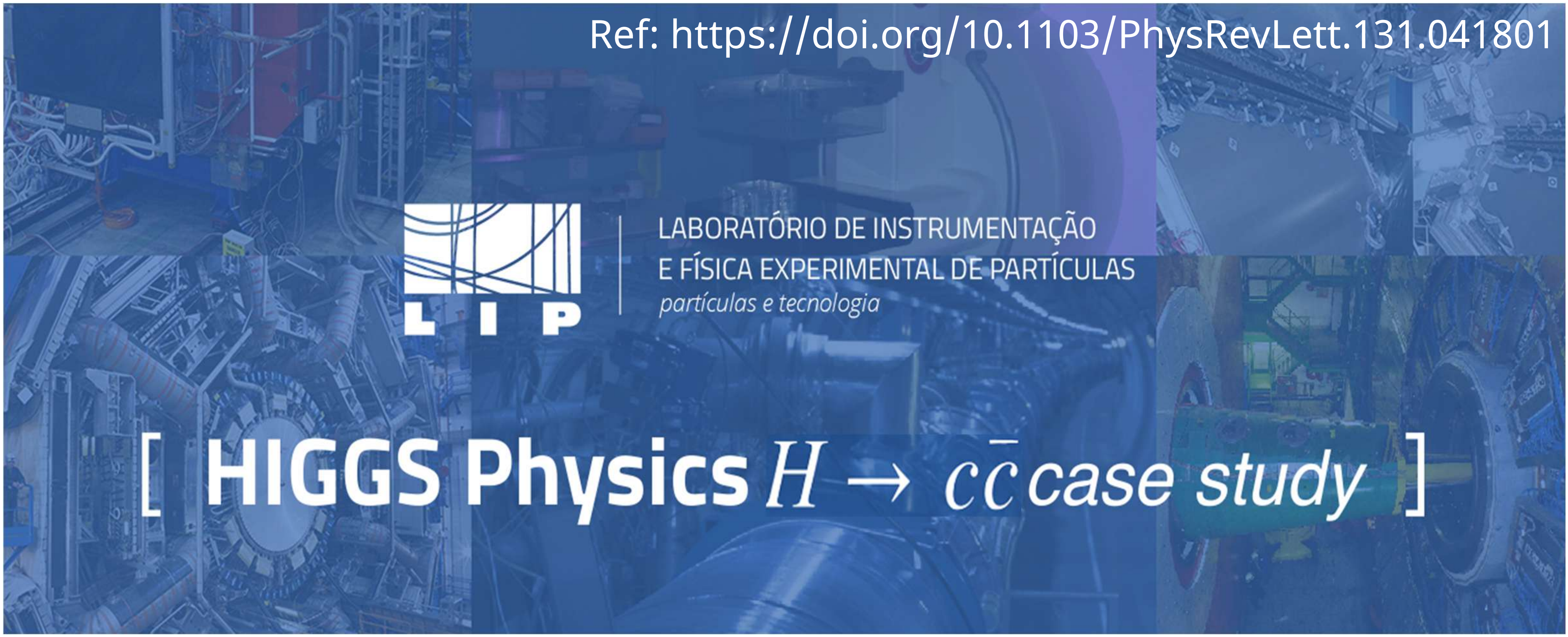
Searching Rare Higgs Decay -

Proton-Proton collisions in boosted topology



04 July 2025

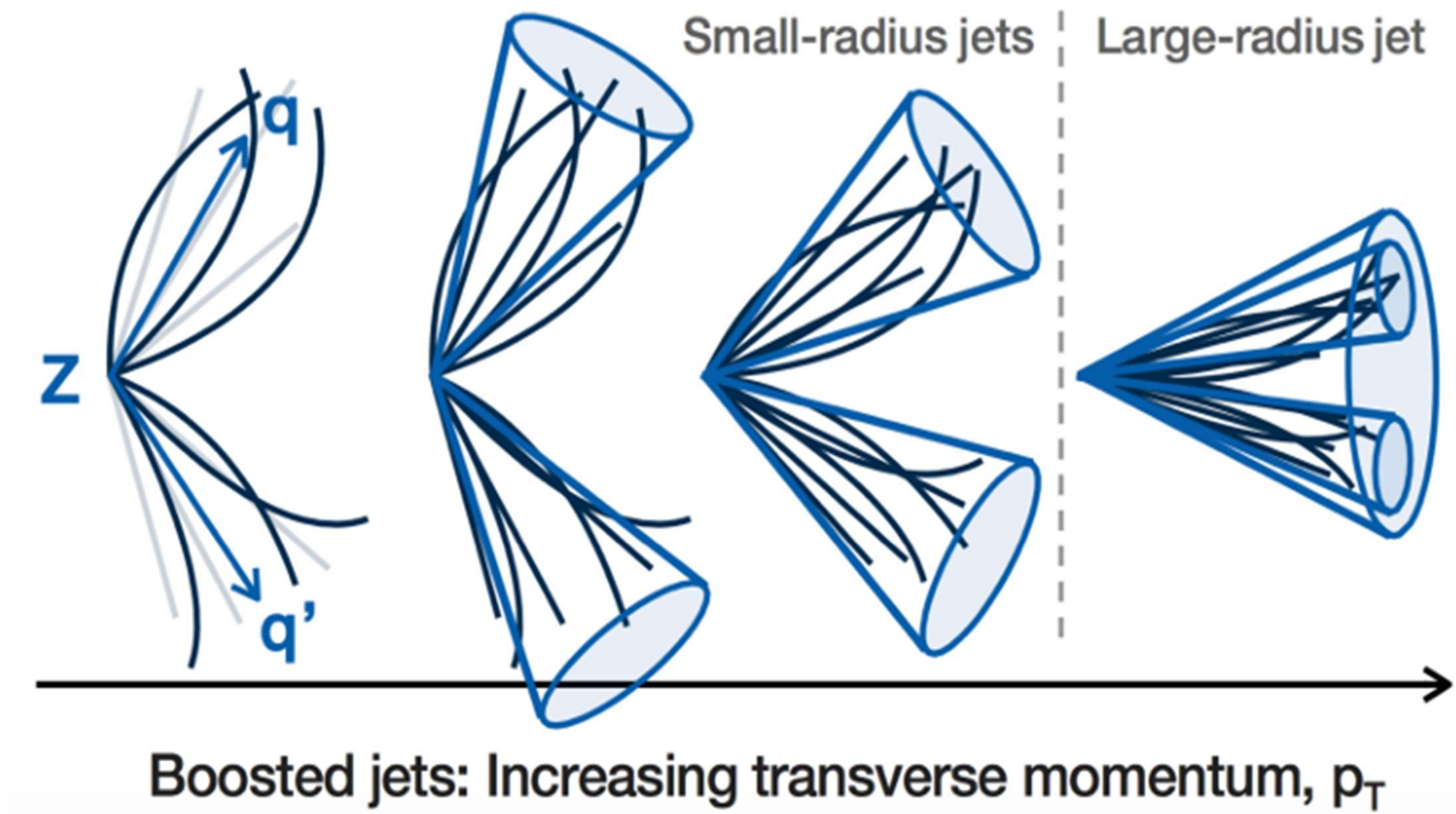
Ankur Goel – Program Alumnus

Program: 14th Course on Physics @ LHC - 2025

Introduction

- This study presents SM Higgs Boson produced at pT > 450 GeV and decaying into charm-anticharm quark pair.
- P-P collisions at √s = 13TeV were conducted by the CMS experiment at the LHC, with luminosity of 138 fb^-1.
- Boosted H → cc(bar) decay products were reconstructed as a large single radius jet and identified using a deep neural network charm tagging technique.
- For validation Z → cc(bar) decay process was measured with jets at high P_t and signal strength of $1.00^{+0.17}_{-0.14}\,(\mathrm{syst}) \pm 0.08\,(\mathrm{theo}) \pm 0.06\,(\mathrm{stat})$
- The observed upper limit on cross section (H)* BR (H → cc(bar)) is set at 47(39) times the SM prediction, with 95% confidence.

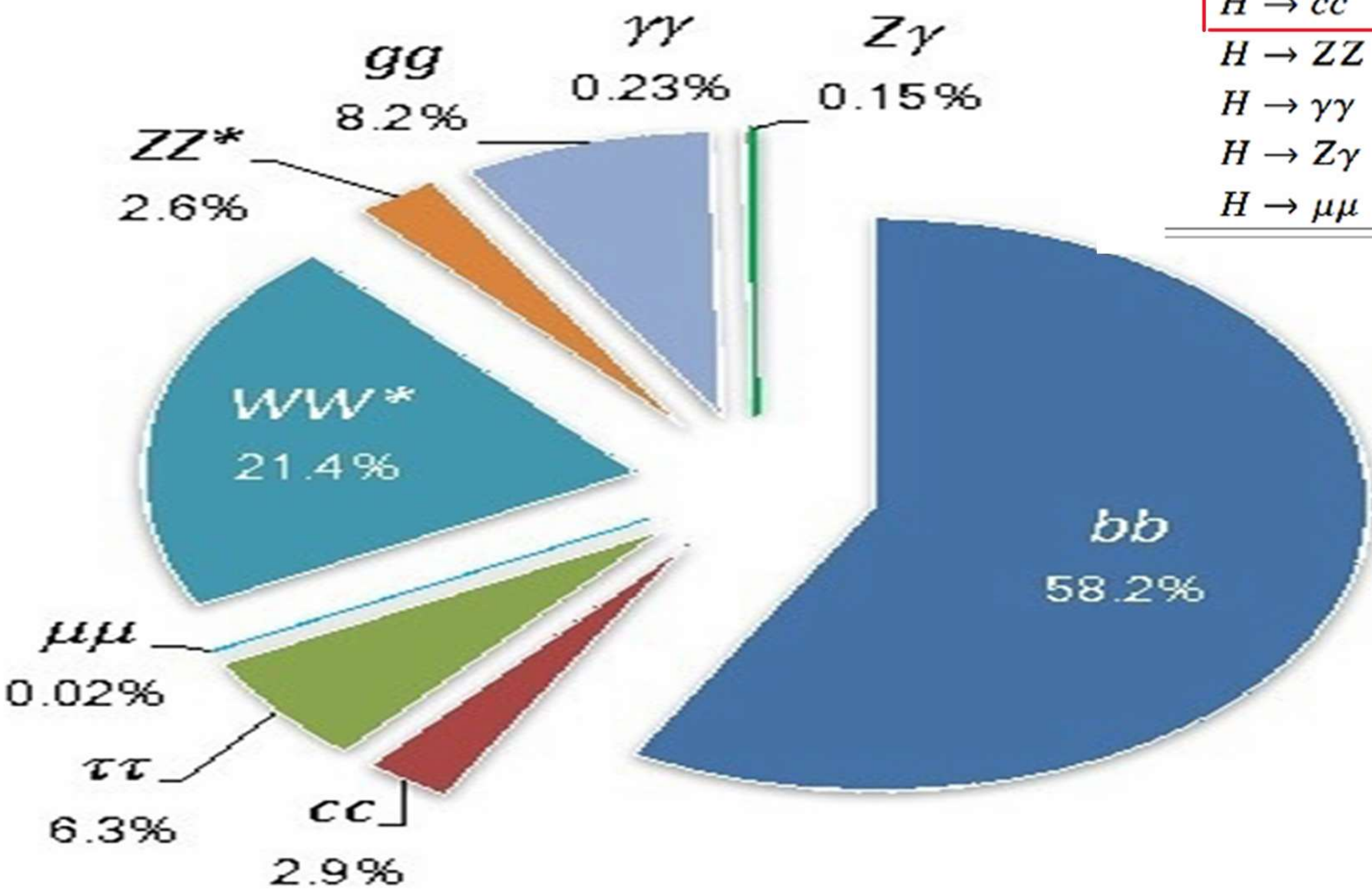
- Quarks/Gluons Hadronize to produce a collimated spray of particles called Jets.
- This study focusses on reconstructing signal from Large-radius jets.



Background – Decay Channels for Higgs

- Couplings and kinematics determine these Branching ratios.

Decay channel	Branching ratio [%]
$H \to b\bar{b}$	57.5 ± 1.9
$H \to WW$	21.6 ± 0.9
$H \to gg$	8.56 ± 0.86
$H \to \tau\tau$	6.30 ± 0.36
$H \to c\bar{c}$	2.90 ± 0.35
$H \to ZZ$	2.67 ± 0.11
$H \to \gamma\gamma$	0.228 ± 0.011
$H \to Z\gamma$	0.155 ± 0.014
$H \to \mu\mu$	0.022 ± 0.001



Ref: H → bb(bar) case study, Rute Pedro, 7th April, LIP

Background – Decay Channels for Higgs

- H → Muon pairs is the most experimentally accessible decays and has been explored by ATLAS and CMS collaborations.
- CMS recently found Higgs couplings to muons.
- While, H → cc(bar) is more challenging to find due to:

  a.) Such decays are extremely rare.

  b.) Extremely large multijet Background especially H → bb(bar).
- What has changed/Improved ?

  a.) Better flavour tagging techniques.

  b.) Advanced Jet substructure.
- Prior searches focussed on H production with a Vector boson (VH, where V could be a W or Z).

Background – Decay Channels for Higgs

- The g-g fusion (ggF) and vector boson fusion (VBF) modes benefits from a larger cross section.

- This study focuses on the events from the ggF production.

- The search strategy is same as that of H → bb(bar) decay channel.

- What's different ?

  → Use of new mass decorrelated discriminators to define a charm enriched signal region.

  → On the existing ATLAS and CMS measurements an additional constraint on the decay process is added.

The Search Mechanism

Peter Higgs (1929 – 2024)

Identification

- The search used a dataset of p-p collisions at $\sqrt{s}$ = 13TeV, collected with CMS detector.
- Candidate events are selected using High pT, large radius jet with substructure observables expected from an H → cc(bar) decay.
- Deep neural Network (DNN) discriminators separates the H signal events from background events, specifically QCD induced multijet events (background events).
- A parallel search for Z → cc(bar) decay is performed to validate the process (as those decays are more abundant).

Identification

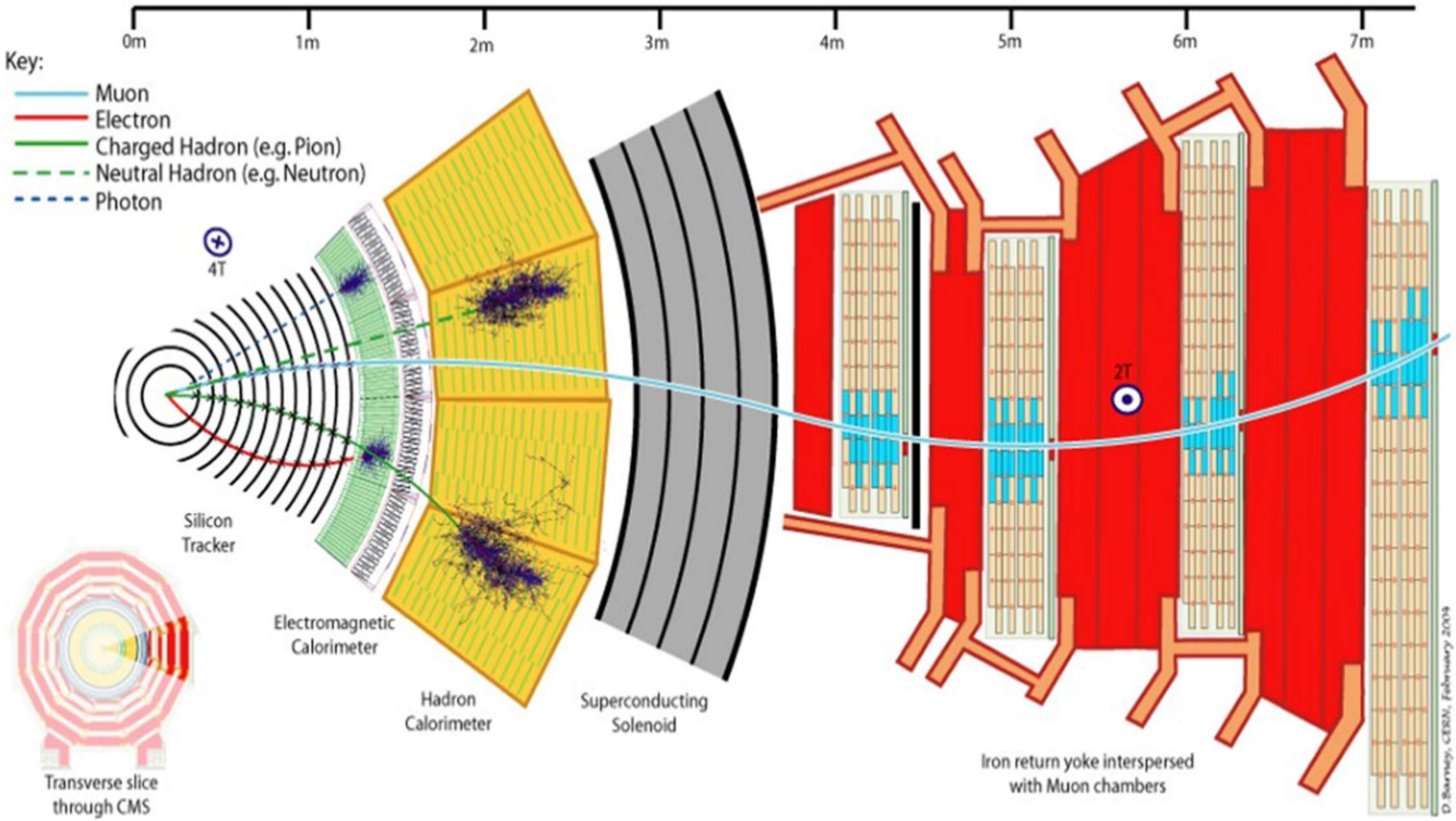
- A simultaneous fit of a model of jet mass distributions is then performed across H → cc(bar), Z→ cc(bar) signals, QCD multijet background events and other background processes across:

- This allows to measure the signal production cross sections.

- The compact muon solenoid (CMS) is a multipurpose detector, nearly hermetic, and can identify electrons, photons and charged and neutral hadrons, apart from muons of course.

- A "particle-flow" algorithm reconstructs all individual particles in an event, combining information from the silicon tracker, the crystal electromagnet and the brass scintillator hadron calorimeters, operating inside a 3.8 T superconducting solenoid.

Identification



https://indico.cern.ch/event/611242/contributions/2464737/attachments/1442943/2222206/SecondHiggsLecture_Conde.pdf

Reconstruction

- The reconstructed particles are used to build leptons, jets, and missing transverse momentum P_t.
- Simulated signal and background events are produced at matrix element level using various Monte Carlo Event Generators (like PYTHIA).
- The QCD multijet and Z+/W+ processes are modelled at QCD LO accuracy using MADGRAPH5_aMC@NLO v2.4.2 generator.
- The tt(bar) and single top quark processes are modelled at NLO using POWHEG 2.0. Diboson processes are modelled at LO using PYTHIA v8.226.
- The ggF Higgs production is simulated using the HJ-MINLO event generator with M_h = 125 GeV, including finite top quark mass effects.
- The POWHEG generator models H production through VBF, VH & tt(bar)H.

Reconstruction

- For Parton showering and hadronization, the POWHEG and MADGRAPH5_aMC@NLO samples are interfaced with PYTHIA v8.205.
- Reconstructed particles are clustered into jets using the anti-Kt algorithm.
- Small radius jets are clustered with distance parameter of 0.4 (AK4 jets) while Large radius jets arising from decays of heavy boosted heavy particles are reconstructed using a distance parameter of 0.8 (AK 8 jets).
- The detector response is modelled with GEANT 4.
- Any additional 'p-p' pile-up is mitigated through charged hadron subtraction and pileup-per-particle identification algorithm.
- Additional jet energy corrections are done as functions of pseudorapidity (η) and pT to account for the detector response.

Reconstruction

- Higgs (H) is constructed as a single AK8 jet with pT > 450 GeV.

- The soft drop (SD) algorithm is applied to the jet mass (mSD) to remove soft and wide angle radiation, which reduces the mass of the jets coming from QCD backgrounds events while preserving the mass of heavy boson decays.

- The range of interest is set to 40 < mSD < 201 GeV.

- To match the tracker acceptance region jets should have $|\eta|$ < 2.5.

- 95% of the expected yield is QCD multijet background.

- Other EW processes including Diboson, triboson and tt(bar)V are estimated from simulations and were found to be negligible.

Reconstruction

- Jet flavor is finally determined by Deep DoubleX DNN algorithm.

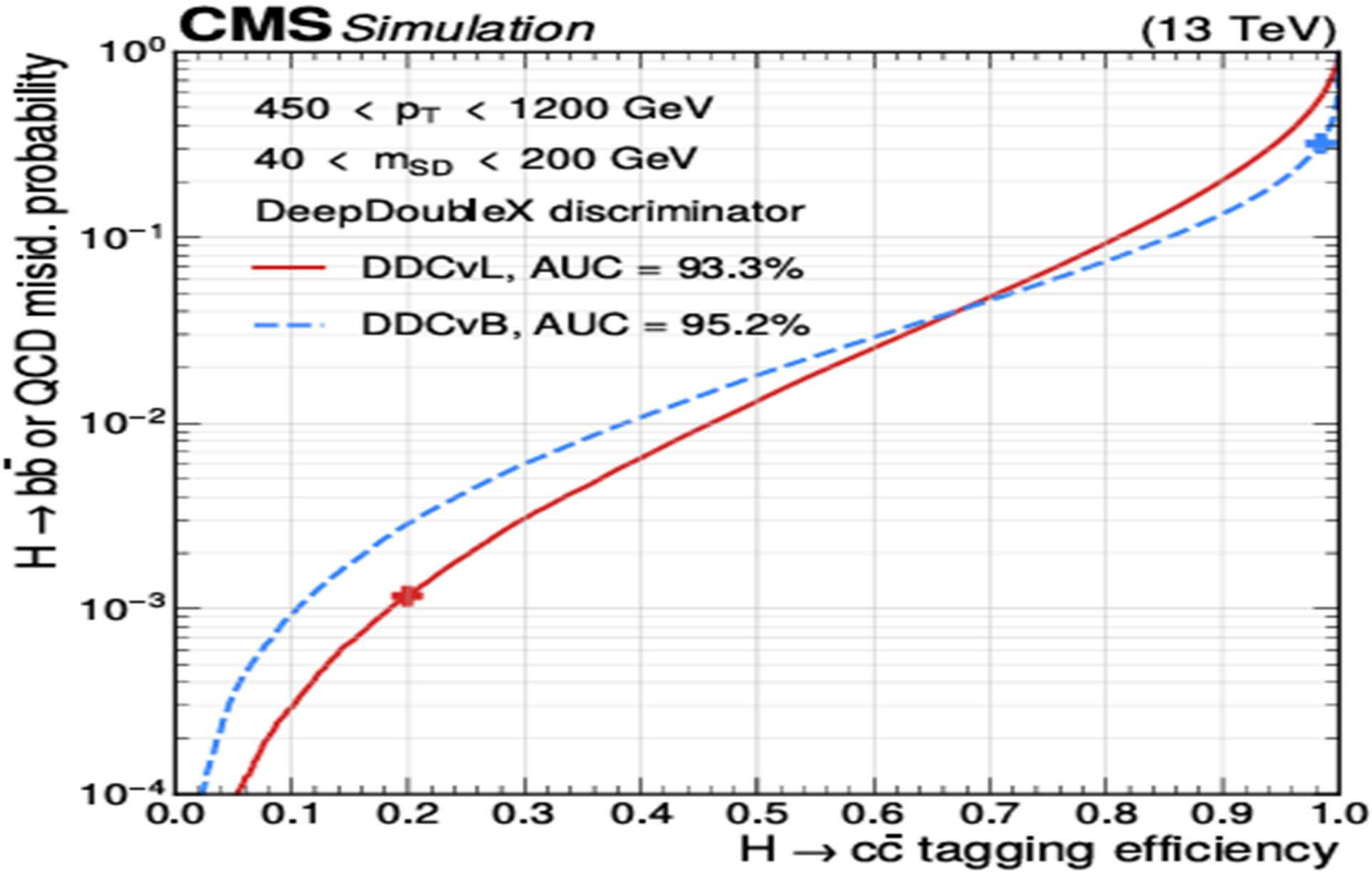
- The model is trained to distinguish between two-prolonged H-like signatures of bottom and charm flavors, as well as the QCD background, yielding two per-jet classifiers: charm versus light, DeepDoubleCvL (DDCvL) and charm versus bottom, DeepDoubleCvB (DDCvB).

- The figure (next slide) details out the tagging efficiencies of the two classifiers. H → cc(bar) has an efficiency of 20.6 % while that of H → bb(bar) is 4.8%.

- These efficiencies means the model will be more sensitive to H → cc(bar) detections and less likely to misidentify a b-jet as a c-jet.

- These efficiencies are also applied to Z → cc(bar) and Z → bb(bar)

  as these classifiers are mass independent.



- The working points are marked with a cross.
- AUC implies to the Area under the curve.
- Both the SR & the CR are divided into 23 evenly spaced bins with jet mSD between 40-200 GeV and 6 pT bins in the range 475 – 1200 GeV. Data taken from year 2016-2018.

Background Estimations & Corrections

- The V+ jet processes and differential tt(bar) are taken from simulation.
- The H→ bb(bar) contribution is taken from simulation and it is fixed to SM expectation.
- The normalizations in the SR and CR are corrected via two scale factors measured in a dedicated tt(bar)-enriched control region.
- This control region is adapted from the SR, by lowering the H candidate pT threshold, requiring exactly one muon and inverting the selection requirements on missing Pt and b-tagged Ak4 jets. This gives predominant semi-leptonic tt(bar) background events in CR.
- The expected SR yield is about 5 times that of H→ cc(bar) signal and its impact is negligible w.r.t overall background uncertainty.
- The corrections are displayed in the table next. These corrections are independent of the jet flavor.

Background Estimations & Corrections

Data period	Jet mass correction [GeV]	Jet mass resolution	$N_2^{1,DDT}$, CvB selection	$t\bar{t}$ normalization	CvL selection $t\bar{t}$	CvL selection (W+jets)	CvL selection (signal)
2016	-1.17 ± 0.22	1.021 ± 0.017	0.89 ± 0.02	0.84 ± 0.05	0.93 ± 0.15	0.62 ± 0.09	1.15 ± 0.25
2017	-1.19 ± 0.23	1.019 ± 0.016	0.90 ± 0.02	0.86 ± 0.09	0.93 ± 0.15	0.64 ± 0.09	0.85 ± 0.16
2018	-0.12 ± 0.21	1.090 ± 0.031	0.92 ± 0.02	0.86 ± 0.08	1.00 ± 0.14	0.72 ± 0.08	0.74 ± 0.20

- The efficiency of DDCvL signal selection is estimated using data and simulations samples enriched in cc(bar) pairs and gluon splitting.

- Signal-like events are selected by requiring each of the two SD sub-jets of an AK8 jet to contain a muon, targeting semileptonic decays of b/c hadrons.

- Other systematic uncertainties are assigned to cover mismodelling of H signal, especially ggF & VBF production modes and higher order corrections to W & Z processes.

- The parameter of interest for us is µH i.e. Signal Strength.

Corrections

- μH = $$\frac{\sigma_{\text{obs}}(pp \to H) \cdot \mathcal{B}(H \to c\bar{c})}{\sigma_{\text{SM}}(pp \to H) \cdot \mathcal{B}_{\text{SM}}(H \to c\bar{c})}$$

$$\mu = \frac{(\sigma \cdot \text{BR})_{\text{observed}}}{(\sigma \cdot \text{BR})_{\text{expected}}}$$

Data

Standard Model

- μH – is dimensionless and is defined as the Observed rate of Higgs production x Branching ratio / SM prediction values.

- If μ = 1, the data matches SM prediction. If μ > 1 ; excess over SM.

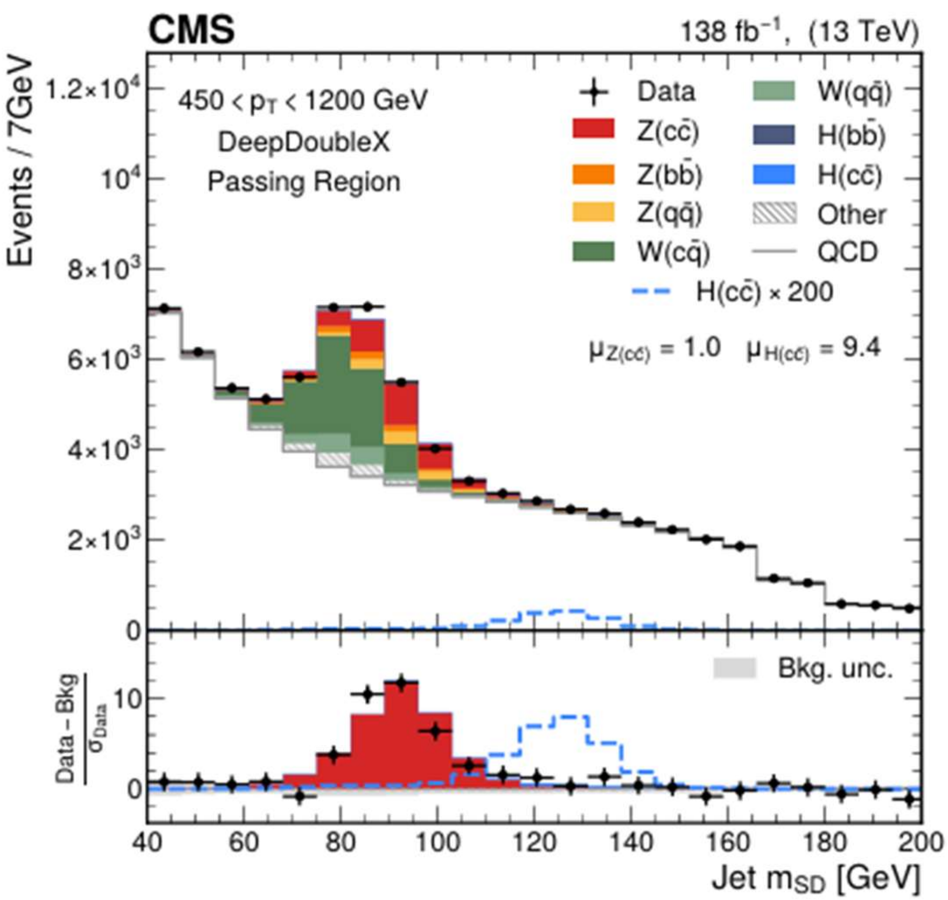
- A maximum likelihood fit is performed to the observed data using binned parameters (mSD, pT), where expected value is sum of the signal contribution + backgrounds accounted by nuisance parameters to account for systematic effects.

- As mentioned earlier to validate the process, μZ is also measured using a profile likelihood fit keeping μH as nuisance parameter. In this case it was found to be $1.00^{+0.17}_{-0.14}\,(\text{syst}) \pm 0.08\,(\text{theo}) \pm 0.06\,(\text{stat})$ which corresponds to an access, both observed and the expected value.

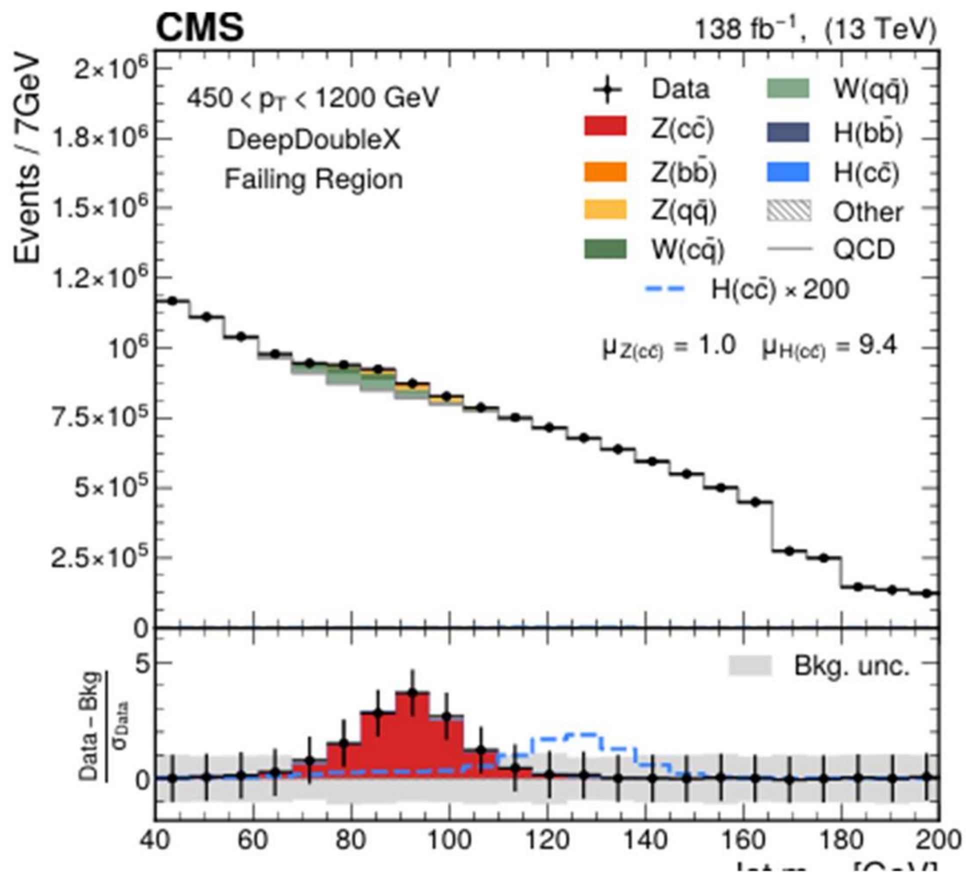
Corrections

- This is primarily due to the systematic uncertainty in the DDCvL signal tagging efficiency.
- To measure µH, we fix µZ ~= 1, by constraining the expected Z contribution to be within the applicable SM value.
- An observed upper limit is defined for µH and found to be 47(39) at 95% confidence level.
- For the best fit value of µH (ccbar) = $9.4^{+20.3}_{-19.9}$, the total mSD distributions in the passing and failing regions are shown in the next chart.
- The tt(bar) backgrounds has the highest contributions labelled as "Other".
- The "blue" dashed line represents H → cc(bar) expectation, multiplied by a factor of 200.

Results



Observed Msd – Passing region



Fitted Msd – Failing region

- The data has been collected from 2016 – 2018 combining all the pT categories.
- The fit is performed using signal-plus-background hypothesis with a single inclusive μH(ccbar).
- Bottom Panel: "Data – Bkg" residuals divided by statistical uncertainty.

To Sum Up…

- We searched a SM (Higgs) (and Z boson for process verification), at High Pt > 450 GeV and later decaying into cc(bar) pairs.

- The integrated Luminosity was kept at $138\,\mathrm{fb}^{-1}$ and √s= 13TeV and the data was collected.

- DNN algorithm identifies the jets originating from cc(bar) pairs.

- The Z→ cc(bar) process has been observed along with jets at the LHC for the first time with signal strength of $1.00^{+0.19}_{-0.17}$ relative to SM.

- This observation has now set an important reference for any future X → cc(bar) searches.

- The upper limit for $\sigma_H \times \mathcal{B}(H \to c\bar{c})$ has been set at 47(39) i.e. observed limit = 47 x SM, and expected limit = 39 x SM predictions

Questions ?

Thank you!