BDT-Based Photon Identification: Calibration and Efficiency Gain in Di-Higgs Searches in the $b\bar{b}\gamma\gamma$ Channel

Hind Taibi

FJPPN Meeting
28th September 2026

Introduction

Introduction

Di-Higgs Searches

$$\mathcal{L}^{\mathrm{SM}} \supset V(\boldsymbol{\Phi}) = \boxed{V_0 - \frac{m_H^2}{2} h^2 + \lambda_{HHH}^{\mathrm{SM}} h^3 + \lambda_{HHHH}^{\mathrm{SM}} h^4}$$

Higgs boson's self-coupling constant

Not accessible at the LHC

Introduction

Di-Higgs Searches

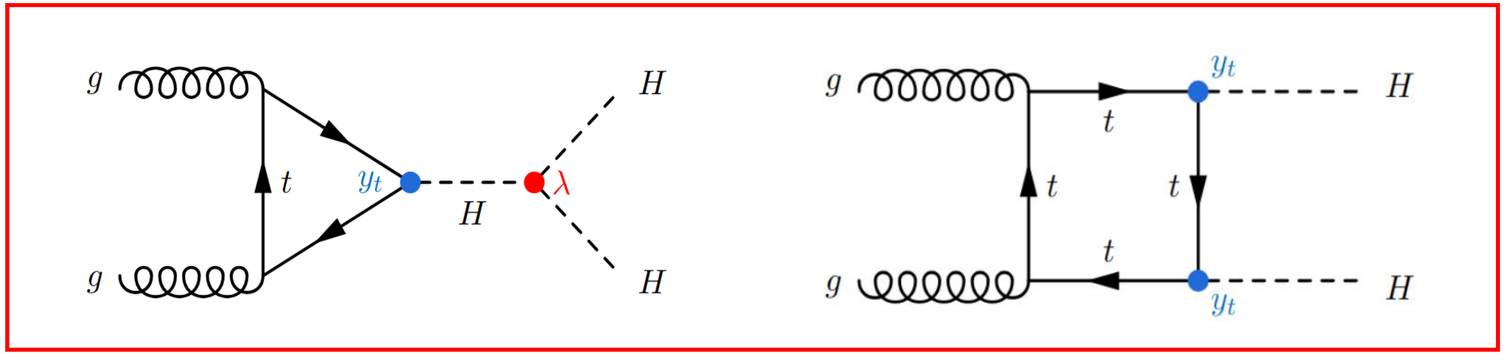
$$\mathcal{L}^{\mathrm{SM}} \supset V(\boldsymbol{\Phi}) = V_0 - \frac{m_H^2}{2} h^2 + \lambda_{HHH}^{\mathrm{SM}} h^3 + \lambda_{HHHH}^{\mathrm{SM}} h^4$$

Latest constraints on the Higgs boson's self-coupling constant:

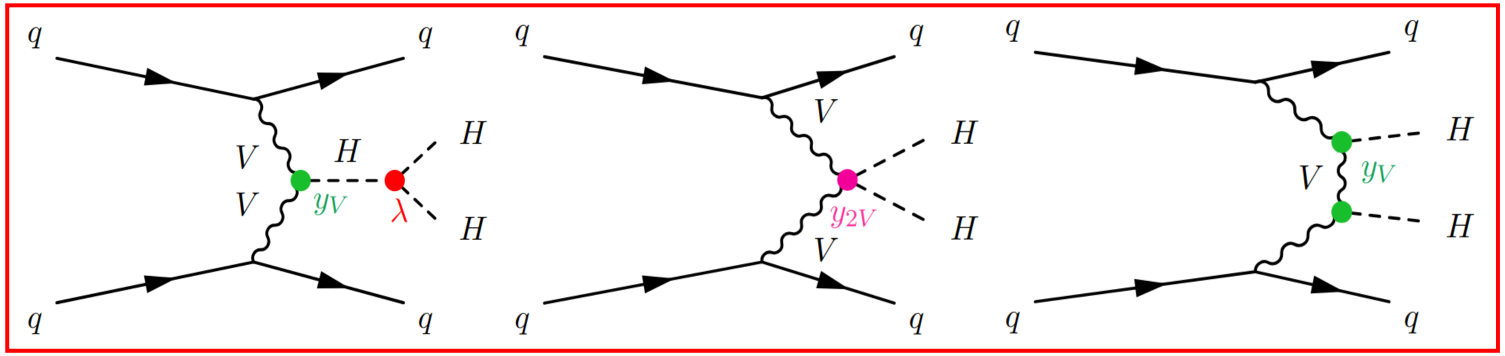
$$-1.3 < \kappa_\lambda < 6.7$$
$$\equiv$$
$$\lambda_{HHH}/\lambda_{HHH}^{\mathrm{SM}}$$

Combination of ATLAS and CMS searches for Higgs boson pair production at $\sqrt{s}$ = 13 TeV with the ATLAS detector

LO ggF Feynman Diagrams



LO VBF Feynman Diagrams



Introduction

The $HH \to b\bar{b}\gamma\gamma$ Channel

High contribution from $HH \to b\bar{b}\gamma\gamma$ in di-Higgs searches

Latest results of the $HH \to b\bar{b}\gamma\gamma$ analysis using full Run 2 and partial Run 3 data:

$$-1.6 < \kappa_\lambda < 6.6$$

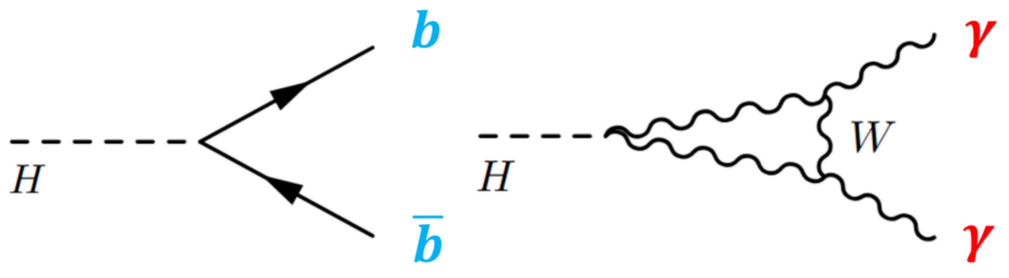
With
- New b-tagging algorithm
- Same photo identification procedure

➡ Implement **new photon ID** for next analysis (e. g. BDT, gives **~5-10 % more efficiency**)

➡ **Calibrate** the new photon ID

[Study of Higgs boson pair production in the $HH \to b\bar{b}\gamma\gamma$ final state with 308 fb^{-1} of data collected at $\sqrt{s} = 13$ TeV and 13.6 by the ATLAS experiment]

	bb	WW	$\tau\tau$	ZZ	$\gamma\gamma$
bb	34%				
WW	25%	4.6%			
$\tau\tau$	7.3%	2.7%	0.39%		
ZZ	3.1%	1.1%	0.33%	0.069%	
$\gamma\gamma$	0.26%	0.10%	0.028%	0.012%	0.0005%



Introduction

Photon ID in $HH \rightarrow b\bar{b}\gamma\gamma$

About **15 %** of HH signal events are lost because of non-optimal cut-based photon ID.

➡ Photon ID should be improved with a new ML-based technique

Run3 ggF HH SM for Full Simulation

Selection	Efficiency [%]
All events	100.00
event clean and dalitz filter	98.71
Pass trigger	59.39
2 loose photons	91.92
trigger matching	99.41
2 tight ID photon	85.77
2 tight ID and Iso photons	91.36
rel. pT cuts	92.45
myy in [105,160]GeV	99.91
Nlep =0	99.68
at least 2 Nj	88.04
Nj central<6	97.70
At least 2 b-jets with 85% WP	49.91

Photon Identification in ATLAS

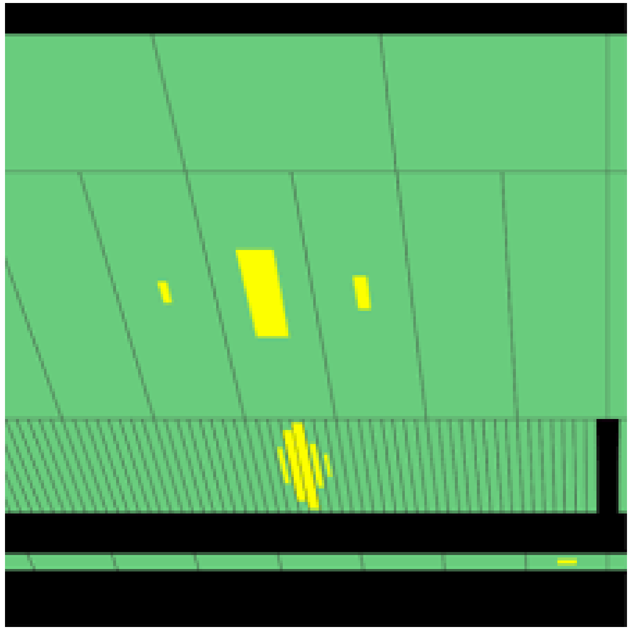
Photon Identification in ATLAS

Prompt (*True*) and Non-Prompt (*Fake*) Photons

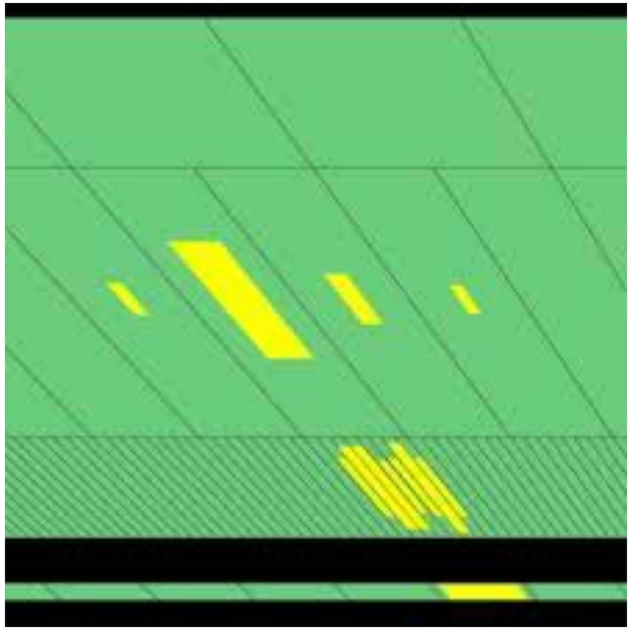
True photons: photons in the final state not coming from hadron decays ($H \rightarrow \gamma\gamma$)

Fake photons: photons coming from the decay of neutral mesons in QCD jets ($\pi^0 \rightarrow \gamma\gamma$)

- ➡ Highly collimated photon pair
- ➡ Dominant source of background in $H \rightarrow \gamma\gamma$



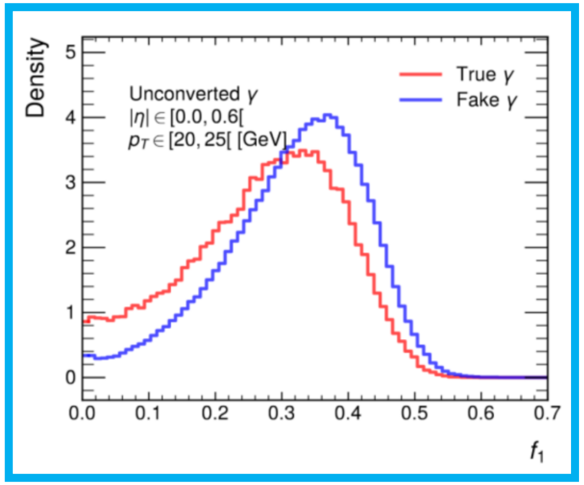
Display of **prompt-photon** event

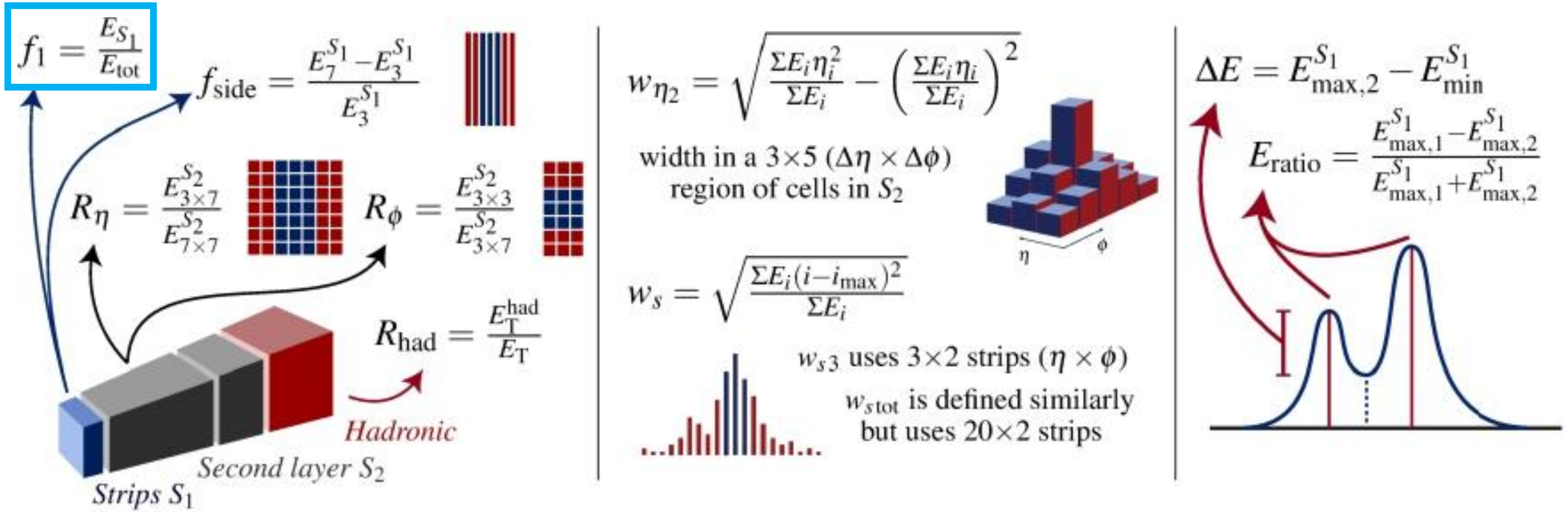


Display of **non-prompt-photon** event

Photon Identification in ATLAS

Shower-Shape Variables for Photon Identification





Legacy approach: rectangular cuts on shower-shape variables (SSV)

➡ ***Loose*** and ***Tight*** photon ID

New approach: use SSV to train a machine-learning (ML) algorithm (e. g. BDT)

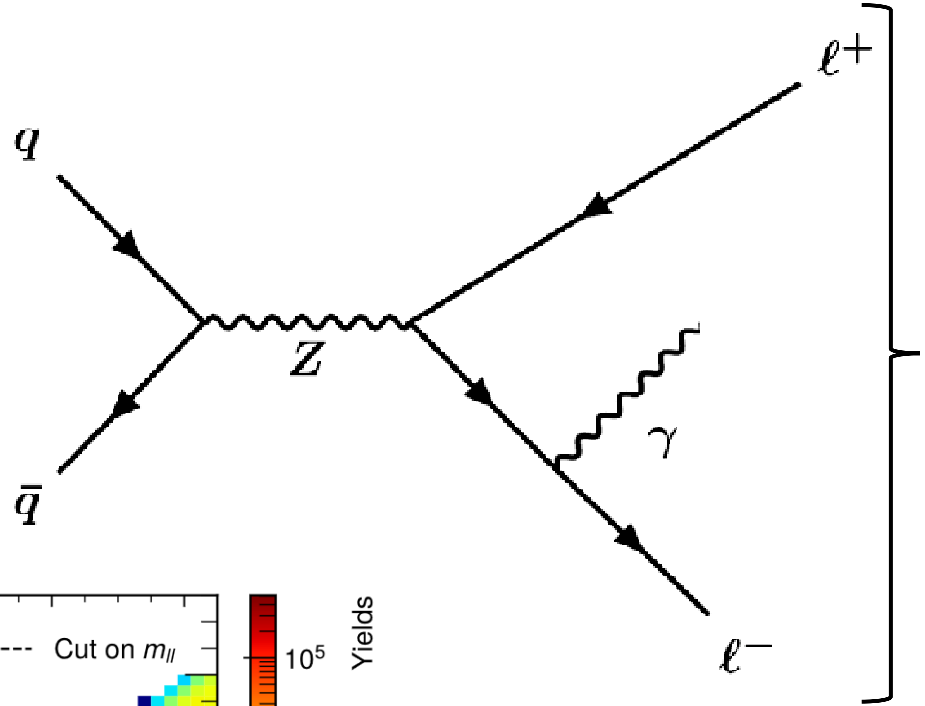
➡ ***ML-based*** photon ID

} Must be calibrated to be used in analyses

Photon Identification Calibration with Radiative Z

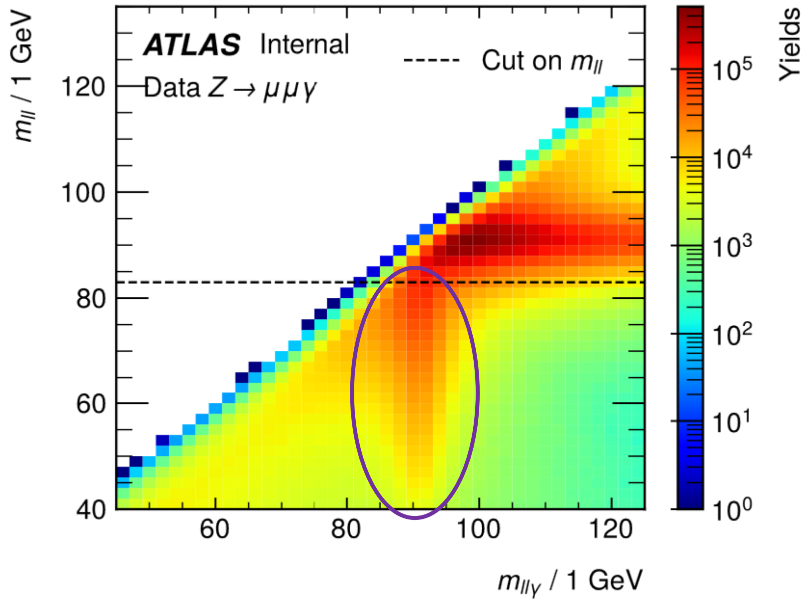
Photon Identification Calibration with Radiative Z

Radiative Z Method



$m^{ll\gamma} = m^{Z}$

➡ **Clear bump at ~91 GeV** in $m^{ll\gamma}$ distribution



A cut on the m^{ll} distribution, alongside other selections on the leptons, removes initial state radiation (ISR)

➡ Pure samples with a true photon in the final state

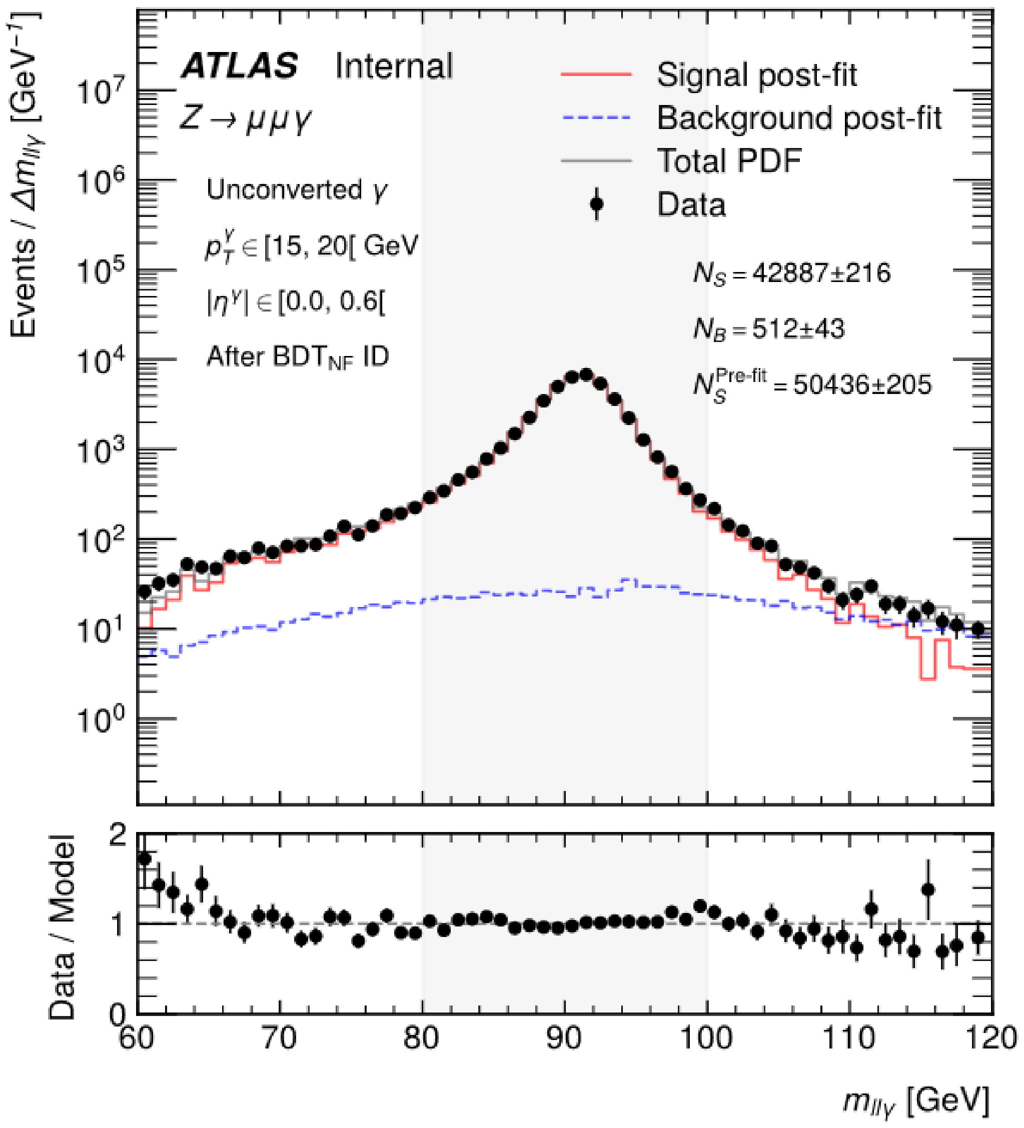
➡ Photon ID efficiency can be estimated from the photon ID performance on these samples

Photon Identification Calibration with Radiative Z

General Procedure

We use the distribution of the invariant mass $m^{ll\gamma}$ in $Z \to ll\gamma$ decays to estimate the number of signal events before and after Photon ID.

$$\boldsymbol{\varepsilon_{\mathrm{Id.}} = \frac{N_S^{ID}}{N_S}}$$



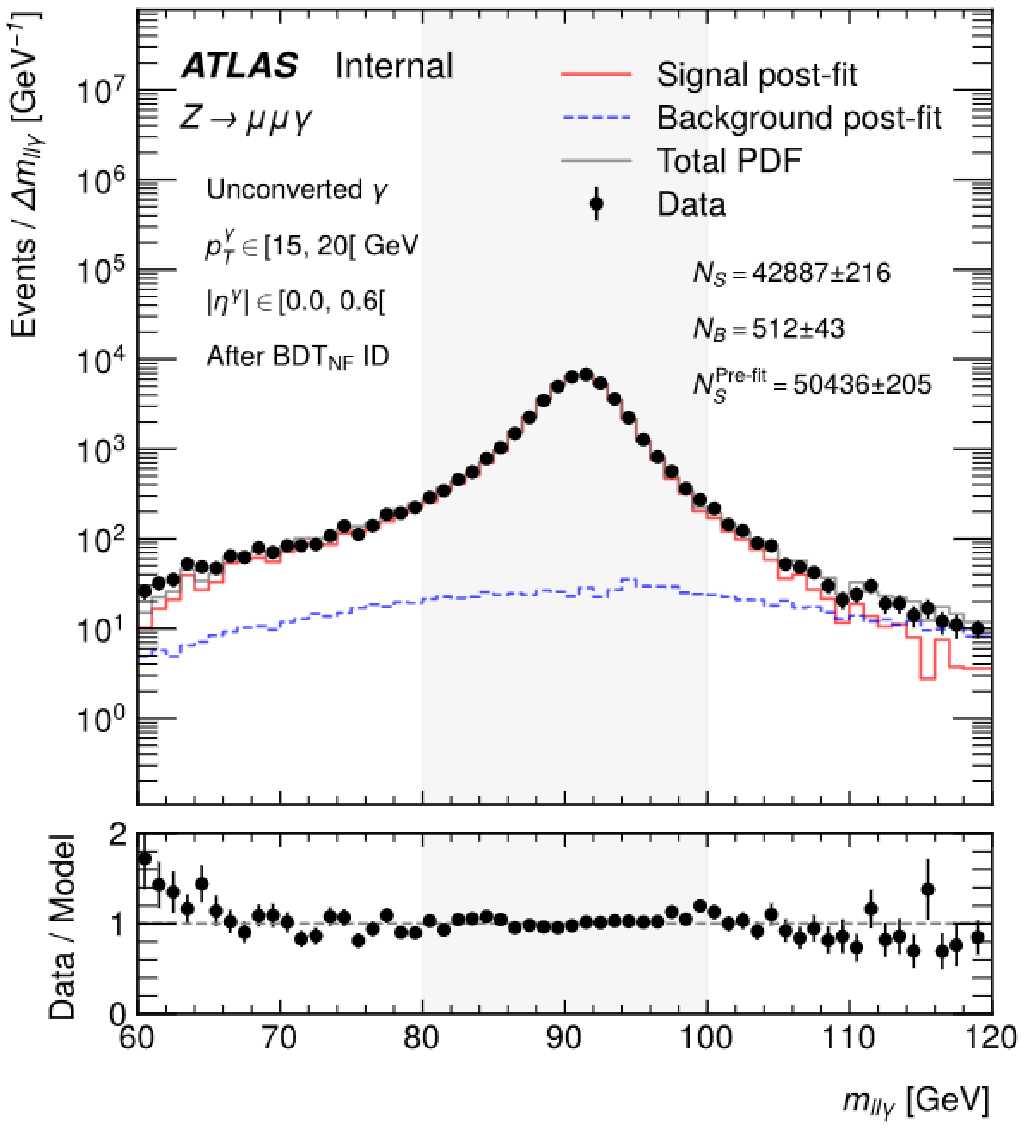
Photon Identification Calibration with Radiative Z

General Procedure

We use the distribution of the invariant mass $m^{ll\gamma}$ in $Z \to ll\gamma$ decays to estimate the number of signal events before and after Photon ID.

$$\varepsilon_{\mathrm{ID}} = \frac{N_S^{ID}}{N_S}$$

In MC the number of signal events is obtained by integrating the $m^{ll\gamma}$ distribution.



Photon Identification Calibration with Radiative Z

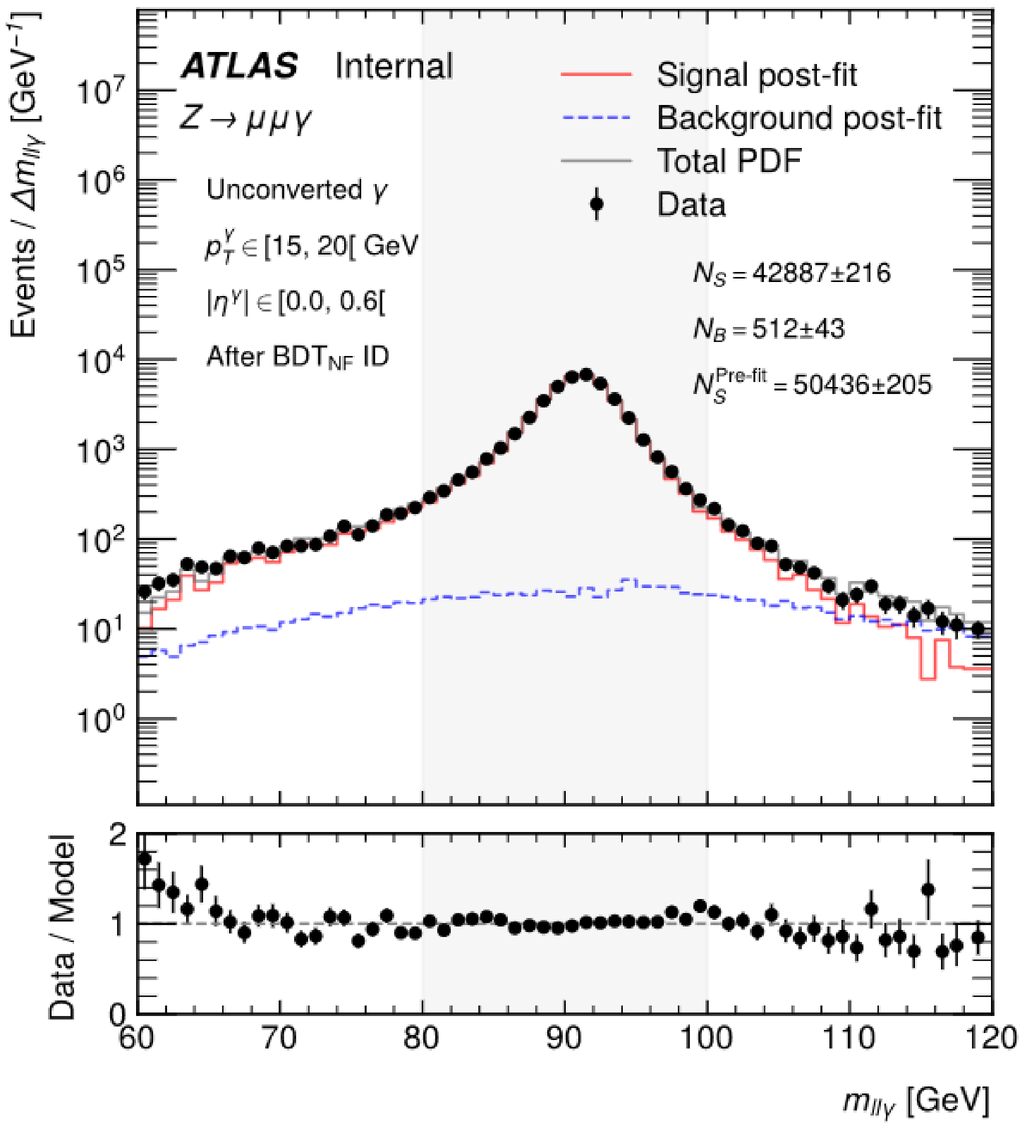
General Procedure

We use the distribution of the invariant mass $m^{ll\gamma}$ in $Z \to ll\gamma$ decays to estimate the number of signal events before and after Photon ID.

$$\varepsilon_{\text{ID}} = \frac{N_S^{ID}}{N_S}$$

In MC the number of signal events is obtained by integrating the $m^{ll\gamma}$ distribution.

In data the number of signal events is estimated with a template fit where $\mathbf{PDF} = \mathbf{PDF}_{\mathbf{MC}} + \mathbf{PDF}_{\mathbf{Bkg}}$ is fitted to the data and yields N_S and N_B are the free parameters of the fit.



Photon Identification Calibration with Radiative Z

General Procedure

We use the distribution of the invariant mass $m^{ll\gamma}$ in $Z \rightarrow ll\gamma$ decays to estimate the number of signal events before and after Photon ID.

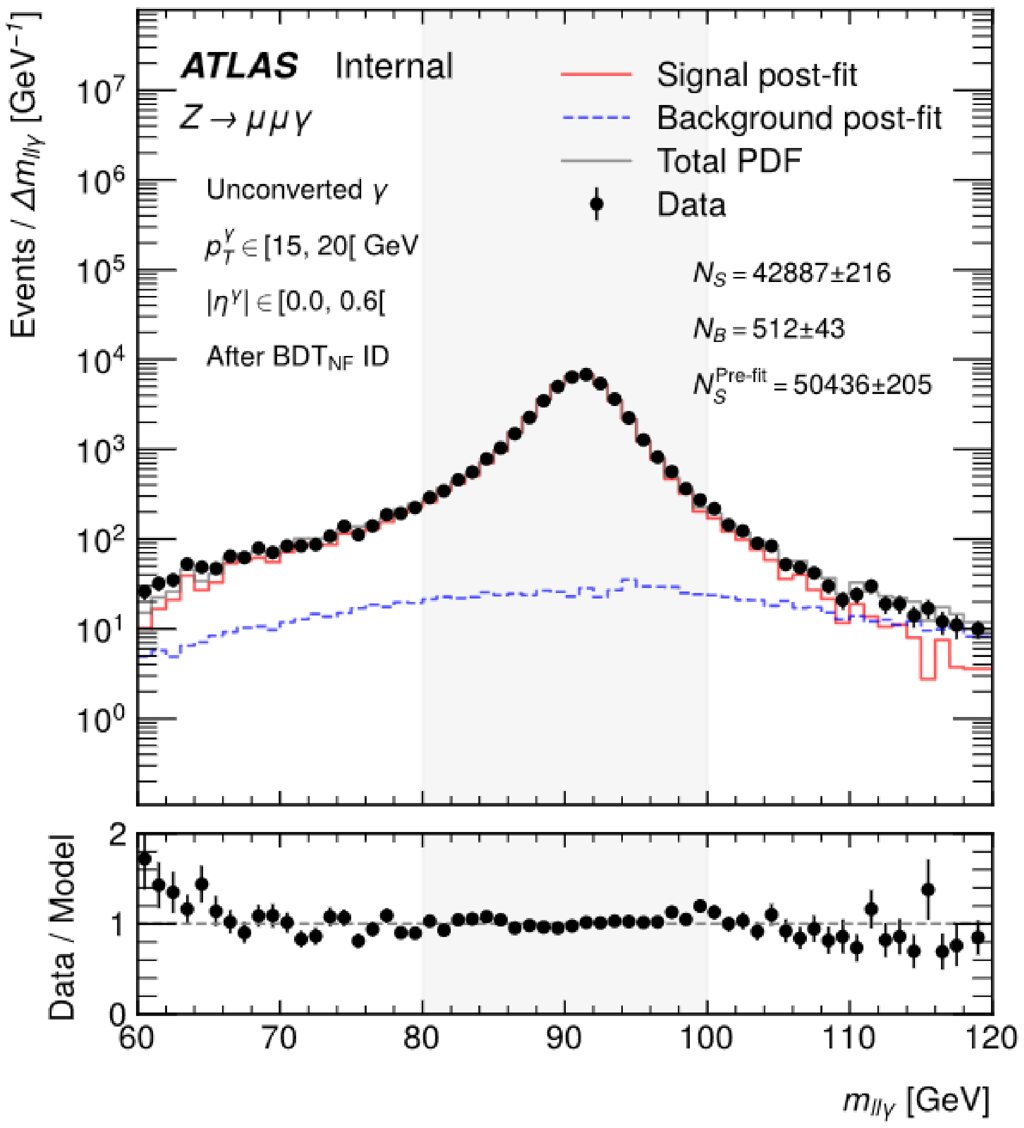
$$\boldsymbol{\varepsilon_{\mathrm{ID}} = \frac{N_S^{ID}}{N_S}}$$

In MC the number of signal events is obtained by integrating the $m^{ll\gamma}$ distribution.

In data the number of signal events is estimated with a template fit where $\mathbf{PDF} = \mathbf{PDF_{MC}} + \mathbf{PDF_{Bkg}}$ is fitted to the data and yields $\boldsymbol{N_S}$ and $\boldsymbol{N_B}$ are the free parameters of the fit.

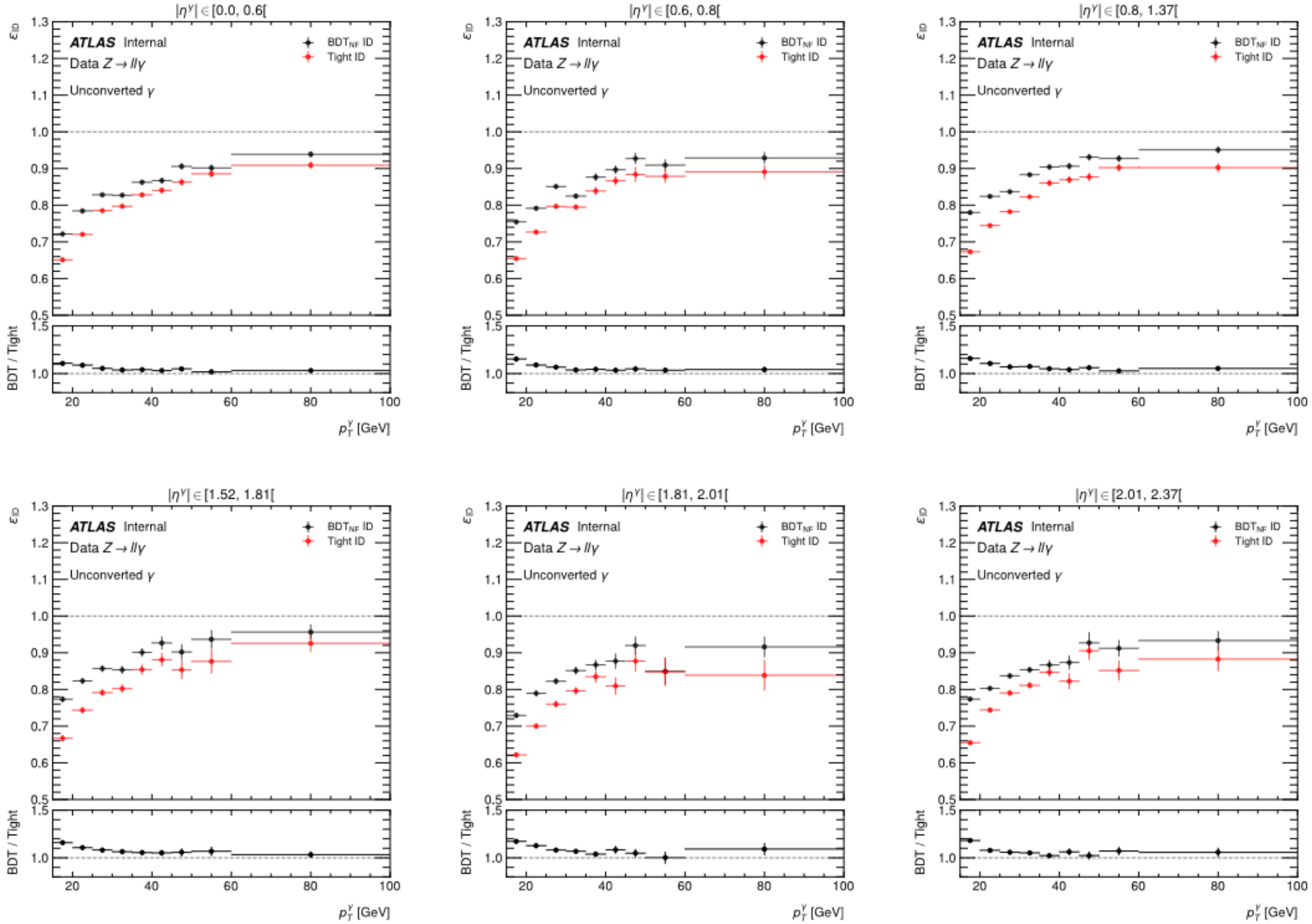
Scale factors to correct the MC difference w. r. t. data:

$$\boldsymbol{SF = \frac{\varepsilon_{\mathrm{ID}}^{\mathrm{Data}}}{\varepsilon_{\mathrm{ID}}^{\mathrm{MC}}}}$$



Photon Identification Calibration with Radiative *Z*

Efficiency



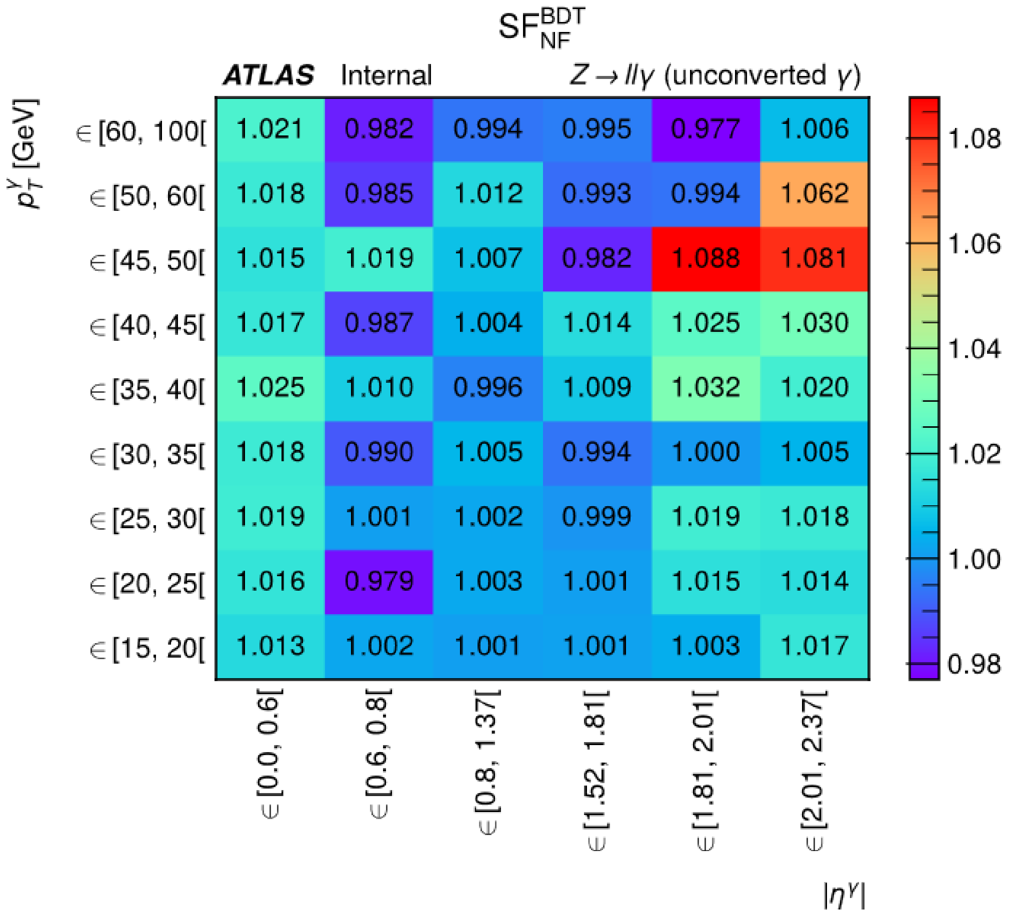
BDT was trained on JF MC samples with shower shapes corrected with notmalizing flows.

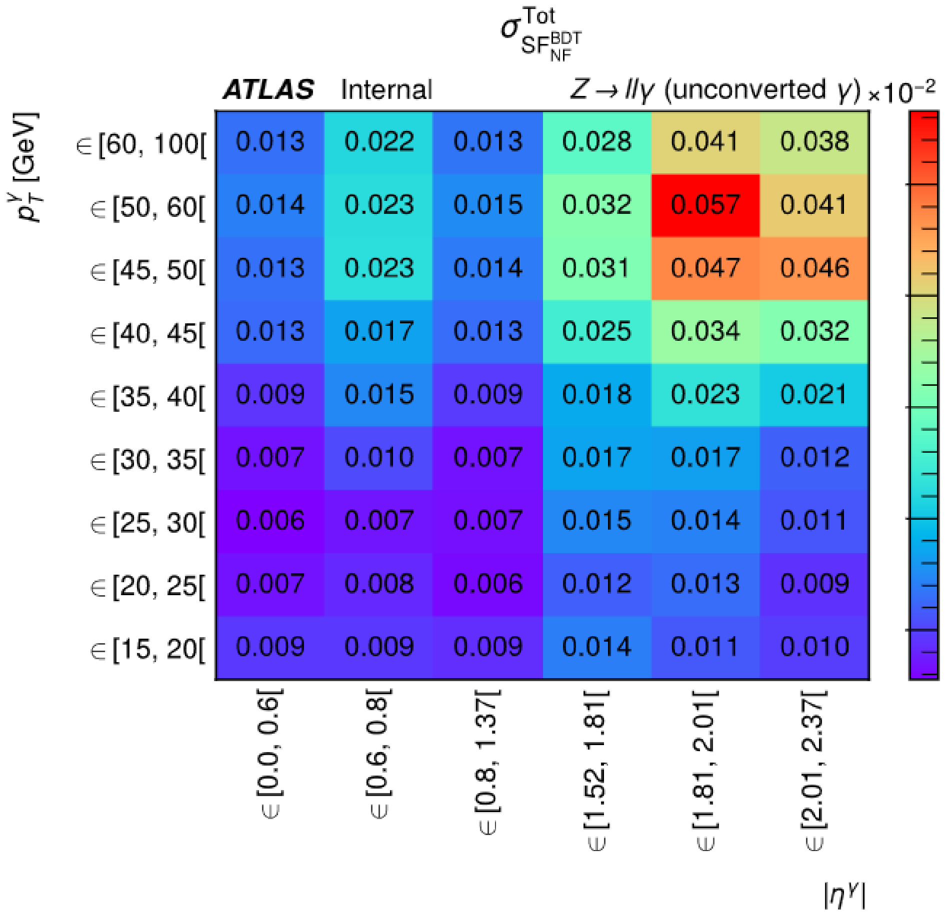
Efficiency of BDT-based photon ID has been computed

➡ 5-20 % improvement on Photon ID efficiecny by using BDT-based ID instead of cut-based ID.

Photon Identification Calibration with Radiative Z

Scale Factors



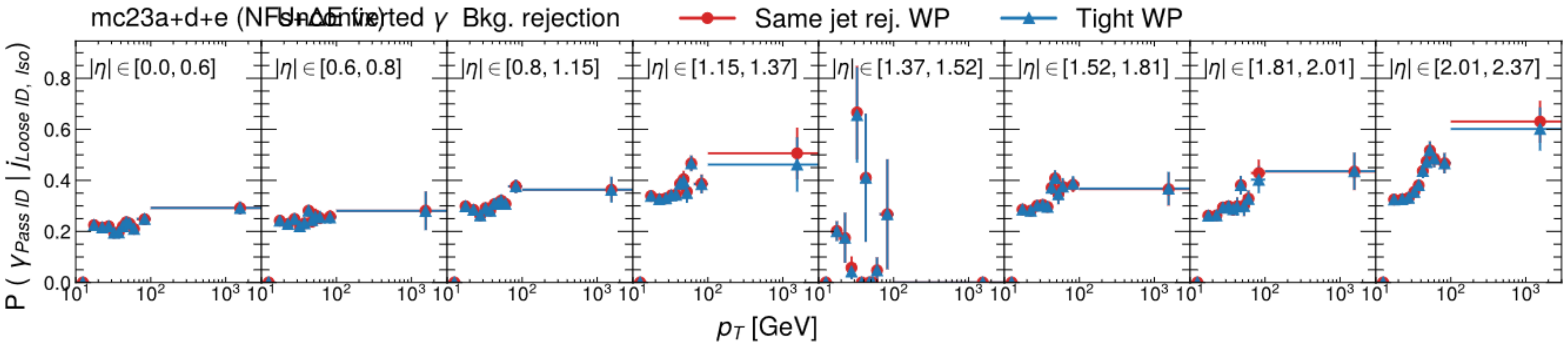


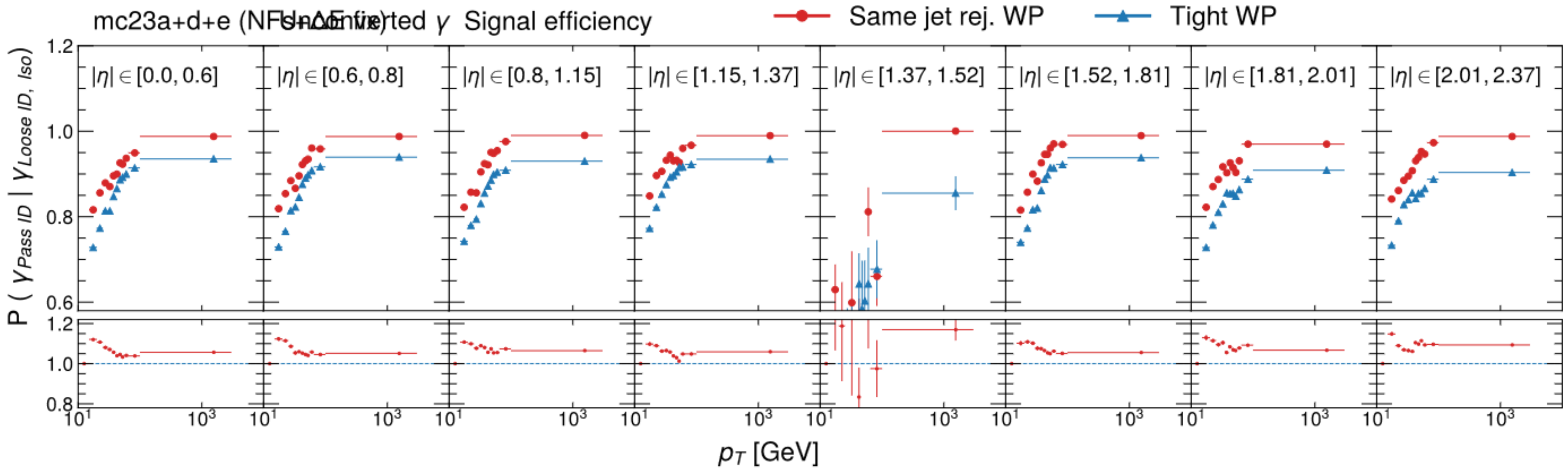
Scale factors are close to 1 with total uceratinty of ~1-6 %

New BDT-Based Photon Identification on $HH \rightarrow b\bar{b}\gamma\gamma$

New BDT-Based Photon Identification on $b\bar{b}\gamma\gamma$

Working-Point Definition





WP optimized such as we keep the same background efficiency as Tight ID

New BDT-Based Photon Identification on $HH \rightarrow b\bar{b}\gamma\gamma$

Improvement on $HH \rightarrow b\bar{b}\gamma\gamma$ Analysis

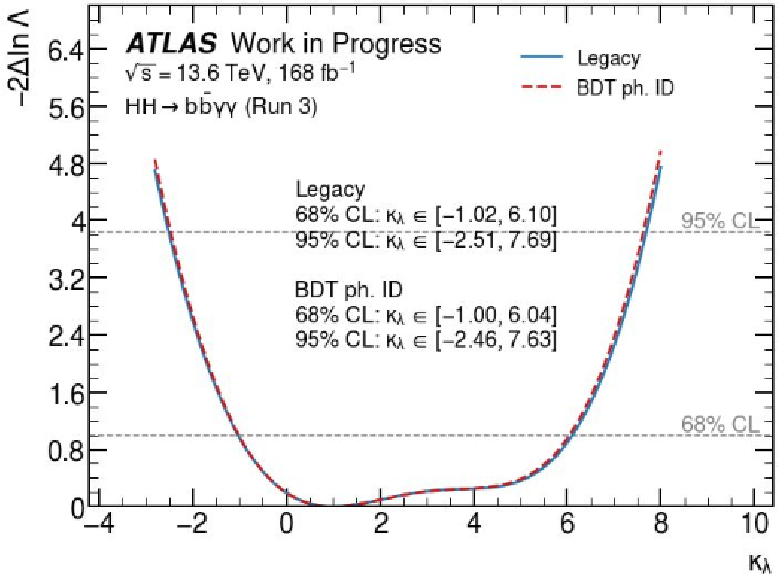
- Expected limits (stat only) are set on the di-Higgs signal strength at 95% CLs.

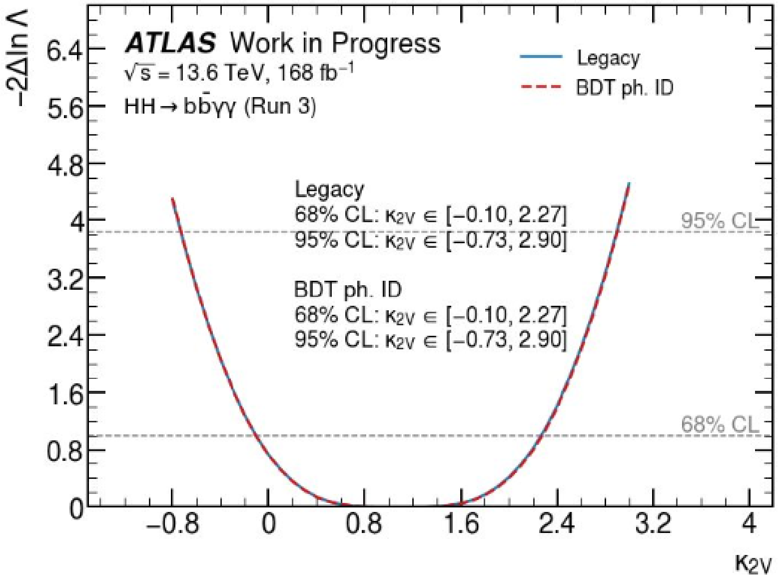
	Tight ID	BDT-based ID	Δ, %
μ(HH) limit	3.44	3.37	-2.0
HH significance	0.775	0.775	0

The same analysis categories were used as for TightID WP

2.0% improvement for μ(HH) upper limit is observed without changing analysis categories.

- Expected 68% and 95% CL limits for kl and k2v are observed via a profile log-likelihood scan.





- About 1% improvement at 95% CL is found for expected kl.
- It seems that there is no improvement for k2v.

Katerina's results

Conclusion

Conclusion

Conclusion

- Generic procedure of photon ID efficiency computation completed,
- 5-10% improvement in photon ID efficiency w. r. t. cut-based ID,
- New BDT-based photon ID gives us 5-10% more signal using the same background rejection working point.
- 1% improvement on expected κ_{λ}.

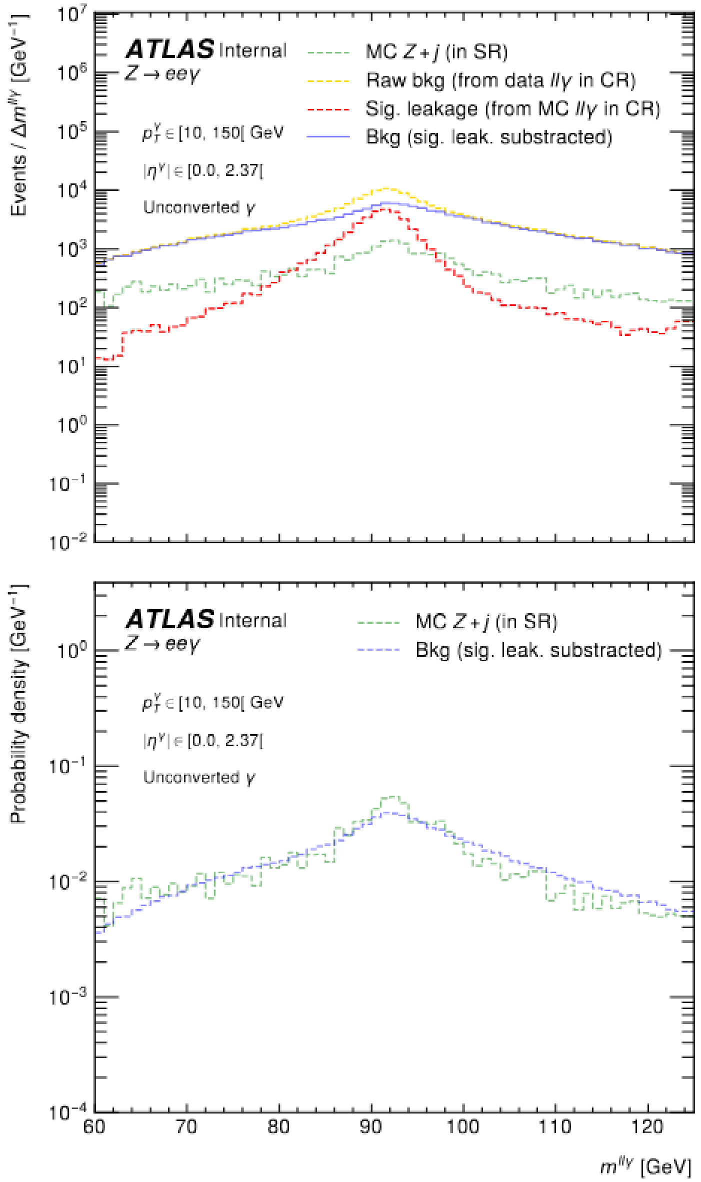
Backup

Backup

Background Template

Not enough statistics of **MC $Z + j$**

➡ We use a **data-driven** background estimation approach



*Control region (CR): anti-id.-and-anti-iso. region
**Signal region (SR): id.-and-iso. region

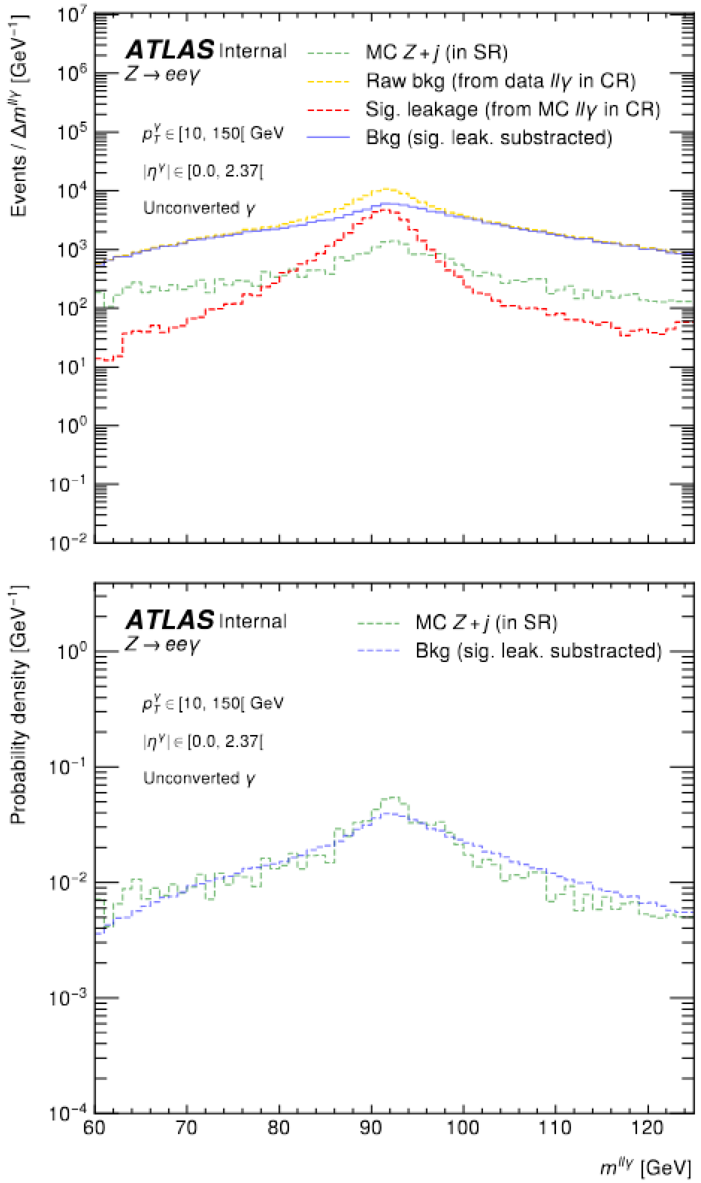
Backup

Background Template

Not enough statistics of **MC** ***Z + j***

➡ We use a **data-driven** background estimation approach

Data $Z \rightarrow ll\gamma$ in CR* ➡ **Raw background**
(= Background + signal leakage)



*Control region (CR): anti-id.-and-anti-iso. region
**Signal region (SR): id.-and-iso. region

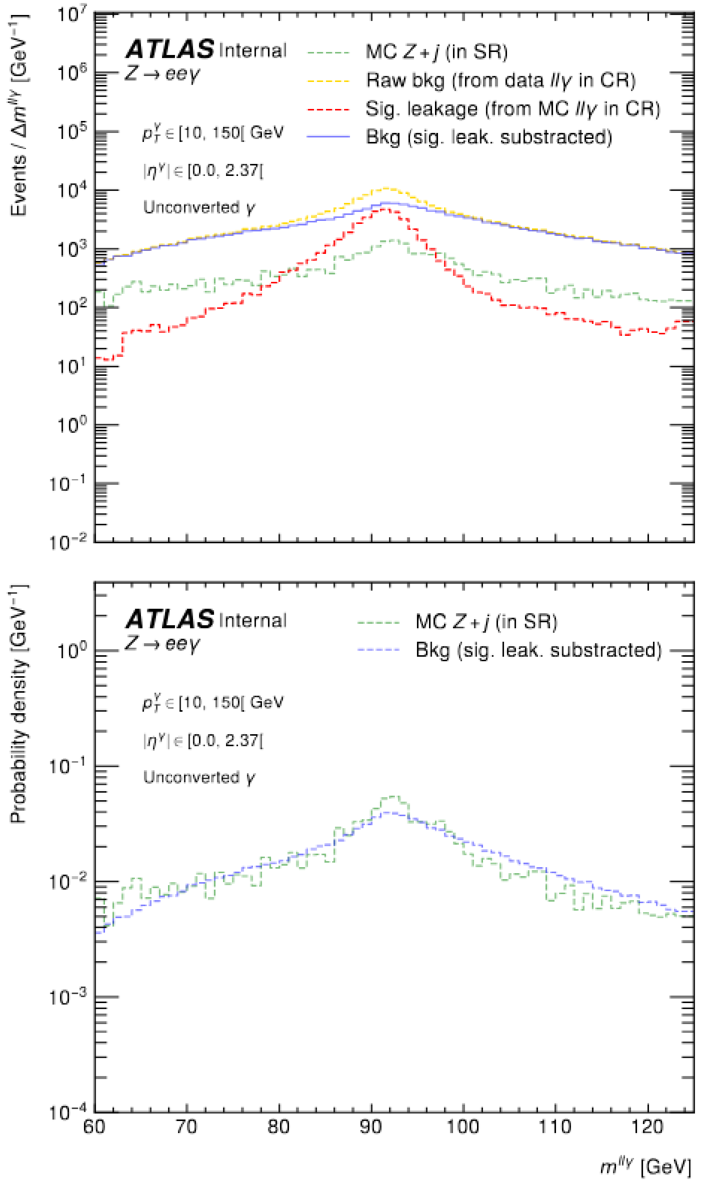
Background Template

Not enough statistics of **MC** ***Z* + *j***

➡ We use a **data-driven** background estimation approach

Data $Z \to ll\gamma$ in CR* ➡ Raw background
(= Background + signal leakage)

MC $Z \to ll\gamma$ in CR* ➡ Signal leakage



*Control region (CR): anti-id.-and-anti-iso. region
**Signal region (SR): id.-and-iso. region

Backup

Background Template

Not enough statistics of **MC** $\boldsymbol{Z+j}$

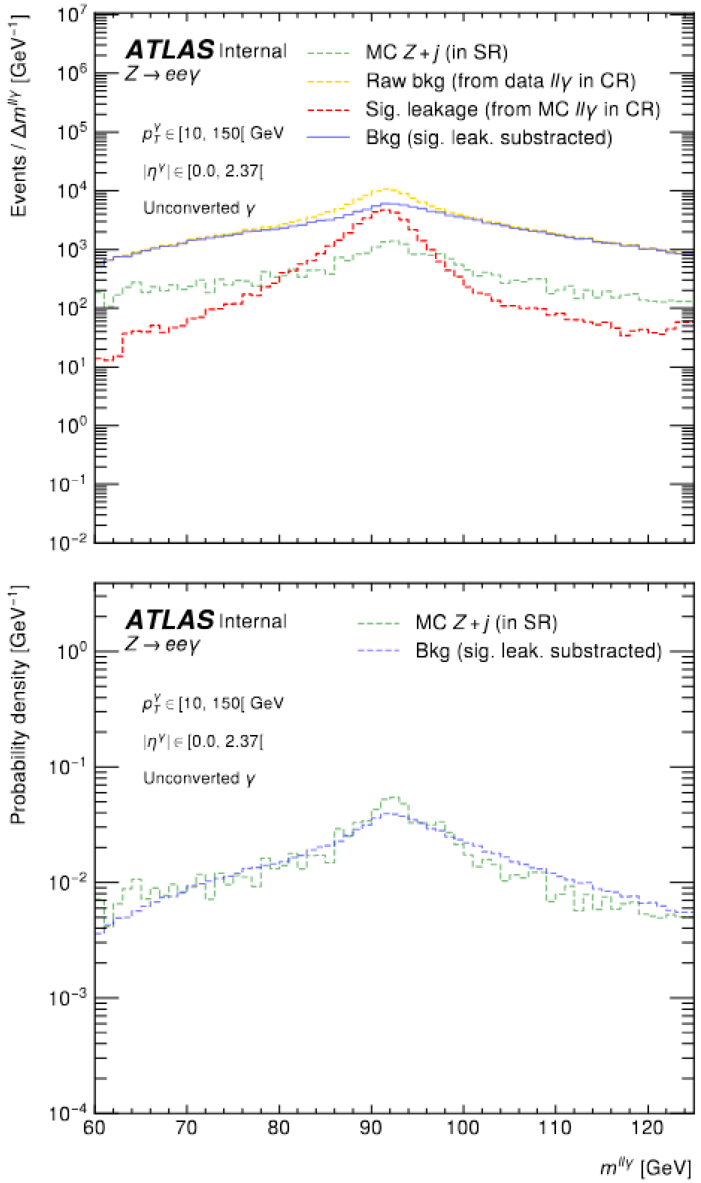
➡ We use a **data-driven** background estimation approach

Data $Z \to ll\gamma$ in CR* ➡ Raw background
(= **Background** + **signal leakage**)

MC $Z \to ll\gamma$ in CR* ➡ Signal leakage

Raw background - **Signal leakage** = **Background template used for the fit**

➡ Shape to be compared with **MC** $\boldsymbol{Z+j}$ in SR**

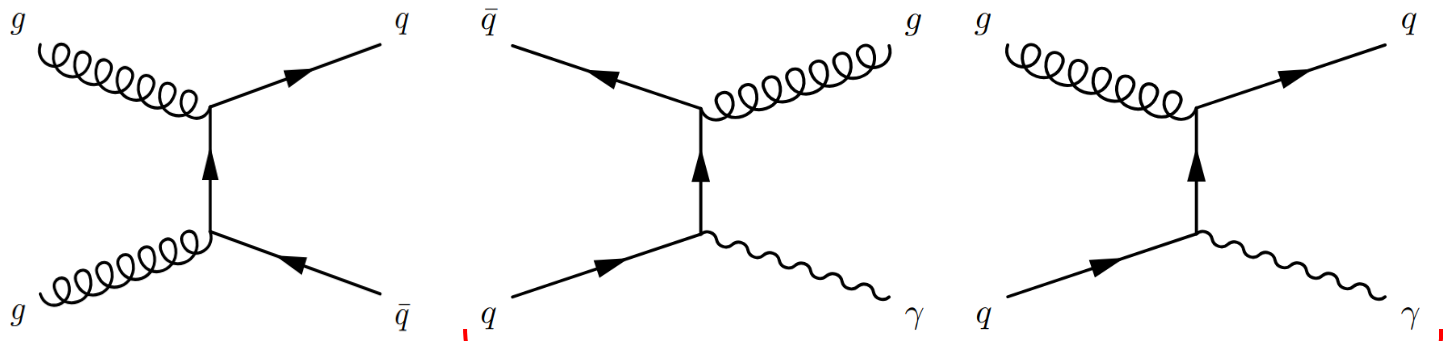


*Control region (CR): anti-id.-and-anti-iso. region
**Signal region (SR): id.-and-iso. region

Backup

BDT Training

- The BDT is trained on γ-jet and jet-jet JF samples.
- Two BDT models: one for converted γ and one for unconverted γ.
- Shower-shape variables corrected with normalizing flows are used as input variables for the training.
- p_T^γ and $|\eta^\gamma|$ are used as splitting variables for the BDT.



Jet-jet samples:
for fake-photon statistics

γ-jet samples:
for true-photon statistics