Search for BSM Higgs bosons decaying into b quarks

Paul Asmuss

CMS BSM Hbb Group: Chayanit Asawatangtrakuldee[2], Sandra Consuegra Rodriguez[1], Elisabetta Gallo[1,3], Clemens Lange[4], Daina Leyva Pernia[1], Rainer Mankel[1], Alexei Raspereza[1], Roberval Walsh[1]

1 - DESY | 2 - Chulalongkorn University | 3 - University of Hamburg
4 - CERN

October 25, 2021

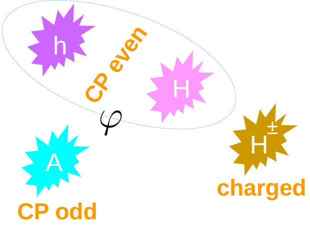
LHC Physics Discussion

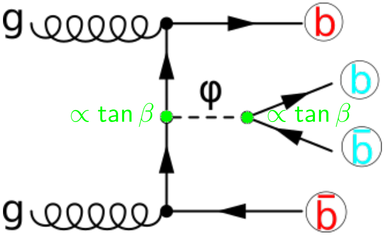
HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

Motivation

- Strong indications:
  Standard Model (**SM**) is incomplete
- Many BSM theories contain two Higgs doublets (**2HDM**)
  - Five physical Higgs bosons
  - Example: Minimal supersymmetric extension of the SM (**MSSM**)
- Coupling to b quarks enhanced
  $\propto \tan\beta = (v_2/v_1)$
  $\Rightarrow$ large $\tan\beta \rightarrow$ enhanced cross section for b-associated production
- Extending and improving public 2016 analysis with 2017 data

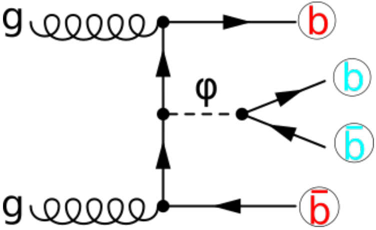
2016 public result:
JHEP08(2018)113





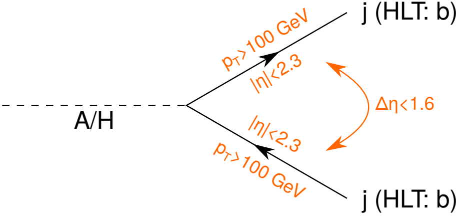
Analysis Strategy

- Search for excess in invariant mass of p_T-leading two jets, m_{12}
  - Candidates for Higgs daughters
  - Main background: **QCD** multijet
- At least 3 b-tagged jets required for signal region (**SR**)
- b-jet energy regression included
- FSR recovered for selected jets



Trigger

- To address huge amount of QCD multi-jet events: dedicated trigger
- Two steps: L1 and HLT
  - Same kinematic requirements (calorimeter jets for L1, PF jets for HLT)
  - HLT includes b-tag (CSVv2 algorithm); working point allowing 0.33 % light-flavor mis-identification probability
- Determination of trigger efficiency in data and Monte Carlo (MC)
  - B-tagging
  - Jet kinematics



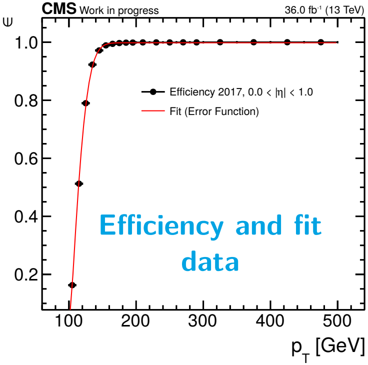
Trigger Efficiency

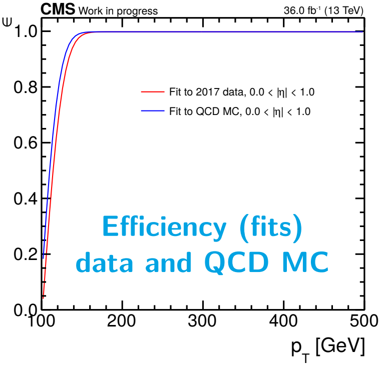
Jet Kinematics

- Use reference trigger objects to calculate efficiency
  - Single jet trigger
  - Select two jets back-to-back $\rightarrow$ unbiased and 100 % efficient reference

$$\epsilon = \frac{\text{reference} \times \text{actual}}{\text{reference}}$$

$$sf = \frac{\epsilon_{\text{data}}}{\epsilon_{\text{MC}}}$$



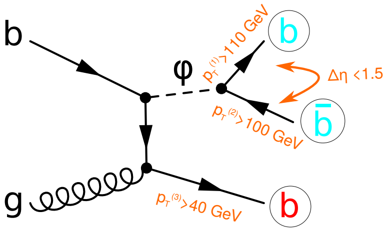


Offline Selection

- $\geq$ 3 jets, anti-k_T algorithm (cone radius parameter 0.4)
- $|\eta| < 2.2$ (leading three jets)
- ΔR_{ij} between leading three jets ≥ 1.0 $(\Delta R_{ij} = \sqrt{\Delta\eta_{ij}^2 - \Delta\phi_{ij}^2})$
- b-tagging, defining analysis regions with similar m_{12} distribution:
  - **SR**: 3 b-tags, leading three jets in p_T (discriminant > medium wp)
  - **CR**: 2 b-tags, third jet in p_T not b-tagged (discriminant < loose wp)
    To determine background shape
  - **VR**: 2 b-tags, third jet intermediate b-tagging discriminant
    To validate fitting procedure

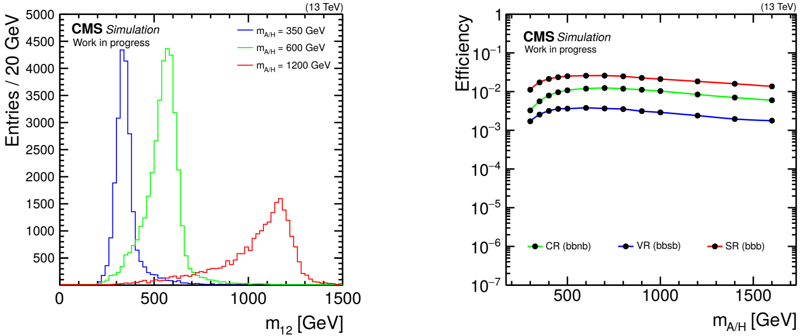
b-tagging: DeepJet algorithm

- Medium working point (wp): Mis-identification probability of light-flavor jets $\sim 1\%$
- Loose wp: mis-identification probability $\sim 10\%$



Signal Model

2017 NLO signal MC

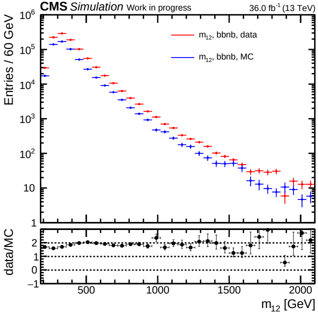


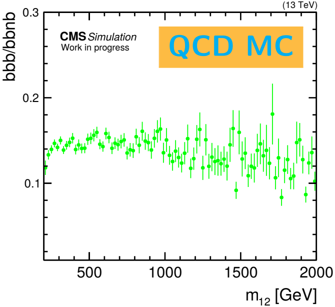
- Signal shapes modeled by NLO MadGraph5 MC
- 13 mass points from 300 to 1600 GeV
- Signal efficiency (SR) larger by factor ~ 2 than in previous analysis
  - b-tagging improved: new pixel detector and new b-tagging algorithm
- Strong signal depletion in both CR and VR

Background Model

Strategy

- Data-driven background model used for reliable description
  - QCD three jet topology challenging to model in MC
- m_{12} shape in CR and SR very similar $\rightarrow$ model small differences with slowly varying transfer factor $TF(m_{12})$
- Simultaneous fit of
  - Background parametrization CR, $B_{CR}(m_{12})$ (relatively complex, orders of magnitude)
  - SR: $B_{CR}(m_{12}) \times TF(m_{12}) + S(m_{12})$ (S: signal template)
- Uncertainties of CR parameters are rigorously taken into account

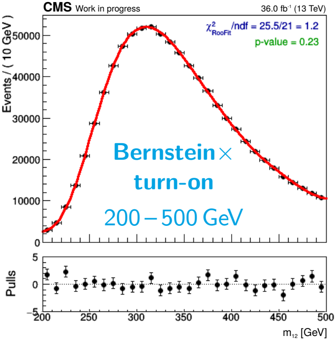




Background Model

Data CR

Data Control Region



Gaussian error function:

$$\mathrm{erf}(\mathrm{m}_{12}) = \frac{2}{\sqrt{\pi}} \int_0^{\mathrm{m}_{12}} \mathrm{e}^{(-t^2)} dt$$

Bernstein polynomial:

$$\mathrm{B}_{i,n}(x) = \sum_{i=0}^{n} c_i \cdot \binom{n}{i} \cdot x^i \cdot (1-x)^{n-i}$$

Used $n = 6$

- To facilitate parametrization in CR: four overlapping fit ranges (FR)
- Turn-on to account for kinematic thresholds: Gaussian error function
- Exploiting commonly used functions
- Very good modeling

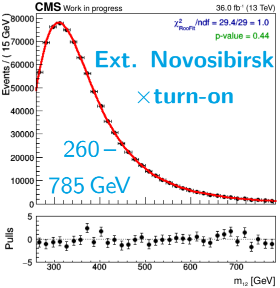
Data CR

Extended Novosibirsk function:

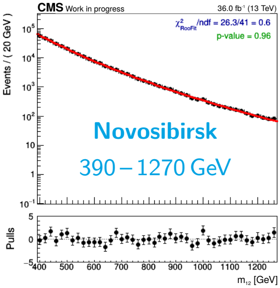
$$\mathrm{EN}(\mathrm{m}_{12}) = N \exp\left(-\frac{1}{2w_0^2}\cdot \ln^2\left(1 - \frac{(\mathrm{m}_{12}-\rho)\cdot\eta}{\nu} - \left(p_1\cdot(\mathrm{m}_{12}-\rho)^2\cdot\frac{\eta}{\nu}\right)\right) - \frac{w_0^2}{2}\right)$$

$w_0 \sim \sinh^{-1}(\eta)$, η: tail, ρ: peak, ν: width

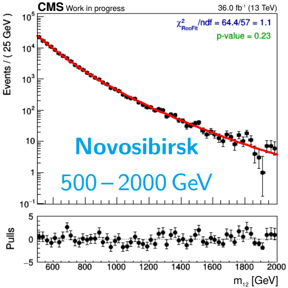
$p_1 \neq 0$



$p_1 = 0$



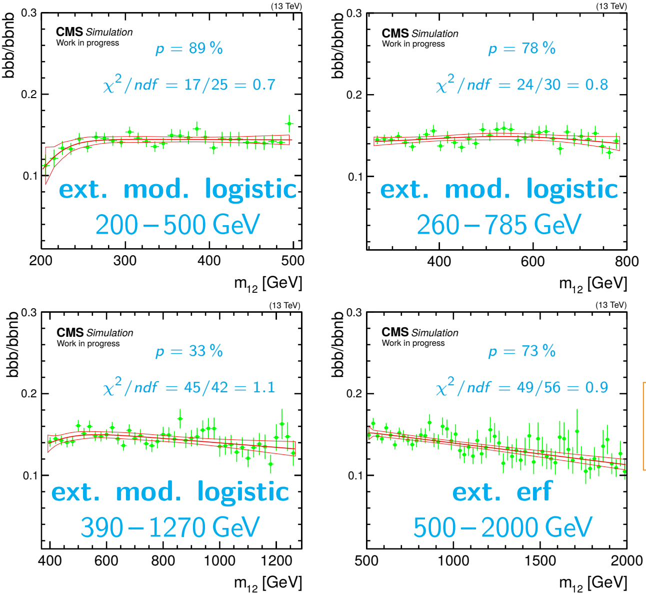
$p_1 = 0$



Background Model

Transfer Factor Parametrization

QCD MC



- Extended modified logistic function in detail on next slide
- FR 4: Extended error function sufficient
  - No significant change of slope observed

Ext. erf:
$f(m_{12}) = A\,\mathrm{erf}(C[m_{12} - B]) \cdot (1 - s \cdot m_{12})$

Background Model

Transfer Factor Requirements

TF = SR/CR
SR = bbb
CR = bbnb

- Address small shape differences between SR and CR m_{12} distribution by a **transfer factor**
  - Slowly varying function, relatively flat in most of the mass range

Functions needs

- Turn-on behavior (low masses) $\rightarrow$ sigmoid function
- Adapt to slight linear decrease (high masses)

$\Rightarrow$ Rely on modification of logistic function

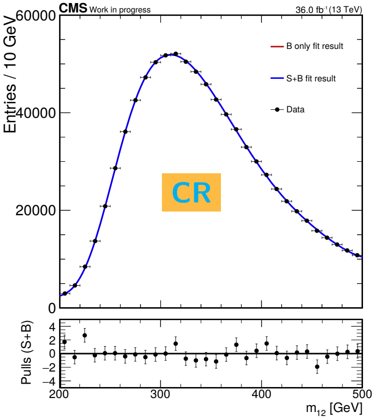
$$f(m_{12}) = \frac{1 + \alpha \exp[-k(m_{12} - x_0)]}{1 + \exp[-k(m_{12} - x_0)]}$$

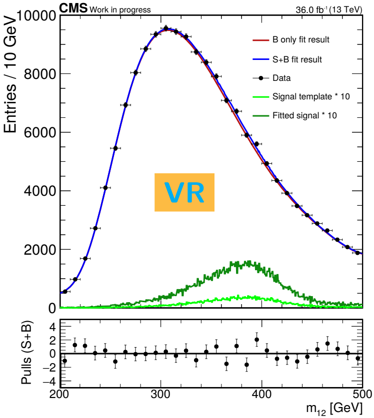
$\alpha = 0$: unmodified logistic

Extension: multiply with $(1 - \mathit{slope} \cdot m_{12})$

Validation of Signal Extraction Procedure

FR 1, 400 GeV





- VR similarities to SR make it suitable to validate approach
- Both regions described very well by simultaneous fit

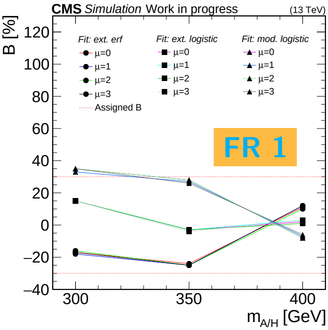
Systematic Uncertainties (Background)

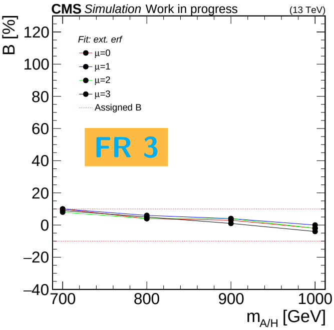
Bias Study

- Target: Evaluate uncertainty due to choice of TF parametrization
- Bias calculated in toy studies with alternative TF functions and various injected signal strengths

$$B = \frac{\hat{\mu}_{inj} - \hat{\mu}_{fit}}{0.5 \cdot (\hat{\sigma}^{+}_{fit} + \hat{\sigma}^{-}_{fit})}$$

- Mean value of all toys μ_G is determined bias
- Bias significantly reduced with respect to previous analyses due to constraints from CR in simultaneous fit



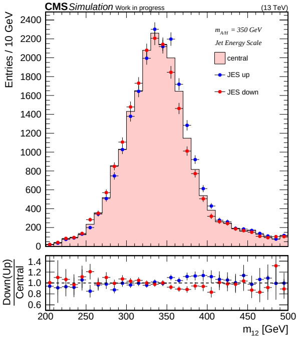


Bias		
FR	2016	2017
1	100 %	30 %
3	25 %	10 %

Systematic Uncertainties (Signal)

Signal Shape and Normalization Uncertainties

JES



- JES/JER impact shape of signal template
- Other uncertainties affect signal yield: **bold = dominant** contribution

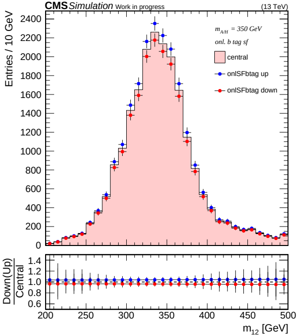
$1 - 2\,\%$	Kinematic trigger turn-on scale factor
$\sim 0.1\,\%$	Pileup reweighting
$4 - 10\,\%$	**Online b-tag** scale factor
$9 - 13\,\%$	**Offline b-tag** scale factor
$2.3\,\%$	Luminosity
$5 - 15\,\%$	In case of model interpretation: **Theory** cross-section predictions

- pdf
- α_S
- Normalization scale
- Factorization scale

Systematic Uncertainties (Signal)

Signal Shape and Normalization Uncertainties

Onl. b-tag sf

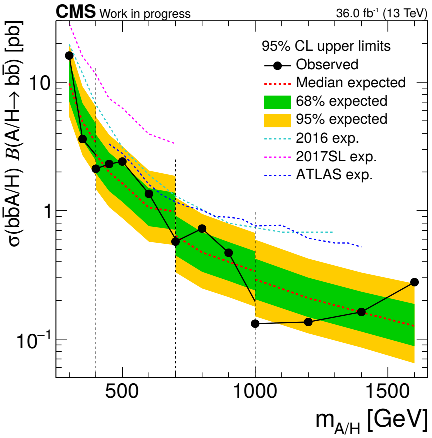


- JES/JER impact shape of signal template
- Other uncertainties affect signal yield:
  bold = dominant contribution

- $1-2\,\%$ — Kinematic trigger turn-on scale factor
- $\sim 0.1\,\%$ — Pileup reweighting
- $4-10\,\%$ — **Online b-tag** scale factor
- $9-13\,\%$ — **Offline b-tag** scale factor
- $2.3\,\%$ — Luminosity
- $5-15\,\%$ — In case of model interpretation: **Theory** cross-section predictions
  - pdf
  - α_S
  - Normalization scale
  - Factorization scale

Results

Cross-Section Limits



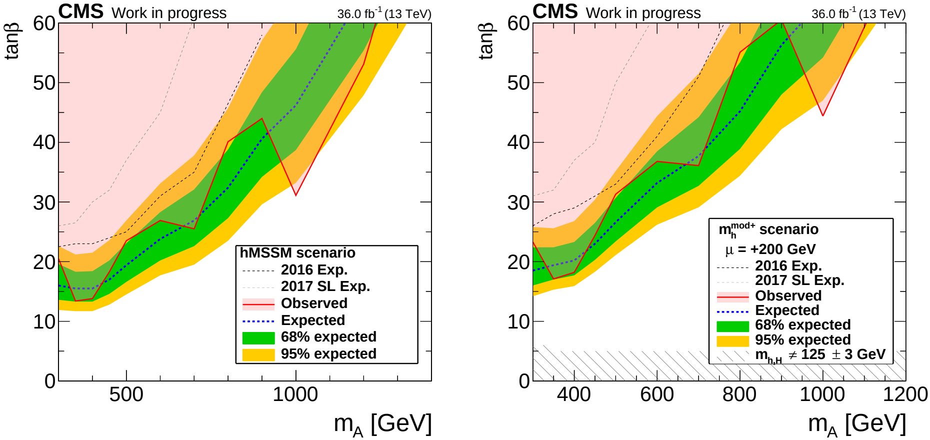
- No excess above 2σ observed
- Extended mass range and improved sensitivity by a factor of ~ 2 with respect to previous analyses
  - CMS 2016, $35.7\,\mathrm{fb}^{-1}$
  - CMS 2017 (semi leptonic), $36.5\,\mathrm{fb}^{-1}$
  - ATLAS 2015 + 2016, $27.8\,\mathrm{fb}^{-1}$
- Profiting from improved b-tagging and background model

References:
CMS 2016: JHEP08(2018)113
CMS 2017: DESY-THESIS-2020-012
ATLAS 15/16: Phys. Rev. D 102, 032004

Results

MSSM

Model Interpretation: Traditional MSSM Benchmark Scenarios

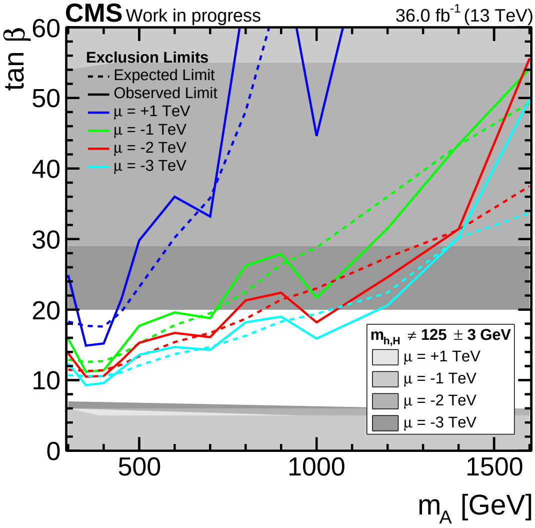


- Models analyzed for comparison with previous analyses
    - Significant improvement in sensitivity
- Limits range down to $\tan\beta = 14$

Results

MSSM

Model Interpretation: Most Recent MSSM Benchmark Scenarios



- μ = higgsino mass parameter
- Particularly strong gain of sensitivity for $A/H \rightarrow b\bar{b}$ with negative μ parameter
  - Results for m_h^{125} and m_h^{mod+} similar
- Observed limits reach down to below $\tan\beta = 10$

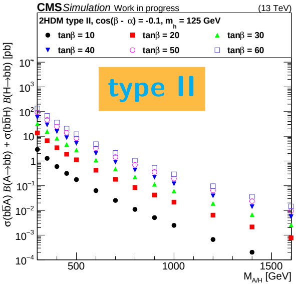
Results

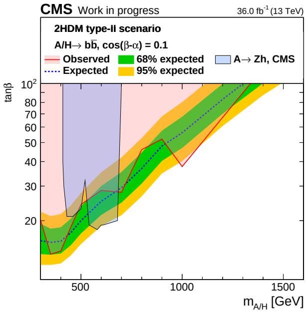
Model Interpretation: 2HDM Scenarios

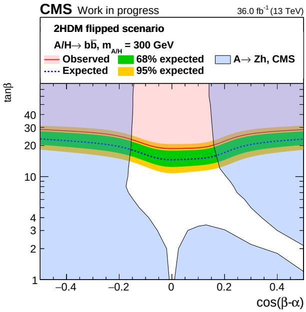
type II

$A \to Zh$ reference: EPJC 79, 564 (2019)

flipped







- Studied parameter space ($\cos(\beta - \alpha)$, $\tan\beta$, $m_{A/H}$)
- 2HDM **type II** scenario at large $\tan\beta$:
  Enhanced Yukawa coupling to down-type quarks and leptons
- 2HDM **flipped** model:
  - Only $A \to Zh$ and $H \to hh$ as well as $A/H \to b\bar{b}$ sensitive
  - Around $\cos(\beta - \alpha) = 0$: Unique sensitivity of $A/H \to b\bar{b}$

Summary and Outlook

- Many BSM theories comprise extended Higgs sector
  - Searches for additional Higgs bosons crucial, b-associated production and decay into b quarks promising as coupling can be enhanced
- Analyzed 2017 data in the fully hadronic channel
- No signal observed (no excess above $2\,\sigma$)
- Limits significantly improved with respect to previous analyses
  - New background model and signal extraction method implemented
  - Successfully validated in dedicated data region
- Interpretation in terms of MSSM and 2HDM scenarios
  - Strong increase of sensitivity for $\mathrm{A/H} \rightarrow \mathrm{b\bar{b}}$ channel in negative μ MSSM scenarios
  - Unique sensitivity of $\mathrm{A/H} \rightarrow \mathrm{b\bar{b}}$ channel around alignment limit in flipped 2HDM
- Results published in DESY-THESIS-2021-012
- **Plan**: Re-analysis of 2017 (SL/FH) with UL; analysis of 2018 data; combination of Run 2 results