Search for H→aa→μμbb

CMS-HIG-21-021

Elham Khazaie

On behalf of the CMS collaboration

Isfahan University of Technology

The 19th Workshop of the LHC Higgs Working Group

Mon, 28 Nov 2022

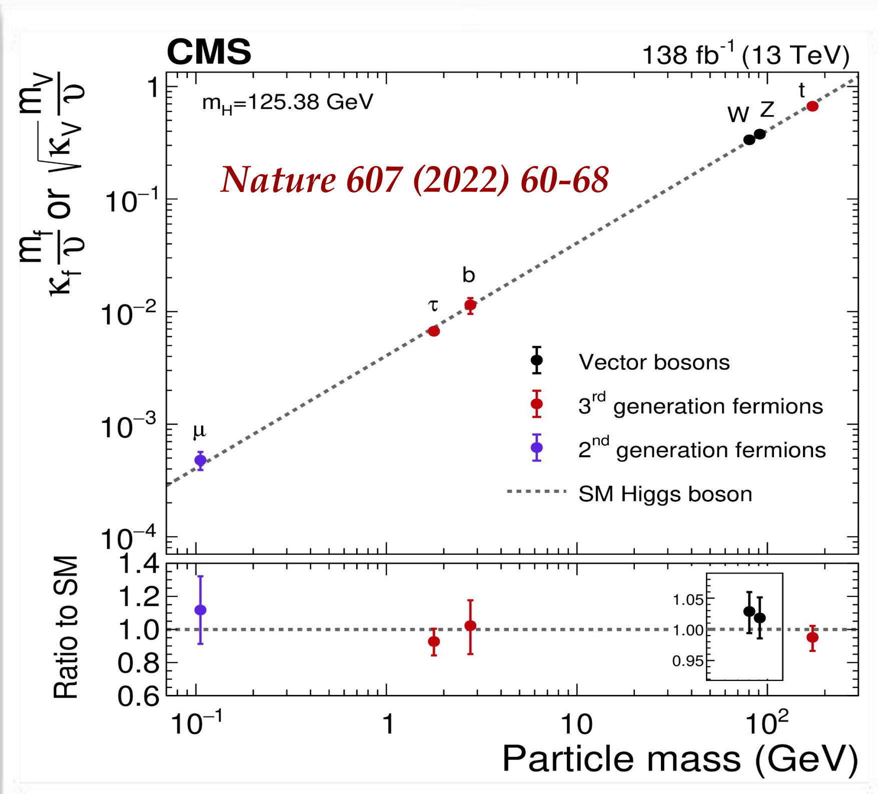
Outline

- General motivation of the search
- Analysis strategy
- Results
- Summary

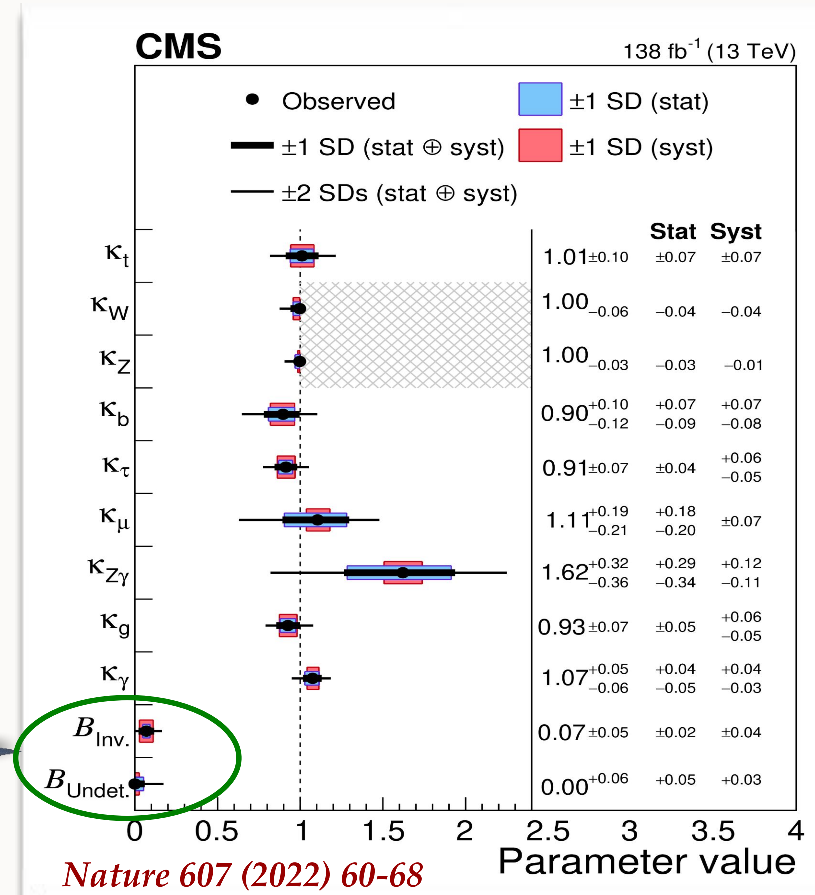
General motivation to search for exotic Higgs decays

CMS-HIG-21-021

- **Journal Nature:** most comprehensive overview of the Higgs boson published by CMS and ATLAS
  - Data in agreement with SM predictions, so far
  - $\mathcal{B}_{\text{Undet.}} < 0.16$ and $\mathcal{B}_{\text{Inv.}} < 0.16$
    - There are rooms for the Higgs boson to connect with BSM



Search for new physics



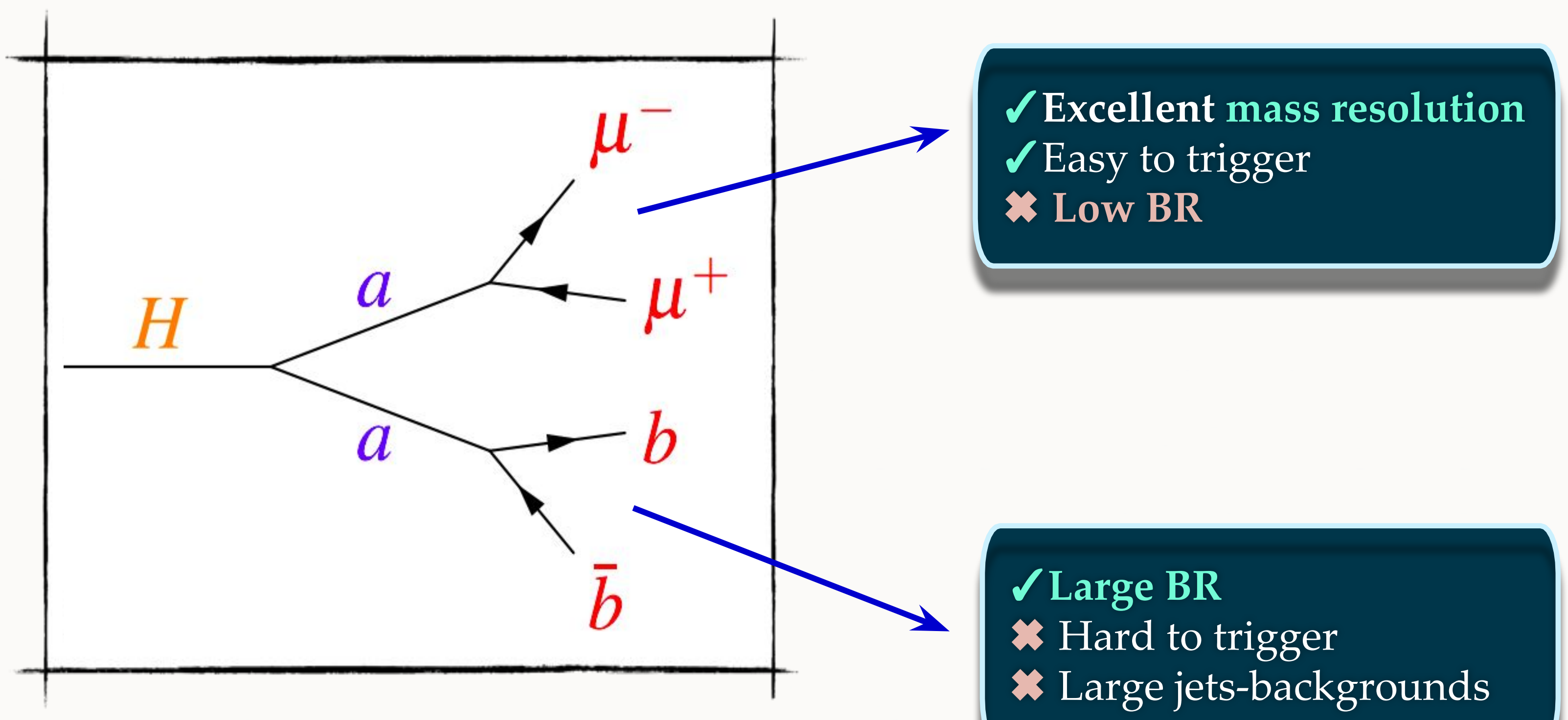
Exotic decays of the Higgs boson provide a natural and efficient way for probing new physics

Exotic Higgs decays : H → (pseudo)scalars

CMS-HIG-21-021

- Some extensions of the SM include Higgs decays via a pair of on-shell (pseudo)scalars, e.g. **2HDM+S**
- The (pseudo)scalar ***a*** decays into fermions via mixing with the Higgs boson
- **$\mathcal{B}$ (*a*→ xx)** depend on the model
- Focus especially on the decay of the pseudoscalar ***aa* → μμbb**
  - **Main backgrounds:** tt and Drell-Yan
  - **Background estimation :** Fully data driven approach

H→aa→μμbb



Analysis strategy

CMS-HIG-21-021

Preselection →

Optimization of kinematics & identification requirements ↓

Model dependent selection

$$\chi_{bb} = \frac{(m_{bb} - m_{\mu\mu})}{\sigma_{bb}}$$

$$\chi_{h} = \frac{(m_{h} - 125)}{\sigma_{h}}$$

↓

Categorization →

Bump hunting over $m_{\mu\mu}$
Fully data-driven background estimation with discrete profiling method, and limit extraction

Analysis strategy

CMS-HIG-21-021

Preselection

Optimization of kinematics & identification requirements

Model dependent selection

$$\chi_{bb} = \frac{(m_{bb} - m_{\mu\mu})}{\sigma_{bb}}$$

$$\chi_h = \frac{(m_h - 125)}{\sigma_h}$$

Categorization

Bump hunting over $m_{\mu\mu}$
Fully data-driven background estimation with discrete profiling method, and limit extraction

Pre-selection and optimization

CMS-HIG-21-021

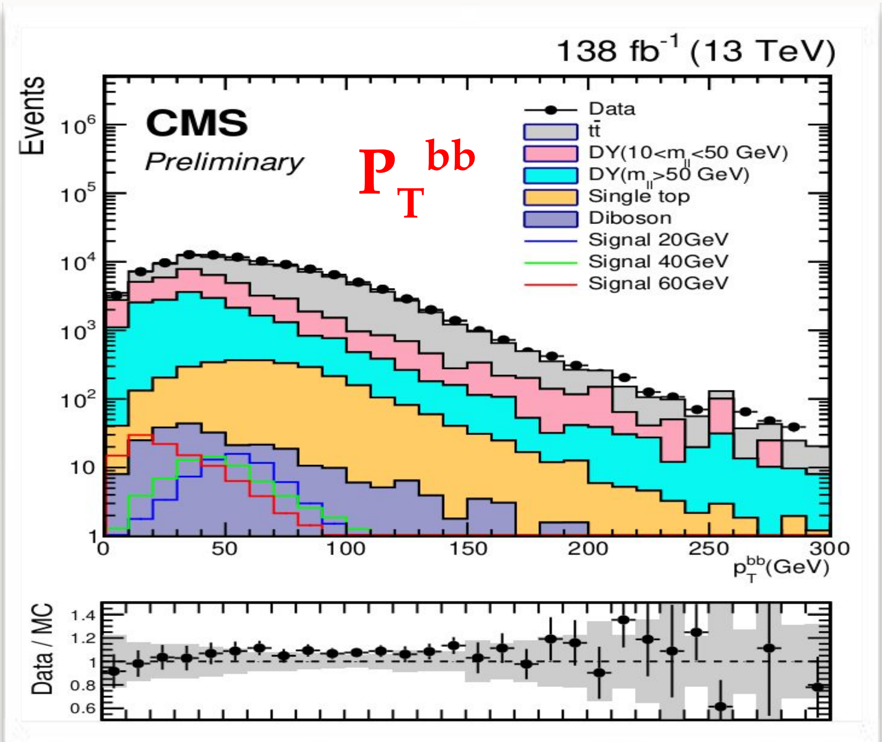
Pre-Selection :

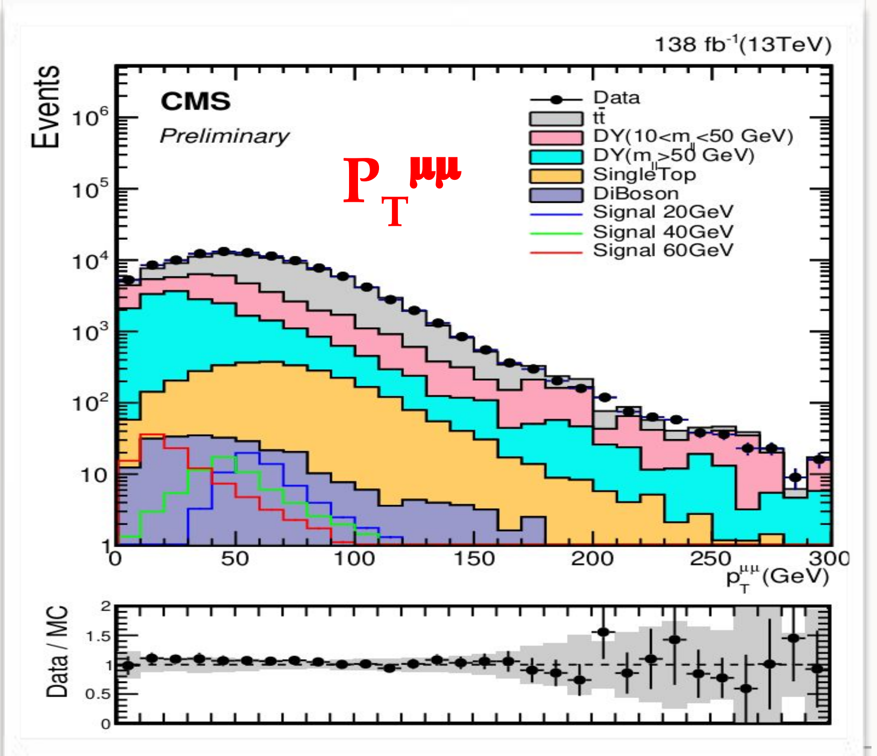
- Required at least two b-jets :
  - $P_T > 15$ GeV
- Required 2 opposite sign Muons :
  - $P_T > (15,17)$ GeV
  - $14 \text{ GeV} < m_{\mu\mu} < 70 \text{ GeV}$
- MET < 60 GeV

Optimization :

- Figure of merit for optimization: **significance** ($S/\sqrt{B}$)
- Optimization:
  - **B-tagging algorithm and working points, muon Id/Iso**
- Try to have a uniform selection vs. m_a
- **Optimal selection:**
  - **Mu: TightIdLooseIso** , **balgo: deepjet** , **bwp: TL**

Data/MC comparison after optimization procedure





LHC Higgs Working Group

Analysis strategy

CMS-HIG-21-021

Preselection → Optimization of kinematics & identification requirements

↓

Model dependent selection

$$\chi_{bb} = \frac{(m_{bb} - m_{\mu\mu})}{\sigma_{bb}}$$

$$\chi_h = \frac{(m_h - 125)}{\sigma_h}$$

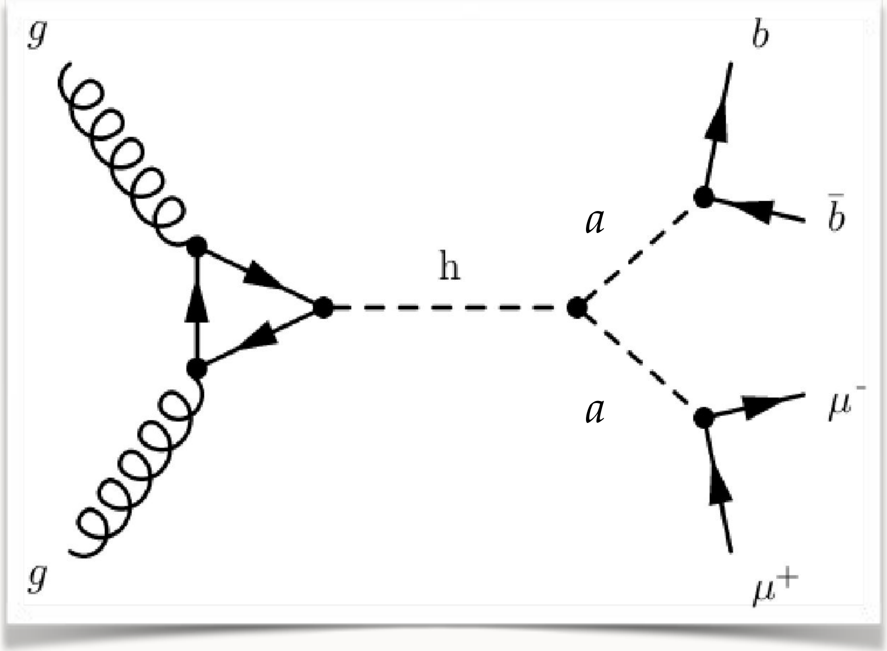
↓

Categorization → **Bump hunting over $m_{\mu\mu}$**
Fully data-driven background estimation with discrete profiling method, and limit extraction

χ^2 definition and selection

CMS-HIG-21-021

- Model dependent selection : exploit two features in signal
    - $m_{bb} = m_{\mu\mu}$
    - $m_{\mu\mu bb} = 125$ GeV



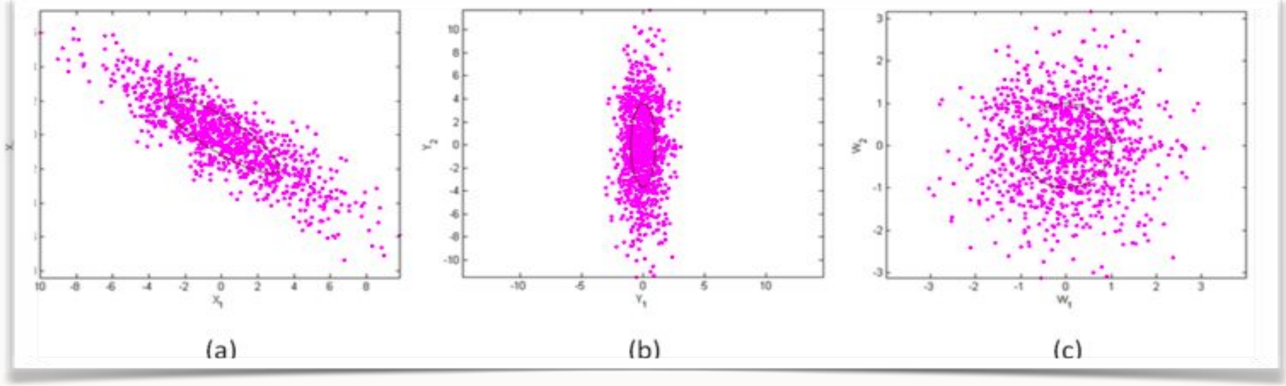
$$\chi_{bb} = \frac{(m_{bb} - m_{\mu\mu})}{\sigma_{bb}} \quad , \quad \chi_h = \frac{(m_h - 125)}{\sigma_h}$$

$$\chi^2 = \chi^2_{bb} + \chi^2_h$$

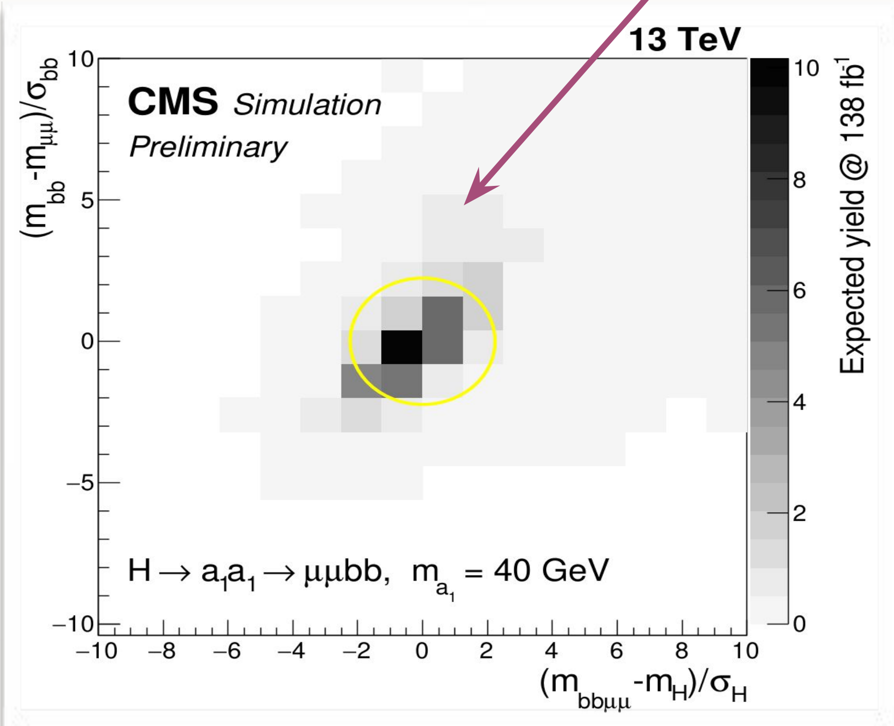
- Resolutions derived from a Gaussian fits to m_{bb} and $m_{\mu\mu bb}$, where objects are matched to the MC truth
- σ_{bb} changes as a function of m_a , consider this change in χ_{bb} calculation
- χ^2 : discriminator variable between signal and backgrounds

χ_{bb} vs χ_h

- Looks like an ellipse and depends on m_a
  - placing a circular cut ($\chi^2 < X$) is suboptimal
- χ_{bb} and χ_h must be decorrelated using PCA method
- New variable χ_d^2 is defined : $\chi_d^2 = \chi_{b,d}^2 + \chi_{h,d}^2$



Two variables are correlated

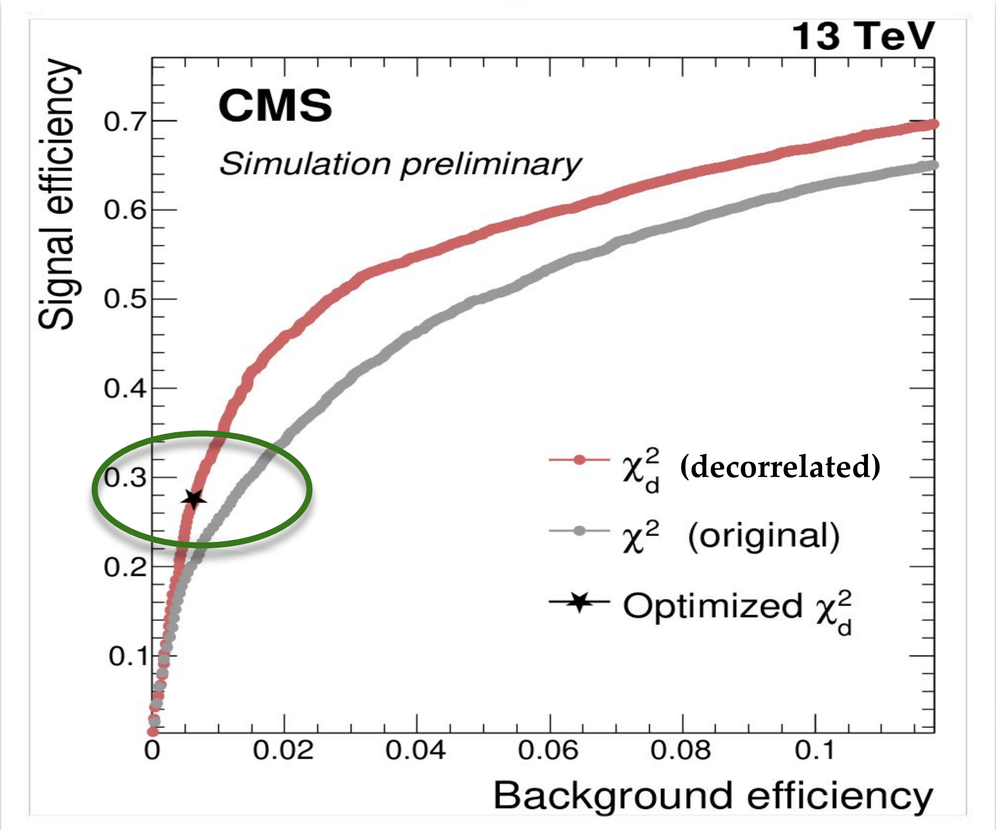


Improved performance of the χ^2_d and χ^2

CMS-HIG-21-021

- Optimized cut value : $\chi_d^2 < 1.5$
- A large fraction of the backgrounds rejected, while most of the signal remains
- Comparison of the selection performance of the χ^2_d and χ^2 in terms of efficiencies for signal (m_a= 40 GeV) and backgrounds

A clear improvement is visible after decorrelation



Analysis strategy

CMS-HIG-21-021

Preselection

Optimization of kinematics & identification requirements

Model dependent selection

$$\chi_{bb} = \frac{(m_{bb} - m_{\mu\mu})}{\sigma_{bb}}$$

$$\chi_h = \frac{(m_h - 125)}{\sigma_h}$$

Categorization

Bump hunting over $m_{\mu\mu}$
Fully data-driven background estimation with discrete profiling method, and limit extraction

Categorization

CMS-HIG-21-021

- A categorization of events is performed, to increase the sensitivity of the analysis
- Events are categorized according to the b-jet P_T, event's compatibility with VBF Higgs production, and the score of the b-jets
- The **Low P_T** category brings extra sensitivity to the signal with lower m_a values.

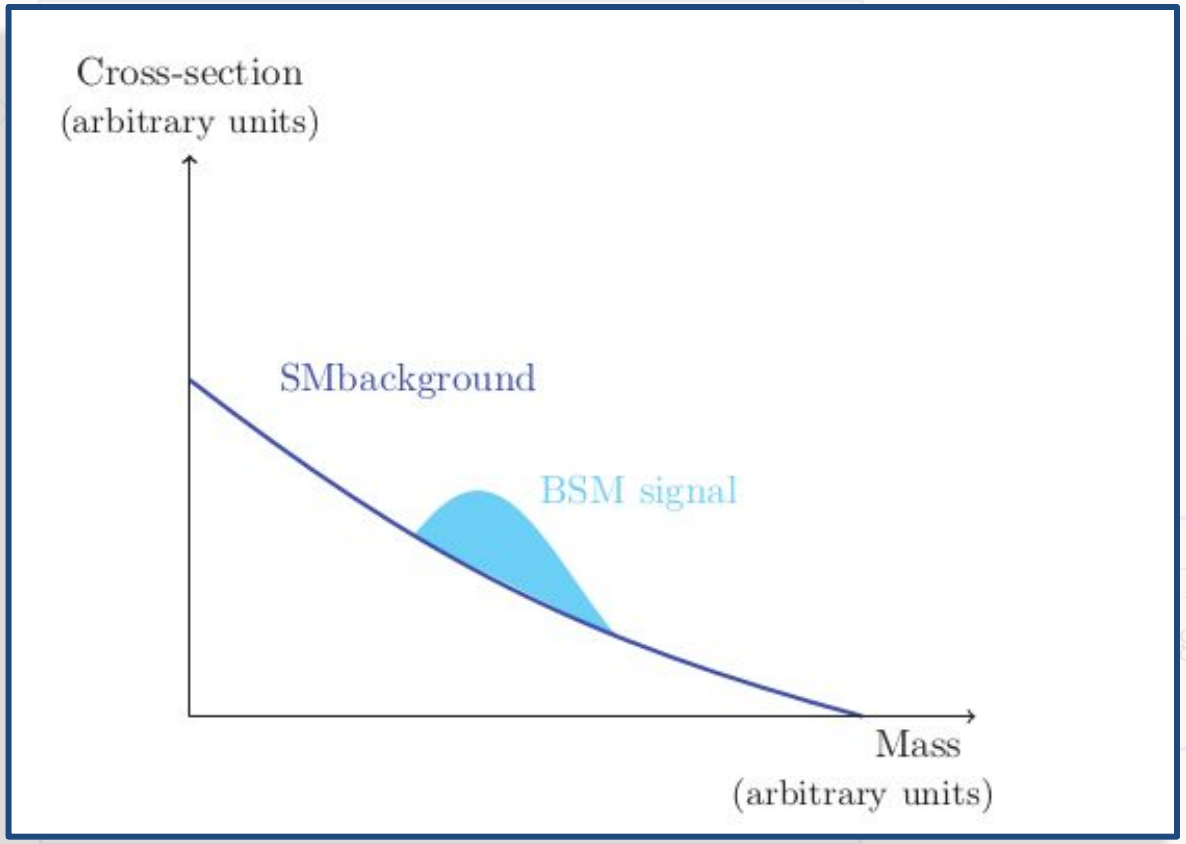
Categories for selected events	
Low P_T	At least one b-jet with $P_T < 20$ GeV
VBF	Two add. Jets with $P_T > 30$ GeV, $\|\eta\| < 4.7$, and $m_{jj} > 250$ GeV
TL	Looser b-jet passes L but fails M
TM	Looser b-jet passes M but fails T
TT	Looser b-jet passes T

- The majority of background events (≈ 70%) fall into the **LowP$_T$** category
- For signal (ggH): ~ 40% in **LowP$_T$**

Input models and limits

CMS-HIG-21-021

Preselection



hile blind, defining a control data
gion to model the background for
expected limits

Categorization based on
b-tagging score of b-jets

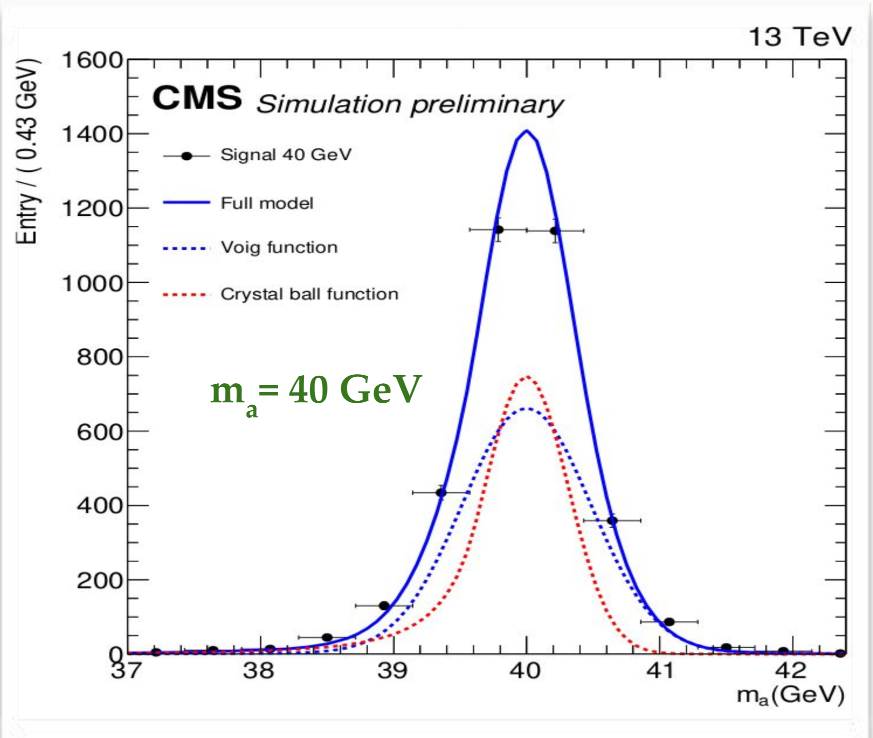
Bump hunting over m$_{\mu\mu}$
Fully data-driven background
estimation with discrete profiling
method, and limit extraction

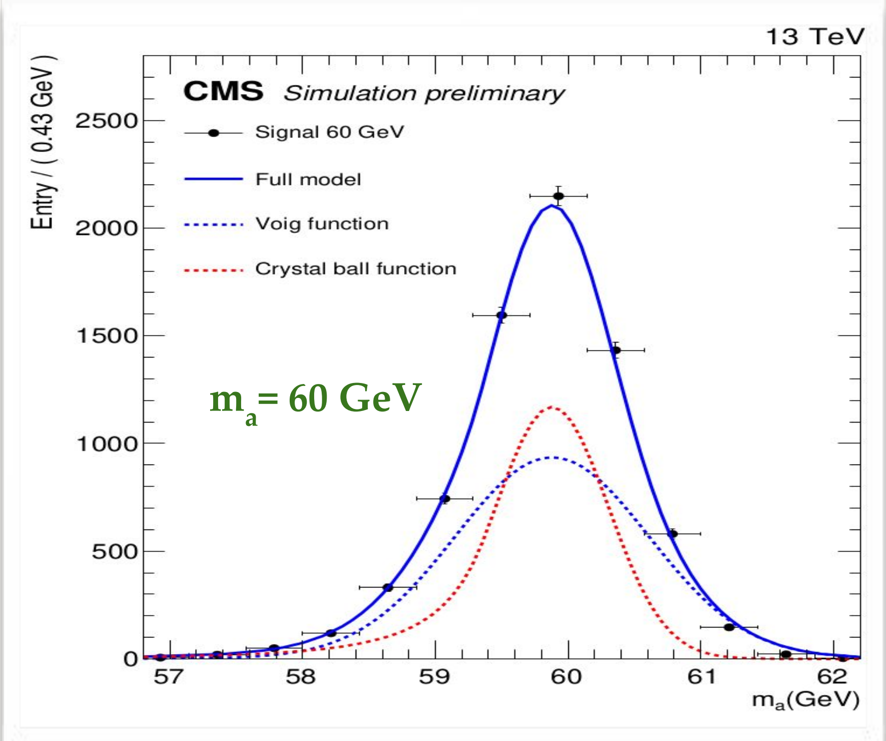
Signal Shape Modeling *(a la HIG-18-011)*

CMS-HIG-21-021

- The shape of signal is estimated from simulation
- Signal is modeled with a combination of a **Voigt** and a **CrystalBall (CB)** profile.

$$S(m_{\mu\mu}|f, p_V, p_{cb}) \equiv f \cdot V(m_{\mu\mu}|p_V) + (1-f) \cdot CB(m_{\mu\mu}|p_{cb})$$

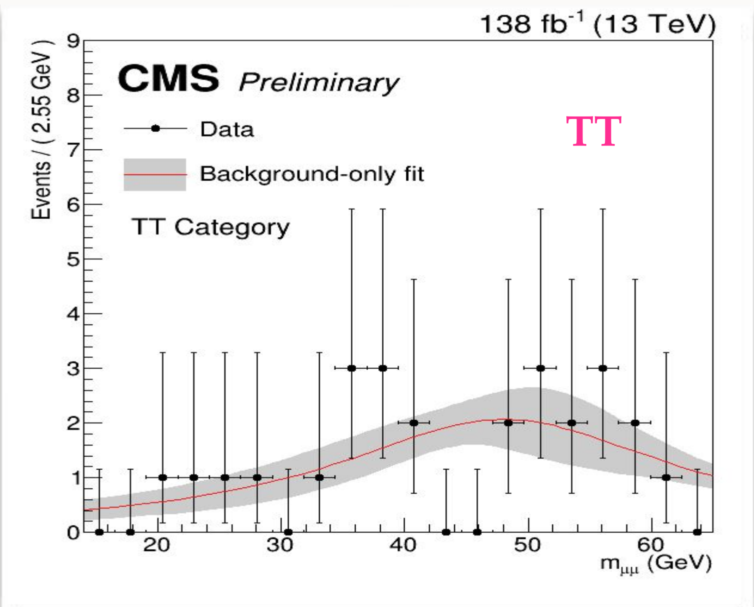


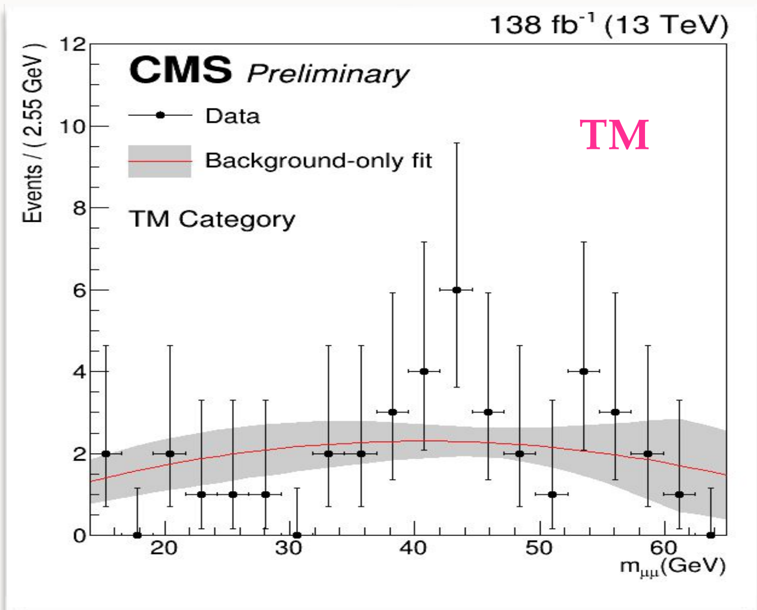


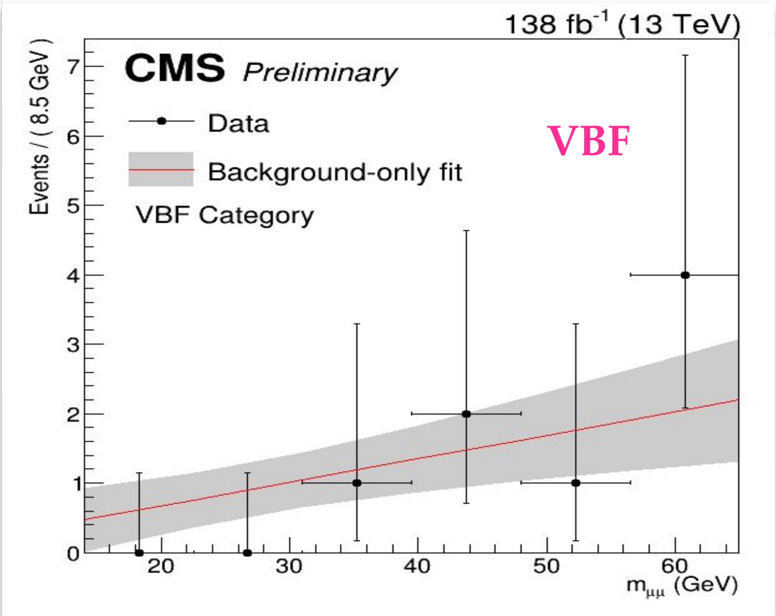
Background Modeling

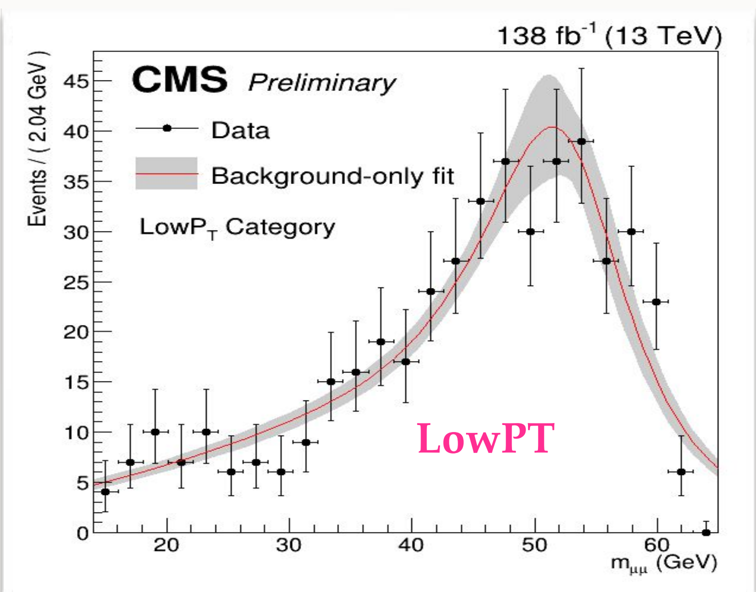
background-only fit results

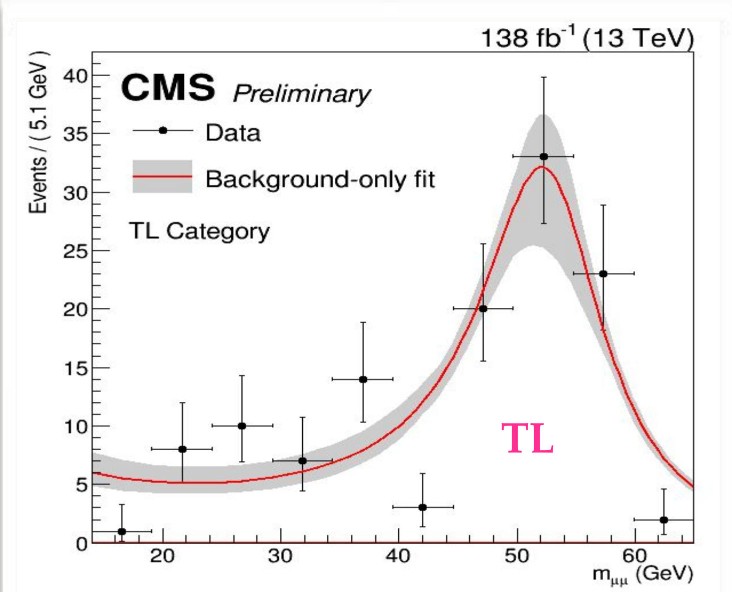
CMS-HIG-21-021









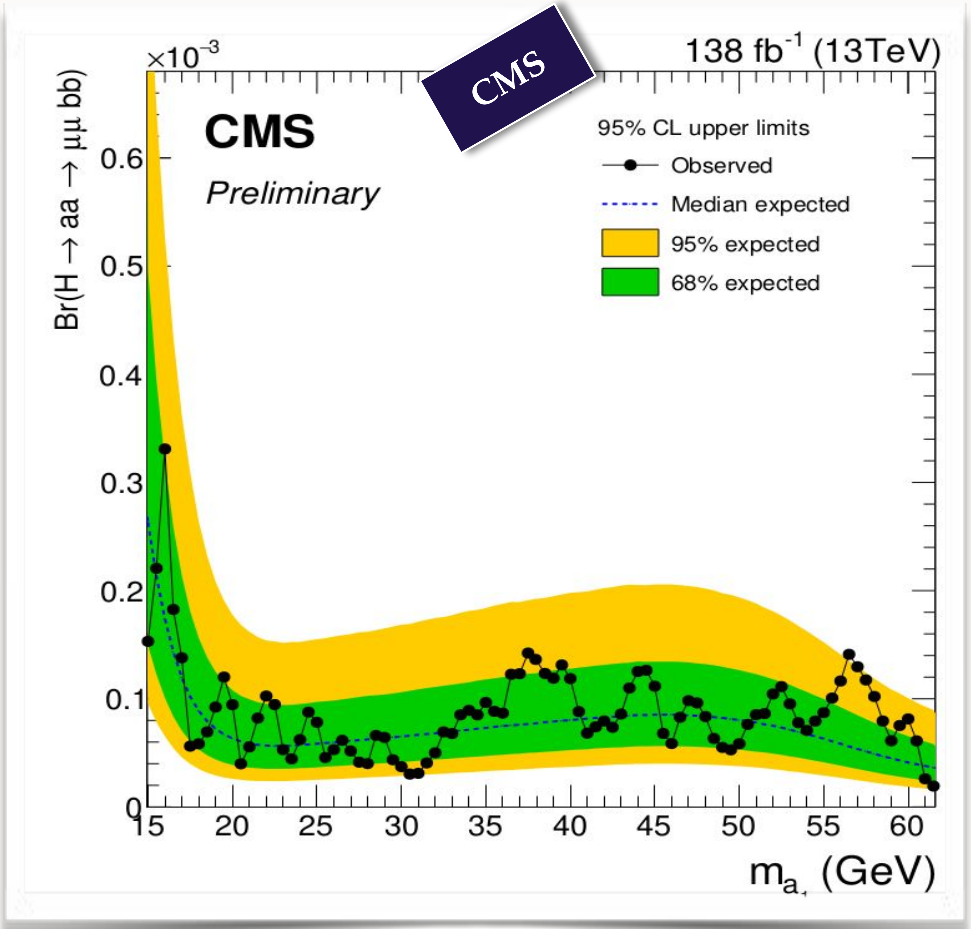


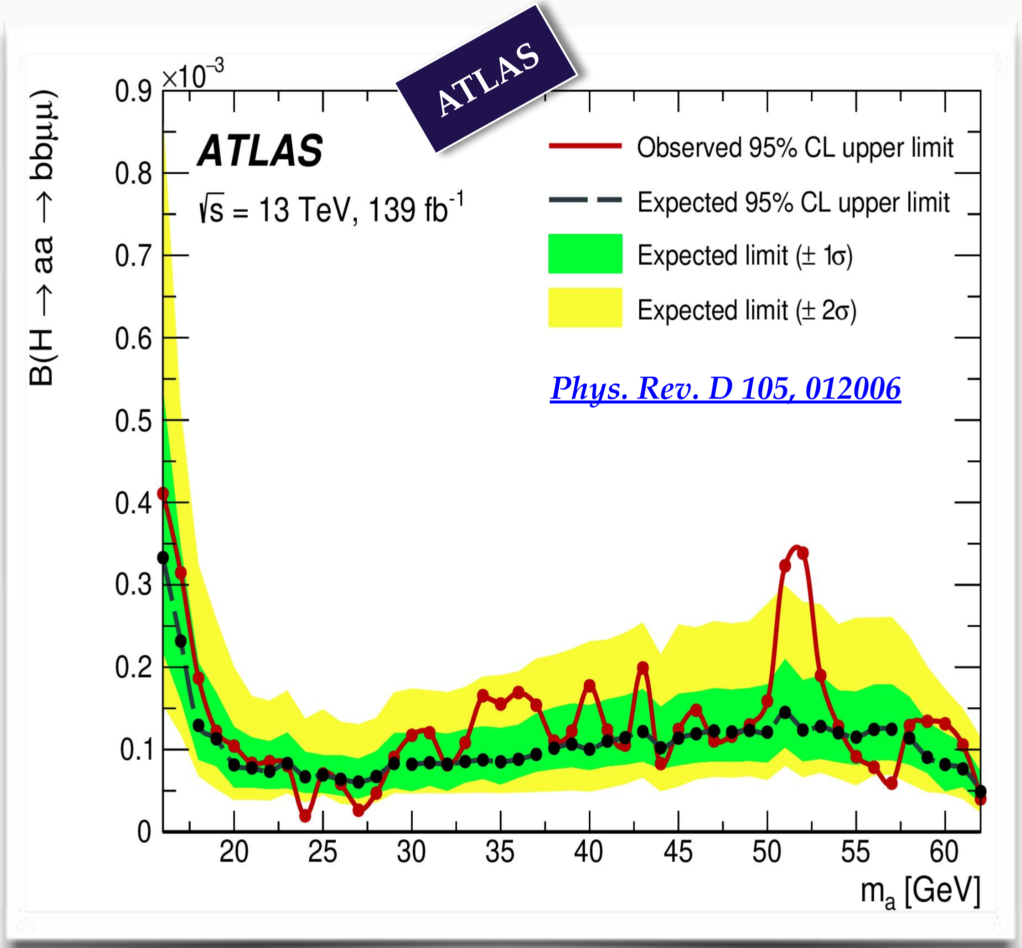
- **Estimated by a fully data-driven approach, using envelope method**
- Results on background only fit for LowP$_T$, TM, TT, TL and VBF categories
- The uncertainty bands correspond to the best-fit background model uncertainty as extracted from the fit to the data

Results
Observed/ Expected Limits

CMS-HIG-21-021

- At 95% CL, upper limits on **$\mathcal{B}$(H →aa→μμbb)** for the mass range 15 to 62.5 GeV:
    - **Observed: (0.17 – 3.3) × 10^{-4}**
    - **Expected: (0.35 – 2.6) × 10^{-4}**
- Large improvement with respect to 2016 beyond the increase of luminosity





Latest ATLAS results :

$\mathcal{B}$ (H →aa→μμbb) < obs. (0.22 – 4) × 10^{-4} exp. (0.5-5) × 10^{-4}

Summary

- Search for exotic Higgs decays in μμbb final state
- Full Run 2 dataset is analyzed
- No excess is found over the SM backgrounds
  - Upper limits are reported on Br ($H \to aa \to \mu\mu bb$)
- Improved analysis strategy resulted in better expected and observed limits compared with 2016 analysis, beyond the effect of luminosity increase
- Looking forward to surprises in LHC Run 3!

"Everything is a mathematical trick,
except what you measure in the lab."
(Prof. Dr. Armin Scrinzi)

Thank you for your attention !

BACKUP

BSM theoretical framework

CMS-HIG-21-021

2HDM:

- 2HDM is an extensions of the SM in the Higgs sector
- Separated in 4 types, depending on how the SM fermions interact with the two Higgs doublets
- After EWKSB, the 2HDM leads to 5 physical states:
    - **Charge scalar pair H±,**
    - **Neutral pseudoscalar A,**
    - **Neutral scalar H0 ,**
    - **Neutral scalar h, $m_h < m_{H0}$, can be SM-like**

	Type-I	Type-II	Type-III	Type-IV
l	φ_2	φ_1	φ_1	φ_2
u	φ_2	φ_2	φ_2	φ_2
d	φ_2	φ_1	φ_2	φ_1

2HDM+S:

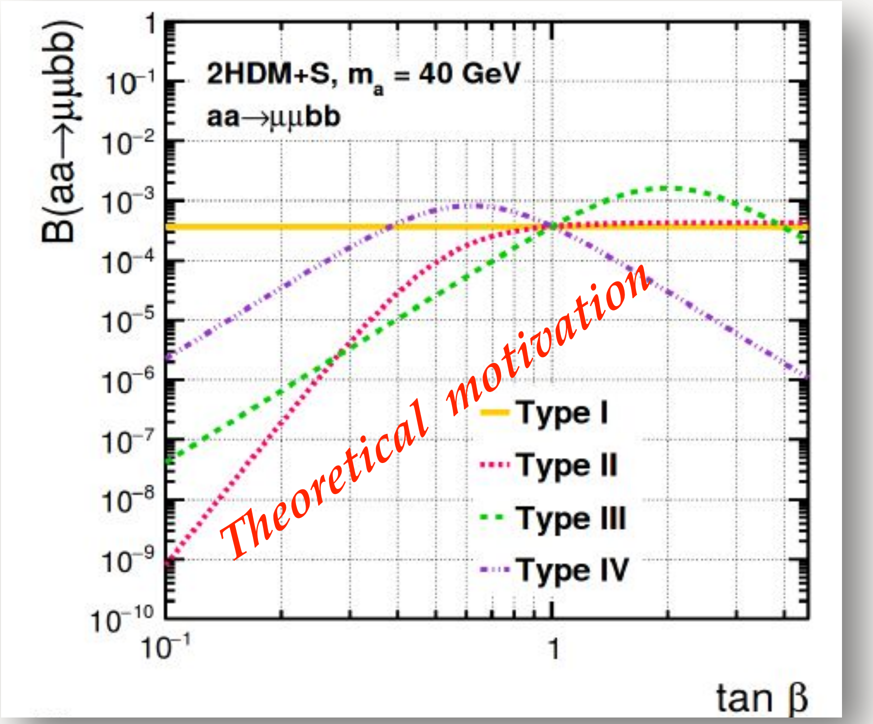
- 2HDM has been constrained from experimental data, one way is to extend the 2HDM
- We consider the extensions of 2HDM **(2HDM+S)** that has **NMSSM** as a special case
- A complex scalar singlet is added to the already present scalar doublets.
- Because of the additional singlet, two new bosons are introduced: *s(a) (pseudo)scalar*
- **h** boson can decay to fermions through *s(a)*

LHC data are used to search for this exotic decay and set limit on Br (h →aa→ff)

PhysRevD .90.075004

The μμbb final state

CMS-HIG-21-021



aa→ μμbb

The largest BR(aa → μμbb) is obtained for tan β >2 in 2HDM+S type III.

Simulated background samples

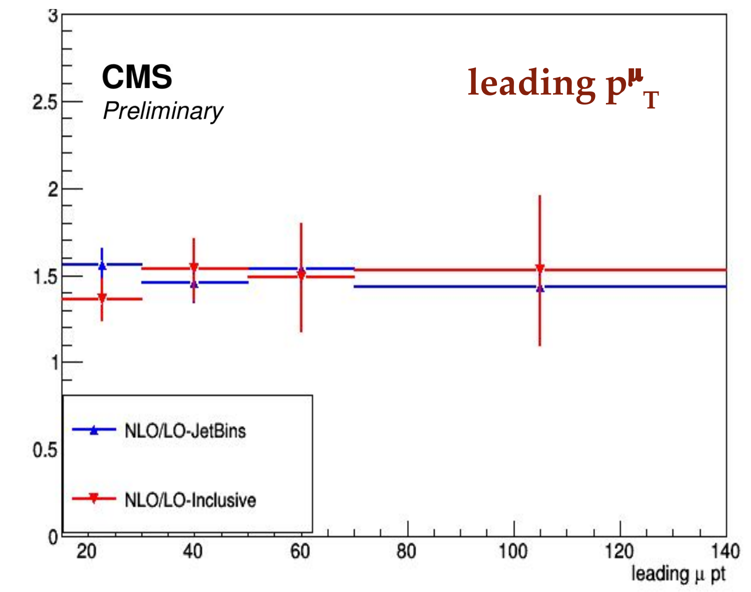
CMS-HIG-21-021

- Summary of simulated background samples :

Samples	2016 Dataset	2017 Dataset	2018 Dataset
Drell-Yan ($10<m_{ll}<50$ GeV)	NLO, LO sample exclusive in number of additional partons (up to four)	Leading Order	Leading Order
Drell-Yan ($m_{ll}>50$)	NLO, exclusive in number of additional partons (up to two)		
ttbar	Next to Leading Order		
Single Top	Next to Leading Order		
Diboson	Leading Order		

- The NLO low-mass Drell-Yan samples are only available in 2016
    - A k-factor of 1.5 is extracted from 2016 samples and applied in all years.
    - An uncertainty of 30% is considered on the k-factor.
- **Note:** Background samples are used only for the selection optimization purposes
- The final background contribution is determined fully based on data with no reference to MC

NLO-to-LO ratios



Principal component analysis (PCA) as a decorrelation method

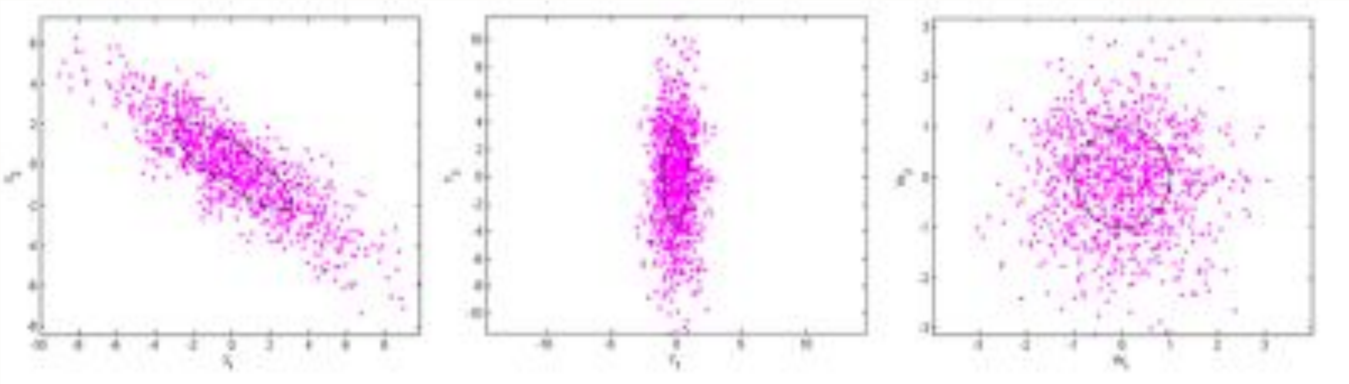
CMS-HIG-21-021

$$\begin{pmatrix}\chi_h \\ \chi_{bb}\end{pmatrix} \longrightarrow \Sigma = \begin{pmatrix} v(\chi_h) & cov(\chi_h,\chi_{bb}) \\ cov(\chi_{bb},\chi_h) & v(\chi_{bb}) \end{pmatrix} \longrightarrow \begin{cases} E0 = \begin{pmatrix} a \\ b \end{pmatrix}, & \lambda_0 \\ E1 = \begin{pmatrix} -b \\ a \end{pmatrix}, & \lambda_1 \end{cases}$$

- **(Σ) decomposed as a sequence of rotation and scaling operations on uncorrelated data (white)**
  - **(E) : Rotation matrix defined by the eigenvectors of Σ .**
  - **(C) : Scaling matrix defined by the eigenvalues of Σ**

$$\begin{cases} C = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \\ E = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \end{cases}$$

$$W = C^{-1/2}E^T \qquad W = \begin{pmatrix} a/\sqrt{\lambda_1} & b/\sqrt{\lambda_1} \\ -b/\sqrt{\lambda_2} & a/\sqrt{\lambda_2} \end{pmatrix}$$



(a) (b) (c)

$$\begin{cases} \chi_{hd} = \dfrac{a}{\sqrt{\lambda_1}}(\chi_h + \chi_{bb}\tan\alpha) \\ \chi_{bbd} = \dfrac{a}{\sqrt{\lambda_2}}(-\chi_h\tan\alpha + \chi_{bb}) \end{cases}$$

Refers to decorrelated variables

Signal and background yields / Categorization

CMS-HIG-21-021

- **Signal yields:** for non-simulated mass points are estimated by extrapolation, using SPLine
  - Benchmark: Br(H -> aa) = 10%, 2Br(a→bb)Br(a→$\mu\mu$)=1.7×10^{-3}
- **Background** is ultimately determined from **data** (envelope method)

Categorization

- For signal (ggH): ~40% in **LowP$_T$**
- Different ratios of signal samples fall into each category.
- The majority of background events (≈ 70%) fall into the **LowP$_T$** category
- (≈20%) of background events fall into the TL category, about 10% pass the TM category, less than 10% can meet the TT criteria
- For signal (vbF) : about 40% in this category

Systematic Uncertainties

CMS-HIG-21-021

Background:

- Estimated by a fully data-driven approach, using discrete profiling method (envelope method)
- Systematic uncertainties on background modeling are taken into account

Signal:

- **Signal shape:**
  - Estimated by a parametric signal model (MC-based)
  - Uncertainties on signal model parameters are found to be negligible (2016, RunI paper)
- **Signal normalization**:
  - Is affected by various sources of systematic uncertainties:

- **Modeling**
  - PS
  - QCD Scales
  - PDF
- **Muon**
  - ID/Iso SF
  - HLT: 1%
- **Jets**
  - btag SF
  - JES/JER
- **Other sources**
  - Prefire
  - PU
  - Lumi: Followed LUM POG recipe
  - Xsection : ggH+VBF, considered for the limit on BR

Uncertainties on signal shape are negligible, their effects on yields are taken into account as nuisance parameters in the fit

Background Modeling

- **Estimated by a fully data-driven approach, using envelope method**
- Evaluated through a fit to the μμ mass distribution in data w/o any reference to MC
- Different types of functional forms are used on data in SR
  - RooPolynomial , RooBernstein, RooChebychev, Inverse polynomial ($1/P_n$)
  - All of them are considered in the extracted limit
- A **F-test** is used to determine the collection of pdfs for each family
- Profile likelihood method is performed using combine
- Results on background only fit for LowP_T , TM, TT, TL and VBF categories.
- The uncertainty band corresponds to the best-fit background model uncertainty as extracted from the fit to the data.

Optimization procedure/results

CMS-HIG-21-021

- Muon working point: **Tight ID/ Loose Iso**
- b-tagging algorithm: **DeepJet** (It outperforms the other taggers)
- The most optimum b-tagging working point :
  - **One Tight/ One Loose**
  - Similar to 2016 published paper
  - Similar working point is optimized for 2016/2017

Optimum selection: Mu: TightIdLooseIso , balgo: deepjet , bwp: TL