Searching for Higgs+charm production in the $\gamma\gamma$+c final state at CMS

Tiziano Bevilacqua

PSI LTP Seminar

11-20-23

Outline

Introduction:

- Motivation.
- State of the art.
- New approaches.

Ingredients:

- Monte Carlo simulation.
- $H \rightarrow \gamma\gamma$ method.
- Framework development.
- Charm tagging.

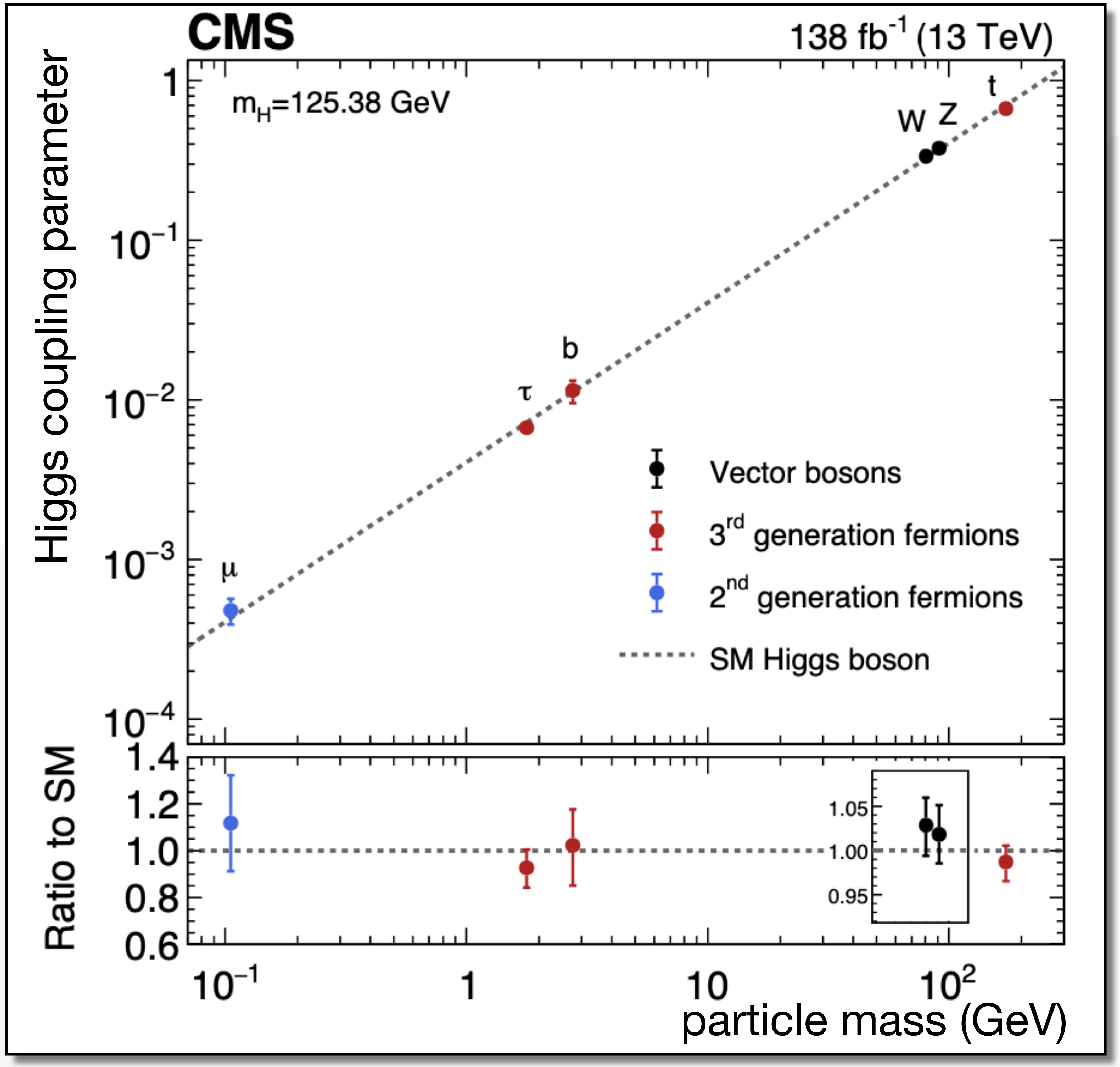
H+c analysis status:

- Analysis strategy.
- Preliminary results.

Motivation

The Higgs boson:

- Last addition to the Standard Model:
    - Theorised in ’64, discovered in 2012 🏅 at the LHC.
- It has a unique role in the SM:
    - **Important benchmark of the theory**.
    - Responsible for the mechanism that gives masses to elementary particles.
- Precise measurements of the Higgs couplings is of crucial importance:
    - The SM is amazing, but it’s an incomplete theory.
    - The Higgs sector could be a **bridge to** understanding **new physics.**

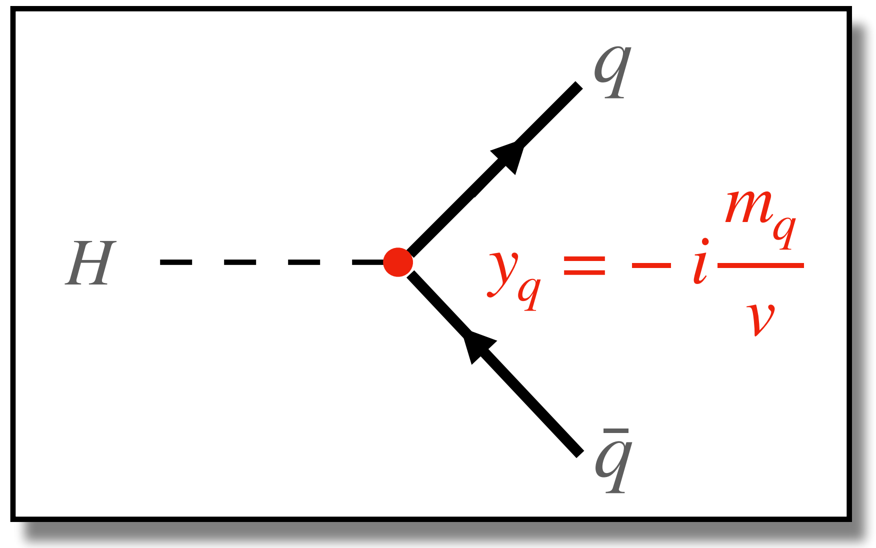


Nature 607, 60–68 (2022)

Motivation

Higgs couplings:

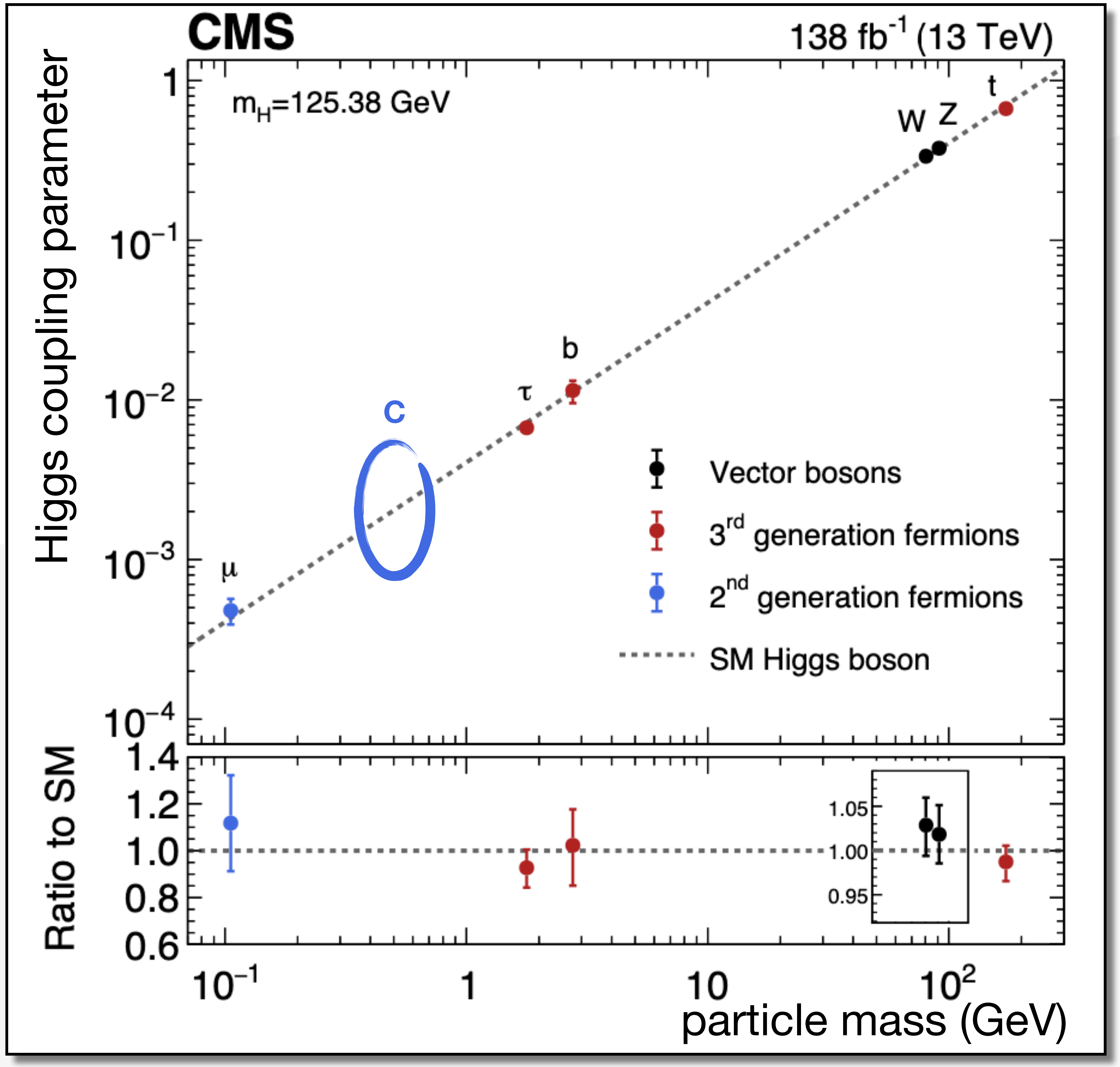
- Couplings to other particles are precisely predicted and **proportional to the particle masses**:



  - BR $H \rightarrow b\bar{b} \sim 58\,\%$
  - BR $H \rightarrow c\bar{c} \sim 3\,\%$

- **3rd generation** couplings already measured, 1 to 3 orders of magnitude bigger than 2nd 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**?

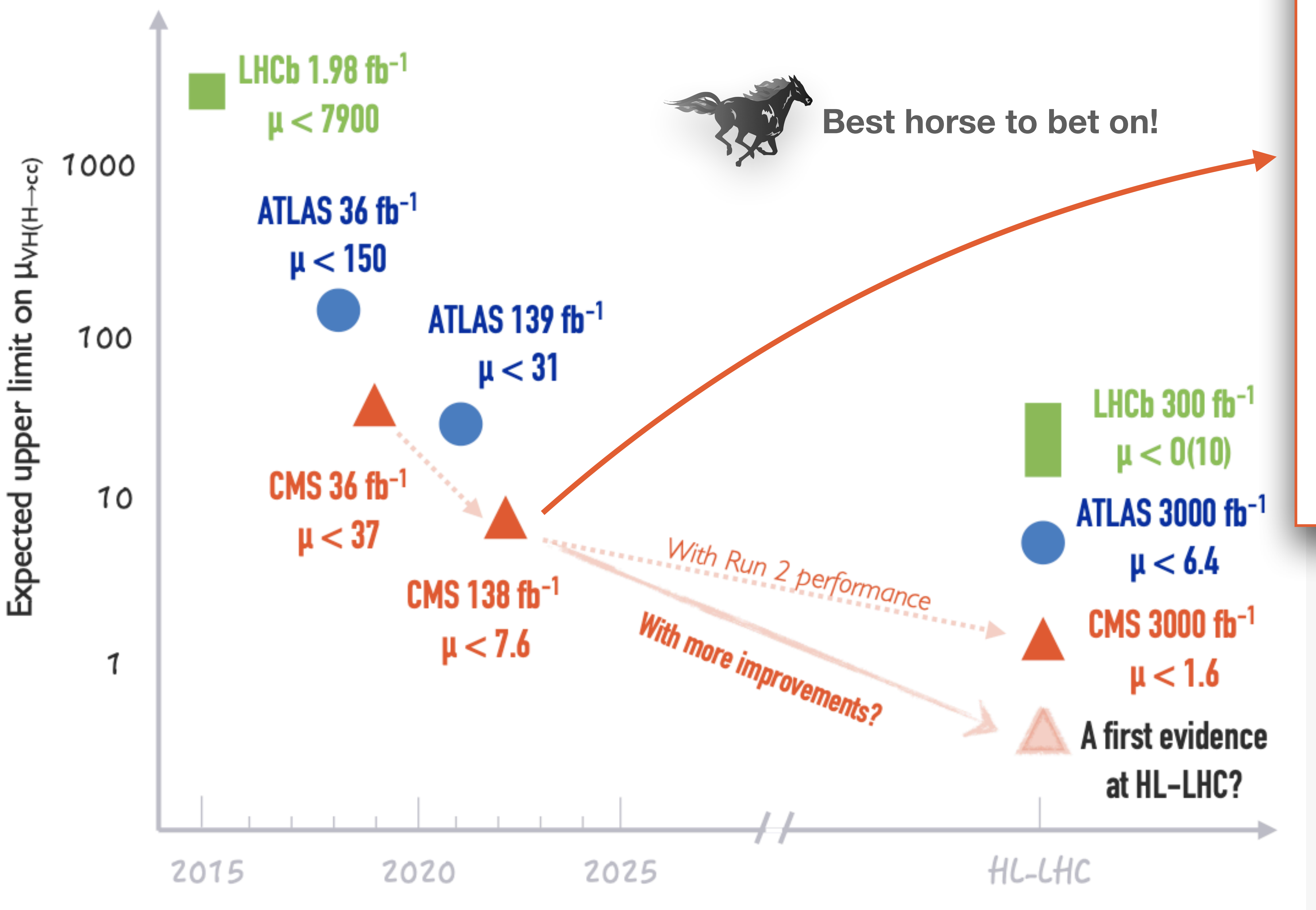


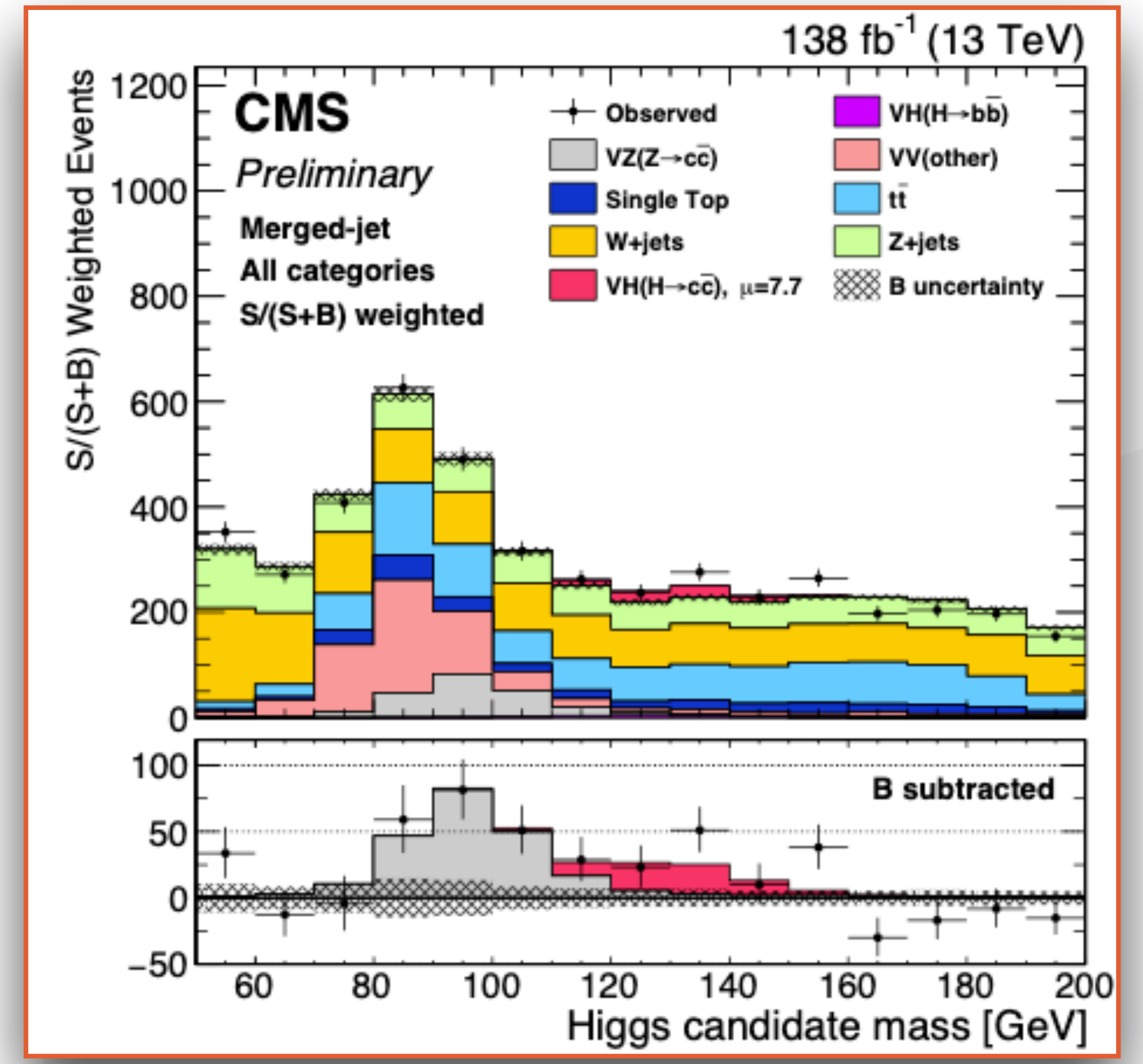
Nature 607, 60–68 (2022)

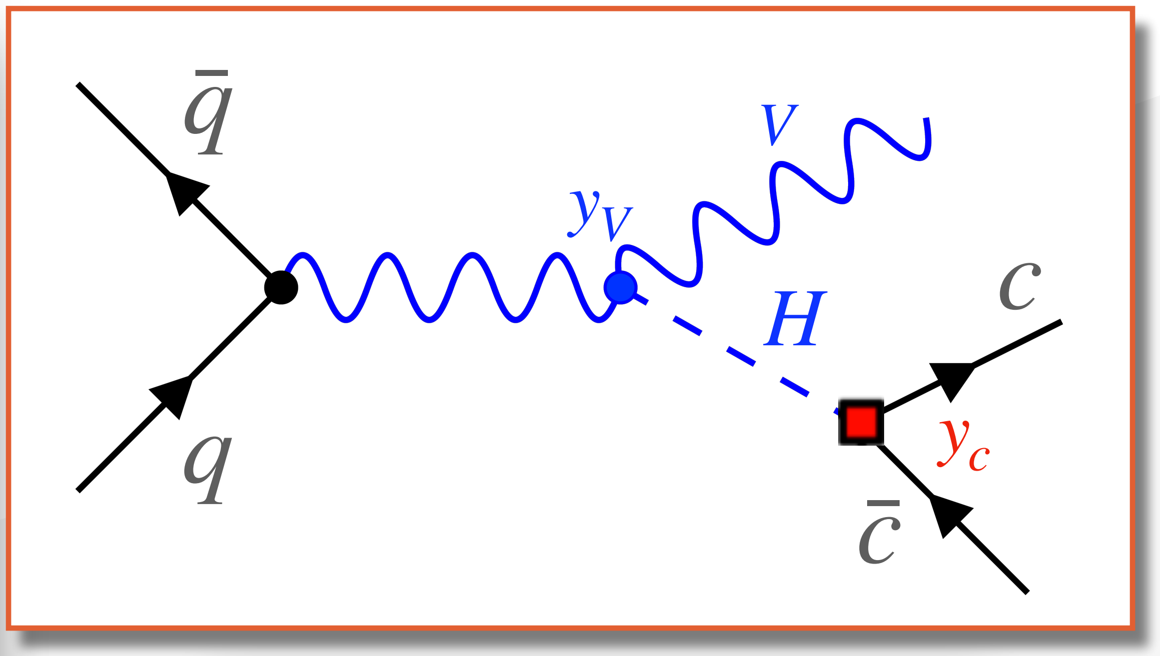
State of the art: Previous results

- **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 …

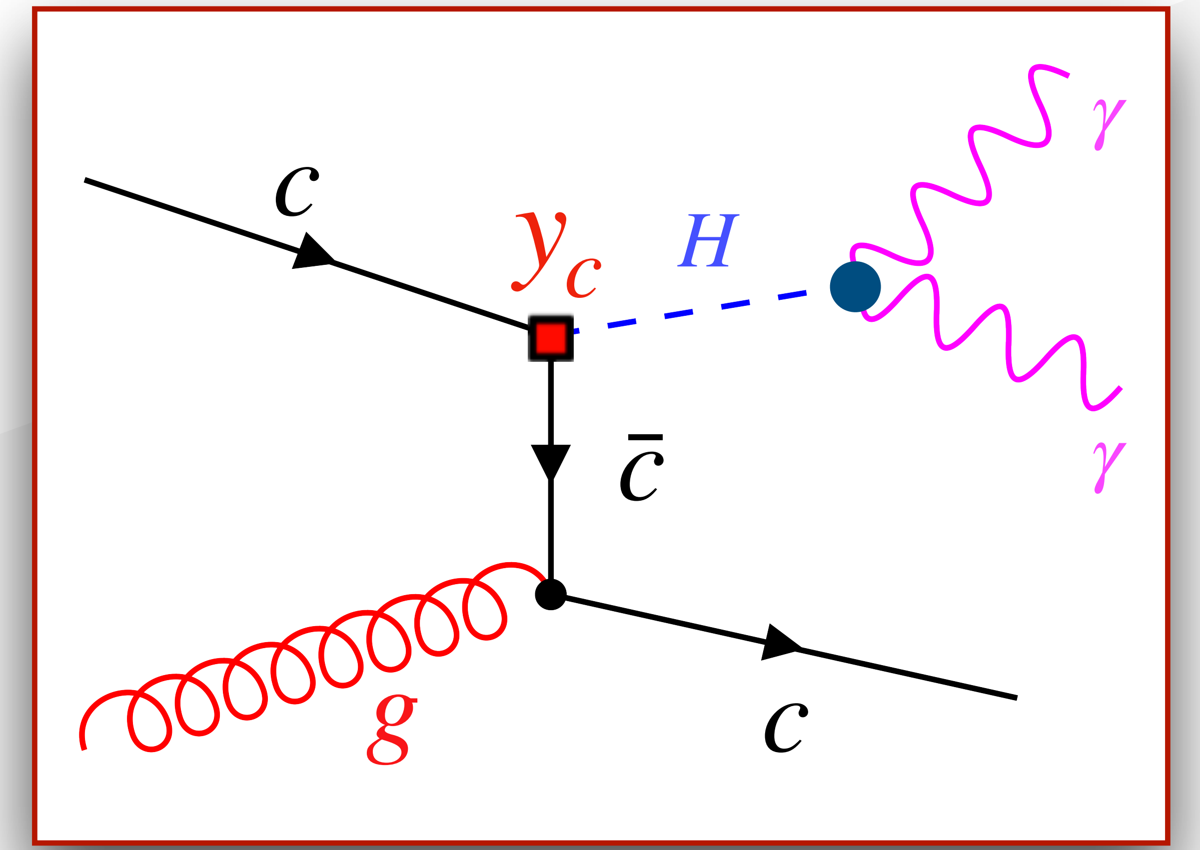
H. Qu

New approaches

H+c associated production:

But it's a group effort!

- Proposed by theorists (Isidori et al. 2015).
- Advantages of this channel:
  - Leading contribution requires only 1 charm to be tagged.
  - **Coupling with charm in production** ⇒ clean Higgs decays $(H \rightarrow \gamma\gamma)$.
  - Uncovered phase space, **complementary to existing** $H \rightarrow c\bar{c}$ **searches**.
- But also a few challenges:
  - Small cross section (~ 0.2 fb for $cH(H \rightarrow \gamma\gamma)$ vs 6.6 fb for $VH(H \rightarrow c\bar{c})$).
  - Non trivial signal MC simulation.
  - Challenging soft c-tagging.



No experimental results yet!

Outline

Introduction:

- Motivation.
- State of the art.
- New approaches.

Ingredients:

- Monte Carlo simulation.
- $H \to \gamma\gamma$ method.
- Framework development.
- Charm tagging.

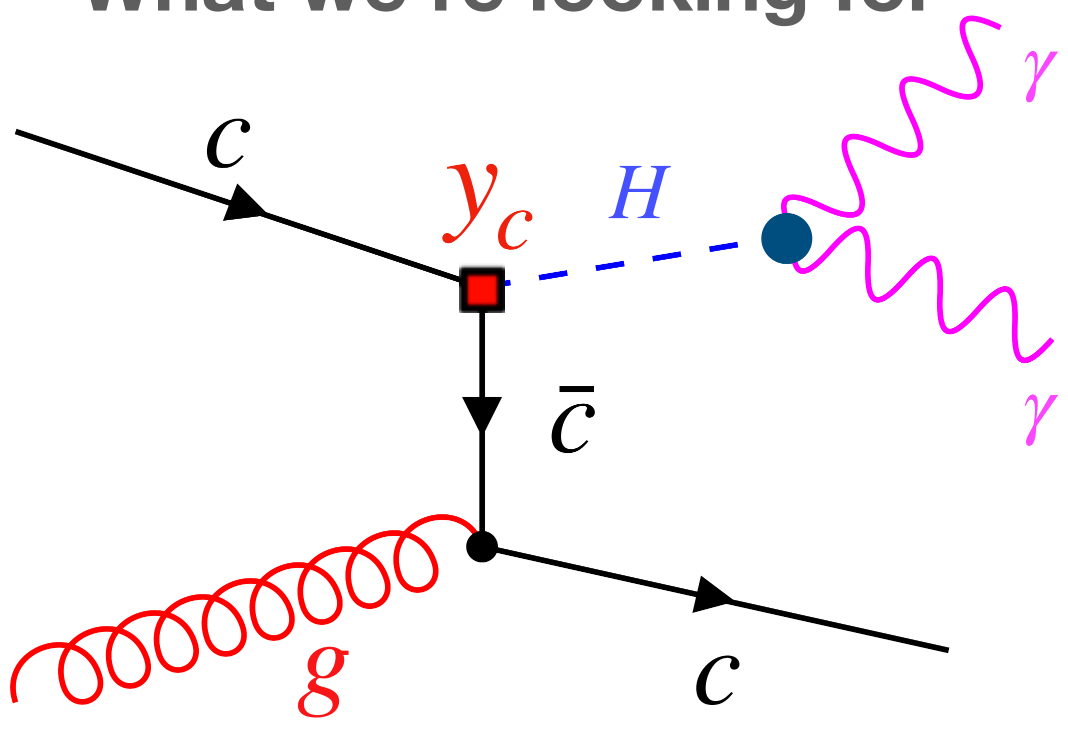
H+c analysis status:

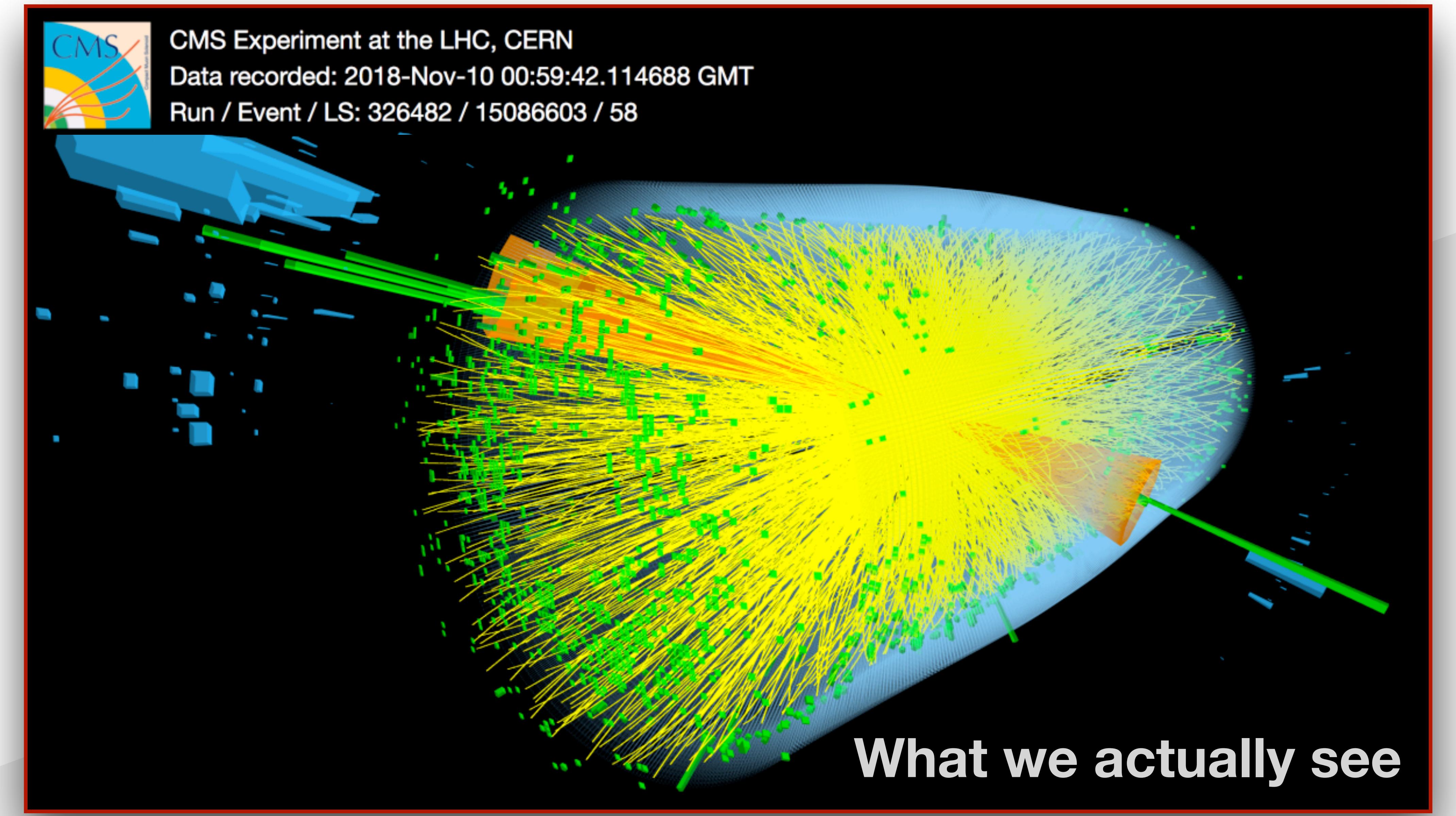
- Analysis strategy.
- Preliminary results.

MC simulation

Why we need them?

What we're looking for





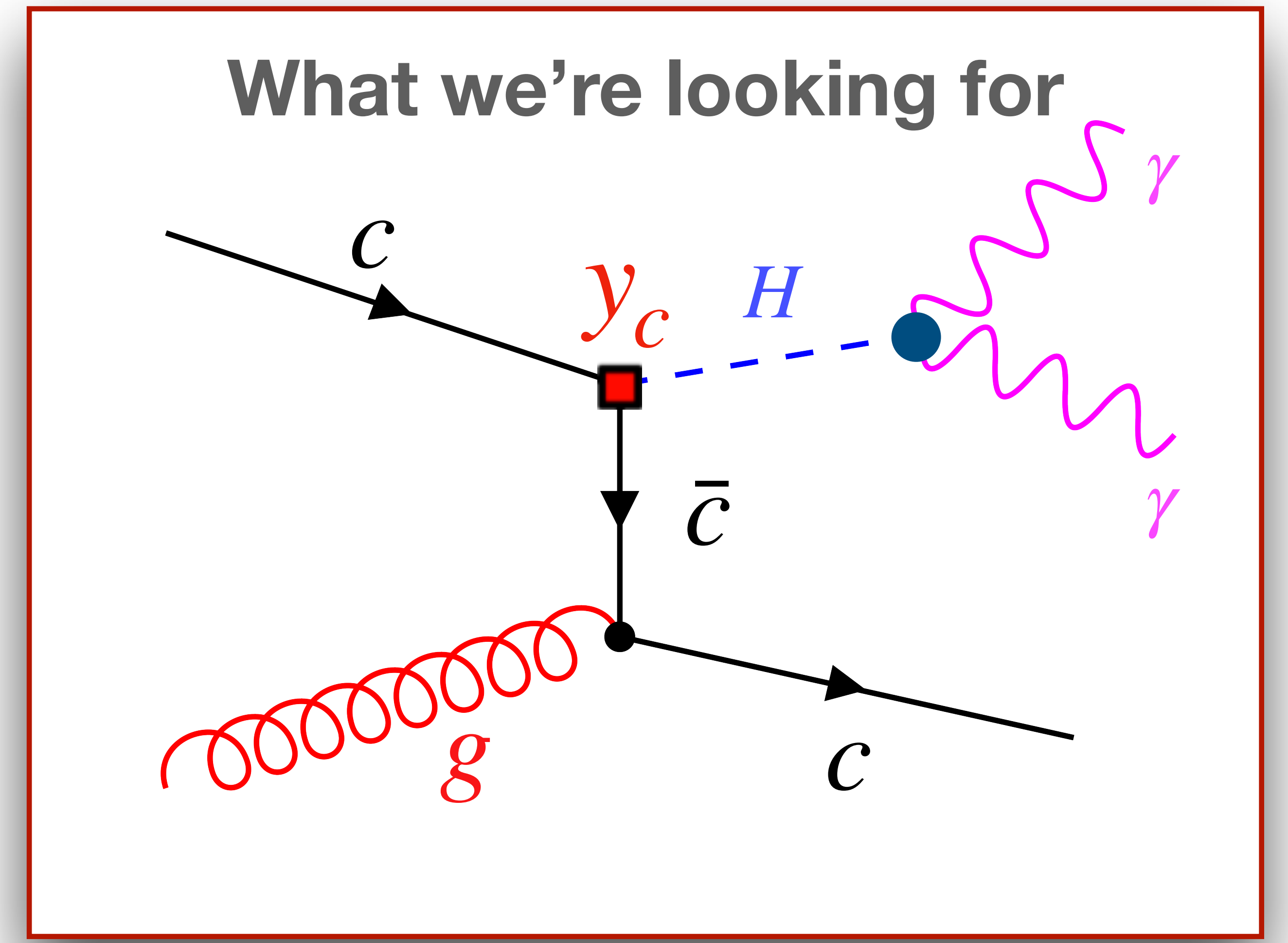
What we actually see

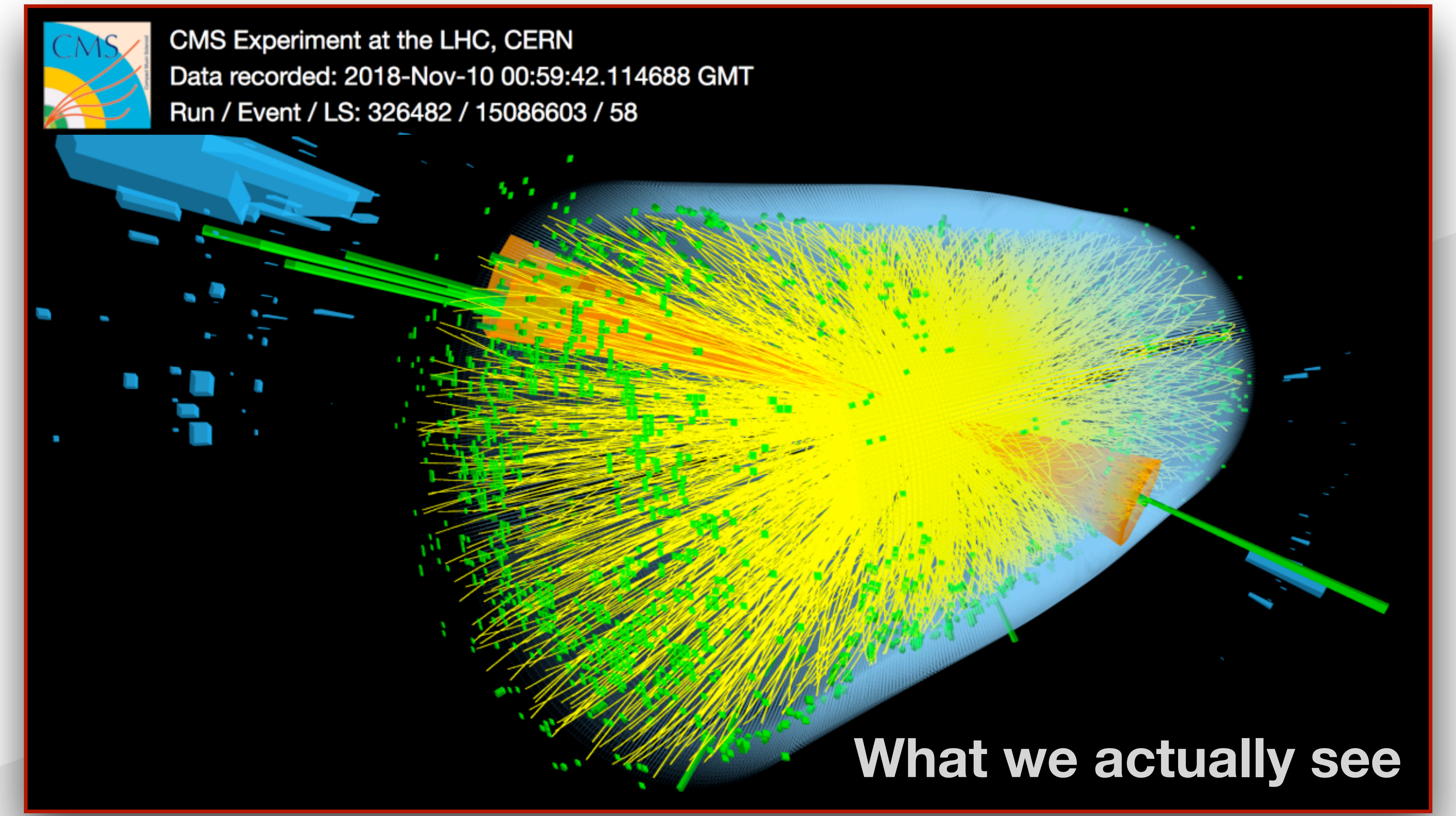
Quantum Chromo Dynamics

HEP experimental physicists

MC simulation

Why we need them?


What we’re looking for



What we actually see


Quantum Chromo Dynamics

HEP experimental physicists

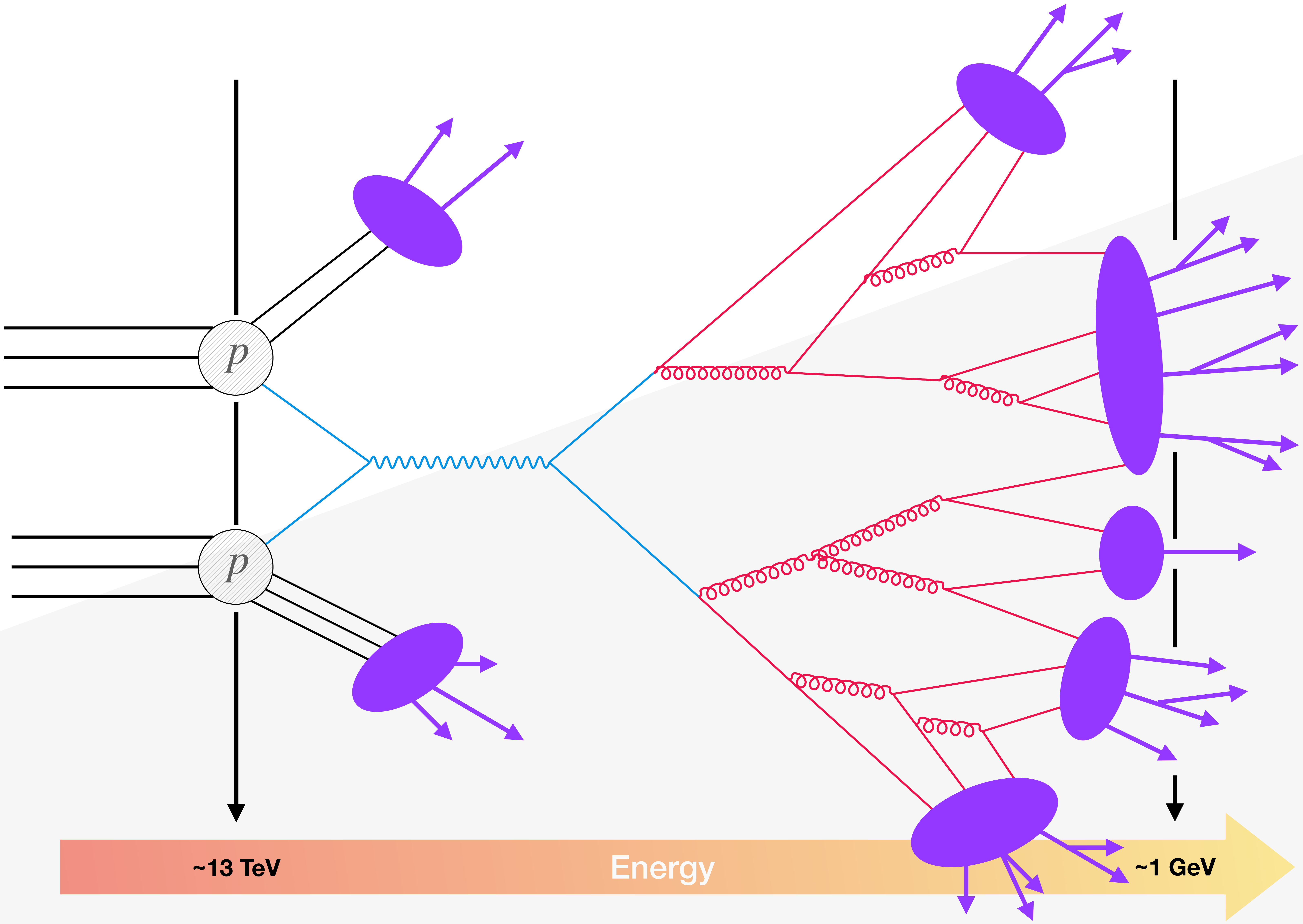
How?

- Theoretical knowledge of fundamental interactions.
- Random number generators give us **events,** extracted **from probability distributions.**

MC simulation

Event generation:

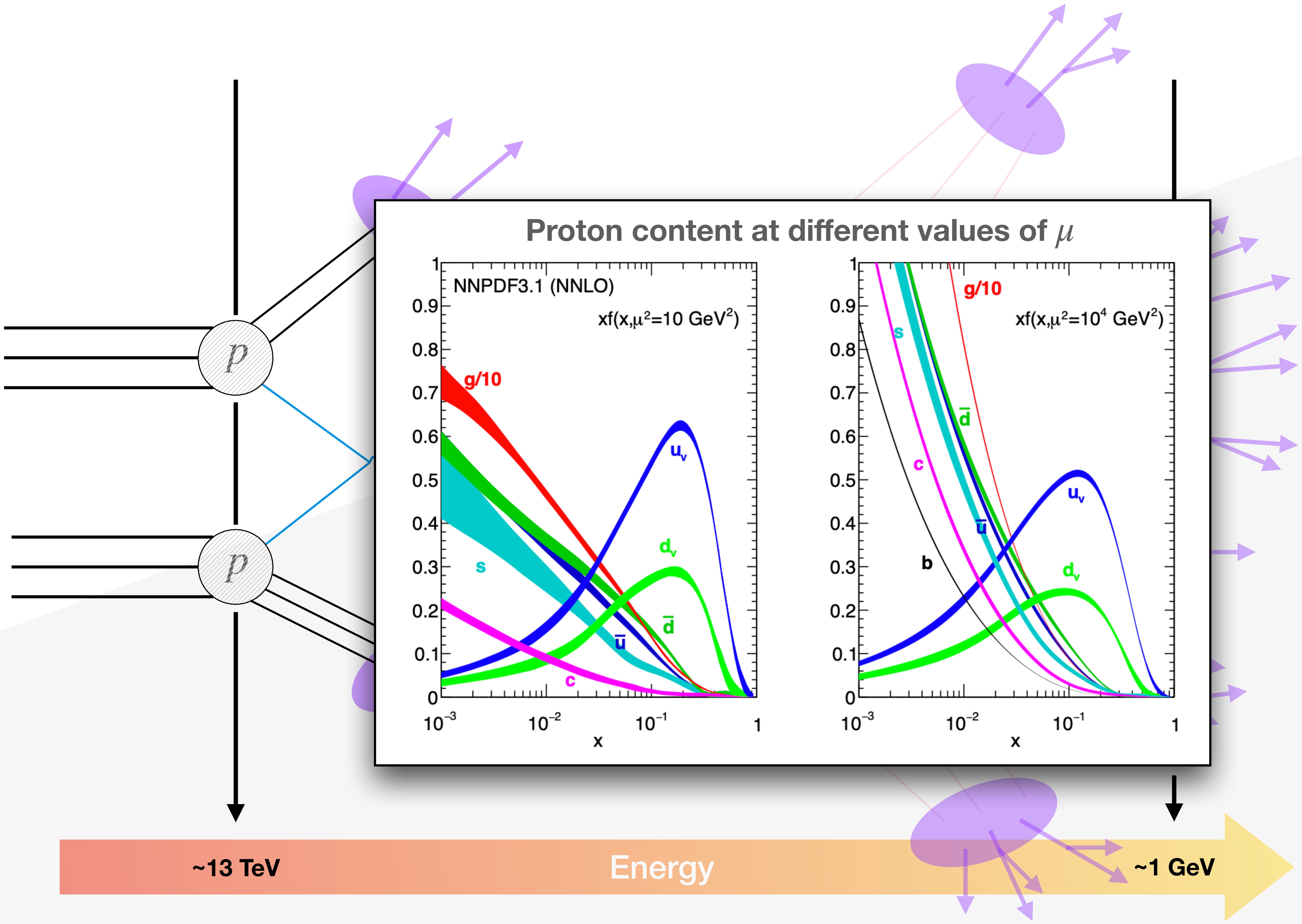
- PDFs.
- ME calculation, high Q^2, perturbative calculation.
- Gluon radiation and splitting (PS) .
- Hadronisation and decay, Q^2~1 GeV (PS).



MC simulation

Event generation:

- PDFs.
- ME calculation, high Q, perturbative calculation.
- Gluon radiation and splitting (PS) .
- Hadronisation and decay, Q~1 GeV (PS).

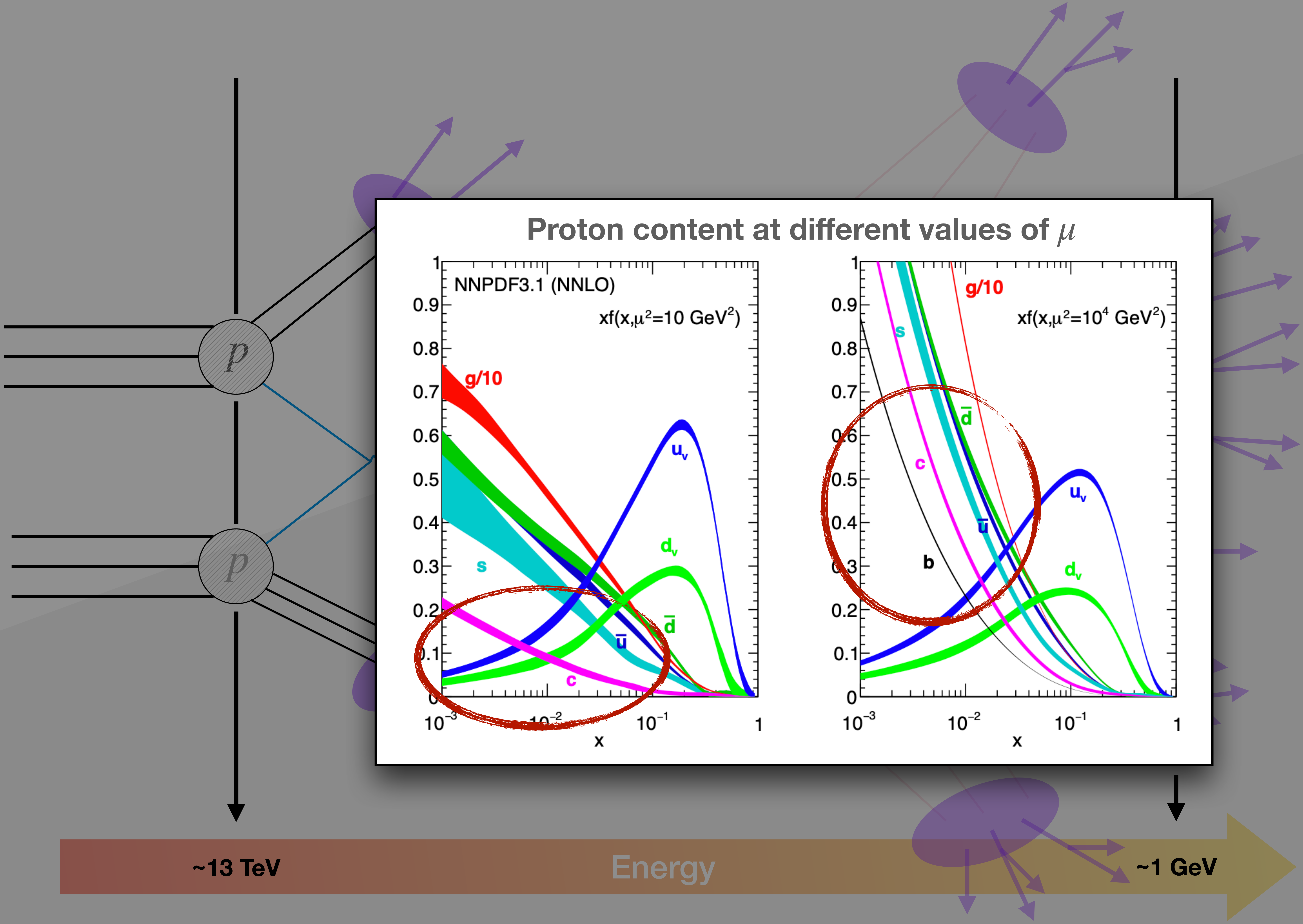


Proton content at different values of μ

MC simulation

Event generation:

- PDFs.
- For **different energy scales** of the process, a **different number of flavours** is relevant.
- Probability to find a Parton of type i inside the proton with a fraction of momentum x_i.

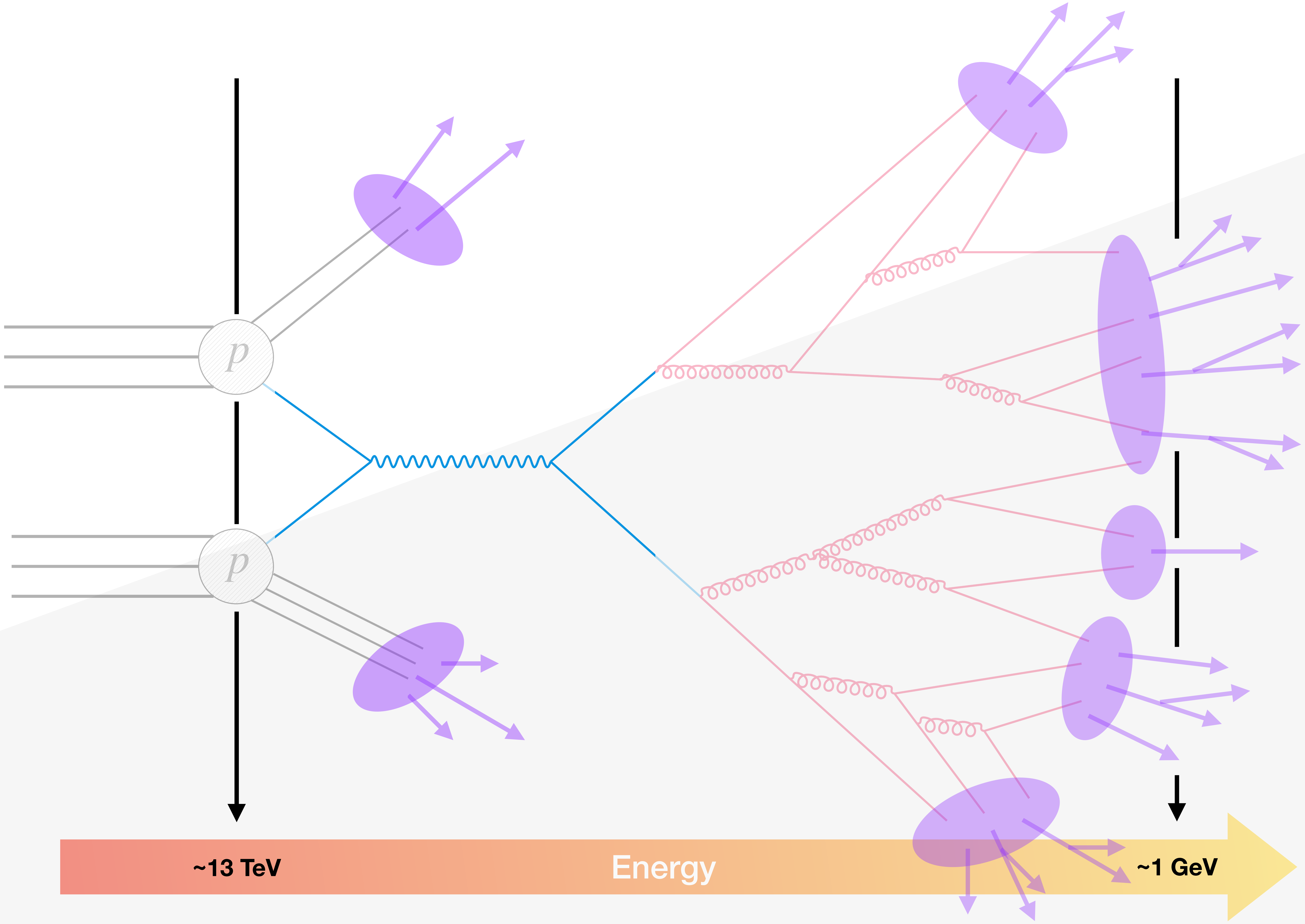


Proton content at different values of μ

MC simulation

Event generation:

- PDFs.
- ME calculation, high Q^2, perturbative calculation.
- Gluon radiation and splitting (PS) .
- Hadronisation and decay, Q^2~1 GeV.



MC simulation

ME calculation: Hadronic cross section

The hadronic cross section producing a final state X:

$$\sigma_{pp\to X} = \sum_{a,b} \int dx_a \cdot dx_b \cdot f_a(x_a, \mu_F^2) \cdot f_b(x_b, \mu_F^2) \cdot \hat{\sigma}_{ab\to X}(x_a p_a, x_b p_b, \mu_F^2, \mu_R^2)$$

- Separation between short- and long-distance effect delimited by a factorization scale μ_F (Factorisation theorem).
- **Parton density functions** f_i for flavour type i with a momentum fraction x_i.
- $\hat{\sigma}_{ab\to X}$ is the **partonic cross section** of the process $ab \to X$ and is computed at fixed order in perturbation theory, introducing a dependence on the renormalization scale μ_R.

MC simulation

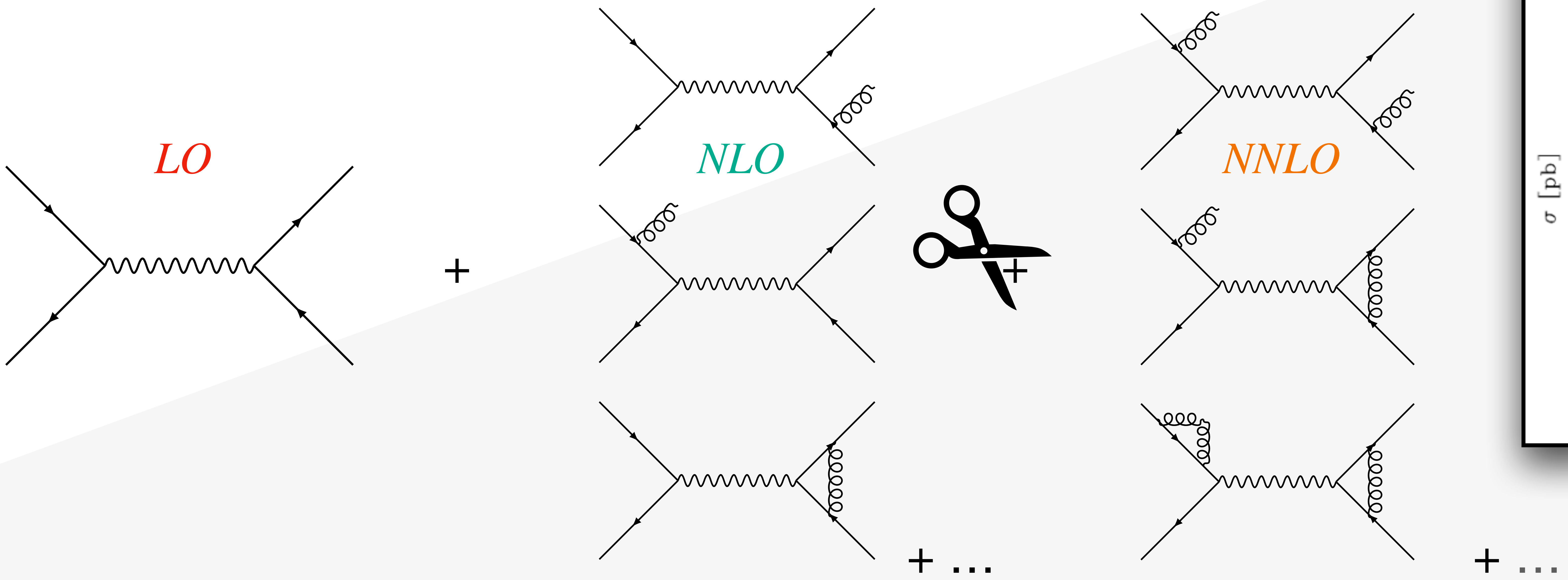
ME fixed order calculations:

Partonic cross section computed as **series expansion** in the strong coupling α_s:

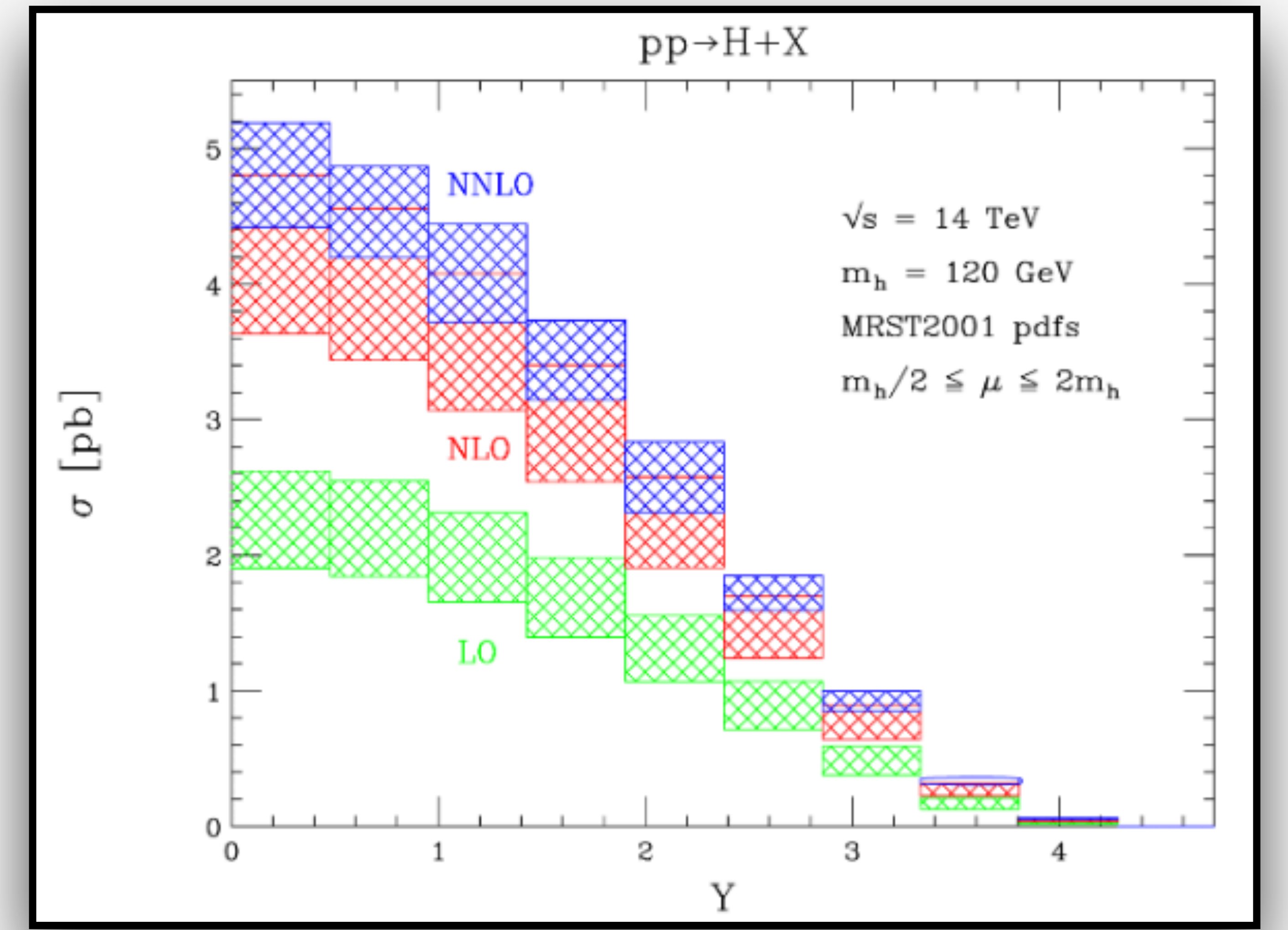
$$\hat{\sigma}_{ab\to X} = \alpha_s^n(\sigma_0 + \alpha_s\cdot\sigma_1 + \alpha_s^2\cdot\sigma_2 + O(\alpha^3))$$

$$LO + NLO + NNLO + \ldots$$

ME generators compute partonic cross section at Fixed Order (FO):



Fixed order cross section

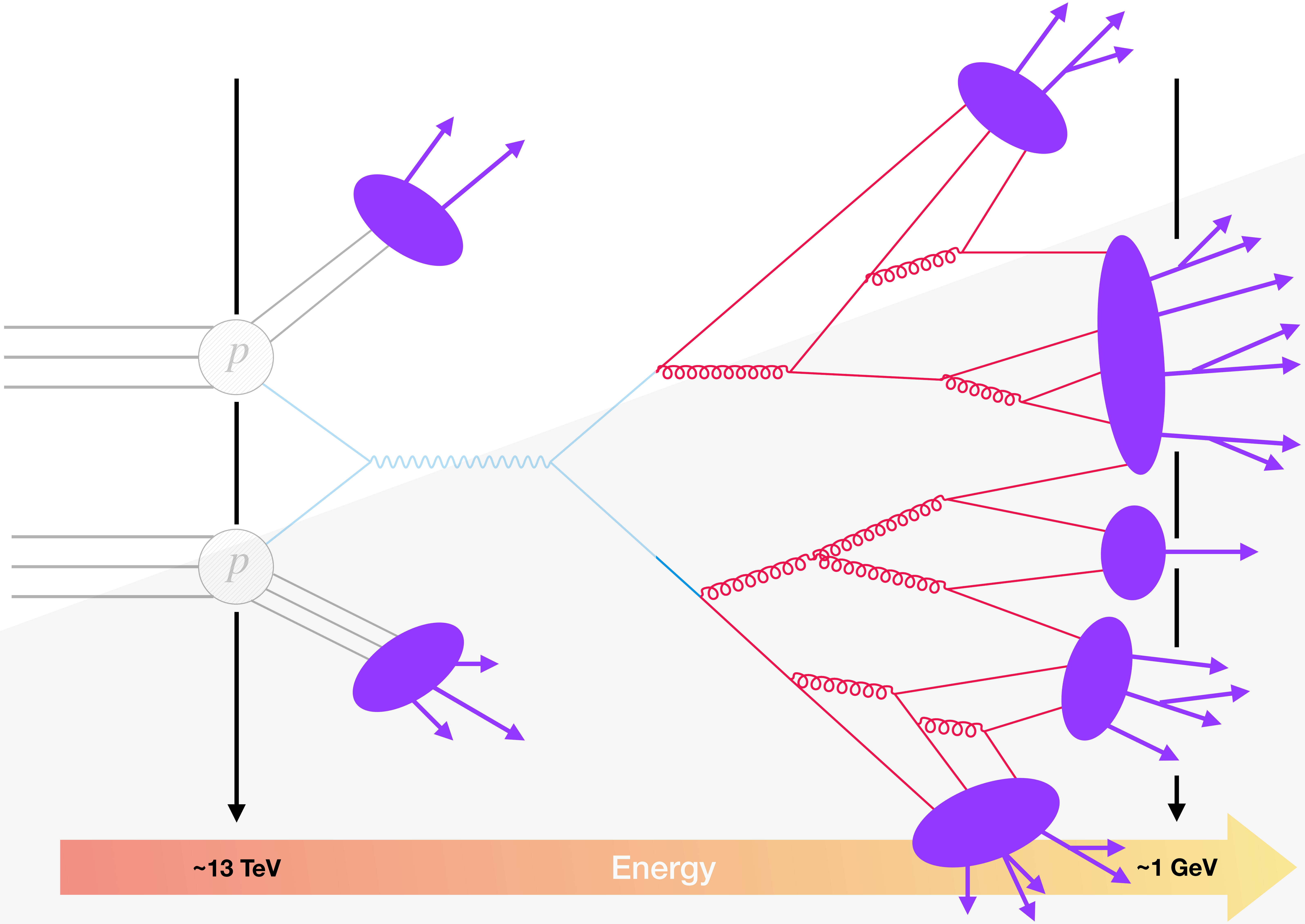


Anastasiou, Melnikov, Petriello, '05

MC simulation

Event generation:

- PDFs.
- ME calculation, high Q^2, perturbative calculation.
- Gluon radiation and splitting (PS) .
- Hadronisation and decay, Q^2~1 GeV (PS).



MC simulation

Parton Shower:

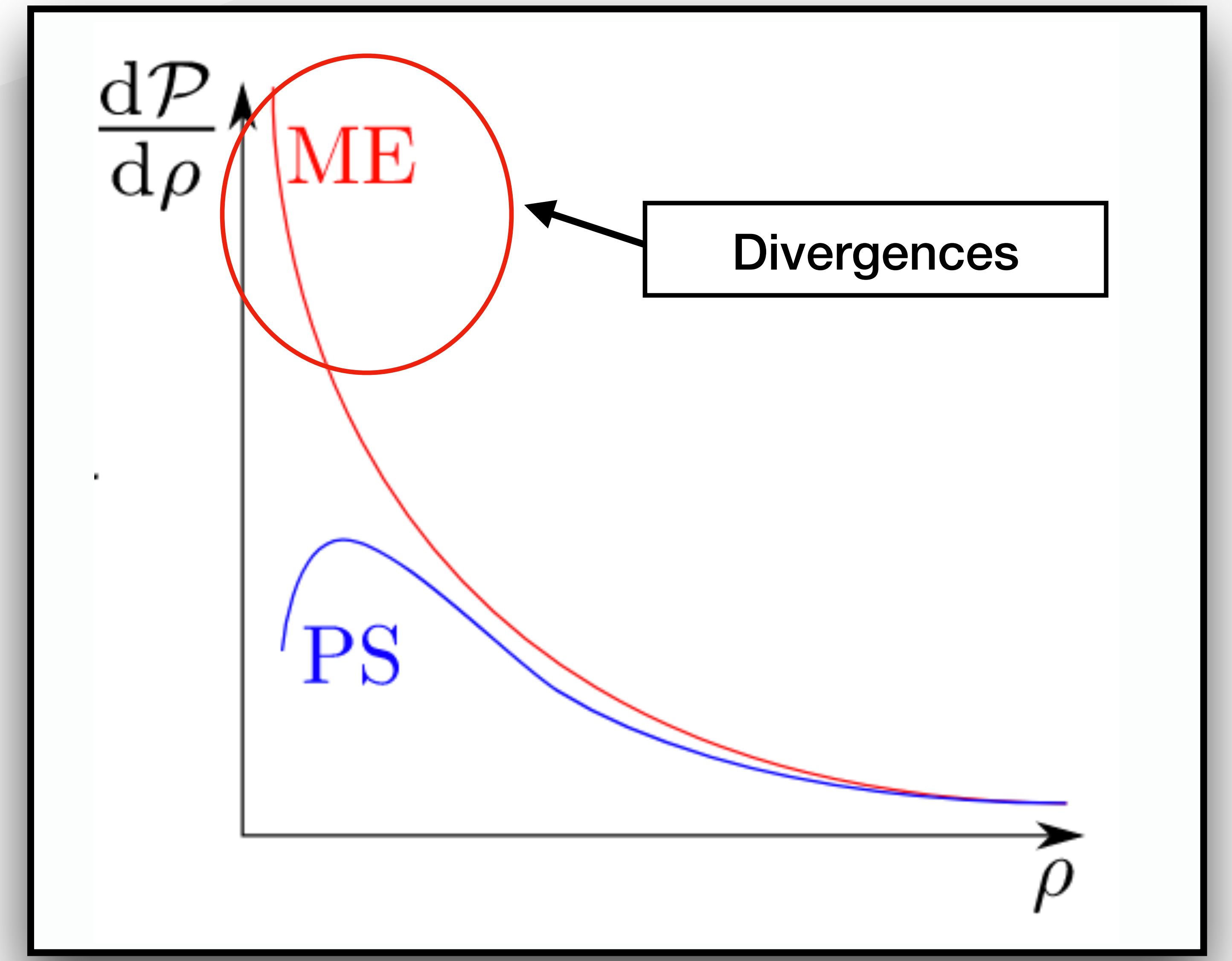
Partonic cross section computed at Fixed Order:

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

$$LO + NLO$$

The rest (additional radiation and showering) is taken care by the PS:

- Account for the rest of higher order contributions.
- It uses approximations to reach high multiplicities.
- Valid in the soft and collinear phase space.
- Always finite.

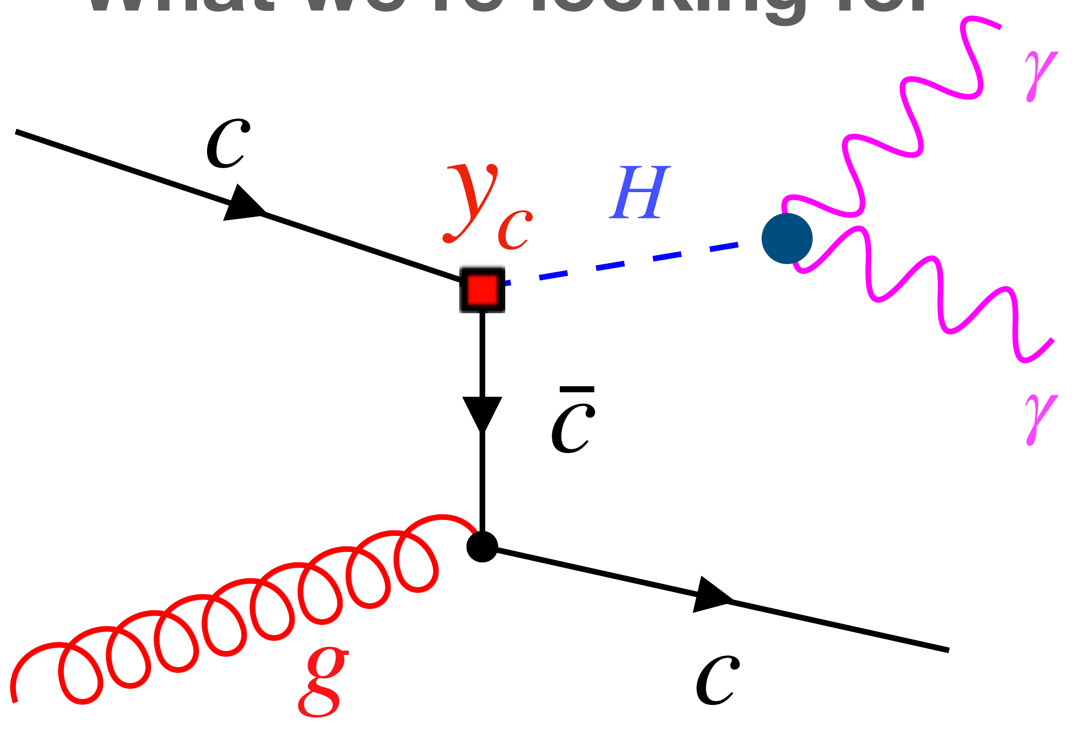


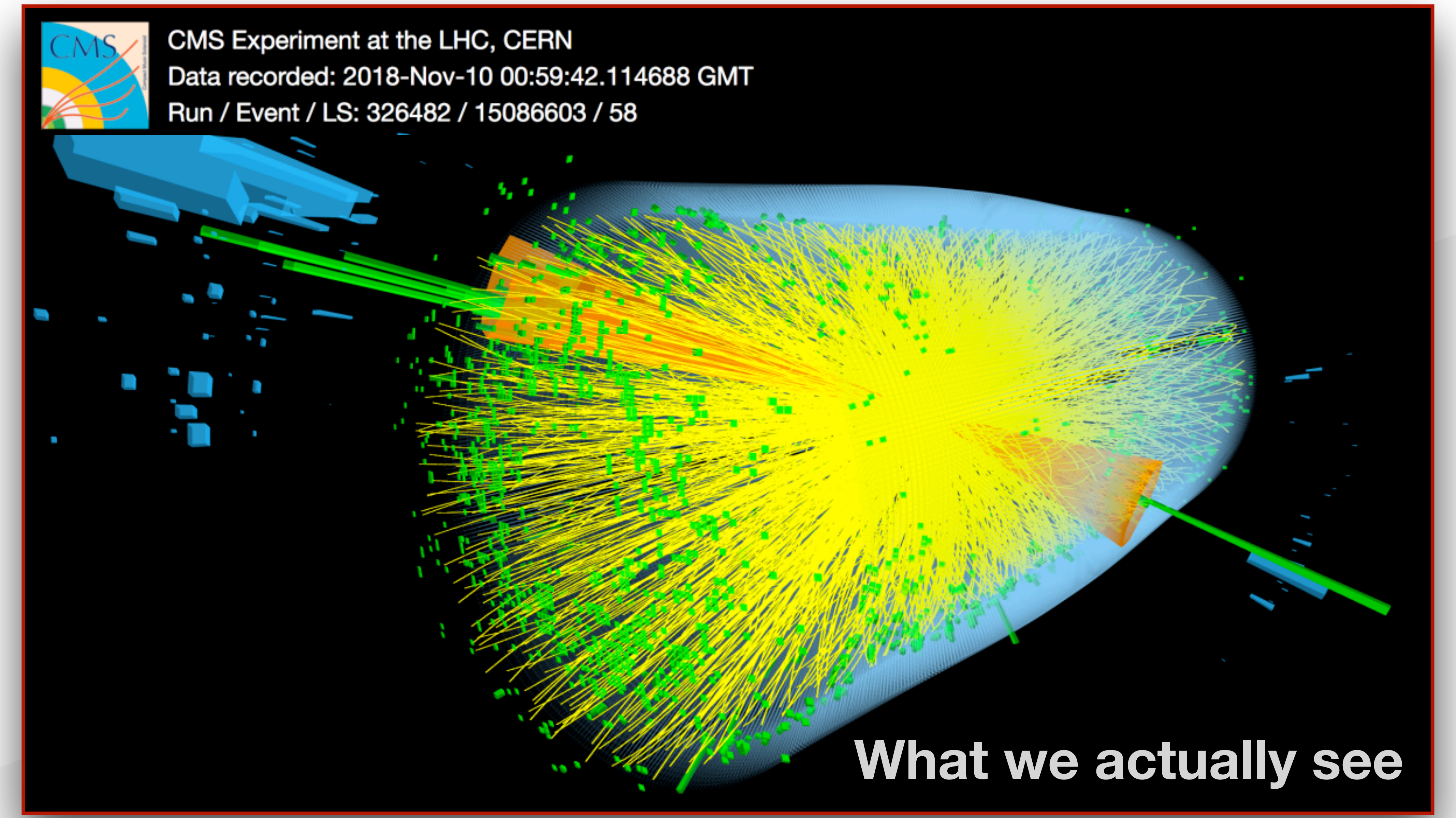
Borrowed from L. Gellersen

MC simulation

Why we need them?

What we’re looking for





What we actually see

Still HEP experimental physicists

- Still challenging but possible!

MC simulation

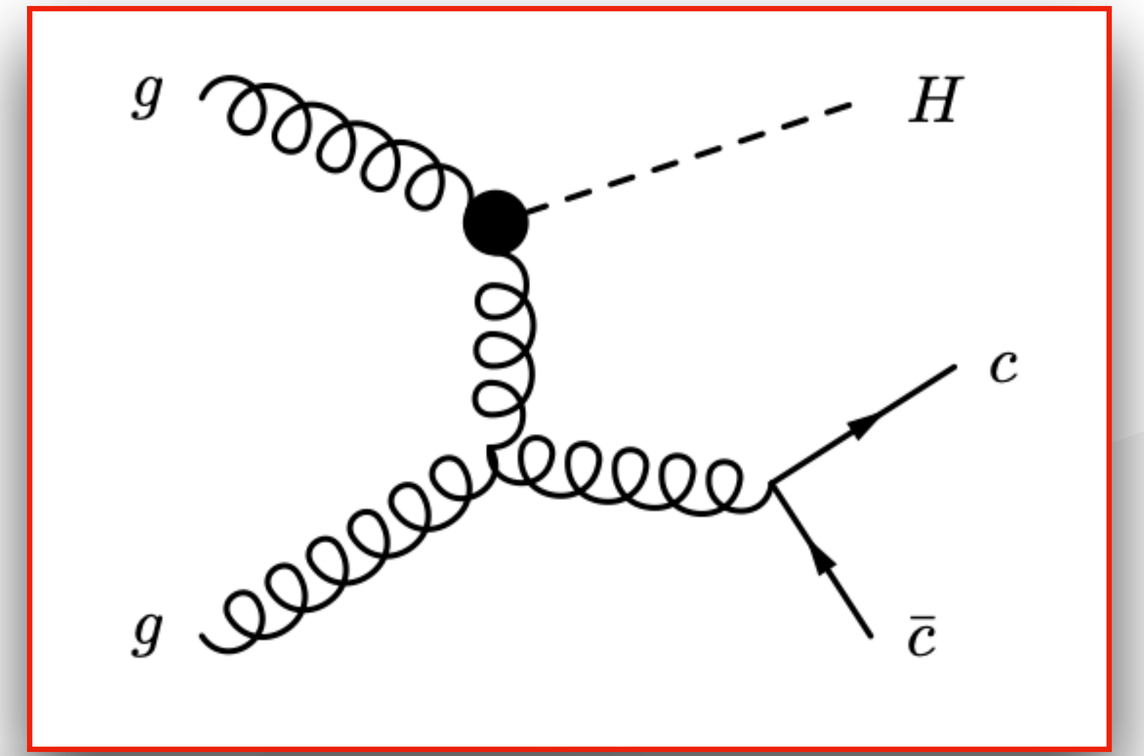
H+c production:

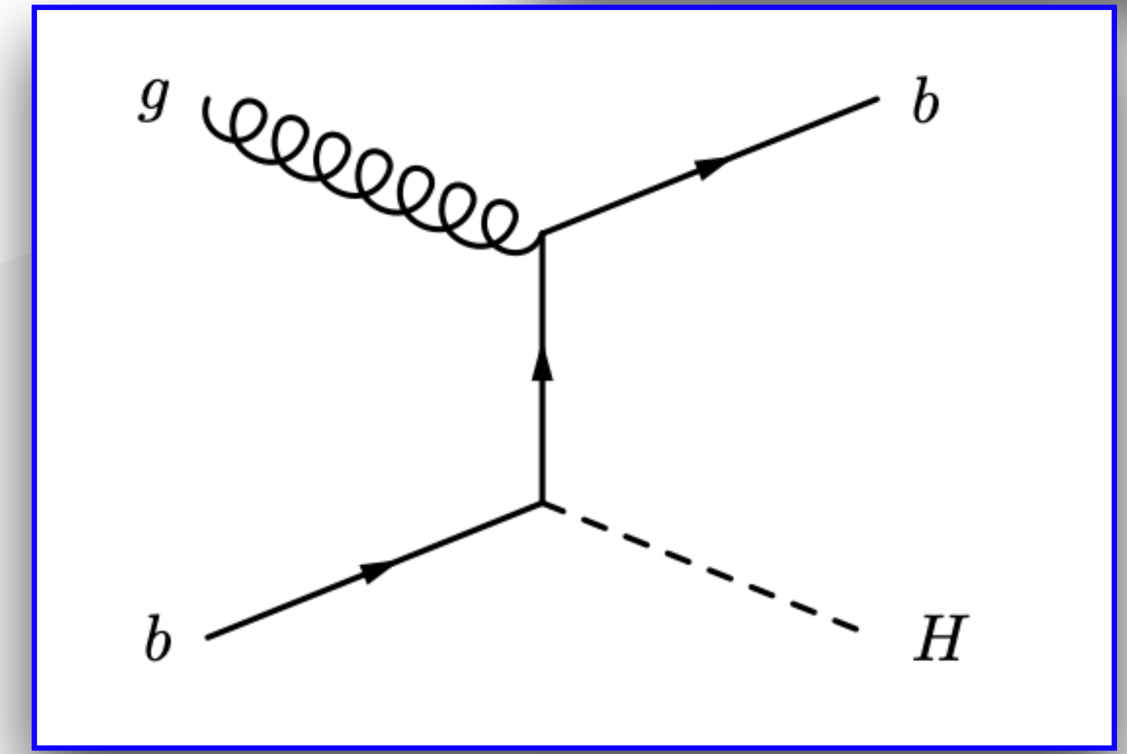
- Includes several contributions that do not depend on y_c :
  - $H+g$ (fakes and $g \to c\bar{c}$)
  - $H+b(b \to c)$
  - $H+c$ not induced by y_c.

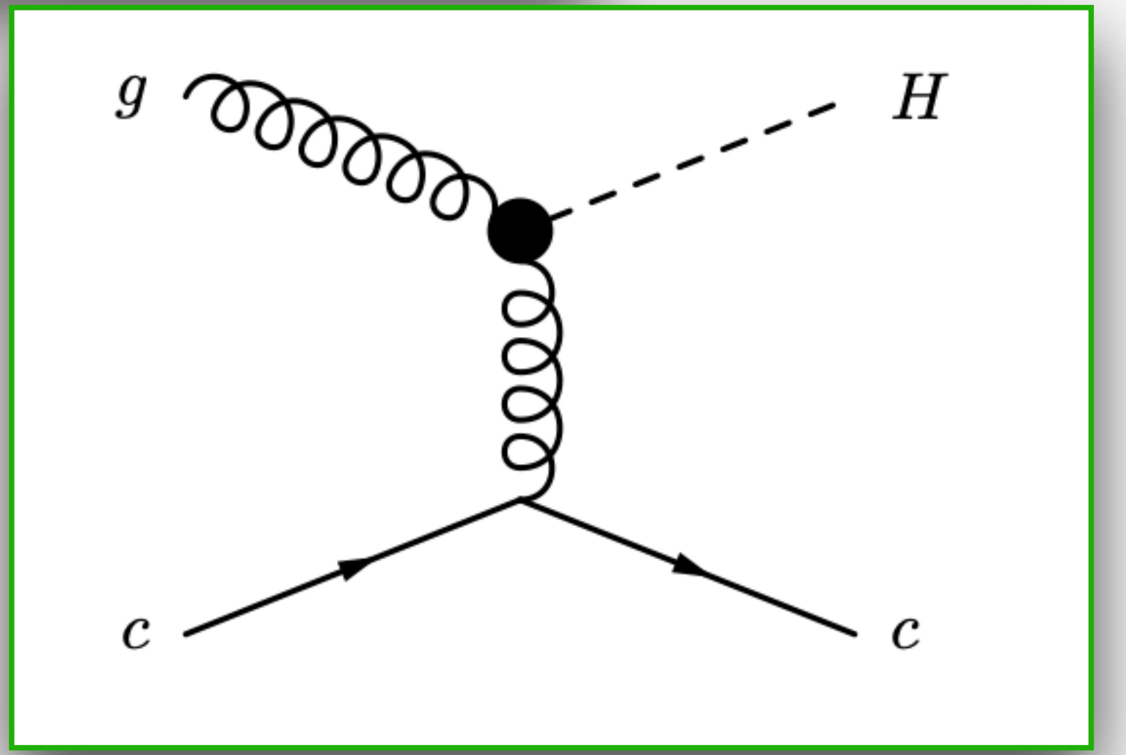
  ⇒ **most of the $H+c$ cross section is not sensitive to y_c.**

- 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:

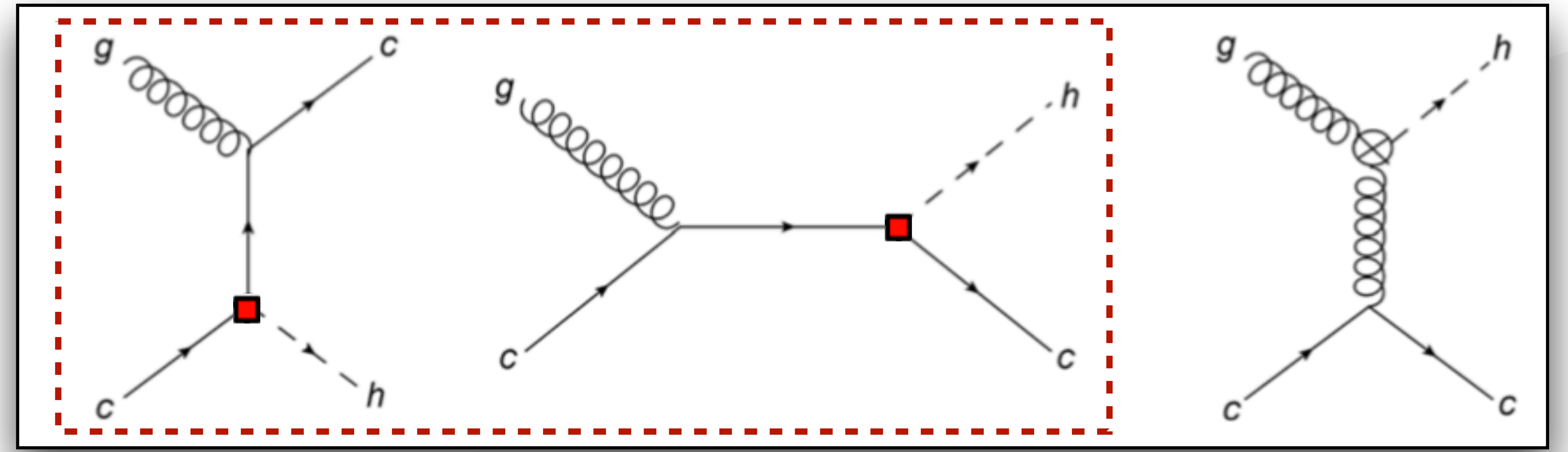
- Focusing on the signal simulation for $H + c$ MC (not available in CMS up to now).

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

	σ [fb]
A	254.5
B	-3.5
C	34.5

[$y_c = y_c^{sm}$]

[GEN charm $p_T > 20$ GeV]

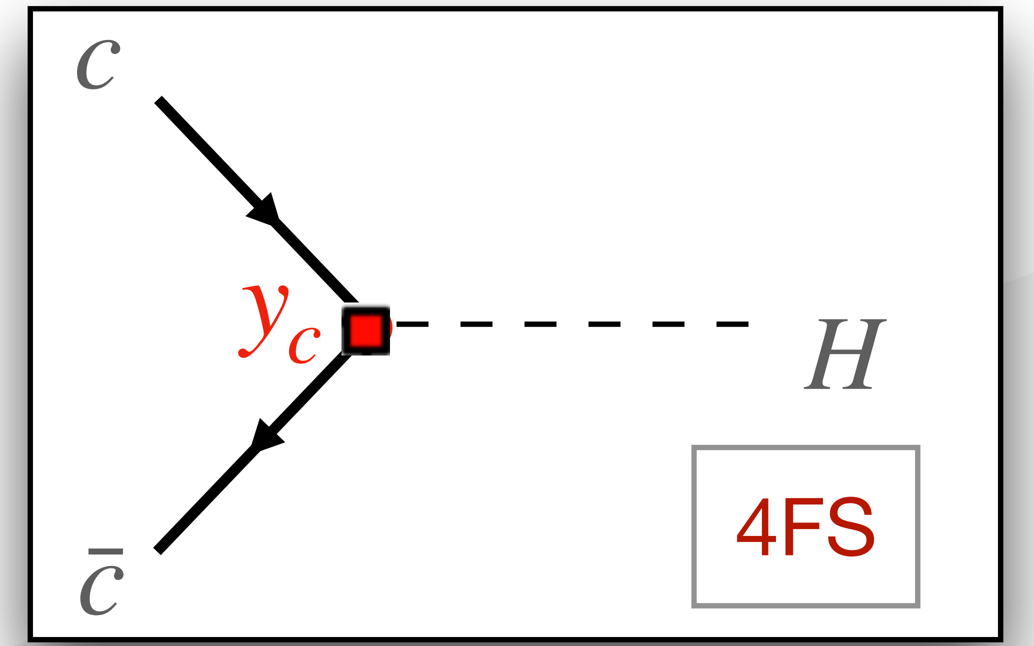


- $\sigma(hc)$ does not scale trivially with y_c, some tests were run with effective ggH coupling at LO.
- Biggest contribution from the term that does not probe y_c, but small y_c proportional interference term (~10 times smaller than the y_c^2 dependent term), for sensitivity $O(10 \cdot SM)$ contribution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

Higgs+charm simulation

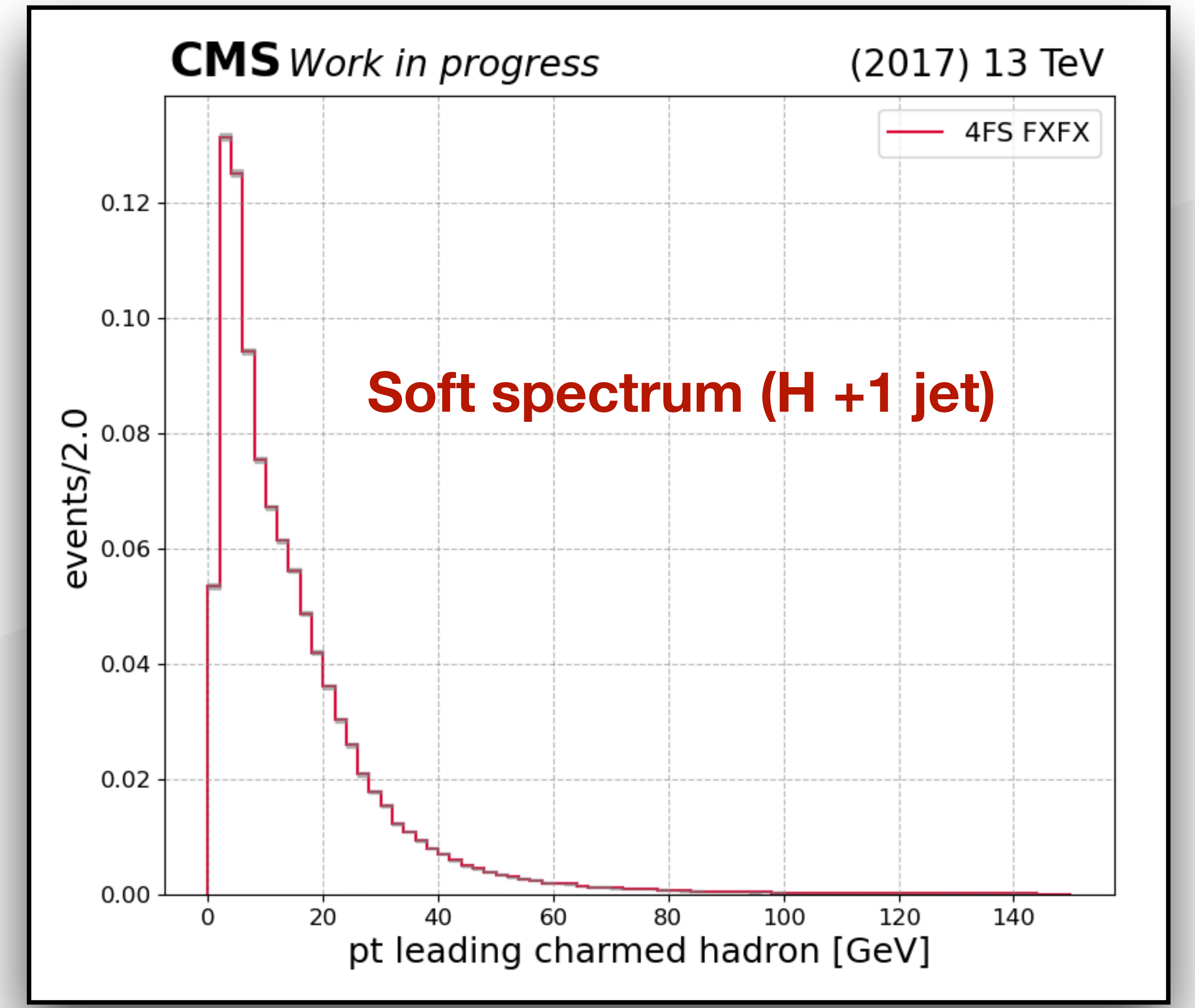
Focus on the y_c^2 term:

- 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.



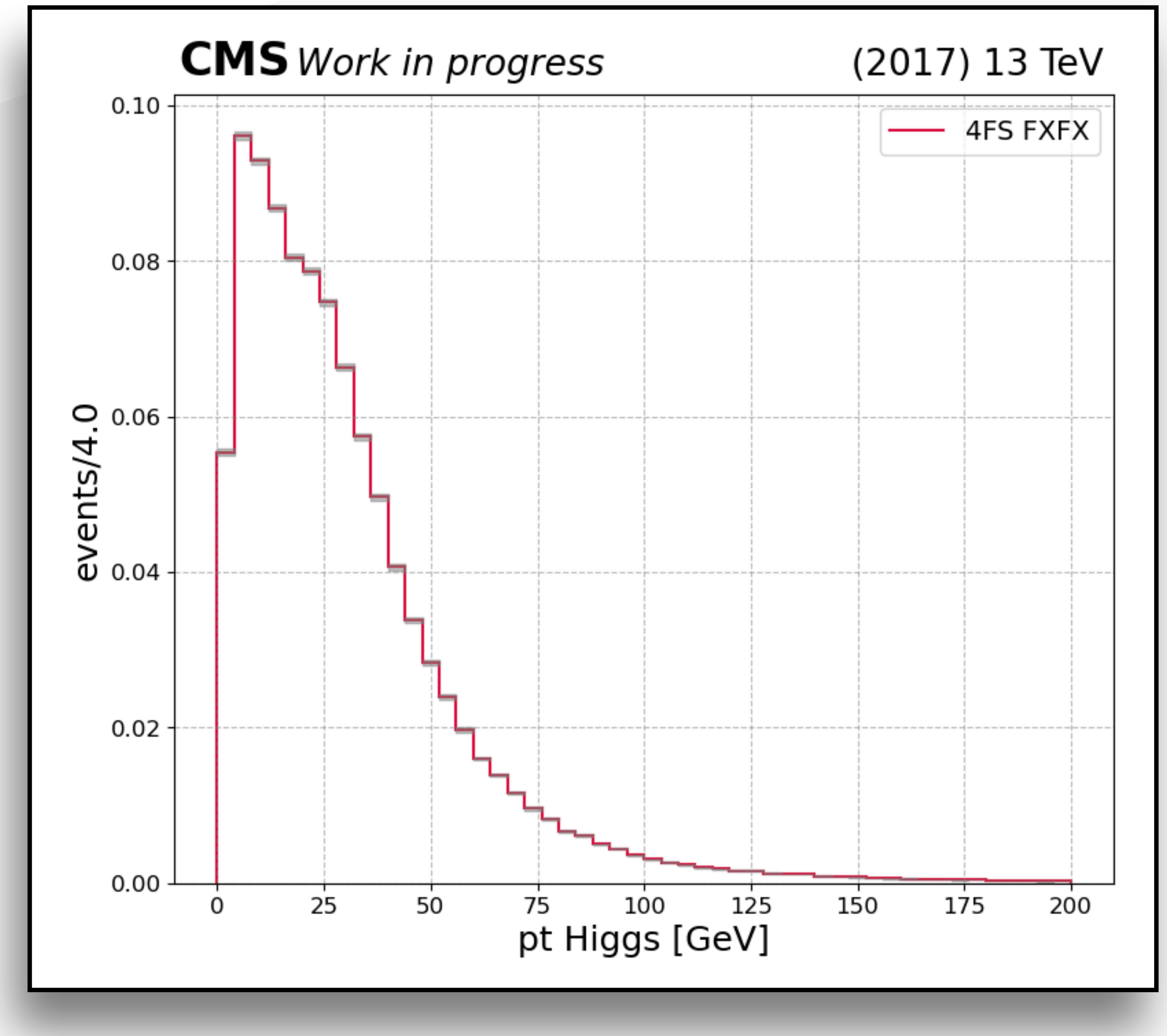
$m_C = 0$

Pt lead charmed hadron



Soft spectrum (H +1 jet)

Pt Higgs

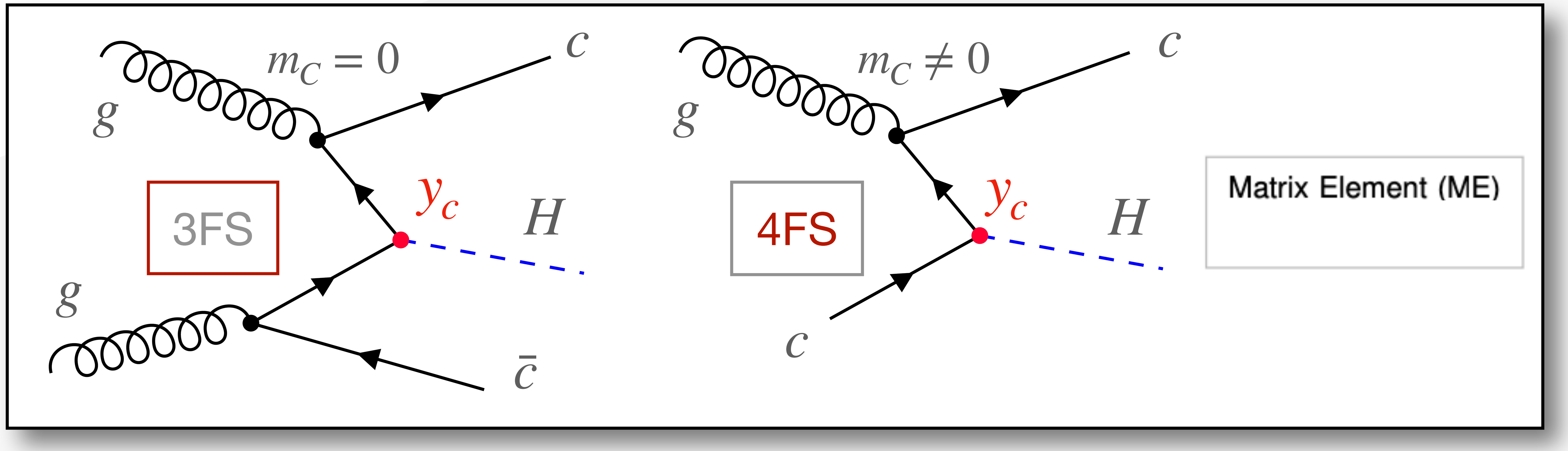


Higgs+charm simulation

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 to assess the additional theory uncertainty we **compare samples produced using both methods:**
  ⇒ FS uncertainty $O(30\%)$ of the yields in analysis categories.
- Uncertainty due to Scales and PDFs are smaller than FS uncertainty.
- We also studied the impact of the choice of simulation input parameters on the two FS:
  ⇒ decide the best choice of theory scales (μ_R, μ_F, and R_{sh}).

LO diagrams

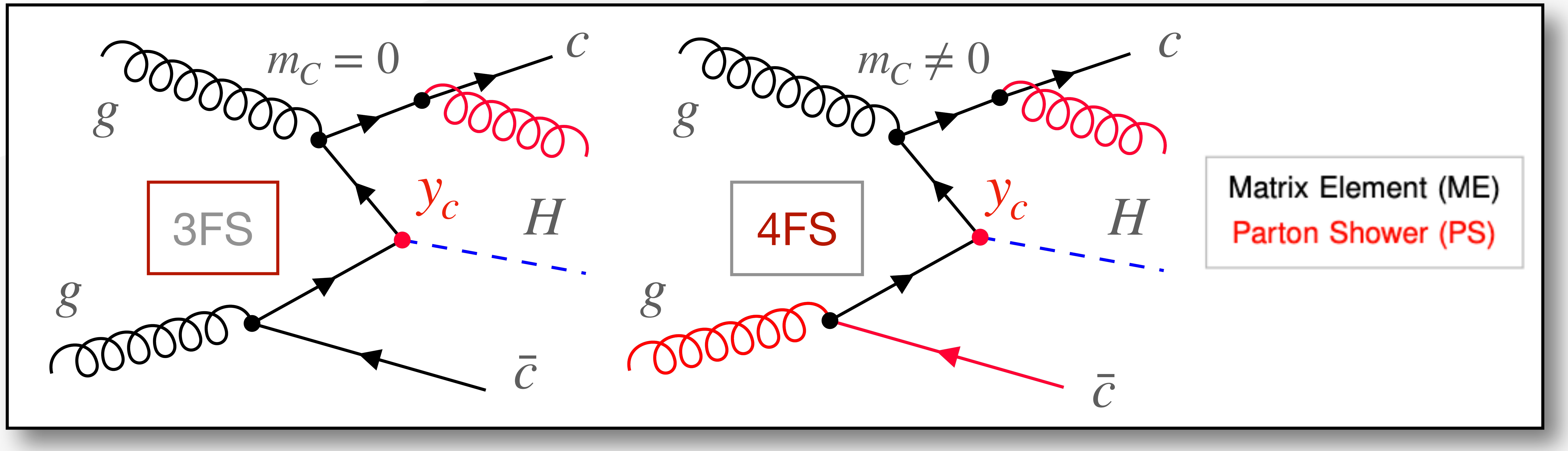


Higgs+charm simulation

Theoretical uncertainty studies:

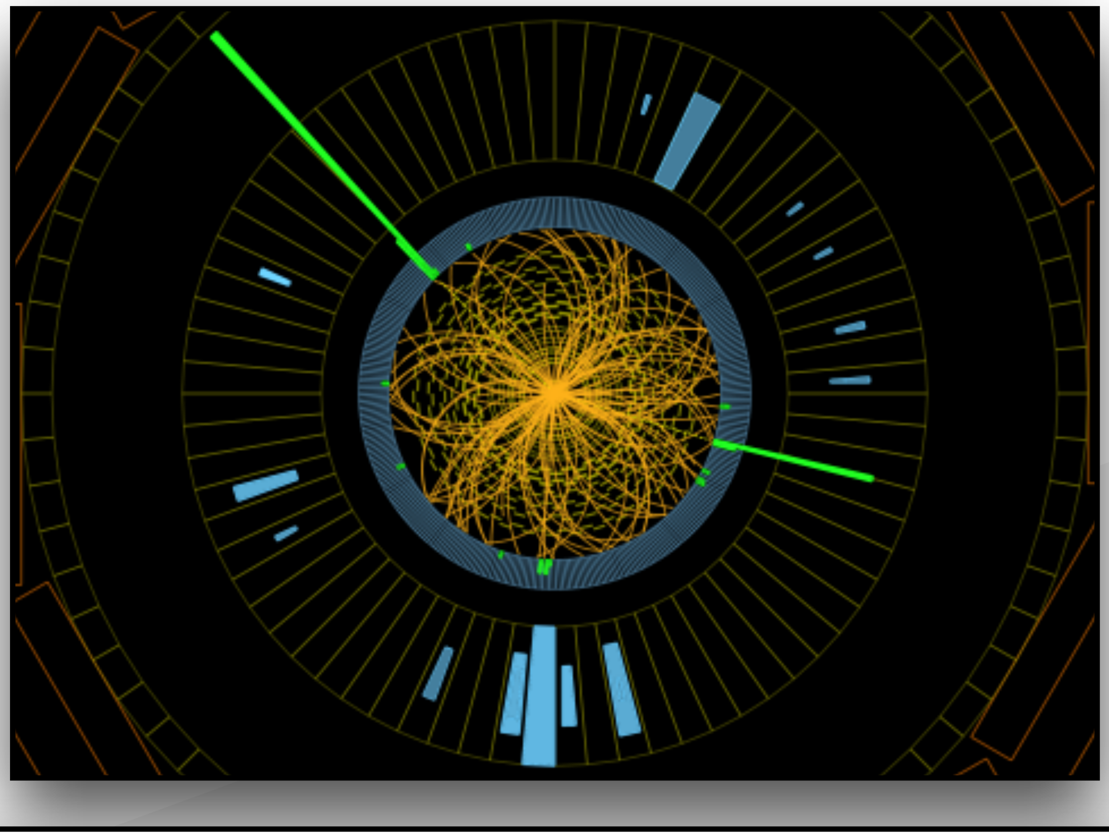
- H+c MC ME calculation can be done in either 3FS or 4FS, i.e. considering the c massive or massless.
- In principle equivalent → In practice to assess the additional theory uncertainty we **compare samples produced using both methods:**
  ⇒ FS uncertainty $O(30\%)$ of the yields in analysis categories.
- Uncertainty due to Scales and PDFs are smaller than FS uncertainty.
- We also studied the impact of the choice of simulation input parameters on the two FS:
  ⇒ decide the best choice of theory scales (μ_R, μ_F, and R_{sh}).

LO diagrams

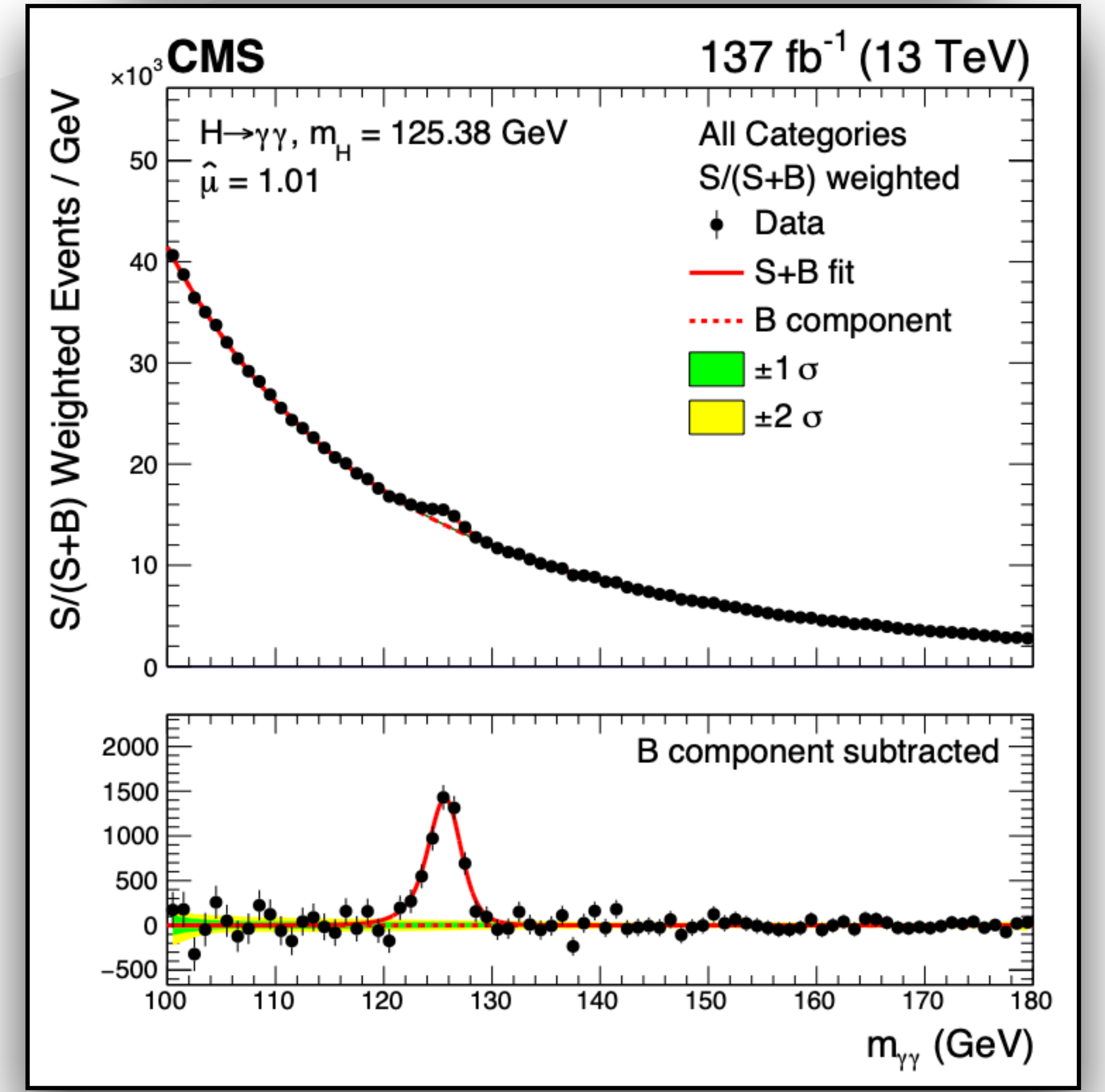


Higgs to $\gamma\gamma$ analysis strategy

Diphoton decay channel:



- Golden channel (since 2012) with a very clear signature: two isolated photons.
- All the possible diphoton candidates are constructed and the one with highest p_T is selected.
- **Fit to the diphoton** candidate **mass spectrum**.
- The signal shows as a **peak over a continuously falling background** from non resonant $\gamma\gamma$ and $\gamma + jets$ events.
- To do an accurate measurement the photon reconstruction systematics have to be kept under control.
- In our H+c we use the full Run 2 dataset (2016, 2017, 2018) of 137 fb^{-1} and we plan to include in the future also Run 3 data.

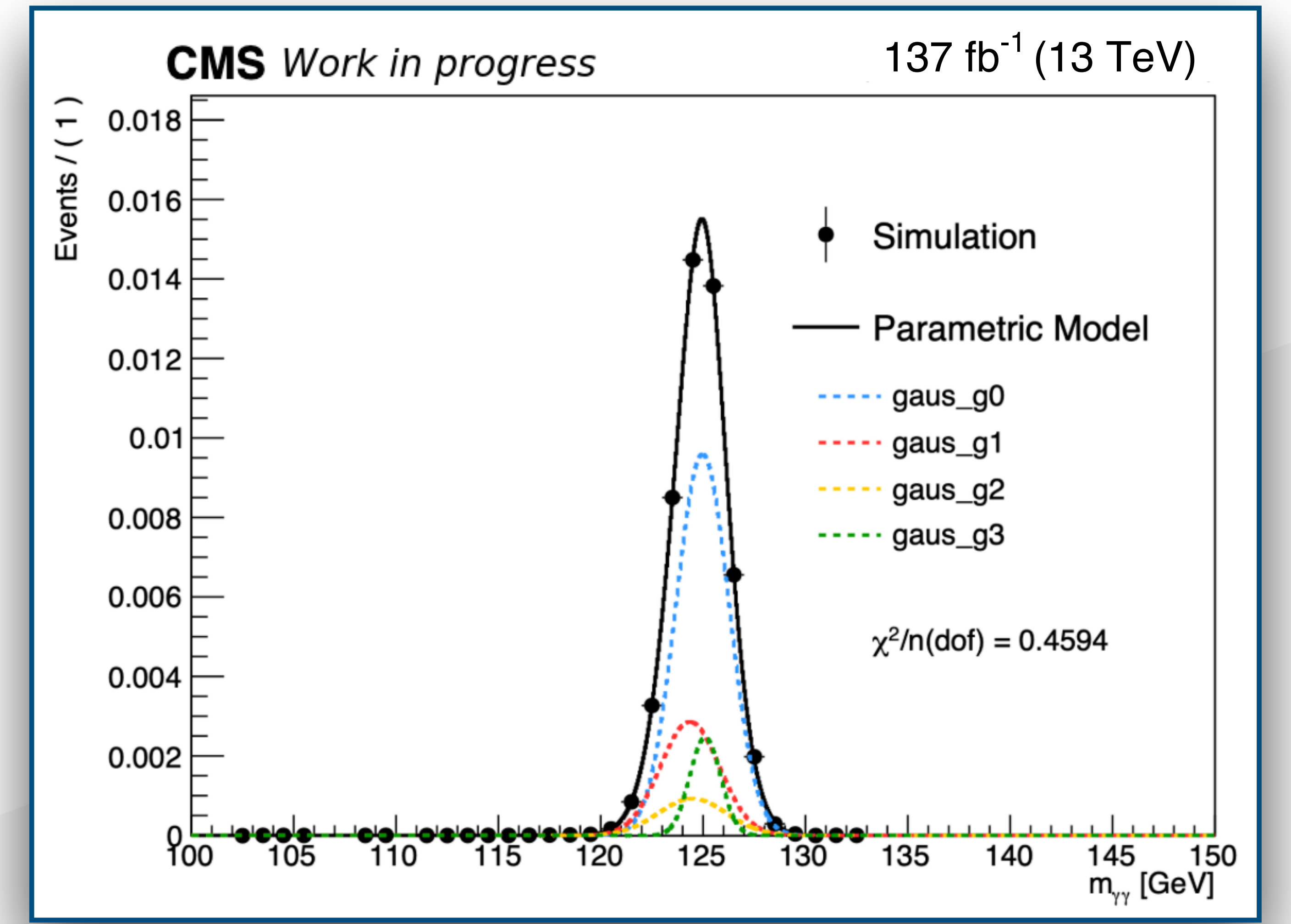


10.1007/JHEP07(2021)027

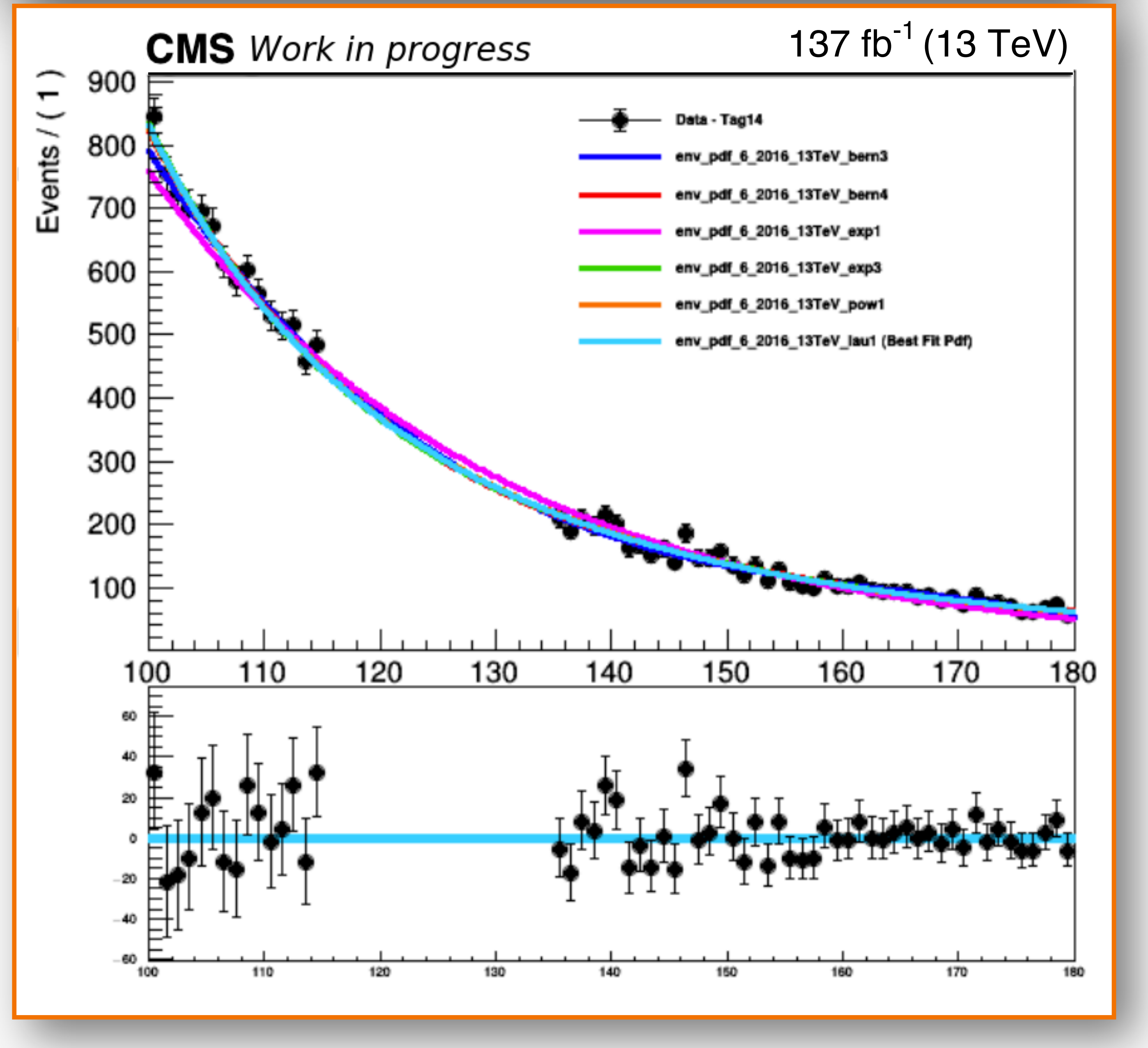
Higgs to $\gamma\gamma$ analysis strategy

Statistical analysis:

- Signal modelling:
  - Higgs processes ($H + c$, ggH, VBF, ttH, VH) are modelled by **fitting** a sum of gaussian to the **MC distribution**.



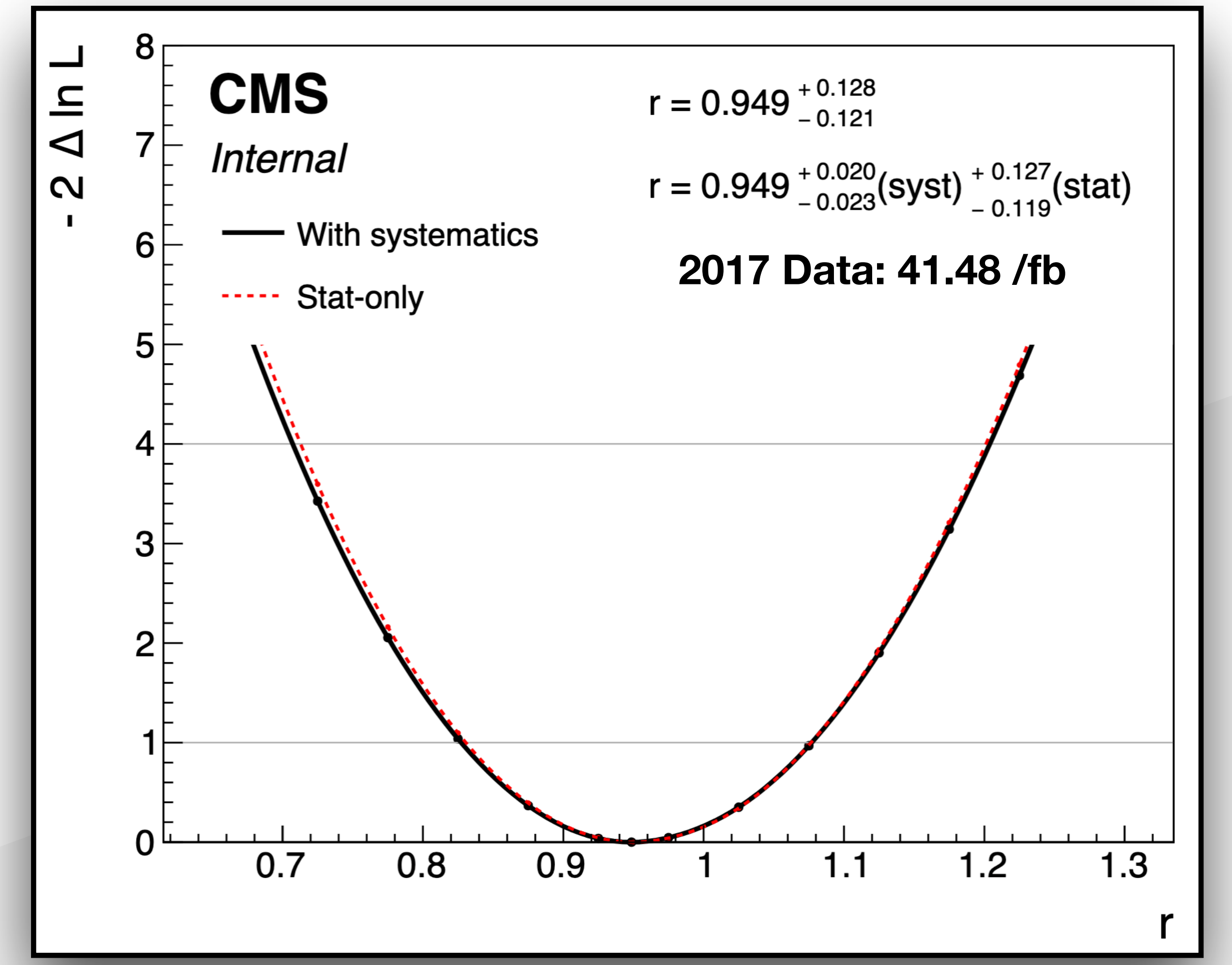
- Background modelling:
  - non-Higgs background is modelled in a **data driven way**, fitting the data sideband to extract the continuum $\gamma\gamma$ functional form.
  - Multiple families of functions are considered, the best one is used in the fit.
  - An uncertainty coming from the choice of the one family over the others is calculated *via* the **discrete profiling** or “envelope” **method**.

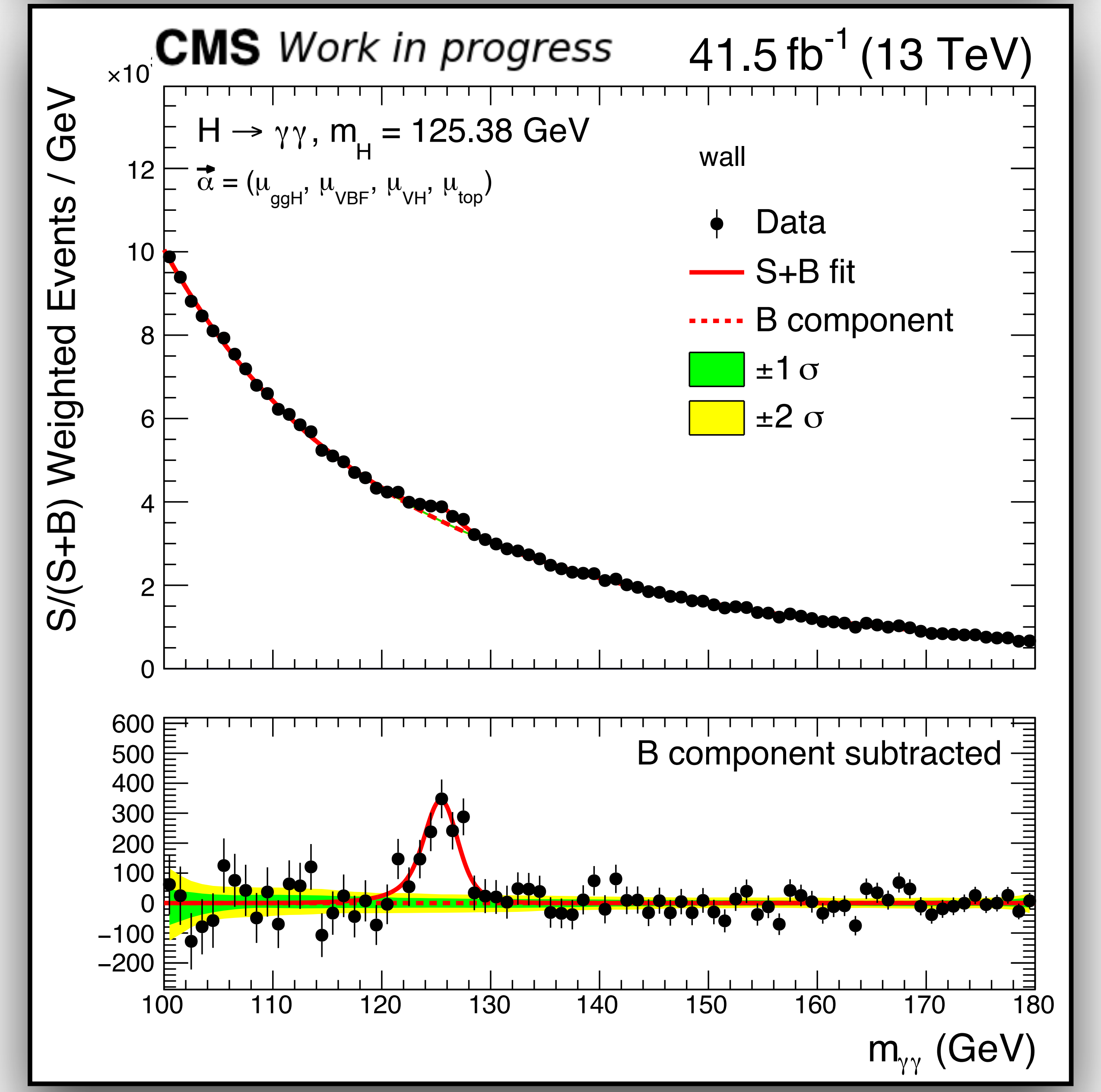


Framework development

HiggsDNA (Diphoton NanoAOD):

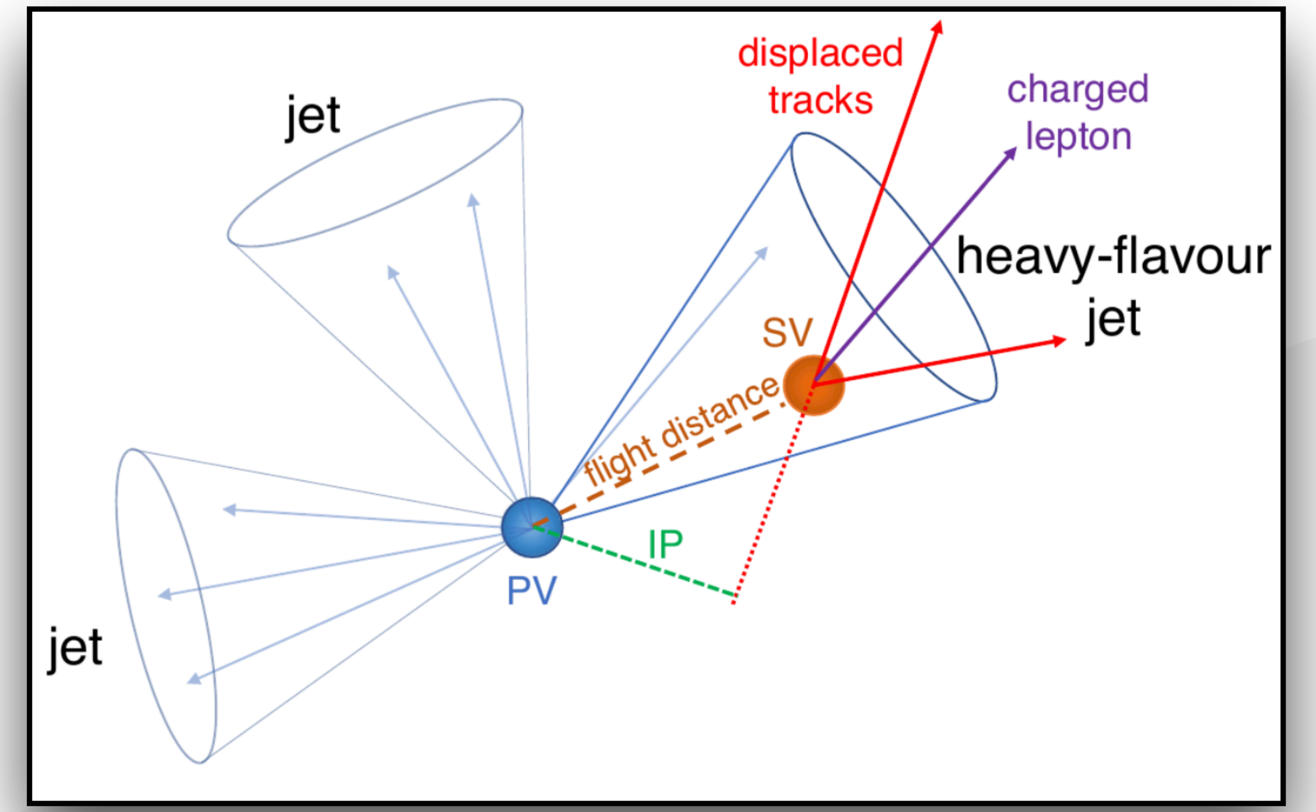
- The CMS $H \rightarrow \gamma\gamma$ group decided to switch to a **new** (Python based) **common framework for Run 3 analysis**.
- Run on flat NTuples (NanoAOD) that are centrally produced.
  $\Rightarrow$ faster running time, columnar analysis.
- Under active development, I joined the group of the core developers to work on the common ingredients.
- I was able to produce the **first complete analysis-like results** reproducing a simplified version of the Run 2 cross section measurement:
  - Event selection.
  - Systematics variations (a few).
  - Statistical fit.



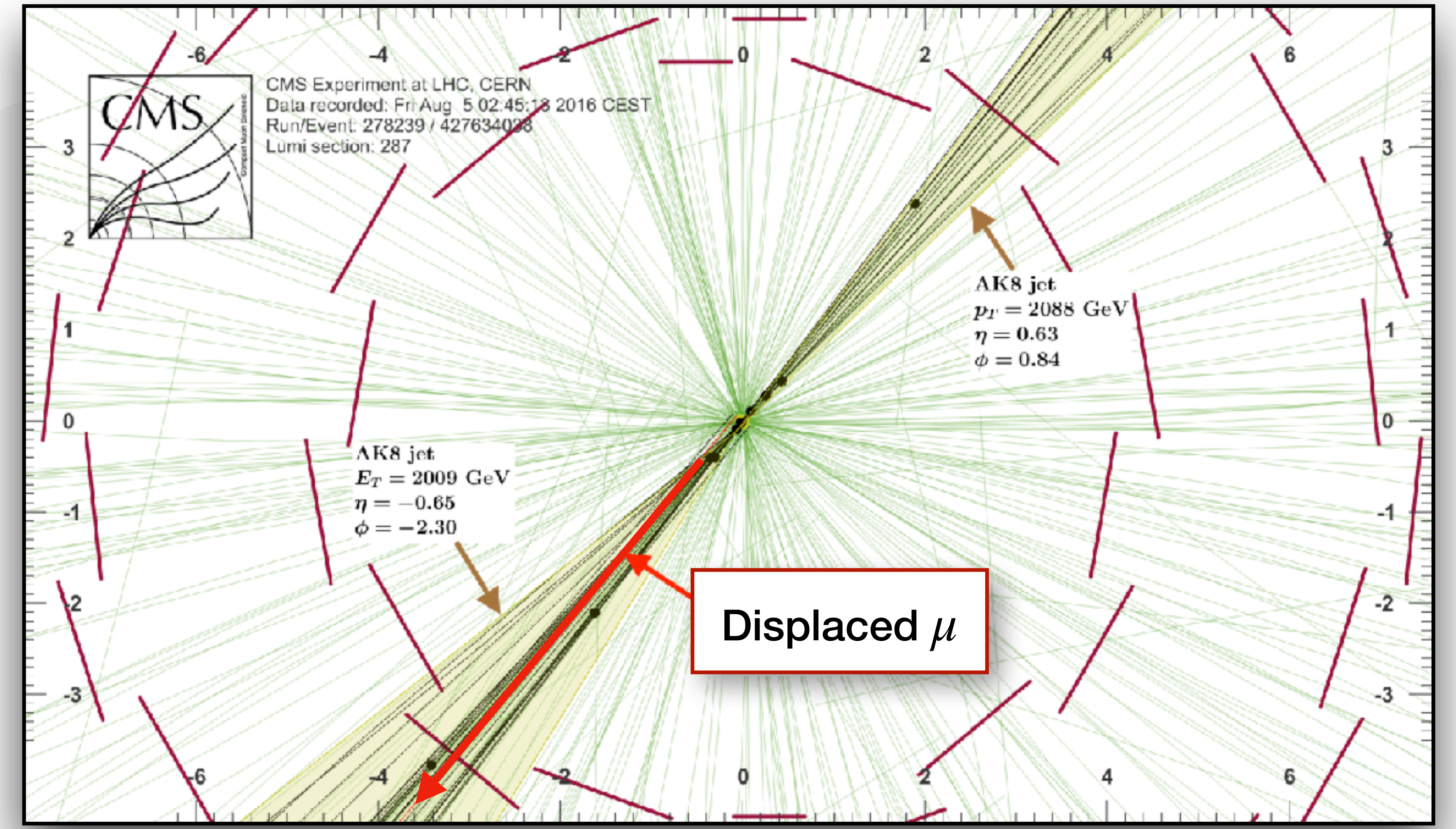


Charm tagging

Heavy flavour hadrons:



- **Have long lifetimes** (for HEP standards!) ~ ps:
  - They travel in the detector ~ mm before decaying.
- The Jets coming from HF quarks **often contain soft non-isolated leptons**.
- CMS is equipped with an excellent Inner Tracker system, the Pixel Detector:
  - 4 Barrel layers (built at PSI) and 6 Forward disks.
  - Excellent 3D spatial resolution ~ 50 μm.
  - Very high granularity ~ 124 M channels.

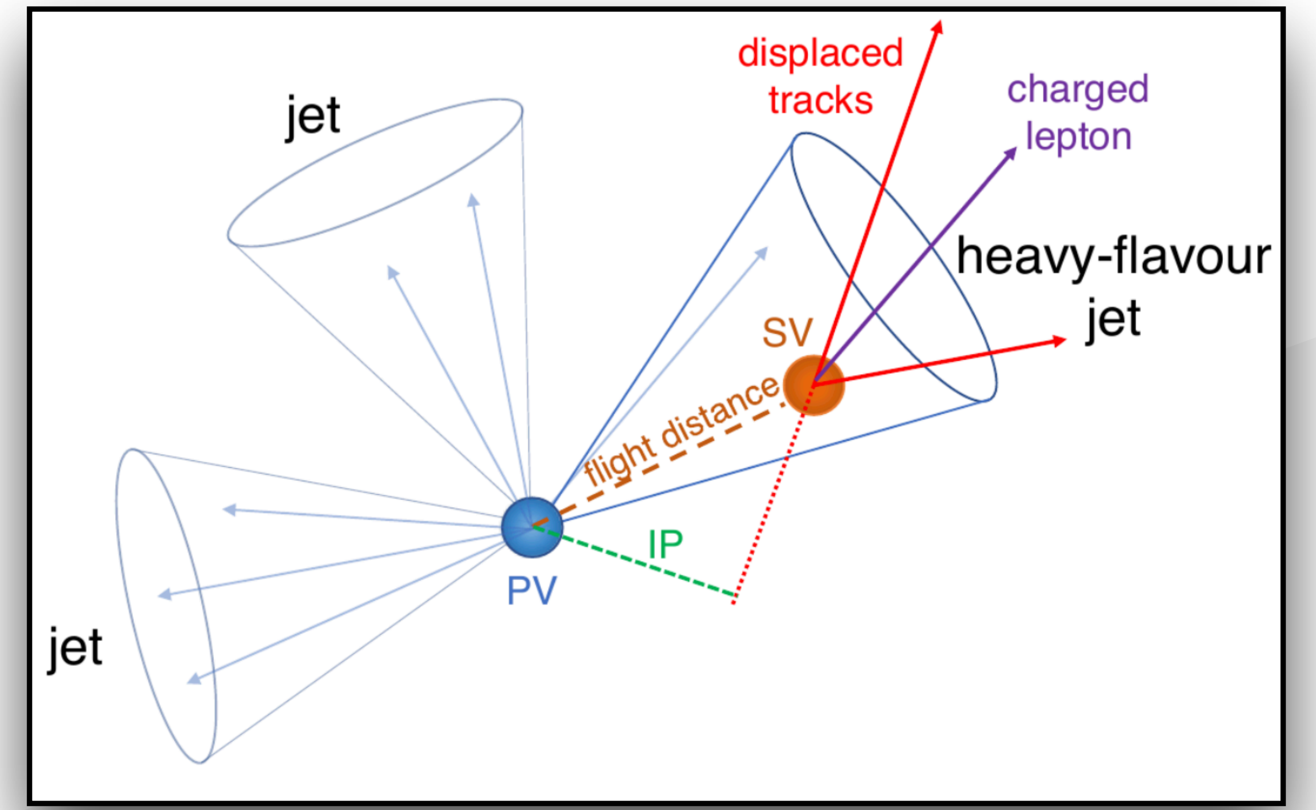


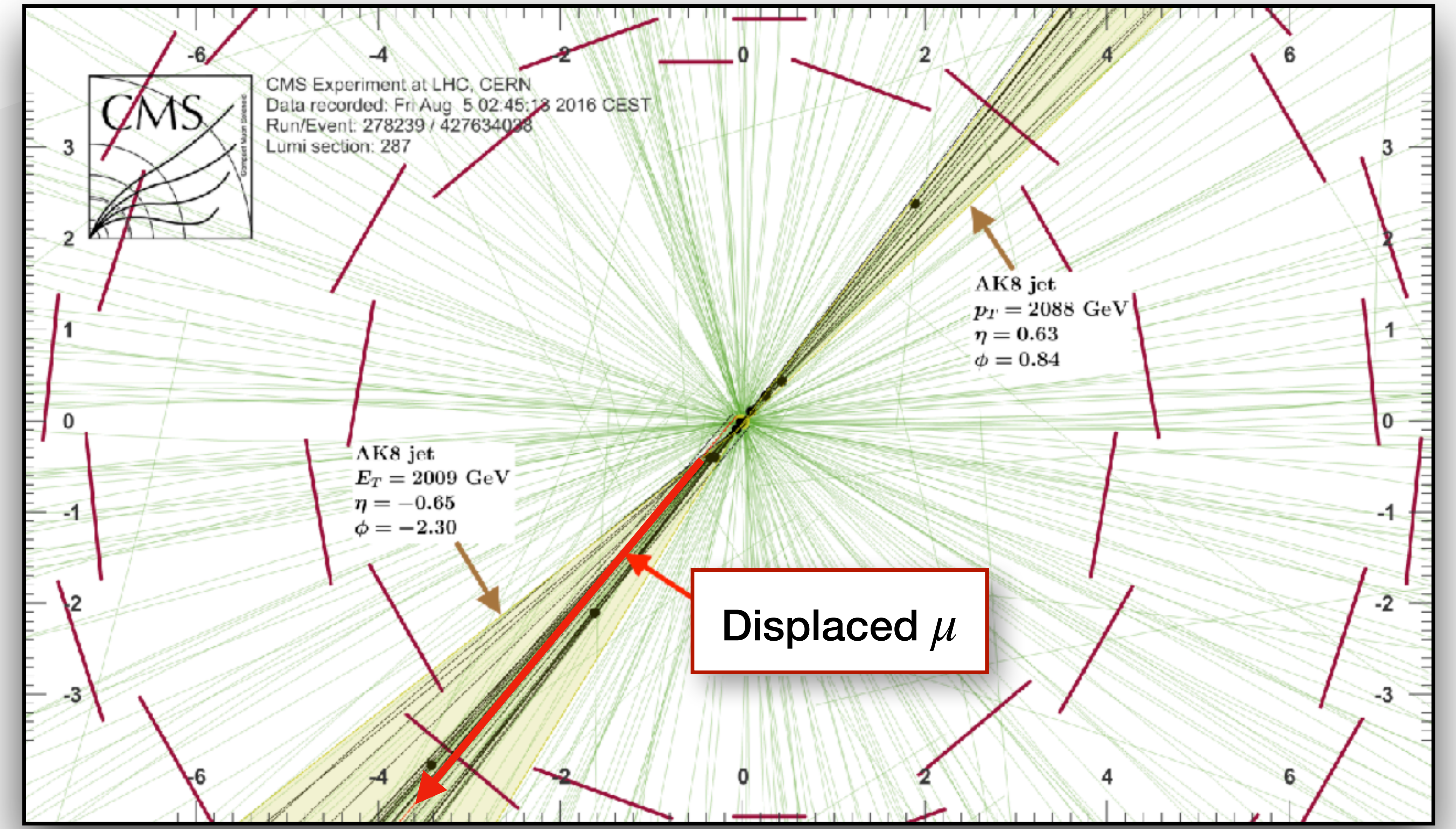
Displaced μ

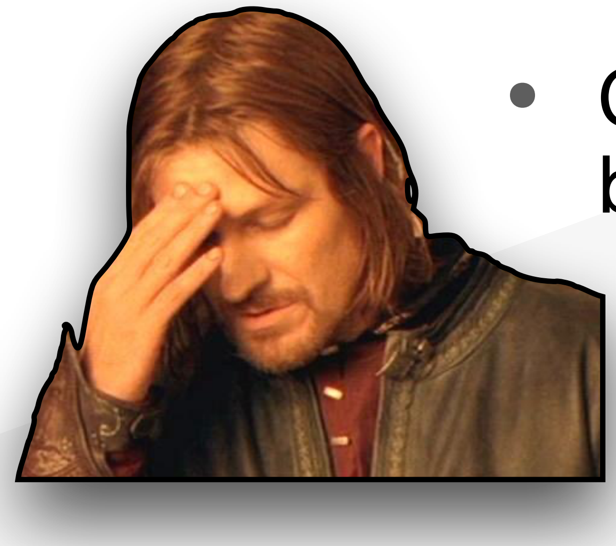
Charm tagging

Heavy flavour hadrons:

- **Have long lifetimes** (for HEP standards!) ~ ps:
    - They travel in the detector ~ mm before decaying.
- The Jets coming from HF quarks **often contain soft non-isolated leptons**.
- CMS is equipped with an excellent Inner Tracker system, the Pixel Detector:
    - 4 Barrel layers (built at PSI) and 6 Forward disks.
    - Excellent 3D spatial resolution ~ 50 μm.
    - Very high granularity ~ 124 M channels.
- Nonetheless one does not simply do Charm tagging:
    - Charmed hadrons have **intermediate properties** between bottom and light Jets.
        - B mesons decay can include D mesons.



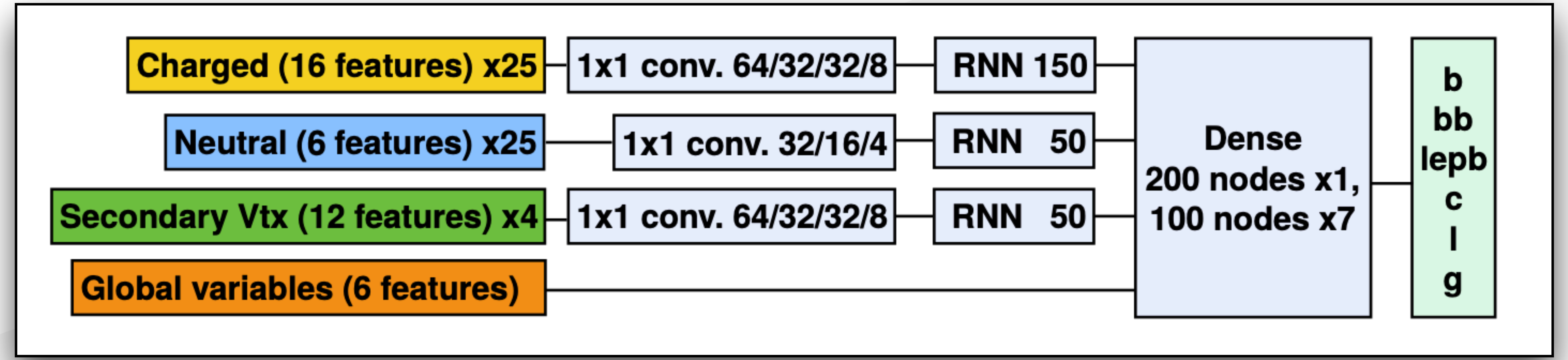

CMS Experiment at LHC, CERN
Data recorded: Fri Aug 5 02:45:18 2016 CEST
Run/Event: 278239 / 427634038
Lumi section: 287



Charm tagging

DeepJet algorithm:

- A complex Neural Network (NN) based discriminator is used to identify charmed jets in CMS.
- It exploits more than 600 input variables:
  - ⇒ global variables,
  - ⇒ charged candidate features,
  - ⇒ neutral candidate features,
  - ⇒ SV features.
- At Medium WP (2017): c-tag efficiency → 60%, b/light-mistag rate → 26%.



Charged (16 features) x25 — 1x1 conv. 64/32/32/8 — RNN 150

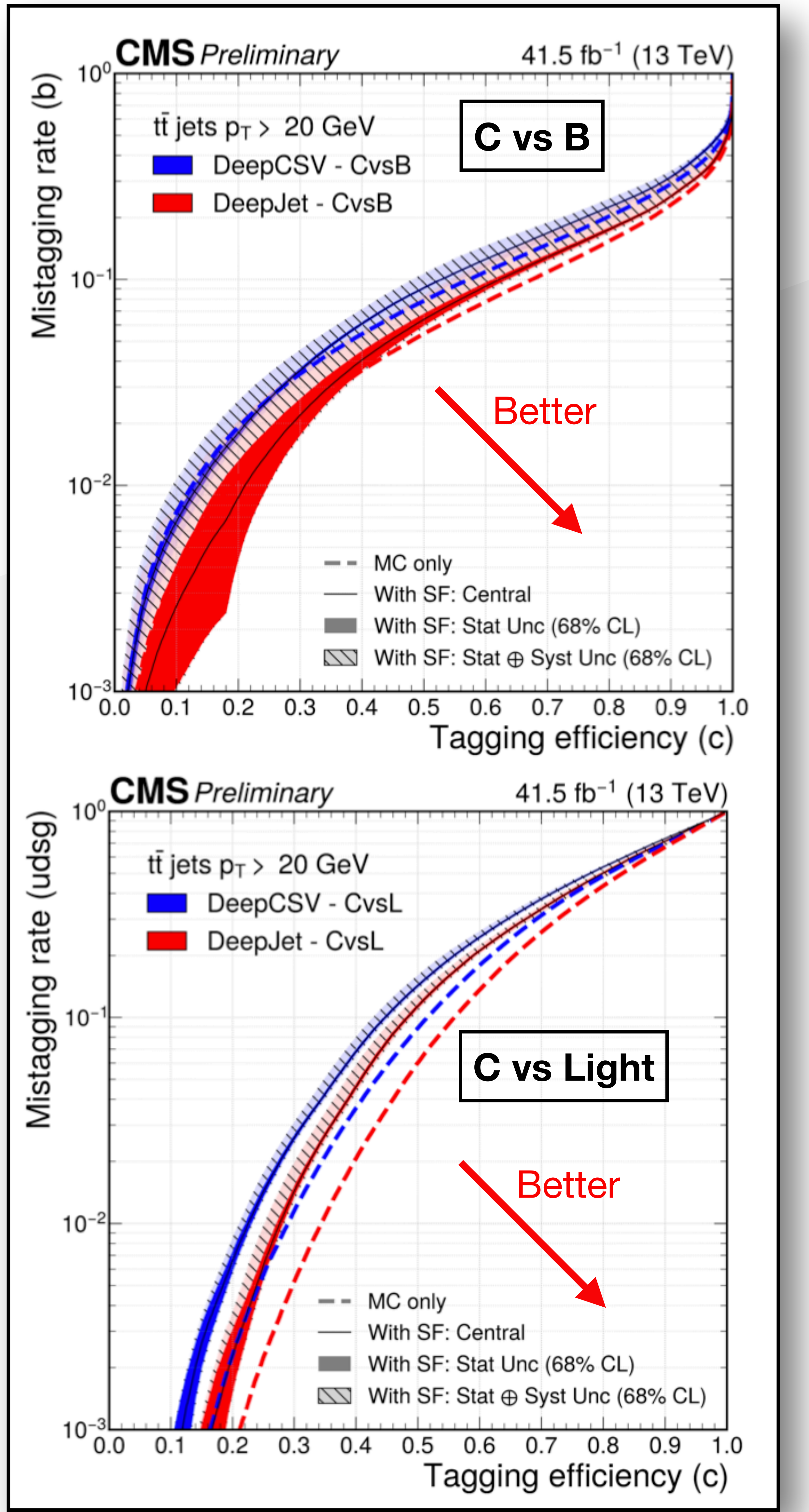
Neutral (6 features) x25 — 1x1 conv. 32/16/4 — RNN 50

Secondary Vtx (12 features) x4 — 1x1 conv. 64/32/32/8 — RNN 50

Global variables (6 features)

Dense 200 nodes x1, 100 nodes x7

b
bb
lepb
c
l
g



Outline

Introduction:

- Motivation.
- State of the art.
- New approaches.

Ingredients:

- Monte Carlo simulation.
- $H \to \gamma\gamma$ method.
- Framework development.
- Charm tagging.

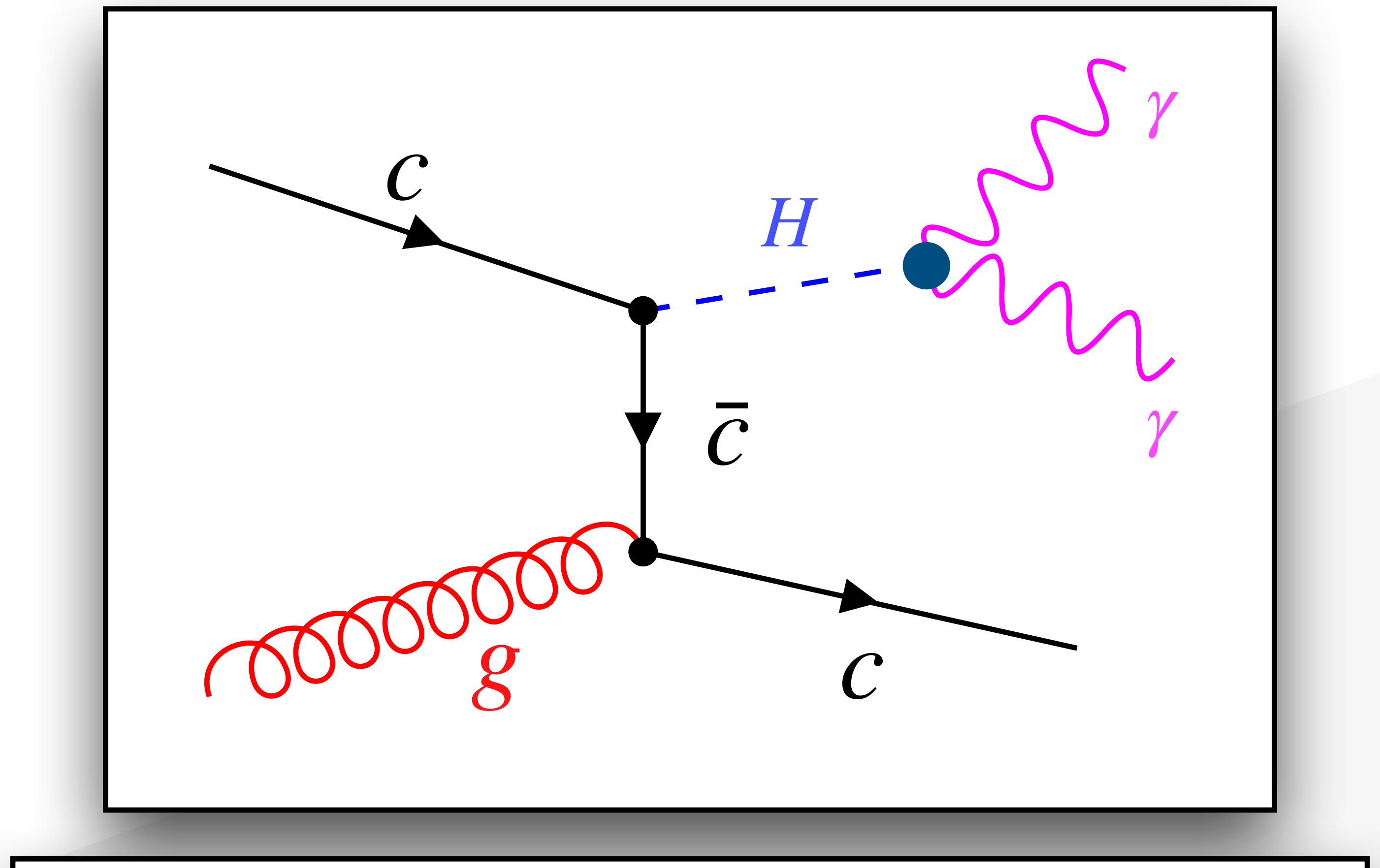
H+c analysis status:

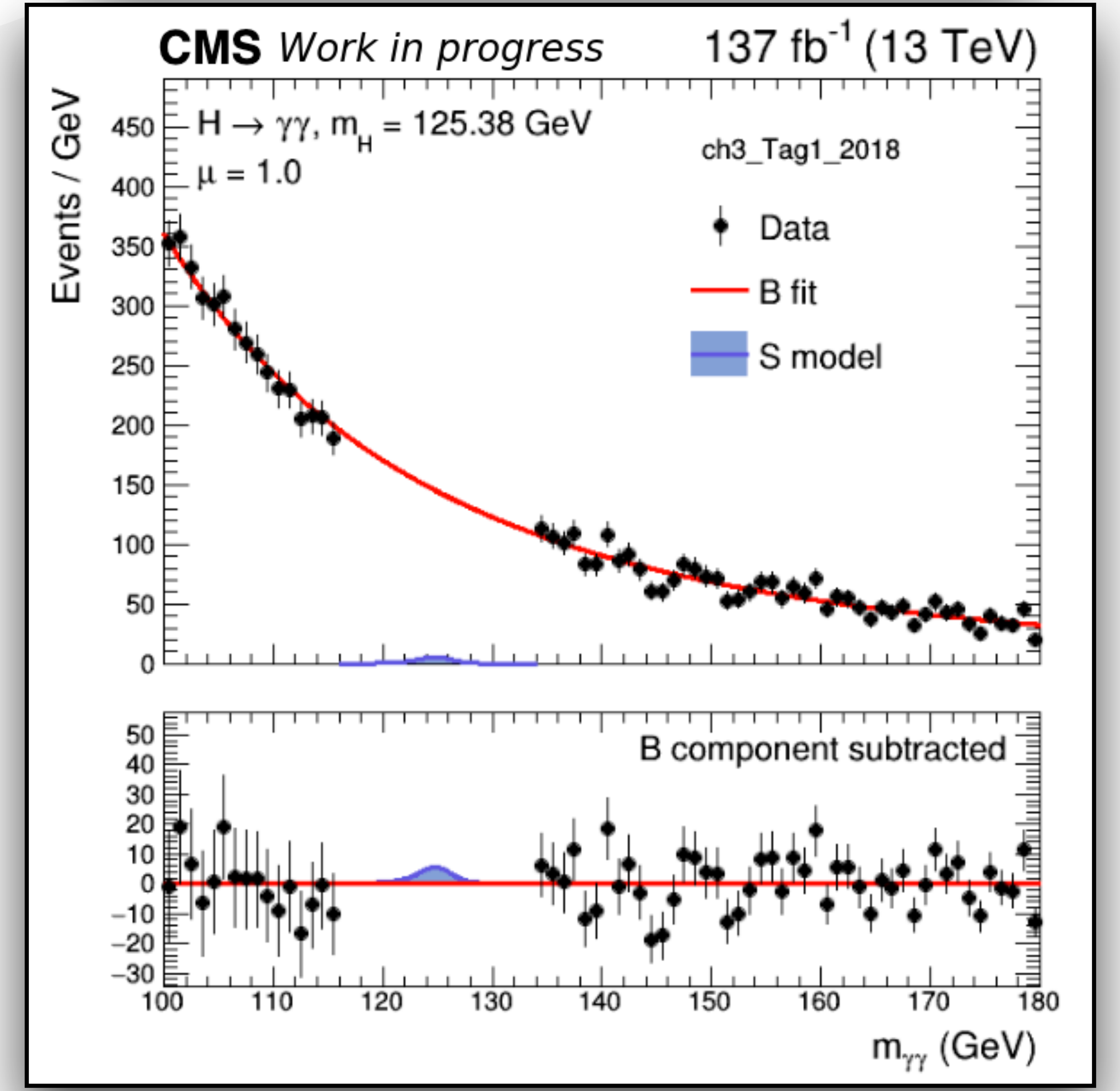
- Analysis strategy.
- Preliminary results.

H+c analysis status:

Analysis strategy:

- We select events with:
  - A good diphoton candidate.
  - At least one c-jet (DeepJet tagger score CvsL > 0.25).
- Main backgrounds:
  - Irreducible “standard” Higgs production through gluon fusion (ggH).
  - Continuous diphoton background (CB) from $\gamma\gamma$ and $\gamma + jets$ events (as in the plain $H \to \gamma\gamma$ analysis).





H+c analysis status:

2D Categories:

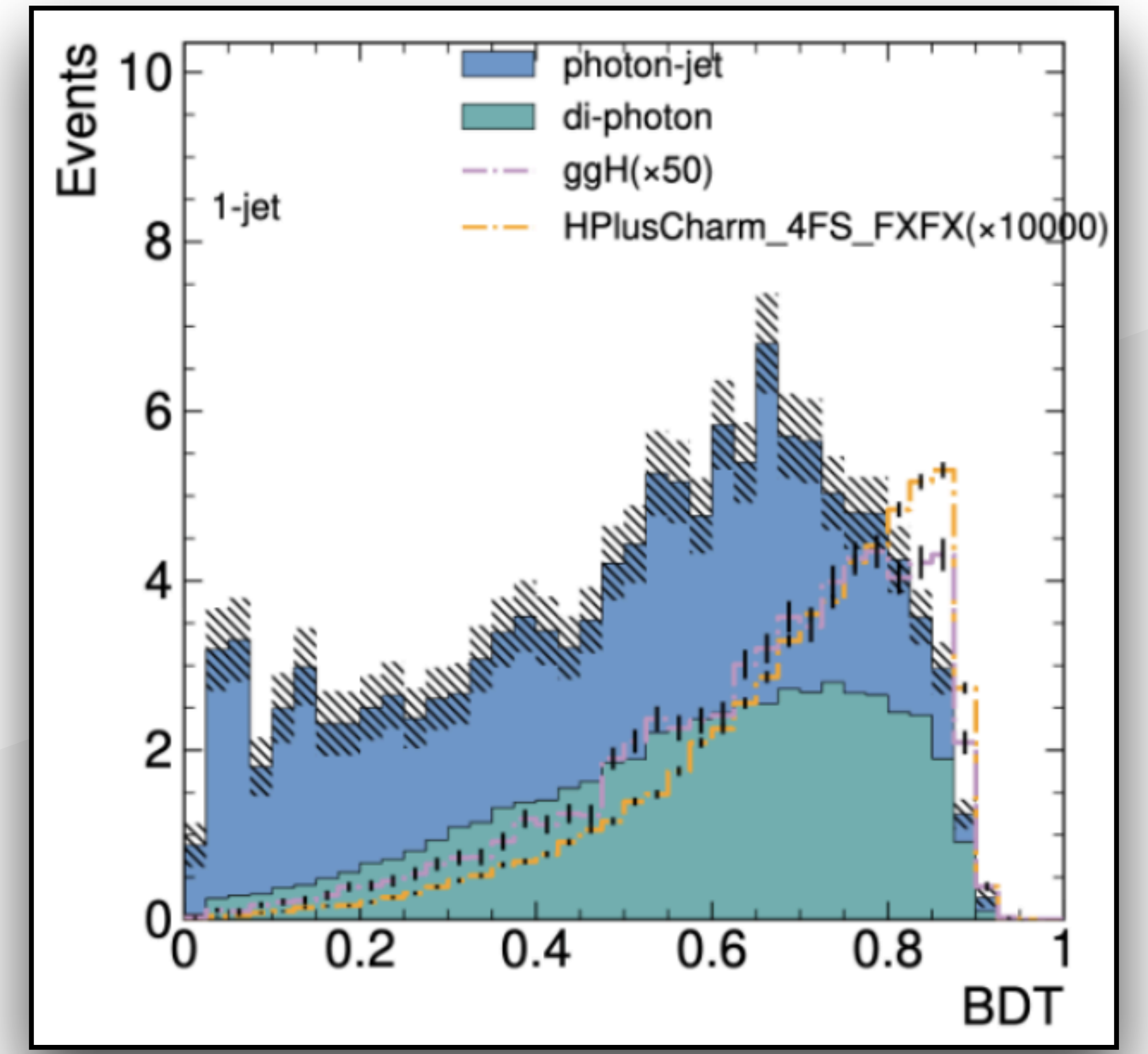
- We separate events in 9 categories according to:
  - A Boosted Decision Tree (BDT) trained to distinguish $H + c$ events from ggH.
  - A BDT trained to distinguish $H + c$ events from the continuous $\gamma\gamma$ background.

BDT training:

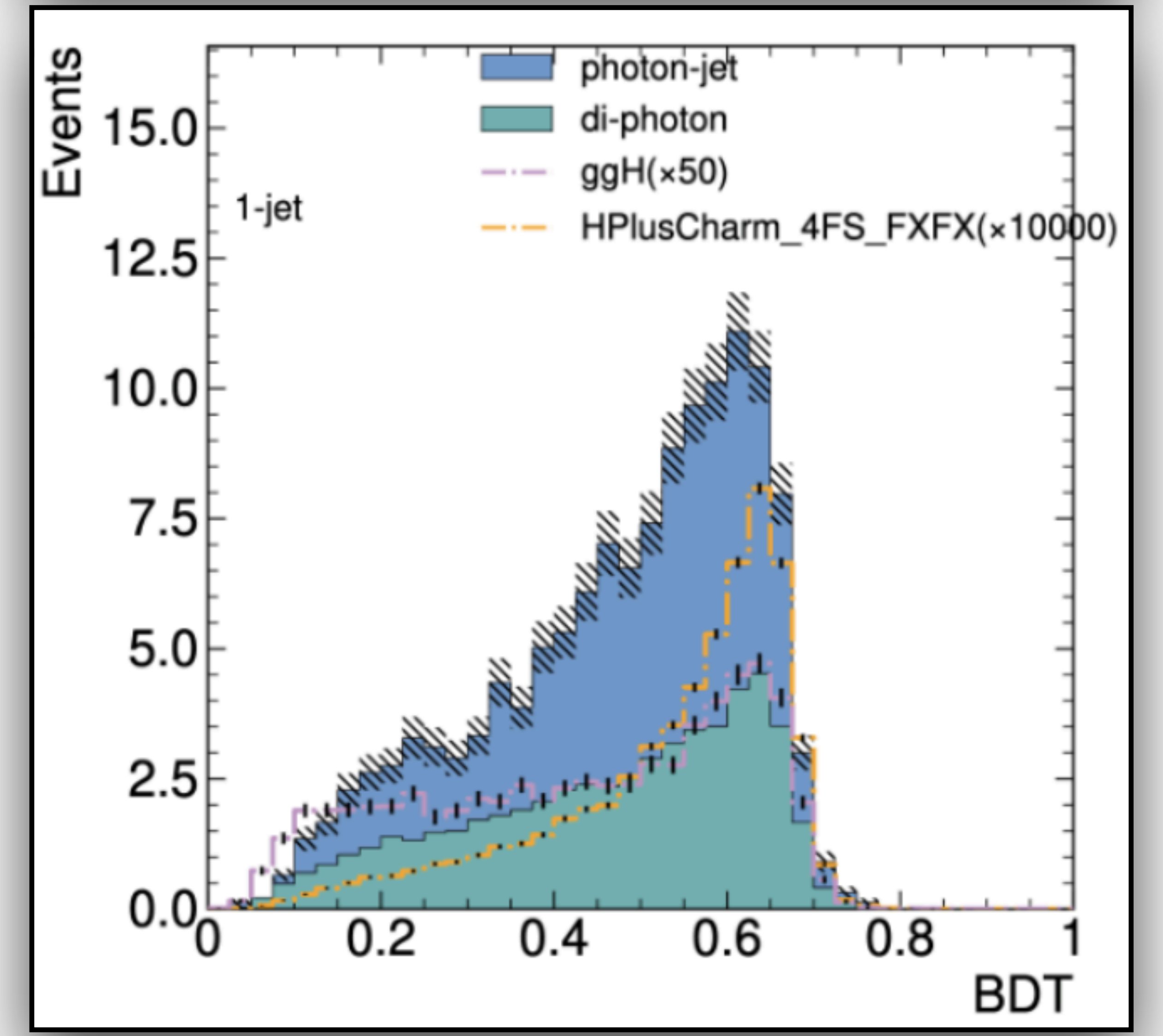
- Separation is achieved exploiting the kinematics of the Photons and Jets in the event.
- To address the irreducible ggH background we avoid using c-tagging information in the BDT training:

  $\Rightarrow$ use low BDT score regions to constrain ggH directly from data.

cH vs CB BDT score



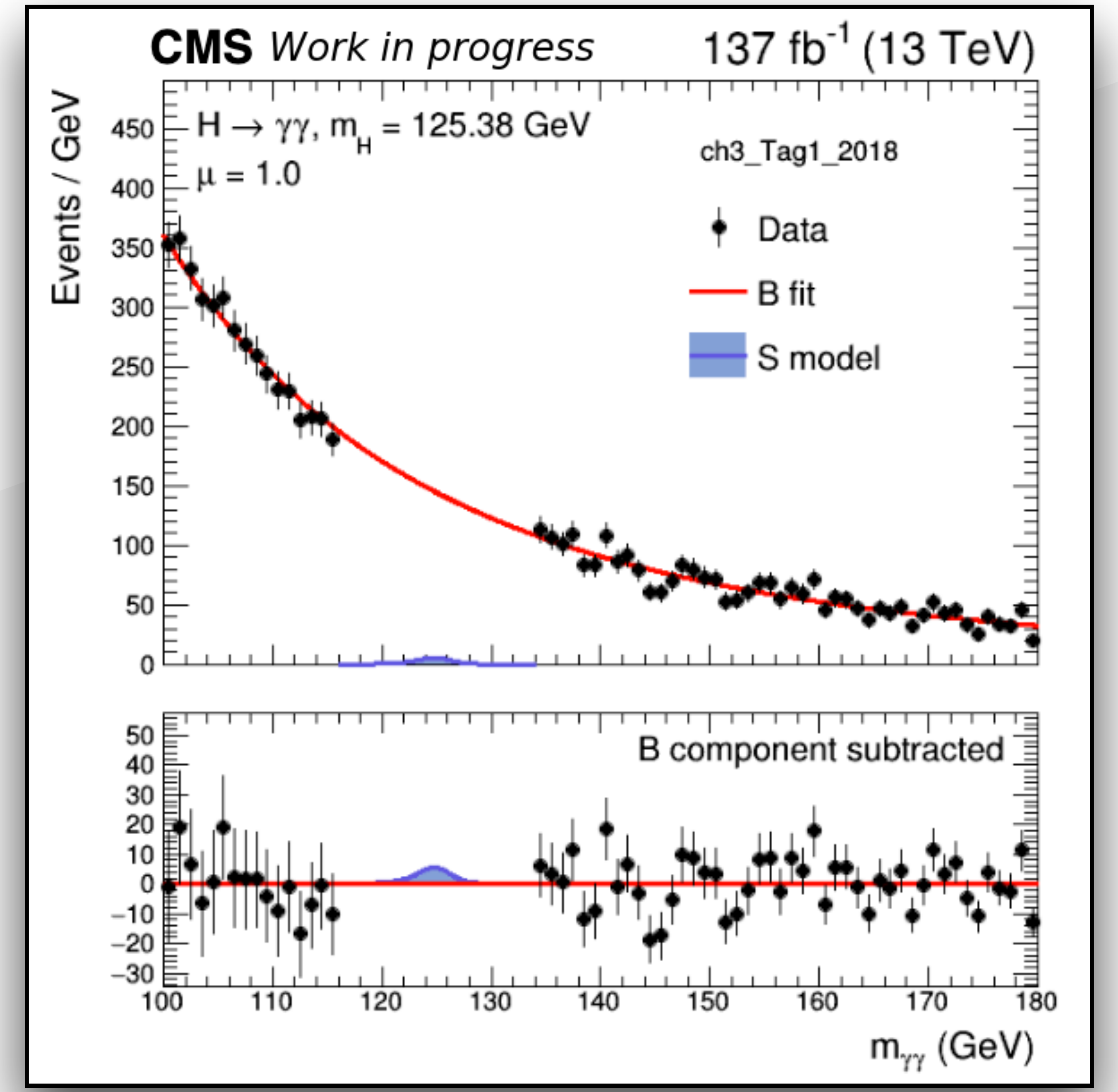
cH vs ggH BDT score



H+c analysis status:

Preliminary results:

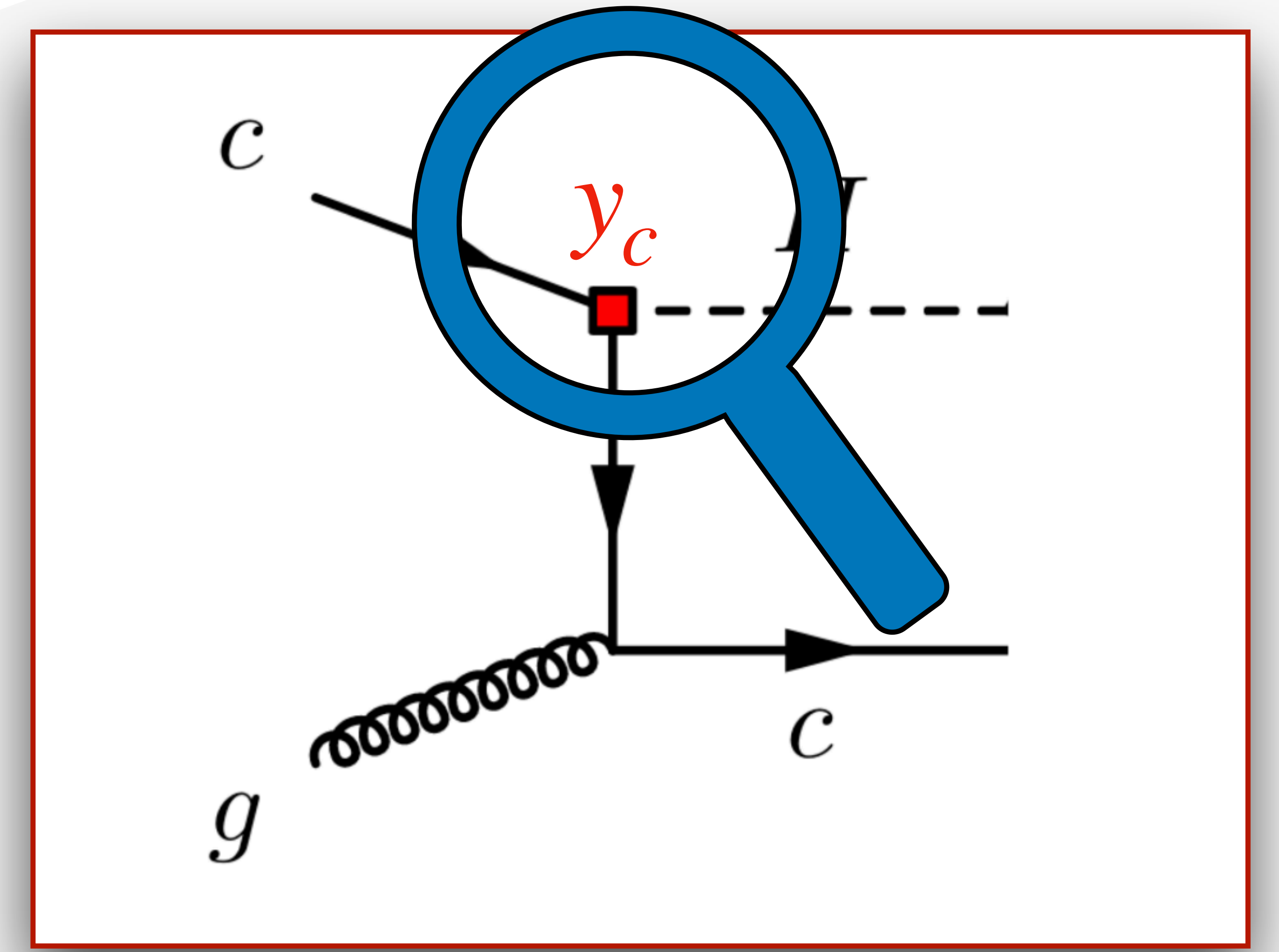
- The Run 2 analysis is still blinded, we’ll be going to CMS internal review soon.
- The expected limit on k_c is of $O(20)$.
- We’re planning of improve the analysis strategy, moving to new, more performing taggers and including Run 3 data.



Conclusions

Summary:

- 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.

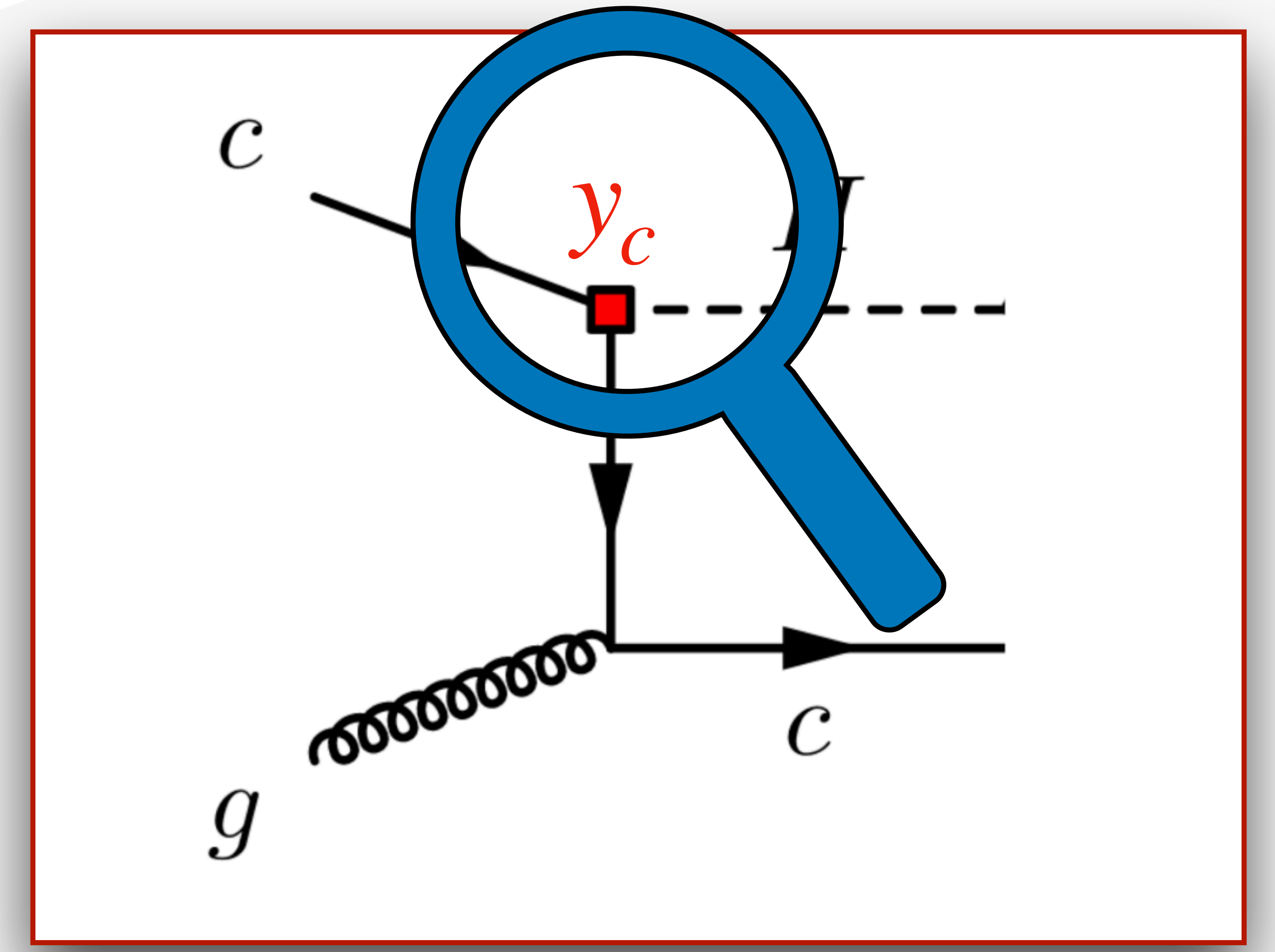


Conclusions

Summary:

- 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

Status: Common tools

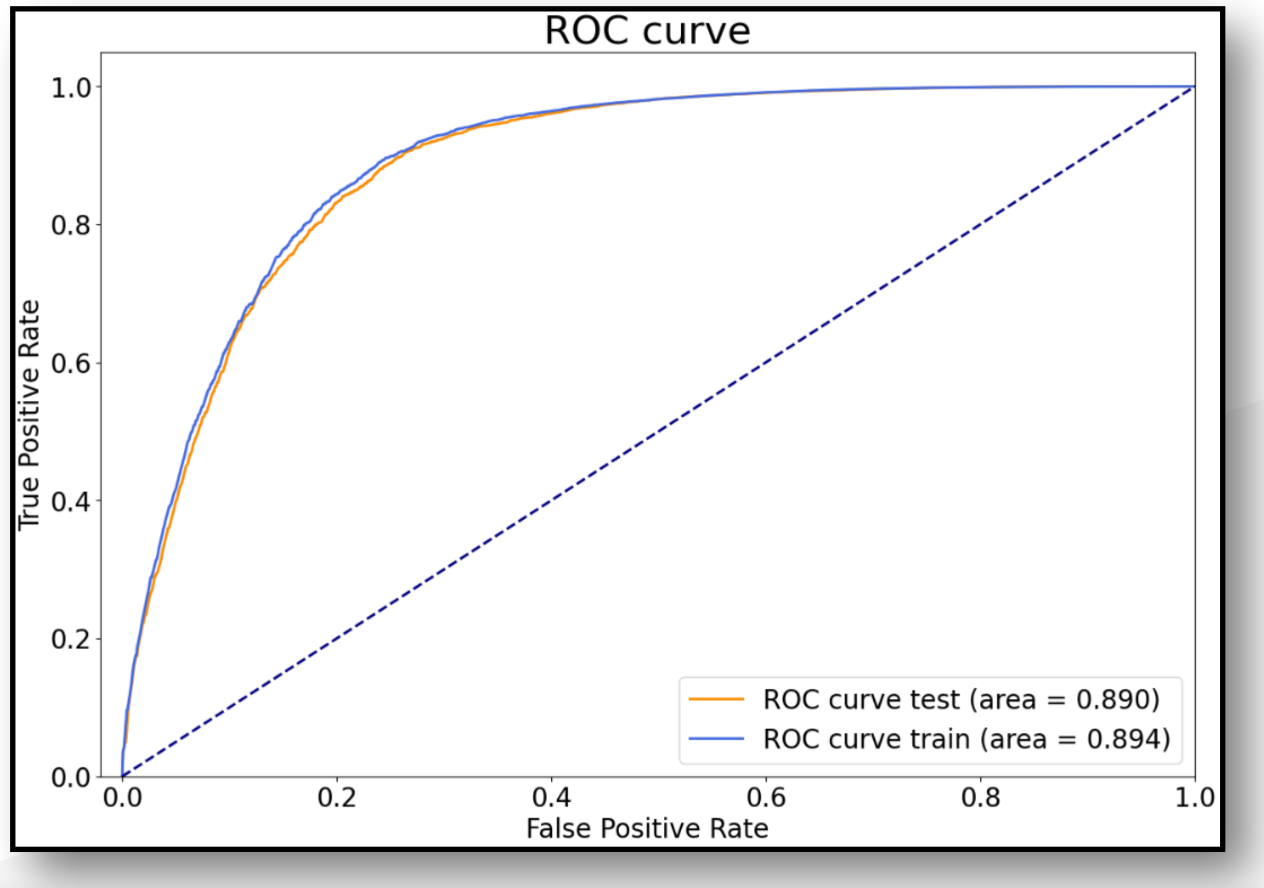
HiggsDNA:

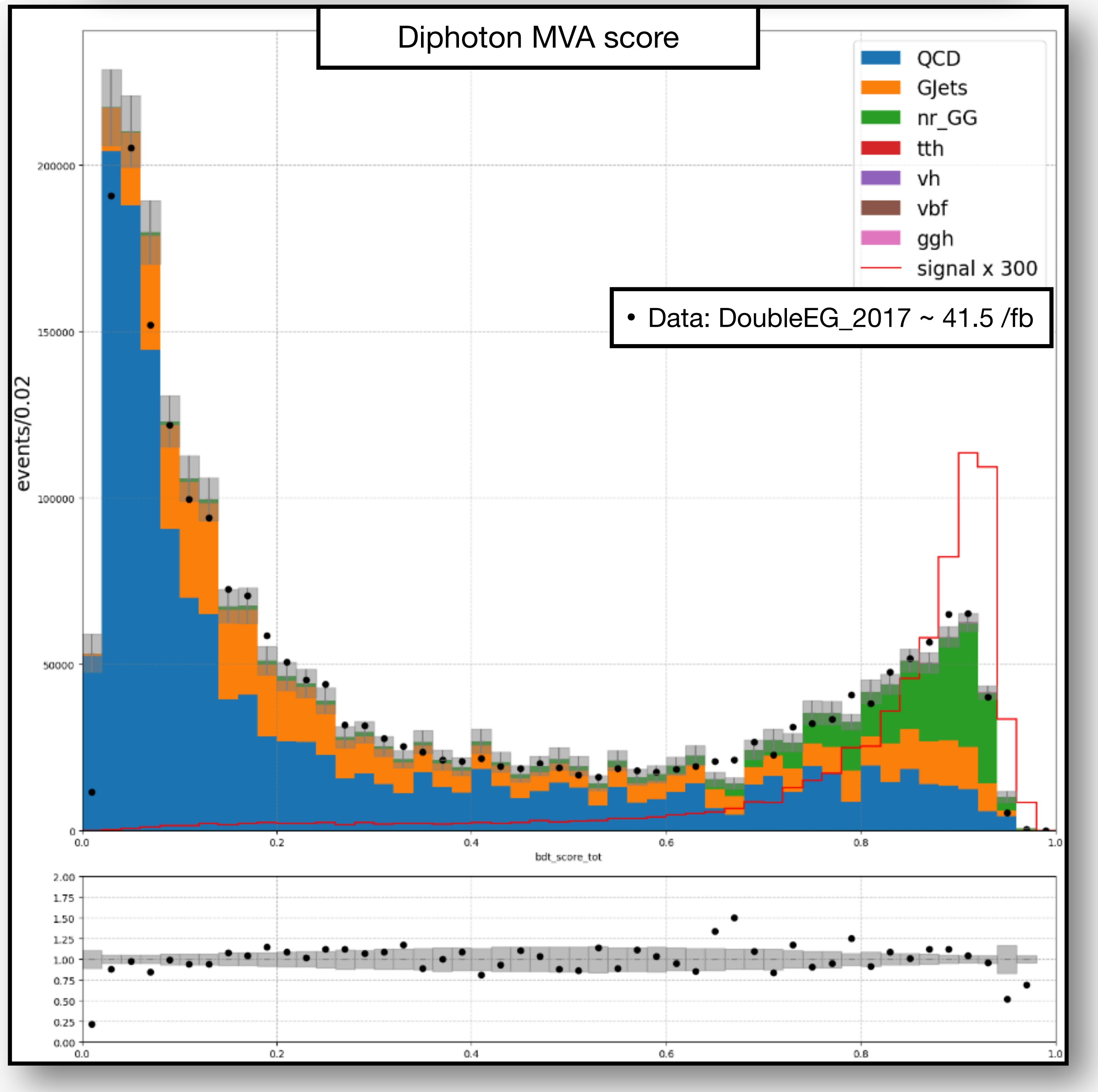
- The framework is **ready to perform an analysis** with Run 3 data.
  - Individual corrections/systematics may still need to be implemented, but the machinery is already in place alongside the main corrections.
  - Development of analysis specific tools is going on.

Diphoton ID BDT:

- Used to discriminate signal and fake γ background.
- In particular I'm focusing on the retraining of the BDT for Run 3:
  - Performed test on 2017 Data starting from nAOD.
  - Results are good but further comparisons with the old framework are needed:

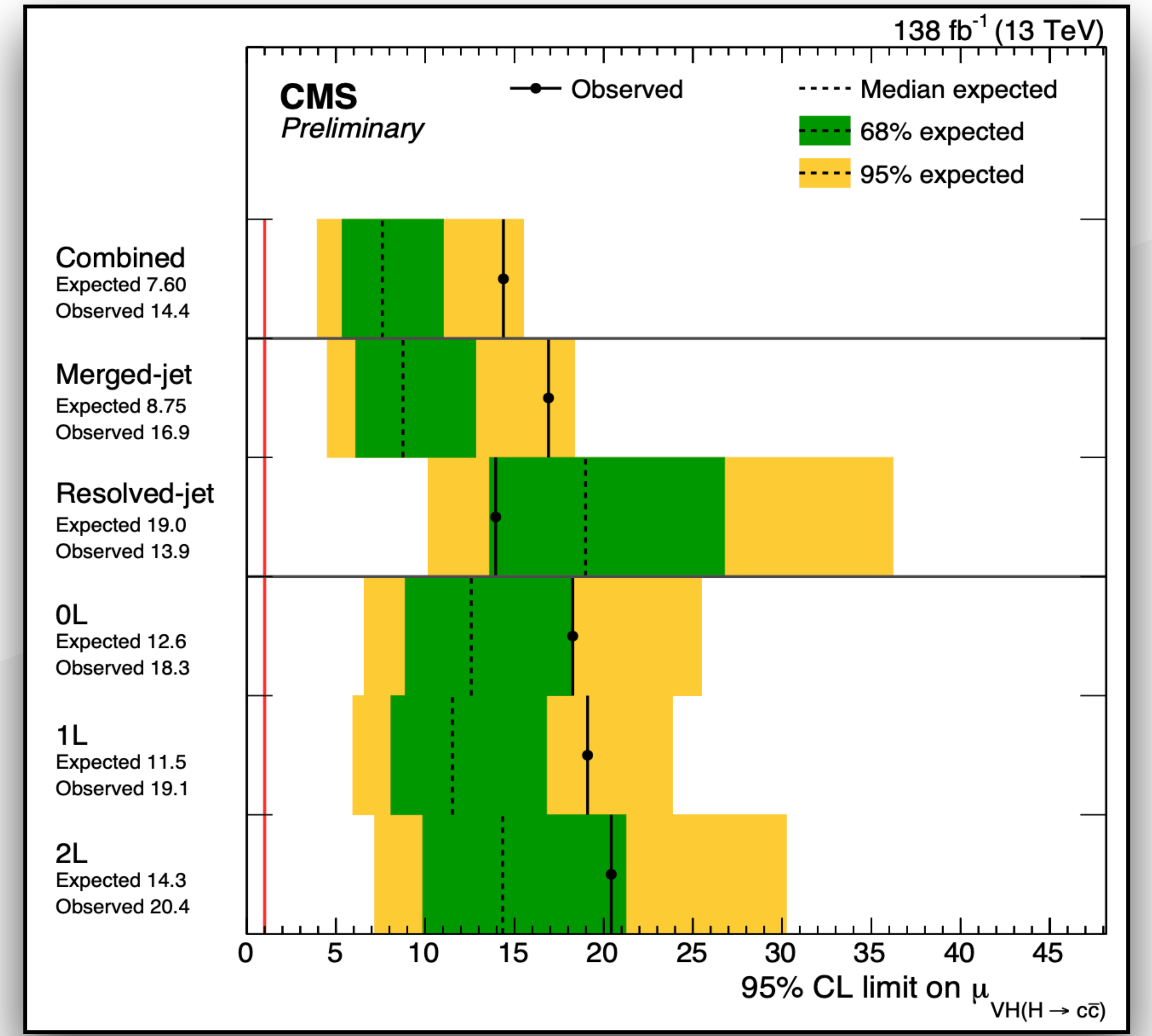
    $\Rightarrow$ **Performances of new framework training are better than with the Run 2 one.**





Previous results

- Direct search for $VH(H \rightarrow c\bar{c})$ arXiv:2205.05550: recent improvements, most stringent limit on $H \rightarrow c\bar{c}$.
  - Upper limit $\mu_{VH(H\rightarrow c\bar{c})} < 14\ (7.6)$ 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 $Z \rightarrow c\bar{c}$ at a hadron collider (5.7σ)
- Boosted $ggH(H \rightarrow c\bar{c})$ HIG-21-012:
  - $\mu < 38\ (45)$ observed (expected) at 95% C.L.
- Exclusive $H \rightarrow J/\Psi + \gamma$ decays, clean signature, $J/\Psi \rightarrow \mu\mu$ 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.

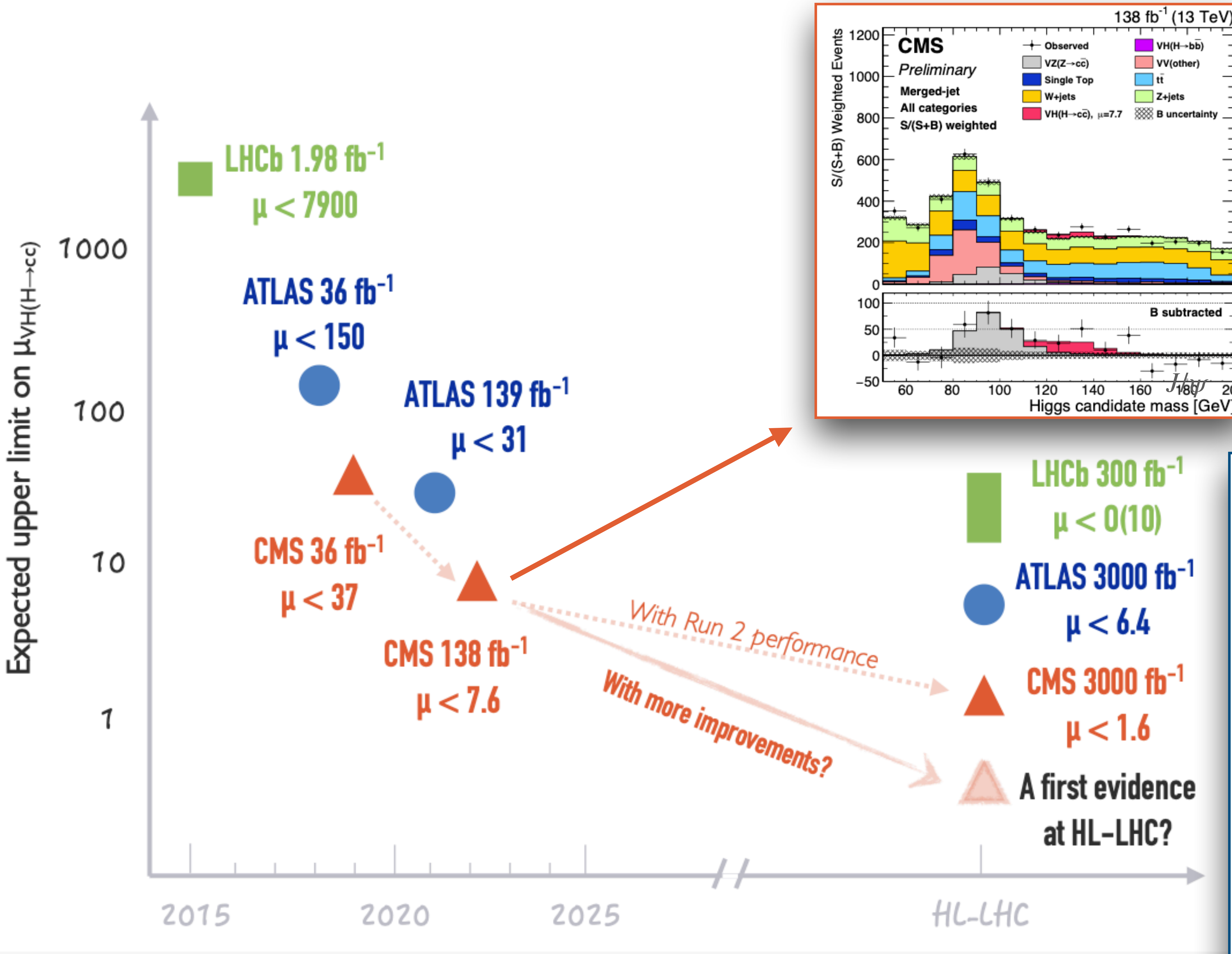


---

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

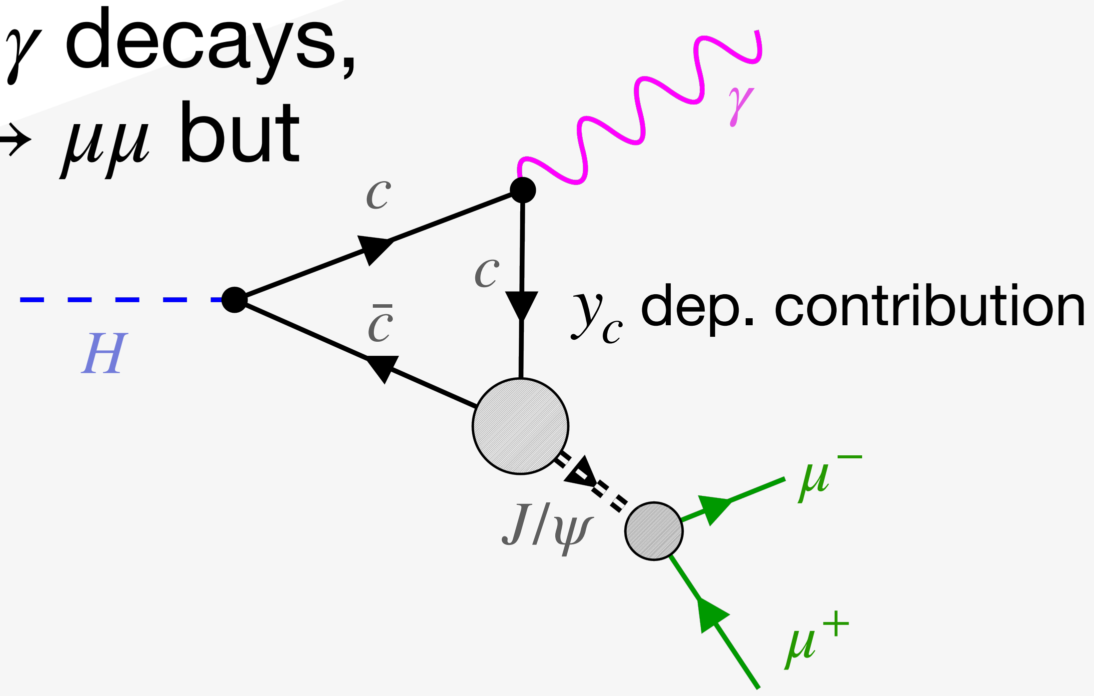
Previous results

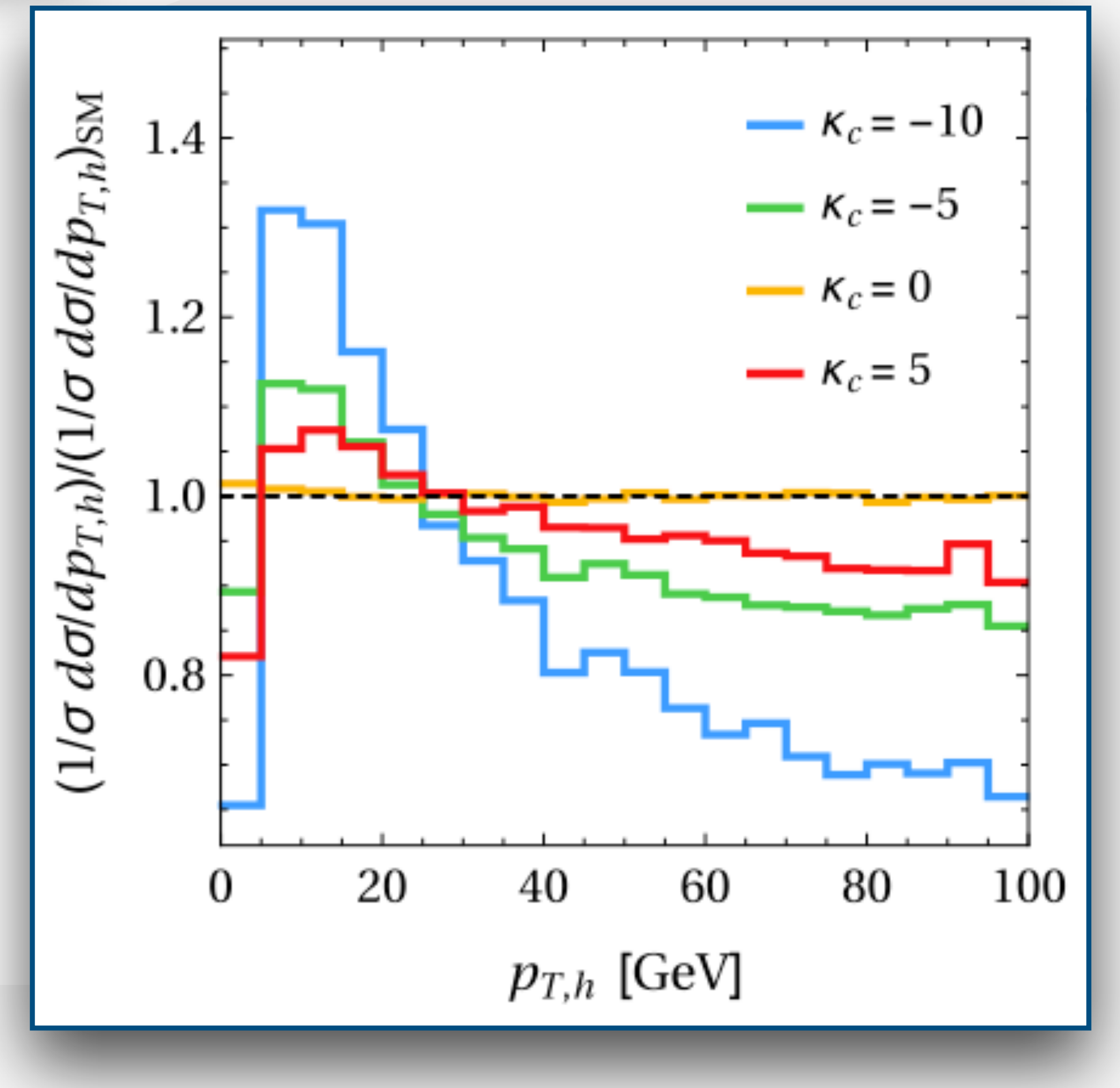
- **Direct search** for $VH(H \to c\bar{c})$ arXiv:2205.05550: recent improvements, most stringent limit on $H \to c\bar{c}$.



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

- **Exclusive** $H \to J/\Psi + \gamma$ decays, clean signature, $J/\Psi \to \mu\mu$ but very rare process:





- H differential measurements, variation of $p_T(H)$ as a function of k_c[*].

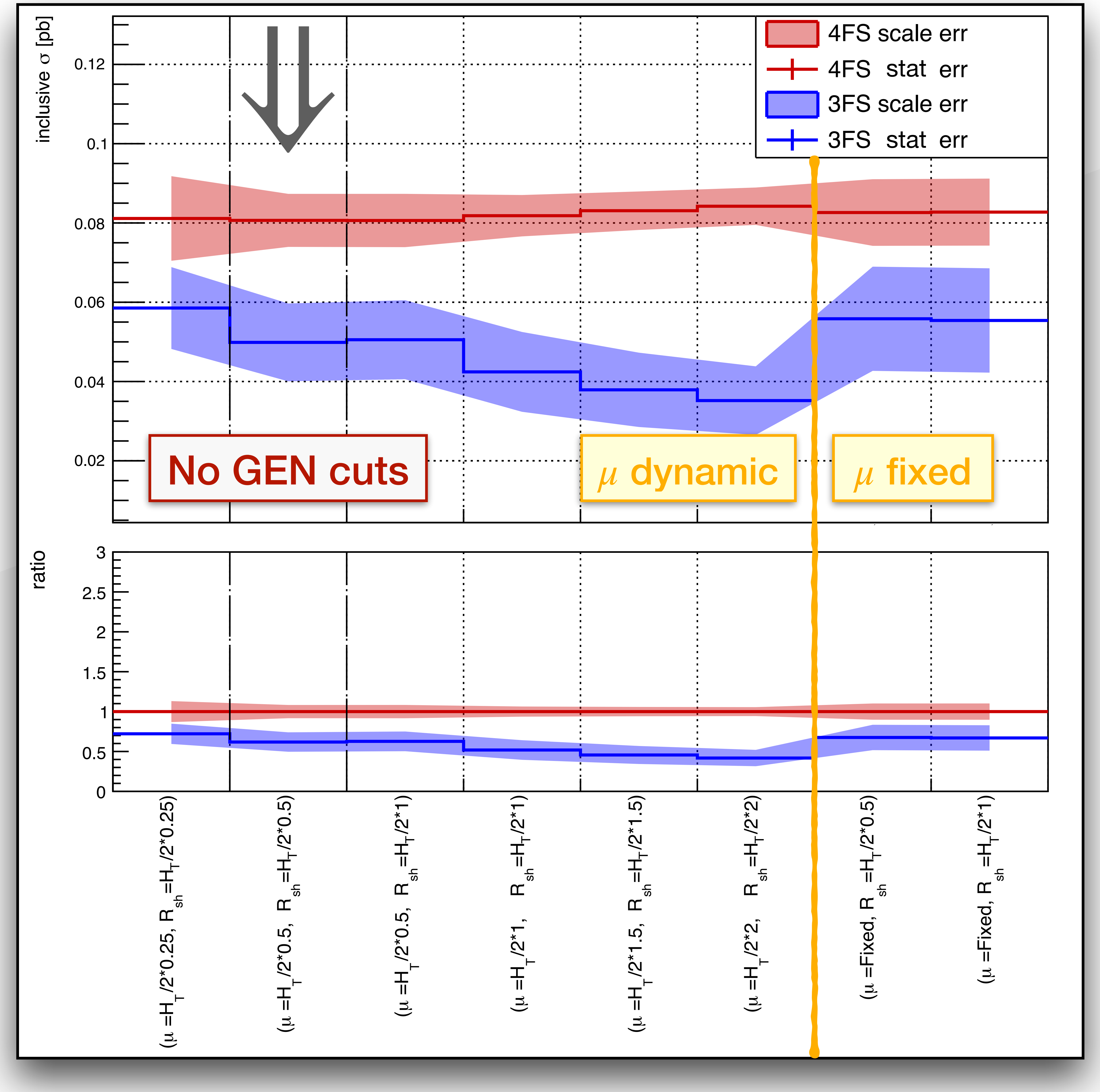
[*] $k_c = y_c / y_c^{SM}$

Phys. Rev. Lett. 118, 121801

GEN-level results

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

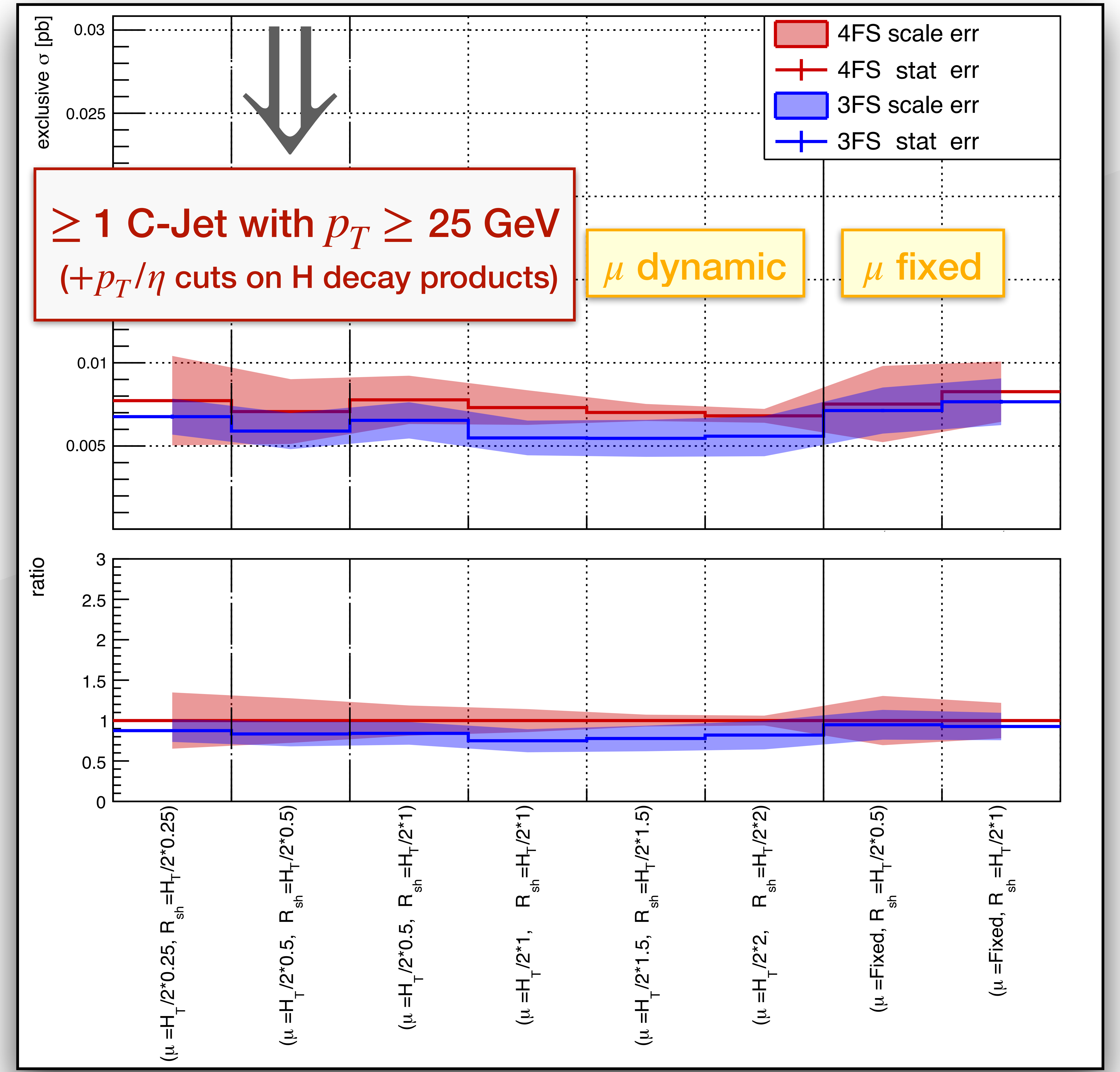
- Large differences between 3FS and 4FS on the inclusive cross sections (up to ~2x).
- To minimise uncertainty on nominal sample (4FS-FXFX) we studied the dependence of the X-section on MG scale parameters (μ_R, μ_F, R_{SH}).
- Uncertainties: μ_R/μ_F scale ~ 15%, PDF ~ 5-10%.



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).



$\geq$ 1 C-Jet with $p_T \geq 25$ GeV

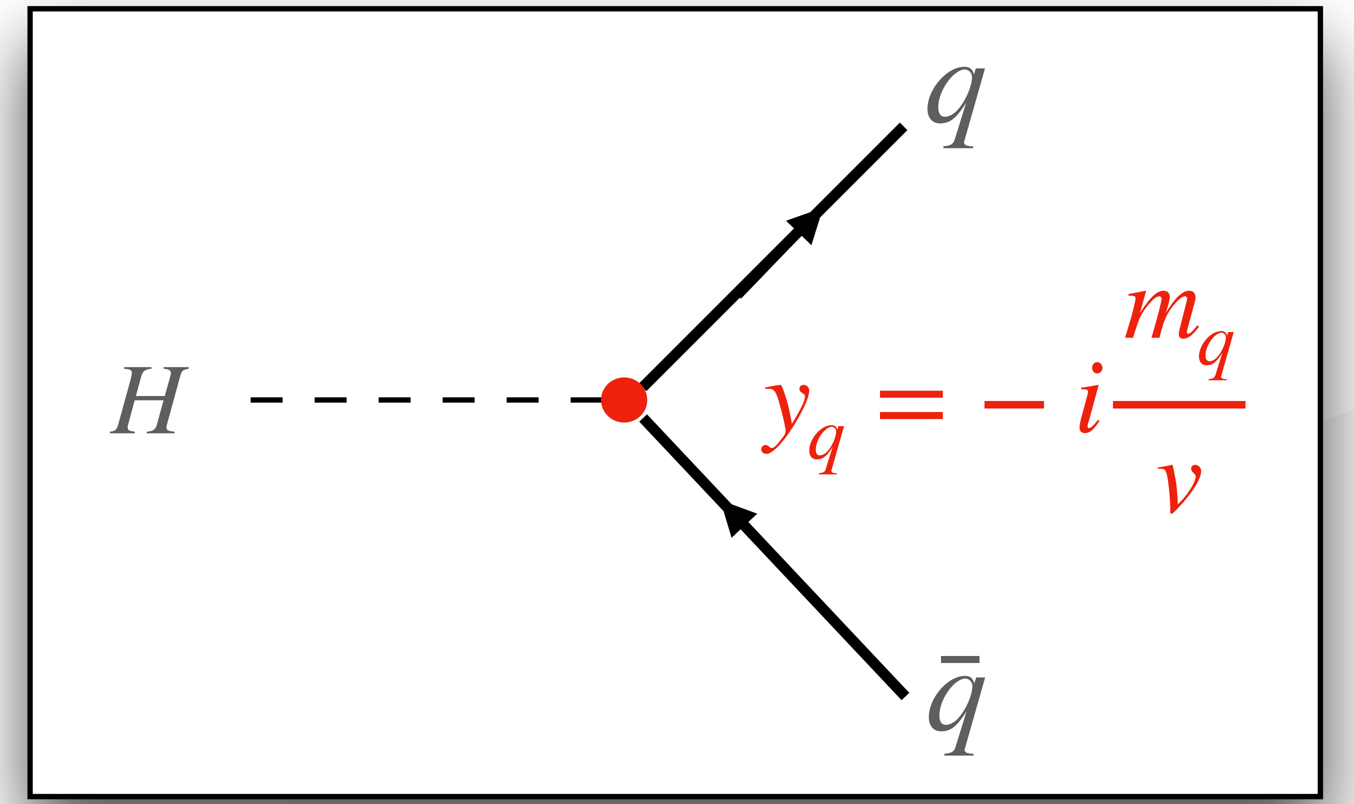
($+p_T/\eta$ cuts on H decay products)

μ dynamic

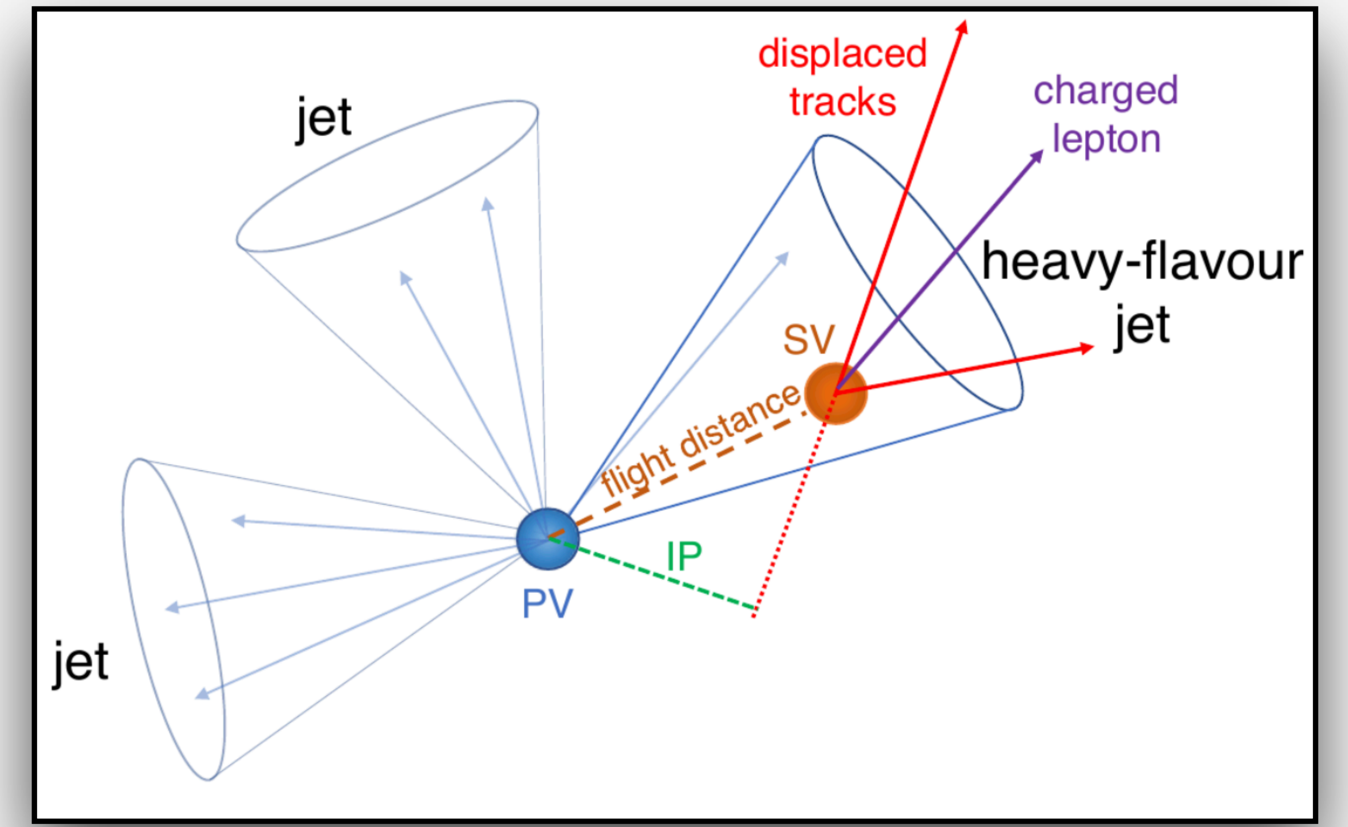
μ fixed

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.



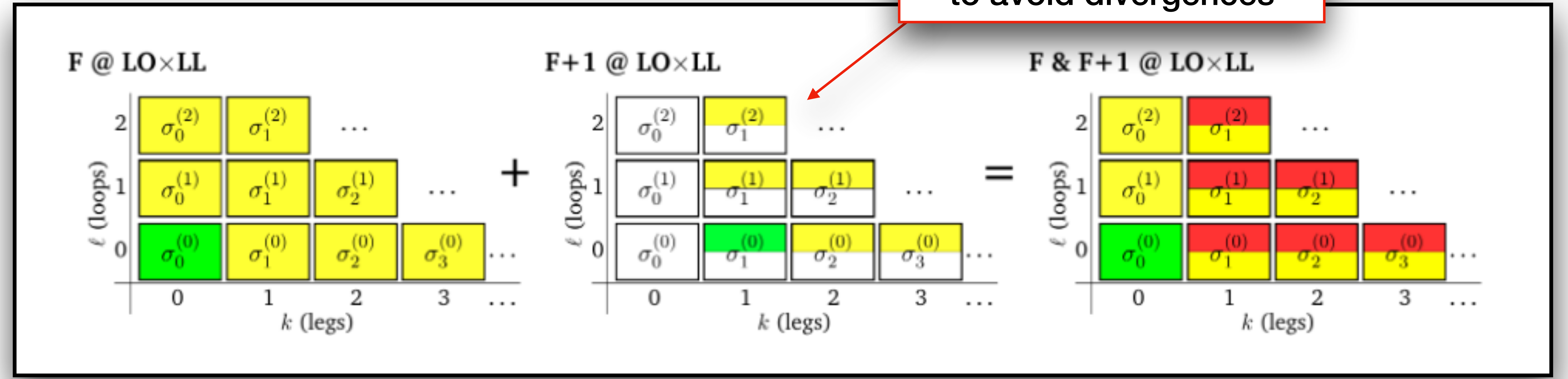
Matching and Merging:

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 !

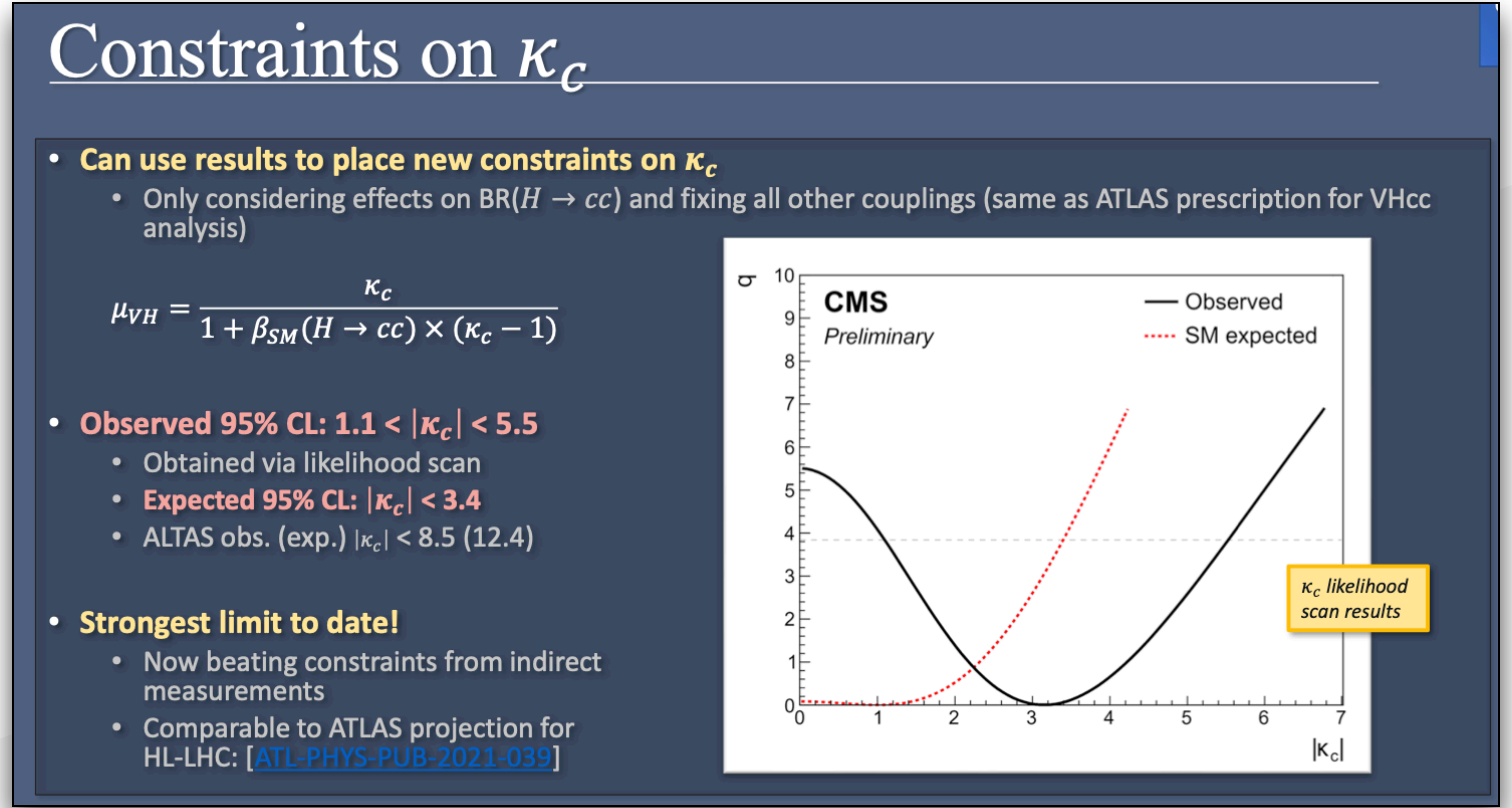


Cuts on the ME emission to avoid divergences

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}.

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.

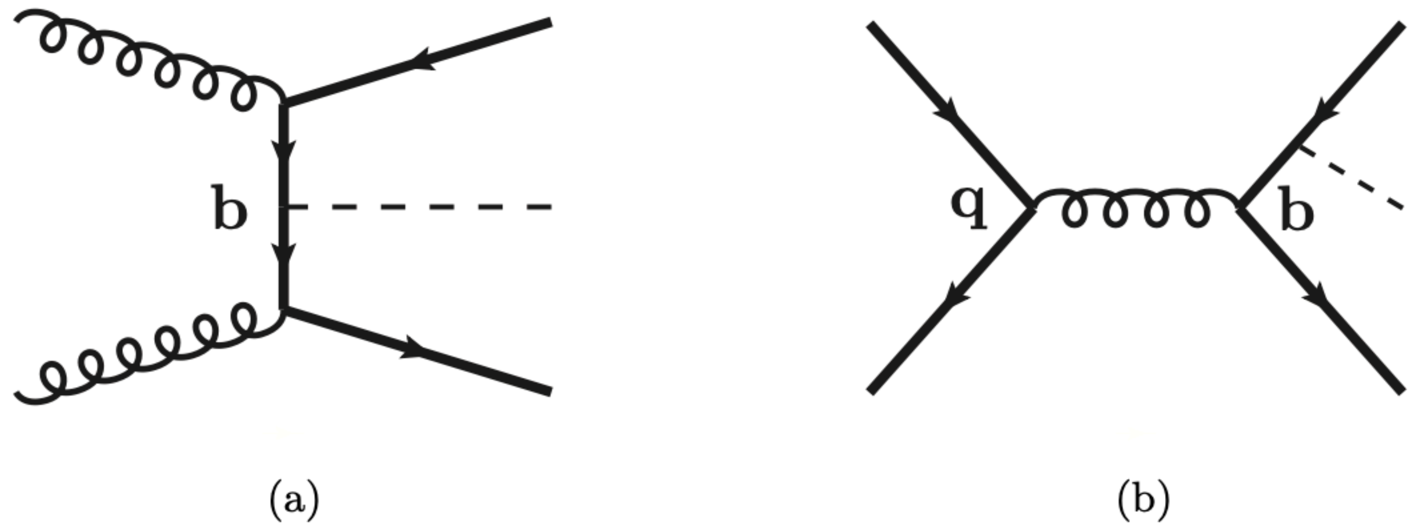


Figure 1: Sample of LO Feynman diagrams for $b\bar{b}H$ production in the four-flavour scheme, for the two relevant classes of partonic subprocesses: (a) $gg \to b\bar{b}H$; (b) $q\bar{q} \to b\bar{b}H$.

from: arXiv:1409.5301v2

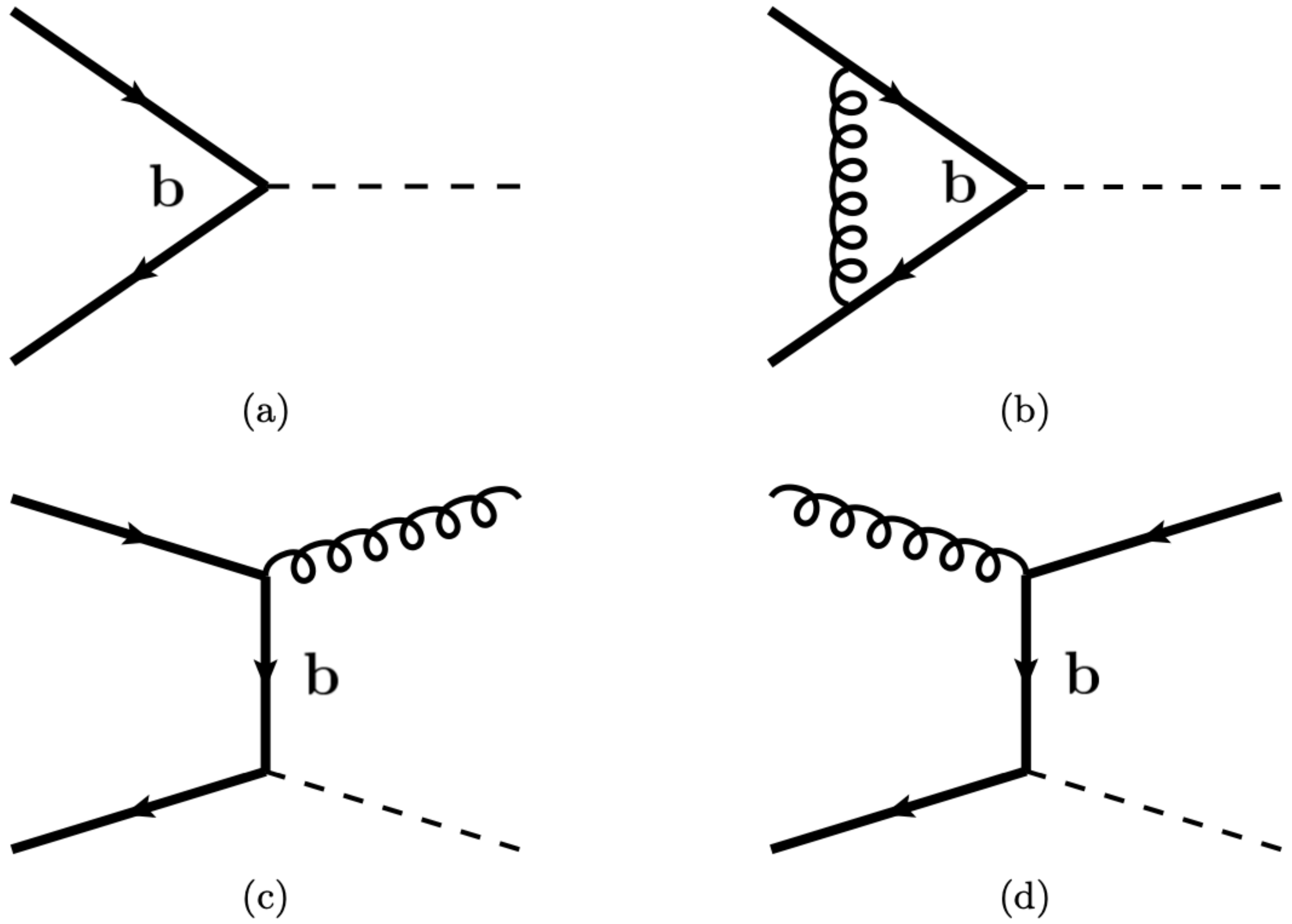


Figure 4: A sample of Feynman diagrams for $b\bar{b}H$ production in the five-flavour scheme: (a) LO; (b) one-loop; (c-d) real emission.