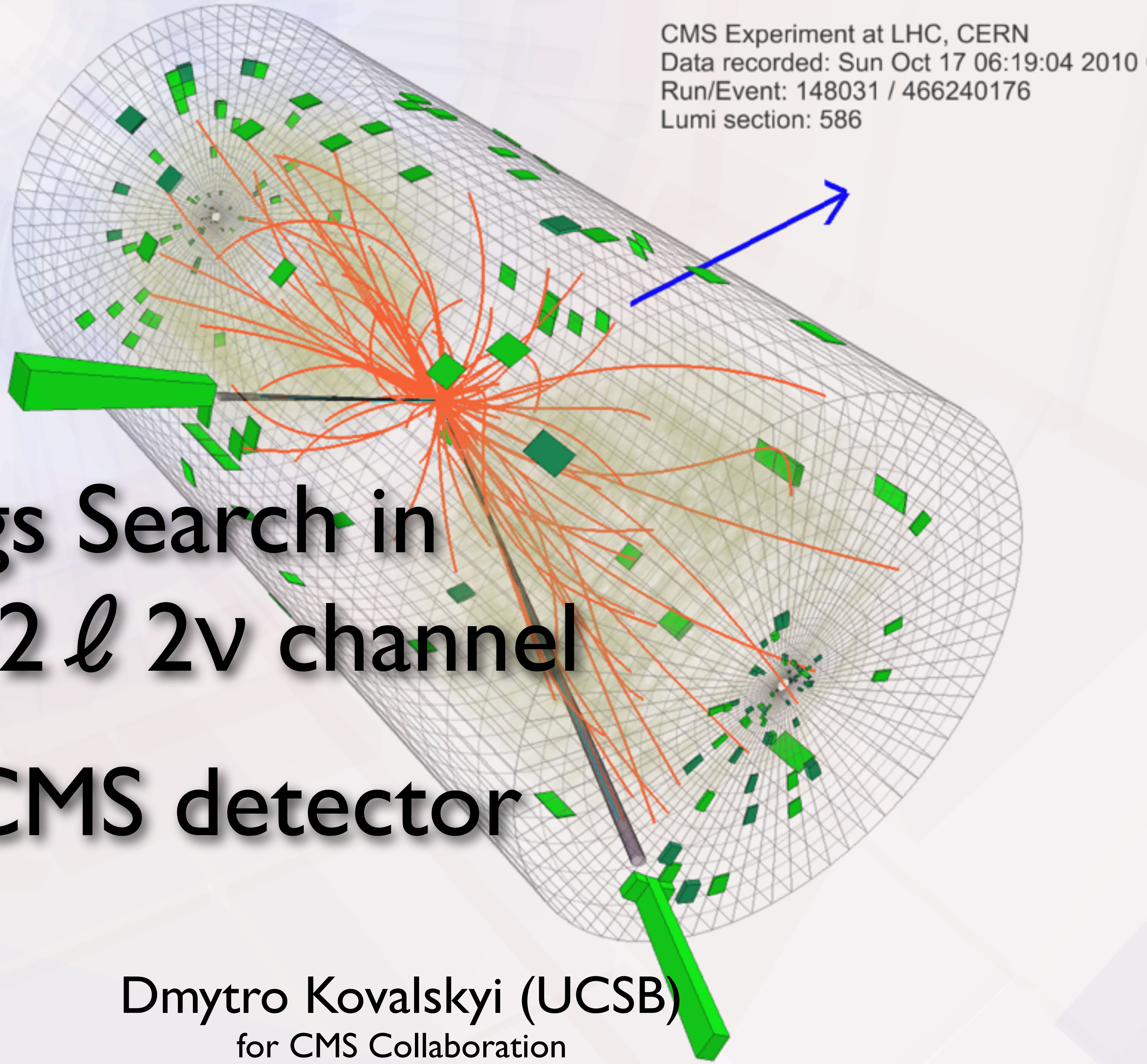


Higgs Search in WW→2 ℓ 2ν channel with CMS detector

Dmytro Kovalskyi (UCSB)
for CMS Collaboration

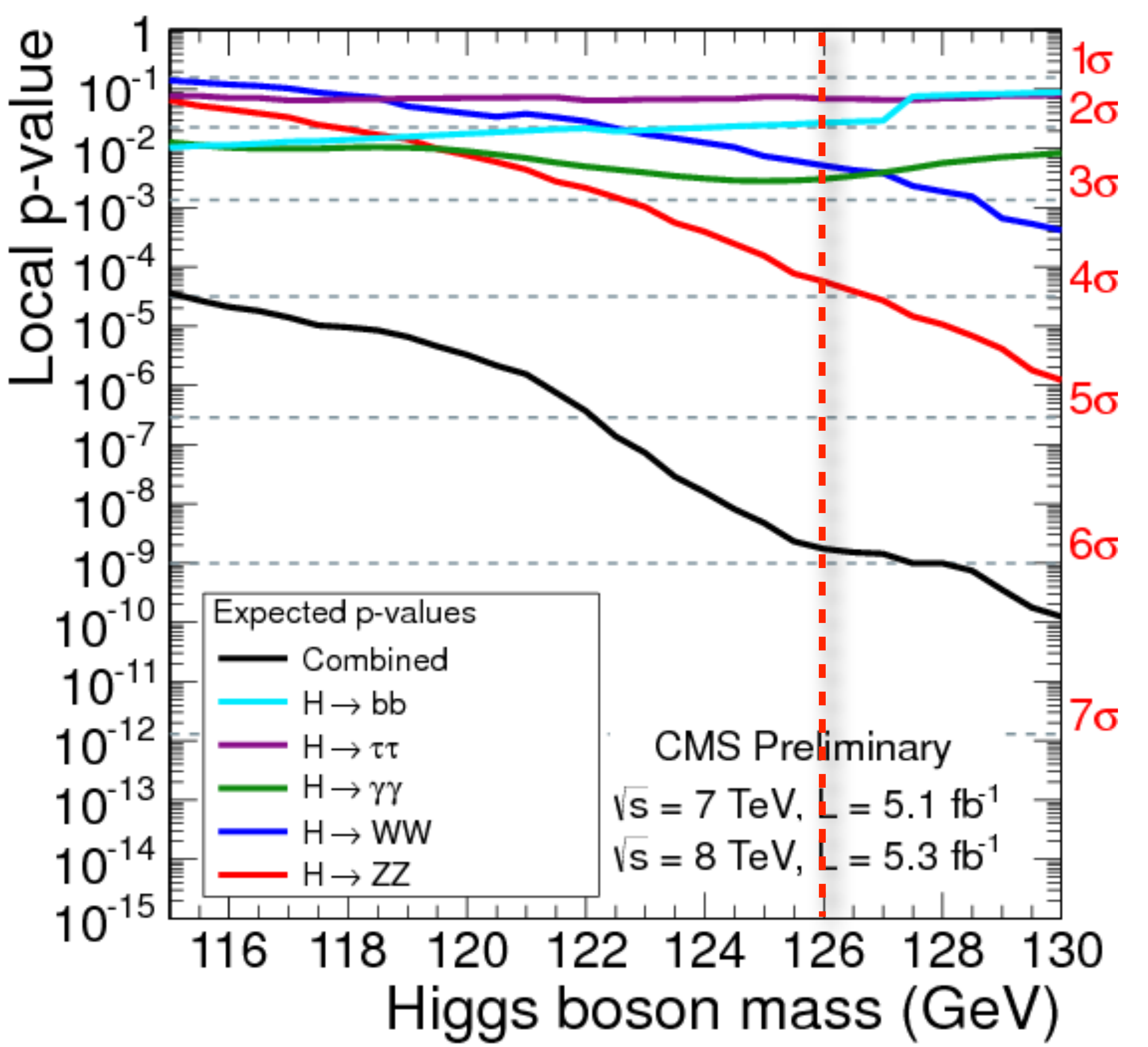
Outline

- Introduction
  - The role of WW in Higgs searches
- Event Selection overview
- Background Estimation
- Higgs Signal Extraction
- Results

Sensitivity to SM Higgs

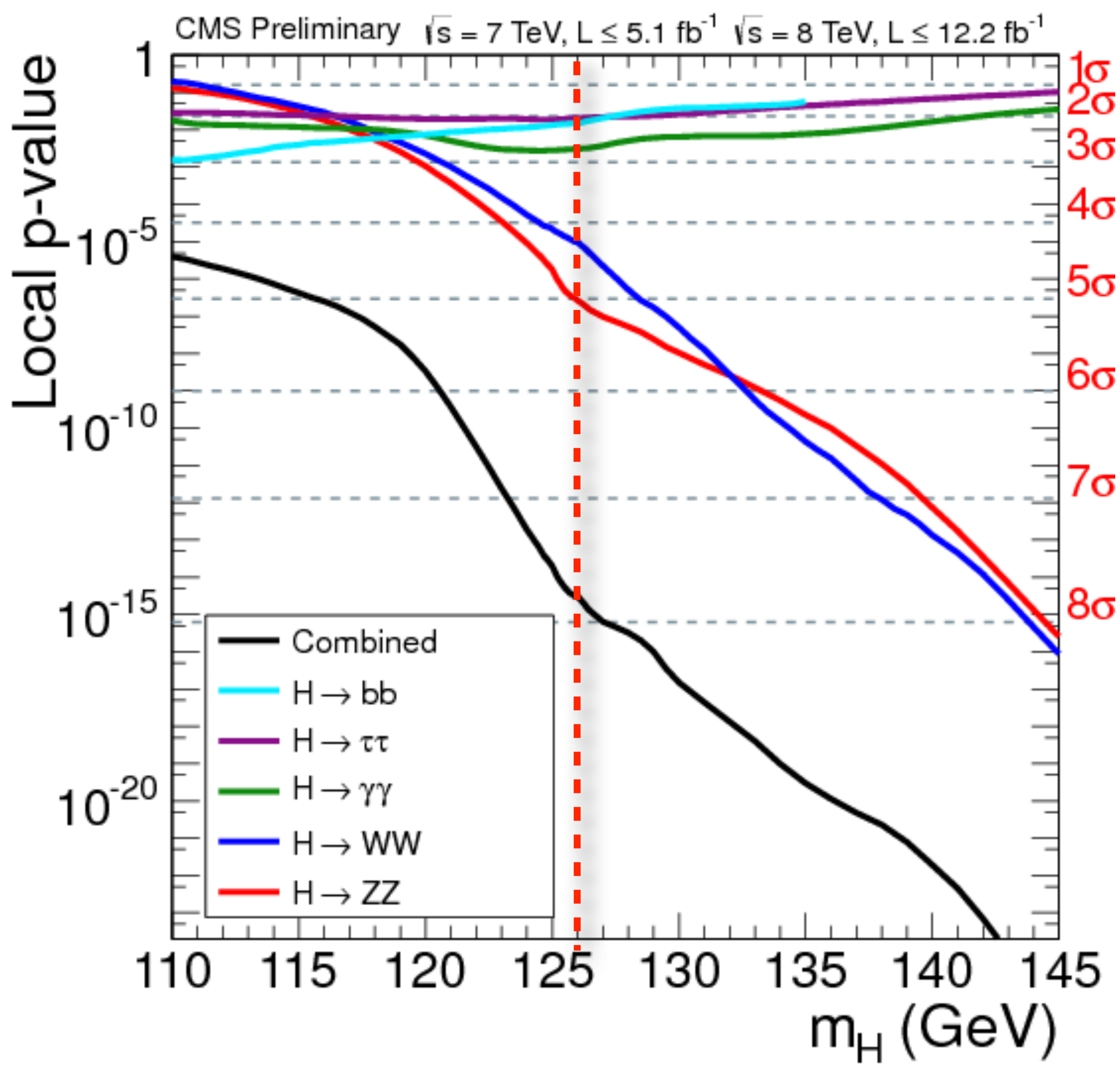
ICHEP 2012 - 12/fb

(WW is cut-based)



HCP 2012 - 17/fb

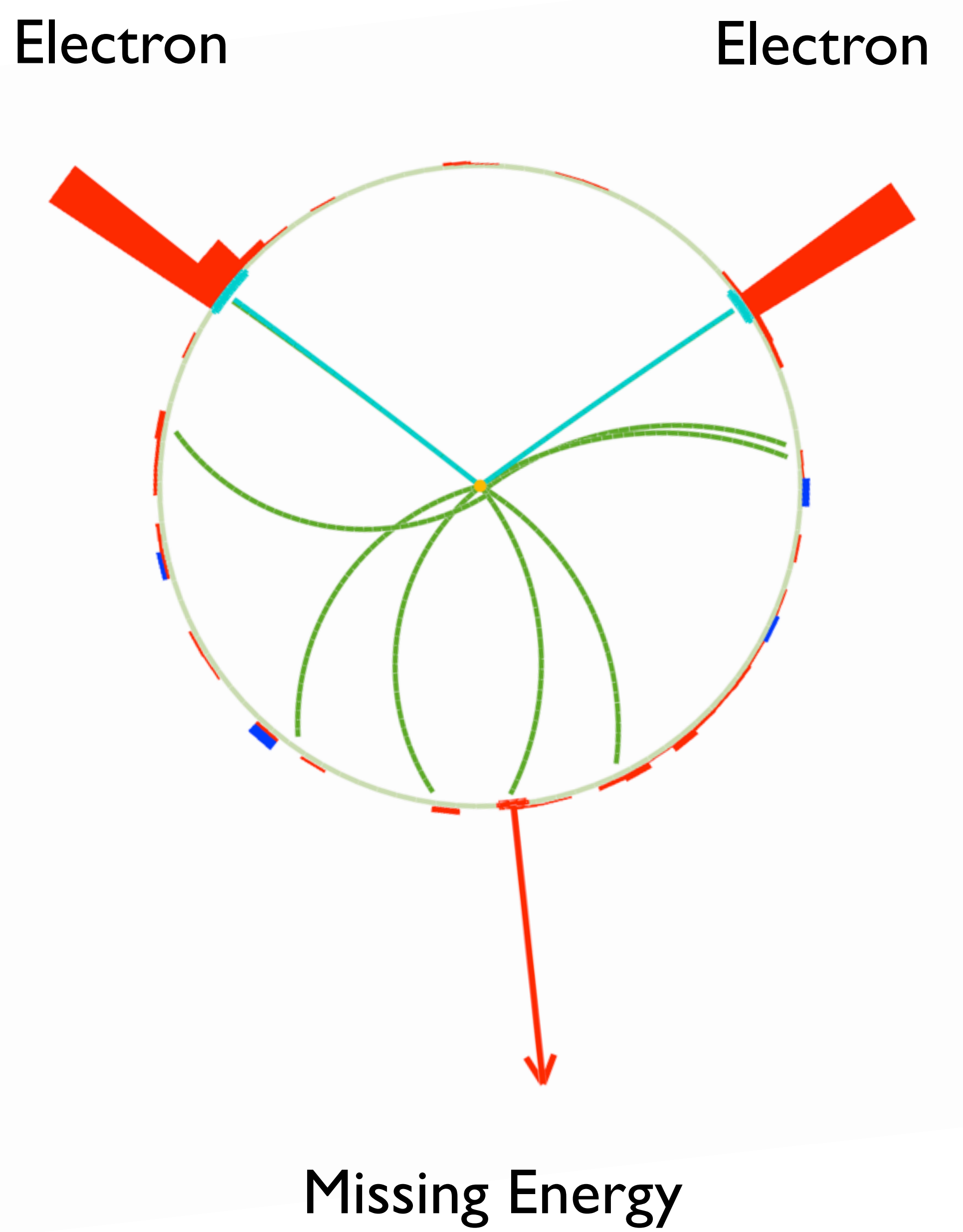
(γγ is just 12/fb)

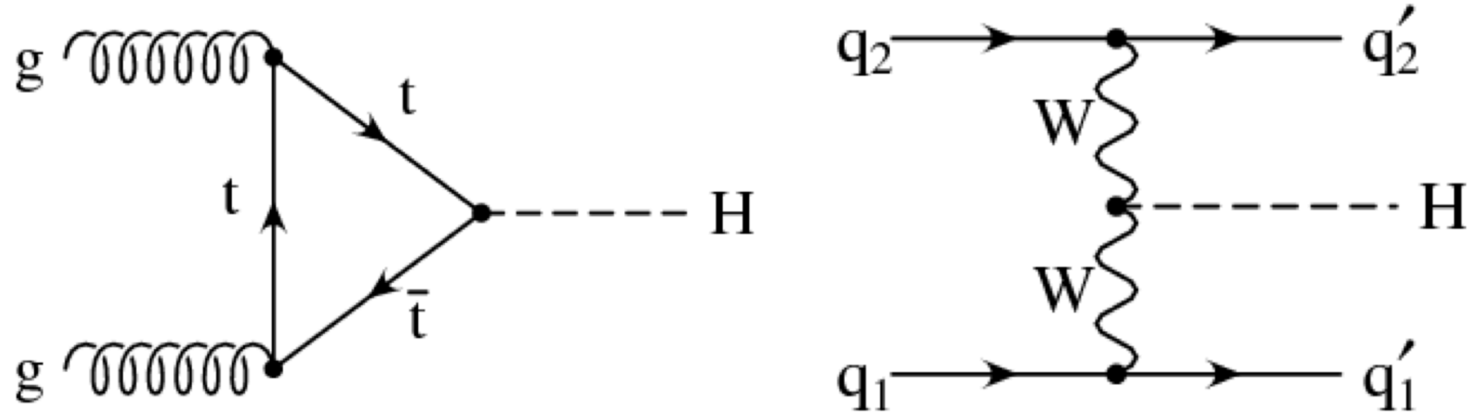


ZZ→4 ℓ ,WW→2 ℓ 2ν and γγ dominate at 126 GeV

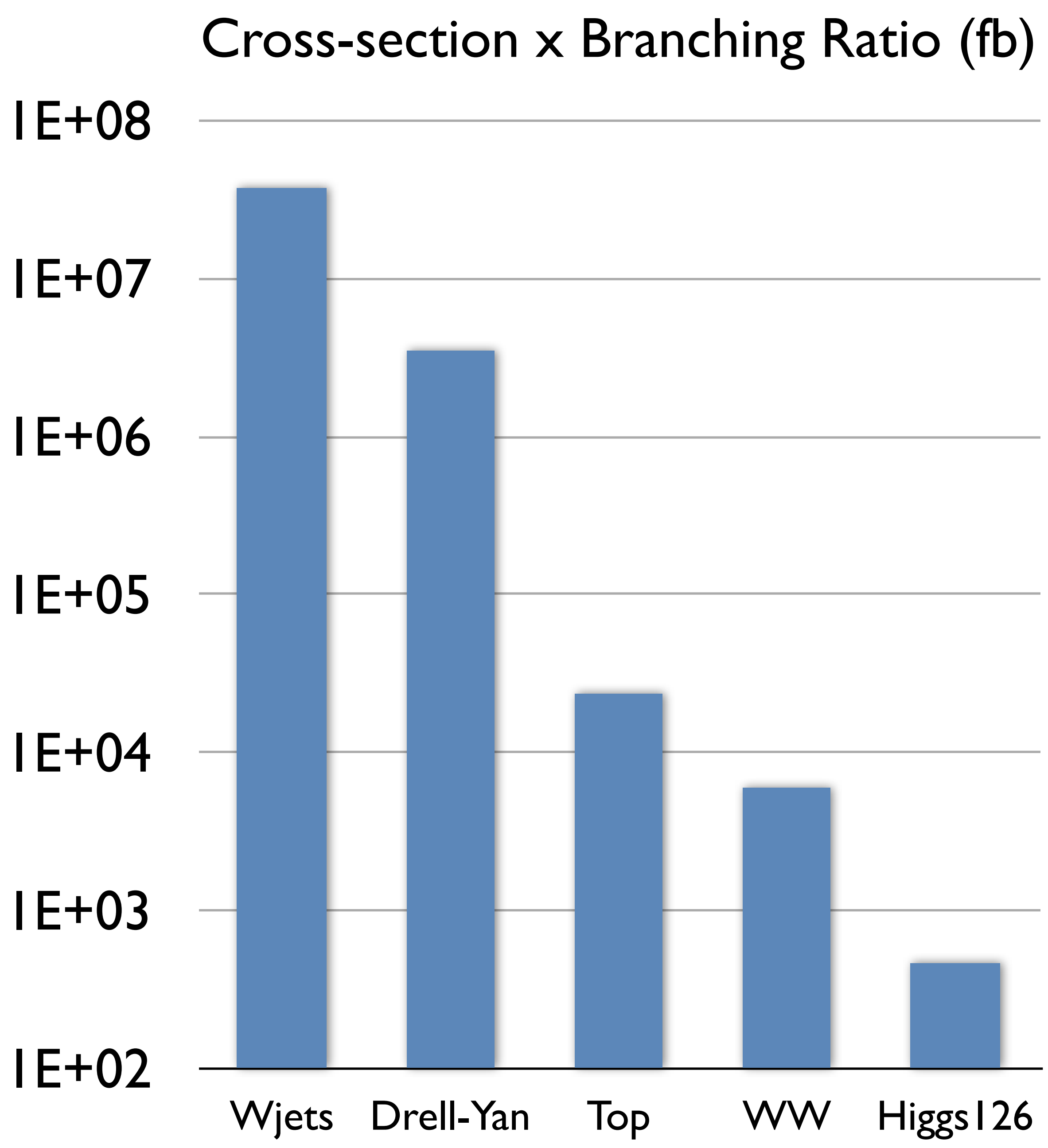
Higgs → WW → 2l2ν

- Higgs Signature:
  - 2 isolated leptons (electron or muon)
  - large missing energy
- Three categories of events:
  - 0, 1 and 2 jets





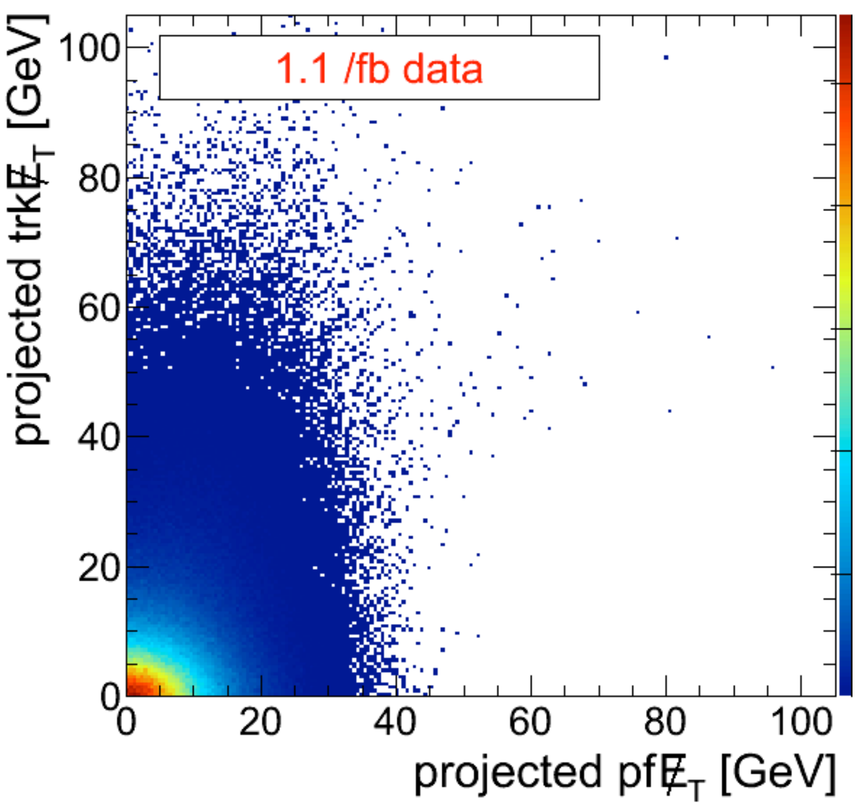
Analysis Challenges

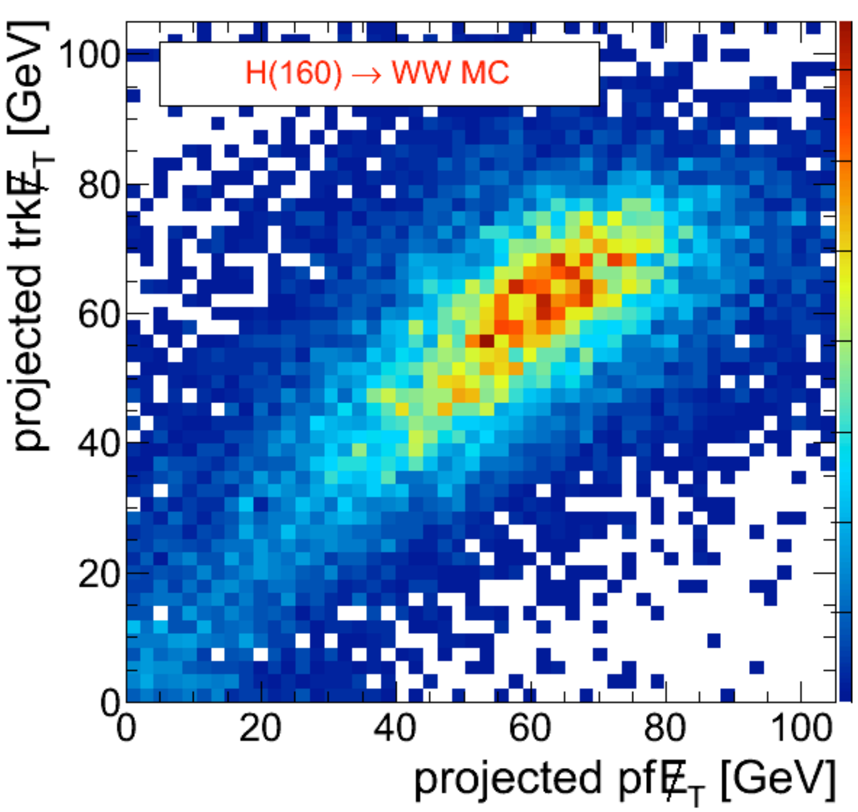


- No mass is reconstructed - essentially a **counting experiment**
- Key selection requirements:
  - lepton pt>10 GeV with tight identification and isolation - **QCD, Wjets**
  - large missing transverse energy (MET) and Z veto - **Drell-Yan**
  - number of jet classification (Pt>30GeV) and b-quark veto - **Top**
  - kinematics (m_{ll}, dφ) - **WW**
- Final step selection requirements are optimized for different Higgs mass hypotheses

Background estimation is the most critical part of the analysis

Drell-Yan background and PileUp





- Pileup condition were changing each year
  - 2010: *~3 interactions per bunch crossing*
  - 2011: *~8 interactions per bunch crossing*
  - 2012: *20-25 interactions per bunch crossing*
- Two different MET variables:
  - nominal - calorimeter and tracker
  - charged tracks based MET
    - not affected by pile up
    - hard to simulate properly - need to get right charge/neutral ratio in the tail of the distribution
- pfMET and trkMET are weakly correlated for backgrounds
  - use the smaller one for each event
  - minMet>40 (same flavor)
  - minMet>20 (opposite flavor)
- 2012 running conditions no longer allow us to perform shape analysis in same-flavor (ee,μμ) final states

Background Estimation

Wjets

A Loose !Tight **QCD**	B Tight **QCD**
C Loose !Tight **Data**	D Tight **Data**

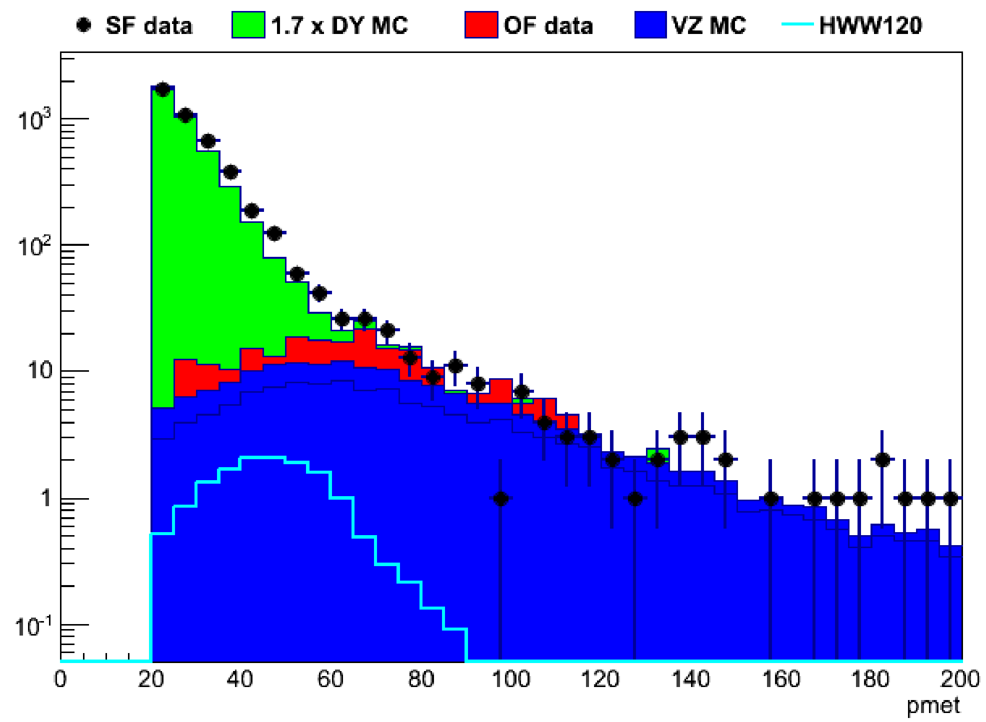
- Jets - main source of fakes
- Requirements: pt, isolation, impact parameter, quality
- Tight →Loose: 10-100 time more fake leptons
- Use QCD sample to measure fake rate: $\varepsilon = N_B/(N_A+N_B)$
- Background estimation:

$$N_D = N_C \frac{\varepsilon}{1-\varepsilon}$$

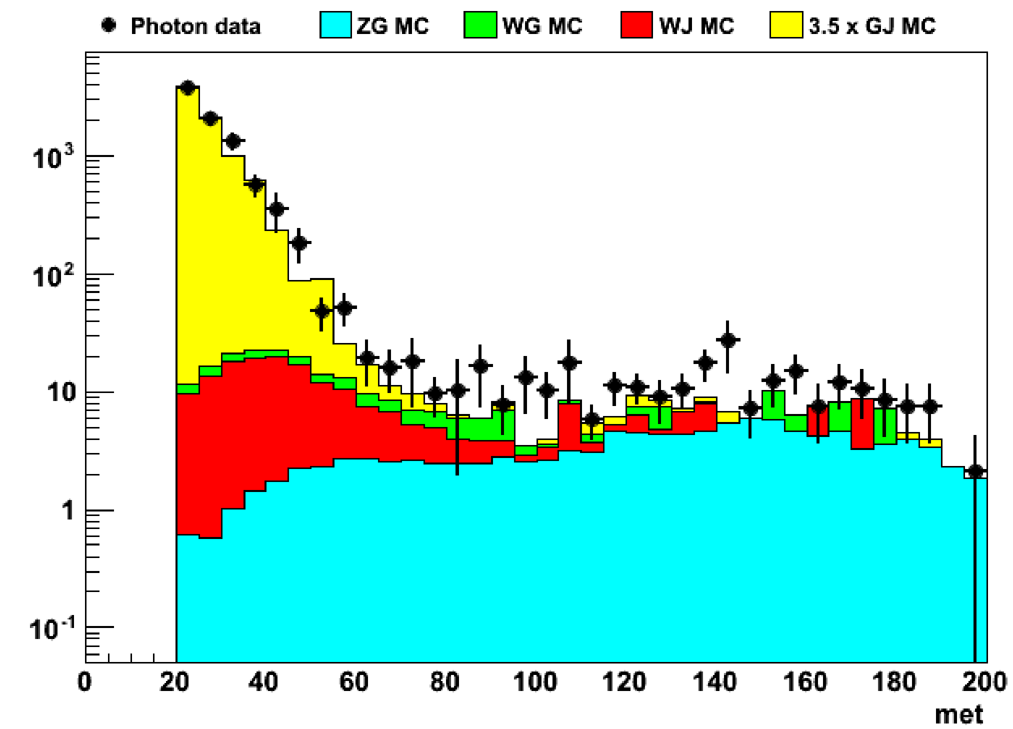
Systematic uncertainty of the method: ~35%

Drell-Yan (ζ-method)

Background	Wjets	Drell-Yan	Wγ*	Top	WW



(a) proj-pfMet in $Z/\gamma^* \to \ell\ell$ sample



(b) pfMet in γ+jets sample

- Similar behavior of MET and trackMET in Z+jets and γ+jets events (reweighted to match the Z pT)
- Compute tight to loose ratio in photon sample (ζ) in bins of pT(γ) and njets:

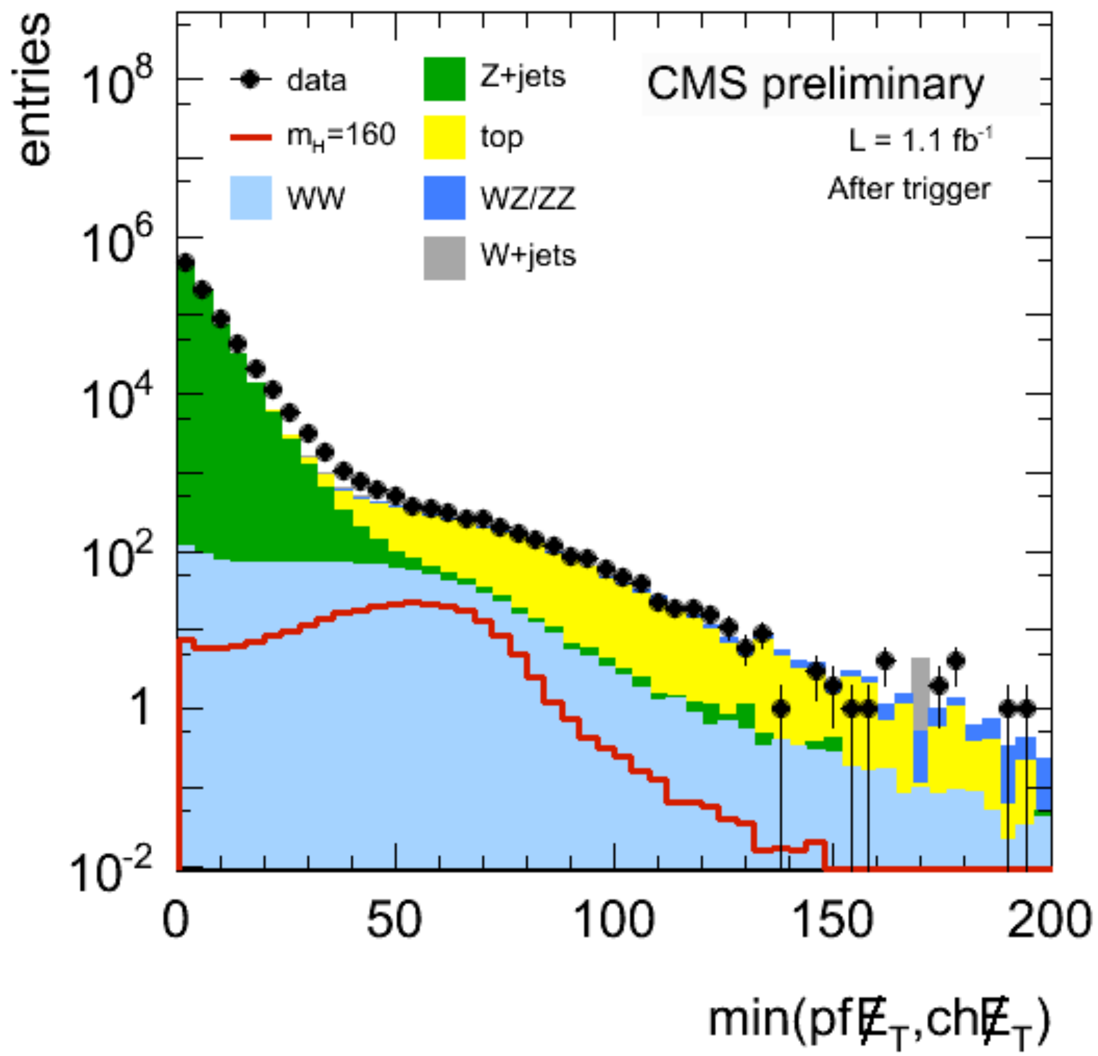
  $\zeta = N_\gamma(MET>45) \,/\, N_\gamma(20<MET<45)$

- Apply ζ to dilepton sample to get DY events after tight cuts:

  $N_{ll}(MET>45) = N_{ll}(20<MET<45) \times \zeta$

Drell-Yan (Z-peak normalization)

Background	Wjets	Drell-Yan	Wγ*	Top	WW



- Drell-Yan: ee/μμ, but not eμ
- Use eμ events to subtract backgrounds
- Narrow Z-peak - little background
- Rout/in is measure both in simulation and in data
- Systematic uncertainties can be as large as 100%

opposite flavor events measured in Z-peak

expected VZ contribution from MC

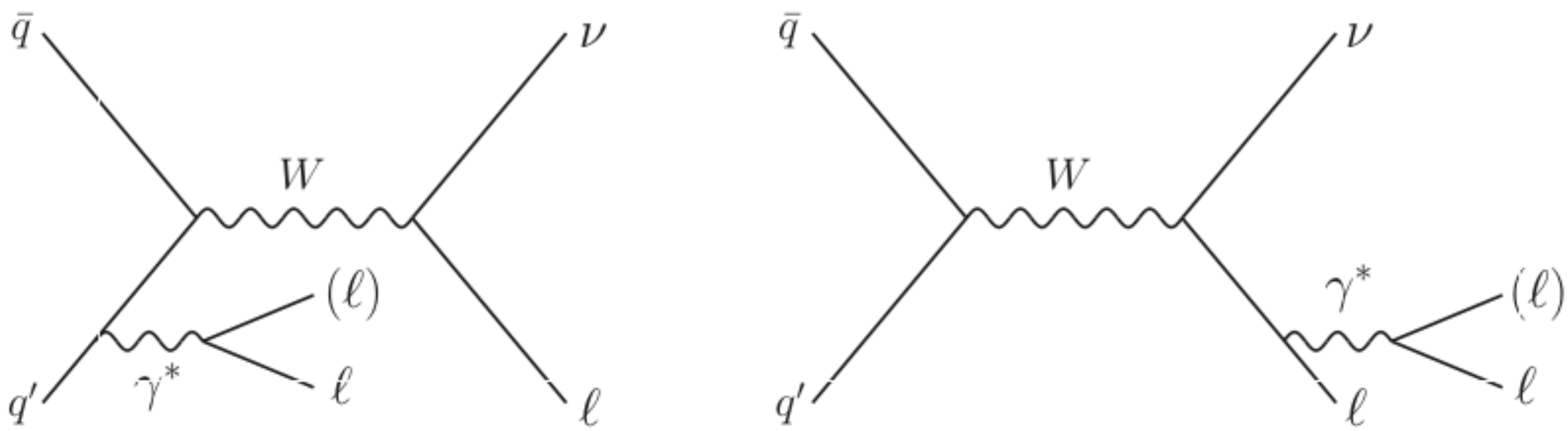
$$N_{out}^{ll,exp} = R_{out/in}^{ll,loose}(N_{in}^{ll} - 0.5N_{in}^{e\mu}k_{ll} - N_{\text{control}}^{\text{ZV, sim.}})$$

same flavor events measured in Z-peak

correction for differences in lepton efficiency

Wγ* Background

R. C. Gray et al., "Backgrounds to Higgs Boson Searches from Asymmetric Internal Conversion", (2011). arXiv:1110.1368.



- One of the leptons from γ* is lost
- Looks similar to Higgs signal
  - Size of the contribution is constrained in same-sign control sample
- Hard to simulated due to "divergence" as m_{γ^*} goes to $2m_\ell$
  - Scott Thomas et all proposed a solution by modifying Madgraph generator

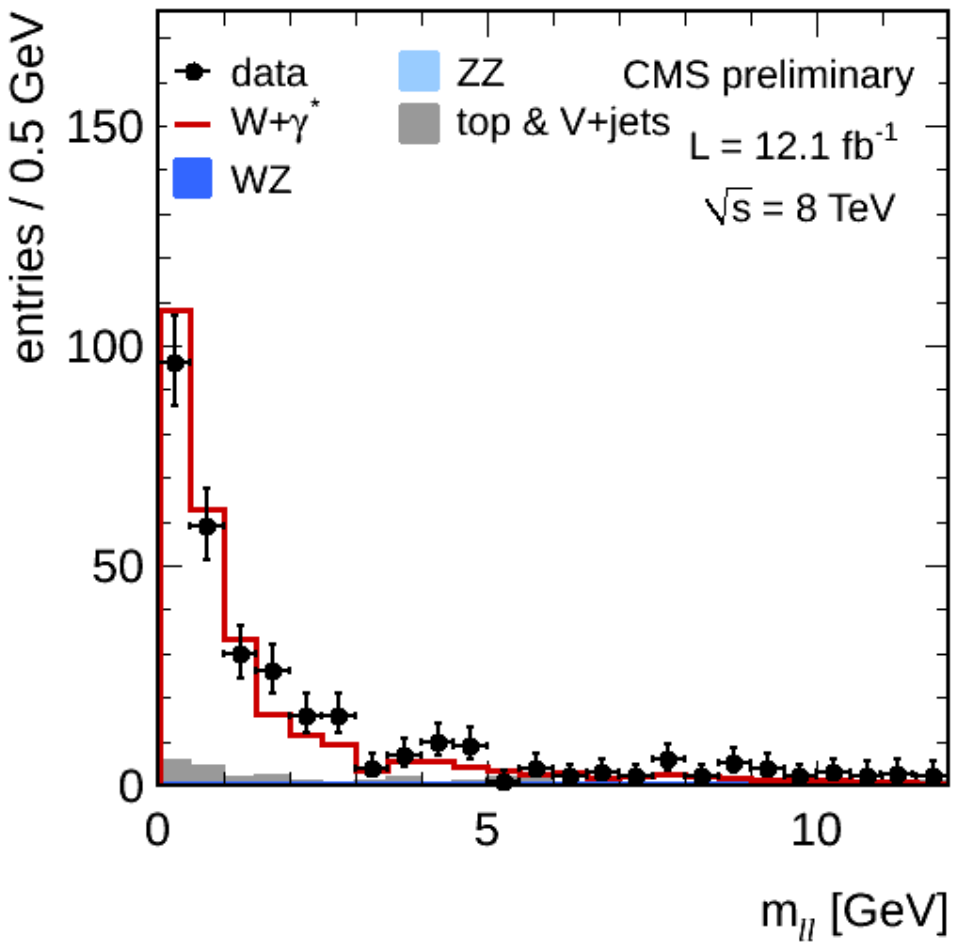
Wγ* - data-driven solution

Background	Wjets	Drell-Yan	Wγ*	Top	WW

- We can isolate 3-lepton events in data and compare the predictions with new MC
  - 1st high Pt lepton comes mostly from W
  - 2 softer leptons (same flavor, opposite charge) mostly from the virtual photon
    - Ranking of leptons can be reshuffled
  - One of the two lepton from γ* at low pT (<10 GeV)
- Only events with γ*→μμ are accessible experimentally at CMS
  - γ*→ee has large background
- Selection requirements:
  - eμμ and μμμ final states: opposite charge is required for γ* muon pair candidate
    - μμ pairing in μμμ case: pair is selected with the smallest mass
  - do not consider the other muon in each muon isolation cone
  - $M_{\mu\mu}<12$, pT>20/10/3, njets<2, anti b-tagging,
  - mT>20 for all lepton-MET pairs
  - mT>45 for the lepton from W-MET pair
  - $|M_{\mu\mu}-3.1|>0.1$ to exclude J/ψ
- Background validation for 3-lepton events: same-sign μμ events

Wγ* corrections

Background	Wjets	Drell-Yan	Wγ*	Top	WW

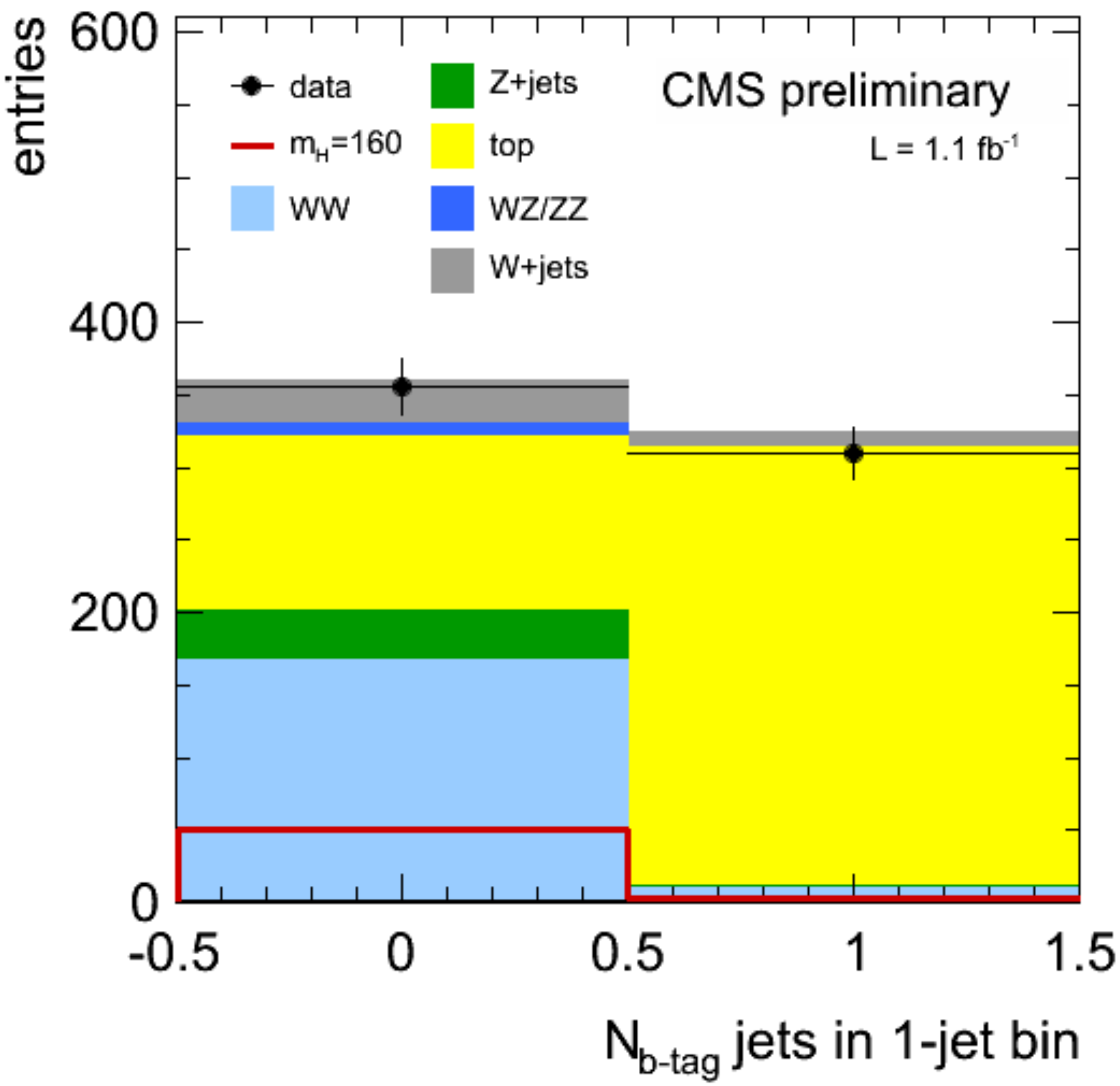


Process	Data	Wγ*	Background	Scale factor
$\ell\mu\mu$, $m_{ll} \in [0,12]$ GeV	319	178.6	32.0	**1.60 ± 0.10**
$e\mu\mu$, $m_{ll} \in [0,2]$ GeV	153	105.8	9.4	1.36 ± 0.12
$e\mu\mu$, $m_{ll} \in [2,12]$ GeV	65	25.2	12.5	2.08 ± 0.32
$\mu\mu\mu$, $m_{ll} \in [0,2]$ GeV	68	32.1	4.5	1.98 ± 0.26
$\mu\mu\mu$, $m_{ll} \in [2,12]$ GeV	33	15.4	5.7	1.77 ± 0.37

- We find some discrepancy in the shape of mll distribution between data and MC
- We use two mll regions [0,2] and [2,12] GeV to estimate the fudge-factors
  - The difference is the systematic uncertainty
- Final scale-factor used in the analysis in $H \to WW \to 2\ell 2\nu$ is **1.6±0.3**
  - Consistent with k-factors NLO/LO for di-boson processes

Top Background

Background	Wjets	Drell-Yan	Wγ*	Top	WW



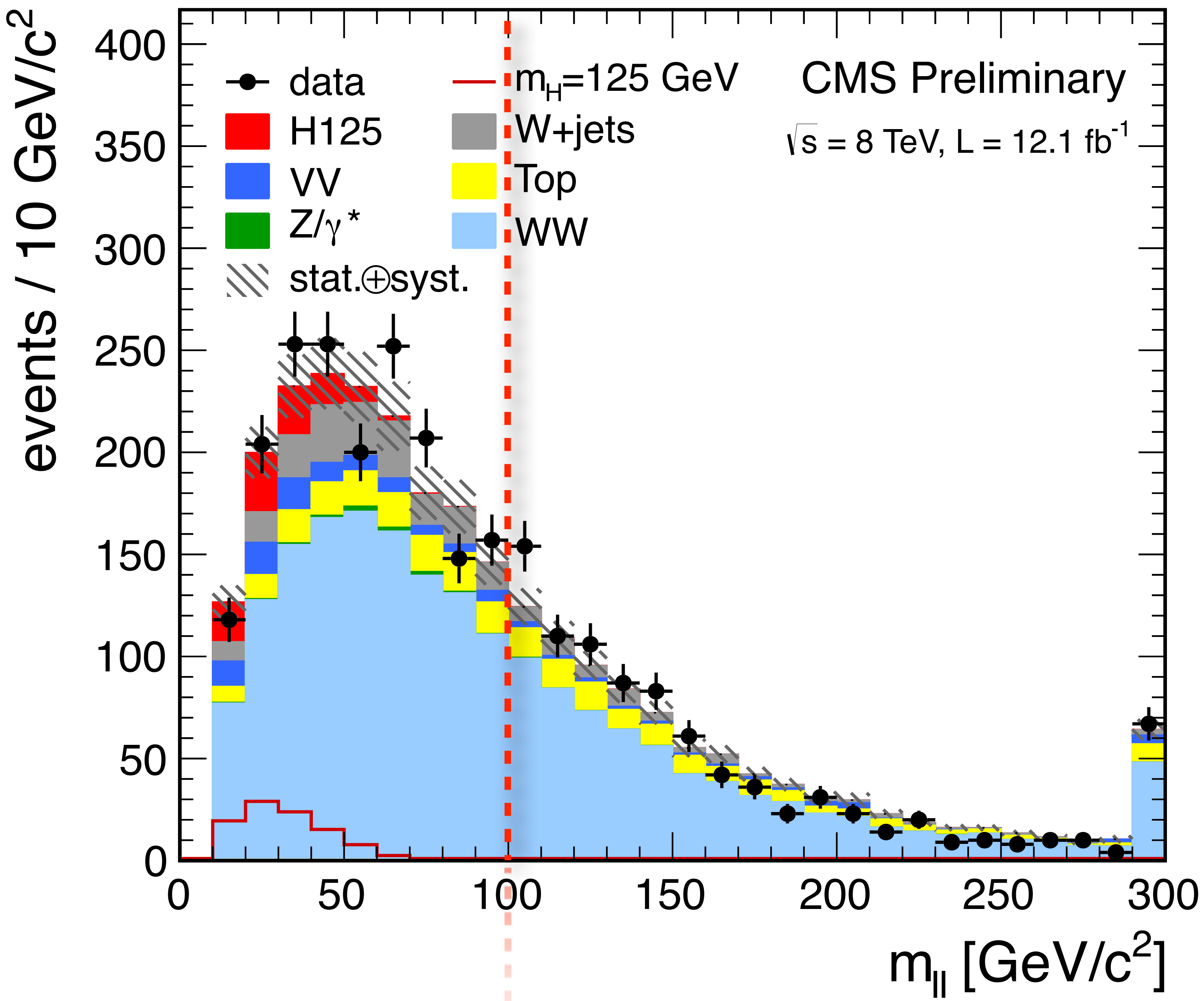
- Jet veto kills top
- Remaining top can be tagged:
  - soft b-jets
  - soft muons
- Top tagging eff is ~50% for 0-jet
- Estimate residual top:

$$N_{top} = N_{tag}\frac{\varepsilon}{1-\varepsilon}$$

- Measure ε in 1-bjet events
  - There mast be another b-quark
- Systematics ~ 20-30%

WW Background

Background	Wjets	Drell-Yan	Wγ*	Top	WW



- WW is an irreducible background - one order of magnitude larger SM Higgs
- Kinematics is the main discriminator:
  - low mass - dPhi, Mll
  - for mH≤130 need to lower lepton pt → larger Wjets background
  - above 200GeV - WW and Higgs harder to distinguish
- Use signal free events to calibrate WW yield
  - mll>100 GeV for low mass Higgs

Signal Extraction

Introduction

- Cut-based analysis
  - Most conservative approach
  - Systematics limited at 10/fb
  - Very minor changes - hard to improve
- Shape-based analysis
  - Extension of the cut-based analysis - set of "cut-based" analyses with different Signal-to-Background ratios
  - "Shape" correlates contributions between different channels
  - Statistics limited
  - MVA shape analysis best sensitivity for a wide range of Higgs mass hypotheses
  - 2D shape analysis using simple observables - comparable to MVA in ~125 GeV region, but simpler interpretation

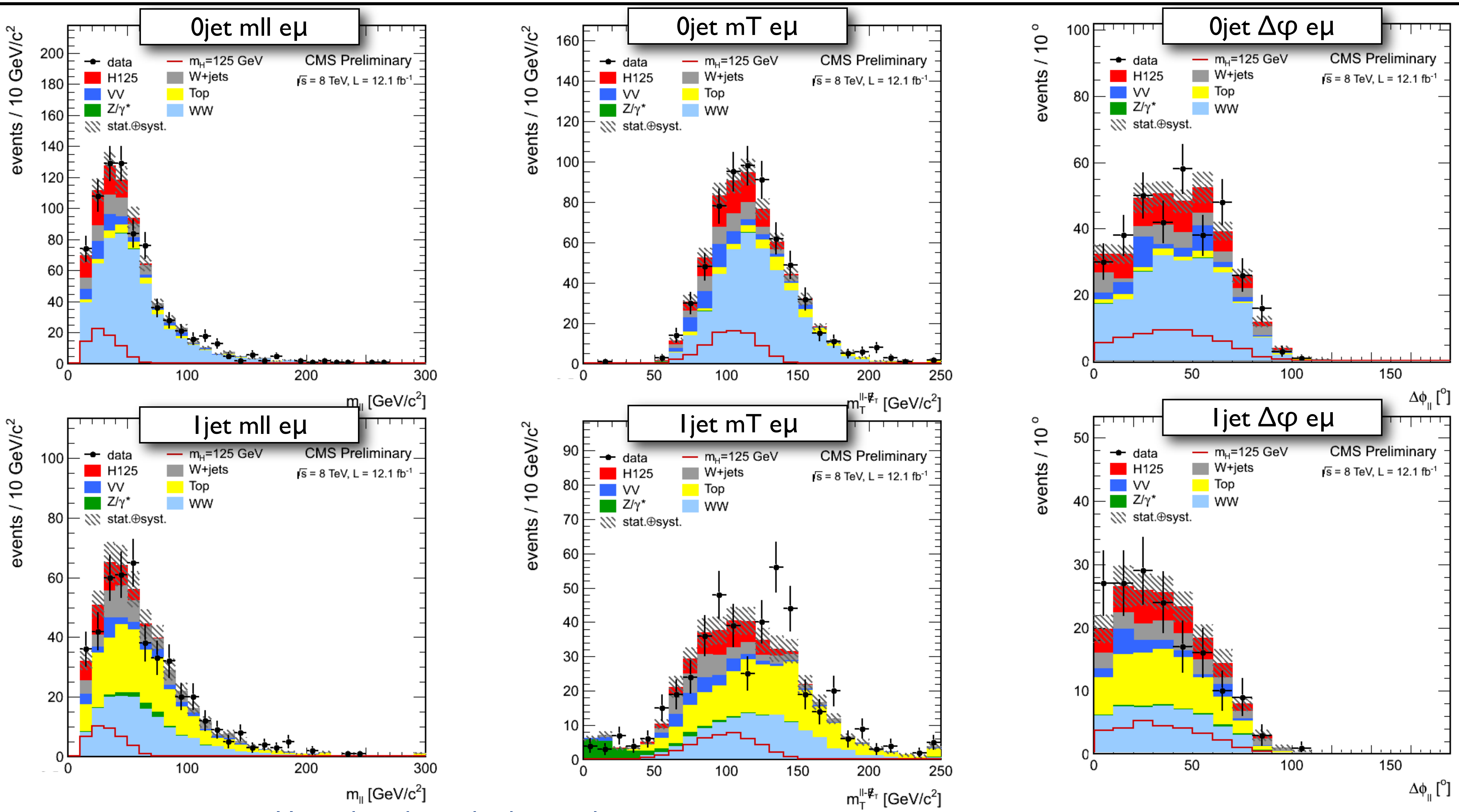
Cut Based Analysis

m_H	$p_T^{\ell,max}$	$p_T^{\ell,min}$	$m_{\ell\ell}$	$\Delta\phi_{\ell\ell}$	$m_T^{\ell\ell E_T^{miss}}$
[GeV/c^2]	[GeV/c]	[GeV/c]	[GeV/c^2]	[dg.]	[GeV/c^2]
	>	>	<	<	[,]
120	20	10(15)	40	115	[80,120]
130	25	10(15)	45	90	[80,125]
160	30	25	50	60	[90,160]
200	40	25	90	100	[120,200]
250	55	25	150	140	[120,250]
300	70	25	200	175	[120,300]
400	90	25	300	175	[120,400]

- Discriminating variables:
  - di-lepton mass
  - angle between two leptons
  - lepton pt
  - transverse mass (dilepton + MET)
- For 2-jets: |Δη|>3.5, mjj>500GeV

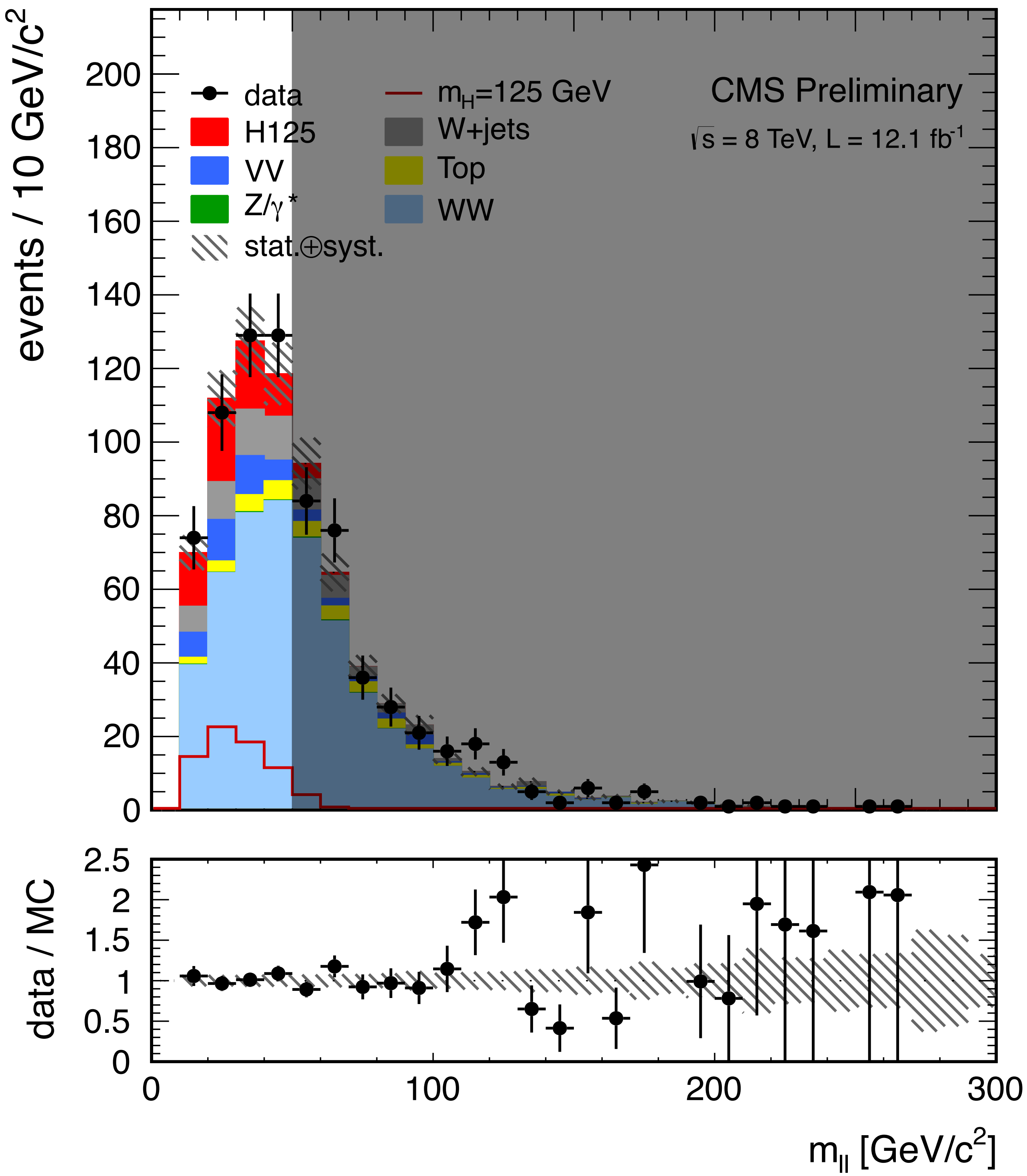
- Background estimation:
  - from data at Higgs selection level: Wjets, Drell-Yan, WW, Top
  - from Monte Carlo and control samples: WZ, ZZ, Wγ(*), Drell-Yan→ττ

Cut-based Results



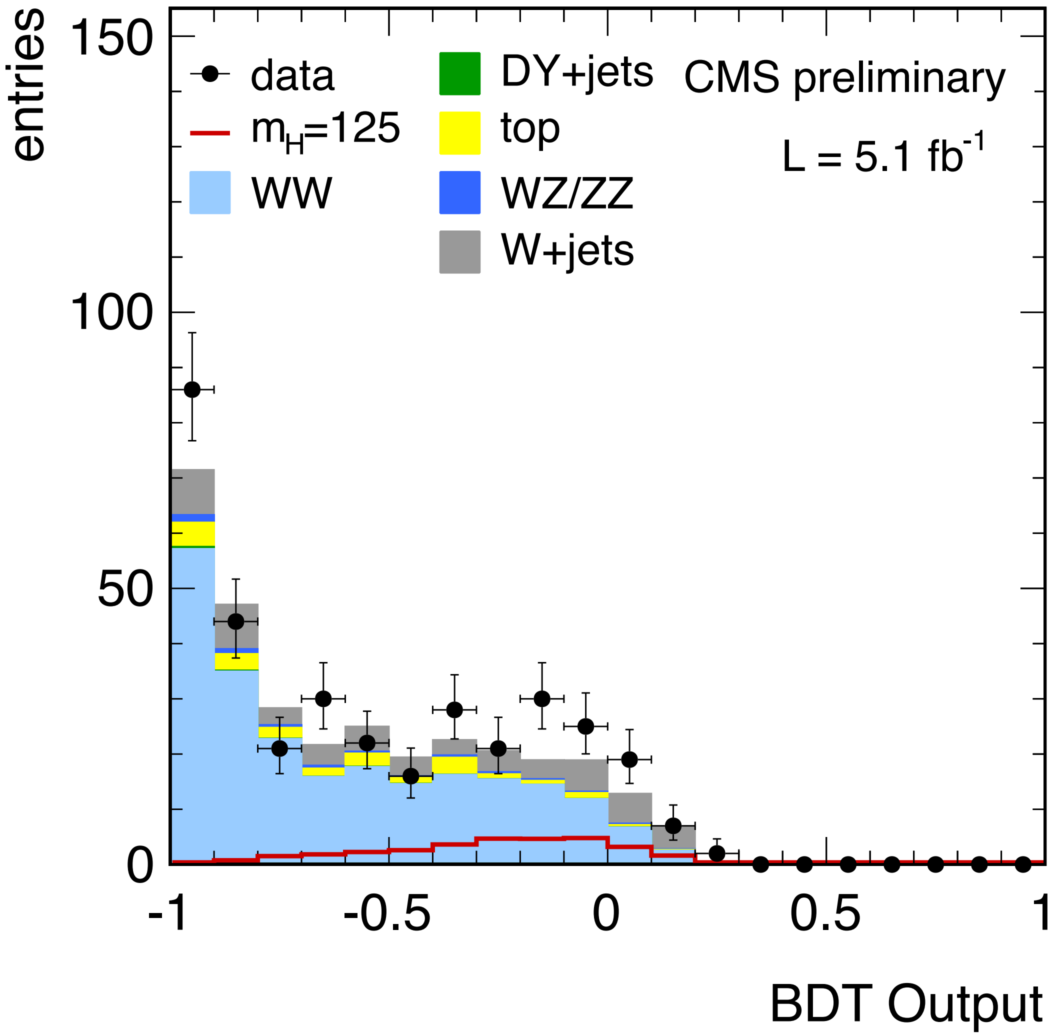
- Using data-driven background estimations
- Expected/observed significance of excess at 125GeV with 8TeV only: 2.4σ/1.7σ
- Best fit value for signal strength: $\mu = 0.80 \pm 0.45$

Why Shape Analysis?



- The plot shows
  - 0-jet category different flavor events
  - cut-based selection, but no cuts on mll
    - $mT \in [80,123]$ GeV
- Cut-based analysis combines the first 4 bins together - S/B is poor
- Shape-based analysis uses all information available giving higher weights to bins with better S/B
- Shape analysis uses sidebands and difference in S/B in the signal box to extract background normalization from data

Where is MVA?

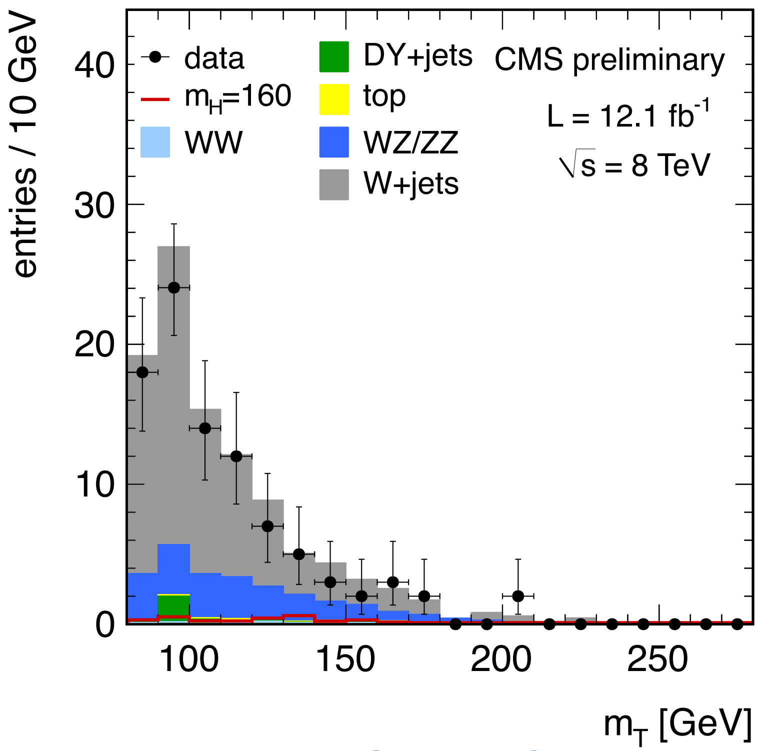


- MVA has a number of advantages:
  - best sensitivity tuned for each Higgs mass
    - MVA find most optimal discriminators
  - single shape to fit
- MVA has some disadvantages:
  - not smooth transitions between mass points
  - it is harder to control systematic effects
  - most important:
    - if data and MC don't agree how do we know what is going on
    - MVA has no simple way to map an excess in the discriminator to kinematic observables like pT, mT, mll etc

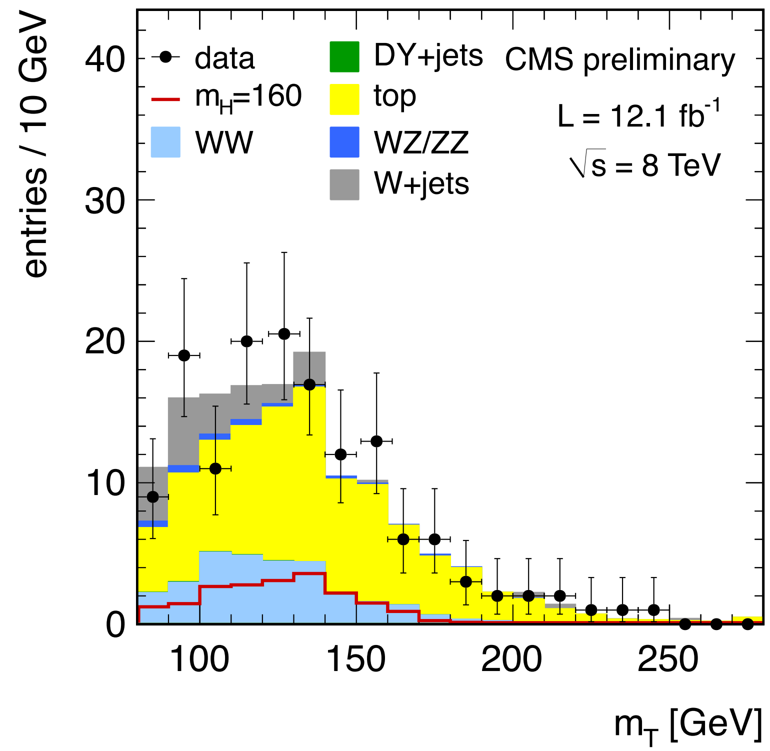
- For Higgs below 140 GeV mll is close in performance to MVA discriminator
- We no longer need ultimate sensitivity in a wide range
  - Time to change to simpler analysis with best sensitivity in the right range

Control Samples

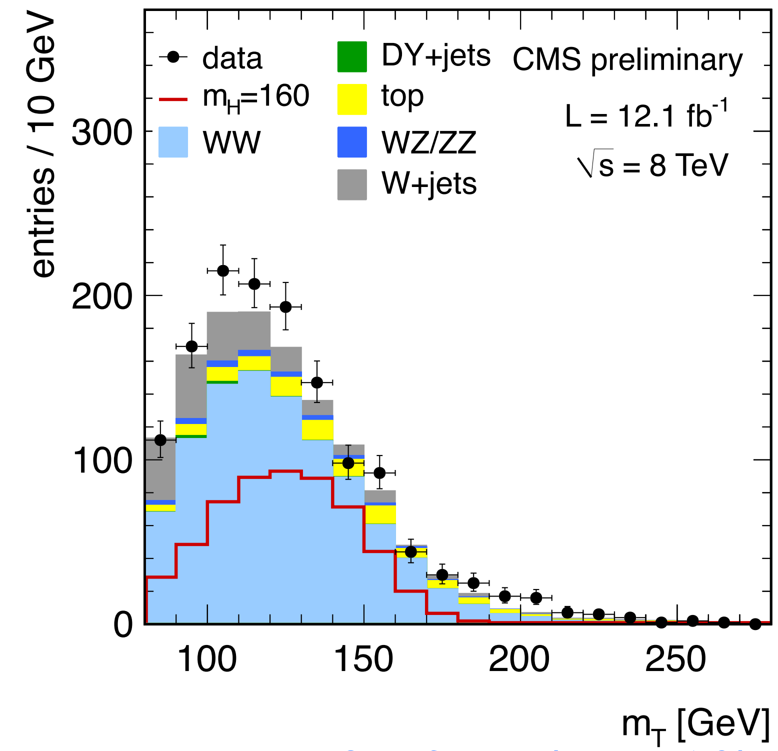
0jet Same-Sign



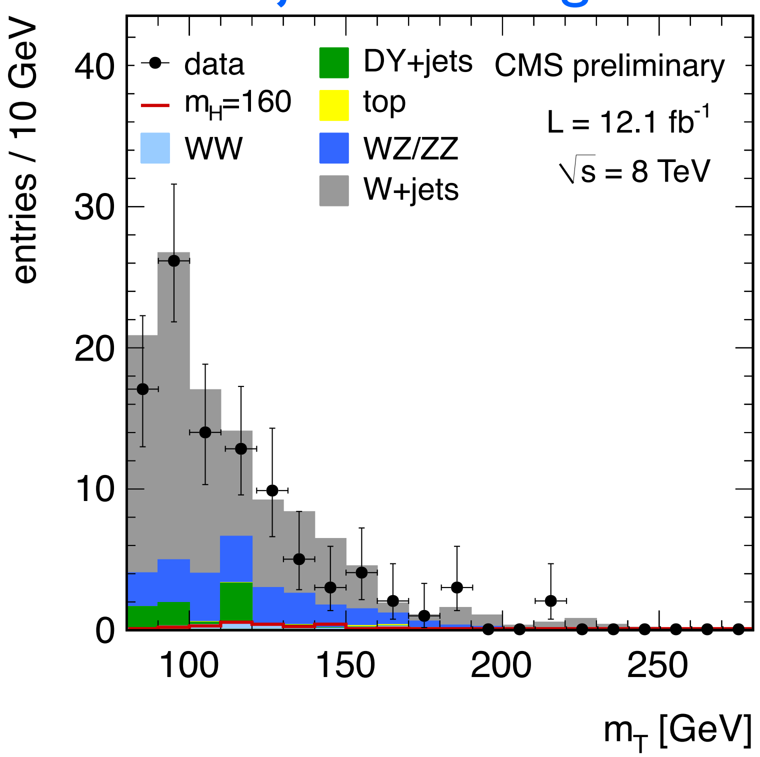
0jet Top-tagged



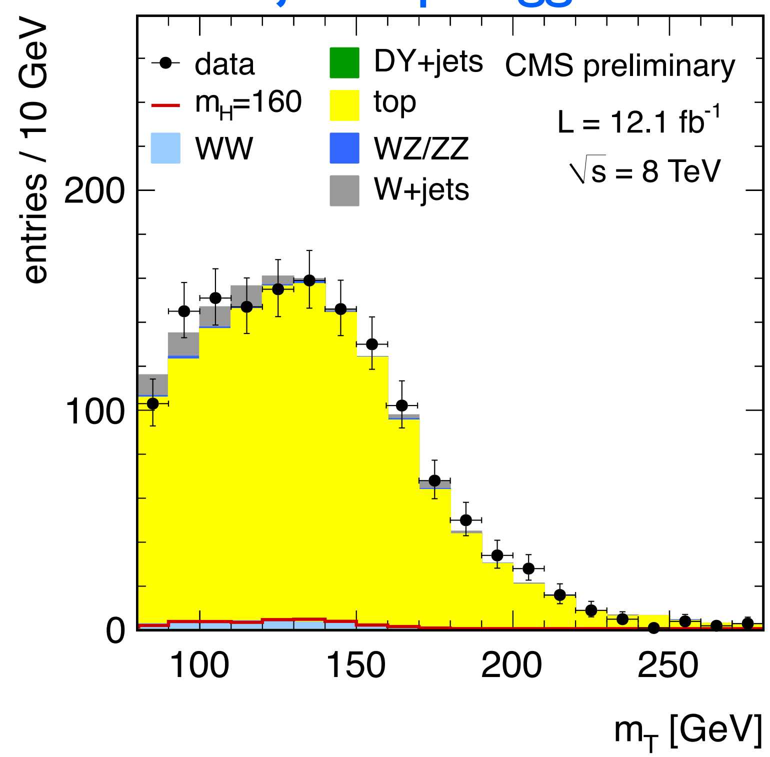
0jet H125 free (mll>60)



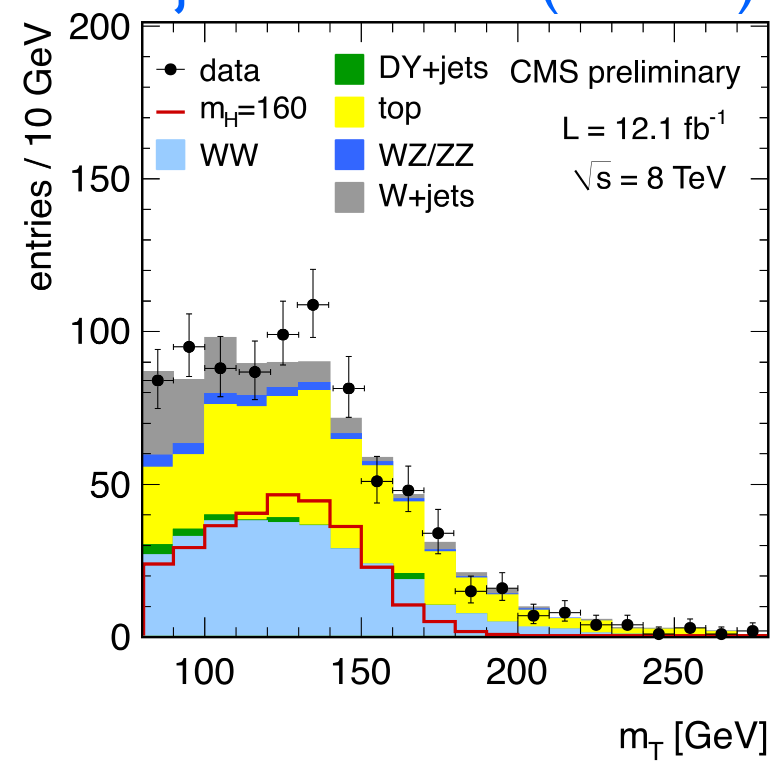
1jet Same-Sign



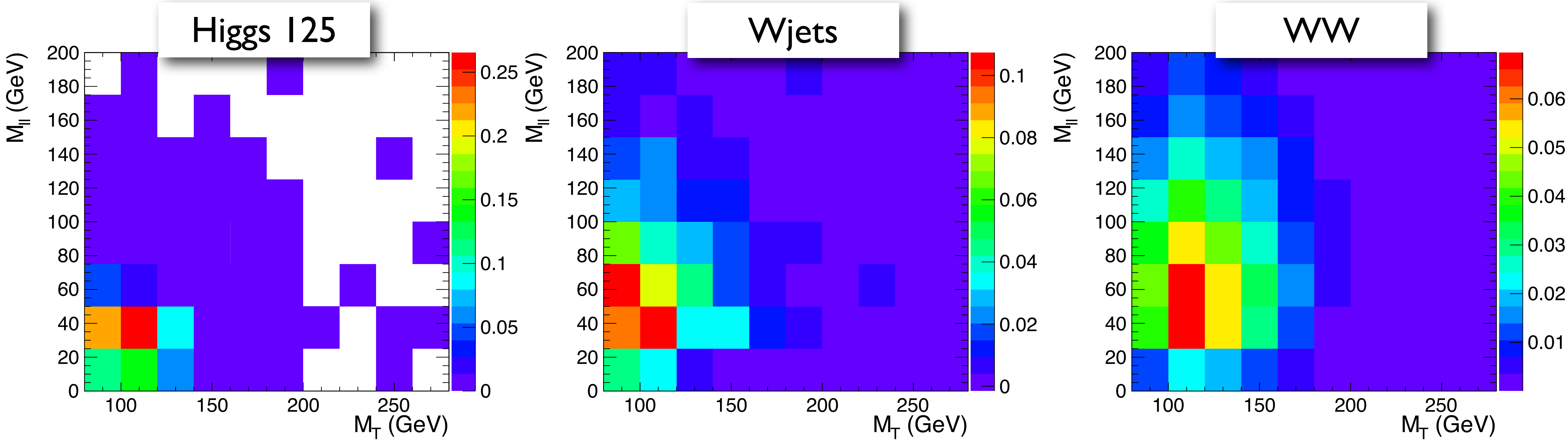
1jet Top-tagged



1jet H125 free (mll>60)



2D analysis

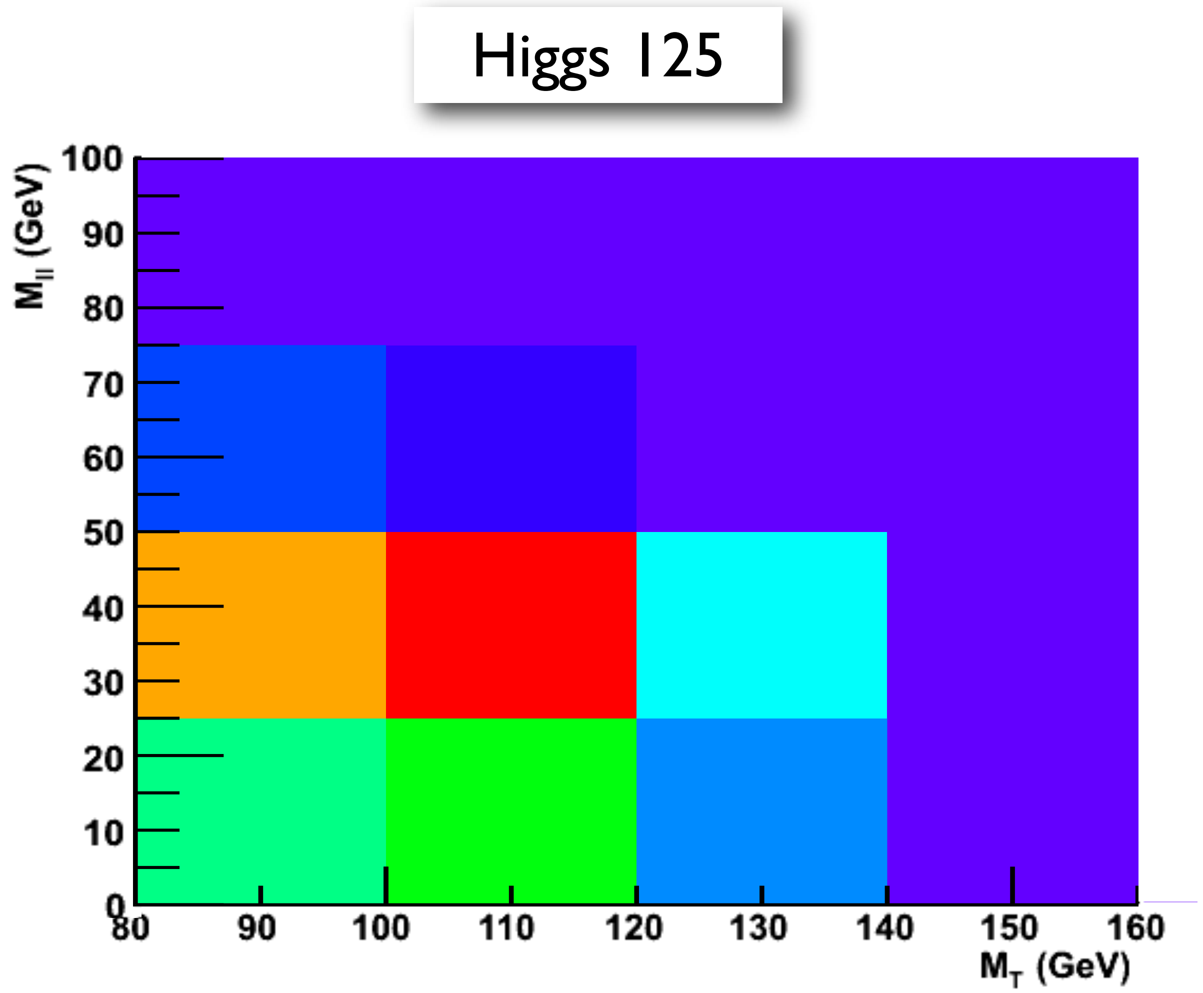


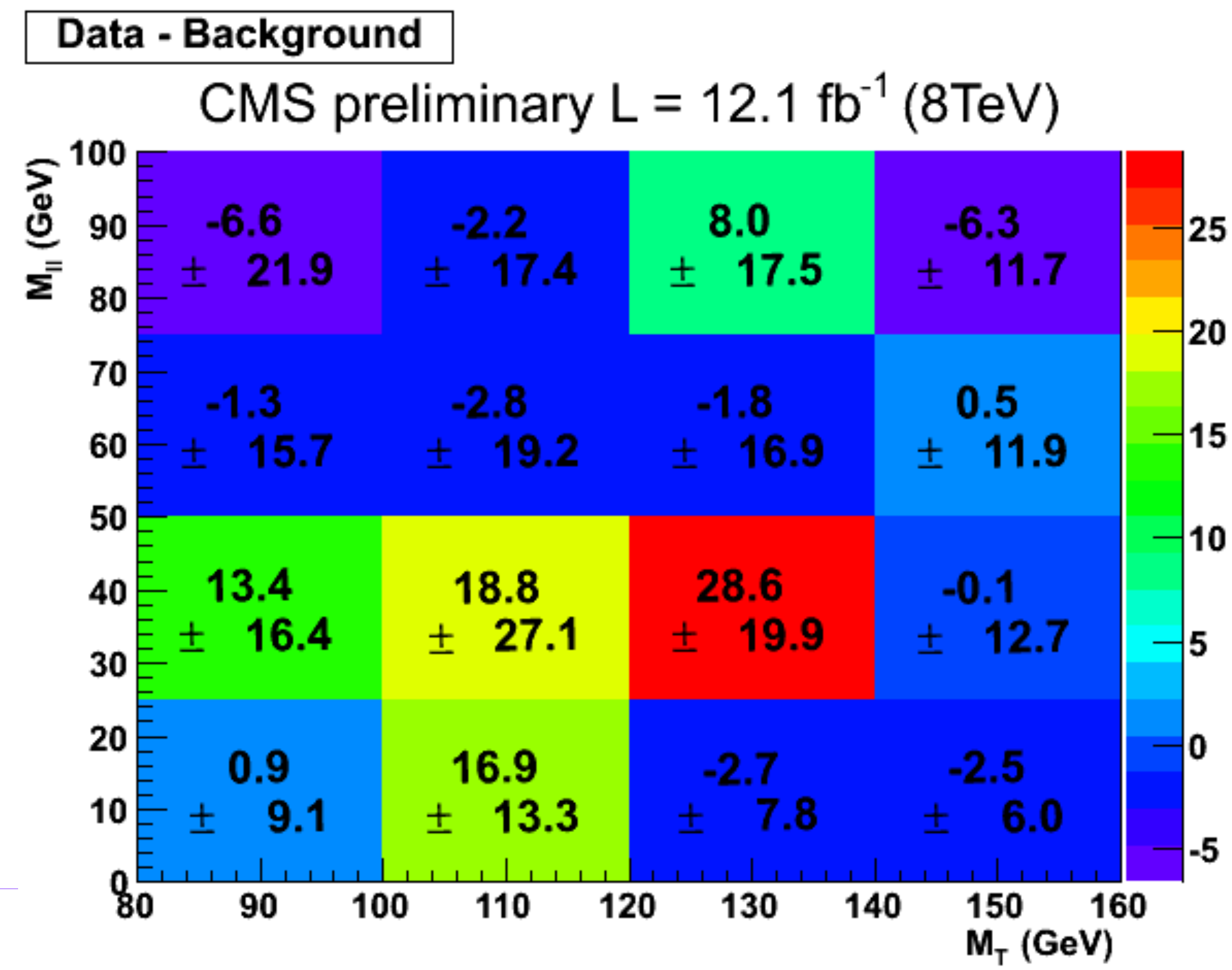
- 2D analysis (mll vs mT) has slightly better sensitivity than 1D MVA at 125GeV
  - Dominant backgrounds populate different regions in 2D plain
  - Reasonable sensitivity at other mass points
- Excesses are easy to interpret

Shape Systematics

- Large number of systematic effects are considered in the analysis. Here is just a brief list of most critical
- Background shapes:
  - WW
    - Default shape from Madgraph
    - Alternatives from MC@NLO with scale variation up/down
      - different showering programs are used Pythia6 vs Herwig6, different PDFs
  - Top
    - Default shape from Powheg
    - Alternative shapes from Madgraph
    - Validated on data
  - Wjets
    - Default from the fake rate method
    - Alternatives from the fakes extracted with different recoiling jet pt in the QCD control sample
- Instrumental uncertainties on basic objects for all shapes determined from MC
  - Lepton pT resolution, MET resolution, Lepton selection efficiency

Shape Analysis Results

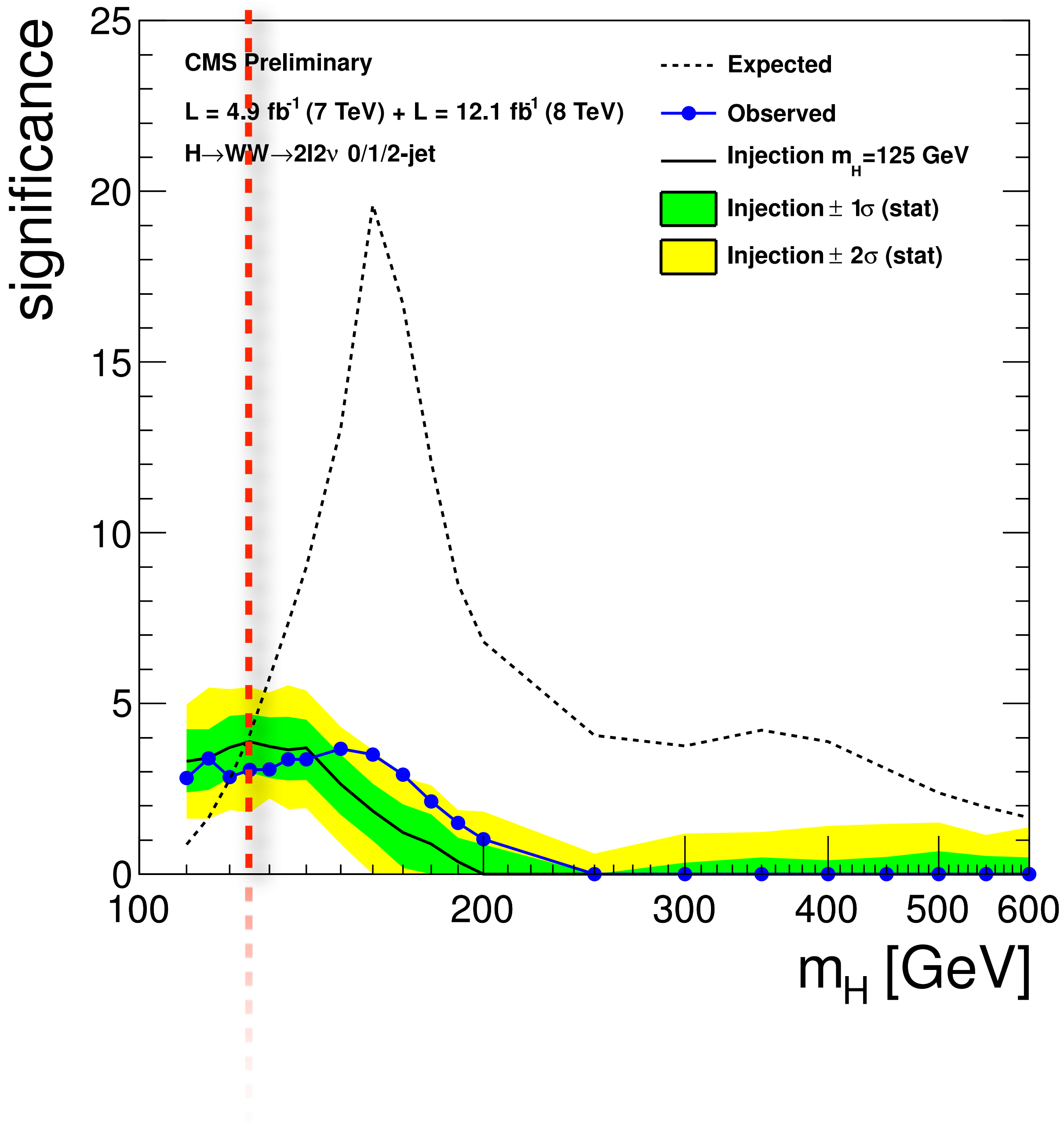


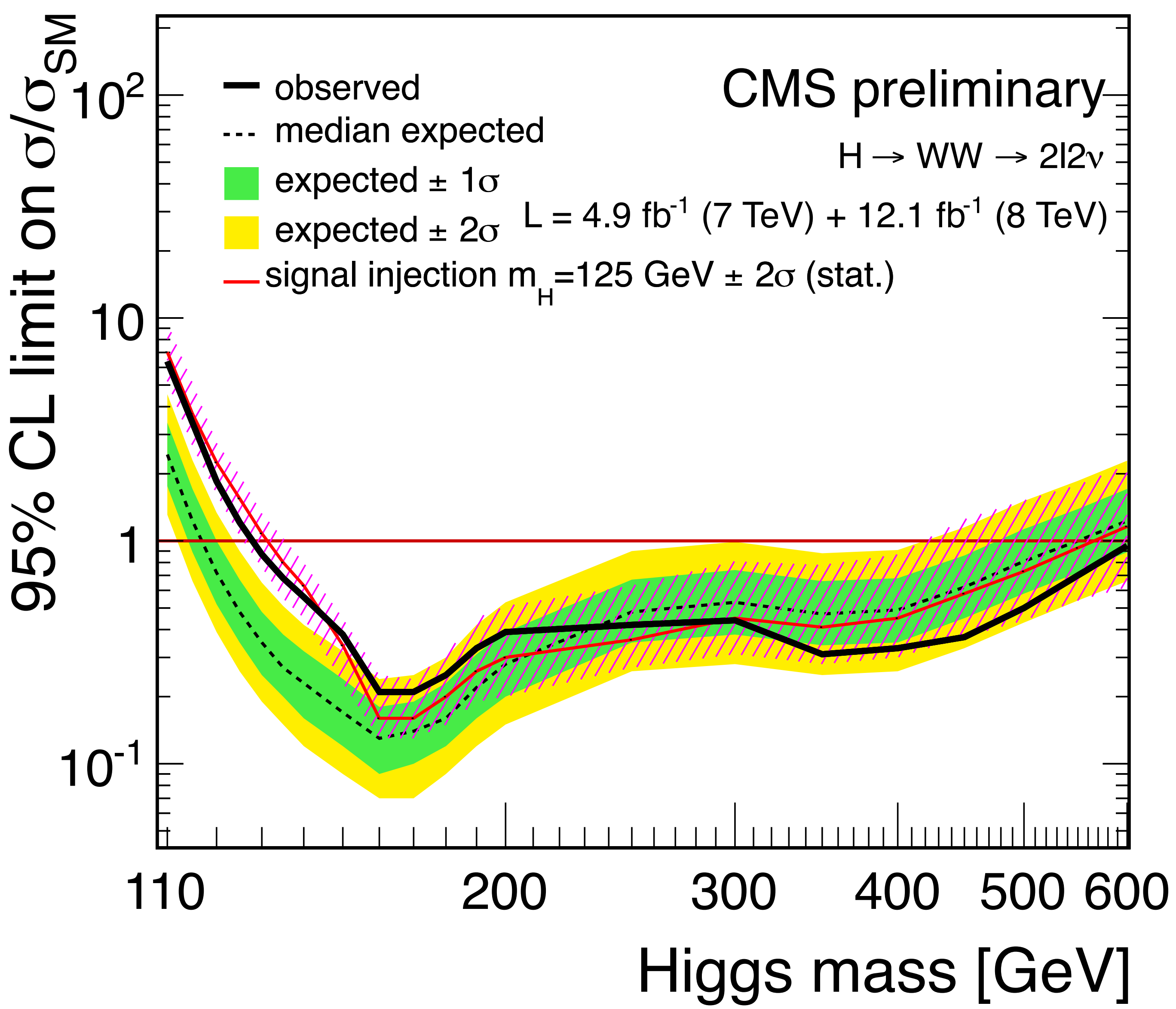


- Zoomed-in version of the most sensitive channel
- 2D analysis is used for 0 and 1jet channels eμ only
- Expected/observed significance of excess at 125GeV for 7+8TeV: 4.1σ/3.1σ
- Best fit value for signal strength: μ = 0.74 ± 0.25

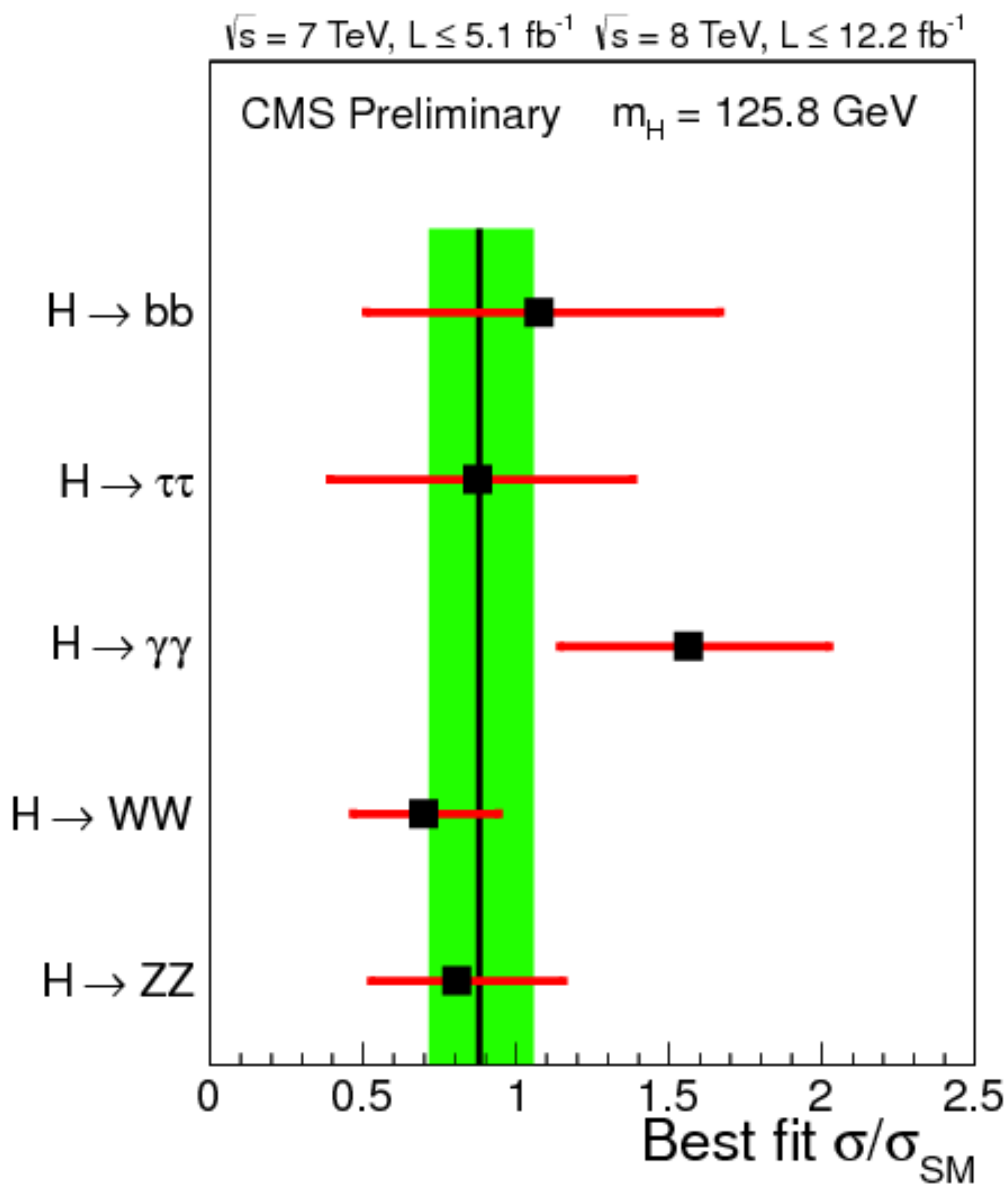
Results

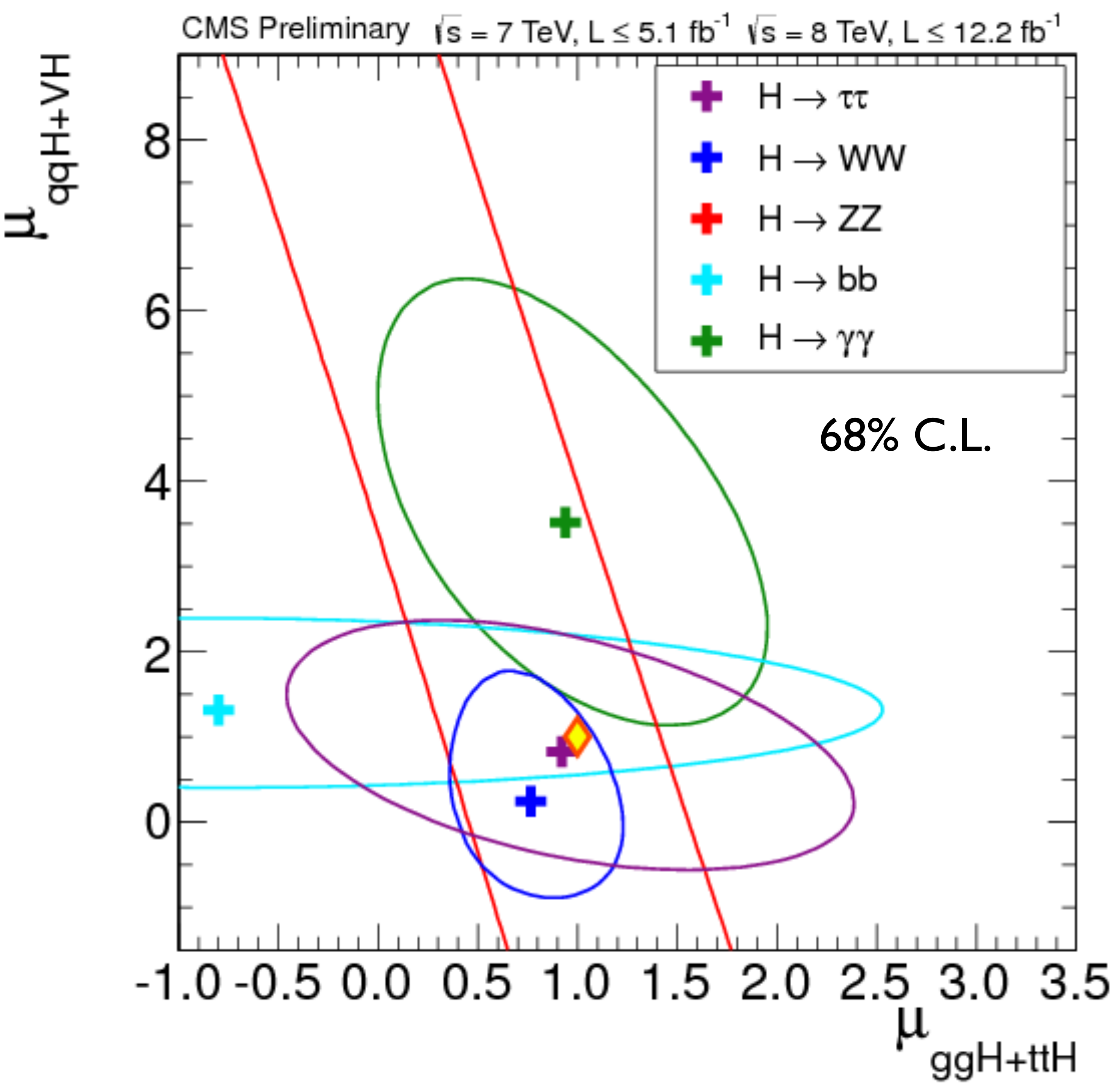
Final Results



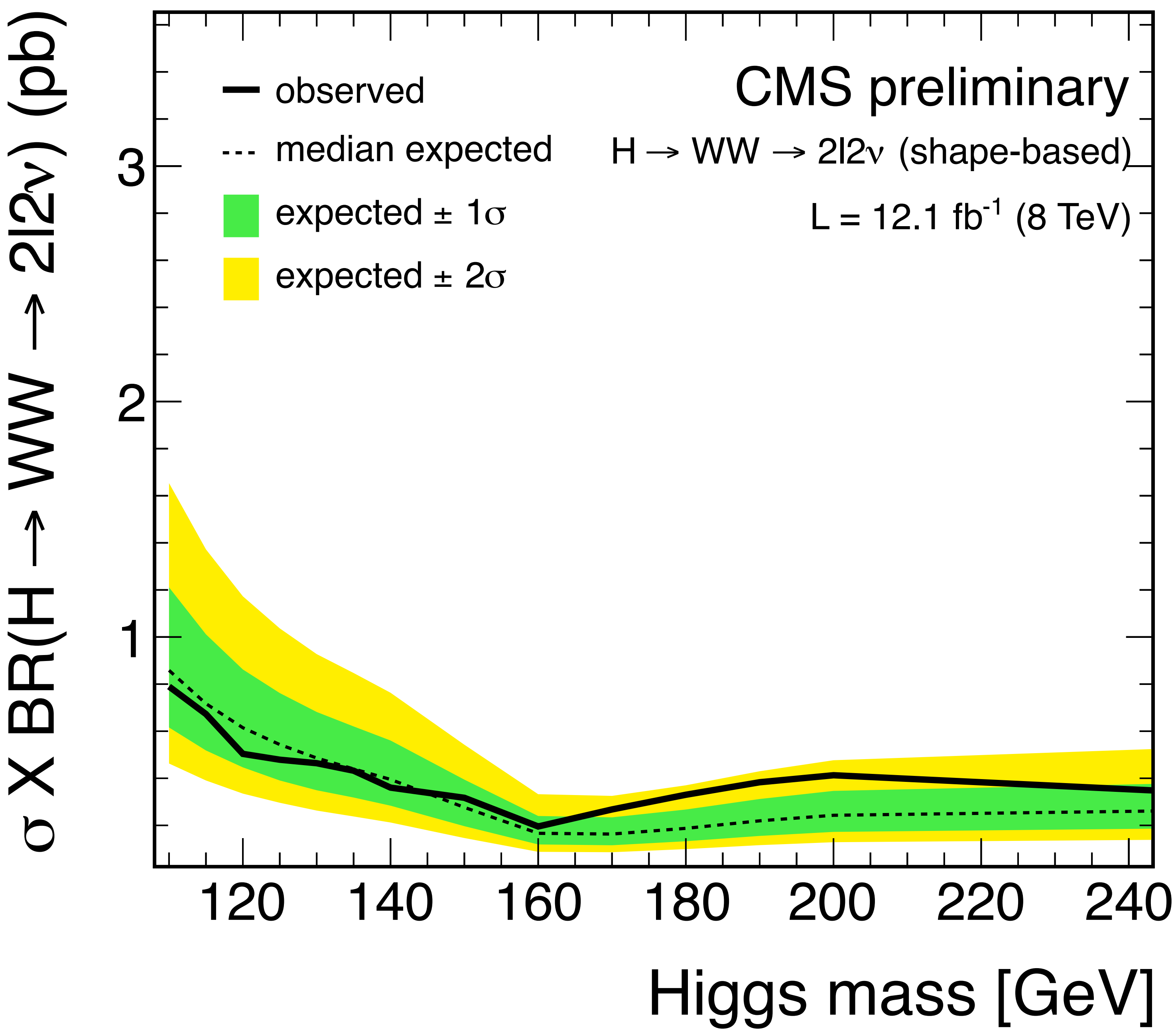


Final Results





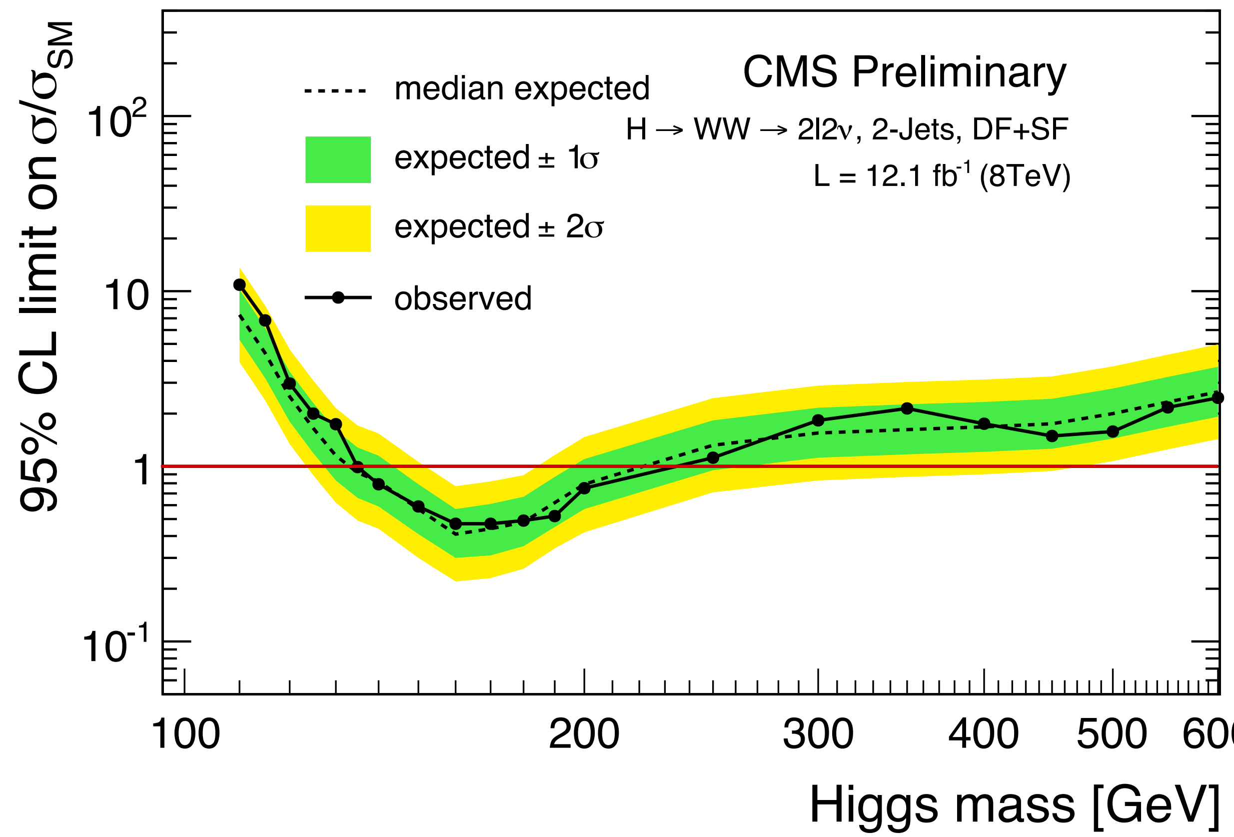
Search for Second "Higgs"



Considering 125GeV Higgs as background, we performed a search for second SM-like Higgs - no significant deviation from background only hypothesis

VBF

m_H	H $\to W^+W^-$	pp $\to W^+W^-$	WZ + ZZ $+Z/\gamma^* \to \ell^+\ell^-$	Top	W + jets	$W\gamma^{(*)}$	all bkg.	data
2-jet category $e\mu$ final state								
120	1.7 ± 0.2	0.8 ± 0.5	0.1 ± 0.0	0.9 ± 0.3	0.3 ± 0.2	0.1 ± 0.1	2.2 ± 0.6	2
125	2.8 ± 0.4	0.9 ± 0.5	0.1 ± 0.0	1.5 ± 0.5	0.3 ± 0.2	0.1 ± 0.1	2.9 ± 0.8	2
130	4.4 ± 0.6	1.3 ± 0.7	0.1 ± 0.0	1.6 ± 0.5	0.3 ± 0.2	0.1 ± 0.1	3.4 ± 0.9	4
160	11.7 ± 1.5	1.2 ± 0.6	0.0 ± 0.0	1.5 ± 0.5	0.0 ± 0.0	0.1 ± 0.1	2.9 ± 0.8	4
200	9.3 ± 1.2	2.5 ± 1.2	1.7 ± 1.6	4.6 ± 1.3	0.3 ± 0.4	0.0 ± 0.0	9.1 ± 2.4	8
400	3.9 ± 0.5	3.5 ± 2.2	1.7 ± 1.6	4.6 ± 1.3	0.0 ± 0.0	0.0 ± 0.0	9.8 ± 3.0	7
600	1.4 ± 0.2	1.6 ± 1.0	0.0 ± 0.0	1.9 ± 0.8	0.3 ± 0.2	0.0 ± 0.0	3.7 ± 1.3	3



- No evidence for Higgs in qq→H→WW final state yet
- Relative contribution of ggH and VBF
  - 0-jet: 99% ggF + 1% VBF
  - 1-jet: 89% ggF + 11% VBF
  - 2-jet: 21% ggF + 79% VBF
- Need more data

Summary

- CMS has successfully updated H→WW results using 5/fb (7TeV) and 12/fb (8TeV) datasets
- We developed a simpler and more intuitive 2D shape analysis
- The results are consistent with Standard Model Higgs boson hypothesis:
  - The expected/observed significance of the excess: 4.1σ/3.1σ
  - The signal strength is found to be $\mu = 0.74 \pm 0.25$
  - A search for second SM Higgs-like boson at low mass shows no significant deviation from a background only hypothesis
- We plan for a full analysis update with full 2012 dataset on a time scale of Moriond 2013 conference

Likelihood Scan for Mass