

# Search for the Standard Model Higgs Boson in the Decay Channel $H \rightarrow ZZ \rightarrow 2l2q$ at CMS

Matthew Kress

Adviser: Daniela Bortoletto  
Ph.D. Defense  
April 11, 2014

# Outline

- 1 LHC Search for Higgs Boson
- 2 Event Selection
- 3 Cross Check and Statistical Analysis
- 4 Results
- 5 Conclusion

# Outline

1 LHC Search for Higgs Boson

2 Event Selection

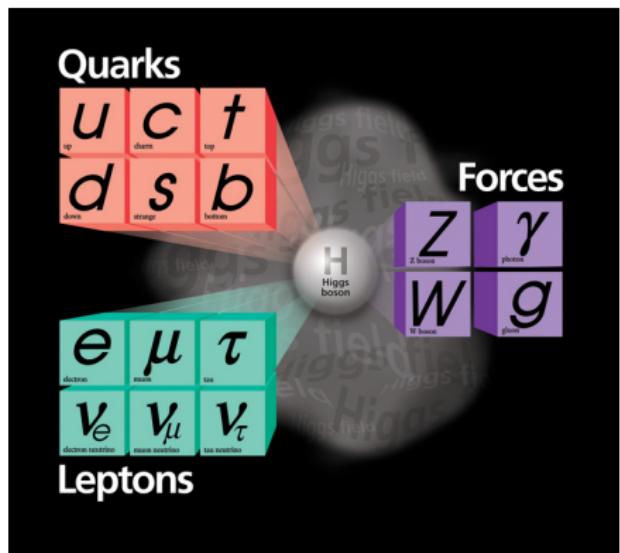
3 Cross Check and Statistical Analysis

4 Results

5 Conclusion

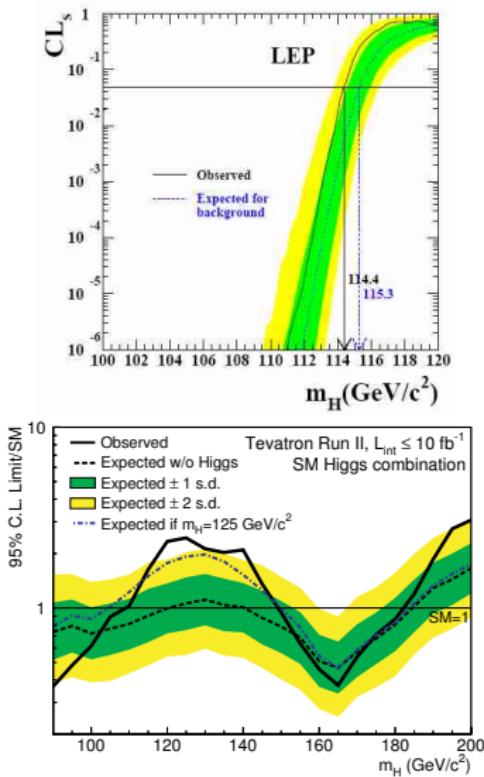
# The Standard Model

The Standard Model is the compilation of over 100 years of scientific discoveries and is in excellent agreement with a wide range of experimental observations.



- In the Standard Model the simplest solution for the nature of the electroweak symmetry breaking is the introduction of the Higgs field.
- At the commencement of the LHC the only free parameter of the Standard Model was the Higgs mass.

# Experimental Constraints on the SM Higgs Mass



Experimental constraints:

- LEP excluded with  $CL_{95\%}$   
 $m_H < 114.4 \text{ GeV}.$
- The latest measurements from Tevatron (July 2013) exclude with  $CL_{95\%}$  at:  
 $90 \text{ GeV} < m_H < 109 \text{ GeV}$   
 $149 \text{ GeV} < m_H < 182 \text{ GeV}.$

# Motivation

## Discovery

In 2012 ATLAS and CMS announced the discovery of a new boson at 126 GeV. In 2013 it was confirmed that this new particle is consistent with a Standard Model Higgs boson.

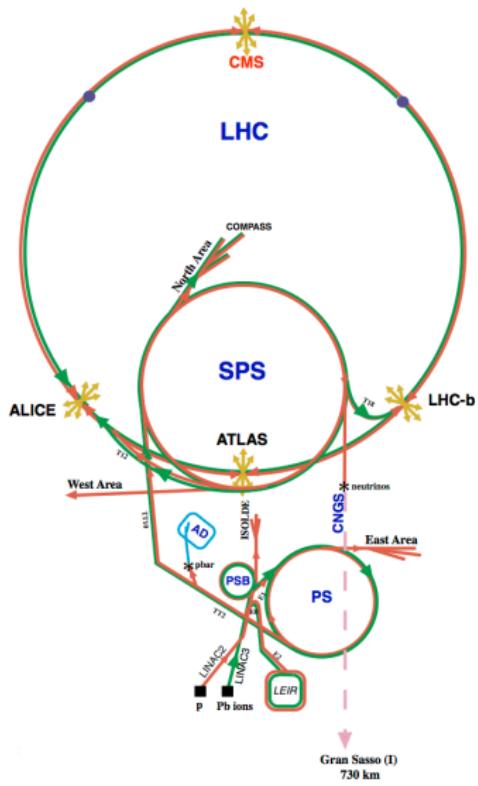
This analysis is sensitive in the 250 to 650 GeV range so why search there?

- When we started this analysis in 2011 we didn't know where the Higgs boson would be. (Needed to look everywhere)
- Is the new particle THE Standard Model Higgs boson?
- Are there other Higgs bosons?

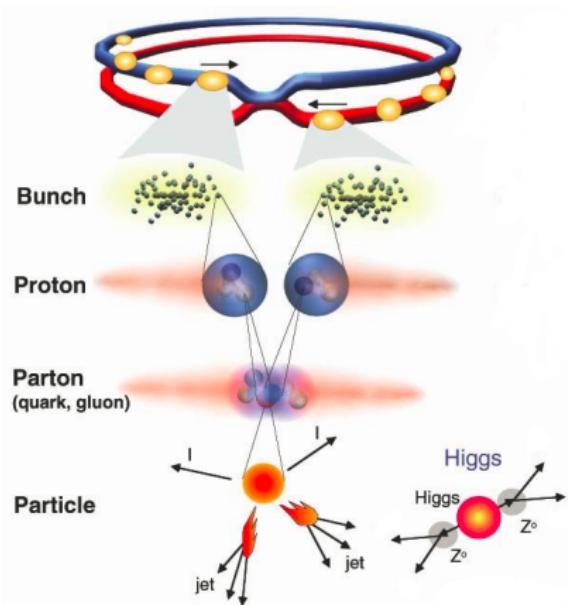
## Beyond the Standard Model

Many Beyond the Standard Model (BSM) theories extend the simple Higgs sector of the Standard Model and lead to more complicated particle spectrum. Very often one of the new particles has properties similar to the Standard Model Higgs.

# LHC Environment



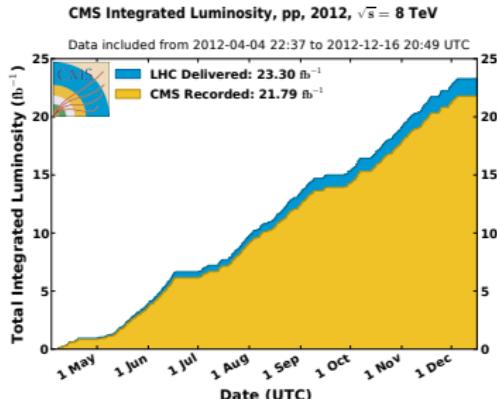
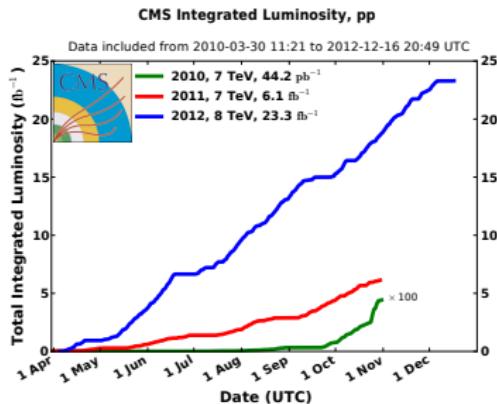
In 2012  $\sqrt{s} = 8 \text{ TeV}$



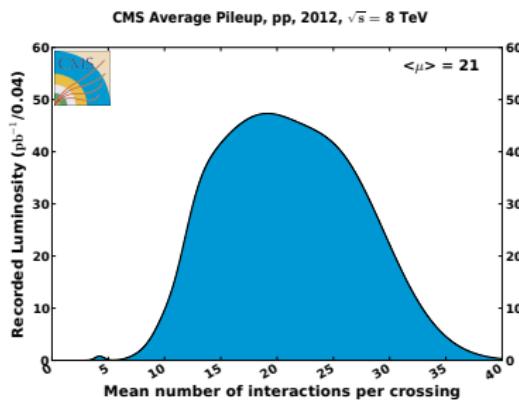
Source: Haijun Yang, Colloquium Shanghai Jiaotong University, Sept 12, 2012

Source: <http://www.quantumdiaries.org/2011/10/24/the-25-na-pumpkin-teeth/>

# LHC Delivered Data



- CMS has done extremely well during Run I (2010-2012).
- Data-taking efficiency was very high (95%)
- Main challenge of the 2012 was how to deal with pileup.

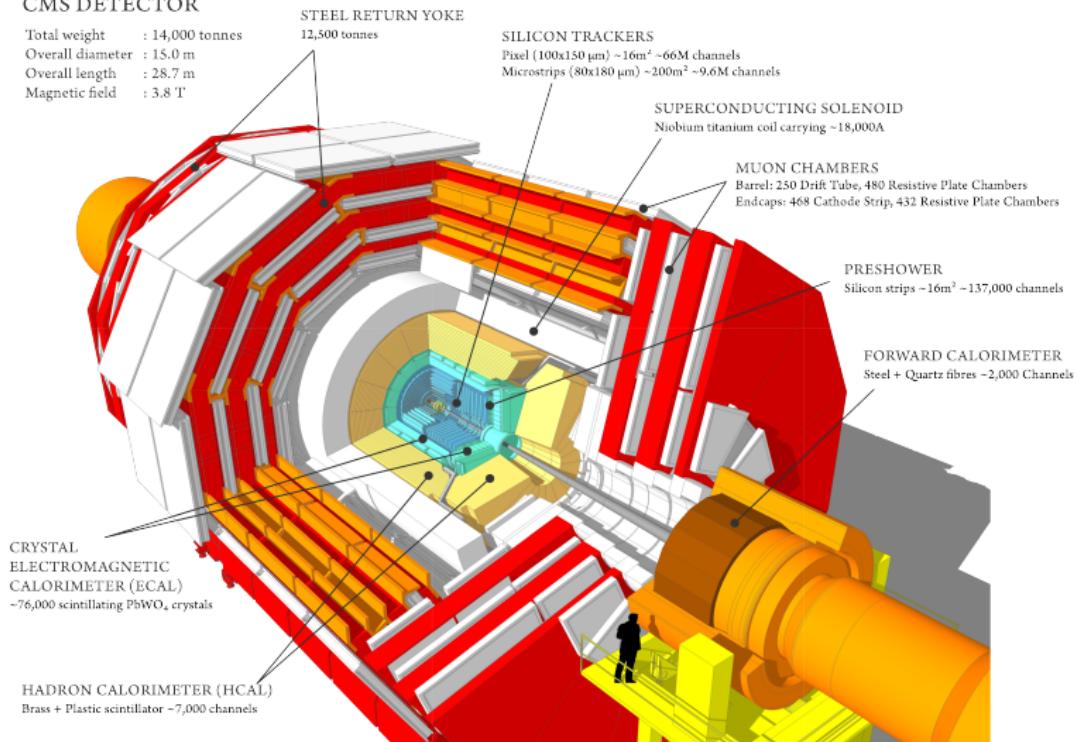


Source:

# Compact Muon Solenoid

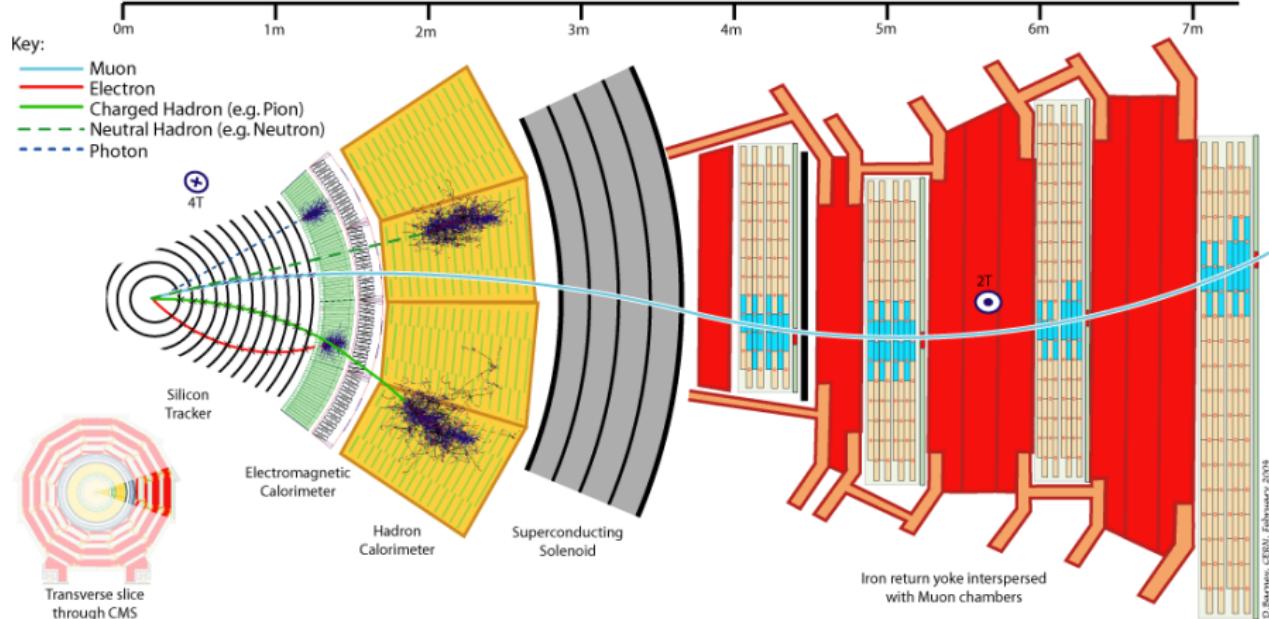
## CMS DETECTOR

Total weight : 14,000 tonnes  
Overall diameter : 15.0 m  
Overall length : 28.7 m  
Magnetic field : 3.8 T



Source: [http://www.fnal.gov/pub/presspass/press\\_releases/2013/Higgs-Boson-20130314.html](http://www.fnal.gov/pub/presspass/press_releases/2013/Higgs-Boson-20130314.html)

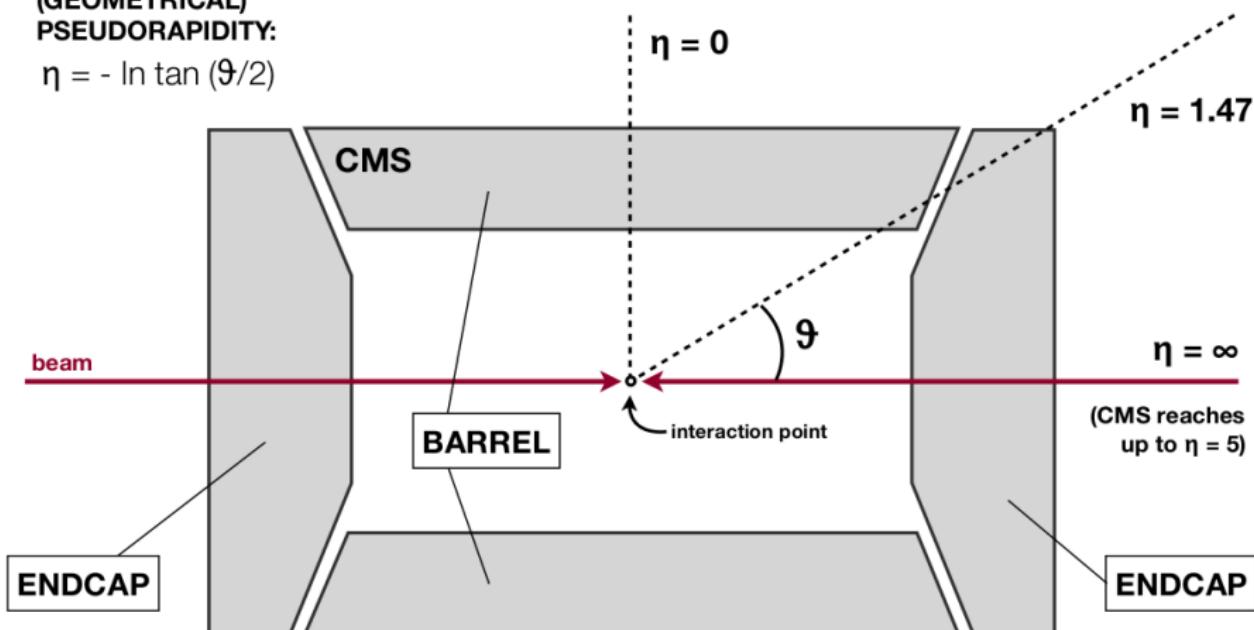
# CMS Slice



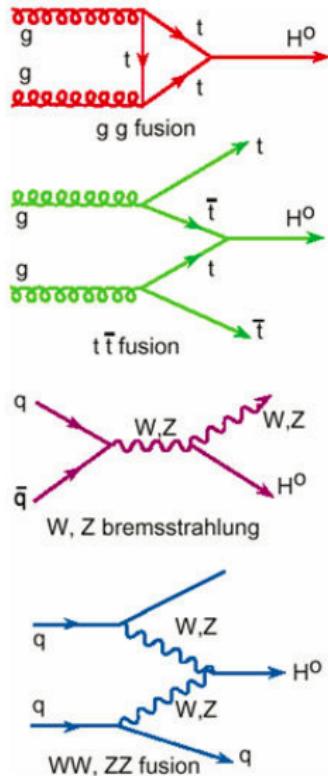
# CMS Detector Glossary

(GEOMETRICAL)  
PSEUDORAPIDITY:

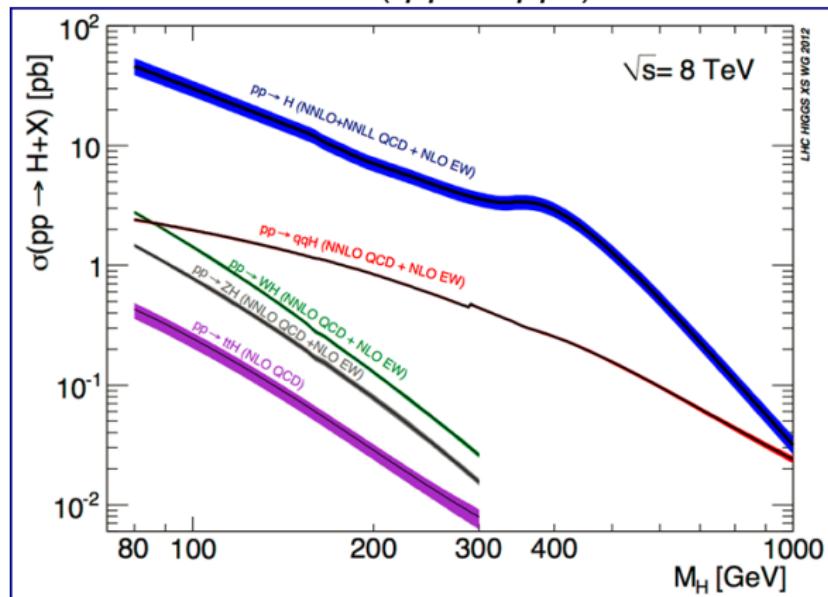
$$\eta = -\ln \tan(\theta/2)$$



# Higgs Production

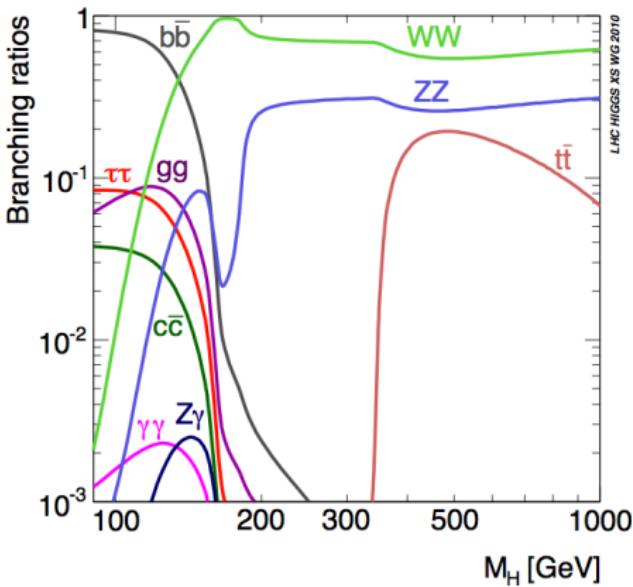


Gluon-gluon fusion ( $gg \rightarrow H$ ) and vector-boson fusion ( $qq \rightarrow qqH$ ) are dominant



# Higgs Decay

- Discovery strategy depends on the available decay channels.
- Decays with leptons provide clean signatures.



Main Discovery Channels

- $H \rightarrow \gamma\gamma$
- $H \rightarrow W^+W^-$
- $H \rightarrow ZZ$

# The $H \rightarrow ZZ \rightarrow llqq$ Channel

There are three sub-channels in  $H \rightarrow ZZ$  that are studied.

$$H \rightarrow ZZ \rightarrow 4l$$

$$H \rightarrow ZZ \rightarrow 2l2\nu$$

$$H \rightarrow ZZ \rightarrow \ell^+\ell^-q\bar{q}$$

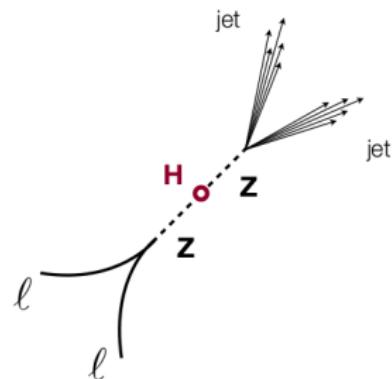
( $\ell = e, \mu$  and  $q = u, d, c, s, b$ )

Pros:

- Large branching ratios
  - $BR(Z \rightarrow qq) = 70\%$
  - $BR(ZZ \rightarrow 2l2q) = 20 \times BR(ZZ \rightarrow 4l)$
- Full decay is reconstructed (closed kinematics)

Cons:

- Quarks hadronize to form “jets.”
- Momentum resolution is worse for jets than for leptons.
- Large background coming from Z production in conjunction with QCD jets ( $\sim 10^5 >$  signal).



# Outline

1 LHC Search for Higgs Boson

2 Event Selection

3 Cross Check and Statistical Analysis

4 Results

5 Conclusion

# Signal MC Samples

$M_H$ GeV	$\sigma \times \text{Br}(H \rightarrow ZZ \rightarrow 2l/2q)$ [pb]
230	0.2278
250	0.2022
275	0.1751
300	0.1563
325	0.1478
350	0.1482
375	0.1360
400	0.1111
425	0.0914
450	0.7311
475	0.6000
500	0.4719
525	0.0380
550	0.0305
575	0.0250
600	0.0201

The signal samples,  
 $H \rightarrow ZZ \rightarrow 2l/2q$  ( $\ell = e, \mu, \tau$ ), were simulated with POWHEG then using PYTHIA to do the final parton showering and hadronization.

The cross section times branching fraction for each  $m_H$  value is listed in pb. Each sample was generated with 300,000 events.

# Background Samples

Background

Process	generator	$\sigma$ [pb]	luminosity [ $\text{fb}^{-1}$ ]
Z+jets (inclusive)	madgraph	3503.71	8.7
Z+1 jet (exclusive)	madgraph	660.6	36.4
Z+2 jet (exclusive)	madgraph	215.1	101.6
Z+3 jet (exclusive)	madgraph	65.79	167.4
Z+4 jet (exclusive)	madgraph	27.59	232.1
t <bar>t</bar>	powheg-pythia6	23.38	461
ZZ	pythia6	17.654	549
WZ	pythia6	22.88	424
WW	pythia6	57.1097	168

# Datasets ( $19.6 \text{ fb}^{-1}$ )

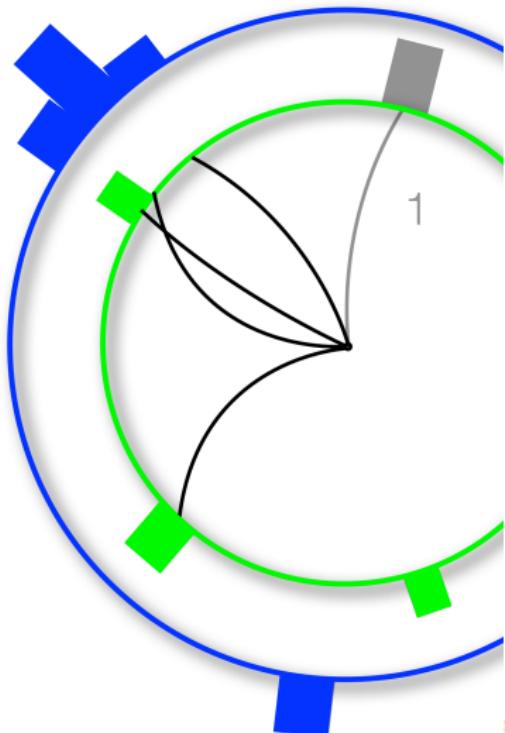
Double Electron and Double Muon datasets from a total of  $19.6 \text{ fb}^{-1}$  collected data in 2012 at  $\sqrt{s} = 8 \text{ TeV}$ .

Channel	Dataset	Luminosity [ $\text{pb}^{-1}$ ]
$2\mu 2q$	/DoubleMu/Run2012A-13Jul2012-v1/AOD	808
	/DoubleMu/Run2012A-recover-06Aug2012-v1/AOD	82
	/DoubleMu/Run2012B-13Jul2012-v4/AOD	4429
	/DoubleMu/Run2012C-24Aug2012-v1/AOD	495
	/DoubleMu/Run2012C-EcalRecover_11Dec2012-v1/AOD	134
	/DoubleMu/Run2012C-PromptReco-v2/AOD	6394
	/DoubleMu/Run2012D-PromptReco-v1/AOD	7274
$2e 2q$	/DoubleElectron/Run2012A-13Jul2012-v1/AOD	808
	/DoubleElectron/Run2012A-recover-06Aug2012-v1/AOD	82
	/DoubleElectron/Run2012B-13Jul2012-v4/AOD	4429
	/DoubleElectron/Run2012C-24Aug2012-v1/AOD	495
	/DoubleElectron/Run2012C-EcalRecover_11Dec2012-v1/AOD	134
	/DoubleElectron/Run2012C-PromptReco-v2/AOD	6394
	/DoubleElectron/Run2012D-PromptReco-v1/AOD	7274

# Particle Flow

The particle flow algorithm attempts to reconstruct all stable final state particles using all of the CMS sub-detectors.

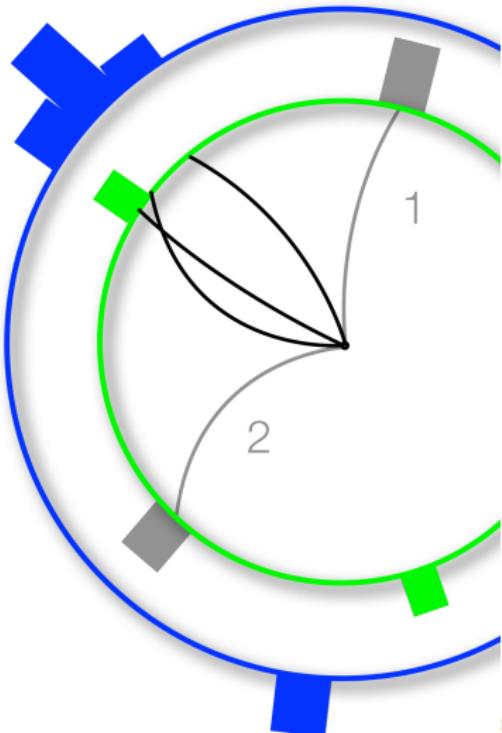
- 1. Create electrons for all tracks linked to a single ECAL cluster that pass the electron ID.



# Particle Flow

The particle flow algorithm attempts to reconstruct all stable final state particles using all of the CMS sub-detectors.

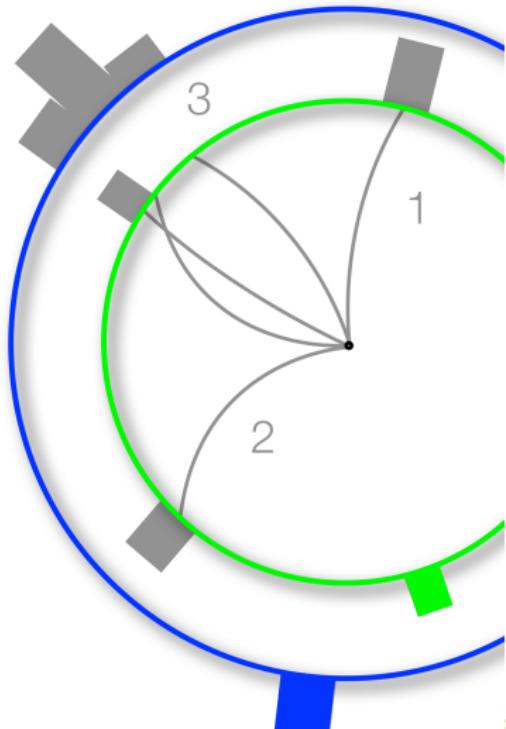
- 1. Create electrons for all tracks linked to a single ECAL cluster that pass the electron ID.
- 2. If the tracks fail the electron ID create a charged hadron, unless there are hits in the muon chambers, then create a muon.



# Particle Flow

The particle flow algorithm attempts to reconstruct all stable final state particles using all of the CMS sub-detectors.

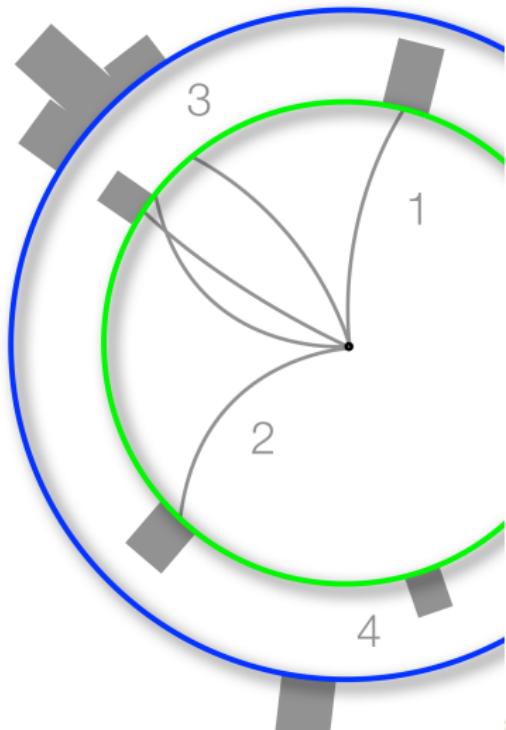
- 1. Create electrons for all tracks linked to a single ECAL cluster that pass the electron ID.
- 2. If the tracks fail the electron ID create a charged hadron, unless there are hits in the muon chambers, then create a muon.
- 3. For each HCAL cluster get all the linked tracks and ECAL clusters. Create a charged hadron for each track. If the energy in the HCAL is greater than the energy in the tracks, create photons or neutral hadrons equal to the missing energy.



# Particle Flow

The particle flow algorithm attempts to reconstruct all stable final state particles using all of the CMS sub-detectors.

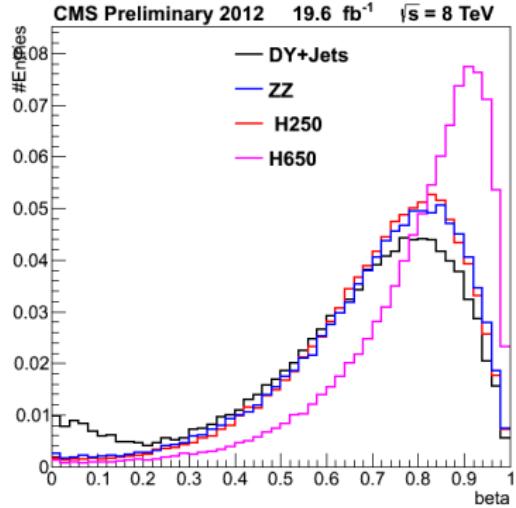
- 1. Create electrons for all tracks linked to a single ECAL cluster that pass the electron ID.
- 2. If the tracks fail the electron ID create a charged hadron, unless there are hits in the muon chambers, then create a muon.
- 3. For each HCAL cluster get all the linked tracks and ECAL clusters. Create a charged hadron for each track. If the energy in the HCAL is greater than the energy in the tracks, create photons or neutral hadrons equal to the missing energy.
- 4. For the remaining ECAL (HCAL) clusters that are not linked to tracks create a photon (neutral hadron).



# PileUp and Isolation

$$\beta = \frac{\sum P_T \text{ CJP from Vertex}}{\sum P_T \text{ all CJP}}$$

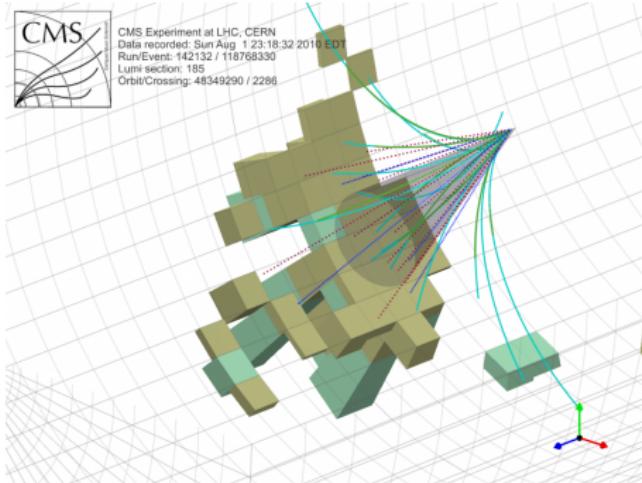
where CJP is charged jet particles.



Jets originating from pile-up interactions are removed by cutting on  $\beta > 0.2$ .

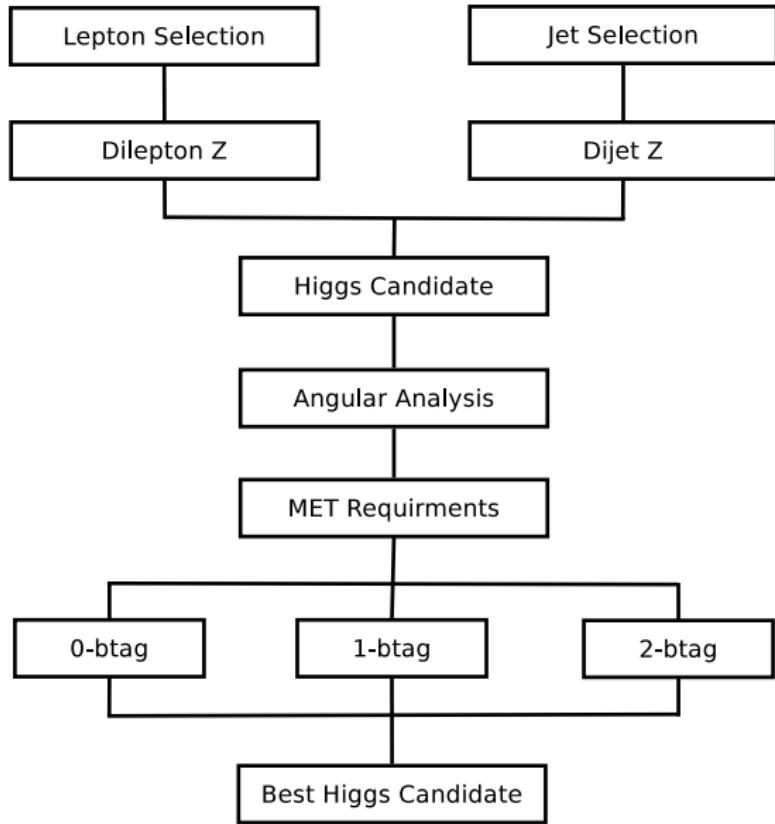
We also require that our objects are isolated by cutting on a cone around the primary vertex with radius:

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} > 0.5$$



Source: <http://www.quantumdiaries.org/2011/06/01/anatomy-of-a-jet-in-cms/>

# Analysis workflow



# Physics Objects and Preselection

## Triggers

- HLT\_Ele17\_CaloIdT\_TrkIdVL\_CaloIsoVL\_TrkIsoVL\_Ele8\_CaloIdT\_TrkIdVL\_CaloIsoVL\_TrkIsoVL
- HLT\_Mu17\_Mu8

We are using the recommendations from the CMS physics object groups for lepton identification and isolation. These values can be seen in the backup slides.

After we have the objects, we apply additional preselection cuts to the leptons.

### Electrons

Physics Object

- Identification and Isolation

#### Preselection

- $p_T > 40/20$  GeV
- $|\eta| < 2.5$

### Muons

Physics Object

- Identification and Isolation

#### Preselection

- $p_T > 40/20$  GeV
- $|\eta| < 2.4$

### Jets

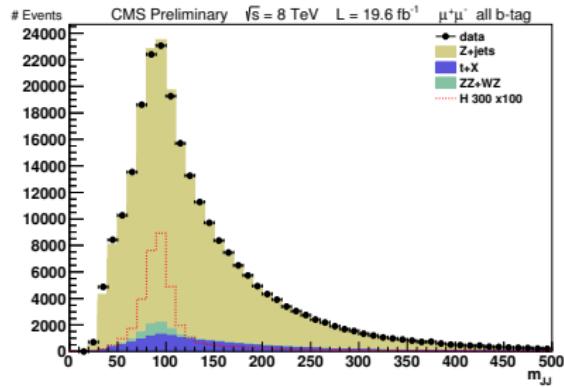
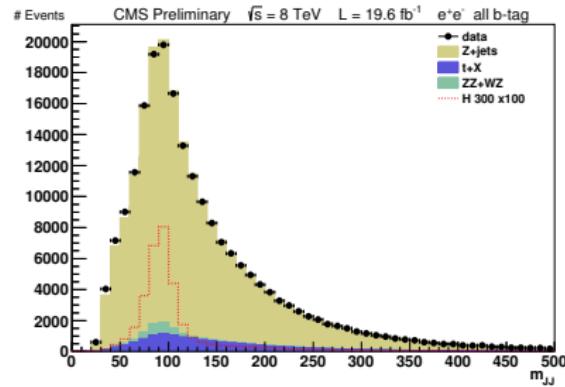
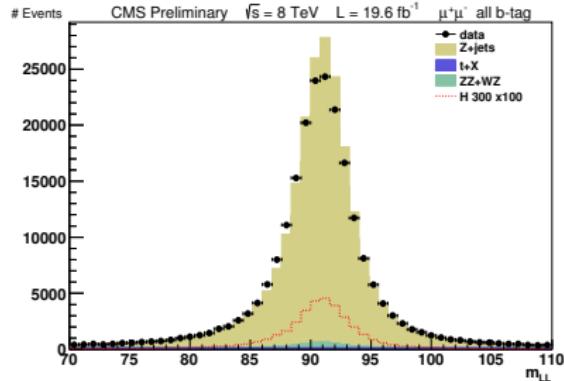
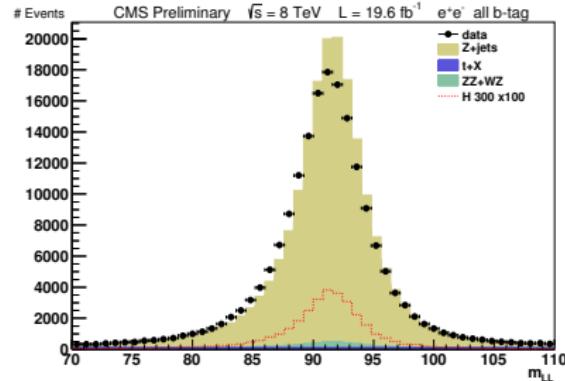
Physics Object

- Identification and Isolation

#### Preselection

- $p_T > 30$  GeV
- $|\eta| < 2.4$

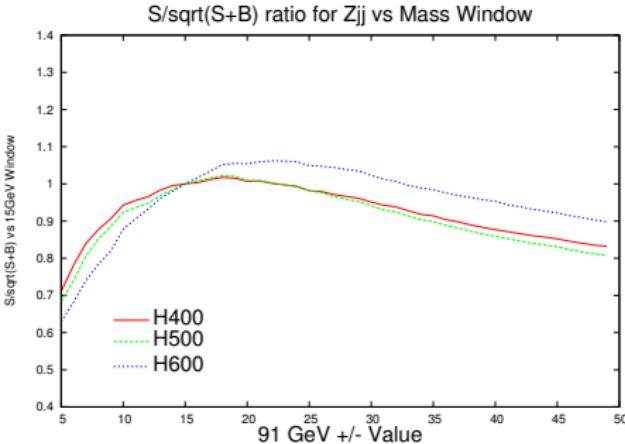
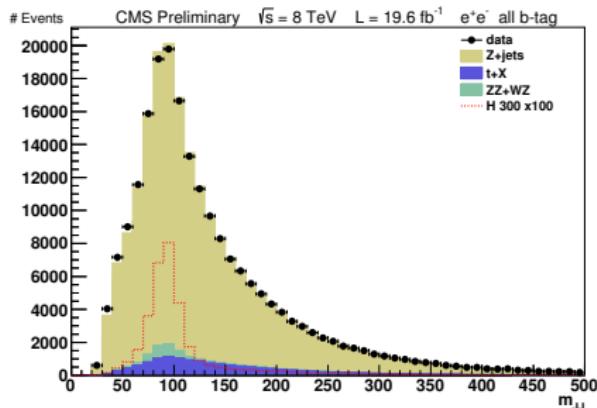
# $m_{\parallel}$ and $m_{jj}$ distributions



# Mass Cut around Z Peaks

Our main background is Z+jets where the Z decays into leptons. For the signal the di-jet mass peak should be at the nominal Z boson mass of 91 GeV. We can cut on either side of this peak to reject a large amount of the Z+jets background. We optimized these cuts to

$$\text{maximize } \frac{S}{\sqrt{S+B}}.$$



The same optimization is done for the  $m_{jj}$  distribution.

# Signal and Sideband Regions

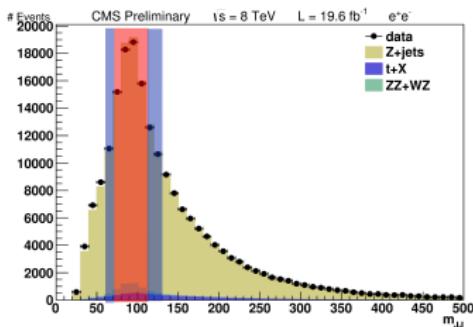
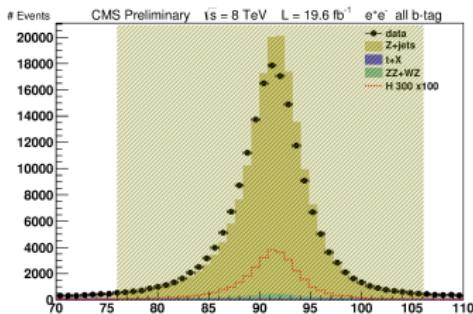
Optimal cuts around the  $m_{\parallel}$  mass peak were also calculated. A side band region is defined in the  $m_{jj}$  mass distribution. This side band region is used to get the Z+jets background from data.

Signal:

- $76 < m_{\parallel} < 106$  GeV
- $71 < m_{jj} < 111$  GeV

Sideband:

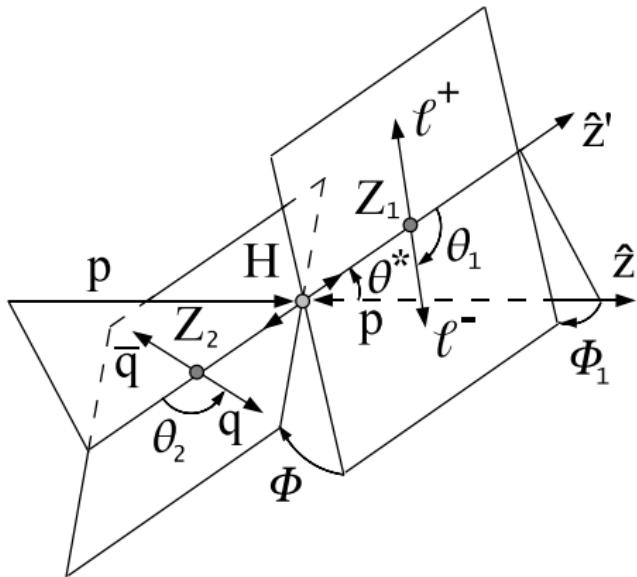
- $76 < m_{\parallel} < 106$  GeV
- $60 < m_{jj} < 71$  GeV
- $111 < m_{jj} < 130$  GeV



At this point we keep each combination that satisfy the previous criteria as a Higgs candidate. There can be multiple Higgs candidates per event at this stage.

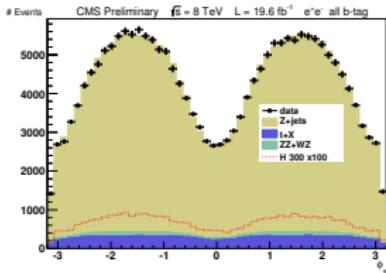
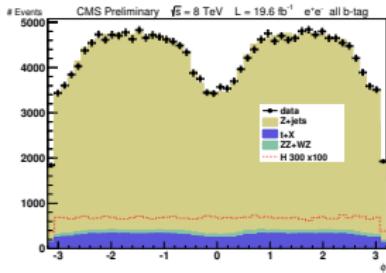
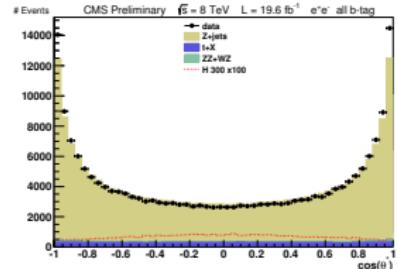
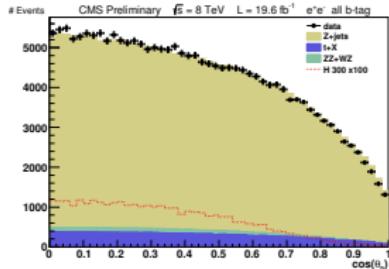
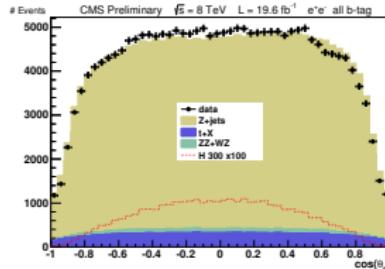
# Helicity and Production Angles

Final state kinematics completely determined by 5 angles.

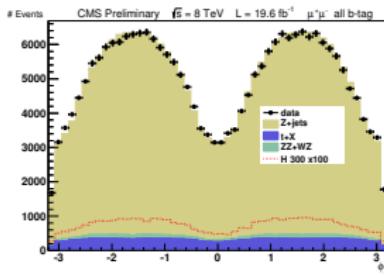
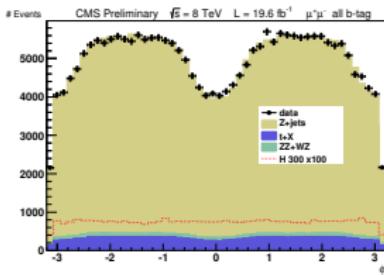
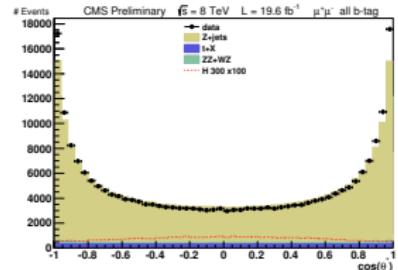
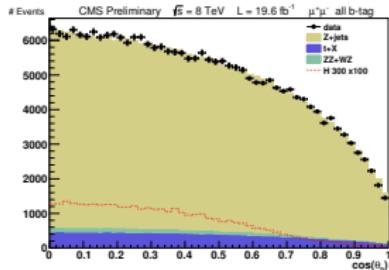
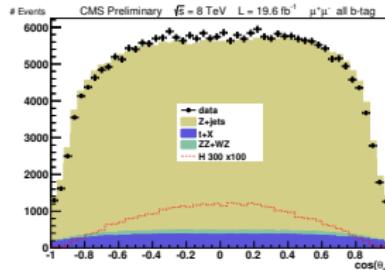


$$\cos(\theta^*), \cos(\theta_1), \cos(\theta_2), \Phi, \Phi_1$$

# Electron Helicity and Production Angles

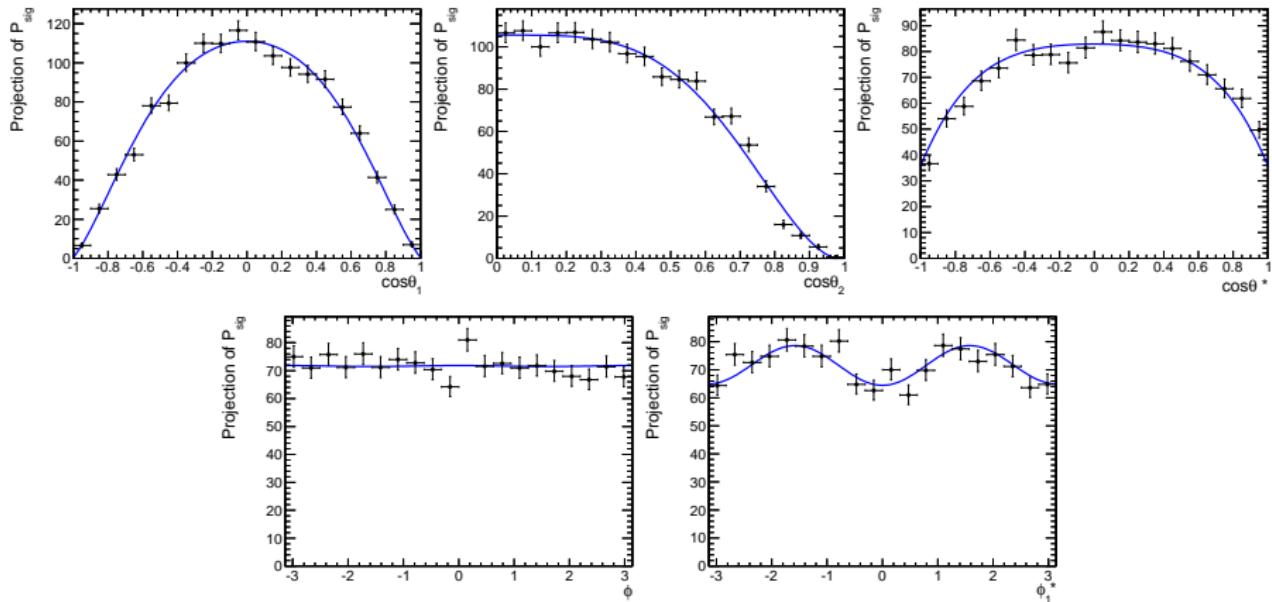


# Muon Helicity and Production Angles



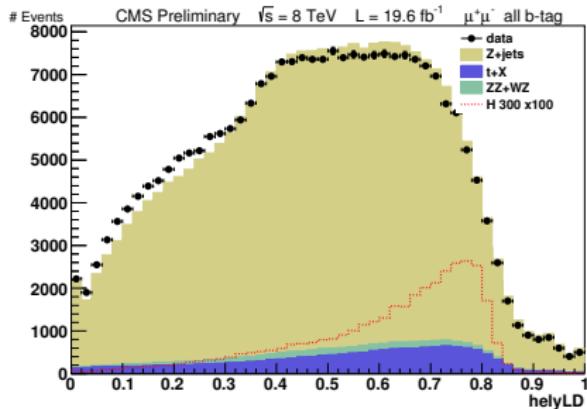
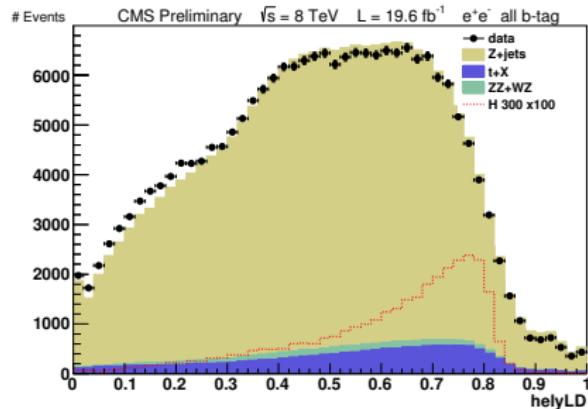
# Signal Angular Distribution Fits

Example fits at for 500 GeV (475 - 550 GeV).



# Helicity LD Distribution and Cuts

$$LD = \frac{P_{sig}}{P_{sig} + P_{bkg}}$$

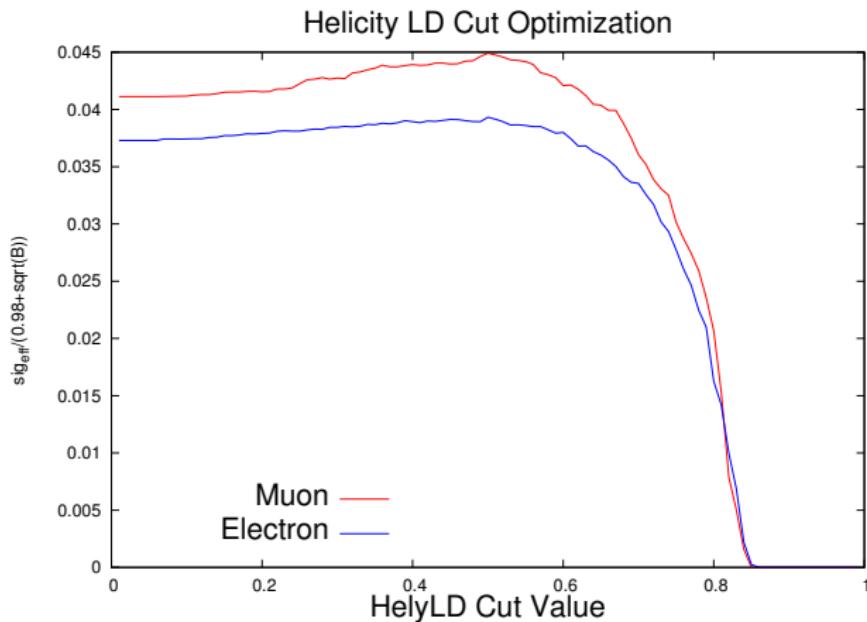


An explanation of the shape discrepancy is given in the backup slides.

# HelyLD Optimization

We do an optimization for the cut on the HelyLD maximizing the Punzi equation:

$$\frac{\# \text{ signal}}{\# \text{ total signal}} / 0.98 + \sqrt{B}$$



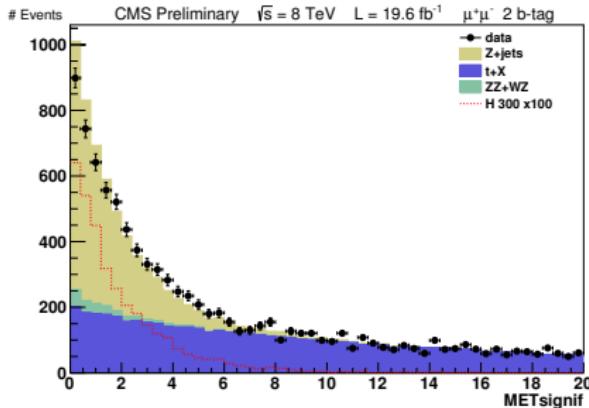
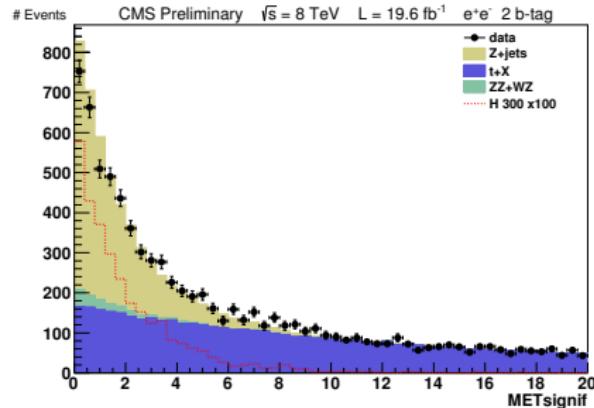
We cut on 0.5 which is within 5% of the optimal value for all Higgs mass values, b-tag regions, and lepton types.

# Missing Transverse Energy

Our signal should not have any Missing Transverse Energy ( $E_T^{miss}$ , MET) from neutrinos. MET significance is a measure of how likely the MET we see is real(not from detector effects).

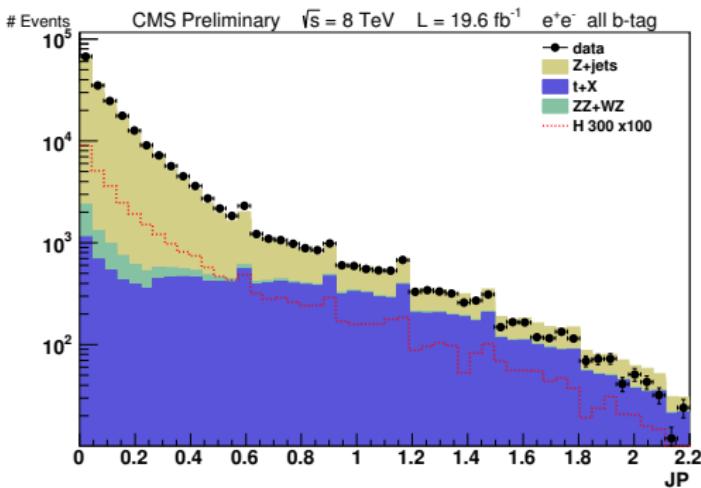
$$MET \text{ Significance} = 2\ln\lambda(MET) = 2\ln \frac{L(MET_{true} = MET_{measured})}{L(MET_{true} = 0)}$$

This is particularly useful to suppress the  $t\bar{t}$  background in the 2-tag region (b-tagging on next slide), but we apply a cut on MET significance  $> 10$  to all categories.



# *b*-tagging

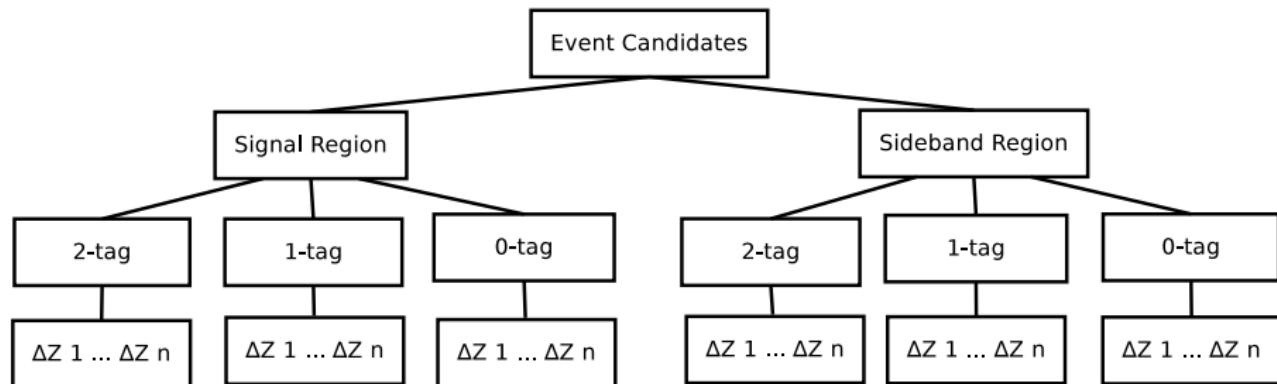
- Using JP algorithm
- We are looking for heavy quarks (b,c)
- There are “loose”(0.275) and “medium”(0.545) working points that are defined by the b-tag group that we use to classify our events.



0 - tag	Both Jets < Loose
1 - tag	> Loose and < Medium
2 - tag	> Loose and > Medium

# Best Higgs Candidate

When there are multiple candidates in the same event we apply a string of logic to select which is the candidate that we are going to keep.



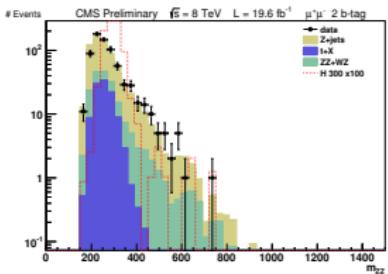
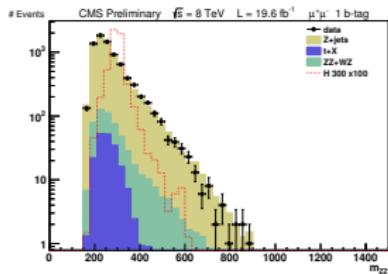
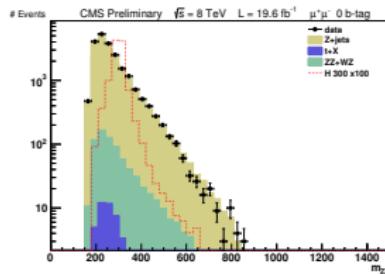
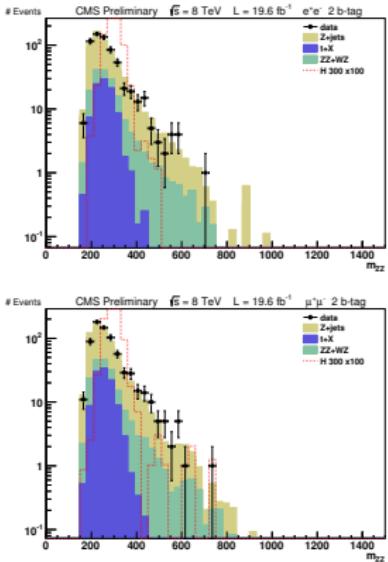
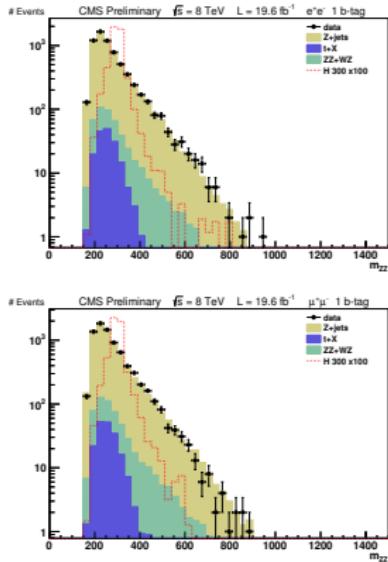
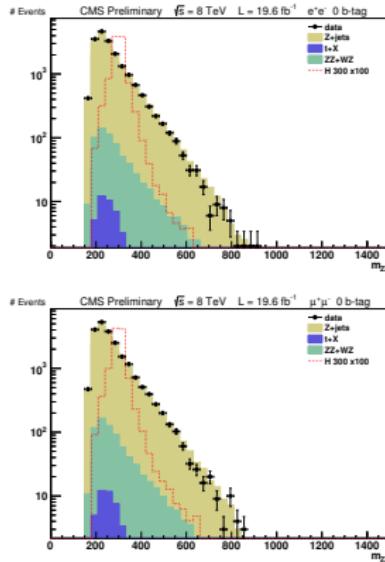
$$\Delta Z = |M_{\ell^+\ell^-} - m_Z| + |M_{J\bar{J}} - m_Z|$$

$m_Z$  is the nominal mass of the Z boson

$\Delta Z_1 \dots Z_n$  is the candidates sorted by  $\Delta Z$  smallest to largest.

When all candidates are sorted we keep the candidate that is furthest to the left.

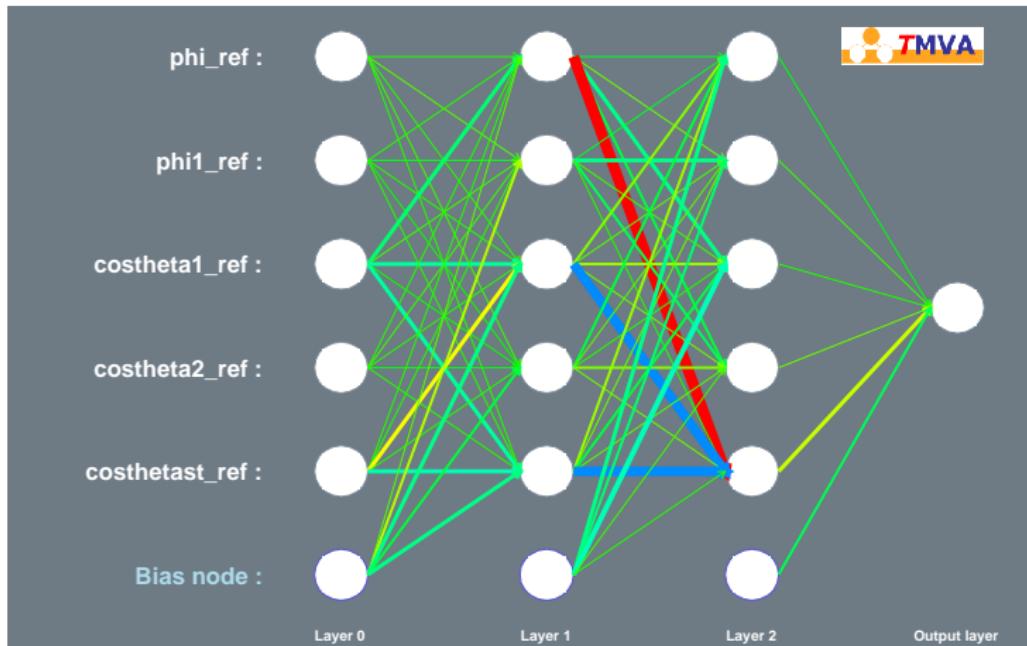
# Final Region $m_{ZZ}$ Plots



# Outline

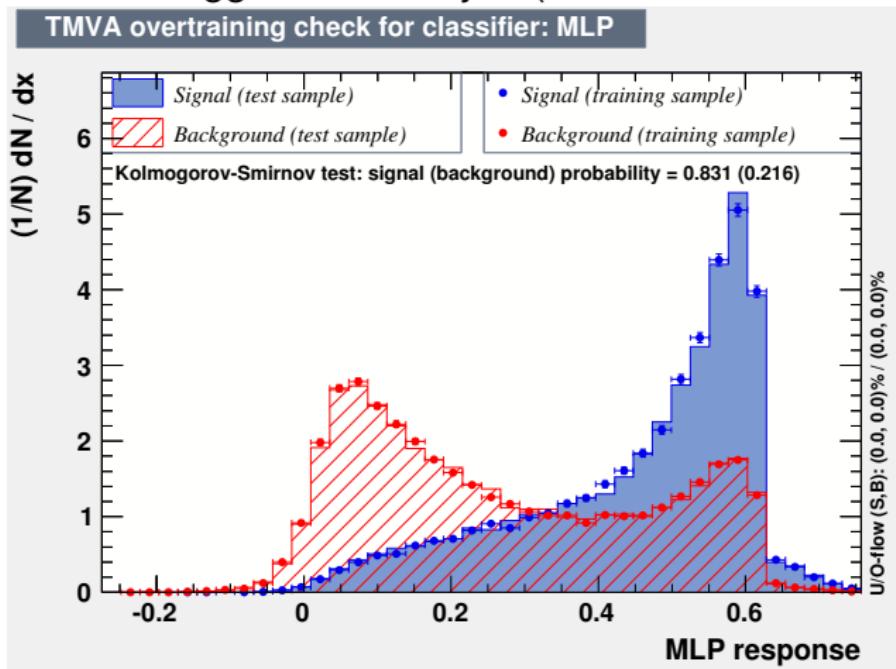
- 1 LHC Search for Higgs Boson
- 2 Event Selection
- 3 Cross Check and Statistical Analysis
- 4 Results
- 5 Conclusion

# Neural Network with TMVA Package



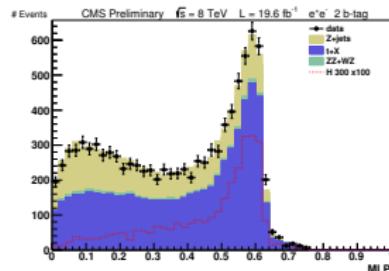
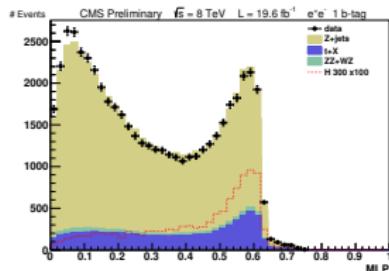
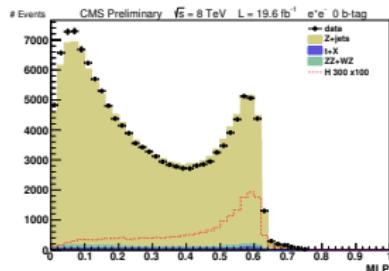
# Neural Network Training and Testing

The trainings are done after preselection and additionally require at least one B-tagged Medium jet. (Trained on H 400 GeV)

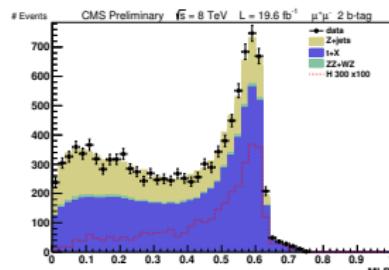
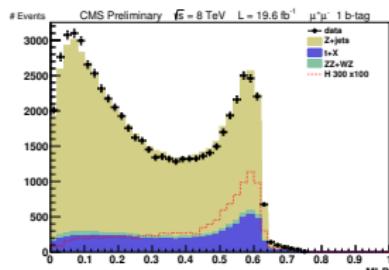
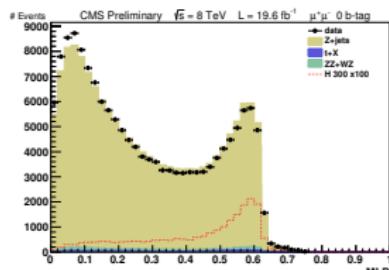


# MLP - In b-tag regions

## Electrons (zero,one,two)

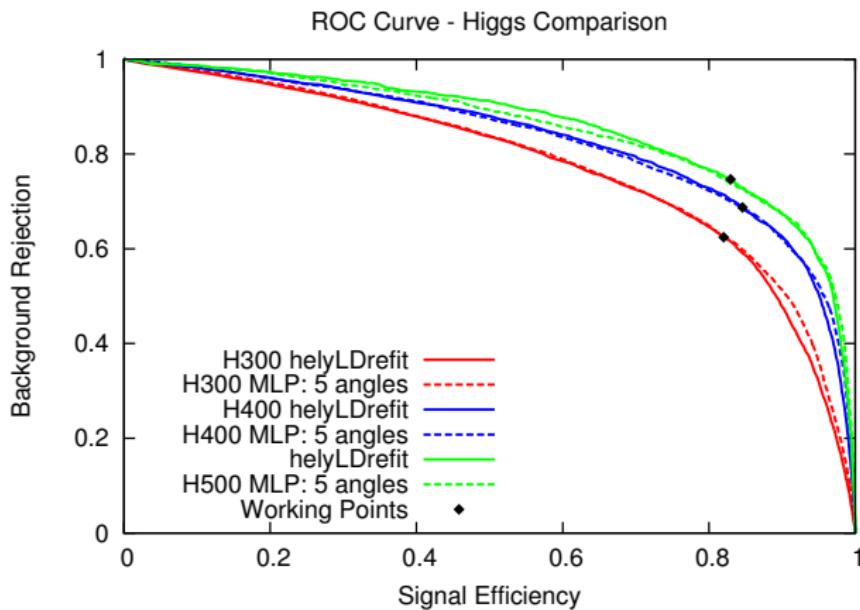


## Muons (zero,one,two)



# MLP vs Helicity LD

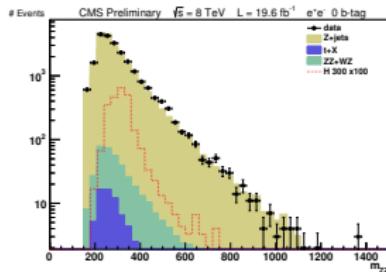
The working point is the background rejection point that we currently achieve in the two tag region in our analysis.



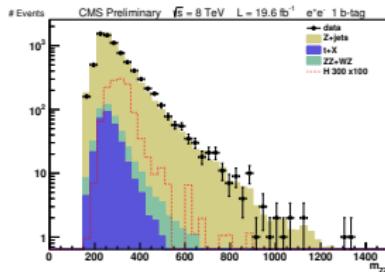
# $m_{ZZ}$ plots - Side Band Region

## Electrons

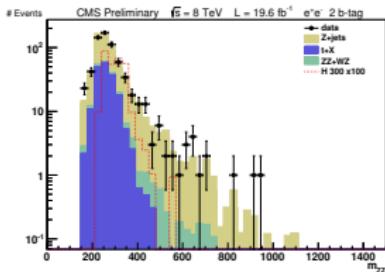
0-tag



1-tag

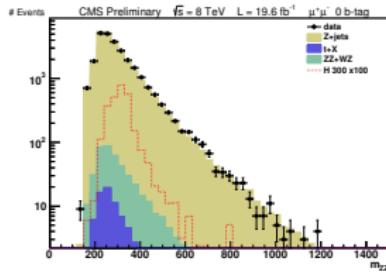


2-tag

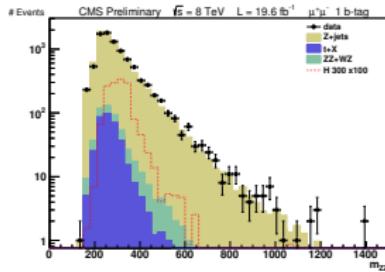


## Muons

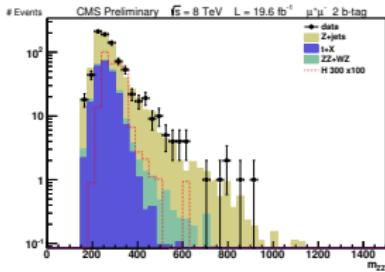
0-tag



1-tag

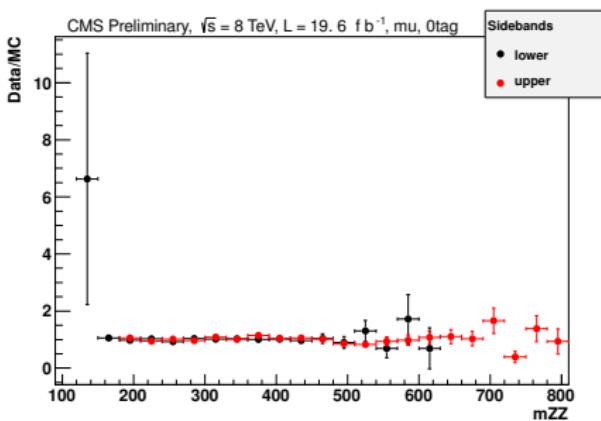
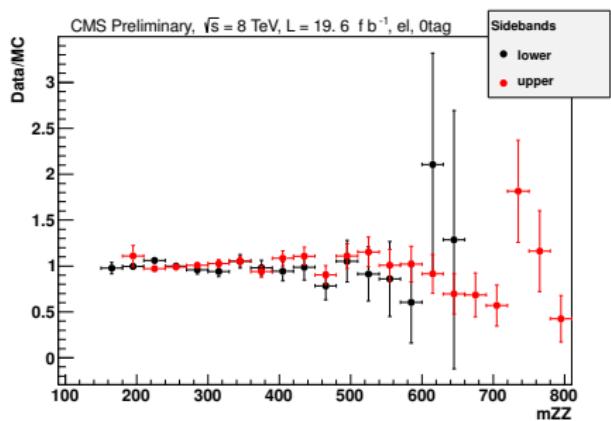


2-tag



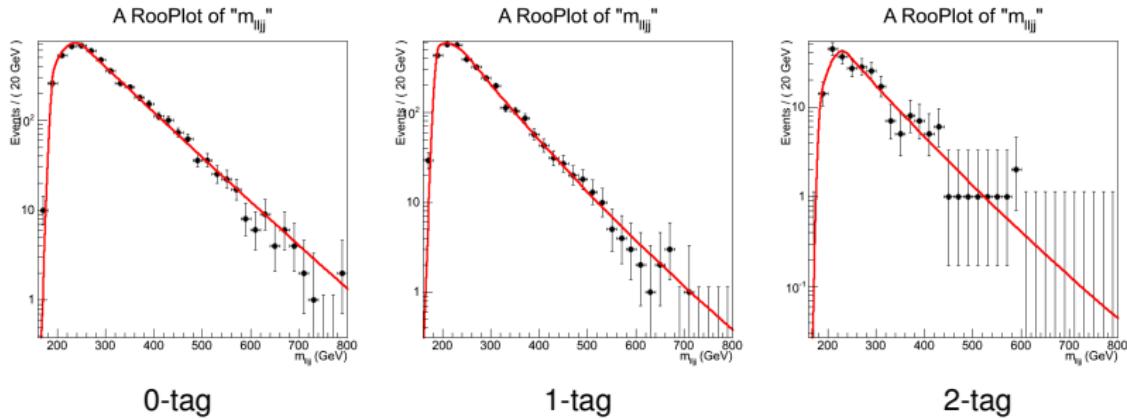
# Upper vs Lower Sideband

We also have very similar performance between the upper and lower sidebands in the  $m_{ZZ}$  distribution so we are able to use them together.



# Background Correction

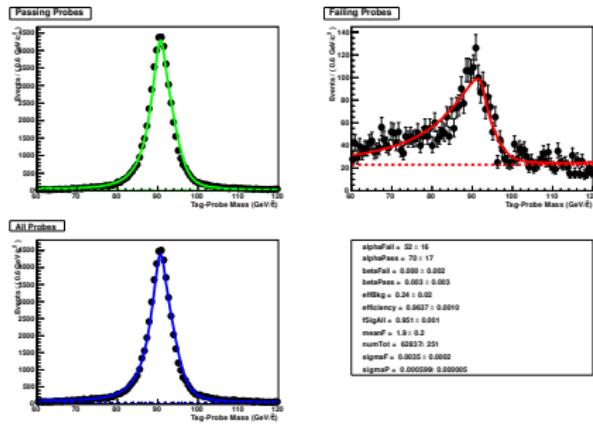
- We use the  $m_{jj}$  sideband in data to get the normalization and a MC shape correction.
- The fit is to a Fermi Function convoluted with a CrystalBall function



# Tag and Probe

- Method to use Z boson to calculate efficiency of Data and Monte Carlo.
- One lepton is “good” (tag) and the other is used for the calculations(probe).
- Comparing Monte Carlo to data gives us Scale Factors.

SuperCluster to GSF Electron



# Signal Systematics

Source	0 b-tag	1 b-tag	2 b-tag
Muon trigger & ID	2.7%		
Electron trigger & ID	2%		
Electron energy scale	0.2%		
Muon momentum scale	0.1%		
Jet reconstruction	1-4%		
<i>b</i> -tagging eff. and mistag rate	1-4%	1-5%	5-8%
MET	< 1%		
Pile-up	1-2%		
Production mechanism (PDF)	1.5%		
Production mechanism (lineshape)	0-3%		
Luminosity	4.4%		
Higgs cross-section	13-15%		

# Background Systematics

Source	Normalization			Shape		
	0 b-tag	1 b-tag	2 b-tag	0 b-tag	1 b-tag	2 b-tag
Muon trigger & ID		2.7%				
Muon momentum scale		0.1%				
Electron trigger & ID		2.0%				
Electron energy scale		0.5%				
Jet energy scale		5.5%				0-4%
<i>b</i> -tagging efficiency SF	+0.4%	-0.8%	-4.5%			
Mistag SF	-1.9%	+7.8%	+6.2%			
MET		0.3%				
Pile-up		0.1%				
$p_T^{\ell\ell jj}$ weighting		0.8%				0-3%
Diboson cross section		15%				
Luminosity		4.4%				
Control Region				0-15%	0-30%	0-40%

# Outline

1 LHC Search for Higgs Boson

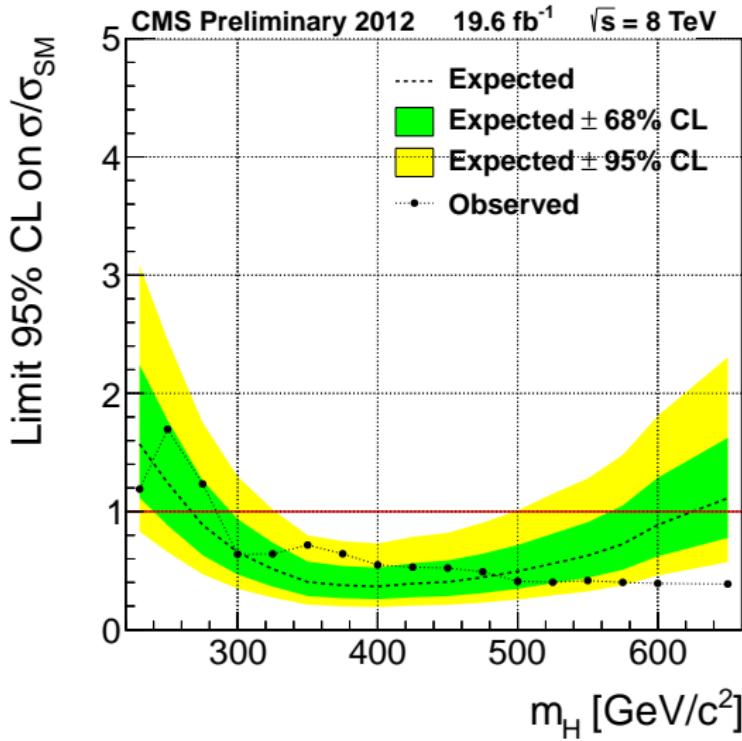
2 Event Selection

3 Cross Check and Statistical Analysis

4 Results

5 Conclusion

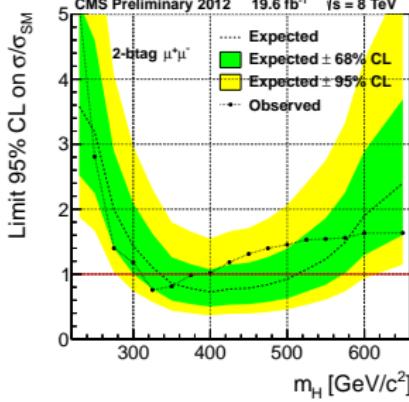
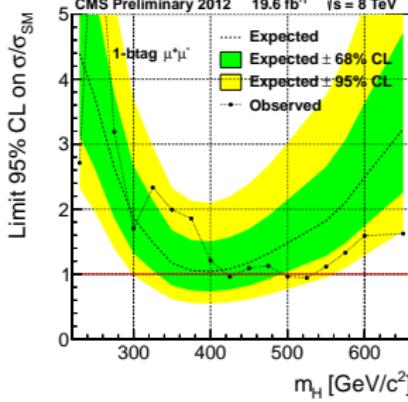
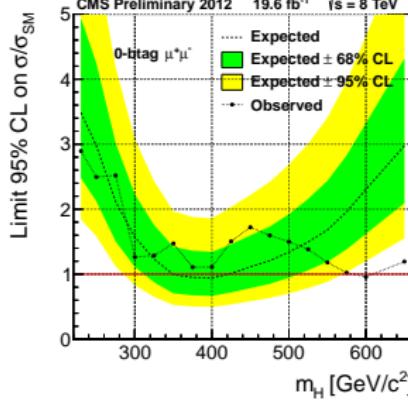
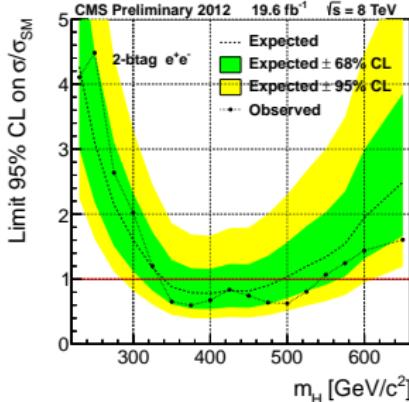
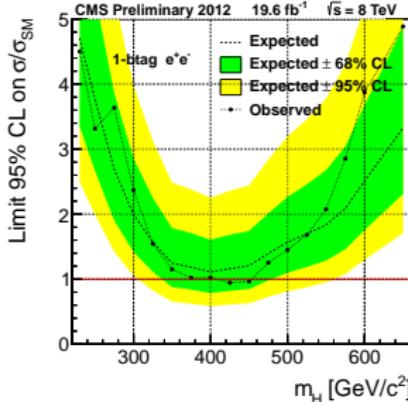
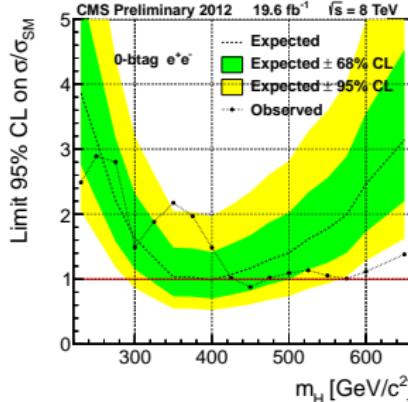
# 8 TeV Results



Observed (solid) and expected (dashed) 95% CL upper limit on the ratio of the production cross section to the SM expectation for the Higgs boson obtained using the  $\text{CL}_s$  technique.

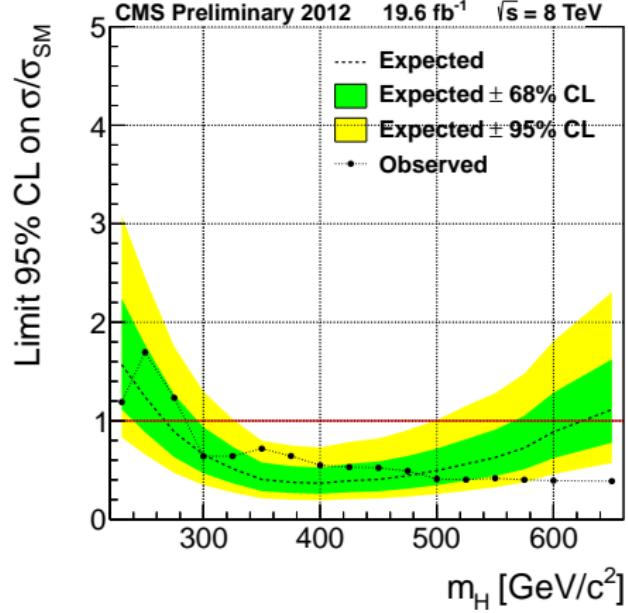
The 68% and 95% ranges of expectation for the background-only model are also shown with green and yellow bands, respectively. The solid line at 1 indicates the expectation for a SM-Higgs-like boson.

# 8 TeV Individual Channel Results

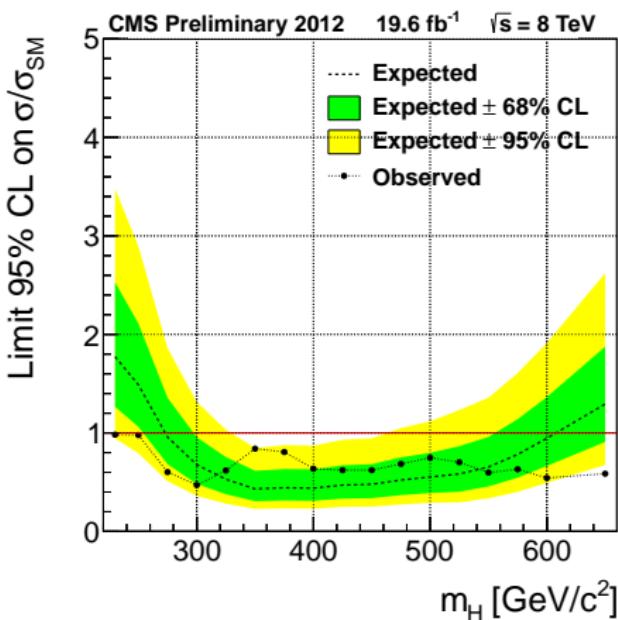


# HelyLD vs MLP

helyLD

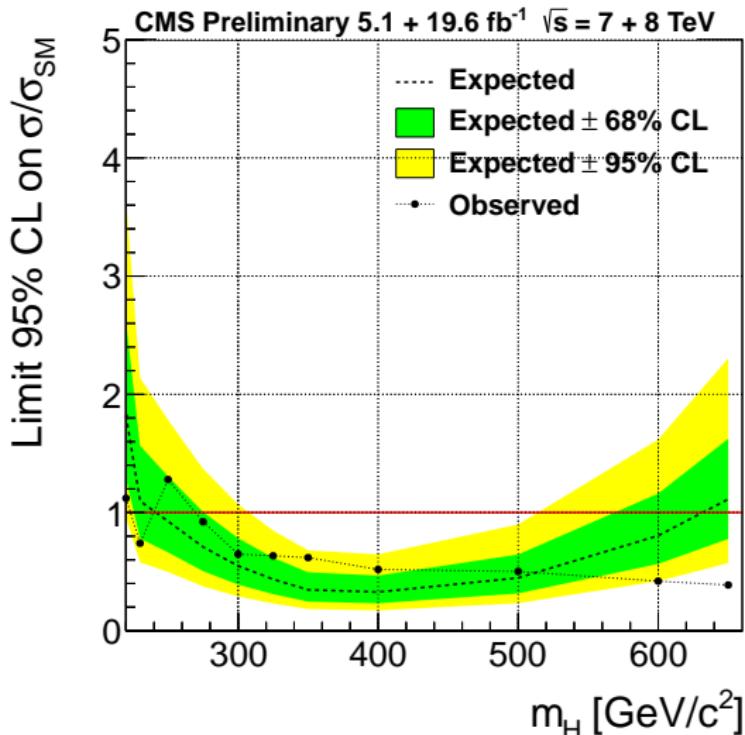


MLP



The helyLD and MLP discriminators give very similar results, with the helyLD doing about 5% better overall.

# 7 TeV and 8 TeV HelyLD Combined Results



The 7 TeV 2l2q group results are in the backup slides.

# Outline

1 LHC Search for Higgs Boson

2 Event Selection

3 Cross Check and Statistical Analysis

4 Results

5 Conclusion

## Summary

- No evidence of a Higgs boson signal is found.
- At  $95\%_{CL}$  a Standard Model Higgs boson is excluded in the range of 285 GeV to 650 GeV.
- Similar Results are obtained using both the HelyLD and the MLP classifier using the kinematic information of the events.

## Future Channel Plans

- Extend analysis range up to 1 TeV.
- Separate treatment of gluon gluon fusion and vector boson fusion.
- Search for additional Higgs.
- Extend analysis to search for the Graviton.

# Thank You

# Backup

## BACKUP

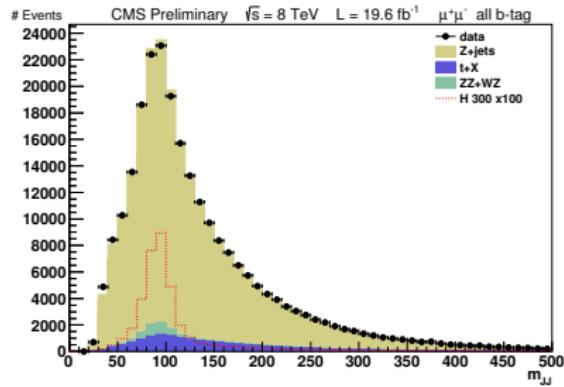
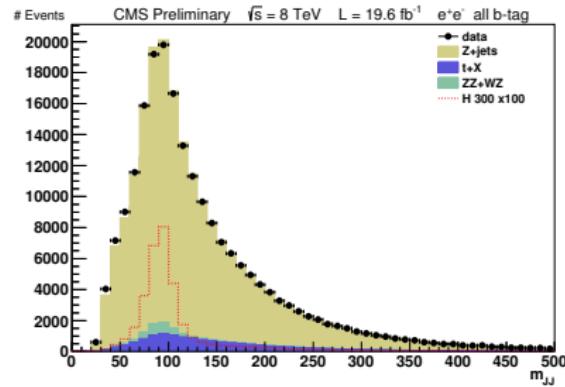
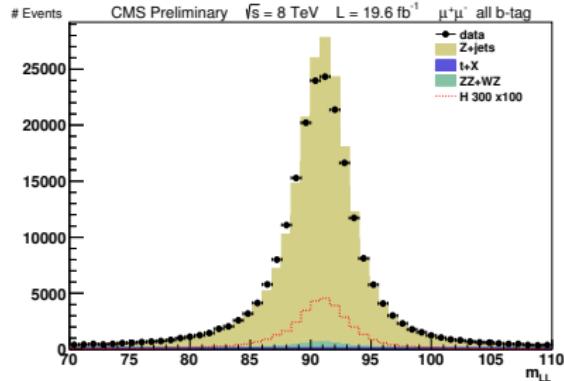
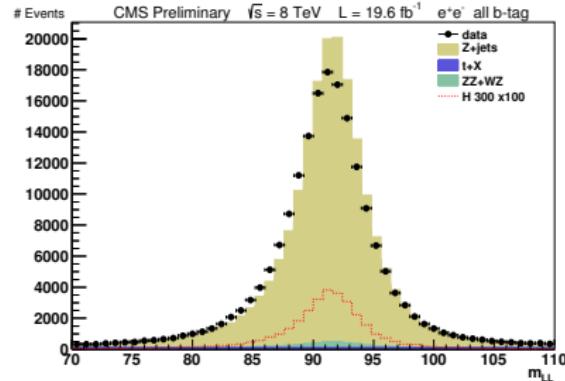
# Electron requirements for the Loose working point

Variable	Barrel cut	Endcap cut
$\Delta\eta_{trk,supercluster}$	< 0.007	< 0.009
$\Delta\phi_{trk,supercluster}$	< 0.15	< 0.1
$\sigma_{i\eta,i\eta}$	< 0.01	< 0.03
$H/E$	< 0.12	< 0.10
$d_0$ (wrt primary vertex)	< 0.2 mm	< 0.2 mm
$d_z$ (wrt primary vertex)	< 2 mm	< 2 mm
$ 1/E - 1/p $	< 0.05	< 0.05
$I_{PF, corr}/p_T$	< 0.15	< 0.15
Missing hits	$\leq 1$	$\leq 1$
Conversion vertex fit prob.	$< 10^{-6}$	$< 10^{-6}$

# Muon requirements for the Tight working point

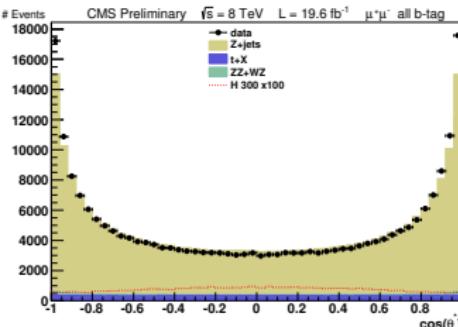
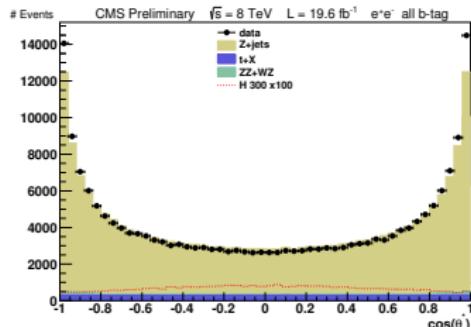
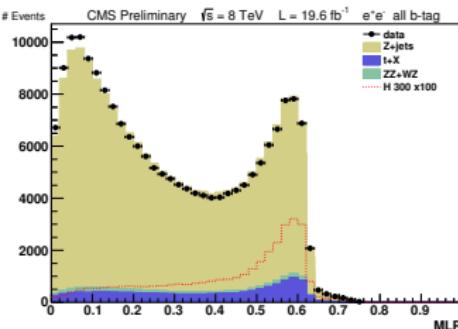
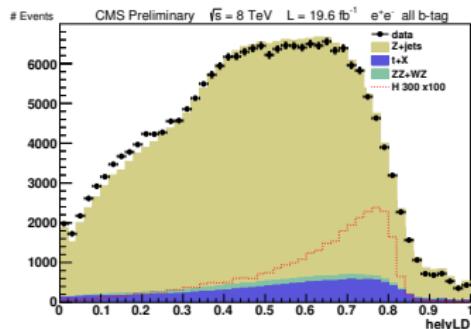
Variable	Cut
isGlobalMuon	True
isPFMuon	True
$\chi^2/ndof$ (global fit)	< 10
Muon chamber hits in global fit	> 0
Muon stations with muon segments	> 1
$d_{xy}$ (from tracker, wrt primary vertex)	< 2 mm
$d_z$ (from tracker, wrt primary vertex)	< 5 mm
Valid pixel hits (tracker track)	> 0
Tracker layers with hits	> 5
$I_{PF, corr}/p_T$	< 0.12

# $m_{\parallel}$ and $m_{jj}$ distributions



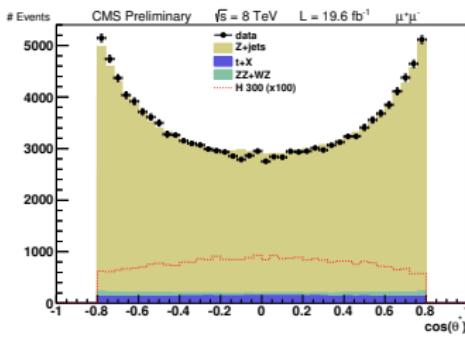
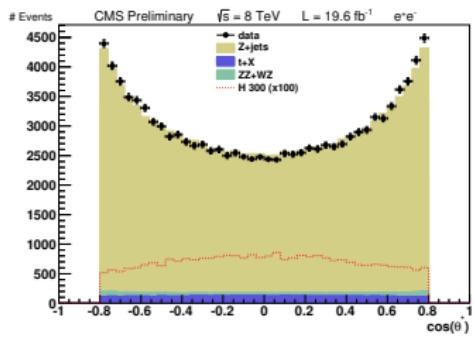
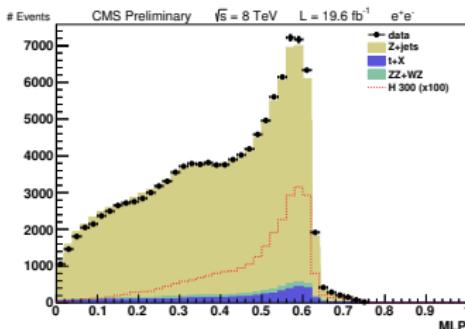
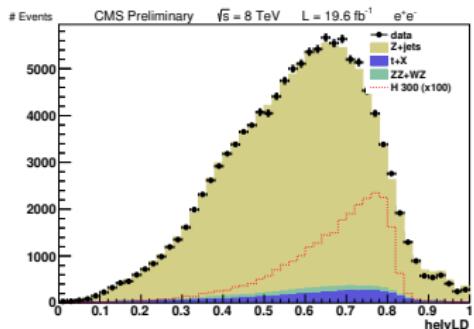
# MLP and HelyLD Data/MC Discrepancy

Both the HelyLD and the MLP don't agree in the background like regions between data and MC. This is because  $\cos(\theta^*)$  is not well modeled in MC.



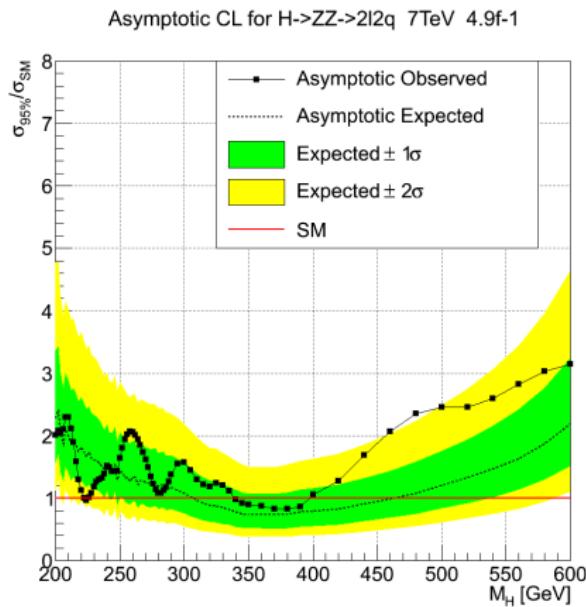
# Solution

This is a MC issue that can be fixed by cutting on  $|\cos(\theta^*)| > 0.8$  or by refitting MC to data.



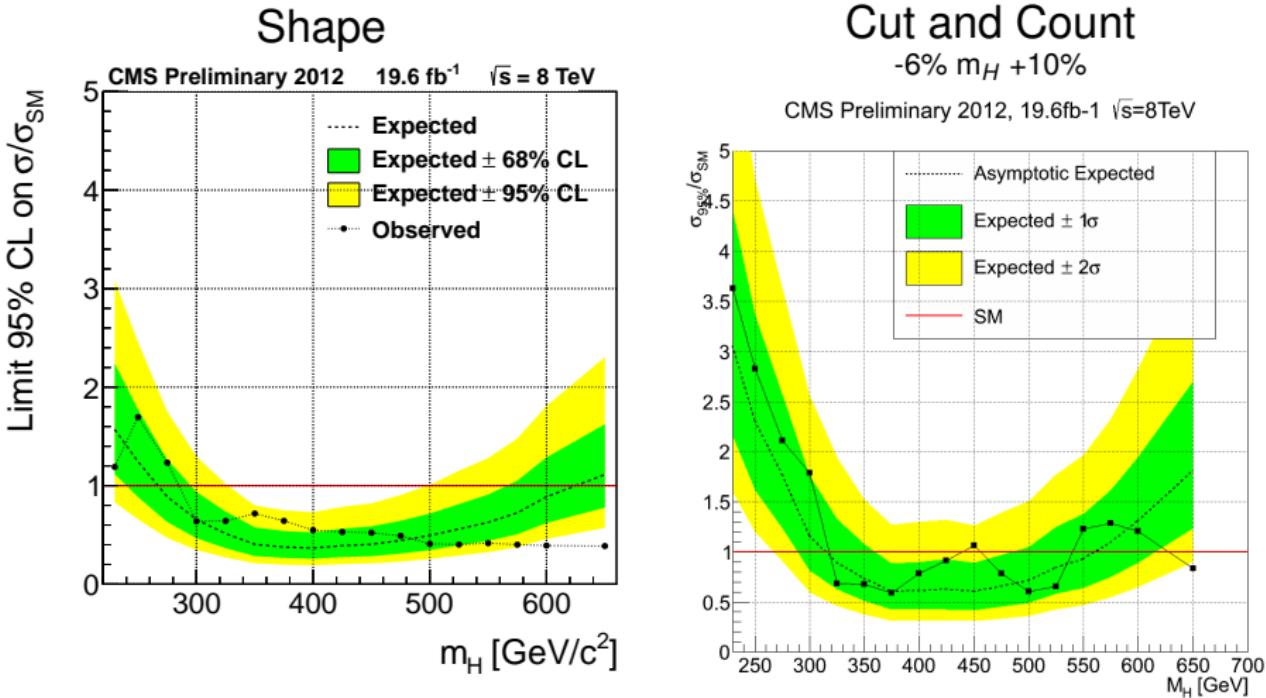
# 2011 Results

Limit on the expected 95% CL upper limit on the product of the Higgs boson production cross section and the branching fraction of  $H \rightarrow ZZ$  (black dots.), for the recorded luminosity in 2011 of  $4.9 \text{ fb}^{-1}$  at 7TeV. Yellow and Green bands represent the 68% and 95% ranges of expectation.



CMS Collaboration, Search for a Higgs boson in the decay channel  $H \rightarrow ZZ \rightarrow \ell^+\ell^- q\bar{q}$ , JHEP 04 (2012) 036, doi:10.1007/JHEP04(2012)036, arXiv:1202.1416.

# Shape vs Cut and Count



The shape analysis does about 15% better than the cut and count, but both give very similar results.