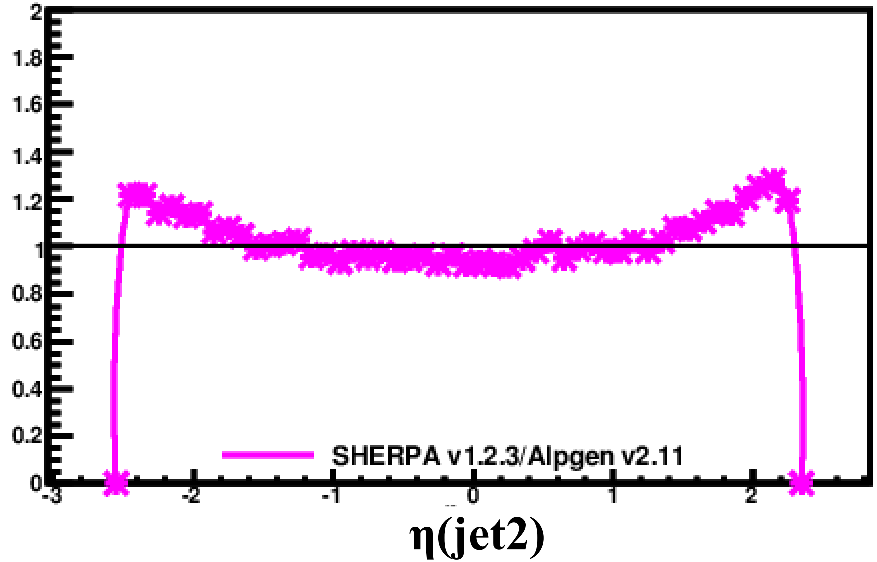


W+jets Modeling

- We know that Alpgen is not the final answer in modeling W+jets
  - Different generators have different predictions

Plots courtesy of Adam Martin



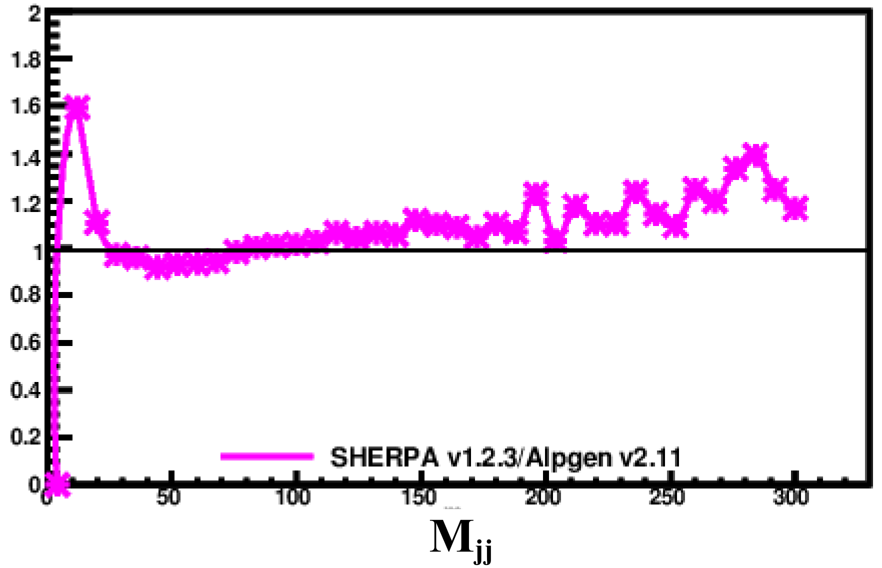


  - In analyses with looser cuts (*e.g.*, WH→lνbb) we see clear discrepancies of exactly this type

W+jets Modeling

- We know that Alpgen is not the final answer in modeling W+jets
  - Different generators have different predictions

Plots courtesy of Adam Martin



  - In analyses with looser cuts (*e.g.*, WH→lνbb) we see clear discrepancies of exactly this type

W+jets Modeling

- In other analyses we use data-driven corrections to fix the modeling of these variables with known discrepancies between predictions
  - Jet η, ΔR(jet1, jet2), and $p_T(W)$
  - However, the relatively tight selection used in this analysis reduces the necessity for these corrections
  - Removes much of the problematic phase space (low $p_T(W)$)

- The CDF analysis did not apply corrections to the Alpgen modeling
  - ⇒ To parallel their analysis, we perform the analysis without these corrections
  - Still including uncertainties one the modeling of these variables

- However, to show that these corrections would not alter the conclusion
  - ⇒ We also present results with these corrections

Fit of SM to Data

- Fit dijet mass distributions for all SM processes to the data
  - Construct a χ^2 function from the ratio of Poisson likelihoods and include prior information on the systematic uncertainties

$$\chi^2(\theta, S, B; D) = 2\sum_{i=0}^{N_{bins}} (B_i + S_i - D_i) - D_i \ln\left(\frac{B_i + S_i}{D_i}\right) + \sum_{k=0}^{N_{syst}} \theta_k^2$$

D = observed number of events

$S(\theta_k)$ = predicted number of signal events

$B(\theta_k)$ = predicted number of background events

θ_k = number of standard deviations systematic k has been pulled away from nominal

Fit of SM to Data

- Fit dijet mass distributions for all SM processes to the data
  - Construct a χ^2 function from the ratio of Poisson likelihoods and include prior information on the systematic uncertainties

$$\chi^2(\theta, S, B; D) = 2\sum_{i=0}^{N_{bins}} (B_i + S_i - D_i) - D_i \ln\left(\frac{B_i + S_i}{D_i}\right) + \sum_{k=0}^{N_{syst}} \theta_k^2$$

D = observed number of events

$S(\theta_k)$ = predicted number of signal events

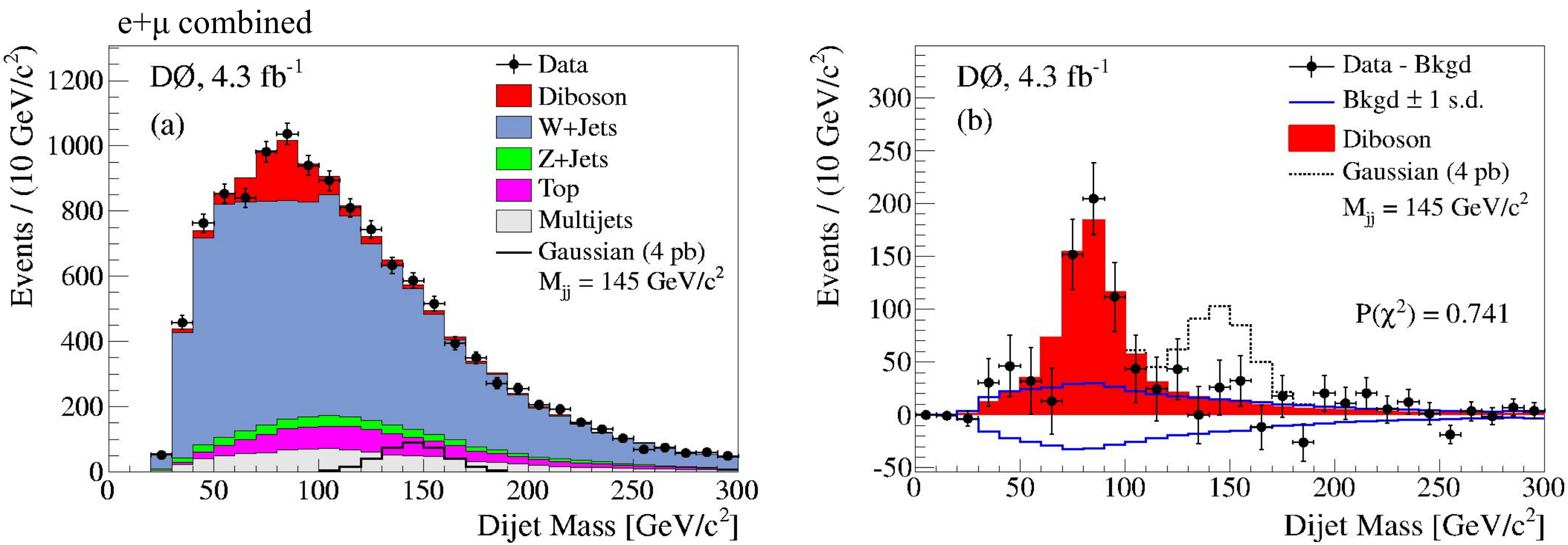
$B(\theta_k)$ = predicted number of background events

θ_k = number of standard deviations systematic k has been pulled away from nominal

- Templates can vary within systematic uncertainties, constrained by Gaussian priors
- Can "float" a parameter by removing the θ^2 prior constraint
  - Float cross sections for Diboson and W+jets contributions

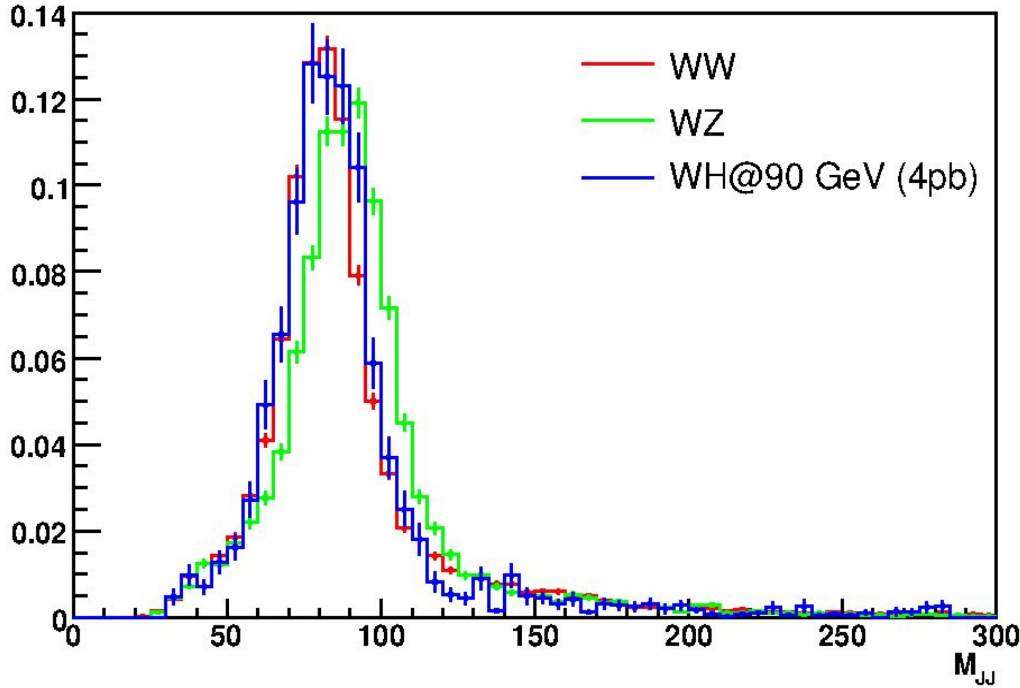
Fit of SM to Data

- The dijet mass distributions after fitting the SM processes to the data
  - **With** Alpgen modeling corrections applied



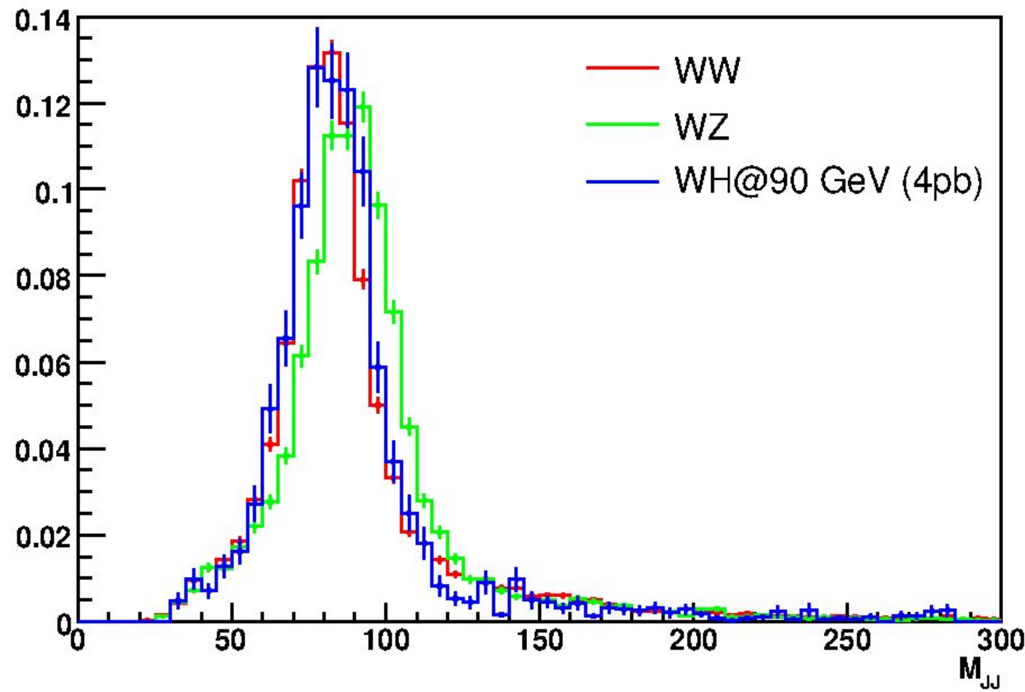
Modeling $WX \to l\nu jj$

- Assume a Gaussian distribution in dijet mass with a width determined by the DØ experimental resolution
  - A simplified model, but a reasonable approximation for a narrow resonance
  - Apples-to-apples comparison to CDF's claim of the excess being consistent with a cross section of ≈ 4 pb
- Width estimated from $WW \to l\nu jj$
  - $\sigma_{jj} = \sigma_{W\to jj} \times \sqrt{m_{jj} / m_{W\to jj}}$
    - For M_{jj} = 145 GeV ⇒ σ_{jj} = 15.7 GeV



Modeling WX→lνjj

- Estimate efficiency from WH→Wbb
  - Assume BR(X→jj) = 1.0
  - Use efficiency from m_H=150 GeV for the Gaussian template with mean of 145 GeV
    - To be consistent with CDF
- Assign systematic uncertainties
  - Jet energy scale uncertainty changes mean by ±1.5 %
  - Jet Resolution uncertainty changes normalization by ±5% and width by ±3%



Limit Setting

- Frequentist approach
  - If the experiment is repeated many times, what fraction would find a more extreme result?
    - Need to simulate repeating the experiment many times
      - Generate ensembles of pseudo-experiments allowing statistical and systematic fluctuations
      - Two hypotheses: Background only and Signal+Background

Limit Setting

- Frequentist approach
  - If the experiment is repeated many times, what fraction would find a more extreme result?
    - Need to simulate repeating the experiment many times
      - Generate ensembles of pseudo-experiments allowing statistical and systematic fluctuations
      - Two hypotheses: Background only and Signal+Background
  - Test statistic: Ratio between S+B fit and B-only fit

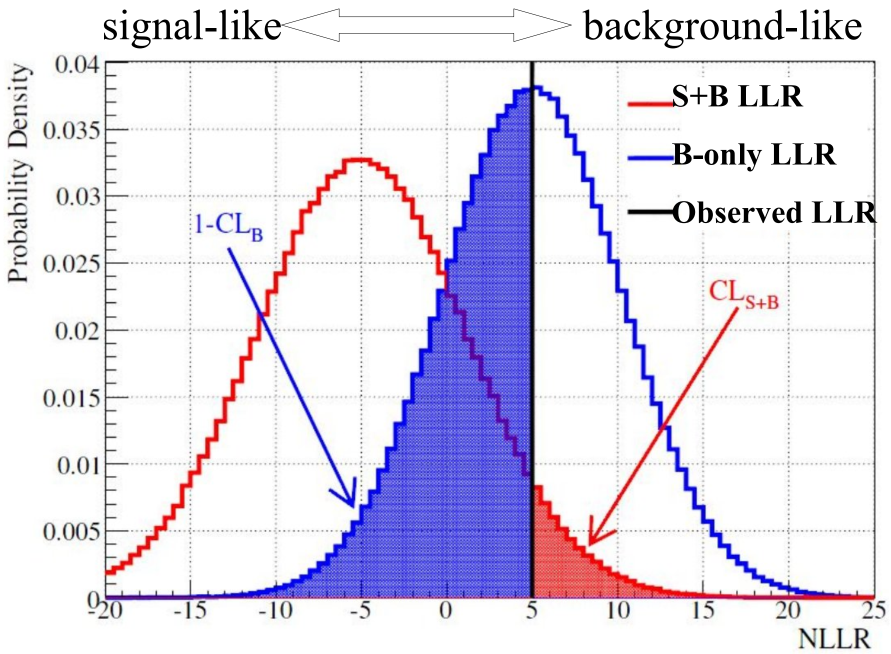
$$LLR = -2\log\left(\frac{P(D;S+B)}{P(D;B)}\right) = \chi^2(D|S+B) - \chi^2(D|B)$$

D = observed number of events
S = predicted number of signal events
B = predicted number of background events

- Construct the LLR probability distribution for each hypothesis and see how they compare to the observed LLR

Limit Setting

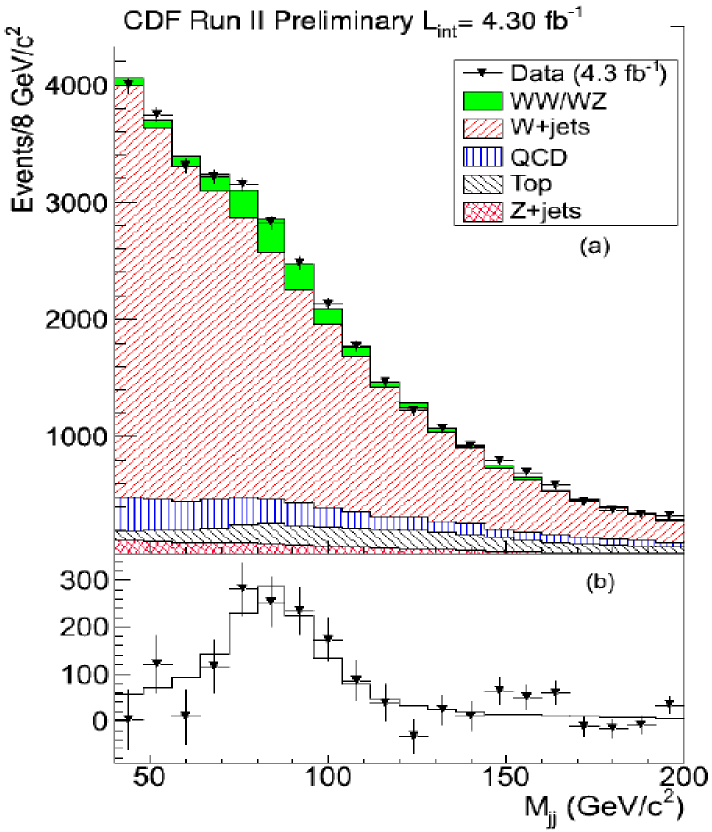
- Set upper limits on the cross section for this model
  - CL_{S+B} = probability of measuring a more background-like result when signal is actually present
  - $1 - CL_B$ = probability of measuring a more signal-like result when there is actually no signal



- 95% CL limit
  - Could use CL_{S+B}
    - The cross section where $CL_{S+B} = 5\%$
    - But what if $1 - CL_B$ is also small?
      - The observed LLR is also inconsistent with the background only hypothesis!
  - Use $CL_S = CL_{S+B}/(1 - CL_B)$
    - The cross section where $CL_S = 5\%$
    - Protects against setting limit too tight when the background is poorly modeled

CDF Dijet Mass Excess

Cross sections measurement



- Use a fit to dijet mass to extrapolate the WW/WZ contribution
- We estimate 1582 ± 275(stat.) $\pm$ 107(syst.) events for a significance of 5.2σ.
- The resulting cross section is

  $\sigma(WW/WZ) = 18.1 \pm 3.3$(stat.)$\pm 2.5$(syst.) pb

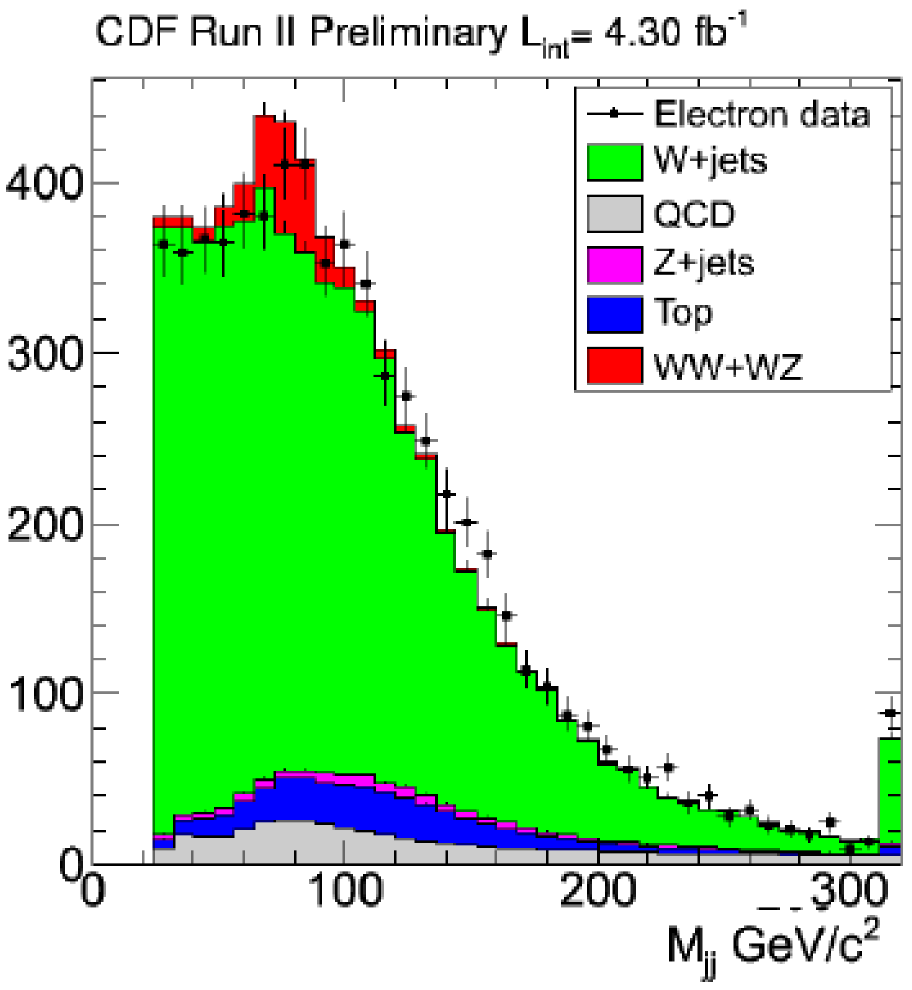
  that is in agreement with SM expectation (15.9 ± 0.9 pb).

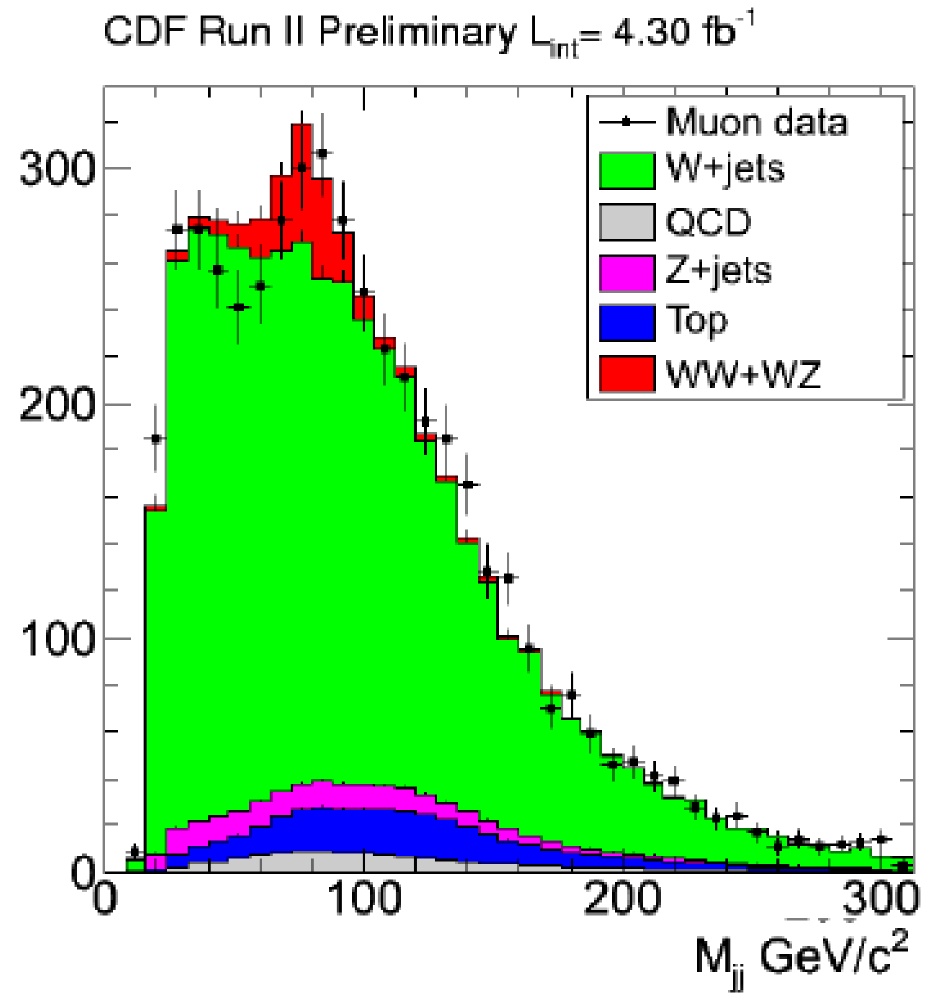
CDF Dijet Mass Excess

Moving to different kinematical region

- Using exactly the same kinematical cuts as the diboson analysis but:
- We require both jets to have $E_T > 30$ GeV
  1. Energetic jets are measured with better accuracy.
  2. Modeling in this region is expected to be more accurate
  3. A possible heavier particle would be characterized by more energetic jets
- Sample modeling using same processes with different relative contribution
- All cuts chosen “a priori”

CDF Dijet Mass Excess

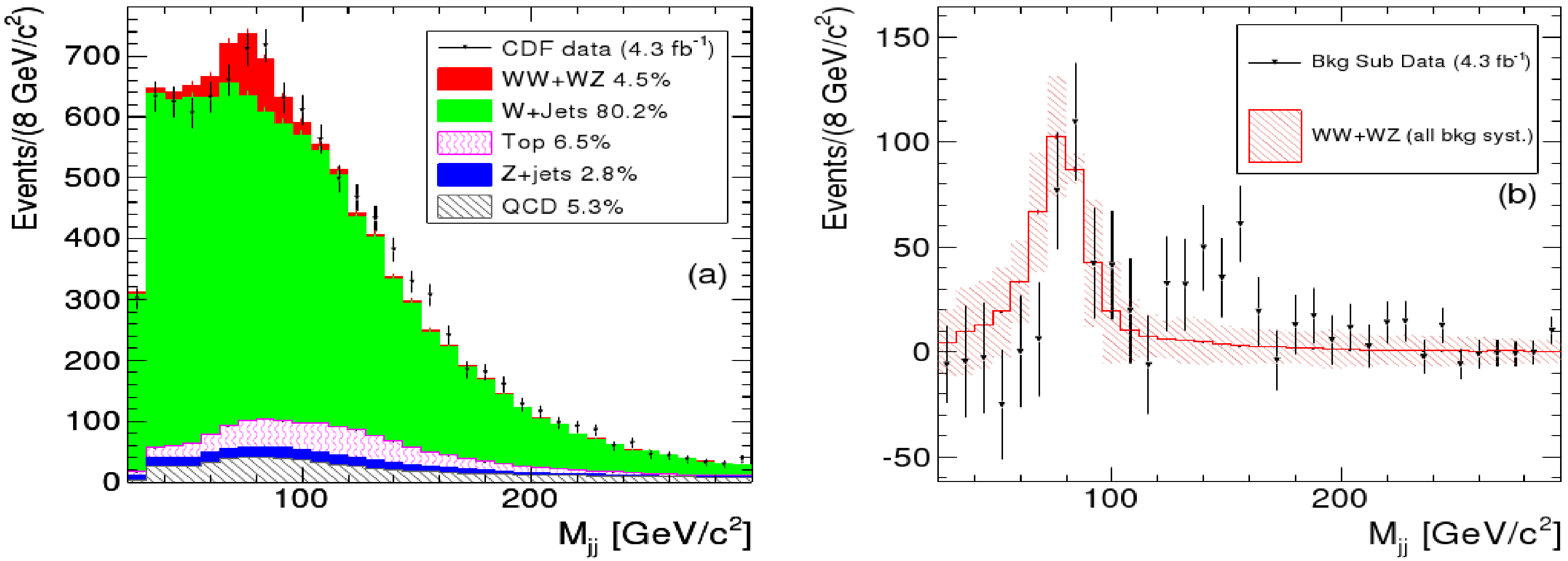




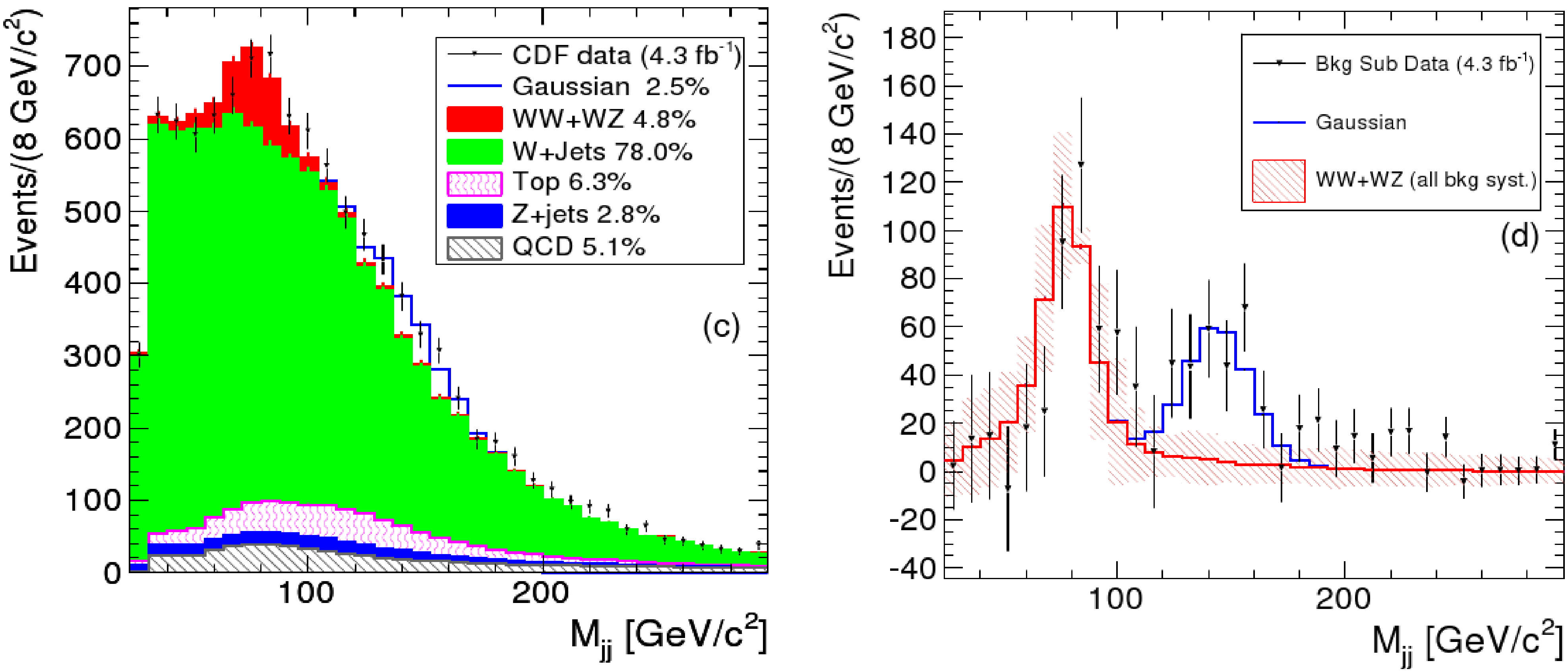
CDF Dijet Mass Excess

Fitting procedure

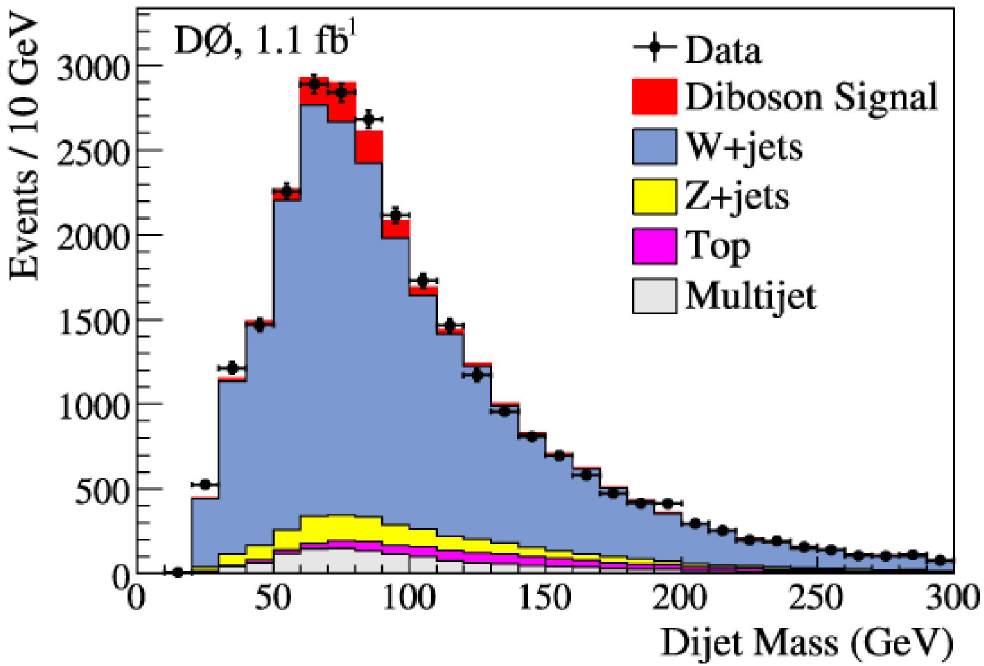
- Combined χ^2 fit to the dijet mass distribution in electron and muon samples.
- 5 templates:
  1. $W+$ jets (uncostrained, normalization determined from the fit)
  2. QCD (normalization constrained to its fraction with 25 % error)
  3. $Z+$ jets (normalization constrained to the measured cross section)
  4. top & single top (normalization constrained to the theoretical cross section)
  5. WW/WZ (normalization constrained to the theoretical cross section)

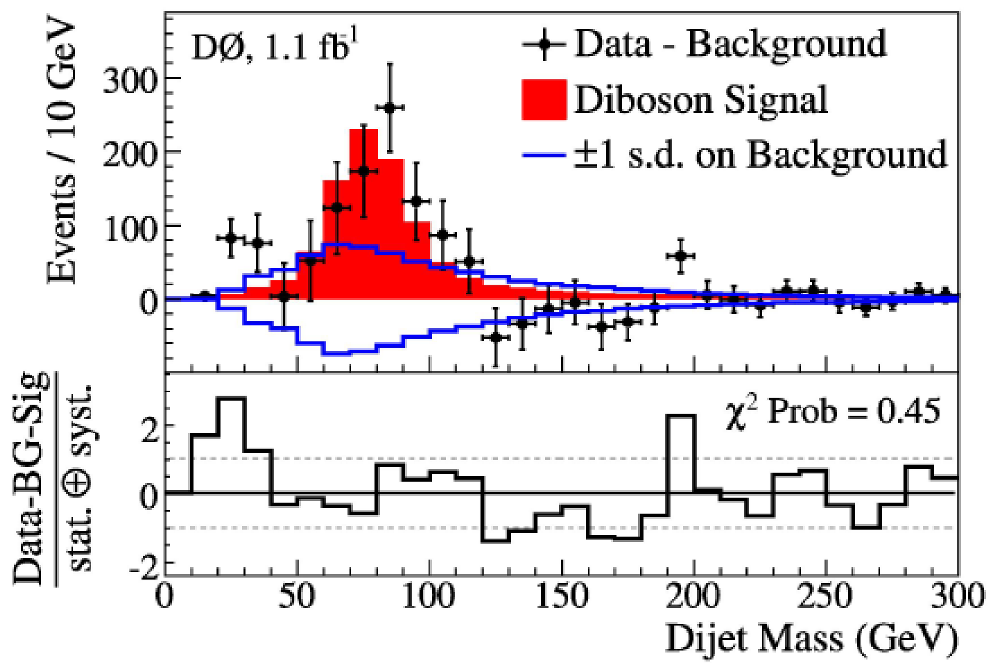


CDF Dijet Mass Excess



DØ lνqq Measurement





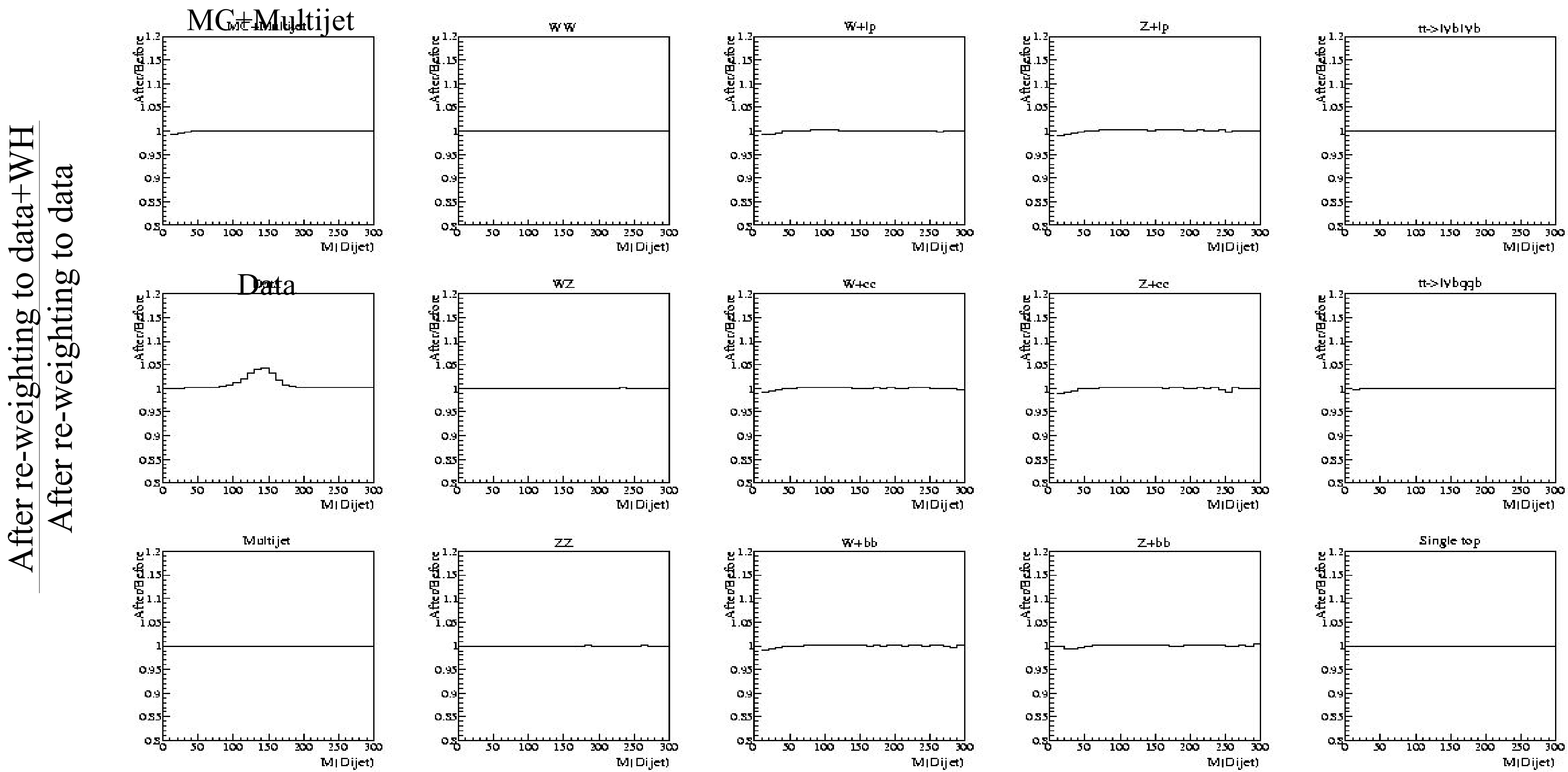
W+jets Modeling

- A concern: Could these data driven re-weightings hide an excess?
  - Instead of correcting to data, correct to data+WH (m_H = 150, σ = 4 pb)
  - See if the resulting re-weighted MC has filled in the WH resonance in data
    - Plot the ratio of MC:

$$\frac{\text{After correcting to data+WH}}{\text{After correcting to data}}$$

Check for Re-weighting Bias

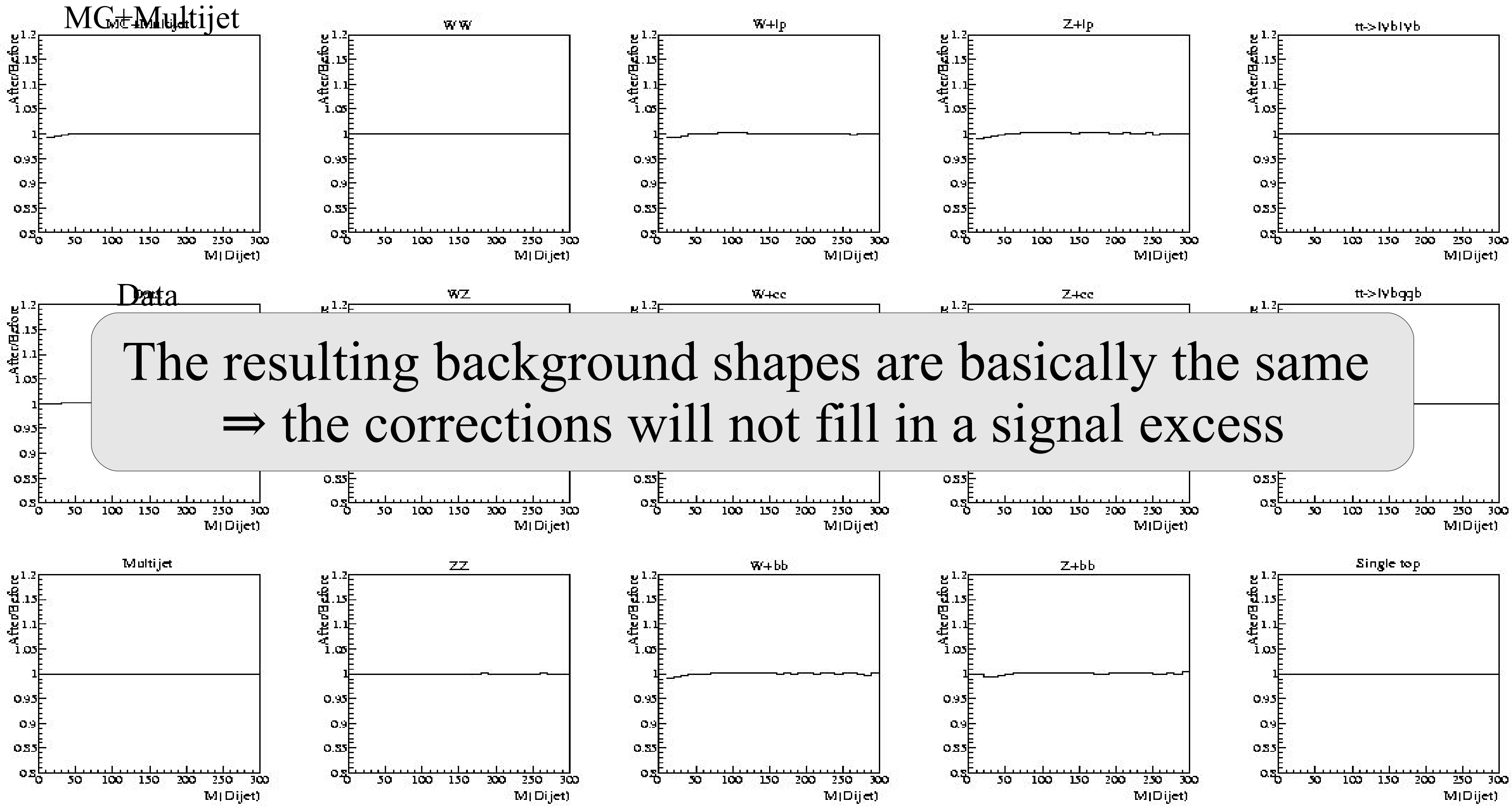
- Would data driven re-weightings cause background to fill in the excess?



Check for Re-weighting Bias

- Would data driven re-weighting cause background to fill in the excess?

$$\frac{\text{After re-weighting to data+WH}}{\text{After re-weighting to data}}$$



The resulting background shapes are basically the same
⇒ the corrections will not fill in a signal excess