Search for the Higgs boson in the $ZZ \rightarrow ll\nu\nu$ channel with CMS

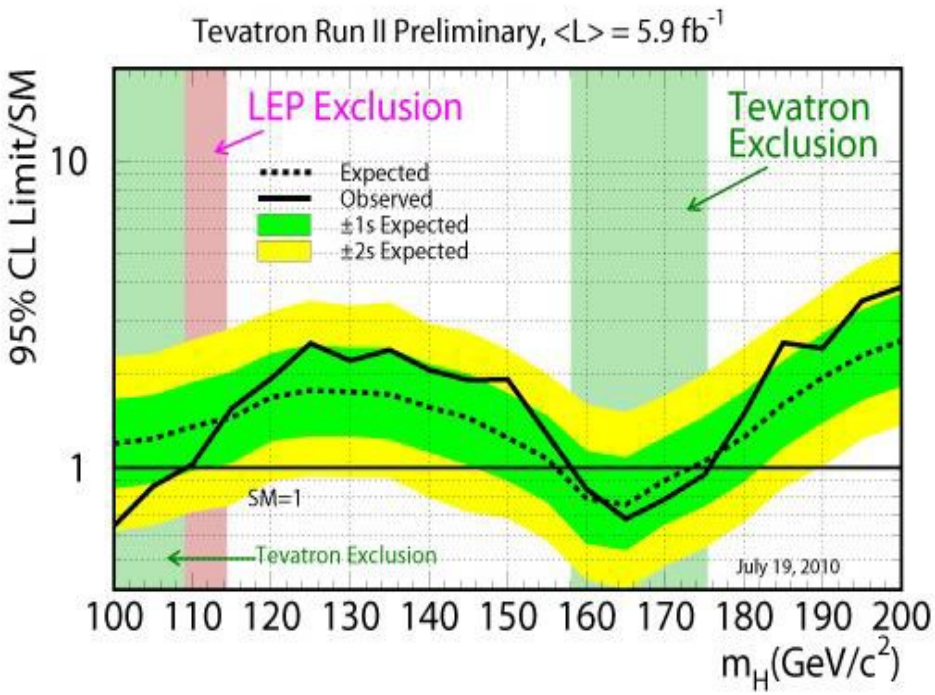
Kevin Sung

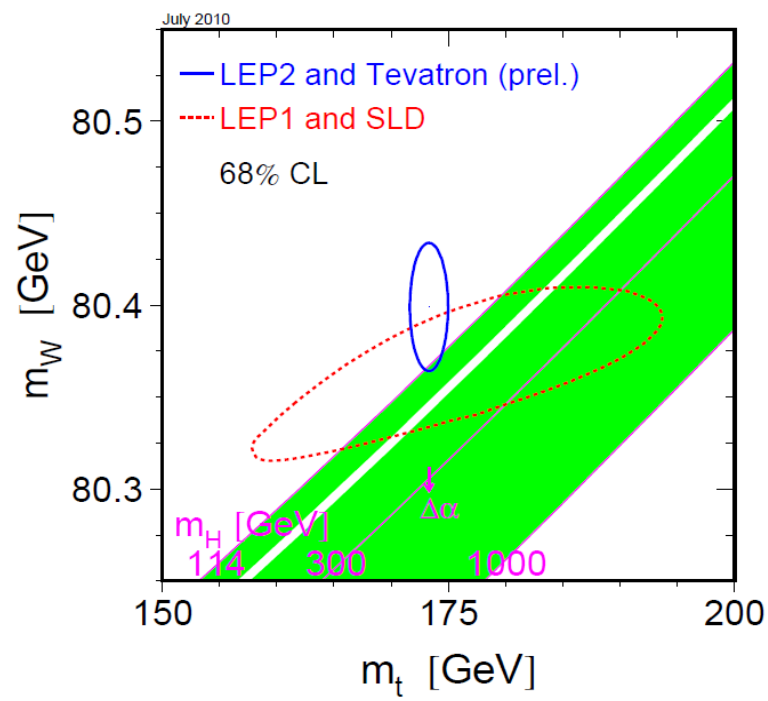
Massachusetts Institute of Technology

FNAL RA Candidate Talk

Introduction

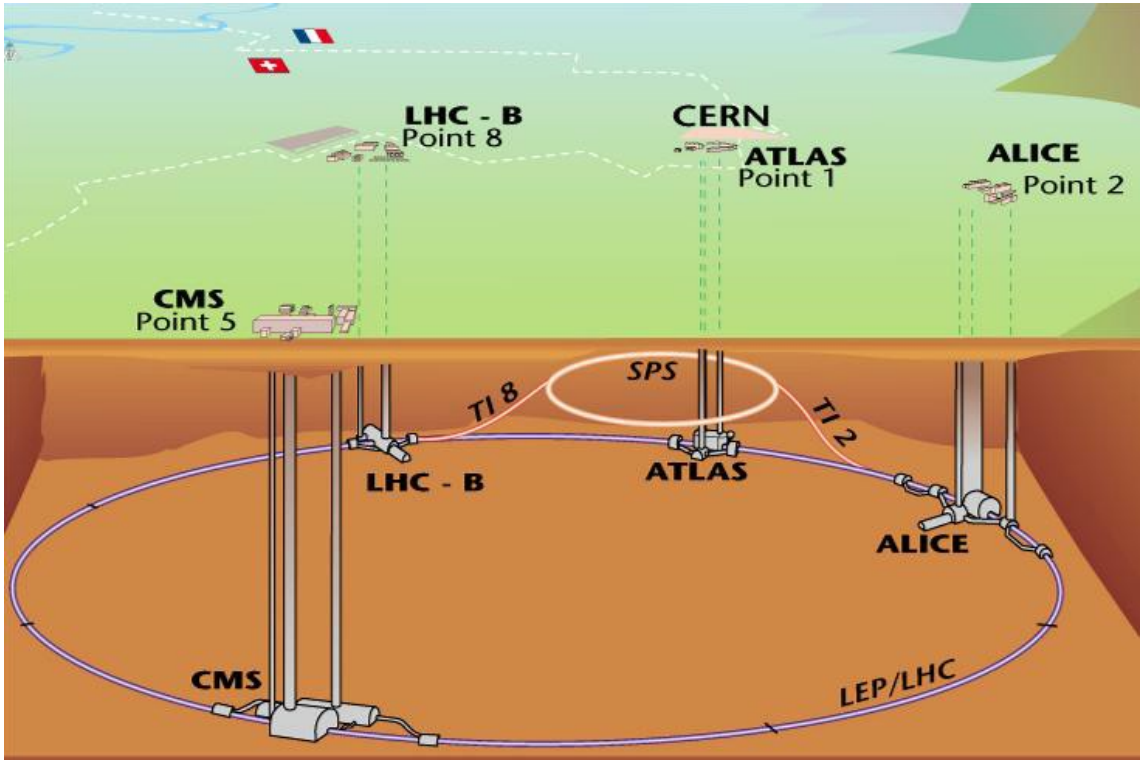
- The search for the Higgs boson is a major component of the LHC physics program
  - Electroweak symmetry breaking mechanism (W and Z masses)
  - Mechanism in Standard Model (SM) to generate fermion masses
    - SM does not predict m_H, constraints from electroweak precision measurements and direct searches at LEP and Tevatron

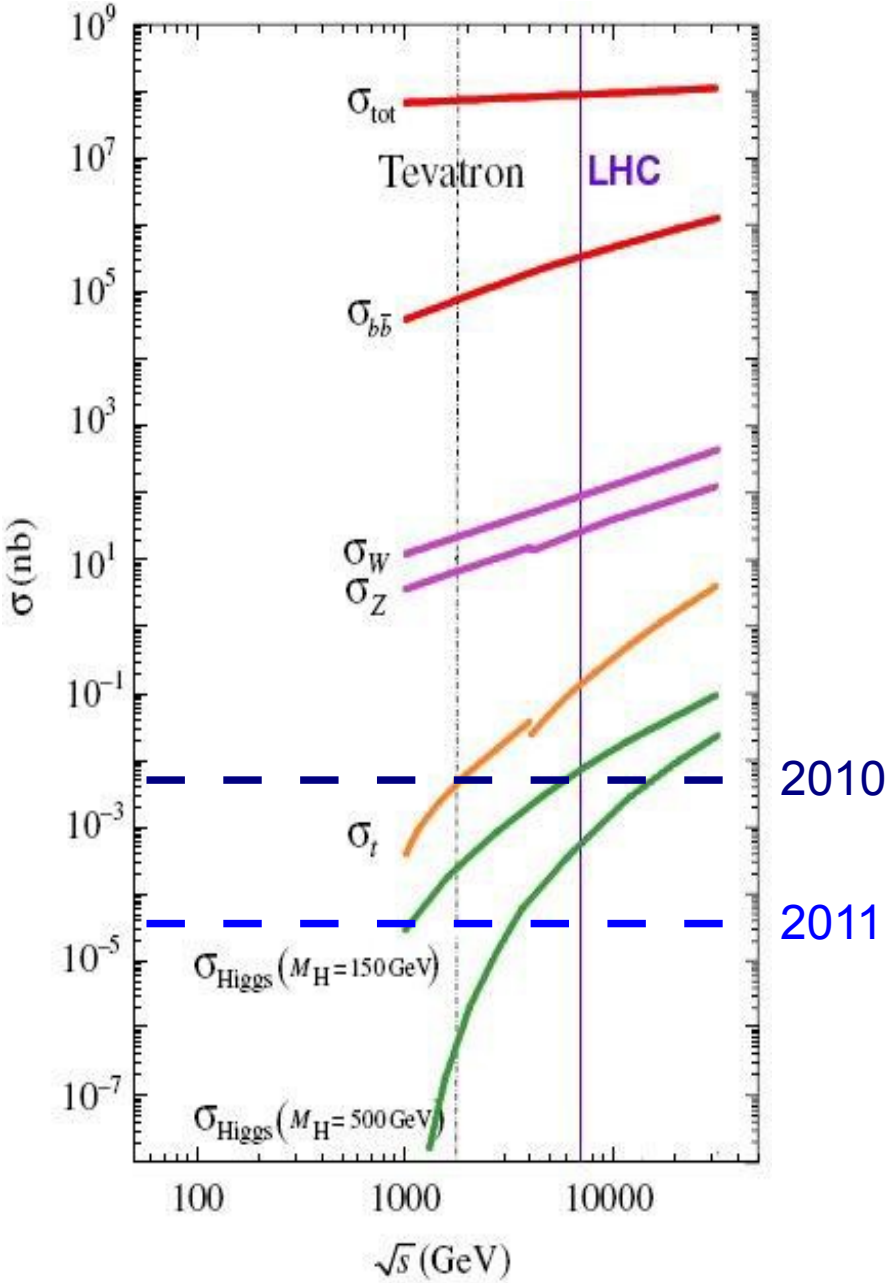




Large Hadron Collider

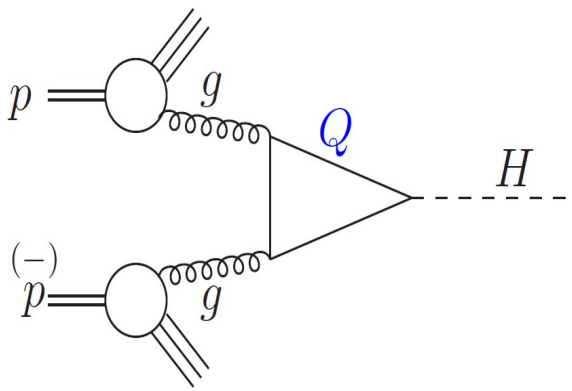
- pp collider at CERN
  - $\sqrt{s} = 7\ TeV$ in 2010 ~ 2011
  - $\sqrt{s} = 8\ TeV$ in 2012
- 4 experiments:
  - ALICE, ATLAS, CMS, LHCb





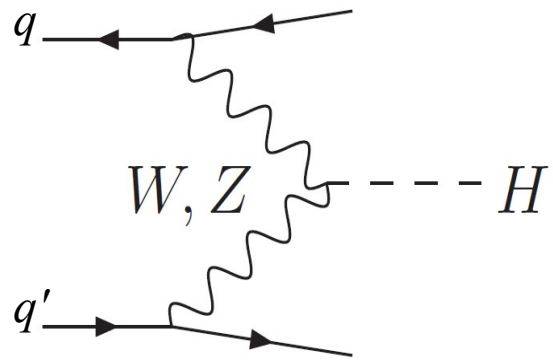
Higgs Production @ LHC

Gluon Fusion



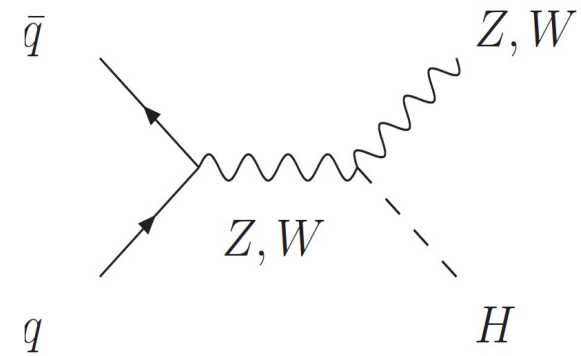
Dominant

Vector Boson Fusion



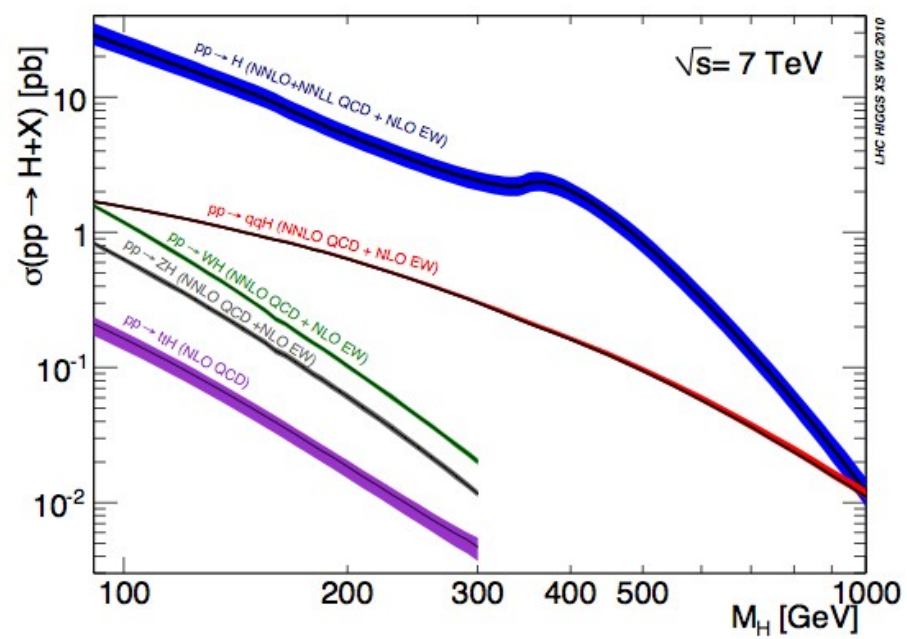
Sub-dominant

Associated Production

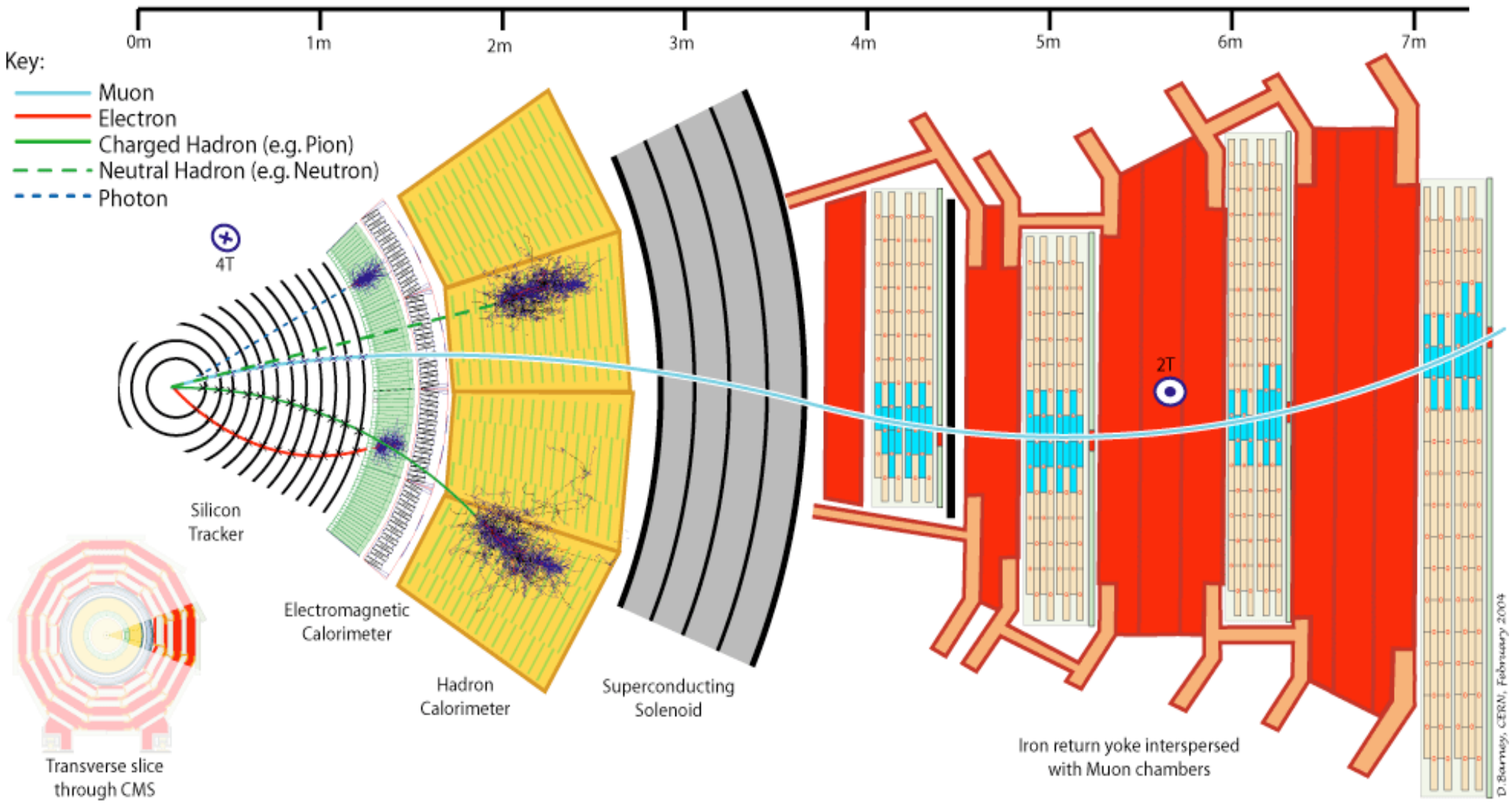


Small rate, but clean signature (useful for $H \to b\bar{b}$)

- Gluon fusion is the most relevant production mode at the LHC (pp collisions)
  - In contrast to associated production ("Higgsstrahlung") at LEP and Tevatron



The CMS Detector

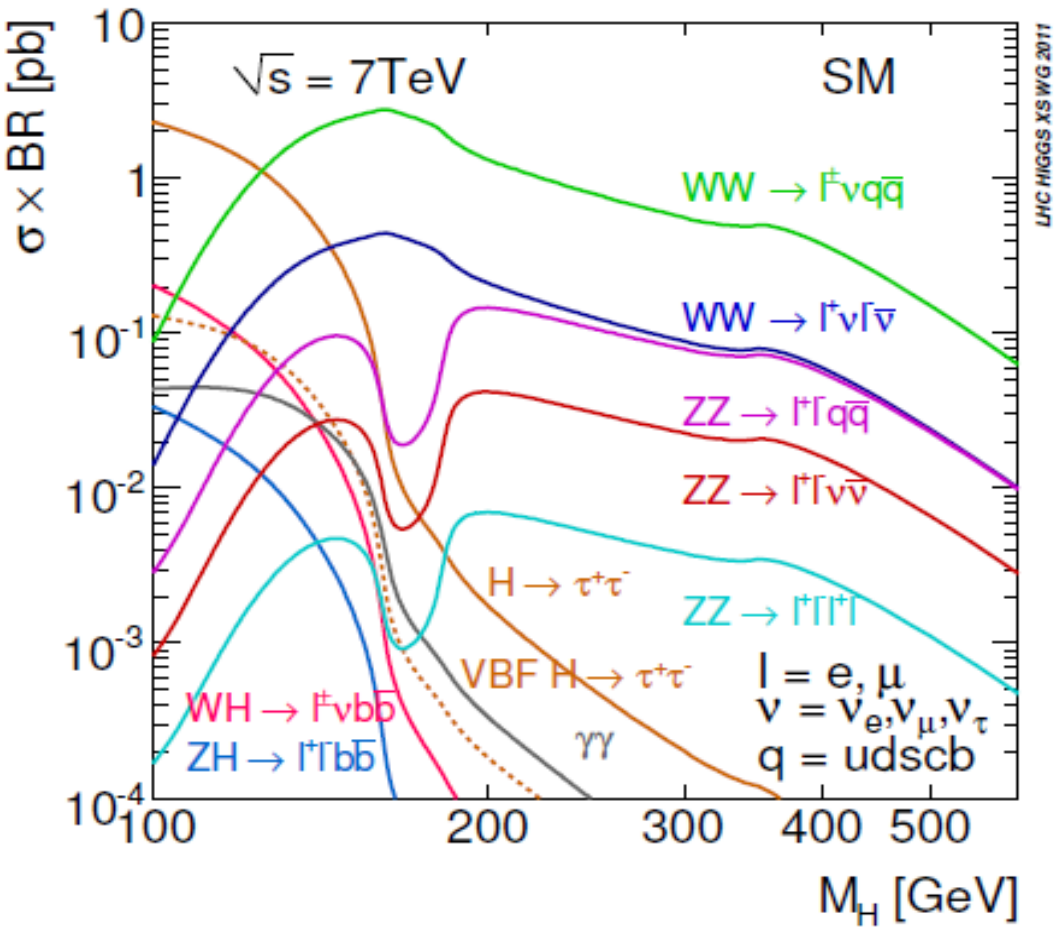


The Search...

$H \rightarrow ZZ \rightarrow ll\nu\nu$

- $H \rightarrow ZZ \rightarrow ll\nu\nu$
  - Clean: leptons in final state, Z mass constraint
  - ~6 x BR($ZZ \rightarrow 4l$)
  - Larger acceptance ($2l$ vs $4l$)

- Search with CMS in 2011:
  - Analyzed $4.6\ fb^{-1}$
  - $H \rightarrow ZZ \rightarrow ll\nu\nu$ analysis targets high mass Standard Model Higgs ($m_H \geq 250\ GeV$)
  - Published in JHEP (DOI: *10.1007/JHEP03(2012)040*)



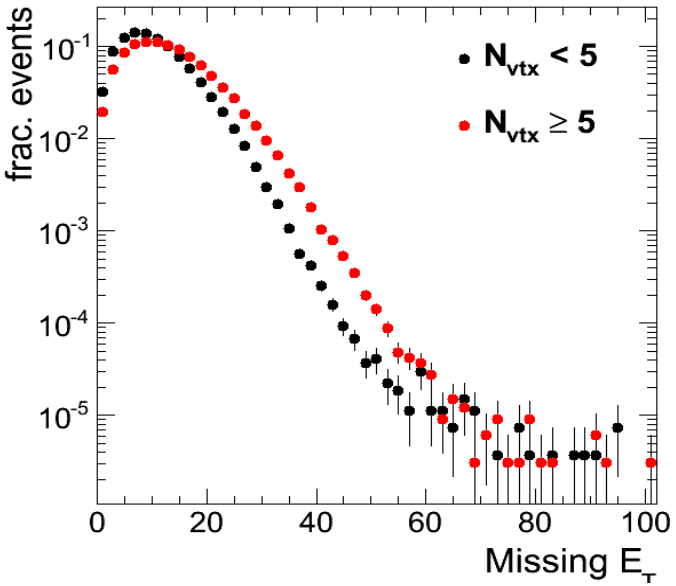
Analysis Strategy

- Select $ZZ \rightarrow ll\upsilon\upsilon$ events:
  - (1) $ee/\mu\mu$ pair consistent with $Z \rightarrow ll$ decay
  - (2) large missing transverse energy (MET) from $Z \rightarrow \upsilon\upsilon$
- Backgrounds:
  - ZZ, WZ, WW, Top, Z+jets
  - $H \rightarrow WW \rightarrow l\upsilon l\upsilon$ treated as background
- *No Higgs mass peak!* Looking for an excess over background expectations...
- Signal extracted in transverse mass (M_T) of ll+MET system

$$M_{\mathrm{T}}^2 = \left(\sqrt{p_{\mathrm{T}}(\ell\ell)^2 + M(\ell\ell)^2} + \sqrt{E_{\mathrm{T}}^{\mathrm{miss}\,2} + M(\ell\ell)^2} \right)^2 - (\vec{p}_T(\ell\ell) + \vec{E}_T^{\mathrm{miss}})^2$$

Pile-up

- To achieve high luminosity, LHC crams as many protons as possible into each bunch...
  - Avg. of ten pp interactions/event in 2011

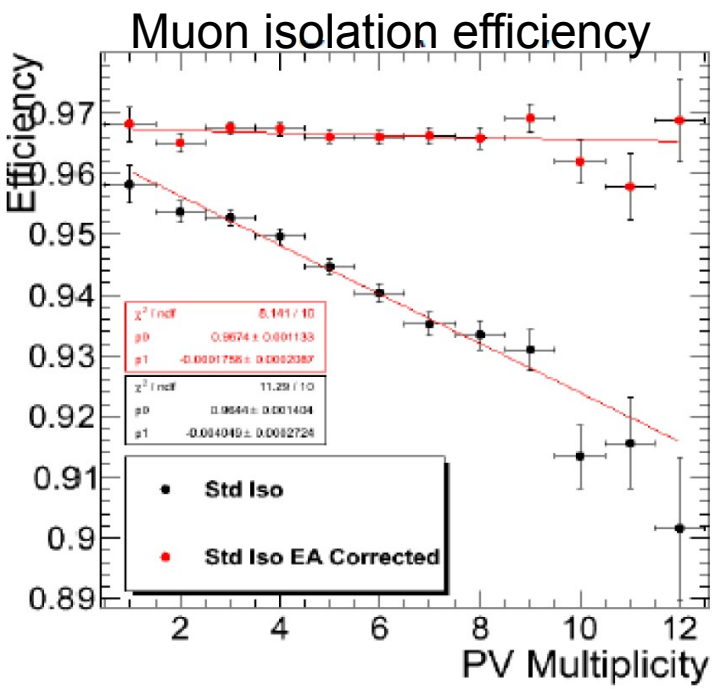


Challenges for analysis:

- Less efficient isolation cuts
- Poorer jet energy resolution
- Poorer MET resolution

Solutions:

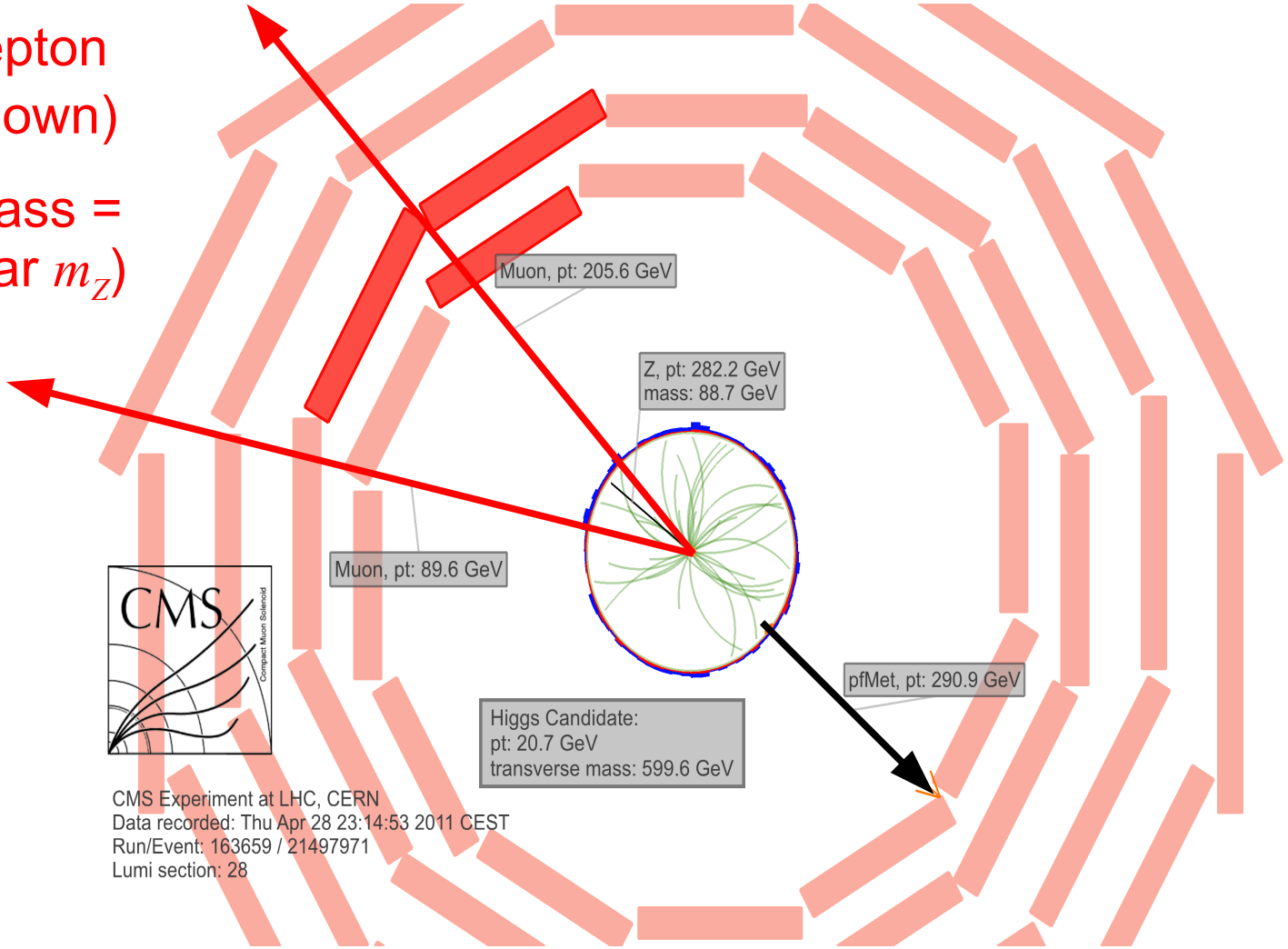
- Exploit excellent tracking resolution to pick out only (charged) particles from primary vertex
- Compute "energy density" to estimate and remove contributions from pile-up



It's even more intense in 2012, but analyzers will adapt!

$H \rightarrow ZZ \rightarrow ll\nu\nu$ Candidate

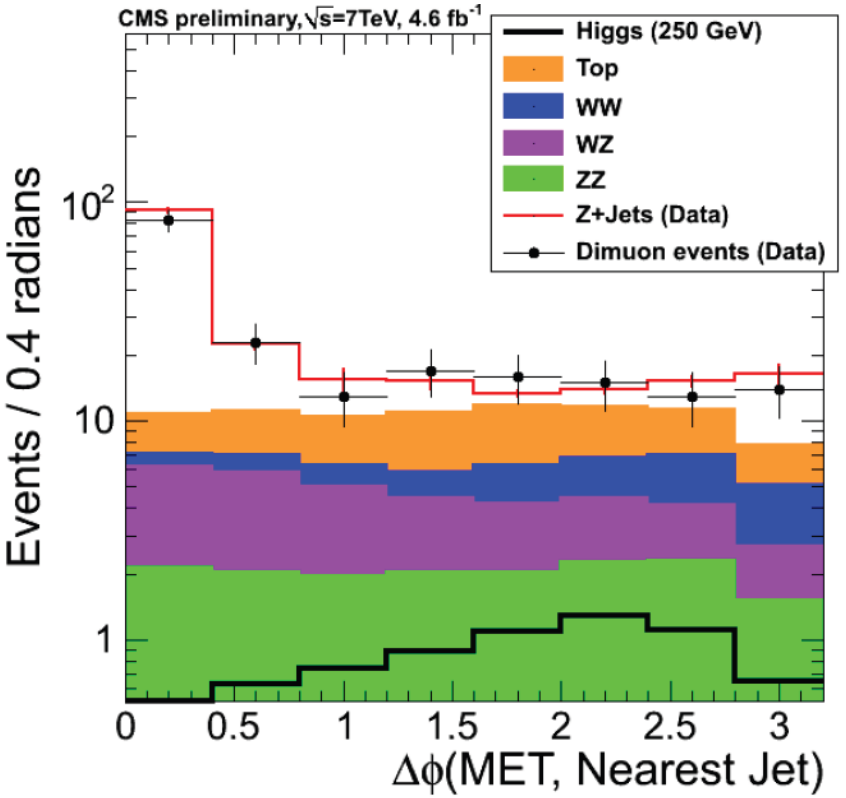
- High p_T dilepton (dimuon shown)
- Dilepton mass = $89\ GeV$ (near m_Z)



- Large MET ($291\ GeV$) and large M_T ($600\ GeV$)

Event Selection

- Acquire events with single/double muon, double electron triggers
  - Per electron (muon) efficiency of ~99% (~95%)
- Two well identified, isolated leptons (e or μ) of opposite charge, $p_T > 20\ GeV$
  - $|m_{ll} - m_Z| < 15\ GeV$
  - Dilepton $p_T > 55\ GeV$
  - MET $> 70\ GeV$
- Reject events with:
  - 3rd well identified, isolated leptons with $p_T > 10\ GeV$ to reduce WZ, ZZ
  - Identified b-decays to reduce top
  - A jet aligned with MET ($\Delta\varphi<0.5$) to reduce Z+jets



Background Estimation

Background Estimation

ZZ, WZ Background

- Estimate using samples generated with MadGraph
  - Scaled to NLO cross section

Non-resonant backgrounds

- Top, WW, $Z \rightarrow \tau\tau$, and $H \rightarrow WW \rightarrow l\upsilon l\upsilon$
- Normalize from control sample of $e\mu$ events under Z peak

Z+jets

- Use γ+jets events to model Z+jets

$e\mu$ Method

Non-resonant backgrounds

- Top, WW, $Z \rightarrow \tau\tau$, and $H \rightarrow WW \rightarrow l\upsilon l\upsilon$
  - Dominated by top and WW
- Such processes are lepton flavor symmetric
- Procedure:
  - Count $e\mu$ events under Z peak:

$$N_{\mu\mu} = \alpha_\mu \times N_{e\mu}, \qquad N_{ee} = \alpha_e \times N_{e\mu}$$

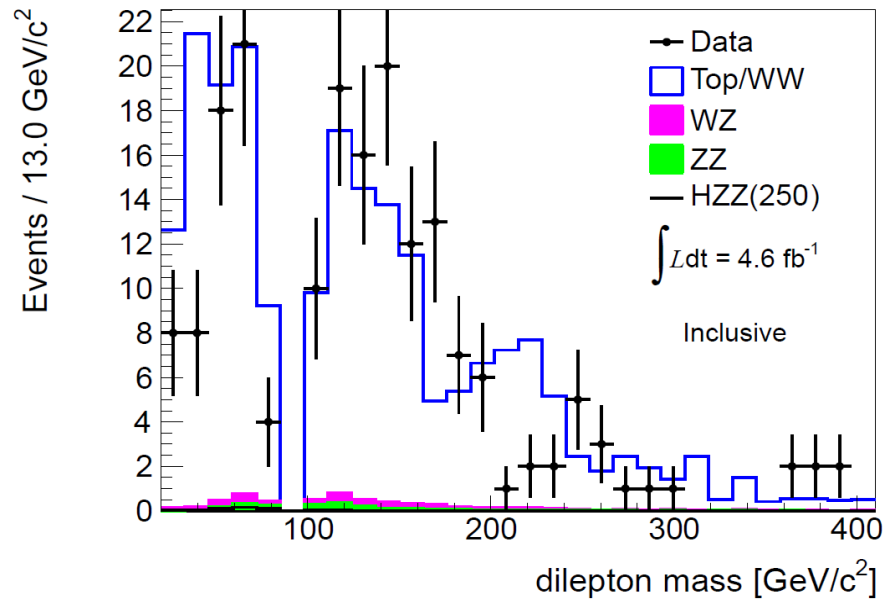
  - α_μ and α_e include combinatoric factor (½) and correction factor for e/μ efficiency differences, obtained from ee and $\mu\mu$ mass side-bands:

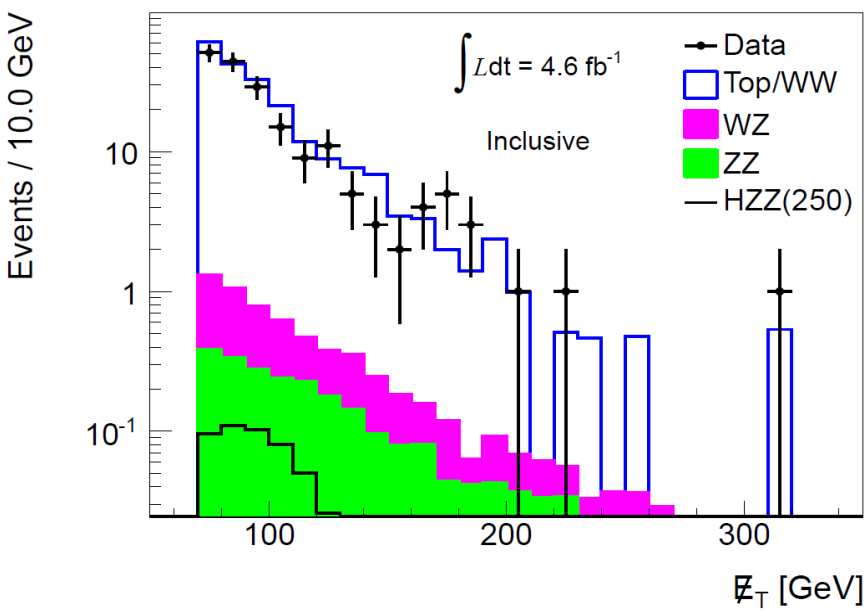
$$\alpha_e = N_{ee}/N_{e\mu} = 0.42 \pm 0.02$$

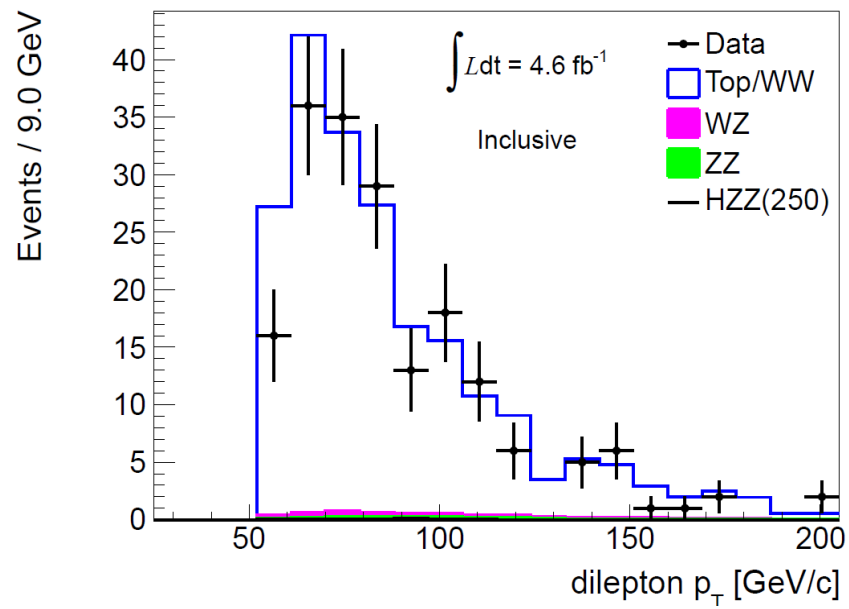
$$\alpha_\mu = N_{\mu\mu}/N_{e\mu} = 0.58 \pm 0.02$$

Top/WW Control Region

- Define a top and WW dominated control region by looking outside the mass window:
  - $|m_{ll} - m_Z| > 15\ GeV$
  - $M_T > 150\ GeV$ (rejects most $Z \rightarrow \tau\tau$)
- Check that $e\mu$ method gives sensible estimate of top and WW
- Dielectron channel shown here





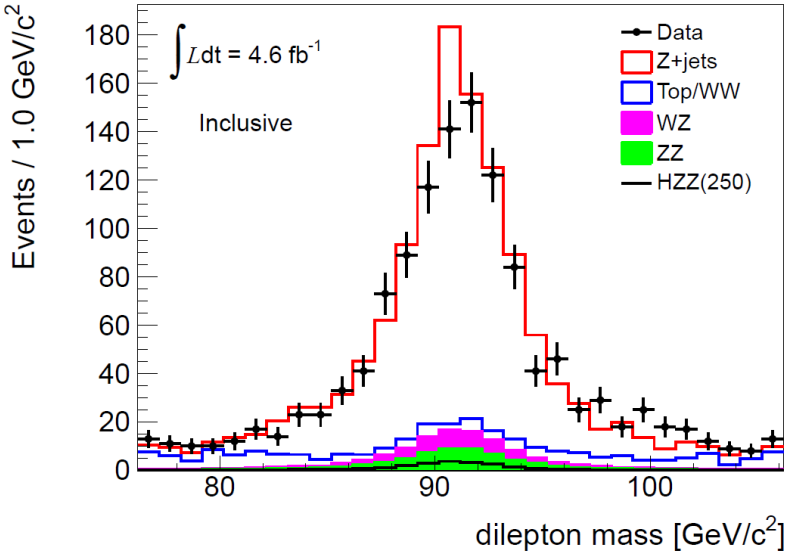


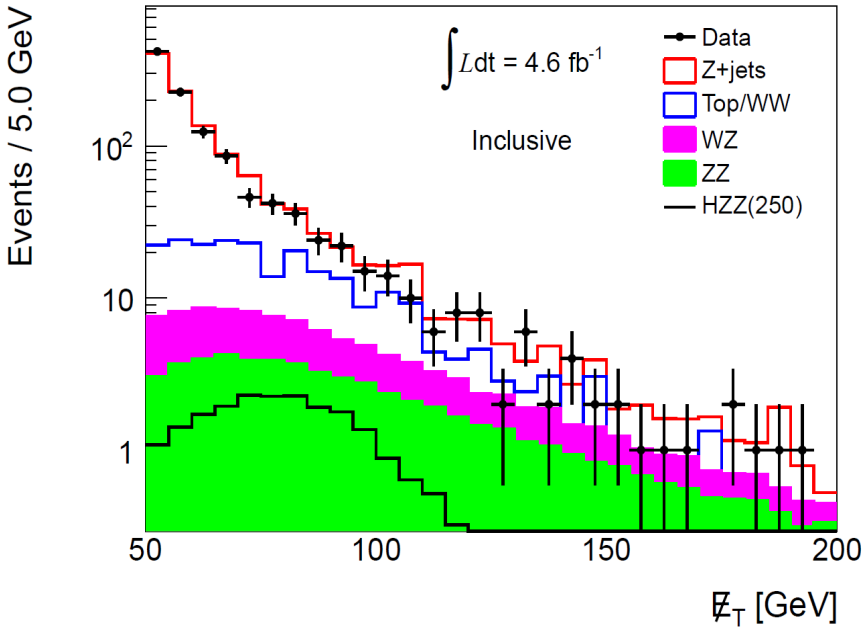
γ+jets Method

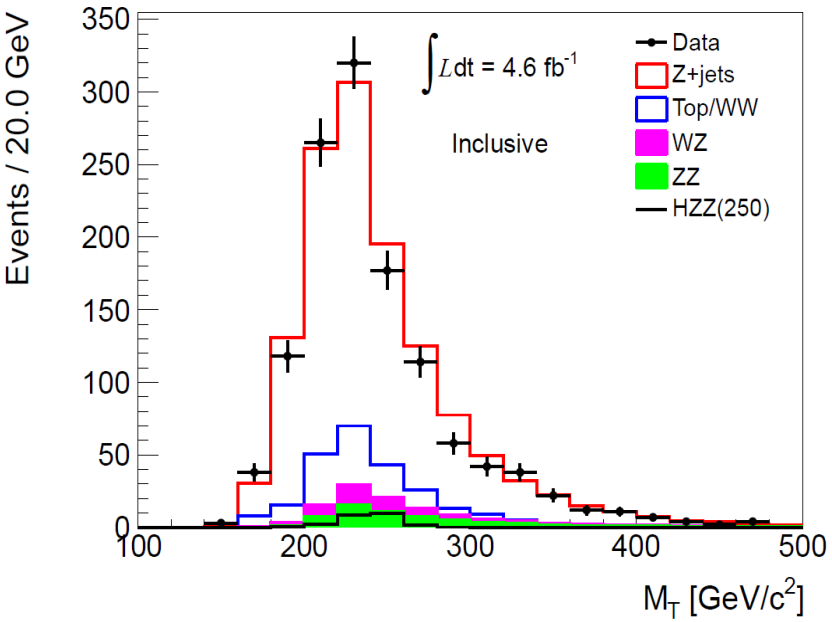
- Use γ+jets events to model Z+jets
  - Assumption: “fake” MET from response/resolution of hadronic recoil
- Select single photon events with $p_T > 55\ GeV$
- Need to re-scale photon sample to mimic Z events:
  - p_T spectra are different
  - Photon events obtained via triggers that have luminosity dependent pre-scales
  - Use data with MET $< 50\ GeV$ to derive photon-dilepton weights in p_T, N_{jet}, and N_{PV}
- Estimate of Z+jets from weighted photon events passing analysis selection (except lepton and mass cuts)
  - M_T computed by assigning a “mass” from sampling simulated Z events

Z+jets Control Region

- Define a Z+jets dominated control region by removing $\Delta\varphi$(jet, MET) cut and applying loose MET and M_T requirements:
  - MET $> 50\ GeV$
  - $M_T > 150\ GeV$
- Dimuon channel shown here

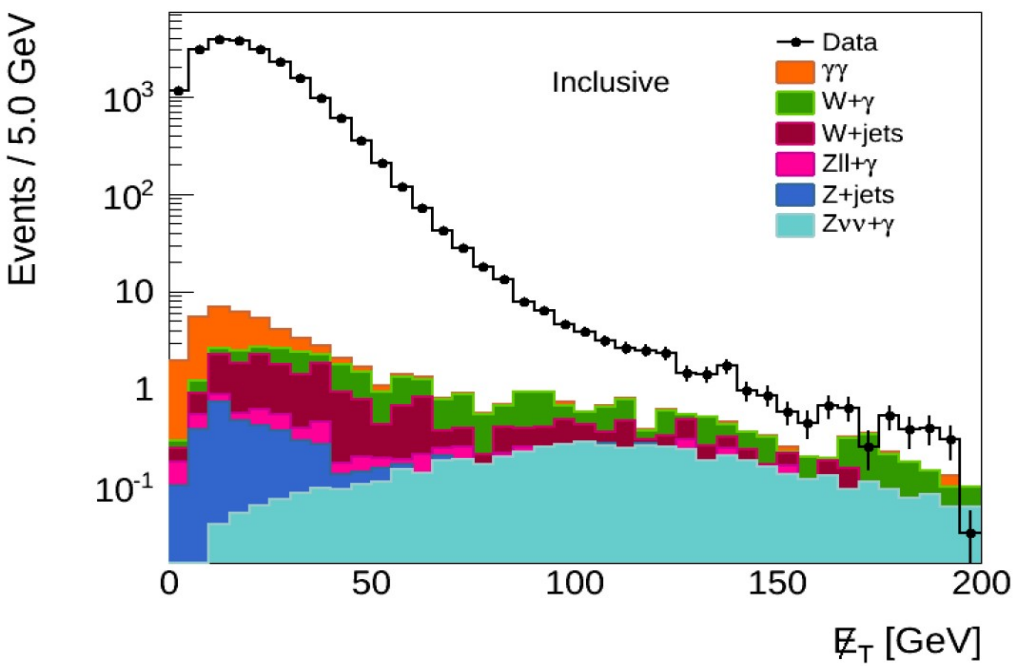






Contamination in γ+jets

- Photon sample contains events with real MET (ex. $W\gamma \rightarrow l\nu\gamma$ and $Z\gamma \rightarrow \nu\nu\gamma$), so *estimate is an upper bound*
- Difficult to define a control region in data to have a handle on normalization
- From simulation, estimate purity of photon sample with MET $> 70\ GeV$ is ~60%



MET distribution of weighted photon sample with expected contamination

- Z+jets prediction is taken as half of γ+jets estimate with 100% relative uncertainty

Signal Extraction

M_T-shape Analysis (1)

$$M_{\mathrm{T}}^2 = \left(\sqrt{p_{\mathrm{T}}(\ell\ell)^2 + M(\ell\ell)^2} + \sqrt{E_{\mathrm{T}}^{\mathrm{miss}\,2} + M(\ell\ell)^2}\right)^2 - (\vec{p}_T(\ell\ell) + \vec{E}_T^{\mathrm{miss}})^2$$

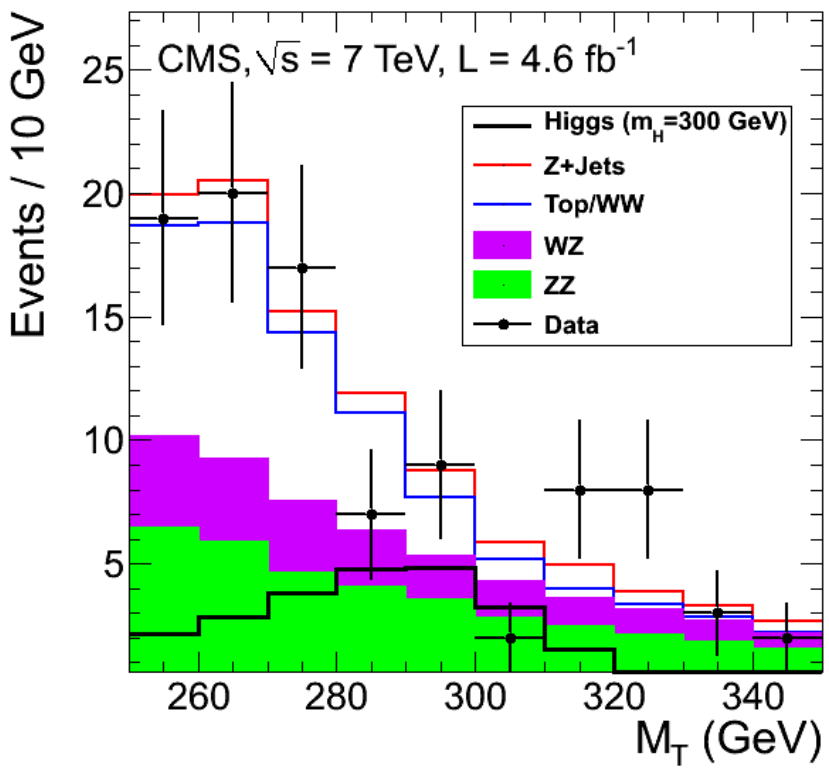
- M_T is transverse mass of dilepton+MET system
  - Assumption: MET from di-neutrino decaying from Z
- <u>Perform a shape analysis:</u> use differences in M_T distribution between signal and background for additional discrimination, improved sensitivity over cut-and-count
  - Final results from binned likelihood fit of M_T
  - Shape templates for SM backgrounds derived from background estimation methods previously discussed

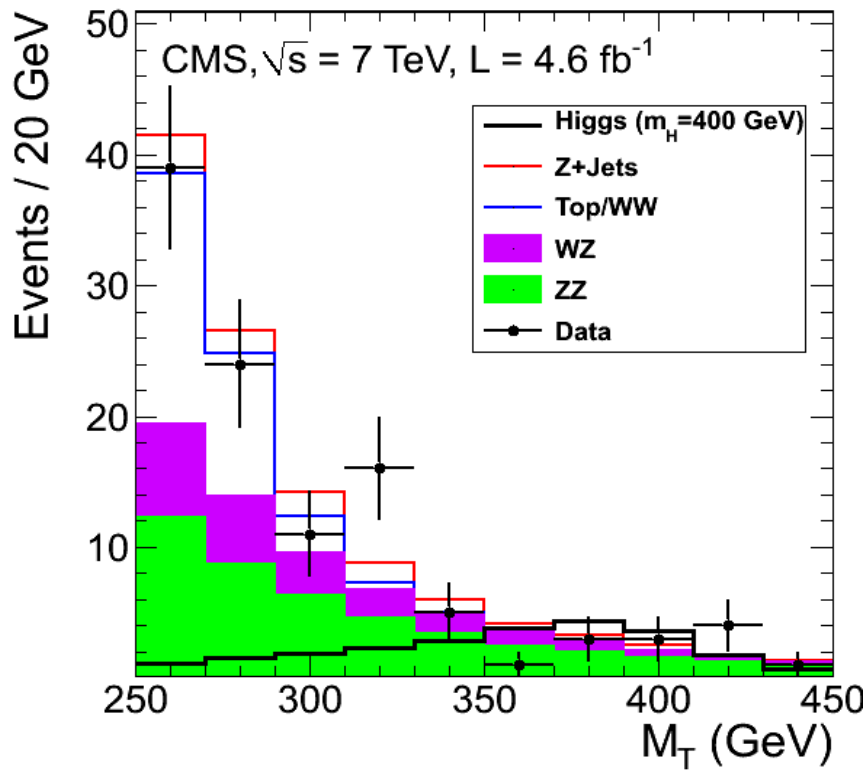
M_T-shape Analysis (2)

- Cuts on MET and M_T dependent on m_H hypothesis

m_H (GeV)	250	300	350	400	500	600
E_T^{miss} (GeV)	> 70	> 80	> 80	> 80	> 80	> 80
M_T (GeV)	[180, 300]	[250, 350]	[250, 400]	[250, 450]	[250, 600]	[250, 750]

- M_T is better discriminator for higher m_H...





M_T-shape Systematics (1)

- Systematic uncertainties in the shape analysis can be categorized to 3 types:

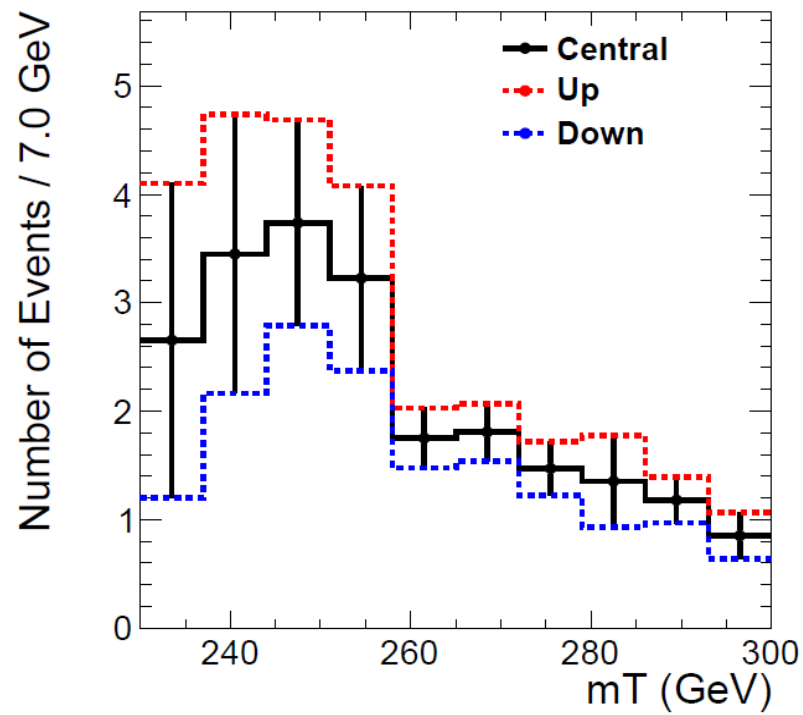
  (1) Normalization: uncertainty on overall shape normalization

  (2) Statistical: account for bin-by-bin statistical uncertainties

  (3) Shape Variation: uncertainty on shape itself

- (1) is accounted for by the uncertainty on predicted yields
- (2) is estimated by varying bins in a correlated way by the statistical uncertainty in each bin

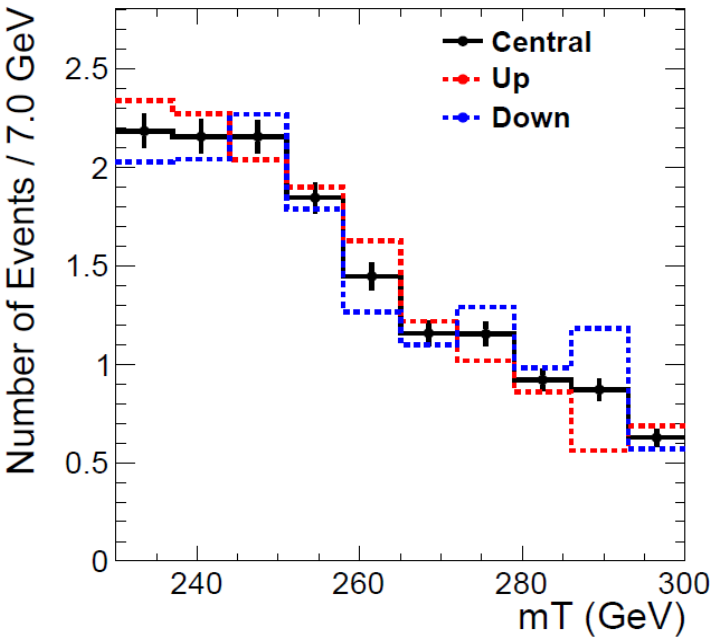
  - "Bounding shapes" to evaluate bin-by-bin statistical uncertainty for Z+jets background
  - The "up" ("down") shape is constructed by varying each bin +1σ (-1σ)



M_T-shape Systematics (2)

- (3) is estimated by comparing with a reasonable alternative shape
  - The variation is "mirrored" to obtain a pair of "bounding shapes"

- "Bounding shapes" to evaluate shape uncertainty for non-resonant background
- Here, the alternative shape is constructed from simulated events



Process	Central Shape	Shape variation
Higgs signal	POWHEG with NLO k-factors	Vary μ_R and μ_F
ZZ	MadGraph	Pythia
WZ	MadGraph	Pythia
Non-resonant	$e\mu$ data	Simulated events
Z+$jets$	γ+jets data	N/A (covered by bin-by-bin statistical)

M_T-shape Systematics (3)

- *Expected limits* on σ/σ_{SM} computed with and without shape uncertainty
  - Performance degrades only up to a few %
- Statistical uncertainty on the shape is the leading effect

	without shape uncertainty			with shape uncertainty		
Mass	Median Expected	68% C.L. band	95% C.L. band	Median Expected	68% C.L. band	95% C.L. band
250	1.38	[1.00, 1.92]	[0.75, 2.55]	1.38	[1.00, 1.92]	[0.75, 2.56]
300	0.92	[0.66, 1.28]	[0.50, 1.69]	0.93	[0.67, 1.29]	[0.51, 1.72]
350	0.62	[0.45, 0.86]	[0.34, 1.15]	0.63	[0.45, 0.87]	[0.34, 1.16]
400	0.63	[0.45, 0.87]	[0.34, 1.16]	0.63	[0.46, 0.88]	[0.34, 1.17]
500	1.06	[0.76, 1.47]	[0.57, 1.95]	1.08	[0.78, 1.50]	[0.58, 1.99]
600	2.20	[1.59, 3.06]	[1.19, 4.06]	2.23	[1.61, 3.10]	[1.21, 4.12]

Results

Event Yields for $4.6\ fb^{-1}$

- Example yields after selection for $m_H = 250, 300, 350$
  - Dominant bkgs are typically ZZ and non-resonant processes
  - Z+jets expectation is taken as midpoint of estimate from γ+jets sample with 100% uncertainty
  - Observed data consistent with SM bkg-only expectations

$m_H = 250$

Channel	ZZ	WZ	Top/WW/W+Jets	Z+Jets	Total	mH(250)	Data
$\mu\mu$	$30 \pm 0.19 \pm 2.9$	$21 \pm 0.3 \pm 2.5$	$63 \pm 3.7 \pm 6$	20 ± 20	133 ± 21	14 ± 1.8	139
ee	$20 \pm 0.15 \pm 2.1$	$14 \pm 0.24 \pm 1.8$	$45 \pm 5.1 \pm 4.4$	13 ± 13	93 ± 15	9.7 ± 1.4	99

$m_H = 300$

Channel	ZZ	WZ	Top/WW/W+Jets	Z+Jets	Total	mH(300)	Data
$\mu\mu$	$21 \pm 0.16 \pm 2$	$11 \pm 0.22 \pm 1.3$	$20 \pm 1.2 \pm 3.4$	5.2 ± 5.2	57 ± 6.8	14 ± 1.9	60
ee	$14 \pm 0.13 \pm 1.5$	$7.8 \pm 0.18 \pm 0.98$	$14 \pm 1.6 \pm 2.4$	3.6 ± 3.6	40 ± 5	10 ± 1.4	35

$m_H = 350$

Channel	ZZ	WZ	Top/WW/W+Jets	Z+Jets	Total	mH(350)	Data
$\mu\mu$	$24 \pm 0.17 \pm 2.3$	$12 \pm 0.23 \pm 1.5$	$20 \pm 1.2 \pm 3.4$	6.1 ± 6.1	62 ± 7.6	16 ± 2.5	62
ee	$17 \pm 0.14 \pm 1.8$	$8.7 \pm 0.19 \pm 1.1$	$14 \pm 1.6 \pm 2.4$	4.2 ± 4.2	44 ± 5.5	12 ± 1.9	38

Systematic Uncertainties

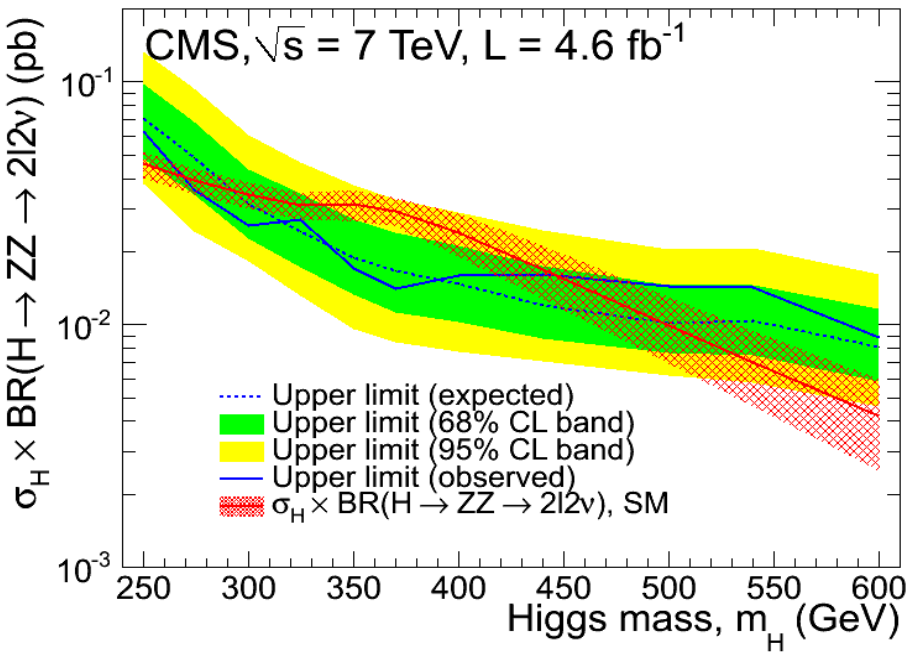
- Uncertainties on signal/background normalizations
- Largest theoretical uncertainty from zero-width line shape in Higgs MC
- Largest experimental uncertainties in estimates of non-resonant bkg (small control region) and Z+jets (γ+jets impurity)

Source	Uncertainty [%]
Luminosity	4.5
PDF for Higgs production (gluon-fusion)	8.2–10.4
PDF for Higgs production (VBF)	3.8–7.6
PDF for $qq \to ZZ$	4.8
PDF for $qq \to WZ$	5.5
QCD scale for Higgs production (gluon-fusion)	7.6–11.1
QCD scale for Higgs production (VBF)	0.4–2
QCD scale for $gg \to ZZ$	20
QCD scale for $qq \to ZZ$	6.2
QCD scale for $qq \to WZ$	8.5
Higgs boson line shape	10–30
Trigger	1 (for ee), 2 (for $\mu\mu$)
Lepton identification and isolation	2
Lepton momentum scale	5 (for ee), 2 (for $\mu\mu$)
Jet energy scale	1–1.5
bjet Veto	1–1.2
Pile-up	1–3
Non-resonant background	15
Z+jets	100

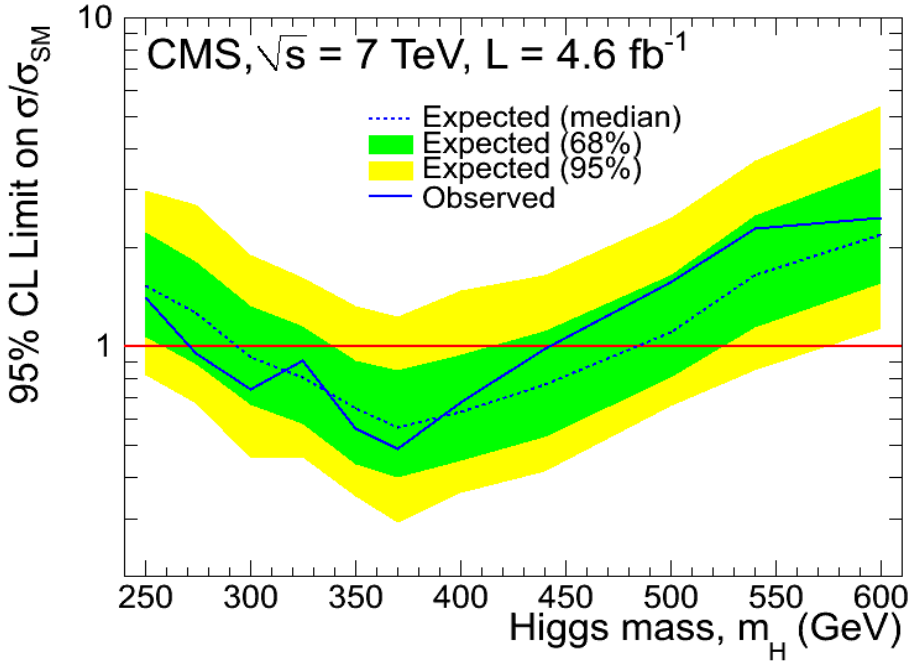
Statistical Analysis

- Apply statistic techniques to quantify search results
  - Characterize exclusion of signal or significance of an excess
- To facilitate comparison and combination, CMS and ATLAS have agreed on a common method: “CLs-LHC”
  - Likelihood ratio method constructed to have several desired statistical properties
- In limit setting procedure, get likelihood profile with “toy experiments” based on expected signal and bkg
  - Systematic uncertainties correspond to nuisance parameters in the maximum likelihood fits of M_T distribution

Limits with *4.6 fb*$^{-1}$



Limit on σ_{SM} x BR(H→ZZ→2l2v)

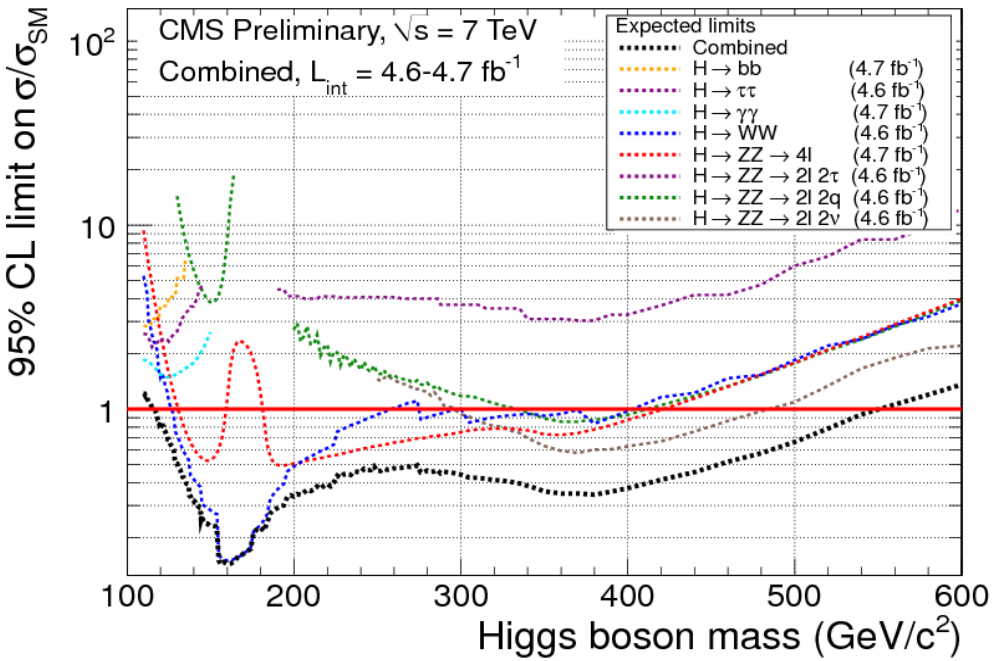


Limit on σ / σ_{SM}

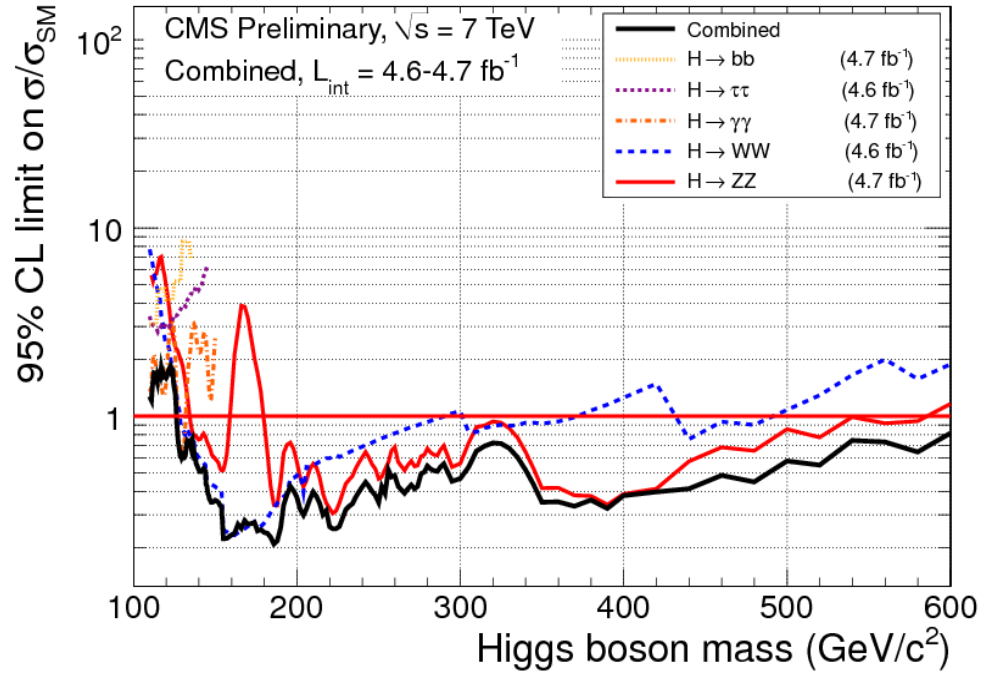
- No significant excess observed
- Expected exclusion region is *[290, 490] GeV* at *95%* C.L.
- Observed exclusion region is *[270, 440] GeV* at *95%* C.L.

CMS Combined Higgs Limits

Expected Limits



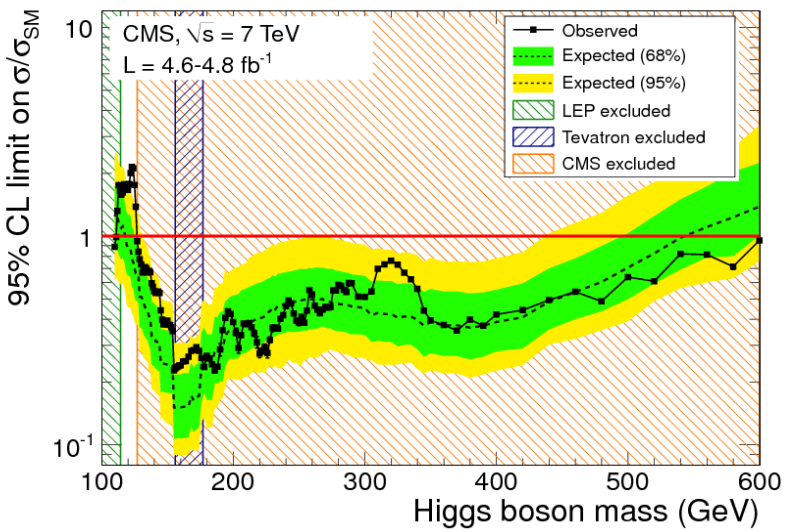
Observed Limits

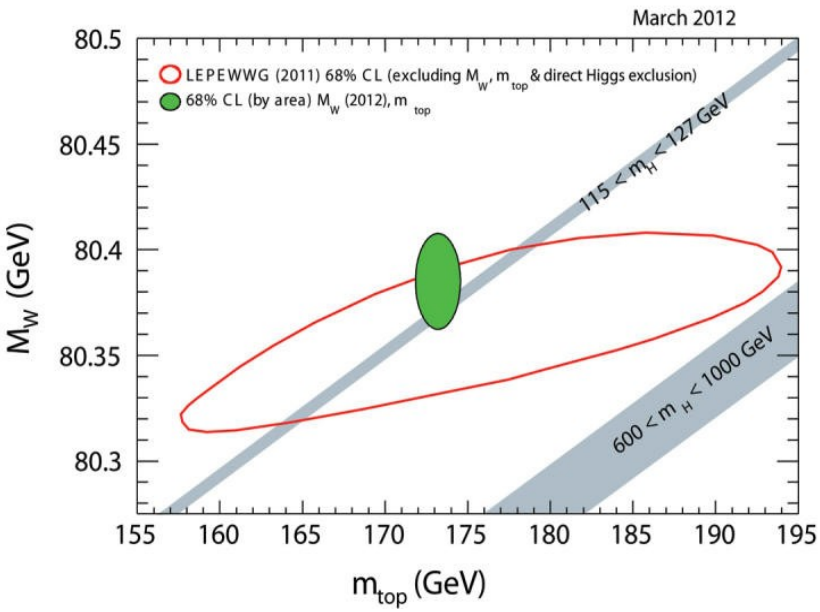


- $H \rightarrow ZZ \rightarrow ll\upsilon\upsilon$ channel has best expected sensitivity for m_H above 320 GeV
- Combining all CMS Higgs channels:
  - Expected exclusion region is *[117, 543] GeV* at *95%* C.L.
  - Observed exclusion region is *[127, 600] GeV* at *95%* C.L.

Summary

- CMS has performed a search for SM Higgs boson on the full 2011 dataset
- No excess observed in the $H \rightarrow ZZ \rightarrow ll\upsilon\upsilon$ channel; exclude SM Higgs boson in mass region $[270, 440]$ GeV at 95% C.L.
- Combining all channels, CMS excludes a SM Higgs boson in the mass region $[127, 600]$ GeV at 95% C.L.

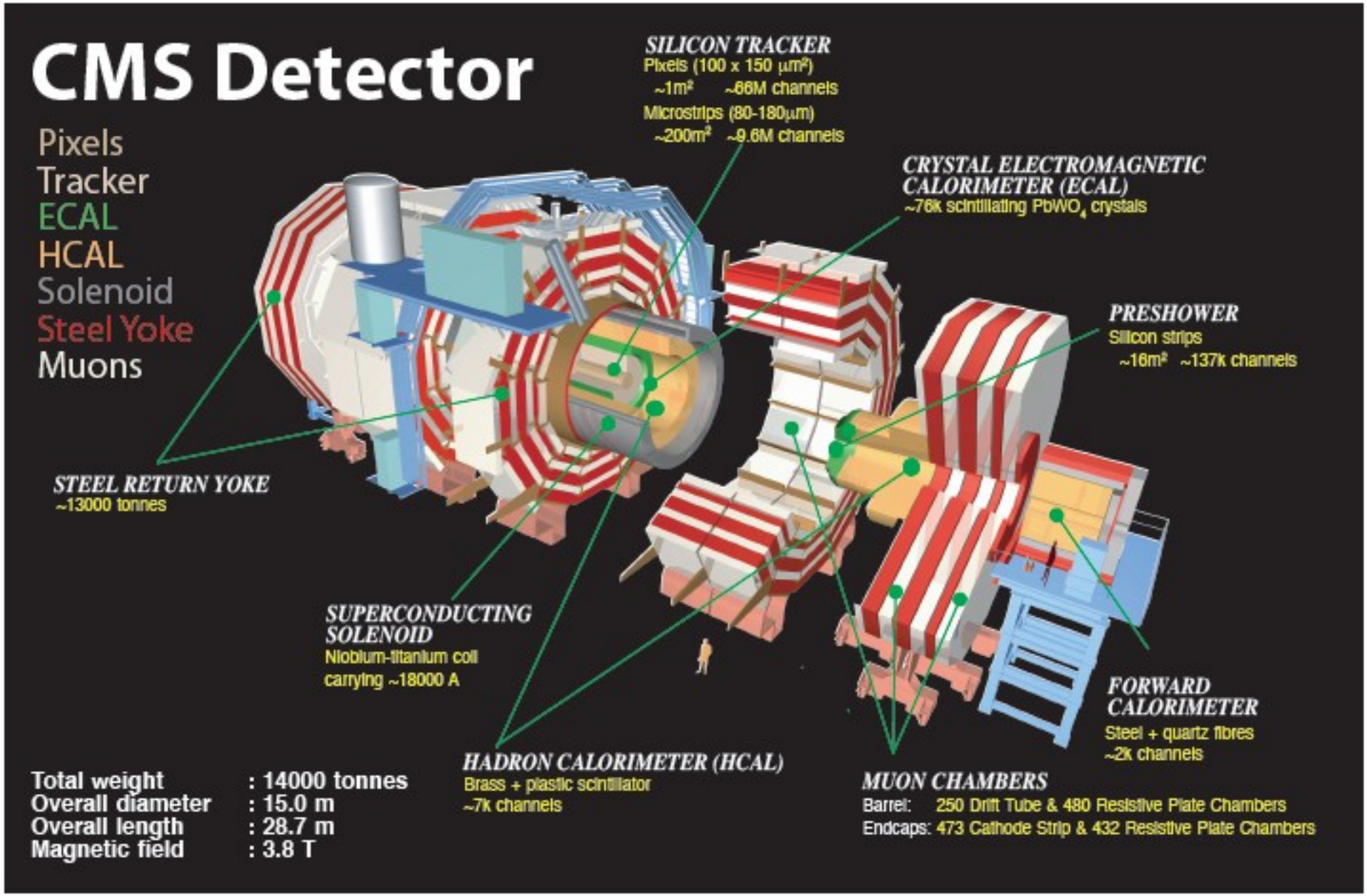




2012 data taking is underway at the LHC, new results soon!

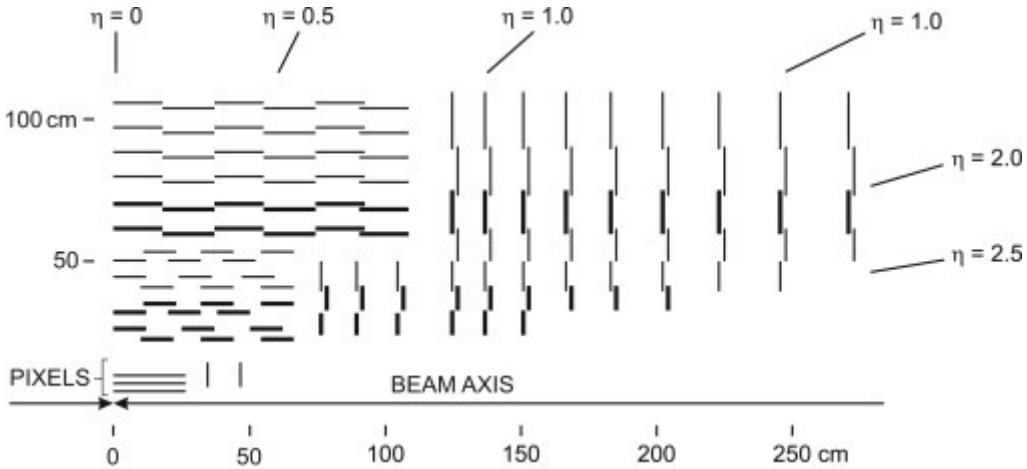
Backup Slides

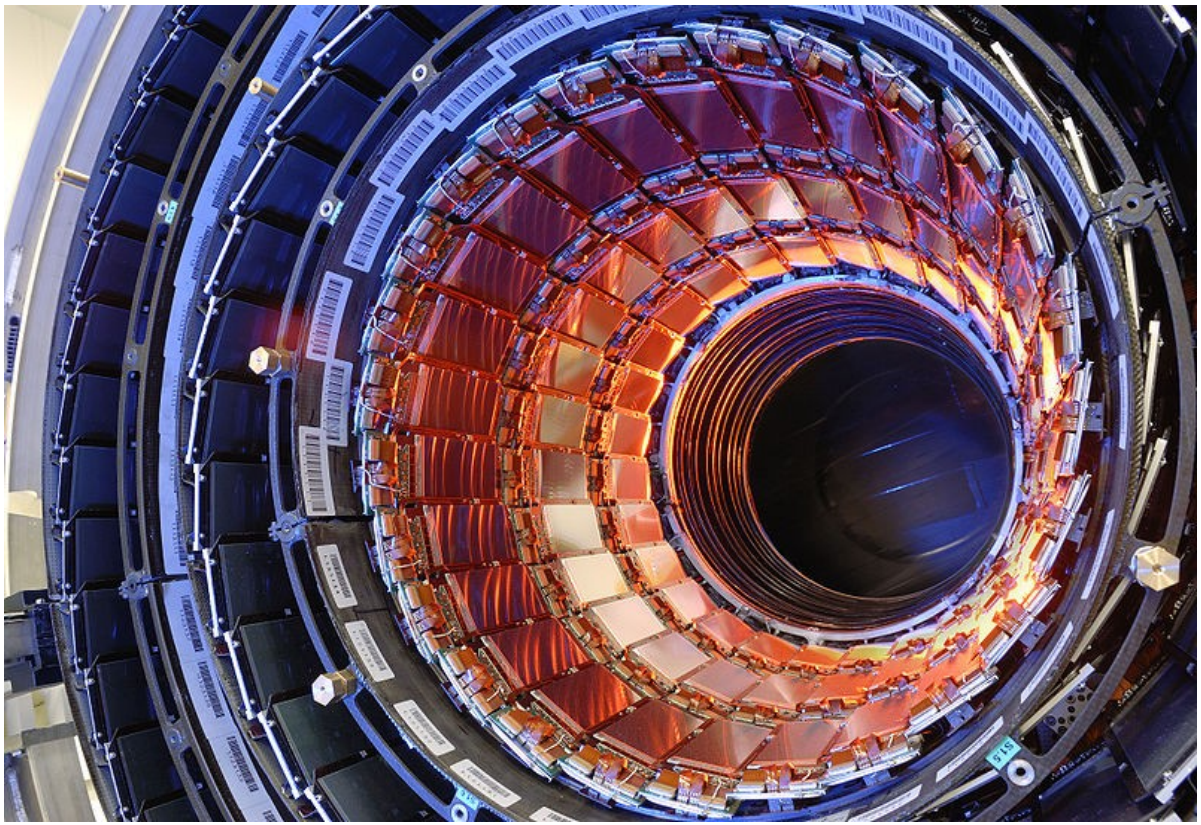
The CMS Detector



Tracker

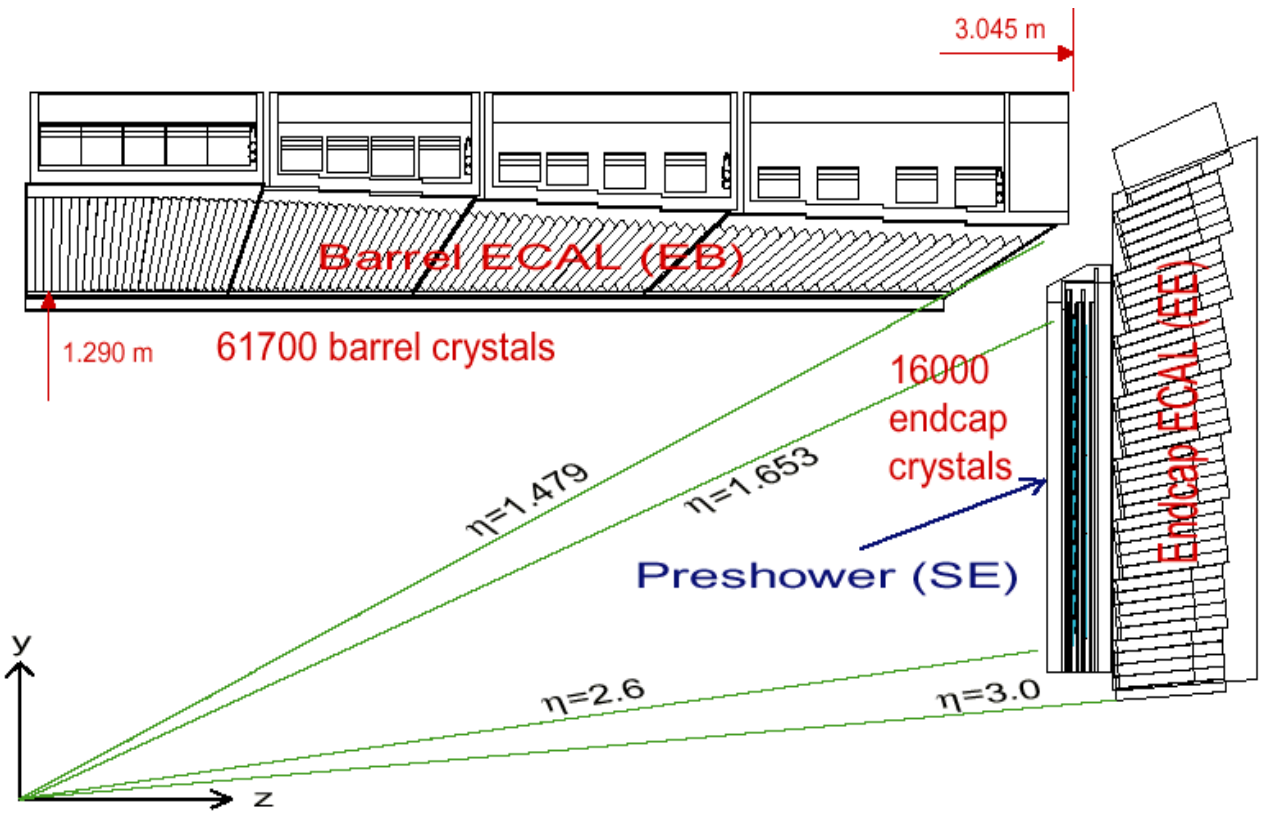
- ~66M pixels
  - 3 barrel layers
  - 2x2 endcap disks
- ~10M strips
  - 10 barrel layers
  - 2x9 endcap disks
- $\sigma(d_0) \sim 15\mu m$
- $\sigma/p_T \sim 1\%$ @ 40 GeV





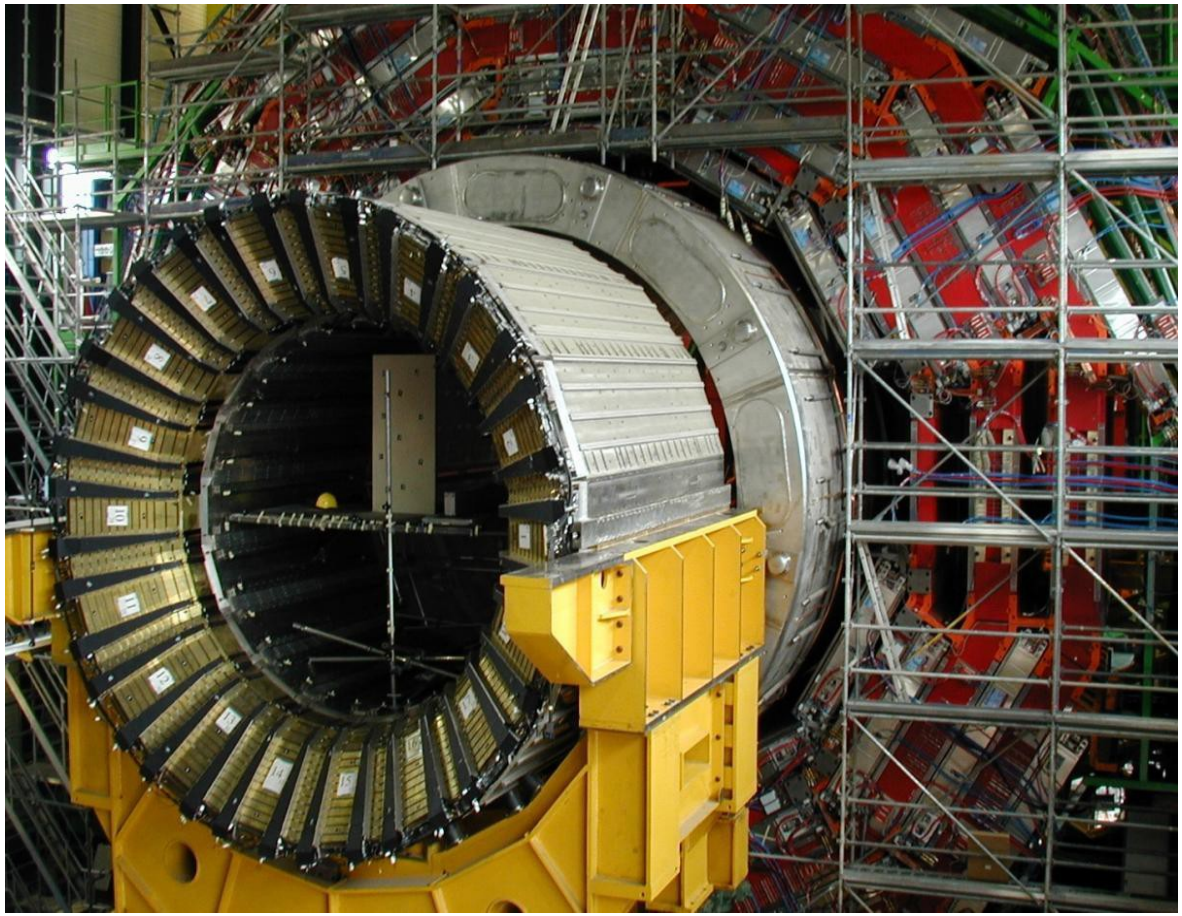
Electromagnetic Calorimeter

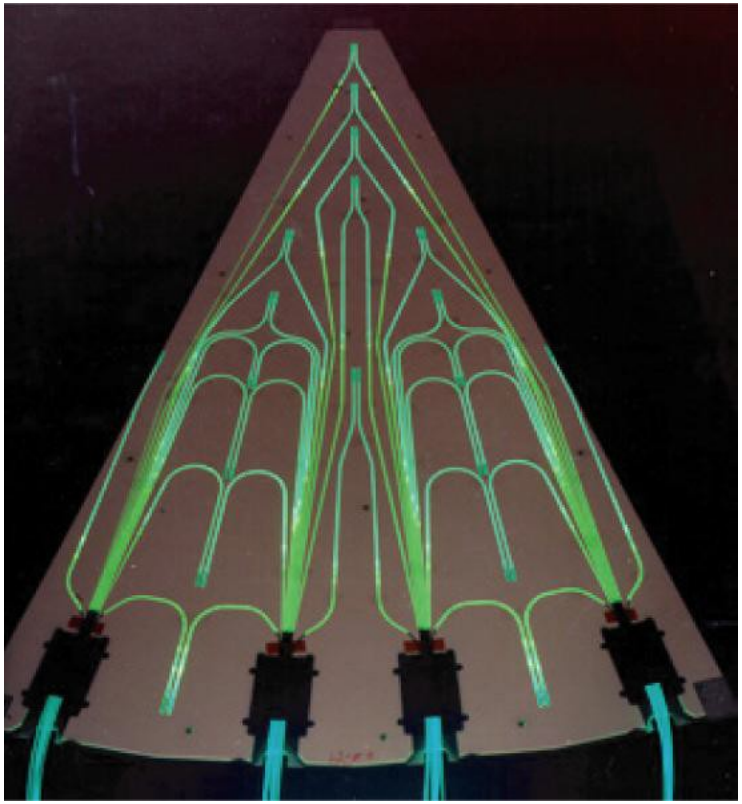
- ~76k $PbWO_4$ crystals
- $\sigma(E)/E \sim 1\%$ @ 50 GeV



Hadronic Calorimeter

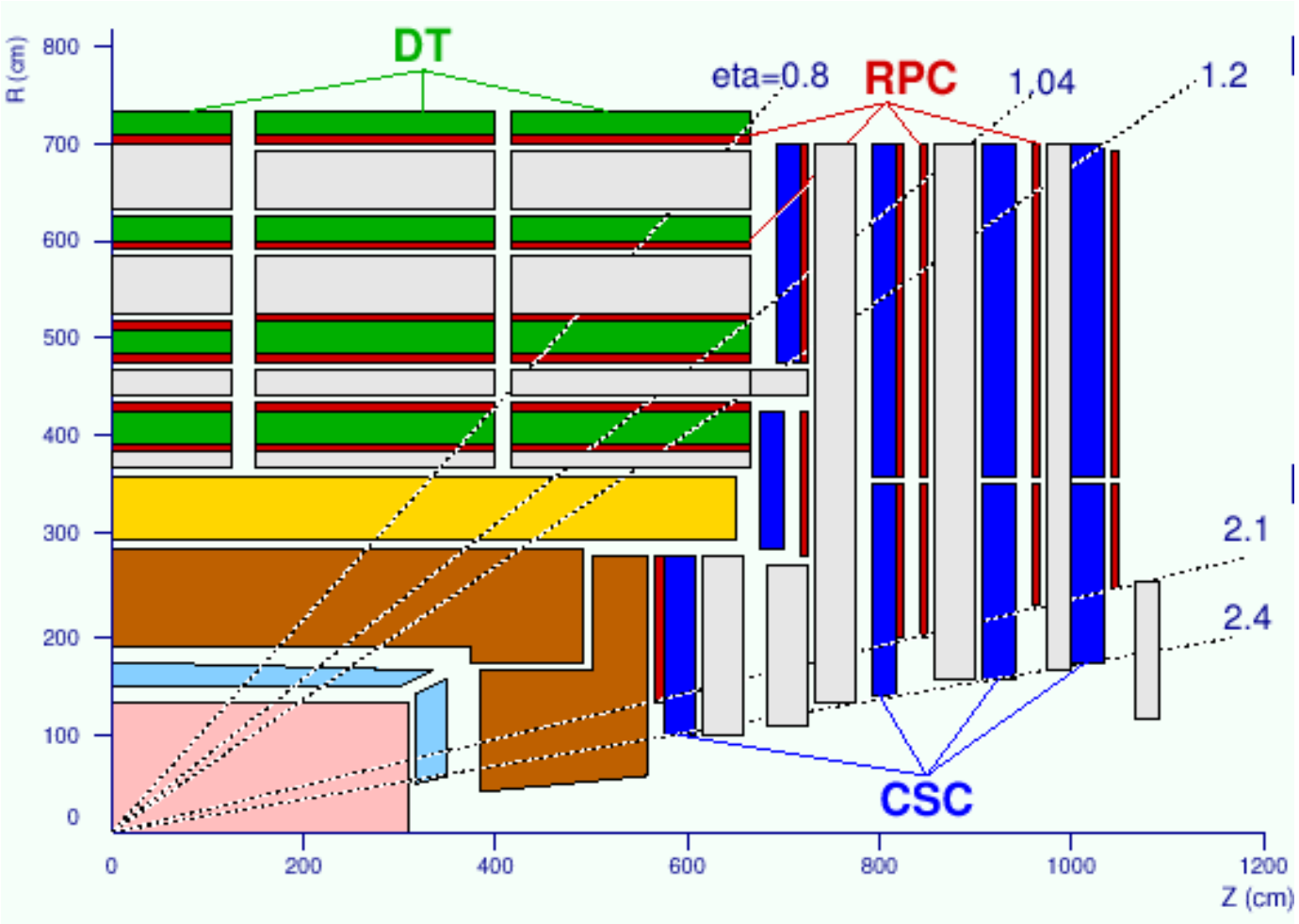
- Sampling calorimeter with brass + plastic scintillator layers
- WLS fibers feed to photodiodes
- $\sigma/E \sim 100\% / \sqrt{E}$





Muon Chambers

- Drift tubes, cathode strip chambers, resistive plate chambers
- $25\ 000\ m^2$ of detection planes
- $O(100\mu m)$ resolution

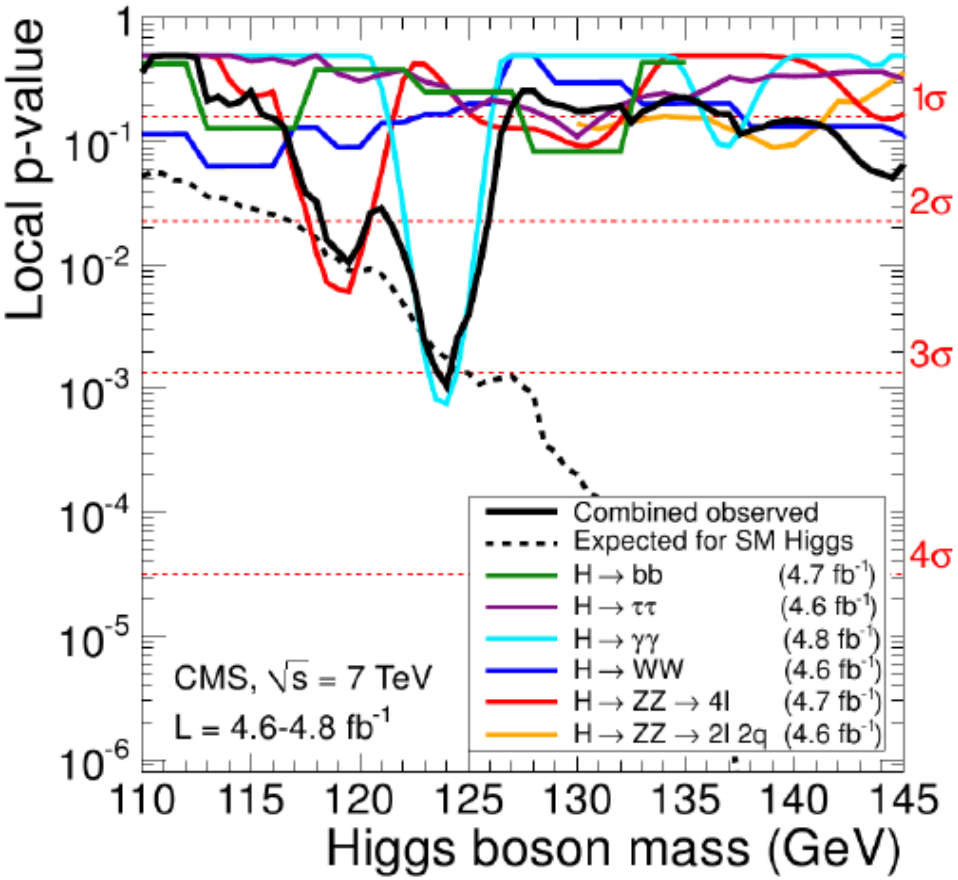


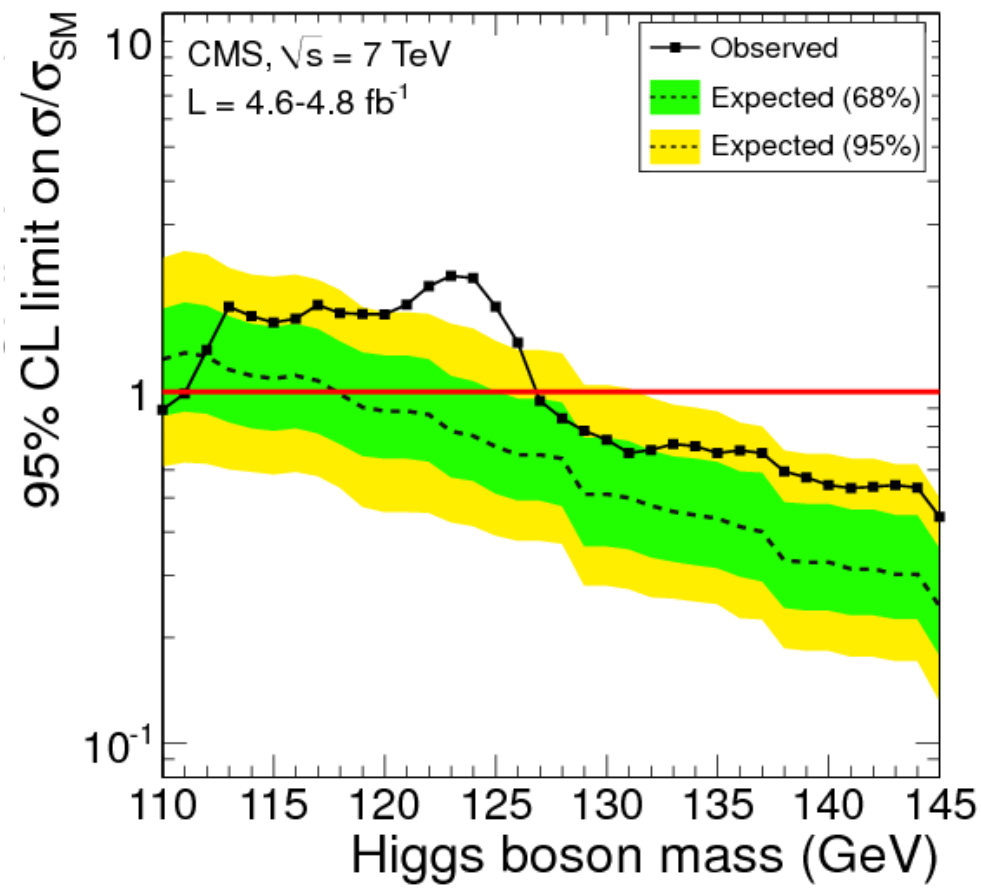
Event Yields for $4.6\ fb^{-1}$

- Observed and expected yields for each m_H hypothesis separated in ee and $\mu\mu$ channels

Channel	ZZ	WZ	Top/WW/W+Jets	Z+Jets	Total	mH(250)	Data
$\mu\mu$	$30 \pm 0.19 \pm 2.9$	$21 \pm 0.3 \pm 2.5$	$63 \pm 3.7 \pm 6$	20 ± 20	133 ± 21	14 ± 1.8	139
ee	$20 \pm 0.15 \pm 2.1$	$14 \pm 0.24 \pm 1.8$	$45 \pm 5.1 \pm 4.4$	13 ± 13	93 ± 15	9.7 ± 1.4	99
Channel	ZZ	WZ	Top/WW/W+Jets	Z+Jets	Total	mH(300)	Data
$\mu\mu$	$21 \pm 0.16 \pm 2$	$11 \pm 0.22 \pm 1.3$	$20 \pm 1.2 \pm 3.4$	5.2 ± 5.2	57 ± 6.8	14 ± 1.9	60
ee	$14 \pm 0.13 \pm 1.5$	$7.8 \pm 0.18 \pm 0.98$	$14 \pm 1.6 \pm 2.4$	3.6 ± 3.6	40 ± 5	10 ± 1.4	35
Channel	ZZ	WZ	Top/WW/W+Jets	Z+Jets	Total	mH(350)	Data
$\mu\mu$	$24 \pm 0.17 \pm 2.3$	$12 \pm 0.23 \pm 1.5$	$20 \pm 1.2 \pm 3.4$	6.1 ± 6.1	62 ± 7.6	16 ± 2.5	62
ee	$17 \pm 0.14 \pm 1.8$	$8.7 \pm 0.19 \pm 1.1$	$14 \pm 1.6 \pm 2.4$	4.2 ± 4.2	44 ± 5.5	12 ± 1.9	38
Channel	ZZ	WZ	Top/WW/W+Jets	Z+Jets	Total	mH(400)	Data
$\mu\mu$	$26 \pm 0.18 \pm 2.5$	$13 \pm 0.23 \pm 1.5$	$20 \pm 1.2 \pm 3.4$	6.4 ± 6.4	64 ± 7.9	14 ± 1.8	65
ee	$18 \pm 0.15 \pm 1.9$	$9.1 \pm 0.19 \pm 1.1$	$14 \pm 1.6 \pm 2.4$	4.5 ± 4.5	45 ± 5.8	9.8 ± 1.4	42
Channel	ZZ	WZ	Top/WW/W+Jets	Z+Jets	Total	mH(500)	Data
$\mu\mu$	$28 \pm 0.19 \pm 2.7$	$13 \pm 0.24 \pm 1.6$	$20 \pm 1.2 \pm 3.4$	6.8 ± 6.8	67 ± 8.3	6.1 ± 0.87	70
ee	$19 \pm 0.15 \pm 2$	$9.4 \pm 0.2 \pm 1.2$	$14 \pm 1.6 \pm 2.4$	4.7 ± 4.7	47 ± 6	4.5 ± 0.67	42
Channel	ZZ	WZ	Top/WW/W+Jets	Z+Jets	Total	mH(600)	Data
$\mu\mu$	$28 \pm 0.19 \pm 2.7$	$14 \pm 0.24 \pm 1.6$	$20 \pm 1.2 \pm 3.4$	6.8 ± 6.8	68 ± 8.3	2.7 ± 0.4	70
ee	$19 \pm 0.15 \pm 2$	$9.5 \pm 0.2 \pm 1.2$	$14 \pm 1.6 \pm 2.4$	4.8 ± 4.8	48 ± 6.1	1.9 ± 0.3	42

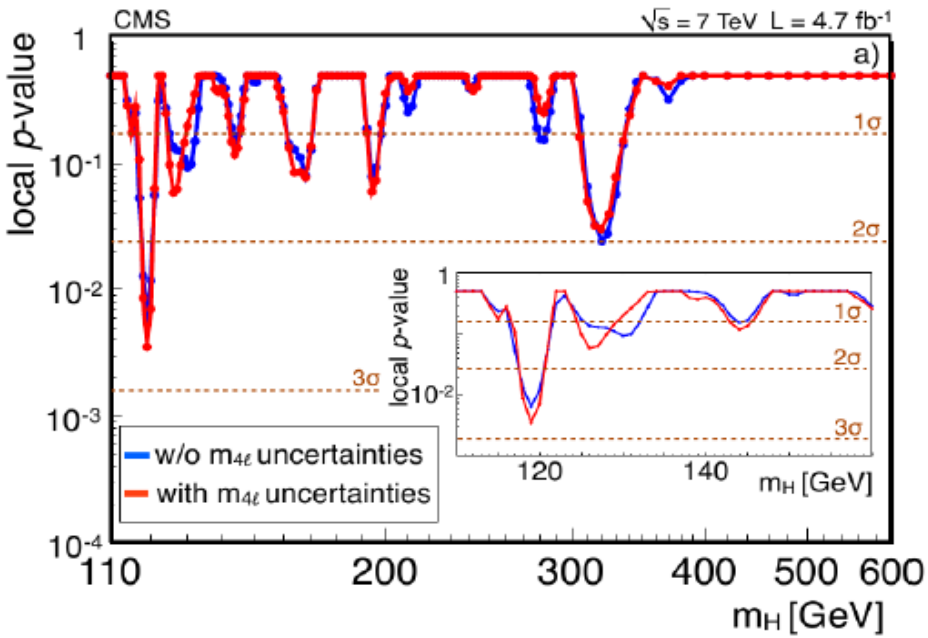
CMS Higgs Combination

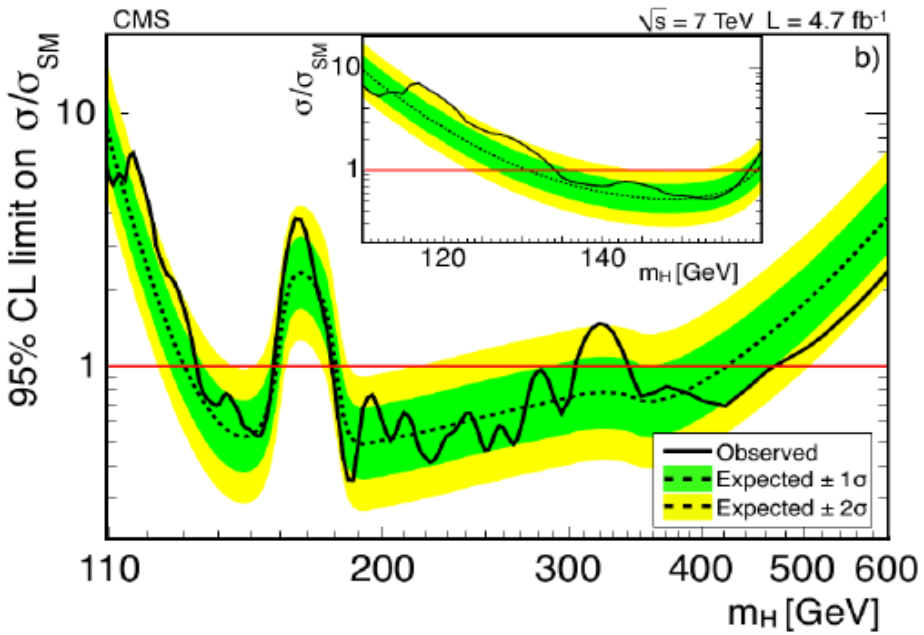




- 95% C.L. exclusion 127 < m_H < 600
- Largest excess at 124 GeV dominated by $H \rightarrow \gamma\gamma$; smaller excess at 119.5 GeV by $H \rightarrow ZZ \rightarrow 4l$ with smaller contribution from $H \rightarrow WW \rightarrow l\nu l\nu$

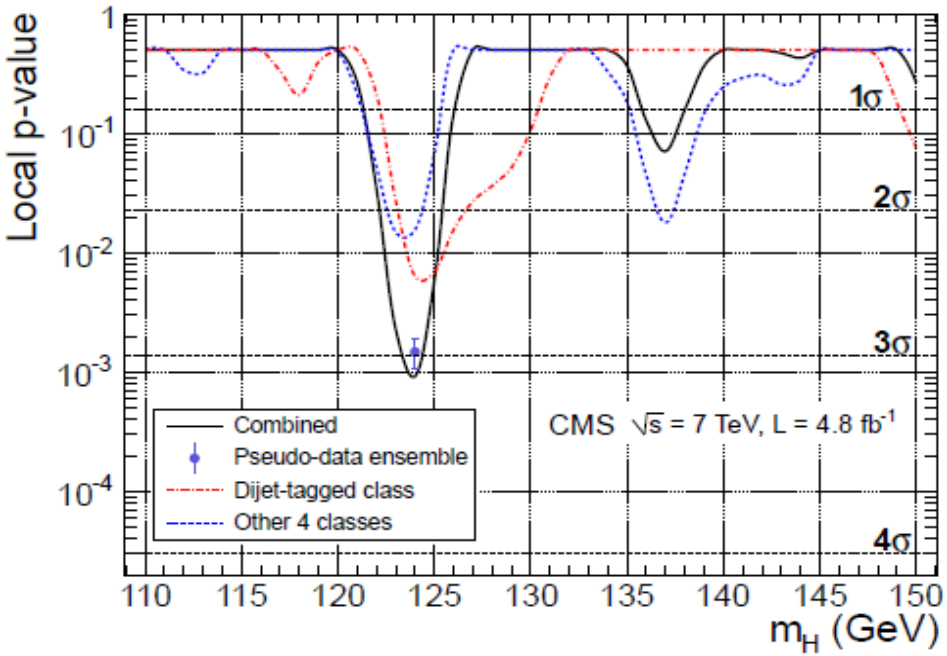
$H \rightarrow ZZ \rightarrow 4l$

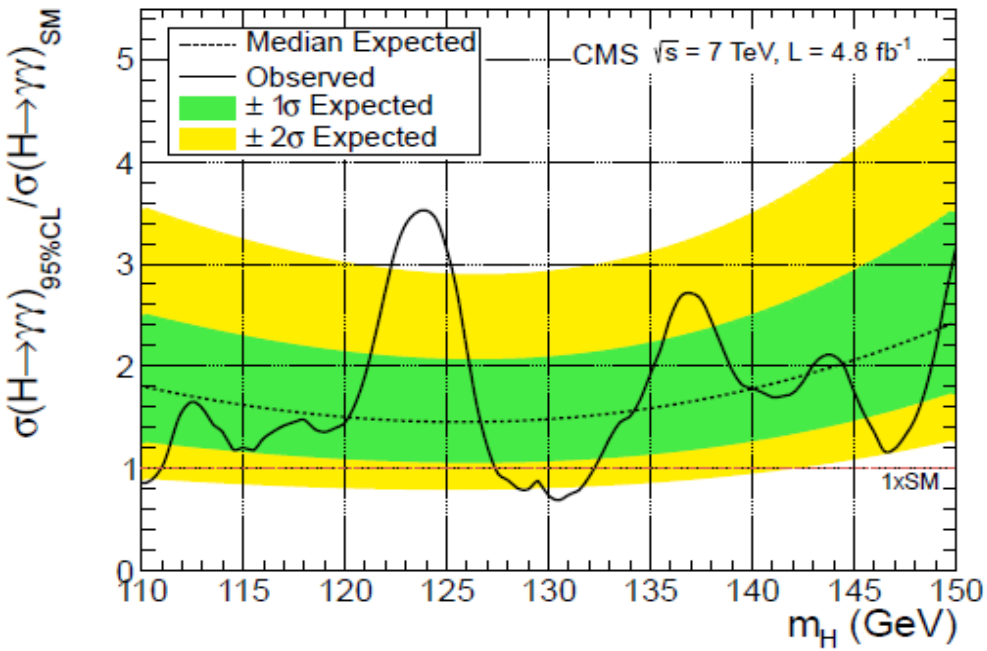




- Limits derived from unbinned likelihood fit to m_{4l} distribution
- Largest observed excesses
  - 119.5 GeV: Global – 1.6σ (Local – 2.5σ)
  - 320.0 GeV: Global – 1.0σ (Local – 2.0σ)
- 95% C.L. exclusion for [134,158], [180,305], [340,465]

$H \rightarrow \gamma\gamma$





- Excess at 124 GeV with contribution from VBF-tagged and non-VBF-tagged events
  - Global – 1.8σ (Local – 3.1σ)
- 95% C.L. exclusion for $128 < m_H < 132$