

Electron reconstruction efficiency measurement with the ATLAS detector at the LHC, and study of the Higgs boson coupling to the top quark with the two same sign leptons channel.

Kevin de Vasconcelos, under the supervision of **Fabrice Hubaut**

CPPM / IN2P3 - Aix-Marseille Université (Marseille, France)

[kevin.devasconcelos@cern.ch](mailto:kevin.devasconcelos@cern.ch)

Journées de Rencontre des Jeunes Chercheurs 2016 - 5<sup>th</sup> December 2016

# Motivations



Leptonic events are « markers of interest » in the harsh pp collision hadronic environment.

- Clear experimental signature.
- Cross section  $> 10^{-6} \sigma_{\text{QCD}}$   
 → Very useful to define triggers

• Tag processes of high interest :

- Z/W physics :
  - Z (+jets)
  - W (+jets)
  - Dibosons
- Higgs :
  - $H \rightarrow ZZ^* \rightarrow 4 \text{ leptons}$
  - $H \rightarrow WW^* \rightarrow 2l + 2\nu$

**Precision measurements**

- Top physics :
  - tt (+jets)
  - single top

- SUSY
- Exotics

**New physics**

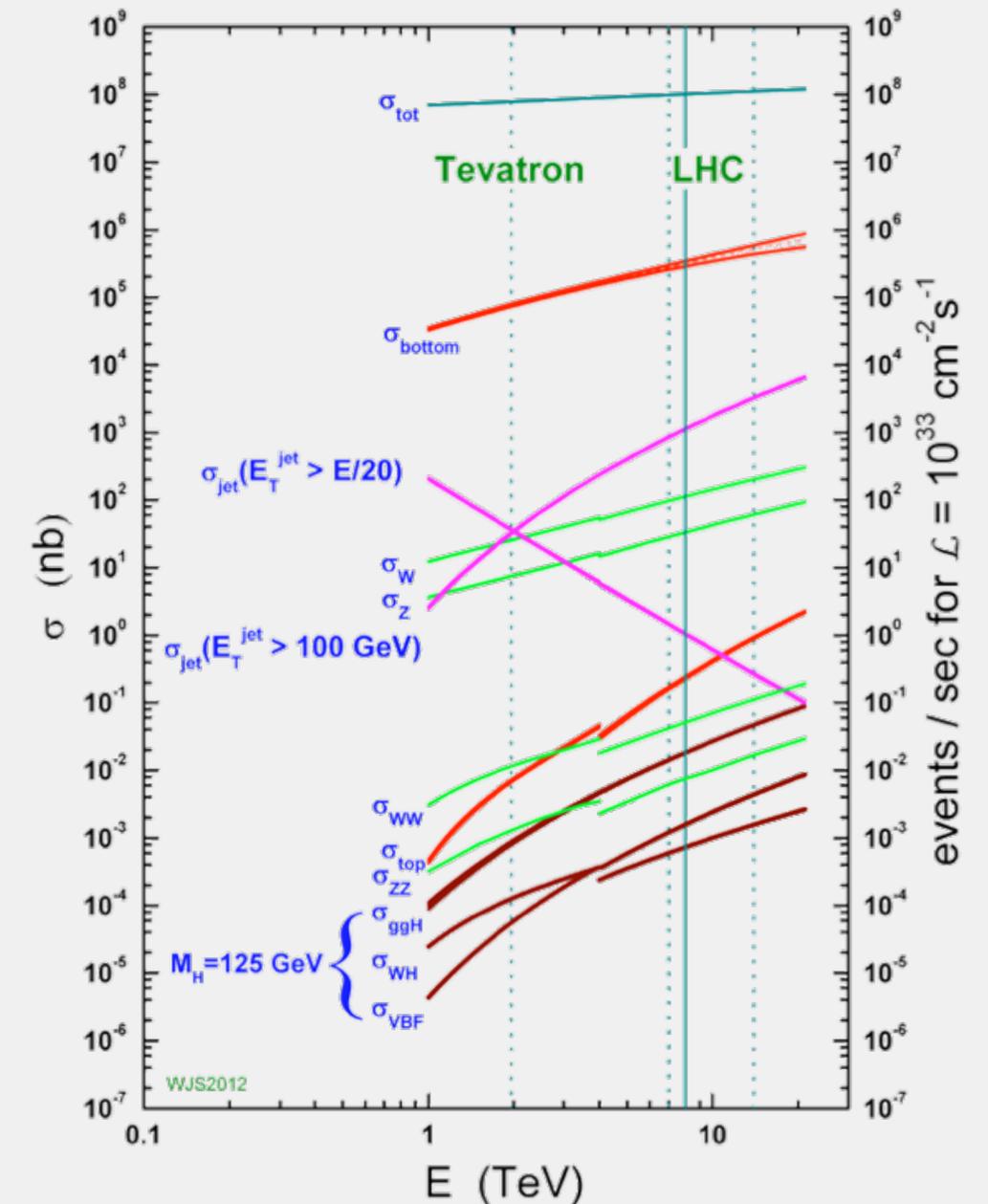
• Needs :

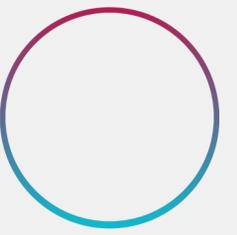
1. low number of events for NP: high selection efficiency
2. high background rejection (ex  $H \rightarrow WW$ )



Require a good understanding of the ATLAS detector.

proton - (anti)proton cross sections





## I

### Experimental context

- 1 Electromagnetic calorimeter
- 2 Tag and probe method
- 3 Background estimate

## II

### Electron reconstruction efficiency

- 1 Comparison with last pre-recommendations
- 2 Stability checks

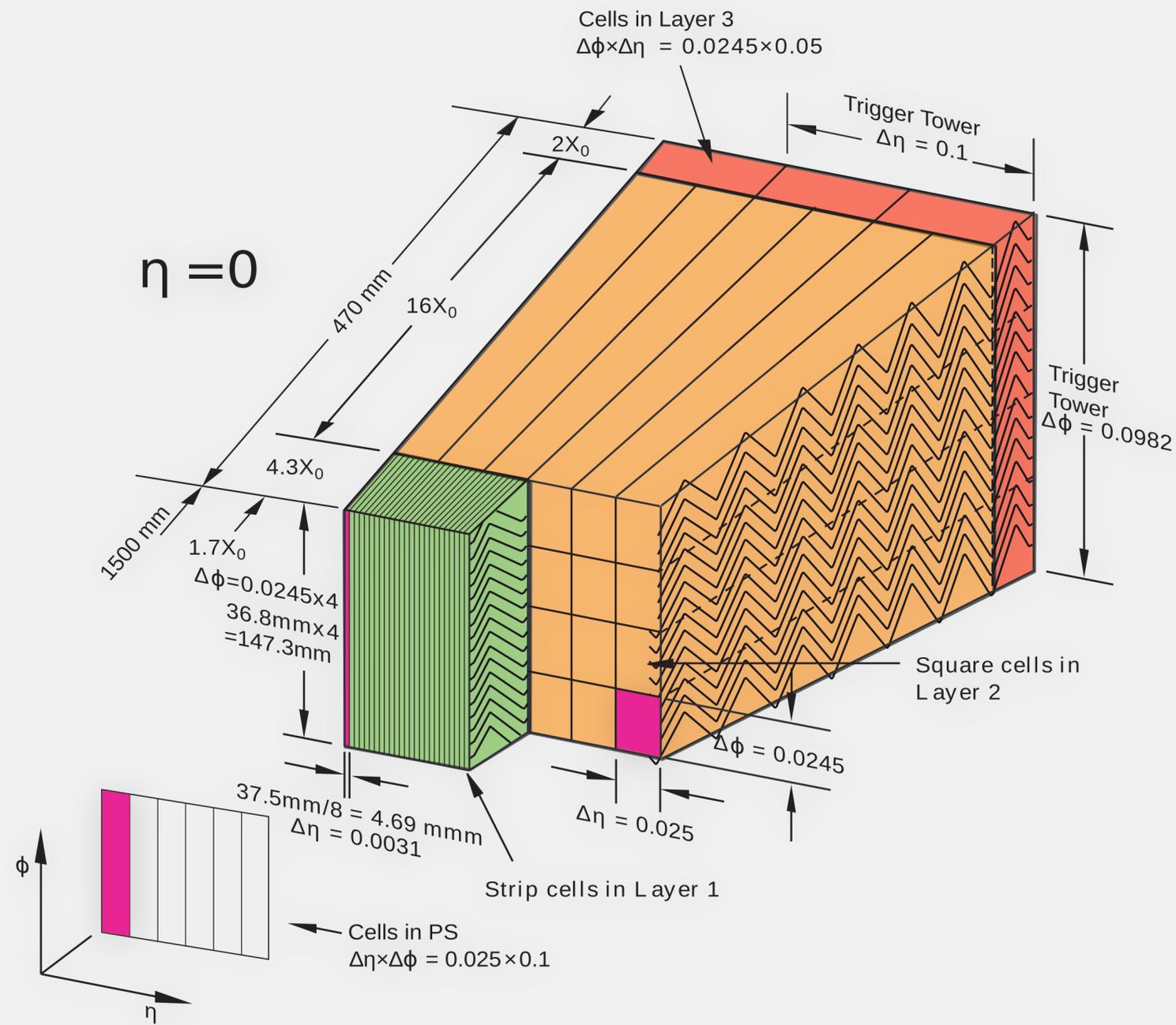
## III

### ttH with two same-sign leptons

- 1 The channel
- 2 Fake estimates
- 3 Multivariate analysis

# Experimental context

Electromagnetic calorimeter



Sketch of the electromagnetic calorimeter granularity

Showers of EM particles generated by the passage of particles through cells.

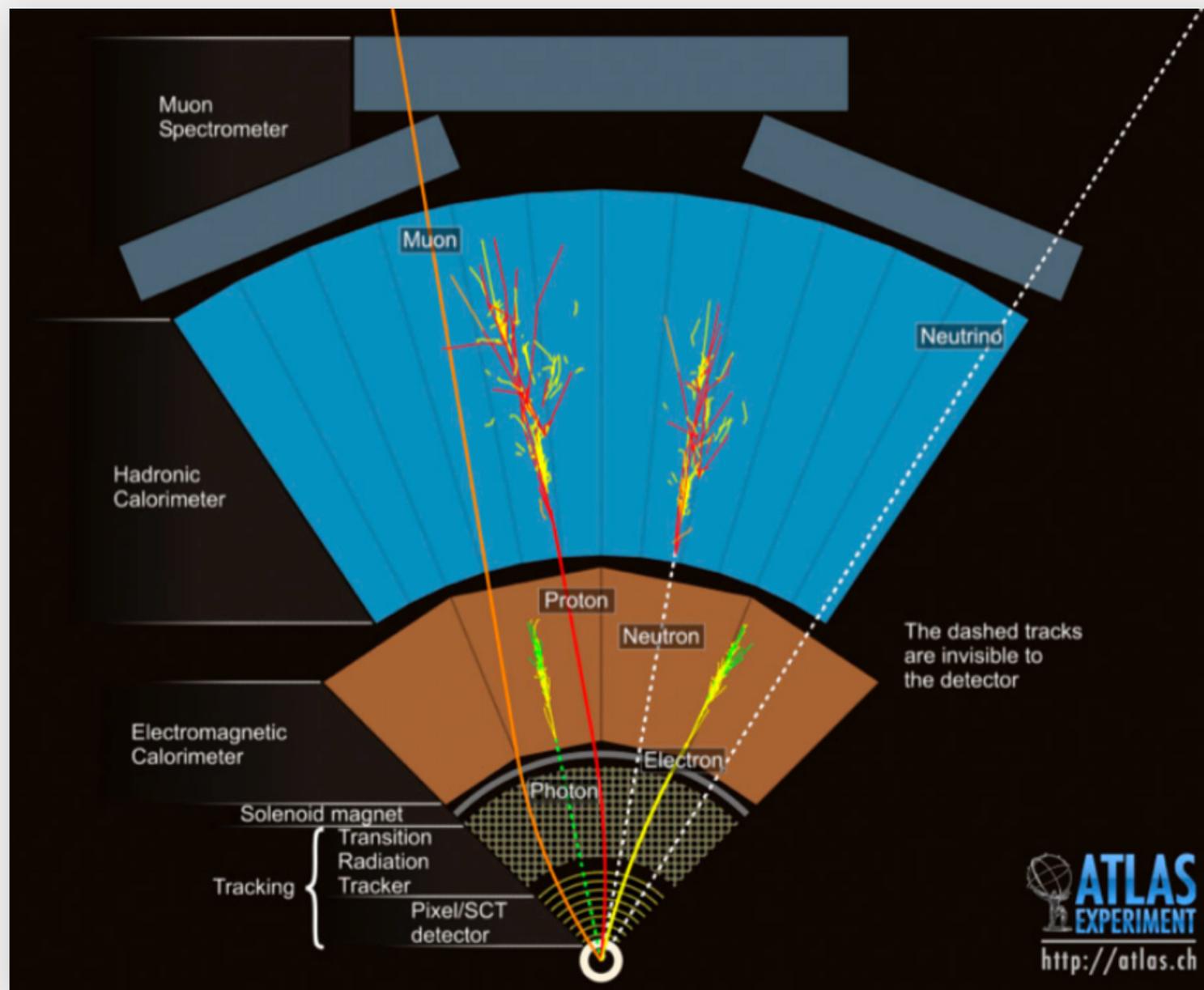
High segmentation : high resolution  
3 layers (S1-3) + pre-sampler (PS)

Accordion geometry : a full coverage and a very good uniformity in  $\phi$

S1 : highest granularity in  $\eta$  : shower shape &  $\eta$  measurements

S2 : accumulates most of the EM shower energy  
Fine granularity  $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$  ( $|\eta| < 2.5$ )

S3 : coarser granularity : collects only the tail shower E



## Principle

- Reconstruction in the central region of the detector ( $|\eta| < 2.47$ )
- starts from energy deposits (clusters) in the EM calorimeter in a fixed  $\Delta\eta \times \Delta\phi$  window
- Deposits are associated to reconstructed tracks of charged particles in the inner detector
  - No track found, or conversion vertex  $\longrightarrow$  photon
  - Track found  $\longrightarrow$  likely an electron

## Challenges for efficiency measurement

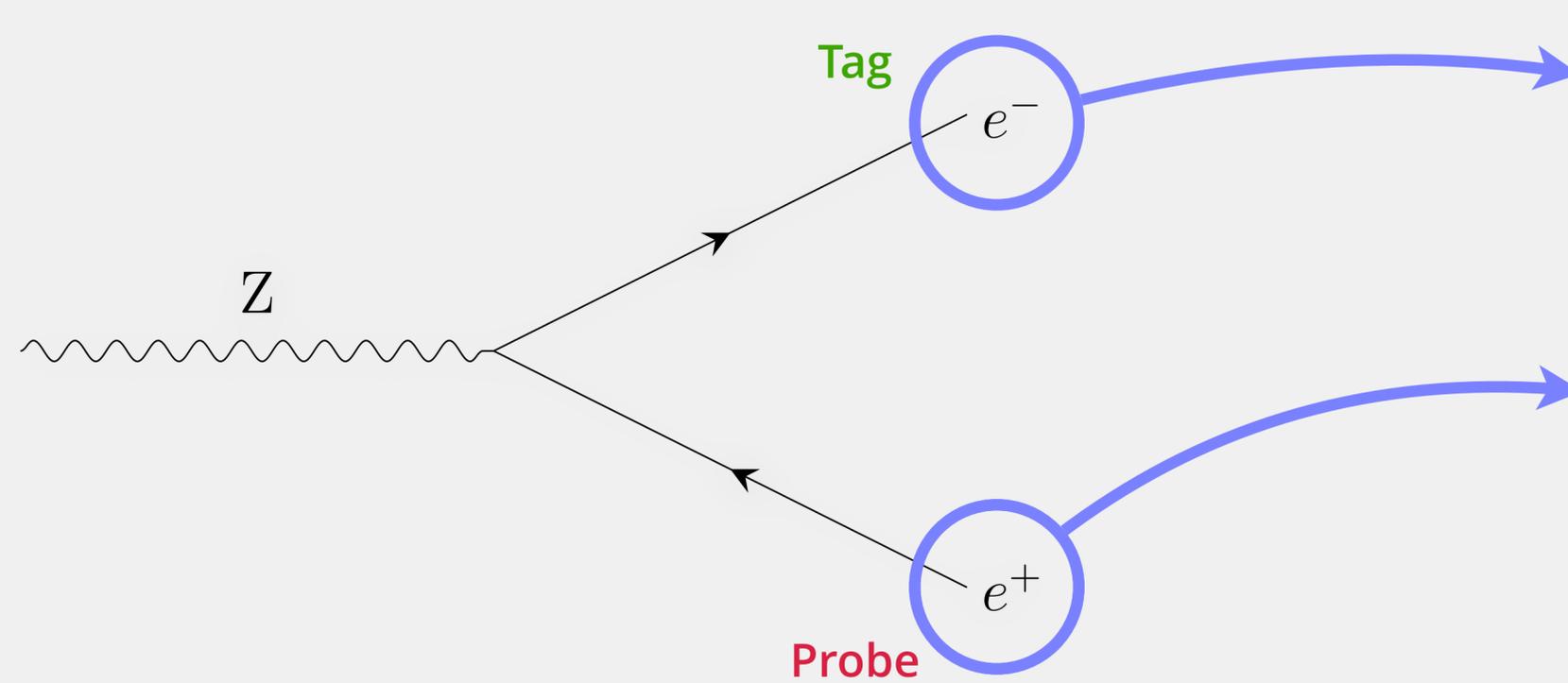
- Main challenge for a precision measurement = background estimate (we use clusters: high and complex background)
- Simulation not reliable enough



Data driven measurements are needed for precision physics measurements such as Z cross section or W mass

# Tag and Probe method

In  $Z \rightarrow ee$  channel



- 1** Pure and unbiased sample  
Strict selection applied on the first electron.  
Identifies the Z decay event
- 2** A looser candidate with  $M_{ee} \sim M_Z$   
Cell cluster reconstructed in the calorimeter  
Used to calculate the efficiency  
+
- 3** Track quality criteria  
At least 1 hit in the pixel detector and 7 in the SCT

Why  $Z \rightarrow e^+e^-$

High production **cross section** (1950 pb at 13 TeV).  
High **purity** thanks to the two electrons.

$$\text{Efficiency} = \frac{\text{Association track - EM cluster} + \text{track quality}}{\text{All clusters with } M_{\text{tag-probe}} \sim M_Z}$$

## 1 Electrons reconstructed as clusters with **an** associated track

### Reversed identification

Background model made by **reversed identification**:  
probe candidates failing selection

→ **Fail to pass at least 2 loose identification cuts and bad isolation**

### Normalization

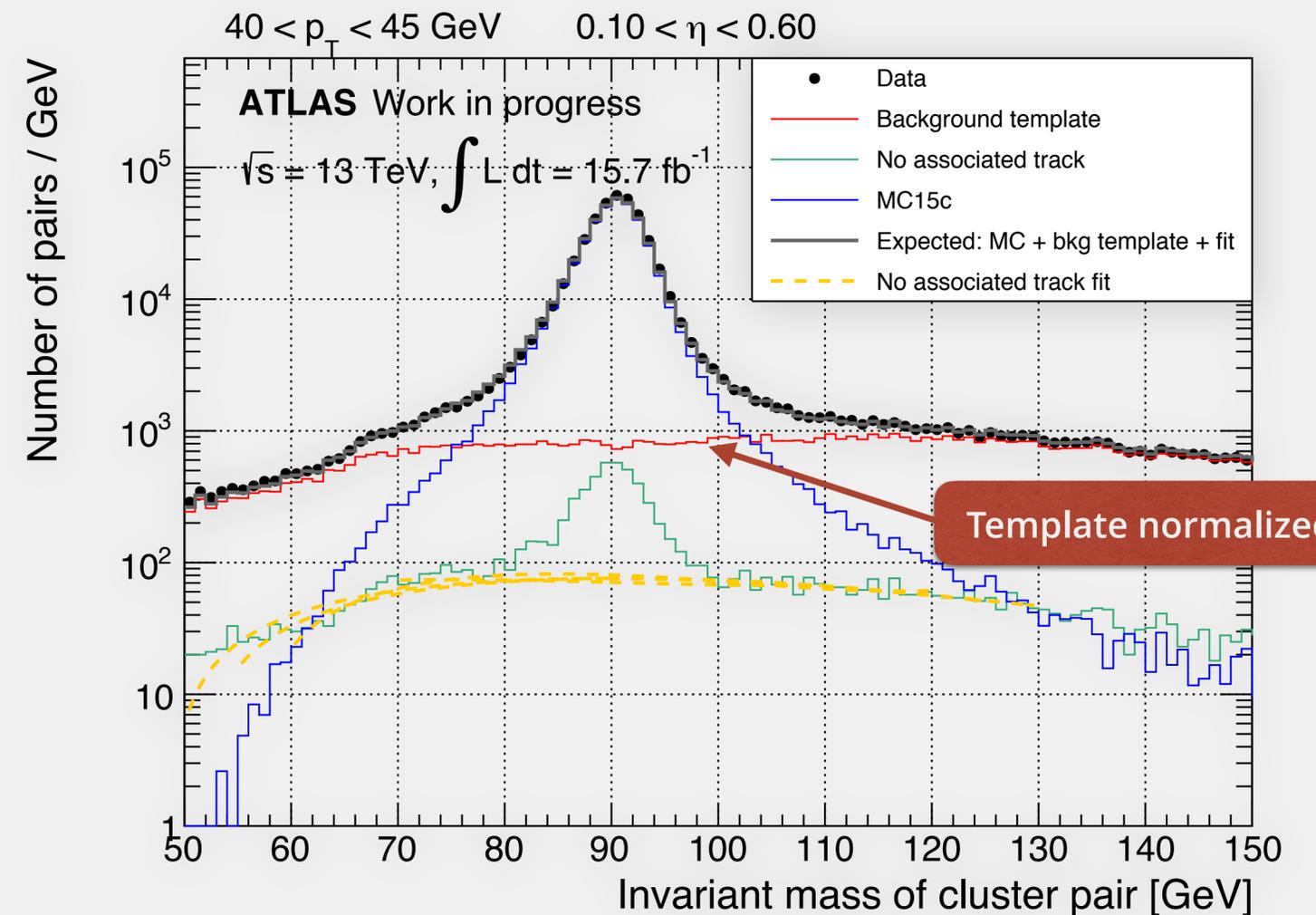
Distribution normalized to the high/low mass data  
distribution:  $M_{ee} > 120 \text{ GeV} / [60, 70] \text{ GeV}$

## 2 Electrons reconstructed as clusters with **no** associated track

Template difficult to build

→ Simpler method

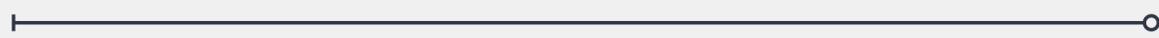
1. Construct invariant mass between electron tag and photon probe.
2. Choose **side band regions** outside the Z mass peak.
3. Signal in the side band regions subtracted from simulation
4. **Fit the ey-mass distribution** with a 3rd order polynomial.





## 1 Electrons reconstructed as clusters with **an** associated track

Reversed identification



Background model made by **reversed identification**:  
probe candidates failing selection

→ **Fail to pass at least 2 loose identification cuts and bad isolation**

Normalization



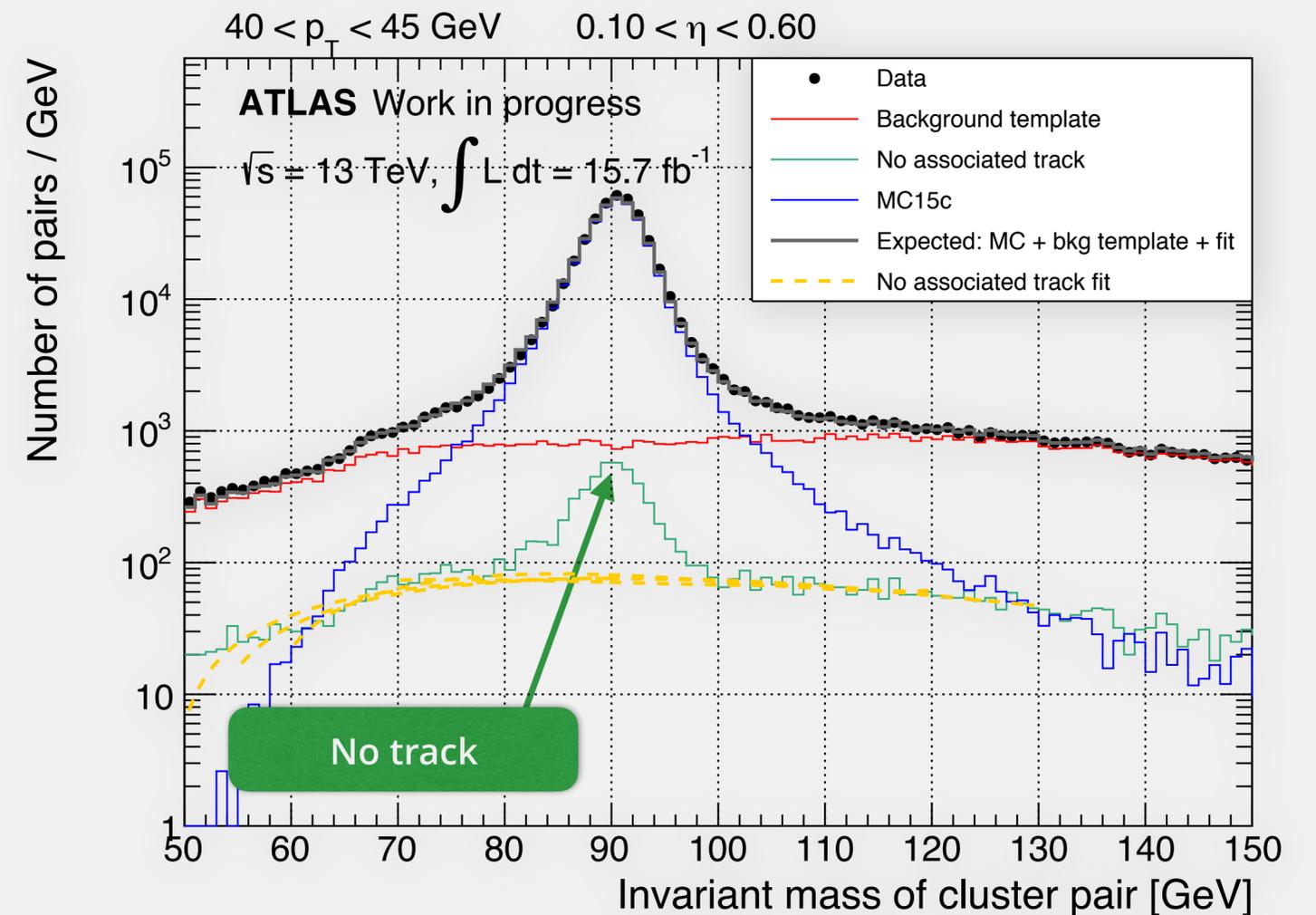
Distribution normalized to the high/low mass data  
distribution:  $M_{ee} > 120 \text{ GeV} / [60, 70] \text{ GeV}$

## 2 Electrons reconstructed as clusters with **no** associated track

Template difficult to build

→ Simpler method

1. Construct invariant mass between electron tag and photon probe.
2. Choose **side band regions** outside the Z mass peak.
3. Signal in the side band regions subtracted from simulation
4. **Fit the ey-mass distribution** with a 3rd order polynomial.



## 1 Electrons reconstructed as clusters with **an** associated track

### Reversed identification

Background model made by **reversed identification**:  
probe candidates failing selection

→ **Fail to pass at least 2 loose identification cuts and bad isolation**

### Normalization

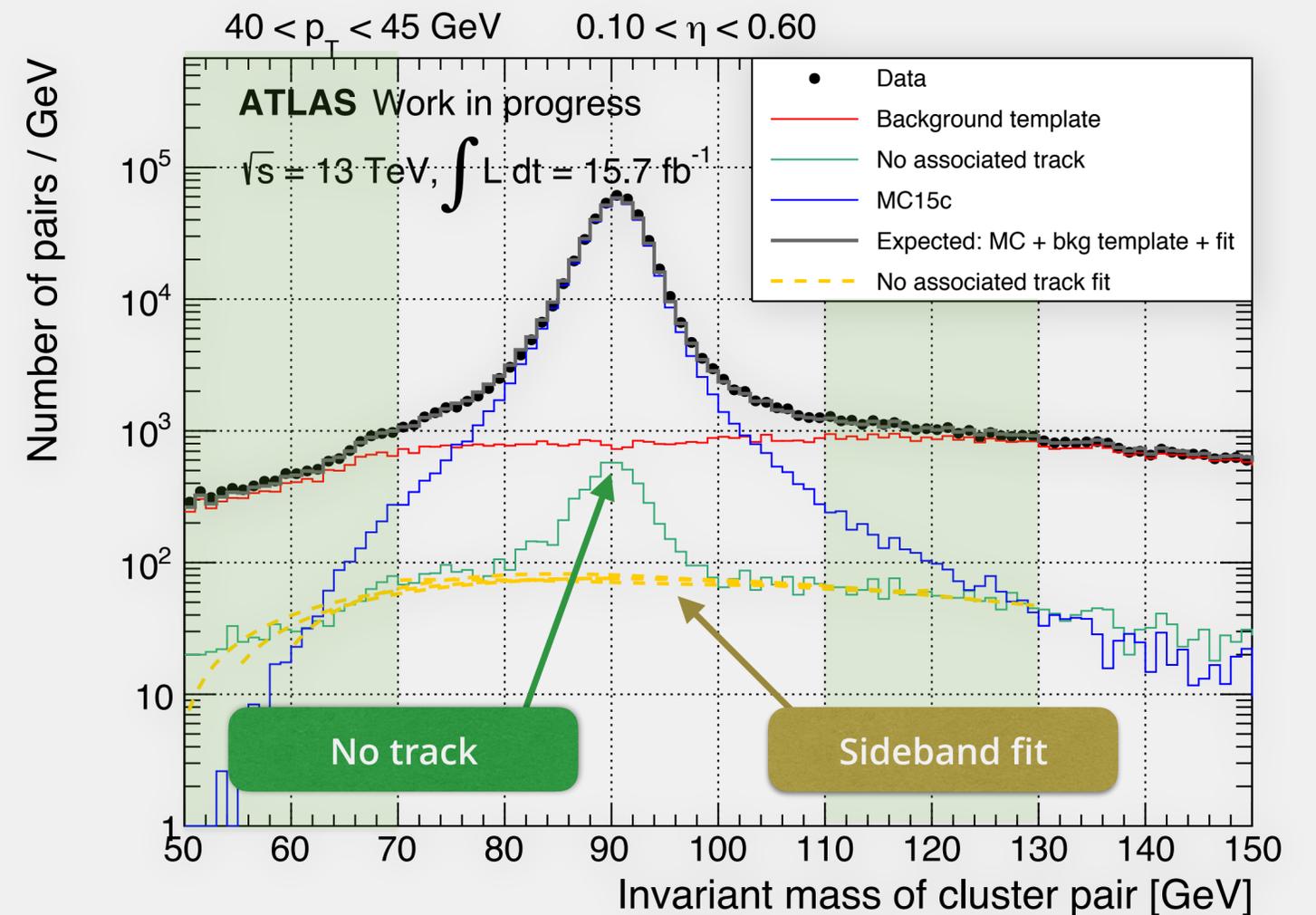
Distribution normalized to the high/low mass data  
distribution:  $M_{ee} > 120 \text{ GeV} / [60, 70] \text{ GeV}$

## 2 Electrons reconstructed as clusters with **no** associated track

Template difficult to build

→ **Simpler method**

1. Construct invariant mass between electron tag and photon probe.
2. Choose **side band regions** outside the Z mass peak.
3. Signal in the side band regions subtracted from simulation
4. **Fit the ey-mass distribution** with a 3rd order polynomial.



# Efficiency computation

Uncertainties on  $\epsilon_{\text{reco}}$



## Systematics

**Problem:** correlated

**Solution:** estimated at the same time by varying tag selection and background parameters.

### Sources of uncertainties:

- Shape of the background
- Composition of the background
- Signal contamination of background templates
- Lower efficiency at low mass (*bremsstrahlung*)
- Background shape for candidates with no associated track.



- Signal contamination in the normalization template region
- Signal contamination in the fit region

## Statistical

Determined by error propagation

### Variations applied:

- x 3 Invariant mass window for the T&P pair
- x 3 TagID = Tag identification criteria
- x 2 Definition of the background template
- x 4 Range for the fit for candidates with no associated track

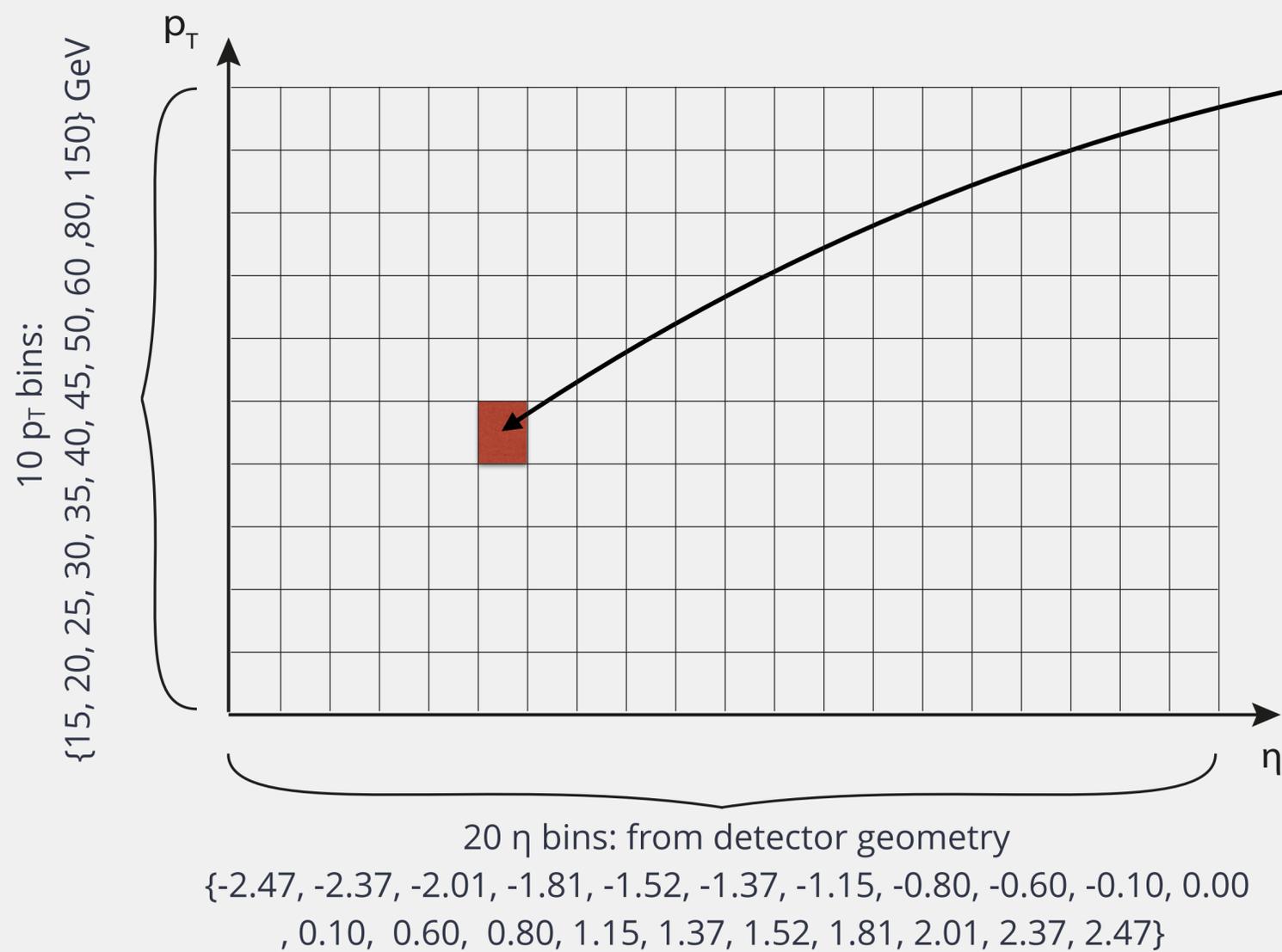
72 variations for data, 9 for MC

$$\epsilon_{\text{reco}} = \epsilon_{\text{reco}} \pm \Delta^{\text{stat}} \epsilon_{\text{reco}} \pm \Delta^{\text{syst}} \epsilon_{\text{reco}}$$

Central value of variations      Error propagation      RMS of variations



Goal: Establish a scale factor map, binned in  $p_T$  and  $\eta$ .



$$\text{scale factor} = \frac{\epsilon_{\text{data}}}{\epsilon_{\text{MC}}}$$

$$\epsilon_{\text{reco}} = \frac{N_{\text{pass}}^{\text{sig}}}{N_{\text{pass}}^{\text{sig}} + N_{\text{fail}}^{\text{sig}}} = \frac{(N_p^e - B_p^e) + (N_F^e - B_F^e) + (N_\gamma - B_\gamma)}{(N_p^e - B_p^e) + (N_F^e - B_F^e) + (N_\gamma - B_\gamma)}$$

Reconstructed candidates passing the probe track quality criteria

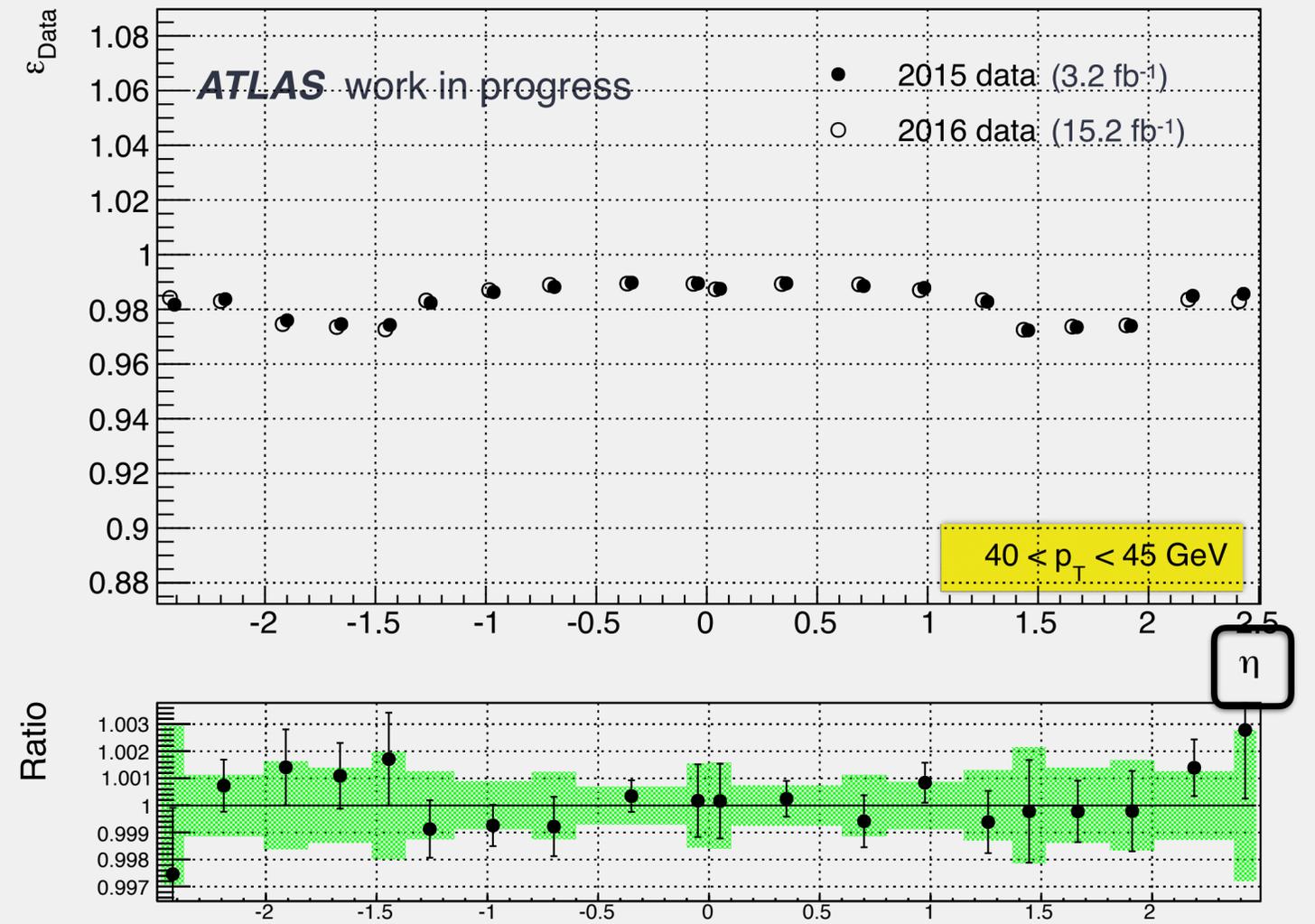
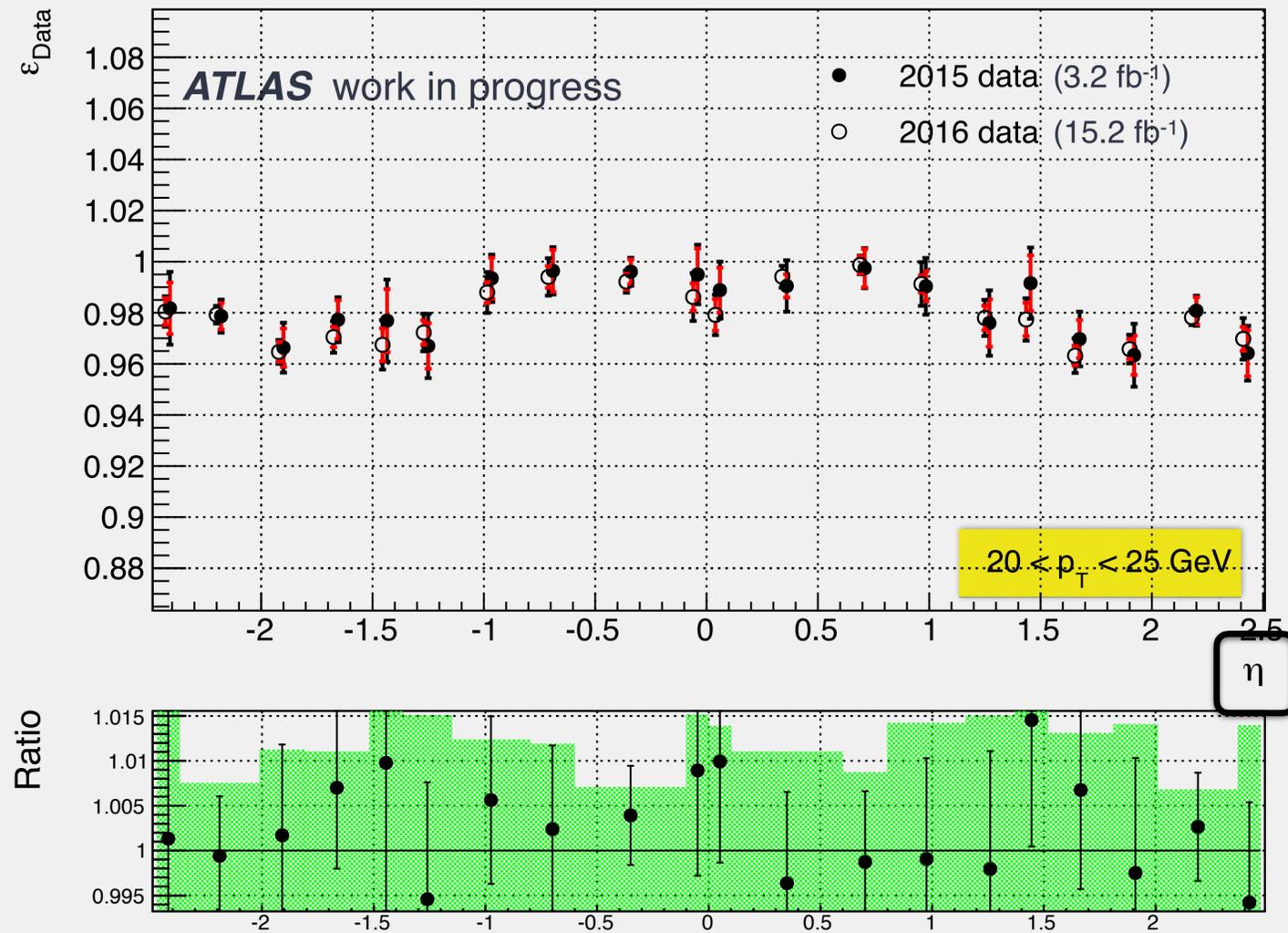
Reconstructed candidates failing track quality criteria

Non-reconstructed candidates

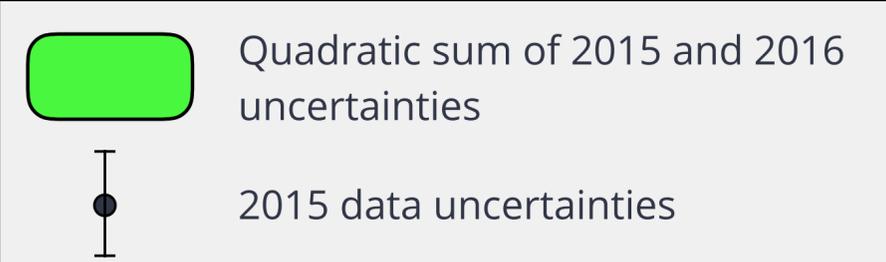
- Historically, CPPM has always been in charge of this measurement.
- Computation code transferred to a centralized framework.
- Regularly new results provided to all physics analyses + regular checks with 2016 data

# Electron reconstruction efficiency in 2016 data

Comparison with 2015 results

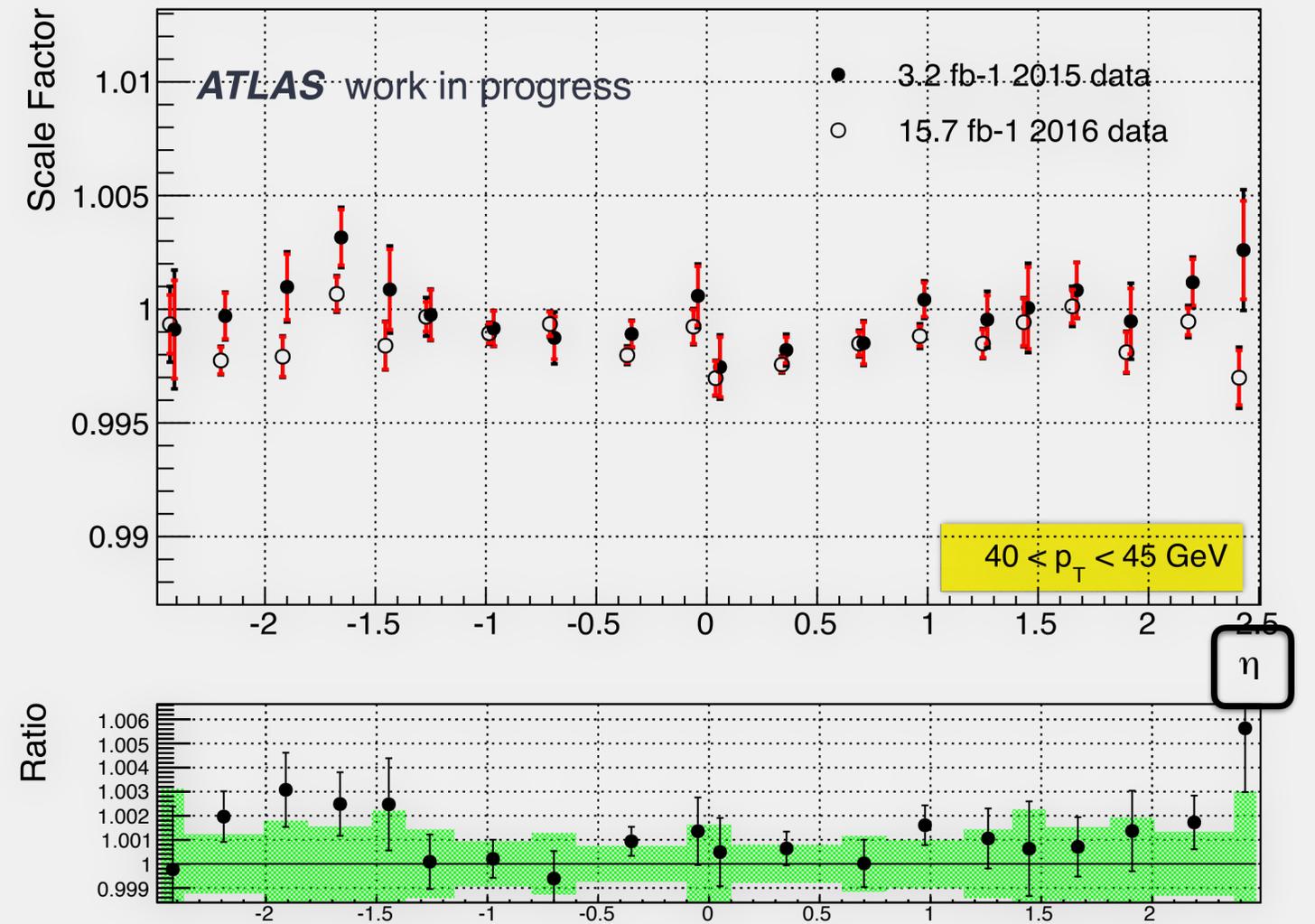
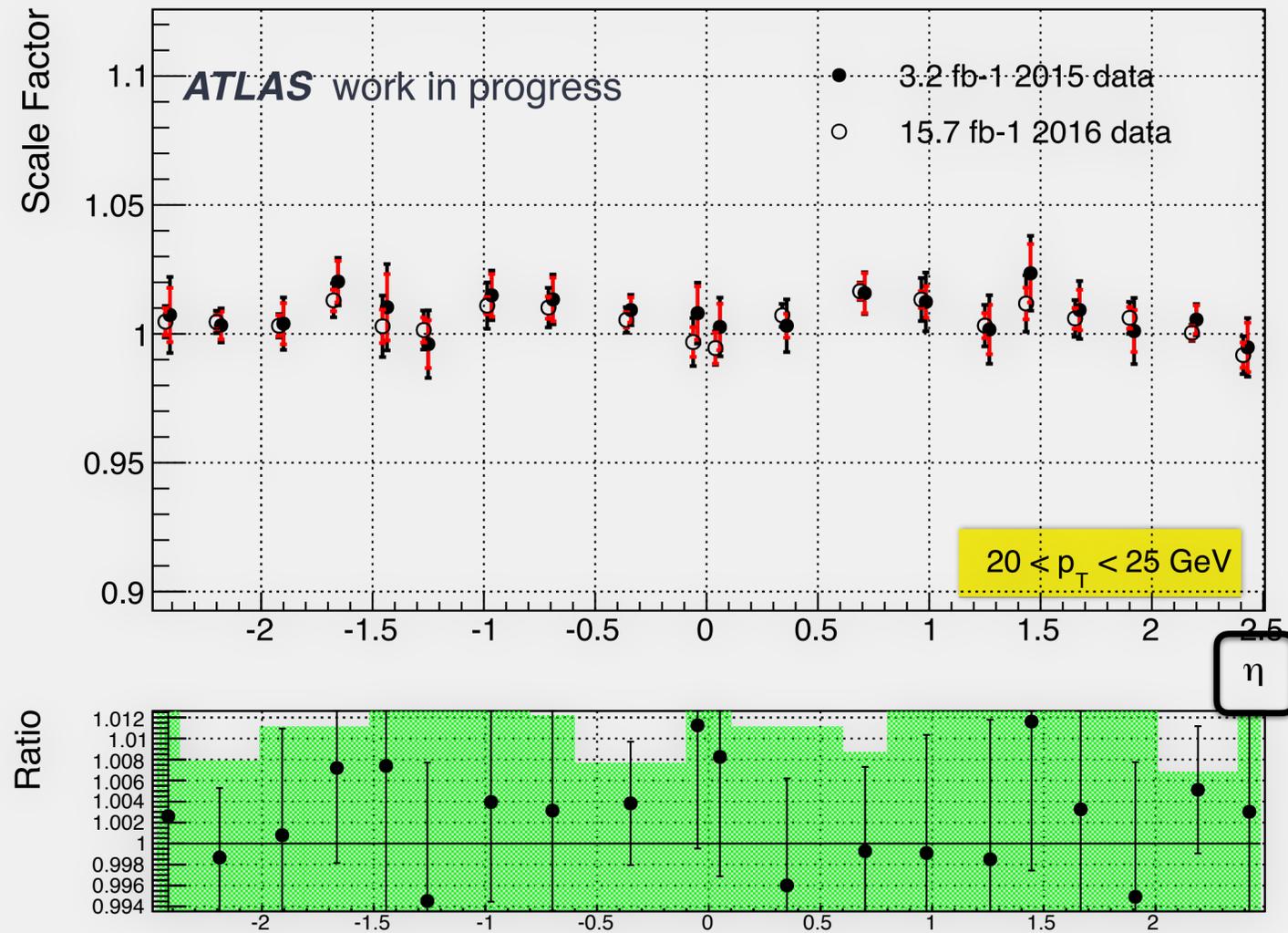


Results in 2016 compatible with 2015 data ones within error in most bins



# Electron reconstruction efficiency in 2016 data

Comparison with 2015 results



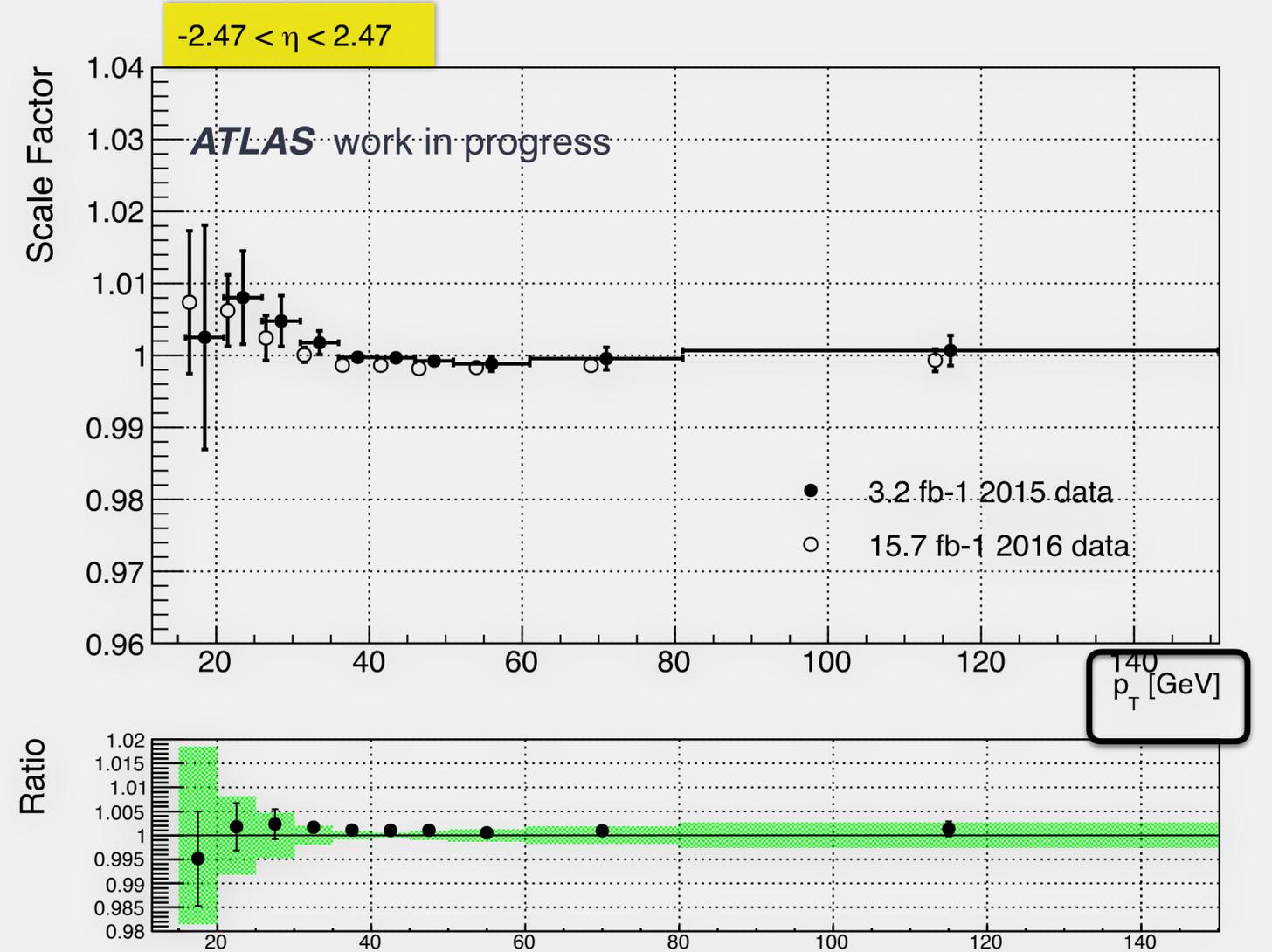
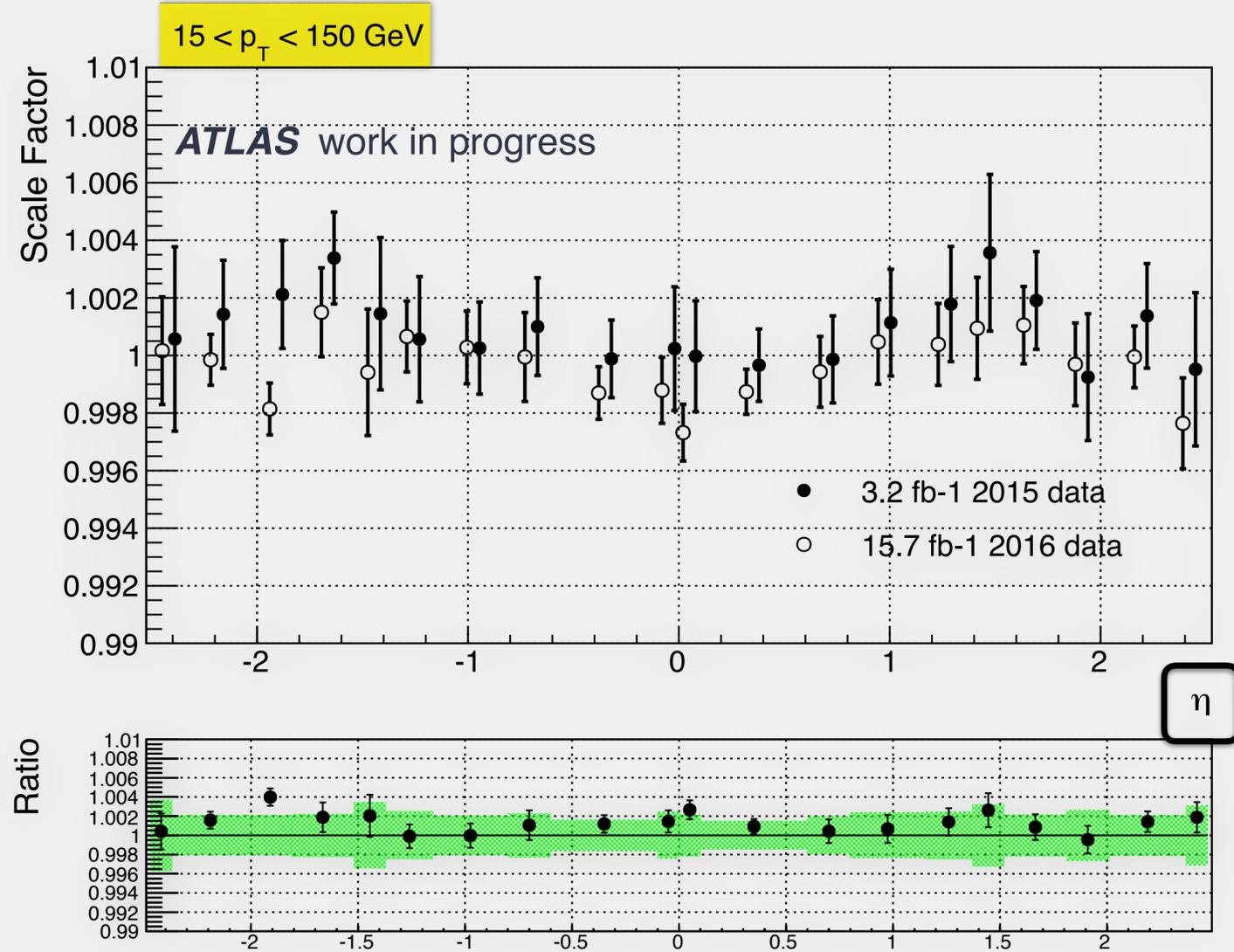
Results in 2016 compatible with 2015 data ones within error in most bins

 Quadratic sum of 2015 and 2016 uncertainties

 2015 data uncertainties

# Electron reconstruction efficiency in 2016 data

Comparison with 2015 results



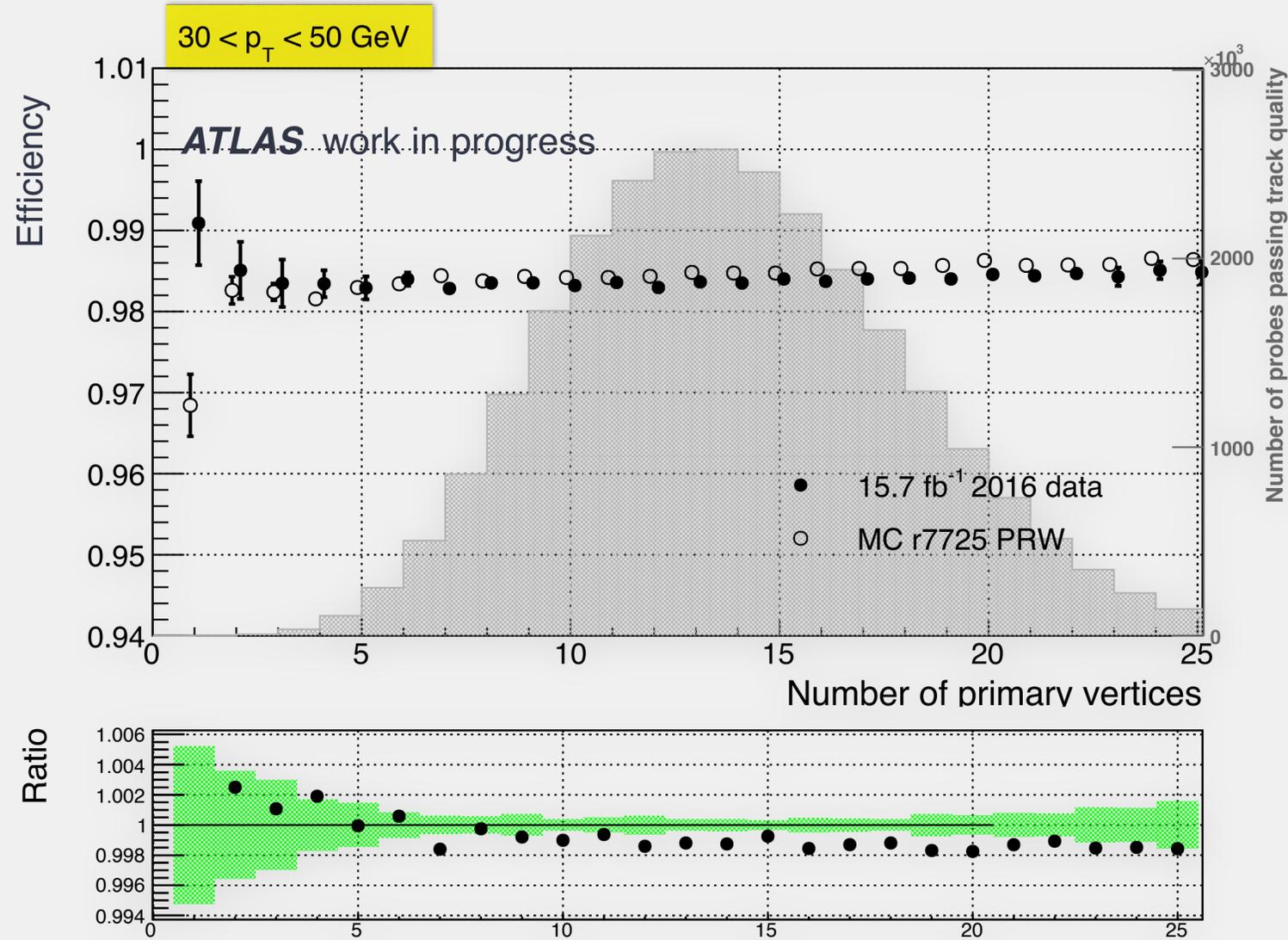
Results in 2016 compatible with 2015 data ones within error in most bins

 Quadratic sum of 2015 and 2016 uncertainties

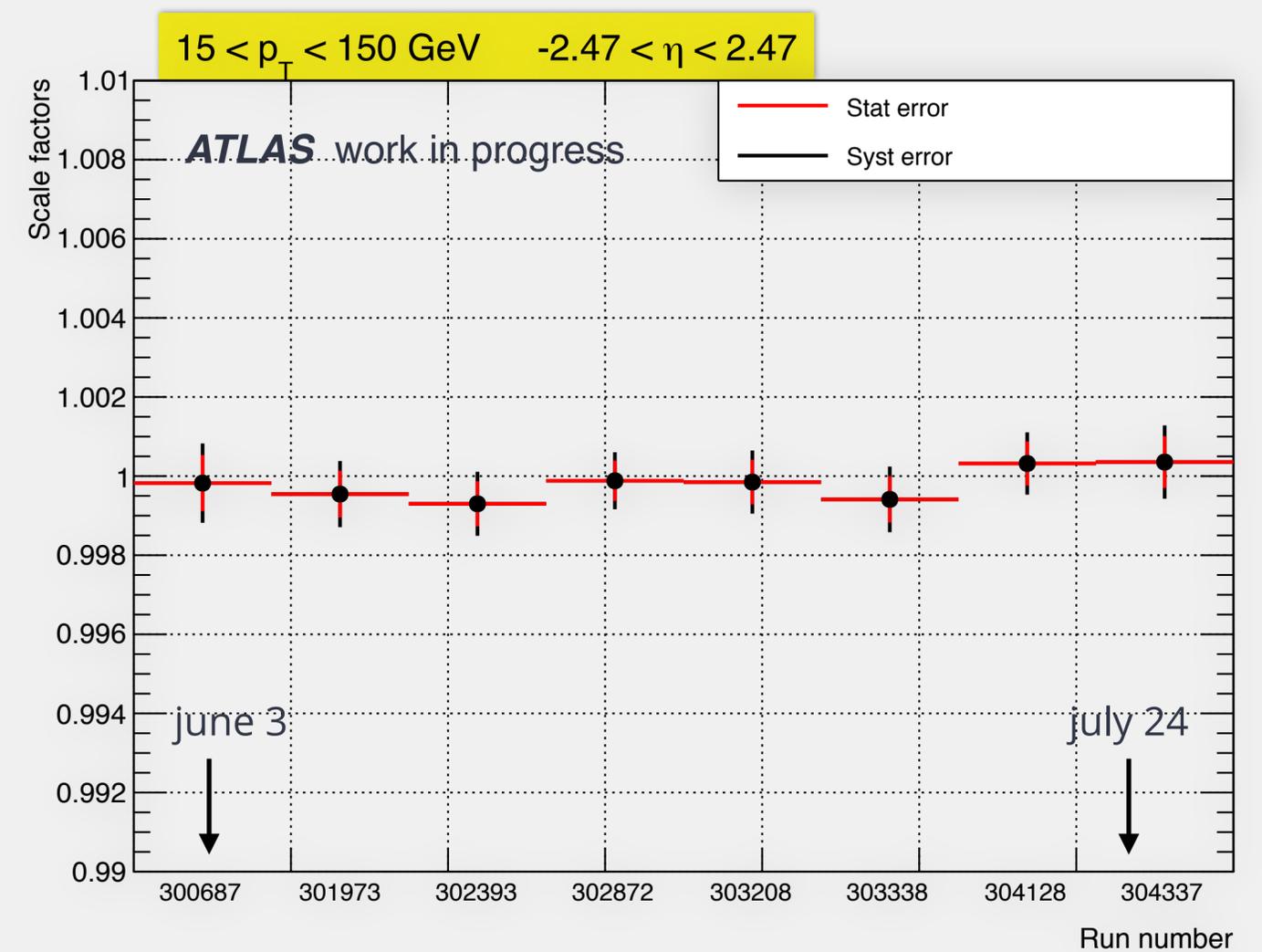
 2015 data uncertainties

# Electron reconstruction efficiency in 2016 data

Stability in pile up and time

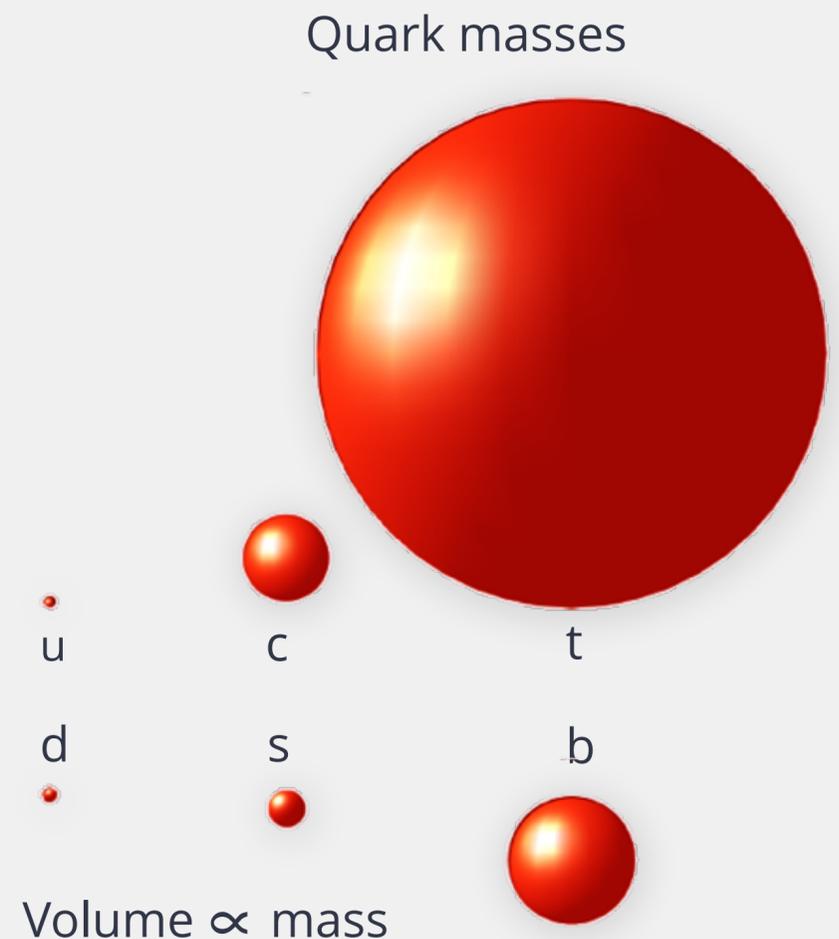


Good agreement between data and MC efficiencies  
Slight increase with pile up as expected



Scale factors stable in time

# Higgs boson coupling to the top quark with the two same sign leptons channel

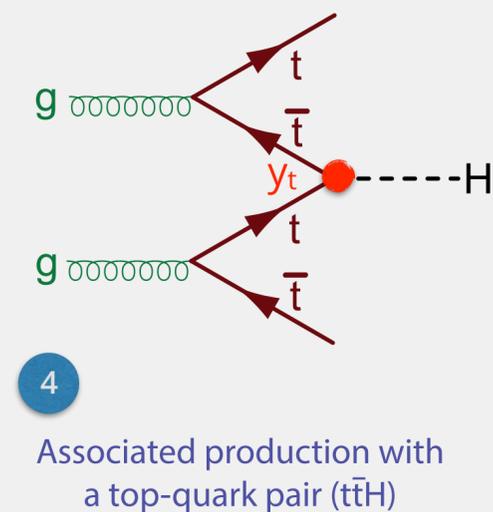
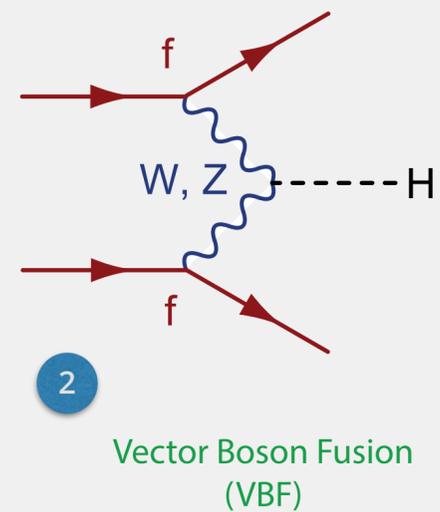
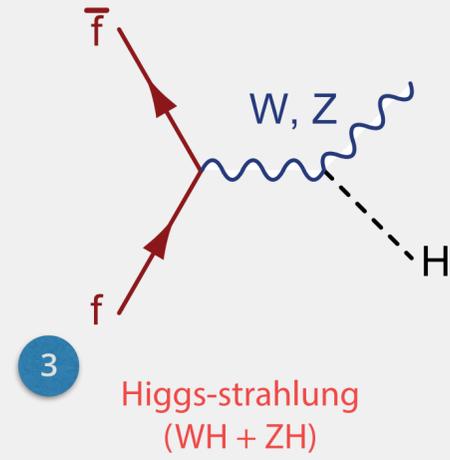
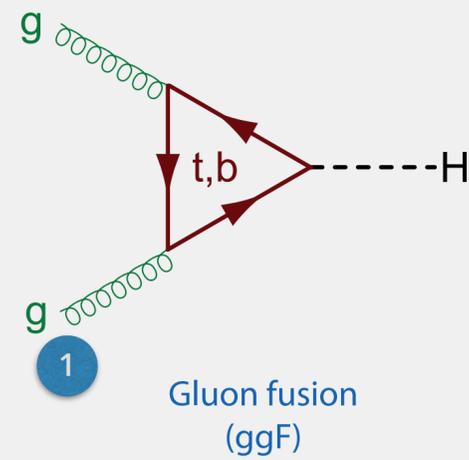


## Why this study ?

- Mass hierarchy problem : one example of Standard Model incompleteness
- Top quark - Higgs coupling :  $y_t = \sqrt{2}m_t/V_{ev} \sim 1$
- Higgs properties : towards New Physics : supersymmetry ? Dark matter candidates ?
- Electroweak symmetry breaking understanding

# Higgs boson

## Production



### 4 production modes at LHC

- **Gluon fusion**: allows an indirect measurement of the top - Higgs coupling
- **Vector boson fusion**: no coupling
- **Higgs-strahlung**: no coupling
- **ttH production**: allows a direct measurement

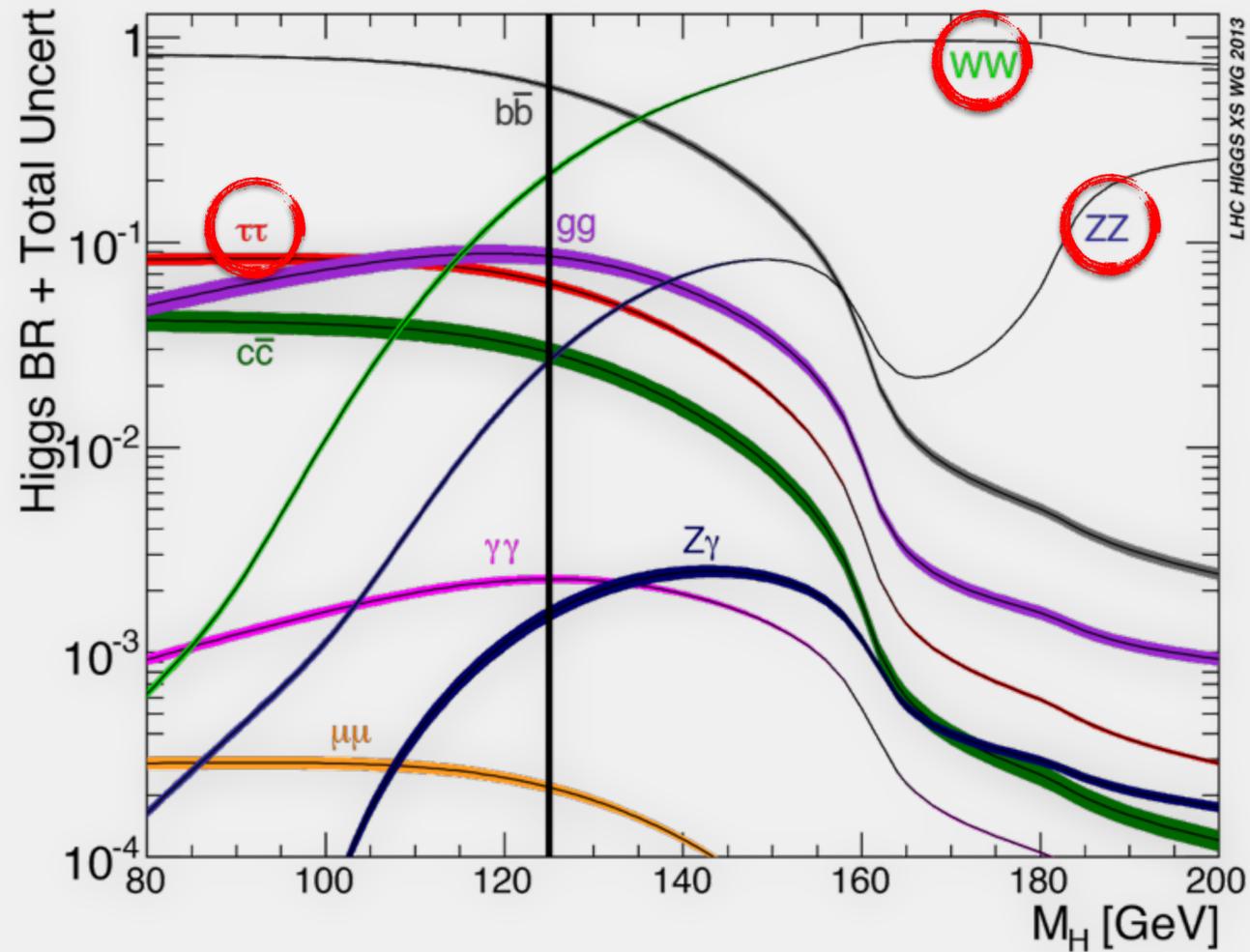
### Low production cross sections

	ggF	WH+ZH	VBF	ttH
Run 1 8 TeV (pb)	19.3	1.12	1.6	0.13
Run 2 13 TeV (pb)	43.9	2.27	3.8	0.51

2 orders of magnitude

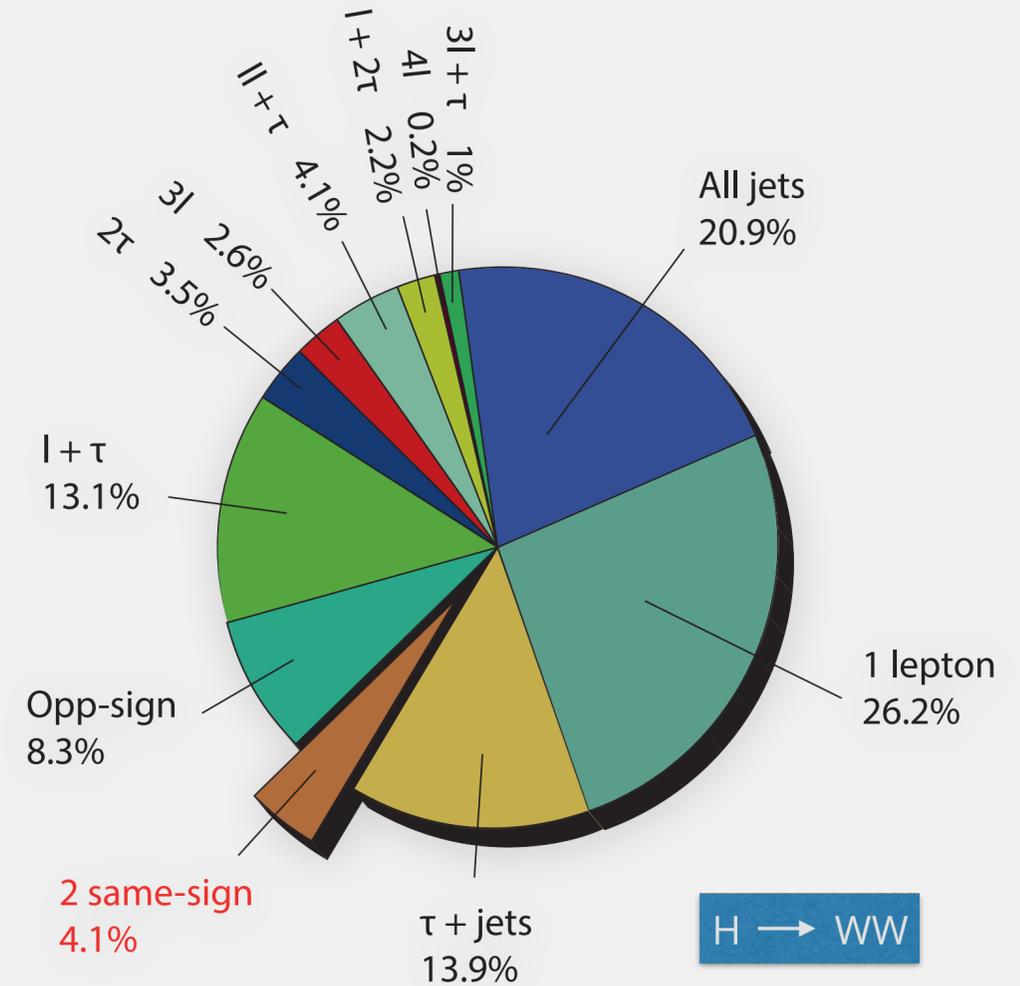
# Higgs boson

Decay



$H \rightarrow WW$  (22%), also  $H \rightarrow \tau\tau$  (6.3%) and  $H \rightarrow ZZ$  (2.6%)

$W \rightarrow l\nu$  (21%),  $\tau \rightarrow l\nu\nu$  (35.2%),  $Z \rightarrow ll$  (6.7%) for  $l = e, \mu$

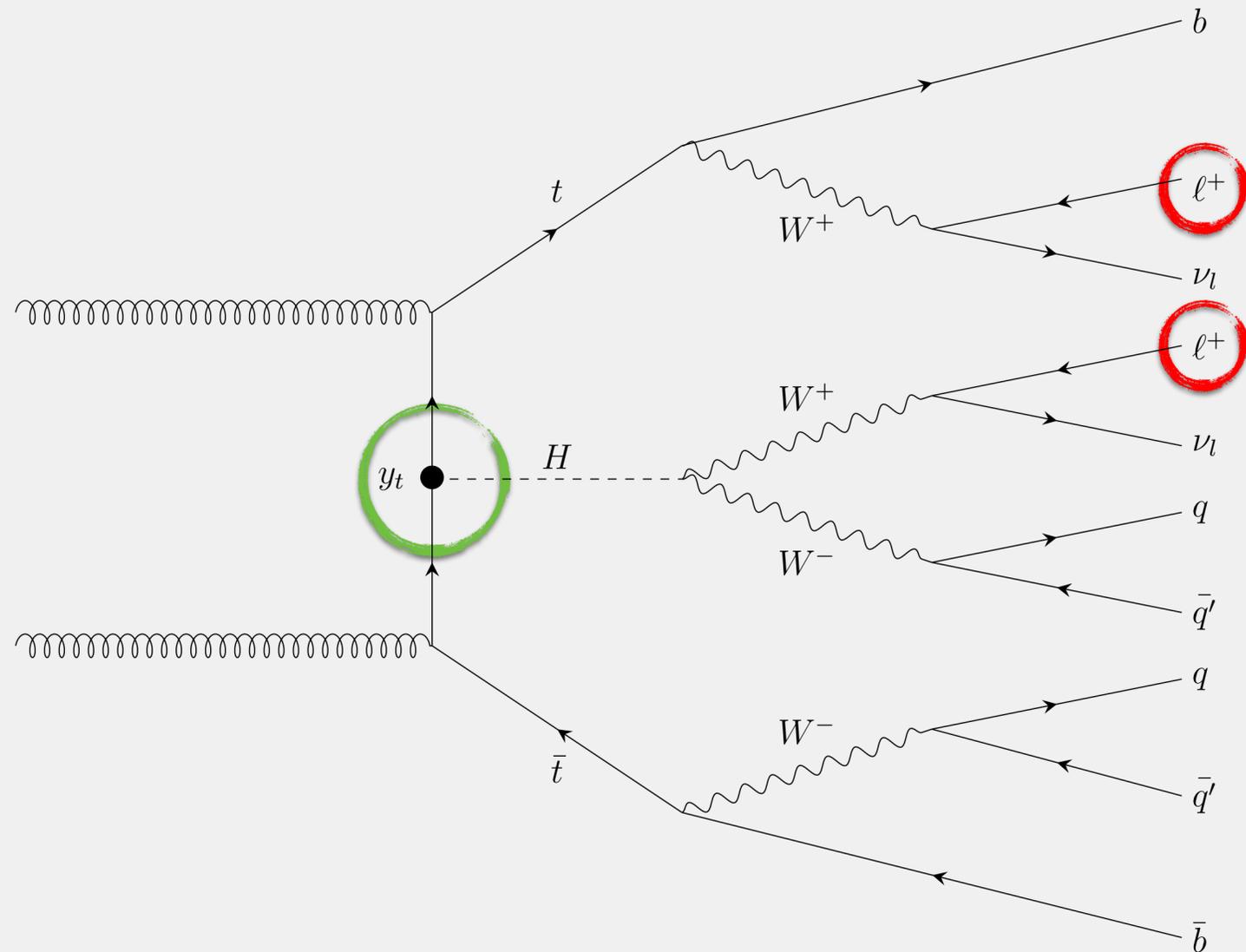


For  $\sqrt{s} = 8$  TeV:  $O(20)$  events,  $\sim 500$  at 13 TeV with  $100 \text{ fb}^{-1}$

$$N_{\text{sig, run2}} = \underbrace{510 \text{ fb}}_{\sigma_{t\bar{t}H(WW)}} \times \underbrace{0.215}_{\text{BR}(H \rightarrow WW)} \times \underbrace{0.041}_{\text{Two same-sign leptons}} \sim 4.3 \text{ events}/\text{fb}^{-1}$$

# ttH in same-sign leptons channel

ttH signal



$H \rightarrow WW$  (22%), also  $H \rightarrow \tau\tau$  (6.3%) and  $H \rightarrow ZZ$  (2.6%)

$W \rightarrow l\nu$  (21%),  $\tau \rightarrow l\nu\nu$  (35.2%),  $Z \rightarrow ll$  (6.7%) for  $l = e, \mu$

## Characteristics

- Three channels :  $ee, e\mu, \mu\mu$
- 6 jets including 2 b-jets
- Missing transverse energy due to neutrinos

## A clear signature

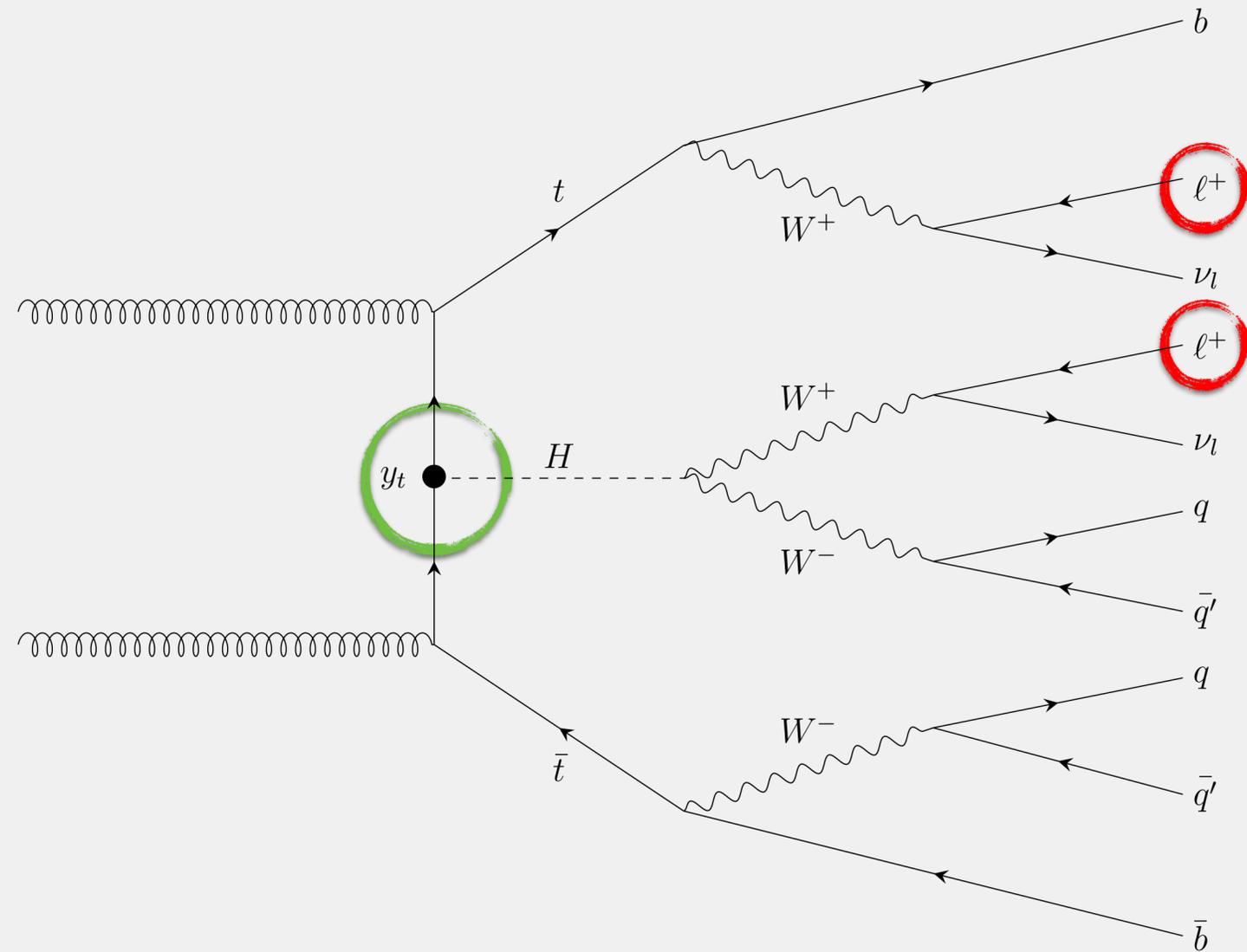
- Standard Model background with a two same-sign leptons final state low but irreducible
- Fake leptons background (mainly from heavy flavour decays) and leptons charge-flip dominant but reducible.

Cross sections (fb) (NLO)	8 TeV	13 TeV
ttH	130	509

$\sigma \times 3,9$

# ttH in same-sign leptons channel

Fake leptons estimation

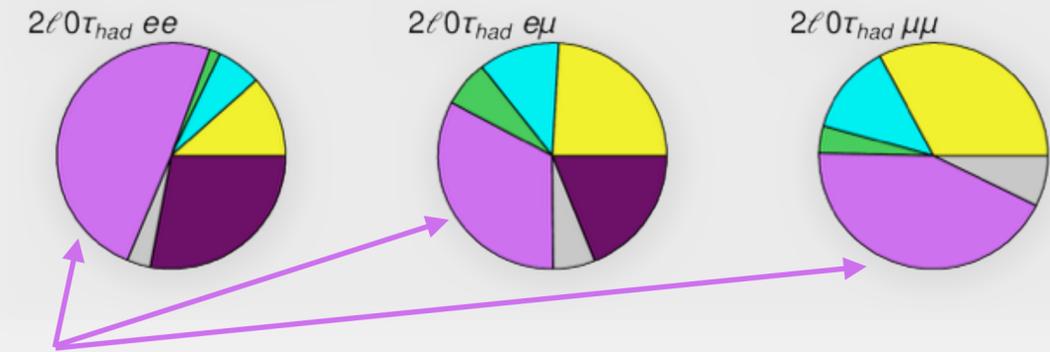


$H \rightarrow WW$  (22%), also  $H \rightarrow \tau\tau$  (6.3%) and  $H \rightarrow ZZ$  (2.6%)

$W \rightarrow l\nu$  (21%),  $\tau \rightarrow l\nu\nu$  (35.2%),  $Z \rightarrow ll$  (6.7%) for  $l = e, \mu$

**ATLAS** Simulation Preliminary  
 $\sqrt{s} = 13 \text{ TeV}$   
 Background composition

■ QMisReco    ■ Other  
■ Non-prompt    ■ Diboson  
■  $t\bar{t}(Z/\gamma^*)$     ■  $t\bar{t}W$



Fake leptons:

- Non negligible contribution to all channels
- Dominate systematics

What are they:

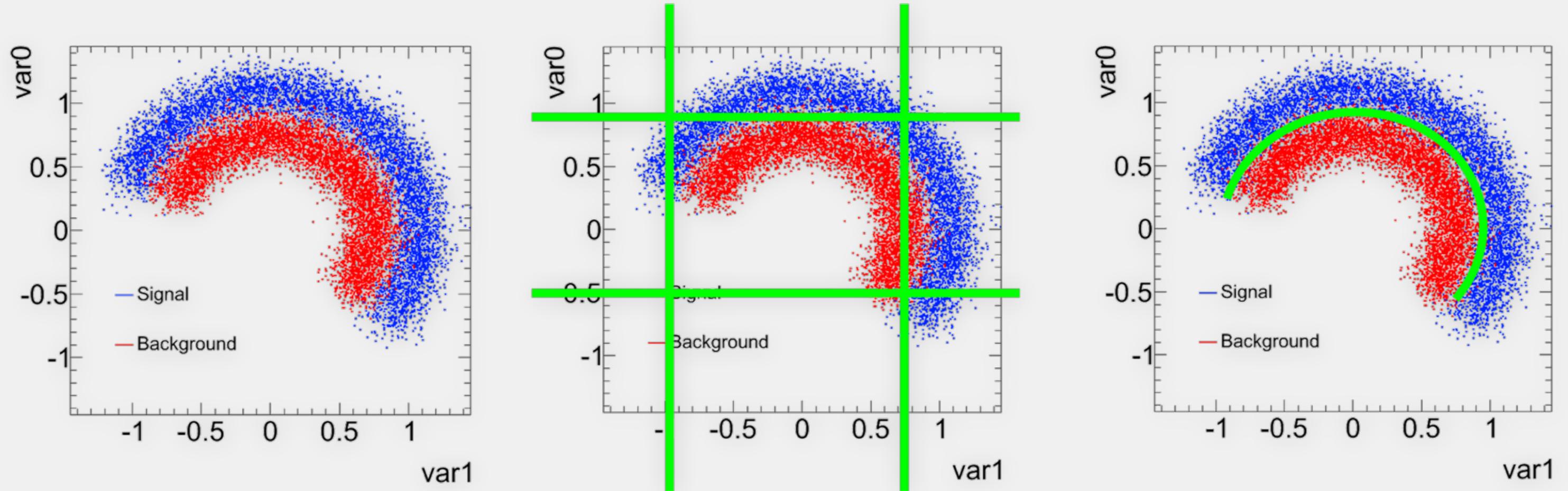
- Instrumental background: misreconstructed object as leptons
- non prompt leptons decaying from heavy hadrons
- electrons from photon conversions



**Data driven method** to estimate the amount of fake leptons, the Matrix Method.  
 Participated to the framework development.

# Multivariate analysis (MVA)

Principle



Lets consider the simple case where we have 2 discriminating variables var0 and var1.

To discriminate the signal and backgrounds one could apply selections on variables to reject as much red points as possible.

→ Unfortunately, selections only allow to do rectangular cuts without considering correlations (middle)

One should use multivariate analysis methods such as neural networks or boosted decision trees (BDT)

→ Better selection (right).

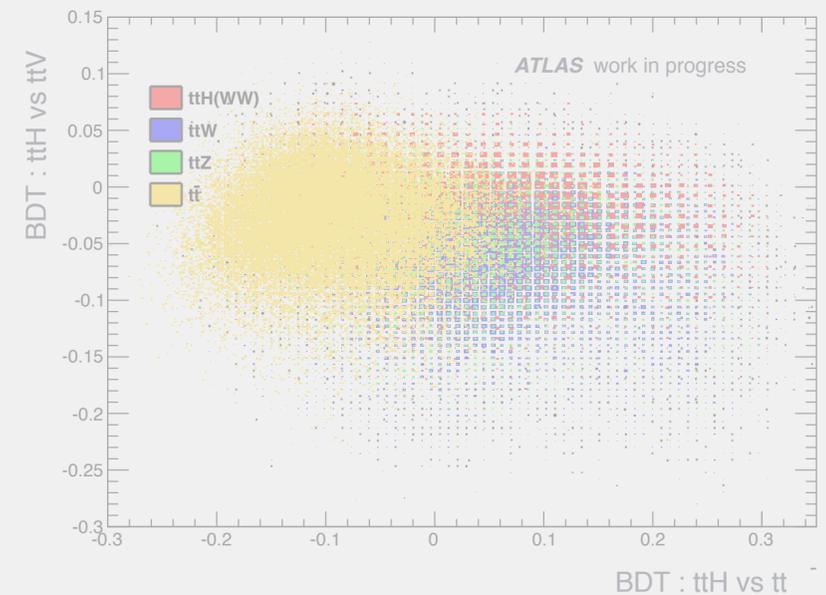
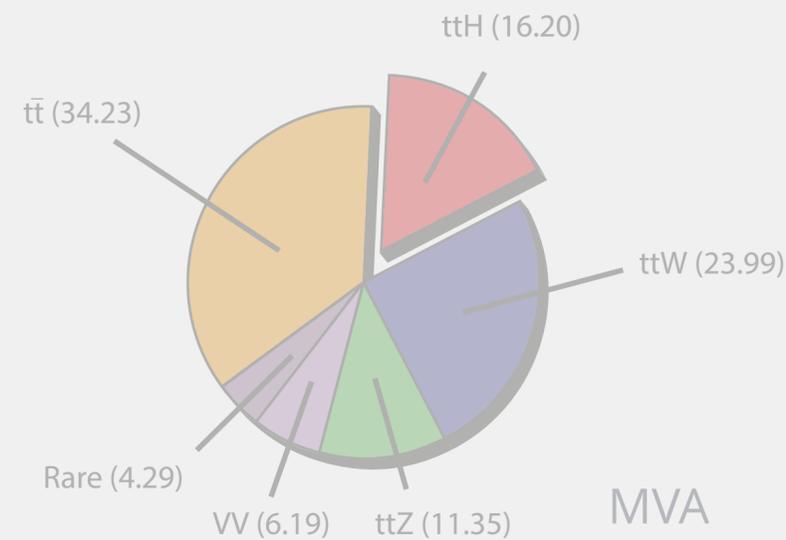
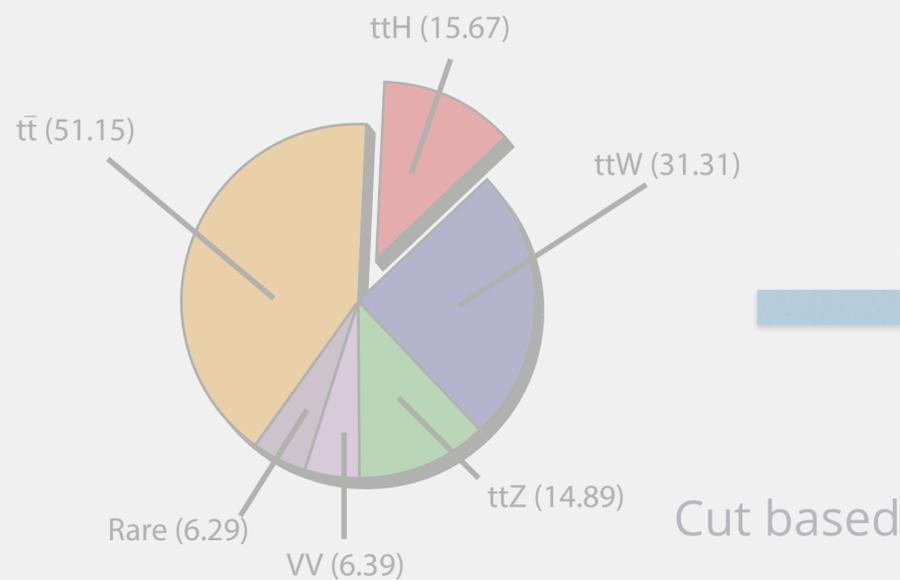
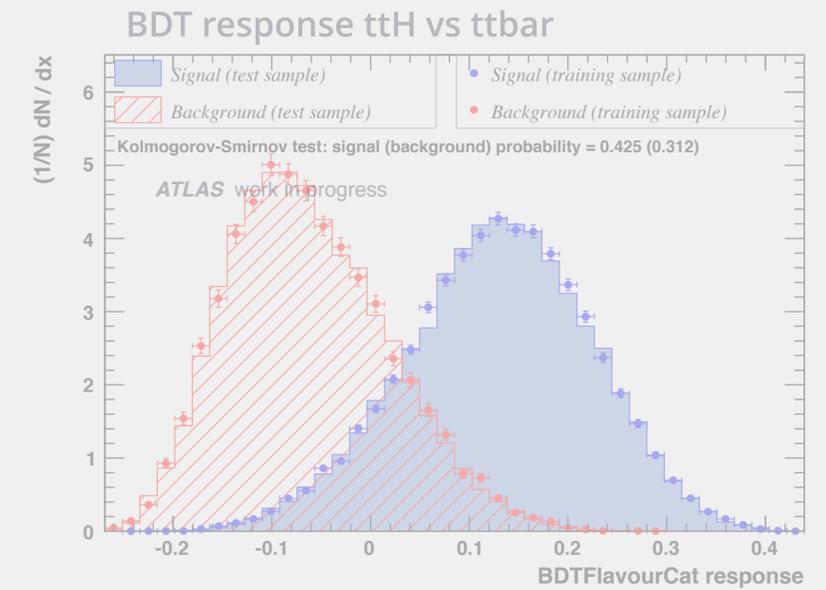
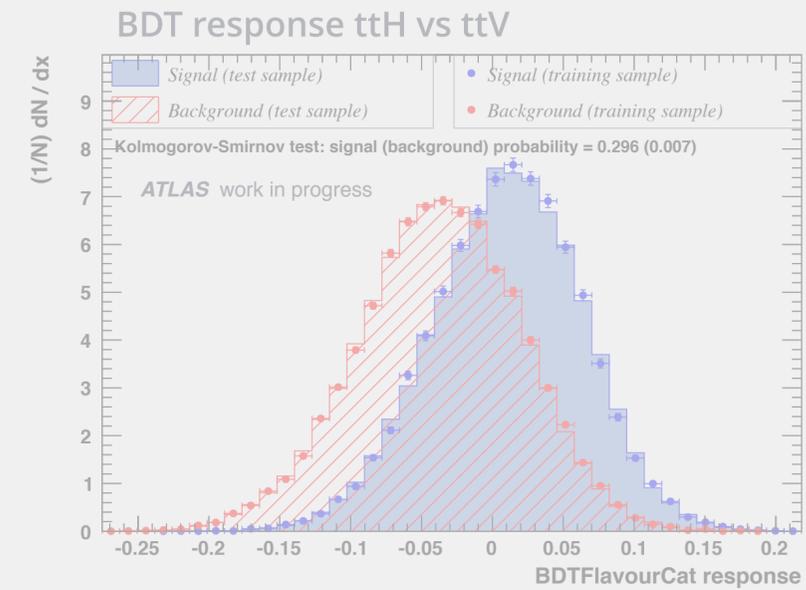
# Multivariate analysis (MVA)

Multivariate analysis in ttH with same-sign leptons channel



First MVA trial in the ttH multileptons group to replace signal regions.

- separate ttH from fakes and ttW + ttZ (ttV) in two separate BDT
- comparison of ICHEP cut based selection and MVA with the same signal efficiency
- results presented in the [HTop workshop](#) (3-4 Oct 2016)



Optimal cut keeping cut based signal efficiency can reduce ttbar by ~35% and ttV by ~20%. Still can be optimized.

Towards a publication around summer 2017

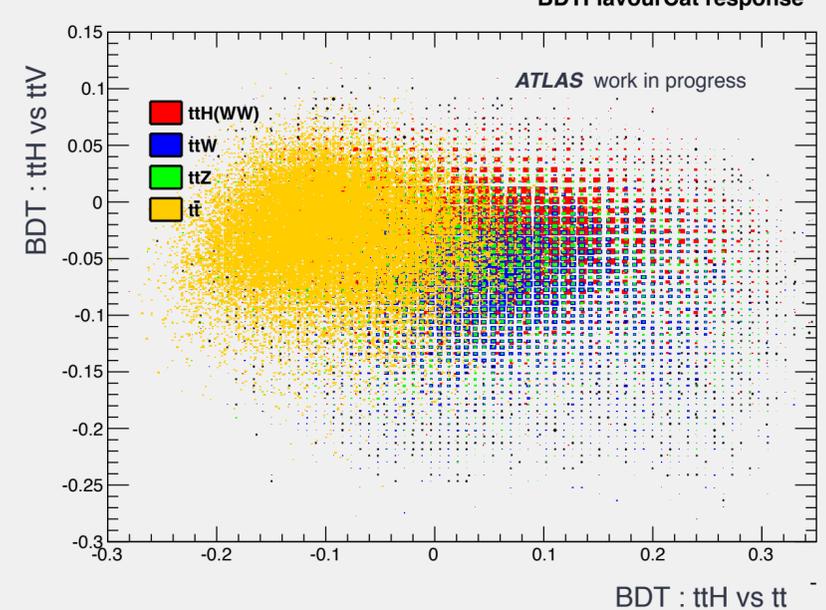
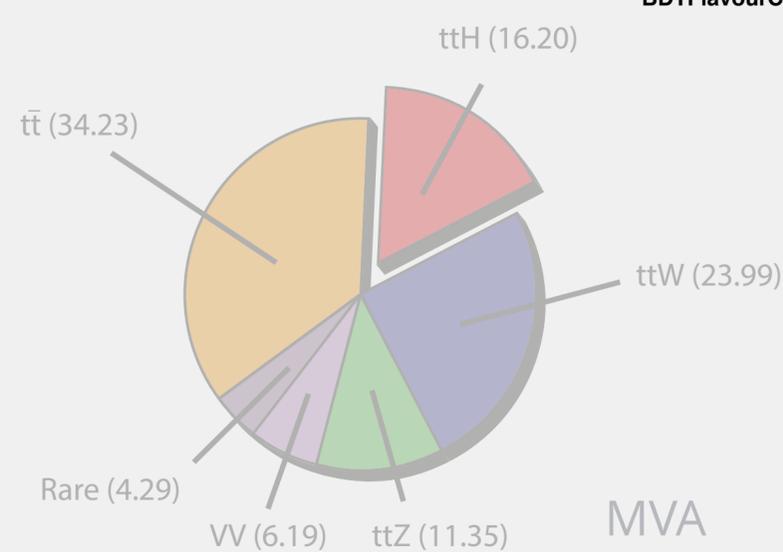
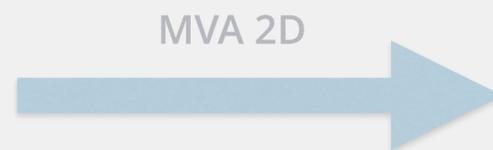
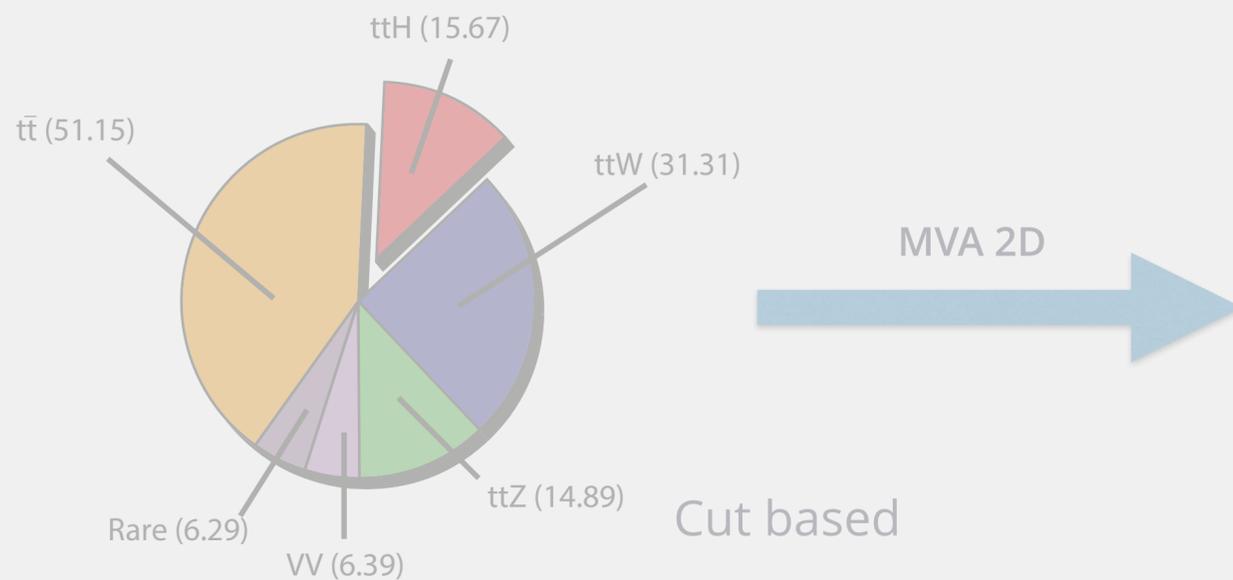
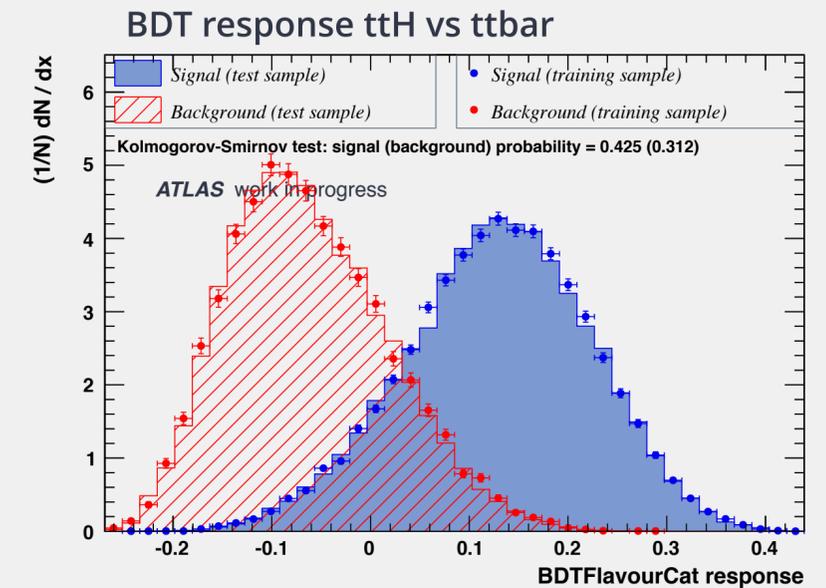
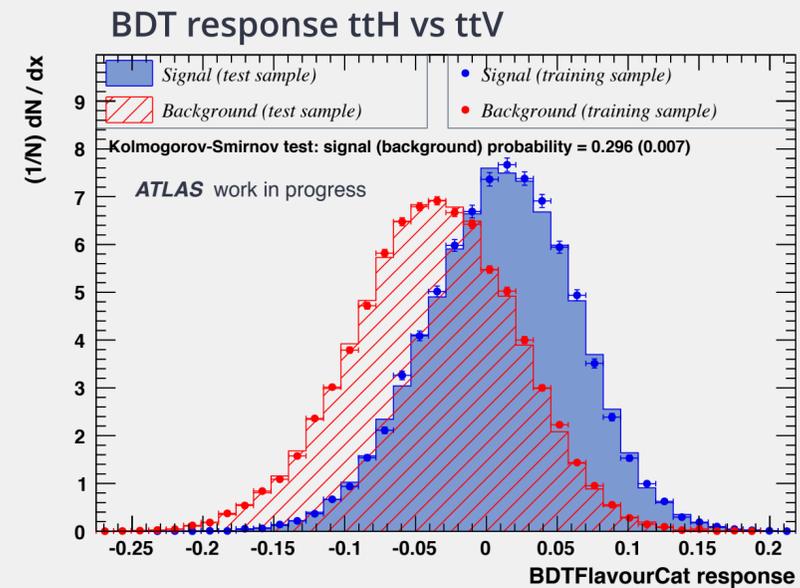
# Multivariate analysis (MVA)

Multivariate analysis in ttH with same-sign leptons channel



First MVA trial in the ttH multileptons group to replace signal regions.

- separate ttH from fakes and ttW + ttZ (ttV) in two separate BDT
- comparison of ICHEP cut based selection and MVA with the same signal efficiency
- results presented in the [HTop workshop](#) (3-4 Oct 2016)



Optimal cut keeping cut based signal efficiency can reduce ttbar by ~35% and ttV by ~20%. Still can be optimized.

Towards a publication around summer 2017

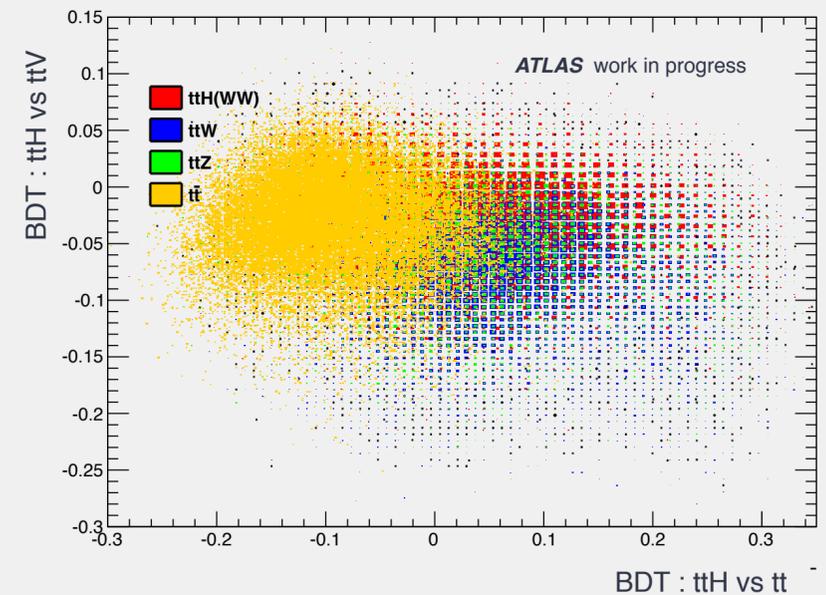
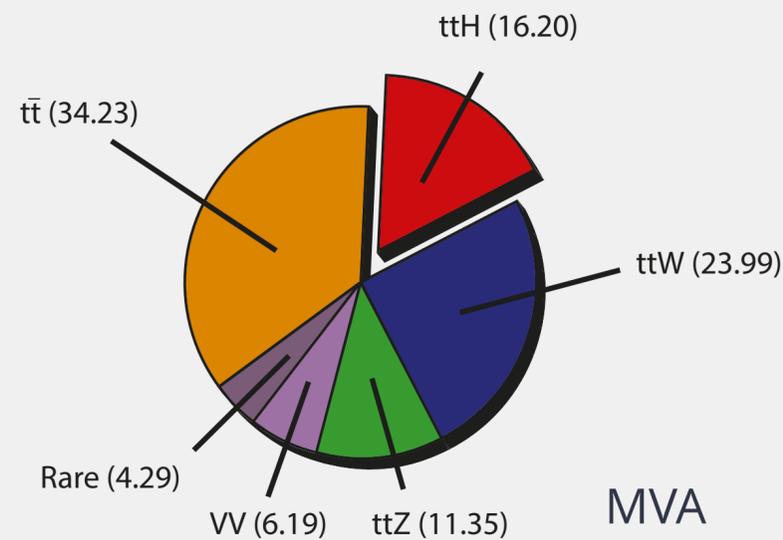
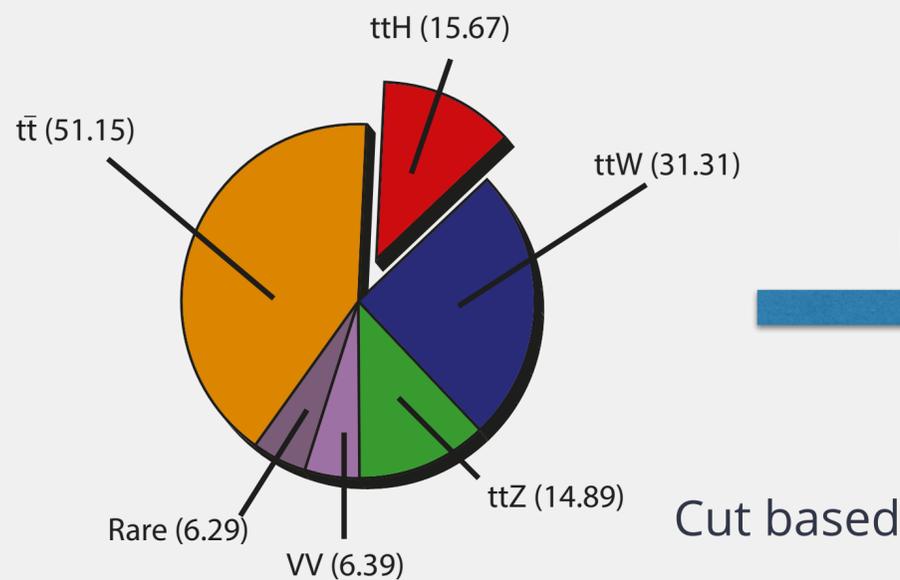
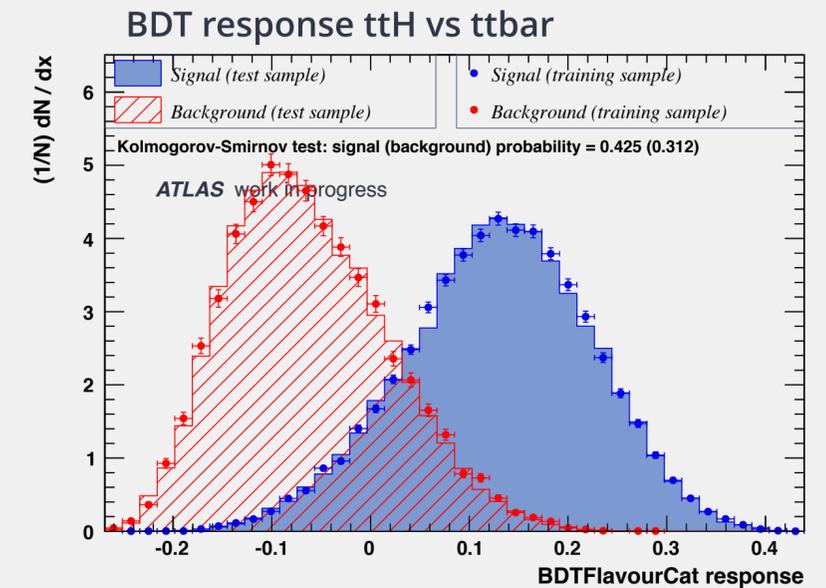
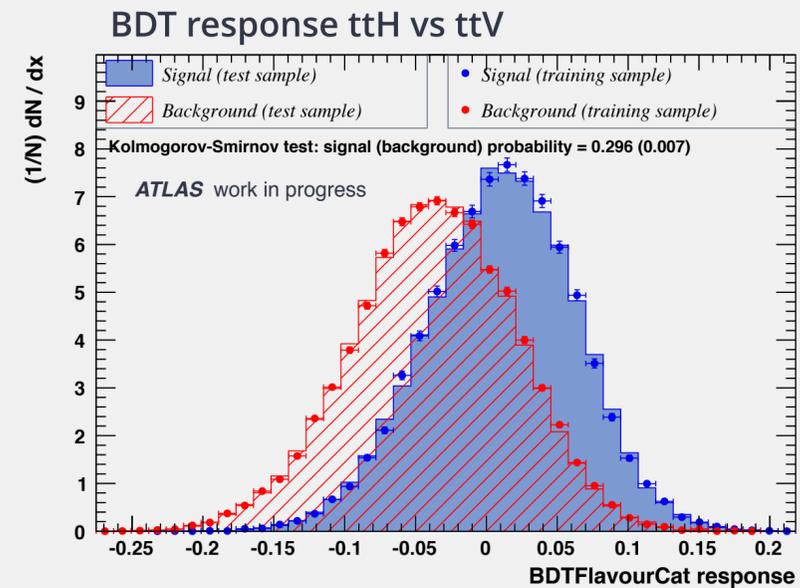
# Multivariate analysis (MVA)

Multivariate analysis in ttH with same-sign leptons channel



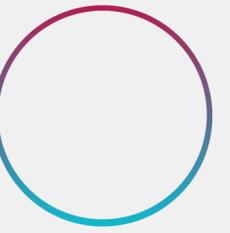
First MVA trial in the ttH multileptons group to replace signal regions.

- separate ttH from fakes and ttW + ttZ (ttV) in two separate BDT
- comparison of ICHEP cut based selection and MVA with the same signal efficiency
- results presented in the [HTop workshop](#) (3-4 Oct 2016)

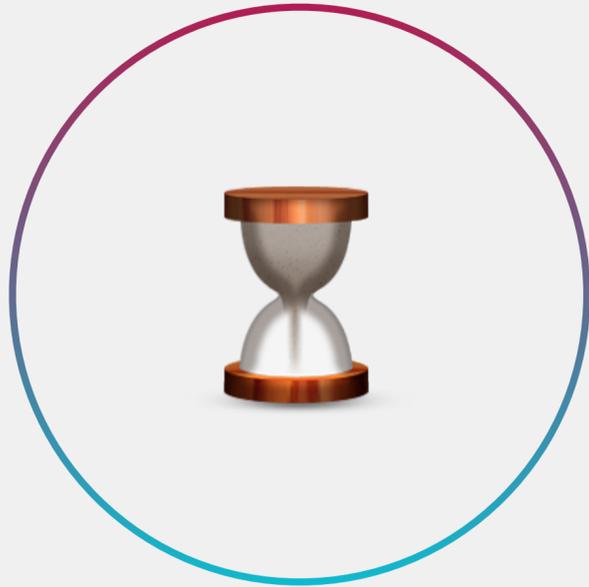


Optimal cut keeping cut based signal efficiency can reduce ttbar by ~35% and ttV by ~20%. Still can be optimized.

Towards a publication around summer 2017



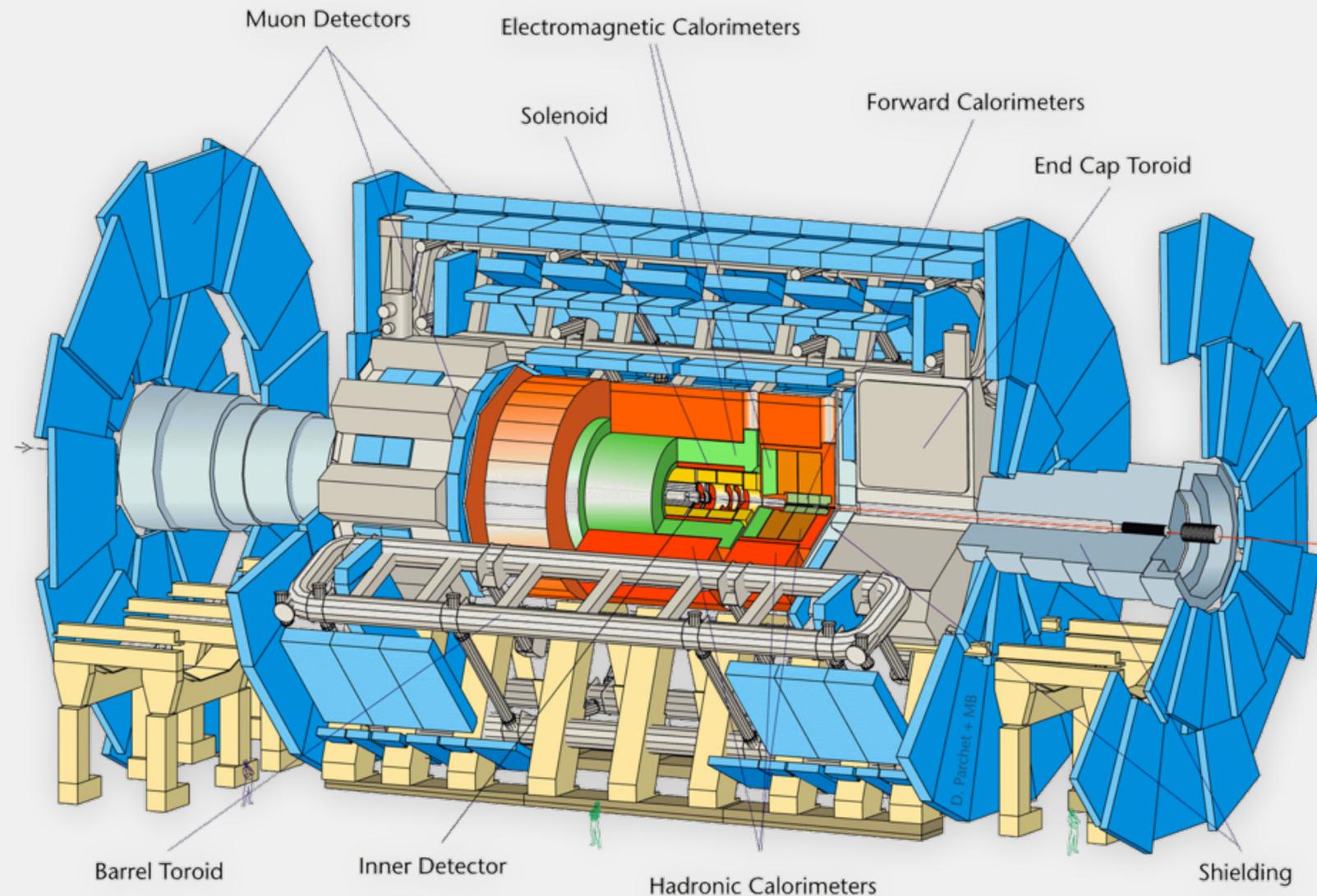
- ▶ Electron reconstruction efficiency measurement = ATLAS qualification task (done)
  - Migrated the code to whole new framework.
  - Participated in 2015 measurements made public in [ATLAS-CONF-2016-024](#)
  - Measurements regularly performed with 2016 data.
  - Under study : improvements in the measurement methodology to further reduce the systematics
  - Results used by all physics analyses of the ATLAS experiment, They will be part of a paper foreseen next spring.
- ▶ Estimate of fake lepton background in ttH with two same sign leptons
  - First results at  $\sqrt{s} = 13$  TeV have gone public last summer at ICHEP conference ([ATLAS-CONF-2016-058](#))
- ▶ Multivariate analysis
  - ttbar reduced by 35% and ttV by 20%  $\longrightarrow$  further used to compute the signal strength  $\mu_{ttH}$
  - Towards a publication around summer 2017



Backup

# Experimental background

ATLAS detector



## Inner detector

- Reconstruction and identification of charged particle tracks
- Trajectory curved by solenoidal magnets.

## Calorimeters

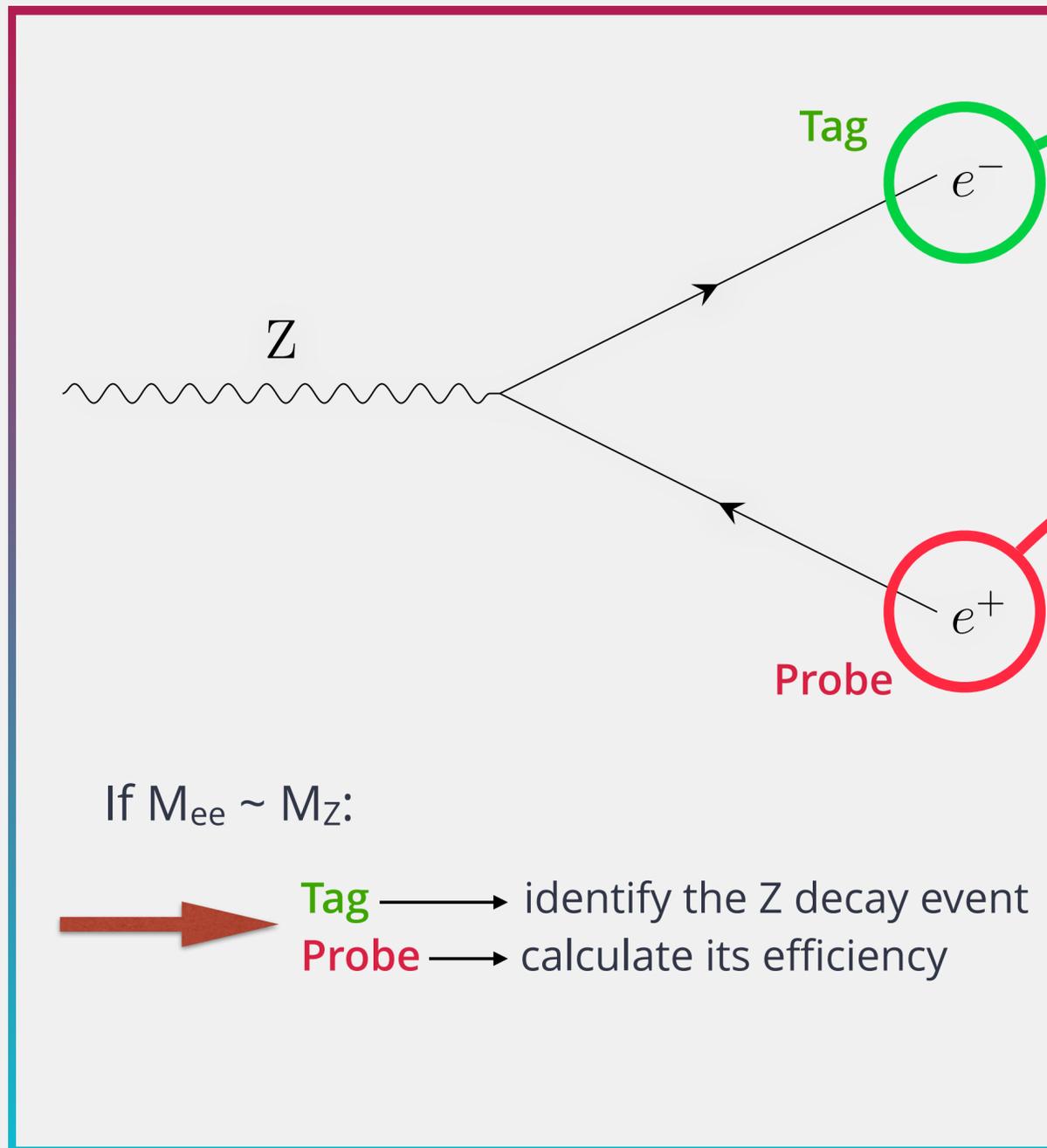
- Electromagnetic calorimeter: photon and electron detection
- Hadronic calorimeter: neutral and charged hadrons
- Missing transverse energy.

## Muon spectrometers

- Measure of muons' impulsion getting through the whole detector
- Trajectory curved by toroidal magnets.

# Tag and Probe method

In Zee channel - Detailed selection



- 1** Need a pure and unbiased electron sampling  
Strict selection on the first electron:
  - tight cuts
  - out of crack region
  - triggered by a  $p_T > 25$  GeV electron
- 2** A looser candidate with  $M_{ee} \sim M_Z$   
Cell cluster reconstructed in the calorimeter
  - $p_T > 15$  GeV
  - $\Delta R > 0.4$  (jet  $p_T > 20$  GeV)
  - $M_{ee}$  in ]80, 100[ GeV
  - + if MC; probe from Z decay (truth level)
- 3** The two electron candidates are tested as tag if they pass the selection criteria
- 4** Association tracks-clusters: track quality requirement (at least 1 hit in the pixel detector and 7 in the SCT)

# Background estimation

Electrons reconstructed as clusters with an associated track



## 1 Reversed identification

Background model made by reversed identification:  
probe candidates failing selection

- Fail to pass at least 2 loose++ cuts
- Bad isolation of the probe

Template	Cuts	$p_T < 30$ GeV	$p_T \geq 30$ GeV
Variation 1	fail at least 2 loose+ + cuts	$\text{topo}E_T^{\text{cone}30}/p_T > 0.02$ $120 < m_{ee} < 250$ GeV	$\text{topo}E_T^{\text{cone}40}/p_T > 0.05$ $120 < m_{ee} < 250$ GeV
Variation 2		$\text{topo}E_T^{\text{cone}30}/p_T > 0.02$ $60 < m_{ee} < 70$ GeV	$\text{topo}E_T^{\text{cone}40}/p_T > 0.20$ $120 < m_{ee} < 250$ GeV

## 2 Exclusion of signal contamination

Signal contribution in the high mass window [120, 250] GeV estimated by simulation:

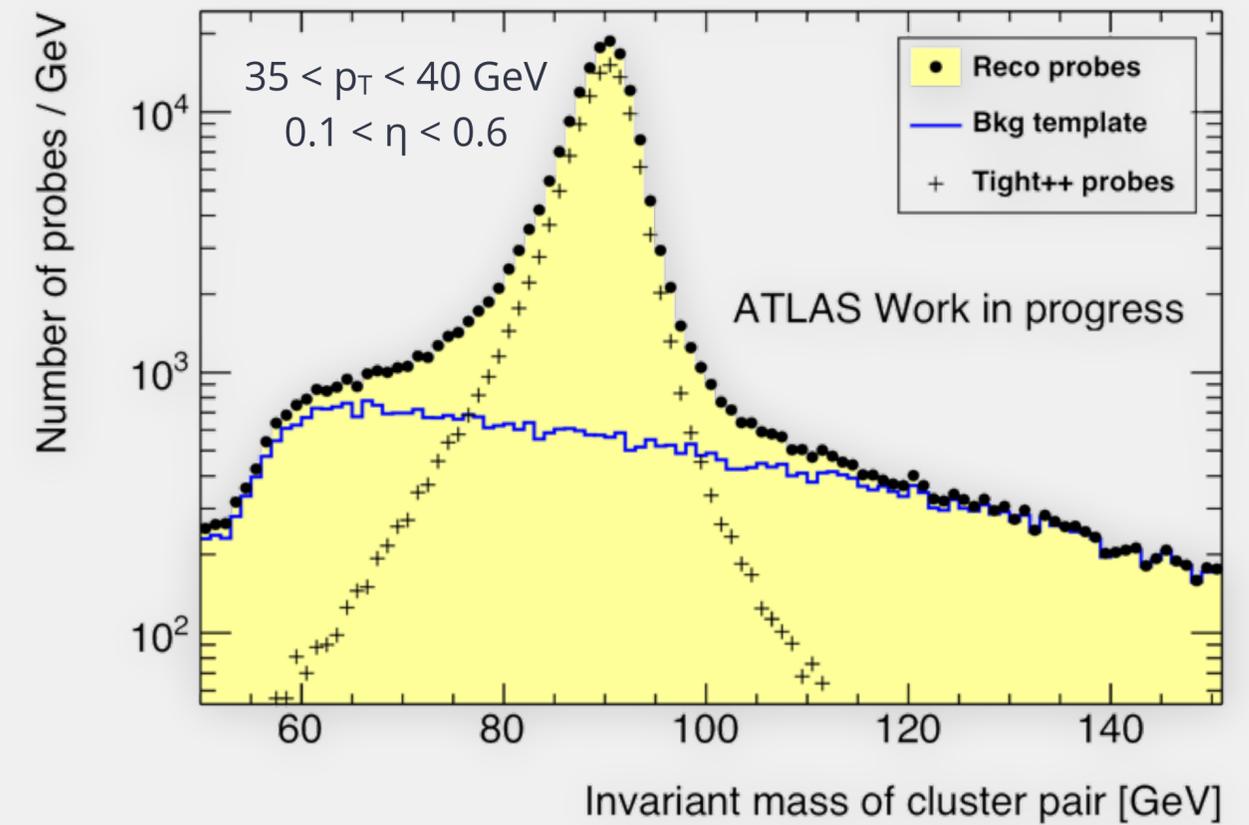
- in the peak region, the number of background is:

$$B^e = N_{\text{peak}}^{\text{template}} \times \frac{N_{\text{tail}}^e - N_{\text{tail}}^{\text{tight}++} / \epsilon_{\text{tail}}^{\text{tight}++}}{N_{\text{tail}}^{\text{template}}} \quad \begin{array}{l} \sim 70\% \text{ for } 15 < p_T < 20 \text{ GeV} \\ \sim 5\% \text{ for } 45 < p_T < 50 \text{ GeV} \end{array}$$

## 3 Normalization

Background efficiency limited

- Distribution normalized to the high/low mass data distribution:  $M_{ee} > 120$  GeV / [60, 70] GeV
- > 50 x less signal in the normalization region compared to the peak



# Background estimation

Electrons reconstructed as clusters with no associated track

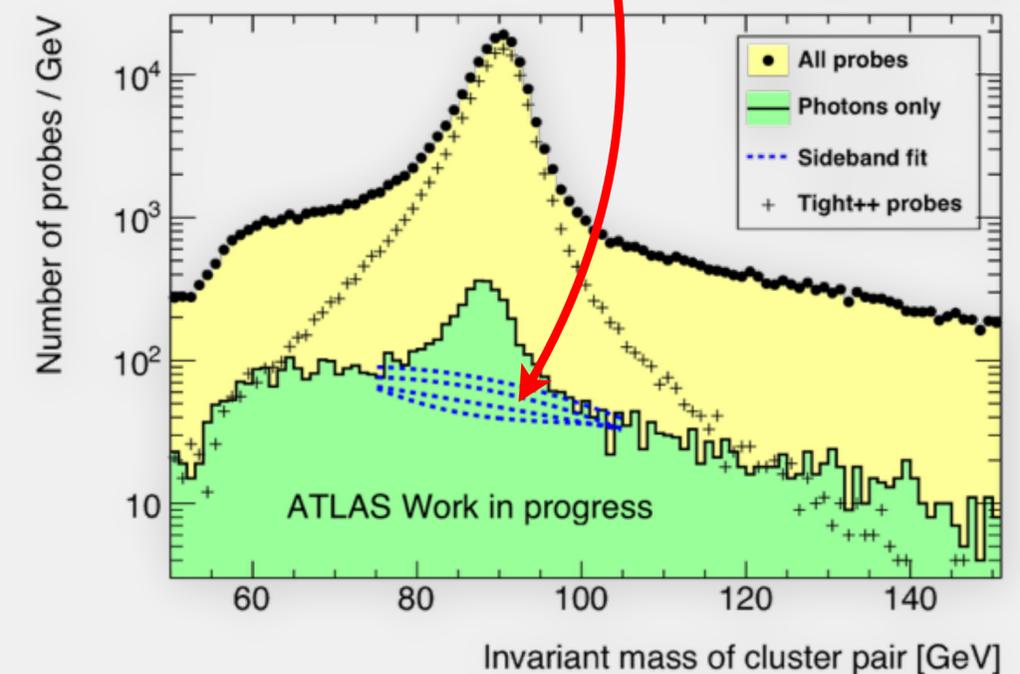
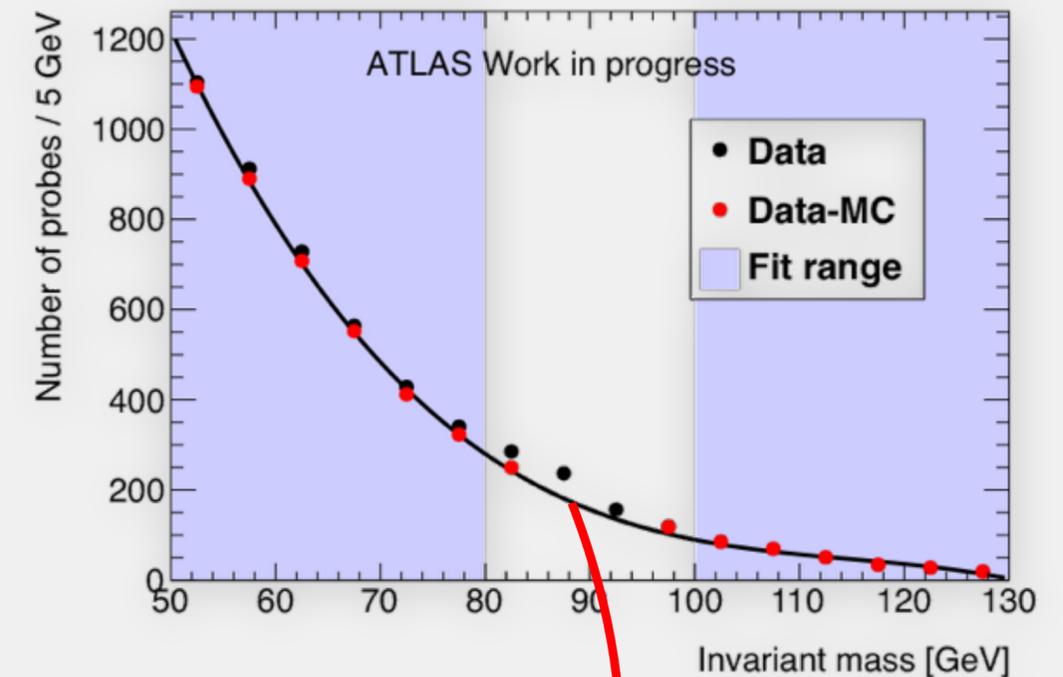


$$\epsilon_{\text{reco}} = \frac{N_{\text{pass}}^{\text{sig}}}{N_{\text{pass}}^{\text{sig}} + N_{\text{fail}}^{\text{sig}}} = \frac{N_p^e - B_p^e}{(N_p^e - B_p^e) + (N_F^e - B_F^e) + \underbrace{(N^\gamma - B^\gamma)}_{\Gamma}}$$

Probe candidates failing electron reconstruction, regarded as photons.  
Here, it is difficult to use reversed cuts or isolation variables.

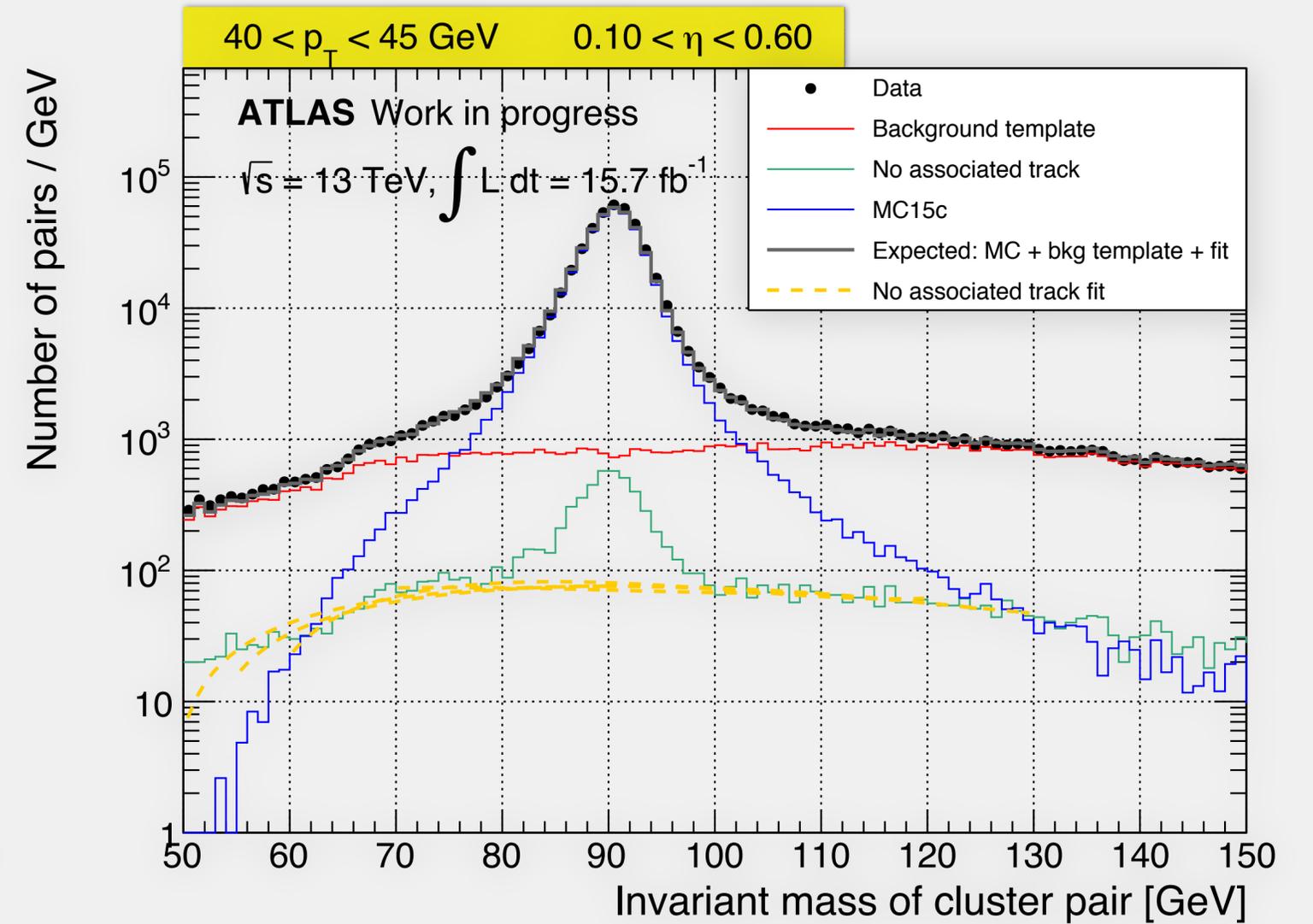
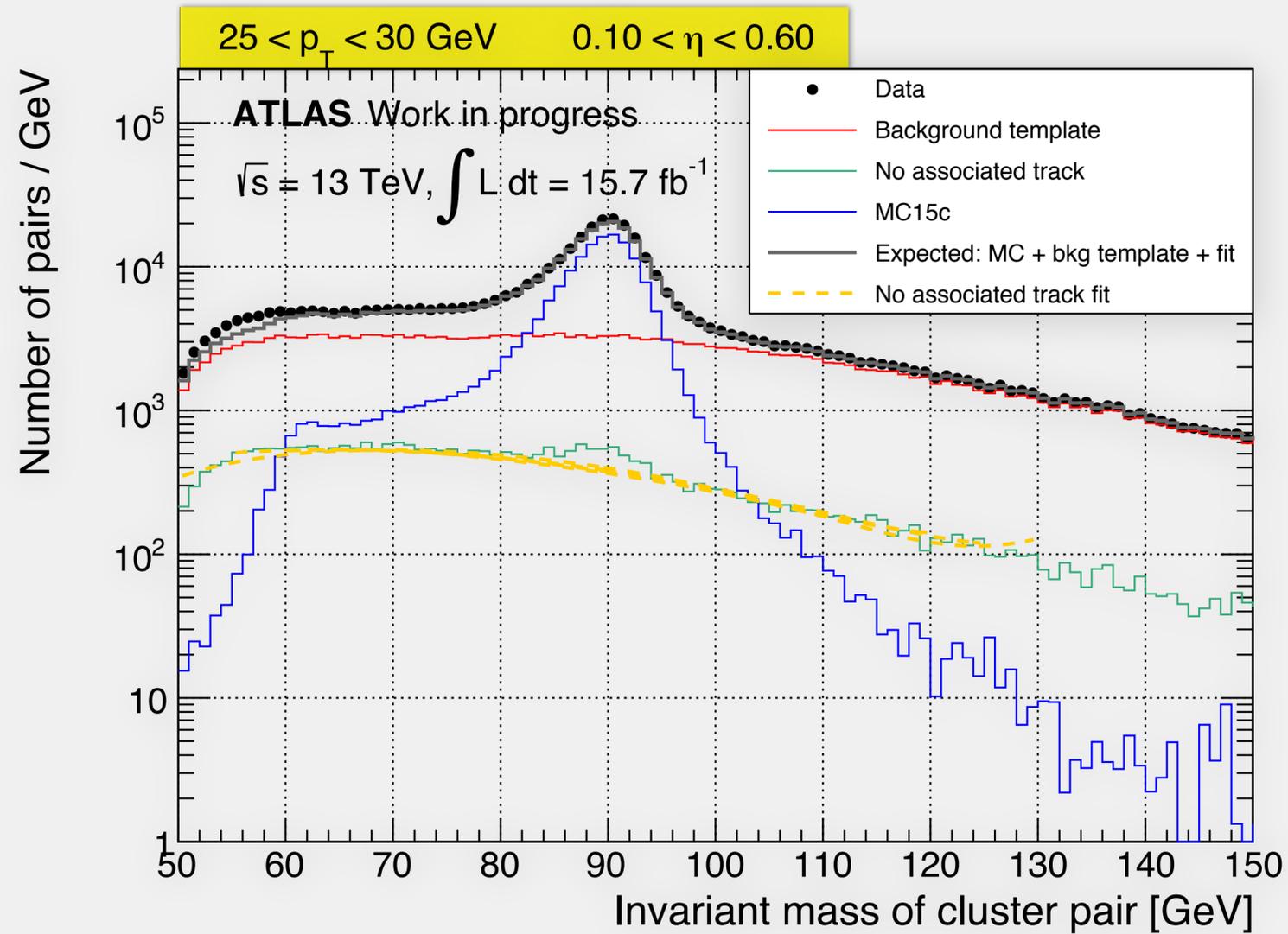
→ Simpler method

1. Construct invariant mass between electron tag and photon probe.
2. Choose side band regions outside the Z mass peak.
3. Signal in the side band regions subtracted from simulation
4. Fit the  $e\gamma$ -mass distribution with a 3rd order polynomial.



# Electron reconstruction efficiency in 2016 data

Denominator detailed distribution

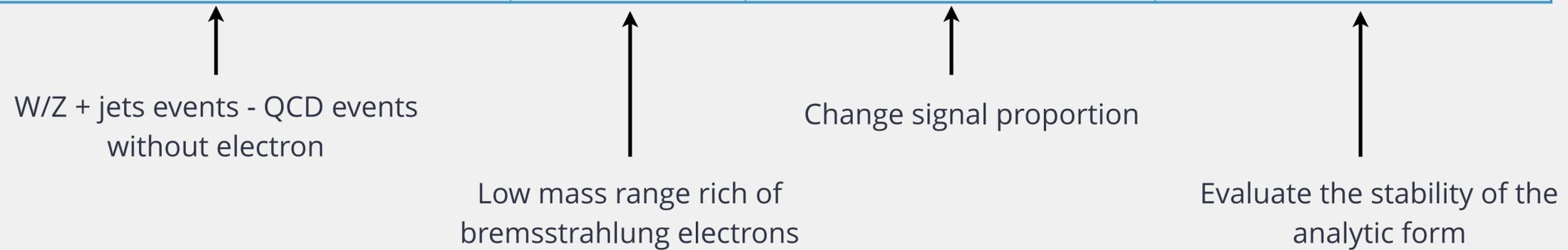


# Efficiency computation

Detailed variations



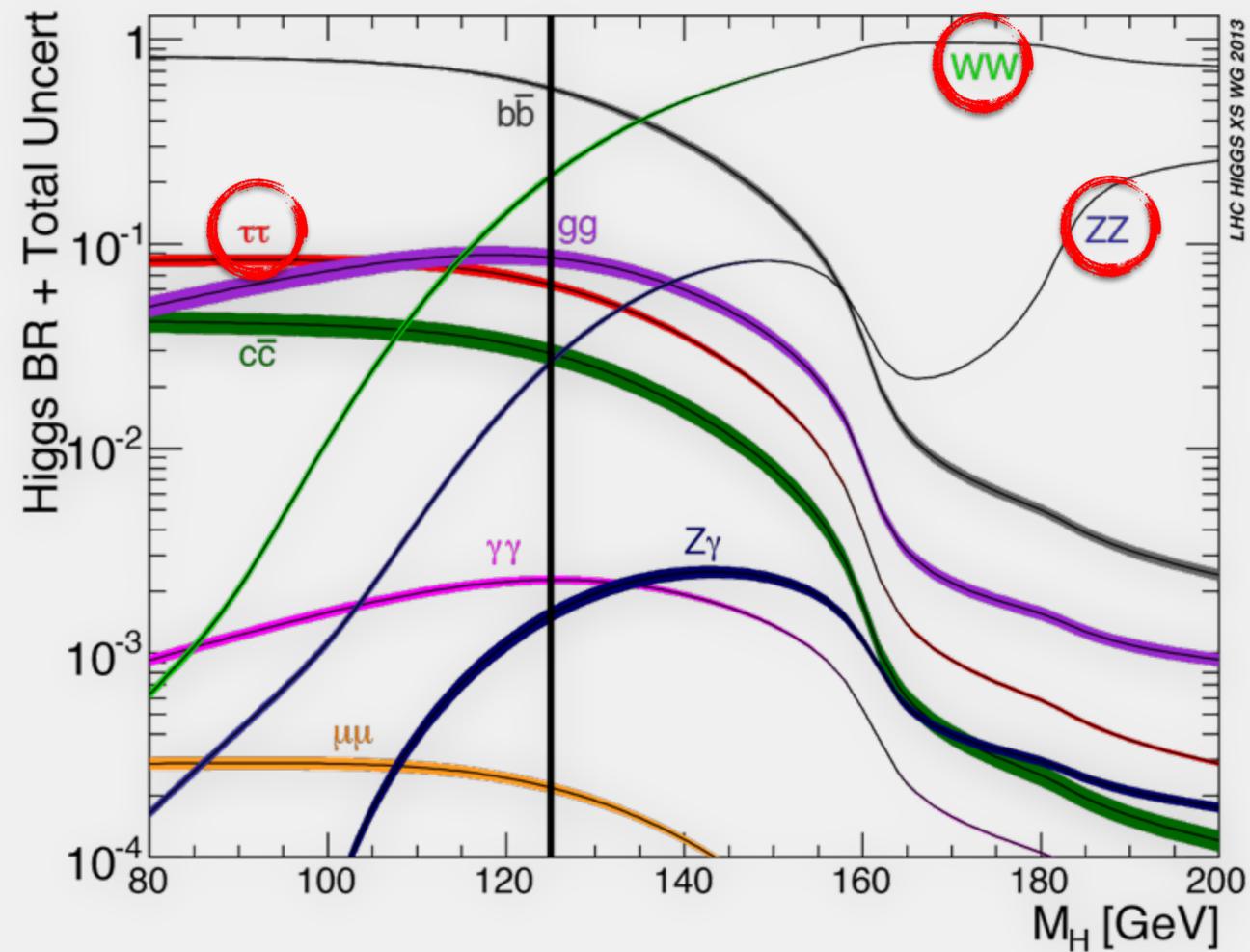
Tag identification variations	Z mass peak windows	Electron background template	Sideband for $e\gamma$ mass fit
Tight LH	]80, 100[ GeV	Variation 1	]70, 80[ U ]100, 110[ GeV
Tight LH & $\text{TopoE}_T^{\text{cone40}} < 5$ GeV	]75, 105[ GeV	Variation 2	]60, 80[ U ]100, 120[ GeV
Medium LH & $\text{TopoE}_T^{\text{cone40}} < 5$ GeV	]85, 95[ GeV		]50, 80[ U ]100, 130[ GeV
			]55, 70[ U ]110, 125[ GeV



Template	Cuts	$p_T < 30$ GeV	$p_T \geq 30$ GeV
Variation 1	fail at least 2 loose+ + cuts	$\text{topoE}_T^{\text{cone30}}/p_T > 0.02$ $120 < m_{ee} < 250$ GeV	$\text{topoE}_T^{\text{cone40}}/p_T > 0.05$ $120 < m_{ee} < 250$ GeV
Variation 2		$\text{topoE}_T^{\text{cone30}}/p_T > 0.02$ $60 < m_{ee} < 70$ GeV	$\text{topoE}_T^{\text{cone40}}/p_T > 0.20$ $120 < m_{ee} < 250$ GeV

# Higgs boson

Decay



## Branching ratios

- Probability to decay in a defined final state

$$BR(H \rightarrow b\bar{b}) = \frac{\Gamma(H \rightarrow b\bar{b})}{\Gamma(H \rightarrow b\bar{b}) + \Gamma(H \rightarrow c\bar{c}) \dots} = 57.7\%$$

## To get two same-sign leptons

- 3 Higgs decays

- H → WW (22%)
- H → ττ (6.3%)
- H → ZZ (2.6%)



- W → lν (21%)
  - τ → lνν (35.2%)
  - Z → ll (6.7%)
- for l = e, μ



## ttH production with a multileptonic final state

Study of five multileptonic final states from ttH production:

Run 1 : 20 fb<sup>-1</sup> data,  $\sqrt{s} = 8$  TeV

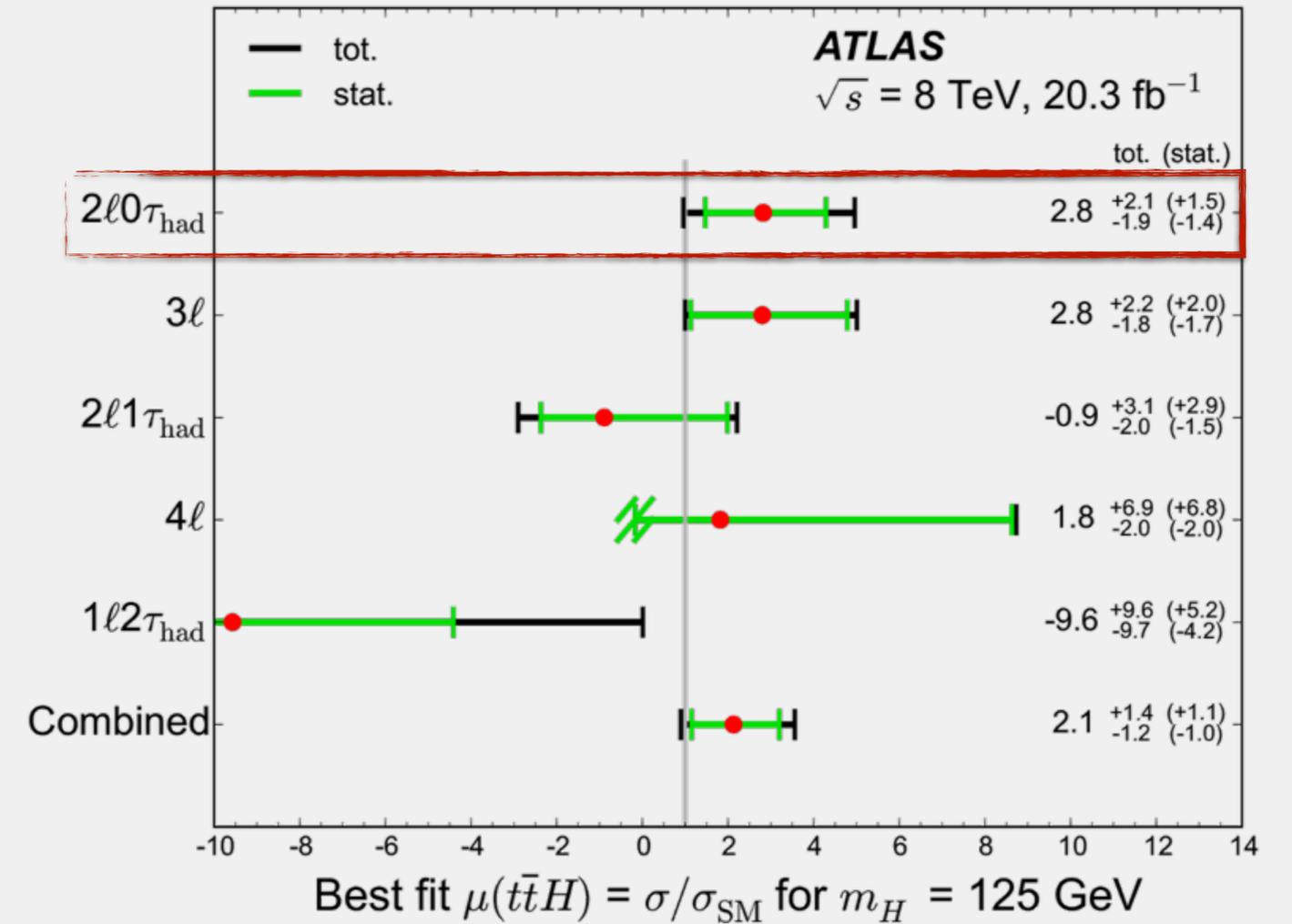
$$\left\{ \begin{array}{l} \mu_{ttH} = \frac{\sigma(pp \rightarrow ttH)}{\sigma_{SM}(pp \rightarrow ttH)} = 2.1^{+1.4}_{-1.2} \\ \mu_{ttH} < 4.7 \text{ with 95\% CI} \end{array} \right.$$

Consistent with the Standard Model, but limited by statistics.

Run 2 : 100 fb<sup>-1</sup> data,  $\sqrt{s} = 13$  TeV

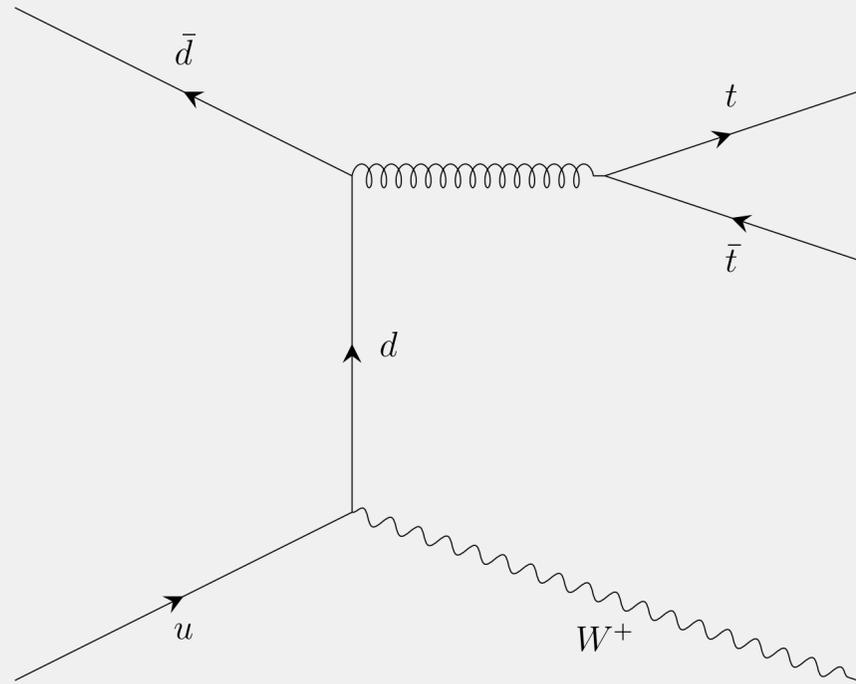
→  $\mu$  measurement possible for the first time with the Run 2 of LHC.  
Great opportunity to get a precise measurement.

Prospect study for the run 2 (2015 - 2018) in the two same-sign leptons channel (e ou  $\mu$ ).



# ttH in same-sign leptons channel

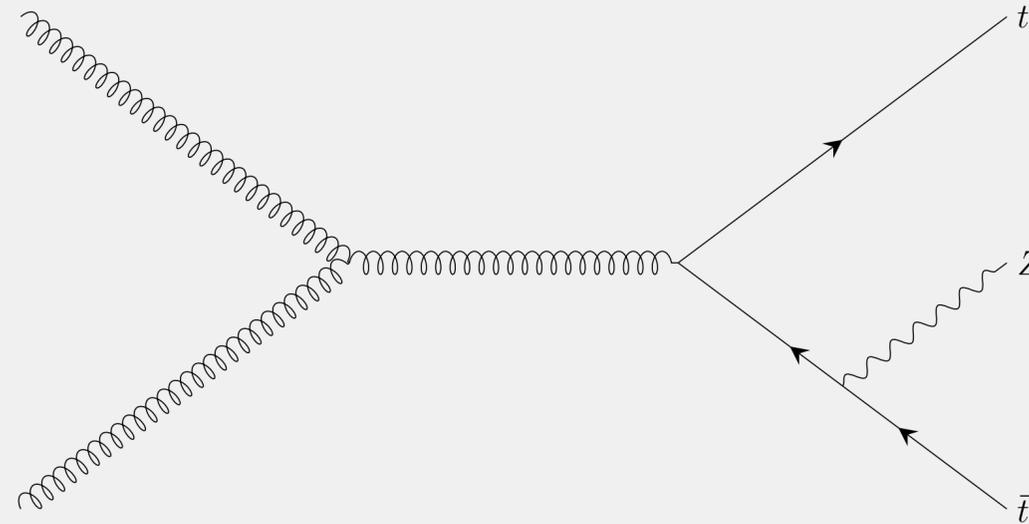
Backgrounds



ttW + 0, 1, 2, jets

This diagram appears with a positive or negative W leading to 2/3 events with two positive leptons and 1/3 with two negative leptons.

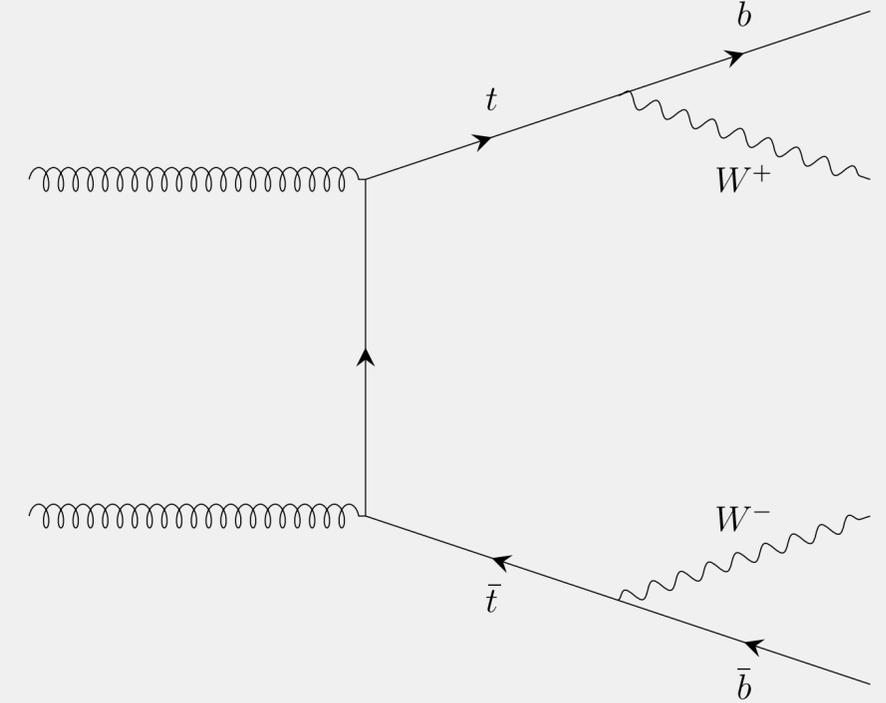
$$\sigma \times 2,9$$



ttZ + 0, 1, 2 jets

ttZ leads to a final state with 3 leptons, but is regarded as a two lepton final state if the third is misreconstructed or out of the detector acceptance.

$$\sigma \times 2,8$$



tt

Dominating background (but reducible) with additional jets and non-prompt leptons.

$$\frac{\sigma(tt)}{\sigma(ttH)} \sim 2000 (1500) \text{ for } \sqrt{s} = 8 \text{ TeV} (13 \text{ TeV})$$

$$\sigma \times 3,3$$

# BDT response to the two trainings



-  = same signal efficiency as cut based
-  = max significance for the defined signal efficiency
-  = max significance for the entire map

