Study of M_{jj} Mass Spectrum and the CDF anomaly in W($\rightarrow$lν)+jj data at CMS

Kalanand Mishra

Fermilab

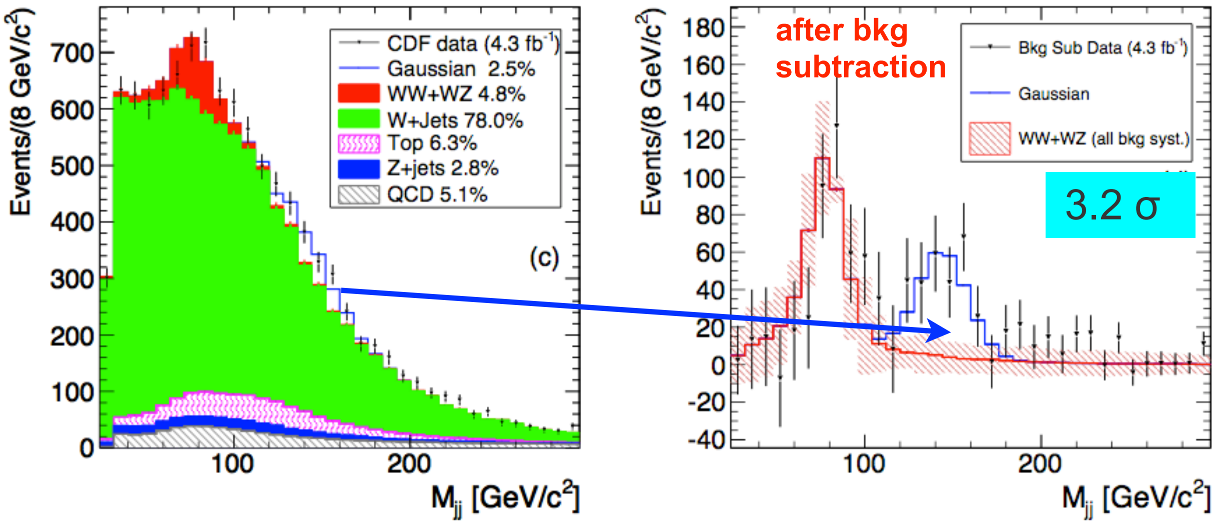
On behalf of CMS Electroweak Mjj subgroup

with inputs from lvjj team members

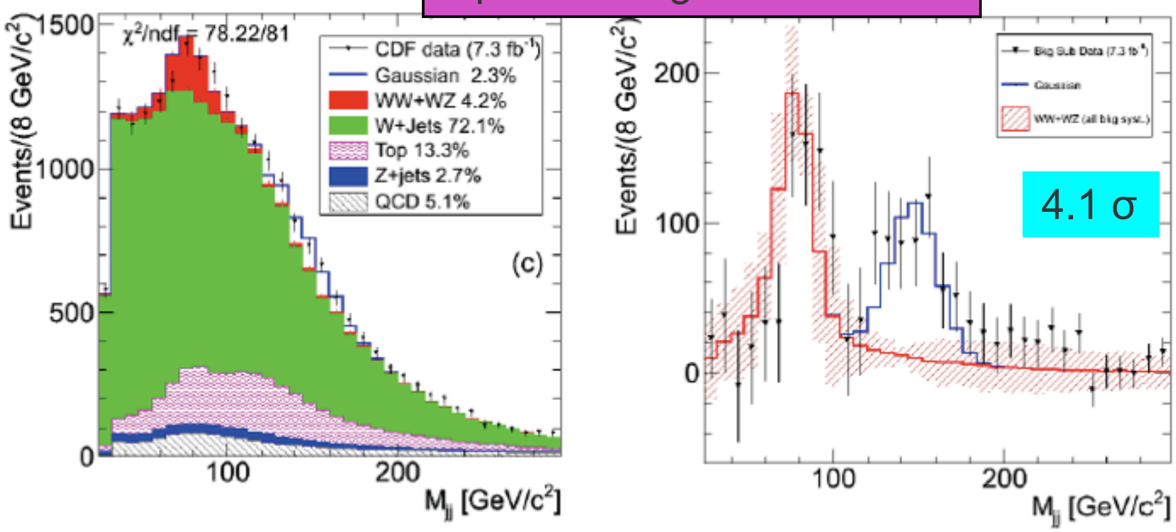
Disclaimer: CMS results/plots presented in these slides are very very preliminary and are NOT approved for presentation outside the CMS.

USCMS meeting, August 5, 2011

Thing that started it all: CDF saw anomaly in M_{jj}



Update using 7.3 fb^{-1} data



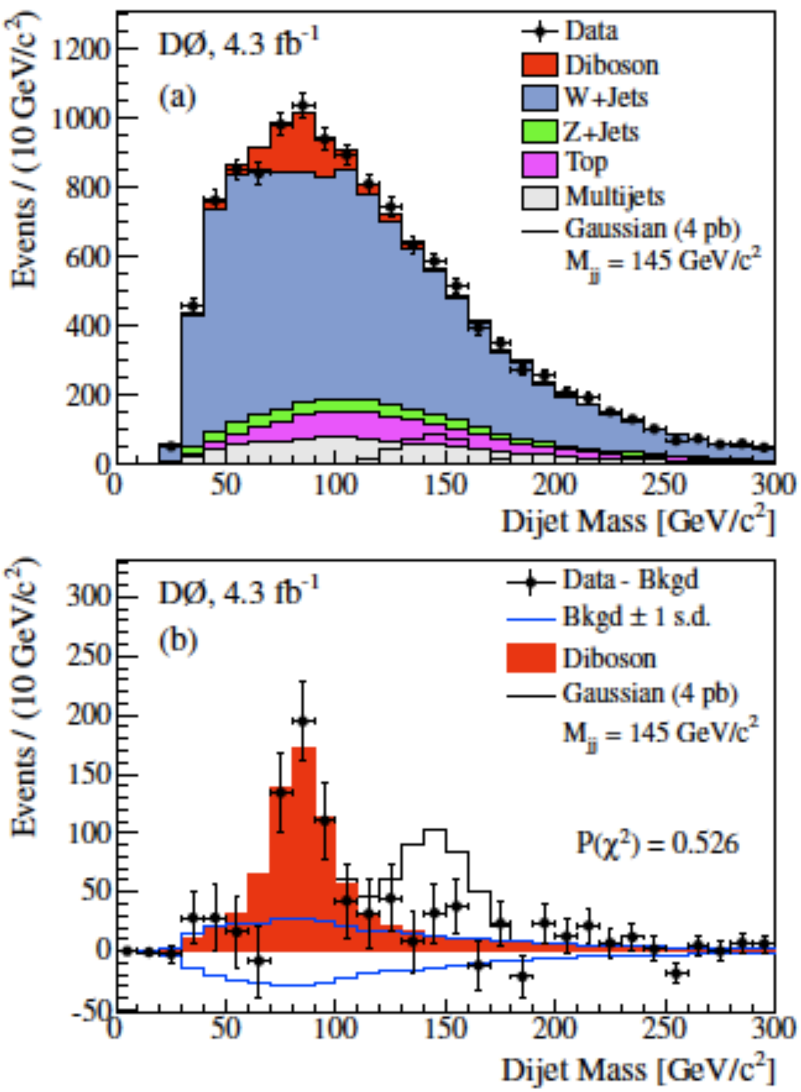
- ✦W+jj data doesn't have the featureless falloff of dijet mass spectrum
- ✦CDF finds an excess of 253 events, peaked at 145 GeV, width = 15 GeV
- ✦Significance 3.2σ, prod cross section 4 pb

arXiv: 1101.6079, Phys. Rev. Lett. 106:171801 (2011)

http://www-cdf.fnal.gov/physics/ewk/2011/wjj/

Significance has been growing with more data ! Statistical significance is not in doubt, everything else is.

D∅ doesn't confirm this anomaly



arXiv: 1106.1921, Phys. Rev. Lett. 107:011804 (2011)

✦W+jj data from D∅ DOES show the featureless falloff of dijet mass spectrum in the range 110−170 GeV

✦ D∅ doesn't find any excess peaked at 145 GeV

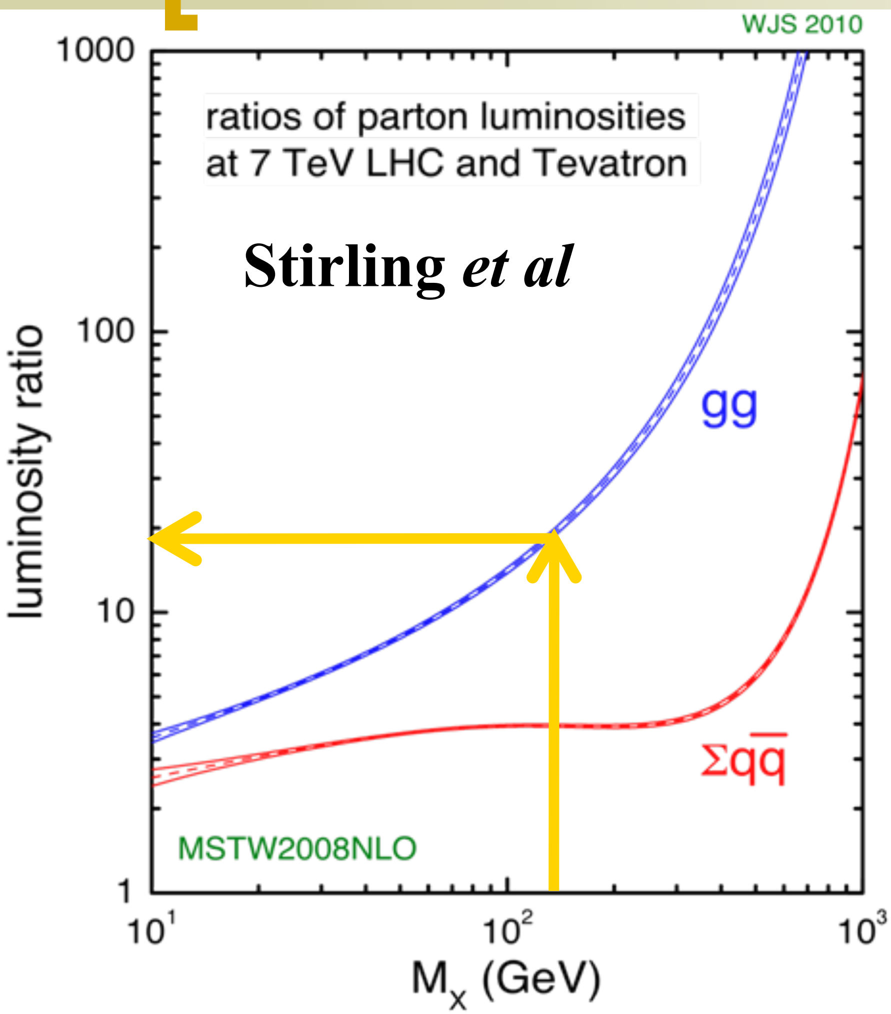
✦Excludes production cross section 4 pb at 99.9999% CL and 1.9 pb at 95% CL

✦D∅ data is fully consistent with the Standard Model W+jets and diboson production in this region

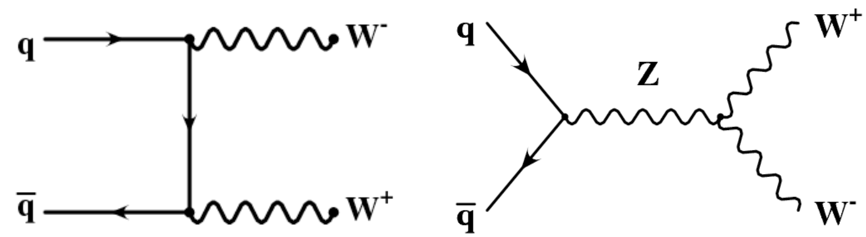
See details at:

http://www-d0.fnal.gov/Run2Physics/WWW/results/final/HIGGS/H11B/

LHC vs Tevatron: partonic luminosity hurts



$q\bar{q} \rightarrow$ WW, WZ cross section at 7 TeV is ~ 3.5 times that at 2 TeV



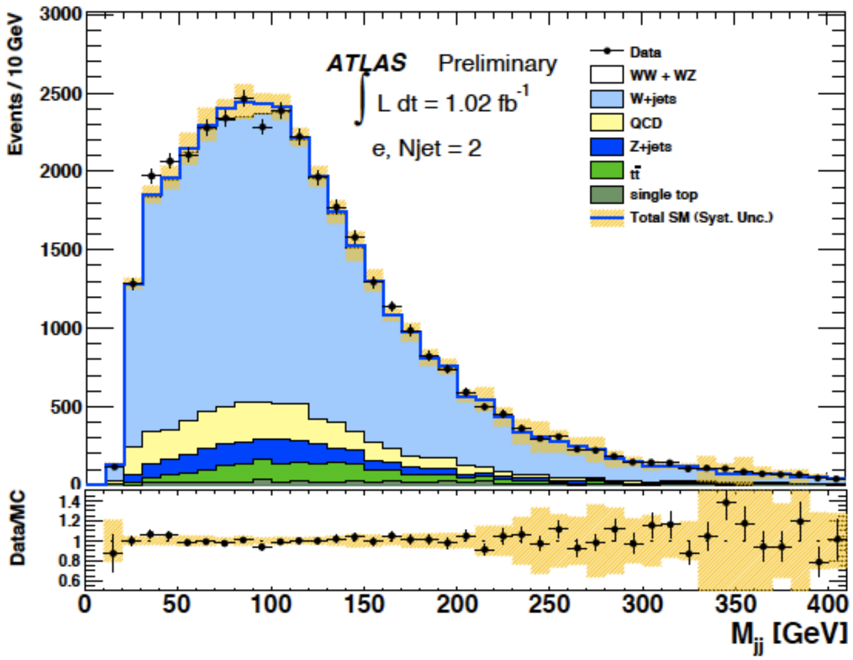
Major backgrounds are W+jets, single top $\&\ t\bar{t}$, QCD multi-jet etc. which rise sharply due to rise in qg and gg cross sections

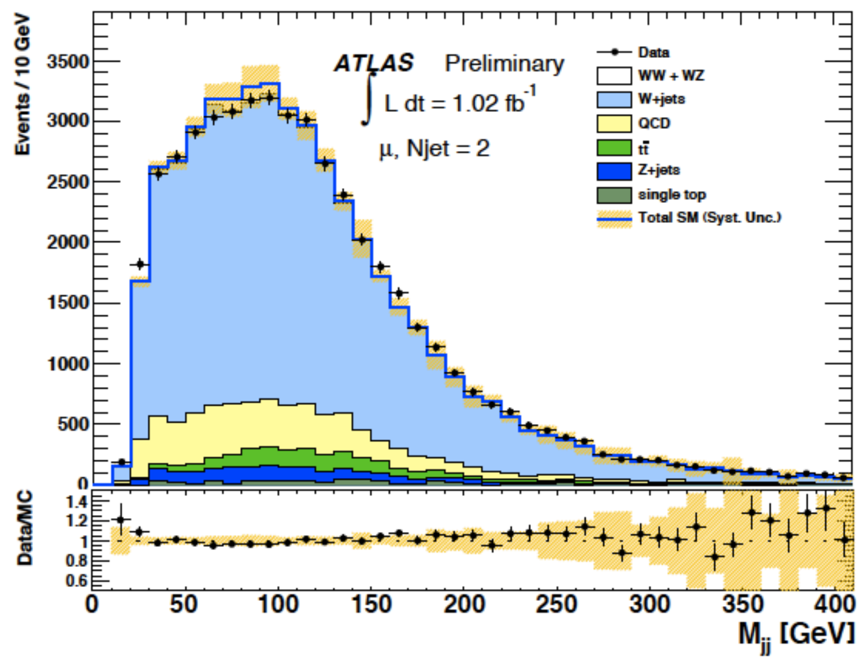
$\Rightarrow$ Small signal, worse S/N

With 3x more signal and 20x more background would have 6 times worse S/B with identical detector. We are doing much better than that.

Life is hard if looking for qqbar signal at LHC

ATLAS result shown in EPS





- Get swamped by W+jets
- **See no diboson peak**, nothing other than W+jets: S/B →0
- Large syst uncertainty, do not even bother to show bkg-subtracted plot
- Instead plot data/MC distribution which obscures any discrepancy

Clearly much worse than Tevatron experiments

Acceptance thresholds and trigger

✦$W \rightarrow l\nu$ reconstruction
- Muon: $p_T > 25$ GeV, $|\eta| < 2.1$, reconstructed as both global & tracker muon
- Electron: $E_T > 30$ GeV, $|\eta| < 2.5$ excluding $1.44 < |\eta| < 1.57$, ECAL seeded gsf electrons

✦Require exactly two PF jets in the event
- corrected $p_T > 30$ GeV and $|\eta| < 2.4$
- $|\Delta R(\text{jet, lepton})| > 0.3$

Triggers

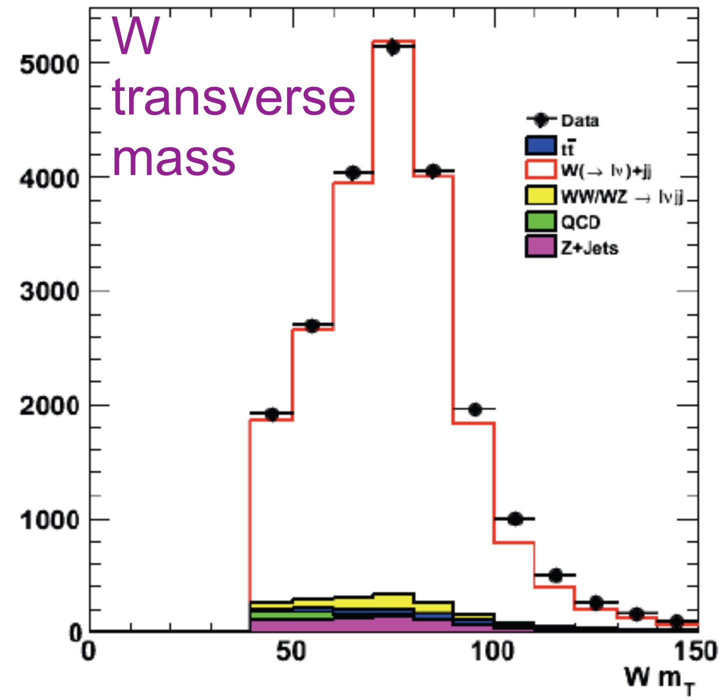
For electron use:	SingleEle || W_inclusive.
For muons:	IsoMu_24 || Mu17

✦For 2010 data (36 pb^{-1}) use single lepton triggers with $p_T > 17$ GeV (or lower)
✦For 2011 data before June TS (~200 pb^{-1}) use single Mu_24 and Ele_27
✦For 2011 data after June TS also use "inclusive" W trigger for electron: keeps electron $E_T > 25$ GeV, pf MHT > 25 GeV, W $m_T > 40$ GeV. Iso_mu_24 for muons.

Lepton selection – muon

✦Quality Requirements

- ≥10 tracker hits, ≥1 pixel hits
- ≥1 good muon chamber hit
- Both inside-out & outside-in reconstruction
- Track matching with ≥2 segments in the muon stations
- $\chi 2/\mathrm{ndf} < 10$ global fit
- Cosmic veto: impact parameter $|d_{xy}| <$ 0.02 cm (w.r.t. the beam spot)



✦Isolation

- Combined relative isolation (R=0.3)

$$I^{\mathrm{rel}}_{\mathrm{comb}} = \left\{ \sum (p_T(tracks) + E_T(em) + E_T(had)) - \pi r^2 . \rho_{\mathrm{Iso}} \right\} / p_T(\mu) \quad < 0.1$$

where ρ_{Iso} = PU density in $|\eta| < 2.5$, and r = radius of isolation cone = 0.3

Selection Efficiency: Efficiency for the above muon selection is 70.1%. A detailed study using Tag&Probe is underway to compute detailed data/MC scale factors.

Lepton selection – electron

Use VBTF working point 70 with a minor change in isolation cut

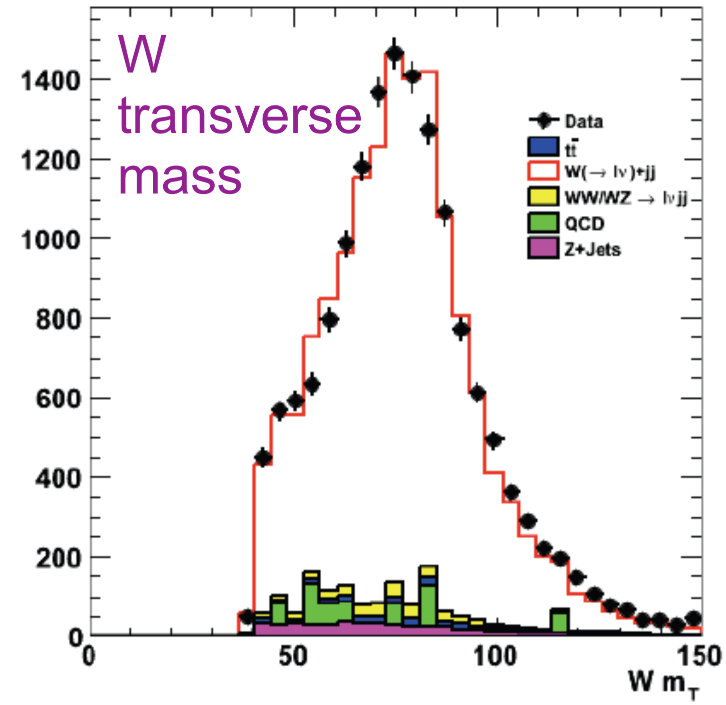
http://twiki.cern.ch/twiki/bin/view/CMS/SimpleCutBasedEleID

Conversion Rejection	
missing hits $\leq$	0
Dist	0.02
$\Delta cot\theta$	0.02

Combined Isolation	0.05

after PU subtraction

Electron ID	EB	EE
$\sigma_{i\eta i\eta}$	0.01	0.03
$\Delta\phi$	0.03	0.02
$\Delta\eta$	0.004	0.005
HoE	0.025	0.025



Reconstruction Efficiency: Electron reco efficiency is consistent with 100%

http://twiki.cern.ch/twiki/bin/viewauth/CMS/EgCommissioningAndPhysicsDeliverables

Selection Efficiency: Efficiency for the above electron selection is 64.3% in MC. A detailed study using Tag&Probe is underway to compute data/MC scale factors.

Jet and MET selection

Jet

Apply JetMET POG recommended default charge hadron subtraction (PF2PAT/PfNoPU) and FastJet PU subtraction. Apply default L2L3 correction and jet Id.

Loose jet Id criteria (recommended by JetMET POG)

- fraction of energy due to neutral hadrons < 0.99;
- fraction of energy due to neutral EM deposits < 0.99;
- number of constituents > 1;
- number of charged hadrons candidates > 0;
- fraction of energy due to charged hadrons candidates > 0;
- fraction of energy due to charged EM deposits < 0.99.

These identification criteria are applied to remove fakes due to calorimeter noise etc. The efficiency of passing these criteria for real jet is ~99.95%.

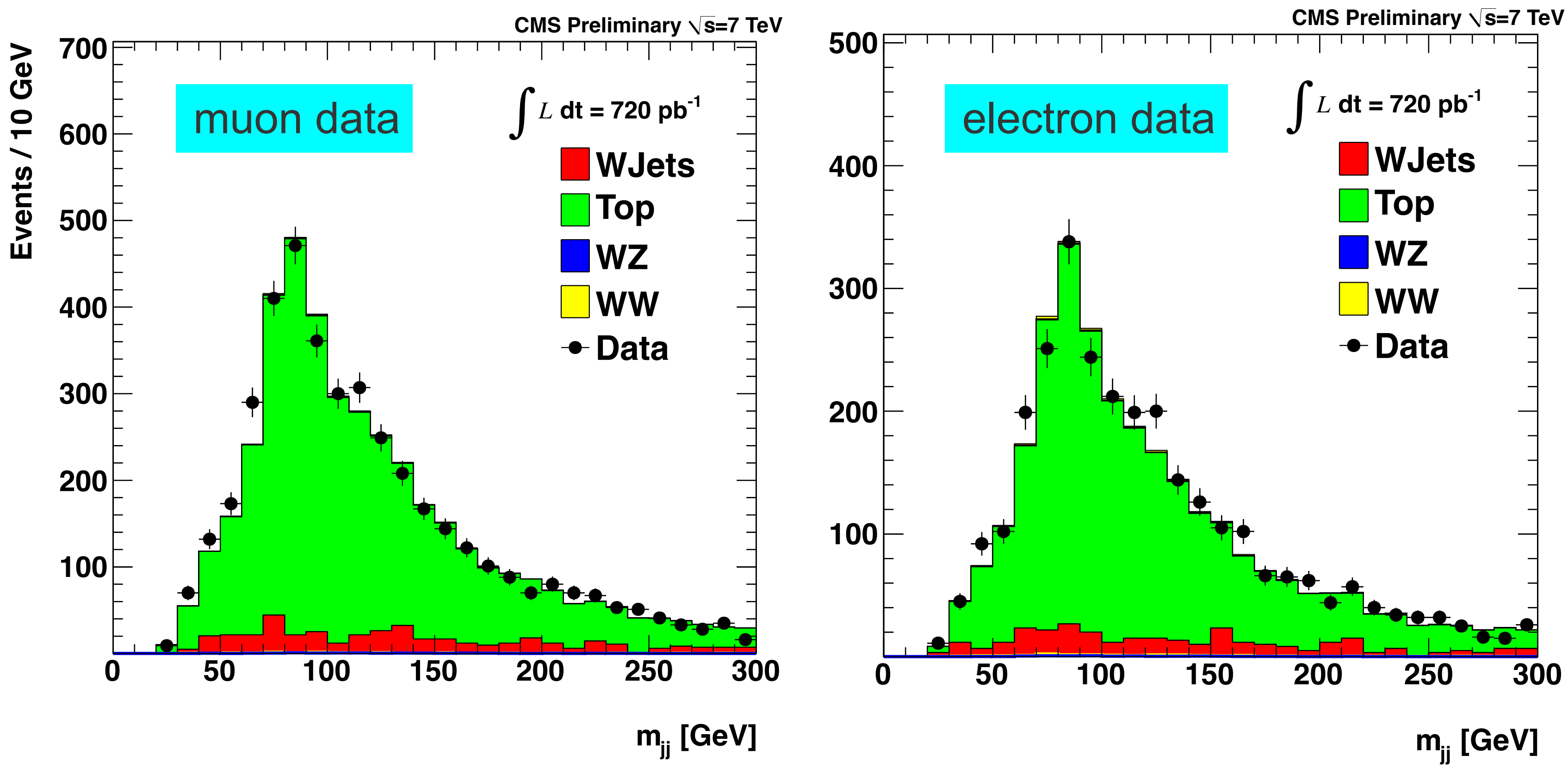
Remove identified leptons reconstructed as jet from the jet collection by requiring ΔR (lepton, jet) > 0.3. In addition, require the jets to be not b-tagged. For this, use "simple secondary vertex high efficiency" loose operating point (tagging efficiency ~ 70%, mistag rate ~ 1%). This greatly reduces top bkg.

MET

Use default particle flow MET. Require MET > 30 GeV.

Can we reconstruct hadronic W in CMS ? Yes

In top events reconstruct clear W peak almost “out-of-box” with good resolution



Just require 4 jets above p_T 30 GeV, 2 loose b-tags, and a leptonic W (muon: p_T>25 GeV or electron: E_T>30 GeV, MET>25 GeV). Then plot m_{jj} of the two jets which are not b-tagged.

Topological cuts

Start with simple topological cuts to suppress W+jets background

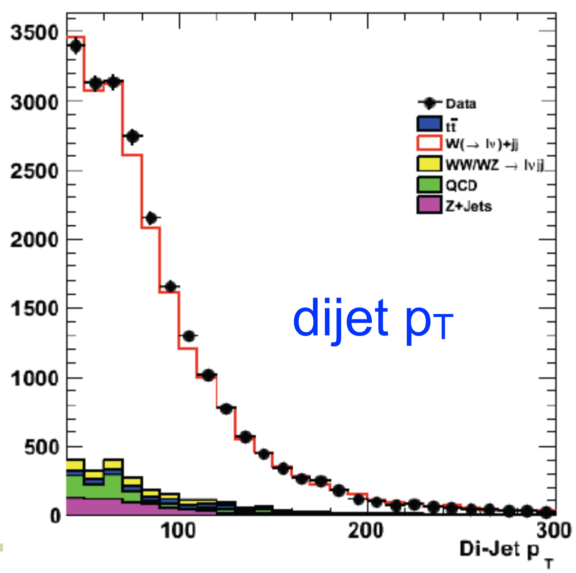
Leptonic W

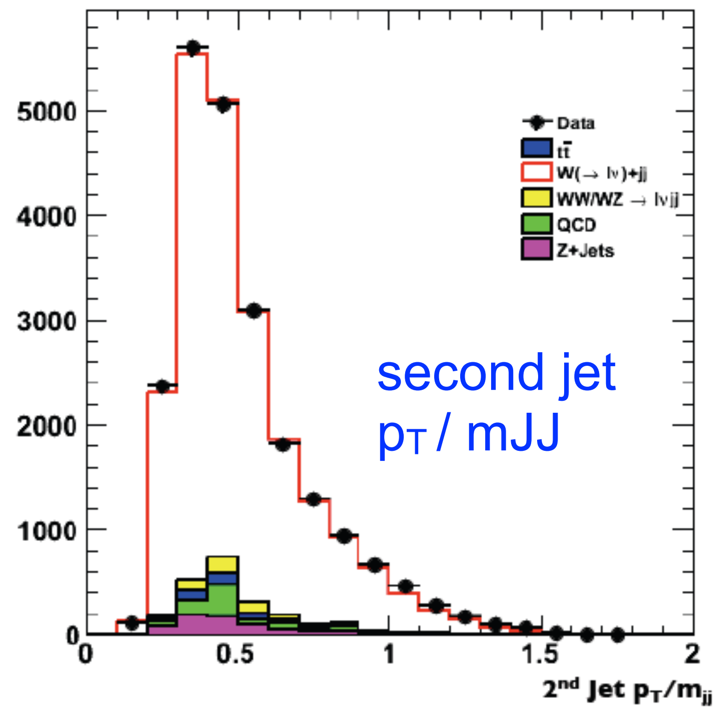
- pf MET > 30 GeV
- W transverse mass > 40 GeV

After this cut essentially pure W events are left : diboson, W+jets, and top pair + single top.

Dijet system

- Dijet p_T > 40 GeV
- $\Delta\eta$ (j1, j2) < 1.5





ELM recommendation to improve S/B @LHC

Testing CDF's Dijet Excess and Technicolor at the LHC

Estia Eichten[1*], Kenneth Lane[2†] and Adam Martin[1‡]
[1]Theoretical Physics Group, Fermi National Accelerator Laboratory
P.O. Box 500, Batavia, Illinois 60510
[2]Department of Physics, Boston University
590 Commonwealth Avenue, Boston, Massachusetts 02215

July 20, 2011

We are in touch with ELM authors. Ken Lane is at CERN these days. He had a private meeting with ATLAS this week. He will talk to our team + Phys Management next week.

Main recommendations

- Lead jet $p_T > 40$ GeV, second jet $p_T > 30$ GeV
- Dijet $p_T > 45$ GeV
- $\Delta\eta$ (j1, j2) < 1.2
- W $p_T > 60$ GeV
- Plus, some model-dependent cuts for TC which we can ignore

All cuts except W p_T are highly effective in reducing W+jets, W p_T can help somewhat but at the cost of large drop in signal efficiency.

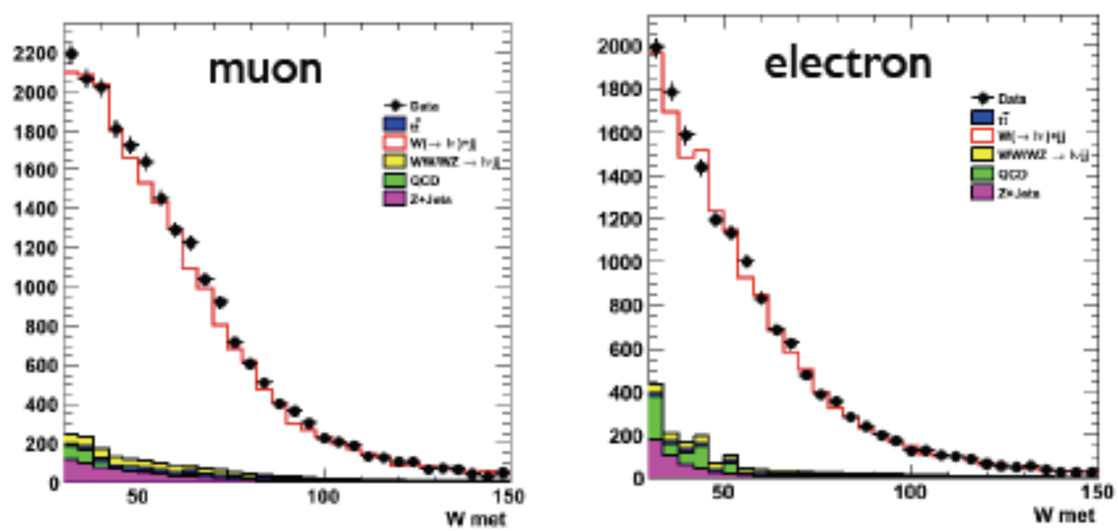
Template fit to extract various contributions

Perform a likelihood fit of the m_{jj} distribution

- in the range 40−300 GeV
- take all shapes except for QCD from MC; for QCD invert iso/id cuts in data
- fit for the absolute number of diboson and W+jets events
- fix top pair, single top, Z+jets, Z→ττ contributions to the NLO cross section
- float the jet energy scale within uncertainty

Then plot background-subtracted (i.e., data − bkg from fit) distribution to visually inspect the quality of the fit.

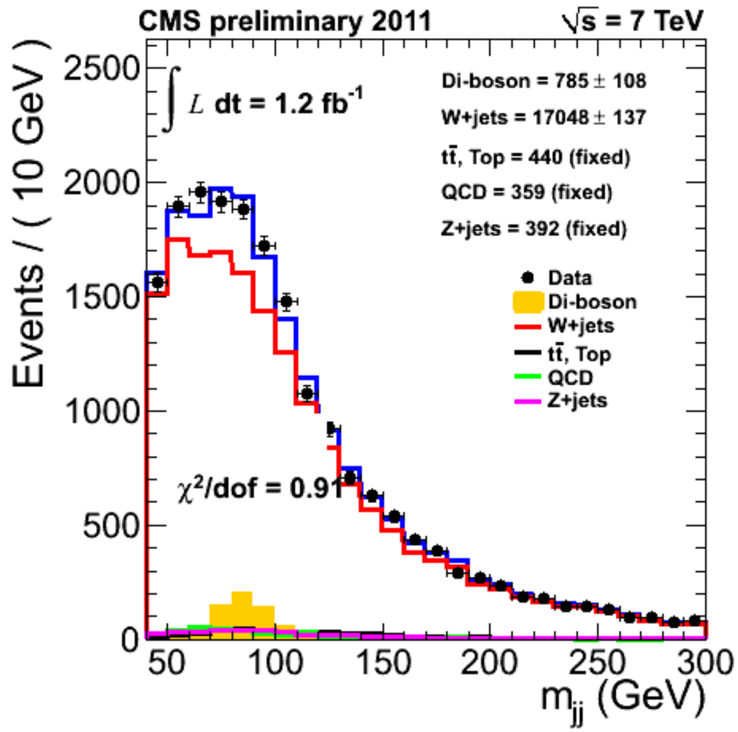
MET distributions in data

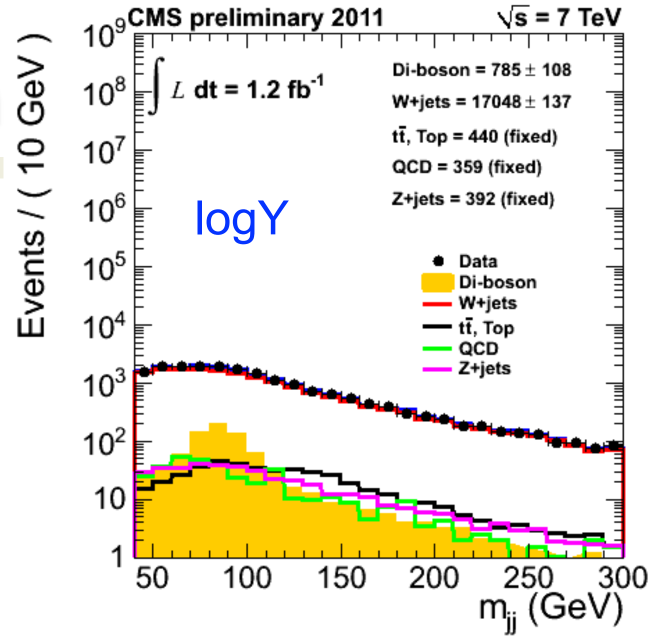


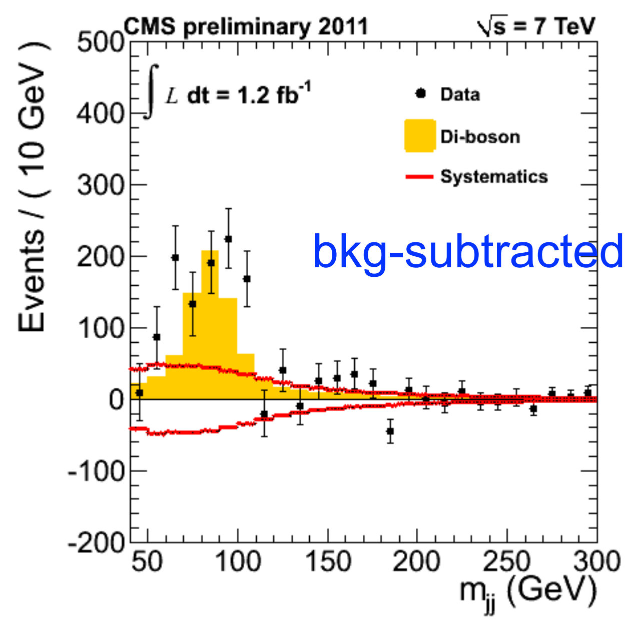
Fit results with default selection

Baseline CDF-like selection and :

- Lead jet p_T>40 GeV
- Dijet p_T > 45 GeV
- $\Delta\eta$ (j1, j2) < 1.2
- p_T^{2ndJet} / m_{JJ} > 0.3
- anti b-tag





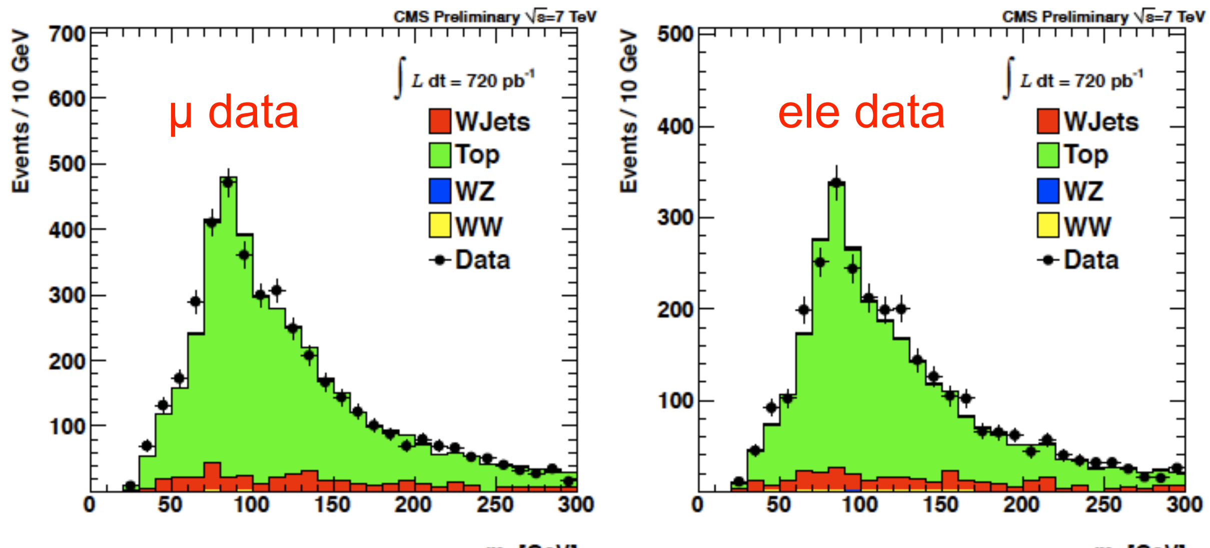


✦Systematics are still being finalized.
✦Working on implementing ~15% broadening of JER observed in jet data (from JetMET POG)
✦Diboson peak established, no anomalous excess

How confident we are about JES ?

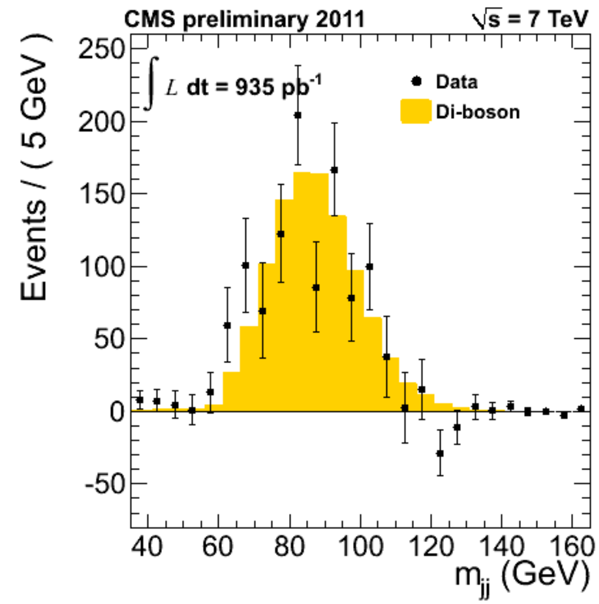
Reasonably confident:

m_{JJ} in top data



(a) (b)

Figure 14: The invariant mass distribution of the two non-b-tagged jets in events containing a high energy lepton, large $\not{E}_T$, and four jets of which two are b-tagged. This sample is dominated by $t\bar{t}$ events. The distribution shown here for both lepton channels (a: muon, b: electron) essentially establishes how well we can reconstruct a hadronically decaying vector boson in CMS detector. The resolution of the reconstructed vector boson is dominated by the jet energy resolution.



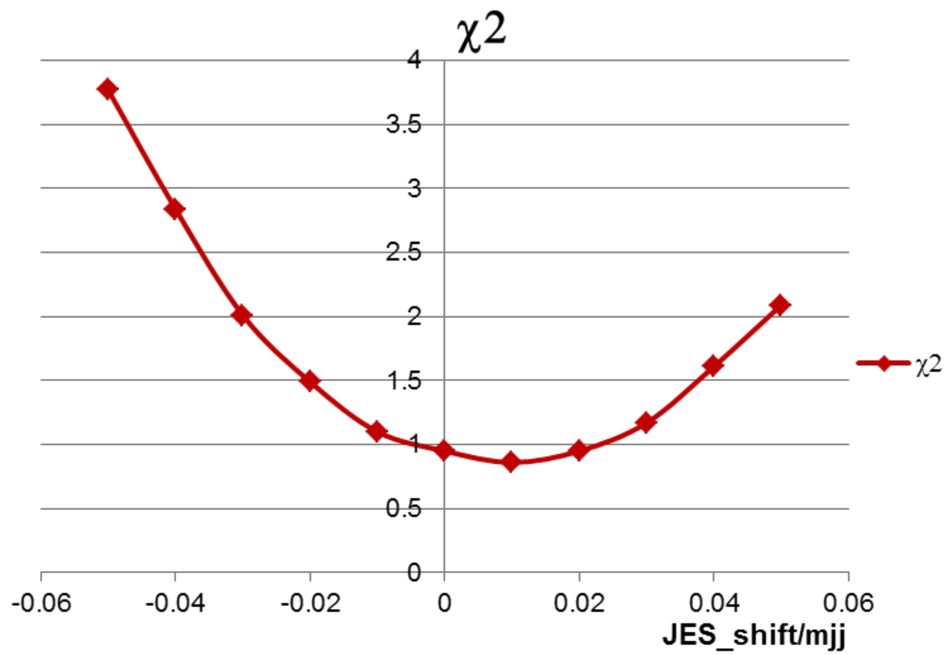
m_{JJ} is diboson data after tight selection

All indications suggest that JES in data and MC agree very well.
Jet resolution is not much worse in data.

Scan of jet energy scale

- The default fit allows the JES to float and returns a value of 0.003 m_{JJ}
- We perform a manual scan by fixing JES and repeating the fit

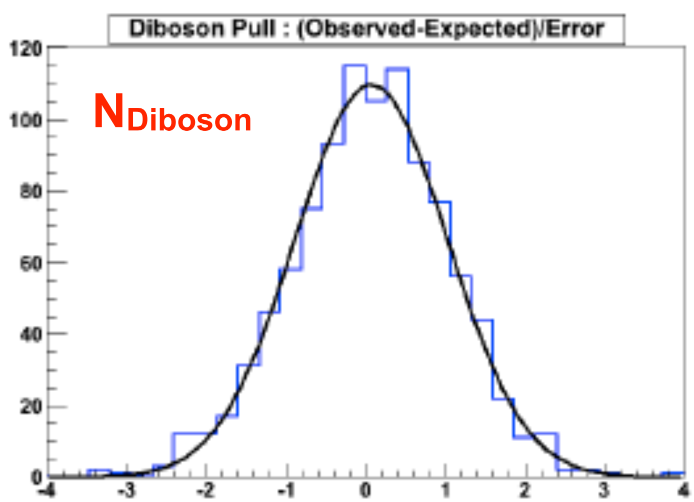
JES_shift/mjj	WW	χ^2
-0.05	1012	3.77
-0.04	960	2.84
-0.03	901	2.01
-0.02	874	1.49
-0.01	877	1.1
0	793	0.95
0.01	676	0.86
0.02	594	0.95
0.03	496	1.17
0.04	354	1.61
0.05	261	2.09

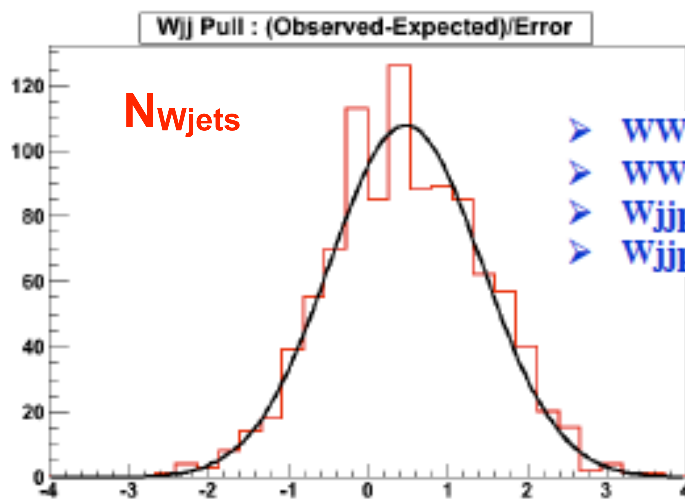


The fit is stable and has a χ^2 minimum near 0.

Validation of fit results

To make sure that the template fit is unbiased and to check the coverage of statistical uncertainty reported by the fit, we need to generate 1000 pseudo experiments (PE) using the shape that best describes the data. Then we fit each of these PE samples using our nominal shape and plot pull distribution for each parameter.





- WWpull_mean=0.07+/-0.03
- WWpull_sigma=0.96+/-0.02
- Wjjpull_mean=0.48+/-0.03
- Wjjpull_sigma=0.96+/-0.02

•Fitter returns consistent results: negligible bias.
•Should correct the final Yields of diboson and Wjj by 13 and 105 events respectively (we have 785 diboson and 17048 Wjj events)

Systematic uncertainties

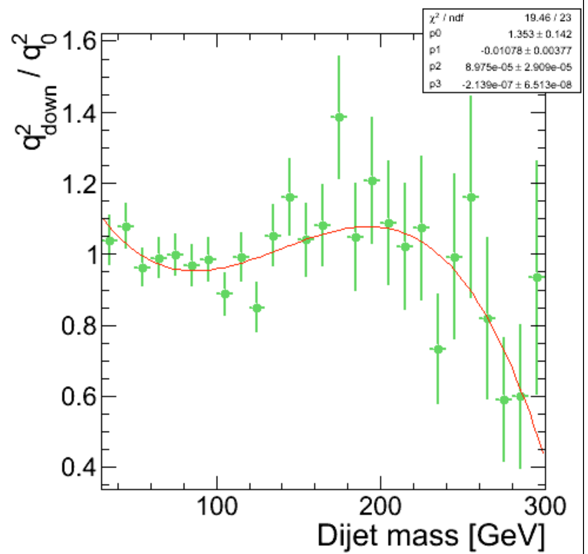
There are three main sources of systematics:

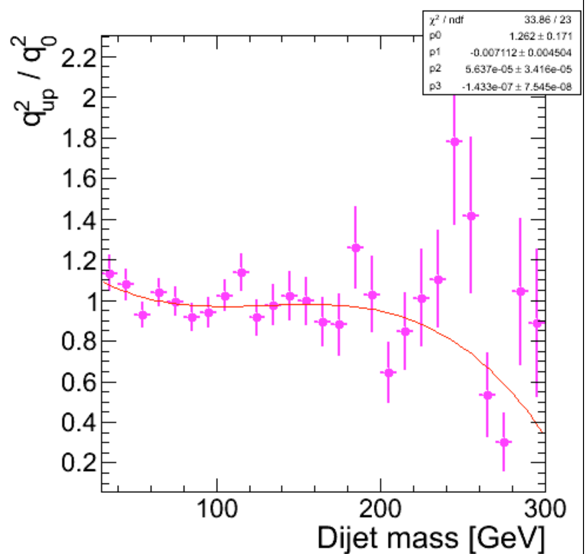
1. W+jets shape: depends on the choice of factorization/ renormalization scheme. Try q^2 up/down variation templates

2. Statistical fluctuation in W+jets shape from MC due to finite (2x data) size of the MC sample: generate 1000 toy MC by fluctuating the bin contents by Poisson statistics and use these as template

3. JES variation

Working to include other smaller systematics: luminosity, uncertainties in top/Zjets/single top/ Ztatau cross section, jet resolution, fit bias etc.

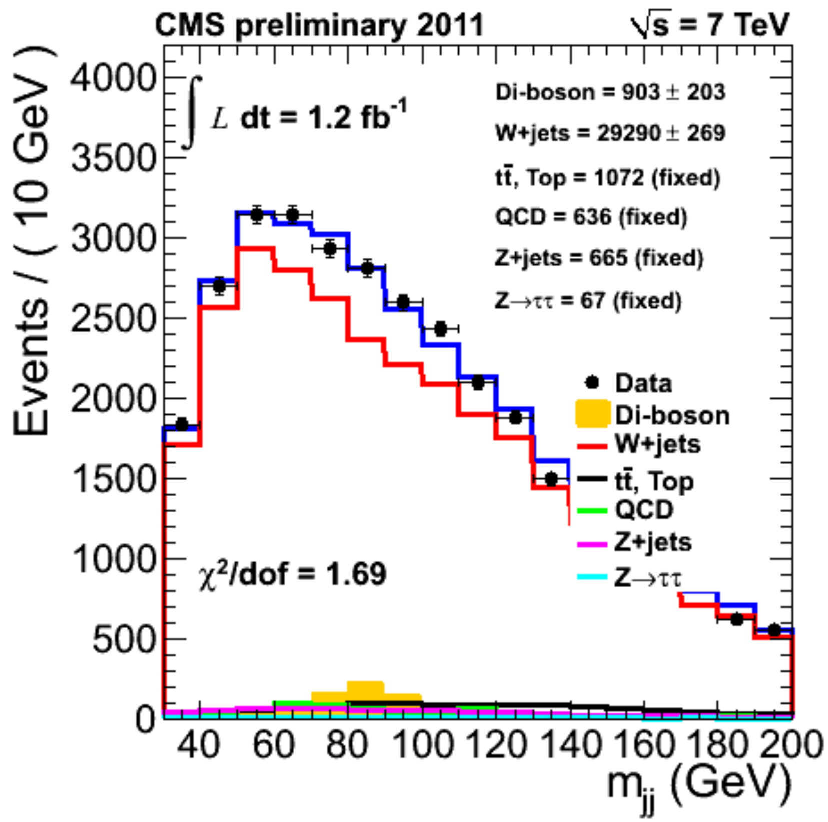


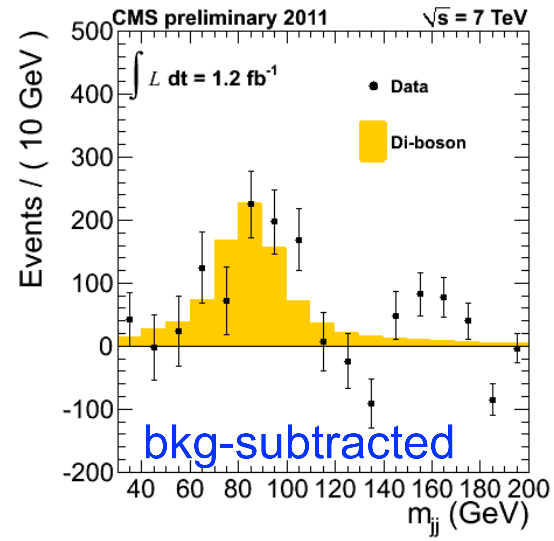


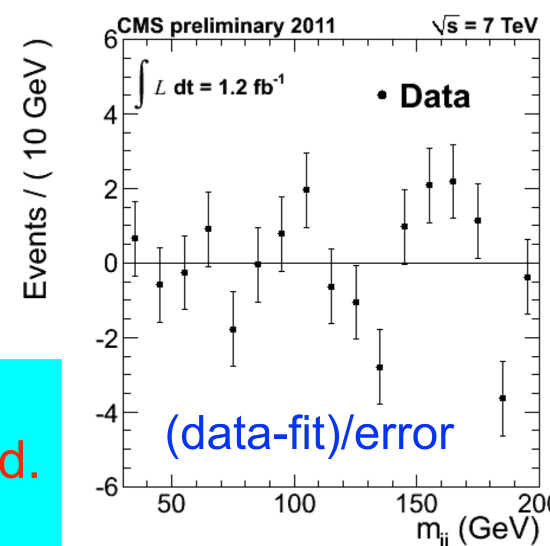
What happens if we go to “CDF-like” selection

Similar cuts as CDF except:

- •Δη(j,j) < 1.5(CDF had 2.5)
- •W m_T > 40 GeV (CDF had 30)
- •MET>30 GeV (CDF had 25)
- •Mu p_T>25 GeV, Ele p_T>30 GeV (CDF had 20, 25)







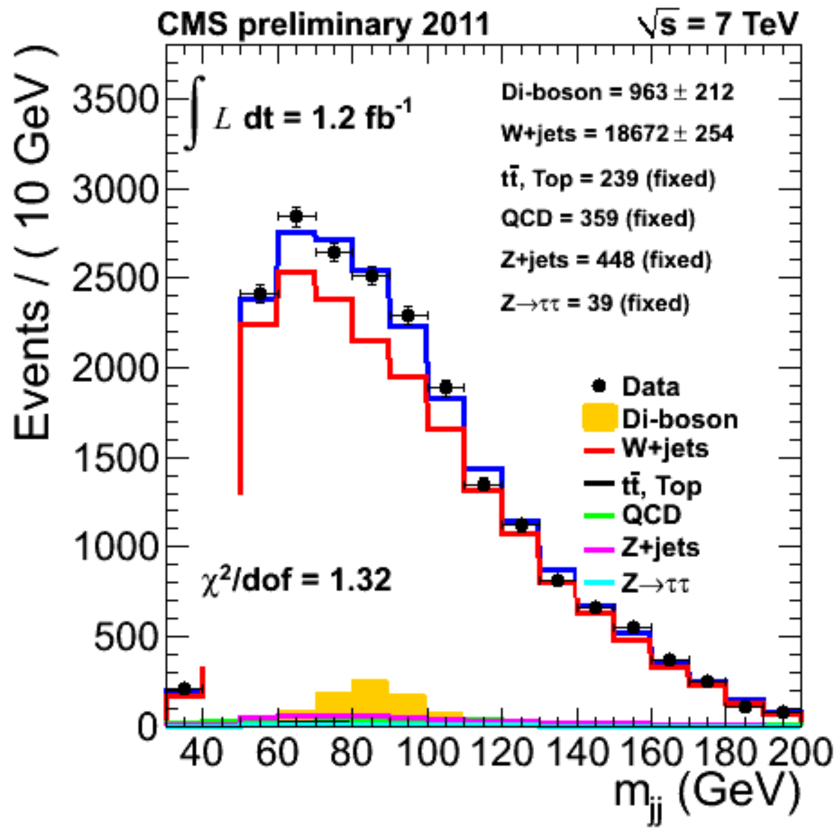
1[st] variation to suppress large Wjj, others due to trigger constraints

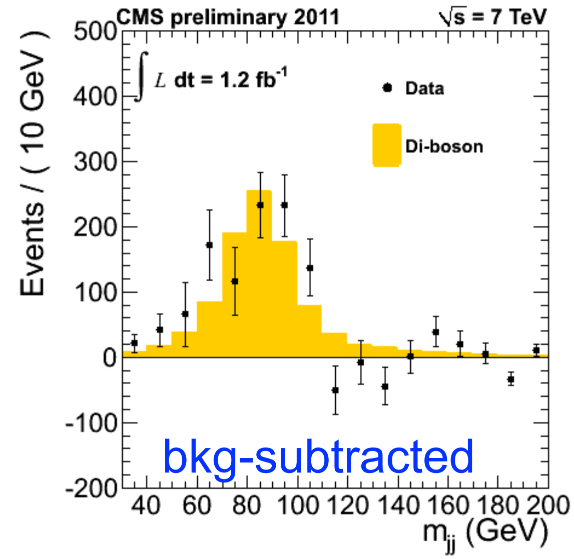
With “CDF-like” loose cut it becomes hard to model the data well and to have all systematics understood. Although, essential features remain the same.

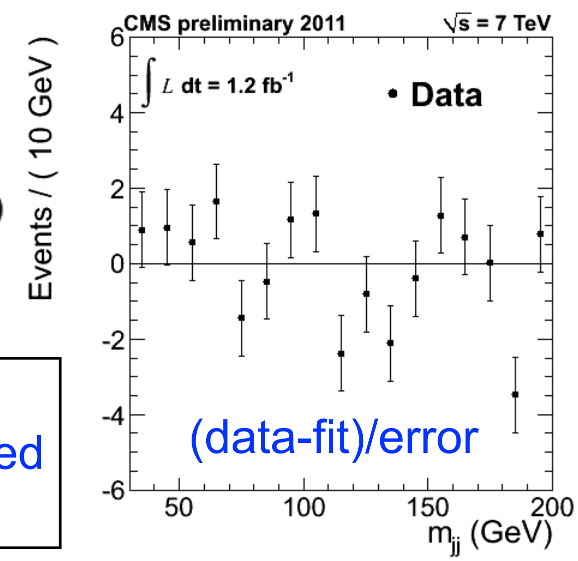
What if we go to "diboson optimized" selection

Additional cuts w.r.t. CDF cuts:

- •Kinematic Fit χ2/NDF<10.0 (not used in generic Mjj analysis)
- •-0.6<cosJacksonAngle<0.8 (not used in Mjj analysis)
- •Jet2pT/mjj>0.3
- •anti b-tag (SSVHE -medium) both jets







- •The fit quality improves with additional cuts
- •Diboson peak becomes more well-defined and pronounced
- •Residual fluctuations are statistically consistent with 0

Does the diboson yield make sense ? Yes

Here is a simple calculation of what we expect:

WW + WZ cross section at NLO = 61 pb
BR(W->lnu) = 2 * 1/9
BR(W/Z -> jj) = 2/3
Luminosity = 1.2 fb^-1
Efficiency x Acceptance = 0.1

Therefore, expected yield ~ 1000
There is 10% uncertainty in cross section and 6% in luminosity.

Our fit gives very close value with an uncertainty of about 20%.

What if we vary one cut at a time

Gradually change the cuts from CDF-like to the ones optimized for extracting the diboson yield

- **Stage 0: CDF-like cuts**
- **Stage 1: Kinematic Fit χ^2/NDF<10.0 (not used in generic Mjj analysis)**
- **Stage 2: -0.6<cosJacksonAngle<0.8 (not used in generic Mjj analysis)**
- **Stage 3: Jet2p_T/m_{jj}>0.3**
- **Stage 4: Anti-btag (SSV-HE-M) both jets [aka diboson selection]**

Since in step 0 we are completely swamped by background, it is hard to say if we do not model the W+jets right or it is statistical fluctuation in the number of W+jets. As the S/B improves, so does our ability to observe qq processes (diboson or otherwise).

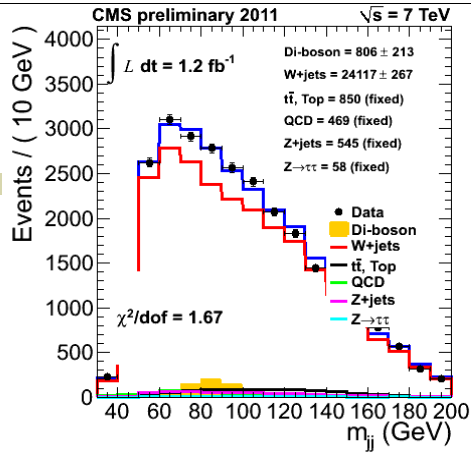
Comparison on the next slide →

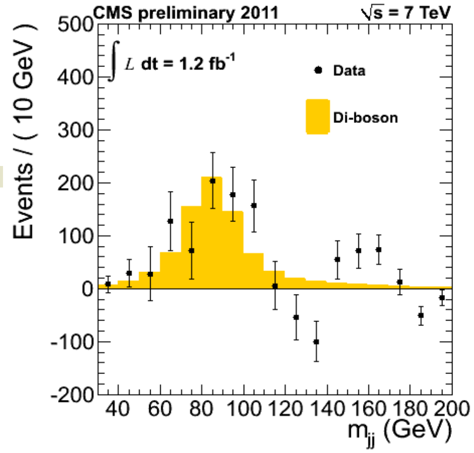
Stage 1:

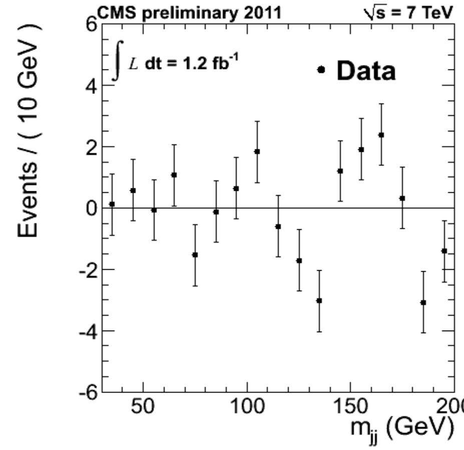
N_{WW}=806±213

N_{Wjj}=24117±267

S/B=0.033





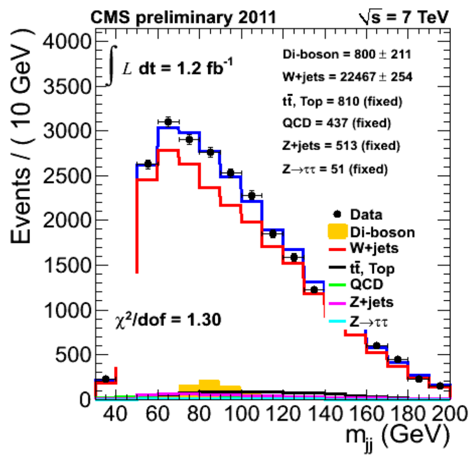


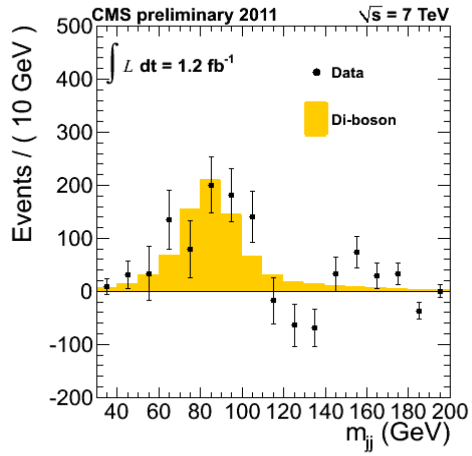
Stage 2:

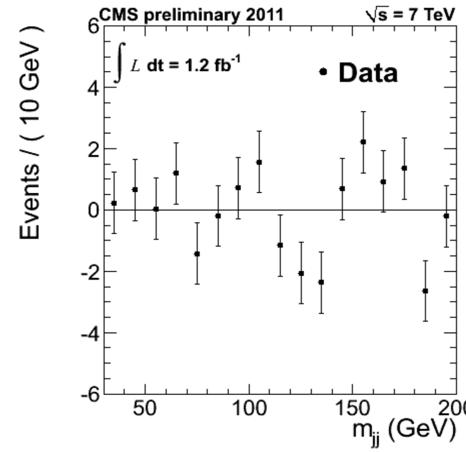
N_{WW}=800±211

N_{Wjj}=22468±255

S/B=0.036





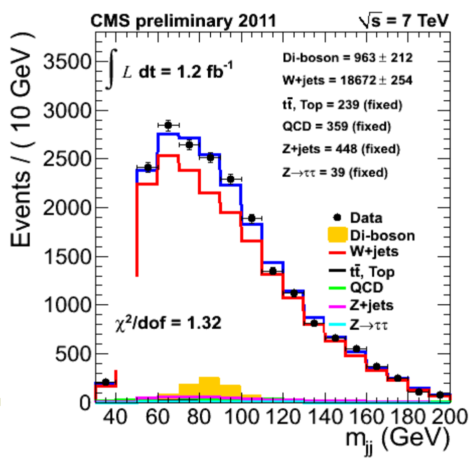


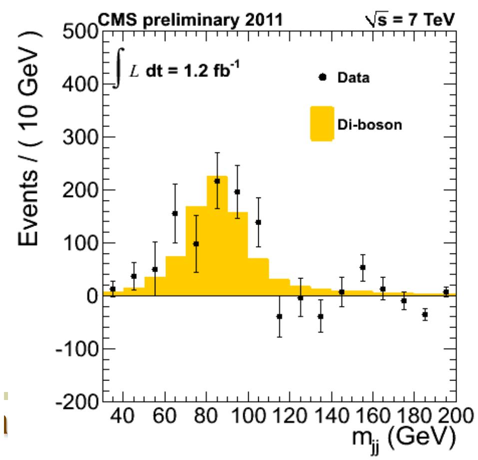
Stage 3:

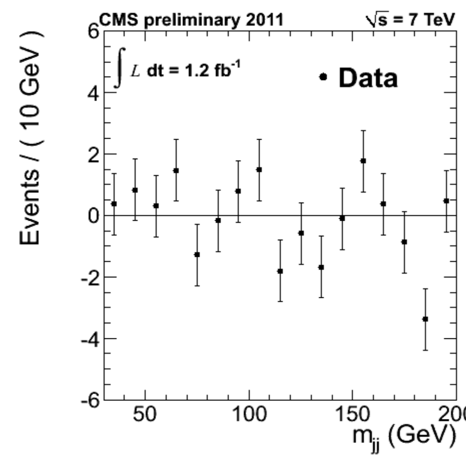
N_{WW}=847±228

N_{Wjj}=20557±272

S/B=0.041

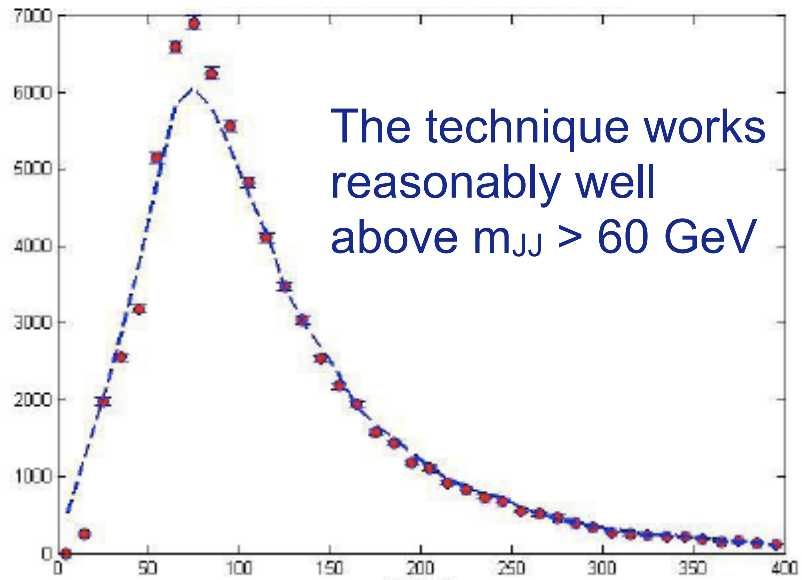


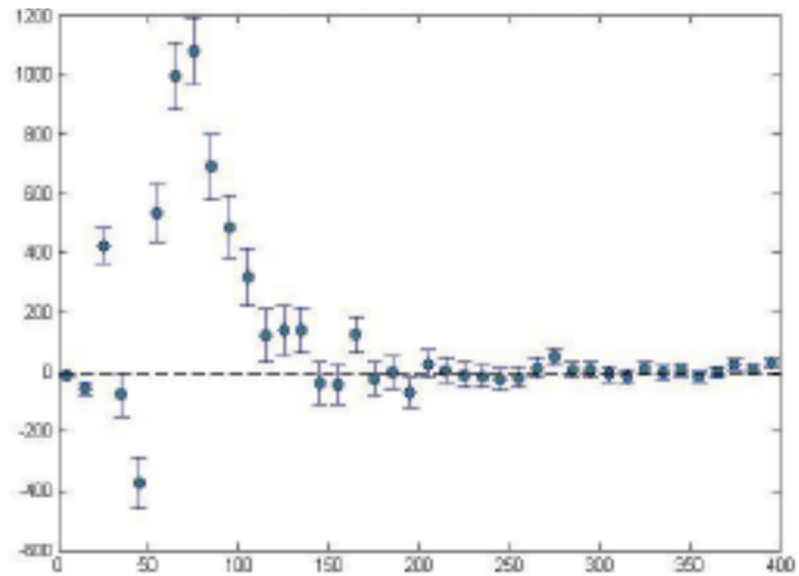




Have we tried data-driven shape for W+jets

Yes: For cross check we take shape using “mixed events”. We make random combination of jets by taking one jet from some other event. This gives large ststistics: N*(N-1). This cannot produce a bump. The only challenge is to manage kinematic correlations.





https://indico.cern.ch/materialDisplay.py?contribId=33&materialId=slides&confId=147015

Work ongoing to improve the technique

Summary

✦Performed study of the dijet mass spectrum in lvjj events using ~1 fb^{-1} data

- start with basic CDF-like loose selection
- try various improvements − including those suggested by ELM − to suppress W+jets and to make qq processes stick out
- settle on selection criteria more appropriate for LHC conditions

✦Di-boson peak in W(lv)+jj channel is clearly established

- This gives us confidence to be able to observe qq processes
- Use as many data-driven inputs as we can, less reliance on MC

✦No clear anomalous peak observed so far

- finalizing the main systematics and residual corrections
- aiming for a preliminary public result soon
- AN-2011-266, PAS: EWK-11-017

BACKUP SLIDES

Data and MC samples

Analyzed 935 pb^{-1} data

- Entire data sample is processed (prompt-/ re-Reco) with CMSSW 4_2_X.
- Few of 4.2.X Summer11 MC samples needed for this analysis are produced. Use 4.1.X Spring11 MC for the rest.

/EG/Run2010A-Apr21ReReco-v1/AOD
/Electron/Run2010B-Apr21ReReco-v1/AOD
/SingleElectron/Run2011A-May10ReReco-v1/AOD
/SingleElectron/Run2011A-PromptReco-v4/AOD
/ElectronHad/Run2011A-PromptReco-v4/AOD
/Mu/Run2010A-Apr21ReReco-v1/AOD
/Mu/Run2010B-Apr21ReReco-v1/AOD
/SingleMu/Run2011A-PromptReco-v4/AOD

MC ID	name	σ LO(NLO) [pb]	lumi LO(NLO) [fb^{-1}]
1000030	WWtoAnything_TuneZ2_7TeV-pythia6-tauola	27.8 (42.9)	73.7(47.8)
1000031	WZtoAnything_TuneZ2_7TeV-pythia6-tauola	10.4 (18.3)	211.0(119.9)
4000041	WJetsToLNu_TuneZ2_7TeV-madgraph-tauola	24640 (31539)	~~0.616 (0.481)~~ ~2 fb^{-1}
9000219	TTToLNu2Q2B_7TeV-powheg-pythia6	65.83 (?)	73.2 (?)
4000040	TTJets_TuneZ2_7TeV-madgraph-tauola	121 (?)	9.6 (?)
4000016	TToBLNu_TuneZ2_s-channel_7TeV-madgraph	0.99 (?)	500 (?)
4000017	TToBLNu_TuneZ2_t-channel_7TeV-madgraph	21.0 (?)	23.05 (?)
4000018	TToBLNu_TuneZ2_tW-channel_7TeV-madgraph	10.56 (?)	46.87 (?)
1000041	QCD_Pt-30to80_EMEnriched_TuneZ2_7TeV-pythia6	3866200	1.858e-02
1000043	QCD_Pt-80to170_EMEnriched_TuneZ2_7TeV-pythia6	139500	5.787e-02
1000040	QCD_Pt-30to80_BCtoE_TuneZ2_7TeV-pythia6	136804	1.459e-02
1000042	QCD_Pt-80to170_BCtoE_TuneZ2_7TeV-pythia6	9360	0.1115
1000039	QCD_Pt-20_MuEnrichedPt-15_TuneZ2_7TeV-pythia6	136804	0.2157