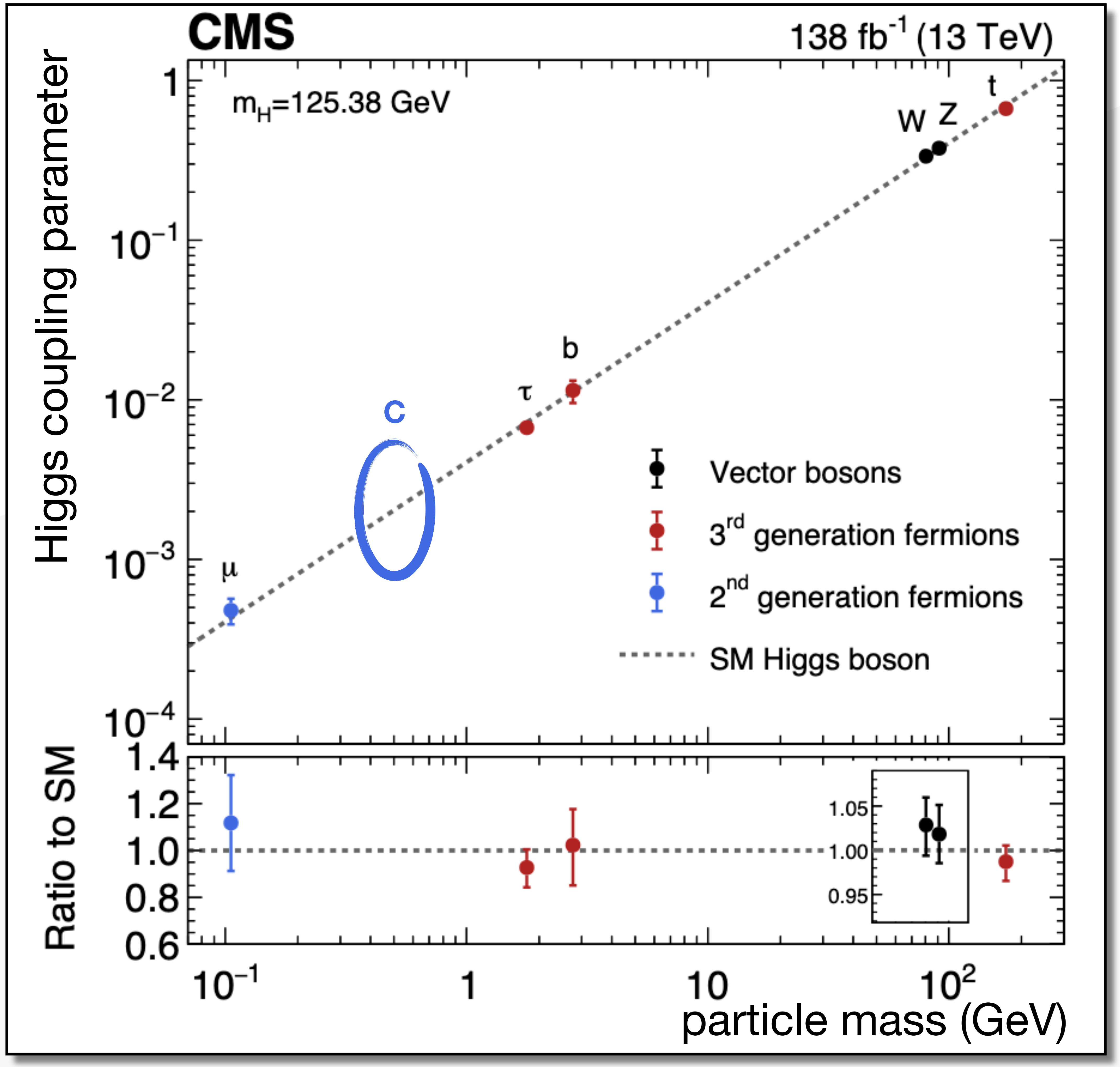
Searching for Higgs+charm production in the diphoton final state at CMS

Tiziano Bevilacqua

SPS conference 2023

H+c coupling

- Higgs boson has a unique role in the Standard Model, couplings to other particles are precisely predicted and proportional to particle mass.
- Precise measurements of Higgs coupling to fermions serve as test of SM consistency.
- Introduction of new particles and forces alters prediction.
- 3rd generation couplings already measured, 1 to 3 orders of magnitude bigger than second generation couplings.
- 2nd generation fermion couplings are one of the primary goals of CMS physics program:
  - $H \rightarrow \mu\mu$: 3σ evidence (JHEP 01 (2021) 148)
  - What about **charm**?

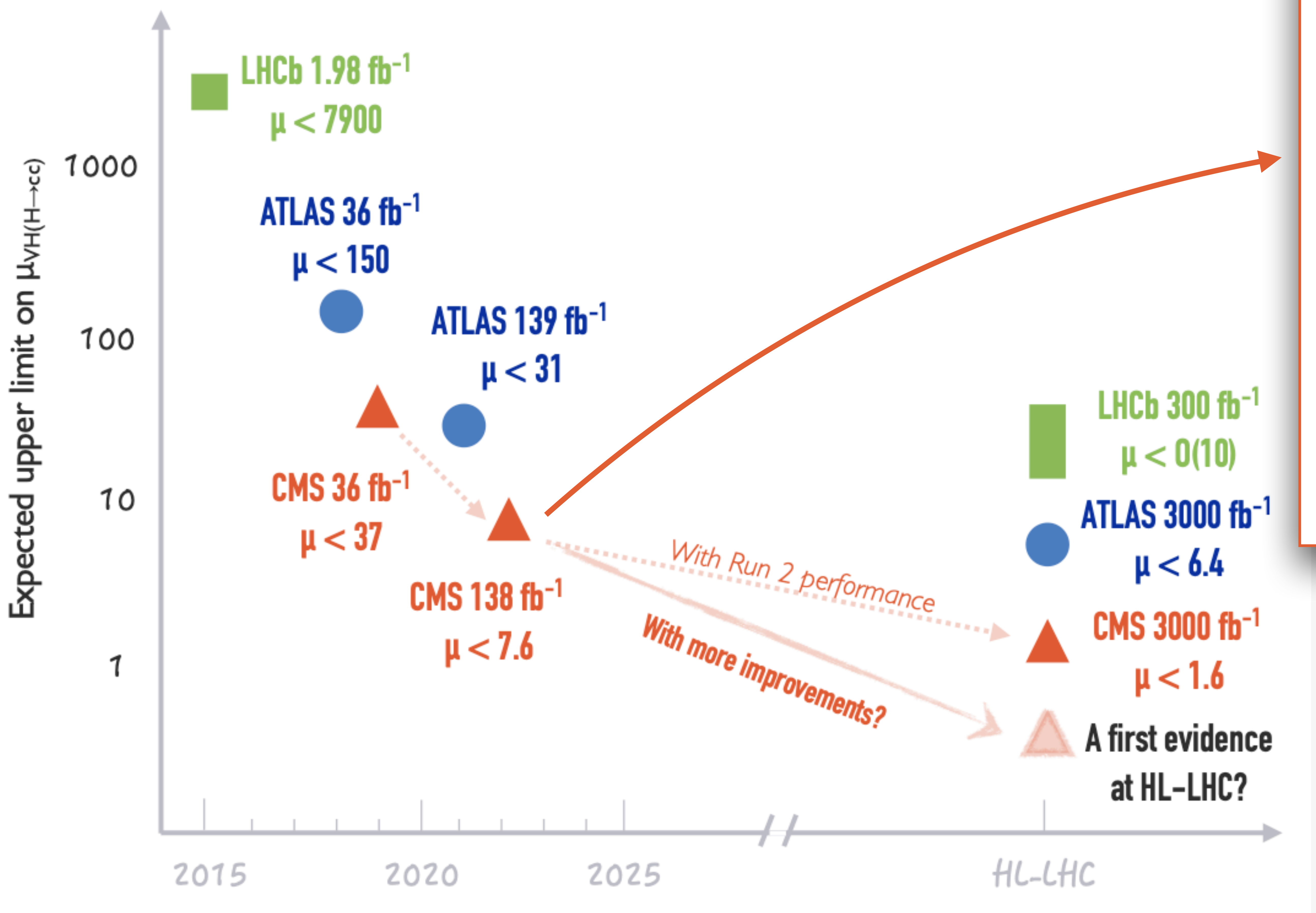


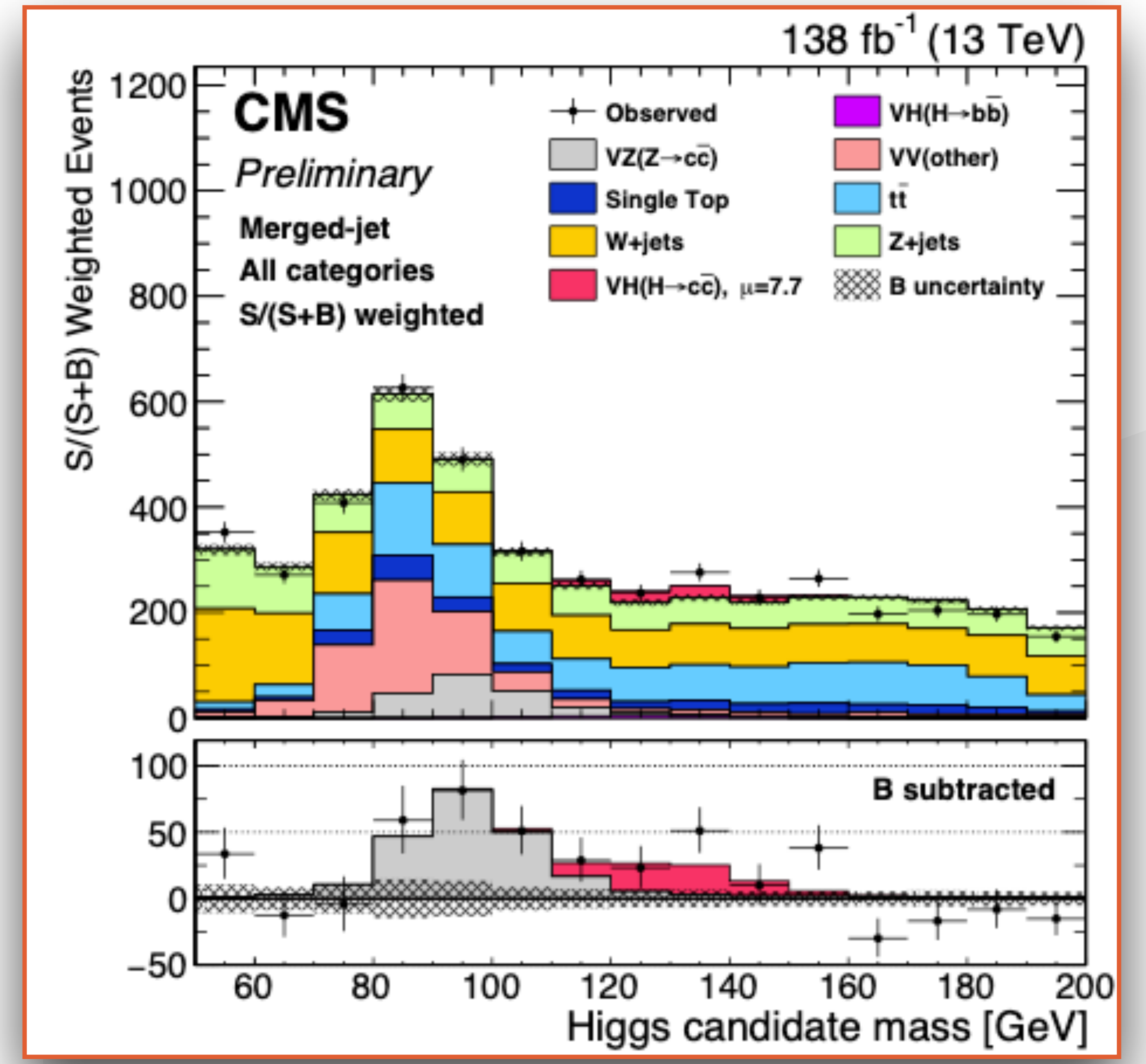
Nature 607, 60–68 (2022)

Previous results

- **Direct search** for $VH(H \to c\bar{c})$ PRL 131 (2023) 061801: recent improvements, most stringent limit on $H \to c\bar{c}$.

Upper limit $\mu_{VH(H\to c\bar{c})} < 14\ (7.6)$ observed (expected) ⇐





$$k_c = y_c / y_c^{SM}$$

$$\mu = \sigma / \sigma_{SM}$$

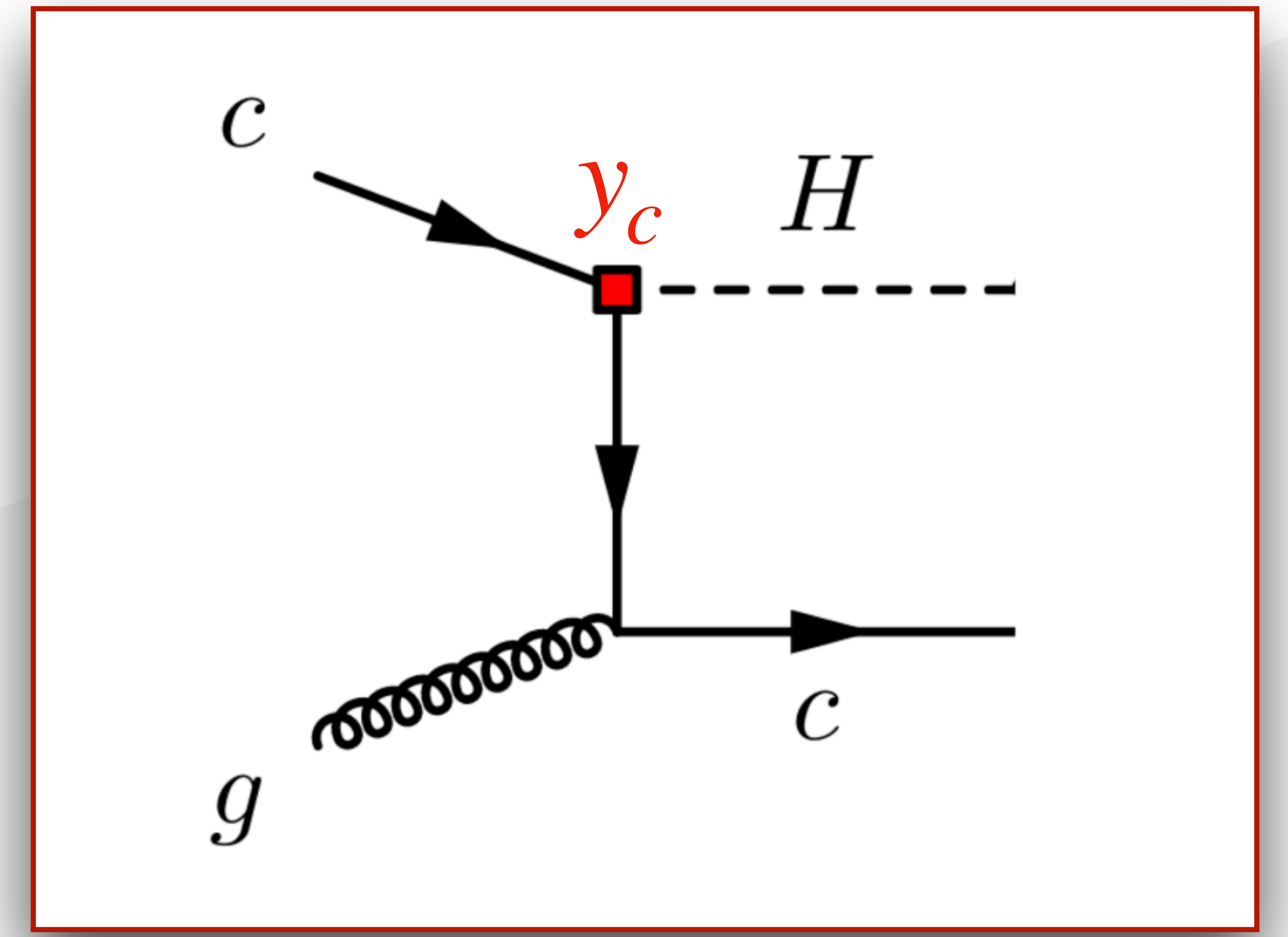
- **Other approaches**: Exclusive rare decays, $p_T(H)$ differential measurements …

H. Qu

New approaches

H+c associated production:

- Proposed in Isidori et al. (2015), constraint of $|k_c| < 3.9$ projected on 3000 fb^{-1}.
- Advantages of this channel:
  - Leading contribution requires only 1 charm to be tagged.
  - All H decays are available, ability to exploit the cleanest ones ($\gamma\gamma/ZZ \to 4L$).
  - Uncovered phase space, complementary to existing $H \to c\bar{c}$ searches.
- But also a few challenges:
  - Small cross section (~ 0.2 fb for $cH(H \to \gamma\gamma)$ vs 6.6 fb for $VH(H \to c\bar{c})$).
  - Non trivial signal MC simulation.
  - Challenging soft c-tagging.



No experimental results yet!

Higgs+charm production

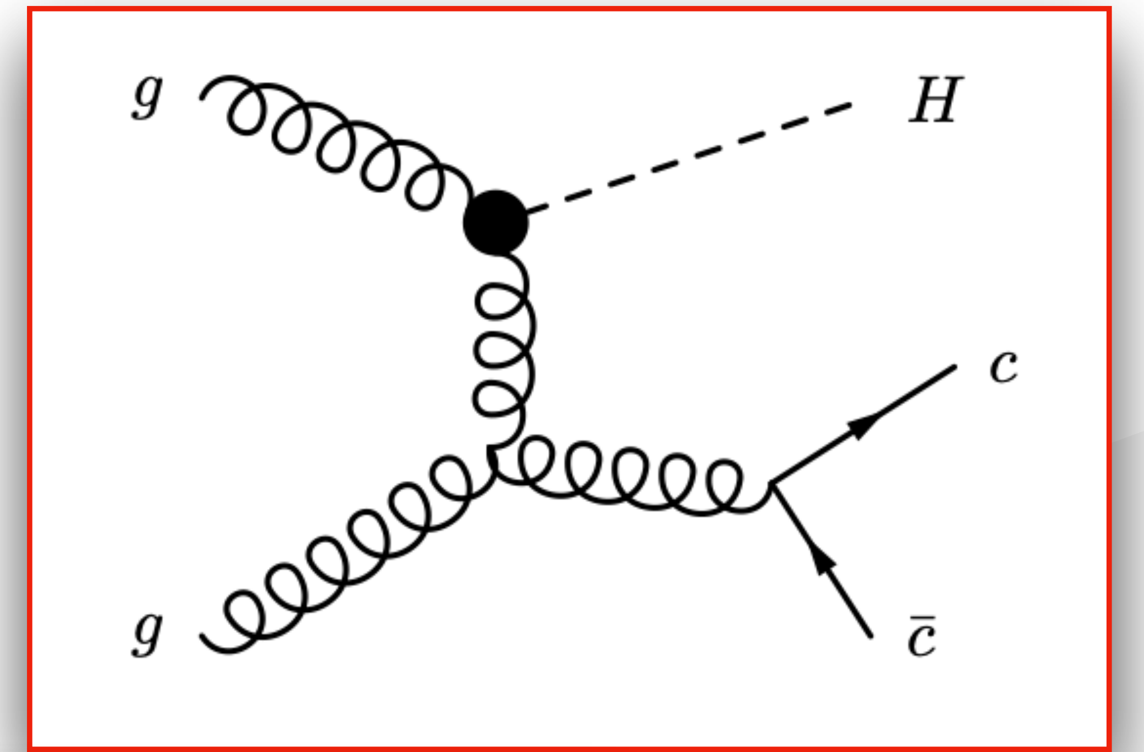
H+jets:

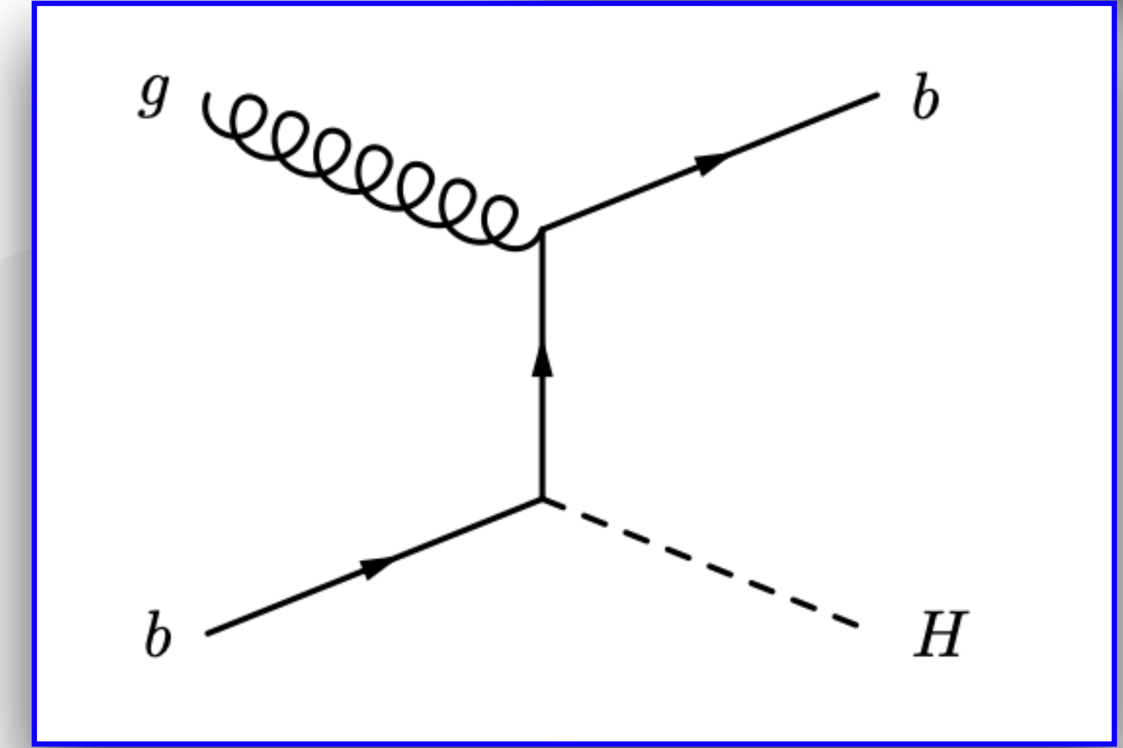
- $H + c$ final state includes several contributions that do not depend on y_c :
  - $H + g$ (fakes and $g \to c\bar{c}$)
  - $H + b(b \to c)$
  - $H + c$ not induced by y_c.

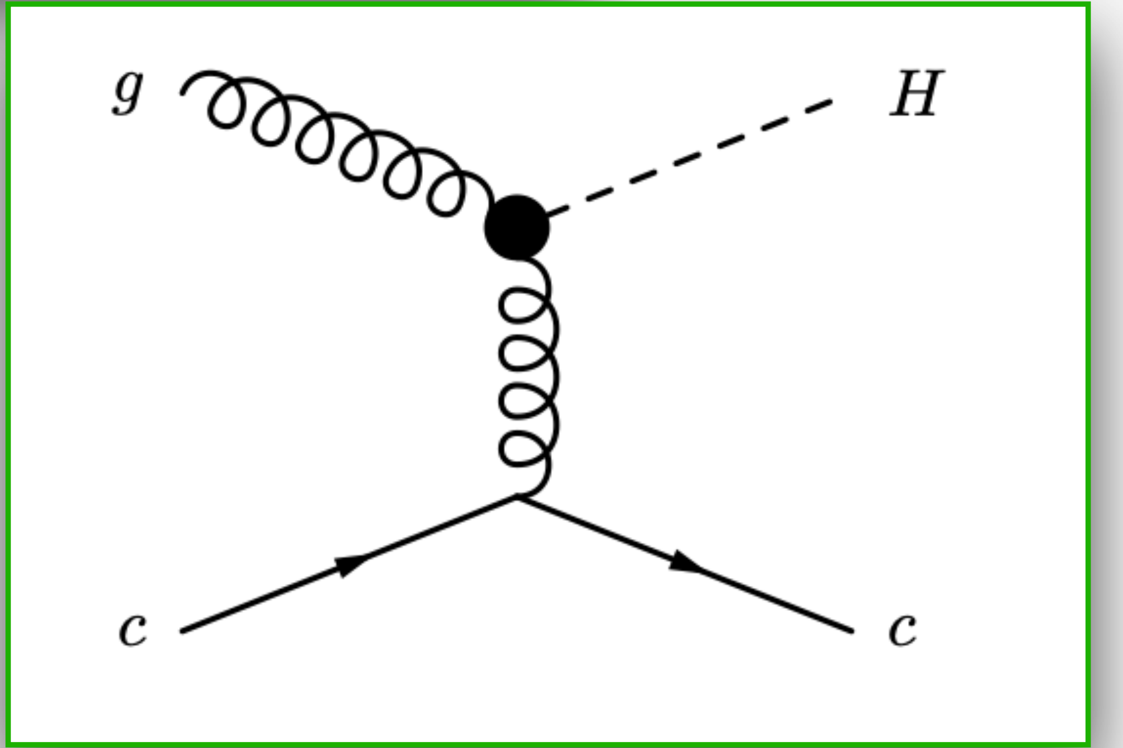
  $\Rightarrow$ most of the $H + c$ cross section is not sensitive to y_c.

- Modelling uncertainties on reducible and irreducible 'Higgs backgrounds' can limit sensitivity to y_c in $H + c$ channel .

- **Open questions:** How to simulate y_c induced H+c?
  - All these non-y_c contributions are already included in $H + jets$ MCs used by experiments (except for $H + b$ component that depends on y_b).
  - Many studies on $H + b$ simulation but none on $H + c$.







Higgs+charm simulation

H+c signal:

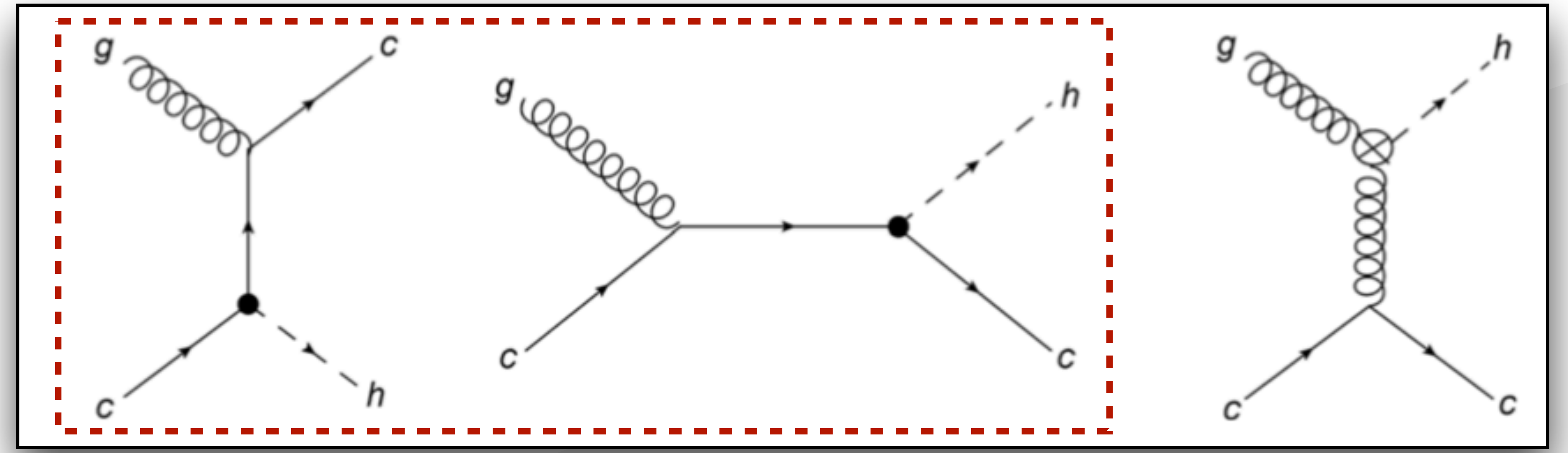
- Focusing on the signal simulation for $H + c$ MC with MadGraph_aMC@NLO (not available in CMS up to now).

$$\sigma(hc) = A + B \cdot y_c + C \cdot y_c^2$$

	σ [fb]
A	254.5
B	-3.5
C	34.5

$[y_c = y_c^{sm}]$

[GEN charm $p_T > 20$ GeV]

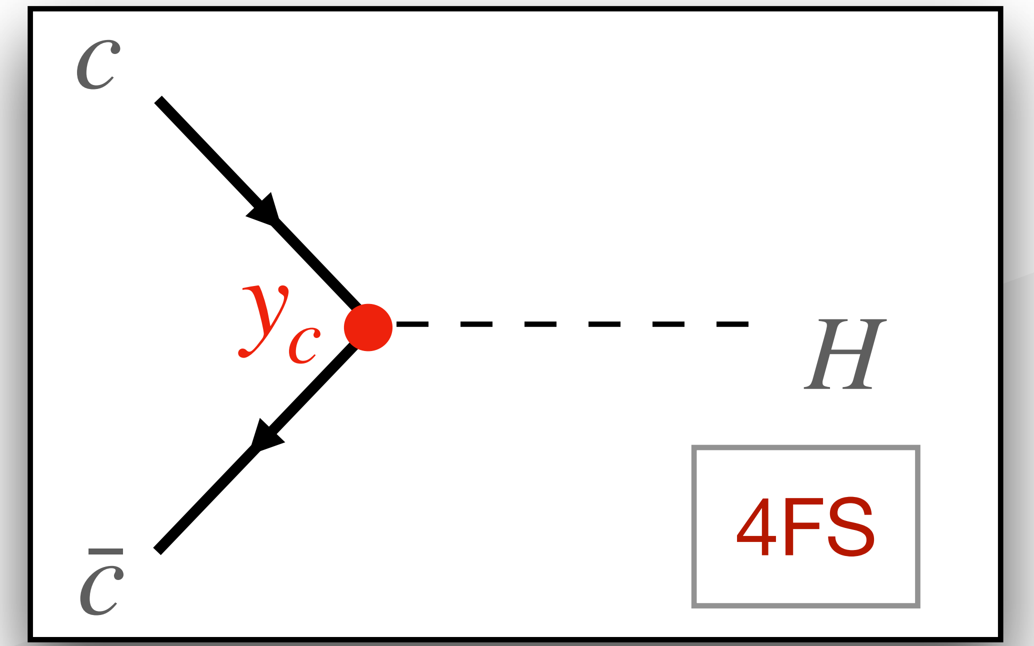


- $\sigma(hc)$ does not scale trivially with y_c, some tests were run with effective ggH coupling at LO.
- Biggest contribution from the term that does not probe y_c, but small y_c proportional interference term (~10 times smaller than the y_c^2 dependent term), for sensitivity $O(10 \cdot SM)$ contribution of ~1%.
- Theory study for bbH shows that the size of the interference term remains similar at NLO QCD.
- As first approximation one can generate signal probing y_c^2 and bgs/interference in separate MC (approach decided in agreement with theory experts), avoid overlap with $H + jets$ MCs.

Higgs+charm simulation

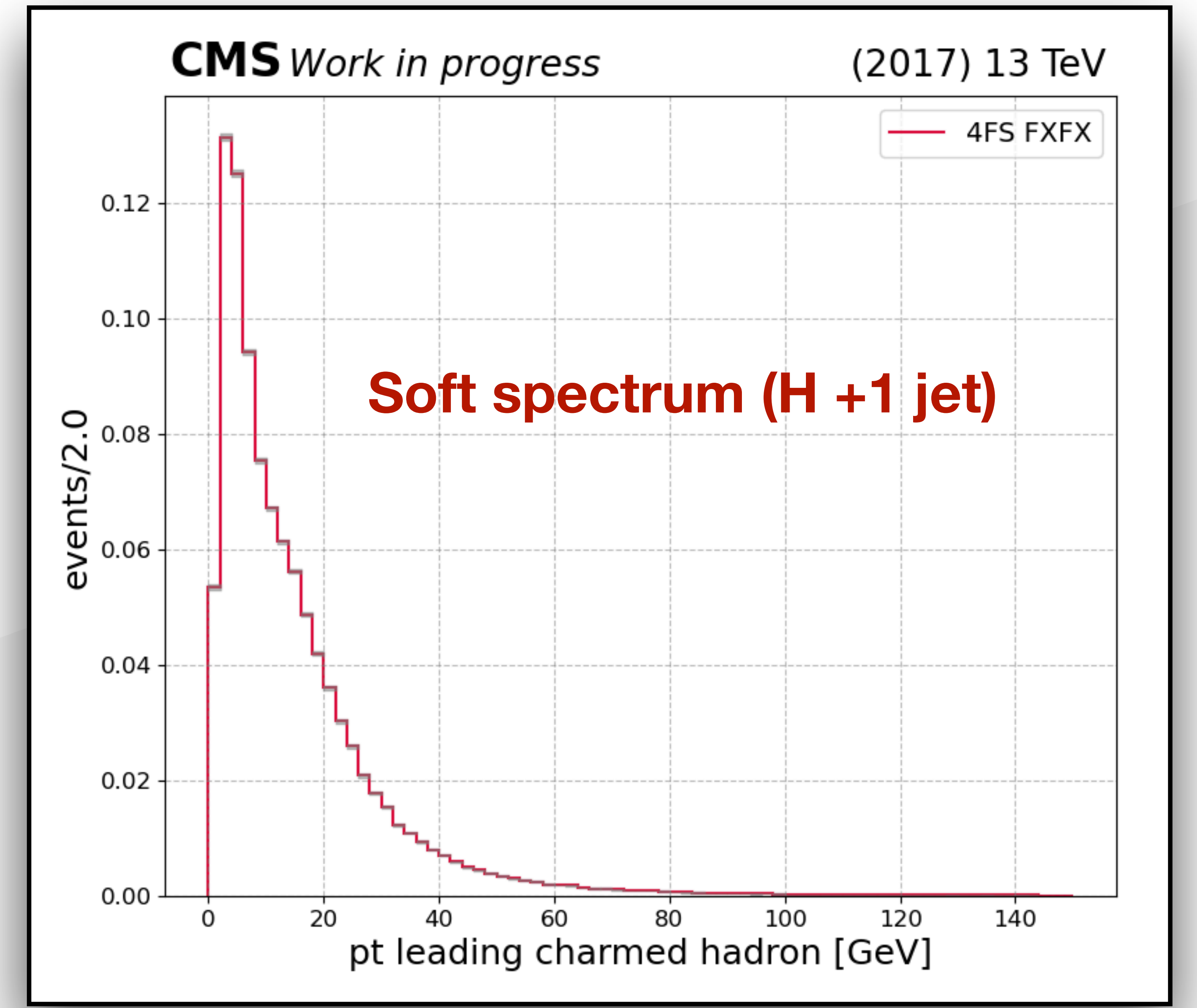
Focus on the y_c^2 term:

- Simulated with MadGraph_aMC@NLO ([QCD] NLO) + Pythia8 Parton Shower.
- Simulated using `loop_sm` model to have y_c in the $\overline{MS}$ renormalisation scheme and include running of $y_c \rightarrow \overline{y}_c(\mu_R)$ and $m_c \rightarrow \overline{m}_c(\mu_R)$.
- Simulated using 4 Flavour Scheme (4FS), to have charm quarks in the initial state, and with FXFX-merging to cover the full phase space.

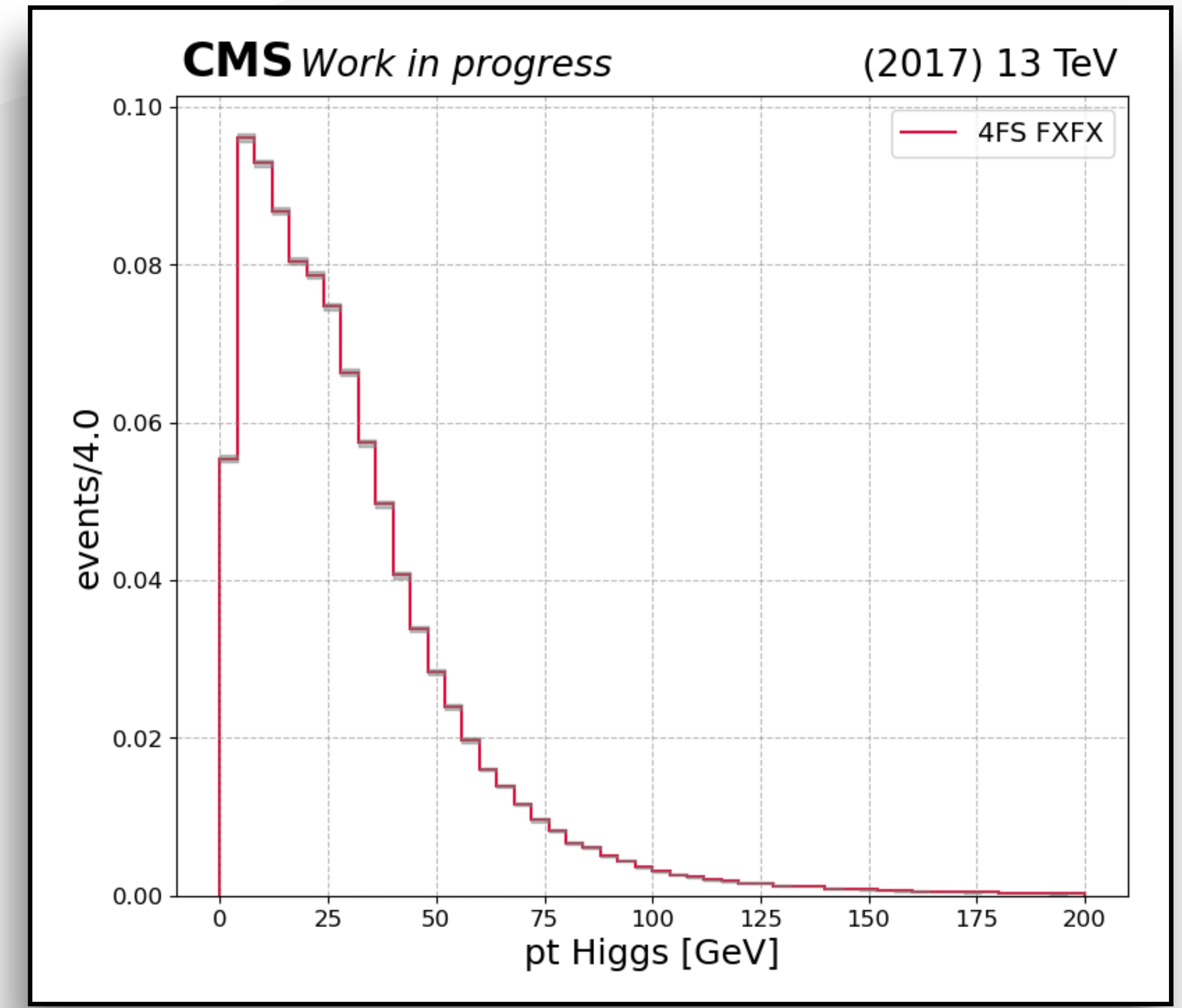


$m_C = 0$

Pt lead charmed hadron



Pt Higgs

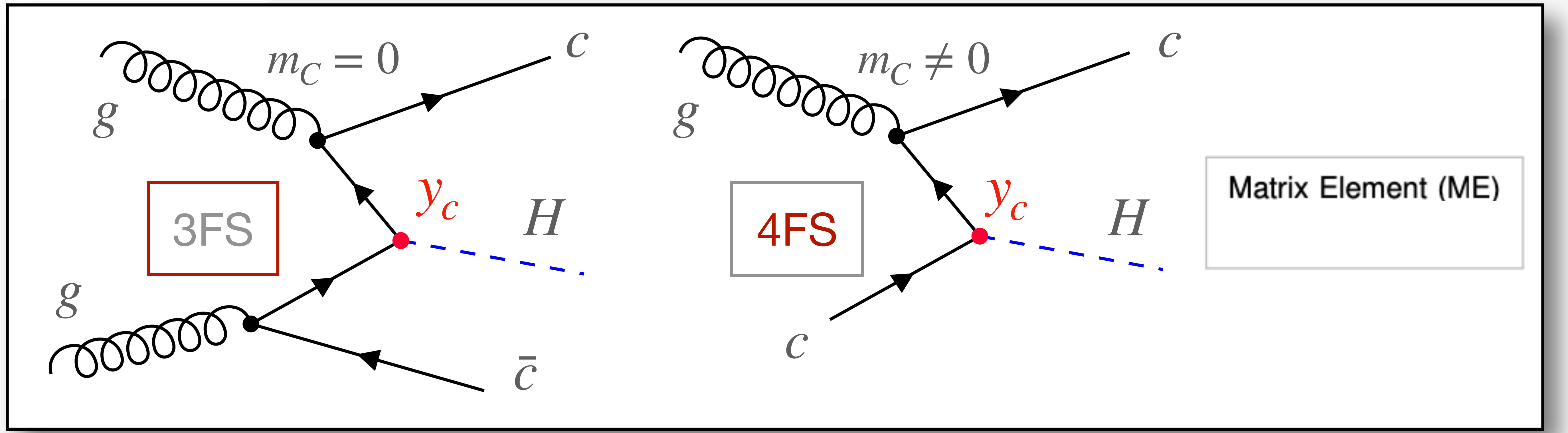


Higgs+charm simulation

Theoretical uncertainty studies:

- H+c MC ME calculation can be done in either 3FS or 4FS, i.e. considering the c massive or massless.
- In principle equivalent, in practice this may not be the case → several studies for bbH production (4FS vs 5FS) but no theory studies for $H + c$.
- To assess the additional theory uncertainty we compare samples produced using both methods:
  ⇒ FS uncertainty $O(30\%)$ of the yields in analysis categories.
- Different impact of the choice of simulation input parameters on the two FS:
  ⇒ study to decide the best choice of theory scales (μ_R, μ_F, and R_{sh}).
  ⇒ PDF ($\sim 5\,\%$) and Scale ($\sim 10\,\%$) uncertainties are less relevant compared to FS.

LO diagrams



Higgs+charm simulation

Theoretical uncertainty studies:

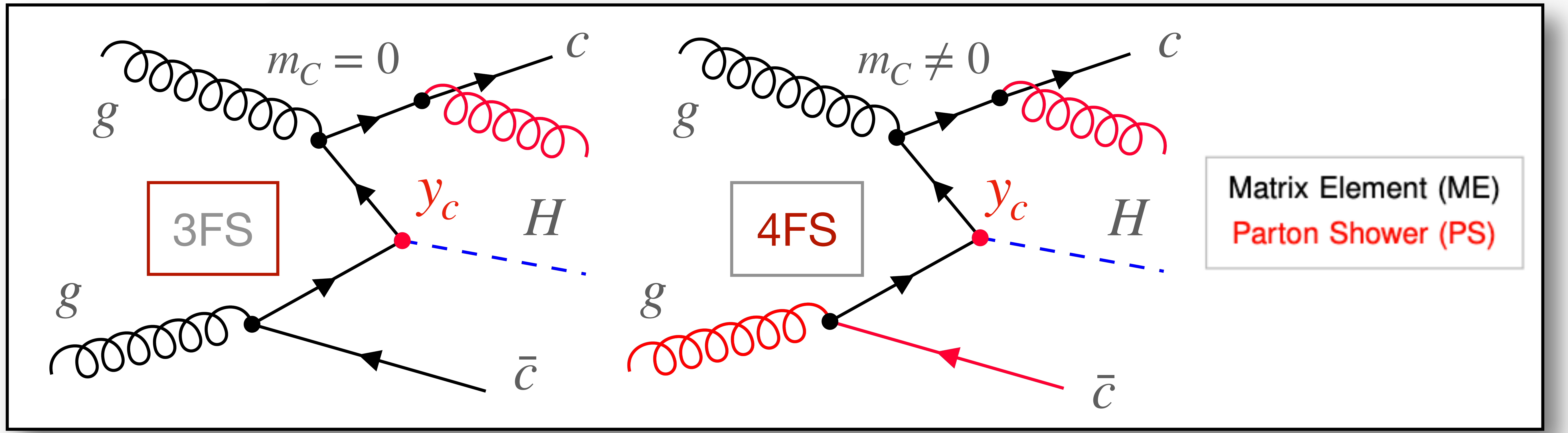
- H+c MC ME calculation can be done in either 3FS or 4FS, i.e. considering the c massive or massless.

- In principle equivalent, in practice this may not be the case → several studies for bbH production (4FS vs 5FS) but no theory studies for $H + c$.

- To assess the additional theory uncertainty we compare samples produced using both methods:
  ⇒ FS uncertainty $O(30\%)$ of the yields in analysis categories.

- Different impact of the choice of simulation input parameters on the two FS:
  ⇒ study to decide the best choice of theory scales (μ_R, μ_F, and R_{sh}).
  ⇒ PDF ($\sim 5\,\%$) and Scale ($\sim 10\,\%$) uncertainties are less relevant compared to FS.

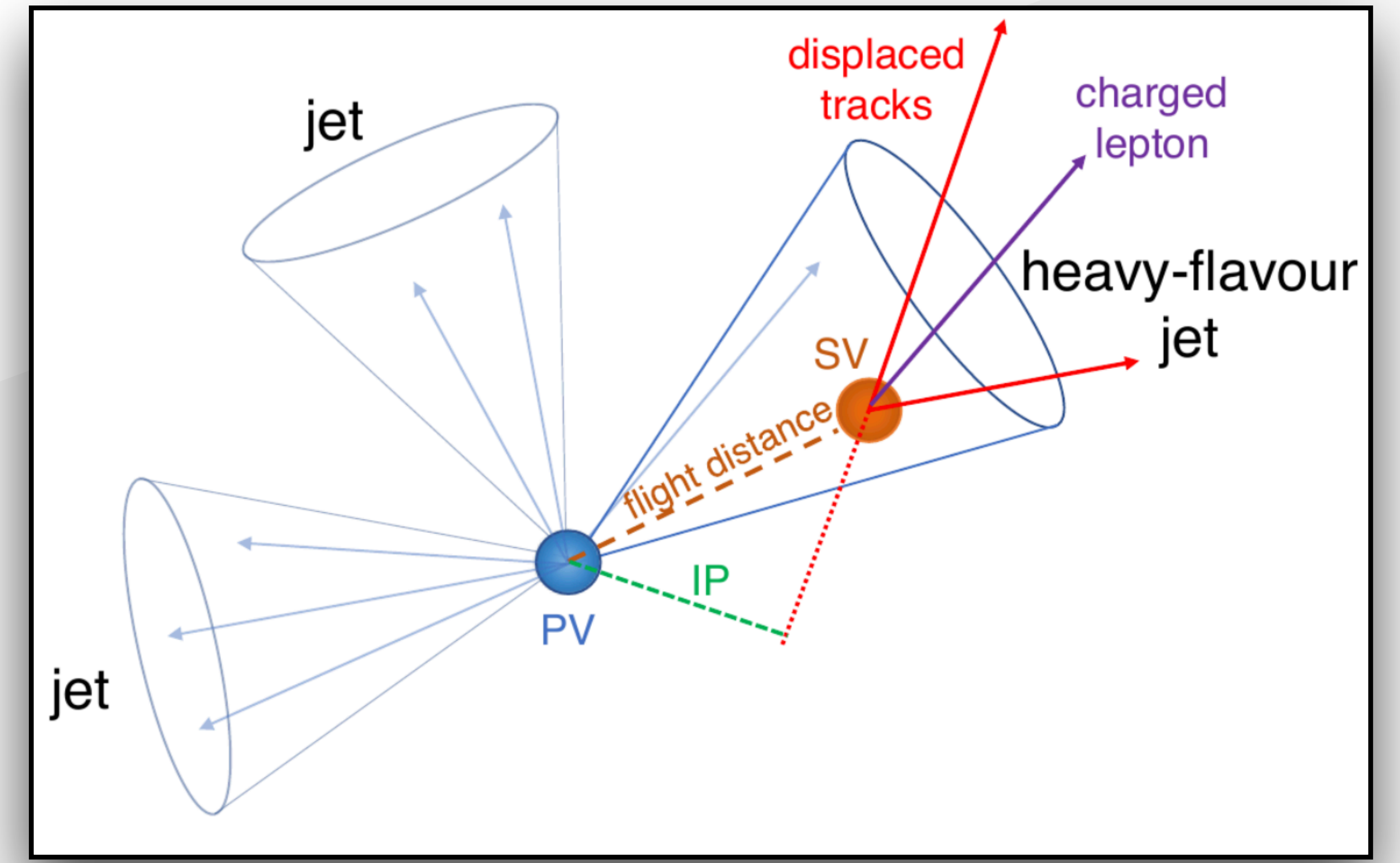
LO diagrams

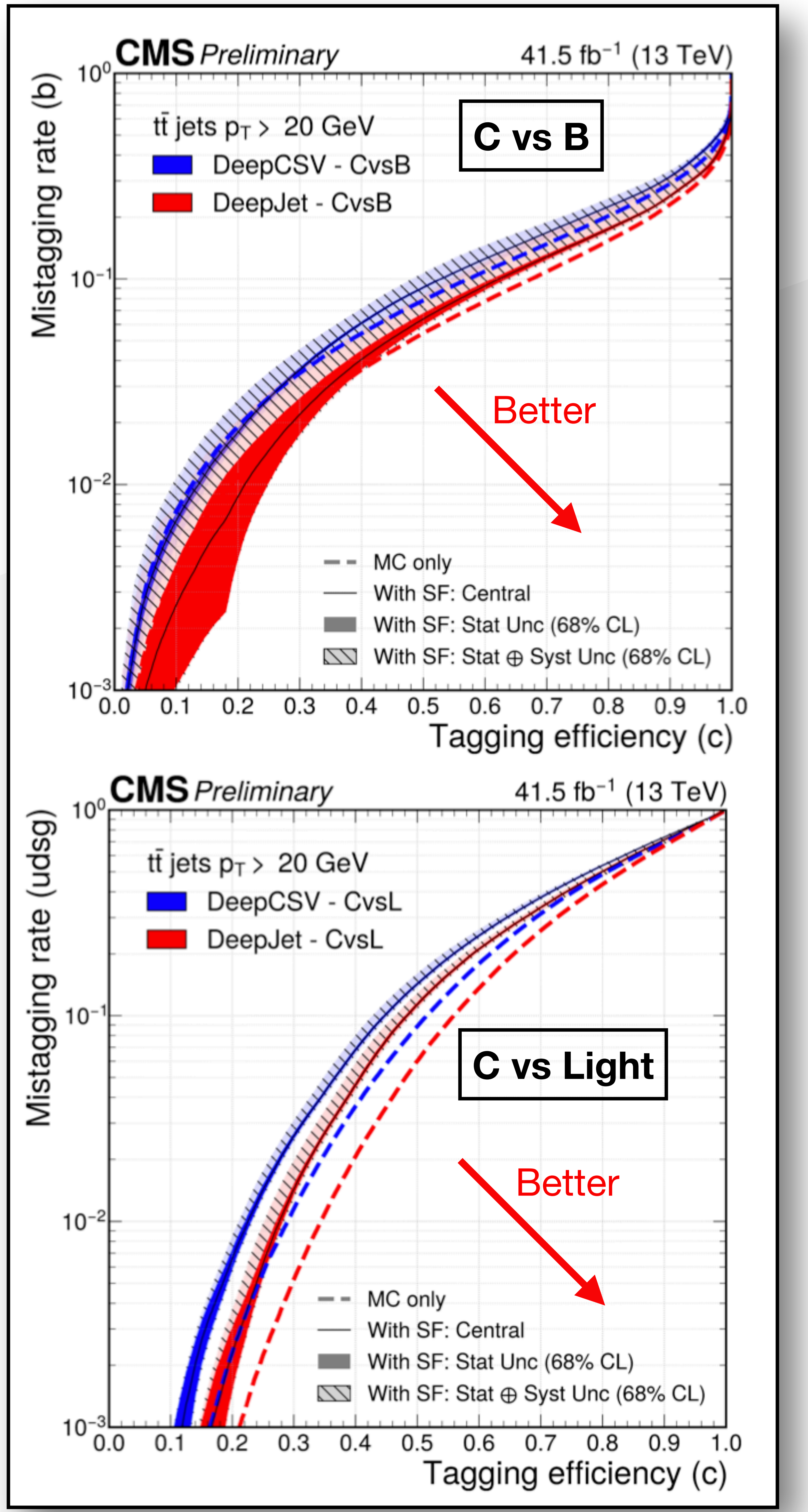


Charm tagging

DeepJet algorithm:

- A complex Neural Network (NN) based discriminator is used to identify charmed jets in CMS.
- It exploits more than 650 input variables from the global event and jet constituents.
- The identification has to discriminate c-jets both against b-jets and light jets.
- At Medium WP (2017): c-tag efficiency → 60%, b/light-mistag rate → 26%.

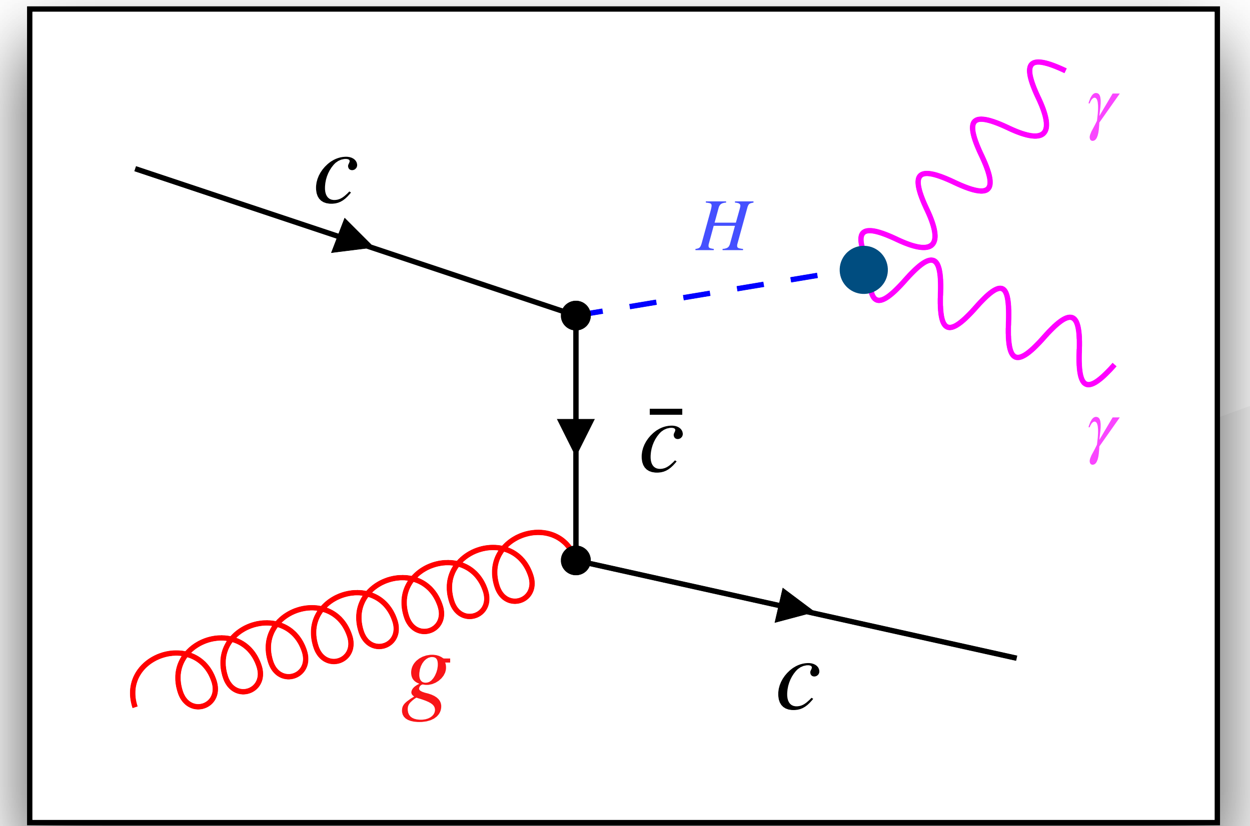


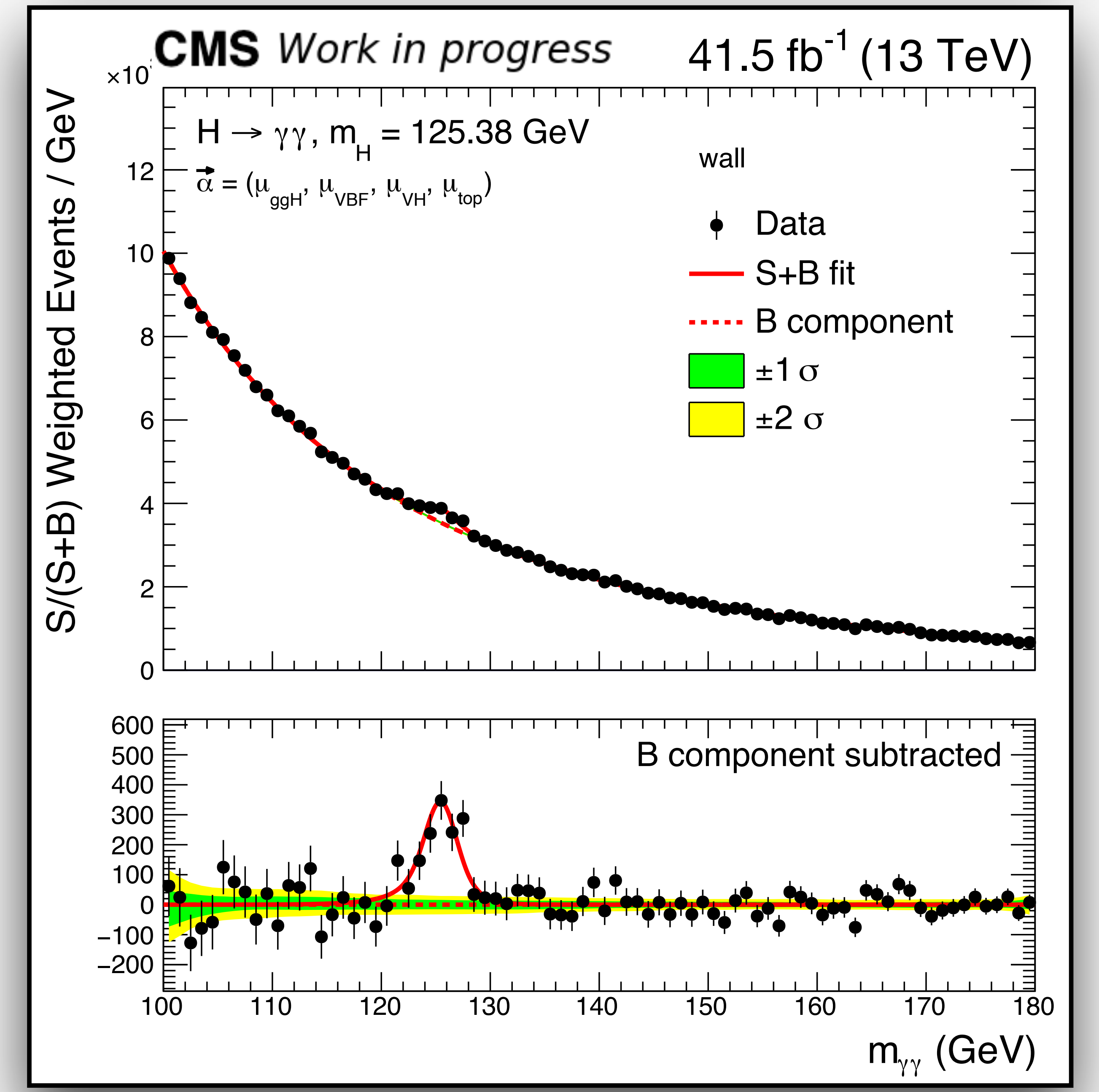


Higgs+charm analysis strategy

Diphoton decay channel:

- H+c coupling at production level ⇒ $H \to \gamma\gamma$ decay channel.
- We select events with at least one jet.
  - Main background is coming from “standard” Higgs production through gluon fusion (ggH) and continuous diphoton background (CB) from $\gamma\gamma$ and $\gamma + jets$ events.
- The analysis follows the standard $H \to \gamma\gamma$ analysis strategy.
  - Fit to the diphoton mass spectrum.
- We use the full Run 2 dataset (2016, 2017, 2018) of 137 fb^{-1}.



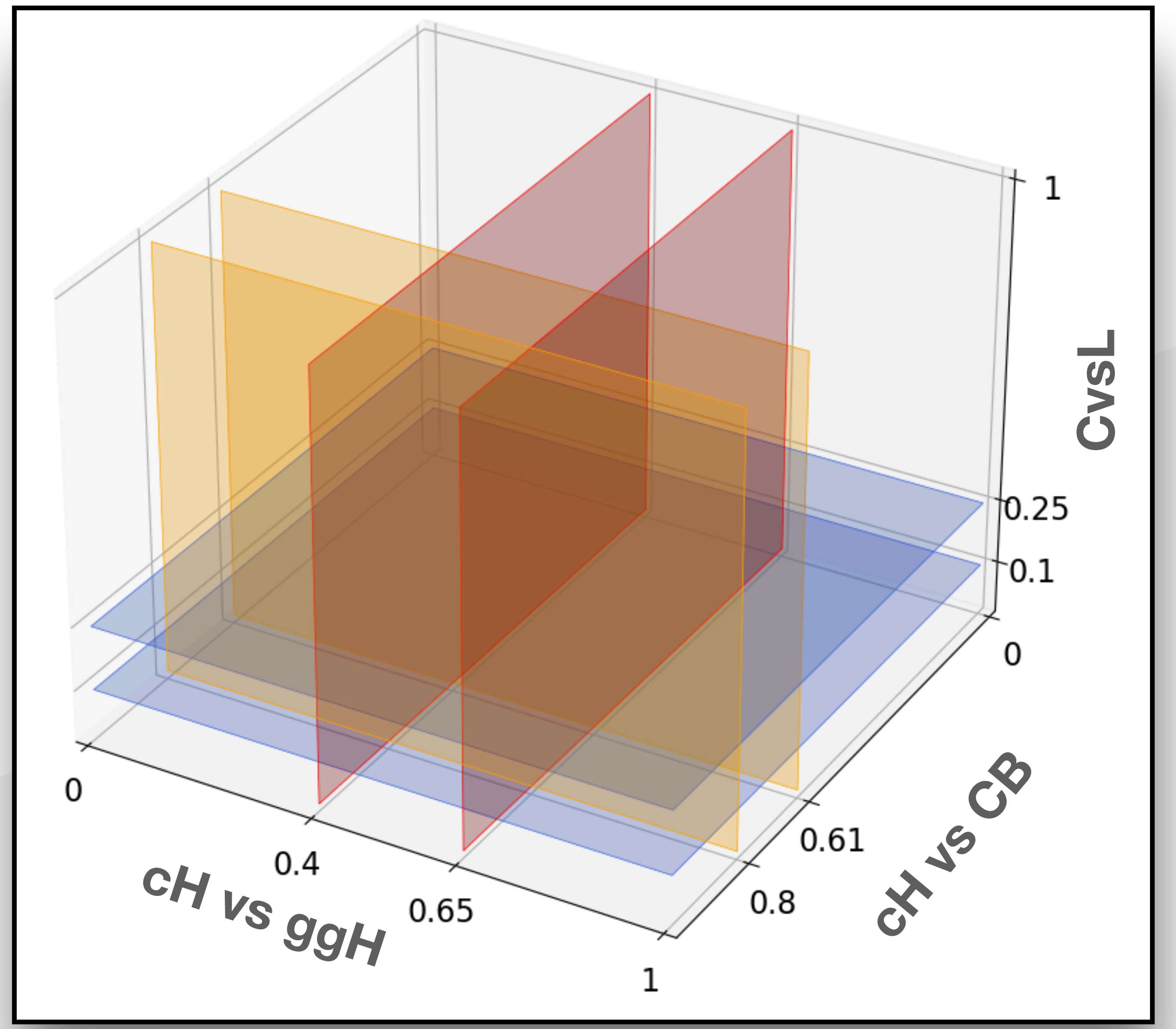


3D event categorisation

Categories:

- We separate events in 27 categories according to:
  - c-tagging score of the leading jet of the event.
  - A Boosted Decision Tree (BDT) trained to distinguish $H + c$ events from ggH.
  - A BDT trained to distinguish $H + c$ events from the continuous $\gamma\gamma$ background.

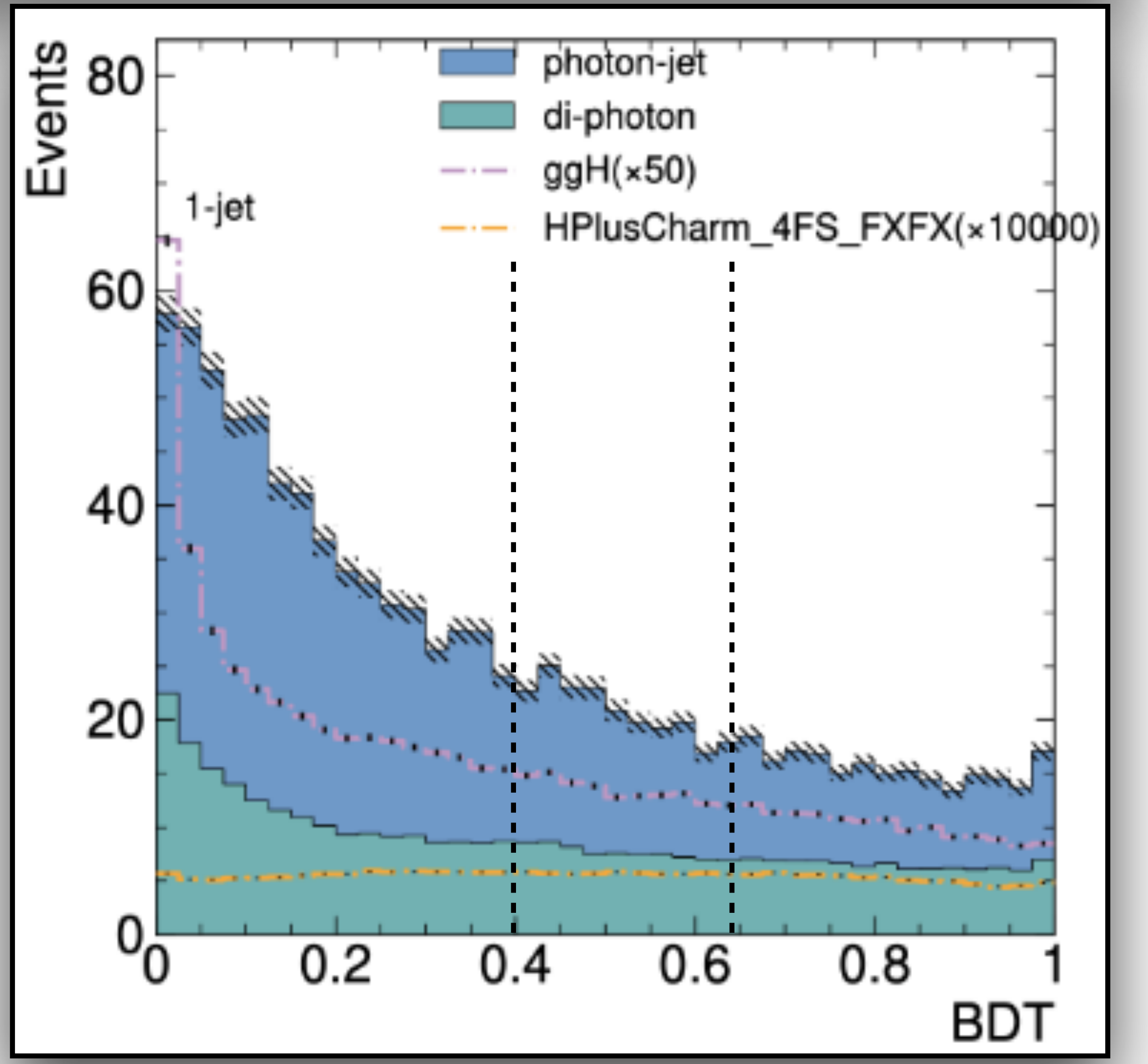
CATEGORISATION



BDT training:

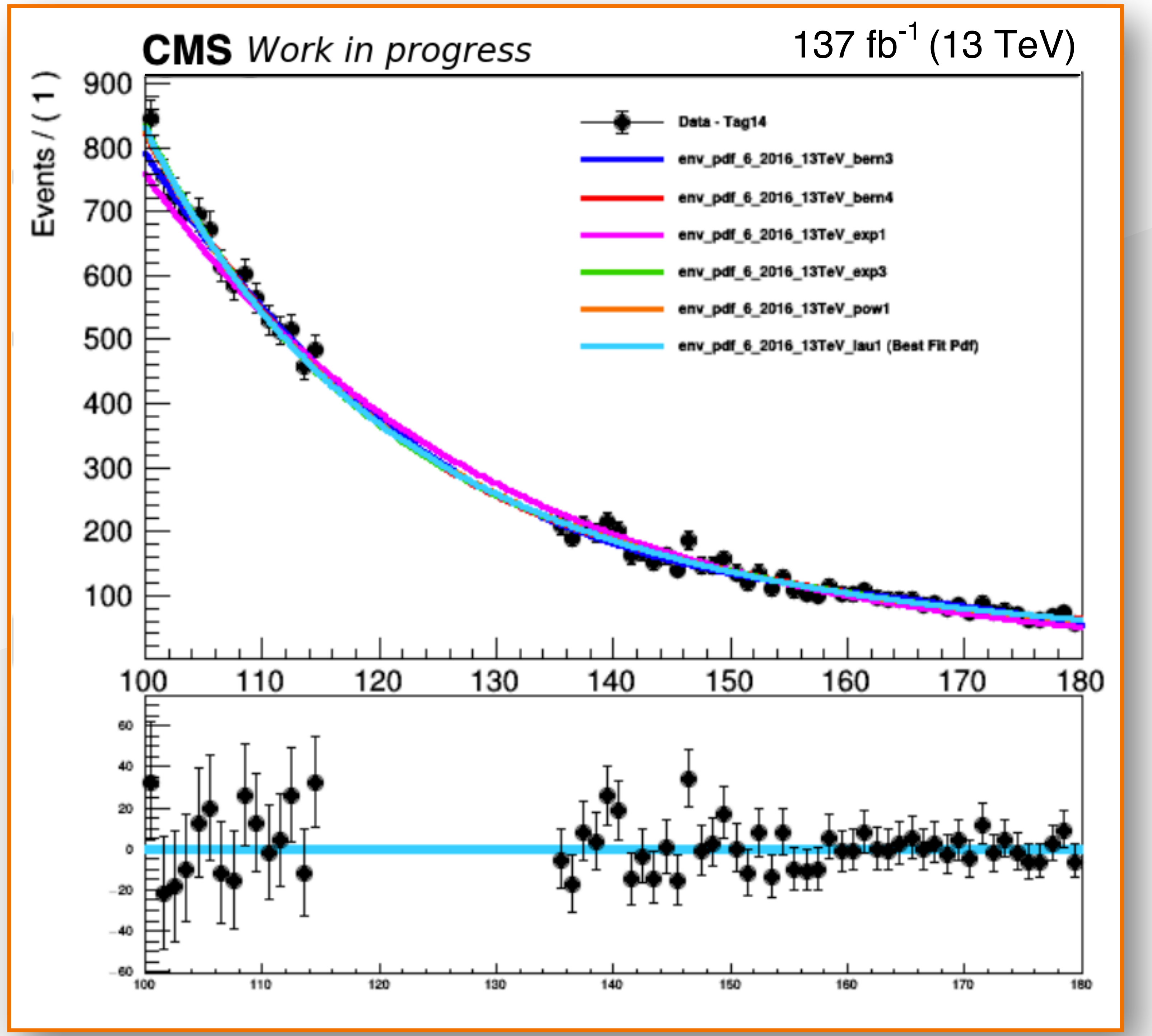
- Separation is achieved exploiting the kinematics of the Photons and Jets in the event.
- To address the irreducible ggH background we avoid using c-tagging information in the BDT training:
  $\Rightarrow$ use low BDT score regions to constrain ggH directly from data.

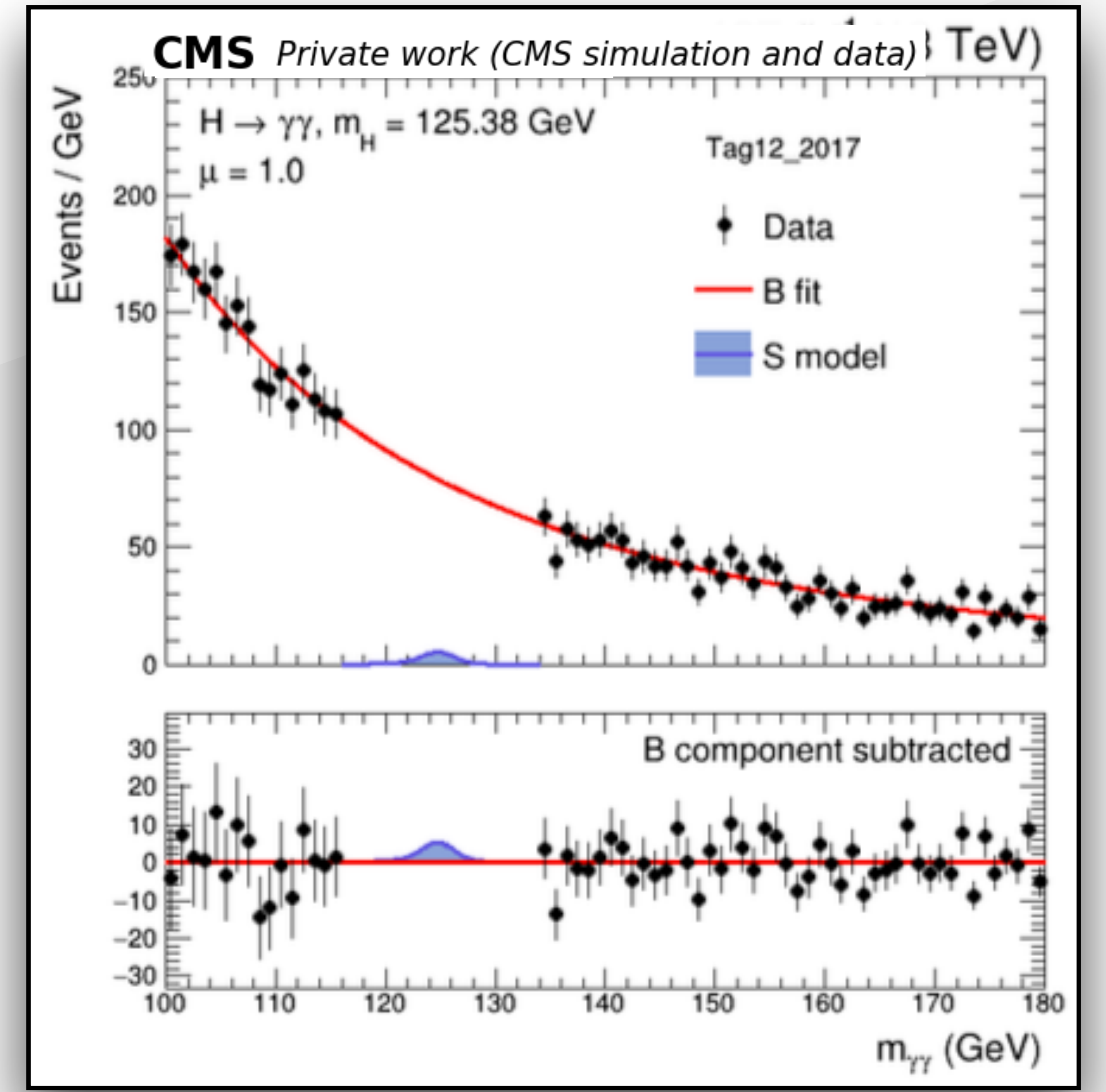
cH vs ggH BDT score



Statistical analysis

- Signal modelling:
  - Higgs related processes ($H + c$, ggH) are modelled by **fitting** a sum of gaussian to the **MC distribution**.
- Background modelling:
  - non-Higgs background is modelled in a **data driven way**, fitting the data sideband to extract the continuum $\gamma\gamma$ functional form.
- We then fit the diphoton mass distribution simultaneously over all categories.
- The expected limit on k_c is of $O(20)$.

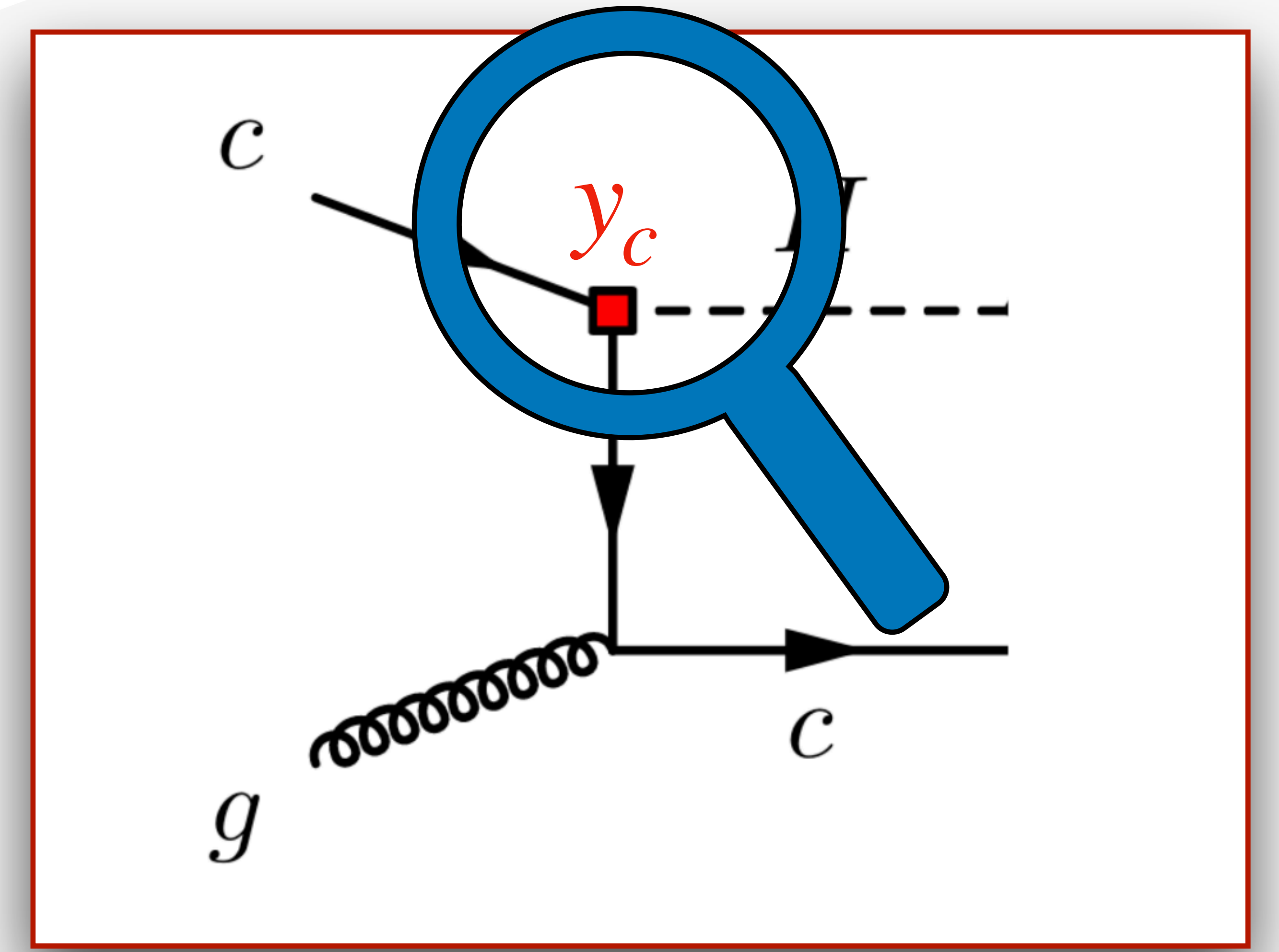




Conclusions

Overview:

- The strategy of the CMS $H + c$ associated production analysis has been presented.
- The analysis presents some challenges (MC simulation, charm-tagging…)
- Nonetheless it is a very interesting channel to explore, given the complementarity of this approach with other existing searches.
- The Run 2 analysis is still blinded, results coming soon.
- **Stay tuned!**

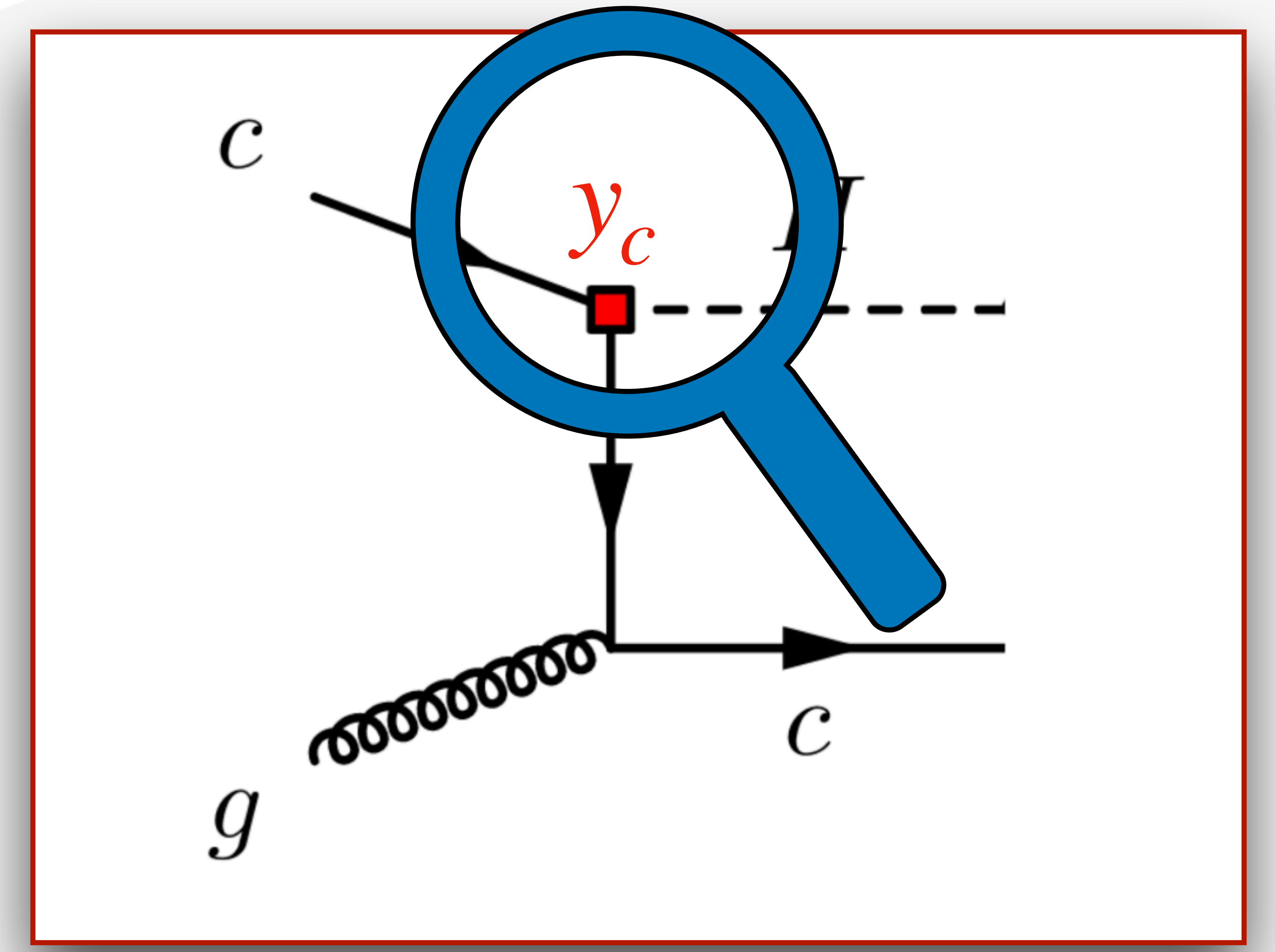


Conclusions

Overview:

- The strategy of the CMS $H + c$ associated production analysis has been presented.
- The analysis presents some challenges (MC simulation, charm-tagging…)
- Nonetheless it is a very interesting channel to explore, given the complementarity of this approach with other existing searches.
- The Run 2 analysis is still blinded, results coming soon.
- **Stay tuned!**

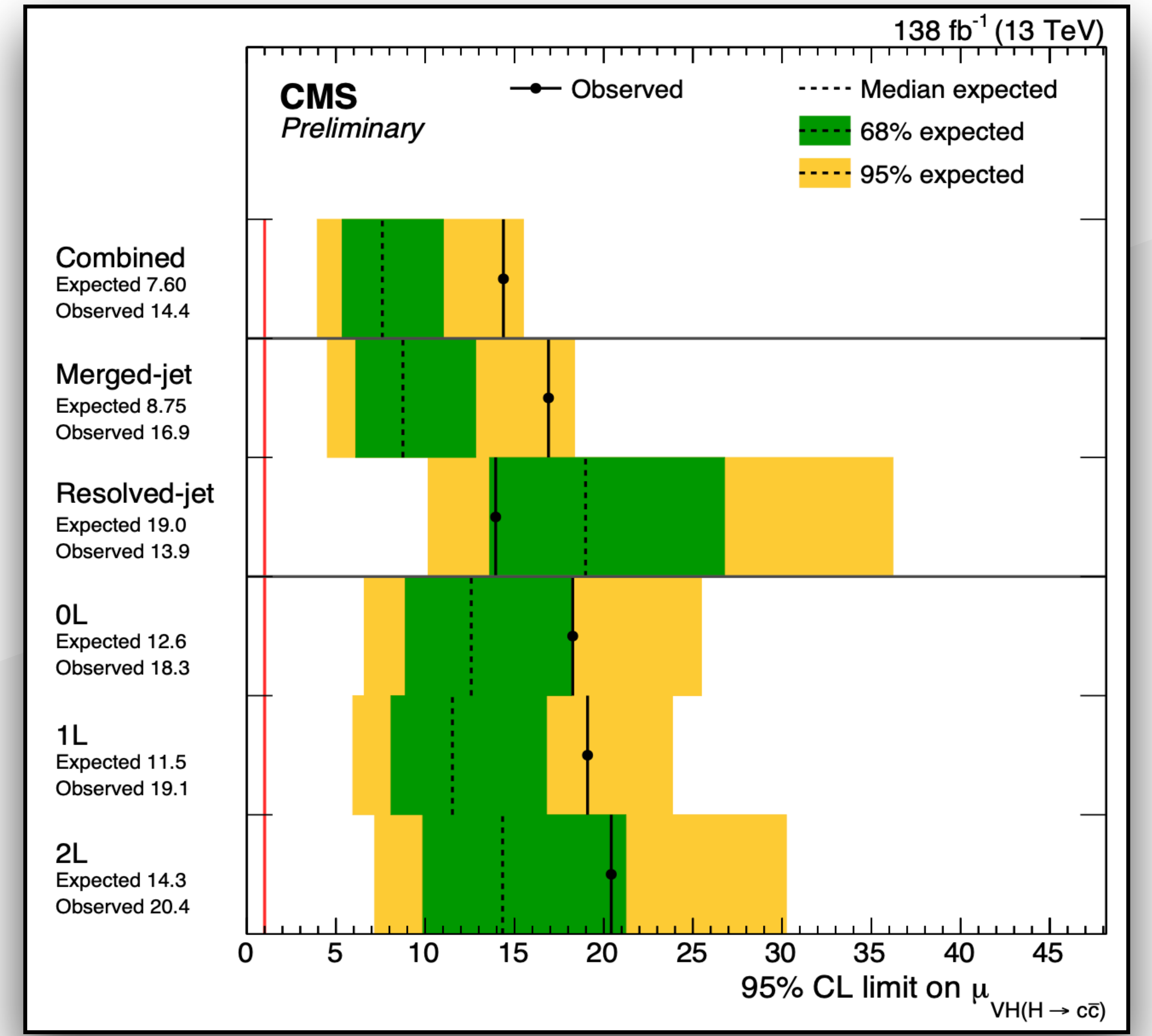
Thank you!



Back up

Previous results

- Direct search for $VH(H \to c\bar{c})$ arXiv:2205.05550: recent improvements, most stringent limit on $H \to c\bar{c}$.
  - Upper limit $\mu_{VH(H \to c\bar{c})} < 14\ (7.6)$ observed (expected).
  - $1.1 < |k_c^{[*]}| < 5.5\ (|k_c| < 3.4)$ observed (expected) at 95% C.L.
    [ATLAS : $|k_c| < 8.5 (12.4)$ obs (exp) at 95% C.L.]
  - First observation of $Z \to c\bar{c}$ at a hadron collider (5.7σ)
- Boosted $ggH(H \to c\bar{c})$ HIG-21-012:
  - $\mu < 38\ (45)$ observed (expected) at 95% C.L.
- Exclusive $H \to J/\Psi + \gamma$ decays, clean signature, $J/\Psi \to \mu\mu$ but very rare process:
  - $BR/BR_{SM} < 220\ (170)$ observed (expected) at 95% C.L.
    [ATLAS : proj. for $3\ ab^{-1}$ $\mu < \mu_{SM}$ at 95% C.L.]
- H differential measurements, variation of $p_T(H)$ as a function of k_c:
  - $-4.9 < k_c < 4.8\ (-6.1 < k_c < 6.0)$ observed (expected) at 95% C.L.

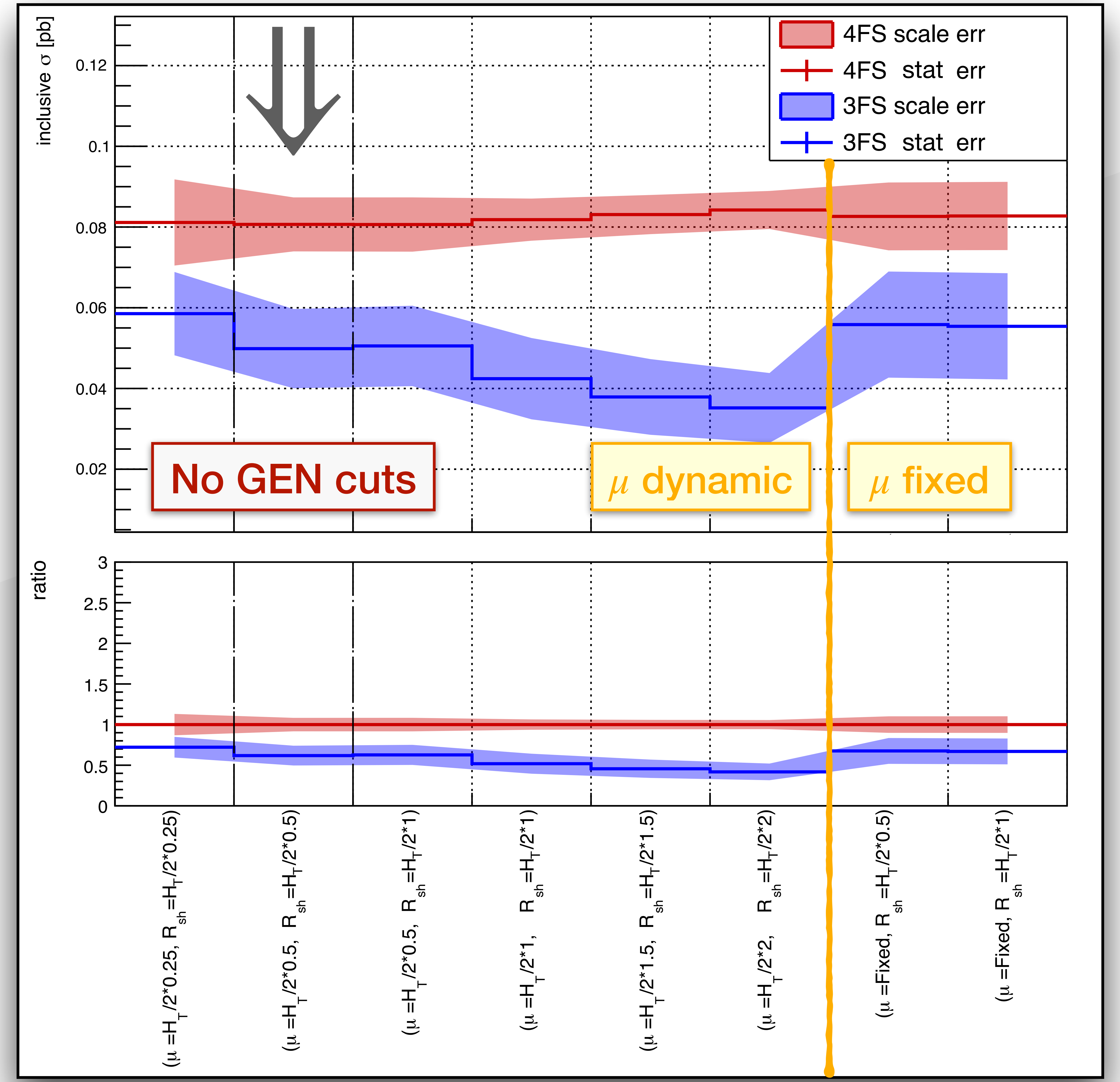


[*] $k_c = y_c / y_c^{SM}$

GEN-level results

$H+c$ cross section for different scale choices:

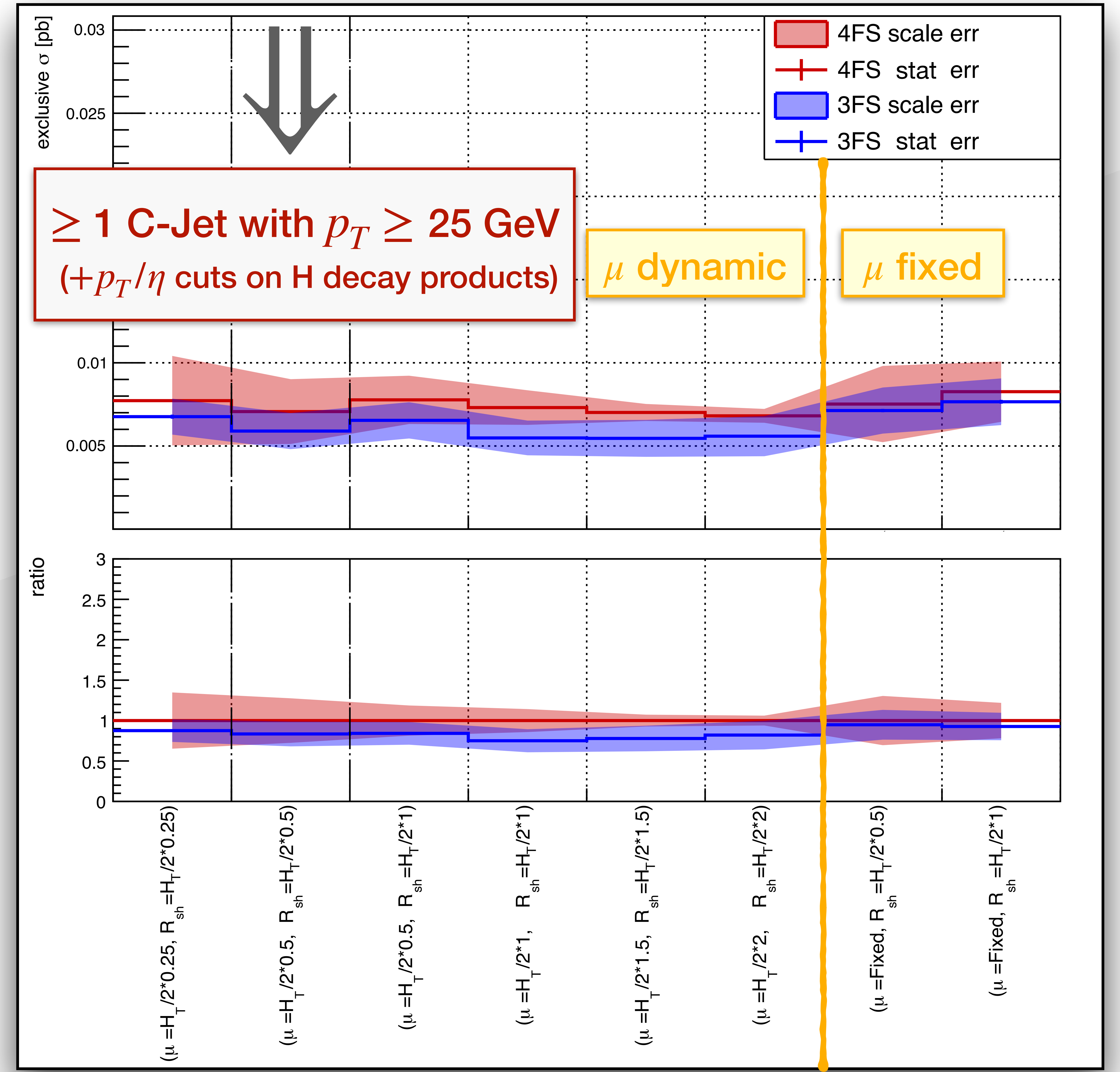
- Large differences between 3FS and 4FS on the inclusive cross sections (up to ~2x).
- To minimise uncertainty on nominal sample (4FS-FXFX) we studied the dependence of the X-section on MG scale parameters (μ_R, μ_F, R_{SH}).
- Uncertainties: μ_R/μ_F scale ~ 15%, PDF ~ 5-10%.



GEN-level results

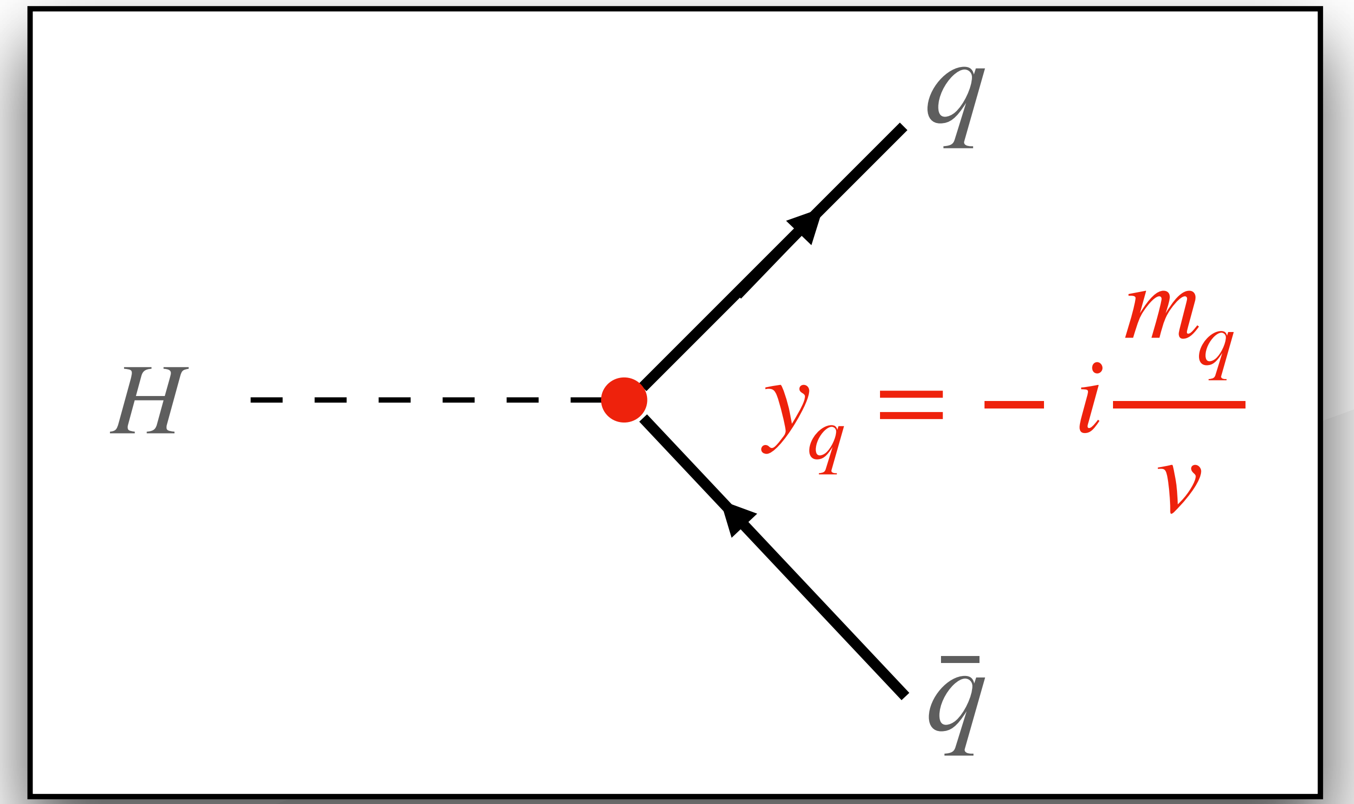
$H + c$ cross section for different scale choices:

- Large differences between 3FS and 4FS on the inclusive cross sections (up to ~2x).
- To minimise uncertainty on nominal sample (4FS-FXFX) we studied the dependence of the X-section on MG scale parameters (μ_R, μ_F, R_{SH}).
- Uncertainties: μ_R/μ_F scale ~ 15%, PDF ~ 5-10%.
- **Smaller (10-20%) differences for analysis-like phase space** (>=1 gen-c-jet w/ $p_T > 25$ GeV).

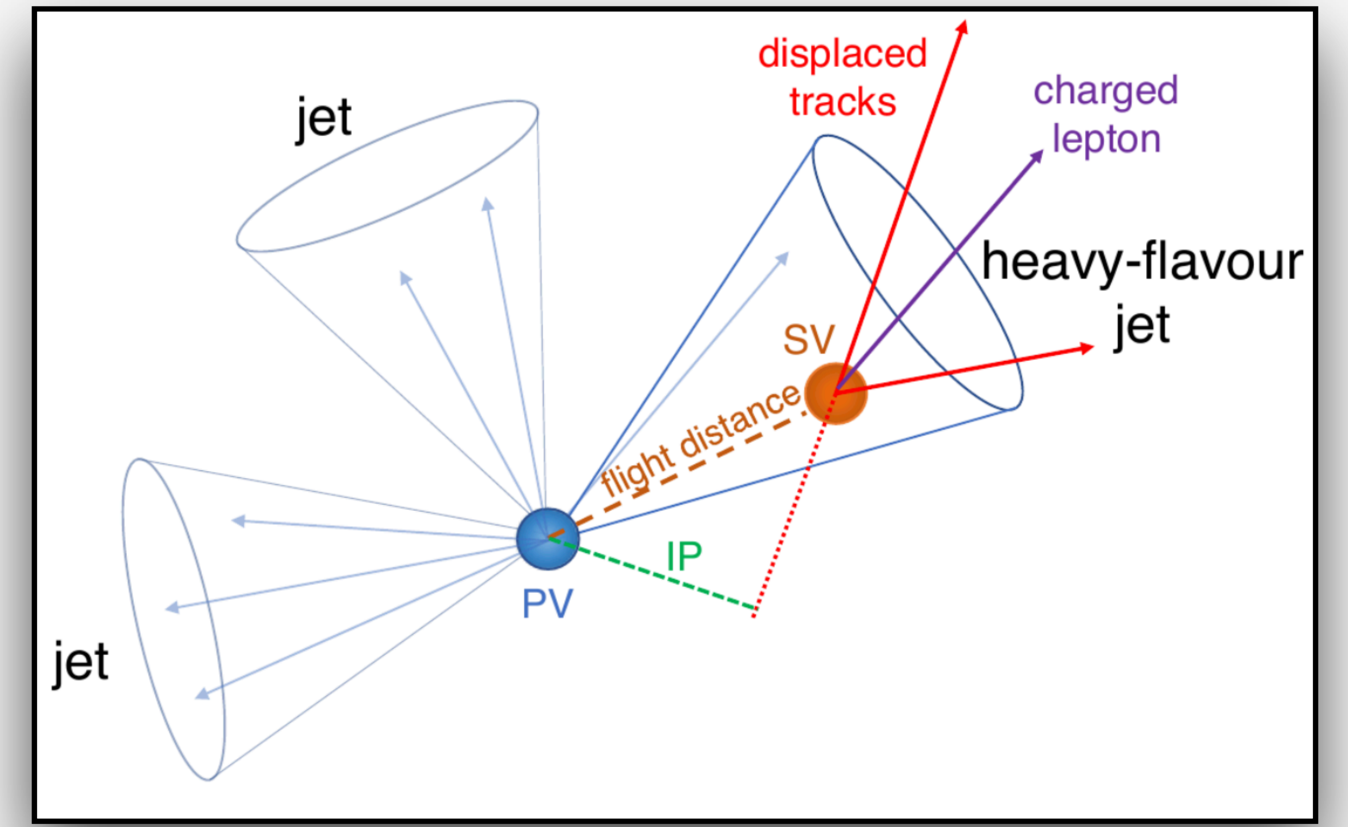


Challenges

- Decay width of Higgs boson to quarks $\Gamma_{H\to q\bar{q}} = 3\frac{m_q^2 \cdot m_H^2}{8\pi v^2}$
  - BR $H \to b\bar{b} \sim 58\,\%$
  - BR $H \to c\bar{c} \sim 3\,\%$



- Discriminate c-flavoured jets from background (b and light jets):
  - D-mesons lifetime ~ 1/2 of B-mesons, less SV displacement.
  - Discrimination wrt light jets more challenging than for b-jets.
  - B-mesons often have decay chains via D-mesons, which can fake c jets.



DeepJet

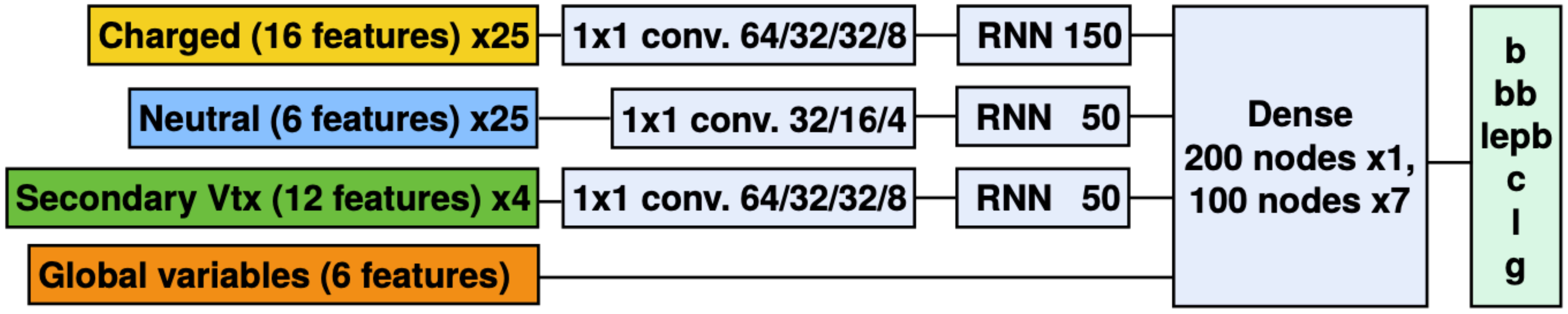
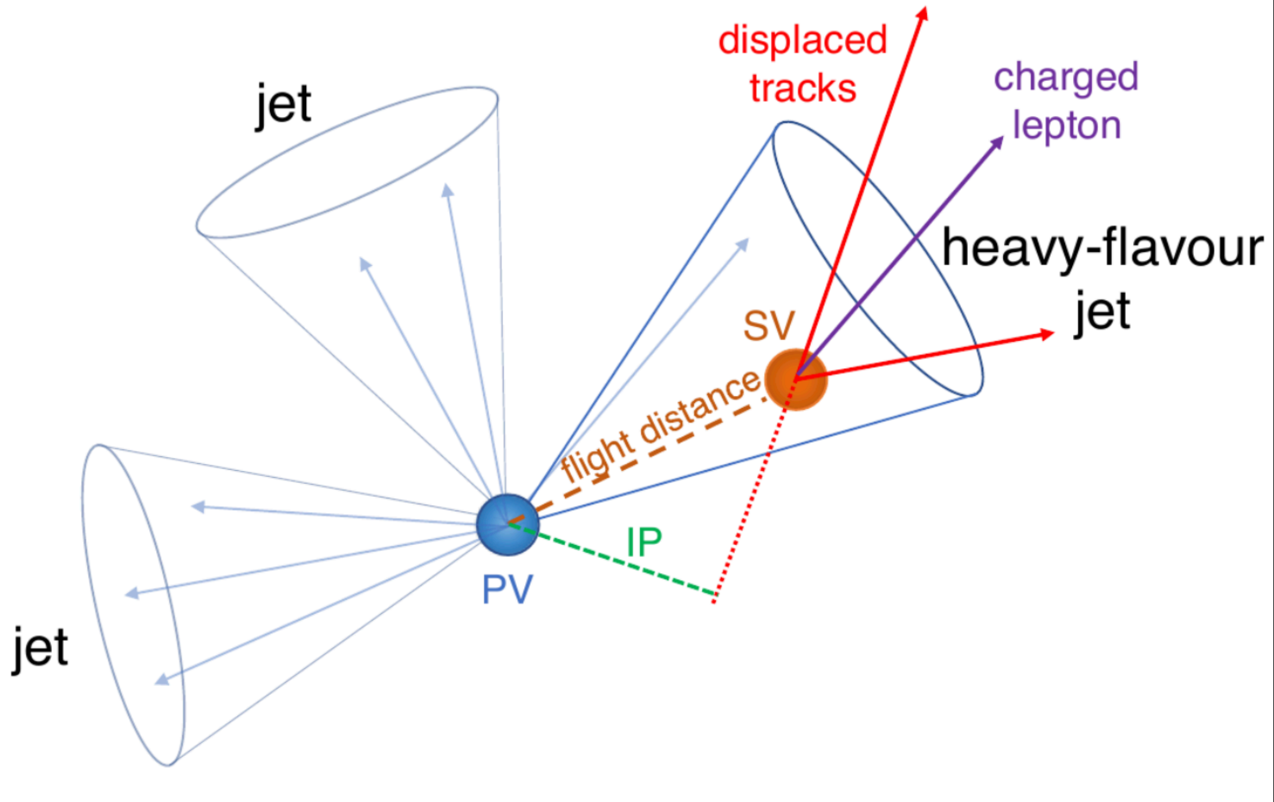


Figure 1. An illustration of the DeepJet architecture. Three seperate branches are used to process charged candidates, neutral candidates and secondary vertices. The algorithm makes use of 1x1 convolutional layers to perform automatic feature engineering for each class of jet constituents. The three RNN (LSTM) layers combine the information for each sequence of constituents. Finally the full jet information is combined using fully connected layers.



- The algorithm exploit more than 650 input variables associated with the jet, divided into four categories:
  - ⇒ global variables,
  - ⇒ charged candidate features,
  - ⇒ neutral candidate features,
  - ⇒ SV features.

Matching and Merging:

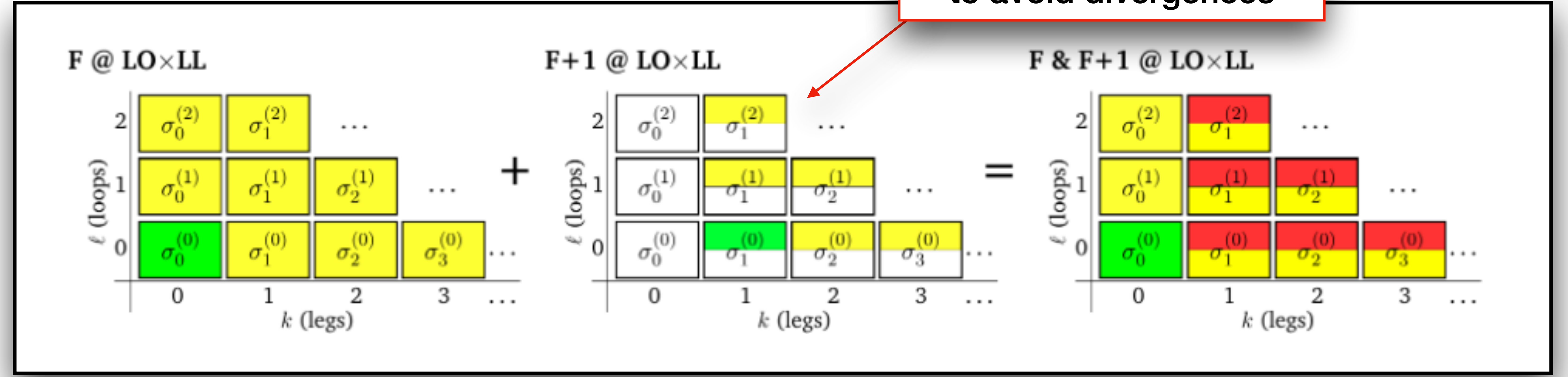
Merging (FXFX):

A separate tree-level calculation is performed for each parton multiplicity of interest. Soft and collinear divergences of the hard matrix elements are regulated by resolution cuts.

- Generate the process $[X]_{ME}$ + PS
- Generate the process $[X + 1\ jet]_{ME}$ + PS
- Generate the process $[X + 2\ jet]_{ME}$ + PS […]

Double counting !

Cuts on the ME emission to avoid divergences



Borrowed from L. Gellersen

Making exclusive by reweighting with no-emission probabilities, i.e. how would PS have produced this configuration, and using normal shower in “soft region” below q_{merg}.

ME fixed order calculations:

High momentum transfer interaction:

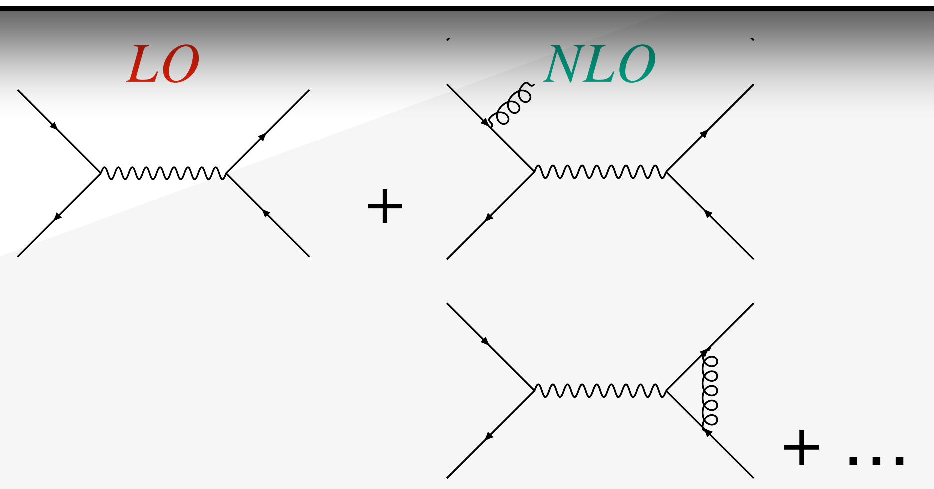
Partonic cross section for hard processes computed as Fixed Order (FO) Matrix Element calculation:

$$\hat{\sigma}_{ab\to X} = \alpha_s^n(\sigma_0 + \alpha_s \cdot \sigma_1)$$

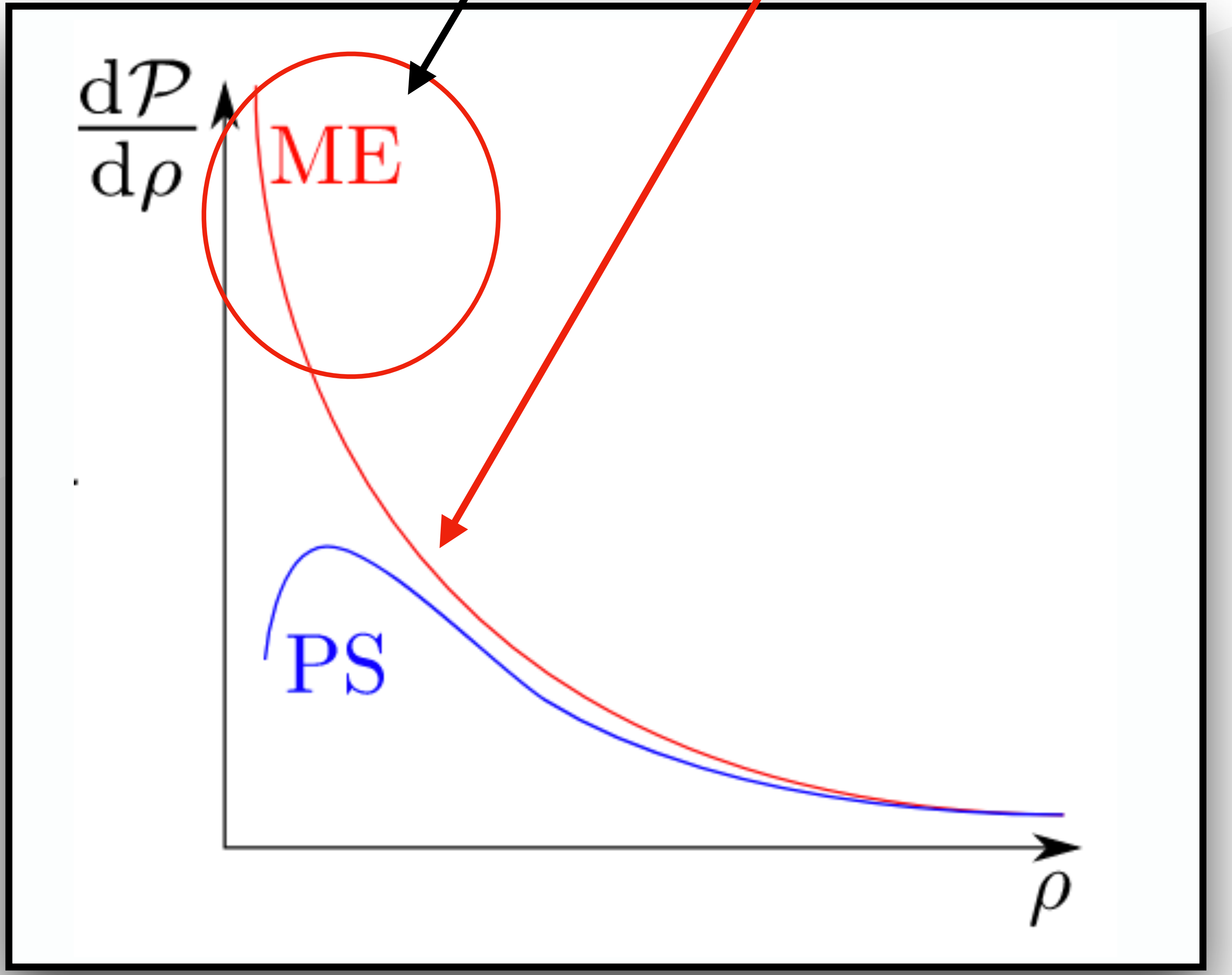
$$LO + NLO$$

Feasible calculation but some problems arise:

- Good description of hard jets.
- Contains all term in a given order of α_s.
- Valid also for high p_T.
- Only feasible for low multiplicities.
- Poor description of soft and collinear emissions.

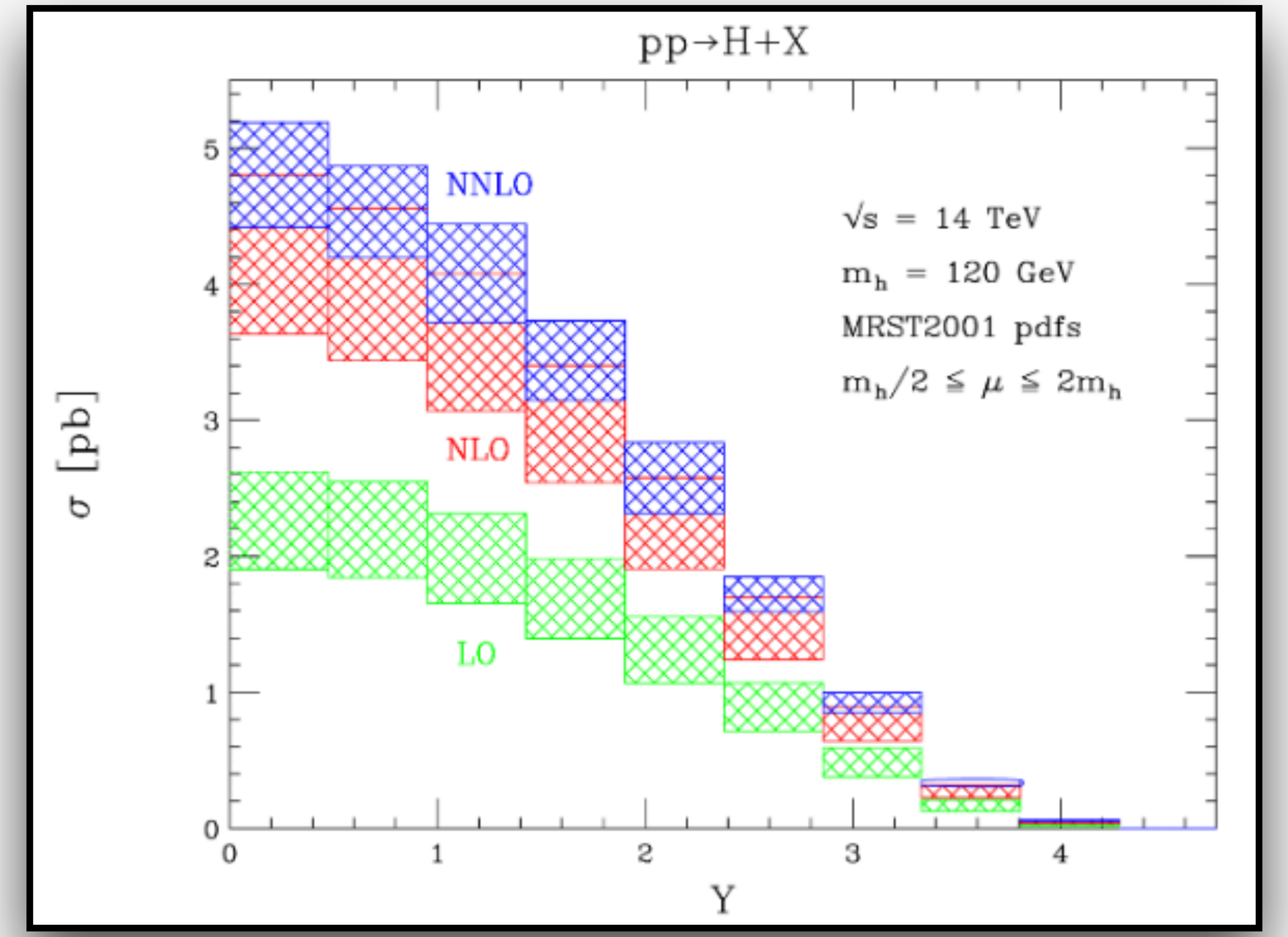


Divergences



Borrowed from L. Gellersen

Less precision



Anastasiou, Melnikov, Petriello, '05

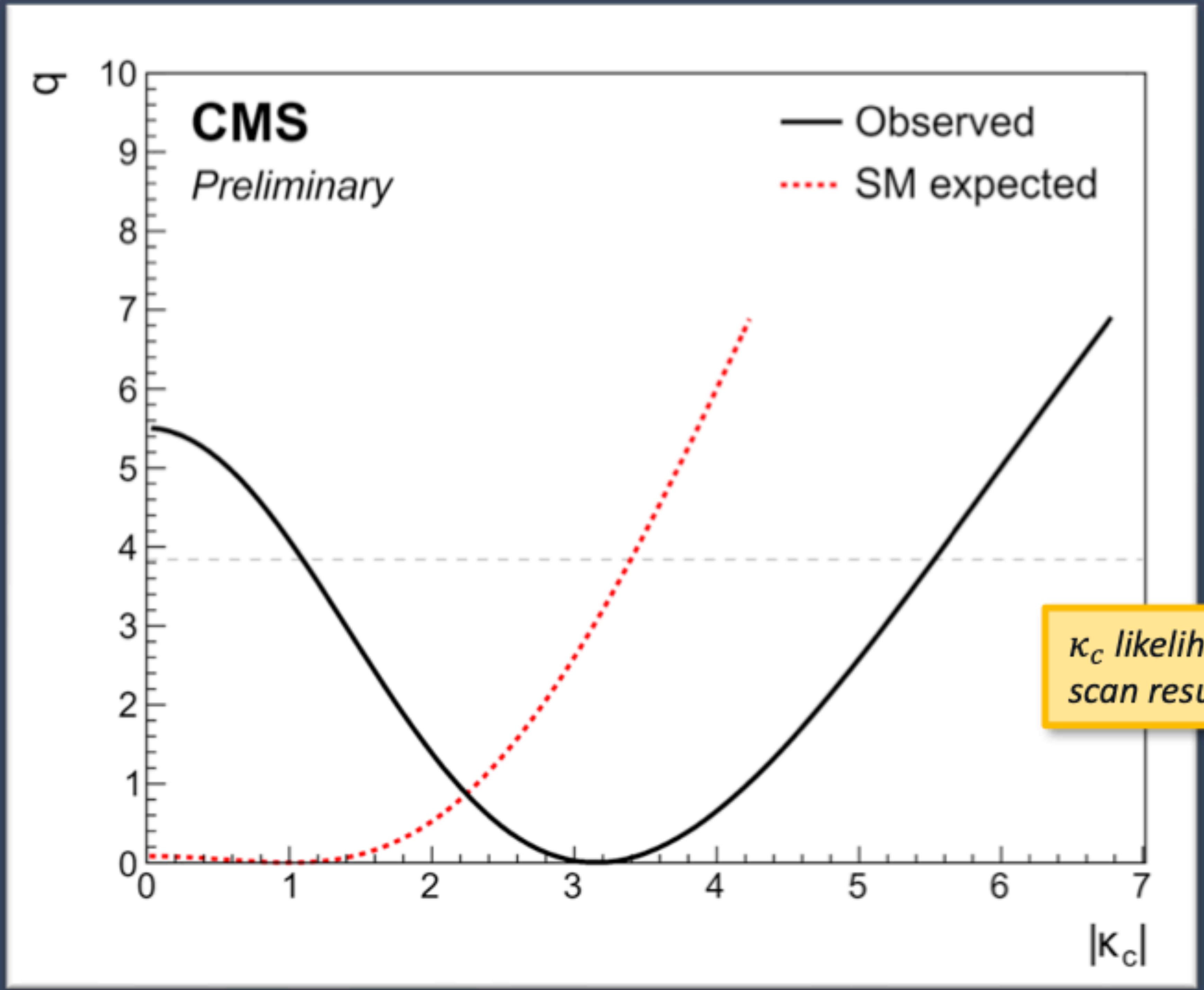
VH(Hcc): from Bjorn Burkle

Constraints on κ_c

- **Can use results to place new constraints on κ_c**
  - Only considering effects on BR($H \to cc$) and fixing all other couplings (same as ATLAS prescription for VHcc analysis)

$$\mu_{VH} = \frac{\kappa_c}{1 + \beta_{SM}(H \to cc) \times (\kappa_c - 1)}$$

- **Observed 95% CL: 1.1 < $|\kappa_c|$ < 5.5**
  - Obtained via likelihood scan
  - **Expected 95% CL: $|\kappa_c|$ < 3.4**
  - ALTAS obs. (exp.) $|\kappa_c|$ < 8.5 (12.4)
- **Strongest limit to date!**
  - Now beating constraints from indirect measurements
  - Comparable to ATLAS projection for HL-LHC: [ATL-PHYS-PUB-2021-039]



κ_c likelihood scan results

FS

- While 4FS results lack logarithmic terms beyond the first few, 5FS results lack power-suppressed terms (mb/Q)n. Which of the two classes of terms is more important depends on the observable studied, that determines the dominant kinematic regime.
- If logarithms are large, the 5FS should be superior to the 4FS; if they are not, and thus power-suppressed terms might be important, then 4FS approaches should be preferred.
- One expects that, for processes and in regions of the phase space where both resummation and mass effects are not dominant, the two approaches should give similar results.

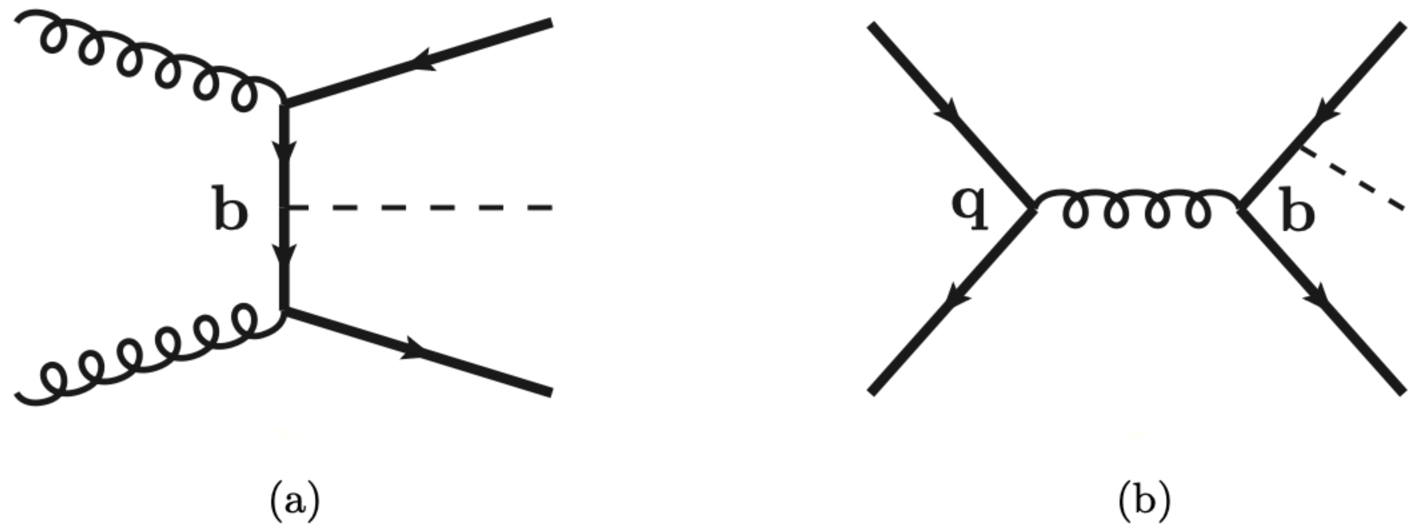


Figure 1: Sample of LO Feynman diagrams for $b\bar{b}H$ production in the four-flavour scheme, for the two relevant classes of partonic subprocesses: (a) $gg \to b\bar{b}H$; (b) $q\bar{q} \to b\bar{b}H$.

from: arXiv:1409.5301v2

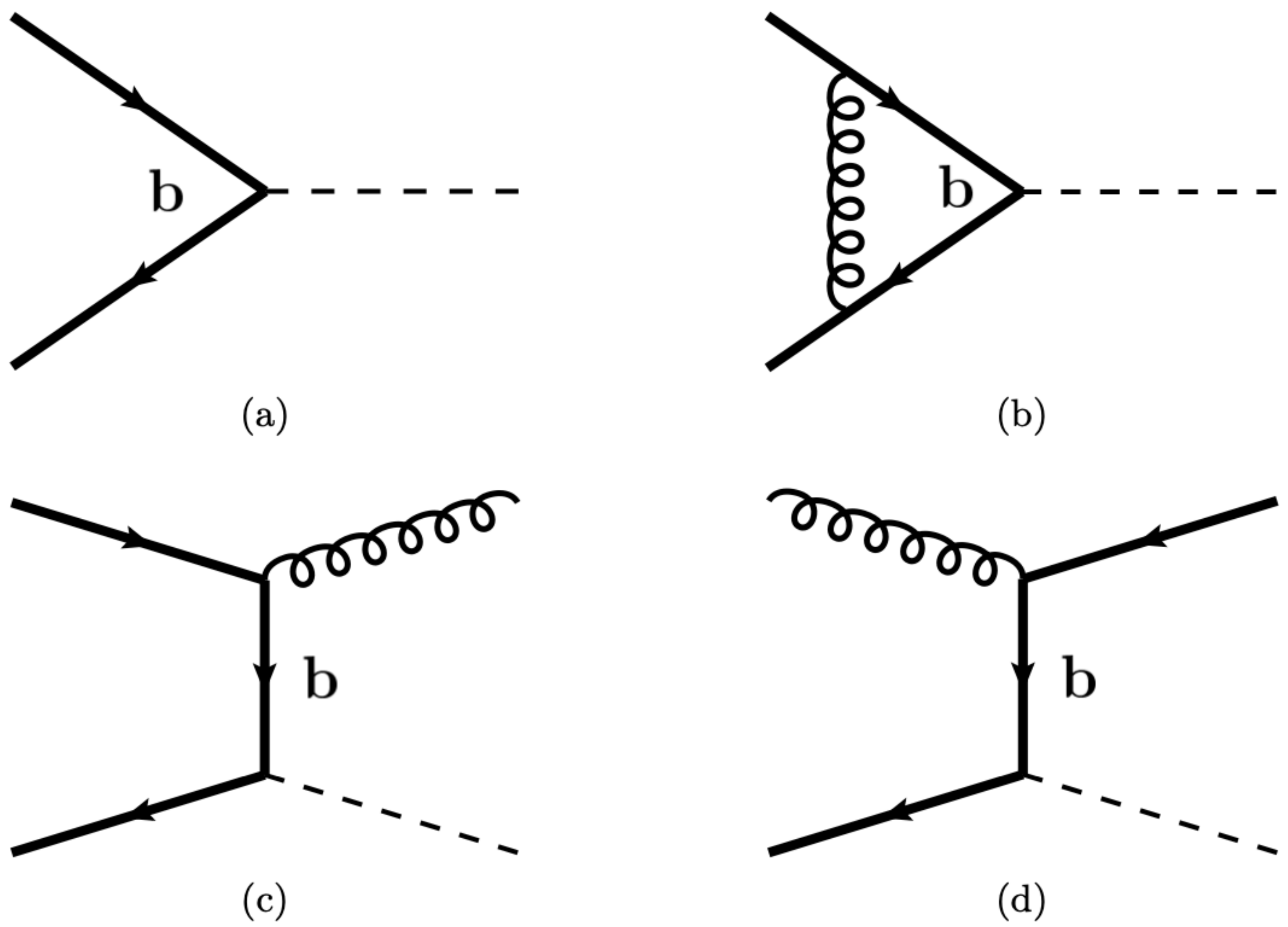


Figure 4: A sample of Feynman diagrams for $b\bar{b}H$ production in the five-flavour scheme: (a) LO; (b) one-loop; (c-d) real emission.