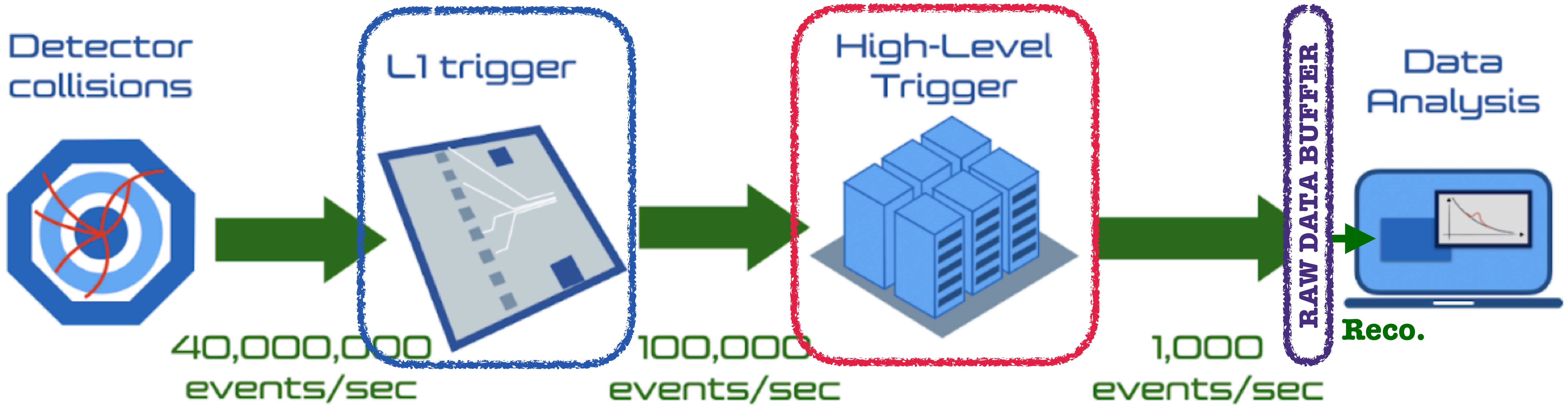
Data Scouting in CMS: catching the physics that escapes triggers

Beatriz Ribeiro Lopes, 24th March 2026

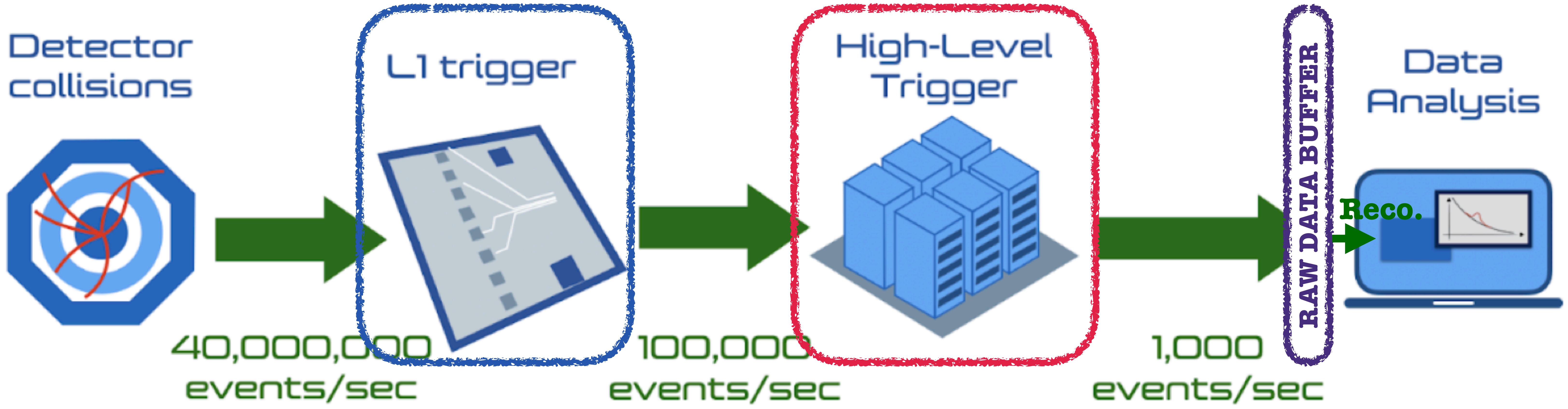
Ghent University
Experimental Particle
Physics and gravity

The trigger system in CMS and its limits



- L1: hardware processors, relies on information from calorimeter and muon detectors to select events up within a fixed latency of ~4 μs — the rate is a hard constraint fixed by the detector design
- HLT: farm of processors running a version of event reconstruction optimised for fast processing (~600ms per event) — less hard constraints on rate, can be expanded with new acquisitions and more modern machines
- DAQ output: limit of the size of temporary raw data storage buffer: hard limit on HLT output rate × event size

The trigger system in CMS and its limits



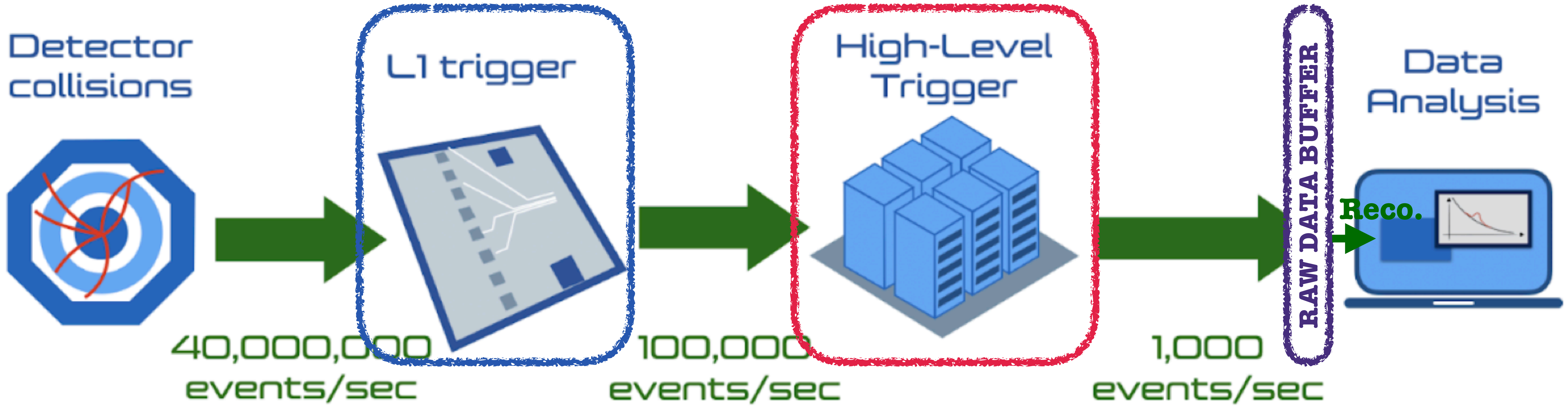
What is the problem?

Intriguing signal events with lower energy and momenta may inadvertently be discarded by imposing tight trigger requirements to keep the rate manageable (e.g. low mass BSM particles decaying into low-energy jets or leptons)

more modern machines

- DAQ output: limit of the size of temporary raw data storage buffer: hard limit on HLT output rate × event size

The trigger system in CMS and its limits



What is the problem?

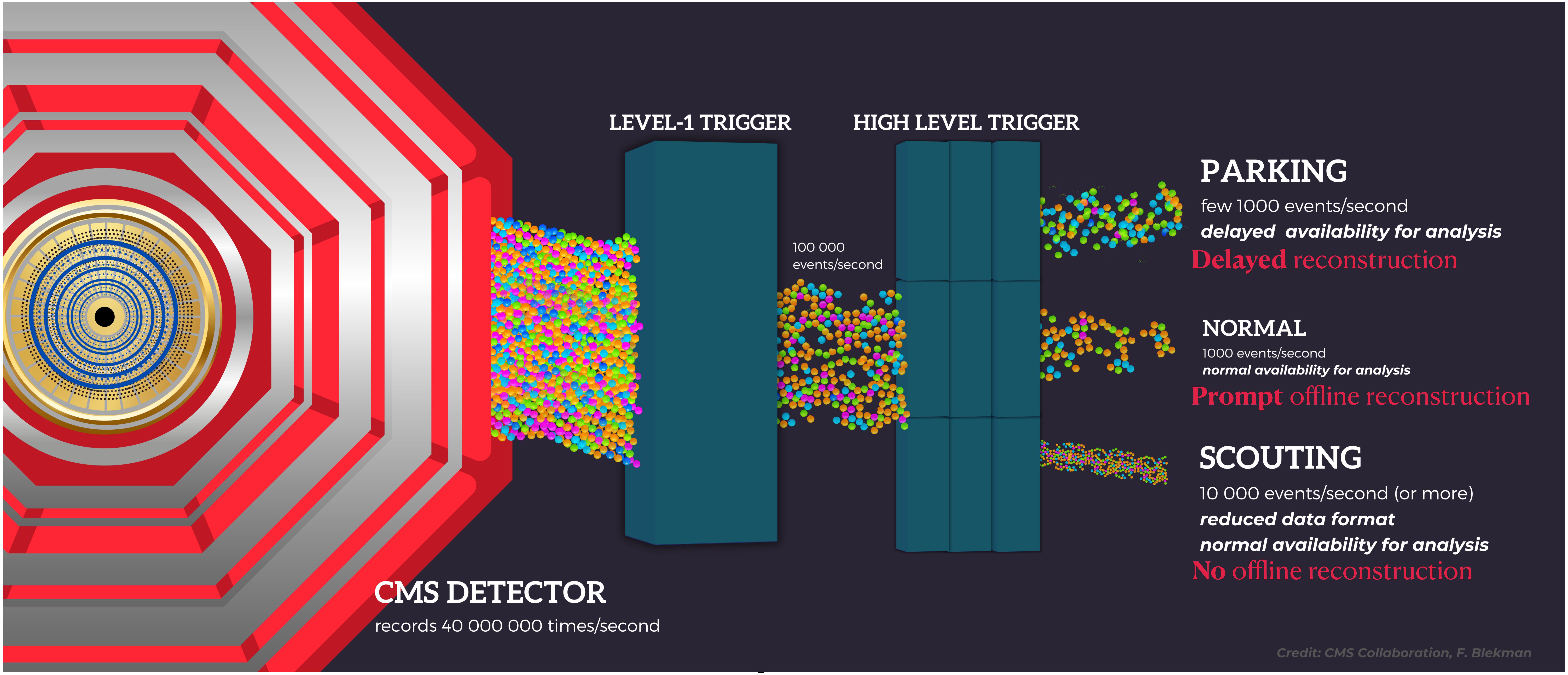
Intriguing signal events with lower energy and momenta may inadvertently be discarded by imposing tight trigger requirements to keep the rate manageable (e.g. low mass BSM particles decaying into low-energy jets or leptons)

more modern machines

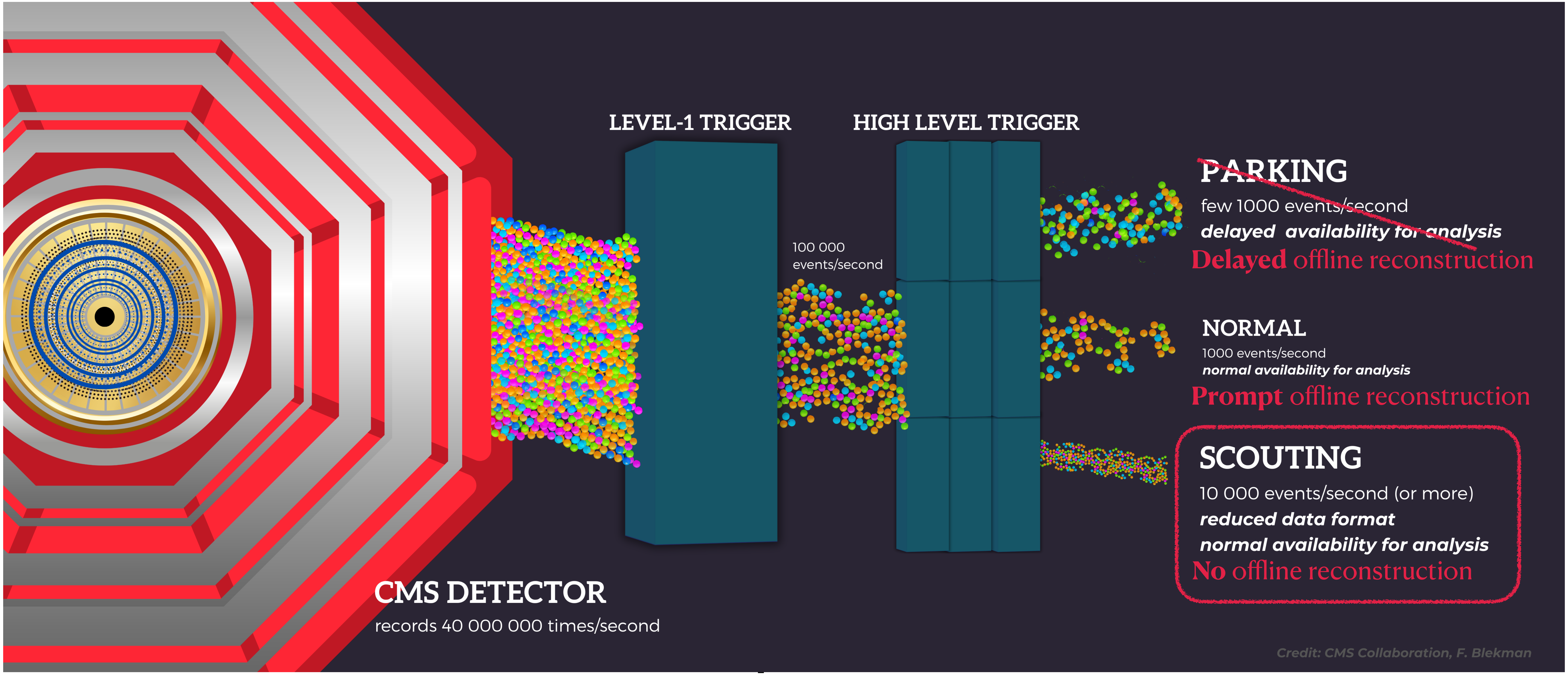
- DAQ output: limit of the size of temporary raw data storage buffer: hard limit on HLT output rate × event size

Idea: we can increase the HLT rate if we reduce event size

Data enhancing techniques: Scouting and Parking



Data enhancing techniques: Scouting and Parking



All discussed in detail in this really nice paper https://arxiv.org/pdf/2403.16134

Data enhancing techniques: Scouting and Parking

Scouting

Trades complete event information for higher event rates

- simplified version of standard HLT reconstruction
- no offline reconstruction
- Stores reduced event content
- keep the data bandwidth within limits

Parking

Stores large amounts of raw detector data with low trigger thresholds to be processed later

- No online event reconstruction
- Reconstruction happens when there is spare computational power (e.g. between Runs)

Data enhancing techniques: Scouting and Parking

Scouting

Trades complete event information for higher event rates

- simplified version of standard HLT reconstruction
- no offline reconstruction
- Stores reduced event content
- keep the data bandwidth within limits

Success of the programme depends on the quality of this reconstruction – can we really do a physics analysis with it?

Parking

Stores large amounts of raw detector data with low trigger thresholds to be processed later

- No online event reconstruction
- Reconstruction happens when there is spare computational power (e.g. between Runs)

What exactly is different in the reconstruction?

Tracks: same algorithm, reducing number of iterations and only running around objects of interest

Primary vertex: only tracks from pixel detector used to find vertex

Electrons and photons: similar algorithms to offline, but only ECAL-driven (and pixel hits for e), calibrations are "old"

Muons: simplified version of same algorithm online and offline, no association with PV, very small difference in performance

Jets: same algorithms online and offline, similar performance, but PF candidates saved are reduced set

B tagging: Only in Run 3 — similar ML algorithms but trained with online objects, performance slightly degraded from offline

Particle flow: simplified version of algorithm, no reco. of tracks from nuclear interactions in the tracker material — candidates are only started starting from Run 3

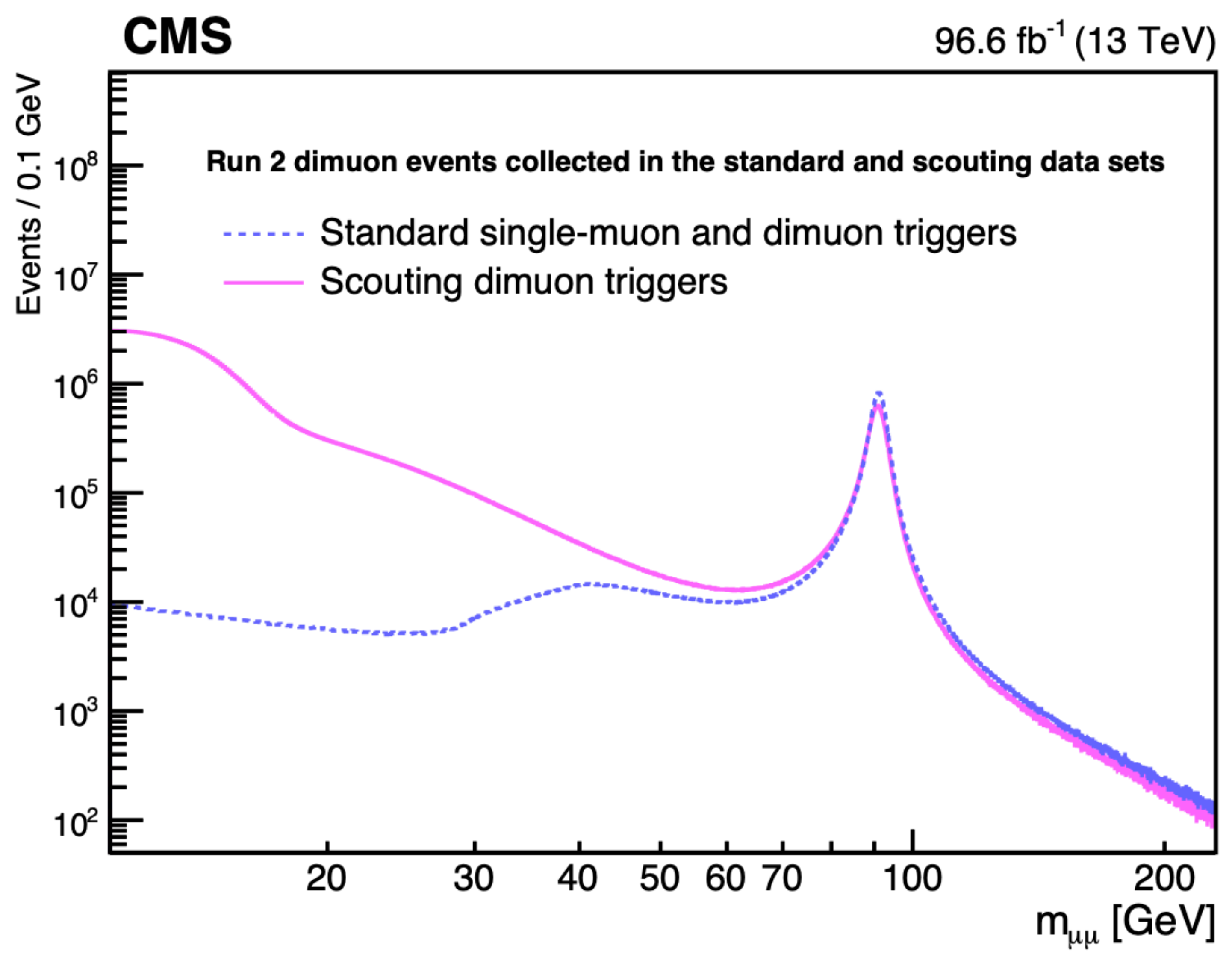
In short, only objects of interest are stored, and ID/reco performances are slightly degraded — but less and less in recent data-taking years!

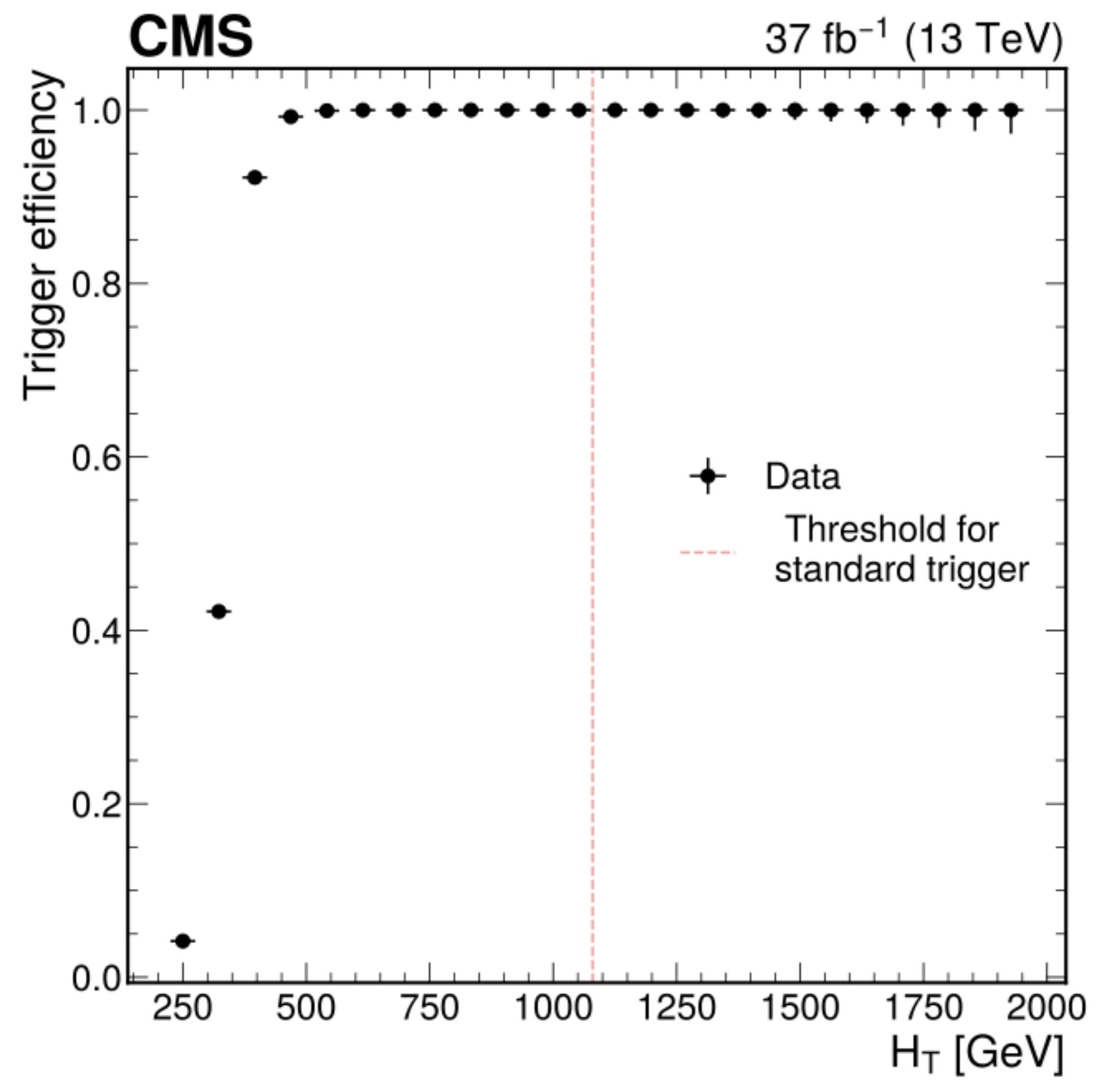
What physics can we do with Scouting?

- Search for low-mass dijet and dilepton resonances
- Study of rare B meson decays involving particles with momenta in the few GeV range
- Search for rare SM processes in fully hadronic final states

Scouting in Run 1 and 2:

- Dijet resonance searches was the clear focus
- Two scouting streams – one based on jets and other on muons
- No PF candidates stored
- Small dataset with scouting and offline collected to make comparisons of performance and validate

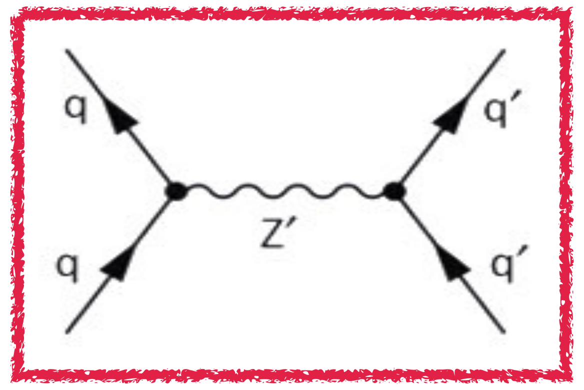


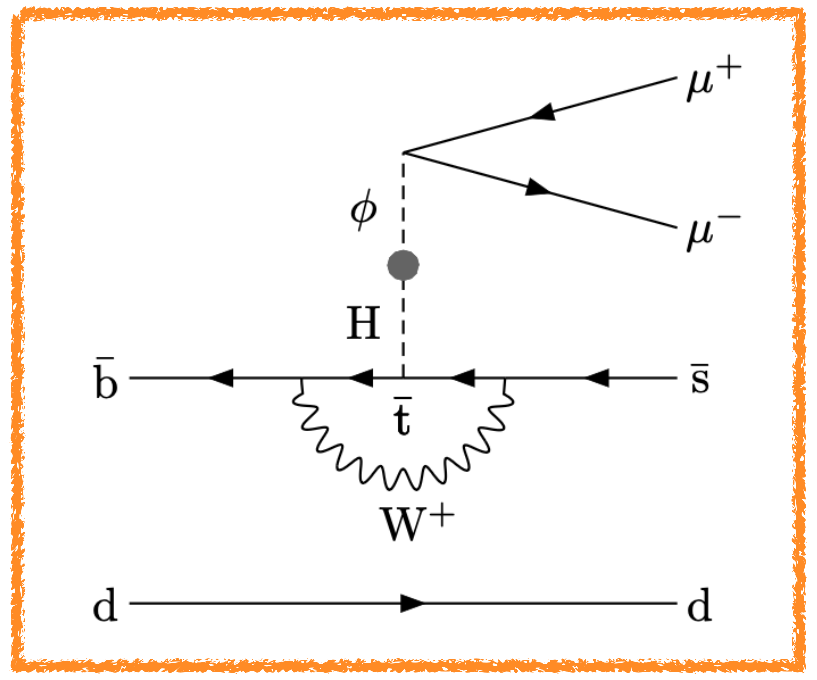


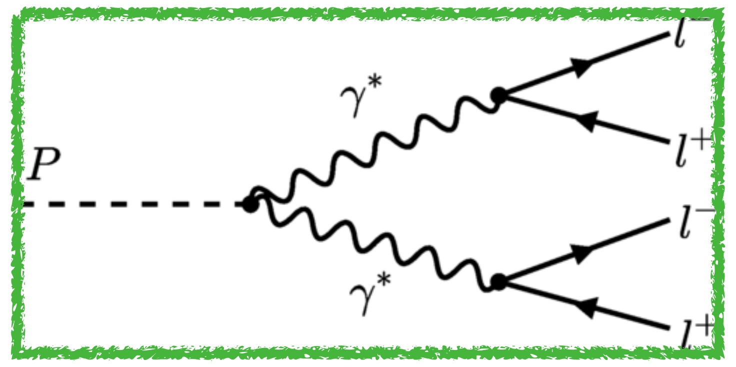
https://cms-hlt-scouting.docs.cern.ch/Publications/PhysicsResults/

Scouting in Run 1 and 2 — Physics results

- CMS-PAS-EXO-11-094: Search for narrow resonances using the Dijet Mass Spectrum in pp Collisions at sqrt(s) of 7 TeV (cds)
- CMS-EXO-14-005: Search for narrow resonances in dijet final states at sqrt(s) = 8 TeV with the novel CMS technique of data scouting (PhysRevLett.117.031802)
- CMS-EXO-16-032: Search for dijet resonances in proton-proton collisions at sqrt(s) = 13 !TeV and constraints on dark matter and other models (j.physletb.2017.02.012)
- CMS-EXO-16-056: Search for narrow and broad dijet resonances in proton-proton collisions at sqrt(s) = 13 TeV and constraints on dark matter mediators and other new particles (JHEP08(2018)130)
- CMS-EXO-17-030: Search for pair-produced three-jet resonances in proton-proton collisions at sqrt(s) = 13 TeV, (PhysRevD.99.012010)
- CMS-EXO-19-004: Search for dijet resonances using events with three jets in proton-proton collisions at sqrt(s) = 13 TeV (j.physletb.2020.135448)
- CMS-EXO-19-018: Search for a narrow resonance lighter than 200 GeV decaying to a pair of muons in proton-proton collisions at sqrt(s) = 13 TeV (PhysRevLett.124.131802)
- CMS-EXO-20-014: Search for long-lived particles decaying into muon pairs in proton-proton collisions at sqrt(s) = 13 TeV collected with a dedicated high-rate data stream (JHEP04(2022)062)
- CMS-EXO-21-005: Search for direct production of GeV-scale resonances decaying to a pair of muons (JHEP12(2023)070
- CMS-EXO-21-004: Searches for pair-produced multijet resonances using data scouting in proton-proton collisions at sqrt(s) = 13 TeV (PhysRevLett.133.201803)
- CMS-BPH-22-003: Observation of the rare decay of the eta meson to four muons (PhysRevLett.131.091903)

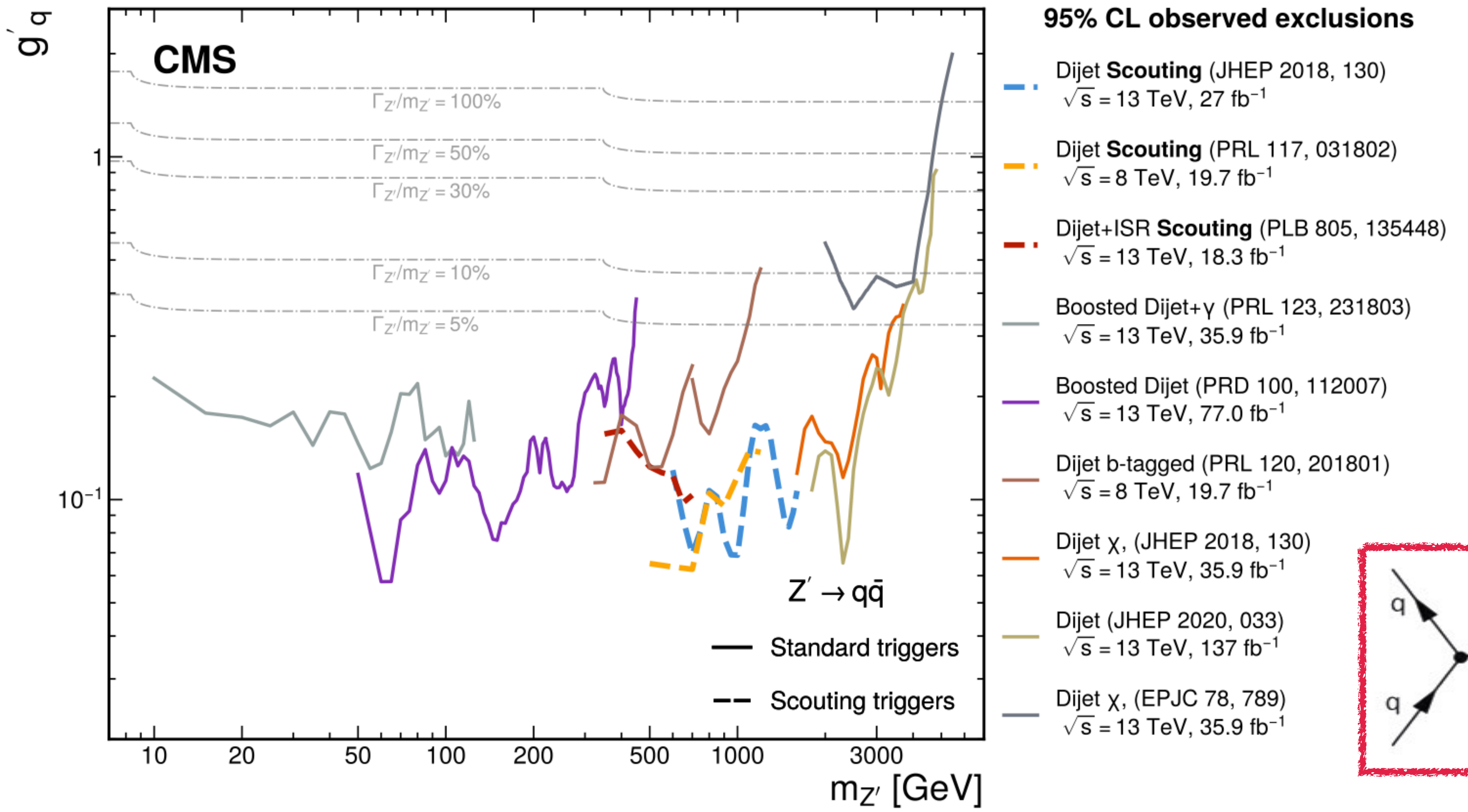


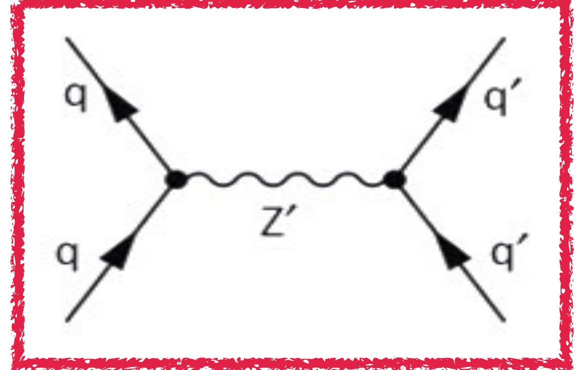


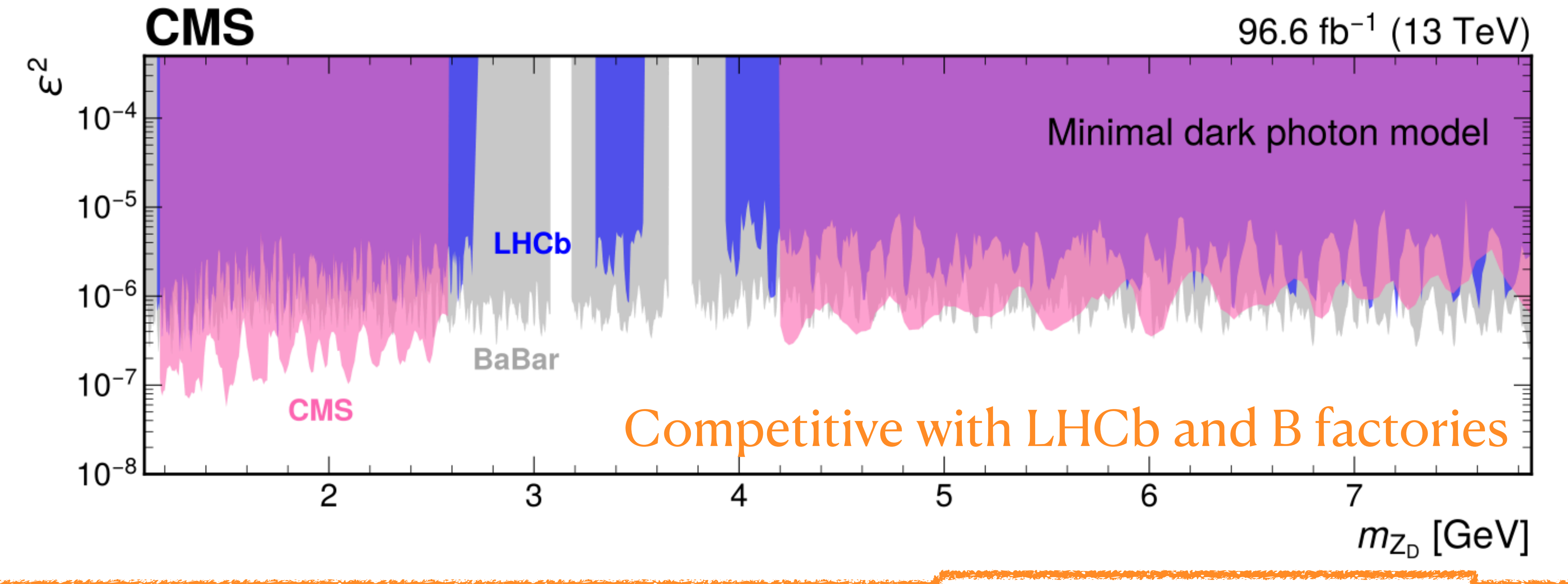


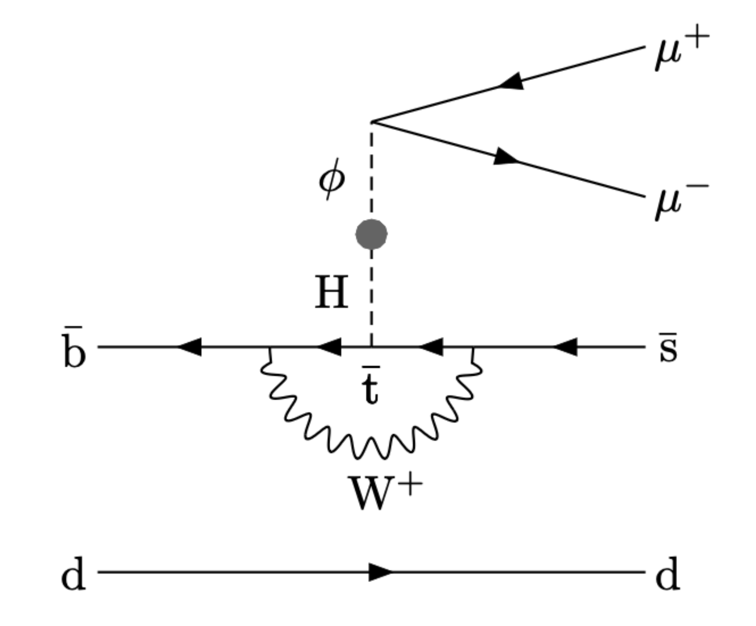
https://cms-hlt-scouting.docs.cern.ch/Publications/PhysicsResults/

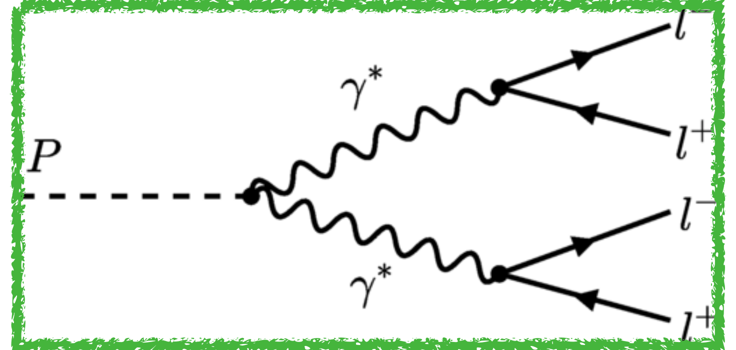
Scouting in Run 1 and 2 — Physics results

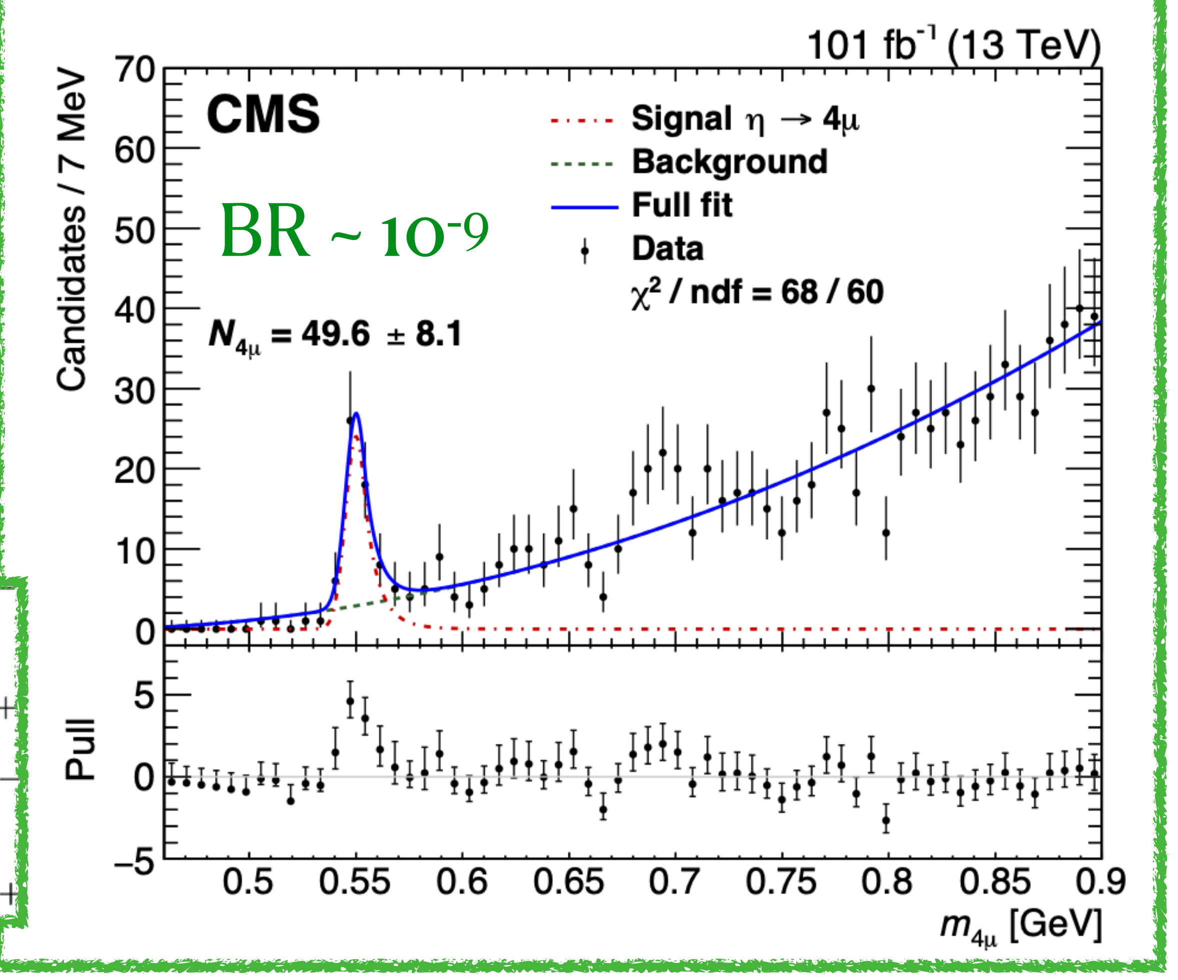




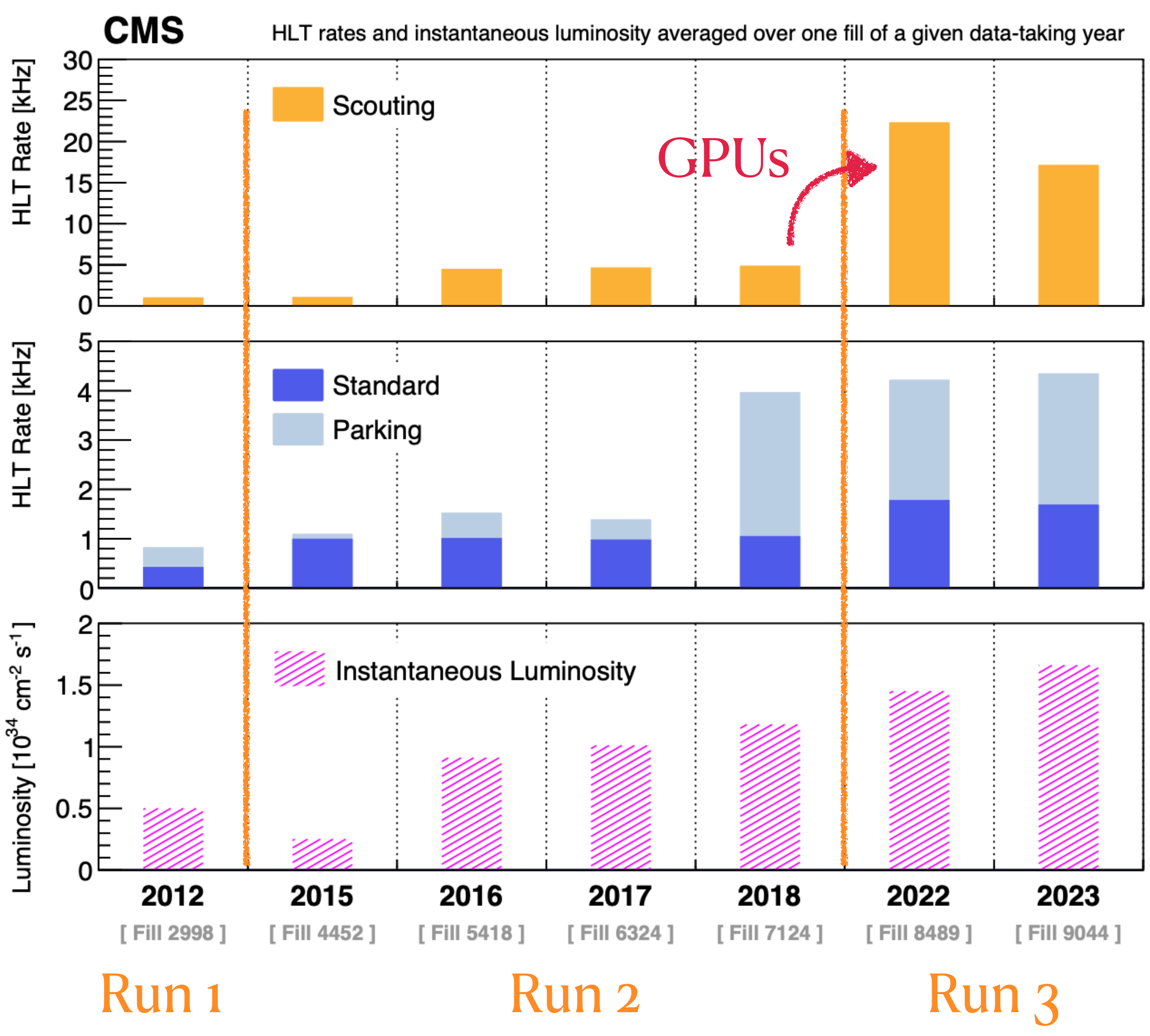








Scouting in Run 3

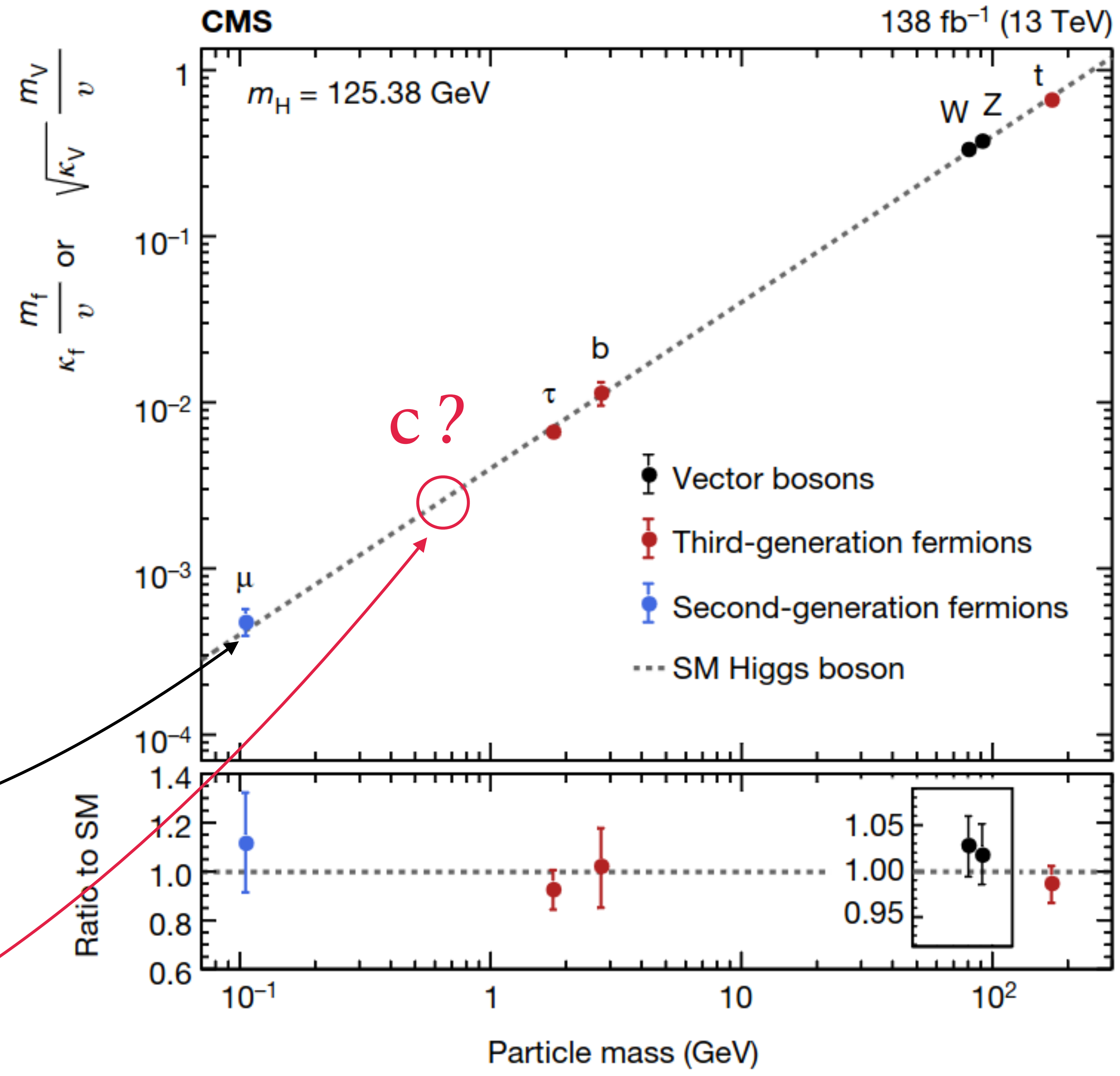


- HLT algorithms redesigned to harness capabilities of parallel architectures — offload HLT steps to GPUs
- Unified data-scouting stream where PF reconstruction is run for all events passing L1 requirements, individual PF candidates stored
- ParticleNet / ParticleTransformer for b/c/bb/cc tagging allows for many physics opportunities, especially Higgs->bb and Higgs->cc

Type	L1 threshold	HLT threshold (**2023**)
e/γ	1 e/γ, $p_T > 30$ GeV, $\|\eta\| < 2.1$	1 SC (loose), $p_T > 30$ GeV
	2 e/γ, $p_T > 18/12$ GeV, $\|\eta\| < 1.5$	2 SC (loose), $p_T > 12$ GeV
μ	2μ, $p_T > 15/7$ GeV	2μ, $p_T > 3$ GeV
	2μ, OS, $p_T > 4.5$ GeV, $\|\eta\| < 2$, $m_{\mu\mu} > 7$ GeV	
	2μ, OS, $p_T > 4$ GeV, $\|\eta\| < 2.5$, $\Delta R < 1.2$	
	2μ, OS, $p_T > 0$ GeV, $\|\eta\| < 1.5$, $\Delta R < 1.4$ (**2023**)	
	3μ, $p_T > 5/3/3$ GeV	
Jets/H_T	$H_T > 280$ (**2023**), 360 (**2022**) GeV	
	1 jet, $p_T > 180$ GeV	
	2 jets, $p_T > 30$ GeV, $\|\eta\| < 2.5$, $\Delta\eta < 1.5$, $m_{jj} > 250$ (**2023**), 300 (**2022**) GeV	

A special example: the Higgs-charm coupling

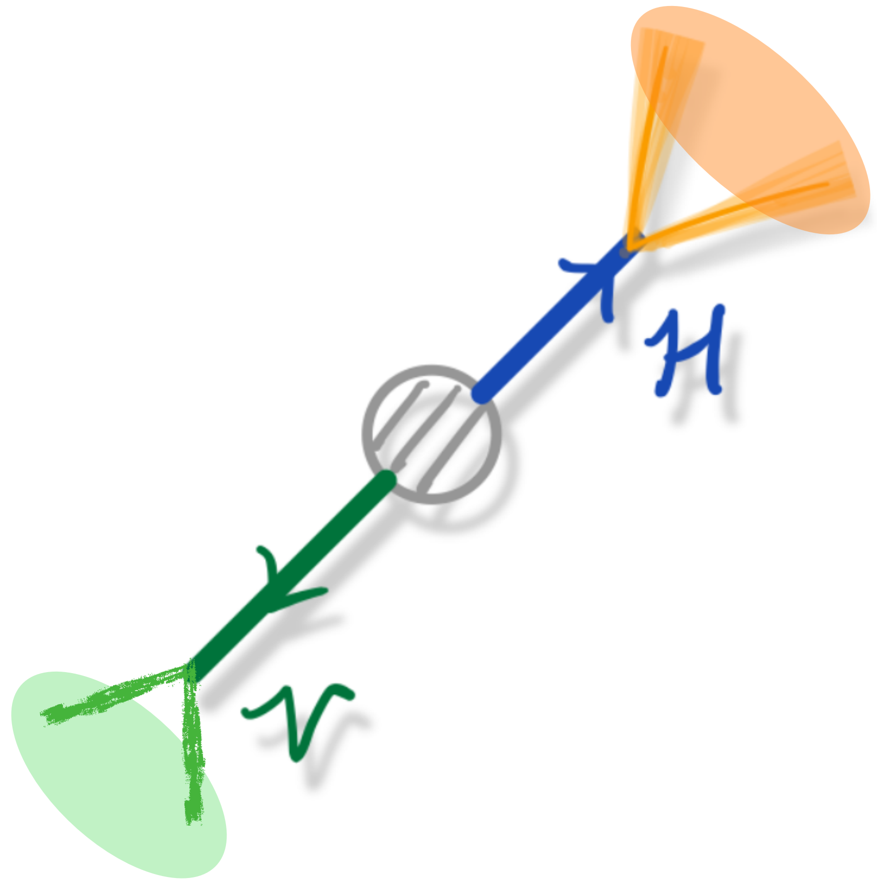
- Measurement of Higgs boson properties among primary LHC goals
  - Self-coupling, couplings to vector bosons and fermions
- Precise measurements possible with <u>3rd generation fermions</u>, and getting better and better with <u>2nd generation</u>
- With the large dataset of Run 3, we can get closer!
- **Simultaneous H→bb/cc** measurement
  - Probe (anti)correlated BSM effects on H couplings to c and b



H→μμ (BR_{SM}~2.2×10^{-4})
evidence at 3.0σ
(2.5σ exp.)

H→cc (BR_{SM}~2.9×10^{-2})
$\mu_{VH(cc)+ttH(cc)}$ < 9.3 (exp 5.6) at 95% CL

VHcc in the hadronic channel



- VH production is a golden channel to look for charm decays, because we can trigger and tag leptonic decays of the Z/W boson
- Decays of the Z/W to hadrons give much larger BR
- Merged-jet topology: Higgs decay and V decays both reconstructed as single large-radius (R=0.8) jets

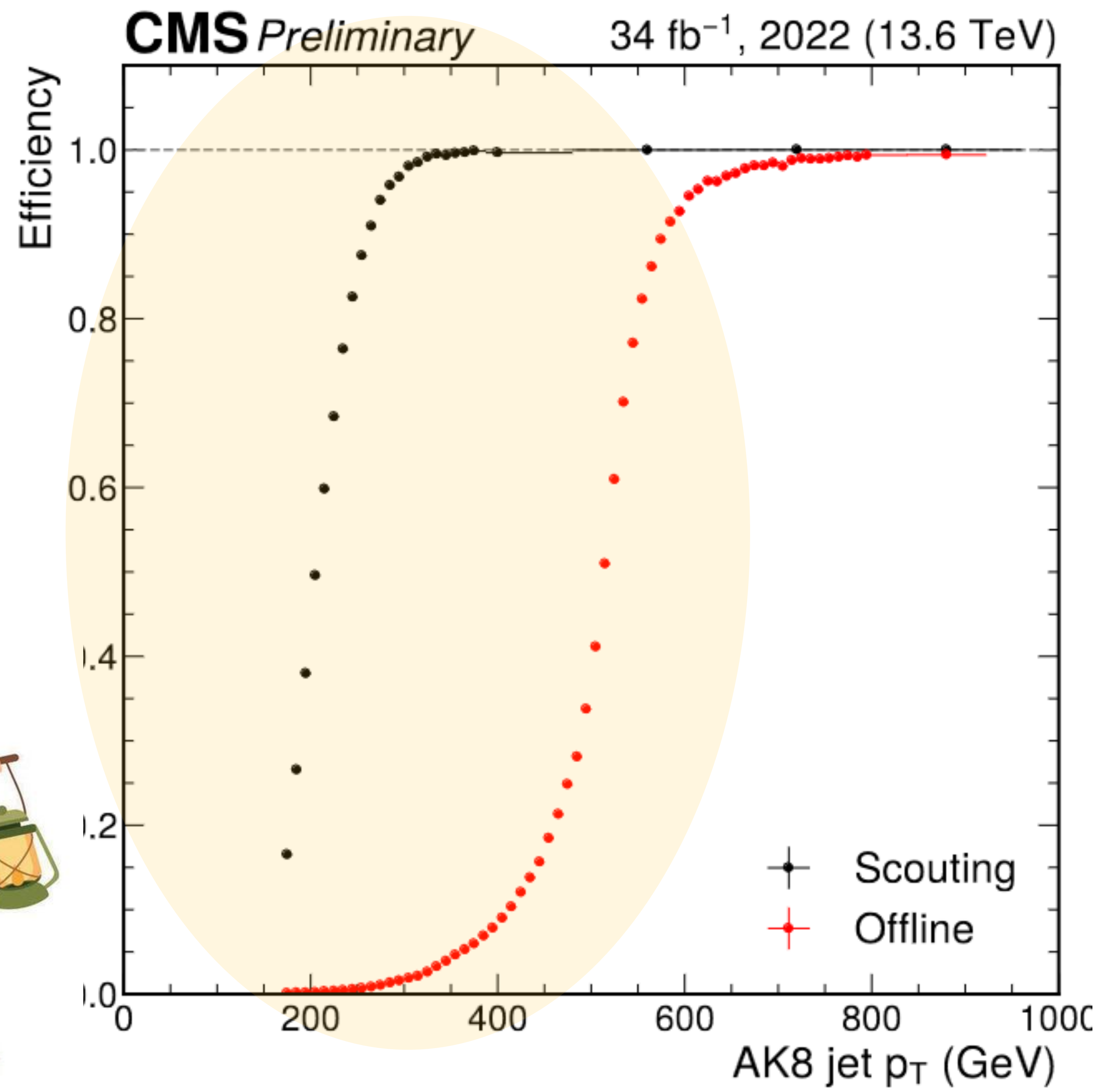
- Very challenging final state: no leptons, overwhelming background from QCD multi jet processes
- How to trigger on this? Standard HLT Jet HT triggers start at very large (1050 GeV) values, the signal efficiency is low

Solution: HLT Scouting data!

- ✦ Much lower HT requirements - but also larger QCD background...
- ✦ Slightly degraded performance for object identification when compared to offline

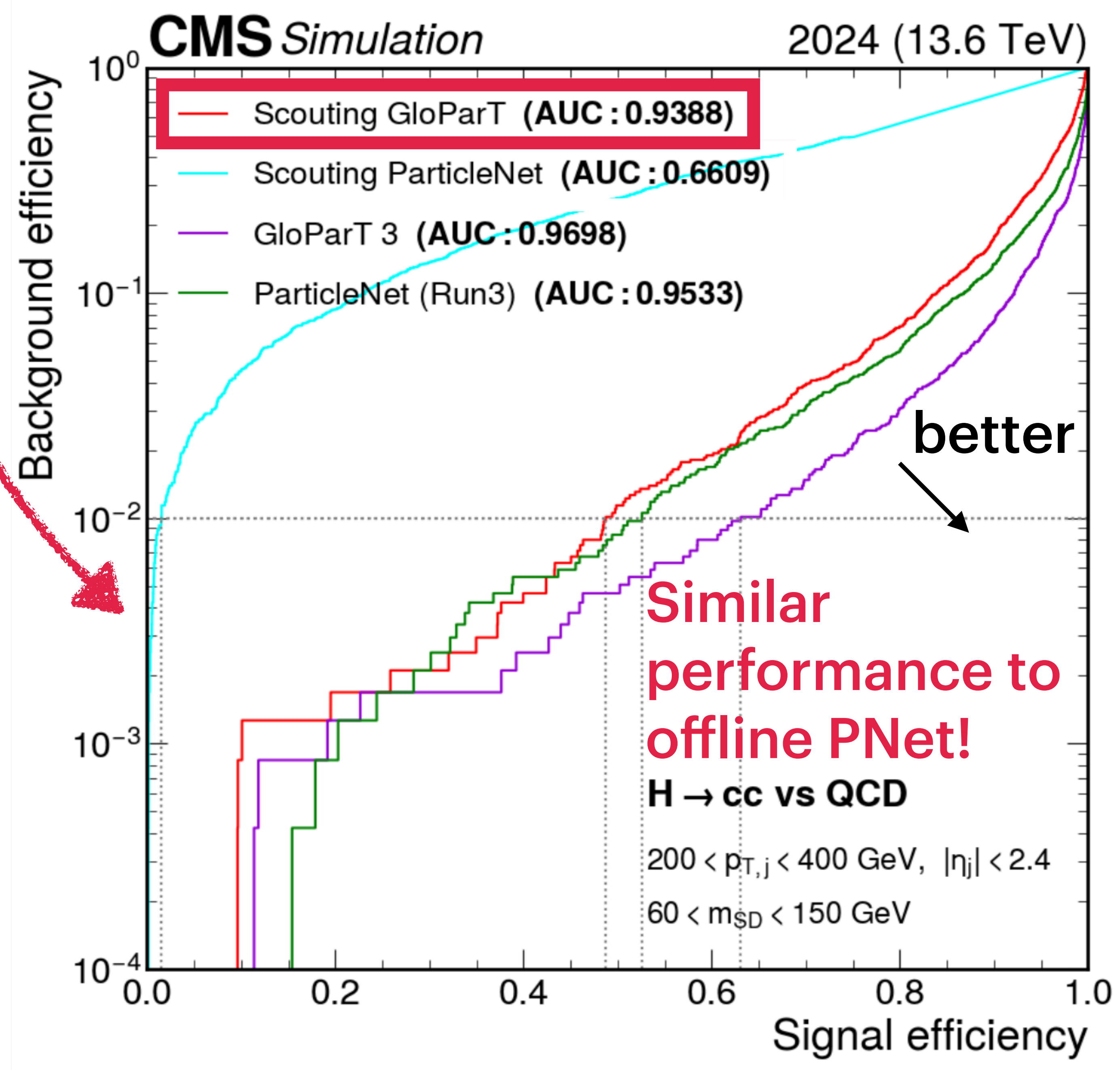
HLT scouting for VHcc hadronic

- Size of 2024 dataset comparable to full Run 2 (~108/fb)
- HLT scouting significantly increases signal efficiency in the fully hadronic channel (HT1050 → HT~400)
- GloParT tagger trained for scouting objects in 2024 data, with very good performance

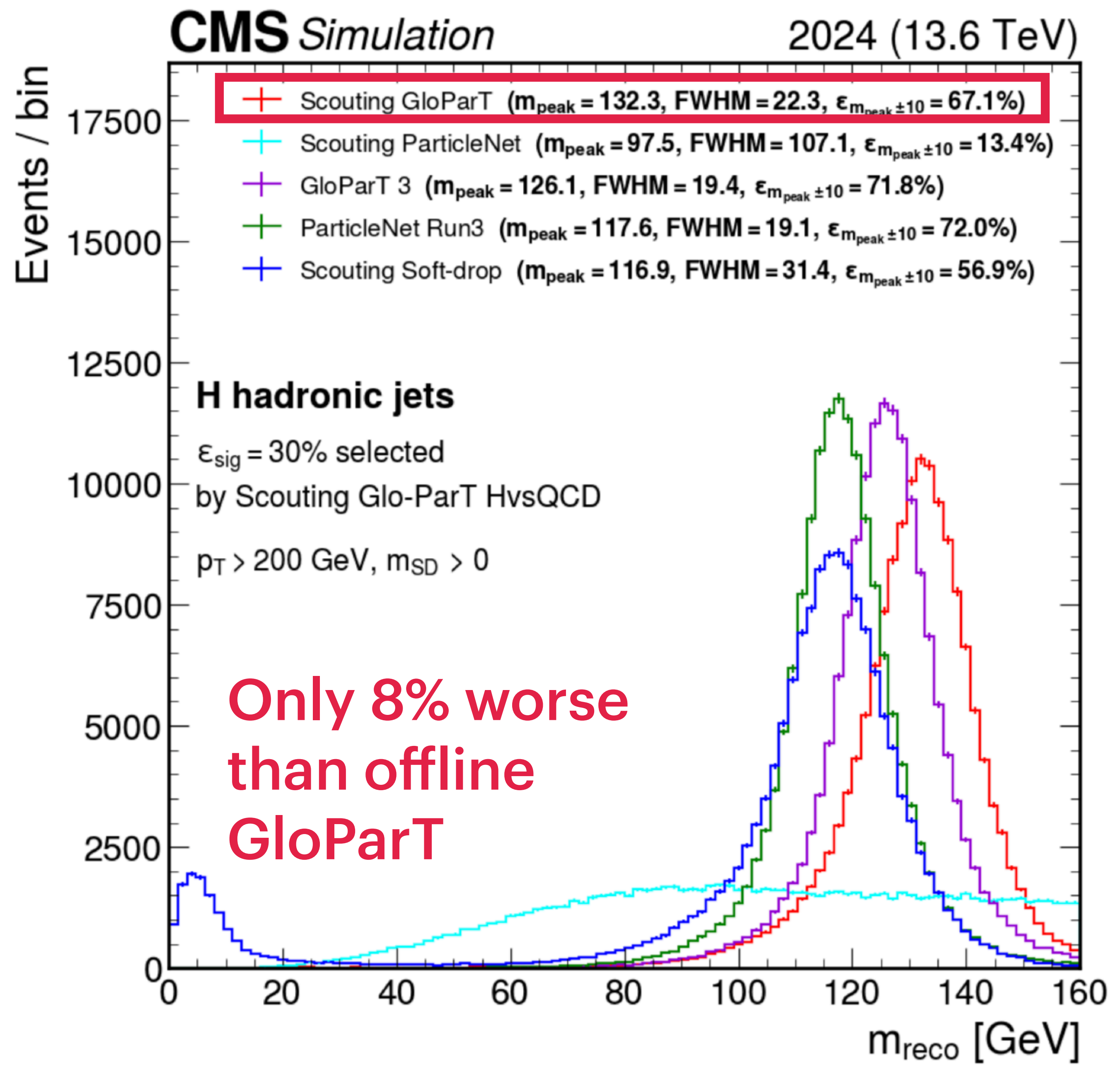


Event selection efficiency vs AK8 p_T in 2022

TAGGING

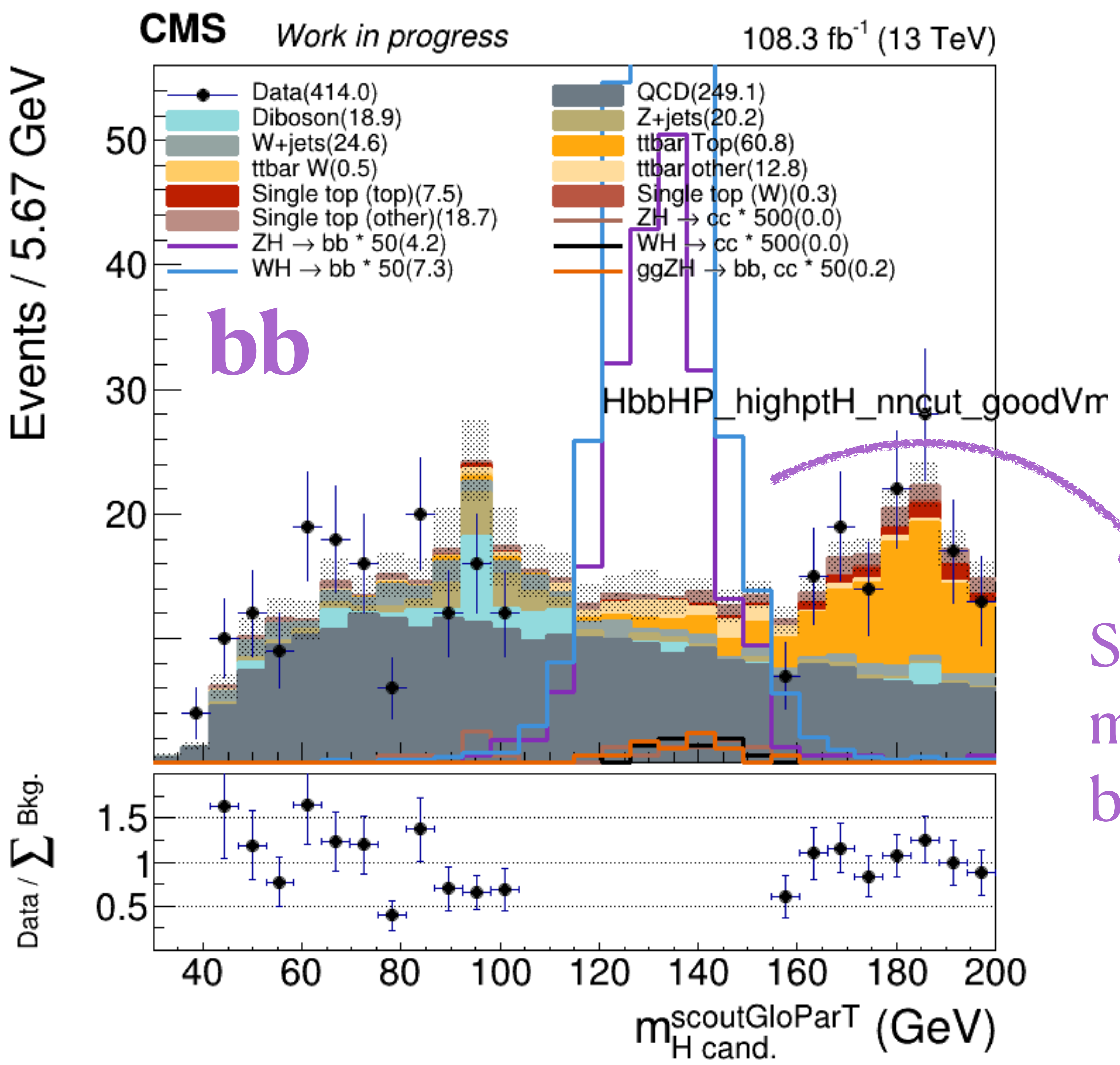


MASS REGRESSION

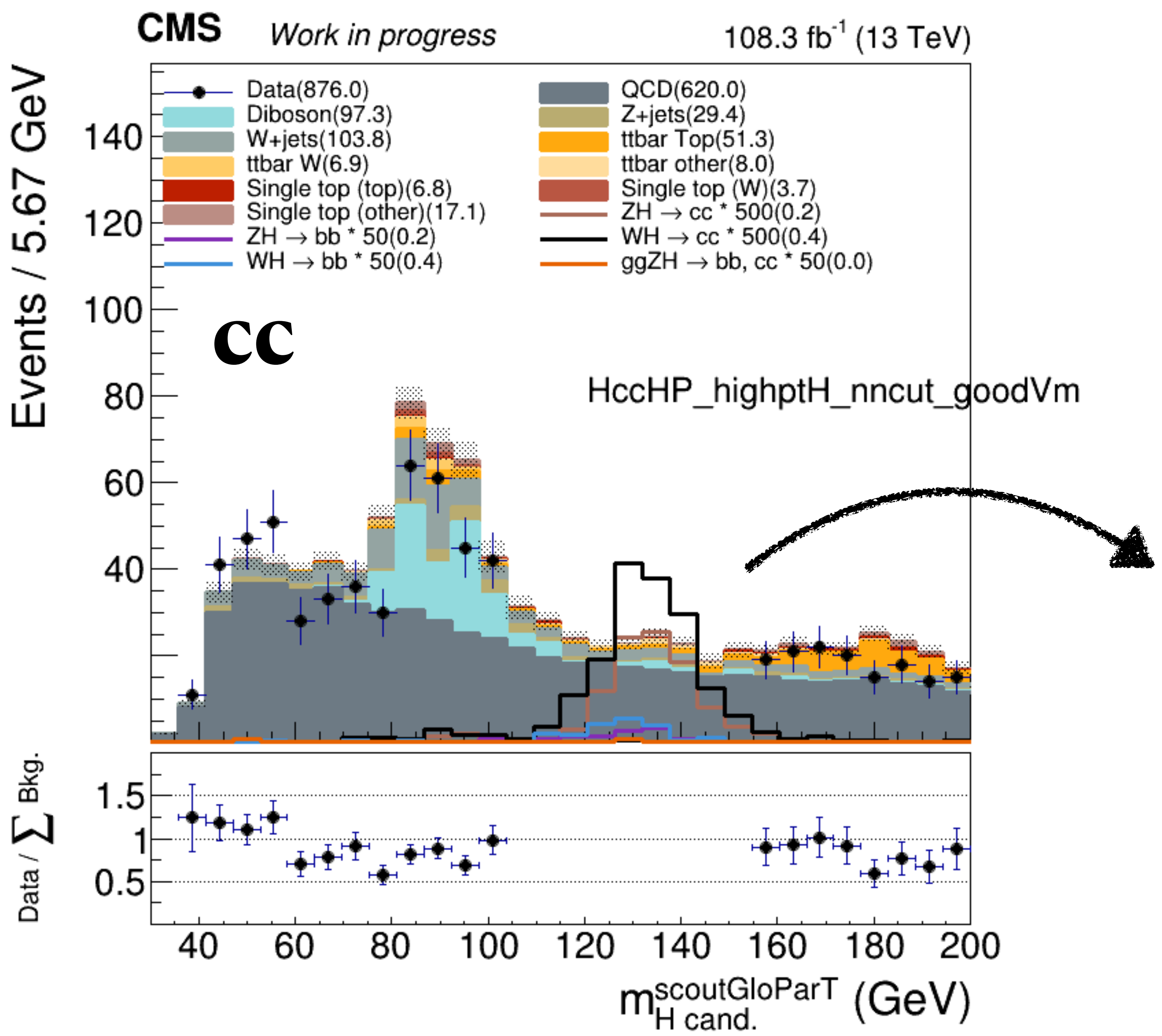


How good is it though?

- Can't really show results yet — work in progress!
- Sensitivity estimates for H->cc: upper limit ~30 x SM
  - best result by far in the fully hadronic channel, not competitive with leptonic but can contribute in a combination



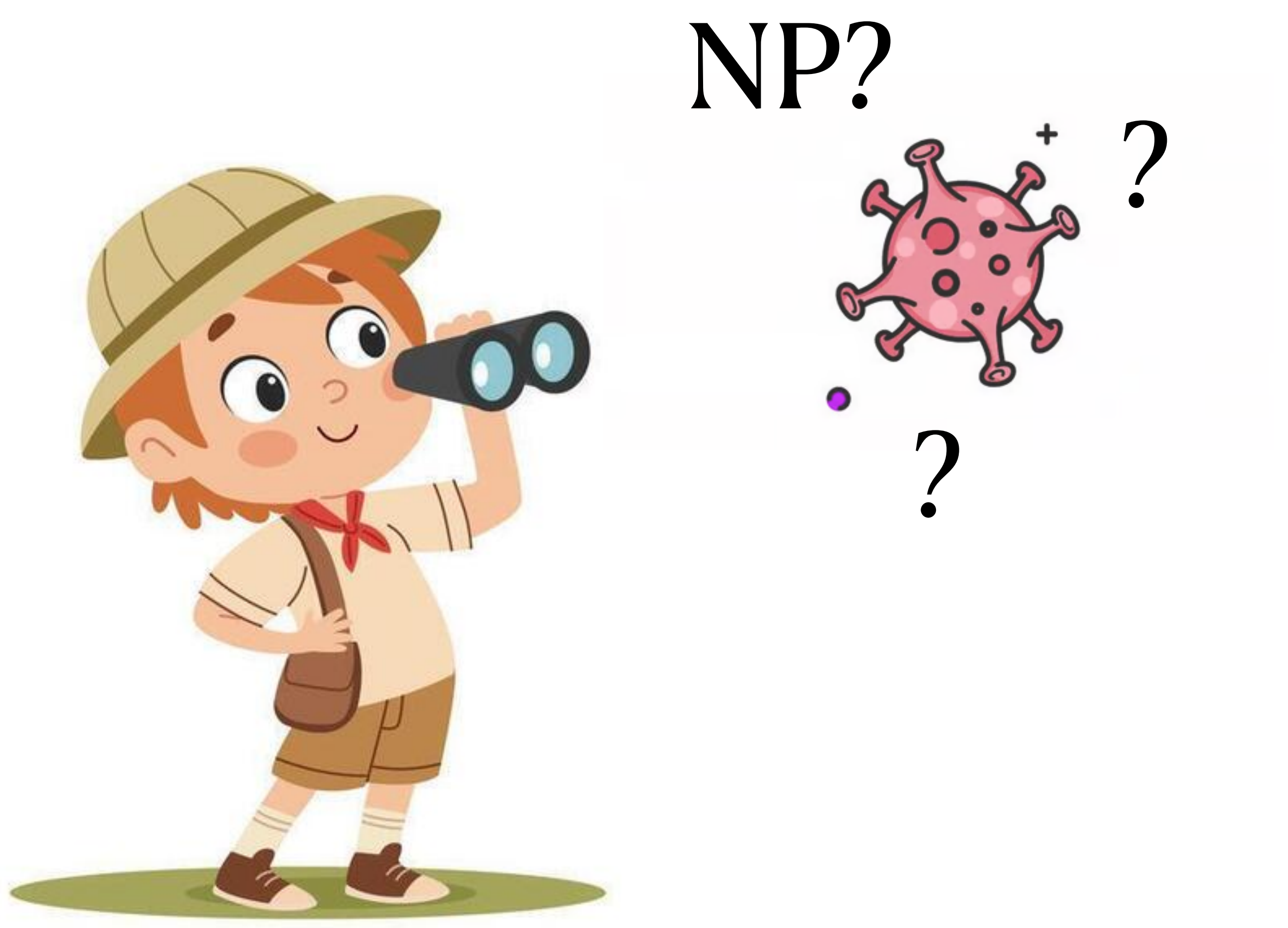
Signal multiplied by 50



Signal multiplied by 500

This is just the beginning...

- HLT Scouting has been used since Run 1 but significantly expanded and improved for Run 3
- 11 physics results with Run 1 and 2 data — very good sensitivity for resonance searches!
- Introduction of tagging algorithms in Run 3 is a game changer — several efforts ongoing to use this
- Higgs physics was my personal example, but many new physics searches in hadronic final states are also possible



Questions?

Beatriz Ribeiro Lopes, 24th March 2026