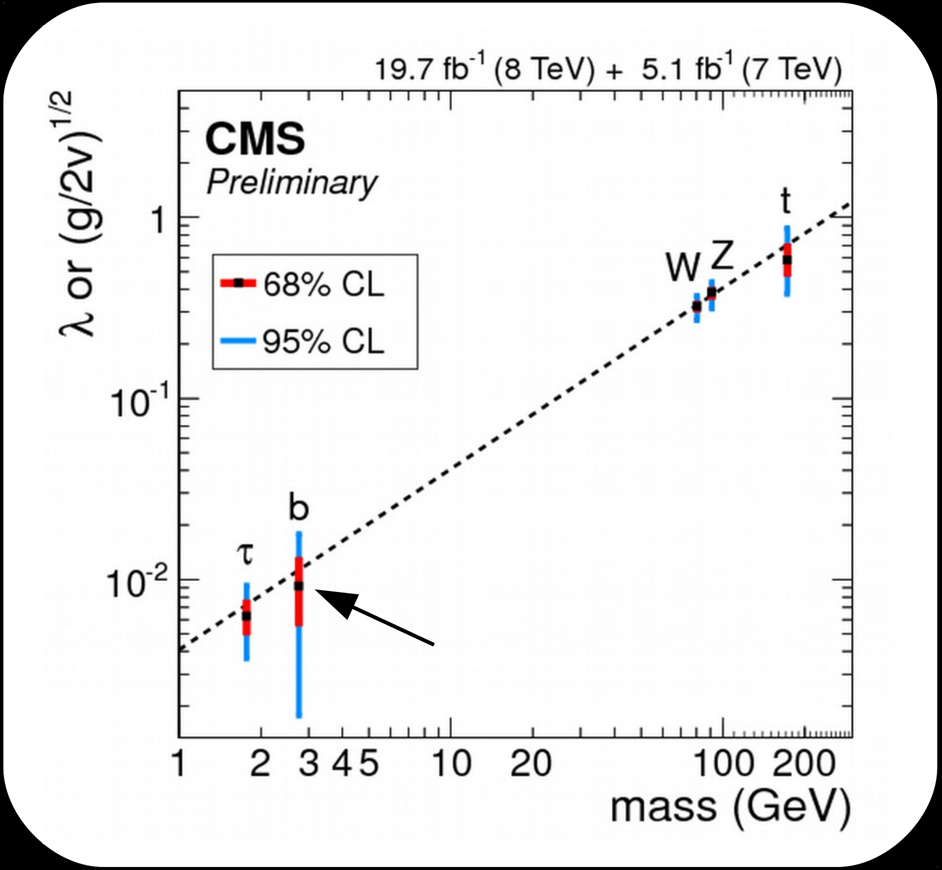
Search for the SM Higgs boson decaying to b quarks with the CMS experiment

Silvio Donato (CERN, SNS, INFN Pisa)

for the CMS Collaboration

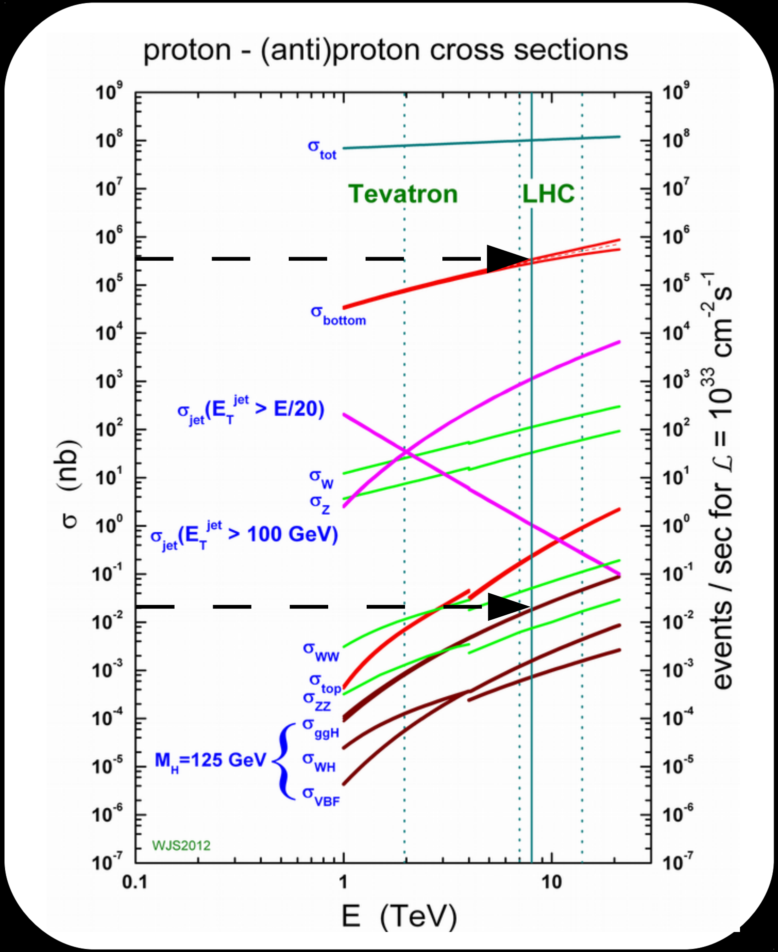
Motivations

- The observation of the Higgs boson decay to b quarks is one of the missing pieces in understanding of the new particle.
- In particular, an anomaly in the Higgs boson coupling would be a hint for New Physics.



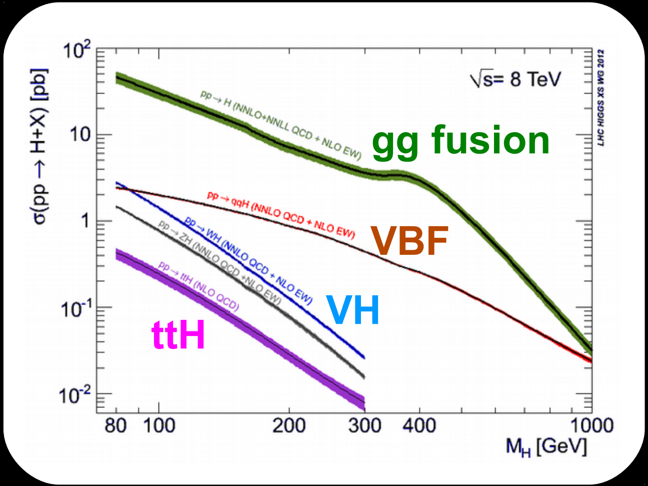
$H \to b\bar{b}$

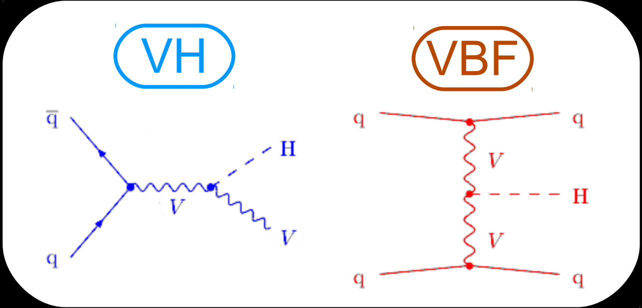
- Although the Higgs boson decay to b quarks have a large branching ratio (~60%) it is very hard to observe the $pp \to H \to b\bar{b}$ process.
- In fact, the QCD background $pp \to b\bar{b}$ is more than 10^7 times larger.



Production channels

- To cope with the QCD background it is useful to exploit the signal topology of:
  - the associated production of the Higgs boson with W/Z boson (**VH**);
  - the Vector Boson Fusion production (**VBF**).
- The VH channel, when the W/Z boson decays leptonically, is characterized by:
  - the presence of two, one or zero isolated energetic charged leptons from the W/Z decay;
  - high missing transverse energy (MET) when the W/Z decay to neutrino(s).
- The VBF channel is marked by:
  - two energetic quark-jets with high mass and η separation;
  - low hadronic activity between them, except for the two jets from Higgs decay.
- In addition, both channel have the two b-jets from the Higgs boson decay.

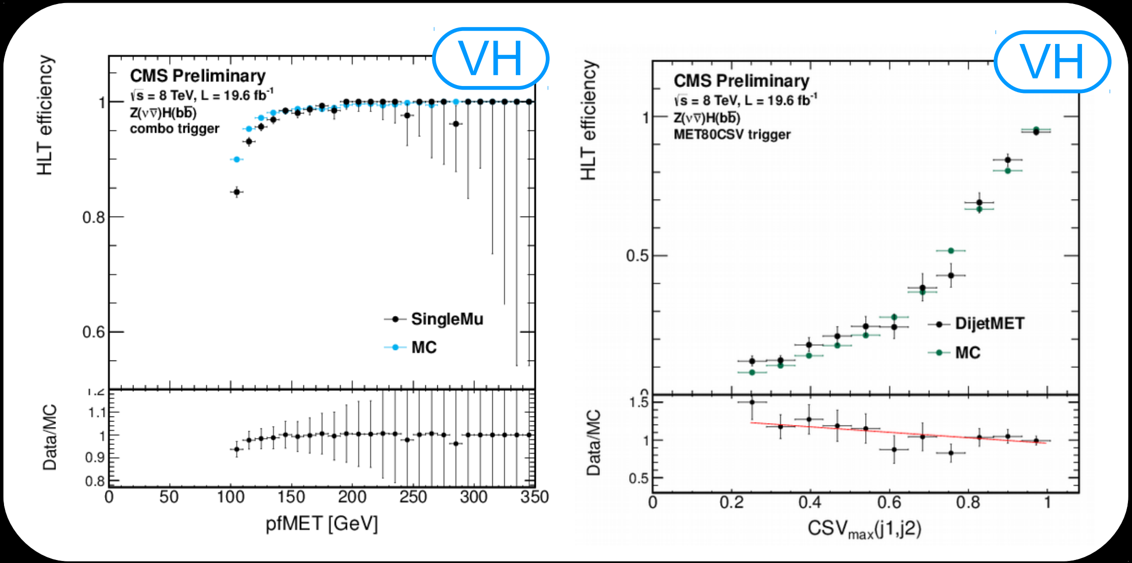




Trigger

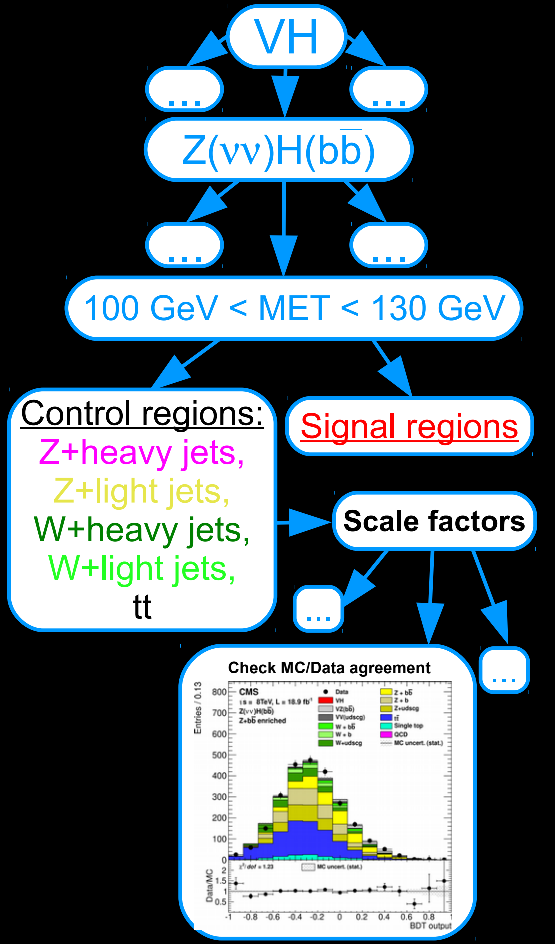
- For the **VH** channel the triggers are based on the presence of the isolated lepton(s), missing transverse energy and jets.
  - To reduce the rate of the Z(νν)H(bb) trigger, it exploits also the on-line b-tagging.

- For the **VBF** channel a set of triggers has been developed in order to cover the signal phase space. They are based on the VBF topology and the on-line b-tagging.



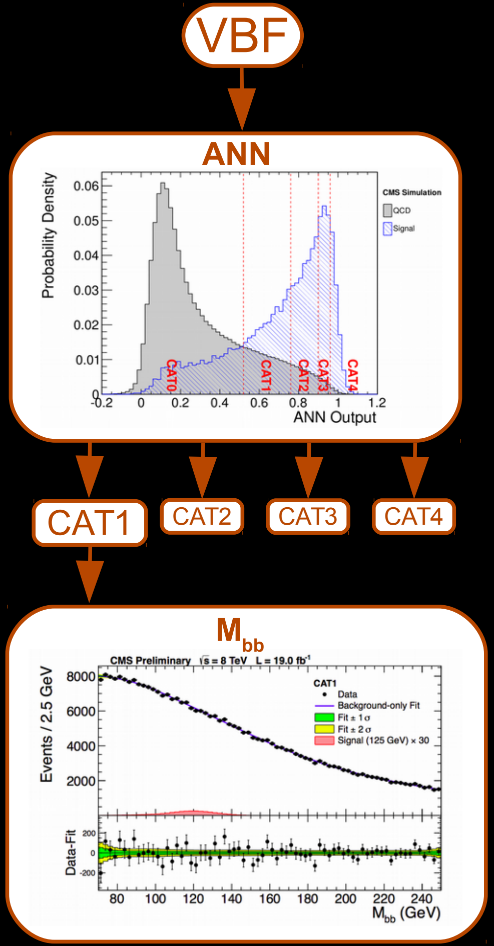
Signal region and background in VH

- In VH the main backgrounds are: W/Z + jets and $t\bar{t}$ productions.
- They are estimated using:
  - Shapes ⟶ simulations; Normalizations ⟶ data-driven.
- The analysis is divided in six channels:
  - W ⟶ eν, μν, τ(1-prong); Z ⟶ ee, μμ and νν.
- Each channel is binned in two or three bins depending on the vector boson p_T.
- In each bin a signal region is defined cutting on: the jet kinematic variables and b-tag discriminants, the lepton momenta, the MET and the number of additional leptons and jets.
- Up to five control regions ($t\bar{t}$, W/Z + heavy/light jets) are defined in each bin inverting some cuts:
  - here scale factors to apply to the main background normalizations (W + 0/1/2 b-jets; Z + 0/1/2 b-jets; $t\bar{t}$) are obtained with a simultaneous fit.

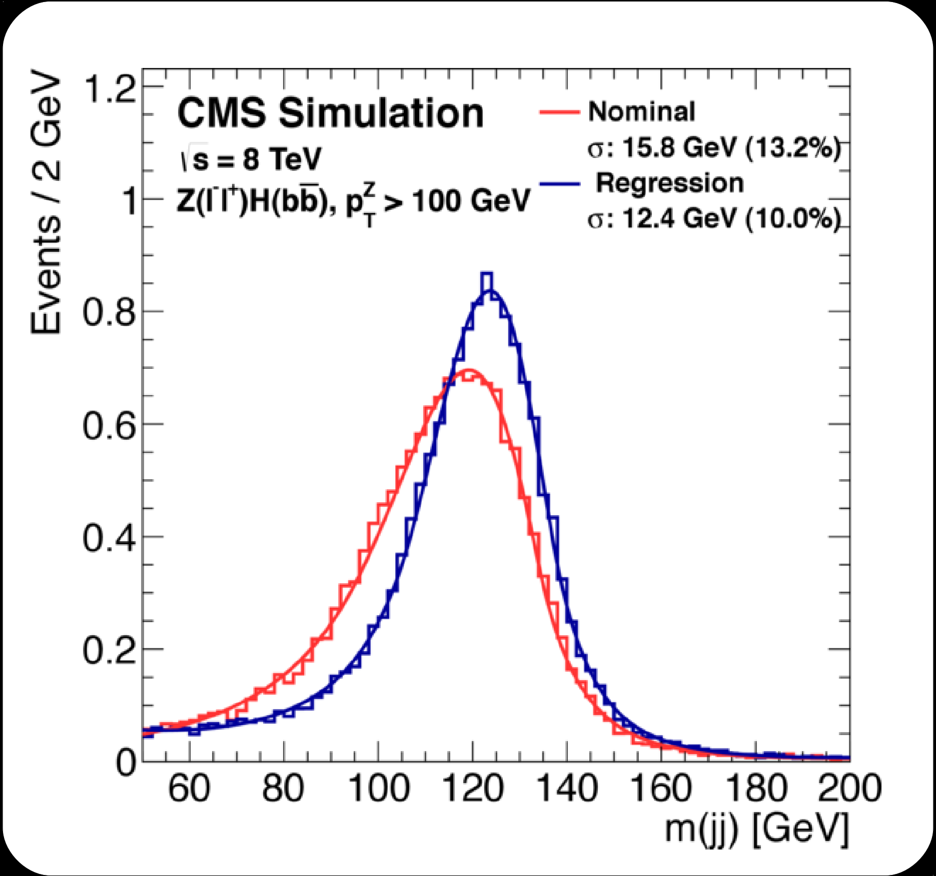


SIGNAL REGION AND BACKGROUNDS IN VBF

- In **VBF** the main background is the multi-jet QCD production and it is estimated by data.
- Additional minor backgrounds are Z/W+jets, $t\bar{t}$ and single top.
- An Artificial Neural Network (ANN) is trained to separate the signal from backgrounds. With the exception of the b-jets kinematics, it exploits the most discriminant variables:
  - $\Delta\eta$ between the most forward/backward jets, b-tagging discriminants and additional hadronic activity in the event.
- The ANN discriminator is used to define five categories of events with different signal/background ratio.
- In each category the Higgs candidate mass distribution will be fitted using:
  - a fifth degree Bernstein polynomial as QCD shape
  - Z/W and top templates from simulation.



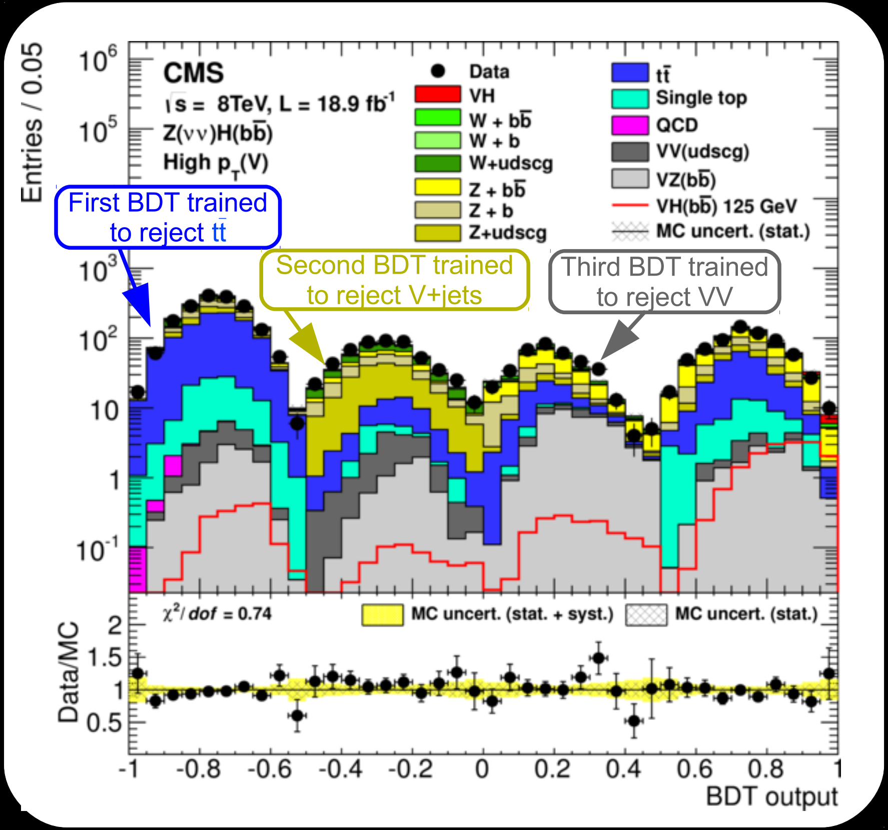
Regression



- Both analyses use a jet energy regression in order to improve the Higgs candidate mass resolution.
- The regression attempts to recalibrate the jet p_T to the true p_T of the particle jet.
- The regression exploits variables like:
  - b-tagging discriminants and secondary vertex informations;
  - jet-energy fractions between charged/neutral hadron and photons, muons and electrons;
  - jet kinematic variables (pT, η);
  - missing transverse energy (only for ZH → llbb and VBF)
- The result of the regression is an improvement on the Higgs mass resolution of about **10 − 20%**.

Multi BDT (VH)

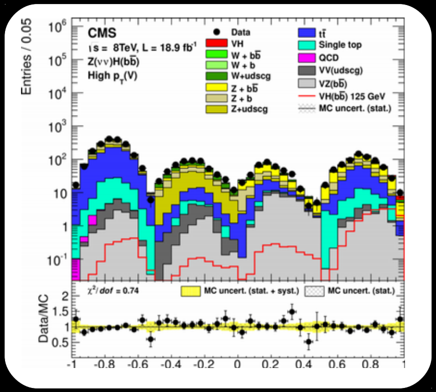
- In the VH channel to reduce the background in the signal region three specialized BDT are trained to reject the $t\bar{t}$, V+jets and VV backgrounds.
- Each BDT is trained with:
  - Higgs candidate mass and pT;
  - b-tagging discriminants;
  - lepton momentum and MET;
  - number of adjunctive leptons and jets;
  - other kinematic variables.
- The three BDT are used to obtain a signal region with an higher signal/background ratio.
- A final BDT is trained against all backgrounds in order to do the final fit.



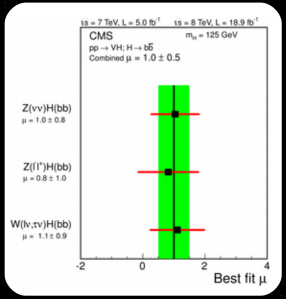
A cross check: the Z peak

- The final fit will be:
  - For VH: a fit, for each bin, of the multi-BDT output using the shapes from simulation and data-driven scale factors;
  - For VBF: a fit, for each category, of the bb dijet mass distribution taking the signal shape from simulation, the QCD shape from the fit of a fifth degree Bernstein polynomial and the Z/W and top shapes from simulation.
- A natural cross check is to measure the peak due to Z ⟶ bb, instead of H ⟶ bb.
- Here, the Z(bb) signal strengths measured are:
  - VH: $\mu = 1.09^{+0.32}_{-0.29}$ (stat. + syst.) ⟶ **Eur. Phys. J. C 74, 2973 (2014)**
  - VBF: $\mu = 0.99 \pm 0.12$ (only stat.)

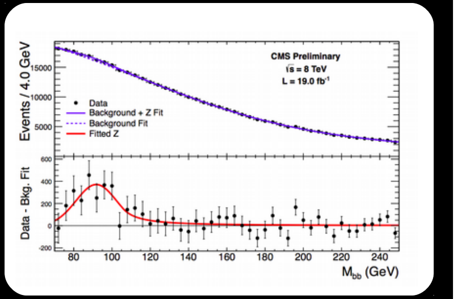
VH



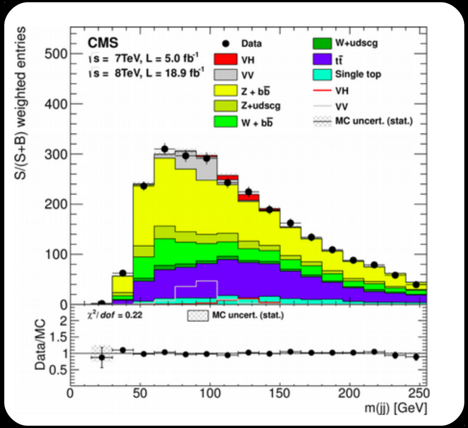
VH

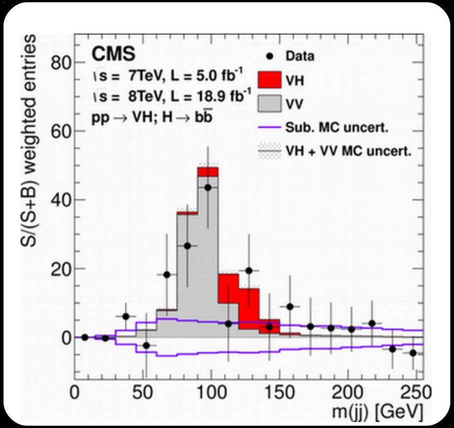


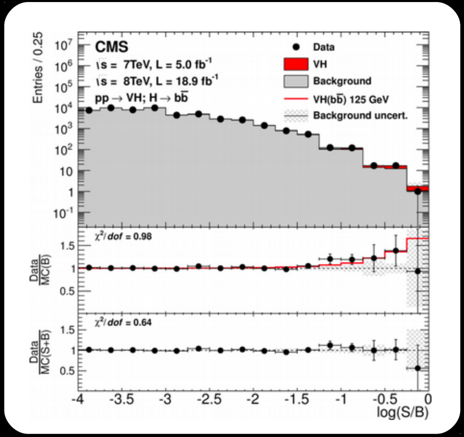
VBF

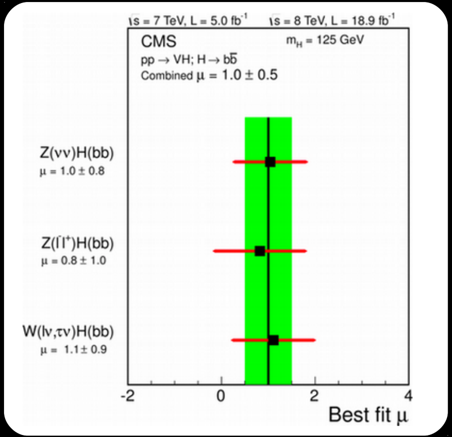


RESULTS: VH









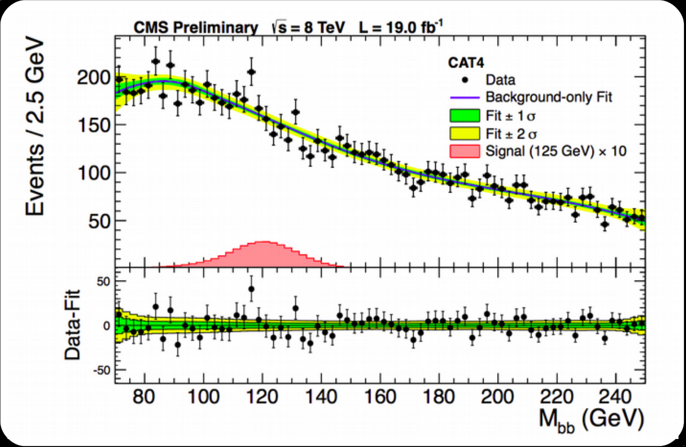
- An upper limit of 1.89 *[0.95]* times the SM Higgs boson cross section at 125 GeV with 95% C.L. has been observed *[expected]* in the VH channel.
- There, an excess of events of 2.1 standard deviations has been reported.
- It corresponds to a signal strength of:
  - $\mu = \sigma_{meas}/\sigma_{SM} = 1.0 \pm 0.5$.
- Compatible results have been obtained in all sub-channels (Z(νν)H(bb), Z(lν)H(bb) and W(lν)H(bb)).

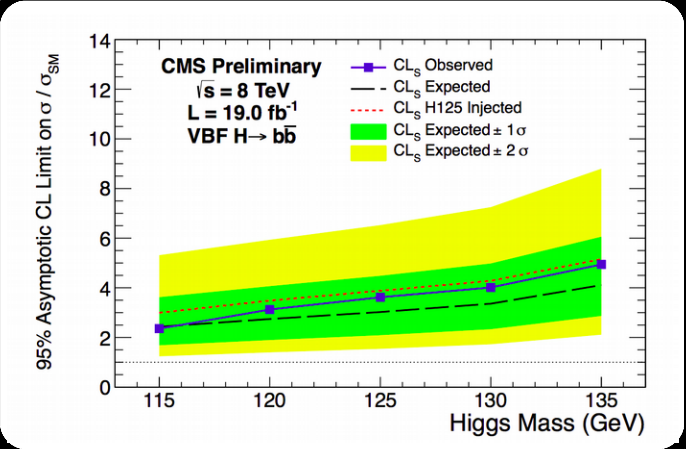
Phys.Rev. D89, 012003 (2014)

RESULTS: VBF

- An upper limit of 3.6 *[3.0]* times the SM Higgs boson cross section at 125 GeV with 95% C.L. has been observed *[expected]* in the **VBF** channel.
- It corresponds to a signal strength of:
  - $\mu = \sigma_{meas}/\sigma_{SM} = 0.7 \pm 1.4.$

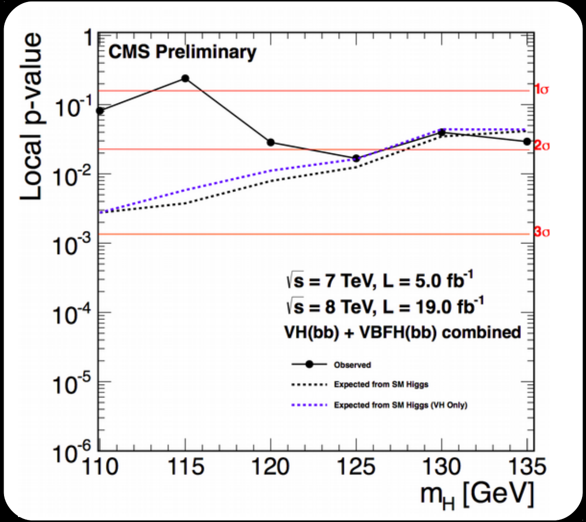
CMS-PAS-HIG-13-011 (2013)





Conclusions

- A search for the standard model Higgs boson decaying to bottom quarks has been presented.
- Two production channels have been studied: the associated production with vector boson decaying to lepton (VH) and the Vector Boson Fusion (VBF).
- The two different analysis strategies have been presented and their background estimation have been described.
- An excess of events of 2.1 standard deviations has been reported in the VH channel. It corresponds to a signal strength of 1.0 ± 0.5.
- In VBF, an upper limit of 3.6 [3.0] times the SM Higgs boson cross section at 125 GeV with 95% C.L. has been observed [expected]. This excess corresponds to a signal strength of 0.7 ± 1.4.
- **Combining the VH and VBF channels an excess of 2.2 standard deviations is reported. It corresponds to a signal strength of μ = 0.97 ± 0.48.**
- The combination with the $H \to \tau\tau$ channel has excess of 3.8 σ (4.4 σ expected) corresponding to μ = 0.83± 0.24. → Nature Physics 10, 557–560 (2014)

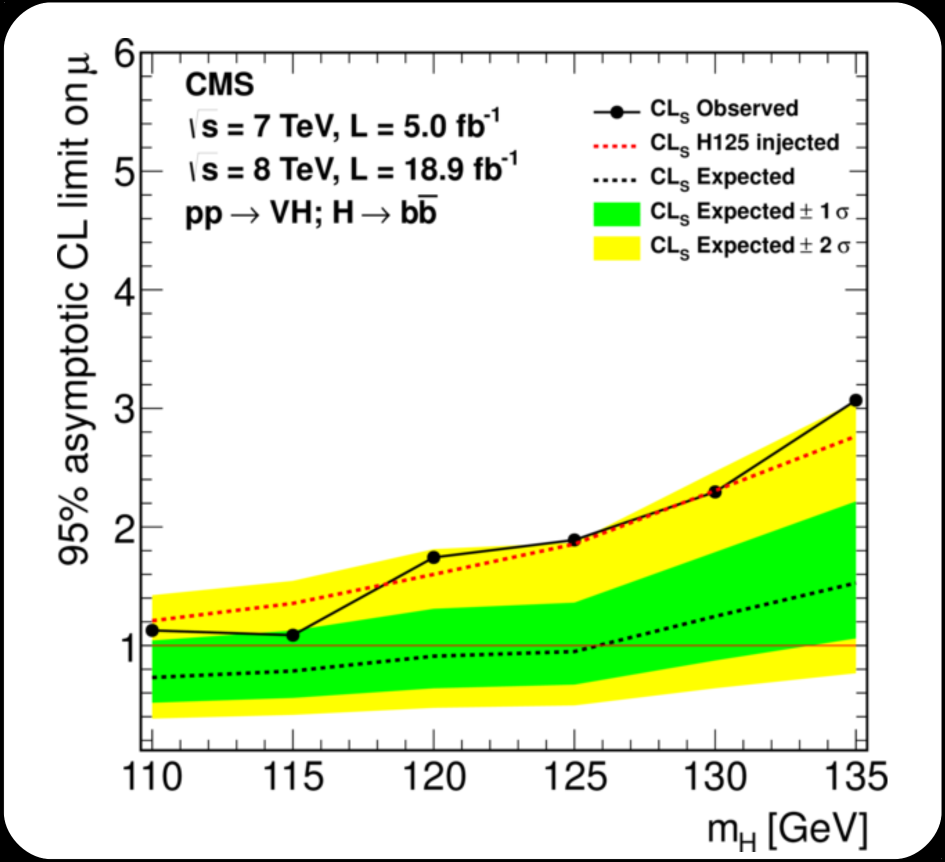


Thank you for your attention!

Questions?

Additional materials

VH: exclusion plot



Systematics

VH

Source	Type	Event yield uncertainty range (%)	Individual contribution to μ uncertainty (%)	Effect of removal on μ uncertainty (%)
Luminosity	norm.	2.2–2.6	<2	<0.1
Lepton efficiency and trigger (per lepton)	norm.	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)	norm.	4	3.9	0.3
Signal cross section (p_T boost, EW/QCD)	norm.	2/5	3.9	0.3
Monte Carlo statistics	shape	1–5	13.3	3.6
Backgrounds (data estimate)	norm.	10	15.9	5.2
Single-top quark (simulation estimate)	norm.	15	5.0	0.5
Dibosons (simulation estimate)	norm.	15	5.0	0.5
MC modeling (V + jets and $t\bar{t}$)	shape	10	7.4	1.1

VBF

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

VH: SIGNAL REGIONS

Variable	$W(\ell\nu)H$	$W(\tau\nu)H$	$Z(\ell\ell)H$	$Z(\nu\nu)H$
$p_{\mathrm{T}}(V)$	[100–130] [130–180] [>180]	[>120]	[50–100] [>100]	[100–130] [130–170] [>170]
$m_{\ell\ell}$	···	···	[75–105]	···
$p_{\mathrm{T}}(\mathrm{j}_1)$	>30	>30	>20	>60
$p_{\mathrm{T}}(\mathrm{j}_2)$	>30	>30	>20	>30
$p_{\mathrm{T}}(\mathrm{jj})$	>100	>120	···	[> 100] [> 130] [> 130]
$m(\mathrm{jj})$	<250	<250	[40–250] [<250]	<250
$E_{\mathrm{T}}^{\mathrm{miss}}$	>45	>80	···	[100–130] [130–170] [>170]
$p_{\mathrm{T}}(\tau)$	···	>40	···	···
p_{T} (track)	···	>20	···	···
$\mathrm{CSV}_{\mathrm{max}}$	>0.40	>0.40	[>0.50] [>0.244]	>0.679
$\mathrm{CSV}_{\mathrm{min}}$	>0.40	>0.40	>0.244	>0.244
N_{aj}	···	···	···	[<2] [···] [···]
$N_{a\ell}$	=0	=0	···	=0
$\Delta\phi(V,H)$	···	···	···	>2.0
$\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, \mathrm{jet})$	···	···	···	[>0.7] [>0.7] [>0.5]
$\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, E_{\mathrm{T}}^{\mathrm{miss}}(\mathrm{tracks}))$	···	···	···	<0.5
$E_{\mathrm{T}}^{\mathrm{miss}}$ significance	···	···	···	[>3] [···] [···]
$\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, \ell)$	$<\pi/2$	···	···	···

VH: variables used in the MVA

Variable
$p_{\mathrm{T}}(\mathrm{j}_1)$, $p_{\mathrm{T}}(\mathrm{j}_2)$: transverse momentum of each Higgs boson daughter
$m(\mathrm{jj})$: dijet invariant mass
$p_{\mathrm{T}}(\mathrm{jj})$: dijet transverse momentum
$p_{\mathrm{T}}(V)$: vector boson transverse momentum (or $E_{\mathrm{T}}^{\mathrm{miss}}$)
N_{aj}: number of additional jets (see caption)
$\mathrm{CSV}_{\mathrm{max}}$: value of CSV for the Higgs boson daughter with largest CSV value
$\mathrm{CSV}_{\mathrm{min}}$: value of CSV for the Higgs boson daughter with second largest CSV value
$\Delta\phi(V, H)$: azimuthal angle between V (or $E_{\mathrm{T}}^{\mathrm{miss}}$) and dijet
$\lvert\Delta\eta(\mathrm{jj})\rvert$: difference in η between Higgs boson daughters
$\Delta R(\mathrm{jj})$: distance in $\eta - \phi$ between Higgs boson daughters
$\Delta\theta_{\mathrm{pull}}$: color pull angle [43]
$\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, \mathrm{jet})$: azimuthal angle between $E_{\mathrm{T}}^{\mathrm{miss}}$ and the closest jet [only for $Z(\nu\nu)H$]
$\max \mathrm{CSV}_{\mathrm{aj}}$: maximum CSV of the additional jets in an event [only for $Z(\nu\nu)H$ and $W(\ell\nu)H$]
$\min \mathrm{CSV}_{\mathrm{aj}}$: minimum distance between an additional jet and the Higgs boson candidate [only for $Z(\nu\nu)H$ and $W(\ell\nu)H$]
Invariant mass of the VH system [only for $Z(\ell\ell)H$]
Cosine of the angle between the direction of the V boson in the rest frame of the VH system and the direction of the VH system in the laboratory frame [only for $Z(\ell\ell)H$]
Cosine of the angle between the direction of one of the leptons in the rest frame of the Z boson and the direction of the Z boson in the laboratory frame [only for $Z(\ell\ell)H$]
Cosine of the angle between the direction of one of the jets in the rest frame of the reconstructed Higgs boson and the direction of the reconstructed Higgs boson in the laboratory frame [only for $Z(\ell\ell)H$]

VH: control regions

Variable	Z + jets	$t\bar{t}$
$m_{\ell\ell}$	[75–105]	$\notin$ [75–105]
$p_T(j_1)$	>20	>20
$p_T(j_2)$	>20	>20
$p_T(V)$	>50	[50–100]
$m(jj)$	<250, $\notin$ [80–150]	<250, $\notin$ [80–150]
CSV_{max}	>0.244	>0.244
CSV_{min}	>0.244	>0.244

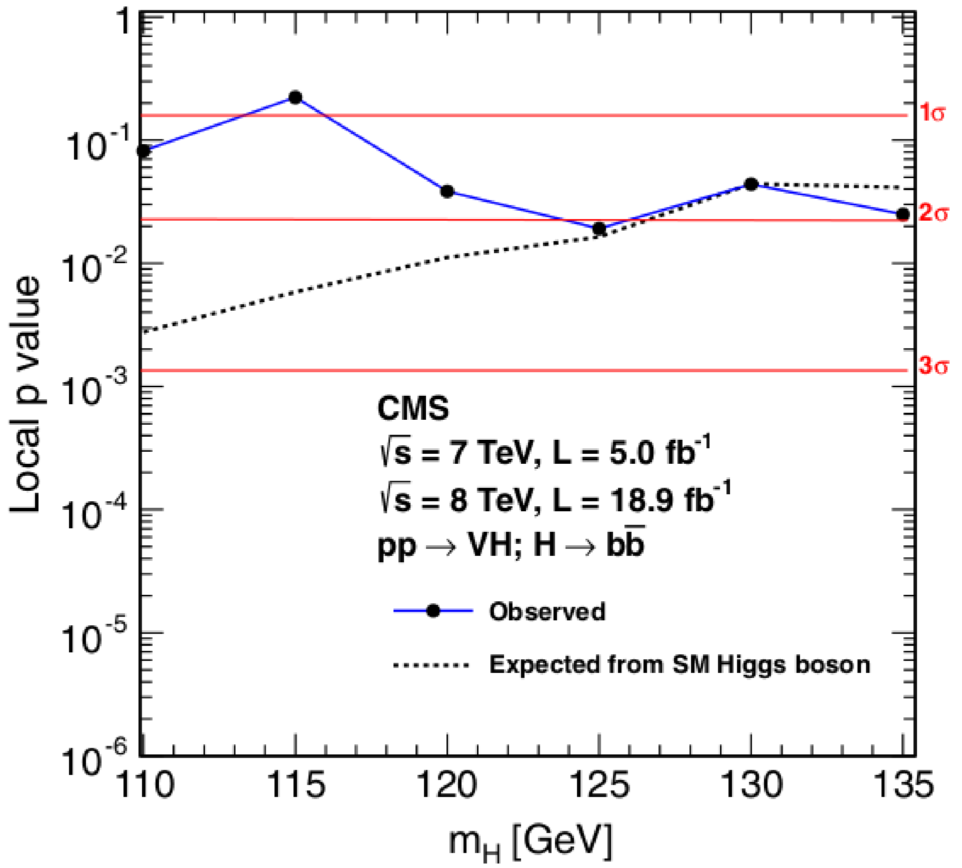
Variable	W + LF	$t\bar{t}$	W + HF
$p_T(j_1)$	>30	>30	>30
$p_T(j_2)$	>30	>30	>30
$p_T(jj)$	>100	>100	>100
$m(jj)$	<250	<250	<250, $\notin$ [90–150]
CSV_{max}	$\in$ [0.244–0.898]	>0.898	>0.898
N_{aj}	<2	>1	=0
$N_{a\ell}$	=0	=0	=0
E_T^{miss}	>45	>45	>45
E_T^{miss} significance	>2.0(μ), >3.0(e)	···	···

Variable E_T^{miss}	Z + LF [100–130] [130–170] [>170]	Z + HF [100–130] [130–170] [>170]	$t\bar{t}$ [100–130] [130–170] [>170]	W + LF [100–130] [130–170] [>170]	W + HF [100–130] [130–170] [>170]
$p_T(j_1)$	>60	>60	>60	>60	>60
$p_T(j_2)$	>30	>30	>30	>30	>30
$p_T(jj)$	[>100] [>130] [>130]	[>100] [>130] [>130]	[>100] [>130] [>130]	[>100] [>130] [>130]	[>100] [>130] [>130]
$m(jj)$	<250	<250, $\notin$ [100–140]	<250, $\notin$ [100–140]	<250	<250, $\notin$ [100–140]
CSV_{max}	[0.244–0.898]	>0.679	>0.898	[0.244–0.898]	>0.679
CSV_{min}	···	>0.244	···	···	>0.244
N_{aj}	[<2] [···] [···]	[<2] [···] [···]	≥1	=0	=0
$N_{a\ell}$	=0	=0	=1	=1	=1
$\Delta\phi(V, H)$	···	>2.0	···	···	>2.0
$\Delta\phi(E_T^{miss}, jet)$	[>0.7] [>0.7] [>0.5]	[>0.7] [>0.7] [>0.5]	[>0.7] [>0.7] [>0.5]	[>0.7] [>0.7] [>0.5]	[>0.7] [>0.7] [>0.5]
$\Delta\phi(E_T^{miss}, E_T^{miss}(tracks))$	<0.5	<0.5	···	···	···
E_T^{miss} significance	[>3] [···] [···]	[>3] [···] [···]	[>3] [···] [···]	[>3] [···] [···]	[>3] [···] [···]

VH: scale factors

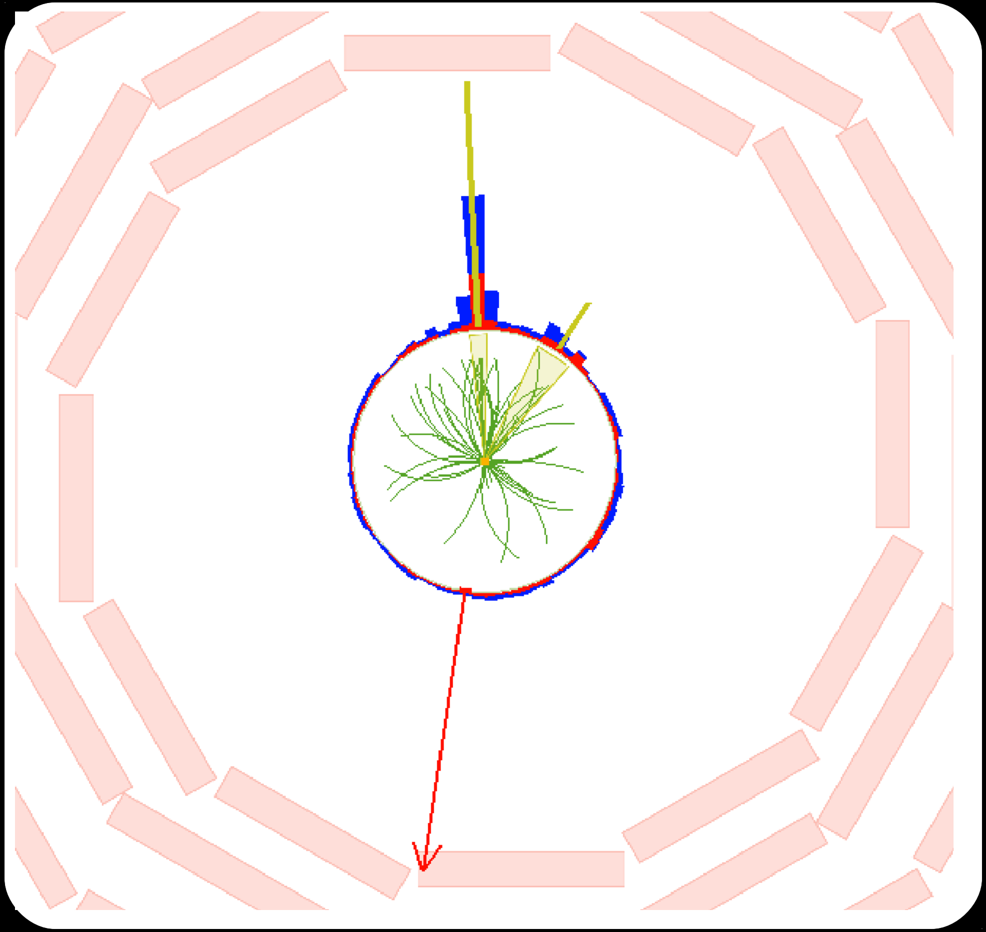
Process	$W(\ell\nu)H$	$Z(\ell\ell)H$	$Z(\nu\nu)H$
Low $p_T(V)$			
$W + udscg$	$1.03 \pm 0.01 \pm 0.05$	···	$0.83 \pm 0.02 \pm 0.04$
$W + b$	$2.22 \pm 0.25 \pm 0.20$	···	$2.30 \pm 0.21 \pm 0.11$
$W + b\bar{b}$	$1.58 \pm 0.26 \pm 0.24$	···	$0.85 \pm 0.24 \pm 0.14$
$Z + udscg$	···	$1.11 \pm 0.04 \pm 0.06$	$1.24 \pm 0.03 \pm 0.09$
$Z + b$	···	$1.59 \pm 0.07 \pm 0.08$	$2.06 \pm 0.06 \pm 0.09$
$Z + b\bar{b}$	···	$0.98 \pm 0.10 \pm 0.08$	$1.25 \pm 0.05 \pm 0.11$
$t\bar{t}$	$1.03 \pm 0.01 \pm 0.04$	$1.10 \pm 0.05 \pm 0.06$	$1.01 \pm 0.02 \pm 0.04$
Intermediate $p_T(V)$			
$W + udscg$	$1.02 \pm 0.01 \pm 0.07$	···	$0.93 \pm 0.02 \pm 0.04$
$W + b$	$2.90 \pm 0.26 \pm 0.20$	···	$2.08 \pm 0.20 \pm 0.12$
$W + b\bar{b}$	$1.30 \pm 0.23 \pm 0.14$	···	$0.75 \pm 0.26 \pm 0.11$
$Z + udscg$	···	···	$1.19 \pm 0.03 \pm 0.07$
$Z + b$	···	···	$2.30 \pm 0.07 \pm 0.08$
$Z + b\bar{b}$	···	···	$1.11 \pm 0.06 \pm 0.12$
$t\bar{t}$	$1.02 \pm 0.01 \pm 0.15$	···	$0.99 \pm 0.02 \pm 0.03$
High $p_T(V)$			
$W + udscg$	$1.04 \pm 0.01 \pm 0.07$	···	$0.93 \pm 0.02 \pm 0.03$
$W + b$	$2.46 \pm 0.33 \pm 0.22$	···	$2.12 \pm 0.22 \pm 0.10$
$W + b\bar{b}$	$0.77 \pm 0.25 \pm 0.08$	···	$0.71 \pm 0.25 \pm 0.15$
$Z + udscg$	···	$1.11 \pm 0.04 \pm 0.06$	$1.17 \pm 0.02 \pm 0.08$
$Z + b$	···	$1.59 \pm 0.07 \pm 0.08$	$2.13 \pm 0.05 \pm 0.07$
$Z + b\bar{b}$	···	$0.98 \pm 0.10 \pm 0.08$	$1.12 \pm 0.04 \pm 0.10$
$t\bar{t}$	$1.00 \pm 0.01 \pm 0.11$	$1.10 \pm 0.05 \pm 0.06$	$0.99 \pm 0.02 \pm 0.03$

VH: p-values

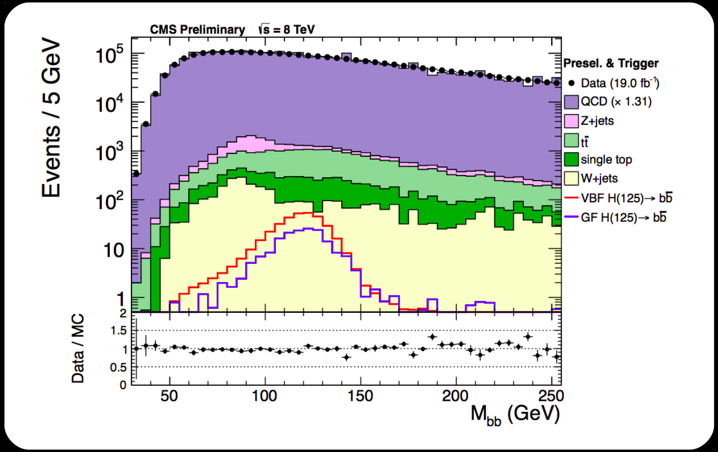


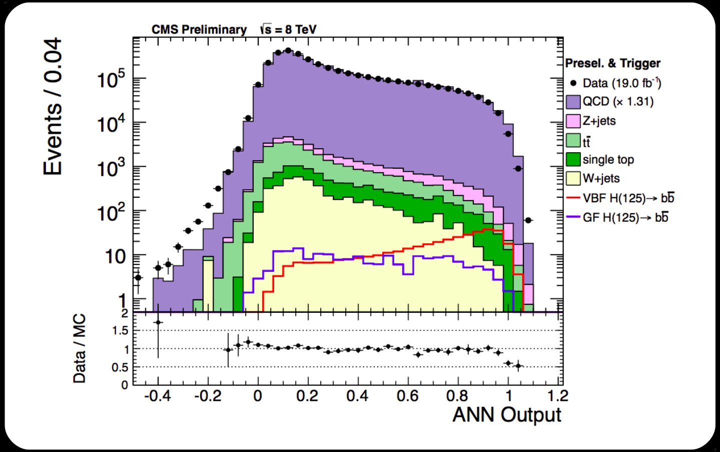
1
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
Local p value
1σ
2σ
3σ
CMS
$\sqrt{s}$ = 7 TeV, L = 5.0 fb^{-1}
$\sqrt{s}$ = 8 TeV, L = 18.9 fb^{-1}
pp → VH; H → $b\bar{b}$
Observed
Expected from SM Higgs boson
110
115
120
125
130
135
m_H [GeV]

Z(νν)H(bb): candidate event



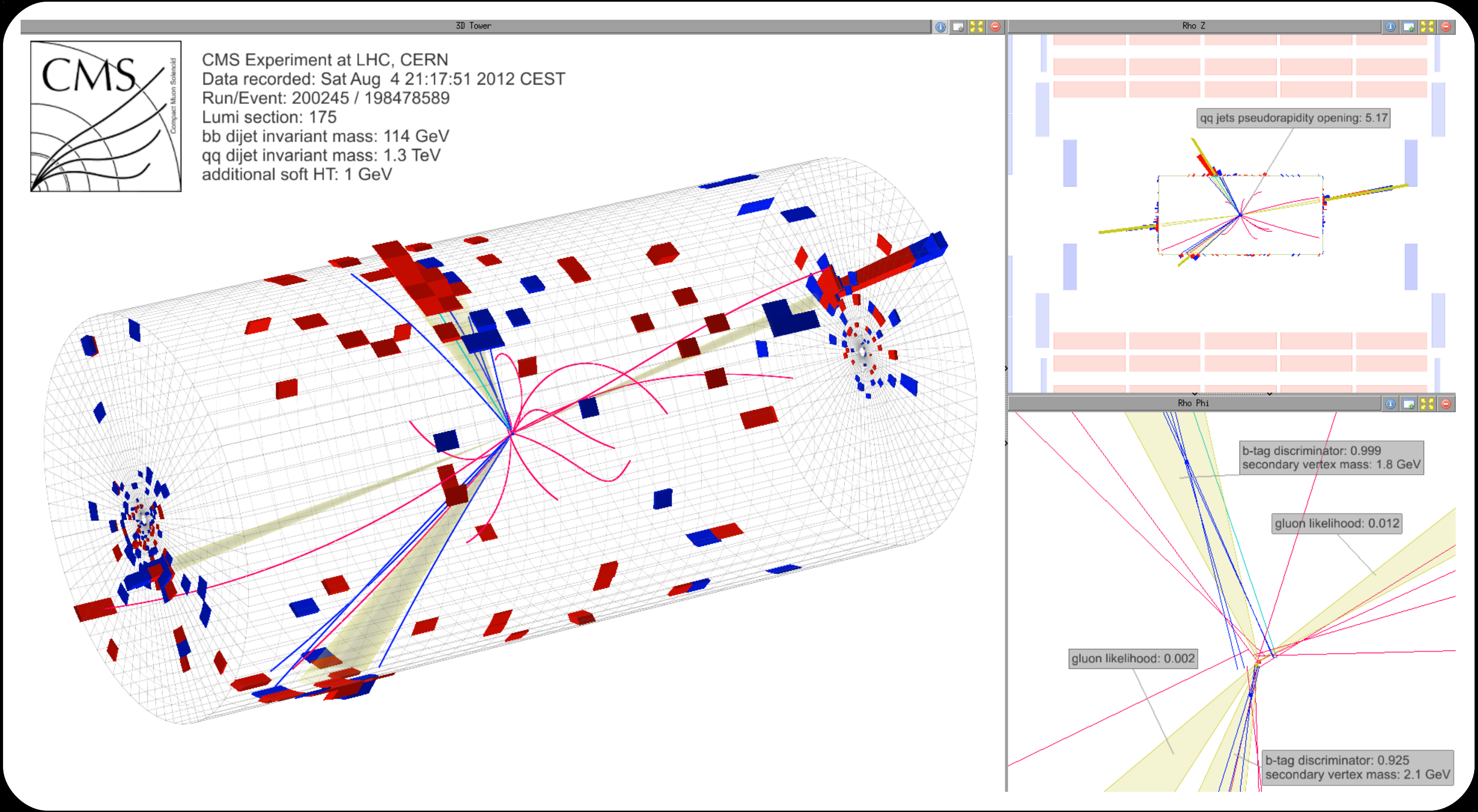
VBF: BACKGROUND DISTRIBUTIONS





Sample/ANN range	< 0.52	0.52 – 0.76	0.76 – 0.90	0.90 – 0.96	> 0.96
QCD	1.9e+6	3.2e+5	1.1e+5	2.7e+4	8.7e+3
$Z + jets$	5531	1222	531	124	54
$t\bar{t}$	12730	1032	190	33	15
t	1839	383	128	25	10
$\bar{t}$	895	226	73	15	7
$W + jets$	2033	226	50	4	<1
VBF $M_H(125)$	66	79	84	49	33
GF $M_H(125)$	94	37	18	6	2

VBF: A CANDIDATE EVENT



VBF: USED VARIABLES

1. $\boldsymbol{\Delta\eta_{qq}}$ the pseudorapidity separation between the b-tag sorted qq jets.
2. $\boldsymbol{\delta\Delta\eta_{qq}}$ the pseudorapidity separation difference between the b-tag sorted and the η sorted qq jets. This difference is expected to be mostly zero for preselected signal events where the non-b-tagged (VBF) jet pair is often the most forward-backward.
3. $\boldsymbol{m_{qq}}$ the invariant mass of the b-tag sorted qq jet pair.
4. $\boldsymbol{\eta_{qq}^{boost}}$ the average pseudorapidity of the b-tag sorted qq jet pair system.
5. $\mathbf{CSV_0}$ the CSV b-tagging output for the most b-tagged jet.
6. $\mathbf{CSV_1}$ the CSV b-tagging output for the second most b-tagged jet.
7. $\mathbf{QGL_2}$ the quark/gluon likelihood discriminator output for the third b-tagged jet.
8. $\mathbf{QGL_3}$ the quark/gluon likelihood discriminator output for the least b-tagged jet.
9. $\boldsymbol{\eta_2}$ the pseudorapidity of the third b-tagged jet.
10. $\boldsymbol{H_T^{soft}}$ the scalar p_T sum of the additional "soft" Track-Jets with $p_T > 1$ GeV.
11. $\mathbf{cos}\,\boldsymbol{\theta}$ the cosine of the polar angle of the vector $\vec{p}_{q_1} \times \vec{p}_{q_2}$ in the Higgs boson rest frame (the frame where the momenta of the two most b-tagged jets are back-to-back), where q_1 and q_2 are the least b-tagged jet pair. The angle θ is essentially the angle between the qq and bb planes.
12. $\mathbf{cos}\,\boldsymbol{\alpha}$ the polar angle of the vector $\vec{p}_{q_1} + \vec{p}_{q_2}$ in the Higgs boson rest frame.