Backgrounds and uncertainties in ttH, H→$\gamma\gamma$

Francesco Micheli

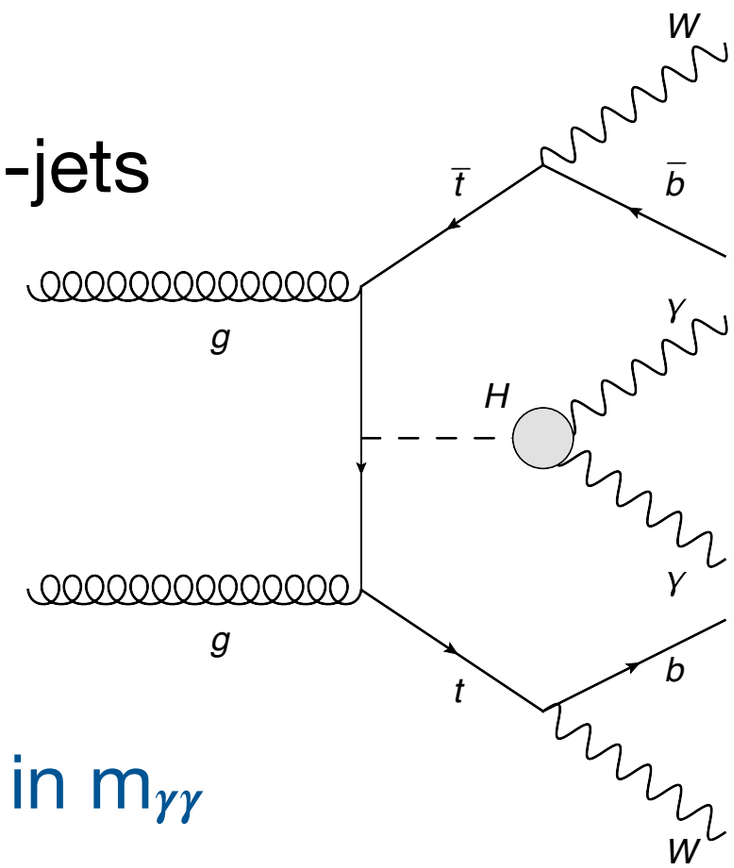
SAPIENZA
UNIVERSITÀ DI ROMA

INFN

CMS
Compact Muon Solenoid

ttH in diphoton decay channel

- ttH, H→$\gamma\gamma$ has a striking signature:
  - Two **photons** (really energetic), **jets** and **b-jets**
- Signal is small for Run1:
  - σ x BR x 20 fb^{-1} ~ 5 evts
- But H→$\gamma\gamma$ is a very sensitive channel:
  - Background is small
  - Full reconstruction of H and peak search in $m_{\gamma\gamma}$



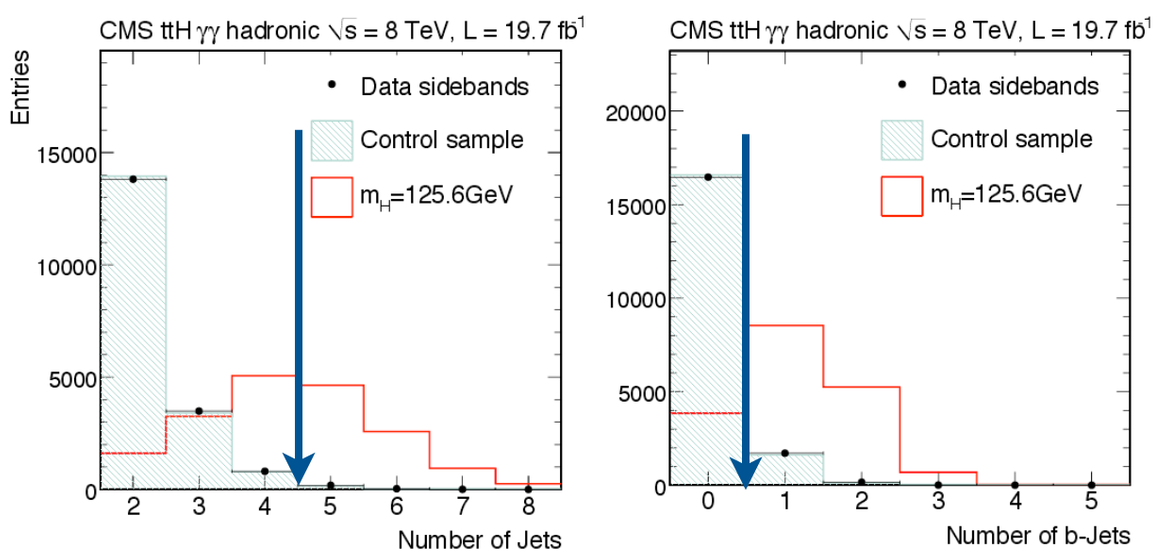
- Backgrounds: everything with photons and jets
  - γ+jets, $\gamma\gamma$+jets, tt+jets, ttγ, tt$\gamma\gamma$, tγ, t$\gamma\gamma$...
- These backgrounds are **not peaking** in $m_{\gamma\gamma}$

Run1 Analysis

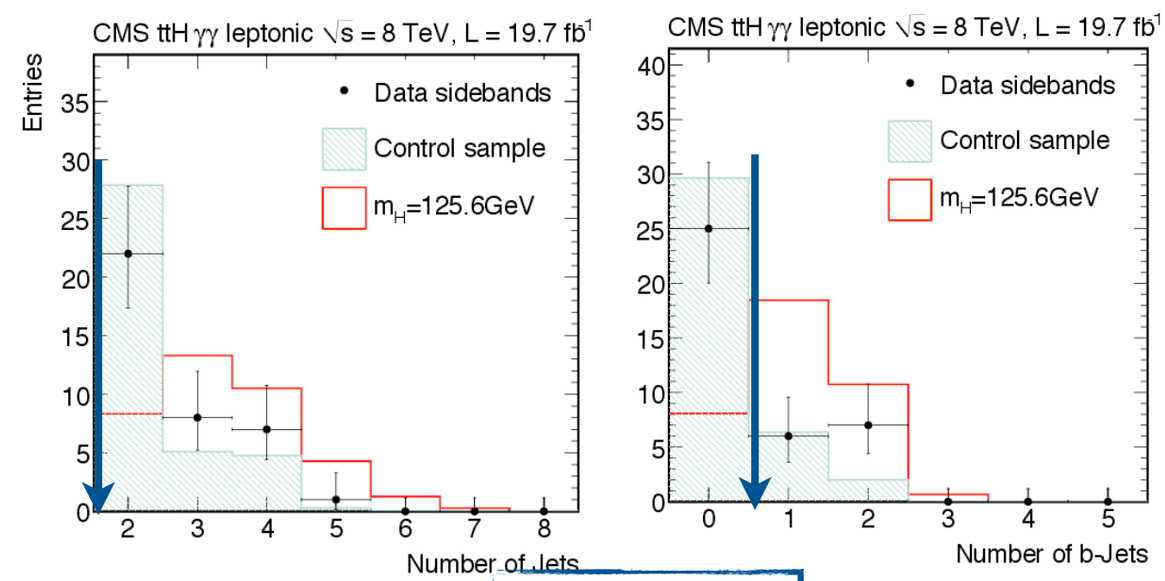
- Analysis was quite simple:
  - We wanted to be **efficient on signal, given** low statistic
  - No possibility of advanced techniques, just **few cuts on basic quantities**
- Photon identification and reconstruction inherited from std H→$\gamma\gamma$:
  - Harder cuts on photon p_T and diphoton MVA
- Cut-based analysis optimized on a **data control sample**

No use of Bkg MC

- Two categories depending on the presence of lepton: **Leptonic** / **Multijet**
- Cuts on **#jets** and **#b-jets** different for the two categories



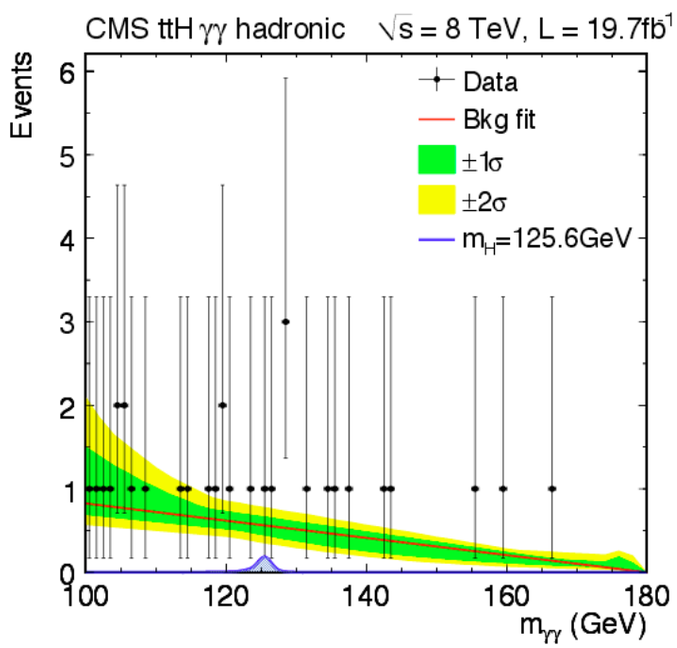
Multijet

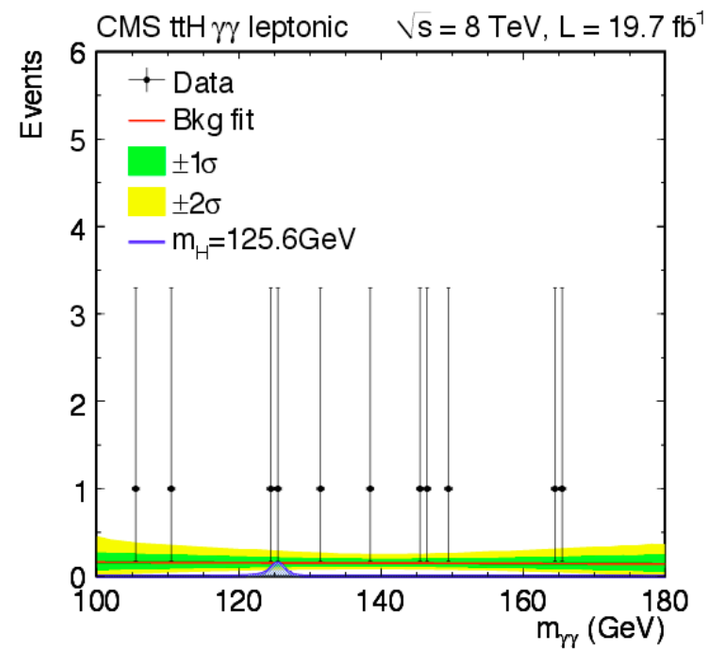


Leptonic

Results

- The main advantage of H→$\gamma\gamma$ is $m_{\gamma\gamma}$:
  - Background and signal estimated fitting $m_{\gamma\gamma}$ distribution
  - Signal model from MC





No use of Bkg MC

- Yields are low:
  - Enough to fit
  - Low impact of syst (stat dominated) ~ 1%

- Background MC in H→$\gamma\gamma$ never used:
  - **Optimization** on data control sample
  - **Bkg estimation** from sidebands

Fine for low-stat Run1 signal.
Is this ok for the future? Could be, but...

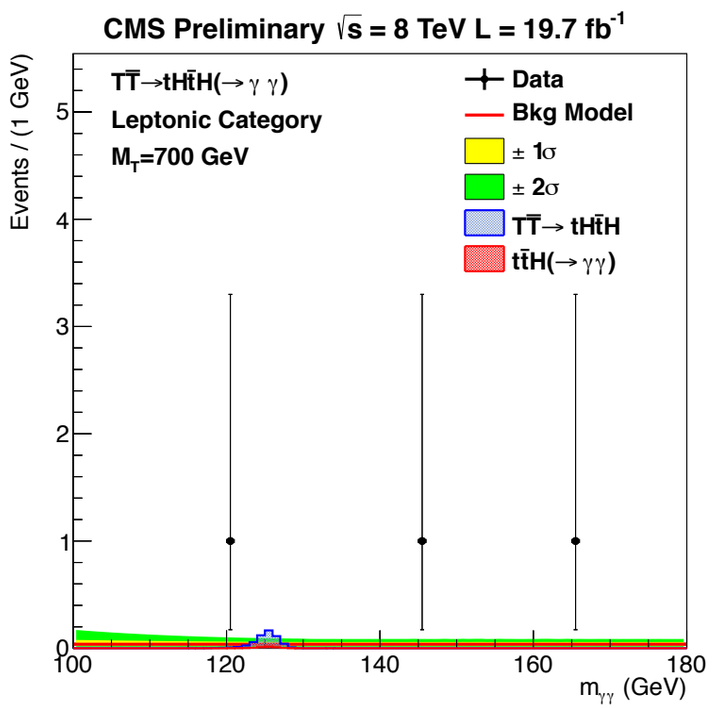
Background MC simulations

- In Leptonic category O(5-10) evts. They are enough but huge uncertainty.
  - Factor 2 fluctuation could be plausible
- We would like to be sure of the composition of this background
  - Hard cuts on jets, what remains? tt$\gamma\gamma$, t$\gamma\gamma$....
- With this low statistic, no advanced reconstruction techniques used
  - In the future lots of improvements possible:
    - Objects full reconstruction
    - MVA techniques for analysis

For all these studies MC crucial

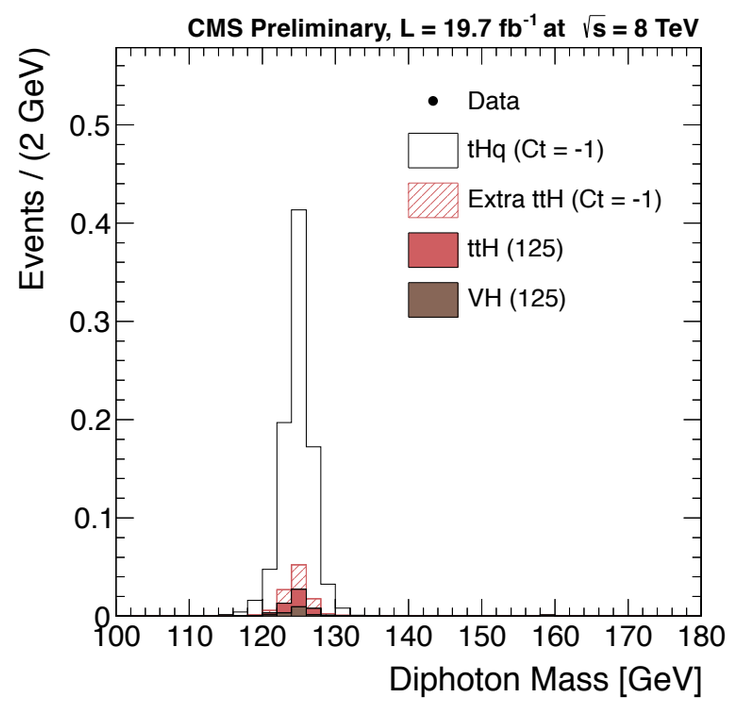
Not only ttH

- Precise MC simulations would be useful not only for ttH
- When we study BSM process similar to ttH, ~0 evts in sidebands



VLQ: TT→tHtH(→$\gamma\gamma$)

3 evts: minimum number of points for a fit



Anomalous t-H coupling: tH(→$\gamma\gamma$)q in leptonic channel

0 evts: background shape from control sample (low stat, big uncertainty)

- All possible data-driven control-samples are **low stat** in this phase space:
  - 1 lepton
  - high number of jets
  - high number of b-jets

If we find an excess and want to believe it robust MC prediction needed

What do we need?

- Up to now low-stat, MC not crucial
- For Run II **high-stat**:
  - MC is needed for precise optimization
  - Signal extraction still data-driven
- Minimal Wish-list:
  - tγ
  - t$\gamma\gamma$
  - ttγ
  - tt$\gamma\gamma$
  - $\gamma\gamma$ + many jets (Sherpa?)
- ttH optimization with BDT and full reconstruction of objects
- tHq and BSM searches would also profit from these:
  - tHq in the **hadronic channel** has to be done

Current status

- Something has changed probably:
  - At the time of the analysis we tried some $tt\gamma\gamma$ and $t\gamma\gamma$ MC but big Data/MC scale factors
  - Now there are **updated calculations**

arXiv.org > hep-ph > arXiv:1408.0278

High Energy Physics – Phenomenology

Hadroproduction of t anti-t pair with two isolated photons with PowHel

Adam Kardos, Zoltán Trócsányi

(Submitted on 1 Aug 2014)

- What has to be studied carefully in MC generation?
  - Matching: for photons and jets. We require a lot of jets in the final state
  - Leptons: in our 'handmade' MC requiring at least a lepton caused disagreement
  - How to sum samples (#jets, #b-jets...)

Conclusions

- Precise background estimation not crucial for $H \rightarrow \gamma\gamma$ Run1 given low-stat:
  - Optimization and background estimation from data
  - Simple cuts on #jets and b-jets (no fine optimization needed)
  - Low impact of systematics
- For Run 2 it could be crucial:
  - ttH optimization
  - BSM searches
- Wish-list :
  - $t\gamma$, $t\gamma\gamma$, $tt\gamma$, $tt\gamma\gamma$