Search for Higgs boson pair production in the $b\bar{b}\gamma\gamma$ final state with the ATLAS detector using Run 2 and Run 3 *pp* collision data

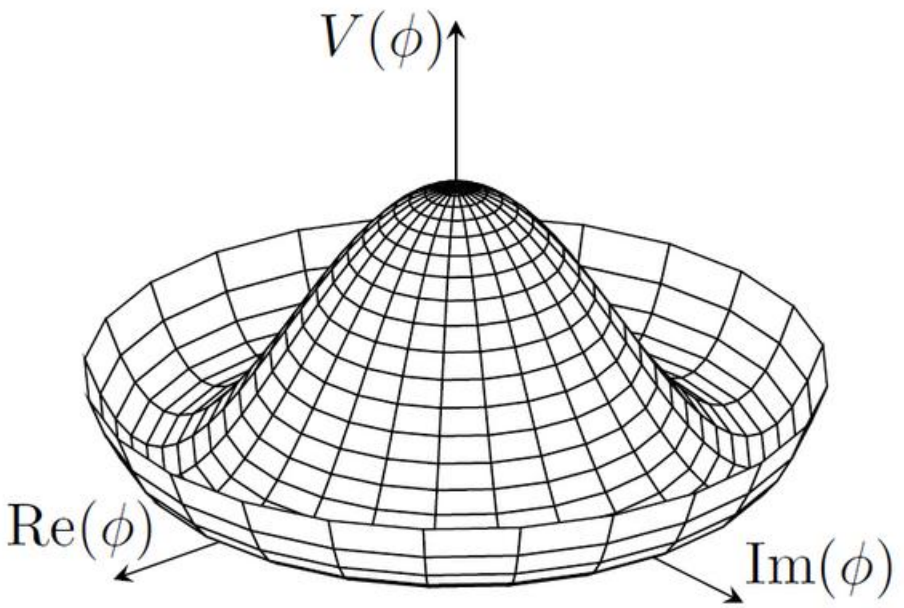
Candidato: Tommaso Cozzi - 28484A
Relatore: Prof. L. Carminati
Correlatori: Dott.ssa E. Mazzeo,
Dott. R. Turra

Università degli Studi di Milano
Facoltà di Scienze e Tecnologie
Corso di laurea magistrale in Fisica
Anno accademico 2024/2025

Higgs mechanism

- The Standard Model (MS) is the theory of the elementary particles and their fundamental interactions, based on the symmetry group $SU(2)_L \times U(1)_Y \times SU(3)_C$.
- It is build exploiting the local gauge invariance principle, which guarantees the renormalizability of a non-Abelian theory ('t Hooft theorem).
- To preserve the gauge invariance, all particles in the SM should be massless.

→ HIGGS MECHANISM



- Idea: the introduction of a scalar field with a self-interacting potential gives rise to mass and new interaction terms.
- In the SM, photon and gluon are massless, while W and Z are massive.

→ Complex neutral scalar $SU(2)_L$ doublet with potential

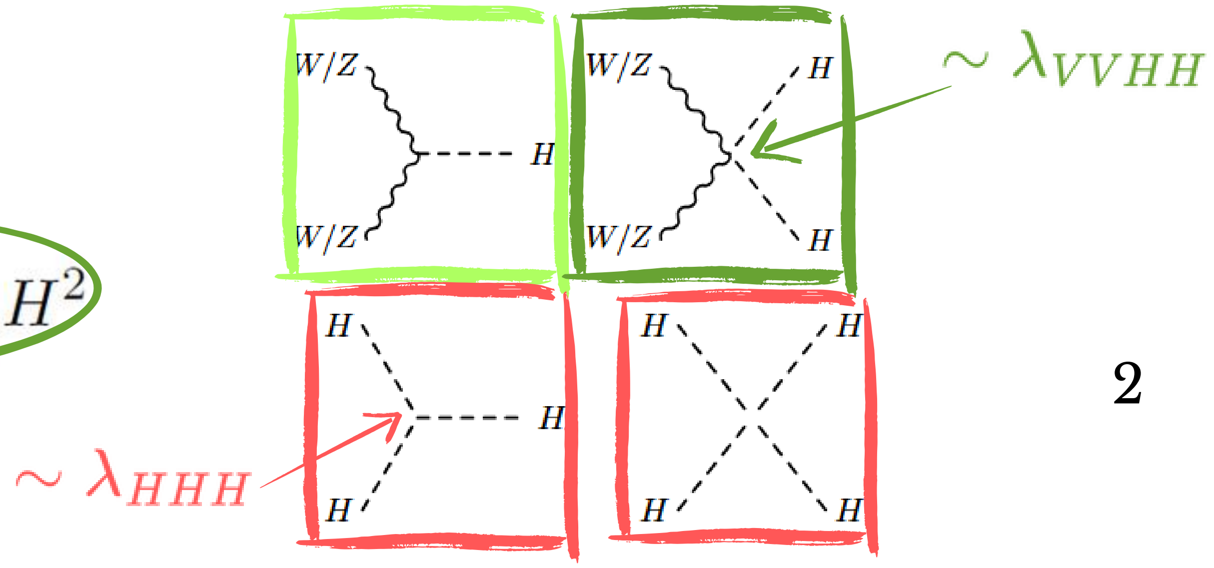
$$V(\Phi) = \mu \left[\Phi^\dagger(x)\Phi(x)\right] + \lambda \left[\Phi^\dagger(x)\Phi(x)\right]^2$$

→ Spontaneous symmetry breaking: choice of ground state, and expansion around that minimum

$$\Phi(x) = \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}$$

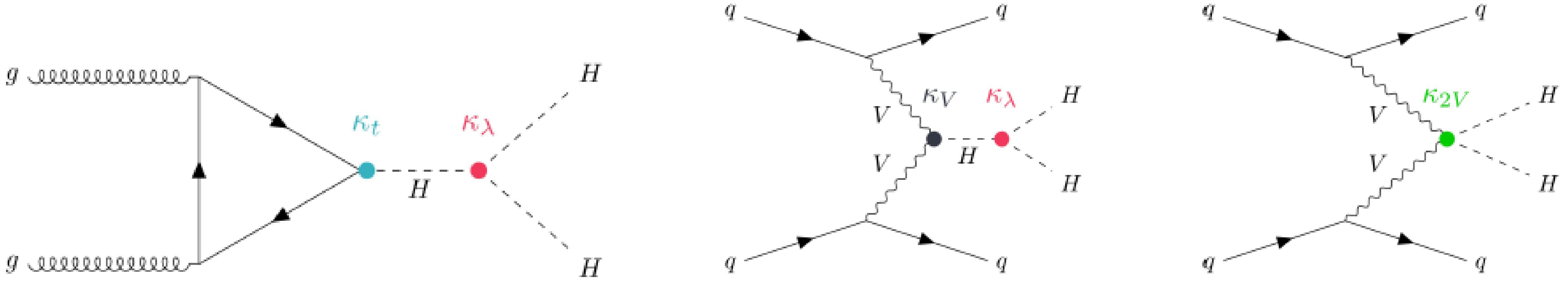
- Mass terms, interactions between H and vector fields, and slef-interaction terms appear.

$$\mathcal{L}_\Phi = +\frac{1}{2}(\partial_\mu H)^2 + \frac{1}{2}\left(-2\mu^2\right)H^2 + \frac{g^2v^2}{4}W^{+\mu}W^-_\mu + \frac{1}{2}\frac{g^2v^2}{4\cos^2\theta_W}Z^\mu Z_\mu$$
$$+ \frac{g^2v}{2}W^{+\mu}W^-_\mu H + \frac{g^2}{4}W^{+\mu}W^-_\mu H^2 + \frac{g^2v}{4\cos^2\theta_W}Z^\mu Z_\mu H + \frac{g^2}{8\cos^2\theta_W}Z^\mu Z_\mu H^2$$
$$+ \lambda v H^3 + \frac{\lambda}{4}H^4$$



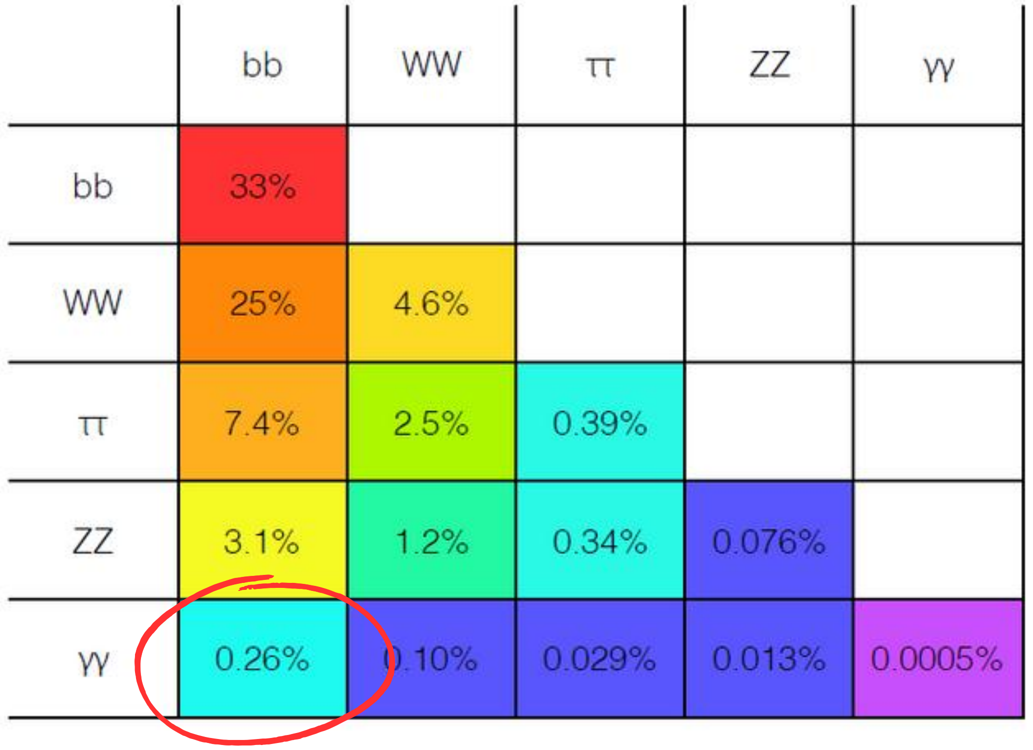
Higgs phenomenology

- The Higgs boson was discovered the July 4th, 2012. Since then, all its key properties (mass, spin, parity, couplings to fermions and vector bosons) have been measured.
- The Higgs self-couplings have not been experimentally observed, yet.
  - → Their observation is of fundamental interest to confirm the shape of the Higgs potential and our understanding of the spontaneous symmetry breaking.
- A direct measurement of the self-coupling can be made through the observation of the Higgs pair production.
- It is a very challenging measurement, due to the extremly low cross-section, ~1000 times smaller than the single-Higgs production.



$$\sigma^{SM}_{ggF}(HH) = 31.02^{+2.2\%}_{-5.0\%} \text{ (Scale) } \pm 3.0\% \text{ (PDF+}\alpha_s) \ \pm 2.6\% \ (m_{top}) \text{ fb}$$

$$\sigma^{SM}_{VBF}(HH) = 1.72^{+0.03\%}_{-0.04\%} \text{ (Scale) } \pm 2.1\% \text{ (PDFs+}\alpha_s) \text{ fb}$$



	bb	WW	ττ	ZZ	γγ
bb	33%				
WW	25%	4.6%			
ττ	7.4%	2.5%	0.39%		
ZZ	3.1%	1.2%	0.34%	0.076%	
γγ	0.26%	0.10%	0.029%	0.013%	0.0005%

- ATLAS and CMS are conducting several searches for the di-Higgs production in different decay channels.
- My thesis focuses on the ATLAS search in the $b\bar{b}\gamma\gamma$ final state.

High $BR(H \rightarrow b\bar{b})$
Excellent di-photon mass resolution

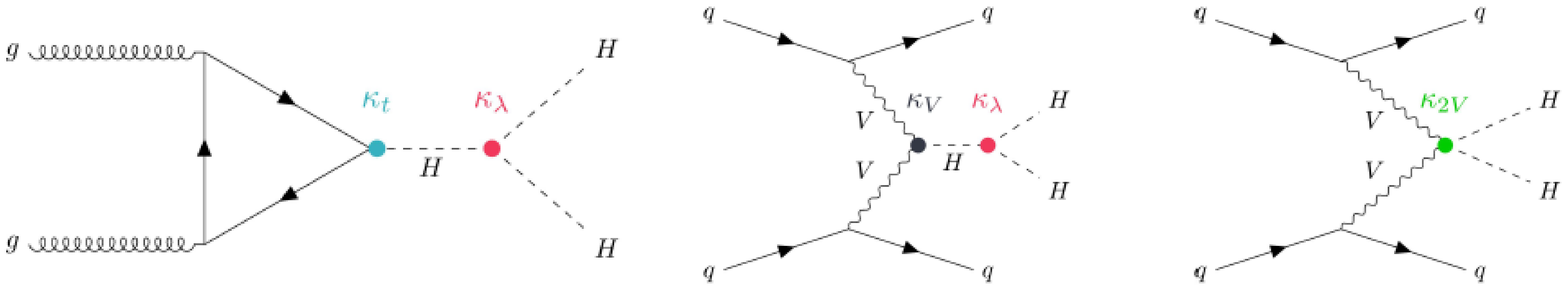
$b\bar{b}\gamma\gamma$ analysis

- Search for HH production in the $b\bar{b}\gamma\gamma$ final state, using pp Run 2 (2015-2018) and partial Run 3 (2022-2024) collision data collected with the ATLAS detector, with an integrated luminosity of 308 fb^{-1}.

→ Observe the di-Higgs production, measuring the signal strength μ_{HH}.
Explore the Higgs potential shape, measuring the coupling modifiers κ_λ and κ_{2V}.

$$\mu_{HH} = \frac{\sigma_{HH}}{\sigma_{HH}^{SM}} \qquad \kappa_\lambda = \frac{\lambda_{HHH}}{\lambda_{HHH}^{SM}} \qquad \kappa_{2V} = \frac{\lambda_{VVHH}}{\lambda_{VVHH}^{SM}}$$

- Signal: ggF and VBF HH production
  Resonant background: single-H processes (ggF, VBF, bbH, ggZH, ttH, qqZH, tHjb, tWH, W$^+$H, W$^-$H)
  Continuum background: yy+jets and yy+bb



- The signal and resonant backgrounds are modeled with MC samples. The continuum background from the data sidebands.

ATLAS

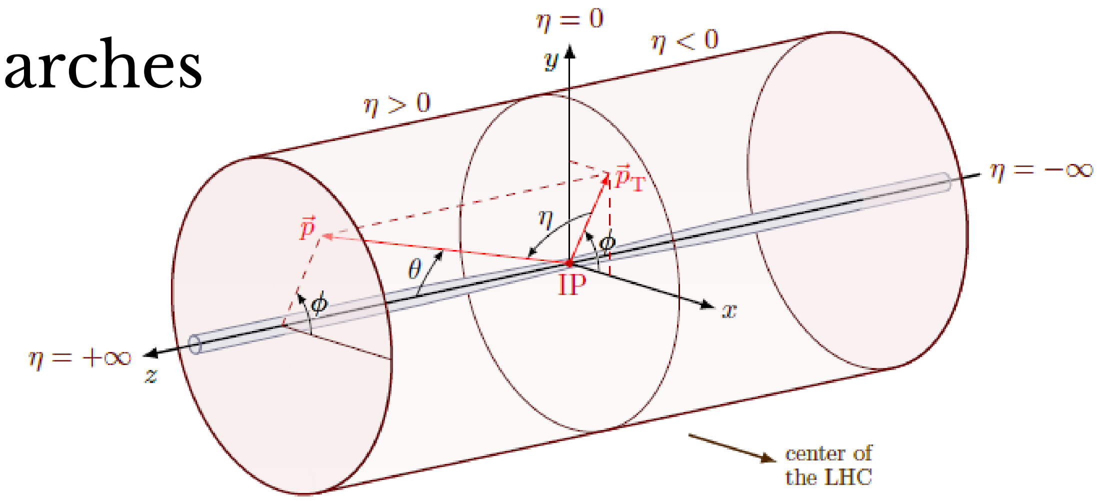
- ALTAS at the LHC is a general-purpose detector.

Precision measurements

SM and BSM searches

- ATLAS coordinate system (z,phi,eta)

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right)$$



Inner detector (ID):

- Tracks and momentum of charged particles
- Primary and secondary verteces
- Particle identification

Calorimeters:

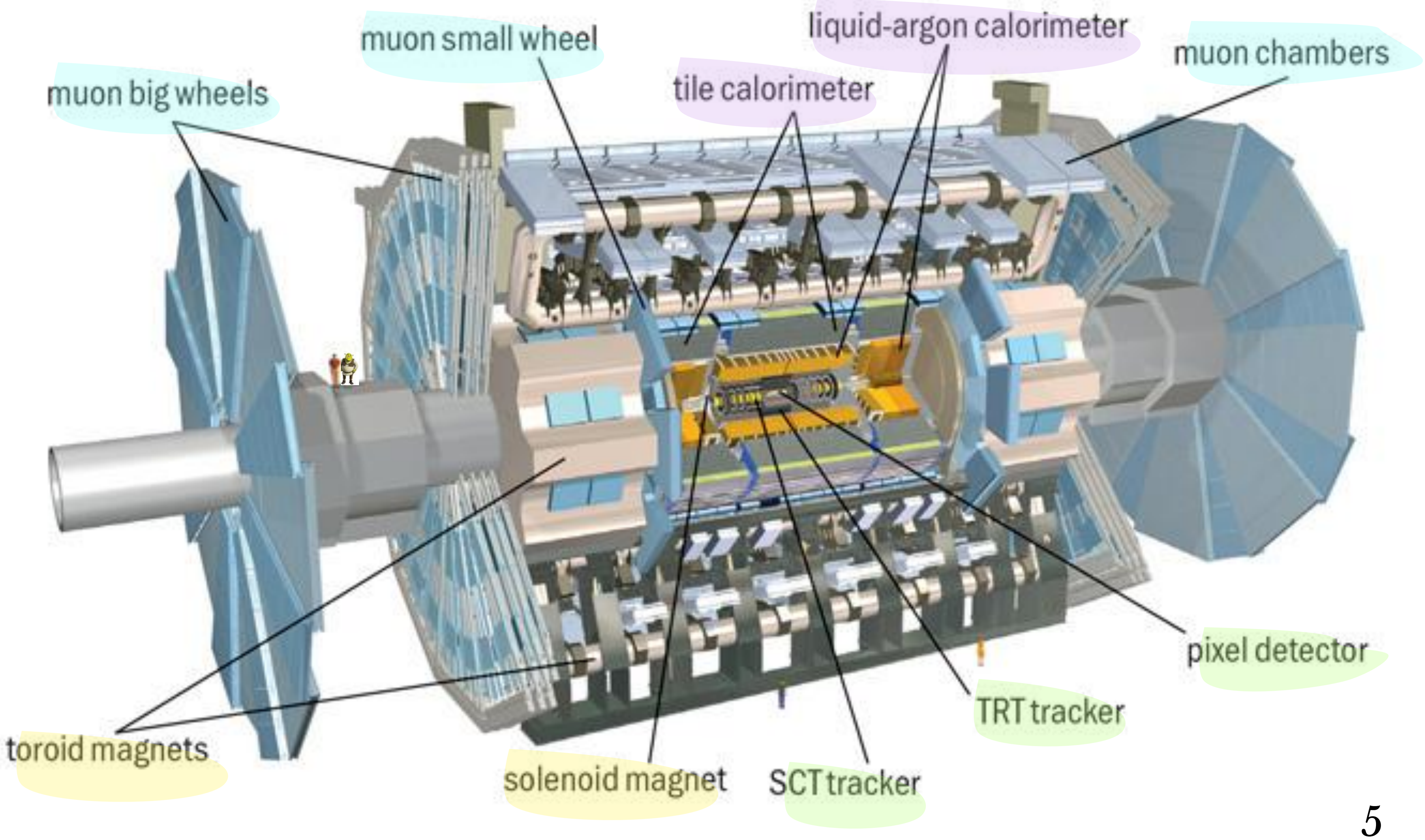
- Sampling calorimeters: electromagnetic (EMCal) and hadronic (HCal)
- Measurement of energy and shape of particle showers
- Particle identification

Muon spectrometer (MS):

- Tracks and momentum of muons
- Particle identification

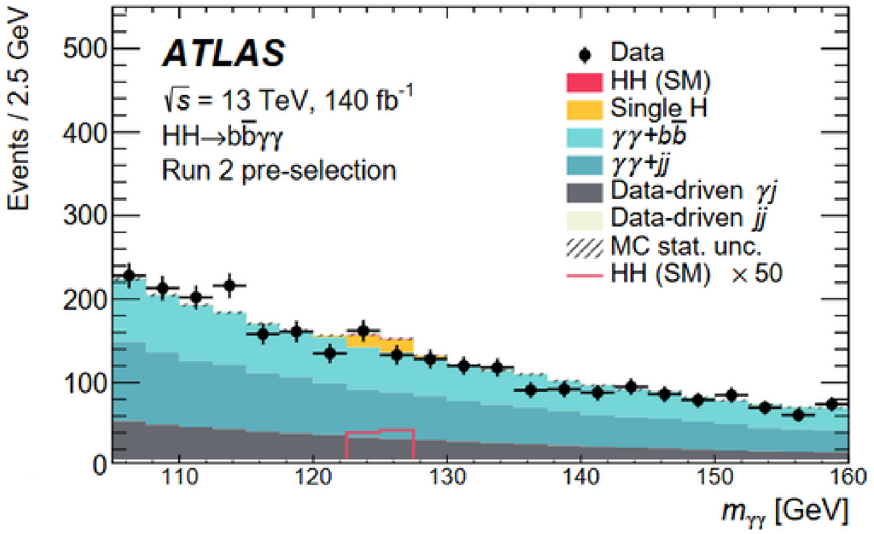
Magnets:

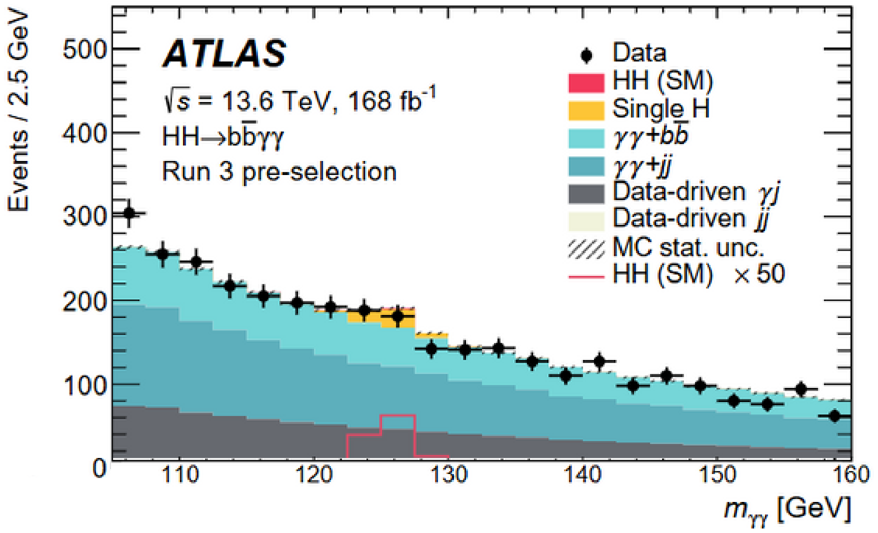
- Trajectory bending
- Solenoidal field in ID, and toroidal in MS



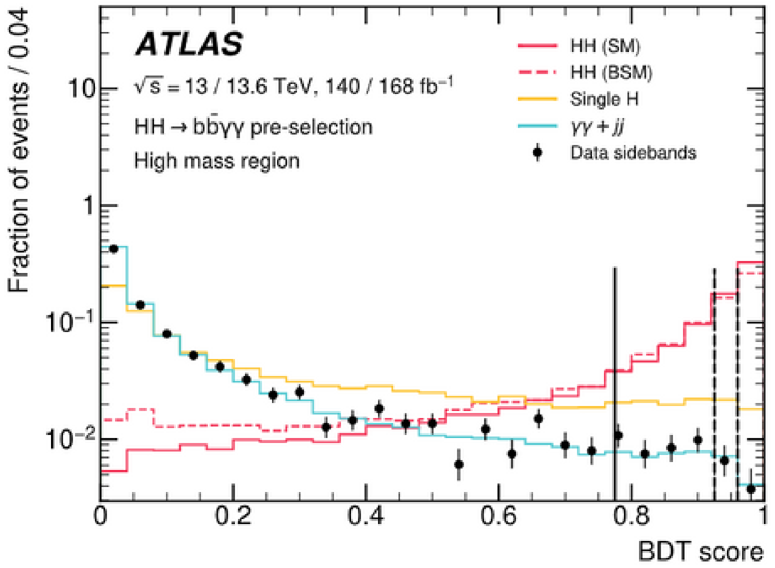
Event selection and categorization

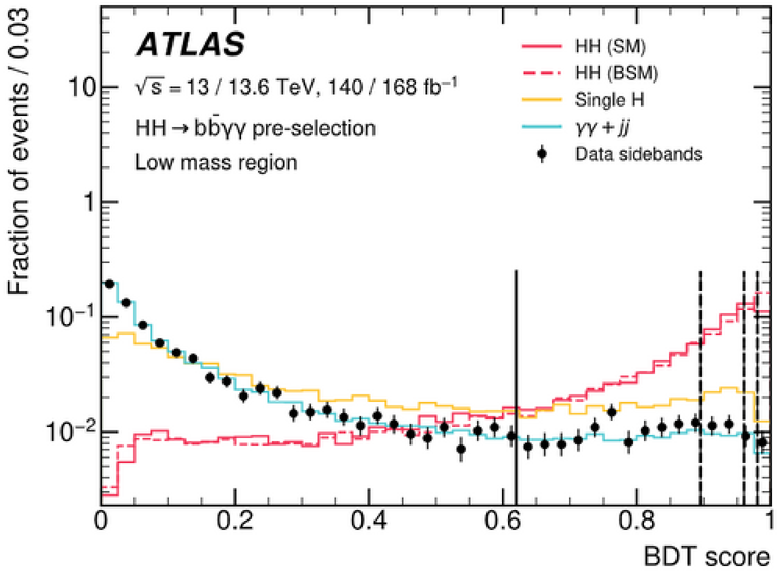
- The selection is optimized to suppress the background while maintaining good signal efficiency. The signal is required to have two well reconstructed photons and at least two b–jet candidates.





- To increase the sensitivity of the analysis, the selected events are divided in categories (7 for Run 2 and 7 for Run 3).
- First, the event are divided in High mass ($m^*_{b\bar{b}\gamma\gamma} > 350$ GeV) and Low mass ($m^*_{b\bar{b}\gamma\gamma} \leq 350$ GeV) regions. The LM region is used to extend sensitivity to a wide range of BSM scenarios.
- In each region a BDT is trained to separate HH signal from the backgrounds. It uses as input kinematic properties of the photon and jet candidates and event-level quantities.
- The BDT score is used to define categories with different signal purities, in order to maximize the overall signal significance.

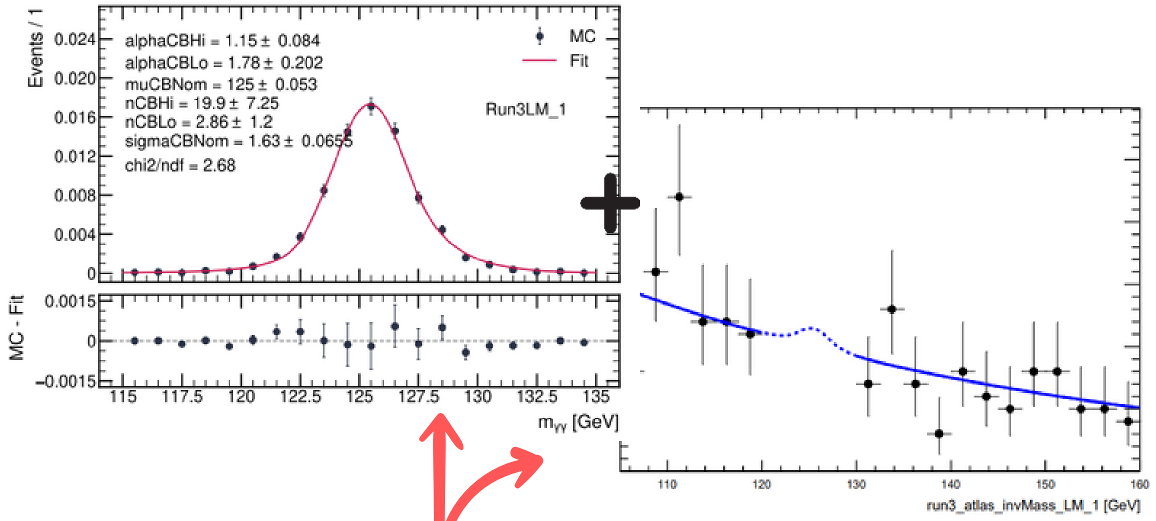




Statistical model

- The results are obtained via an unbinned maximum likelihood fit on the $m_{\gamma\gamma}$ distribution in the $105 < m_{\gamma\gamma} < 160$ GeV, performed simultaneously over all the categories.

The signal and resonant background are modelled with a DSCB, with parameters from MC fits. The continuum background with an exponential, with parameters from a data sidebands fit.



$$\mathcal{L} = \prod_c \left(\mathrm{Pois}\left(n_c | N_c(\boldsymbol{\theta})\right) \cdot \prod_{i=1}^{n_c} f_c\left(m^i_{\gamma\gamma}, \boldsymbol{\theta}\right) \cdot G(\boldsymbol{\theta}) \right)$$

$$N_c(\boldsymbol{\theta}) = \mu_{HH} N_c^{HH}\left(\boldsymbol{\theta}^{HH}_{\mathrm{yield}}, \kappa_\lambda, \kappa_{2V}\right) + N_c^{H}\left(\boldsymbol{\theta}^{H}_{\mathrm{yield}}\right) + N_c^{SS} + N_c^{\gamma\gamma}$$

The sources of systematic uncertainties enter the model through the NP θ. They are constrained by a gaussian PDF.

The observed yields are described by a poissonian with expected yied as central value. Expected yields are the sum of signal, resonant and continuum background, and spurious signal events.

The systematics enter the model multiplying the yield extracted from the MC by a response function $r(\theta)$: $N_c^p(\theta, \kappa_\lambda, \kappa_{2V}) = N_c^p(\kappa_\lambda, \kappa_{2V}) \cdot r(\theta)$.

$$r(\theta) = (1 + \delta \cdot \theta)$$

FTAG uncertainties

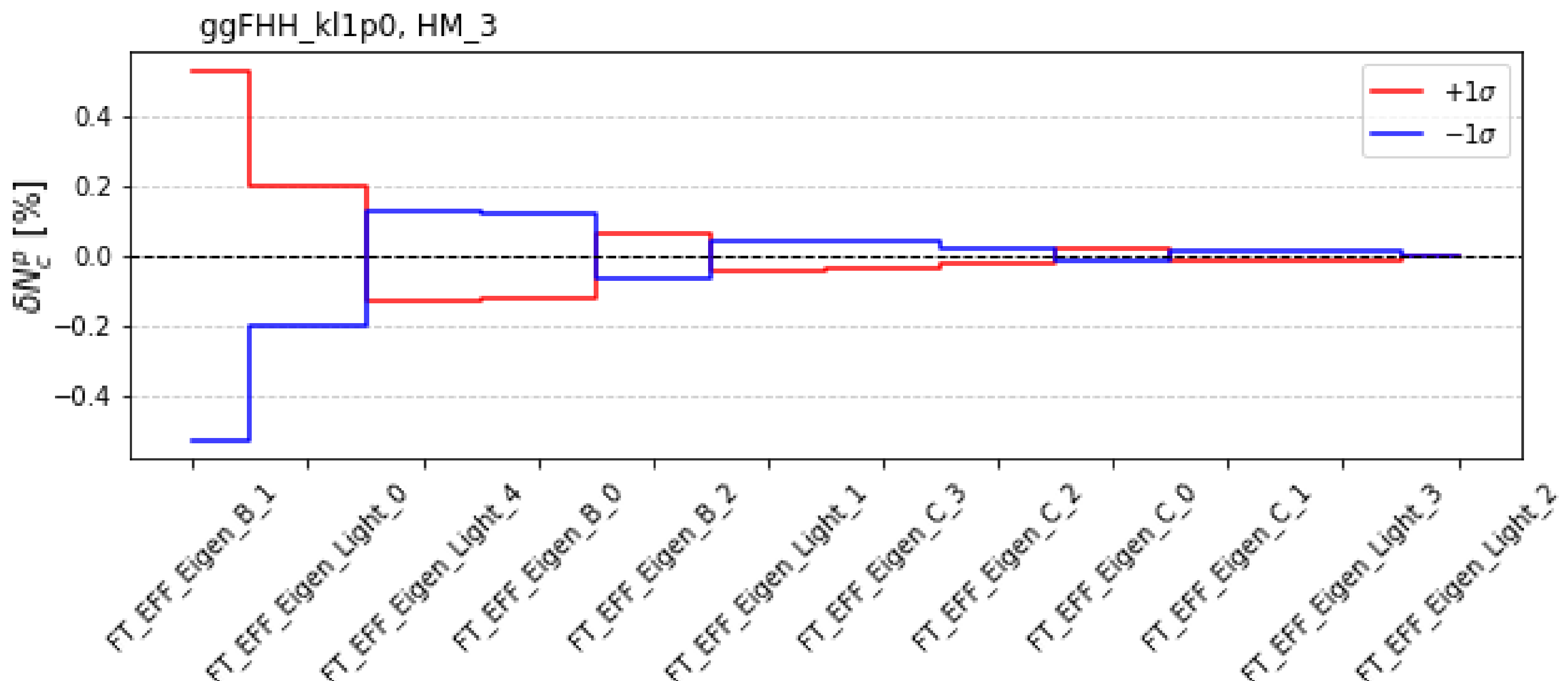
- Study of the FTAG systematic uncertainties for the inclusion of 2024 data into the analysis.
- To evaluate their impact in including 2024 data, the systematics are evaluated on 22+23 samples, and injected in the 22+23+24 dataset.
- The yield variations are evaluated on the signals and on the most relevant resonant backgrounds as:

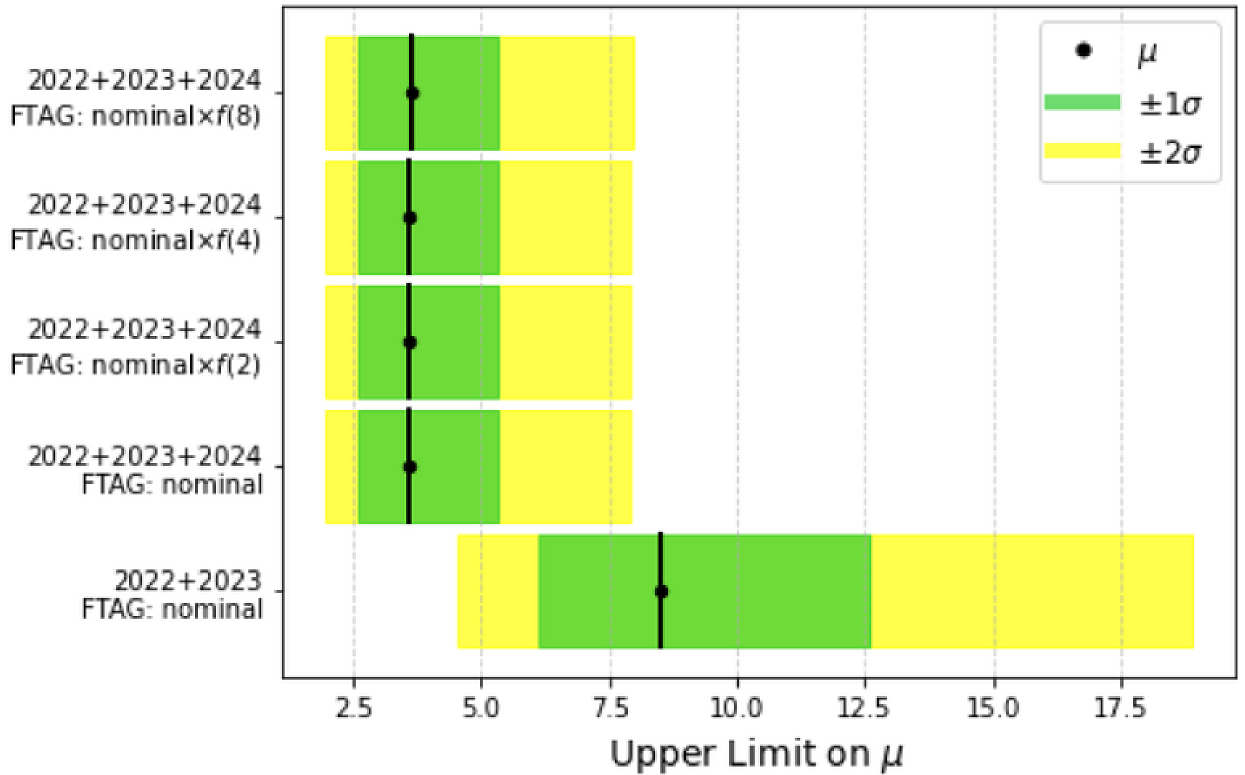
$$\delta N_c^i(\pm 1\sigma) = \frac{N_c^i(\pm 1\sigma)}{N_c^i} - 1$$

- The 22+23 FTAG systematic are enhanced to study if they could be significative on the analysis if 2024 had larger uncertainties.

$$f(N) = \frac{L(2022+2023) + N \cdot L(2024)}{L(2022+2023+2024)}$$

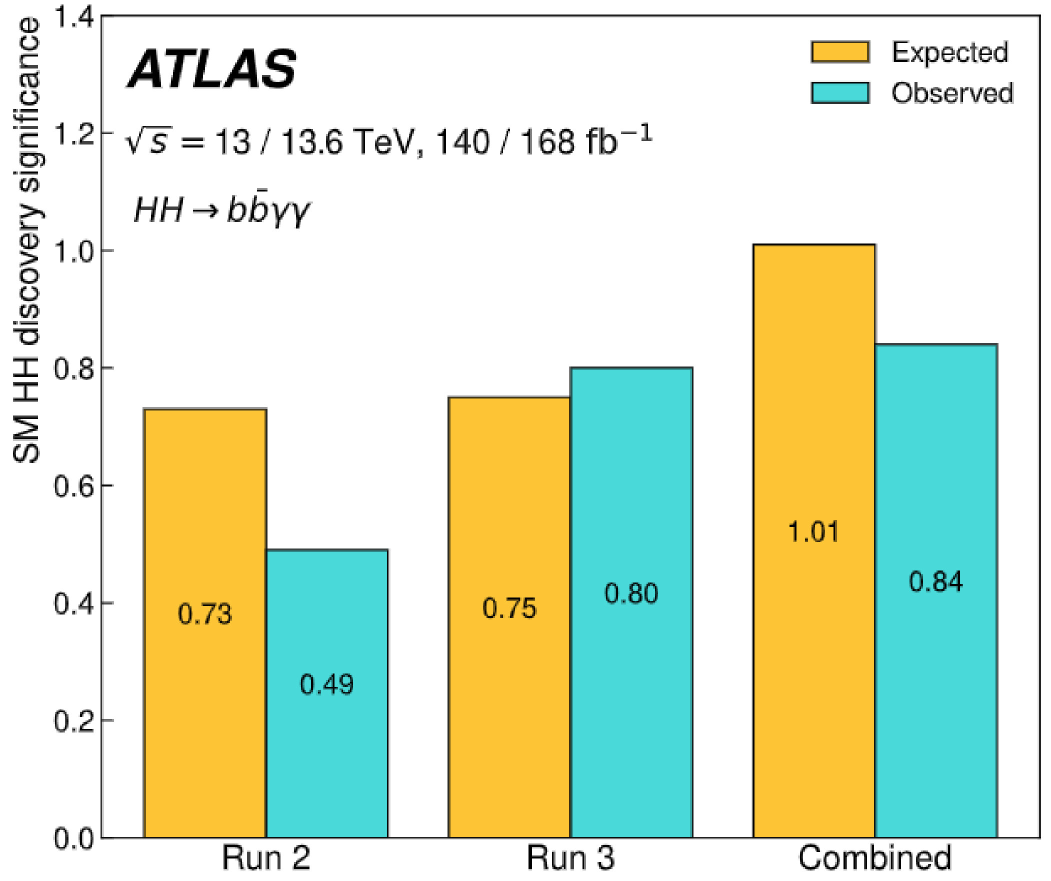
- The signal strength upper limit found in each systematic configurations appears to have a small relative difference with respect to the nominal case (<0.5%).
- The flavor tagging performance group measured for 2024 data the same uncertainties as 22+23, and the 2024 data have been used in the analysis.

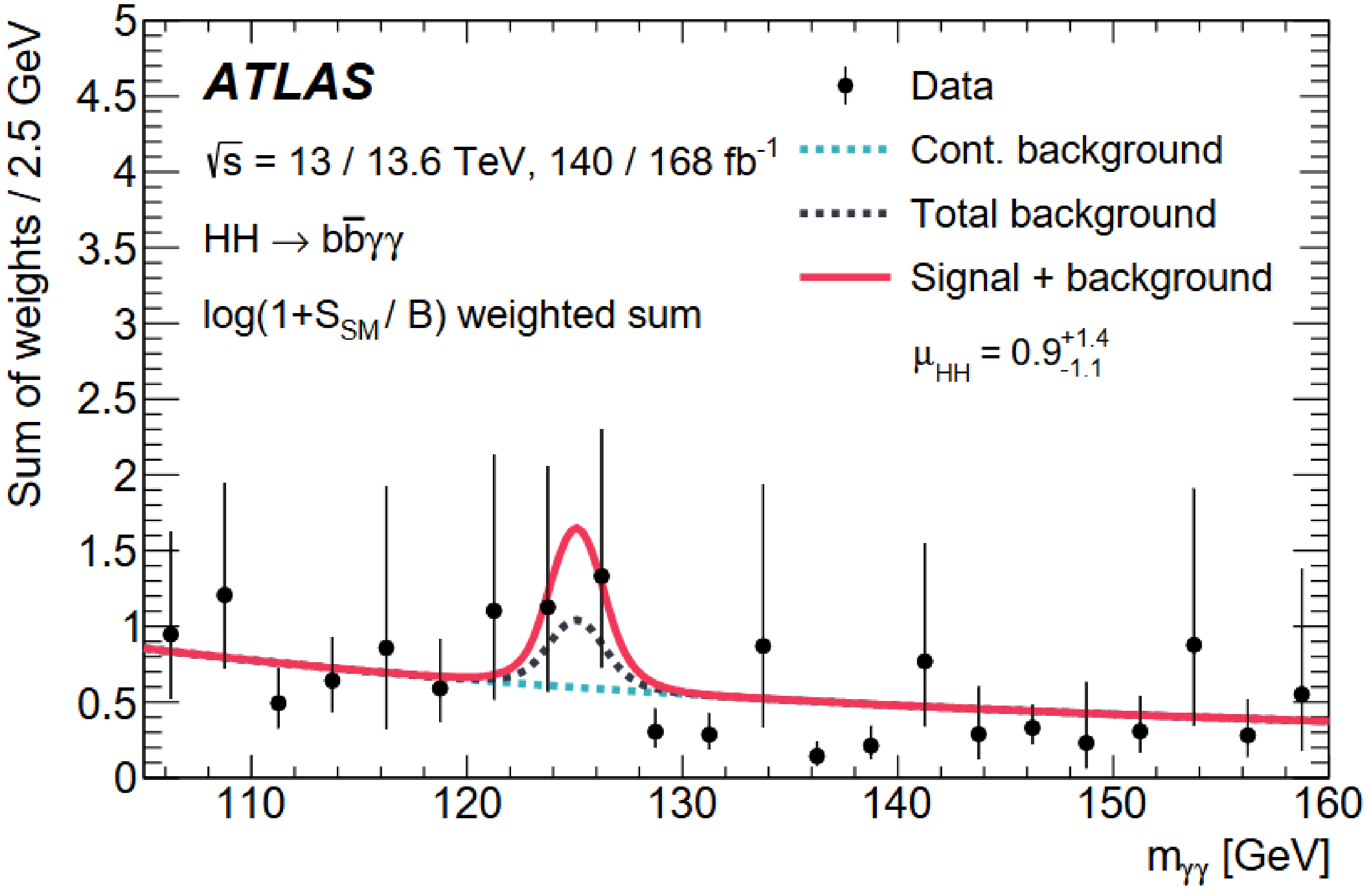


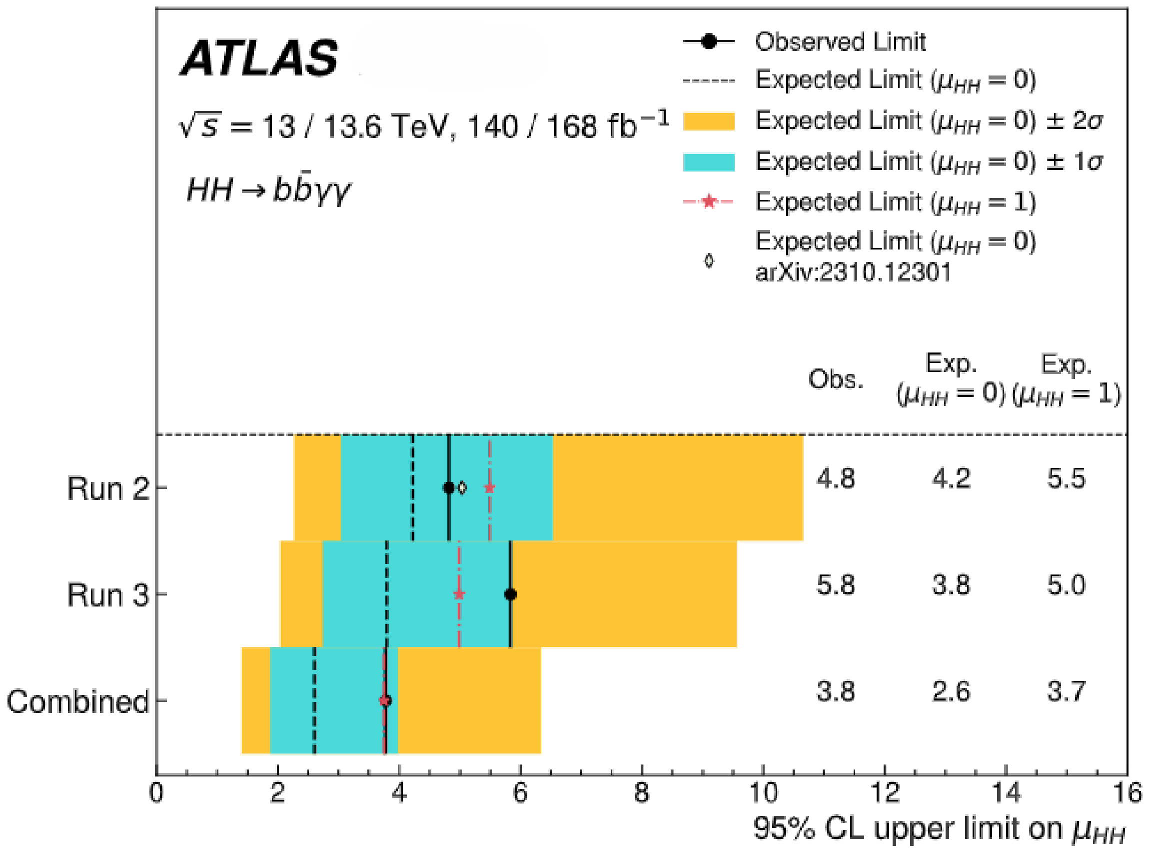


Results

- The search results in significance of 0.84, and a signal strength, compatible with the SM of $\mu_{HH} = 0.9^{+1.3}_{-1.0}(stat.)^{+0.6}_{-0.5}(syst.)$, and a corresponding upper limit of 3.8.
- These results improve the limit by almost a factor two with respect to the previous Run 2 only analysis.

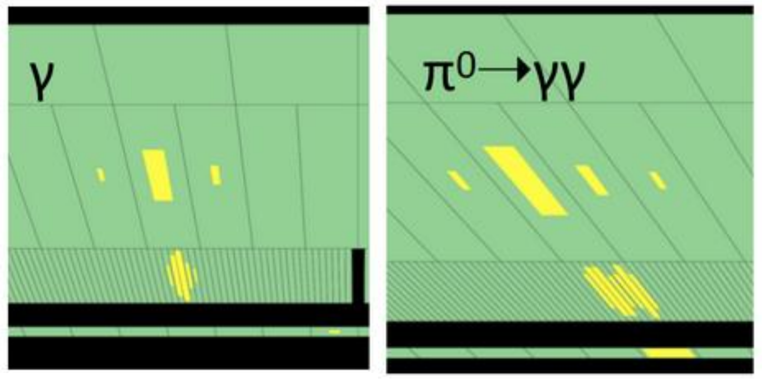






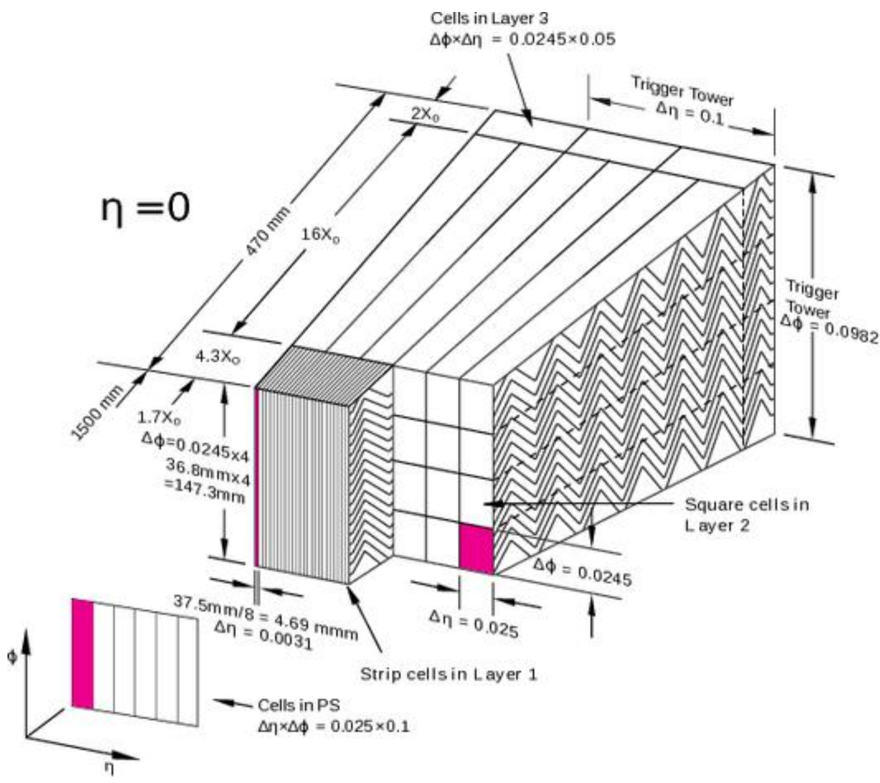
Photon identification

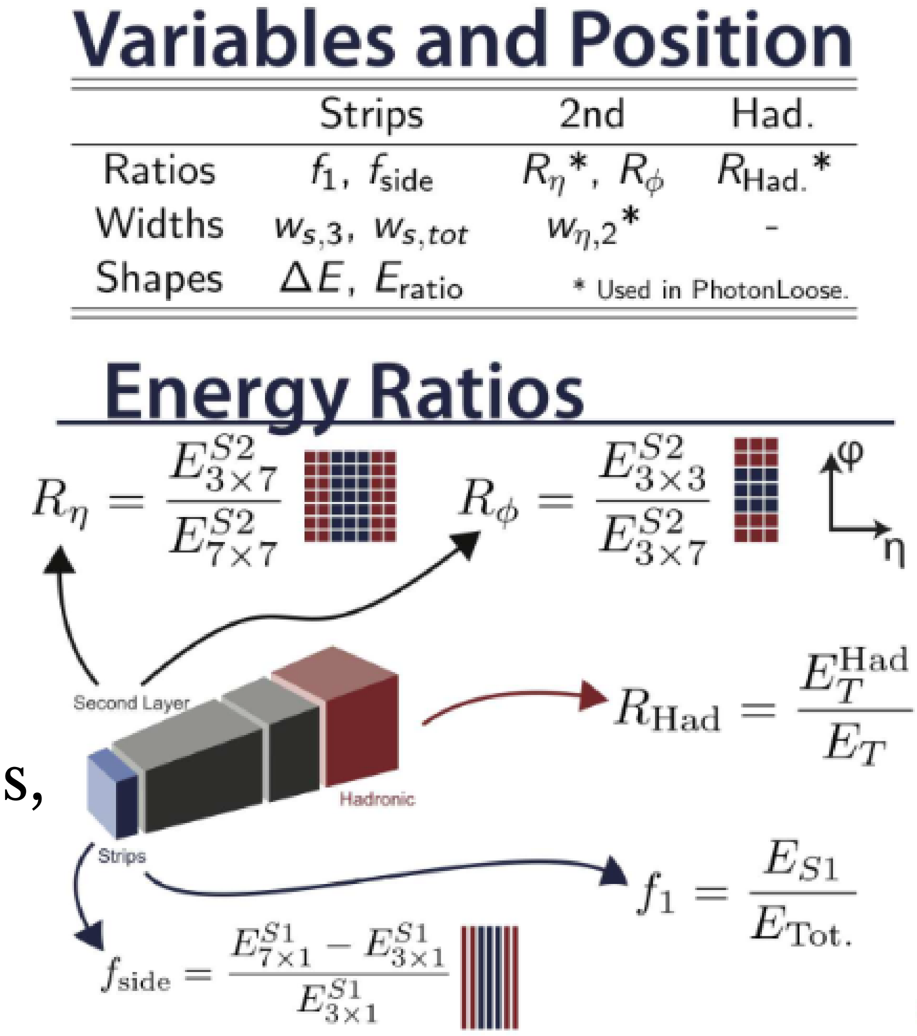
- Given the importance of photon reconstruction in this search, I investigated the possibility to increase the efficiency of photon identification (PID) to improve the results of the analysis.
- Photons are detected in the calorimeters as energy clusters deposited by the EM showers.
- The ATLAS calorimeter is segmented both longitudinally and transversally. The combination of informations from the cells produces the *shower shape* (SS) variables.
- SS are used in PID as discriminant variables.
- PID criteria are optimized to select prompt photons, and to suppress jet or electron faking photons background.



- Independent rectangular selections are applied on SS (three working points: *Loose, Medium, Tight*).
- The Tight working point (WP) is optimized in $|\eta|$ and E_T bins, and in conversion status.

Photon can convert into e^+e^- pairs in the ID





Variables and Position

	Strips	2nd	Had.
Ratios	f_1, f_{side}	R_η*, R_ϕ	$R_{\text{Had.}}$*
Widths	$w_{s,3}$, $w_{s,tot}$	$w_{\eta,2}$*	-
Shapes	ΔE, E_{ratio}		

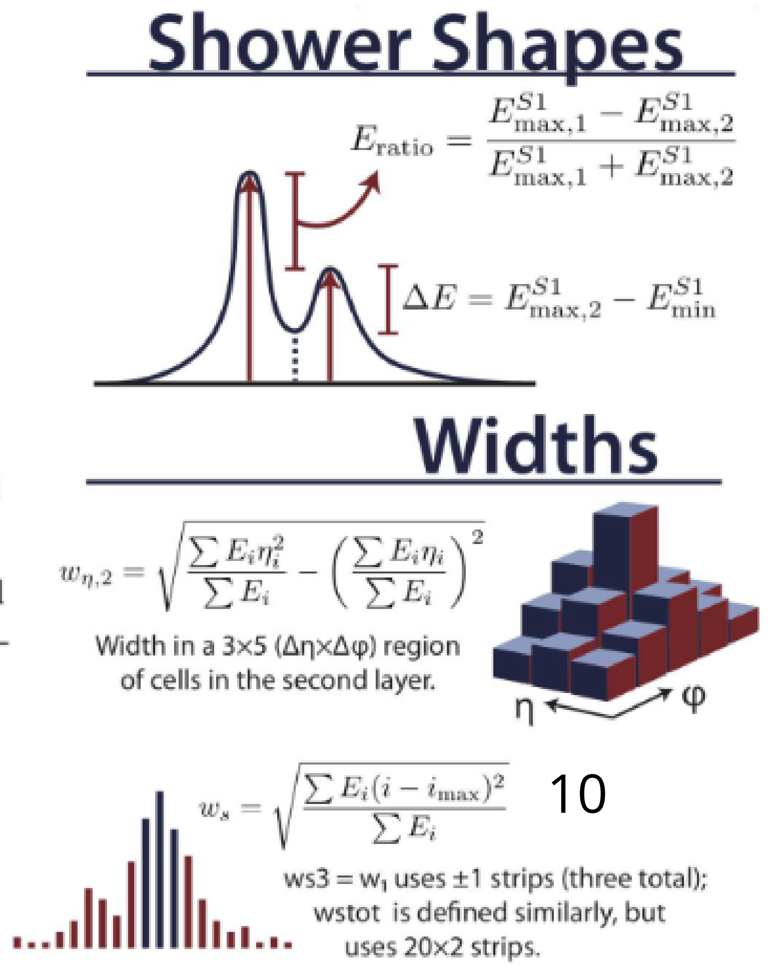
* Used in PhotonLoose.

Energy Ratios

$$R_\eta = \frac{E^{S2}_{3\times7}}{E^{S2}_{7\times7}} \qquad R_\phi = \frac{E^{S2}_{3\times3}}{E^{S2}_{3\times7}}$$

$$R_{\text{Had}} = \frac{E_T^{\text{Had}}}{E_T} \qquad f_1 = \frac{E_{S1}}{E_{\text{Tot.}}}$$

$$f_{\text{side}} = \frac{E^{S1}_{7\times1} - E^{S1}_{3\times1}}{E^{S1}_{3\times1}}$$



Shower Shapes

$$E_{\text{ratio}} = \frac{E^{S1}_{\max,1} - E^{S1}_{\max,2}}{E^{S1}_{\max,1} + E^{S1}_{\max,2}}$$

$$\Delta E = E^{S1}_{\max,2} - E^{S1}_{\min}$$

Widths

$$w_{\eta,2} = \sqrt{\frac{\sum E_i \eta_i^2}{\sum E_i} - \left(\frac{\sum E_i \eta_i}{\sum E_i}\right)^2}$$

Width in a 3×5 (Δη×Δφ) region of cells in the second layer.

$$w_s = \sqrt{\frac{\sum E_i (i - i_{\max})^2}{\sum E_i}}$$

ws3 = w_1 uses ±1 strips (three total); wstot is defined similarly, but uses 20×2 strips.

BDT PID

Idea: tune a new PID WP to replace Tight WP in the analysis.

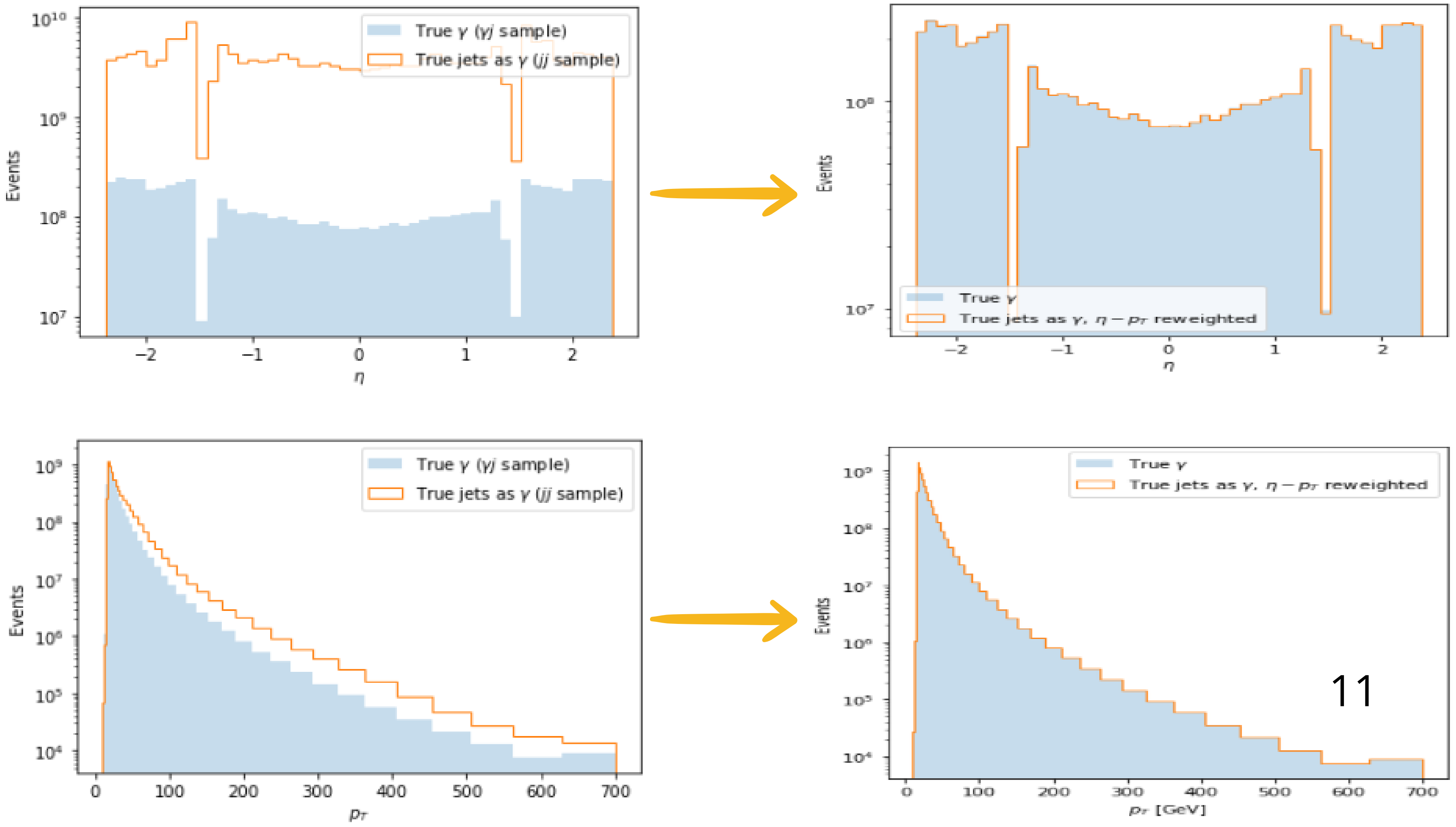
- BDTs to separate photons and jets.

Jets and photons from hadron decay from $jet + jet$ samples .

Prompt-photons (no photons from hadron decay) from $\gamma + jet$ sample.

- Selection applied for the training: truth + Loose ID
- Input variables: kinematic + SS
- Kinematic reweighting: the BDT must not learn how distinguish signal and background from their kinematic features.

Creation of wieghts to apply to the bkg sample such that the kinematic distributions of sig and bkg coincide

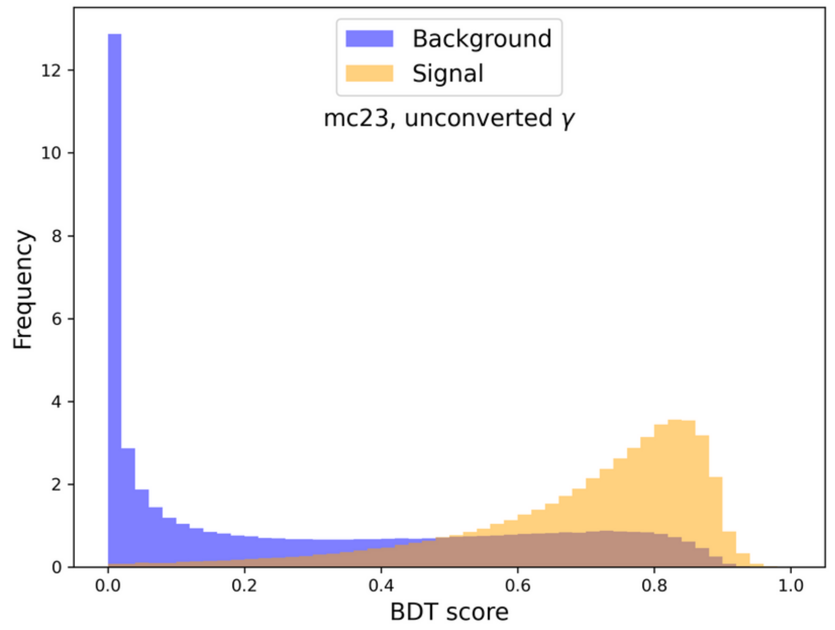


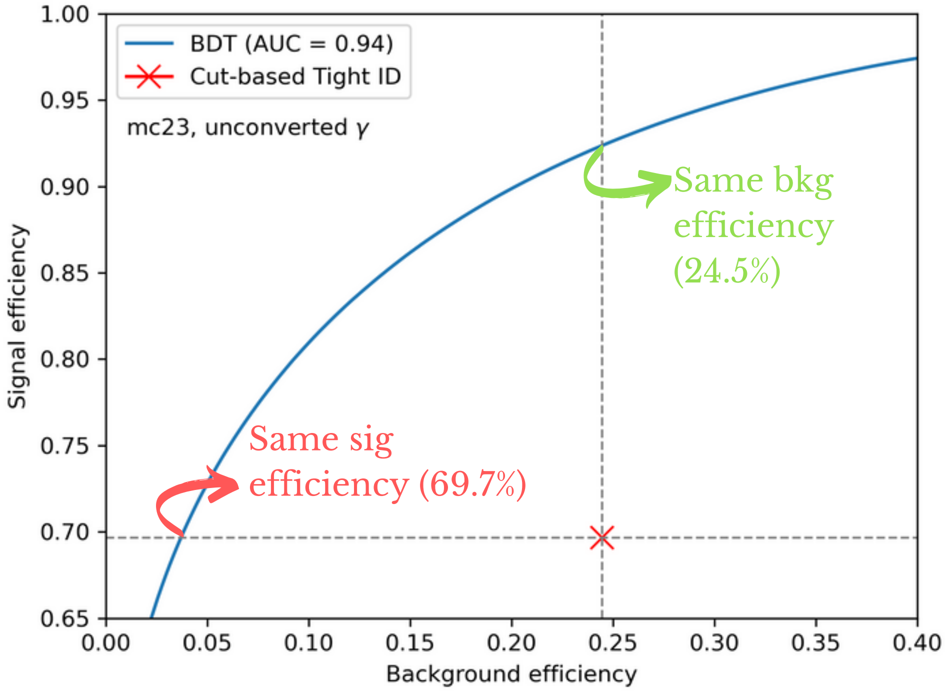
- BDTs inclusive over p_T and $|\eta|$.

- Dedicated BDT for converted and unconverted, and for data-taking period.

Converted γ Run 2
Unconverted γ Run 2
Converted γ Run 3
Unconverted γ Run 3

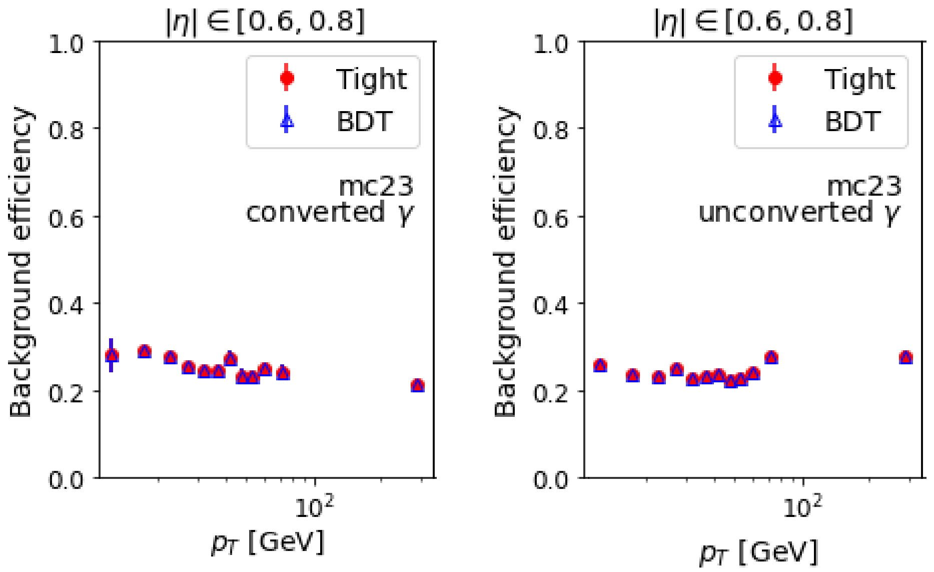
BDT PID





Same bkg efficiency (24.5%)

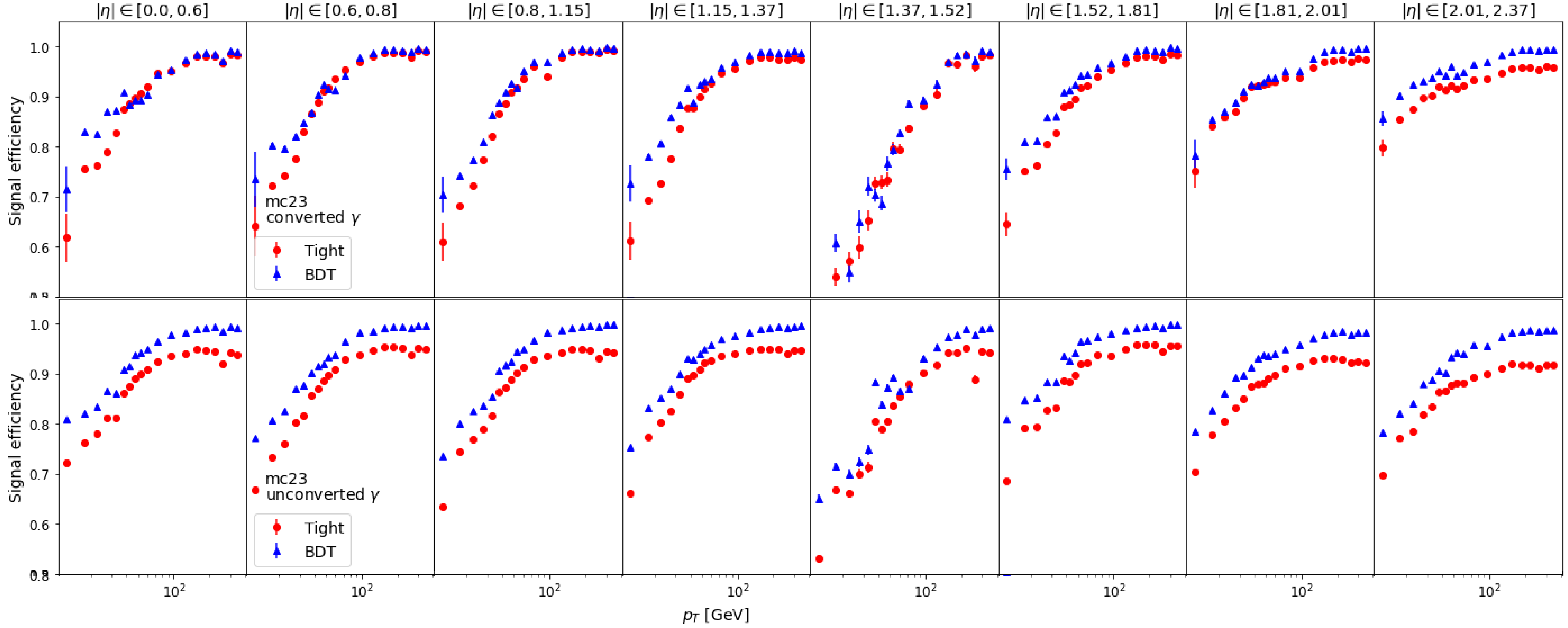
Same sig efficiency (69.7%)



- The BDT score shows a good separation between signal and background.
- Two possible choices for the WP
  - Improve sig efficiency maintaining the same bkg efficiency as Tight
  - Improve bkg rejection maintaining the same sig efficiency as Tight

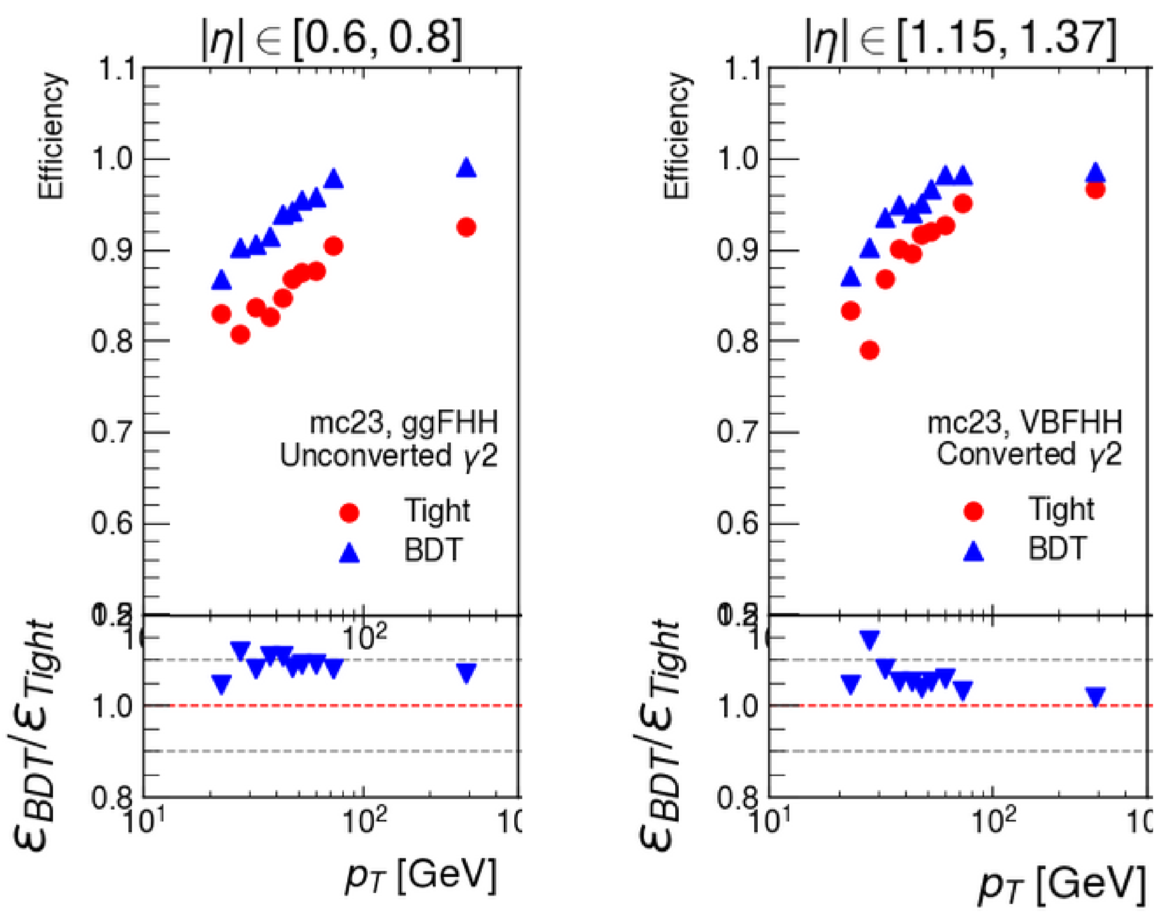
- Tuning a WP for each BDT.
- For every $(|\eta|, p_T)$ bin a threshold on the score has been evaluated such to have the same bkg efficiency as Tight. Events with score higher than threshold pass the BDT WP.

$$\varepsilon_{sig} = \frac{\gamma\,(passPID)}{\gamma\,(Reco, LooseID, LooseISO)} \qquad \varepsilon_{bkg} = \frac{jet\,(passPID)}{jet\,(Reco, LooseID, LooseISO)}$$

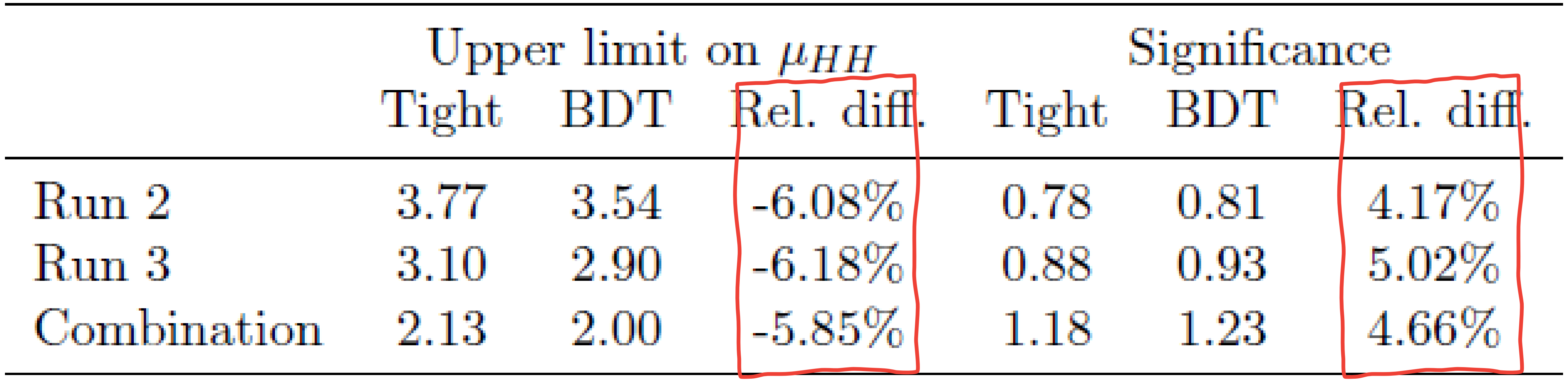


BDT PID impact on the analysis

- The new BDT-based PID is applied to the analysis samples, reproducing the same improvement on efficiencies as observed in the test samples.
- Two assumptions are made to replicate the analysis with the new PID, to avoid large statistical fluctuation induce by the bkg modeling strategy:
  - The PID efficiency on data scales as in MC
  - The background is composed only by irreducible background
- Hence, the normalization of the continuum background inside the models with the BDT PID is fixed to:



$$N_c(\text{bkg, BDT}) = N_c(\text{bkg, Tight}) \cdot \frac{N_c(\text{sig, BDT})}{N_c(\text{sig, Tight})}$$



	Upper limit on μ_{HH}			Significance		
	Tight	BDT	Rel. diff.	Tight	BDT	Rel. diff.
Run 2	3.77	3.54	-6.08%	0.78	0.81	4.17%
Run 3	3.10	2.90	-6.18%	0.88	0.93	5.02%
Combination	2.13	2.00	-5.85%	1.18	1.23	4.66%

Conclusions

- I presented the latest search for di-Higgs production in $b\bar{b}\gamma\gamma$ final state using Run 2 and partial Run 3 ATLAS data.
  - Confirm the shape of the Higgs potential
  - Observe the HH production

- Best fit for μ_{HH} compatible with the SM prediction, with an observed upper limit of 3.8 and an observed significance of 0.84. Put constrains on coupling modifiers κ_λ and κ_{2V}, finding as best fits values compatibles with the SM.

- A new PID based on a BDT has been implemented, showing an improvement from few percent up to more than 10%, depending on $(|\eta|, p_T)$ bin and conversion status.
- The new PID has then been applied to the $b\bar{b}\gamma\gamma$ analysis showing an improvement of about 6% in upper limit of signal strength and of about 4.5% in significance.

- These results are encouraging and set the basis for the use of the new PID in the next round of the analysis.

Thank you for the attention!

And happy H-day to everybody

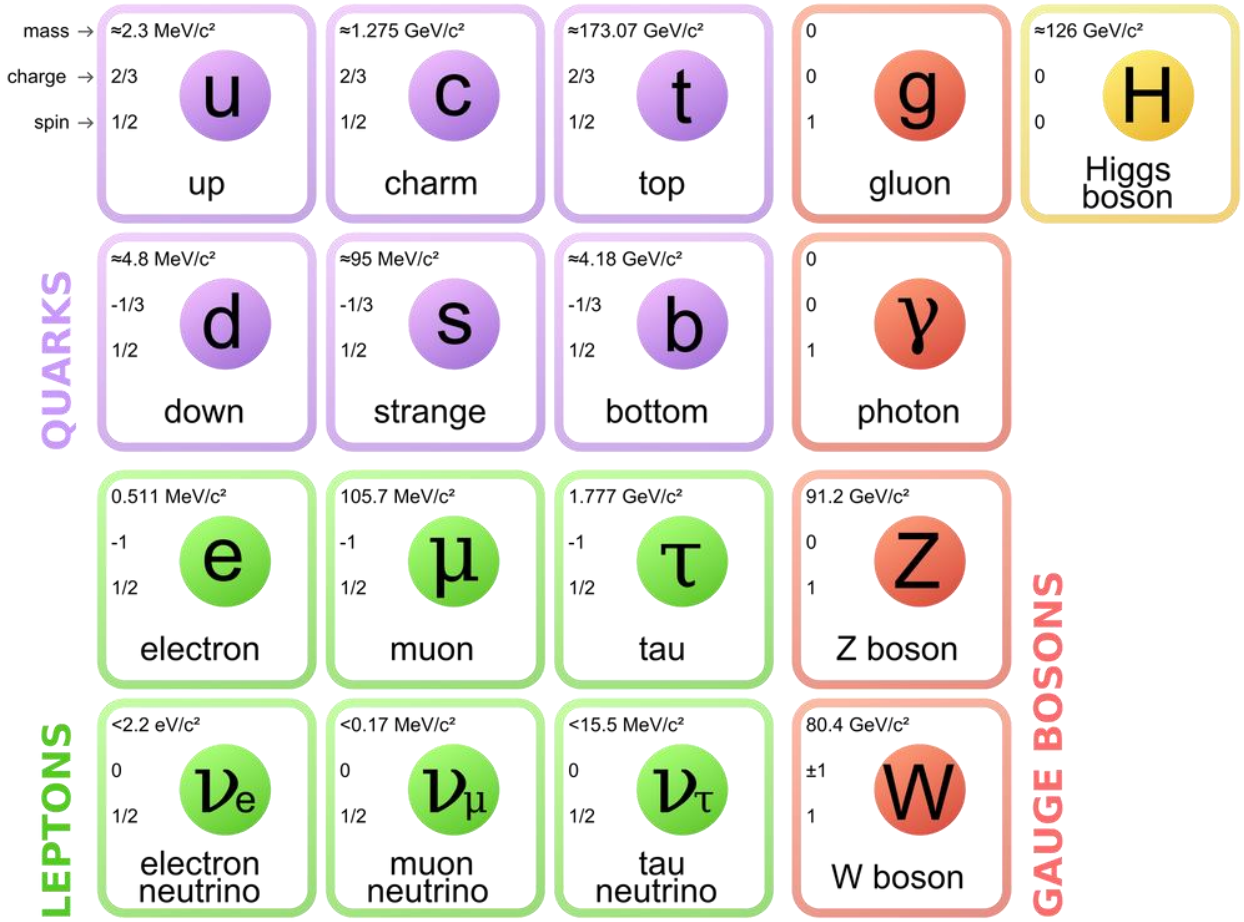
Standard Model

The Standard Model (SM) describes the dynamics of the fundamental particles and their interactions

$$SU(2)_L \times U(1)_Y \times SU(3)_C$$

Electroweak

QCD



Gauge theory: let's start with Lagrangian for a U(1) symmetry group (QED)

U(1) gauge transformation:

$$\psi(x) \to e^{i\theta}\psi(x)$$

$$\bar{\psi}(x) \to e^{-i\theta}\bar{\psi}(x)$$

$$\mathcal{L} = i\bar{\psi}\partial\!\!\!/\psi - m\bar{\psi}\psi$$

$$\mathcal{L}' = \mathcal{L} - \bar{\psi}\gamma^\mu\psi\partial_\mu\theta$$

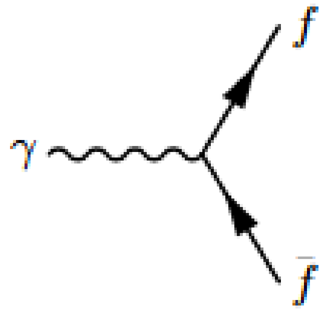
Making it local

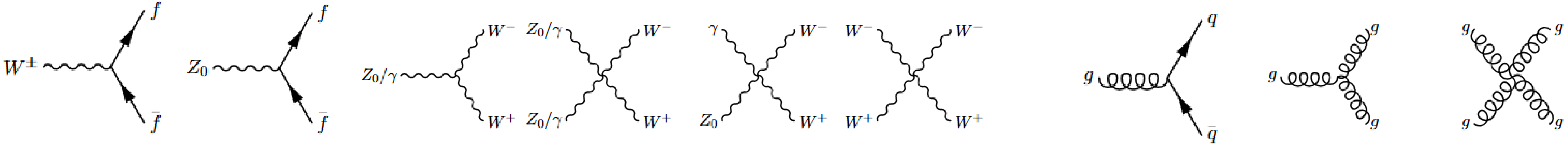
To restore the gauge invariance a new field is introduced through the covariant derivative:

$$D_\mu(x) \equiv \partial_\mu + iqA_\mu(x) \qquad A_\mu(x) \to A_\mu(x) - \frac{1}{q}\partial_\mu\theta(x)$$

In this way an interaction term between fermion and vector fields appears:

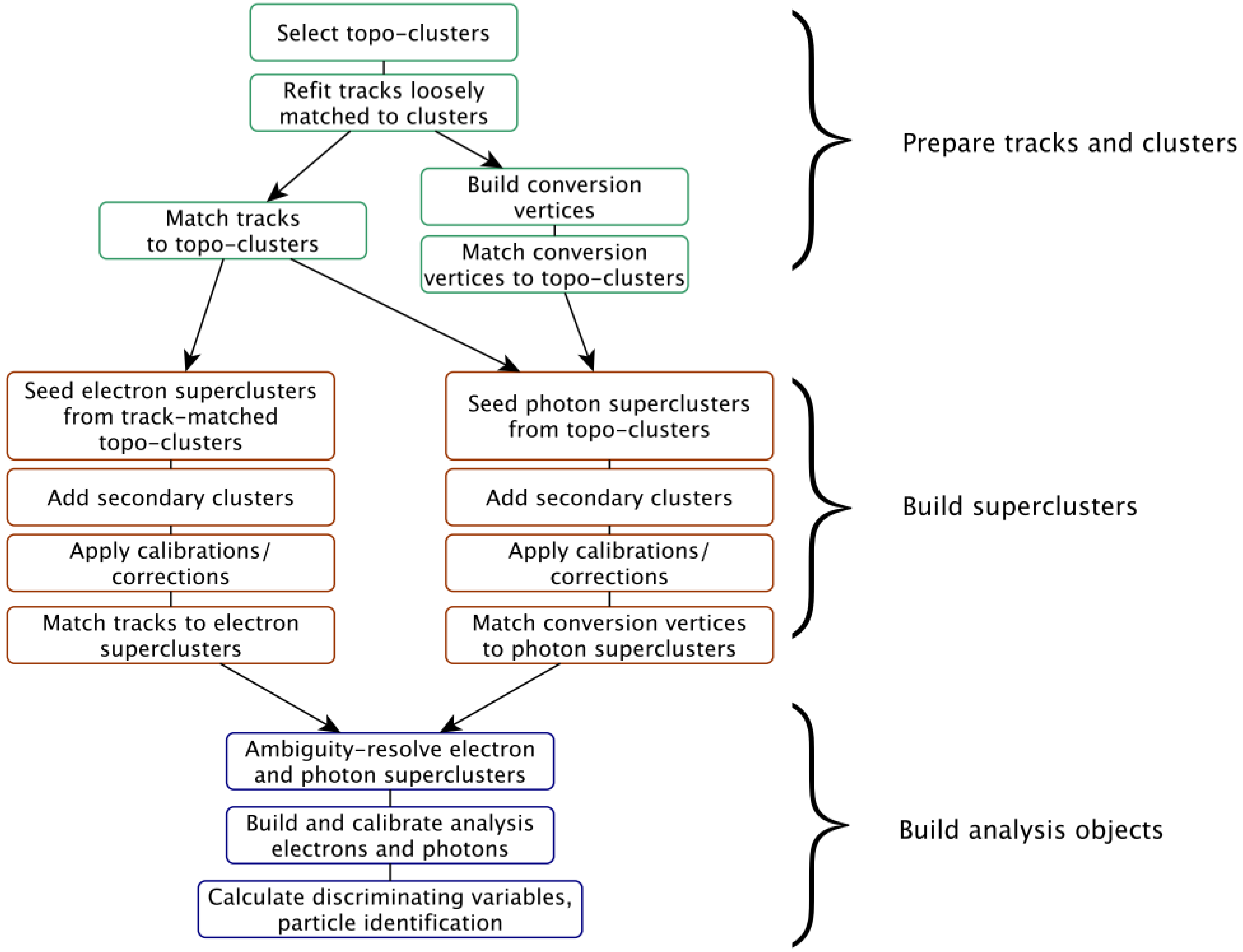
$$\mathcal{L}_{QED} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + i\bar{\psi}D\!\!\!/\psi - m\bar{\psi}\psi = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + i\bar{\psi}\partial\!\!\!/\psi - m\bar{\psi}\psi - qA_\mu\bar{\psi}\gamma^\mu\psi$$





Object reconstruction

- Photon reconstruction
- Energy calibration: correct the energy losses that affect the linearity and resolution of the measurements
- Identification (PID): differentiate the prompt and non-prompt photons
- Isolation: reduce the contribution from fake photons from nearby hadronic activity



Select topo-clusters

Refit tracks loosely matched to clusters

Match tracks to topo-clusters

Build conversion vertices

Match conversion vertices to topo-clusters

Prepare tracks and clusters

Seed electron superclusters from track-matched topo-clusters

Add secondary clusters

Apply calibrations/ corrections

Match tracks to electron superclusters

Seed photon superclusters from topo-clusters

Add secondary clusters

Apply calibrations/ corrections

Match conversion vertices to photon superclusters

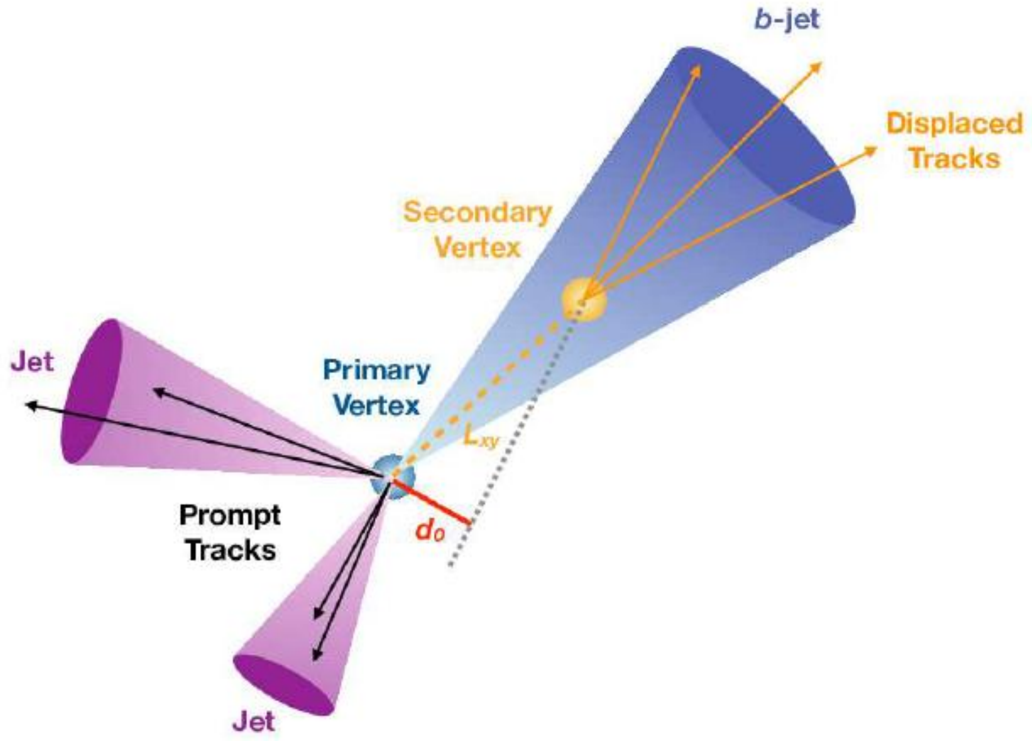
Build superclusters

Ambiguity-resolve electron and photon superclusters

Build and calibrate analysis electrons and photons

Calculate discriminating variables, particle identification

Build analysis objects



A jet is a bunch of collimated hadrons generated through parton shower and hadronization of an initial parton.

- Particle Flow: reconstruct a jet matching tracks to topo-cluster in the HCal
- JVT: remove background jets from pile-up
- b-jet tagging: multivariate technique which exploits characteristics of b-jets.
  - Secondary verteces
  - Impact parameter
  - Soft leptons
- Energy calibration: correction of the dijet invariant mass
  - Mu-in-jet
  - PtReco

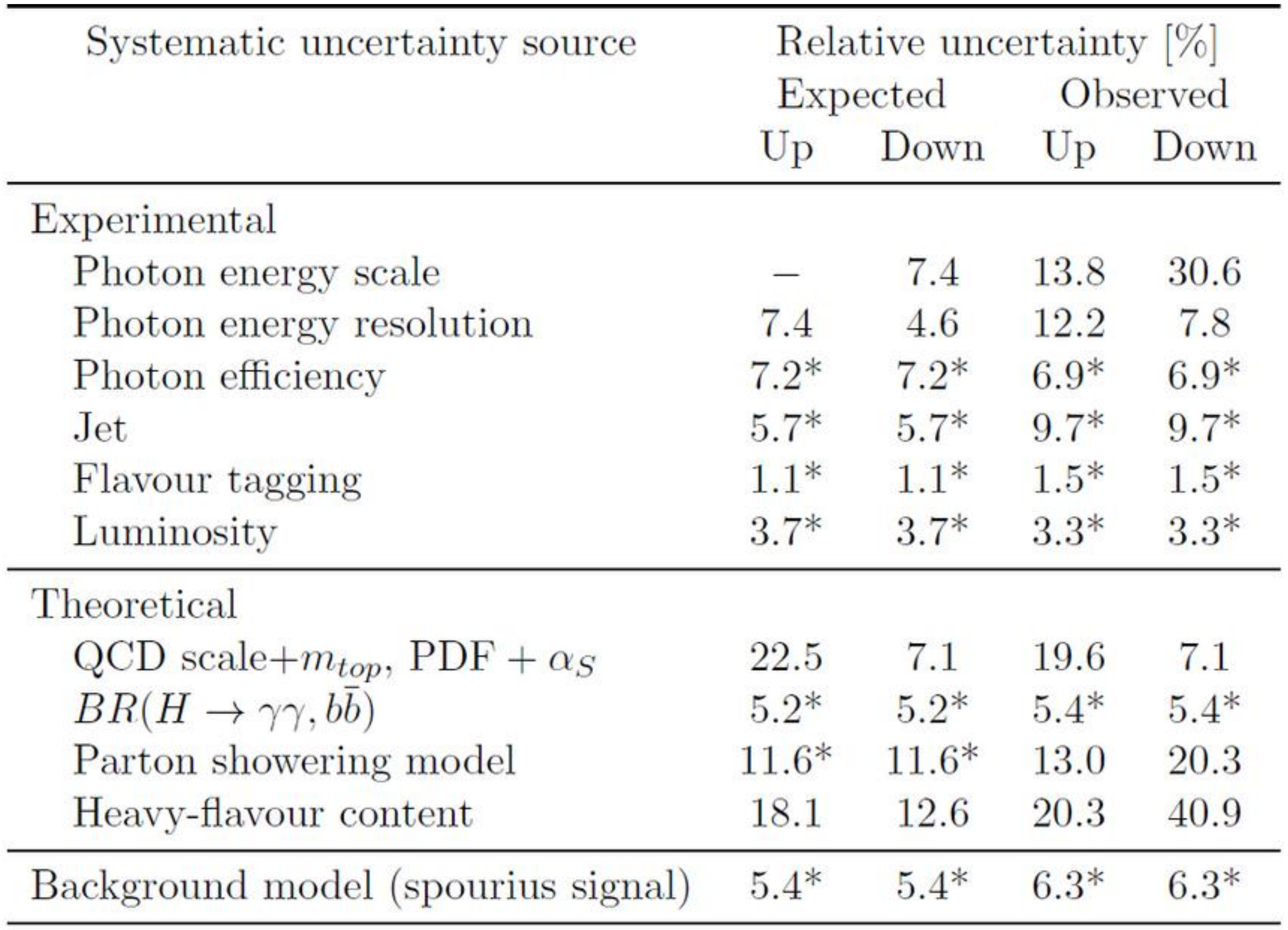
Systematic uncertainties

- The sensitivity of this analysis is limited by the statistical precision. Systematic uncertainties are nevertheless essential part of a measurement.
- The impact of each source is quantified by propagating its uncertainty troughout the full analysis, and quoted as the relative difference between the nomilan and varied results.

- The experimental uncertainties, from auxiliary measurement in model construction or from object reconstruction, can affect the yields and the shapes of the resonat processes.

- The theoretical uncertainties, maily from limitations in QCD calculations, affect the predicted yields of the resonant processes.

- Spurious signal only systematic related to the continuum background. It's the bias on the fitted signal yield on the backgroud-only template.



Systematic uncertainty source	Relative uncertainty [%]			
	Expected		Observed	
	Up	Down	Up	Down
Experimental				
Photon energy scale	–	7.4	13.8	30.6
Photon energy resolution	7.4	4.6	12.2	7.8
Photon efficiency	7.2*	7.2*	6.9*	6.9*
Jet	5.7*	5.7*	9.7*	9.7*
Flavour tagging	1.1*	1.1*	1.5*	1.5*
Luminosity	3.7*	3.7*	3.3*	3.3*
Theoretical				
QCD scale+m_{top}, PDF + α_S	22.5	7.1	19.6	7.1
$BR(H \to \gamma\gamma, b\bar{b})$	5.2*	5.2*	5.4*	5.4*
Parton showering model	11.6*	11.6*	13.0	20.3
Heavy-flavour content	18.1	12.6	20.3	40.9
Background model (spourius signal)	5.4*	5.4*	6.3*	6.3*

Results

➯ Additional fits constrain the trilinear Higgs and VVHH couplings to -17<kl<6.6 and -0.5<k2v<2.6 at 95% CL.