Inclusive $H \to b\bar{b}/c\bar{c}$ with Run 2 data from LHCb

for questions/comments: davide.zuliani@cern.ch

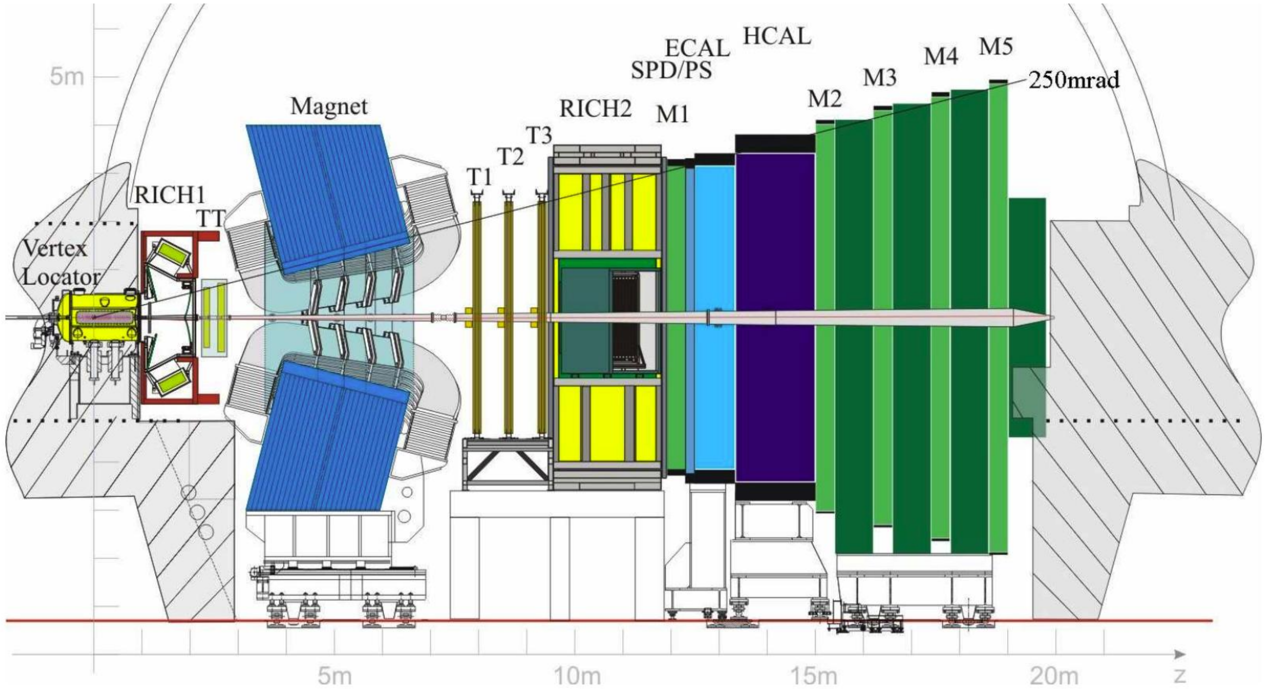
Menglin Xu
CERN
on behalf of the LHCb Collaboration

EPS, 7 July 2025, Marseille

The LHCb Run 2 Detector

- Small angle spectrometer, covering the forward region, $2 < \eta < 5$
- The detector has high-precision tracking system
- Excellent track momentum resolution, $\Delta p/p \sim 0.5\%$
- Excellent vertex recontruction

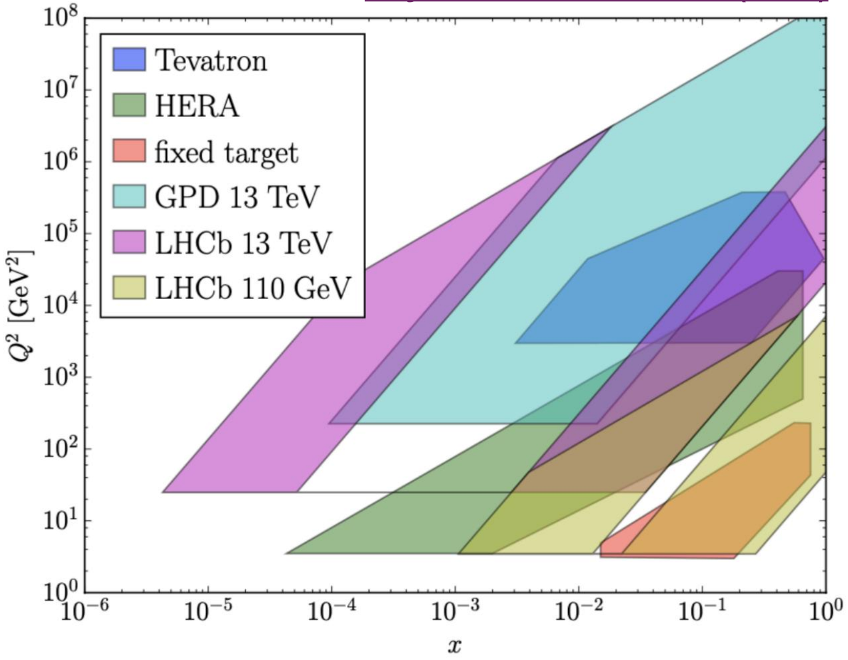
JINST 3 (2008) S08005

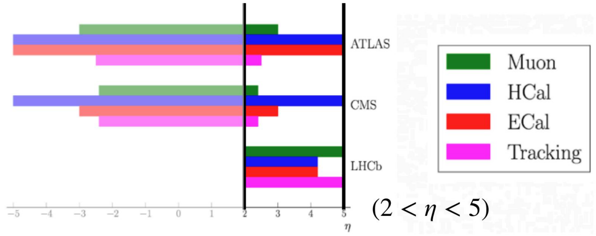


Why LHCb

- Experimental strengths of LHCb
  - Lower pile-up: cleaner events
  - Low energy/momentum threshold triggers
  - Excellent secondary vertex (SV) reconstruction performance helps with *b*- and *c*-jet tagging
- Unique physics research in the forward region
  - LHCb enables tests of perturbative QCD prediction in the unique region, complementary to ATLAS and CMS
  - PDFs and proton structure can be studied in regions not accessible to other LHC experiments

Phys. Rev. D 93, 074008 (2016)





Higgs Search @ LHCb

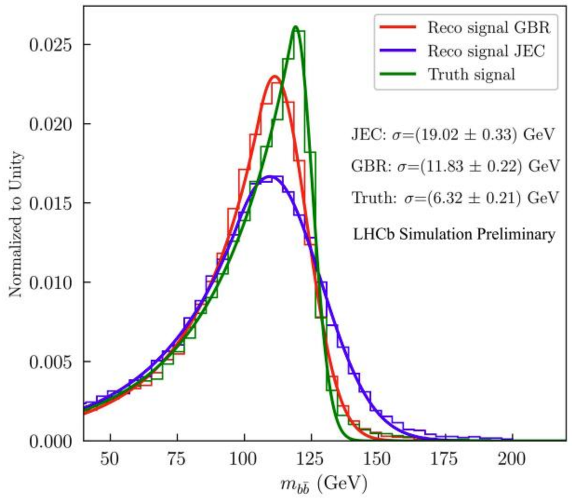
- LHCb Run1: search for $b\overline{b}$ and $c\overline{c}$ + lepton signatures with 2 fb^{-1} dataset LHCb-CONF-2016-006
  - no signals were observed, upper limits on Yukawa couplings: $y^b < 7\,y^b_{\text{SM}},\ y^c < 80\,y^c_{\text{SM}}$
- Pathfinder: search for $H \to b\overline{b}$ and $H \to c\overline{c}$ in the inclusive final state from pp collisions at 13 TeV, using 2016 data, $\mathcal{L} = 1.6\ \text{fb}^{-1}$ LHCb-PAPER-2025-034 (in preparation)
- A model-independent approach, the selection is independent from the Higgs production mechanism
  - just two jets are required, no requirements are applied to additional objects, e.g. μ
- New techniques have been used to
  - correct reconstructed jet energy
    - search based on a fit to invariant mass
  - identify jets flavour

Regression Technique for Jet Energy Correction

- b- and c-jets might be reconstructed differently, w.r.t light jets
  - semileptonic decay of b and c hadrons with undetected neutrinos
  - leading to jet energy underestimation and degraded energy resolution

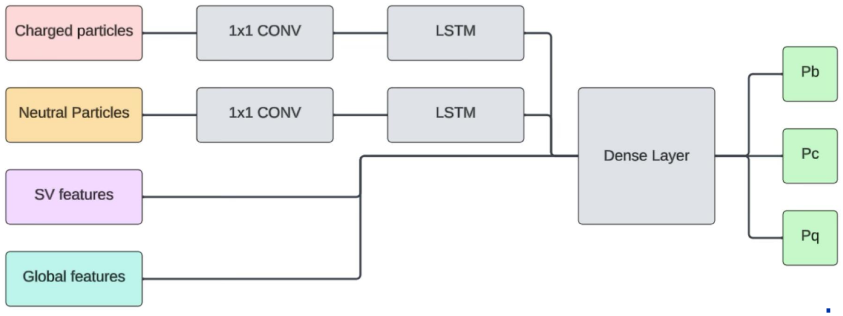
- Run1 measurement used the standard Jet Energy Correction (JEC) with $k = \frac{E(\mathrm{jet_{truth}})}{E(\mathrm{jet_{reco}})}$
- New method: Gradient Boosting Regressor (GBR)
  - inputs: 51 observables from the jet kinematic and substructure
  - ~50% improvement on the Higgs invariant mass resolution
  - ~22% improvement in the Higgs signal significance

LHCb-FIGURE-2023-029



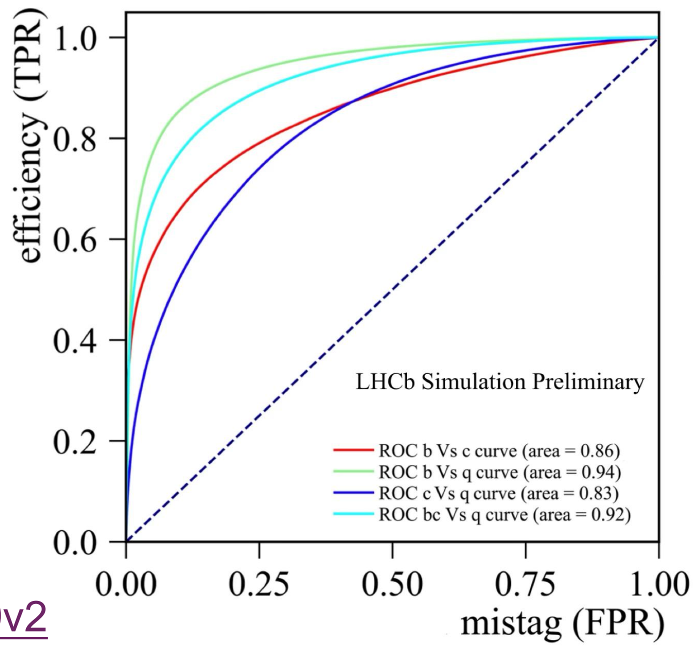
Deep Neural Network for Jet Identification

- In the previous study, jet tagging relies on SV identification and BDTs to distinguish jet flavours
  - by reconstructing secondary vertices from significantly displaced tracks inside and outside the jet
- New approach: Deep Neural Network (DNN)
  - ~400 jet observables are used as input, related to jet constituents and sub-structure
  - SV is not strictly required
  - trained with $b\overline{b}$, $c\overline{c}$ and light dijets simulation
  - 3 outputs: b-, c- and light- probability, P_b, P_c and P_q



inspired by CMS DeepJet, arXiv:2008.10519v2

LHCb-FIGURE-2023-029

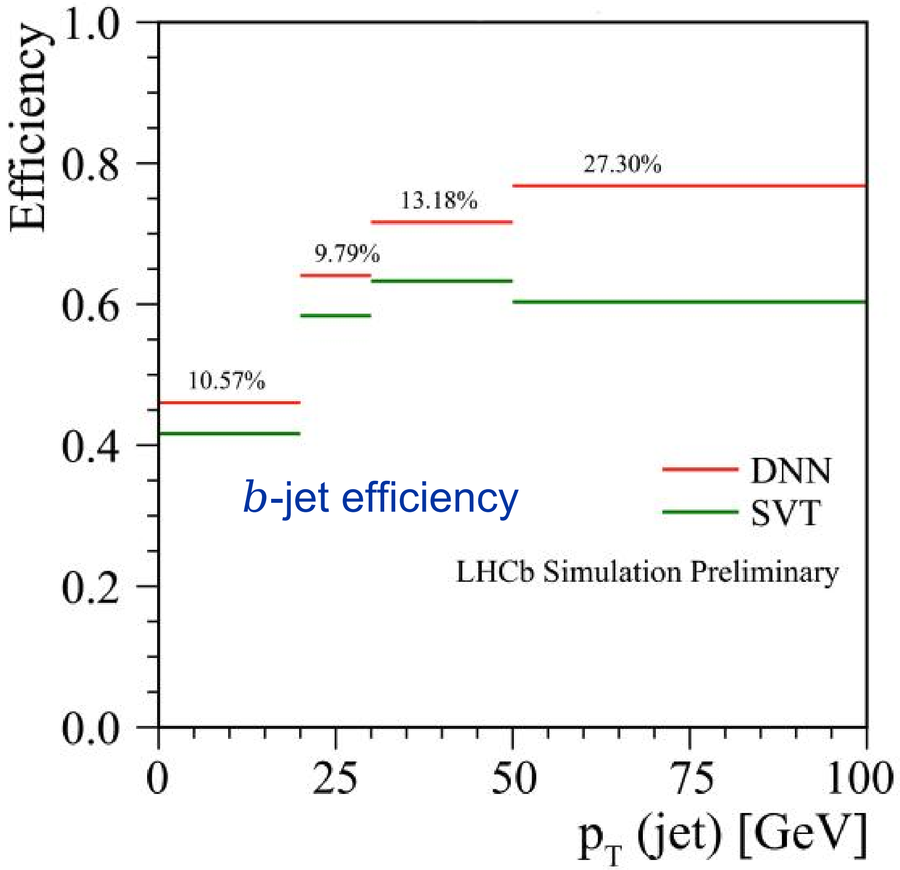


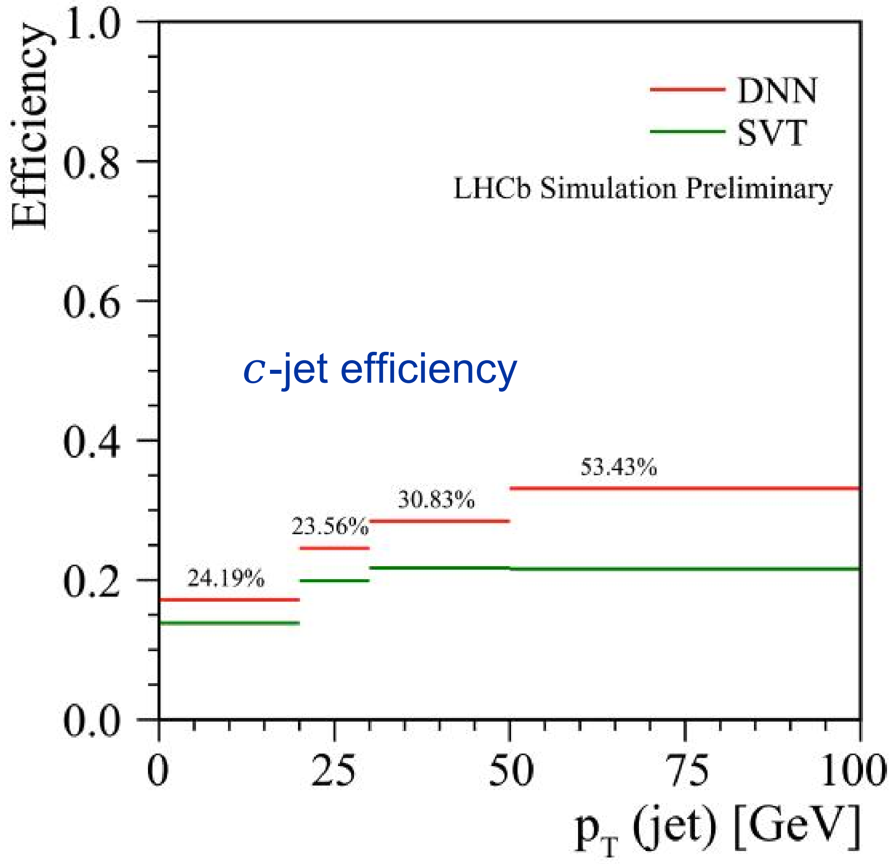
Deep Neural Network for Jet Identifiaction

LHCb-PAPER-2025-034 (in preparation)

- > 20(9)% improvement for $c\,(b)$-jet tagging with respect to standard SV tagging (SVT) algorithm

$$\varepsilon = \frac{N_{\text{tag}}}{N_{\text{rec}}}$$





Analysis Region

LHCb-PAPER-2025-034 (in preparation)

- Event selection:
  - two reconstructed jets, originating from the same PV
  - $p_{\mathrm{T}}(j) > 20$ GeV, $2.2 < \eta(j) < 4.2$, the difference in the azimuthal angle between the jets >1.5
- The main challenge is the accurate modelling the multi-jet QCD background: data-driven method
- Use tagging DNN probabilities to define
  - signal region (SR), optimized for the signal significance
  - control region (CR), used to characterise the properties of the multi-jet QCD background

LHCb Preliminary

Search	$H \to b\bar{b}$	$H \to c\bar{c}$
SR requirements	$P_b > 0.25$ and SV-tag on both jets	$P_c > 0.15$ and SV-tag on both jets
CR requirements	$P_b > 0.25$ and SV-tag on one jet $P_b < 0.25$ on the other jet	$P_c > 0.15$ and SV-tag on one jet $P_c < 0.02$ and SV-tag on the other jet
R_p	5.5	35

$$R_p = \frac{N_{\mathrm{sig}}^{\mathrm{SR}} / N_{\mathrm{data}}^{\mathrm{SR}}}{N_{\mathrm{sig}}^{\mathrm{CR}} / N_{\mathrm{data}}^{\mathrm{CR}}}$$

Simulation Correction

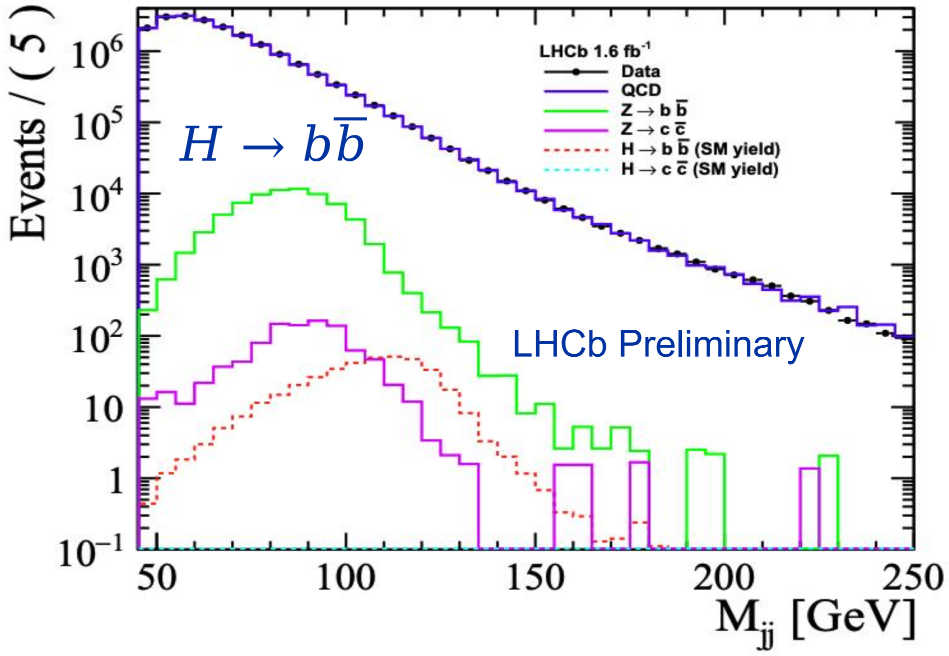
LHCb-PAPER-2025-034 (in preparation)

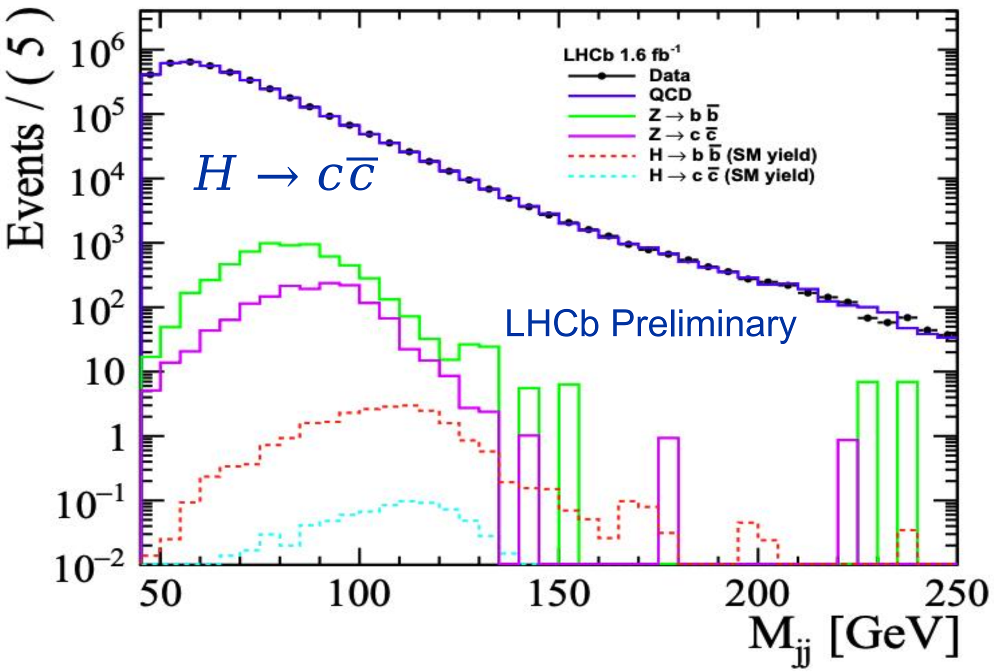
- MC samples are used to
  - determine selection efficiencies
  - model the signal dijet invariant mass distribution
  - investigate the properties of backgrounds
  - evaluate the performance of the GBR algorithm and DNN
- How is MC corrected
  - SV-tagging efficiency → p_{T} dependent reweight
  - DNN selection efficiency → tag and probe method, apply data/MC scale factors
  - Global event cut efficiency → reweight nSPD hit distribution to data
  - Jet energy (GBR) → also applied regression

Fit to Data

LHCb-PAPER-2025-034 (in preparation)

- Binned maximum likelihood fit to the dijet invariant mass in the SR
  - the background is modelled using data in the CRs, together with a transfer function to account for kinematic differences between CRs and SRs
- The fitted signal yields are compatible with zero within statistical uncertainties





Systematic Uncertainties

LHCb-PAPER-2025-034 (in preparation)

- The expected upper limit: all sources are implemented as nuisance parameters in the fit
- Impact is evaluated by removing each nuisance one at a time and re-fitting
- Systematic uncertainty dominated by SV-tagging and L0+HLT1 efficiencies

Systematic source	$H \to b\bar{b}$ search	$H \to c\bar{c}$ search
GEC	0.4 %	0.1 %
Trigger	3.2 %	3.6 %
Jet SV-tagging	15 %	19 %
Jet Identification	0.7 %	0.1 %
Jet Energy Resolution	0.03 %	0.2 %
Jet Energy Scale	1.2 %	0.3 %
Background model	~ 1 %	~ 1.2 %
Z boson yield	2%	1%
Signal region contamination	$< 1\%$	$< 1\%$
DNN	0.13 %	1.8 %
SM cross sec. + Acceptance	1.8%	1.7%
Luminosity	1.2 %	0.2 %

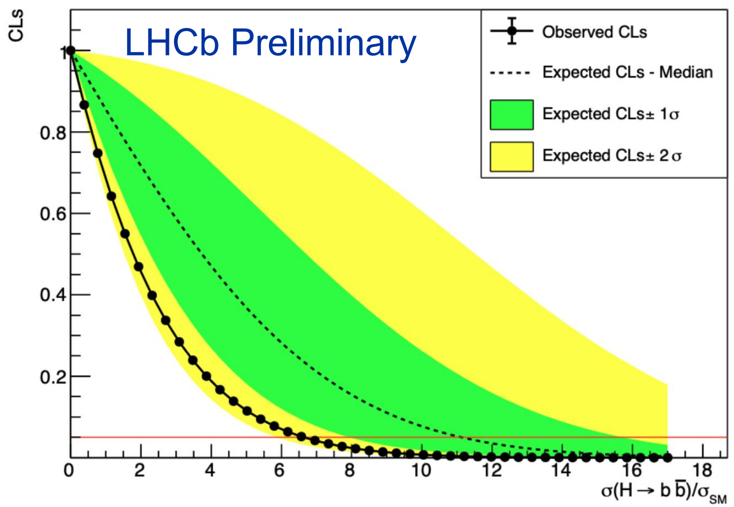
LHCb Preliminary

Observed Upper Limits

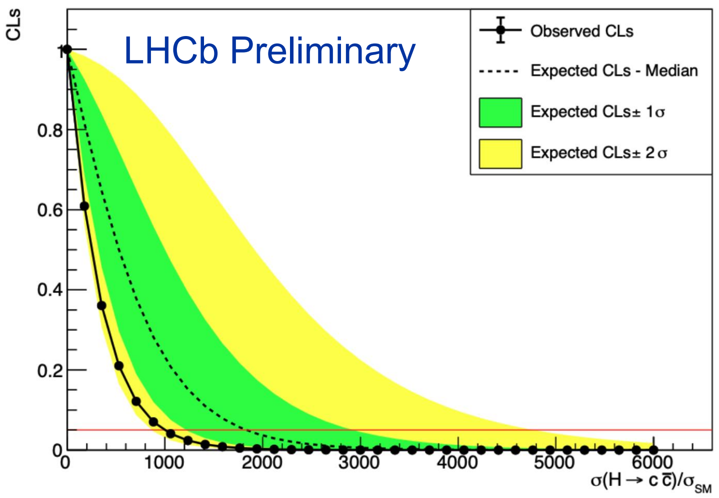
LHCb-PAPER-2025-034 (in preparation)

- Upper limits are derived using the CLs method with 95% confidence level

H → bb upper limit



H → cc upper limit



Process	$\sigma_{\mathrm{UP}}/\sigma_{\mathrm{SM}}$	N_{UP}	σ_{UP} [pb]
$H \to b\bar{b}$	11.1	2652	353
$H \to c\bar{c}$	1834	1414	2939

Expected upper limits

Process	$\sigma_{\mathrm{UP}}/\sigma_{\mathrm{SM}}$	N_{UP}	σ_{UP} [pb]
$H \to b\bar{b}$	6.64	1587	211
$H \to c\bar{c}$	1003	722	1605

Observed upper limits

In Run 1 conference note (LHCB-CONF-2016-006) observed upper limits (2 fb^{-1} at 8 TeV)

- $V + H \to b\bar{b}$: $\sigma < 50\ \sigma_{\mathrm{SM}}$
- $V + H \to c\bar{c}$: $\sigma < 6400\ \sigma_{\mathrm{SM}}$

Prospects

LHCb-PAPER-2025-034 (in preparation)

- Dedicated trigger lines: trigger won’t reply on SV-tagging, apply DNN at trigger level
- Removal of GEC: more statistics
- Improved tracking: increase SV-tagging efficiency
- HCAL removal: slightly degrades the performance of jets

extrapolated limits

LHCb Preliminary

	$\sigma_{\mathrm{UP}}/\sigma_{\mathrm{SM}}$ 2016	$\sigma_{\mathrm{UP}}/\sigma_{\mathrm{SM}}$ Run 2	$\sigma_{\mathrm{UP}}/\sigma_{\mathrm{SM}}$ Run 4	$\sigma_{\mathrm{UP}}/\sigma_{\mathrm{SM}}$ U2
$H \to b\bar{b}$	11.1	6.0	1.1	0.38
$H \to c\bar{c}$	1834	990	141	45

LHCb Preliminary

y_c/y_c^{SM} 2016	y_c/y_c^{SM} Run 2	y_c/y_c^{SM} Run 4	y_c/y_c^{SM} U2
43	31	12	6.7

Conclusions

LHCb-PAPER-2025-034 (in preparation)

- LHCb-PAPER-2025-034 provides the first LHCb limits on the search for $H \to b\bar{b}$ and $H \to c\bar{c}$ in inclusive final state from the pp collisions using 2016 data, $\mathcal{L} = 1.6\ \text{fb}^{-1}$
- New techniques have been used to
  - correct reconstructed jet energy
  - identify jet flavour
- New upper limits significantly improve upon LHCb-CONF-2016-006
- The first evidence of the production of $H \to b\bar{b}$ in the forward region could be possible with Run 4 data
- The Upgrade II (Run 5) with 300 fb^{-1} of integrated luminosity will be able to measure the $H \to b\bar{b}$ properties in more details

LHCb-TDR-023

LHC Run Year	Integrated luminosity fb^{-1}		
$\text{cm}^{-1}\text{s}^{-1}$	1.0×10^{34}	1.5×10^{34}	2.0×10^{34}
Run 1-4	50	50	50
LS4	–	–	–
Run 5 Year 1	21	25	26
Run 5 Year 2	43	50	51
Run 5 Year 3	43	50	51
LS5	–	–	–
Run 6 Year 1	43	50	51
Run 6 Year 2	43	50	51
Run 6 Year 3	43	50	51
Total	284	325	331

Back Up

home.cern

Fit Function

LHCb-PAPER-2025-034 (in preparation)

$$f(x) = N_{QCD,SR} \cdot B_{CR}(x) \cdot TF(x) + N_{Hbb,SR} \cdot B_{Hbb}(x) + N_{Hcc,SR} \cdot B_{Hcc}(x) + N^{\mathrm{SM}}_{Zbb,SR} \cdot R_Z \cdot B_{Zbb}(x) + N^{\mathrm{SM}}_{Zcc,SR} \cdot R_Z \cdot B_{Zcc}(x)$$

- $N_{QCD,SR}$: multi-jet QCD yield (free paramter)
- $B_{CR}(x)$: CR mass distribution
- $TF(x)$: transfer function (fitted in data)
- $B_{Hbb(cc)}(x)$: $H \to b\bar{b}(c\bar{c})$ mass model (from simulation)
- $N^{SM}_{Zbb(cc),SR}$: $Z \to bb(cc)$ yield, from SM expectation
- $B_{Zbb(cc)}(x)$: $Z \to bb(cc)$ mass model (from simulation)

$$TF(x) = N \cdot \sum_{i=0}^{n} c_i B_{i,n}(x),$$

$$B_{i,n}(x) = \binom{n}{i} x^i (1-x)^{n-i}.$$