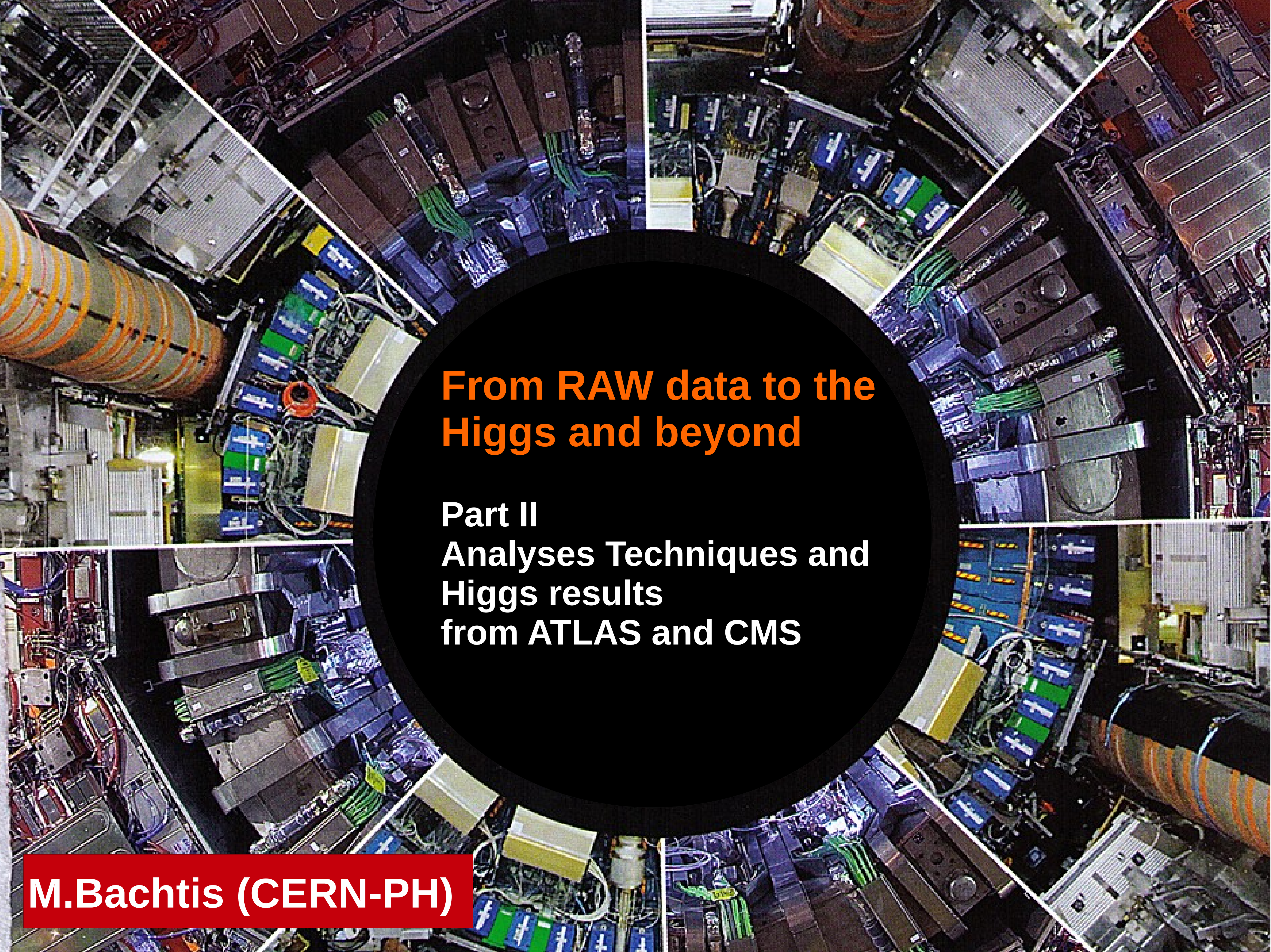


From RAW data to the Higgs and beyond

Part II
Analyses Techniques and Higgs results from ATLAS and CMS

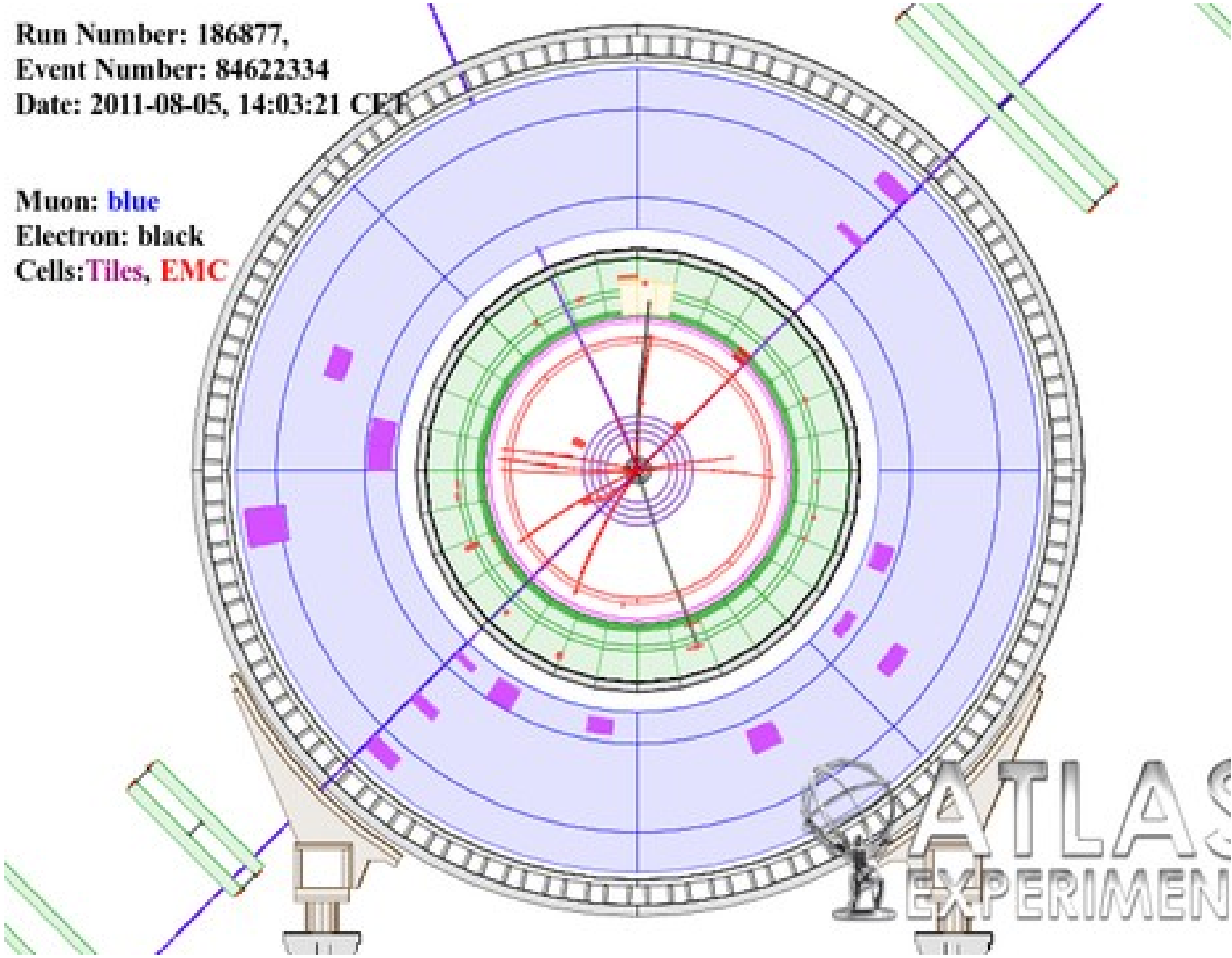
M.Bachtis (CERN-PH)

Outline

- Review of the big 5 Higgs analyses + ttH
  - H → ZZ → 4l
  - H → gamma gamma
  - H → WW → 2l2nu
  - H → tautau
  - H → bb
- Specific analyses techniques used in each final state
- Common analysis techniques
  - Mass measurements, signal extraction

The golden channel (H → ZZ* → 4l)

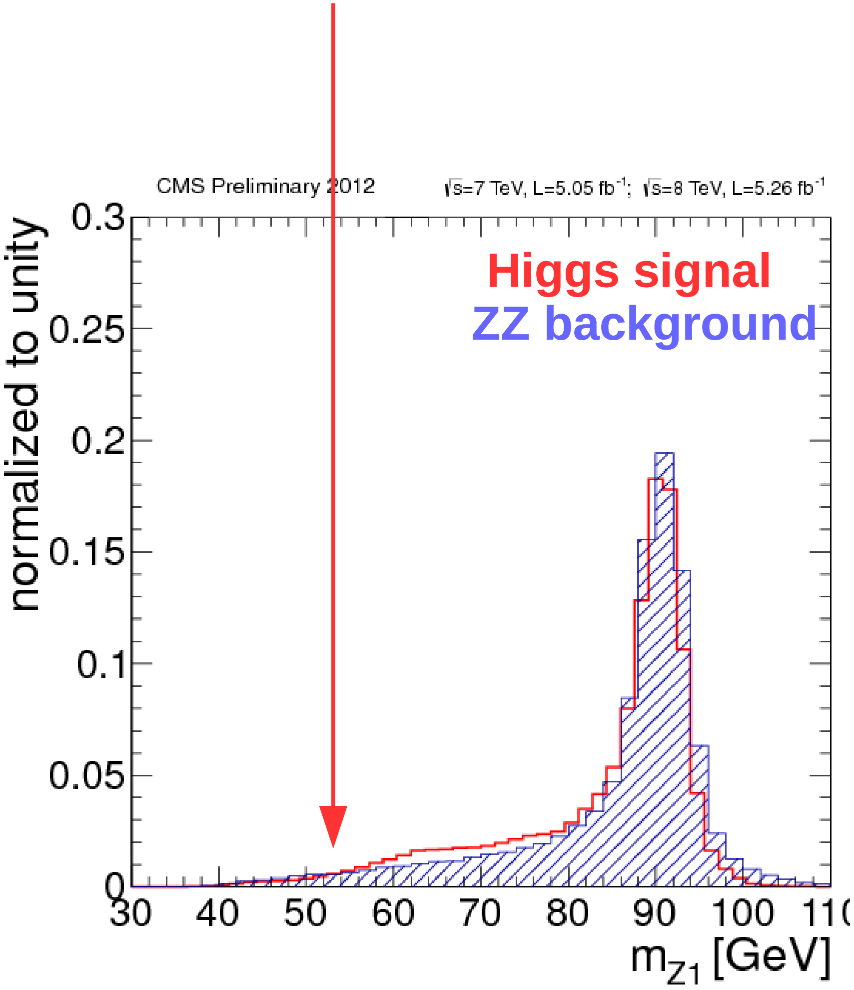
- Require 4 high quality leptons
  - Electrons or muons
- Pt thresholds need to be small!
  - O(5 GeV) to not loose efficiency
- 4 leptons can be grouped to two Z pairs
  - Usually Z_1 defined as closest to Z mass

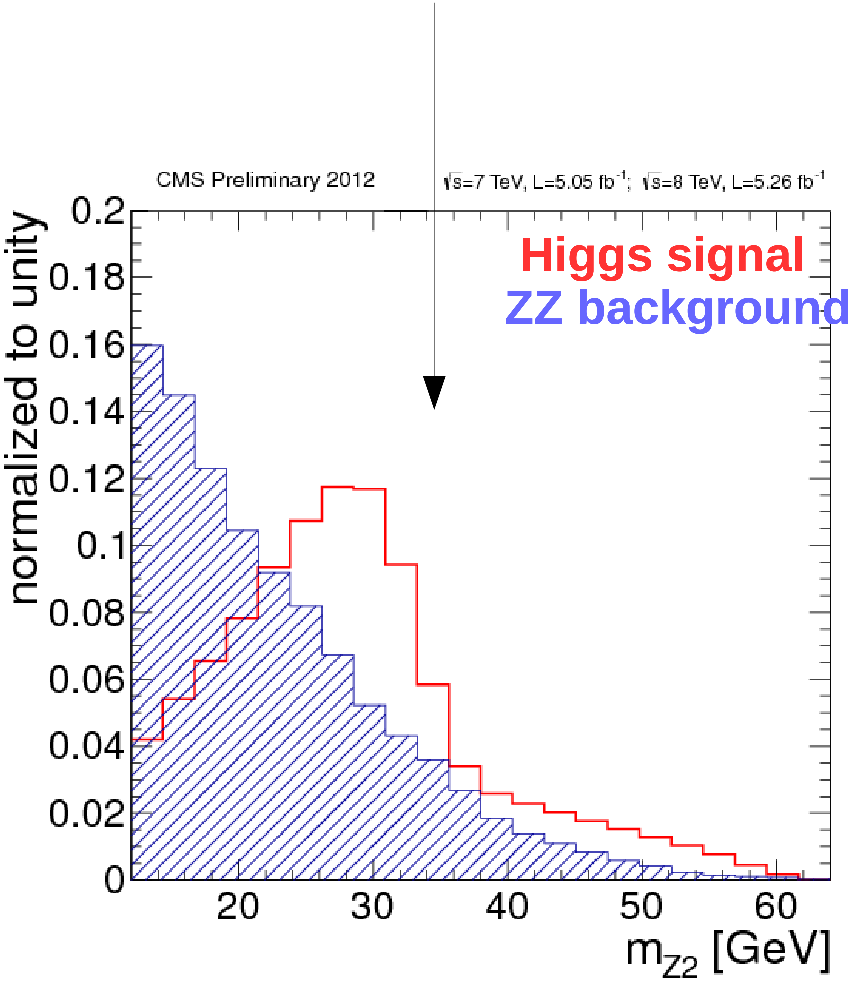


Z_1 and Z_2

25% of the cases both Z virtual
Nominal cut of 40 GeV to not loose efficiency

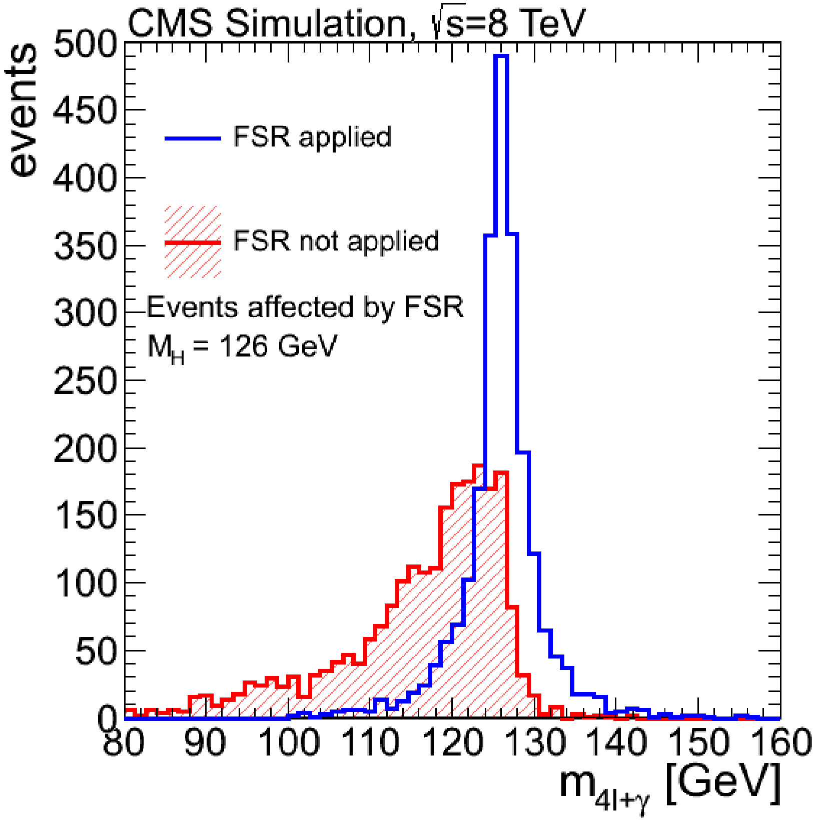
Virtual Z spectrum for signal
Constrained by the Higgs mass
O(10 GeV) to reject background

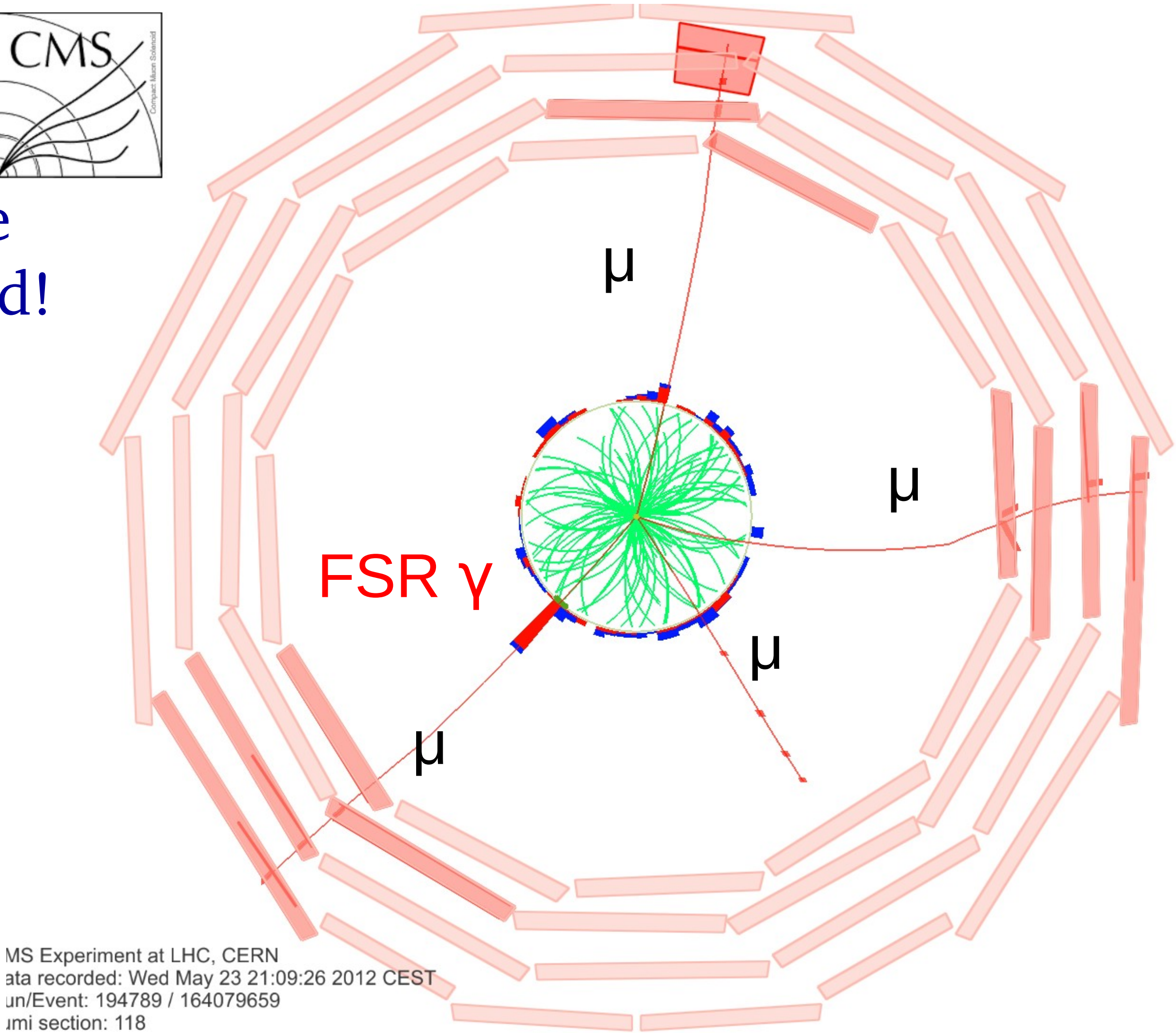




Recovery of final state radiation

- Leptons radiate photons
  - In case of high energy radiation, photon can be detected and recombined!
- ~5-10% of the ZZ* events affected



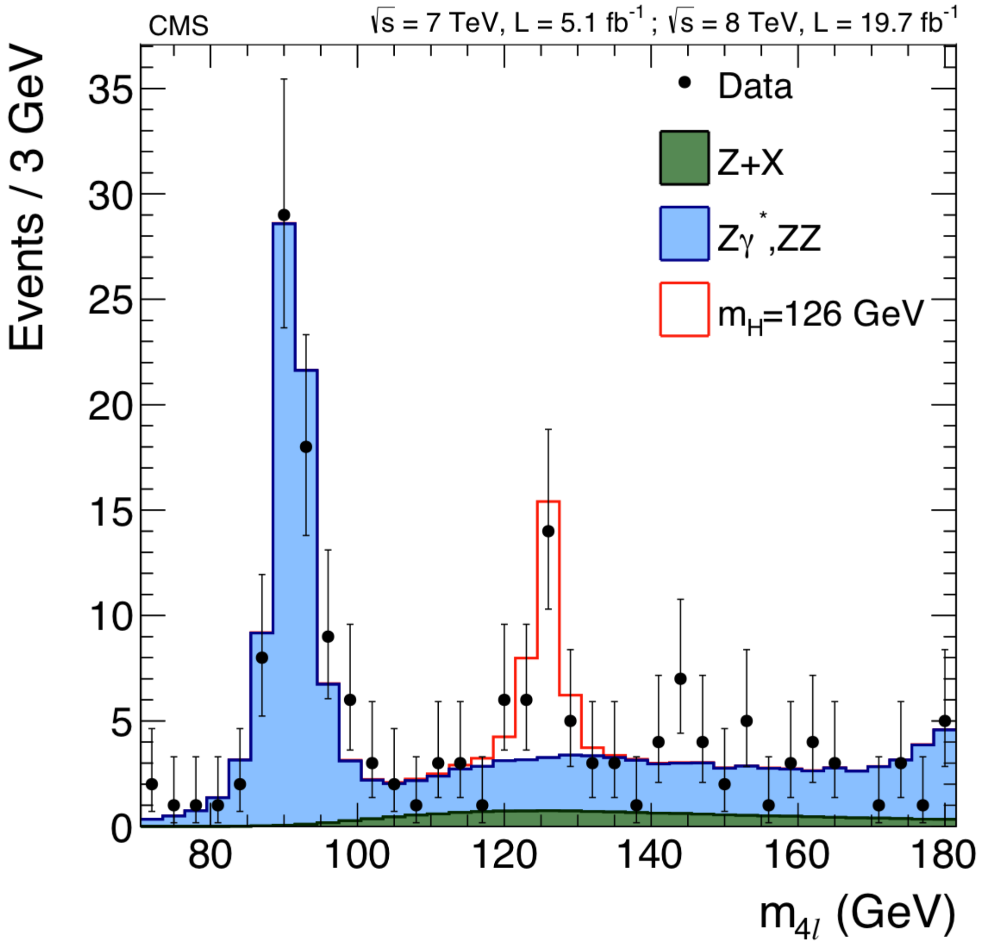


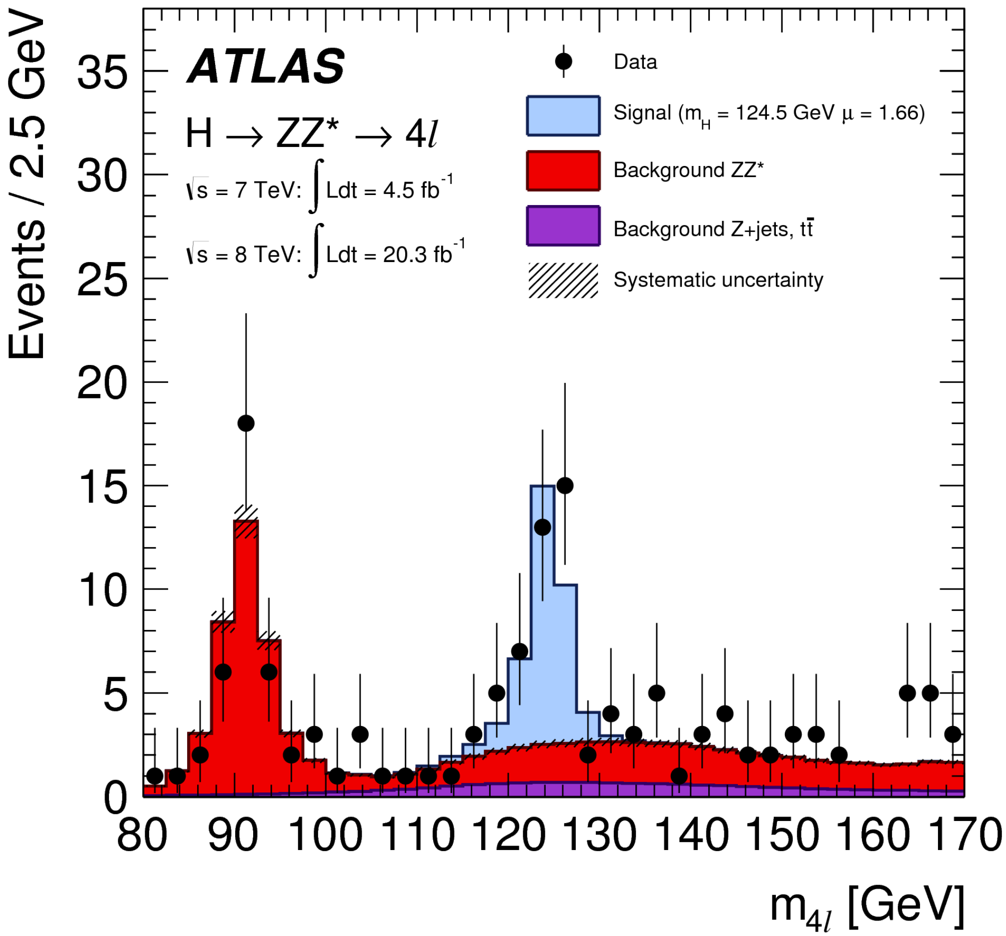
Example event with FSR @ CMS
Mass moves from 116 → 126 GeV

Reducible backgrounds

- The main reducible backgrounds consist of a Z and two (real or fake) leptons
  - In the case of muons, main source Z+ bb/cc or tt where muons come from semi-leptonic decays
    - Also Z+jets with decays in flight
  - In the case of electrons main backgrounds are Z+jets [faking electrons] and Z+jet+gamma where photon converts
- Background rejection relies on object ID and isolation
- Case of low mass resonances [I.e Z+ j/psi] rejected by mass cuts

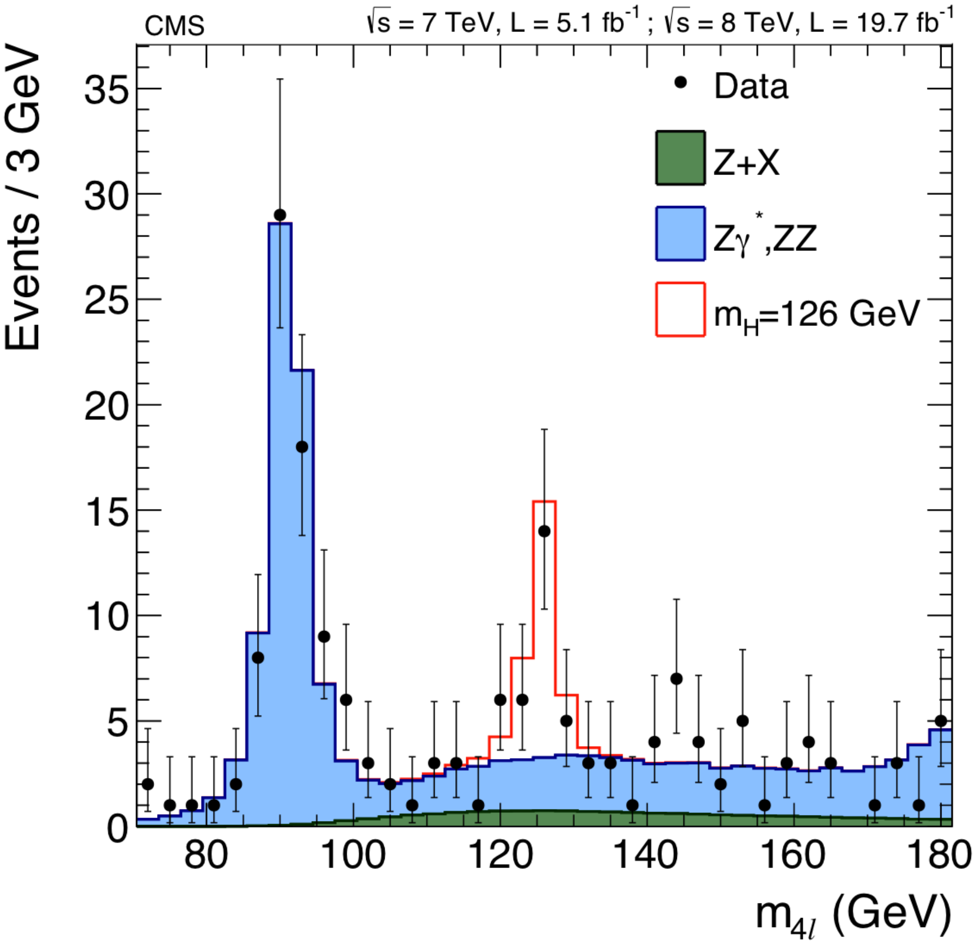
Mass distributions

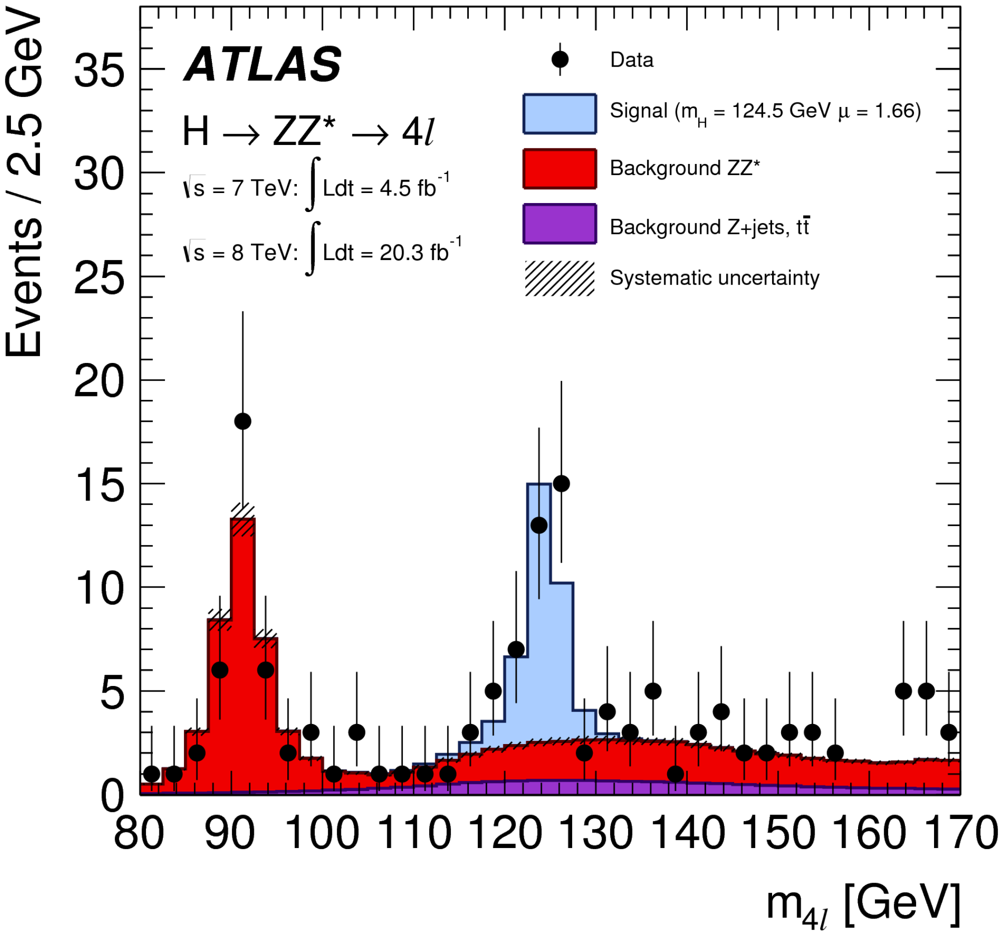




- Both experiments observe Higgs signal in the same place on top of the smooth background

Mass distributions



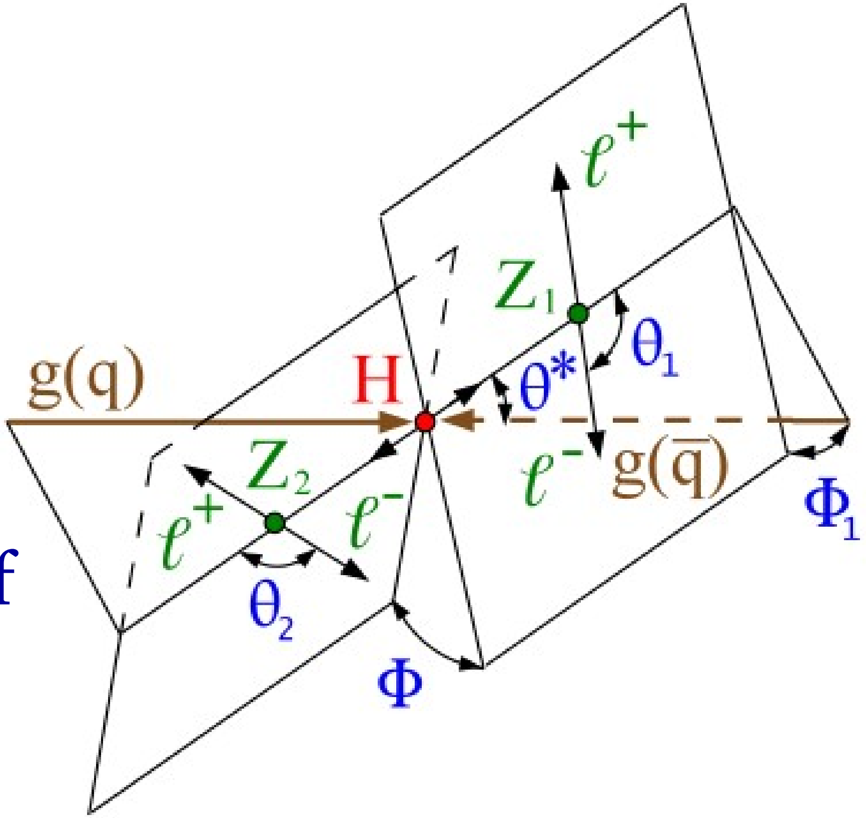


- Both experiments observe Higgs signal in the same place on top of the smooth background

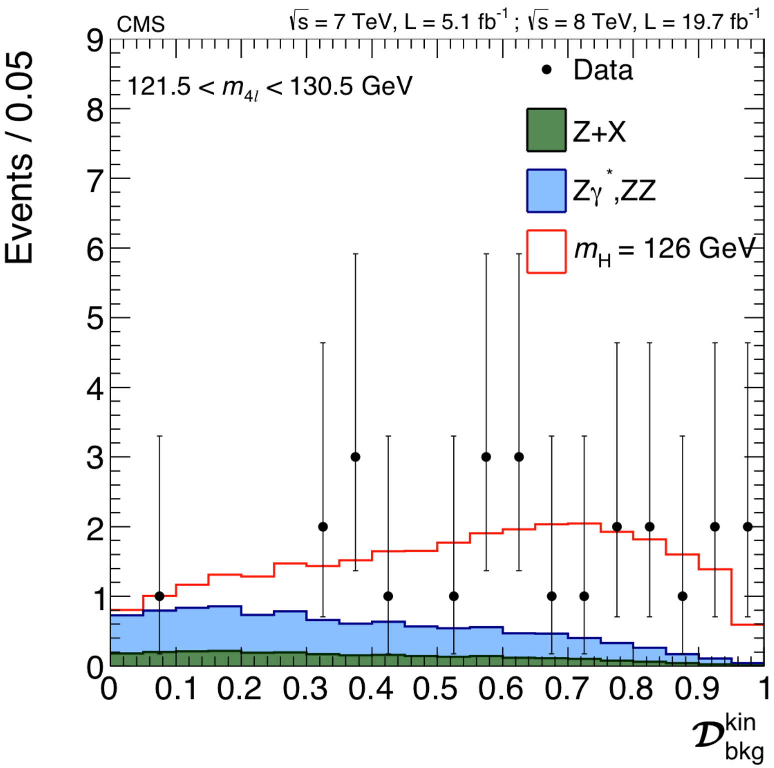
- Z → 4l peak can be used as a standard candle to test the experimental methods

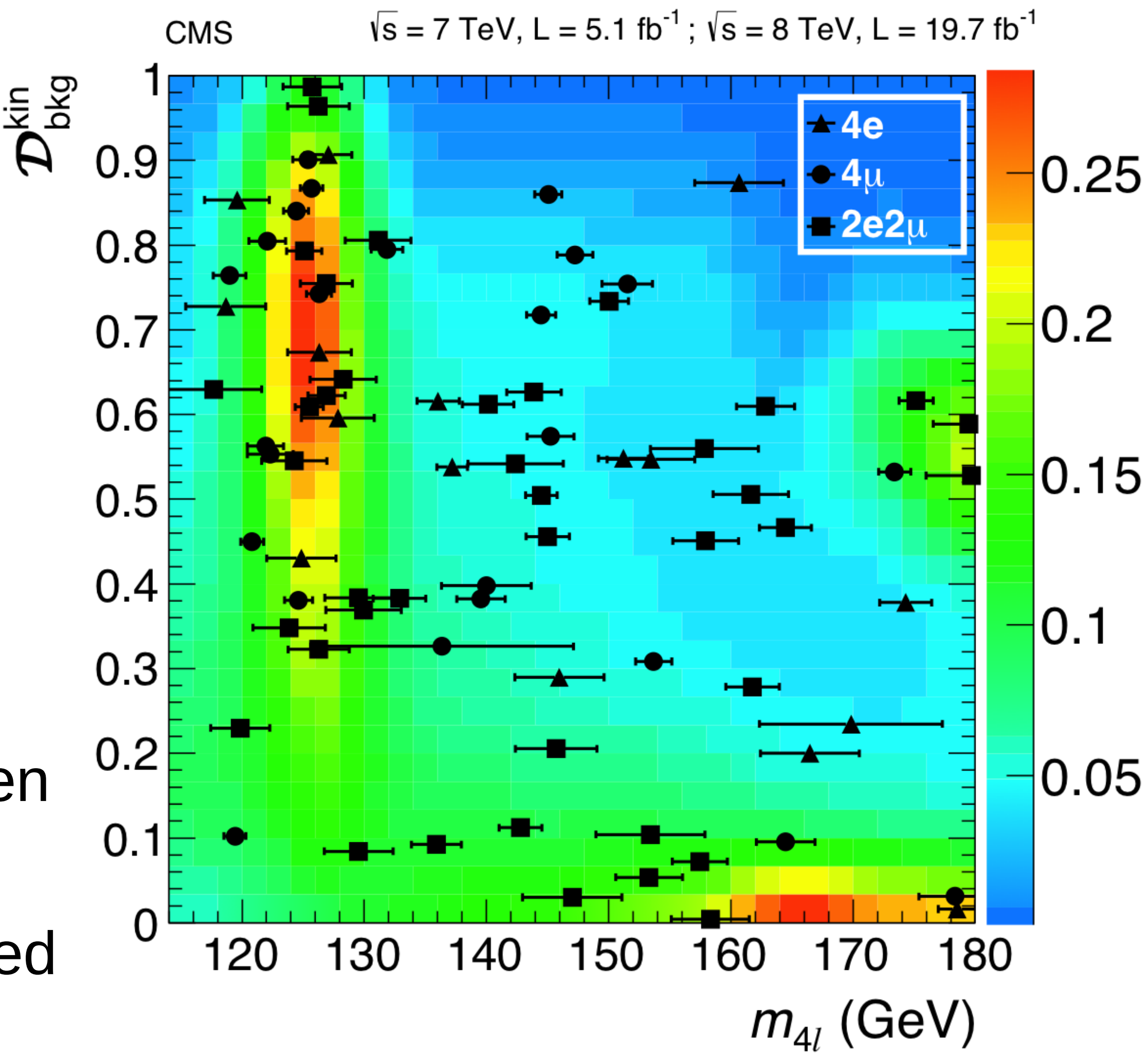
Using the angles

- Fully reconstructed final state
  - Described by 5 angles and the masses of Z and Z*
- Exploiting the zero spin of the Higgs enhances separation vs ZZ background
  - Can be implemented in terms of kinematic discriminant
    - Using generator level matrix element [CMS]
    - Using multivariate techniques [ATLAS]



Kinematic Discriminants





- Good discrimination between ZZ and Higgs
- Data peaking at the expected Higgs region

Extracting the signal

- The amount of signal is extracted by a fit on the data using a model on signal and background
  - Signal model taken from simulation
  - Background model usually taken from data or from simulation

Number of signal events

Signal shape

Number of background events

$$M = N_s P_s(m_{4l}) + \sum_b N_b P_b(m_{4l})$$

Background shape

- Signal shape can be a parametric form or a histogram
- In the case of 4l parametric form preferred
  - Backgrounds can be fitted easily and fit → unbinned

Cross section and efficiency

- If we are measuring cross section, the model can be modified as:

Efficiency

Cross section

$$M = \sigma \mathcal{L} \epsilon P_s(m_{4l}) + \sum_b N_b P_b(m_{4l})$$

Luminosity

- The efficiency can be expressed as MC efficiency and a correction factor to match the simulation efficiency to the data

$$M = \sigma \mathcal{L} \epsilon_{MC} \rho P_s(m_{4l}) + \sum_b N_b P_b(m_{4l})$$

Efficiency in MC

Data MC correction factor

The signal strength

- In the case of Higgs searches the signal strength is used (many prod. Mechanisms , common prescription)

$$\mu = \frac{\sigma \times BR}{\sigma_{SM} \times BR_{SM}}$$

Values in data

Values from theory

- The model is then modified to be

$$M = \mu N_{exp} P_s(m_{4l}) + \sum_b N_b P_b(m_{4l})$$

- The number of expected events is

$$N_{exp} = N_{MC}\rho$$

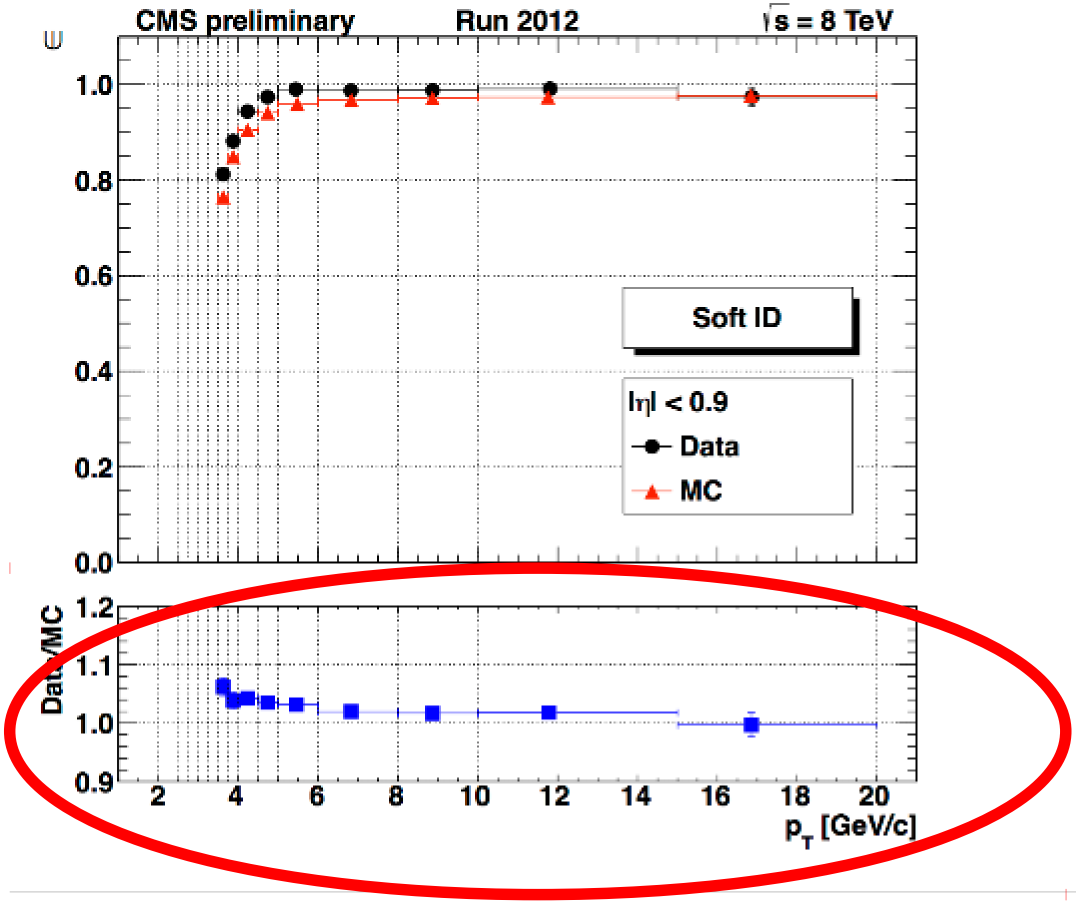
Events expected in simulation
At the current luminosity

Correction factor (Data/Sim)

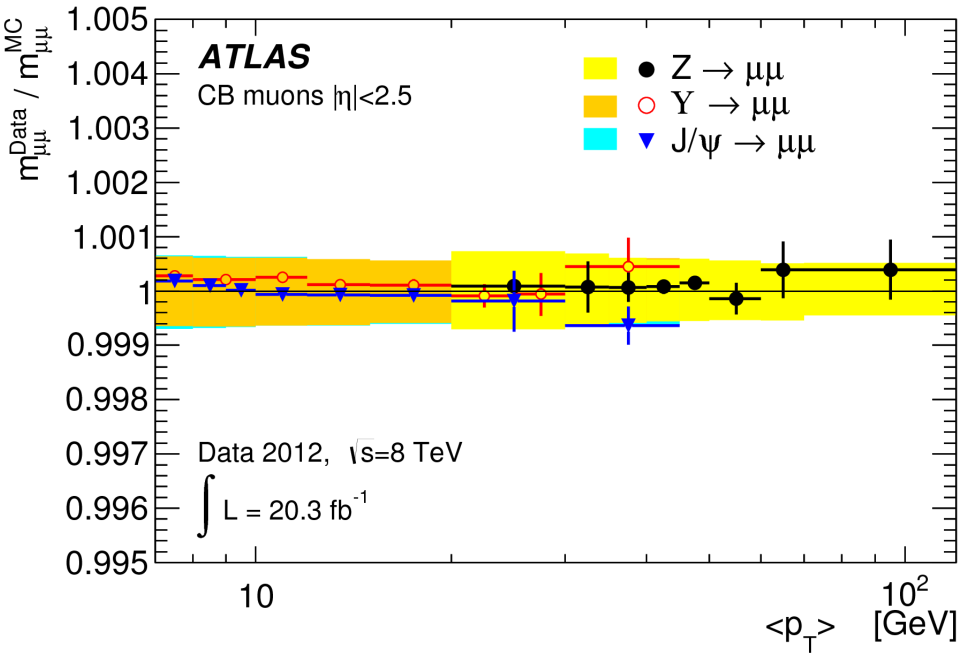
Data vs MC corrections

- MC is corrected to match the data using known processes (I.e Z → mumu/ee, J/psi → mumu/ee)

- Example: muon efficiency using Z



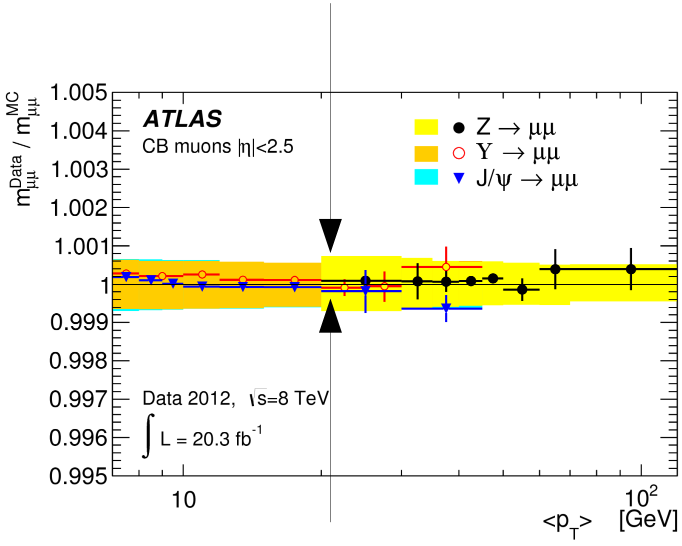
- Example: muon momentum scale using di-muon resonances



Systematic uncertainties

Two types regarding implementation

- Affecting normalization of signal or background
  - Efficiency correction factor uncertainties
    - Example: the statistical uncertainty of the efficiency measurements in data is a systematic uncertainty for the Higgs search
- Affecting shape of distribution(i.e moving signal peak, or making it broader)
  - Lepton momentum scale and resolution
    - The statistical and systematic errors from the calibration of the momentum scale and resolution are systematic unccertainties in the Higgs search



Based on this plot, the systematic effect of each lepton In the Higgs mass is an effect of ~ 0.5 per mil

Systematics in the fit

- Systematics are included in the fit as constrained nuisance parameters

Normalization syst.

shape syst.

Constrain term

$$M = \left[\mu N_{exp}\rho(\theta_i)P_s(m_{4l},\theta_j) + \sum_b N_b P_b(m_{4l},\theta_j)\right] \times \prod_{i,j} G(\theta,\bar{\theta},\sigma_\theta)$$

Mean measured value of the systematic effect

Systematic uncertaInty

- With this treatment, uncertainties are floating in the fit and the total error is automatically calculated by the fit its self
- G is usually a Gaussian or Lognormal function

“Expected” vs “Observed”

- Analyses selections need to be optimized for maximum performance
- A figure of merit(FOM) is defined based on the scope
  - In case of discovery → **expected** significance
  - In case of exclusion → **expected** limit
  - In case of precision measurements → **expected** measurement uncertainty
- What is “expected”?
  - Expected FOM is defined as the value of the FOM assuming that the signal is present on top of the background as described by theory and after adding all our knowledge of detector performance
    - Defined without the real data
- And what is observed?
  - Observed is the value of the FOM given the real data

How do we optimize an analysis?

- We pick the figure of merit
- We calculate the expected figure of merit for different analysis strategies by doing the **full** signal extraction **including the systematic uncertainties**
- We choose the strategy that optimizes the expected figure of merit
- Things to be careful about:
  - Over-optimizing → Over tuning our selection to the theory model → Not good if we dont know what the signal is!
    - Example: we should not exploit the angular variables if we dont know the CP properties
      - OR... we do the analyses for different CP assumptions ...

Signal significance

- We define a test statistic

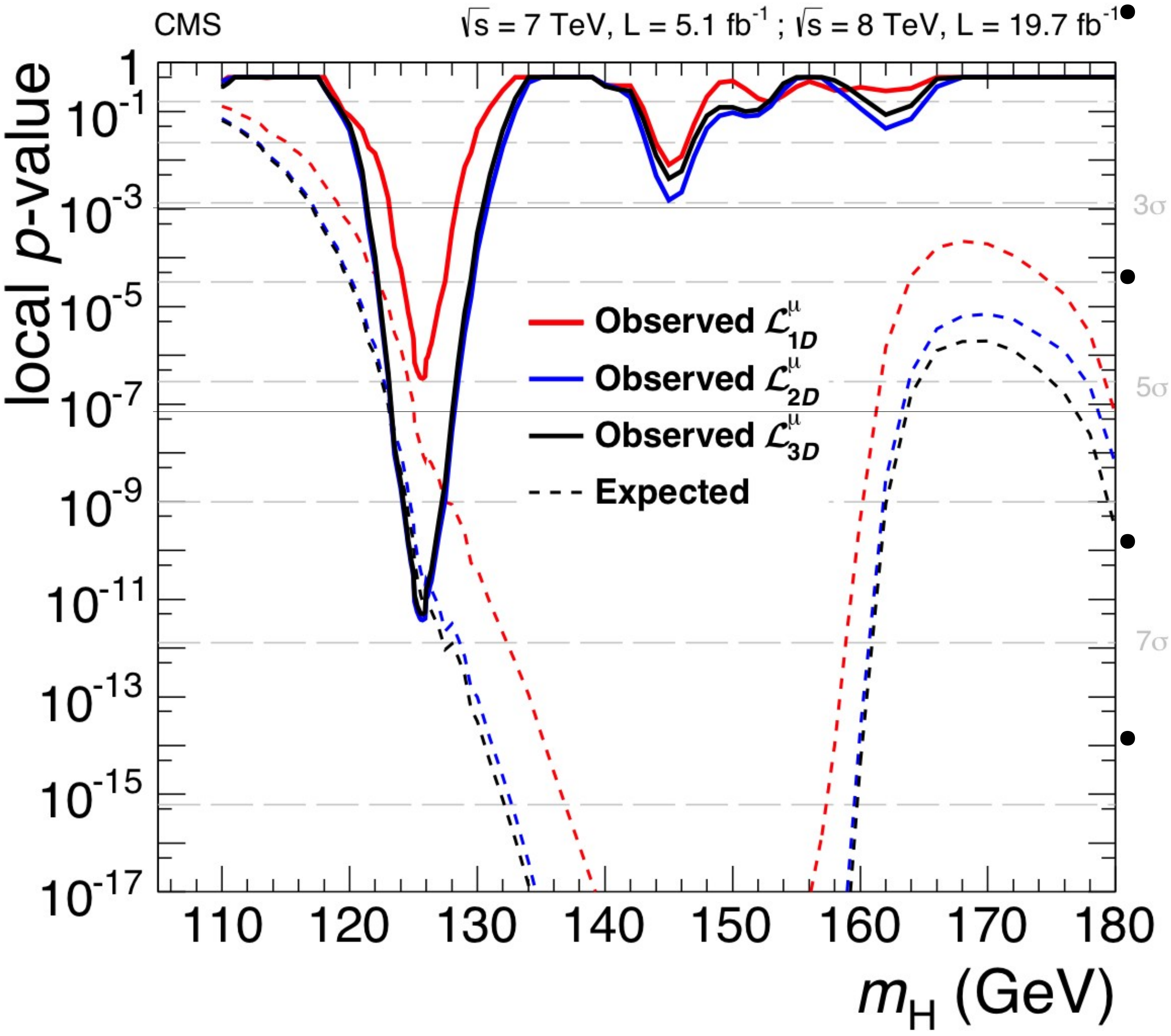
Maximum value of likelihood with signal fixed to zero in the data

$$q = -2ln\frac{L_b}{L_{s+b}}$$

Maximum value of likelihood for any signal

- Then we define as p-value the probability that q is larger than the one observed in the data
- Practically: the probability that background describes the data
- Can be expressed in terms of standard deviations [significance]
  - For a gaussian distributed experiment S /SQRT(S+B)

Significance in H → ZZ* → 4l



- CMS
  - Expecting 6.7 sigma, observed 6.8 sigma
- ATLAS
  - Expecting 4.4 sigma, observed 6.6 sigma
- 3 sigma considered an evidence,
  - 5 sigma an observation
- Both experiments observe the new signal in the ZZ final state stadalone

Signal strength results

- ATLAS: $1.66^{+0.45}_{-0.38}$
- CMS: $0.93^{+0.26}_{-0.23}$
- ATLAS sees more signal than expected [at 2 sigma] CMS consistent [within 1 sigma]
- Possibilities
  - Higgs signal strength is higher than theory and CMS under-fluctuated
  - Higgs signal strength is as theory and ATLAS over-fluctuated
  - The measurements are consistent with them and with the SM ?
- How do we answer this question?

Compatibility

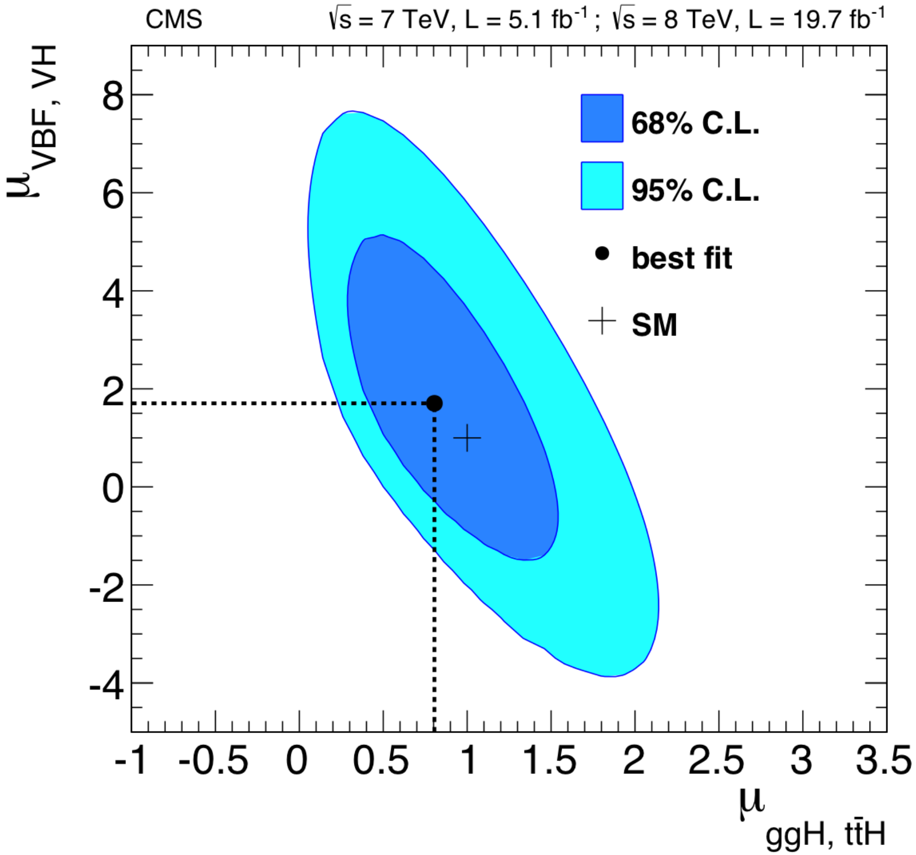
- Neglecting correlations
  - There are some since the same theoretical model is used!

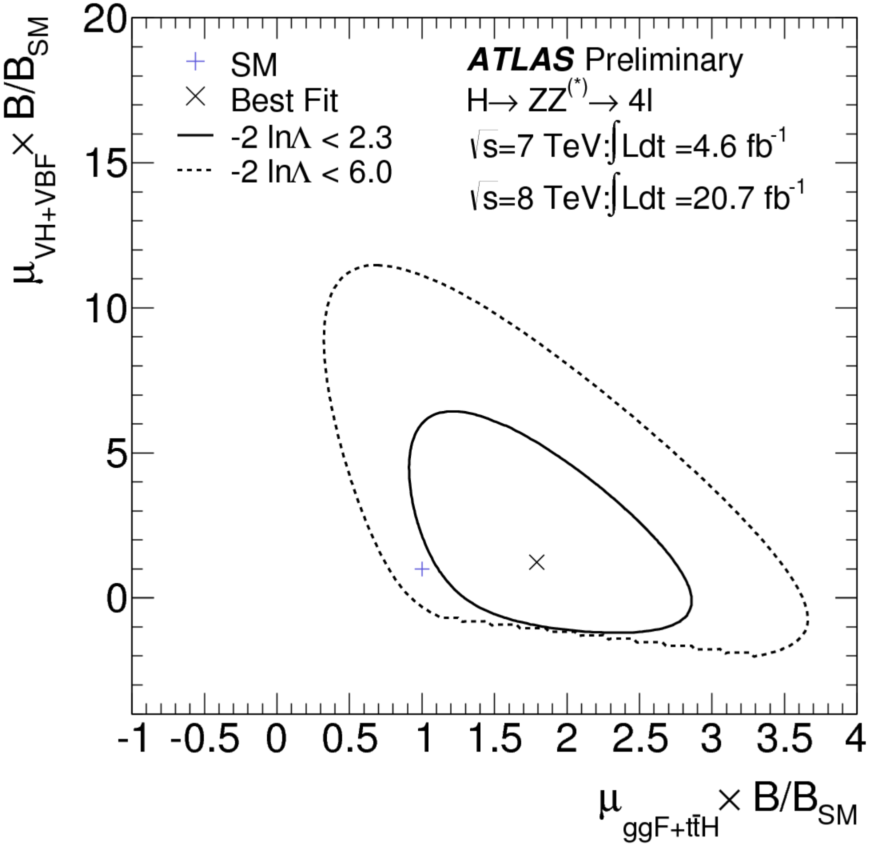
$$\mu_{ATLAS} - \mu_{CMS} = 0.73 \pm 0.48$$

- Compatibility at 1.5 sigma level
  - 7% probability → not significant concerning the number of measurements we do!
- Lets think differently and combine them!
  - Combined signal strength: 1.09 +- 0.21
  - Compatible with SM Higgs
  - Again neglecting correlations + assuming measurements are at the same mass but result at the correct ballpark

Signal strength per production mode

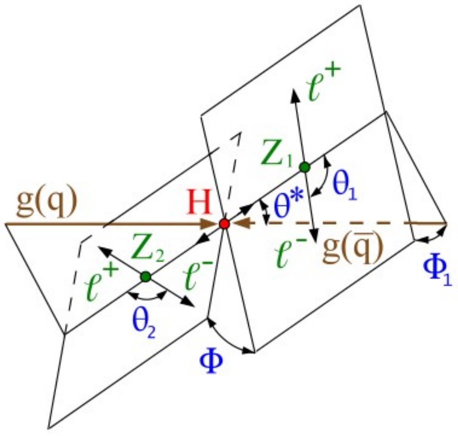
- Signal strength can be split into fermion induced mechanisms [ggH,ttH] and vector boson induced mechanisms (VBF/VH)

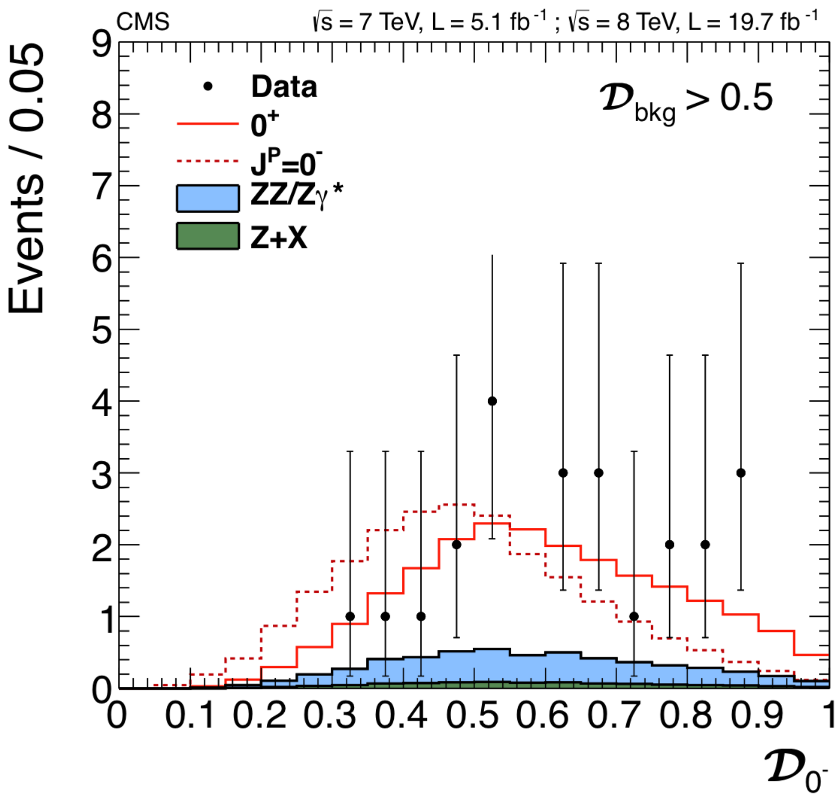




Spin/Parity studies

- Previously we used the angular properties to distinguish between signal and background
  - We can use it to also discriminate against different hypotheses
  - Building discriminants for each hypothesis separation
- Example CP even VS CP odd Higgs



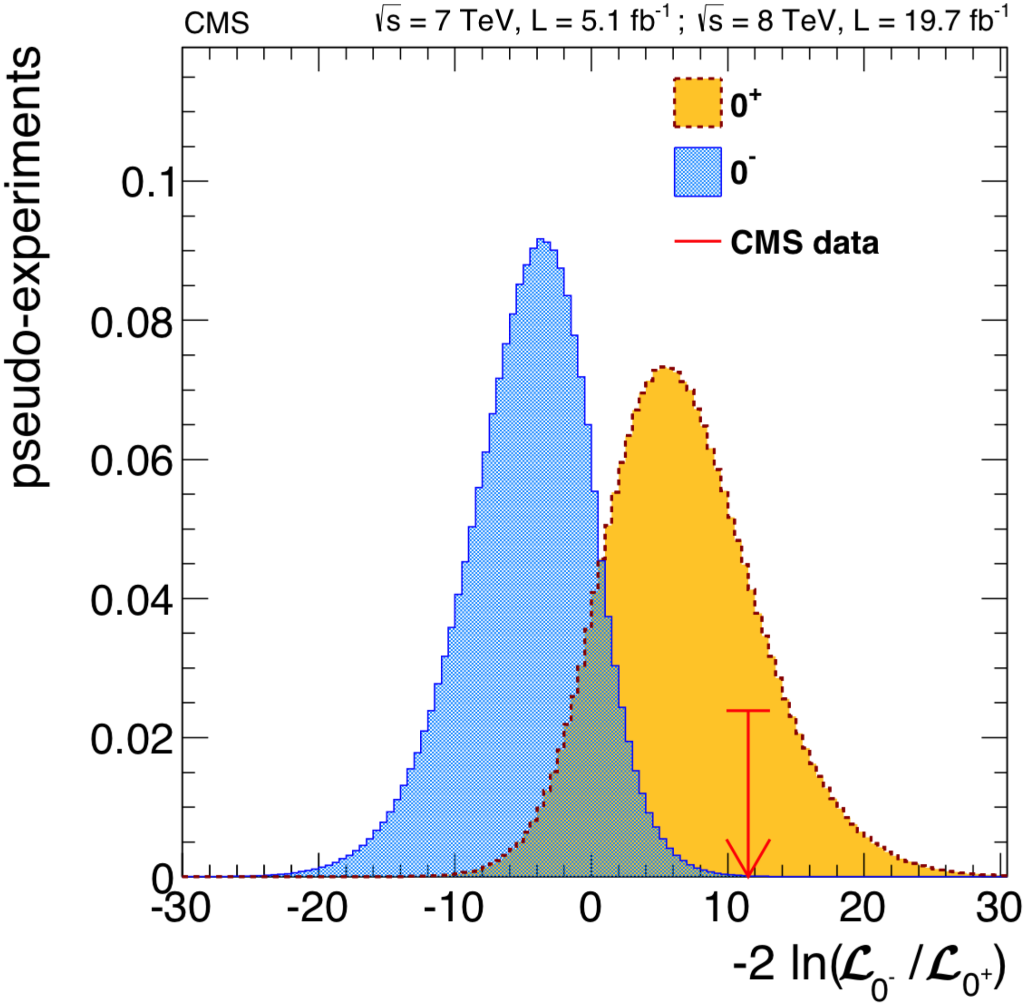


Hypothesis tests

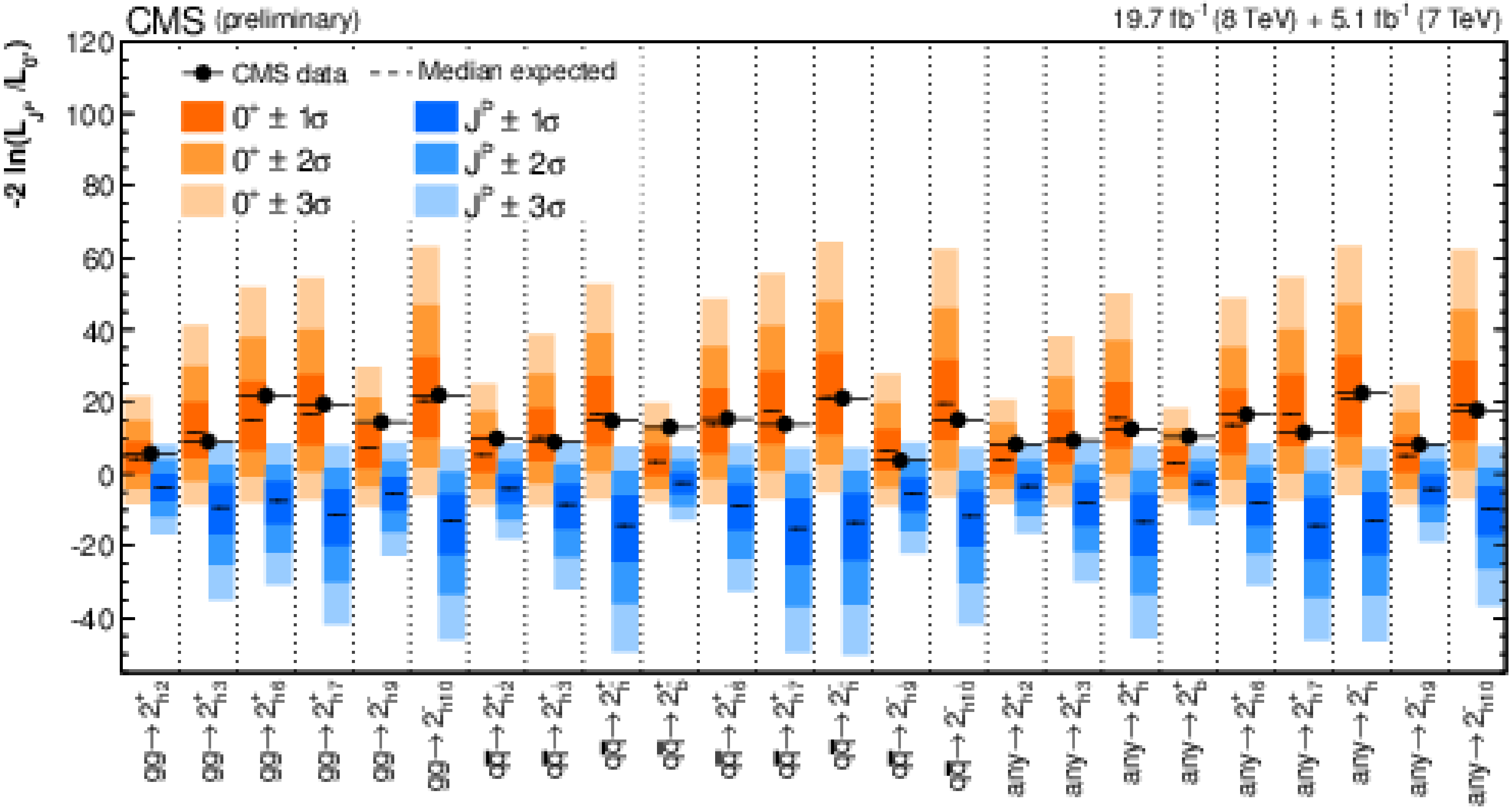
- We build a signal + background likelihood for each hypothesis
- We define a test statistic:

$$q = \frac{\mathcal{L}_{0^-}}{\mathcal{L}_{0^+}}$$

- We do toy experiments for each hypothesis and make distributions
- We calculate q in data and see how far it is from each hypothesis



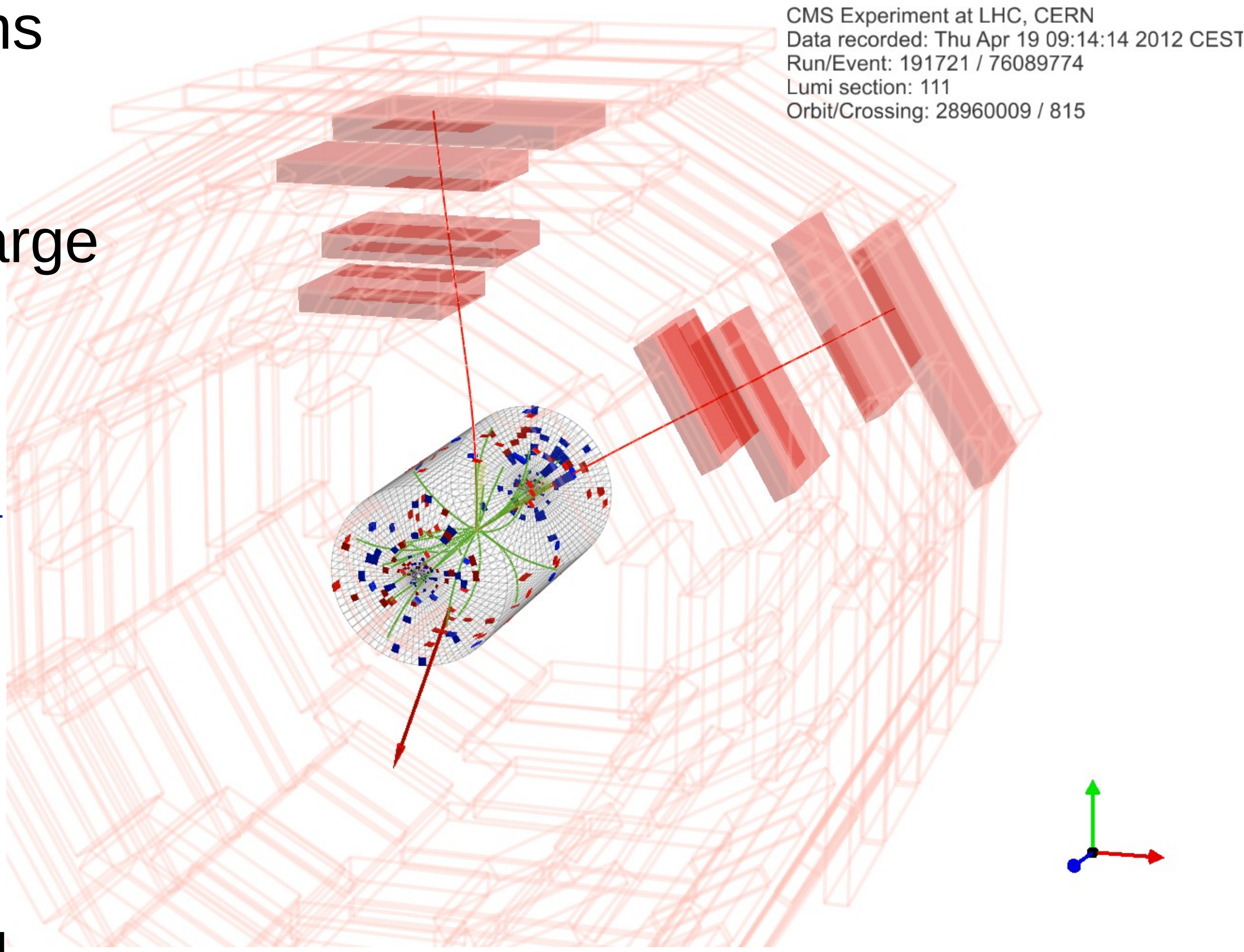
Spin – Parity Results



- Recent CMS measurements show compatibility of spin and parity with the SM
  - Similar results from ATLAS

$H \to WW^*$

- Two high quality OS leptons
- 2 neutrinos → MET
- Low mass resolution but large yield!
    - Transverse mass of the di-lepton +MET system used
- Main backgrounds
    - WW,top, DY, W+jets
- Production mechanisms studied: ggH, VBF, WH,ZH

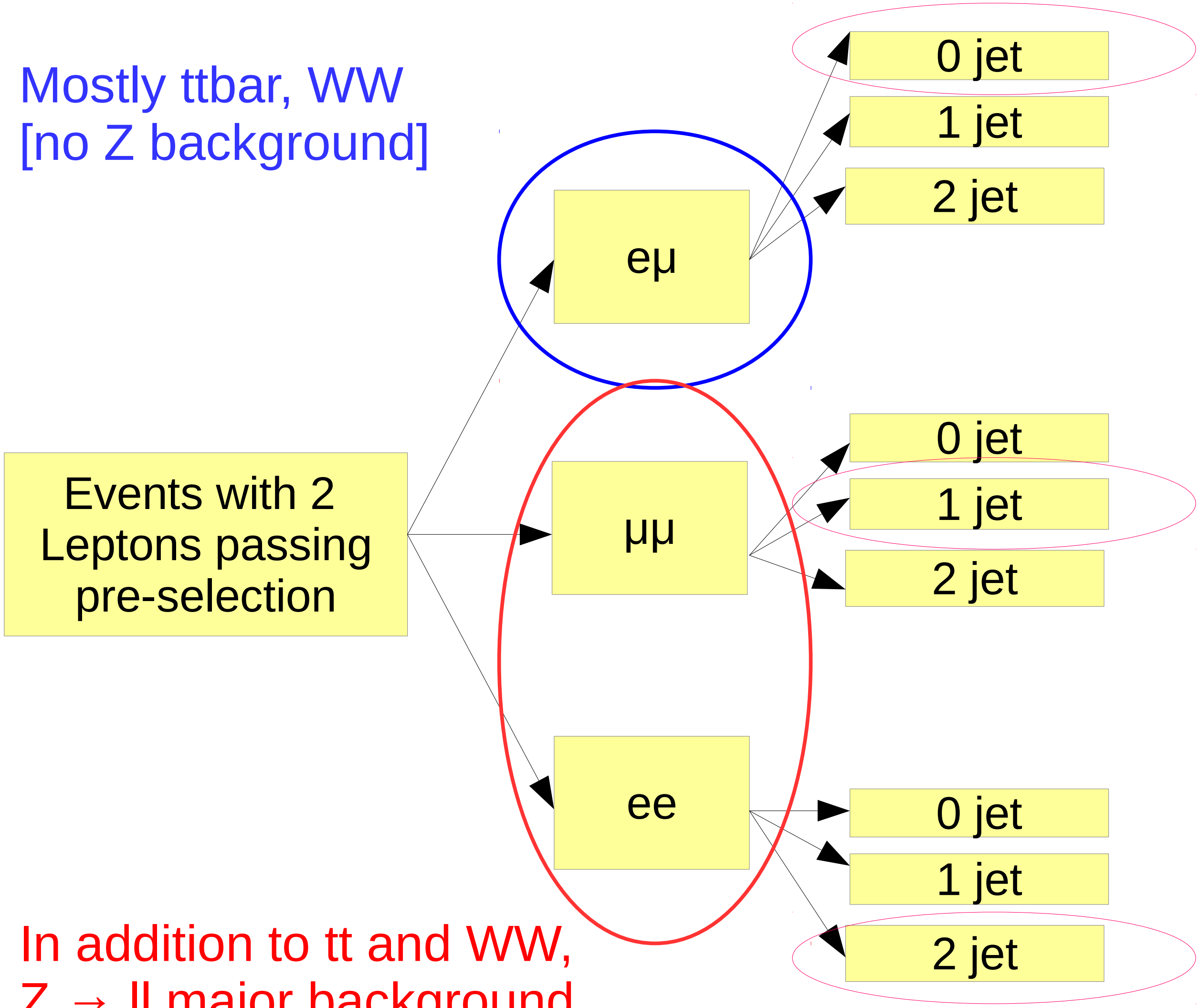


$$M_T^2 = (E_T^{\ell\ell} + E_T^{\text{miss}})^2 - (\vec{p}_T^{\,\ell\ell} + \vec{E}_T^{\text{miss}})^2$$

Event categorization technique

- In complicated final states with many backgrounds events are split into categories
- Each category has specific sensitivity and specific background composition and is analyzed separately
- The categories then are combined to estimate the total result
- Using event categorization allows to boost sensitivity without throwing away events
  - It is equivalent to adding a new dimension in the fit

Example: Categorization in H → WW*



Mostly ttbar, WW
[no Z background]

0 jet requirement supresses ttbar

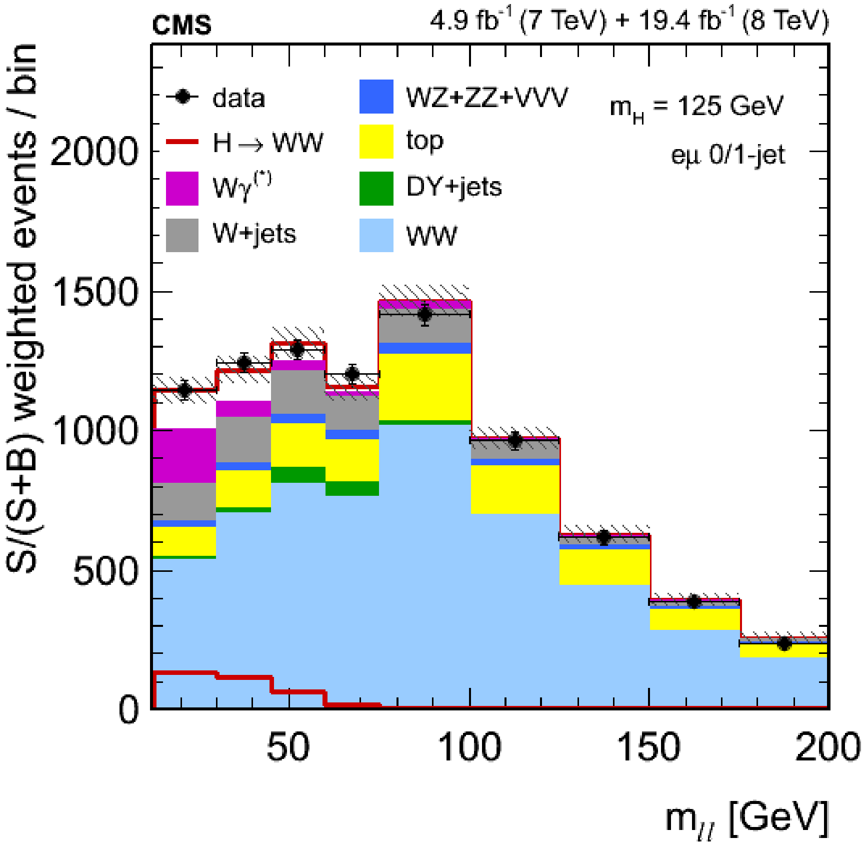
1 jet requirement. Some ttbar Present. Lower sensitivity than 0 jets

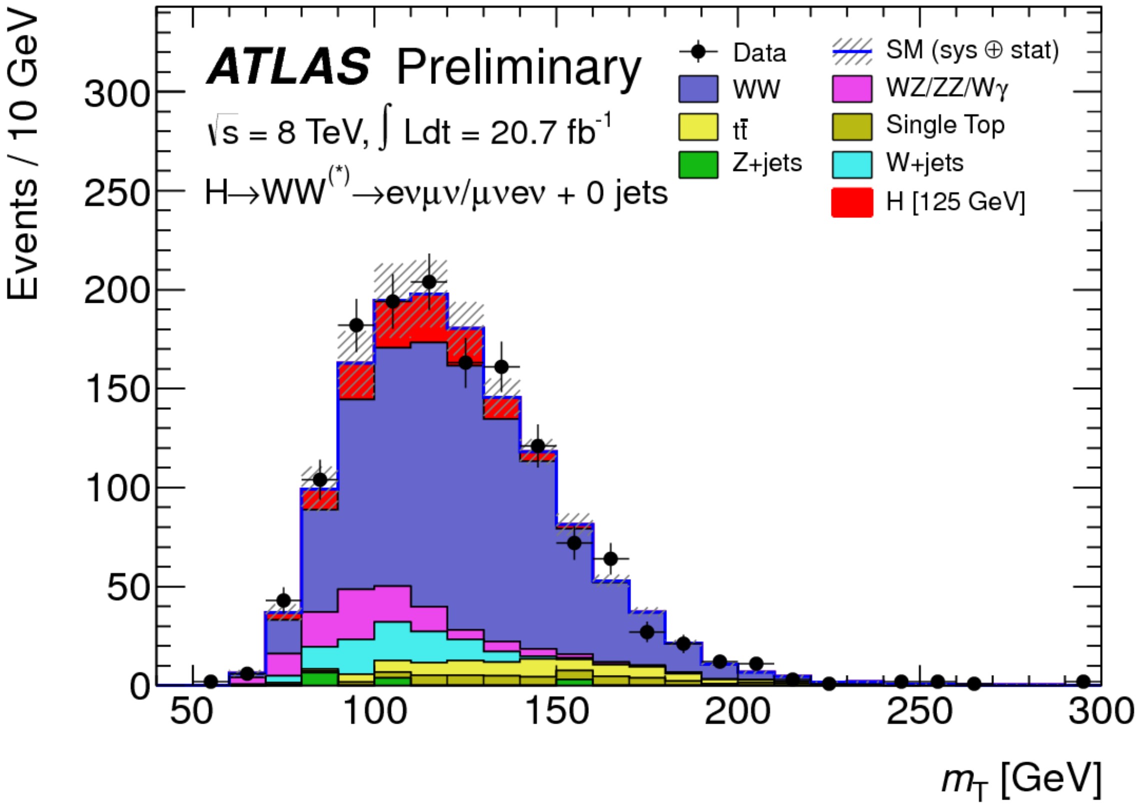
In addition to tt and WW,
Z → ll major background

2 jets has a lot of ttbar background Target VBF and apply tight VBF requirements

Here sample is split into 9 categories
0 jet e/mu and 2 jet VBF seem to be most sensitive

H → WW* distributions





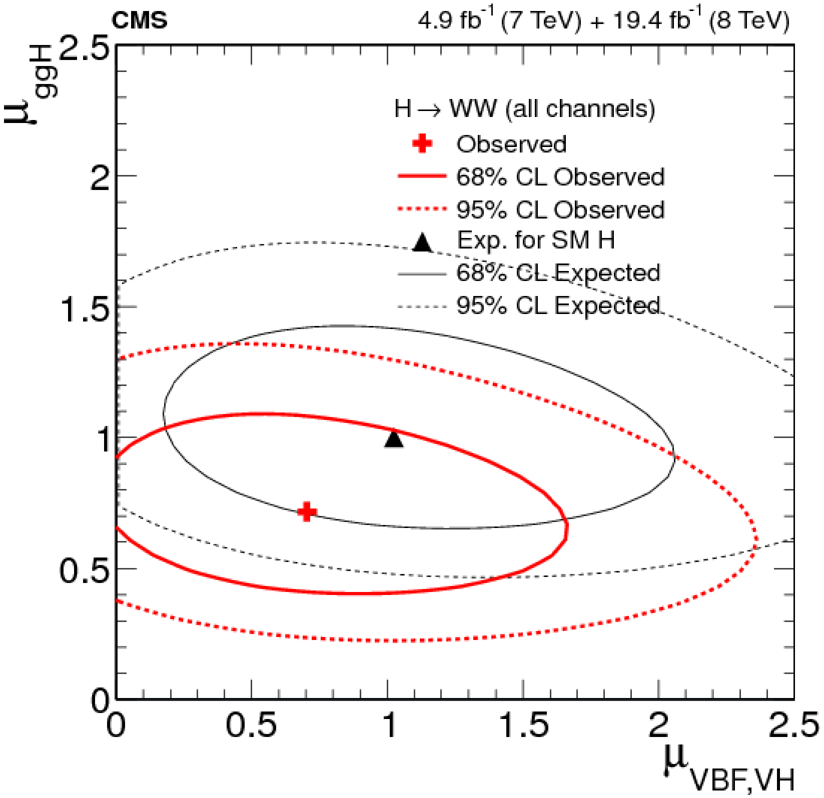
- Signal visible on top of large broad background
  - Significance of 4.3σ in CMS (exp. 5.8) and 3.8 σ in ATLAS (exp. 3.7)
- Signal strength

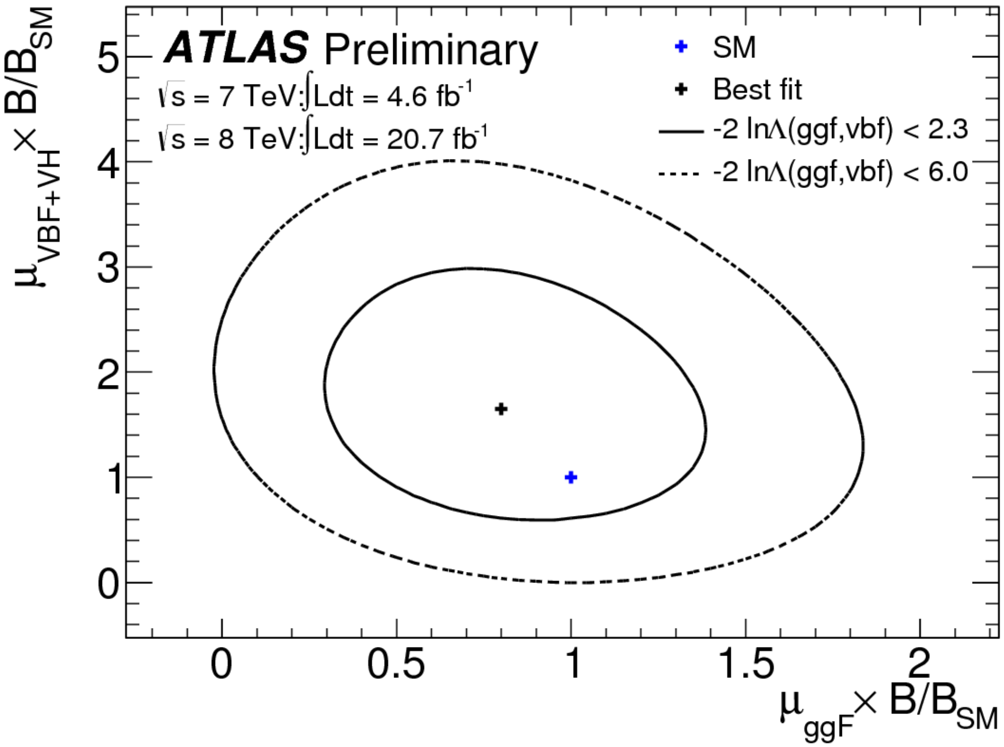
$$\mu_{ATLAS} = 0.99^{+0.30}_{-0.30}$$

$$\mu_{CMS} = 0.72^{+0.20}_{-0.18}$$

→ Exercise: How inconsistent they are?

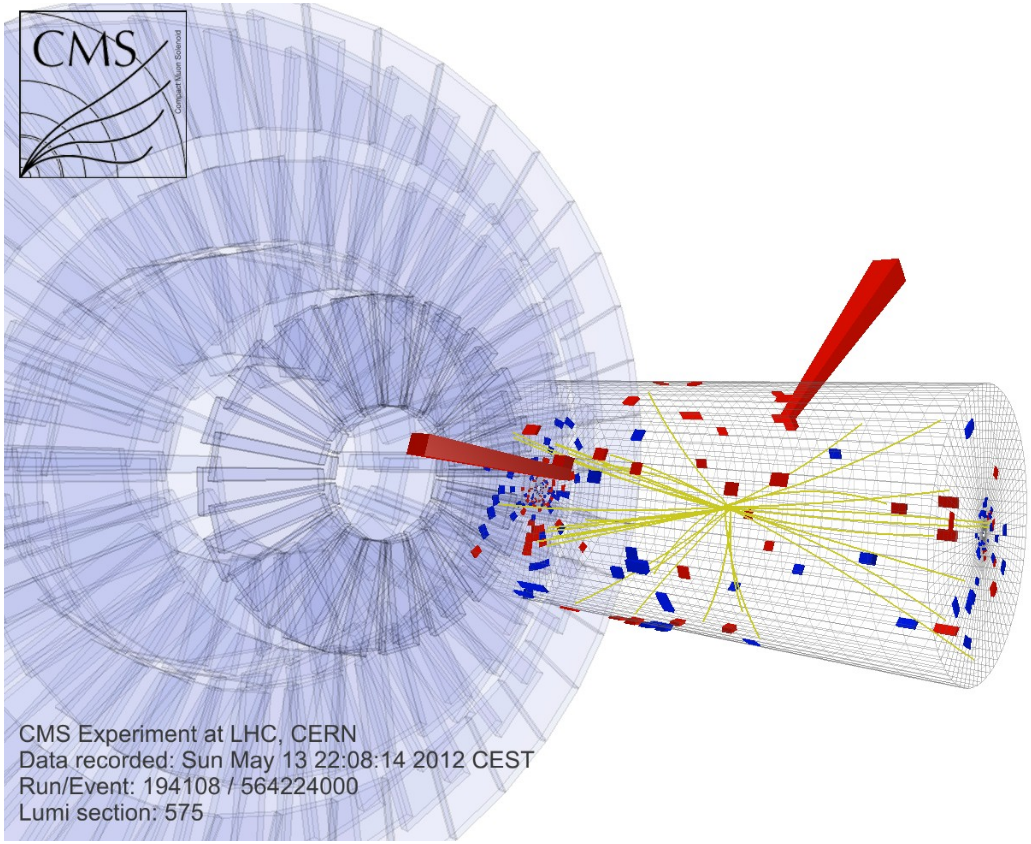
Production mechanisms

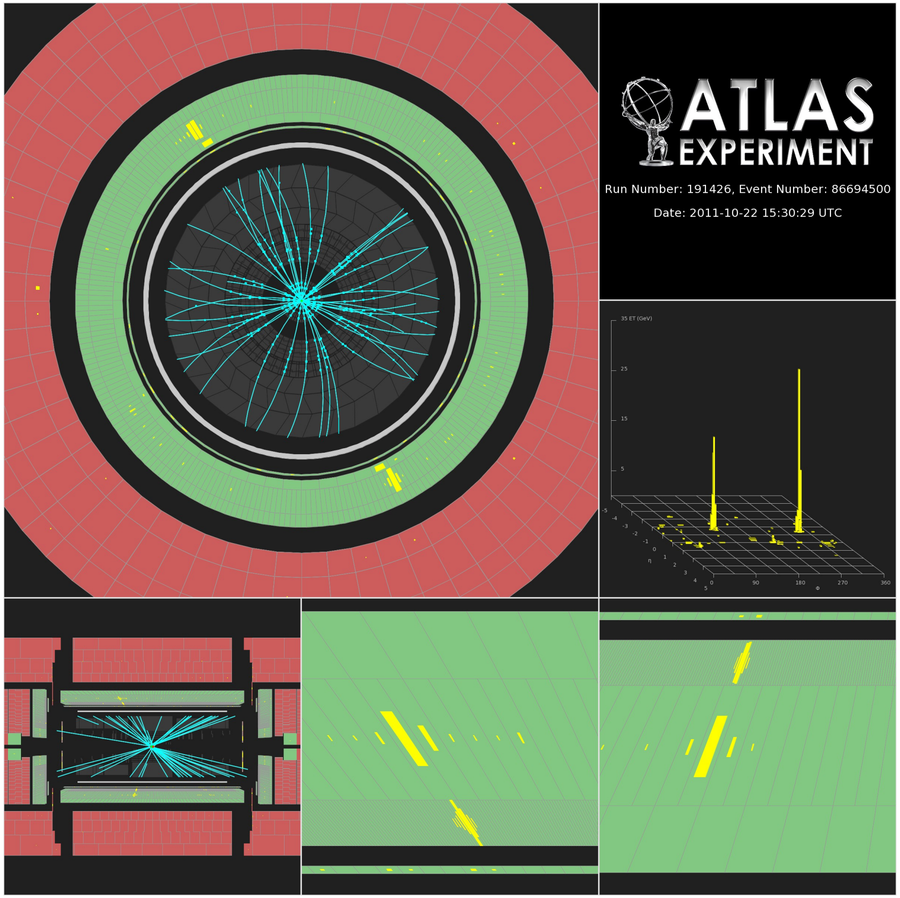




- Both experiments compatible within 1 sigma

H → γγ





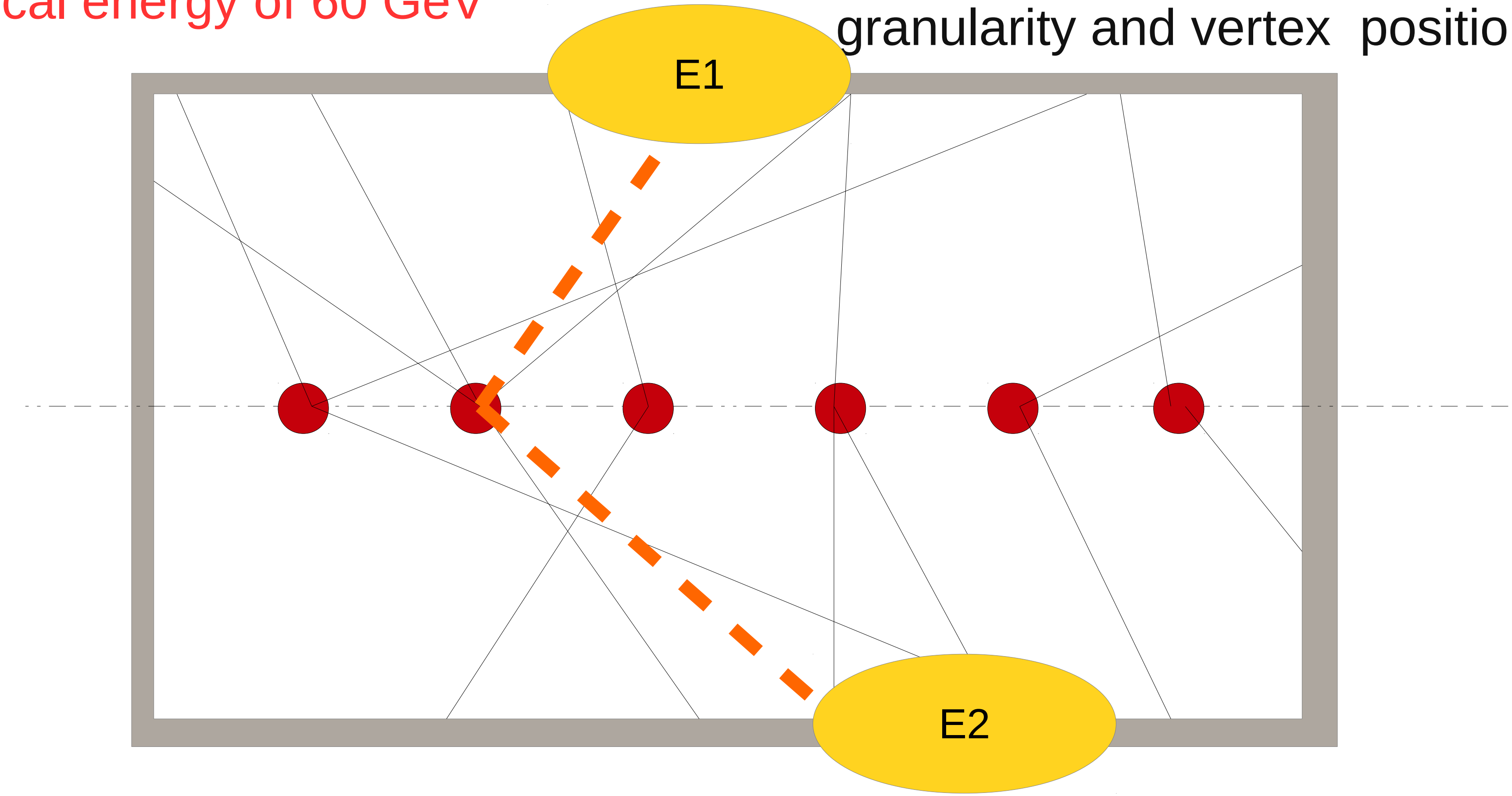
- Reconstruct a pair of photons
  - Selection tight to reduce fake photon background
- Dominant background → SM di-photon production
  - Detector resolution crucial to identify the signal

Reconstructing di-photon mass

$$M_{\gamma\gamma} = \sqrt{2E_1E_2(1 - cos\theta_{12})}$$

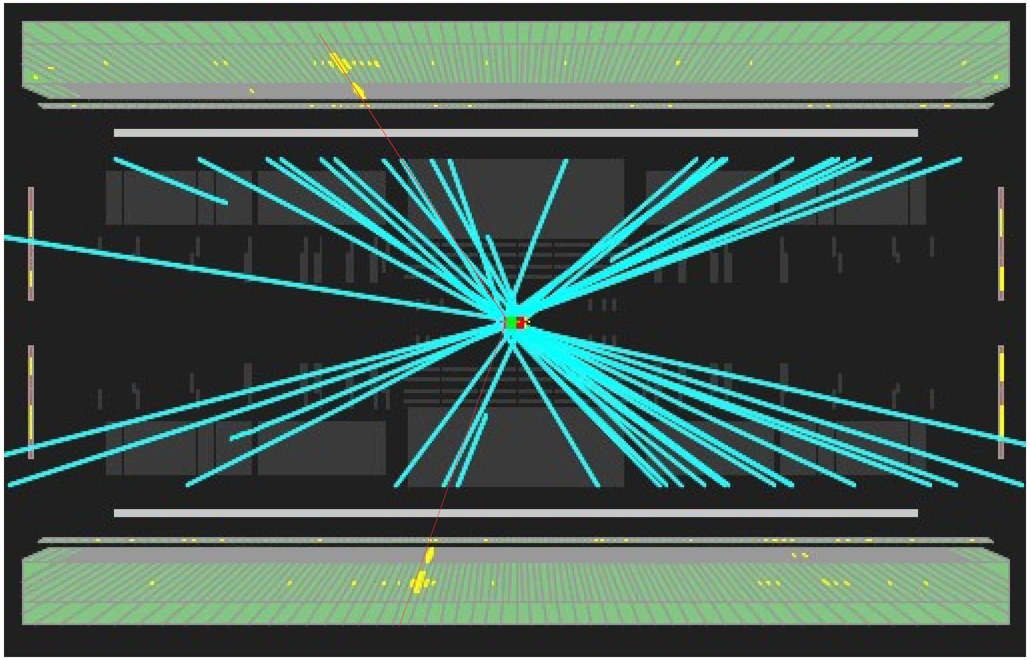
Energies measured by calorimeters
Typical energy of 60 GeV

Angle resolution by calorimeter
granularity and vertex position

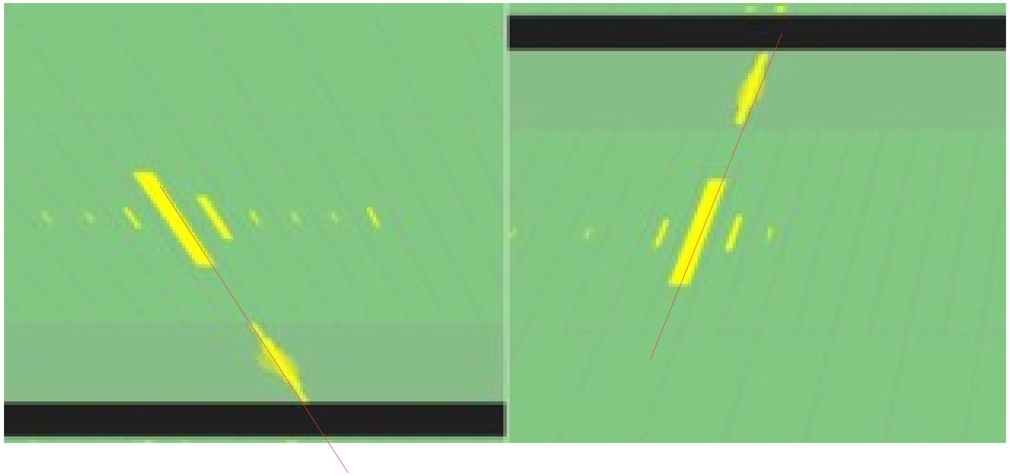


Contributions to the resolution

- Calorimeter resolution defines the precision for E1 and E2
  - Typical values(design)
    - 0.3 GeV for CMS
    - 0.7 GeV for ATLAS

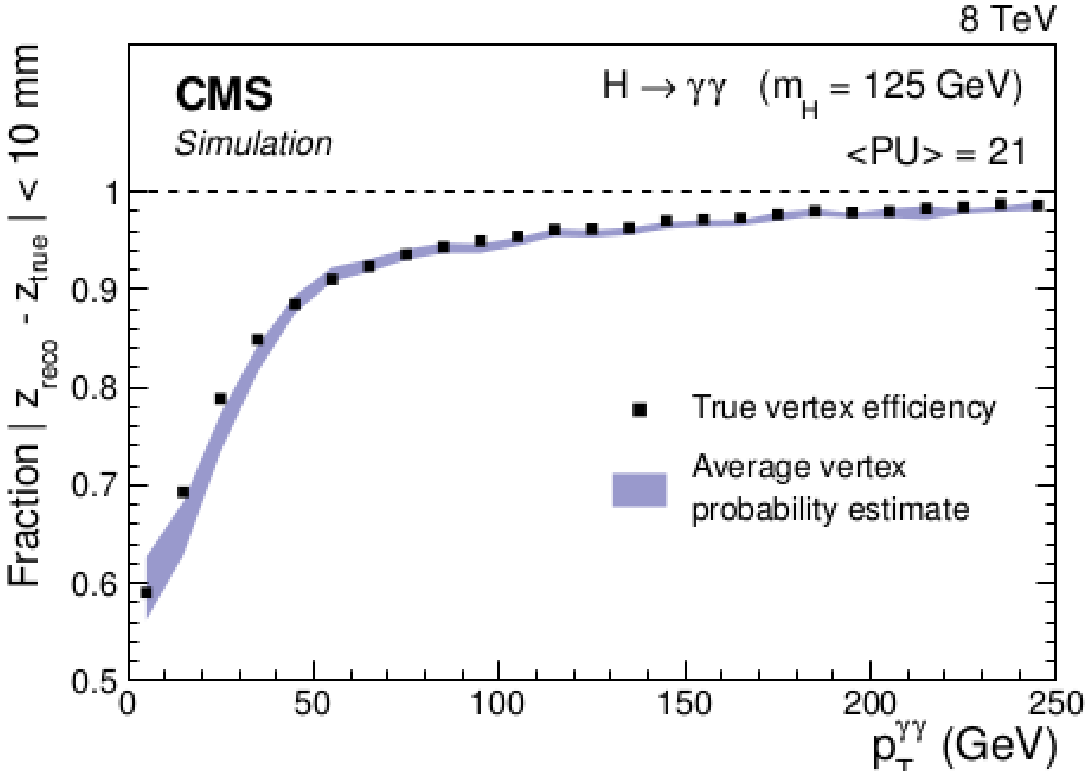


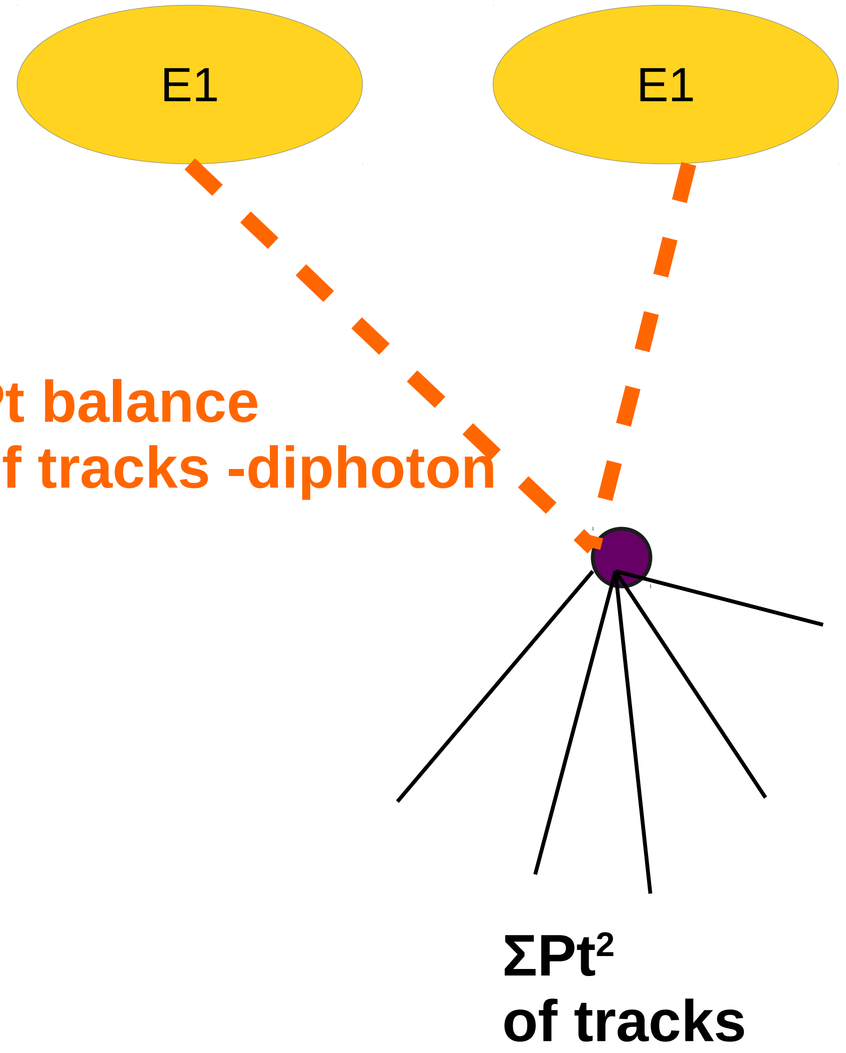
- For the angular distribution one needs to pick the right vertex!
  - ATLAS longitudinal segmentation allows to extrapolate back and find the good one
  - CMS needs to use more information to find it



Di-photon vertex finding

- In a nominal event primary vertex has largest ΣPt^2 of associated tracks
  - In di-photon event use the balance of the diphoton
  - + pointing information in case of ATLAS

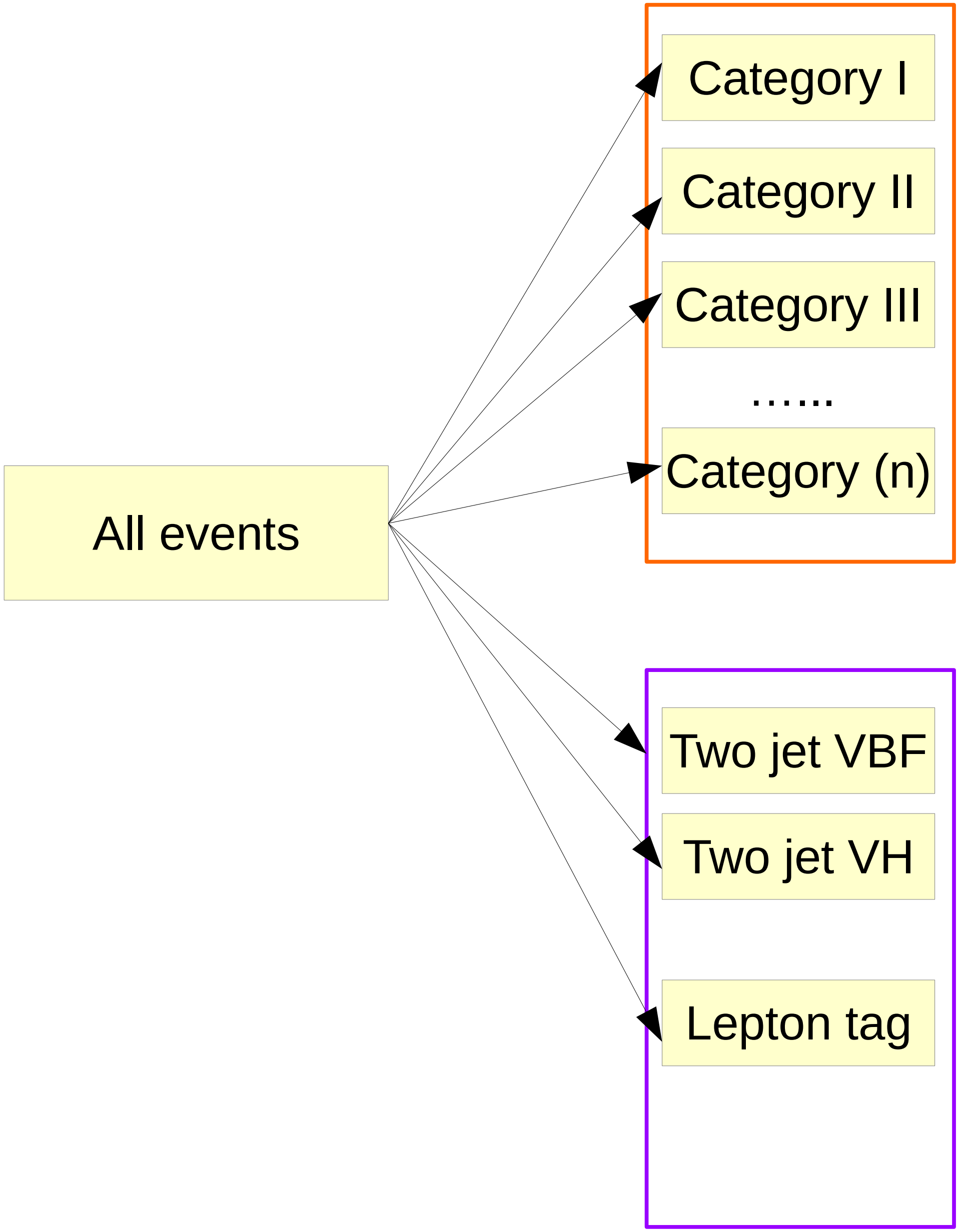




Pt balance of tracks -diphoton

ΣPt^2 of tracks

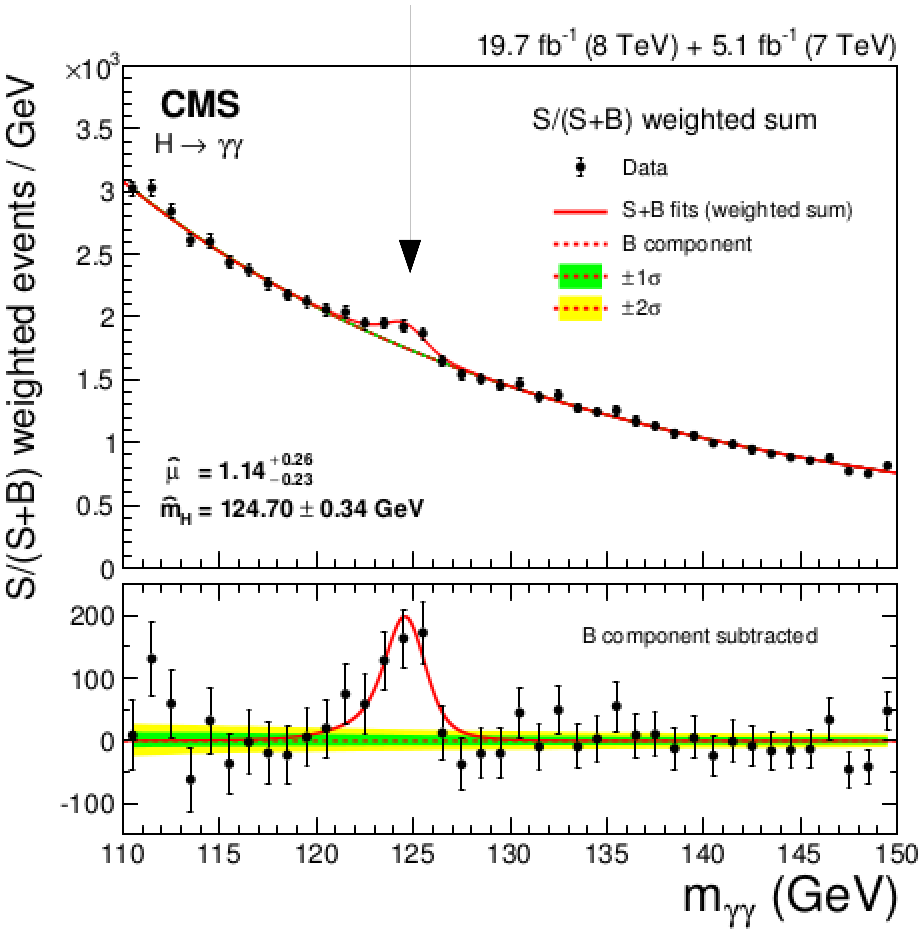
Event categorization (again!)

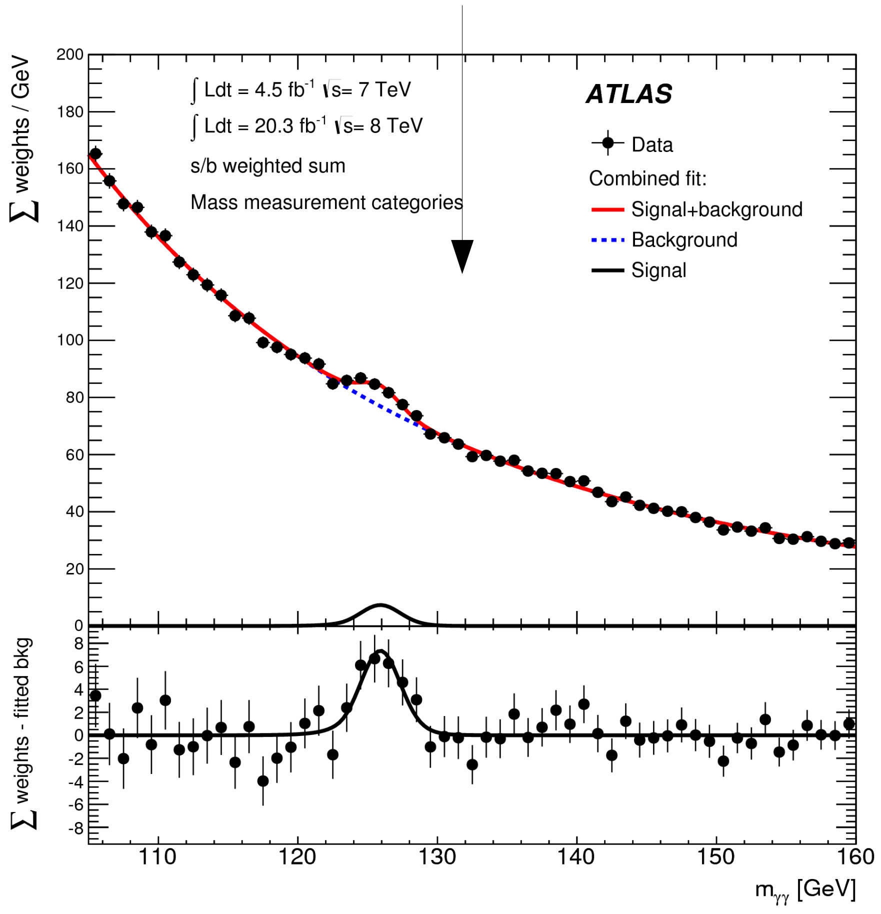


Categorization based on the expected resolution . Different Categories for golden photons, converted photons etc

Exclusive categories for VBF and VH /ttH tagging (similar strategy as in WW)

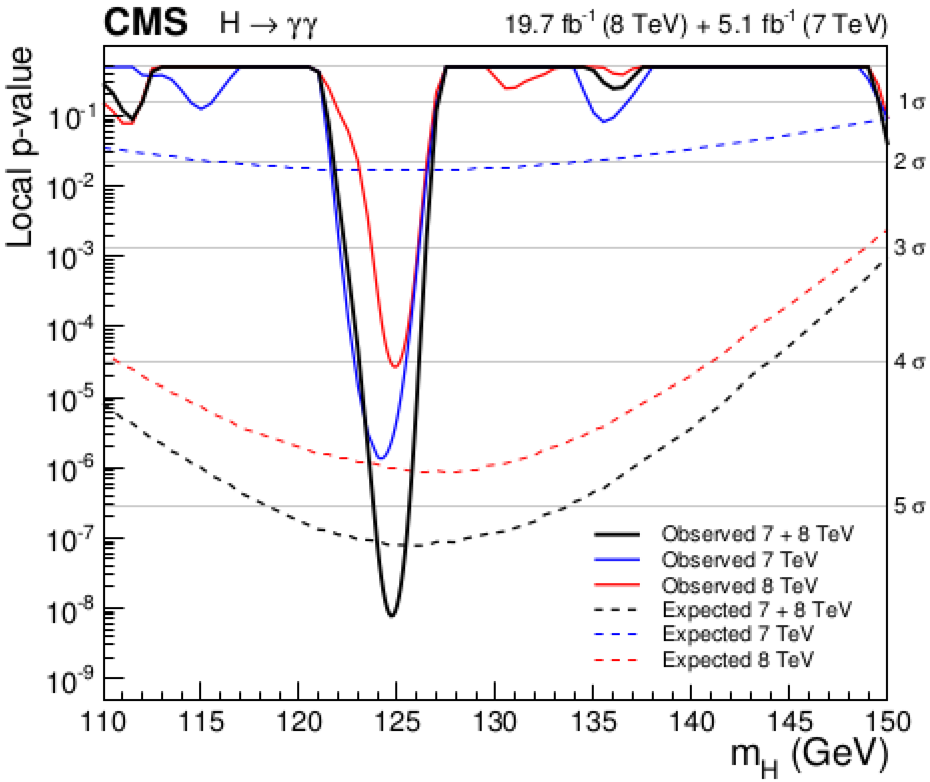
Combined mass distributions

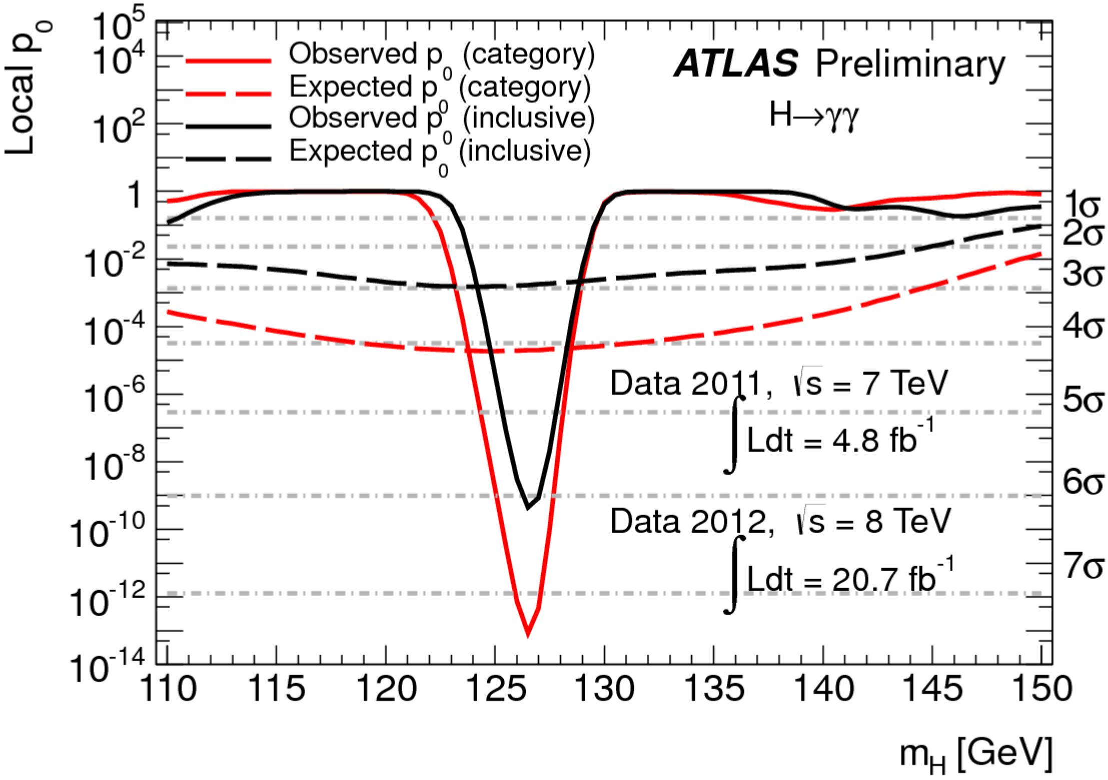




- Signal bump visible on top of background spectrum for both experiments

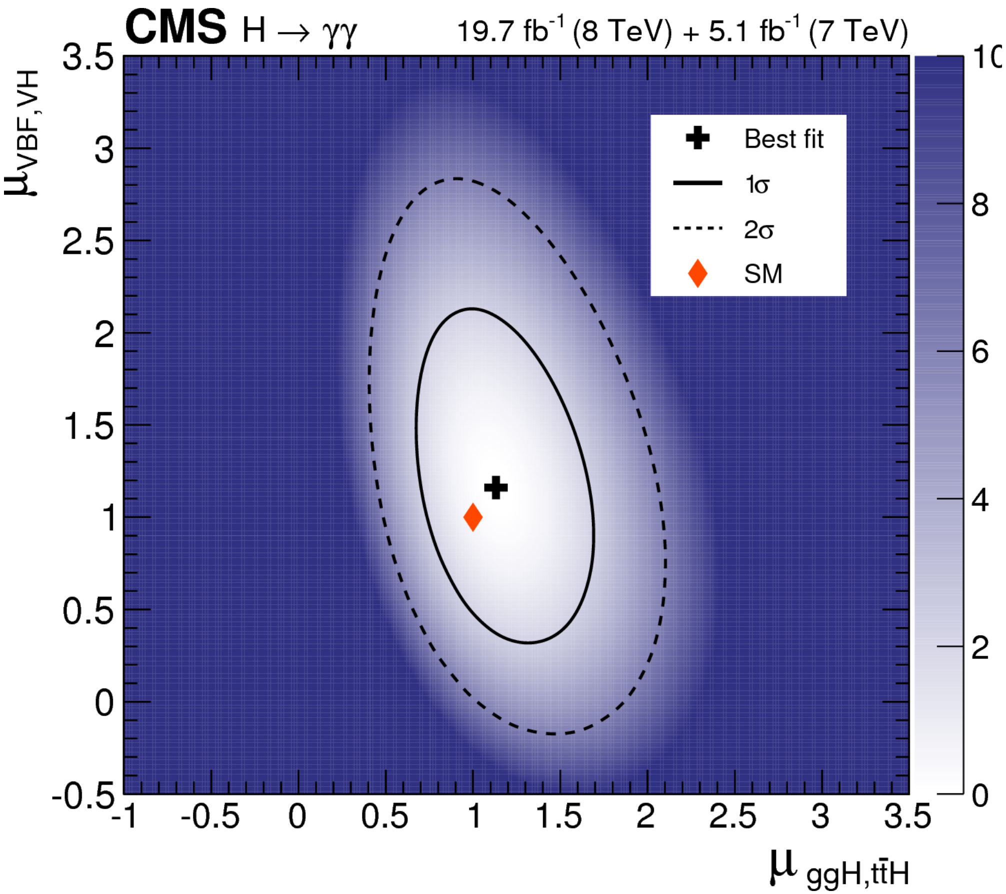
Significance and signal strength

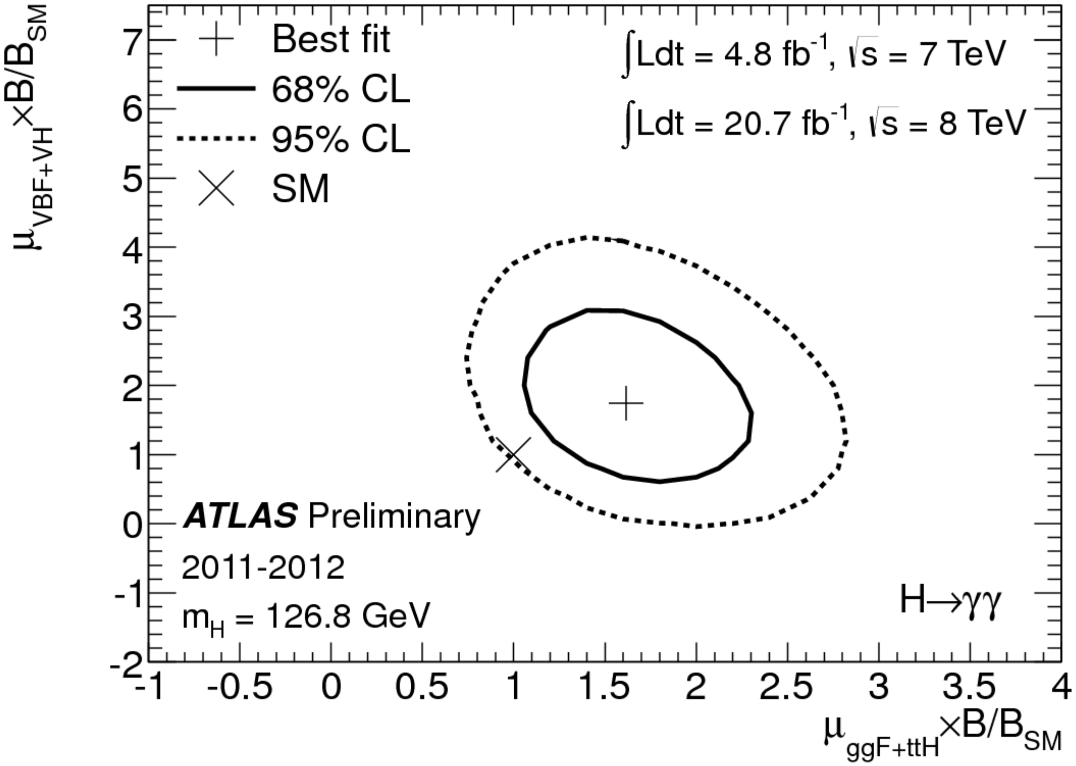




- ATLAS $\mu = 1.66 \pm 0.32$
  - Expected 4.1sigma, Observed 7.4 sigma
- CMS $\mu = 1.14 \pm 0.31$
  - Expected 5.2 sigma, observed 5.7 sigma

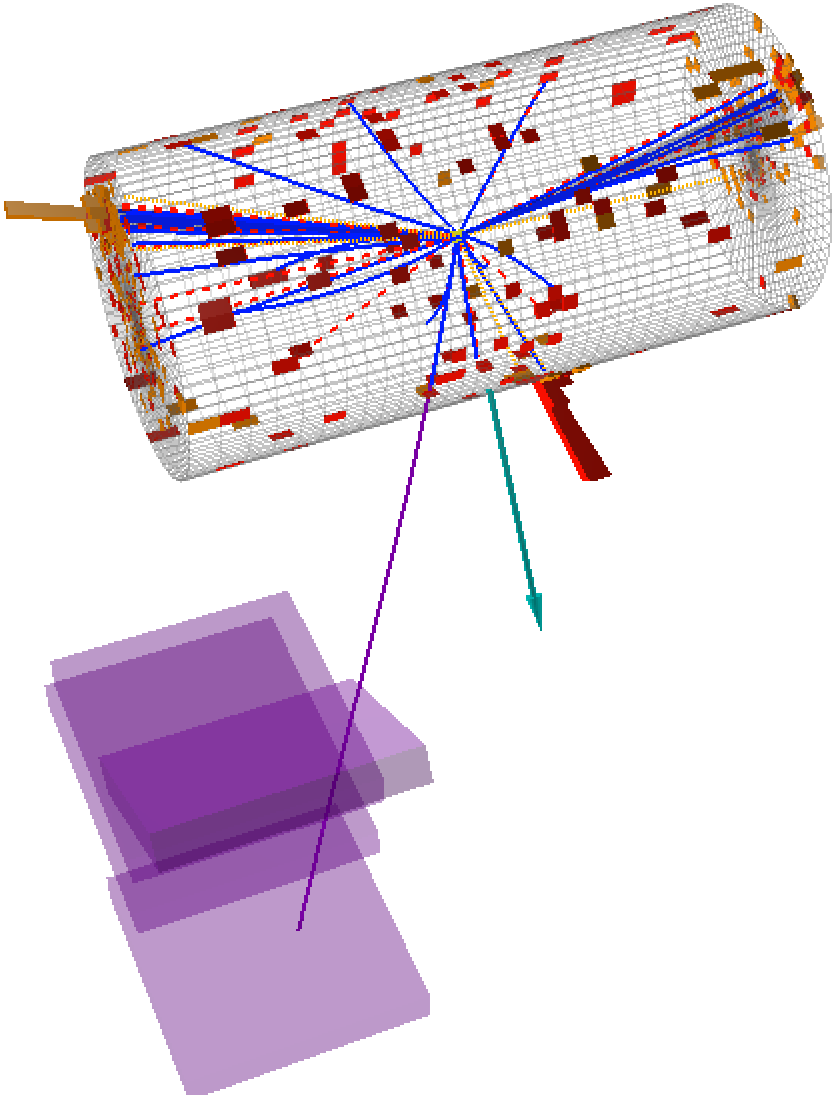
Production modes





- ATLAS consistent within 2 sigma
- CMS consistent within 1 sigma

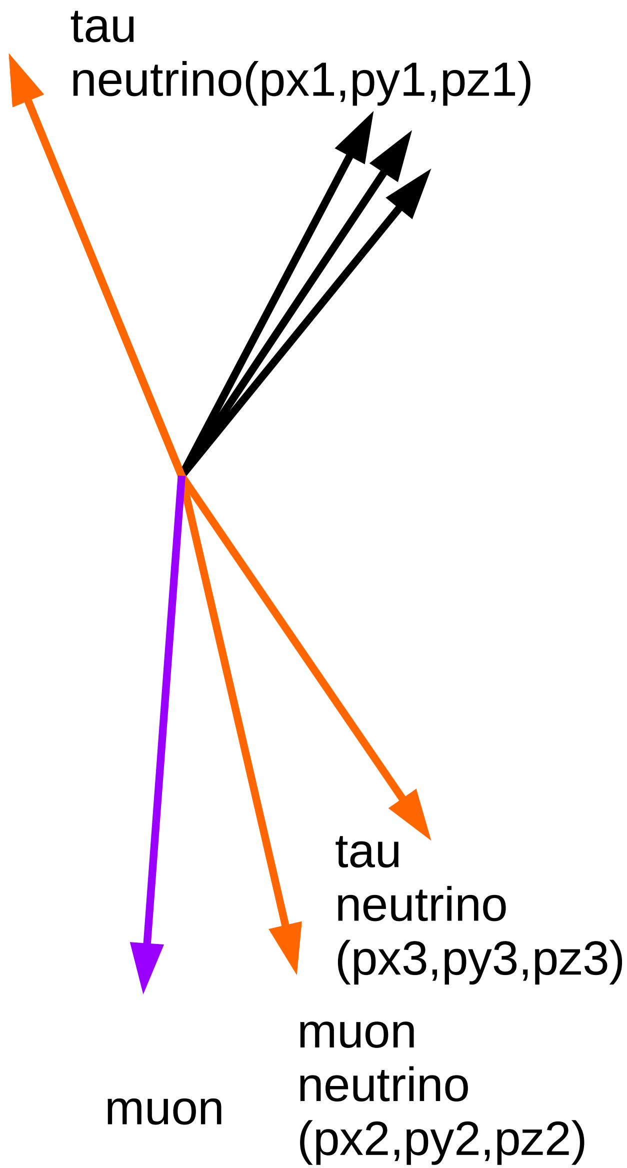
H → ττ



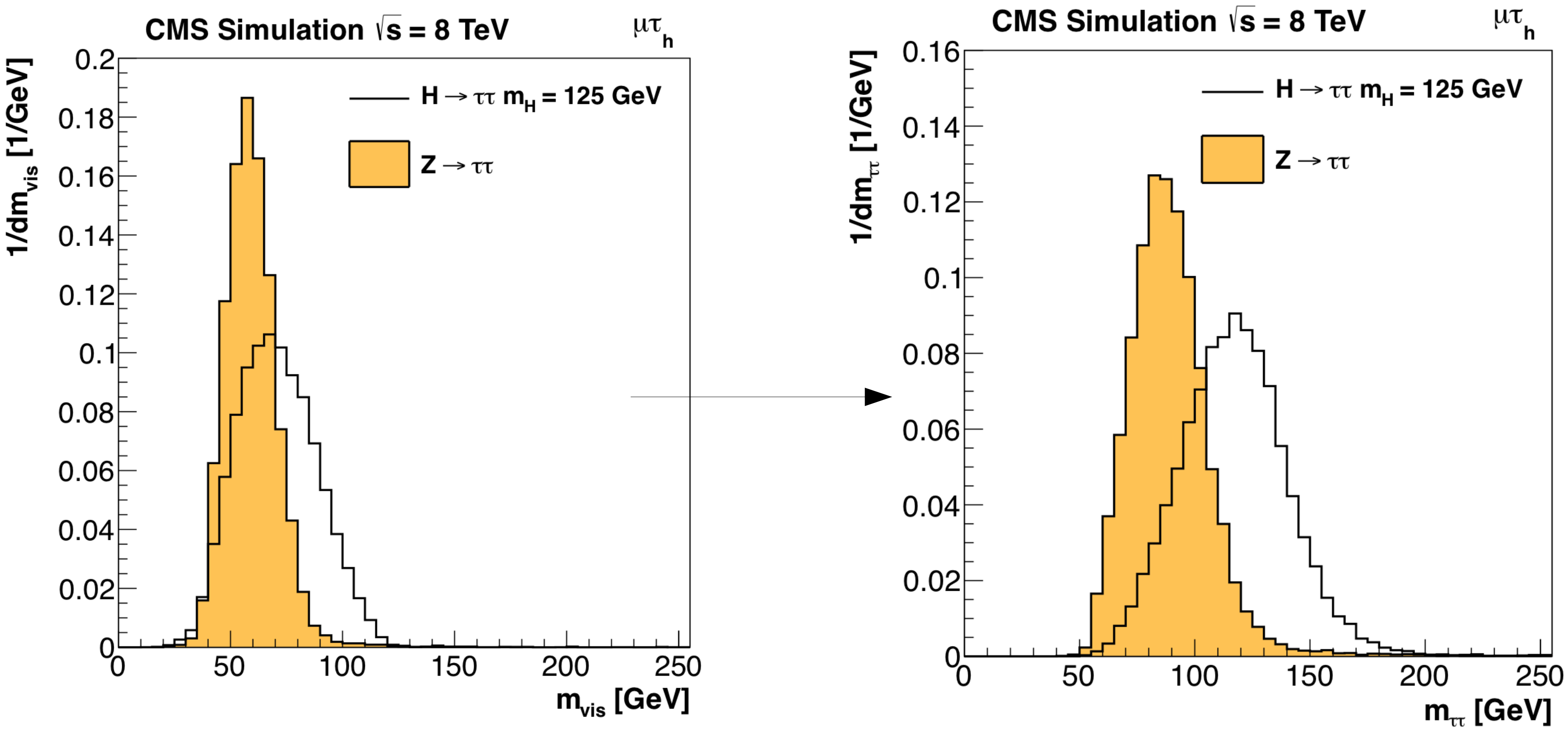
- Excellent probe to fermionic couplings!
- Require a tau pair and split sample in categories
    - Tau leptons can decay hadronically or leptonically
- Main backgrounds
    - Z → ττ, W+jets, QCD
- Major experimental challenges
    - Identify taus
    - Separate Higgs signal from Z background

Reconstructing the di-tau mass

- 2-4 neutrinos in the final state
  - 9 unknown variables!
- The two neutrinos in the muon side can be combined into one particle (px2,py2,pz2,E2) with an effective mass → 7 unknowns
- The tau mass is known (1.7 GeV) so for the two taus we go to 5 unknowns
- We have the missing transverse energy that constrains px and py → 3 unknowns
- **For the rest of the unknowns we use probability based on the way the tau decays!**



Di-Tau Mass reconstruction

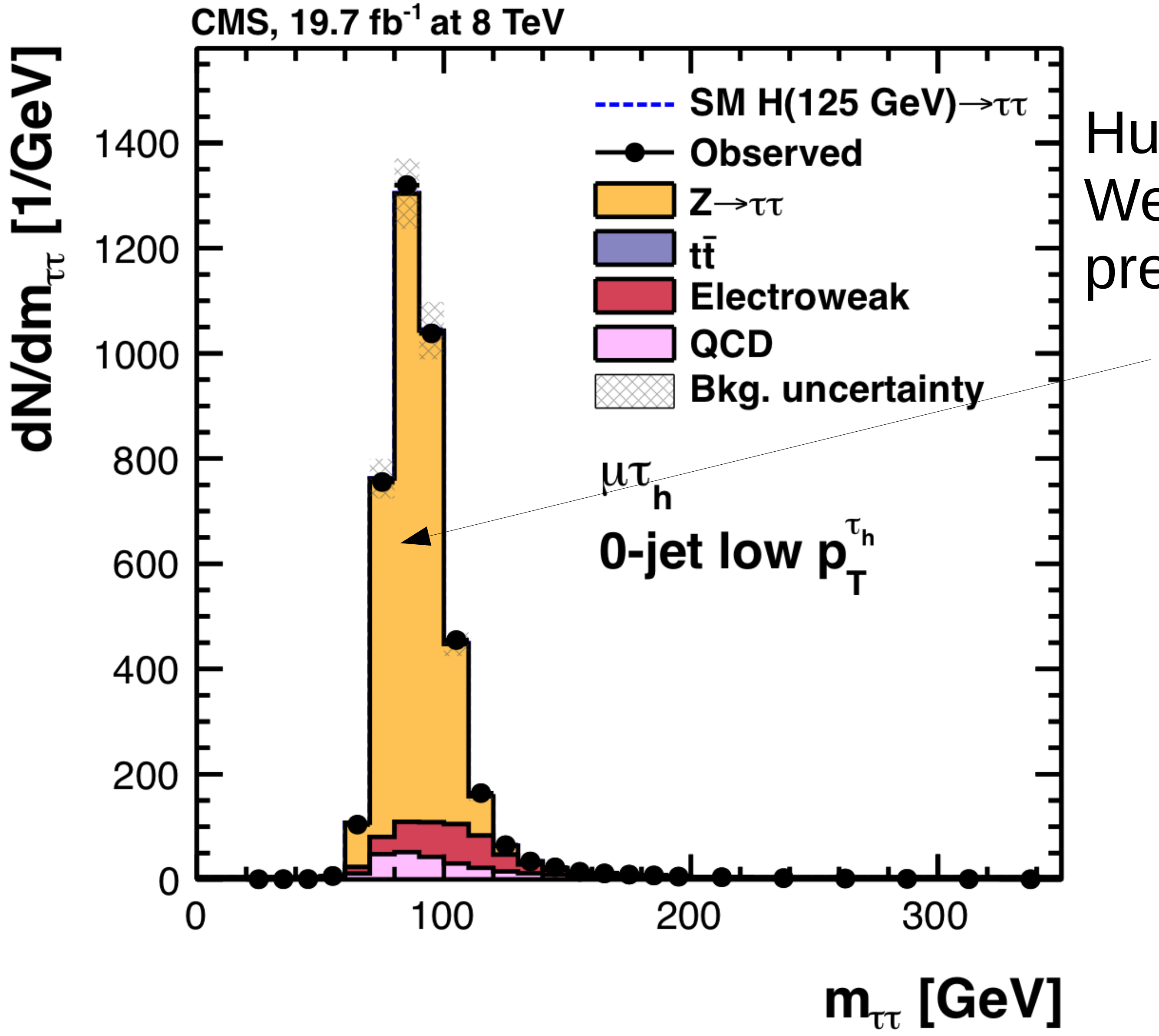


Mass of visible products

Full reconstructed mass

Better discrimination between Z and Higgs !

Inclusive distributions

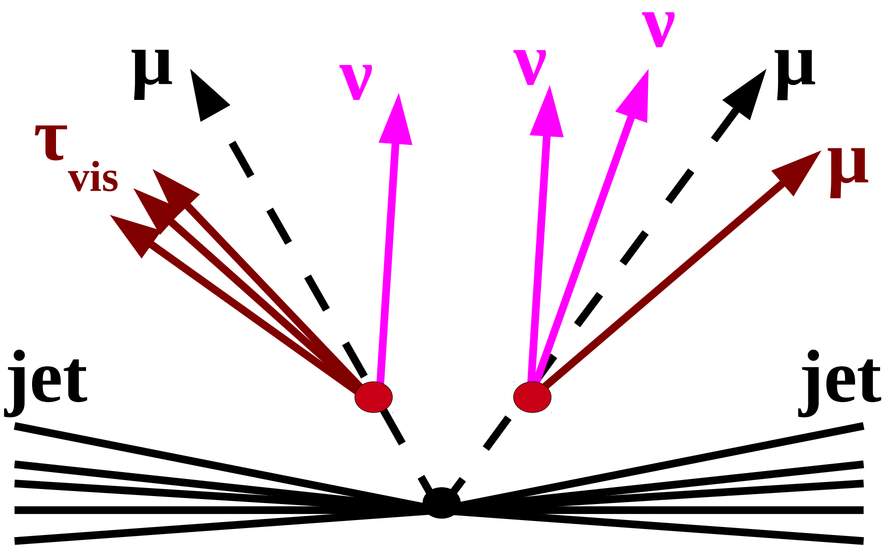


Huge contribution from Z
We need to estimate it precisely

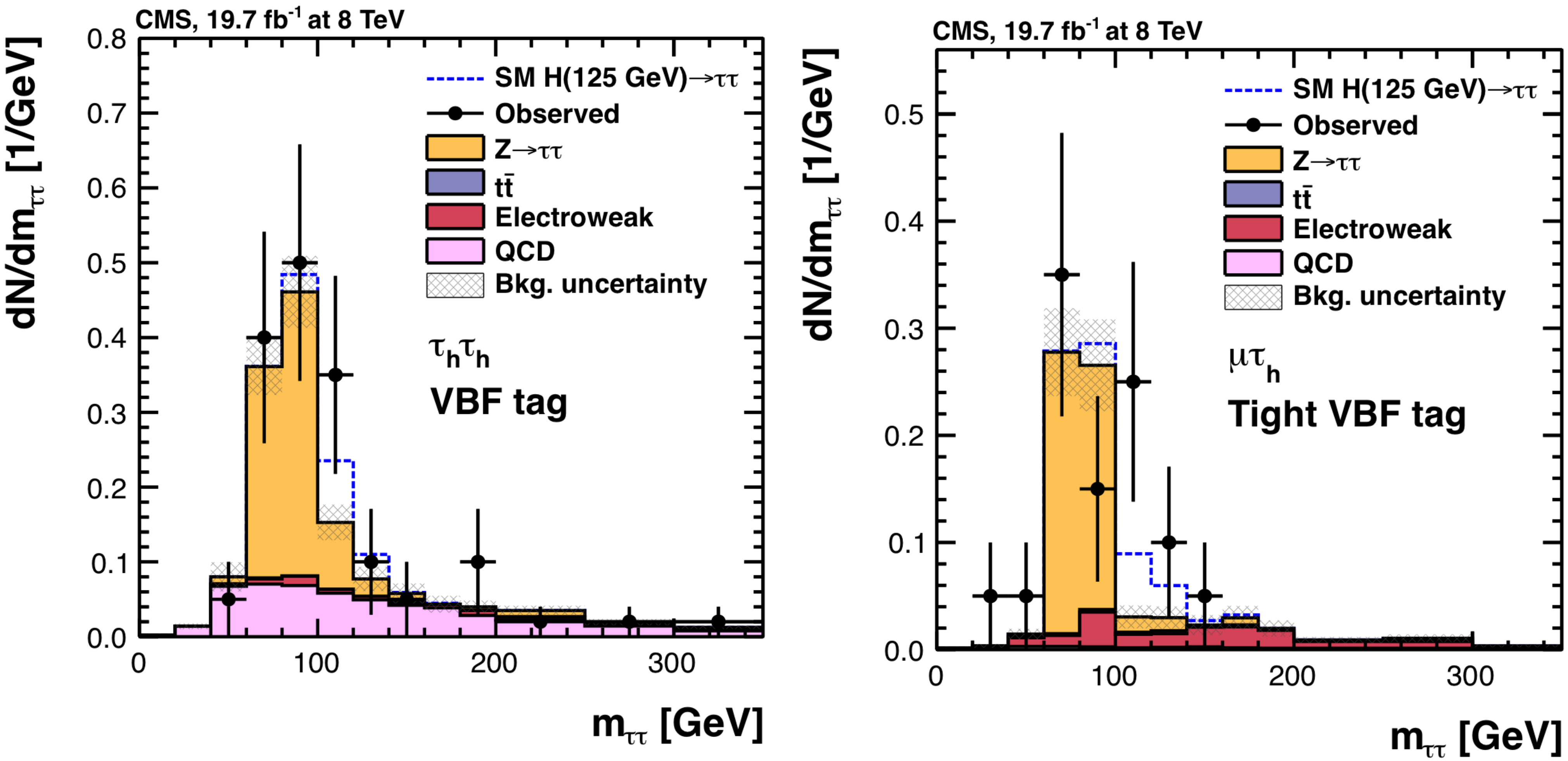
Z → ττ Background Estimation

- Reconstruct Z → μμ events in data
- Replace μ with tau and decay the event usin simulation
- Mix the simulated tau pair event with the initial events without the muon
- PU/UE and jets from data!

Embedding Technique

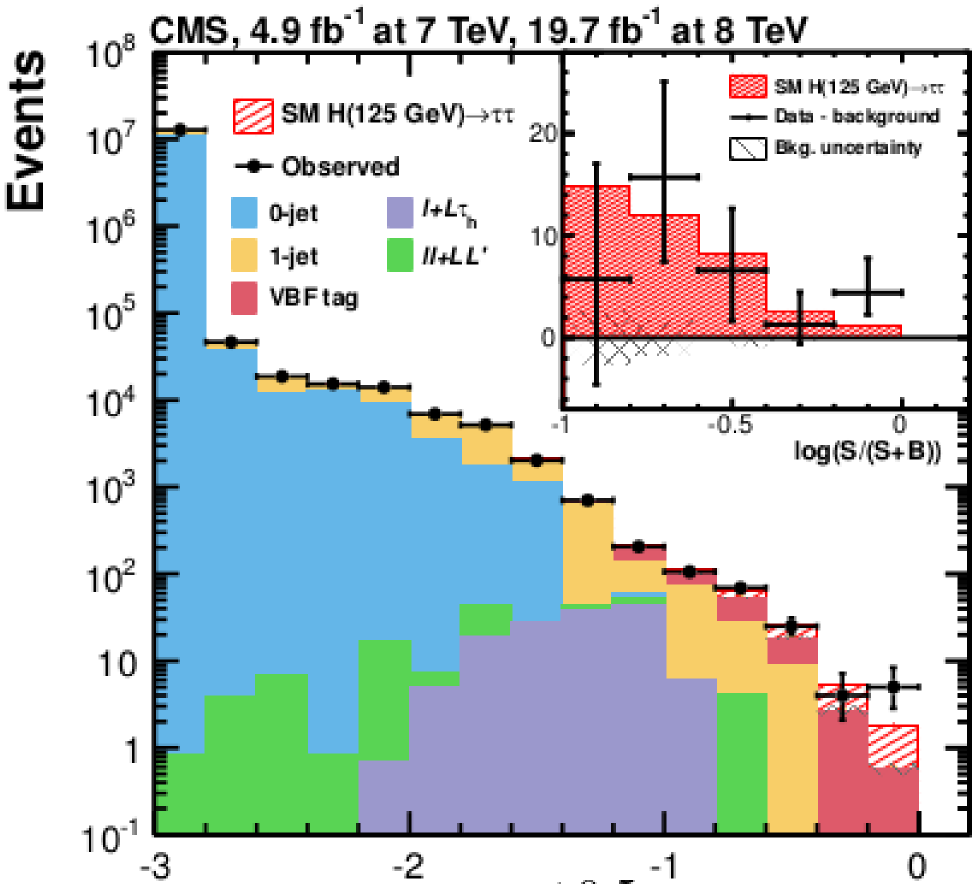


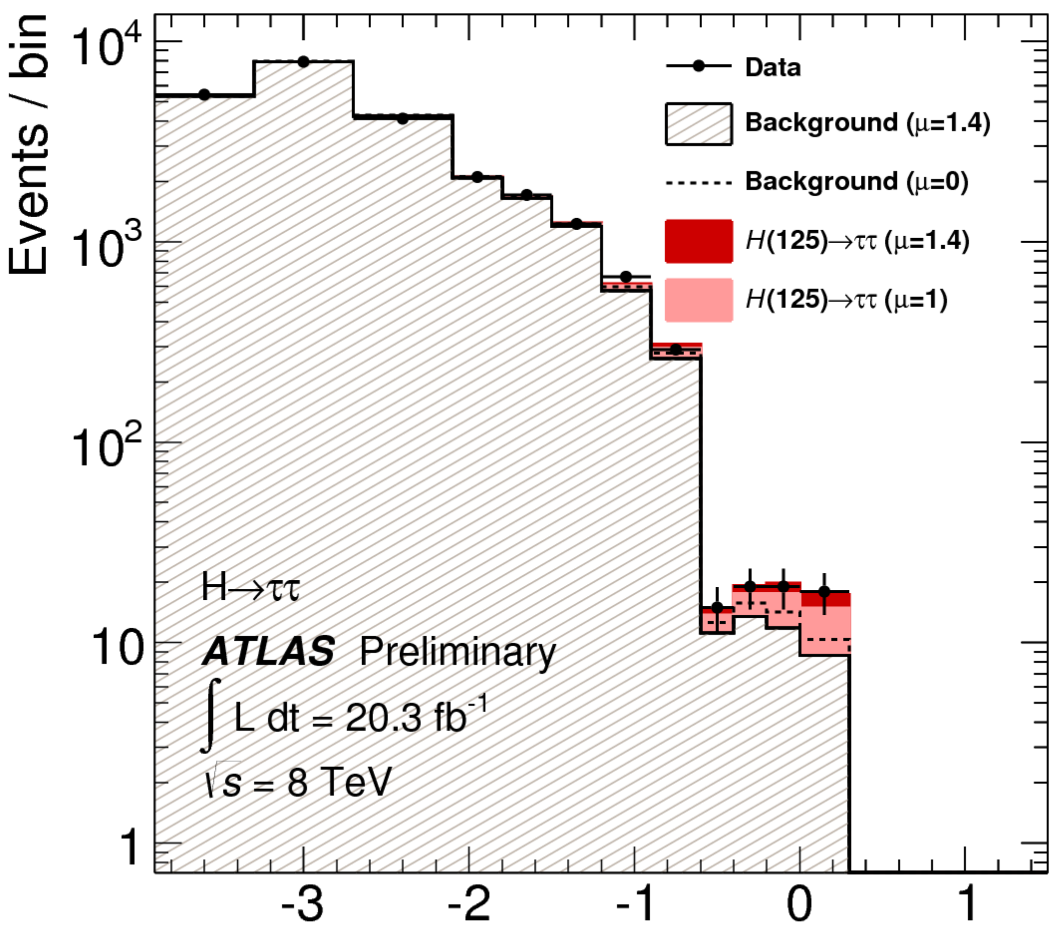
Some mass distributions (CMS)



- Excesses appearing in the expected bins!

Results

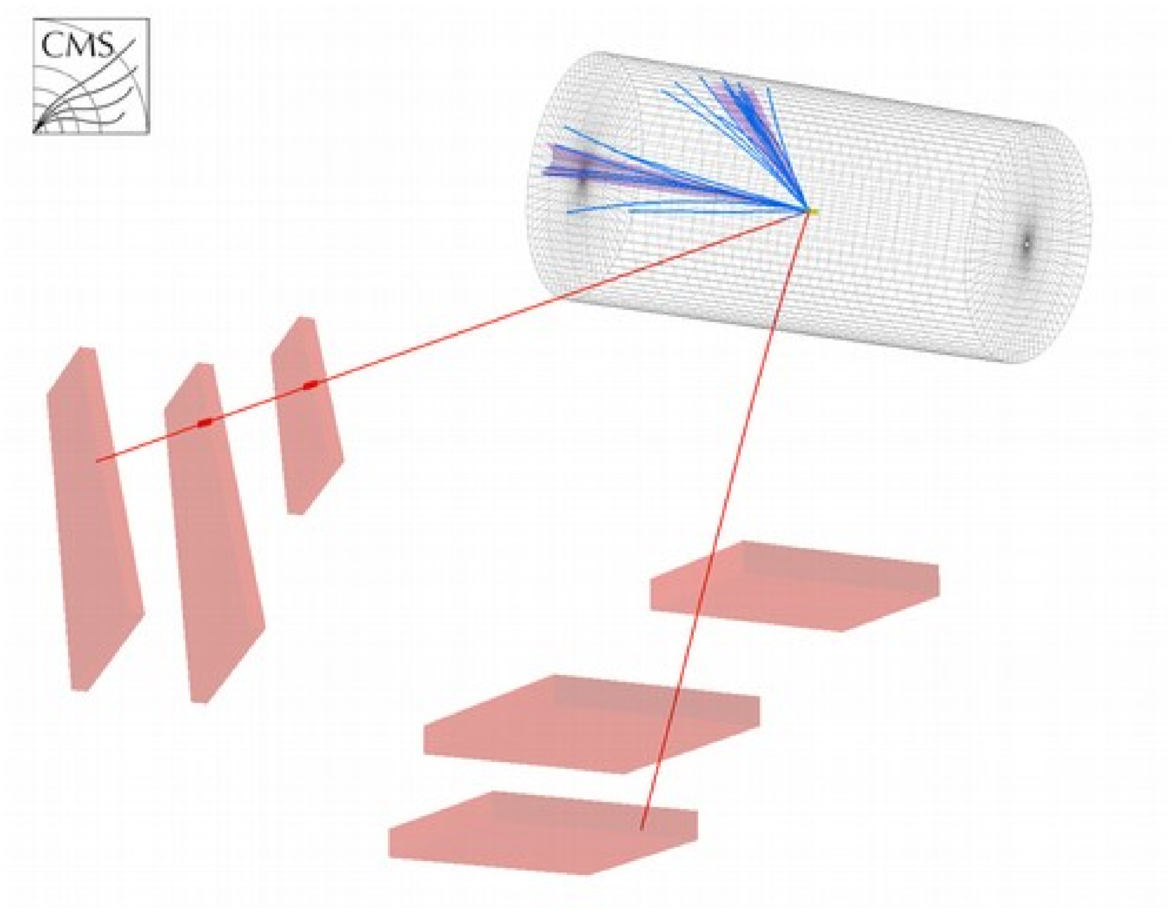




- ATLAS $\mu = 1.4^{+0.5}_{-0.4}$
  - Expected 3.2 sigma, observed 4.1 sigma
- CMS $\mu = 0.8 \pm 0.3$
  - Expected 3.7 sigma, observed 3.2 sigma

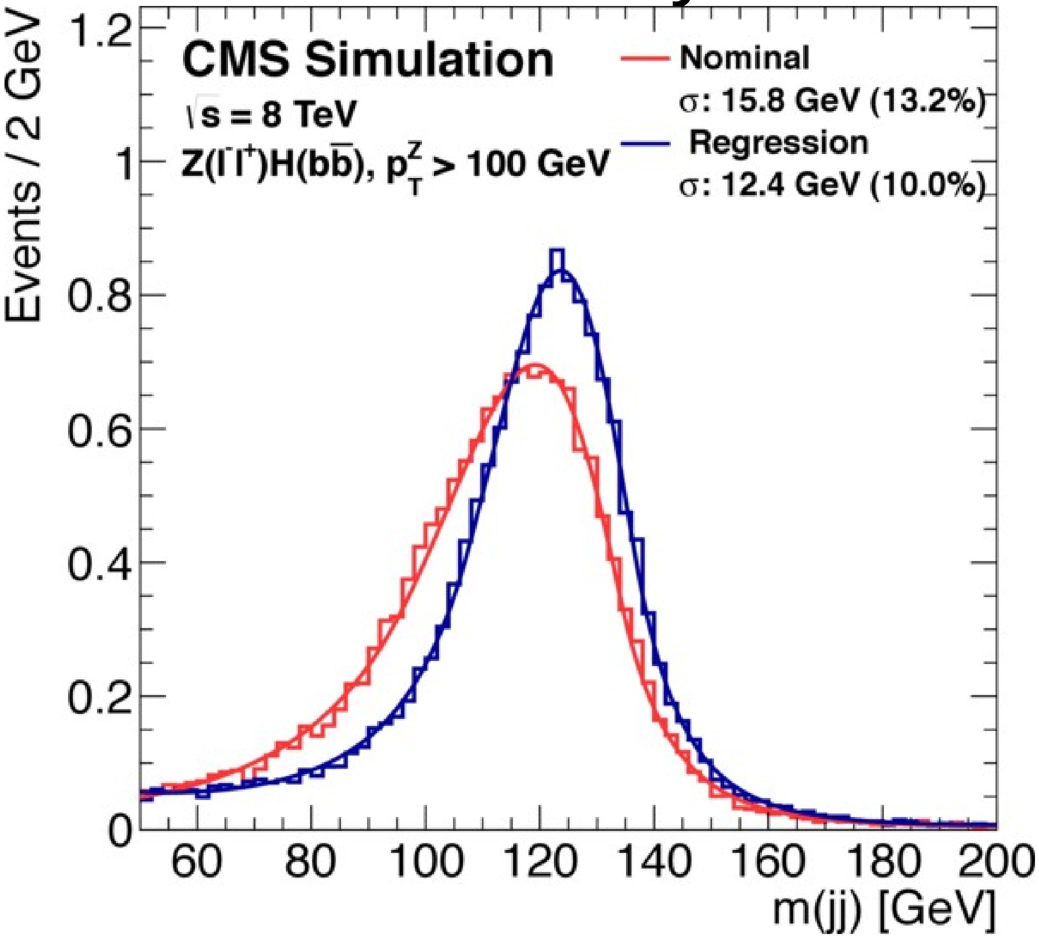
H → bb

- Vector boson signature + 2 bjets
- Large background
- Mass resolution depending on jet resolution
- Higgs and vector boson recoil away from each other
    - Higgs system boosted

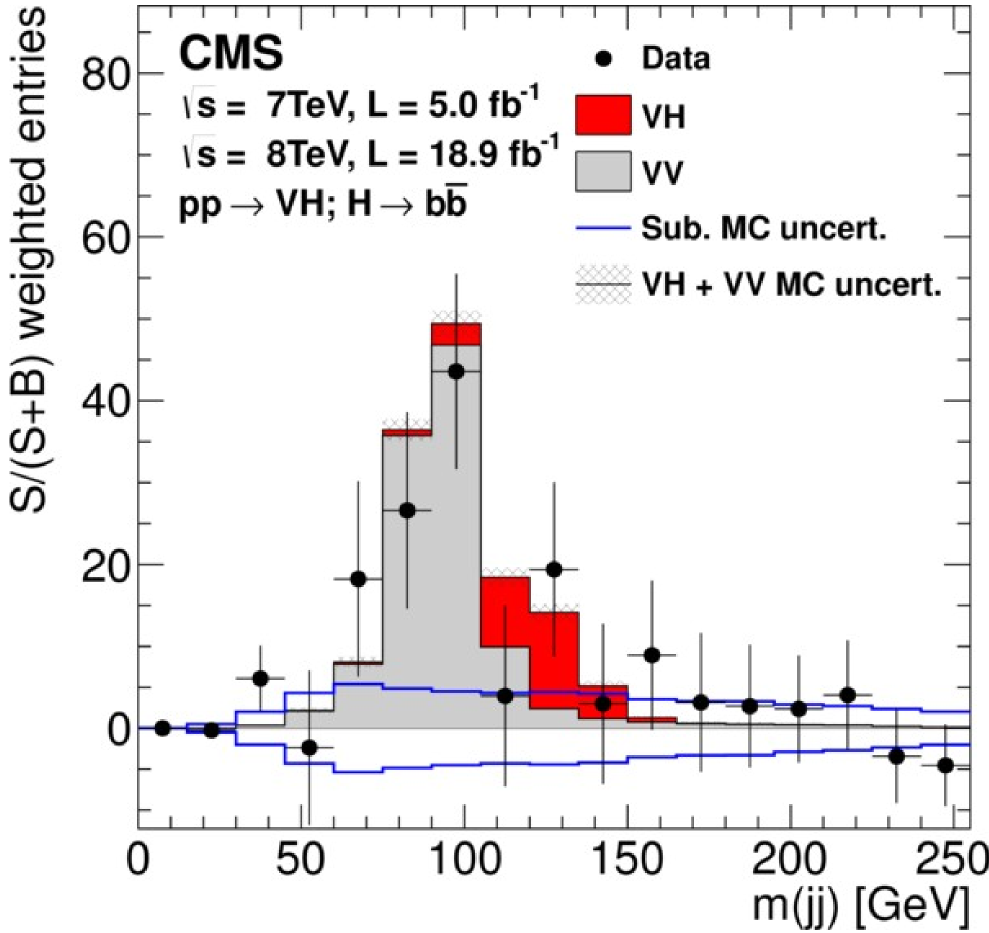


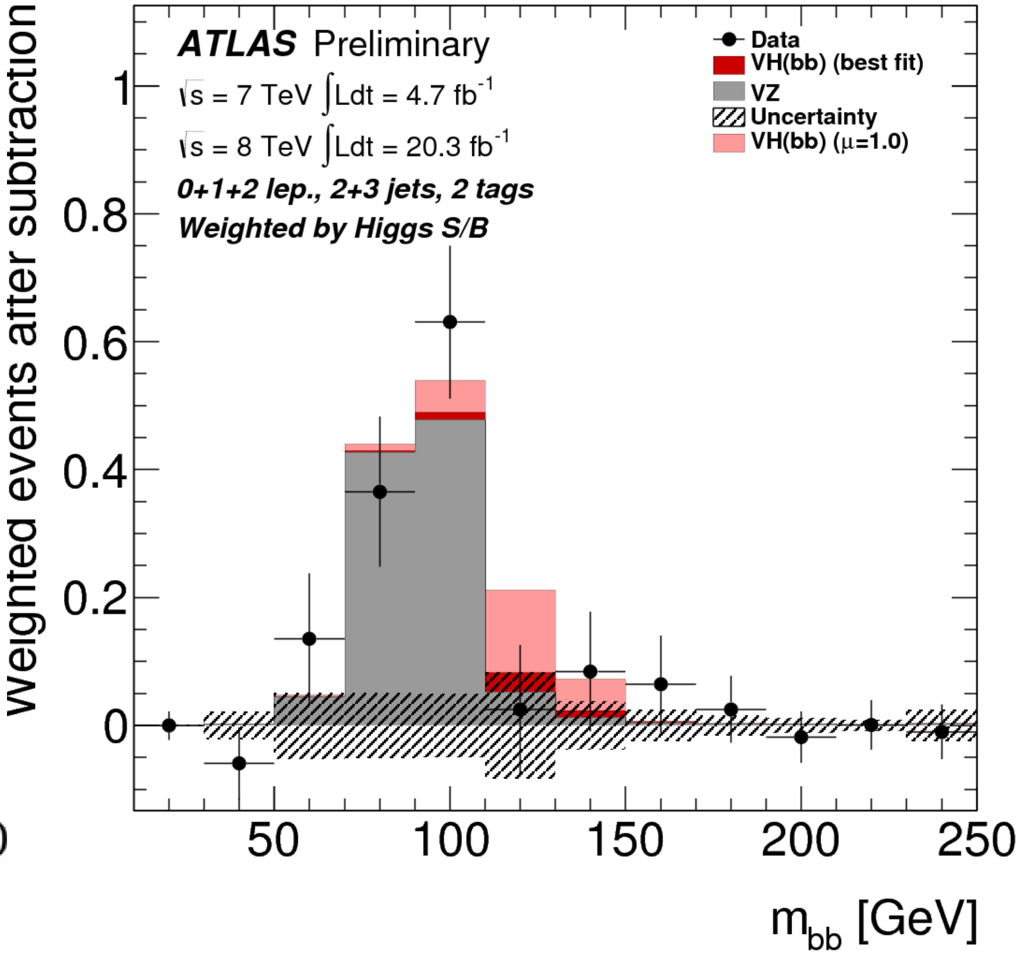
Analysis strategy

- Separate events in categories of vector boson (or Higgs) pt and try to reconstruct the bb mass
- Multivariate regression techniques used to improve the mass resolution of the bb system



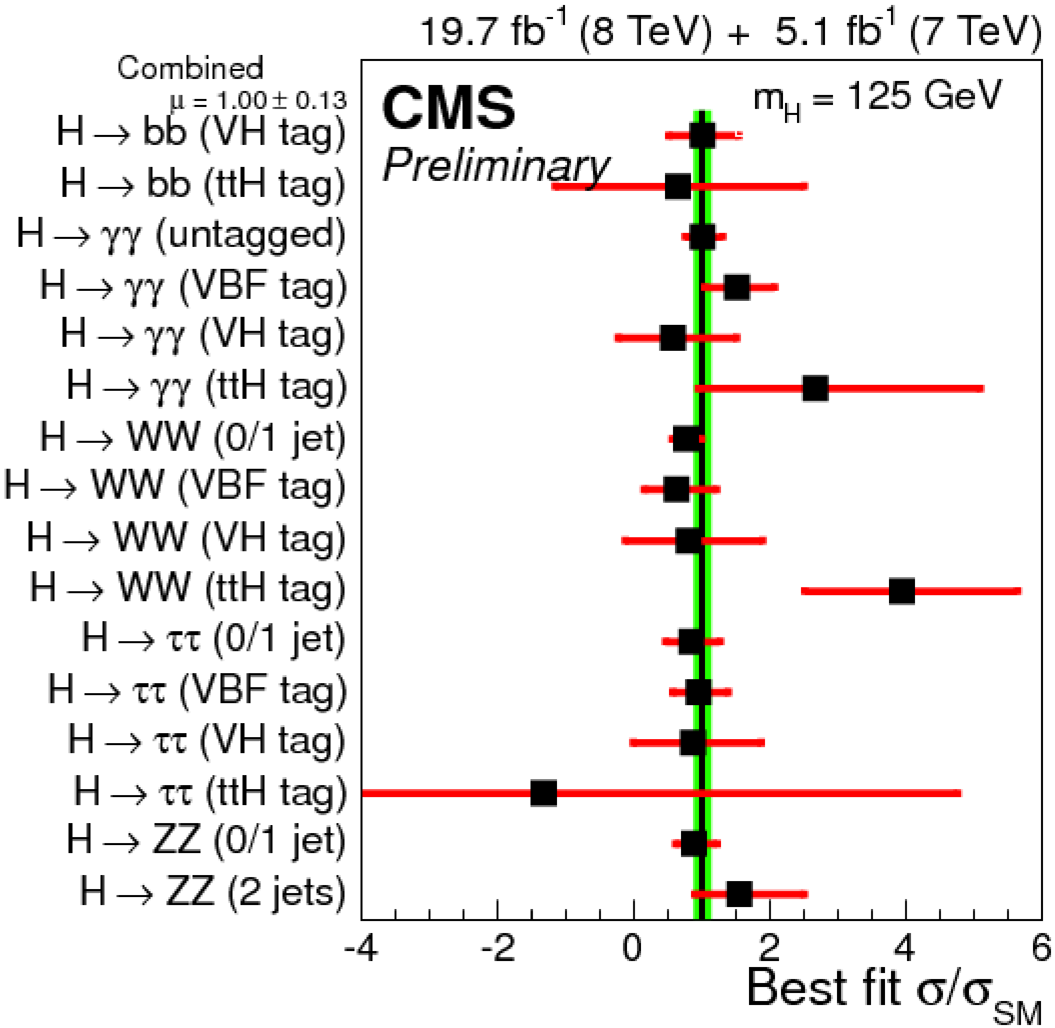
Results

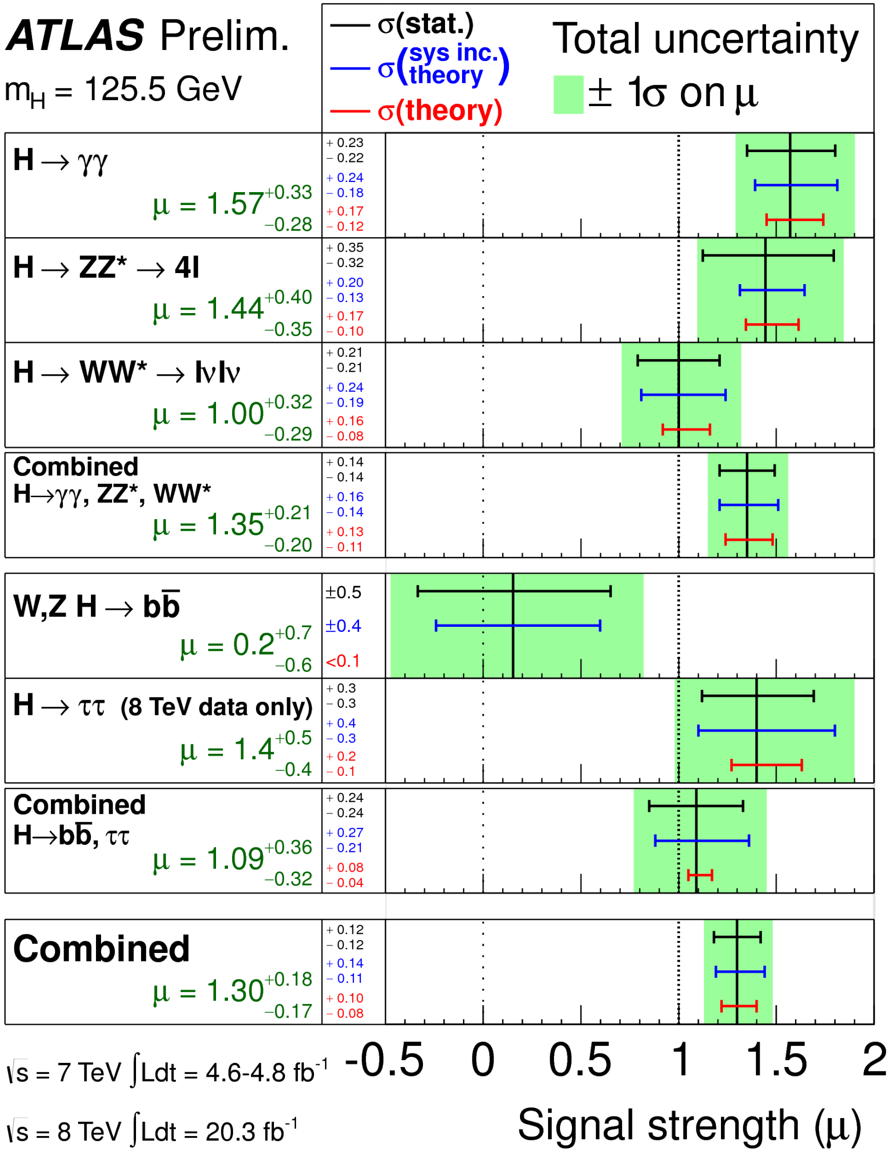




- Very difficult analysis
  - CMS observes 2sigma excess [ATLAS ~ 1 sigma]
  - More data needed to observe it!

Summary of measurements



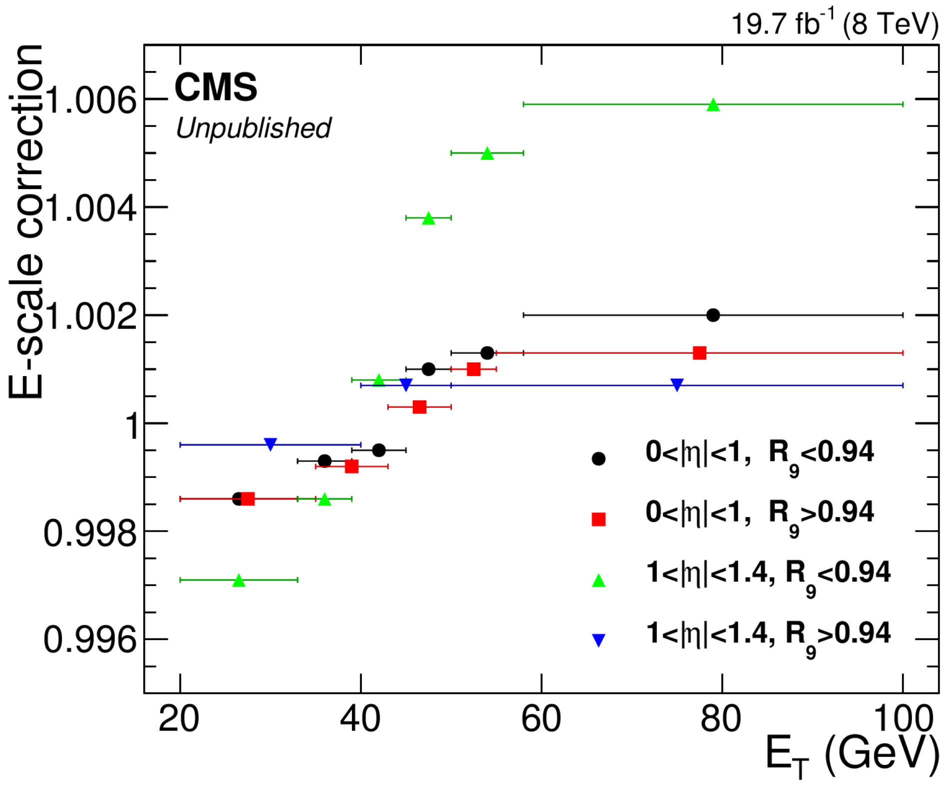


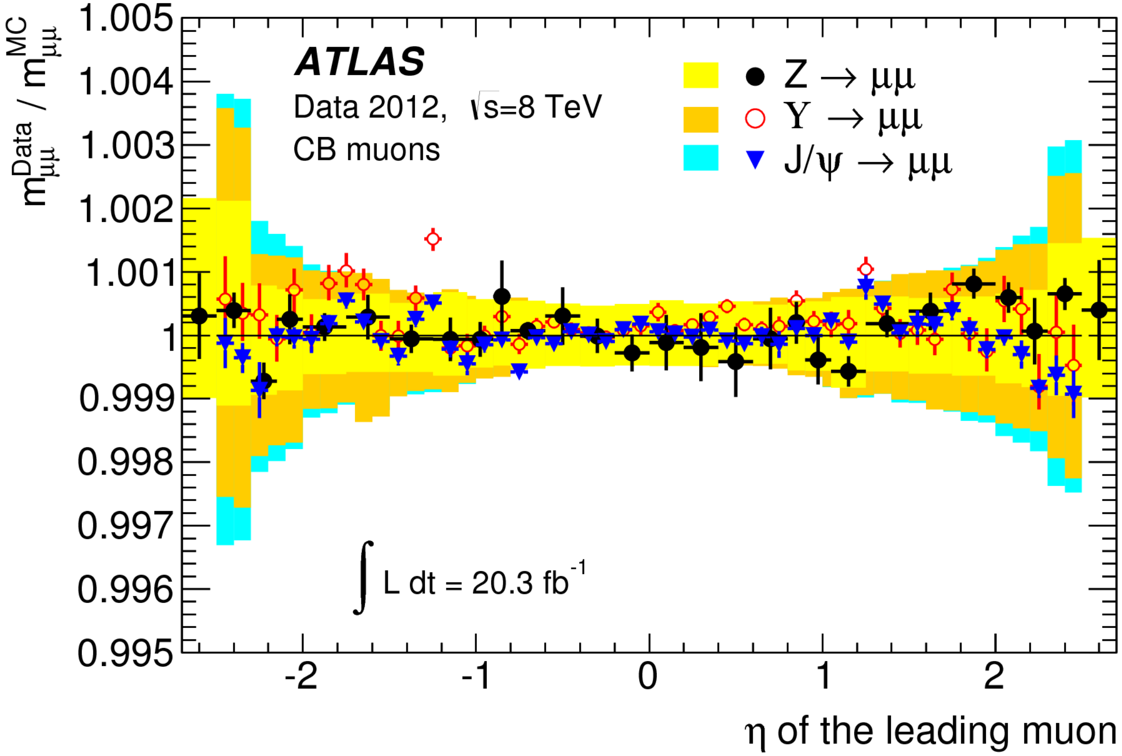
Combined = 1.00 +- 0.13

Combined = 1.30 +- 0.18
ATLAS combination to be updated soon

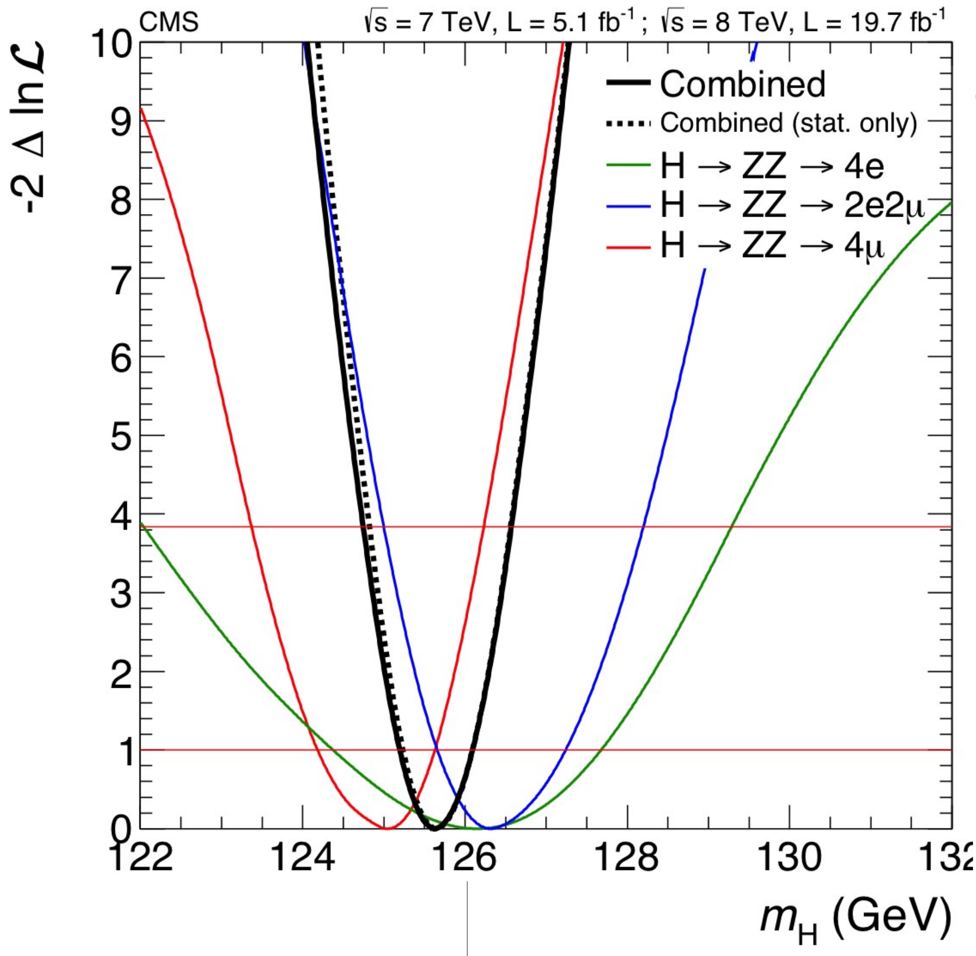
Measurement of the Higgs mass

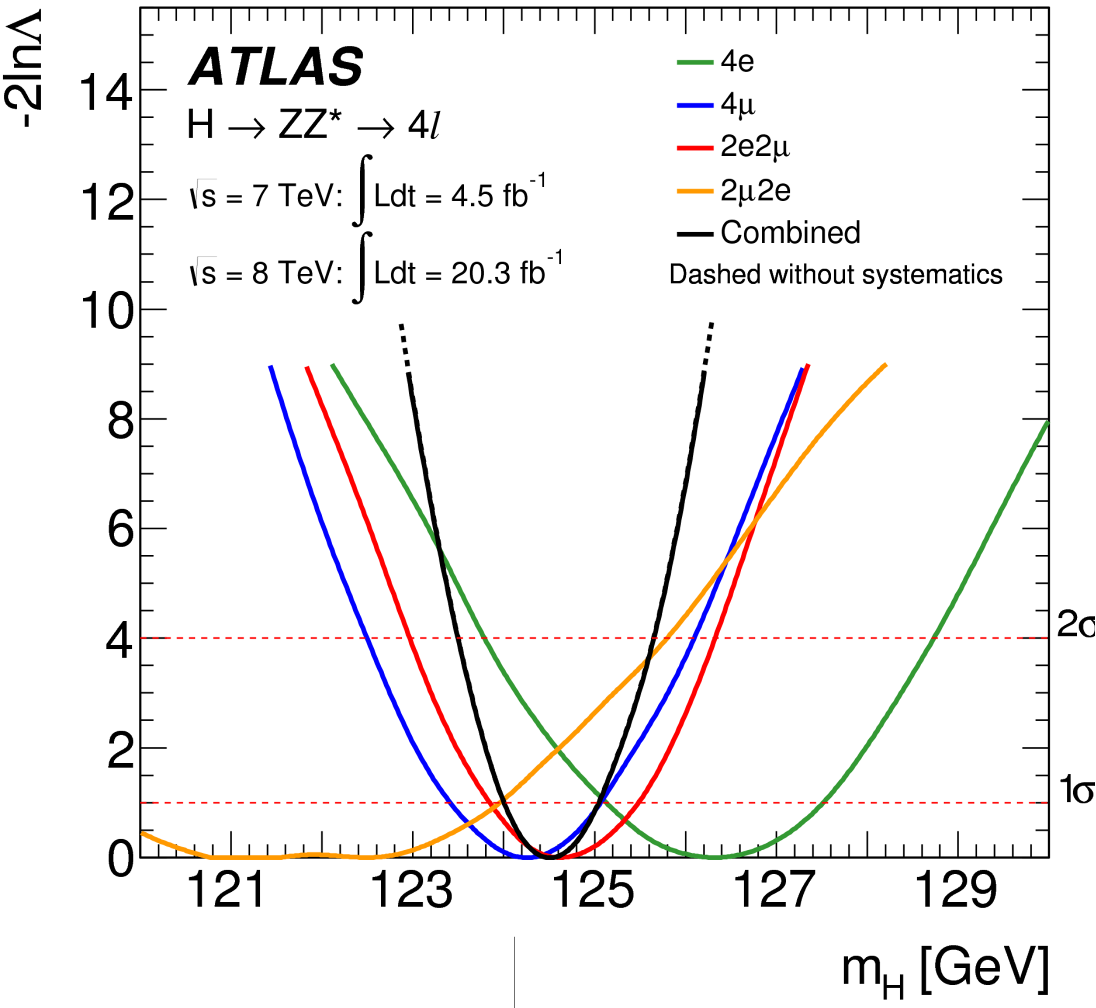
- The mass is measured by the final states that have good resolution (di-photon and four lepton)
- Excellent detector calibration is needed!





Four lepton mass measurement

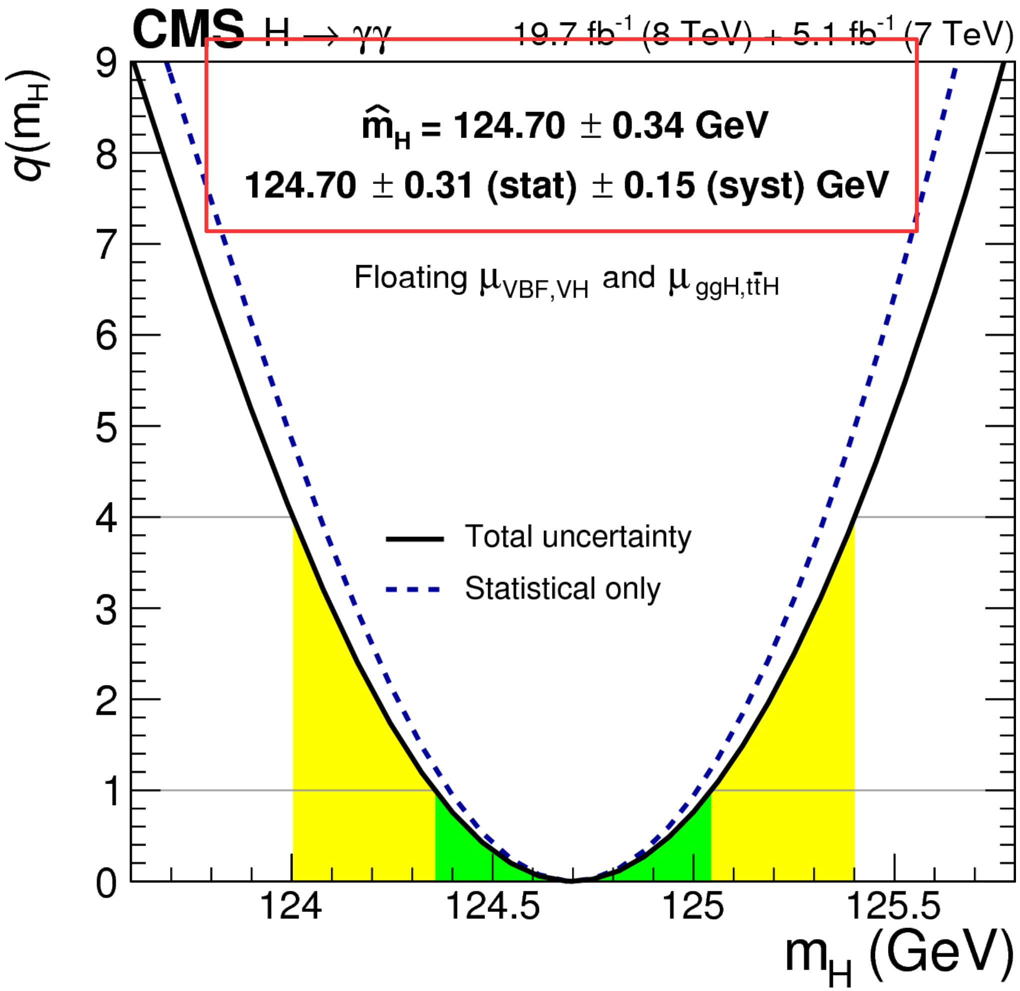


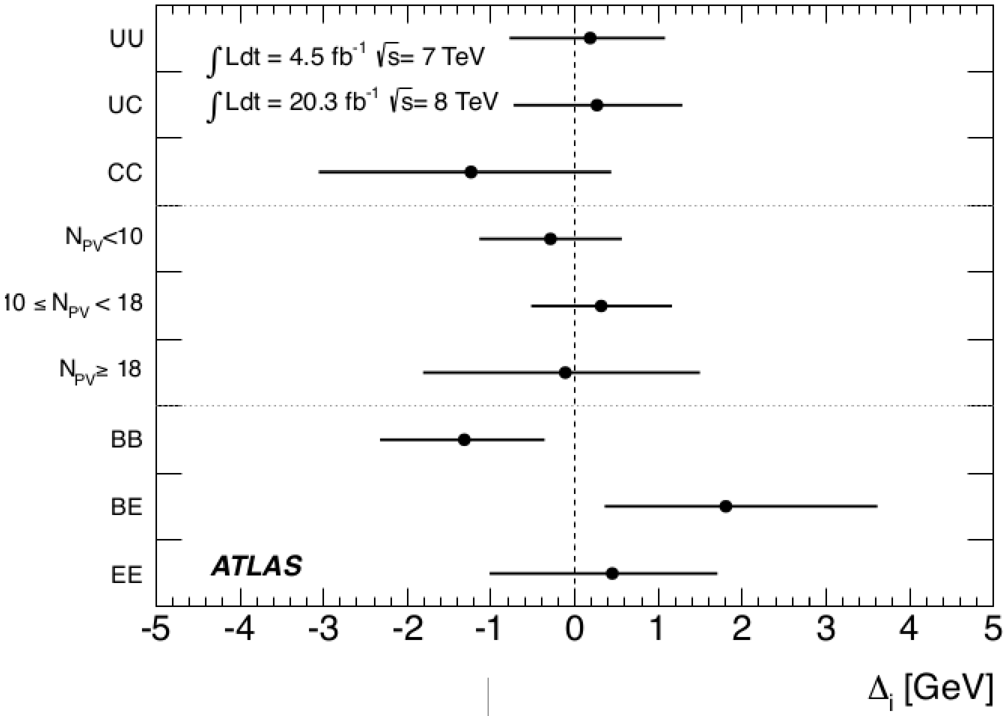


$$m_H = 125.6 \pm 0.4(stat) \pm 0.2(syst)\ \text{GeV}$$

$$m_H = 124.51 \pm 0.52(stat) \pm 0.06(syst)\ \text{GeV}$$

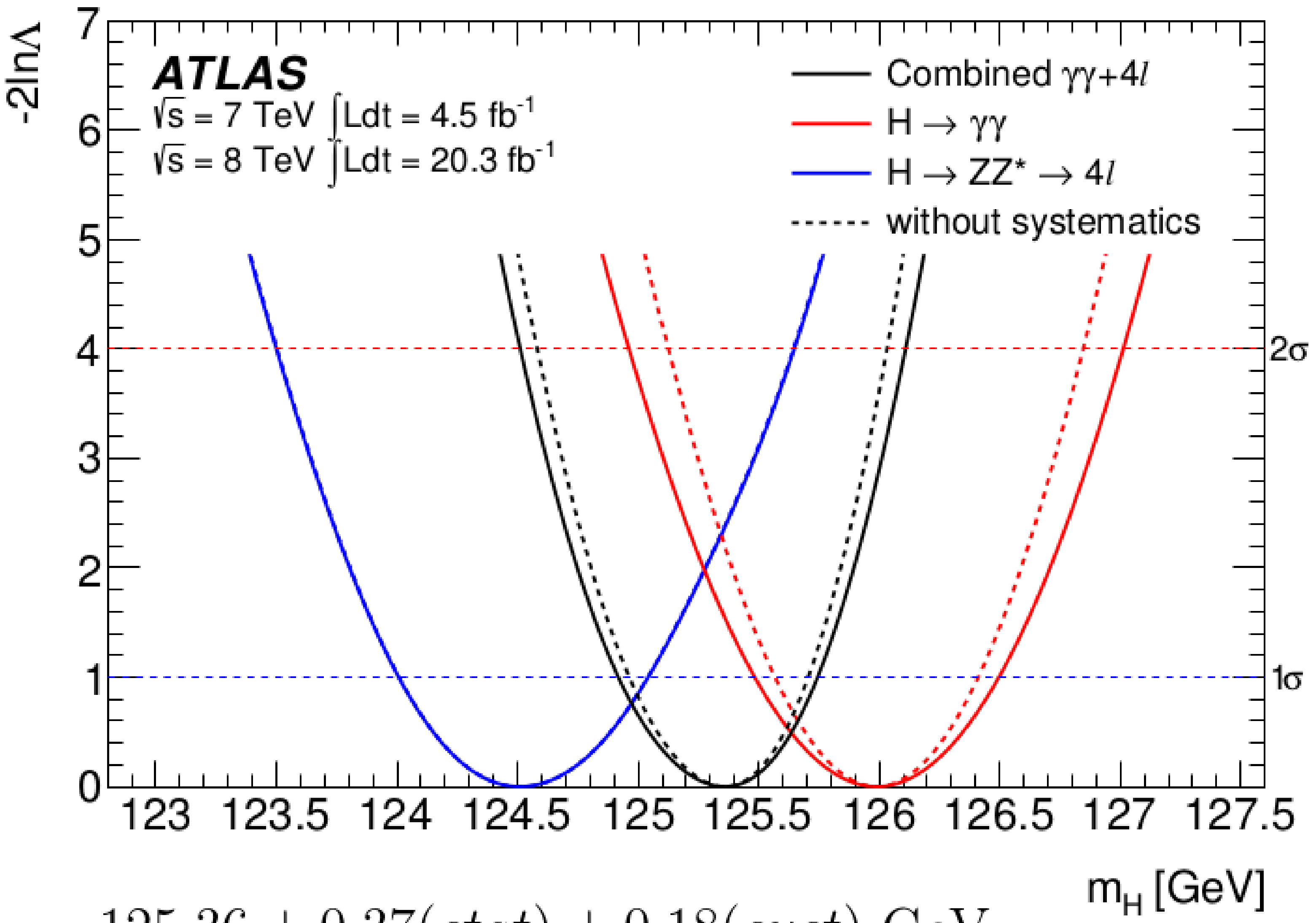
Di photon mass measurement





$$m_H = 125.98 \pm 0.32(stat) \pm 0.28(syst)\ \text{GeV}$$

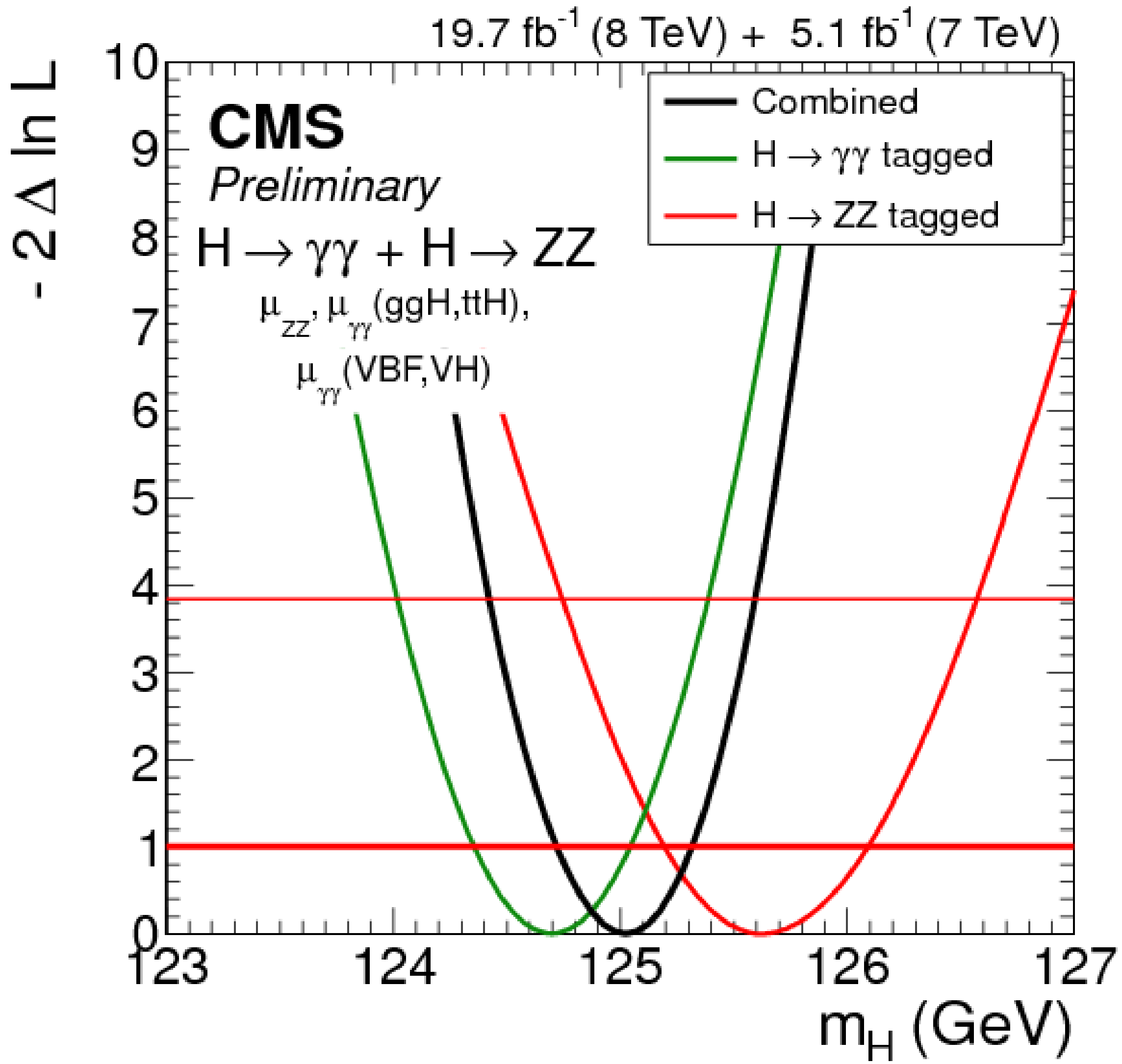
Mass combination (ATLAS)



$$m_H = 125.36 \pm 0.37(stat) \pm 0.18(syst)\ \mathrm{GeV}$$

Compatibility of the two final states at 2-2.5 sigma level

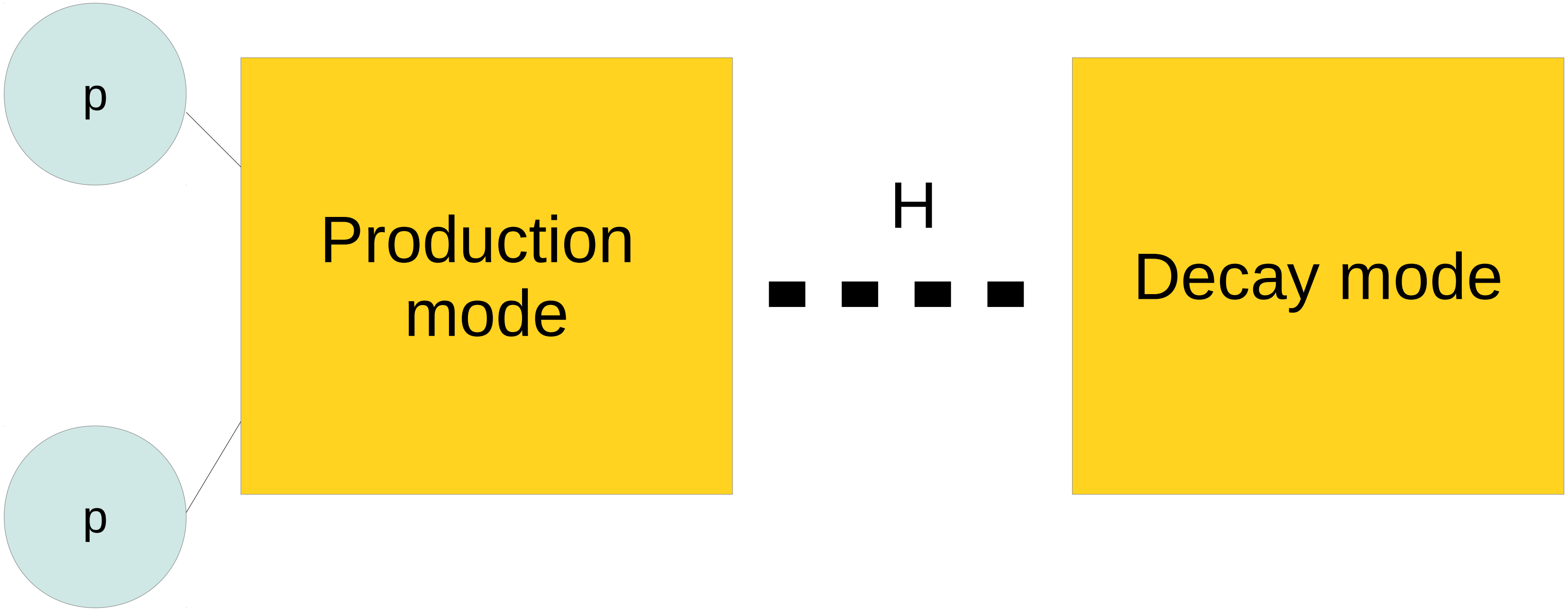
Mass combination (CMS)



$$m_H = 125.03^{+0.26}_{-0.27}(stat)^{+0.13}_{-0.15}(syst)\ \text{GeV}$$

- Agreement between final states at 1.6 sigma level

How do we measure the couplings?

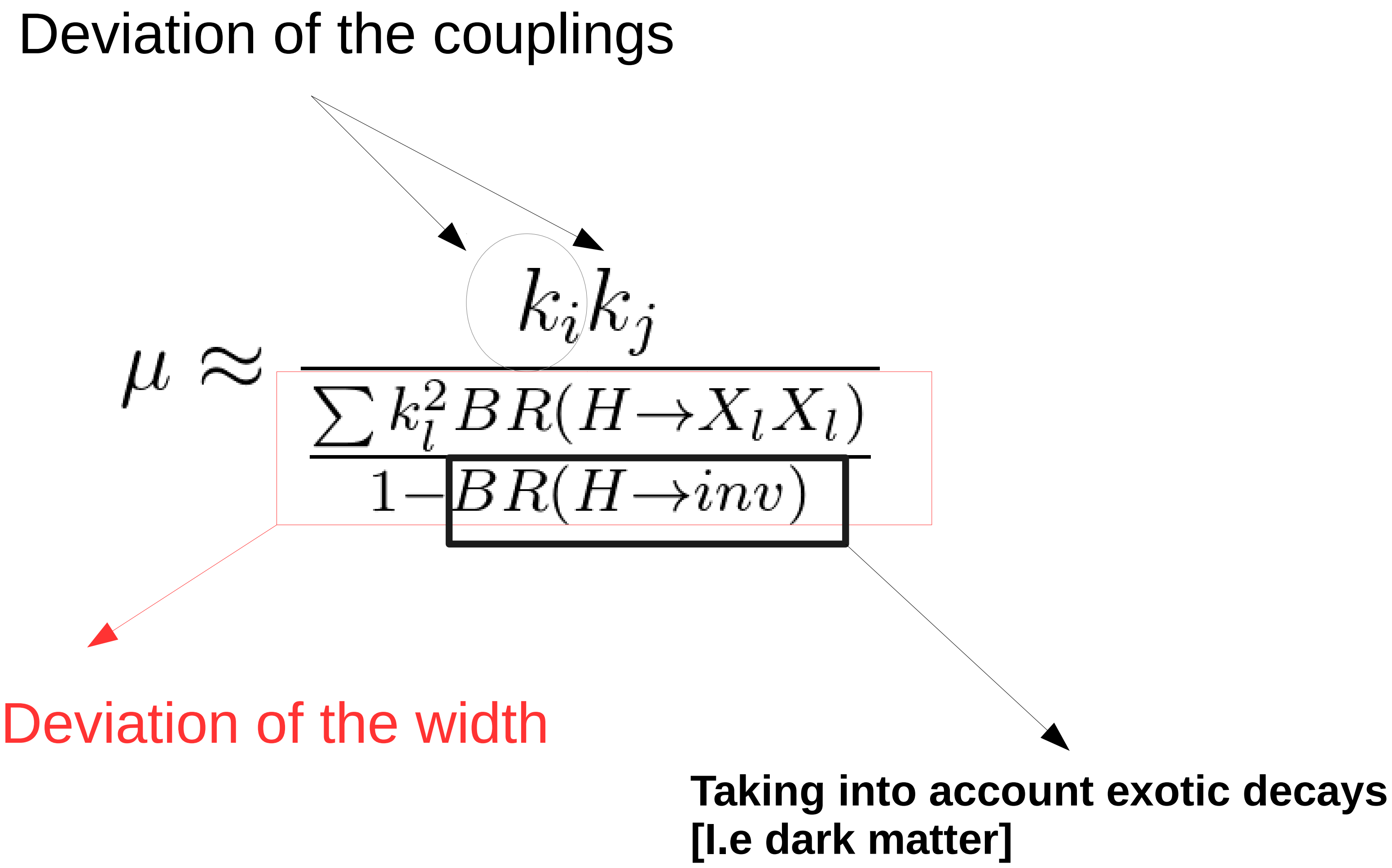


- The coupling of the Higgs to the individual particles affects both production and decay mode!
- For a mode: pp → X → H → YY the total signal strength is proportional to

Couplings → $$\frac{g_X^2 g_Y^2}{\Gamma}$$ → Higgs decay width

How do we measure the couplings

- We define deviations K from the expected numbers



Deviation of the couplings

$$\mu \approx \frac{k_i k_j}{\frac{\sum k_l^2 BR(H\rightarrow X_l X_l)}{1-\boxed{BR(H\rightarrow inv)}}}$$

Deviation of the width

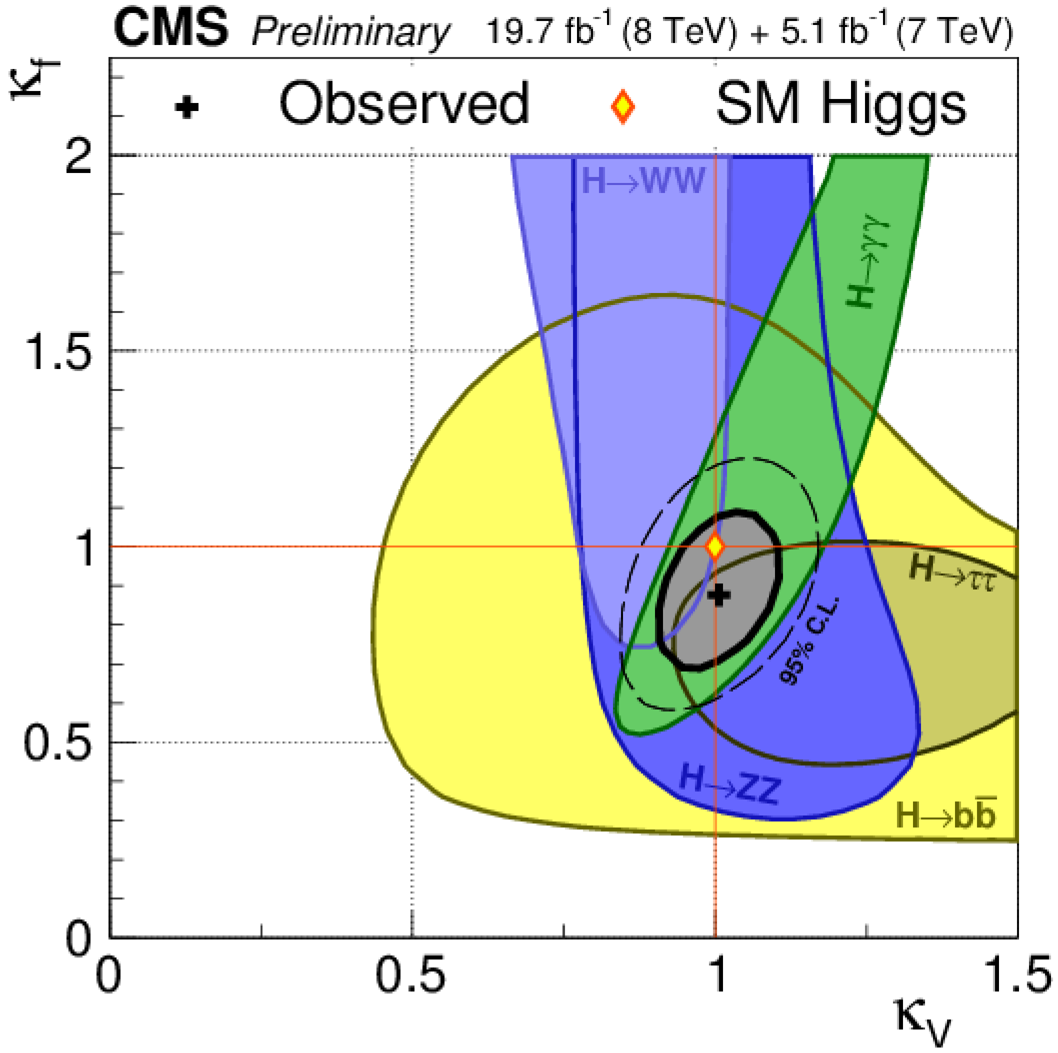
Taking into account exotic decays [I.e dark matter]

Examples

- ggH , H → bb ~ [k_g,k_b]
- WH H → WW ~ [k_W]

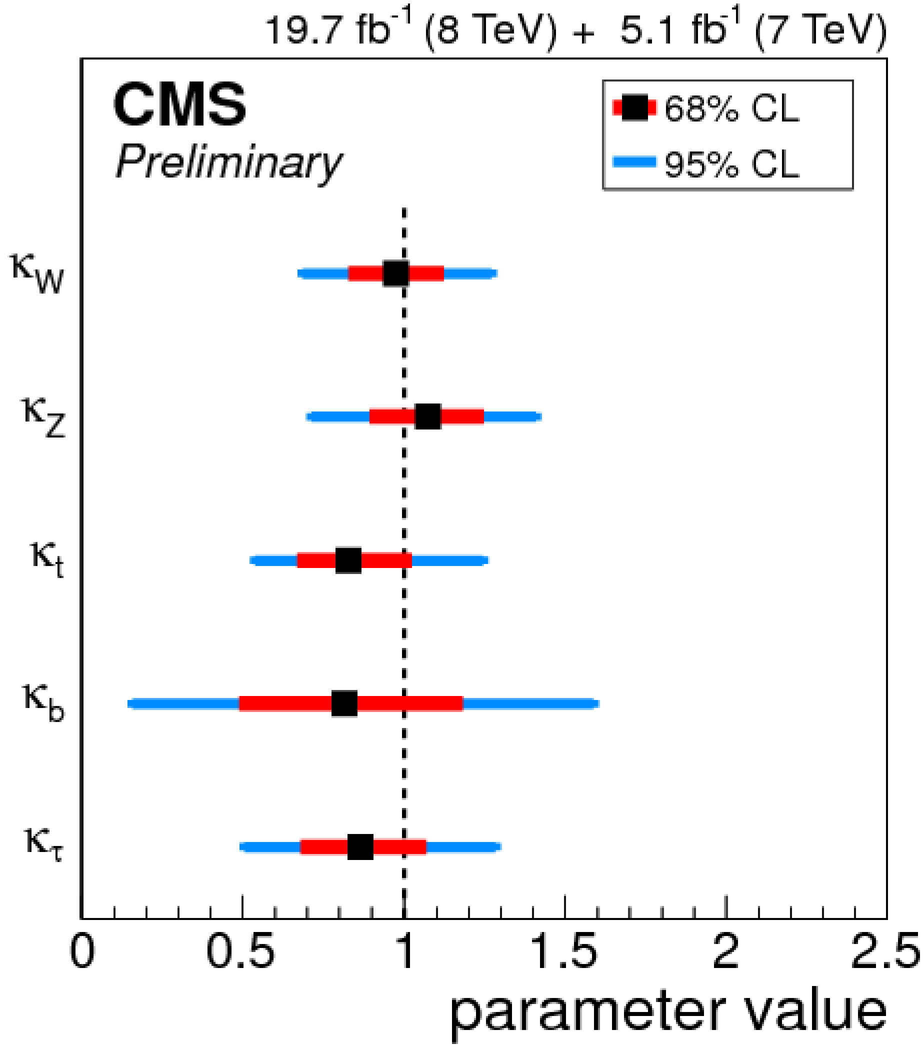
The simplest approach

- Scale all vector induced couplings by k_V and all fermionic ones by k_F

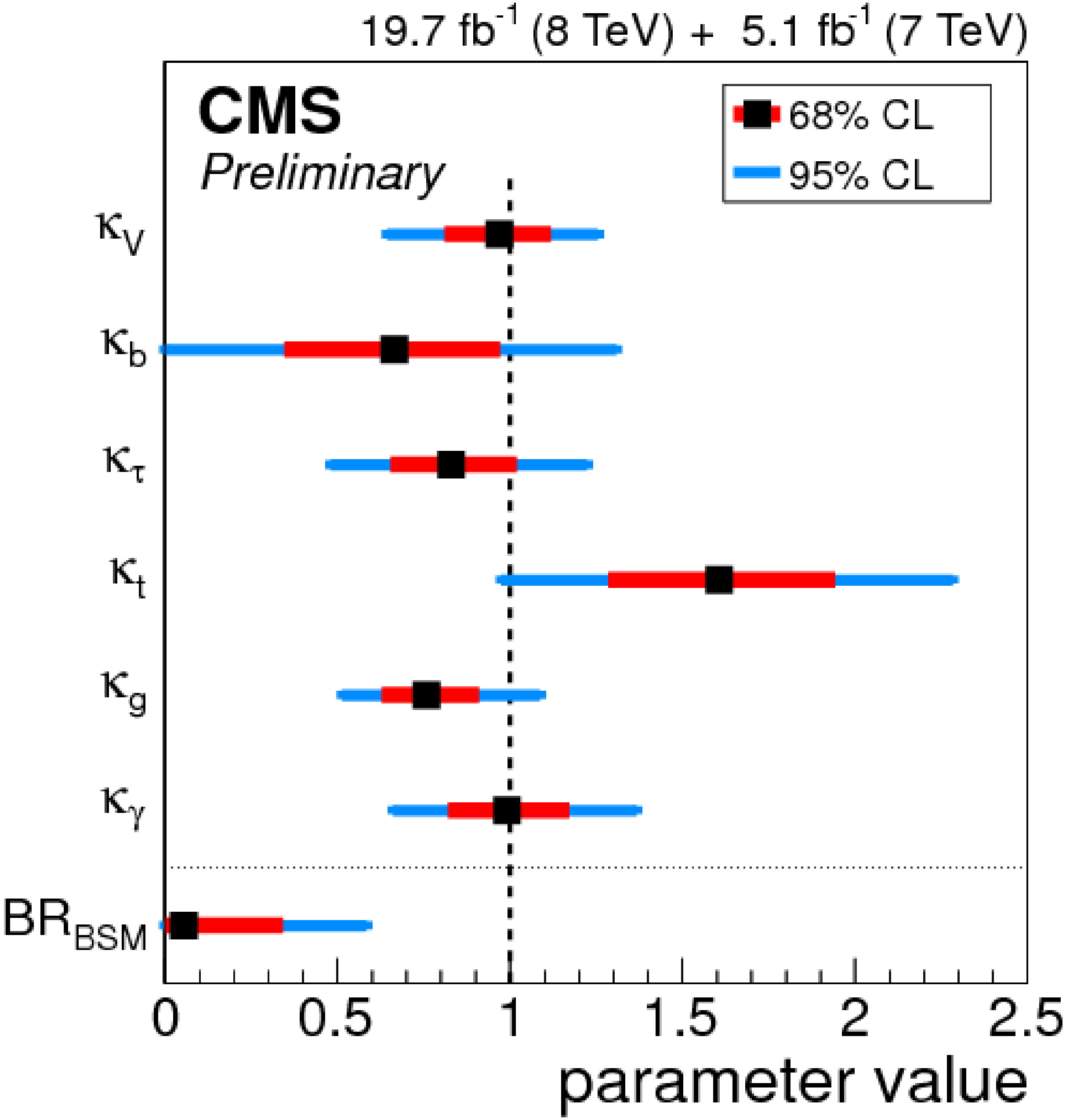


The semi full approach

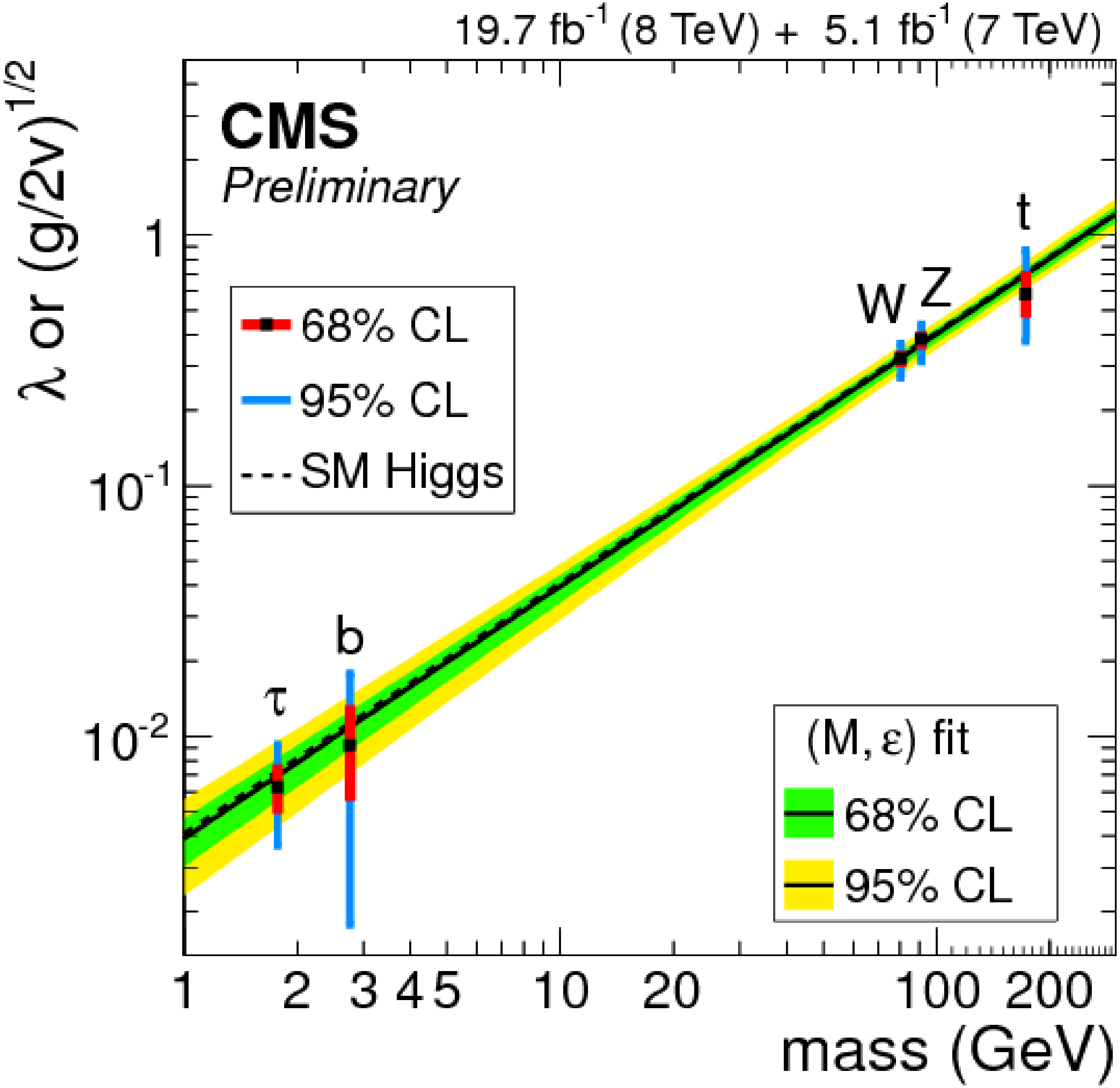
- Here the photon and gluon fusion loop have been resolved to the couplings of top and W



The almost full approach

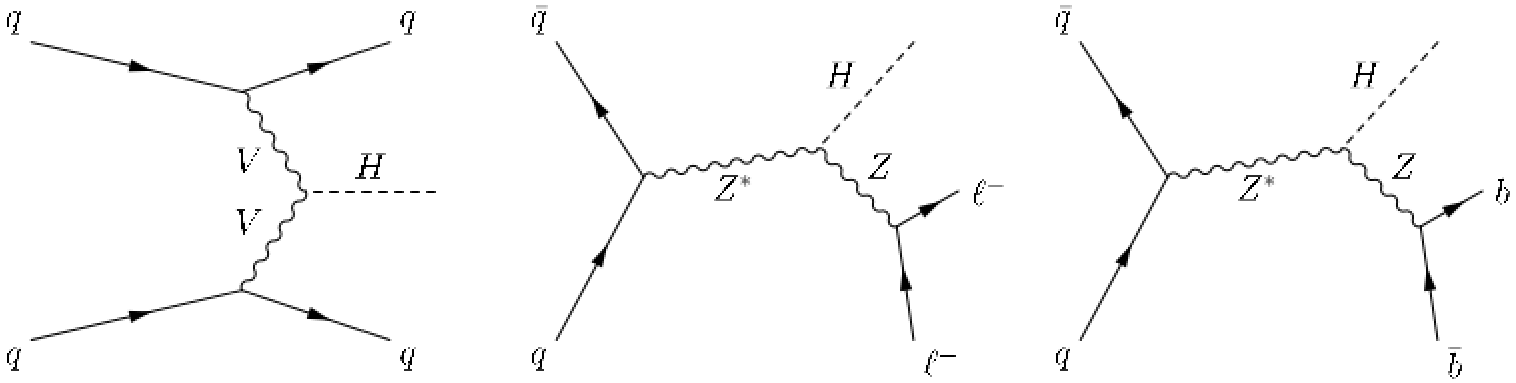


Couplings summary

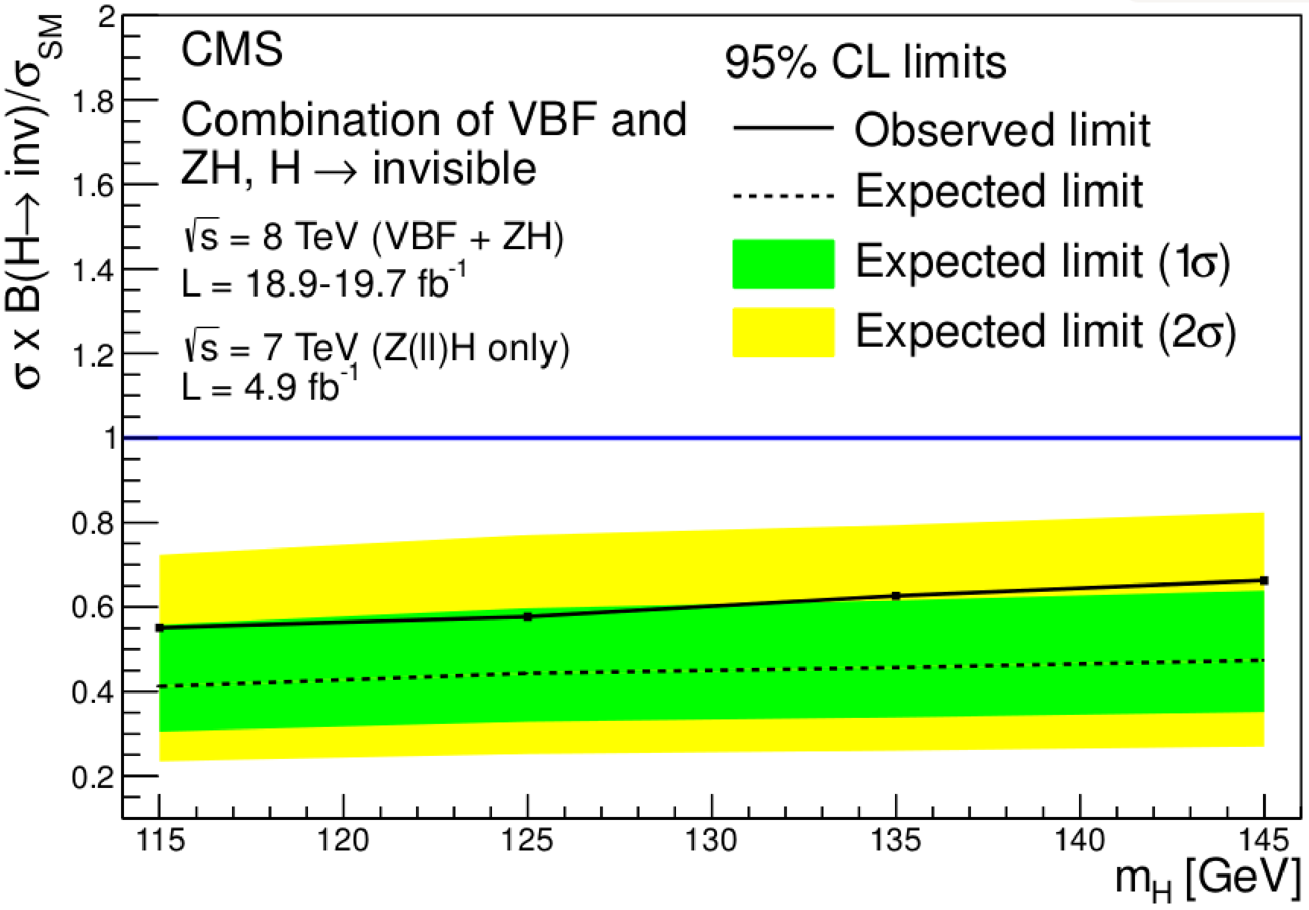


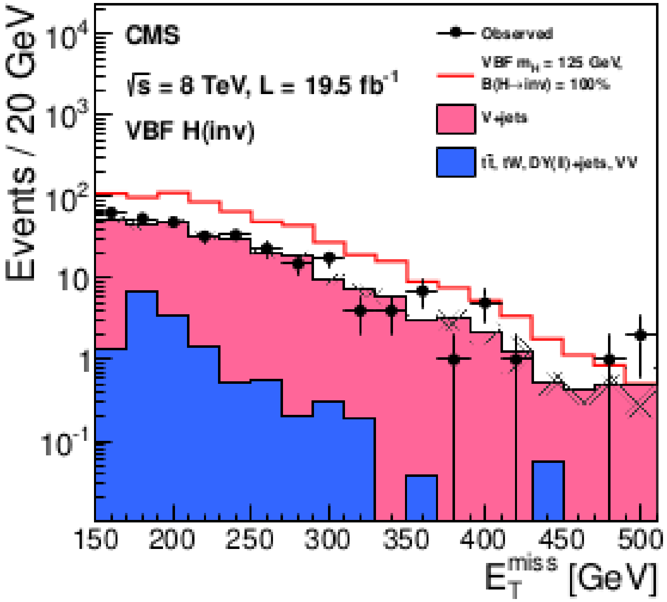
Higgs as a probe to dark matter

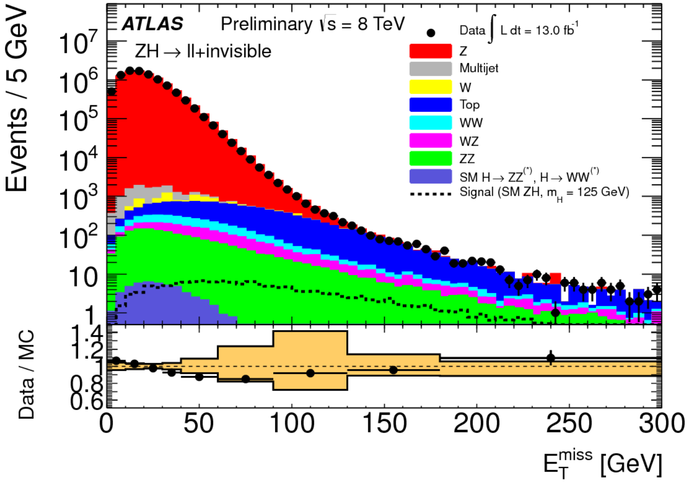
- Invisible decays of Higgs → New physics
  - Higgs → portal to the Dark Matter sector via invisible decays
- BR (H → inv) constrained by global couplings fit
- Direct searches also possible @ LHC
  - Mostly on ZH and VBF production modes



Invisible Decays: Results

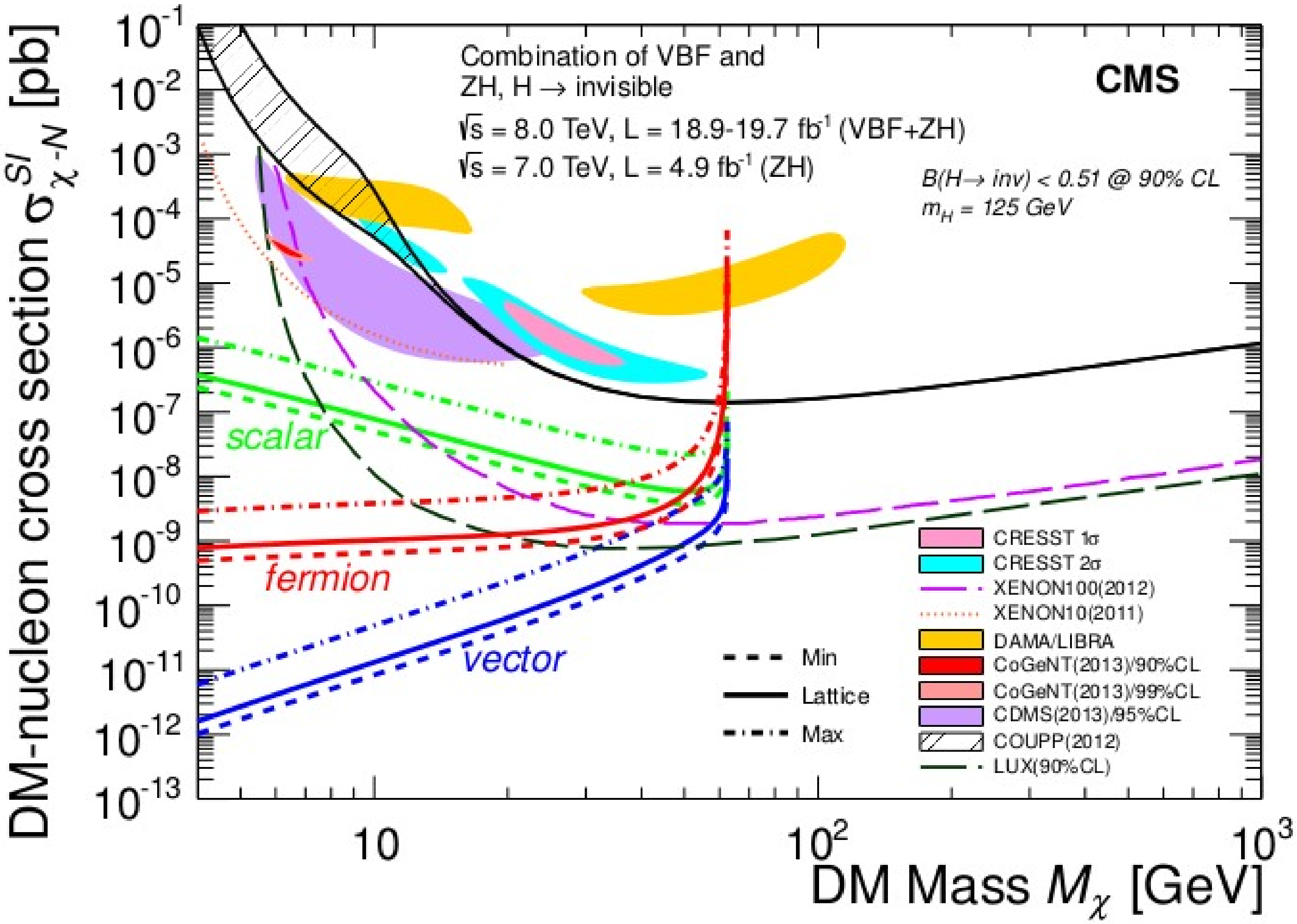






- CMS Expected limit of 44% in the invisible BR @ 95% CL
  - Assuming SM production cross section
- CMS Observed limit of 51%

Dark Matter Interactions



- BR results reinterpreted in the context of Higgs portal of DM interactions
- LHC Higgs search extends reach at low DM mass

Summary

- 5 major analyses to search for the Higgs
- Very sophisticated experimental methods to extract the signal
- Impressive results from both experiments showing signal compatible with the SM Higgs boson
- The measurement program just started
- Looking forward to the next LHC run