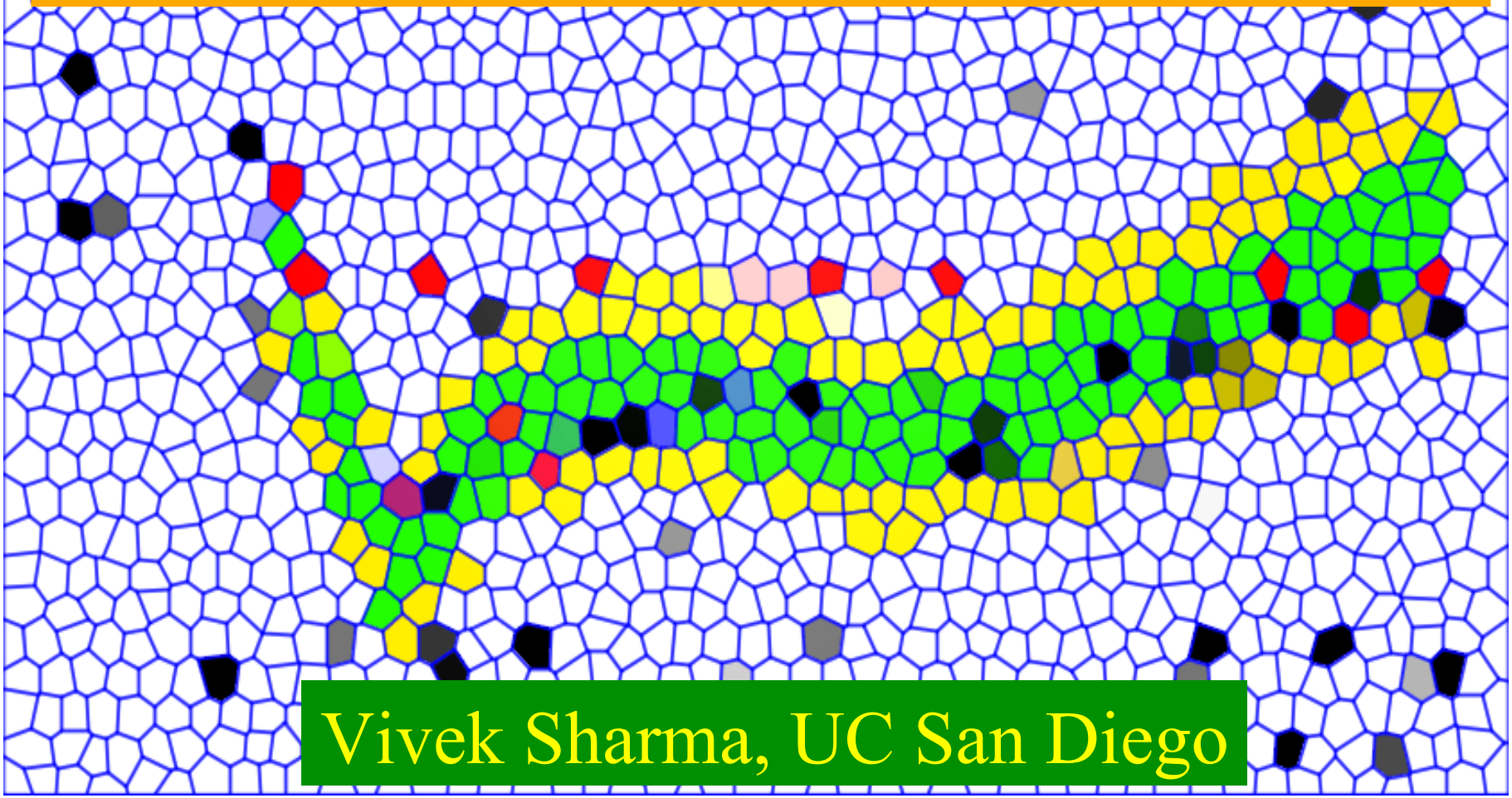
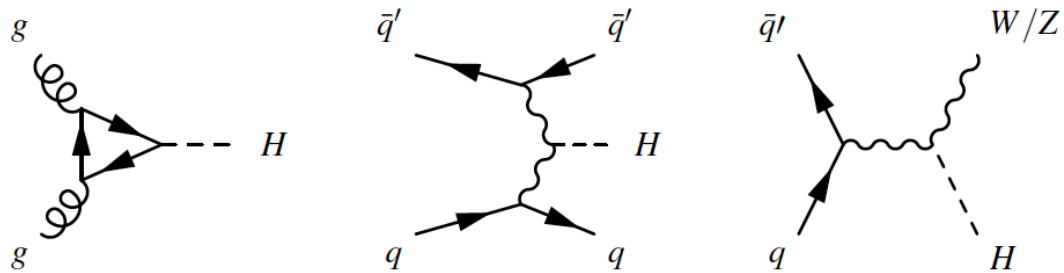
Search For The SM Higgs Boson With the CMS Detector

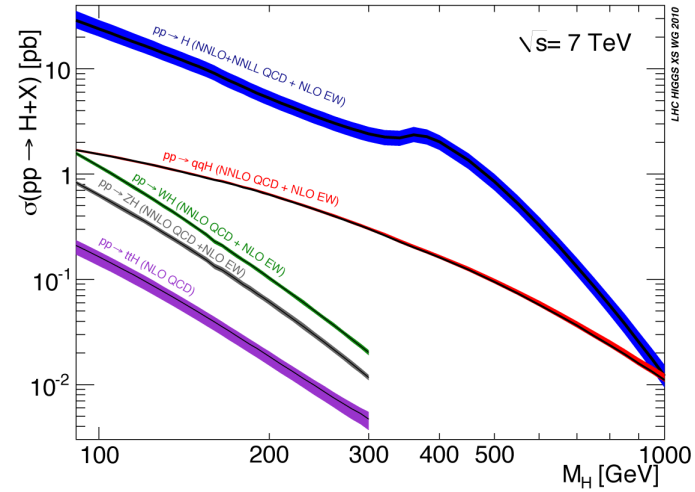


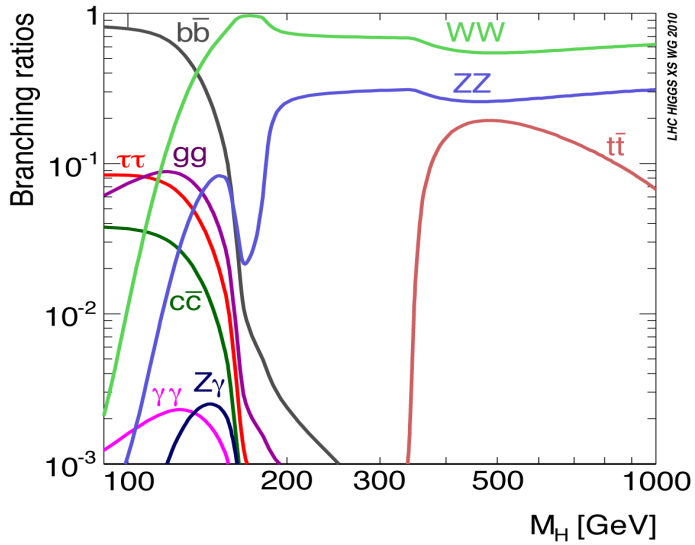
Vivek Sharma, UC San Diego

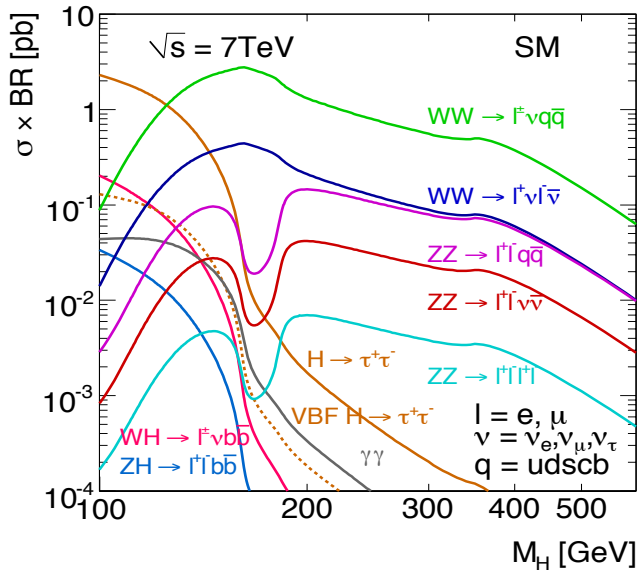
Portrait Of SM Higgs Production and Decay



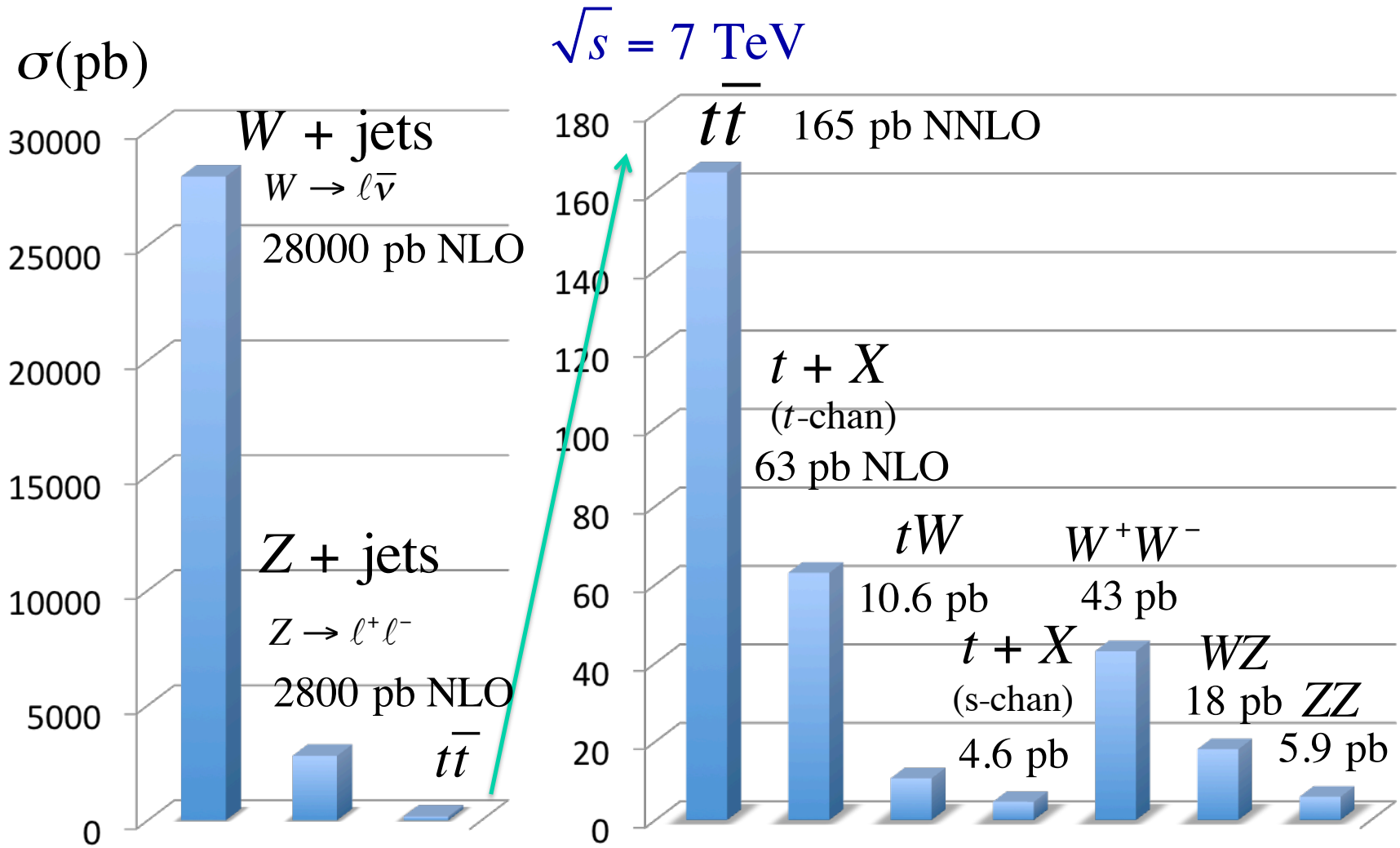
(a) $gg \to H$ (b) VBF (c) VH





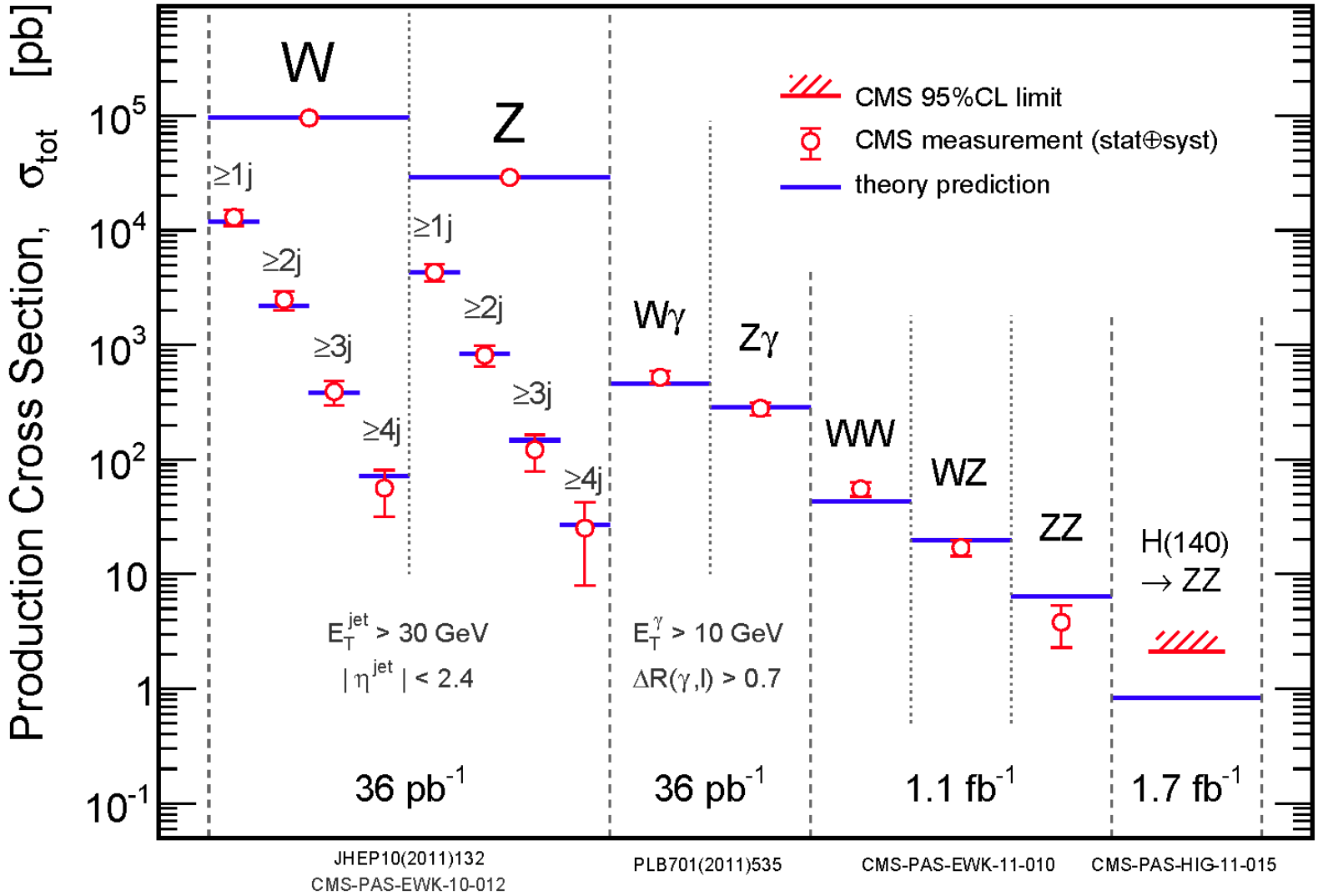


Cross Sections for Key SM Background Processes

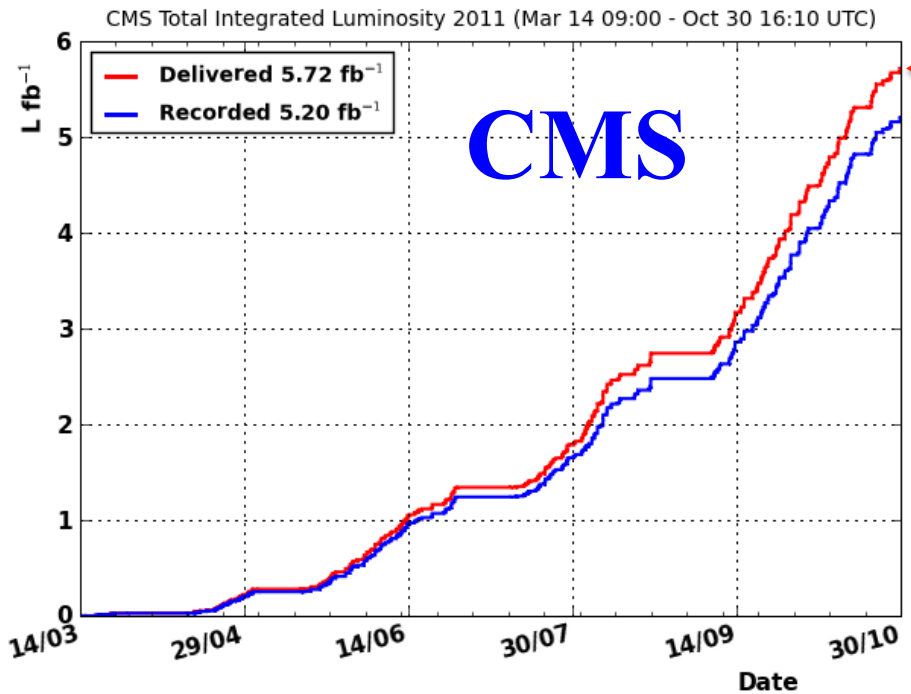


Strong CMS effort to measure these cross sections

Some Key Background Measurements With CMS



LHC And CMS Performance



>5 fb^{-1} delivered !
LHC target for 2011 was 1 fb^{-1} per exp.

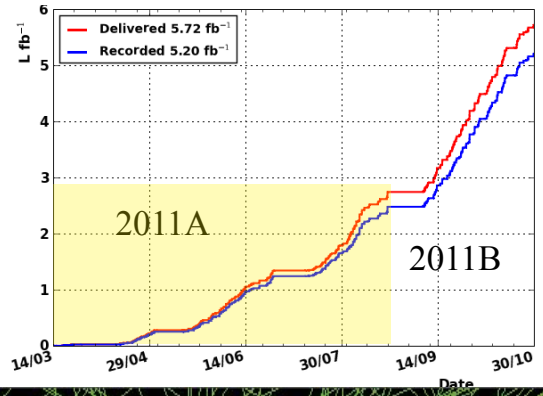
Max luminosity
$L = 3.5 \times 10^{33}\ cm^{-2}\ s^{-1}$

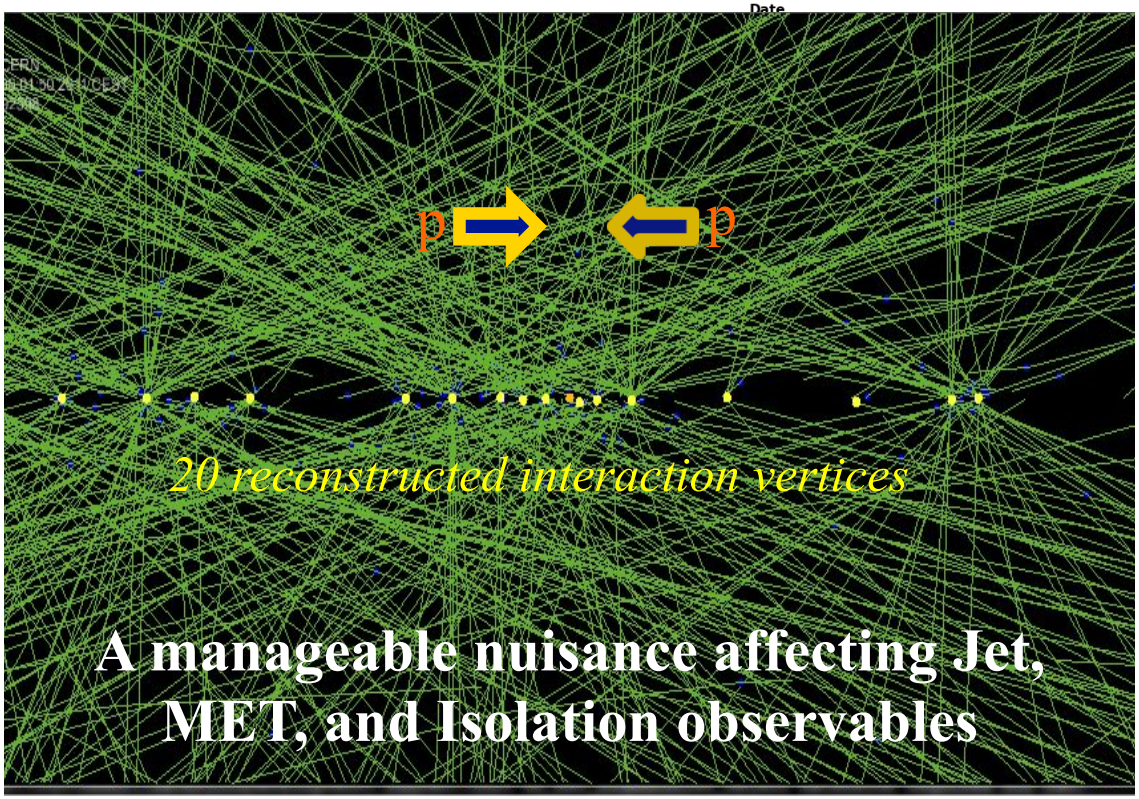
Rapid increase in instantaneous luminosity:
April ($L=2\times10^{32}\ cm^{-2}s^{-1}$) – October ($3.5\times10^{33}\ cm^{-2}s^{-1}$)

1 day in October'11 = more data than 4 x (entire 2010 run)

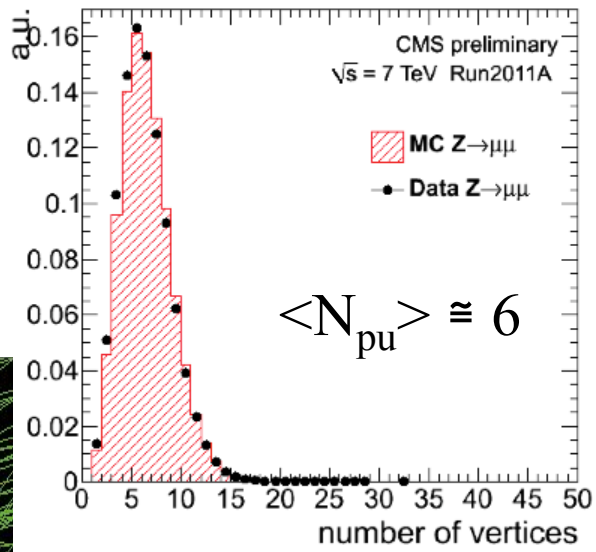
The challenge of trigger budget & pileup successfully met

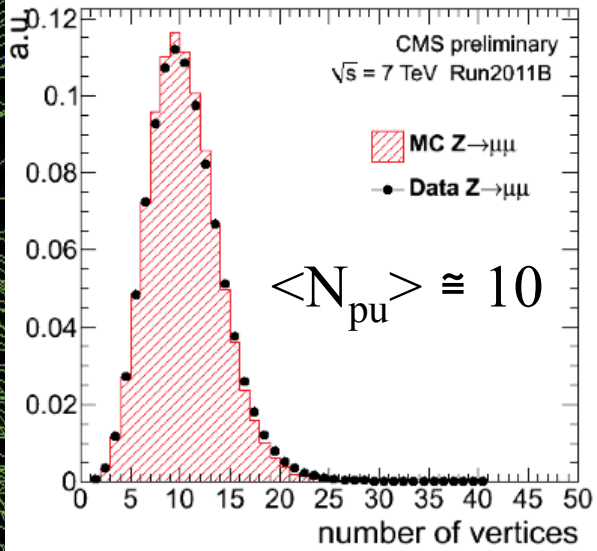
Pileup: A Key Feature in 2011 Data





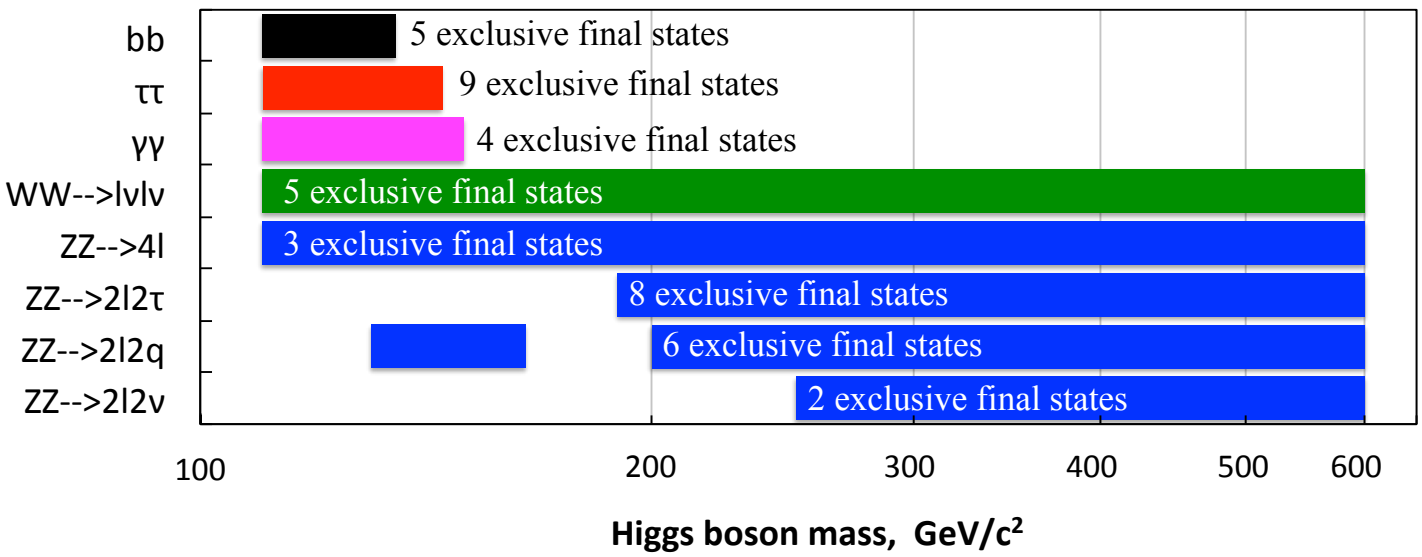
A manageable nuisance affecting Jet, MET, and Isolation observables



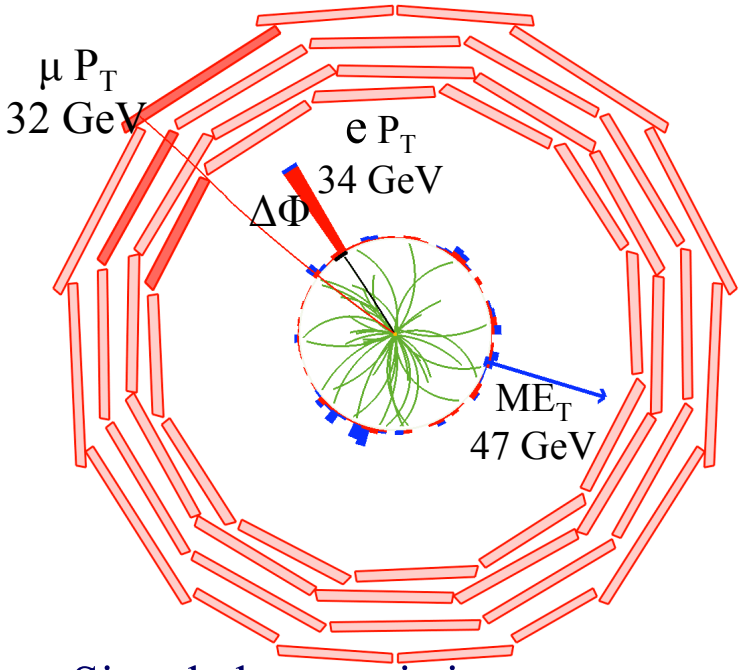


Search Channels For SM Higgs Boson

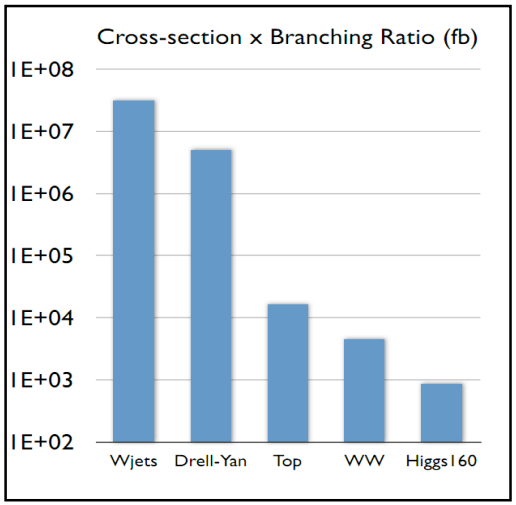
Mode	Mass Range	Data Used (fb^{-1})	Mass resolution	CMS Document
$H \rightarrow \gamma\gamma$	110-150	4.7	**1-3 %**	HIG-11-030
$H \rightarrow bb$	110-135	4.7	10 %	HIG-11-031
$H \rightarrow \tau\tau$	110-145	4.6	20 %	HIG-11-029
$H \rightarrow WW \rightarrow 2l\ 2\nu$	110-600	4.6	20%	HIG-11-024
$H \rightarrow ZZ \rightarrow 4l$	110-600	4.7	**1-2%**	HIG-11-025
$H \rightarrow ZZ \rightarrow 2l2\tau$	190-600	4.7	10-15%	HIG-11-028
$H \rightarrow ZZ \rightarrow 2l2j$	130-165/200-600	4.6	3%	HIG-11-027
$H \rightarrow ZZ \rightarrow 2l2\nu$	250-600	4.6	7%	HIG-11-026



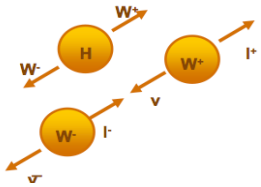
H→ WW → 2l 2ν



- No signal mass peak (missing νν) ➔ Counting expt.
- Challenge is to remove & control large backgrounds



- Signal characteristics:
  - Only 2 opposite sign, isolated leptons
  - significant ME_T ➔ No mass peak
  - No b-jets, no additional low P_T μ
  - With additional 0, 1 or 2 jets (VBF)
  - Small ΔΦ (l^+l^-) ← Higgs scalarity

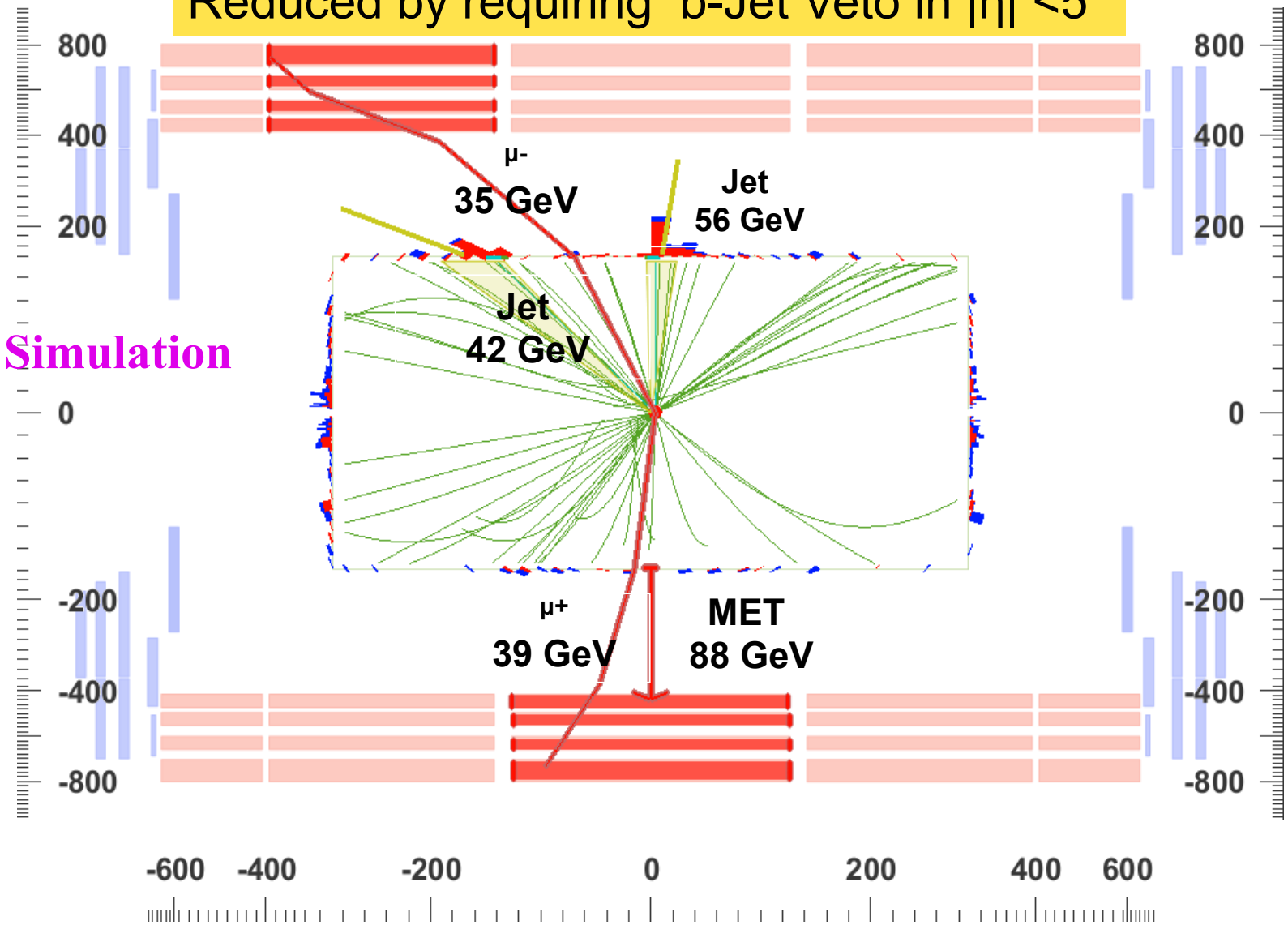


Major requirements:

- Lepton P_t > 15 GeV, tight ID & Isolation
  - removes QCD & W+jets contamination
- Large ME_T & Z → μμ, ee veto
  - removes Drell-Yan contamination
- Classification by # of jets (P_T > 30 GeV) & b-jet veto
  - removes Top contamination
- Kinematic discriminants: M_{ll} & ΔΦ (l^+l^-)
  - mitigates pp → ww background
- M_H–dependent cut optimization

Major Background: ttbar

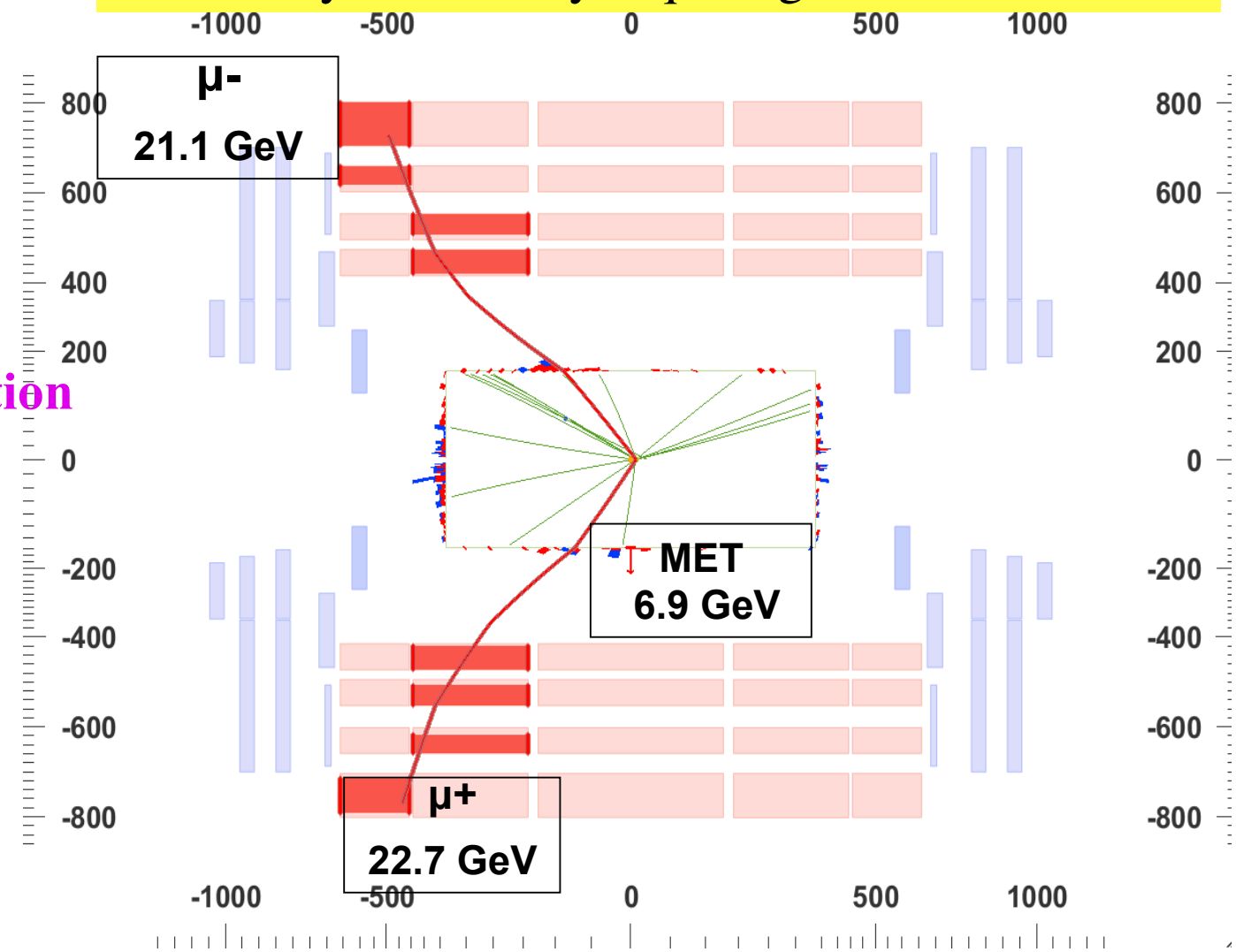
Reduced by requiring b-Jet Veto in |η| <5



Major Background: Drell-Yan

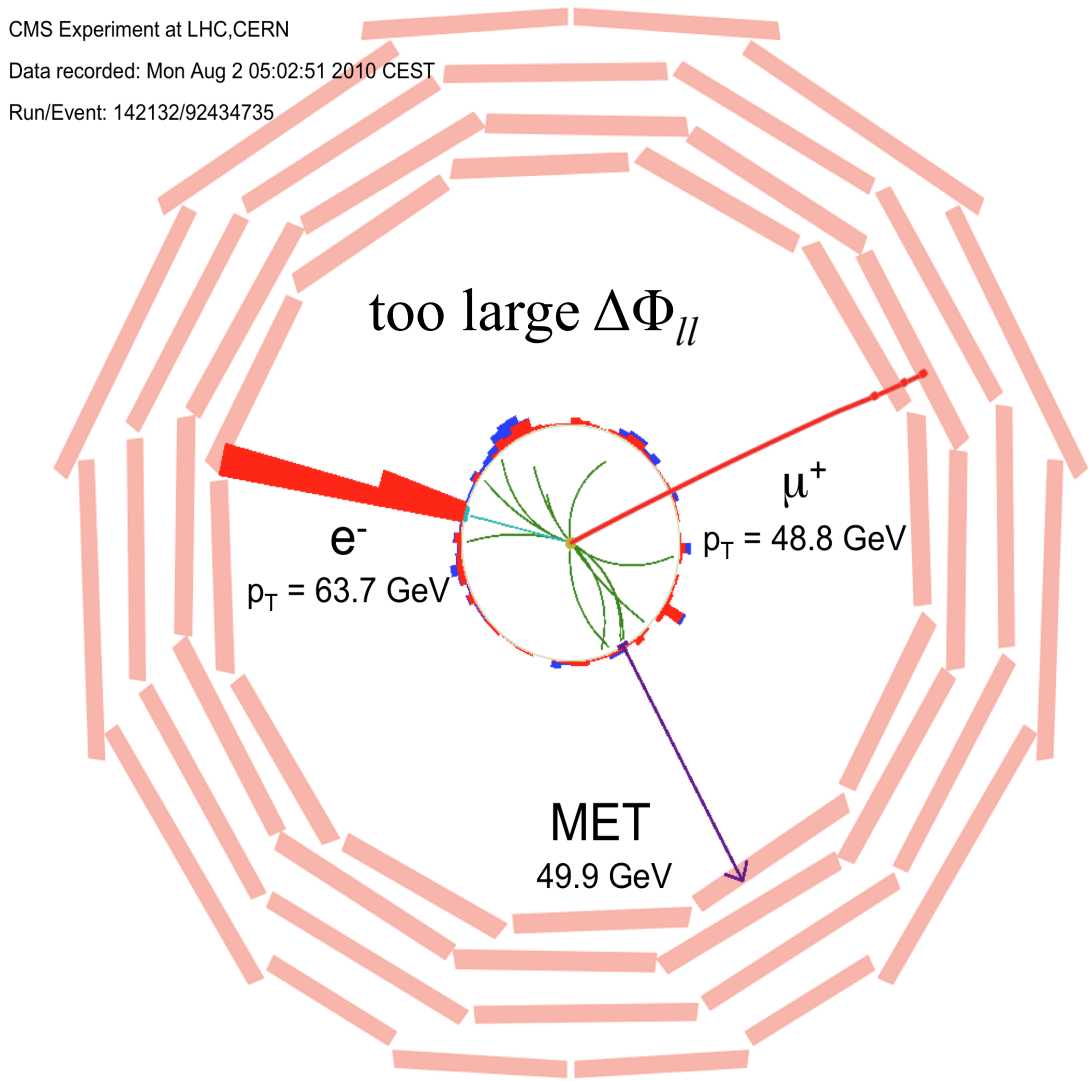
Drastically reduced by requiring MET in the event

Simulation



pp → WW Is The Major Irreducible Background

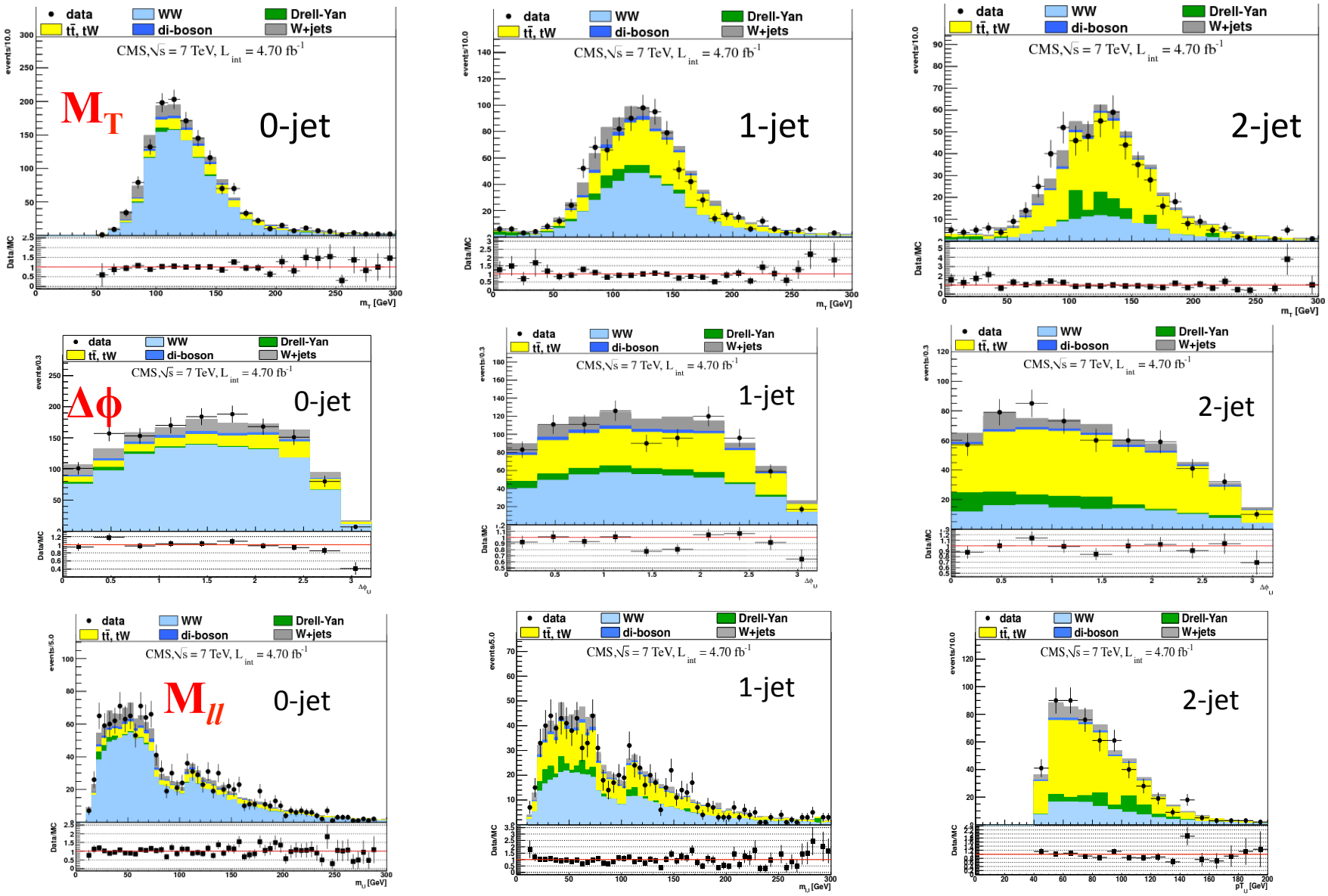
Data



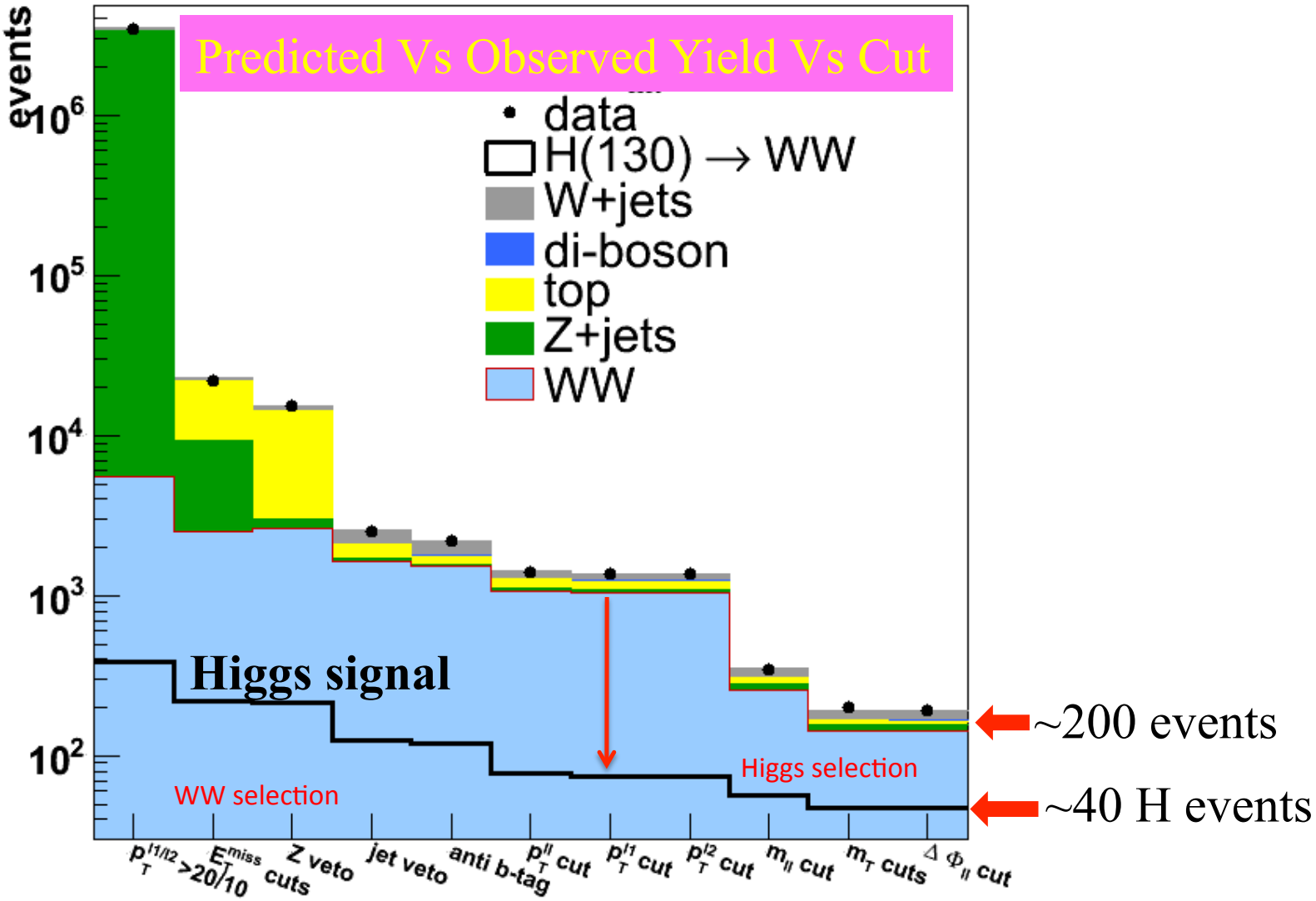
H→ WW → 2l 2ν: Improvements Since LP'11

- Selection: Focus on PU mitigation and extend low mass reach
  - Pileup dependent MET cut
  - Minimum dilepton mass cut from 12→20 GeV for same-flavor events (ee, μμ)
  - Refined top tagging procedure (less sensitive to PU)
  - Trailing lepton p_T from 10 → 15 GeV for SF and dilepton p_T cut at 45 GeV (suppressing Drell-Yan and W+jets)
- Backgrounds:
  - Refined DY and top estimation
  - Data driven normalization of DY→ττ and Wγ*
- **Major effort in understanding BDT Shape based analysis:**
  - **Shape variation for systematic uncertainty evaluation**
  - **Additional cross-checks: single variable shape analyses; matrix element analysis**

Distributions At WW Selection level



Cut Flow Evolution (0-jet Bin As An Example)

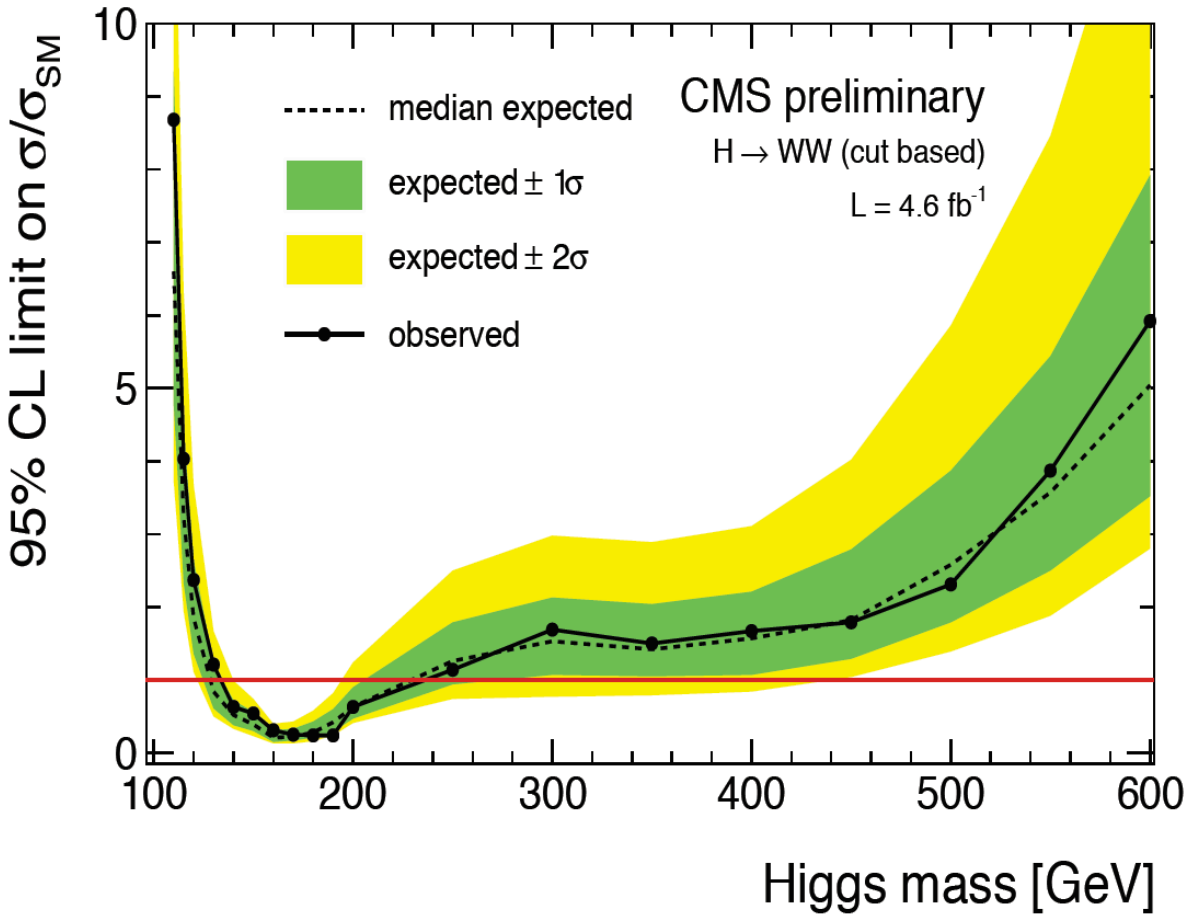


Data describes predicted background well, no excess

Event Yield Tally For Cut Based Analysis

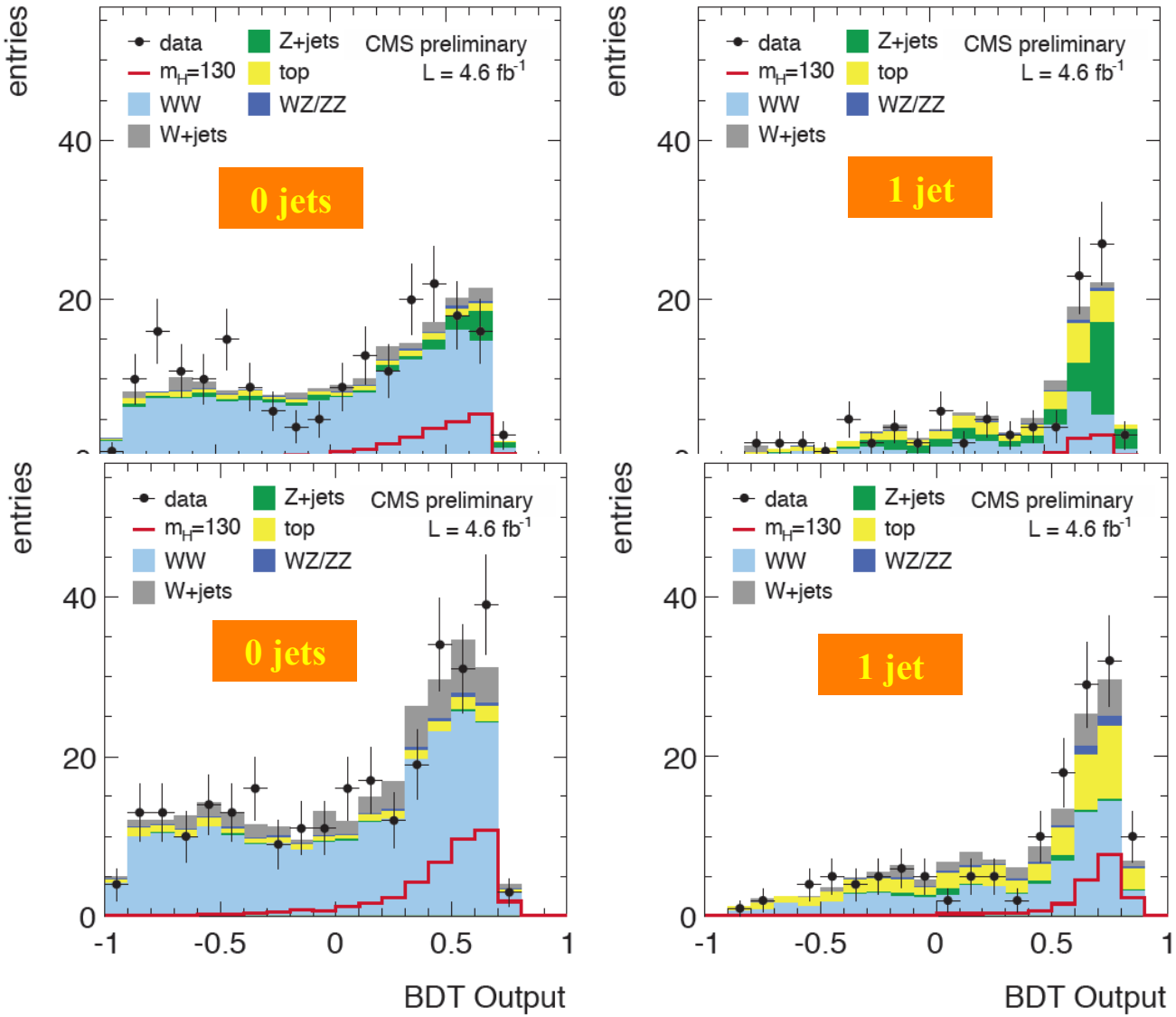
m_H	$Z/\gamma^* \to \ell^+\ell^-$	top	W + jets	WZ + ZZ + $W\gamma$	pp→ W^+W^-	all bkg.	$H \to W^+W^-$	data
cut-based approach 0-jet category								
120	8.8 ± 9.2	6.7 ± 1.0	14.7 ± 4.7	6.1 ± 1.5	100.3 ± 7.2	136.7 ± 12.7	15.7 ± 0.8	136
130	13.7 ± 7.8	10.6 ± 1.6	17.6 ± 5.5	7.4 ± 1.6	142.2 ± 10.0	191.5 ± 14.0	45.2 ± 2.1	193
160	3.4 ± 3.4	10.5 ± 1.4	3.0 ± 1.5	2.2 ± 0.4	82.6 ± 5.4	101.7 ± 6.8	122.9 ± 5.6	111
200	2.7 ± 3.7	23.3 ± 3.1	3.4 ± 1.5	3.2 ± 0.3	108.2 ± 4.5	140.8 ± 6.8	48.8 ± 2.2	159
250	0.3 ± 0.6	36.2 ± 4.8	6.7 ± 2.1	5.7 ± 0.7	101.8 ± 4.5	150.8 ± 6.9	23.5 ± 1.1	152
300	0.7 ± 1.9	41.6 ± 5.4	6.5 ± 2.1	7.0 ± 0.7	87.5 ± 3.9	143.3 ± 7.2	20.2 ± 0.9	147
400	0.2 ± 0.2	35.9 ± 4.7	5.5 ± 1.8	9.3 ± 1.1	59.8 ± 2.7	110.8 ± 5.8	17.5 ± 0.8	109
cut-based approach 1-jet category								
120	6.6 ± 2.3	17.2 ± 1.0	5.4 ± 2.4	3.2 ± 0.6	27.0 ± 4.7	59.5 ± 5.9	6.5 ± 0.3	72
130	5.3 ± 2.5	25.6 ± 1.4	6.5 ± 2.5	4.0 ± 0.6	38.5 ± 6.6	79.9 ± 7.7	17.6 ± 0.8	105
160	4.2 ± 1.4	27.9 ± 1.4	3.2 ± 1.4	1.9 ± 0.3	33.7 ± 5.5	70.8 ± 6.0	60.2 ± 2.6	86
200	14.6 ± 5.3	59.4 ± 2.8	5.2 ± 1.8	2.2 ± 0.1	49.3 ± 2.2	130.8 ± 6.7	25.8 ± 1.1	111
250	12.9 ± 6.8	83.8 ± 3.9	5.9 ± 2.1	3.3 ± 0.2	60.3 ± 2.8	166.2 ± 8.6	14.8 ± 0.6	158
300	12.8 ± 4.8	83.6 ± 3.9	6.2 ± 2.2	3.8 ± 0.4	57.5 ± 2.7	163.9 ± 7.1	13.7 ± 0.6	168
400	8.3 ± 3.2	60.6 ± 2.9	6.2 ± 2.1	3.9 ± 0.5	44.6 ± 2.2	123.6 ± 5.3	12.2 ± 0.5	128

Limits From H→ WW → 2l 2ν: Cut Based



Expected 95% CL limit: $129 < M_H < 236$ GeV
Observed 95% CL limit: $132 < M_H < 238$ GeV

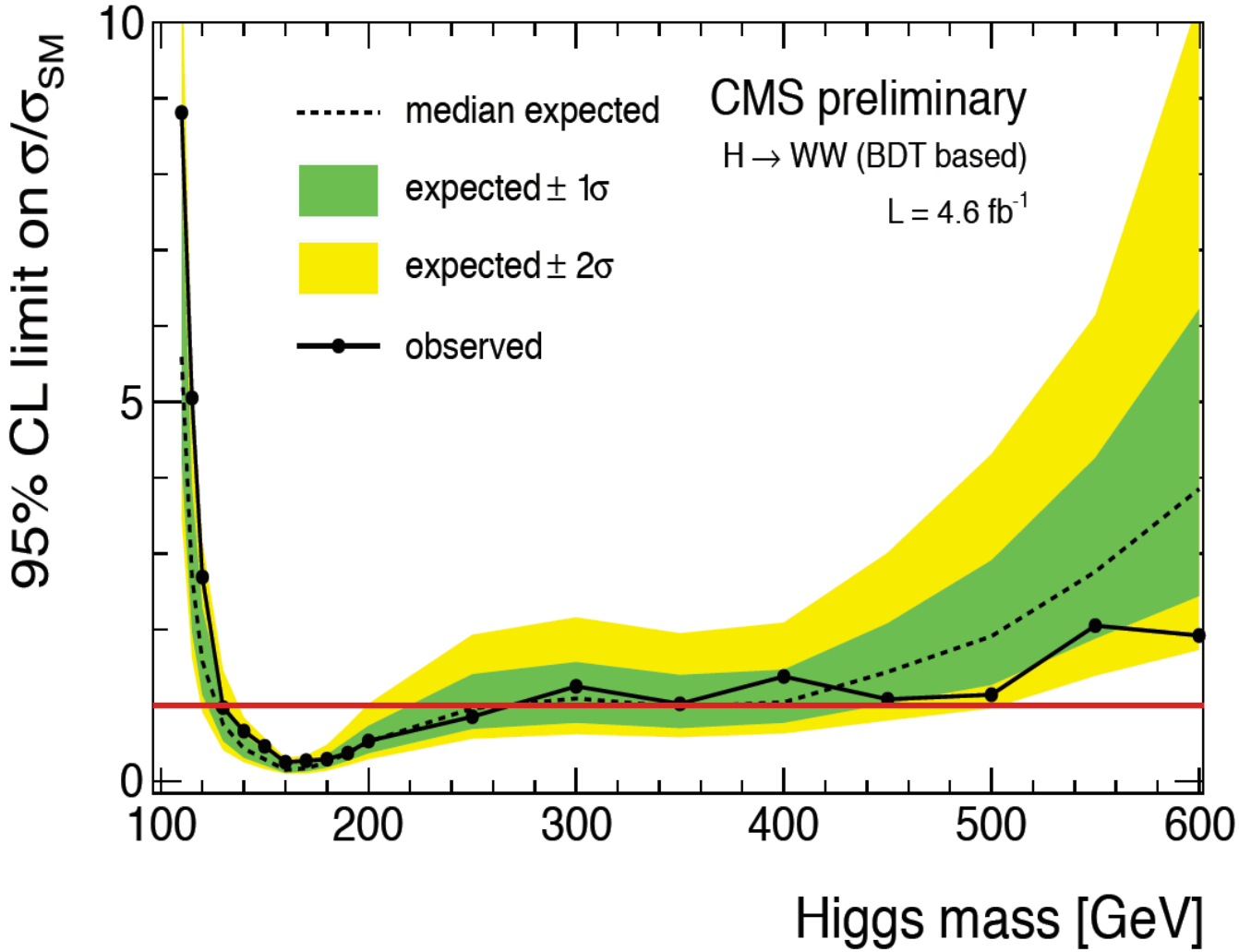
Multivariate (BDT) Classifier



ee
+μμ

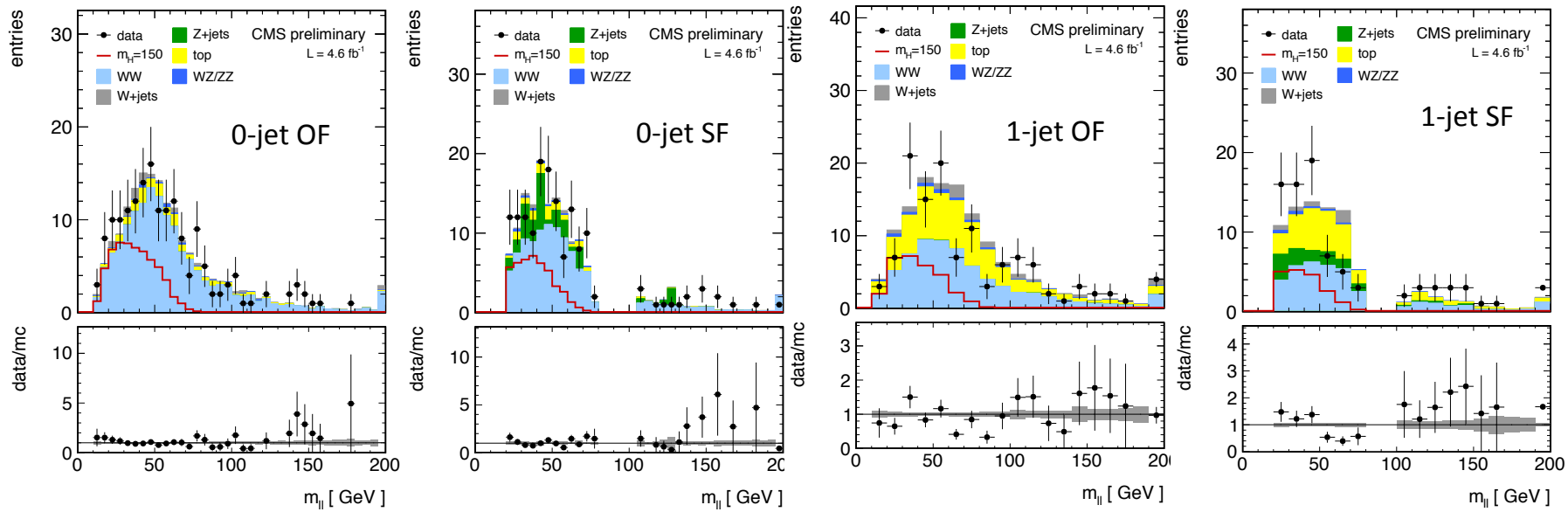
eμ

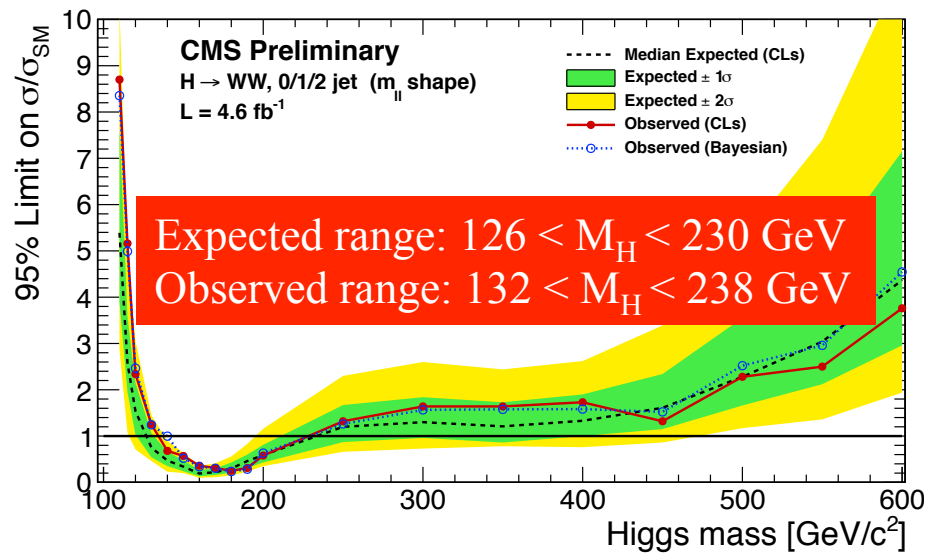
Limits From H→ WW → 2l 2ν: BDT Analysis



Expected limit: $127 < M_H < 270$ GeV
Observed limit: $129 < M_H < 270$ GeV

Cross Check: Single Variable (M_{ll}) Shape Analysis

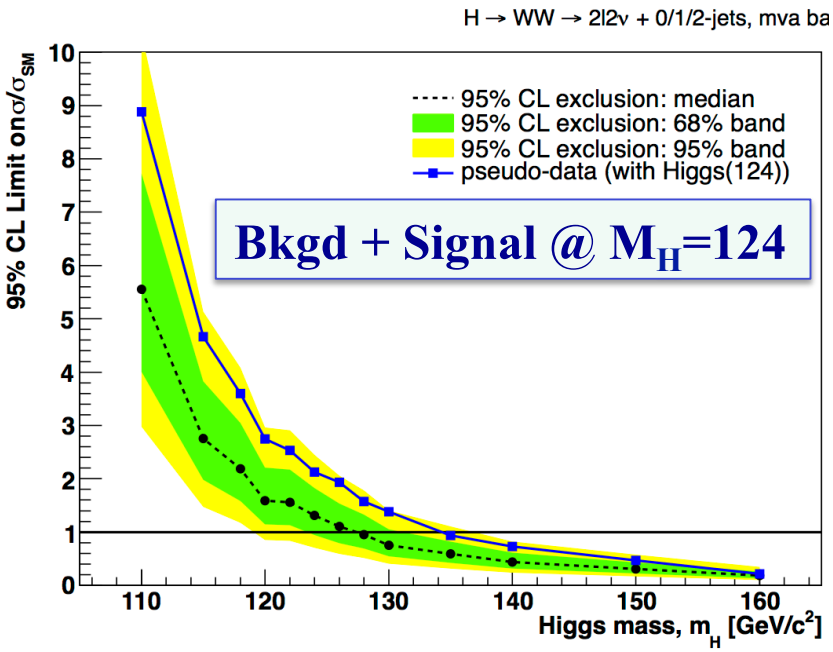


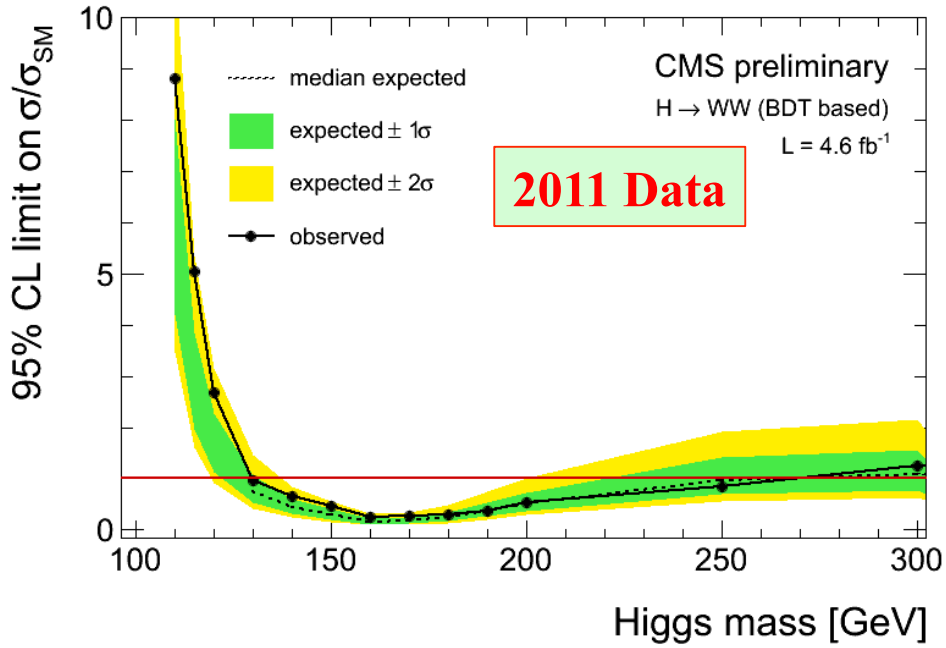


Result consistent With BDT

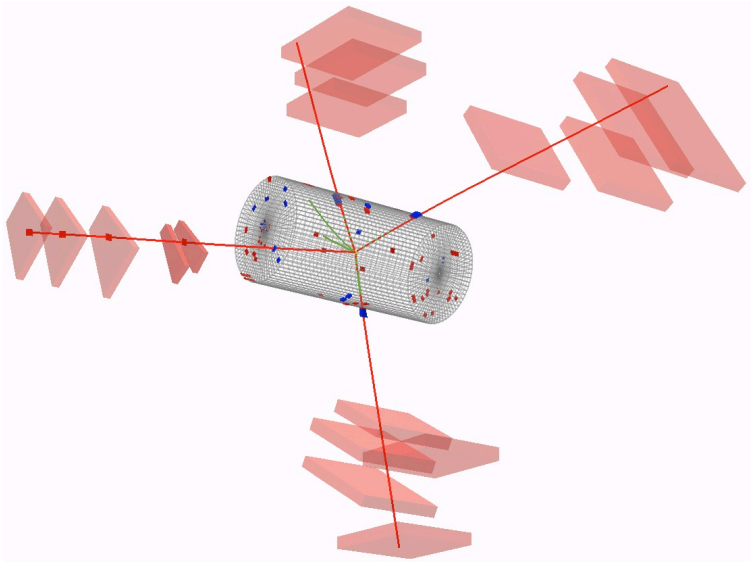
H→ WW → 2l 2ν: What To Expect For M_H=124

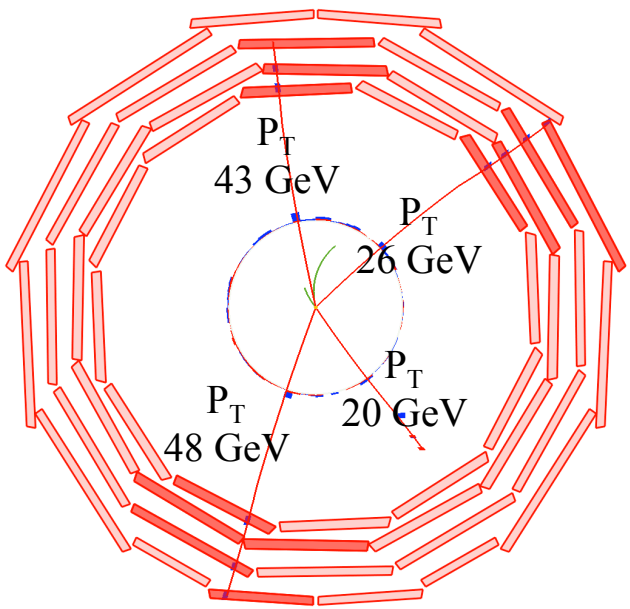
Inject a SM like Higgs signal over predicted backgrounds





H → ZZ → 4e, 4μ, 2e 2μ : The Golden Channels

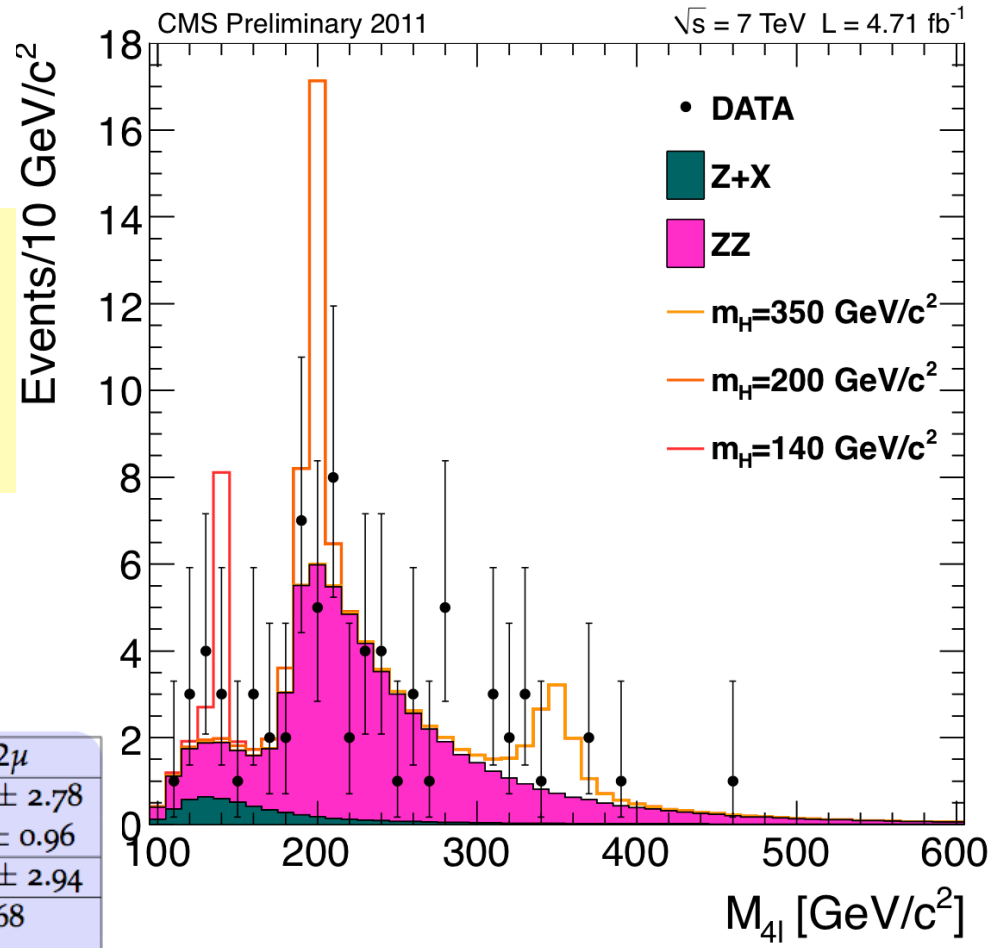




- Signal: 4 isolated lepton from common vertex
- Fully reconstructed, Mass resolution ~ 1%
- Reducible backgrounds:
  - ttbar → 2l2υ2b ; Z+bb
  - Removed by Isolation & Impact parameter requirements
- <u>Irreducible background: pp → ZZ Continuum</u>
- Event Selection: Same Flavor, opposite charge
  - Z_1 : $P_T(\min) > 10$, $P_T(\max) > 20$ GeV, $50 < M_{ll} < 120$
  - Z_2: $12 < M_{ll} < 120$ GeV
  - $M_{4l} > 120$ GeV
  - Impact parameter significance > 4
- Reducible background contribution from data
- **ZZ Continuum:**
  - Shape known at NLO, corrected for gg → ZZ →4l evaluated with MCFM
  - Rate obtained from Z yield in data & theoretical prediction for ratio of ZZ to Z cross sections

H → ZZ → 4l : Expected & Observed Yields

M_{4l} > 100 GeV/c^2
Observed events: 72
Expected events: 67.1 ± 6.0

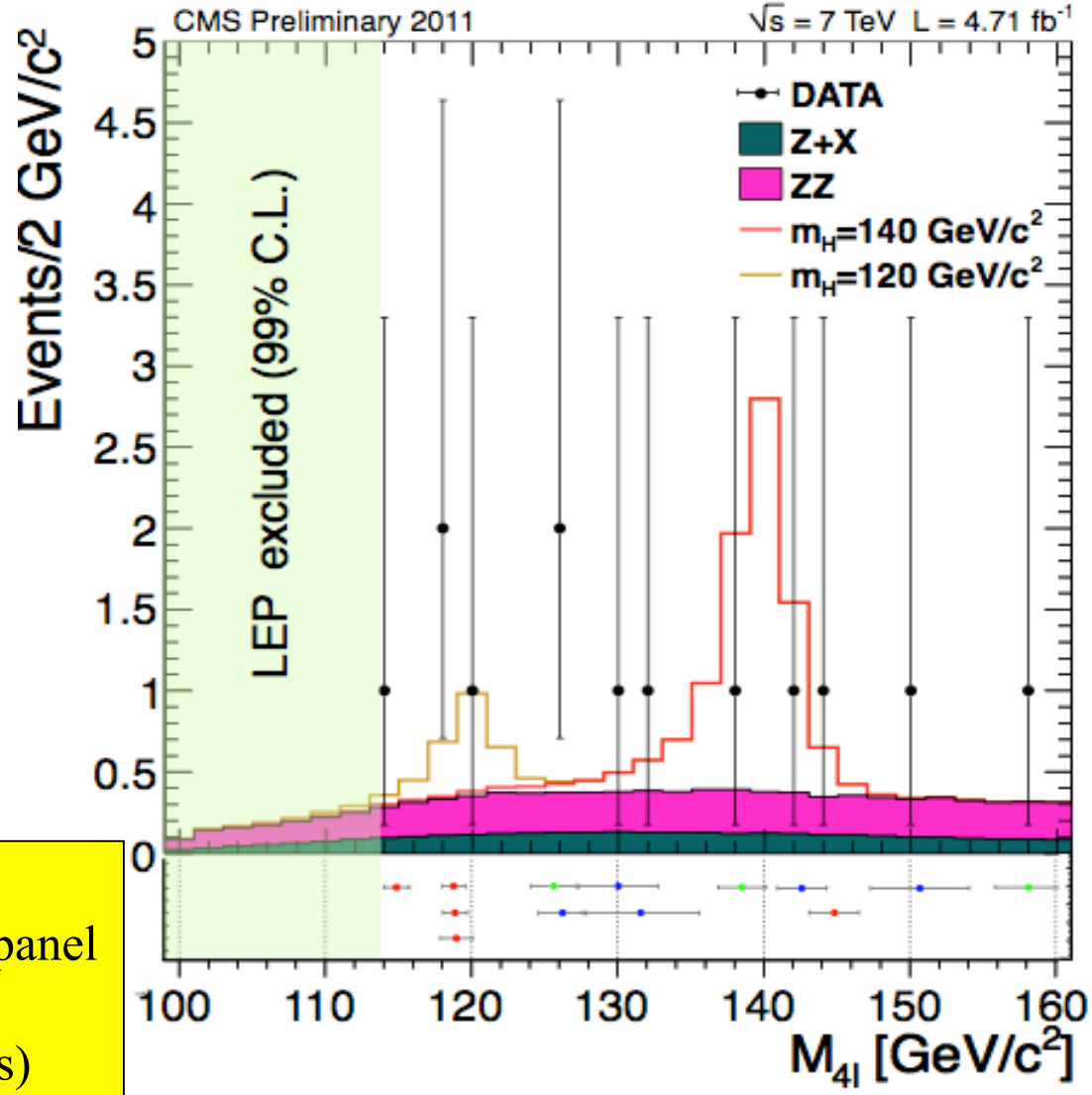


Baseline	$4e$	4μ	$2e2\mu$
ZZ	12.27 ± 1.16	19.11 ± 1.75	30.25 ± 2.78
Z+X	1.67 ± 0.55	1.13 ± 0.55	2.71 ± 0.96
All background	13.94 ± 1.28	20.24 ± 1.83	32.96 ± 2.94
$m_H = 120$ GeV/c^2	0.25	0.62	0.68
$m_H = 140$ GeV/c^2	1.32	2.48	3.37
$m_H = 350$ GeV/c^2	1.95	2.61	4.64
Observed	12	23	37

H → ZZ → 4l : Zoom Of Low Mass Range

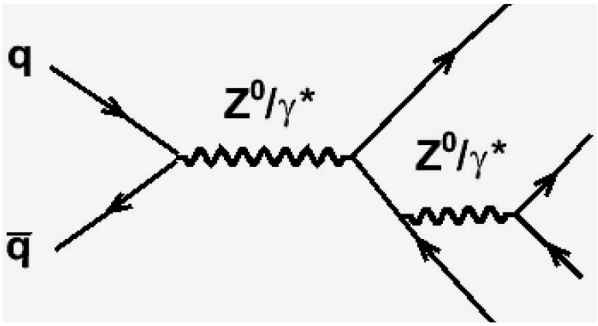
100 < M_{4l} < 160 GeV/c^2
Obs. events: 13
Exp. events: 9.5 ± 1.3

Final state:	4e	4μ	2e2μ
Obs. events:	3	5	5
Exp. events:	1.7	3.3	4.5

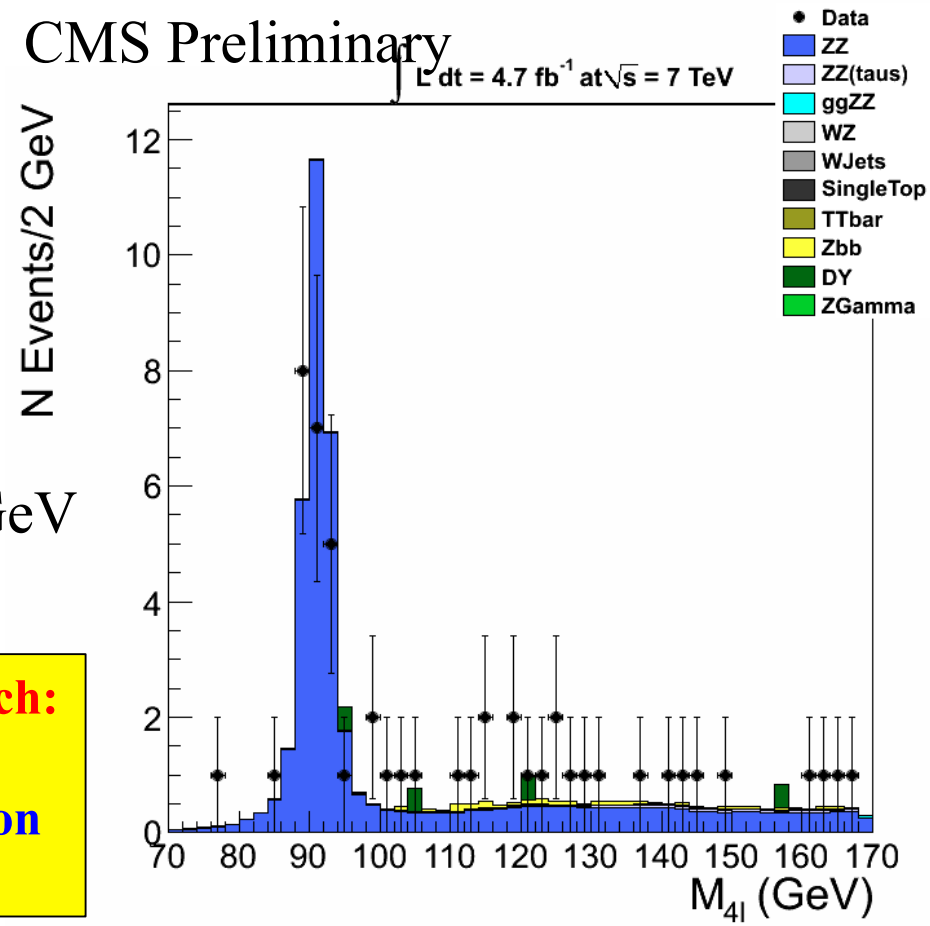


Note:
- unbinned events in the bottom panel
- **4e**, **4μ**, **2e2μ**
- Event-by-event mass error (bars)

Along The Way: Observation of Z → 4l



CMS Preliminary

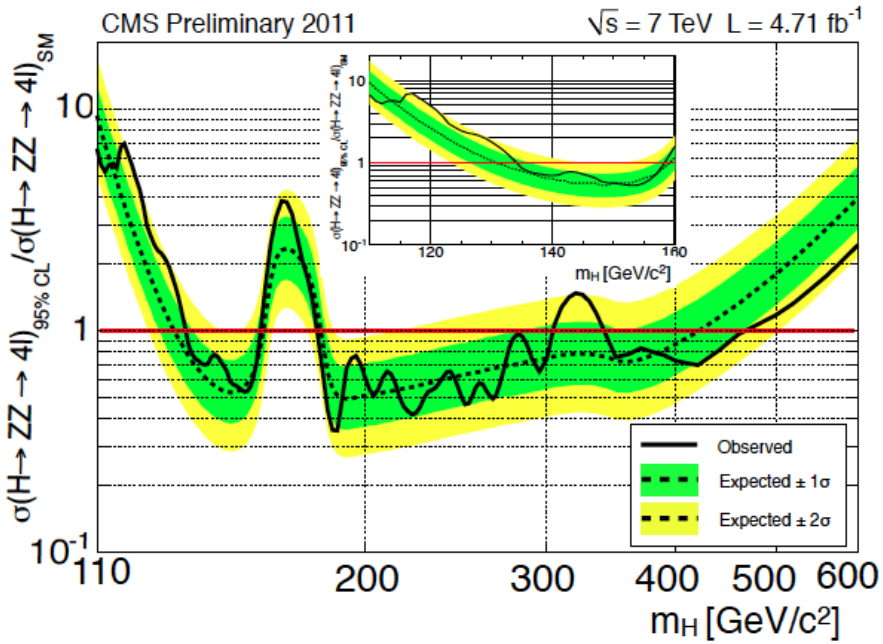


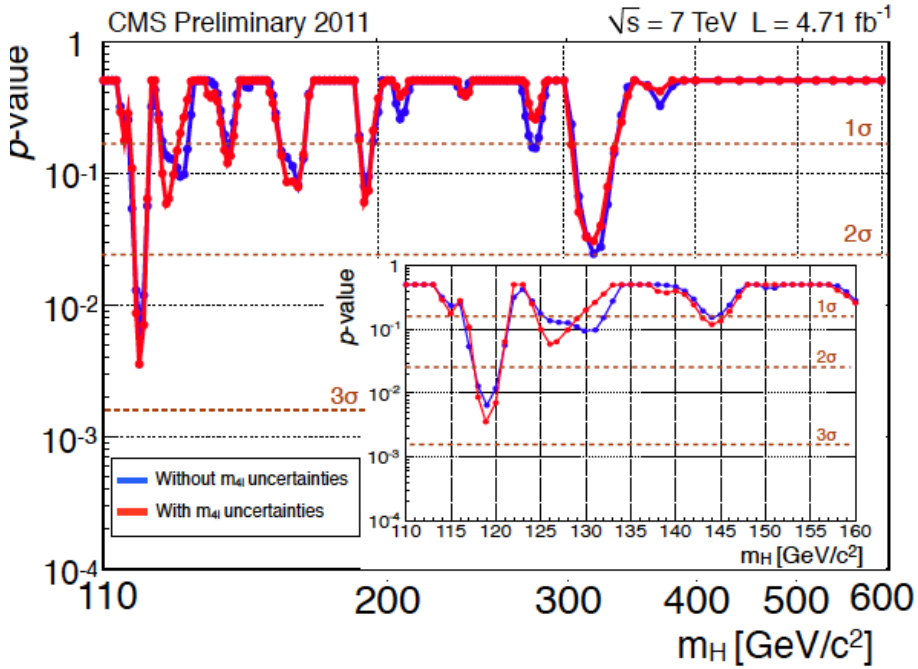
Di-lepton mass cut relaxed to 4 GeV

Standard Candle for H→ZZ→4l search:

- **direct calibration of M_{4l} scale**
- **direct measurement of M_{4l} resolution**
- **M_{4l}=91.3±0.6 GeV (stat)**

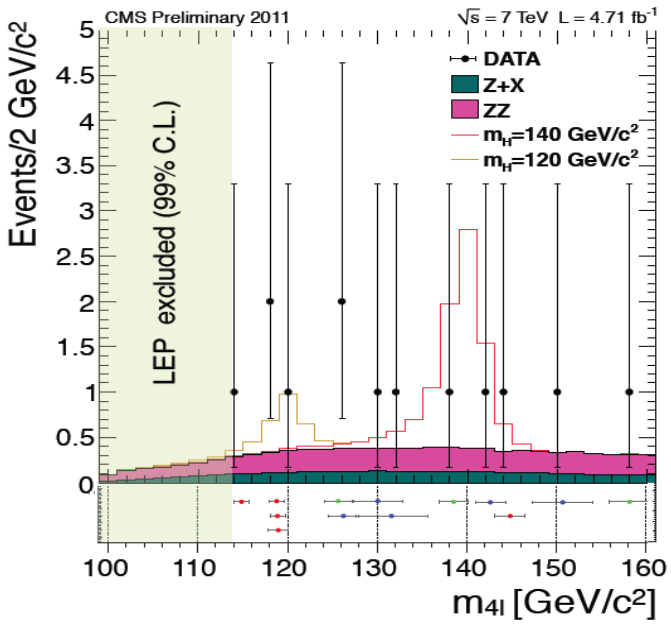
H → ZZ → 4l : Results

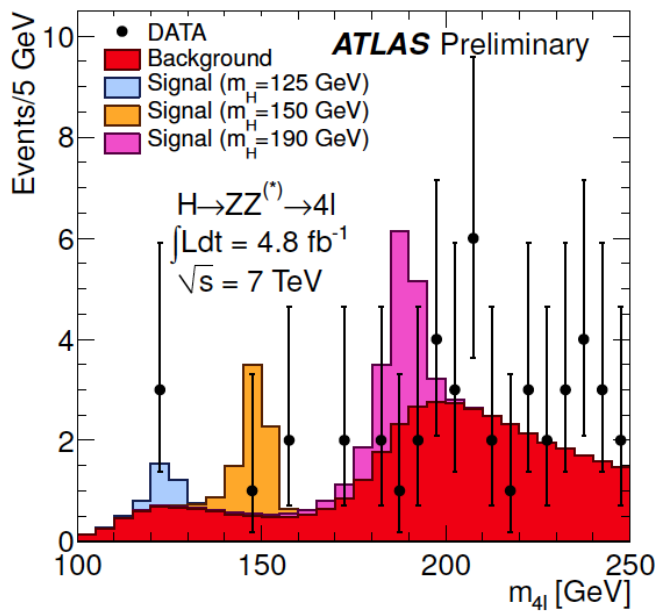




LEE trials factor ≈ 40 for the full mass range
Hence, 2σ deviations are not very significant

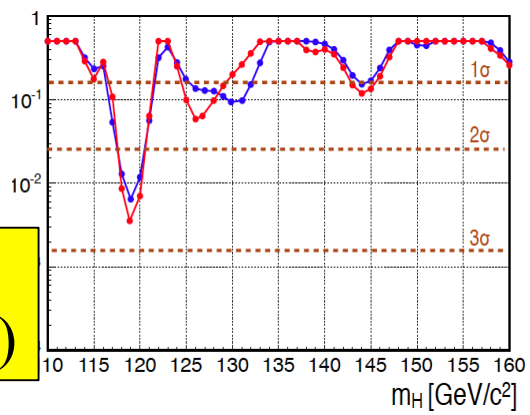
H → ZZ → 4l : Comparing CMS & ATLAS

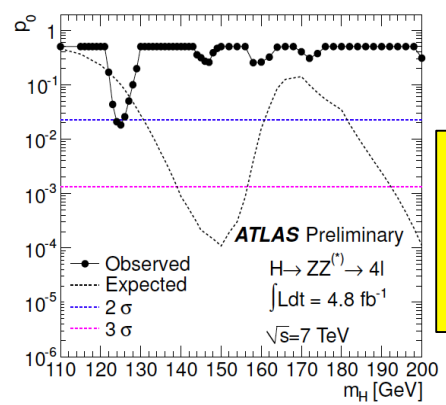




M=118.8 (4μ)
M=118.9 (4μ)
M=119.0 (4μ)

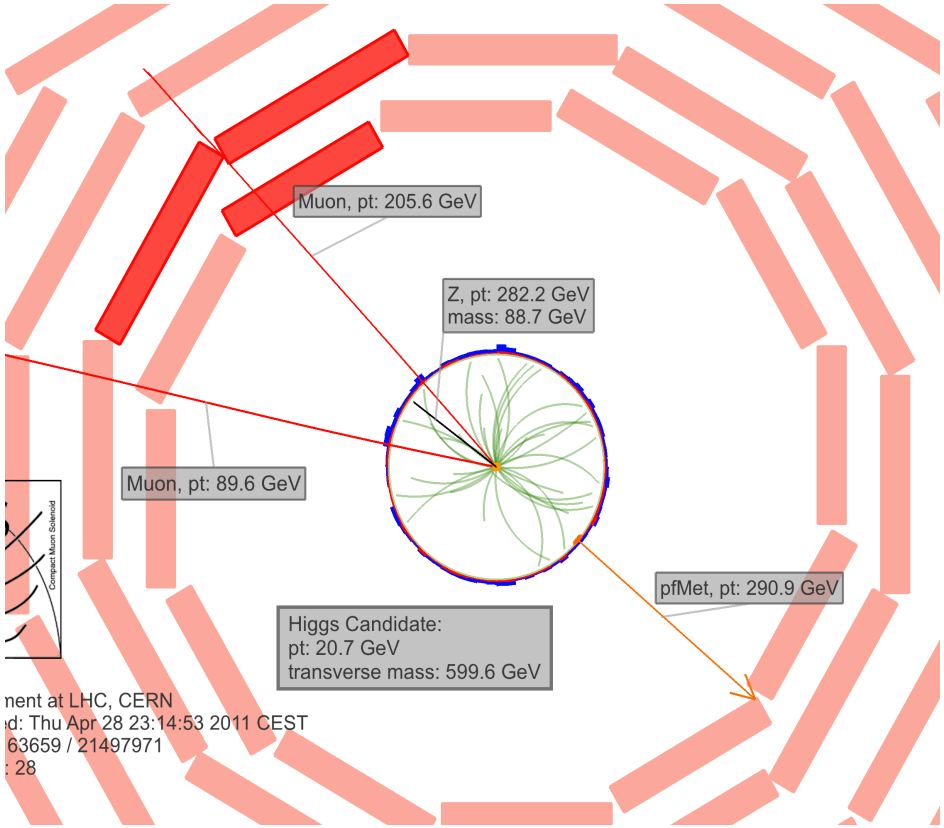
M=125.7 (4e)
M=126.2 (2e2μ)





M=123.6 (2μ2e)
M=124.3 (2e2μ)
M=124.6 (4μ)

High Mass Higgs: H → ZZ → 2l 2ν



M_T= 600 GeV

Event Selection:

- 2 isolated leptons: Z(2e), Z(2μ)
- no impact parameter
- $p_T(ll) > 55$ GeV (large)
- large MET, not aligned with jets or leptons
- **Final discriminant: <u>M_T shape</u>**

$$M_T{}^2 = (\sqrt{P_{TZ}{}^2 + M_Z{}^2} + \sqrt{MET^2 + M_Z{}^2})^2 - (\vec{P_{TZ}} + \vec{MET})^2$$

Mass resolution: 7%

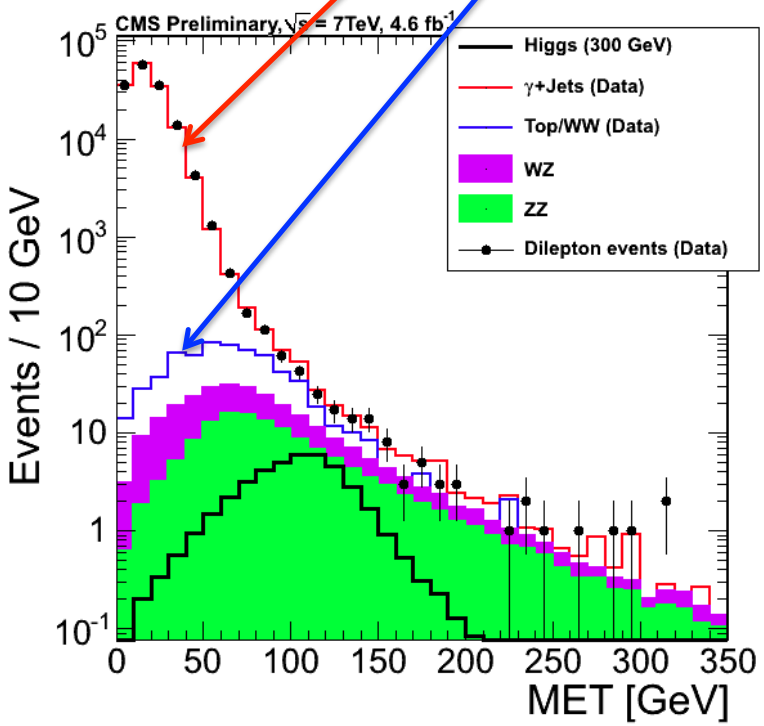
Main backgrounds:

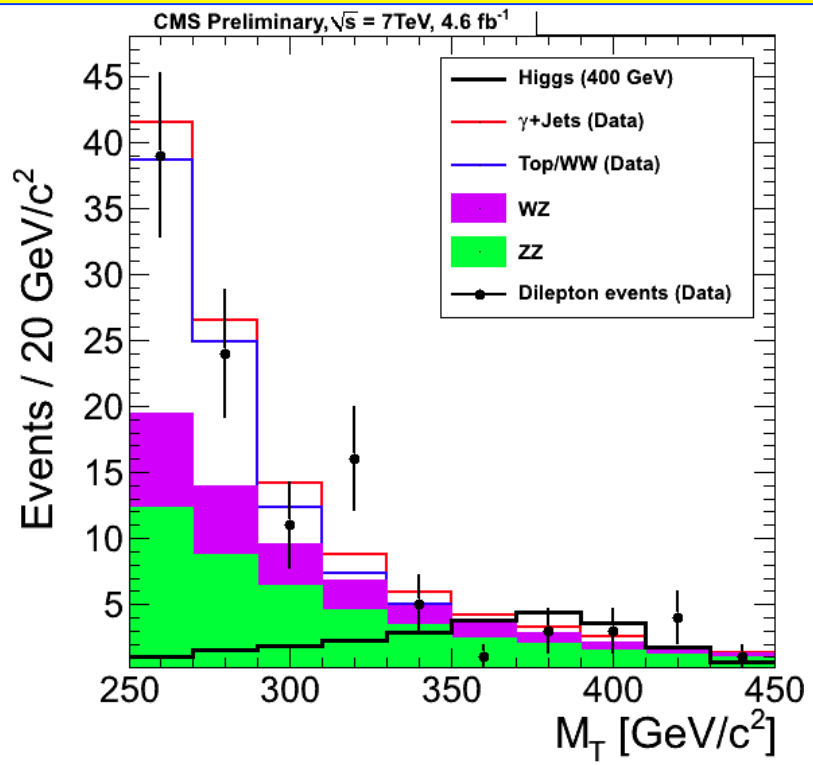
- Z+jets (sig:bkgd = $1:10^5$): from data (γ+jets)
- tt, WW, Wjets: from data (off Z-peak)
- ZZ, WZ: from MC

H → ZZ → 2l 2ν: Distributions

Data-driven prediction for Z+jets bkgd, using γ+jets events

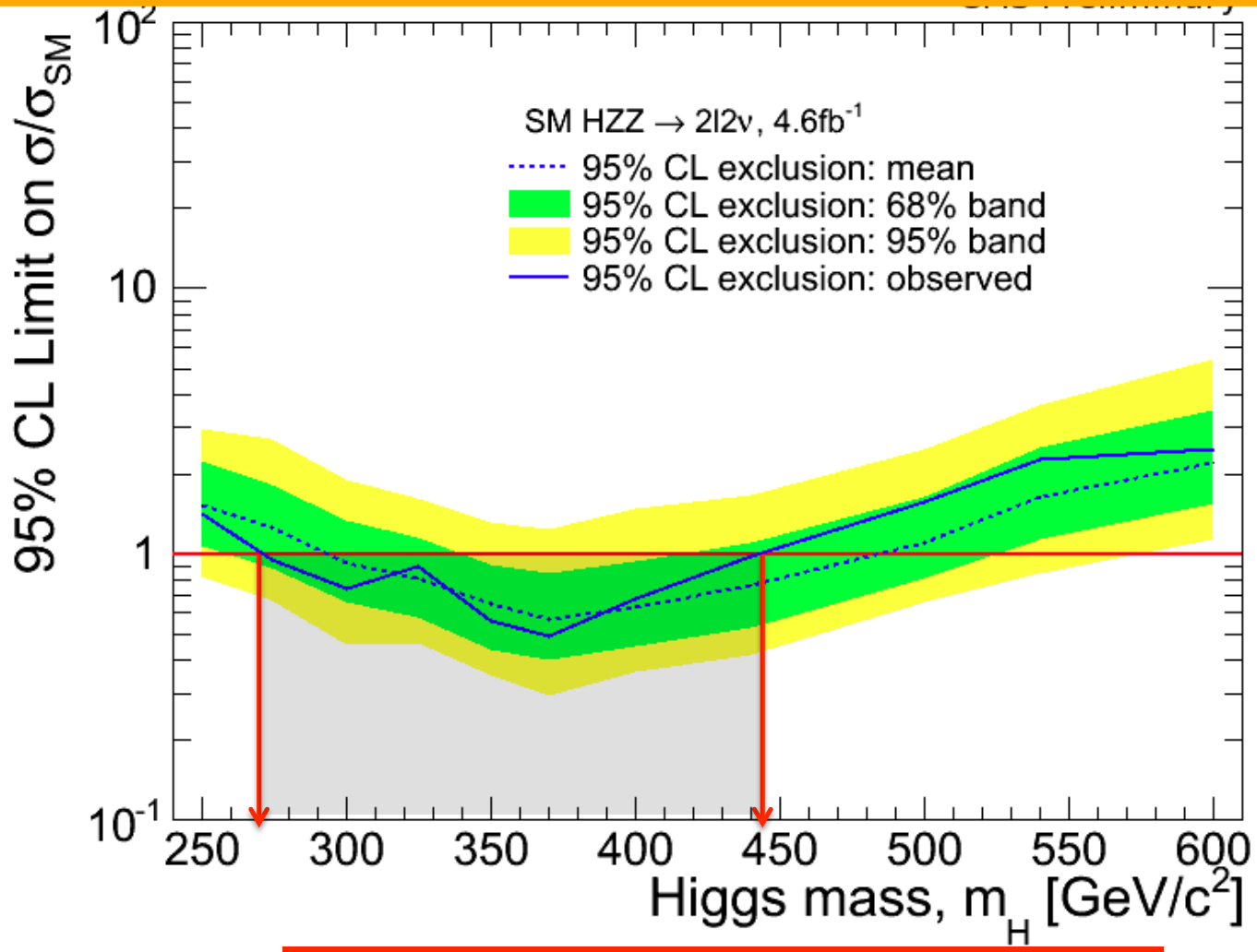
Data-driven prediction for tt/WW bkgd, using m_Z sideband





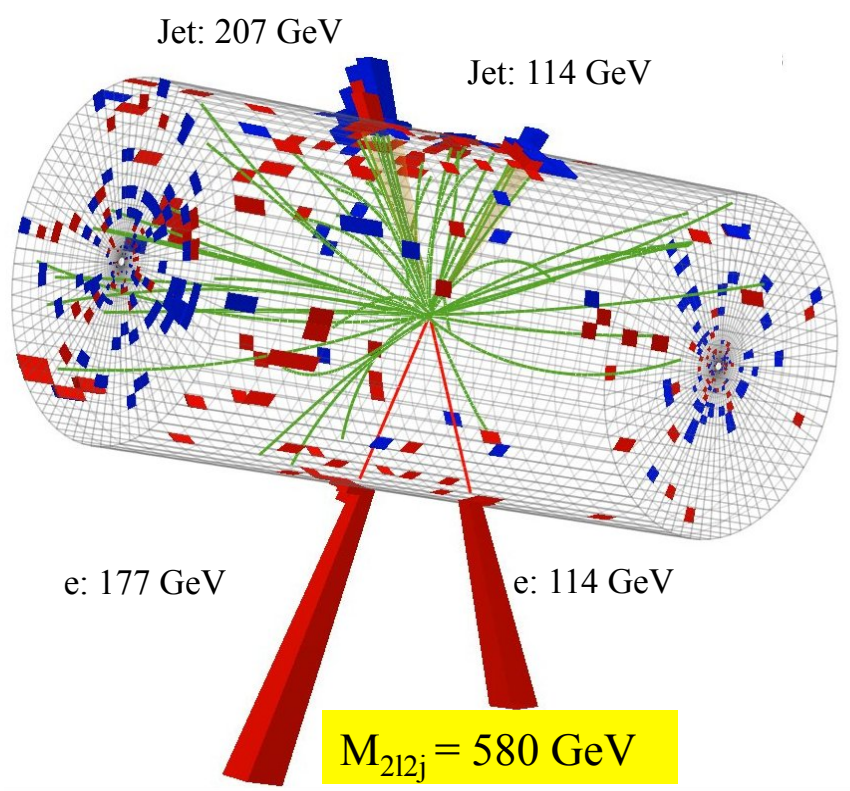
After 250 < MET < 450 GeV

H → ZZ → 2l 2ν: Limits



Expected limit: $290 < M_H < 490$ GeV
Observed limit: $270 < M_H < 440$ GeV

High Mass Higgs: H → ZZ → 2l 2q (or 2b)



Selection:

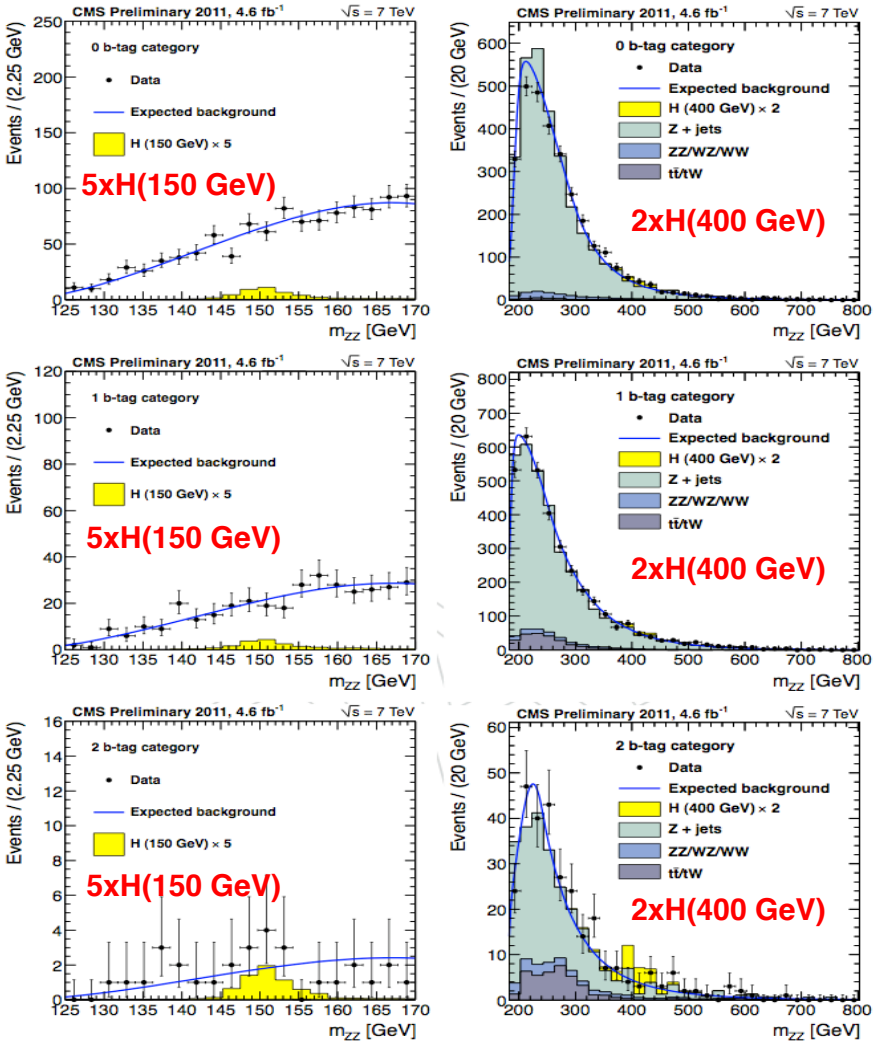
- 2 isolated prompt leptons making Z→2e, Z→2μ
- two jets: Z(jj) with 0, 1, 2 b-tags
- most of sensitivity from 2 b-tag category
- no MET
- cut on angular topology (ME-based)
- Final discriminant: m_{lljj} mass distribution

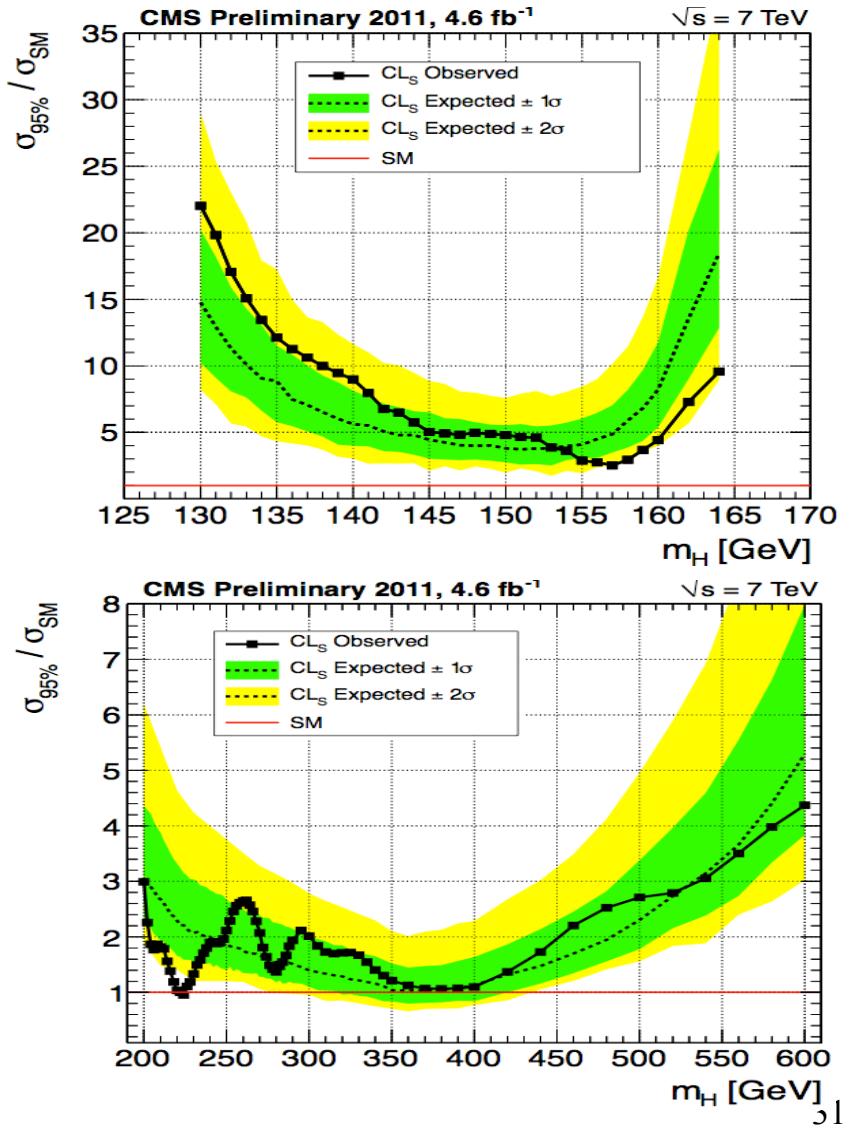
Mass resolution: 3%

Main backgrounds: from sidebands

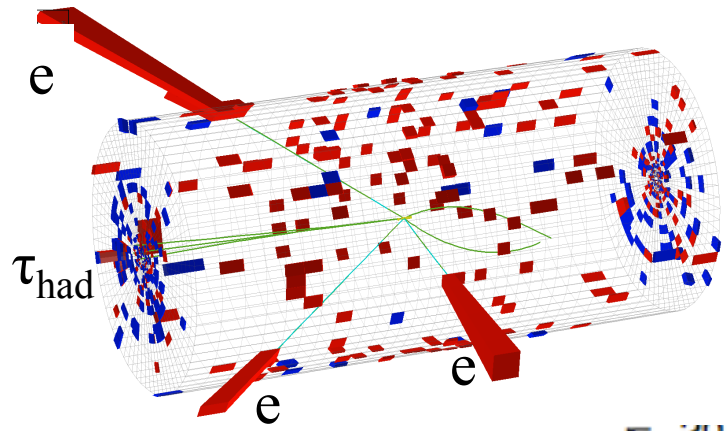
- Z+jets (including heavy flavor jets)
- WZ, ZZ
- ttbar, WW

H → ZZ → 2l 2q : Distributions & Limits

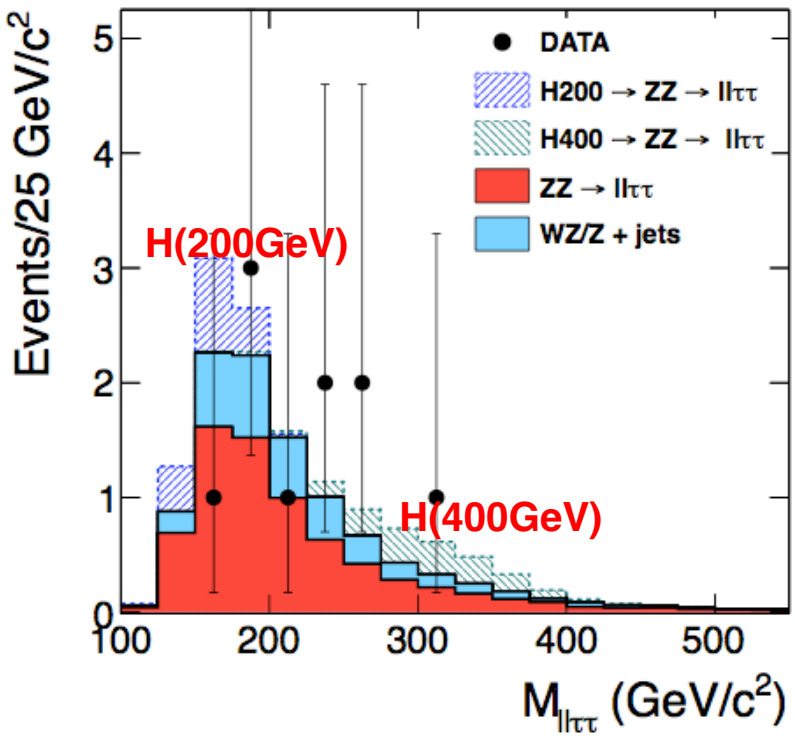




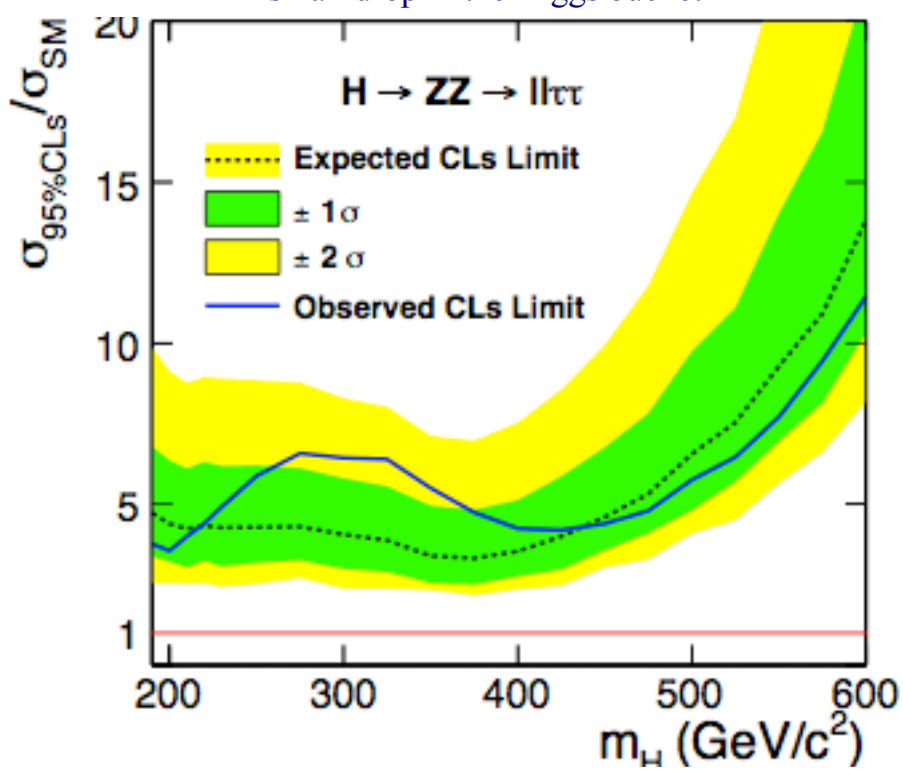
H → ZZ → 2l 2τ



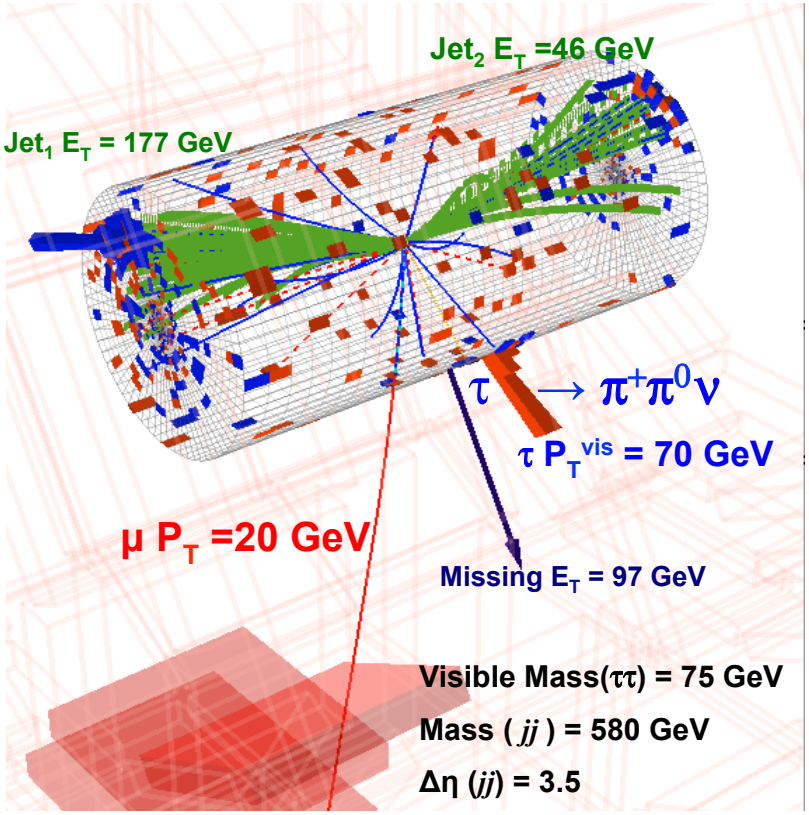
H → ZZ→ eeeτ candidate

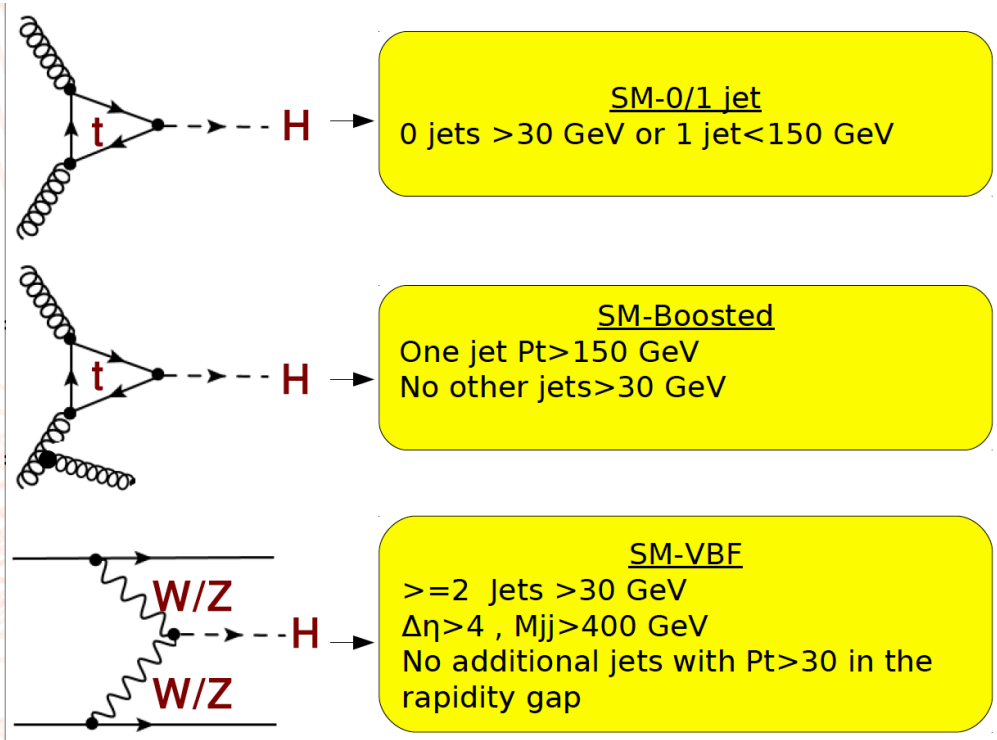


A small drop in the Higgs bucket



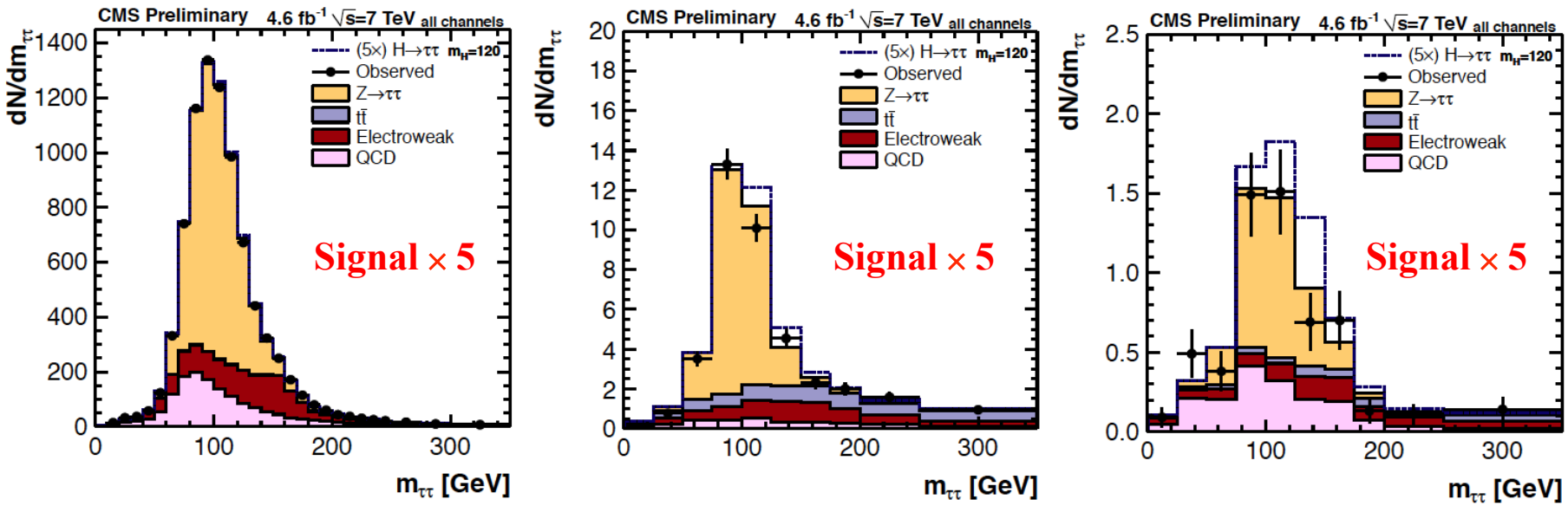
Low Mass Higgs Search : H→ ττ

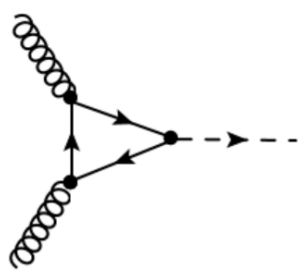


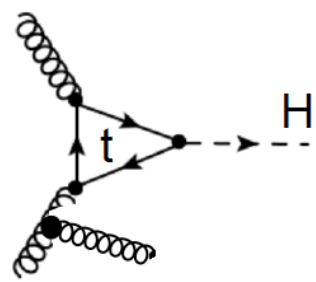


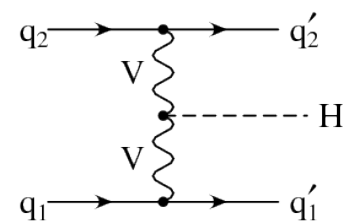
- $\tau\tau$ selection: $\mu+\tau_{had}$, $\mu+\tau_{had}$, μ+e
- SM-Boosted mode added since summer
- <u>VBF mode cleanest, most sensitive</u>

H→ ττ : Mass Spectrum By Categories

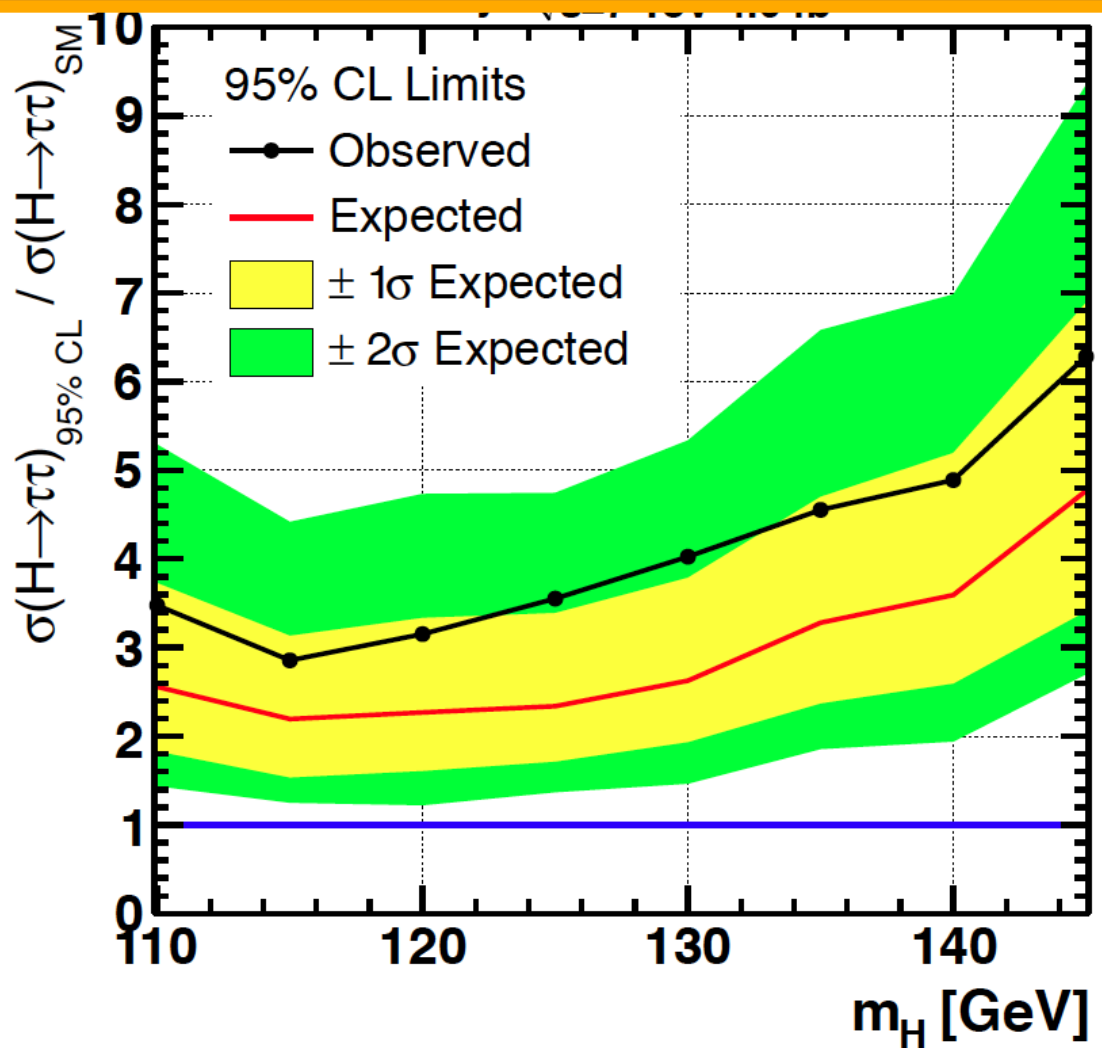






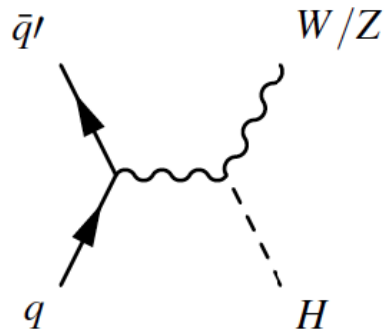


H→ ττ : Limits Becoming Interesting



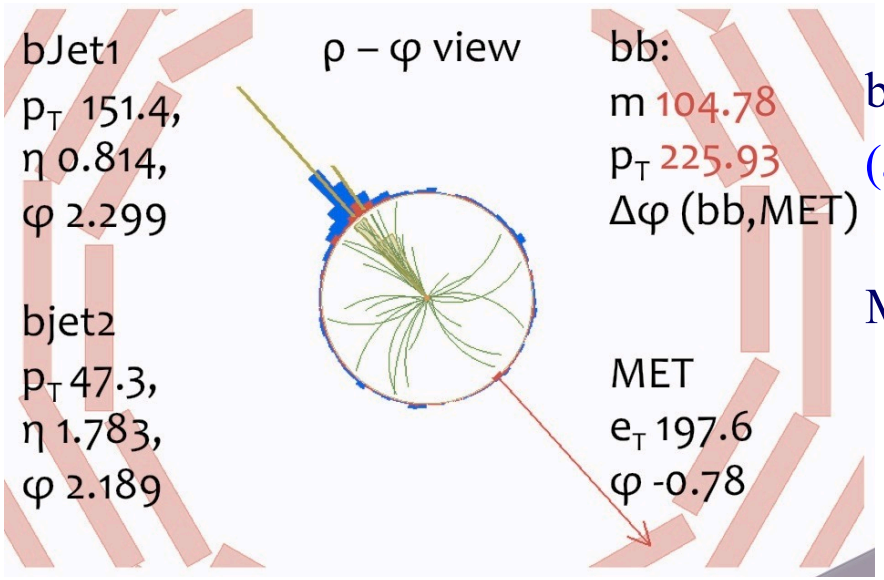
At M_H=120 GeV : Sensitivity = 2.3xσ_{SM}, Observed = 3.2xσ_{SM}

Low Mass Higgs Search : H→ b $\bar{b}$



Selection:

- 5 channels: W→lν, Z→*ll*, Z→νν
- high MET quality for W→lν and Z→νν
- two jets with tight b-tags
- V+H(bb) topology: back-to-back, Δφ(V,H) > 3
- p_T(bb)> 100-160 GeV (*but not super boosted*)
- Final discriminant: MVA output shape



bb mass resolution: 10%

(aided by the boost of bbbar system)

Main backgrounds:

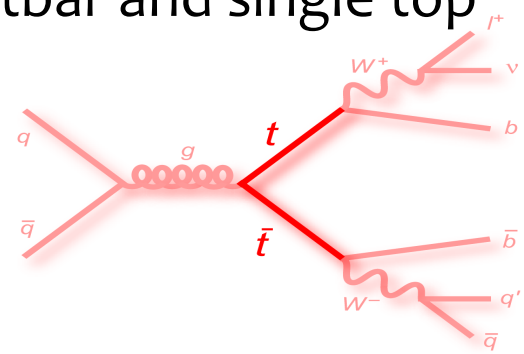
- Vbb: from data (invert p_T(bb) boost)
- V+jets: from data (invert b-tag)
- tt: from data (require extra jet)
- QCD: from data (require small Δφ(MET,jet), …)
- W+Z(bb) and Z+Z(bb): from MC

H→ b b : Data Driven Way of Dealing With Bkgnd

Reducible backgrounds

QCD, V+udscg

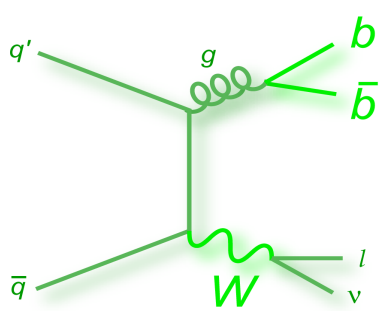
ttbar and single top



Irreducible backgrounds

V+bb @ high p_T

ZZ(bb), W(lν)Z(bb)

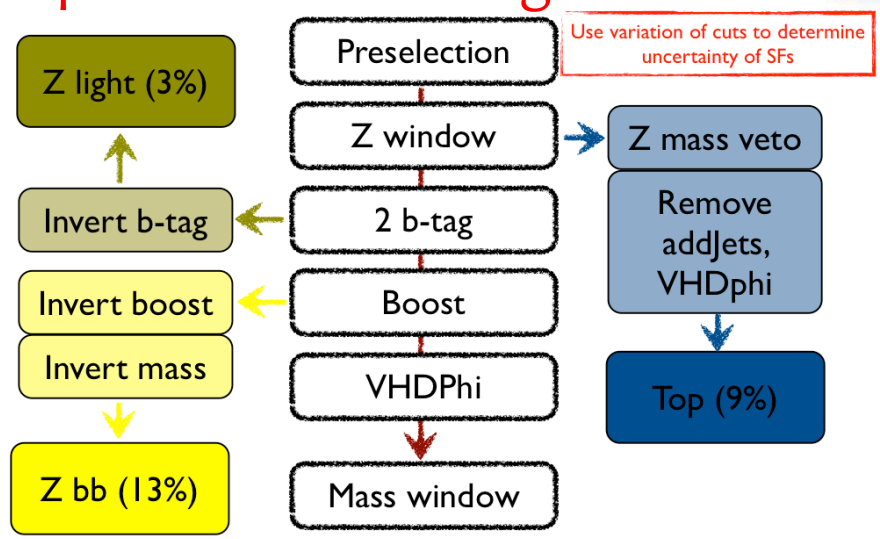


Definition of Control Regions (CR) crucial element of the analysis

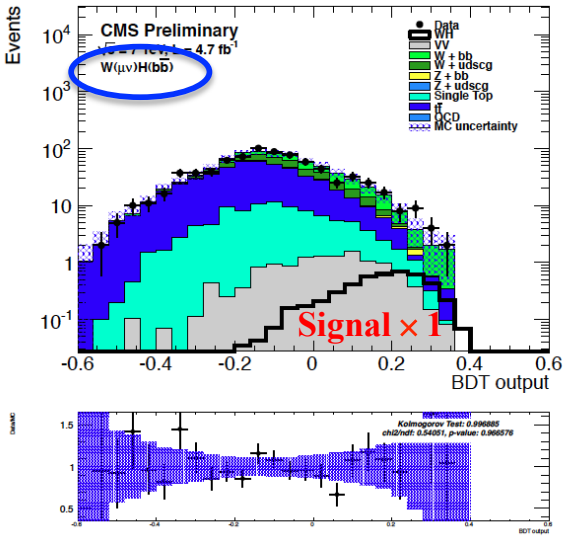
Define several CR enriched in different background components:

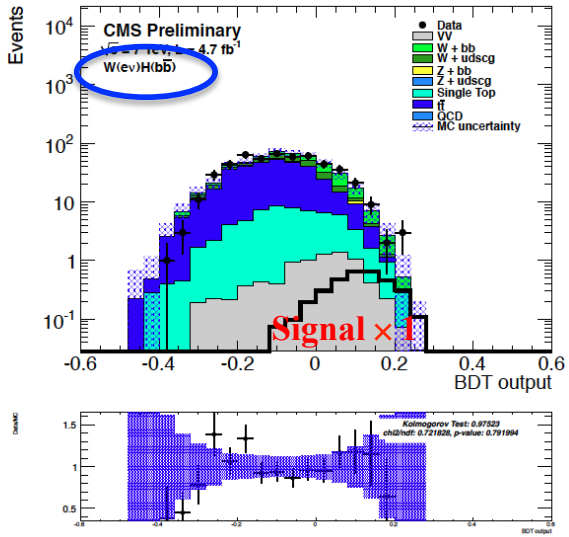
Control regions cuts as close as possible to the signal region

Example: Zee control region definition

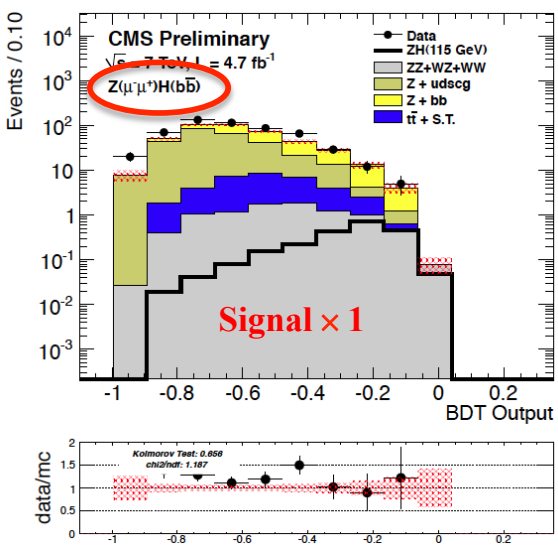


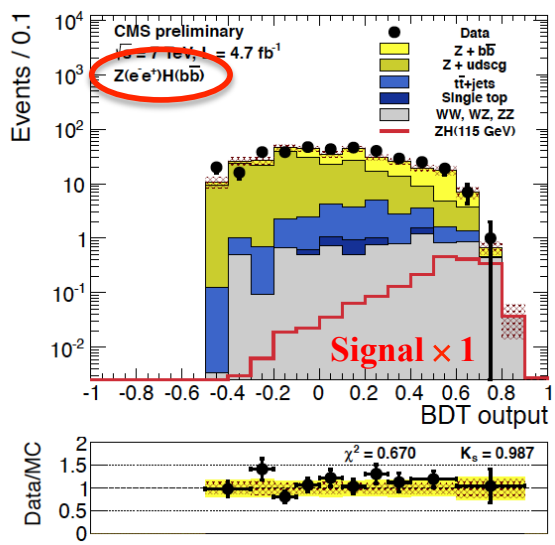
Low Mass Higgs Search : H→ b $\bar{b}$

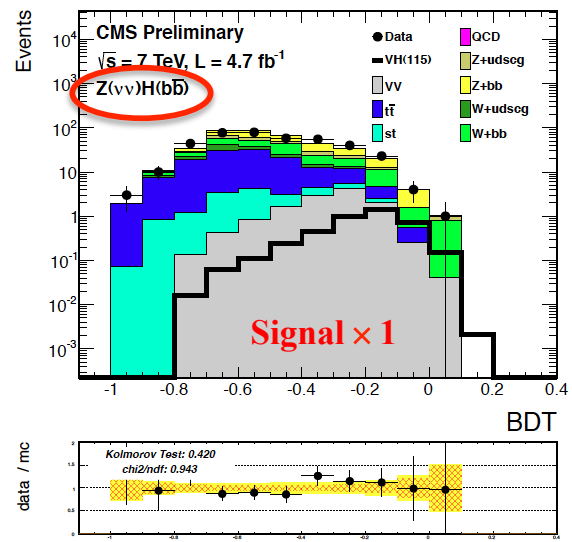




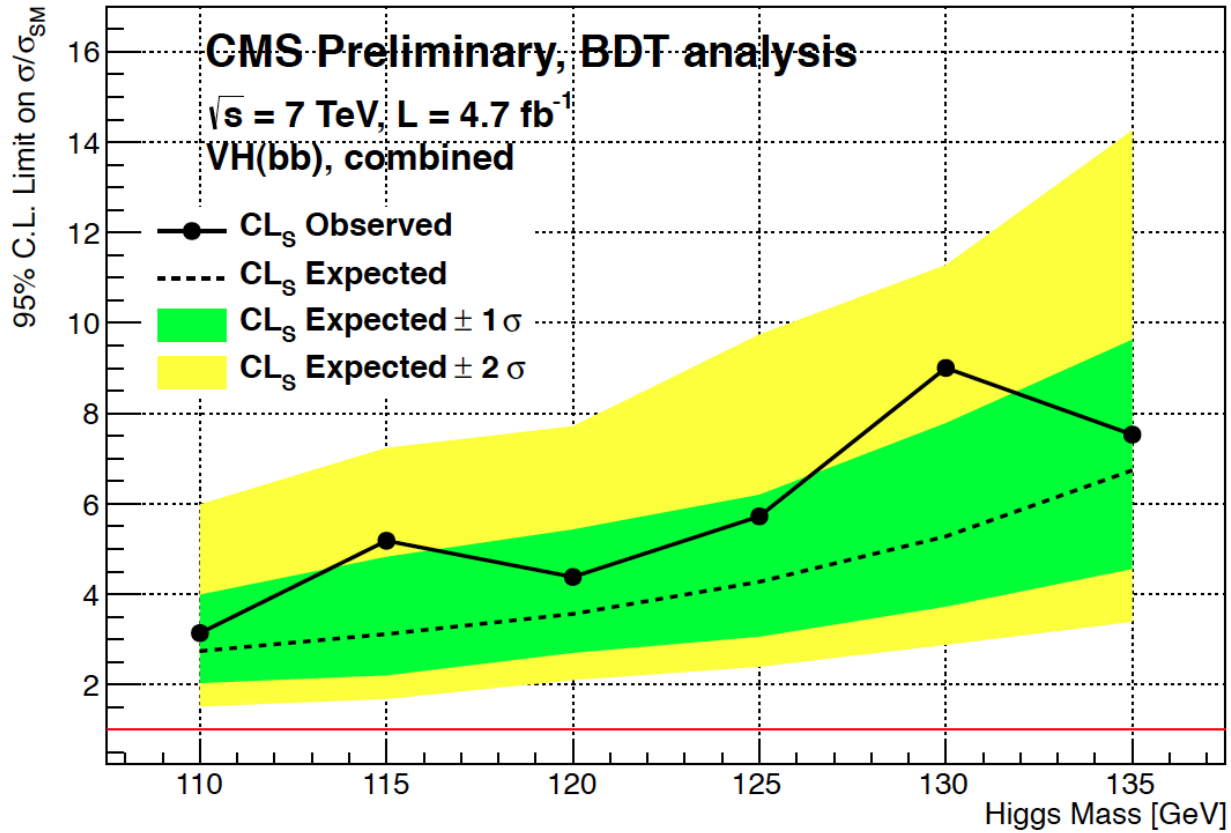
Systematic errors are a persistent challenge as the integrated lumi increases







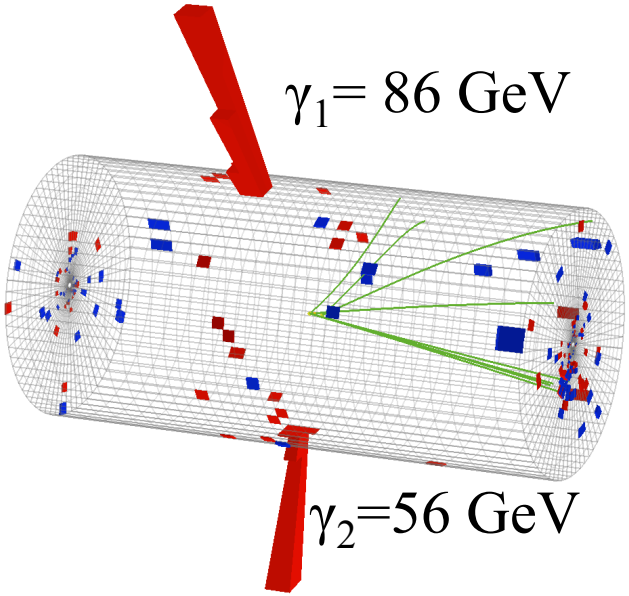
Low Mass Higgs Search : $H \rightarrow b\bar{b}$



BDT shape analysis for m_H=120 GeV

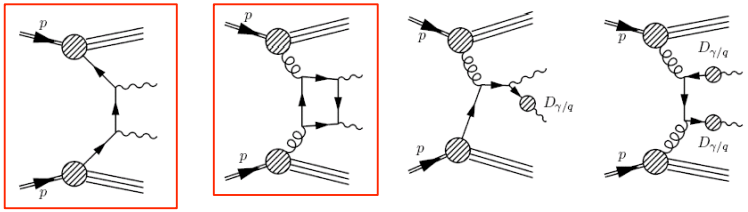
- Sensitivity ≈ $3\times\sigma_{SM}$
- Observed ≈ $4\times\sigma_{SM}$

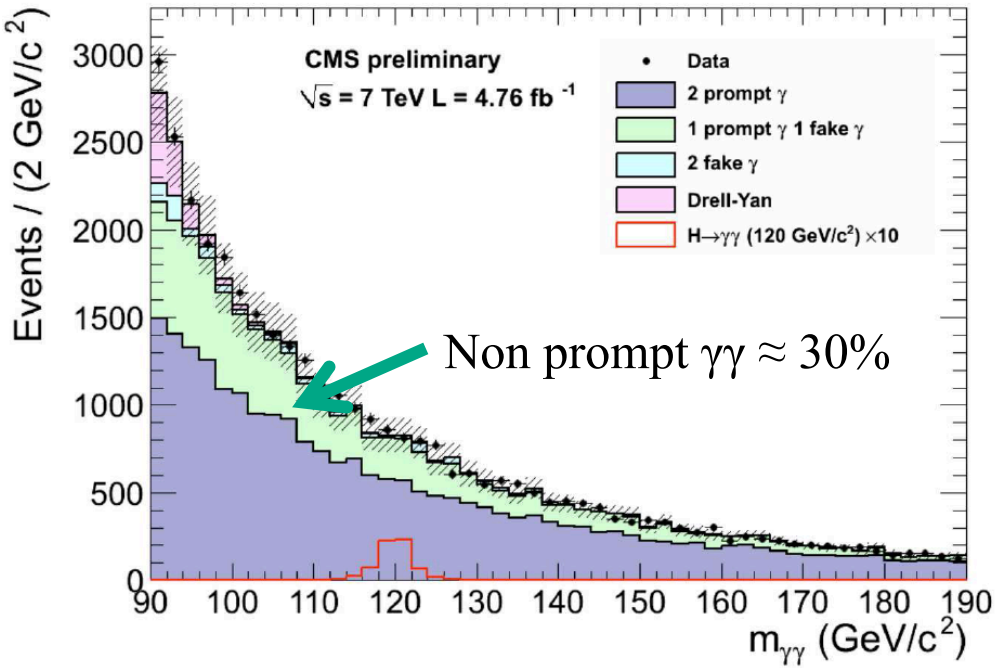
Low Mass Higgs Search : H→ γγ



Signature: 2 energetic, isolated γ, narrow mass peak

Background: Large & partly irreducible QCD. Measured from $M_{\gamma\gamma}$ sidebands in data

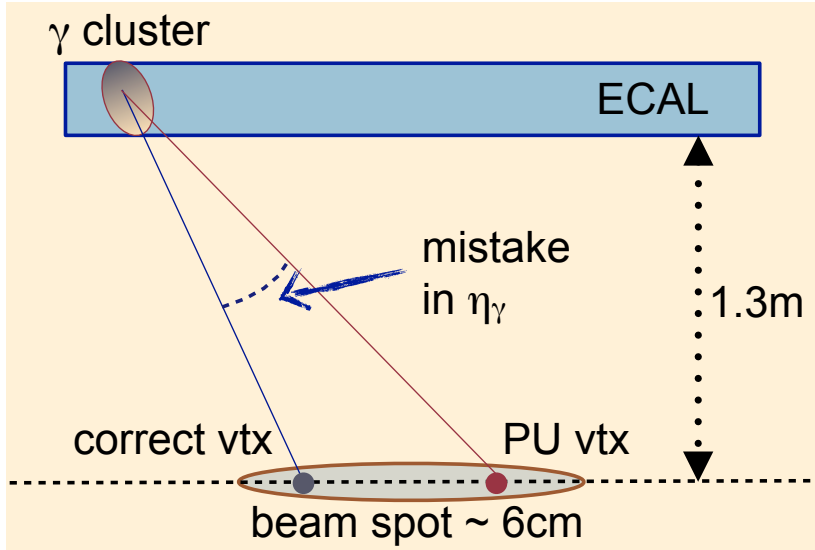


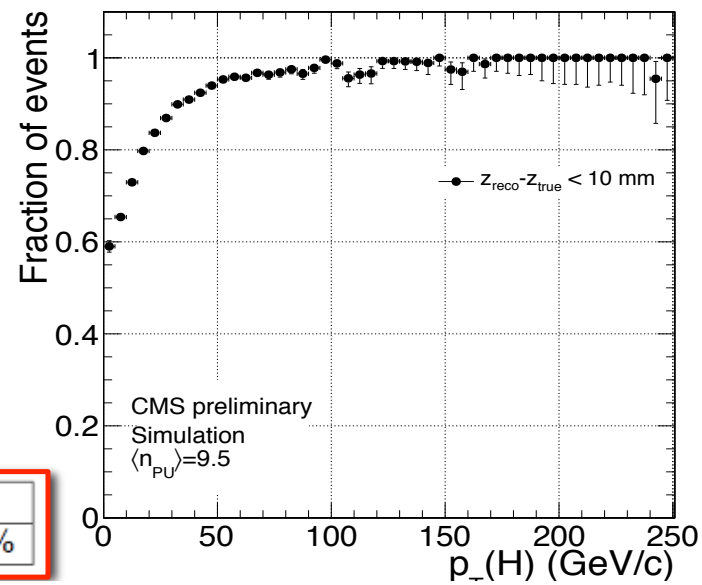


- Data divided in 4 categories depending on γγ mass resolution
- Background shape fitted by a 5th order polynomial constrained to be positive
  - Several other forms also tried
- Signal shape : Sum of 3 Gaussians, parameters determined from Z → ee data

H →γγ Event Vertex Determination: Impact of Pileup

- **"large" pile-up conditions**
  - ➪ $\langle N_{PU} \rangle \sim 10$
- di-photon invariant **mass resolution affected by vertex** choice
- **vextex determination** based on
  - **tracks belonging to vertex** combined **with di-photon kinematics**
    - use of $\Sigma p_T^2{}_{trk}$ and p_T balancing
  - **conversion-track** finding and projection on beam spot
- performance **cross-checked using** $\mathbf{Z \rightarrow \mu^+\mu^-}$ after removing muon tracks





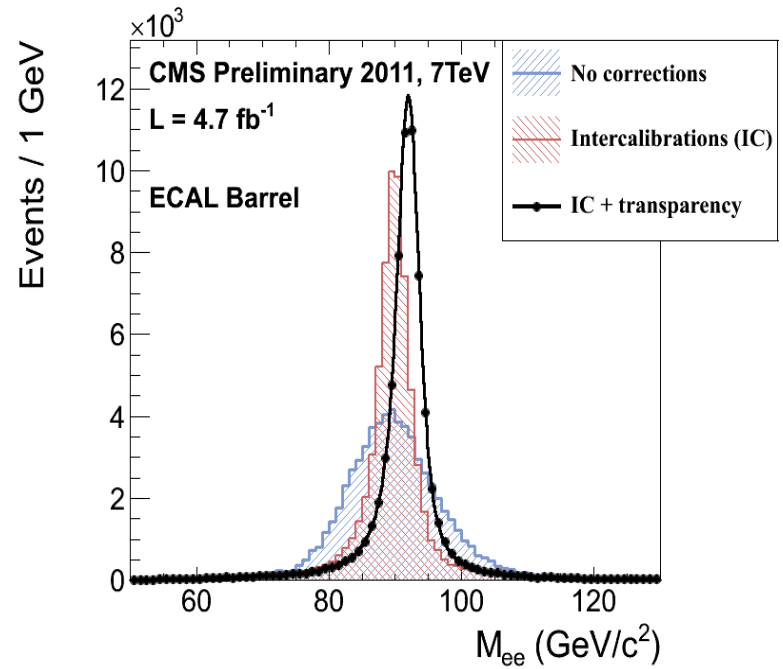
2011A	2011B	2011
86.3% ± 0.2% ± 0.4%	79.8% ± 0.2% ± 0.5%	83.0% ± 0.2% ± 0.4%

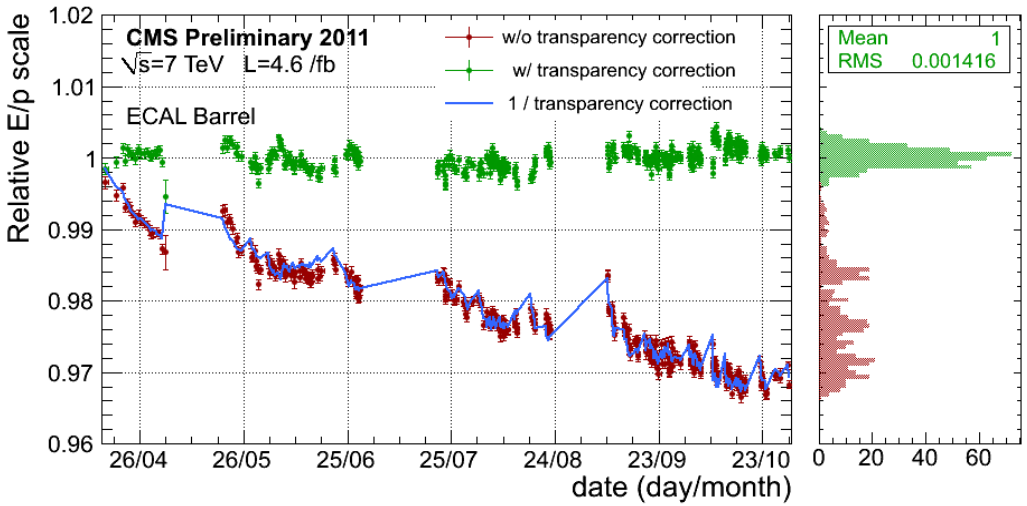
Photon Energy Scale & Resolution

Energy resolution studied with Z →ee, W →eν , π^0 inter-calibrations and laser signals for transparency corrections

Resolution in data improves typically by 10% in barrel (|η|<1, R9>0.94) w.r.t LP'11 data

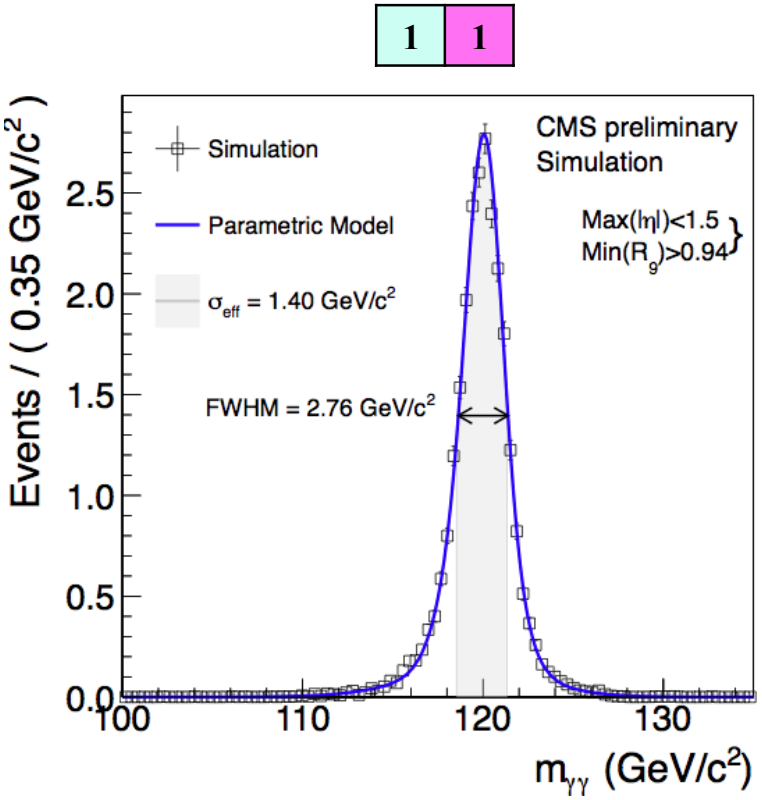
Instrumental resolution in the best Barrel category = 0.99 ±0.01 GeV

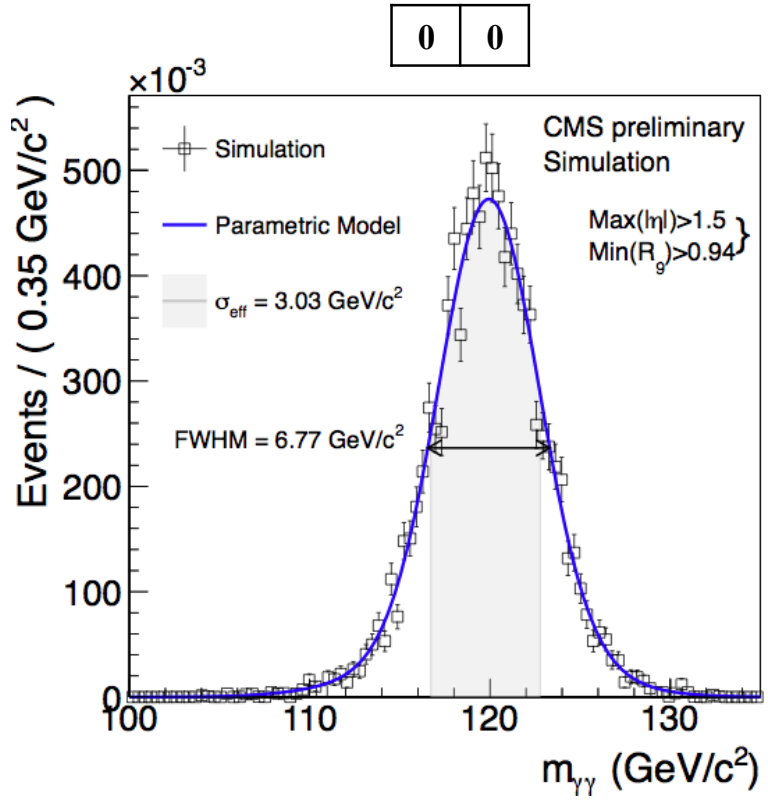




Energy scale for W→eν and Z→ee stable throughout 2011 at the level of 0.2% in barrel and 0.7% in endcap

$M_{\gamma\gamma}$ Resolution In Four Catagories





	both γ in barrel		one or two γ in endcap	
	$min(R_9)$ >0.94	$min(R_9)$ <0.94	$min(R_9)$ >0.94	$min(R_9)$ <0.94
	unconverted γ		unconverted γ	
signal	31.1%	40.3%	12.2%	16.4%
data	23.0%	33.8%	17.8%	25.4%
resol. σ_{eff}	1.4 GeV	1.8 GeV	2.8 GeV	3.2 GeV

H→ γγ : Systematic Uncertainties

Source		Uncertainty	
Photon identification efficiency:	barrel	1.0%	
	endcap	2.6%	
R_9 >0.94 efficiency	barrel	4%	
(results in class migration)	endcap	6.5%	
		$R_9 > 0.94$	$R_9 < 0.94$
Energy resolution ($\Delta\sigma/E_{MC}$):	barrel low η, high η	0.22%, 0.61%	0.24%, 0.59%
	endcap low η, high η	0.91%, 0.34%	0.30%, 0.53%
Energy scale ($(E_{data} - E_{MC})/E_{MC}$)	barrel low η, high η	0.19%, 0.71%	0.13%, 0.51%
	endcap low η, high η	0.88%, 0.19%	0.18%, 0.28%
Integrated luminosity		4.5%	
Trigger efficiency: One or more photons $R_9 < 0.94$ in endcap		0.4%	
	Other events	0.1%	
Vertex finding efficiency		0.4%	
gluon fusion process cross section (scale)		+12.5% -8.2%	
gluon fusion process cross section (PDF)		+7.9% -7.7%	
Vector boson fusion process cross section (scale)		+0.5% -0.3%	
Vector boson fusion process cross section (PDF)		+2.7% -2.1%	
Associated production with W/Z cross section (scale)		+1.8% -1.8%	
Associated production with W/Z cross section (PDF)		+4.2% -4.2%	
Associated production with $t\bar{t}$ cross section (scale)		+3.6% -9.5%	
Associated production with $t\bar{t}$ cross section (PDF)		+8.5% -8.5%	

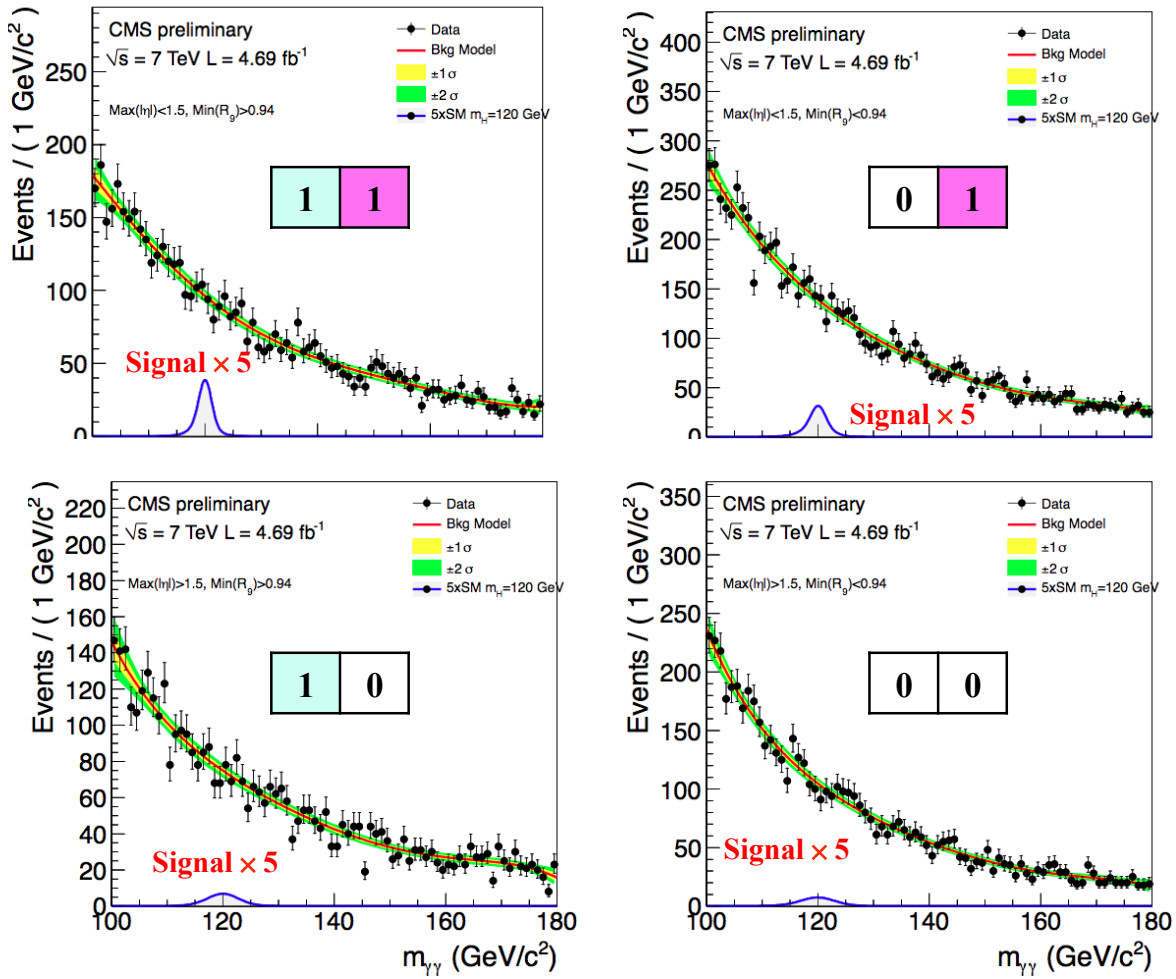
H→ γγ Search Categories

Both photons of high quality?
Both photons in barrel?

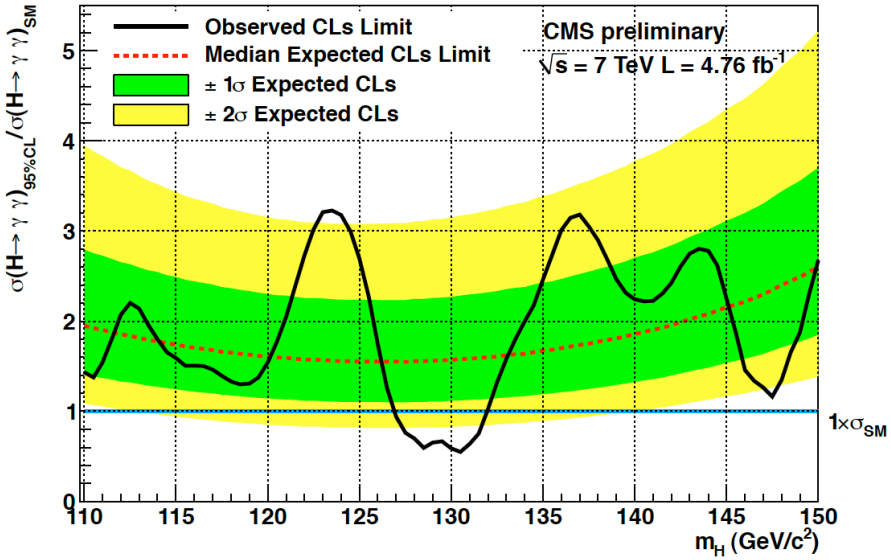
1	1

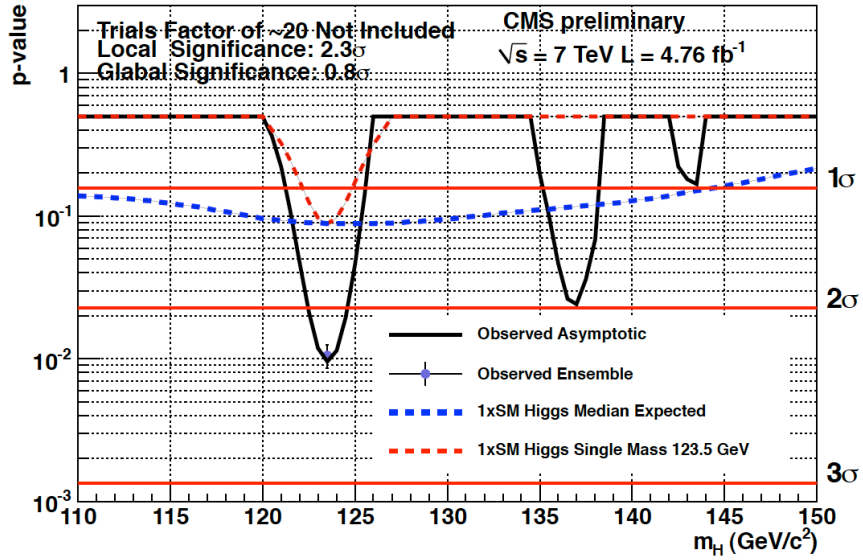
Important:
- $M_{\gamma\gamma}$ resolutions
- S/B ratios

Categorization helps with both



H→ γγ : Results



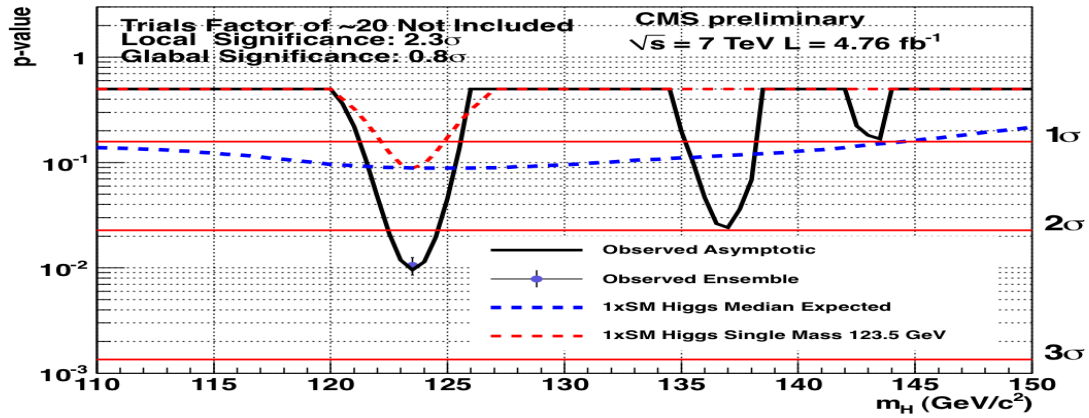


Expected exclusion limits between 1.5 and 2.0 times the Standard Model cross-section in the mass range of interest

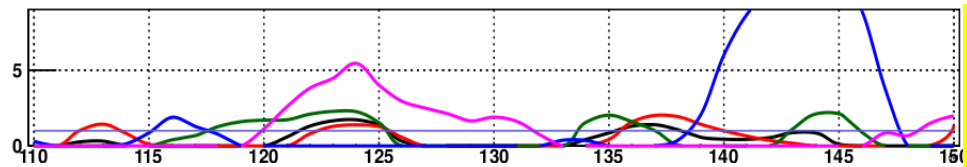
Observed excess at ~123.5 GeV has a local significance of 2.3 σ, but only 0.8 σ accounting for the Look-Elsewhere-Effect over the 110-150 GeV search range Signal Strength σ/σ_{SM}= 1.7±0.8 at $M_{\gamma\gamma}$=123.5 GeV is

Exclude (lucky) $127 < M_H < 131$ GeV Higgs mass range @ 95% CL

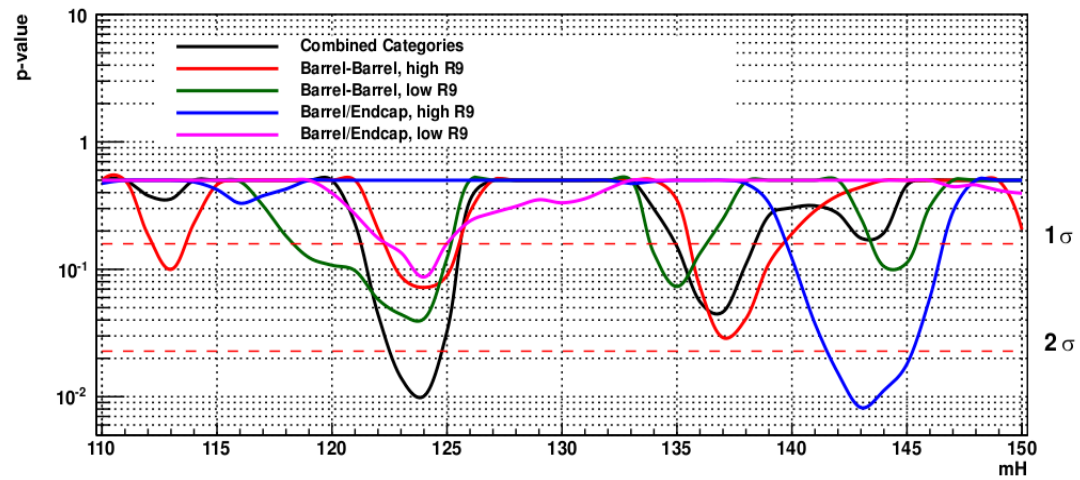
Staring At It : Contributions From the 4 Categories



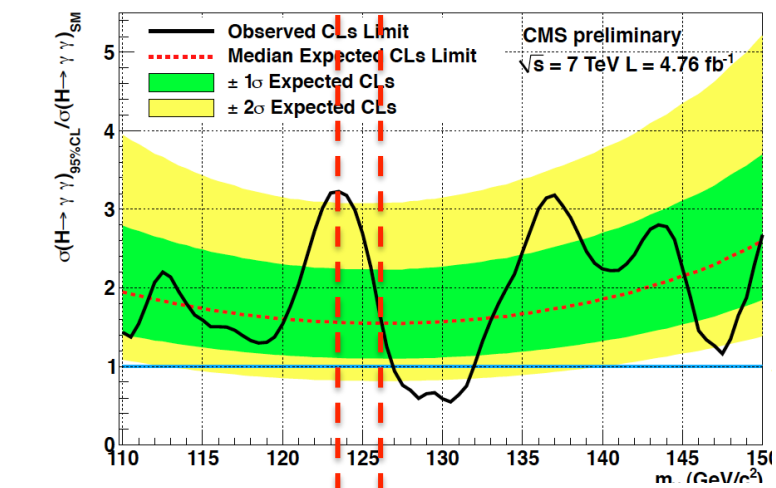
$\mu=\sigma/\sigma_{SM}$

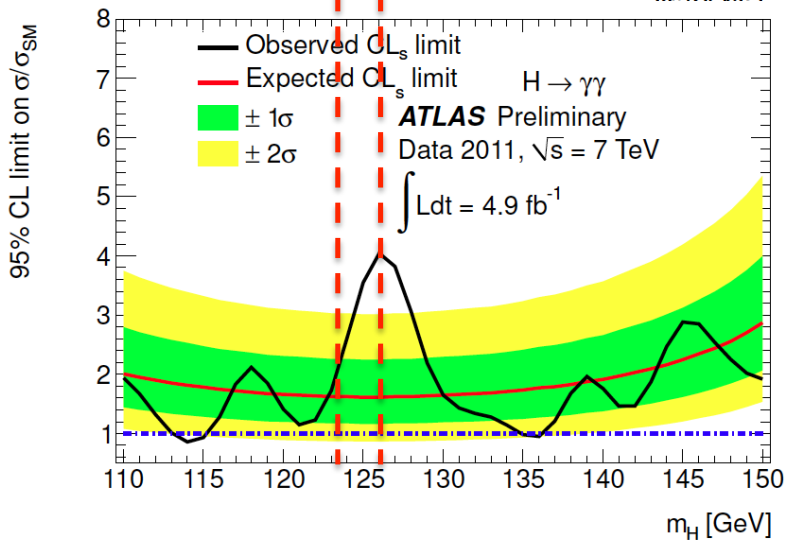


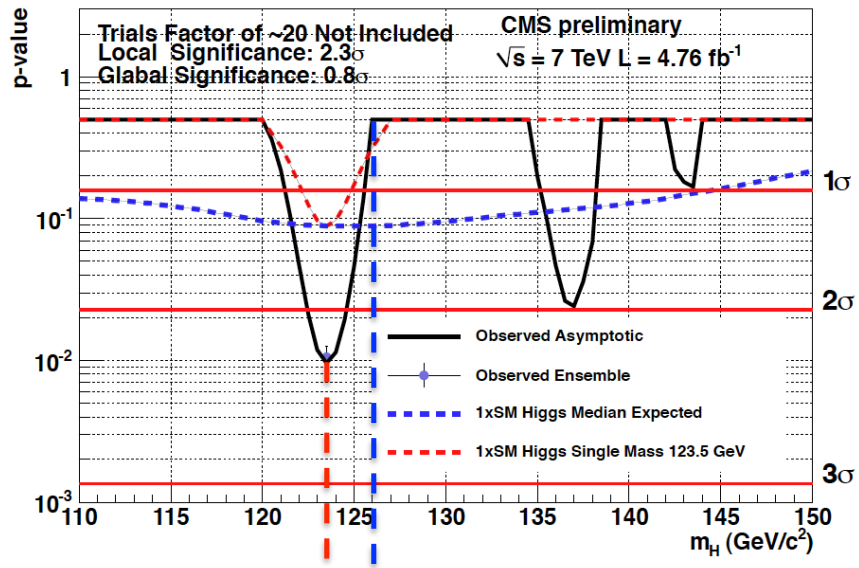
σ/σ_{SM}= 1.7±0.8 at $M_{\gamma\gamma}$=123.5

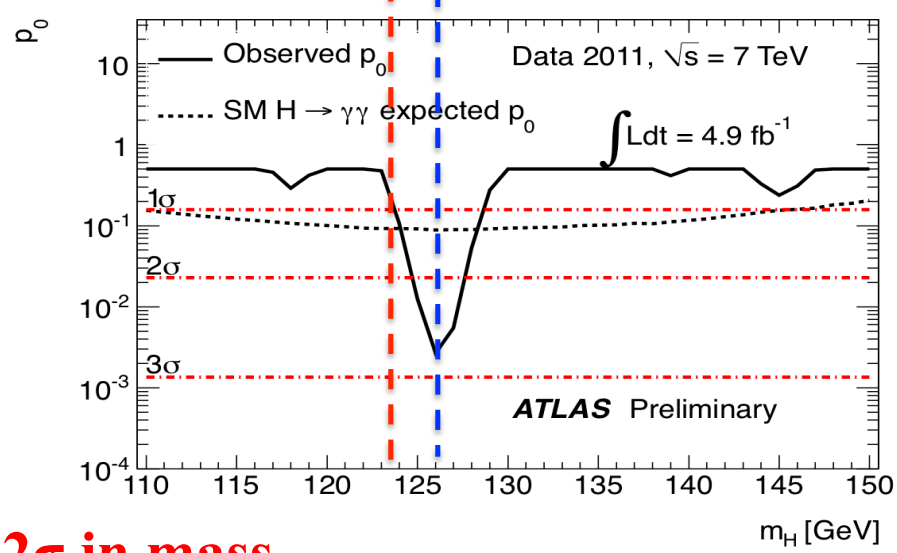


H→ γγ : CMS & ATLAS Compared

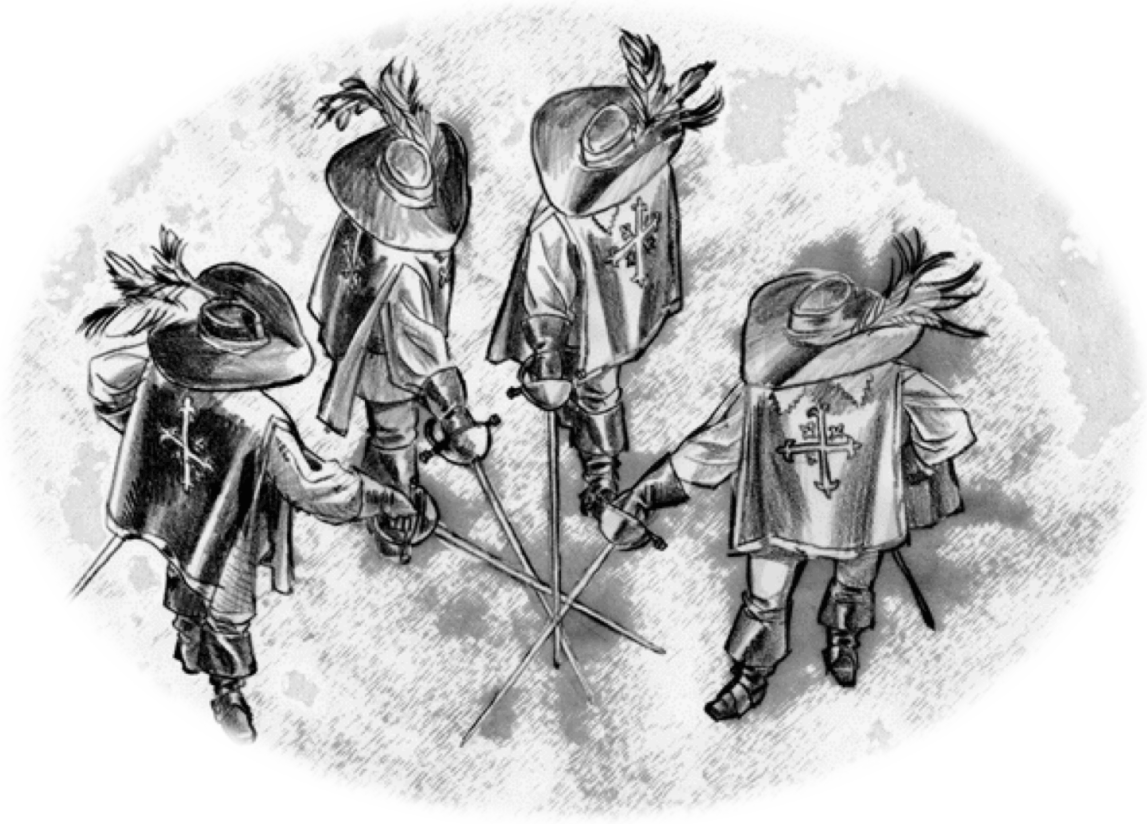






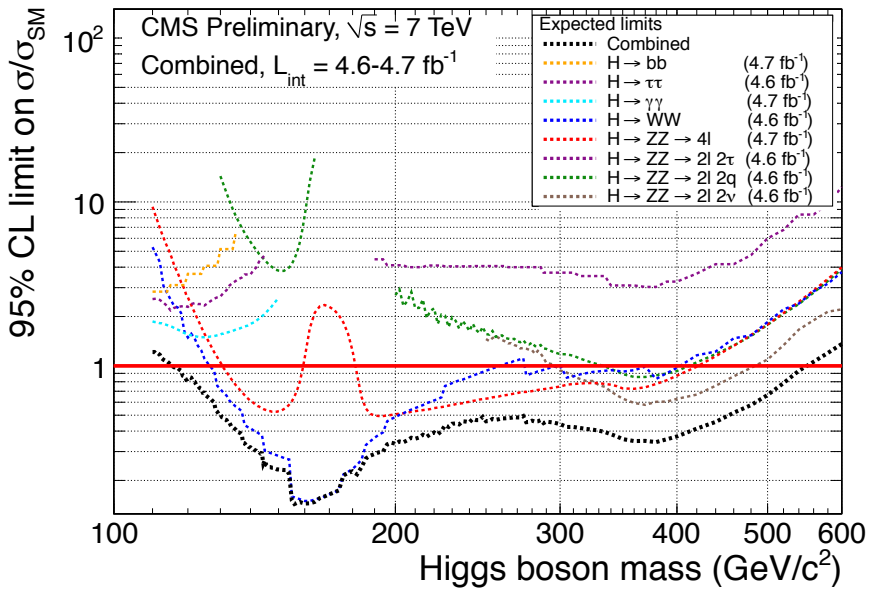


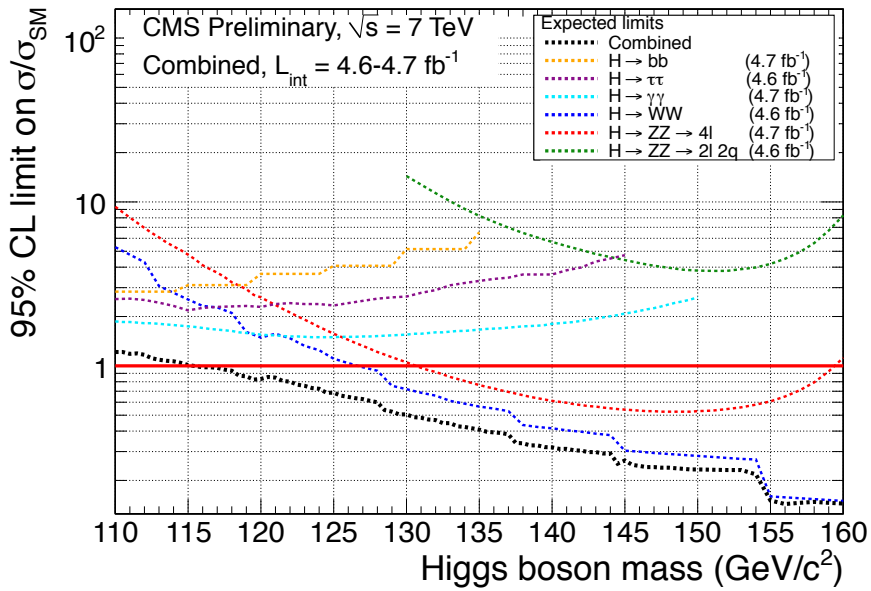
Different by 2σ in mass



Combination Of CMS SM Higgs Searches

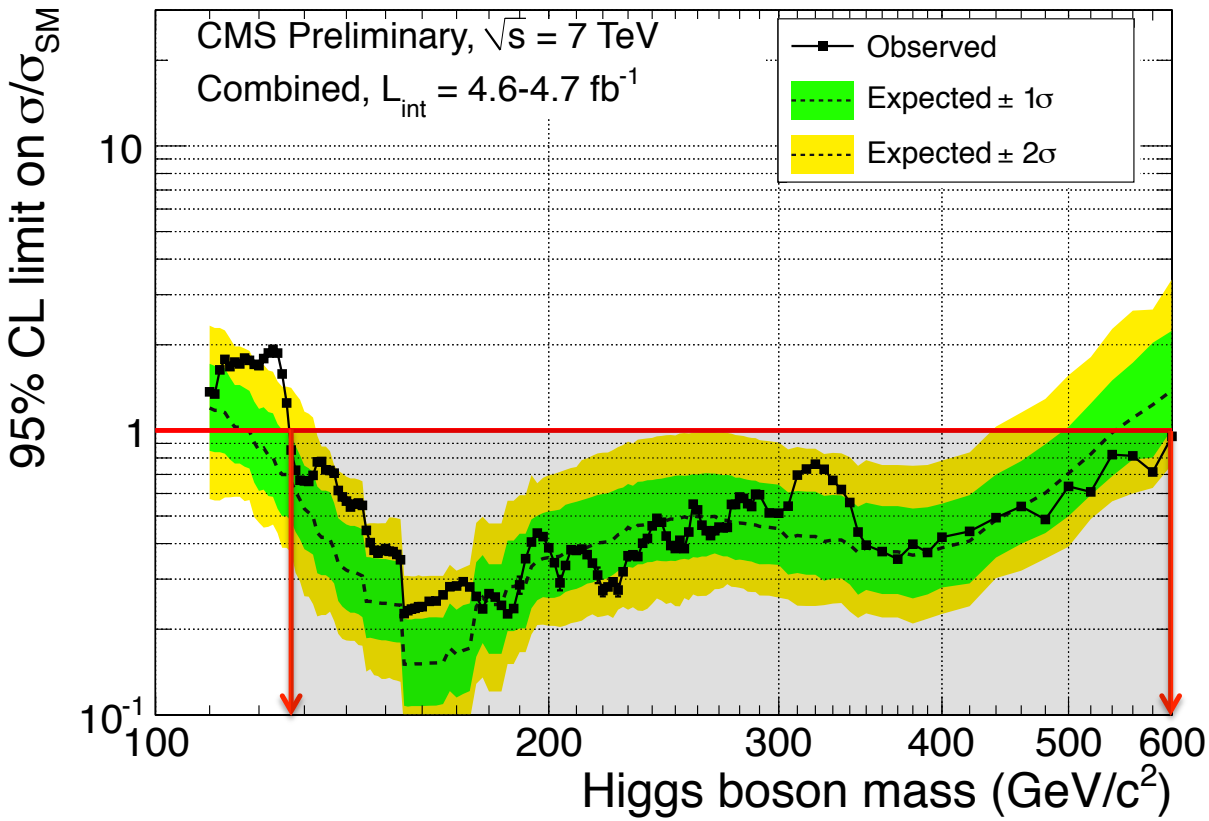
Expected Sensitivity with 4.7 fb^{-1}





95% CL expected sensitivity: 117—543 GeV

Exclusion Limits on $\mu = \sigma/\sigma_{SM}$

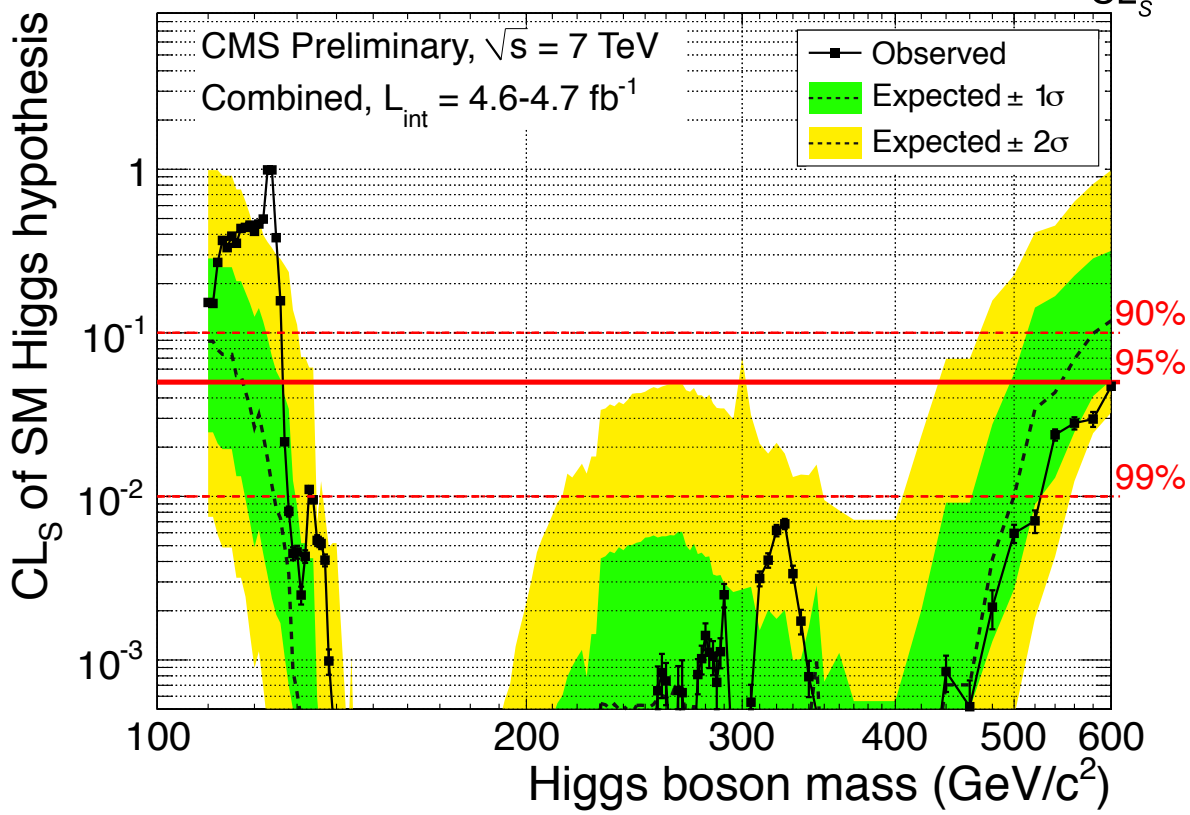


95% CL expected limits: 117—543 GeV

95% CL observed limits: 127—600 GeV

Observed Exclusion Confidence Levels

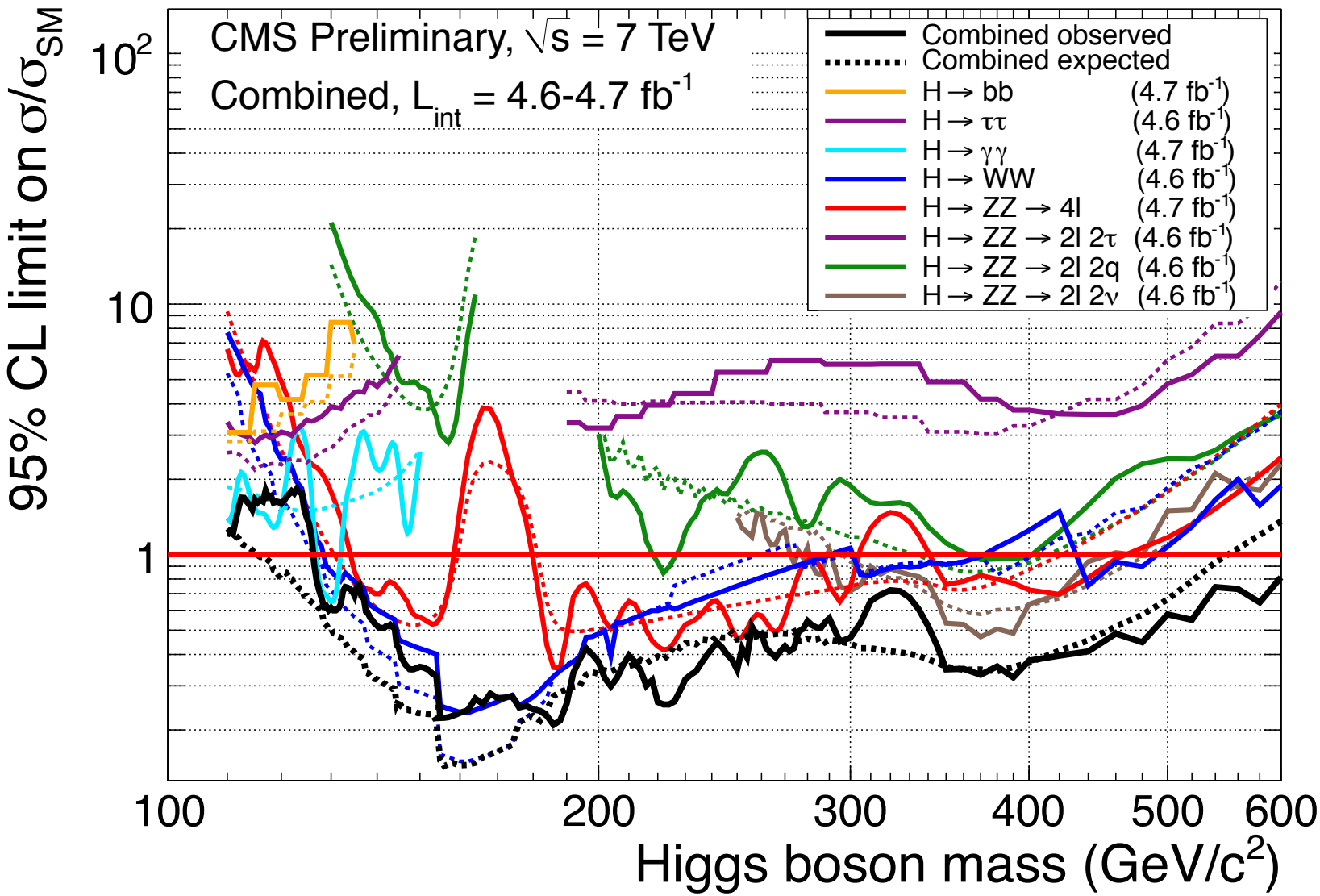
$$CL_S = \frac{P(\, n \le n_{obs} \mid b + s_{SM}\,)}{P(\, n \le n_{obs} \mid b\,)}$$



95% CL observed limits: 127—600 GeV

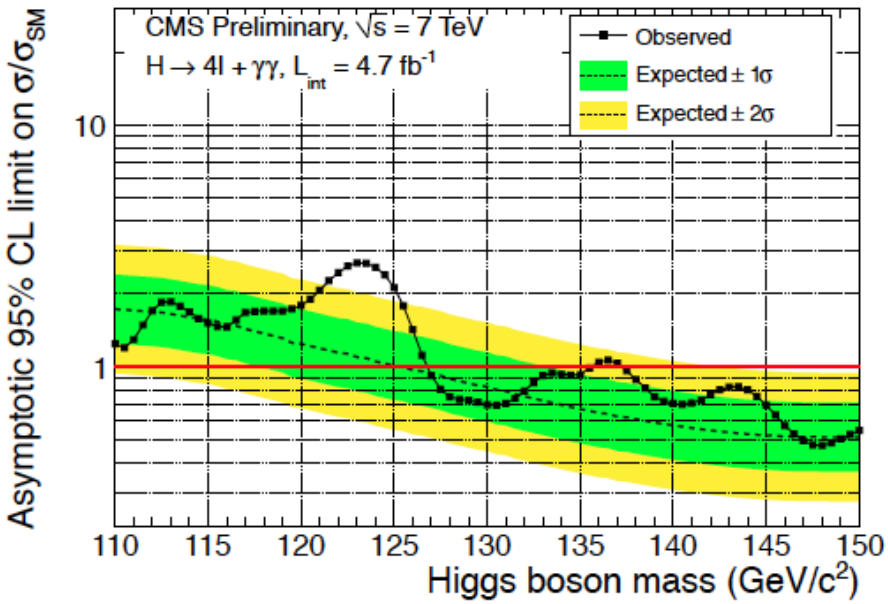
99% CL observed limits: 128 —525 GeV

Expected & Observed Limit : Channel By Channel

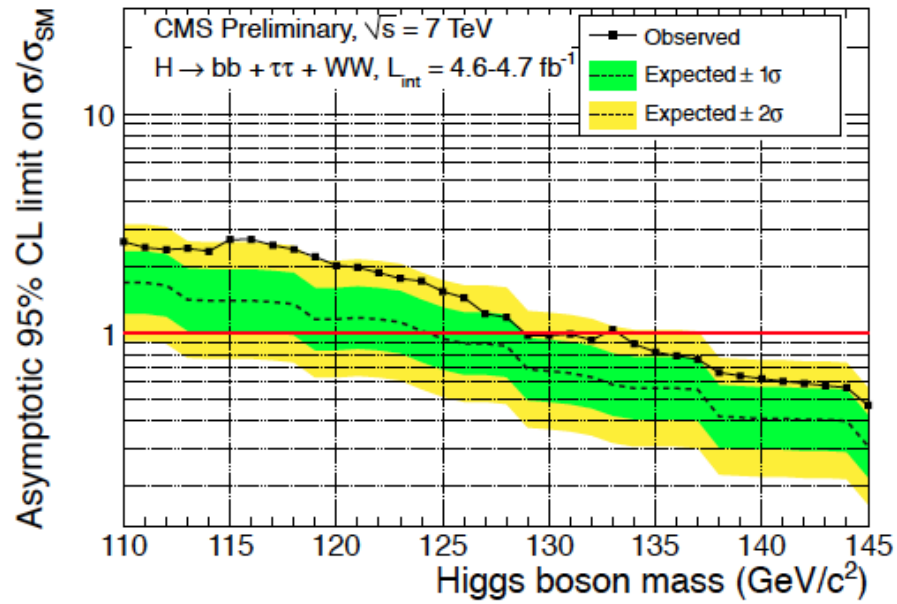


By High & Low Mass Resolution Channels

High mass resolution channels: γγ + 4l

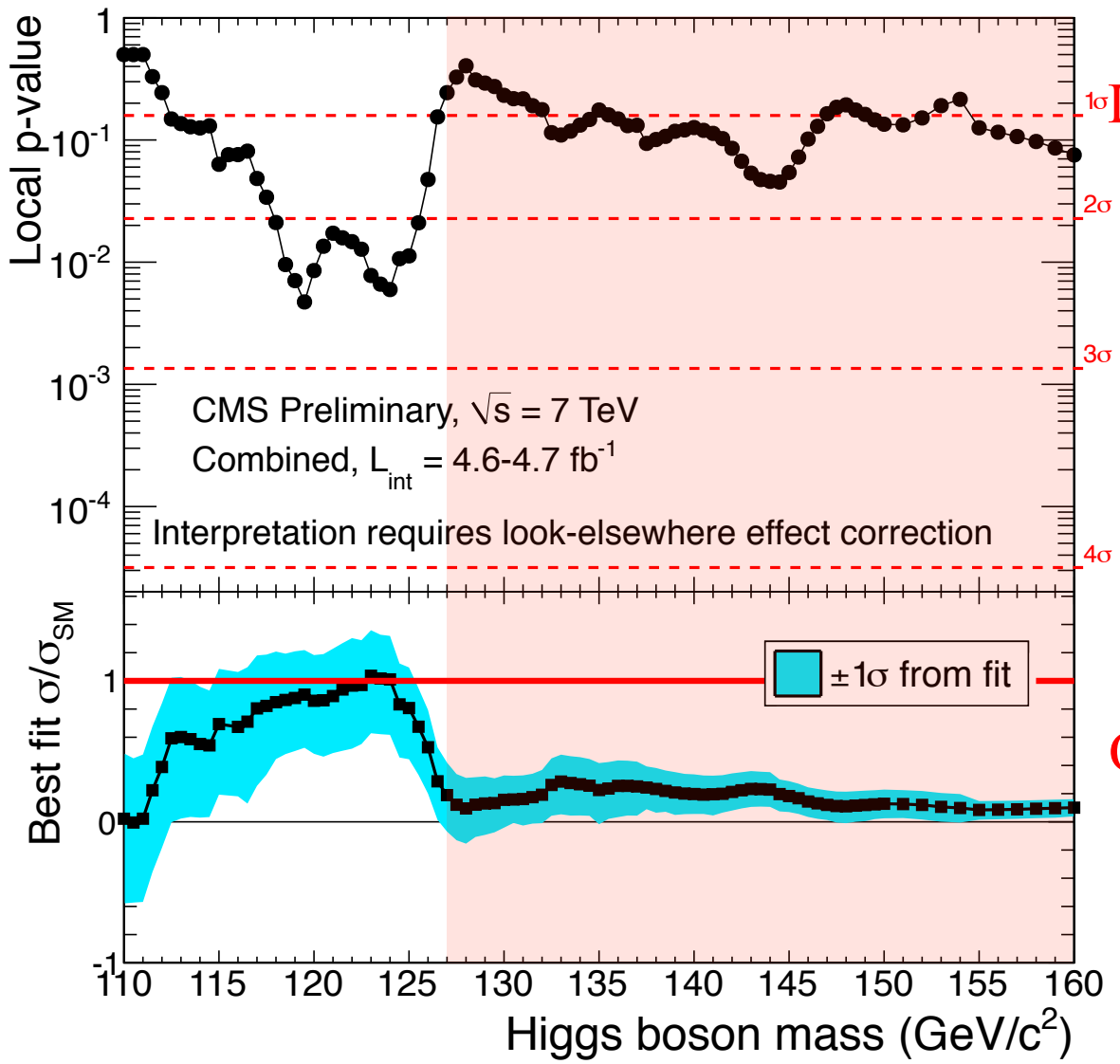


Poor mass resolution channels: WW + ττ + bb



- The two sets have nearly identical sensitivity
- The γγ+4l group shows a localized excess >2σ around m_H=121-125 GeV
- The WW+ττ+bb group shows a broad excess, reaching 2σ around 115-125 GeV

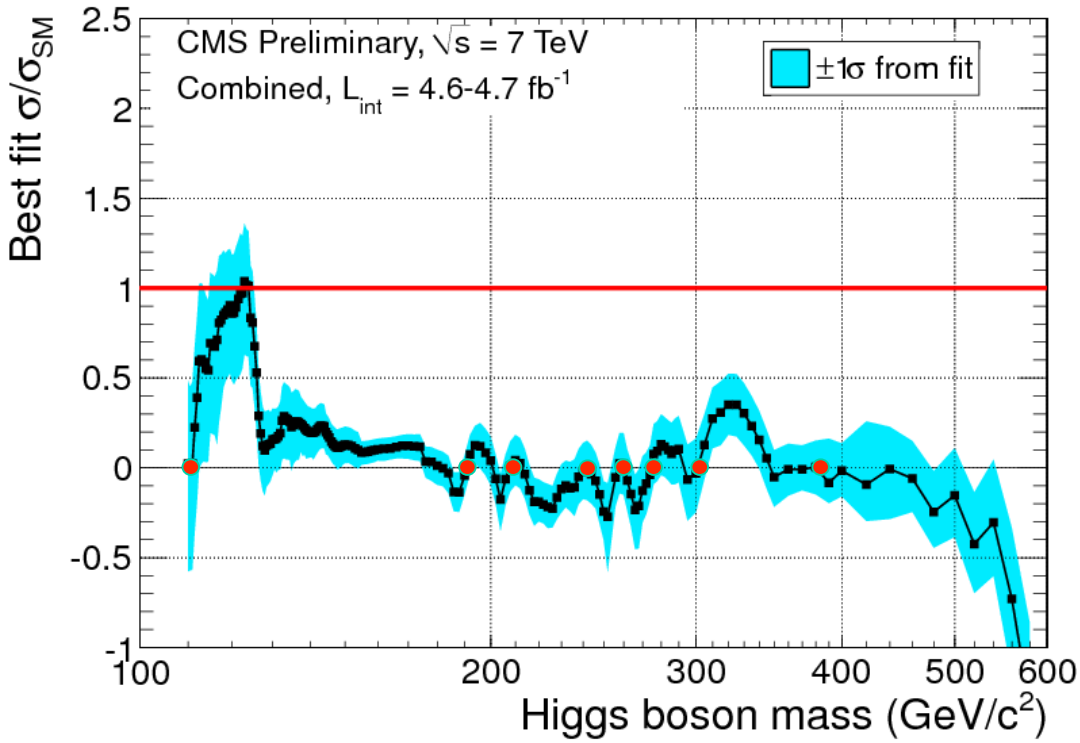
Quantifying The Low Mass Excess



Largest observed excess
Z_{max}=2.6; p_{min}=0.005

Observed excess at 124 not inconsistent (±1σ) with SM Higgs

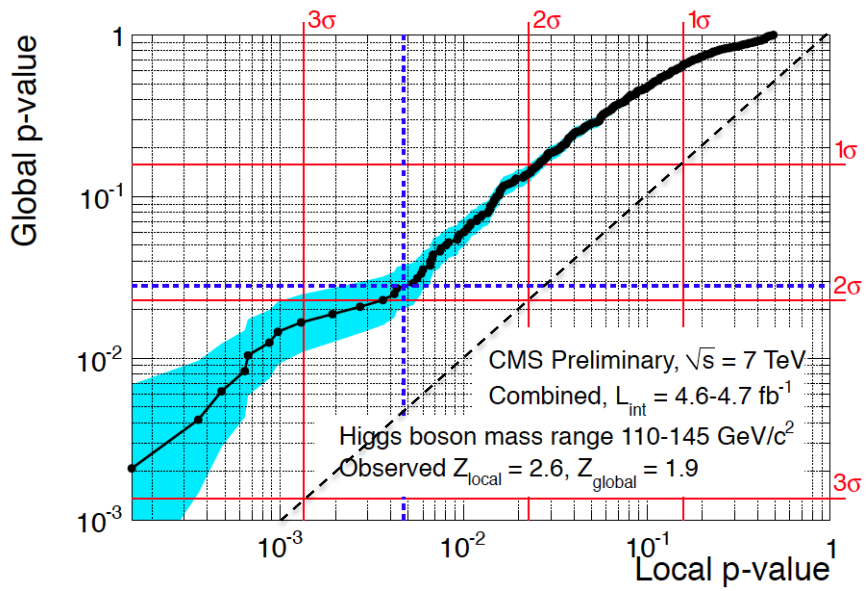
With Look Elsewhere Effect (LEE) In Full Mass Range



- Maximum local significance: $Z_{max} = 2.6$
- Count number of upcrossings at $\sigma/\sigma_{SM} = 0$: $N_0=8$
- Global p-value can be estimated as $p_0^{\text{global}} \sim p_0^{\text{min}} + N_0\, e^{-\frac{1}{2} Z_{\text{max}}^2}$

- This corresponds to global p-value=0.28, global Z=0.6

With Look Elsewhere Effect (LEE) In Low Mass Range

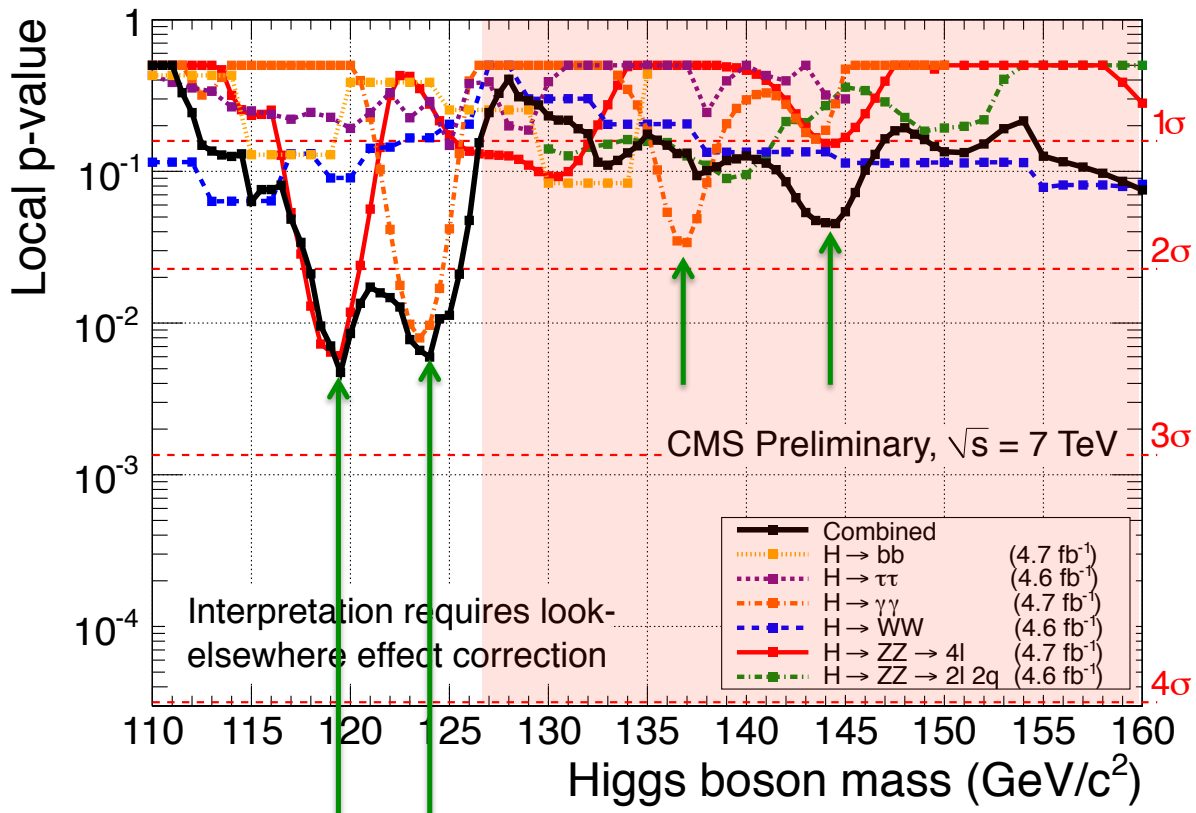


Restricted mass range: 110-145 GeV

Generated 500 pseudo-data sets

$Z_{local} = 2.6\sigma \rightarrow Z_{global} = 1.9\sigma$

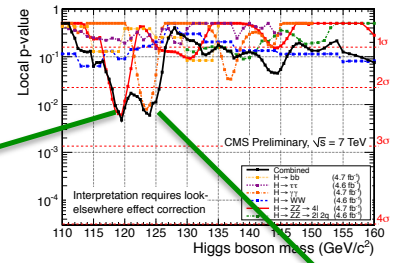
Investigating Observed Excesses

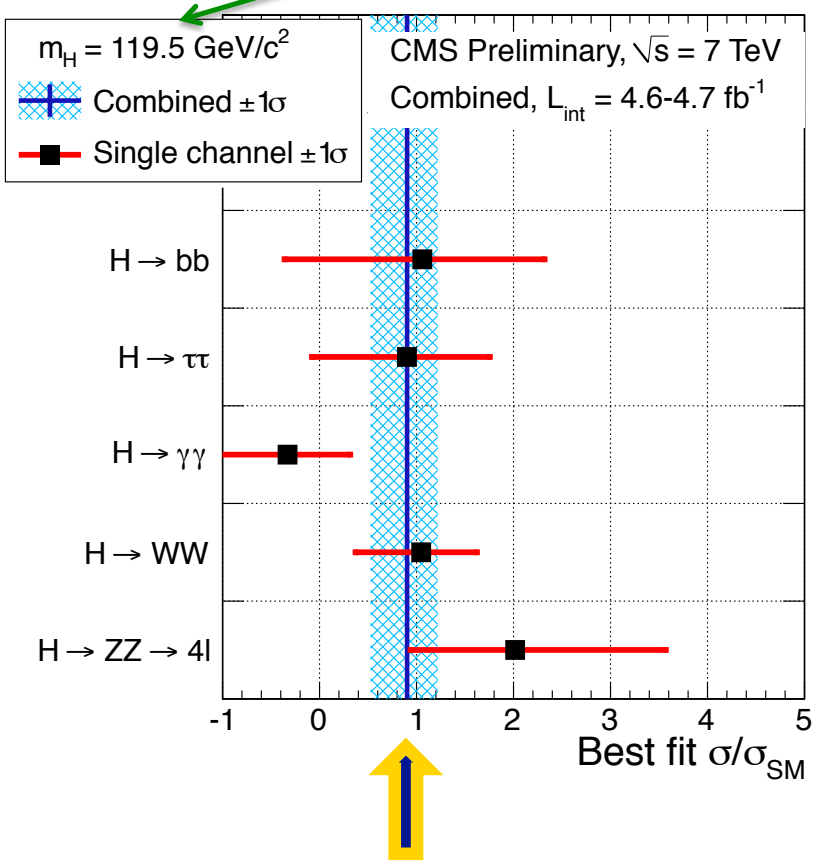


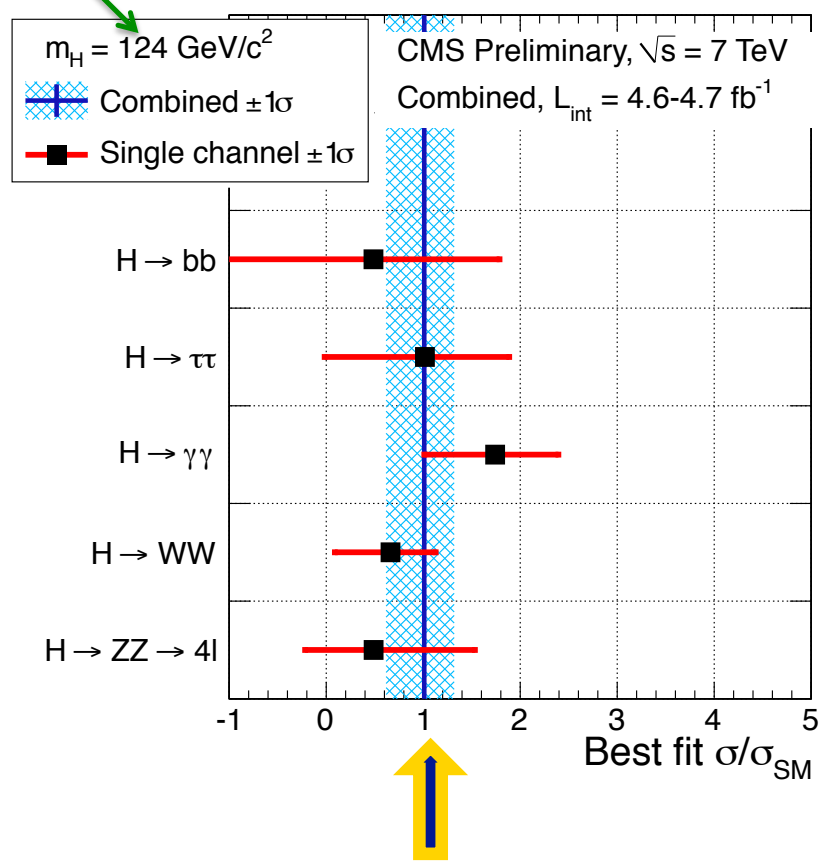
three 4l events
But deficit in γγ

excess in γγ
two 4l events not too far

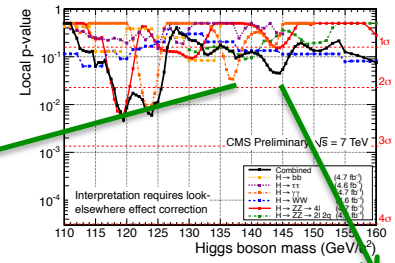
Observed Rate Of Excesses at M_H= 119.5, 124 GeV

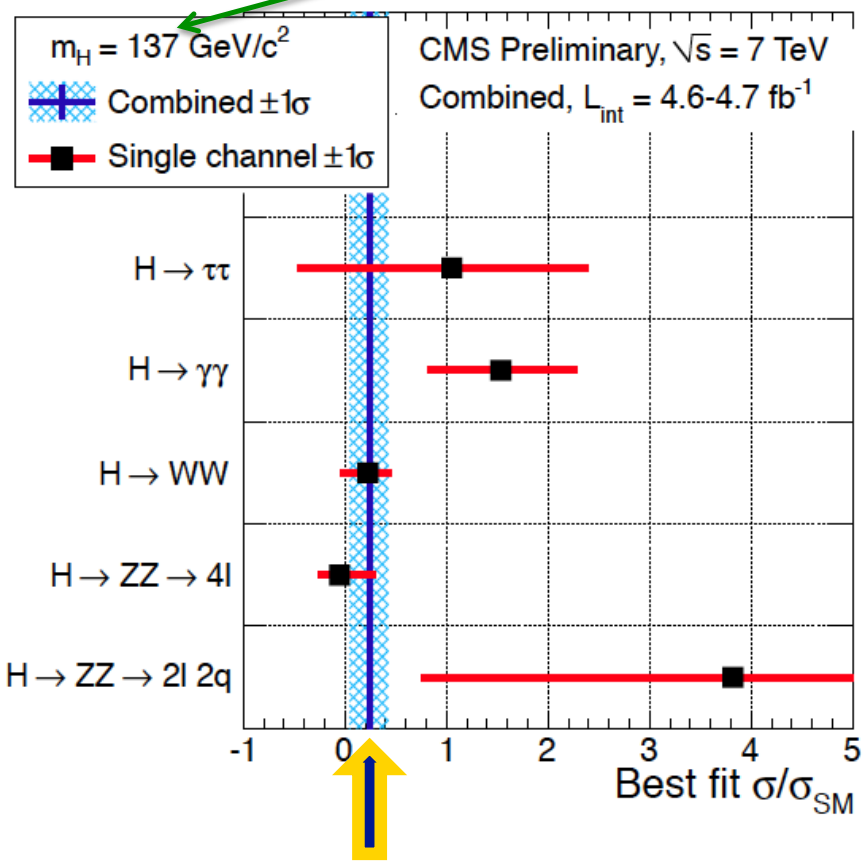


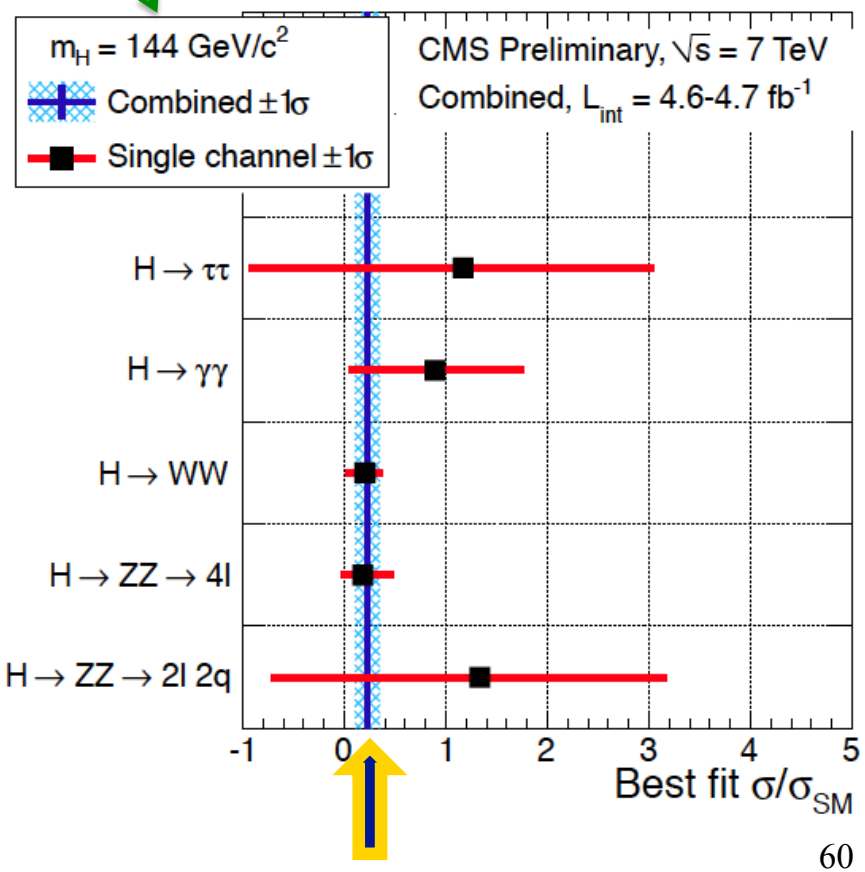




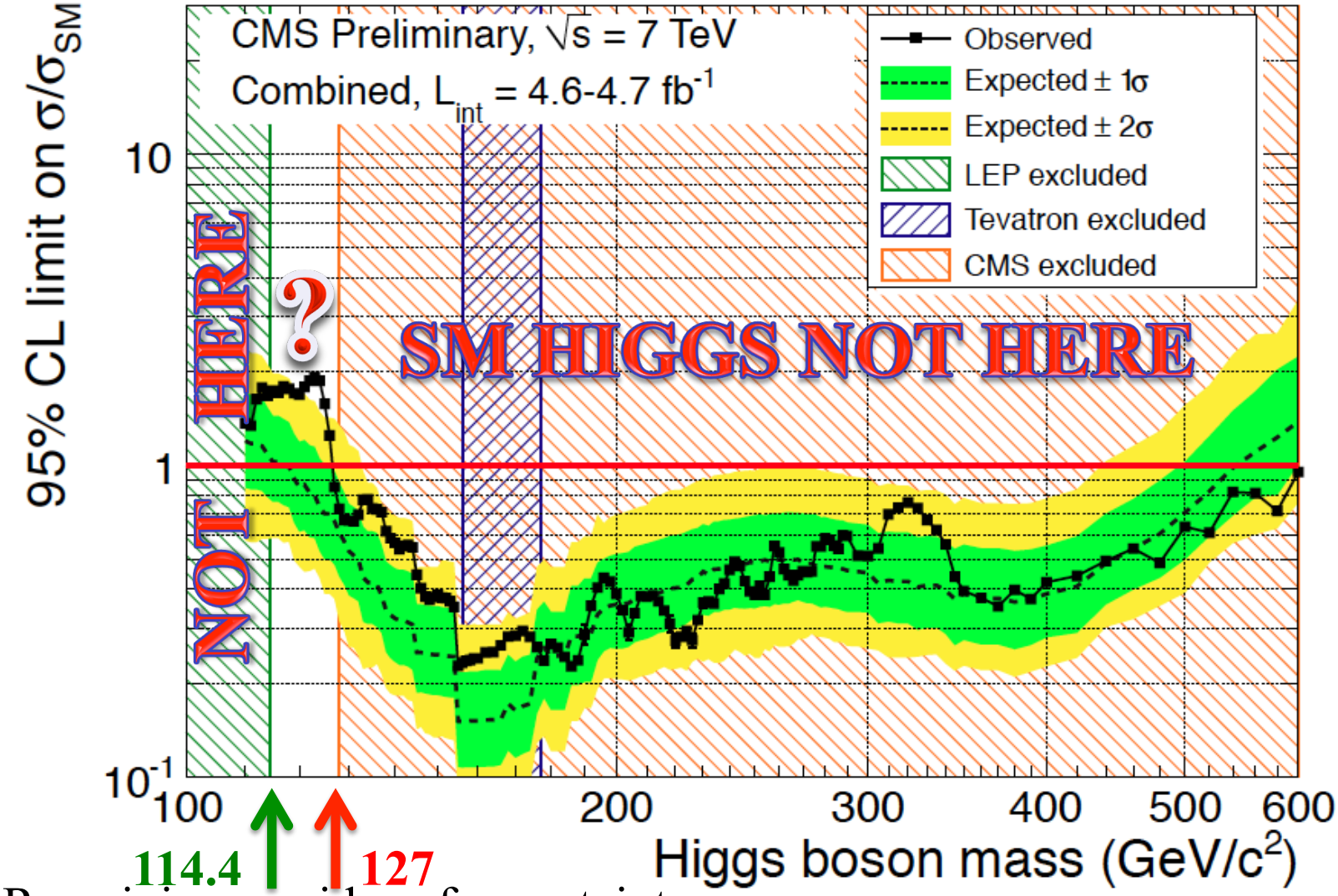
Observed Rate Of Excesses at M_H =137, 144 GeV





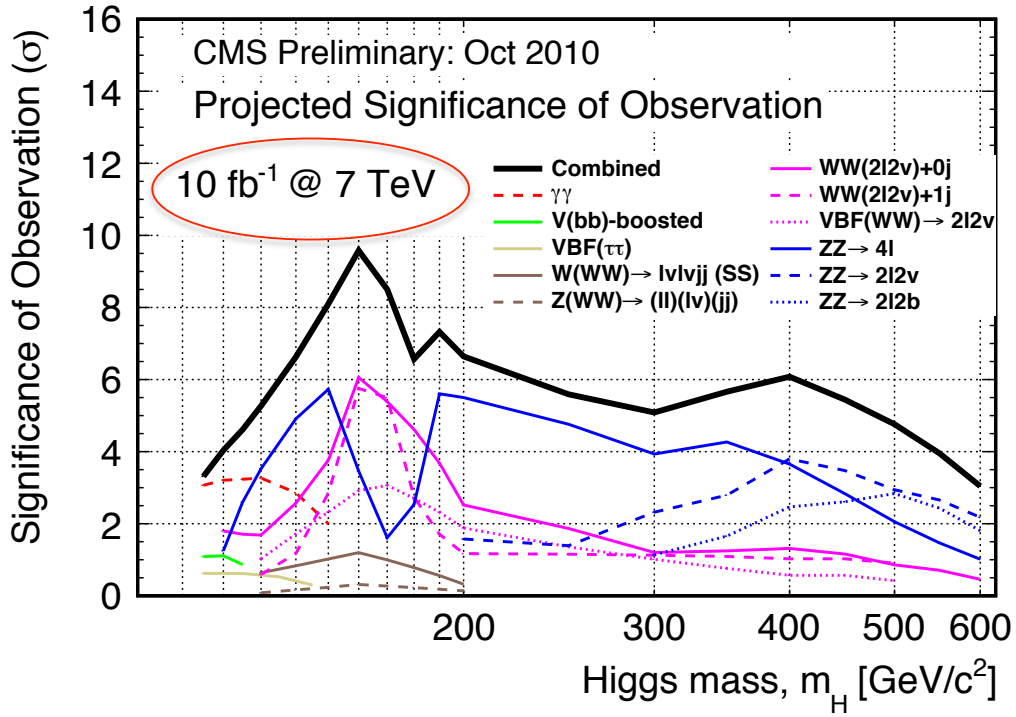


Summary Of CMS 2011 Higgs Searches



Remaining corridor of uncertainty

What 2012 LHC Run Will Tell Us



Expect $\approx 20\ fb^{-1}$ data in 2012

If SM Higgs exists, it will be discovered this year

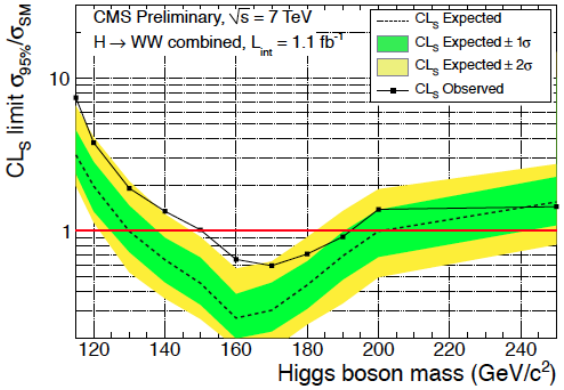
Else it will be ruled out over the 114-600 GeV mass range

Summary Of CMS Higgs Boson Search

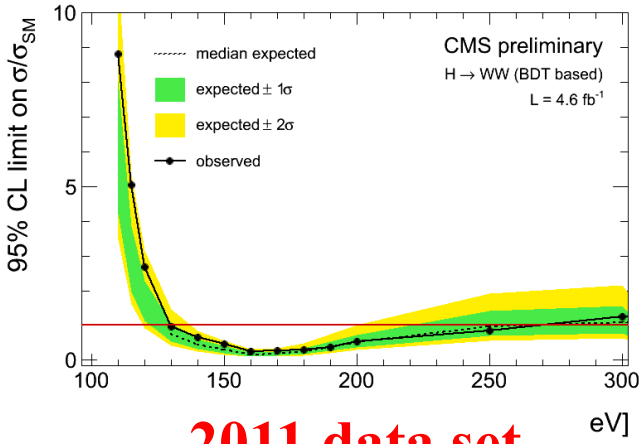
- CMS has searched for SM Higgs in 8 distinct Higgs boson decay channels with up to 4.7 fb^{-1} data
- The Higgs search mass range spans 110-600 GeV
- Expected 95% CL exclusion range, in absence of Higgs, is **117-543 GeV**
- Observed 95% CL exclusion range is: **127-600 GeV**
- A small excess in the low mass range seen
- More data are required to determine the source of this excess.
- Hopefully the Hunt for the SM Higgs will come to a decisive end in 2012

Backup Slides

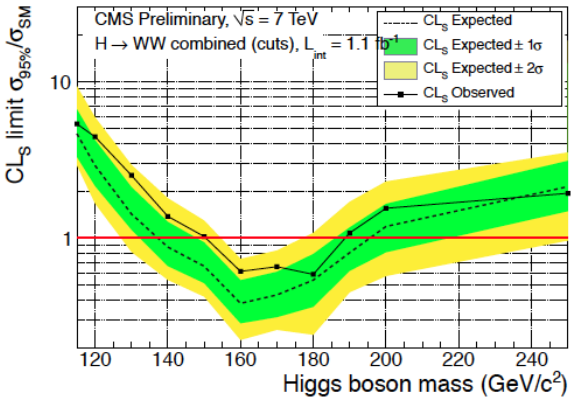
H->WW->lvlv: Evolution of Search Limits



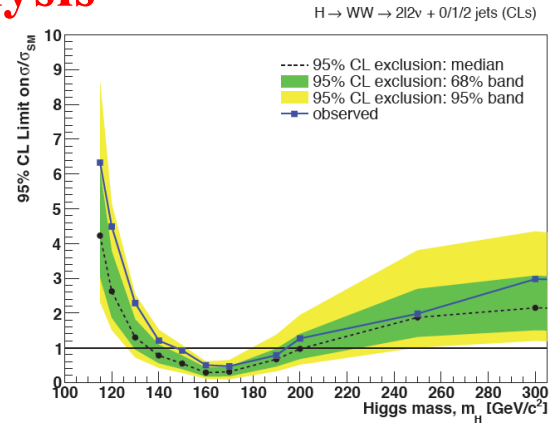
EPS'11
MVA-based analysis



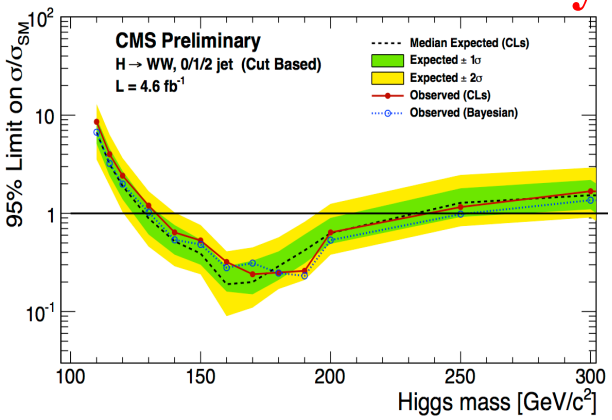
2011 data set
MVA-based analys



EPS'11
Cut-based analysis



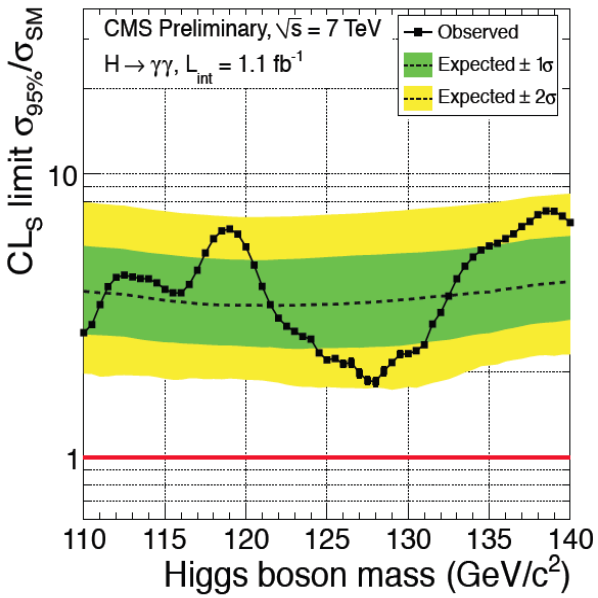
LP'11
Cut-based analysis



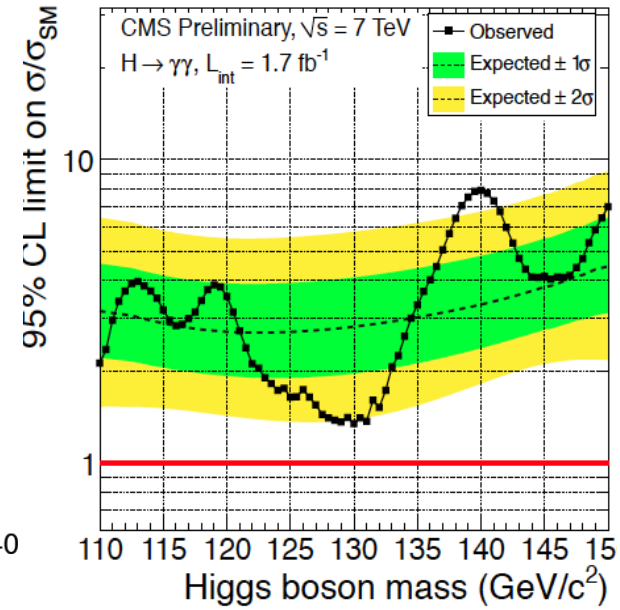
2011 data set
Cut-based analysis

H->γγ:Evolution Of Search Limits

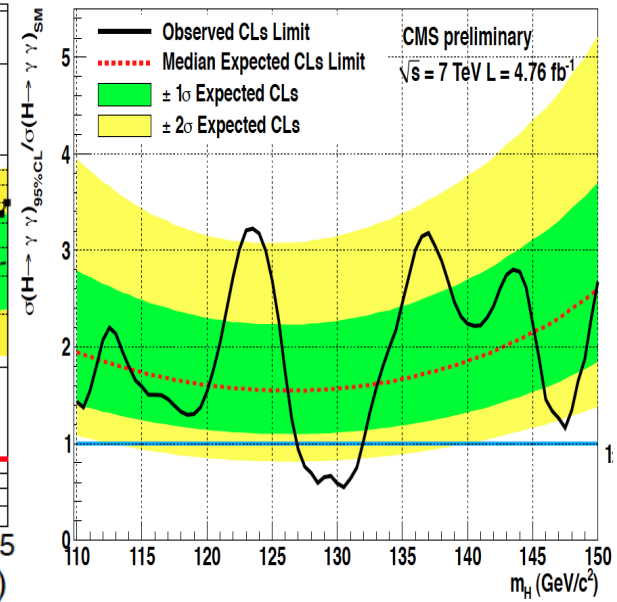
EPS'11



LP'11

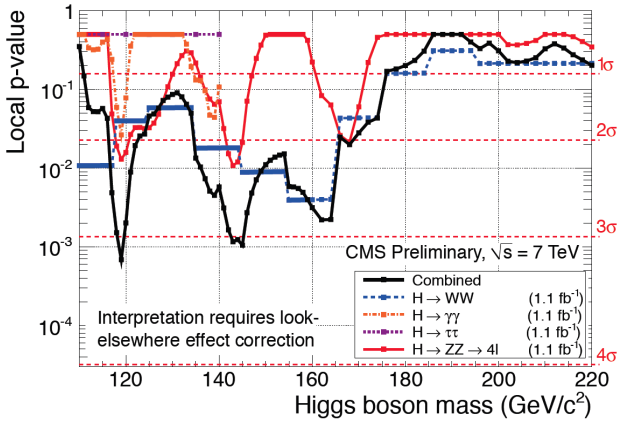


2011

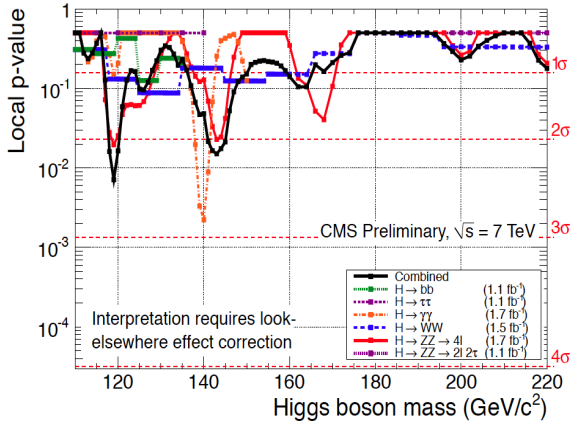


Evolution Of CMS Higgs Combination in 2011

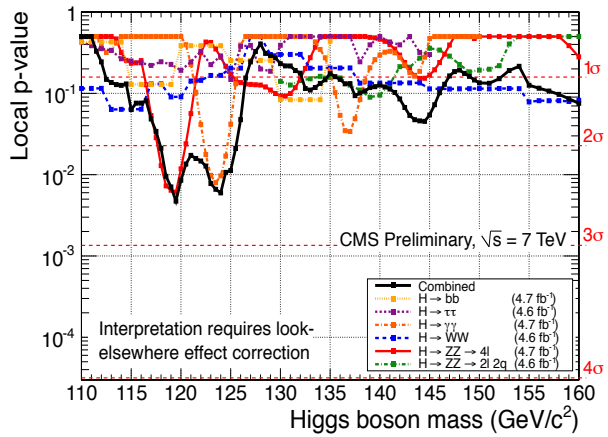
EPS'11



LP'11



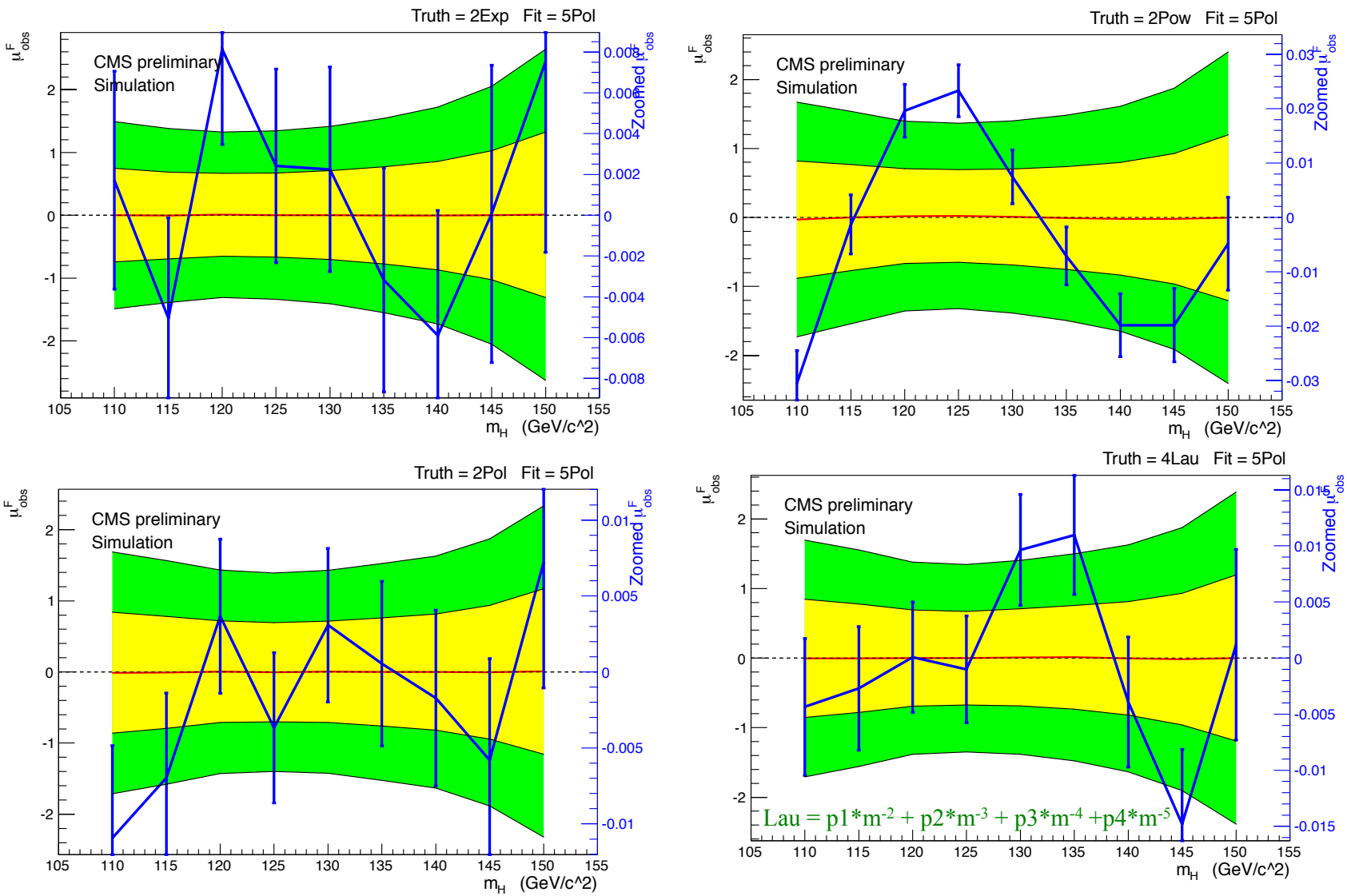
2011 data



$H \rightarrow \gamma\gamma$: Studies of Background Model

- The distribution of background events in $M_{\gamma\gamma}$ is assumed to behave as a continuous function. Since no prior knowledge is assumed of the specific form of this function we choose a parameterization which is demonstrably flexible enough to describe any reasonable form without introducing a bias. Studied several test functional forms all of which provide good fits to MC and data.
- Choose some functional form as truth model (used here 2nd Order polynomial, double exponential, double power law and 4-term "partial Laurent series")
- Fit chosen truth form to background MC scaled to $5fb^{-1}$ and throw **background only** toys.
- Fit test functional form + signal model and evaluate signal strength μ (in units of σ_{SM})
- **Bias defined as mean value of signal strength μ, should be 0 for background only toys**

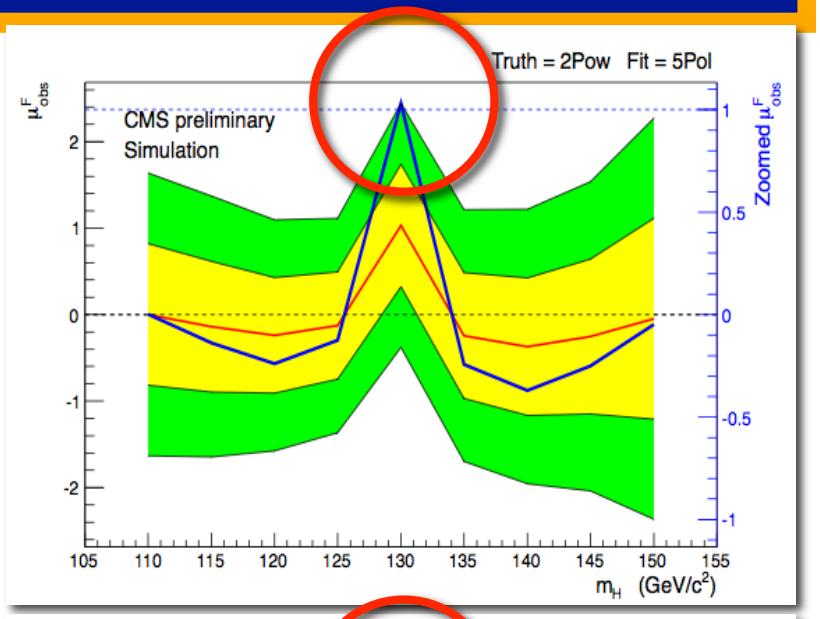
Robustness of Background Model (5th Order Poly)

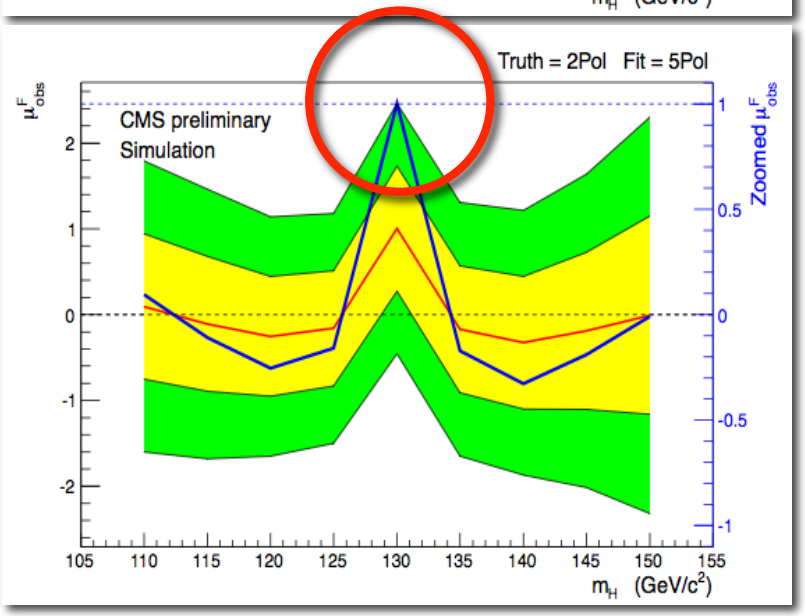


No perceptible bias, small loss in sensitivity (not shown)

BG+Signal Toys

- 5th order polynomial showed best performance, both in background-only and in signal+background tests





Fitting With 2 Exponential Background Model

$$2\mathrm{Exp}(m) := f e^{p_1 m} + (1-f) e^{p_2 m}$$

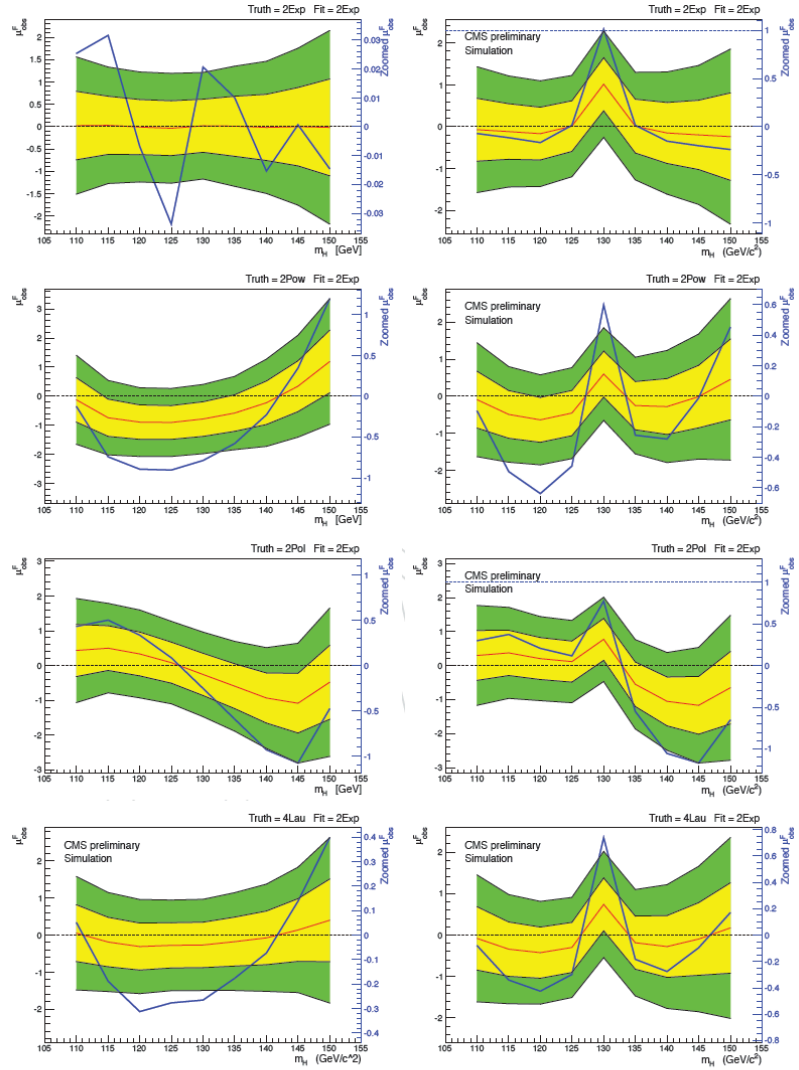


Figure 39: 2Exp fit-model: mass-hypothesis dependent bias on the fitted signal-strength with respect to the several truths (one per line). Left column: BG-only toys. Right column: Signal+BG toys (m_H =130 GeV/c^2).