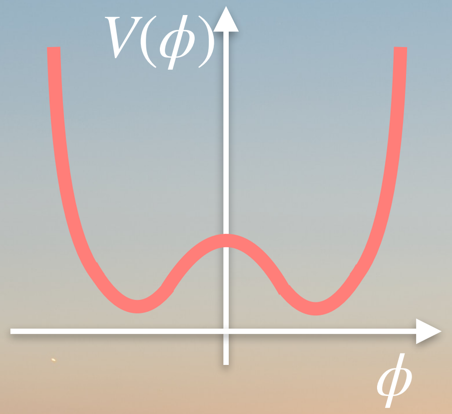
Hunting for di-Higgs in hadronic final states



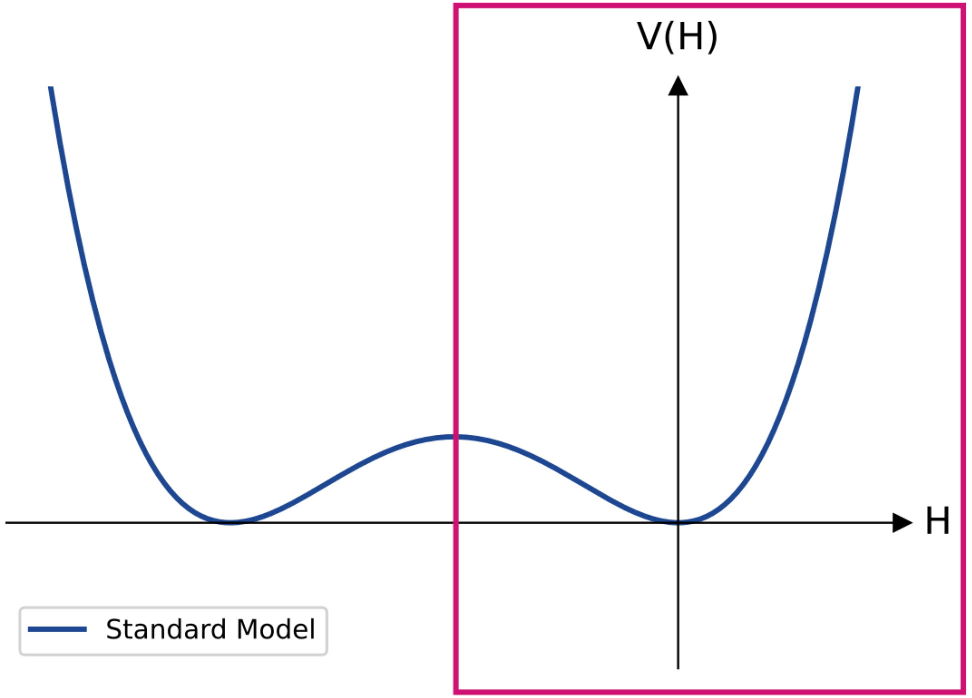
Nicole Hartman
nicole.hartman@tum.de

Workshop on Machine Learning and High Energy Physics
Austrian Academy of Sciences
13 Dec 2023

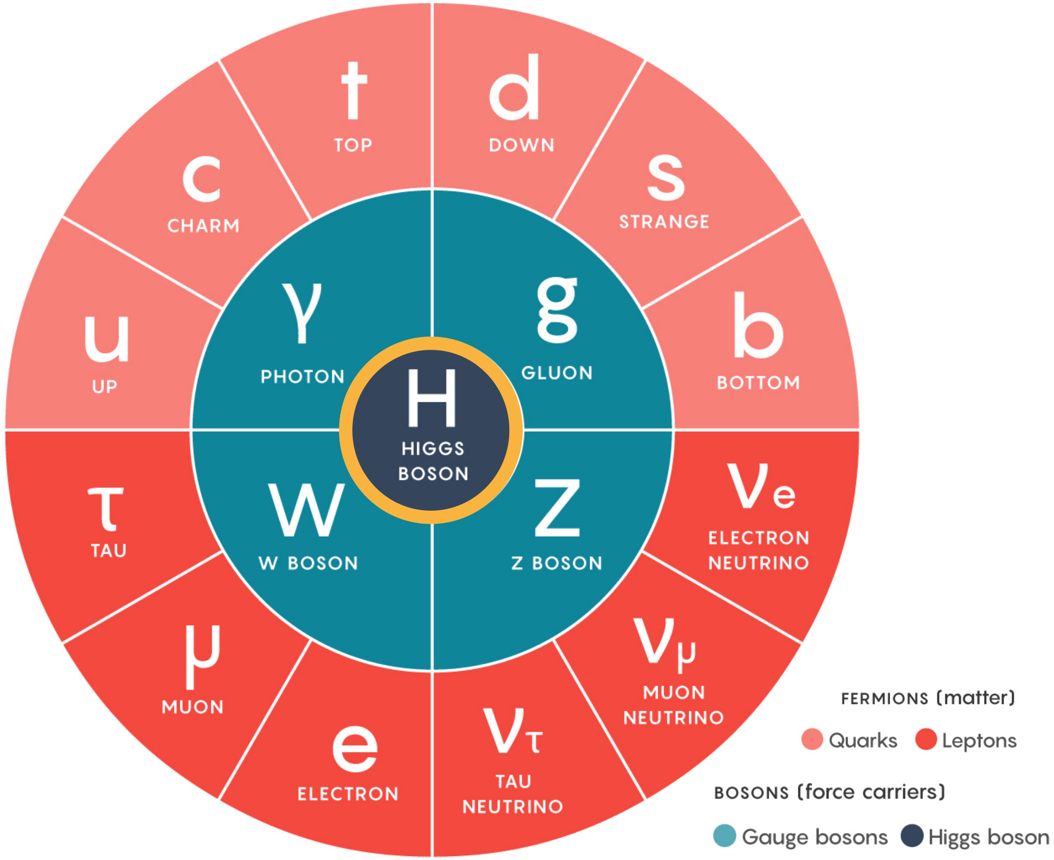
Photo by Jacek Dylag on Unsplash

Motivation

$$\mathscr{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}\gamma_\mu D^\mu\psi + |D_\mu\phi|^2 - \boxed{V(\phi)} + (y_{ij}\bar{\psi}_i\psi_j + \text{h.c.})$$

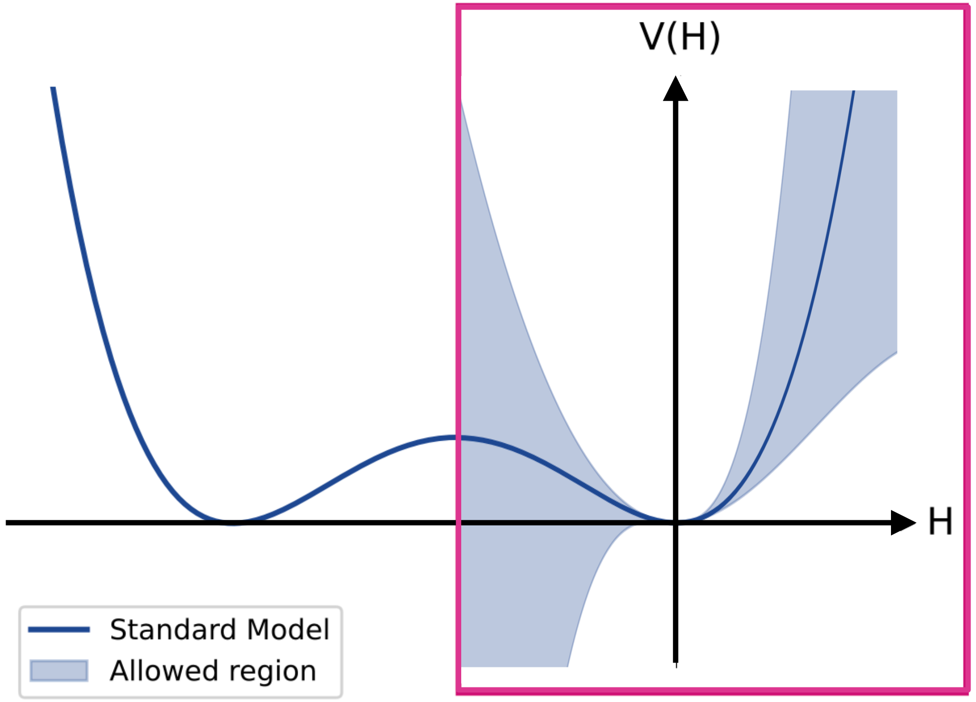


Graphic from B. Moser

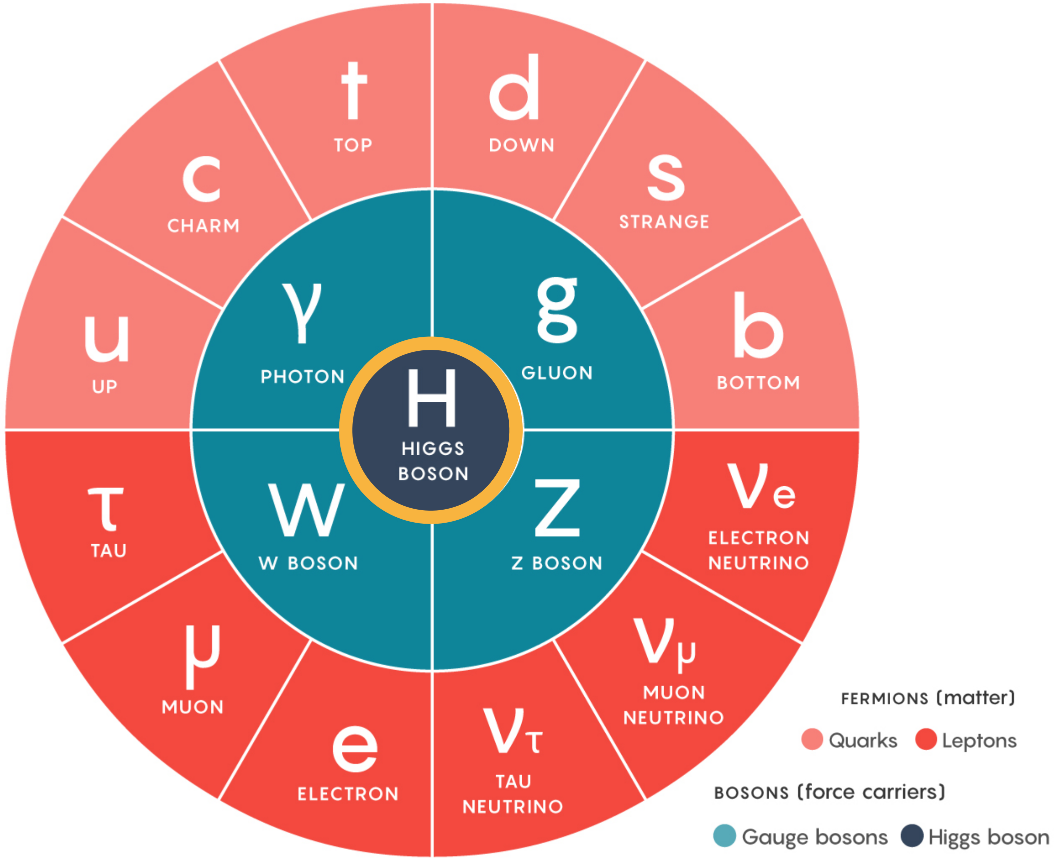


Motivation

$$\mathscr{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}\gamma_\mu D^\mu\psi + |D_\mu\phi|^2 - \boxed{V(\phi)} + (y_{ij}\bar{\psi}_i\psi_j + \text{h.c.})$$

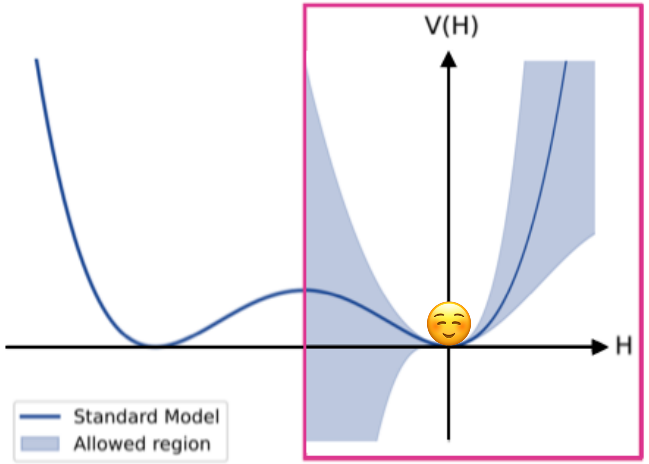


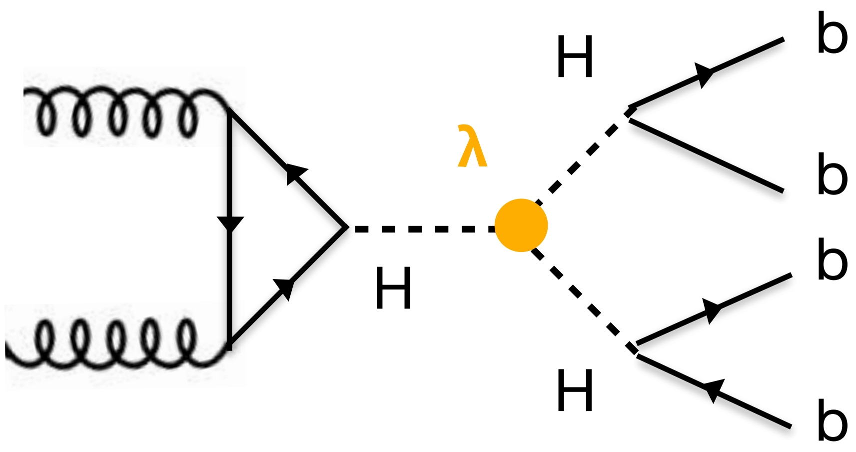
Graphic from B. Moser



Where to look

$$V(\phi) = \mu^2 h(x)^2 + \lambda v h(x)^3 + \frac{1}{4}\lambda h(x)^4$$





Focus for today!

Higgs 1 decay

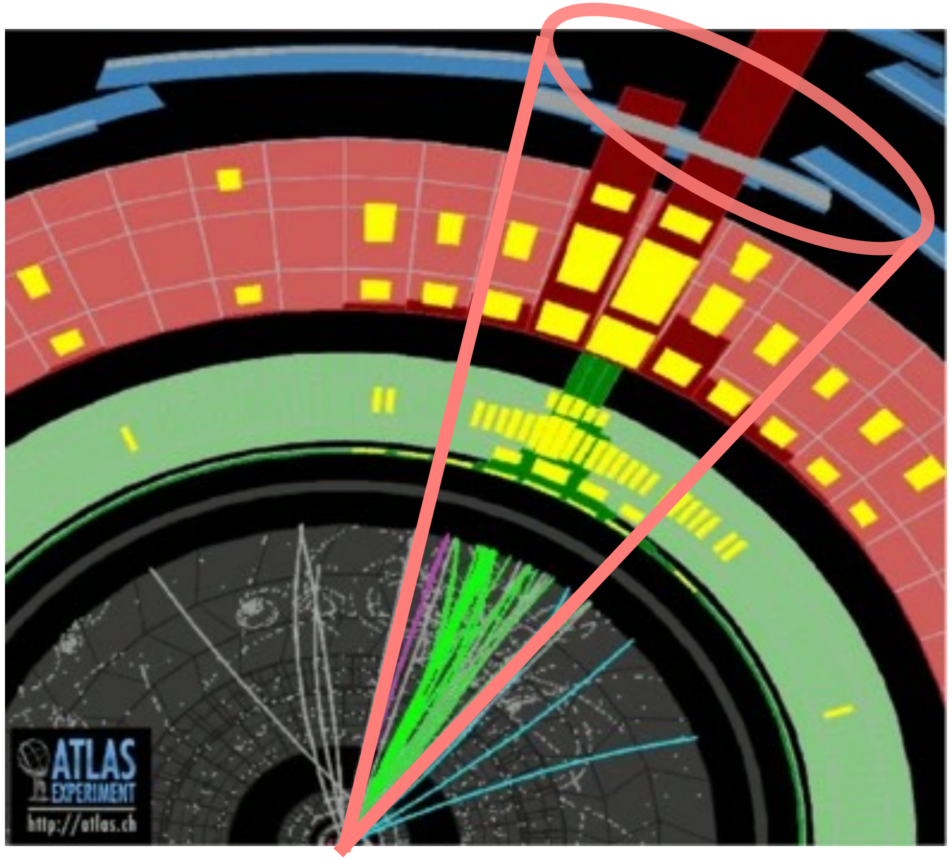
Higgs 2 decay

	bb	WW	ττ	ZZ	γγ
bb	34%				
WW	25%	4.6%			
ττ	7.3%	2.7%	0.39%		
ZZ	3.1%	1.1%	0.33%	0.069%	
γγ	0.26%	0.10%	0.028%	0.012%	0.0005%

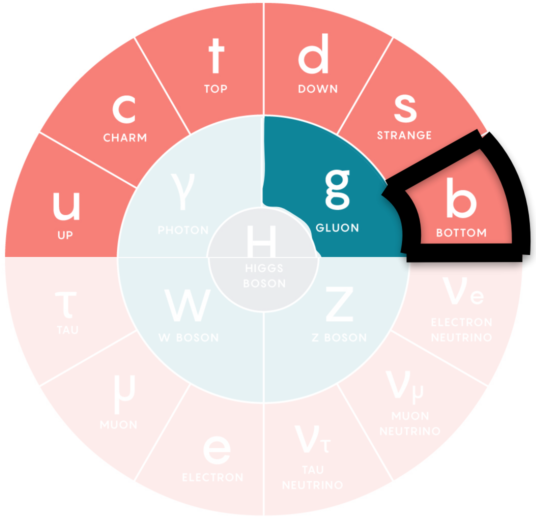
Other main HH channels

The European Physical Journal C
Vol 73 3 (2013) 2304

Jets

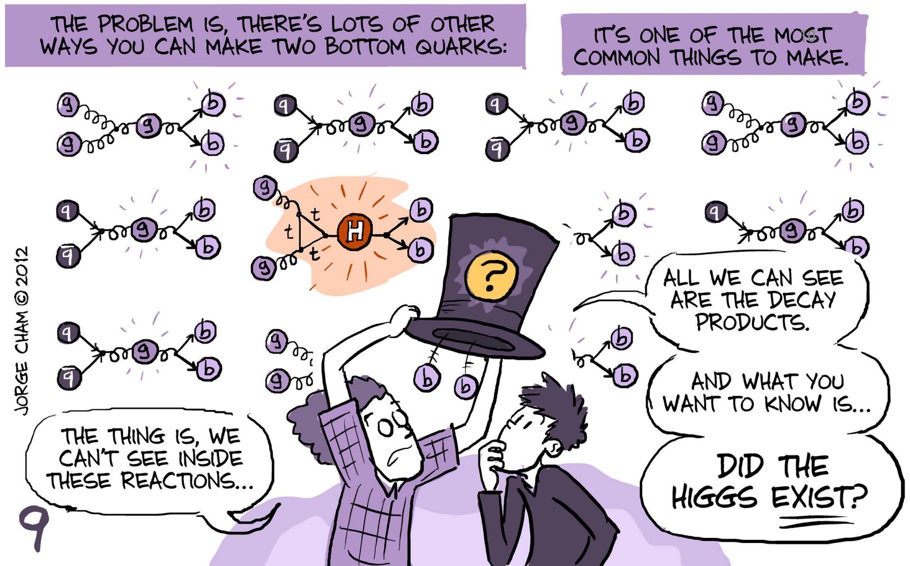


In the detector, quarks create jets



Our favorite jets

Also lots of ways to get **b-jets** other than Higgs bosons!

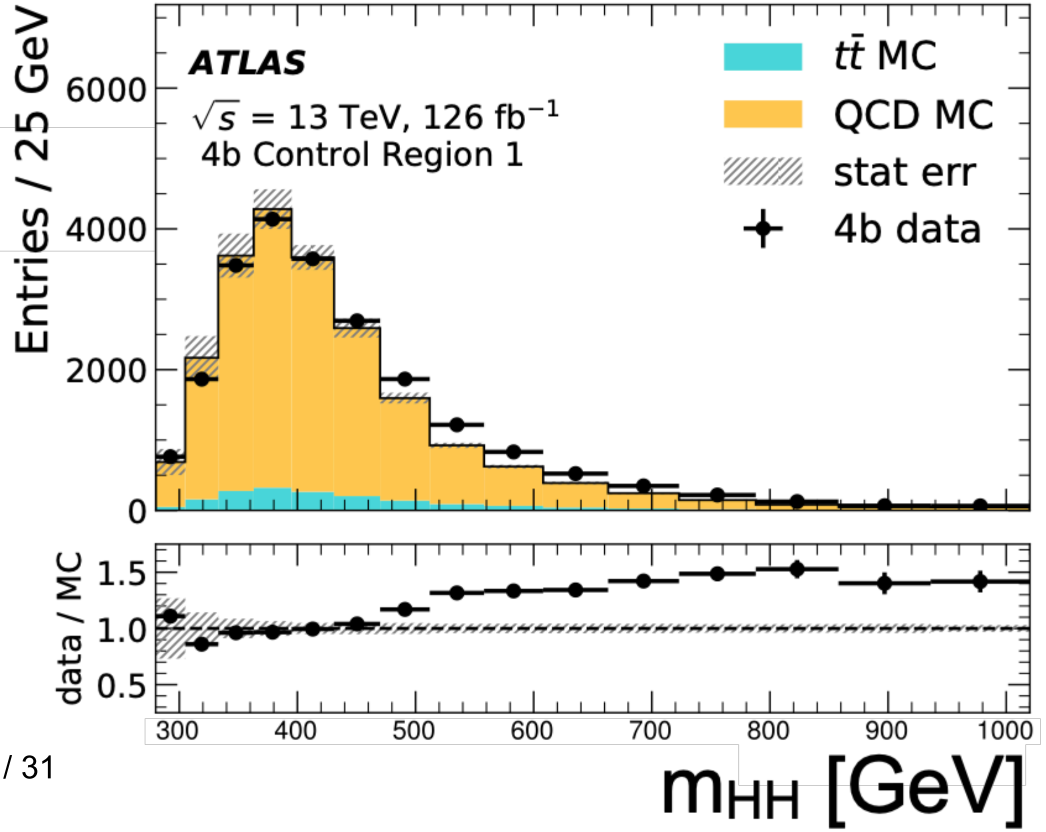


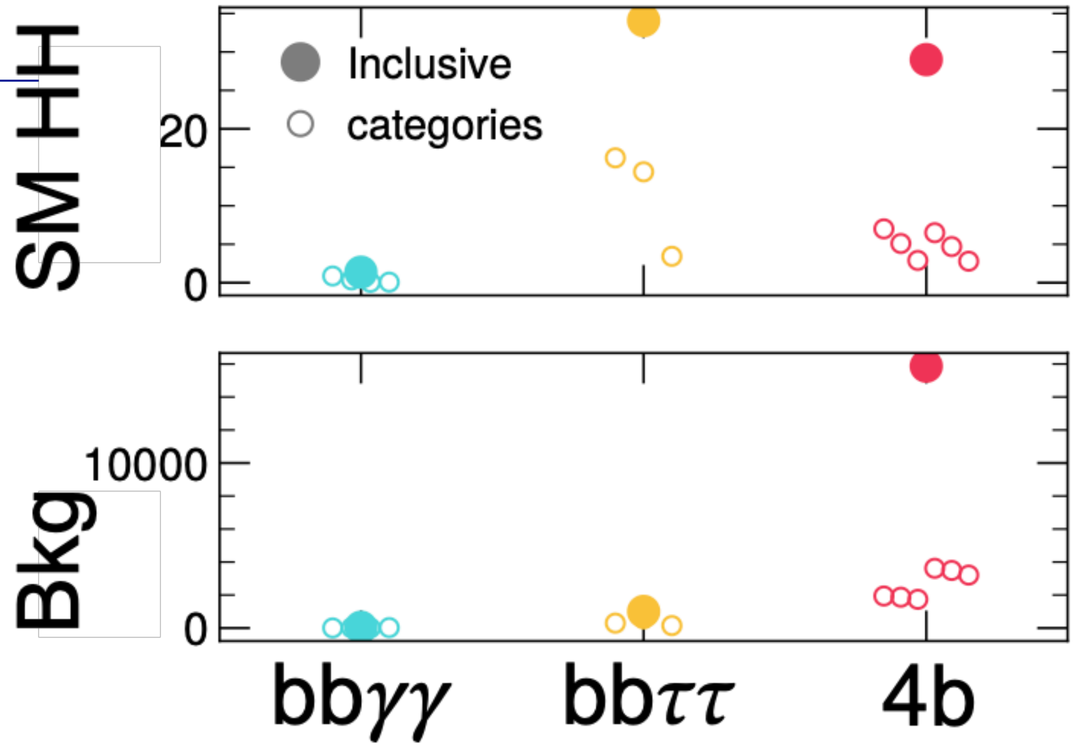
From ATLAS $bb\gamma\gamma$, $bb\tau\tau$, and 4b analyses.

Challenges

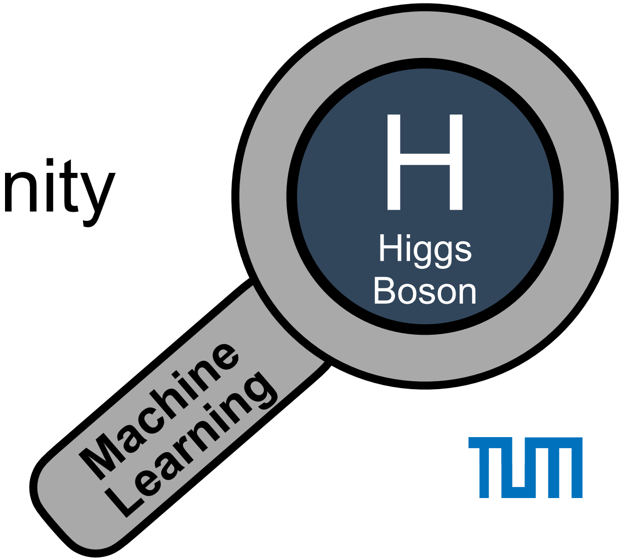
Analyses with jets have large backgrounds ➡

which are challenging to simulate from first principles ⬇





Key opportunity for ML

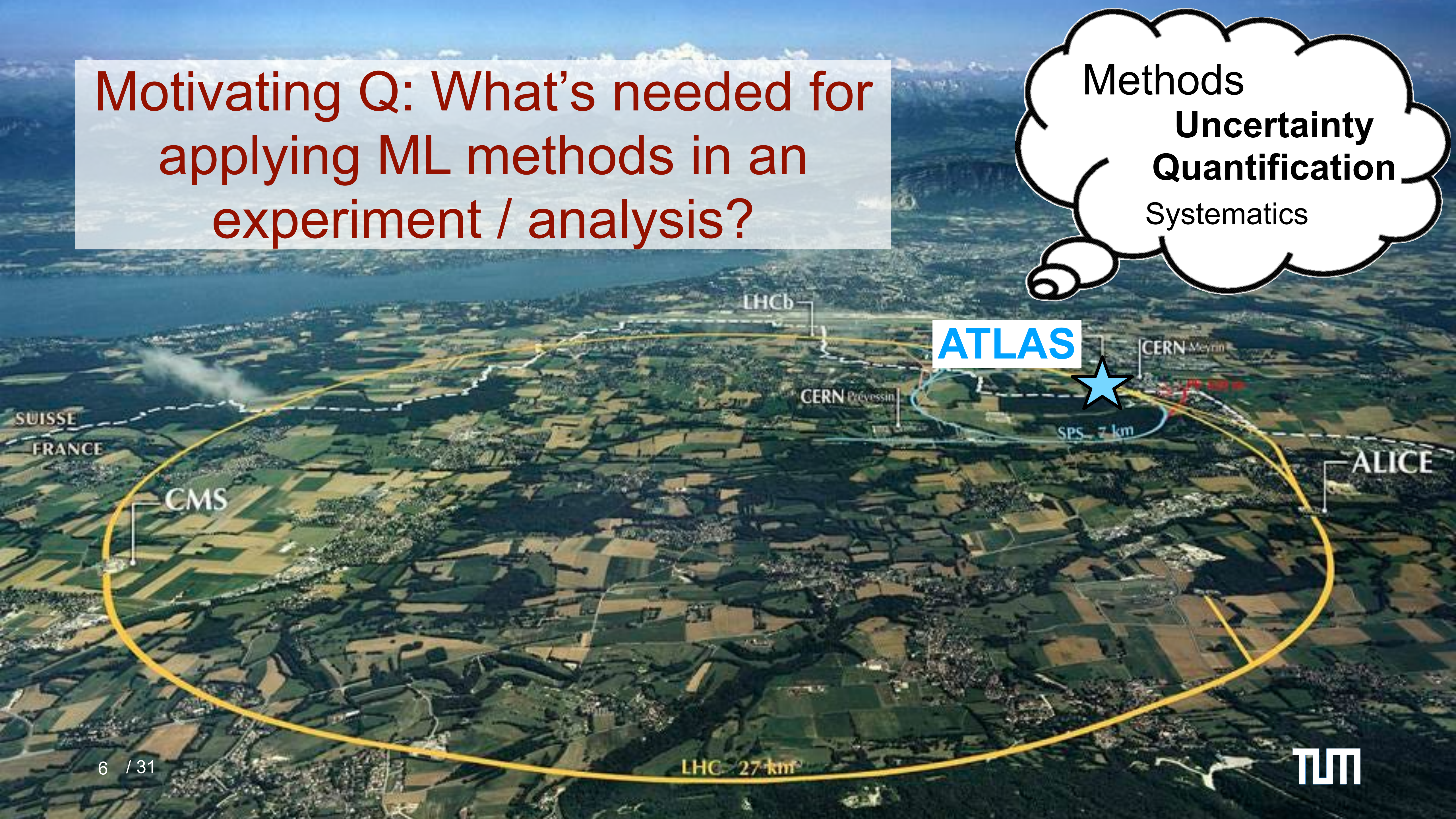


Motivating Q: What's needed for applying ML methods in an experiment / analysis?

Methods

Uncertainty Quantification

Systematics



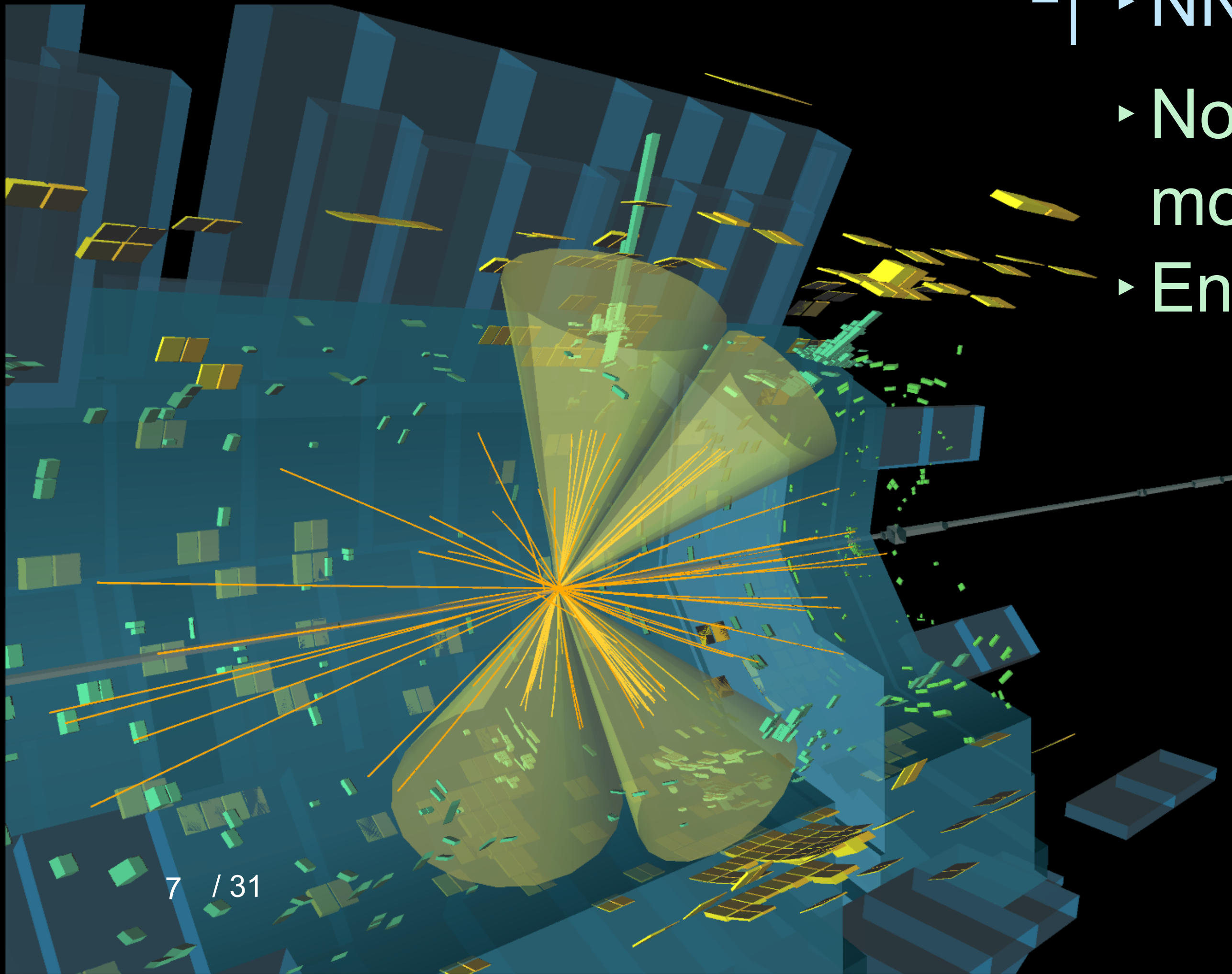
HH → 4b analysis

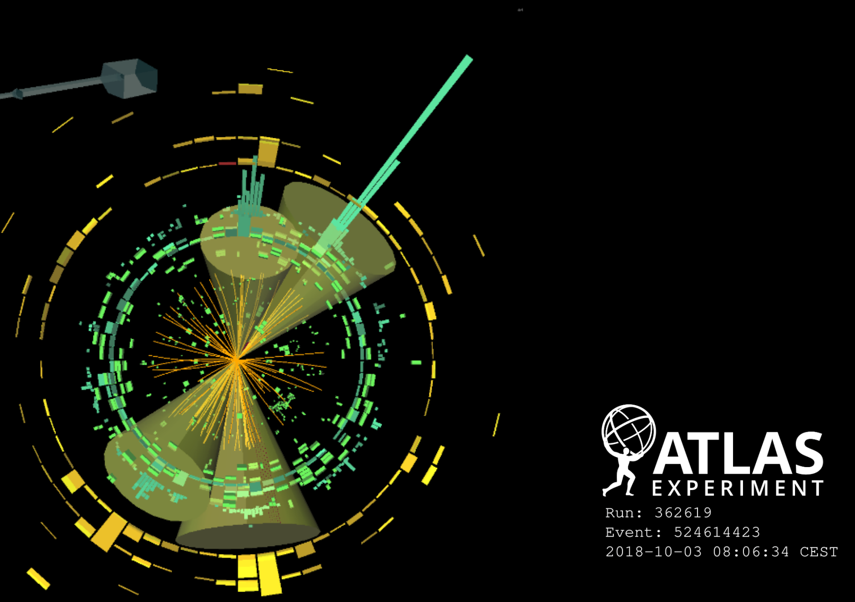
Run 2

- Birds-eye view of analysis
- NN reweighting for background modelling

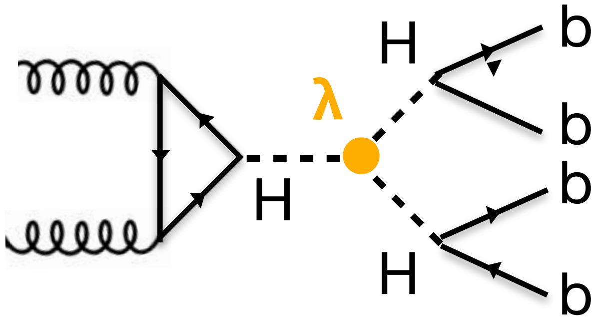
- Normalizing Flows for background modelling
- End-to-end analysis

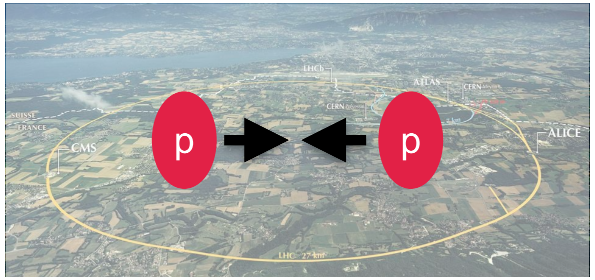
Run 3



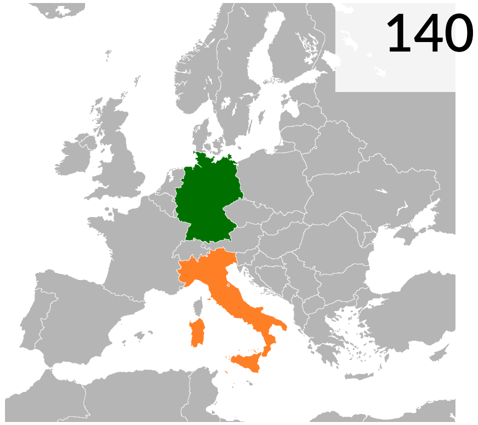


The scale of the problem





$$\frac{\text{1 signal event / (3 hour)}}{\text{1 billion collisions / second}} = 10^{-13}\ \frac{\text{signal}}{\text{background}}$$



140 million people in Germany + Italy

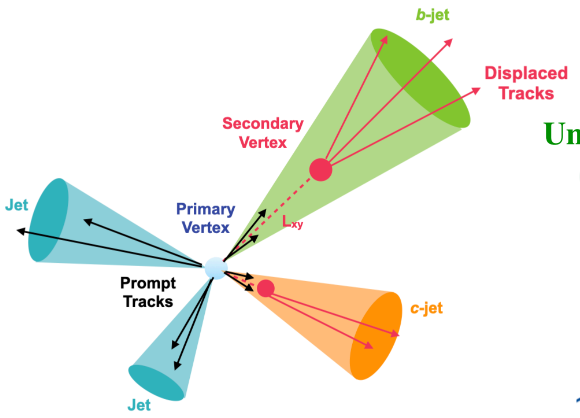
100,000 hairs / person

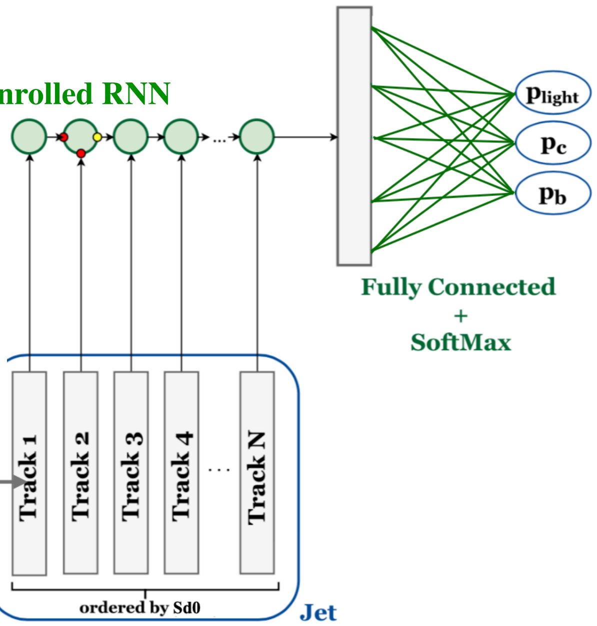
Analysis overview

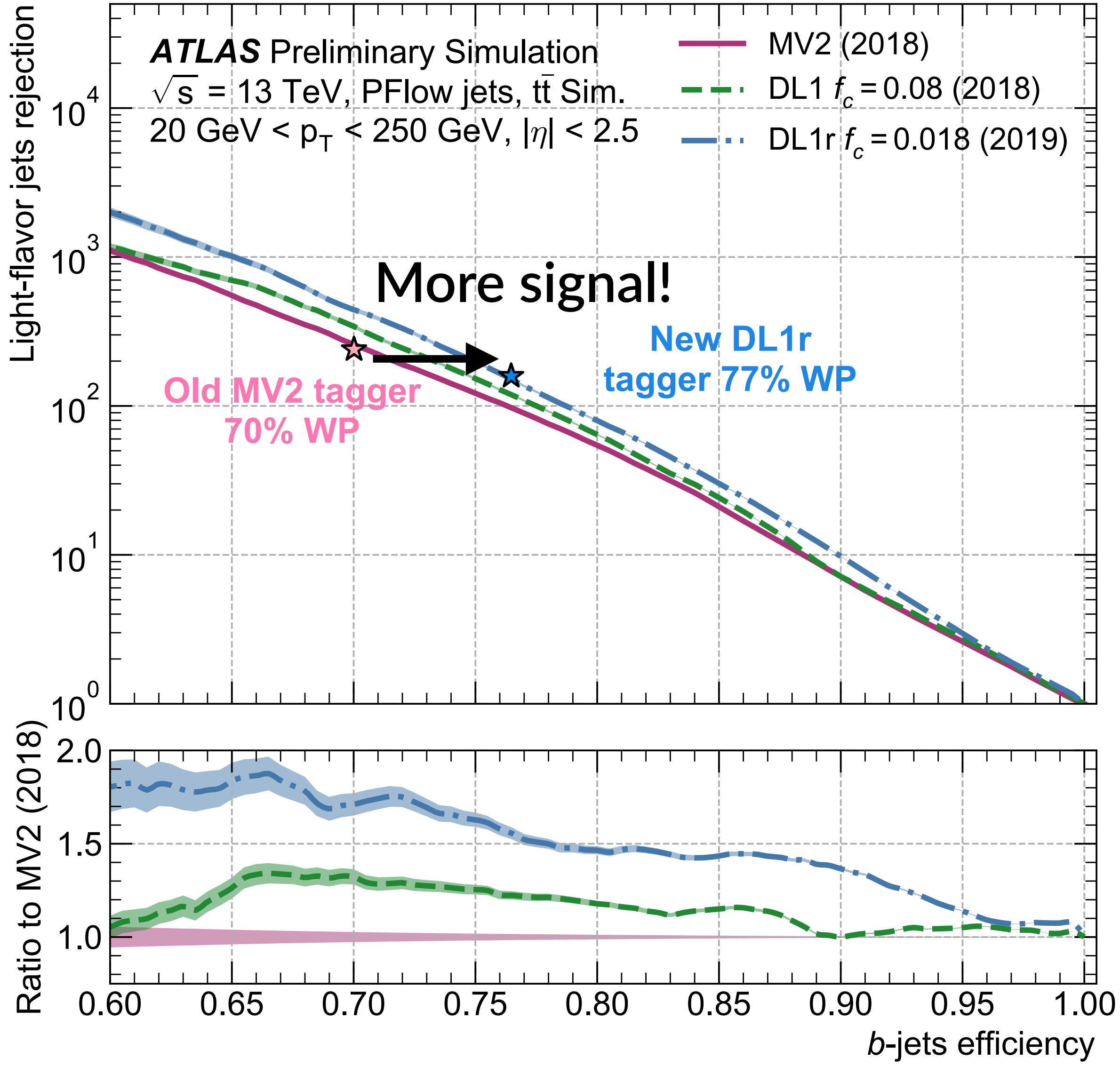
Trigger

↓

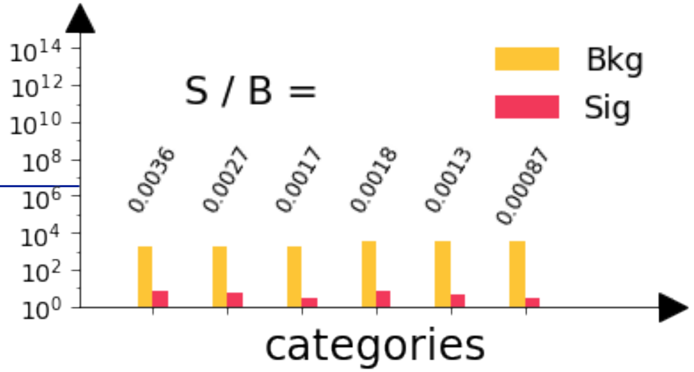
b-jets







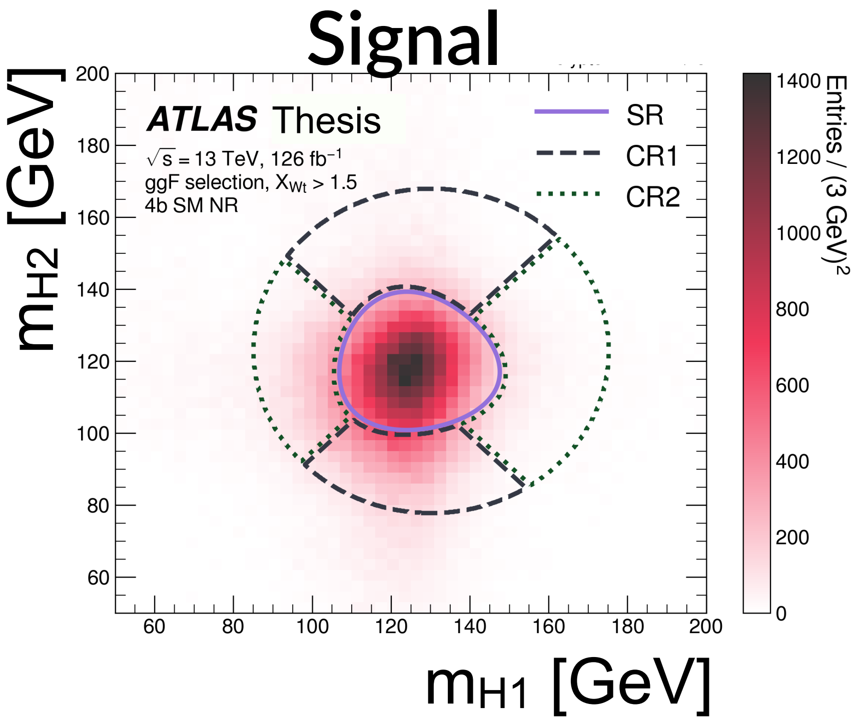
Analysis overview

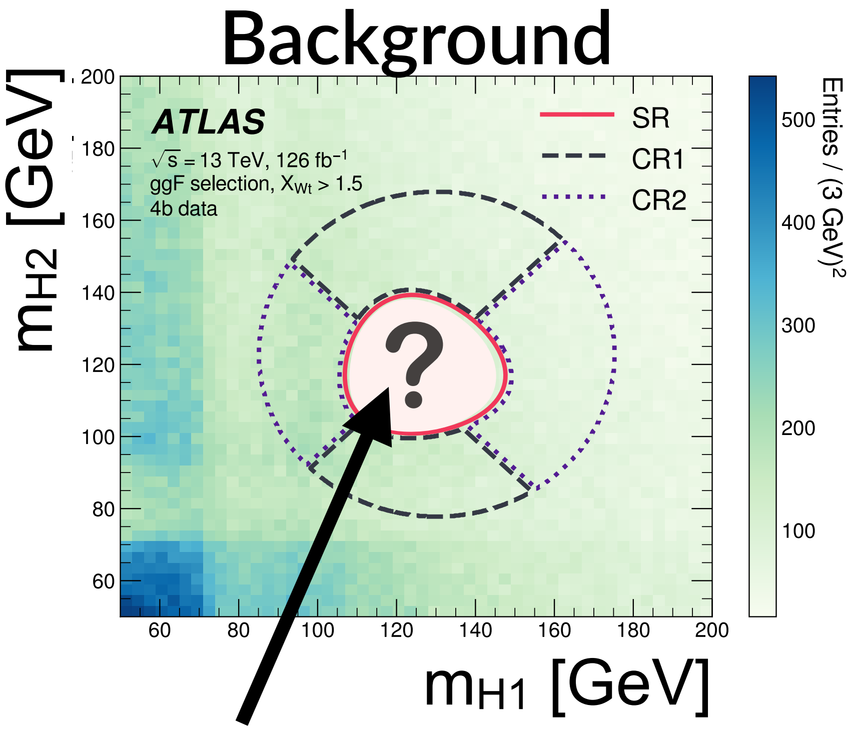


Trigger → b-jets → $\Delta\eta_{HH} < 1.5$ → SR

$$X_{HH} = \sqrt{\left(\frac{m_{H1} - 124\ \text{GeV}}{0.1 m_{H1}}\right)^2 + \left(\frac{m_{H2} - 117\ \text{GeV}}{0.1 m_{H2}}\right)^2}$$

SR: $X_{HH} < 1.6$





How to predict the background here?

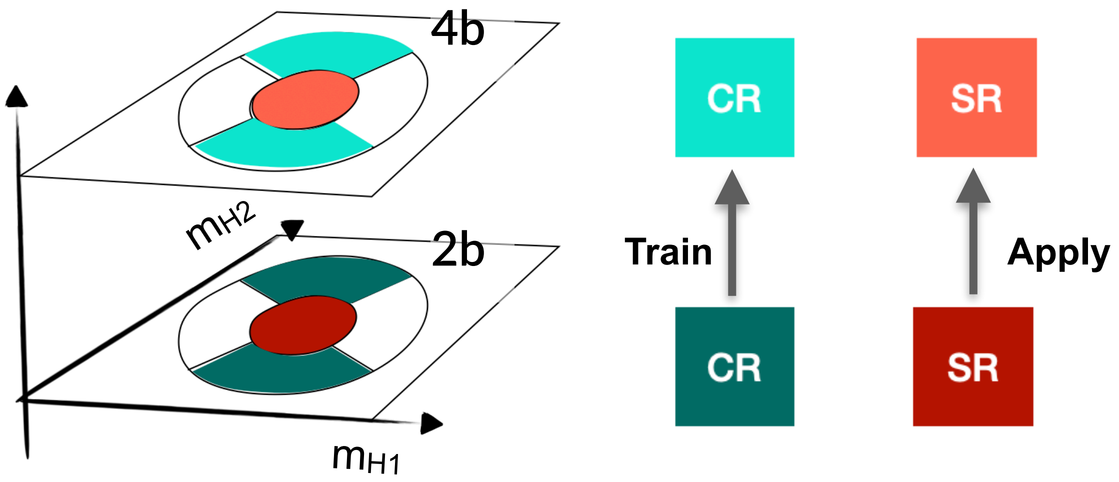
Want to fit a shape with categories… need a multidimensional description of the data.

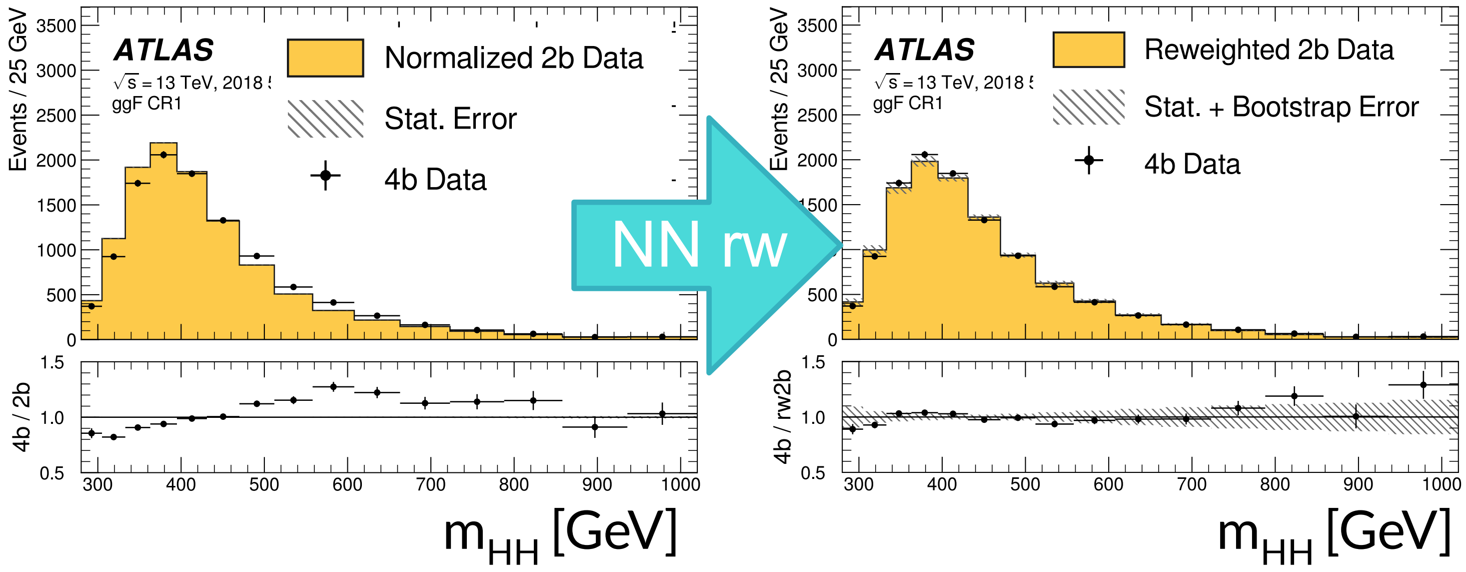
1911.00405
2202.07288; 2301.03212
Graphic from Lukas Heinrich

Reweighting for background estimation

Use classifiers to learn likelihood ratios, e.g, $p_{4b}(x)/p_{2b}(x)$

$$p_{4b} = \boxed{w(x)} \cdot p_{2b}(x),\ x \in \mathbb{R}^d$$



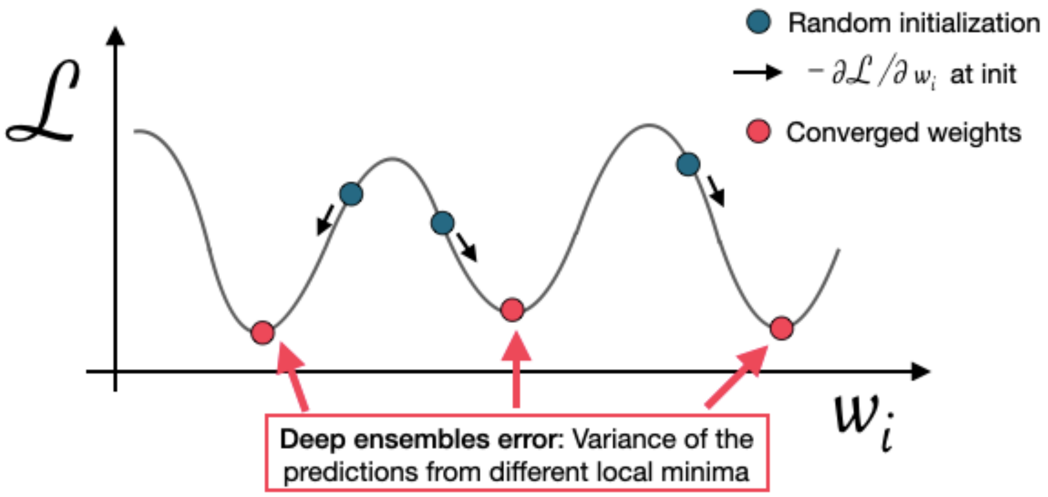


🥳 It works!

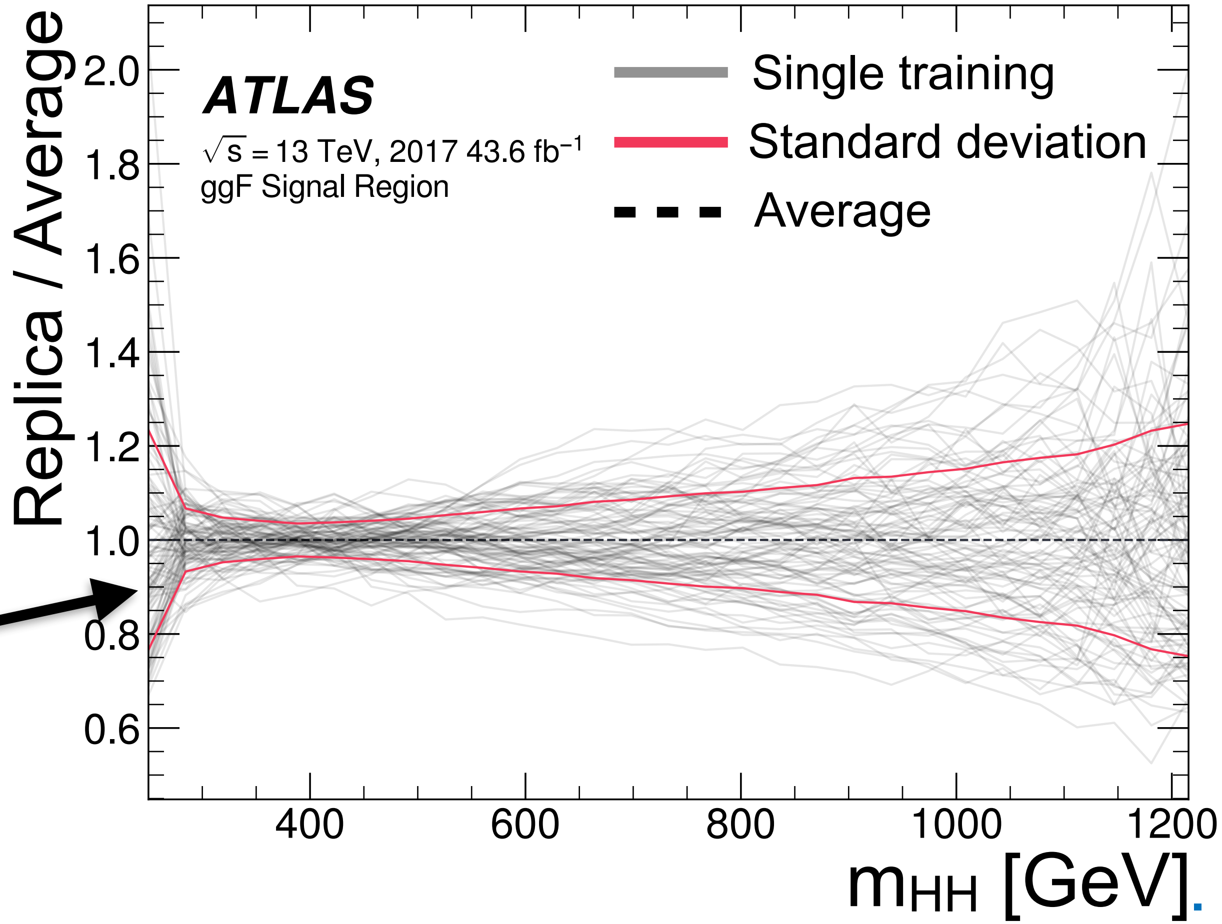
Note m_{HH} was not used in the set of reweighting features x.

Deep Ensembles: 1612.01474 & 1912.02757
2301.03212

Uncertainty: NN training

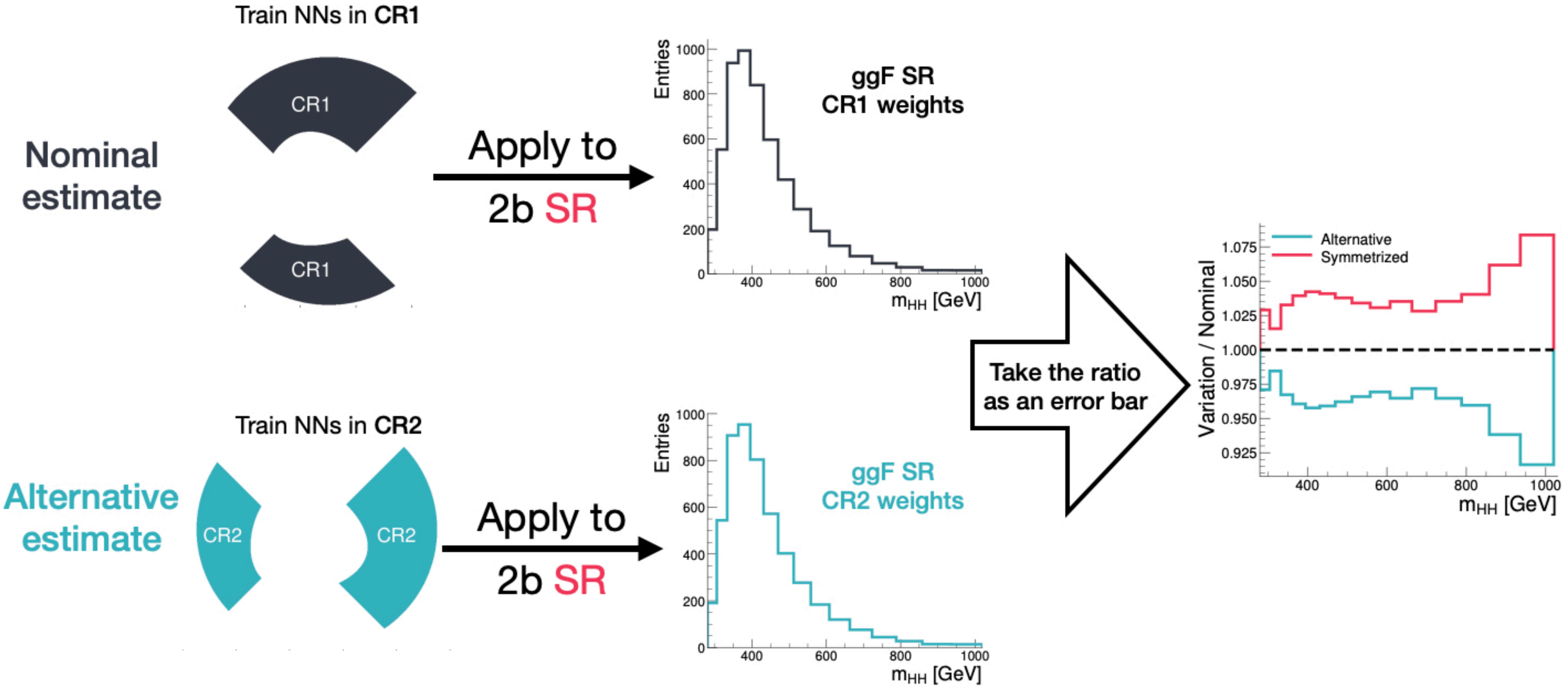


Retraining the NN 100x captures the uncertainty from multiple local minima



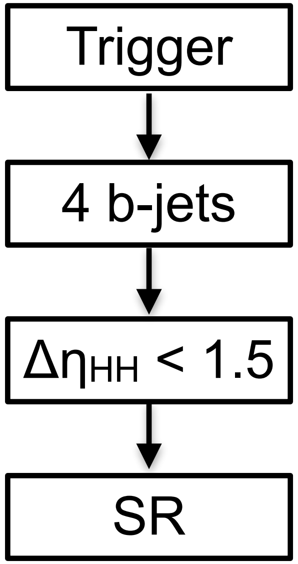
Use the variation of trainings as a nuisance parameter

Uncertainty: choice of control region



Background validation

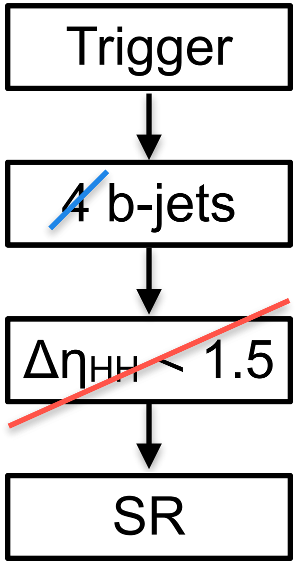
Q: Have we accounted for all uncertainties?



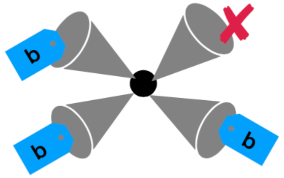
Background validation

Q: Have we accounted for all uncertainties?

→ Invert *every cut!!*

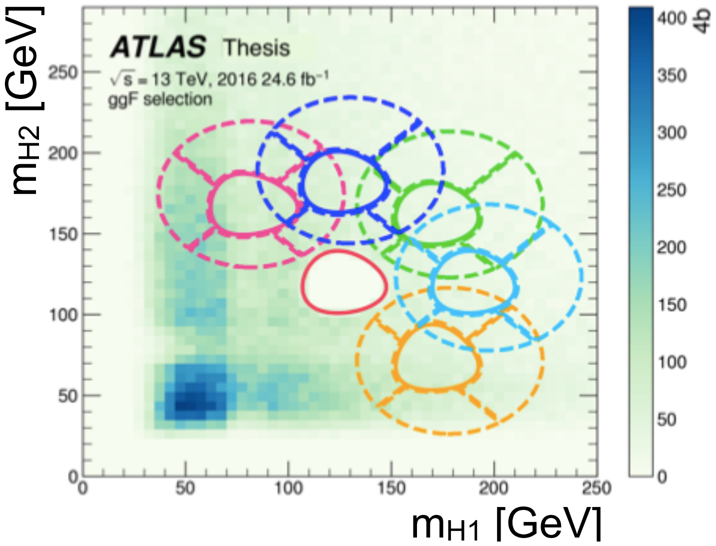


3 b-jets



$\Delta\eta_{HH} > 1.5$

Shift the center (5x)



Background validation

Q: Have we accounted for all uncertainties?

→ Invert *every cut!!*

Trigger

↓

~~4~~ b-jets

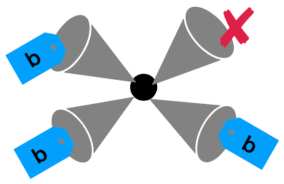
↓

~~$\Delta\eta_{HH} < 1.5$~~

↓

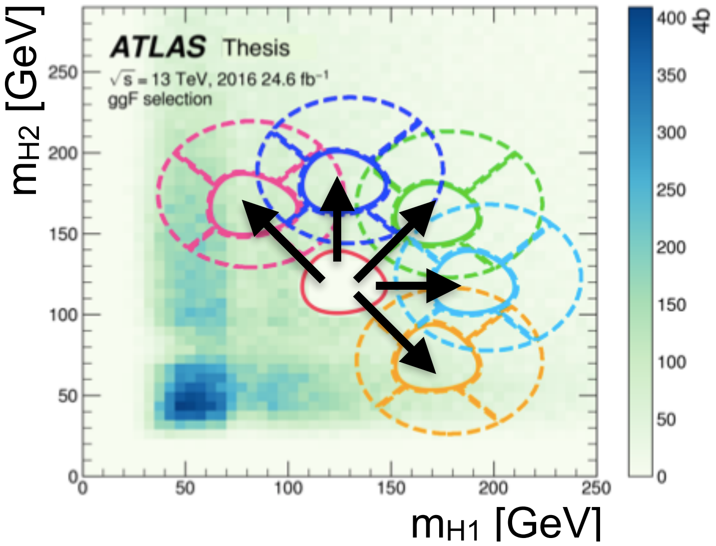
SR

3 b-jets



$\Delta\eta_{HH} > 1.5$

Shift the center (5x)



Background validation

Q: Have we accounted for all uncertainties?

→ Invert *every cut!!*

Trigger

↓

~~4~~ b-jets

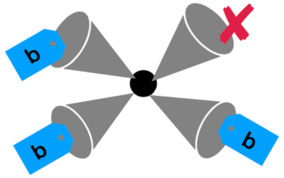
↓

~~$\Delta\eta_{HH} < 1.5$~~

↓

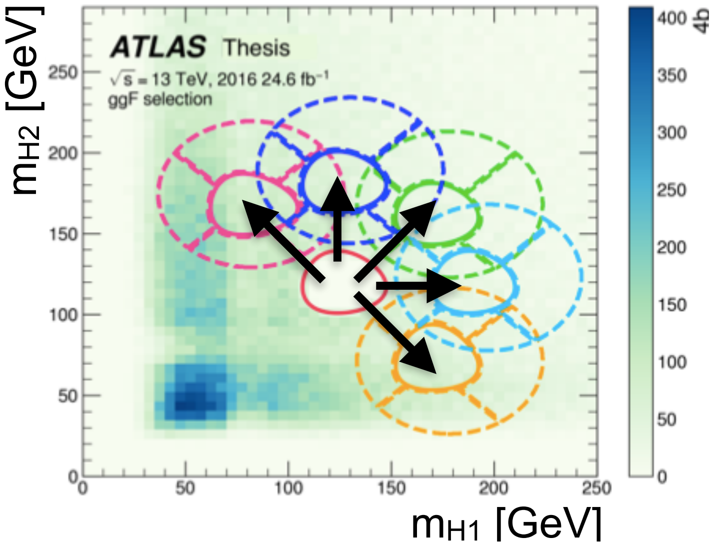
SR

3 b-jets



$\Delta\eta_{HH} > 1.5$

Shift the center (5x)



✅ Modeled by existing uncertainties

Background validation

Q: Have we accounted for all uncertainties?

→ Invert *every cut!!*

Trigger

↓

~~4~~ b-jets

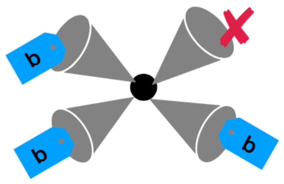
↓

~~$\Delta\eta_{HH} < 1.5$~~

↓

SR

3 b-jets

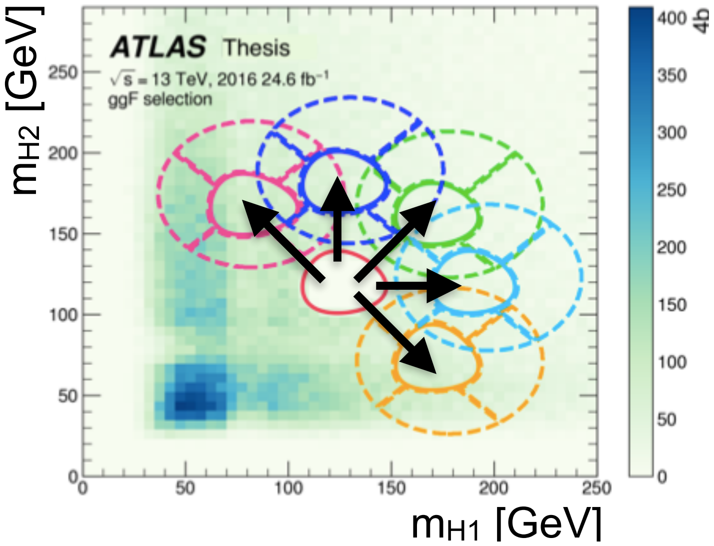


❌ Some mismodeling

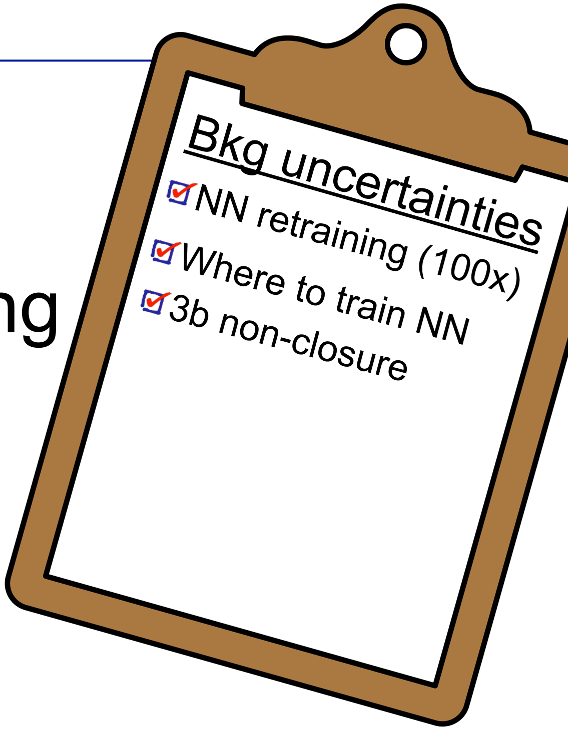
→ Add uncertainty

$\Delta\eta_{HH} > 1.5$

Shift the center (5x)

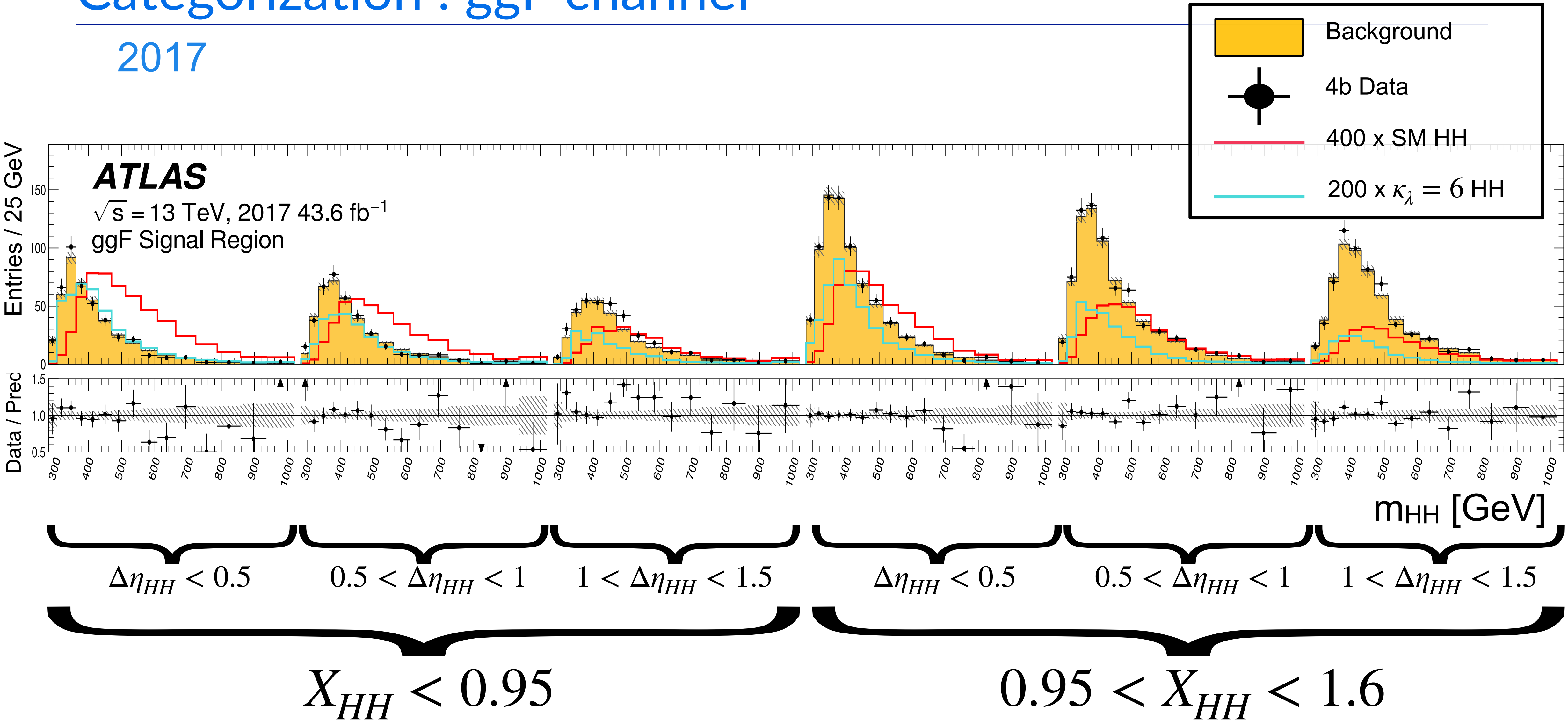


✅ Modeled by existing uncertainties



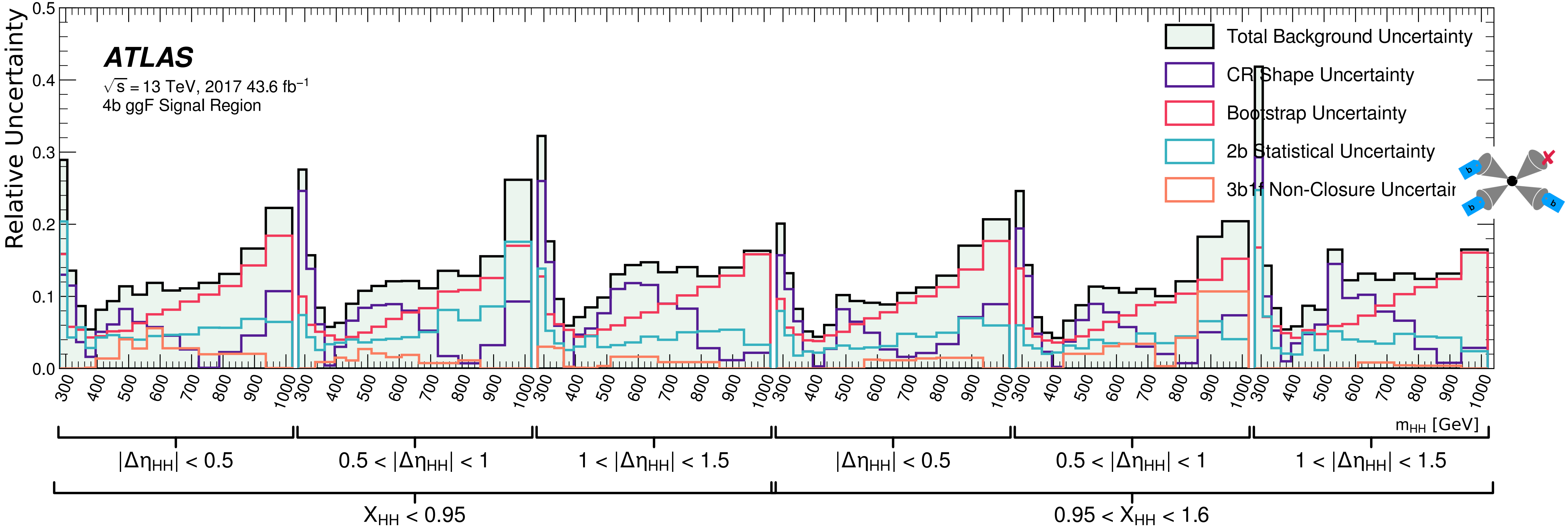
Categorization : ggF channel

2017

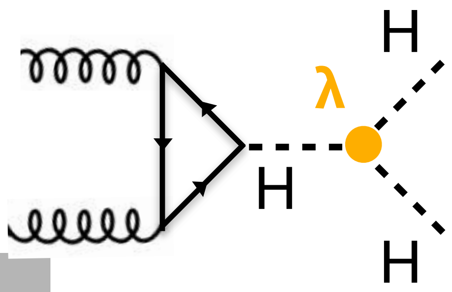


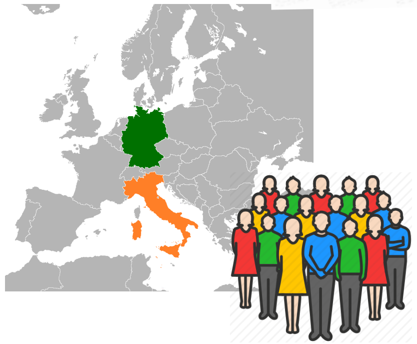
Systematics

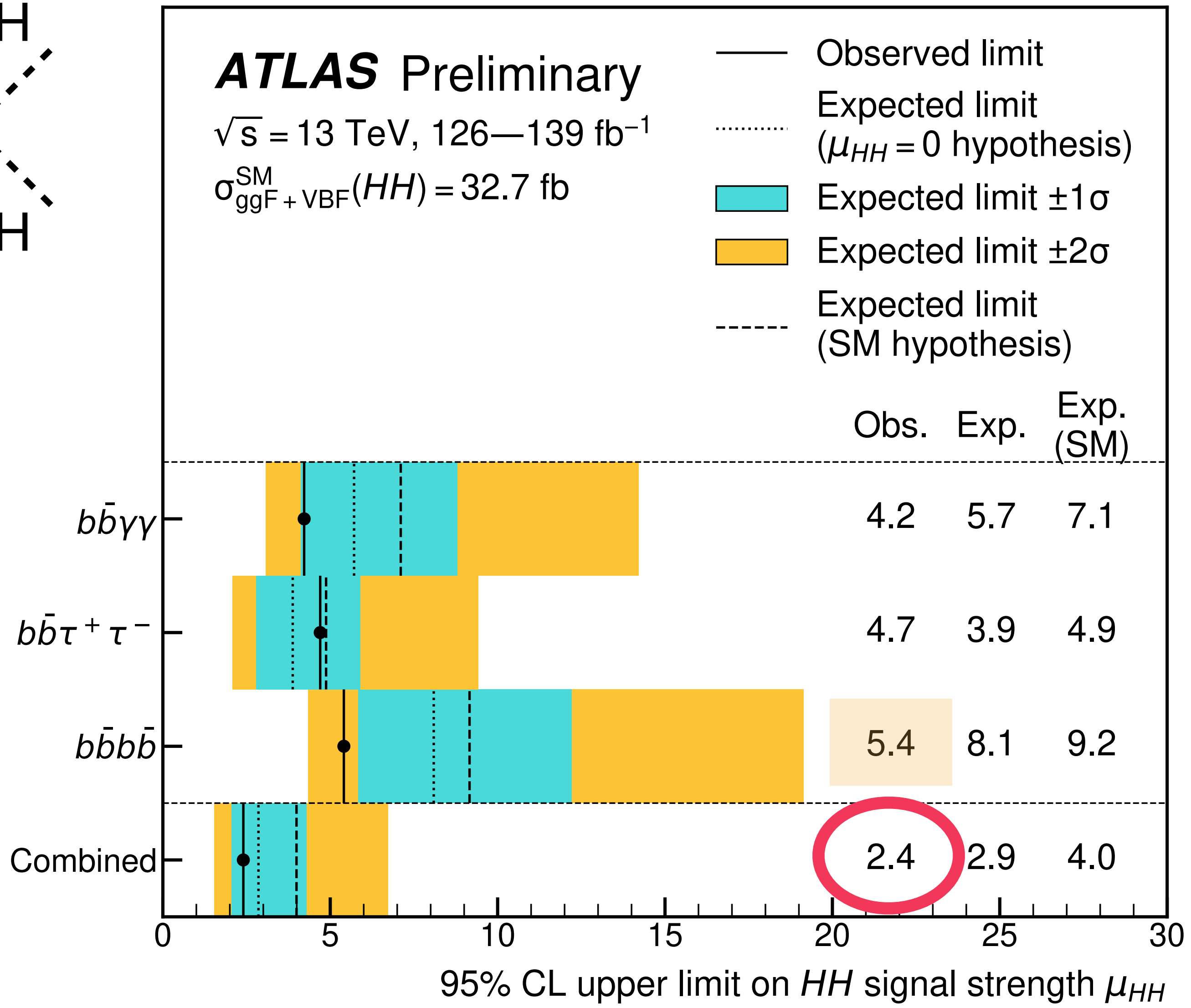
2017



Impact in the combination

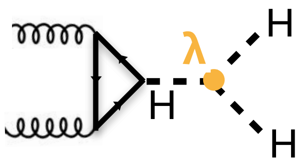






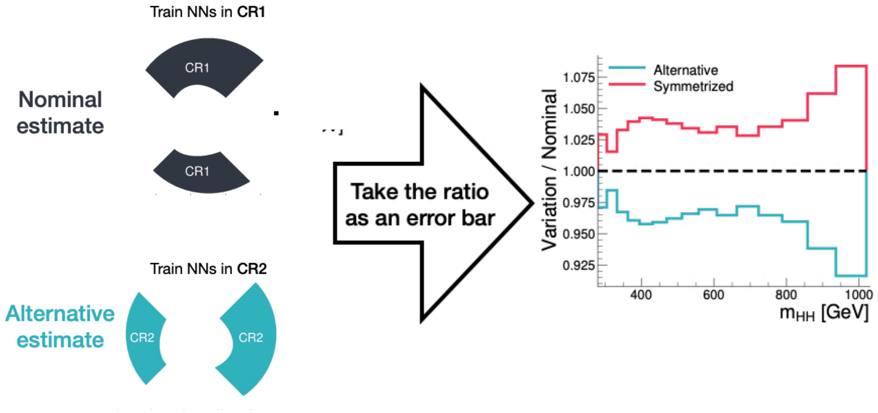
Impact in one analysis

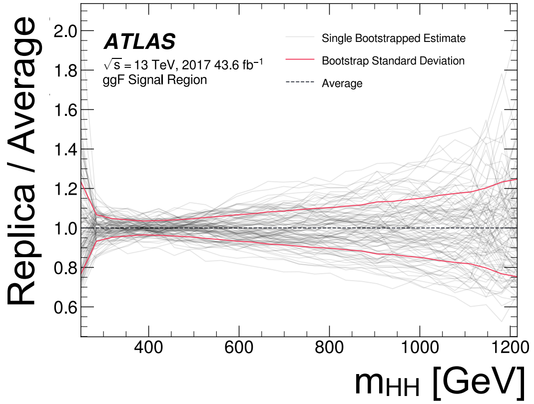
Impact on upper limit on σ(ggF HH)



Source of uncertainty		$\|\Delta\mu\|/\mu$
Theoretical	Uncertainty on signal rate	9.0%
	All other theory uncertainties	1.4%
Background modeling	Choice of CR	7.5%
	NN retraining (100x)	7.1%
	3b non-closure	2.0%

All other experimental systematics [FTAG, Jets, …] < %-level impact

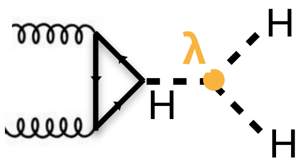




2301.03212

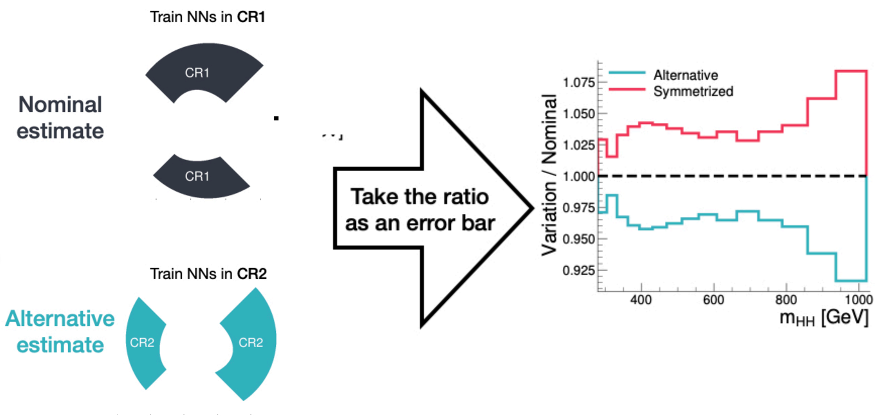
Impact in one analysis

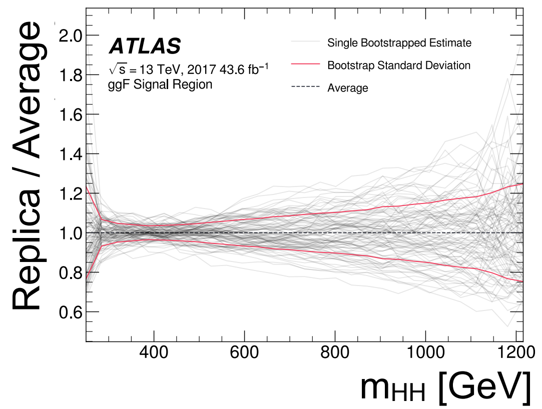
Impact on upper limit on σ(ggF HH)



Source of uncertainty		$\|\Delta\mu\|/\mu$
Theoretical	Uncertainty on signal rate	9.0%
	All other theory uncertainties	1.4%
Background modeling	Choice of CR	7.5%
	NN retraining (100x)	7.1%
	3b non-closure	2.0%

All other experimental systematics [FTAG, Jets, …] < %-level impact





Better background modelling **key** for improving this analysis!

How will ML continue to improve our analysis?

Better b-taggers

End-to-end optimization strategies

Flows for background modelling

NH thesis with S. Gasiorowski, R. Teixeira de Lima, M. Kagan

Flows for high dimensional interpolation

Hierarchical model: $p(x, m_{H1}, m_{H2}) = p(x| m_{H1}, m_{H2})\; p(m_{H1}, m_{H2})$

Normalizing flow — $p(x| m_{H1}, m_{H2})$

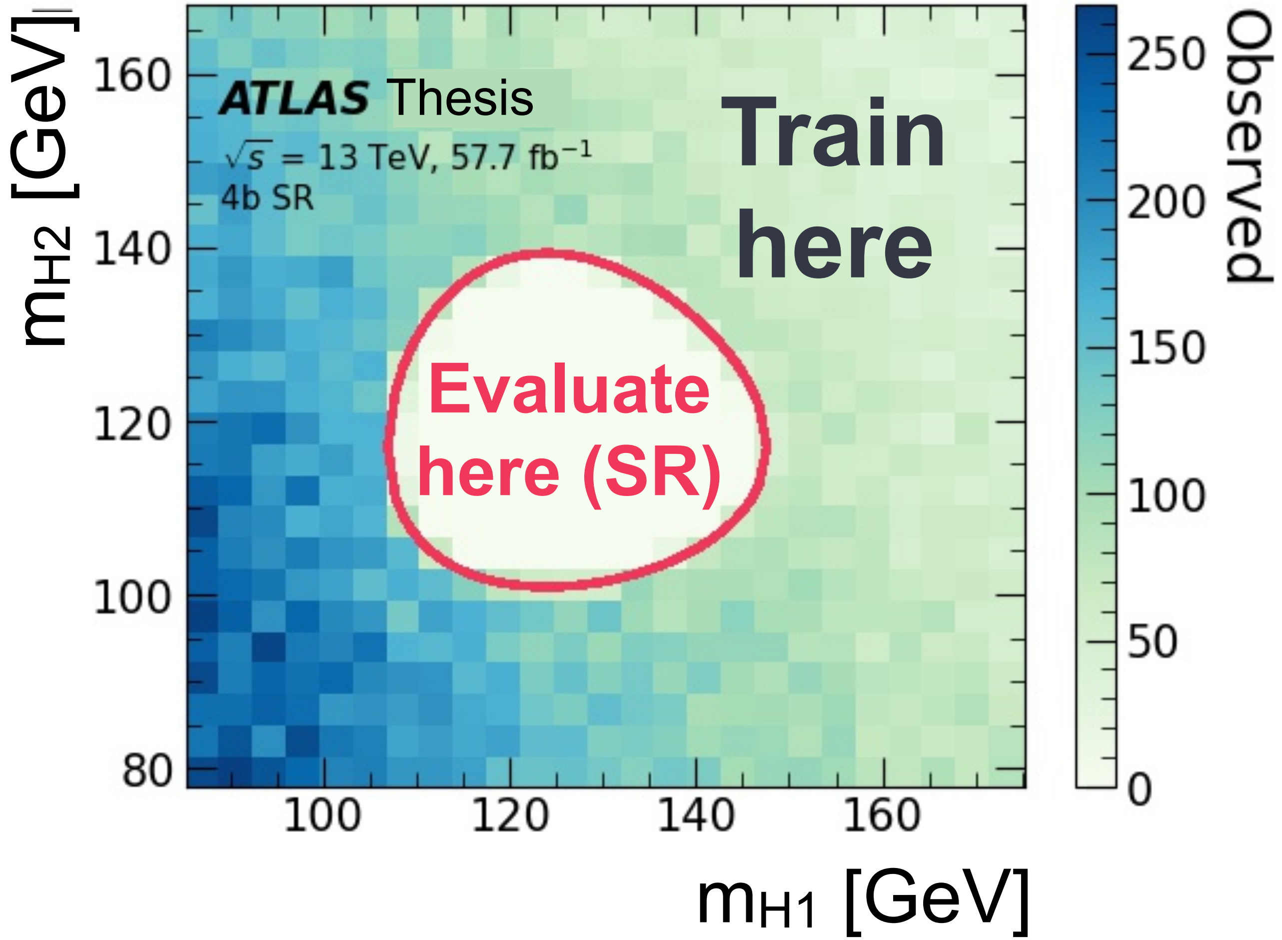
Gaussian process — $p(m_{H1}, m_{H2})$

x: Event kinematics

$p_{T,H1}$, $p_{T,H2}$, η_{H1}, η_{H2}, $\Delta\phi_{HH}$, X_{Wt} [top veto]

Use the smoothly varying (m_{H1}, m_{H2}) to predict SR kinematics.

Conceptually identical to CATHODE



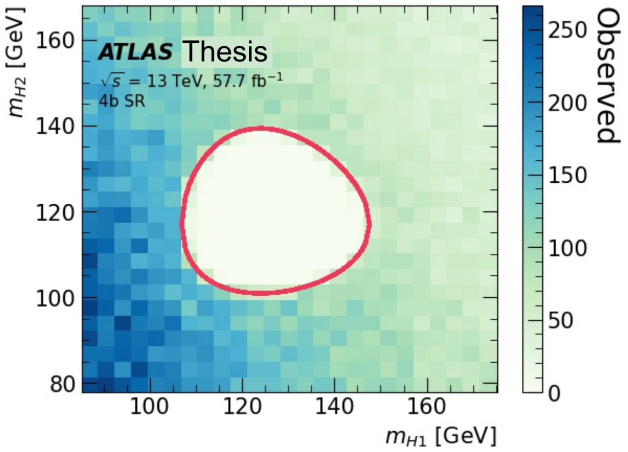
$$p(x, m_{H1}, m_{H2}) = p(x \mid m_{H1}, m_{H2})\ p(m_{H1}, m_{H2})$$

1) GP fits

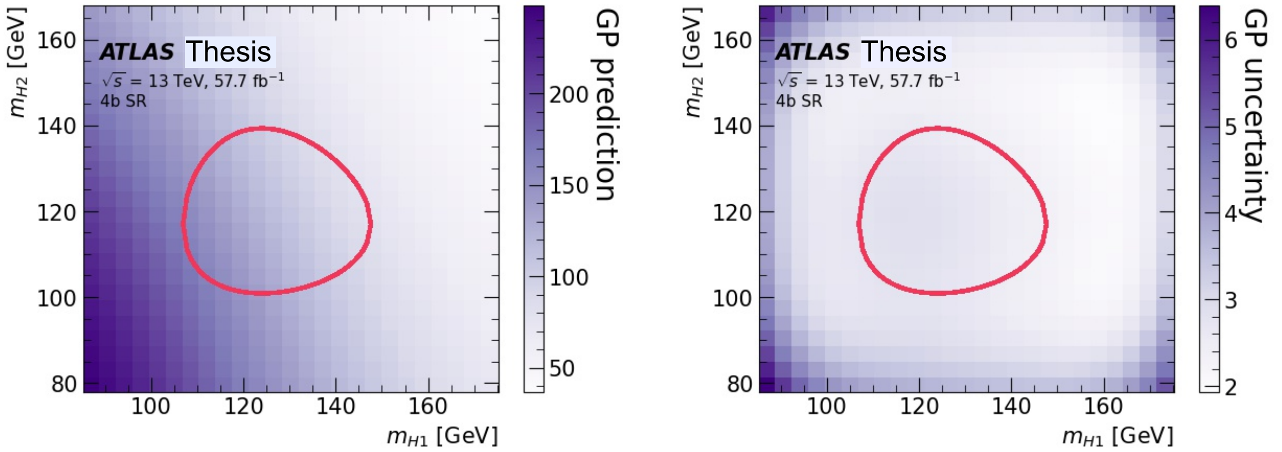
Fit a GP to the 2d (m_{H1}, m_{H2}) histogram

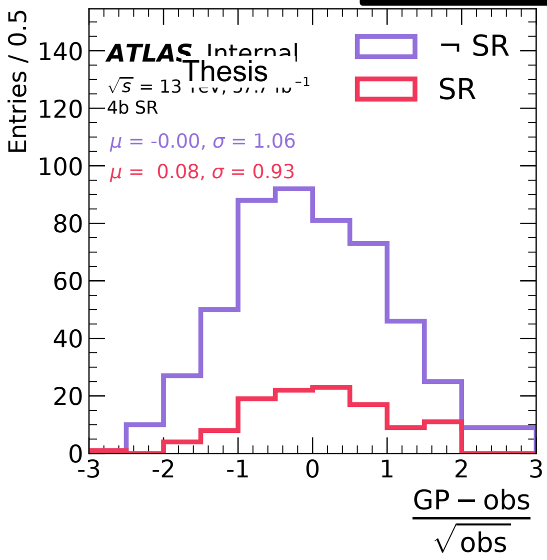
Radial basis function kernel, 2d length scale

Input



Output



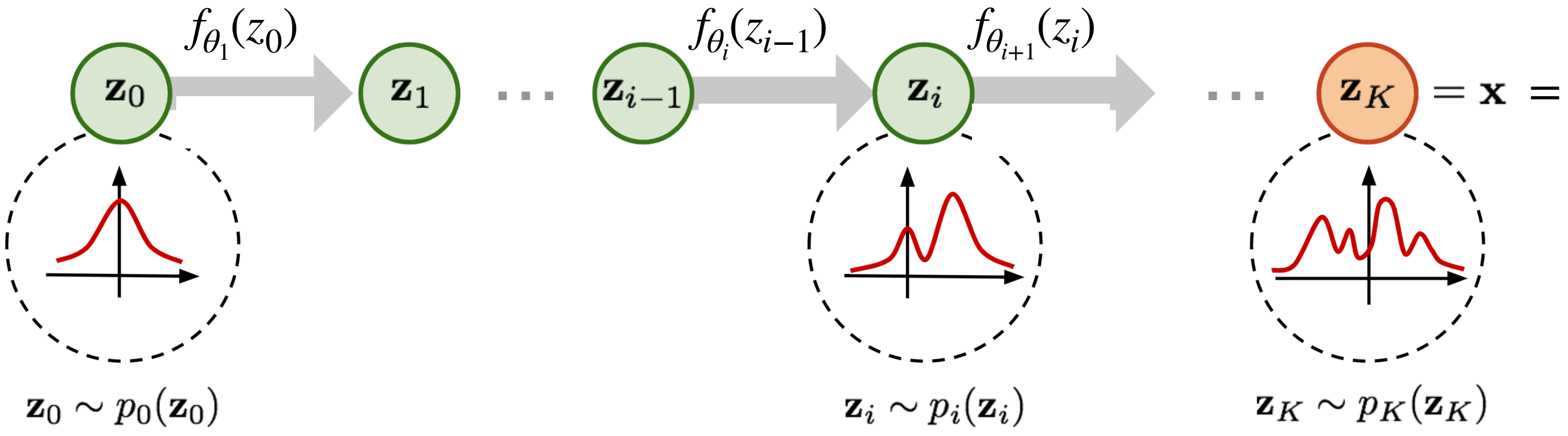


Set the normalization from massplane fit:

$$\text{SR yield} = \text{obs}(\text{not SR})\ \frac{\text{pred(SR)}}{\text{pred}(\neg\ \text{SR})}$$

Normalizing Flow

Gaussian,
e.g, $z \in \mathbb{R}^6$



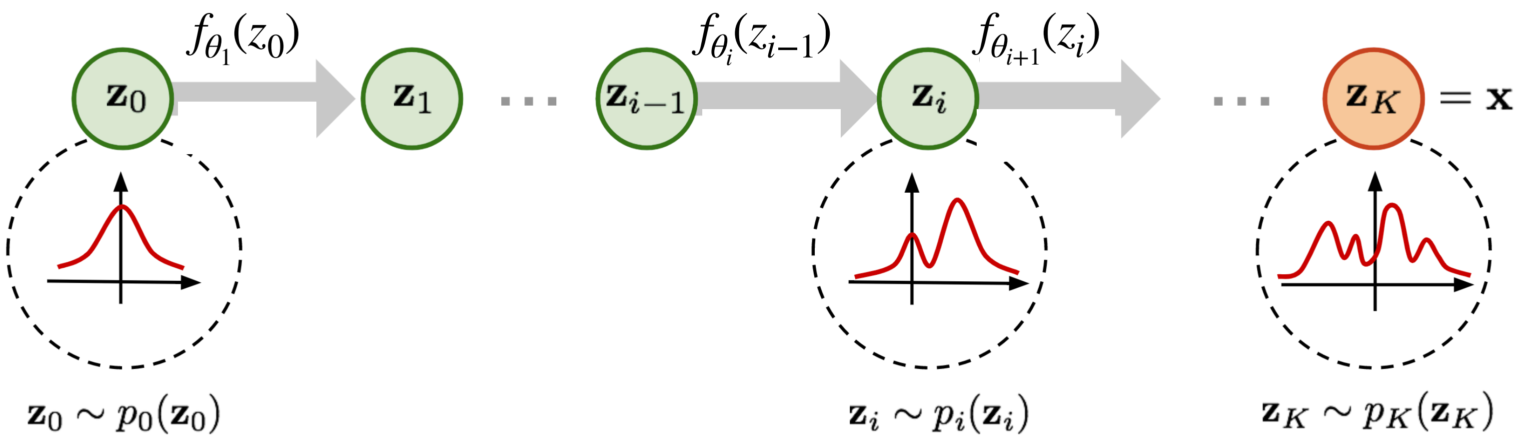
$$\begin{pmatrix} \log p_{T,H1} \\ \log p_{T,H2} \\ \eta_{H1} \\ \eta_{H2} \\ \log(\pi - \Delta\phi_{HH}) \\ X_{Wt} \end{pmatrix}$$

Predict Higgs
Candidates

Flow gradually from one distribution into another.

Normalizing Flow

Gaussian,
e.g, $z \in \mathbb{R}^6$



$$= \mathbf{x} = \begin{pmatrix} \log p_{T,H1} \\ \log p_{T,H2} \\ \eta_{H1} \\ \eta_{H2} \\ \log(\pi - \Delta\phi_{HH}) \\ X_{Wt} \end{pmatrix}$$

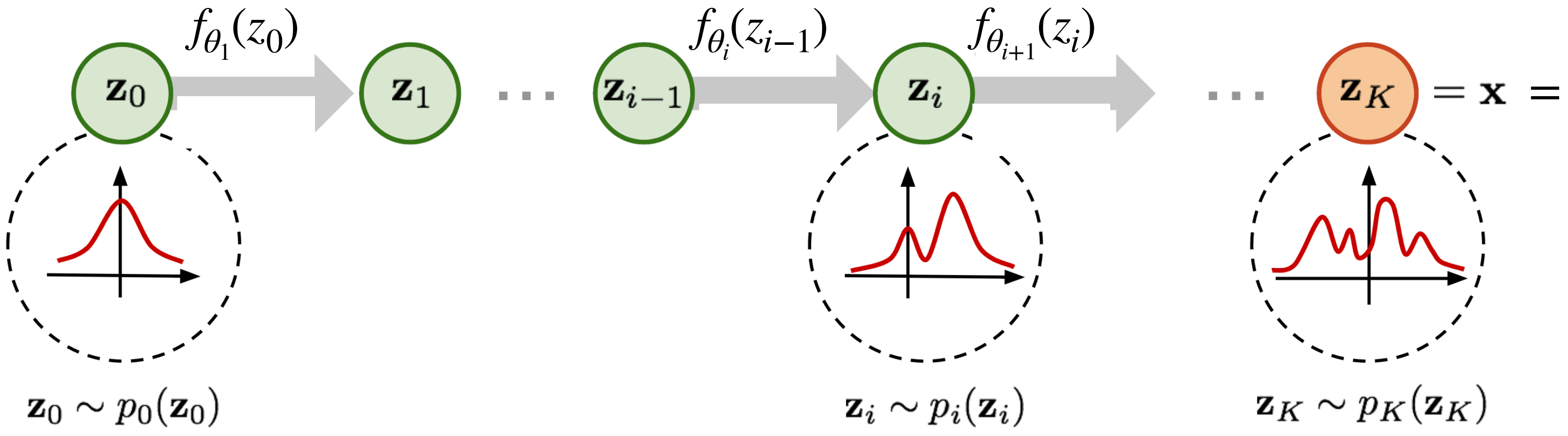
Predict Higgs Candidates

Flow gradually from one distribution into another.

- $f_\theta = f_{\theta L} \circ \cdots \circ f_{\theta 2} \circ f_{\theta 1}$
- Invertible f_{θ_i} allow us to get the density of training samples

Normalizing Flow

Gaussian,
e.g, $z \in \mathbb{R}^6$



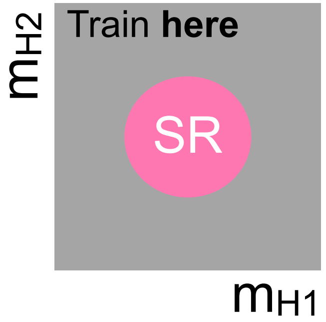
$$= \mathbf{x} = \begin{pmatrix} \log p_{T,H1} \\ \log p_{T,H2} \\ \eta_{H1} \\ \eta_{H2} \\ \log(\pi - \Delta\phi_{HH}) \\ X_{Wt} \end{pmatrix}$$

Predict Higgs Candidates

Flow gradually from one distribution into another.

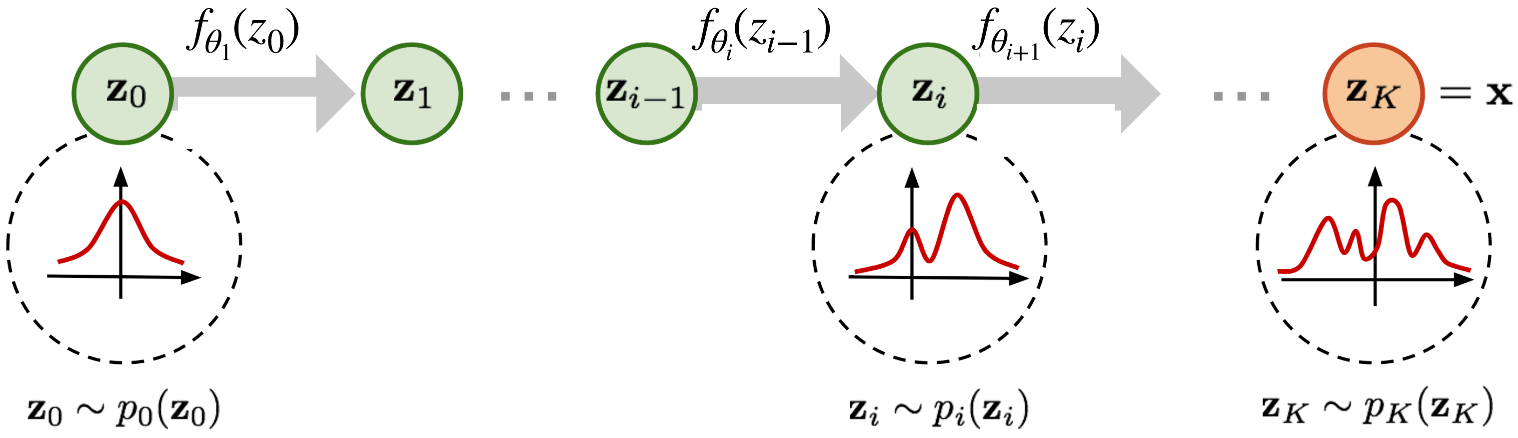
- $f_\theta = f_{\theta L} \circ \cdots \circ f_{\theta 2} \circ f_{\theta 1}$
- Invertible f_{θ_i} allow us to get the density of training samples
- Conditional generative model $p_\theta(x \mid y),\ y = (m_{H1}, m_{H2})$

Key for interpolation!



Normalizing Flow

Gaussian,
e.g, $z \in \mathbb{R}^6$



$$= \mathbf{x} = \begin{pmatrix} \log p_{T,H1} \\ \log p_{T,H2} \\ \eta_{H1} \\ \eta_{H2} \\ \log(\pi - \Delta\phi_{HH}) \\ X_{Wt} \end{pmatrix}$$

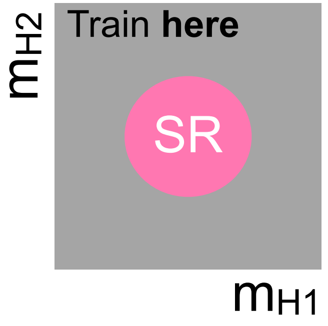
Predict Higgs Candidates

Flow gradually from one distribution into another.

- $f_\theta = f_{\theta L} \circ \cdots \circ f_{\theta 2} \circ f_{\theta 1}$
- Invertible f_{θ_i} allow us to get the density of training samples
- Conditional generative model $p_\theta(x \mid y),\ y = (m_{H1}, m_{H2})$

$$\mathcal{L}oss = -\log p_\theta(x \mid y) = -\log p_z(f_\theta^{-1}(x \mid y)) - \sum_{i=1}^{K} \left| \frac{\partial f_{\theta_i}^{-1}}{\partial z_i^T} \right|$$

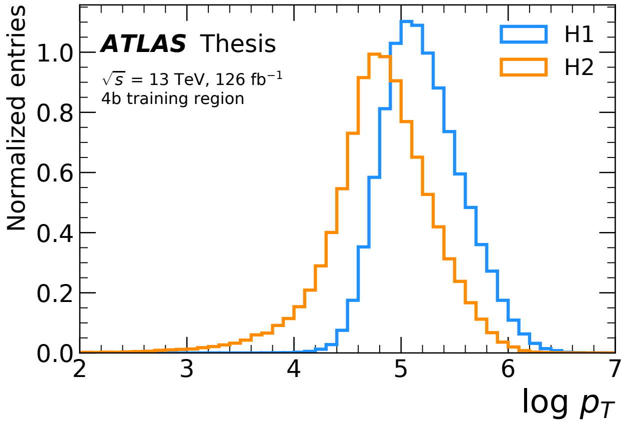
Key for interpolation!

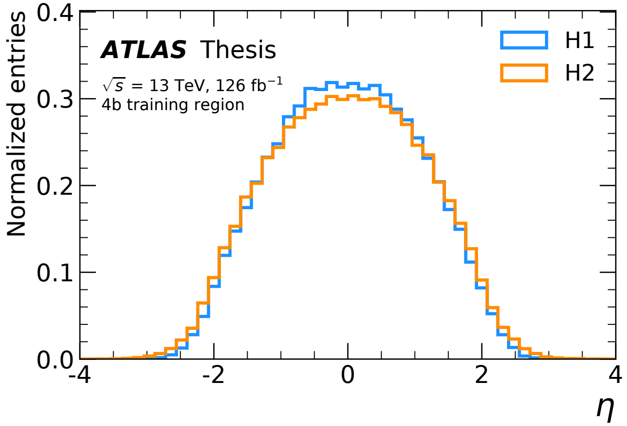


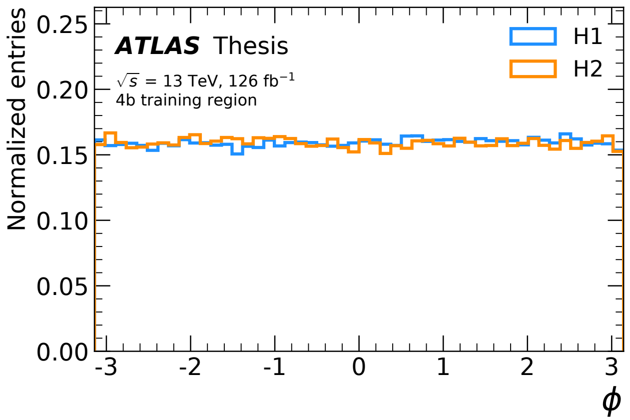
Input processing: HH -> 4b background modeling

Conditioning on m_{h1}, m_{h2} → ~~6~~ 5 variables left to model

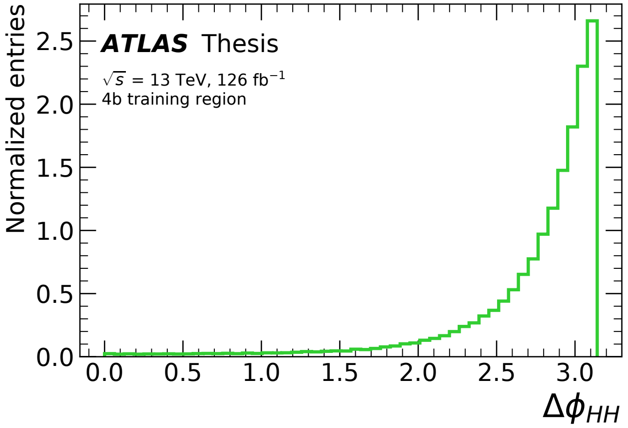
Concerned about modeling b/c no way to know that -π → π.

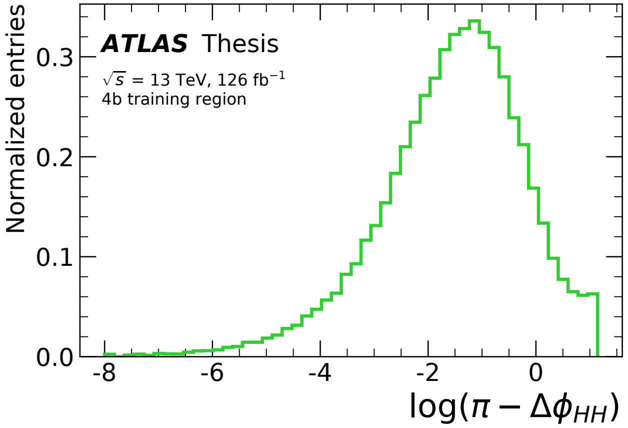




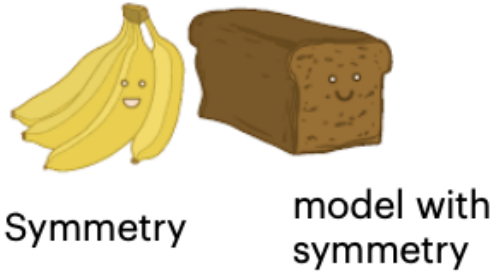


Constant $\Delta\Phi_{HH}$ - will give the same m_{HH}





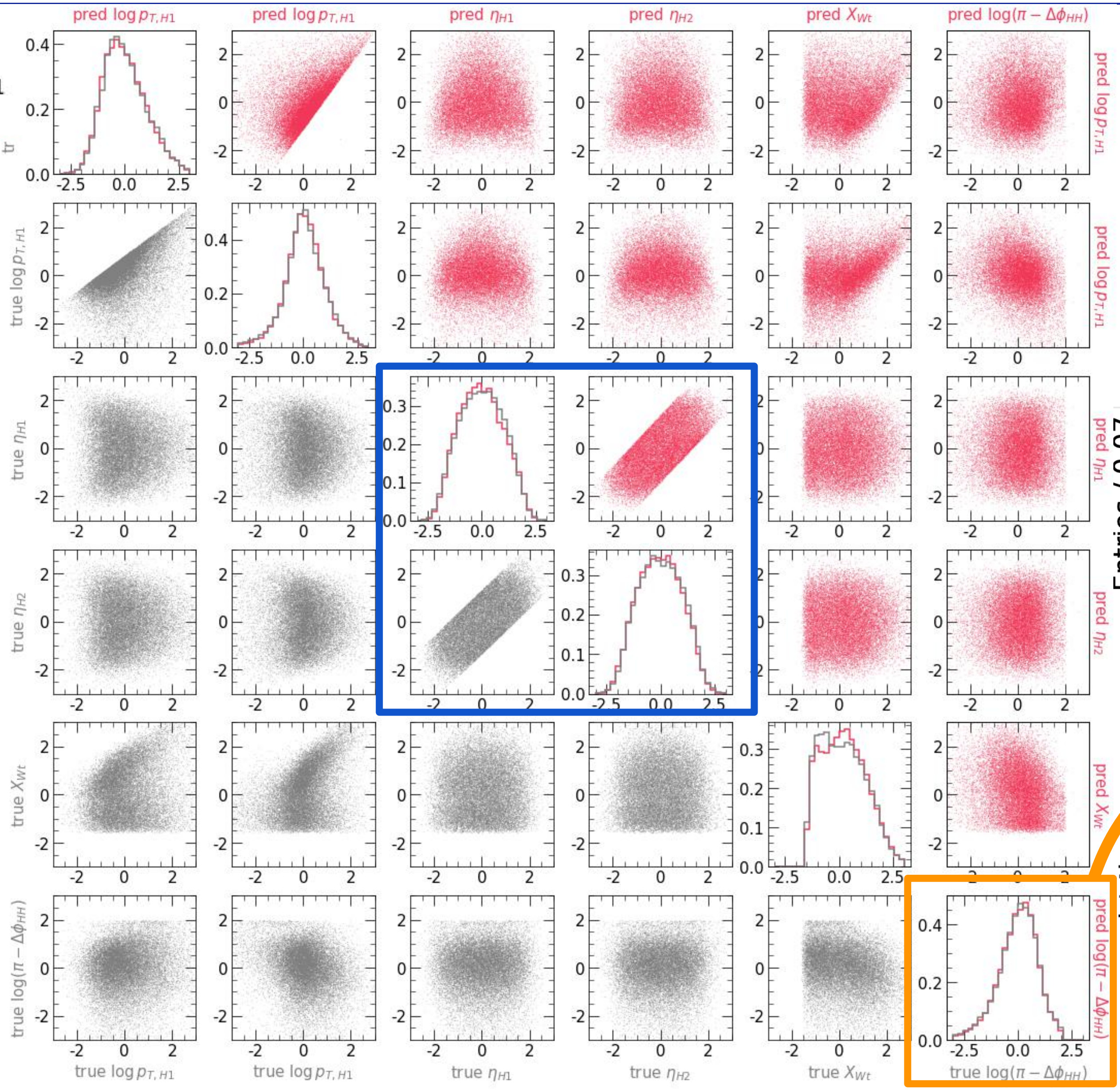
Azimuthal symmetry baked into the model



Input processing : HH -> 4b background modeling

ATLAS Thesis

$\sqrt{s}$ = 13 TeV, 126 fb^{-1}
4b SR

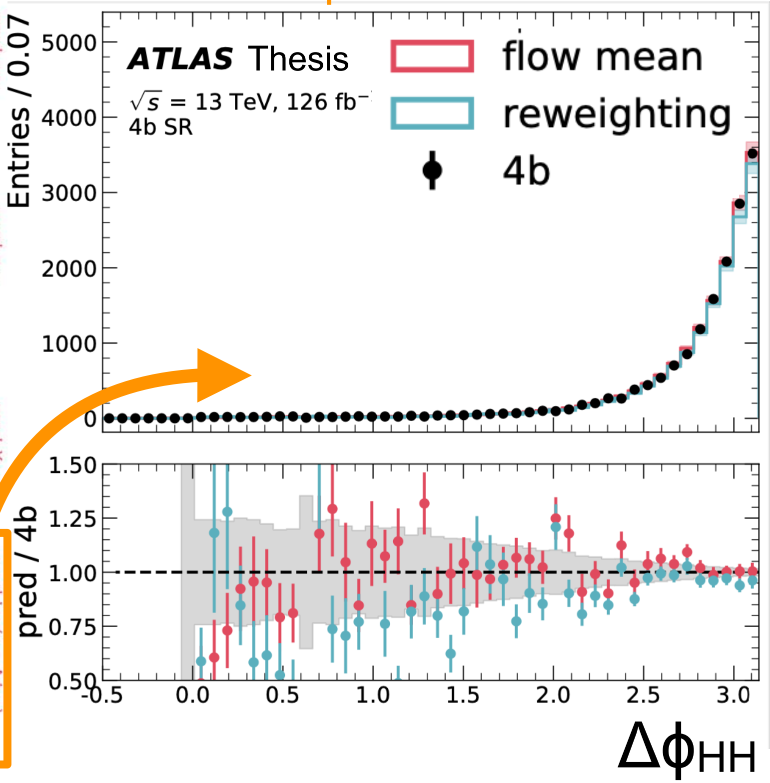


flow prediction

Learns about preprocessing cut ($\Delta\eta_{HH}$ < 1.5)

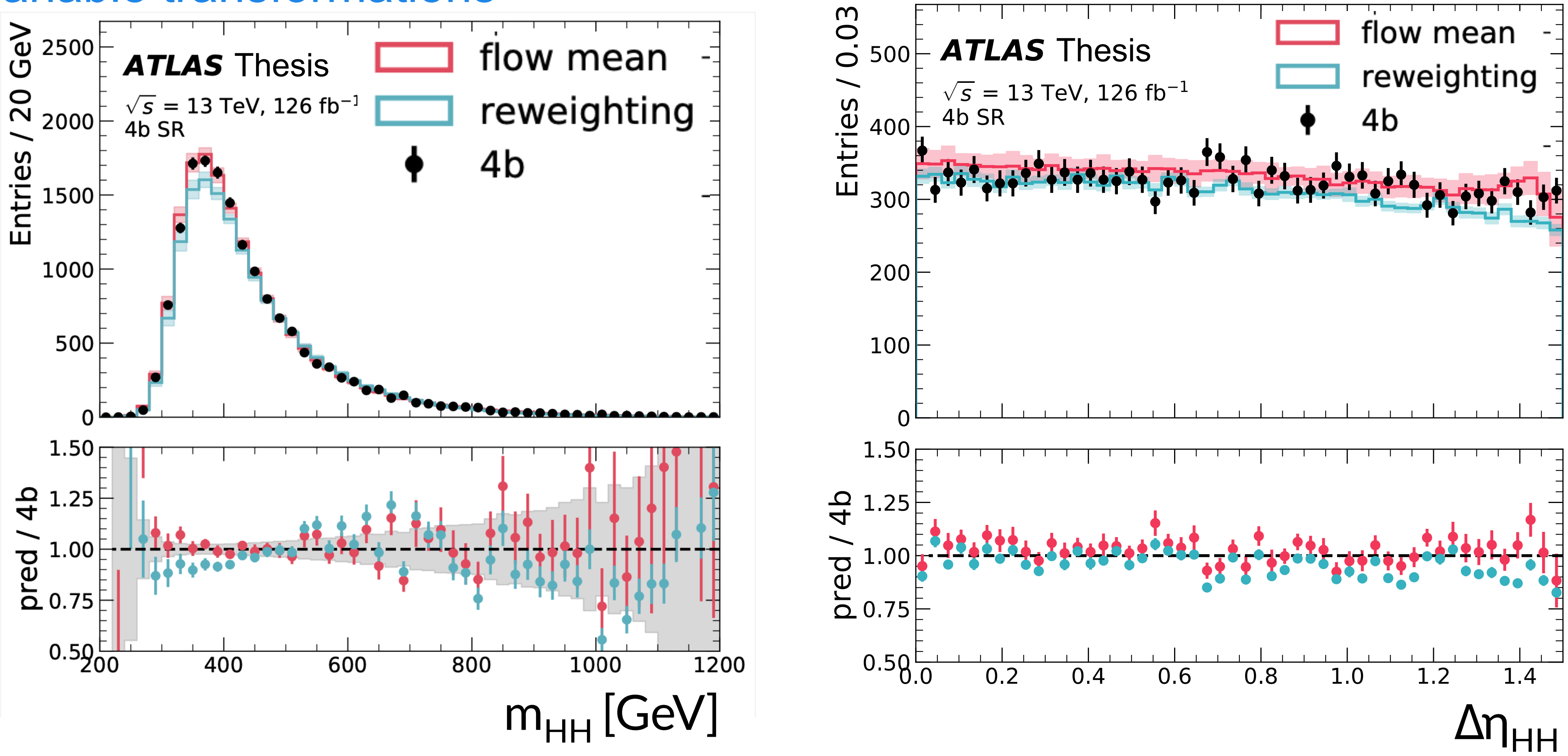
4b data

Good transformations help for complex distributions.



Input processing : HH → 4b background modeling

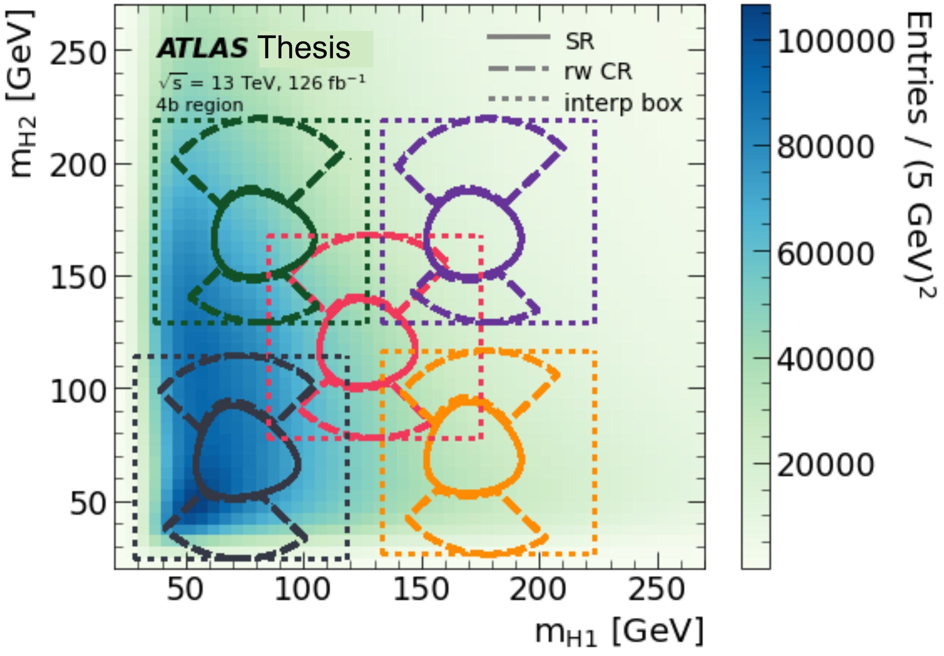
Variable transformations

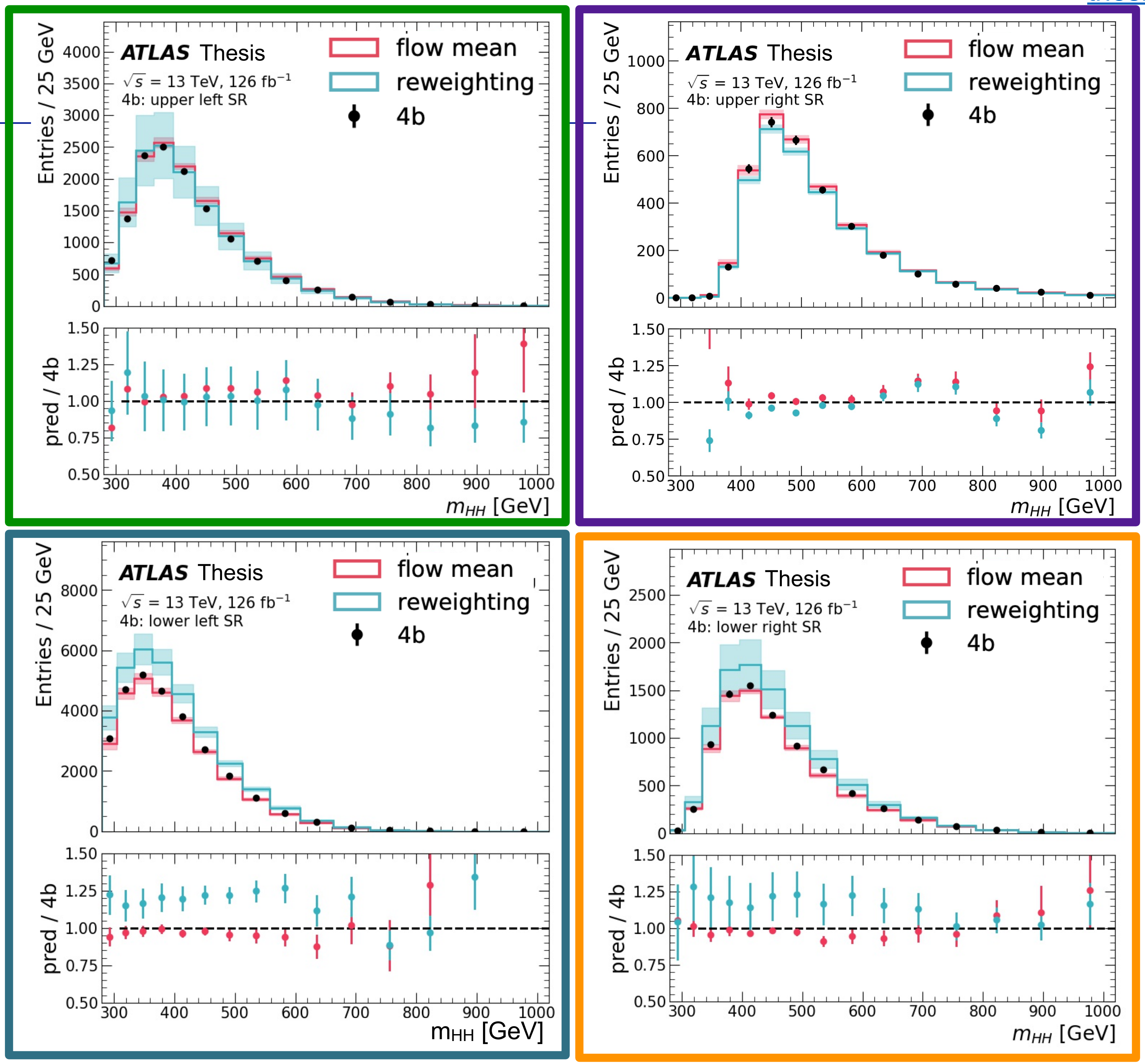


Promising... better closure.

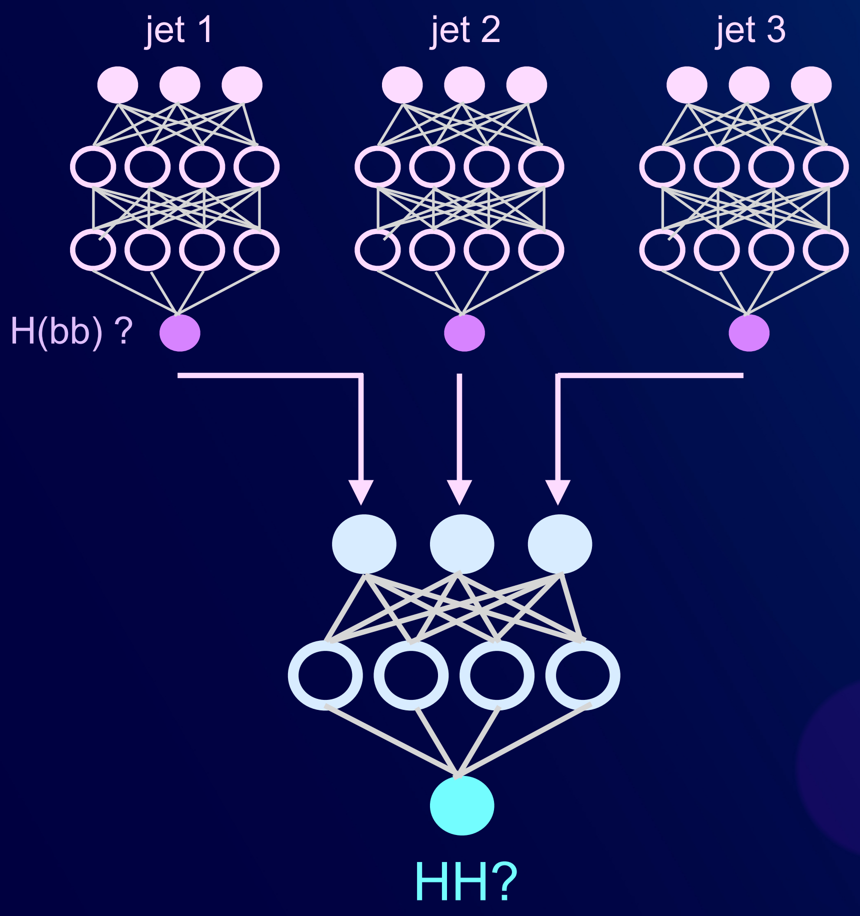
[thesis](#)

Shifted regions

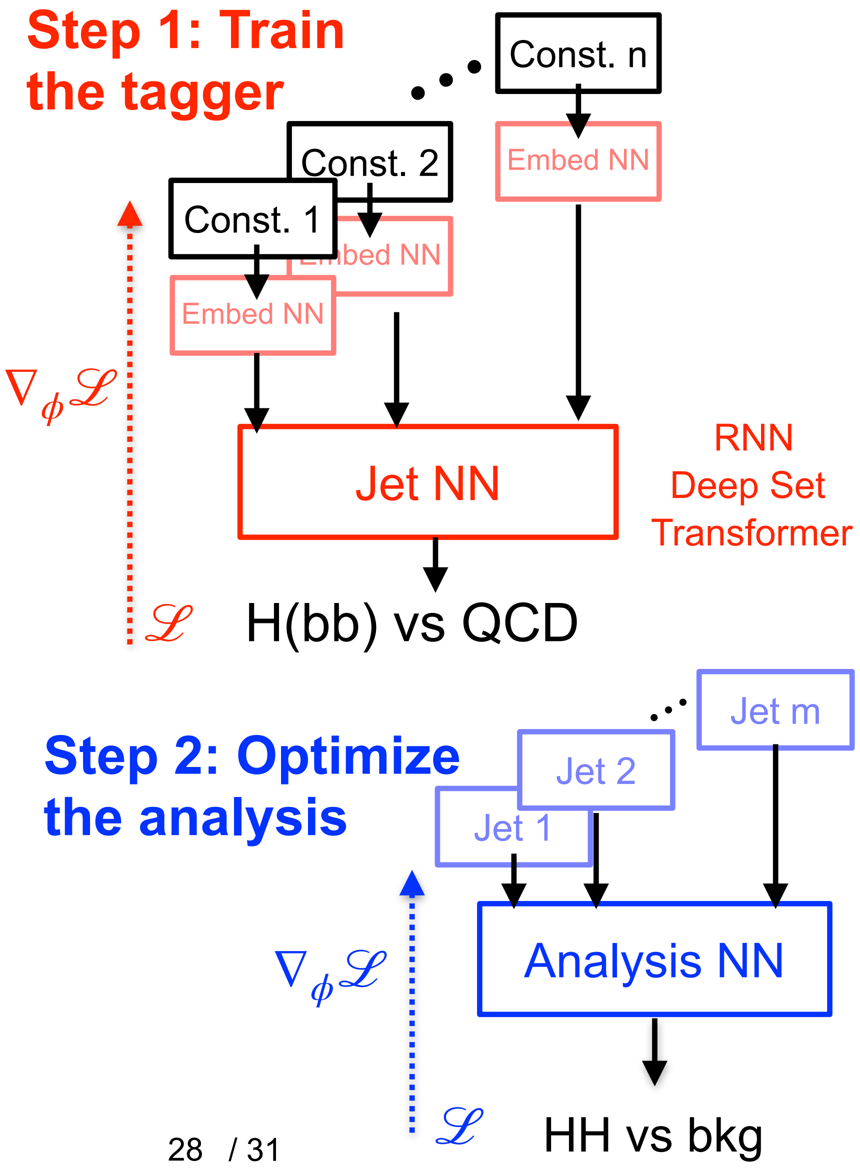




Can we use gradient optimization more?



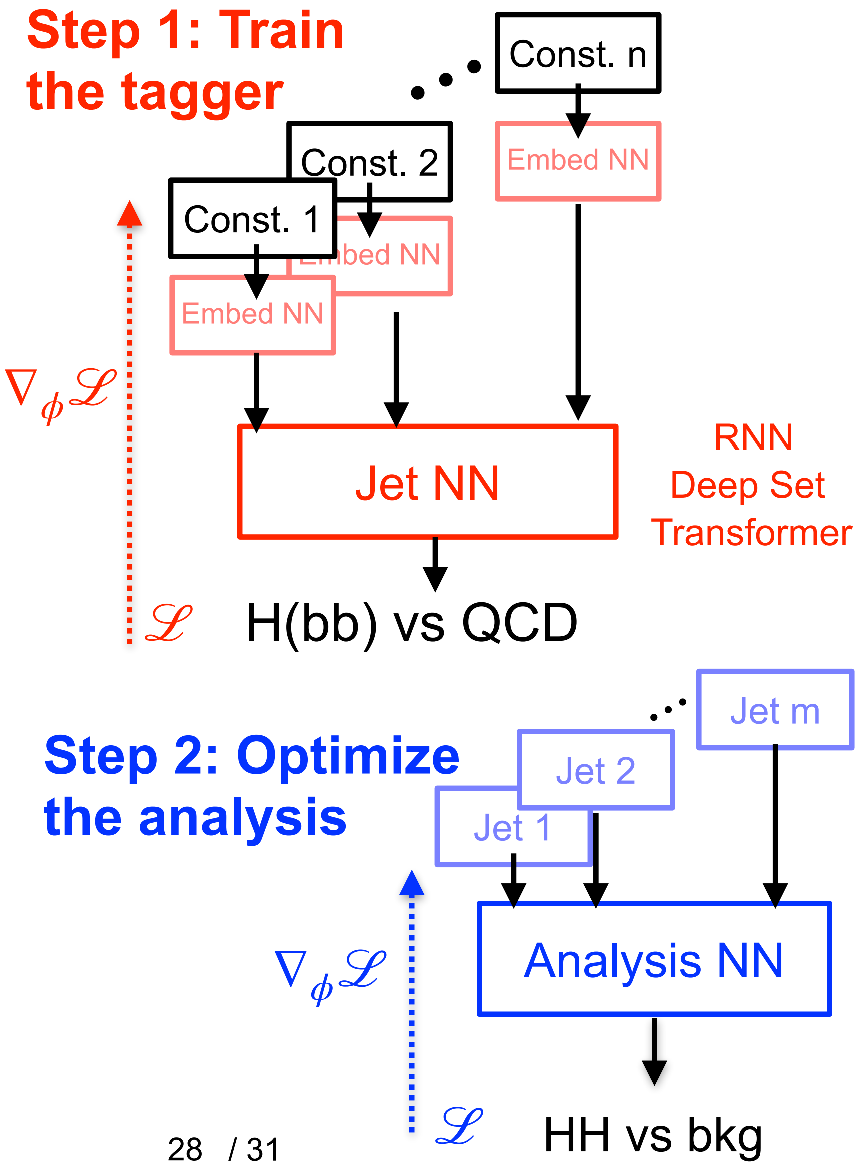
Traditional Analysis

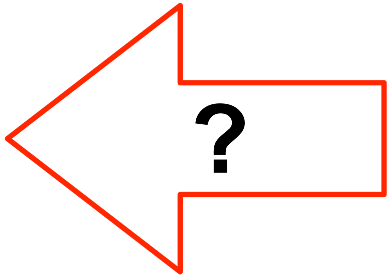



Lukas Heinrich


Matthias Vigl

Traditional Analysis





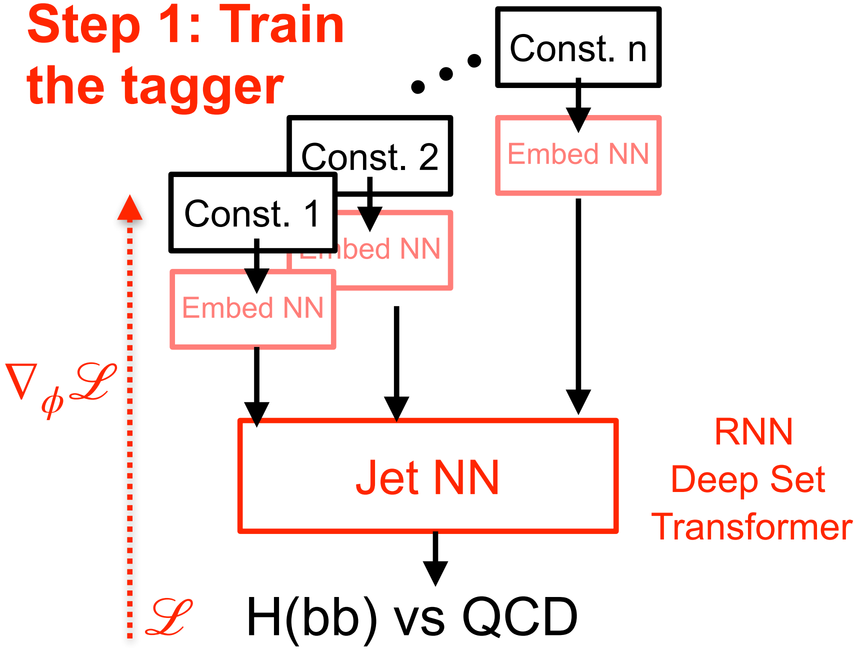
How do you know we this particle classifier is optimal for **every analysis**?

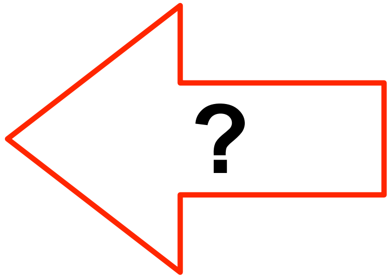
Lukas Heinrich

Matthias Vigl

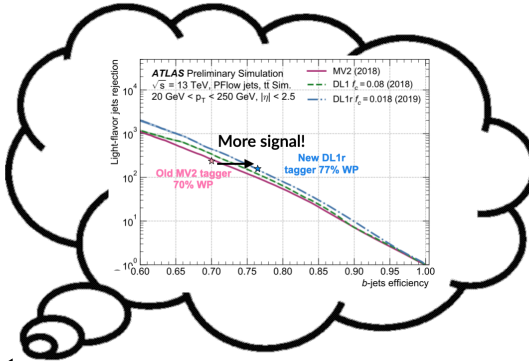
Traditional Analysis

Step 1: Train the tagger

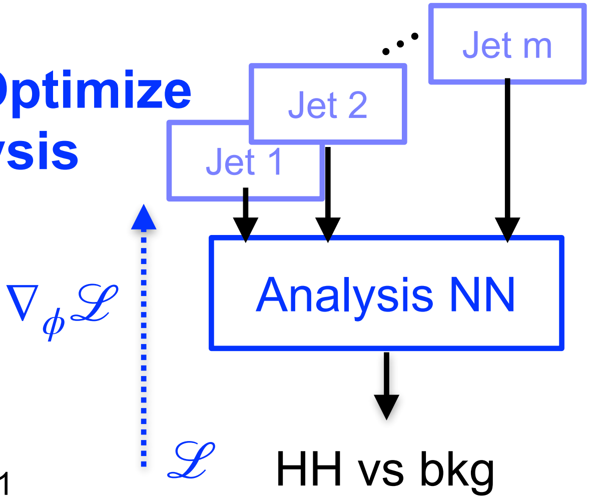




How do you know we this particle classifier is optimal for **every analysis**?



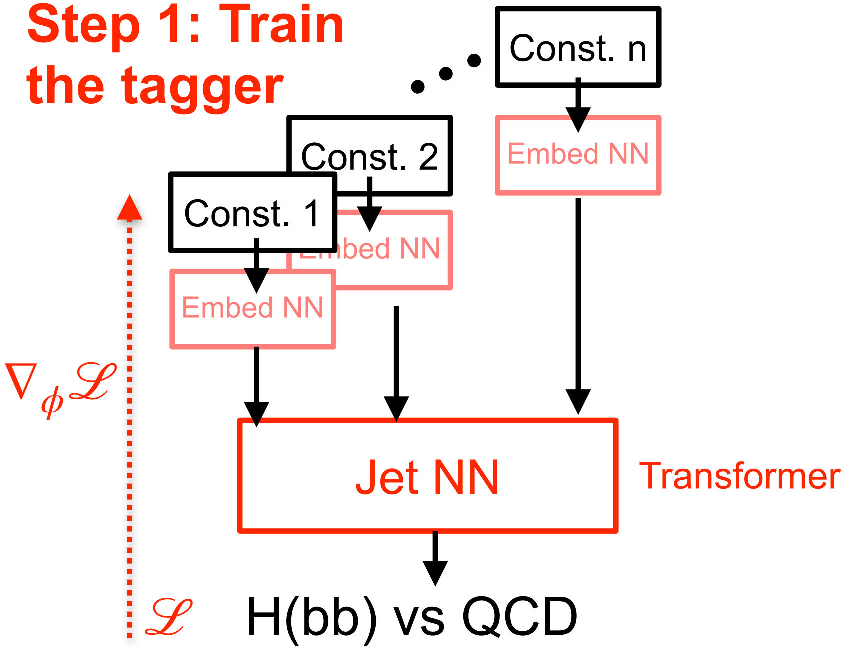
Step 2: Optimize the analysis



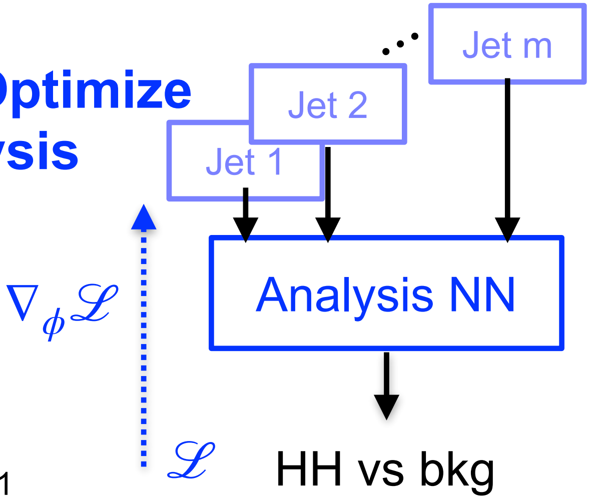
In some sense… we know it's not… analyses optimize tagger "working points"

Traditional Analysis

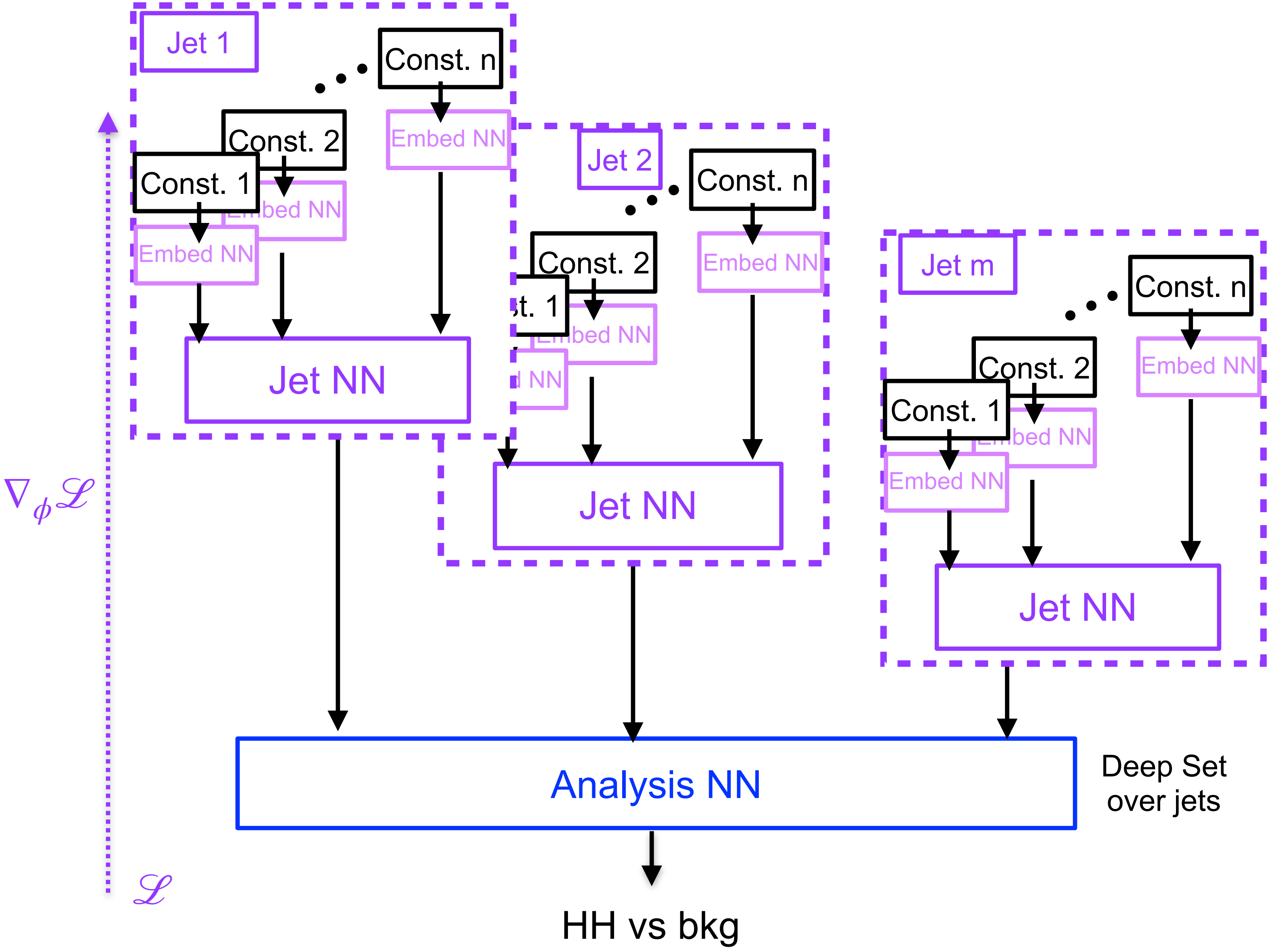
Step 1: Train the tagger



Step 2: Optimize the analysis

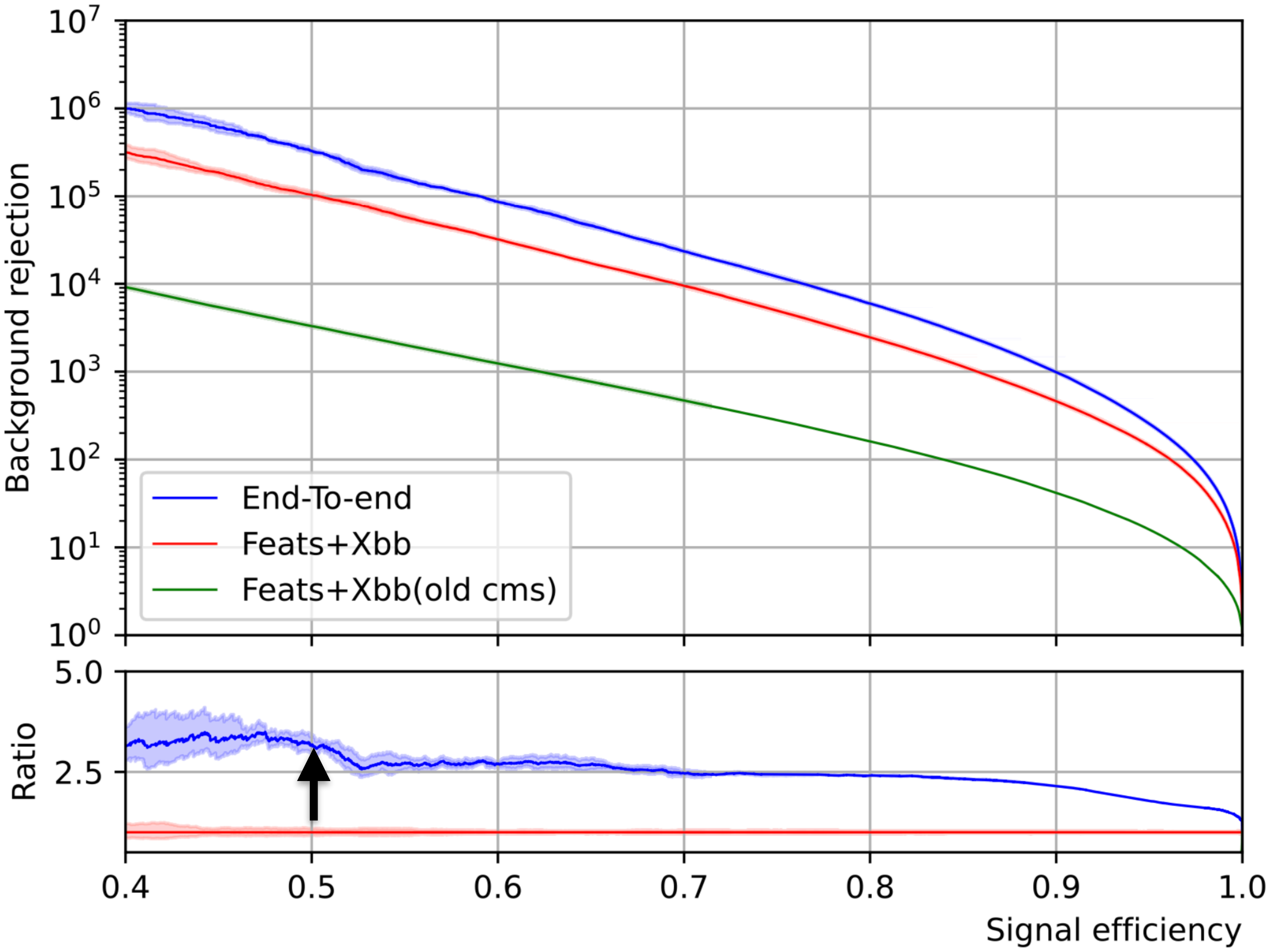


New paradigm



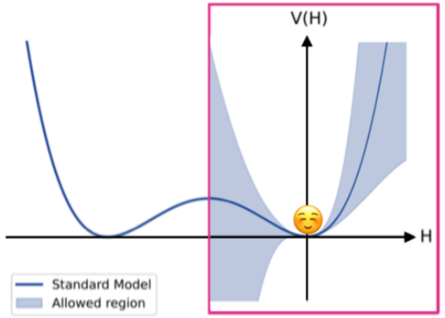
Dataset: CMS Open Data
Signal: X → HH → 4b
Bkg: QCD simulation

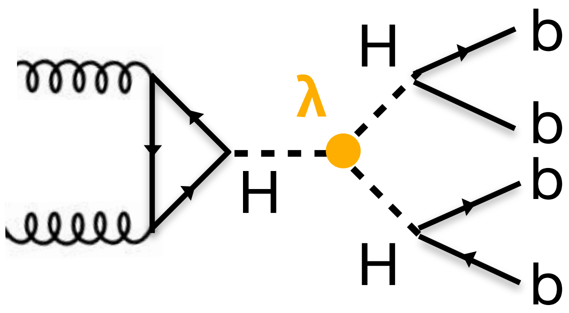
Jet Tagger: Transformer, ParT [2202.03772], start from published weights (pre-training with 100m jets in JetClass dataset)



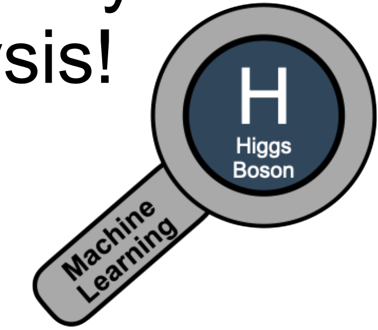
There *is* a gain with an end-to-end analysis!

Conclusion

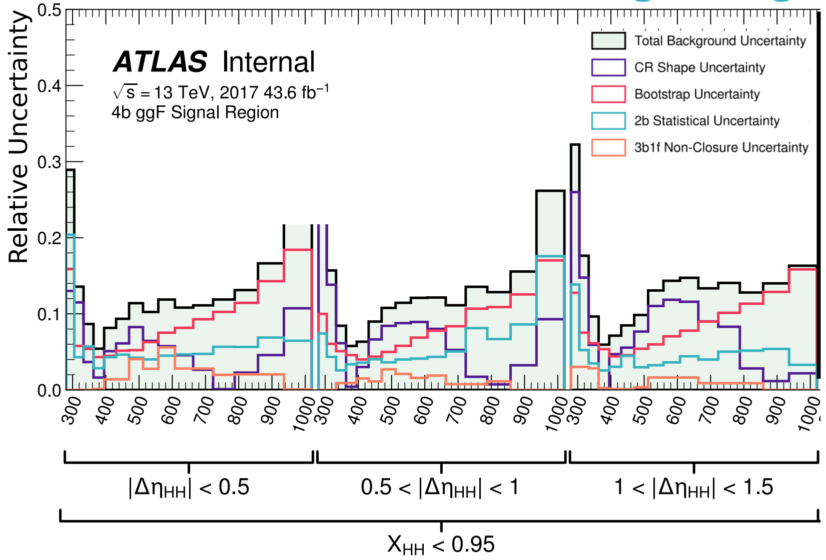


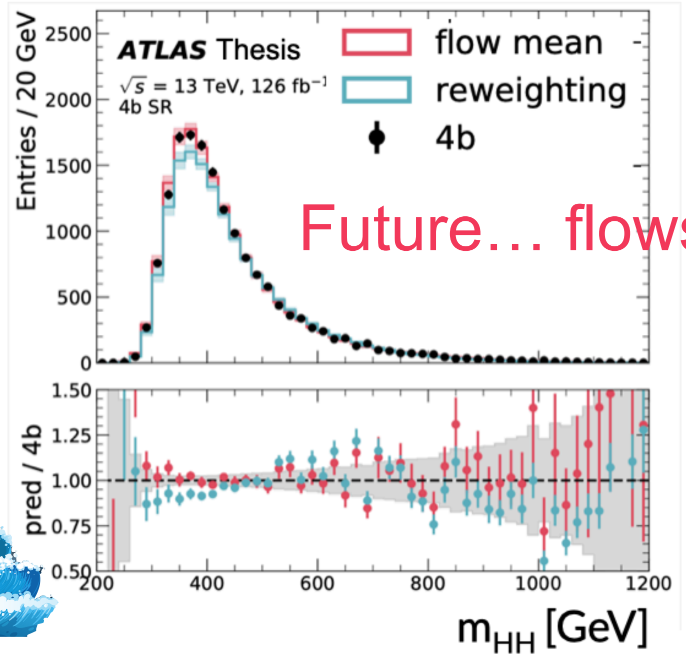


A background model with an accurate uncertainty is **key** to this analysis!



Run 2, neural reweighting

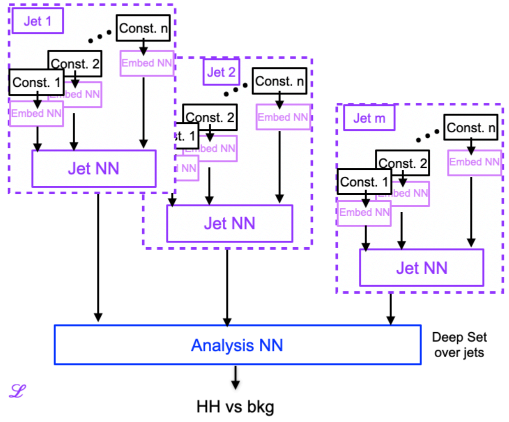




Future… flows?

Very exciting time for ML in Run 3 and bbbbeyond!

Backup



And exciting new ways to start rethinking our analyses…

How can the next generation of ML improve this?

Limitations of reweighting methods:

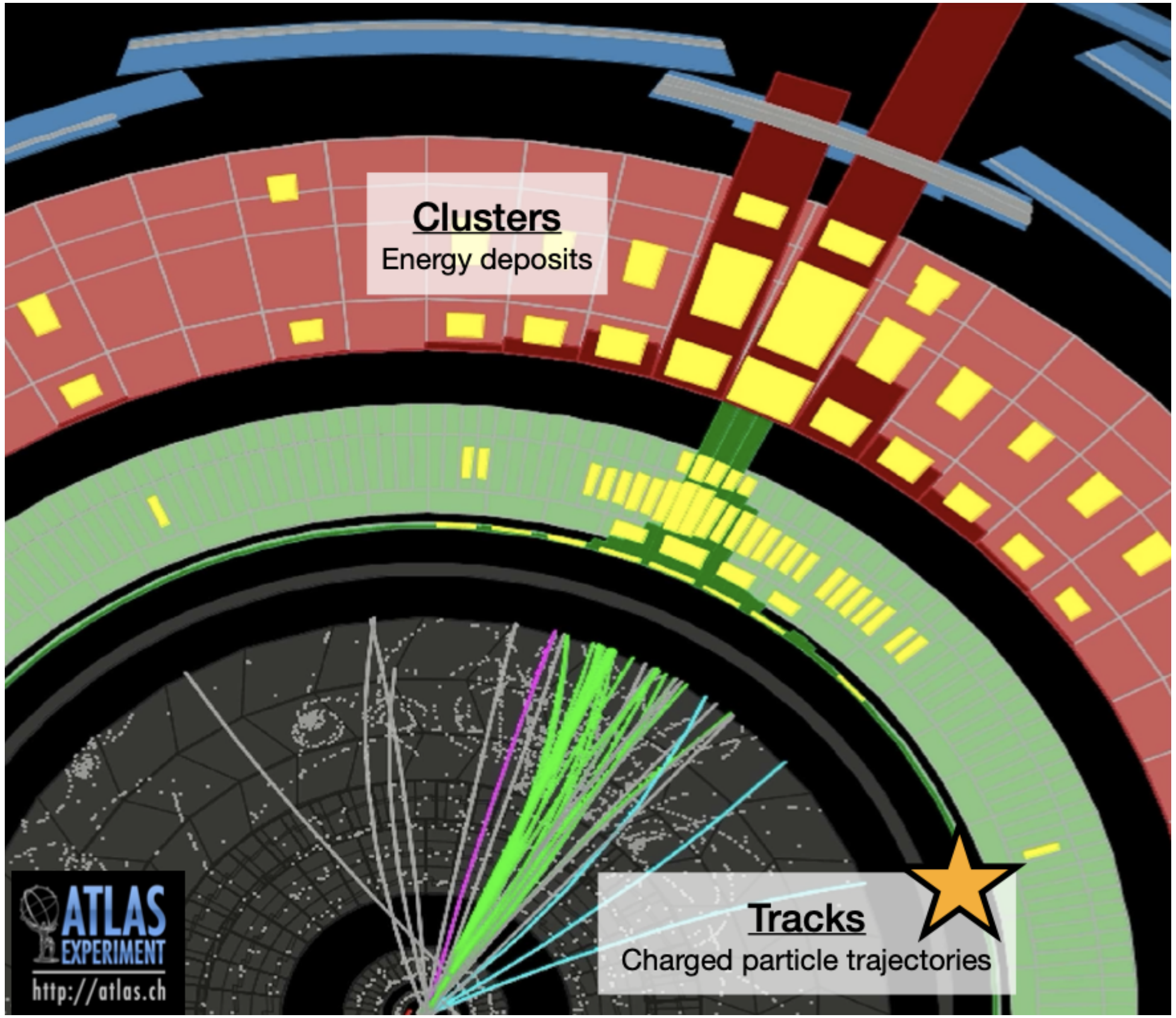
1. Choice of CR

2. Suffers in regions of finite support

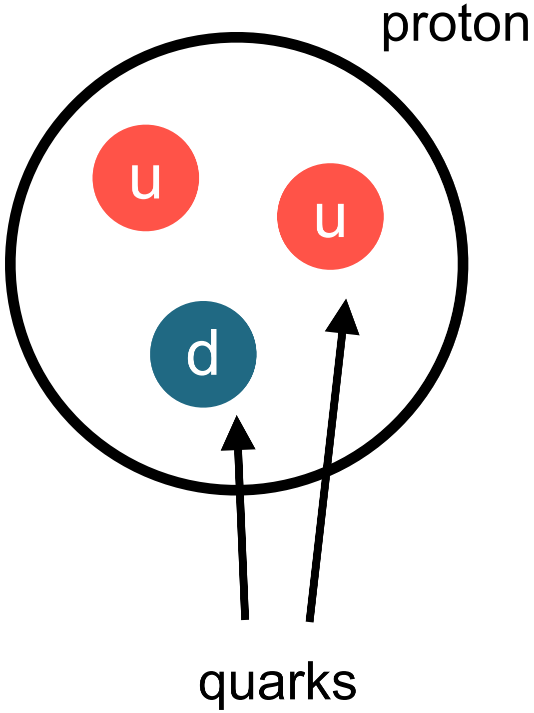
3. Limited by the stats of the “source distribution”

The European Physical Journal C
Vol 73 3 (2013) 2304

Quark signature



Quark ➜ reconstructed as collimated spray of particles



Non resonant signal

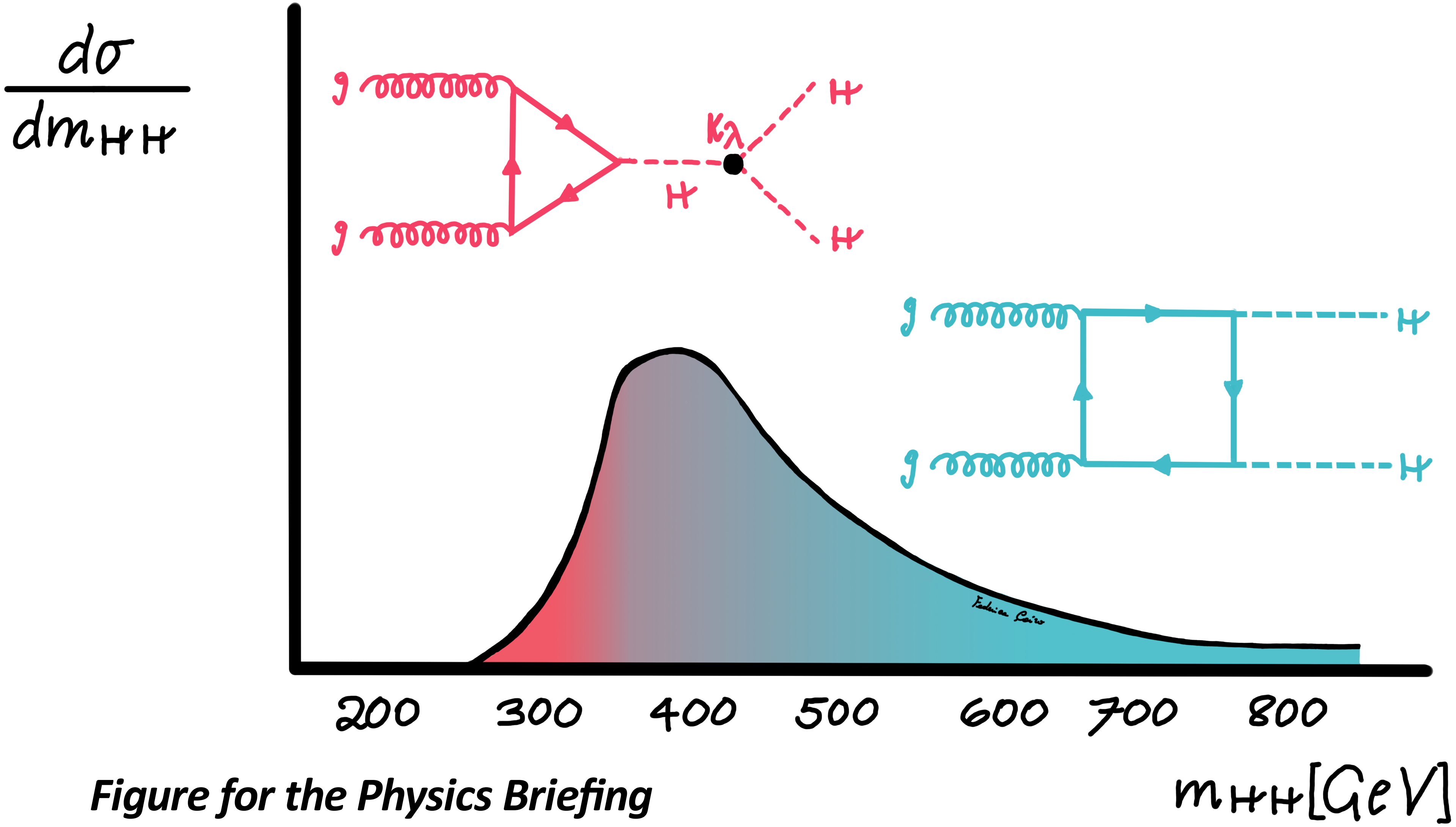
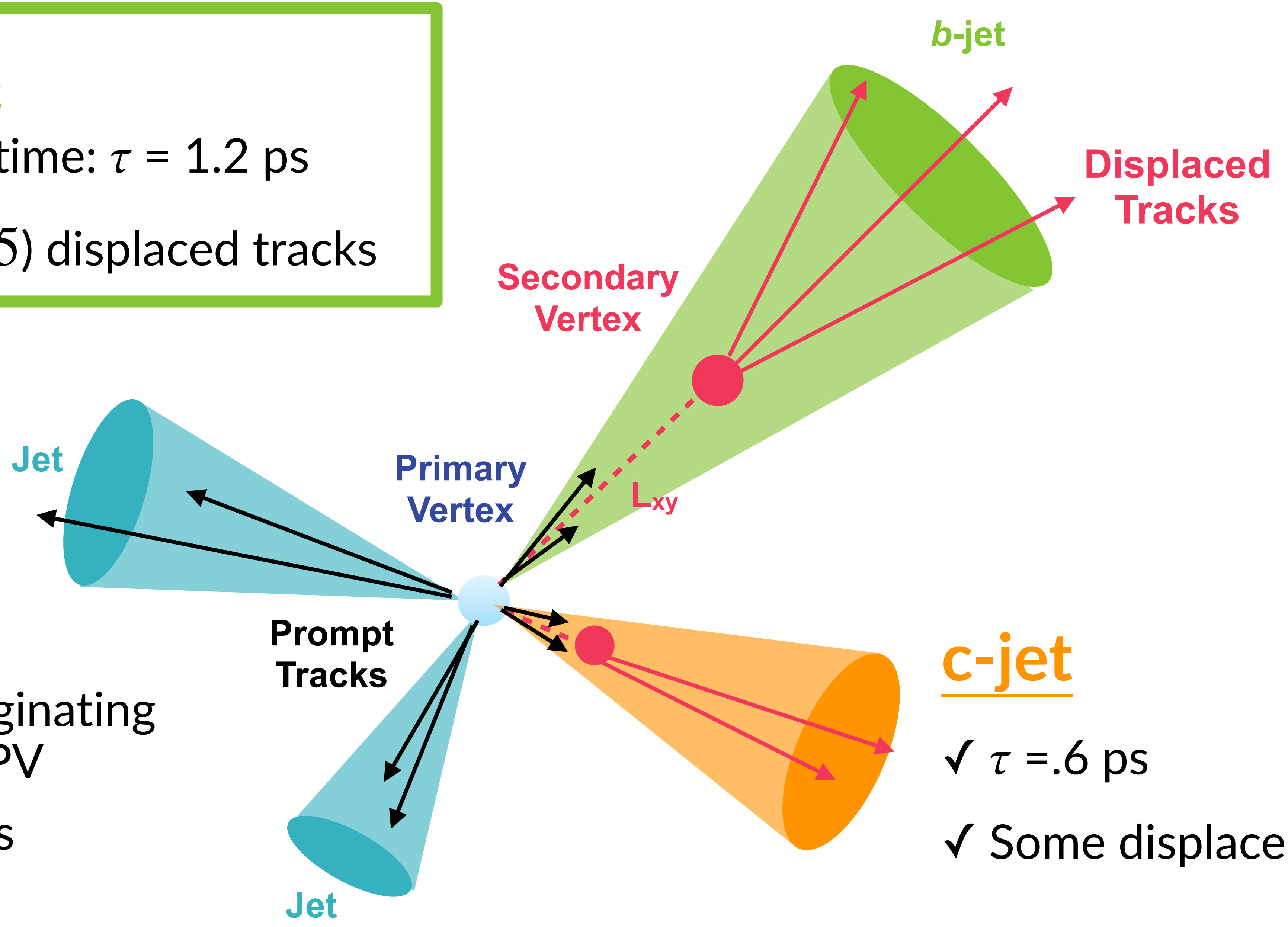


Figure for the Physics Briefing

😍 b-jet

✓ "Long" lifetime: τ = 1.2 ps

✓ Many ($\approx$ 5) displaced tracks



light jet

✓ Tracks originating from the PV

✓ Few tracks

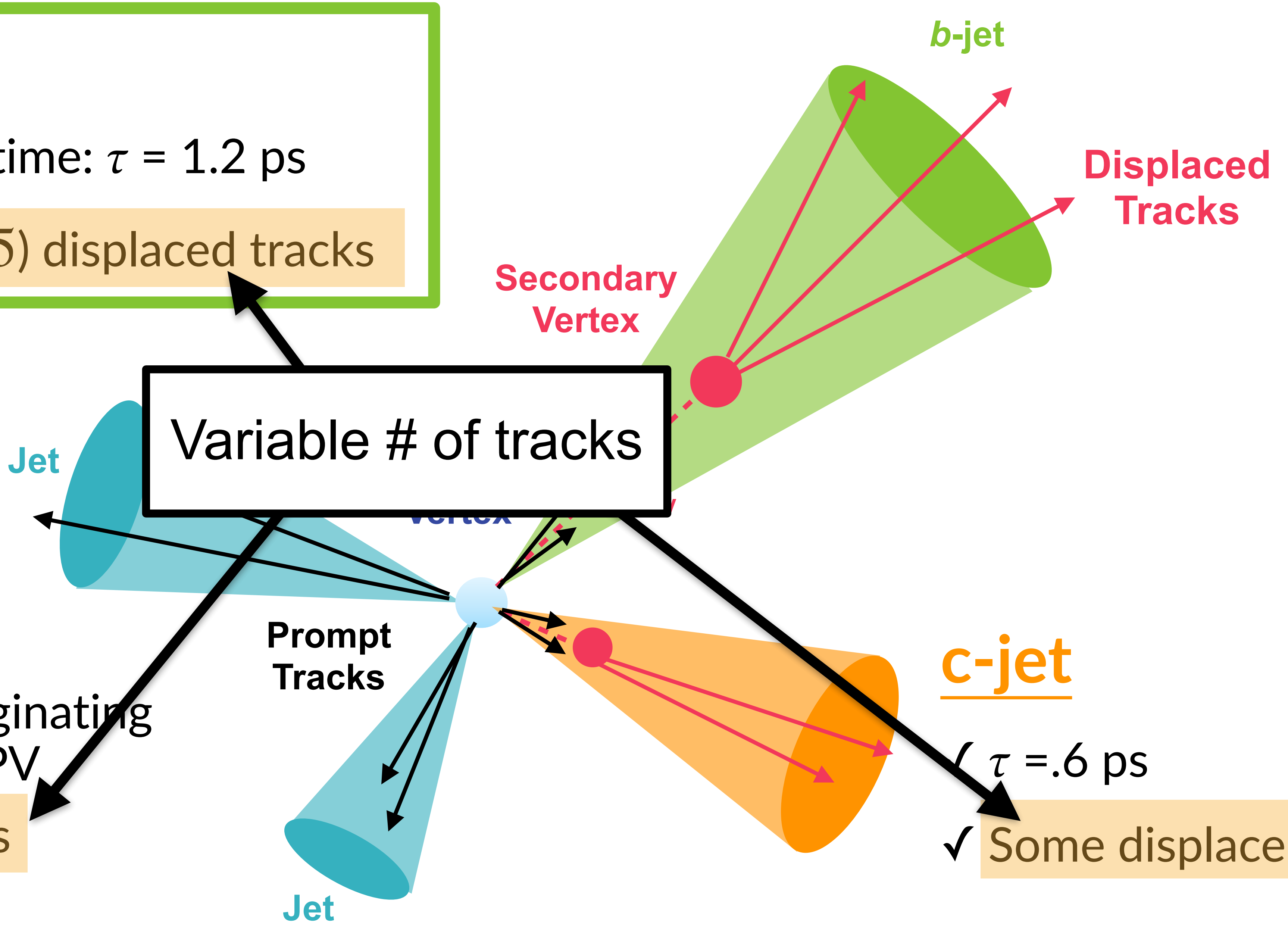
c-jet

✓ τ =.6 ps

✓ Some displaced tracks

b-jet

- ✓ "Long" lifetime: τ = 1.2 ps
- ✓ Many ($\approx$ 5) displaced tracks



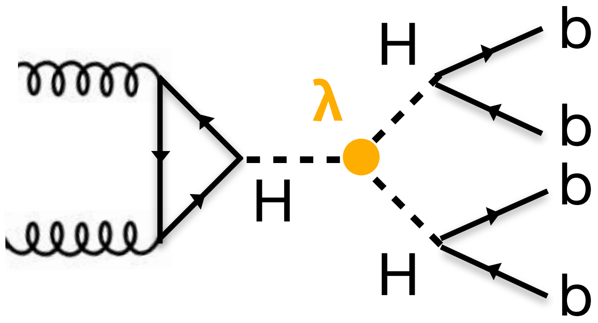
light jet

- ✓ Tracks originating from the PV
- ✓ Few tracks

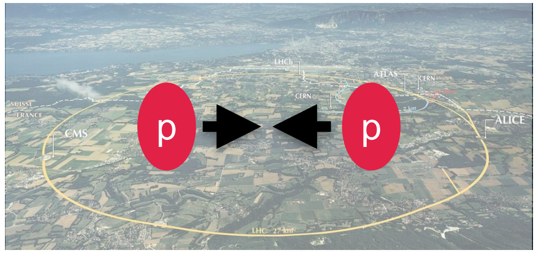
c-jet

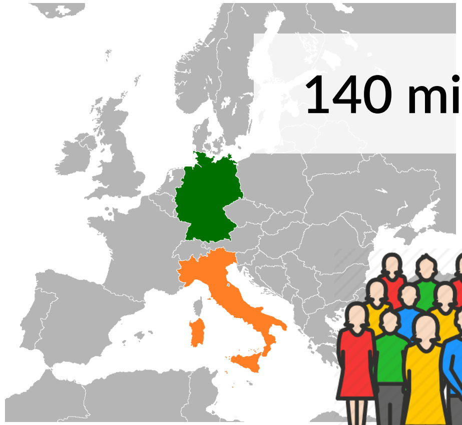
- ✓ τ =.6 ps
- ✓ Some displaced tracks

My research... guess who?



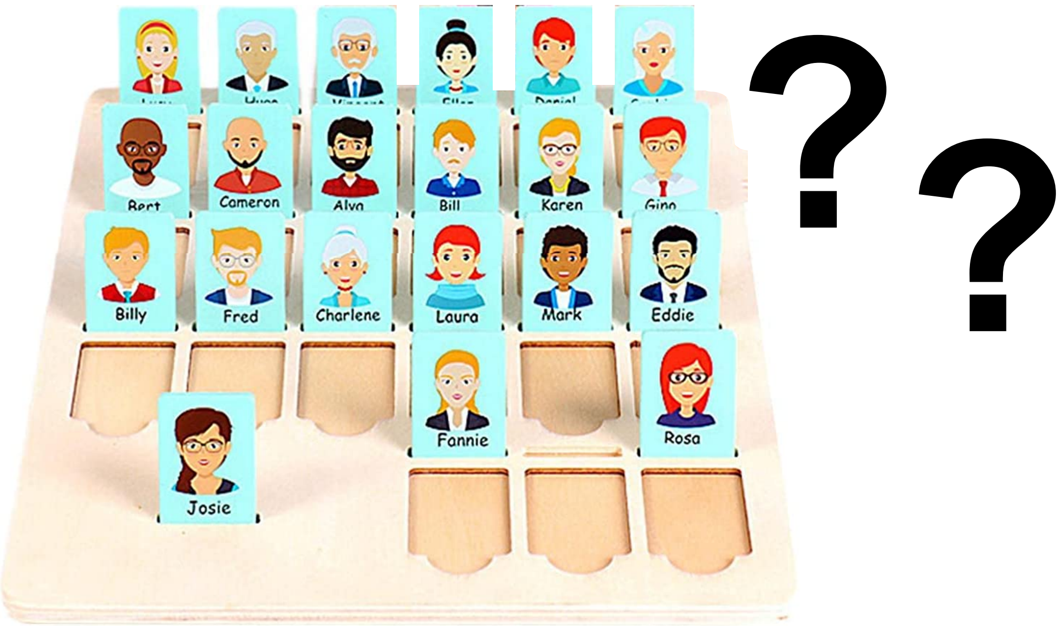
$$\frac{\text{1 signal event / (3 hour)}}{\text{1 billion collisions / second}} = 10^{-13} \; \frac{\text{signal}}{\text{background}}$$





140 million people in Germany + Italy

100,000 hairs / person



Rate mal wer ich bin??
Indovina Chi?

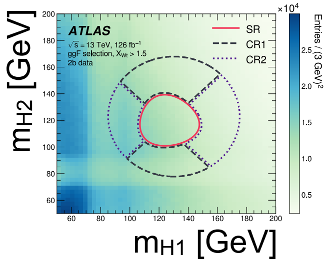
Examples of reweighting for backgrounds

NN

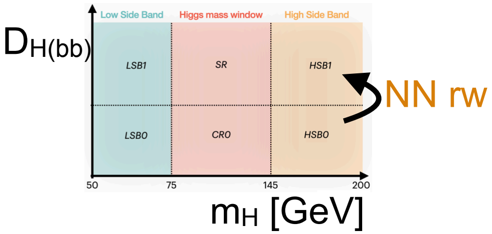
BDT

Low → high b-tag

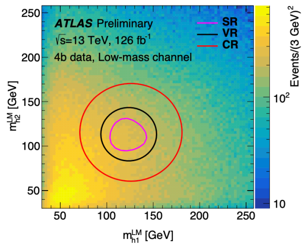
HH -> 4b
NR [2301.03212](2301.03212)
resonant [2202.07288](2202.07288)



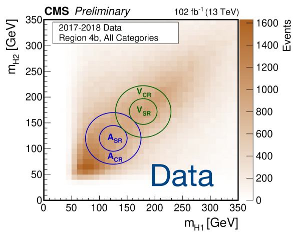
Y -> XH [2306.03637](2306.03637)
Anomalous jets analysis



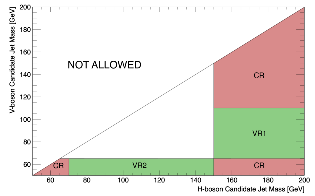
HH4b + MET
[1806.04030](1806.04030) and ATLAS-CONF-2023-048



CMS HH4b NR
[2202.09617](2202.09617)
Higgs Pairs 2022 talk

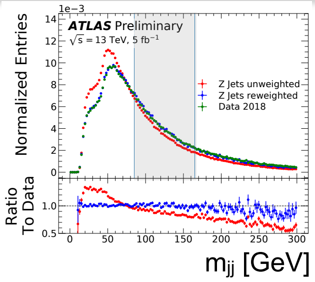


VH(qqbb)
[2007.05293](2007.05293)



MC → data

gg tagger
Match MC -> data
ATL-PHYS-PUB-2021-027



SALAD for CWoLA
Simulation Assisted Likelihood-free Anomaly Detection
MC -> data rw in sideband
ATL-COM-PHYS-2023-452

+ *many others!*

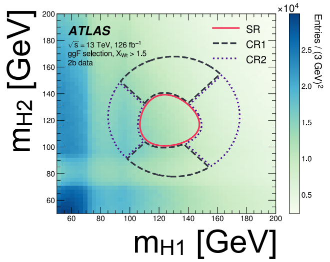
Q: How to quanitfy the uncertainty on the background model?

Examples of reweighting for backgrounds

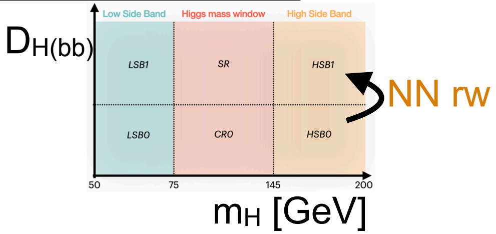
NN

Low → high b-tag

HH -> 4b
NR [2301.03212](2301.03212)
resonant [2202.07288](2202.07288)

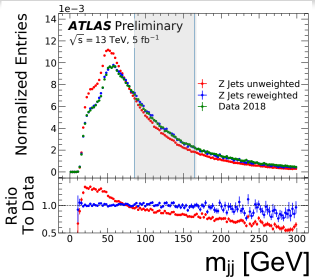


Y -> XH [2306.03637](2306.03637)
Anomalous jets analysis



MC → data

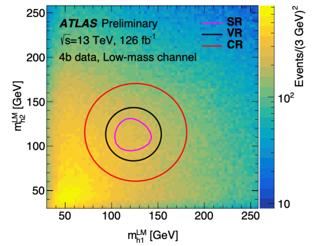
gg tagger
Match MC -> data
ATL-PHYS-PUB-2021-027



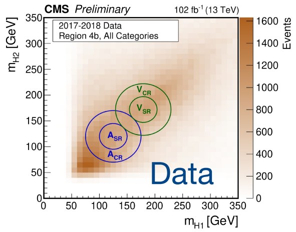
SALAD for CWoLA
Simulation Assisted Likelihood-free Anomaly Detection
MC -> data rw in sideband
ATL-COM-PHYS-2023-452

BDT

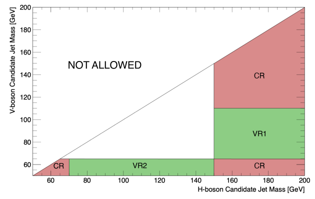
HH4b + MET
[1806.04030](1806.04030) and ATLAS-CONF-2023-048



CMS HH4b NR
[2202.09617](2202.09617)
Higgs Pairs 2022 talk



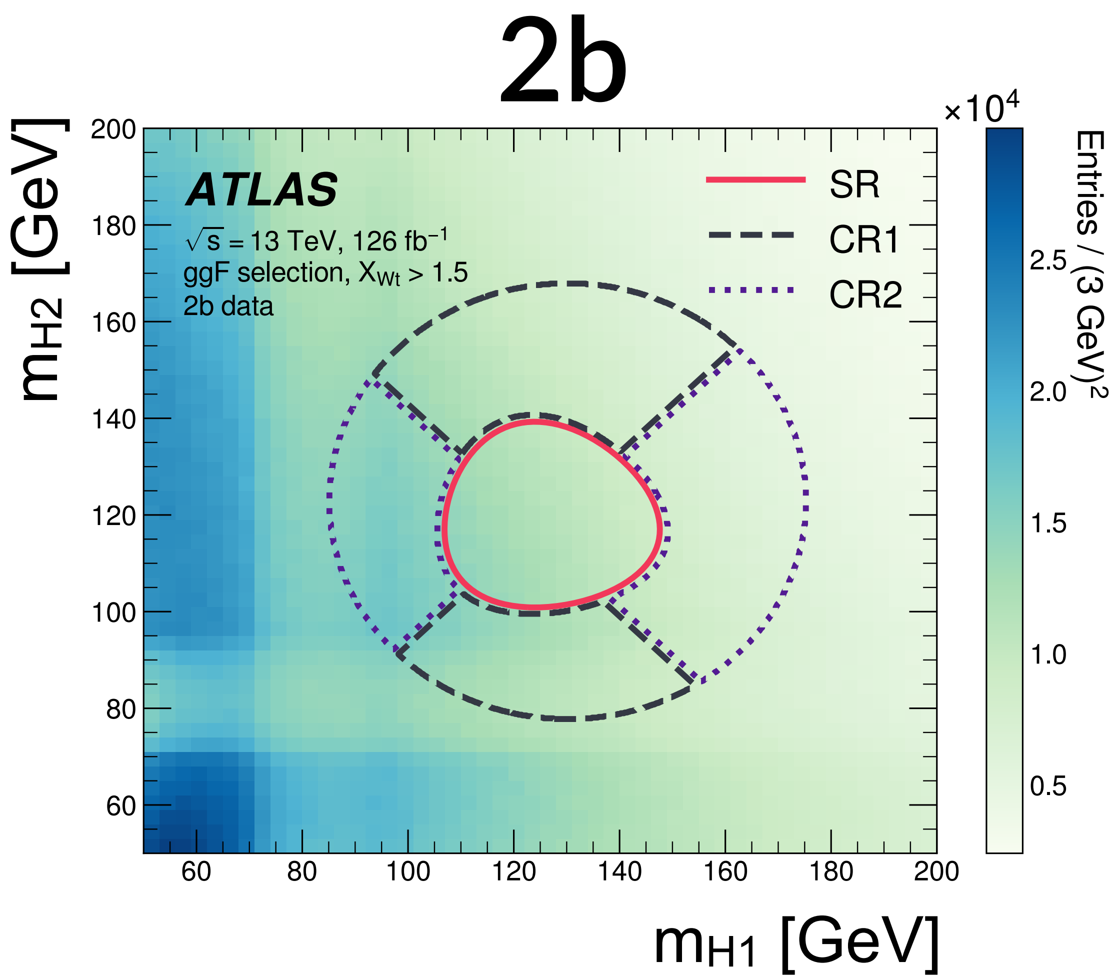
VH(qqbb)
[2007.05293](2007.05293)

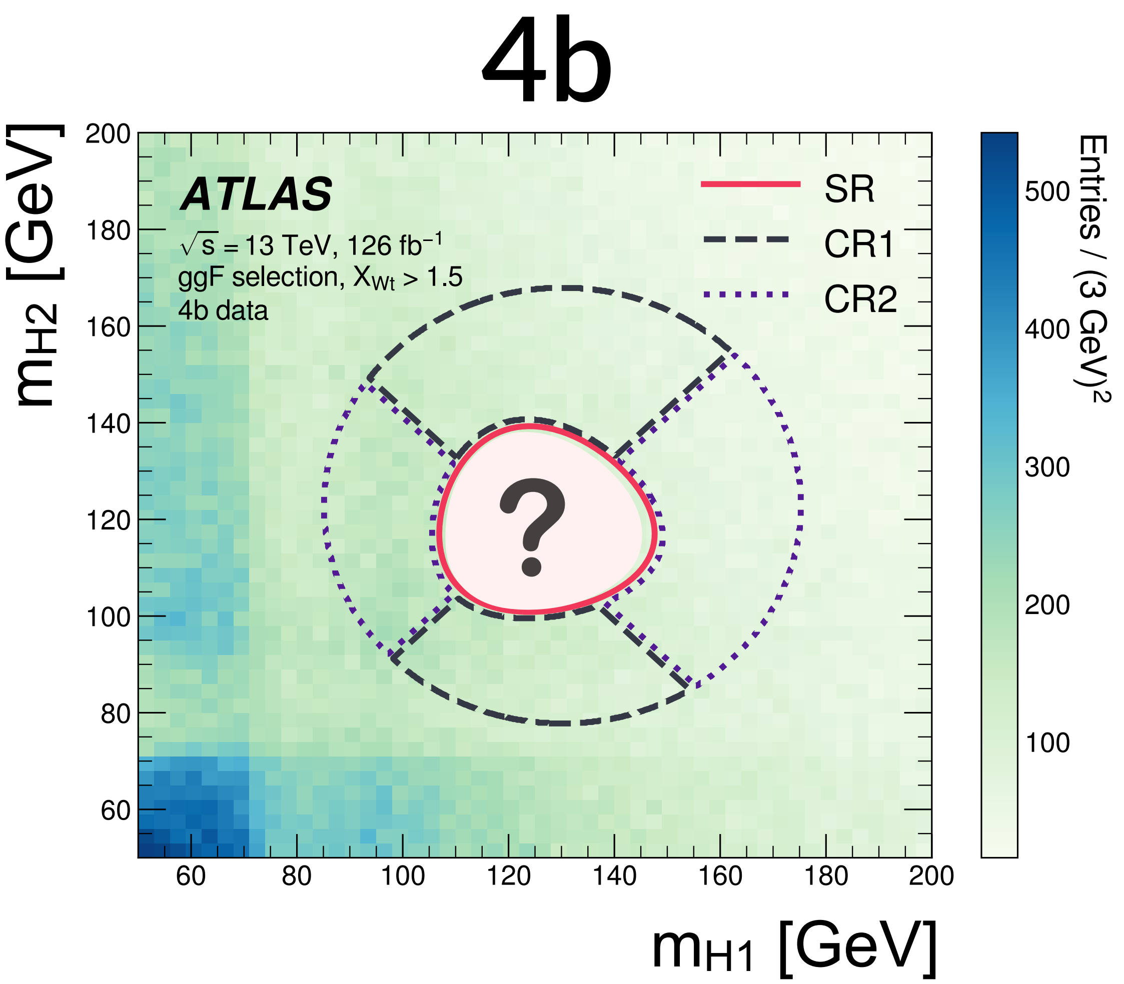


+ *many others!*

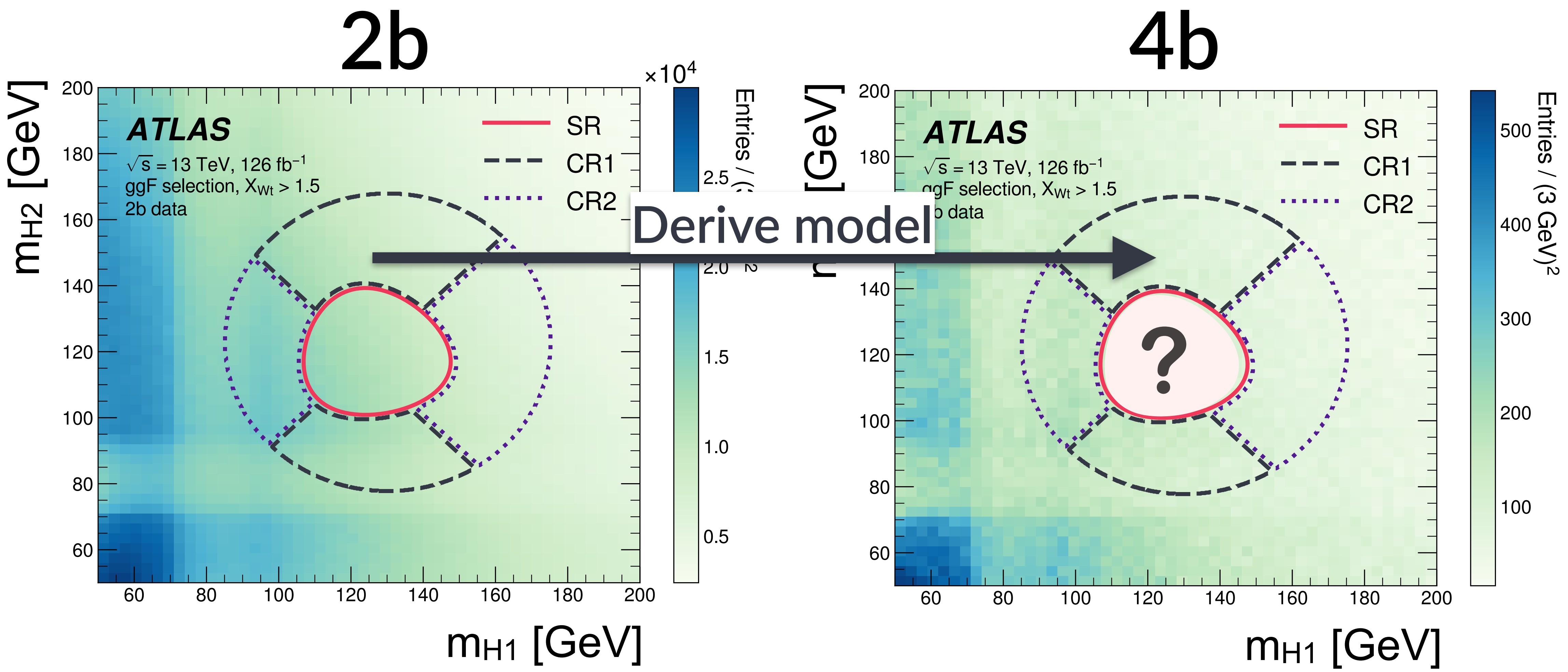
Q: How to quanitfy the uncertainty on the background model?

The background model

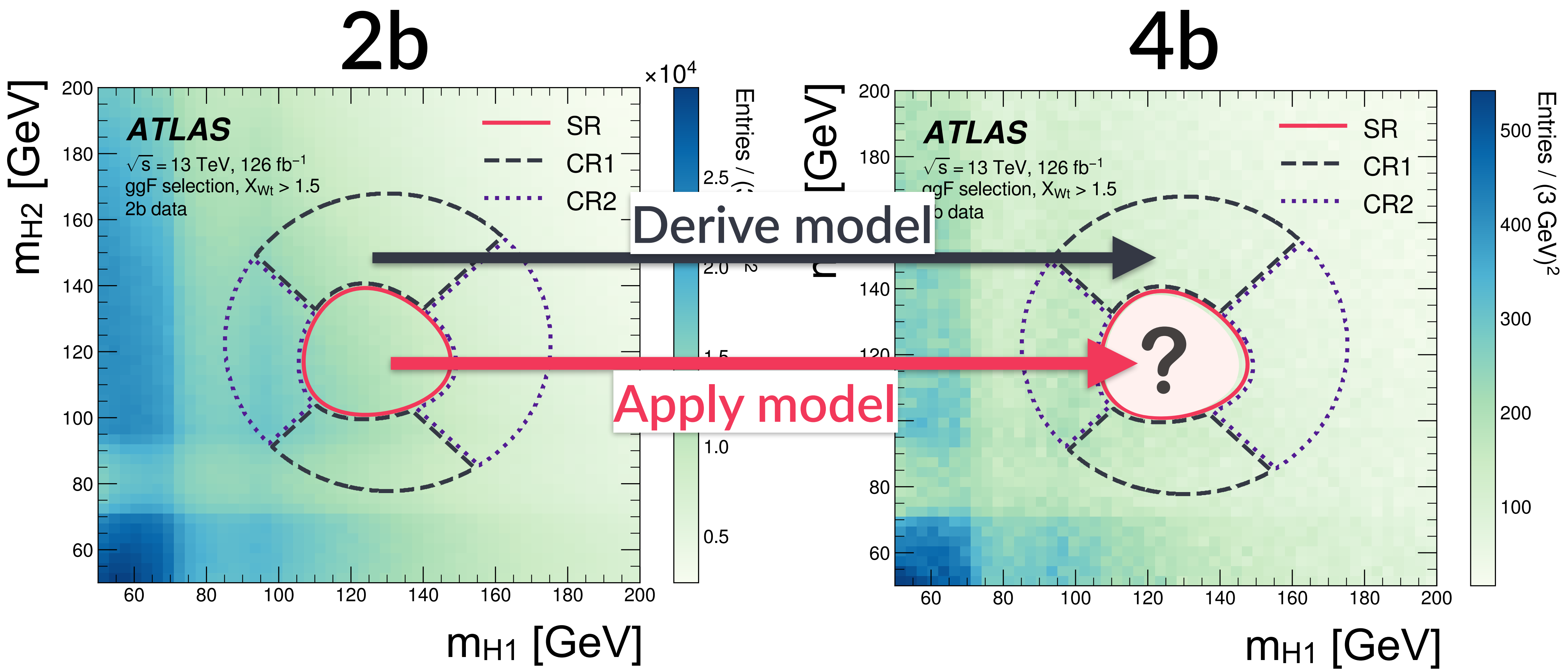




The background model



The background model



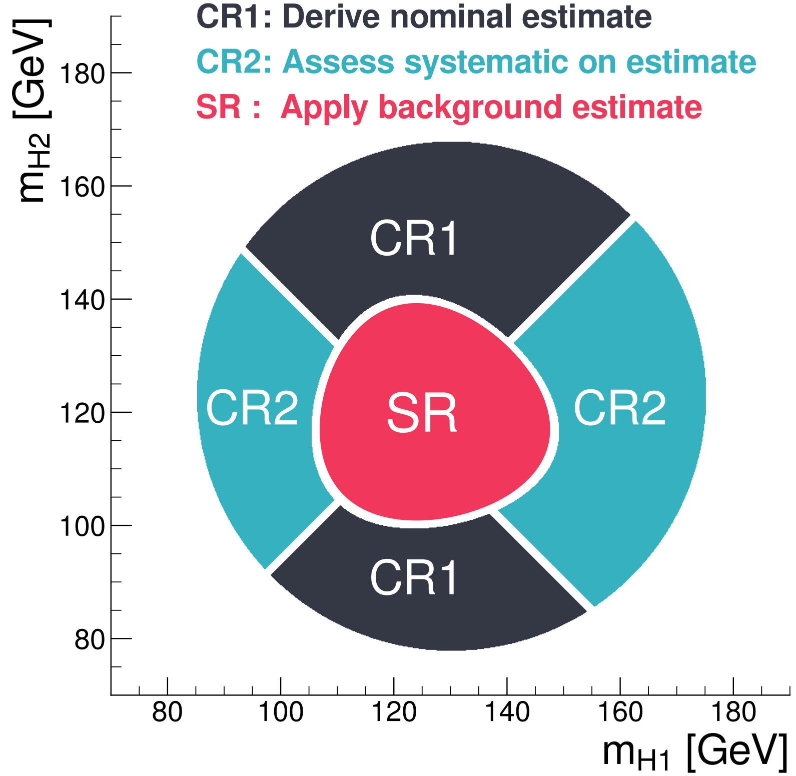
Reweighting — key idea

Multi-dimensional reweighting 2b ➜ 4b .

source distribution → target distribution

$$p_{4b}(x) = w(x) \cdot p_{2b}(x)$$

Want to learn



Let Q(x) be a NN mapping from 2b ➜ 4b.

$$\mathscr{L}[Q(x)] = \mathbb{E}_{x\sim p_{2b}}\left[\exp\left(\frac{Q(x)}{2}\right)\right] + \mathbb{E}_{x\sim p_{4b}}\left[\exp\left(-\frac{Q(x)}{2}\right)\right]$$

Minimize loss in **Control Region**.

$$Q^*(x) = \arg\min_Q \mathscr{L}[Q(x)] = \log \frac{p_{4b}(x)}{p_{2b}(x)}$$

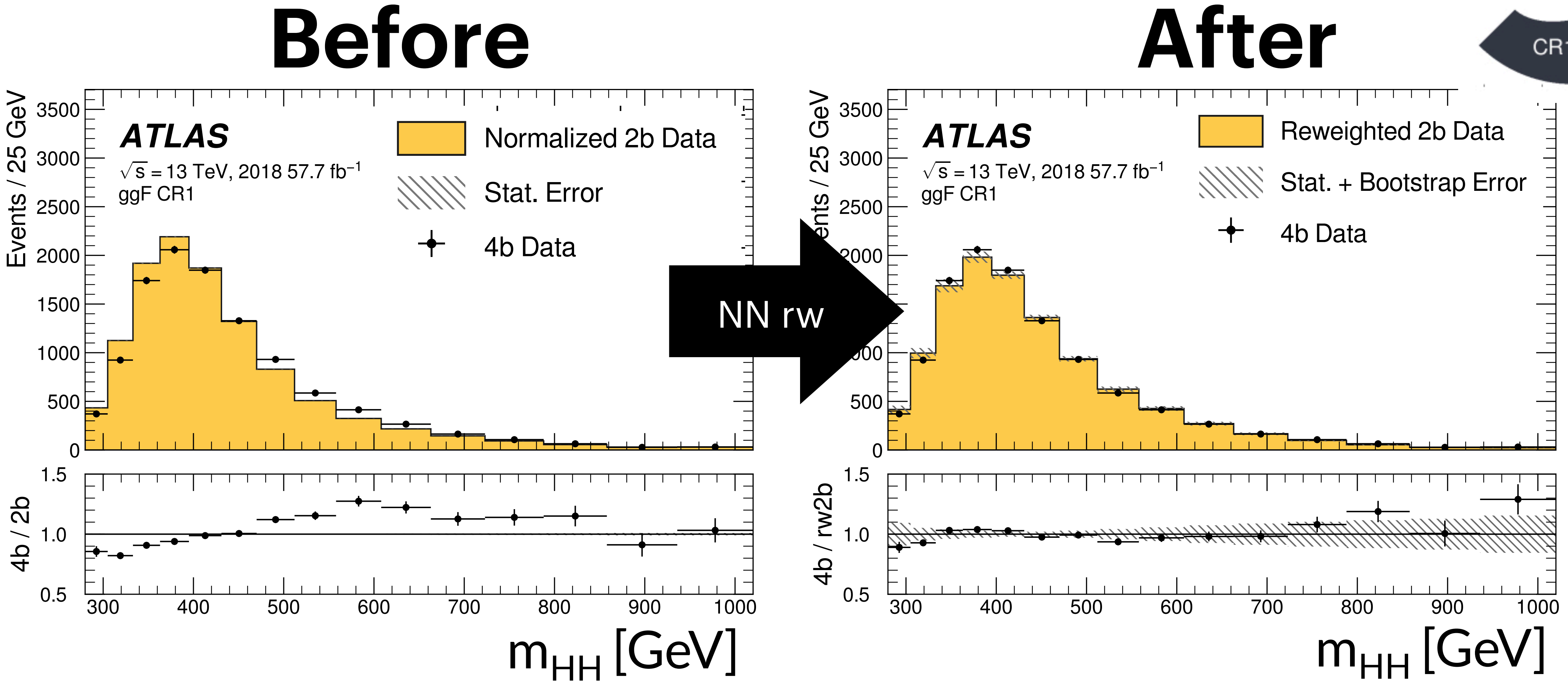
$$\implies w(x) = e^{Q^*(x)}$$

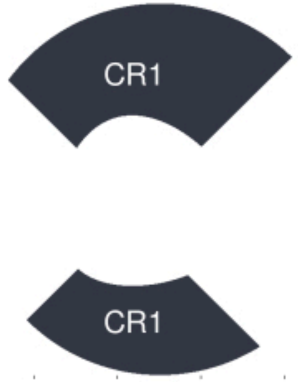
Apply w(x) to 2b Signal Region to get 4b prediction

2301.03212

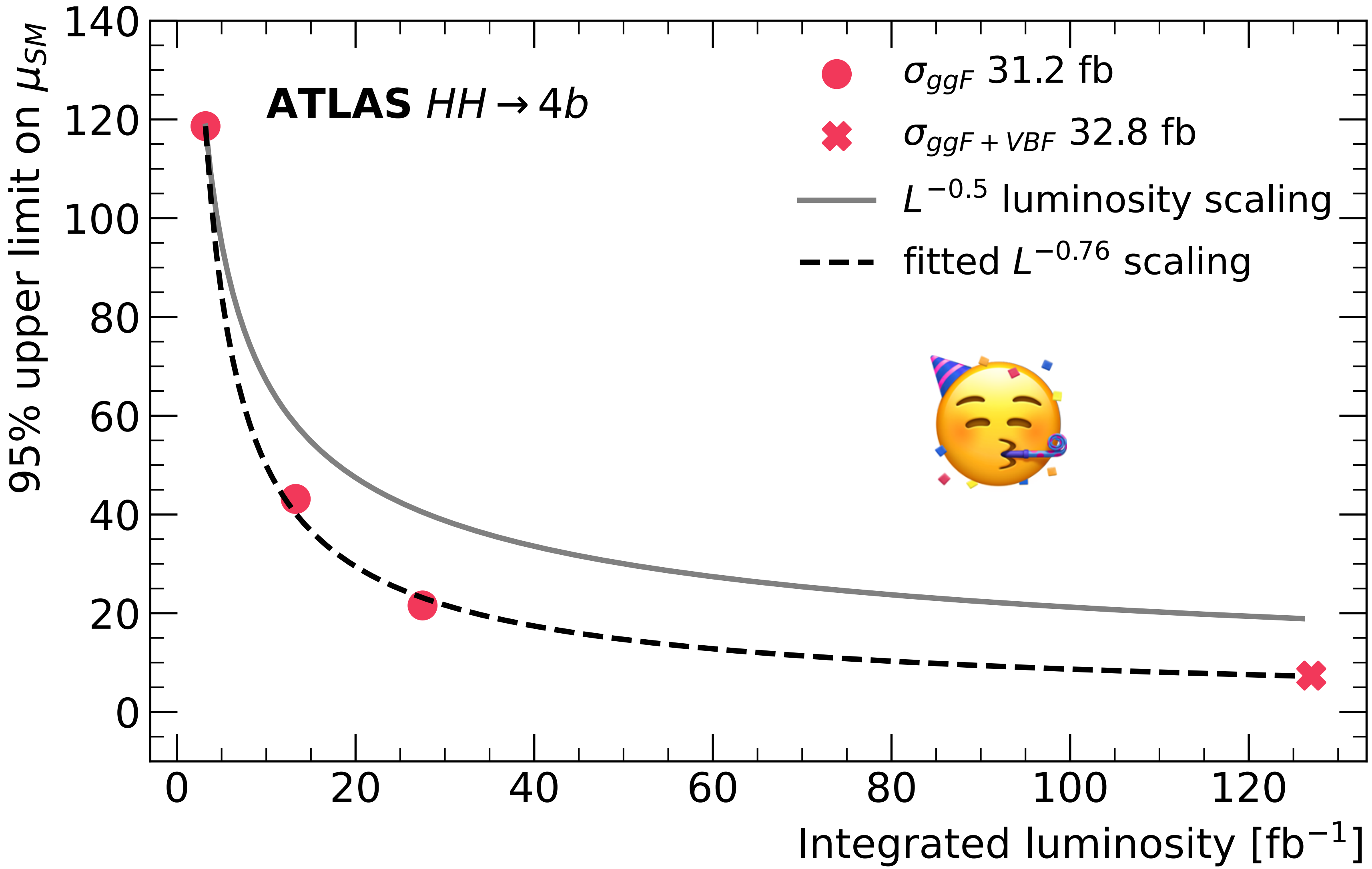
How does it work?

Closure test



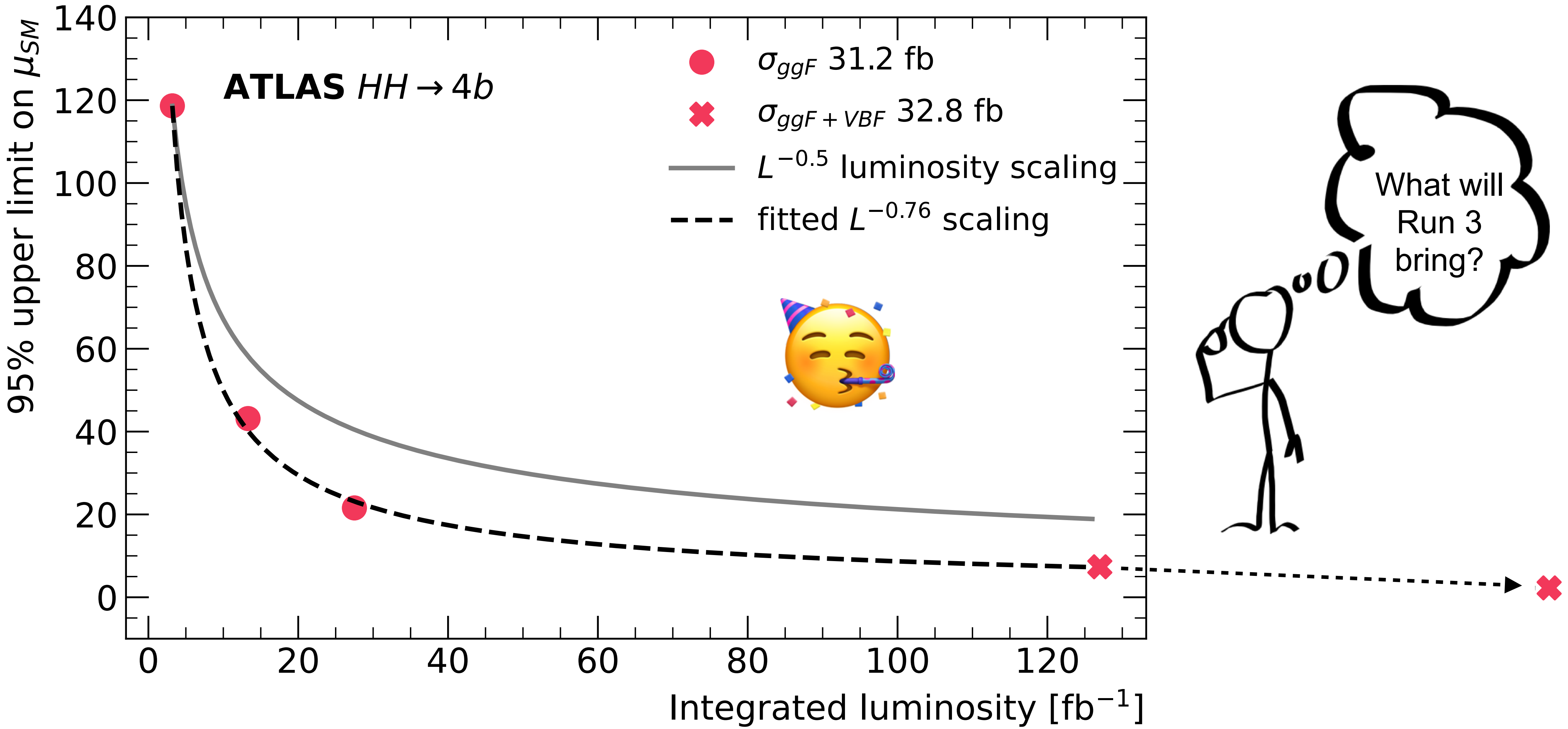


What's next?

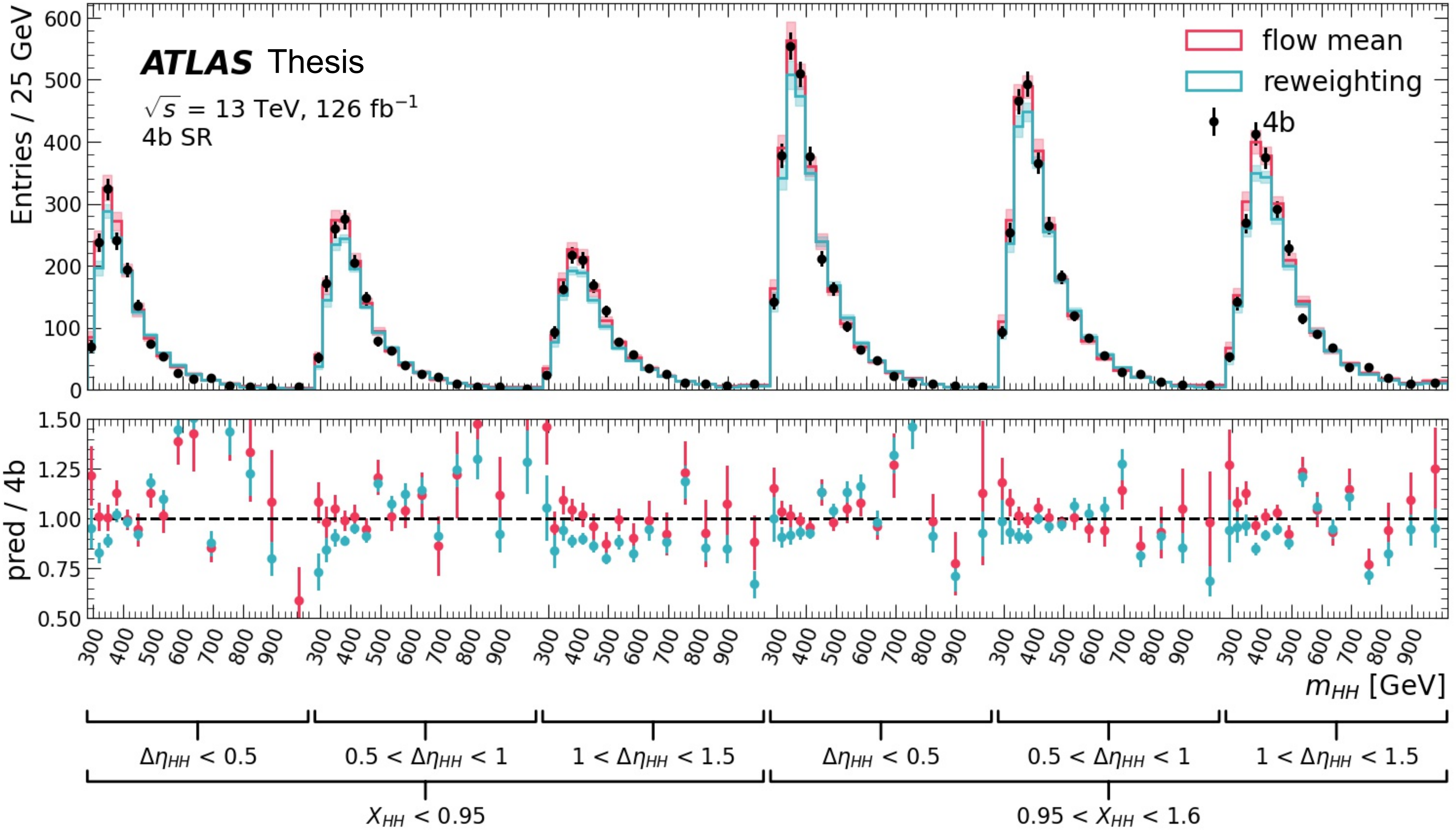


ATLAS $HH \to 4b$
σ_{ggF} 31.2 fb
$\sigma_{ggF+VBF}$ 32.8 fb
$L^{-0.5}$ luminosity scaling
fitted $L^{-0.76}$ scaling
95% upper limit on μ_{SM}
Integrated luminosity [fb^{-1}]
0
20
40
60
80
100
120
140

What's next?

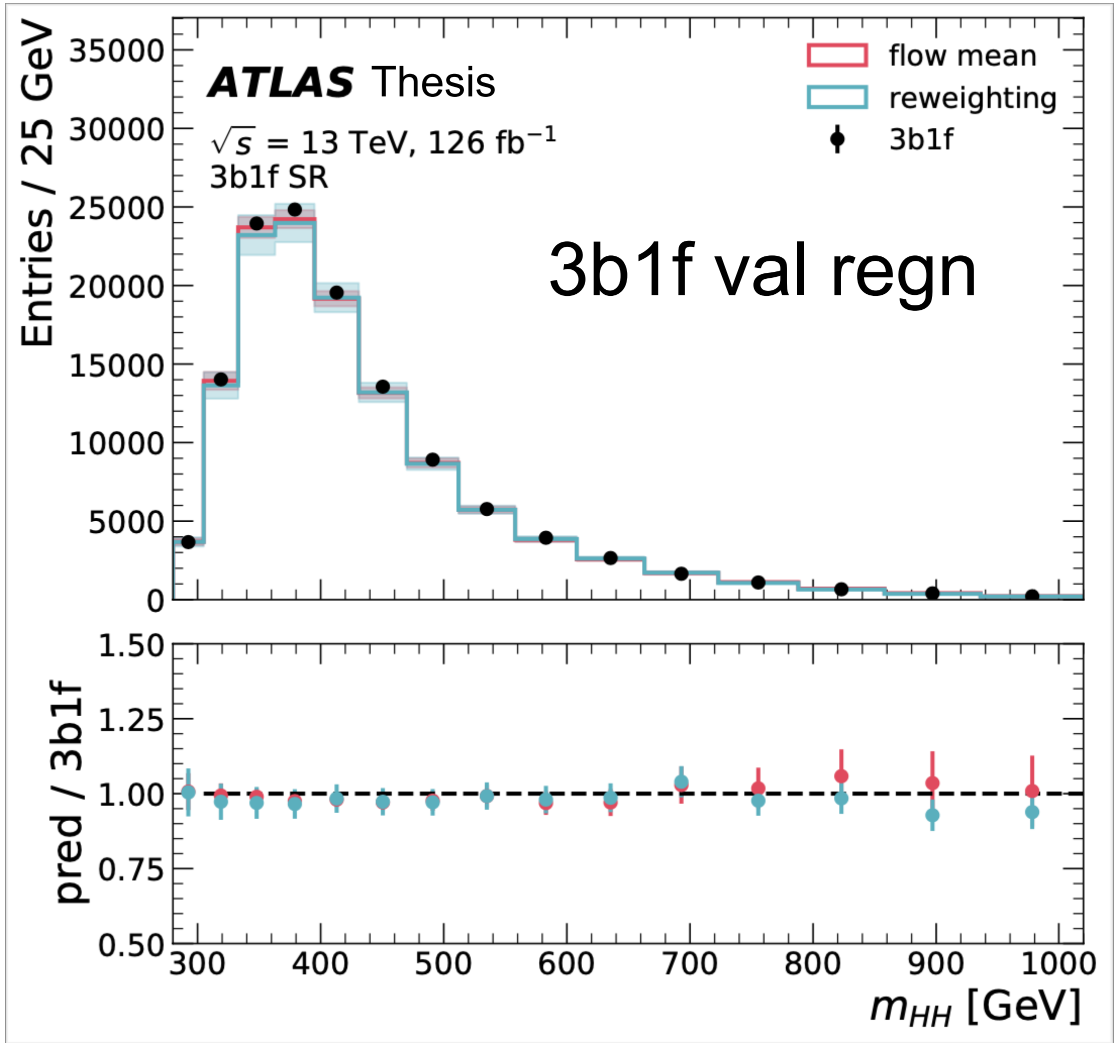


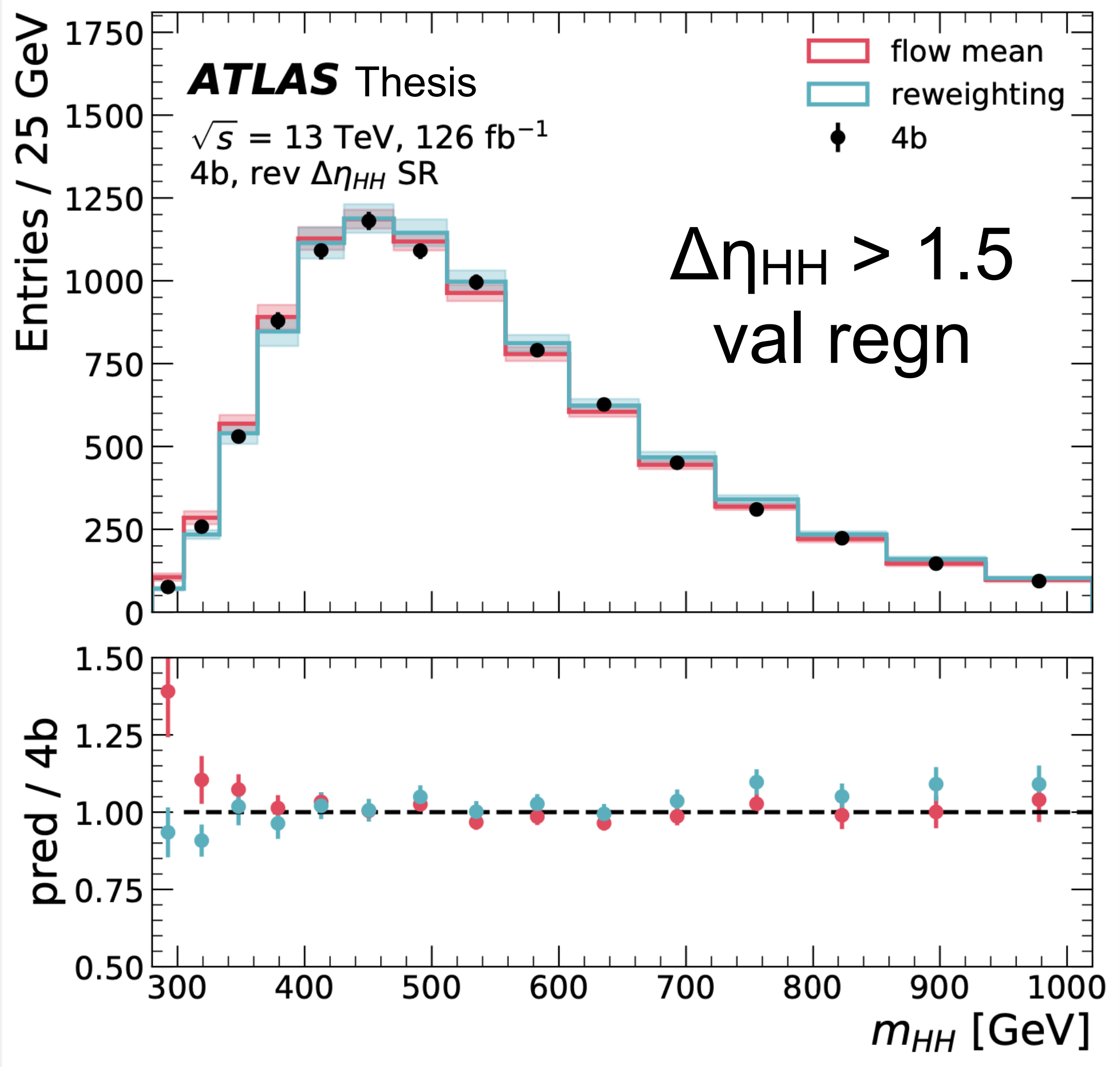
95% upper limit on μ_{SM}
140
120
100
80
60
40
20
0
ATLAS $HH \rightarrow 4b$
σ_{ggF} 31.2 fb
$\sigma_{ggF+VBF}$ 32.8 fb
$L^{-0.5}$ luminosity scaling
fitted $L^{-0.76}$ scaling
0
20
40
60
80
100
120
Integrated luminosity [fb^{-1}]
What will Run 3 bring?



ATLAS Thesis
$\sqrt{s}$ = 13 TeV, 126 fb^{-1}
4b SR
flow mean
reweighting
4b
Entries / 25 GeV
pred / 4b
m_{HH} [GeV]
$\Delta\eta_{HH}$ < 0.5
0.5 < $\Delta\eta_{HH}$ < 1
1 < $\Delta\eta_{HH}$ < 1.5
$\Delta\eta_{HH}$ < 0.5
0.5 < $\Delta\eta_{HH}$ < 1
1 < $\Delta\eta_{HH}$ < 1.5
X_{HH} < 0.95
0.95 < X_{HH} < 1.6

3b + rev $\Delta\eta_{HH}$





Event reconstruction

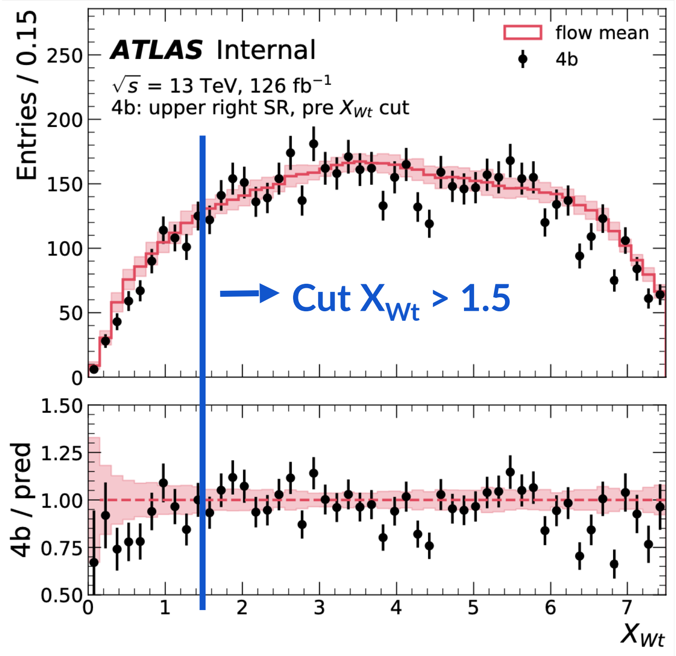
1 $m_{h1}, m_{h2} \sim p(m_{h1}, m_{h2})$ [GP]

2 $\begin{bmatrix} p_{T,H1}, p_{T,H2} \\ \eta_{H1}, \eta_{H2} \\ \Delta\Phi_{HH}, X_{Wt} \end{bmatrix} \sim p(\text{evt vars}|\ m_{h1},m_{h2})$ [flow]

3 Construct Higgs 4-vectors

HC 1: $(p_{T,\,H1}, \eta_{H1},\ 0,\ m_{H1})$
HC 2: $(p_{T,\,H2}, \eta_{H2},\ \Delta\Phi_{HH},\ m_{H2})$

4



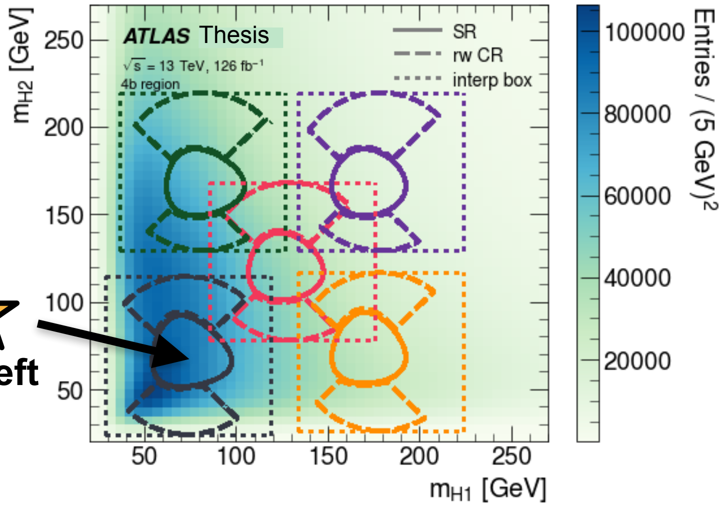
reweighting better
GP+flow better

C.f. the 4b NR bkg validation regions

Compare the observed and predicted yields

shifted regions

	obs	rw	flow	1 - rw / obs [%]	1 - flow / obs [%]
3b1f	180044	175817.9	175416.8	2.3	2.6
rev $\|\Delta\eta_{HH}\|$	16113	16462.7	16185.9	-2.2	-0.5
lower left	40578	48708.9	39252.2	-20.0	3.3
lower right	12377	14648.5	11982.7	-18.4	3.2
upper right	5751	5543.0	5825.9	3.6	-1.3
upper left	19075	19504.7	19833.4	-2.3	-4.0
4b	16171	15423.7	16564.8	4.6	-2.4



lower left

Shape modelling

Normalize histograms to the target, and compare shapes

reweighting better
GP+flow better

shifted regions

	3b1f		rev $\|\Delta\eta_{HH}\|$		lower left		lower right		upper right		upper left	
	rw	flow	rw	flow	rw	flow	rw	flow	rw	flow	rw	flow
3d disc	1.85	2.24	1.37	1.25	5.52	2.19	1.32	0.98	1.35	0.93	2.54	1.60
m_{HH}	2.29	3.77	1.34	1.60	2.70	1.98	1.26	0.78	1.30	0.91	4.44	4.70
$m_{HH,cor2}$	2.21	3.81	1.63	2.34	8.26	2.21	1.61	1.03	1.97	1.64	2.85	4.49
$\Delta\eta_{HH}$	1.38	1.94	1.16	3.90	6.95	1.37	1.52	1.07	1.51	1.63	1.99	0.98
$p_{T,H1}$	1.61	1.58	1.42	1.48	1.76	0.95	0.70	0.58	1.07	1.14	2.85	1.77
$p_{T,H2}$	2.64	2.50	1.74	1.51	2.81	1.83	2.83	1.17	1.82	1.06	0.92	1.49
η_{H1}	1.99	2.07	2.74	1.51	2.33	1.54	2.24	0.87	1.45	0.82	2.36	1.74
η_{H2}	1.37	1.36	2.07	1.17	2.62	1.35	4.93	1.11	0.85	0.75	1.33	1.16
$\Delta\phi_{HH}$	1.88	1.85	2.08	1.16	9.64	1.46	18.81	1.04	3.23	2.61	2.24	1.40
X_{Wt}	2.49	3.73	1.74	1.71	17.64	1.27	5.29	0.79	1.39	1.45	1.33	0.82
$p_{T,HH}$	2.54	0.93	2.35	0.87	6.06	1.13	1.76	1.00	1.50	1.23	1.80	0.80

GP fitted length scales

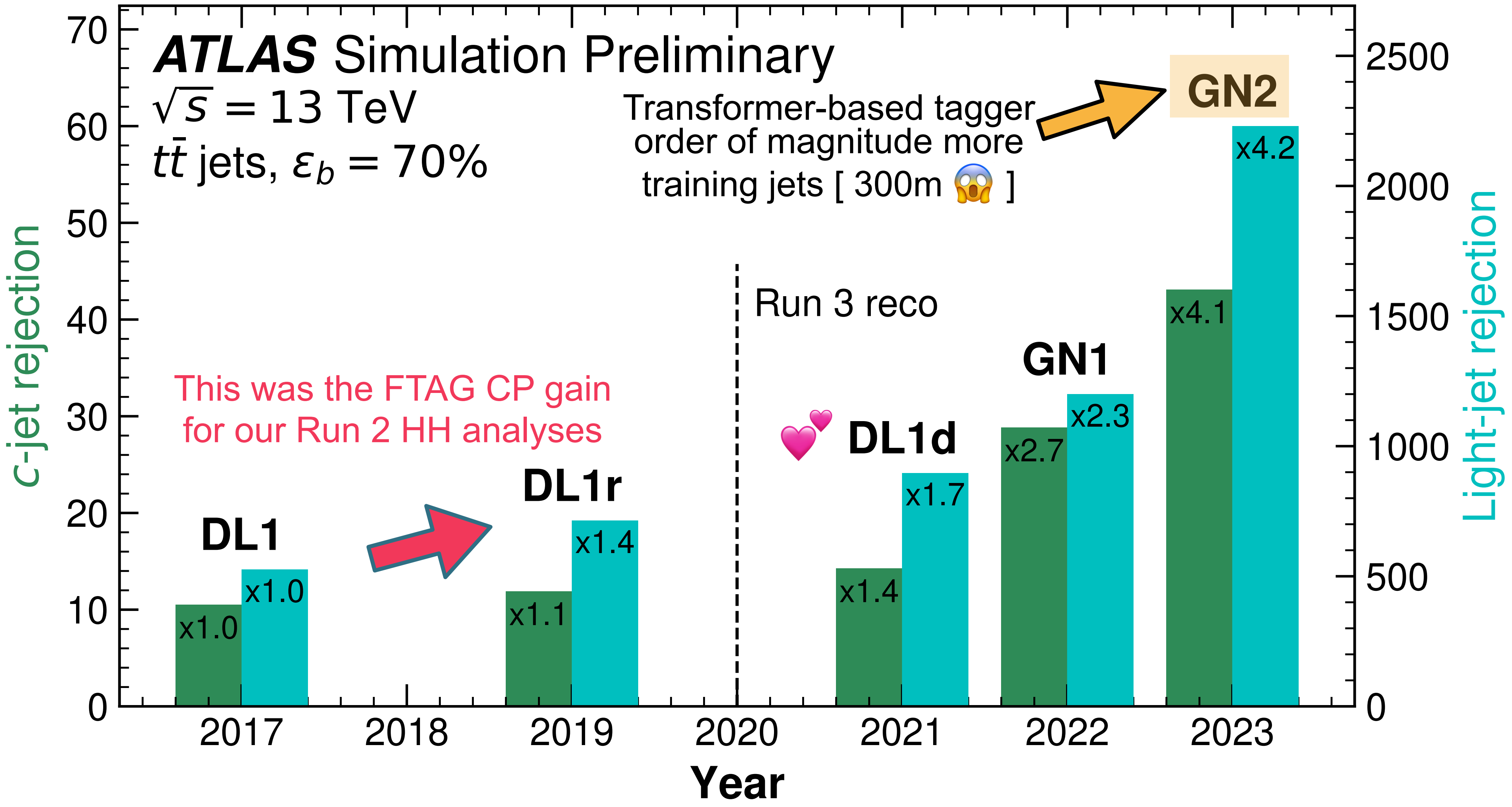
	m_{H1} [GeV]	m_{H2} [GeV]
2016	108.3	50.6
2017	106.1	64.4
2018	105.4	56.4

Table 13.1: The fitted length scales for 4b data.

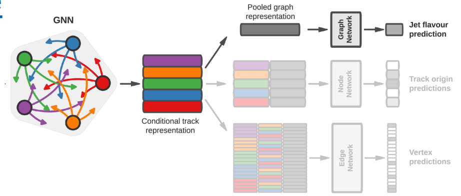
More info in S. Stroud's CERN Data Science seminar

GN2v01

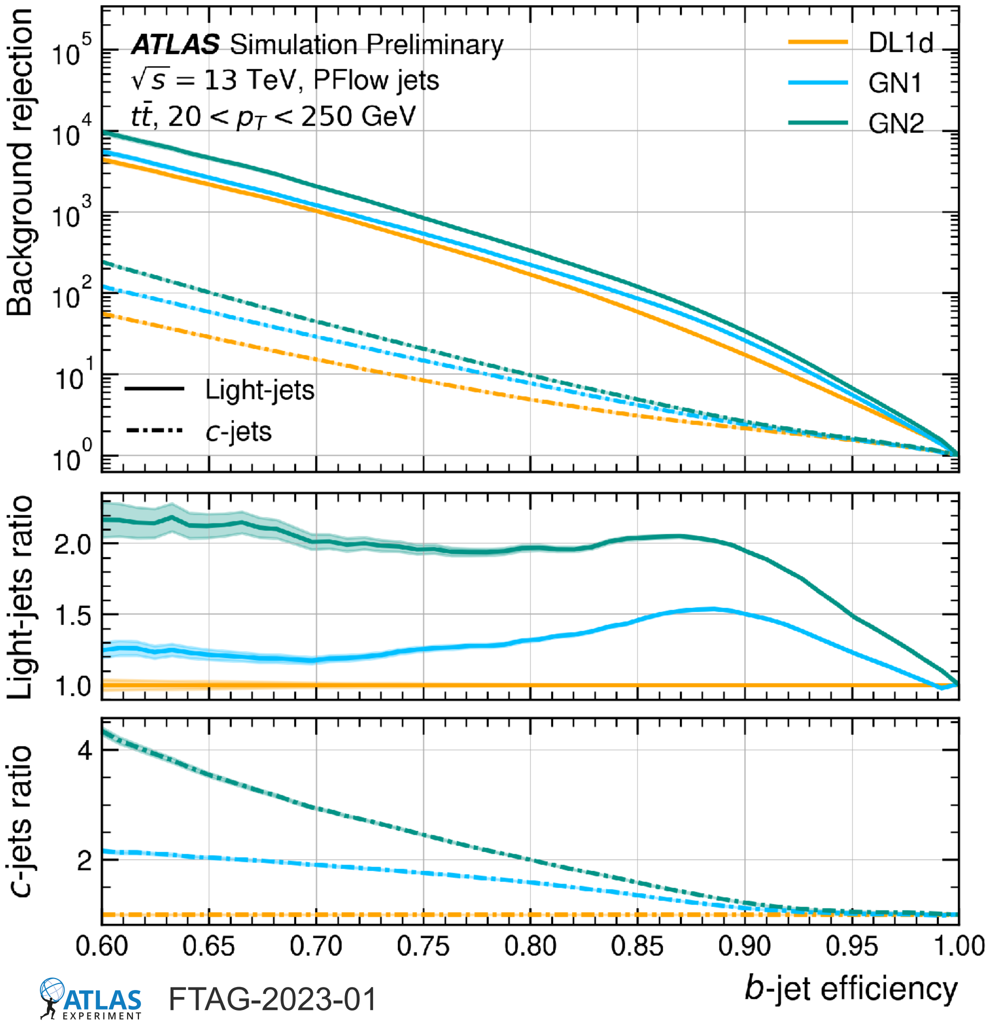
FTAG-2023-01

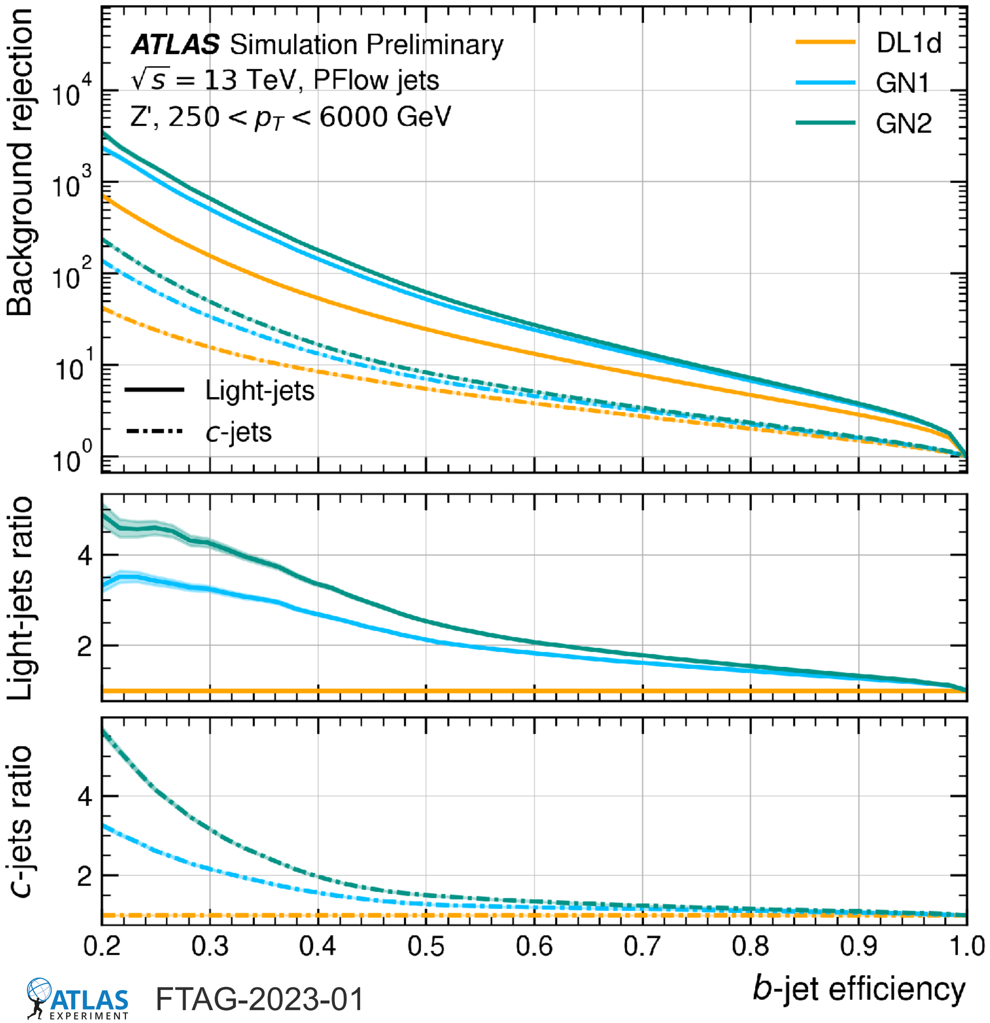


Philipp's Gadow's slide



GN2 b-tagging performance

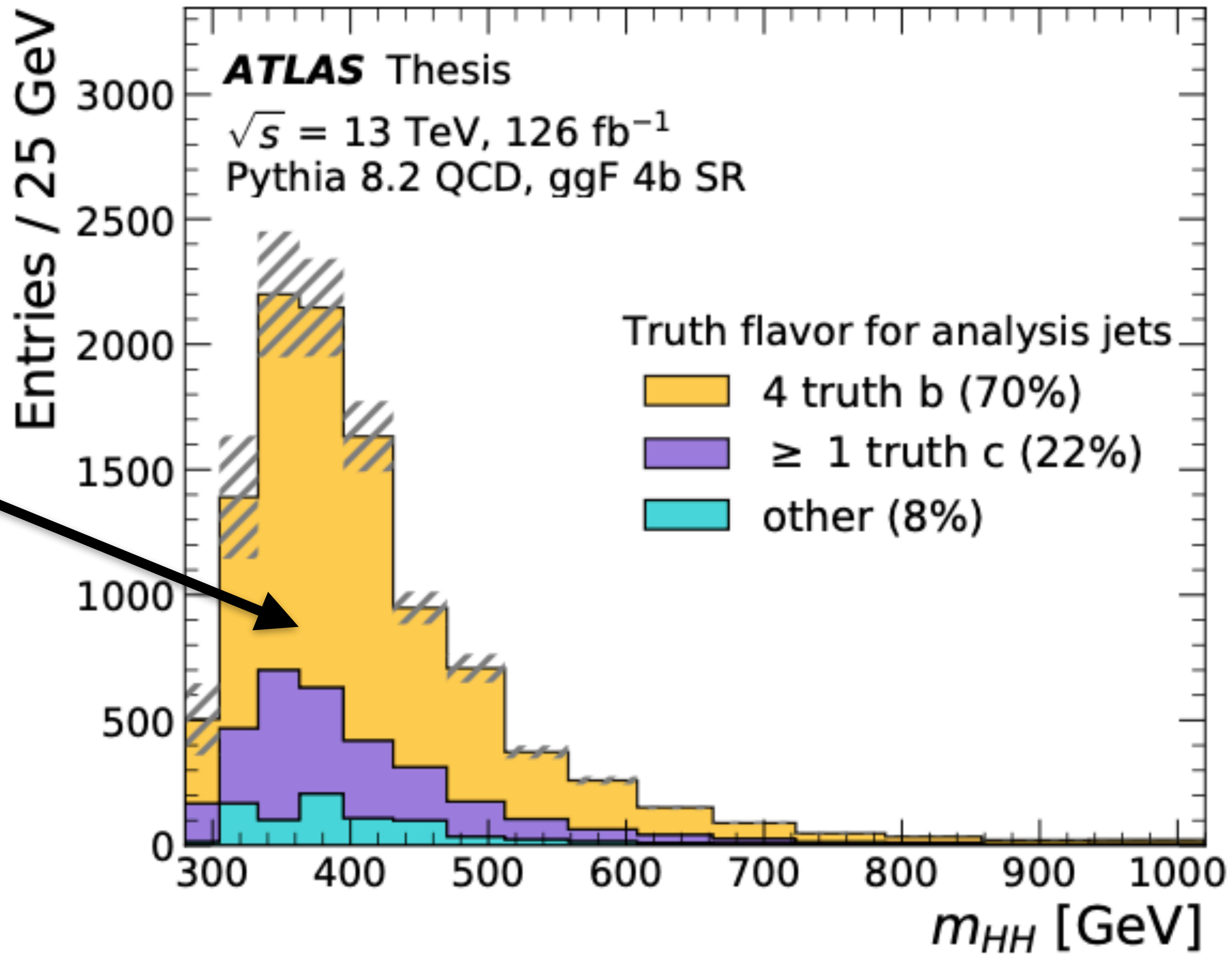




How can we learn more?

70% of our background is *irreducible* at this point (actual 4b events)

Can we leverage event level information to elucidate more signal?

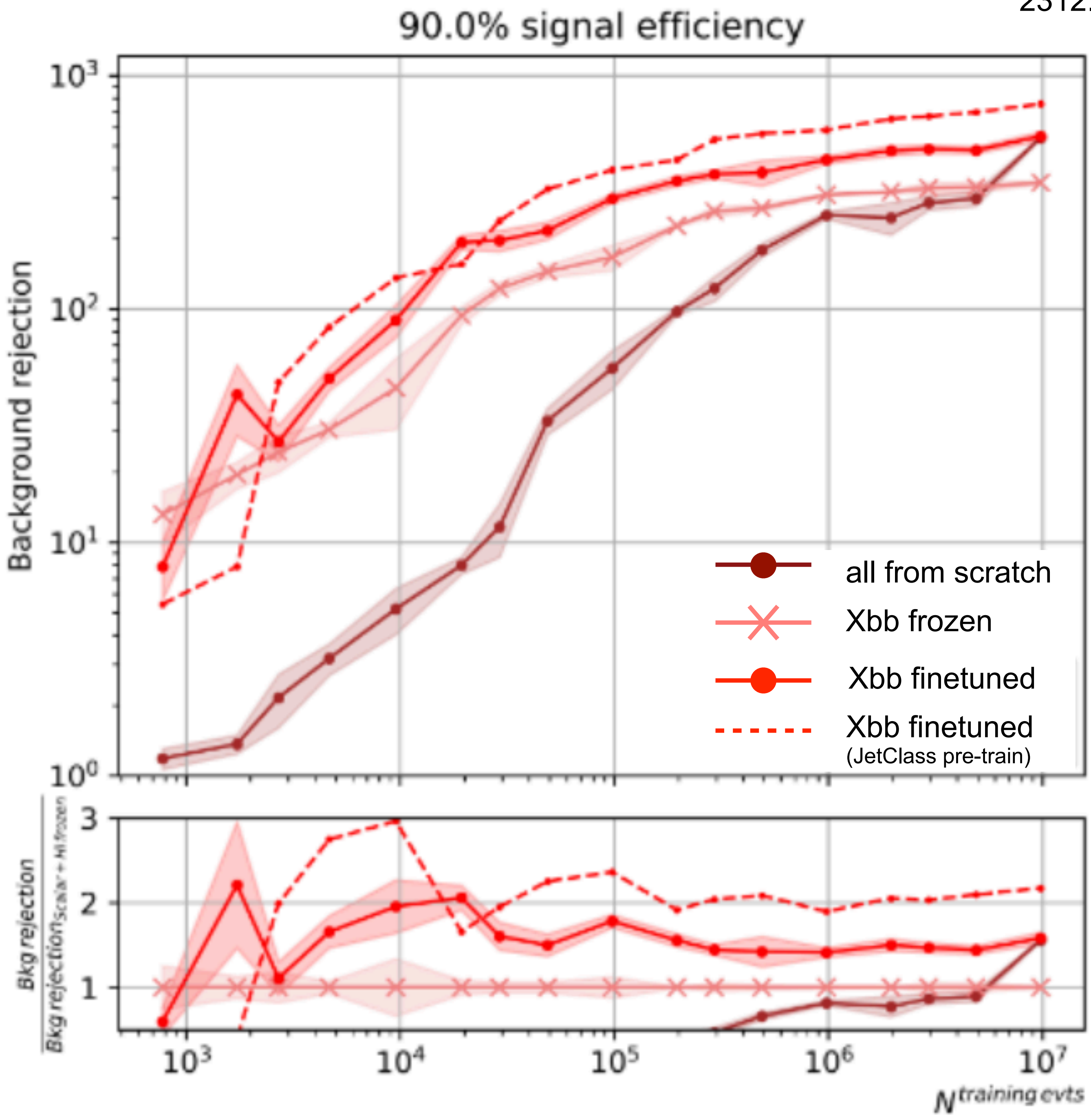


Data efficiency

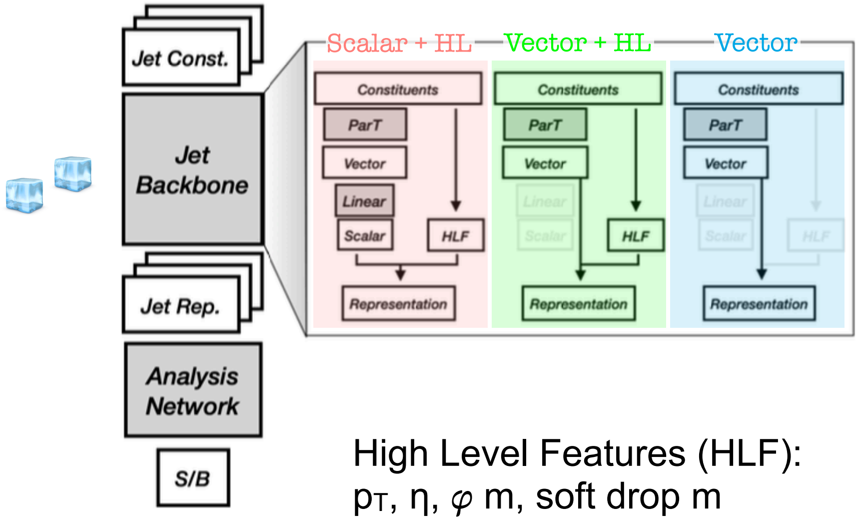
When you train the whole thing from scratch (all on the 10m jets CMS dataset), you eventually get the *same level of performance* as the finetuned model.

The frozen Xbb score is consistently worse than the finetuned model.

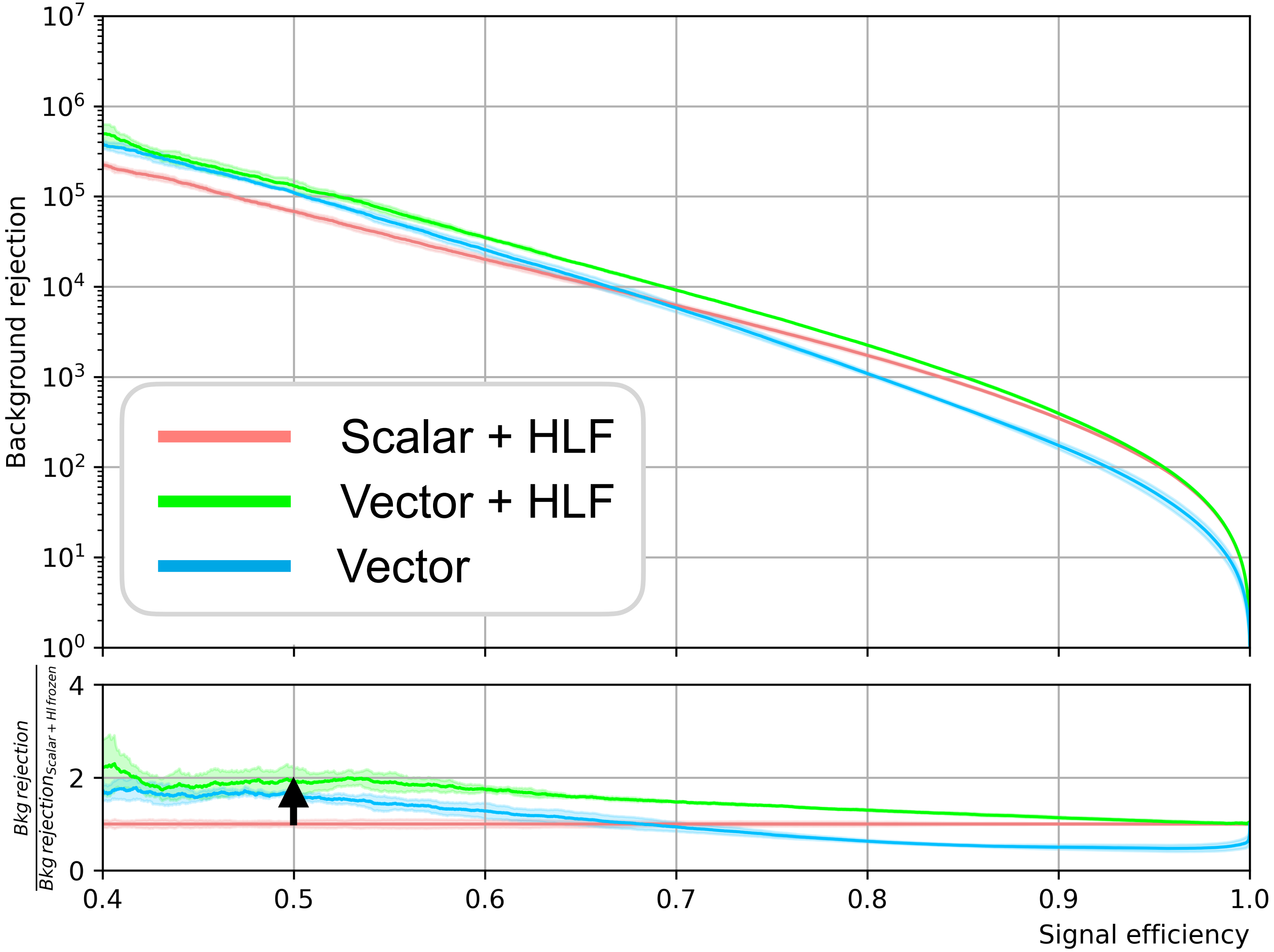
The fine-tuning with pretaining has the best performance once you have more than 10k jets to fine-tune the model on.



Utility of higher dimensional representation

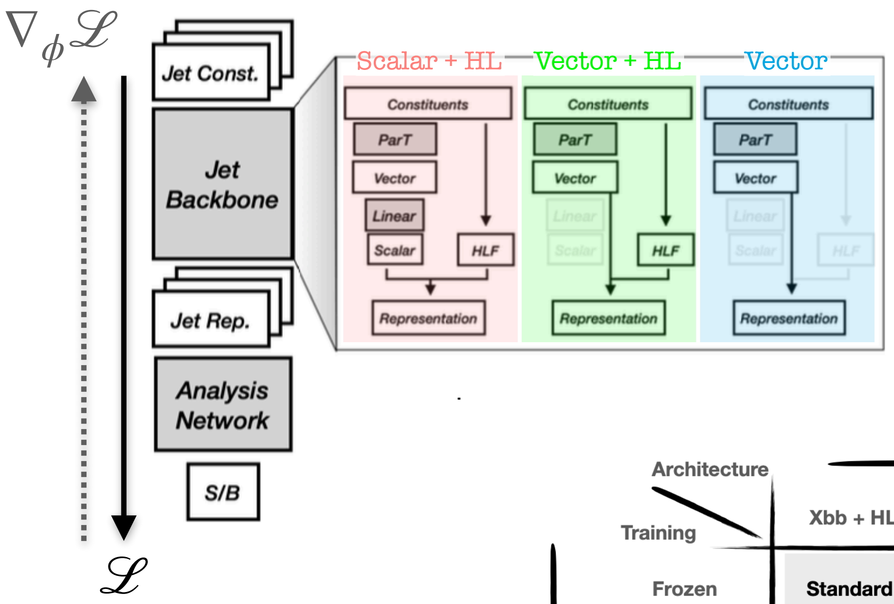


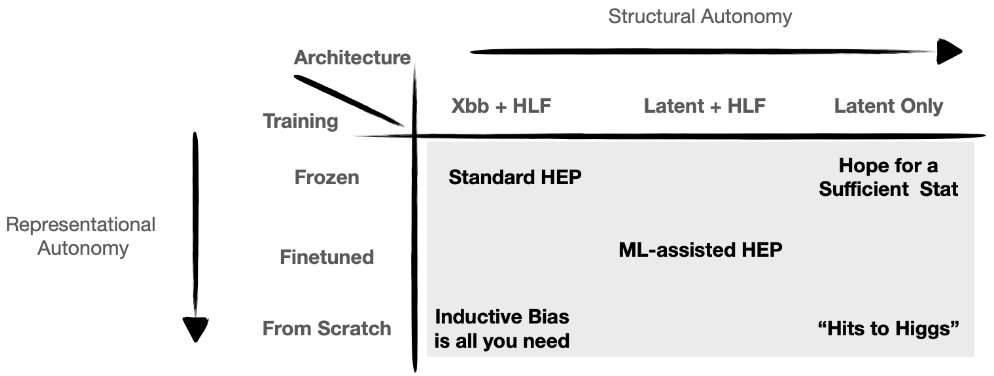
High Level Features (HLF):
p_T, η, φ m, soft drop m

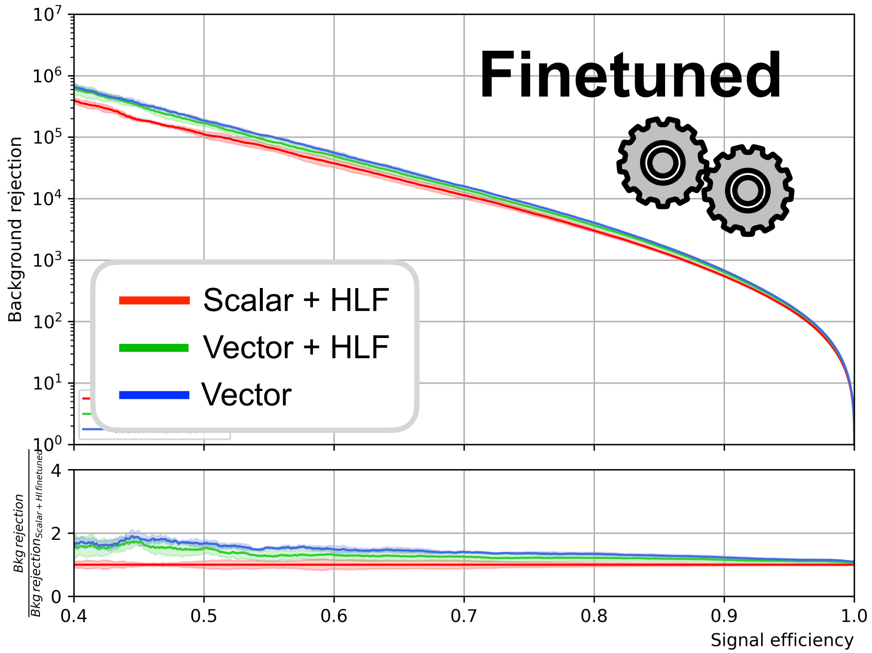


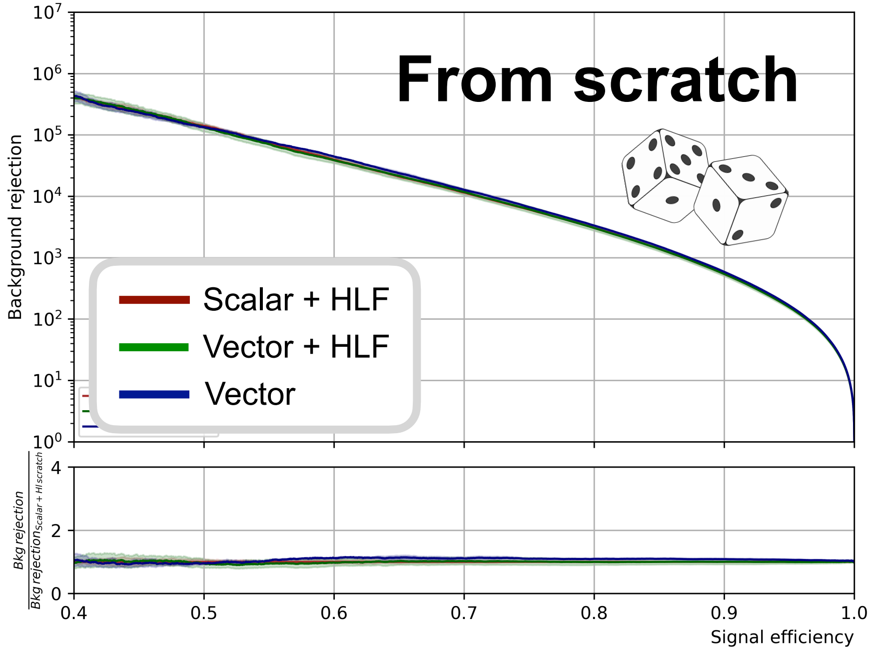
If the jet level representation is frozen, vector performs better than single scalar

And with the gradient flow?









With a finetuned model, some benefit from vector info.
In the from scratch phase, *no benefit from the additional information*.