



UF



Improved CMS results on Higgs boson pair production in the 4b final state with Run 3 data

EVAN KOENIG

YOUNG SCIENTISTS FORUM



Search for $HH \rightarrow 4b$

Broad program at CMS to measure and understand the Higgs potential



Main HH production mechanisms:

- **Gluon-fusion (ggHH)**
- **Vector-boson-fusion (qqHH)**

Different HH final-states:

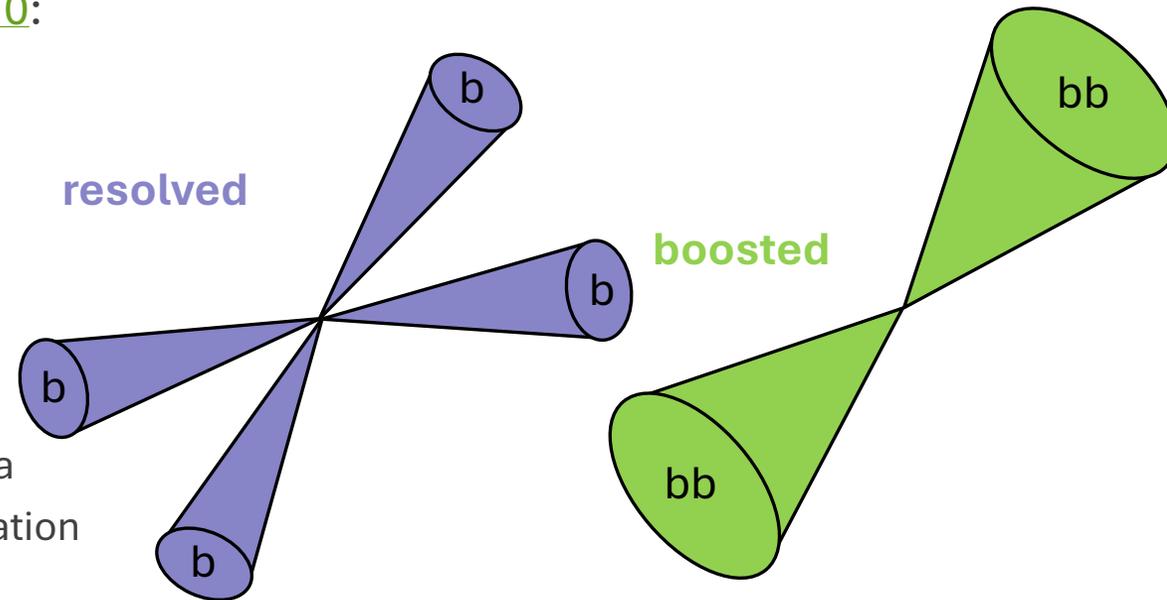
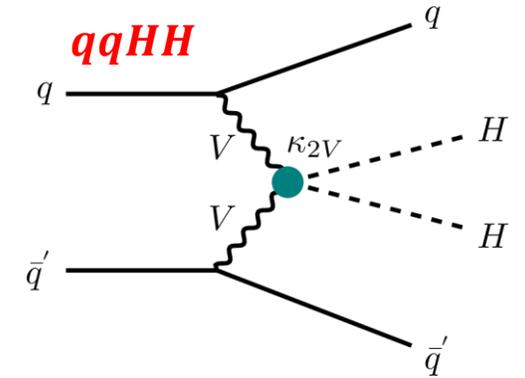
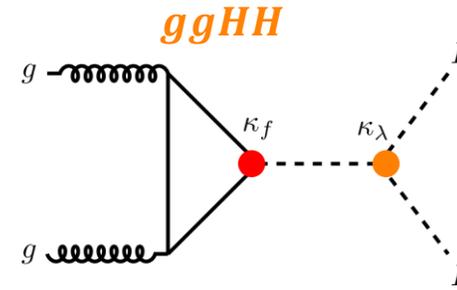
bbbb (largest branching ratio), $bb\tau\tau$, $bb\gamma\gamma$, $bbVV$, etc.

$HH \rightarrow 4b$ results presented from [CMS-PAS-HIG-24-010](#):

- Re-analysis with Run 2 (138 fb^{-1})
- Two approaches with partial Run 3 (62 fb^{-1})
- Search in 2 distinct topologies: **resolved** and **boosted**

Run 3 **improvements** coming from:

- High level trigger (HLT)
- Object tagging – b-tagging and bb-tagging
- ML techniques for modeling **large** QCD multi-jet with data
- ML techniques for Signal vs background (SvsB) discrimination





Improved Tools: Online/Offline Flavor Tagging

Higher signal efficiency for same background rejection



Online:

Incorporates a graph neural network for flavor tagging

AK4 b-tagging – [ParticleNet](#)

- **Resolved** trigger requires 2 b-tags

AK8 bb-tagging – [ParticleNet](#)

- **Boosted** trigger requires 1 bb-tag

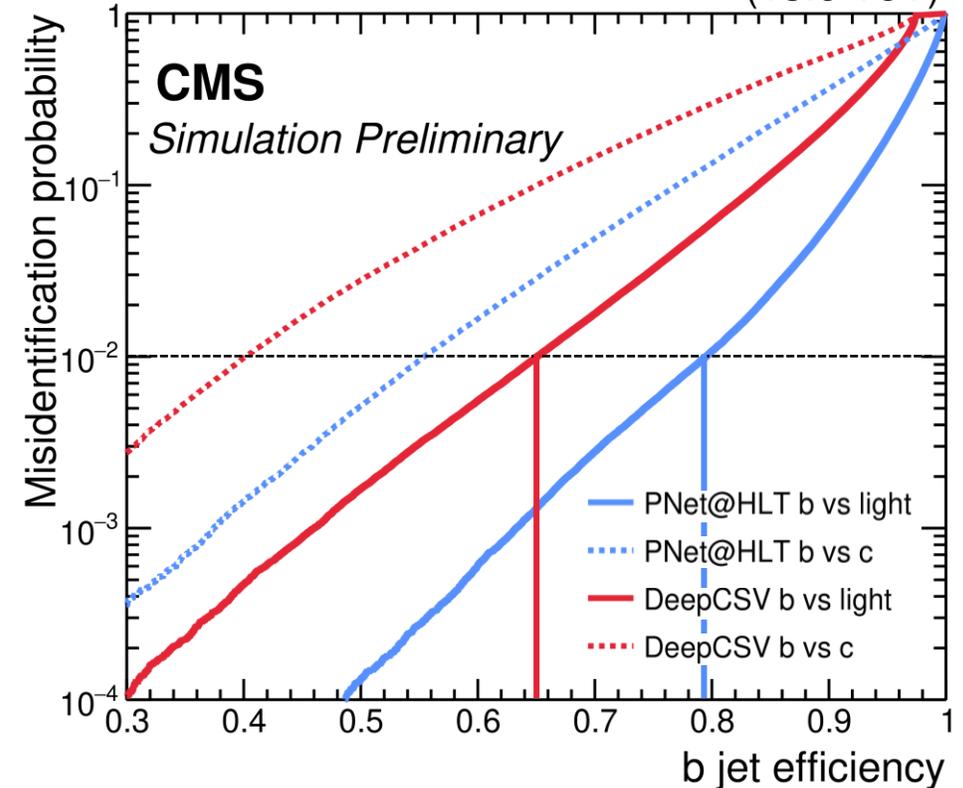
Offline:

AK4 b-tagging – [ParticleNet](#)

AK8 bb-tagging – New transformer-based [GloParT](#)

Online AK4 b-tagging – **resolved**

(13.6 TeV)





Improved Tools: High Level Trigger



With the improvement in flavor tagging, we can reduce online p_T thresholds

- **Resolved** – requires AK4 jets as low as 30 GeV (40 in Run 2)
- **Boosted** – requires AK8 jets as low as 230 GeV (330 in Run 2)

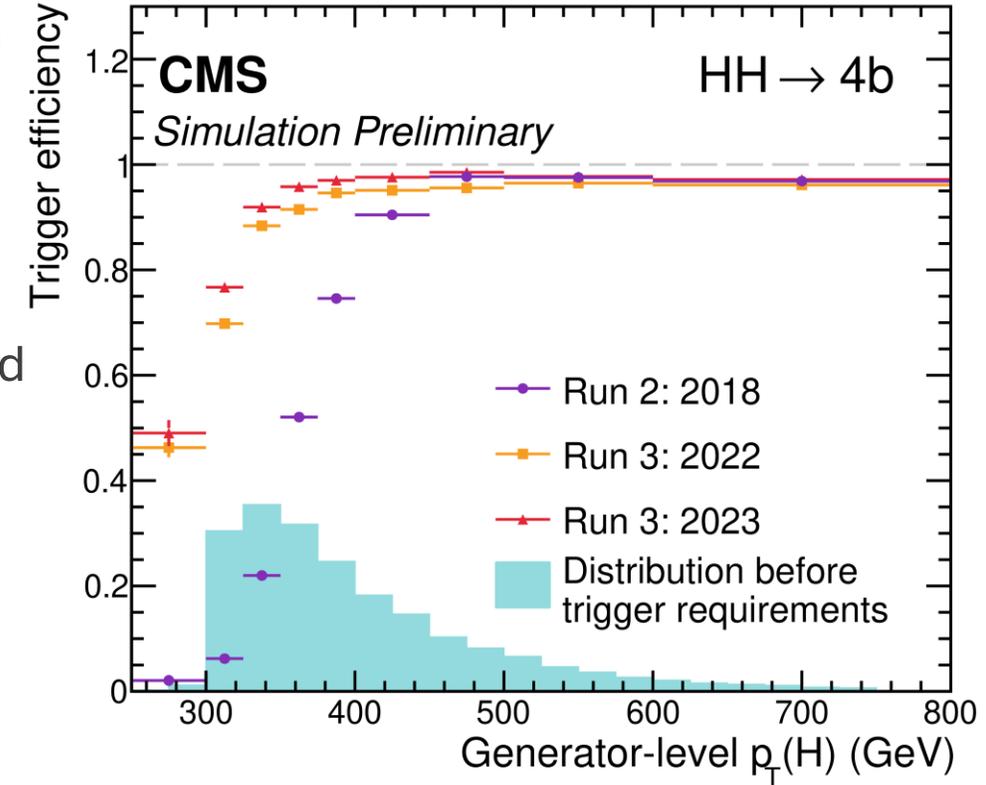
Significant gain in signal efficiency at HLT

Implemented the **parking** strategy in 2023 for the resolved trigger

- allows a larger rate allocation

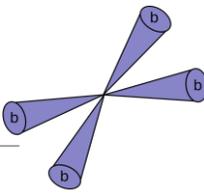
Resolved Trigger	2018	2022	2023
Rate	15 Hz	60 Hz	150 Hz

Online AK8 trigger – **boosted**





Run 2 Resolved Re-analysis



Signal (SR) and control (CR) regions defined in M_H plane

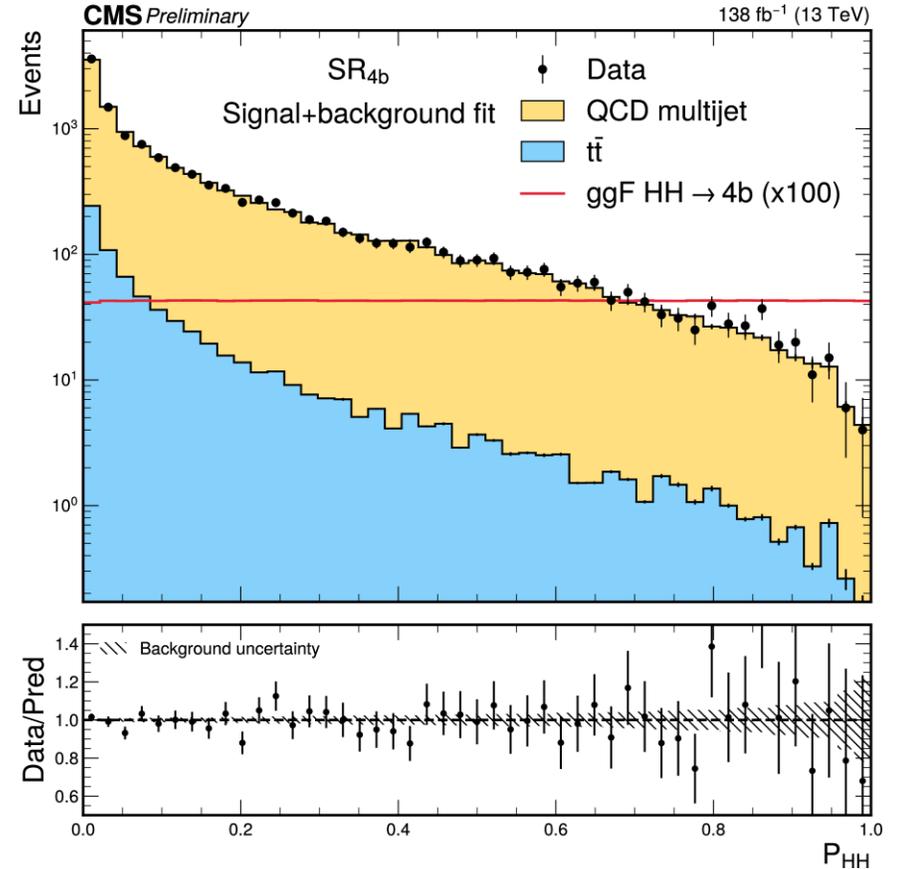
4 b-tagged QCD Multi-jet modeled from 3 b-tagged data using NN reweighting in CR

SvsB discriminant used to extract results

Inclusion of ZZ/ZH enriched regions for validation

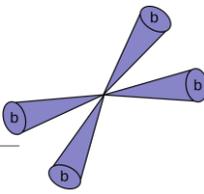
Largest systematics coming from statistics and background systematics

Observed (Expected) Upper Limit at 95% CL	μ_{ggHH}
Re-analysis	10 (5.9)
Compared to the previous Run 2 result : 3.9 (7.8) 25% improvement	





Run 3 Resolved Analysis



Similar approach to Run 2 re-analysis

More robust NN architecture for NN reweighting and SvsB

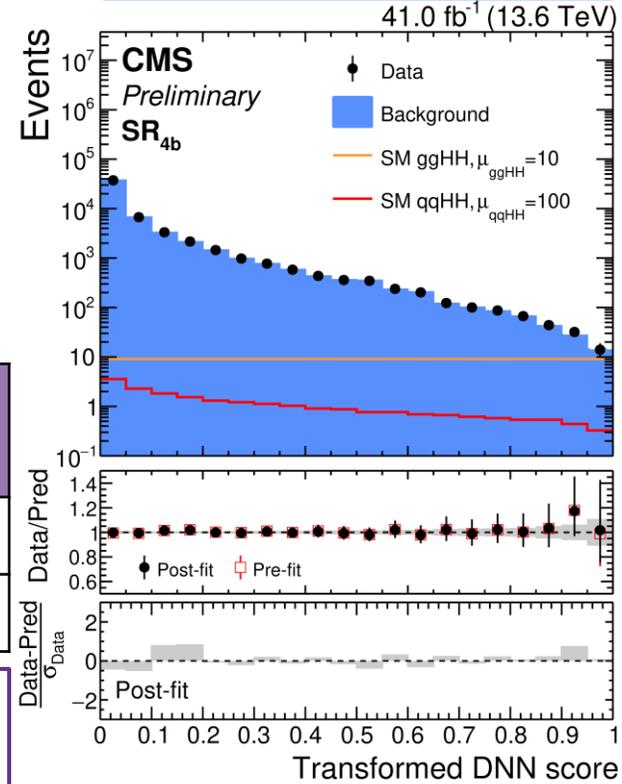
Looser triggers allow the 4 b-tagged background to be modeled by 2 b-tagged data

Largest systematics coming from statistics and background systematics

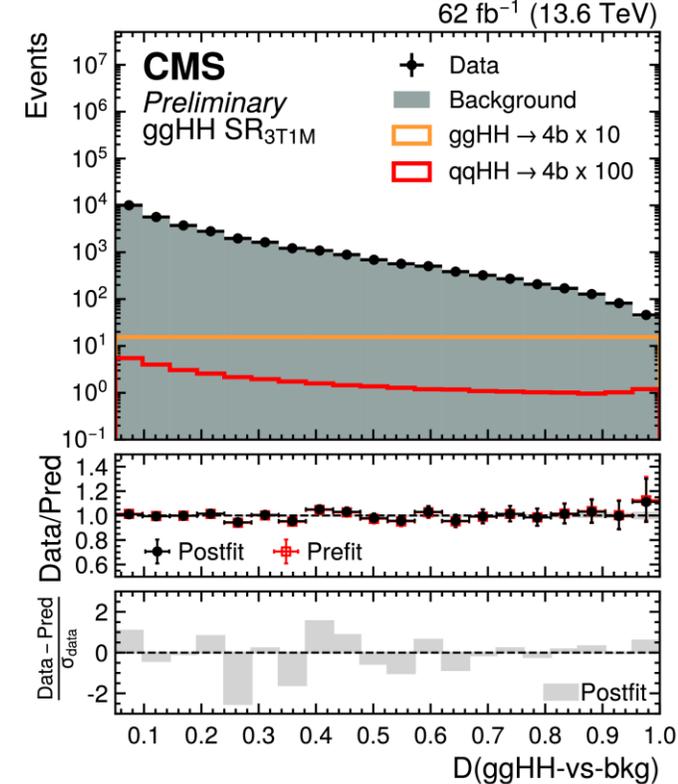
Observed (Expected) Upper Limit at 95% CL	μ_{HH}
First Approach	6.3 (5.3)
Second Approach	5.0 (6.1)

Scaling 5.3 to $138 \text{ fb}^{-1} \rightarrow 3.5$
 Compared to the Run 2 re-analysis: 3.9 (7.8)
40% improvement

First Approach
Focus on HH

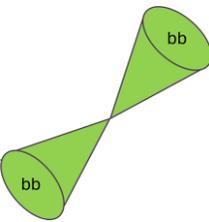


Second Approach
Includes ZZ/ZH





Run 3 Boosted Analysis



SRs and CRs defined with M_{H1} and bb-tag score

QCD Multi-jet modeled in two ways:

1. NN reweighting like in the resolved
2. Parametric alphabet method

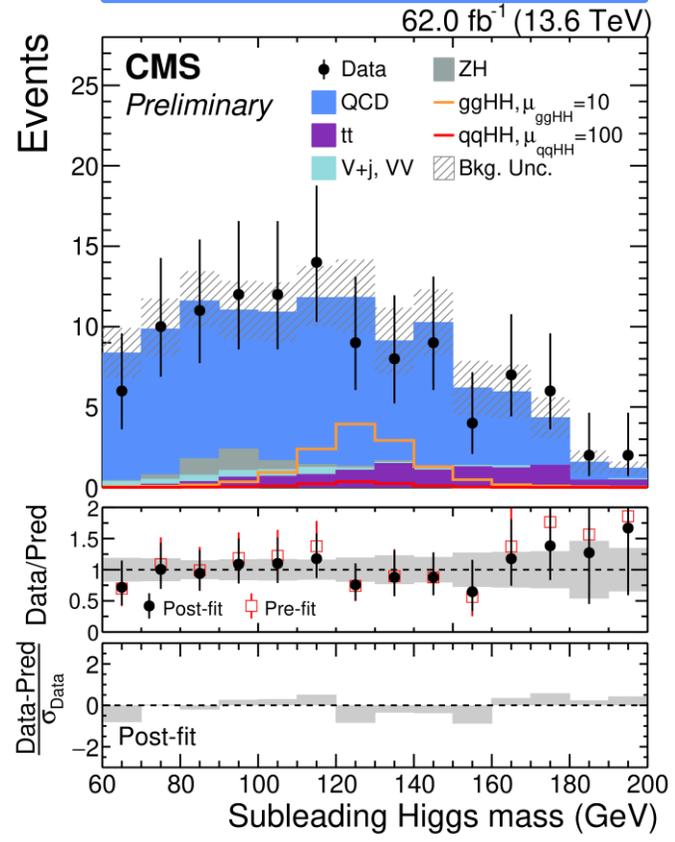
SvsB used to reduce background

Largest systematics coming from statistics

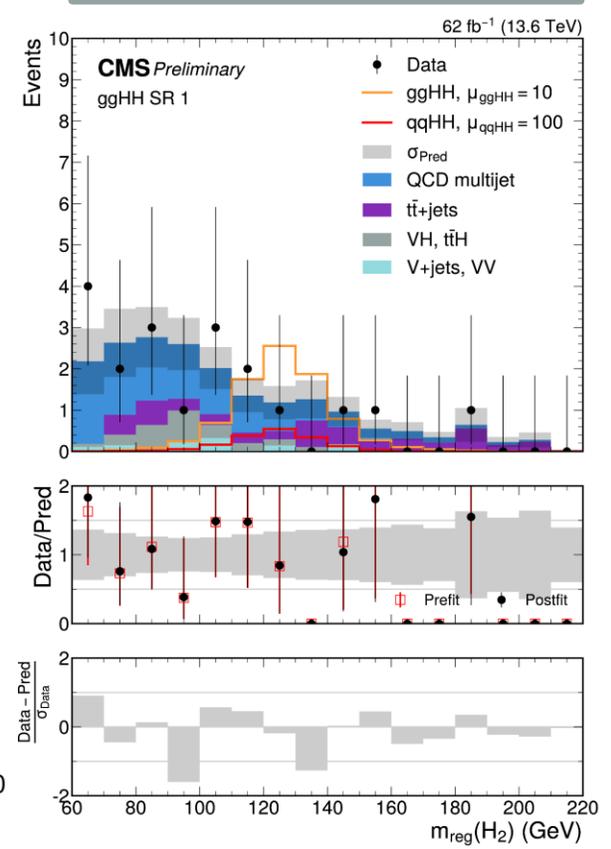
Observed (Expected) Upper Limit at 95% CL	μ_{HH}
First Approach	5.3 (7.1)
Second Approach	11 (8.1)

Scaling 7.1 to 138 fb⁻¹ → 4.4
 Compared to the previous Run 2 result: 9.9 (5.1)
13% improvement

First Approach NN QCD



Second Approach Parametric QCD





Combined Run 3 Results



Considerable improvements have been made in both **resolved** and **boosted** topologies

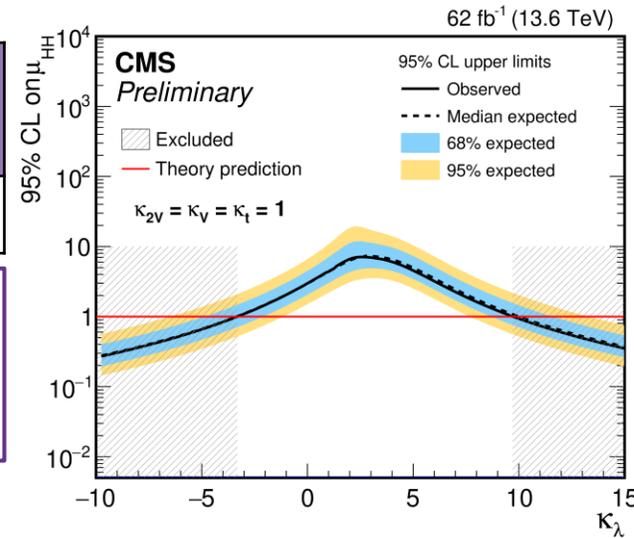
Two different approaches provide panorama of the different analysis components

For equivalent luminosity, Run 3 shows improvements of 13 – 40% when compared to Run 2

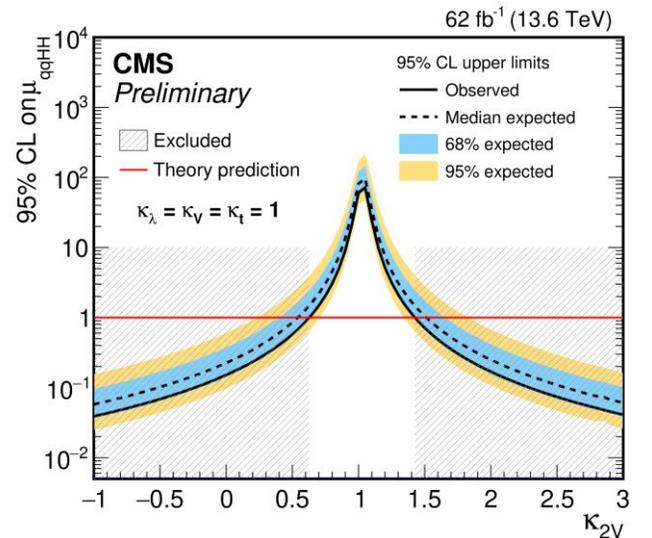
Combination of the most sensitive $HH \rightarrow 4b$ results from both **resolved** and **boosted**

Observed (Expected) Upper Limit at 95% CL	μ_{HH}
Run 3 Combination	4.4 (4.4)
Scaling 4.4 to 138 fb ⁻¹ → 2.9 Compared to the previous Run 2 result: 7.0 (4.3) 30% improvement	

Expected 95% CL constraints
 $\kappa_\lambda \in [-3.4, 10.0]$



Expected 95% CL constraints
 $\kappa_{2V} \in [-0.54, 1.51]$





Presented early results in the $HH \rightarrow 4b$ channel

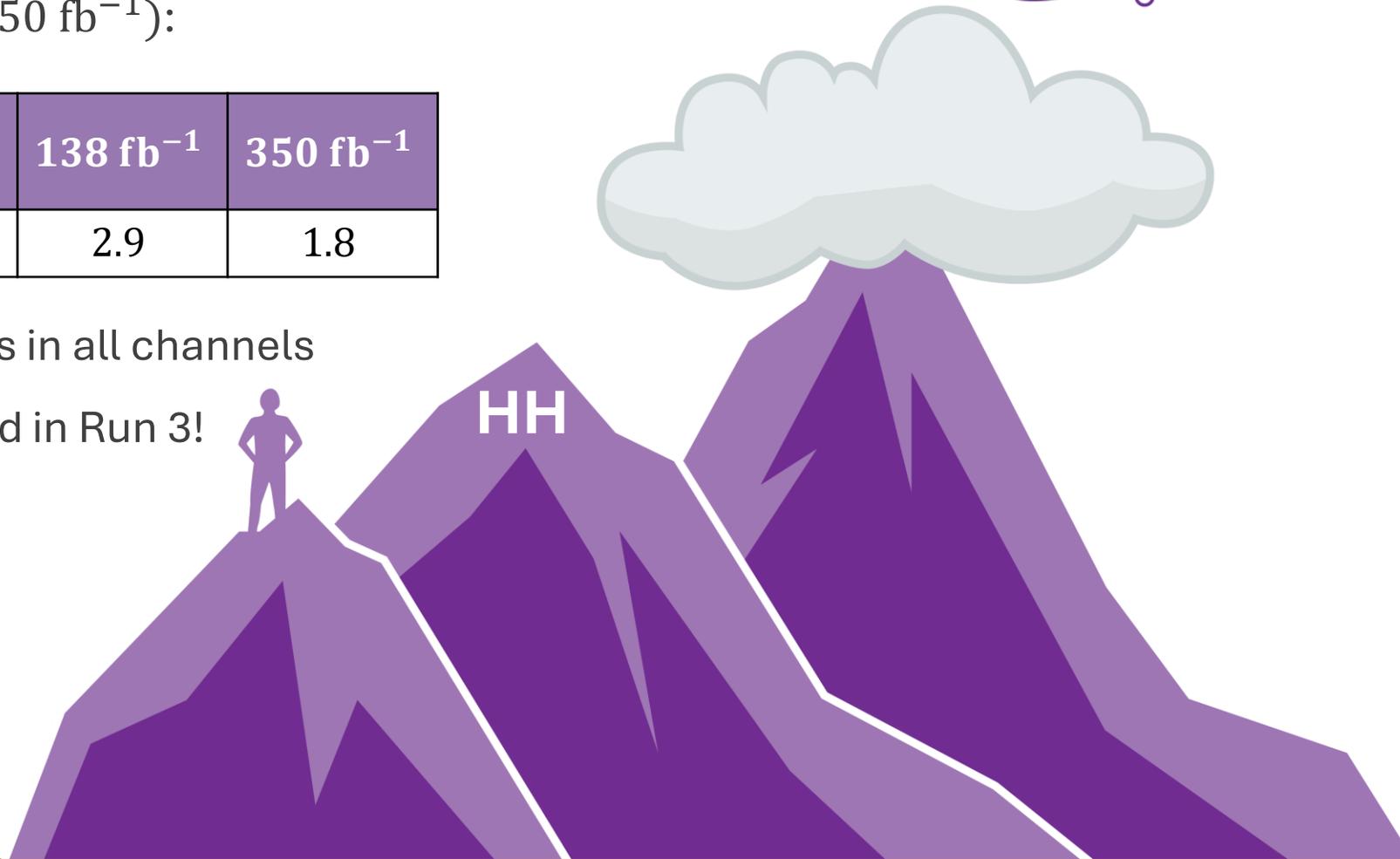
Run 3 proving to be one of the most sensitive datasets

Projections for full Run 3 ($\sim 350 \text{ fb}^{-1}$):

Expected Upper Limit at 95% CL	62 fb^{-1}	138 fb^{-1}	350 fb^{-1}
μ_{HH}	4.4	2.9	1.8

Run 3 is seeing improvements in all channels

SM sensitivity may be reached in Run 3!



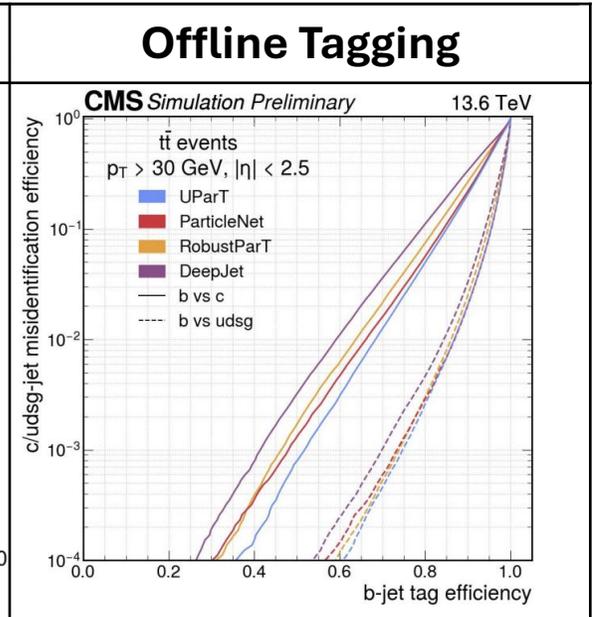
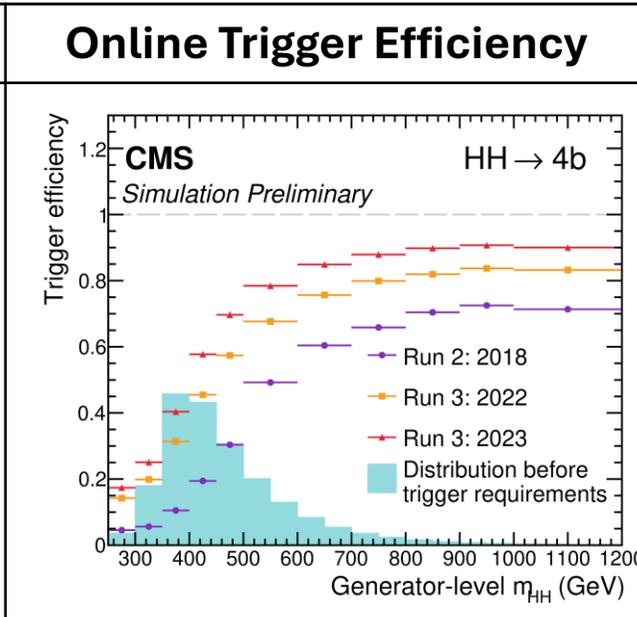
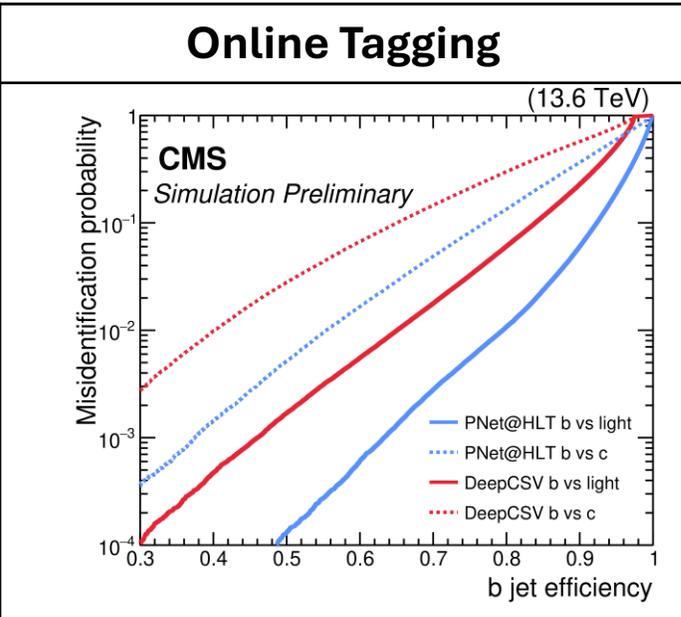
Backup



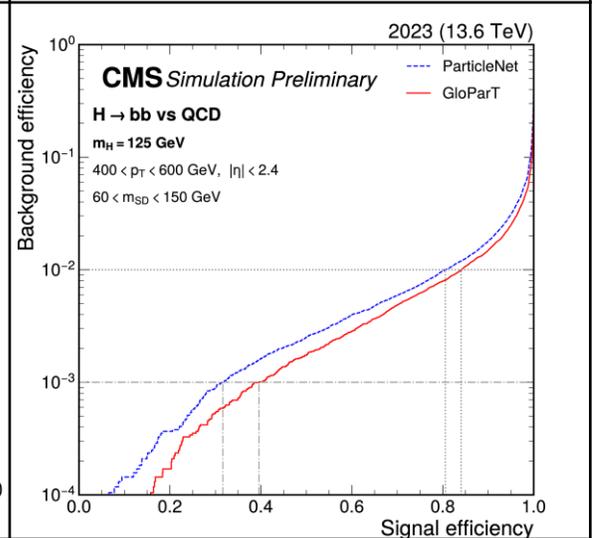
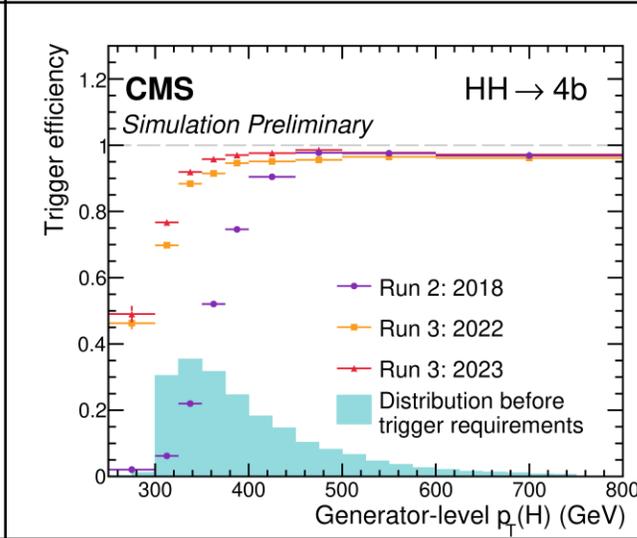
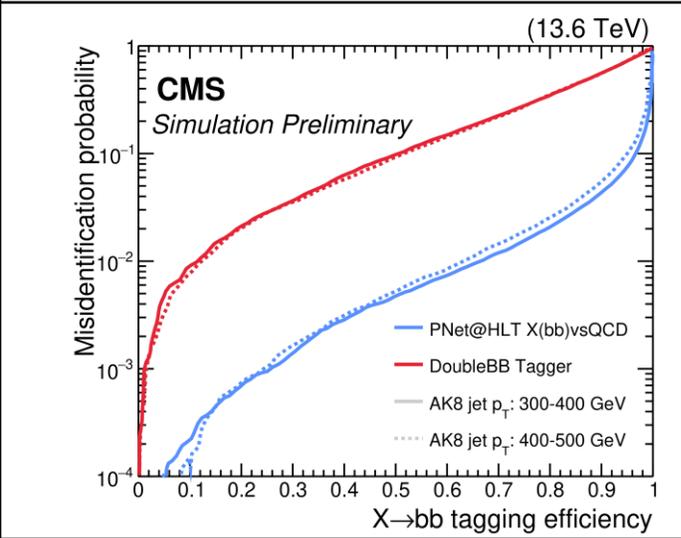
Run 3 Improvements in Trigger and Tagging

UF

Resolved



Boosted





NN Reweighting



Background is estimated using a NN classifier to reweight 2b data

The classifier is trained in the CR to classify CR_{2b} data from CR_{4b} data

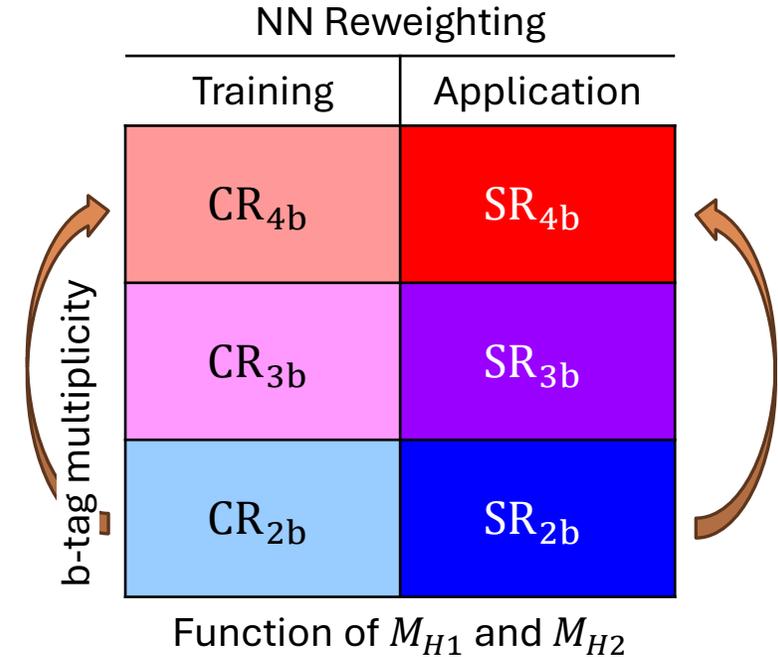
Event weights are derived using the likelihood ratio:

$$w_{4b}(\vec{x}) = \frac{P(\vec{x}|CR_{4b})}{P(\vec{x}|CR_{2b})}$$

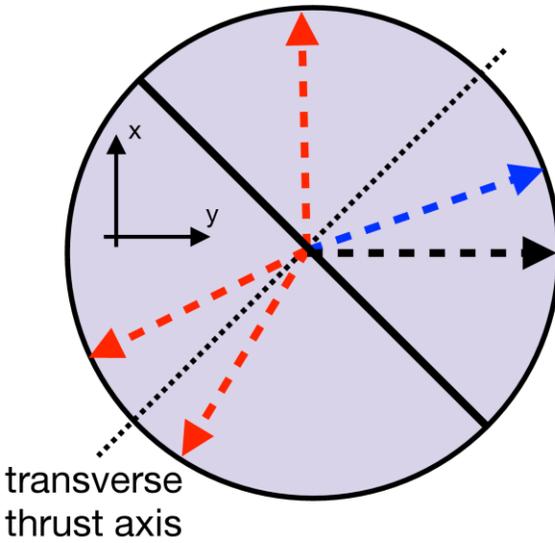
These weights are used to reweight the SR_{2b} data to model the background in the SR_{4b} region

A second training is performed between the 2b and 3b data as validation

The closure in the 3b data is used to derive systematics

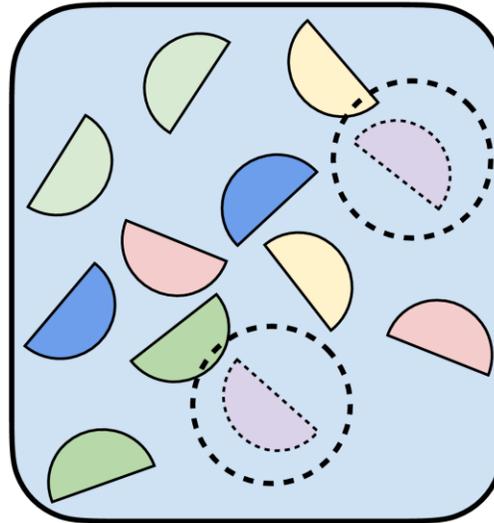


Original three-tag event
split into two hemispheres



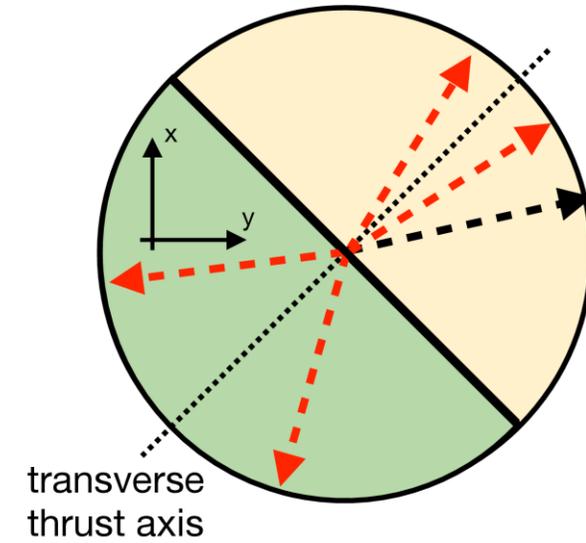
---▶ b-tagged jets

Hemisphere library
made from four-tag events
filled in 1st pass, queried on 2nd



---▶ pseudo-tagged jets

Mixed Event
using replaced hemispheres



---▶ non b-tagged jets



Systematic Uncertainties



Systematic uncertainties on the background model arise from:

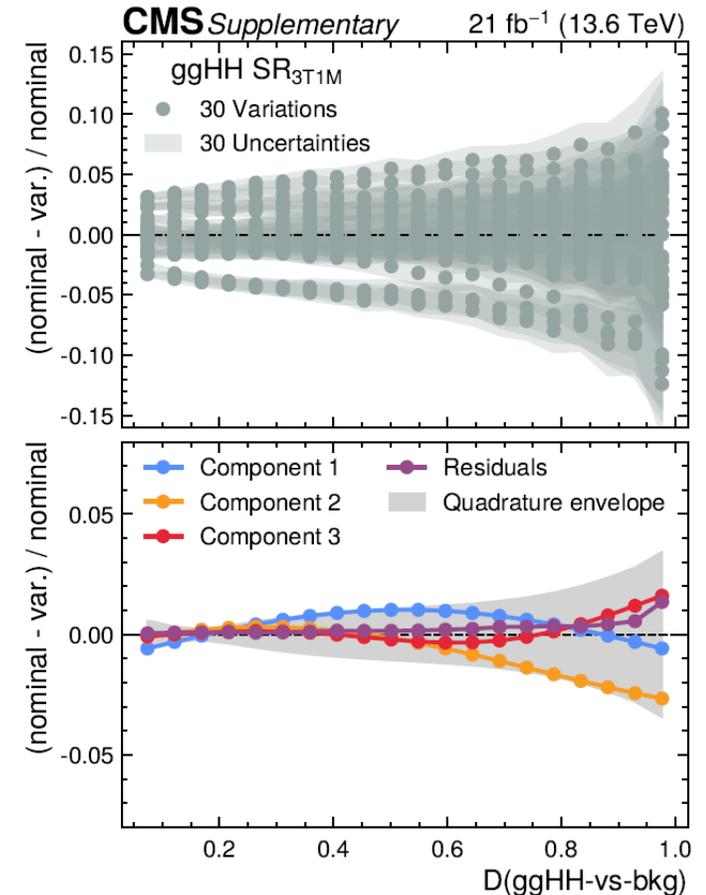
- Variance: due to limited statistics in the CRs
- Bias: due to extrapolations from the CRs to the SRs

Variance is estimated using k-fold cross validation.

- Differences are parameterized with Fourier basis functions
- Fit each of the k-fold models with linear combination of basis functions
- Take RMS of the fit coefficients → Uncorrelated shape uncertainties
- Residual non-closure is added as additional bin-by-bin uncertainty

Bias is estimated by fitting 3 b-tagged data with background model

- Bias quantified with the pull of each Fourier coefficient
- Include a duplicate of each Fourier component scaled by the bias term
- Treated uncorrelated from Variance uncertainties





Parametric Alphabet Method



Data-driven background estimation for QCD using the parametric alphabet method.

Define a **transfer factor** $R_{p/f}$ between the SR and CR_{QCD} as a function of M_{H2}

$$N_{SR}^{QCD}(M_{H2}) = \epsilon^{QCD} R_{p/f}(M_{H2}) N_{CR}^{QCD}(M_{H2})$$

$R_{p/f}$ is parameterized as a Bernstein polynomial and fit to data in the CR_{QCD}

$$R_{p/f}(M_{H2}) = \sum_{k=0}^n a_k b_{k,n}(M_{H2})$$

$$b_{k,n}(x) = \binom{n}{k} x^k (1-x)^{n-k}$$

Parametric Method

