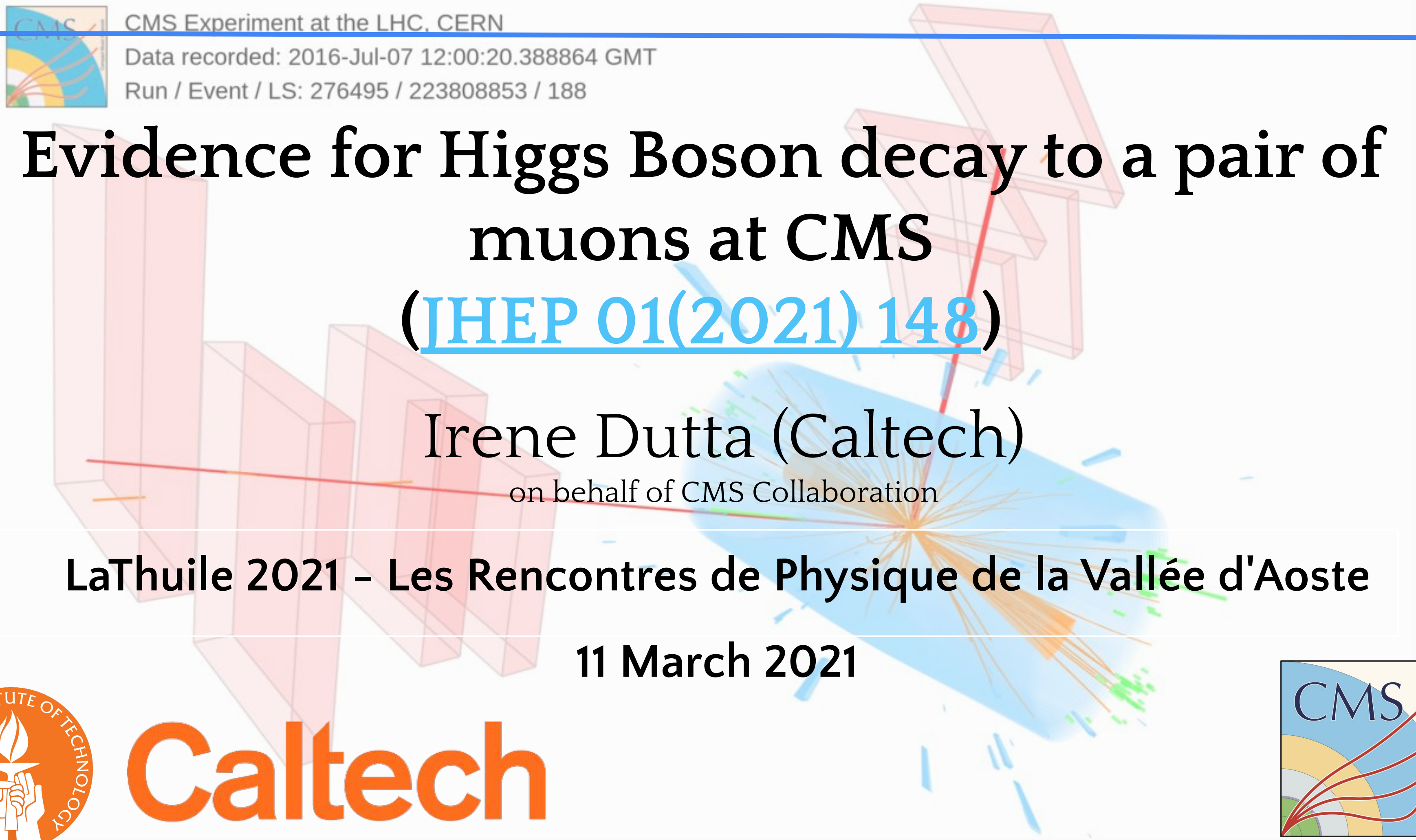


Evidence for Higgs Boson decay to a pair of muons at CMS

(JHEP 01(2021) 148)

Irene Dutta (Caltech)

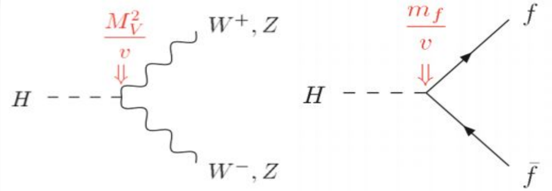
on behalf of CMS Collaboration

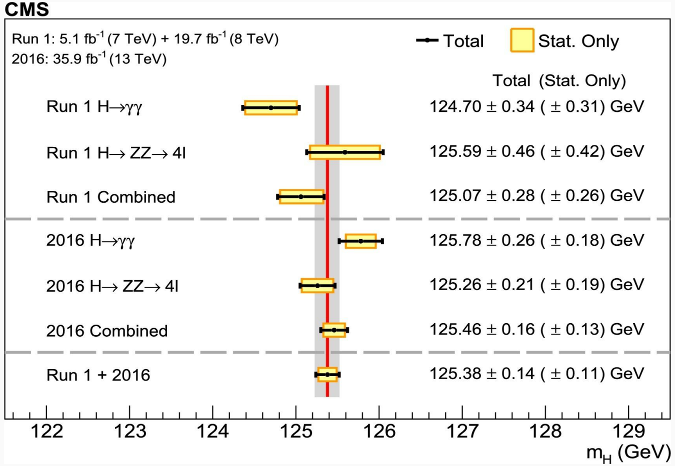
LaThuile 2021 – Les Rencontres de Physique de la Vallée d'Aoste

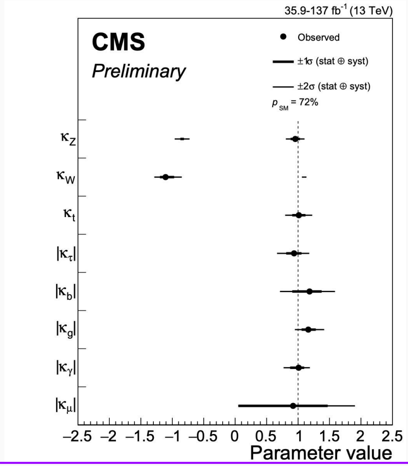
11 March 2021

The Higgs Boson properties

- Gauge Bosons in the SM acquire mass by electroweak symmetry breaking
- Fermions acquire masses through Yukawa couplings







Up Next- Yukawa coupling to 2nd generation fermions

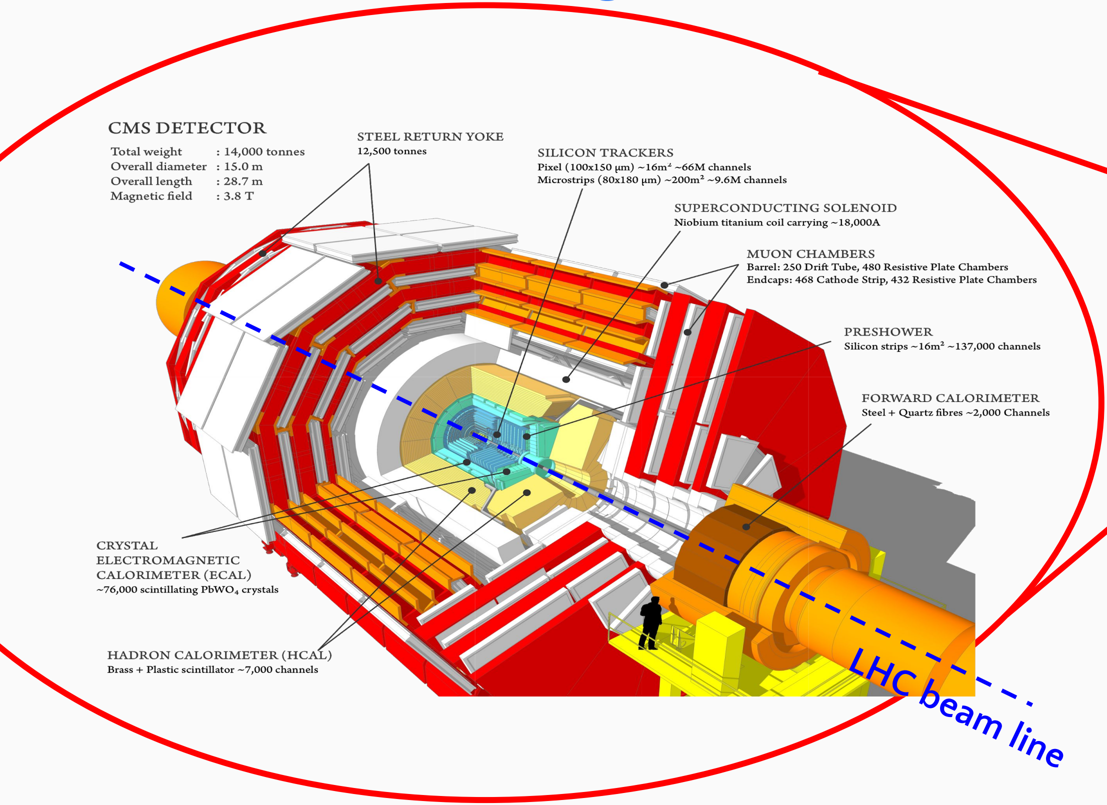
Most precise measurement of m_H = 125.38 ± 0.14 GeV
Phys. Lett. B. 805 135425

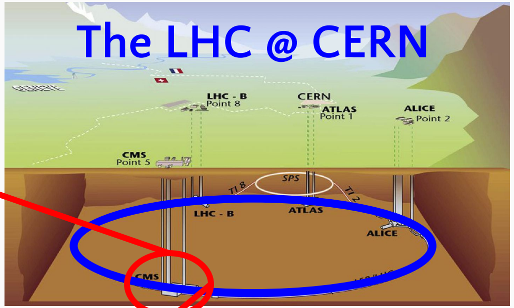
Higgs couplings to the third generation fermions (t/b/τ) and to the SM bosons (W/Z/γ) are well measured - CMS PAS HIG-19-005

Higgs couplings to 2nd generation fermions

- $H \to cc$ has the largest BR, but also more background contamination
- $H \to \mu\mu$ is currently the cleanest probe for second generation Yukawa coupling at the LHC
  - $BR(H \to \mu\mu) \sim 2.15\times10^{-4}$ for $M_H = 125.38$ GeV
  - Mass peak resolution : 1.5~2.5 GeV
- Large background (dominated by Drell-Yan $Z \to \mu\mu$, electro-weak Z, others include top, diboson and triboson production).
- S/B~1/500 in mass window 120-130 GeV.

CMS: a detector designed for muons (and more!)





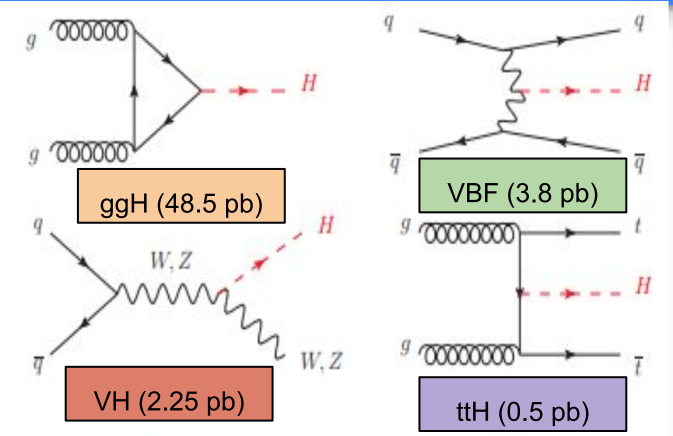
The muon p_T is precisely measured due to

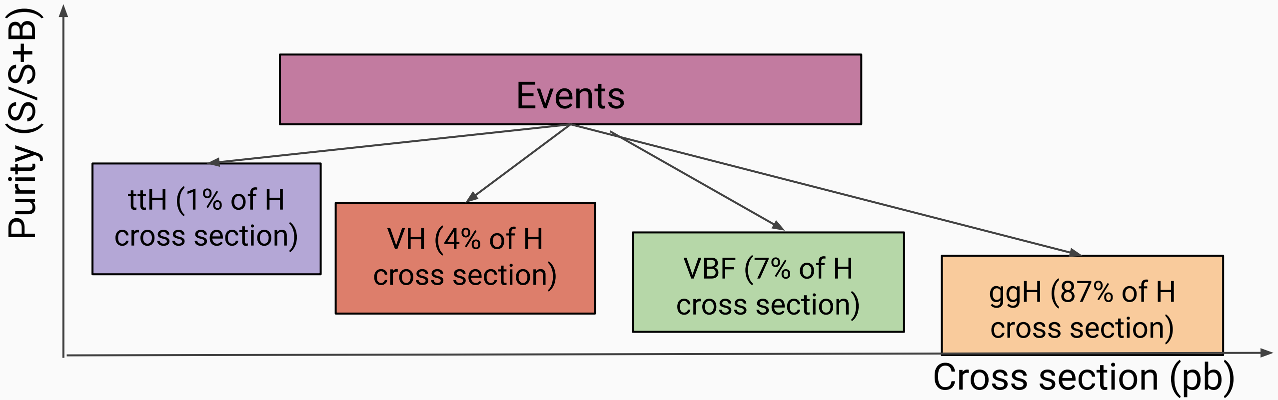
- high magnetic field (3.8 T)
- excellent tracking performances from inner tracker and muon chambers

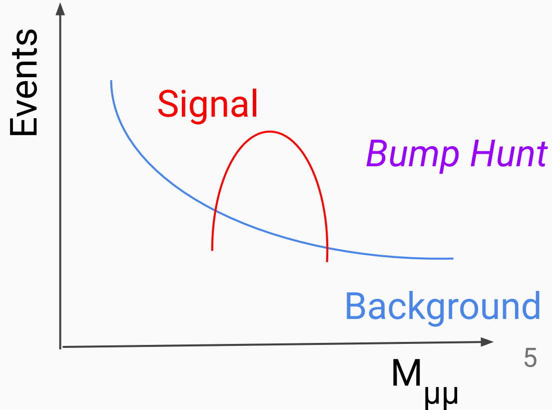
Muon $\Delta p/p$ varies from ~1% at 50 GeV to ~10% at 1 TeV

Search strategy

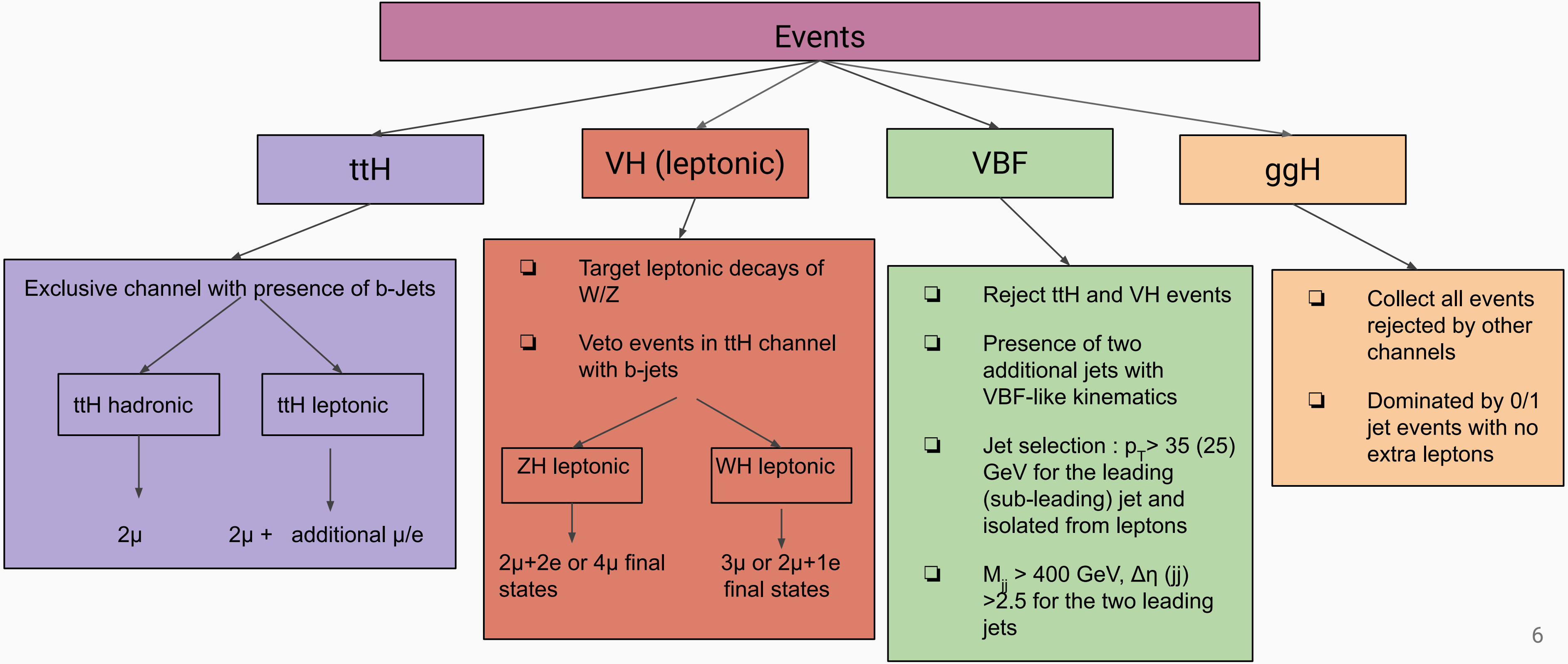
- Two oppositely charged muons that are well isolated and have the largest sum p_T.
- Higgs-candidate $M_{\mu\mu} \in [110, 150]$ GeV
- Exploit different kinematic features of different production mechanisms



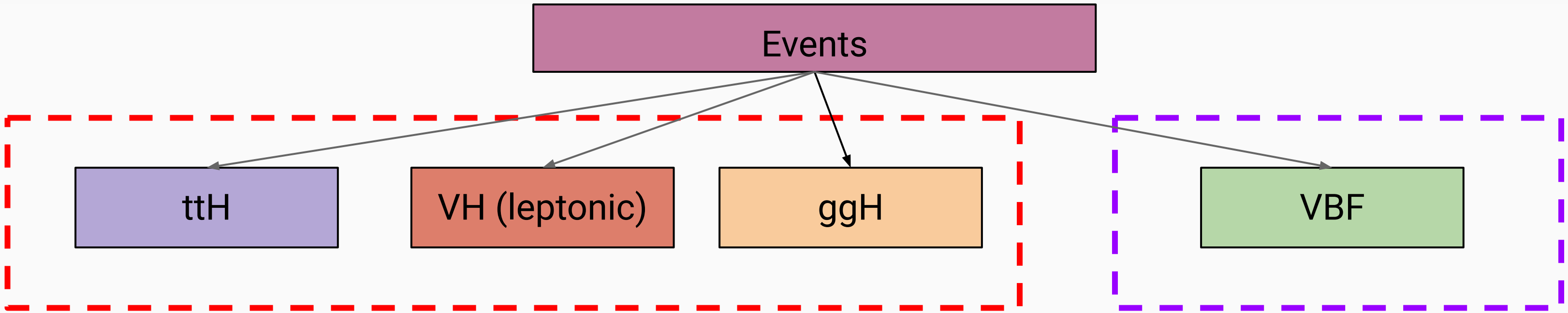




Event separation



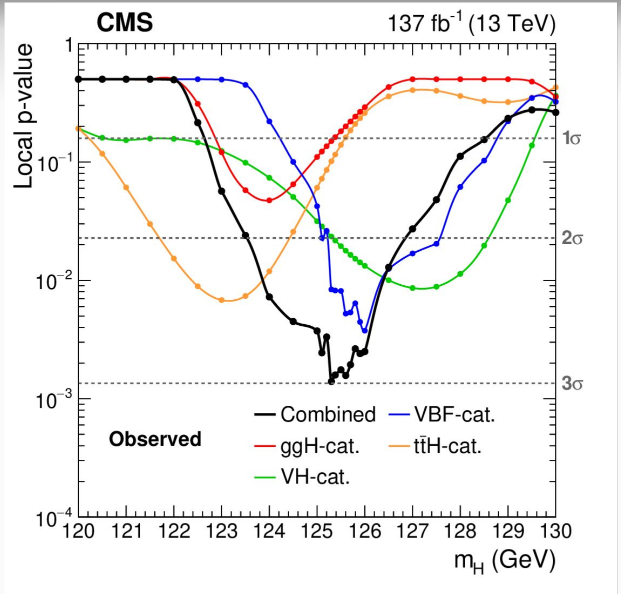
Event separation

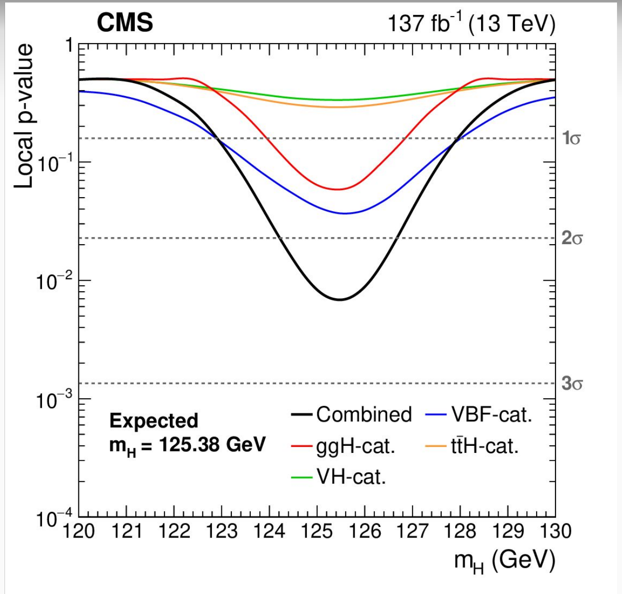


- Train independent BDTs for each region
- Background modelled with discrete likelihood profile of physics inspired and empirical functions.
- Signal peak modelled with a Gaussian function with power-law tails on both sides
- Perform a parametric fit to $M_{\mu\mu}$ spectrum

- Train a deep neural network
- Perform a simulation based template fit to DNN score output

Evidence of H→μμ



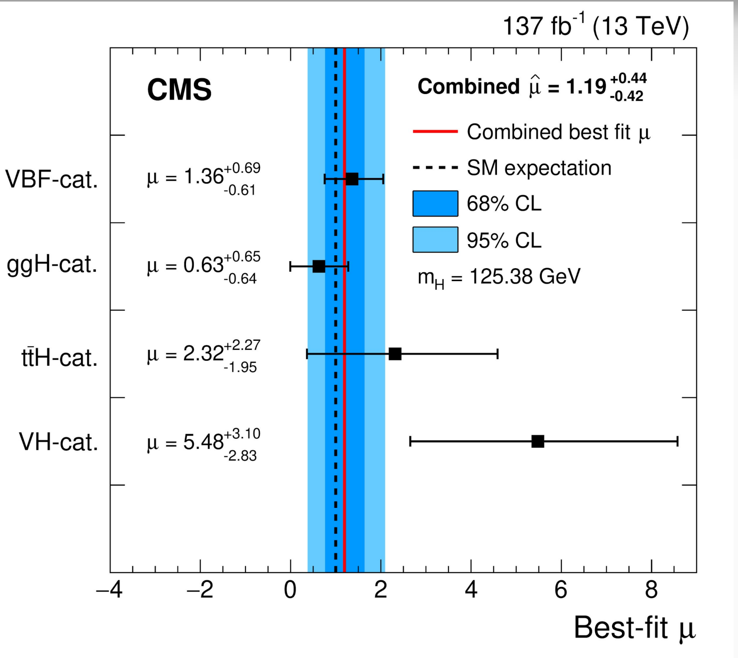


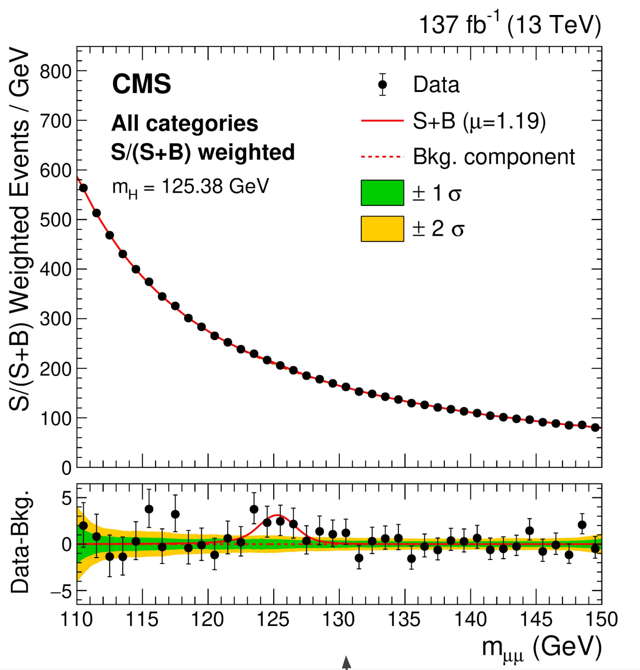
- For ggH, VH, and ttH channels - interpolate signal shape parameters analytically from m_H = 120, 125, and 130 GeV signal samples.
- For VBF channel, re-evaluate DNN for each mass point with the mass input fixed to that value and perform a new fit
- Observed jitter in VBF due to shuffling of data events in high score DNN bins as mass changes (binned fit).

Production category	Observed (expected) Signif.	Observed (expected) UL on μ
VBF	2.40 (1.77)	2.57 (1.22)
ggH	0.99 (1.56)	1.77 (1.28)
$t\bar{t}$H	1.20 (0.54)	6.48 (4.20)
VH	2.02 (0.42)	10.8 (5.13)
Combined $\sqrt{s} = 13$ TeV	2.95 (2.46)	1.94 (0.82)
Combined $\sqrt{s} = 7, 8, 13$ TeV	2.98 (2.48)	1.93 (0.81)

M_H = 125.38 GeV

H→μμ signal strength





Simultaneous fit performed across all channels with **μ = [σ*BR$_{obs}$]/[σ*BR$_{SM}$]**

$\mu = 1.19\ ^{+0.41}_{-0.39}(\text{stat})\ ^{+0.17}_{-0.16}(\text{syst})$ for M_H= 125.38 GeV

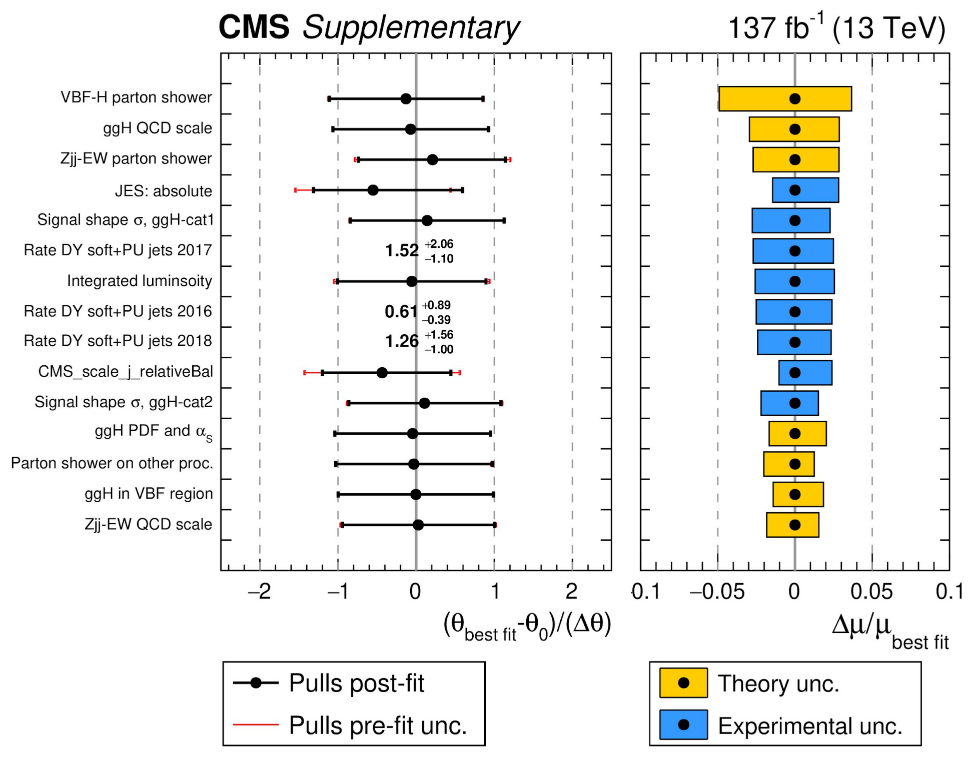
Not representative of the fit results; VBF fits to DNN score

Systematic uncertainties on Run 2 combination

Largest systematic uncertainty impacts from

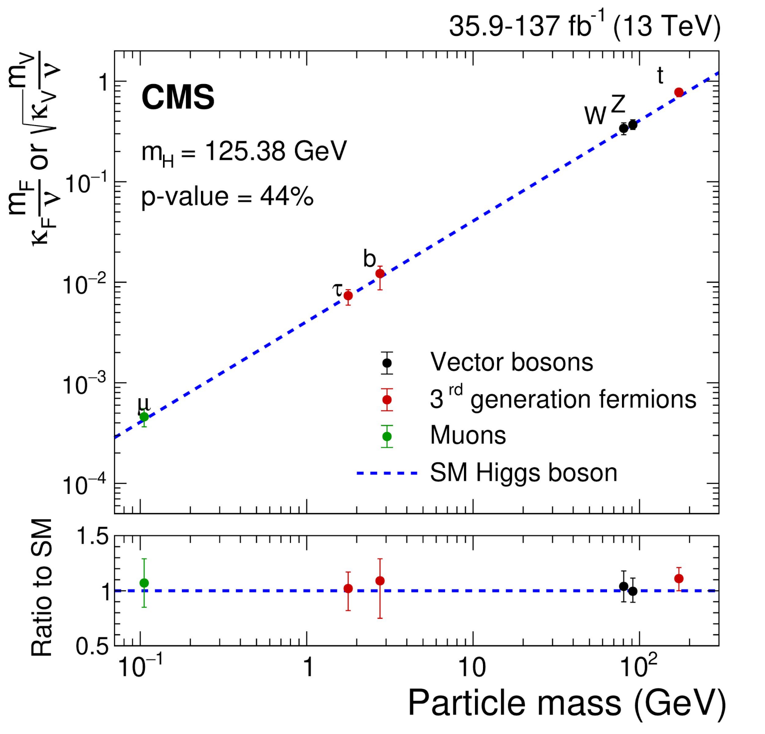
- limited statistics in data.
- the signal and background theory modeling
- Main experimental uncertainties include jet energy scale and resolution uncertainties.

Uncertainty source	$\Delta\mu$	
Total uncertainty	+0.44	−0.42
Statistical uncertainty	+0.41	−0.39
Total systematic uncertainty	+0.17	−0.16
Size of simulated samples	+0.07	−0.06
Total experimental uncertainty	+0.12	−0.10
Total theoretical uncertainty	+0.10	−0.11



Summary

- Measurement of H→μμ presented, combing four separate analyses targeting ggH, VBF, VH, and ttH production.
- Observed (expected) excess in data over bkg-only prediction of 3.0σ (2.5σ) at m_H = 125.38 GeV.
- $\kappa_\mu \in [0.62, 1.52]$ at 95% CL



Measured κ_μ directly by combining with HIG-19-005 (PAS-only)

Backup

Some common variable definitions

- **H_T** : Scalar sum p_T of all jets in an event
- **H_T^{miss}** : The negative vector sum of the p_T of all jets in the event with p_T > 30 GeV and $|\eta|$ < 4.7
- **E_T^{miss}** : The negative vector sum of the p_T of all reconstructed particles
- **M_T** : Transverse mass or $\sqrt{(E^2 - p_z^2)}$
- Collins - Soper angles :
    - **cos(θ*)** : The cosine of the angle between the collinear muons (μ) in the dimuon rest frame and the line that bisects the acute angle between the colliding partons (p)
    - **cos(φ*)** : The cosine of the angle between the dimuon plane and the plane of the colliding partons

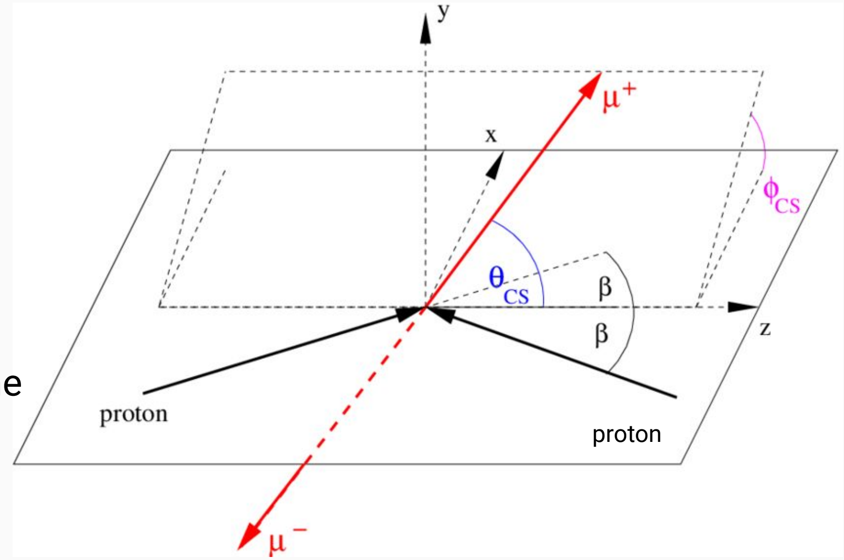
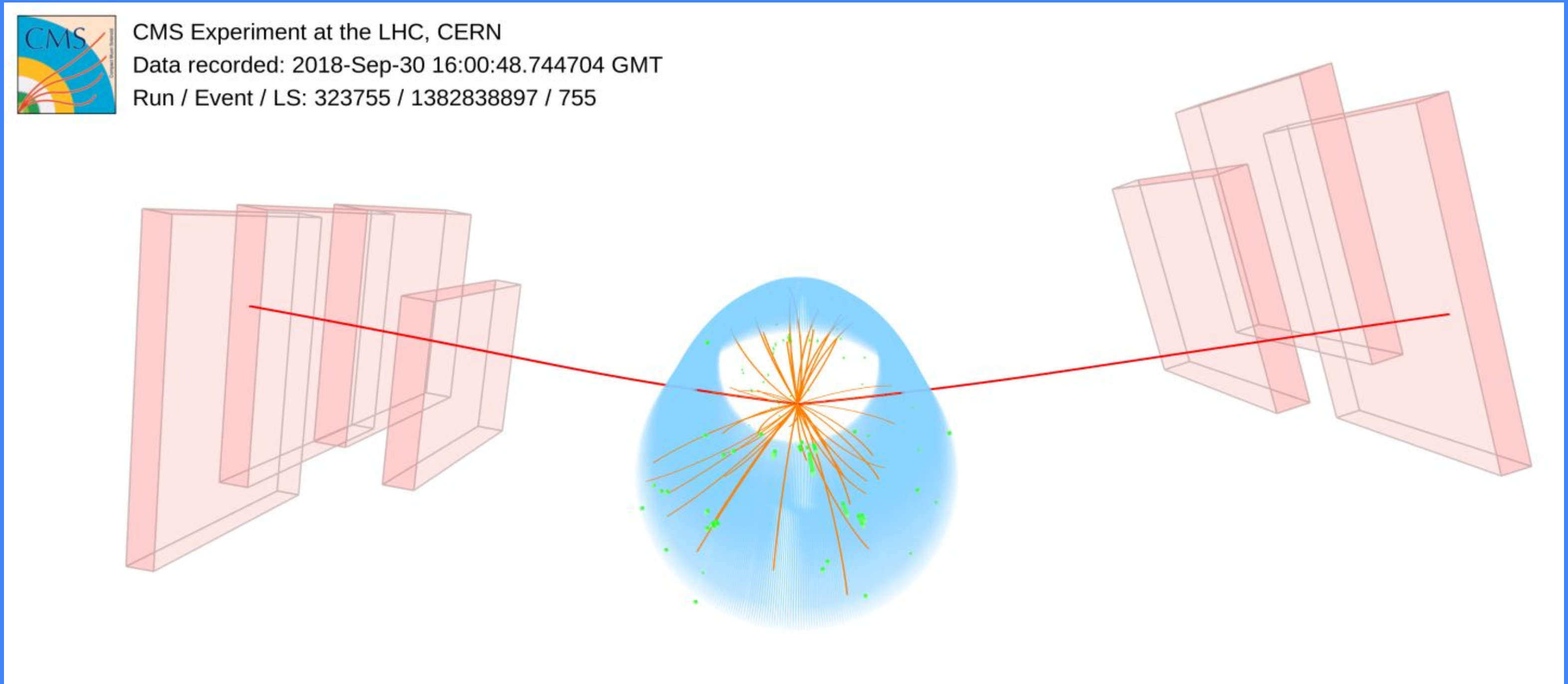


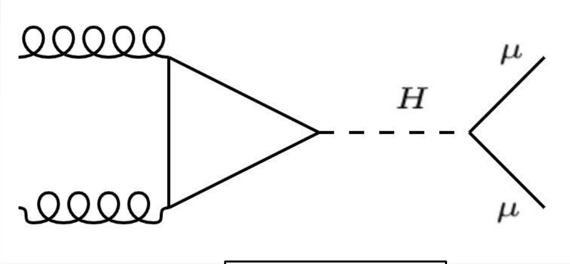
Image credit : J Kessler


CMS
CMS Experiment at the LHC, CERN
Data recorded: 2018-Sep-30 16:00:48.744704 GMT
Run / Event / LS: 323755 / 1382838897 / 755

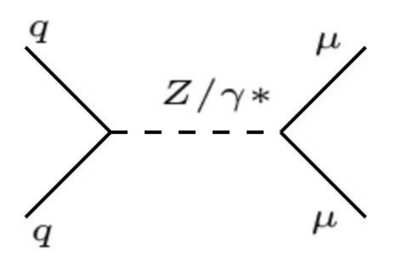
ggH channel

ggH

- The ggH channel collects events not selected by the other channels (ttH, VH, and VBF)
- This channel collects the largest amount of H(μμ) events but has the smallest S/B → challenging search
- Main background is Drell-Yan+jets
- Machine learning classifier (a.k.a. Boosted Decision Tree or BDT) used to separate Higgs signals from expected backgrounds
- Perform a parametric fit to the M(μμ) spectrum



ggH → μμ



Drell Yan: Z/γ*→ μμ

ggH BDT

Dimuon system
- $p_T(\mu\mu)$
- $y(\mu\mu)$
- Colins-Soper angles

Single muon
- $p_T(\mu)/m(\mu\mu)$
- $\eta(\mu)$

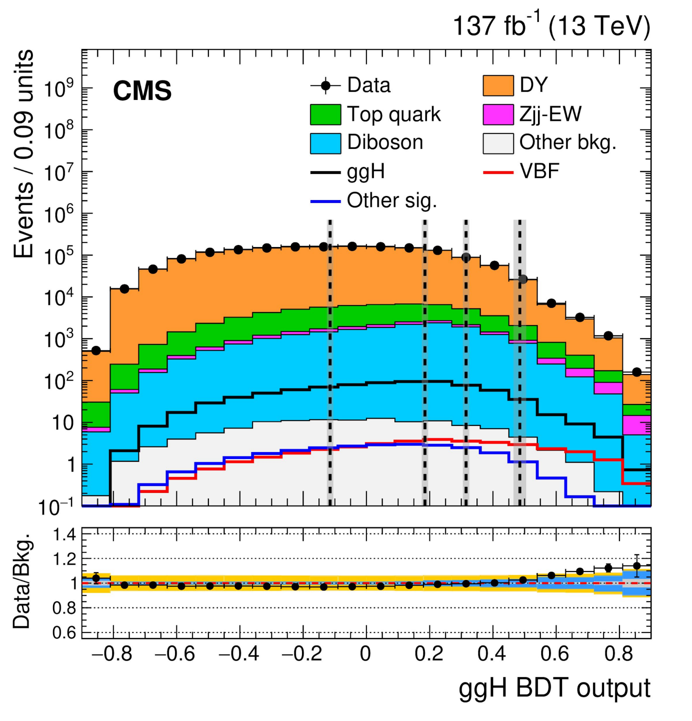
Event variables
- Njets

Weight signal events in BDT training by $1/(\sigma_{\mu\mu}/M_{\mu\mu})$

Dijet system
- $m(jj)$
- $\Delta\eta(jj)$
- $\Delta\varphi(jj)$
- Zeppenfeld: $Z = \eta_{\mu\mu} - 0.5*(\eta_{j1}+\eta_{j2})$
- $p_T(j_2)$
- min-$\Delta\eta(\mu\mu,j)$
- min-$\Delta\varphi(\mu\mu,j)$

For targeting residual VBF and V(qq)H events for $N_{jets} \geq 2$

Leading jet
- $p_T(j_1)$
- $\eta(j_1)$
- $\Delta\eta(\mu\mu, j_1)$
- $\Delta\varphi(\mu\mu, j_2)$

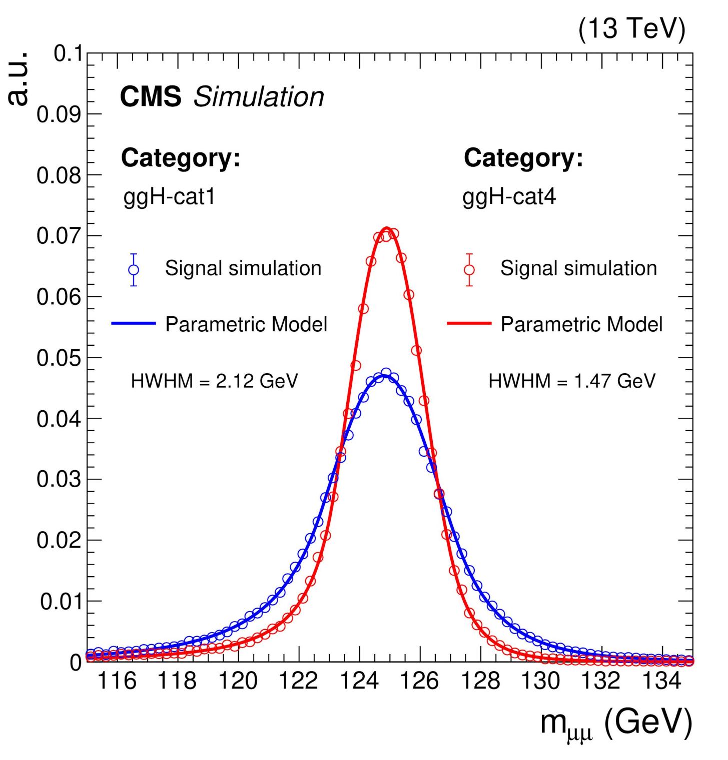


Bin the continuous BDT output into different boundaries by optimizing S/B

Signal modeling for ggH

- Need to extract the signal peak from the continuously falling background shape
- Signal model: Double sided crystal ball function
    - Gaussian with power-law tails on both sides

$$\mathrm{DCB}(m_{\mu\mu}) = \begin{cases} e^{-(m_{\mu\mu}-\hat{m})^2/2\sigma^2} & -\alpha_\mathrm{L} < \frac{m_{\mu\mu}-\hat{m}}{\sigma} < \alpha_\mathrm{R} \\ \left(\frac{n_\mathrm{L}}{|\alpha_\mathrm{L}|}\right)^{n_\mathrm{L}} e^{-\alpha_\mathrm{L}^2/2}\left(\frac{n_\mathrm{L}}{|\alpha_\mathrm{L}|} - |\alpha_\mathrm{L}| - \frac{m_{\mu\mu}-\hat{m}}{\sigma}\right)^{-n_\mathrm{L}} & \frac{m_{\mu\mu}-\hat{m}}{\sigma} \leq -\alpha_\mathrm{L} \\ \left(\frac{n_\mathrm{R}}{|\alpha_\mathrm{R}|}\right)^{n_\mathrm{R}} e^{-\alpha_\mathrm{R}^2/2}\left(\frac{n_\mathrm{R}}{|\alpha_\mathrm{R}|} - |\alpha_\mathrm{R}| - \frac{m_{\mu\mu}-\hat{m}}{\sigma}\right)^{-n_\mathrm{R}} & \frac{m_{\mu\mu}-\hat{m}}{\sigma} \geq \alpha_\mathrm{R} \end{cases}$$



Background modeling for ggH

- Background modeling through Core-PDF method
- No prior assumption on background shape or normalization

$$\boxed{B_{cat}(m_{\mu\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = N_B \times F_{core}(m_{\mu\mu}, \boldsymbol{\alpha}) \times T_{SMF}(m_{\mu\mu}, \boldsymbol{\beta})}$$

Number of bkg. events

Core-PDF

- discrete likelihood profile of physics inspired and empirical functions.
- Models the background DY spectrum
- Common parameters across all categories

Polynomial shape modifier

- “morphs” the core-component
- Accounts for shape variation across categories
- Different parameters in different categories

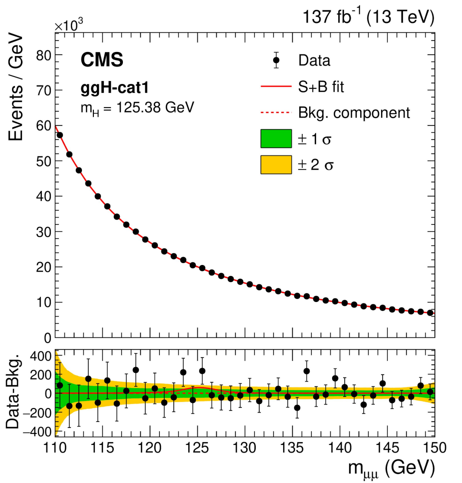
Background modeling for ggH

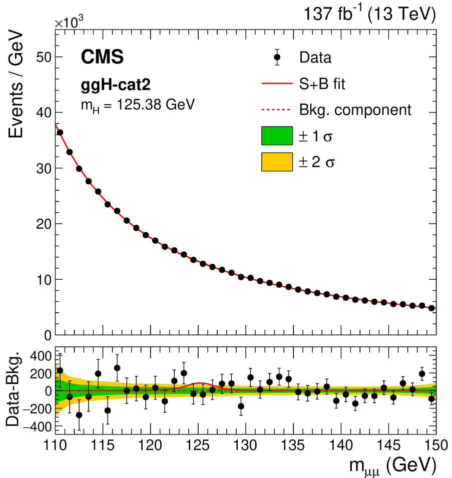
Bias due to choice of background function studied

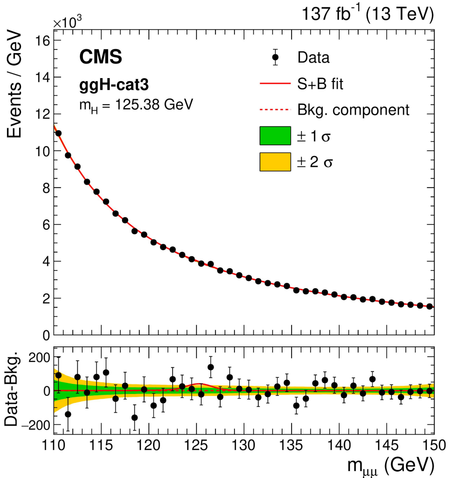
- Select multiple physics-inspired and agnostic background functions
- For each choice of function: perform B-only fits to data in each category with independent and uncorrelated parameters
- From the post-fit shapes → generate pseudo-data across all categories
- S+B fits are performed on the pseudo-data sets using the core-pdf method with a chosen signal strength (μ)
- Measured bias in each category as **median[b = $(\mu-\mu_{inj})/\sigma_{\mu}$]** → always < 20% (negligible impact on result).

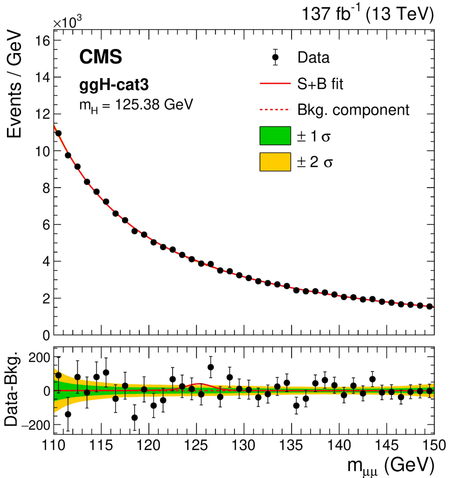
Fitting strategy for ggH

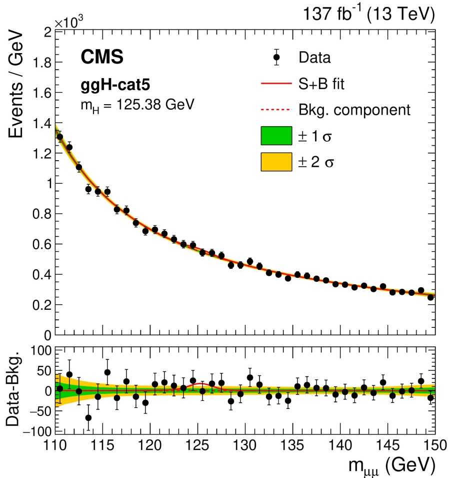
- Perform independent S+B fits to M(μμ) in each BDT category
- S+B fit describes the data well throughout the M(μμ) spectrum in all categories.

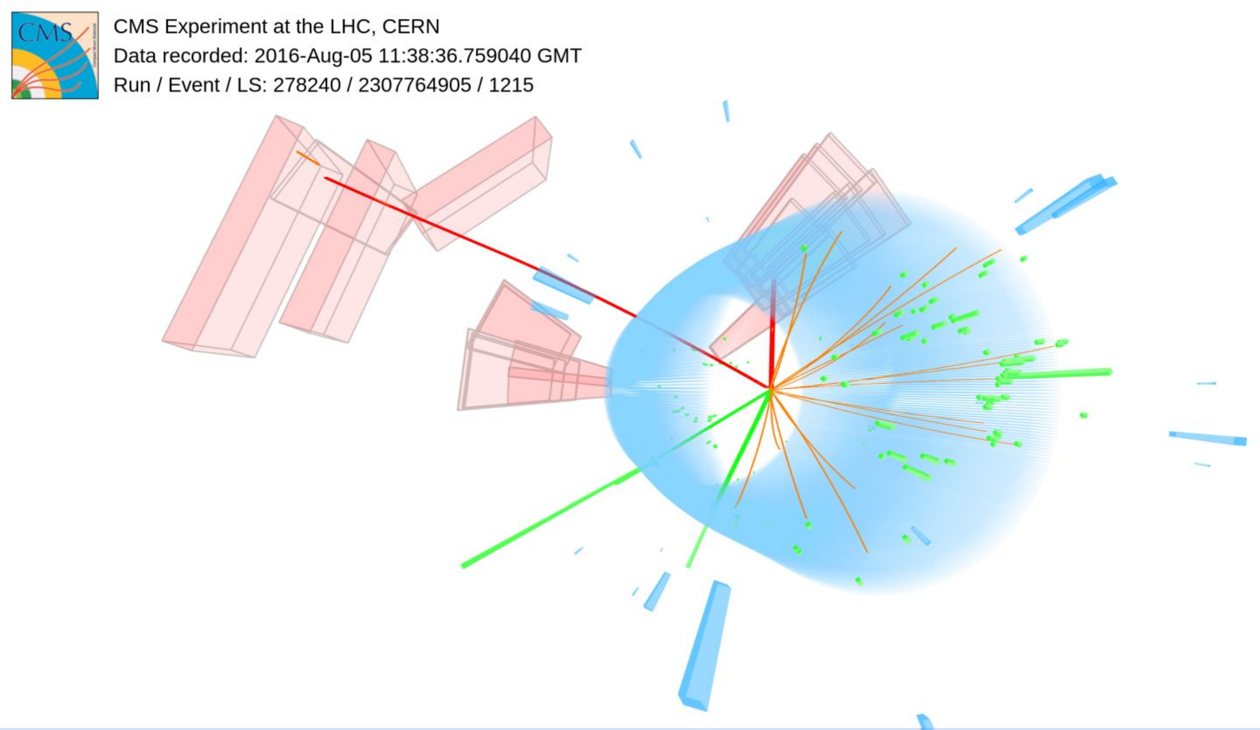











CMS Experiment at the LHC, CERN
Data recorded: 2016-Aug-05 11:38:36.759040 GMT
Run / Event / LS: 278240 / 2307764905 / 1215

VH channel

VH

- No two same flavour leptons with $m_{ll} < 12$ GeV in the event
- Main backgrounds are WZ, qq→ZZ and gg→ZZ
- Train two independent BDTs for WH and ZH

ZH leptonic

↓

2μ2e or 4μ final final states

↓

Z-candidate:

- $M_{\mu\mu} \in [81, 101]$ GeV
- $M_{ee} \in [71, 111]$ GeV

WH leptonic

↓

3μ or 2μ1e states

↓

3μ : No $M_{\mu\mu} \in [81, 101]$ GeV in the event

WH BDT

Dimuon system

- $p_T(\mu\mu)$

Single muon

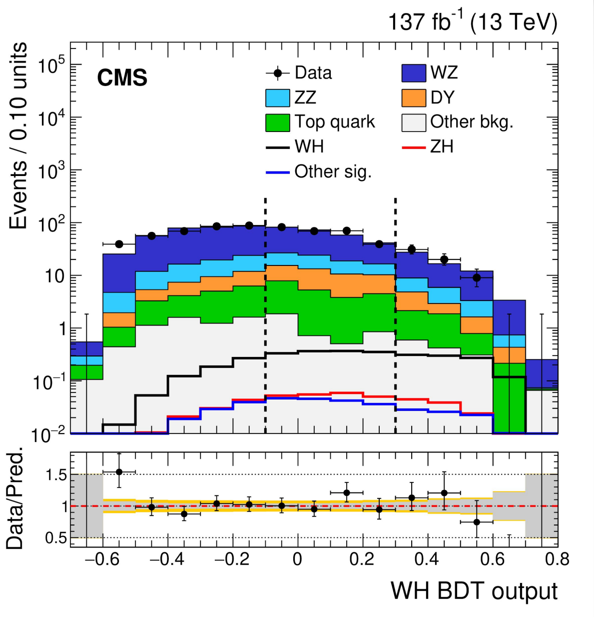
- $\eta(\mu_1)$, $\eta(\mu_2)$
- $\Delta R(\mu_1,\mu_2)$

Additional lepton (with charge c)

- Flavour , p_T, η
- $\Delta R(\mu\mu, l)$, $\Delta\eta(\mu\mu, l)$

Others

- $\cos(\theta^*)$ between (μ_1, l) and (μ_2, l)
- $\Delta\eta(\mu_1, l)$ and $\Delta\eta(\mu_2, l)$
- $\Delta R(\mu^{-c}, l^c)$
- M_T of the extra lepton and $\mathbf{H_T^{(miss)}}$
- M_T of the μ^C and the $\mathbf{H_T^{(miss)}}$
- $\Delta\varphi(\mathbf{H_T^{(miss)}}, l)$



Weight signal events in BDT training by $1/(\sigma_{\mu\mu}/M_{\mu\mu})$

Bin the continuous BDT output into different boundaries by optimizing S/B

ZH BDT

Dimuon system

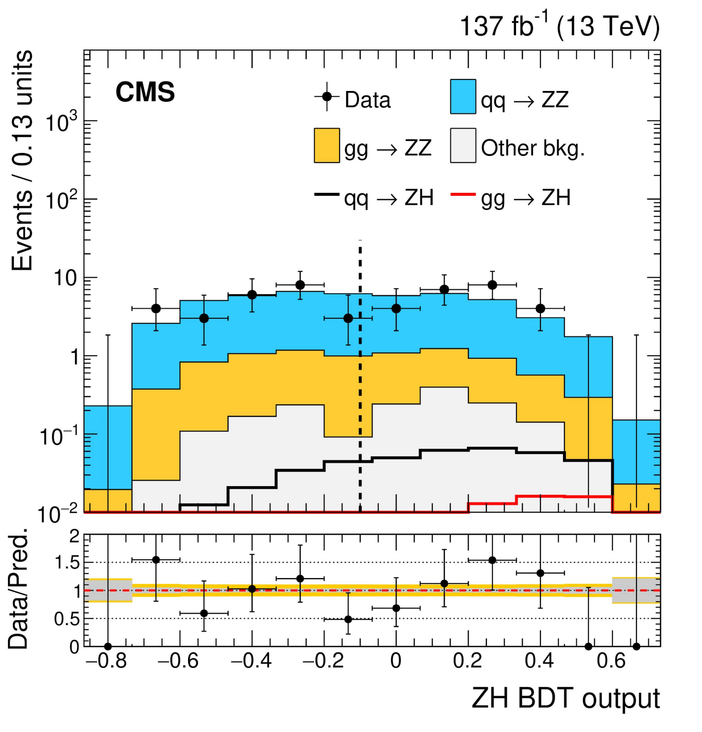
- $p_T(\mu\mu)$, $\eta(\mu\mu)$
- $\Delta\varphi(\mu_1, \mu\mu)$, $\Delta\varphi(\mu_1, \mu\mu)$

Other lepton pair

- M(ll), p_T, η, flavour
- $\Delta R(l_1,l_2)$

Others

- cos(θ*) between μμ and ll
- Δη(μμ, ll)

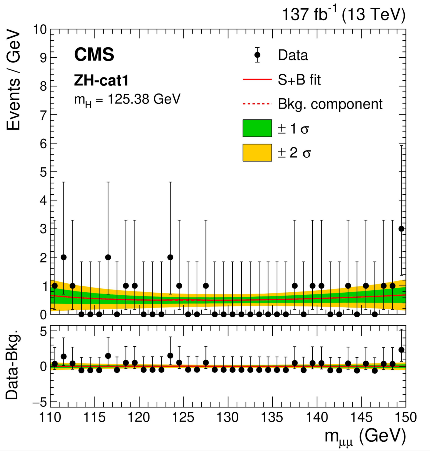


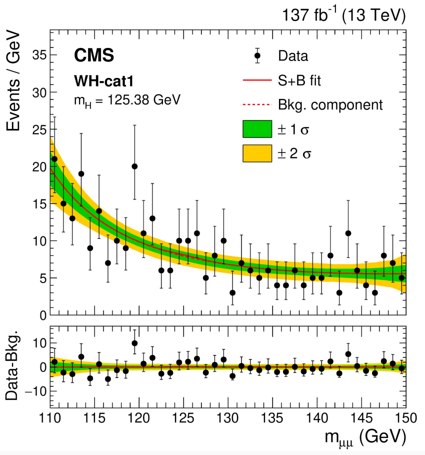
Weight signal events in BDT training by $1/(\sigma_{\mu\mu}/M_{\mu\mu})$

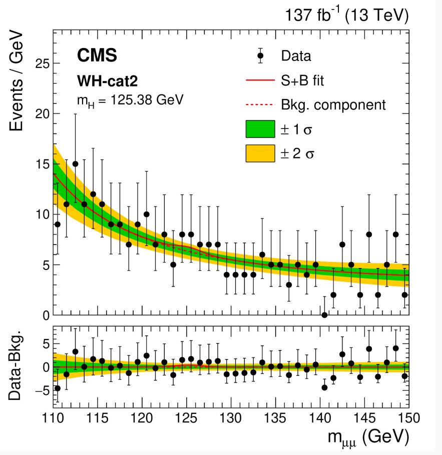
Bin the continuous BDT output into different boundaries by optimizing S/B

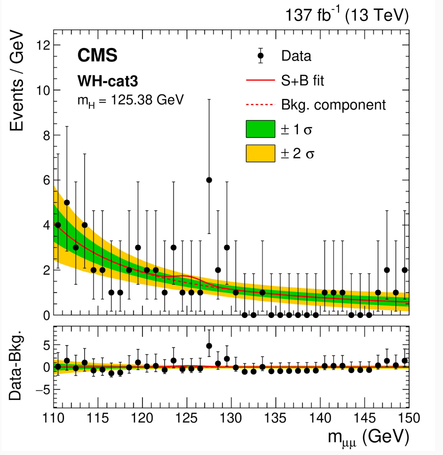
VH fitting strategy

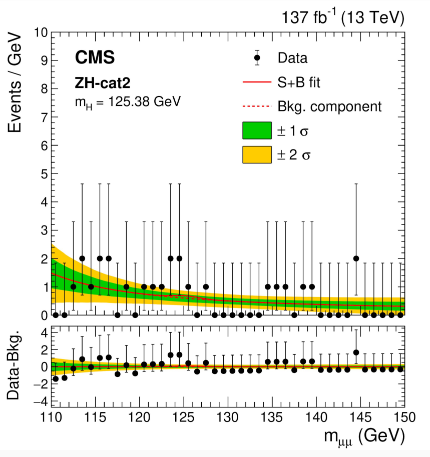
- Perform independent fits to M(μμ) in each BDT category
- Signal model: Double sided crystal ball function
- Background fit with analytical functions

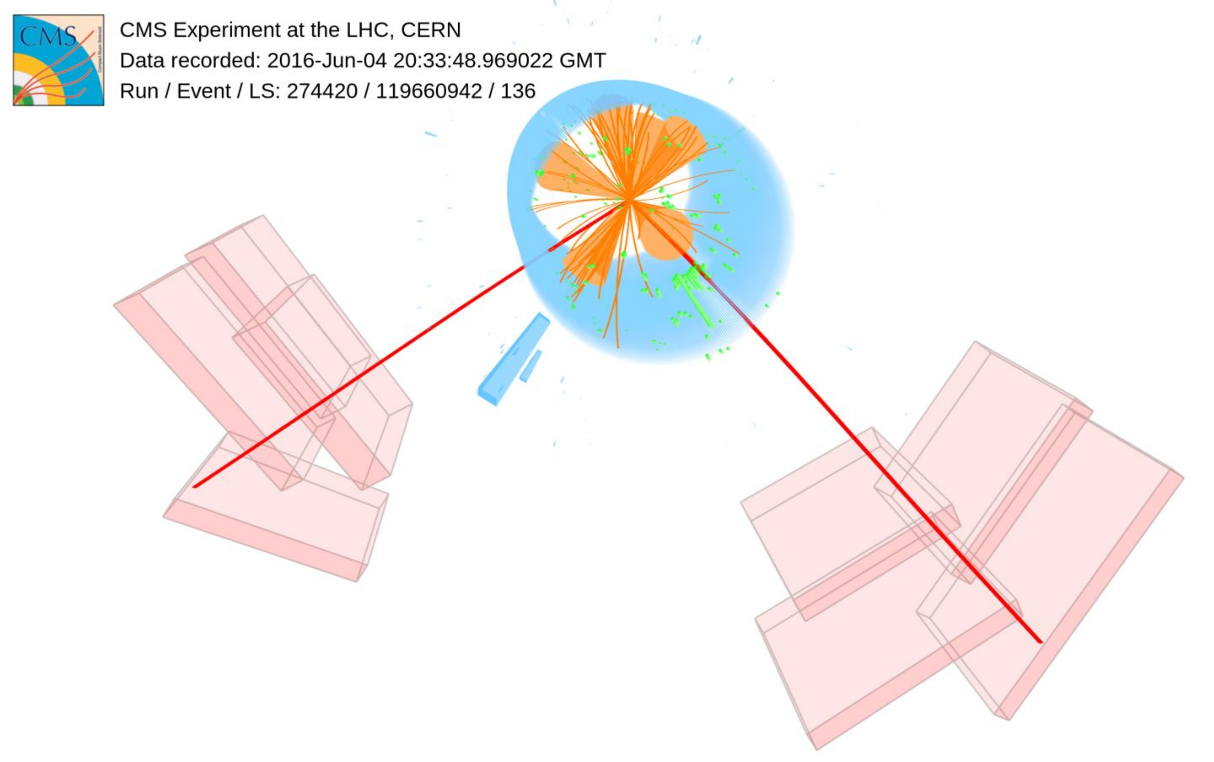








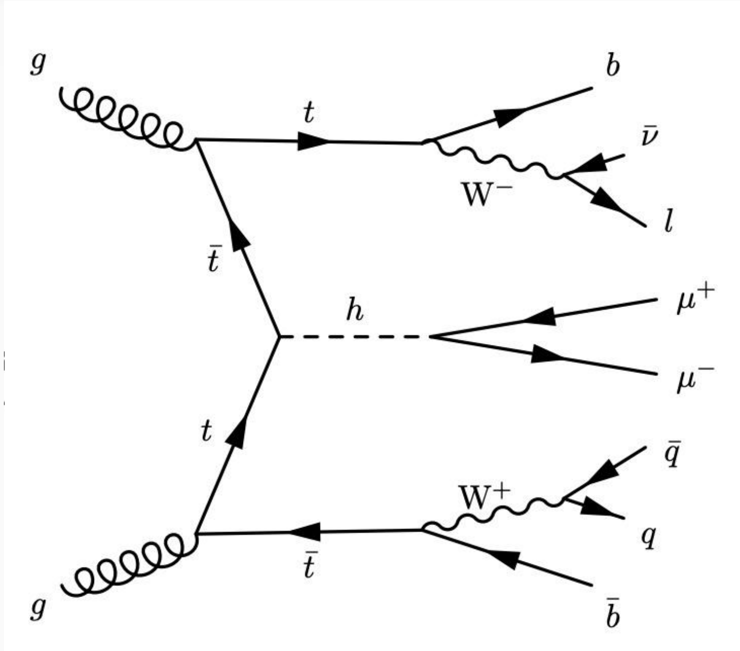



CMS Experiment at the LHC, CERN
Data recorded: 2016-Jun-04 20:33:48.969022 GMT
Run / Event / LS: 274420 / 119660942 / 136

ttH channel

ttH

- Characterized by ≥ 1 b tagged jet
- Split into two orthogonal channels: hadronic (2μ) and leptonic (2μ + additional μ/e)
- Large backgrounds from
  - di-leptonic $t\bar{t}$ production and
  - Drell Yan
  - ttZ production
- Two independent BDTs are trained for ttH-hadronic and ttH-leptonic



ttH BDT

Dimuon system
- $p_T(\mu\mu)$
- $y(\mu\mu)$
- Collin-Sopper angles

Single muon
- $p_T(\mu)/m(\mu\mu)$
- $\eta(\mu)$

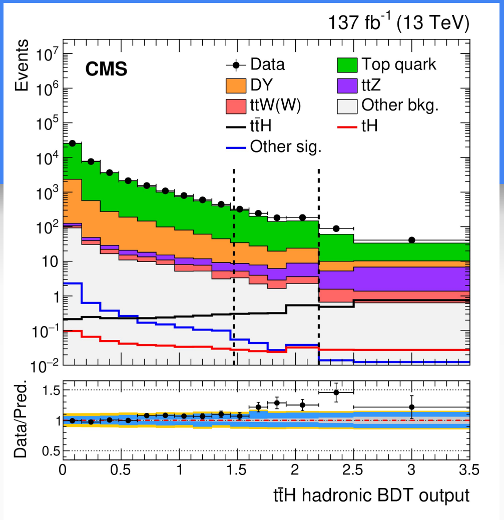
Weight signal events in BDT training by $1/\sigma(m_{\mu\mu})$

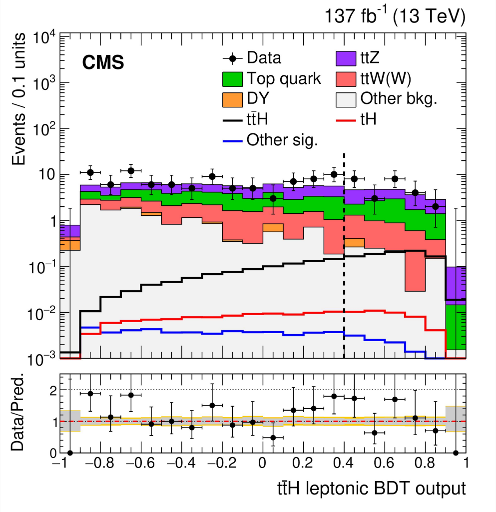
Event activity
- H_T , $H_T^{(miss)}$, $E_T^{(miss)}$, N(jets), p_T(jet), η(jet), b-tagging score
- ζ(projection of $E_T^{(miss)}$ on bisector of dimuon system)

Adding either observables aiming at the hadronic decay or the leptonic decay of the top quark

leptonic top
- m(lb)
- $\Delta\phi(\mu\mu, l)$
- $m_T(l,\ E_T^{(miss)}$
)

hadronic top
- RHTT -a BDT trained to identify top quarks decaying to three resolved jets
- p_T(top)

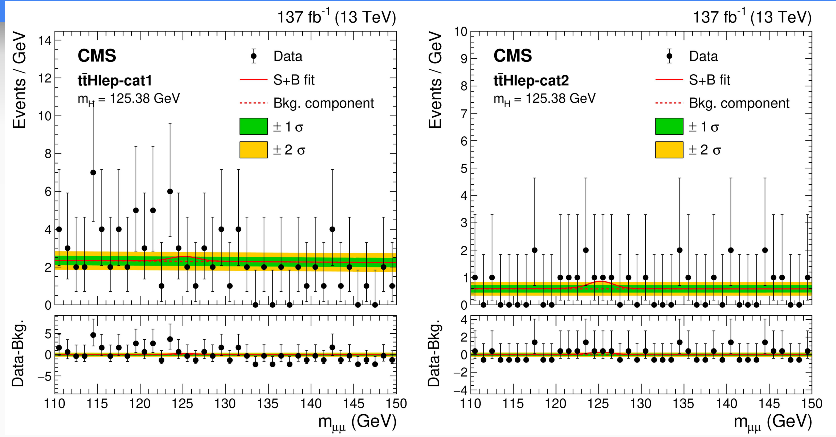


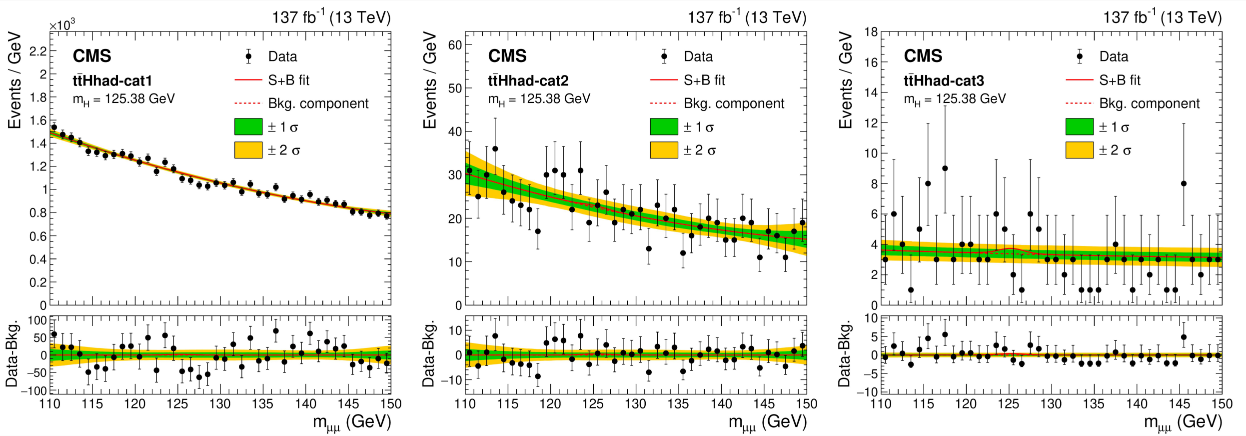


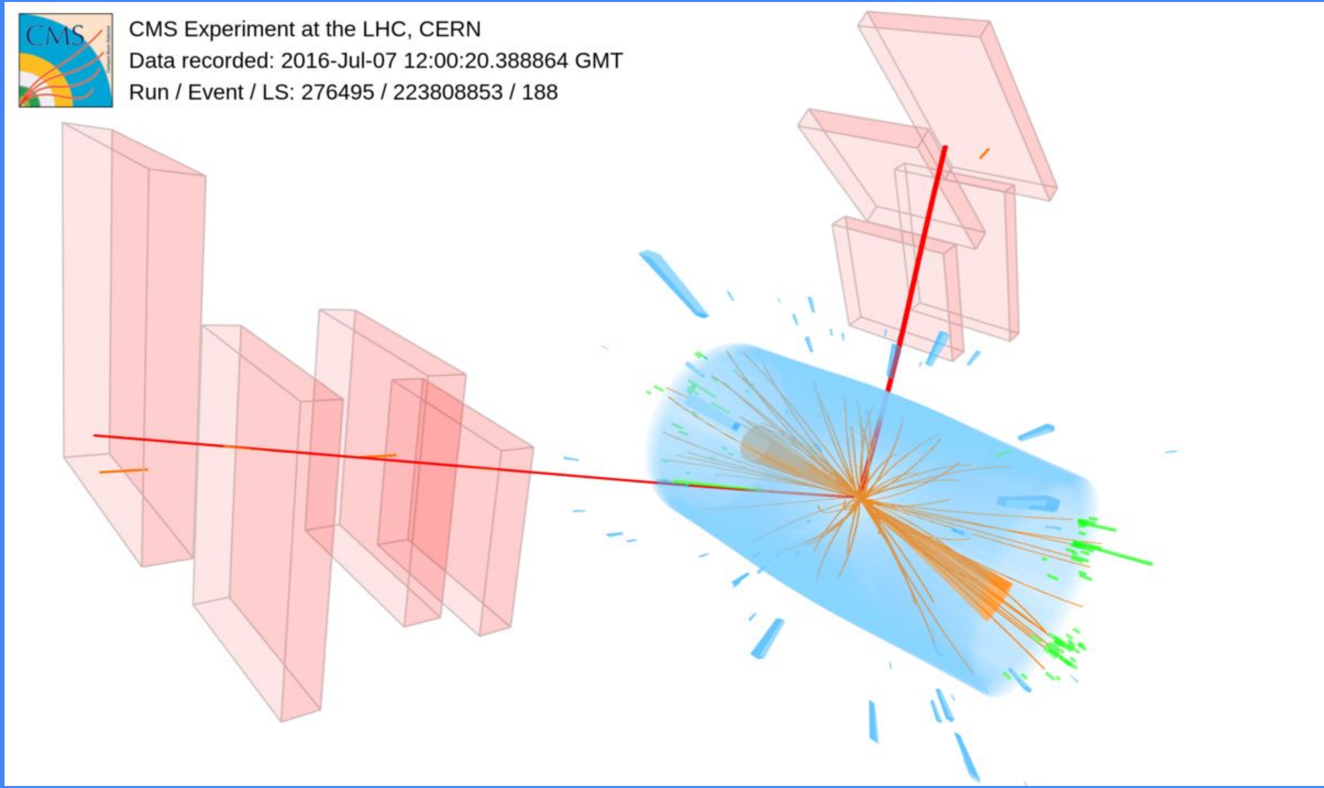
Bin each of the continuous BDT outputs into different boundaries by optimizing S/B

ttH fitting strategy

- Perform independent fits to M(μμ) in each BDT category
- Signal model: Double sided crystal ball function
- Background fit with simple exponential (ttH-lep) or polynomial (ttH-had).



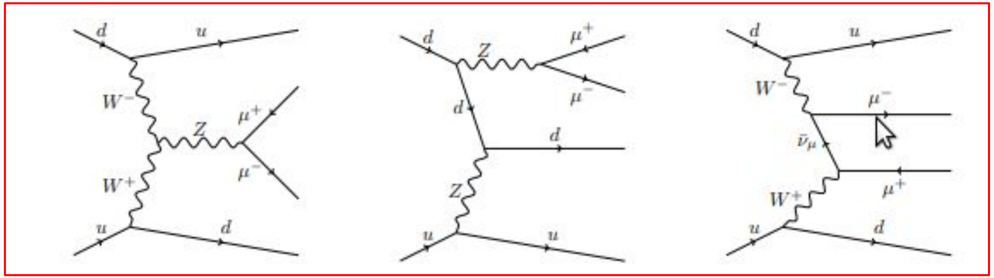



CMS Experiment at the LHC, CERN
Data recorded: 2016-Jul-07 12:00:20.388864 GMT
Run / Event / LS: 276495 / 223808853 / 188

VBF channel

Template based VBF channel

- Jet selection : p_T> 35 (25) GeV for the leading (sub-leading) jet and isolated from leptons
- At least two jets with M_{jj} > 400 GeV, $\Delta\eta$ (jj) >2.5
- Train a supervised machine learning classifier (a.k.a deep neural network or DNN) with M(μμ)
- **MC template based fit to DNN score** - Performance depends on statistical power of background MC in signal region and data/MC agreement
- Major backgrounds are Drell-Yan Z and electroweak production of Z



Electroweak production of Z+jj

VBF DNN

Dimuon variables

- $M(\mu\mu)$, $\delta M(\mu\mu)$, $\delta M(\mu\mu)/M(\mu\mu)$
- $p_T(\mu\mu)$, $\log(p_T(\mu\mu))$,
- η $(\mu\mu)$

Dijet variables

- $\log(M(jj))$, $M(jj)$,
- $\Delta\eta$ (jj)

Jet kinematics

- j_1 p_T, j_2 p_T
- j_1 η, j_2 η

Quark-Gluon Likelihood: exploit jet constituents and shape to discriminate between quark-jet and gluon-jet (j1 qgl, j2 qgl)

Collins-Soper variables

- $\cos(\theta^*)$
- $\cos(\varphi^*)$

Dimuon + Dijet system

- $Z^*(\mu\mu) = Z/(|\eta_{j1} - \eta_{j2}|)$, where $Z = \eta_{\mu\mu} - 0.5^*(\eta_{j1} + \eta_{j2})$
- $\min[\Delta\eta(\mu\mu, j1), \Delta\eta(\mu\mu, j2)]$

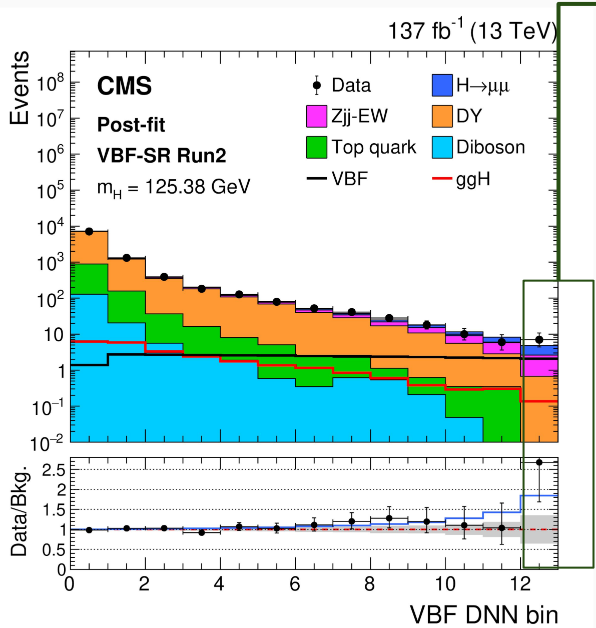
Soft jet variables (jets reconstructed by only charged tracks associated to the PV that are not associated with the 2 leading jets or the selected muons)

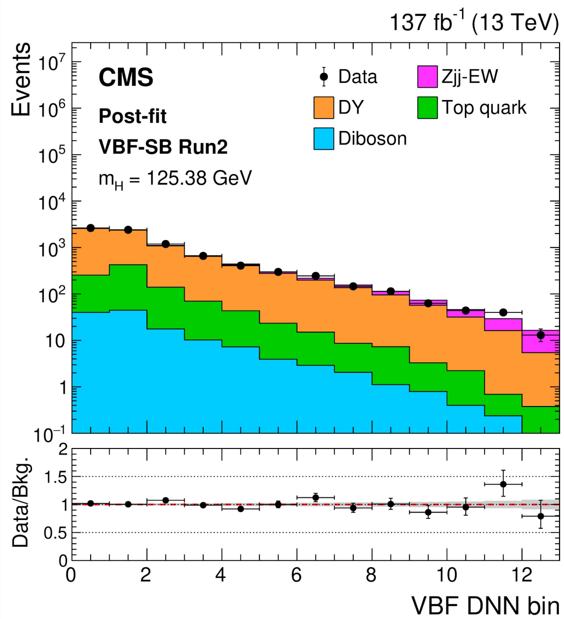
- HT of soft EWK jets with pt > 2 GeV
- Number of soft EWK jets with pt > 5 GeV

VBF fitting strategy

Two analysis regions based on $M_{\mu\mu}$

- Higgs peak: $M_{\mu\mu} \in [115,135]$ GeV
- Higgs sideband: $M_{\mu\mu} \in [110, 115]$ GeV $\cup$ $[135, 150]$ GeV





Binned template fit to DNN score

- last DNN bin signal purity 42%
- Simulation provides a background prediction with better precision, including systematics
- 20% improvement in sensitivity compared to a parametric fit to $M_{\mu\mu}$
- Fit performed simultaneously across data-taking periods and in signal and sideband regions

Inverting the VBF DNN score fit to get $M_{\mu\mu}$

- $M_{\mu\mu}$ distribution for the weighted combination of VBF-SB and VBF-SR events.
- Each event is weighted proportionally to the S/(S + B) ratio, calculated as a function of the mass-decorrelated DNN output (replacing the dimuon mass input to the DNN with a fixed value of 125 GeV).
- Disclaimer - This plot isn't used to derive the final combined result - just a way to visualize the $M_{\mu\mu}$ spectrum