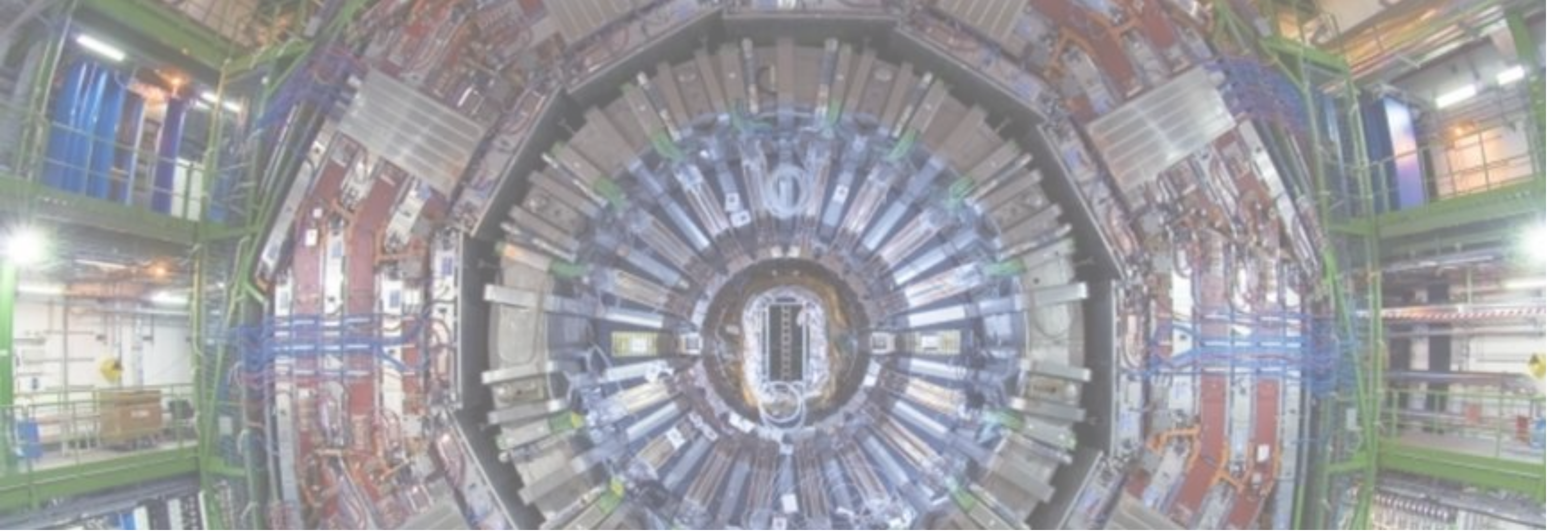


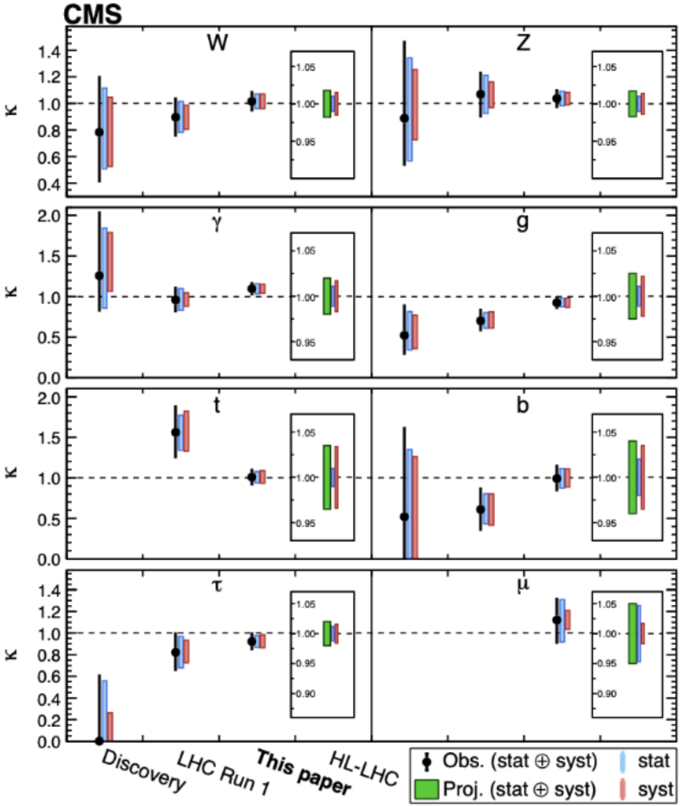
Search for $H \to c\bar{c}$ at CMS in VBF production with Run-3 data

PhD defense – XXXVII cycle
07/04/2025

PhD candidate: Angela Zaza

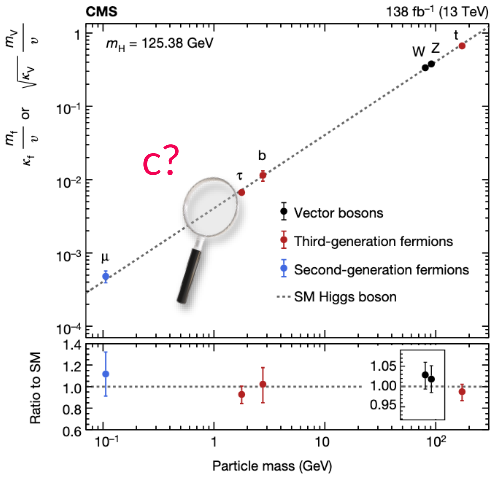
Supervisors: Prof. Anna Colaleo
Prof. Rosamaria Venditti

Motivation of the search



[Nature **607**, pages 60–68 (2022)](#))

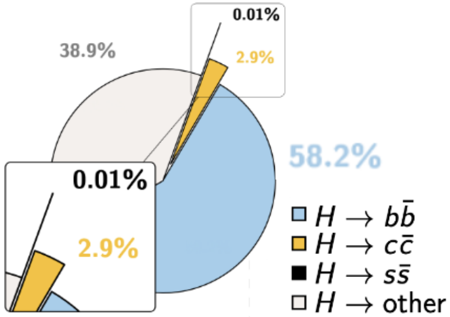
- Higgs boson coupling with gauge bosons and third generation fermions measured with precision of ~10%
- First evidence for $H \rightarrow \mu\mu$ at 3σ with Run-2 data
- Higgs coupling to **second generation quarks** still out of reach



  - one of the highest priority goals of CMS and ATLAS physics program
  - Discrepancies from SM prediction could provide valuable insights into new physics *(BSM model in backup)*

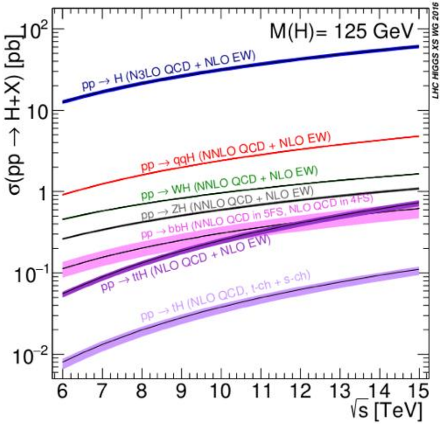
State of the art: CMS and ATLAS (1)

Higgs decay modes

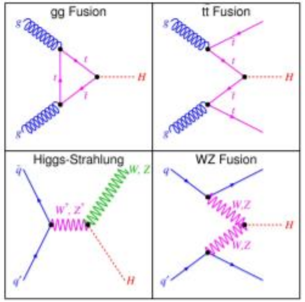


SM $m_H = 125$ GeV $H \to q\bar{q}$ branching fractions,
$H \to u\bar{u}/d\bar{d}$ too small to show!

Higgs production at LHC



ggF



ttH (and bbH)

VH

VBF

$H \to c\bar{c}$ searches performed by CMS and ATLAS with **Run-2** data (140 fb^{-1})

- CMS: $\frac{\sigma(VH)\cdot B(H\to c\bar{c})}{\sigma(VH)_{SM}\cdot B(H\to c\bar{c})_{SM}} < 14 \quad @95\%CL \qquad 1.1 < |k_c| < 5.5$

  NEW $\boxed{\frac{\sigma(ttH)\cdot B(H\to c\bar{c})}{\sigma(ttH)_{SM}\cdot B(H\to c\bar{c})_{SM}} < 7.8 \quad @95\%\,CL \qquad |k_c| < 3.0}$

- ATLAS: $\frac{\sigma(VH)\cdot B(H\to c\bar{c})}{\sigma(VH)_{SM}\cdot B(H\to c\bar{c})_{SM}} < 11.5 \quad @95\%CL \qquad |k_c| < 4.2$

Most stringent constraints to date

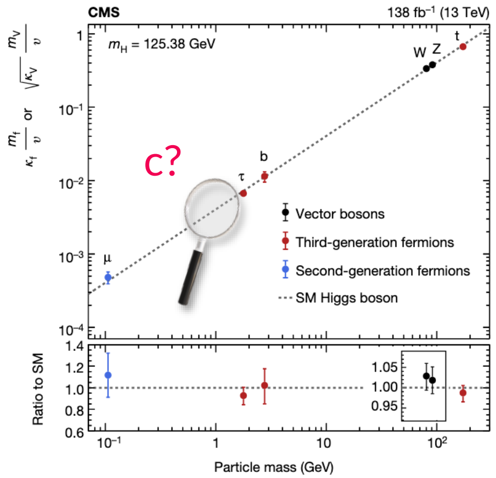
! Search for $H \to c\bar{c}$ **extremely challenging** because of the **small branching ratio** and the **overwhelming QCD multijet background**

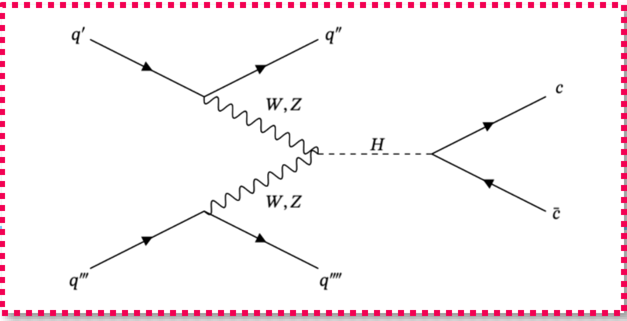
Heavy-flavour tagging crucial

Search for VBF $H \to c\bar{c}$ at CMS

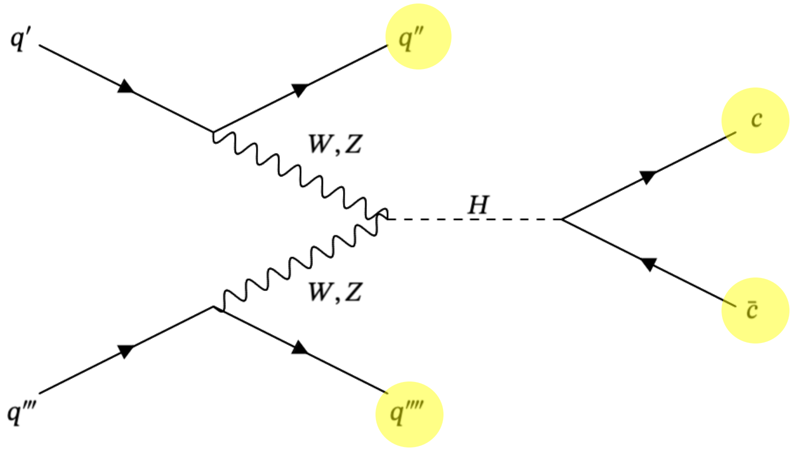
Challenges of this search

- Lack of a suitable trigger before this study
  - development of a **dedicated trigger** with c-tagging online
- Identification of c jets
  - novel **ParticleNet** heavy-flavour tagger
- Overwhelming QCD background
  - **machine learning** approaches for signal vs background discrimination
- Large number of resonant backgrounds: H→bb, Z→qq, W→qq
  - Sophisticated **signal extraction**



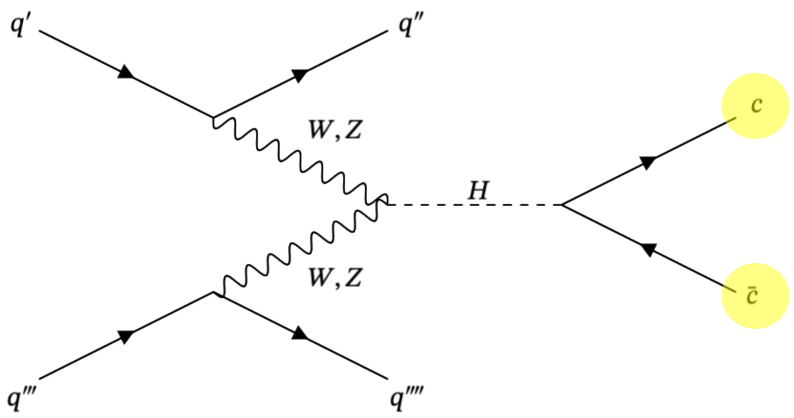


Signal topology



Four quarks in the final state:

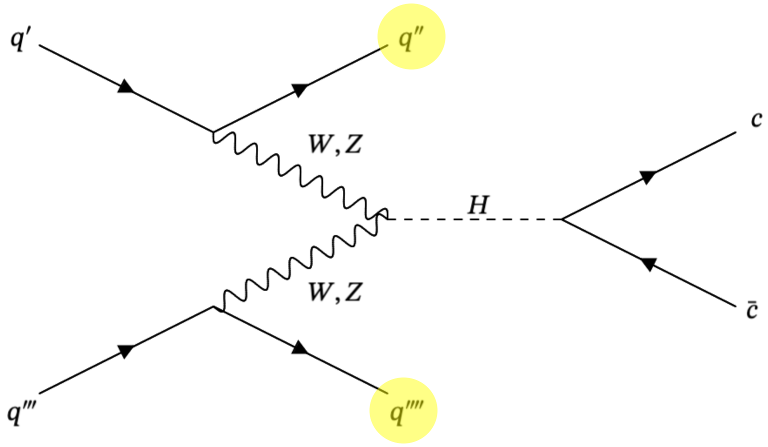
Signal topology



Four quarks in the final state:

- two **charm quarks** produced by the Higgs boson decay
  emitted centrally

Signal topology

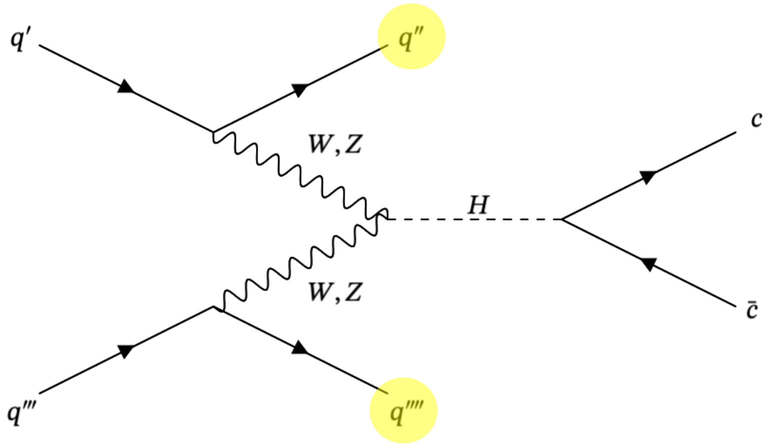


Four quarks in the final state:

- two **charm quarks** produced by the Higgs boson decay
  emitted centrally

- two **quarks** produced by the **VBF** emission
  emitted **forward and backward**, with large aperture

Signal topology



Four quarks in the final state:

- two **charm quarks** produced by the Higgs boson decay
  emitted centrally

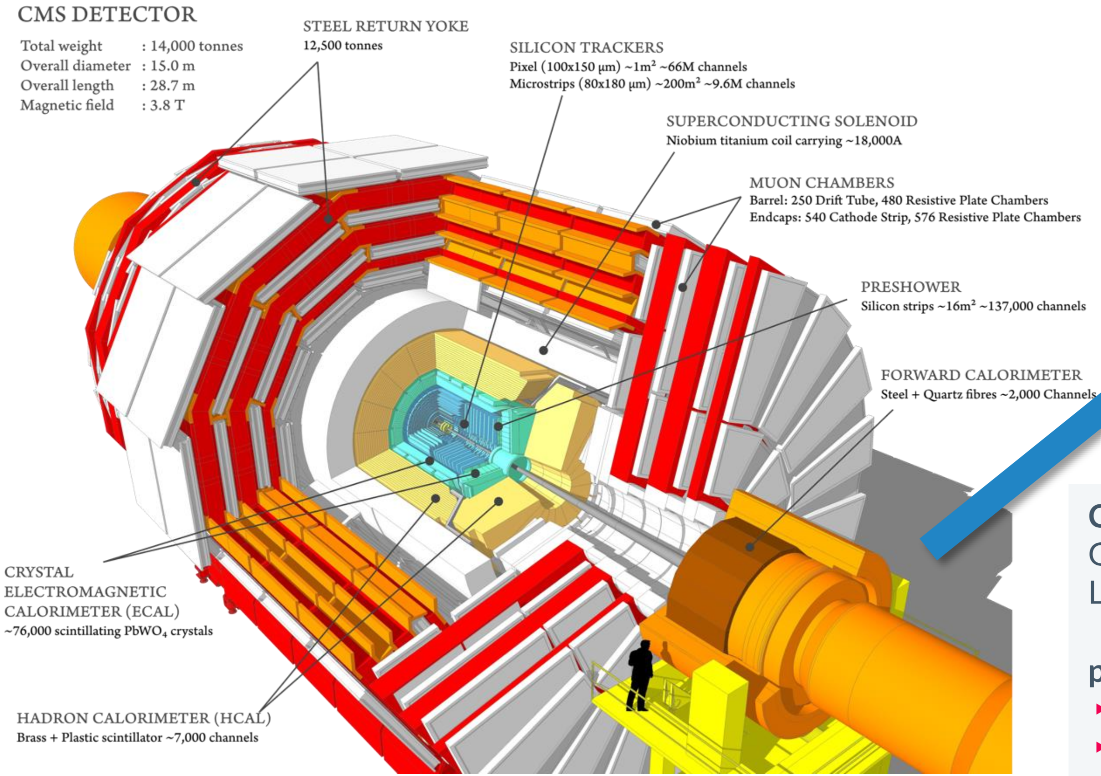
- two **quarks** produced by the **VBF** emission
  emitted **forward and backward**, with large aperture

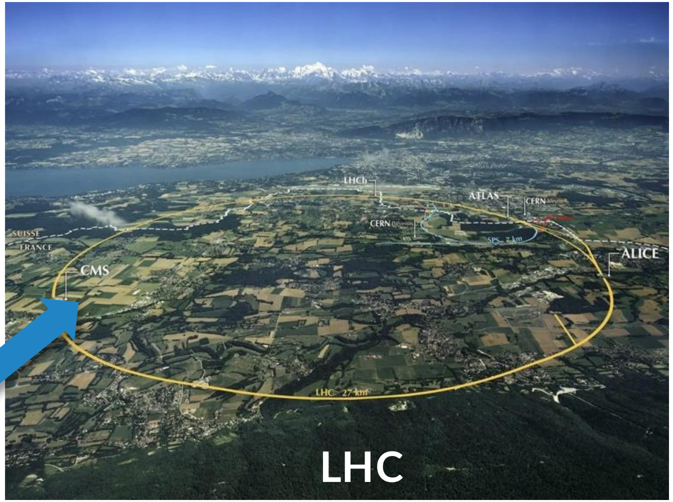
- Quarks hadronize into spray of particles reconstructed as jets

- Jet properties exploited in **trigger** development and **BDT** algorithm implementation

Search for $H \rightarrow c\bar{c}$ at CMS: Object reconstruction and Trigger studies

CMS Experiment

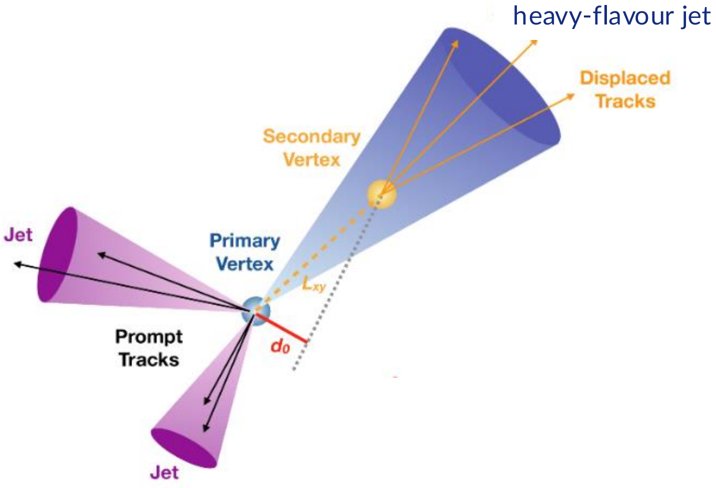




Compact Muon Solenoid (CMS)
General purpose detector at the Large Hadron Collider (**LHC**)

proton-proton collisions in Run-3:
- center-of-mass energy: 13.6 TeV
- peak instantaneous luminosity: $2 \cdot 10^{34}$ cm^{-2} s^{-1}

Jets and heavy-flavour tagging

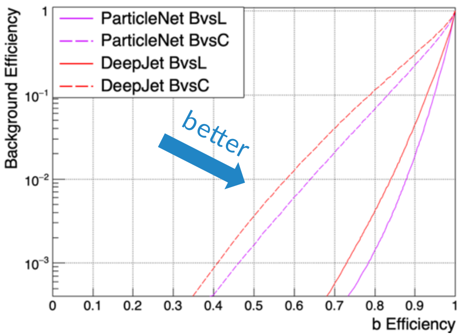


Heavy-flavour jets:

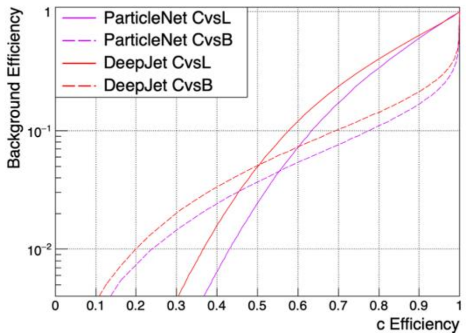
- Lifetime of b (c) hadrons ~ 1.5 ps (~1 ps)
  → **displaced tracks from PV** (impact parameter)
  → **seconday vertex**
- Larger mass and harder fragmentation wrt light quarks and gluons
  → **larger p_T of the decay products**
- Presence of a **muon or electron** in 20% (10%) of the cases

ParticleNet

- Customized dynamic graph convolutional neural network (DGCNN)
- Each jet represented as an unordered, permutation-invariant set of particles → **particle cloud**
- Developed for boosted jets in 2020, then adapted to small-radius jets
- Outstanding improvements wrt DeepJet (default tagger in Run-2)

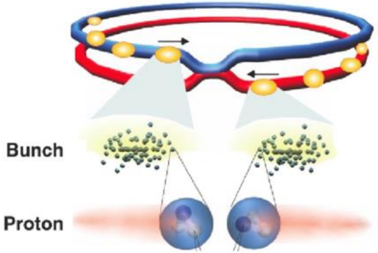


b-tagging



c-tagging

CMS Trigger system

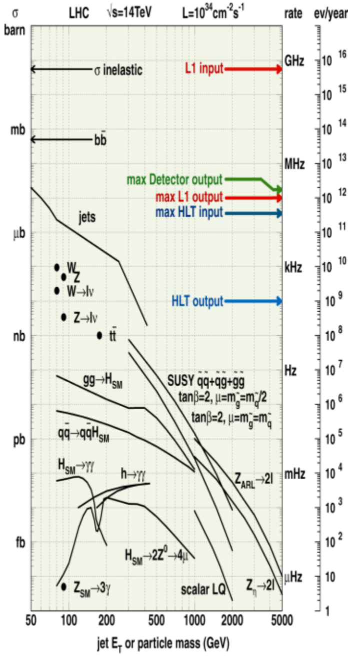


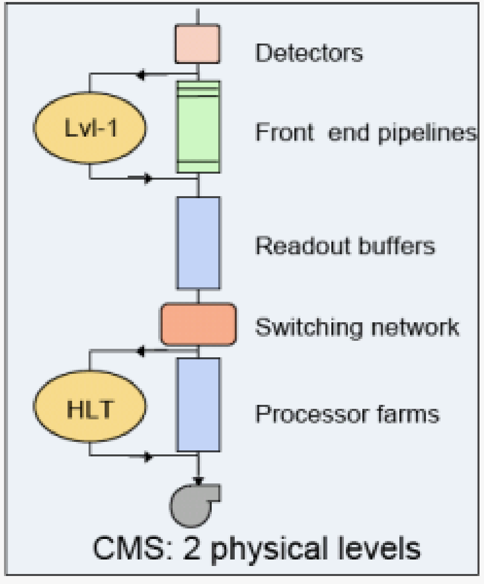
Proton-Proton	2835 bunch/beam
Protons/bunch	10^{11}
Beam energy	7 TeV (7×10^{12} eV)
Luminosity	10^{34} cm^{-2} s^{-1}
Crossing rate	40 MHz
Collisions ≈	10^7 - 10^9 Hz

data

$$V \sim 10^{23}\ Bytes/year$$

Impossible to store!



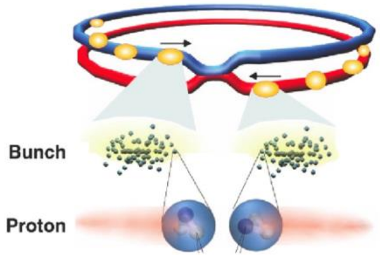


2-level trigger:

Level 1 (L1) → **HLT**

- **Hardware**
- Exploits information only from **calorimeters** and **muon system**
- Latency: 4 μs
- **Rate from 40 MHz to maximum ~110 kHz**

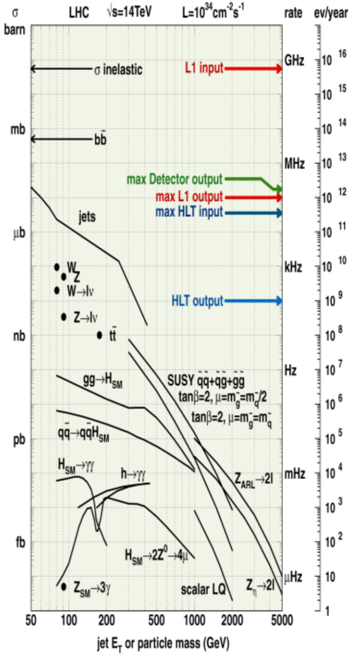
CMS Trigger system

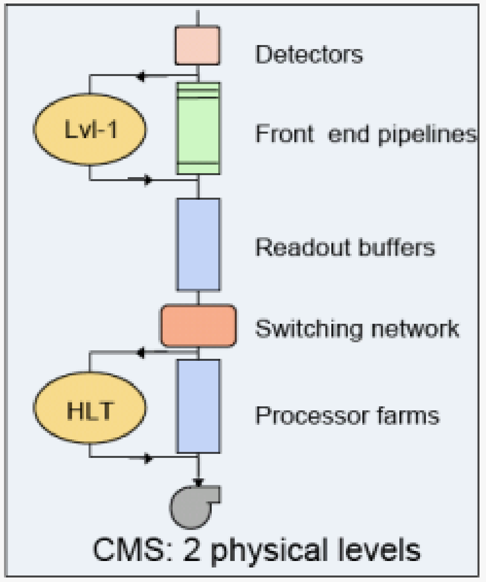


Proton-Proton	2835 bunch/beam
Protons/bunch	10^{11}
Beam energy	7 TeV (7×10^{12} eV)
Luminosity	10^{34} cm^{-2} s^{-1}
Crossing rate	40 MHz
Collisions ≈	10^7 - 10^9 Hz

data → $V \sim 10^{23}\ Bytes/year$

Impossible to store!



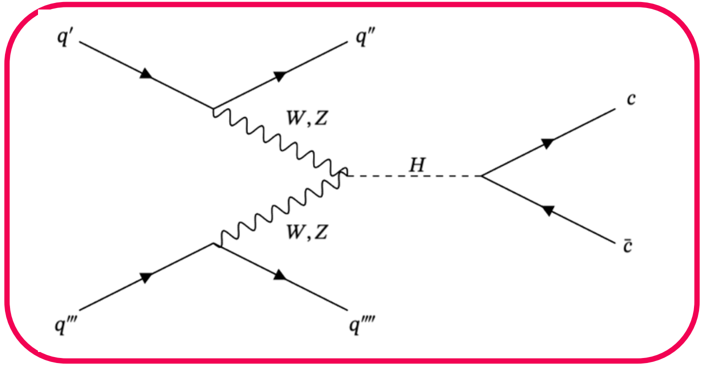


2-level trigger:

L1 → **High Level Trigger (HLT)**

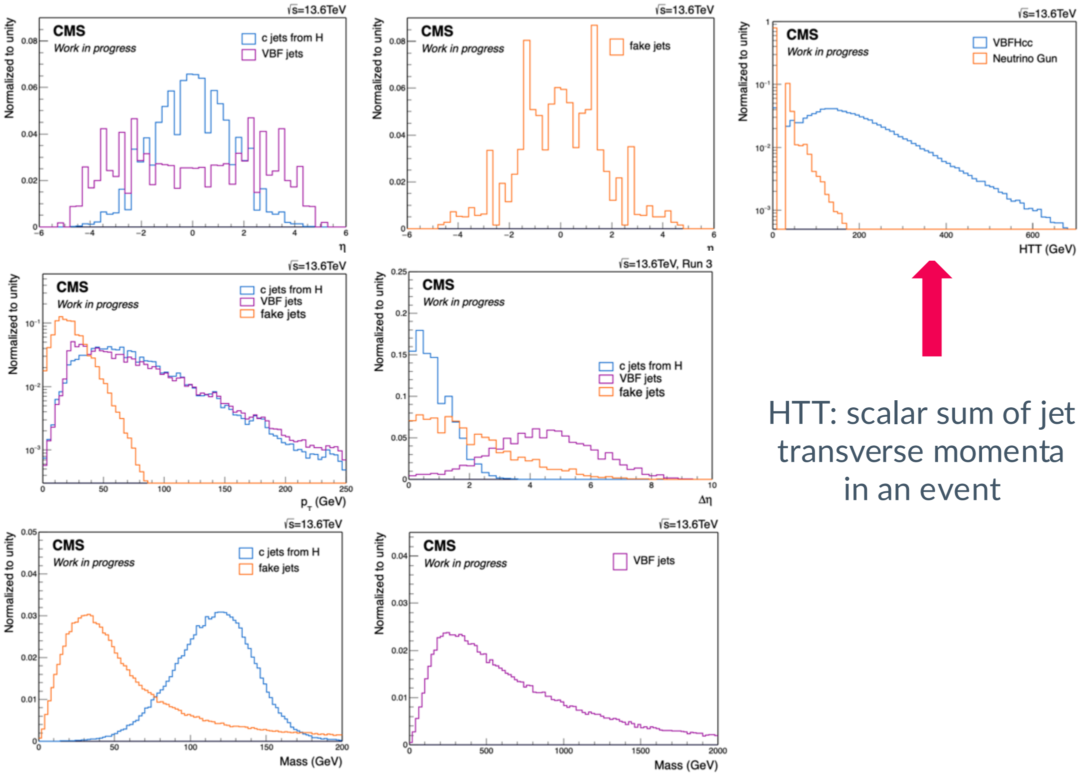
- **Software**
- Full detector information exploited
- Algorithms implemented to run both on **CPU**s and **GPU**s
  - automatically directed to GPU if available
- Rate from 110 kHz to
  2.6 kHz prompt + 3 kHz parking

Trigger studies: L1 seeds (1)



First step for the implementation of the HLT path:
Definition of a set of L1 seeds

Large number of fake jets from Pile-up interactions → study of kinematic distributions of signal and fake jets for discrimination



HTT: scalar sum of jet transverse momenta in an event

Jets reconstructed with L1 information geometrically matched with quarks from MC truth

Trigger studies: L1 seeds (2)

List of L1 seeds included in the HLT path dedicated to VBF $H \rightarrow c\bar{c}$

L1 seed	Selection	
L1_TripleJet_95_75_65_DoubleJet_75_65_er2p5	3 jets with • p_T > 95, 75, 65 GeV	2 jets with • p_T > 75, 65 GeV • $\|\eta\|$ < 2.5
L1_TripleJet_100_80_70_DoubleJet_80_70_er2p5	3 jets with • p_T > 100, 80, 70 GeV	2 jets with • p_T > 80, 70 GeV • $\|\eta\|$ < 2.5
L1_SingleJet180	1 jet with p_T > 180 GeV	
L1_SingleJet200	1 jet with p_T > 200 GeV	
L1_DoubleJet_110_35_DoubleJet35_Mass_ _Min620	2 jets with • p_T > 110, 35 GeV	2 jets with • p_T > 35 GeV • Inv mass > 620 GeV
L1_QuadJet_95_75_65_20_DoubleJet_75_65_ _er2p5_Jet20_FWD3p0	4 jets with • p_T > 95, 75, 65, 20 GeV 1 jet with p_T > 20 GeV and $\|\eta\|$ > 3	2 jets with • p_T > 75, 65 GeV • $\|\eta\|$ < 2.5
L1_HTT360er	HTT > 360 GeV	

Trigger path for VBF $H \to c\bar{c}$

HLT_QuadPFJet100_88_70_30_PNetTag1CvsAll0p5_VBF3Tight

1 **Kinematic sequence**

4 jets with p_T > 100, 88, 70, 30 GeV
2 jets with p_T > 30 GeV η < 2.5

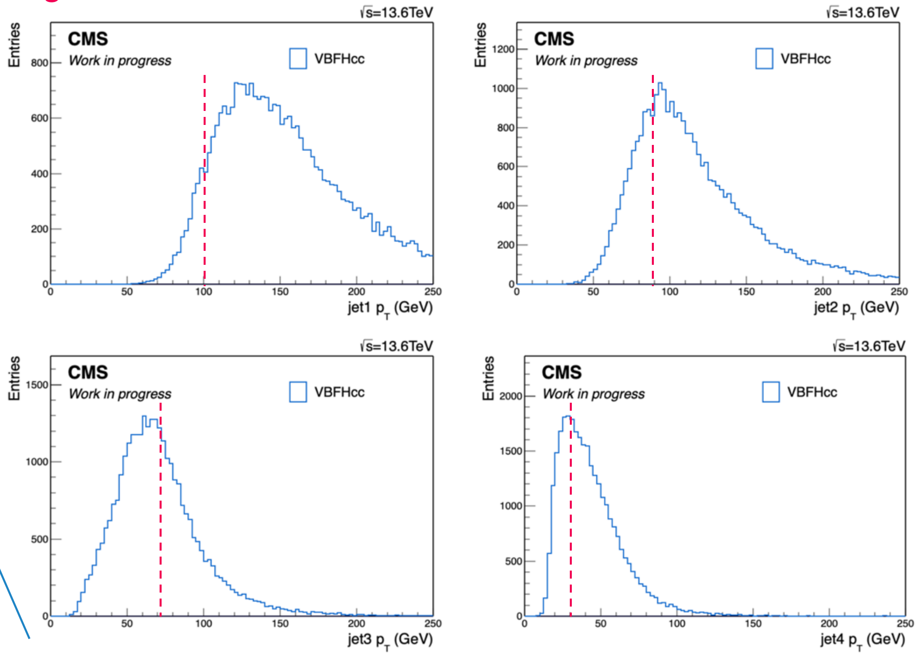
2 **C-tagging sequence**

1 jet with p_T > 30 GeV,
η < 2.5 and CvsAll>0.5

3 **VBF-tagging sequence**

Among the 4 p_T-leading jets:

- 2 jets with highest CvsAll: **c-jets**
- other 2 jets: **VBF jets** (qq), with
  $\Delta\eta(qq) > 3.5$
  $M(qq) > 460\ GeV$



Trigger path for VBF $H \to c\bar{c}$

HLT_QuadPFJet100_88_70_30_PNetTag1CvsAll0p5_VBF3Tight

Kinematic sequence (1)

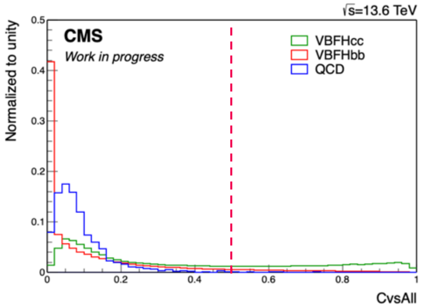
4 jets with p_T > 100, 88, 70, 30 GeV
2 jets with p_T > 30 GeV η < 2.5

C-tagging sequence (2)

1 jet with p_T > 30 GeV,
η < 2.5 and CvsAll>0.5

VBF-tagging sequence (3)

Among the 4 p_T-leading jets:
- 2 jets with highest CvsAll: **c-jets**
- other 2 jets: **VBF jets** (qq), with
  $\Delta\eta(qq) > 3.5$
  $M(qq) > 460\ GeV$



ParticleNet @HLT

- **first** trigger path with **c-tagging at HLT**
- one of the first trigger paths with ParticleNet

$$\text{CvsAll} = \frac{\text{prob(c)}}{\text{prob(c)+prob(b)+prob(uds)+prob(g)+prob(tauh)}}$$

- All paths with heavy-flavour tagging online moved to ParticleNet in 2024

Trigger path for VBF $H \to c\bar{c}$

HLT_QuadPFJet100_88_70_30_PNetTag1CvsAll0p5_VBF3Tight

1 Kinematic sequence

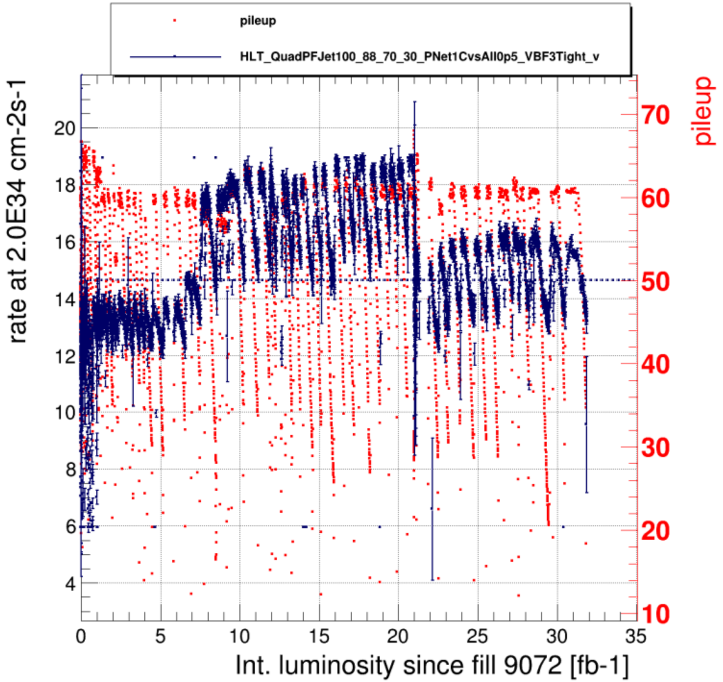
4 jets with p_T > 100, 88, 70, 30 GeV
2 jets with p_T > 30 GeV η < 2.5

2 C-tagging sequence

1 jet with p_T > 30 GeV,
η < 2.5 and CvsAll>0.5

3 VBF-tagging sequence

Among the 4 p_T-leading jets:
- 2 jets with highest CvsAll: **c-jets**
- other 2 jets: **VBF jets** (qq), with
  $\Delta\eta(qq) > 3.5$
  $M(qq) > 460\ GeV$



Trigger acceptance: 1.8%

- Deployed online for 2023 data taking and collecting data since!

Trigger performance studies - p_T

MC do not replicate **data** perfectly → **mismatching** in trigger performance to be accounted for with proper **scale factors** (SF)

$$SF = SF_{pT} \cdot SF_{ctag} \cdot SF_{VBF}$$

1. **p_T trigger SFs:**

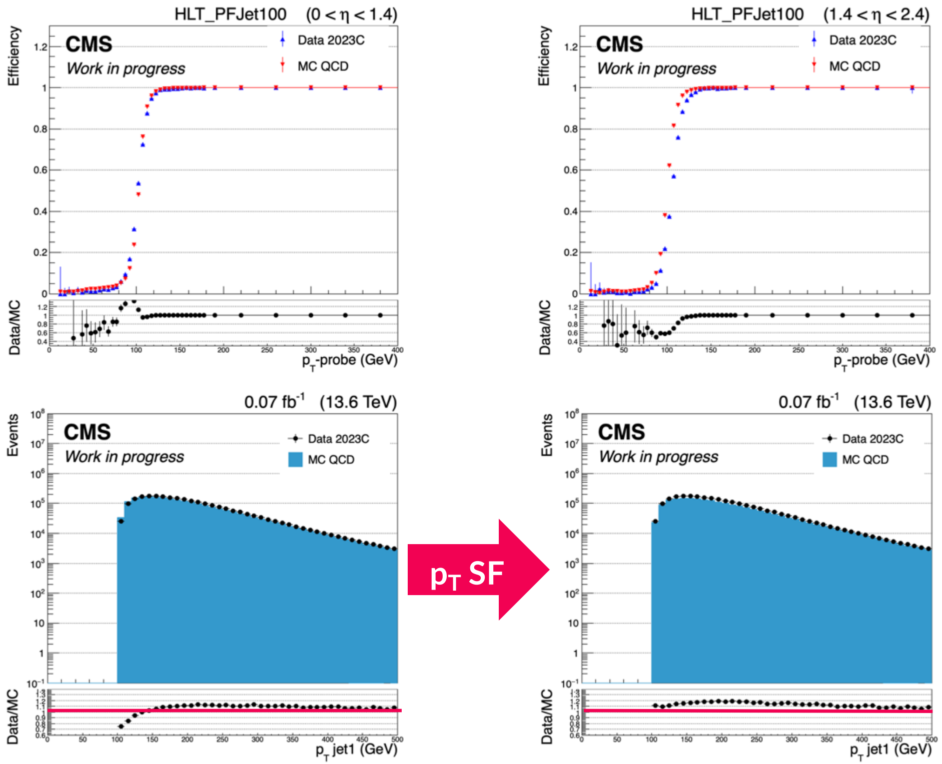
Tag and probe method

- Events with back-to-back di-jet topology selected
- Tag: leading offline jet matched to an HLT object
- Probe: subleading offline jet

For each p_T threshold (thr)
Efficiency:

$$\frac{p_T\,probe, matched\ to\ HLT\ jet\ with\ p_T > thr}{p_T\,probe}$$

- validated on events selected with the **control path** HLT_QuadPFJet100_88_70_30: exactly the same as the signal path, without c-tagging and VBF sequences



Trigger performance studies - ctag

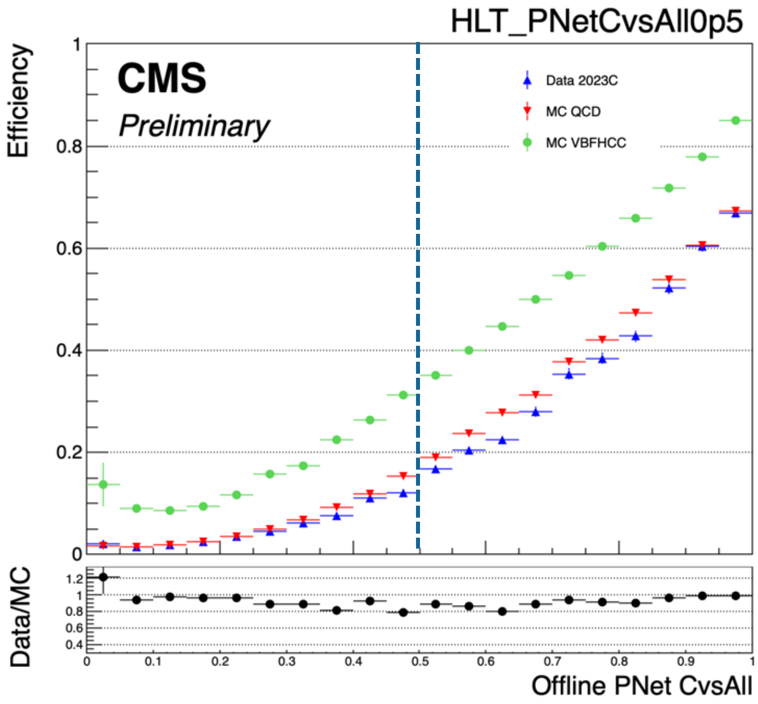
MC do not replicate **data** perfectly → **mismatching** in trigger performance to be accounted for with proper **scale factors** (SF)

$$SF = SF_{pT} \cdot SF_{ctag} \cdot SF_{VBF}$$

2. **c-tag trigger SFs:**
   - evaluated on data and QCD MC events selected with control path HLT_QuadPFJet100_88_70_30 and VBF offline selection

   $$\text{Efficiency} = \frac{\text{PNet offline score of the most c−tagged jet, } \textbf{HLT:signal+control}}{\text{PNet offline score of the most c−tagged jet, } \textbf{HLT:control}}$$

3. **VBF trigger SFs:**
   - Assumed to be ~1

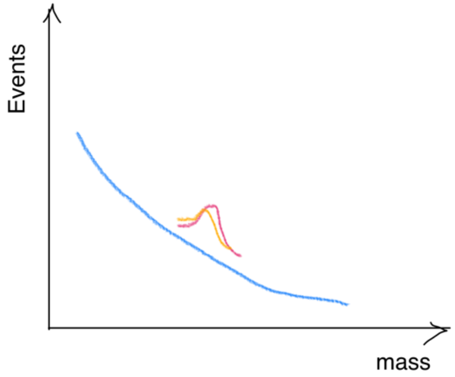


Search for VBF $H \to c\bar{c}$ at CMS: Statistical analysis of 2023 data ($27\ \mathrm{fb}^{-1}$)

Analysis strategy

Signal: VBF $H \to cc$
gg $H \to cc$

Main Backgrounds: QCD (dominant) } continuum
VBF $H \to bb$, gg $H \to bb$
QCD $Z/W \to qq$, EWK $Z/W \to qq$ } resonant



Offline pre-selection (similar to trigger) → BDT for signal vs QCD discrimination → Signal and background modelling → Final result: expected upper limit on the signal strength

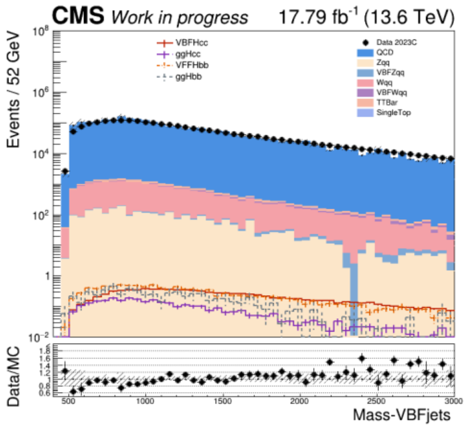
Signal and background modelling ↔ QCD → data-driven; resonant bkg → MC

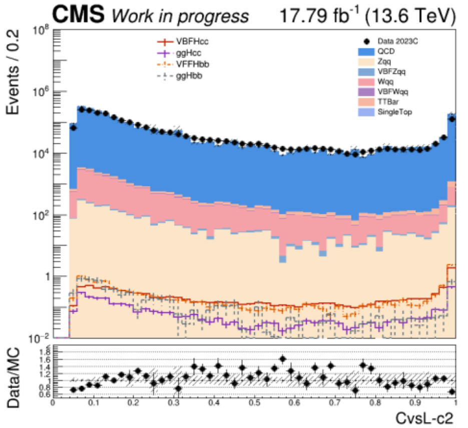
- Blind analysis: only data in the Higgs mass sidebands are accessible

Offline pre-selection

- Trigger HLT_QuadPFJet100_88_70_30_PNet1CvsAll0p5_VBF3Tight_v
- Electron/Muon veto
- MET p_T < 170 GeV
- 4 leading p_T jets with p_T > 105, 90, 75, 35 GeV and $|\eta|$ < 4.7 matching with HLT objects
- 2 jets with highest **PNet** CvsL score and $|\eta|$ < 2.4: c-jets respectively medium and tight c-tagging WPs are applied
- Other 2 jets: VBF-jets
- VBF jets: invariant mass > 500 GeV, $\Delta\eta$ > 3.8

Data/MC corrections: Gen weights, PU reweighting, trigger SFs, JECs



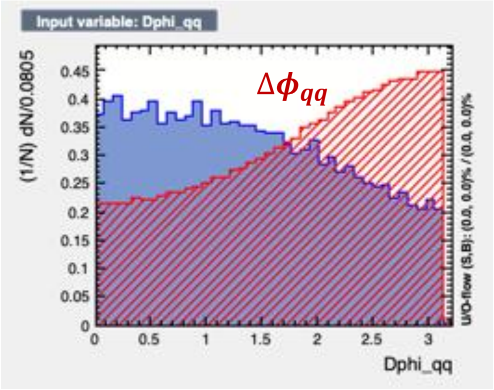


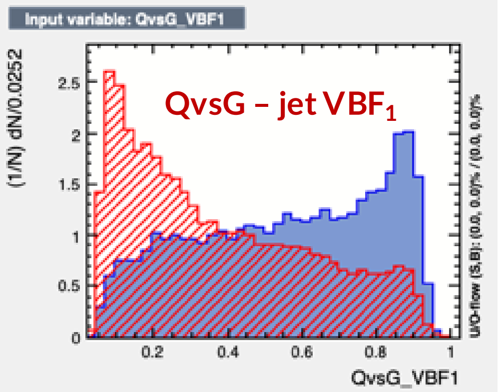
$H \to c\bar{c}$ vs QCD discriminator (1/4)

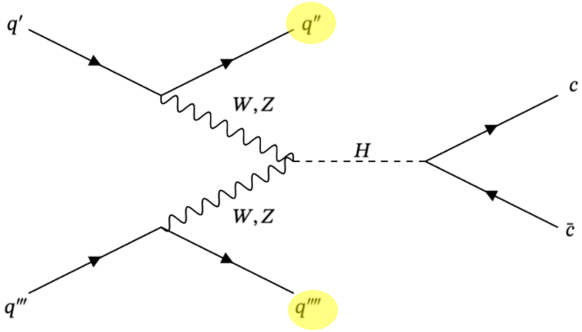
BDT trained over signal from MC and bkg from data in sidebands with:

VBF related variables

- m_{qq}: invariant mass of the two VBF jets
- $|\Delta\eta_{qq}|$: absolute pseudorapidity difference of the two VBF jets
- $\Delta\phi_{qq}$: absolute azimuthal angle difference of the two VBF jets
- α_{qq}: Min(α_{q1}, α_{q2}), where $\alpha_{q1/q2}$ is the angle between the lead/sublead VBF jet and the boosted system of the VBF jet pair
- QvsG: PNet QvsG score of the two VBF jets





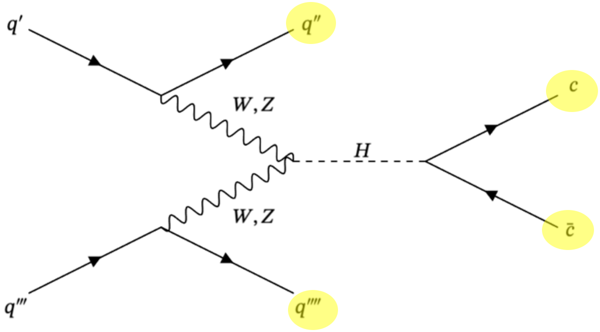


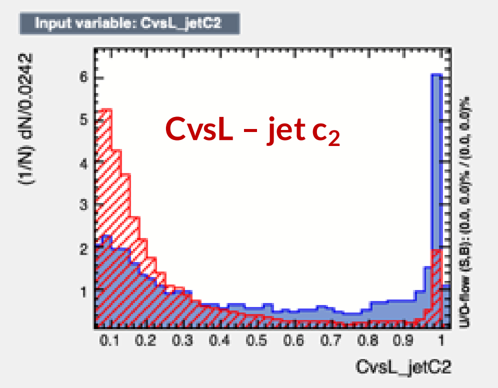
H → $c\bar{c}$ vs QCD discriminator (2/4)

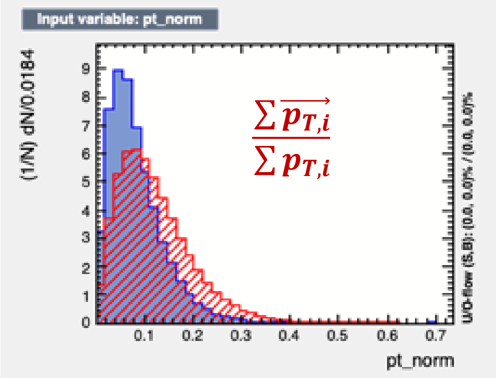
BDT trained over signal from MC and bkg from data in sidebands with:

Higgs related variables

- c-tagging: CvsL and CvsB PNet scores of the two c jets
- total longitudinal momentum of the selected four jets
- normalized sum of the transverse momenta of the selected four jets
- angular distance ΔR between the Higgs boson candidate and the lead and sublead VBF jets
- $\Delta(\phi_{qq} - \phi_{cc})$: difference of azimuthal angle of the VBF jet pair system and the c jet pair system





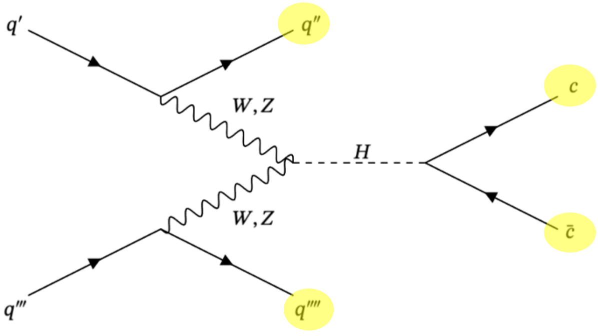


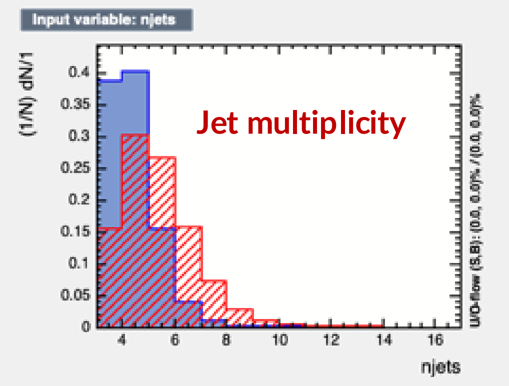
H → $c\bar{c}$ vs QCD discriminator (3/4)

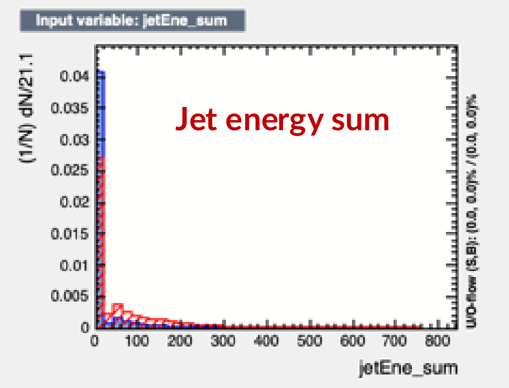
BDT trained over signal from MC and bkg from data in sidebands with:

Event related variables

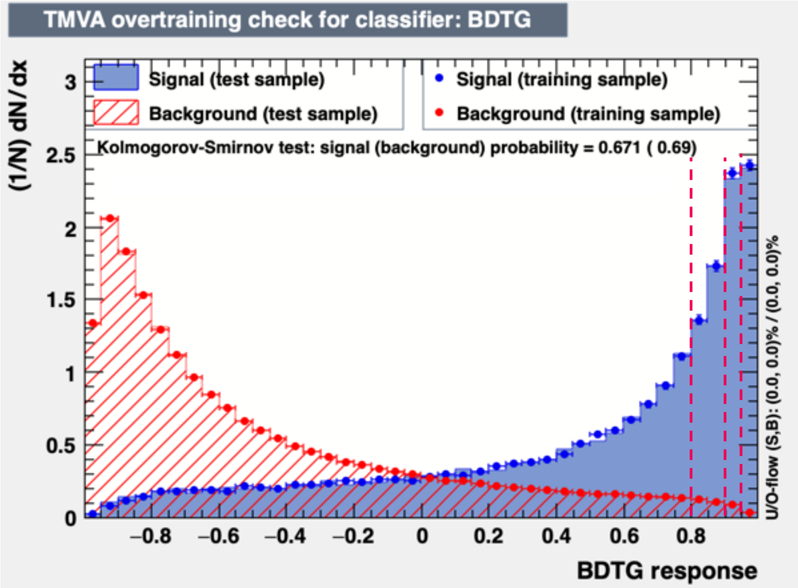
- jet multiplicity in the region $|\eta| < 2.4$ above 20 GeV
- sum of energy and transverse momentum of all the jets with $p_T > 30$ GeV and $|\eta| < 2.4$ excluding the selected four jets

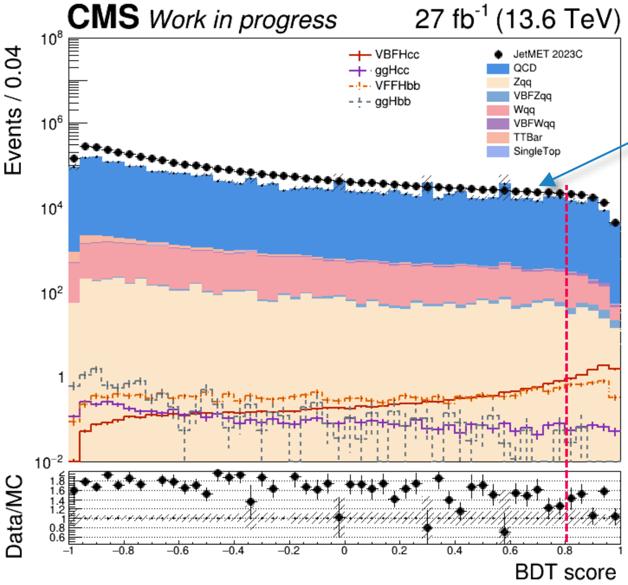






H → $c\bar{c}$ vs QCD discriminator (4/4)



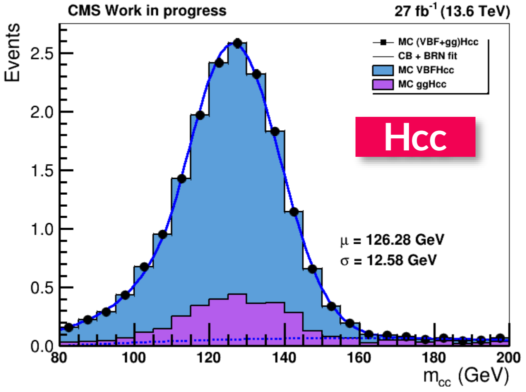


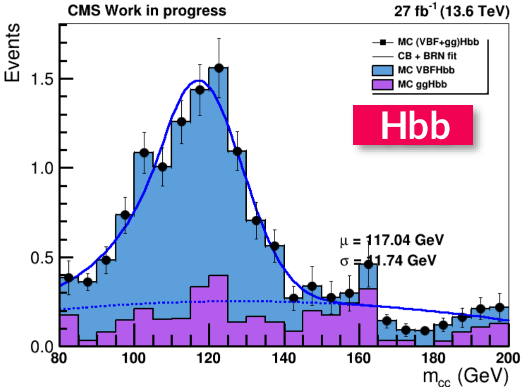
- QCD from MC (mismodeling effects) better agreement in the signal region

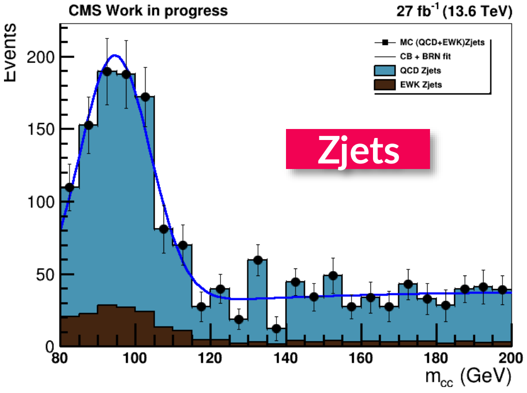
- BDT trained over signal from MC and bkg from data sidebands
- Overall good agreement between training and test distributions
- Three categories based on the BDT score

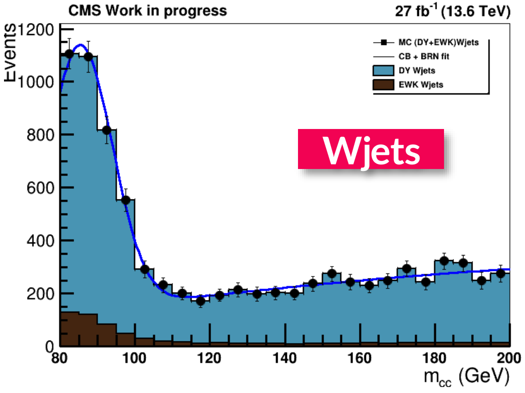
category	BDT score
CAT0	0.8 – 0.9
CAT1	0.9 – 0.95
CAT2	0.95 – 1

MC shapes (Hcc, Hbb, Zqq, Wqq)









No thresholds applied on the BDT score

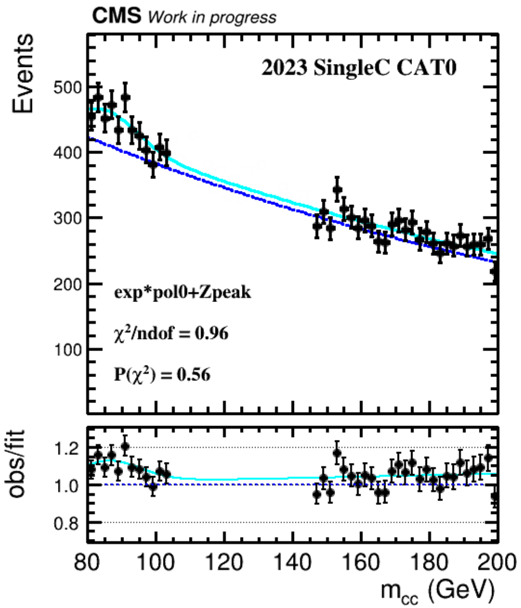
Yields in each category

process	CAT0	CAT1	CAT2
VBFHcc	2.66 ± 0.02	2.13 ± 0.02	2.01 ± 0.02
ggHcc	0.14 ± 0.01	0.09 ± 0.01	0.06 ± 0.01
VBFHbb	1.39 ± 0.07	0.98 ± 0.06	0.51 ± 0.04
ggHbb	0.06 ± 0.04	-	-
Zjets	92.02 ± 1.61	45.80 ± 1.41	24.71 ± 1.21
Wjets	241.06 ± 1.98	52.89 ± 3.42	24.21 ± 1.74

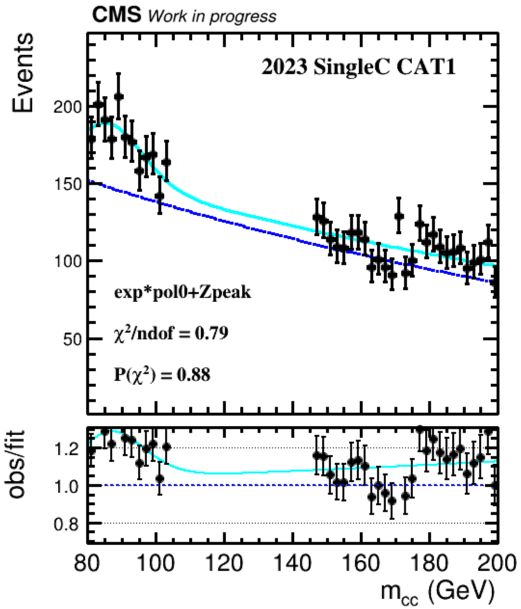
MC distributions fitted with CB and Bernstein polynomial function

Continuum background modelling

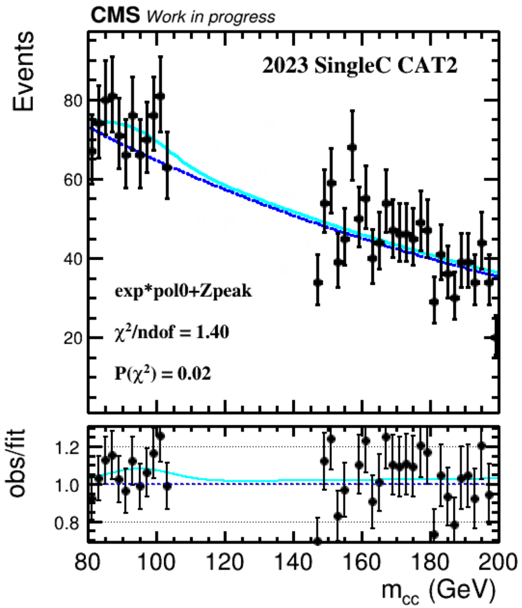
- ▷ The continuum background shape is extracted from an exponential fit of the mass spectrum in the sidebands of Higgs nominal mass: [80,104] – [146,200] GeV
- ▷ QCD, Z peak and W peak fitted to data sidebands simultaneously (Z/W model taken from fit to MC histograms)



QCD yield: 19491.0 ± 153.6



QCD yield: 8130.8 ± 99.2



QCD yield: 3119.7 ± 61.9

Systematic uncertainties (1/2)

- **Fit bias uncertainty:**
  - related to the choice of the continuum background fitting function
  - assumed a conservative 20% on the signal (spurious signal method)

Source of uncertainty	$VBF\ H \to c\bar{c}$	$gg\ H \to c\bar{c}$	$VBF\ H \to b\bar{b}$	$gg\ H \to b\bar{b}$	Z/W+jets
Luminosity	1.4%	1.4%	1.4%	1.4%	1.4%
VBF model	8%	-	- 8%	-	-
Trigger	3%	3%	4%	1%	3%
c tagging	10%	10%	10%	10%	10%
Higgs XS QCD scale	$^{+0.4\%}_{-0.3\%}$	$^{+4.6\%}_{-6.7\%}$	$^{+0.4\%}_{-0.3\%}$	$^{+4.6\%}_{-6.7\%}$	-
Higgs XS PDF + α_s	2.1%	3.2%	2.1%	3.2%	-
Higgs decay BR	$^{+5\%}_{-3\%}$	$^{+5\%}_{-3\%}$	-	-	-

Systematic uncertainties (1/2)

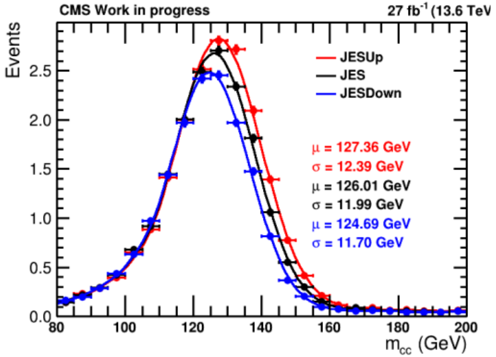
- **Fit bias uncertainty:**
  - related to the choice of the continuum background fitting function
  - assumed a conservative 20% on the signal (spurious signal method)
- **c tagging**:
  - assumed a conservative 10% uncertainty on the yields

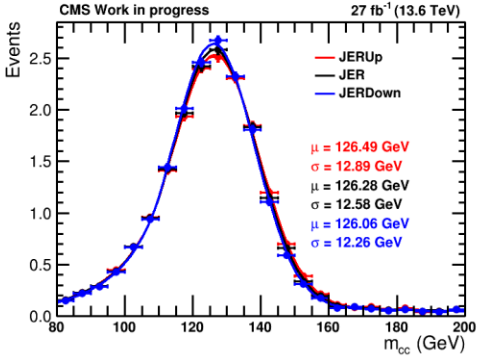
Source of uncertainty	$VBF\,H \to c\bar{c}$	$gg\,H \to c\bar{c}$	$VBF\,H \to b\bar{b}$	$gg\,H \to b\bar{b}$	Z/W+jets
Luminosity	1.4%	1.4%	1.4%	1.4%	1.4%
VBF model	8%	-	- 8%	-	-
Trigger	3%	3%	4%	1%	3%
c tagging	10%	10%	10%	10%	10%
Higgs XS QCD scale	$^{+0.4\%}_{-0.3\%}$	$^{+4.6\%}_{-6.7\%}$	$^{+0.4\%}_{-0.3\%}$	$^{+4.6\%}_{-6.7\%}$	-
Higgs XS PDF + α_s	2.1%	3.2%	2.1%	3.2%	-
Higgs decay BR	$^{+5\%}_{-3\%}$	$^{+5\%}_{-3\%}$	-	-	-

Systematic uncertainties (1/2)

- **Fit bias uncertainty:**
  - related to the choice of the continuum background fitting function
  - assumed a conservative 20% on the signal (spurious signal method)
- **c tagging**:
  - assumed a conservative 10% uncertainty on the yields
- **JES and JER:**
  - affects the shape of the reconstructed Higgs boson candidate mass and is therefore treated as a source of **shape systematic uncertainty**
  - whole analysis reprocessed with up and down variations

Source of uncertainty	$VBF\,H \to c\bar{c}$	$gg\,H \to c\bar{c}$	$VBF\,H \to b\bar{b}$	$gg\,H \to b\bar{b}$	Z/W+jets
Luminosity	1.4%	1.4%	1.4%	1.4%	1.4%
VBF model	8%	-	- 8%	-	-
Trigger	3%	3%	4%	1%	3%
c tagging	10%	10%	10%	10%	10%
Higgs XS QCD scale	$^{+0.4\%}_{-0.3\%}$	$^{+4.6\%}_{-6.7\%}$	$^{+0.4\%}_{-0.3\%}$	$^{+4.6\%}_{-6.7\%}$	-
Higgs XS PDF + α_s	2.1%	3.2%	2.1%	3.2%	-
Higgs decay BR	$^{+5\%}_{-3\%}$	$^{+5\%}_{-3\%}$	-	-	-





Systematic uncertainties (1/2)

- **Fit bias uncertainty:**
  - related to the choice of the continuum background fitting function
  - assumed a conservative 20% on the signal (spurious signal method)
- **c tagging**:
  - assumed a conservative 10% uncertainty on the yields
- **JES and JER:**
  - affects the shape of the reconstructed Higgs boson candidate mass and is therefore treated as a source of shape systematic uncertainty
  - whole analysis reprocessed with up and down variations
- **Parton showering and hadronization model for VBF production:**
  - estimated by comparing the signal acceptance of two generators: PYTHIA and HERWIG

Source of uncertainty	$VBF\ H \to c\bar{c}$	$gg\ H \to c\bar{c}$	$VBF\ H \to b\bar{b}$	$gg\ H \to b\bar{b}$	Z/W+jets
Luminosity	1.4%	1.4%	1.4%	1.4%	1.4%
VBF model	8%	-	- 8%	-	-
Trigger	3%	3%	4%	1%	3%
c tagging	10%	10%	10%	10%	10%
Higgs XS QCD scale	$^{+0.4\%}_{-0.3\%}$	$^{+4.6\%}_{-6.7\%}$	$^{+0.4\%}_{-0.3\%}$	$^{+4.6\%}_{-6.7\%}$	-
Higgs XS PDF + α_s	2.1%	3.2%	2.1%	3.2%	-
Higgs decay BR	$^{+5\%}_{-3\%}$	$^{+5\%}_{-3\%}$	-	-	-

Systematic uncertainties (1/2)

- **Fit bias uncertainty:**
  - related to the choice of the continuum background fitting function
  - assumed a conservative 20% on the signal (spurious signal method)
- **c tagging**:
  - assumed a conservative 10% uncertainty on the yields
- **JES and JER:**
  - affects the shape of the reconstructed Higgs boson candidate mass and is therefore treated as a source of shape systematic uncertainty
  - whole analysis reprocessed with up and down variations
- **Parton showering and hadronization model for VBF production:**
  - estimated by comparing the signal acceptance of two generators: PYTHIA and HERWIG
- **Trigger:**
  - estimated with trigger SFs up and down variations

Source of uncertainty	$VBF\,H \to c\bar{c}$	$gg\,H \to c\bar{c}$	$VBF\,H \to b\bar{b}$	$gg\,H \to b\bar{b}$	Z/W+jets
Luminosity	1.4%	1.4%	1.4%	1.4%	1.4%
VBF model	8%	-	- 8%	-	-
Trigger	3%	3%	4%	1%	3%
c tagging	10%	10%	10%	10%	10%
Higgs XS QCD scale	$^{+0.4\%}_{-0.3\%}$	$^{+4.6\%}_{-6.7\%}$	$^{+0.4\%}_{-0.3\%}$	$^{+4.6\%}_{-6.7\%}$	-
Higgs XS PDF + α_s	2.1%	3.2%	2.1%	3.2%	-
Higgs decay BR	$^{+5\%}_{-3\%}$	$^{+5\%}_{-3\%}$	-	-	-

Systematic uncertainties (1/2)

- **Fit bias uncertainty:**
  - related to the choice of the continuum background fitting function
  - assumed a conservative 20% on the signal (spurious signal method)
- **c tagging**:
  - assumed a conservative 10% uncertainty on the yields
- **JES and JER:**
  - affects the shape of the reconstructed Higgs boson candidate mass and is therefore treated as a source of shape systematic uncertainty
  - whole analysis reprocessed with up and down variations
- **Parton showering and hadronization model for VBF production:**
  - estimated by comparing the signal acceptance of two generators: PYTHIA and HERWIG
- **Trigger:**
  - estimated with trigger SFs up and down variations
- **Integrated luminosity**

Source of uncertainty	$VBF\,H \to c\bar{c}$	$gg\,H \to c\bar{c}$	$VBF\,H \to b\bar{b}$	$gg\,H \to b\bar{b}$	Z/W+jets
Luminosity	1.4%	1.4%	1.4%	1.4%	1.4%
VBF model	8%	-	- 8%	-	-
Trigger	3%	3%	4%	1%	3%
c tagging	10%	10%	10%	10%	10%
Higgs XS QCD scale	$^{+0.4\%}_{-0.3\%}$	$^{+4.6\%}_{-6.7\%}$	$^{+0.4\%}_{-0.3\%}$	$^{+4.6\%}_{-6.7\%}$	-
Higgs XS PDF + α_s	2.1%	3.2%	2.1%	3.2%	-
Higgs decay BR	$^{+5\%}_{-3\%}$	$^{+5\%}_{-3\%}$	-	-	-

A. Zaza

Systematic uncertainties (2/2)

- **Theoretical uncertainty on Higgs production XS:**
  - uncertainty arising from
    - approximations used in perturbative calculations of QCD
    - uncertainty on the parton distribution functions (PDFs)
    - uncertainty on α_S

Source of uncertainty	$VBF\,H \to c\bar{c}$	$gg\,H \to c\bar{c}$	$VBF\,H \to b\bar{b}$	$gg\,H \to b\bar{b}$	Z/W+jets
Luminosity	1.4%	1.4%	1.4%	1.4%	1.4%
VBF model	8%	-	- 8%	-	-
Trigger	3%	3%	4%	1%	3%
c tagging	10%	10%	10%	10%	10%
Higgs XS QCD scale	$^{+0.4\%}_{-0.3\%}$	$^{+4.6\%}_{-6.7\%}$	$^{+0.4\%}_{-0.3\%}$	$^{+4.6\%}_{-6.7\%}$	-
Higgs XS PDF + α_s	2.1%	3.2%	2.1%	3.2%	-
Higgs decay BR	$^{+5\%}_{-3\%}$	$^{+5\%}_{-3\%}$	-	-	-

Systematic uncertainties (2/2)

- **Theoretical uncertainty on Higgs production XS:**
  - uncertainty arising from
    - approximations used in perturbative calculations of QCD
    - uncertainty on the parton distribution functions (PDFs)
    - uncertainty on α_S
- **Theoretical uncertainty on $H \rightarrow c\bar{c}$ decay BR:**
  - uncertainty arising from
    - higher order QCD and electroweak corrections considered in theoretical calculation
    - uncertainty on the c quark mass
    - uncertainty on α_S

Source of uncertainty	$VBF\ H \rightarrow c\bar{c}$	$gg\ H \rightarrow c\bar{c}$	$VBF\ H \rightarrow b\bar{b}$	$gg\ H \rightarrow b\bar{b}$	Z/W+jets
Luminosity	1.4%	1.4%	1.4%	1.4%	1.4%
VBF model	8%	-	- 8%	-	-
Trigger	3%	3%	4%	1%	3%
c tagging	10%	10%	10%	10%	10%
Higgs XS QCD scale	$^{+0.4\%}_{-0.3\%}$	$^{+4.6\%}_{-6.7\%}$	$^{+0.4\%}_{-0.3\%}$	$^{+4.6\%}_{-6.7\%}$	-
Higgs XS PDF + α_s	2.1%	3.2%	2.1%	3.2%	-
Higgs decay BR	$^{+5\%}_{-3\%}$	$^{+5\%}_{-3\%}$	-	-	-

Results

Final result: expected upper limit on the signal strength μ at 95% CL with 27 fb^{-1}

- estimated with **CLs method** by giving as input
  - the parametric shapes modelling the signal and the background
  - the expected yields estimated for signal and peaking backgrounds from MC simulation and for the QCD multijet background from the fit to data sidebands
- most relevant systematic uncertainties taken into account

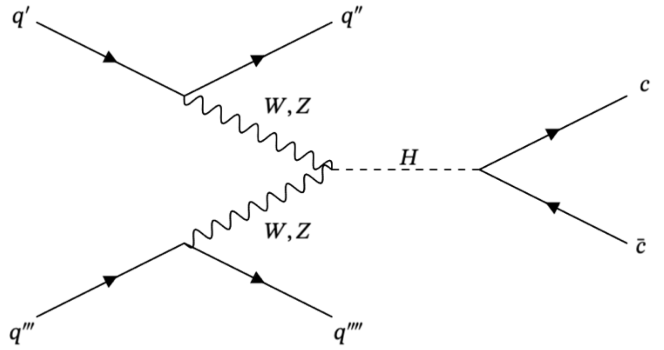
$$\mu = \frac{\sigma(VBF) \cdot B(H \to c\bar{c})}{\sigma(VBF)_{SM} \cdot B(H \to c\bar{c})_{SM}} < 30.87 \quad @95\%\ CL$$

Category	Upper Limit
CAT0	75.88
CAT1	60.00
CAT2	40.25
Combination	30.87

Conclusions

- **VBF** $H \to c\bar{c}$ investigated for the **first time** at CMS
- Developed **new trigger** specific for this search
  - first trigger with c-tagging online
  - one of the first triggers with **ParticleNet**
  - deployed online for 2023 data-taking
- Implemented strategy to analyse 2023 data (27 fb^{-1})
  - offline preselection similar to trigger
  - BDT for signal vs QCD background discrimination
  - signal and background modelling
  - statistical analysis for **expected upper limit** extraction
- Results

$$\frac{\sigma(VBF)\cdot B(H\to c\bar{c})}{\sigma(VBF)_{SM}\cdot B(H\to c\bar{c})_{SM}} < 30.87 \quad @95\%\ CL$$



Projections

- Projection full **Run-3**
  - expected L = **360 fb^{-1}** to be collected by CMS in 2023-2026

$$\frac{\sigma(VBF)\cdot B(H\rightarrow c\bar{c})}{\sigma(VBF)_{SM}\cdot B(H\rightarrow c\bar{c})_{SM}} < 8 \qquad @95\%\ CL$$

Competitive with VH results!

- Projection **HL-LHC**
  - expected L = **3000 fb^{-1}** to be collected by CMS until 2041

$$\frac{\sigma(VBF)\cdot B(H\rightarrow c\bar{c})}{\sigma(VBF)_{SM}\cdot B(H\rightarrow c\bar{c})_{SM}} < 3 \qquad @95\%\ CL$$

  - reasonable to assume a $\sqrt{3}$ improvement from VH, VBF and ttH combination

$$\frac{\sigma\cdot B(H\rightarrow c\bar{c})}{\sigma_{SM}\cdot B(H\rightarrow c\bar{c})_{SM}} < 2 \qquad @95\%\ CL$$

Further improvements on flavour tagging performance and analysis techniques expected in the next future

Backup

angela.zaza@cern.ch

Standard Model

$$\mathcal{L}_{SM} = -\frac{1}{4}G_{\lambda\rho}G^{\lambda\rho} - \frac{1}{4}W_{\lambda\rho}W^{\lambda\rho} - \frac{1}{4}B_{\lambda\rho}B^{\lambda\rho} + \bar{\psi}i\gamma_\lambda D^\lambda \psi$$
$$+(D_\lambda\phi)^\dagger(D^\lambda\phi) - V(\phi)$$
$$+\mathcal{L}_{Yukawa} + h.c.$$

Symmetry group:
SU(3) × SU(2) × U(1)

$$D_\lambda = \partial_\lambda + ig_s G_\lambda + igW_\lambda + ig' B_\lambda Y$$

Interaction between Gauge fields

Higgs term

Interaction between fermions and the Higgs field

$$\mathcal{L}_{Yukawa,l1} = -c_e \bar{e}_R \phi^\dagger \begin{pmatrix} \upsilon_{eL} \\ e_L \end{pmatrix} + h.c. \qquad m_e = -c_e \frac{\rho_0}{\sqrt{2}}$$

electrons

$$\mathcal{L}_{Yukawa,q1} = -c'_q \bar{u}_R \phi^T \begin{pmatrix} u \\ d' \end{pmatrix}_L - c_q \bar{d'}_R \phi^\dagger \begin{pmatrix} u \\ d' \end{pmatrix}_L + h.c.$$

First-generation quarks

BSM theory - EFT

Basic premise of EFT: dynamics at **low energies** does not depend on the details of the dynamic at **high energies**

→ low energy physics can be described using an effective Lagrangian that contains only a few degrees of freedom, ignoring additional degrees of freedom present at higher energies

$$S = \int \mathrm{d}^{\mathrm{d}}x \; \mathscr{L}(x)$$

In d spacetime dimensions, the Lagrangian density has dimension d

$$\Rightarrow [\mathscr{L}(x)] = \mathrm{d}$$

$$\mathscr{L}(x) = \sum_i c_i \, O_i(x)$$

O_i: local, gauge invariant and Lorentz invariant operators

The operator dimension is denoted by $\mathcal{D}$, and its coefficient has dimension d-$\mathcal{D}$

BSM theory - EFT

$$\mathscr{L}_{\mathrm{EFT}} = \sum_{\mathscr{D}\geq 0,i} \frac{c_i^{(\mathscr{D})} O_i^{(\mathscr{D})}}{\Lambda^{\mathscr{D}-d}} = \sum_{\mathscr{D}\geq 0} \frac{\mathscr{L}_{\mathscr{D}}}{\Lambda^{\mathscr{D}-d}} \Longrightarrow \mathscr{L}_{\mathrm{EFT}} = \mathscr{L}_{\mathscr{D}\leq 4} + \frac{\mathscr{L}_5}{\Lambda} + \frac{\mathscr{L}_6}{\Lambda^2} + \ldots$$

Λ: scale introduced so that the coefficients $c_i^{(\mathcal{D})}$ are dimensionless (scale at which new physics occurs)

Standard Model EFT (SMEFT)

$$\boxed{\mathcal{L}_{\mathrm{eff}} = \mathcal{L}_{\mathrm{SM}} + \sum_i \frac{c_i^{(5)}}{\Lambda}\mathcal{O}_i^{(5)} + \sum_i \frac{c_i^{(6)}}{\Lambda^2}\mathcal{O}_i^{(6)} + \sum_i \frac{c_i^{(7)}}{\Lambda^3}\mathcal{O}_i^{(7)} + \sum_i \frac{c_i^{(8)}}{\Lambda^4}\mathcal{O}_i^{(8)} + \cdots,}$$

- O_i^D: SU(3) × SU(2) × U(1) invariant operator of dimension D
- c_i^D: **Wilson coefficients**

BSM theory - EFT

- Switch from the k-parametrization to the one given in terms of Wilson coefficients
- It is possible to identify a one way mapping between the { k_i } and { C_i }

$k_i^2 = 1 + \Delta k_i$, with Δk_i a linear combination of EFT parameters

Example: $h \to bb$

$$\kappa_b^2 = \frac{\mathcal{A}^2(h \to \bar{b}b)_{\mathrm{SMEFT}}}{\mathcal{A}^2(h \to \bar{b}b)_{\mathrm{SM}}} = 1 + \Delta\kappa_b \,,$$

$$\Delta\kappa_b = 2\,\bar{v}_T^2 \left(C_{H\Box} - \frac{C_{HD}}{4} - C_{Hl}^{(3)} + \frac{C'_{ll}}{2} - \frac{C_{dH}}{[Y_d]_{33}} \right)$$

Only direct d=6 contribution, which is due to the operator Q_{dH}, that perturbs the Yukawa coupling

BSM theory – Two Higgs Doublet Model

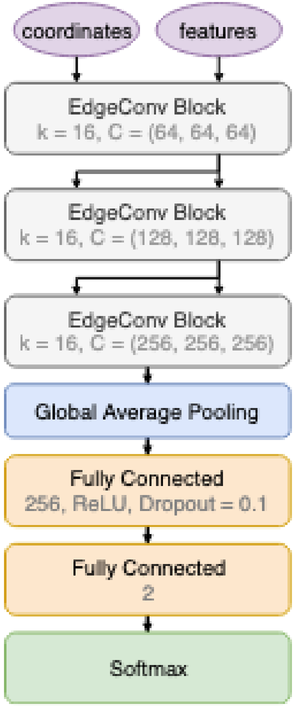
- Model within the Spontaneous Flavour Violation (SFV) framework
- The Higgs sector is extended with an additional doublet

Results of this study

- Possible to have a second Higgs with Yukawa couplings of O(10^{-1}) with any up-type quark

- For instance, it is possible to have a **new Higgs** that **couples at this strength only to the charm quarks**
  When this new Higgs has a non-zero mixing angle with the SM Higgs, this allows for large **deviations of O(10) to O(10^4) to the SM** charm and up Yukawa couplings consistent with measurements of flavour changing neutral currents (FCNC)

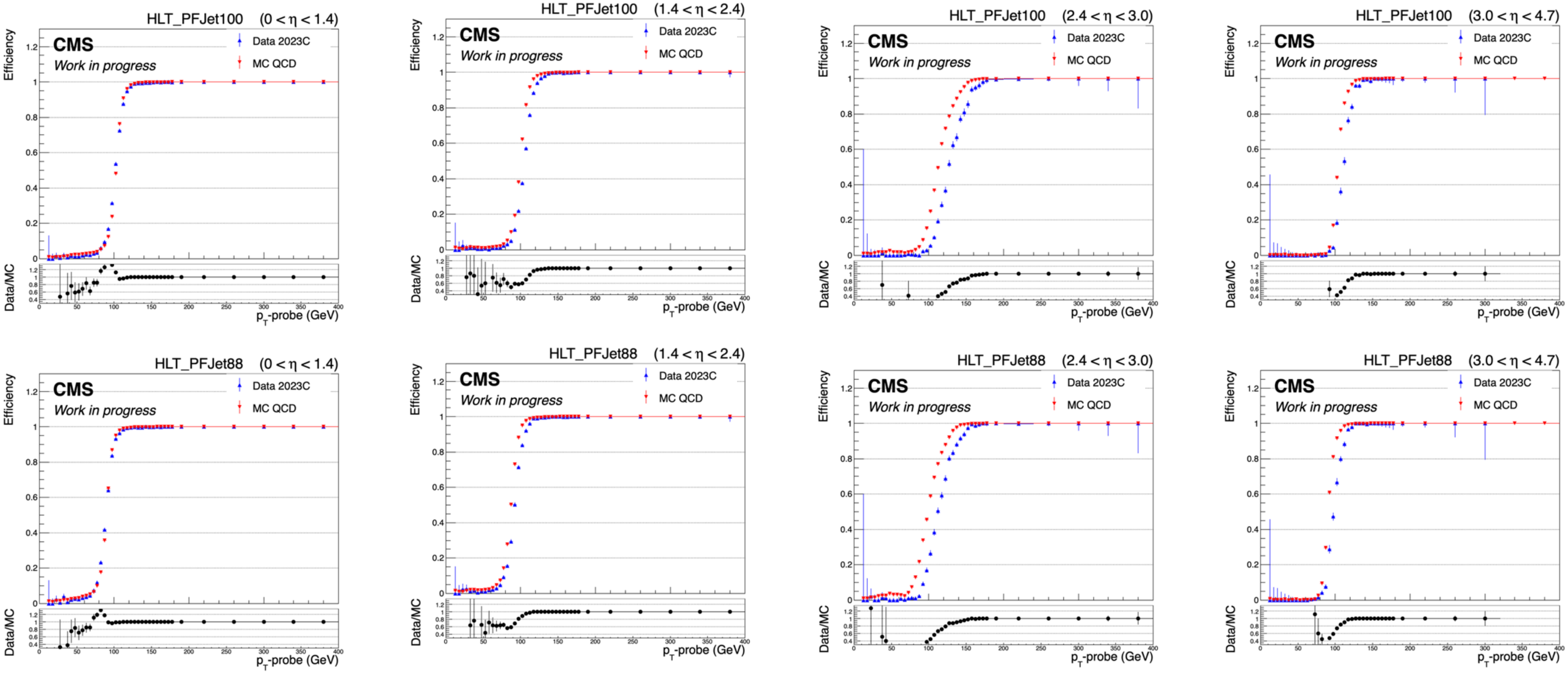
ParticleNet - architecture



(a) ParticleNet

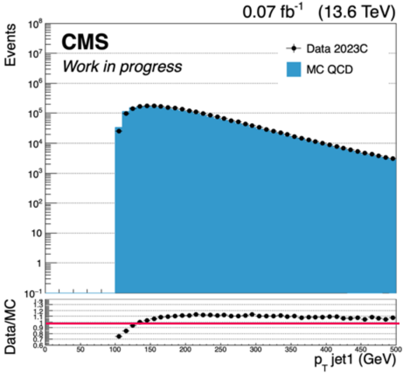
- Three EdgeConv blocks allow the model to learn hierarchical jet substructures
  - for each particle the EdgeConv block identifies the k nearest neighboring particles and builds *edges* (reletionships between nearby particles)
- The learned features from all the particles are combined by a global average pooling operation, followed by a fully connected layer
- Finally, a fully connected layer with two units and a softmax function provide the output for a binary classification task (PNet was first evaluated on two jet tagging benchmarks)

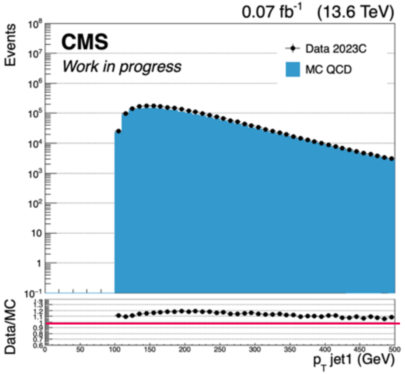
Trigger SFs

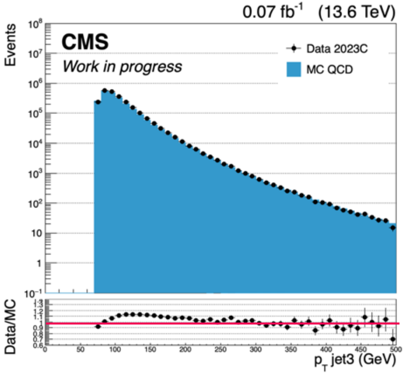

HLT_PFJet100 $(0 < \eta < 1.4)$
HLT_PFJet100 $(1.4 < \eta < 2.4)$
HLT_PFJet100 $(2.4 < \eta < 3.0)$
HLT_PFJet100 $(3.0 < \eta < 4.7)$
HLT_PFJet88 $(0 < \eta < 1.4)$
HLT_PFJet88 $(1.4 < \eta < 2.4)$
HLT_PFJet88 $(2.4 < \eta < 3.0)$
HLT_PFJet88 $(3.0 < \eta < 4.7)$
CMS
Work in progress
Data 2023C
MC QCD
Efficiency
Data/MC
p_T-probe (GeV)

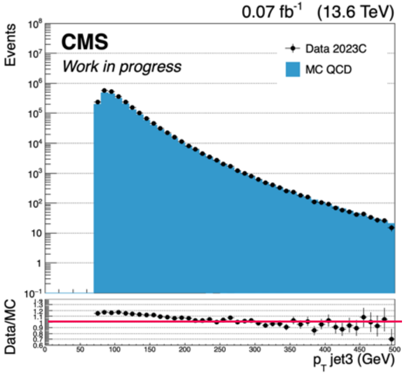
Validation plots for HLT p_T SFs

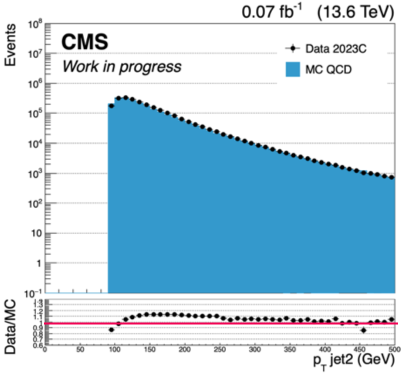
Before SF | After SF | Before SF | After SF

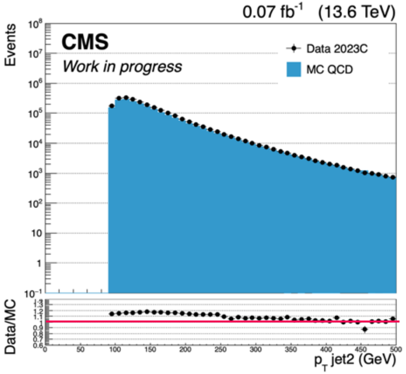


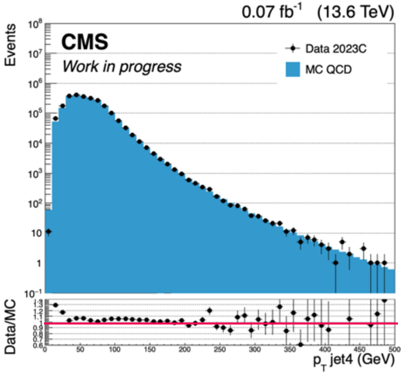


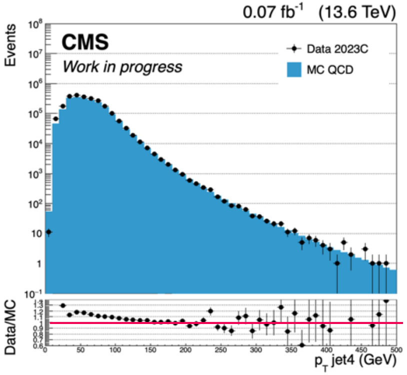






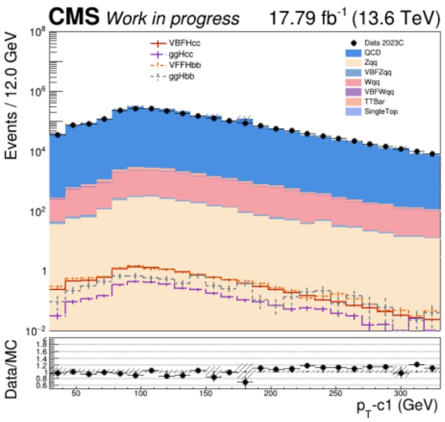


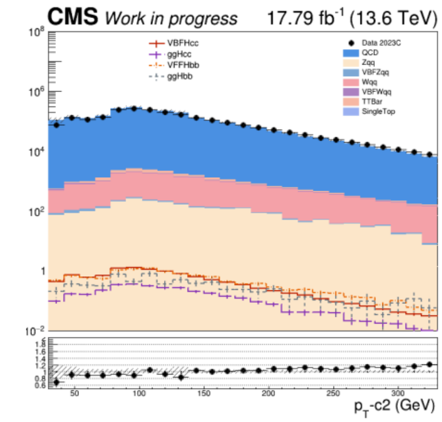


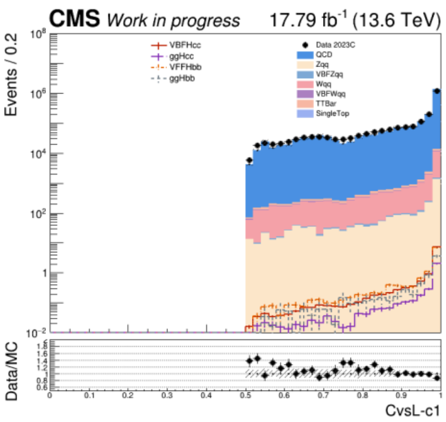


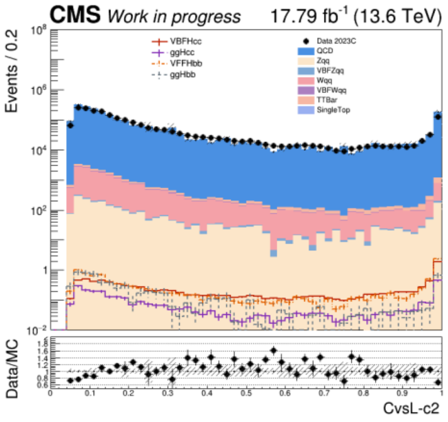
→ Trigger p_T SFs correct the mismatch between data and MC

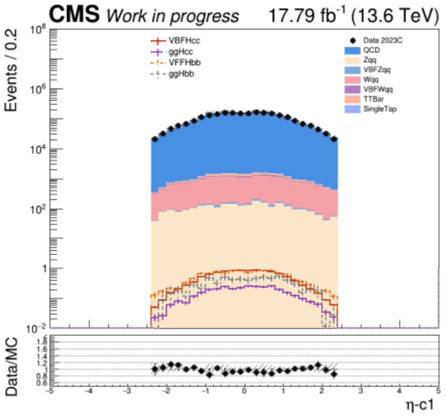
Plot Data/MC

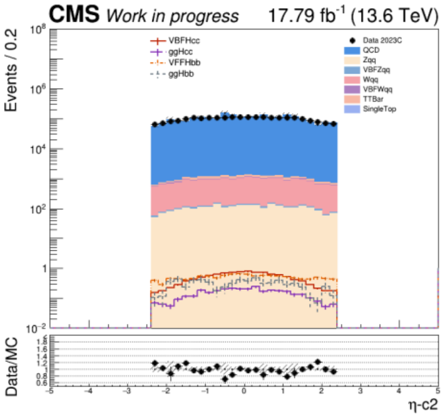

CMS Work in progress
17.79 fb-1 (13.6 TeV)
VBFHcc
ggHcc
VFFHbb
ggHbb
Data 2023C
QCD
Zqq
VBFZqq
Wqq
VBFWqq
TTBar
SingleTop
Events / 12.0 GeV
Data/MC
p_T-c1 (GeV)

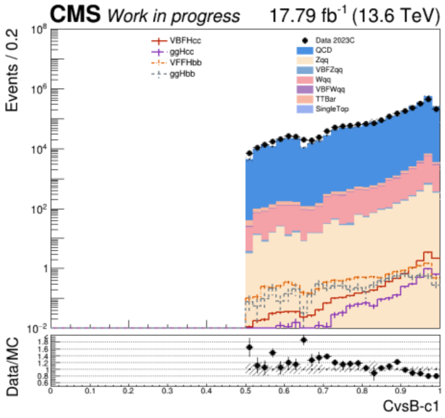

CMS Work in progress
17.79 fb-1 (13.6 TeV)
VBFHcc
ggHcc
VFFHbb
ggHbb
Data 2023C
QCD
Zqq
VBFZqq
Wqq
VBFWqq
TTBar
SingleTop
Data/MC
p_T-c2 (GeV)

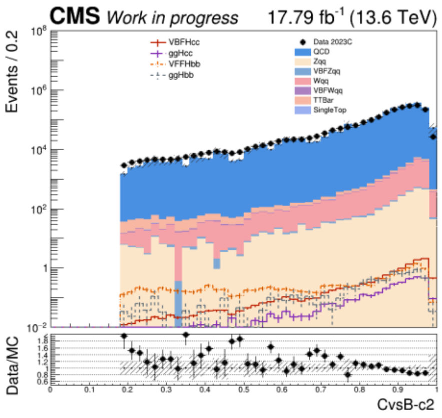

CMS Work in progress
17.79 fb-1 (13.6 TeV)
VBFHcc
ggHcc
VFFHbb
ggHbb
Data 2023C
QCD
Zqq
VBFZqq
Wqq
VBFWqq
TTBar
SingleTop
Events / 0.2
Data/MC
CvsL-c1


CMS Work in progress
17.79 fb-1 (13.6 TeV)
VBFHcc
ggHcc
VFFHbb
ggHbb
Data 2023C
QCD
Zqq
VBFZqq
Wqq
VBFWqq
TTBar
SingleTop
Events / 0.2
Data/MC
CvsL-c2

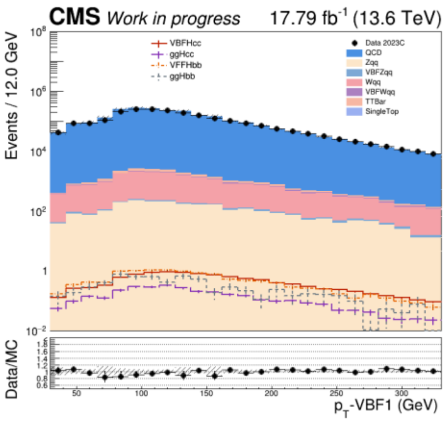

CMS Work in progress
17.79 fb-1 (13.6 TeV)
VBFHcc
ggHcc
VFFHbb
ggHbb
Data 2023C
QCD
Zqq
VBFZqq
Wqq
VBFWqq
TTBar
SingleTop
Events / 0.2
Data/MC
η-c1


CMS Work in progress
17.79 fb-1 (13.6 TeV)
VBFHcc
ggHcc
VFFHbb
ggHbb
Data 2023C
QCD
Zqq
VBFZqq
Wqq
VBFWqq
TTBar
SingleTop
Events / 0.2
Data/MC
η-c2

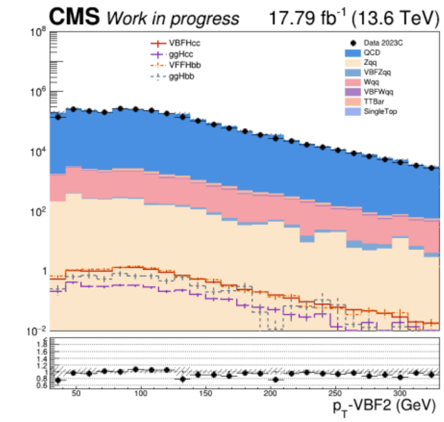

CMS Work in progress
17.79 fb-1 (13.6 TeV)
VBFHcc
ggHcc
VFFHbb
ggHbb
Data 2023C
QCD
Zqq
VBFZqq
Wqq
VBFWqq
TTBar
SingleTop
Events / 0.2
Data/MC
CvsB-c1


CMS Work in progress
17.79 fb-1 (13.6 TeV)
VBFHcc
ggHcc
VFFHbb
ggHbb
Data 2023C
QCD
Zqq
VBFZqq
Wqq
VBFWqq
TTBar
SingleTop
Events / 0.2
Data/MC
CvsB-c2

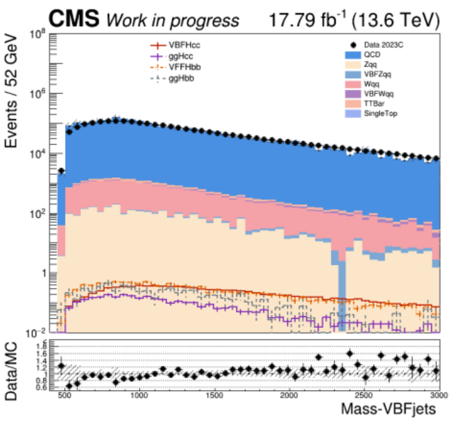
Plot Data/MC



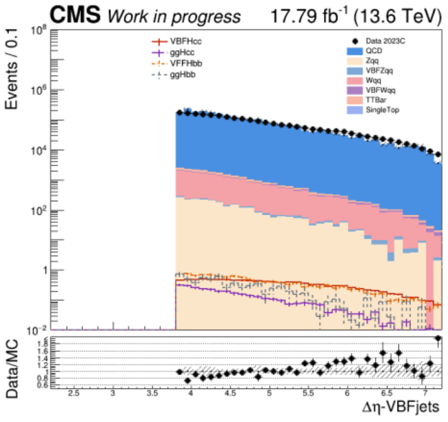
CMS Work in progress
17.79 fb^{-1} (13.6 TeV)
VBFHcc
ggHcc
VFFHbb
ggHbb
Data 2023C
QCD
Zqq
VBFZqq
Wqq
VBFWqq
TTBar
SingleTop
Events / 12.0 GeV
Data/MC
p_T-VBF1 (GeV)



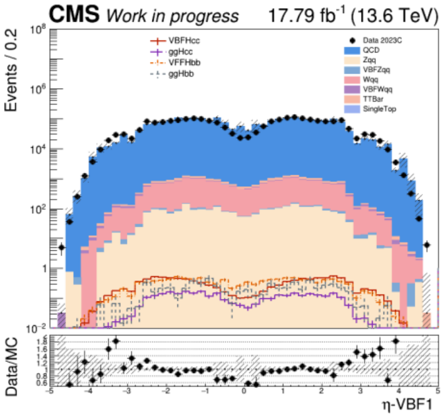
CMS Work in progress
17.79 fb^{-1} (13.6 TeV)
VBFHcc
ggHcc
VFFHbb
ggHbb
Data 2023C
QCD
Zqq
VBFZqq
Wqq
VBFWqq
TTBar
SingleTop
Data/MC
p_T-VBF2 (GeV)



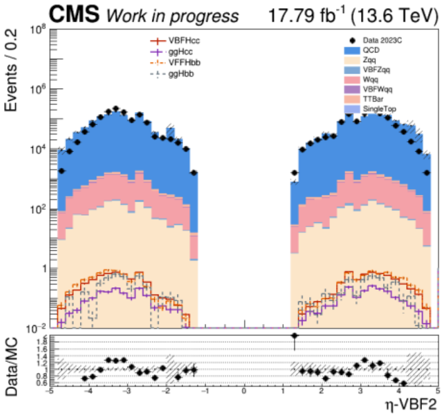
CMS Work in progress
17.79 fb^{-1} (13.6 TeV)
VBFHcc
ggHcc
VFFHbb
ggHbb
Data 2023C
QCD
Zqq
VBFZqq
Wqq
VBFWqq
TTBar
SingleTop
Events / 52 GeV
Data/MC
Mass-VBFjets



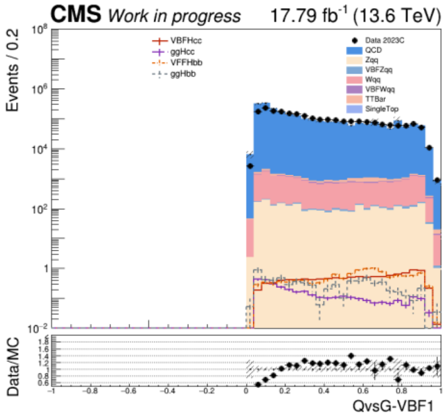
CMS Work in progress
17.79 fb^{-1} (13.6 TeV)
VBFHcc
ggHcc
VFFHbb
ggHbb
Data 2023C
QCD
Zqq
VBFZqq
Wqq
VBFWqq
TTBar
SingleTop
Events / 0.1
Data/MC
$\Delta\eta$-VBFjets



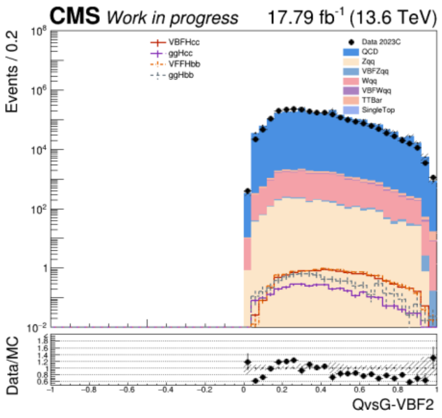
CMS Work in progress
17.79 fb^{-1} (13.6 TeV)
VBFHcc
ggHcc
VFFHbb
ggHbb
Data 2023C
QCD
Zqq
VBFZqq
Wqq
VBFWqq
TTBar
SingleTop
Events / 0.2
Data/MC
η-VBF1



CMS Work in progress
17.79 fb^{-1} (13.6 TeV)
VBFHcc
ggHcc
VFFHbb
ggHbb
Data 2023C
QCD
Zqq
VBFZqq
Wqq
VBFWqq
TTBar
SingleTop
Events / 0.2
Data/MC
η-VBF2

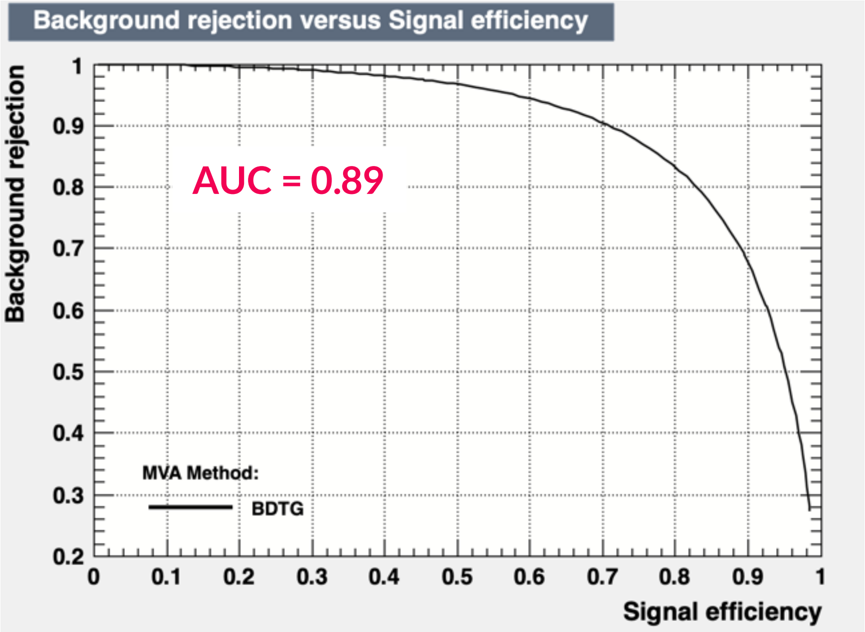


CMS Work in progress
17.79 fb^{-1} (13.6 TeV)
VBFHcc
ggHcc
VFFHbb
ggHbb
Data 2023C
QCD
Zqq
VBFZqq
Wqq
VBFWqq
TTBar
SingleTop
Events / 0.2
Data/MC
QvsG-VBF1



CMS Work in progress
17.79 fb^{-1} (13.6 TeV)
VBFHcc
ggHcc
VFFHbb
ggHbb
Data 2023C
QCD
Zqq
VBFZqq
Wqq
VBFWqq
TTBar
SingleTop
Events / 0.2
Data/MC
QvsG-VBF2

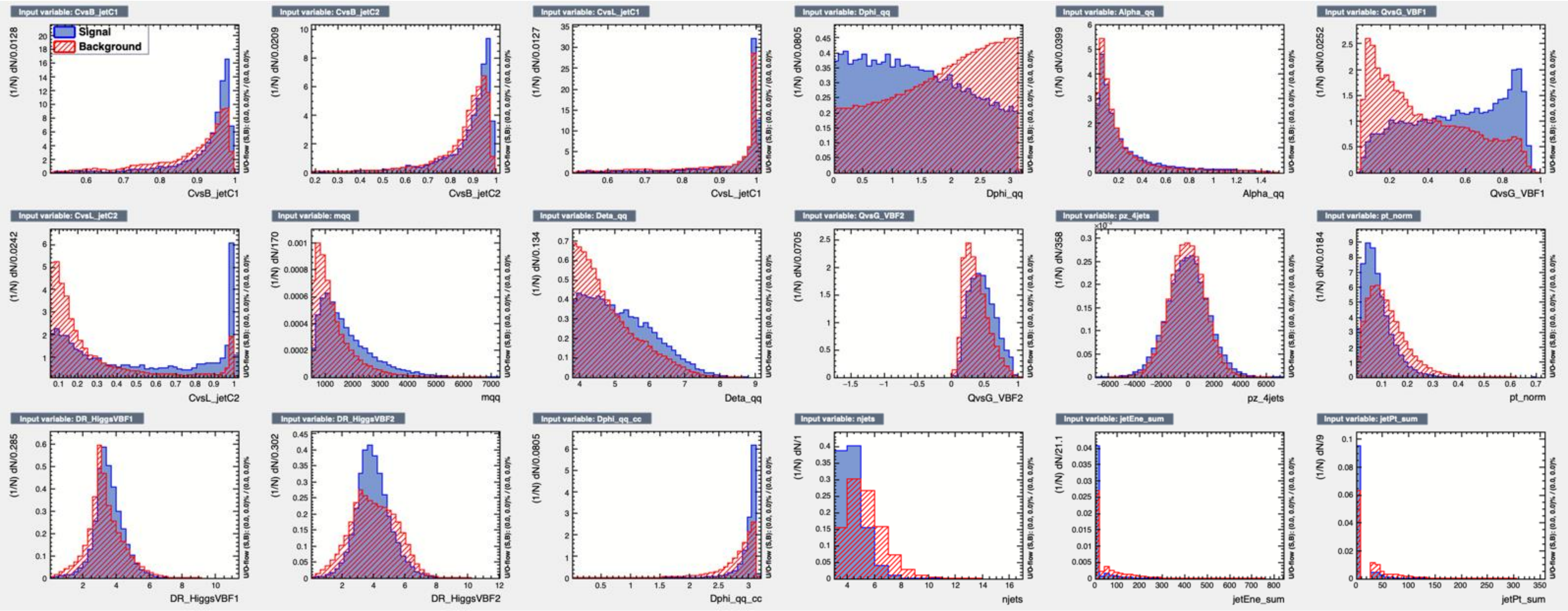
BDT for QCD bkg mitigation



Feature importance

Rank	Variable	Variable Importance
1	DR_HiggsVBF2	6.892e-02
2	QvsG_VBF1	6.812e-02
3	Dphi_qq_cc	6.620e-02
4	mqq	6.617e-02
5	CvsL_jetC2	6.540e-02
6	Dphi_qq	6.101e-02
7	CvsB_jetC1	6.093e-02
8	CvsB_jetC2	6.082e-02
9	Deta_qq	6.061e-02
10	DR_HiggsVBF1	6.022e-02
11	CvsL_jetC1	5.234e-02
12	jetPt_sum	5.184e-02
13	Alpha_qq	5.115e-02
14	njets	4.943e-02
15	pt_norm	4.683e-02
16	pz_4jets	4.195e-02
17	QvsG_VBF2	3.959e-02
18	jetEne_sum	2.846e-02

H $\rightarrow c\bar{c}$ vs QCD discriminator



A. Zaza

UL estimation

Test statistic

$$q_\mu = -2\ln\lambda(\mu)$$

Likelihood ratio

$$\lambda(\mu) = \frac{L(\mu,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})}$$

μ: signal strength

L : likelihood function

θ: nuisance parameters

Counting experiment

$$L(\mu,\theta) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)}$$

p-value: probability of observing a disagreement with the null hypothesis as large as the one observed in data

$$p_\mu = \int_{q_{\mu,\mathrm{obs}}}^{\infty} f(q_\mu,\mu)\,dq_\mu$$

$f(q_\mu,\mu)$: pdf of q_μ under the assumption of the signal strength μ

For the calculation of the UL, the modified test statistic for upper limit is used

$$\tilde{q}_\mu = \begin{cases} -2\ln\frac{L(\mu,\hat{\hat{\theta}})}{L(0,\hat{\hat{\theta}}(0))} & \text{if } \hat{\mu} < 0, \\ -2\ln\frac{L(\mu,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})} & \text{if } 0 \le \hat{\mu} \le \mu, \\ 0 & \text{if } \hat{\mu} > \mu. \end{cases}$$

This way:

- the data is assumed to show lack of agreement with the hypothesized μ only if $\hat{\mu} < \mu$
- If $\hat{\mu}$ <0 (number of data events smaller than the one expected from background only because of statistical fluctuations), the ML value of μ is set to 0

UL estimation

The Upper Limit is the largest value of μ such that the p-value is larger than or equal to a fixed threshold (0.05 for 95% CL)

$$p_\mu = \int_{q_{\mu,\mathrm{obs}}}^{\infty} f(q_\mu, \mu) dq_\mu < 1 - \mathrm{CL}$$

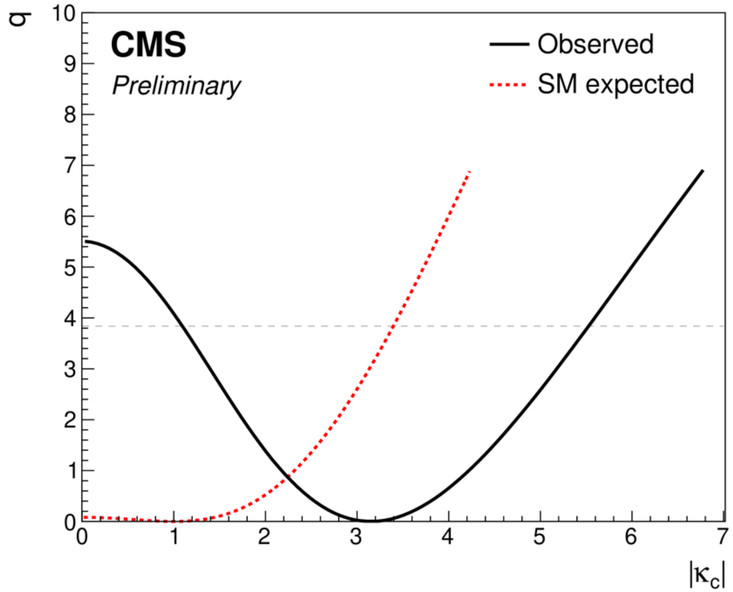
CLs method:

$$\frac{p_\mu}{1 - p_b} < 1 - CL, \text{with } p_b = \int_{q_0}^{\infty} f(q_0|\mu = 0) dq_0$$

The CLs method is used to prevent aggressive exclusion of signal hypotheses, especially in cases where data shows fewer events than expected from bkg alone.
In such cases, 1-p_b becomes small

Kc constraints – Likelihood fit

VHcc



ttHcc – VHcc combined