Wjj analysis: news, status, plans

Monika Grothe, Kalanand Mishra

On behalf of the Wjj working group

June 7, 2011

Likely timeline for preliminary result for EPS

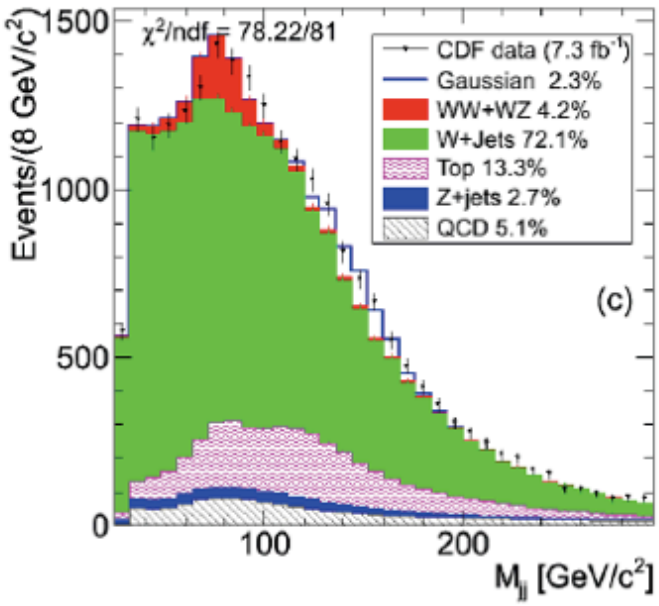
This is our current understanding of the deadlines for EPS approval

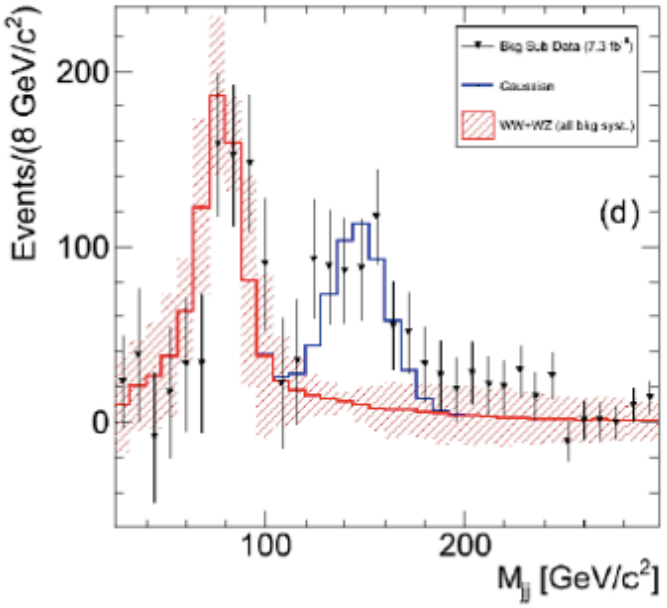
- For presentation at EPS, a week of approvals will be organized by the EWK PAG around mid-July.
- This requires a pre-approval of the analysis in the last week of June, around June 27th.
- The analysis documentation must be frozen one week before pre-approval, which is around June 20.

This means we have less than 2 weeks to finalize analysis details and complete the documentation (AN and PAS) !!!

Meanwhile CDF announced ~5σ significance

Updated W-jj with 7.3fb^{-1}





- Now closer to 5 sigma
- It was not just a statistical fluctuation
- Serious issue for CDF to understand this.
- Larger sample now allows for more detailed studies - stay tuned for updates.

slides 30−35 of G. Punzi's talk at Blois 2011

Task list twiki, analysis steps, etc

https://twiki.cern.ch/twiki/bin/view/CMS/EWKMjjinVplusJets

CDF Mjj anomaly in CMS data

Minimal plan:

Immediate goal is to produce a statement on the CDF bump for the summer conferences, ideally already for EPS.

- Apply diboson selection for lnujj
- Fit the Mjj spectrum with the V+jets background shape taken from a MC template. (Should eventually be taken from data, this is an obvious point to improve upon, but that needs more time).
- Extract W+Jets Mjj templates from Madgraph/Alpgen/MCFM/Sherpa MC generators and compare effect in fit to data.
- If using both lepton flavors proves to require too much time, concentrate on muon channel

Urgent list of tasks for minimal plan:

- As a very first step, to get an idea how sensitive the W+jets Mjj spectrum is to the actual generator, compare Mjj on gen level between the various generators.
- For sys checks: Understand from MC authors which MC parameters may impact the Mjj spectrum and what their possible range is. Extract W+jets templates with MC parameters varied within possible range.
- Monitor in steadily increasing data set the effect of the diboson selection
- Monitor in steadily increasing data set the effect of changing trigger conditions

Tasks list

Data & selection used in this presentation

Acceptance

- Tight lepton selection from top PAG
- Exactly two jets with p_T > 25 GeV (using PF2PAT cleaning)
- pf MET > 25 GeV
- W transverse mass > 50 GeV

4 kinematic cuts to suppress W+jets:

With just a single cut the S/B is low enough that the fit runs into instability. With three additional cuts S/B ~1/5.

- $|\Delta\phi(W, W) - \pi| < 0.3$ (our original single cut)
- $\Delta\eta(j1, j2) < 1.8$
- $\Delta\phi(j1, \text{muon}) > 2.1$, $\Delta\phi(j1, \text{electron}) > 1.8$
- $\Delta\phi(j2, \text{muon}) > 2.3$, $\Delta\phi(j2, \text{electron}) > 2.0$

These cuts are not necessarily optimal or final. Alexx Perloff and KM are working on this.

See Alexx's talk

Processed ~ 350 pb^{-1} of data so far (340 pb^{-1} for electron, 360 pb^{-1} for muon). Still use 4.1.X MC. Big concern: MadGraph W+jets MC is only about 0.4 fb^{-1}, observe same statistical jittering in MC as in data. Hard to get good template.

We take m_{jj} and m_{lvjj} shape from MC

Problem

We do not have large enough W+jets MC sample to make a good template. The MadGraph sample corresponds to 700 pb^{-1} which is only ~ 2 times larger than our data size. Once we process full 0.6 fb^{-1}, the MC and data will have about the same statistics. This creates large statistical jitter if one takes shape from a simple uniformly-binned histogram of MC events.

Current solution

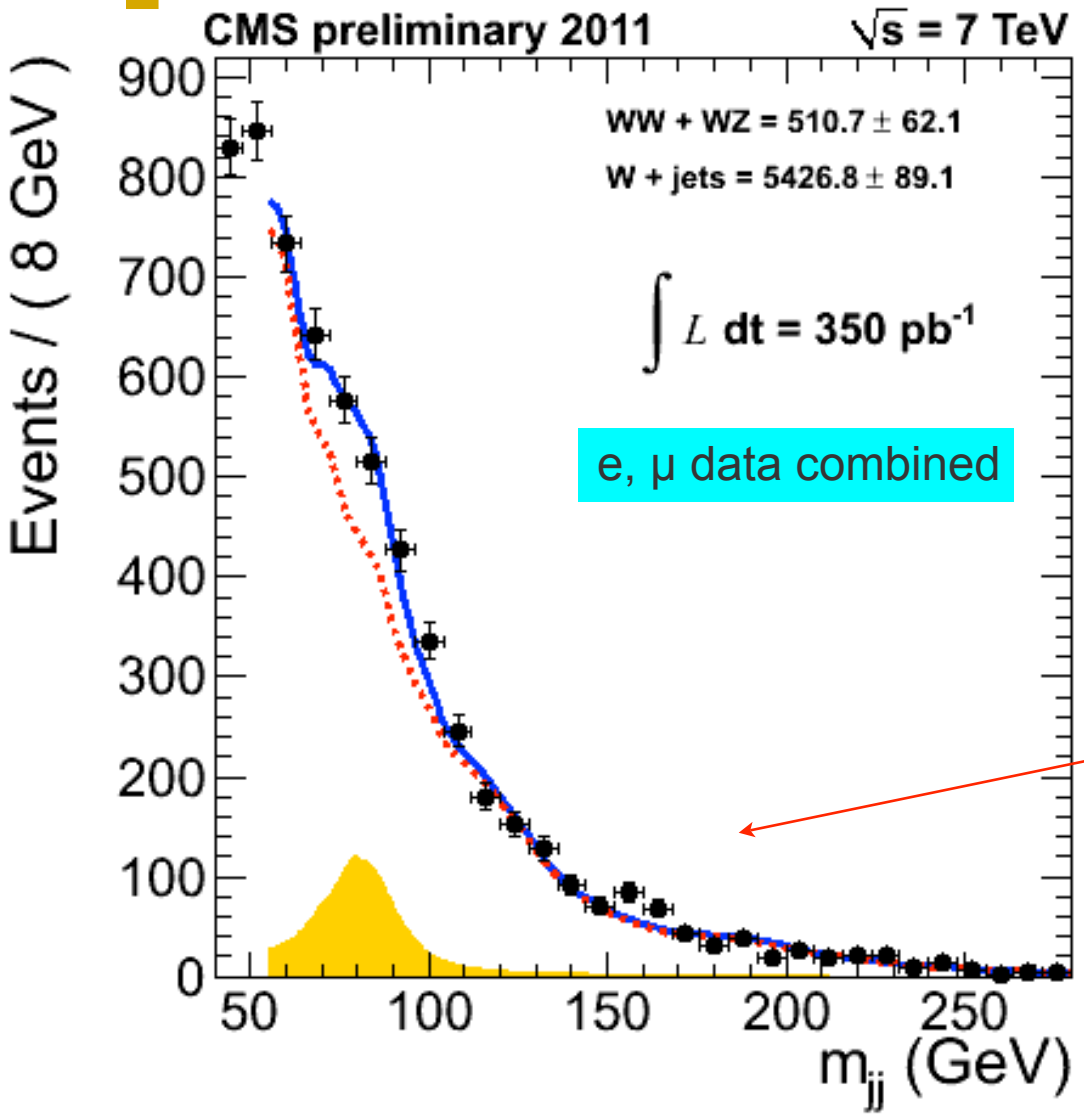
Instead of using fixed bin histograms to derive templates, I use a ROOT functionality called 'RooKeysPdf'. This class is useful if one has to deal with histograms with poor statistics and the trade-offs between having too large bins and having spikes in the plots. It's a class that behaves like a histogram, but internally saves the un-binned events and finally produces a smooth histogram.

Documentation of RooKeysPdf: http://root.cern.ch/root/htmldoc/RooKeysPdf.html

CMS Higgs combination group also uses this class for templates

see for example: HiggsAnalysis/CombinedLimit/interface/TH1Keys.h

Template fit to m_{jj} in W+2 jet events



MC predicts ~ 400 WW + WZ events

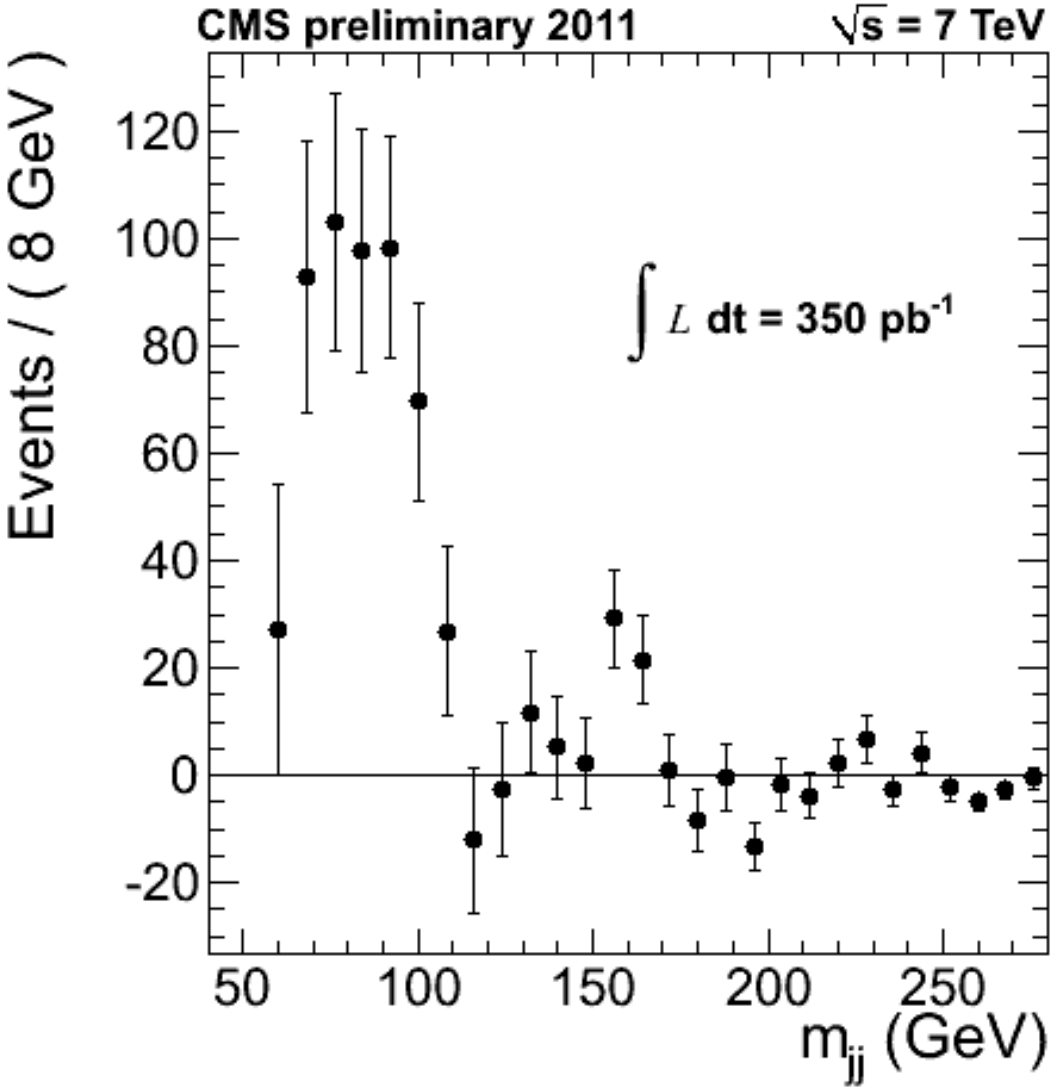
Take the shape from MC. Just fit for the normalization. Blue curve shows the fit to data, dashed red curve is W+jets component, shaded area in orange is di-boson component.

The fit is completely unbinned. Only the plots have binning.

Haven't included single top events in the fit which peak around 150 GeV. Working on it.

Background subtracted distribution on the next slide

m_{jj} in W+2 jet events after bkg subtraction

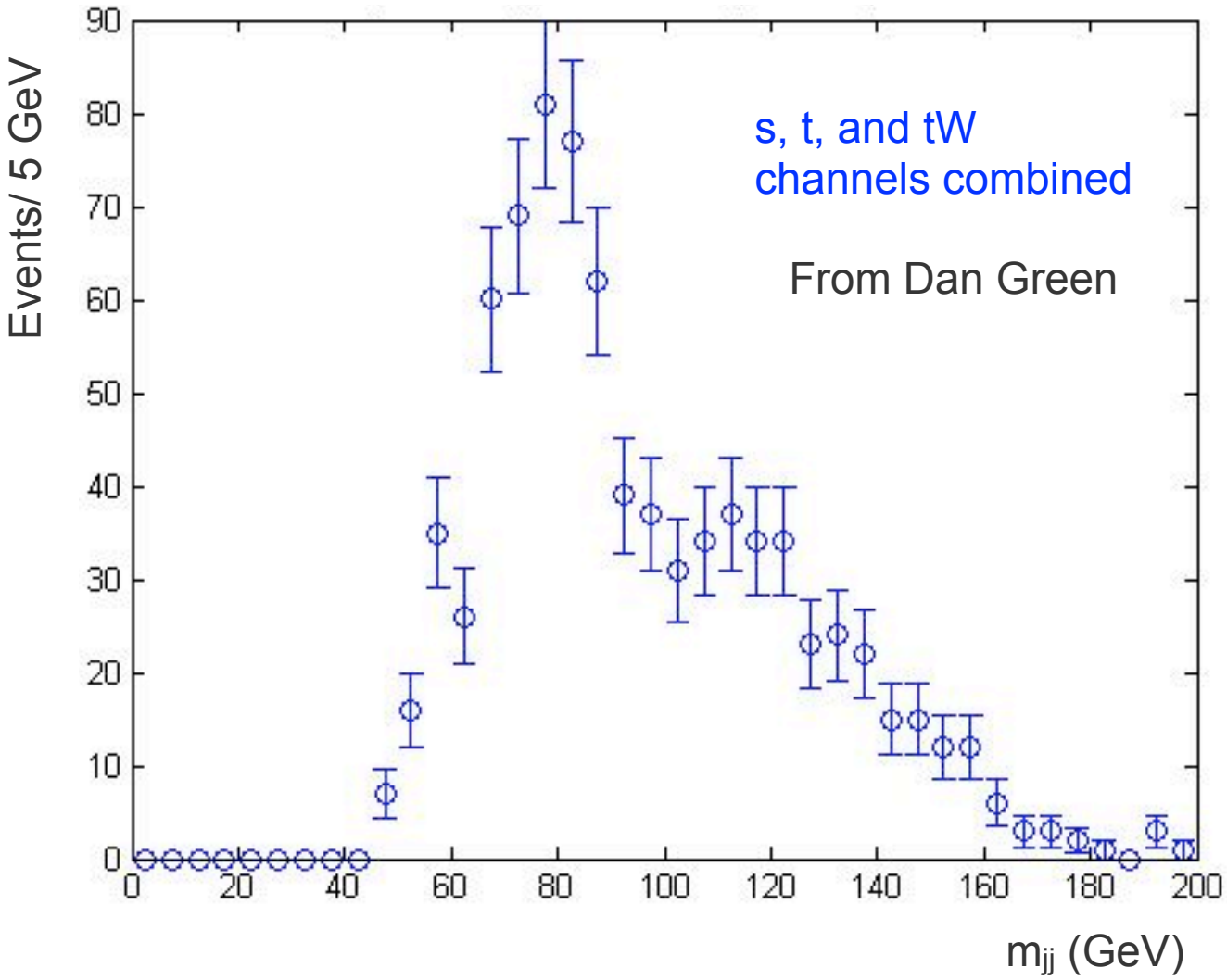


In the W mass window 65 < m_{jj} < 95 GeV we get:

338 di-boson events
1702 W+jets events

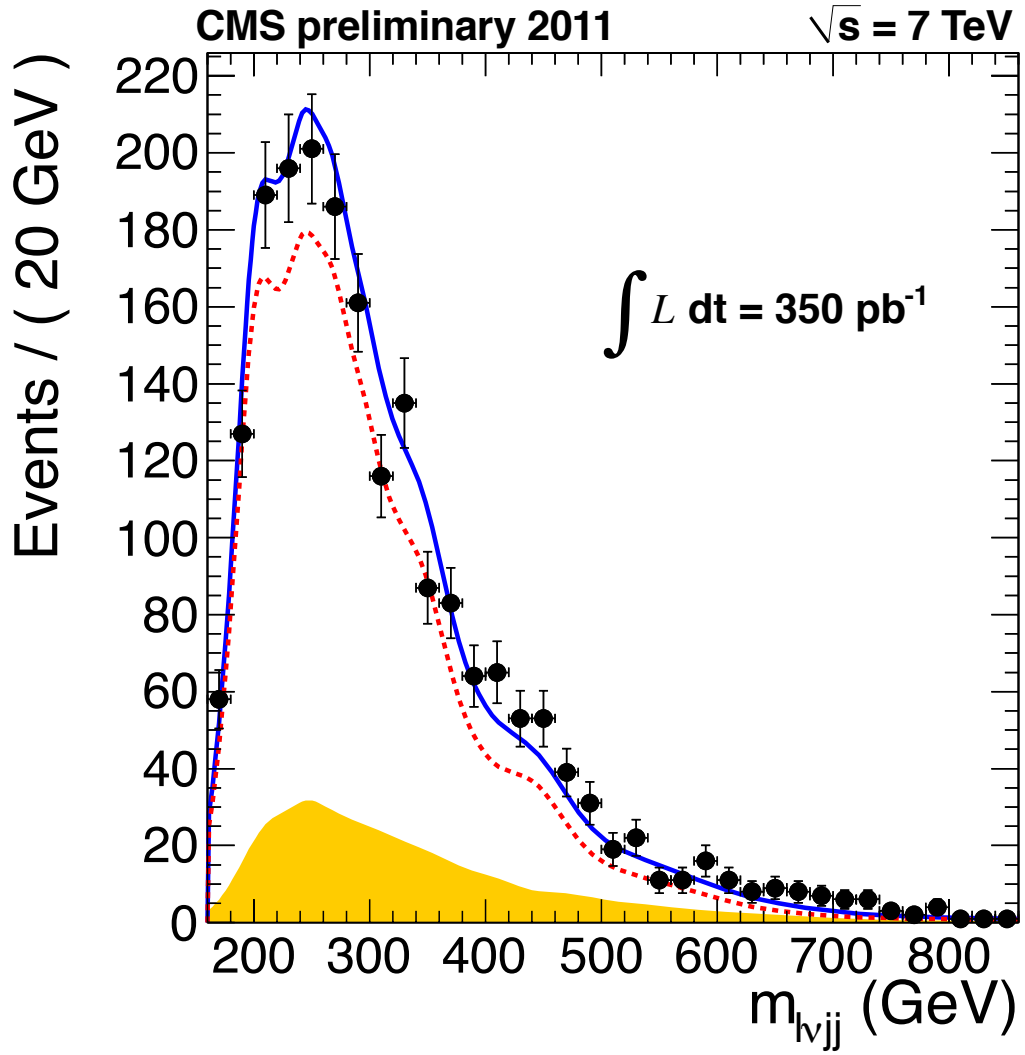
We will fix the normalization in the m_{lvjj} fit to above yields. On the next slide.

Single top has a broad peak in 100–180 GeV

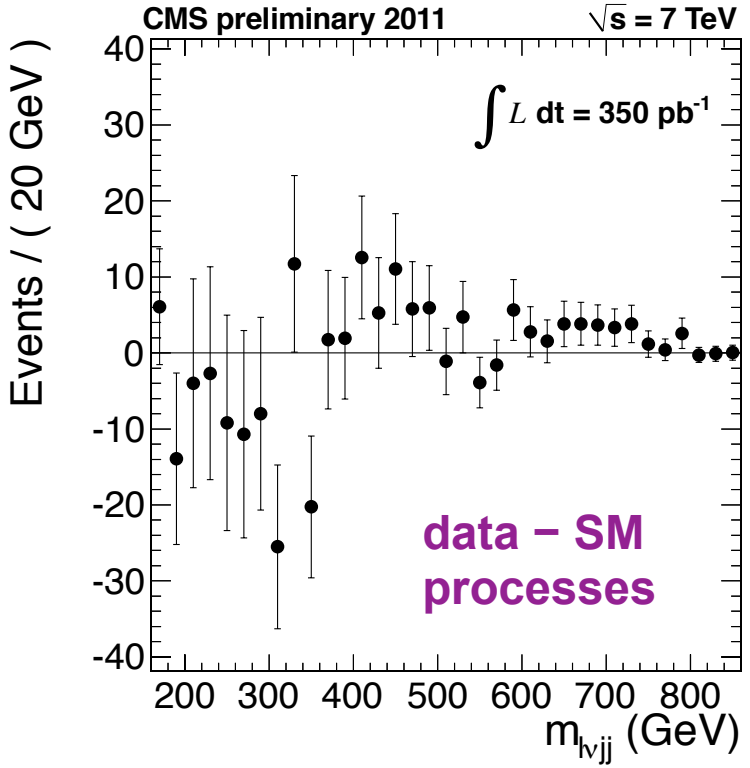


We are working to include this background in our template fit

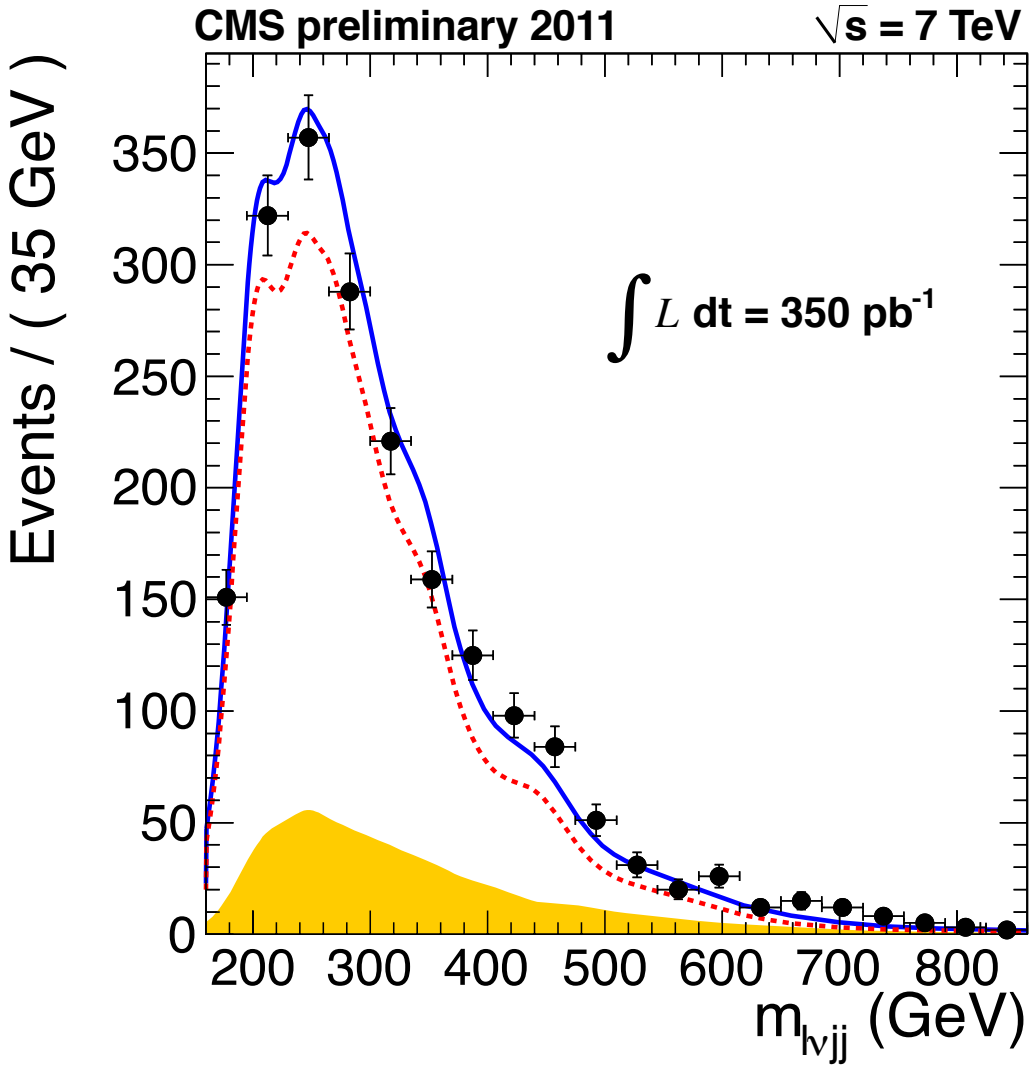
$m_{l\nu jj}$ distribution: $65 < m_{jj} < 95$ GeV



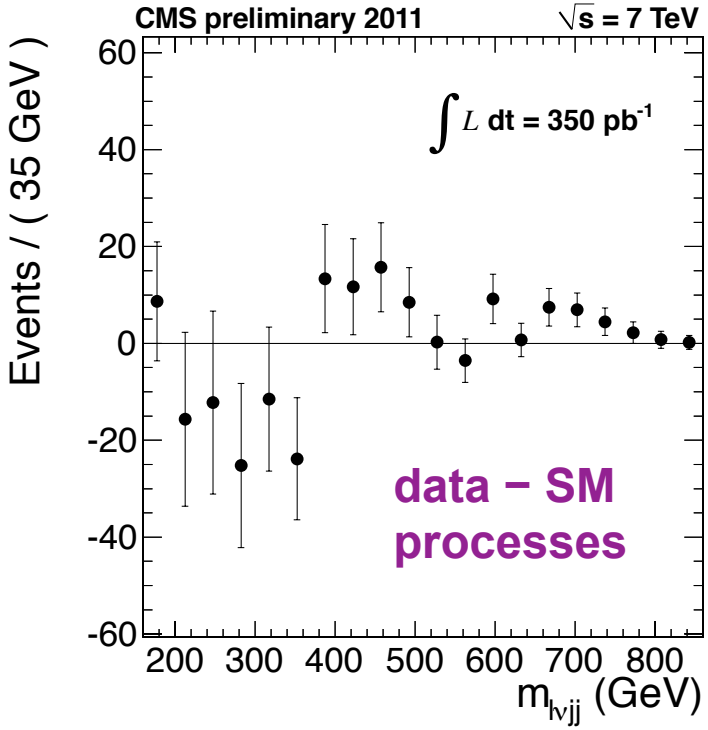
Take the shape directly from MC without any correction.
Additionally, fix the normalization from fit to m_{jj} distribution in data.



$m_{l\nu jj}$ distribution: $65 < m_{jj} < 95$ GeV (larger bins)



Working on kinematic fit to improve the resolution in 4-body mass



Next steps

1. Converge on the m_{jj} fit: try W+jets shape from data, functional forms motivated by MC but fit on data.

2. Compute efficiency and acceptance

3. Try some alternative physics models which would produce bump in m_{jj} spectrum. Needed to compute sensitivity or limit for "CDF bump".

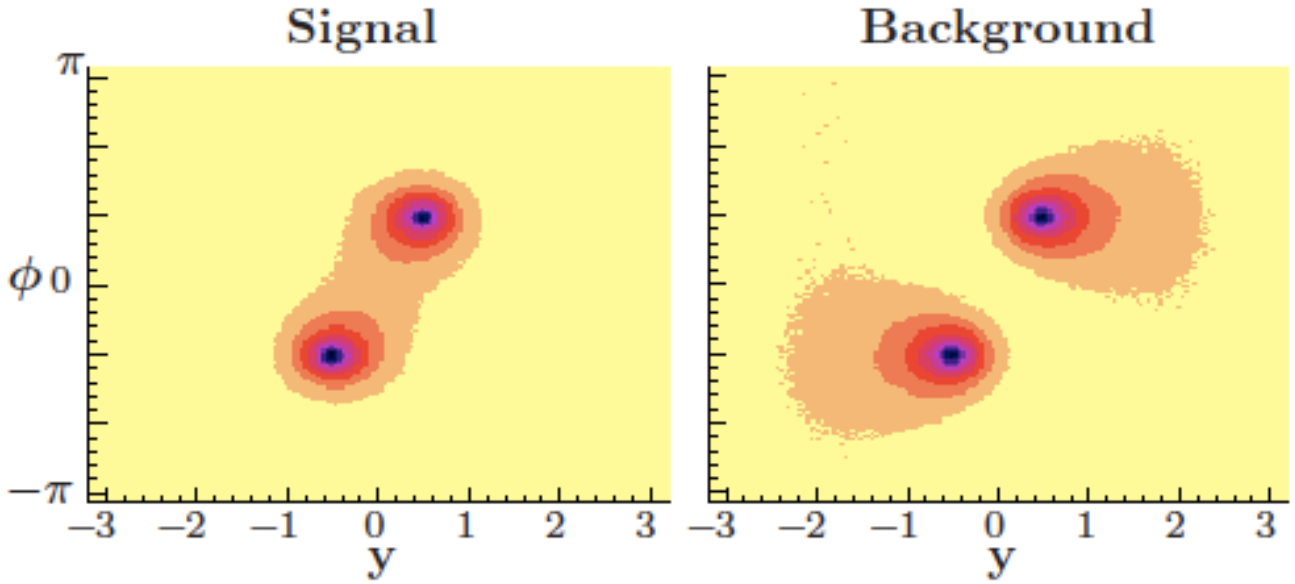
4. Include systematics in the likelihood
   - JES/JER are easy to include
   - For uncertainty in template due to NLO effect need NLO MC
   - Similarly, need MC with Q^2 up/down variation
   - Include single top, QCD multi-jet, top etc. contributions

Besides, need to have AN and PAS (even with place holders) written soon.

Some future improvements

- Use information about color correlation between between the two jets

Should give us another 20−30% more discrimination.



arXiv:1001.5027

color pull:

$$\vec{t} = \sum_{i \in \text{jet}} \frac{p_T^i \, |r_i|}{p_T^{\text{jet}}} \vec{r}_i \, .$$

backup slides