Understanding the Performance Gains of a Full-Event HH→4b Search

Congqiao Li, Qiang Li, **Yipin Wang**, Tianyi Yang, Chen Zhou (PKU)
Huilin Qu (SJTU), Javier Duarte (UCSD), Cristina Mantilla Suárez (UVA)

Step 14, 2026

Motivation

Higgs discovered in 2012 — self-interaction still unexplored

$$V(h) = V_0 + \frac{1}{2}\boxed{m_h^2}h^2 + \boxed{\kappa_\lambda \lambda_3^{SM}}h^3 + \frac{\boxed{\lambda_4^{SM}}}{4}h^4$$

HHH production

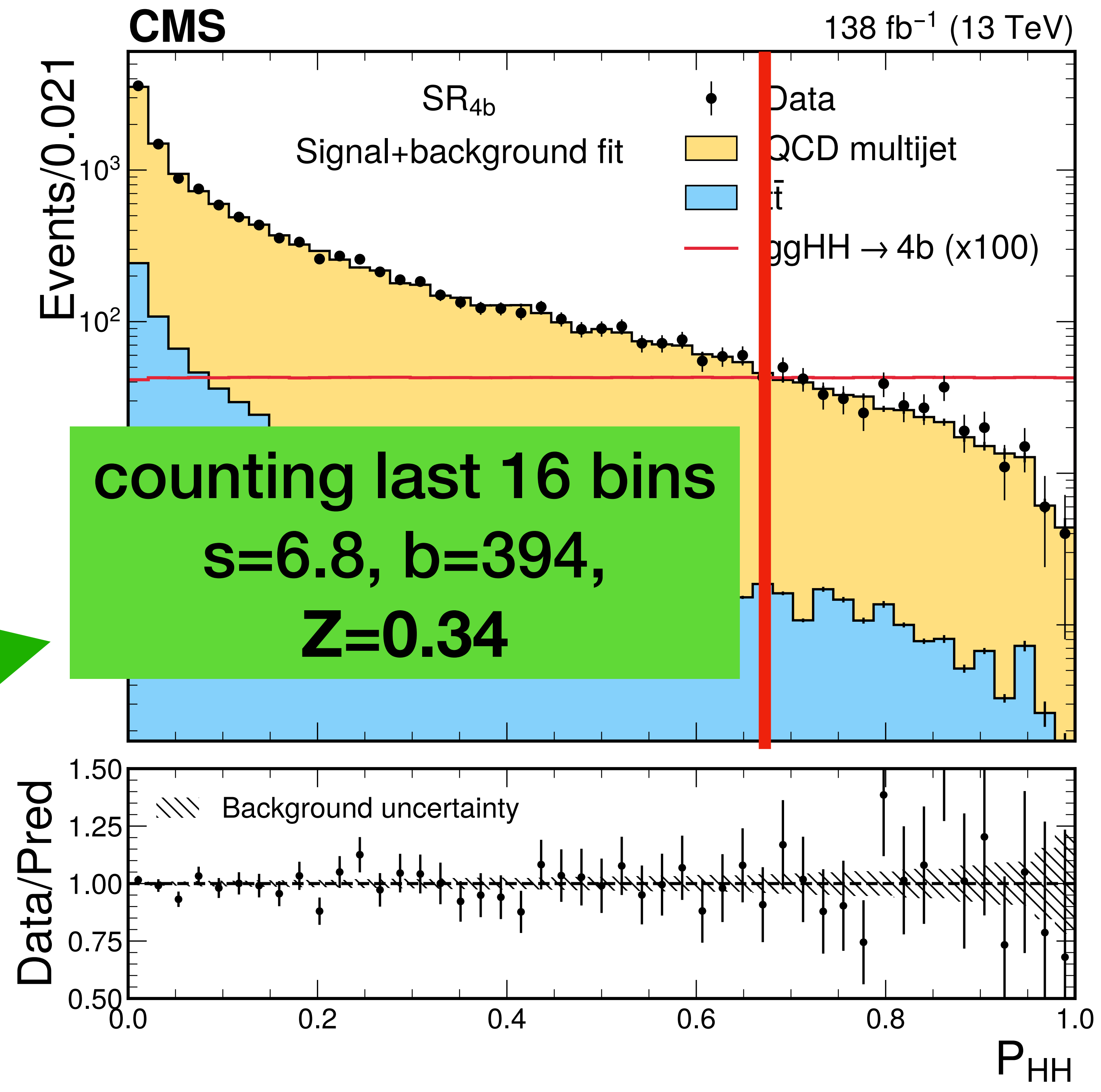
Higgs mass

HH production, $\kappa_\lambda = \lambda_3/\lambda_3^{SM}$

$$\sigma(ggHH) = \sigma_{SM}(A\kappa_\lambda^2 + B\kappa_\lambda + C)$$

HH production probes the Higgs self-coupling

CMS HH4b Run2 Subset



Current LHC Sensitivity

Run2 ATLAS and CMS HH Search Combination (arXiv:2602.23991):

- Observed significance: 1.1σ
- 95% CL: –0.71 < κλ < 6.1

CMS HH → 4b Analysis (arXiv:2604.27044):

- Run2 subset: simple event-counting estimate gives Z≈0.34

Current HH→4b sensitivity remains limited
Can more powerful ML methods do better?

yipin.wang@cern.ch

Full-Event HH→4b Search

ML4Jets 2025: A new phenomenological projection for HH→4b using an event-level Transformer (Yang & Li (PKU); Indico; arxiv: 2508.15048):

Conventional approach:

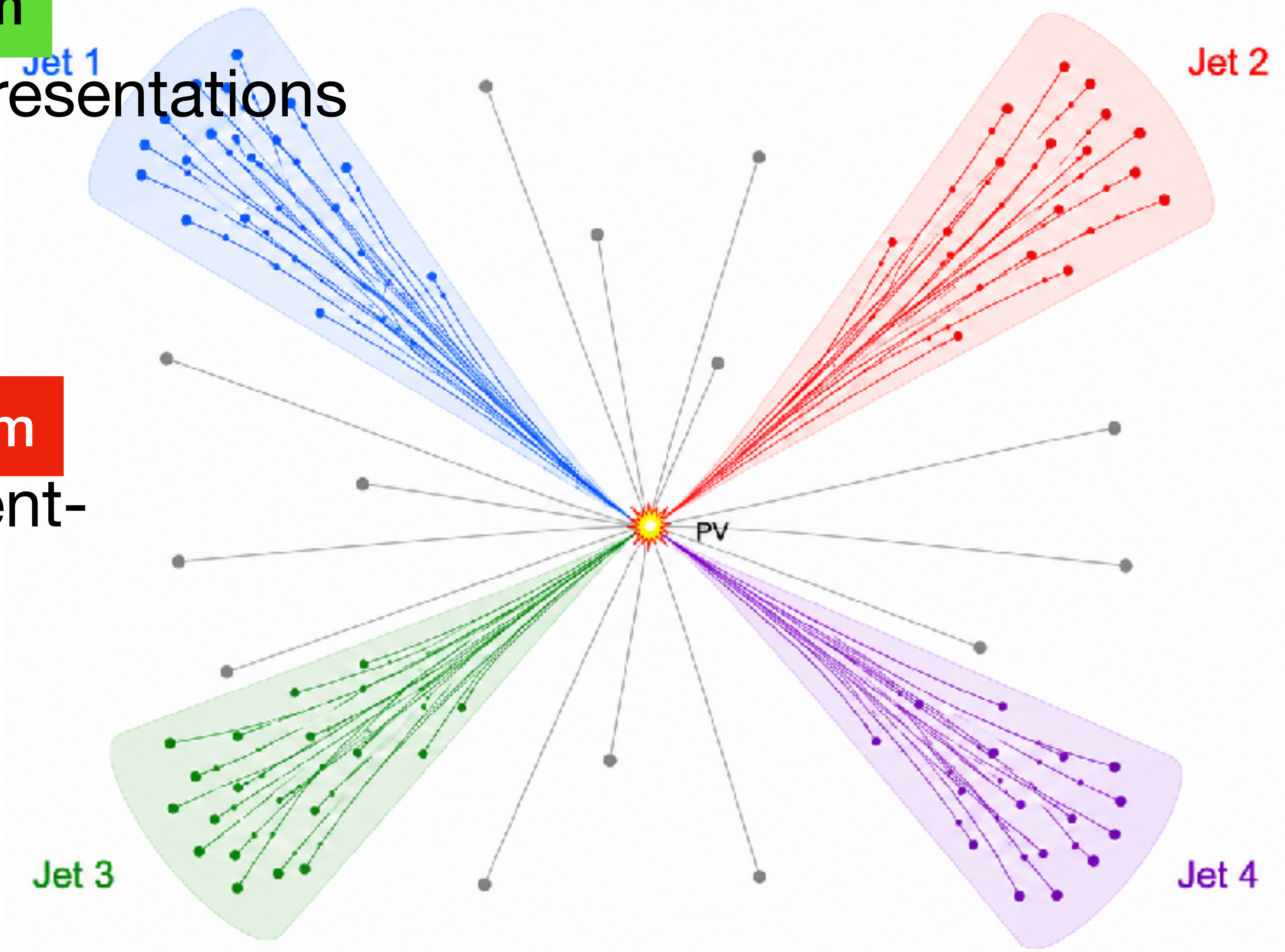
PF Candidates $\xrightarrow{\text{Clustering}}$ Jets $\xrightarrow{\text{Selection}}$ Di-Higgs Mass Spectrum

Drawback: particle-level information is compressed into jet-level representations

New approach:

PF Candidates $\xrightarrow{\text{Event-Level Transformer Model}}$ Di-Higgs Mass Spectrum

Advantage: retains particle-level information and directly models event-wide correlations, achieves O(10) background suppression.



450 fb^{-1} projection: 2.8σ, –0.81 < κλ < 8.7

The gain is striking — but where does it come from? → Let's disentangle it step by step

Analysis Setup

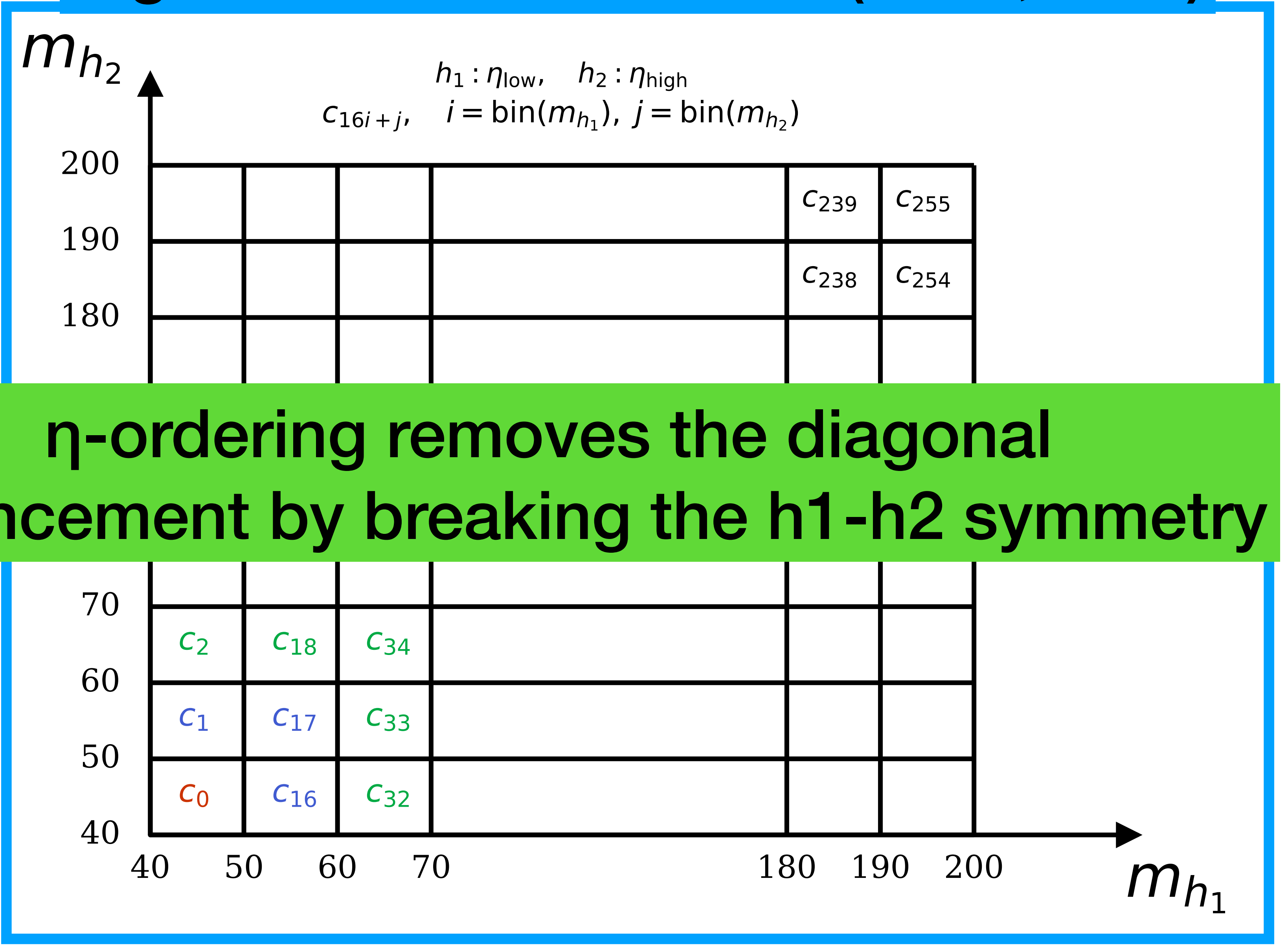
Training:

- Dataset: X→Y_1Y_2→4b(Signal Proxy, 50M), QCD(50M) and TTbar
- Model: Transformer (~9M parameters)
- Class: 256 signal score + QCD score + $t\bar{t}$ score
- Input Variables:
  - PF features: kinematics, particle ID, and track variables
  - PF four-vectors to build pairwise features:
    lnkt , lnz, lnΔR, lnm2, lnds2 , cosθ, Δy, Δϕ

Score Usage:

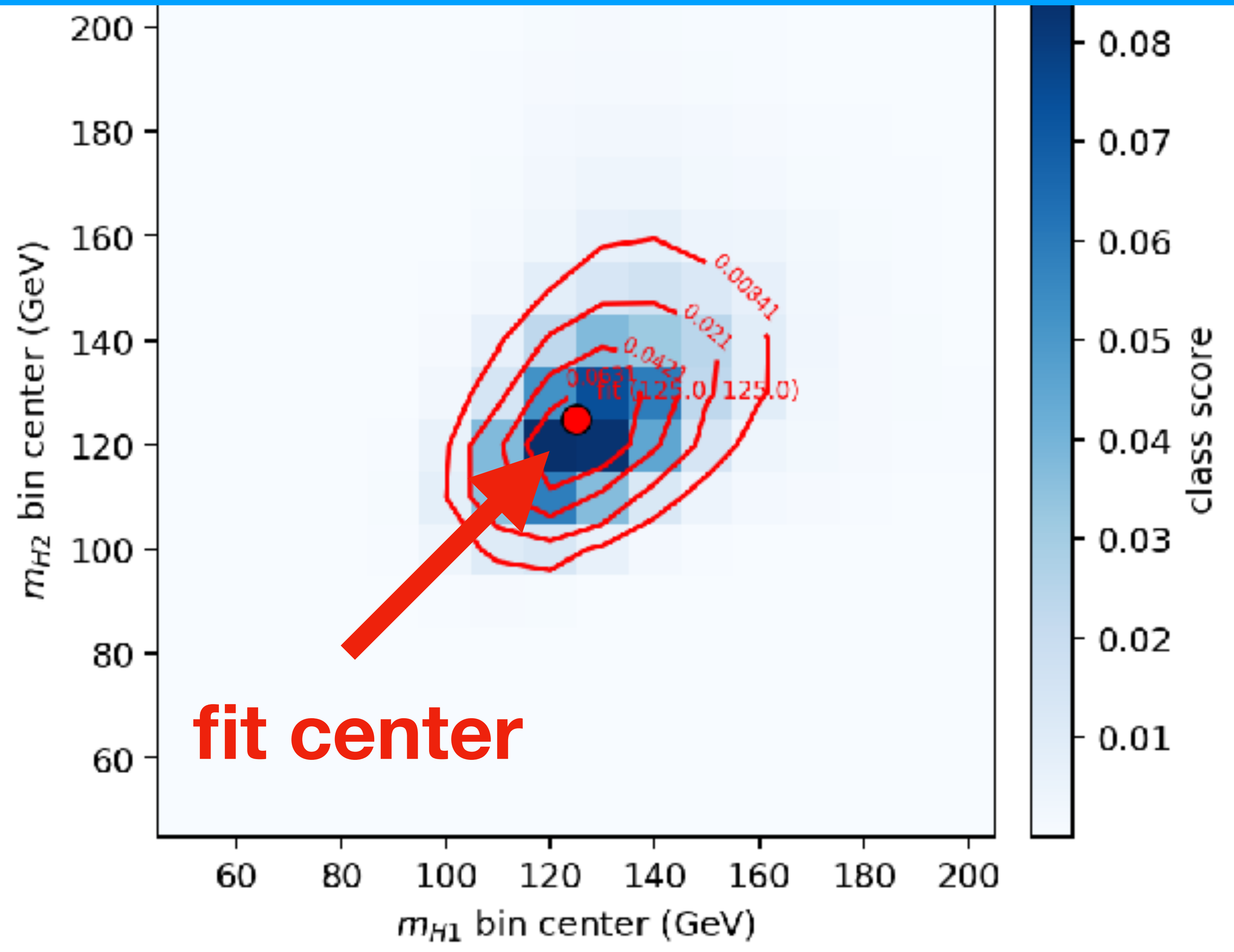
- Event selection: cut on $\mathrm{Disc} = \dfrac{\sum_{i=1}^{256} S_i}{\sum_{i=1}^{256} S_i + QCD + t\bar{t}}$
- Mass reconstruction: fit the 256 signal scores with a 2D Gaussian
- Results: 1-bin counting with S and B yields in the signal region.

Signal-class definition in (mh1 ,mh2)



η-ordering removes the diagonal enhancement by breaking the h1-h2 symmetry

Example event: score map + 2D Gaussian fit



yipin.wang@cern.ch

Projected Sensitivity of the Full-Event Model

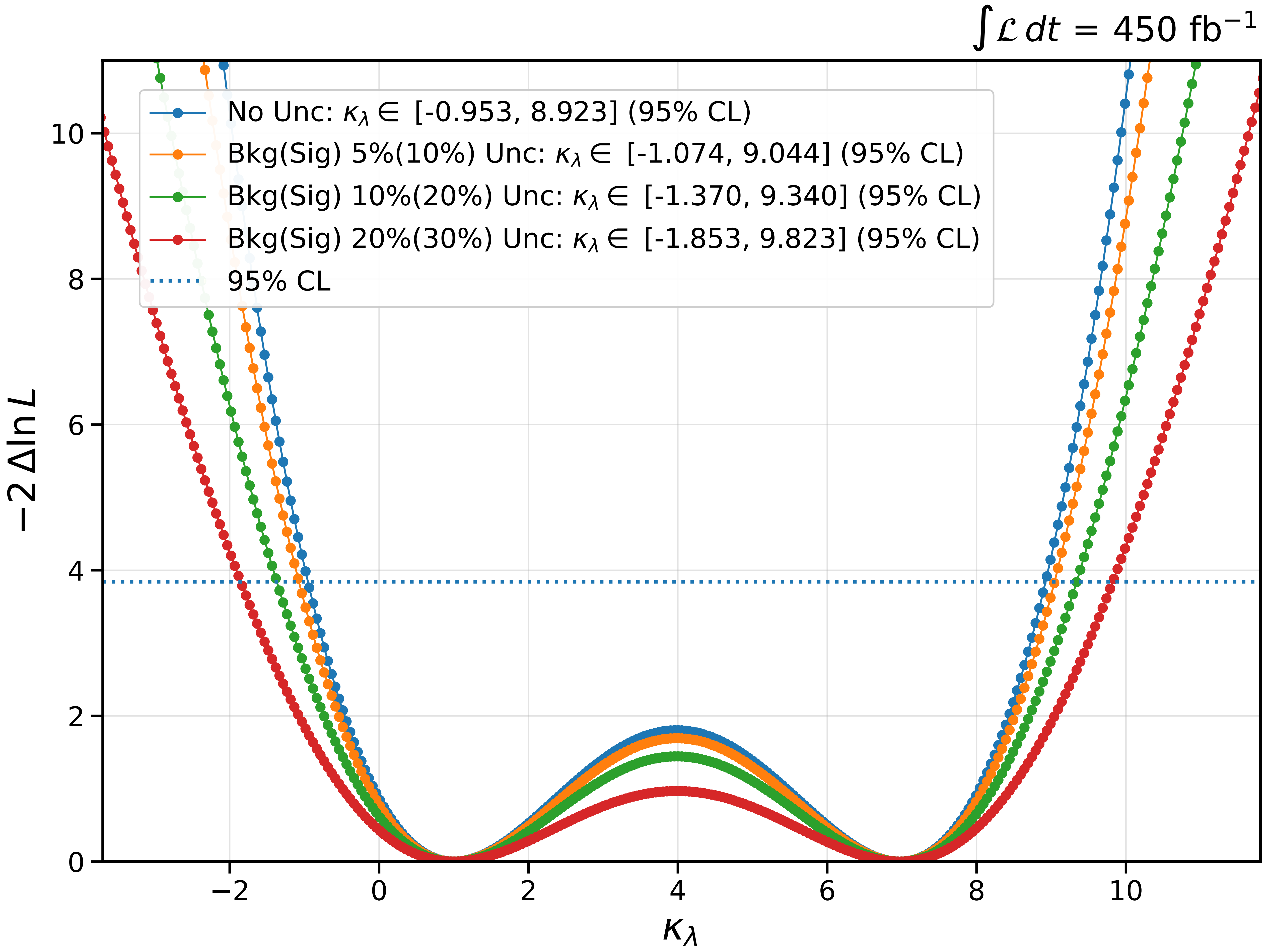
Working point: D > 0.9998760252

Backgrounds: QCD + top/electroweak/Higgs backgrounds.

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

Process	Expected Yield
ggHH, κλ = 0	19.728
ggHH, κλ = 1 (SM)	13.818
ggHH, κλ = 5	5.797
qqHH	0.0713
QCD	24.169
Total Background	25.579

Background Uncertainty	Asimov Significance
0%	2.5414σ
10%	**2.2307σ**
20%	1.7182σ



The ML2025 sensitivity was partly enhanced by an unphysical diagonal feature; removing it in the 256-class model gives slightly lower but more reliable sensitivity

Understanding Performance Gains of a Full-Event HH→4b

Progressive Ablation From the Full Model

Question: Where does the performance gain of the full-event approach come from?
Strategy: Start from the full model and progressively remove information or modeling capability, while keeping the same training and evaluation framework

Step	Model / Representation	What it isolates
Reference	Full-event PF — Large / 100M	Full-model performance
Step 1	Full-event PF — Small / 10M	Training scale
Step 2	In-Jet PF	PF candidates outside reconstructed jets
Step 3	Trainable Jet Embedding	Cross-jet PF correlations
Step 4	Frozen Jet Embedding	HH→4b-specific in-jet optimization
Step 5	Jet features	Conventional jet-level baseline / validation against CMS Latest HH4b Results arXiv:2604.27044

Each step isolates a different source of performance gain

yipin.wang@cern.ch

Understanding Performance Gains of a Full-Event HH→4b

Step 1: How Much Does Training Scale Matter?

Reduced Training:

- Goal: separate the gain from training scale from the gain due to the input representation, using a training scale closer to typical analyses (O(10M) events)

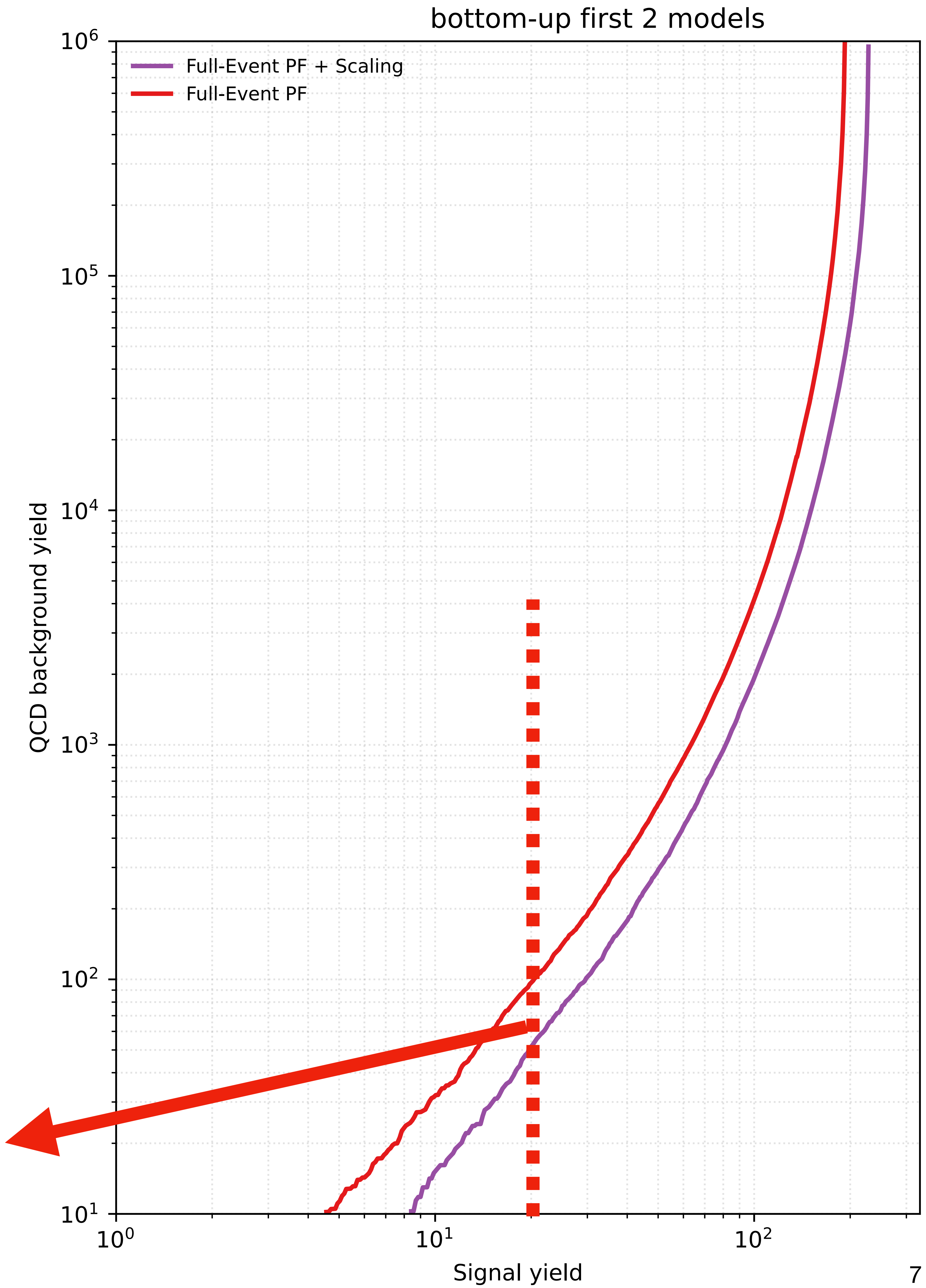
Changes from the full size model:

- Training samples: 100M → 10M events
- Model parameters: ~9M (Large) → ~1M parameters (Small)

Model	Transformer	Particle embedding	Pairwise embedding	FC head
Large	8 layers × 16 heads	256 → 1024 → 256	64 → 64 → 64	1024
Small	6 layers × 8 heads	96 → 384 → 96	32 → 32 → 32	256

Full-Event PF candidates
(10M training events)

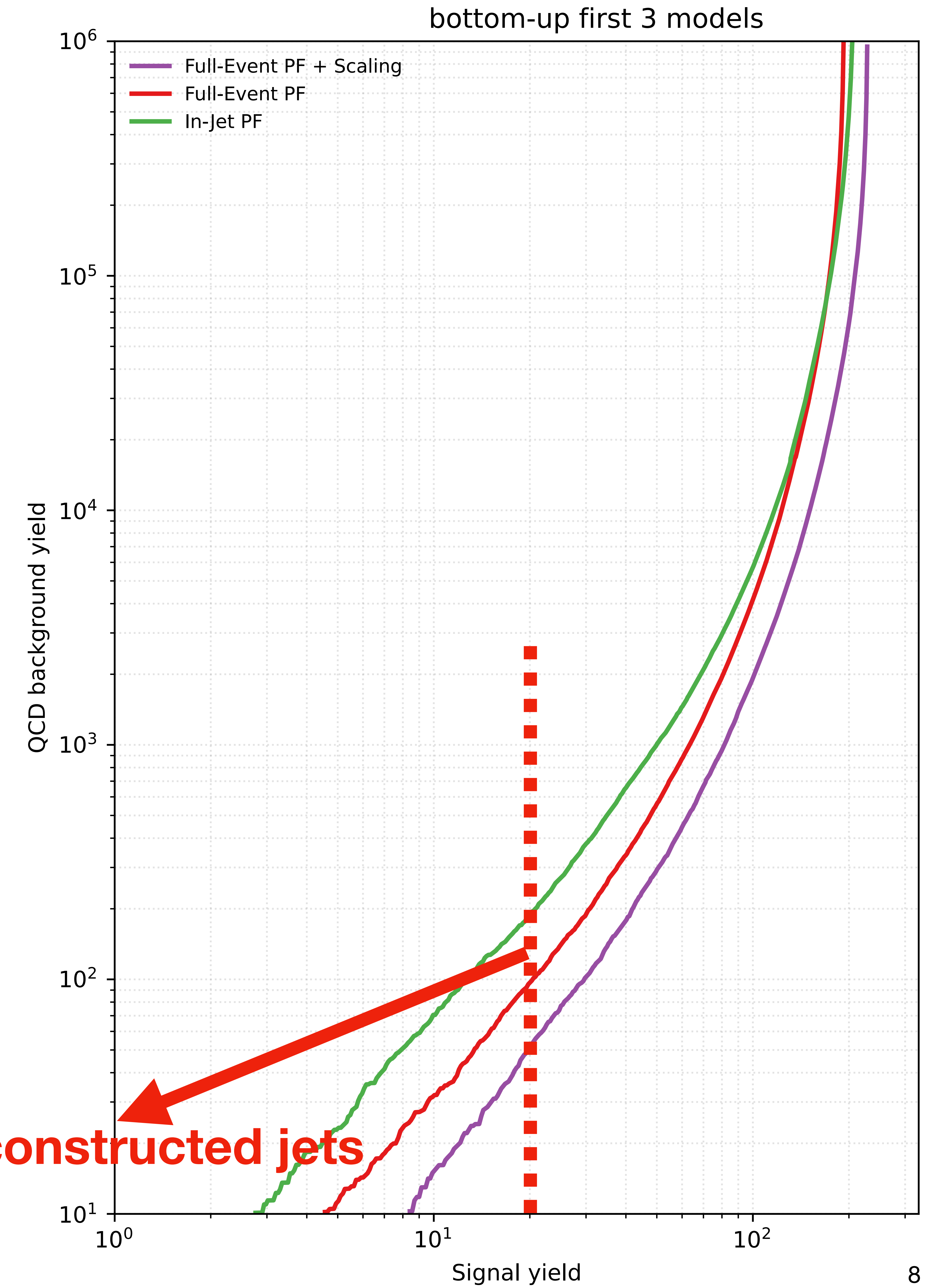
Small event-level Transformer



~2× stronger background rejection from the larger training scale

Understanding Performance Gains of a Full-Event HH→4b

Step 2: What Do PF Candidates Outside Jets Add?



Restricting PF Candidates to Reconstructed Jets

- Goal: quantify the contribution from PF candidates outside the selected reconstructed jets

Change from Step 1:

- Full-event PF → PF candidates inside jets

PF candidates in jet
(10M training events)

Same small event-level Transformer as in Step 1

~2x stronger background rejection from PF candidates outside reconstructed jets

yipin.wang@cern.ch

Understanding Performance Gains of a Full-Event HH→4b

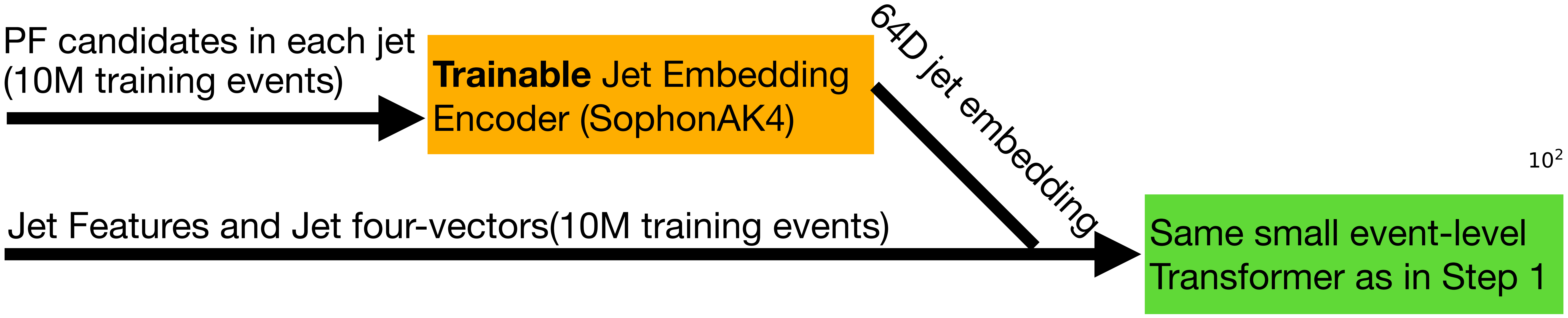
Step 3: How Much Do Cross-Jet PF Correlations Help?

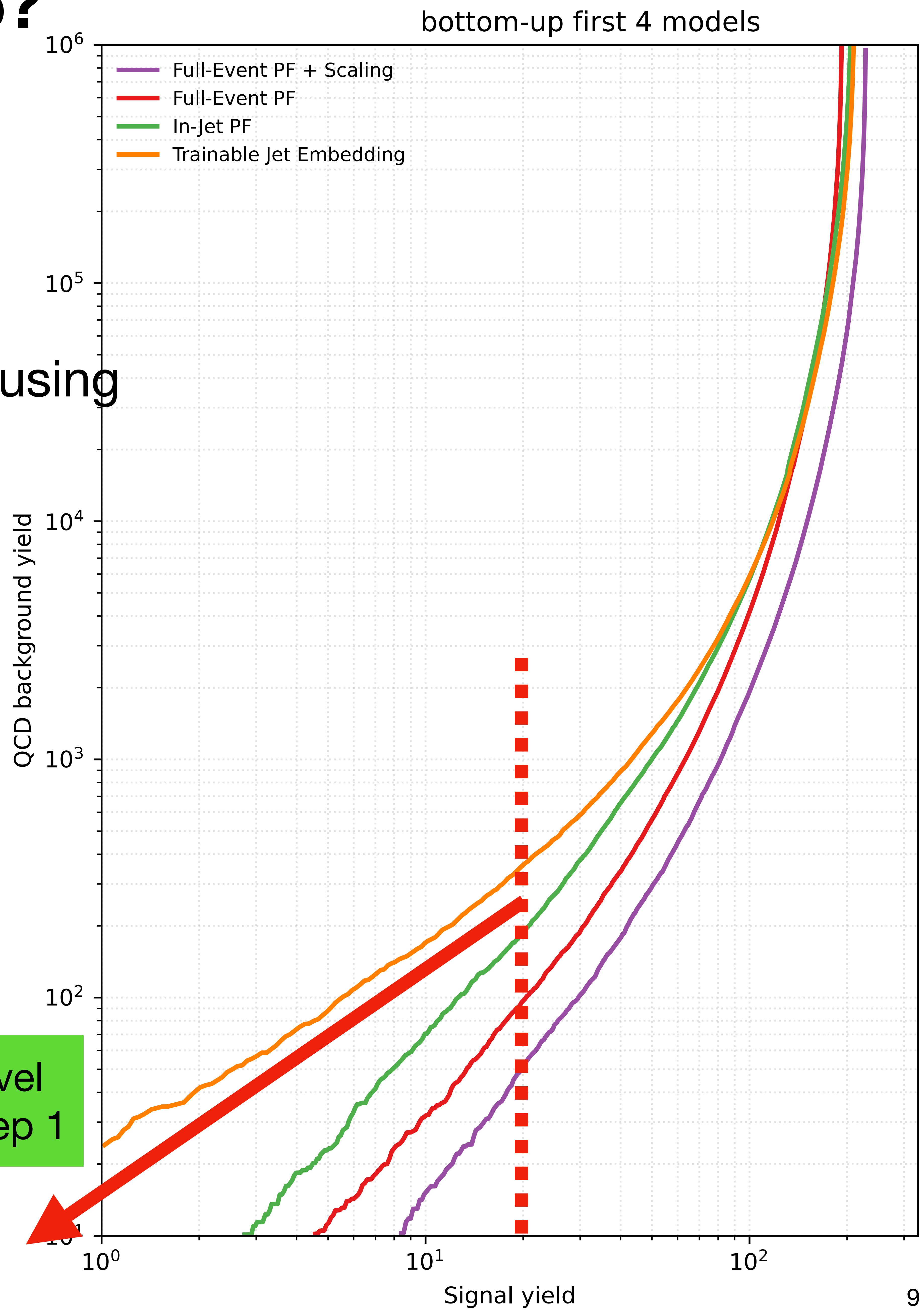
Per-Jet PF Encoding: Removing Direct Cross-Jet PF Correlations

- **Goal:** quantify the gain from directly modeling pairwise correlations between PF candidates across different jets.

Training Setup:

- Extract a 64D learned embedding from PF candidates inside each jet using a pretrained and Trainable jet model (SophonAK4 tagger) as encoder
- Input Variables:
  - Jet Features: kinematics, jet tagger scores (B,QCD)
  - Jet embedding: 64D output from the jet encoder
  - Jet four-vectors to build 8D pairwise features
- Size: 10M Training samples and 1M parameters Model



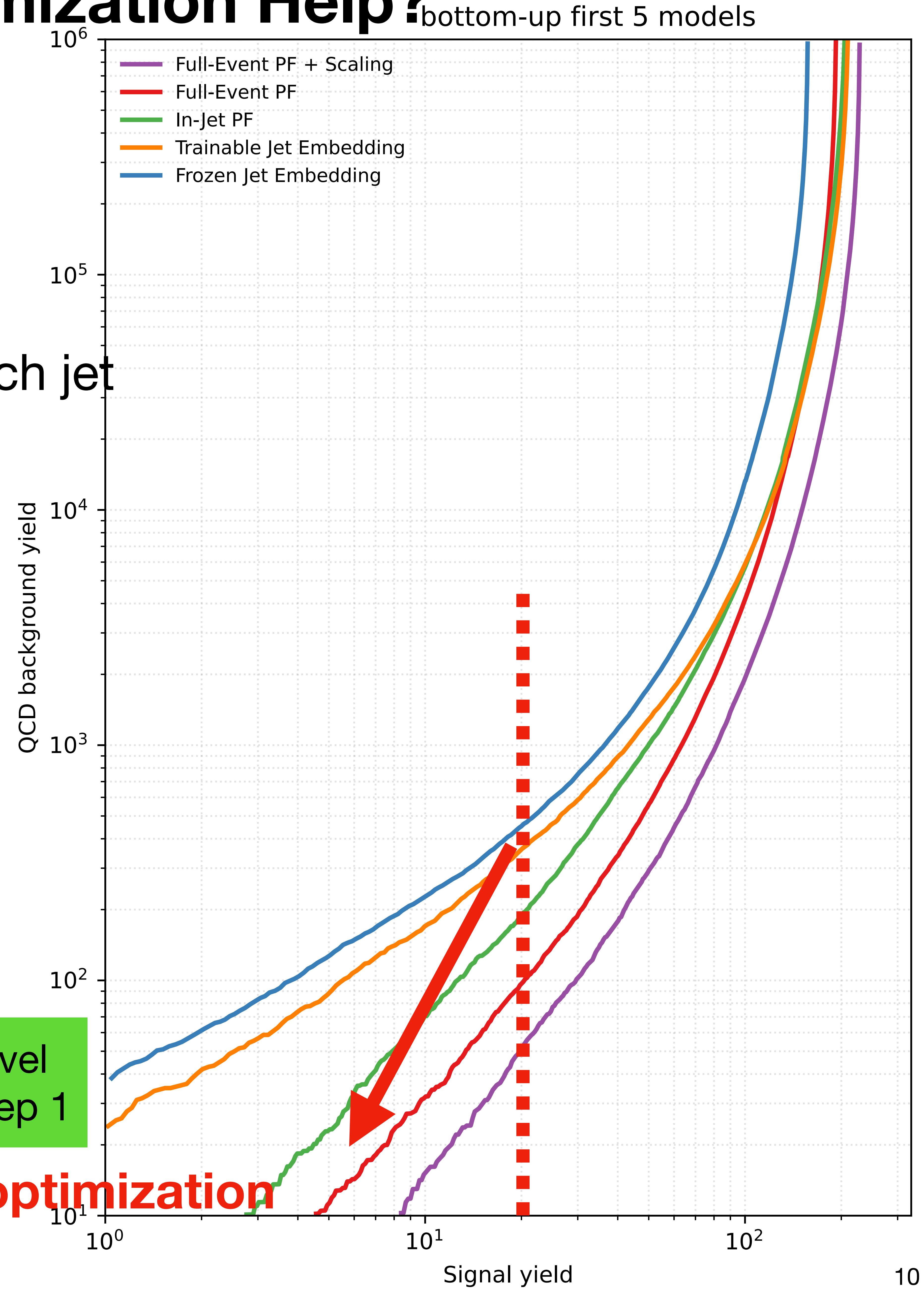


~2× stronger background rejection from cross-jet PF correlations

yipin.wang@cern.ch

Understanding Performance Gains of a Full-Event HH→4b

Step 4: How Much Does HH→4b-Specific In-Jet Optimization Help?

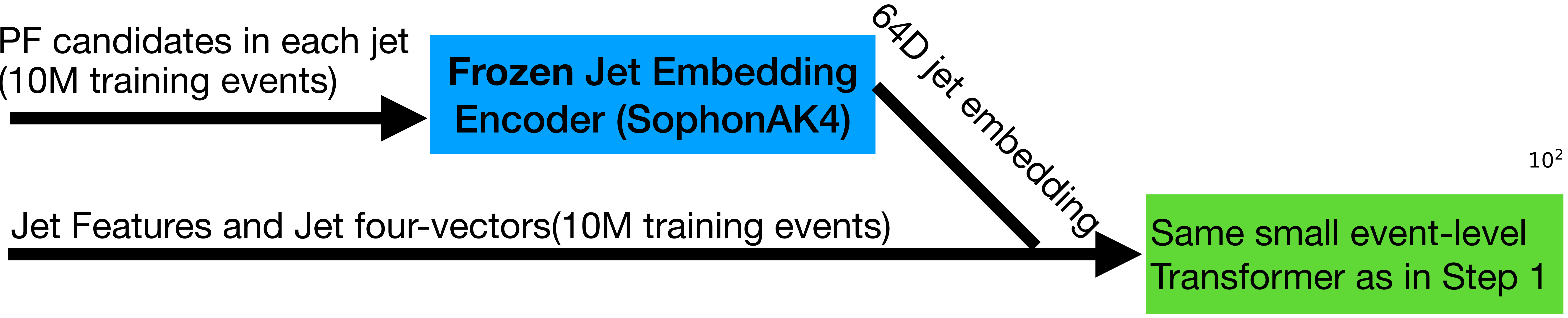


Frozen Jet Embedding: Removing HH→4b-Specific Optimization

- **Goal:** isolate the gain from HH→4b-specific optimization within each jet

Change from Step3:

- SophonAK4 encoder: Trainable → Frozen



~1.25× stronger background rejection from HH→4b-specific in-jet optimization

yipin.wang@cern.ch

Understanding Performance Gains of a Full-Event HH→4b

Step 5: Can We Reproduce the CMS Jet-Level Baseline?

Jet-Level Baseline Validation and Benchmark

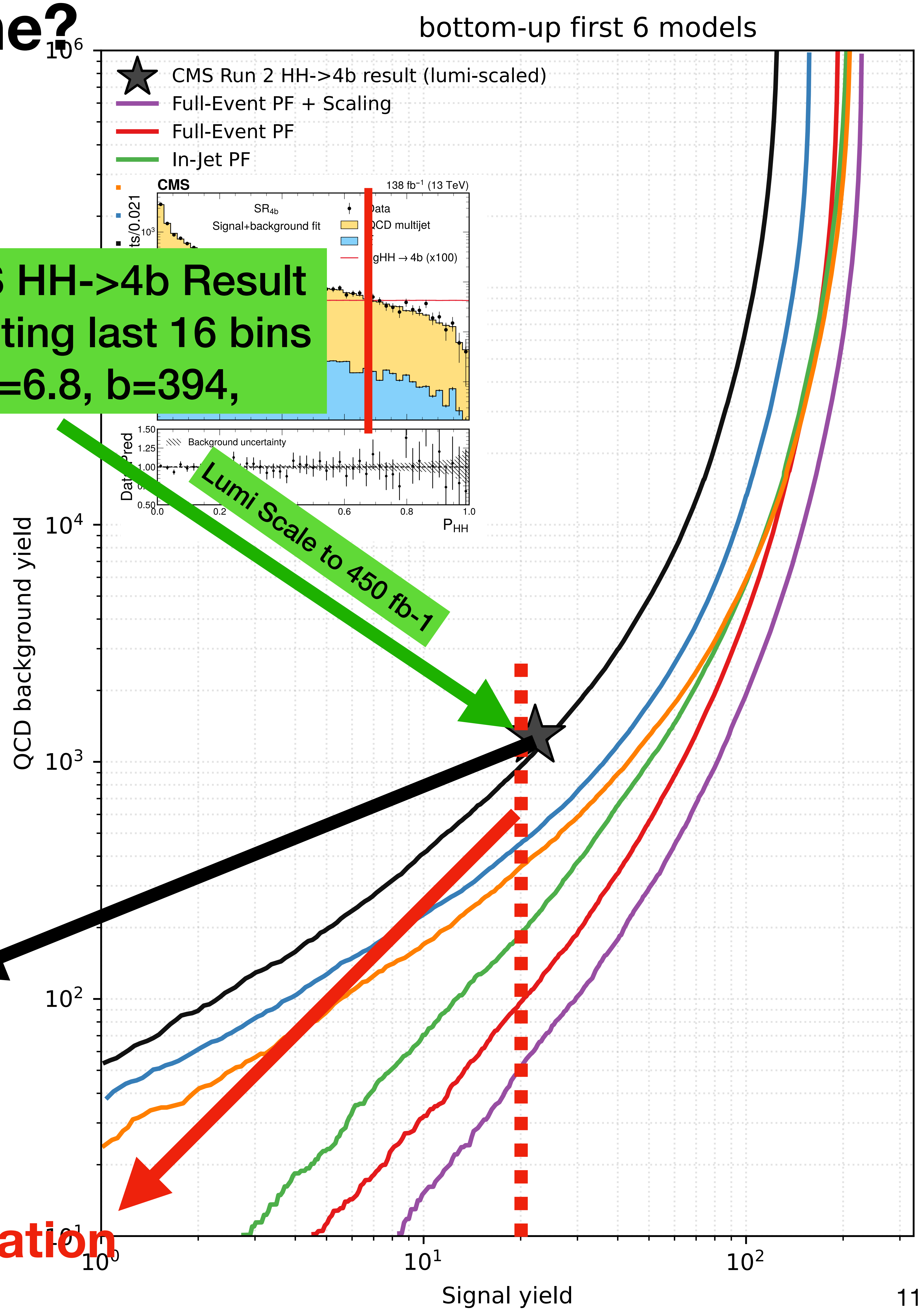
- **Goal:** validate our training framework by reproducing the performance of the CMS jet-level HH4b analysis, and then quantify the improvement from a richer jet-level representation over conventional jet features.

Changes from step4:

- Remove Jet embedding Encoder
- Keep jet tagger score in jet feature

Jet Features and Jet four-vectors(10M training events) → Same small event-level Transformer as in Step 1

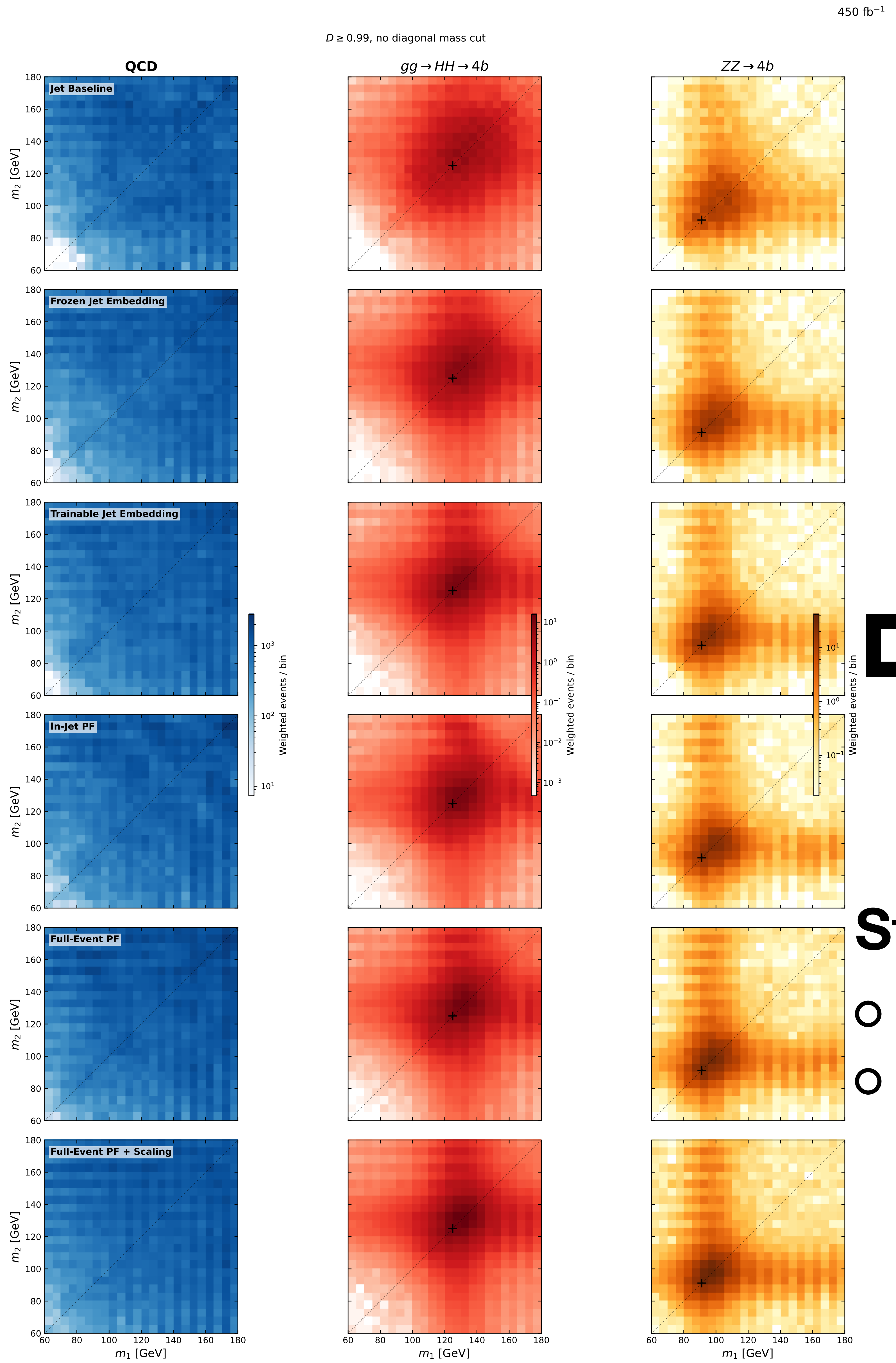
The conventional jet baseline reproduces CMS HH4b results, validating our training framework



~2× stronger background rejection from a richer jet-level representation

Understanding Performance Gains of a Full-Event HH→4b

Mass Plot Evolution Across the Ablation Steps



Common working point: D > 0.99

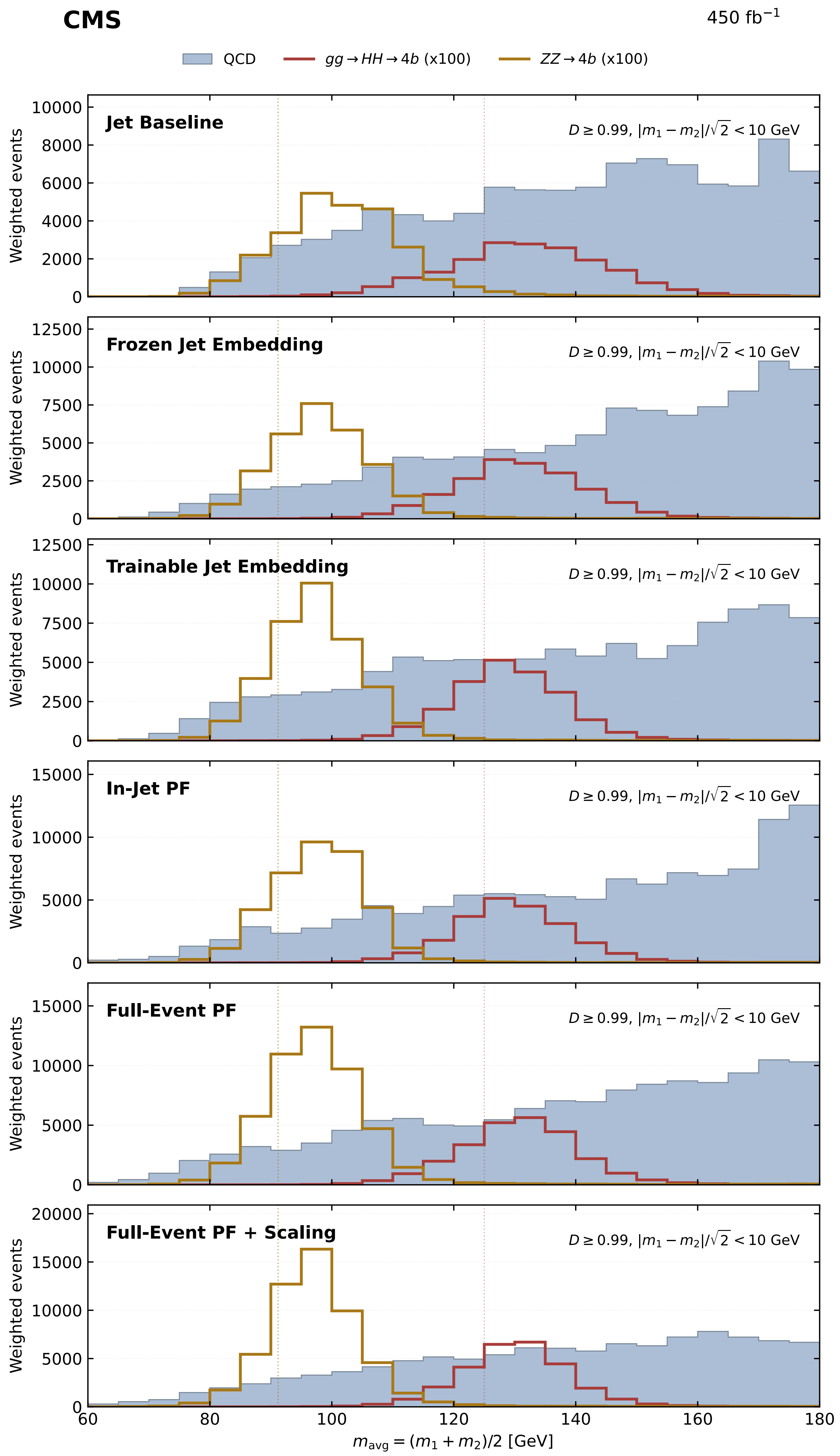
Left: 2D mass distributions in (m1, m2)

Right: diagonal projection, $m_{\text{avg}} = \dfrac{m_1 + m_2}{2}$

Selection: $\dfrac{|m_1 - m_2|}{\sqrt{2}} < 10\ \text{GeV}$

Step-by-step improvements:

- Smoother QCD mass distributions
- Narrower and more pronounced signal peaks



Step by step, better mass resolution, better QCD suppression

yipin.wang@cern.ch

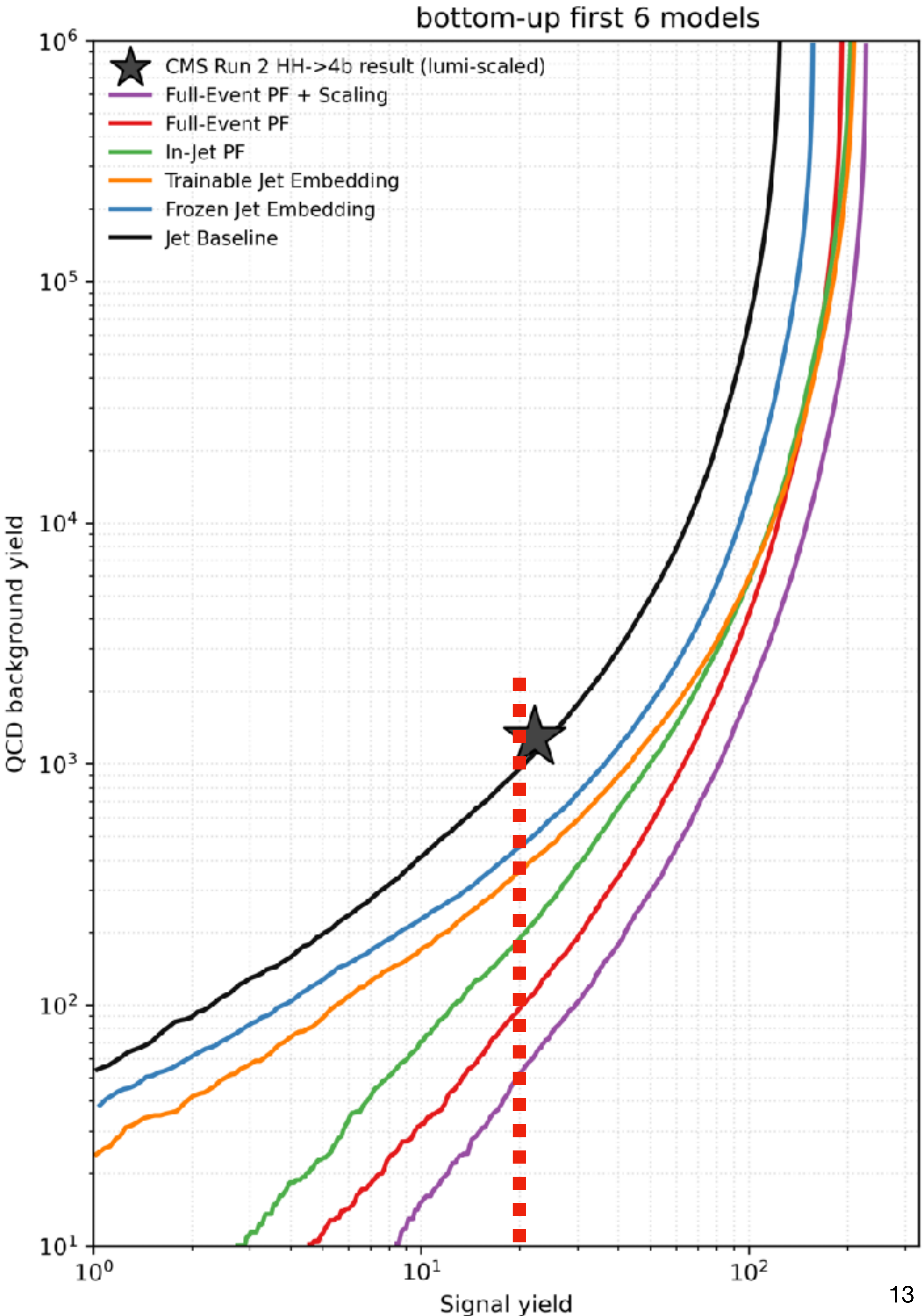
Summary

Full-event particle-level modeling

- Full-event particle-level modeling preserves more information and captures event-wide correlations

Model	Source of Gain	Rejection Gain
Full-Event PF + Scaling	Training scale	~2×
Full-Event PF	PF candidates outside reconstructed jets	~2×
In-Jet PF	Cross-jet PF correlations	~2×
Trainable Jet Embedding	HH→4b-specific in-jet optimization	~1.25×
Frozen Jet Embedding	Richer jet-level representation	~2×
Jet Baseline	Conventional jet-level baseline / validation against CMS Latest HH4b Results [arXiv:2604.27044]()	Reference

~20× stronger QCD background rejection at ~20 signal events

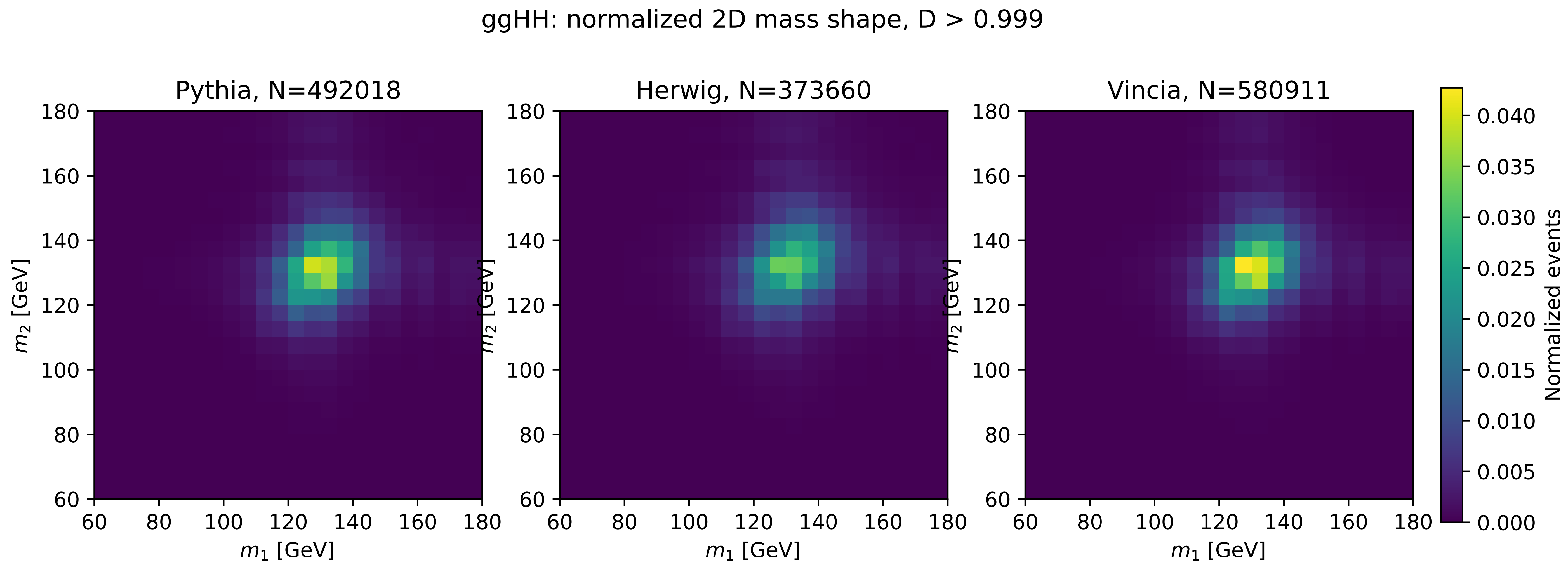


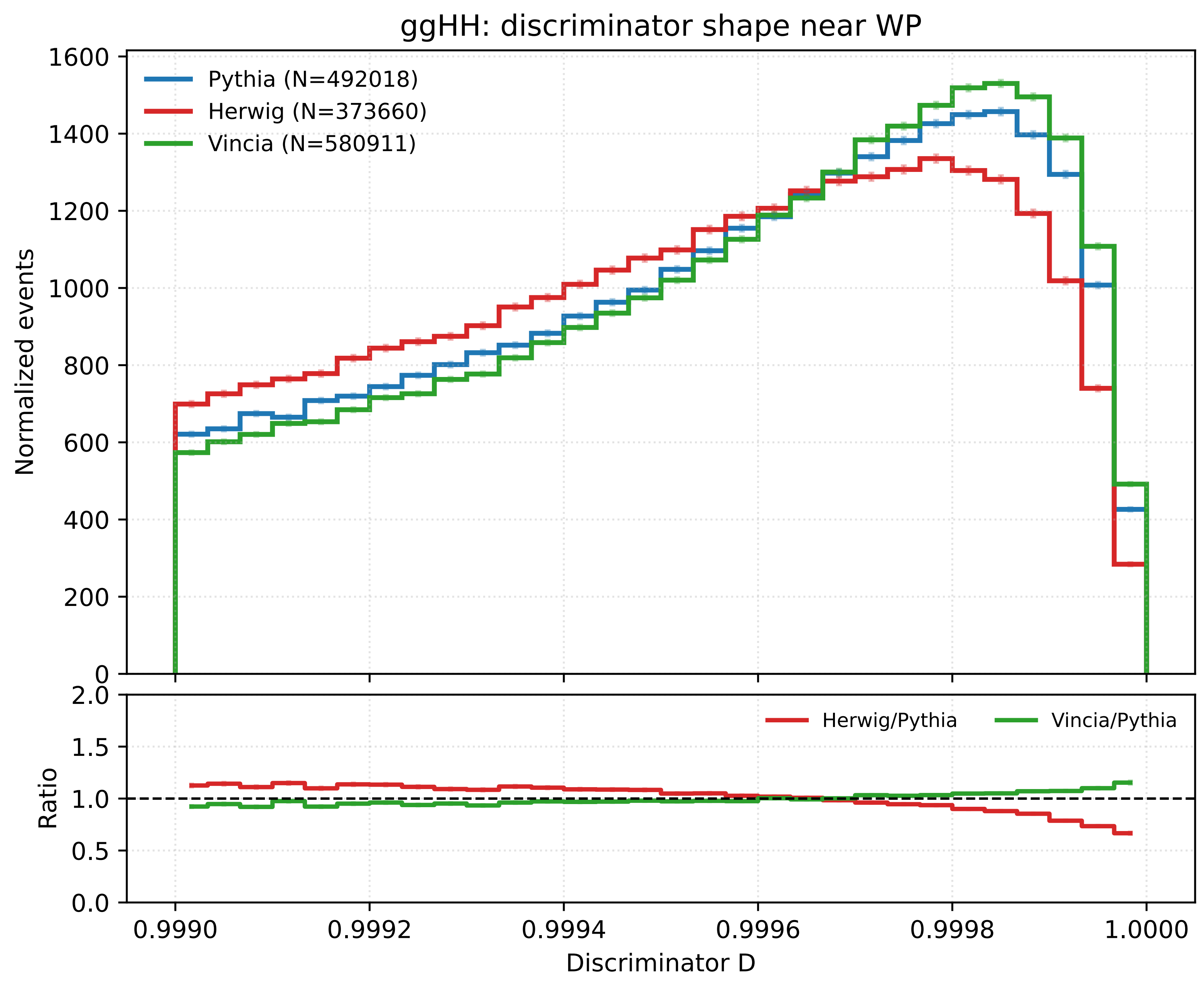
THANK YOU

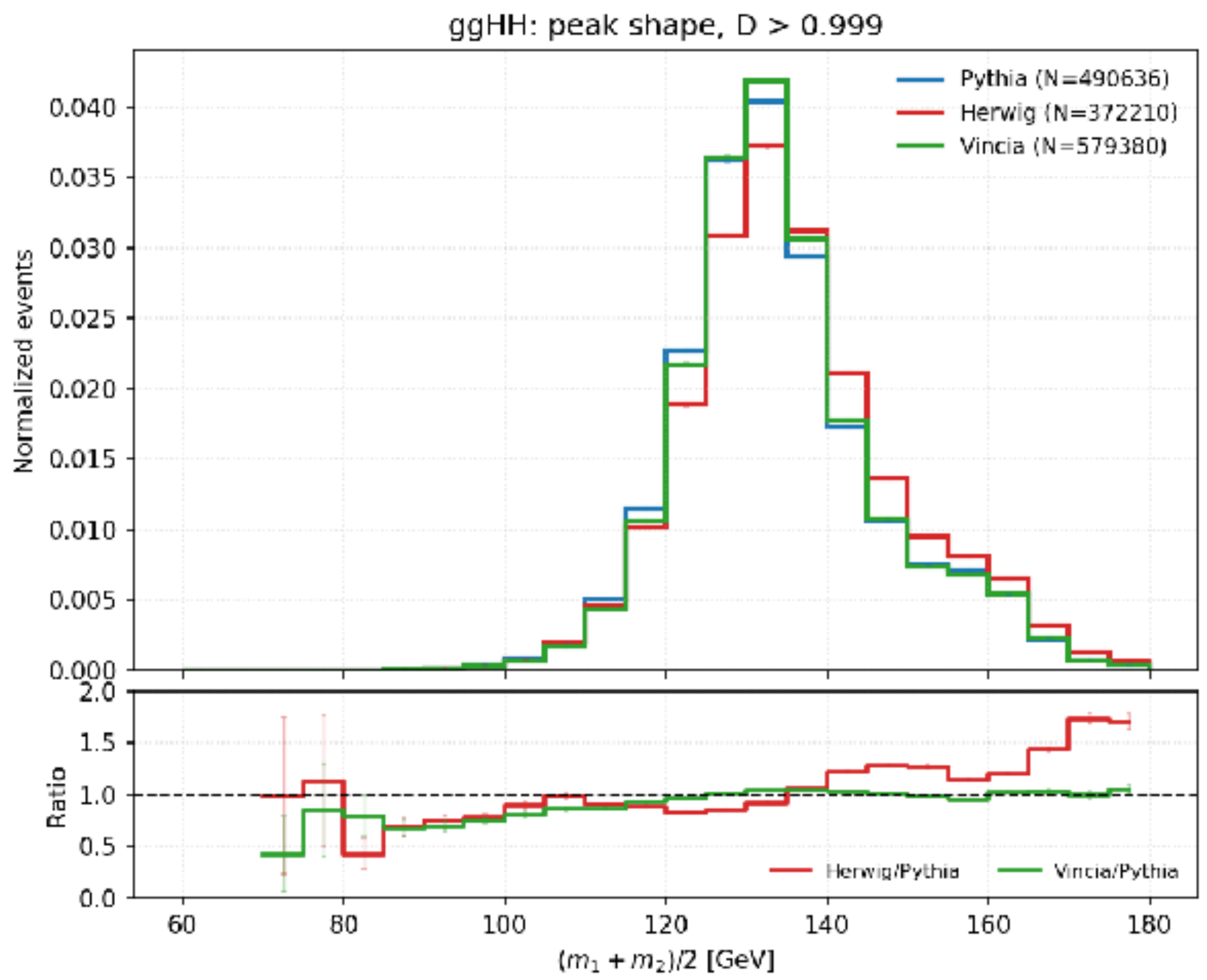
Back up

Generater Dependence

How sensitive is the model to generator/shower modeling?







yipin.wang@cern.ch

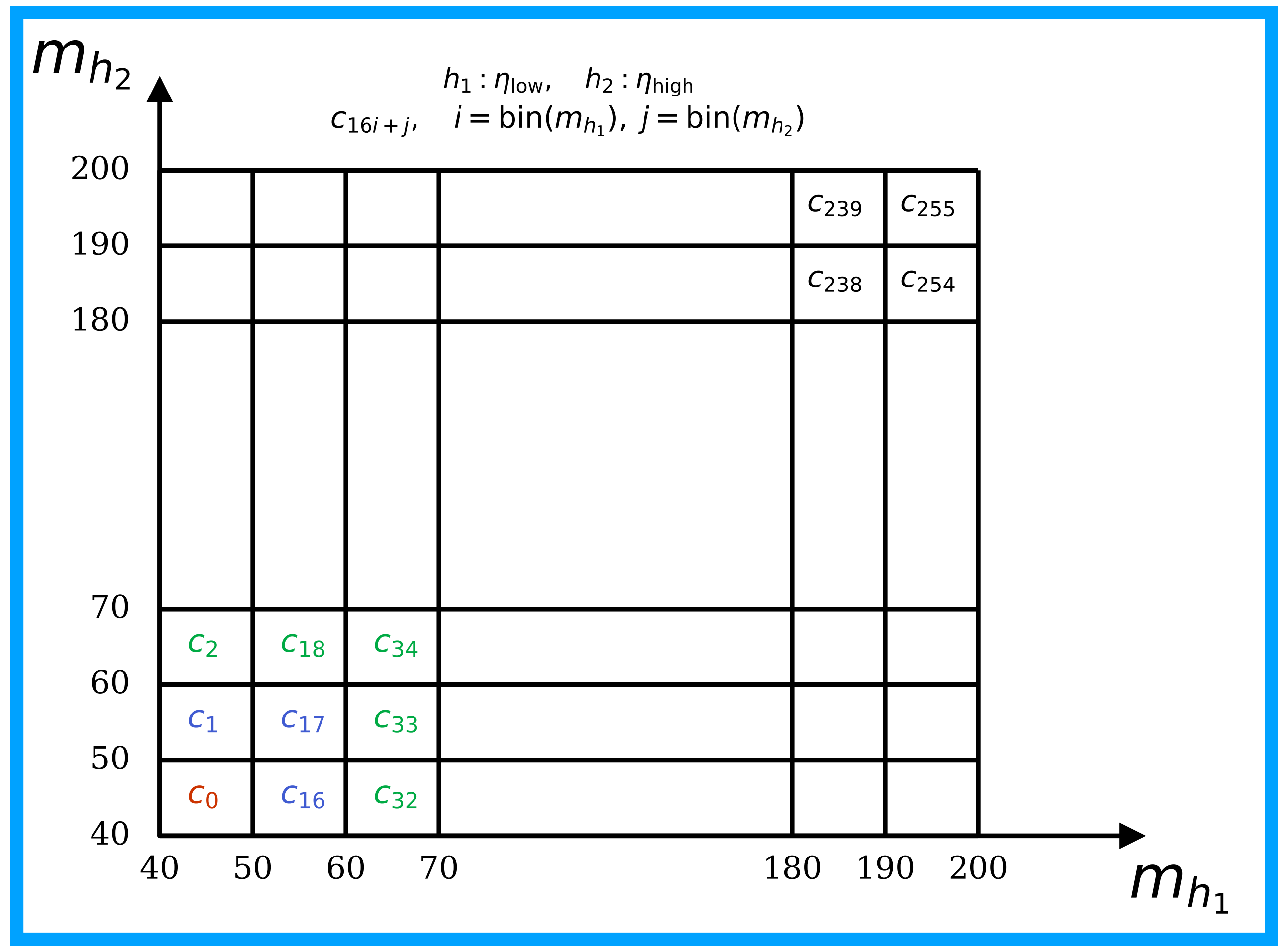
Training Sample

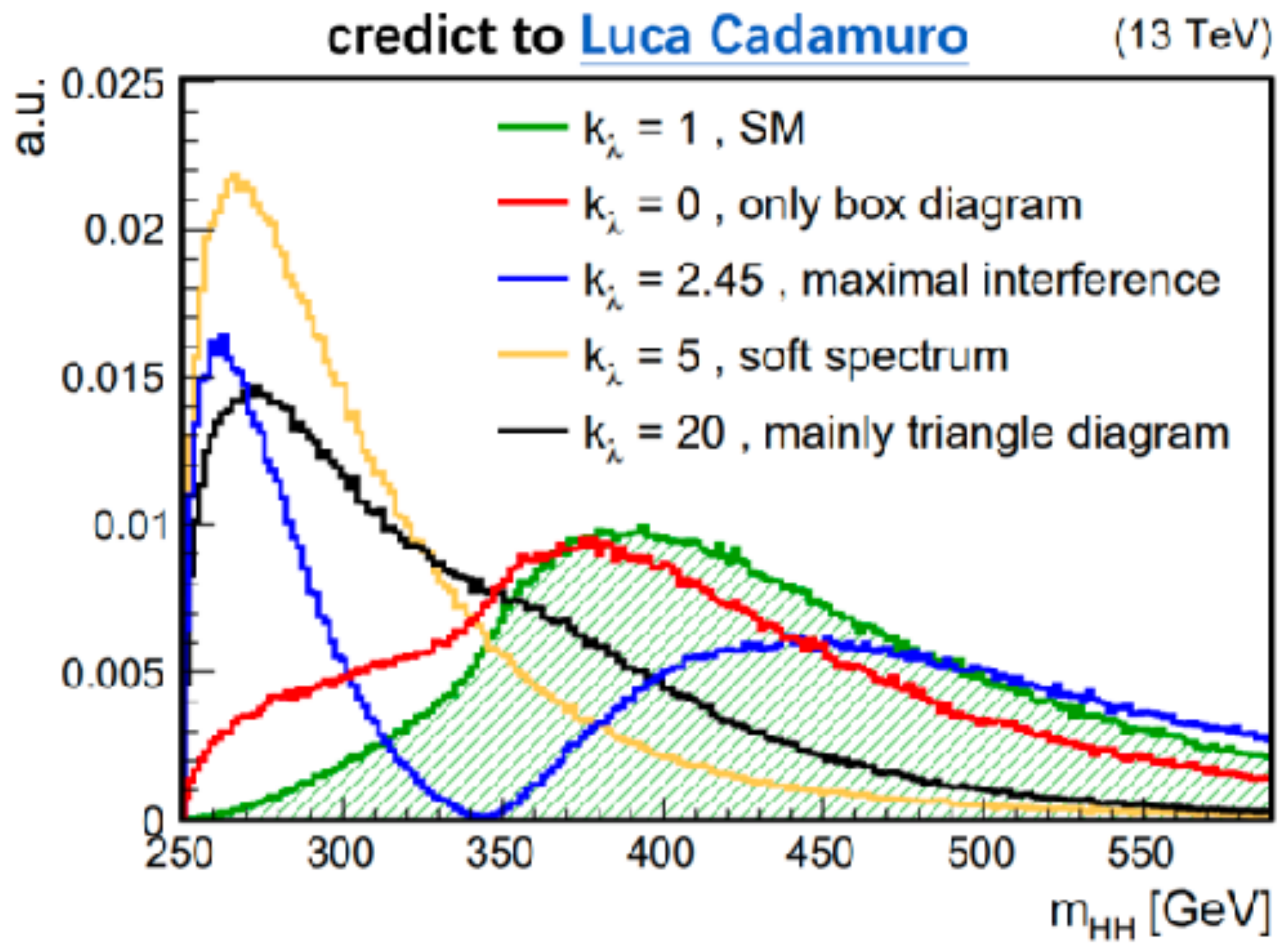
Three-process training: $X \to Y_1Y_2 \to 4b$(Signal Proxy), QCD and TTbar(BKG)

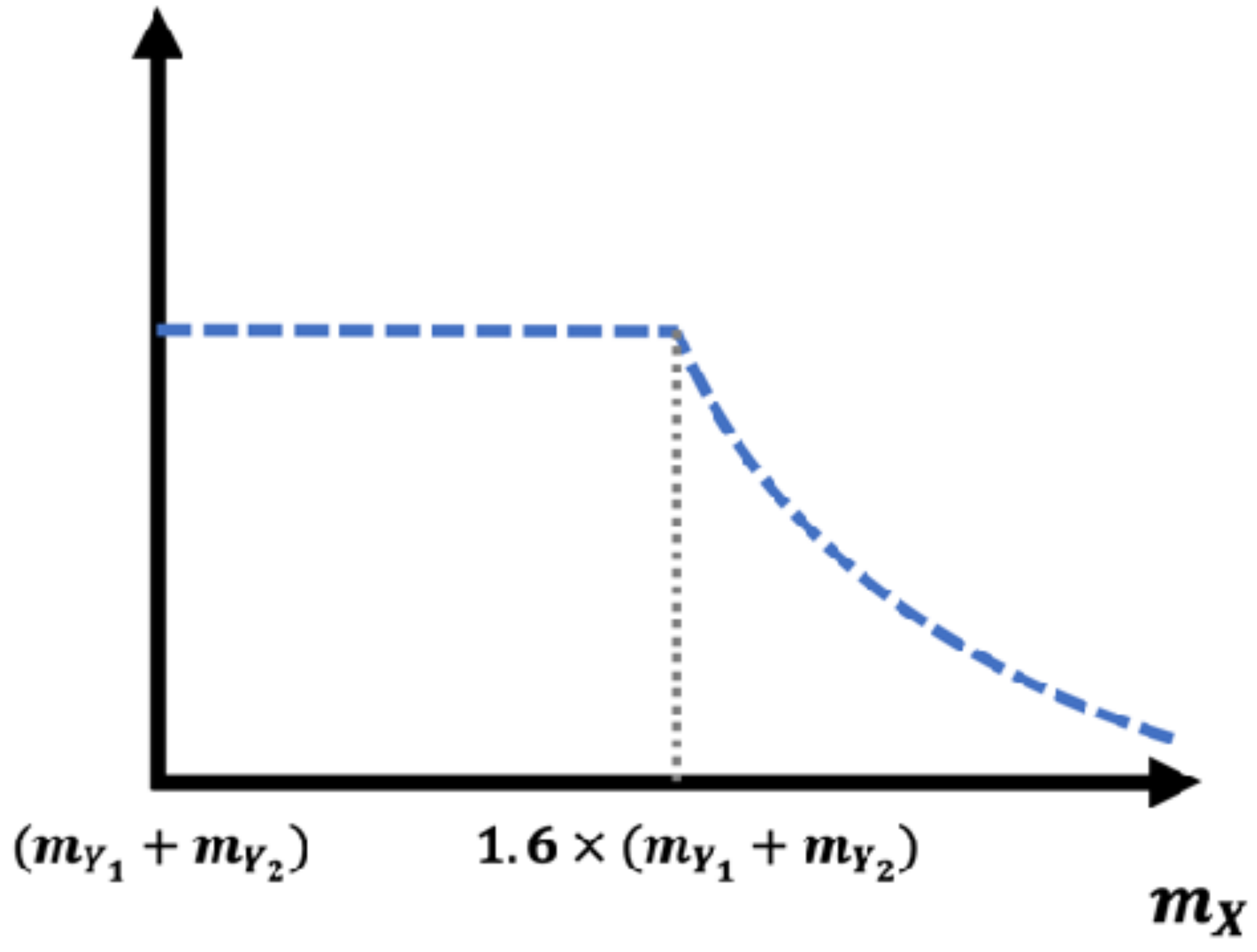
- Events are required pass resolved “4j3b” trigger preselection
- Main dataset Size: $X \to Y_1Y_2 \to 4b$(50M); QCD(50M)

$X \to Y_1Y_2 \to 4b$ as a variable-mass proxy for the HH → 4b signal:

- Y_1 and Y_2 masses: uniformly distributed in [40, 200] GeV
- X mass used for m_{HH} spectrum control: taking various HH→4b topologies as a reference







Training Setup

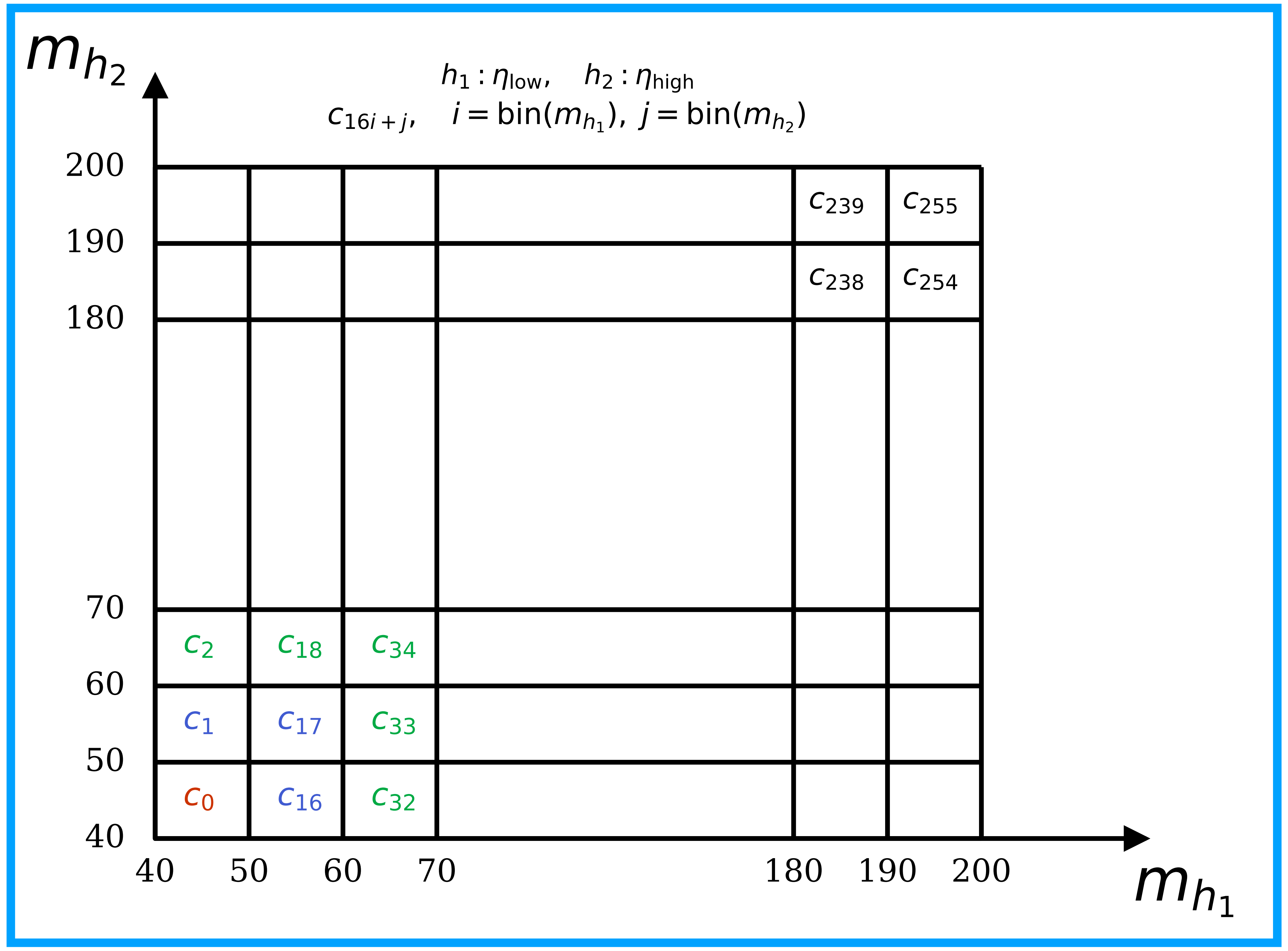
Preselection:

- Events Passing Resolved "4j3b" Trigger (100M events pass preselection)

Framework & Architecture

- Softwork: Weaver—a streamlined PyTorch-based ML framework for HEP training and inference
- Trasfomer Parameter:

Model	Transformer	Particle embedding	Pairwise embedding	FC head
Large	8 layers × 16 heads	256 → 1024 → 256	64 → 64 → 64	1024
Small	6 layers × 8 heads	96 → 384 → 96	32 → 32 → 32	256



Training Class:

- Signal: 16 × 16 (mh1, mh2) mass classes
- Backgrounds: QCD and TTbar

Input Variables:

- PF features: kinematics, particle ID, and track variables
- PF four-vectors to build pairwise features: lnkt , lnz, lnΔR, lnm2, lnds2 , cosθ, Δy, Δϕ

Background Yield Estimation for ROC

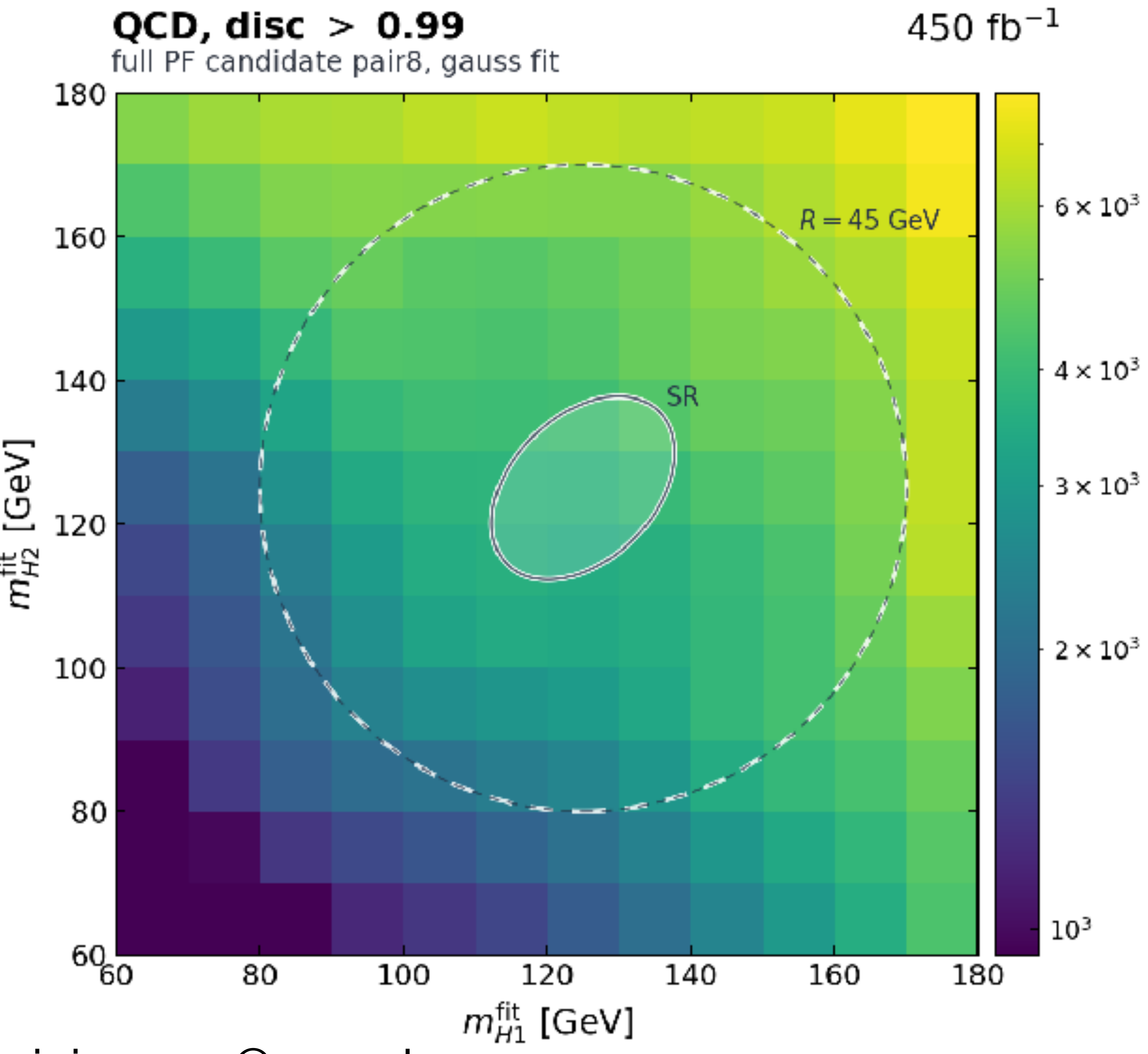
Signal Region: Elliptical region with semi-axes 10 and 15 GeV

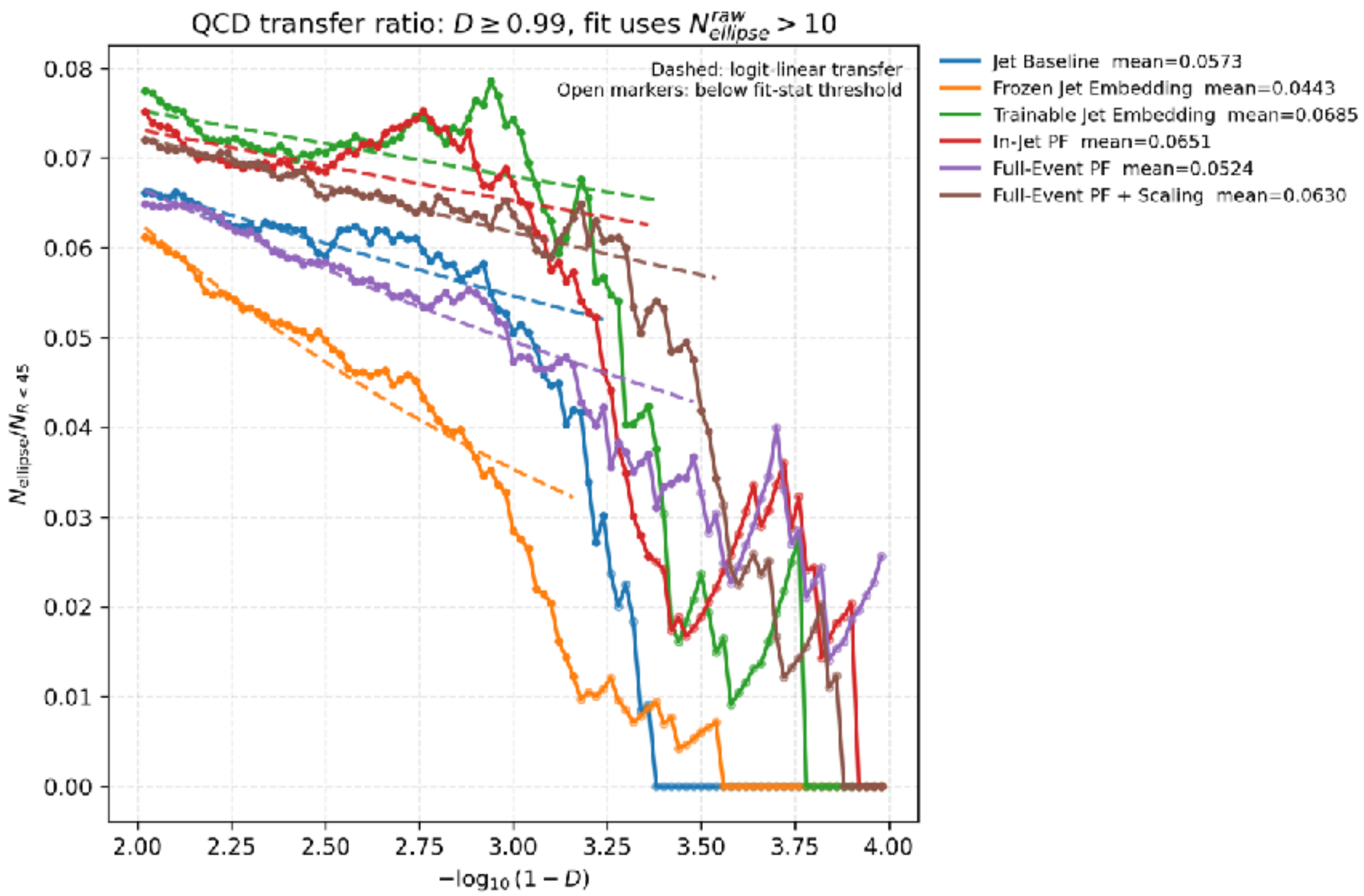
Issue at the ROC Tail

- limited background statistics lead to large fluctuations in the estimated background yield

Background Proxy

- Estimate the background yield in a larger **45 GeV proxy region**
- Scale it back to the nominal SR using a **transfer factor**





yipin.wang@cern.ch

Model Score Usage Strategy

Output Scores: 256 signal score + QCD score + $t\bar{t}$ score

$$\text{Disc} = \frac{\sum_{i=1}^{256} S_i}{\sum_{i=1}^{256} S_i + \sum_{j\in \text{BKG}} B_j \sigma_j \epsilon_j^{\text{trig}}}$$

For selected events, derive m_1 and m_2 from a Gaussian fit to the 256 signal scores

$$S(m_1, m_2) \propto \exp\left[-\frac{(m_1-\mu_1)^2 + (m_2-\mu_2)^2}{2\sigma^2}\right]$$

Gaussian fitting reduces score fluctuations and provides smoother mass estimates