Summary of CMS Checks of CDF Anomaly in W(→lν) +jj Events

Kalanand Mishra

Fermilab

with inputs from EWK,EXO,HIG W(→lν)+jj analysis team members

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

CMS Joint EWK-EXO-HIG Meeting

(April 19, 2011)

Why look at W(→lν)+jj ?

✦W(→lν)+jj is important for variety of reasons

✓Establish electroweak WW diboson production at CMS using W→jj

✓BR is 6 times larger than WW→lνlν yielding more events

✓Main search mode for SM Higgs with mass above 200 GeV

-Can still contribute significantly if m_H > 160 GeV

✓Search for resonances/bumps from new physics in dijet mass in the ~few hundred GeV range produced in association with W

✦WW (and WZ)→lνjj is fully reconstructible: can look for WW "bumps"

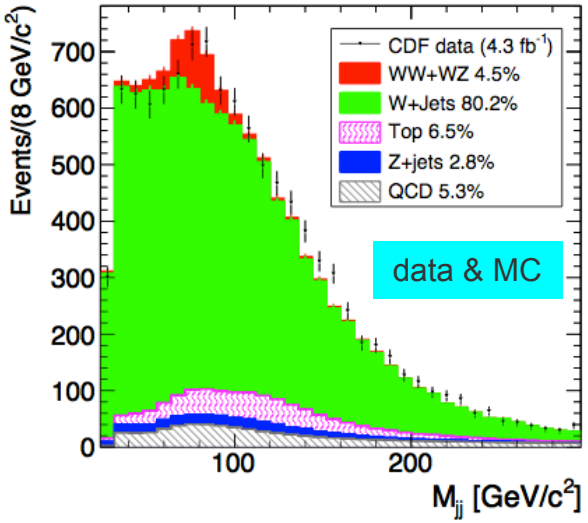
•W pair with muon/electron + MET + exactly 2 jets. Solve for kinematics. Most events have solutions with 2 real roots for p_z of ν.

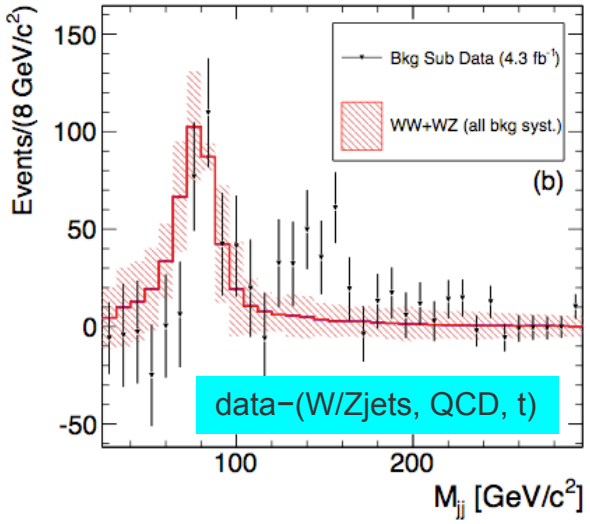
•Main backgrounds are W+jj and top.

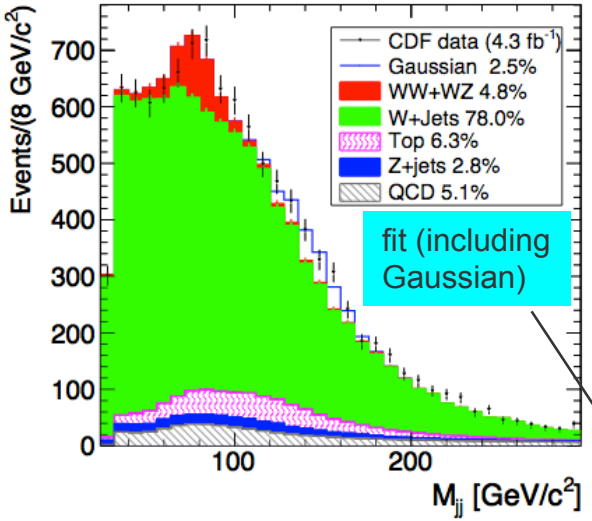
-Top background can be reduced by asking for no b tags and for exactly 2 jets ("top veto").

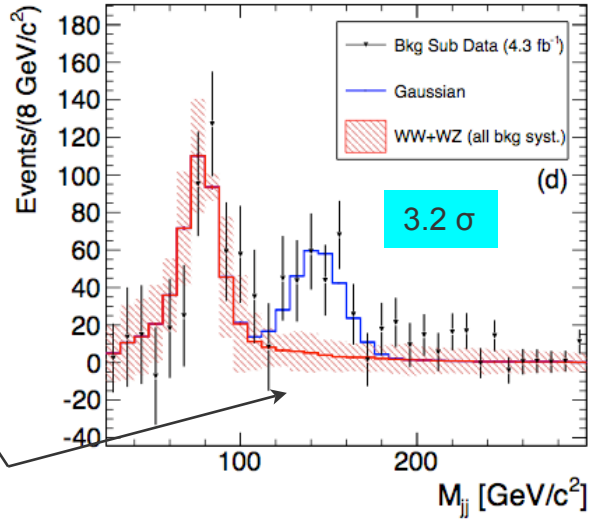
-W+jj background can be reduced by requiring a W mass for the jet pair and other kinematic cuts

CDF finds anomaly in W(→lν)+jj events









arXiv: 1101.6079 (Apr 4)

✦CDF has a dijet W signal on a large Wjj bkg.

✦However, the Wjj MC doesn't reproduce the featureless falloff of dijet mass spectrum.

✦CDF finds an excess of 253 events in the dijet mass window 120−160 GeV. Significance is 3.2σ.

✦The bump is compatible with a particle of mass 150 GeV and width 14 GeV. Not compatible with SM Higgs boson.

CDF W+jj analysis: some details

Event selection

- W→lν reconstruction
  - E_T^{lepton} > 20 GeV, MET > 25 GeV
  - m_T^W > 30 GeV, Z veto
- 2 jets with p_T >30 GeV and |η| <2.4
  - |Δφ(jet1, MET)| > 0.4
  - |Δη(jet1, jet2)| < 2.5
  - p_T^{dijet} >40 GeV

Template fit: shapes and normalizations

Process	Model	σ (pb)
WW/WZ inclusive	PYTHIA	15.9 ± 0.9
$Z \to e, \mu, \tau$+jets	ALPGEN+ PYTHIA	787 ± 85
$t\bar{t}$	PYTHIA	7.5 ± 0.83
single top	MADEVENT + PYTHIA	2.86 ± 0.36
W+jets	ALPGEN+ PYTHIA	from data
QCD multijet	from data	from data

QCD template derived from inverting the muon isolation or inverting two Id cuts for electrons

Signal extraction

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)

"bump" significance

Compute p-value of the fit with and w/o Gaussian component. Largest p-value is significance. Include syst. and trial factor.

More details: Wine & Cheese talk on April 6 http://theory.fnal.gov/jetp/talks/Viviana.pdf

CDF web site: http://www-cdf.fnal.gov/physics/ewk/2010/WW_WZ/

Student thesis: http://lss.fnal.gov/archive/thesis/fermilab-thesis-2010-51.pdf

New York Times: http://www.nytimes.com/2011/04/06/science/06particle.html?_r=2&hp

Some characteristics of excess events

Production cross section **= 4 pb** i.e., very large, ≈ 50% of ttbar cross section

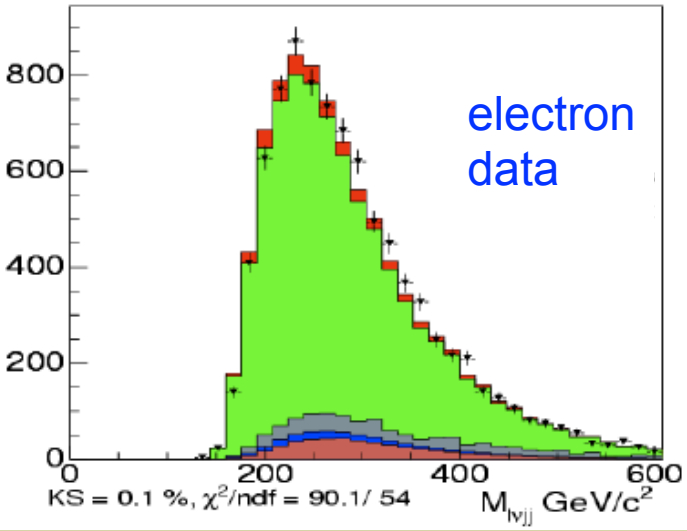
Excess is NOT from $b\bar{b}$

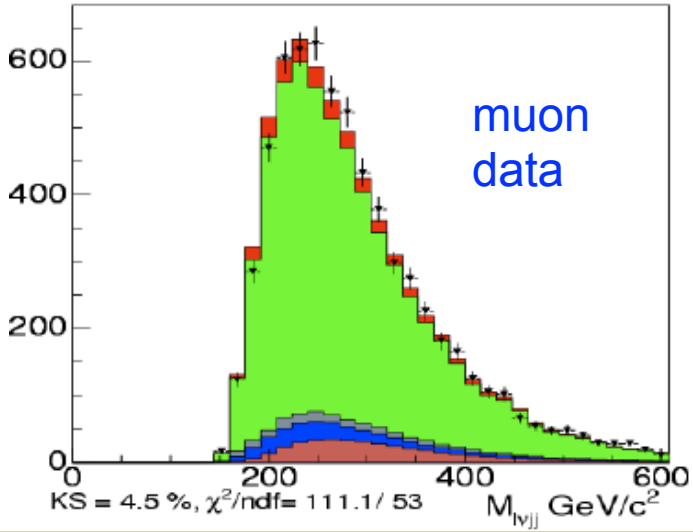
Fraction of events with b jets

Tag requirement	Excess region (120−160 GeV)	Sideband region
Muons		
1 tag	0.1027 ± 0.0112	0.0813 ± 0.0096
2 tag	0.0078 ± 0.0030	0.0084 ± 0.0030
Electrons		
1 tag	0.0897 ± 0.0088	0.0945 ± 0.0087
2 tag	0.0110 ± 0.0030	0.0095 ± 0.0026

No significant difference is observed in b jet fraction between excess region and sideband

No obvious structure in $M_{(\mathrm{lepton},\nu,jj)}$ and $M_{(\mathrm{lepton},\nu,jj)}$-$M_{jj}$ distributions





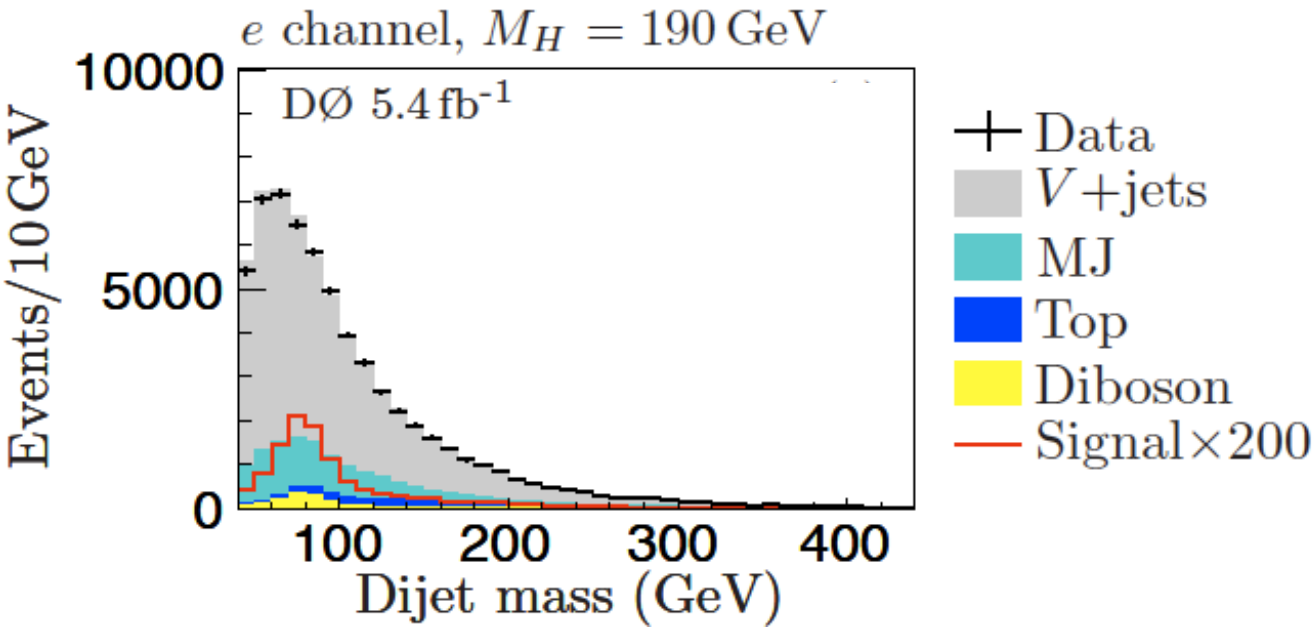
Distributions are compatible with "bkg. only" hypothesis

i.e, not from a quasi 2-body lνjj resonance.

DØ public result on the topic

Higgs search in lvjj

arXiv: 1101.6079 (Feb 2, '11)



This is a multi-variate analysis. The paper has plots of some variables for e- and other variables for μ-channel. I couldn't find the corresponding μ plot.

Event selection

- W→lν reconstruction
  - E_T^l > 15 GeV, MET > 15 GeV
  - m_T > (40 − 0.5 * MET) GeV
- ≥2 jets with p_T >20 GeV and |η| <2.5

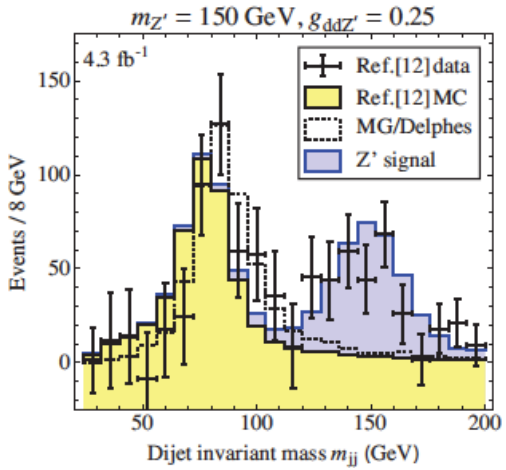
DØ promised during W&C seminar that they will quickly (~ 2 weeks) analyze data for any bump.

No obvious bump visible in DØ data in the same mass range as CDF.
But then it is hard to tell in this plot.

Theoretical interpretations of CDF bump

Light Z' boson

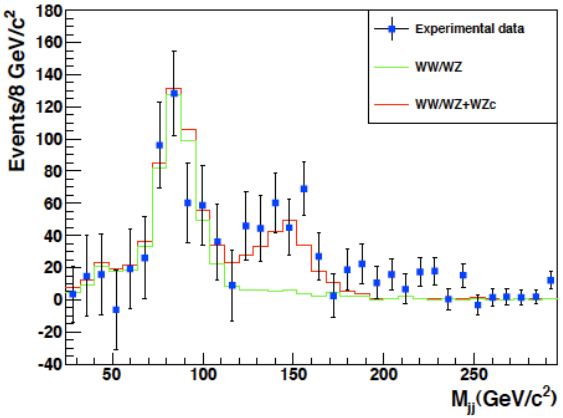
arXiv: 1103.6035, Buckley et al
arXiv: 1104.3139, Jung et al



This Z' has suppressed coupling to leptons, but substantial couplings to quark. Can also explain top F-B asymmetry.

Color octet axial-vector boson

arXiv: 1101.6079, Wang et al



Also explains top F-B asymmetry.

Intrinsic s in nucleon

arXiv: 1104.1894, He & Ma

Nucleon intrinsic s quarks contribute to W boson production in association with single top. This causes excess in dijet mass in W+jj but not in Z+jj.

SUSY particle

arXiv: 1104.2014, Sato et al

An MSSM particle with tanβ≈10

First observation of a SUSY particle !

Technicolor

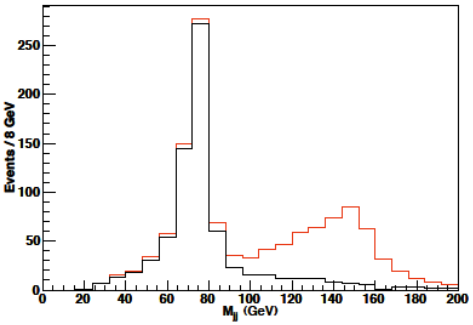
arXiv: 1104.1255, Andersen et al
arXiv: 1104.0976, Eichten et al

Techni-π/ρ→WZ can enhance Wjj with low m_{jj}

Unified flavor symmetry

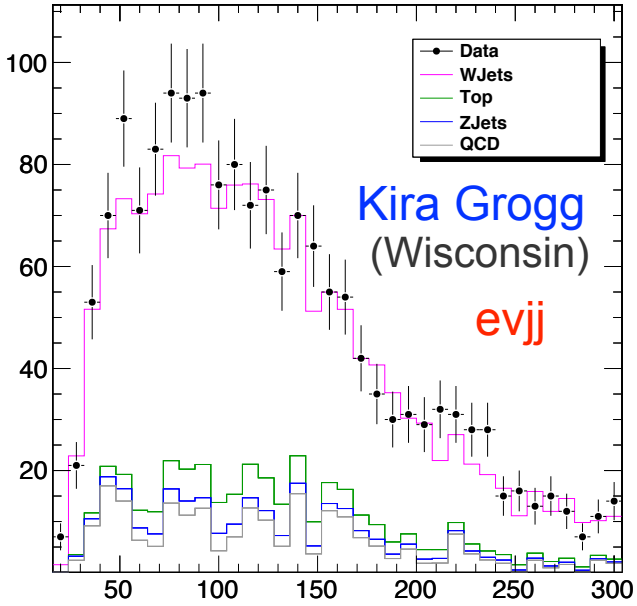
arXiv: 1104.2030, Nelson et al

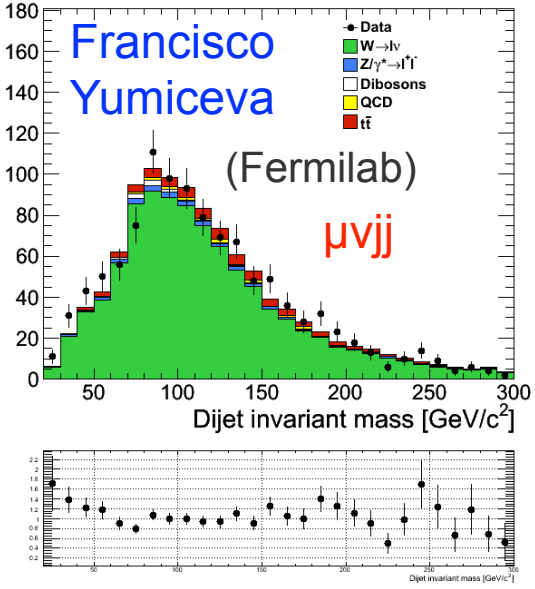
Propose a subgroup of SU(3). Predict comparable F-B asymmetry in tt and cc but suppressed in bb system. Excess in Wjj but suppressed in Zjj.

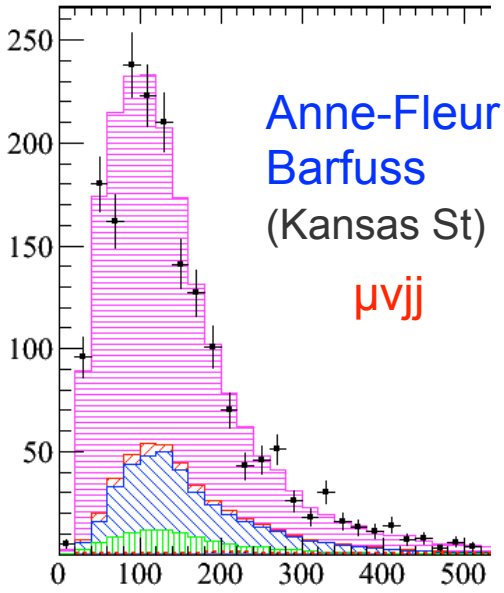


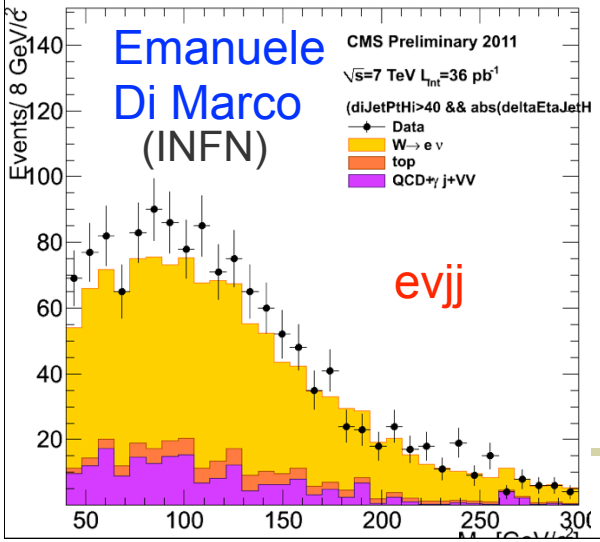
Huge excitement in CMS over CDF bump

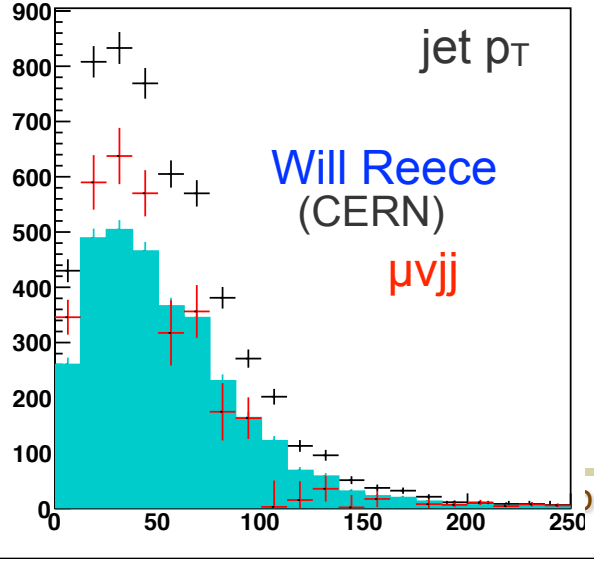
Lot of activity within CMS to reproduce CDF analysis within 24 hours !

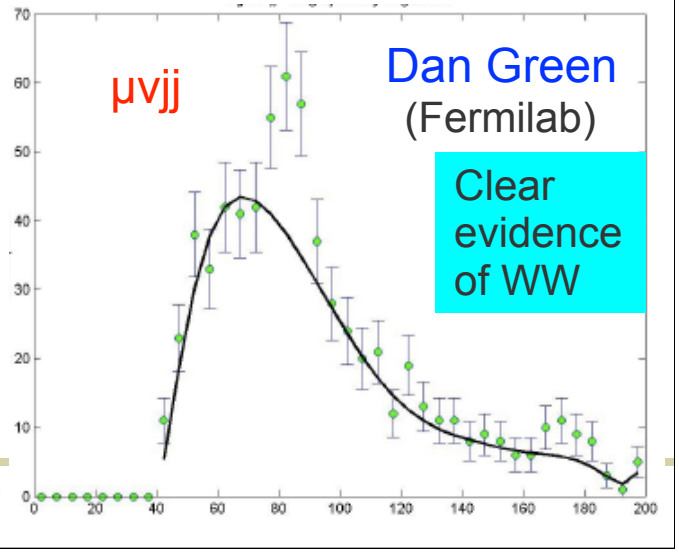




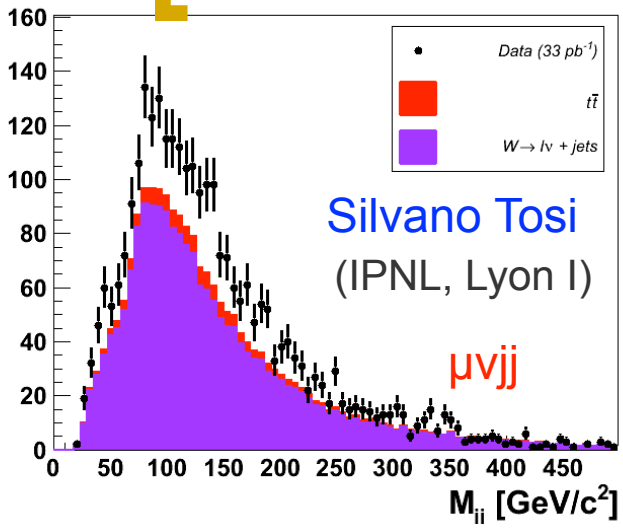


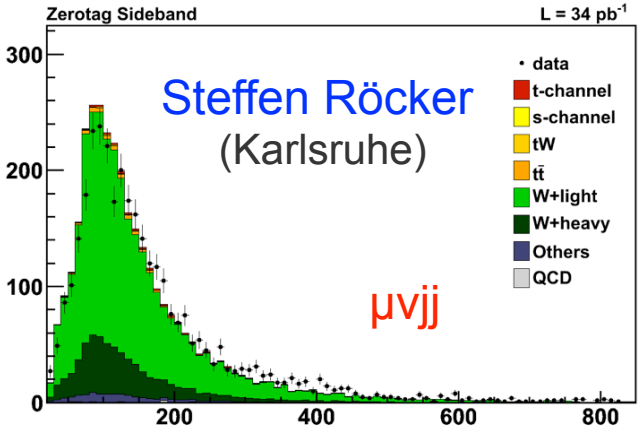


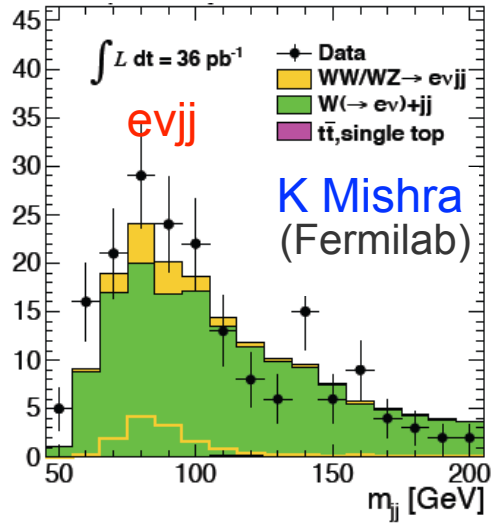




Huge excitement in CMS over CDF bump !



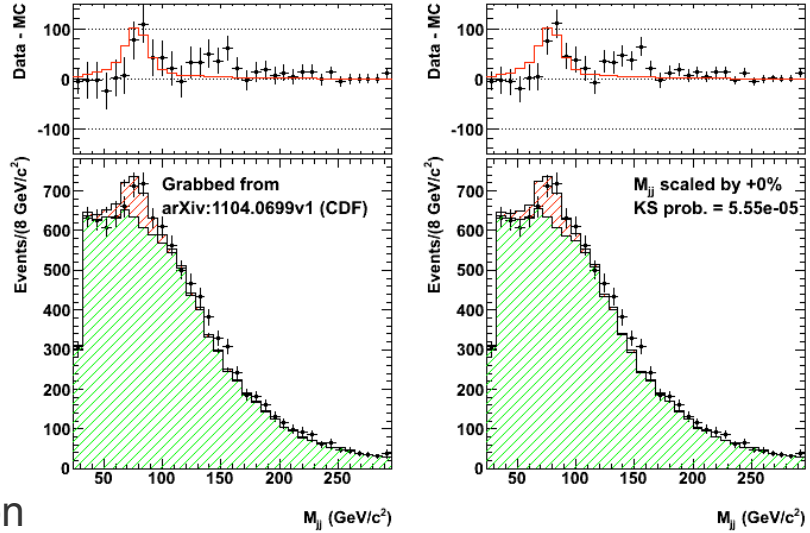


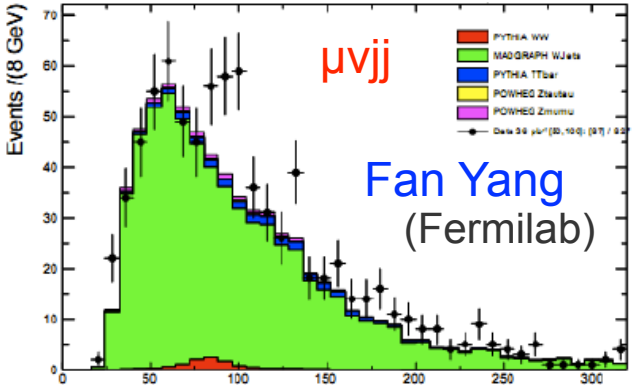


Some interesting observations from

Tommaso Tabarelli

(Milano-Bicocca)





Follow discussion in V+jets HN: https://hypernews.cern.ch/HyperNews/CMS/get/ewk-vplusjets.html

Does LHC have sensitivity with L=50 pb^{-1} ?

Process	Cross section: Tevatron (pb)	Cross section: LHC (pb)
WW	12	43
WZ	3.5	18

NLO cross sections

Ratio of WW+WZ cross section: Tevatron/LHC = 1/4
Ratio of integrated luminosity: Tevatron/LHC = 4300 / 50 = 86

- We need about 20 times more data to have the same sensitivity.
- With 4.3 fb^{-1} CDF has ~1500 WW+WZ(→lνjj) events from both e,μ channels
  - We should have about 60 events from 2010 data (WW: 43 pb * 0.33 acceptance * 0.5 efficiency * 2/9 * 2/3 BR * 36 pb^{-1} lumi = 45, similarly for WZ).
- CDF finds an excess of 253 events over a background of ~2000 in the mass window 120−160 GeV
  - Scaling to 2010 luminosity we expect an excess of 8−9 events with similar S/B
  - Clearly, not significant enough to generate excitement
  - Need ~ 1 fb^{-1} to be competitive with Tevatron. So by summer ....

However, maybe the enhanced partonic luminosity (gg and qg) at LHC can produce this access at higher rate or via other mechanism. Need to watch out.

Quick analysis of CMS W(→lν)+jj data

✦W→lν reconstruction
- E_T > 25 GeV, |η| <2.5 (2.1) for electron (muon)
- Passing Top PAG recommended isolation/Id criteria
-W transverse mass: m_T > 50 GeV, PF MET > 25 GeV
-Z veto

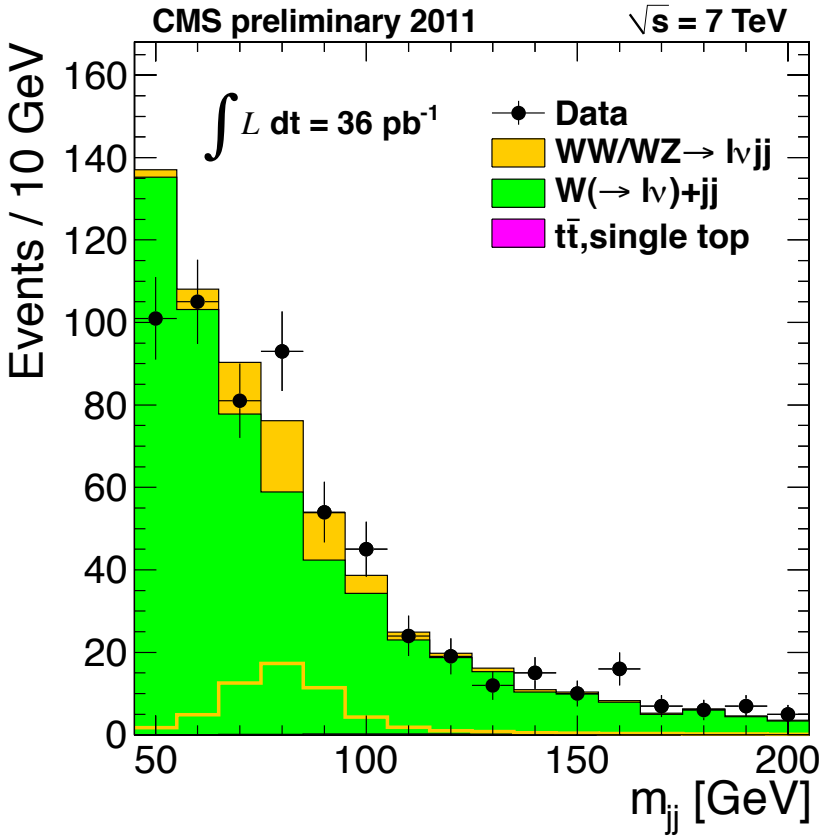
✦Require <u>two</u> PF jets in the event
-each jet with corrected p_T >20 GeV and |η| <2.4
-|cosθ*|< 0.4, |Δφ(jet1, jet2)| > 1.5
-|Δφ(jet1, MET)| > 0.6
-|Δφ(W, W) − π| < 0.2
-No b-tagged jets

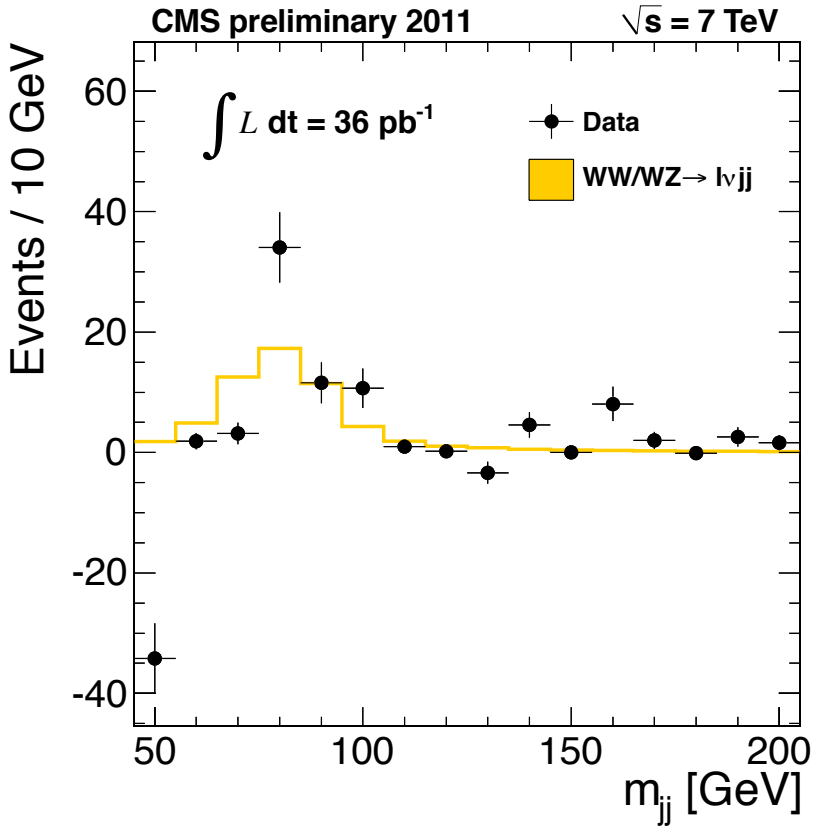
Apply standard "L2 L3" correction and "residual correction" in data

Using 36 pb^{-1} data from 2010 run

MC: W+jets: Madgraph, Top: Powheg, WW+WZ: Pythia (all with pileup conditions observed in 2010 data)

m_{jj} distribution in CMS data (e,μ combined)





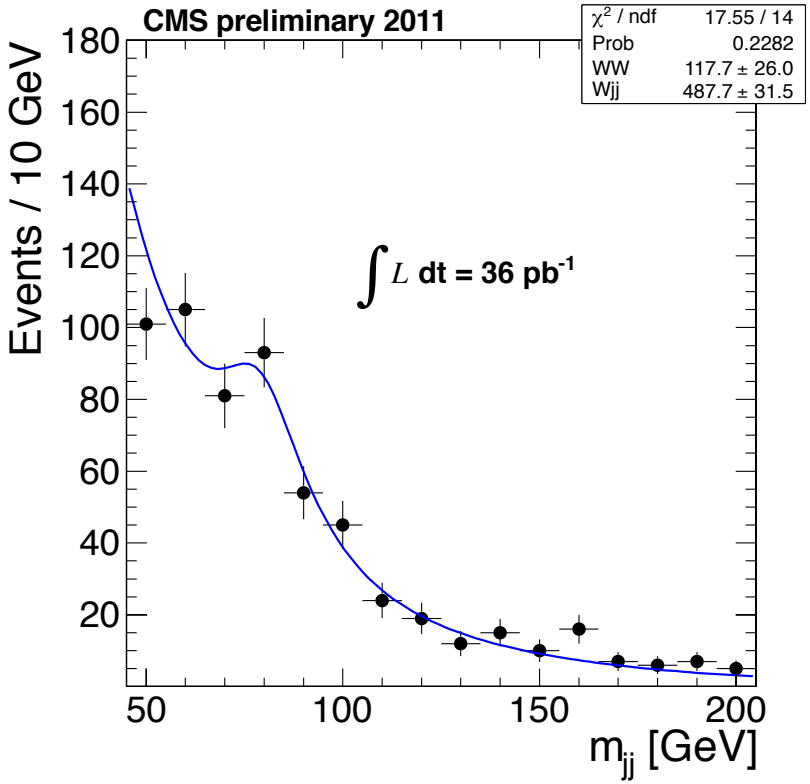
observed events in data = 600
MC predicts:
W+jj = 554
Ttbar, single top = 2
WW + WZ = 58

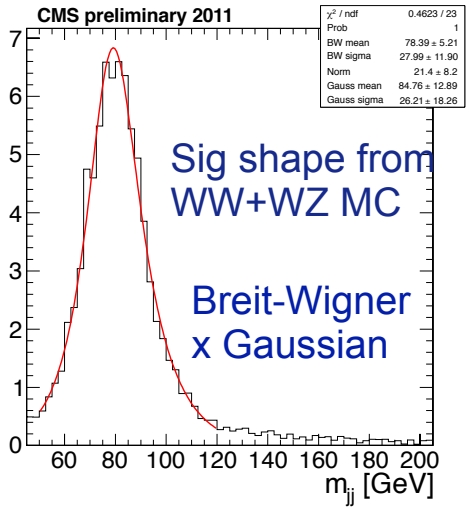
subtracting W+jj +top bkg from MC

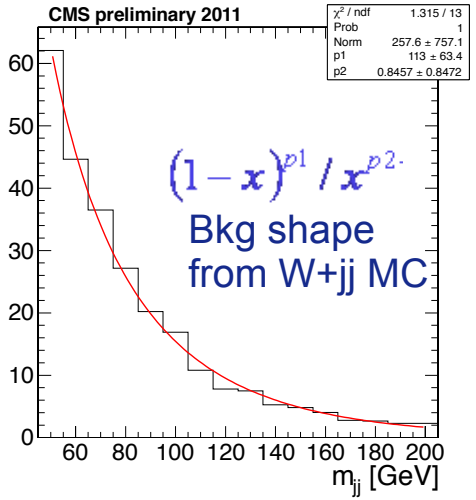
CMS data: WW+WZ signal estimation

(e,μ data combined)

Shape derived from MC.
Fit for the normalization.





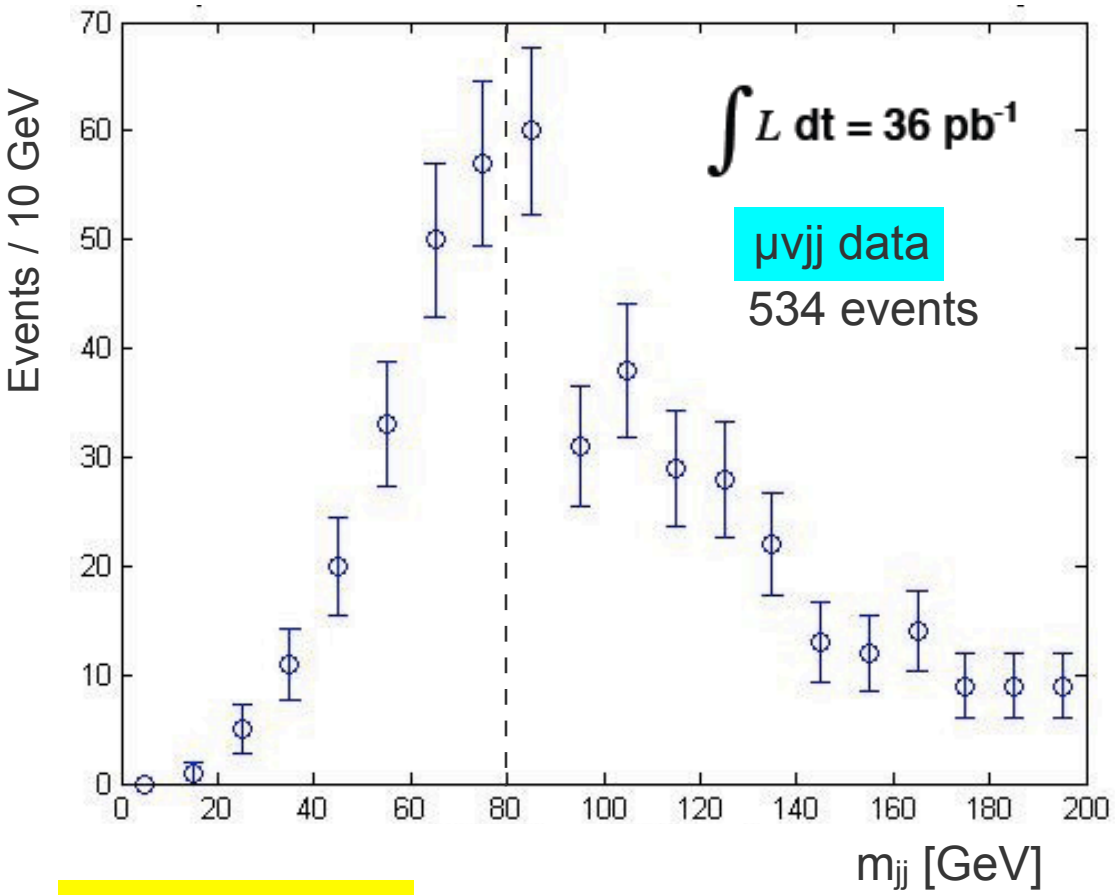


WW + WZ yield = 118 ± 26 (stat)
W+jets yield = 488 ± 32 (stat)

Clear evidence of diboson production in lνjj final state in CMS

Any excess between 120−160 GeV is not significant, but is consistent with CDF result.

Can we see W→jj peak in top events ? Yes



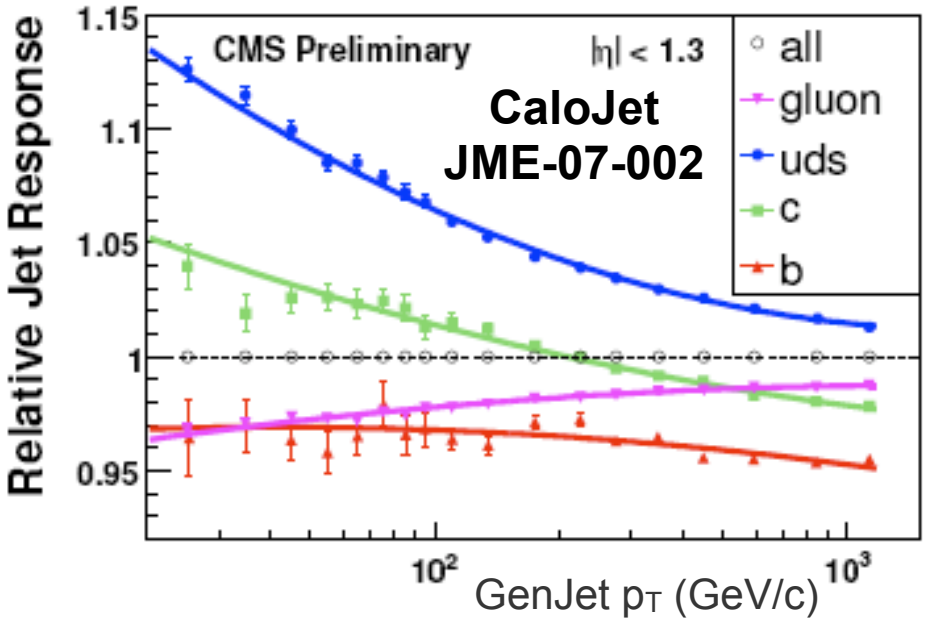
From Dan Green

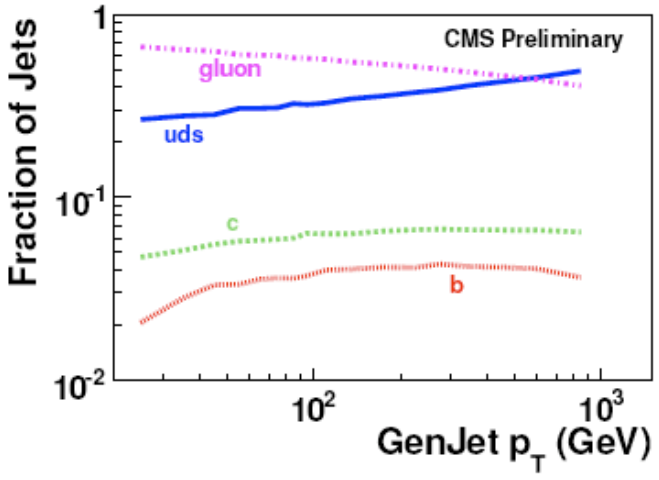
- Look at all 2010 data
- Ask for > 3 jets (p_T > 25 GeV) and ≥1 SSV – medium b tag
- Ask for no b tag for the dijet combination being plotted
- There is clear W peak near the W mass
  - -mass resolution is ~ 10 GeV
  - -peak is centered at > 80 GeV indicating over-correction of JES for light-flavored jets
  - -When jet p_T threshold is lowered to 20 GeV the over-correction is even larger
- Top is an excellent calibration sample for our purposes

Flavor dependence of jet correction

- Assuming the jet originated from a specific parton flavor
  - Light quarks have higher response than gluons because they fragment into higher momentum particles
  - QCD dijet events have mostly gluons. Default L2L3 correction corrects jets assuming same flavor composition as in QCD
  - Therefore, light flavor jets get overcorrected

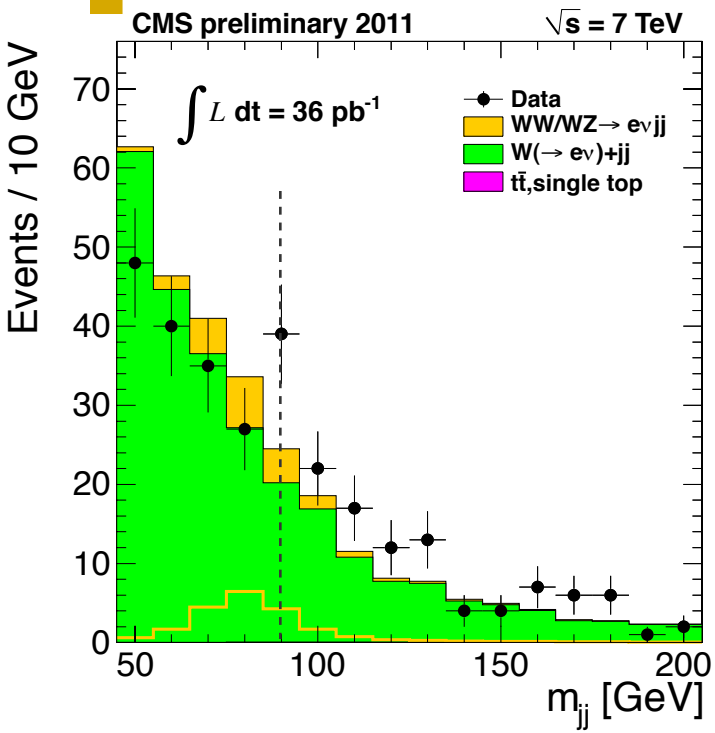
- Flavor-dependence of JEC should be smaller for PFJets. But we do not have dedicated studies. Empirical observation in data suggests significant effect at low p_T



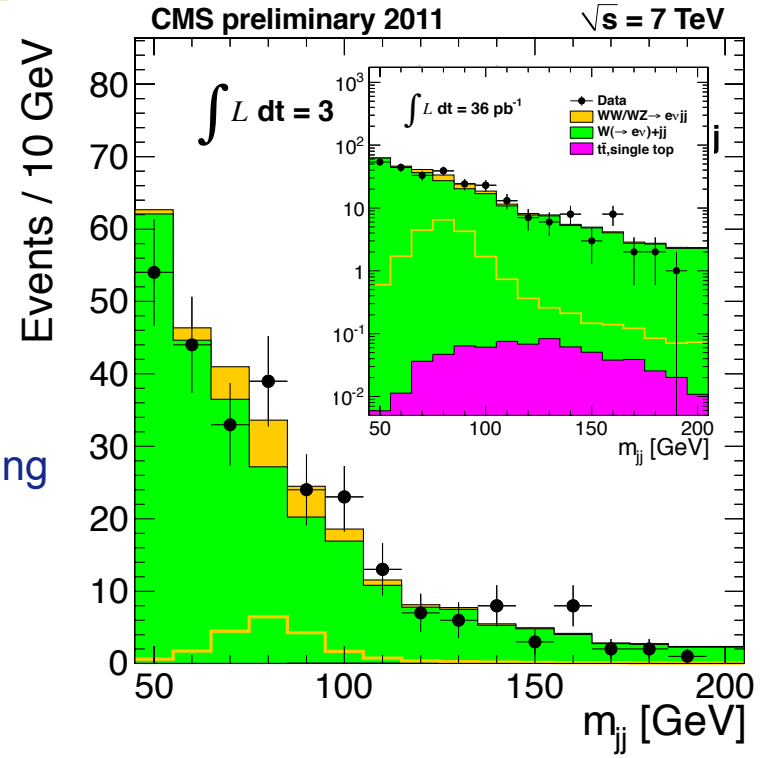


The fraction of jets from each flavor in the QCD dijet sample

m_{jj} distribution for W(→ev)+jj events



After adjusting JES



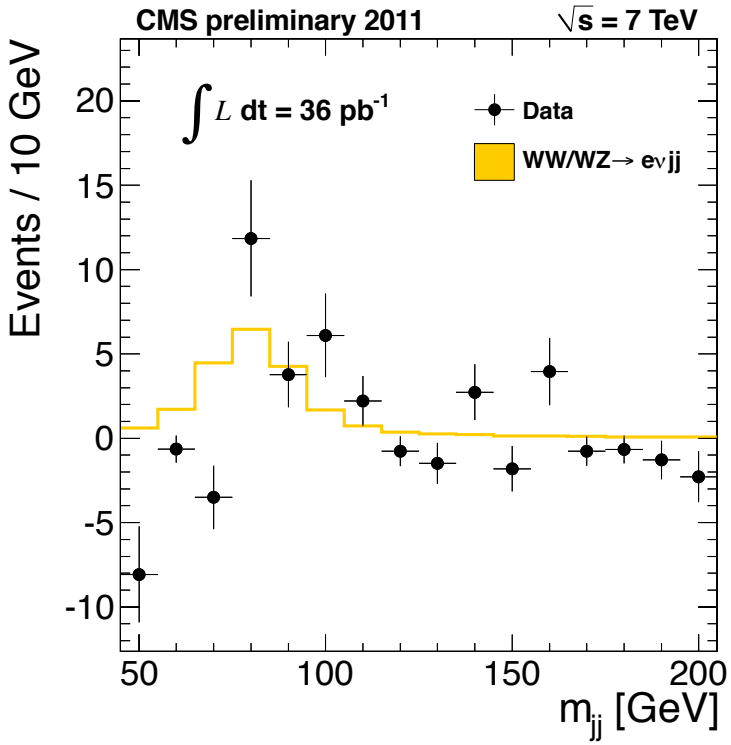
- See clear W bump
- The bump is shifted right
- Need to adjust JES to have W peak at 80 GeV

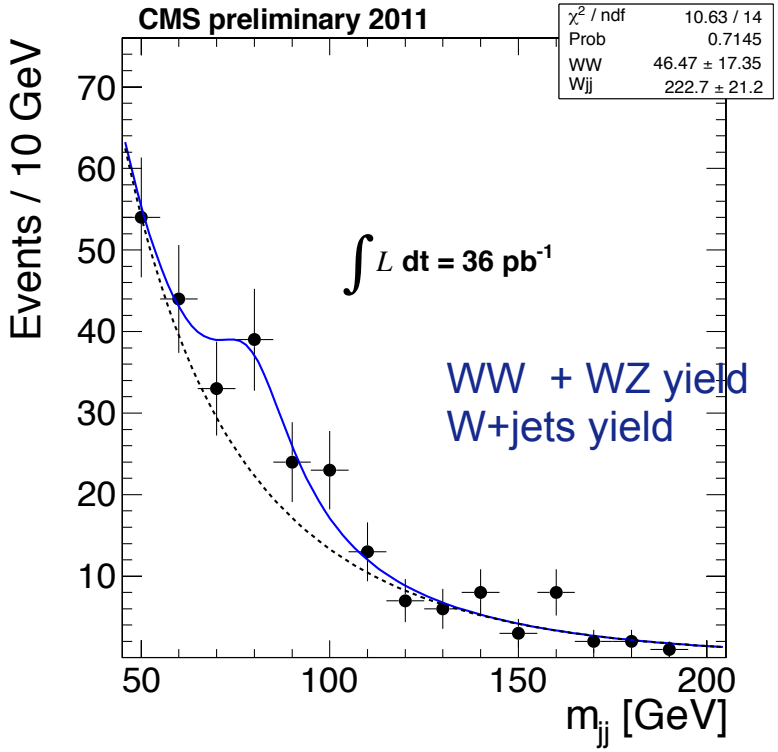
observed events in data = 267
MC predicts:
W+jj = 257
Ttbar, single top = 0.7
WW + WZ = 21

m_{jj} distribution after background subtraction

after subtracting W+jj +top bkg from MC

Shape derived from MC.
Fit for the normalization.



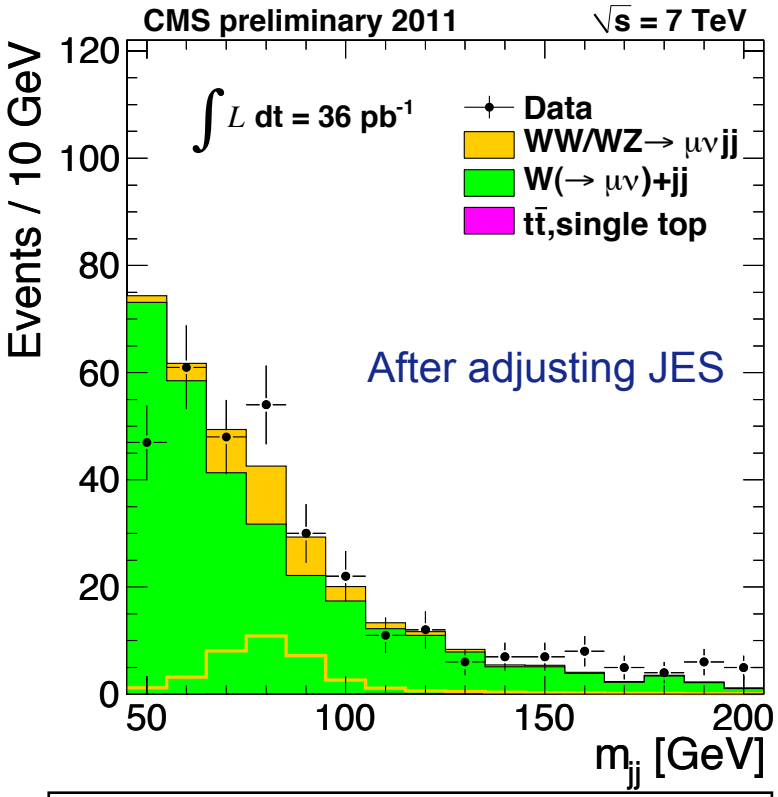


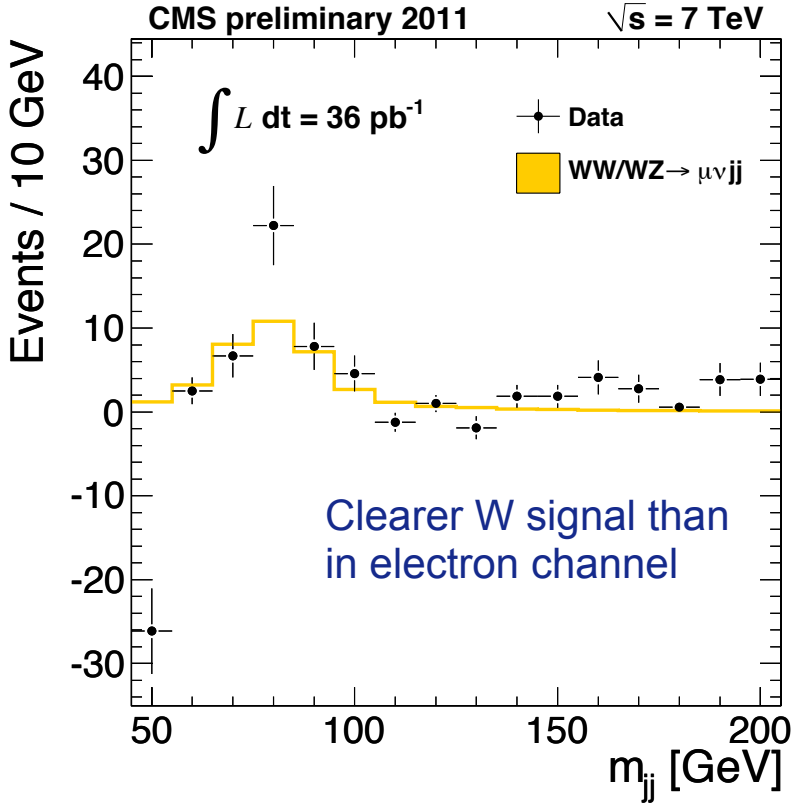
WW + WZ yield = 46 ± 17
W+jets yield = 223 ± 21

Systematic effects haven't been studied. Potentially large.

Need to understand W+jets tail and effect of JES.

m_{jj} distribution for W(→μν)+jj events





observed events in data = 333
MC predicts:
W+jj = 297
Ttbar, single top = 1.3
WW + WZ = 37

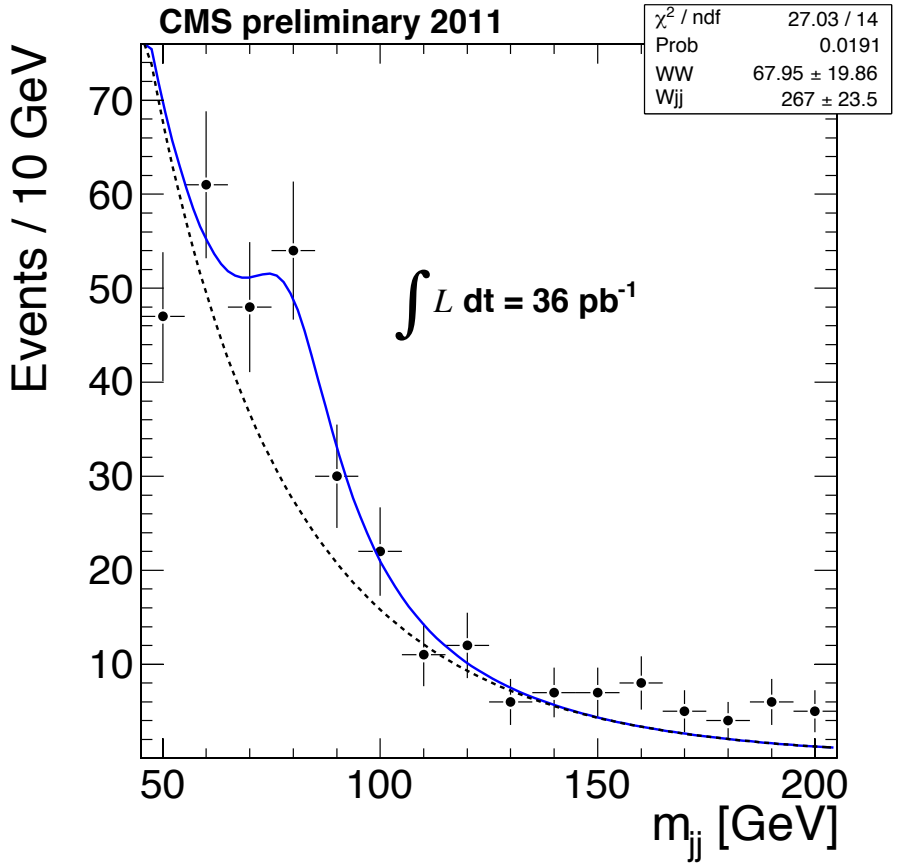
subtracting W+jj +top bkg from MC

From Fan Yang

m_{jj} distribution: WW+WZ signal estimation

From Fan Yang

Shape derived from MC.
Fit for the normalization.



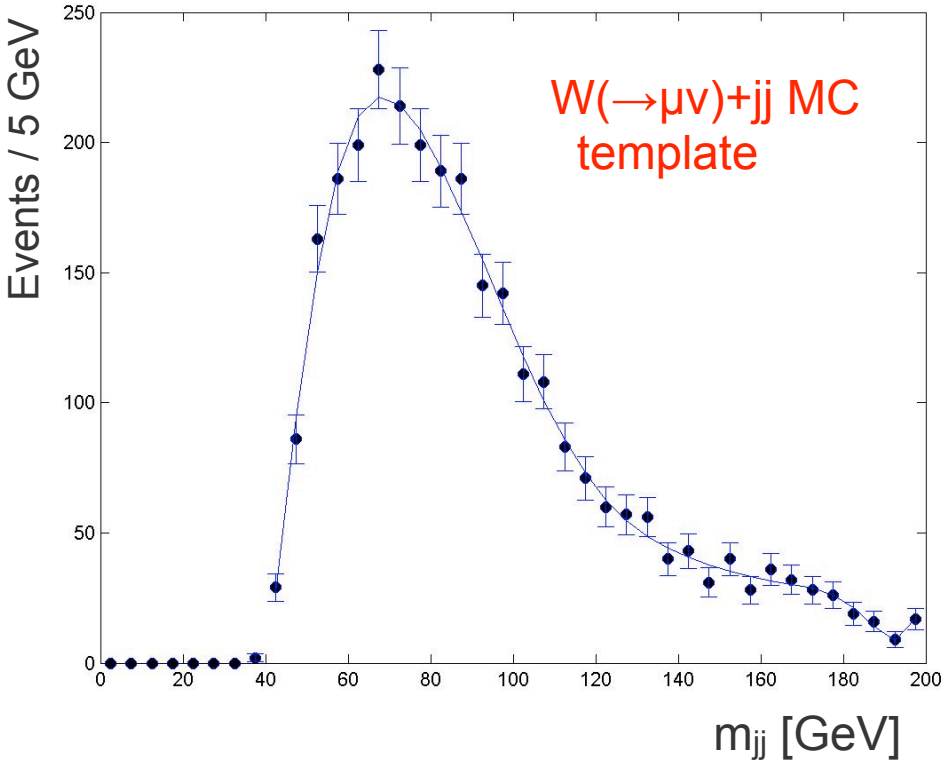
WW + WZ yield = 68 ± 20 (stat)
W+jets yield = 267 ± 24 (stat)

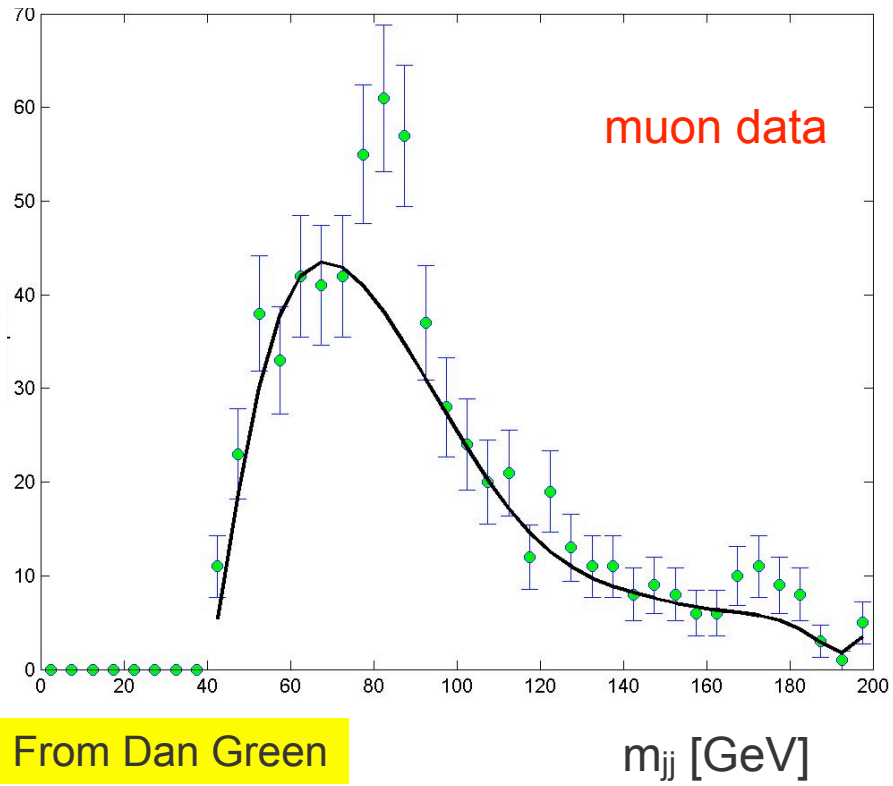
Systematic effects haven't been studied. Potentially large.

Clearly need to understand W+jets tail and effect of JES

WW signal is robust w.r.t. selection cuts (I)

Dan Green performed analysis using top PAG skim of muon data, and somewhat different selection cuts. He gets the same number of signal events.



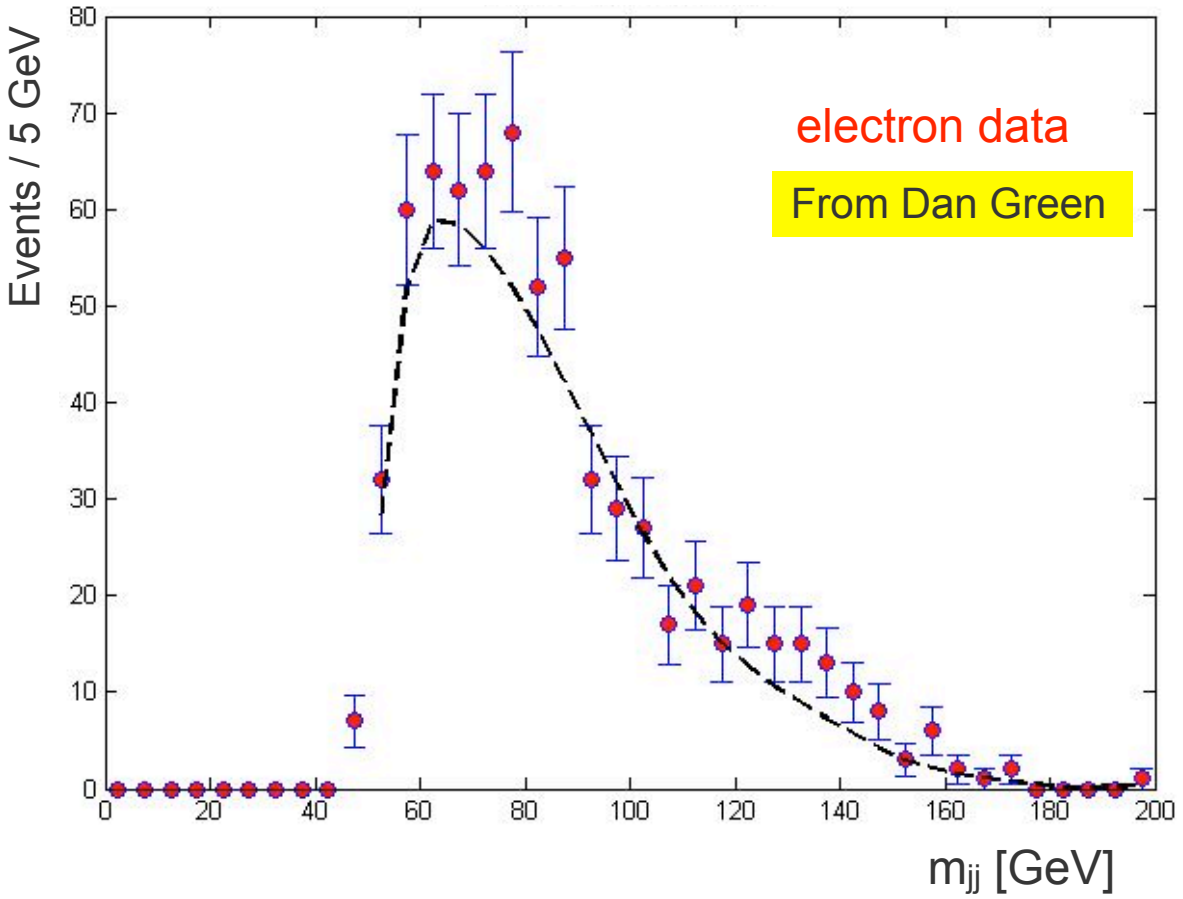


From Dan Green

The fit after all cuts is to the Wjj MC shape. The evidence remains for a W signal in the μ data with ~ 60 ± 20 events above a smooth background.

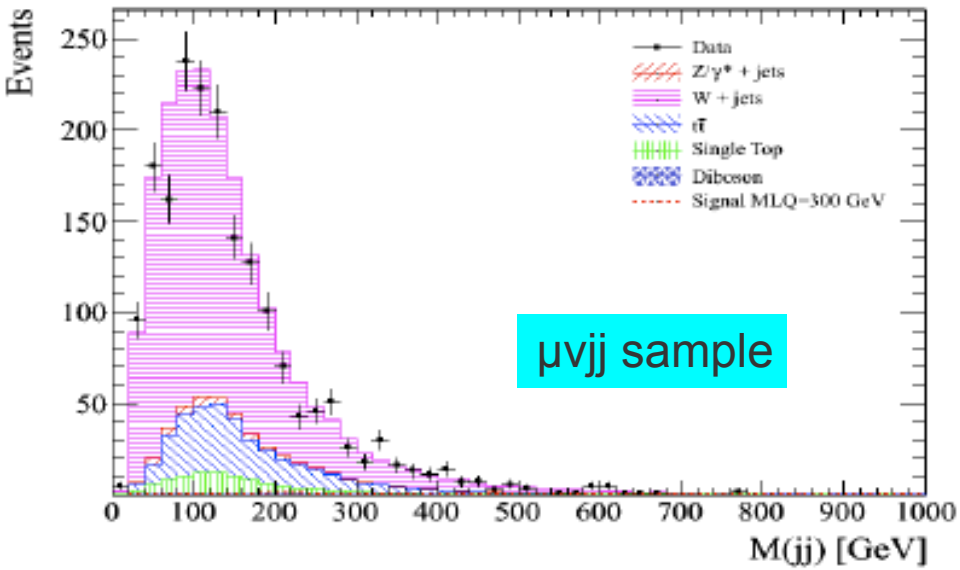
WW signal is robust w.r.t. selection cuts (II)

Dan's analysis of electron data using top PAG skim and somewhat different selection cuts. He gets the similar number of signal events.



Exotica analysis by lepto-quark group: μ data

From Anne-Fleur Barfuss



μνjj sample

Excellent cross check, although limited by existing ntuple/skim cuts

See details at:

https://indico.cern.ch/conferenceDisplay.py?confId=131845 (Anne-Fleur Barfuss)

- 1 muon Pt>30 GeV
- MET>25 GeV, MT(μν)>30 GeV
- DeltaR(μ,jet)>0.5
- At least two PFjets with ET>30 GeV, $|\Delta\eta|<2.5$
- PT(jj)>40 GeV
- $\Delta\phi$(MET,1st jet)>0.4

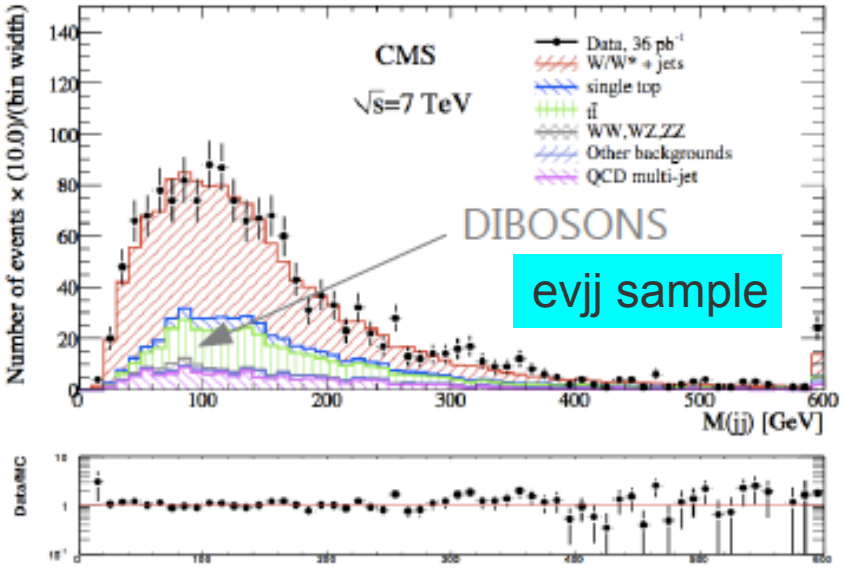
- No isolated electron with PT>15 GeV
- Tighter PT lepton selection than CDF due to existing ntuple skims
- Background shapes from MC
- W+jets rescaling from data in range 50<MT<110 GeV

Find no excess in data compared to MC

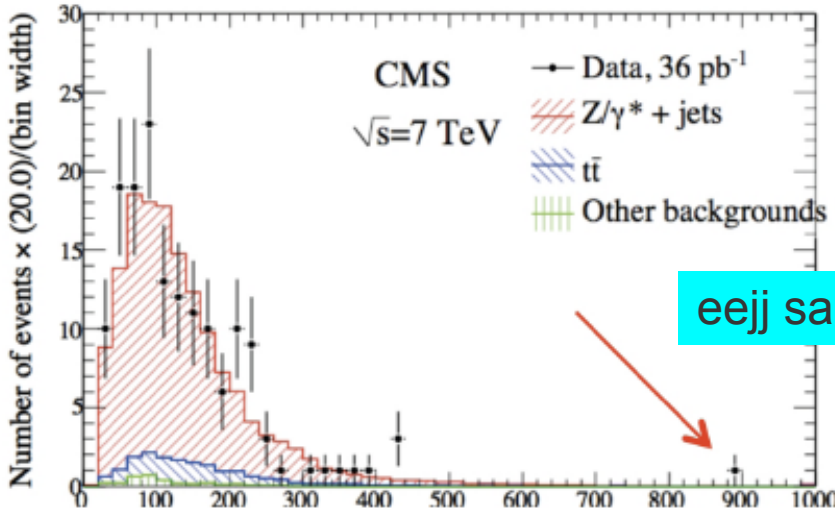
Exotica analysis by lepto-quark group: ele data

Find no excess in data below 200 GeV

From Anne-Fleur Barfuss



evjj sample



eejj sample

1 ele Pt>35 GeV, |η|<2.2

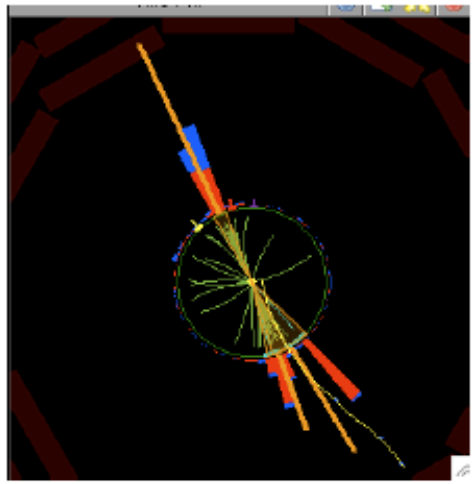
MET>30 GeV, MT(eν)>30 GeV

DeltaR(ele,jet)>0.7

At least two PFjets with ET>30 GeV, |η|<3.0, |Δη|<2.5

PT(jj)>40 GeV

Δφ(MET,1st jet)>0.5



M_{jj} (GeV)

DeltaR(ele,jet)>0.7

At least two PFjets with ET>30 GeV |η|<2.4, |Δη|<2.5

PT(jj)>40 GeV

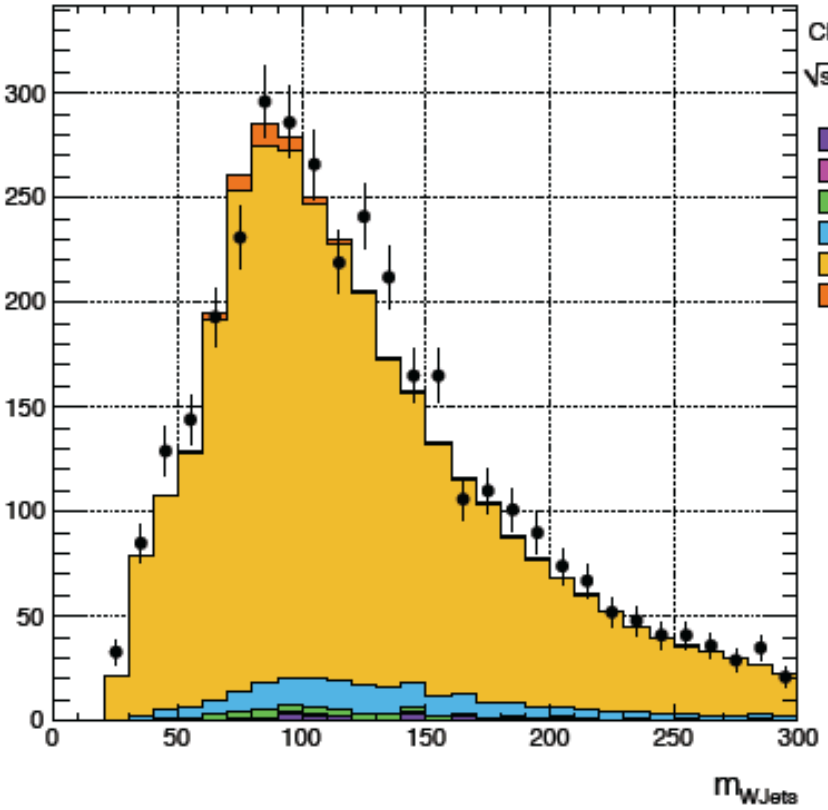
Mee>50 GeV

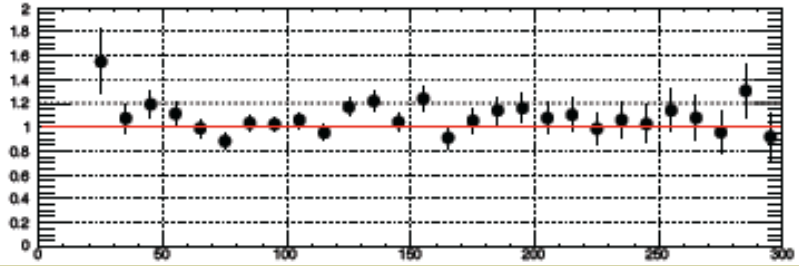
Background shapes from MC

No data-driven rescaling applied

Analysis by Higgs WW(lvjj) group (I)

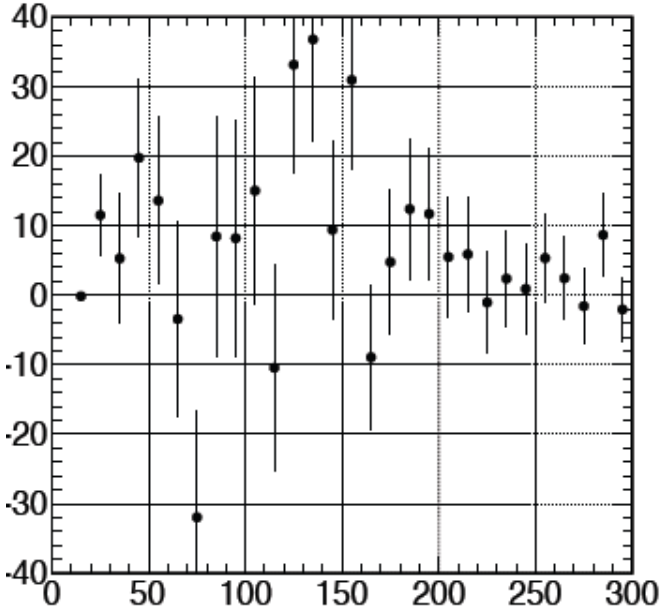
From Andrea Benaglia





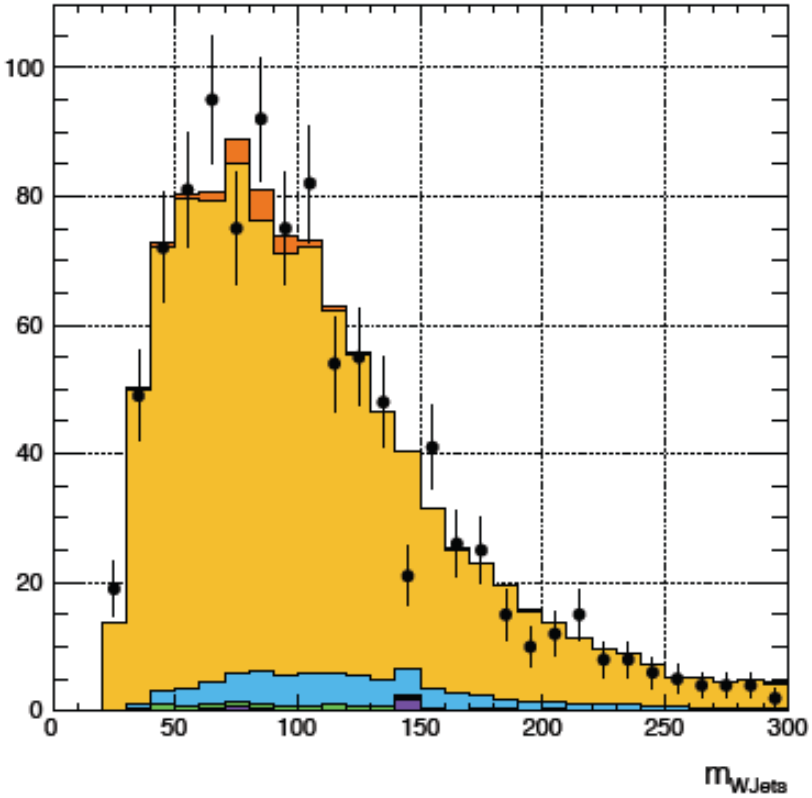
Event selection

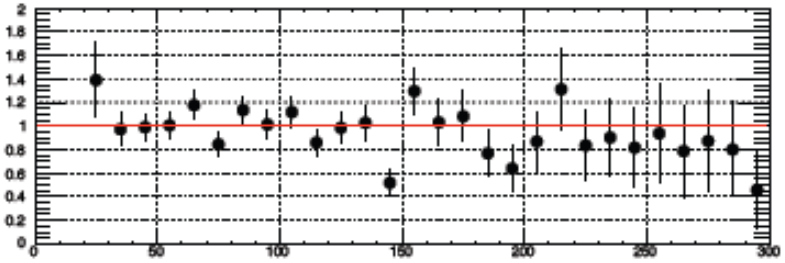
✦W→lν reconstruction: $E_T^l > 20$ GeV, $|\eta| < 1.5$, MET > 25 GeV, $m_T > 40$ GeV
✦2 jets with $p_T > 30$ GeV and $|\eta| < 2.4$



After bkg subtraction

Analysis by Higgs WW(lvjj) group (II)

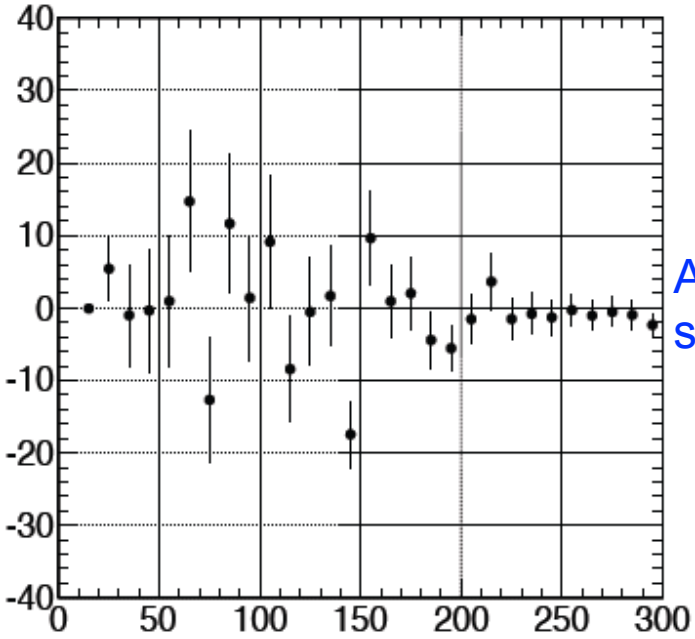




From Andrea Benaglia

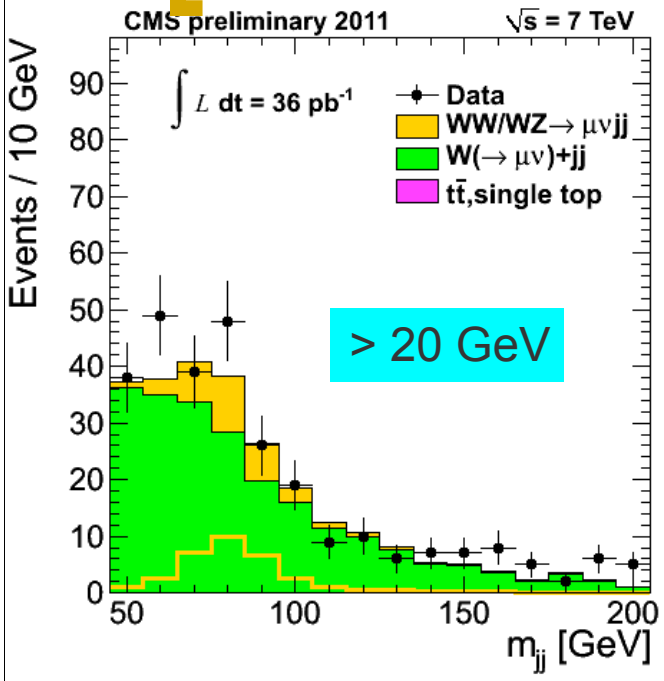
Tighten cuts to enhance diboson peak

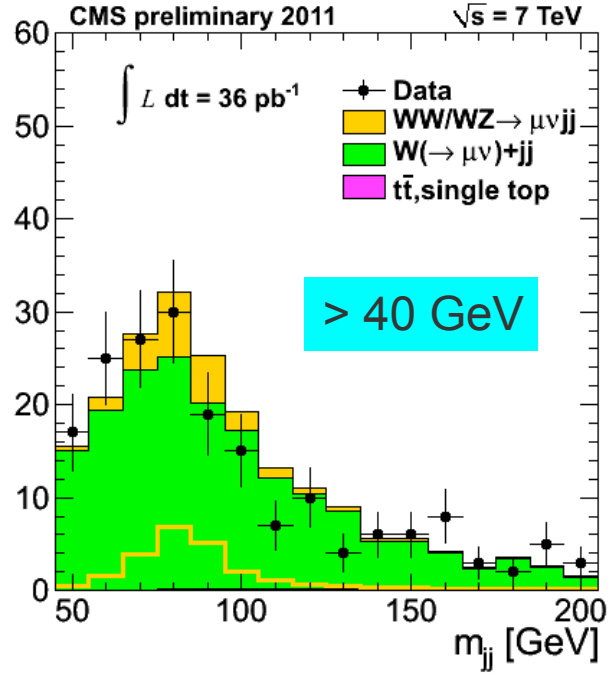
- $|\Delta\varphi(\text{lepton},E_T^{miss}) - \pi| < 0.2$
- $|\Delta\eta_{jj}| < 1.5$
- very small enhanchement, in fact

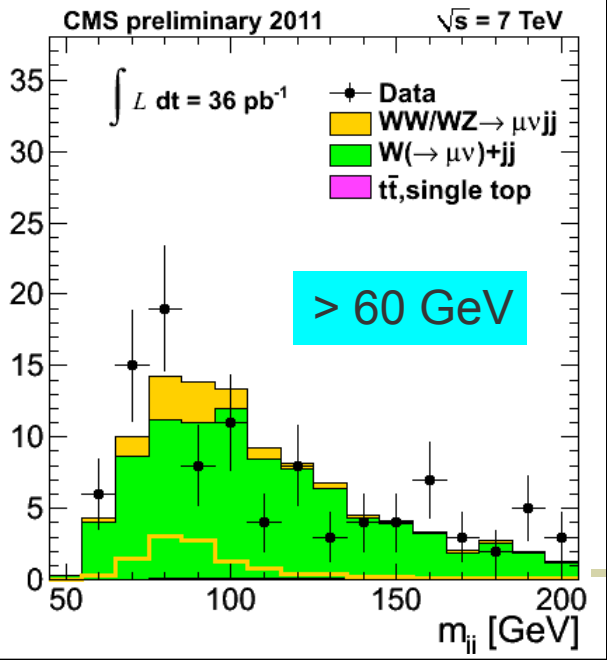


After bkg subtraction

Effect of p_T^{dijet} cut







The dijet p_T cut does not seem to be enhancing any resonance in 120−160 GeV.

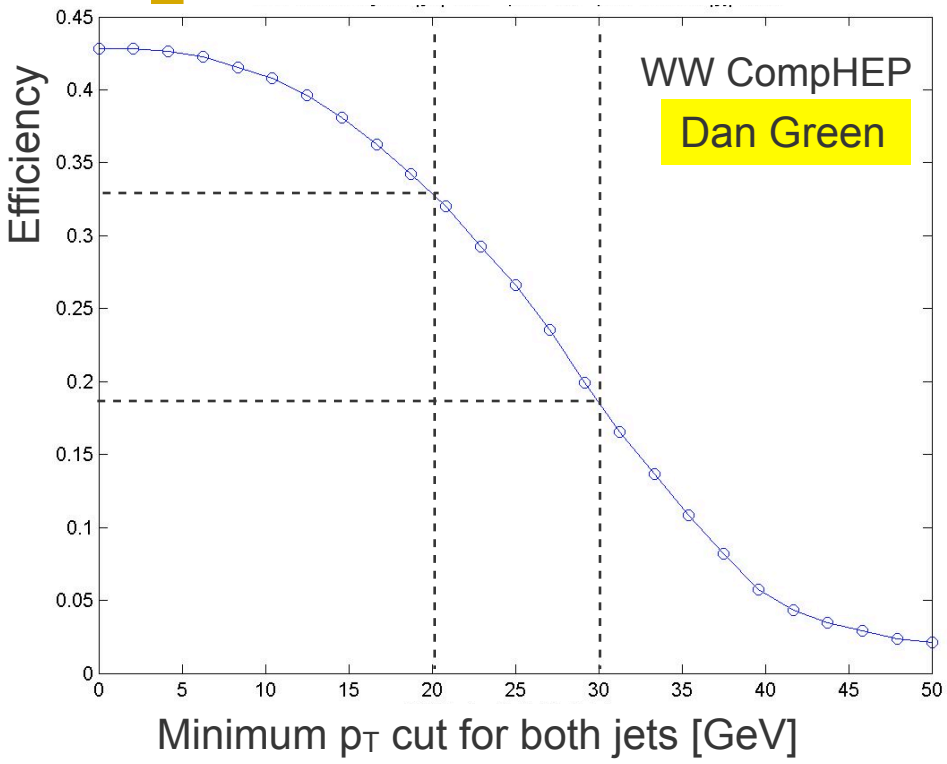
Next steps for CMS

- Measure the electroweak WW+WZ(→lvjj) production rate
  - With both 0 and 1 extra jet
  - This will help constrain the JES and nuisance parameters
  - Thoroughly understand the shape of the W+jets background

- Examine the 2011 data as the integrated luminosity increases
  - Trigger is going to be the most critical element in the high lumi high pileup scenario
  - we live by the trigger or die with the trigger
  - Jets and MET with PU are likely to be crucial to optimize.

- With a firm understanding of the WW production
  - Will be well positioned to search for resonances (a la CDF bump) and Higgs
  - Need ~ 1 fb^{-1} to have sensitivity in the interesting kinematic regime
  - Should have it by the end of summer, i.e., soon.

Keeping W+jj events: big trigger challenge



- WW reconstruction efficiency is a strong function of the cut on jet p_T.
  - For a cut on both jets of 25 GeV, the overall efficiency is ~25%.
  - 18% for 30 GeV cut
- We lose ~1.5% of efficiency for each 1 GeV increase in jet p_T threshold

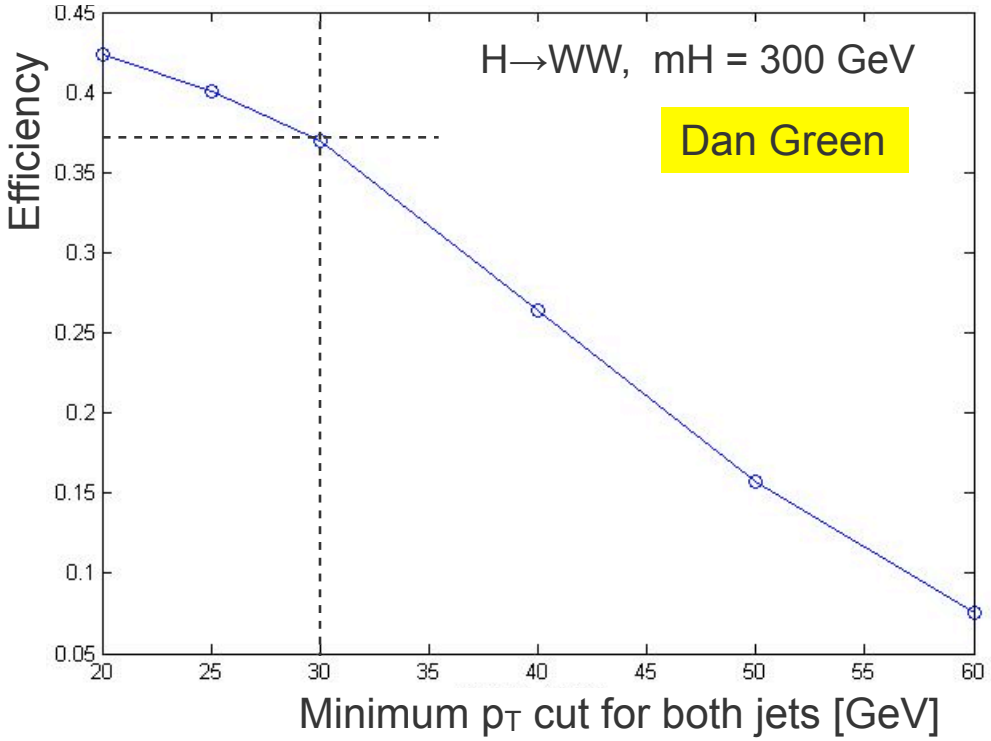
HLT trigger rate (Hz)

Andrea Benaglia & Leonardo di Matteo

L=	5.00E+32	1.00E+33	2.00E+33
Ele15_Jet30_Jet30	4.05	8.1	16.2
Ele17_Jet30_Jet30	3.70	7.4	14.8
Ele22_Jet30_Jet30	2.98	5.96	11.92

Adding MET >25 GeV will reduce the rate by a factor 3.5 ➜ viable for 2E33 menu with electron E_T>22 GeV. But MET is uncertain in high PU/OOP. Jet threshold still too high.

Effect of jet threshold on H→WW analysis



- For a Higgs this heavy a jet threshold of 30 GeV in p_T remains reasonably efficient.
- However, the light Higgs should not be thrown away if it can be retained.
- Even in the mass range 160-200 GeV semileptonic decay mode can add significantly to Higgs sensitivity.

Summary

✦Huge excitement over CDF result in Wjj
- in both experimental and theoretical community
- CMS needs an order of magnitude more data for similar sensitivity
- with 36 pb^{-1} data ~ 20−30 WW candidates in each lepton channel
- making progress on signal extraction techniques, S/B optimization, ...

→ efforts ongoing to understand W+jj background shape

→ and effect of PU on sig/bkg, MET resolution, jet pT, lepton iso,...

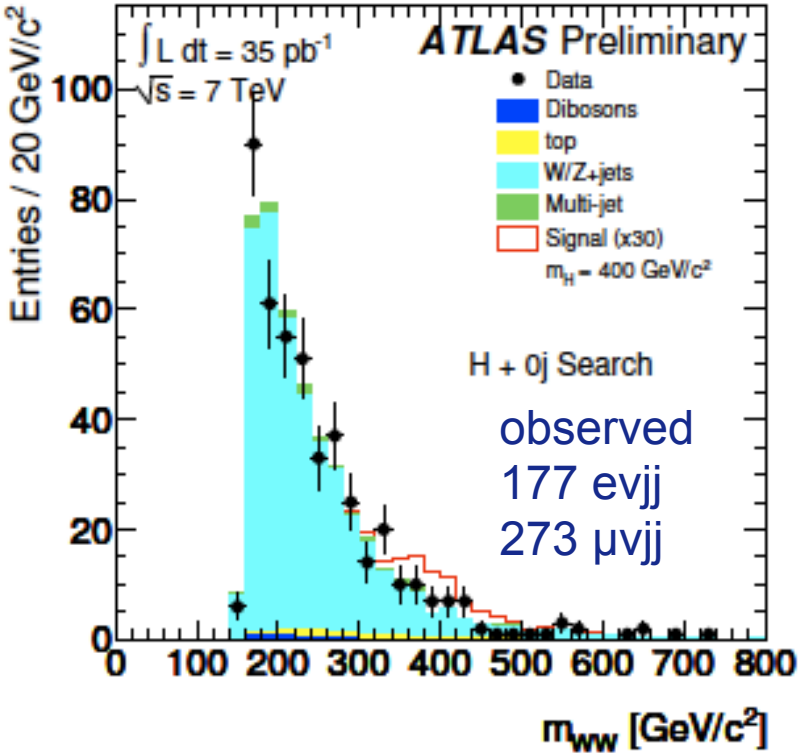
✦Analysis strategy for summer
- work hard to have sensible trigger in place to keep most of the interesting Wjj events in high luminosity and pileup conditions
- measure WW production rate, discover of H→WW and dijet resonance in Wjj events

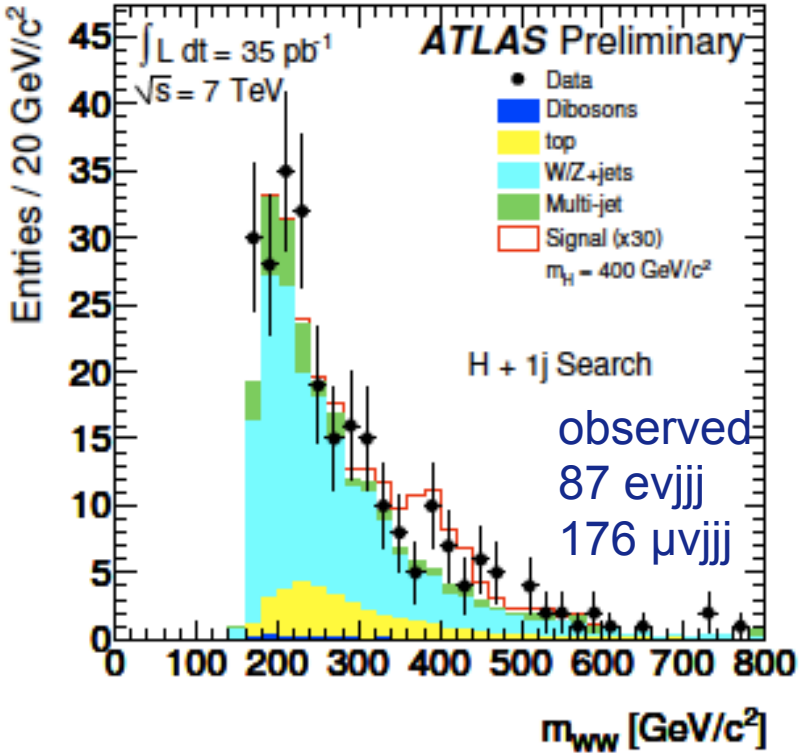
✦Many thanks to Jeff Berryhill, Andrea Benaglia, Dan Green, Leonardo di Matteo, Fan Yang, Francisco Yumiceva, everyone mentioned on slides 6−7, and the PAG conveners for providing plots/results/insights.

BACKUP SLIDES

ATLAS Higgs search limit

ATLAS-CONF-2011-052 (Apr 5, '11)

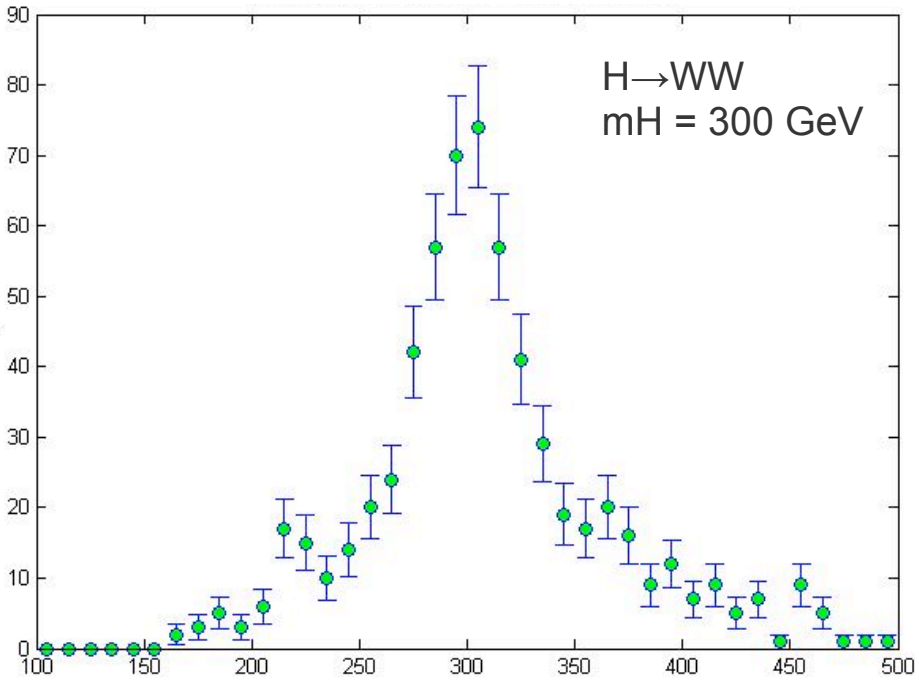




- W→lν reconstruction
  E_T^l > 30 GeV, MET > 30 GeV, Veto any second lepton in the event
- 2 or 3 jets with p_T >30 GeV, |η| <4.5
  both jets from W decay should have |η| <2.8 and 71 < mjj < 91 GeV

CMS m_{WW} mass resolution for mH = 300 GeV

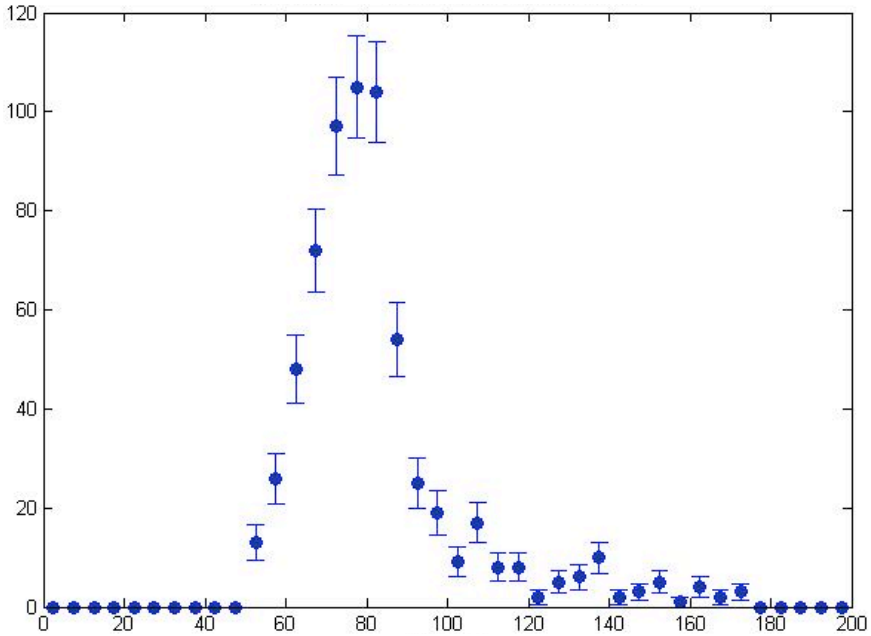
From Dan Green



m_{lvjj} [GeV]

Once the WW signal is established, look for resonant substructure.

The dM/M resolution in the dijet mass is ~ 13%.

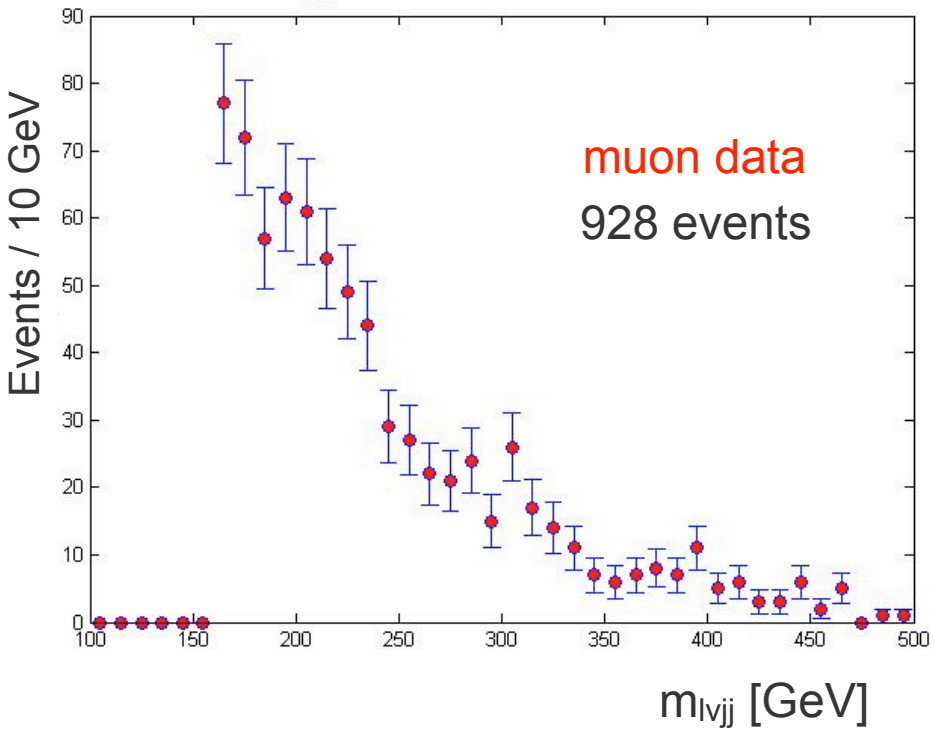


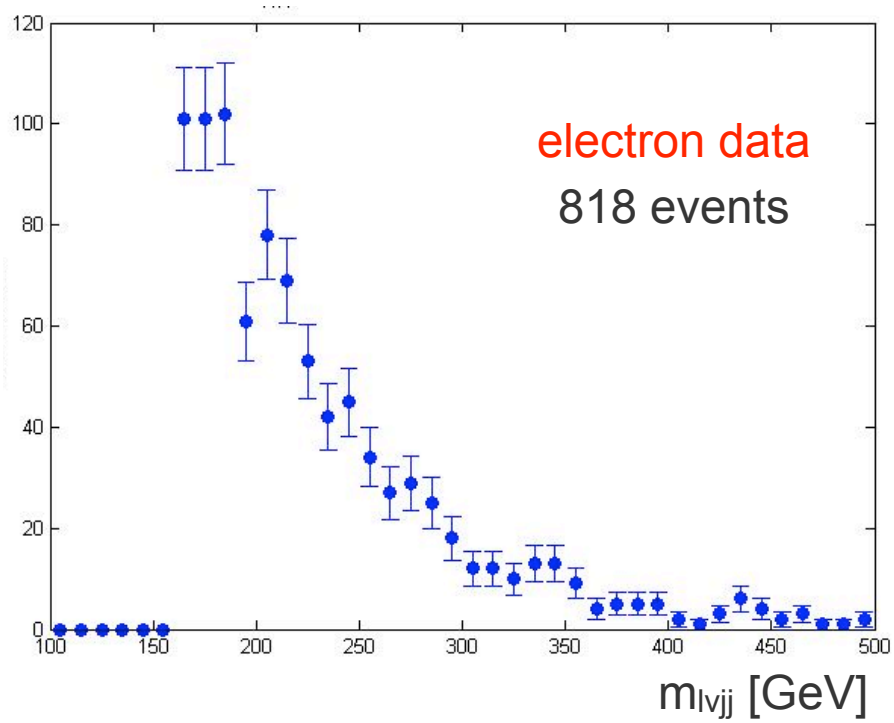
m_{jj} [GeV]

After the kinematic fit, and asking for no b tags of the 2 jets, the Higgs mass distribution has a ~20 GeV resolution, or dM/M ~ 7 %.

CMS m_{WW} distribution in data

From Dan Green

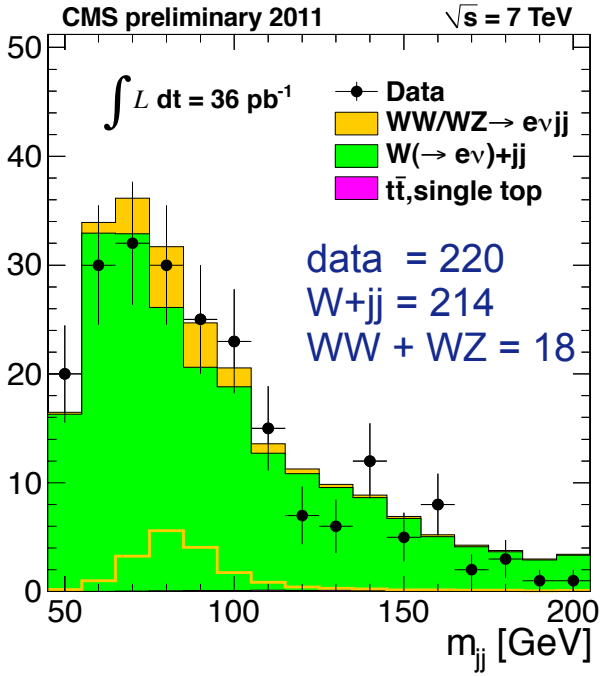




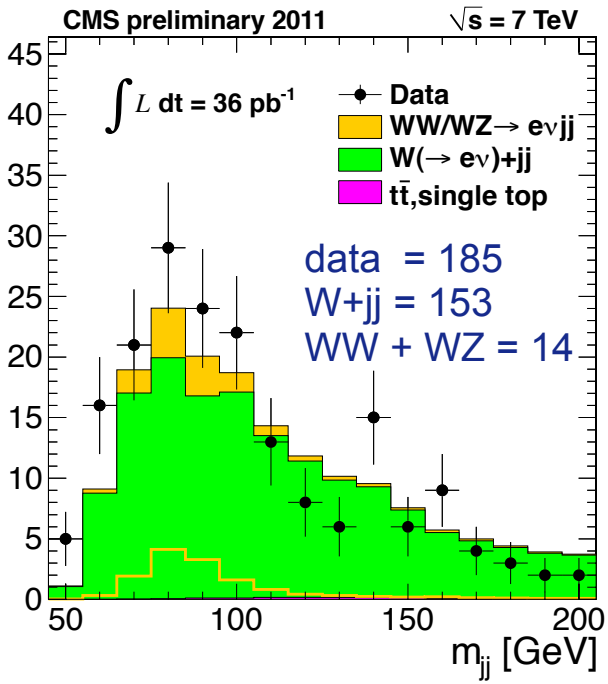
Not directly comparable with ATLAS because of different threshold cuts.
However, the distributions look similar.

Another consequence of higher jet threshold

jet p_T > 25 GeV

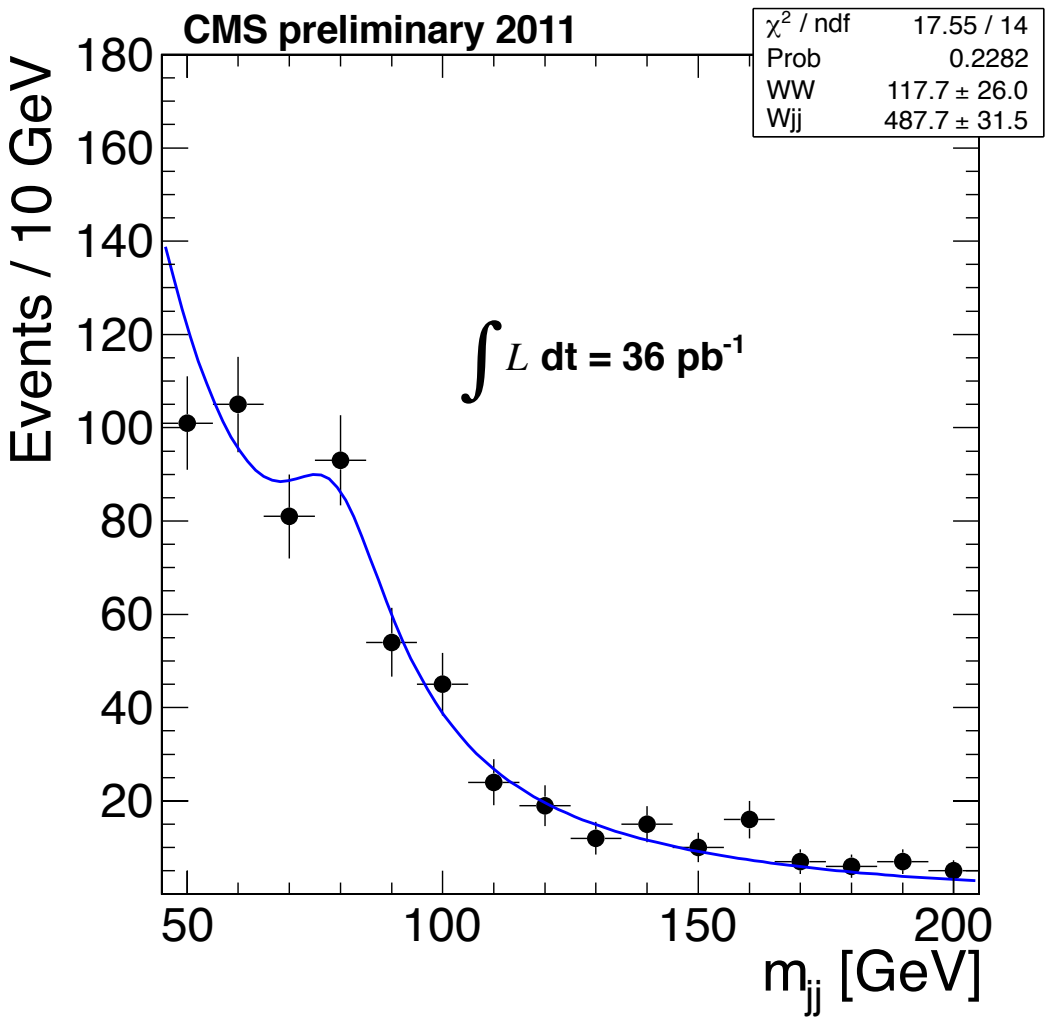


jet p_T > 30 GeV



W+jj and WW are peaking close to each other. It is harder to discriminate between their shape although the absolute number of W+jj has gone down.

CMS data: WW+WZ signal estimation



(e,μ data combined)

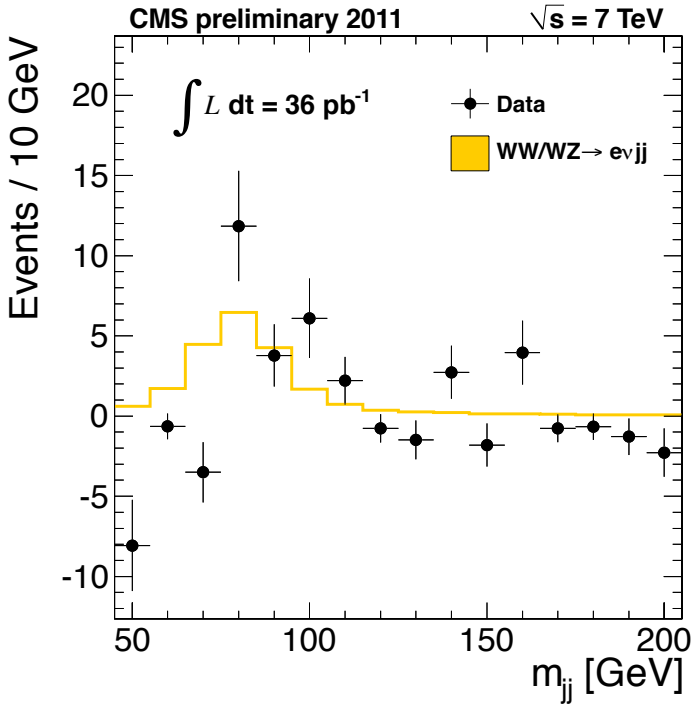
Shape derived from MC.
Fit for the normalization.

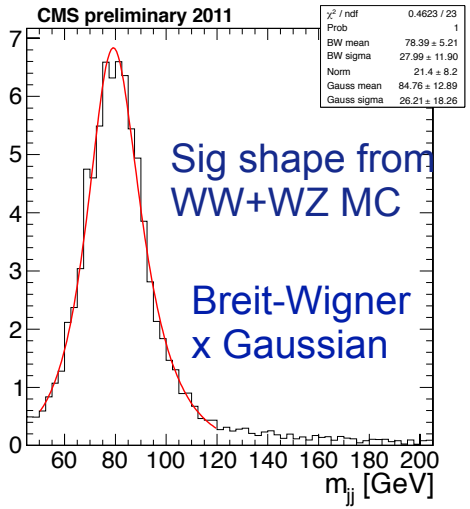
WW + WZ yield = 118 ± 26
W+jets yield = 488 ± 32

Clear evidence of diboson production in lνjj final state in CMS

m_{jj} distribution after background subtraction

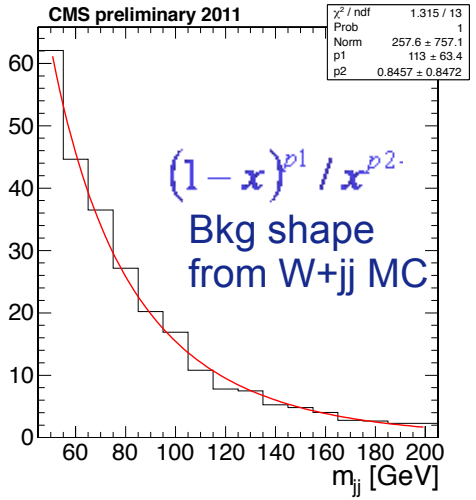
subtracting W+jj +top bkg from MC



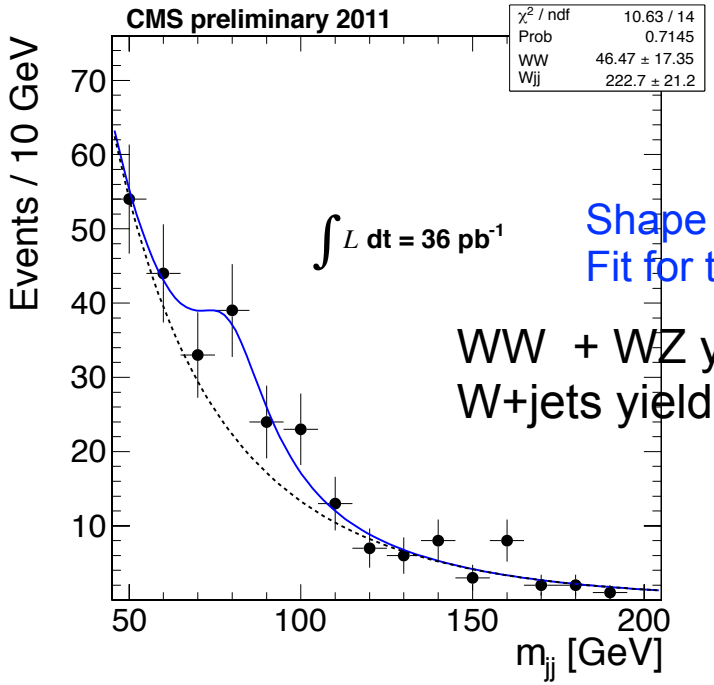


Sig shape from WW+WZ MC

Breit-Wigner x Gaussian



Bkg shape from W+jj MC

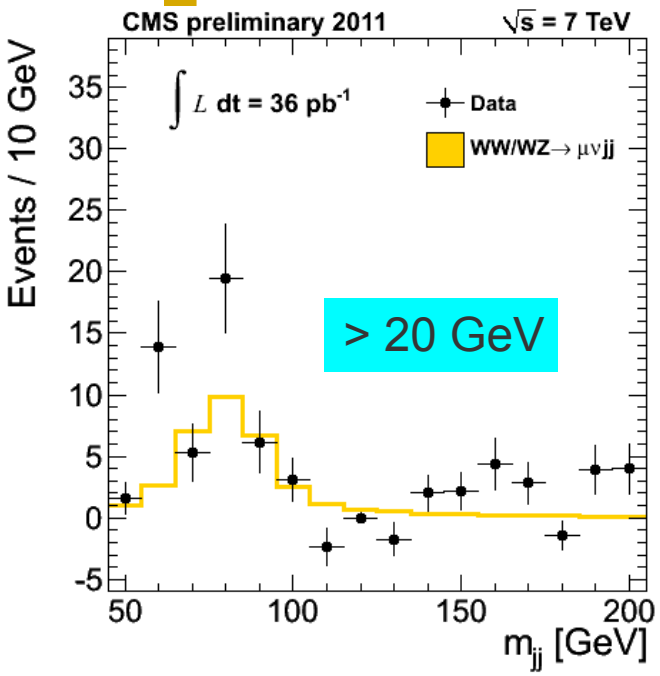


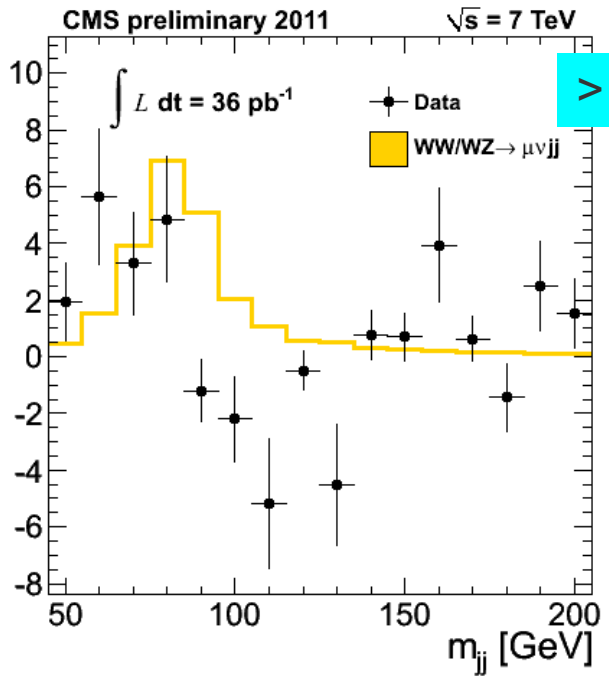
Shape derived from MC.
Fit for the normalization.

WW + WZ yield = 46 ± 17
W+jets yield = 223 ± 21

Any excess between 120−160 GeV is not significant, but is consistent with CDF result.

Effect of p_T^{dijet} cut (after bkg subtraction)





Same conclusion as on the last slide.
Are we seeing the failure of Madgraph W+jets MC to describe data for these boosted events ?