SILAFAE

XI SIMPOSIO LATINOAMERICANO DE FÍSICA DE ALTAS ENERGÍAS

Higgs To Two Photons at $\sqrt{s} = 13$ TeV using an integrated luminosity of 12.9 fb^{-1} in CMS

Manuel Olmedo

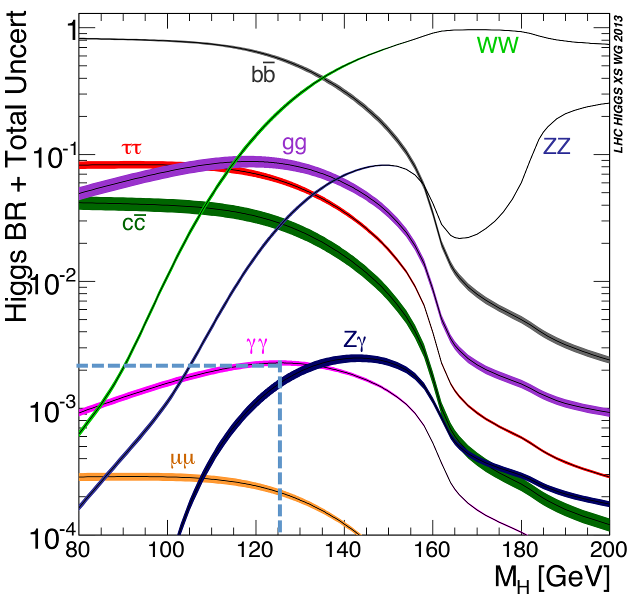
University of California Riverside on behalf of CMS

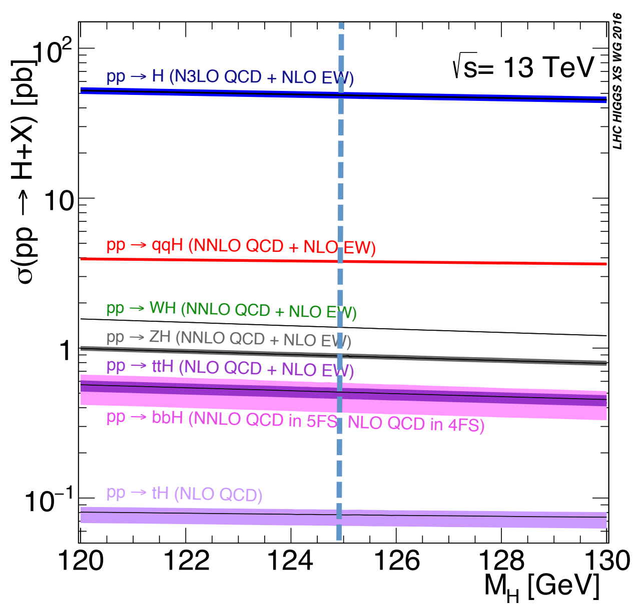
Outline Of Presentation

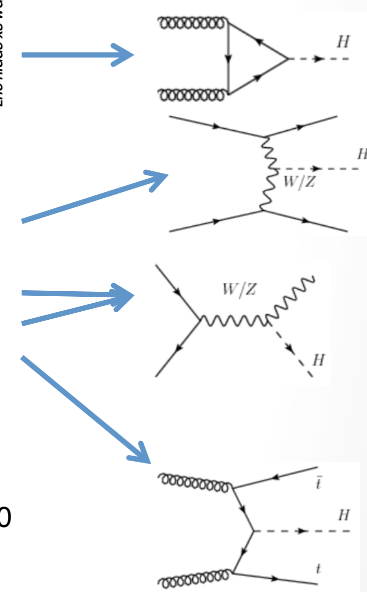
- Introduction
  - Object Selection.
  - Object optimization.
  - Event classification & interpretation.
  - Statistical methods.
  - Results.

Higgs to two photons channel

Higgs to two photon branching ratio in the order of < 1%, though it has one of the cleanest signal topologies so it is easier to reconstruct.

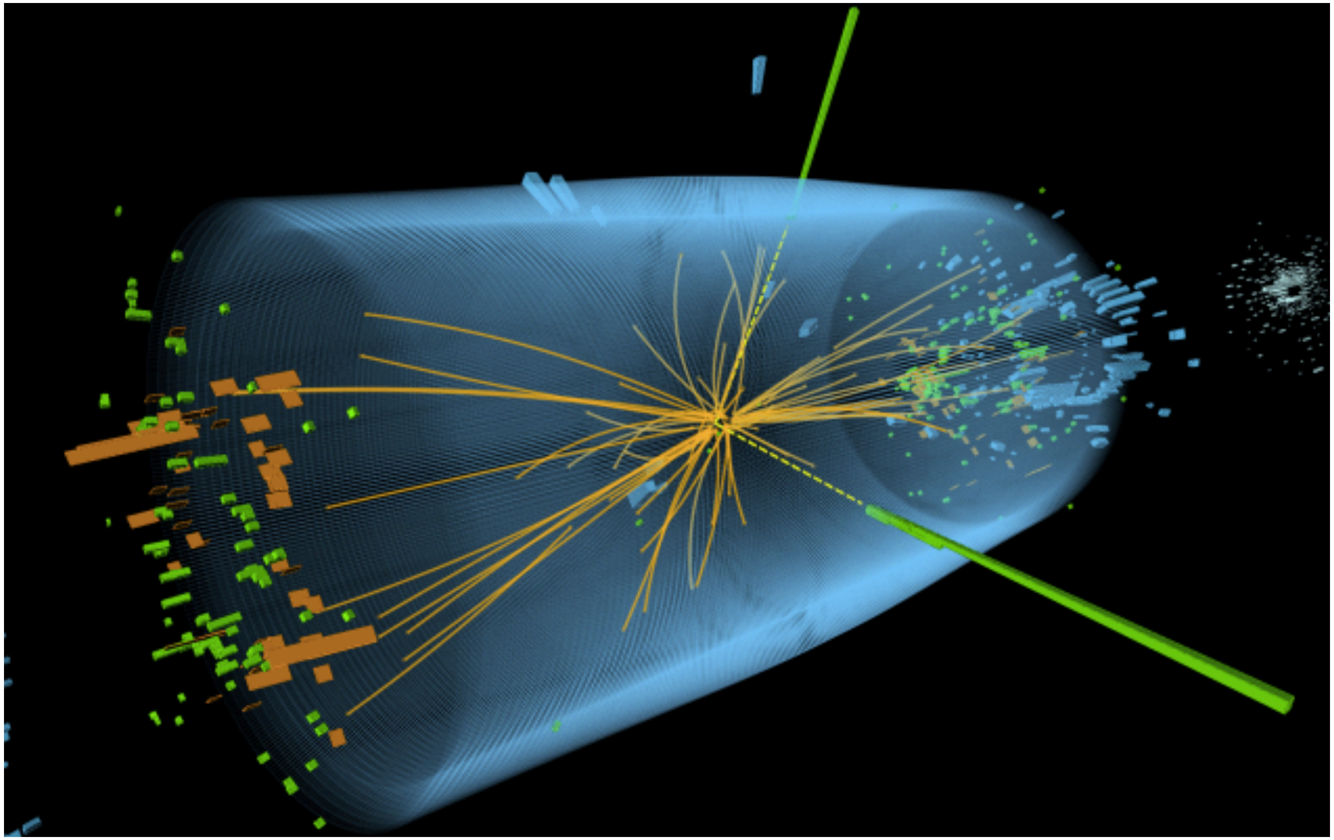


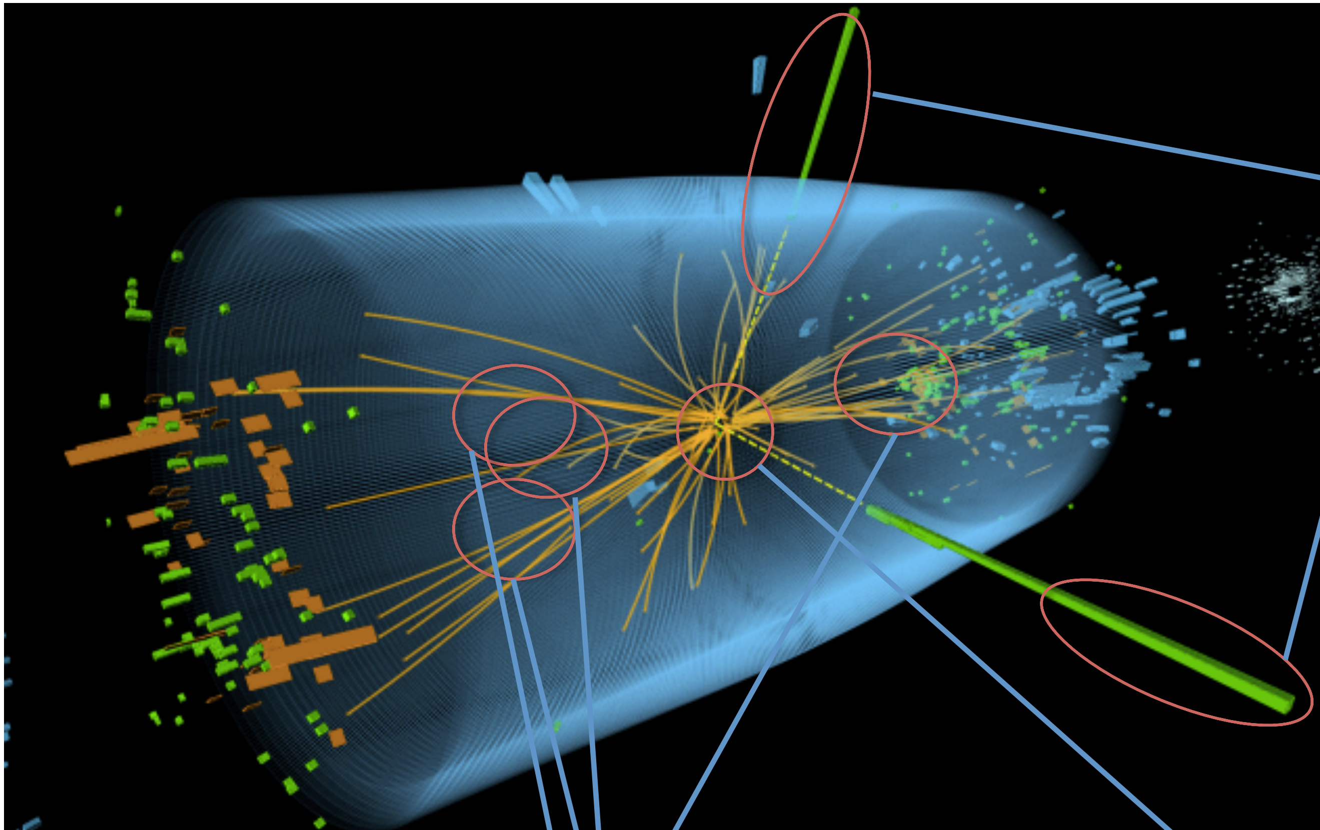




Simple invariant mass reconstruction :

$m_{\gamma\gamma}= \sqrt{2E_1E_2(1-\cos\theta)}$





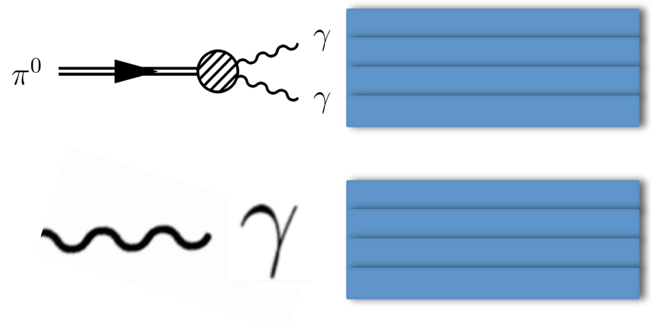
Photons!
(ECAL hits)

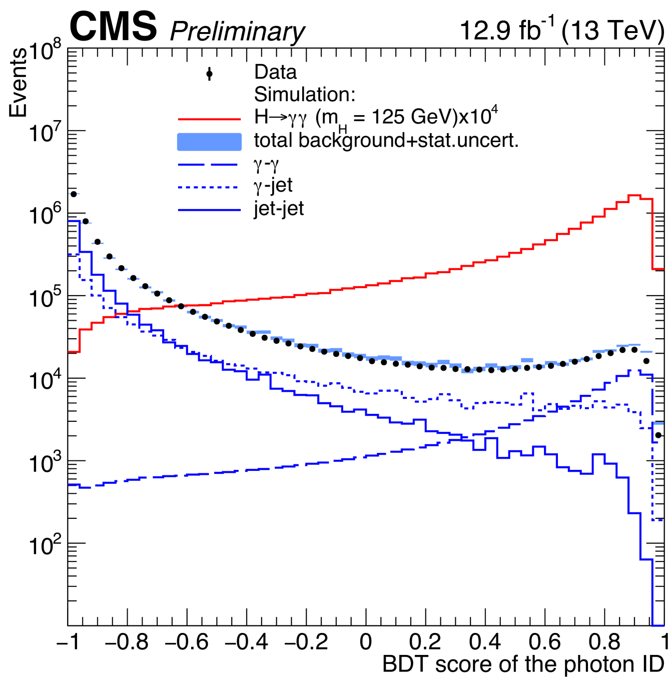
Event interpretation

Vertex!

Photon Identification

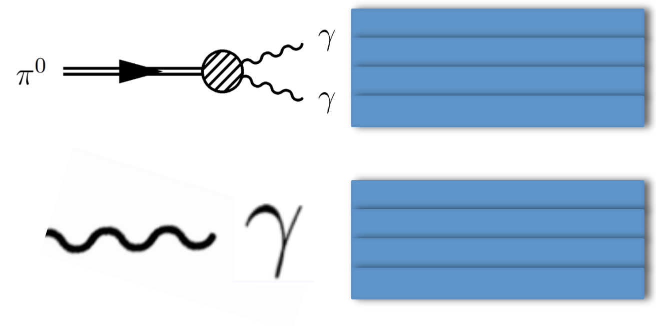
- Differentiating photons from background (fake photons)
  - These mainly come from pion decays with collimated photons.
  - Use of a Boosted Decision Tree (BDT) to differentiate the shapes of the shower in the ECAL.
  - BDT trained using photon enriched sample, γ + *jet*. Prompt photons used for signal, non-prompt (fake) photons used for background.
- Since it is a di-photon analysis, the photons also need to pass through a high level di-photon trigger.
  - Plus a pre-selection.
- So far γ_{good}= trigger + preselection + good photon id

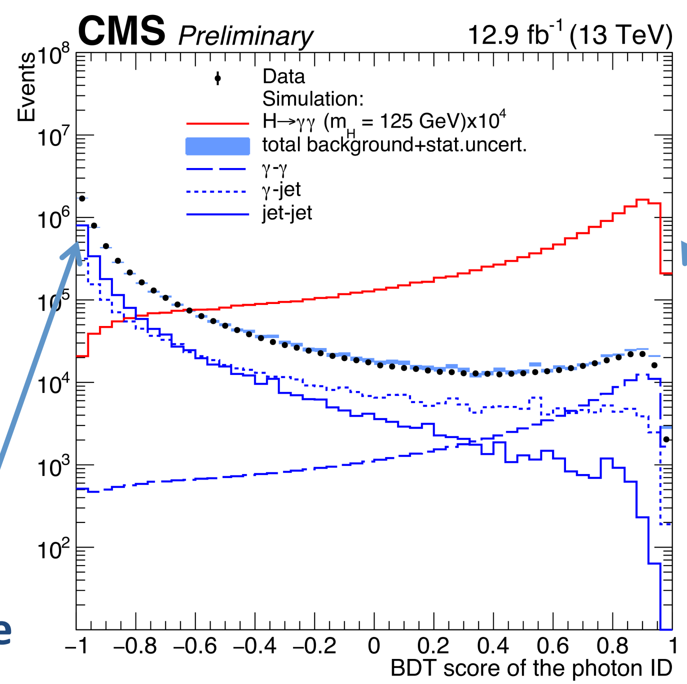




Photon Identification

- Differentiating photons from background (fake photons)
  - These mainly come from pion decays with collimated photons.
  - Use of a Boosted Decision Tree (BDT) to differentiate the shapes of the shower in the ECAL.
  - BDT trained using photon enriched sample, γ + *jet*. Prompt photons used for signal, non-prompt (fake) photons used for background.
- Since it is a di-photon analysis, the photons also need to pass through a high level di-photon trigger.
  - Plus a pre-selection.
- So far γ_{good}= trigger + preselection + good photon id



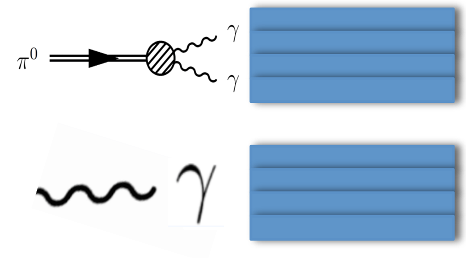


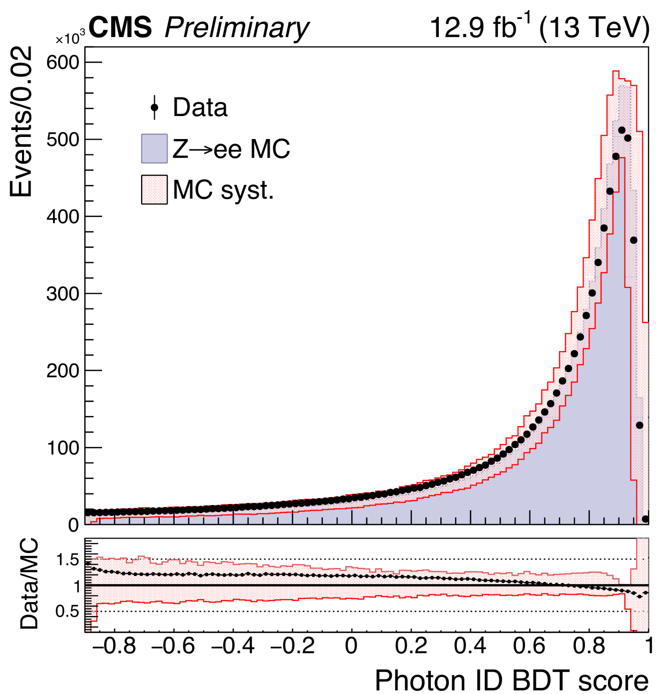
Background like

Signal like!

Photon Identification Validation

- Validation is done using z->ee events both in data & MC.
- The idea being that the kinematics of z->ee events are similar to h->γγ events (without the tracks).
- The MC events are taken from a Drell-Yan sample after properly selecting the events.

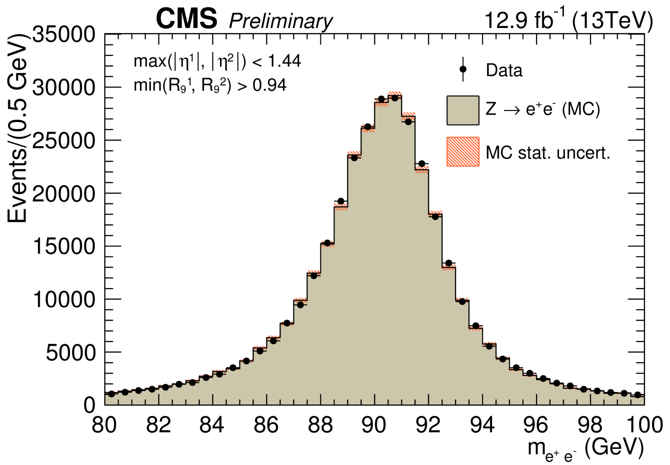




Photon Energy

$$E_{e,\gamma} = F_{e,\gamma}(\eta) \cdot \sum_{cluster\ crystals} G(GeV/ADC) \cdot S_i(T,t) \cdot c_i \cdot A_i$$

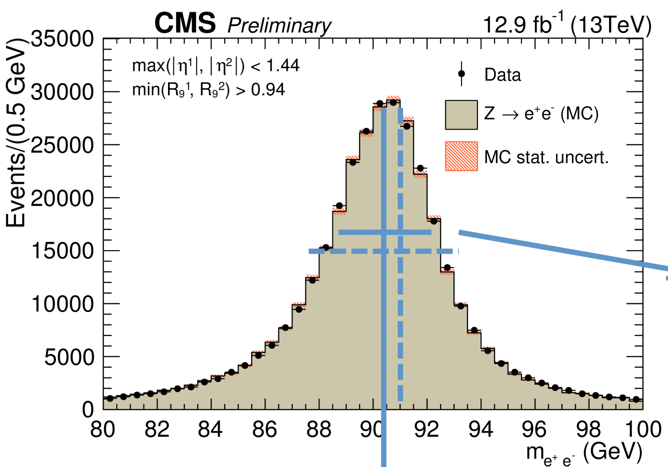
Signal crystal amplitude



Energy scale (energy/ADC counts)
Here the Z boson invariant mass peak in data is fitted to get scale and resolution.

There are higher level corrections to the energy of a photon. To obtain these corrections the energy response of a single photon is essentially estimated/predicted using multi-variate techniques (regression). The corrections are then derived from these function.

Photon Energy



Δm

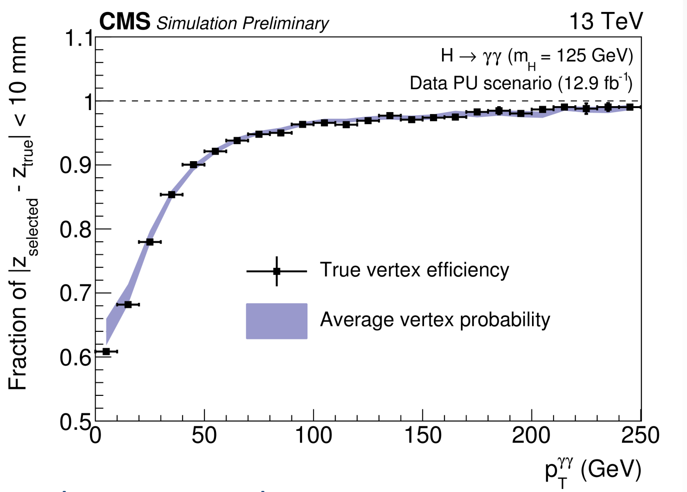
Δσ

Even after the energy corrections are applied, there are still some left over data/MC differences. These corrections are then applied to the data to match the simulation.

These corrections vary in detector region and run number.

Vertex Identification

- Also a BDT, taking advantage of the recoil and conversion information.
- Choosing the correct vertex as opposed to the primary vertex benefits the creation of the di-photon object.



$$sumpt2 = \sum_i |\vec{p}_T^{\,i}|^2$$

Primary vertex is the vertex with maximum sumpt2, tends to be the 0th vertex.

$$ptbal = -\sum_i (\vec{p}_T^{\,i} \cdot \frac{\vec{p}_T^{\,\gamma\gamma}}{|\vec{p}_T^{\,\gamma\gamma}|})$$

$$ptasym = (|\sum_i \vec{p}_T^{\,i}| - p_T^{\gamma\gamma})/(|\sum_i \vec{p}_T^{\,i}| + p_T^{\gamma\gamma})$$

Recoil variables

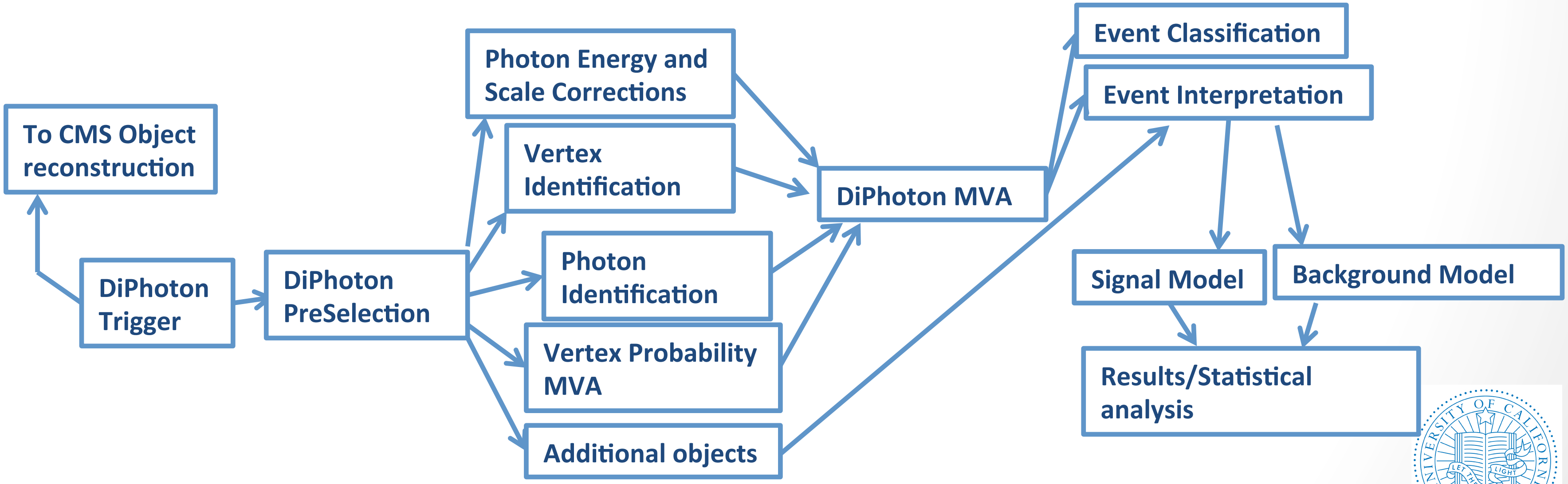
$$Pull_{\mathrm{conv}} = \frac{|z_{\mathrm{PV}} - z_{\mathrm{PV}}^{\mathrm{conv}}|}{\sigma_{\mathrm{conv}}}$$

Additional variable if a conversion is present

Validation of this technique in data is done using Z->μμ events, here you get two nice tracks. The photon is emulated by removing the tracks of the muon.

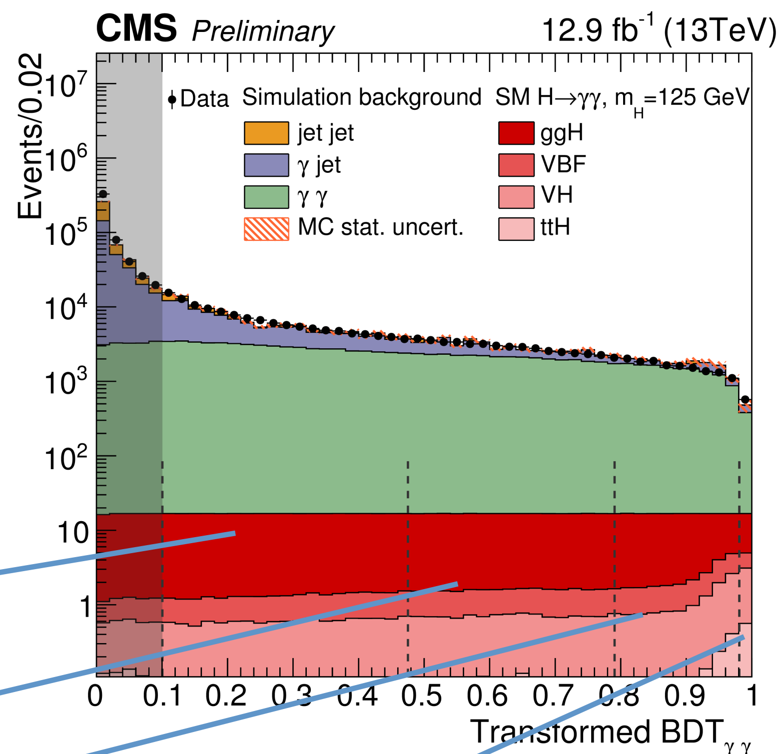
Analysis Strategy

- With a clear definition and optimization of our basic objects we can now begin to look and interpret di-photon events in data!
- The main idea is to divide the data into categories with good/bad S/B and mass resolution.
  - To do this one more BDT is used. The di-photon BDT has as input variables di-photon kinematics and the output from the vertex BDT and the photon BDT plus and additional one which I won't mention here.
  - Beyond this, if the event has good reconstructed objects such as leptons,jets or missing energy the event can be classified according to it's production process + the particular decay -> hadronic or leptonic.



Di-Photon BDT

- Di-Photon events are classified using a BDT. Di-photon events with a good mass resolution and good di-photon kinematics get a score closer to 1.
- By optimizing S/B regions in the BDT one can define regions/categories of different S/B ranges.



Untagged 0 :
Low resolution and low S/B

Untagged 1

Untagged 2

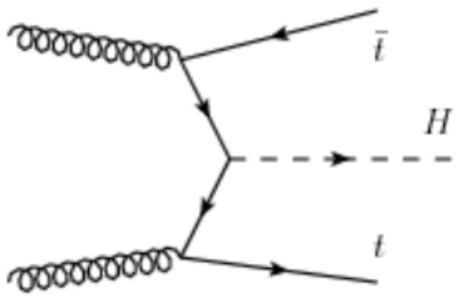
Untagged 3:
High resolution and high S/B

The BDT is trained with signal (ggf,vbf,vh,tth) and background (γγ,γ*j*,*jj*)

Event interpretation

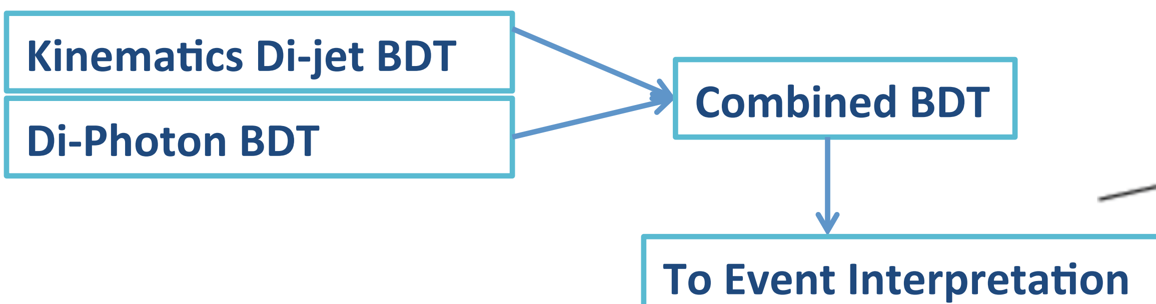
- To identify different production processes and gain sensitivity, additional objects are identified within the event.
  - Leptons, Jets, missing energy.

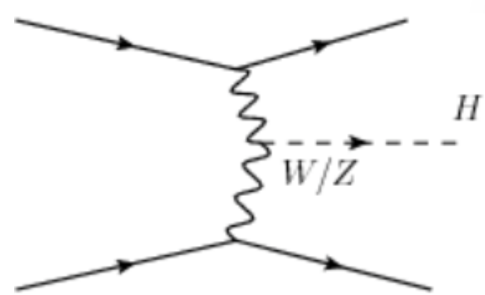
- The TTH process is split into leptonic decays and hadronic decays.
  - Hadronic decay selection:
    - At least 1 reconstructed b-jet .
    - A minimum of 5 jets in the event.
  - Leptonic decays selection:
    - A minimum of one good lepton.
    - At least 1 b jet in the event.
    - At least 3 jets in the event



Event interpretation

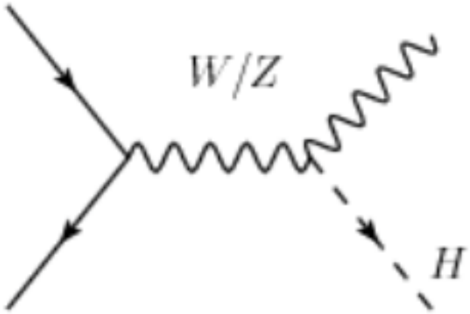
- VBF has a more complicated procedure.
  - Two new BDTs are used to identify an event as vbf.
  - The di-jet kinematics BDT and the combine di-jet and di-photon BDT (takes di-photon and and di-jet BDT output as input).
    - The di-jet kinematics BDT is used to identify a good di-jet pair.
    - The combined BDT is used to further discriminate the di-jet signal from background.
  - The di-jet kinematics BDT is trained using MC VBF signal sample @ 125 GeV and for background the standard processes are used.



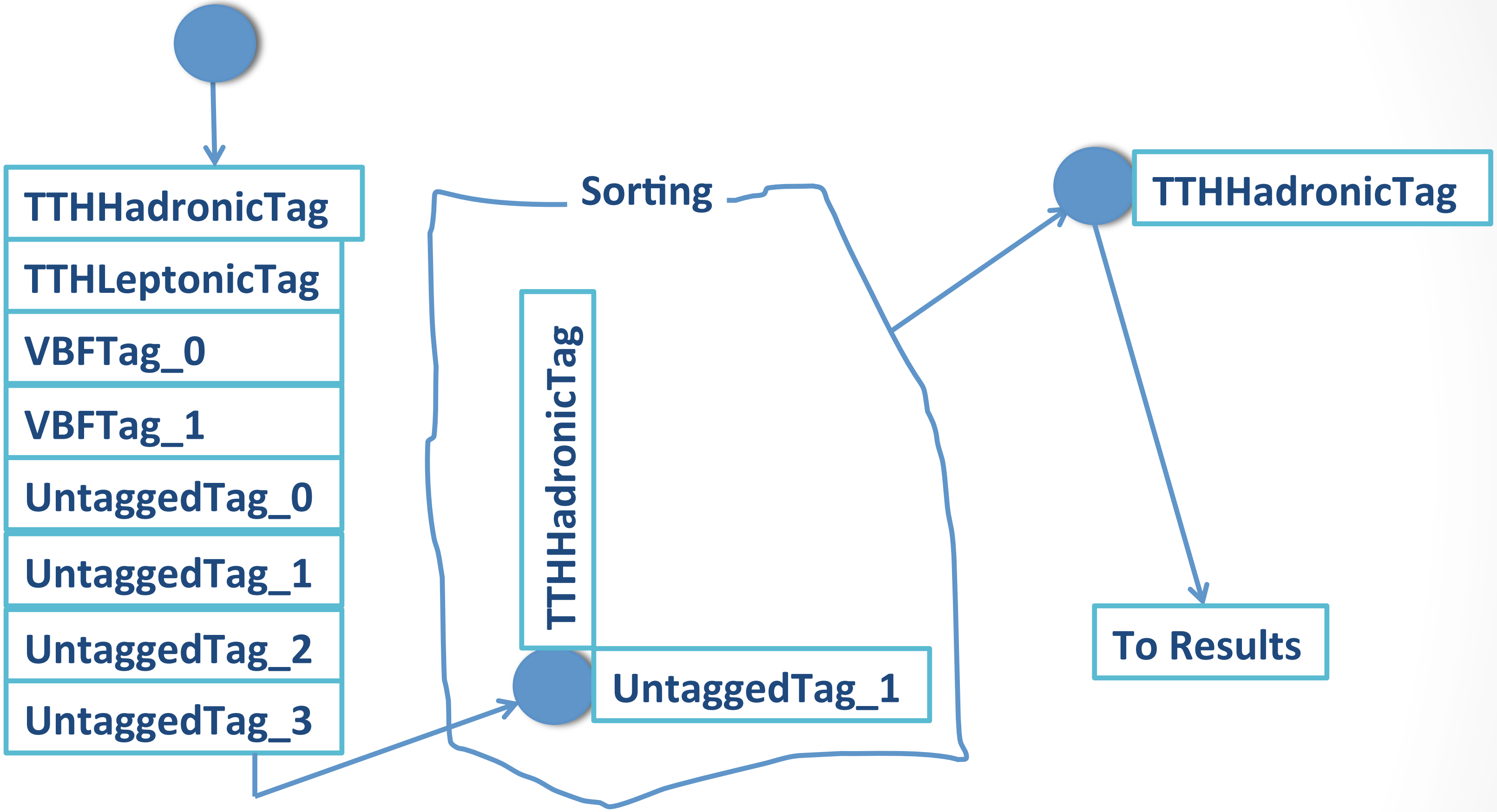


Event interpretation

- No Associated production (Higgstrahlung) yet…

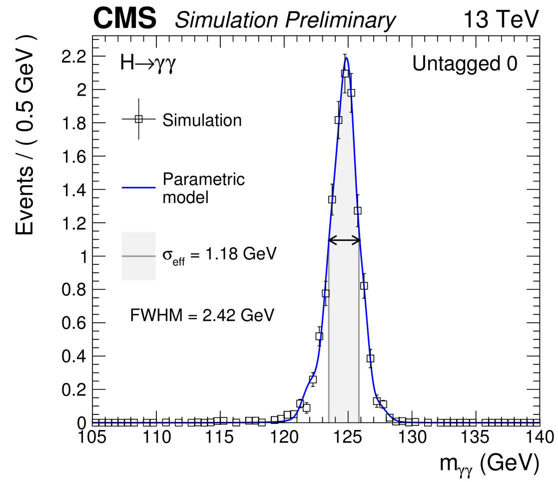


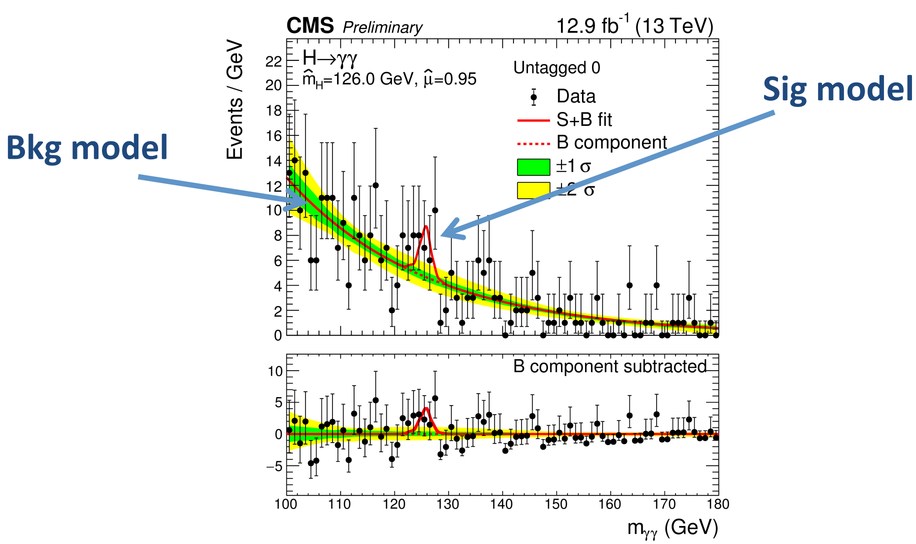
Event interpretation

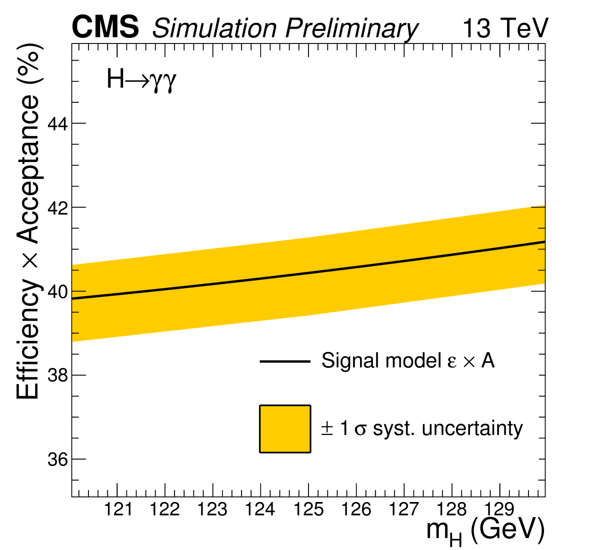


Signal And Background model

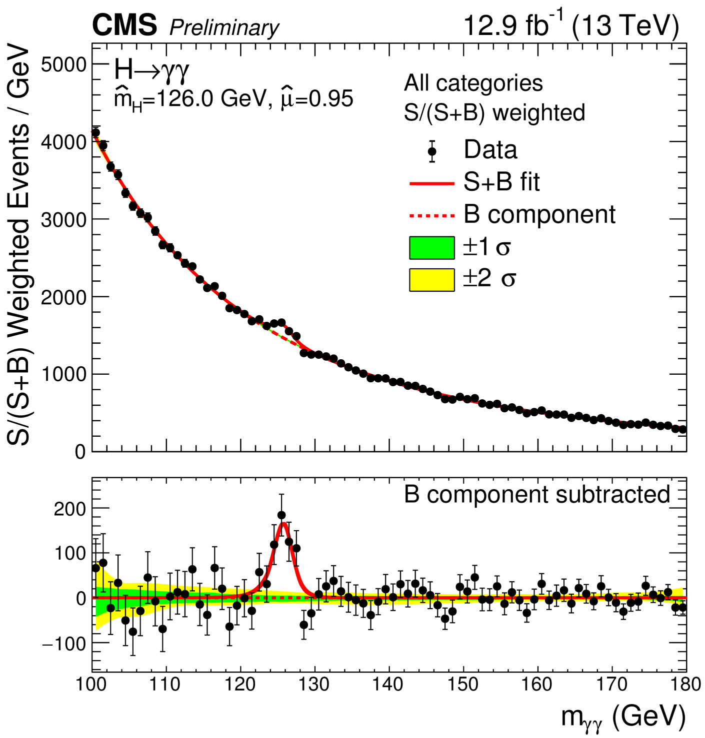
- Fully parametric signal model.
  - Take selected events from signal MC and build a gaussian function as the signal (μ,σ).
- Data driven background model .
  - Use selected data and choose the best fit from a family of functions. The choice of function is treated as a nuisance parameter.



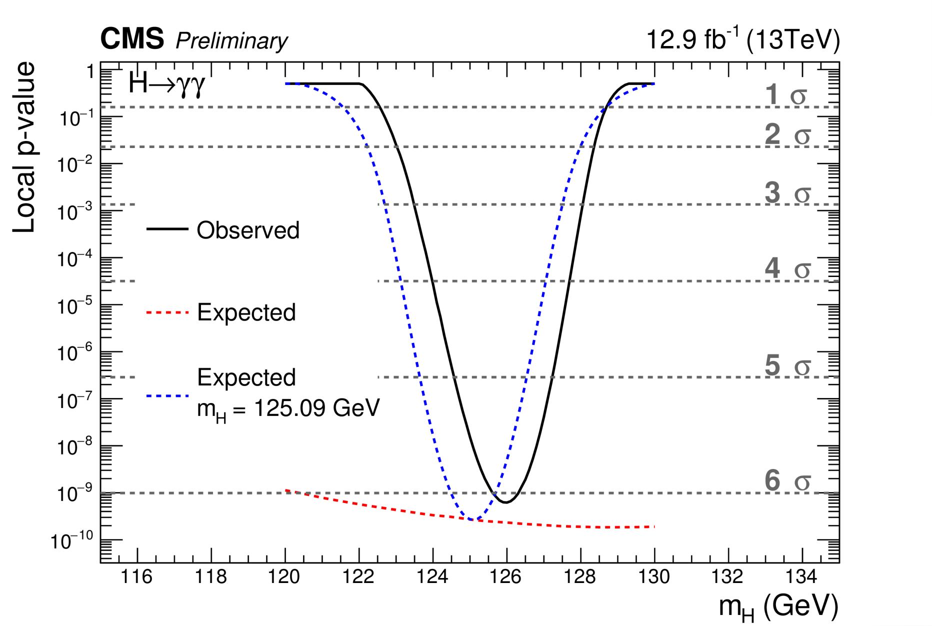




Results

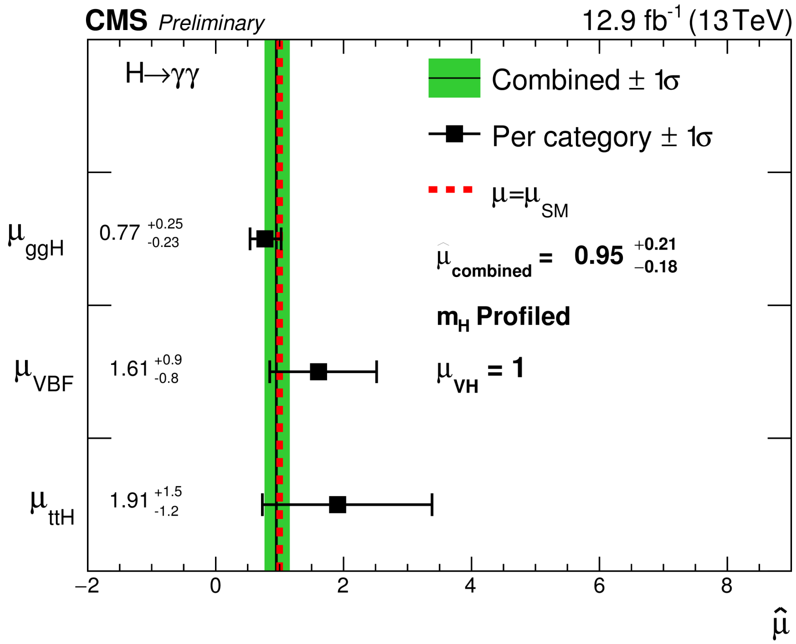


Di-Photon invariant mass spectrum
Famous bump reappeared !

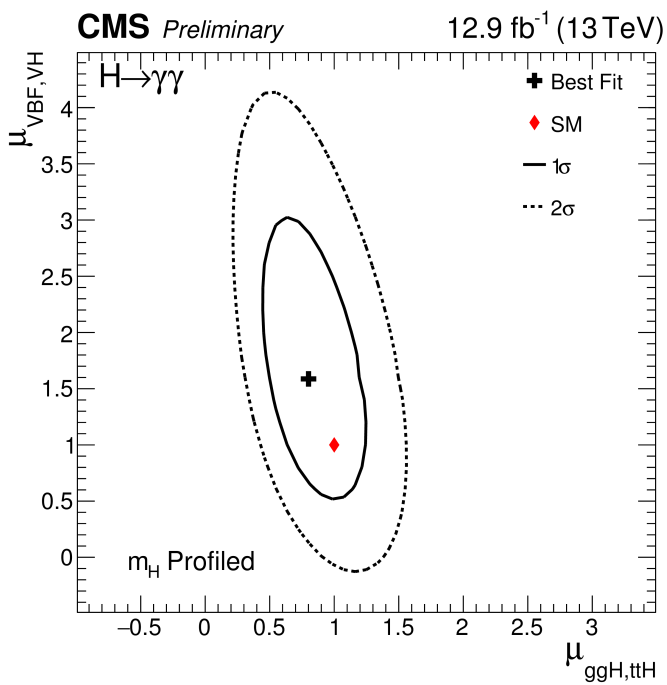


P-value scan, significance at m_H= 125.09 GeV is 5.6σ. Highest observed is 6.1σ at 126 GeV.

Results



Relative signal strengths, all within SM prediction!

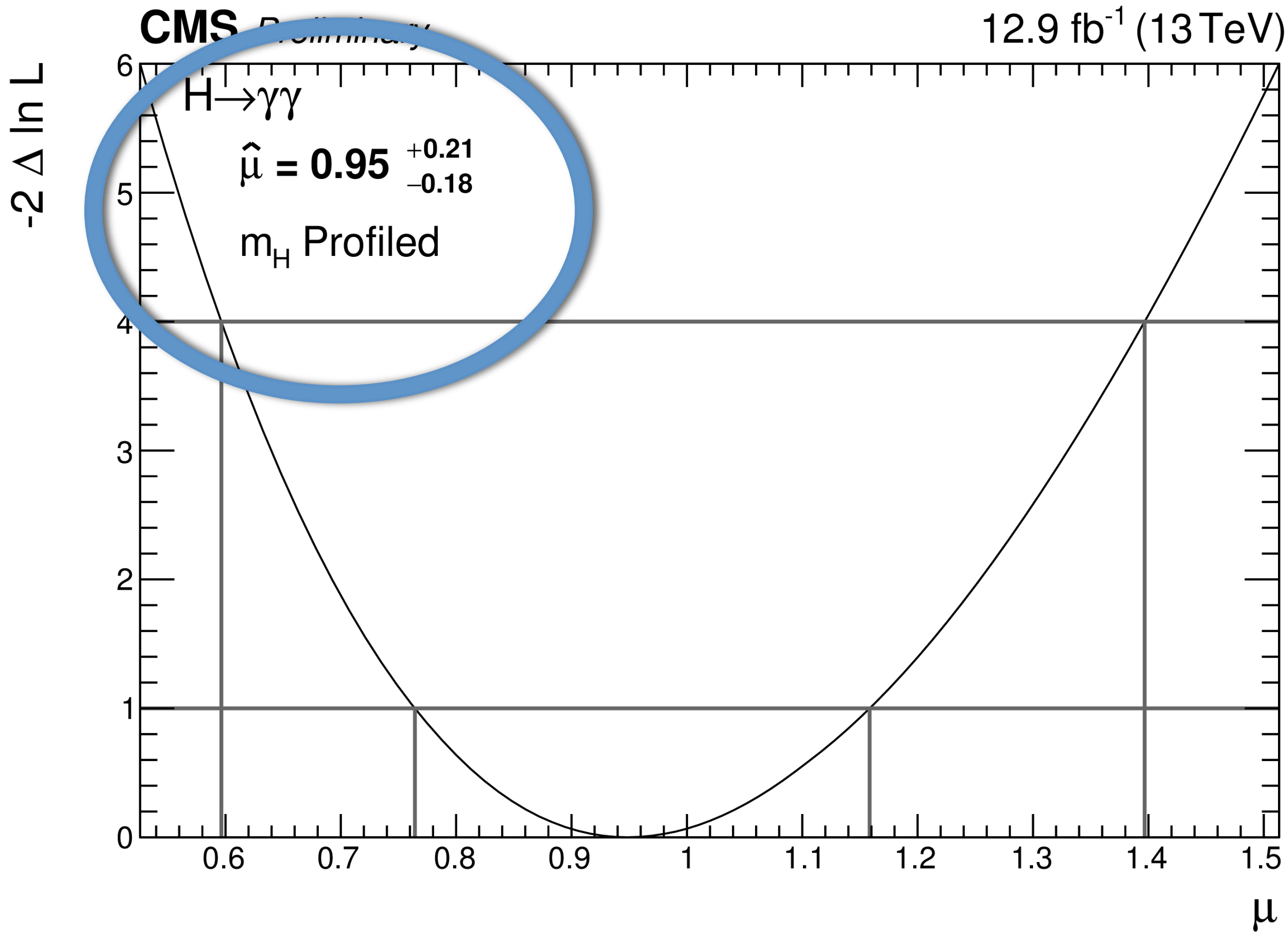


Best signal strength split between fermionic and bosonic components

$\mu_{ggH,ttH} = .80^{+0.14}_{-0.18}$

$\mu_{VBF,VH} = 1.59^{+0.73}_{-0.45}$

Results


CMS Preliminary
12.9 fb-1 (13 TeV)
H→γγ
μ̂ = 0.95 +0.21 −0.18
mH Profiled
-2 Δ ln L
μ
0
1
2
3
4
5
6
0.6
0.7
0.8
0.9
1
1.1
1.2
1.3
1.4
1.5

Conclusion

- **First results with 2016 data have been shown.**
  - **Reliable mass measurement still in the works (need to understand 13TeV detector conditions better, add VH channel)**
  - **Machinery has been built and centralized, measurement turn around time has been reduced.**
- **Centralization within CMS is imperative as the LHC moves into HL stage.**
- **This collaboration always welcomes a fresh pair of eyes!**
  - **If you are interested in contributing contact :** *Seth Zenz (Imperial College), Federico Ferrie (SACLAY) Martina Malberti (Milano-Biccoca).*

References

- **CMS Collaboration, "Updated measurements of Higgs boson production in the diphoton decay channel at sqrt(s) = 13 TeV in pp collisions at CMS." CMS-PAS-HIG-16-020**

Back up

11.3 Kinematic Dijet MVA

The kinematic Dijet MVA is built using the following variables:

- the transverse momenta of the leading and subleading photons divided by the invariant mass of the di-photon candidate: $p_{\mathrm{T}}^{\gamma_1}/m_{\gamma\gamma}$ and $p_{\mathrm{T}}^{\gamma_2}/m_{\gamma\gamma}$;
- the transverse momenta of the leading and subleading jets: $p_{\mathrm{T}}^{j_1}$ and $p_{\mathrm{T}}^{j_2}$;
- the di-jet invariant mass, $m_{j_1 j_2}$;
- the difference in pseudorapidity between the two jets, $\Delta\eta_{j_1 j_2}$;
- the so-called *Zeppenfeld* variable [17], defined as $\eta^* = \eta|_{obs} - \frac{\eta(j_1)+\eta(j_2)}{2}'$, where $\eta|_{obs} = \eta(\gamma_1 + \gamma_2)$;
- the difference in azimuthal angle between the dijet and the diphoton, $\Delta\phi_{(j_1 j_2, \gamma\gamma)}$.

Back up

11.4 Combined MVA and Categorization Scheme

The Combined Dijet MVA is built using the Kinematic Dijet MVA, the Diphoton MVA, and the $p_{\mathrm{T}}^{\gamma_1\gamma_2}/m_{\gamma\gamma}$ as inputs. The purposes of this variable is to maximally discriminate the VBF dijet signal from background utilizing information from all relevant objects tagged in the event. $p_{\mathrm{T}}^{\gamma_1\gamma_2}/m_{\gamma\gamma}$ is included as an input because of its significant correlation to both the Dijet MVA and the Diphoton MVA.

The information is combined by means of a BDTG algorithm. The BDTG is trained on simulated events: a VBF $\mathrm{H}\to\gamma\gamma$ sample with m_H =125 GeV is used as signal. All standard backgrounds are used in the training. Background rejection suffers when gluon fusion is used as a training background and it is therefore not used in the Combined MVA training. Figure 65 shows the BDT output for VBF signal, gluon fusion, the background simulation and data.

A tight (VBFTag0) and loose (VBFTag1) category are created by two selection requirements on the Combined MVA. These categories are optimized by first choosing the requirement that maximizes the $S/\sqrt{S+B}$ of the tight bin, then excluding those events and repeating the procedure for the loose bin. Of signal events accepted by the tight category, 72% are VBF and 27% luon-fusion; for the loose category these two signals each make up 49% of the total accepted signal. (The remaining $\sim$ 1% in each case is from the other production modes.)

Back up

The following variables are used as input to the event classifier:

- the transverse momenta for both photons, rescaled for the diphoton mass, $p_{\mathrm{T}}^{1,(2)}/m_{\gamma\gamma}$;
- the pseudorapidities of both photons, $\eta^{1(2)}$;
- the cosine of the angle between the two photons in the transverse plane, $\cos(\Delta\phi)$;
- the identification BDT score for both photons;
- the per-event relative mass resolution estimate, under the hypothesis that the mass has been reconstructed using the correct primary vertex (σ_{rv});
- the per-event relative mass resolution estimate, under the hypothesis that the mass has been reconstructed using an incorrect primary vertex (σ_{wv});
- the per-event probability estimate that the correct primary vertex has been used to reconstruct the mass, based on the event-level vertex selection MVA as described in Section in 7.2.

Back up

8.1.2 Photon Id MVA input variables

The following variables are used as input to the BDT and described in detail in Section 4.0.1:

$\sigma_{i\eta i\eta}$, $cov_{i\eta i\Phi}$, E_{2x2}/E_{5x5}, R_9, σ_η, σ_ϕ, Preshower σ_{RR}, PF Photon ISO, PF Charged ISO (worst vertex), PF Charged ISO (selected vertex), ρ, supercluster η, and supercluster E_{RAW}.