Potential of di-Higgs observation via a calibratable jet-free *HH*(4*b*) search framework

arXiv:2508.15048

Tianyi Yang (杨天一) *and* **Congqiao Li (李聪乔)**

(Peking University)

TeV 物理前沿专题研讨会暨第 31 届

LHC Mini-Workshop · Qingdao

10 October, 2025

Key points of this talk

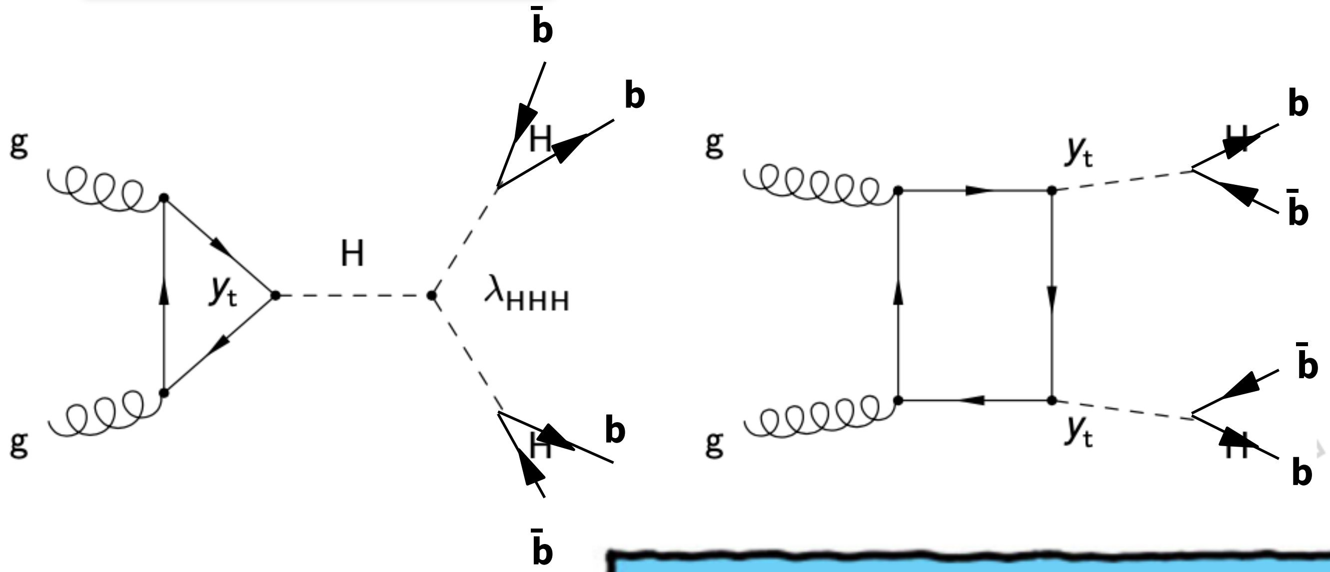
TL;DR

- ➔ We find that CMS/ATLAS can achieve >5× sensitivity improvement for di-Higgs in the *HH*→4*b* final state, with potential to observe HH even before the HL-LHC
- ➔ This gain is driven entirely by AI engineering
  - ❖ **"scaling up" improves performance** (larger models, larger datasets) and training **directly on low-level objects** (going beyond modularized objects like jets)
- ➔ We validated the method's reliability
  - ❖ successfully reproduced two prior CMS results
- ➔ We aim to deliver this within the next three years. If successful, it will be transformative for the LHC, reshaping expectations for what physics measurements are achievable

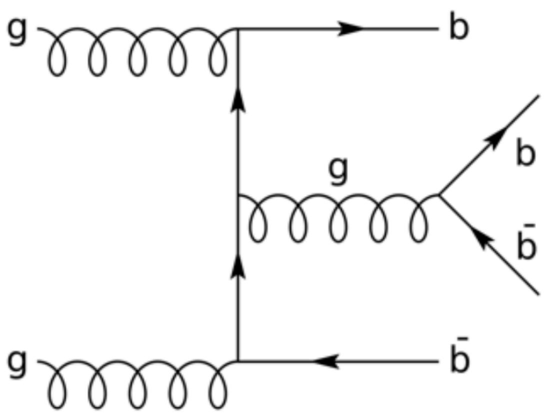
based on arXiv:2508.15048

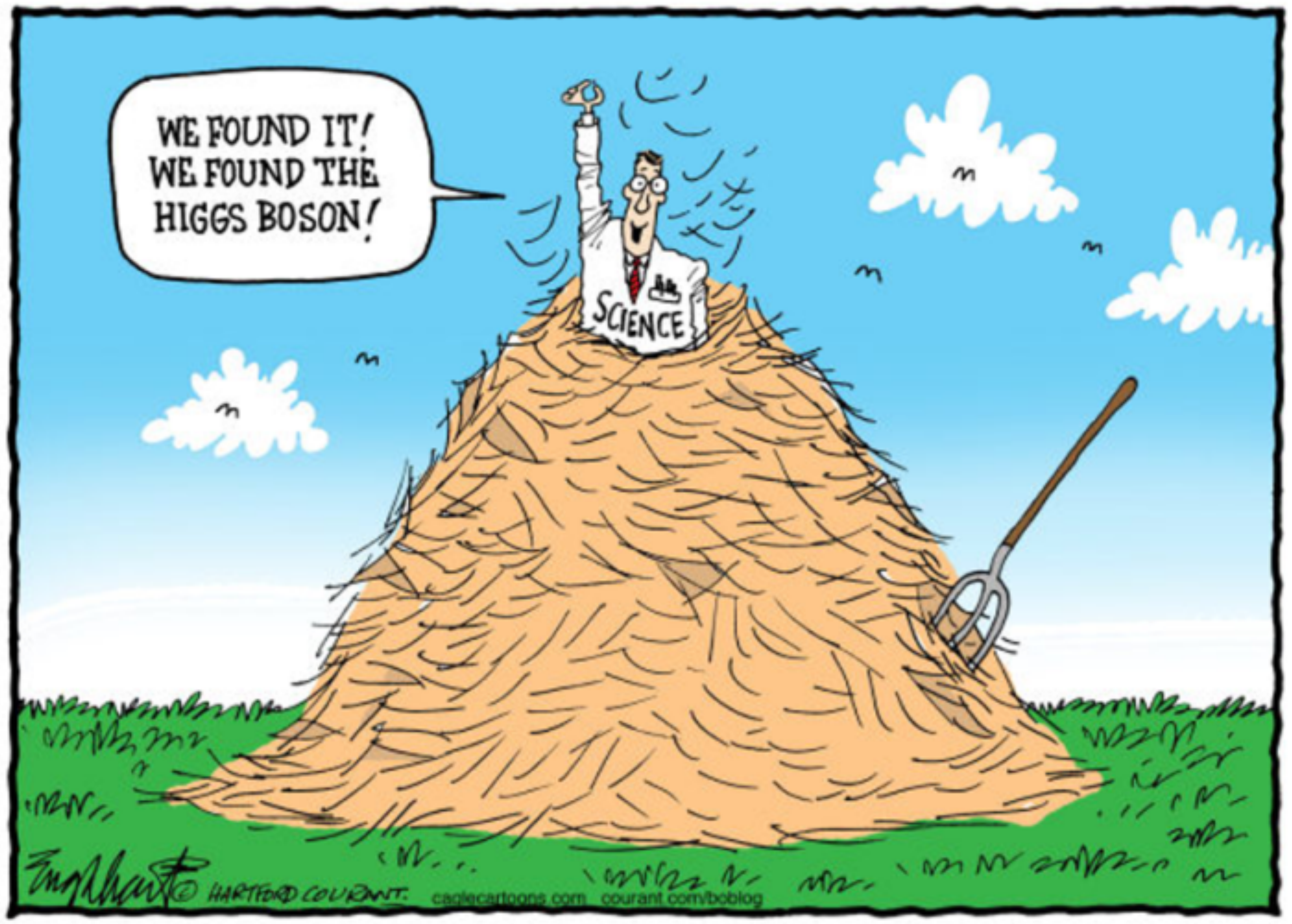
Illustrative view of the *HH*(4*b*) search

_HH_(4*b*) signal



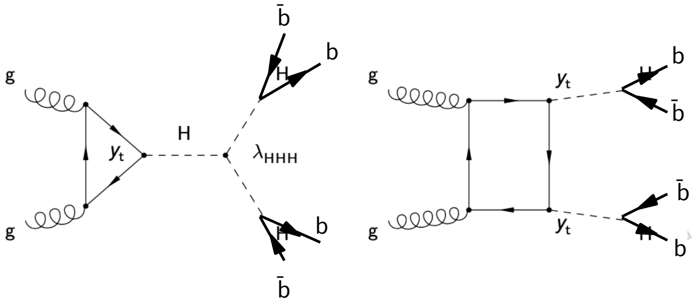
QCD background

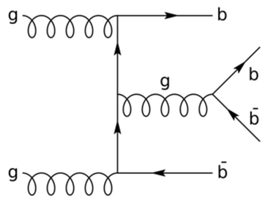




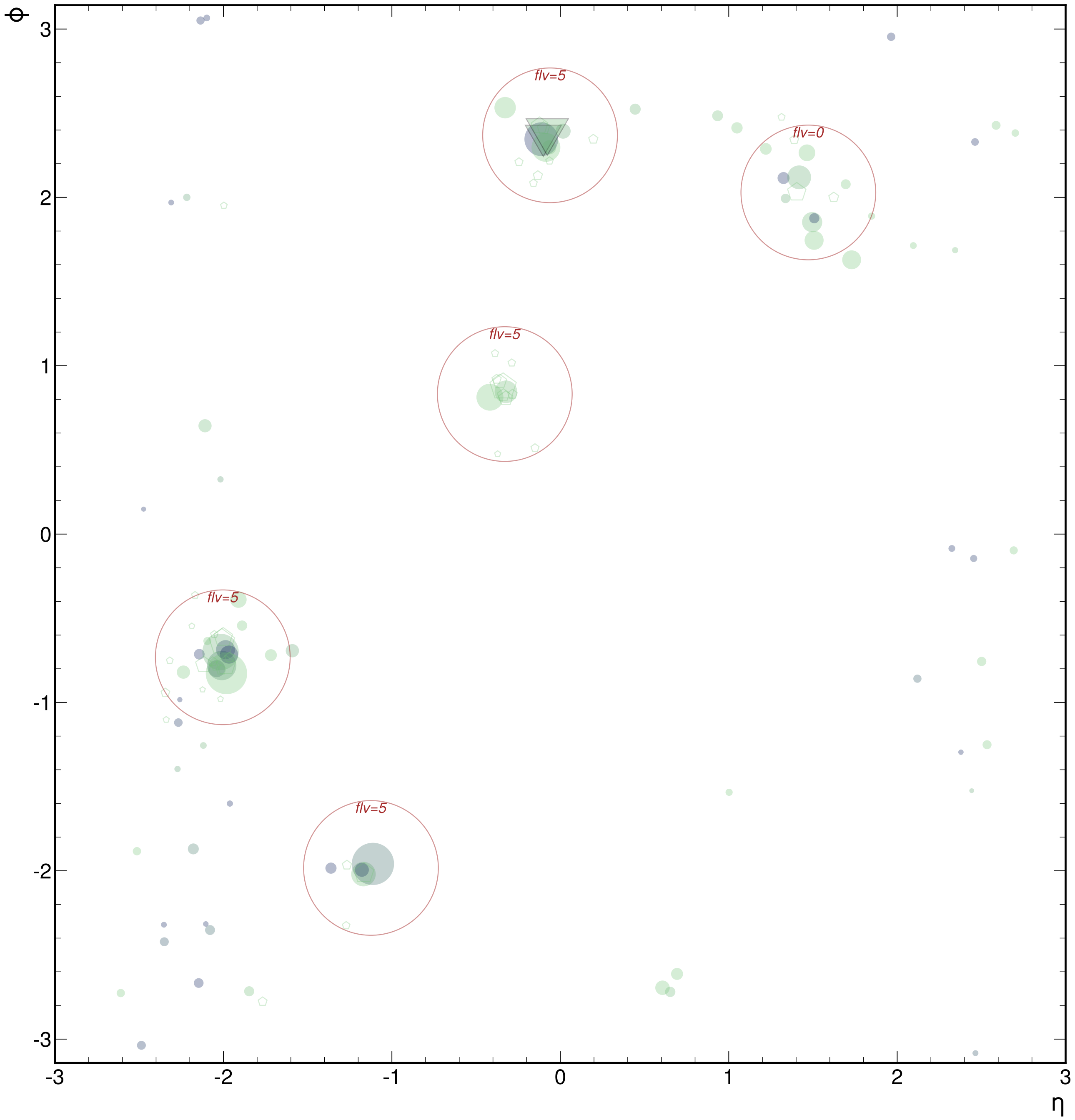
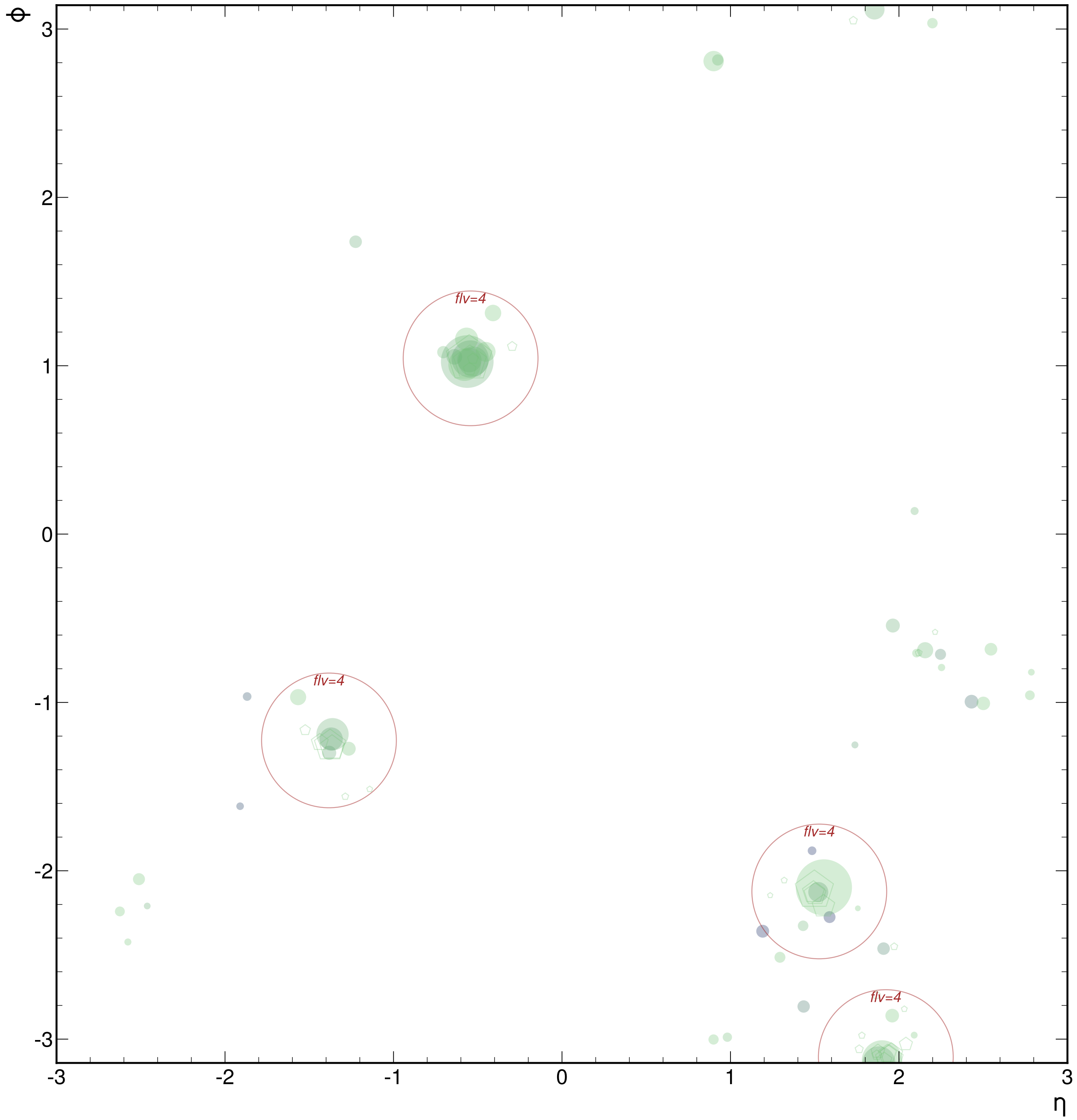
Illustrative view of the *HH*(4*b*) search

***HH*(4*b*) signal**



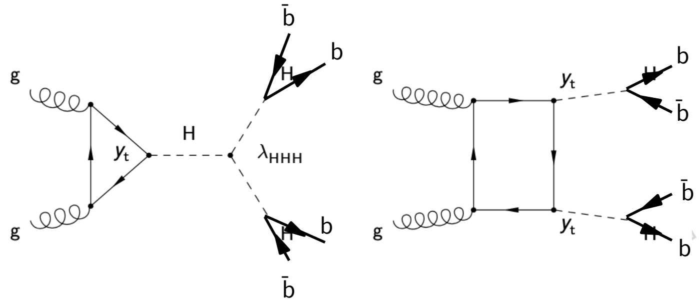


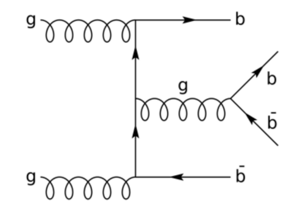
QCD multijet bkg.

Illustrative view of the *HH*(4*b*) search

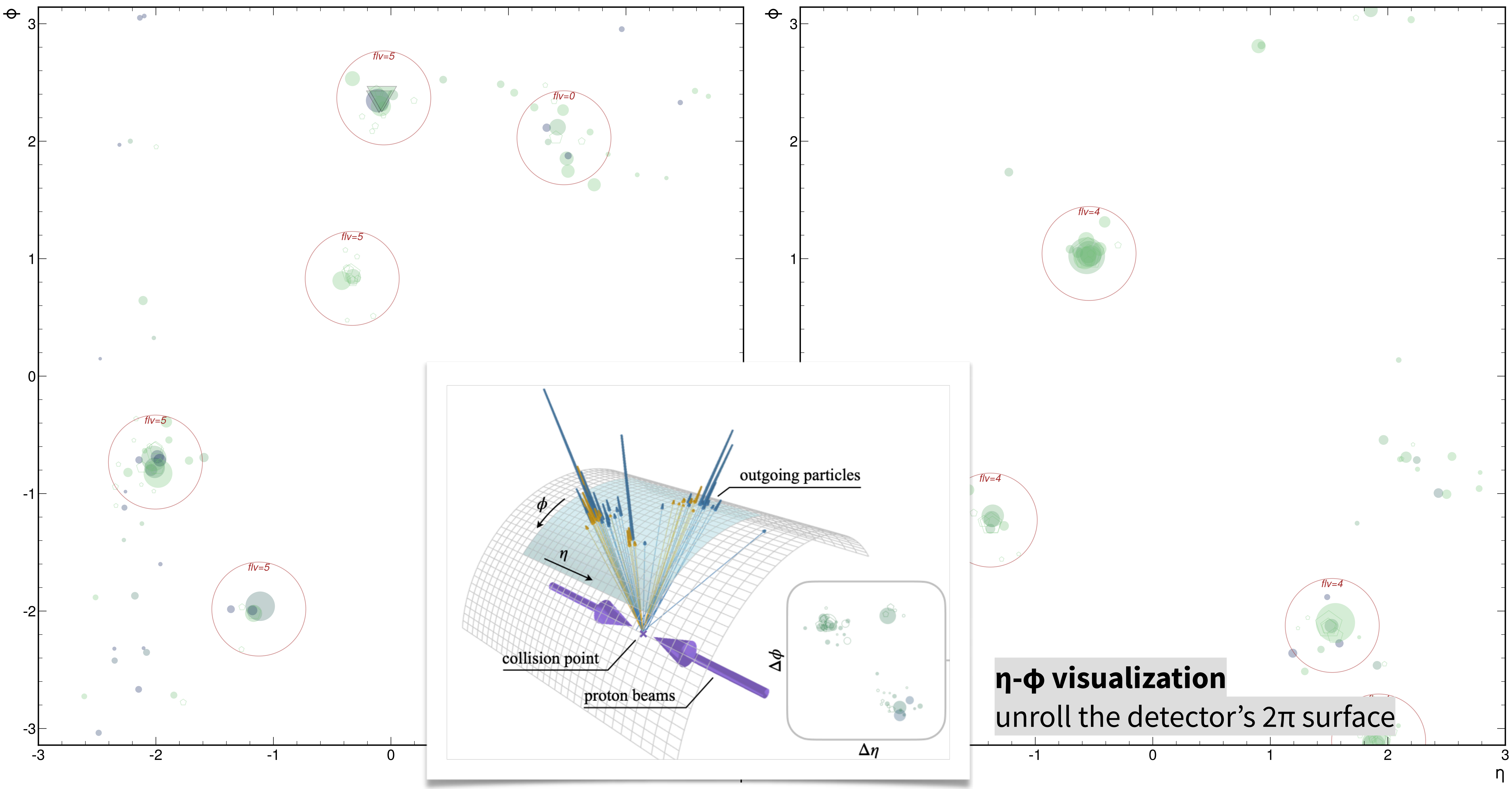
***HH*(4*b*) signal**





QCD multijet bkg.

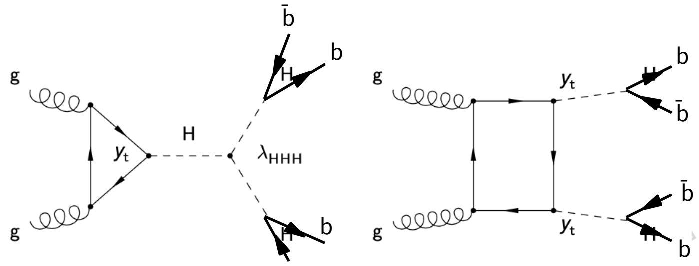
…



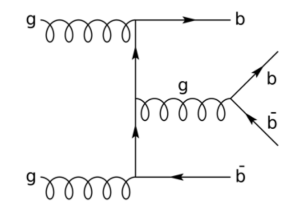
η-φ visualization
unroll the detector's 2π surface

Illustrative view of the *HH*(4*b*) search

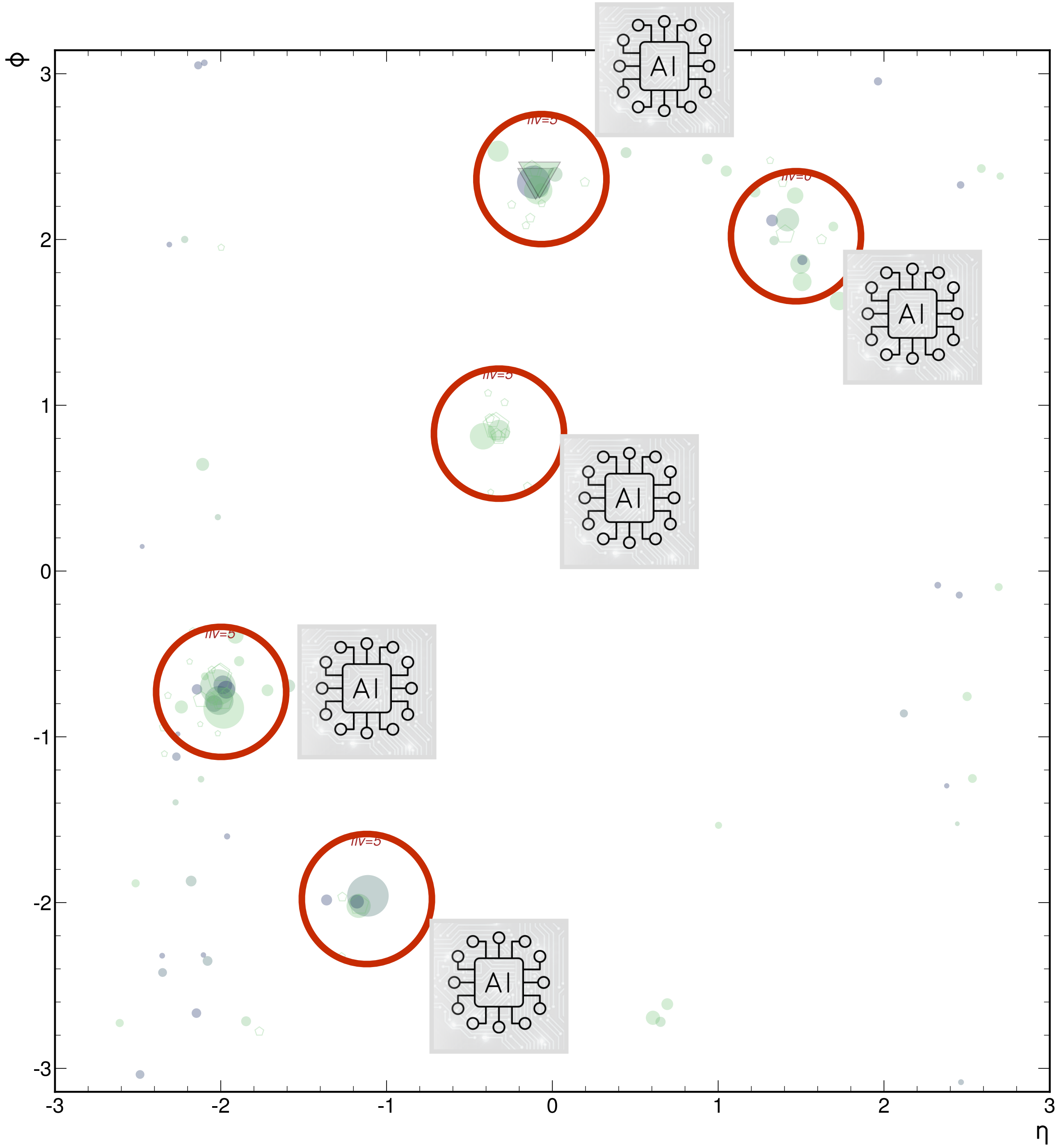
***HH*(4*b*) signal**

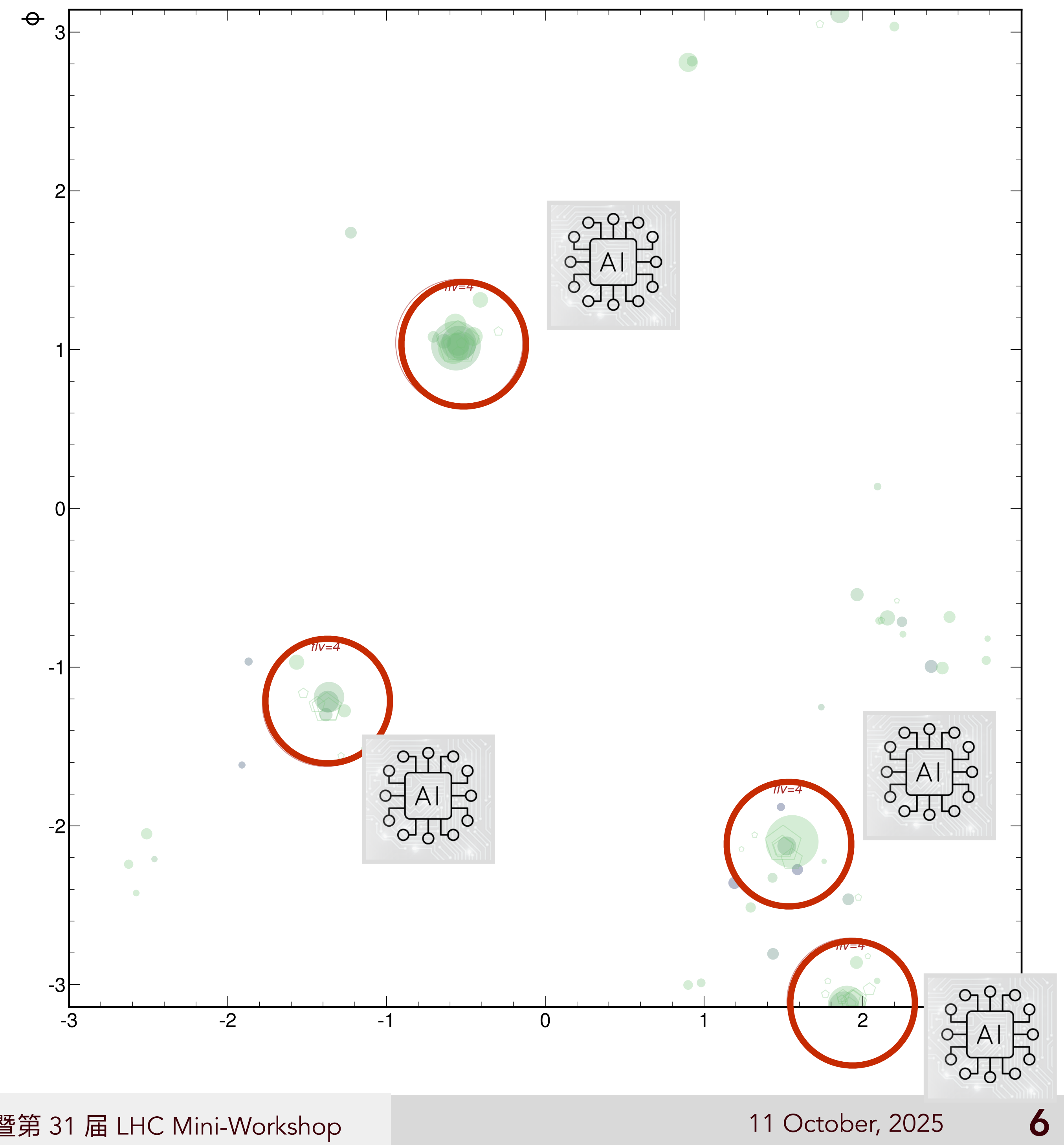


QCD multijet bkg.



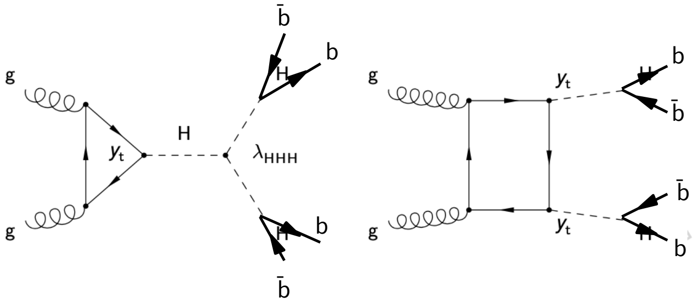
...



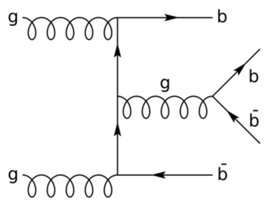


Illustrative view of the *HH*(4*b*) search

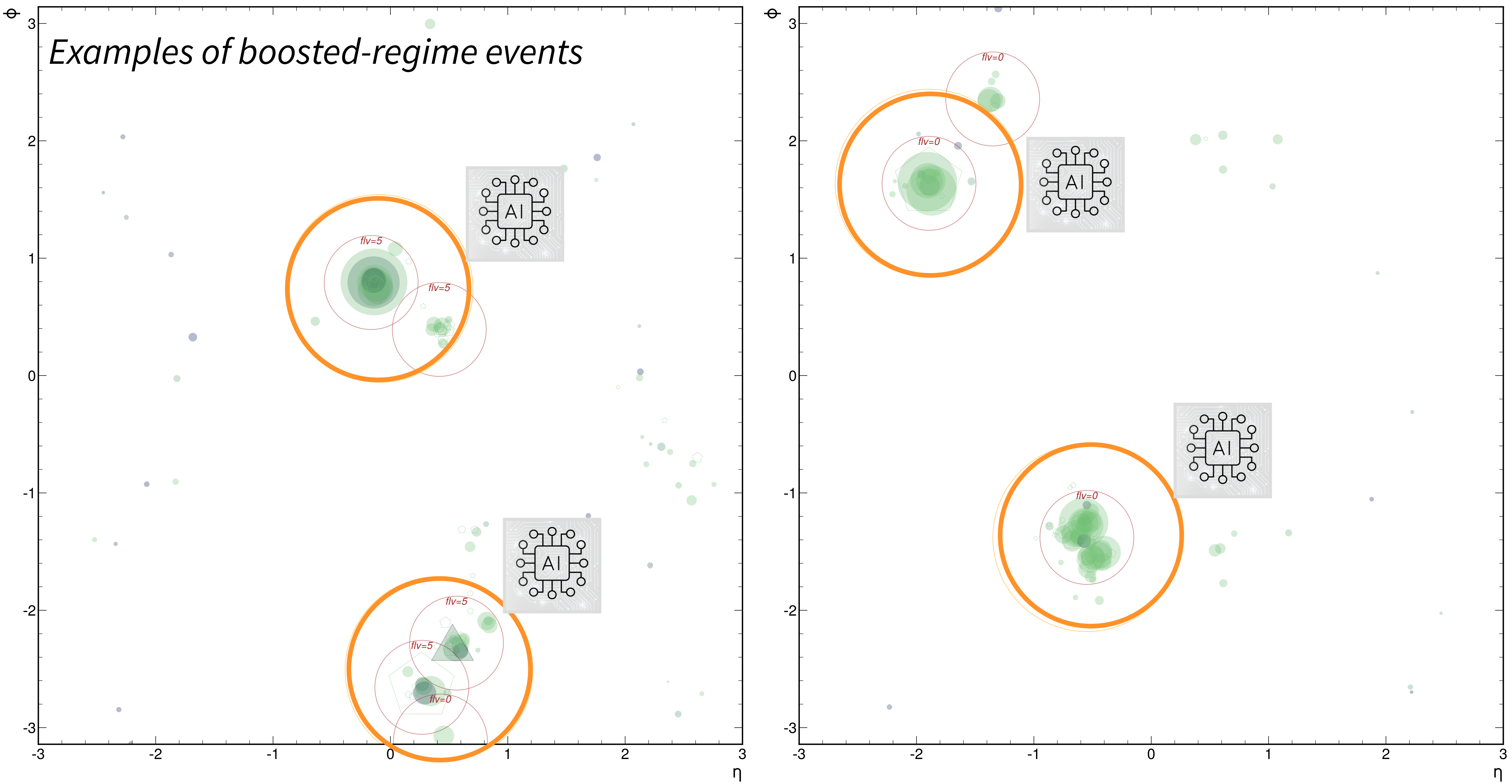
***HH*(4*b*) signal**



QCD multijet bkg.

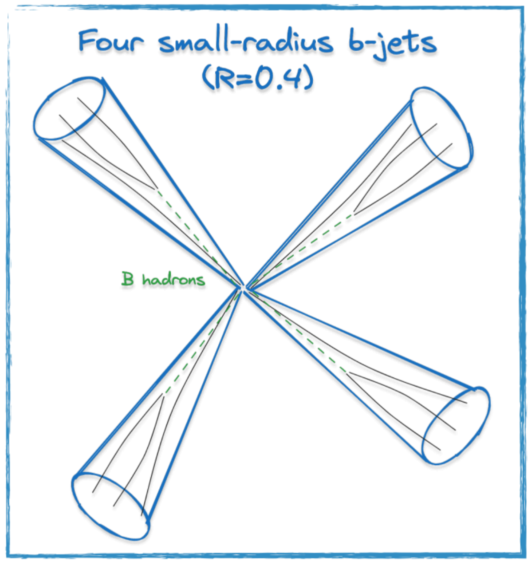


...



Resolved vs boosted?

Resolved regime



ATLAS papers:
1. ggF+VBF resolved
2. VBF boosted

PRL 129 (2022) 081802

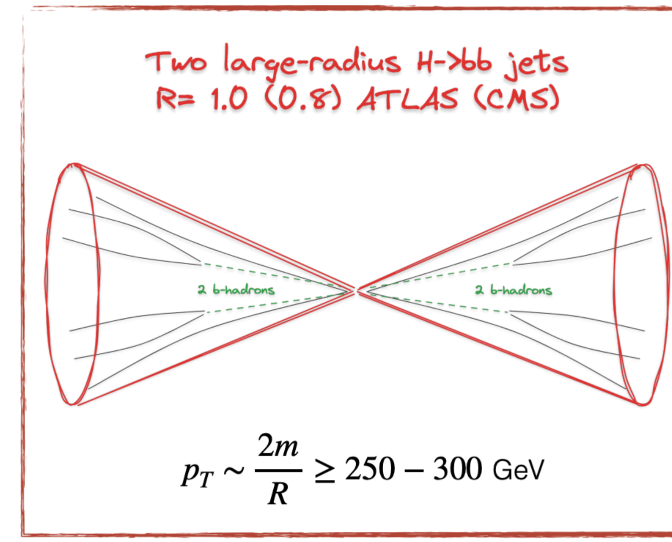
results website

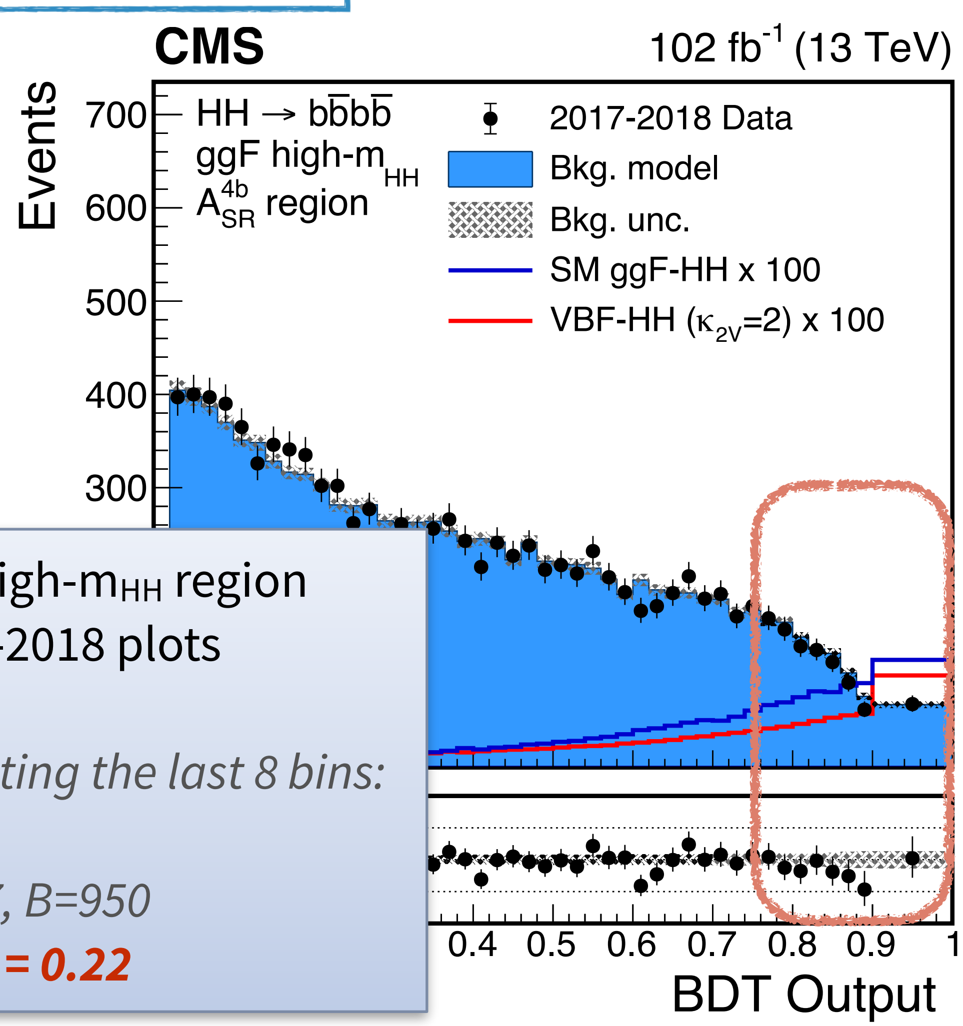
CMS papers:
1. ggF+VBF resolved
2. ggF+VBF boosted
3. ZZ/ZH in 4b resolved

PRL 131 (2023) 041803

results website

Boosted regime



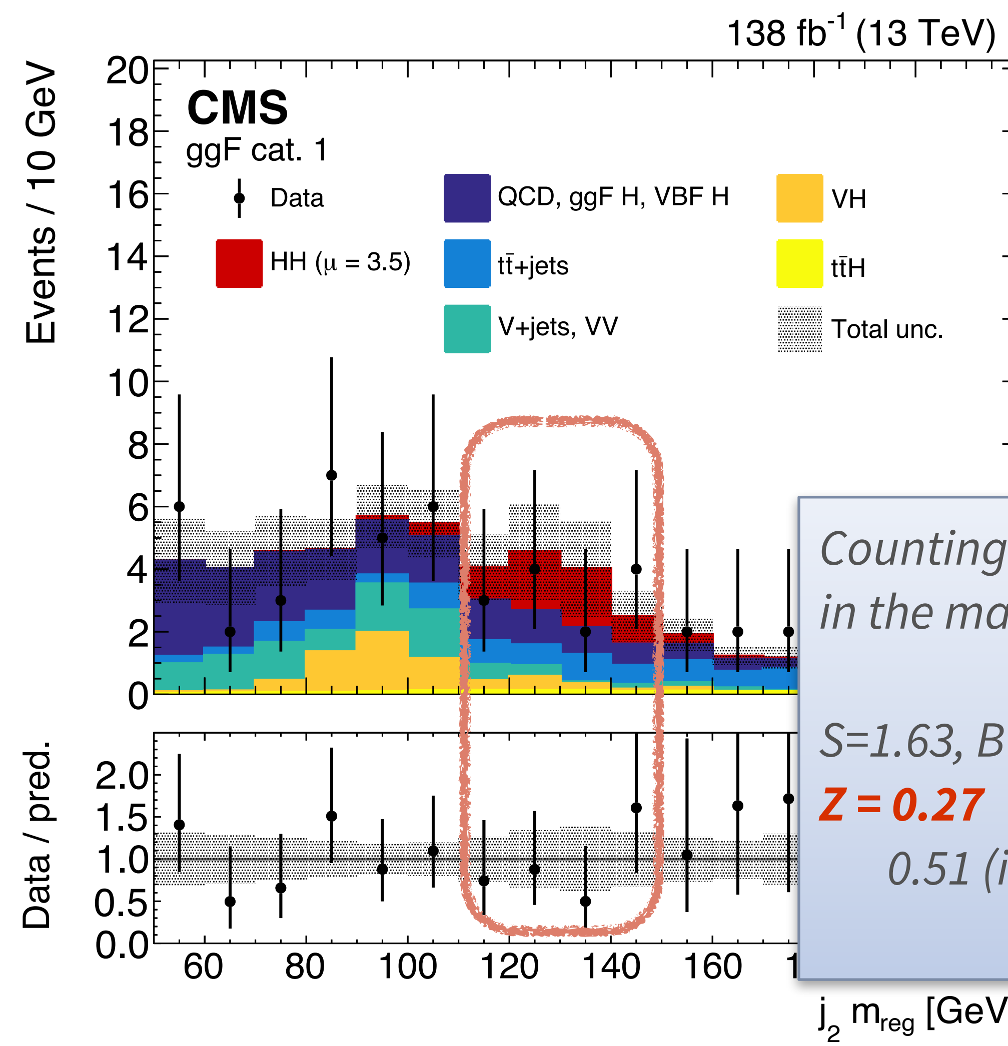


the high-m_{HH} region
2017-2018 plots

Counting the last 8 bins:

S=6.7, B=950

S/√B = 0.22



Counting the 4 bins in the mass window:

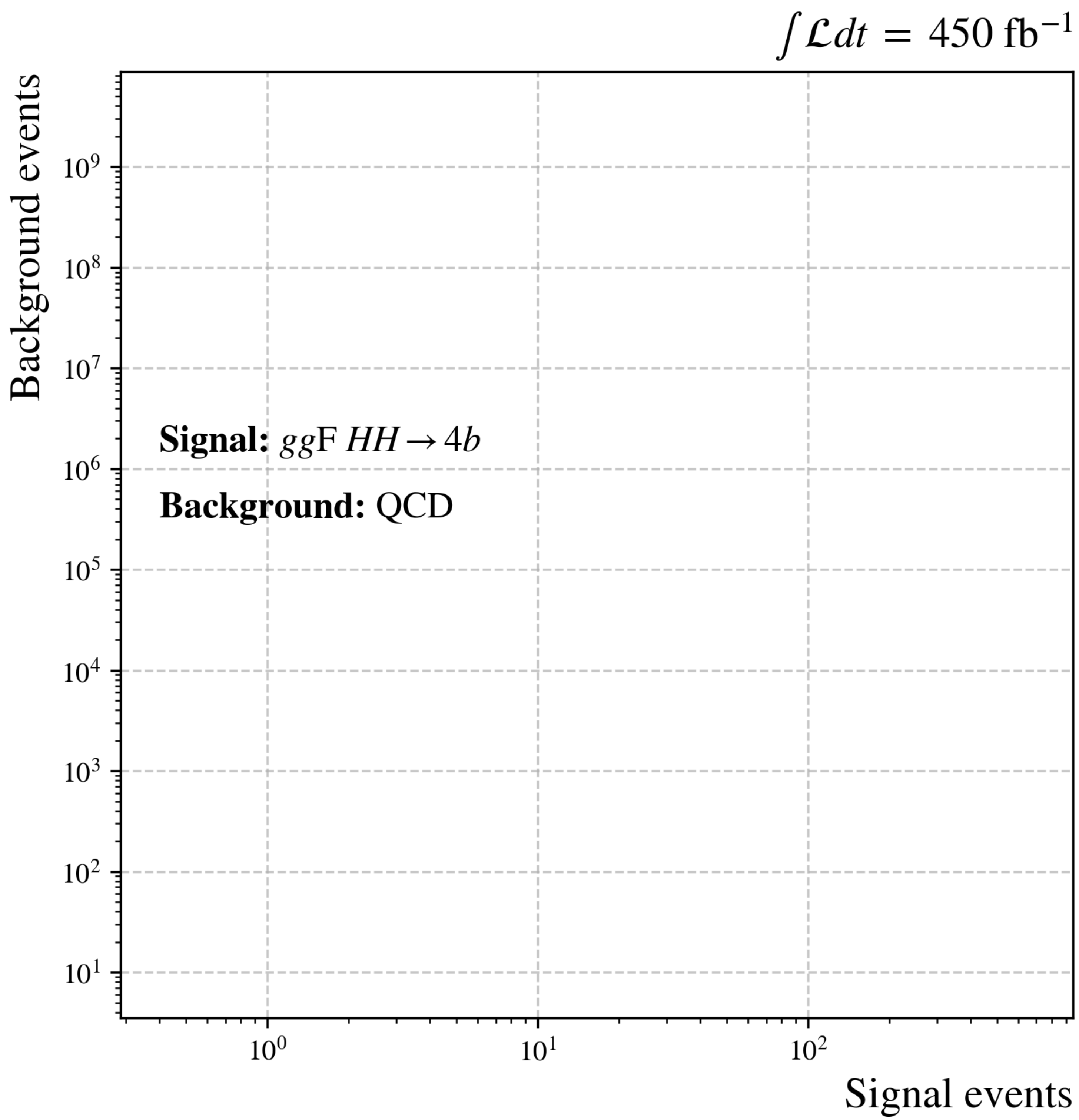
S=1.63, B=9.5, σ_B=4.9

Z = 0.27

0.51 (if B is certain)

Reproducing CMS's *HH*→4*b* results

→ First, reproduce CMS's ***HH*→4*b*** results using our highly realistic Delphes simulation



introducing the N(signal) vs N(bkg.) canvas

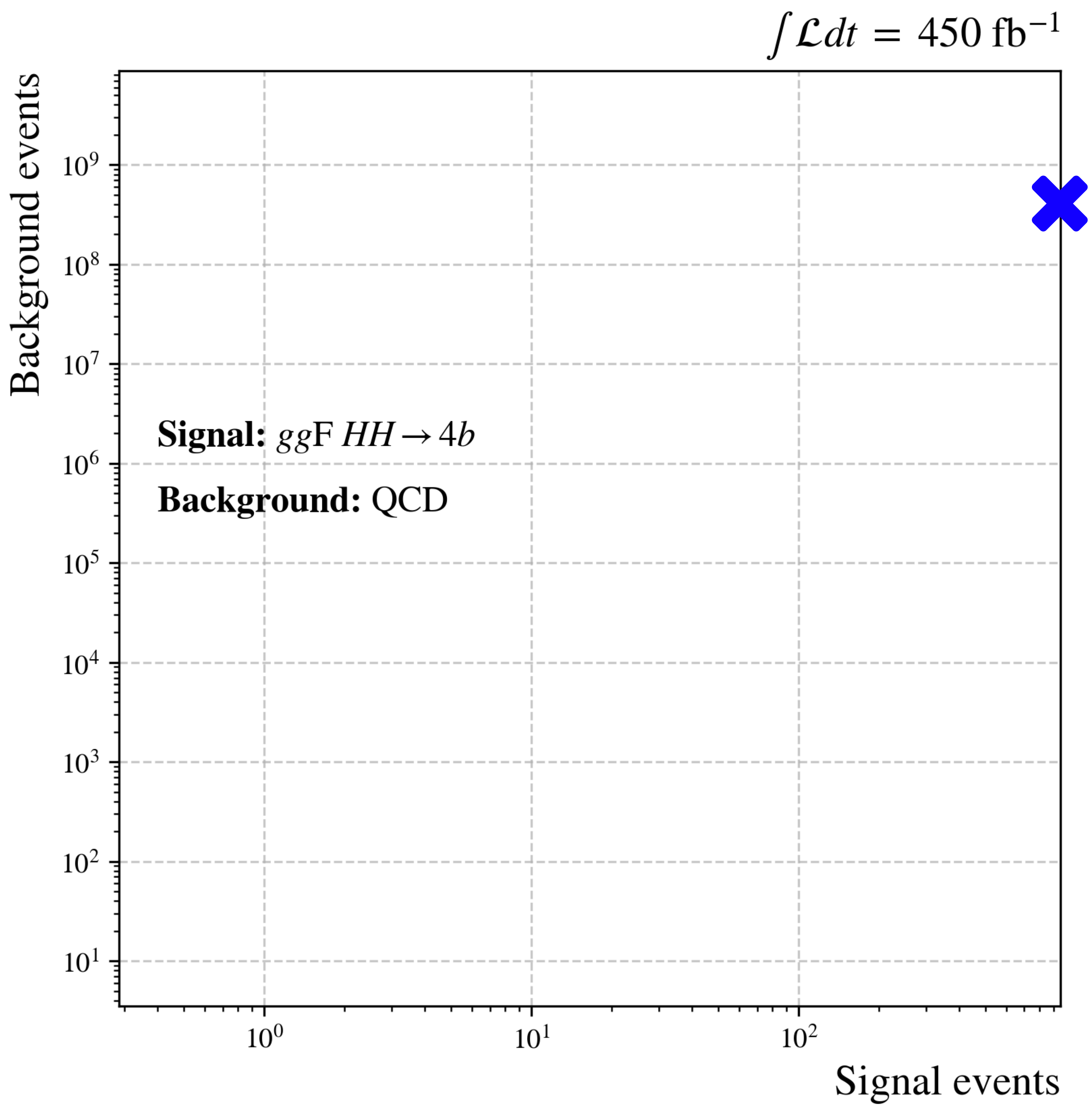
Concept:

✦ each "cut" can be visualized as a point on this canvas.

✦ **a "dynamic cut" forms a curve**—effectively a ROC curve scaled by total signal/bkg. event counts

Reproducing CMS's *HH→4b* results

➔ First, reproduce CMS's ***HH→4b*** results using our highly realistic Delphes simulation



Events (100 fb^{-1}) passing resolved *HH4b* triggers

Mimic CMS's 2018 trigger ("4j3b"):

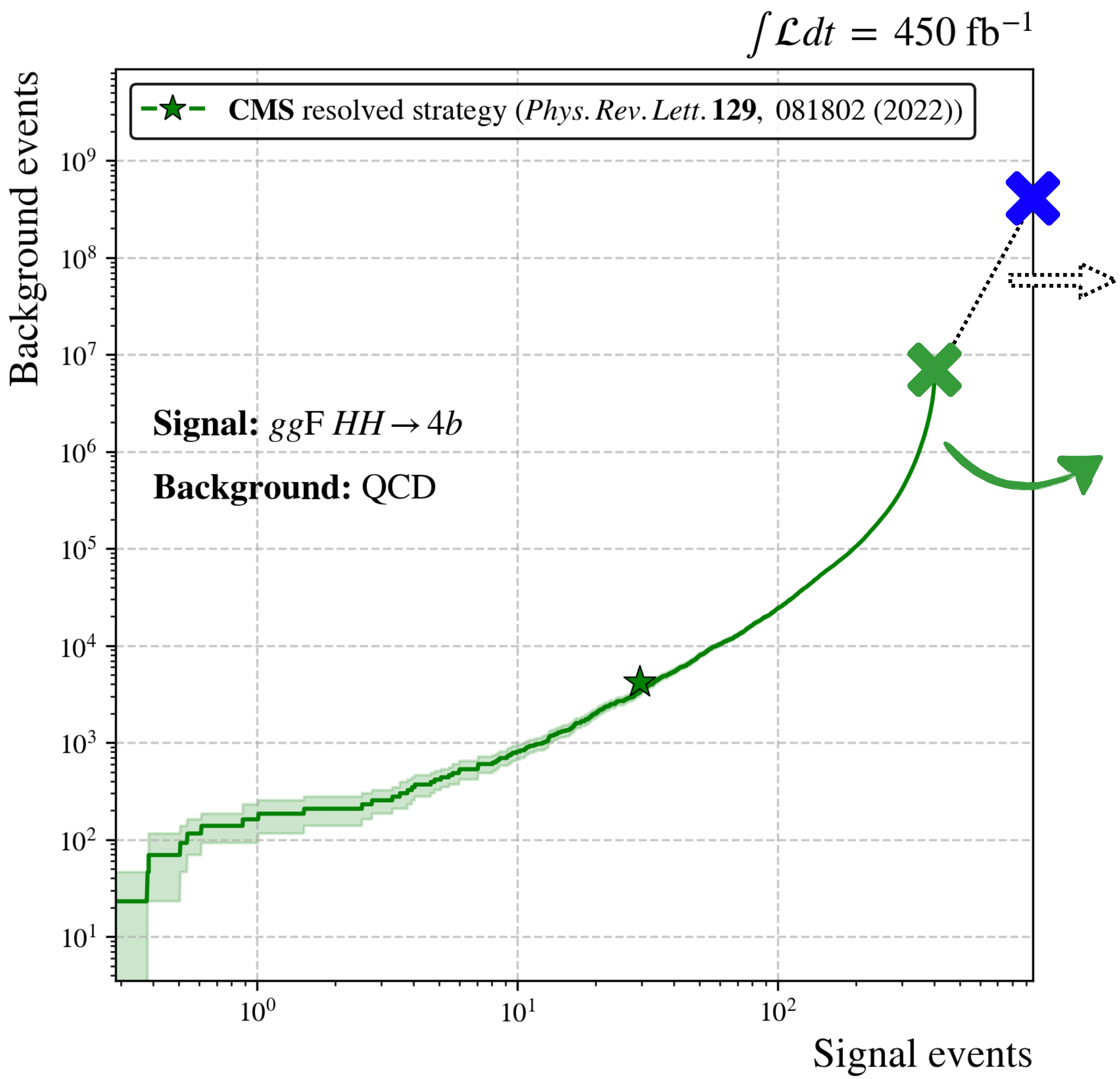
- ✦ H_T (AK4 p_T>30 GeV) > 330 GeV
- ✦ four jet p_T > 75/60/45/40 GeV
- ✦ ≥ 3 b-tagged jets (SophonAK4 loose b-tag WP: ε_l = 10%)

After trigger selection (under 450 fb^{-1})**:**

Signal (ggF *HH4b*): 960
Bkg. (QCD multijets): 3.5e8 (rate@2e34 ≈ 18 Hz)
pretty realistic!

Reproducing CMS's *HH→4b* results

➔ First, reproduce CMS's ***HH→4b*** results using our highly realistic Delphes simulation

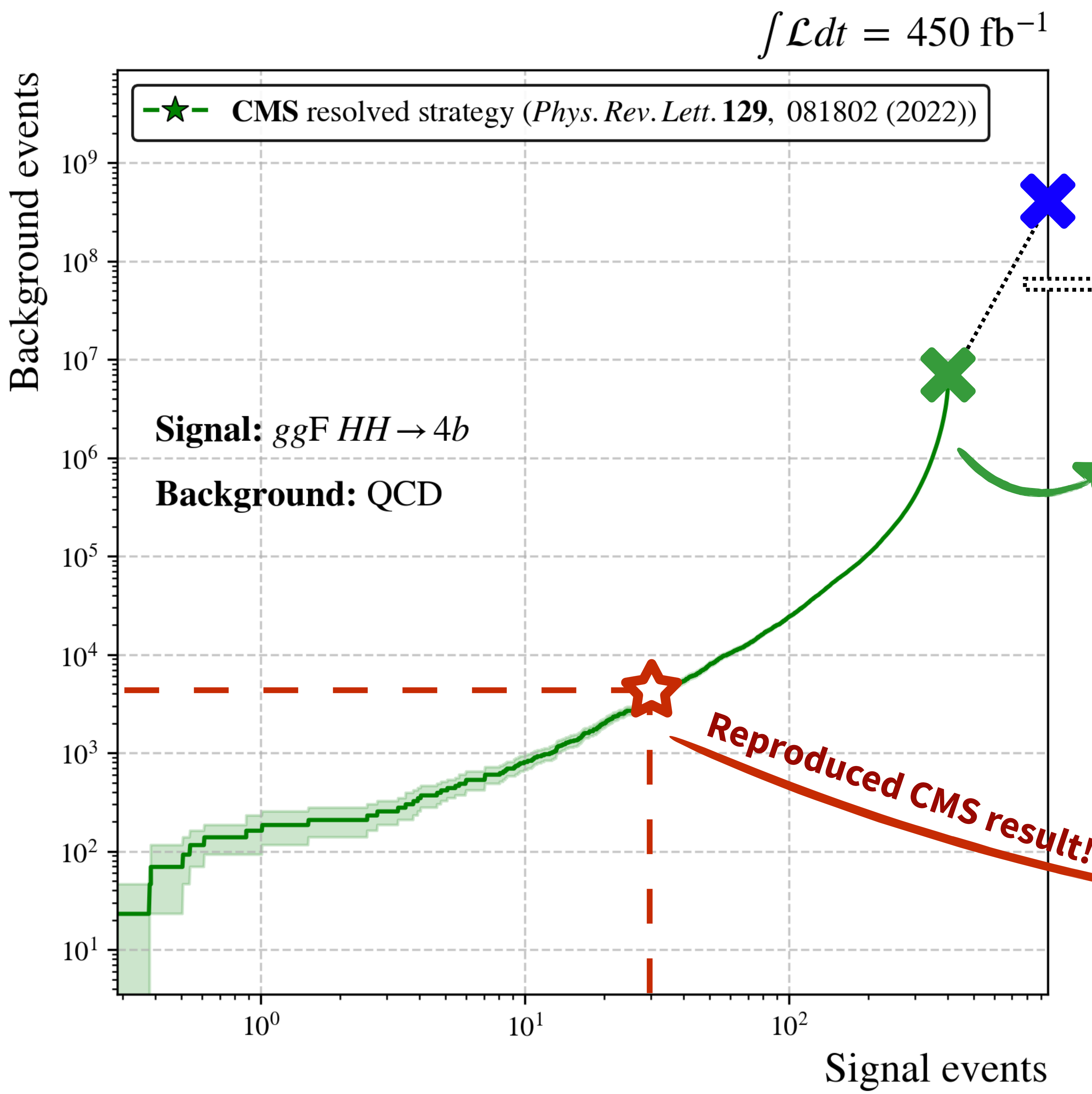


Further require all four leading jets to be b-tagged at *medium WP* (ε_l = 1%)

Then train a ParT classifier, using four jets' kinematic inputs to distinguish signal events vs QCD.
(ParT algorithm constructs ΔR, m_{ij} of all jet pairs, effectively enabling jet-pair association.)

Reproducing CMS's *HH→4b* results

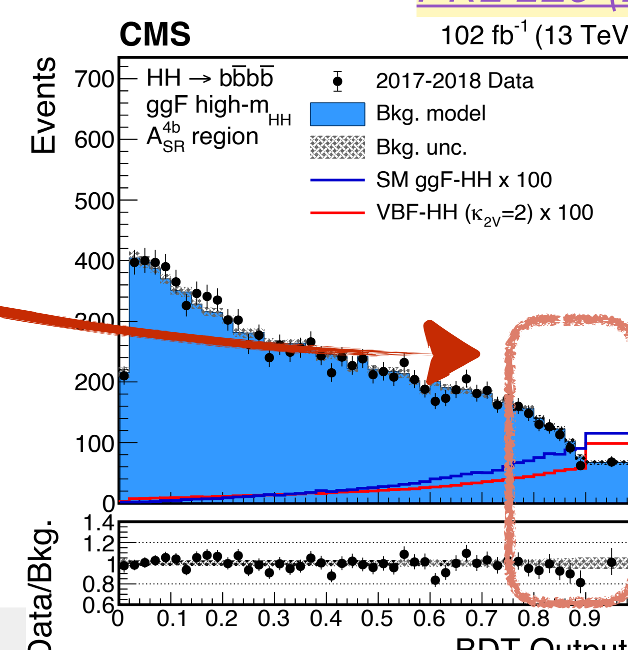
➔ First, reproduce CMS's ***HH→4b*** results using our highly realistic Delphes simulation



Further require all four leading jets to be b-tagged at *medium WP* (ε_l = 1%)

Then train a ParT classifier, using four jets' kinematic inputs to distinguish signal events vs QCD.
(ParT algorithm constructs ΔR, m_{ij} of all jet pairs, effectively enabling jet-pair association.)

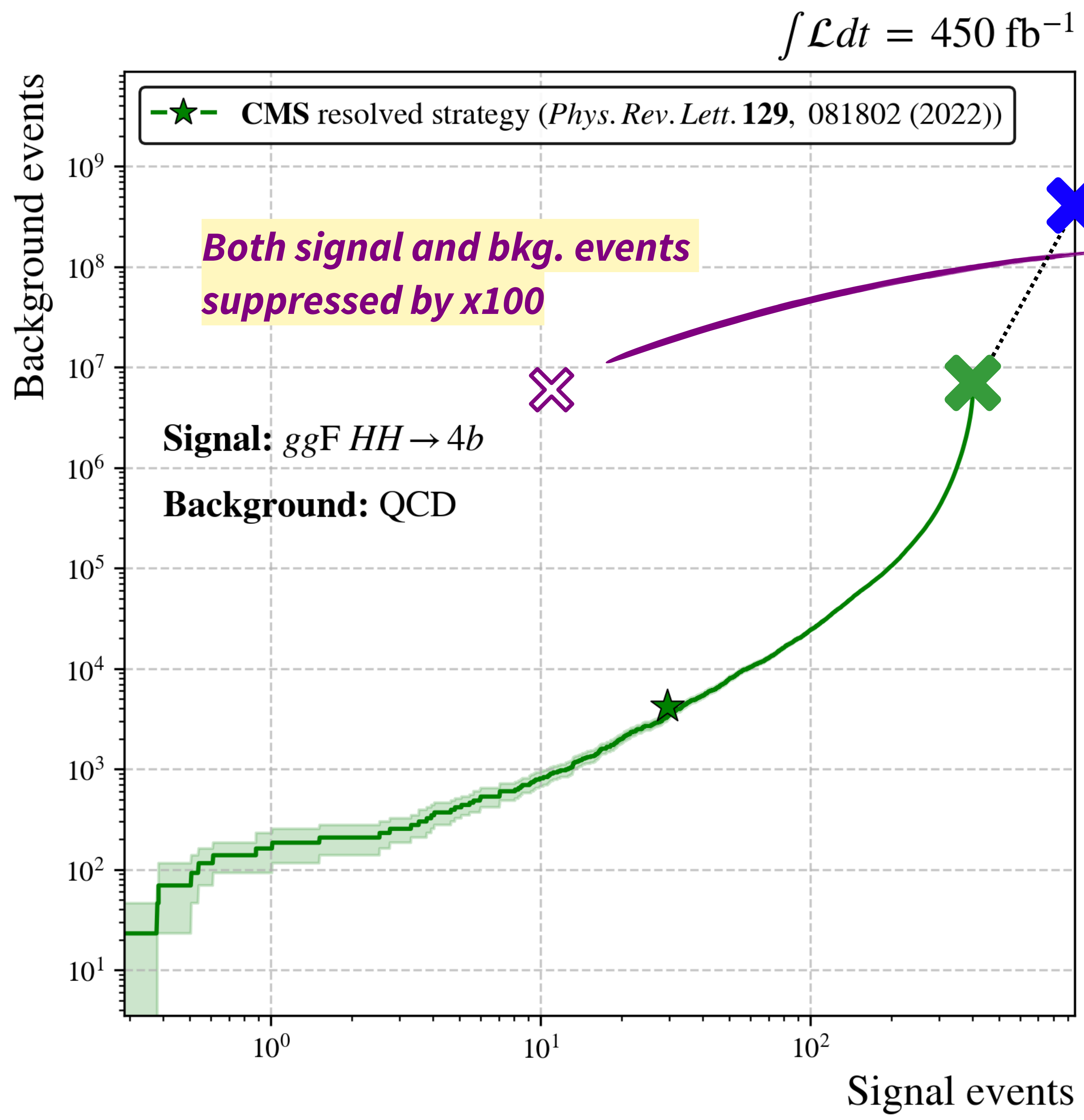
PRL 129 (2022) 081802



S=6.7
B=950

Reproducing CMS's *HH→4b* results

➔ First, reproduce CMS's ***HH→4b*** results using our highly realistic Delphes simulation

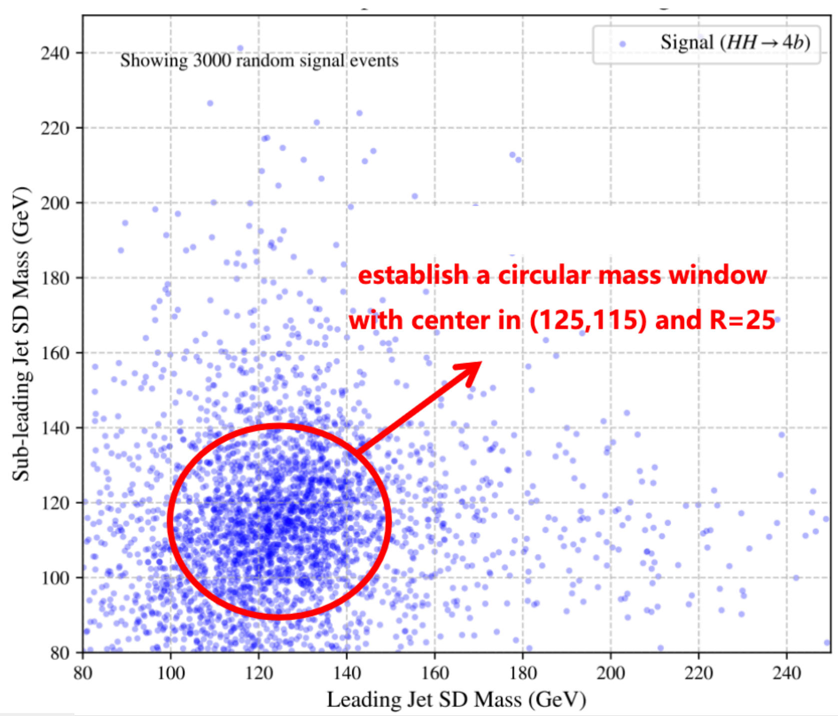


Then, we reproduce boosted *HH4b* results

① **Events (100 fb^{-1}) passing boosted-jet triggers**

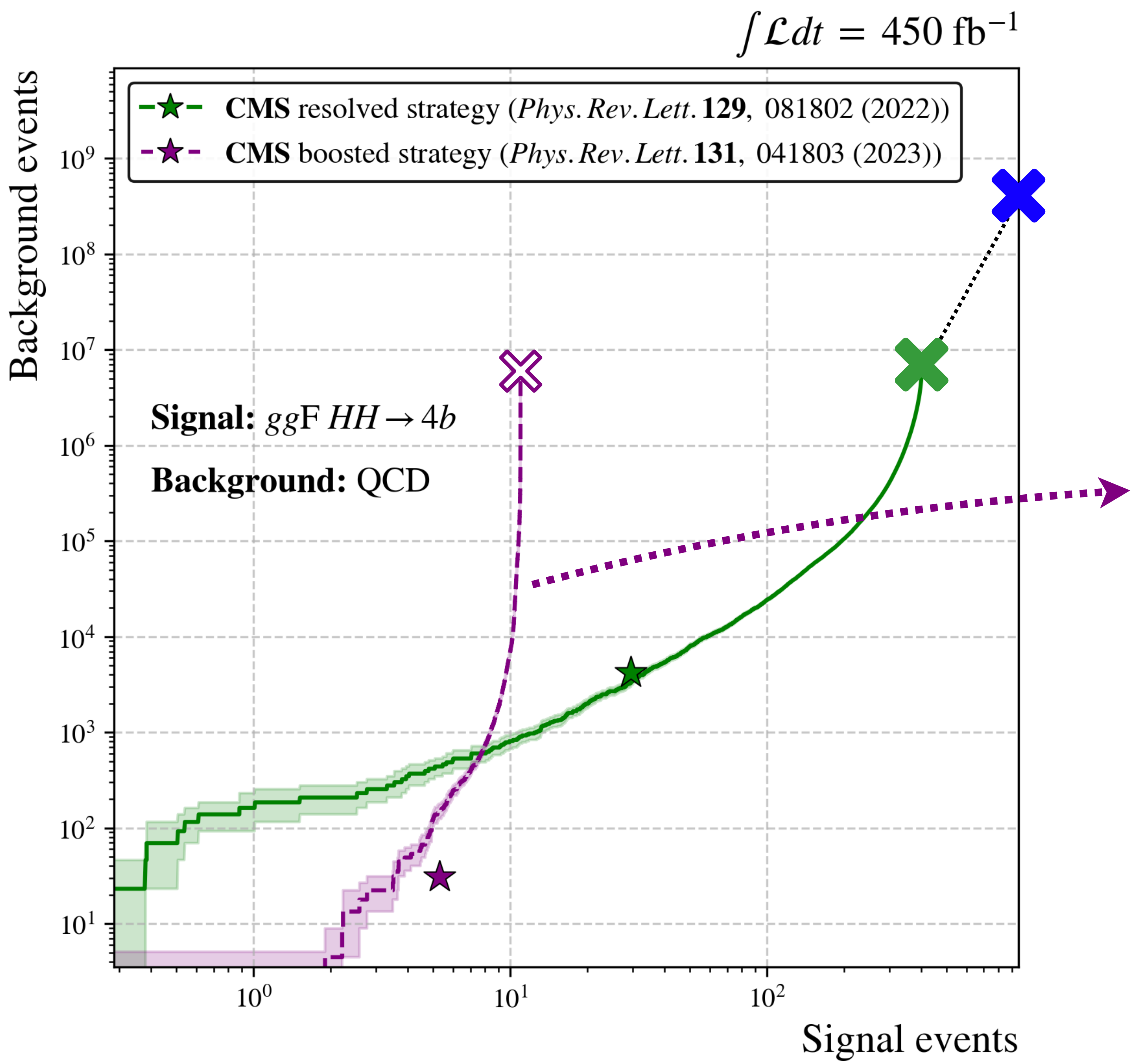
Mimic CMS's typical boosted-jet trigger:
- ✦ H_T (AK8 p_T>200 GeV) > 800 GeV
- ✦ leading AK8 jet trimmed mass > 50 GeV

② **And further require two leading jets to have soft-drop masses within the circle:**
- centered at (125, 115) with *R*=25 GeV



Reproducing CMS's *HH*→4*b* results

➔ First, reproduce CMS's ***HH*→4*b*** results using our highly realistic Delphes simulation

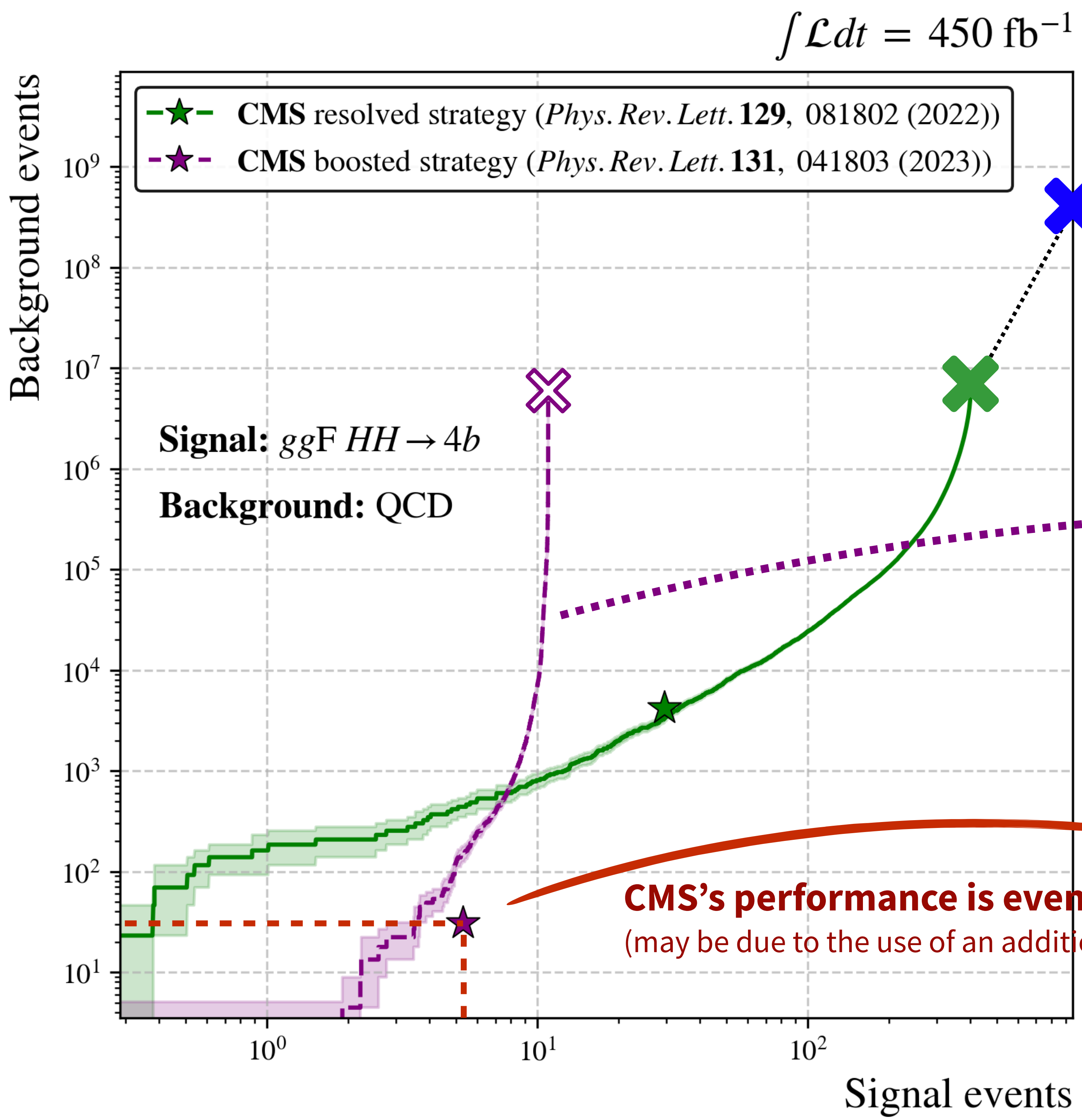


But the QCD suppression power is "astonishingly" powerful

using the sum of two AK8 jets' **Sophon (AK8) XbbvsQCD scores** as the discriminator

Reproducing CMS's *HH→4b* results

➔ First, reproduce CMS's ***HH→4b*** results using our highly realistic Delphes simulation

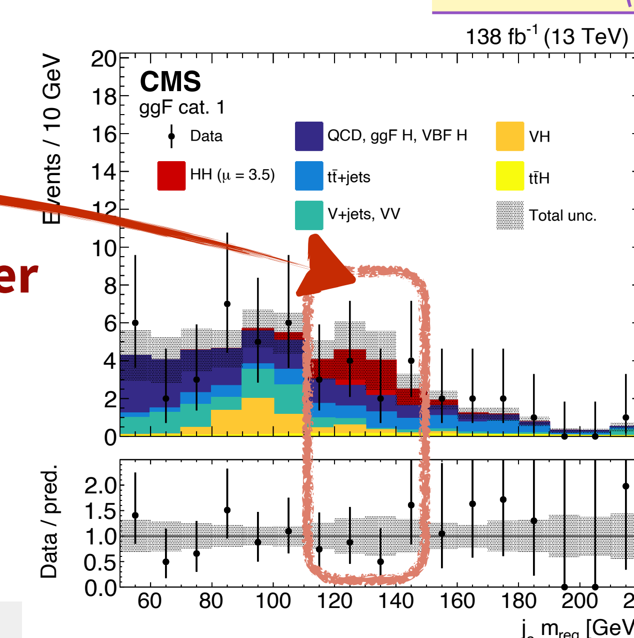


CMS's performance is even stronger
(may be due to the use of an additional BDT)

But the QCD suppression power is "astonishingly" powerful

using the sum of two AK8 jets' **Sophon (AK8) XbbvsQCD scores** as the discriminator

PRL 131 (2023) 041803

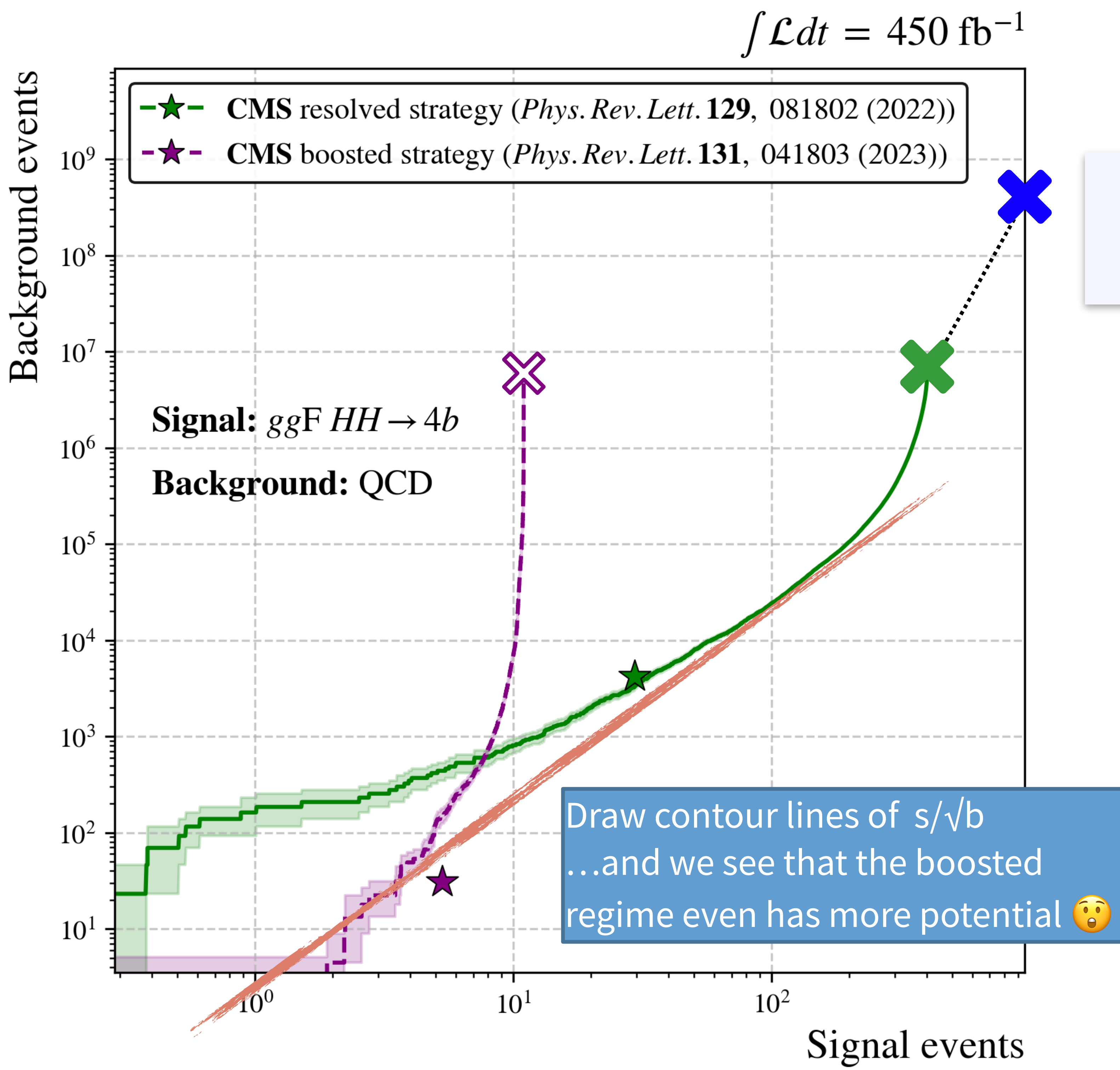


S=1.63

B(QCD)≈4.1

Reproducing CMS's *HH→4b* results

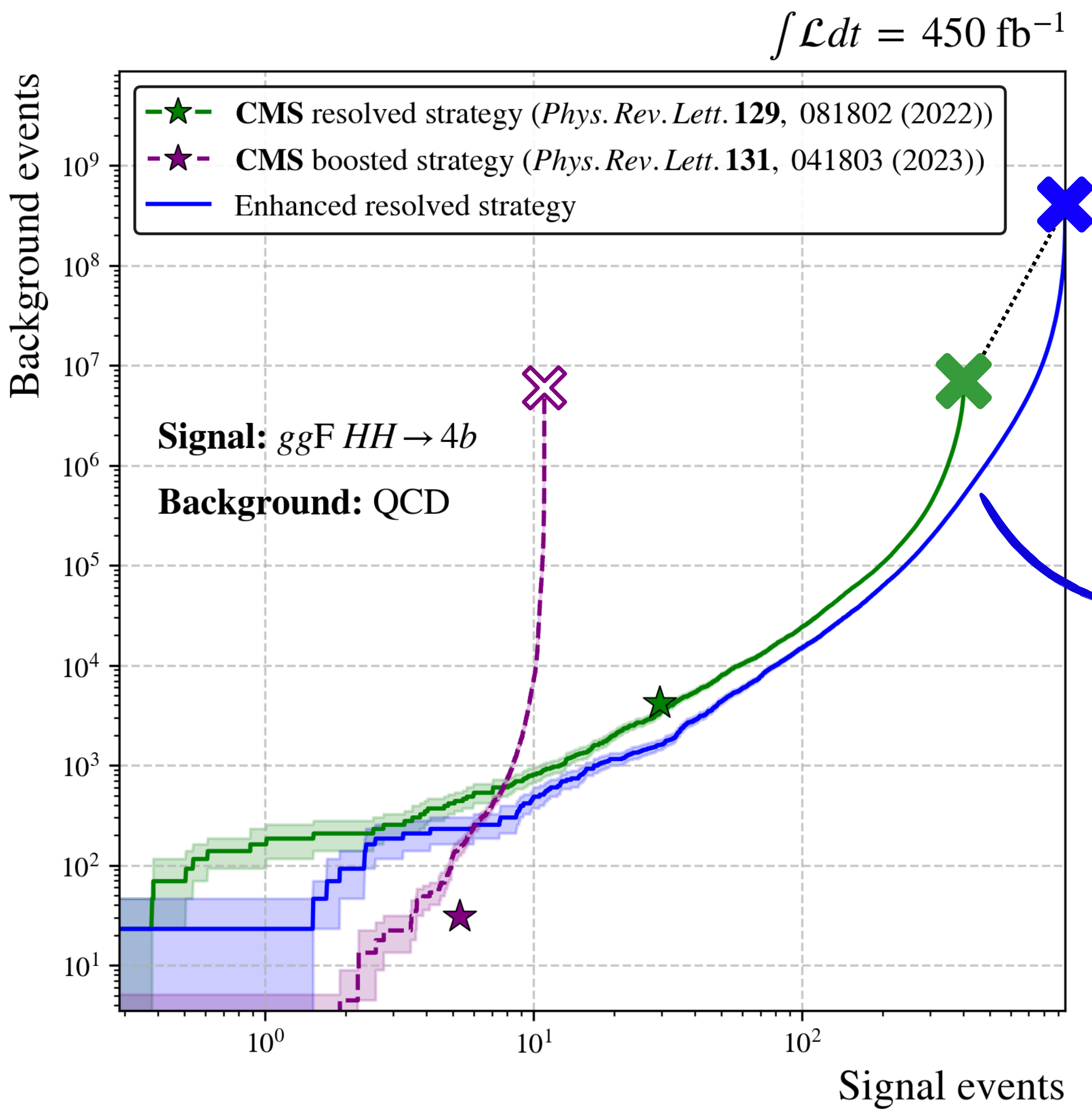
➔ First, reproduce CMS's ***HH→4b*** results using our highly realistic Delphes simulation



Remarks from comparison of CMS's resolved & boosted analyses

Reproducing CMS's *HH*→4*b* results

➔ First, reproduce CMS's ***HH*→4*b*** results using our highly realistic Delphes simulation



Remarks from comparison of CMS's resolved & boosted analyses

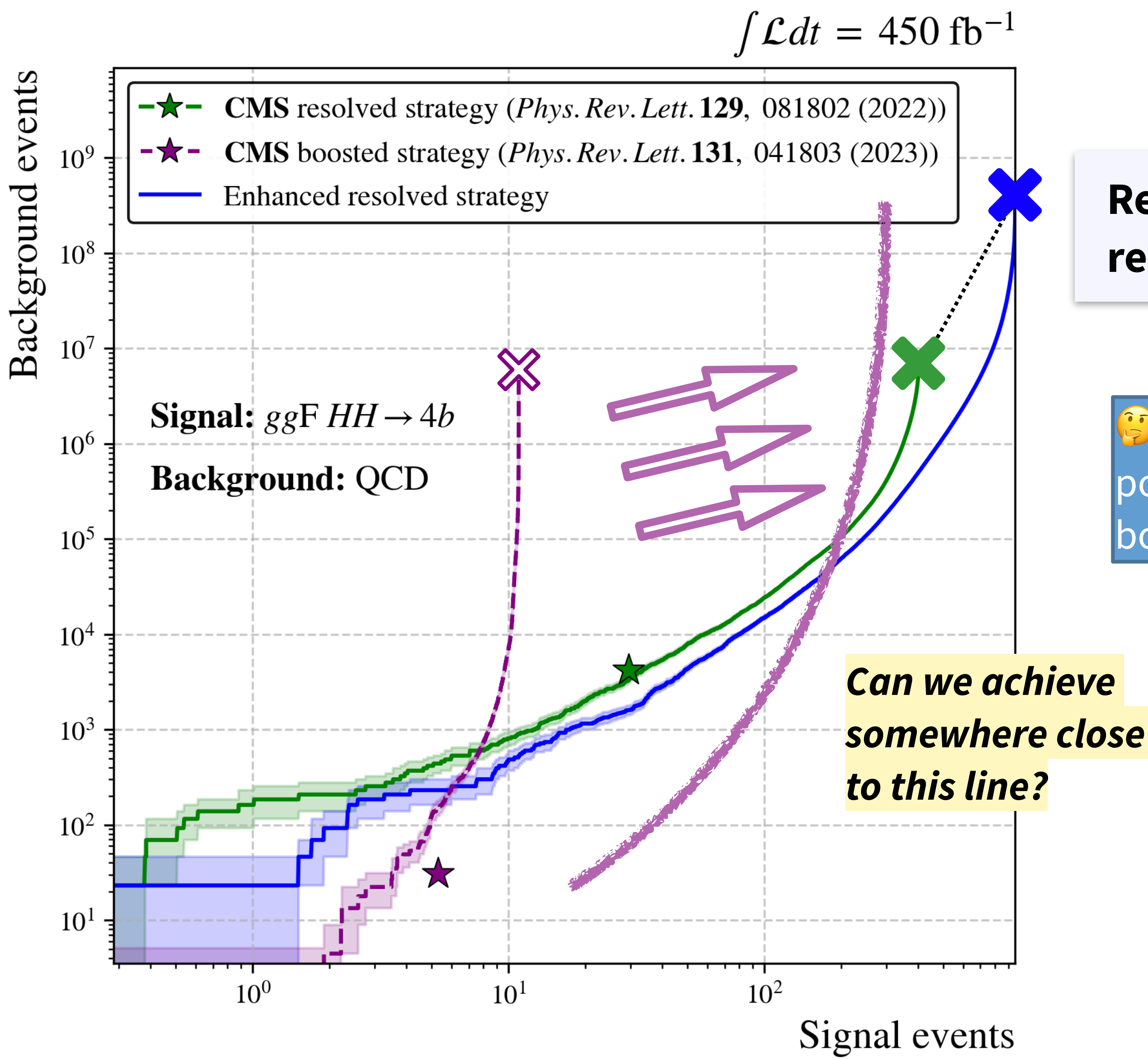
But… can we improve the resolved-regime strategy?
😕 Unfortunately, if we insist on using AK4 jets in the resolved regime, we yield only marginal gains.

Here, we benchmark the best achievable performance relying on small-*R* jets

Train a ParT on events passing resolved triggers, utilizing full AK4 jet information—including tagging scores

Reproducing CMS's *HH*→4*b* results

➔ First, reproduce CMS's ***HH*→4*b*** results using our highly realistic Delphes simulation

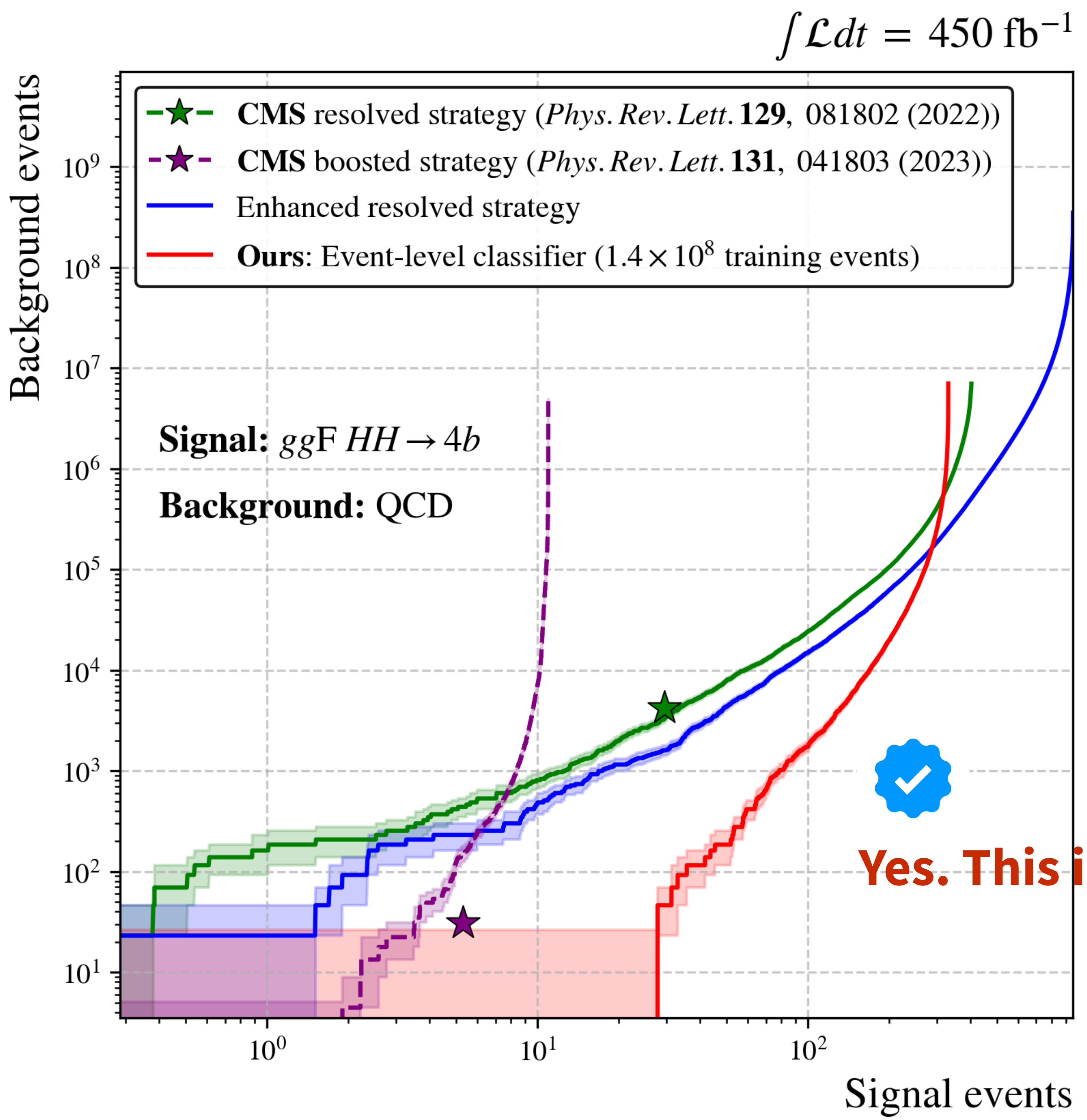


Remarks from comparison of CMS's resolved & boosted analyses

🤔 Naively, the strong discriminative power should generalize across all boosting conditions.

Reproducing CMS's *HH→4b* results

→ First, reproduce CMS's ***HH→4b*** results using our highly realistic Delphes simulation



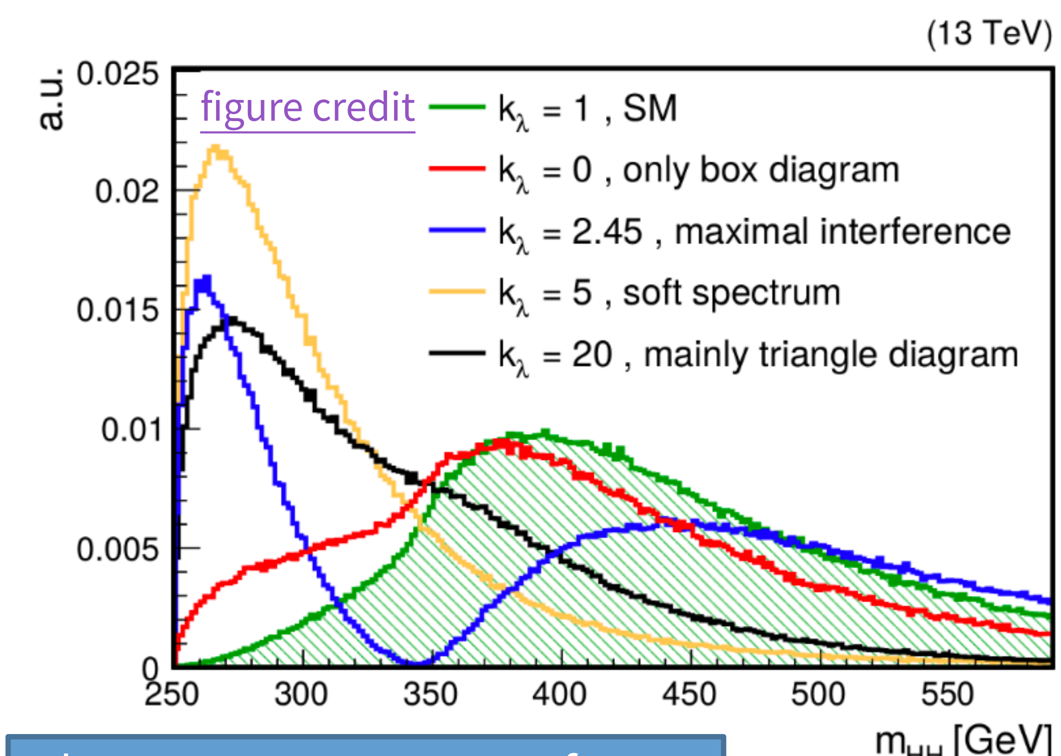
Yes. This is achievable.

Introducing our *HH4b* strategy

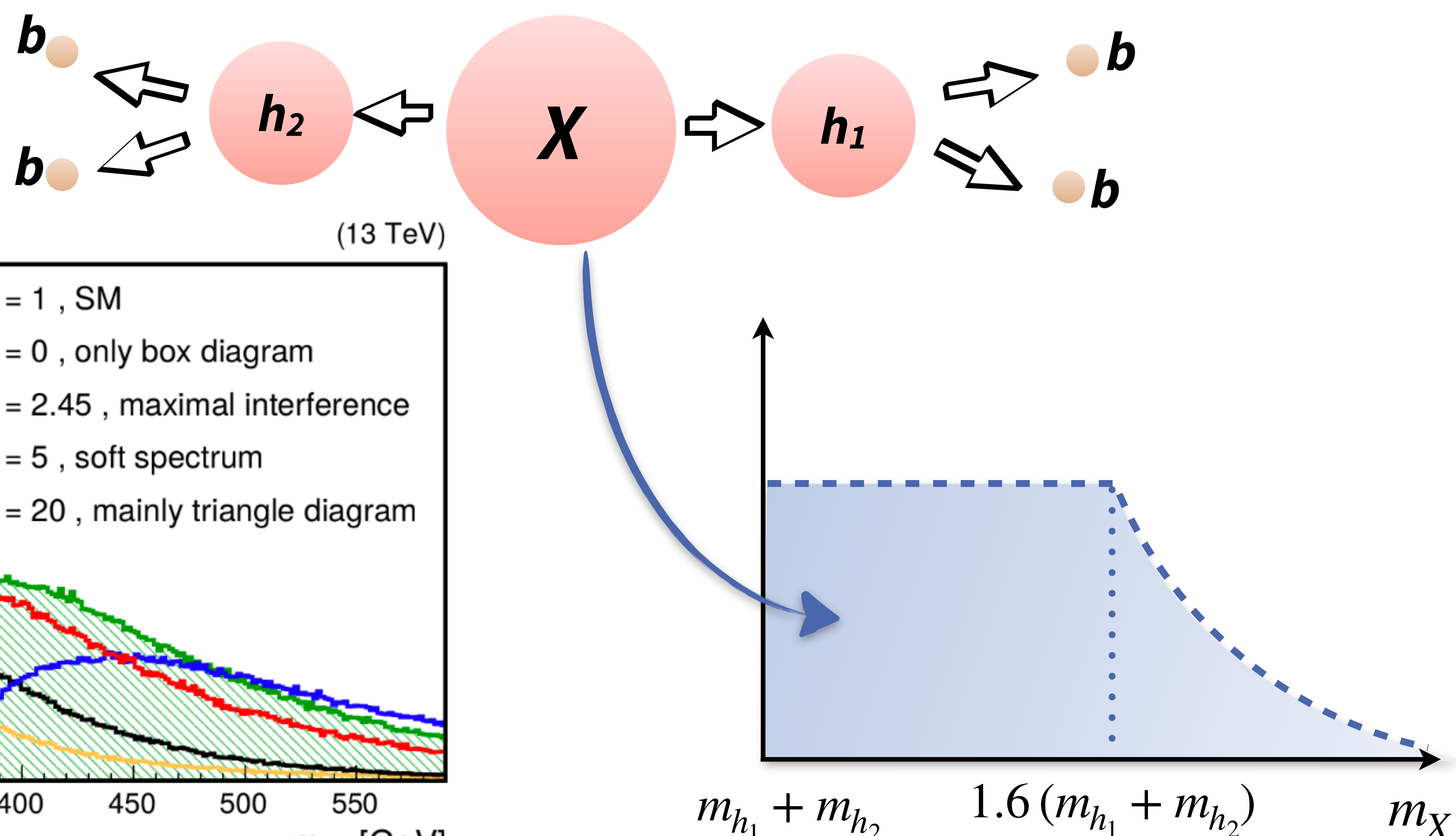
Analysis strategy: training a $X\to h_1h_2\to 4b$ vs bkg. classifier

➔ Train "$X\to h_1h_2\to 4b$" signal versus QCD and $t\bar{t}$ (passing the "4j3b" resolved trigger!)

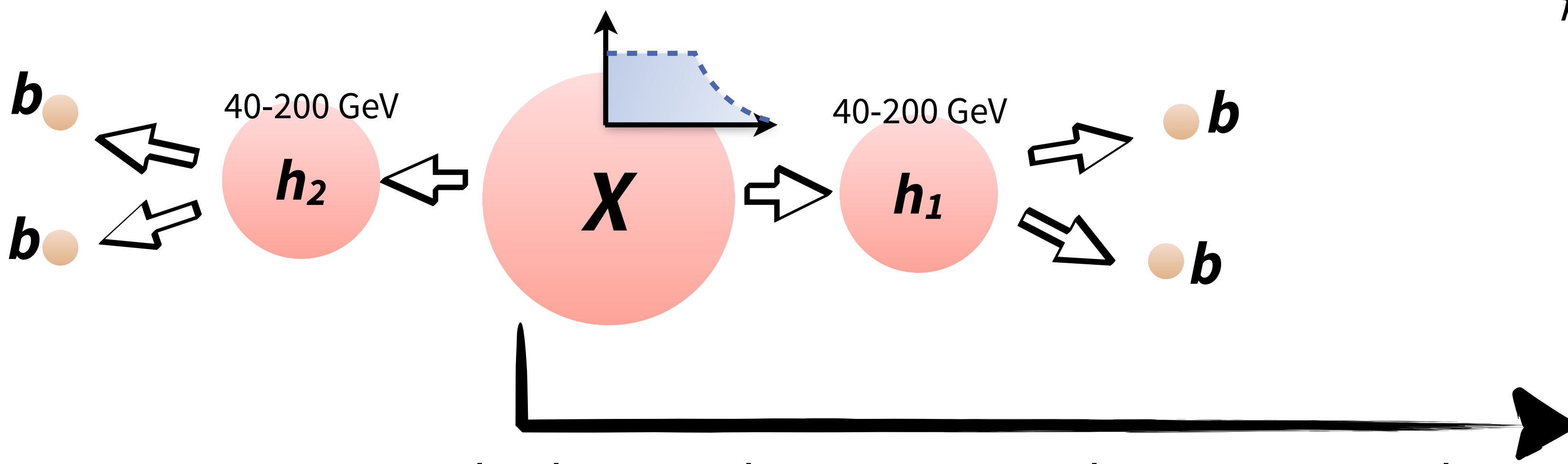
- h_1 and h_2 masses: uniformly distributed in [40, 200] GeV
- X mass → taking various HH→4b topologies as a reference
- using full-event particle-level inputs



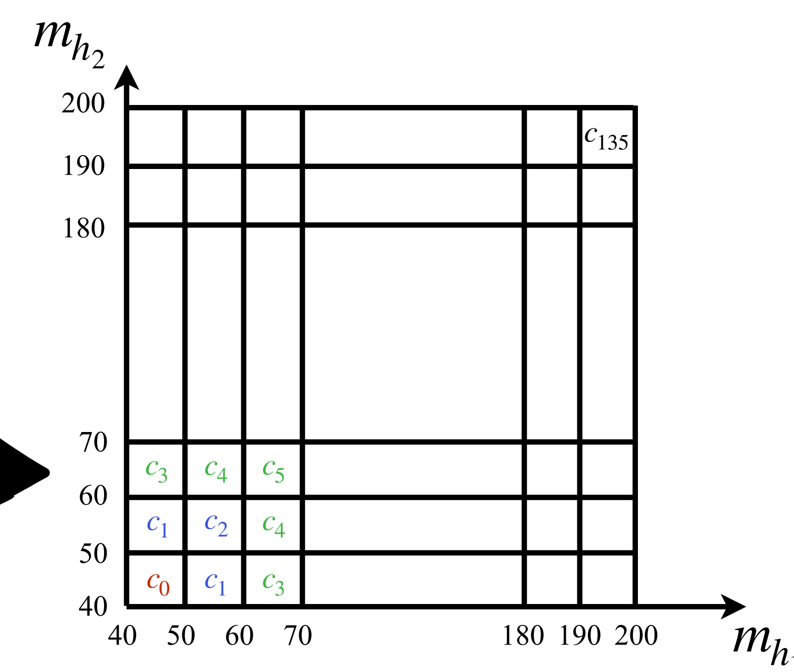
Take m_{HH} spectrum as a reference



Analysis strategy: training a $X\to h_1 h_2 \to 4b$ vs bkg. classifier

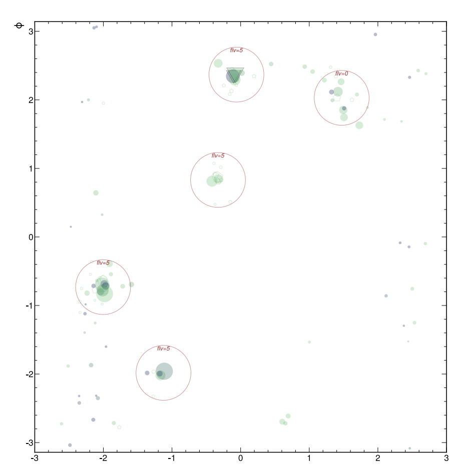


Split the signal training samples into 136 grids (classes), based on (m_{h_1}, m_{h_2})

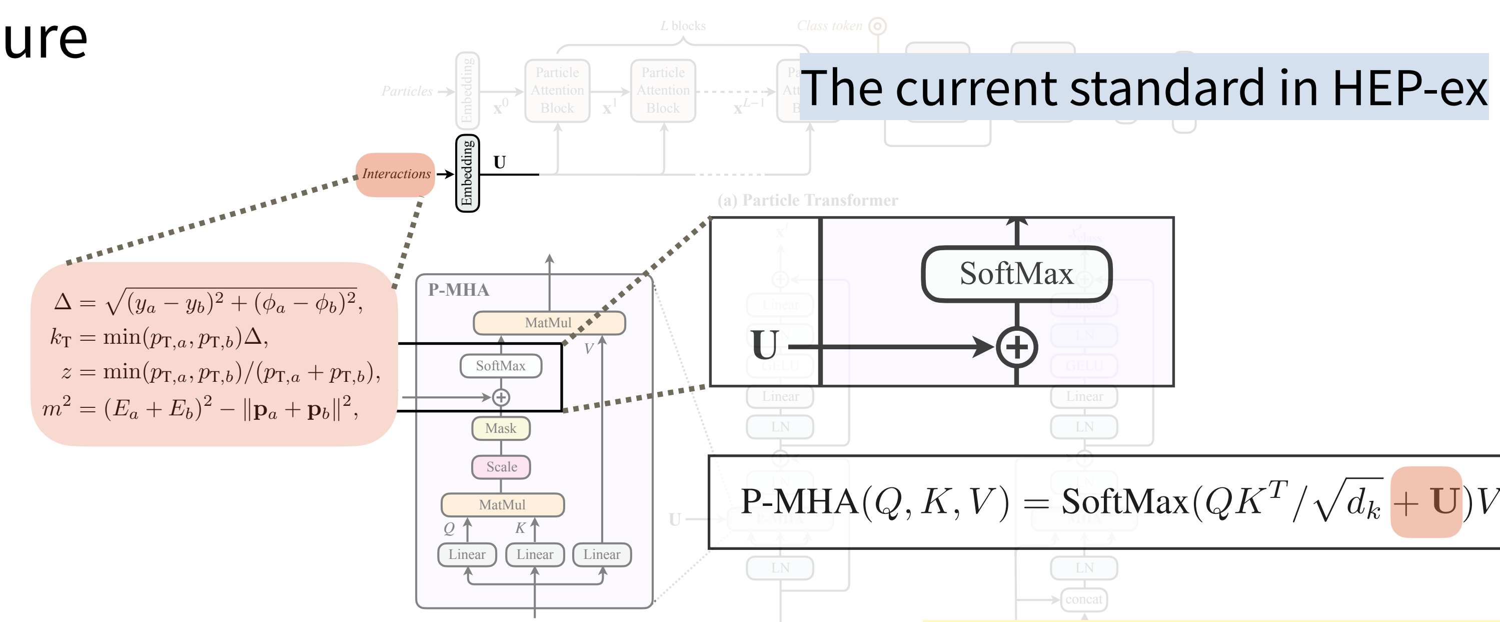


Train a "136+2 class" classifier

- on full-event reconstructed particle inputs (pileup-mitigated)
- using a **Particle Transformer** architecture

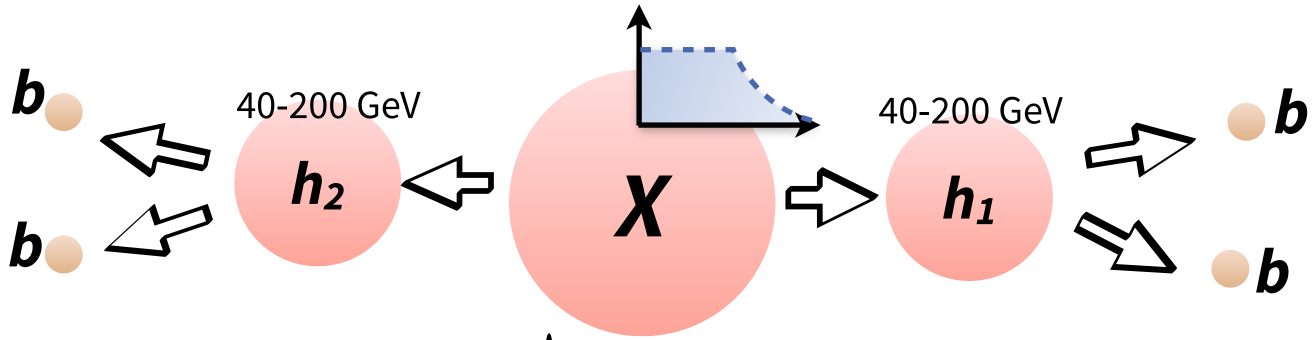


full-event particles

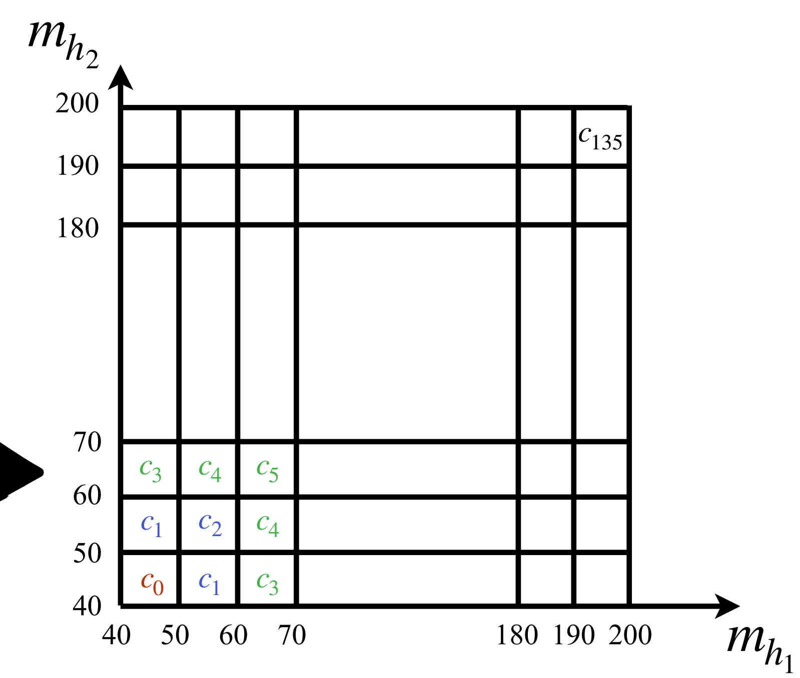


H. Qu, CL, S. Qian. ICML 2022

Analysis strategy: training a $X \to h_1 h_2 \to 4b$ vs bkg. classifier



Split the signal training samples into 136 grids (classes), based on (m_{h_1}, m_{h_2})

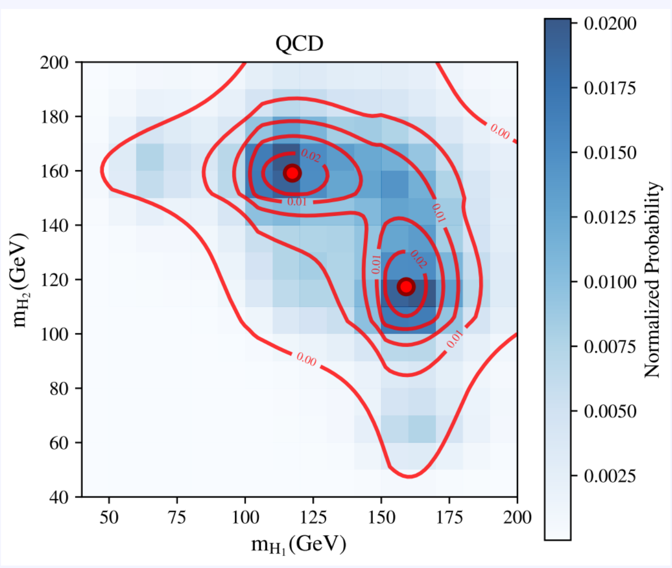


The "136+2" classifier offers two analysis tools

A powerful bkg-veto classifier

- p_{sig} = the sum of 136 signal-class scores, defining **discr = $p_{sig} / (p_{sig} + p_{QCD} + p_{ttbar})$** provides a powerful background-veto discriminant
- naturally mass-decorrelated due to the flat prior of (m_{h_1}, m_{h_2})

An (m_{h_1}, m_{h_2}) estimator

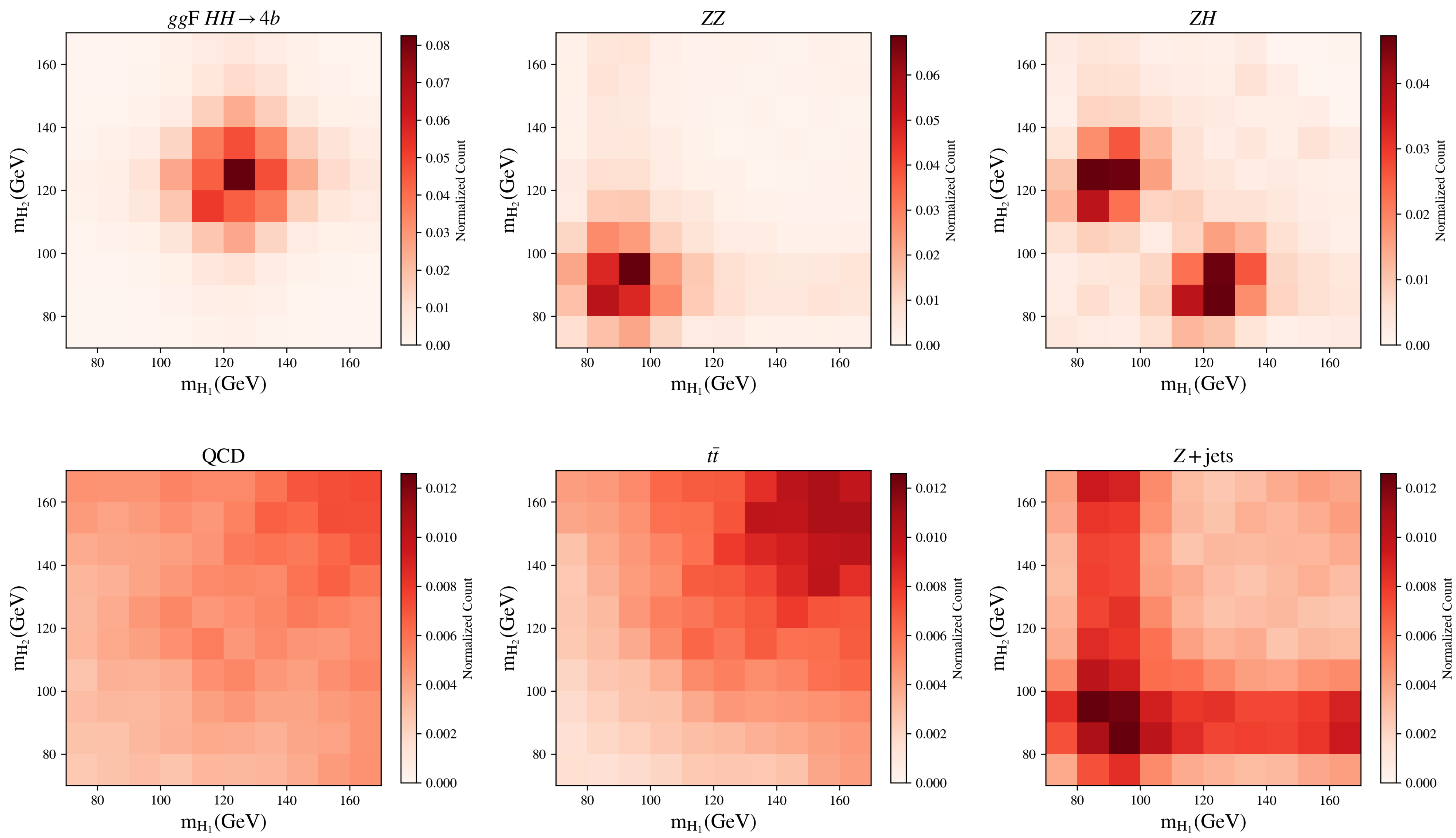


- 136-grid NN output estimated a discrete probability density ratio (by Neyman–Pearson lemma)
- **Perform a fit to extract the most probable point $(m^0_{h_1}, m^0_{h_2})$**

Mass distributions

➔ Distributions on (m_{h_1}, m_{h_2}) plane after selection

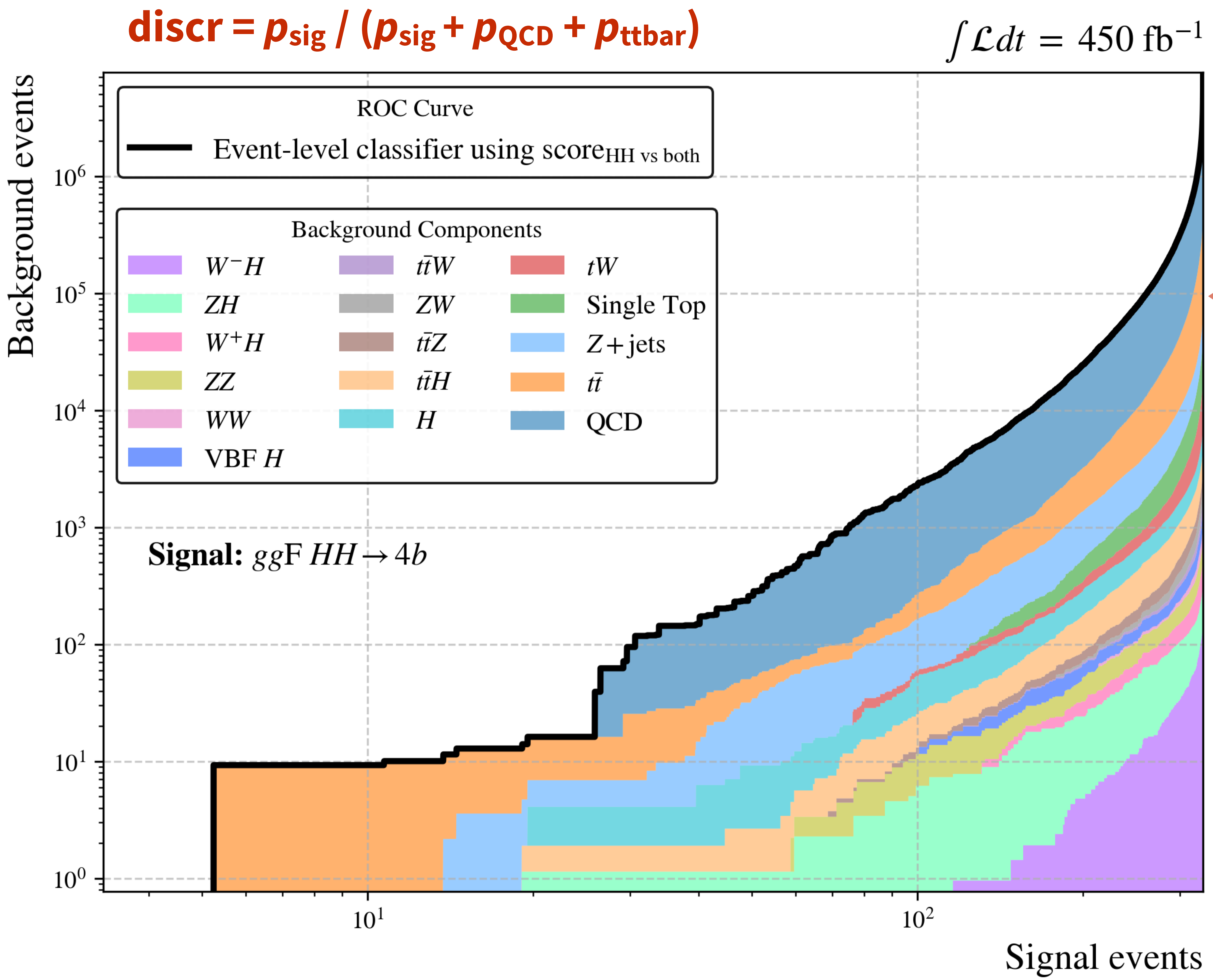
discr = p_{sig} / (p_{sig} + p_{QCD} + p_{ttbar}) > 0.9

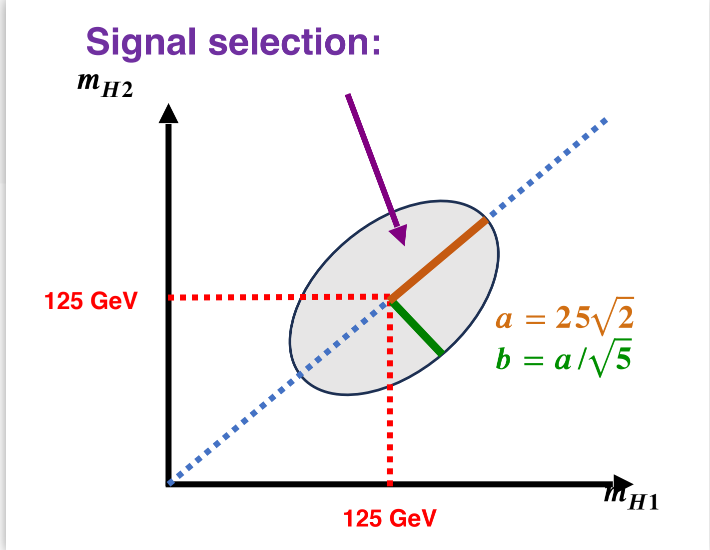


Expected shapes

- **smooth QCD and $t\bar{t}$ spectrum**
- single-Z/H: form ridge-like structures
- ZZ/ZH/HH: 2D peak-like structures

Effect of analysis selection

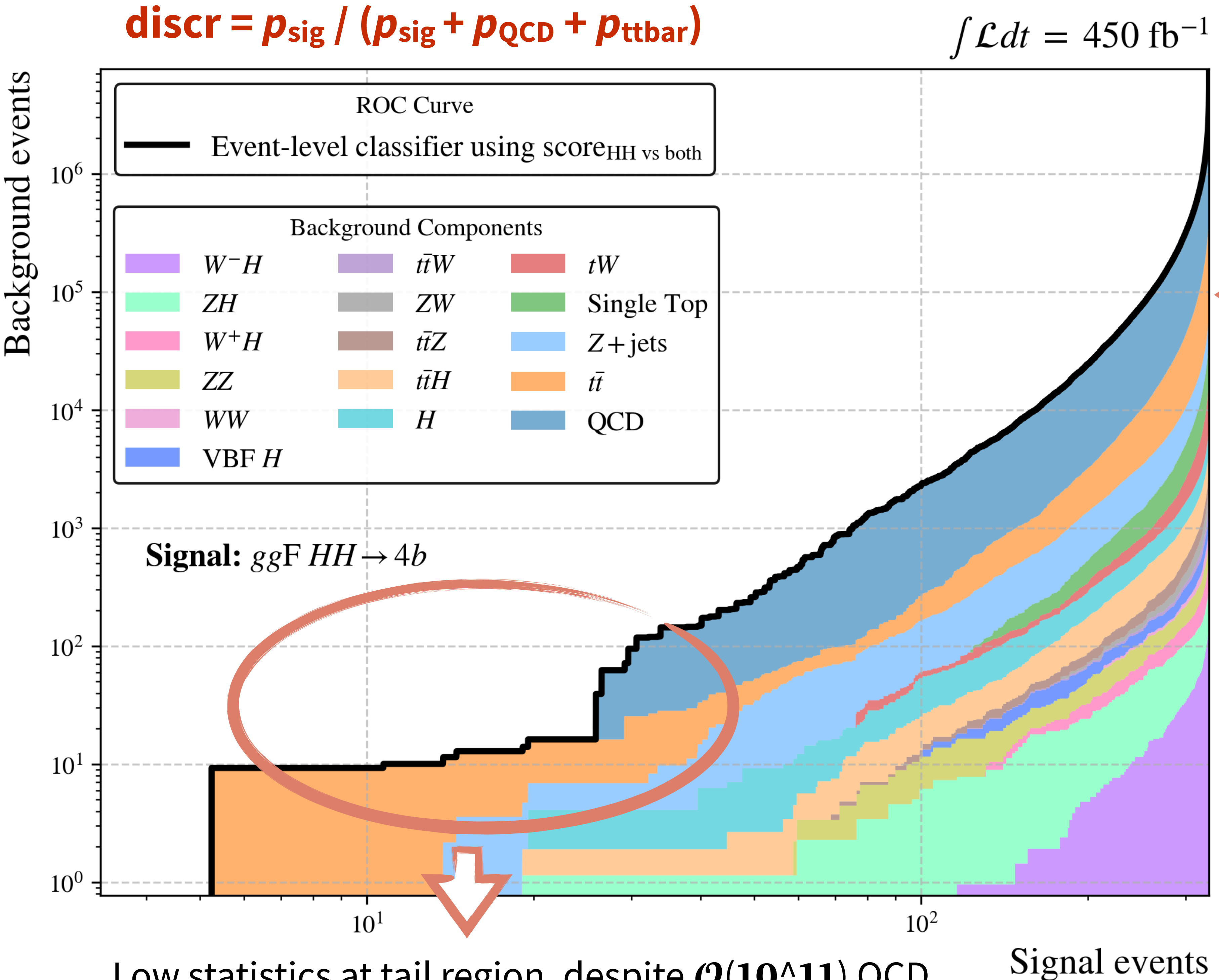


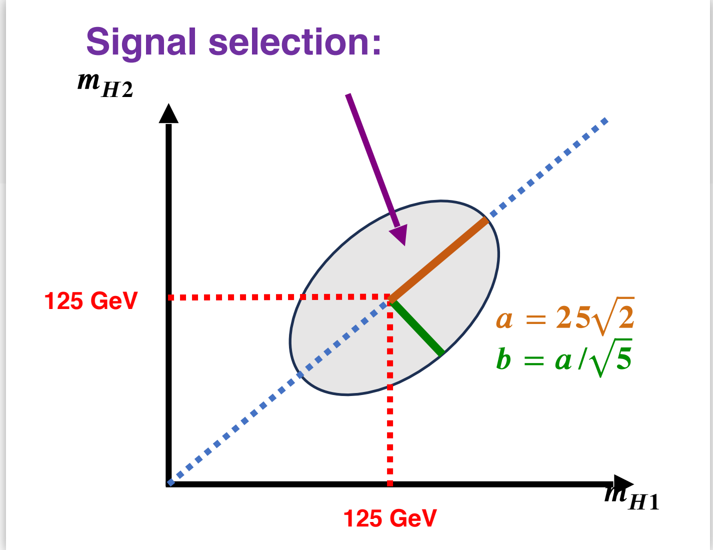


Select events in the HH mass window, and count the yields of each physics process

Effect of analysis selection

$\text{discr} = p_{\text{sig}} / (p_{\text{sig}} + p_{\text{QCD}} + p_{\text{ttbar}})$

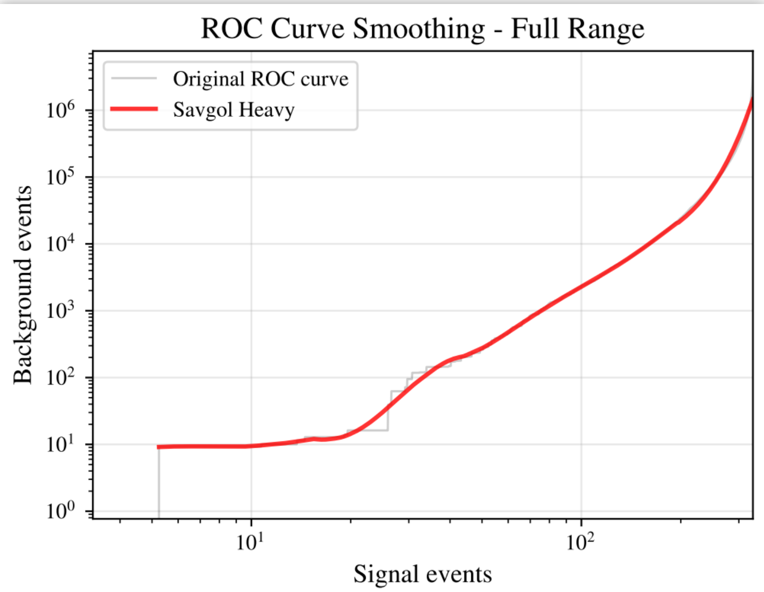




Select events in the HH mass window, and count the yields of each physics process

Low statistics at tail region, despite $\mathcal{O}(10^{11})$ QCD events before trigger selections are generated

smoothed ROC curve



Expected result

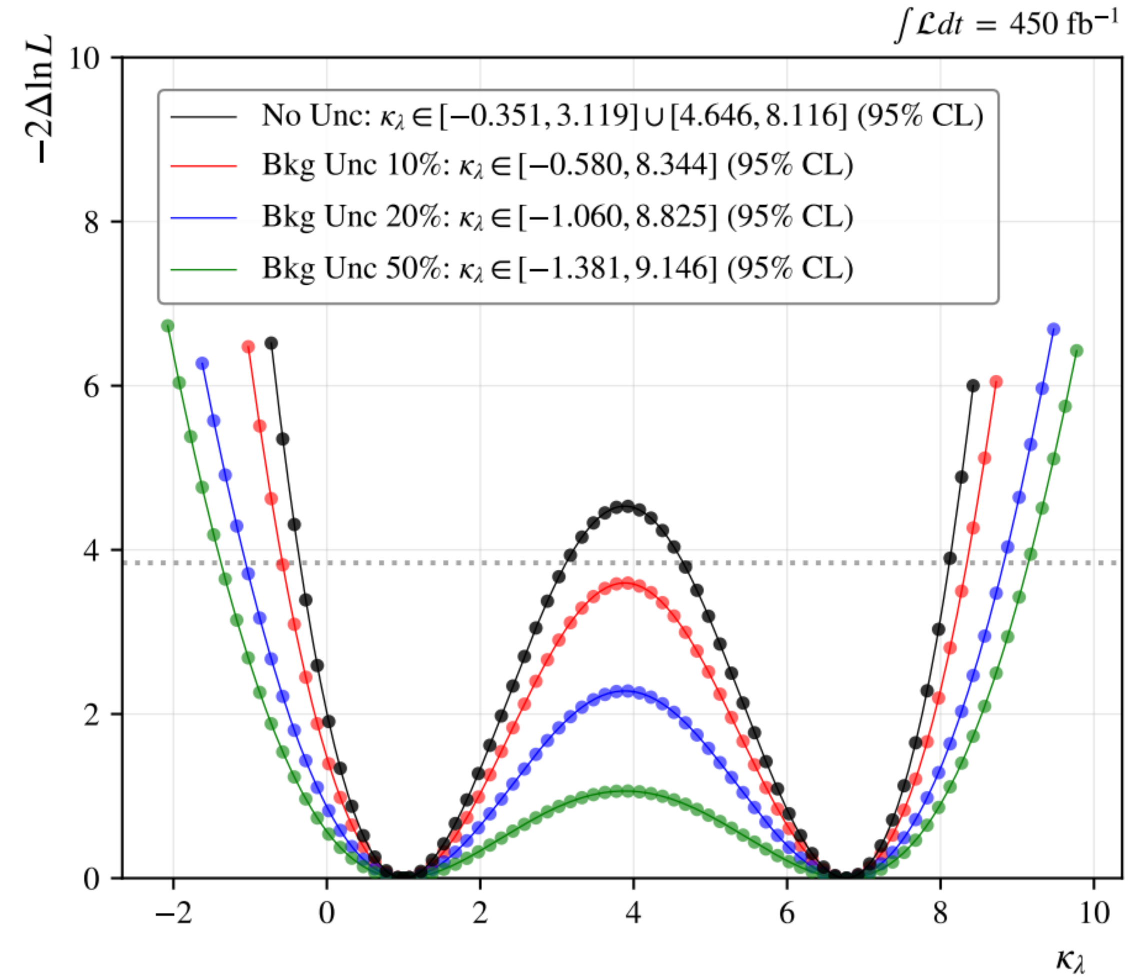
- The significance is calculated for each points at smoothed ROC:

$$Z_A = \sqrt{2\left\{(s+b)\ln\left[\frac{(s+b)(b+\sigma_b^2)}{b^2+(s+b)\sigma_b^2}\right] - \frac{b^2}{\sigma_b^2}\ln\left[1+\frac{\sigma_b^2 s}{b(b+\sigma_b^2)}\right]\right\}}$$

- Best significance at **S = 20.5 and B = 15.2**
- Different settings for $\boldsymbol{\sigma_b}$ are considered
  - 0%, 10%, 20%, 50%

Background (signal) uncertainties	SM significance (σ)	95% CL κ_λ range
0% (0%)	4.5	$[-0.35, 3.1] \cup [4.6, 8.1]$
10% (10%)	4.1	$[-0.58, 8.3]$
20% (20%)	3.3	$[-1.1, 8.8]$
50% (20%)	1.8	$[-1.4, 9.1]$

TABLE I. Summary of the expected SM di-Higgs statistical significance and 95% CL constraints on κ_λ under different signal and background uncertainty scenarios, for an integrated luminosity of 450 fb^{-1}.

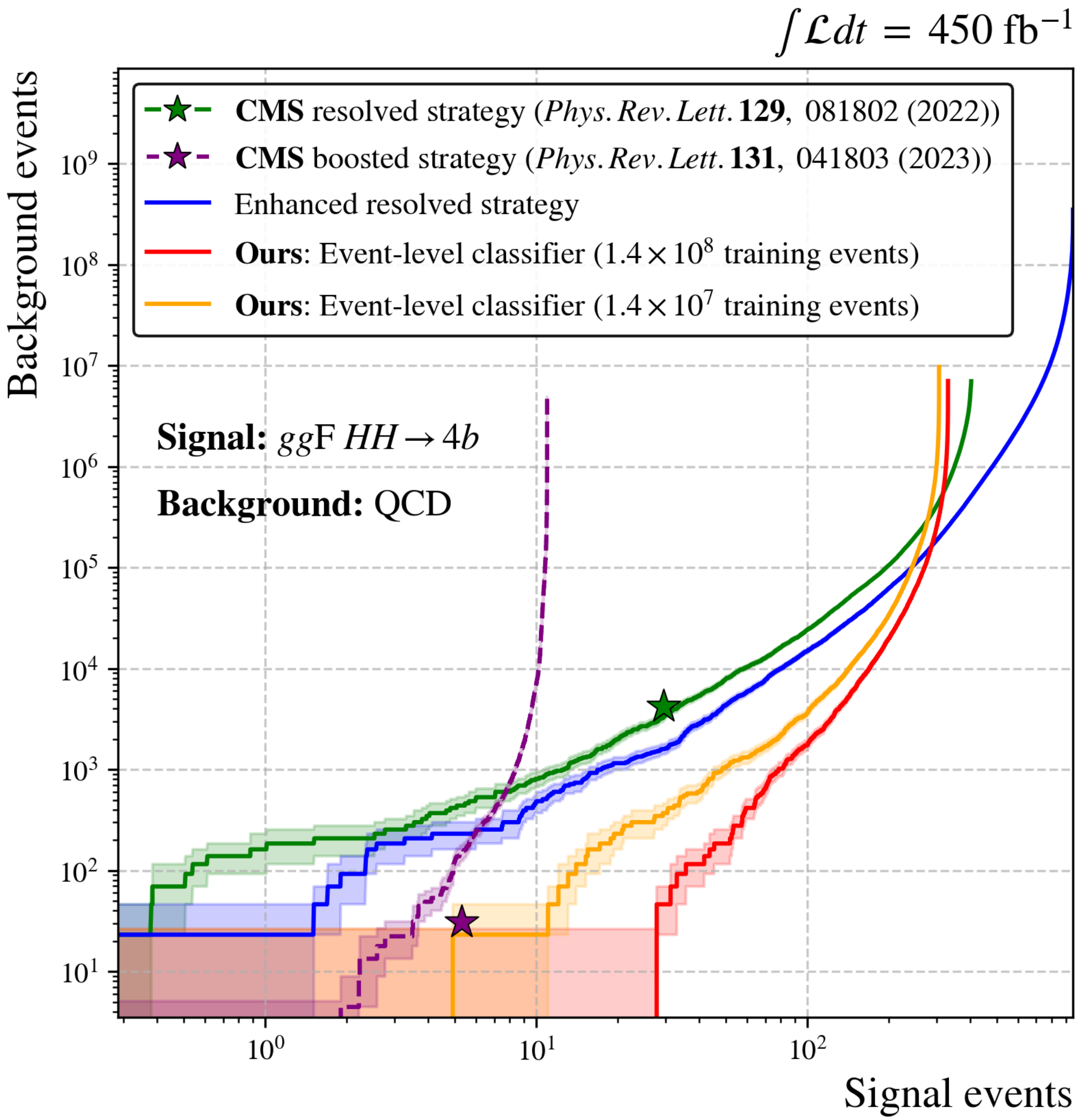


- >5x improvement in significance over existing approach!

Validity of the new method?

Further investigation into the classifier

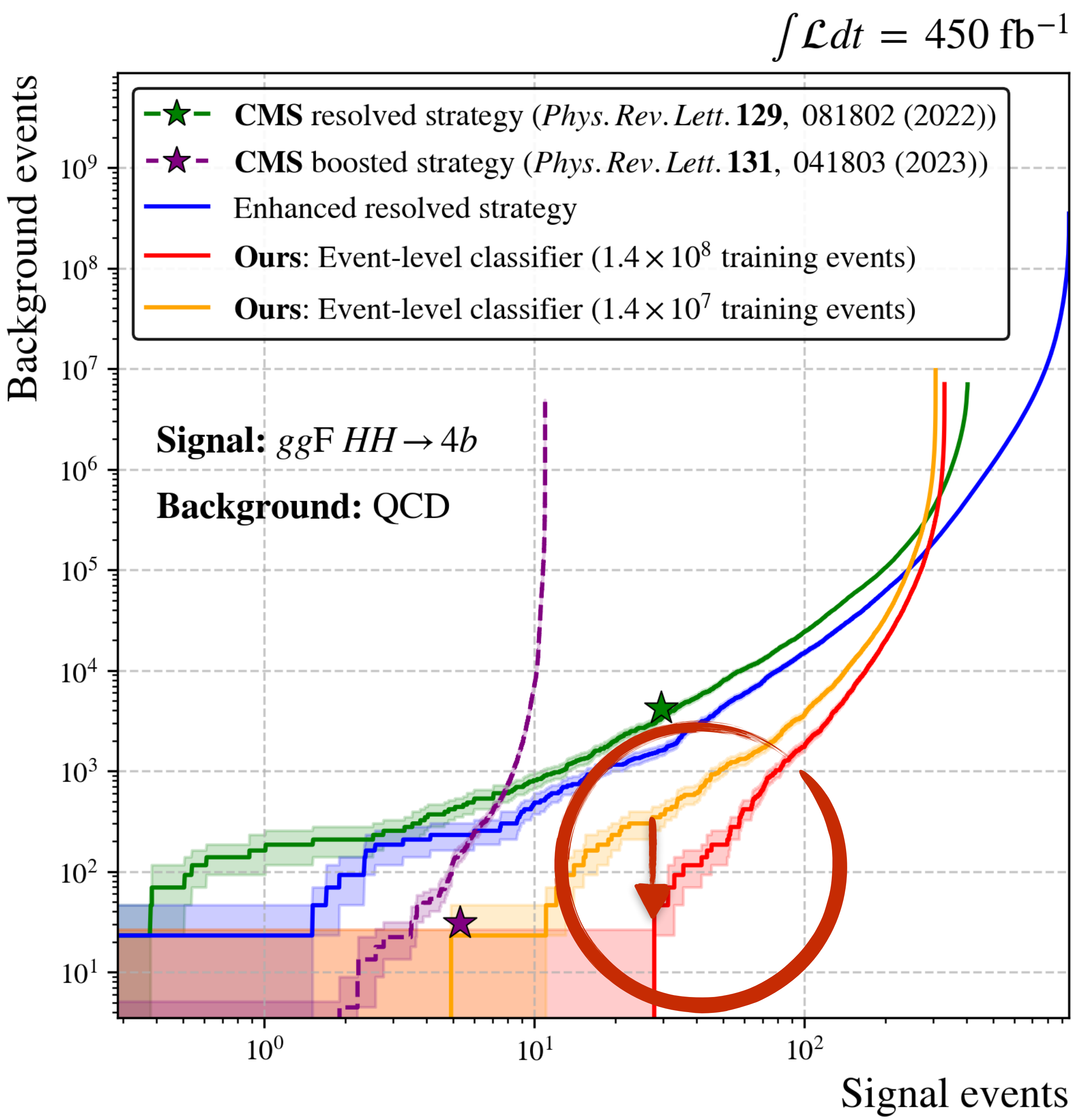
1 **What does it take to train a powerful classifier? Data size is important!**



- The nominal model (denoted SophonHH) is trained on the **full dataset of 140M events**
  - ParT with 10M parameters (4x default size)
- To study data dependence, we also trained on reduced subsets: **1/10 (14M)** of the full dataset.
- Number of training batches scale down accordingly to avoid overfitting (by 5x)

Further investigation into the classifier

1 **What does it take to train a powerful classifier?**
Data size is important!

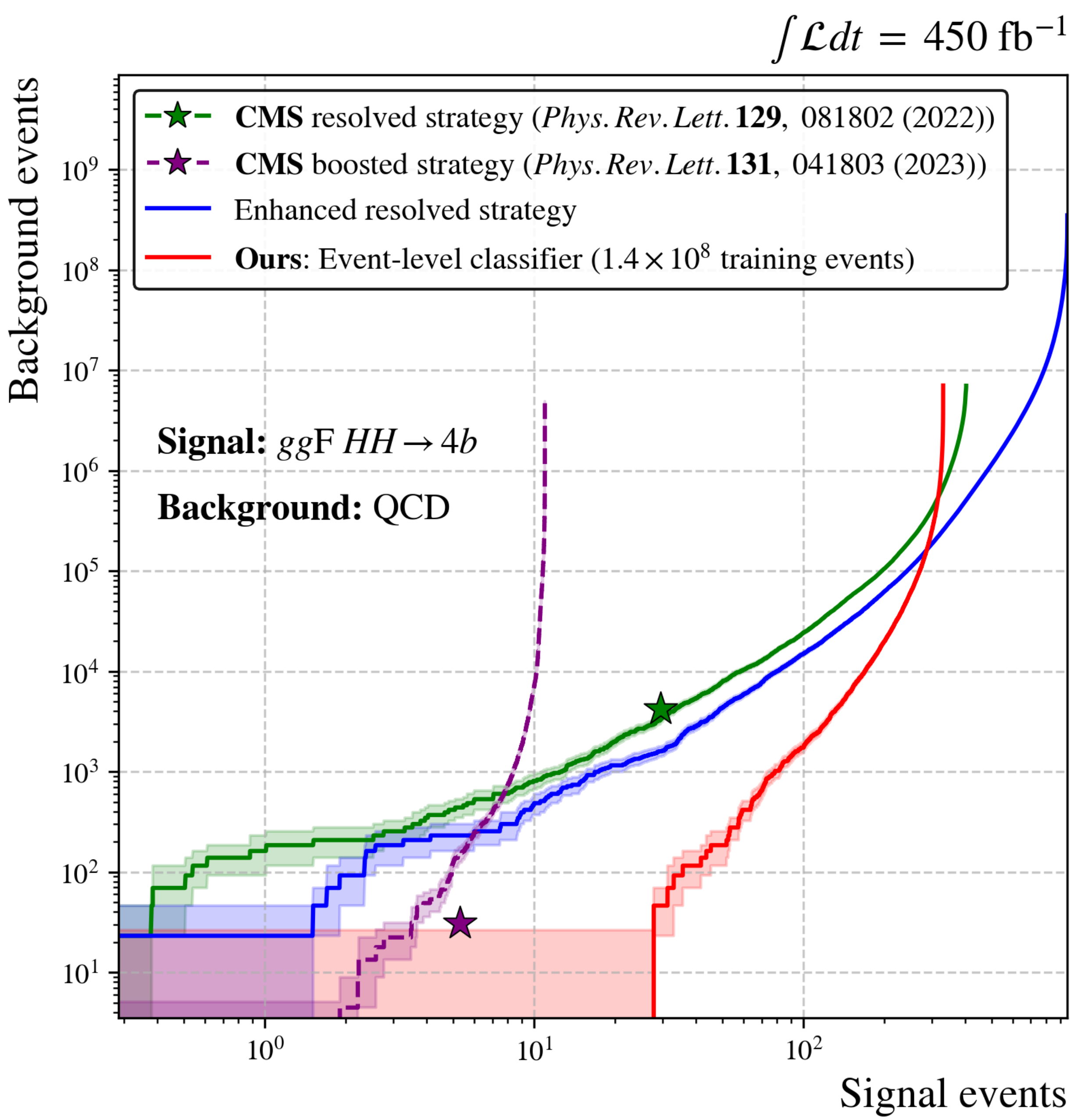


- ✦ The nominal model (denoted SophonHH) is trained on the **full dataset of 140M events**
  - ParT with 10M parameters (4x default size)
- ✦ To study data dependence, we also trained on reduced subsets: **1/10 (14M)** of the full dataset.
- ✦ Number of training batches scale down accordingly to avoid overfitting (by 5x)

Data size is crucial to the classifier's performance!
From 13M to 130M: bkg rejection improved by 4-8x

Further investigation into the classifier

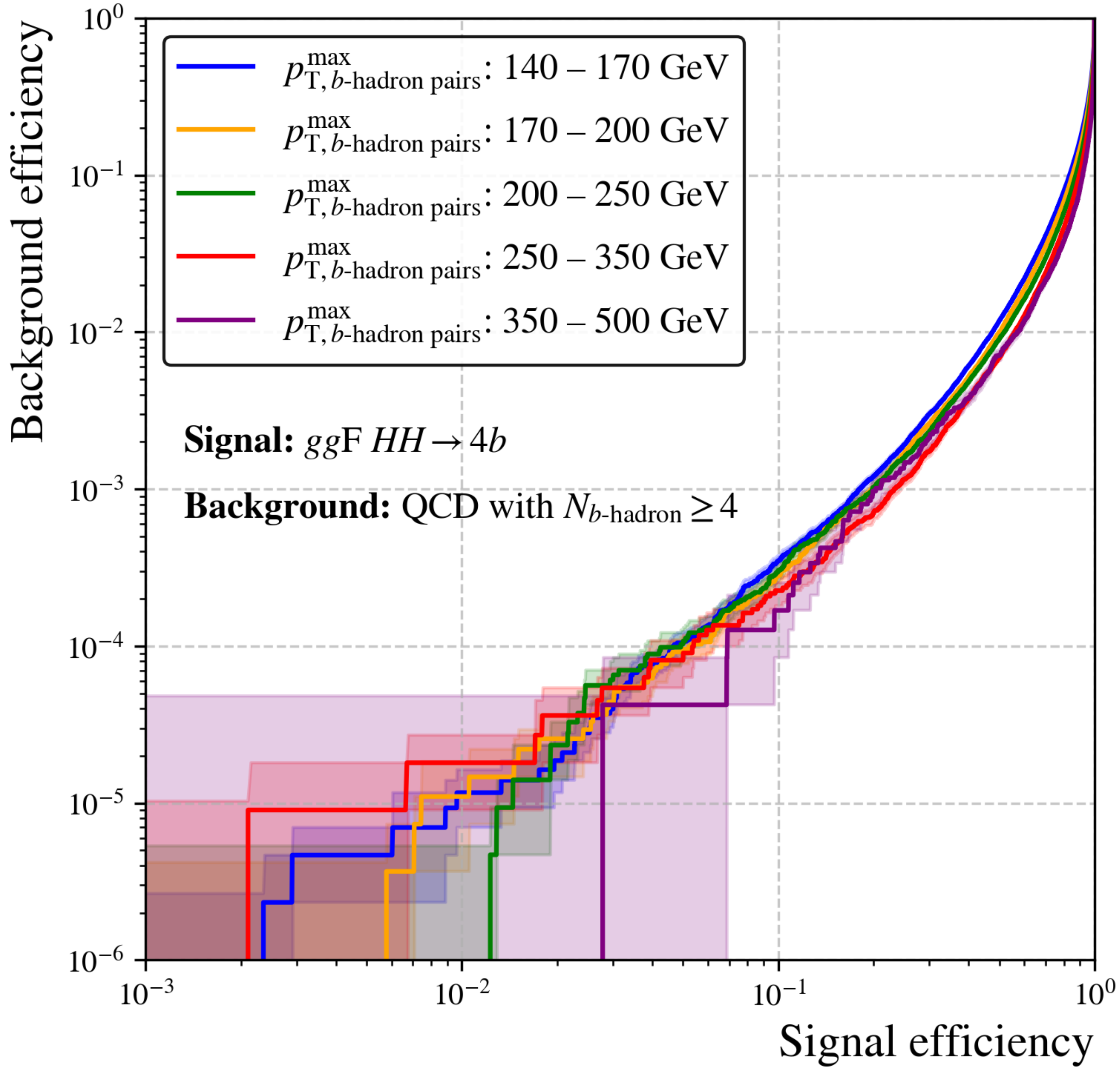
2 Is the QCD background truly irreducible? -No



- From signal vs QCD ROC curves, we observe that we can **suppress QCD bkgs to a non-negligible level,** even in the resolved regime.
- Using solely jet-based info, it's challenging to reach the same stage
- **Soft radiation pattern matters in vetoing QCD!**

Further investigation into the classifier

3 Should HH4b vs bkgs have a similar separation power across all Lorentz boost conditions? –Yes



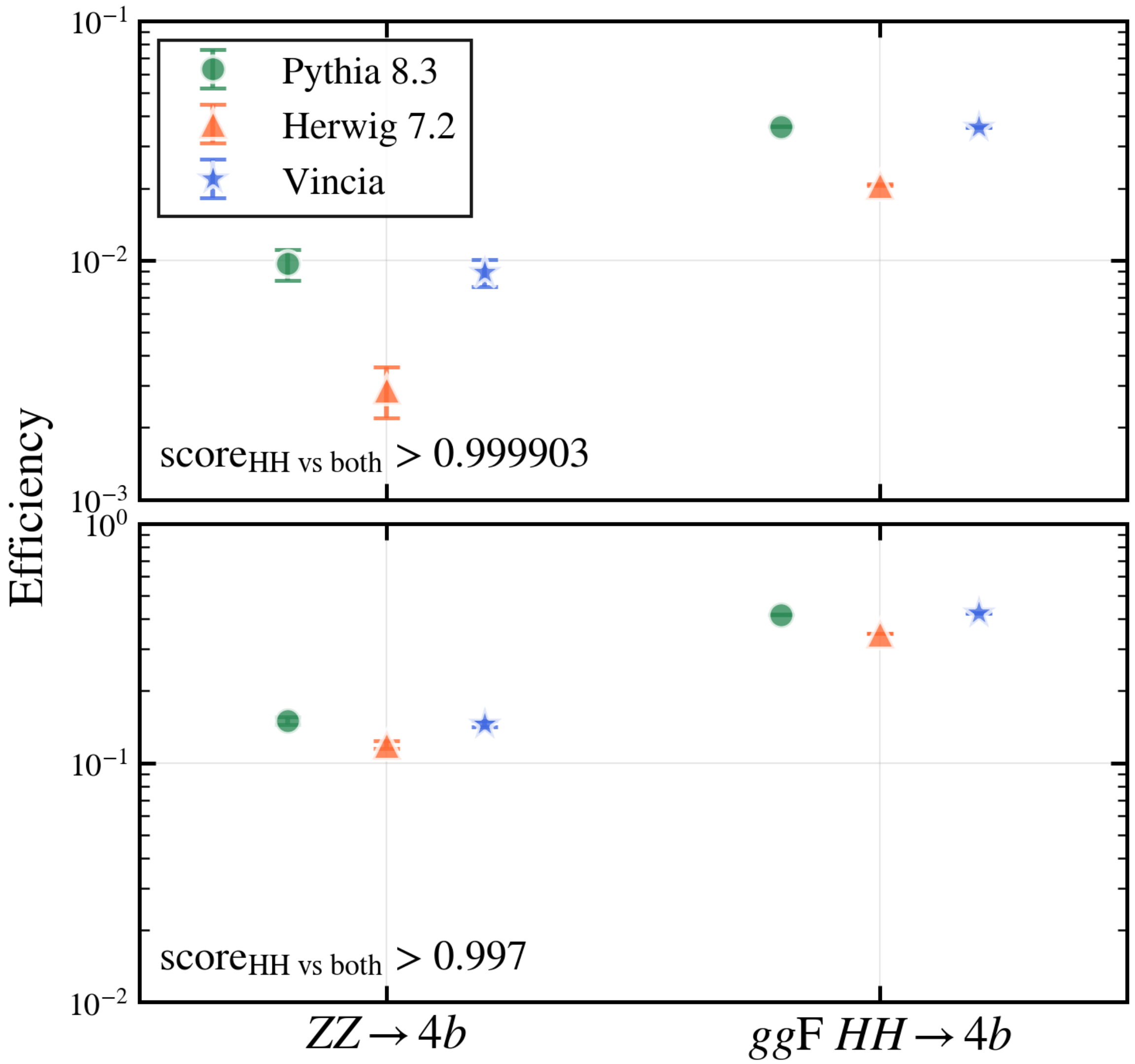
- Focus on ggHH signal vs QCD that contains exactly 4 gen-*b*-hadrons
- Use the **maximum p_T vector sum of all b-hadron pairs** as an indicator to characterize the event's Lorentz boost
- Result: **separation ability is consistent across different boosting conditions**!

Further investigation into the classifier

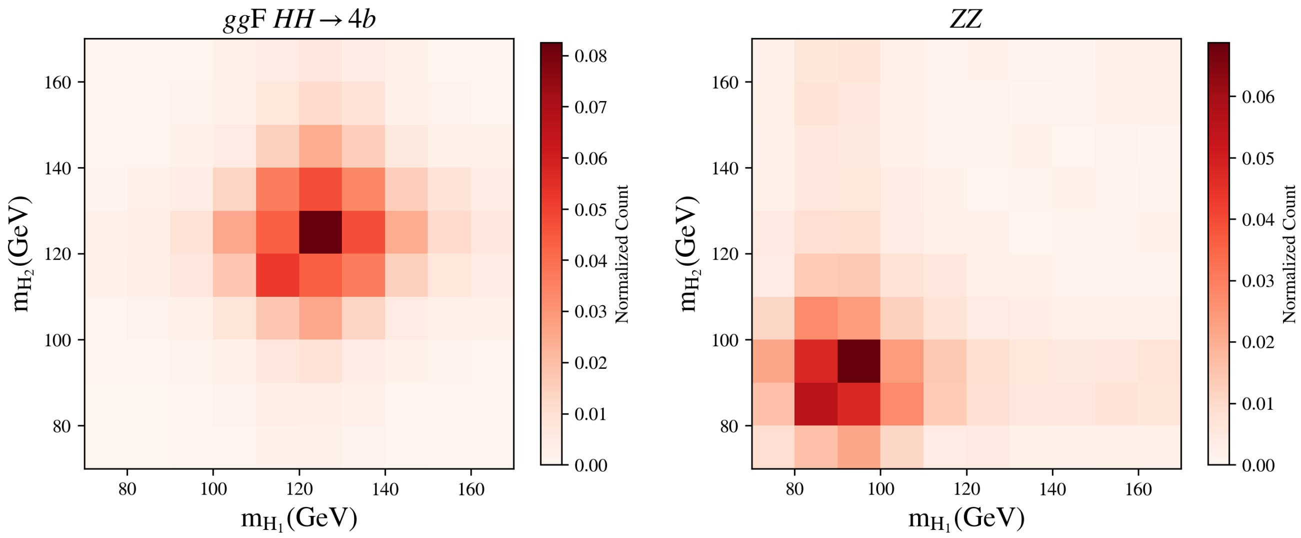
4

Robustness across hadronization models, ZZ vs HH

Comparing the efficiencies of Pythia 8.3, Herwig 7.2 and Vincia



- Tighter classifier working points enlarge discrepancies, most notably between Herwig 7.2 and the others
- No significant efficiency loss is observed
- **Similar behavior between and HH vs ZZ** (allowing to use ZZ→4b as a "calibration candle")



Summary & outlook

arXiv:2508.15048

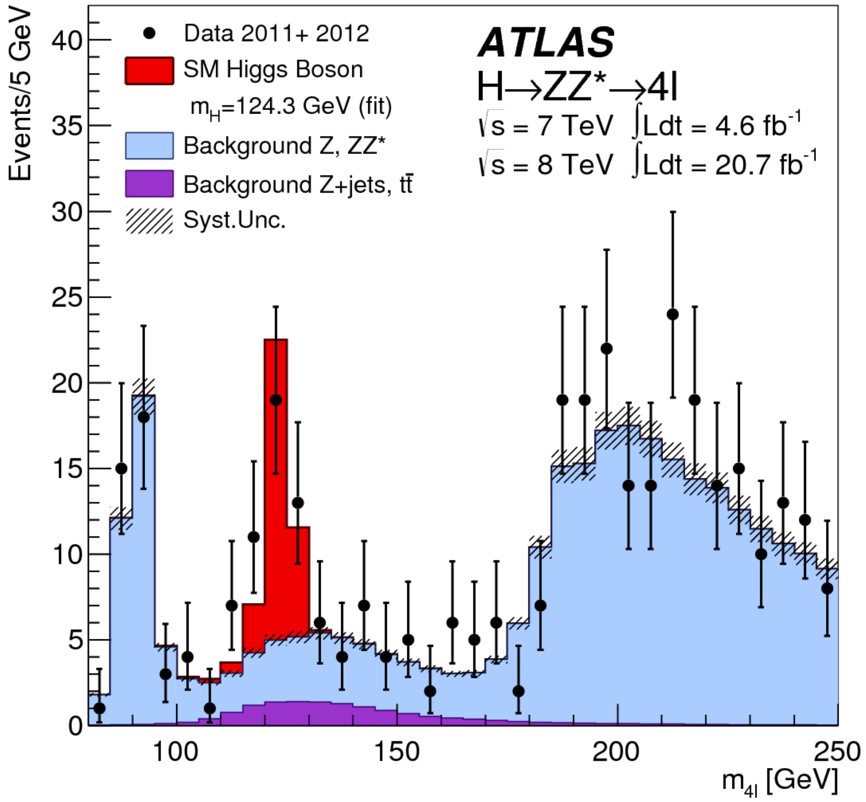
- ➔ Take home:
  - ❖ We show that CMS/ATLAS can achieve >5× sensitivity improvement for di-Higgs in the *HH*→4*b* final state, with potential to observe HH even before the HL-LHC
  - ❖ This gain is driven entirely by AI engineering
    - ‣ “scaling up” improves performance; training directly on low-level objects
  - ❖ also validated the method’s reliability
- ➔ These findings challenge the conventional understanding of di-Higgs!
- ➔ Aim to deliver this search within the next three years

Summary & outlook

arXiv:2508.15048

- → Take home:
  - ❖ We show that CMS/ATLAS can achieve >5× sensitivity improvement for di-Higgs in the *HH→4b* final state, with potential to observe HH even before the HL-LHC
  - ❖ This gain is driven entirely by AI engineering
    - “scaling up” improves performance; training directly on low-level objects
  - ❖ also validated the method’s reliability
- → These findings challenge the conventional understanding of di-Higgs!
- → Aim to deliver this search within the next three years

A conceptual sketch
- We are using advanced AI to make the fully hadronic channel “clean”
- As clean as H→4ℓ observation 🙂



Backup

Interlude: Our journey toward a "more realistic Delphes"

Interlude: Our journey toward a "more realistic Delphes"

H. Qu, CL, S. Qian, ICML 2022*

February '22

JetClass [data card]

Realistic particle-level features:
(+impact params d0, dz, d0_err, dz_err)

⇒ enable realistic flavour-tagging

CL et al. arXiv:2405.12972*

May '24

JetClass-II [data card]

emulate **pileup** ($\langle\mu\rangle$=50) &
pileup mitigation with PUPPI

Interlude: Our journey toward a "more realistic Delphes"

H. Qu, CL, S. Qian, ICML 2022*

February '22

JetClass [data card]

Realistic particle-level features:
(+impact params d0, dz, d0_err, dz_err)

⇒ enable realistic flavour-tagging

CL et al. arXiv:2405.12972*

May '24

JetClass-II [data card]

emulate **pileup** (<μ>=50) &
pileup mitigation with PUPPI

Y. Zhao, CL et al. arXiv:2503.00118*

March '25

JetClass-II
+ realistic small/large-R jet tagger (anti-k_T R=0.4/0.8)

[SophonAK4], [Sophon (AK8)**]**

Interlude: Our journey toward a "more realistic Delphes"

H. Qu, CL, S. Qian, ICML 2022*

CL et al. arXiv:2405.12972*

Y. Zhao, CL et al. arXiv:2503.00118*

February '22

JetClass [data card]

Realistic particle-level features:
(+impact params d0, dz, d0_err, dz_err)

⇒ enable realistic flavour-tagging

May '24

JetClass-II [data card]

emulate **pileup** (<μ>=50) &
pileup mitigation with PUPPI

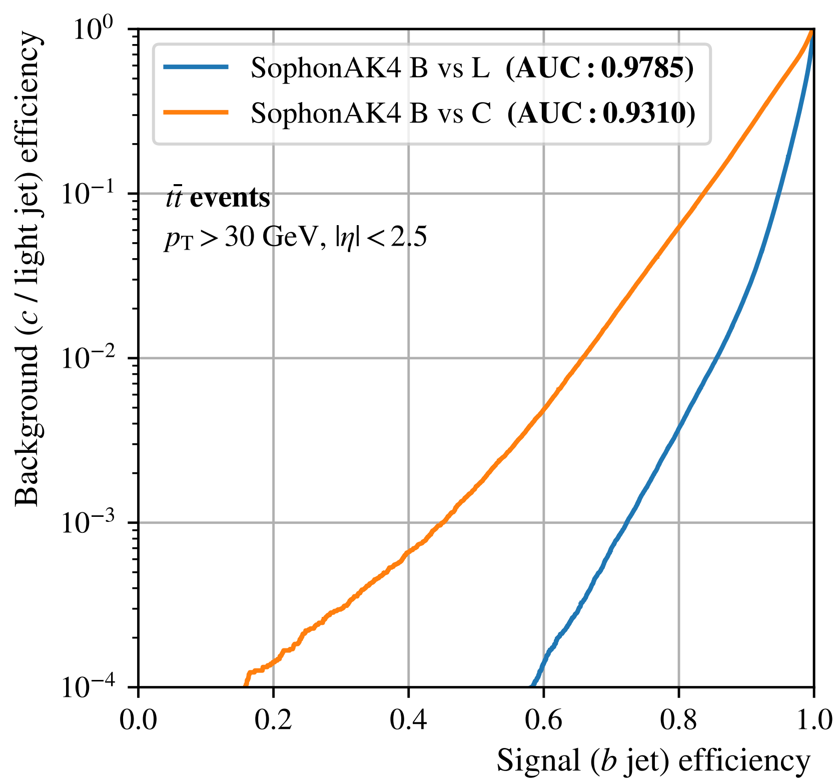
March '25

JetClass-II
+ realistic small/large-R jet tagger (anti-k_T R=0.4/0.8)

[SophonAK4], [Sophon (AK8)**]**

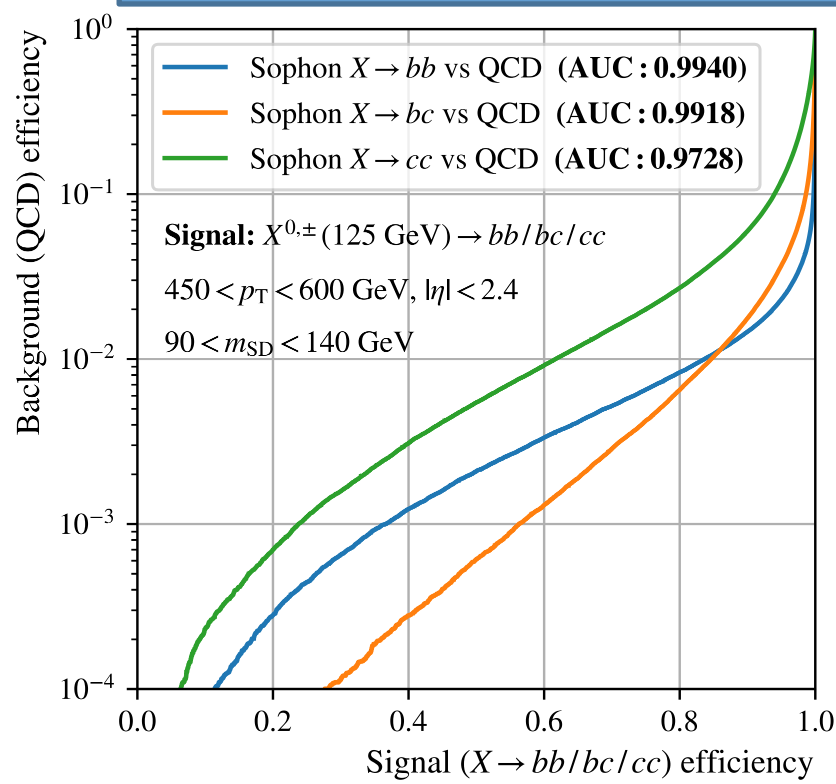
arXiv:2503.00118

Performance of b vs light/c tagging



arXiv:2503.00118

Performance of X→bb vs QCD tagging



Interlude: Our journey toward a "more realistic Delphes"

H. Qu, CL, S. Qian, ICML 2022* — **February '22**

JetClass [data card]

Realistic particle-level features:
(+impact params d0, dz, d0_err, dz_err)
⇒ enable realistic flavour-tagging

CL et al. arXiv:2405.12972* — **May '24**

JetClass-II [data card]

emulate **pileup** (<μ>=50) &
pileup mitigation with PUPPI

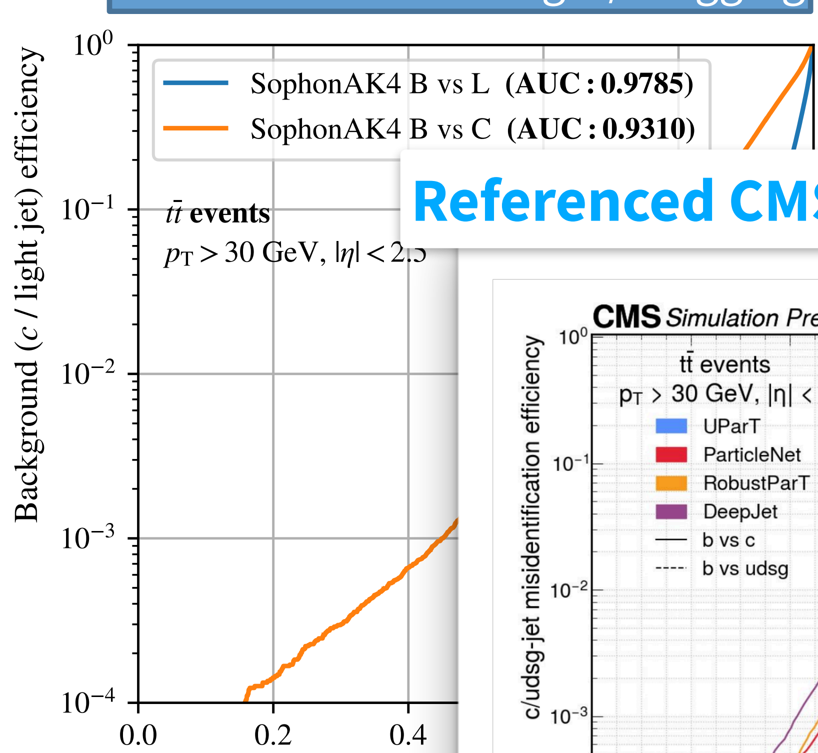
Y. Zhao, CL et al. arXiv:2503.00118* — **March '25**

JetClass-II
+ realistic small/large-R jet tagger (anti-k_T R=0.4/0.8)

[SophonAK4], [Sophon (AK8)**]**

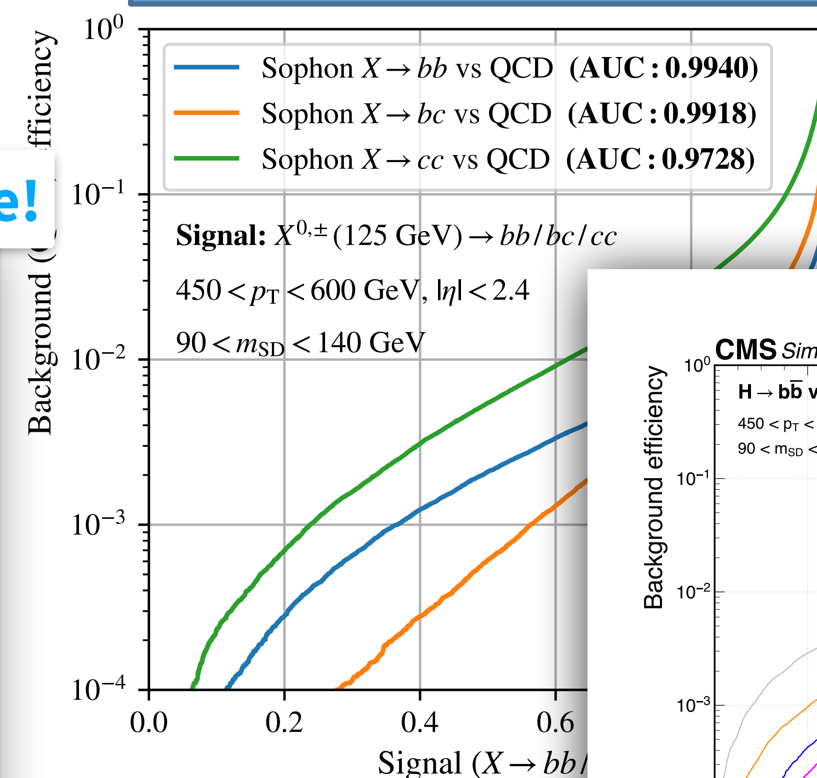
arXiv:2503.00118

Performance of b vs light/c tagging



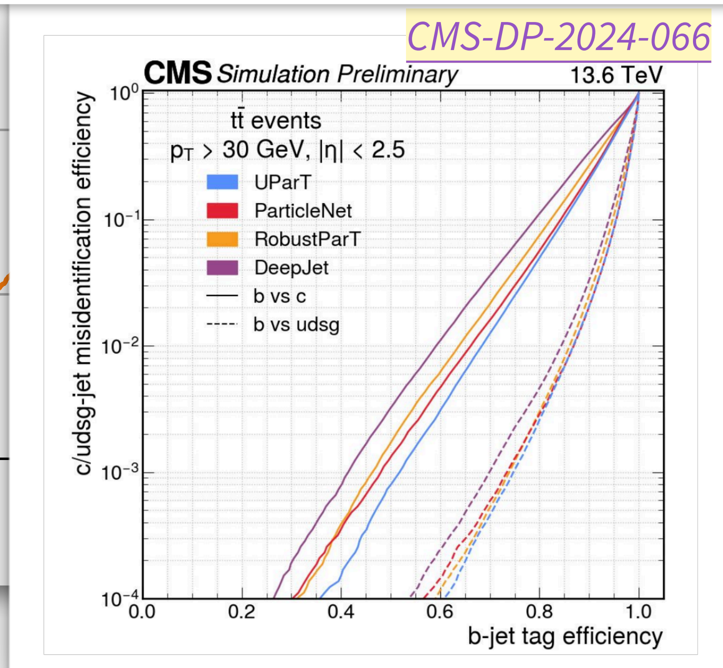
arXiv:2503.00118

Performance of X→bb vs QCD tagging

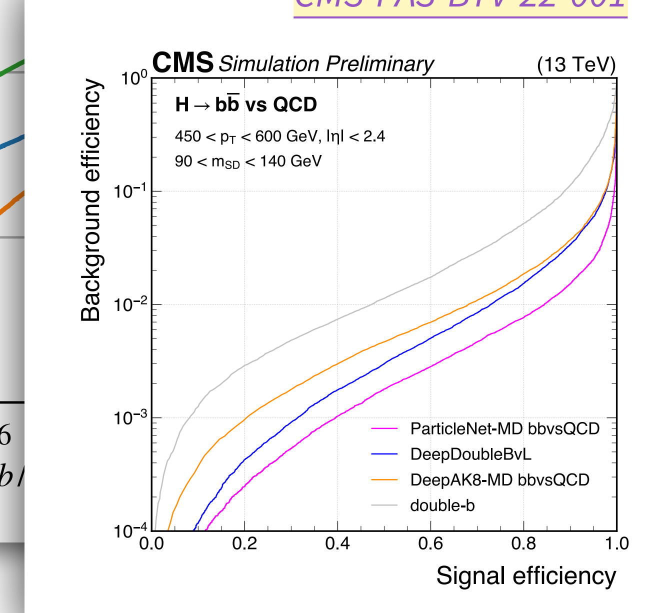


Referenced CMS performance!

CMS-DP-2024-066



CMS-PAS-BTV-22-001



Interlude: Our journey toward a "more realistic Delphes"

H. Qu, CL, S. Qian, ICML 2022* — *CL* et al. arXiv:2405.12972* — *Y. Zhao, CL* et al. arXiv:2503.00118*

February '22

JetClass [data card]

Realistic particle-level features:
(+impact params d0, dz, d0_err, dz_err)

⇒ enable realistic flavour-tagging

May '24

JetClass-II [data card]

emulate **pileup** (<μ>=50) &
pileup mitigation with PUPPI

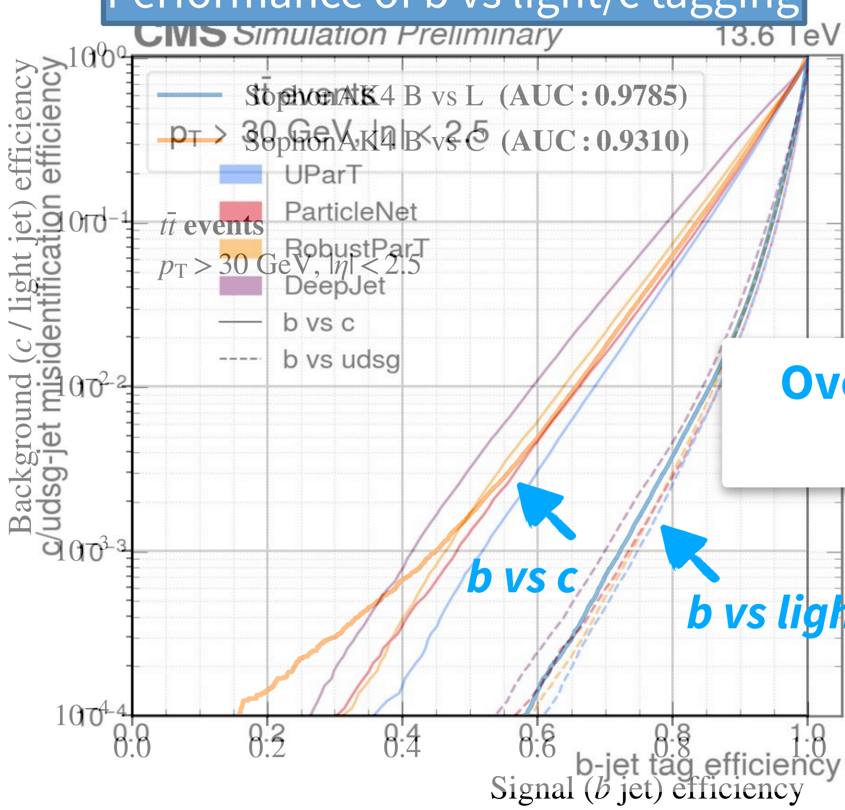
March '25

JetClass-II
+ realistic small/large-R jet tagger (anti-k_T R=0.4/0.8)

[SophonAK4], [Sophon (AK8)]

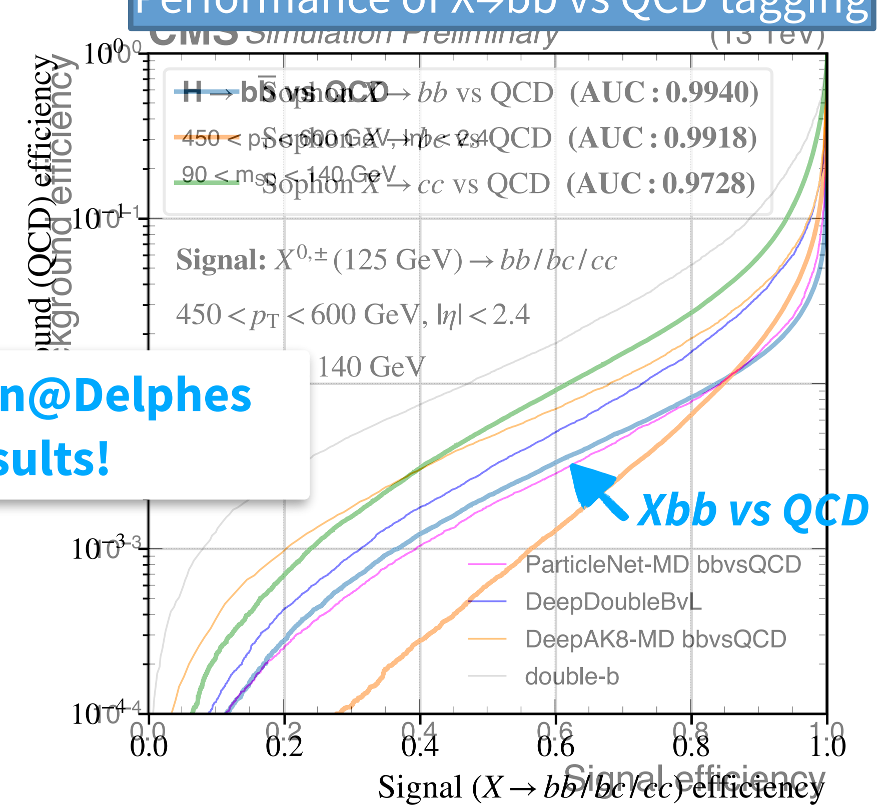
arXiv:2503.00118

Performance of b vs light/c tagging



arXiv:2503.00118

Performance of X→bb vs QCD tagging



Overlaying Sophon@Delphes and CMS results!

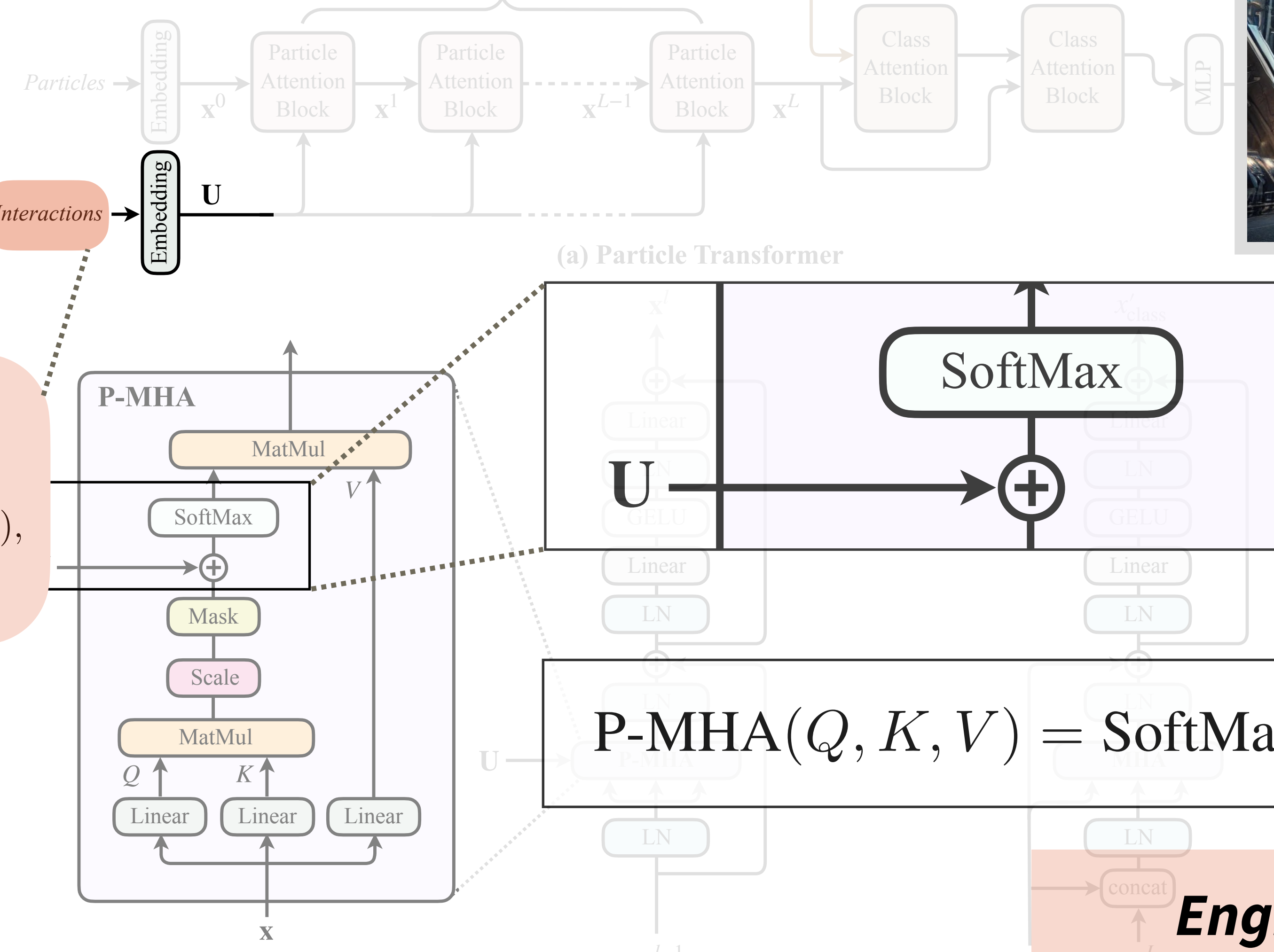
Sophon/SophonAK4 is really close to CMS's state-of-the-art taggers

Particle Transformer

➔ Transformer tailored for particle physics

❖ featuring its "attention bias" that **embed pairwise features respecting different levels of Lorentz symmetry**

H. Qu, CL, S. Qian. ICML 2022



$$\Delta = \sqrt{(y_a - y_b)^2 + (\phi_a - \phi_b)^2},$$
$$k_T = \min(p_{T,a}, p_{T,b})\Delta,$$
$$z = \min(p_{T,a}, p_{T,b})/(p_{T,a} + p_{T,b}),$$
$$m^2 = (E_a + E_b)^2 - \|\mathbf{p}_a + \mathbf{p}_b\|^2,$$

$$\text{P-MHA}(Q, K, V) = \text{SoftMax}(QK^T/\sqrt{d_k} + \mathbf{U})V,$$

- Now a new standard in the HEP field

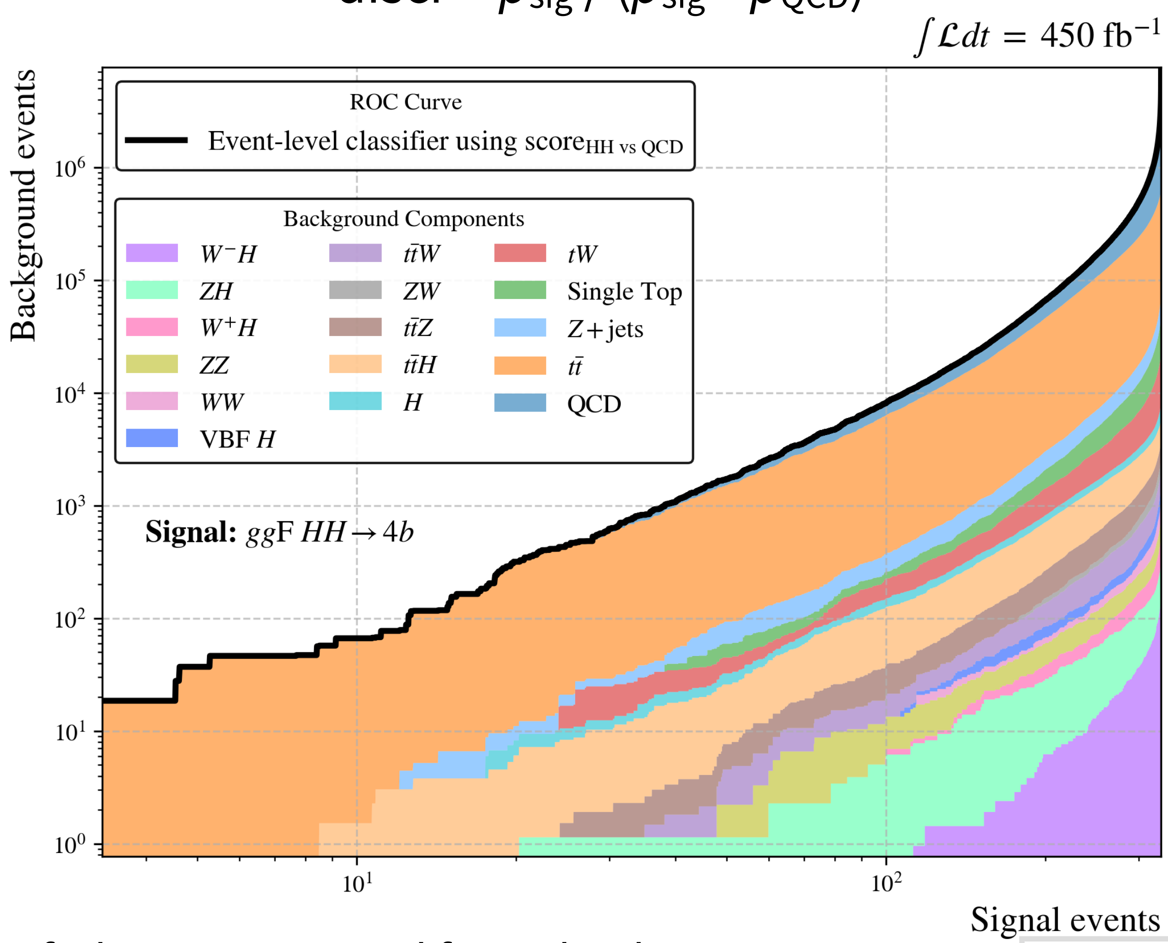
Engineering blueprint

Plain Transformer + Inductive bias from "jet symmetries"

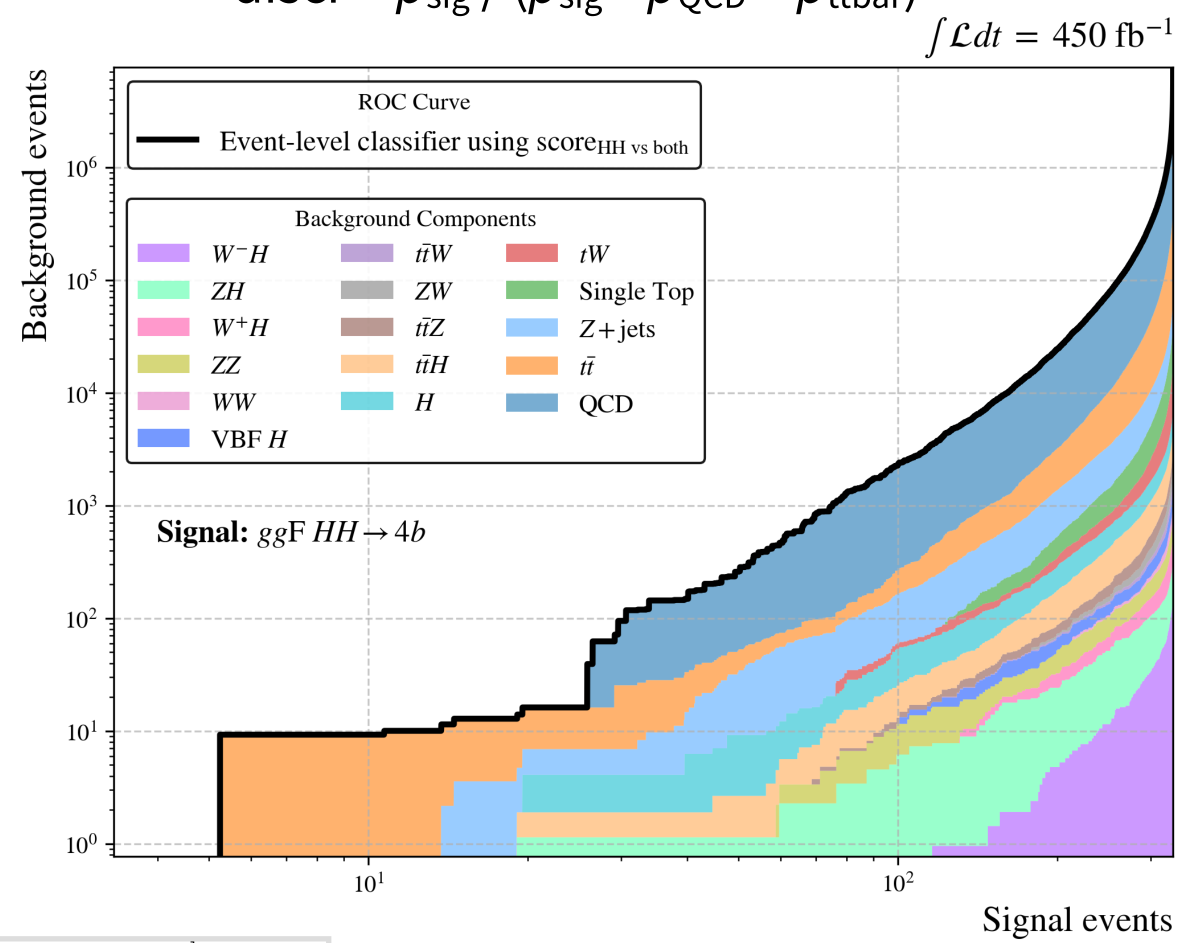
Further investigation into the classifier

Which background dominates? QCD or ttbar?

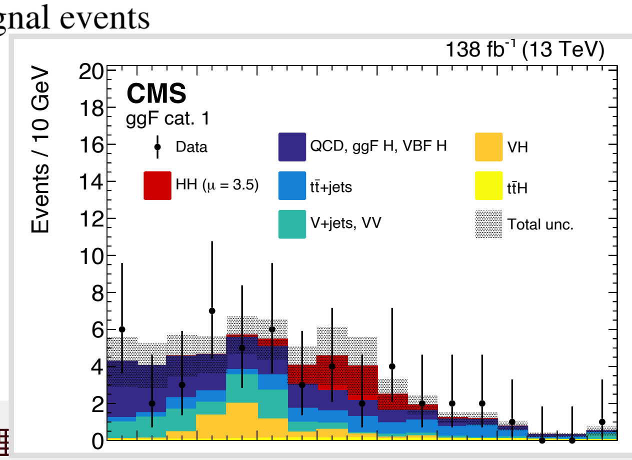
discr = p_{sig} / (p_{sig} + p_{QCD})



discr = p_{sig} / (p_{sig} + p_{QCD} + p_{ttbar})



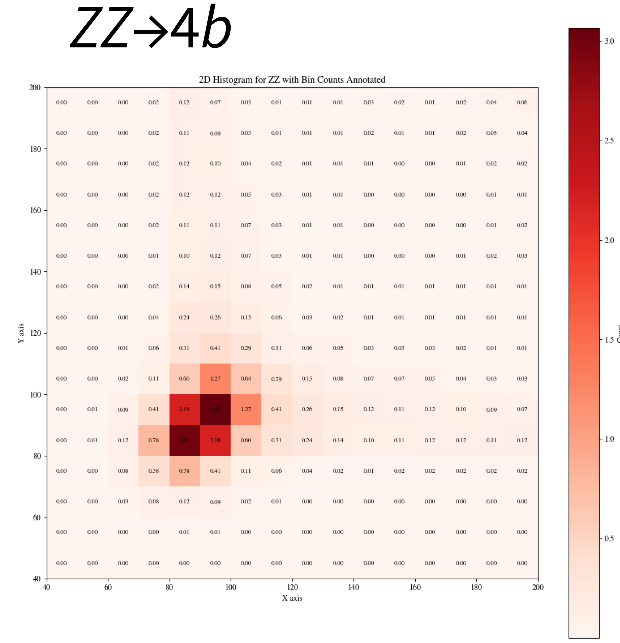
- If ttbar is not vetoed from the discriminant definition, it becomes dominant
- Suggest that boosted-channel search can be further improved by incorporating a top-like jet veto (currently using a XbbvsQCD discriminant)



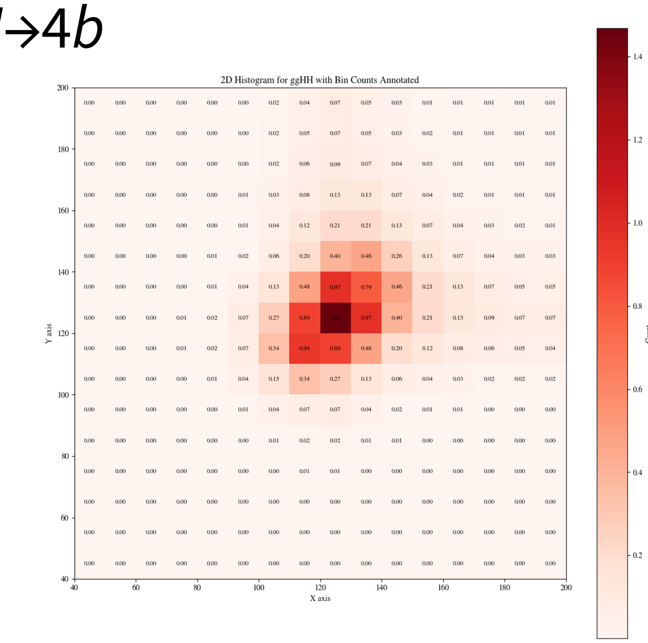
Calibration strategy

We assume their MC-to-data scale factors, peak shifts, and smearing effects are the same

ZZ→4b



HH→4b



A valid strategy

ZZ→4b extraction
(with SF, peak shifts and smearing obtained from the fit)

⇔

HH→4b extraction
(using shared factors)

An even safer strategy

extracting fake "ZZ→4b" obtained from event mixing
(with SF, peak shifts and smearing obtained from the fit in a different phase space)

⇔

ZZ→4b extraction
(using shared factors)

⇔

HH→4b extraction
(using shared factors)

Validation region

Sample production

- ➔ QCD:
  - ❖ Pythia-generated
  - ❖ for resolved studies: only kept events passing the resolved "4j3b" trigger → enabled GEN-level filters to improve event-generation efficiency: GEN H_T and GEN jet p_T cuts (looser than reco-level cuts to avoid over-removing phase space)
  - ❖ use the Pythia calculated cross-section
- ➔ Z(qq)+jets
  - ❖ generated by MadGraph LO with the addition +0/1/2 jets and MLM matching with partons. Z→qq decay implemented in MadGraph
  - ❖ for resolved studies: only kept events passing the resolved "4j3b" trigger → use matrix-element-level cuts, on $p_{T,j1/2/3}$ and H_T
  - ❖ use the MadGraph calculated cross-section
- ➔ ttbar, single-top, tW, $t\bar{t}W$, $t\bar{t}Z$, WW, WZ, ZZ
  - ❖ generated by MadGraph LO with the addition +0/1/2 jets, or +0/1 jet, or 0 jet (depending on the processes)
  - ❖ consider inclusive "boson -> fermion fermion" decay modes
  - ❖ use the highest-order cross sections recommended
- ➔ ggF Higgs, VBF Higgs, VH, $t\bar{t}H$ & ggHH signal
  - ❖ generated by Powheg at NLO
  - ❖ consider inclusive decay
  - ❖ use the Powheg cross sections

TeV 物理前沿专题研讨会暨第 31 届 LHC Mini-Workshop