The ATLAS H→γγ analysis

Chris Hengler
DESY ATLAS group

LHC Physics Discussion, 10/06/2013

Outline

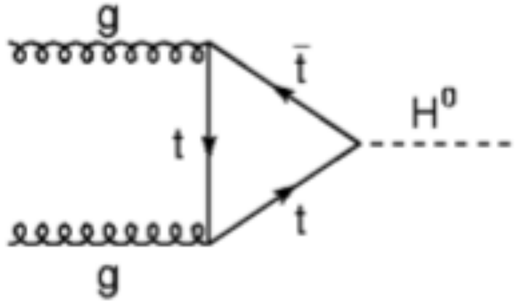
- Higgs production and decay
- ATLAS $H \rightarrow \gamma\gamma$ analysis
  - ATLAS dataset
  - Photon reconstruction & identification
  - Mass determination
  - Event categories
  - Background parameterisation
  - Results
  - Outlook

Higgs Production

- Gluon-gluon fusion

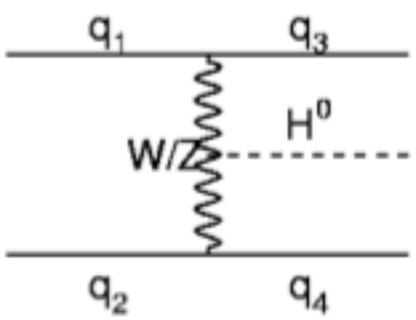
  σ = 19 pb

  lower p_T than other processes



- Vector boson fusion

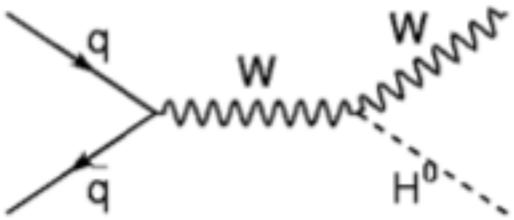
  σ = 1.5 pb



- Associated Higgs production

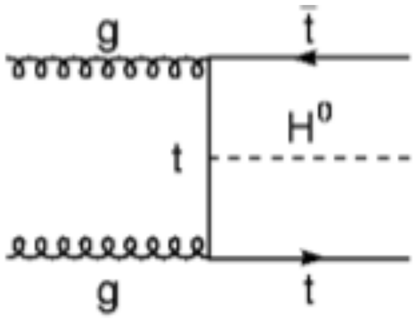
  σ = 0.7 pb (WH)

  σ = 0.4 pb (ZH)



- Higgs + tt-bar

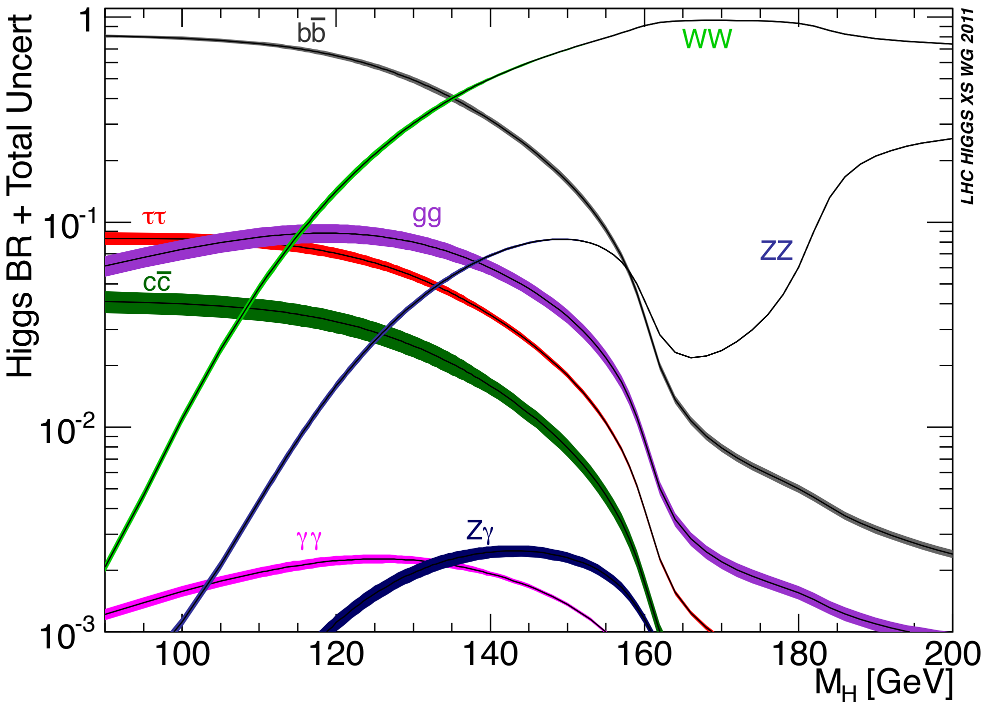
  σ = 0.12 pb

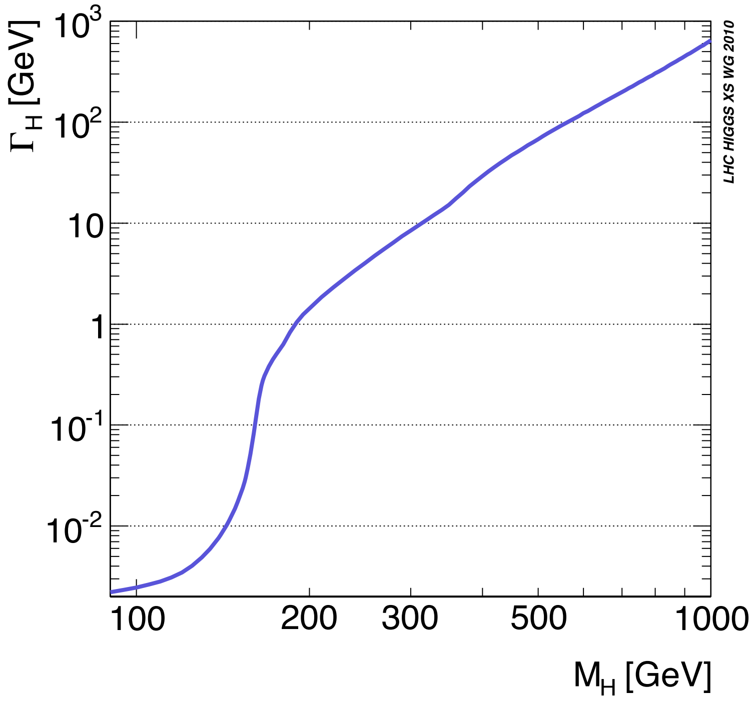


cross-sections are standard model predictions at M_H = 125.5 GeV with $\sqrt{s}$ = 8 TeV

Higgs Decay

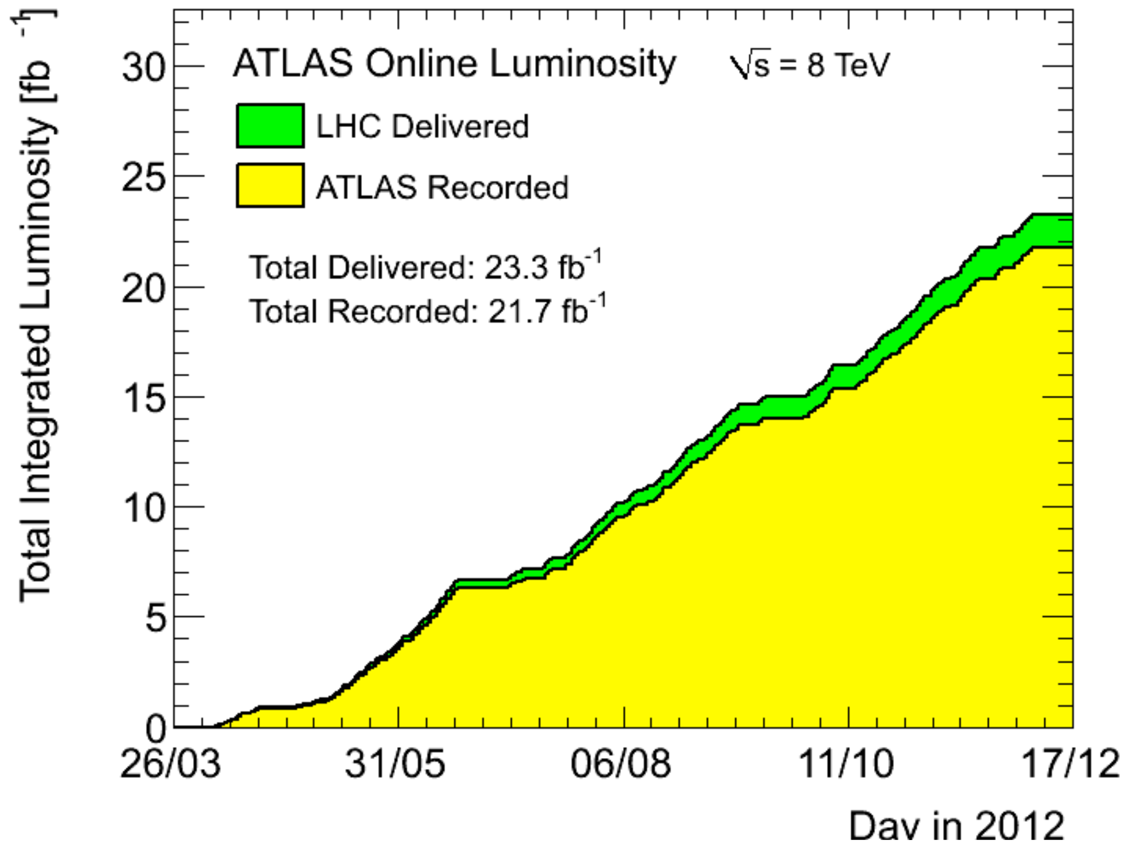
- Many decay channels, branching ratios vary with M_H
- bb-bar dominates below 130 GeV
- WW dominant above 130 GeV
- Diphoton branching ratio small (0.23%), but simple final state makes it quite powerful
- SM predicts narrow decay width (4MeV at M_H = 125 GeV)

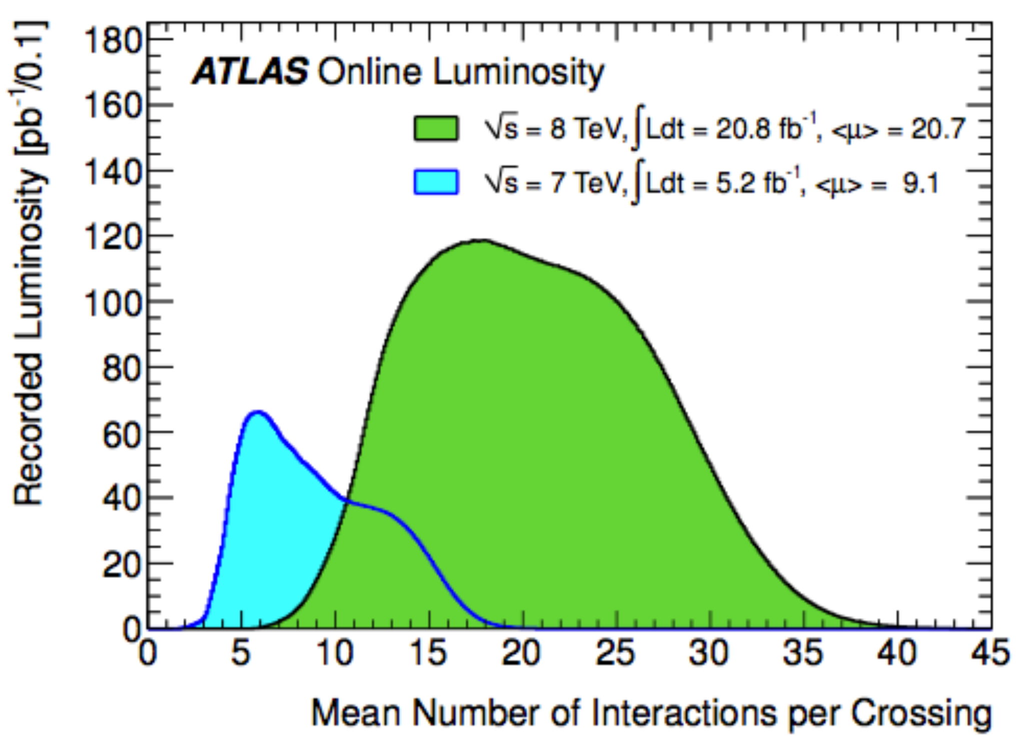


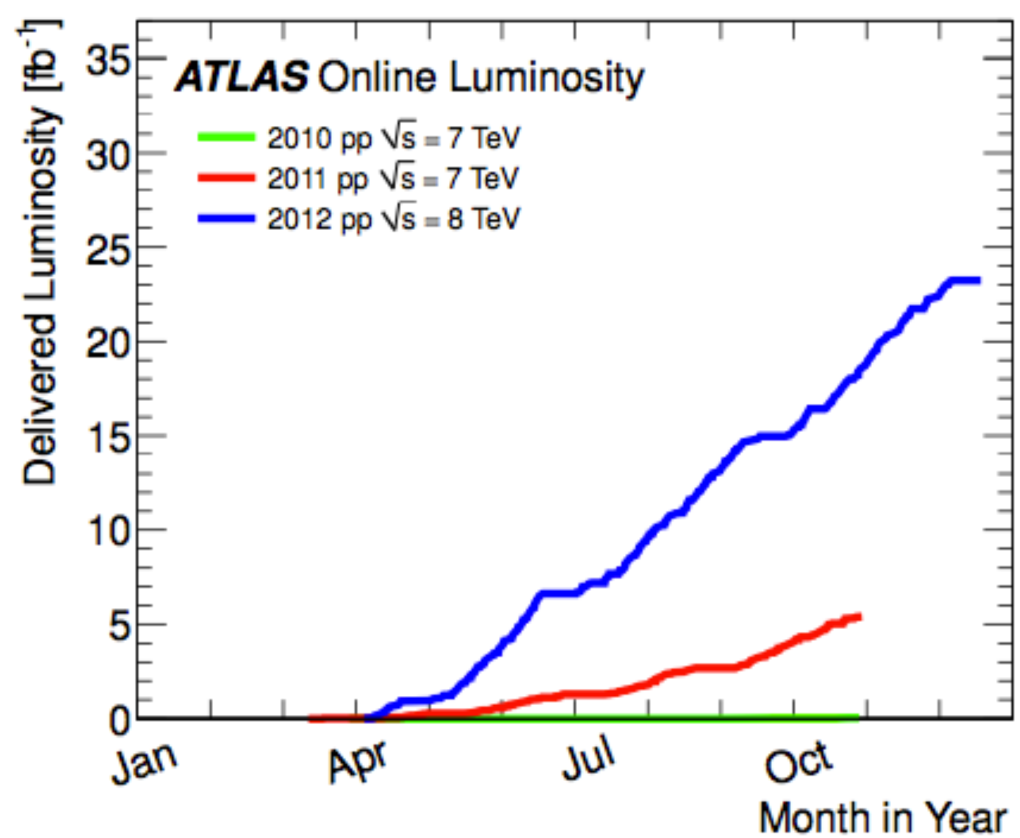


ATLAS Dataset

- LHC delivered ~23 fb^{-1} in 2012 at $\sqrt{s}$ = 8 TeV
- 21.7 fb^{-1} recorded by ATLAS
- A further ~5 fb^{-1} recorded at $\sqrt{s}$ = 7 TeV
- Instantaneous lumi 7.7 x 10^{33} $cm^{-2}s^{-1}$ (2012)
- Large pileup: challenging for trigger, reco & analysis

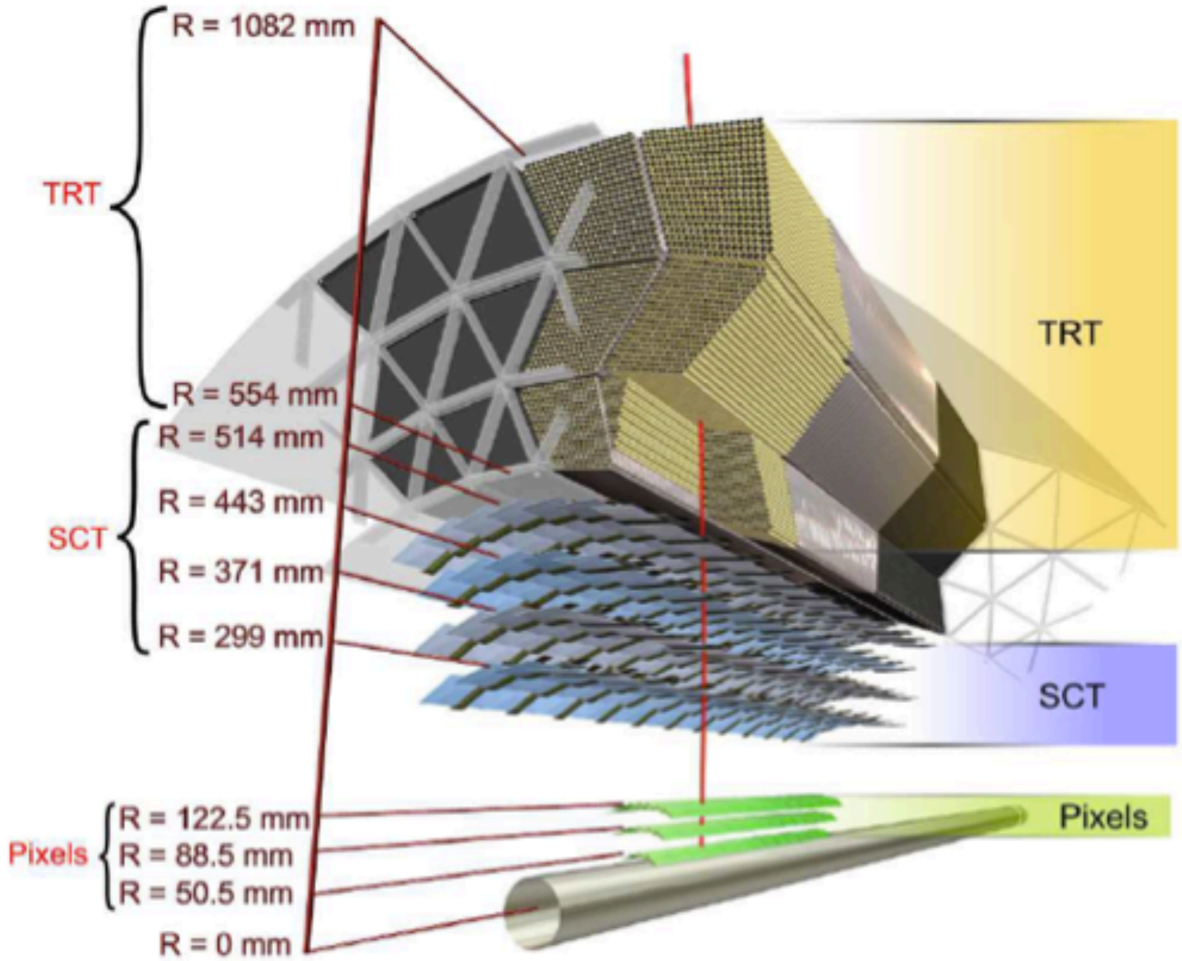






ATLAS Detector

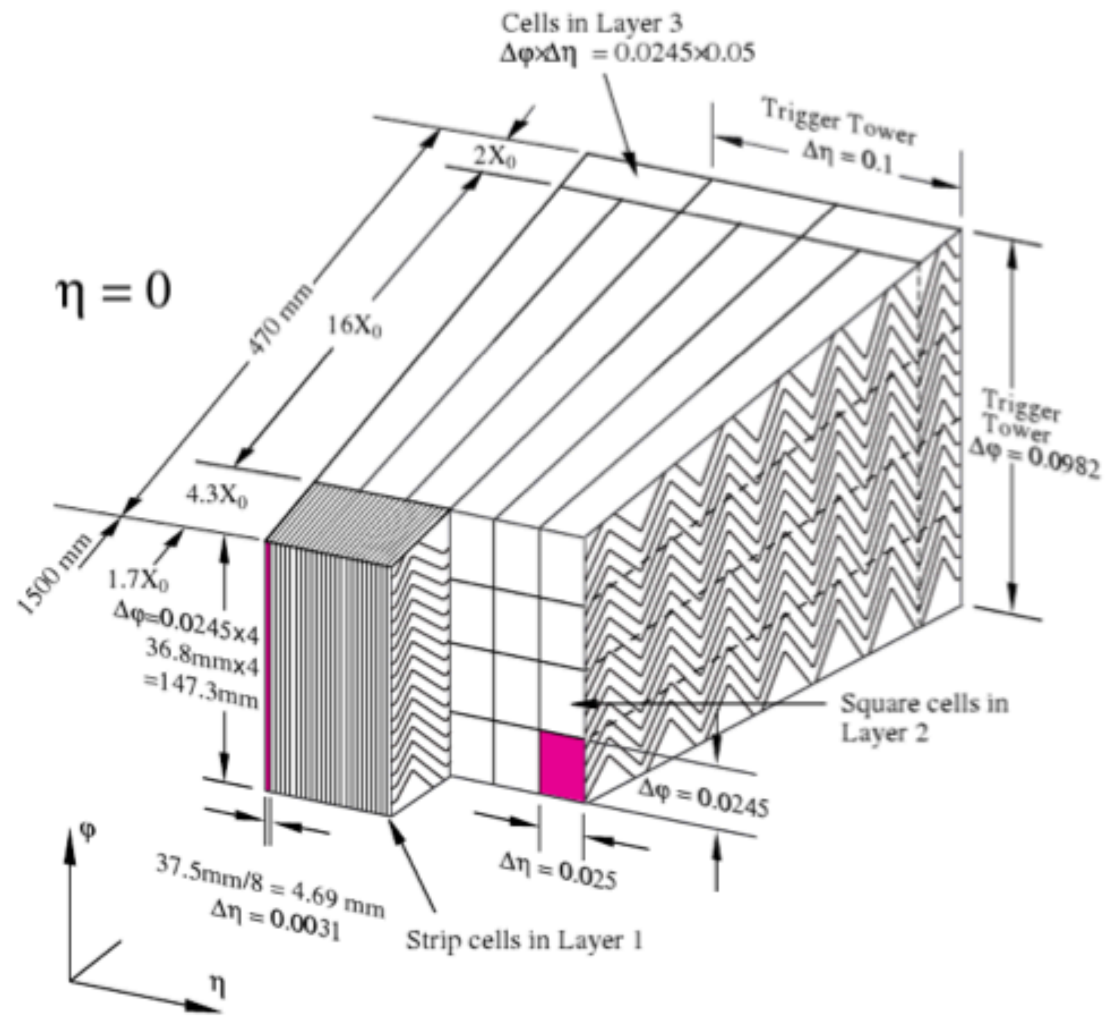
Inner Detector



- $|\eta| < 2.5$

- 3 Si pixel layers
- 4 Si strip layers
- Transition radiation tracker

EM Calorimeter



- $|\eta| < 3.2$

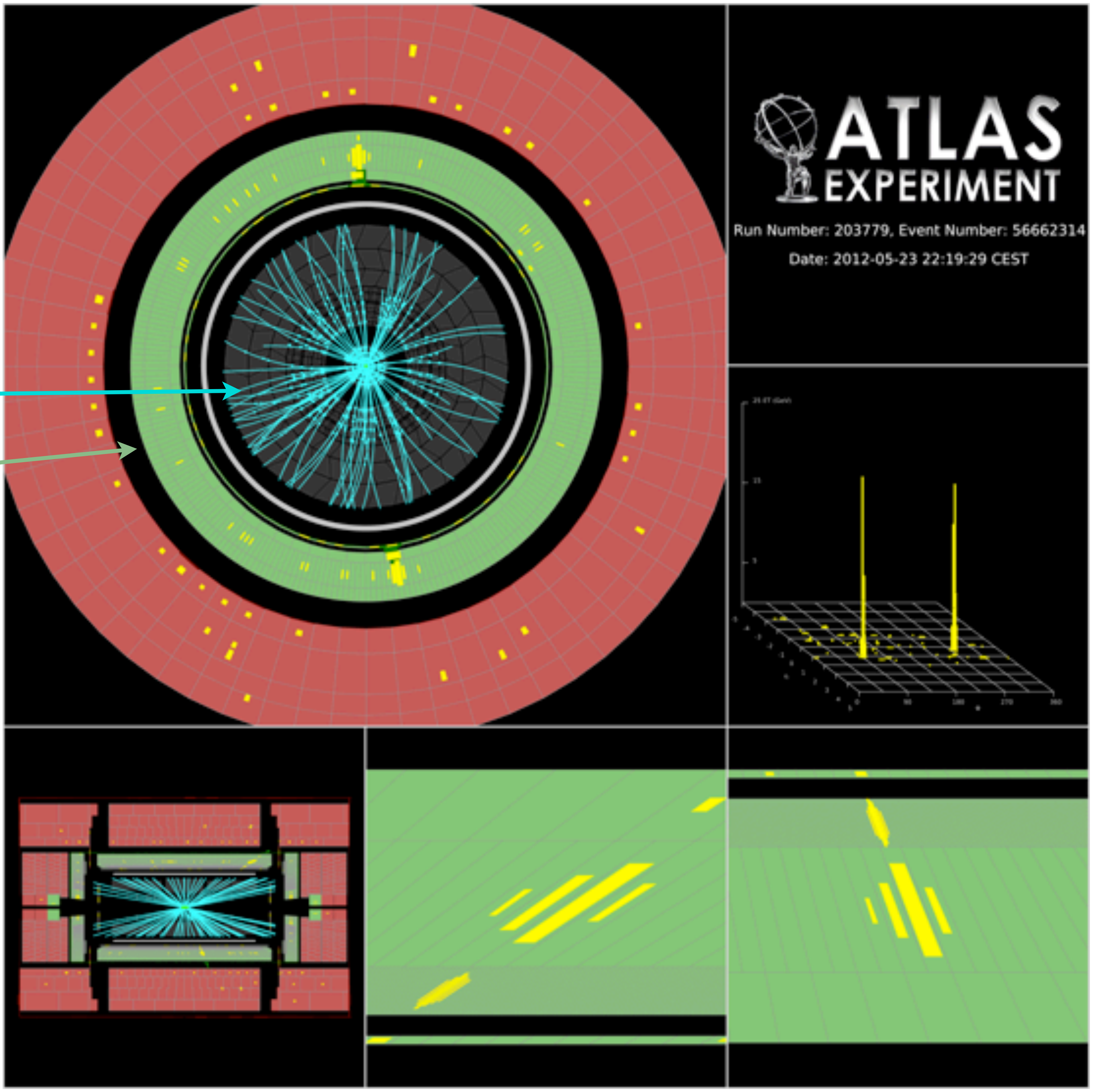
- Liquid argon calorimeter
- 3 layers
- Layer 1 strips, for high eta granularity

ATLAS H→γγ Analysis

- Relatively simple event signature:
    - 2 isolated high p_T photons
    - $m_{\gamma\gamma}^2 = 2E_1E_2(1 - \cos\alpha)$

Inner Detector:

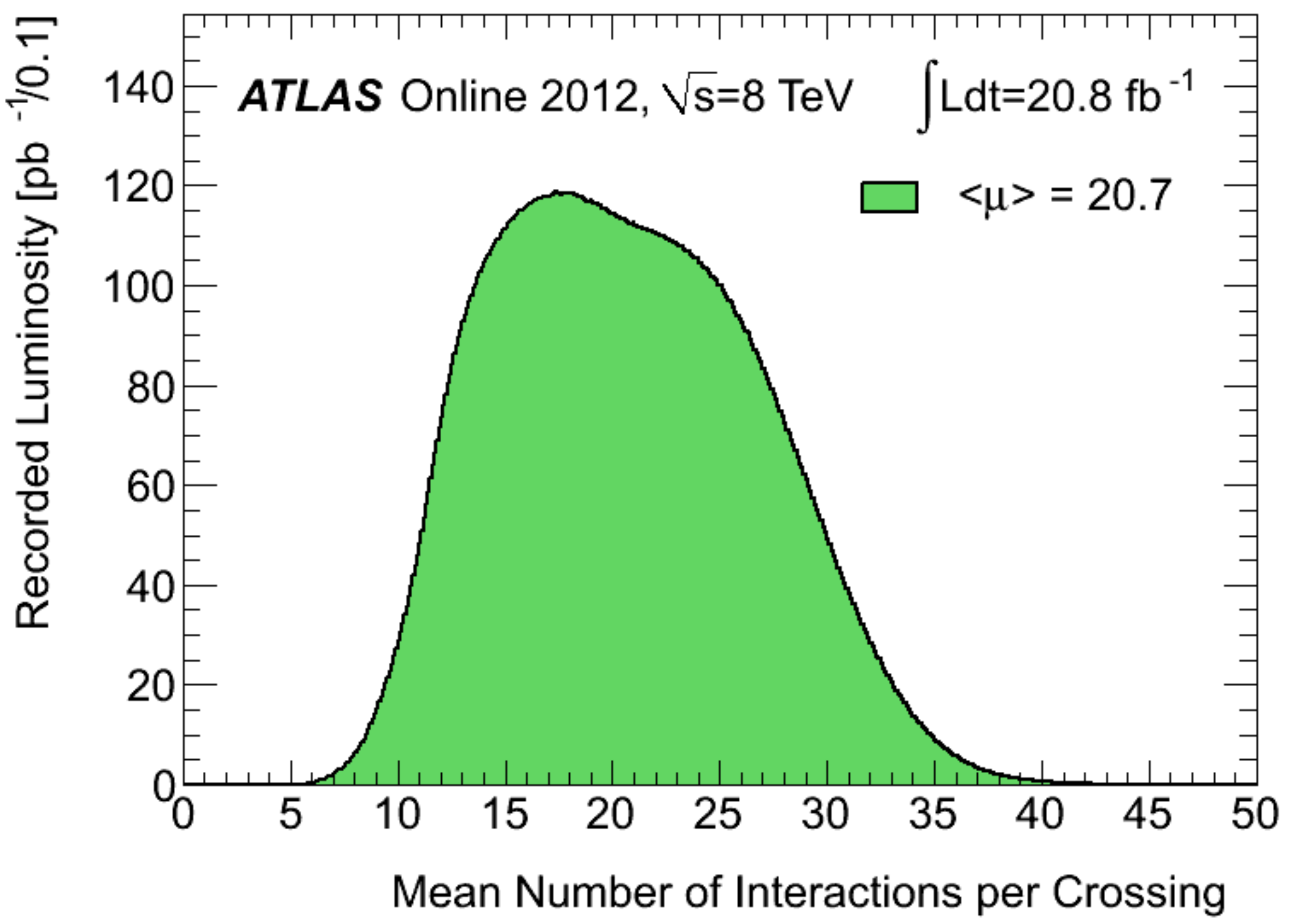
EM Calorimeter:



- Challenges
    - Large QCD background
        - Jets can fake photons
        - Good rejection is crucial
    - High Pileup

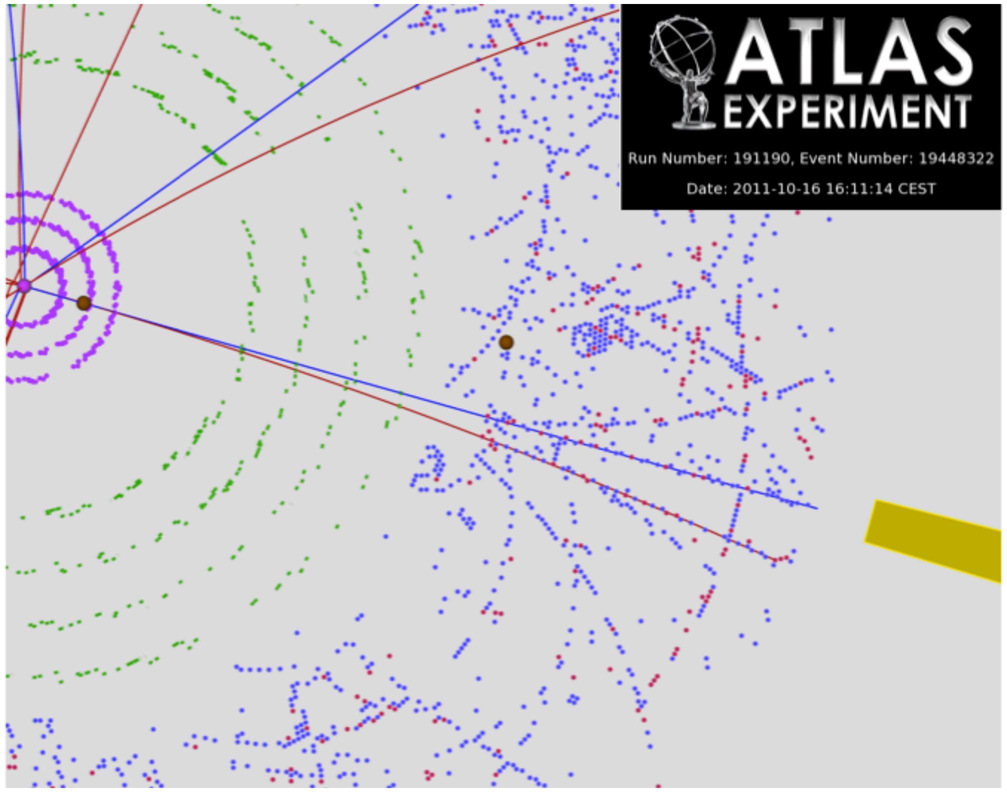
ATLAS H→γγ Analysis

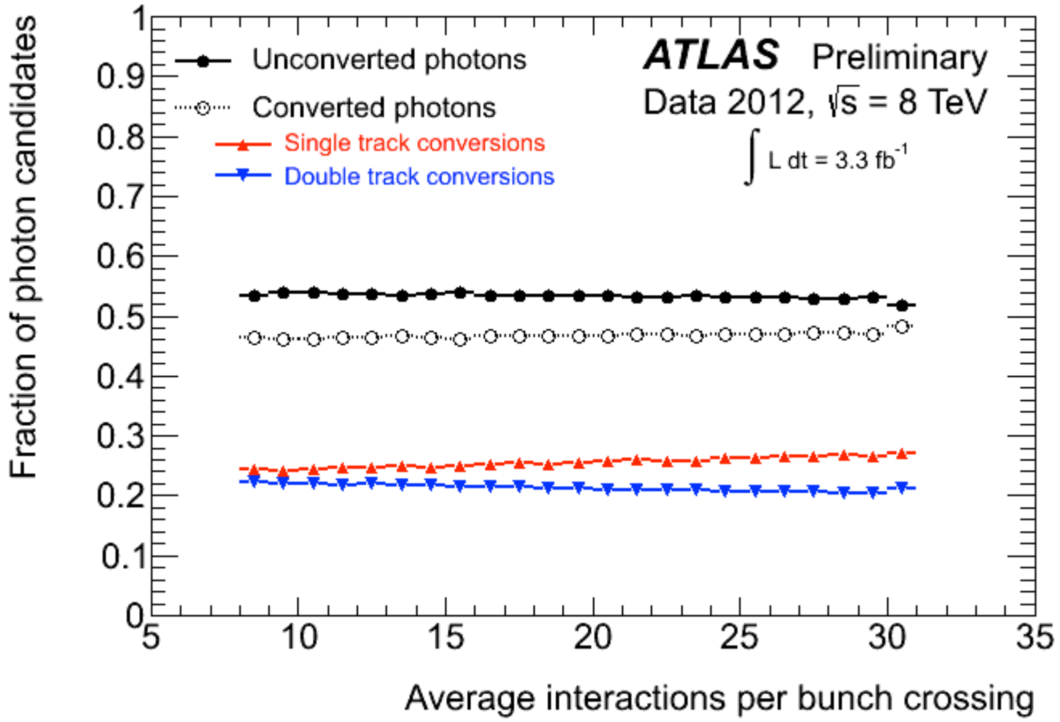
- Relatively simple event signature:
  - 2 isolated high p_T photons
  - $m_{\gamma\gamma}^2 = 2E_1E_2(1 - \cos \alpha)$
- Challenges
  - Large QCD background
    - Jets can fake photons
    - Good rejection is crucial
  - High Pileup



Photon reconstruction

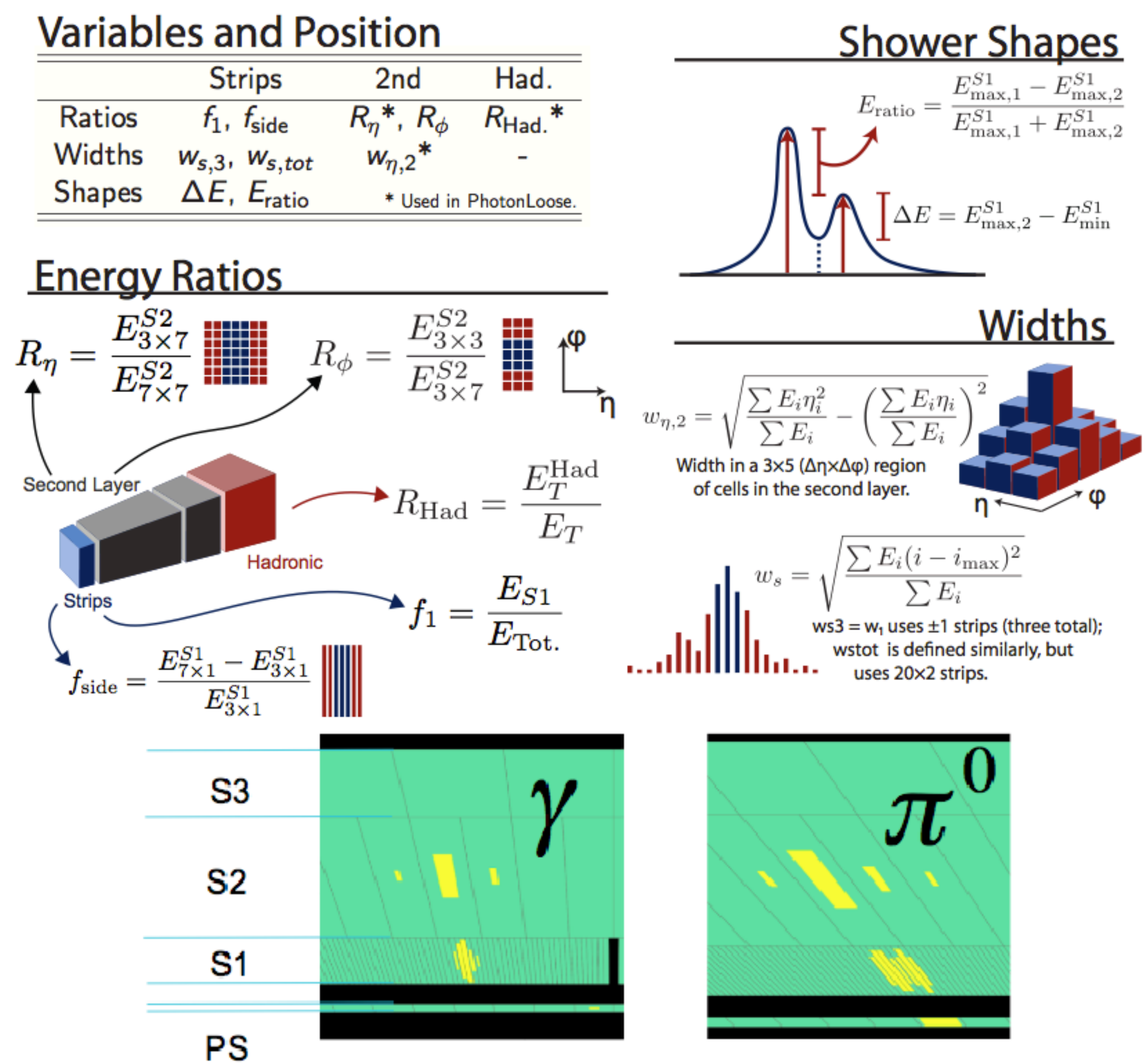
- Photons leave energy deposit in EM calorimeter
- Fixed-window algorithm searches for energy clusters in calo
    - photons may convert in Inner Detector
    - possible ambiguity with electrons: look at tracks and vertices
    - vertex matched to shower? converted photon candidate
- Photon reconstruction must be robust against pileup for both converted and unconverted





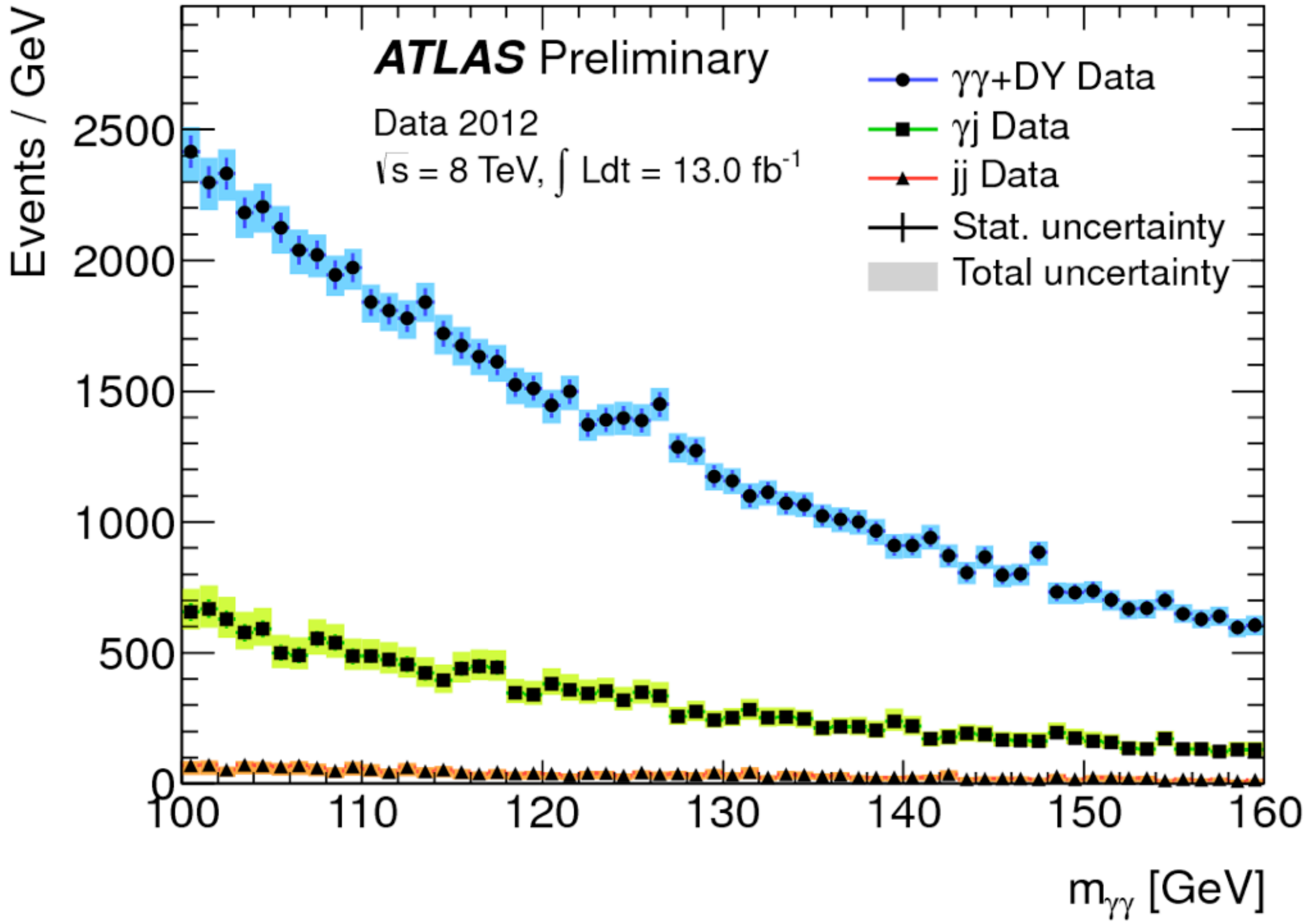
Photon identification

- Large QCD background to deal with, rejection of around 10^4 needed
- High granularity of calorimeter allows for detailed look at structure of showers ('shower shape' variables)
- Jet showers look different - e.g. typically wider, less isolated
- Devise cut-based photon selection based on MC studies



Photon identification

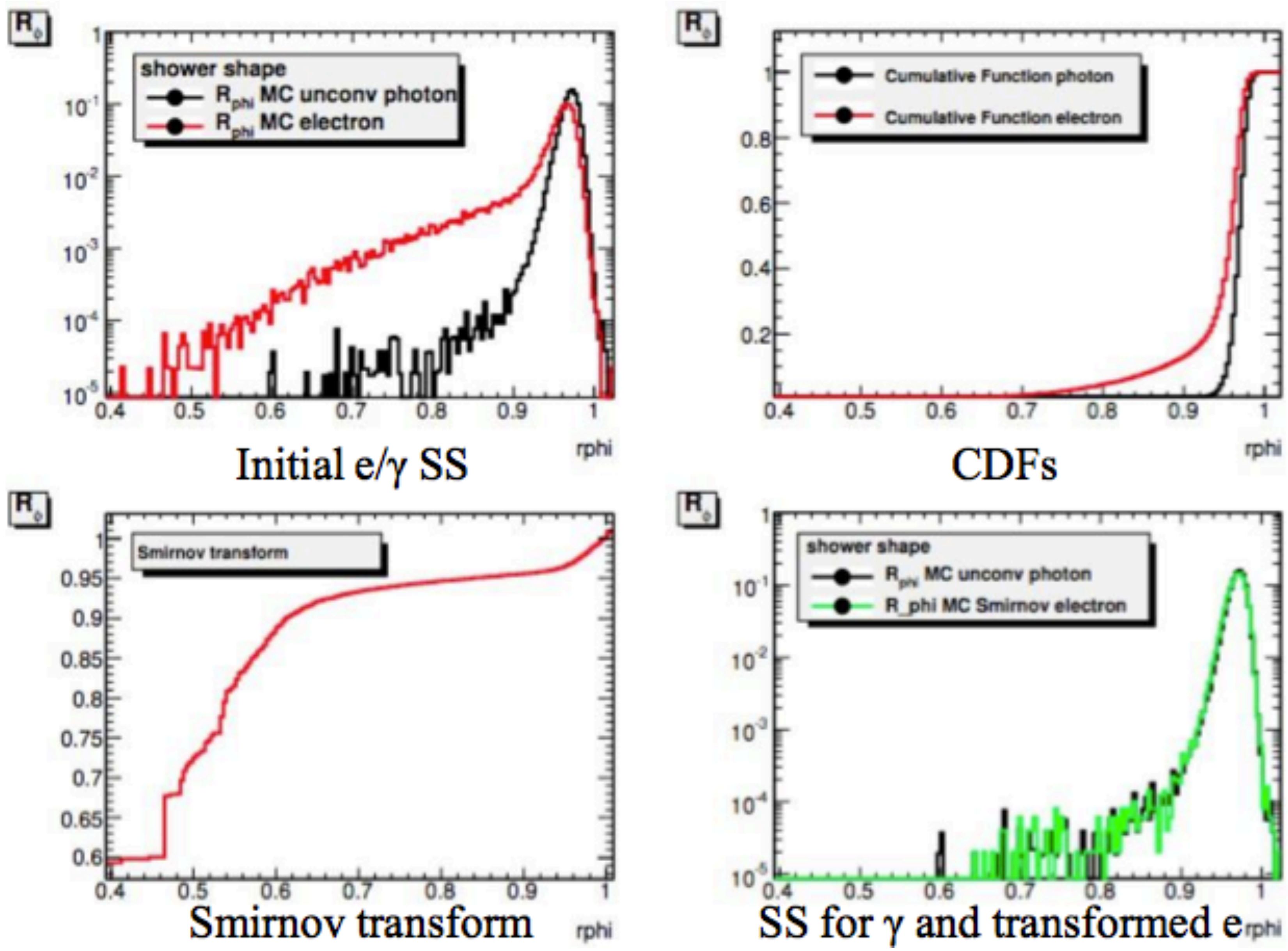
- Large QCD background to deal with, rejection of around 10^4 needed
- High granularity of calorimeter allows for detailed look at structure of showers ('shower shape' variables)
- Jet showers look different - e.g. typically wider, less isolated
- Devise cut-based photon selection based on MC studies



After photon ID, jet-jet background reduced to ~3%, ~22% gamma-jet, 75% irreducible gamma-gamma

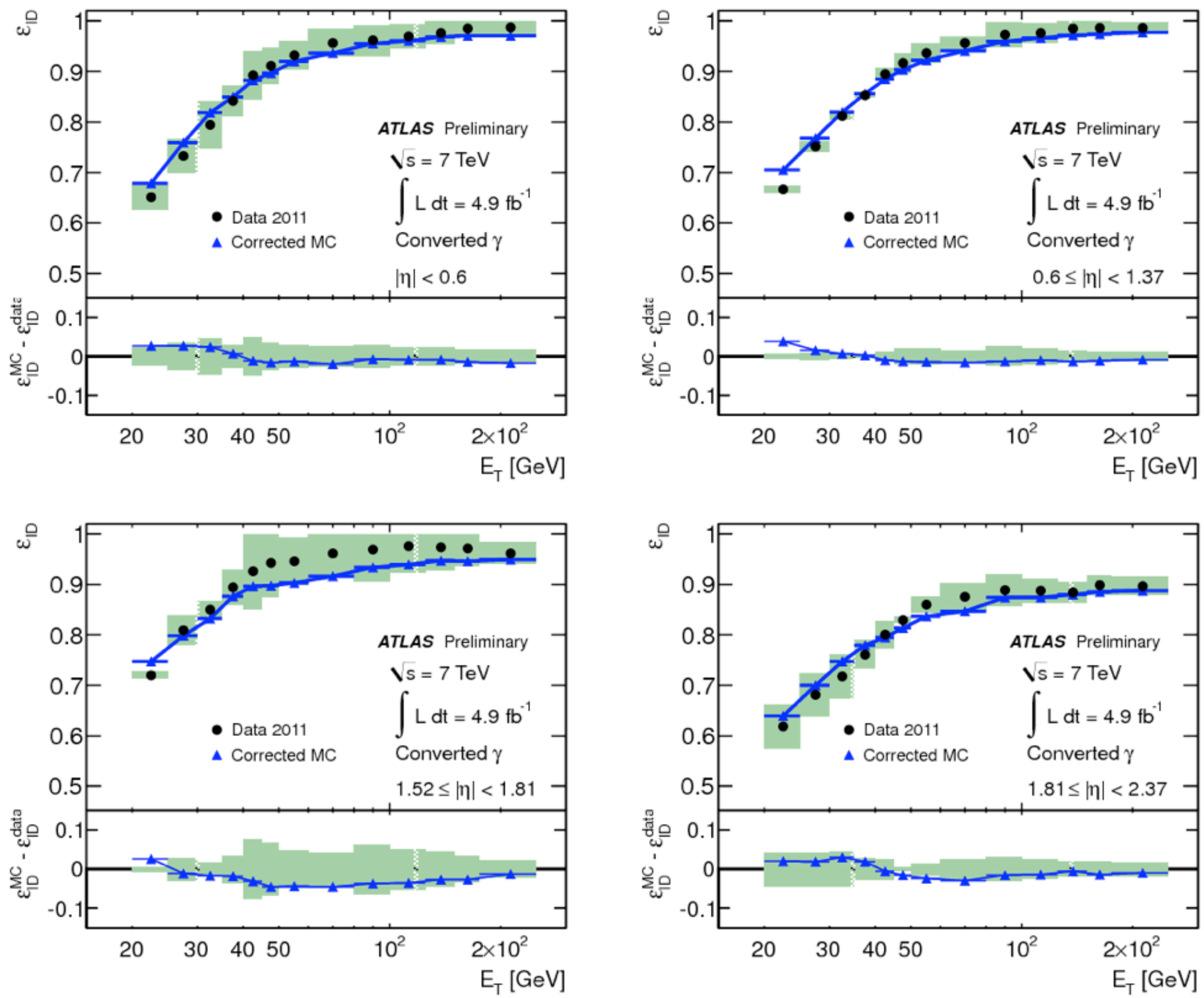
Photon ID efficiency: electron extrapolation

- One of three methods used for photon ID efficiency studies in ATLAS
- Use $Z \rightarrow e^+e^-$ events to extract electron shower shapes from data and extrapolate to photons
- Cumulative distribution frequencies taken from electron & photon shower shapes in MC
- Smirnov transform ('inverse probability transform') obtained from CDFs
- Applying transformation to data electrons gives us 'photons' which can be used for efficiency studies



Photon ID efficiency: electron extrapolation

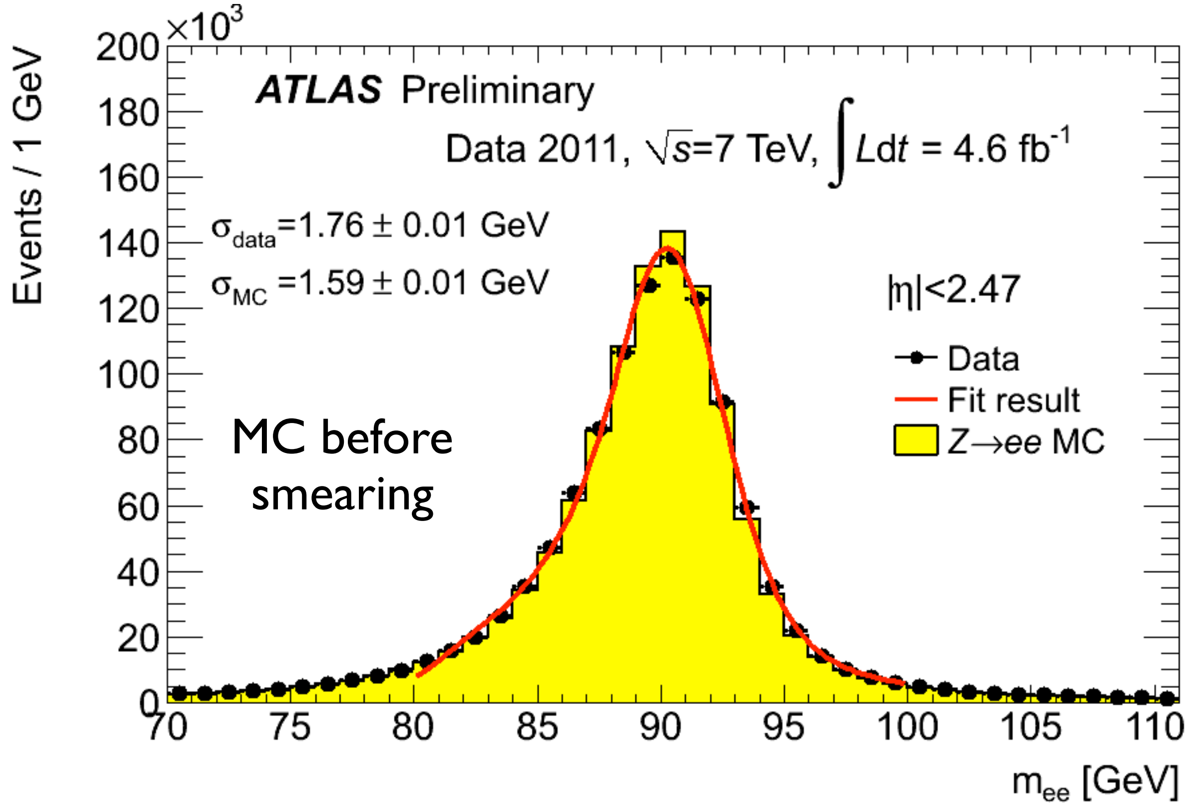
- One of three methods used for photon ID efficiency studies in ATLAS
- Use $Z \rightarrow e^+e^-$ events to extract electron shower shapes from data and extrapolate to photons
- Cumulative distribution frequencies taken from electron & photon shower shapes in MC
- Smirnov transform ('inverse probability transform') obtained from CDFs
- Applying transformation to data electrons gives us 'photons' which can be used for efficiency studies

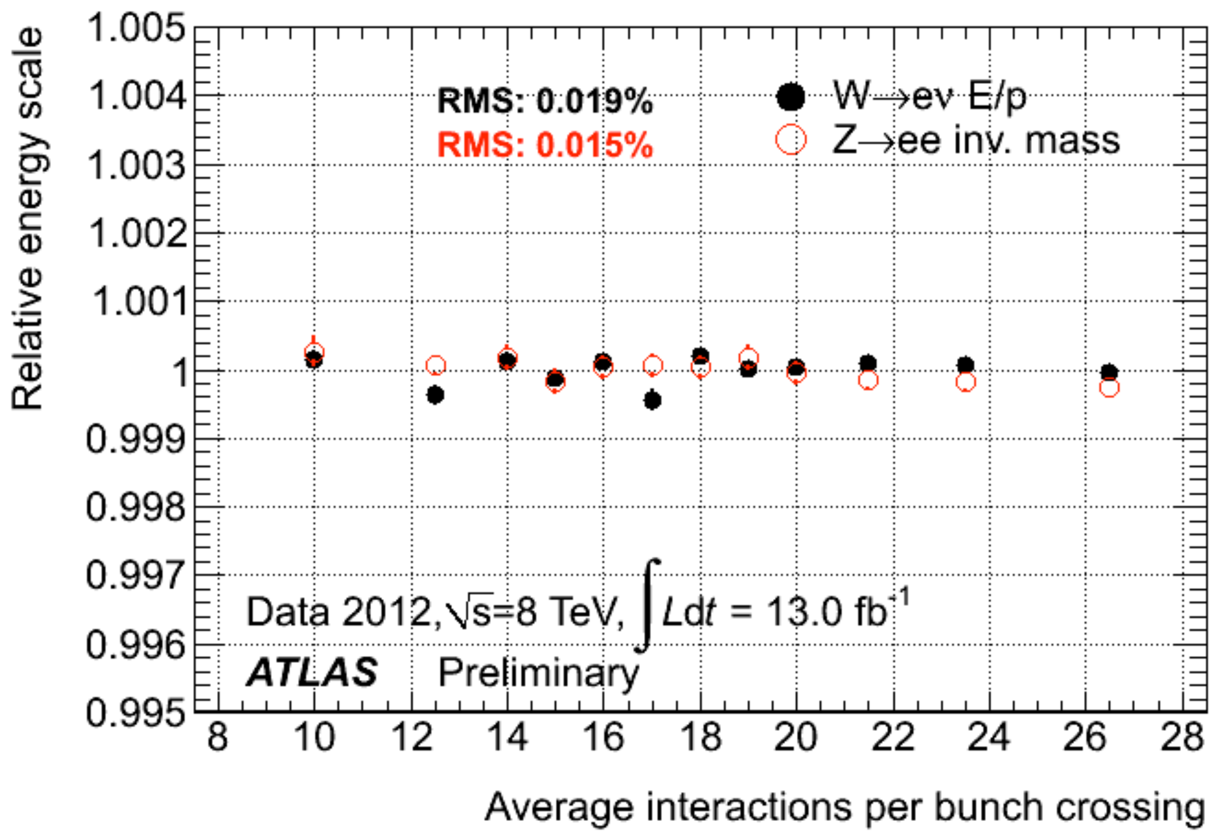


Energy calibration

$m_{\gamma\gamma}^2 = 2E_1E_2(1 - \cos \alpha)$

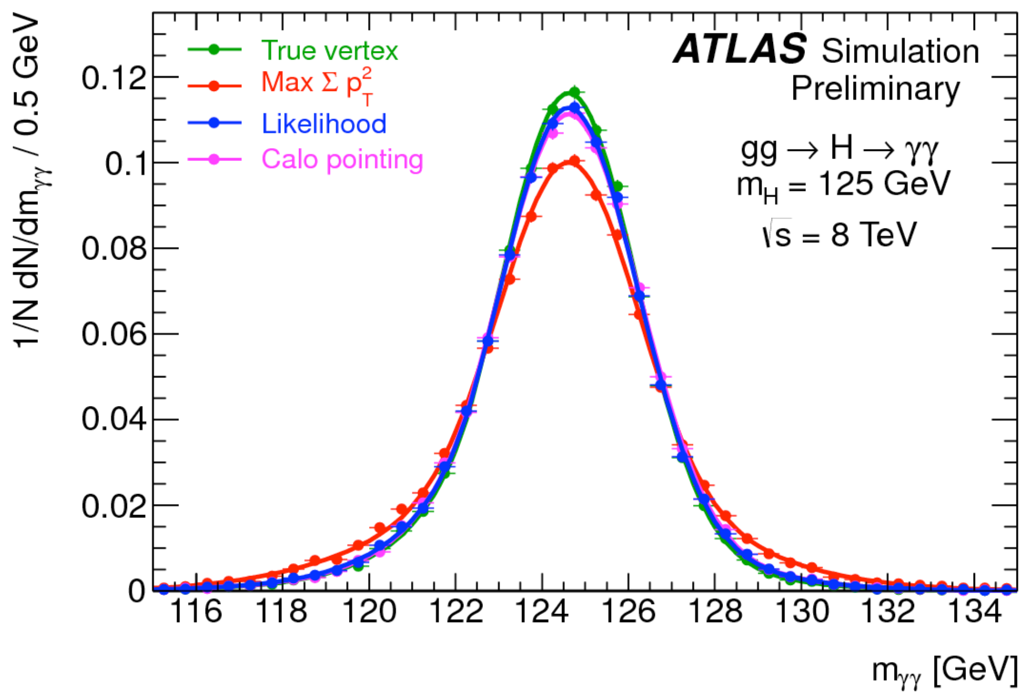
- Multiple stages of calibration:
  - calibrate energy response of calorimeter cells
  - correct for energy loss due to material before calorimeter and leakage out of cluster
  - final calibration and energy resolution extracted from $Z \rightarrow e^+e^-$ studies
- Calorimeter energy response stable with pileup to within 0.1%

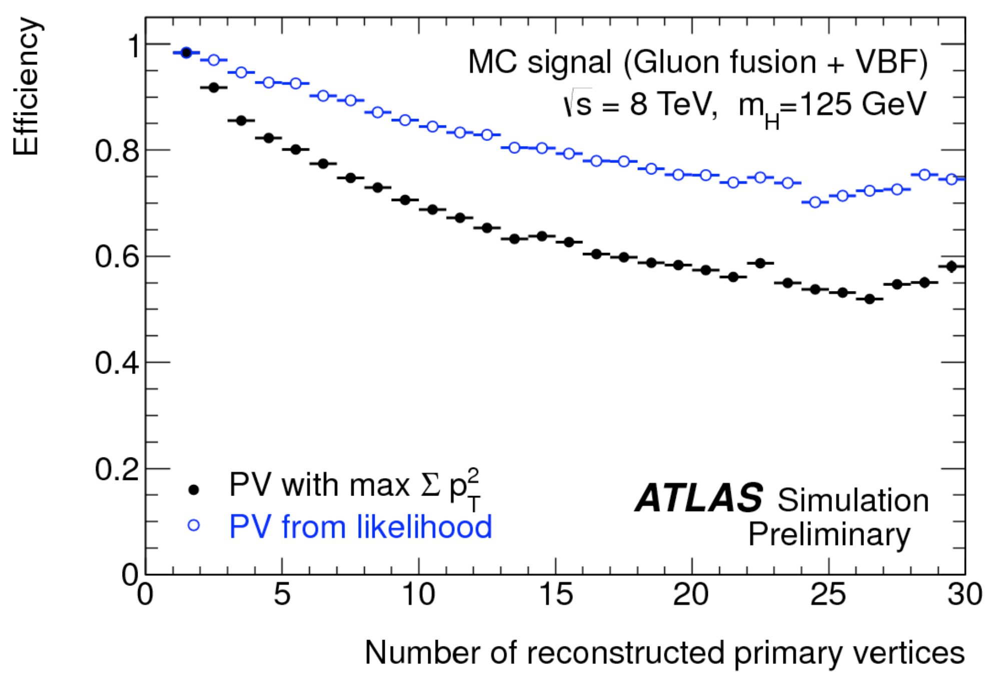




ATLAS H→γγ PV measurement

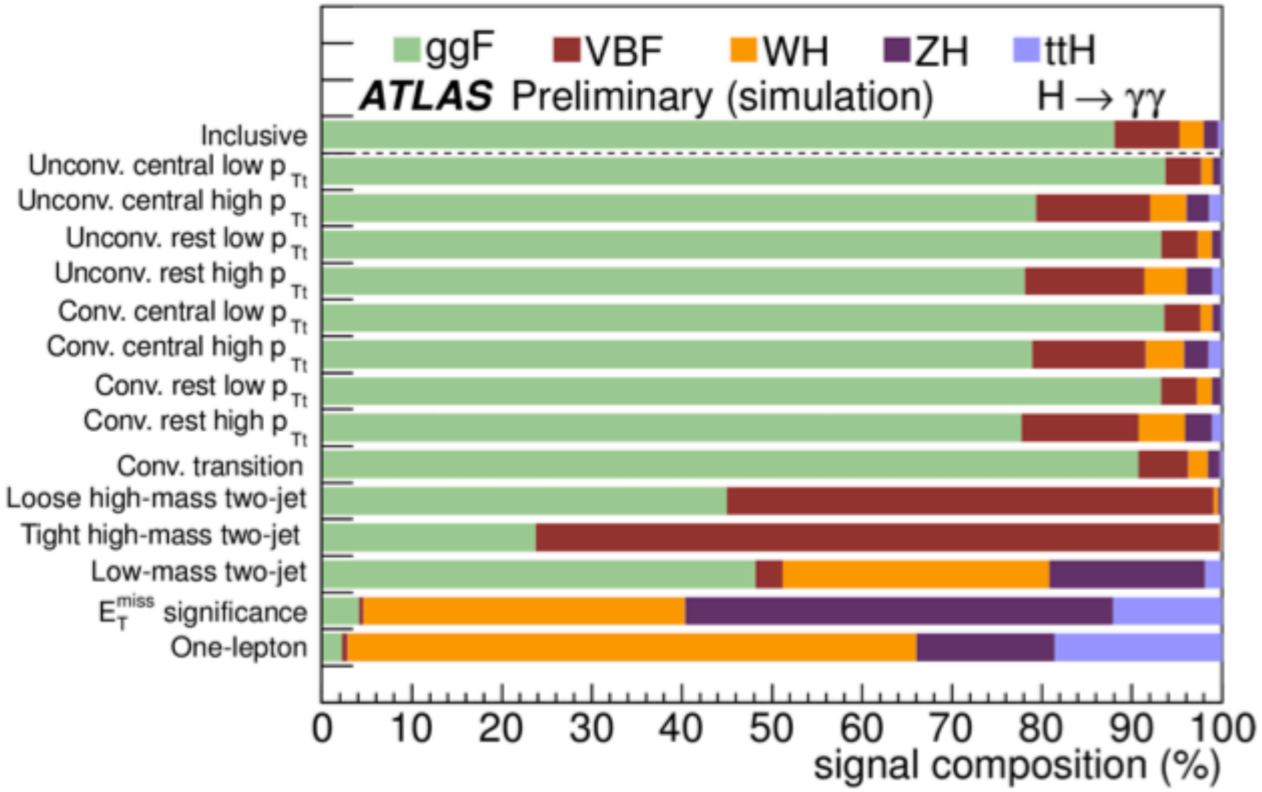
- $m_{\gamma\gamma}^2 = 2E_1E_2(1 - \cos\alpha)$
  - α = angle between photons
  - *Primary vertex reconstruction hugely important*
  - Determined for 8 TeV by Neural Network using several inputs
    - calorimeter pointing
    - conversion vertex information
    - $\Sigma\, p_T^2$
    - scalar $\Sigma\, p_T$
  - Studied in MC:
    - Calo pointing alone provides good estimate
    - Efficiency >80% of selecting vertex with Δz < 0.2mm from hard interaction vertex

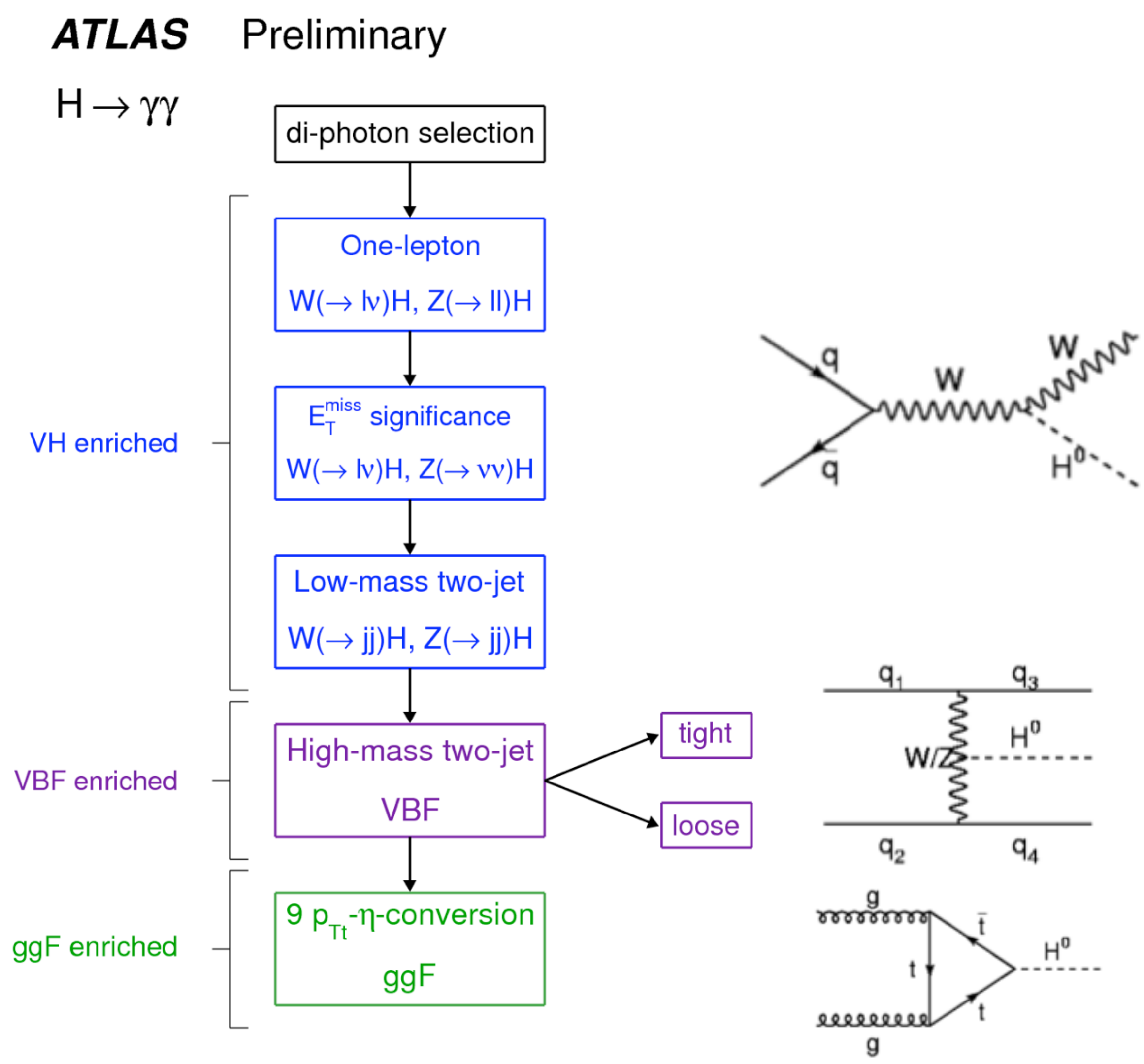




Signal categories

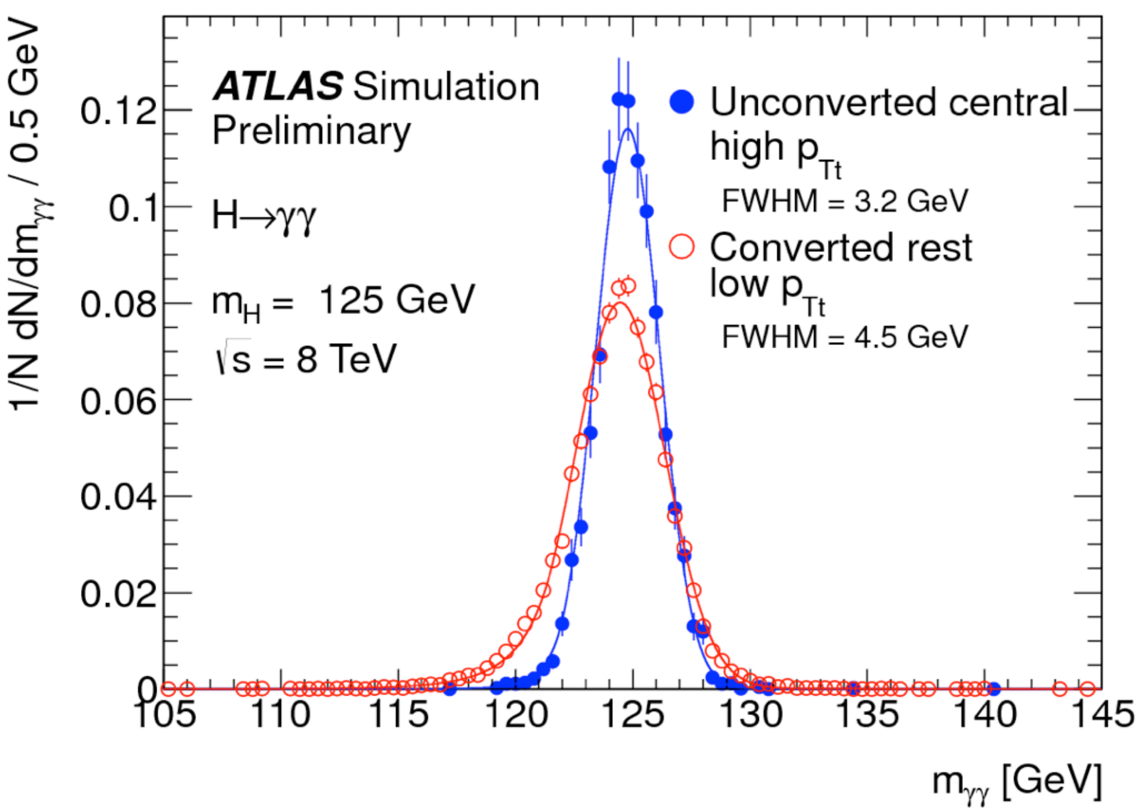
- Candidate events are categorised to select samples enriched with various production modes
- Mass resolution varies as function of p_T, η, and photon conversion status
- Remaining events are further categorised into regions of signal resolution and signal/background

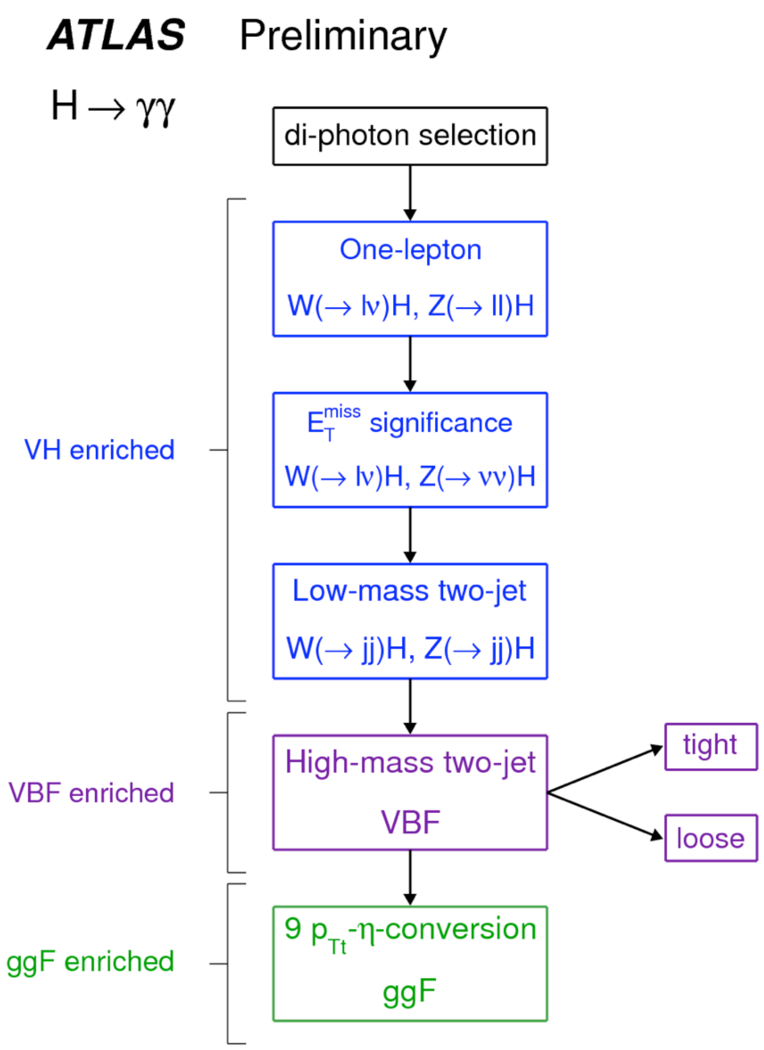


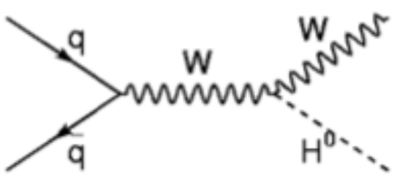


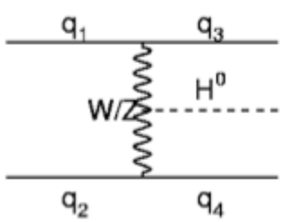
Signal categories

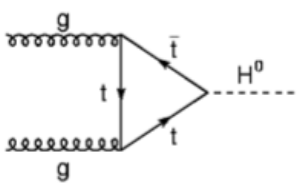
- Candidate events are categorised to select samples enriched with various production modes
- Mass resolution varies as function of p_T, η, and photon conversion status
- Remaining events are further categorised into regions of signal resolution and signal/background







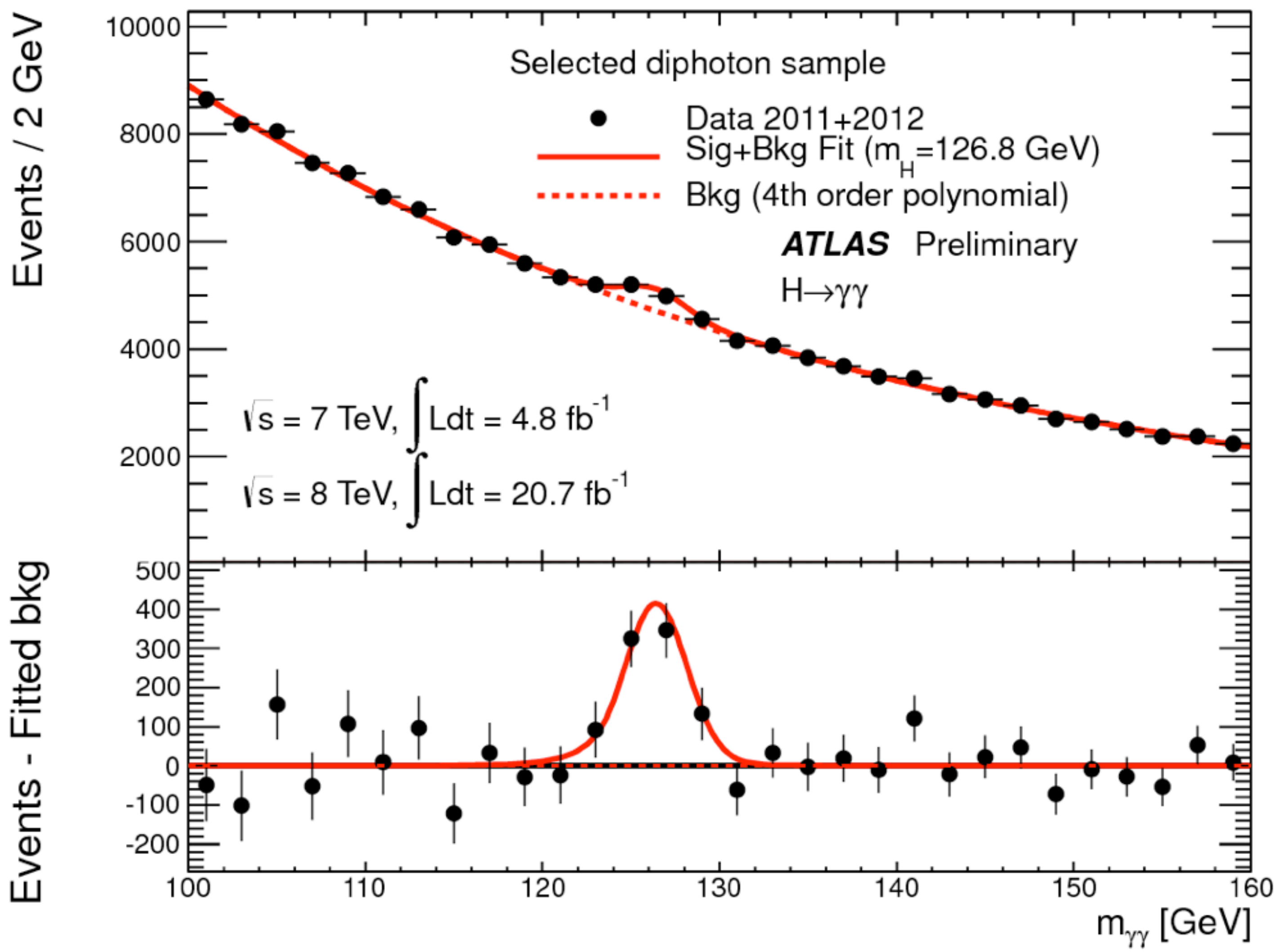




Mass spectrum and background parameterisation

- Signal modelled as crystal ball + gaussian
- Background as 4th order Bernstein polynomial, exponential of second order polynomial, or exponential, depending on category

Category	Parametrisation	Uncertainty [N_{evt}] $\sqrt{s}$ = 7 TeV	Uncertainty [N_{evt}] $\sqrt{s}$ = 8 TeV
Inclusive	4th order pol.	7.3	12.0
Unconverted central, low p_{Tt}	Exp. of 2nd order pol.	2.1	4.6
Unconverted central, high p_{Tt}	Exponential	0.2	0.8
Unconverted rest, low p_{Tt}	4th order pol.	2.2	11.4
Unconverted rest, high p_{Tt}	Exponential	0.5	2.0
Converted central, low p_{Tt}	Exp. of 2nd order pol.	1.6	2.4
Converted central, high p_{Tt}	Exponential	0.3	0.8
Converted rest, low p_{Tt}	4th order pol.	4.6	8.0
Converted rest, high p_{Tt}	Exponential	0.5	1.1
Converted transition	Exp. of 2nd order pol.	3.2	9.1
Loose high-mass two-jet	Exponential	0.4	1.1
Tight high-mass two-jet	Exponential	-	0.3
Low-mass two-jet	Exponential	-	0.6
E_T^{miss} significance	Exponential	-	0.1
One-lepton	Exponential	-	0.3



Results: All categories

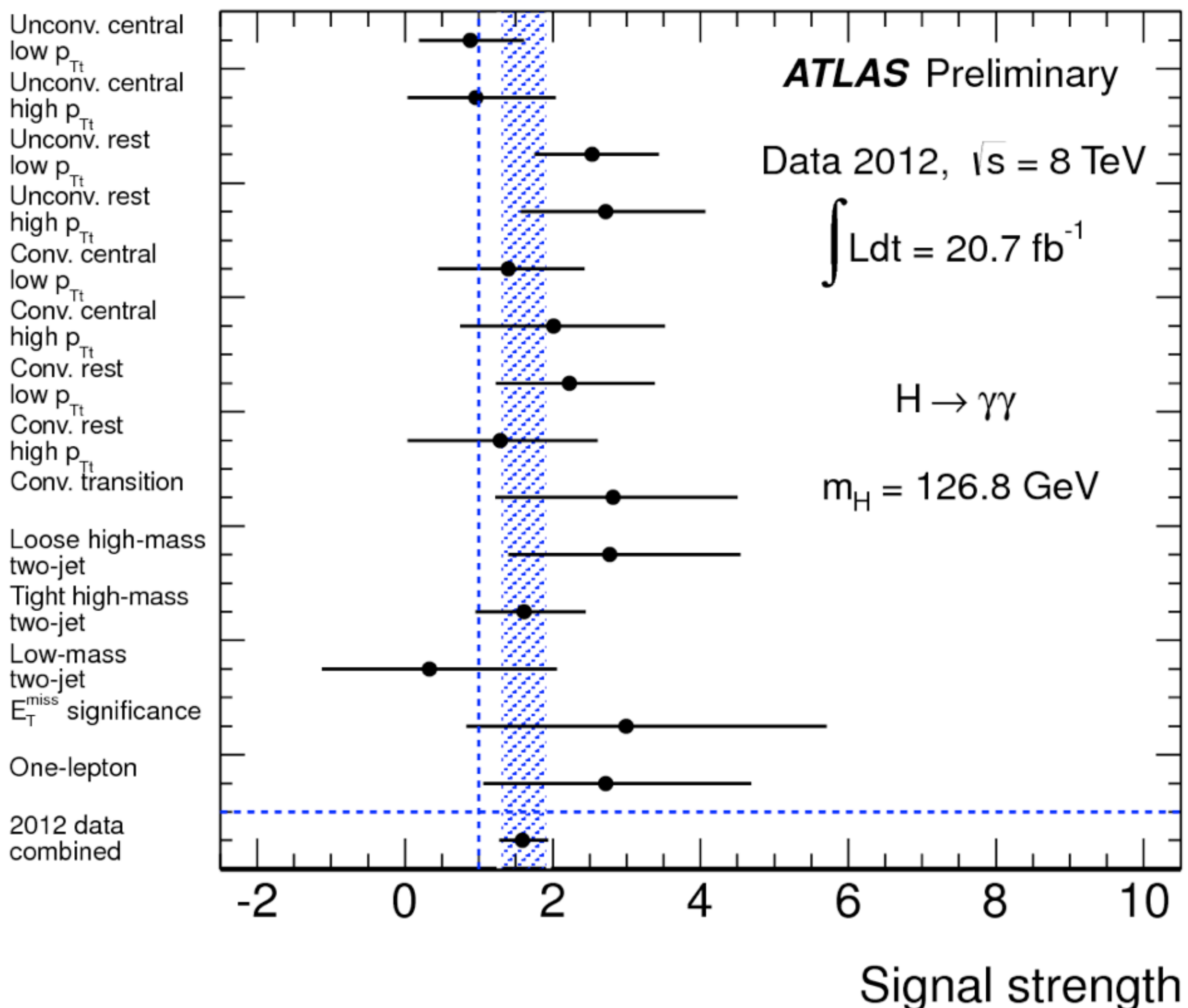
$\sqrt{s}$	8 TeV				
Category	σ_{CB}(GeV)	Observed	N_S	N_B	N_S/N_B
Unconv. central, low p_{Tt}	1.50	911	46.6	881	0.05
Unconv. central, high p_{Tt}	1.40	49	7.1	44	0.16
Unconv. rest, low p_{Tt}	1.74	4611	97.1	4347	0.02
Unconv. rest, high p_{Tt}	1.69	292	14.4	247	0.06
Conv. central, low p_{Tt}	1.68	722	29.8	687	0.04
Conv. central, high p_{Tt}	1.54	39	4.6	31	0.15
Conv. rest, low p_{Tt}	2.01	4865	88.0	4657	0.02
Conv. rest, high p_{Tt}	1.87	276	12.9	266	0.05
Conv. transition	2.52	2554	36.1	2499	0.01
Loose High-mass two-jet	1.71	40	4.8	28	0.17
Tight High-mass two-jet	1.64	24	7.3	13	0.57
Low-mass two-jet	1.62	21	3.0	21	0.14
E_T^{miss} significance	1.74	8	1.1	4	0.24
One-lepton	1.75	19	2.6	12	0.20
Inclusive	1.77	14025	355.5	13280	0.03

σ_{CB} : signal mass resolution
N_S : number of expected signal events
N_B : number of expected background events

Expected numbers based on MC with M_H = 126.5 GeV, in window containing 90% of signal events

Best S/B expected in VBF enriched category (tight high-mass two-jet)

Signal strength defined as ratio between observed events and expected events from SM Higgs



$$\mu = 1.65^{+0.34}_{-0.30}$$

(combination of all categories)

Results:
All categories

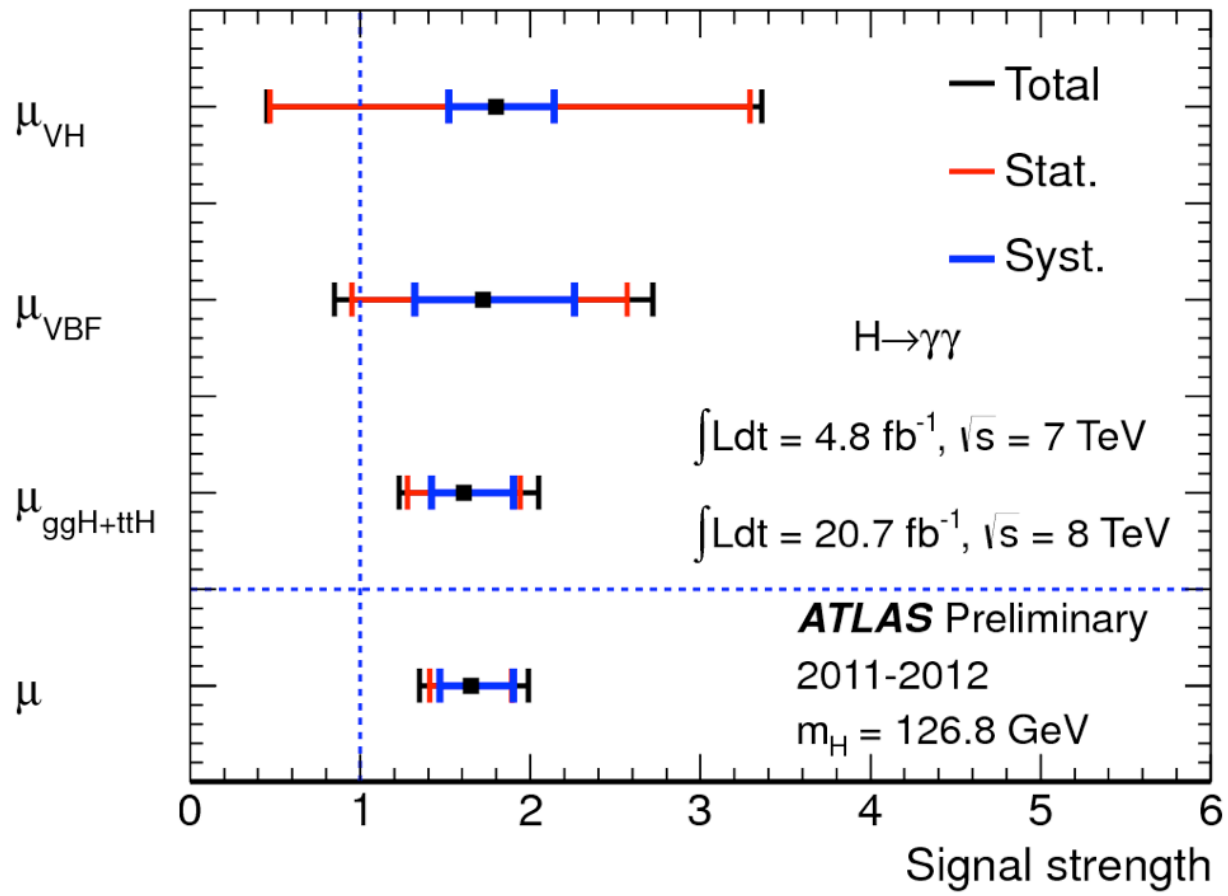
$\sqrt{s}$	8 TeV				
Category	σ_{CB}(GeV)	Observed	N_S	N_B	N_S/N_B
Unconv. central, low p_{Tt}	1.50	911	46.6	881	0.05
Unconv. central, high p_{Tt}	1.40	49	7.1	44	0.16
Unconv. rest, low p_{Tt}	1.74	4611	97.1	4347	0.02
Unconv. rest, high p_{Tt}	1.69	292	14.4	247	0.06
Conv. central, low p_{Tt}	1.68	722	29.8	687	0.04
Conv. central, high p_{Tt}	1.54	39	4.6	31	0.15
Conv. rest, low p_{Tt}	2.01	4865	88.0	4657	0.02
Conv. rest, high p_{Tt}	1.87	276	12.9	266	0.05
Conv. transition	2.52	2554	36.1	2499	0.01
Loose High-mass two-jet	1.71	40	4.8	28	0.17
Tight High-mass two-jet	1.64	24	7.3	13	0.57
Low-mass two-jet	1.62	21	3.0	21	0.14
E_T^{miss} significance	1.74	8	1.1	4	0.24
One-lepton	1.75	19	2.6	12	0.20
Inclusive	1.77	14025	355.5	13280	0.03

σ_{CB} : signal mass resolution
N_S : number of expected signal events
N_B : number of expected background events

Expected numbers based on MC with M_H = 126.5 GeV, in window containing 90% of signal events

Best S/B expected in VBF enriched category (tight high-mass two-jet)

Signal strength defined as ratio between observed events and expected events from SM Higgs

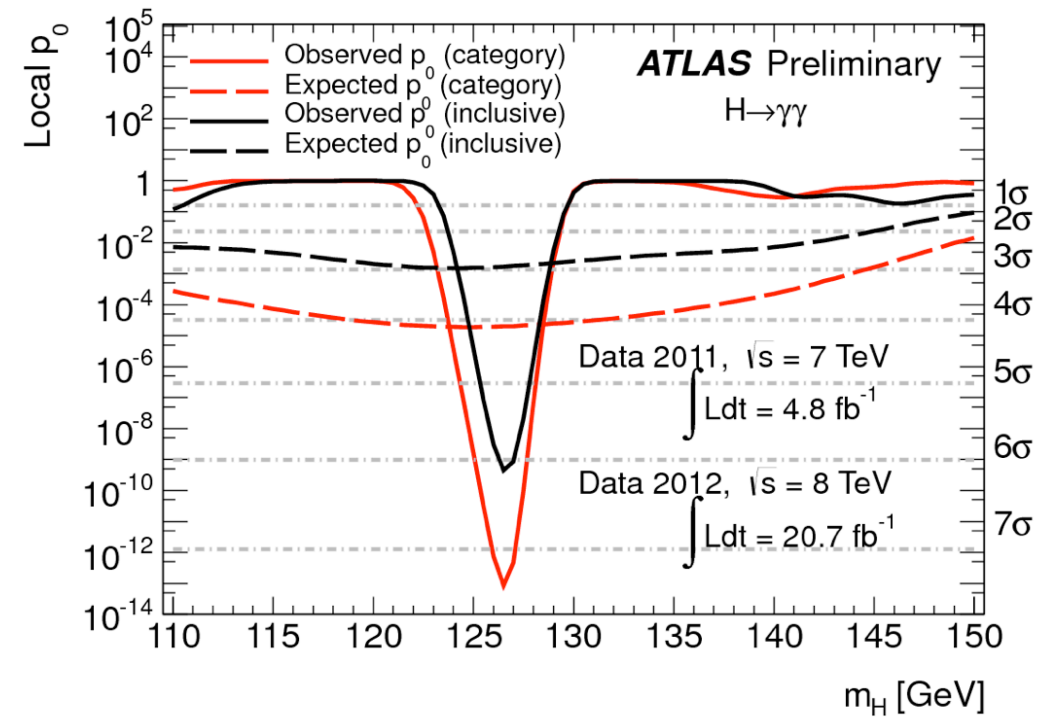


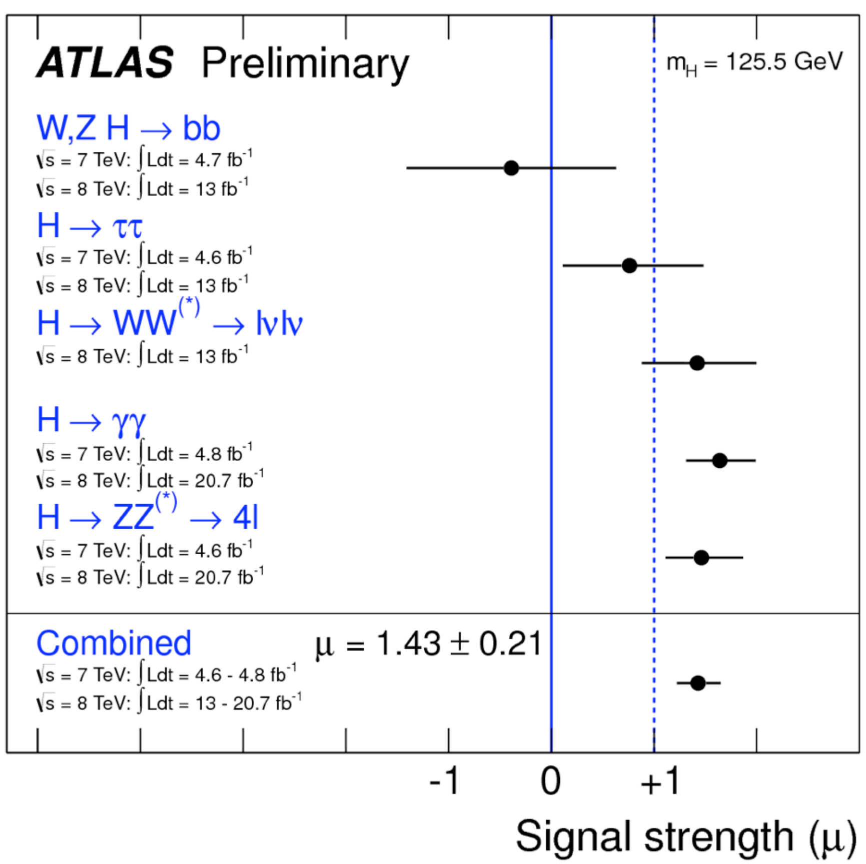
Sources of uncertainty

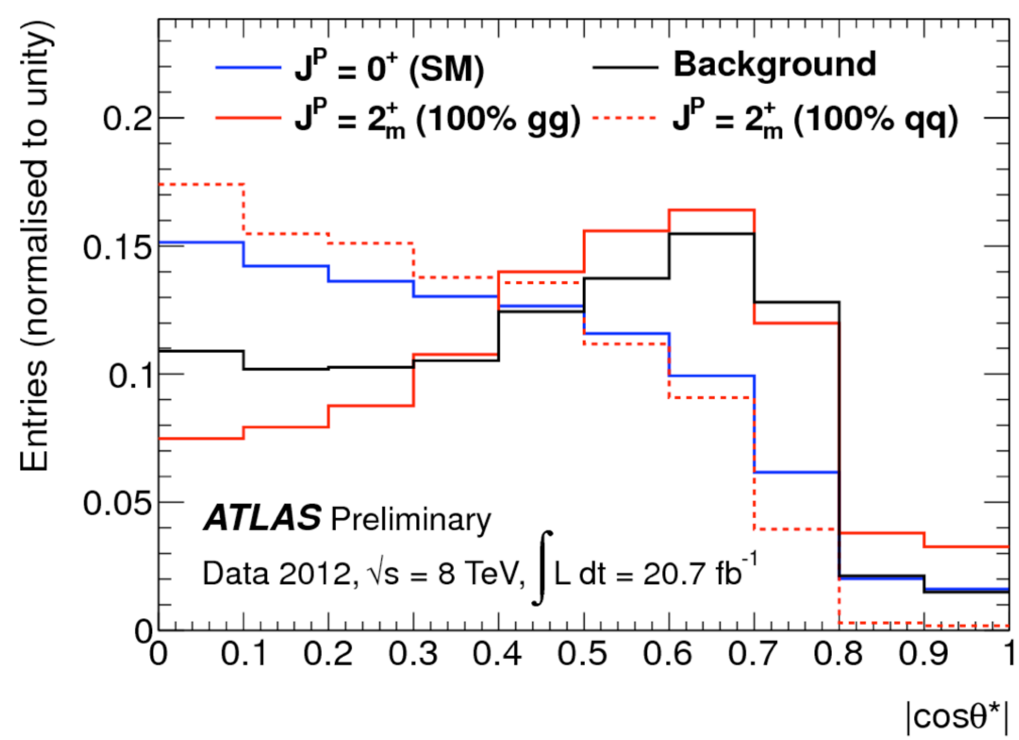
	7 TeV	8 TeV
Photon id efficiency	8.4%	2.4%
Luminosity	1.8%	3.6%
Theory	up to 8%	
	each scale and PDF+α_s	
	up to 25%	up to 48%
	($gg \to H + 2$ jets)	
Jet E-scale (2-jets)	4-20%	
Underl. evt. (2-jets)	6-30%	2-13%
Higgs p_T	up to 12.5%	
Dijet modeling		up to 12%
Bkgd Param (evts)	0.2-4.6	0.1-11.4
$m_{\gamma\gamma}$ resolution	14%	14-23%
mass scale	0.6%	0.55%

Summary & Outlook

- H→γγ channel alone able to claim observation of new boson (7.4σ)
- Observations also in several other channels
- Detailed studies of properties have begun
  - SM Higgs?
  - non-SM Higgs?
  - Something completely different?



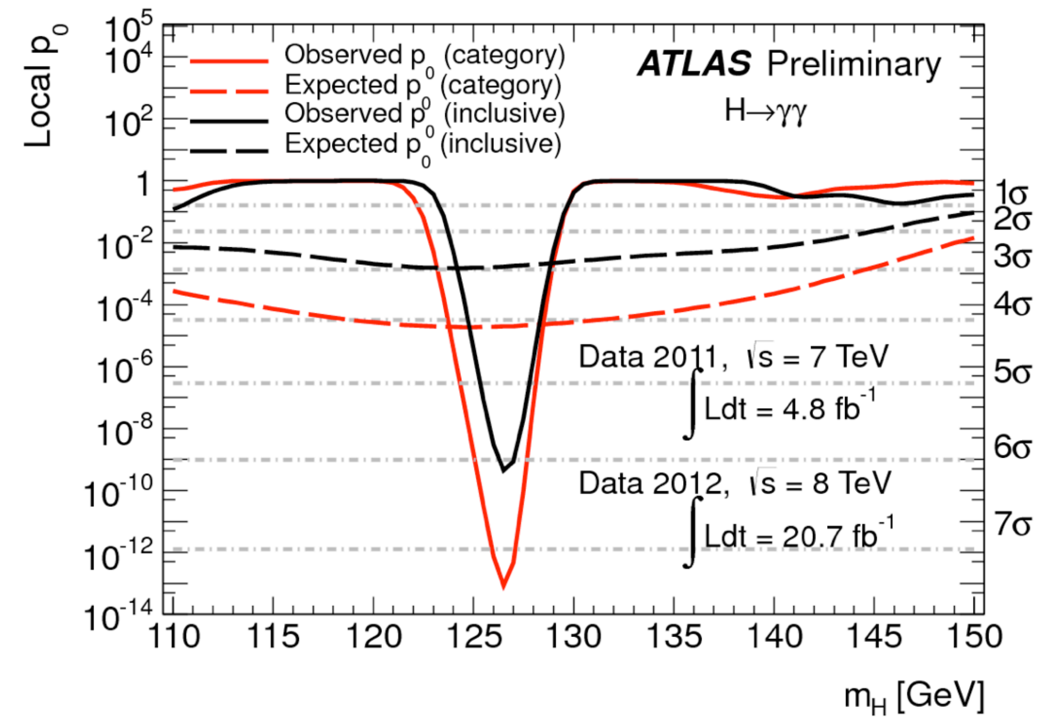


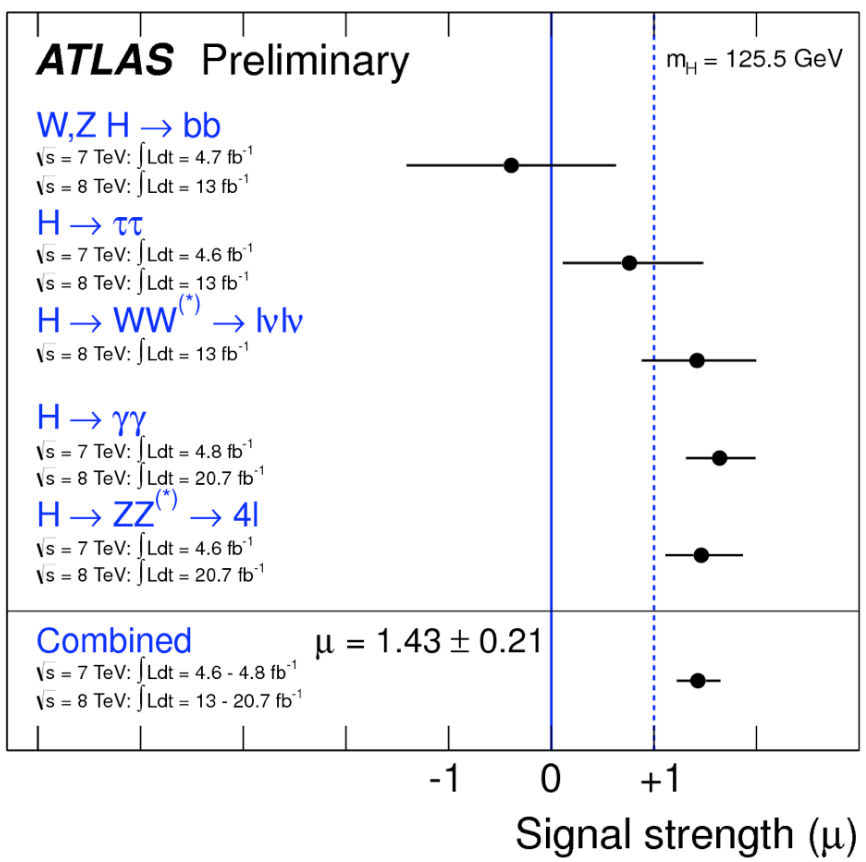


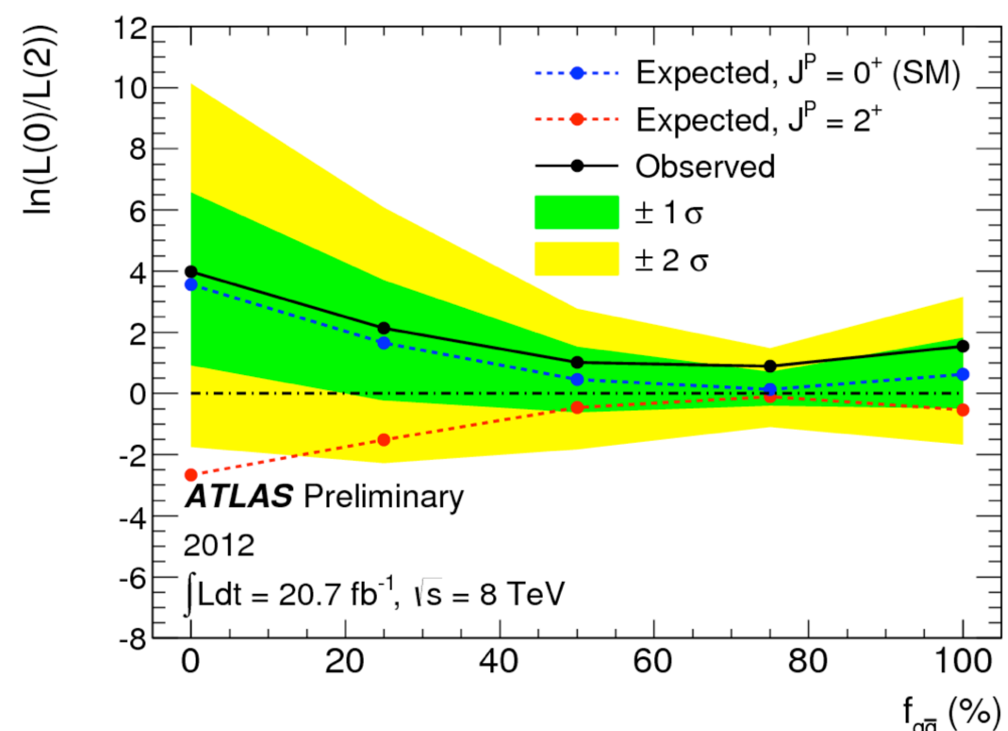
Higgs diphoton spin analysis

Summary & Outlook

- H→γγ channel alone able to claim observation of new boson (7.4σ)
- Observations also in several other channels
- Detailed studies of properties have begun
  - SM Higgs?
  - non-SM Higgs?
  - Something completely different?



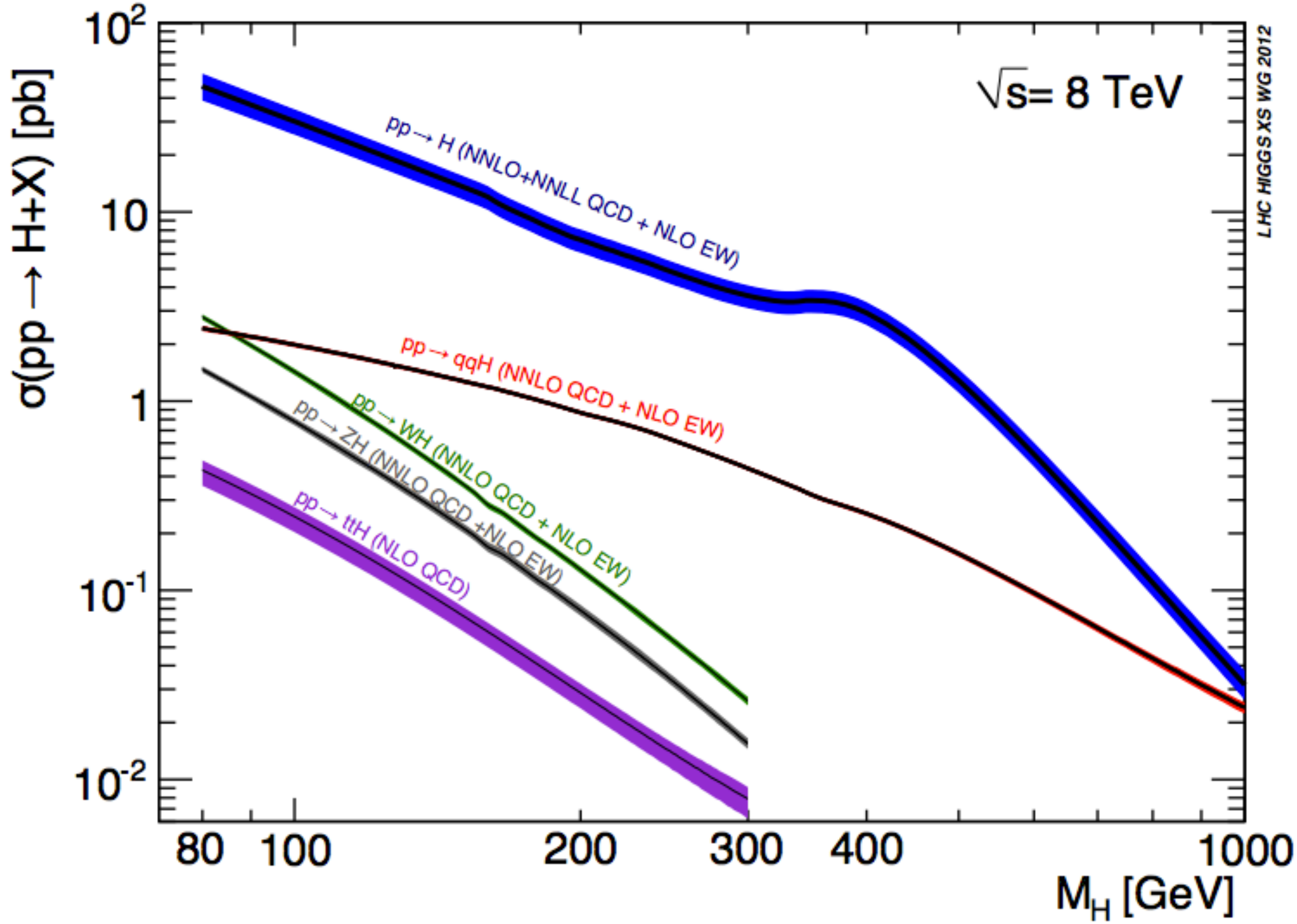




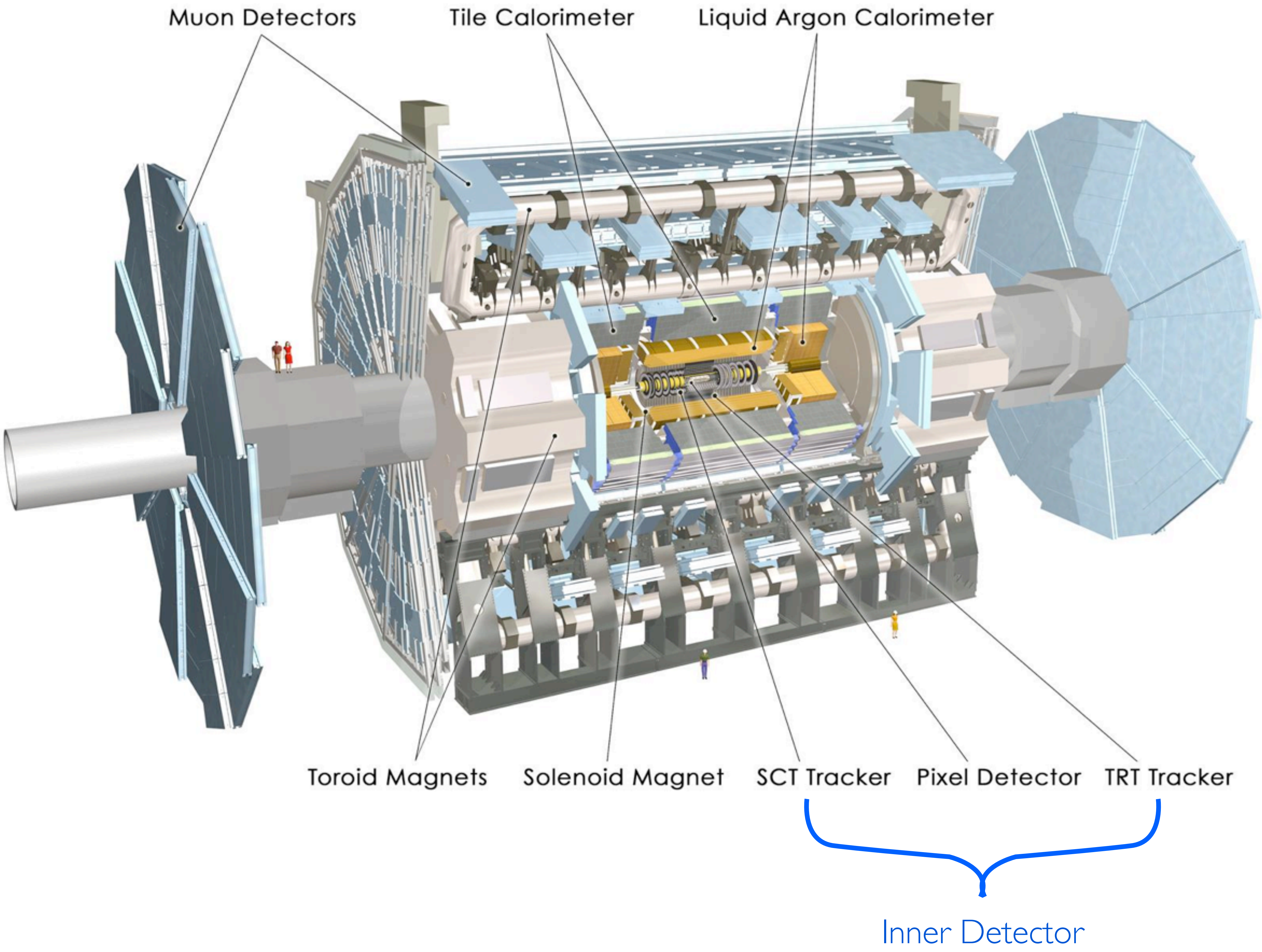
Higgs diphoton spin analysis

Backup

Higgs Production

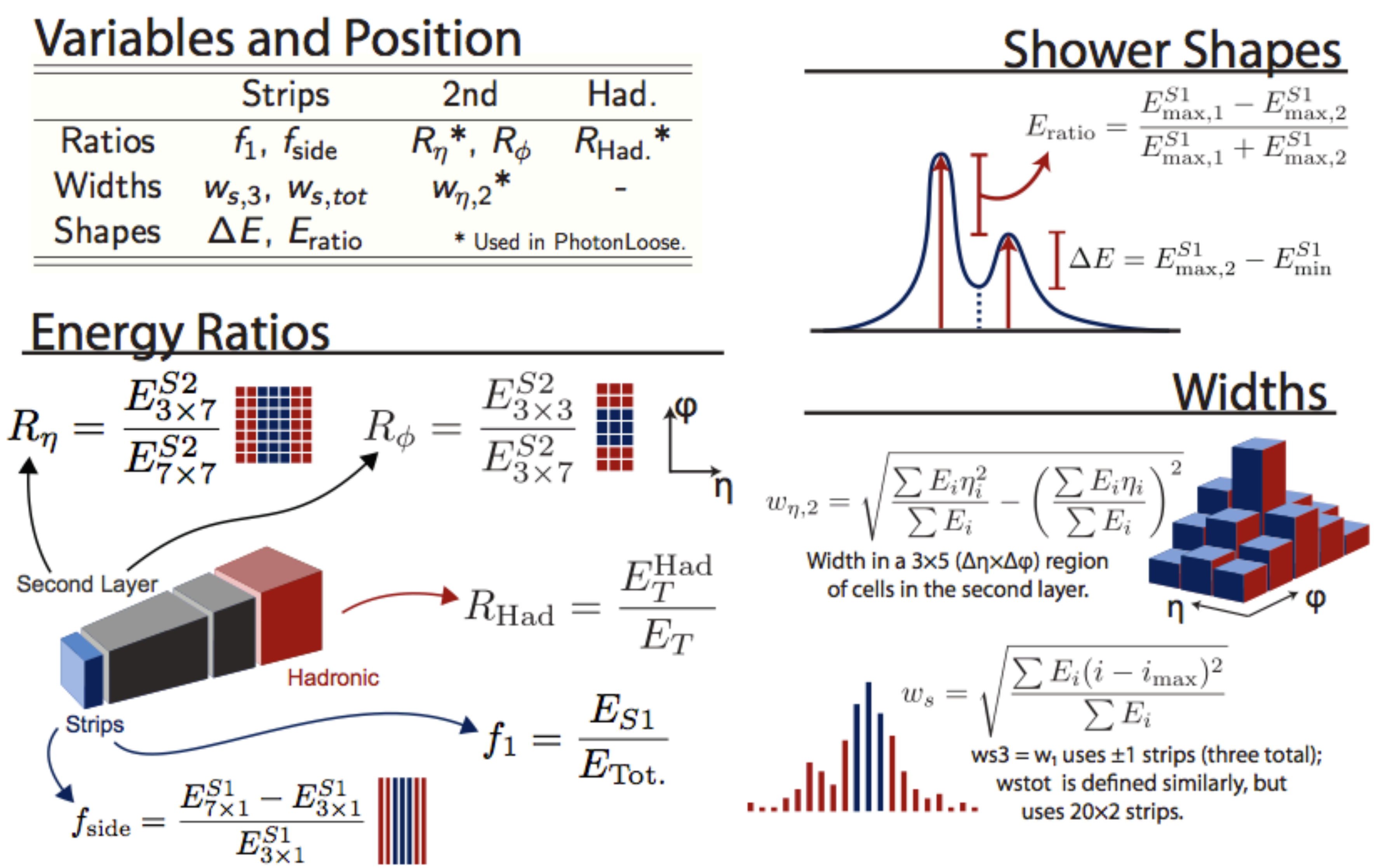


ATLAS Detector and Dataset



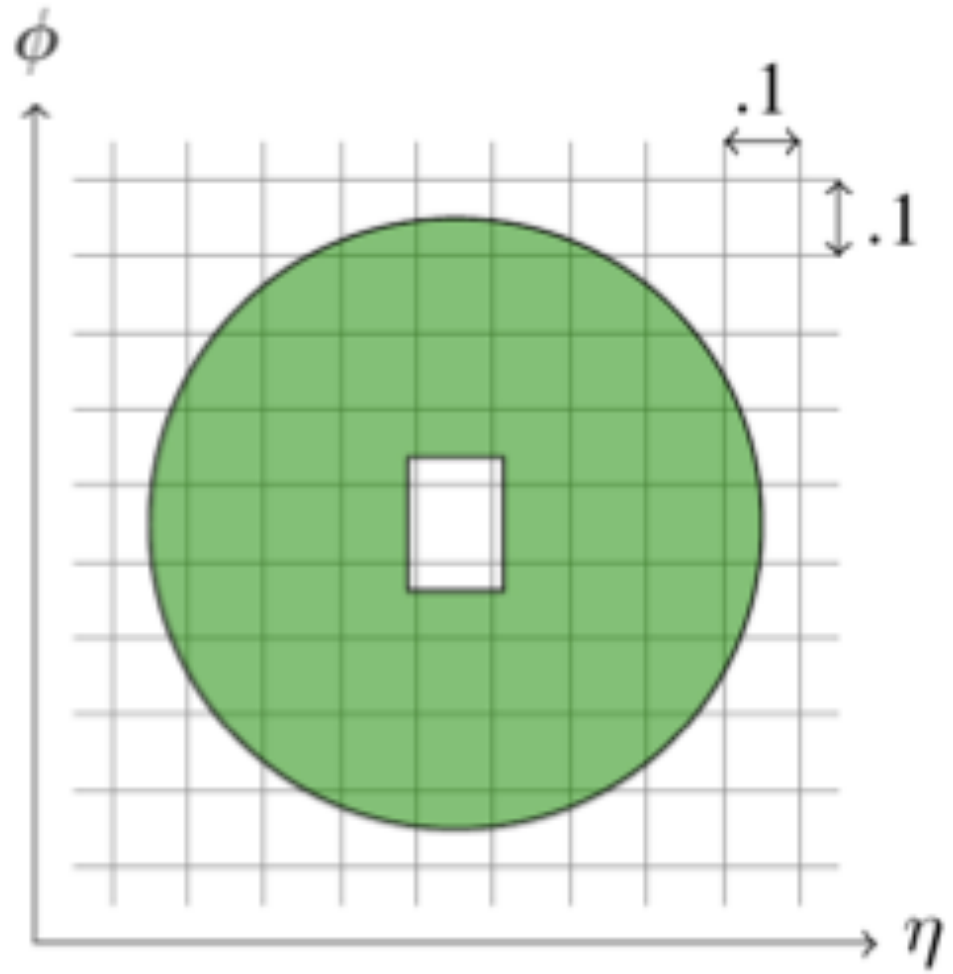
Shower shapes

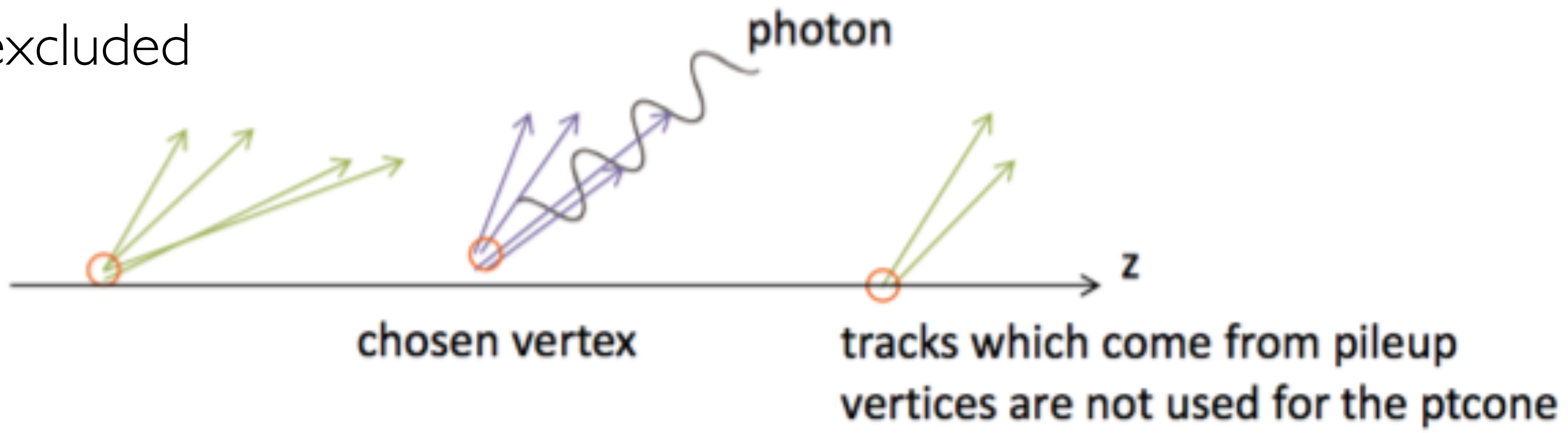
- Showers in ATLAS are described by variables called ‘shower shapes’
- Measure the distribution of energy within a shower - e.g. how wide is it in η,φ?



Photon isolation

- Reconstructed photon candidates have a very large hadronic background
- Require candidates to be isolated in calorimeter
- isolation computed within cone $\Delta R = 0.4$ around photon cluster
  - sum of energy within cone, excluding photon window
  - corrections for pileup
- track isolation also used in 8 TeV data
  - sum of pT of tracks within cone $\Delta R = 0.2$
  - tracks associated to photon are excluded
  - tracks from pileup vertices are excluded





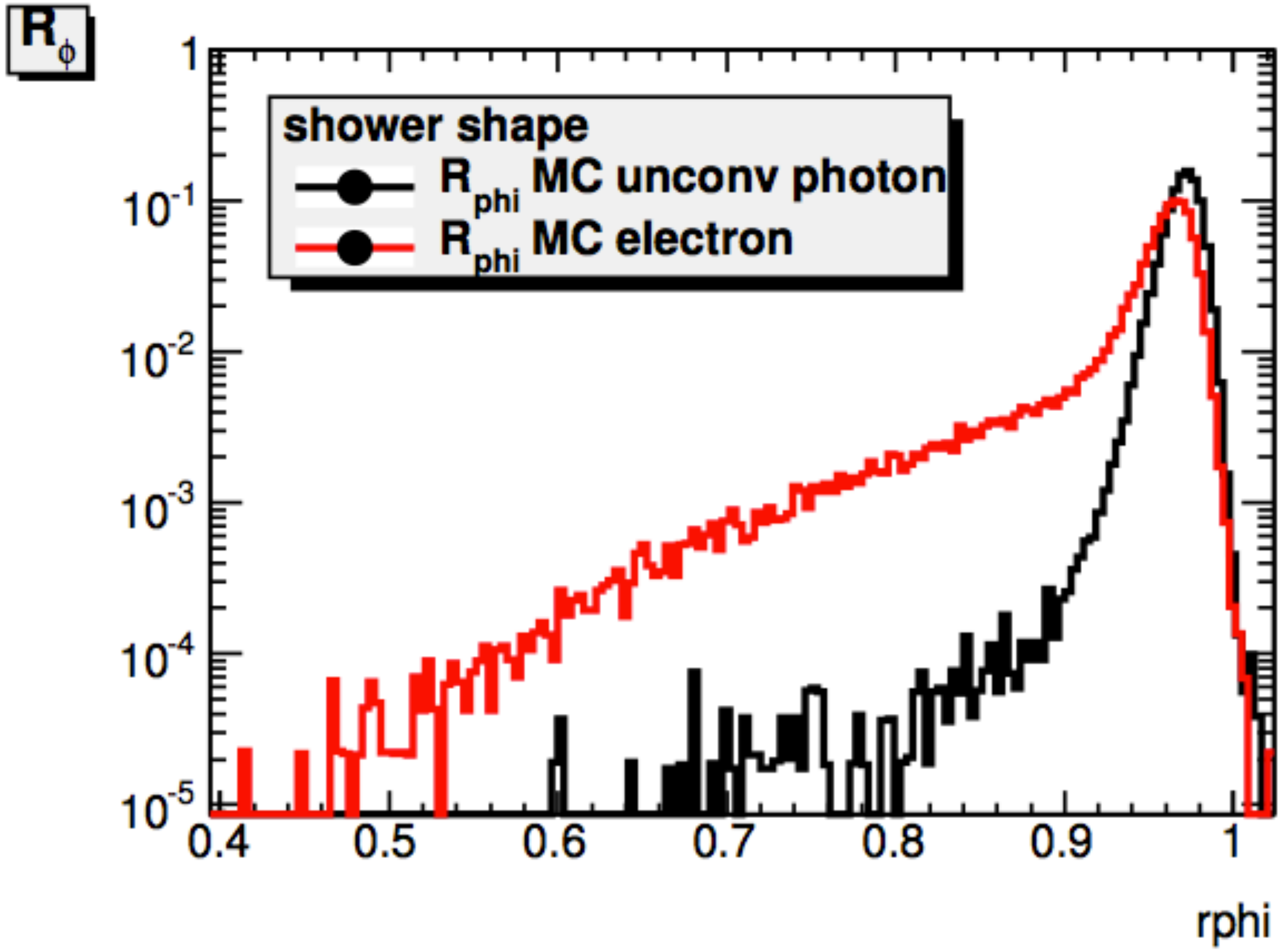
Photon ID efficiency

- photons in range 25 GeV $< p_T <$ 100 GeV are key in the $H \to \gamma\gamma$ analysis, so we need a good understanding of ID efficiency
- data-driven measurement is tough, as it is hard to select a pure sample of photons in this range
- $Z \to e^+e^-$ peak makes it easy to select high-purity electron sample
- both leave similar showers in EM calorimeter, transformations can improve similarity further

Can we take advantage of this?

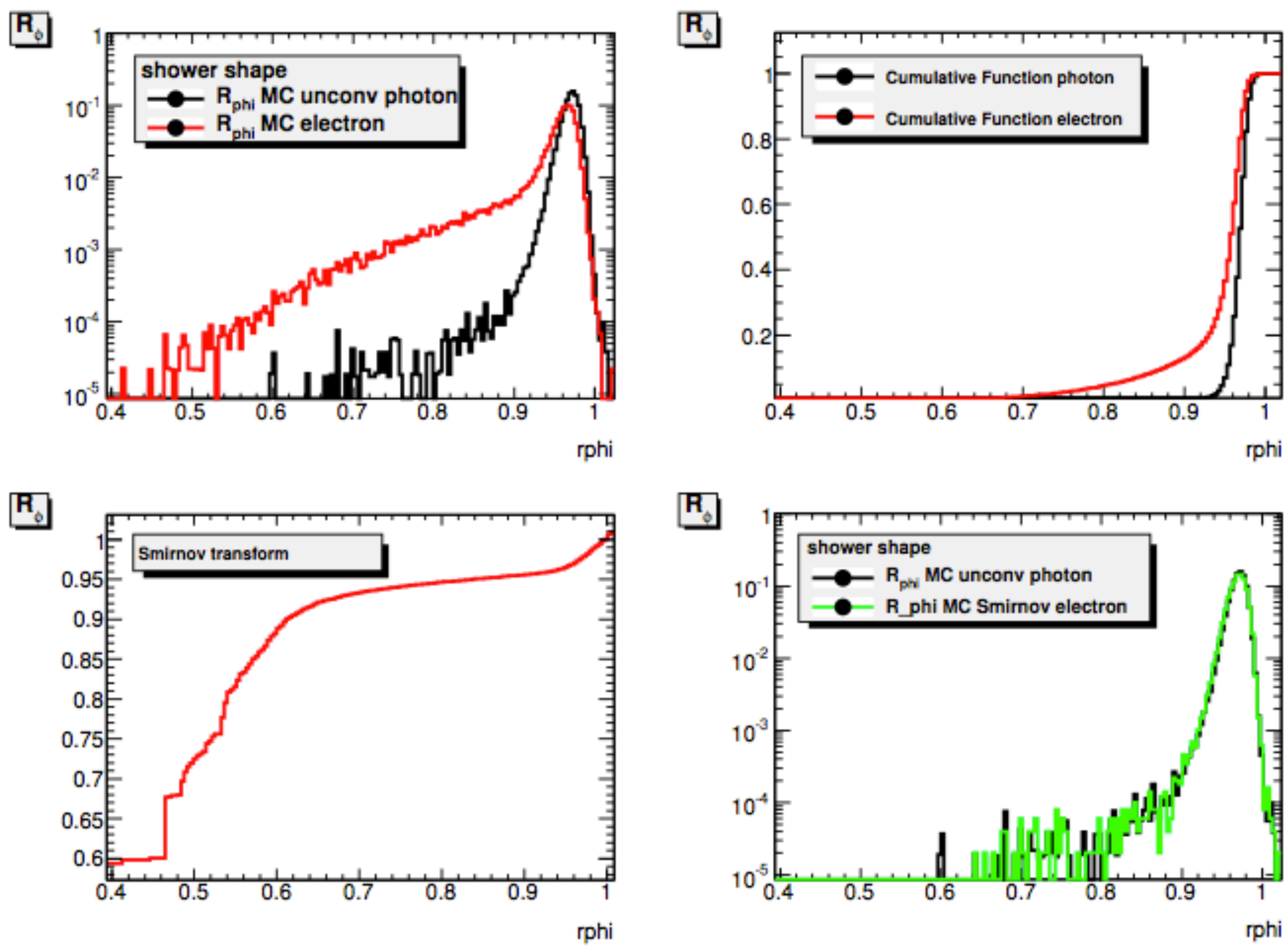
Electron extrapolation

- Same shower shapes for electrons and photons, but distributions can be very different
- We need photon distributions
- Smirnov transforms:
- Cumulative distribution functions (CDF) are calculated for each shower shape, based on MC samples
- A transformation can then be derived, from electron distribution to photon distribution
- Applying this transformation to data electrons gives us 'photons'
- Process must be done individually for each shower shape, in each pT, η bin, for converted and unconverted photons



Electron extrapolation

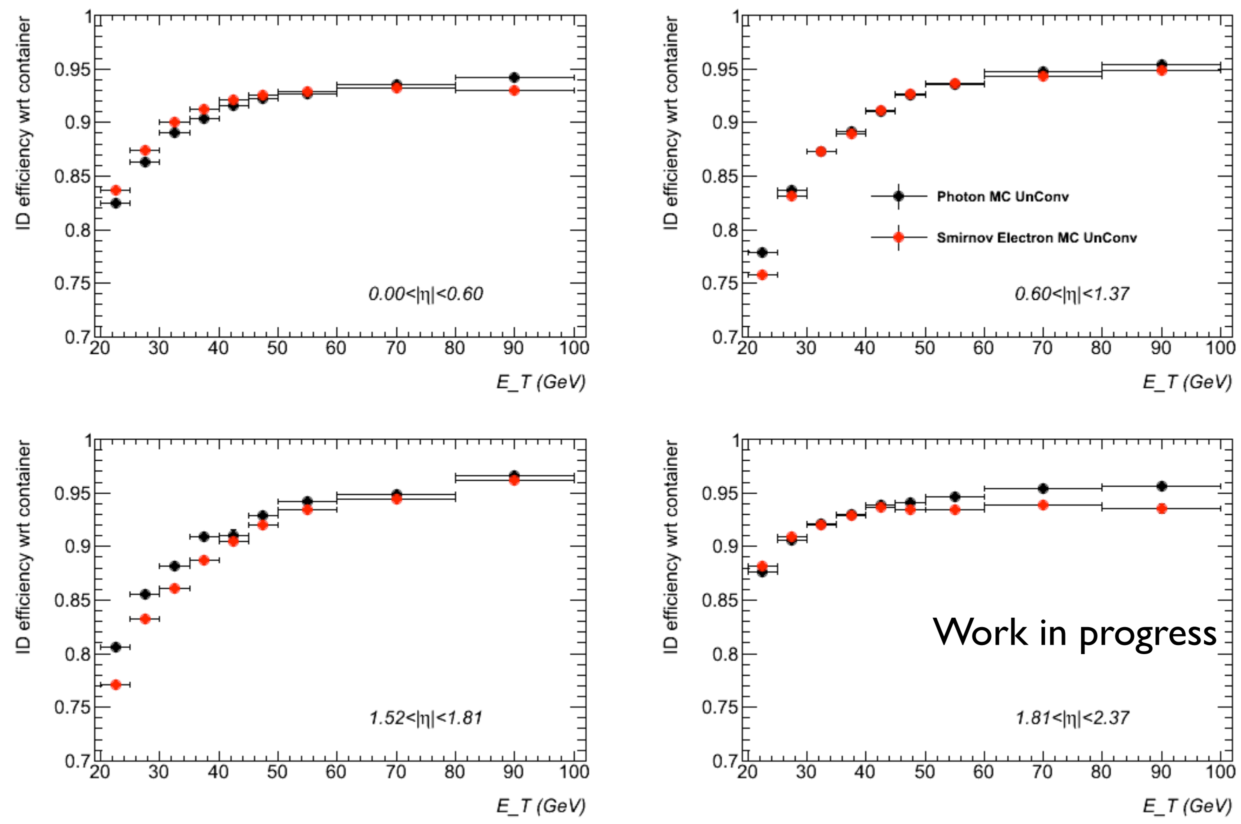
- Same shower shapes for electrons and photons, but distributions can be very different
- We need photon distributions
- Smirnov transforms:
- Cumulative distribution functions (CDF) are calculated for each shower shape, based on MC samples
- A transformation can then be derived, from electron distribution to photon distribution
- Applying this transformation to data electrons gives us 'photons'
- Process must be done individually for each shower shape, in each pT, η bin, for converted and unconverted photons

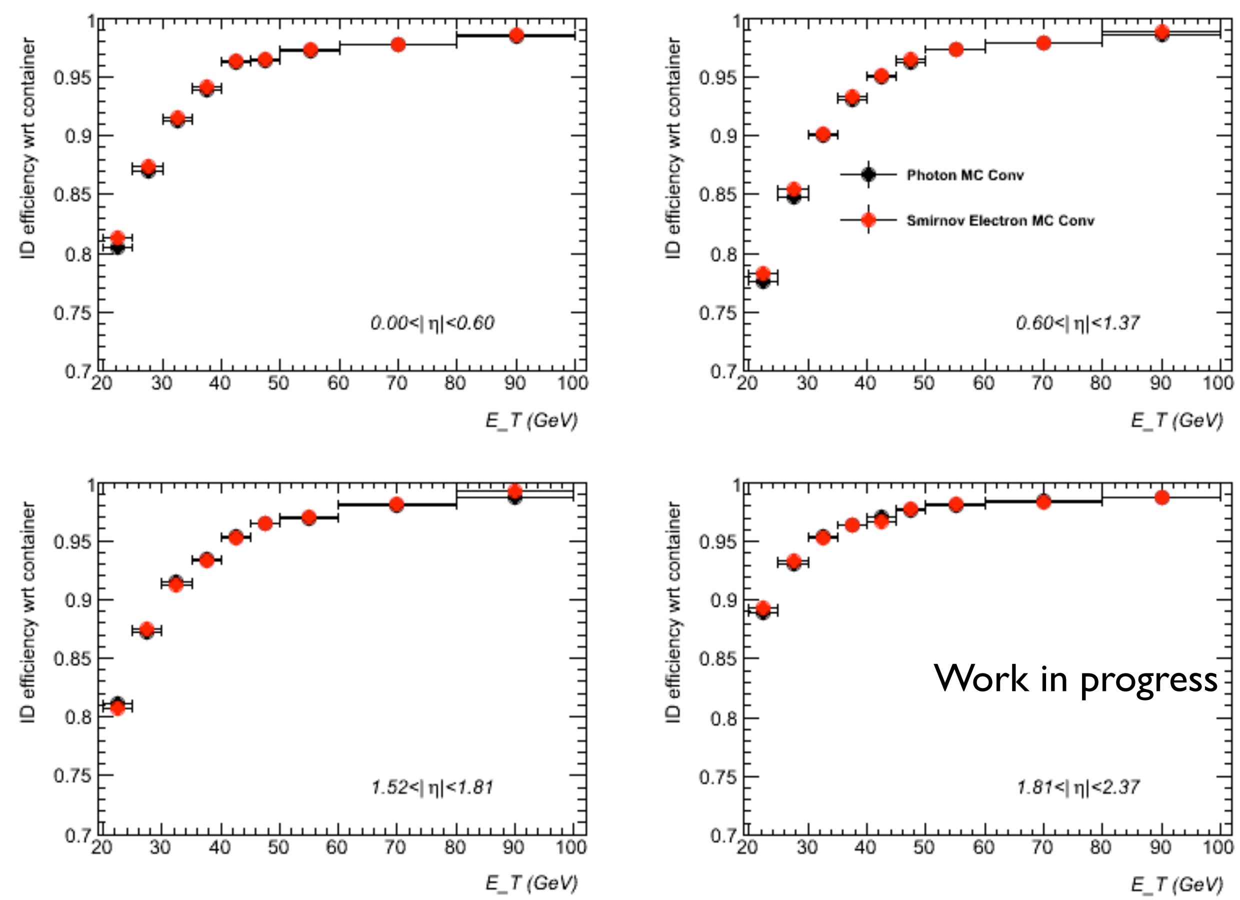


Electron selection

- Tag & probe selection, optimised for high purity
- Placing strict cuts to 'tag' electron allows high confidence selection of $Z \rightarrow e^+e^-$ events
    - Tag selection: shower shapes, tracking, isolation
- Electron ID shower shape cuts are not applied to probe, to avoid biasing sample
    - Probe selection: tracking, isolation

Closure test

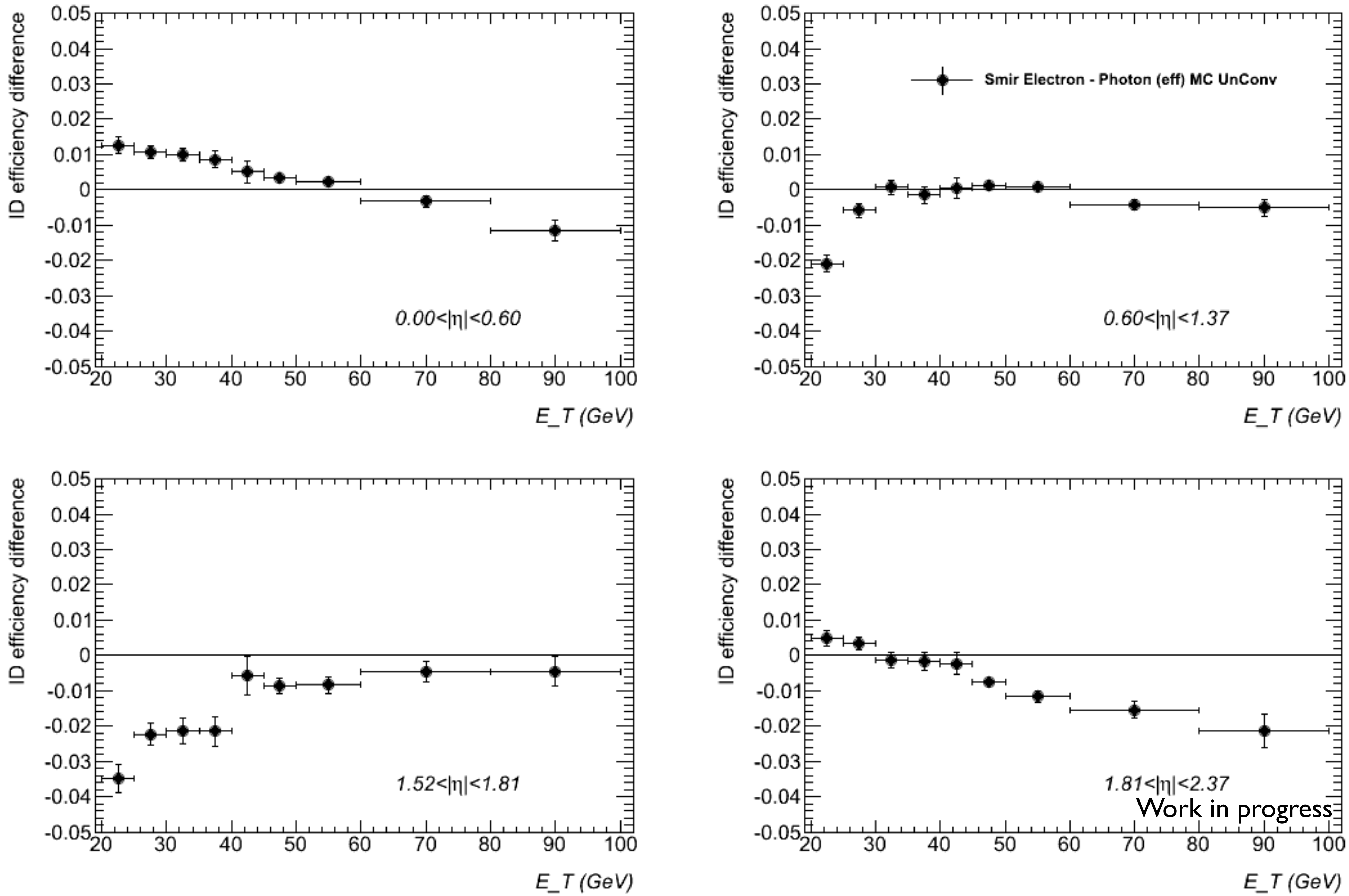




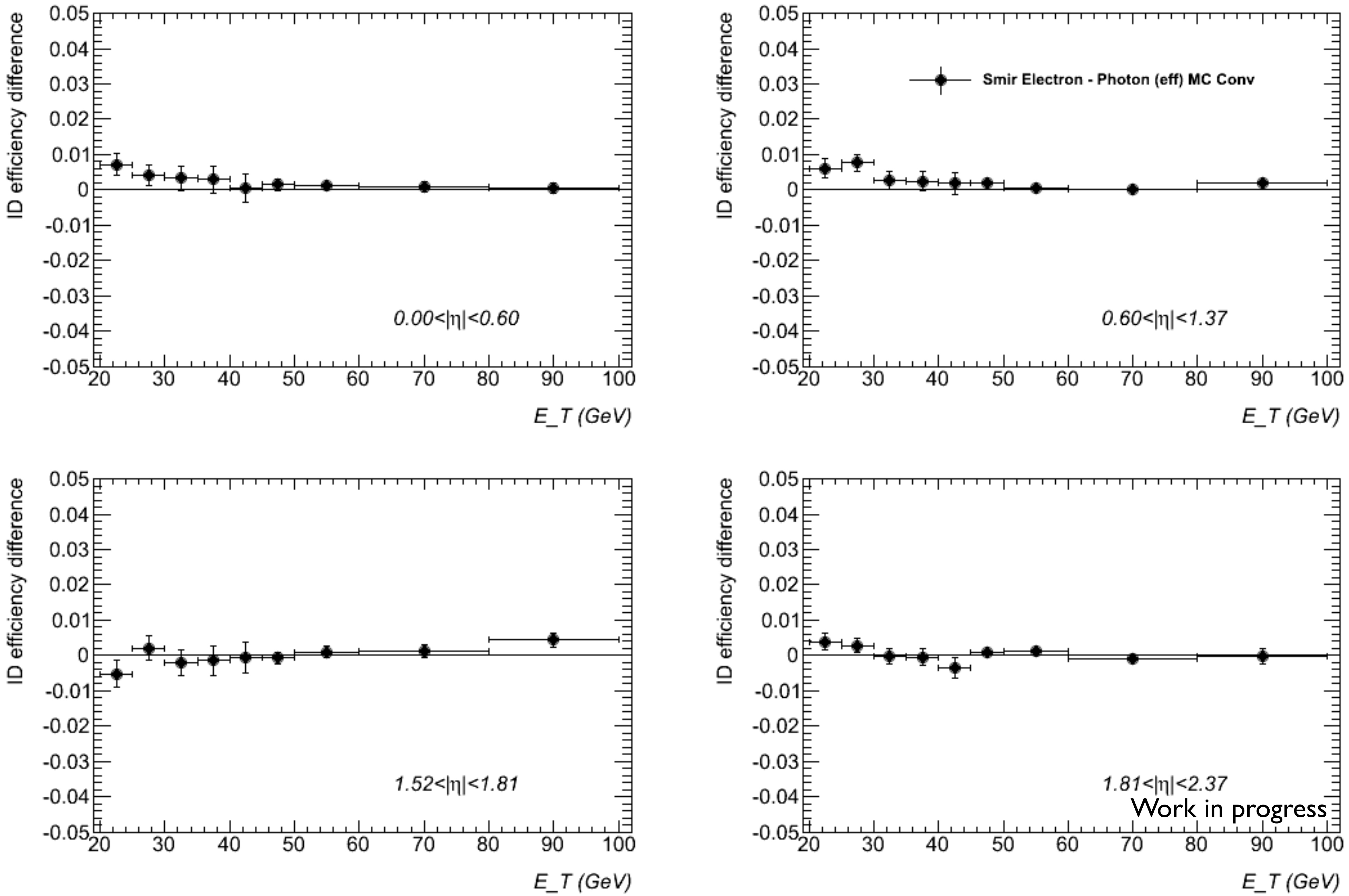
Closure test compares efficiency of MC electron sample after transformation to efficiency of raw MC photon sample

top: unconverted
bottom: converted

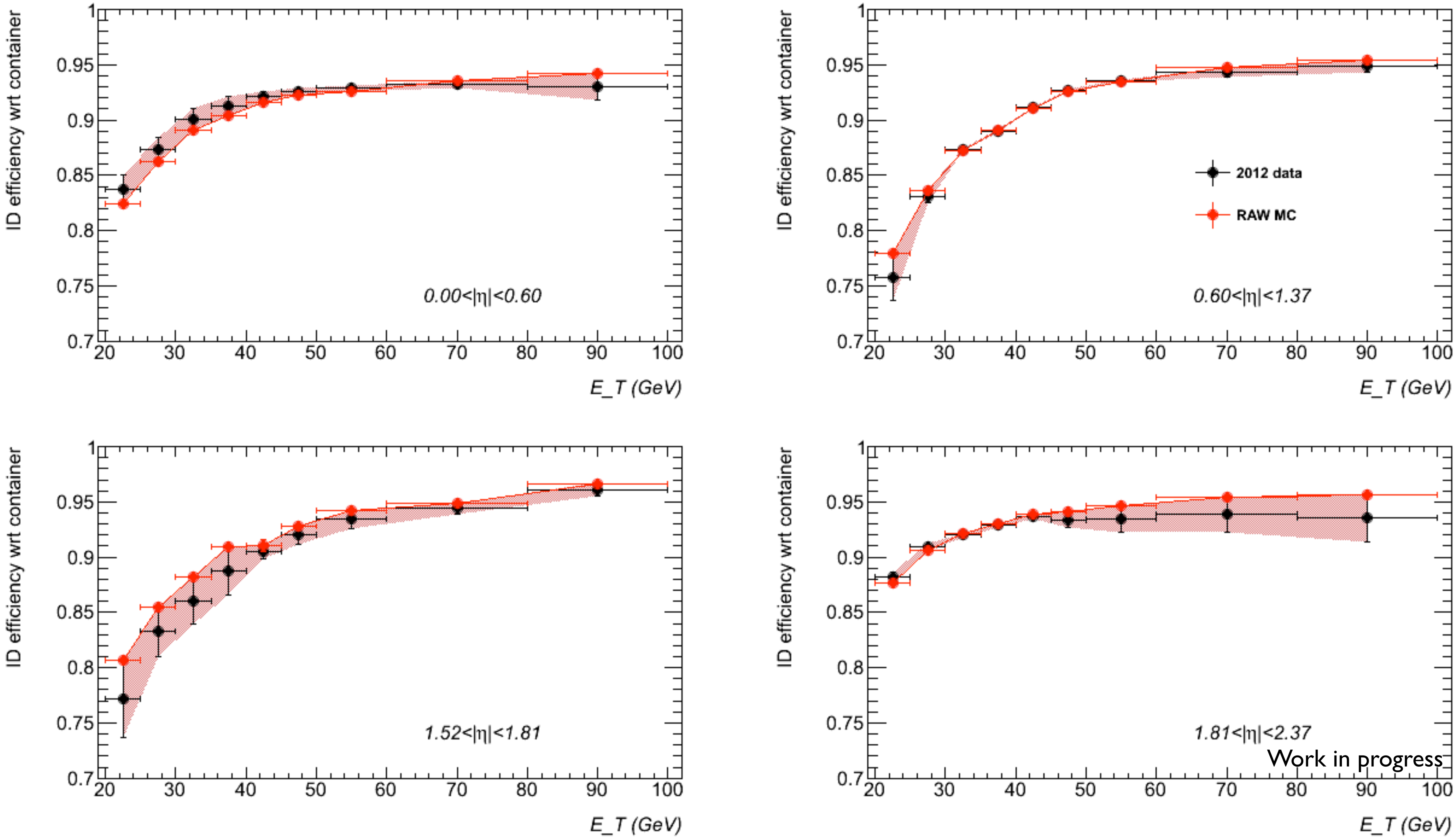
Closure test: unconverted photons



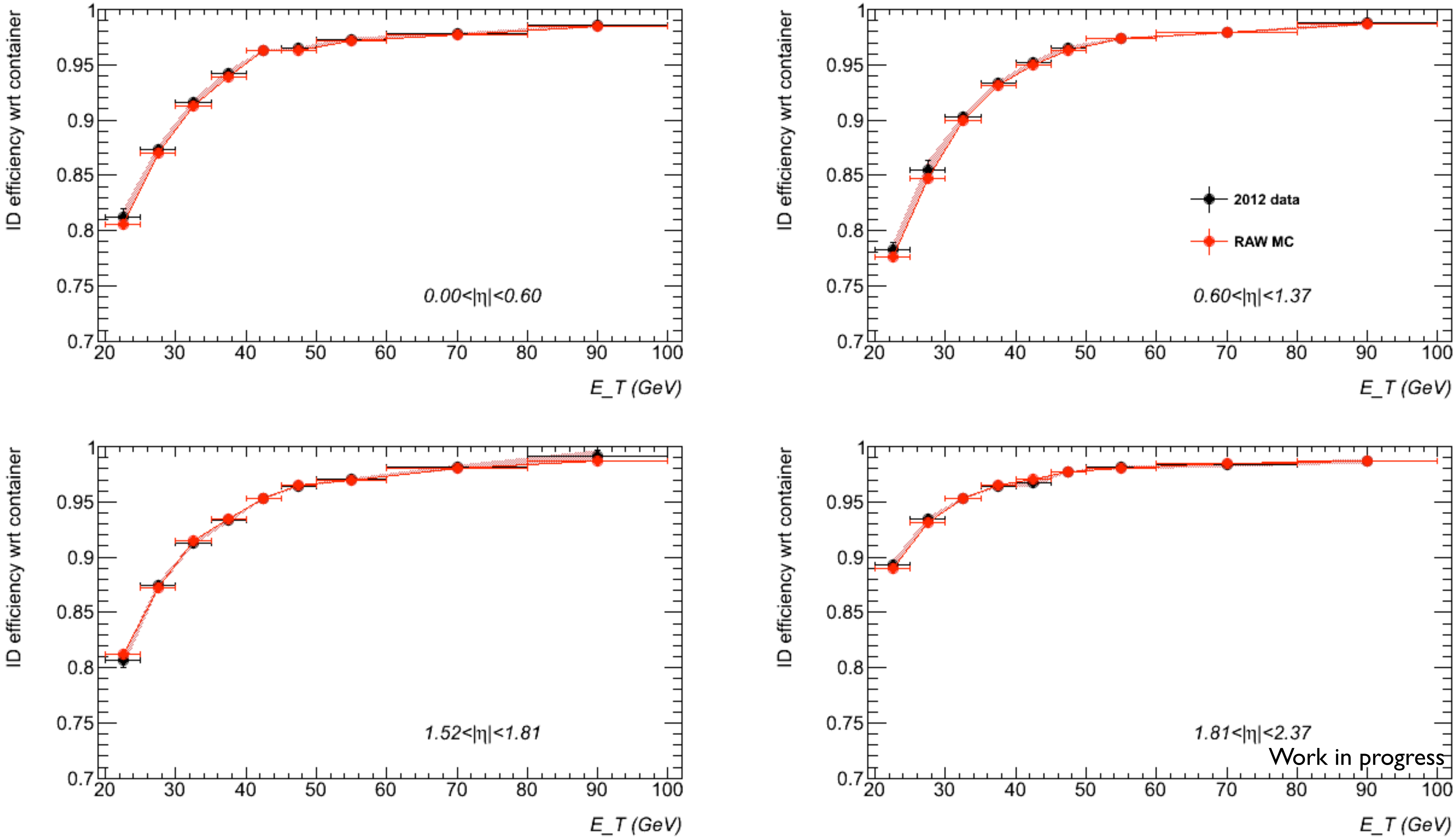
Closure test: converted photons



Photon ID efficiency: unconverted



Photon ID efficiency: converted



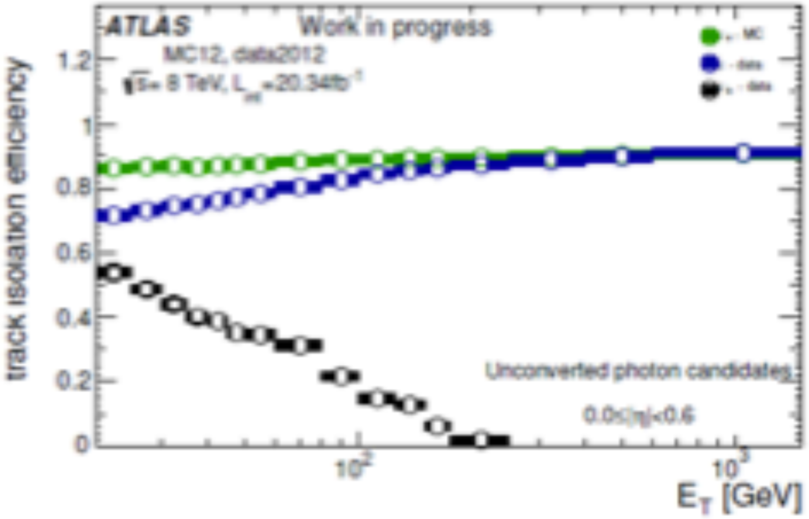
PID efficiency from "matrix" method

<u>Main idea:</u> to use all reconstructed photon candidates any energy, but initially evaluate purity for them before and after tight ID using additional discrimaiating variable (track isolation)

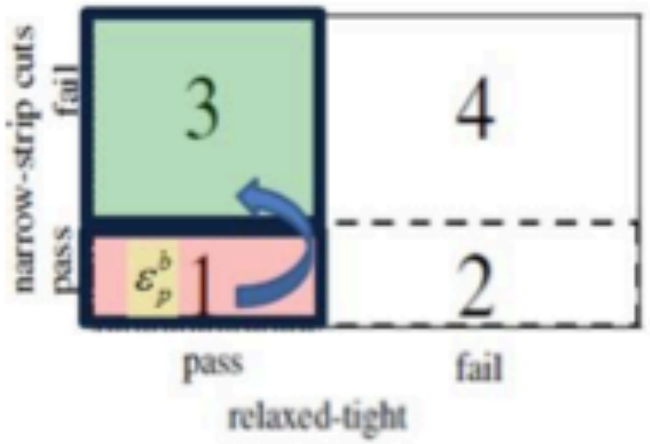
Covers 20-1500 GeV region

$$\frac{N^S_{pass\_tight}}{N^S_{pass\_tight}+N^B_{pass\_tight}} = P = \frac{\epsilon_p-\epsilon^b_p}{\epsilon^s_p-\epsilon^b_p}$$ - Purity after tight

$$\frac{N^S_{all}}{N^S_{all}+N^B_{all}} = A = \frac{\epsilon_a-\epsilon^b_a}{\epsilon^s_a-\epsilon^b_a}$$ - Purity before tight



- ✓ ϵ^s - Calculate signal isolation efficiency from MC (green)
- ✓ ϵ - Calculate isolation efficiency from data (blue)
- ✓ ϵ^b - Calculate fakes isolation efficiency from data (black)



✓ bkg track iso efficiency for region 1 is estimated from region 3 (same, but fail narrow-strip cuts)

✓ regions 2+3+4 are used to estimate track iso efficiency of the sample in region 1+2+3+4

<u>And PID efficiency:</u>

$$\epsilon^{tight\_ID} = \frac{N^S_{pass\_tight}}{N^S_{all}} = \frac{N^{data}_{pass\_tight}\times P}{N^{data}_{all}\times A}$$

Systematic uncertainties:

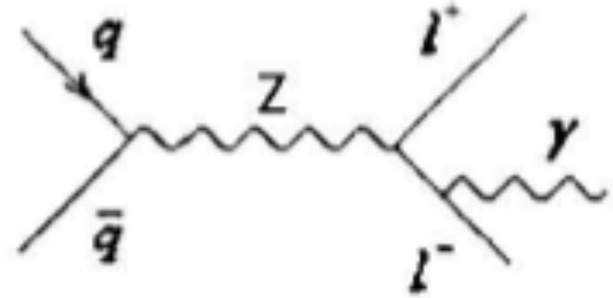
➢ Background ε^{iso} uncertainty from difference of the bkg iso efficiency in control region and in signal region from JF samples (**up to 5% error before tight, and up to 10% for fakes passed tight selection**)
mostly for low energies

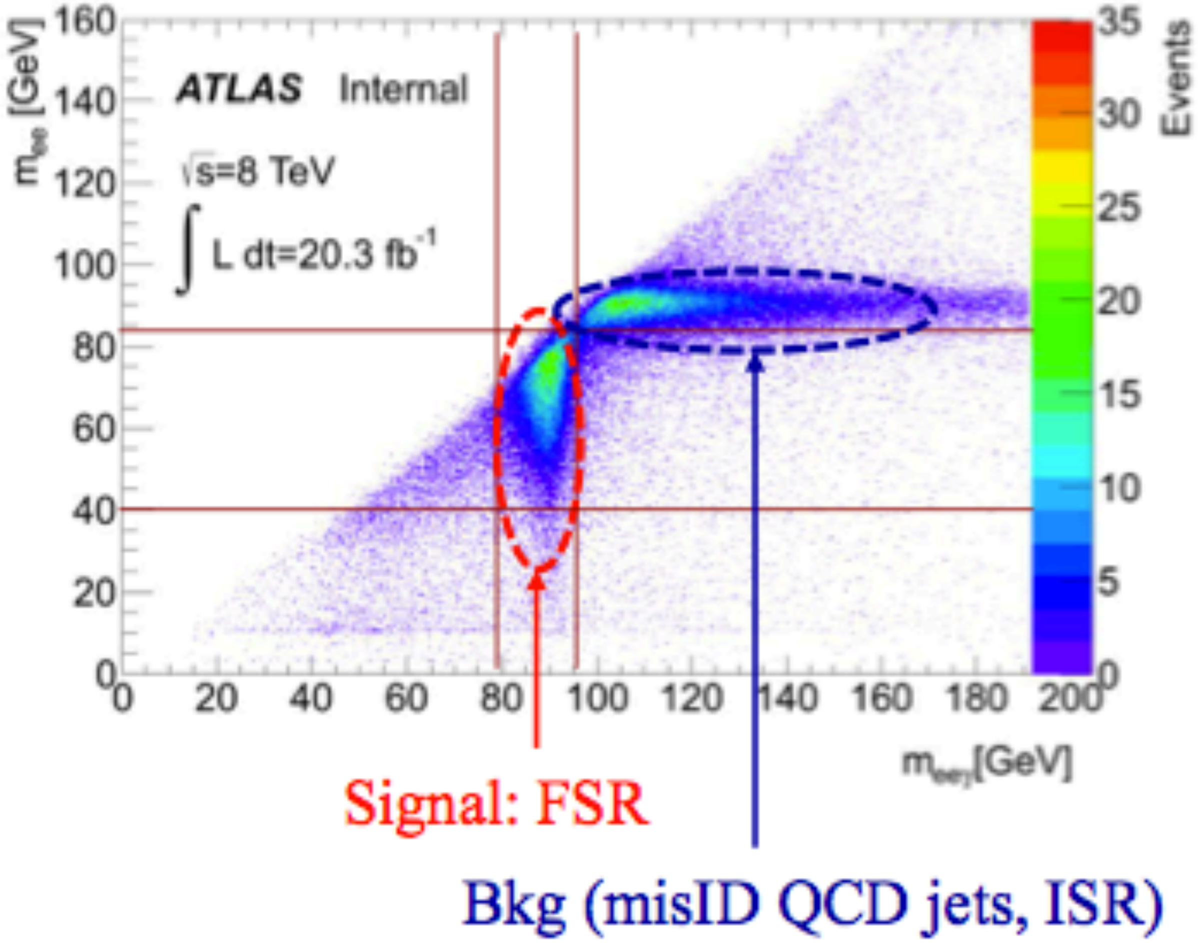
➢ Signal ε^{iso} uncertainty from comparison data/MC distributions for tag&probe $Z\rightarrow ee$ (**<1%**)

➢ Correlation btw track isolation and narrow strip variables can pull the efficiency downwards
still needs to be checked!

PID efficiency measurement using Z→llγ decay

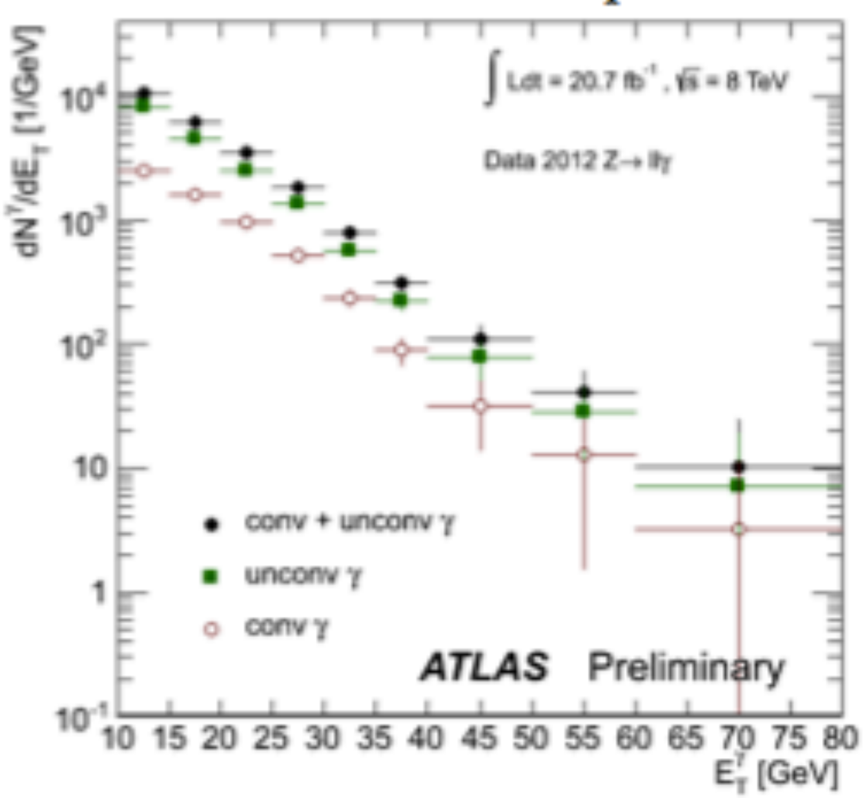
Main idea: obtain a pure (P ~ 96-97 %) photon sample selected from Z radiative decays using M_{ll}, $M_{ll\gamma}$ kinematic selections, which are not biasing shower-shape variables.





Covers 10-80 GeV region

Selected events spectra:



- **Main problem**: small cross-section ⇒ low statistics and decrease of E_T spectra towards high energies
  - Statistical uncertainty is the dominating one
  - Measurement of efficiency possible **up to ~40 GeV** (2011 data) and **up to ~80 GeV** (2012 data)