Study of the boosted Higgs boson production

Jie Xiao *IP2I–Lyon*
On behalf of the CMS collaboration

Higgs 2023 conference
IHEP CAS, Beijing, China
27 Nov - 2 Dec 2023

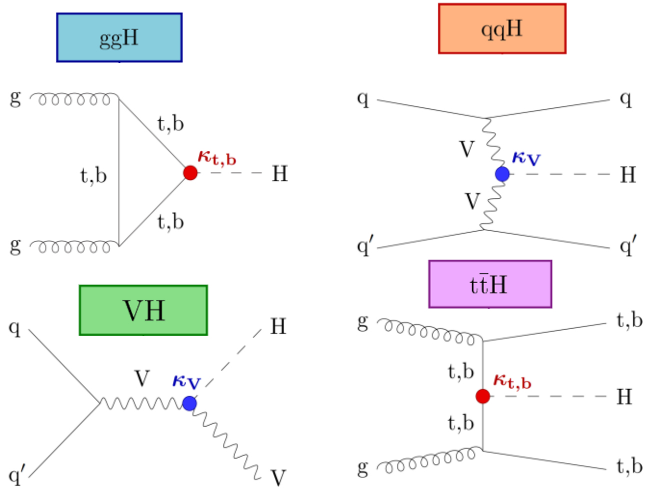
Why boosted Higgs?

Higgs boson production at **large transverse momentum (boosted Higgs)**

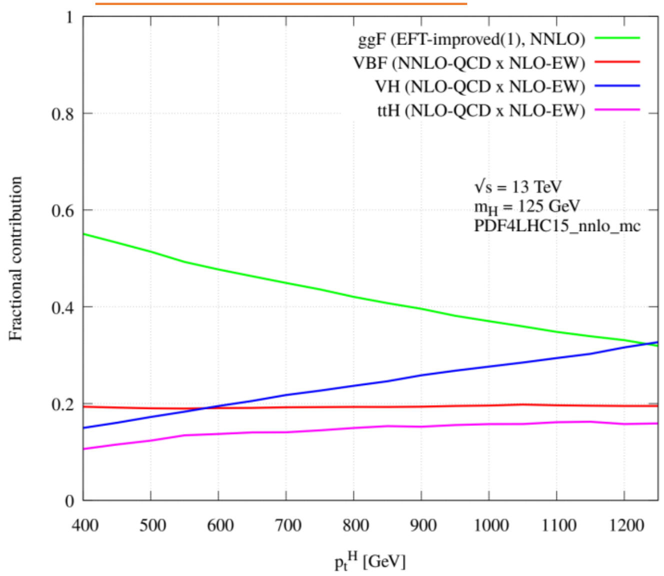
- A new frontier for the LHC Higgs physics

Measuring the properties of boosted Higgs bosons can test

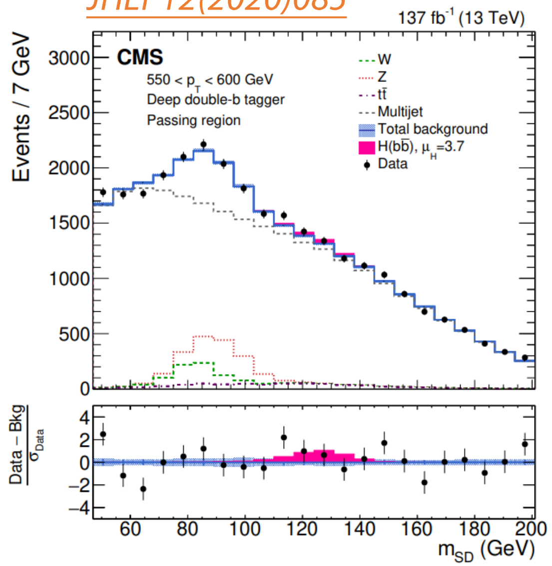
- Existence of physics **beyond the SM (BSM)** in the scalar sector
- **Higher order EW radiative corrections** in Higgs production



LHCHXSWG-2019-002



JHEP12(2020)085



Full Run2 inclusive Boosted H → bb measurement

Why boosted Higgs?

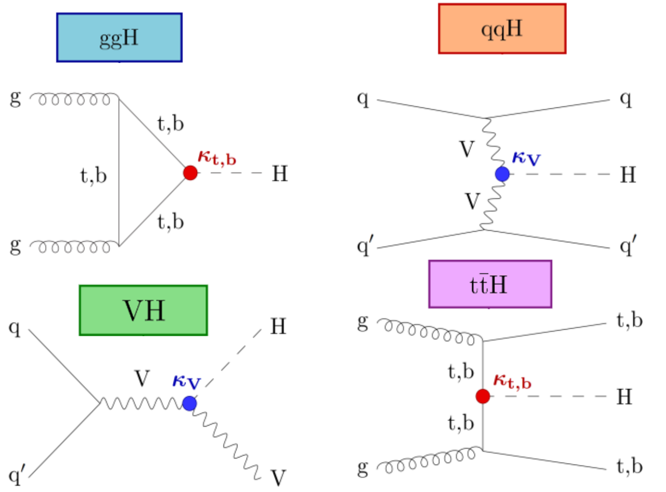
Higgs boson production at **large transverse momentum (boosted Higgs)**

- A new frontier for the LHC Higgs physics

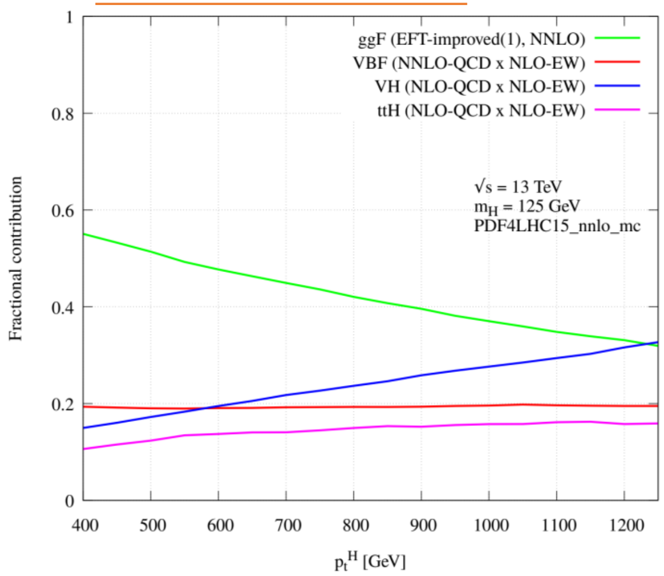
Measuring the properties of boosted Higgs bosons can test

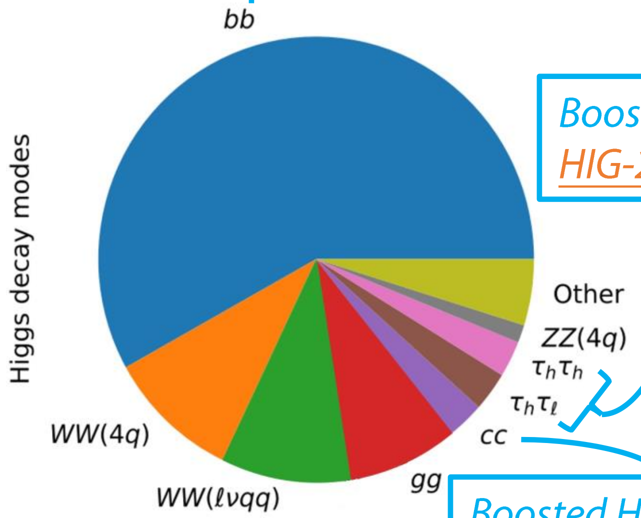
- Existence of physics **beyond the SM (BSM)** in the scalar sector
- **Higher order EW radiative corrections** in Higgs production

Focus on **recent boosted Higgs measurements** with CMS



LHCHXSWG-2019-002





Boosted $H \rightarrow bb$
HIG-21-020

Boosted $H \rightarrow \tau\tau$
HIG-21-017

Boosted $H \rightarrow cc$
PRL 131 (2023) 041801

Boosted H → bb via VBF or ggF

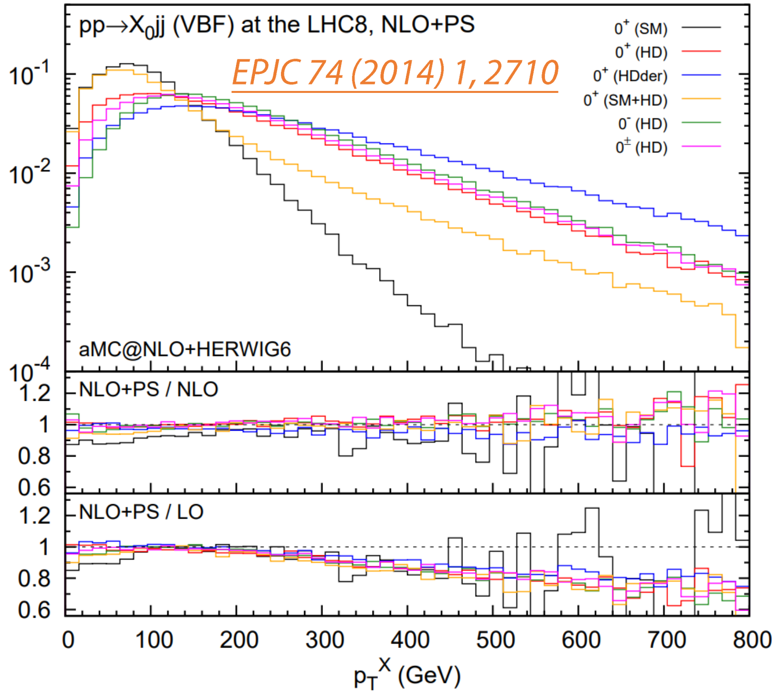
HIG-21-020

Motivation

HIG-21-020

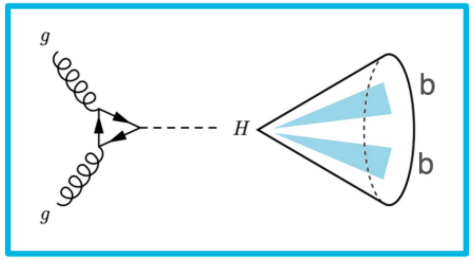
Predictions for different boosted Higgs production

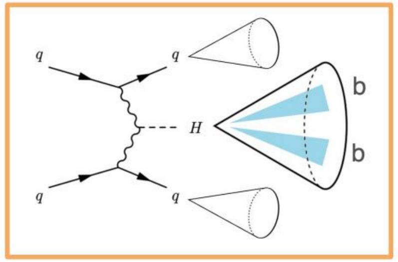
- **ggF** is still the **largest production** mode, but less dominant
- **VBF** production is the **sub-dominant**
- Probe other production modes that offer sensitivity to **different couplings**
- **Sensitive to new physics** at high energy scales



H→bb has the **highest branching ratio** (57.5%)

- **Higgs appears as a single large-R jet** in boosted regime
- Reduced background
- **Better mass resolution** compared to resolved decay
- Identifiable by **two-prong substructure**, and **Deep Double b-tag (DDB)**





Event selection

HIG-21-020

Higgs candidate: **AK8* jet with p_T > 450 GeV**, $|\eta| < 2.5$, passing tight ID & pileup ID

- Requirement on **two-prong substructure** using N_2 variable decorrelated with mass (**N_2^{DDT}**)
- Restrict jet **$\rho = \ln(m_{SD}^2/p_T^2)$** to ensure Higgs decay is merged into jet cone
- If more than one jet qualifies, select the one with **highest DDB score**

AK8*: *anti-kT algorithm with a distance parameter of 0.8*

Event selection

HIG-21-020

Higgs ca[...] < 2.5, passing tight ID & pileu[...]

Soft-drop algorithm is applied to the **jet mass (m_{SD})** to remove soft and wide-angle radiation

- Requirement on two-prong substructure using N_2 variable decorrelated with mass (**N_2^{DDT}**)
- Restrict jet **$\rho = \ln(m_{SD}^2/p_T^2)$** to ensure Higgs decay is merged into jet cone
- If more than one jet qualifies, select the one with **highest DDB score**

Substructure variable **N_2** (N_2^1) identifies **two-prong jets** by energy correlation functions

Designing decorrelated taggers (DDT) to mitigate distorting the background jet mass [1]

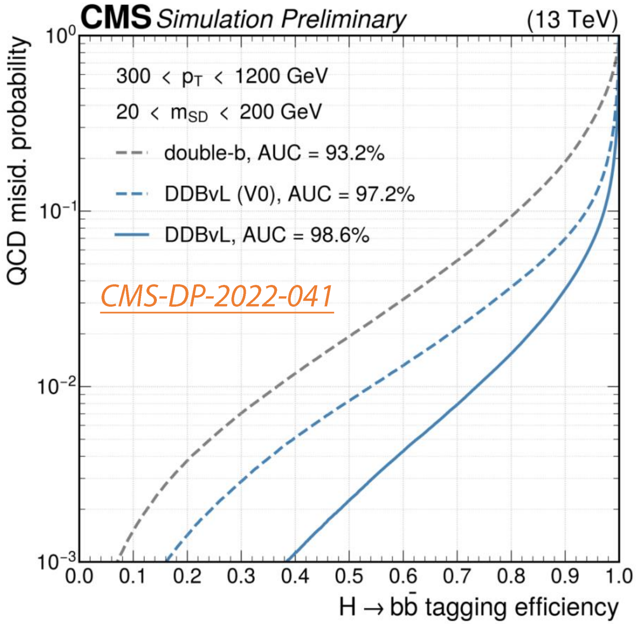
$$N_2^{DDT} \equiv N_2 - q_{0.26}(p_T, \rho) < 0$$

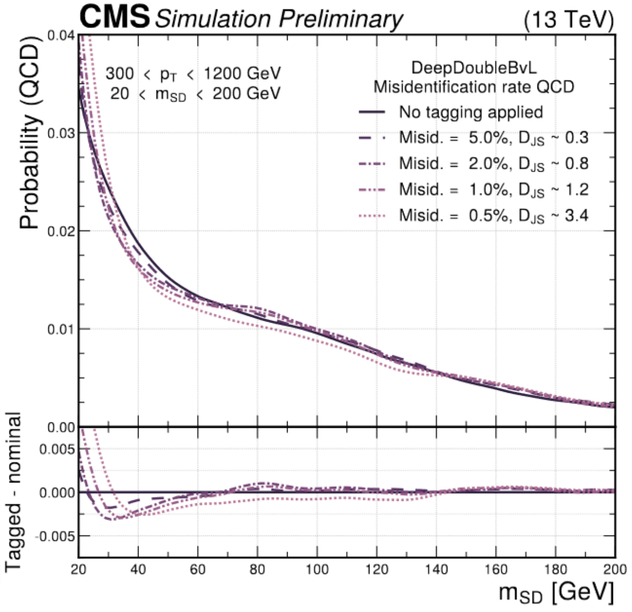
- Decorrelated from jet mass Function $q_{0.26}(p_T, \rho)$ → **26% efficiency on QCD jets**

Background **Signal**

DeepDoubleX tagger v2 for jet flavor tagging

- **DeepDoubleBvL (DDB)** - **×2 improvement** in significance compared to version I
- Trained signal MC that is **flat in m_{SD}** → reduced mass sculpting





Event selection

Higgs candidate: **AK8* jet with p_T > 450 GeV**, $|\eta| < 2.5$, passing tight ID & pileup ID

AK8*: *anti-kT algorithm with a distance parameter of 0.8*

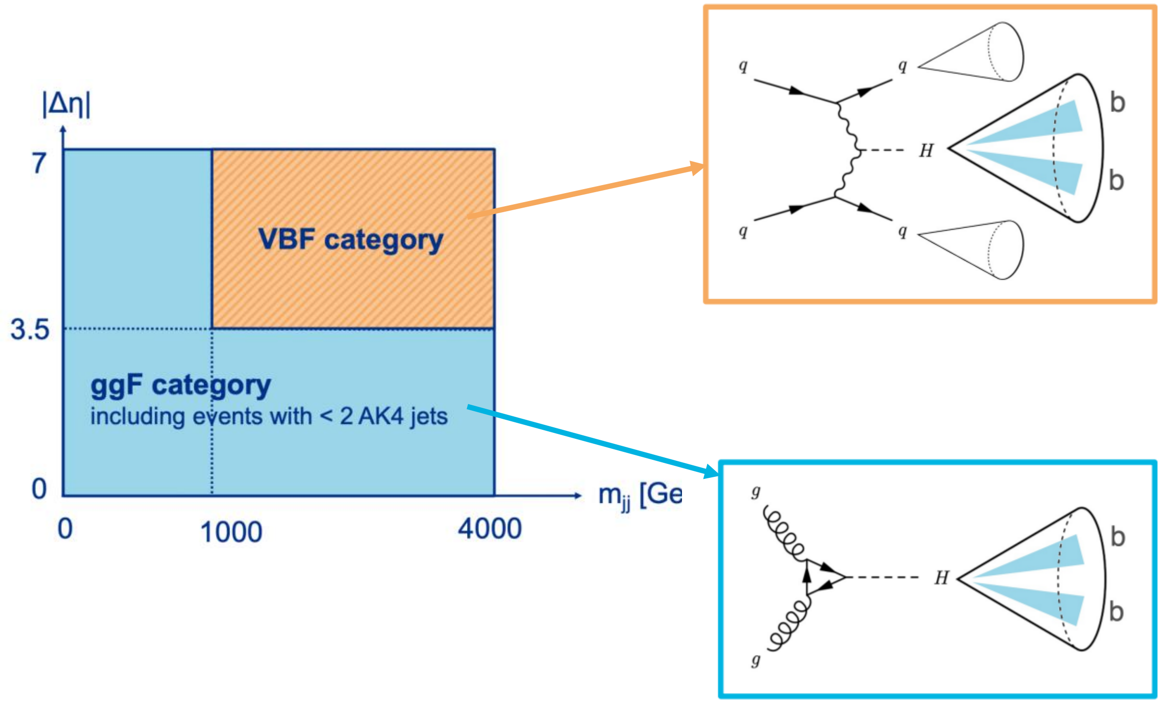
- Requirement on **two-prong substructure** using N_2 variable decorrelated with mass (**N_2^{DDT}**)
- Restrict jet **$\rho = \ln(m_{SD}^2/p_T^2)$** to ensure Higgs decay is merged into jet cone
- If more than one jet qualifies, select the one with **highest DDB score**

No leptons

Dedicated cuts to **reject ttbar**

- **p_T^{miss} < 140 GeV**, **b-veto** in hemisphere opposite H candidate

2 more AK4 jets with **$\Delta\eta_{jj}$ > 3.5 and m_{jj} > 1 TeV**
⟶ **VBF category** Else ⟶ **ggF category**



Event selection

HIG-21-020

Higgs candidate: **AK8* jet with p_T > 450 GeV**, $|\eta| < 2.5$, passing tight ID & pileup ID

AK8*: *anti-kT algorithm with a distance parameter of 0.8*

- Requirement on **two-prong substructure** using N_2 variable decorrelated with mass (**N_2^{DDT}**)
- Restrict jet **$\rho = \ln(m_{SD}^2/p_T^2)$** to ensure Higgs decay is merged into jet cone
- If more than one jet qualifies, select the one with **highest DDB score**

No leptons

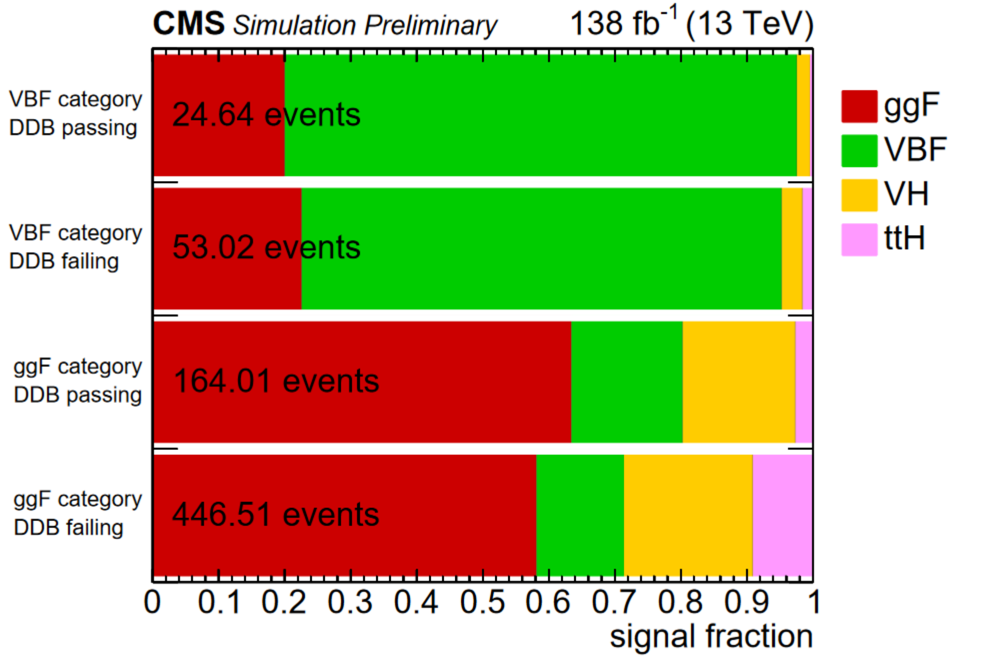
Dedicated cuts to **reject ttbar**

- **p_T^{miss} < 140 GeV**, **b-veto** in hemisphere opposite H candidate

2 more AK4 jets with **$\Delta\eta_{jj}$ > 3.5 and m_{jj} > 1 TeV**
⟶ **VBF category** Else ⟶ **ggF category**

Split into **DDB pass and fail regions**

- **DDB > 0.64** ⟶ DDB **pass** region, chosen to optimize VBF sensitivity
- **DDB < 0.64** ⟶ DDB **fail** region, used to estimate QCD background



MC relative contribution of each production mode

QCD data-driven estimation

HIG-21-020

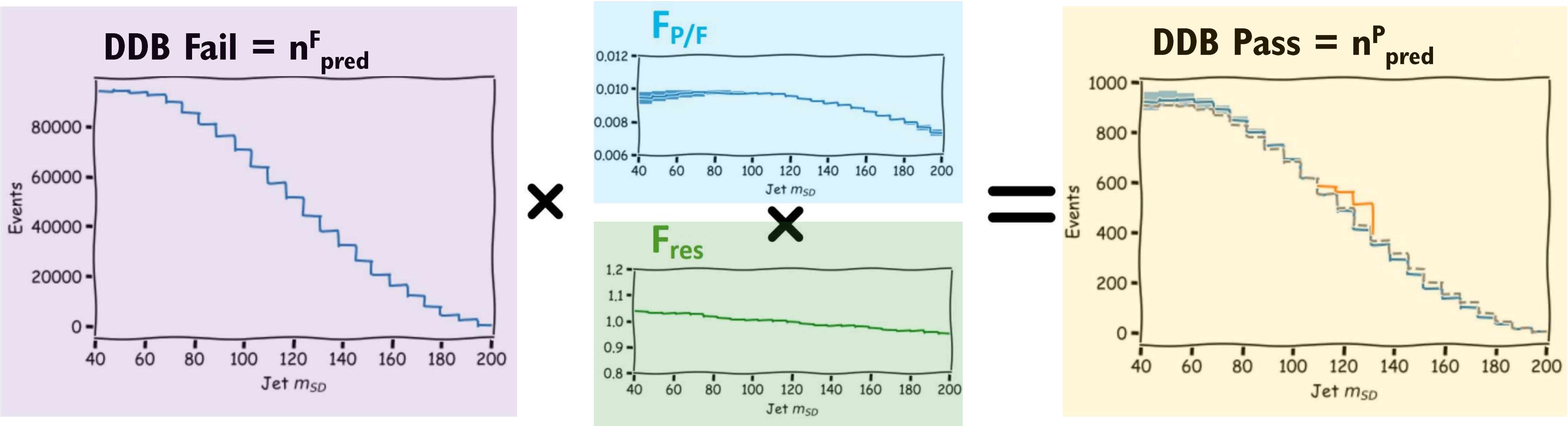
Define **two transfer factors**

- $\mathbf{F_{P/F}}$ translates QCD **from DDB failing to passing** regions by fitting on MC only
- $\mathbf{F_{res}}$ translates **from QCD MC to data** (fit in DDB pass/fail regions simultaneously)

The predicted number of QCD events in bin i in the DDB passing region

$$N_P^i = R_{P/F}^{MC} F_{P/F}^i F_{res}^i N_F^{data,i}$$

Number of total simulated QCD events in the DDB pass region divided by that in the failing region



Fit strategy and results

HIG-21-020

Fit to the observed $\mathbf{m_{SD}}$ distributions

- Both the **DDB passing and failing** regions with **bins**
  - 2 Bins of **VBF jets mass for VBF**
  - 6 bins of $\mathbf{p_T^H}$ **for ggF**
- Single muon control region – to control top background

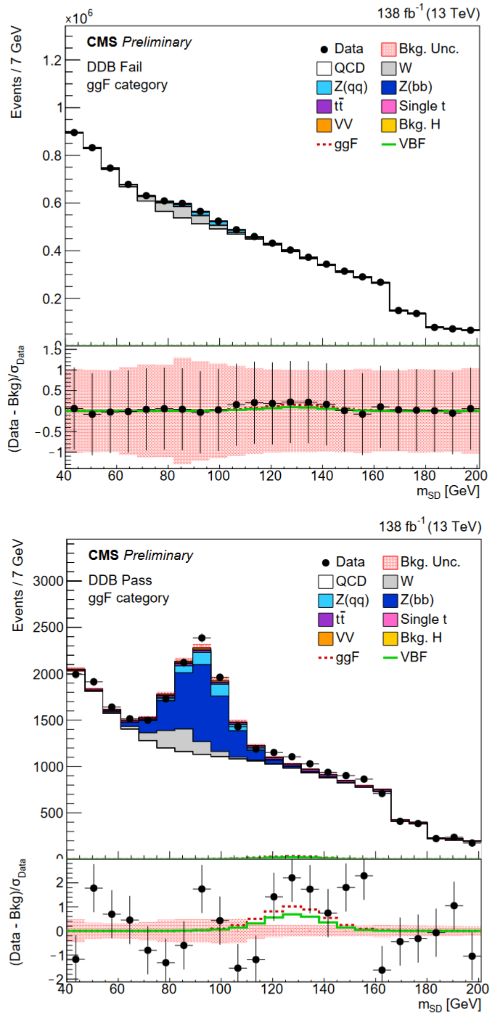
Signal strength of boosted H→bb

- VBF process: $\mu_{VBF} = 5.0^{+2.1}_{-1.8}$
- ggF process: $\mu_{ggF} = 2.1^{+1.9}_{-1.7}$

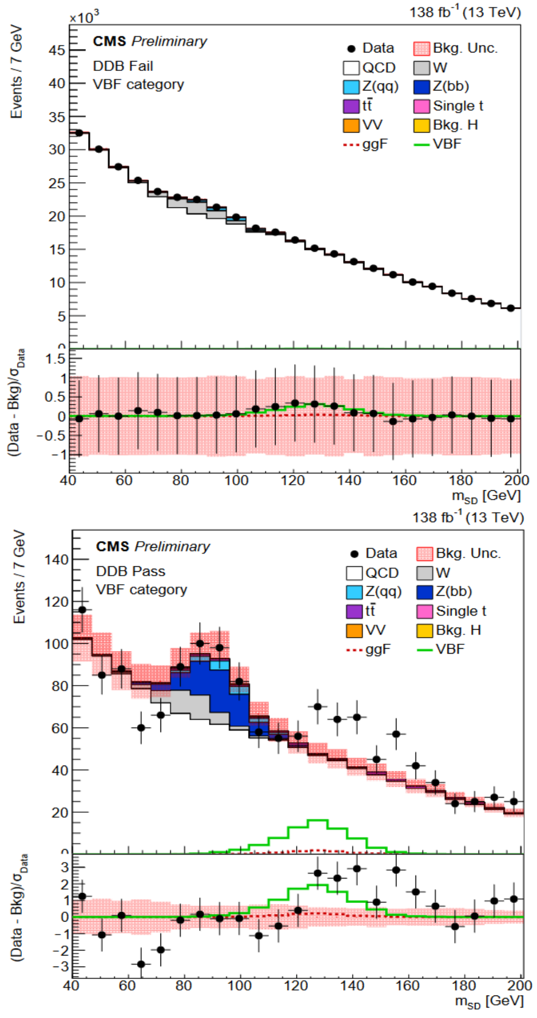
Observed (expected) **significance**

- **VBF is 3.0σ (0.9σ)**
- **ggF is 1.2σ (0.9σ)**

ggF category



VBF category



Fit strategy and results

HIG-21-020

Fit to the observed m_{SD} distributions

- Both the **DDB passing and failing** regions with **bins**
  - 2 Bins of **VBF jets mass for VBF**
  - 6 bins of **p_T^H for ggF**
- Single muon control region – to control top background

Signal strength of boosted H→bb

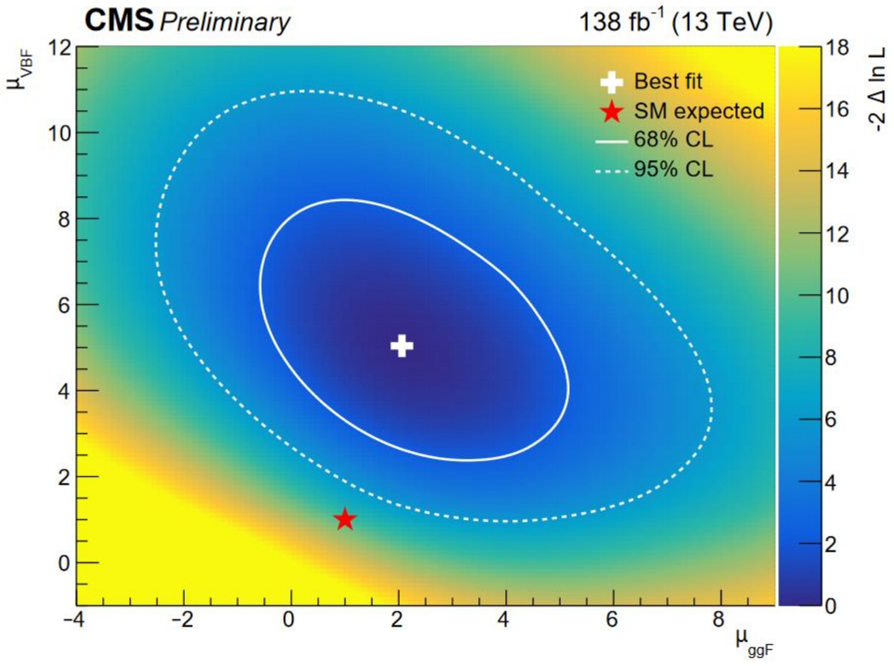
- VBF process: $\mu_{VBF} = 5.0^{+2.1}_{-1.8}$
- ggF process: $\mu_{ggF} = 2.1^{+1.9}_{-1.7}$

Observed (expected) **significance**

- **VBF is 3.0σ (0.9σ)**
- **ggF is 1.2σ (0.9σ)**

2D likelihood scan

- Best fit differs from **SM** by **2.6σ**



Two-dimensional likelihood contour of the ggF and VBF signal strengths

Boosted H → cc

PRL 131 (2023) 041801

Motivation

PRL 131 (2023) 041801

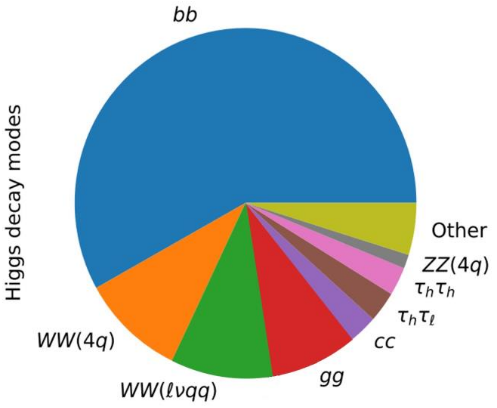
Constrain **Higgs-charm coupling**

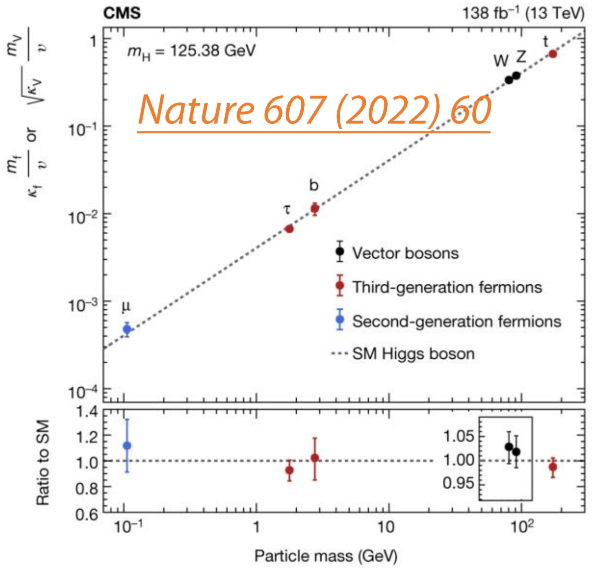
- **Largest BR without evidence**

Second-generation coupling sensitivity

- **3σ** evidence for **H(μμ)** with HIG-19-006
- Upper limit on **VH(cc)** < 14×SM with HIG-21-008

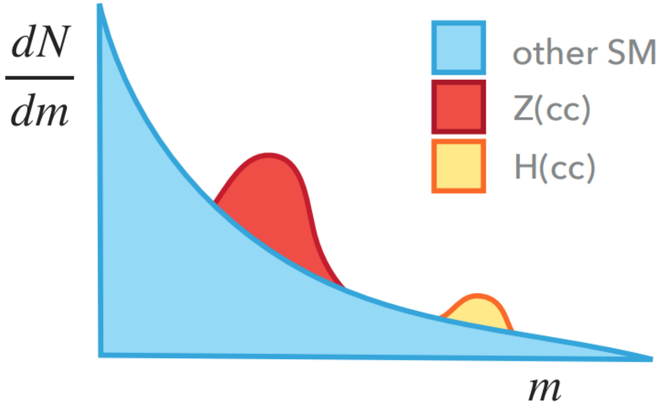
Complementary to VH(cc)





Z(cc) SM candle **observation**

- **First single-boson observation** at hadron collider
- VZ(cc) observed with HIG-21-008



Event Selection

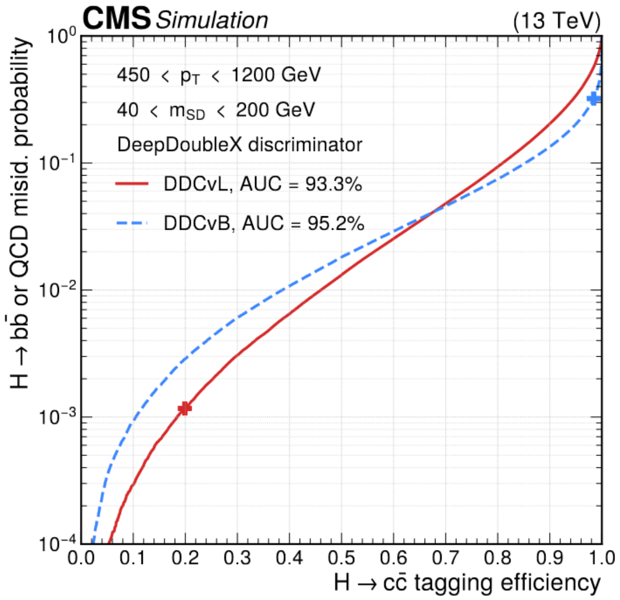
PRL 131 (2023) 041801

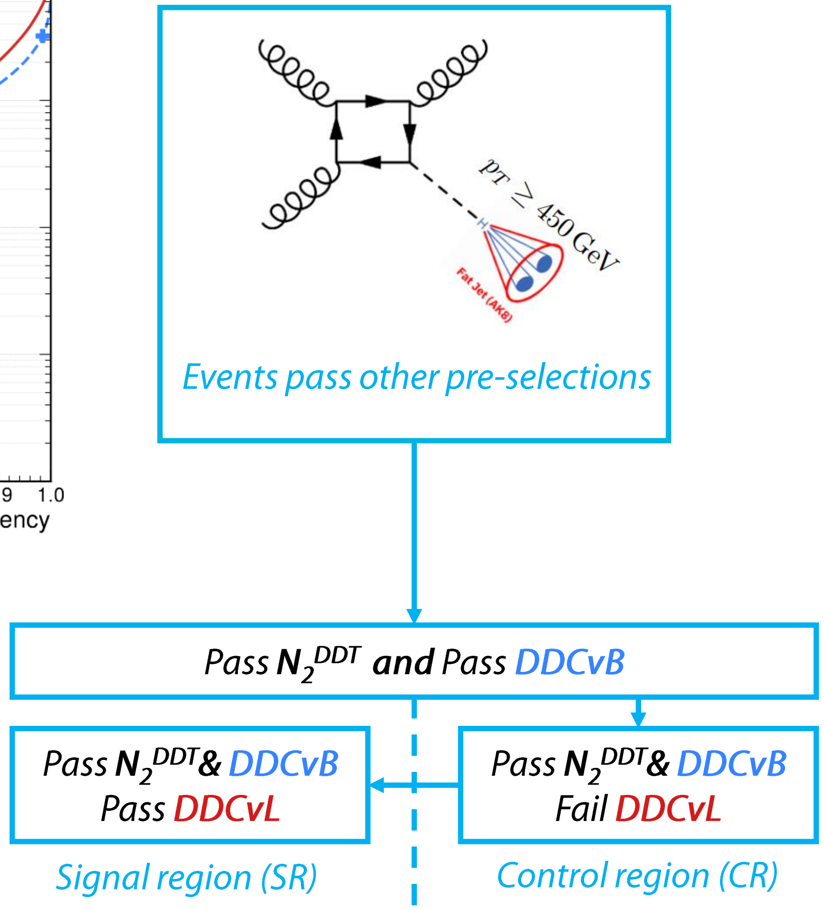
Candidate jet – **AK8 jet**

- Tight ID, **jet p_T > 450 GeV**, jet $|\eta|$ < 2.5
- Selections on $\mathbf{m_{SD}}$, **ρ**, and $\mathbf{N_2^{DDT}}$

No leptons

Dedicated cuts to **reject ttbar**





DeepDoubleX tagger v2 for jet flavor tagging

- **DeepDoubleCvL** – charm vs light
- **DeepDoubleCvB** – charm vs bottom

SR and CR regions are binned

- **23** evenly spaced **bins** of **jet $\mathbf{m_{SD}}$** – range (40, 201) GeV
- **6 bins** of **jet $\mathbf{p_T}$** – (450 or 475, 1200) GeV

Fit strategy and results

PRL 131 (2023) 041801

QCD estimation based on the **same data-driven method** as **boosted H → bb** HIG-21-020

Fit is performed simultaneously across

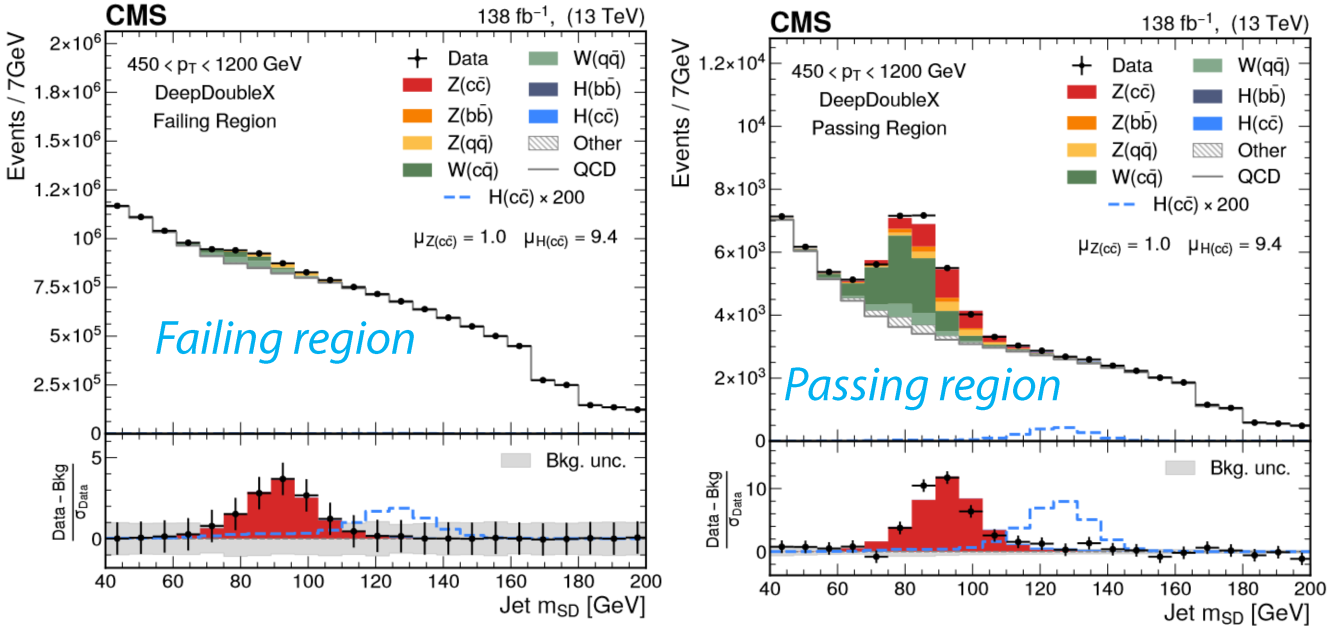
- All subdivisions of the SR and CR
- Single muon control region – to control top background

Fit to Z to measure **Z(cc) signal strength** - Float also H(cc)

- Signal strength $\mu_Z = 1.00^{+0.19}_{-0.17}$, a significance of well over **5σ**
- First observation of Z → cc process in association with jets

Fit to H

- Signal strength $\mu_H = 9.4^{+20.3}_{-19.9}$
- 95% confidence level **observed (expected) upper limit** on the $\sigma_{H\ production} \times BR_{H\to cc}$: **47 (39) times the SM expectation**



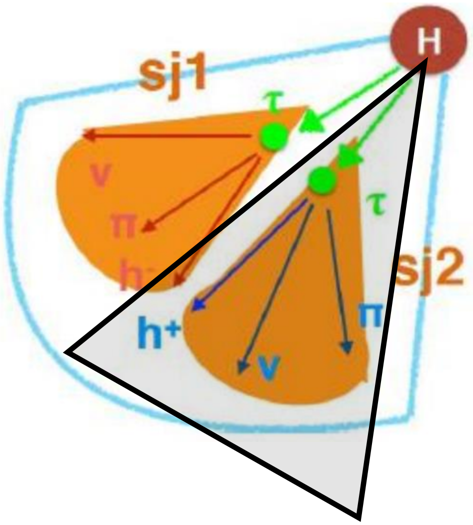
Boosted H → ττ

HIG-21-017

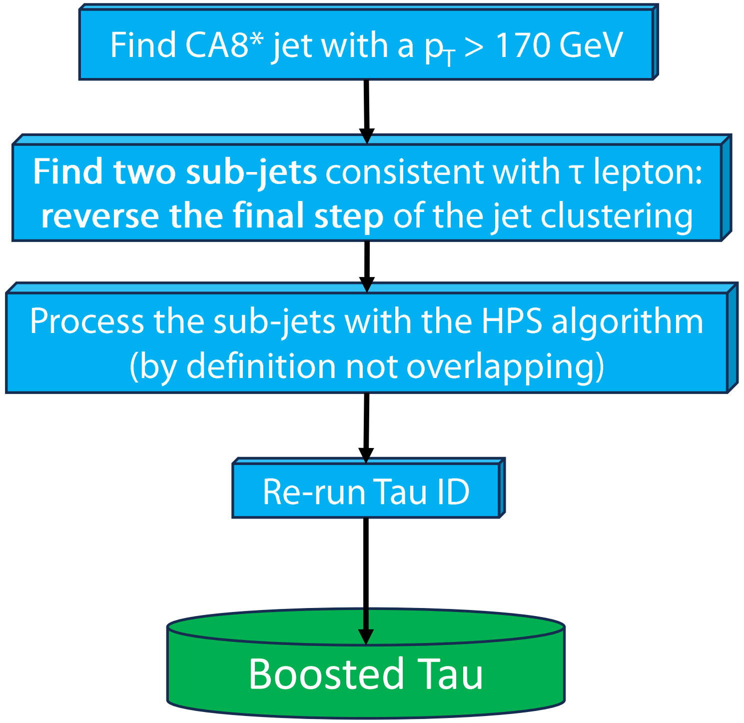
Motivation

[HIG-21-017](HIG-21-017)

- First measurement in **H → ττ** in the **boosted** regime
- The **sensitivity at higher p_T^H is improved** by the **H → ττ and H → bb** because of larger branching fractions
- Improved algorithm relying on the jet substructure technique for **boosted tau** *JINST 13 (2018) P10005*



Particle-flow constituent of one tau spoils the isolation of the next tau



CA8*: *Cambridge–Aachen algorithm with a distance parameter of 0.8*
HPS**: *hadrons-plus-strips*

Event Selection

HIG-21-017

Four final states are considered: **$\mu\tau_h$, $e\tau_h$, $e\mu$, $\tau_h\tau_h$**

Common selections

- Pair of leptons with **$dR_{ll} > 0.1$** and **$dR_{ll} < 0.8$**
- **Higgs p_T** (reconstructed with visible tau candidates and MET) **> 250 GeV**
- **Selections on H_T** (scalar sum of jet p_T) and **$m_{\tau\tau}$**
- Two selected leptons are **opposite sign**
- **No b tagged jet** for **$\mu\tau_h$, $e\tau_h$, $e\mu$**
- **No additional electron or muon** to avoid any overlap across the channels

Different additional selections for different final states

- Dedicated triggers for electrons, muons, or jets
- Kinematic cuts according to the triggers
- Identification selections on the selected leptons
- …

Fake background estimation

HIG-21-017

Example of the background estimation of **$\mu\tau_h$, $e\tau_h$, $\tau_h\tau_h$** final states

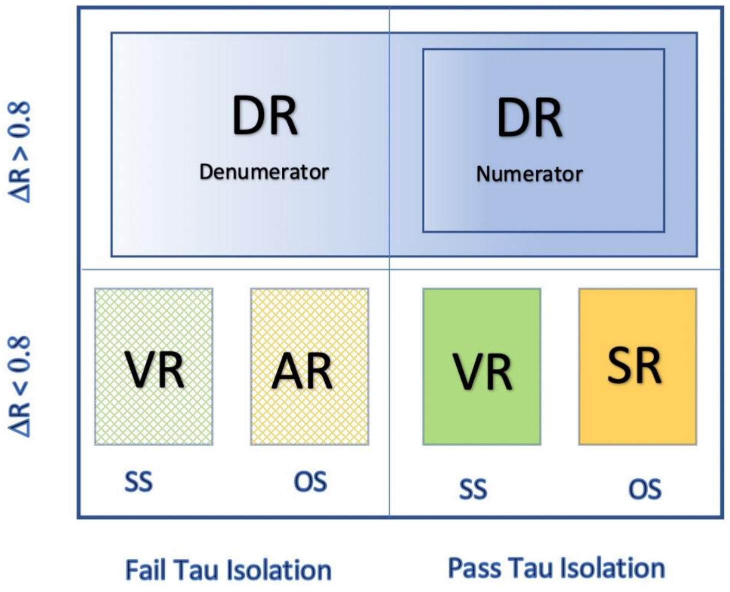
QCD and W+jet backgrounds are estimated by a data-driven method

Events phase space are split to exclusively mutual regions

- **SR**: signal region, aim to measure fake contribution
- **DR**: determination region, to determine the **probability of a jet passing tau isolation WP**
- **AR**: application region, fake bkg. is estimated by **reweighting the events** in this region
- **VR**: validation region, perform a **closure test** to validate the bkg. estimation method

p_T dependent fake rate is applied as **a weight per event**

The **shape of the fake background** is derived **from a same-sign (SS) region**



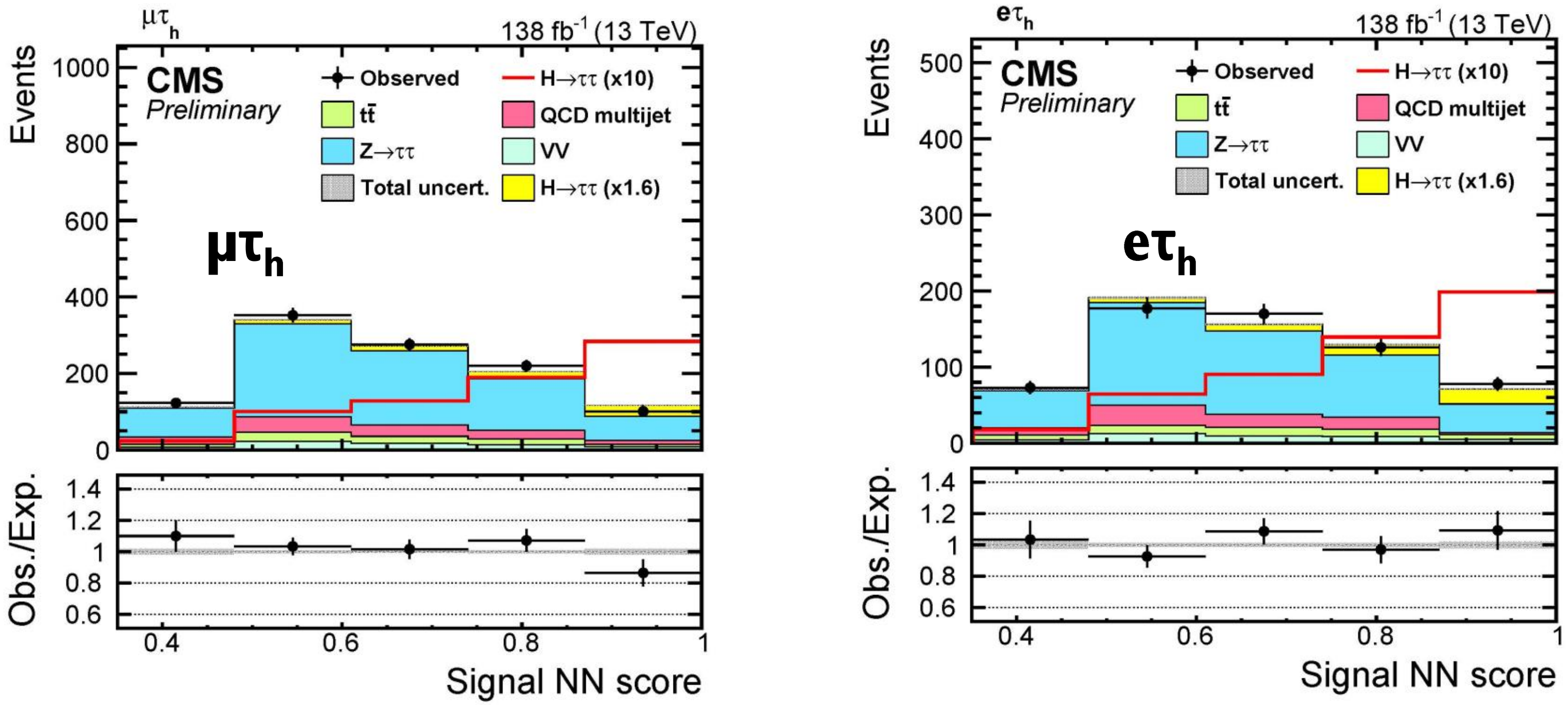
Fit strategy and results

HIG-21-017

A **multi-class NN** is used to **separate the signal** from the Drell-Yan and mis-ID (Fake) backgrounds

The signal region is split into **4 bins of p_T^H**: 250, 350, 450, and 600 GeV

Fit the DNN score in all regions



Best-fit signal strength modifier $\mathbf{1.64^{+0.68}_{-0.54}}$. The sensitivity is driven by the $\boldsymbol{\mu\tau_h}$ and $\boldsymbol{e\tau_h}$ channels

Fiducial differential measurements

HIG-21-017

Inclusive fiducial cross section: $\mathbf{1.96^{+0.86}_{-0.69}}$ **fb**

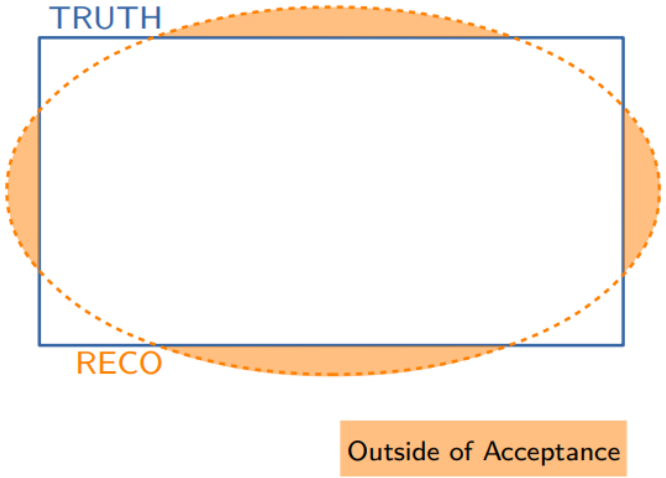
- SM prediction: 1.20±0.20 fb

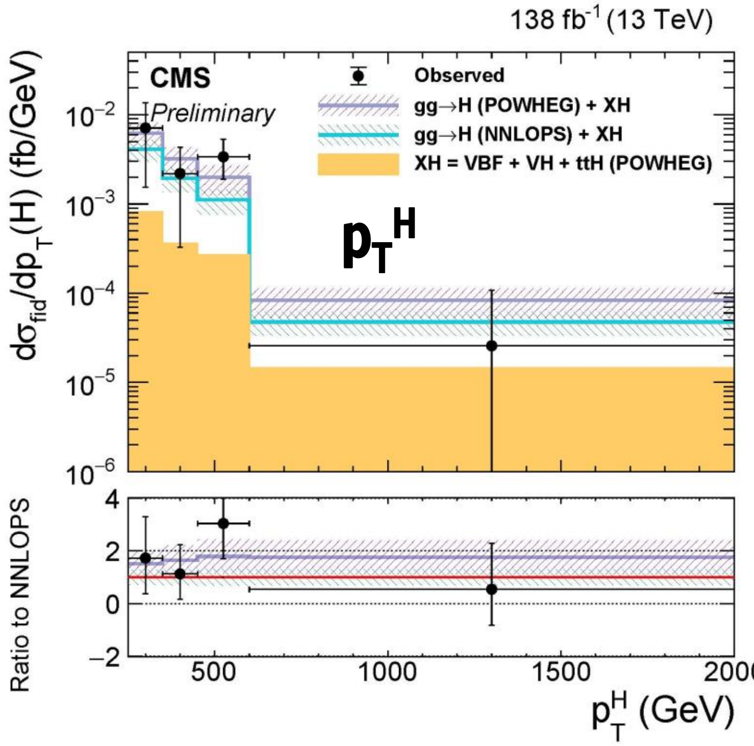
Differential fiducial cross sections as a function of $\mathbf{p_T^H}$ and $\mathbf{p_T^{j1}}$

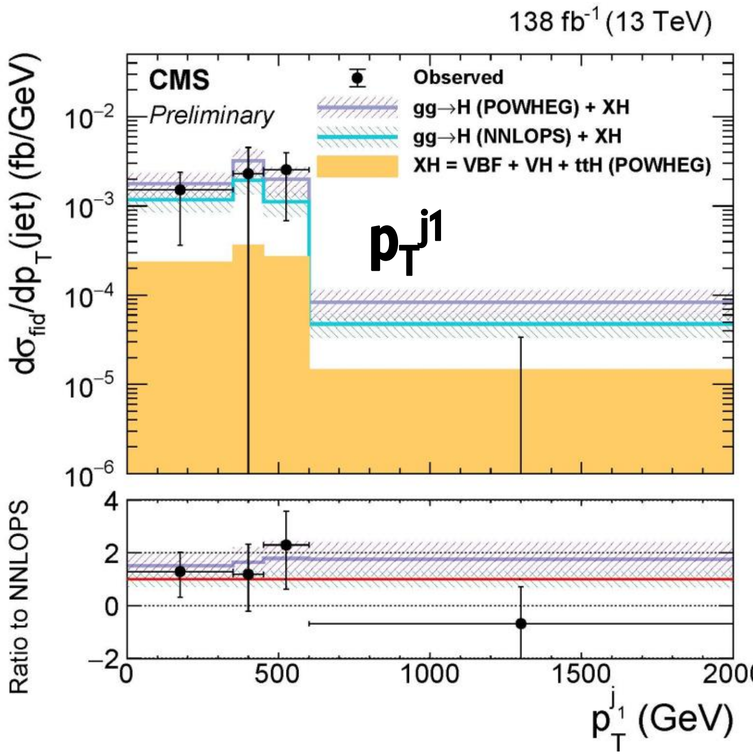
Example: $\mu\tau_h$ fiducial region definition

Generator level selections on

- Muon p_T and p_T^{miss}
- τ_h p_T
- p_T^H







Summary

Recent boosted Higgs measurements with CMS Run2 data are presented

HIG-21-020: boosted $H \rightarrow bb$ produced in the ggF and VBF production modes with $p_T^H > 450$ GeV

- Go beyond the inclusive boosted $H \rightarrow bb$ measurements *JHEP12(2020)085*
- Significances of the VBF and ggF processes are shown separately

PRL 131 (2023) 041801: SM Z and Higgs with $p_T > 450$ GeV and decay to charm quark-antiquark pairs

- First observation of $Z \rightarrow cc$ process in association with jets at a hadron collider
- Set an upper limit on $\sigma_{H\ production} \times BR_{H \rightarrow cc}$

HIG-21-017: first measurement of boosted $H \rightarrow \tau\tau$ signal with $p_T^H > 250$ GeV

- Measured inclusive and differential fiducial cross sections

Continuing to explore the boosted Higgs phase space with Run2 and Run3 data