Higgs boson – fermionic decays

Christina Reissel

on behalf of the CMS collaboration



Higgs Hunting, September 12th 2022



Introduction

- Higgs boson coupling to fermions proportional to fermion's mass in the Standard Model (Yukawa coupling)
- unresolved questions, e.g. CP properties of Yukawa interactions





- most sensitive Higgs production channel for $H \rightarrow b\overline{b}$ (suppression of multi-jet background)
- event selection targets 3 control regions enriched in the main backgrounds
 V+light flavour jets (LF), V+heavy flavour jets (HF), tt
- improvements:
 - categories for Higgs boson decays reconstructed as a single merged jet (boosted topology)
 - optimization of overlap between boosted and resolved topology
 - V+jets modelling (NLO, flavour treatment)
 - interpretation of results in STXS framework
- simultaneous template fit using **MVA output scores** in the signal regions and V+HF dominated control regions
- V+HF control region: multiclassifier for main backgrounds (V+bb, V+c, V+LF, single top, tt)



ETH zürich



$\overrightarrow{\text{UPDATE}} VH \rightarrow b\overline{b}$

• inclusive signal strength:

 $\mu = 0.58^{+0.19}_{-0.18}$ with observed (expected) significance of 3.3 σ (5.2 σ)

- background scale factors extracted insitu from simultaneous fit of signal and control regions
- dominant uncertainties: MC statistics and signal modelling theory uncertainties
- cross check analysis:
 - direct visualization of excess using dijet invariant mass: $\mu = 0.34 \pm 0.34$ (mass-decorrelated DNN for event categorization)
 - VZ analysis: $\mu = 1.16 \pm 0.13$



ETH zürich



CMS-HIG-20-001



- STXS is a combination of fully fiducial cross sections and direct fits
- reduction of impact of theoretical uncertainties $ZH, 150 < p_T^{V \le 250 \text{ GeV}, \ge 1J}$ on Higgs boson measurements $ZH, 150 < p_T^{V \le 250 \text{ GeV}, \ge 0J}$





$VH \rightarrow CC$

CMS-HIG-21-008





- difficulty of **charm jet identification** in high jet environment at LHC JINST 17 P03014
- event categorization: exploration of merged-jet ($p_{\tau} > 300 \text{ GeV}$) and resolved-jet topologies
- discriminant: •

mass of the Higgs boson candidate (merged-jet topology) and BDT output score (resolved-jet topology)

- dominant **uncertainties**: statistical, limited MC statistics, c jet identification efficiencies
- cross check: $VZ \rightarrow c\overline{c}$ leading to first observation at a hadron collider
- observed (expected) **upper limit** on $\sigma(VH) \times BR(H \rightarrow cc)$ • 0.94 (0.50^{+0.22}_{-0.15}) pb at 95% CL corresponding to 14 (7.6^{+3.4})xSM
- most stringent constraint on Higgs-charm Yukawa coupling modifier

ETH zürich

2L

$ggH \to c\overline{c}$

- Higgs production with p_T (Higgs) > 450 GeV in inclusive Higgs production mode
- similar analysis strategy to ggF H \rightarrow bb analysis JHEP12 (2020) 085
- orthogonal constraint on the charm coupling w.r.t. $VH \rightarrow c\overline{c}$ analysis
- exploration of single large radius jets tagged by the DeepDoubleX tagger CMS-DP-2018-046
- DDCvL used to define signal and control regions event categorisation: p_{T} , DDCvL pass/fail
- dominant uncertainties: QCD pass-fail ratio, statistical
- **first observation** of $Z \rightarrow c\overline{c}$ at a hadron collider in Z+jets production mode and boosted topology
- upper limit 45(38)xSM at 95% CL



ETH zürich

arXiv:2204.12957



- $H \rightarrow TT$
- relative large branching ratio (~6%) and reduced background • compared to $H \rightarrow b\overline{b}$
- two analysis focussing on ggF and VBF production: • cut based (CB) and neural network based (NN)
- event selection based on number of e, μ and $\tau_{_{\rm h}}$ candidates
- event categorization: • five classes (CB) and multiclassification output (NN)
- major backgrounds:
 - $Z \rightarrow \tau \overline{\tau}$ (estimated with embedding technique) JINST 14 (2019)
 - jets misidentified as τ_h, e, μ (estimated with fake factor technique)

discriminant:

CB: m_{π} vs. $p_{\tau}(\tau) / m_{\mu} / p_{\tau}(H)$ NN: classifier output in each node

30 – 40% stronger constraints with NN based analysis



ETH zürich



Observed: CB-analysis

Observed: NN-analysis



combination with analysis focusing on Higgs production in association with W/Z

 $H \rightarrow \tau \tau$

- purely leptonic τ pair decays not considered (small branching fraction, overlap with H → WW)
- **2D discriminant**: m_{π} vs. $p_{T}(V)$
- together with VBF and ggH analysis: 16 kinematic regions tested (STXS stage 1.2)







$H \rightarrow \tau \overline{\tau}$ (differential)

- extension of the H $\rightarrow \tau \bar{\tau}$ analysis to a differential cross section measurement in terms of p_{τ} (Higgs), jet multiplicity and p_{τ} (leading jet)
- first fiducial differential measurements in final states with τ 's
- significant improvement over other final states in phase space with large jet multiplicity or high $p_{_{T}}(\text{Higgs})$
- signal extracted using **2D-distributions** of m_{π} vs. differential variable



 $H \rightarrow \mu \mu$

JHEP01 (2021) 148



- first evidence for decays of the Higgs boson to 2nd gen fermions with observed significance of 3.0σ
- most precise measurement of the Higgs boson couplings to µ's reported to date
- four exclusive categories targeting the different production channels of the Higgs boson
- backgrounds: estimated from data with discrete profiling method JINST 10 (2015)

category	event categorization	discriminant
VBF	m(µµ)	DNN output score
ggH, tṫH and VH	BDTs	m(µµ)

measurement statistically dominated



-2

0

2

ETH zürich

Christina Reissel

ETH zürich



- analysis targets Higgs boson production via ggF and VBF
- BDT classifier used to define analysis categories
- simultaneous fit to m(e⁺e⁻) distribution
 - \rightarrow backgrounds modelled using discrete profiling method JINST 10 (2015)
- most stringent upper limit up to date: 3.0x10⁻⁴ (3.0x10⁻⁴) at 95 % CL







Summary & Outlook

- measurements of the Higgs-fermion coupling **crucial tests** of the SM
- approaching the era of precision measurements for $H \rightarrow b\overline{b}$ and $H \rightarrow \tau \overline{\tau}$
- major updates with full Run II datasets
 - improvements due to larger statistics and improved analysis strategies
 - new interpretations (STXS, effective couplings)
- fermions used in BSM searches, searches for double-Higgs production and for anomalous coupling measurements in the Higgs boson production

PDATE new result: $VH \rightarrow b\overline{b}$

- $H \rightarrow b\overline{b}$ drives uncertainty of the total Higgs boson width
- update in the most sensitive channel for measuring $H \rightarrow b\overline{b}$ decays!
- stay tuned! still outstanding Run II results & improvements with larger Run III dataset
- possible thanks to the fantastic performance of LHC and CMS



ETH zürich

Backup





CMS-HIG-20-001

b jet energy regression Comput Softw Big Sci 4, 10 (2020)

- DNN regression on b jet specific kinematic properties for jet momentum reconstruction
- **dedicated smearing and scaling** correction to account for different jet resolution in data and MC
 - fit uses 2-jet event topology in which the jet resolution can be measured by the jet system balance against the Z in the transverse plane

Kinematic fit (2 lepton channel)



- possible due to better momentum resolution of leptons than jets and absence of intrinsic MET in 2 lepton channel
- fit of lepton and jet kinematics takes into account uncertainties





ETH zürich





- boosted measurement employs a single fat jet (reconstructed with anti- k_{t} and $\Delta R = 0.8$)
- soft-drop corrections applied to jet mass
- double b-tagger algorithm for selection and variable in multivariate discriminant used for signal extraction
- double b-tagger: DeepAK8
 - significant improvement with respect to double-B algorithm (usage of a DNN and low-level information)
 - output decorrelated from the soft-drop mass
 - multiple output classes: analysis uses bb vs. light output node
 - DeepAK8 SF's extracted from $g \rightarrow b\overline{b}$ topology (extrapolated to $H \rightarrow b\overline{b}$ phase-space) on the signal and freely-floating in-situ rate parameters for background DeepAK8 SF's



ETH zürich





Event Selection (resolved topology)







Event Selection (boosted topology)



- overlap treatment with resolved analysis optimized by maximizing Asimov sensitivity
- all events assigned to resolved analysis, unless event categorized as resolved CR and boosted SR



ETH zürich



$VH \to b\overline{b}$

	SR	t ī CR	V+LF CR	V+HF CR
0-lepton, resolved	DNN	$p_{\mathrm{T}}(\mathrm{V})$	$p_{\mathrm{T}}(\mathrm{V})$	HFDNN
0-lepton, boosted	BDT	DeepAK8bbVsLight	DeepAK8bbVsLight	DeepAK8bbVsLight
1-lepton, resolved	DNN	$p_{\mathrm{T}}(\mathrm{V})$	$p_{\mathrm{T}}(\mathrm{V})$	HFDNN
1-lepton, boosted	BDT	DeepAK8bbVsLight	DeepAK8bbVsLight	DeepAK8bbVsLight
2-lepton, resolved	DNN	$p_{\mathrm{T}}(\mathrm{V})$	$p_{\mathrm{T}}(\mathrm{V})$	DeepCSV scores
2-lepton, boosted	BDT	DeepAK8bbVsLight	DeepAK8bbVsLight	DeepAK8bbVsLight



0-lepton

1-lepton

2-lepton





ETH zürich







previous result (partial Run II result):

- expected sensitivity 4.2σ
- µ = 1.06 ± 0.26 (2016+2017)
- µ = 1.08 ± 0.34 (2017)

CMS-PAS-HIG-18-016



- Jackknife resampling is a non-parametric method of estimating uncertainty on a parameter by removing partitions from total event dataset.
- we divide the 2017 dataset combined from 2 analyses into g equal-sized orthogonal partitions
- for each partition *i*:
 - Remove that set of events from each analyses datacards
 - Redo both the fits to get μ_i
- jackknife estimate of the variance on $\Delta \mu$ is calculated from the variance of $\Delta \mu_i$:

$$var_{I}(\Delta \mu) = \frac{g-1}{g} \sum (\Delta \mu_{(i)} - \overline{\Delta \mu_{d}})^{2} = \frac{(g-1)^{2}}{g} var(\Delta \mu_{(i)})$$

• Compute disagreement on $\Delta \mu$ ($\sigma_{_{\Delta \mu}}$)

$$\sigma_{\Delta\mu} = \frac{\overline{\Delta\mu}}{\sqrt{var(\Delta\mu)}} = \frac{\overline{\Delta\mu}}{\frac{(g-1)}{\sqrt{g}} \times std. \, dev_{\Delta\mu}}$$

$$\sigma_{\Delta\mu} = 2\sigma$$
 correlation $\rho \sim 0.5$

ETHzürich



$t\bar{t}H \rightarrow b\bar{b}$ (partial Run II) CMS-PAS-18-030

- events are classified according to number of jets, number of b tagged jets and output node of ANN (artificial neural network)
- inputs to ANN: kinematics of leptons, jets and missing transverse energy, but also highlevel variables (e.g. event shape or Matrix-Element Method discriminant)
- ANN discriminant output used also as final discriminant in fit
- evidence for $t\bar{t}H(b\bar{b})$ found with observed (expected) significance of 3.9 σ (3.5 σ) using data collected in 2016/2017
- signal strength found to be $\hat{\mu} = 1.15 \, {}^{+0.15}_{-0.15}(\text{stat}) \, {}^{+0.28}_{-0.25}(\text{syst})$





ggF $H \rightarrow b\overline{b}$ (boosted)

- inclusive Higgs production with p_{τ} (Higgs) > 450 GeV
- Higgs bosons reconstructed as single large-R jet with two prong structure
- application of dedicated b-tagging algorithms to target boosted topologies (DDBT) CMS-DP-2018-046
- major **backgrounds**:

QCD/tt (estimated in control regions) W/Z+jets resonance (used to constrain uncertainties)

- event categorisation: p_{T} , DDBT pass/fail
- improvements: ggF p_τ(Higgs) modelling, improved b tagging algorithm for large-R jets
- dominant uncertainties: statistical, QCD pass-fail ratio
- observed (expected) significance: 2.5σ (0.7 σ) with 1.9 σ above SM expectation
- differential measurement interpretation in terms of p_{τ} (Higgs) in the boosted H \rightarrow bb region



5

10

 $\mu_{\rm L} = 3.7^{+2.7}_{-2.6}$

 $[500, 550] GeV \\ \mu_{H} = -3.6^{+2.6}_{-2.8} \\ [450, 500] GeV \\ \mu_{u} = -0.5^{+2.7}_{-2.7} \\$

-10

--5

n

JHEP12 (2020) 085



Christina Reissel

15

 μ_{μ}

ggF $H \rightarrow b\overline{b}$ (boosted)





JHEP 2012 (2020) 085





 $\frac{\tau_{\rm h} \tau_{\rm h}}{\text{genuine } \tau}$

jet $\rightarrow \tau_h$

misc

misc

misc

Observables

 $m_{\tau\tau}$ (e μ)

 $m_{\tau\tau}, m_{ii}$

 $m_{\tau\tau}, m_{ii}$

 $m_{\tau\tau}, \hat{p}_{T}^{H}$

 $m_{ au au}, \hat{p}_{\mathrm{T}}^{\mathrm{\hat{H}}}$

 $m_{\tau\tau}, m_{\parallel}$

 $m_{\tau\tau}, m_{ii}$

 $m_{\tau\tau}, \hat{p}_{\tau}^{\rm E}$

 $m_{\tau\tau}, \hat{p}$

 $m_{\tau\tau}$

 $m_{\tau\tau}, p_{\rm T}^{\tau_{\rm h}} (\ell \tau_{\rm h})$

 $\mu \tau_{\rm h}$

jet $\rightarrow \tau_{\rm h}$

tt

zll

misc



ETH zürich

arXiv:2204.12957

 $H \rightarrow \tau \tau$





ETH zurich



- $H \rightarrow \tau \tau$
- comparison CB and NN based analysis for stage-0 STXS:
 - compatible constrains on $\mu_{\mbox{\tiny incl}}$ and $\mu_{\mbox{\tiny ggH}}$:
 - inclusive result driven by μ_{ggH} which is dominated by systematic uncertainties \rightarrow not taken into account during the training process of the NN
 - separation very difficult between ggH vs. $qqH/Z \rightarrow \tau \overline{\tau}$
- comparison CB and NN based analysis for stage-1.2 STXS:
 - systematic variations smaller in NN-based approach

 \rightarrow NN uses 14D input space, so systematic variations of single features impact result less

- NN includes discriminating distributions for background

 → high purity control regions help to constrain nuisance parameters
- larger samples used for training of NN approach
- STXS bins in CB approach based on event selection, NN targets best possible separation with multiclassification approach



cleaned event.

ETH zürich



$H \rightarrow \tau \tau \tau (CP)$

- first measurement of the CP structure of the Higgs boson coupling to τ leptons
- multi-class DNN/BDT discriminant (categories: Higgs, two genuine τ leptons, hadronic jet misidentified as τ_h)
- extraction of results with **2D fit** to maximum multiclass output score and angular correlation between decay planes of τ 's
- measurement dominanted by statistical uncertainty
- disfavour pure CP-odd scenario at 3σ \rightarrow compatible with SM Higgs boson



ETH zürich

Christina Reissel

-2

-1

31

Reconstruction of CP-sensitive observables:

- impact parameter method
- neutral-pion method
- combined method ٠
- polarimetric vector method



99.7% CL -

- 20

15 Jol 02-10 Log τ

5

0

0

2

 κ_{τ}

90

45 $\alpha^{H\tau\tau}(degrees)$





 $H \rightarrow \tau \tau (CP)$





0

1

0

SM

-45



CMS

Post-fit

VBF category

m_H = 125.38 GeV

S/(S+B) Weighted Events / GeV

Data-Bkg.

110

115

120

125

130

135

140

145

m_{μμ} (GeV)

150

25

20

10

$H \to \mu \overline{\mu}$

 $\Delta \mu$

+0.44 -0.42

-0.40

-0.16

-0.11

-0.11

-0.06

+0.41

+0.17

+0.12

+0.10

+0.07

Uncertainty source

Post-fit uncertainty

Statistical uncertainty

Systematic uncertainty

Experimental uncertainty

Size of simulated samples

Theoretical uncertainty

JHEP01 (2021) 148



Coupling strength κ_{μ}



137 fb⁻¹ (13 TeV)

🗌 Н→µµ

ggH

DY

Top quark Diboson

🛉 Data

– VBF

Zjj-EW





ETH zürich

Christina Reissel



CMS-HIG-21-008



Uncertainty source	$\Delta \mu / (\Delta \mu)_{tot}$
Statistical	85%
Background normalizations	37%
Experimental	48%
Sizes of the simulated samples	37%
c jet identification efficiencies	23%
Jet energy scale and resolution	15%
Simulation modeling	11%
Integrated luminosity	6%
Lepton identification efficiencies	4%
Theory	22%
Backgrounds	17%
Signal	15%

 $1.1 < |\kappa_{\rm c}| < 5.5 \; (|\kappa_{\rm c}| < 3.4)$

ETH zürich



$ggH \rightarrow cc$

CMS-HIG-21-012



arXiv:2208.00265



ETH zürich

$H \rightarrow e^+e^-$







Combination plots

Nature 607, 60-68 (2022)







ETH zürich



Combination plots

Nature 607, 60-68 (2022)



ETH zürich



Combination plots

