Study of diboson WW,WZ production in W(→lν)+jj events

Analysis Note: AN-11-151, PAS: EWK-11-011

Jake Anderson[1], Jeffrey Berryhill[1], Ricardo Eusebi[2], Dan Green[1], Pratima Jindal[3], Kalanand Mishra[1], Ilya Osipenkov[2], Alexx Perloff[2], Andre Sznajder[4], Fan Yang[1], and Francisco Yumiceva[1]

[1] Fermi National Accelerator Laboratory, Batavia, Illinois, USA
[2] Texas AM University, College Station, Texas, USA
[3] University of Nebraska at Lincoln, Nebraska, USA
[4] Universidade do Estado do Rio de Janeiro (UERJ), Brazil

EWK pre-approval meeting
(July 5, 2011)

Outline

- ✦The goal of this analysis
- ✦Introduction to di-boson production
- ✦Analysis steps, data / MC samples, trigger
- ✦Event selection, acceptance, efficiency
- ✦Signal extraction
- ✦Cross section computation
- ✦Likely improvements
- ✦Conclusion

Goal of this analysis

Ultimately three things:

1. Establish Standard Model electroweak WW+WZ(→lvjj) production
   - with both 0 and 1 extra jet − if feasible
     - but focus on 0 extra jet for now

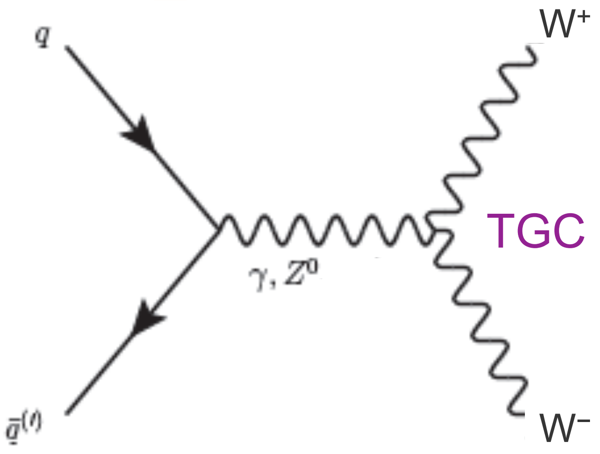
2. Measure cross section for di-boson production with this decay mode
   - crucial input for H→WW (lvjj) analysis and Wjj bump hunt

3. Understand basic features of di-boson production
   - set limit on anomalous gauge coupling

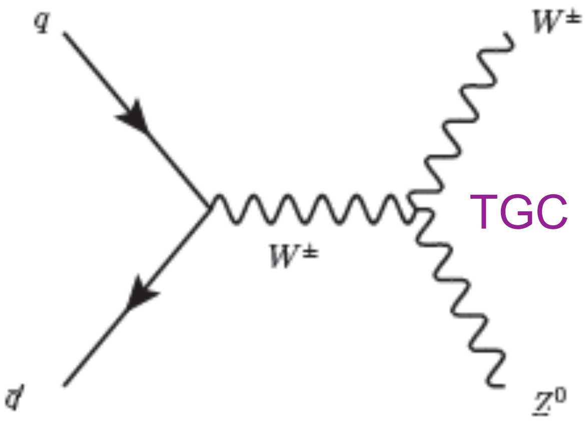
For EPS (summer conferences) focus on #1 and preliminary result for #2. By August aim to get #3 done and complete the analysis in time for Fall conferences.

Diboson production at LHC

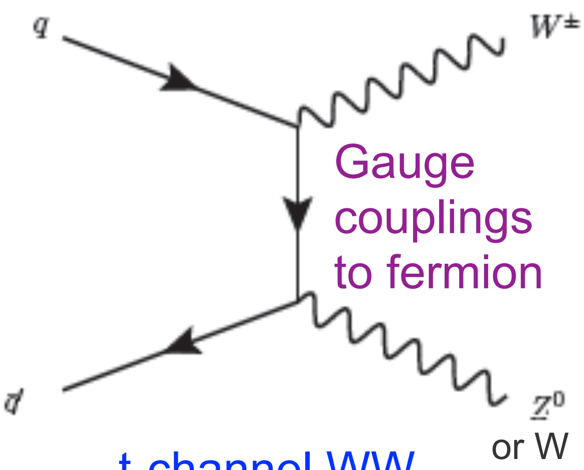
at Leading Order in α_S



s-channel WW production



s-channel WZ production



t-channel WW, WZ production

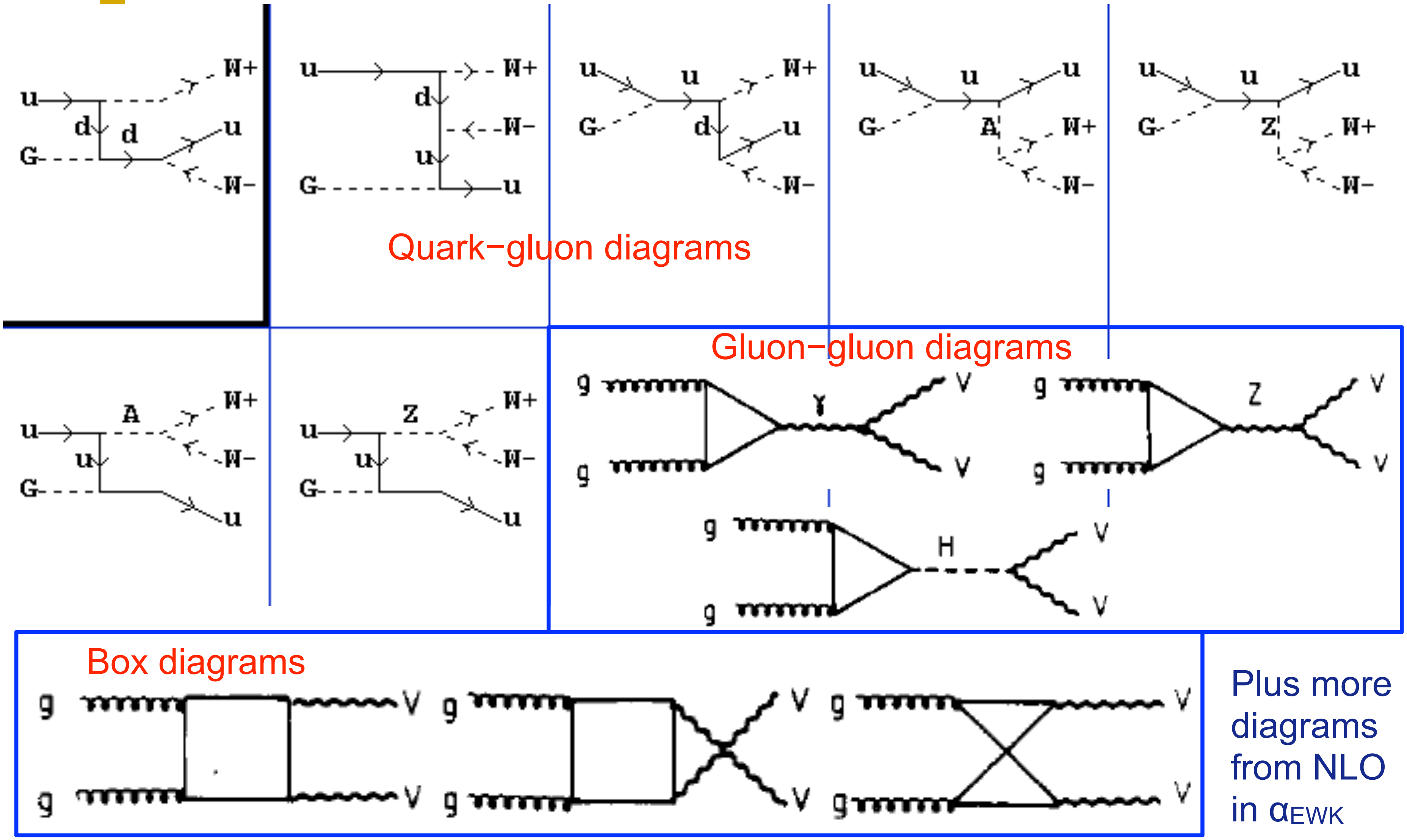
Some quick Observations

1. In standard model, the s- and t-channel diagrams are both divergent but when combined together divergencies cancel out miraculously. Sensitive to gauge coupling.

2. Because of qqbar initial state the production rate at LHC is only ~ 3x Tevatron.

3. One W or Z decays hadronically. Although W and Z boson masses differ by 10 GeV the dijet mass resolution is ~ 10 GeV ➜ cannot distinguish between WW and WZ.

Diboson production mechanism at NLO

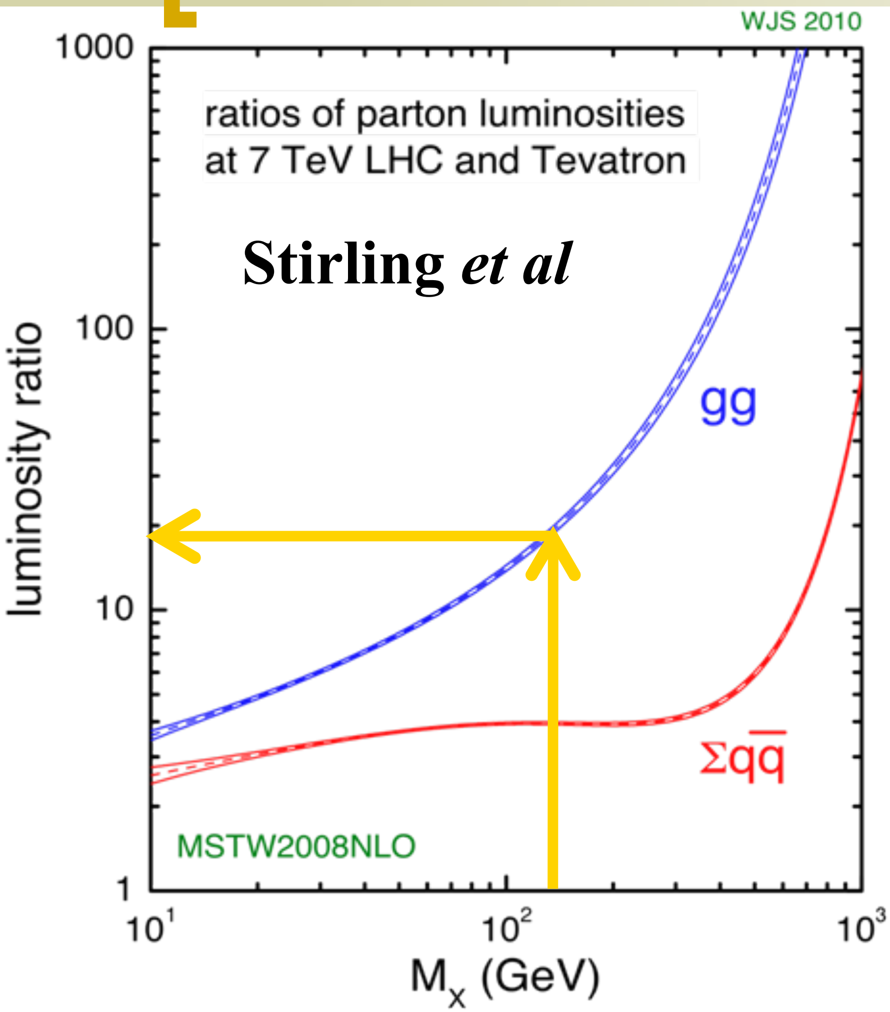

Quark−gluon diagrams

Gluon−gluon diagrams

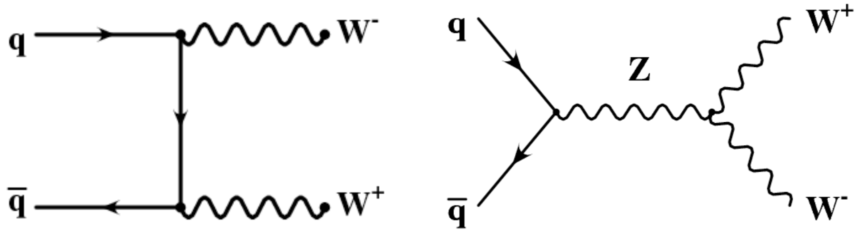
Box diagrams

Plus more diagrams from NLO in α_{EWK}

LHC vs Tevatron: partonic luminosity hurts



$q\bar{q}$ → WW, WZ cross section at 7 TeV is ~ 3 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

⇒ 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.

Di-boson analysis steps

step 0: Use baseline W selection using recipe from top PAG

step 1: Use pfJet and pfMET as recommended by JetMET POG
- charge hadron subtraction: get rid of jets generated from PU
- fastJet area subtraction: remove contribution to jet energy from PU
- default “relative”, “absolute”, and “residual” jet corrections

step 2: Establish di-boson production using appropriate selection criteria

step 3: Converge on reasonable selection to enhance S/B if needed
step 4: Compute efficiency and acceptance for the selection used
step 5: Perform signal extraction, compute cross section

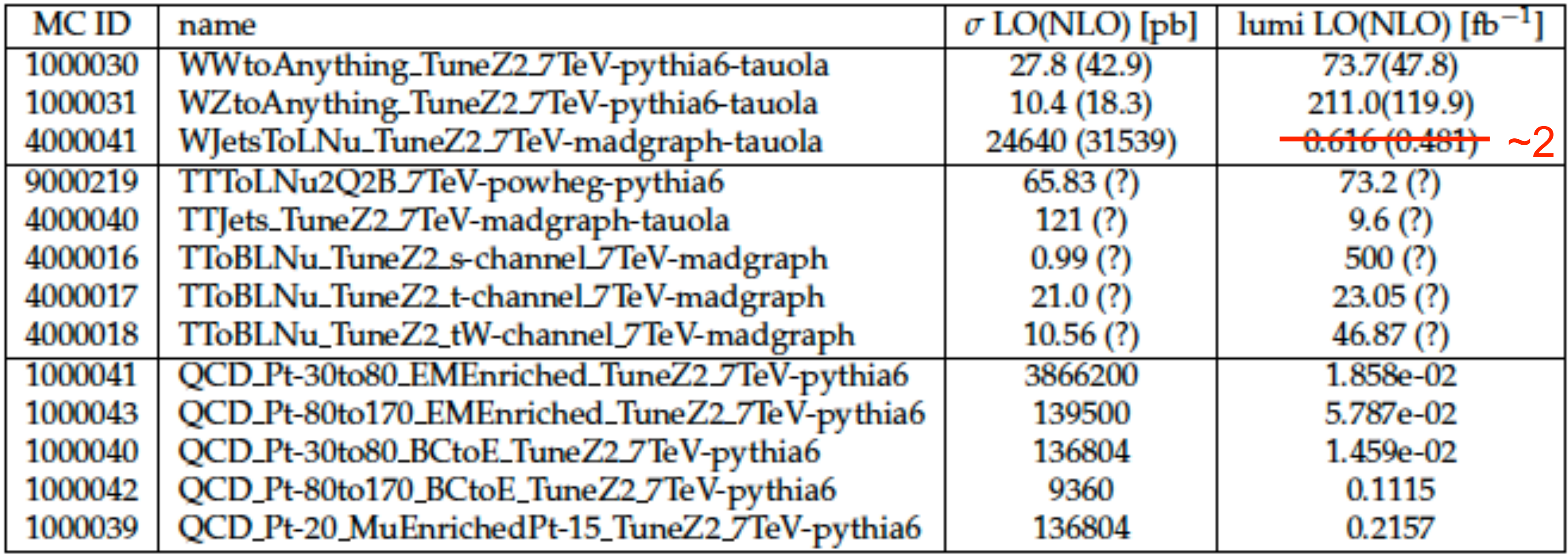
step 6: Set limit on anomalous gauge couplings

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

Are we triggering on the events we need ?

Yes, but life is a little complicated ....

✦For 2010 data (36 pb^{-1}) use single lepton triggers with p_T > 17 GeV (or lower)

✦For 2011A data before technical stop (~200 pb^{-1}) still rely on single lepton triggers: Mu_24 (non isolated) and Ele_27 (CaloId, TightIso)

-So, had to go to offline cuts: mu p_T > 25 GeV, electron E_T > 30 GeV

✦For 2011 data after technical stop there is dedicated "Ele17_CentralJet30_CentralJet25_MHT15" trigger. For muon still rely on Mu_24

-go to following offline cuts to be minimally tighter than trigger:

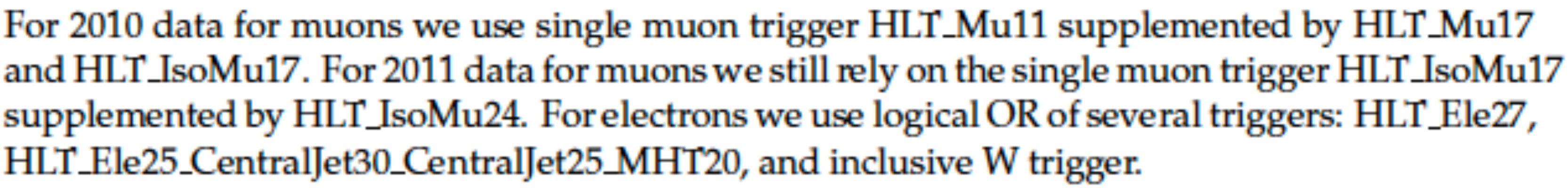
pfMET>25 GeV, W transverse mass > 40 GeV
leading pf jet pt > 30 GeV, second jet pt > 30 GeV.

✦ Now also have an "inclusive" W trigger in the menu for electron: keeps electron E_T > 25 GeV, pf MET > 25 GeV, W m_T > 40 GeV.

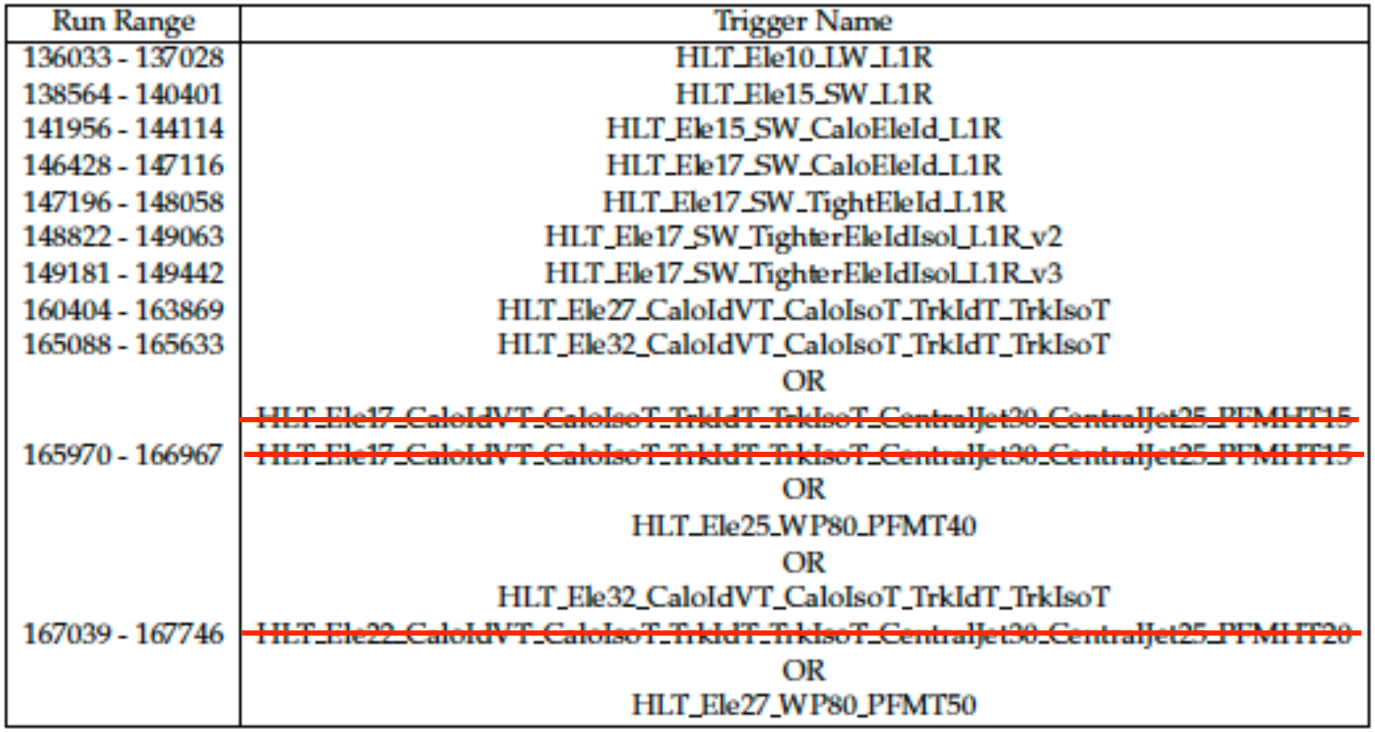
✦Submitted proposal for TSG's consideration for Mu+jj trigger

For electron use: SingleEle || Ele+MHT+jj || W_inclusive.
For muons: IsoMu_24 || Mu17

Summary of HLT paths used for this analysis



For 2010 data for muons we use single muon trigger HLT_Mu11 supplemented by HLT_Mu17 and HLT_IsoMu17. For 2011 data for muons we still rely on the single muon trigger HLT_IsoMu17 supplemented by HLT_IsoMu24. For electrons we use logical OR of several triggers: HLT_Ele27, HLT_Ele25_CentralJet30_CentralJet25_MHT20, and inclusive W trigger.



Run Range	Trigger Name
136033 - 137028	HLT_Ele10_LW_L1R
138564 - 140401	HLT_Ele15_SW_L1R
141956 - 144114	HLT_Ele15_SW_CaloEleId_L1R
146428 - 147116	HLT_Ele17_SW_CaloEleId_L1R
147196 - 148058	HLT_Ele17_SW_TightEleId_L1R
148822 - 149063	HLT_Ele17_SW_TighterEleIdIsol_L1R_v2
149181 - 149442	HLT_Ele17_SW_TighterEleIdIsol_L1R_v3
160404 - 163869	HLT_Ele27_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT
165088 - 165633	HLT_Ele32_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT
165970 - 166967	OR ~~HLT_Ele17_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT_CentralJet30_CentralJet25_PFMHT15~~ ~~HLT_Ele17_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT_CentralJet30_CentralJet25_PFMHT15~~ OR HLT_Ele25_WP80_PFMT40
167039 - 167746	OR HLT_Ele32_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT ~~HLT_Ele22_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT_CentralJet30_CentralJet25_PFMHT20~~ OR HLT_Ele27_WP80_PFMT50

For the analysis shown in this presentation we didn't use the data from ele+jj trigger. We plan to add this data soon.

Acceptance thresholds

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

Systematic uncertainties are being evaluated

Computed as reco/gen ratio. Affected by escale and resolution.

WW muon:

Muon Acceptance = 0.5903
Jets Acceptance = 0.4158
Mu+Jets Acceptance = 0.3738

WW electron:

Electron Acceptance = 0.5503
Jets Acceptance = 0.5825
Electron+Jets Acceptance = 0.4882

WZ muon:

Muon Acceptance = 0.6133
Jets Acceptance = 0.3855
Mu+Jets Acceptance = 0.3501

WZ electron:

Electron Acceptance = 0.5717
Jets Acceptance = 0.5849
Electron+Jets Acceptance = 0.5061

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
- -$\chi2/\mathrm{ndf} < 10$ 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

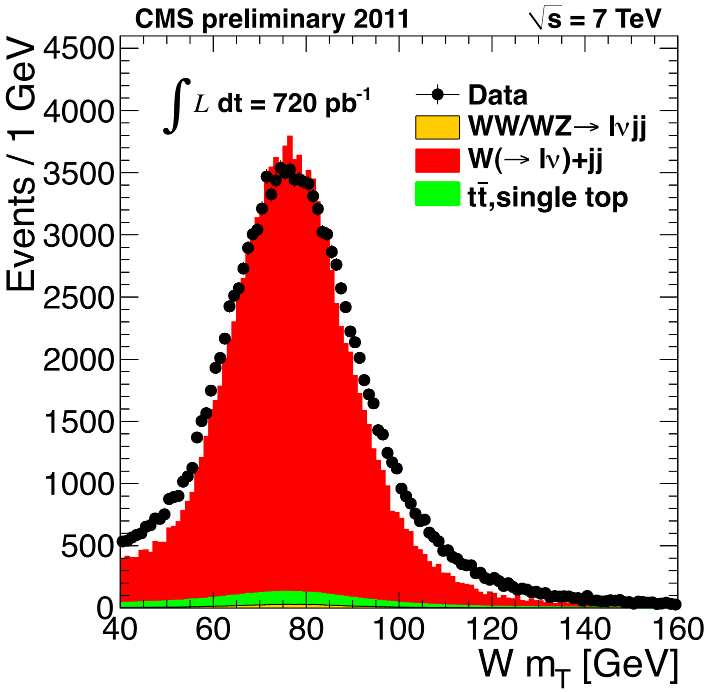
http://twiki.cern.ch/twiki/bin/view/CMS/SimpleCutBasedEleID

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

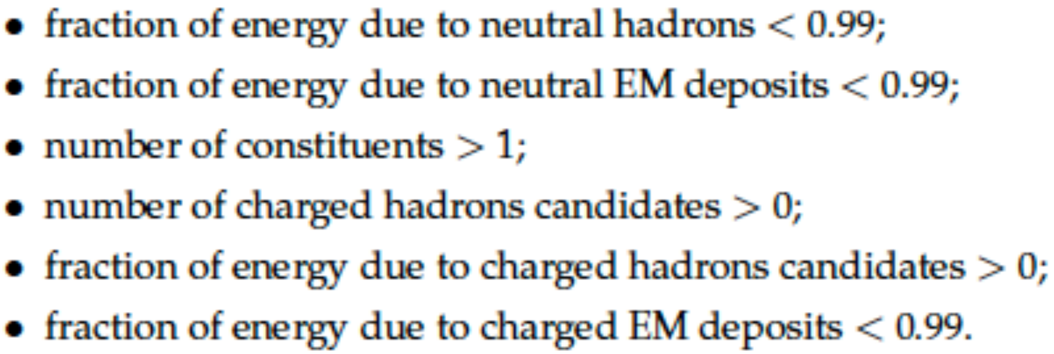
Selection Efficiency: Efficiency for the above electron selection is 64.3% in MC. A detailed study using Tag&Probe is underway to compute data/MC scale factors.

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%.

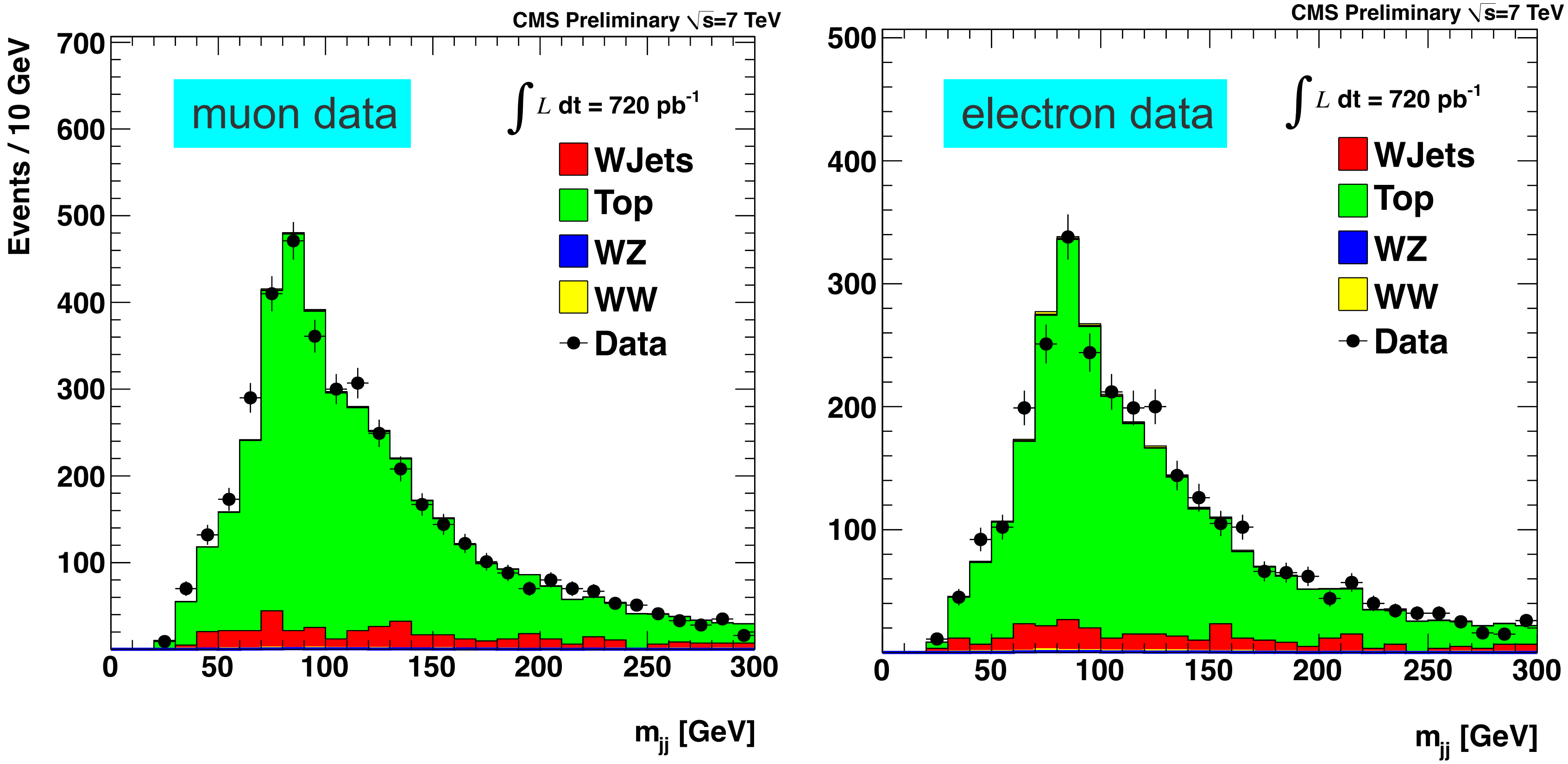
Remove identified leptons reconstructed as jet from the jet collection by requiring ΔR (lepton, jet) > 0.3. In addition, require the jets to be not b-tagged. For this, use “simple secondary vertex high efficiency” loose operating point (tagging efficiency > 70%, mistag rate ~ 1%). This greatly reduces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

MET

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

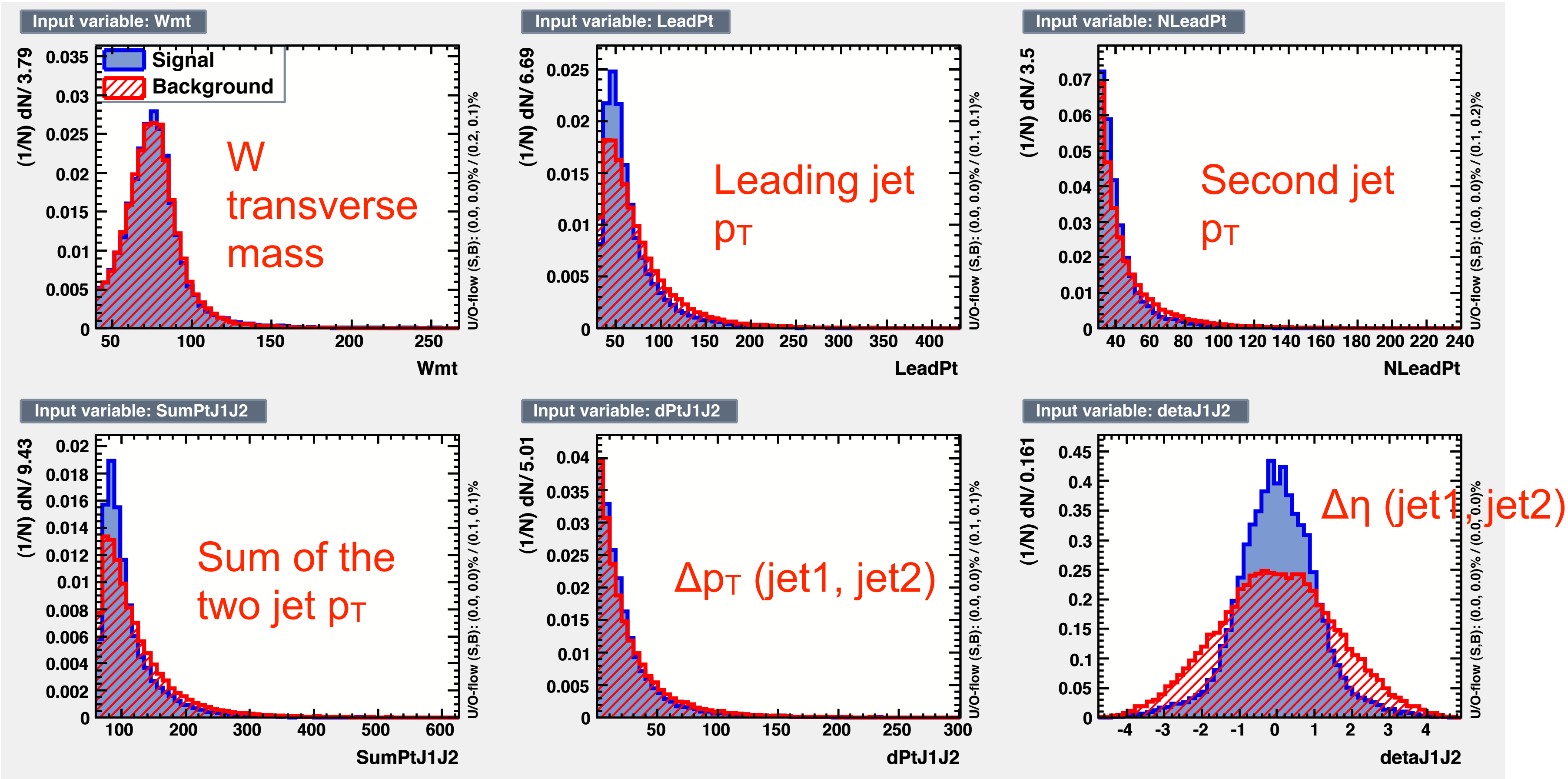
Can we reconstruct hadronic W in CMS ? Yes

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



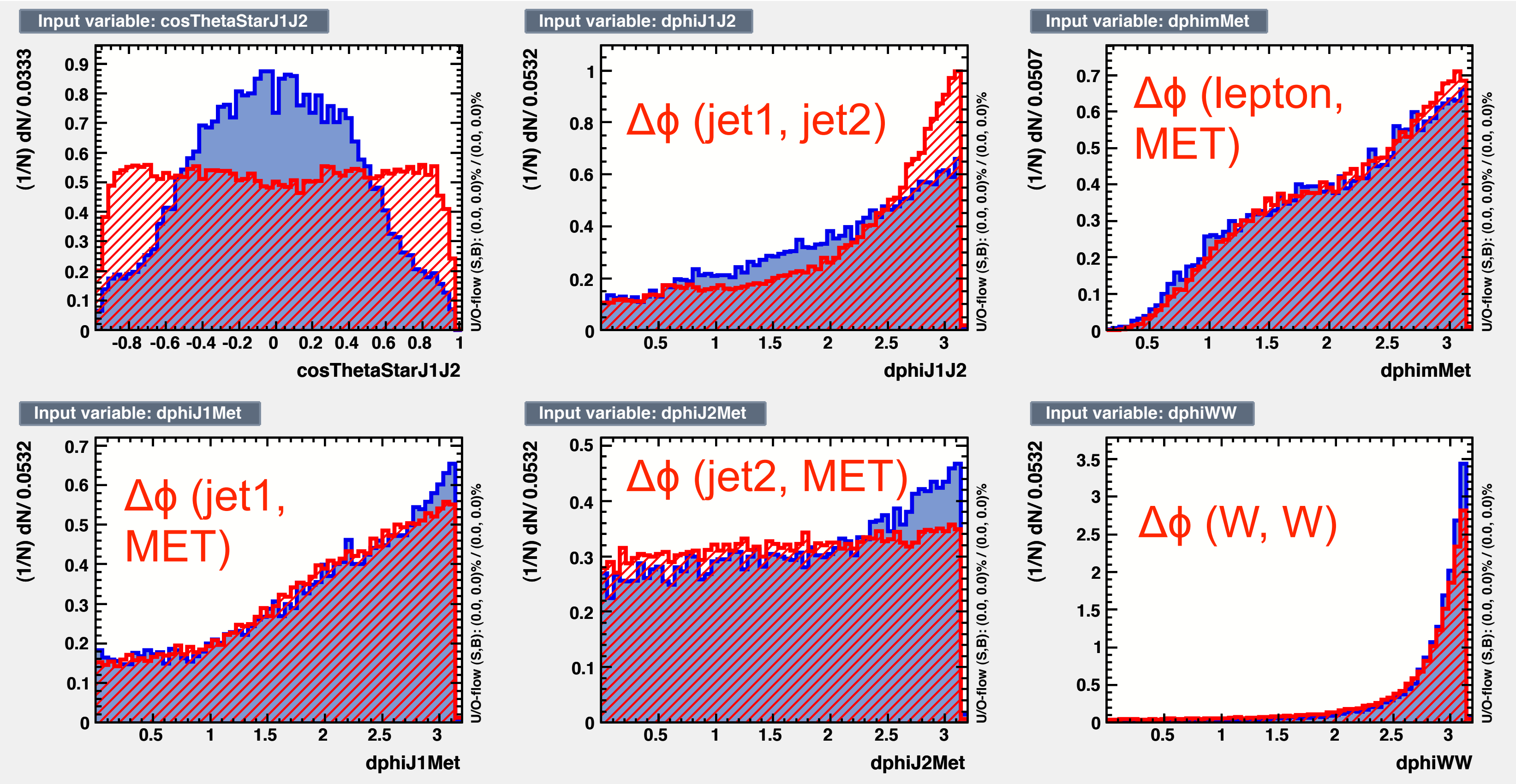
Just require 4 jets above p_T 30 GeV, 2 loose b-tags, and a leptonic W (muon: p_T>25 GeV or electron: E_T>30 GeV, MET>25 GeV). Then plot m_{jj} of the two jets which are not b-tagged.

Comparison of kinematic variables (I)



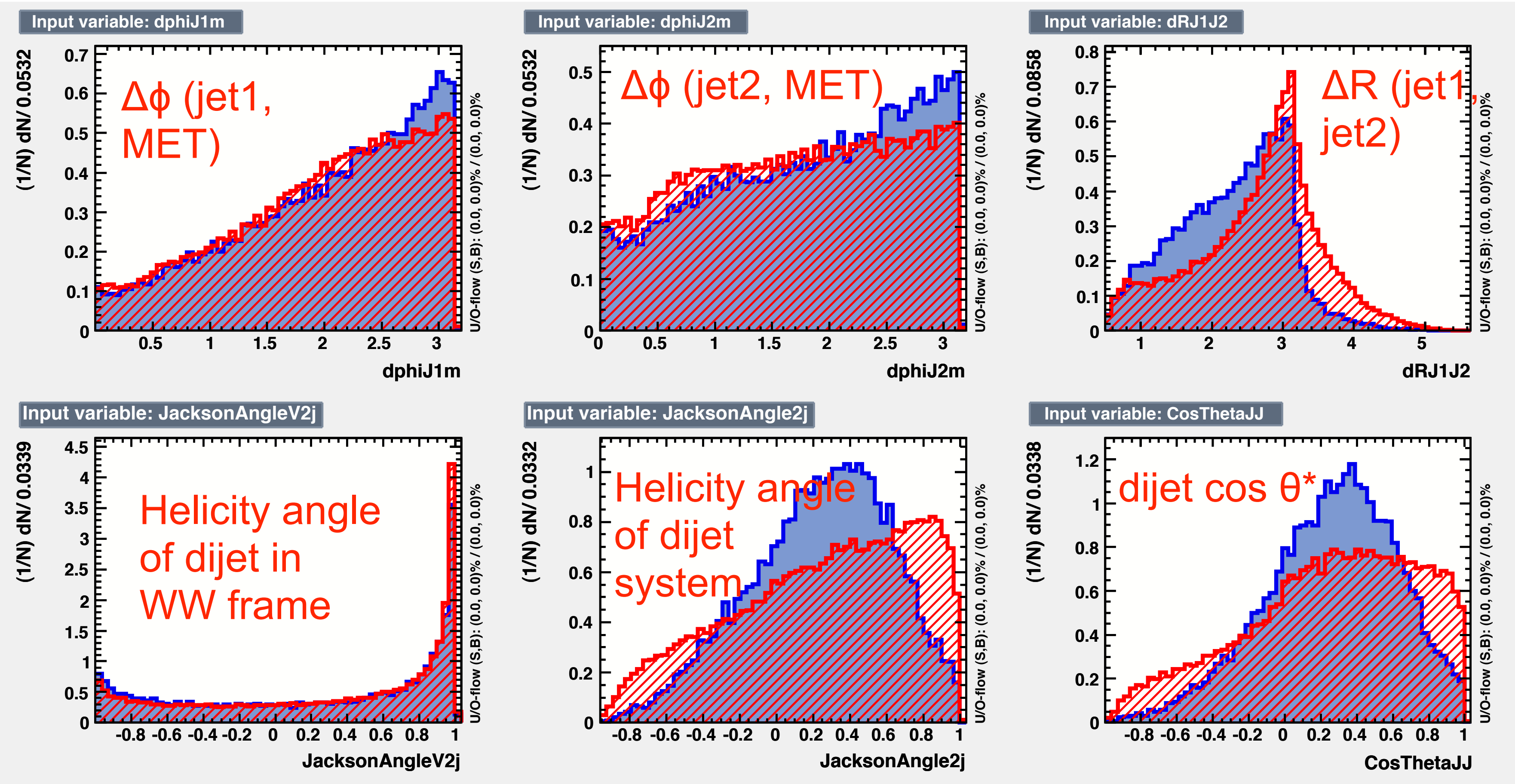
signal = WW, background = W+jets

Comparison of kinematic variables (II)



signal = WW, background = W+jets

Comparison of kinematic variables (III)



signal = WW, background = W+jets

Topological cuts

Apply some simple topological cuts to suppress backgrounds and enhance S/B for the di-boson signal. These are listed below.

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
- Cut on helicity angle of the leading jet in the dijet frame
  $-0.6 < \cos\theta^* < 0.8$

The S/B after these selections in the dijet W mass window 65−95 GeV is ~1/12. The efficiency for these cuts is diboson signal 28.66%.

We are currently working on a multi-variate optimization which will improve S/B further for the same efficiency. More details on a later slide.

Signal extraction procedure

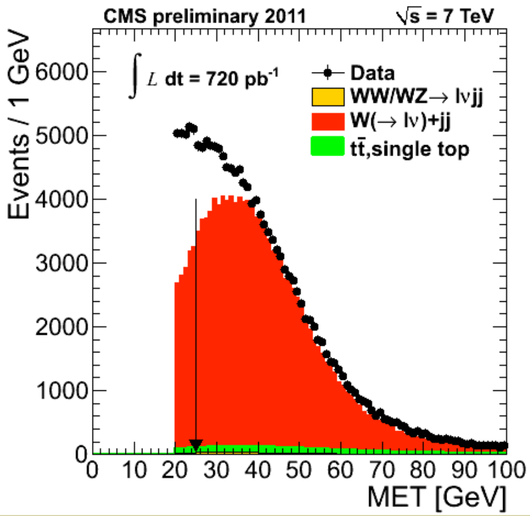
Perform a likelihood fit of the m_{jj} distribution

- in the range 30−135 GeV
- take all shapes from MC
- fit for the absolute number of diboson and W+jets events
- fix the top pair and single top contribution to the measured cross section
- float the effect of jet energy scale (position of the peak) within its uncertainty

⟵ to avoid any mixup with “CDF bump”

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

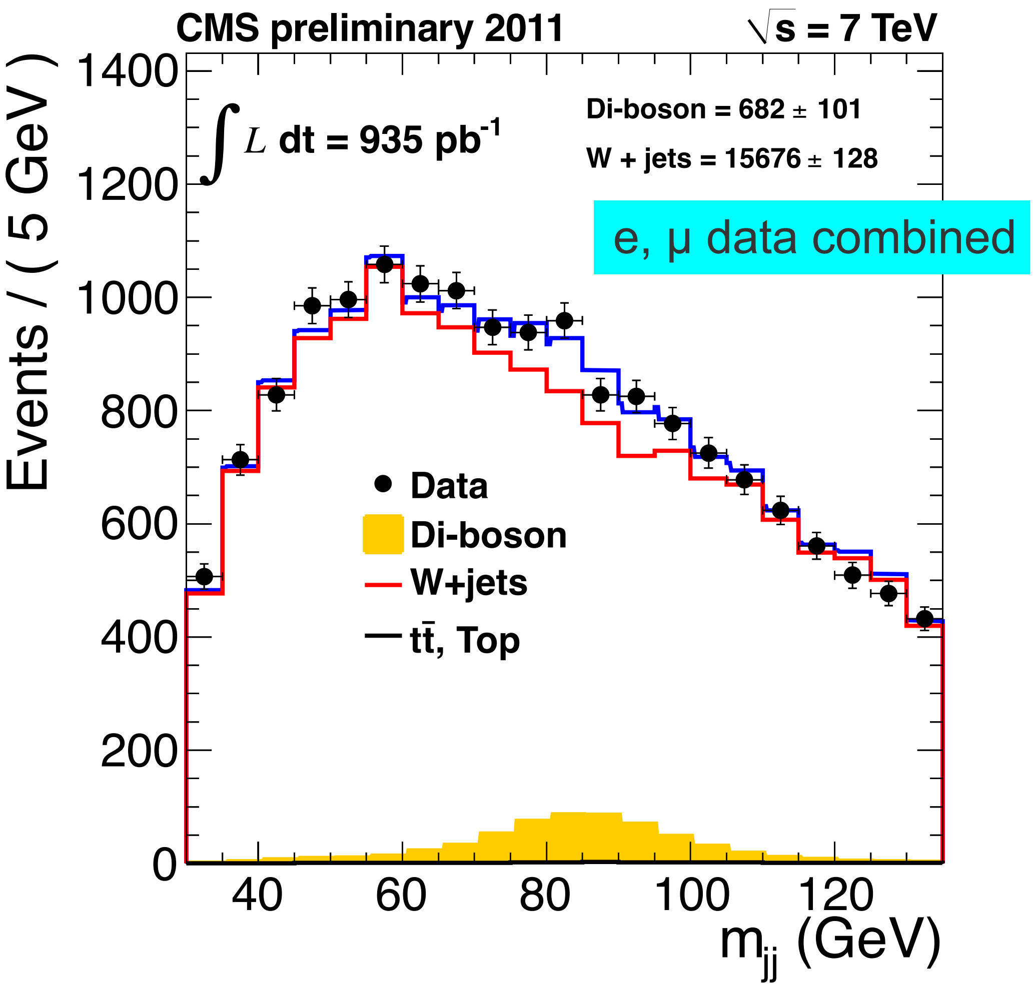
Regarding QCD multi-jet background



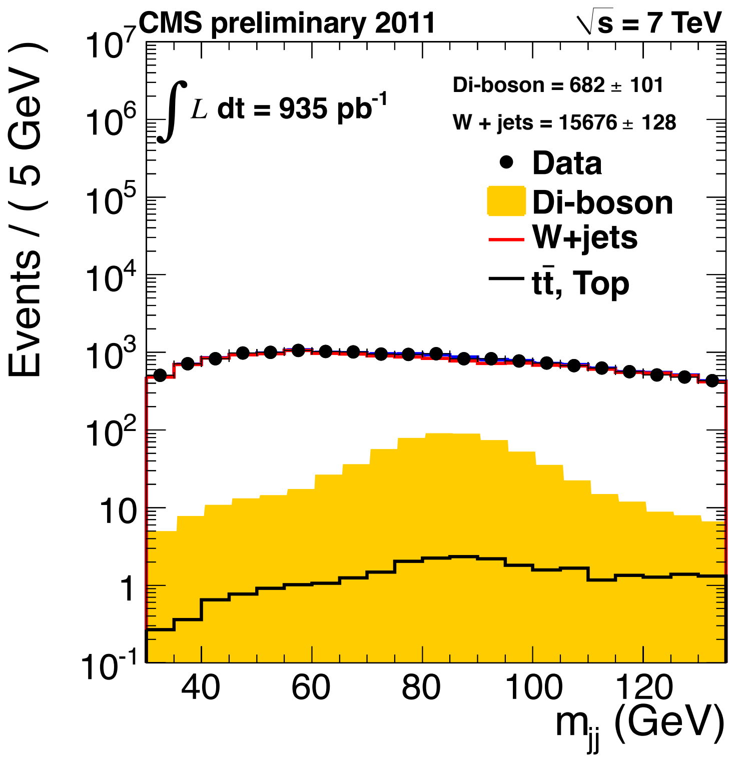
The QCD background is insignificant. We have been running jobs on grid for a week but for various reasons not all jobs have finished. We intend to take QCD shape from MC and obtain its normalization from data by fitting the full spectrum of the MET distribution.

← MET distribution in electron data before topological cuts

Signal extraction (1)



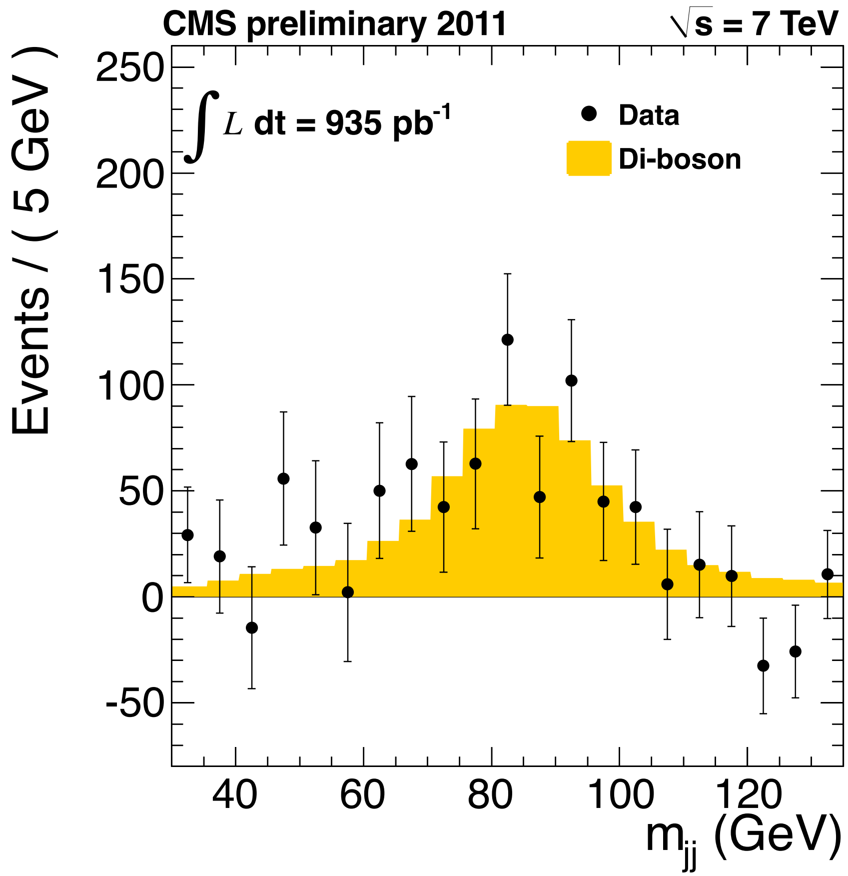
e, μ data combined



Fit results: on the next slide

Will include insignificant backgrounds due to QCD, Z(→ll)+jets, Z→ττ (l+had) once CRAB stabilizes.

Signal extraction (2)



Fit result

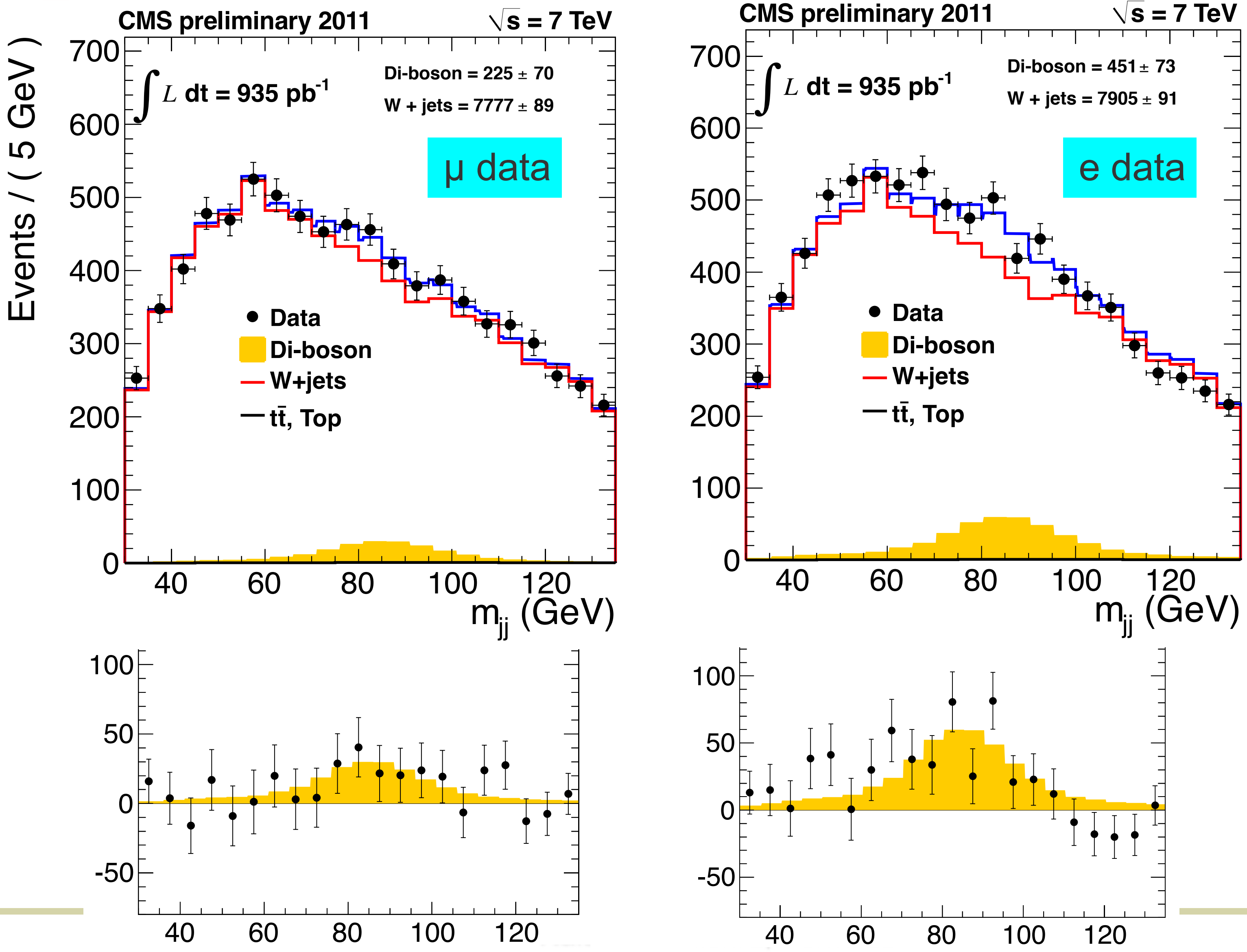
massShift	-0.54 ± 0.85
nWjets	15676 ± 128
nDiboson	683 ± 101

Correlation matrix

NO.	GLOBAL	1	2	3
1	0.20753	1.000	0.166	-0.207
2	0.77949	0.166	1.000	-0.779
3	0.78350	-0.207	-0.779	1.000

S/B in the W mass window 65−95 GeV = 422/5053 ≈ 1/12

Signal extraction (3)



Cross section measurement

$$\sigma \cdot Br = \frac{N_{candidates} - N_{background}}{Acceptance \cdot Efficiency \cdot L}$$

From MC

Uncertainty is dominated by JES and jet resolution

Currently take from MC

= lepton selection efficiency x trigger efficiency x event selection cut efficiency

External input

6% uncertainty

Lepton efficiency will come from Tag&Probe measurements in data and MC

$$\epsilon_{X} = \epsilon_{MC\text{-}X} \times \rho_{eff\text{-}X},$$

$$\rho_{eff\text{-}X} = \frac{\epsilon_{TNP\text{-}X}(data)}{\epsilon_{TNP\text{-}X}(MC)}$$

Since we cannot distinguish between WW and WZ we take the weighted acceptance for now: acceptance = $(2 . A_{WW} + A_{WZ}) / 3$

Cross section results (very preliminary)

Muon data:
signal events = 225 ± 70
acceptance = 0.3659
efficiency = 0.701 x 0.2866
luminosity = 935 pb^{-1}
branching ratio = 0.108 * 0.68
cross section x BR = 44.6 ± 13.9 pb (stat)

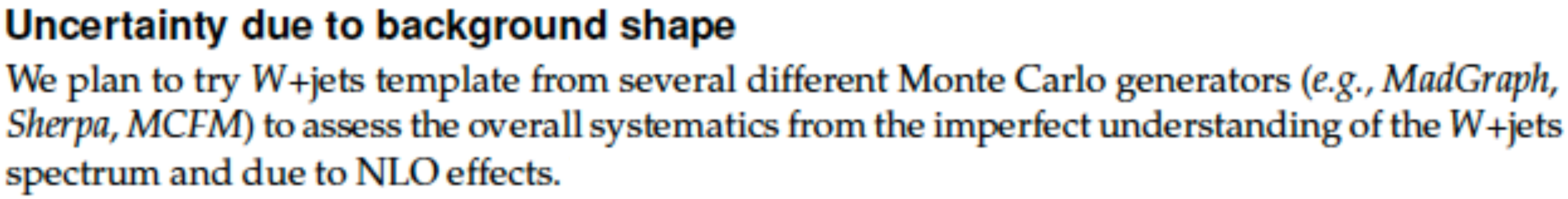
Electron data:
signal events = 451 ± 73
acceptance = 0.4942
efficiency = 0.643 x 0.2866
luminosity = 935 pb^{-1}
branching ratio = 0.108 * 0.68
cross section x BR = 72.0 ± 11.6 pb (stat)

Standard Model electroweak production cross section at NLO:
$\sigma(WW) = 42.9$ pb, $\sigma(WZ) = 18.3$ pb

Many of the above ingredients − acceptance and efficiency − were computed in the last few days. They are being cross-checked. Also, efficiency numbers are purely from MC, these need to be confirmed from data.

Systematic uncertainties

So far, the only uncertainty we have evaluated is due to JES.
Another important source of systematics is in W+jets shape due to variation of normalization scale ($\mu = \sqrt{(m_W^2 + p_T{}^2{}_{jet1})}$) which we are attempting to compute soon. Besides these there are various other sources of systematic uncertainties that we need to compute.



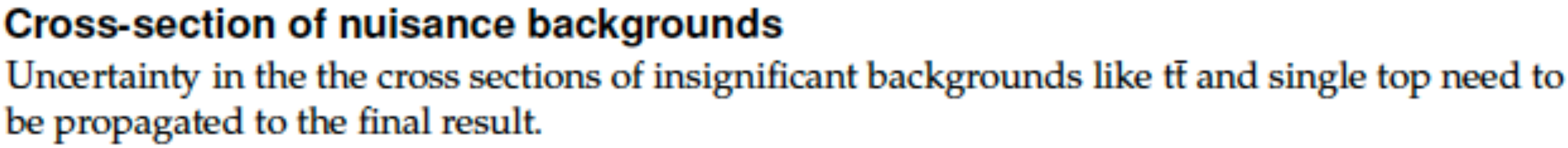
Uncertainty due to background shape

We plan to try W+jets template from several different Monte Carlo generators (*e.g.*, *MadGraph*, *Sherpa*, *MCFM*) to assess the overall systematics from the imperfect understanding of the W+jets spectrum and due to NLO effects.

Lepton energy scale, resolution, selection, and trigger

MET uncertainty

Vary the MET by ±1σ and compute the change in signal yield.



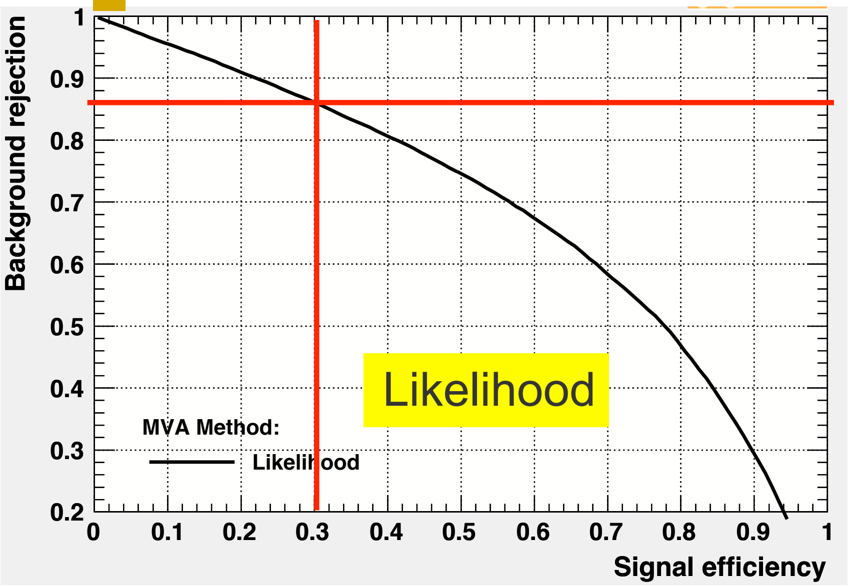
Cross-section of nuisance backgrounds

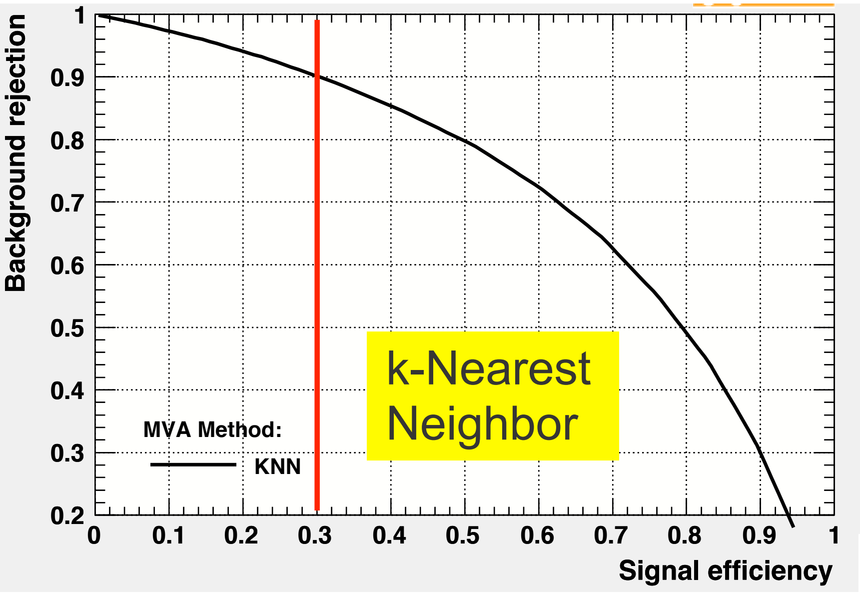
Uncertainty in the the cross sections of insignificant backgrounds like $t\bar{t}$ and single top need to be propagated to the final result.

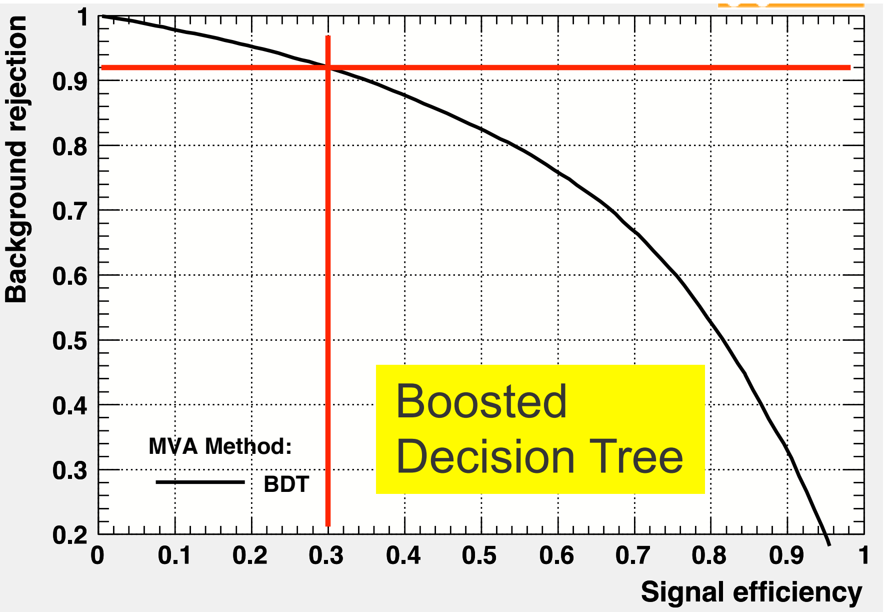
Luminosity uncertainty currently 6%

Likely improvement in event selection

Trying several multi-variate classifier. To begin with the S/B before topological cuts is ~ 1/100. Simple cut gives us S/B ~ 1/12 for 30% efficiency. Initial indication is that with the same signal efficiency we may be able to get S/B ~ 1/10 or better using MVA.







Other ideas being explored

Kinematic fit

- Constrain the leptonic W mass. Then plot the m_{jj} of the dijet system.

- This should improve the resolution and hopefully give better discrimination between diboson and W+jets.

Quark-gluon separator

- The second jet in W+jets is almost always gluon initiated, whereas the second jet in the di-boson is almost always light flavor.

- A classifier can be trained on quark enriched (say, photon+jet or Z+jet events) and gluon enriched (say, QCD dijet events) samples to discriminate between the two.

- The inputs to the classifier would be various energy fractions (charged energy fraction, neutral energy fraction, e/mu/photon/had fractions) and/or multiplicities, jet area, eta- and phi- moments etc.

Summary

✦Performed the first study of the SM EWK WW,WZ→lvjj production in CMS using ~1 fb^{-1} data

✦Di-boson signal in W(lv)+jj channel is established
- basic selection in place, further minor optimization possible
- signal extraction technique / template fit is working

✦Most ingredients in place to compute cross section
- preliminary estimation agrees with the SM prediction

✦Working to estimate systematic uncertainties
- need to crosscheck all the numbers
- JES systematics included in the likelihood
- will work through the rest

BACKUP SLIDES