



# The search for Higgs decaying to $b\bar{b}$ in CMS

Michele de Gruttola (U of Florida)  
on behalf of CMS collaboration

EPS 2013, Stockholm 18/07/2013



# Outline



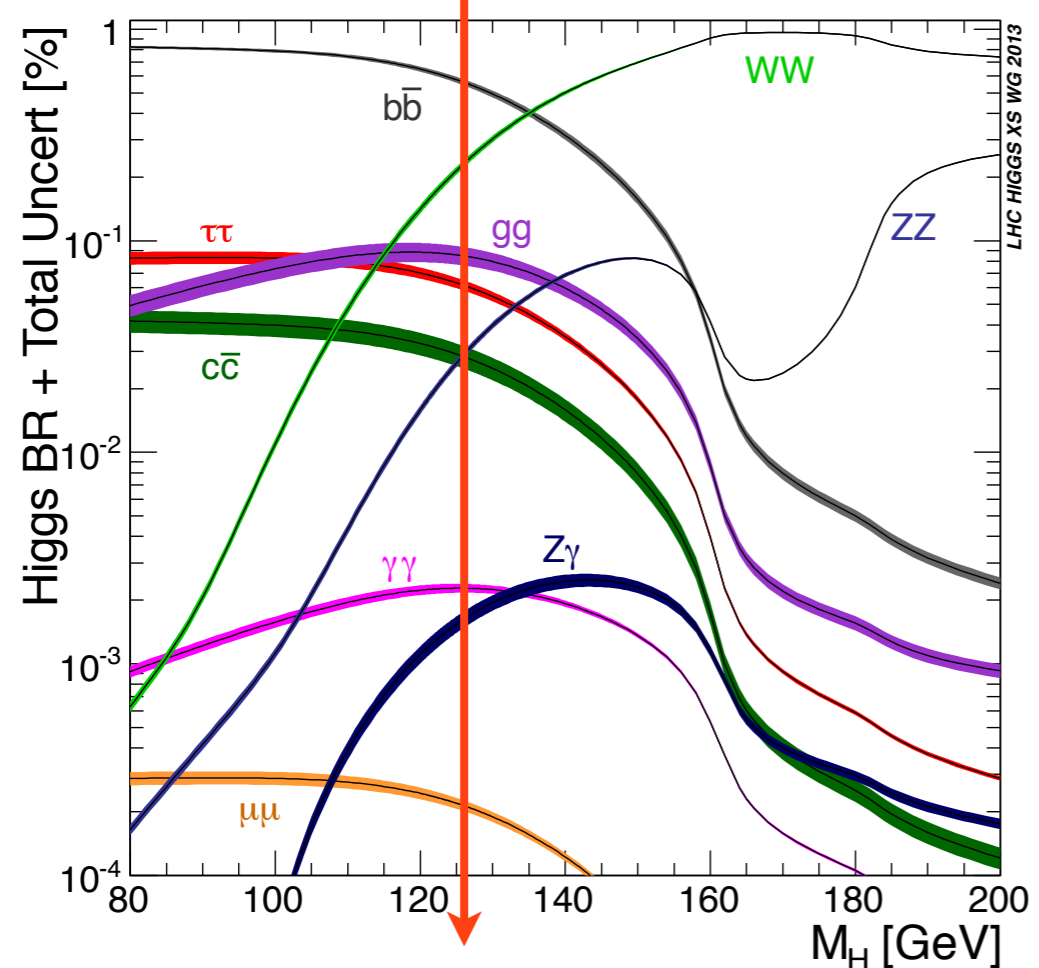
- **H->bb@LHC**
  - *why it is so important and challenges*
- **H->bb at CMS**
  - *Hbb in association with a Z,W (VHbb, CMS-HIG-13-12)*
    - ▶ strategy, main ingredients of the analysis, results
  - *Hbb in association with two jets from Vector Boson Fusion (VBFHbb, CMS-HIG-13-11)*
    - ▶ strategy, ingredients, results
  - *Hbb in association with top pair (ttHbb, arxiv.1303.0763: ttH with H(bb))*
    - ▶ not here, see soon talk of F.Margaroli
- **preliminary results on full 2011+2012 data and implications**



# Higgs BRs



Michele de Gruttola



@125:  
 $BR(H \Rightarrow bb) \sim 58\%$   
 $BR(H \Rightarrow WW) \sim 22\%$   
 $BR(H \Rightarrow \tau\tau) \sim 6\%$   
 $BR(H \Rightarrow ZZ^*) \sim 3\%$   
 $BR(H \Rightarrow \gamma\gamma) \sim 0.22\%$

*b's and tau's: test the Higgs couplings to fermions*

*vital part of the Higgs identification and Higgs property studies*

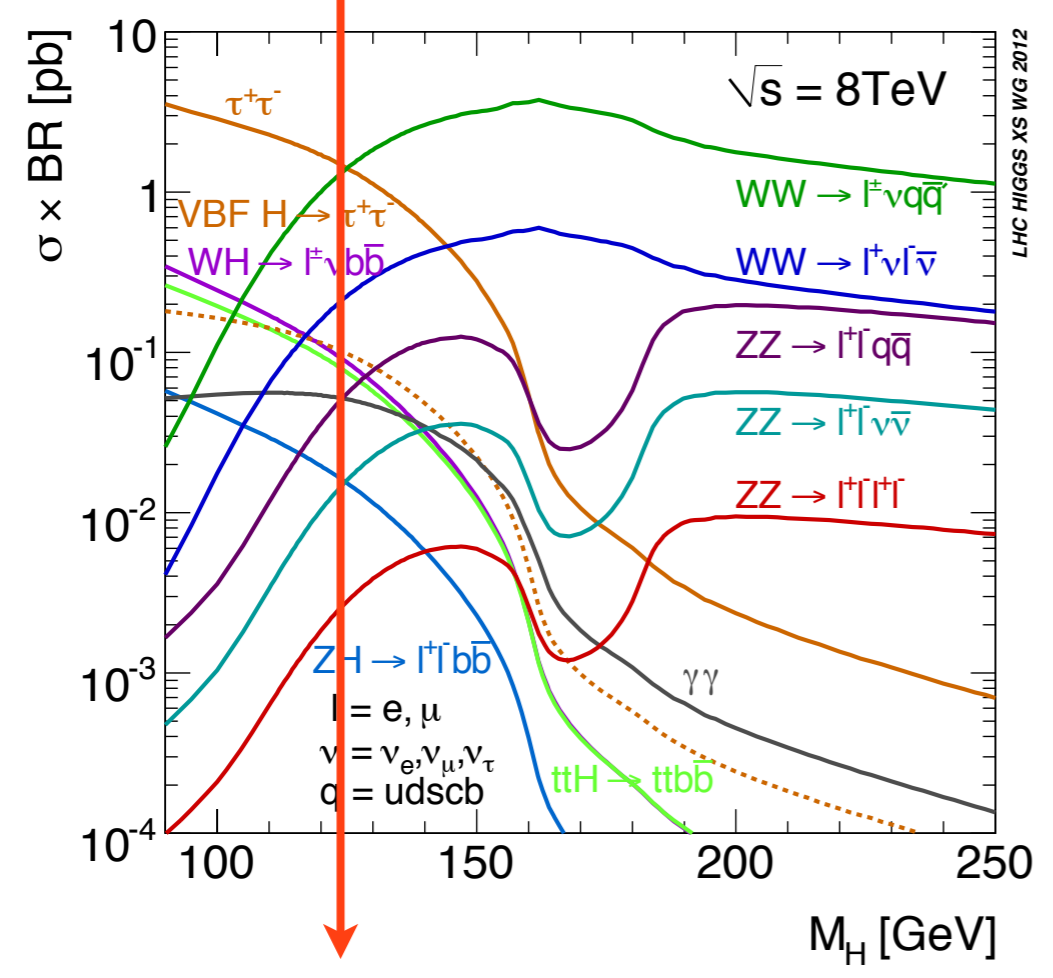
*But life is hard...*



# Higgs BRs



Michele de Gruttola

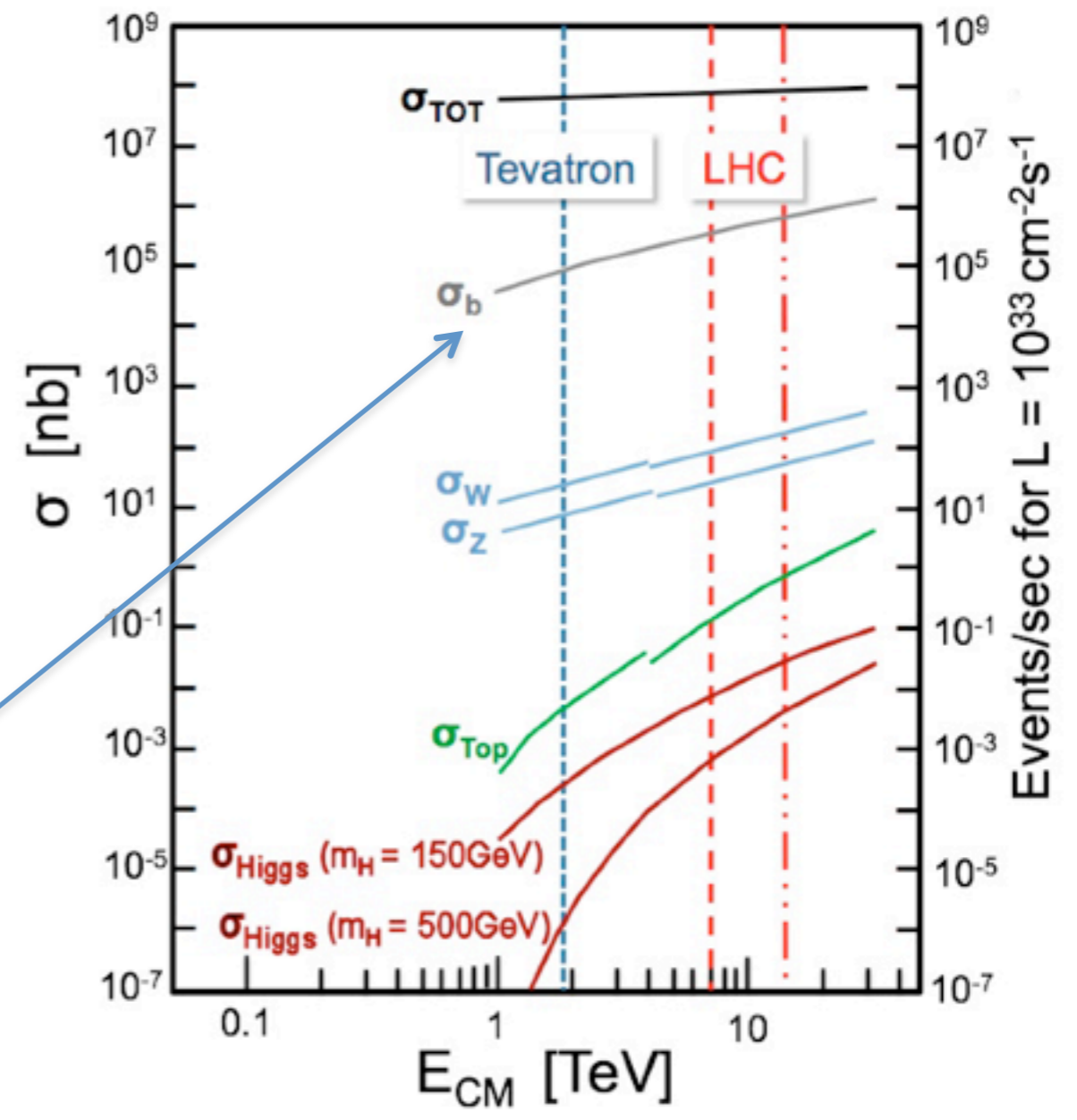


b's and tau's: test the Higgs couplings to fermions

vital part of the Higgs identification and Higgs property studies

But life is hard...

But  $\sigma_{bb}(\text{QCD}) \sim 10^7$   
 $\sigma \times \text{BR}(H \rightarrow bb)$ !



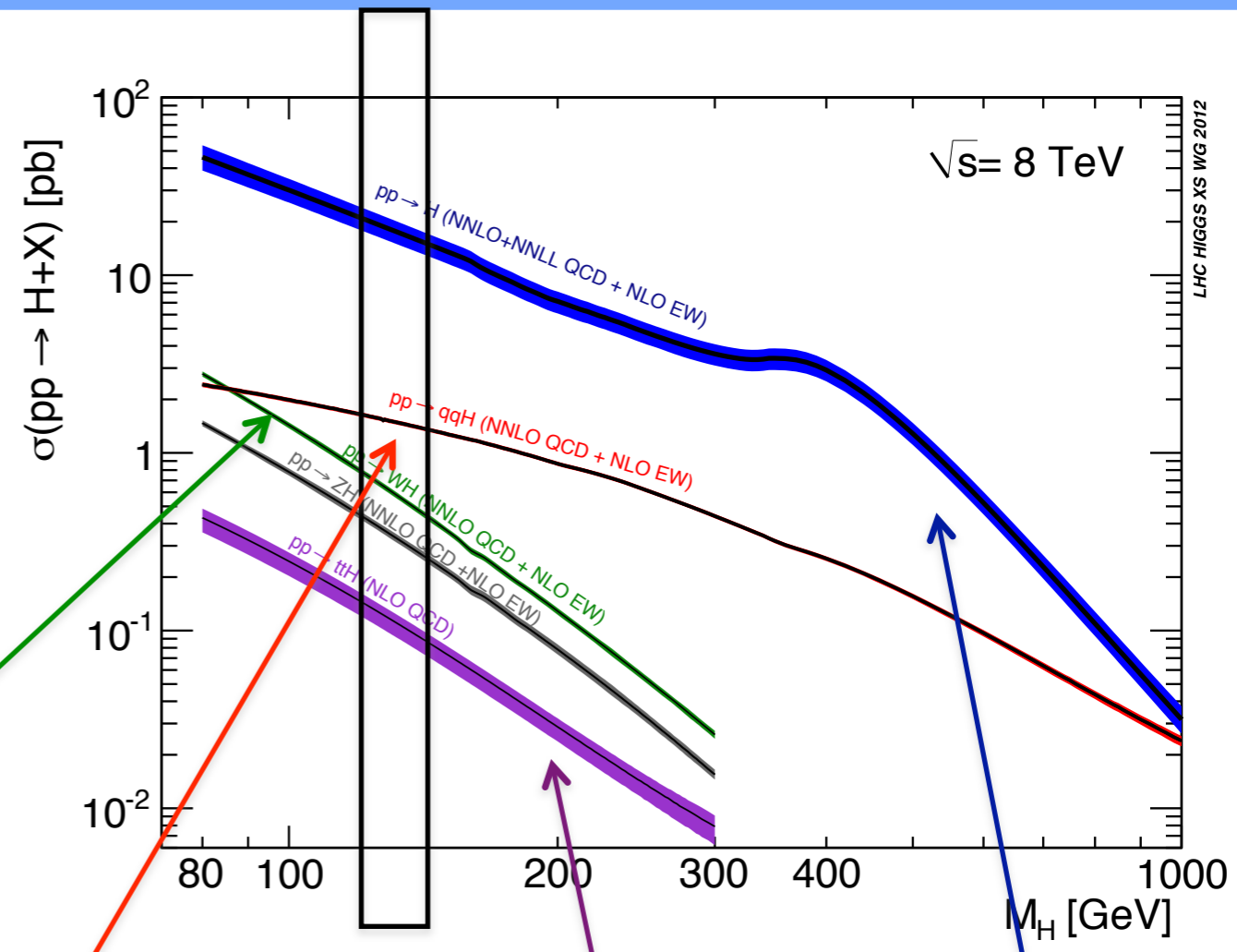


# Associated production

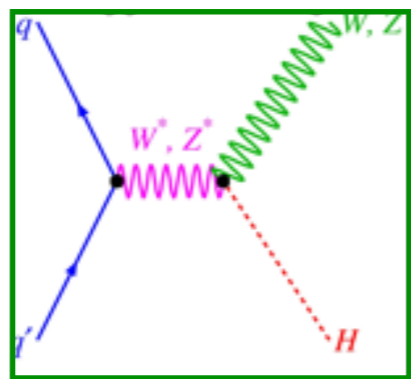


We need to search  $t\bar{t}H$  in association with other objects

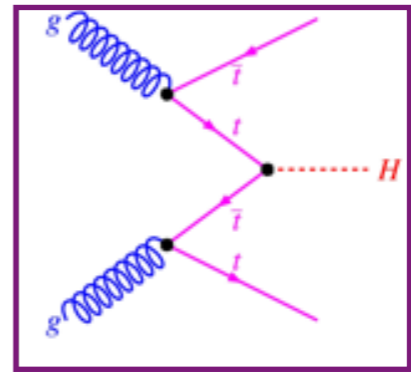
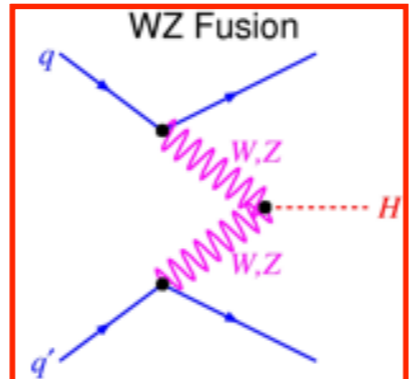
- gluon fusion impossible :(
- VBF cleaner, but 4jets QCD still high
- $VH$  - work horse can use lepton/MET and topology
- $t\bar{t}H$  - even less rate, look at topology, but  $t\bar{t}$  bkg well understood



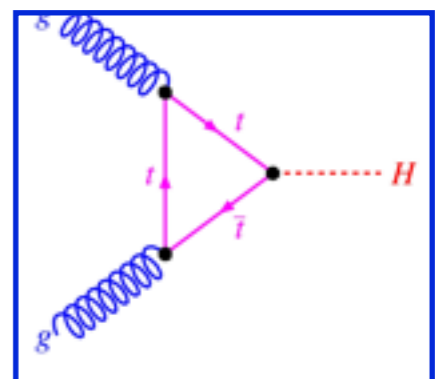
Association with Z/W



Vector boson fusion

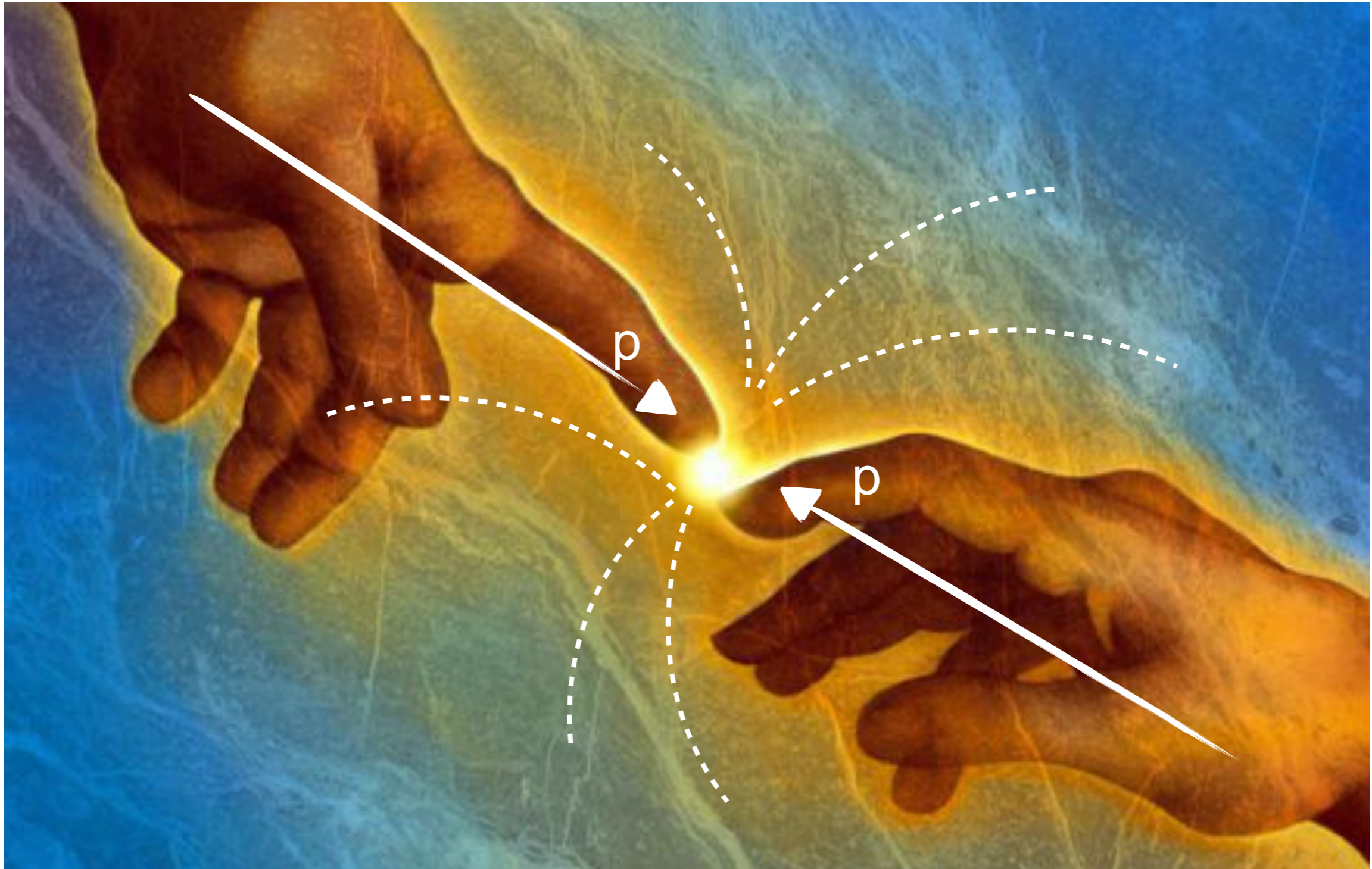


with  $t\bar{t}$



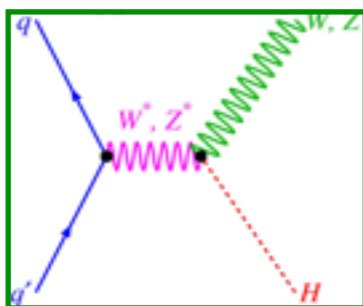
gluon fusion

Michele de Gruttola

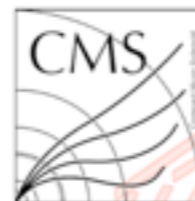


# VH → Vbb

# V → lv, ll, vv



- Most sensitive channel with b as final state
- interesting to compare with Tevatron results (2.8@125 excess for  $VHbb$ )
- can calibrate/test the results looking at SM candle  $VZbb$



CMS Experiment at LHC, CERN  
 Data: Mon Jun 27 02:59:42 2011 CEST  
 Run/Event: 167807 / 149404739  
 Lumi section: 134  
 Orbit/Crossing: 35103256 / 2259

$M_{bb} = 128 \text{ GeV}$   
 $p_T(bb) = 181 \text{ GeV}$

$ZH \rightarrow \mu\mu bb$



b-jet  
 CSV~0.92

Muon 0  
 $p_T = 161.78 \text{ GeV/c}$

Muon 1  
 $p_T = 27.25 \text{ GeV/c}$

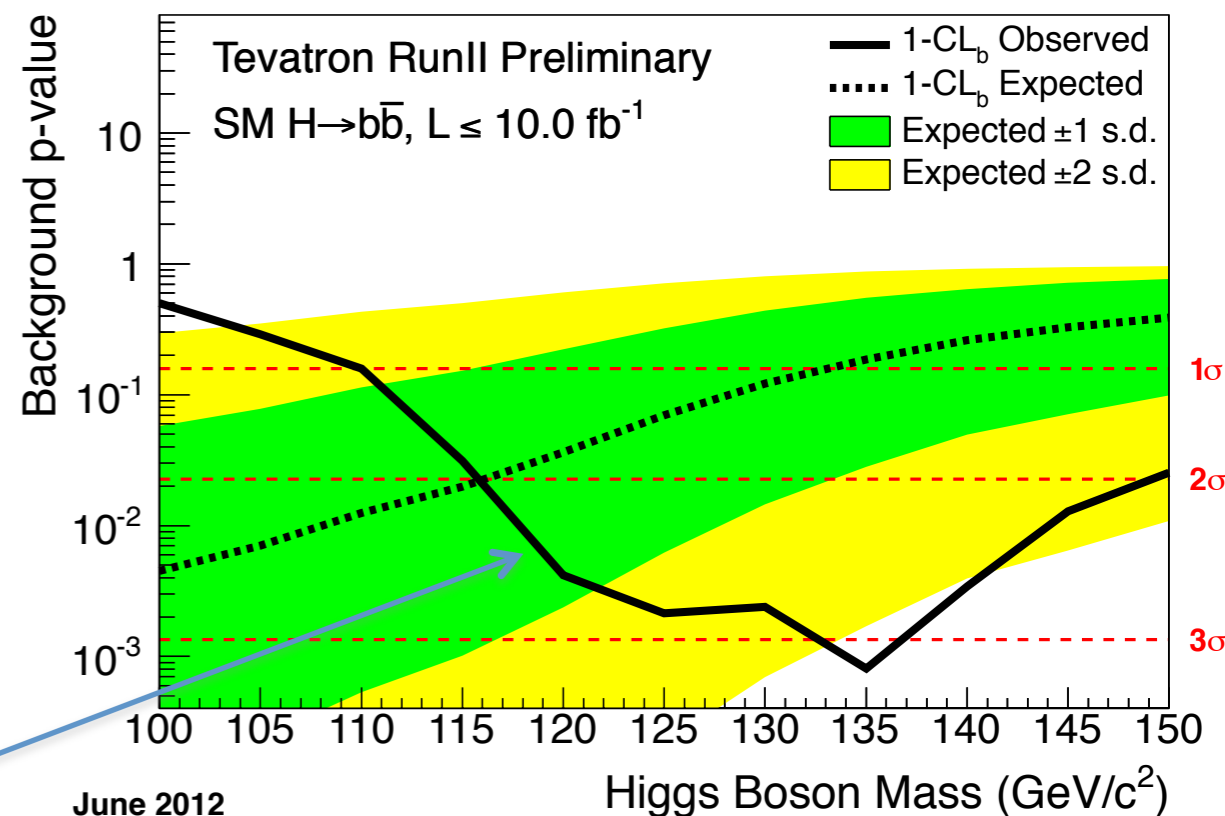
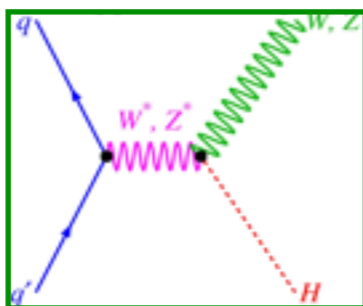
Second B Jet Candidate  
 $p_T = 48.44 \text{ GeV/c}$

Combined Secondary

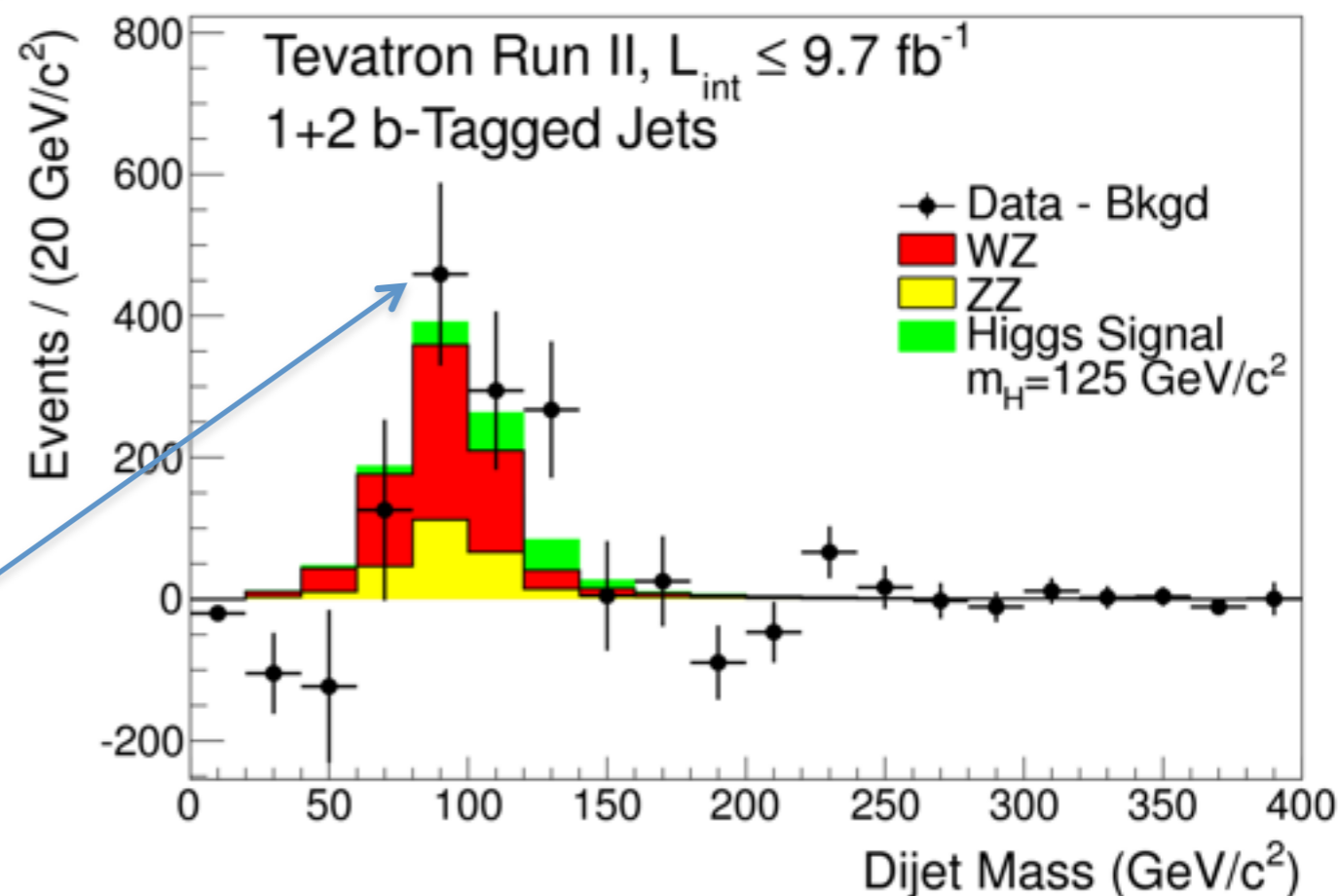
b-jet CSV~0.99

# VH → Vbb

# V → lv, ll, vv



- Most sensitive channel with  $b$  as final state
- interesting to compare with Tevatron results (2.8@125 excess for  $VHbb$ )
- can calibrate/test the results looking at SM candle  $VZbb$





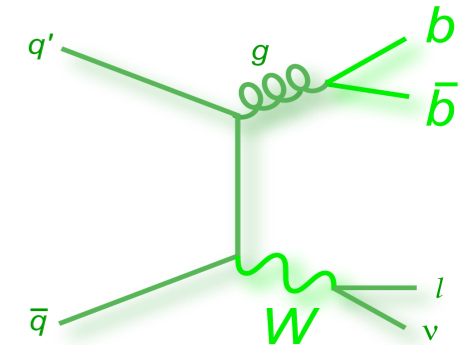
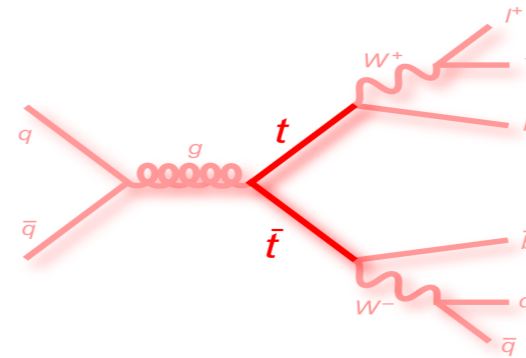


# VHbb strategy



## ■ backgrounds

- *not peaking: top (pairs),  $\nu$  + bb,  $\nu$  + udscg, QCD*
- normalized from data control regions
- *peaking:  $\nu$ Zbb*



## ■ Search done in 6 channels

- Z(l)Hbb (l=e,μ)
- Z(νν)Hbb
- W(lν)Hbb (l=e,μ,τ)

## ■ categories in Vector boson pt

- *high ( $\nu$ pt ≥ 150) and medium pt ( $\nu$ pt ≥ 100)*

## ■ V and H back to back

## ■ 2 b-tag

- **CSV**: likelihood tagger using many jet properties: secondary vertex (if any), tracks impact parameters, etc.

## ■ reconstruct $m_{bb}$

- *most background are not peaking*
- *we gain a lot from improving the mass resolution*

## ■ likelihood fit to MVA (using BDT) final discriminant

Michele de Gruttola

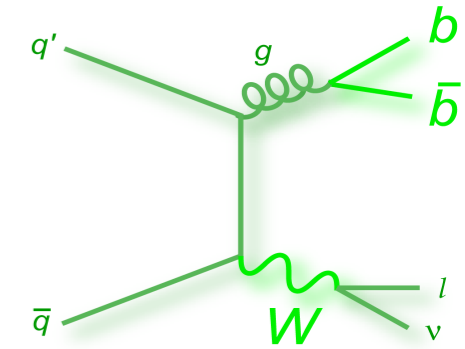
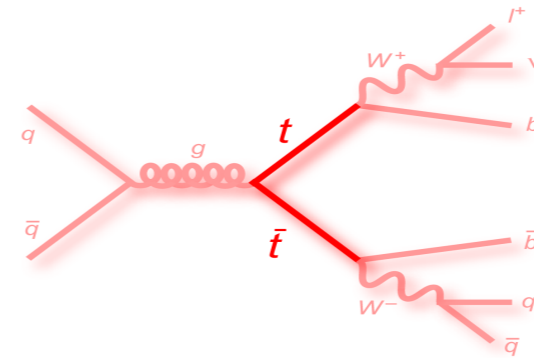


# VHbb strategy



## backgrounds

- *not peaking: top (pairs),  $\nu$  + bb,  $\nu$  + udscg, QCD*
- normalized from data control regions
- *peaking:  $\nu$ Zbb*



## Search done in 6 channels

- Z(l l)Hbb (l=e,  $\mu$ )
- Z( $\nu\nu$ )Hbb
- W(l $\nu$ )Hbb (l=e,  $\mu$ ,  $\tau$ )

## categories in Vector boson pt

- *high ( $\nu$ pt  $\geq 150$ ) and medium pt ( $\nu$ pt  $\geq 100$ )*

## V and H back to back

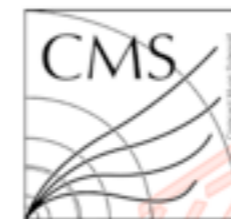
## 2 b-tag

- **CSV**: likelihood tagger using many jet properties: secondary vertex (if any), tracks impact parameters, etc.

## reconstruct $m_{bb}$

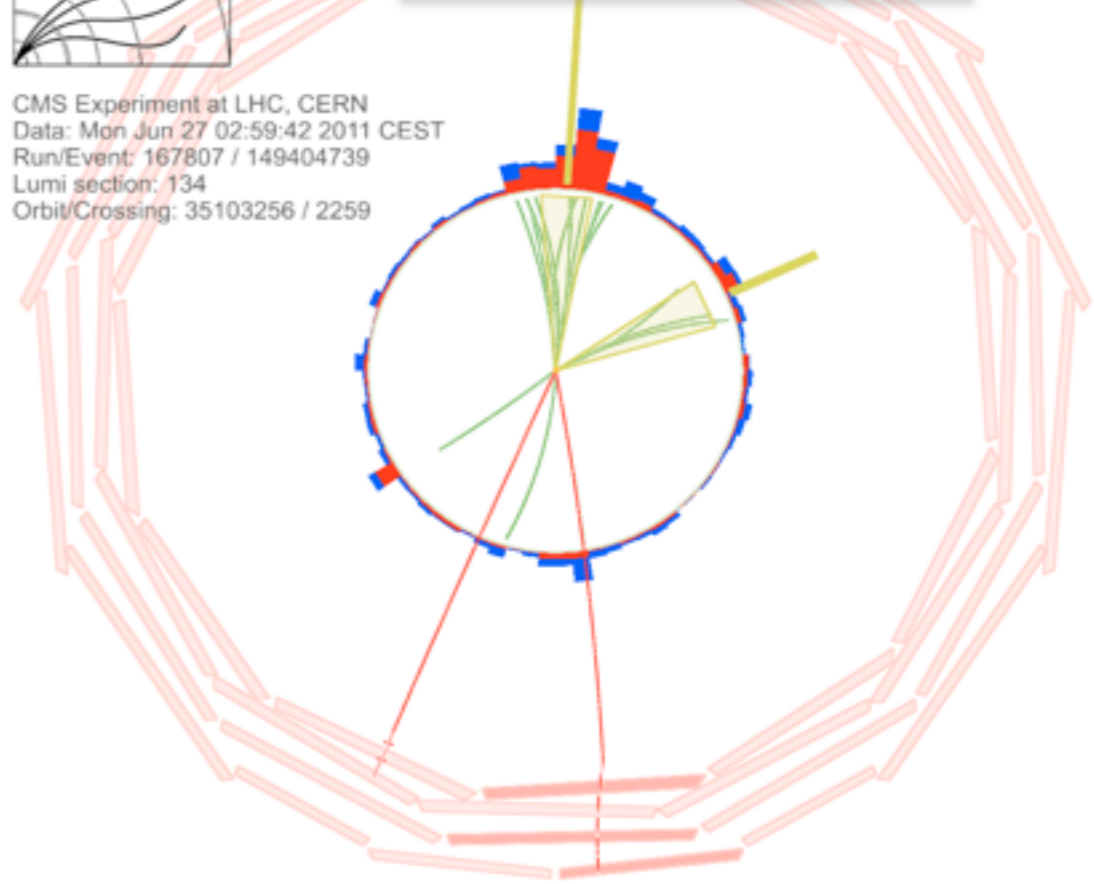
- *most background are not peaking*
- *we gain a lot from improving the mass resolution*

## likelihood fit to MVA (using BDT) final discriminant



CMS Experiment at LHC, CERN  
 Data: Mon Jun 27 02:59:42 2011 CEST  
 Run/Event: 167807 / 149404739  
 Lumi section: 134  
 Orbit/Crossing: 35103256 / 2259

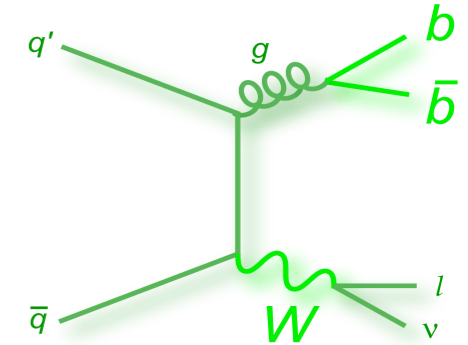
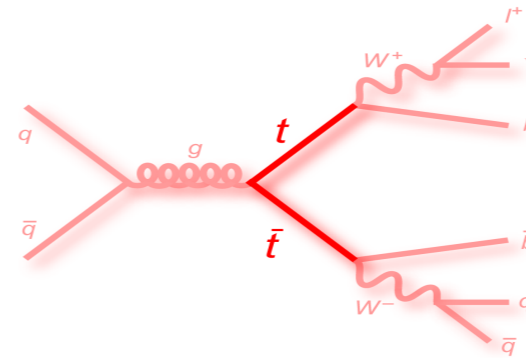
### ZH- $\rightarrow\mu\mu$ bb



Michele de Gruttola

## backgrounds

- *not peaking: top (pairs),  $\nu$  + bb,  $\nu$  + udscg, QCD*
- normalized from data control regions
- *peaking:  $\nu$ Zbb*



## Search done in 6 channels

- Z(l $\bar{l}$ )Hbb (l=e, $\mu$ )
- Z( $\nu\nu$ )Hbb
- W(l $\bar{\nu}$ )Hbb (l=e, $\mu$ , $\tau$ )

## categories in Vector boson pt

- *high ( $\nu_{pt} \geq 150$ ) and medium pt ( $\nu_{pt} \geq 10$ )*

## V and H back to back

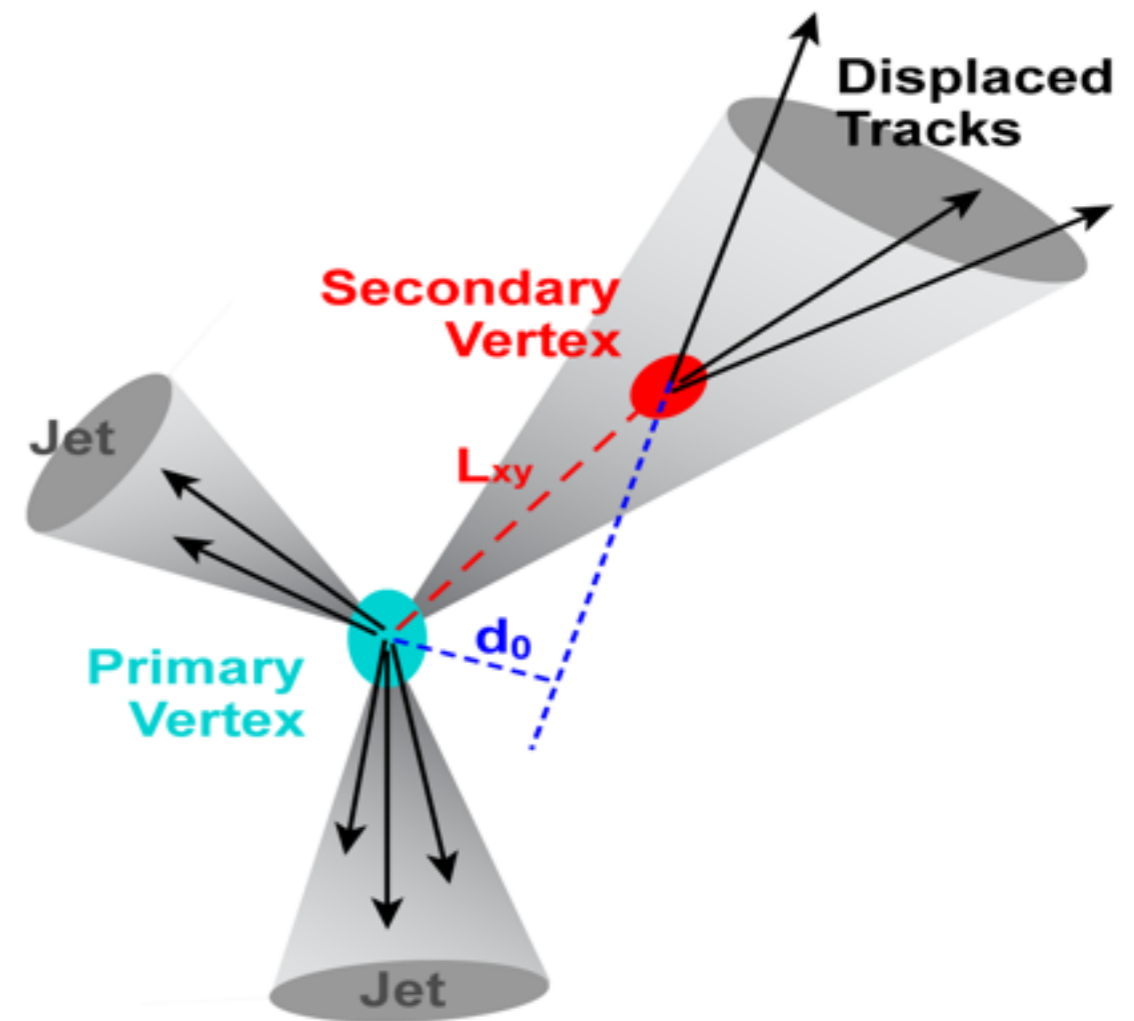
## 2 b-tag

- CSV: likelihood tagger using many jet properties: secondary vertex (if any), tracks impact parameters, etc.

## reconstruct $m_{bb}$

- *most background are not peaking*
- *we gain a lot from improving the mass resolution*

## likelihood fit to MVA (using BDT) final discriminant



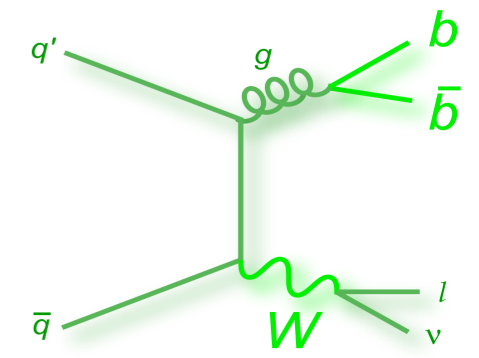
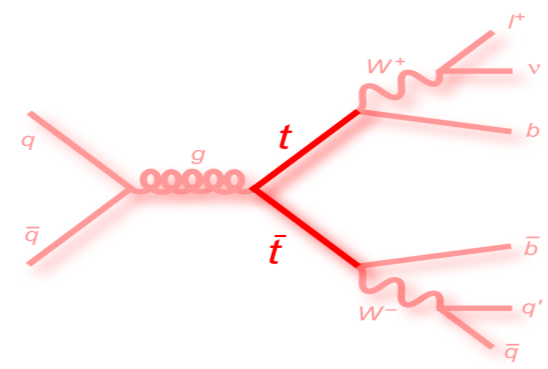


# VHbb strategy



## backgrounds

- not peaking: top (pairs),  $\nu$  + bb,  $\nu$  + udscg, QCD
- normalized from data control regions
- peaking:  $\nu$ Zbb



## Search done in 6 channels

- Z(l $\bar{l}$ )Hbb (l=e, $\mu$ )
- Z( $\nu\nu$ )Hbb
- W(l $\bar{\nu}$ )Hbb (l=e, $\mu$ , $\tau$ )

## categories in Vector boson pt

- high ( $\nu$ pt $\geq$ 150) and medium pt ( $\nu$ pt $\geq$ 10)

## V and H back to back

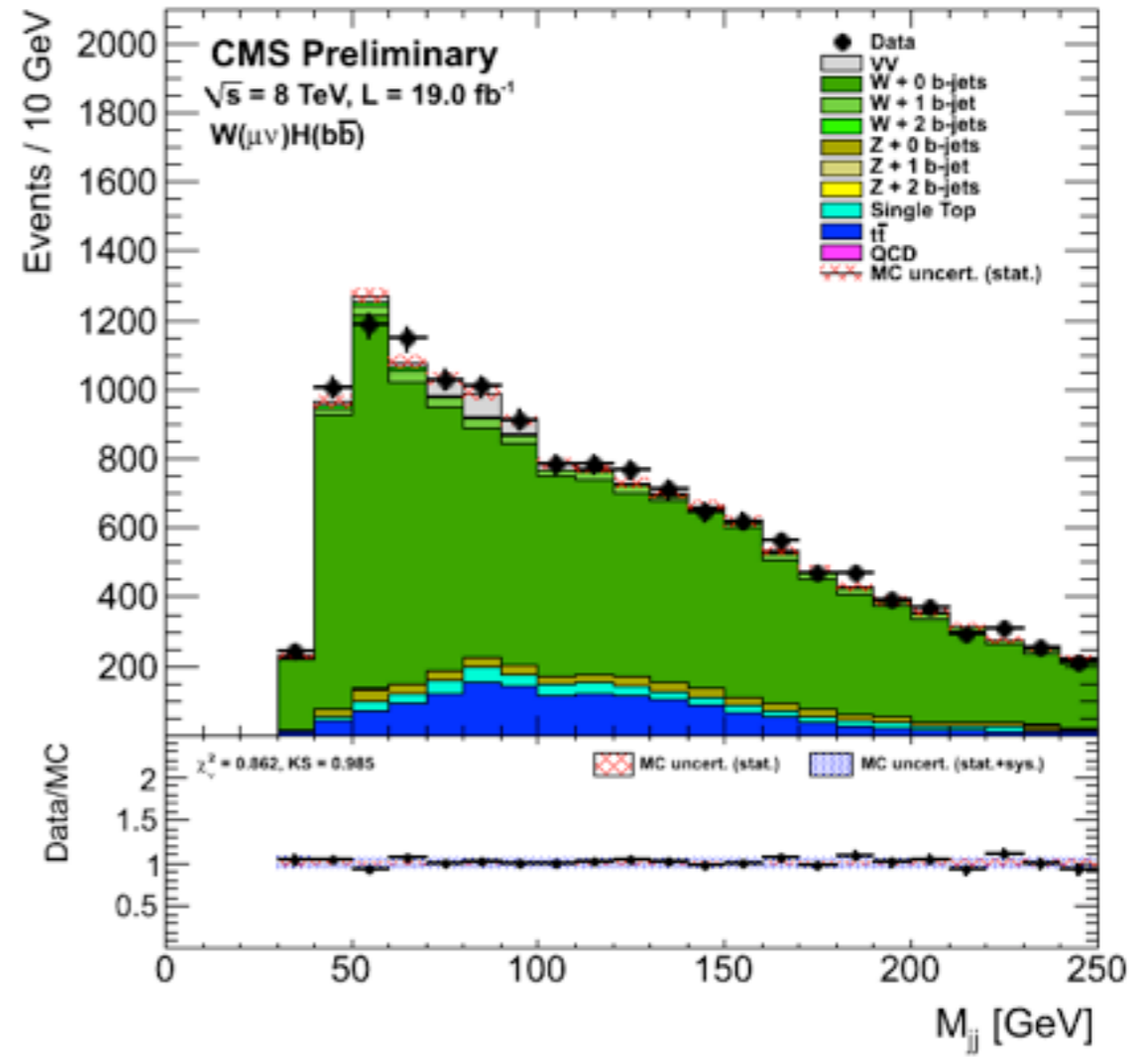
## 2 b-tag

- CSV**: likelihood tagger using many jet properties: secondary vertex (if any), tracks impact parameters, etc.

## reconstruct $m_{bb}$

- most background are not peaking
- we gain a lot from improving the mass resolution

## likelihood fit to MVA (using BDT) final discriminant



Michele de Gruttola

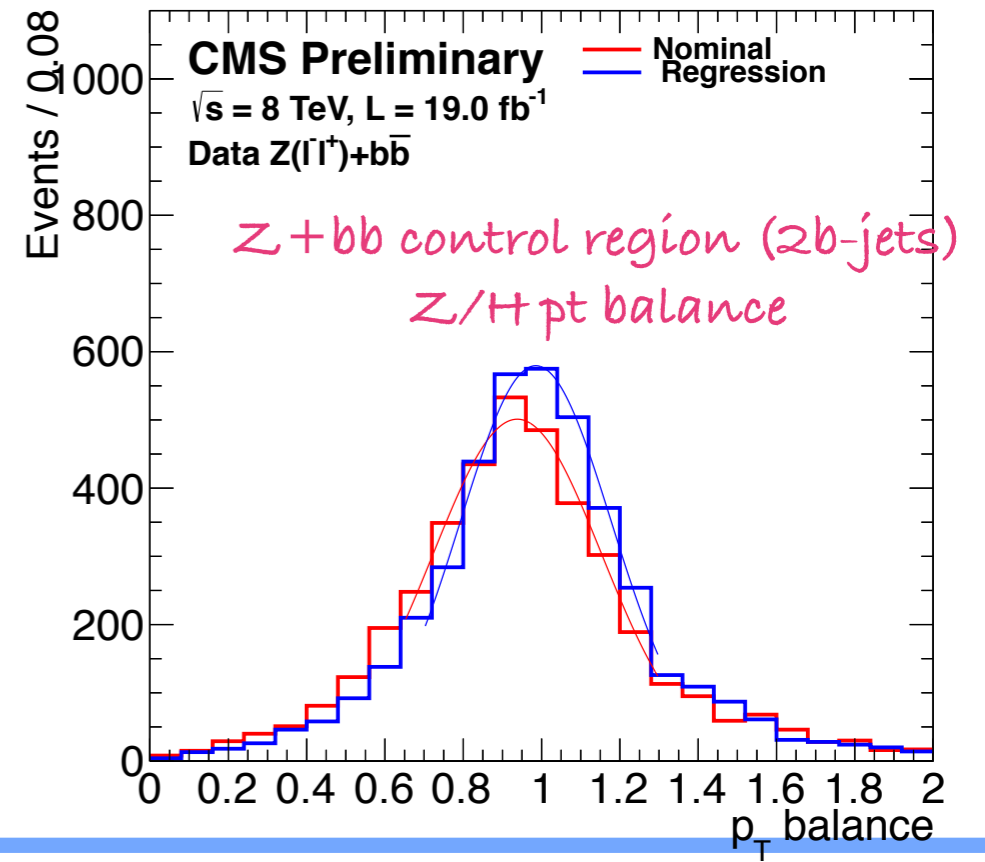
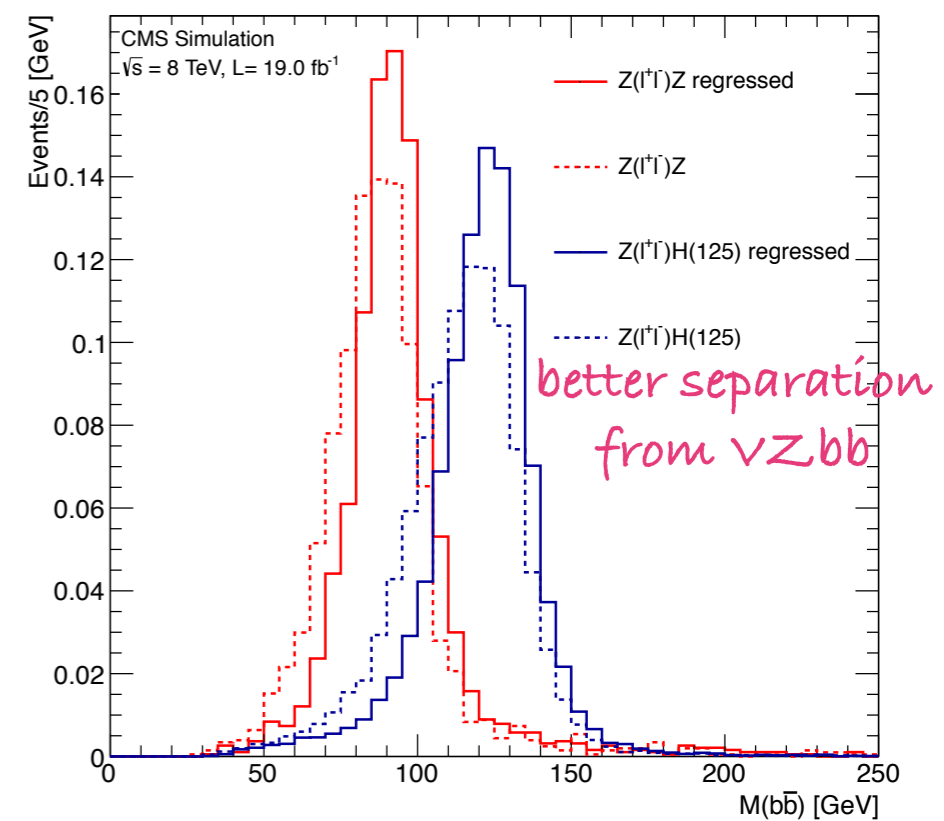


# VHbb $m_{bb}$ improvement



Michele de Gruttola

- b-jet energy resolution improvement (and hence mass)
  - *use BDT regression*
    - target generated parton energy
    - use jet info
      - ▶ kinematics:  $p_{t,b}$ , energy, JEC unc, mt raw pt, n of charged tracks
      - ▶ b-jet specific: soft lepton pt, ptRel,  $p_{t,vtx}$
      - ▶ topology: average energy, MET (for Z/Hbb only)
- validation in data using Z + dijets pt balance and top mass





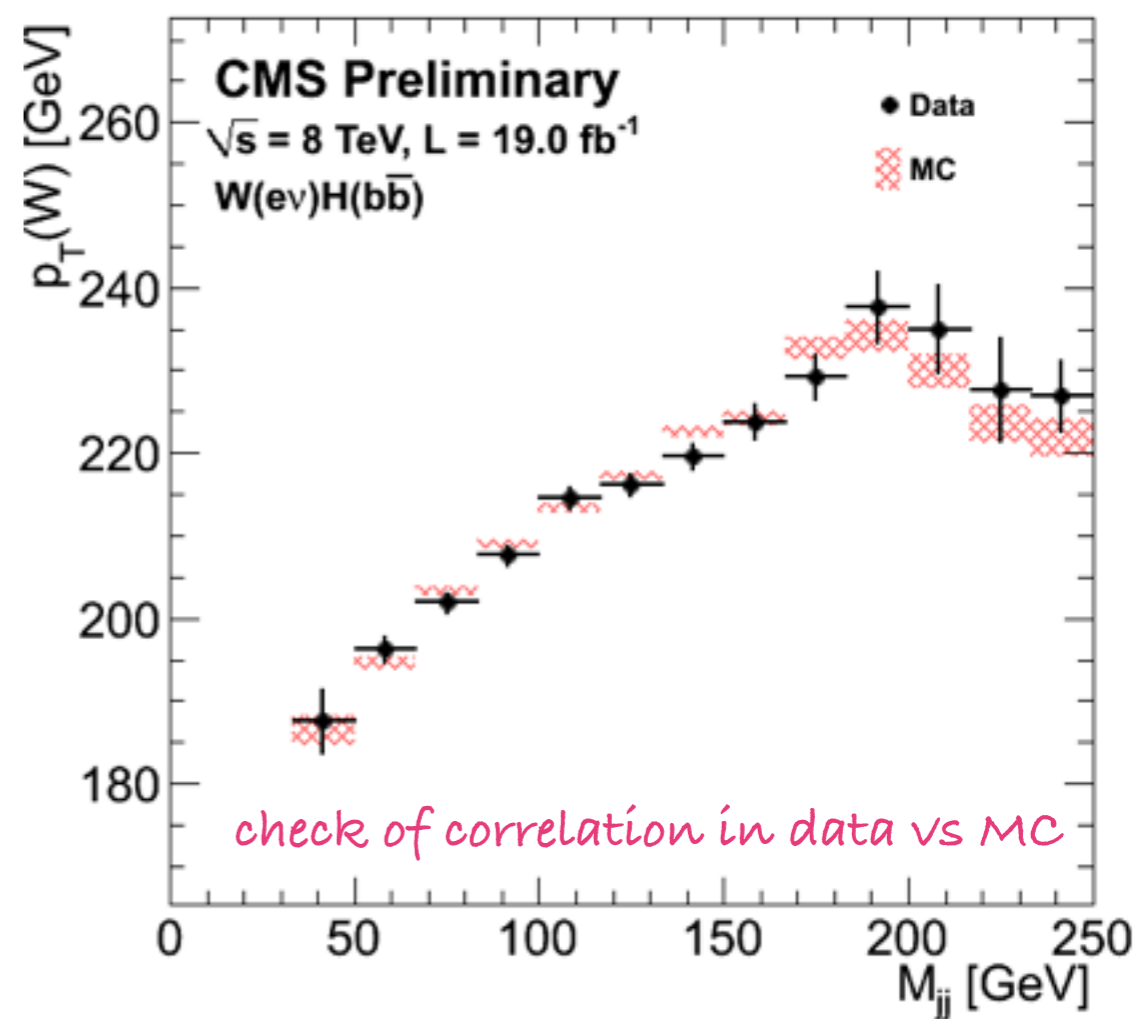
- Use fit to BDT discriminant
  - *using as inputs*
    - kinematics:  $m_{bb}$ ,  $pt_{bb}$ ,  $pt_b$
    - b-jet specific:  $btag_{1,2}$
    - topology: additional jets,  $\Delta\Phi_{bb}$ , ...
- check correlations in data vs MC
- in total 14 BDT (8TeV) + 10 BDT (7TeV)
  - *some of them trained separately for the each individual background)*



# VHbb Event classification



- Use fit to BDT discriminant
  - *using as inputs*
    - kinematics:  $m_{bb}$ ,  $pt_{bb}$ ,  $pt_b$
    - b-jet specific:  $btag_{1,2}$
    - topology: additional jets,  $\Delta\Phi_{bb}$ , ...
- check correlations in data vs MC
- in total 14 BDT (8TeV) + 10 BDT (7TeV)
  - *some of them trained separately for the each individual background)*





# VHbb Event classification

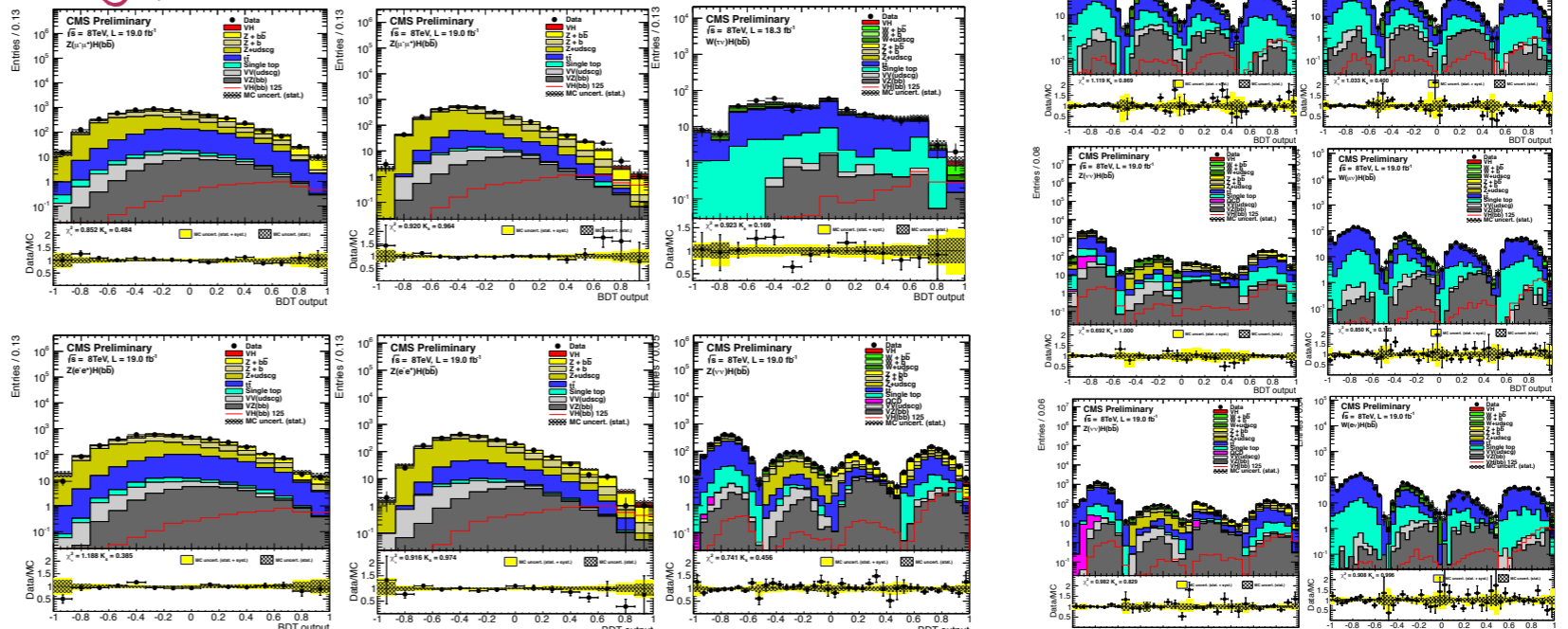


Michele de Gruttola

- Use fit to BDT discriminant
  - *using as inputs*
    - kinematics:  $m_{bb}$ ,  $pt_{bb}$ ,  $pt_b$
    - b-jet specific:  $btag_{1,2}$
    - topology: additional jets,  $\Delta\Phi_{bb}$ , ...

*hard to tell from here...*

- check correlations in data vs MC
- in total 14 BDT (8TeV) + 10 BDT (7TeV)
  - *some of them trained separately for the each individual background*



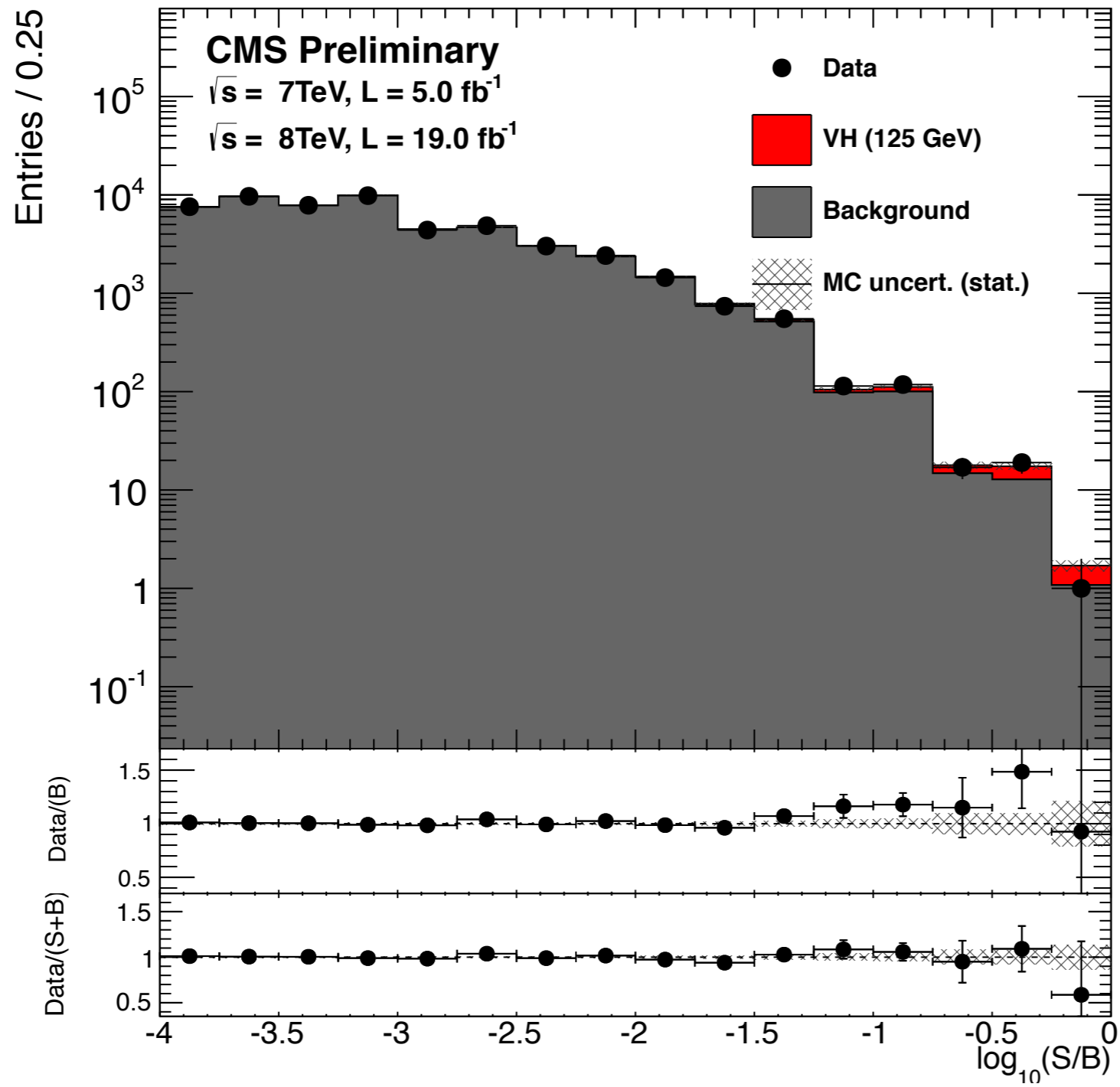




# VHbb BDT summary



- Summarize BDT analysis by sorting bins in S/B



*excess of events in the high s/b bins*

Michele de Gruttola

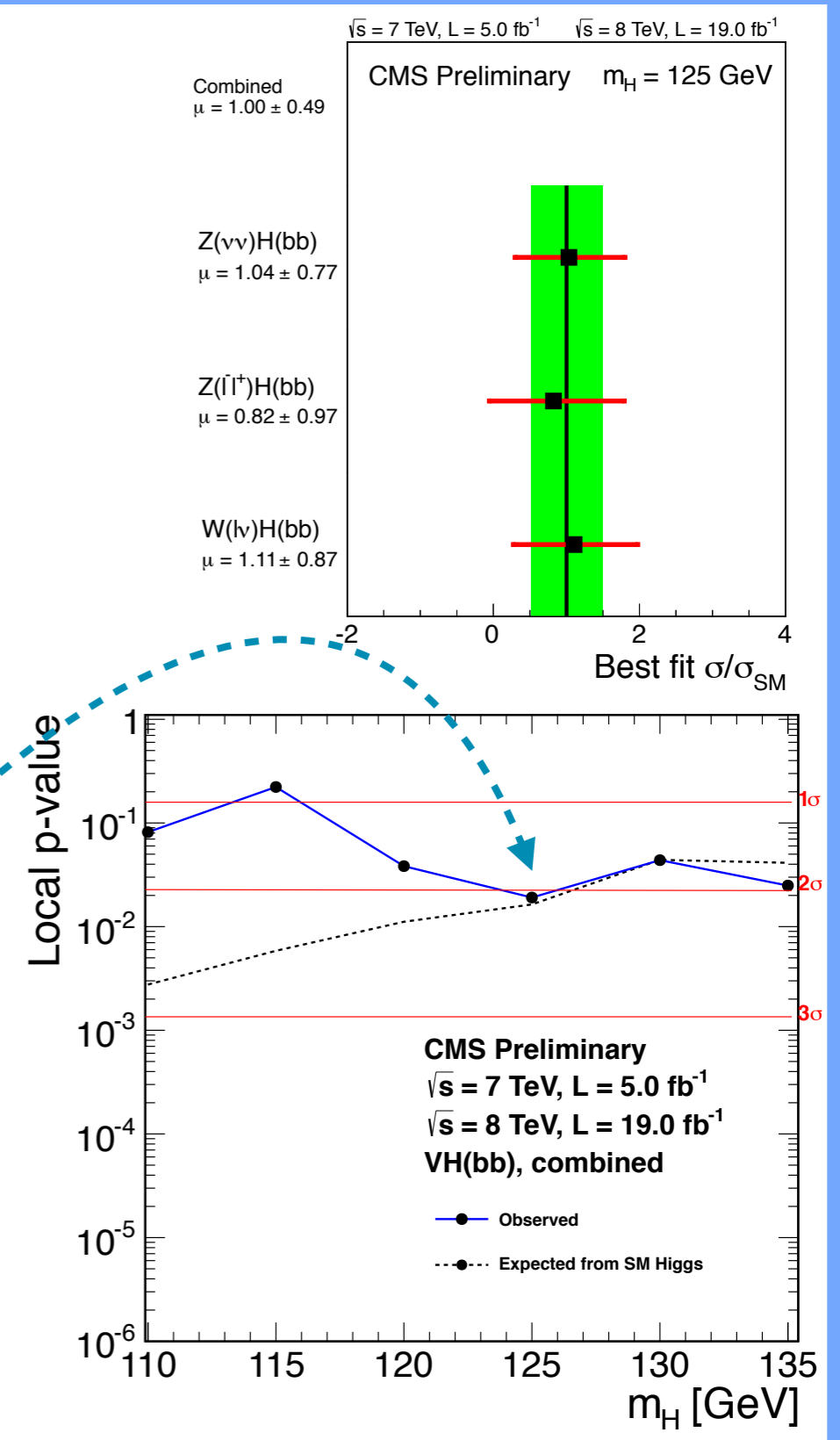
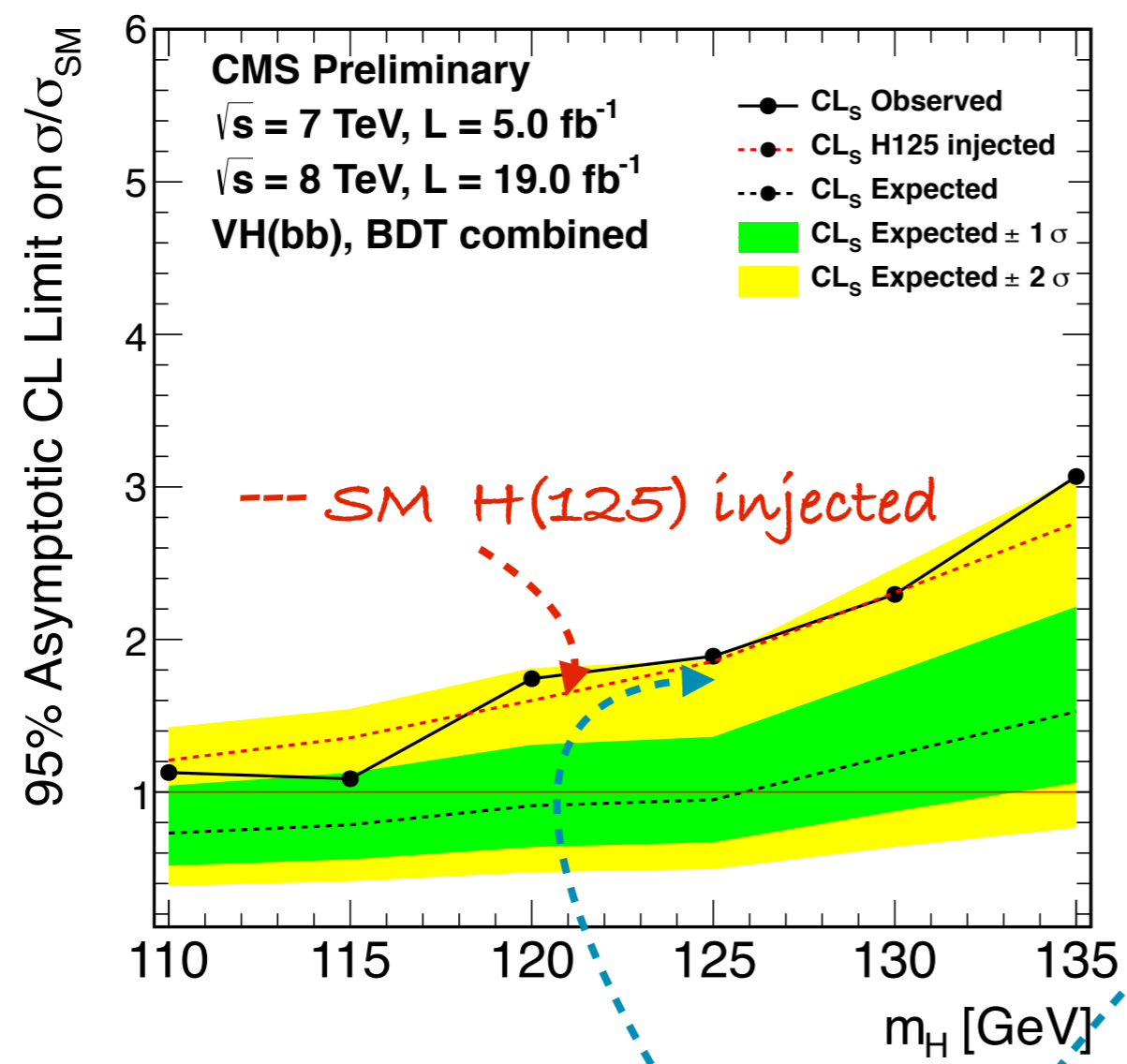


# VHbb

# Results



Michele de Gruttola



- Observed (expected) limit of **1.89(0.95)xSM** at 125 GeV
- Observed (expected) local significance of **2.1 $\sigma$  (2.1 $\sigma$ )** for  $m_H=125 \text{ GeV}$
- Combined best-fit  **$\mu=1.0 \pm 0.5$**

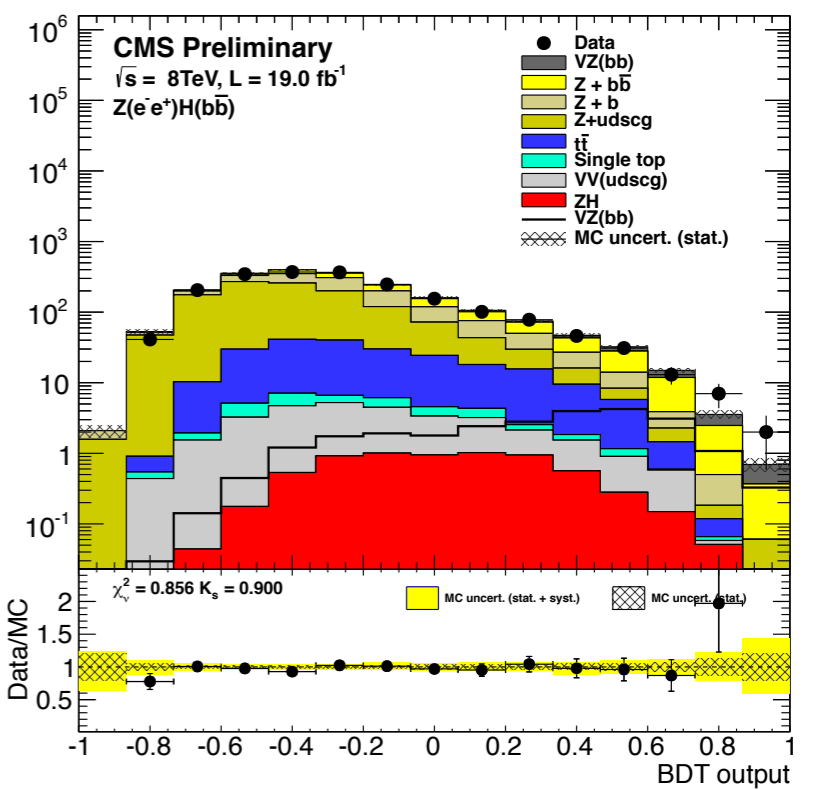
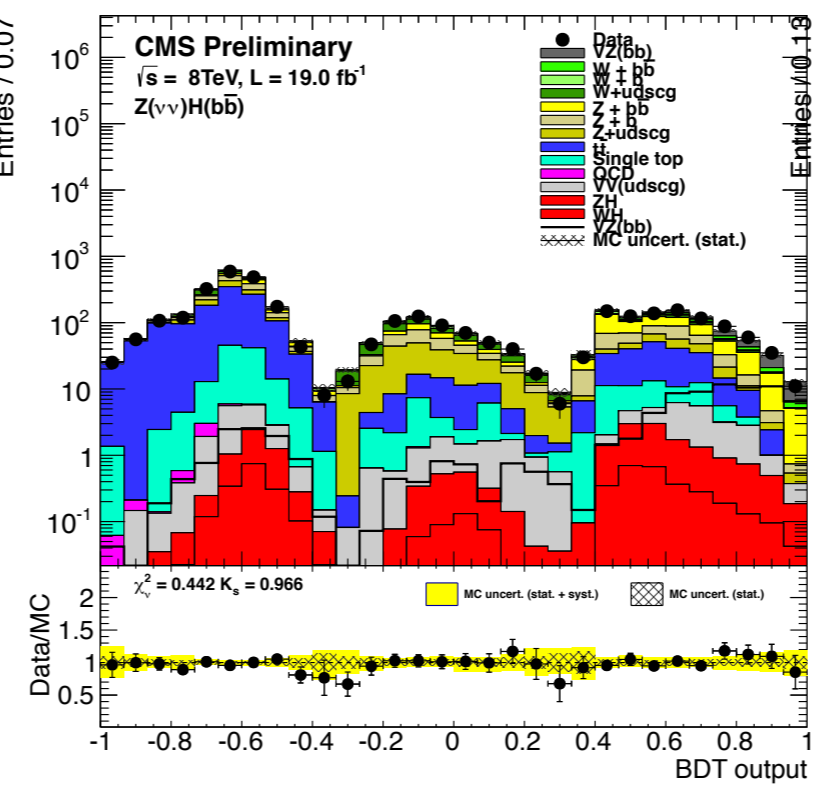
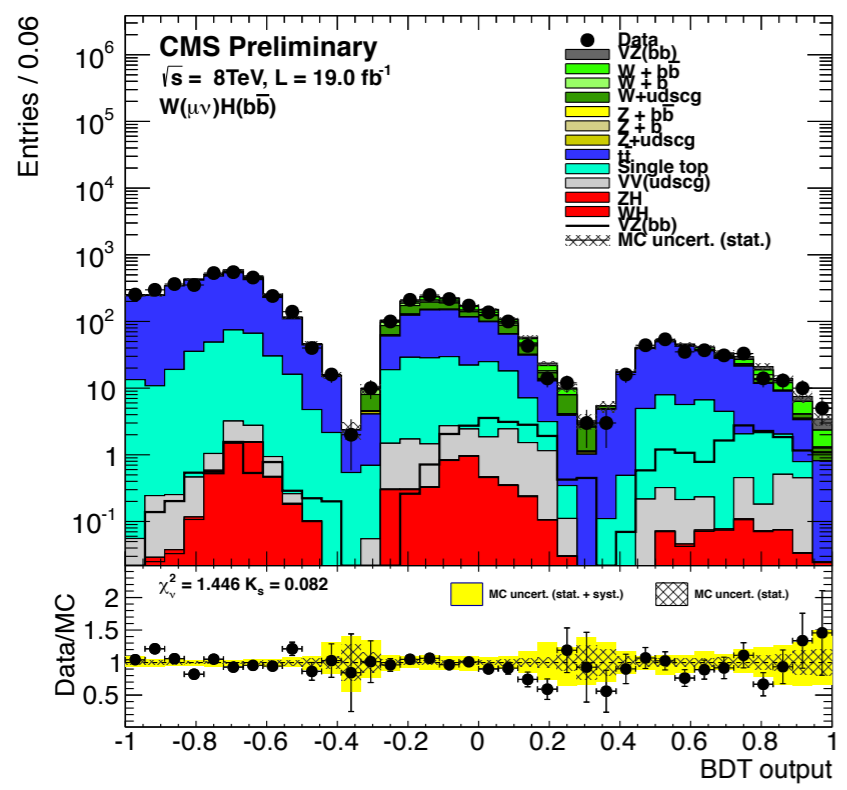


# VZ(bb) results



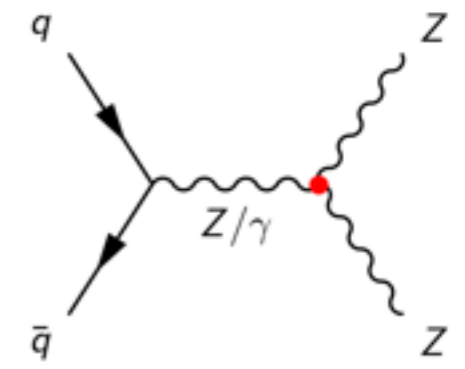
- same technique as in VH for 8TeV data for a bkg sitting close by

Michele de Gruttola



- we measure this SM candle with the right xsection

expected signif	6.30
obs signif	7.50
best fit $\mu$	$1.19^{+0.27}_{-0.23}$

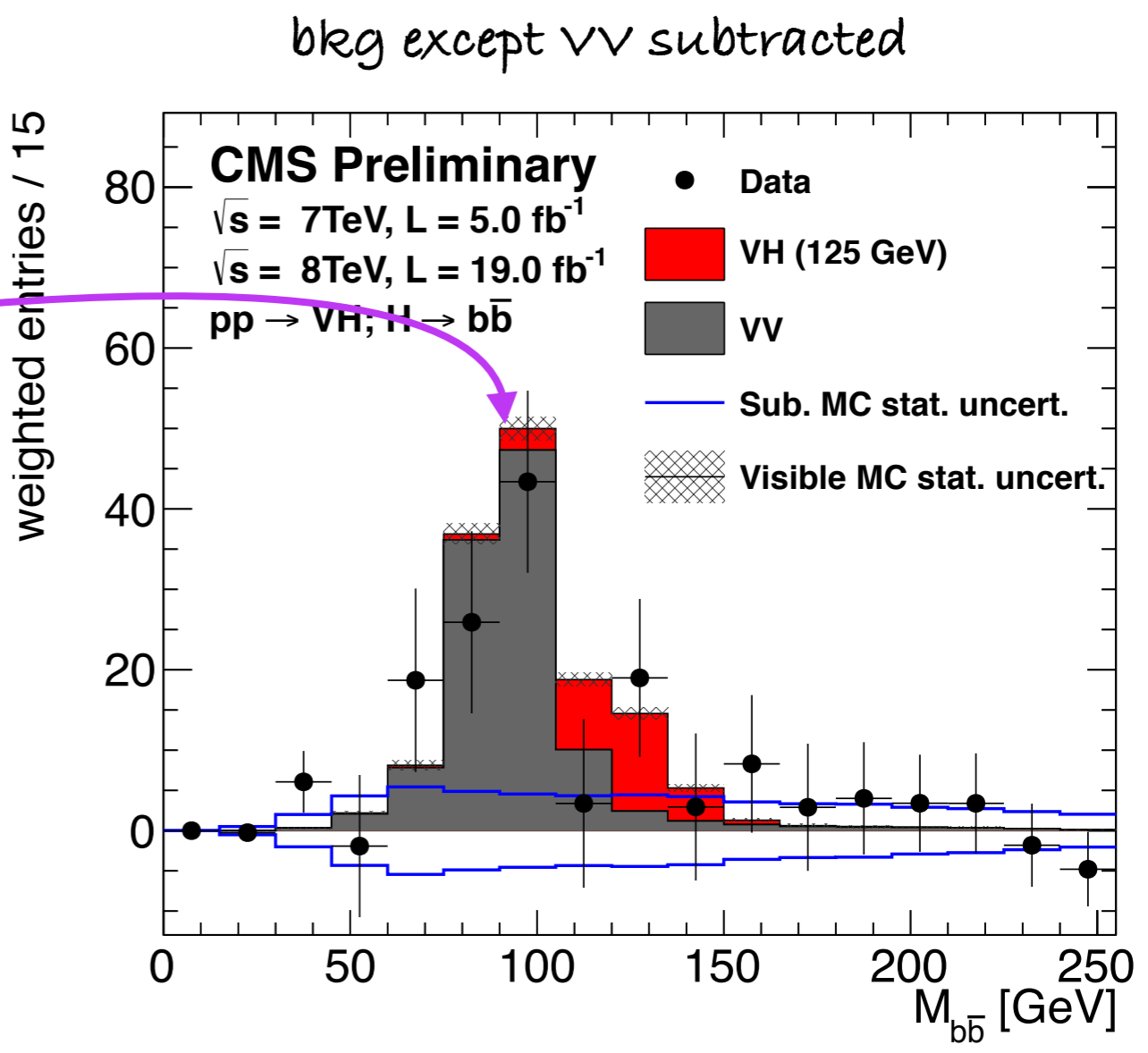
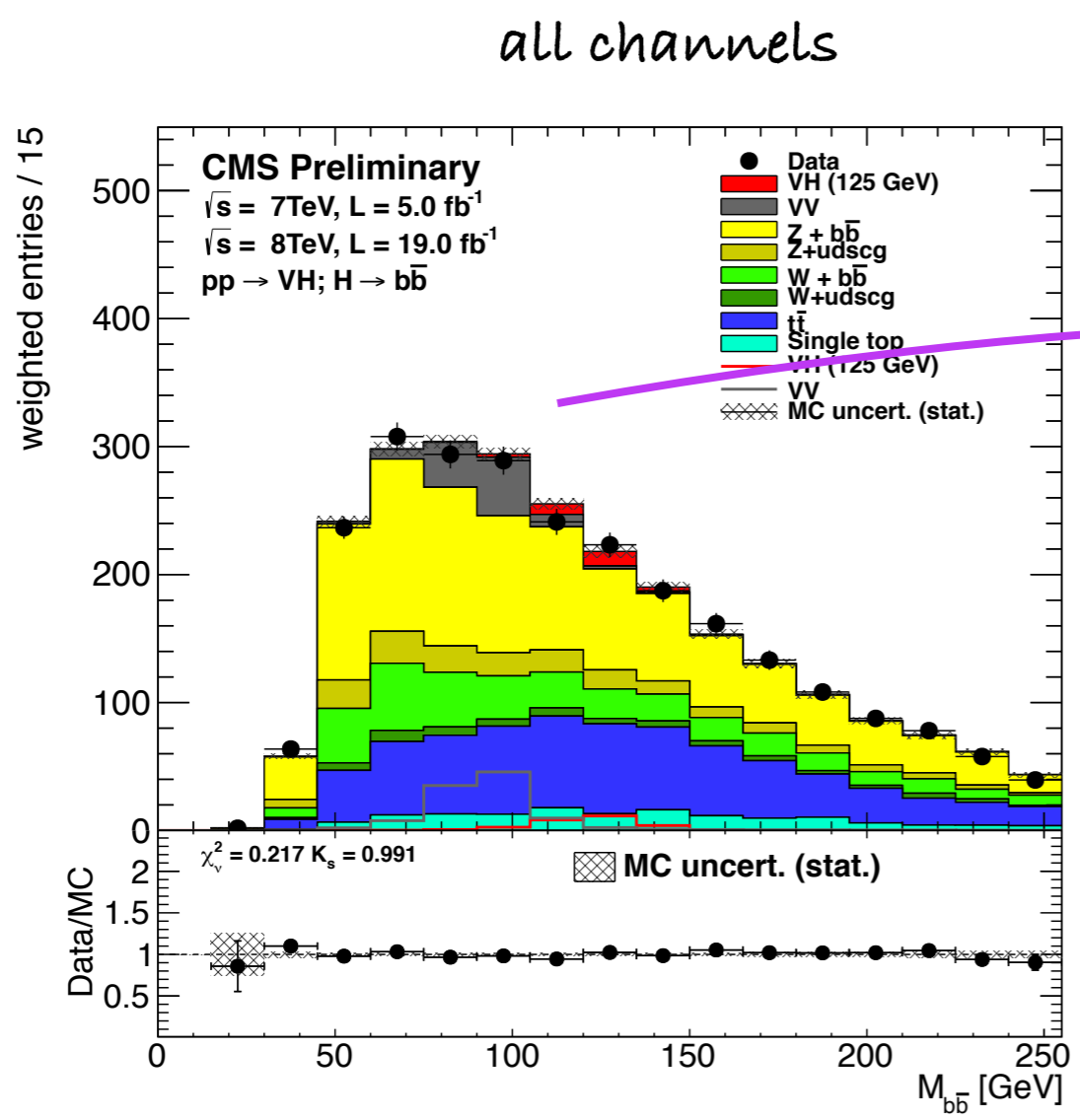




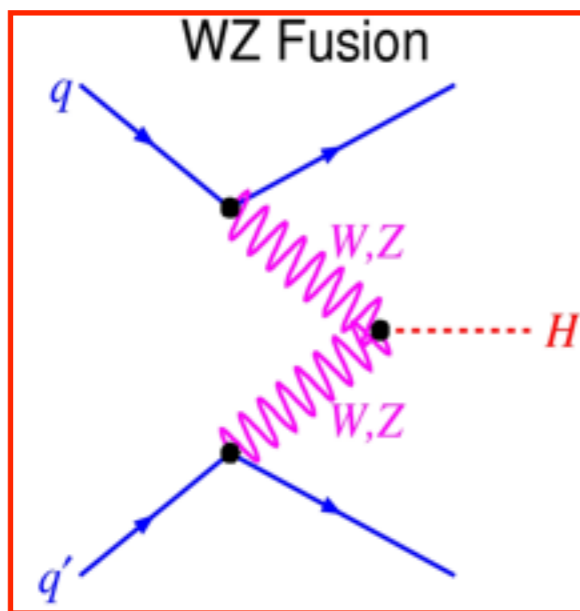
# VHbb $M_{bb}$ distribution



Michele de Gruttola



- Tighter selection than used in BDT analysis
- weight each category by  $s/(s+b)$  in 105-150 window
- consistent with diboson expectation + small excess in signal region



# VBFH(bb)

new comers in Hbb,  
testing Higgs properties



# VBFHbb Strategy



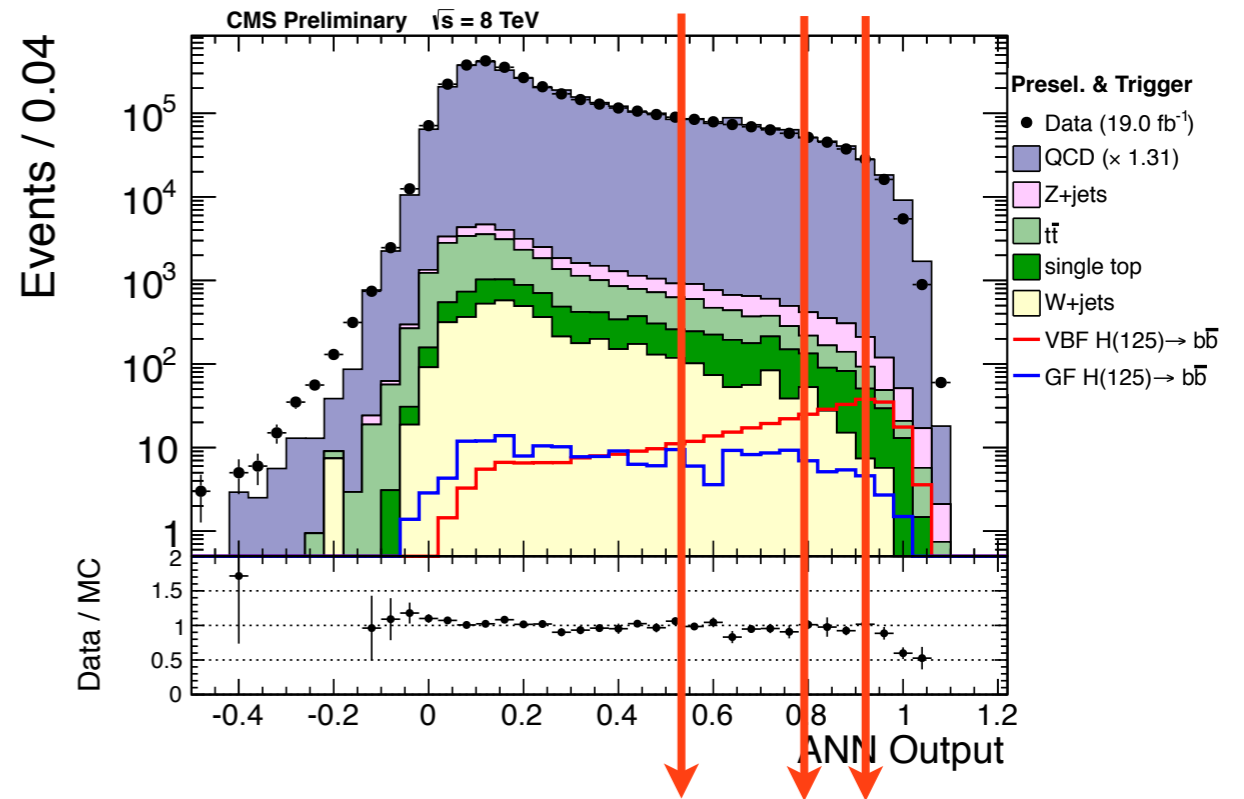
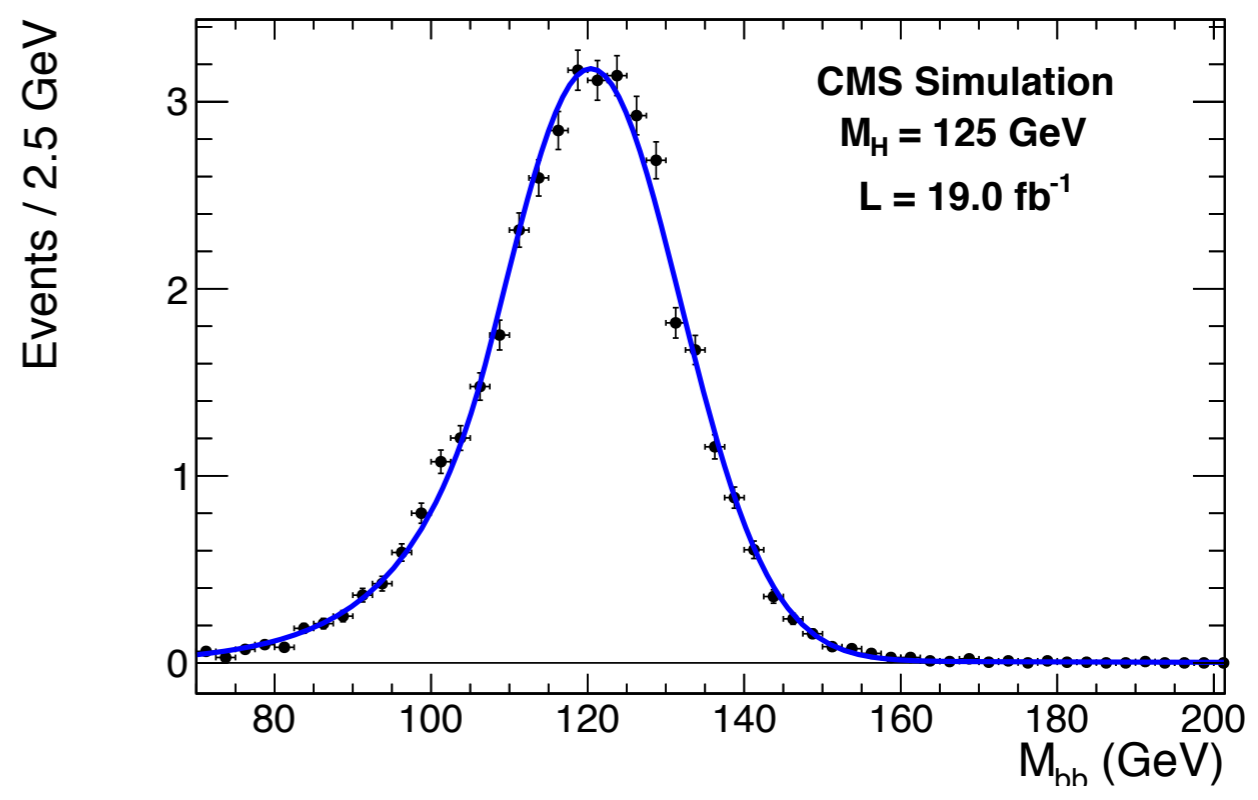
Michele de Gruttola

## ■ Fully hadronic final state (and trigger)

- 2012 entire 8TeV dataset
- QCD (4jets) is the main background
- b-tag (2jets) and VBF light jets

## ■ Best b-pair mass for extracting signal

- regression to improve resolution
- quark/gluon discriminant for back-forward jets up to  $|\eta| < 4.7$
- events categorized in 4 cat. based on ANN (no mass, no b-tag)
- fit on the  $m_{bb}$  falling spectrum



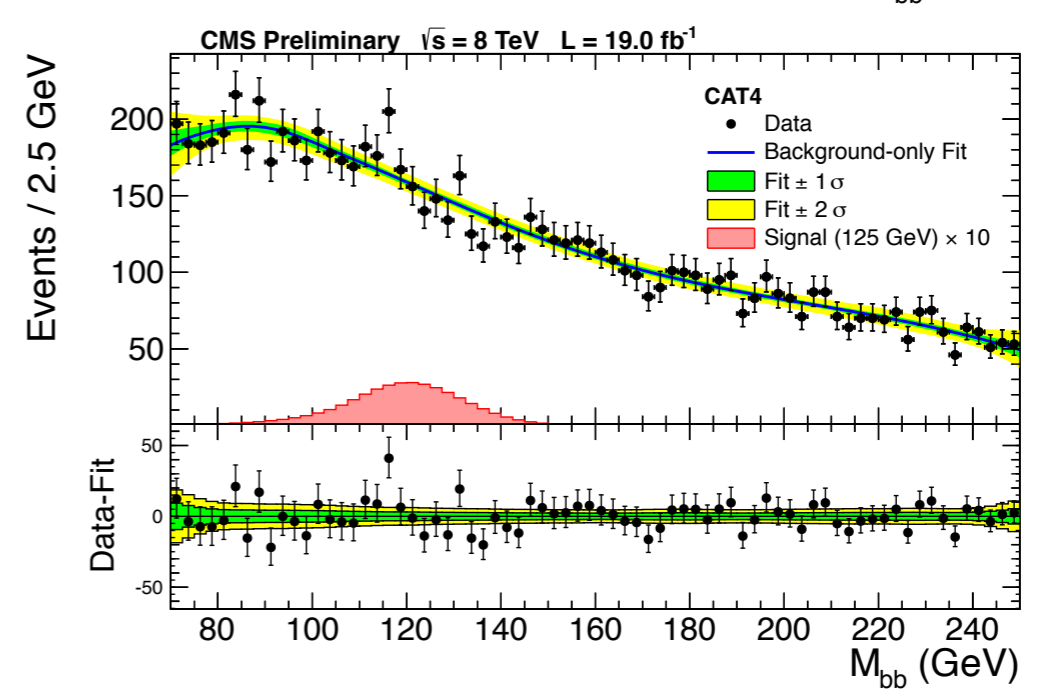
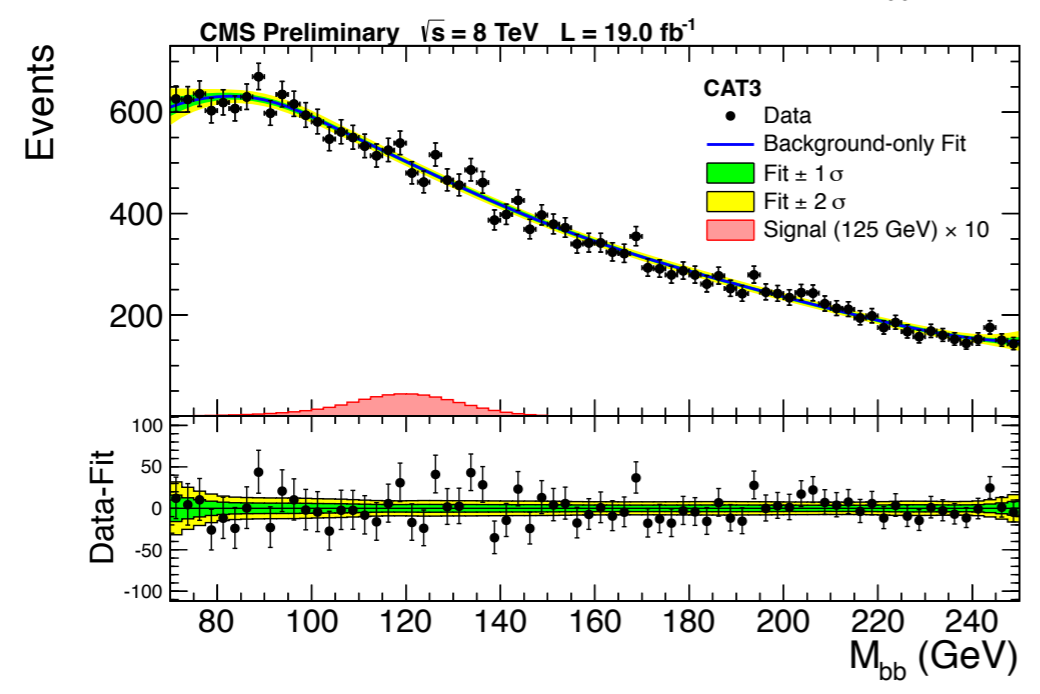
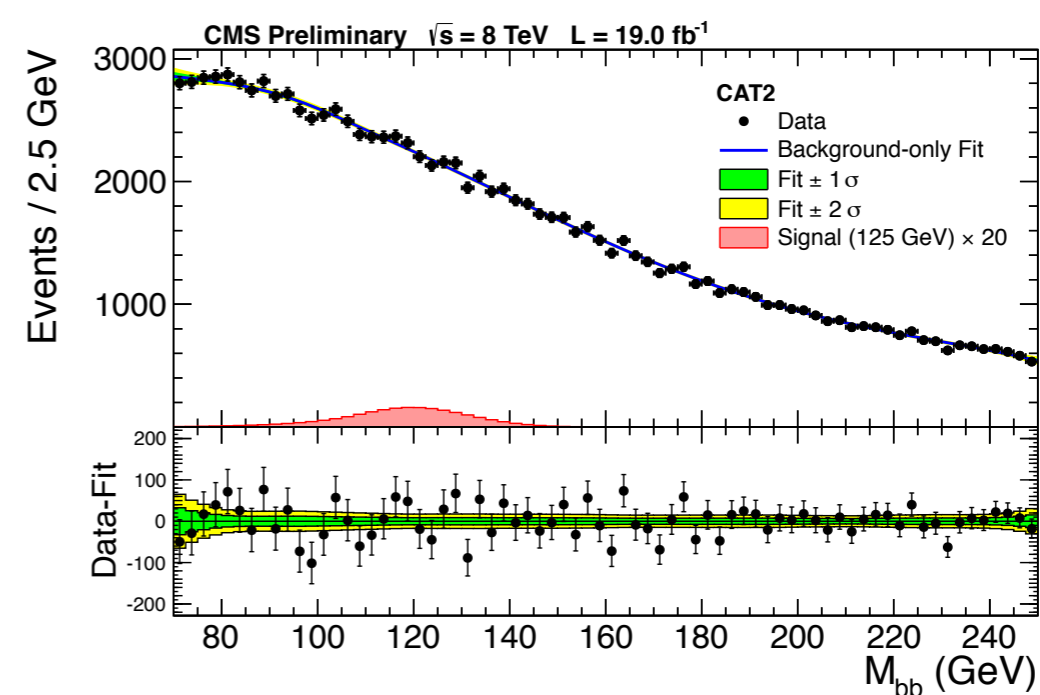
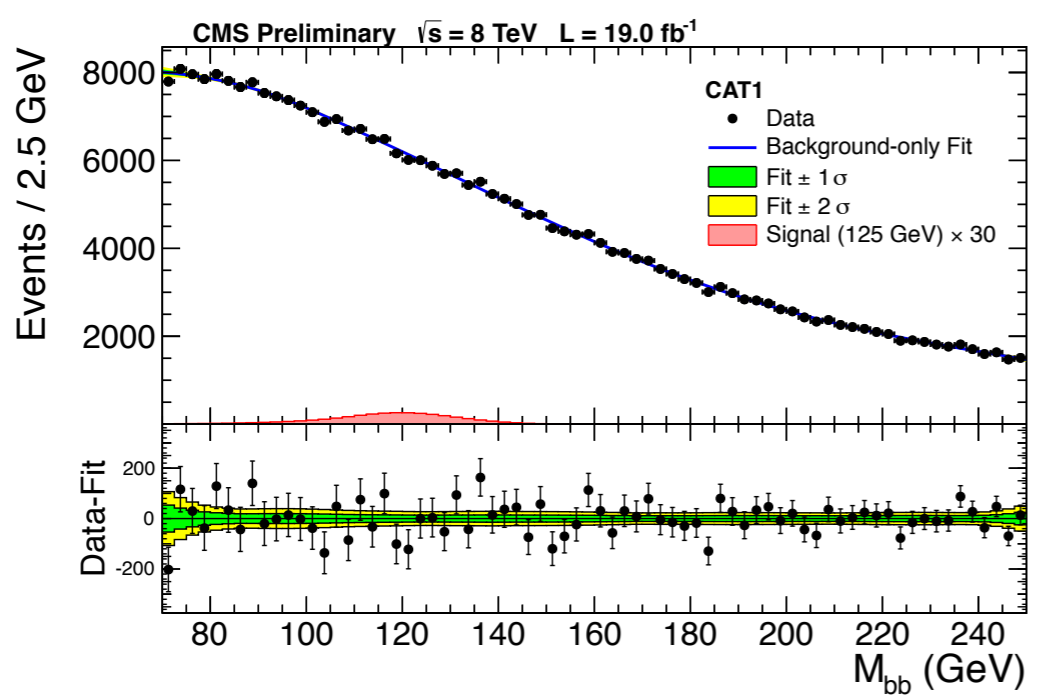


# VBFHbb Signal extraction



## ■ Likelihood fit on $m_{bb}$

○ using 5th order polynomial for QCD and MC templates for EWK and top



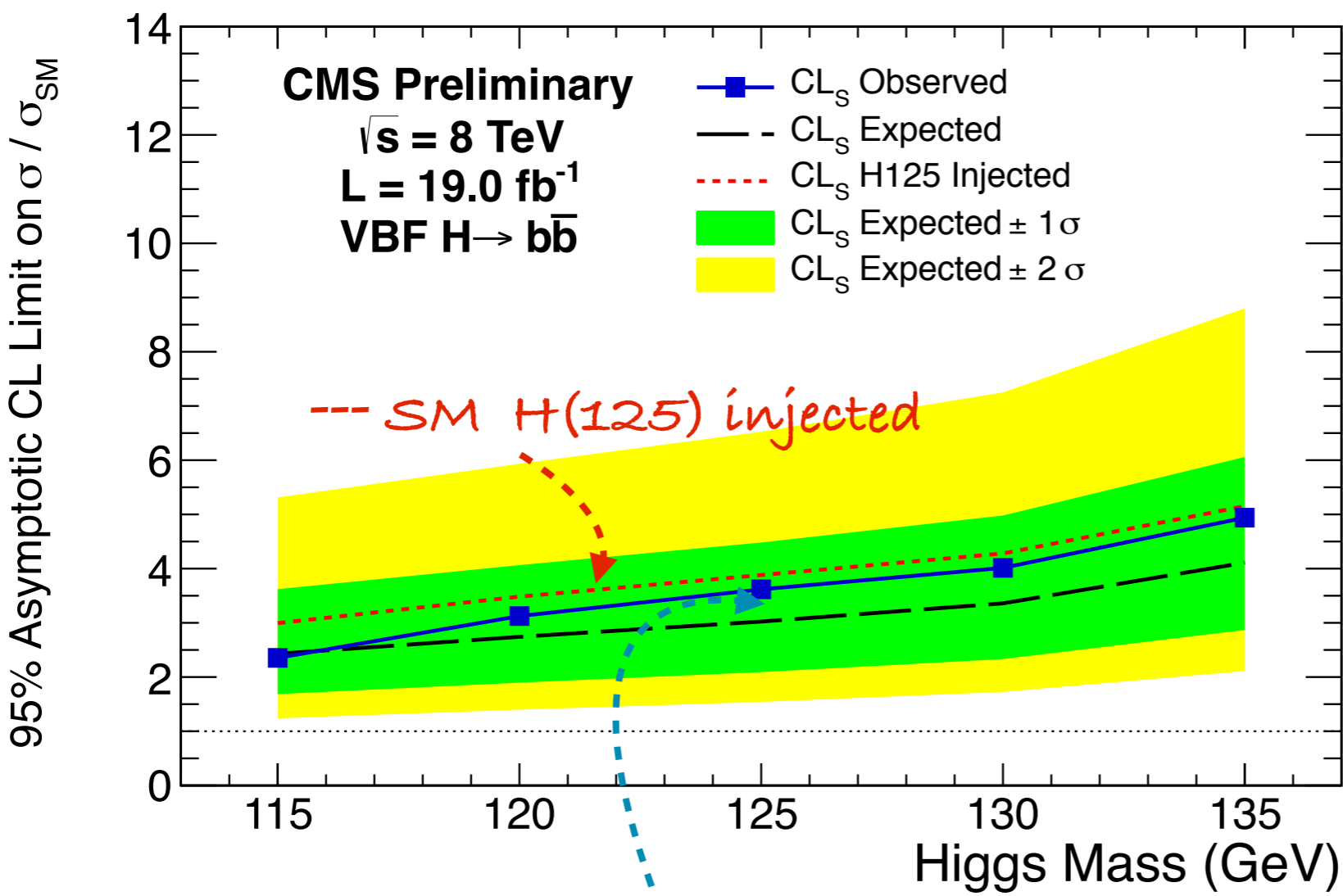
Michele de Gruttola



# VBFHbb Results



Michele de Gruttola



- Observed (expected) limit of **3.6(3.0)**xSM at 125 GeV
- Observed (expected) local significance of **0.5 $\sigma$  (0.7 $\sigma$ )** for  $m_H=125$  GeV
- Combined best-fit  **$\mu=0.7 \pm 1.4$**

*nice addition in the CMS Hbb program*

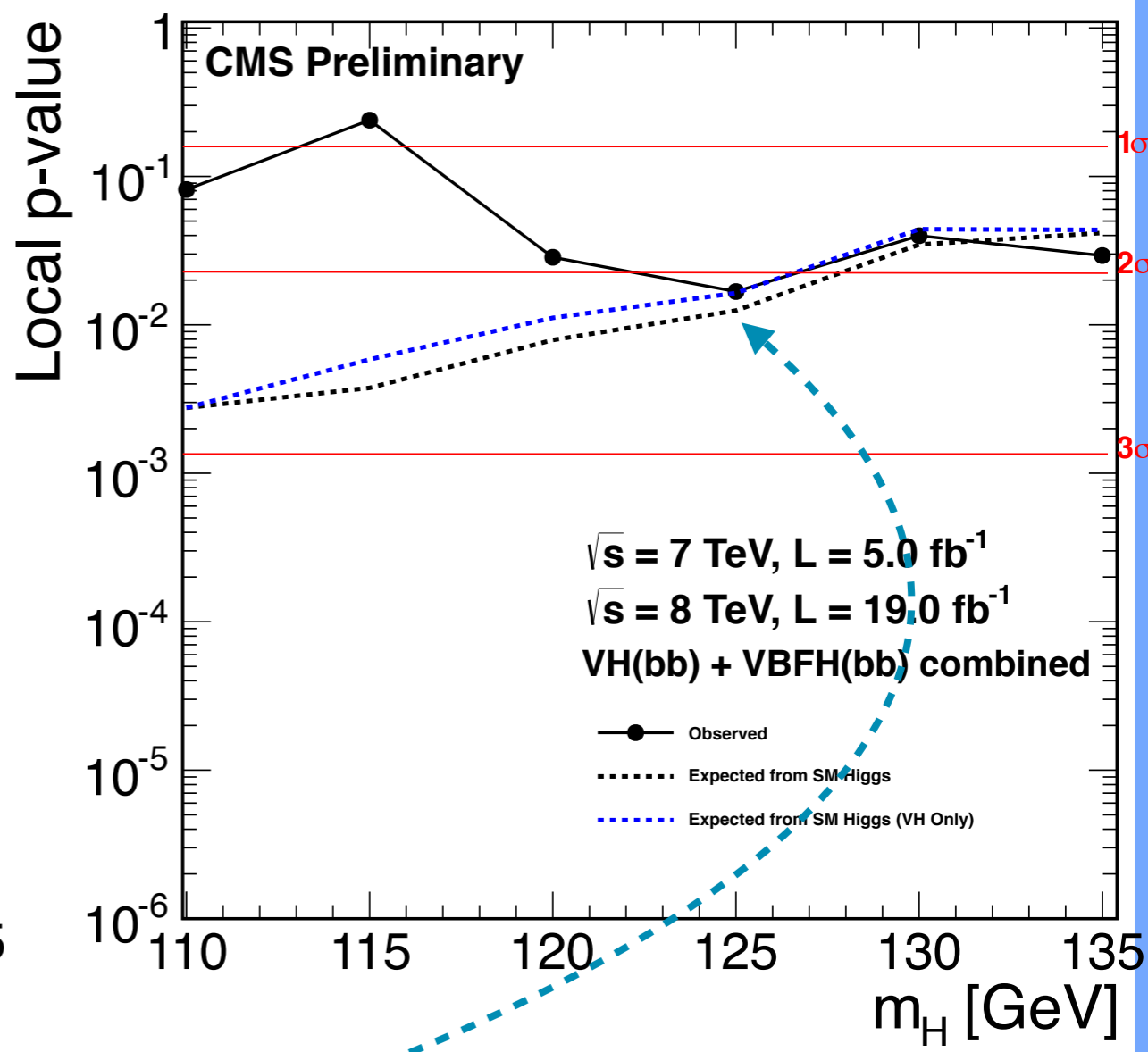
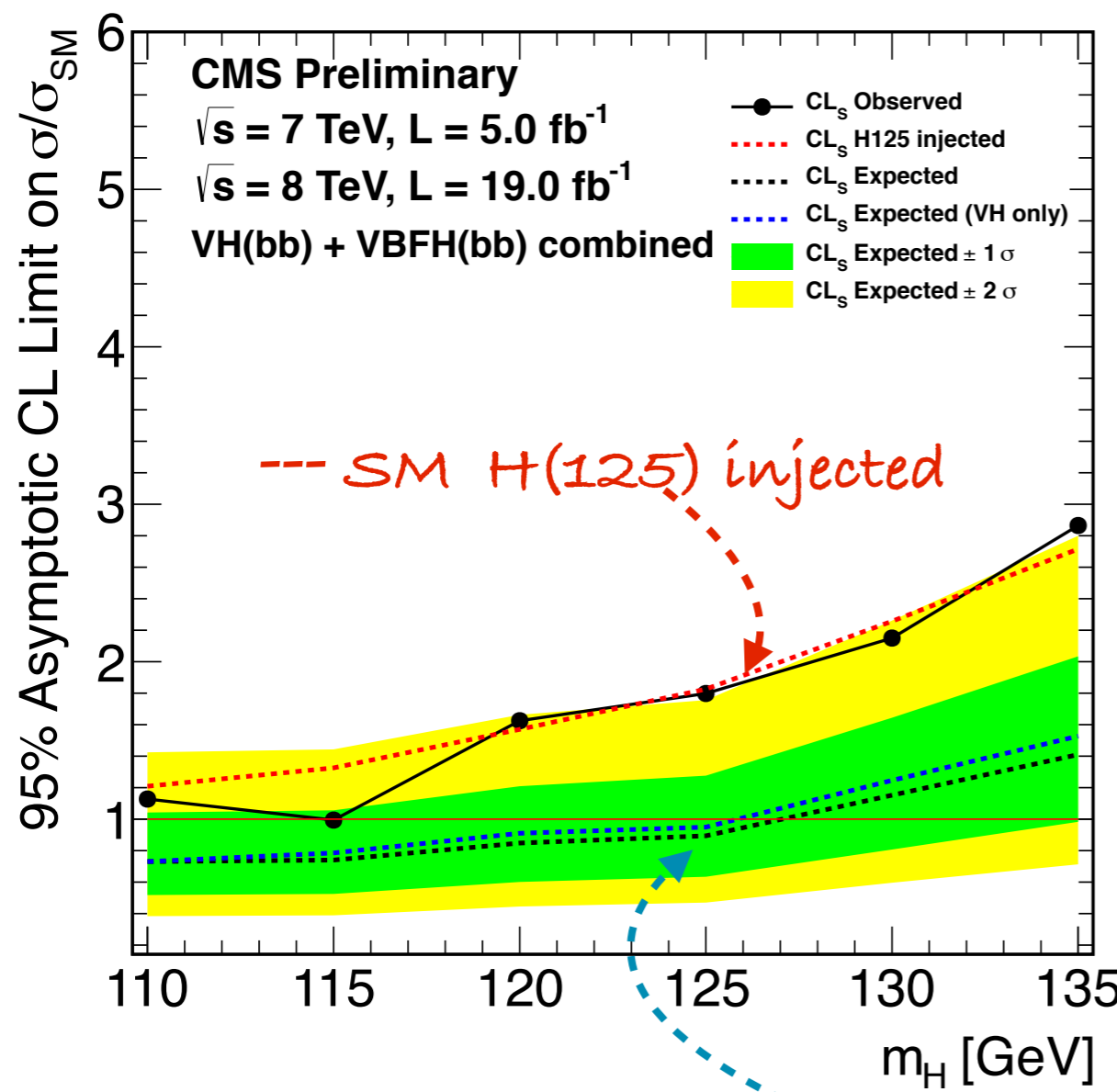




# VBFbb + VHbb combination



Michele de Gruttola



- Observed (expected) limit of 1.79(0.89)xSM at 125 GeV
- Observed (expected) local significance of 2.1 $\sigma$  (2.2 $\sigma$ ) for  $m_H=125$  GeV
- Combined best-fit  $\mu=0.97 \pm 0.48$

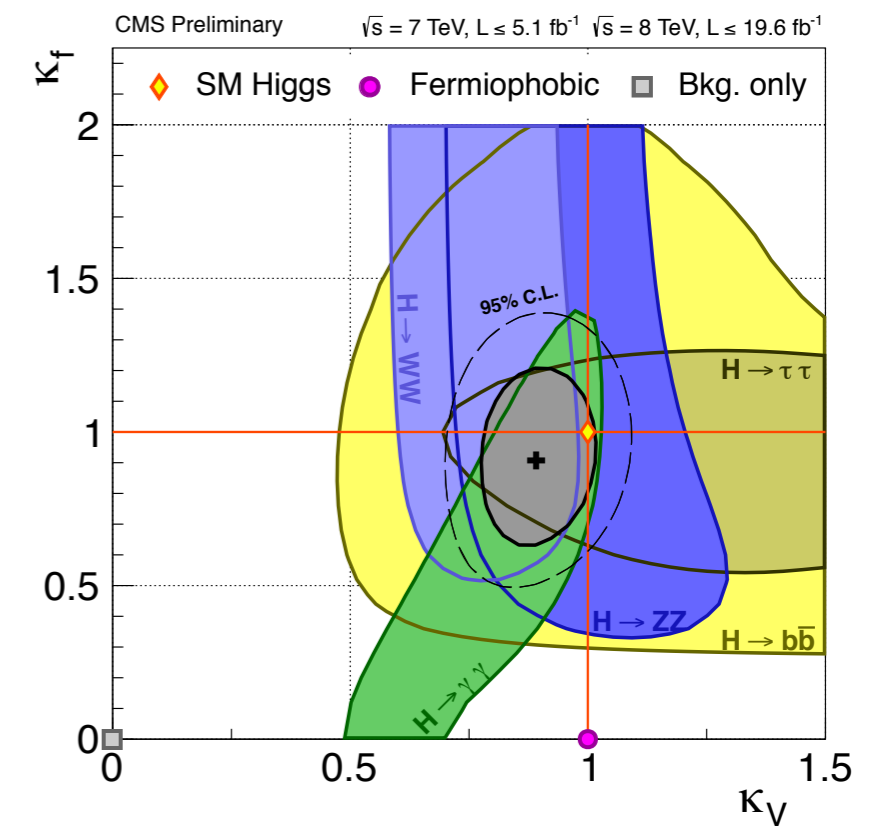
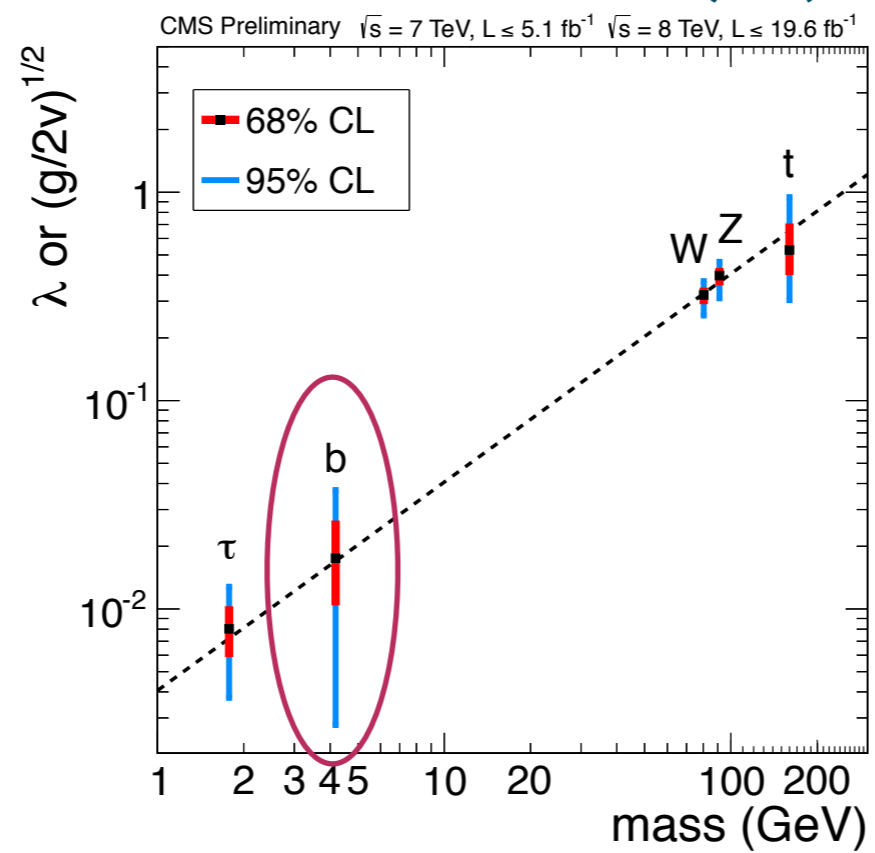
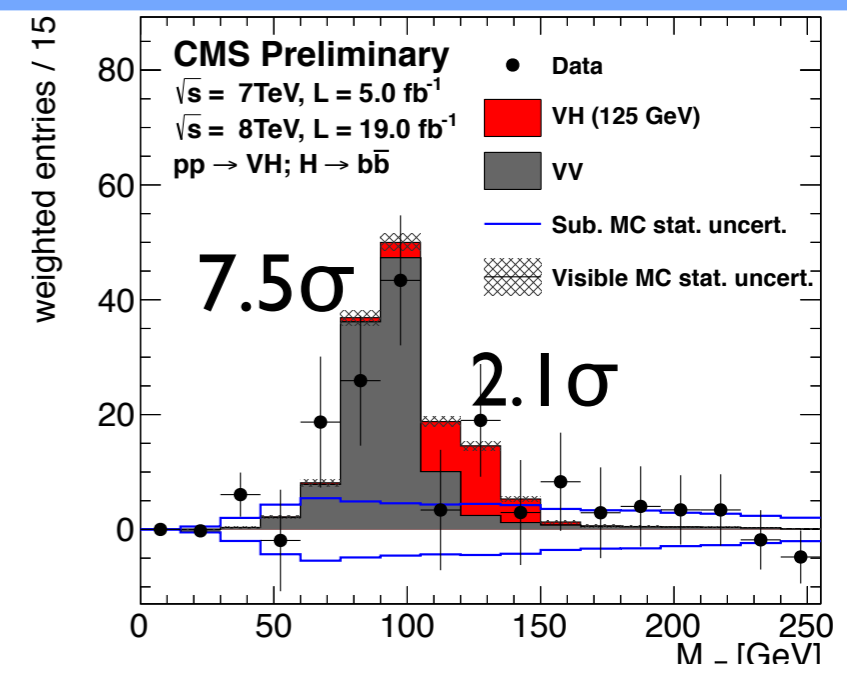
VBFbb add 5-10% to VHbb



# Conclusions



- CMS showed preliminary results on full 2011 + 2012 data ( $\sim 23 \text{ fb}^{-1}$ ) for VHbb and 8TeV dataset for VBFHbb
- $\geq 2\sigma$  excess in Hbb compatible with SM prediction
  - with signal strength  $1.0 \pm 0.5$
- Along with VHbb we measure a the VZ(bb) candle with  $7\sigma$



the more we analyze the data the more it looks like a SM Higgs!

Michele de Gruttola



# BACKUP

## Definition of Control Regions (CR) crucial element of the analysis

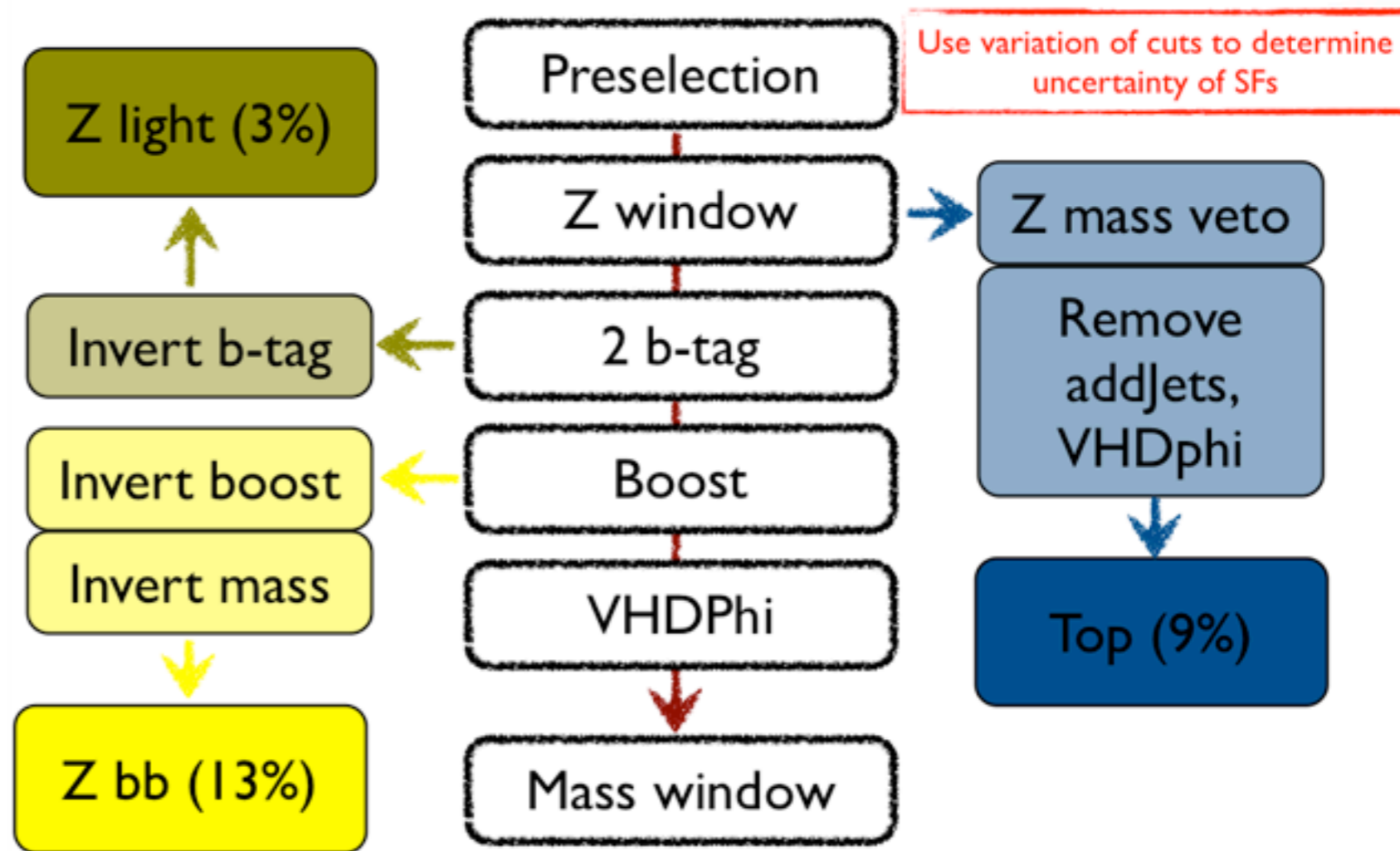
Define several CR enriched in different background components:

Control regions cuts as close as possible to the signal region

Renormalize MC simulation yields and extrapolate to signal region

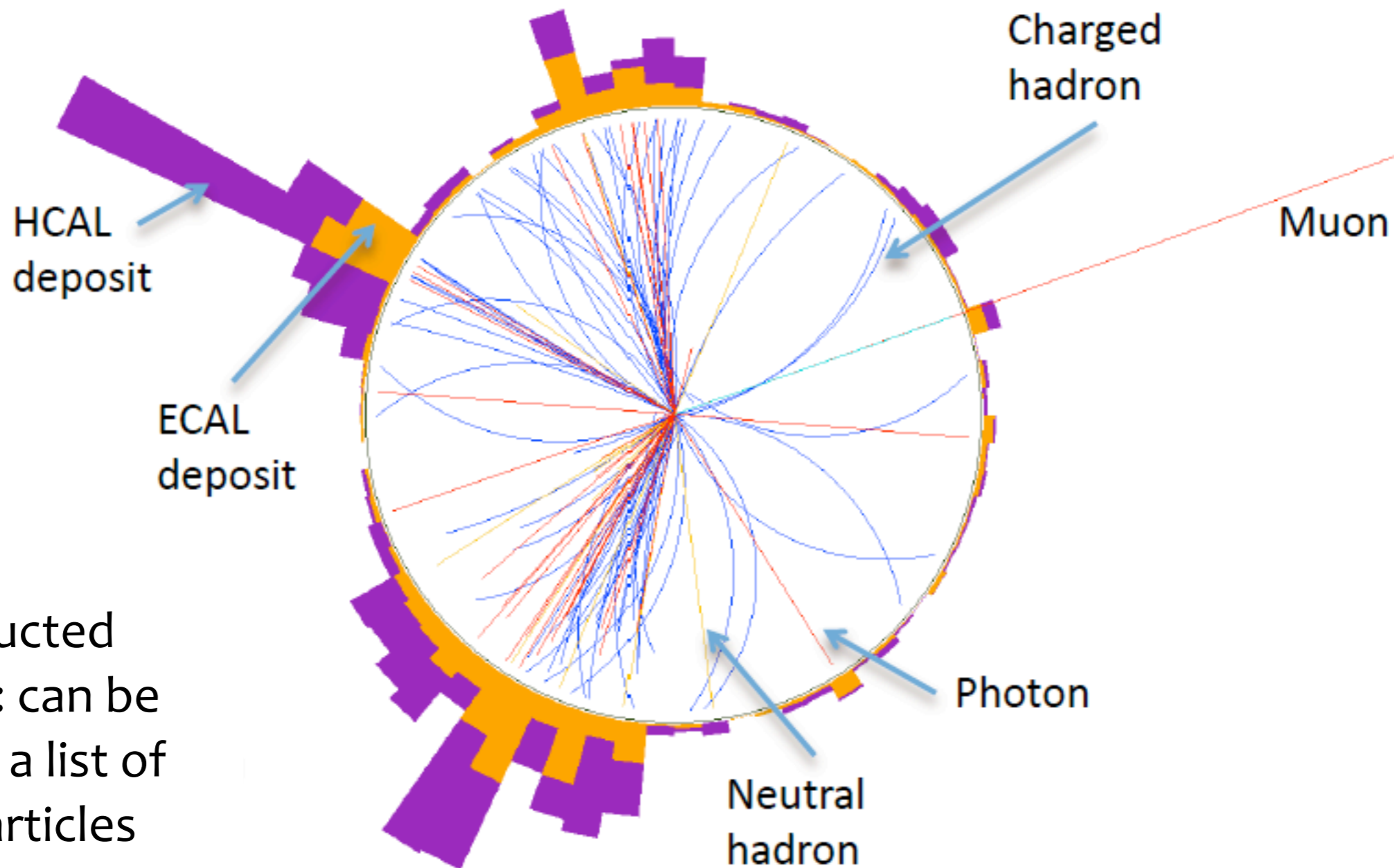
Account for associated stat. and syst. uncertainties

Example: Zee control region definition



(%)= uncertainty

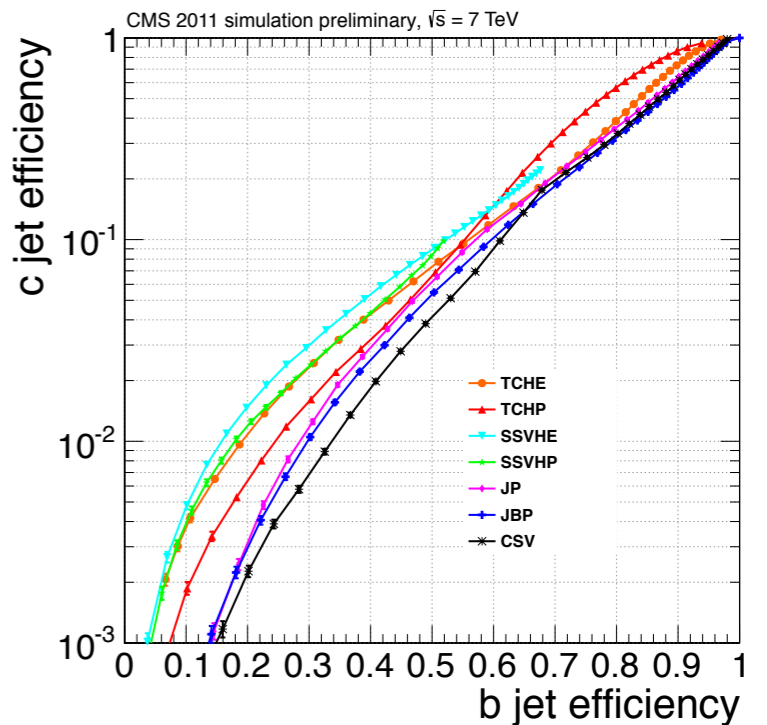
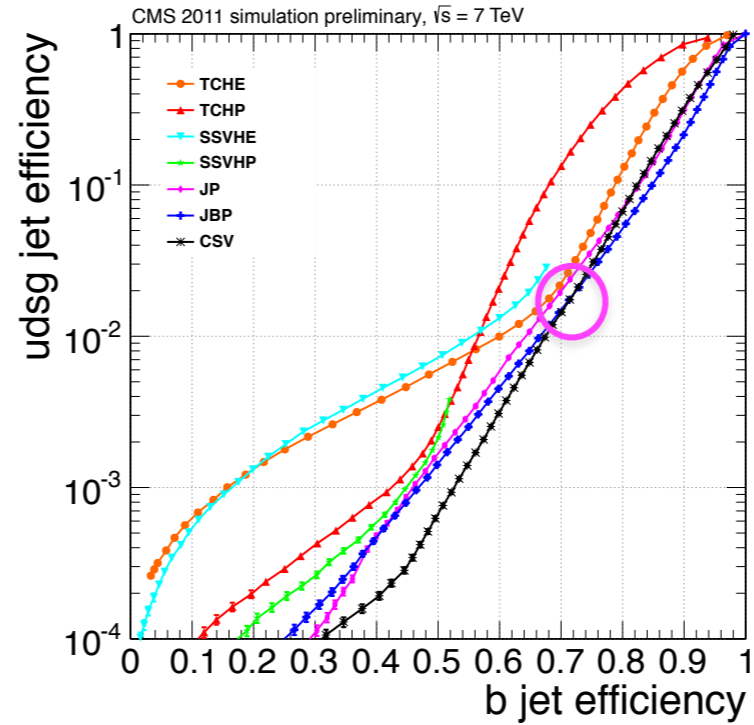
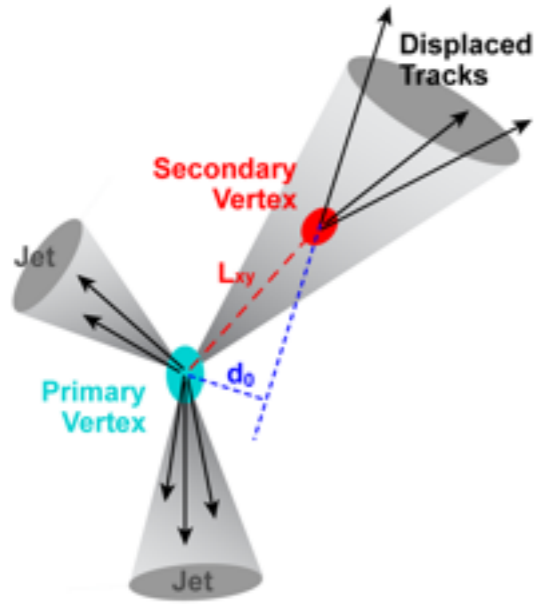
- CMS Particle flow reconstruction



List of reconstructed particles: can be used like a list of stable particles from a generator

# B-jets at CMS

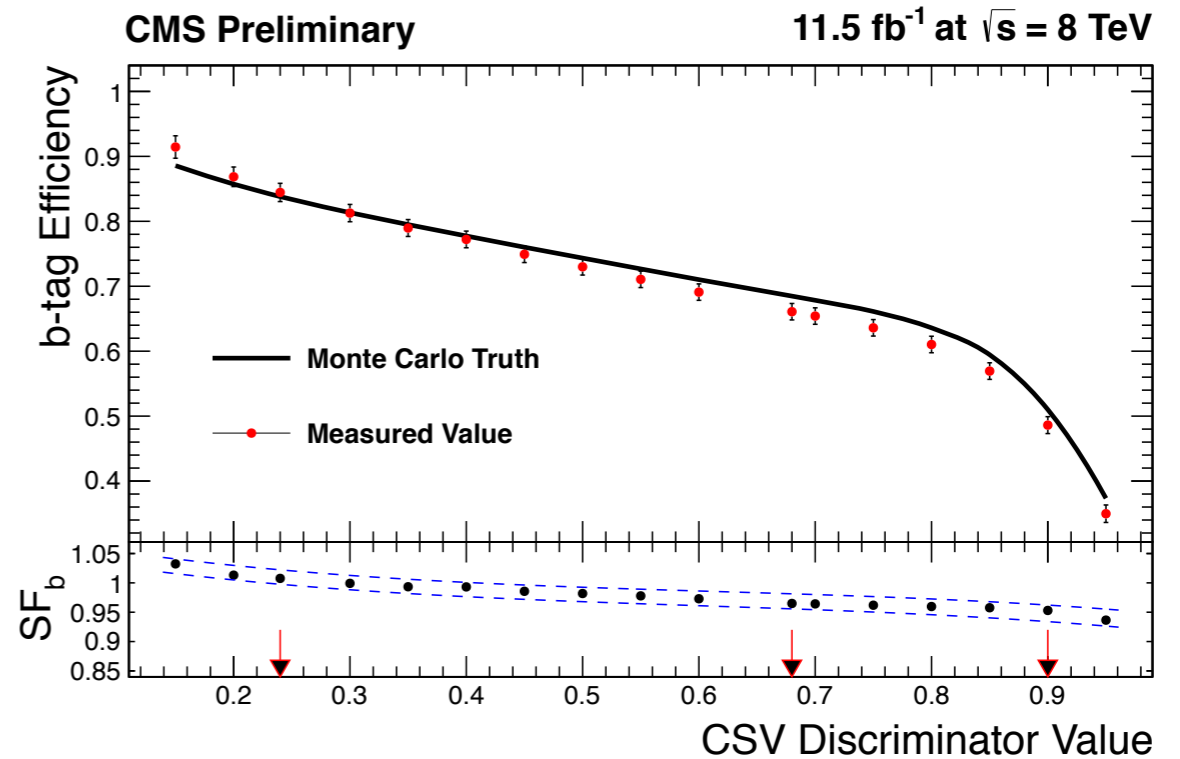
Michele de Gruttola



Eff of ~70% for a fake-rate of ~2%

- Many algorithms deployed at CMS
- Best separation from udscg and c jets
  - **CSV**: likelihood tagger using many jet properties: secondary vertex (if any), tracks impact parameters, etc.

eff and fake-rate from  $t\bar{t}$  samples and muon plus jets





# VHbb systematics



Source	Type	Yield uncertainty (%) range	Contribution to uncertainty (%)	Removal effect on total uncertainty (%)
Luminosity	normalization	2.2-4.4	< 2	< 0.1
Lepton efficiency and trigger (per lepton)	normalization	3	< 2	< 0.1
Z( $\nu\nu$ )H triggers	shape	3	< 2	< 0.1
Jet energy scale	shape	2-3	5.0	0.5
Jet energy resolution	shape	3-6	5.9	0.7
Missing transverse energy	shape	3	3.2	0.2
b-tagging	shape	3-15	10.2	2.1
Signal cross section (scale and PDF)	normalization	4	3.9	0.3
Signal cross section ( $p_T$ boost, EWK/QCD)	normalization	2/5	3.9	0.3
Monte Carlo statistics	shape	1-5	13.3	3.6
Backgrounds (data estimate)	normalization	10	15.9	5.2
Single-top (simulation estimate)	normalization	15	5.0	0.5
Dibosons (simulation estimate)	normalization	15	5.0	0.5
MC modeling (V+jets and $t\bar{t}$ )	shape	10	7.4	1.1



- total effect of 15% on limits/significance



# VBFbb systematics



Source	Uncertainty
Background fit	depending on the statistics of each category
Z+jets cross section	$\pm 20\%$
top cross section	$\pm 20\%$
Signal and Z peak position (JES)	$\pm 1.5\%$
Signal and Z resolution	$\pm 10\%$
Luminosity	$\pm 4.4\%$
Trigger efficiency	$\pm 5 - 8\%$
Signal acceptance due to JES	$\pm 10\%$
Signal acceptance due to JER	$\pm 2\%$
VBF cross section	$\pm 3\%$
VBF Monte Carlo acceptance	$\pm 10\%$
PDF	$\pm 5\%$
VBF ANN shape due to b-tag	$\pm 2\%$
VBF ANN shape due to quark-gluon discriminator	$\pm 2\%$
VBF ANN shape due to UE modeling	$-8 - +2\%$
GF cross section	$\pm 15\%$
GF Monte Carlo acceptance	$\pm 50\%$
GF ANN shape	$\pm 50\%$

- total effect of 20% on limits/significance
  - *~20% on bkg and 15% on signal MC*





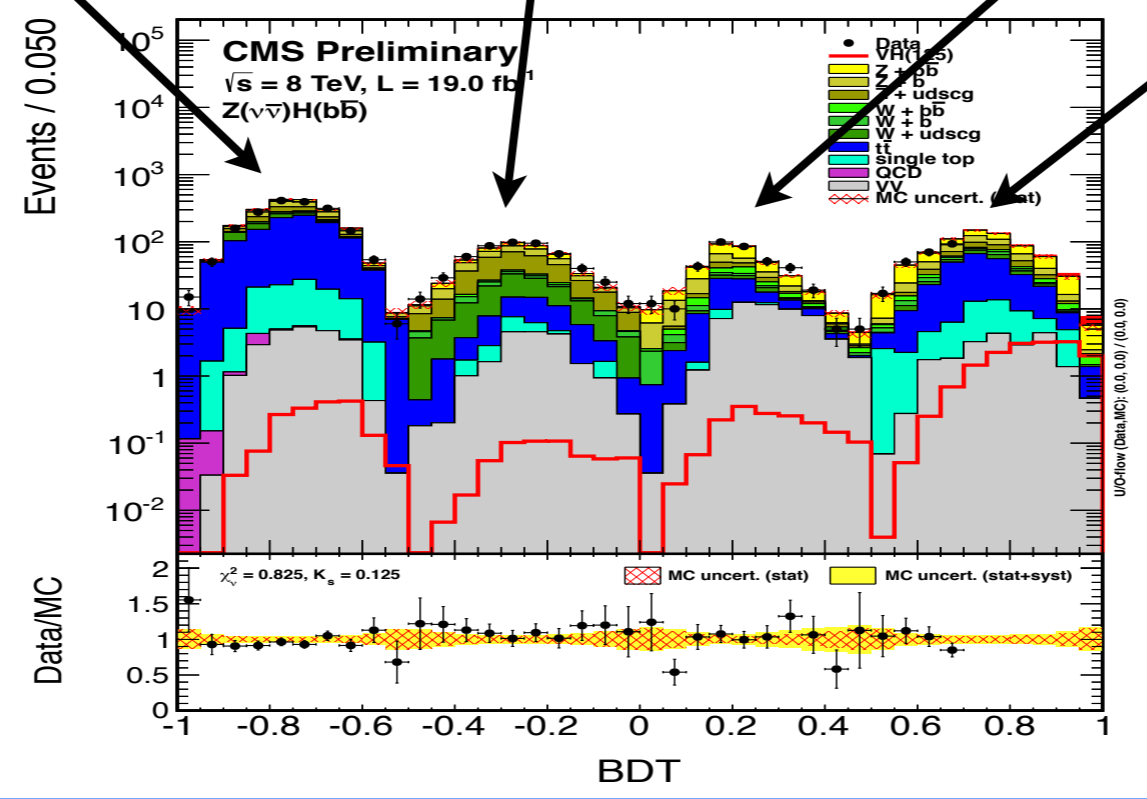
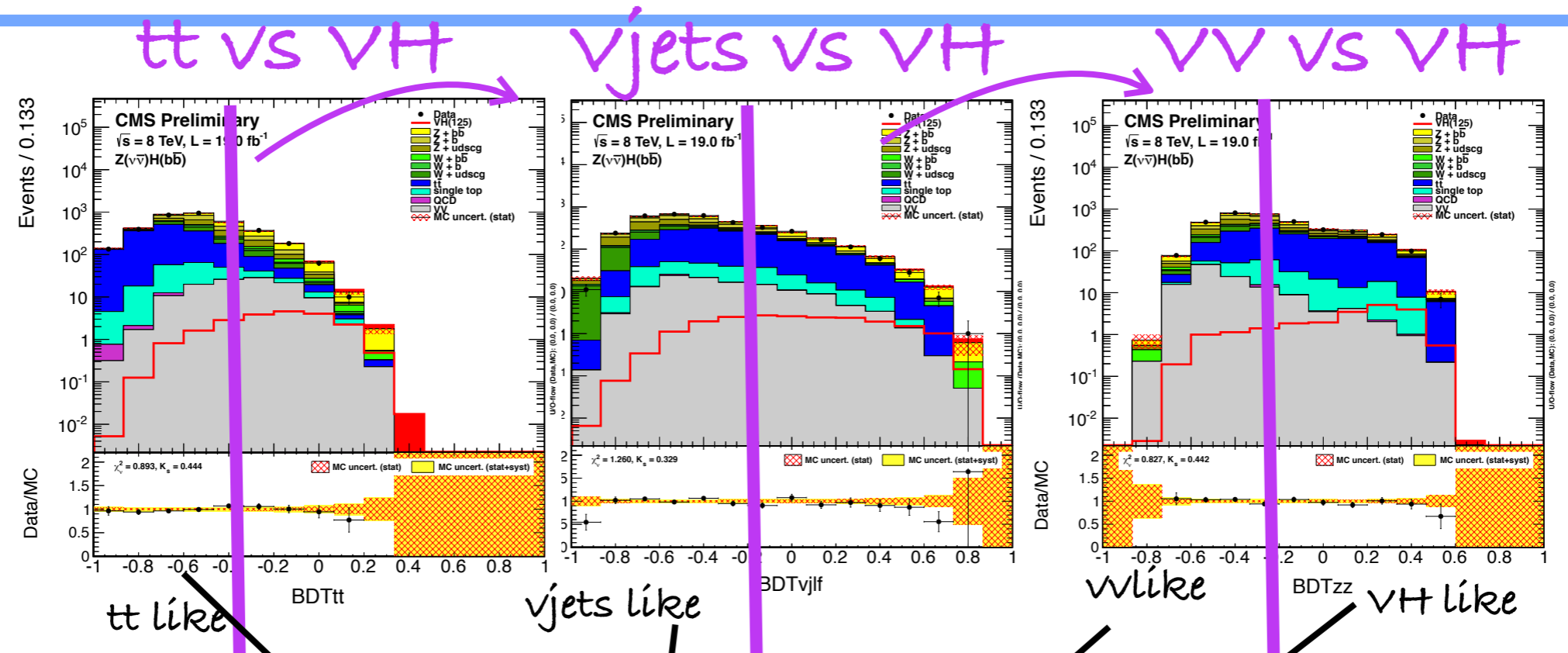
# Bkg specific BDT



- Use all events [after evt sel] to train main BDT to distinguish VH from all BCK
- Train tt-specific BDT with all tt and VH events and pick cut value.
  - Run all events from all processes through tt-BDT and for those below tt cut value, plot the value of main BDT in the left-most partition of final plot, after scaling transform  $[-1,1] \Rightarrow [-1.,-0.5]$
- Train Vjets-specific BDT with all Vj and VH events that pass tt-BDT cut and pick cut value
  - Run all events from all processes, that pass tt-BDT cut, through Vj-BDT and for those below Vj cut value, plot the value of main BDT in the second partition, from left, of final plot, after scaling transform  $[-1,1] \Rightarrow [-0.5,0]$
- Train VV-specific BDT with all VV and VH events that pass tt-BDT and Vj-BDT cuts and pick cut value
  - Run all events from all processes, that pass tt-BDT and Vj-BDT cuts, through VV-BDT and for those below VV cut value, plot the value of main BDT in the third partition, from left, of final plot, after scaling transform  $[-1,1] \Rightarrow [0,0.50]$
- Run all events from all processes, that pass tt-BDT, Vj-BDT and VV-BDT cuts, through main BDT and plot value in the right-most partition of final plot, after scaling transform  $[-1,1] \Rightarrow [0.5,1.0]$



# Bkg specific BDT



through main BDT

Michele de Gruttola