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, Fabrice Hubaut

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

kevin.devasconcelos@cern.ch

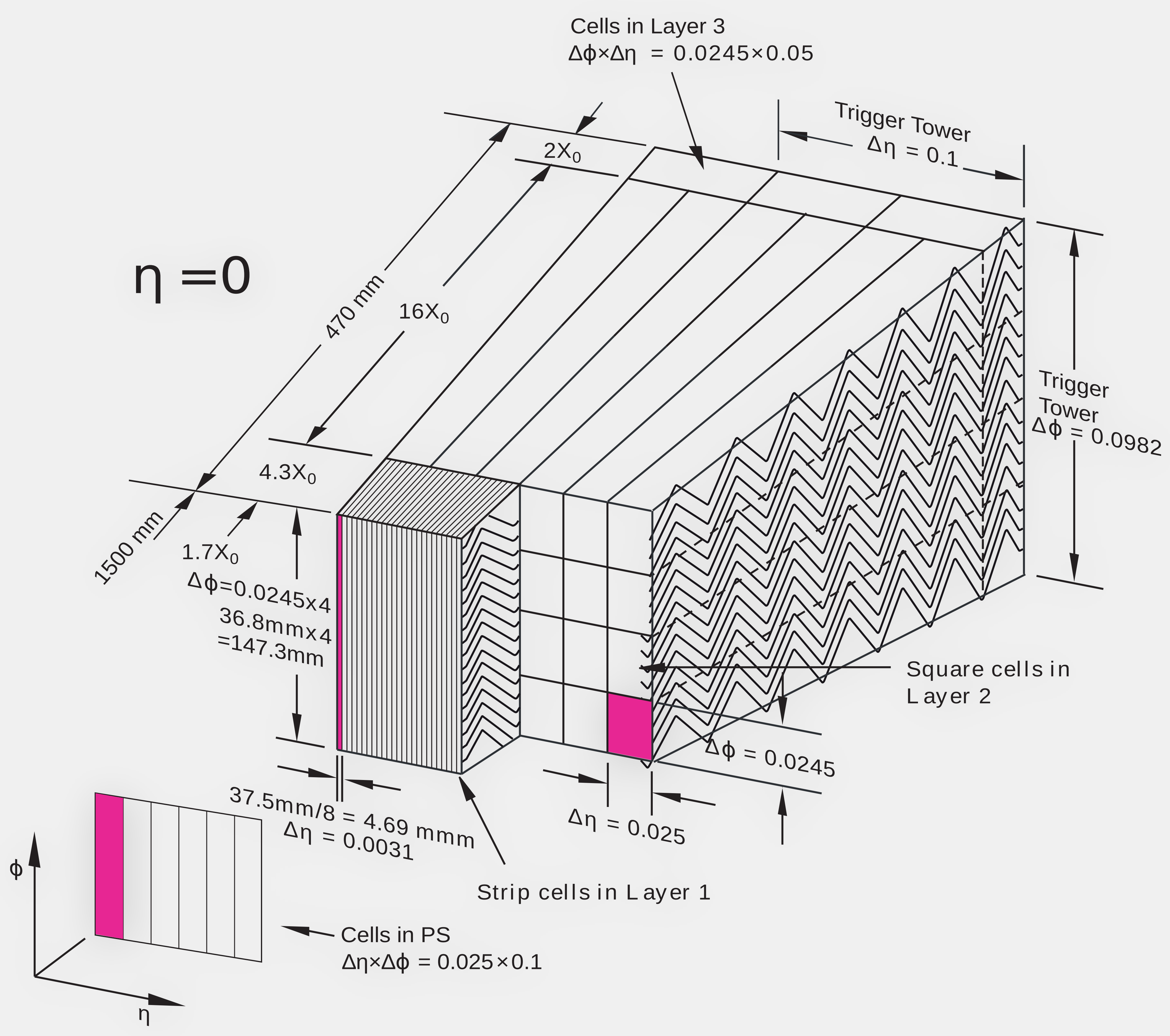
CPPM PhD day seminar - 7th October 2016

ATLAS EXPERIMENT

CENTRE DE PHYSIQUE DES PARTICULES DE MARSEILLE CPPM

Motivation

I



Sketch of the electromagnetic calorimeter granularity

Electron reconstruction

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

$\longrightarrow$ New Physics (NP)

$\longrightarrow$ Precision measurements

Needs :

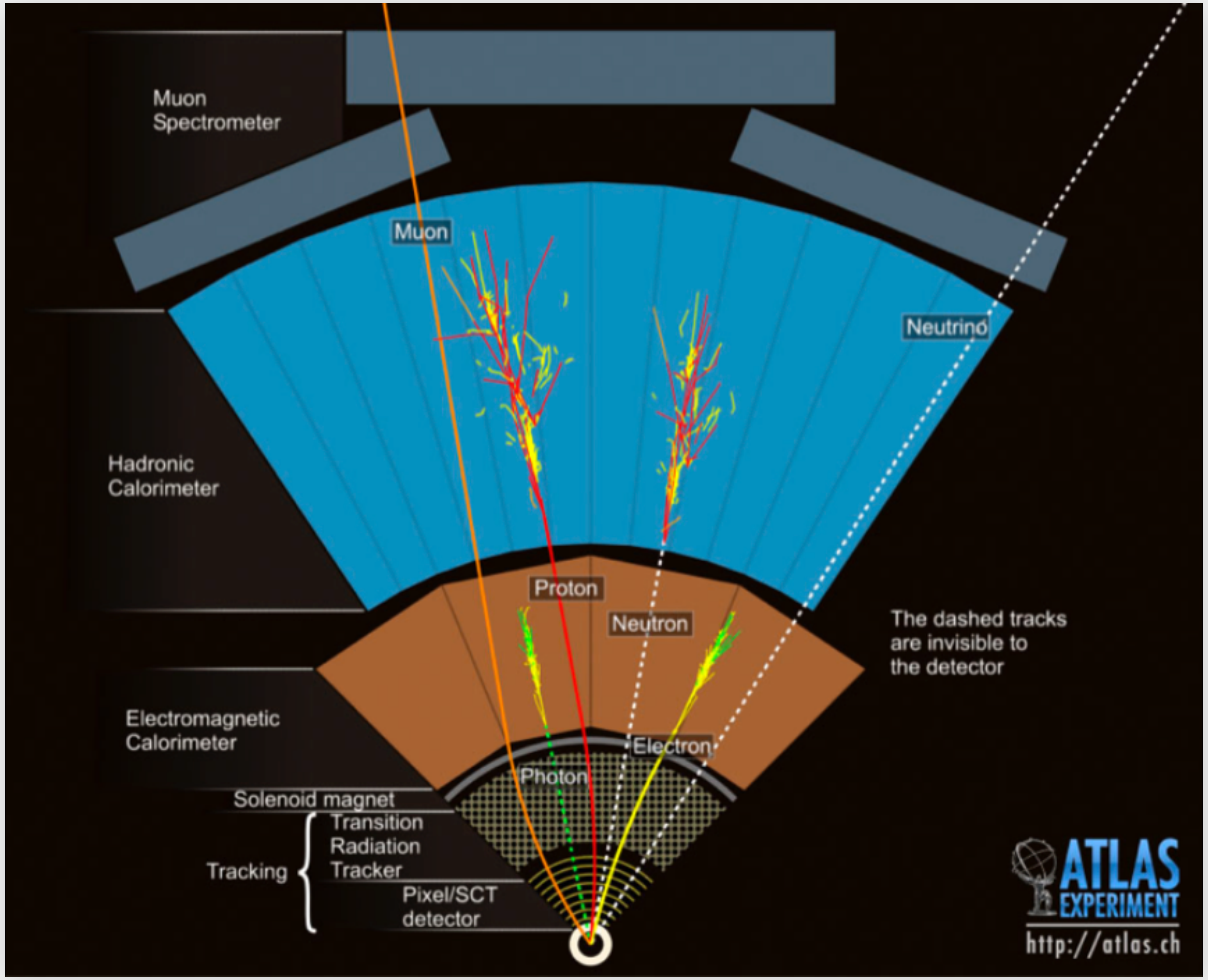
1. low number of events for NP: high selection efficiency
2. high background rejection (ex $H \longrightarrow WW$)

$\longrightarrow$ Understanding of the ATLAS detector.

As trigger and identification, reconstruction of electrons needs a precise efficiency measurement for all SM studies or to set limits.

$$\varepsilon_{\text{total}} = \boxed{\varepsilon_{\text{reco}}} \times \varepsilon_{\text{id}} \times \varepsilon_{\text{trigger}} \times \varepsilon_{\text{other}}$$

Electron reconstruction



Principle

- Reconstruction in the central region of the detector ($|\eta| < 2.47$)
- starts from energy deposits (clusters) in the EM calorimeter in a fixed $\Delta\eta \times \Delta\varphi$ 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

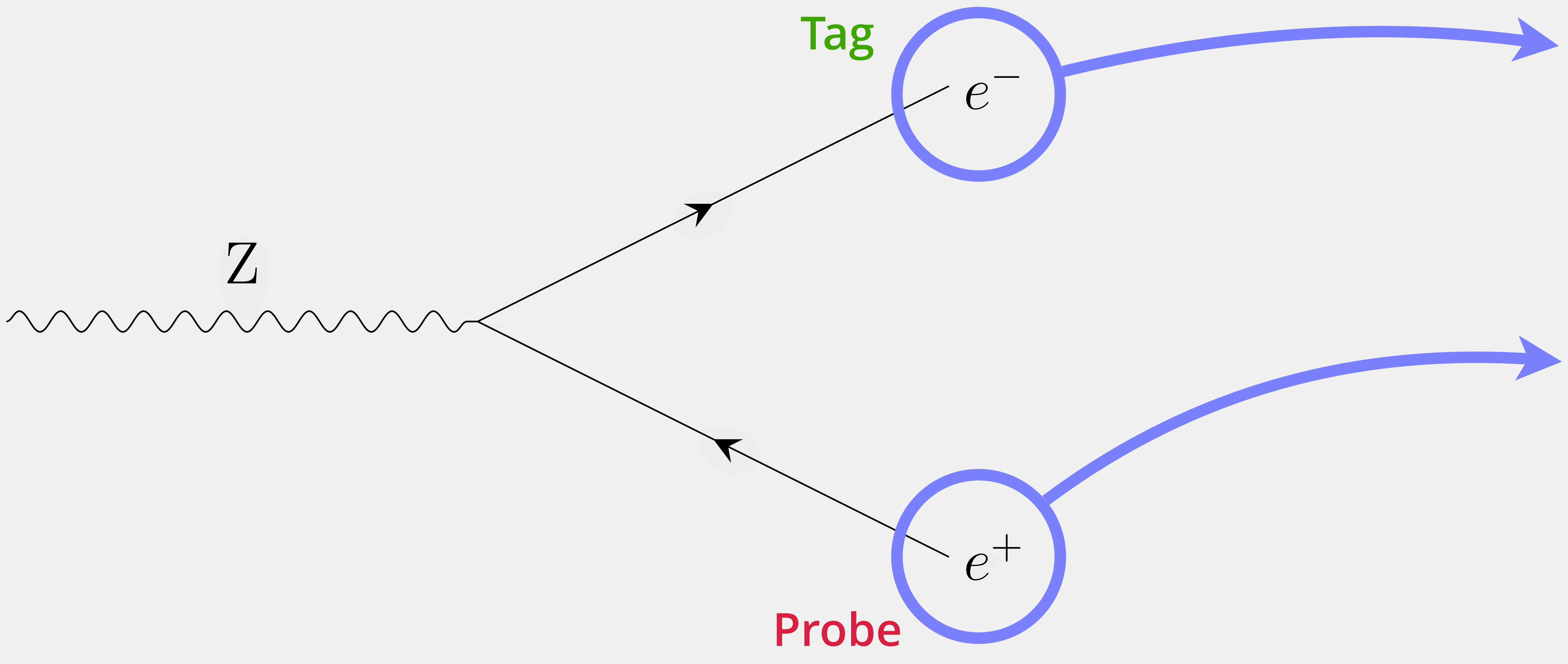
Challenges

- 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 a % or sub-% level accuracy.

Tag and Probe method

In Z ⟶ 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

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

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

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

Background estimate

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++ cuts** and bad isolation

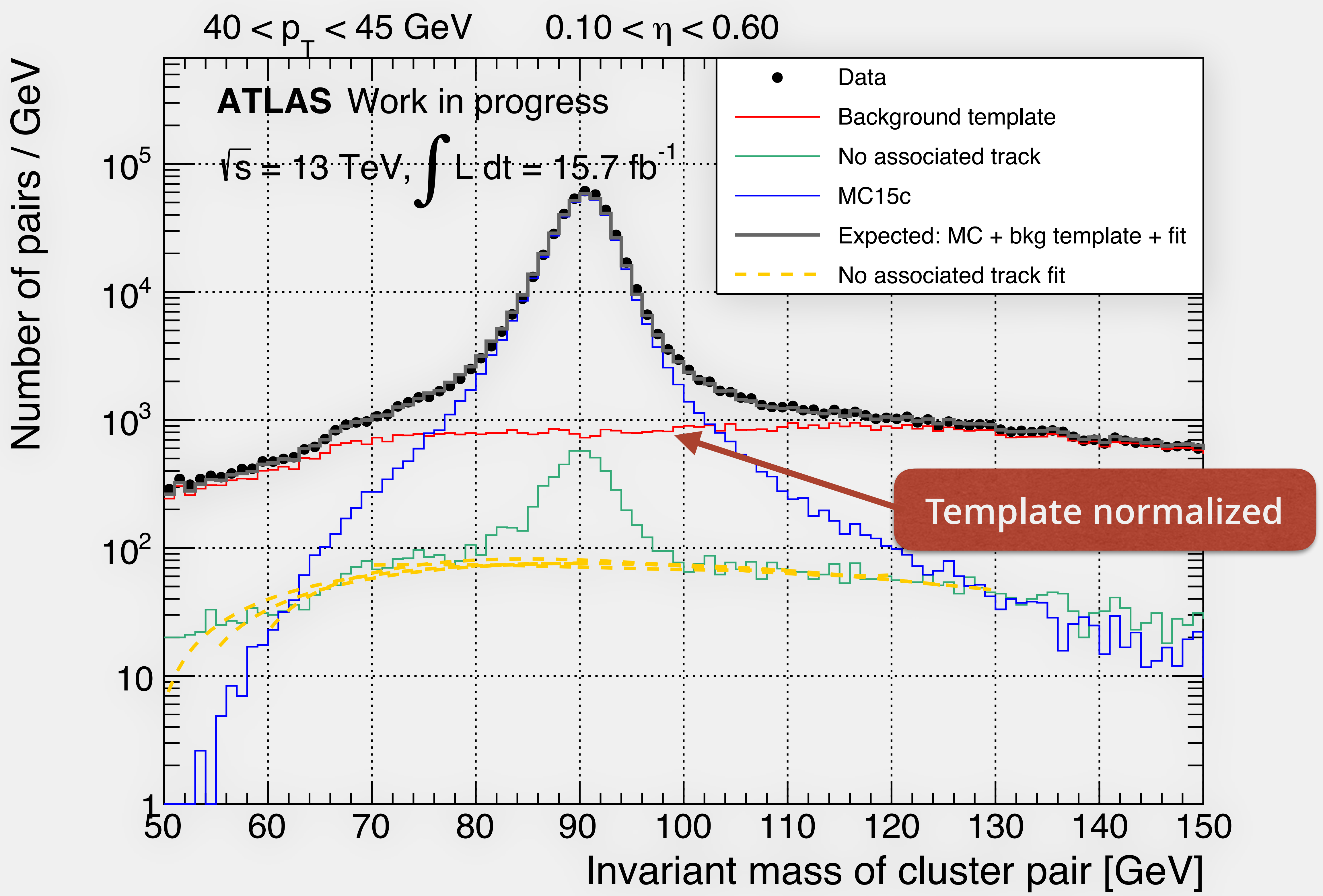
Normalization

Distribution normalized to the high/low mass data distribution: M_{ee} > 120 GeV / [60, 70] 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 eγ-mass distribution** with a 3rd order polynomial.



Background estimate

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++ cuts** and bad isolation

Normalization

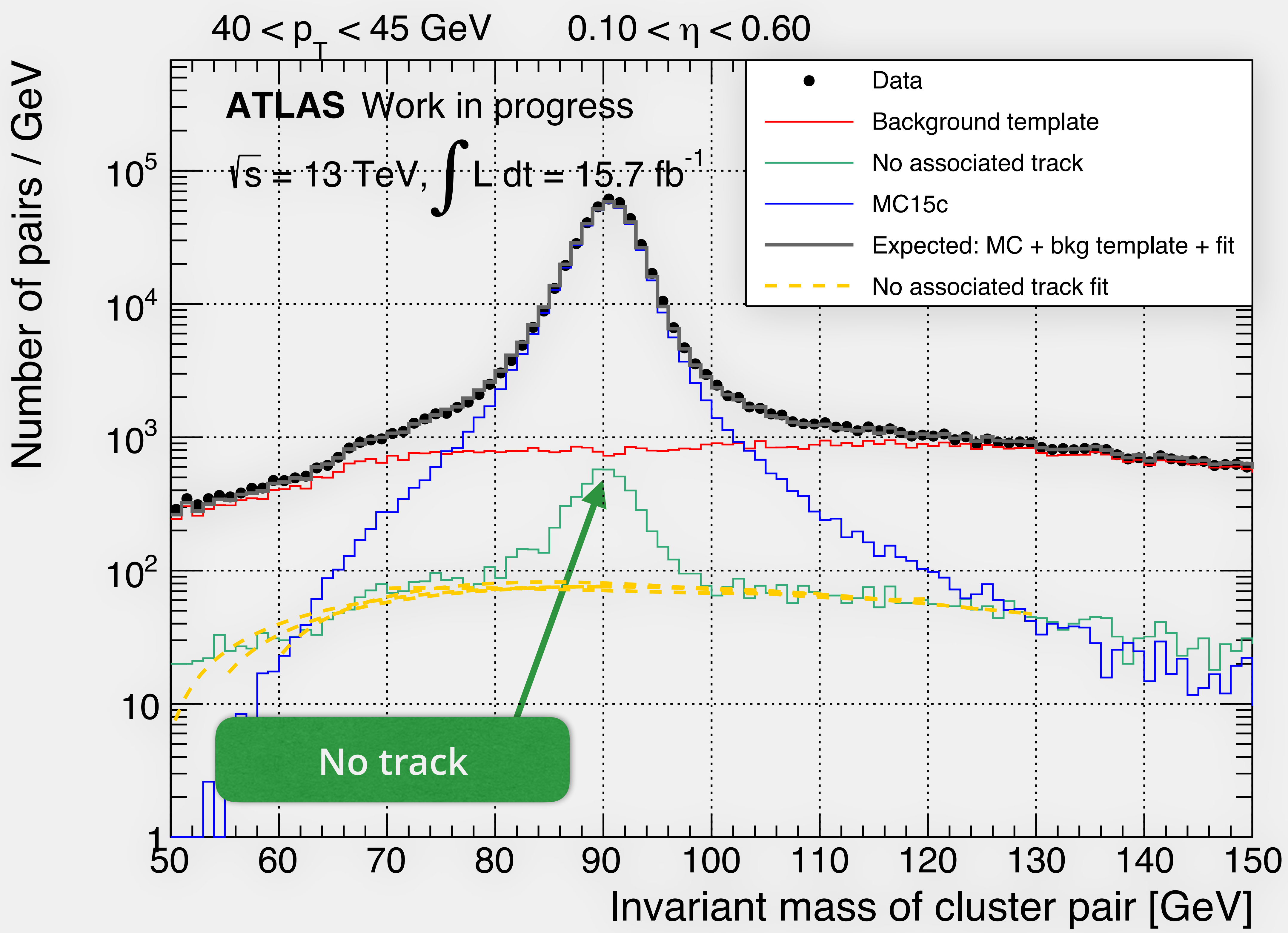
Distribution normalized to the high/low mass data distribution: M_{ee} > 120 GeV / [60, 70] GeV

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 eγ-mass distribution** with a 3rd order polynomial.



Background estimate

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++ cuts** and bad isolation

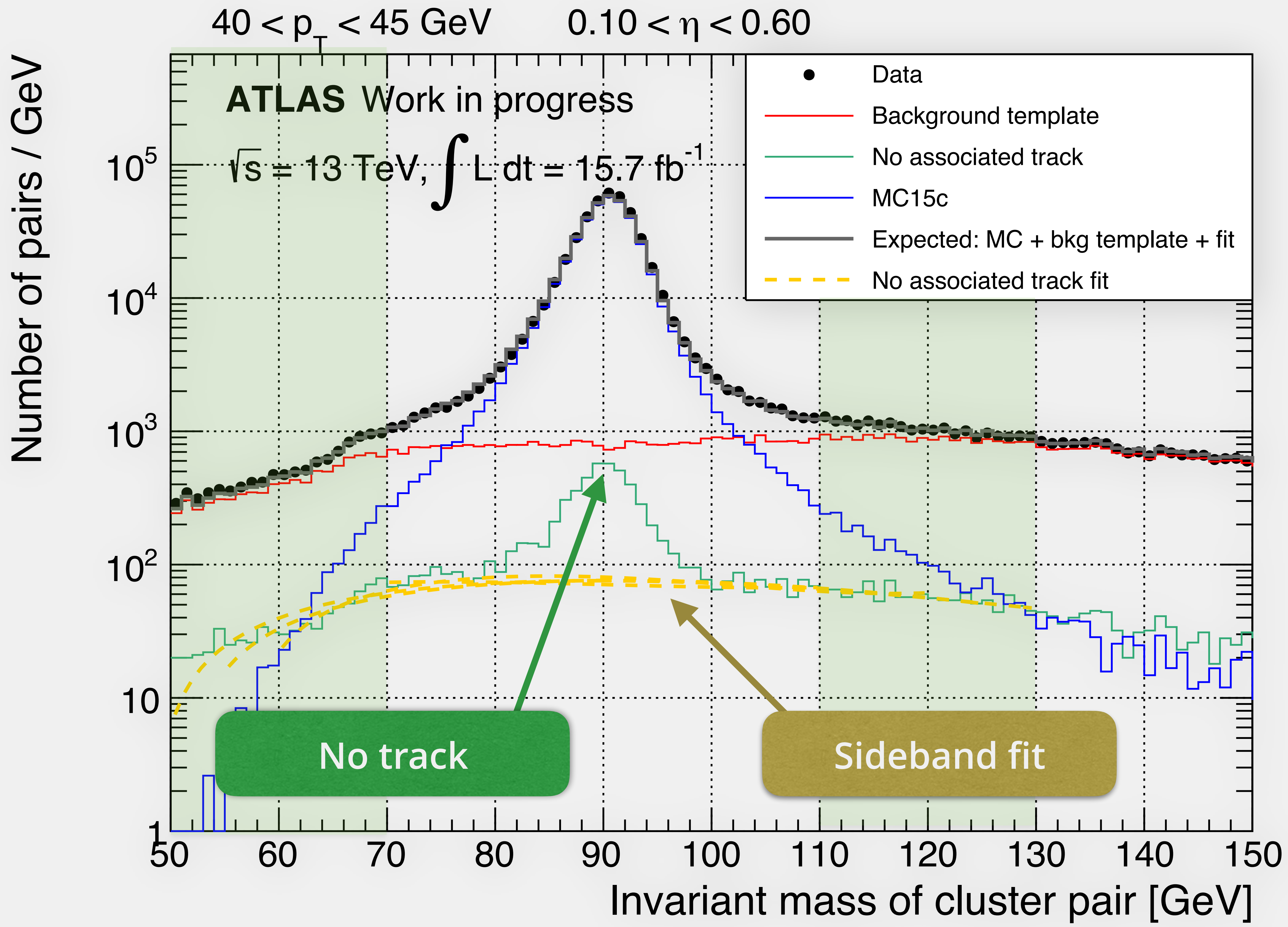
Normalization

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

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

Template difficult to build
→ Simpler method

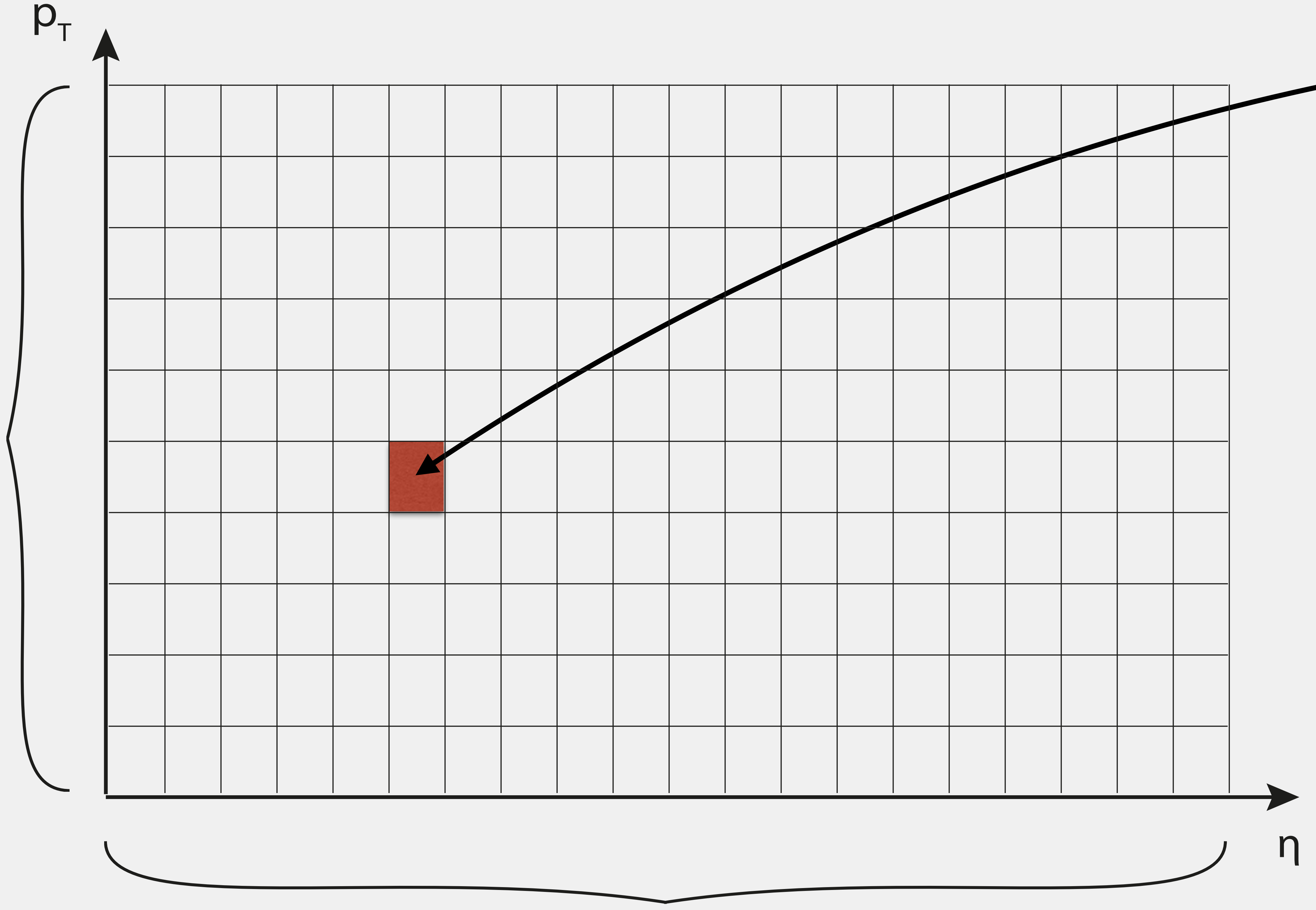
1. Construct invariant mass between electron tag and photon probe.
2. Choose **side band regions** outside the Z mass peak.
3. Signal in the side band regions subtracted from simulation
4. **Fit the eγ-mass distribution** with a 3rd order polynomial.



Measurement

Goal: Establish a scale factor map, binned in p_T and η.

10 p_T bins:
{15, 20, 25, 30, 35, 40, 45, 50, 60 ,80, 150} GeV



20 η bins: from detector geometry
{-2.47, -2.37, -2.01, -1.81, -1.52, -1.37, -1.15, -0.80, -0.60, -0.10, 0.00 , 0.10, 0.60, 0.80, 1.15, 1.37, 1.52, 1.81, 2.01, 2.37, 2.47}

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

$$\varepsilon_{\text{reco}} = \frac{N^{\text{sig}}_{\text{pass}}}{N^{\text{sig}}_{\text{pass}} + N^{\text{sig}}_{\text{fail}}} = \frac{\boxed{N^e_p - B^e_p}}{\boxed{(N^e_p - B^e_p)} + \boxed{(N^e_F - B^e_F)} + \boxed{(N^\gamma - B^\gamma)}}$$

Reconstructed candidates passing the probe track quality criteria

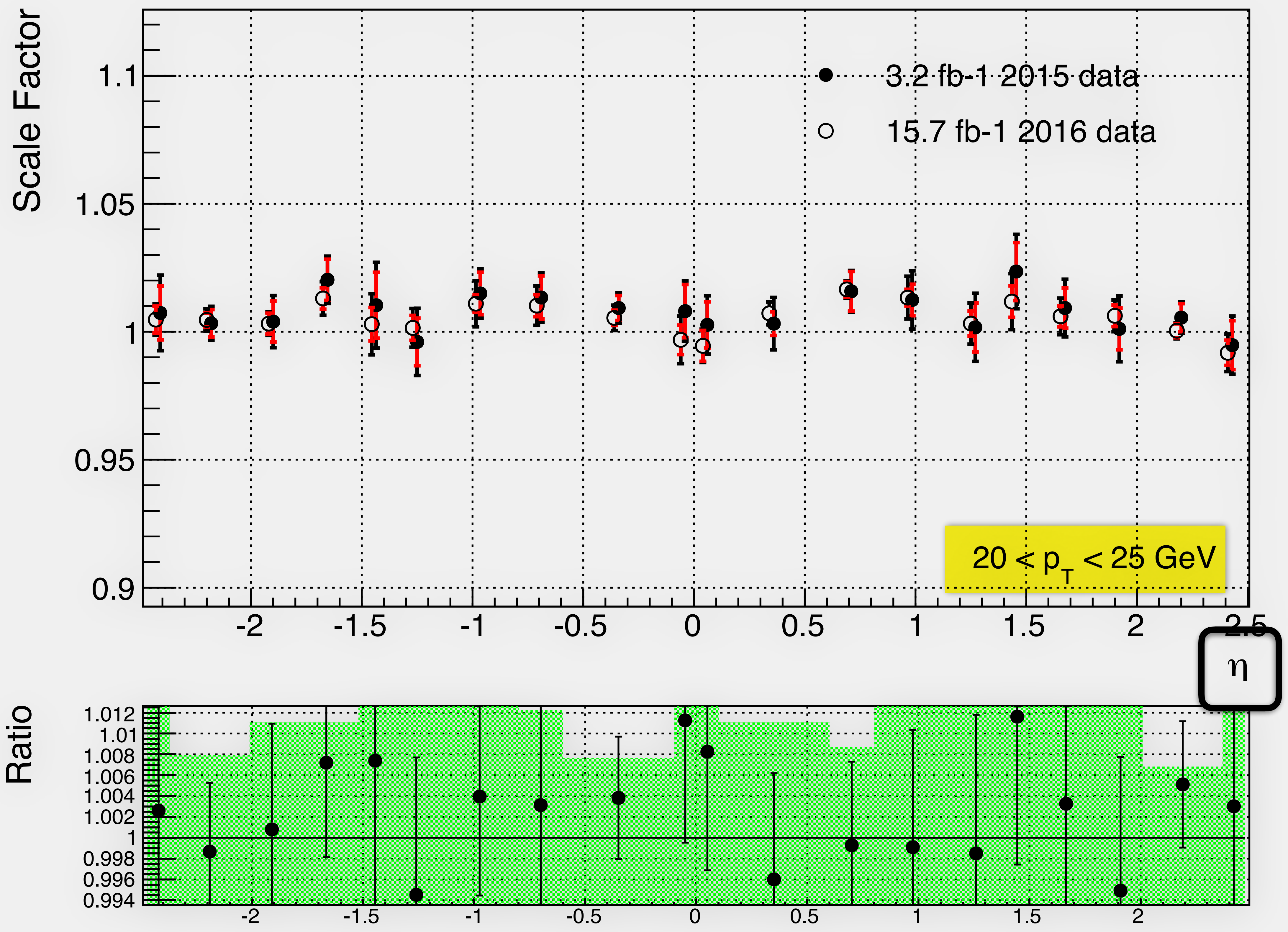
Reconstructed candidates failing track quality criteria

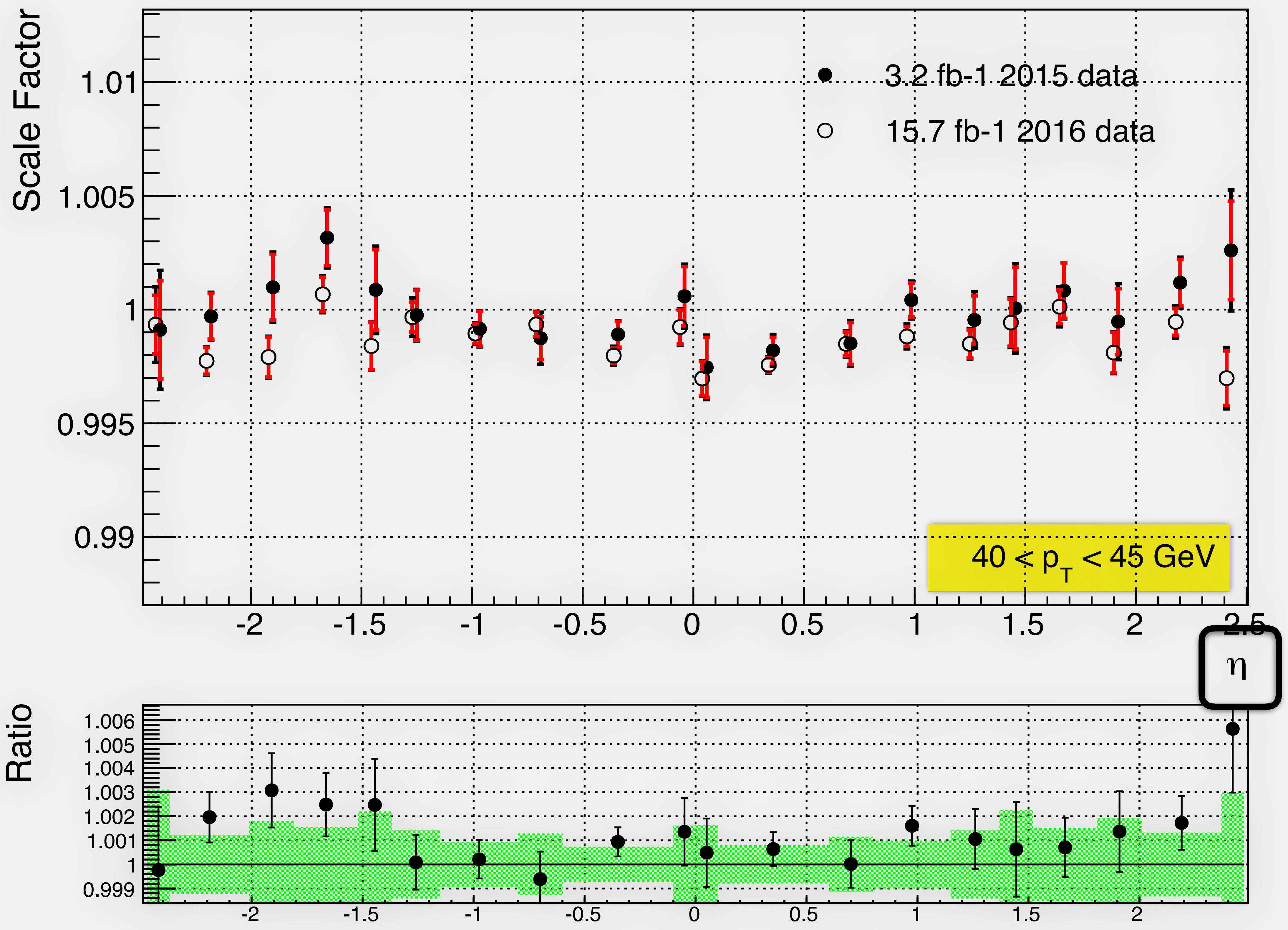
Non-reconstructed candidates

- Historically, CPPM has always been in charge of this measurement (Julien, Otilia, Asma, me, Grigore).
- 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





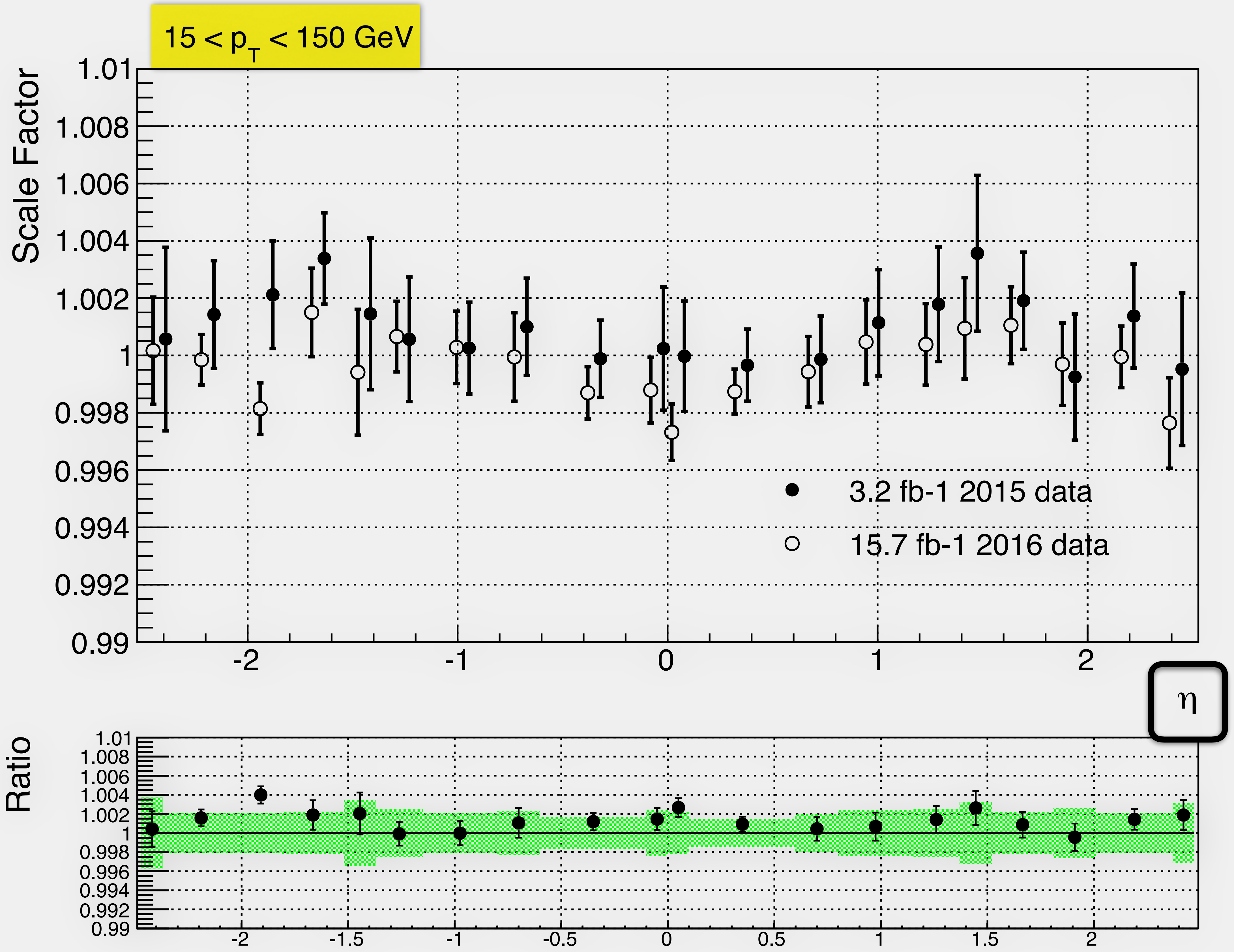
Results in 2016 compatible with 2015 data ones within error in most bins

Quadratic sum of 2015 and 2016 uncertainties

2015 data uncertainties

Electron reconstruction efficiency in 2016 data

Comparison with 2015 results



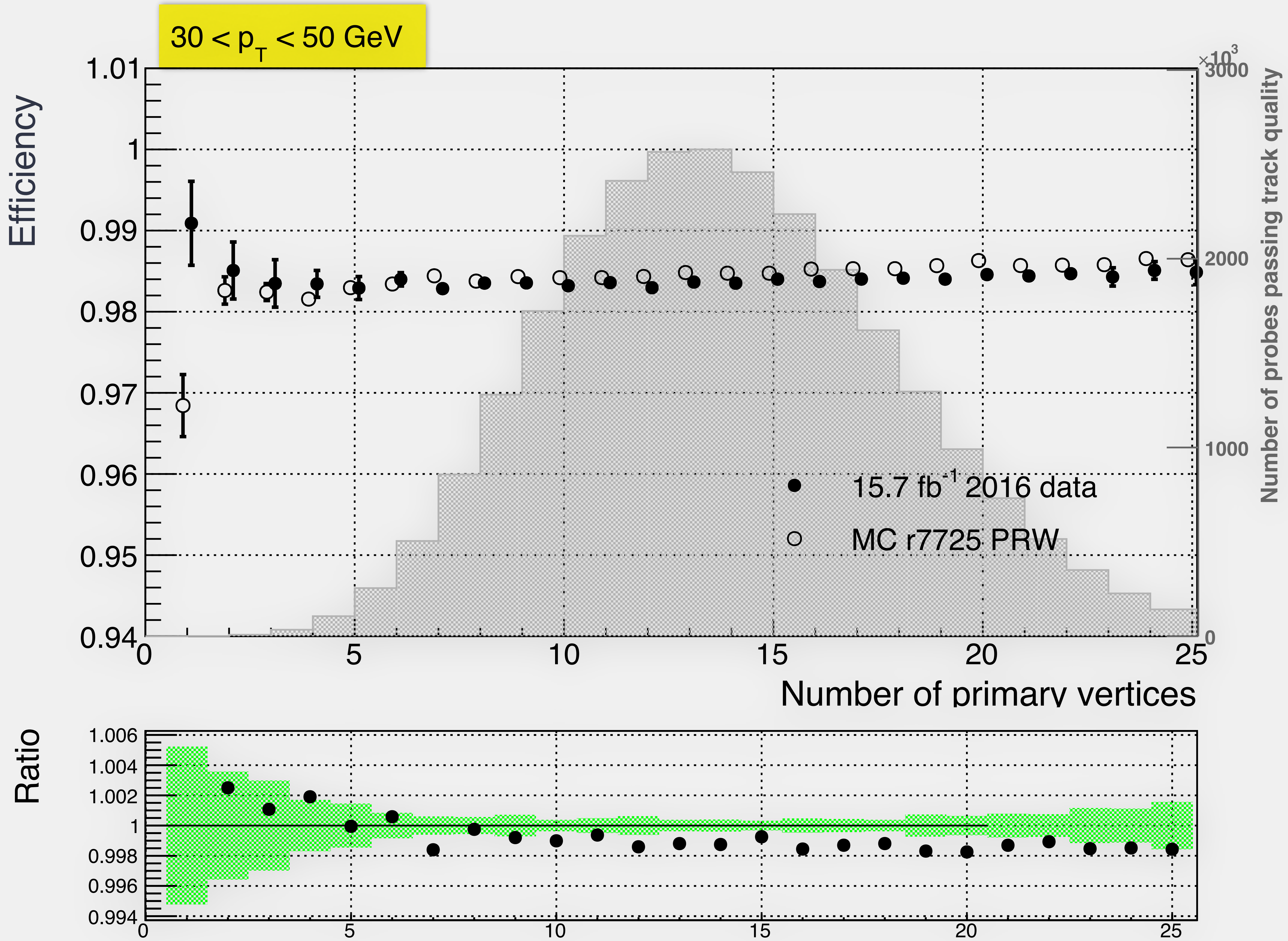
Results in 2016 compatible with 2015 data ones within error in most bins

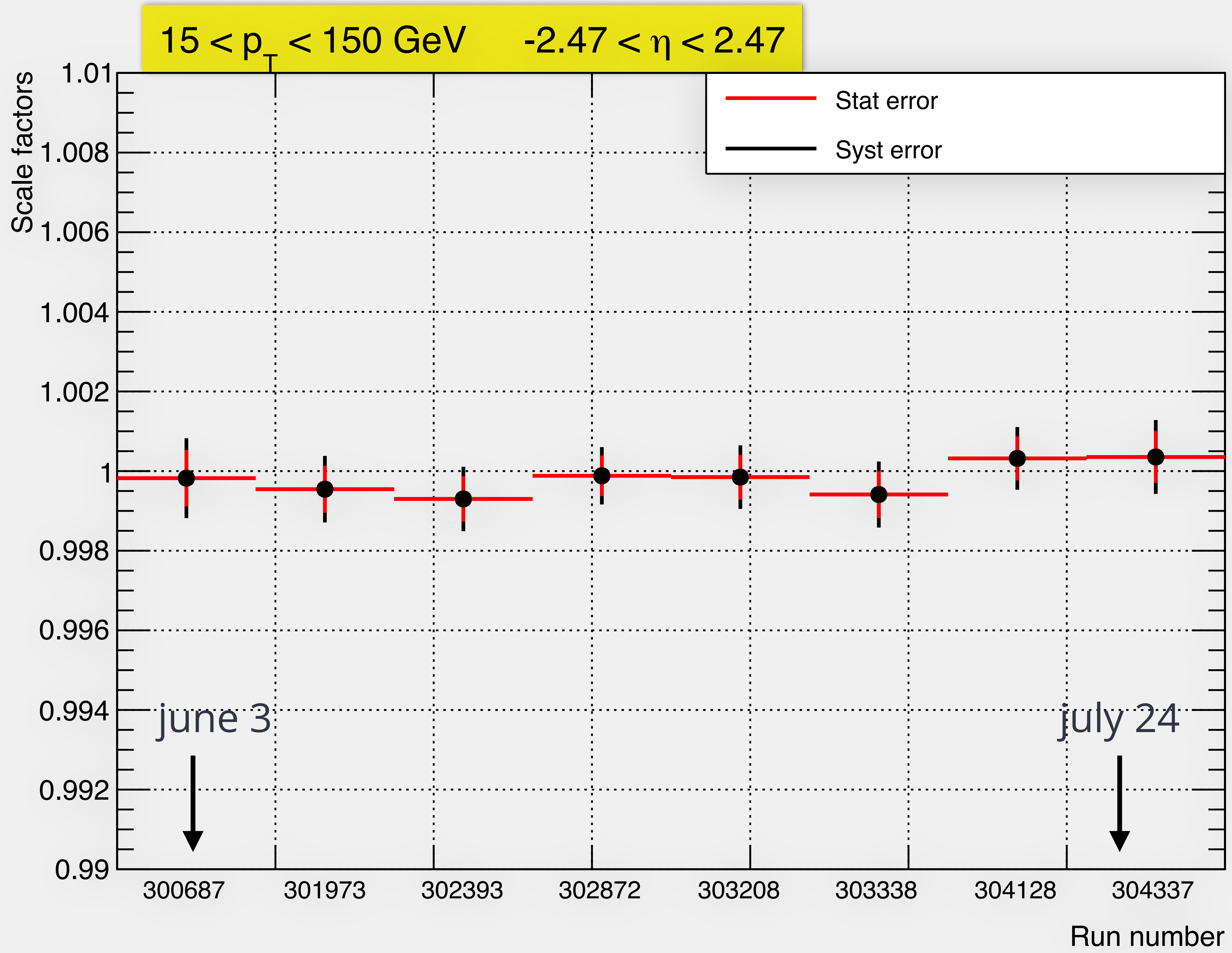
Quadratic sum of 2015 and 2016 uncertainties

2015 data uncertainties

Electron reconstruction efficiency in 2016 data

Stability in pile up and time



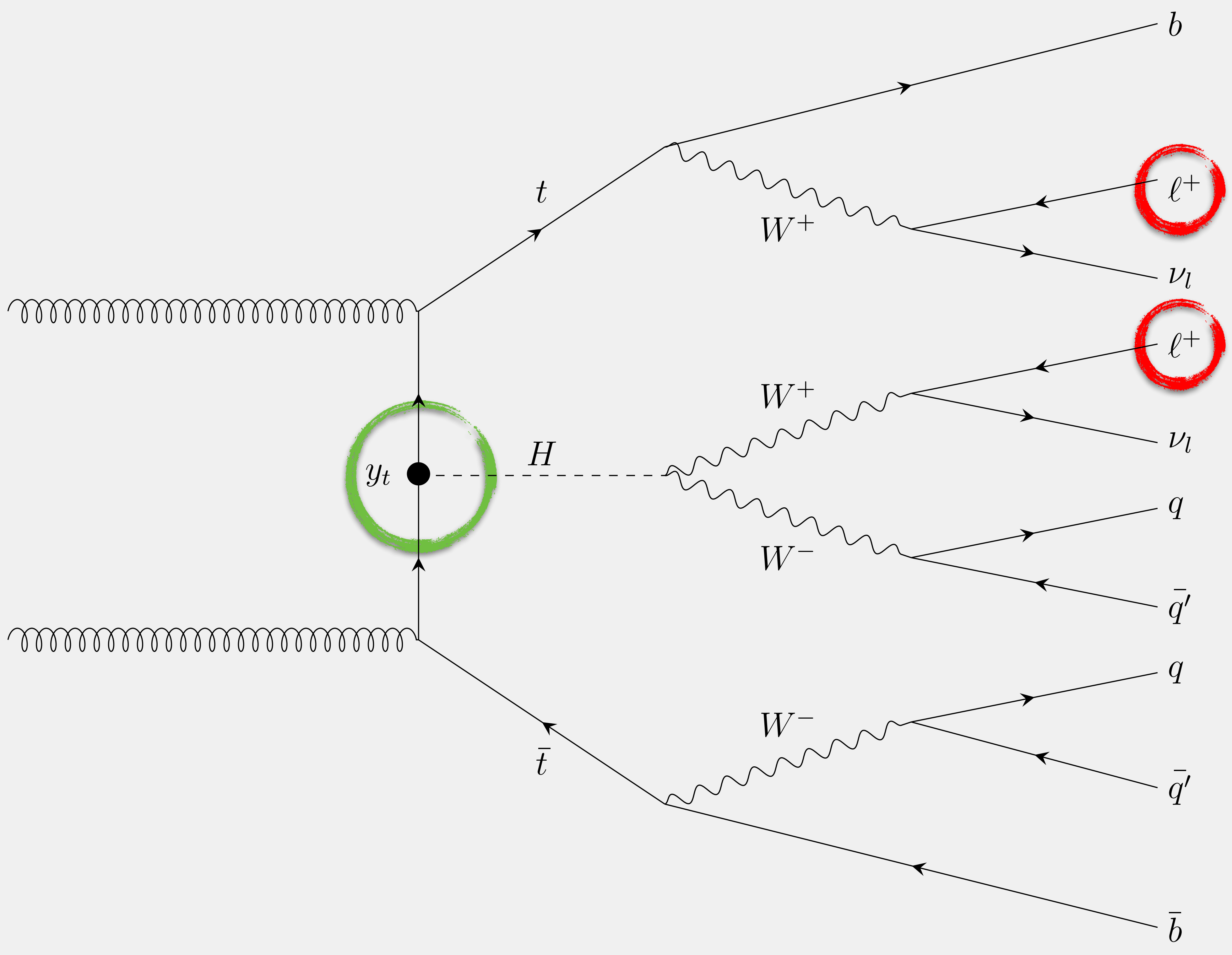


Good agreement between data and MC efficiencies
Slight increase with pile up as expected

Scale factors stable in time

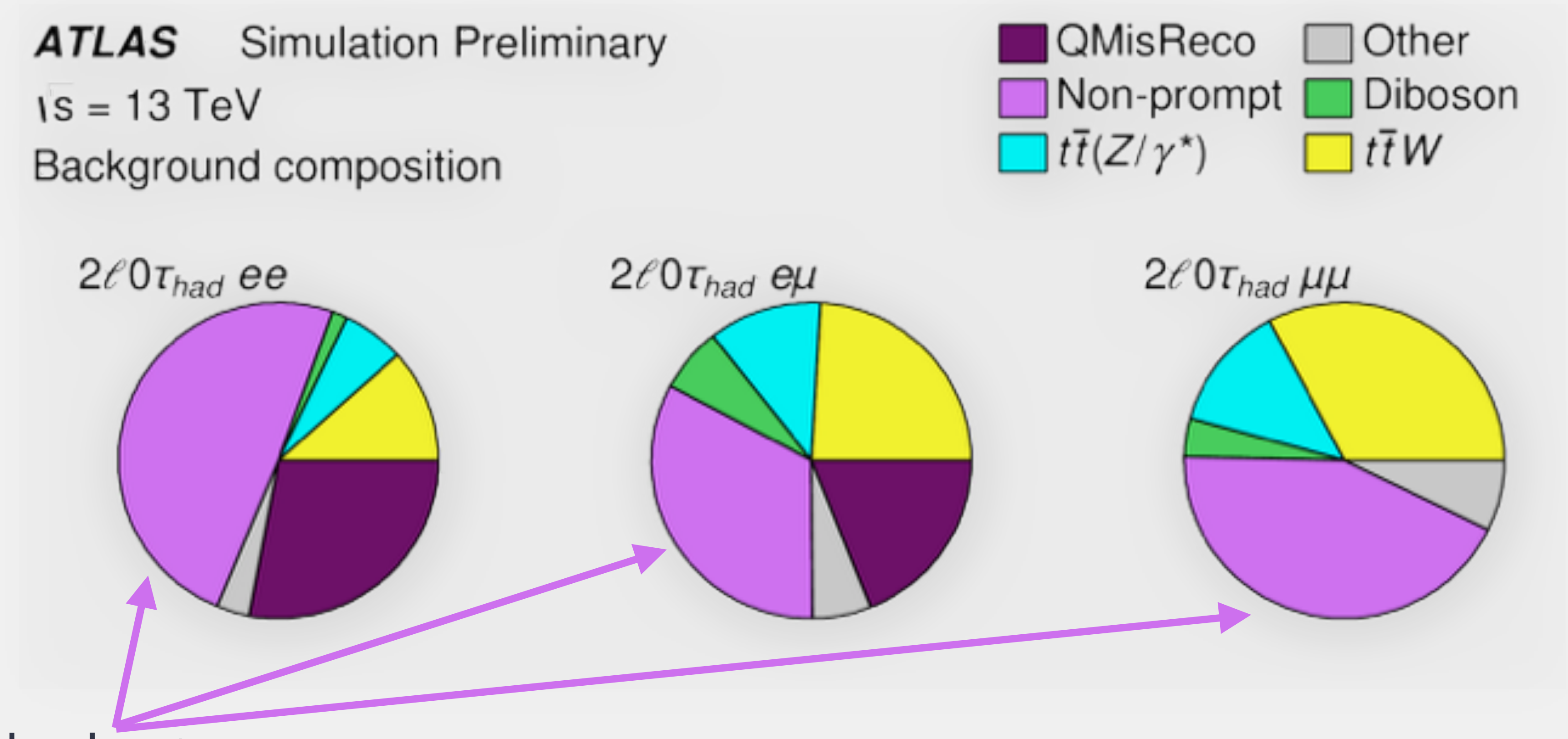
ttH in same-sign leptons channel

Fake leptons estimation



H ⟶ WW (22%), also H ⟶ ττ (6.3%) and H ⟶ ZZ (2.6%)

W ⟶ lν (21%), τ ⟶ l$\nu\nu$(35.2%), Z ⟶ ll (6.7%) for l = e, μ



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.

Conclusions and outlooks

- 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.

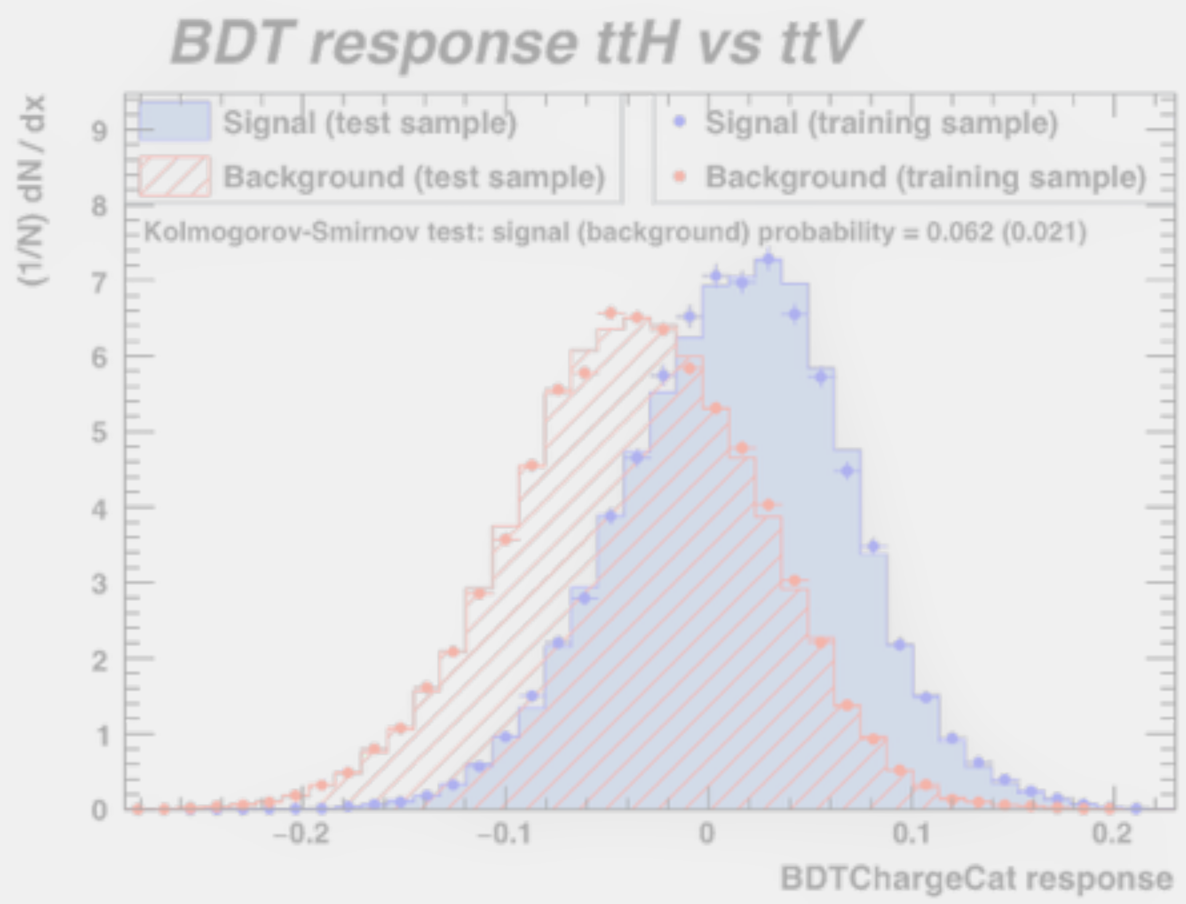
- Estimate of fake lepton background in ttH with two same sign leptons
  - First results at $\sqrt{s}$ = 13 TeV have gone public last summer at ICHEP conference (ATLAS-CONF-2016-058)

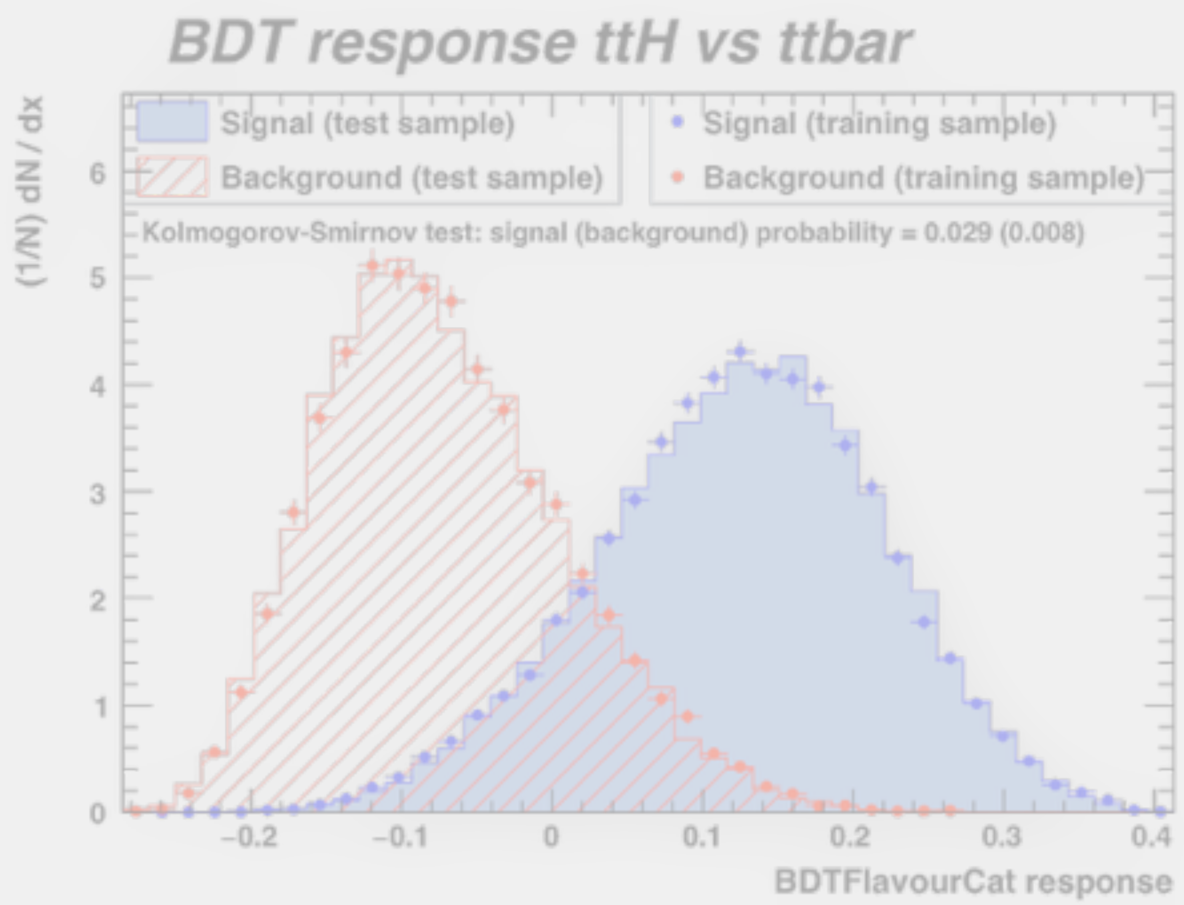
What’s next ?

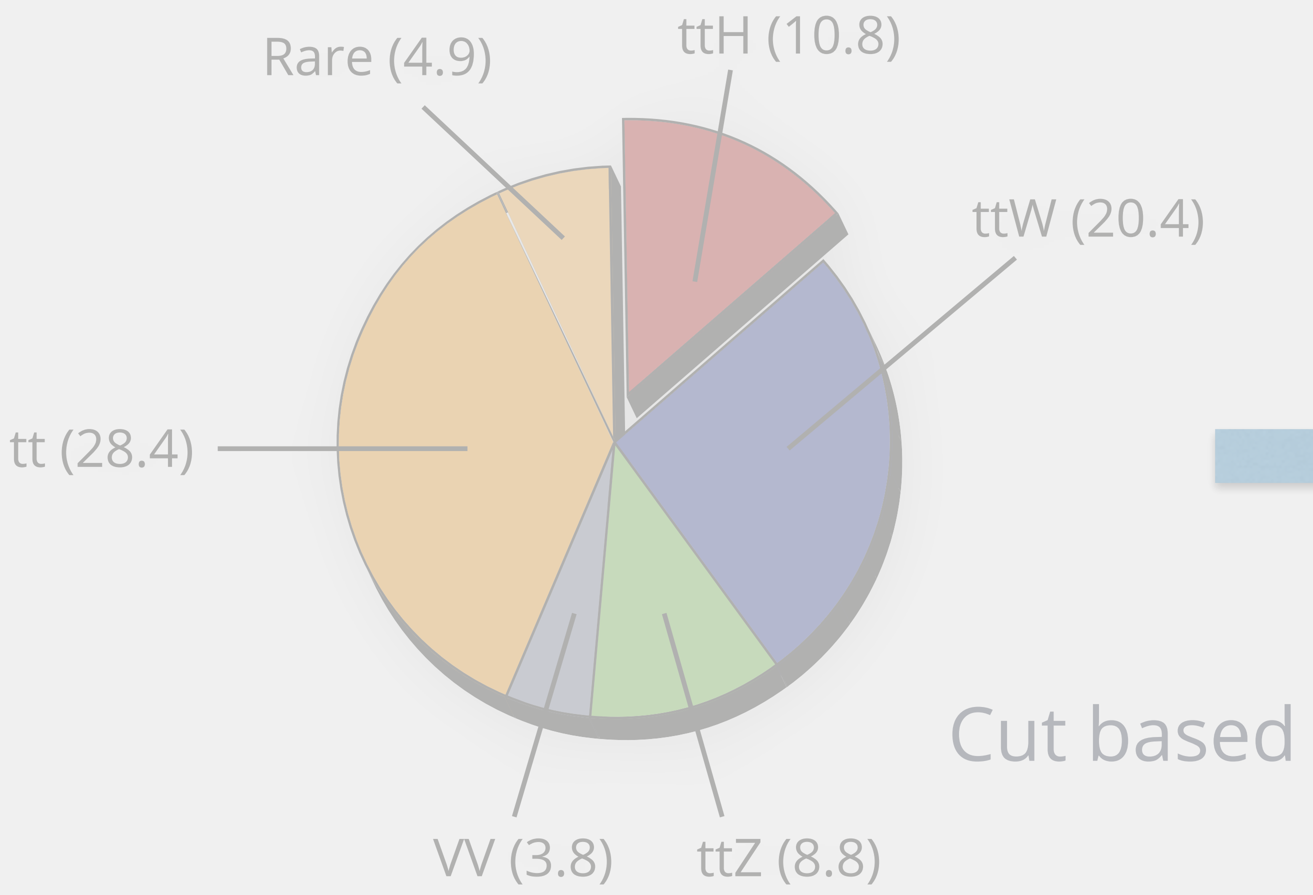
Multivariate analysis in ttH with same-sign leptons channel

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

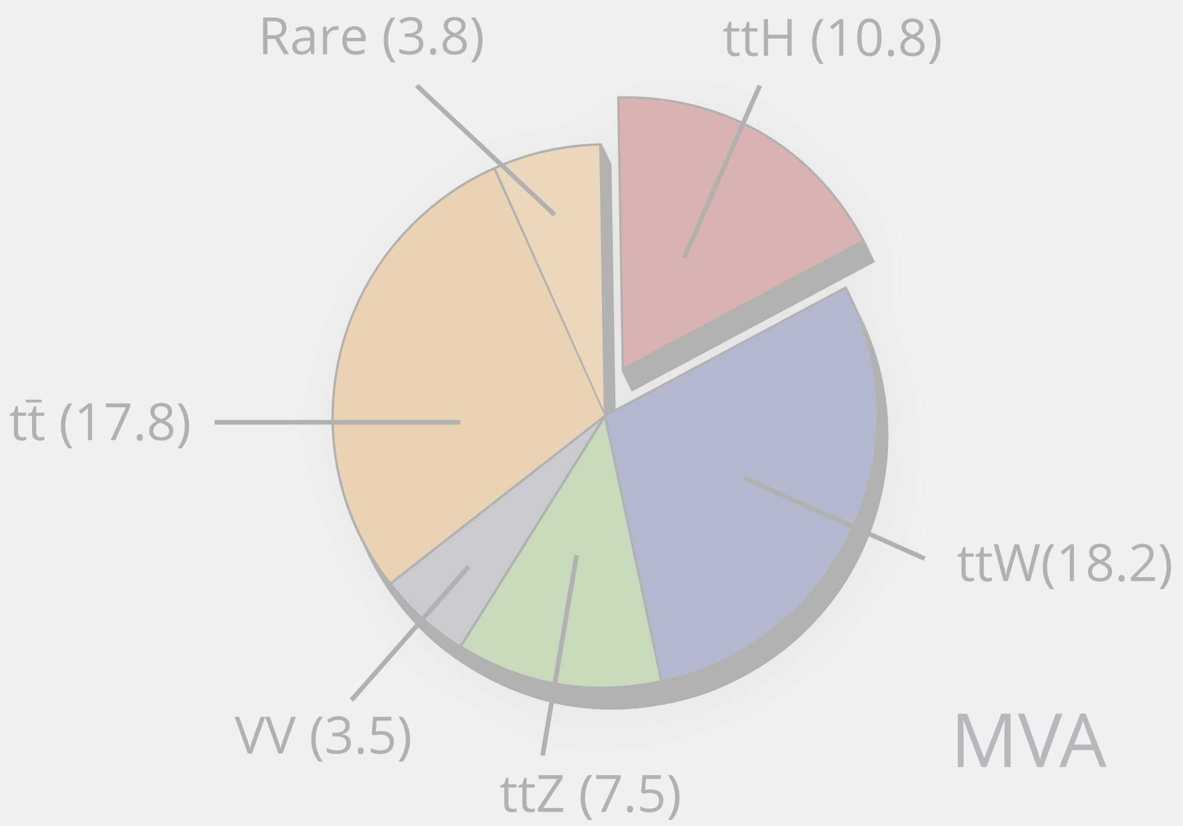
- separate ttH from fakes and ttW + ttZ (ttV) in two separate BDT
- comparison of ICHEP (13.2 fb^{-1}) cut based selection and MVA with the same signal efficiency
- results presented in the HTop workshop (3-4 Oct 2016)
- towards a publication around summer 2017

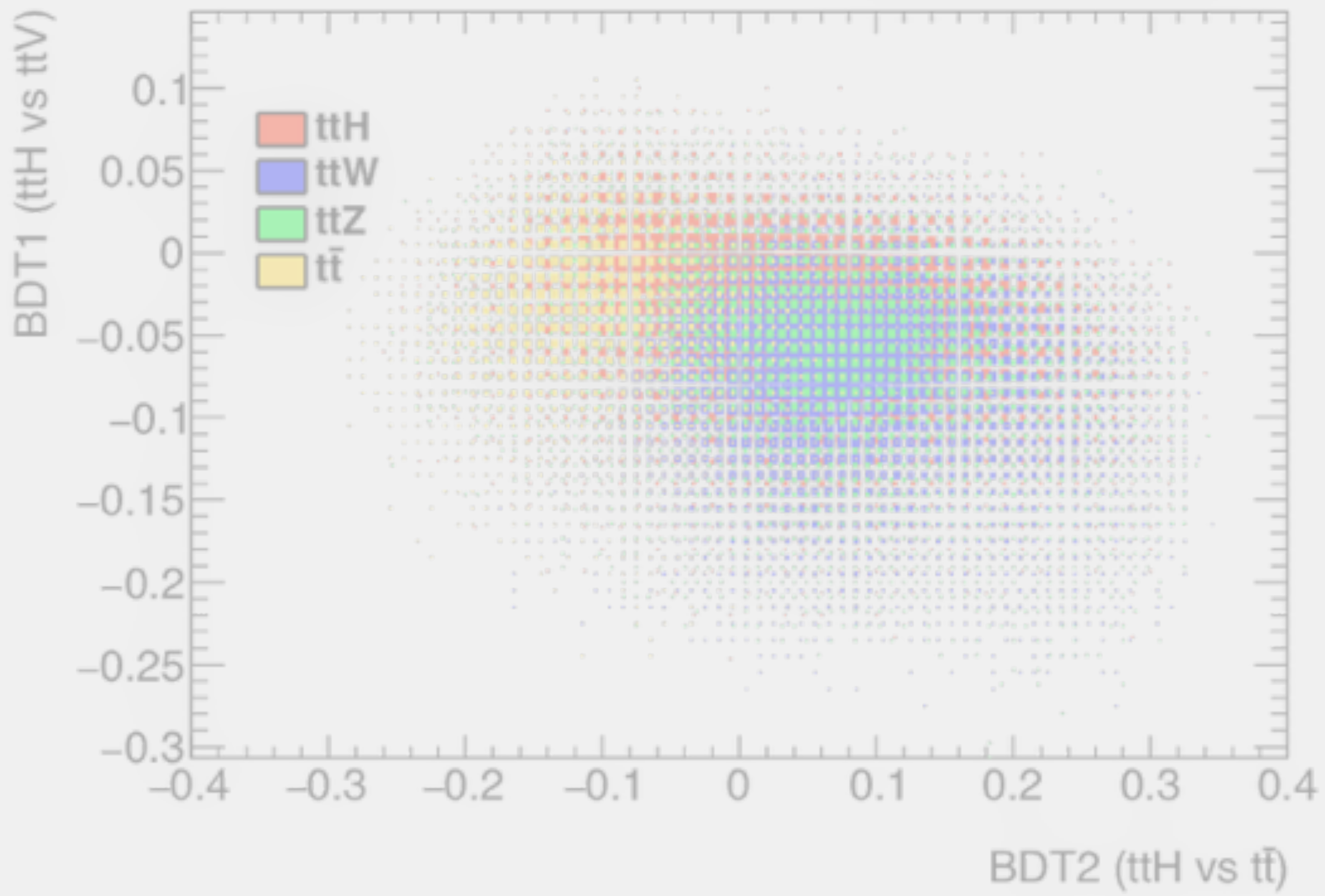






MVA 2D





Arbitrary cuts (by eye), can reduce ttbar by ~40% and ttV by ~12%. Still can be optimized.

Towards a publication around summer 2017

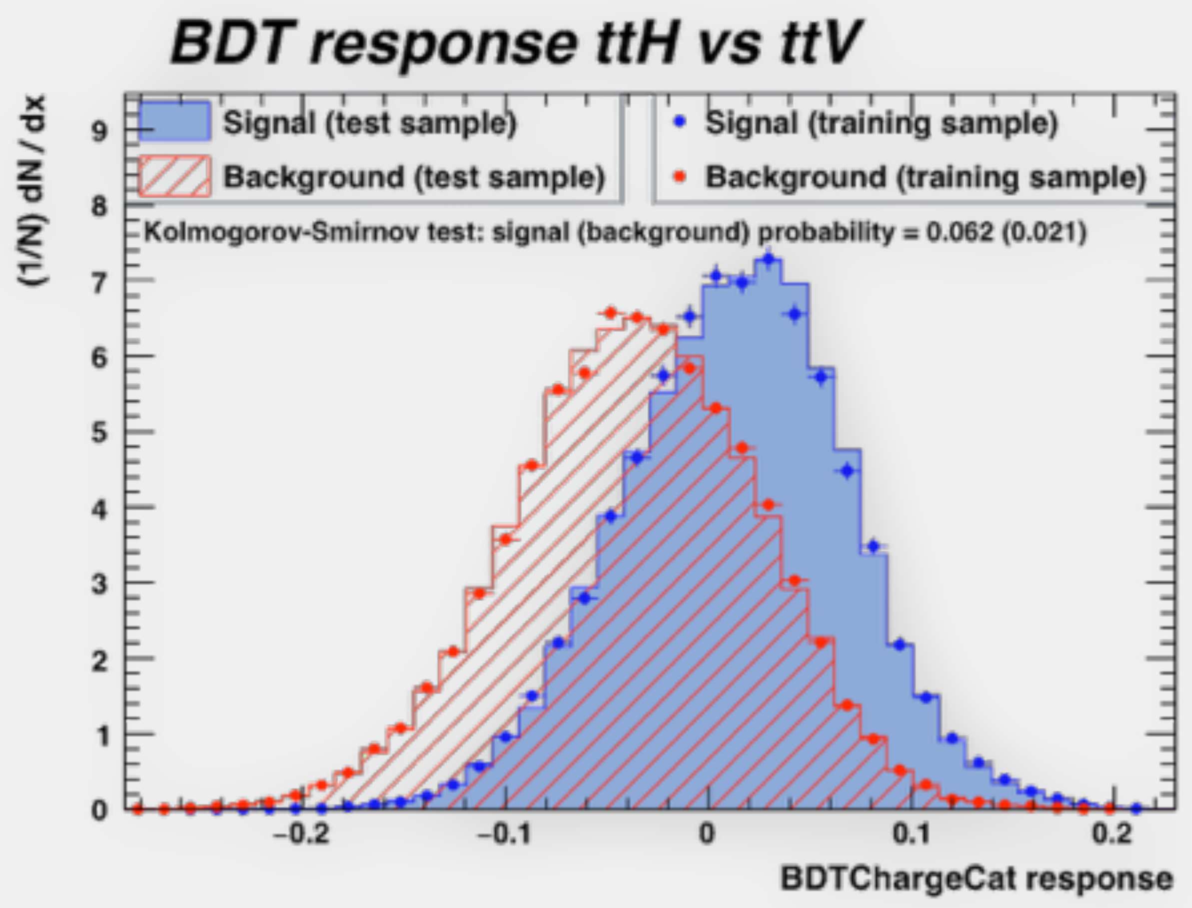
What's next ?

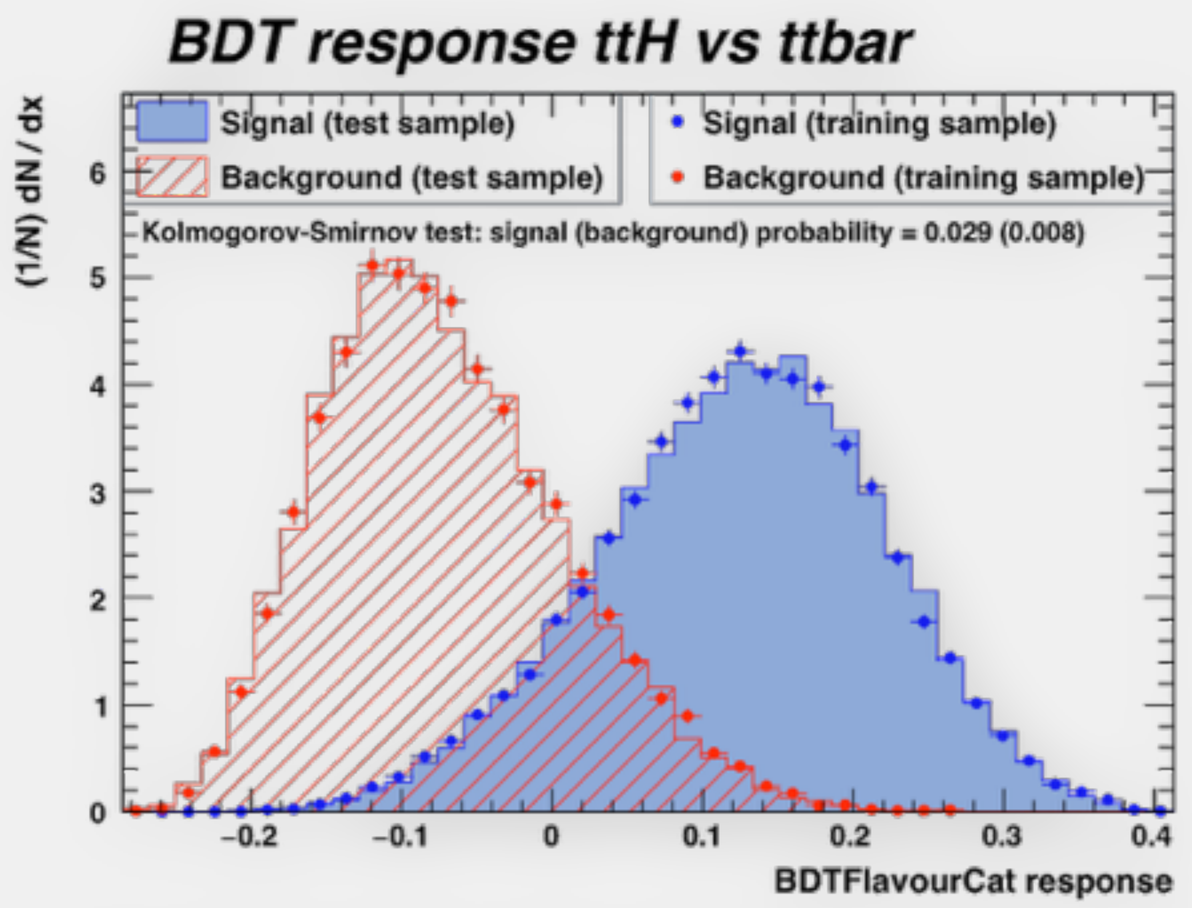
Multivariate analysis in ttH with same-sign leptons channel

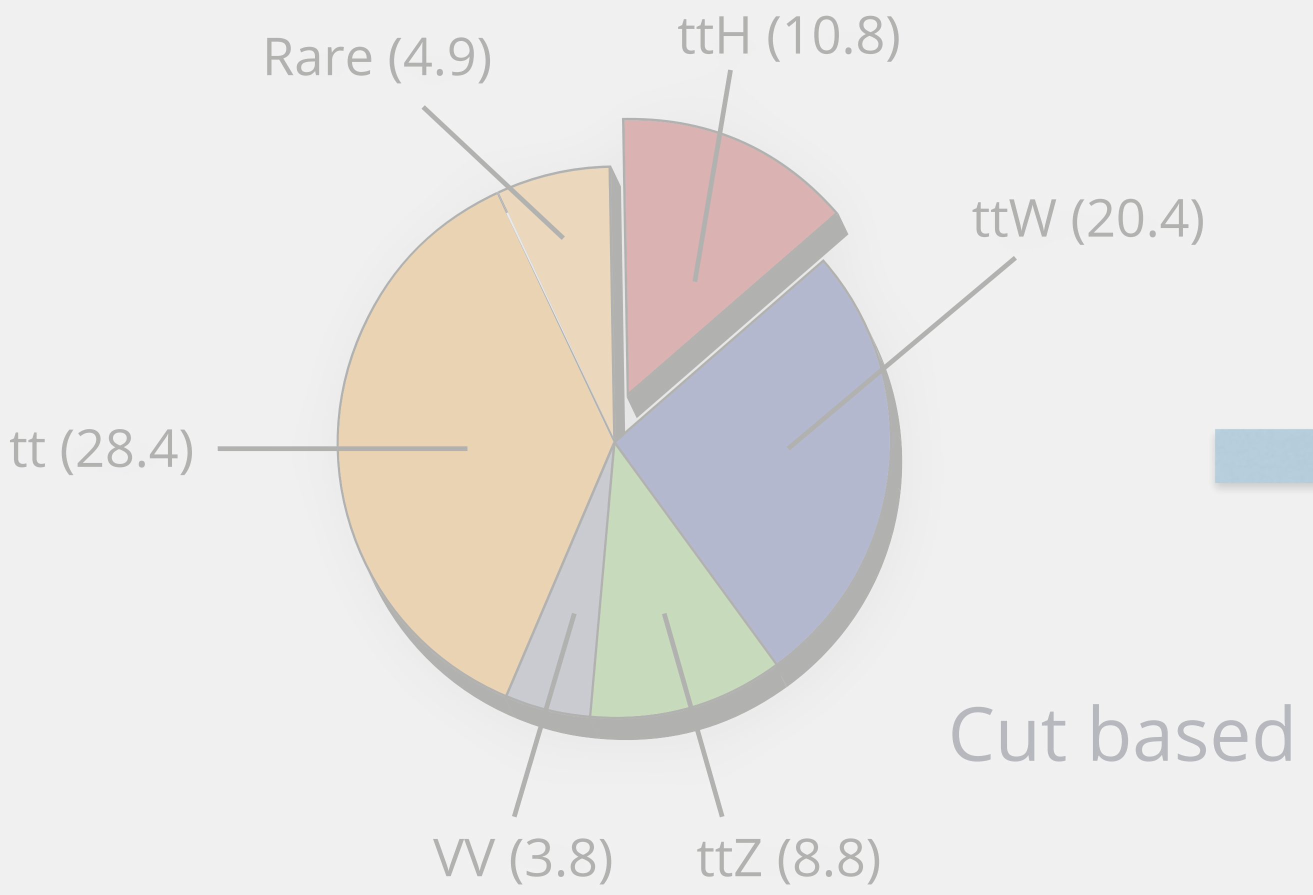
IV

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

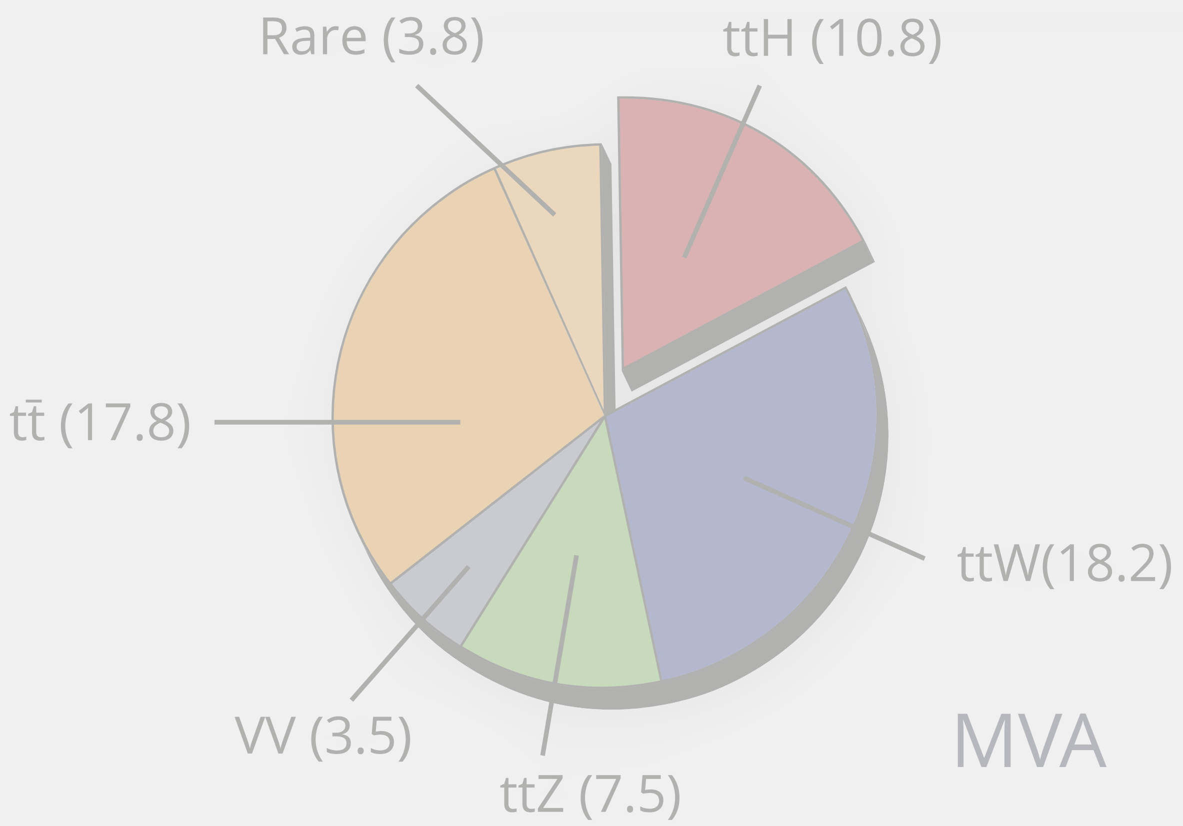
- separate ttH from fakes and ttW + ttZ (ttV) in two separate BDT
- comparison of ICHEP (13.2 fb^{-1}) cut based selection and MVA with the same signal efficiency
- results presented in the HTop workshop (3-4 Oct 2016)
- towards a publication around summer 2017

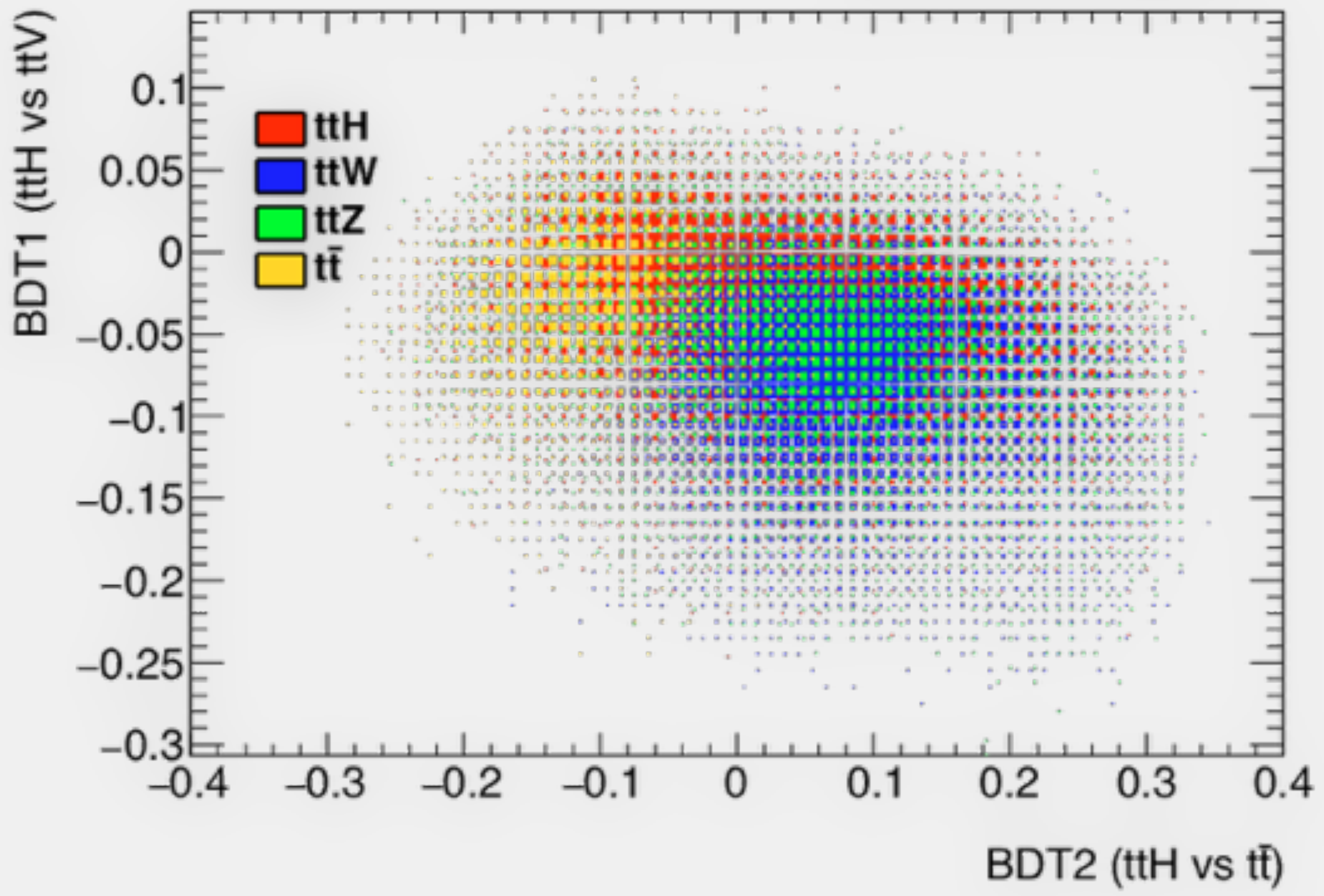






MVA 2D





Arbitrary cuts (by eye), can reduce ttbar by ~40% and ttV by ~12%. Still can be optimized.

Towards a publication around summer 2017

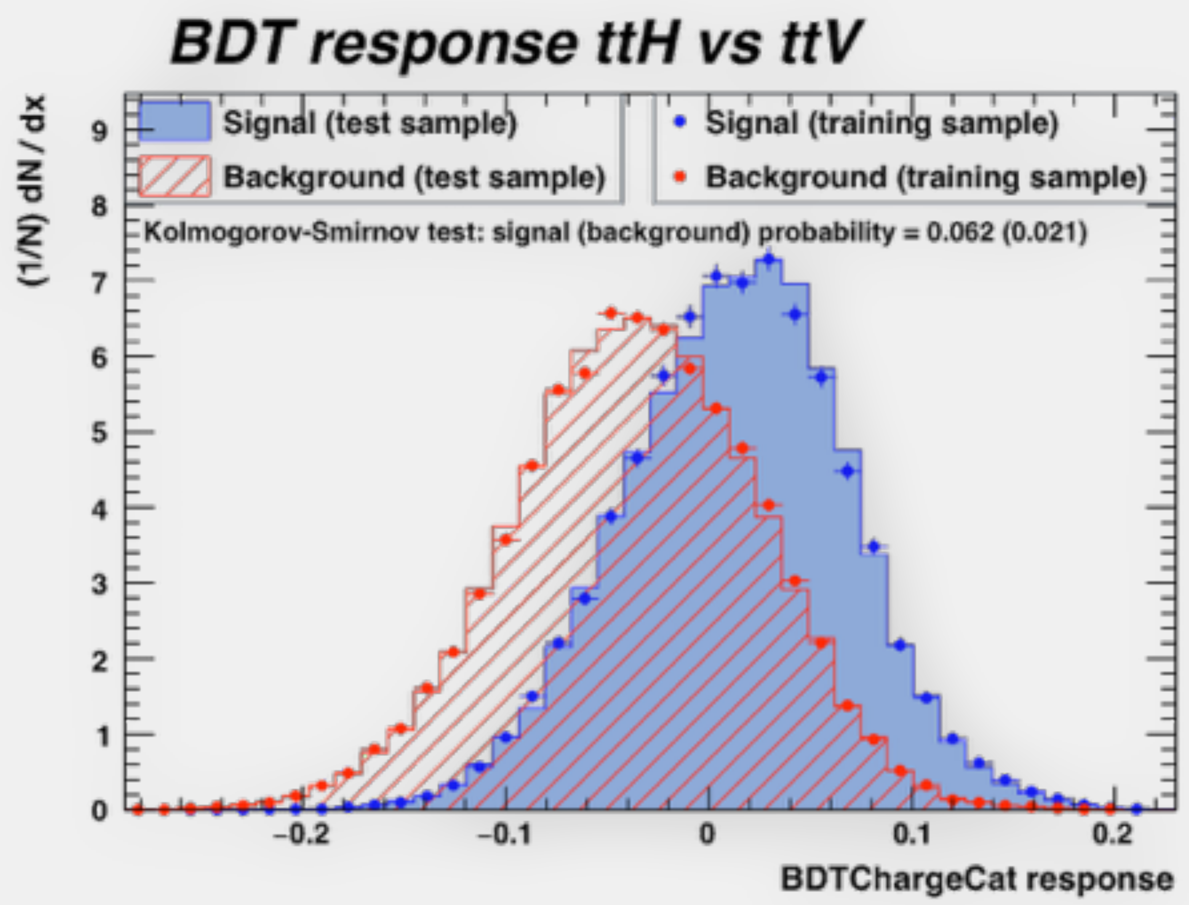
What’s next ?

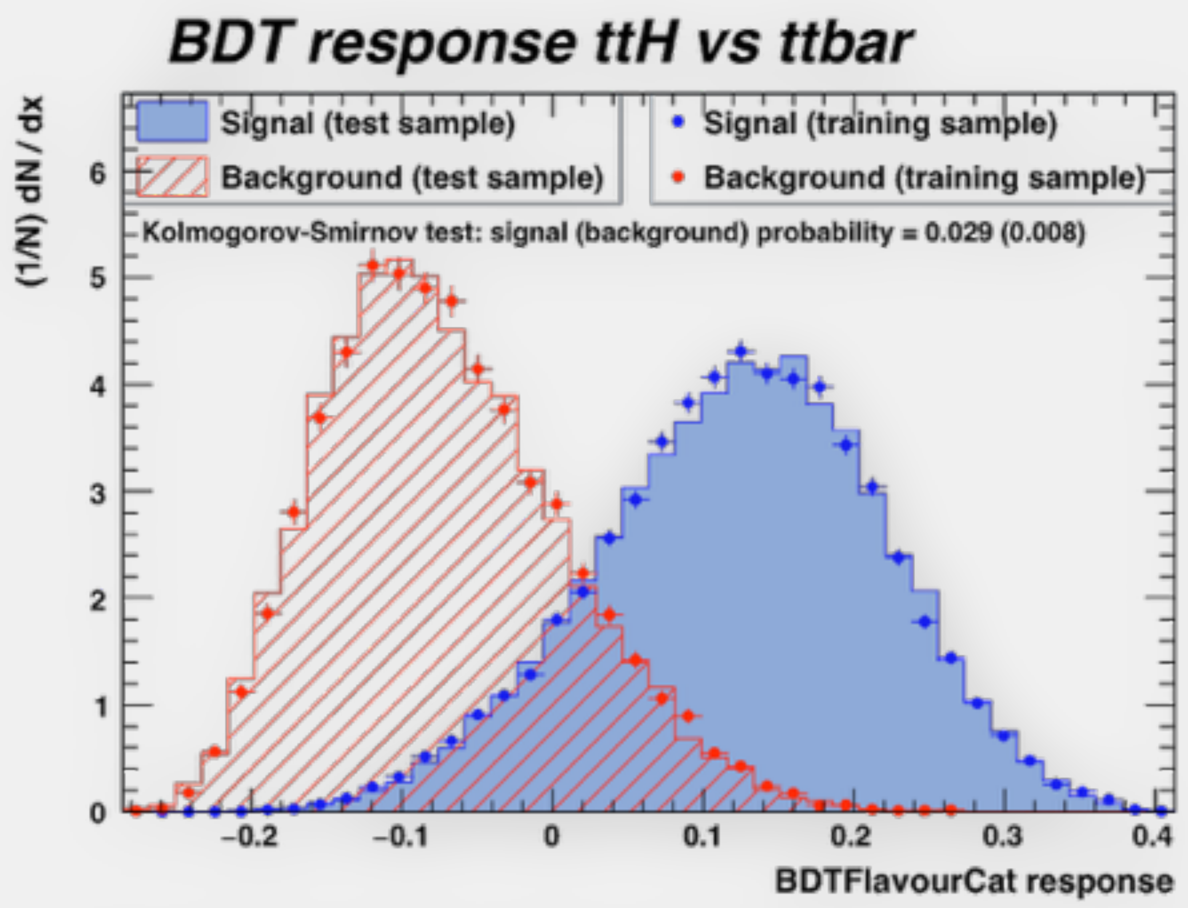
Multivariate analysis in ttH with same-sign leptons channel

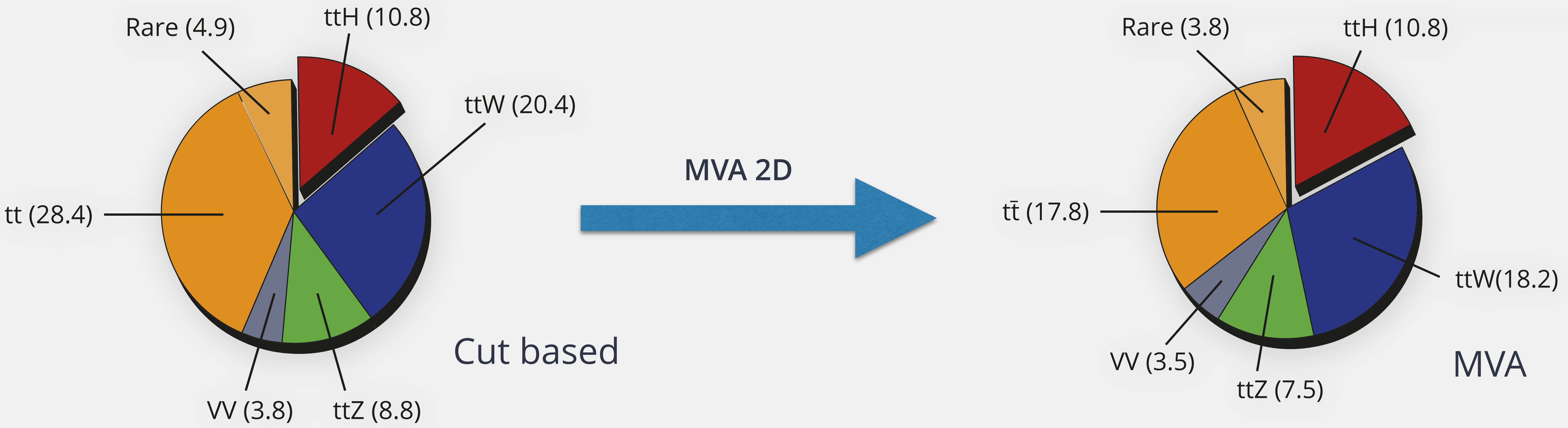
IV

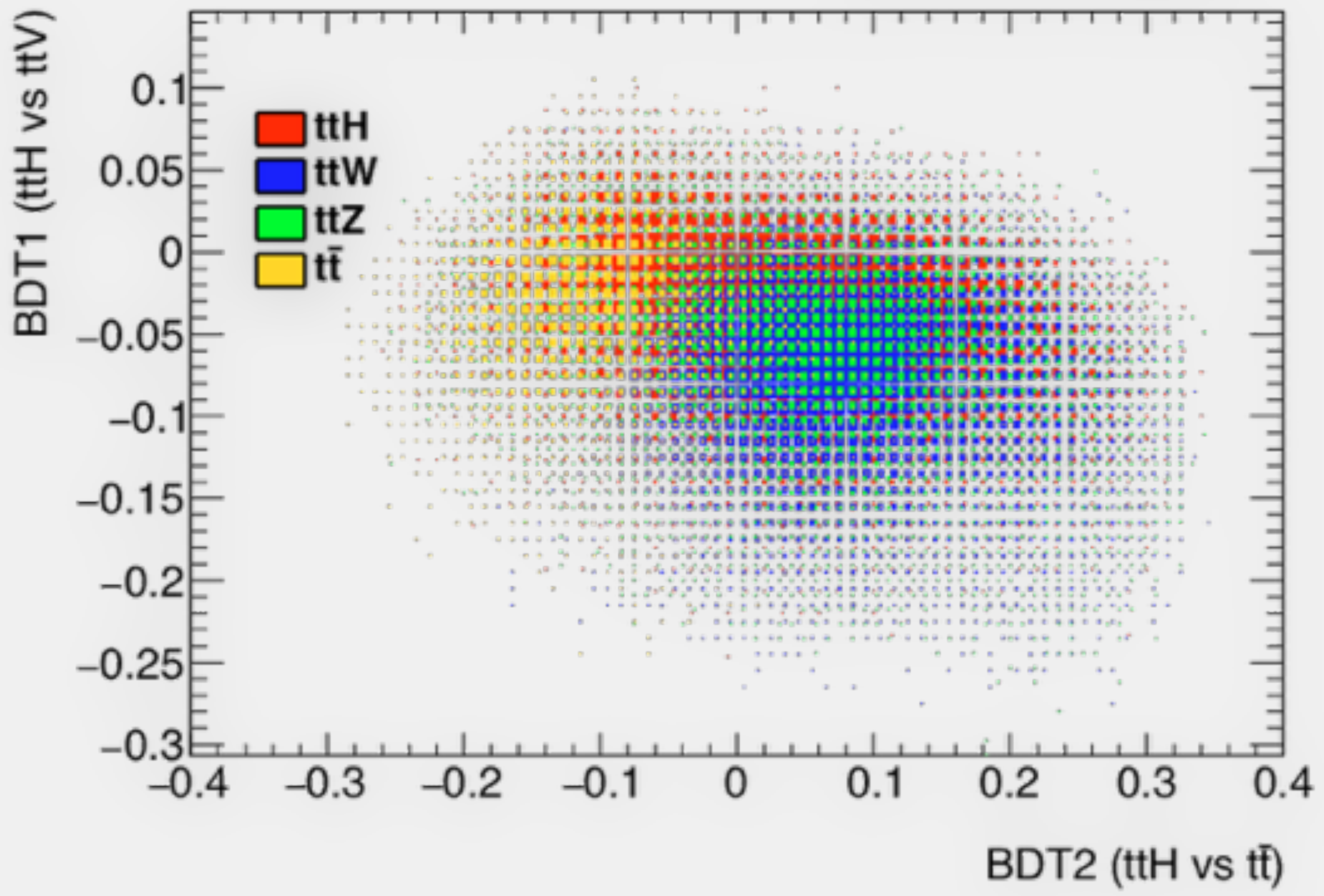
First MVA trial in the ttH multileptons group to replace signal regions.

- separate ttH from fakes and ttW + ttZ (ttV) in two separate BDT
- comparison of ICHEP (13.2 fb^{-1}) cut based selection and MVA with the same signal efficiency
- results presented in the HTop workshop (3-4 Oct 2016)









Arbitrary cuts (by eye), can reduce ttbar by ~40% and ttV by ~12%. Still can be optimized.

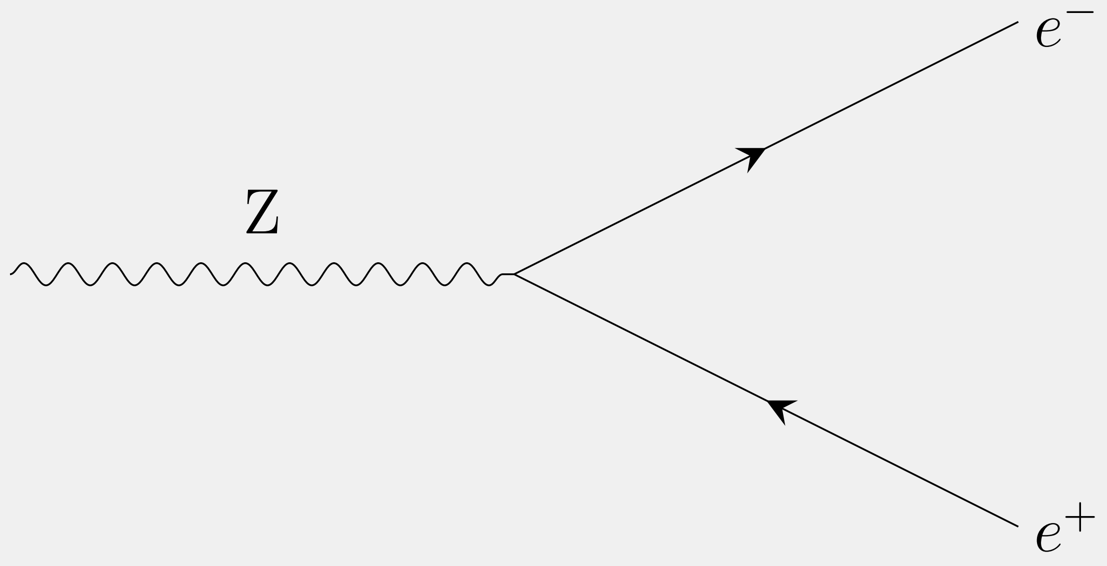
Towards a publication around summer 2017

Backup

Tag and Probe method

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

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



Other channels

Higher production cross section (1 order of magnitude):
$\longrightarrow$ $W^\pm \longrightarrow e^\pm\nu$
But more complex to study.

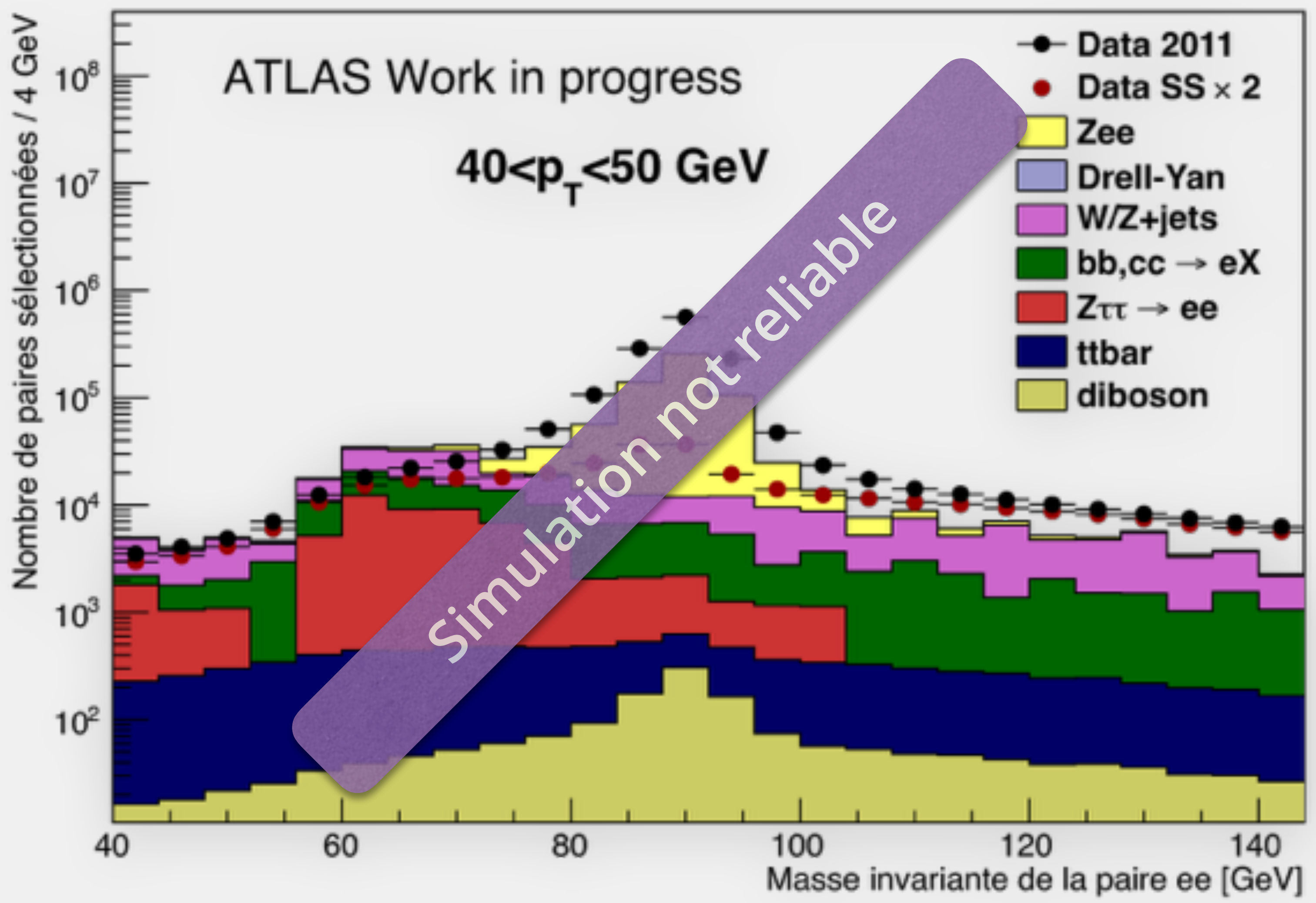
High hadronic background for low pT (< 15 GeV) and few Zee events $\longrightarrow$ $J/\psi \longrightarrow e^+e^-$

ATLAS measurements combine those three channels for ID

Main challenge for a precision measurement :
$\longrightarrow$ Background estimate
But: For electron reconstruction: probe is a cluster
$\longrightarrow$ High and complex background

$\Rightarrow$ Only Zee is usable for reco **+** only for E_T > 15 GeV

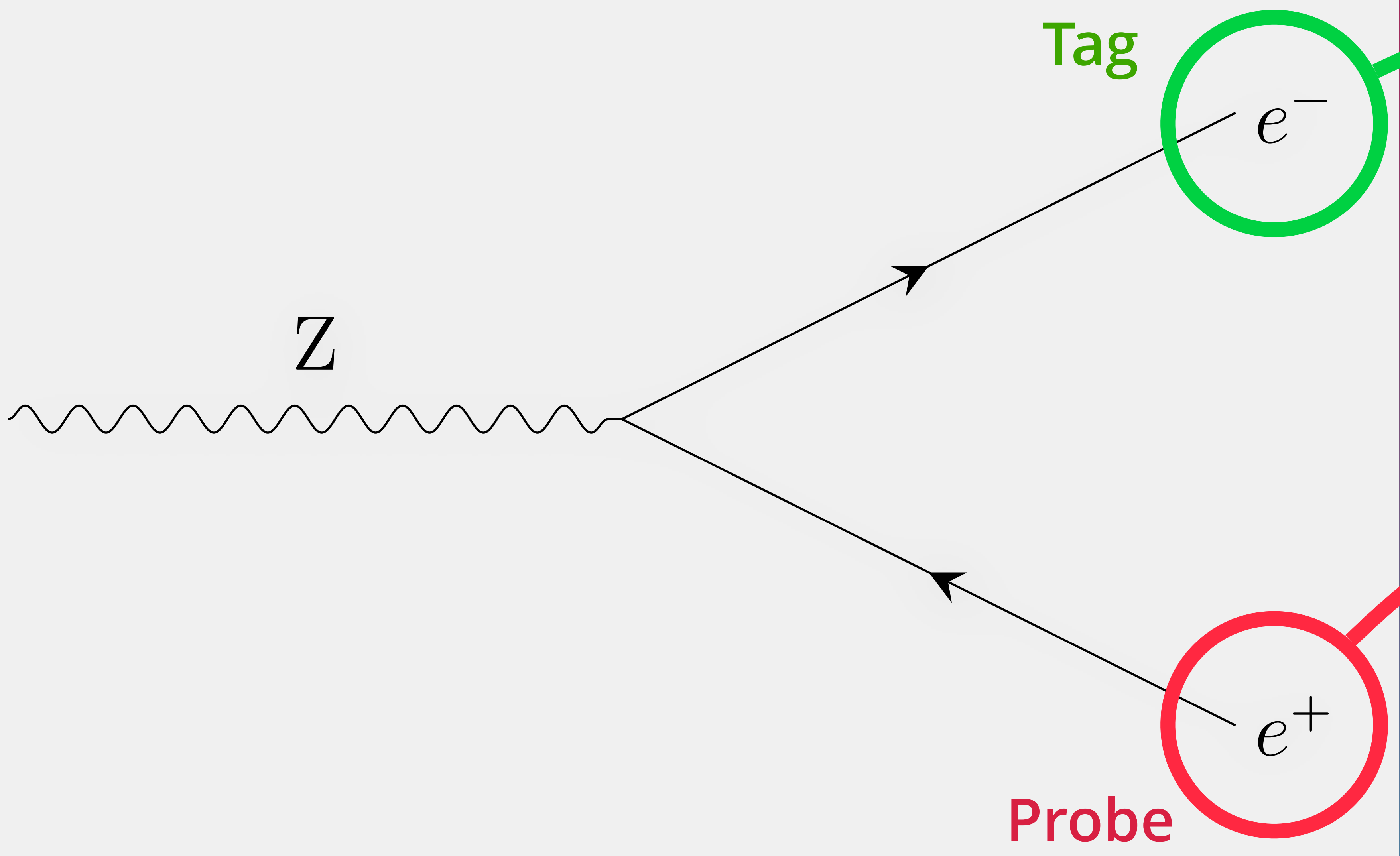
+



$\Rightarrow$ **Data driven measurements** are needed for a % or sub-% level accuracy.

Tag and Probe method

In Zee channel - Detailed selection



If $M_{ee} \sim M_Z$:

Tag ⟶ identify the Z decay event
Probe ⟶ calculate its efficiency

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

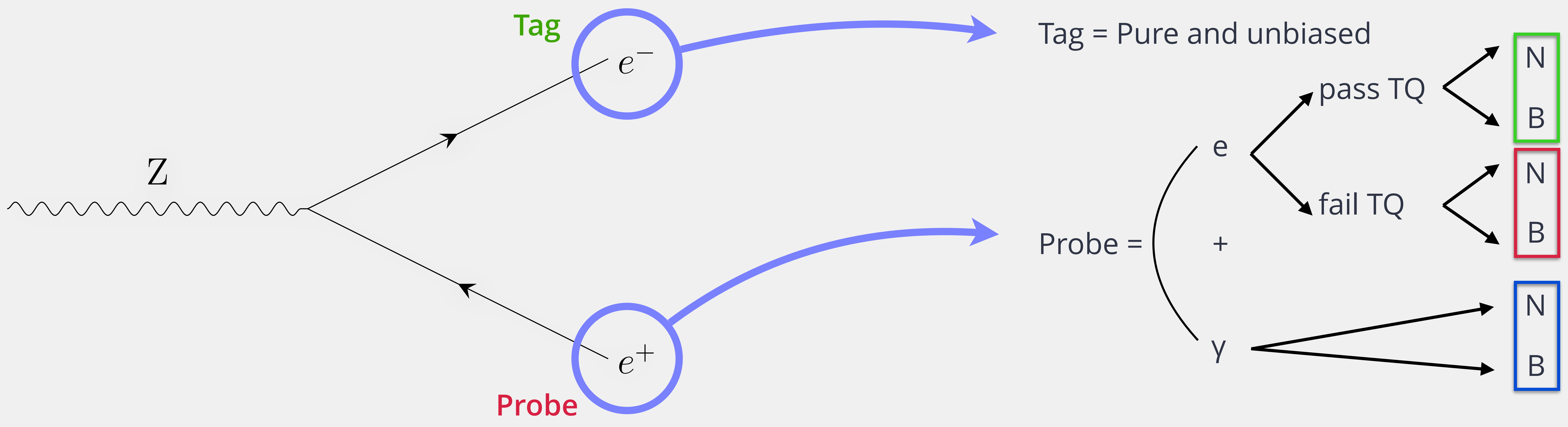
- $p_T > 15$ GeV
- $\Delta R > 0.4$ (jet pT > 20 GeV)
- M_{ee} in]80, 100[GeV
- + if MC; probe from Z decay (truth level)

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)

Tag and Probe method

In Zee channel



$$\varepsilon_{\mathrm{reco}} = \frac{N^{\mathrm{sig}}_{\mathrm{pass}}}{N^{\mathrm{sig}}_{\mathrm{pass}} + N^{\mathrm{sig}}_{\mathrm{fail}}} = \frac{\boxed{N^e_p - B^e_p}}{\boxed{(N^e_p - B^e_p)} + \boxed{(N^e_F - B^e_F)} + \boxed{(N^\gamma - B^\gamma)}}$$

Reconstructed candidates passing the probe track quality criteria

Reconstructed candidates failing track quality criteria

Electrons reconstructed as clusters with no associated track

Efficiency = $\left(\dfrac{\text{Association track - EM cluster} + \text{track quality}}{\text{All clusters with } M_{\text{tag-probe}} \sim M_Z} \right)$

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 >= 30 GeV
Variation 1	fail at least 2 loose++ cuts	$topoE_T^{cone30}/p_T > 0.02$ *$120 < m_{ee} < 250$ GeV*	$topoE_T^{cone40}/p_T > 0.05$ *$120 < m_{ee} < 250$ GeV*
Variation 2		$topoE_T^{cone30}/p_T > 0.02$ *$60 < m_{ee} < 70$ GeV*	$topoE_T^{cone40}/p_T > 0.20$ *$120 < m_{ee} < 250$ 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:

$$B^e = N_{\mathrm{peak}}^{\mathrm{template}} \times \frac{N_{\mathrm{tail}}^{e} - N_{\mathrm{tail}}^{tight++}/\varepsilon_{\mathrm{tail}}^{tight++}}{N_{\mathrm{tail}}^{\mathrm{template}}}$$

~70% for 15 < p_T < 20 GeV

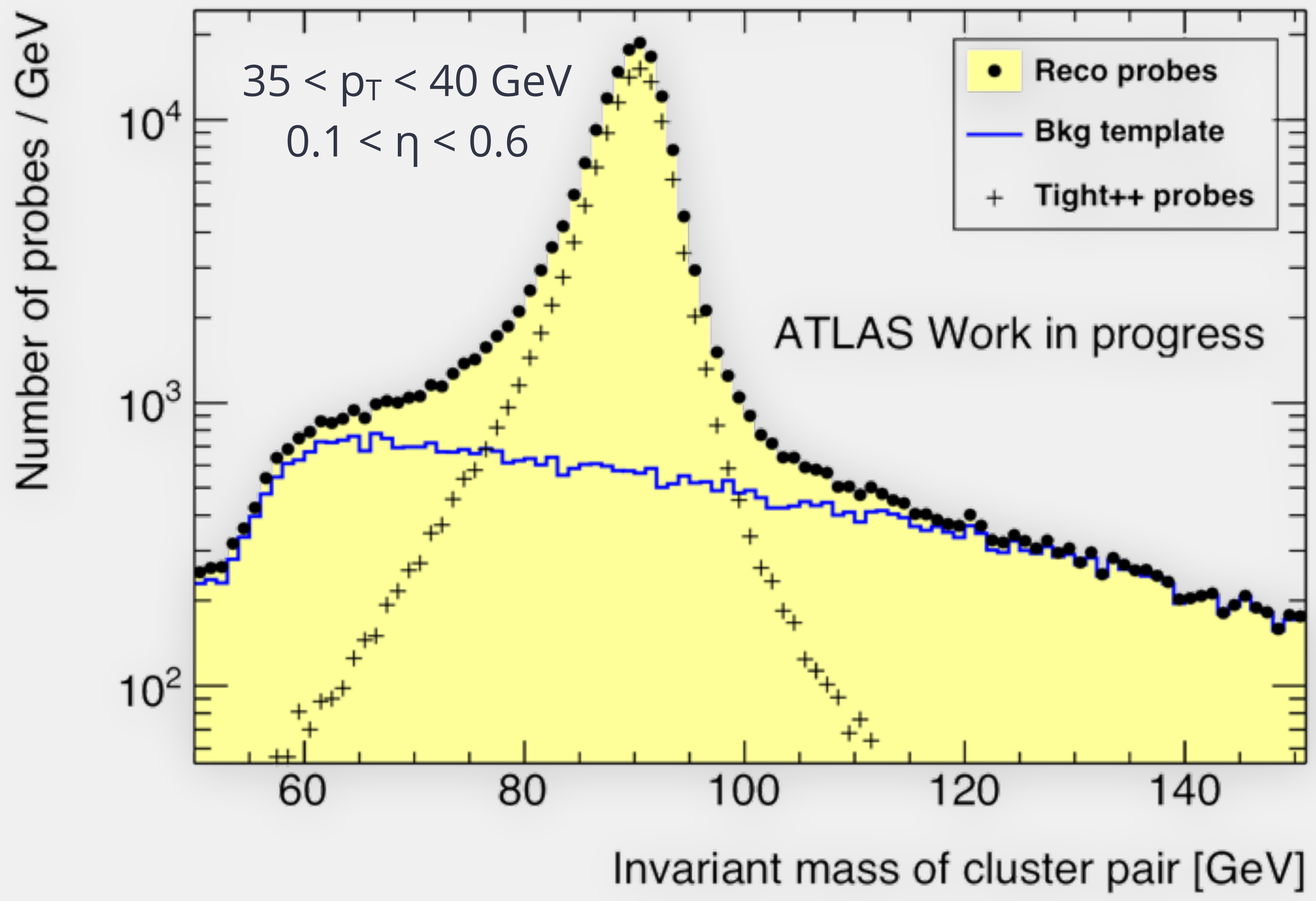
~ 5% for 45 < p_T < 50 GeV

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

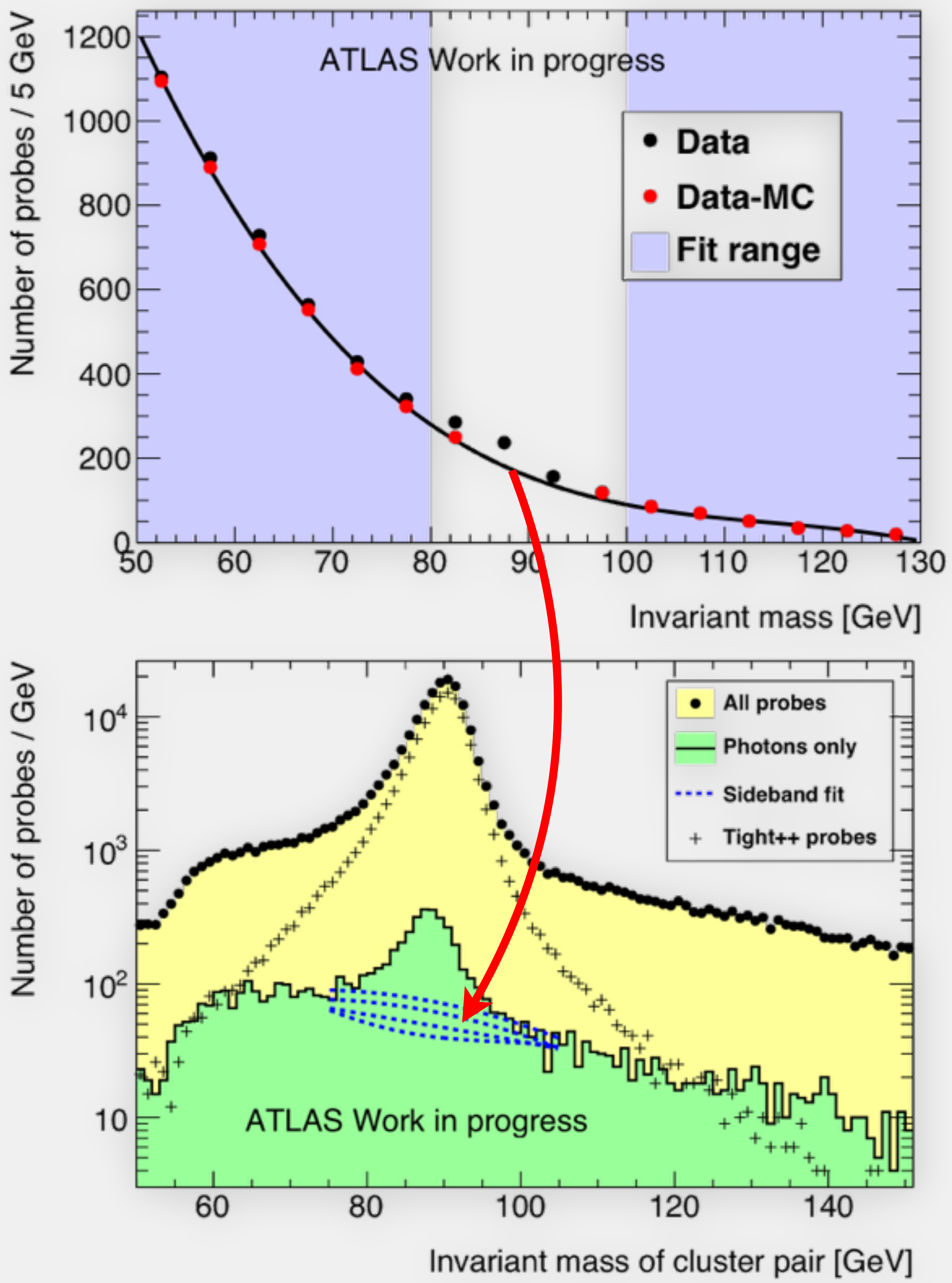
Electrons reconstructed as clusters with no associated track

$$\varepsilon_{\mathrm{reco}} = \frac{N^{\mathrm{sig}}_{\mathrm{pass}}}{N^{\mathrm{sig}}_{\mathrm{pass}} + N^{\mathrm{sig}}_{\mathrm{fail}}} = \frac{N^e_p - B^e_p}{(N^e_p - B^e_p) + (N^e_F - B^e_F) + \underbrace{\boxed{(N^\gamma - B^\gamma)}}_{\Gamma}}$$

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

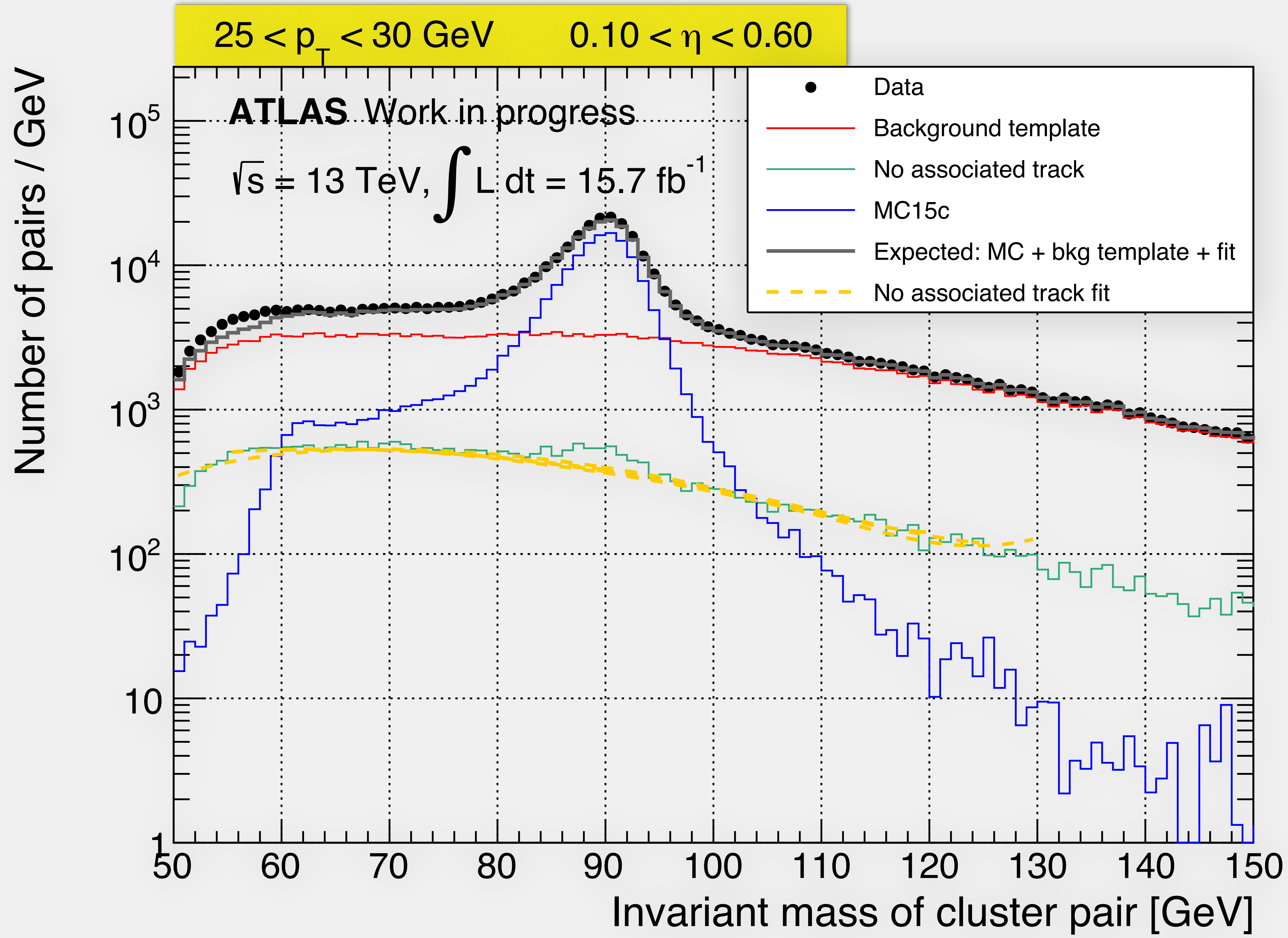
$\longrightarrow$ Simpler method

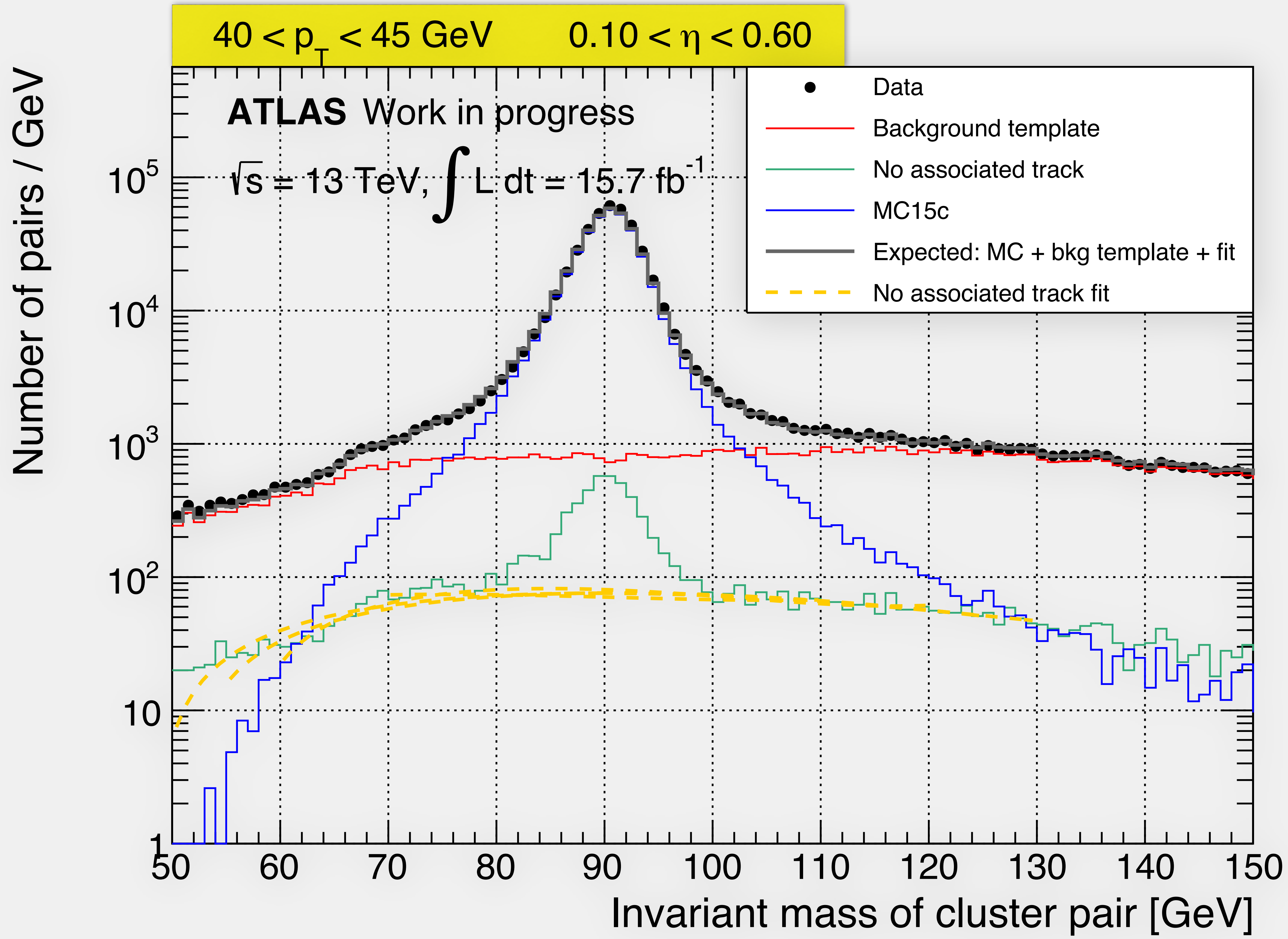
1. Construct invariant mass between electron tag and photon probe.
2. Choose side band regions outside the Z mass peak.
3. Signal in the side band regions subtracted from simulation
4. Fit the eγ-mass distribution with a 3rd order polynomial.



Electron reconstruction efficiency in 2016 data

Denominator detailed distribution





Efficiency computation

Uncertainties on ε_{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
- **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

$$\varepsilon_{\mathrm{reco}} = \varepsilon_{\mathrm{reco}} \pm \Delta^{\mathrm{stat}}\varepsilon_{\mathrm{reco}} \pm \Delta^{\mathrm{syst}}\varepsilon_{\mathrm{reco}}$$

Central value of variations

Error propagation

RMS of variations

Efficiency computation

Detailed variations

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

W/Z + jets events - QCD events without electron

Low mass range rich of bremsstrahlung electrons

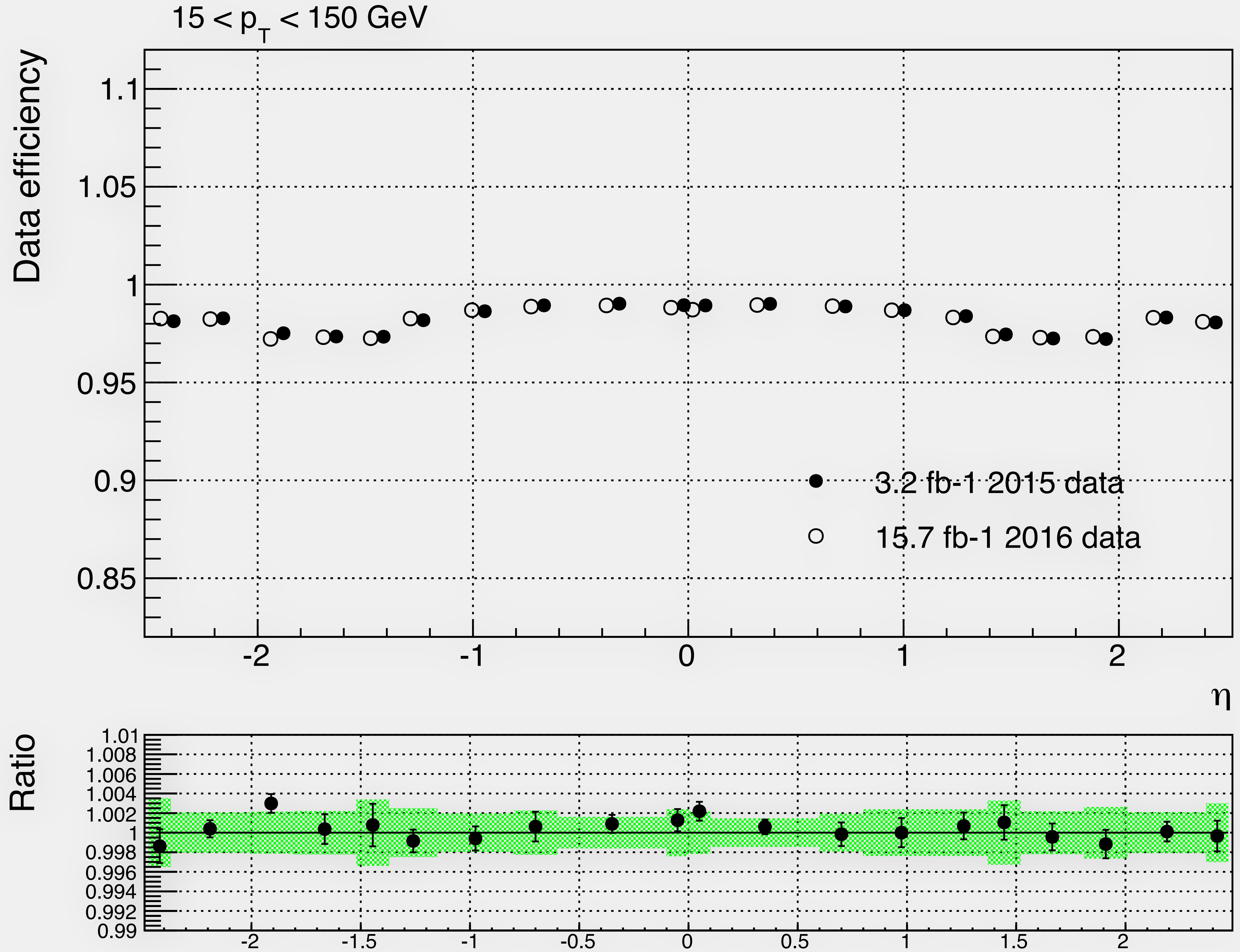
Change signal proportion

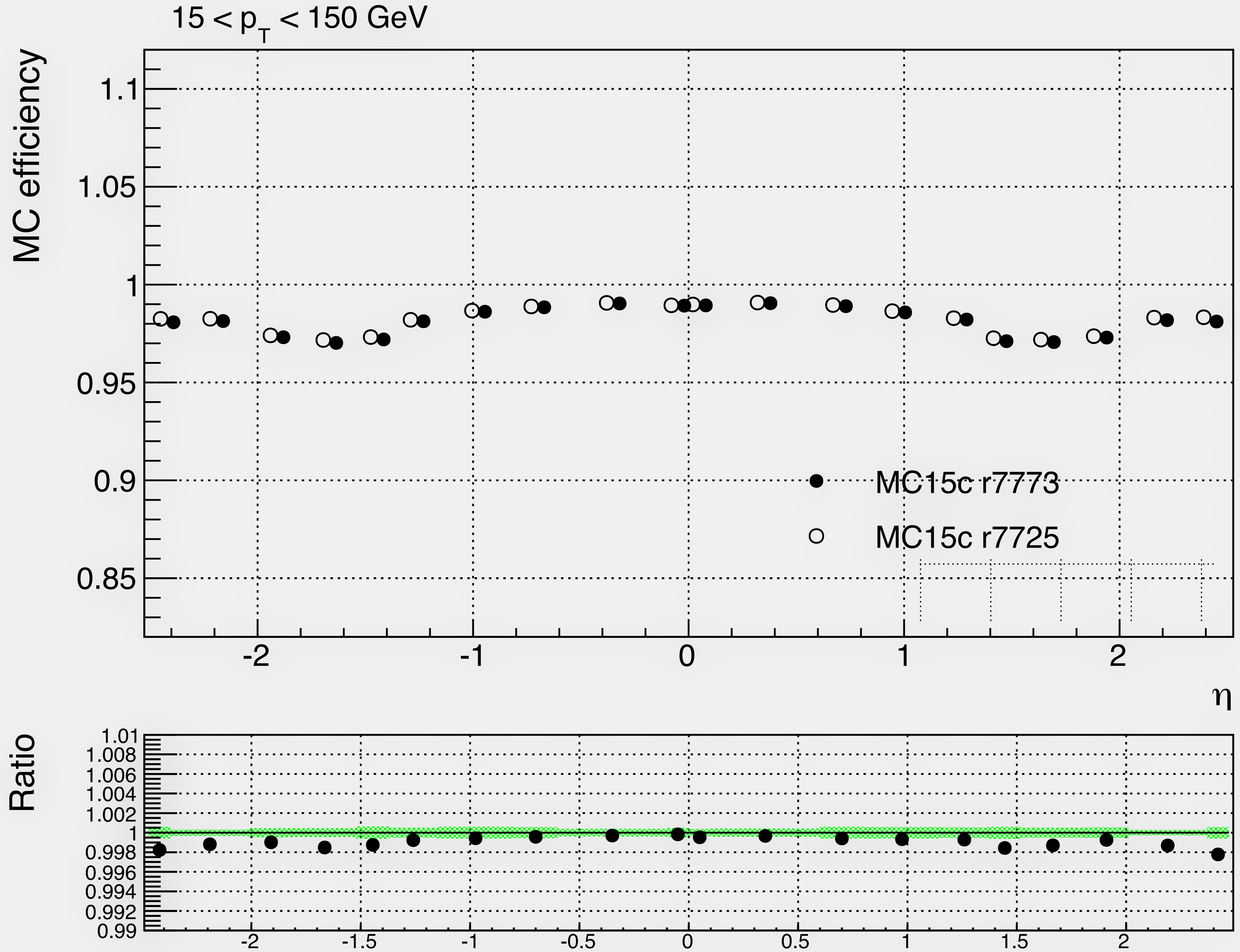
Evaluate the stability of the analytic form

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

Data - MC efficiencies

Comparison with pre-recommendations





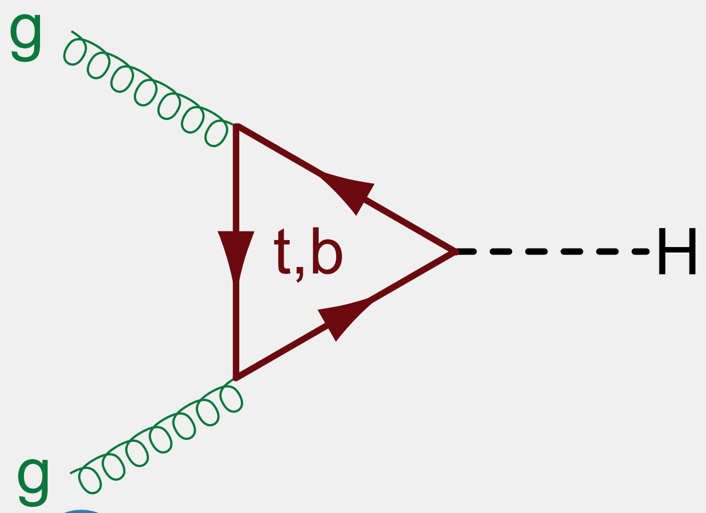
No systematic effect for data efficiency
MC efficiency slightly higher for MC r7725 resulting in slightly lower scale factors for 2016 data

Quadratic sum of 2015 and 2016 / MC r7773 - 7725 uncertainties

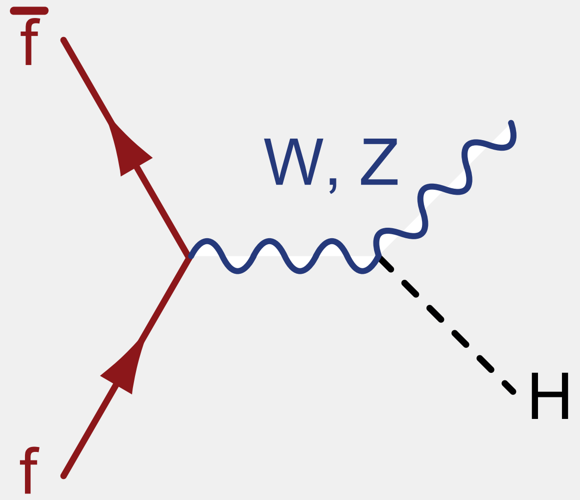
2015 data uncertainties

Higgs boson

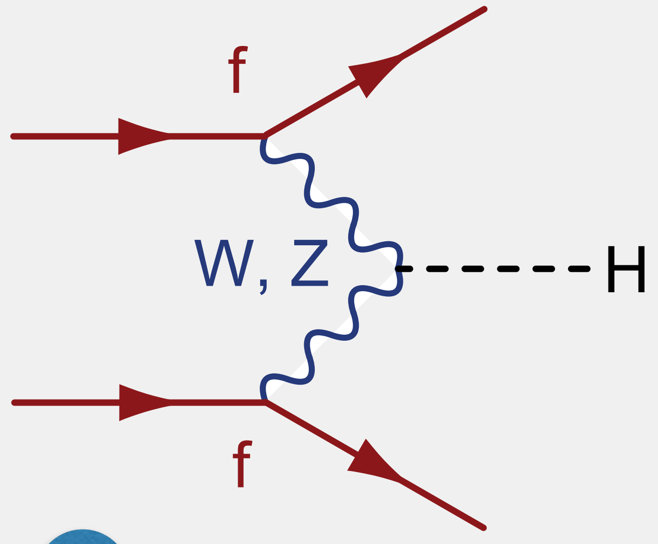
Production



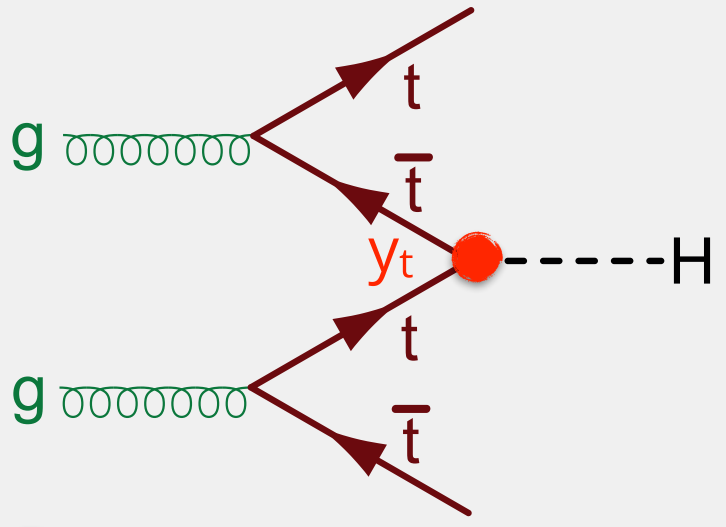
1 Gluon fusion (ggF)



3 Higgs-strahlung (WH + ZH)



2 Vector Boson Fusion (VBF)



4 Associated production with a top-quark pair ($t\bar{t}H$)

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**

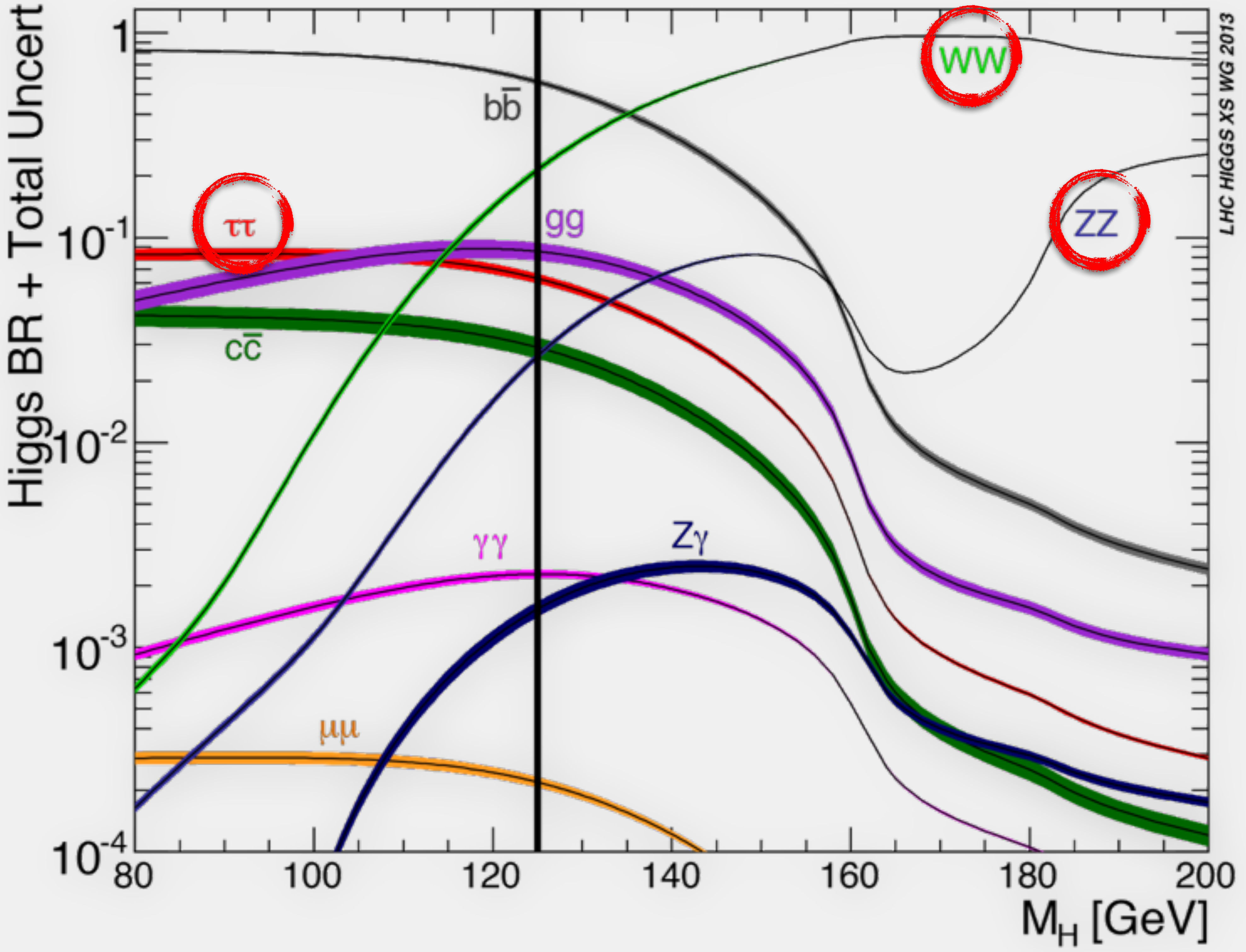
Low production cross sections

	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



×

Branching ratios

- Probability to decay in a defined final state

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

To get two same-sign leptons

- 3 Higgs decays

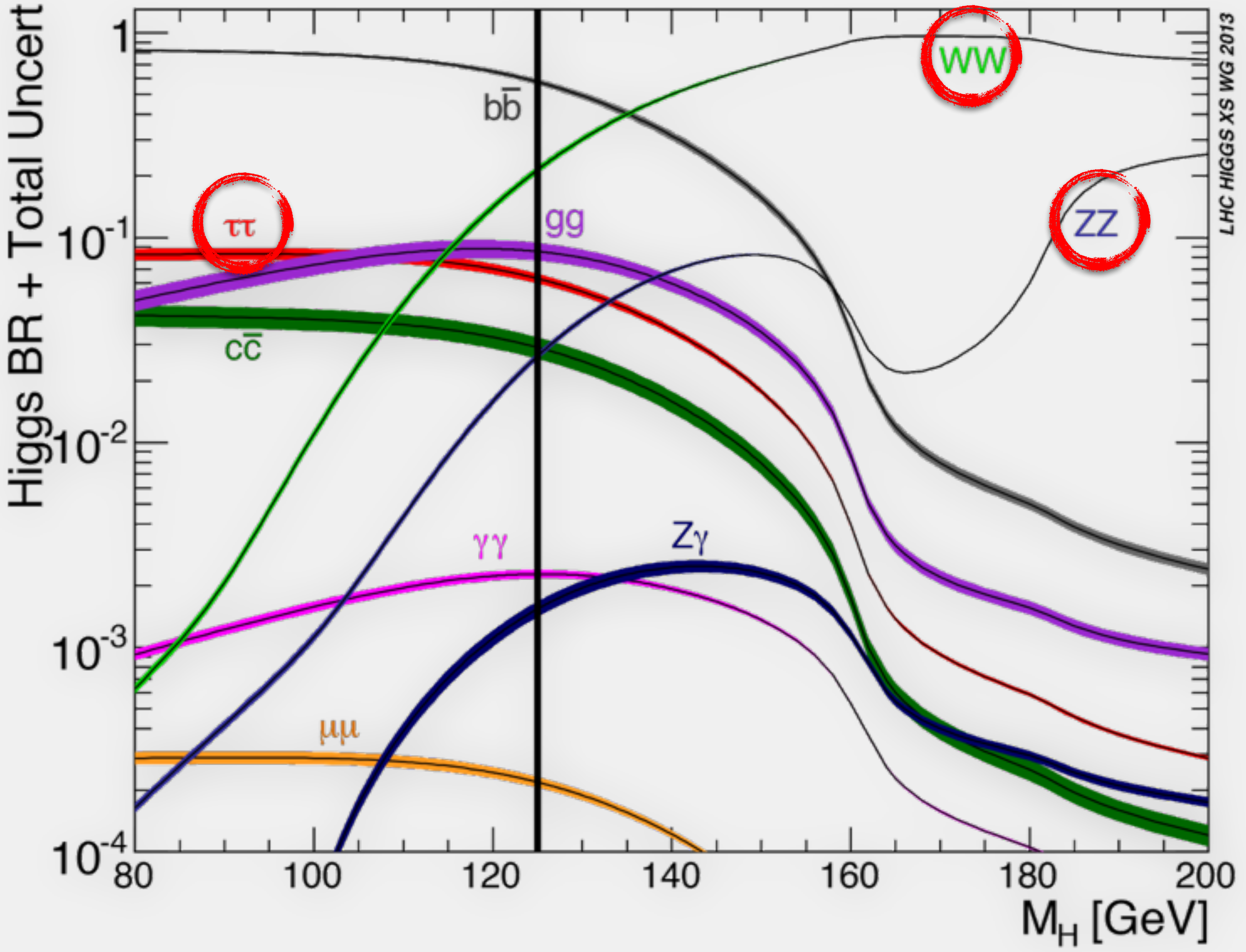
H ⟶ WW (22%)
H ⟶ ττ (6.3%)
H ⟶ ZZ (2.6%)

×

W ⟶ $l\nu$ (21%)
τ ⟶ $l\nu\nu$ (35.2%)
Z ⟶ ll (6.7%)
for l = e, μ

Higgs boson

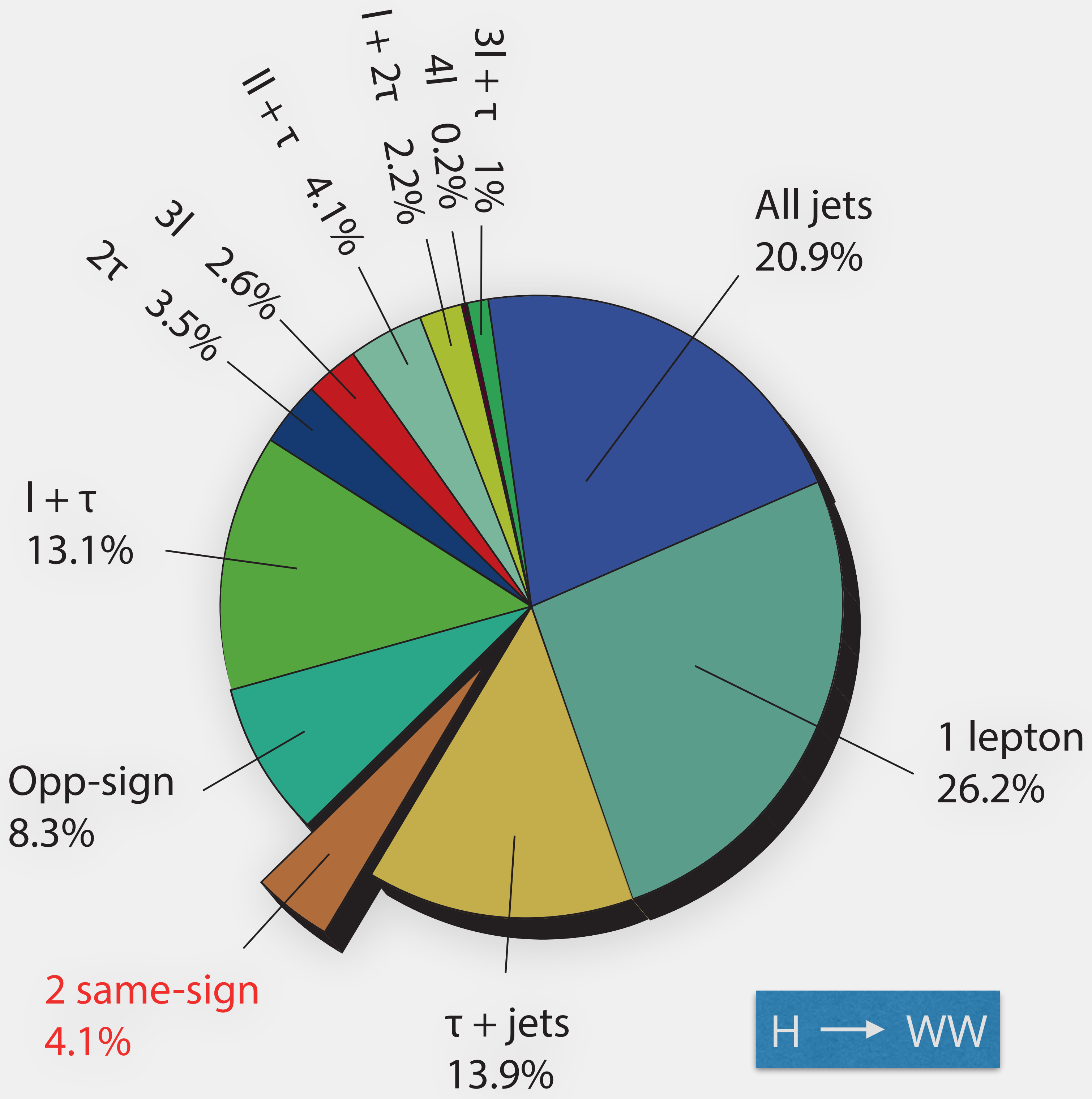
Decay



H ⟶ WW (22%), also H ⟶ ττ (6.3%) and H ⟶ ZZ (2.6%)

W ⟶ lν (21%), τ ⟶ l$\nu\nu$(35.2%), Z ⟶ ll (6.7%) for l = e, μ

×



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

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

Run 1 ATLAS results

ttH production with a multileptonic final state

Study of **five multileptonic final states** from ttH production:

Run 1 : 20 fb^{-1} data, $\sqrt{s} = 8$ TeV

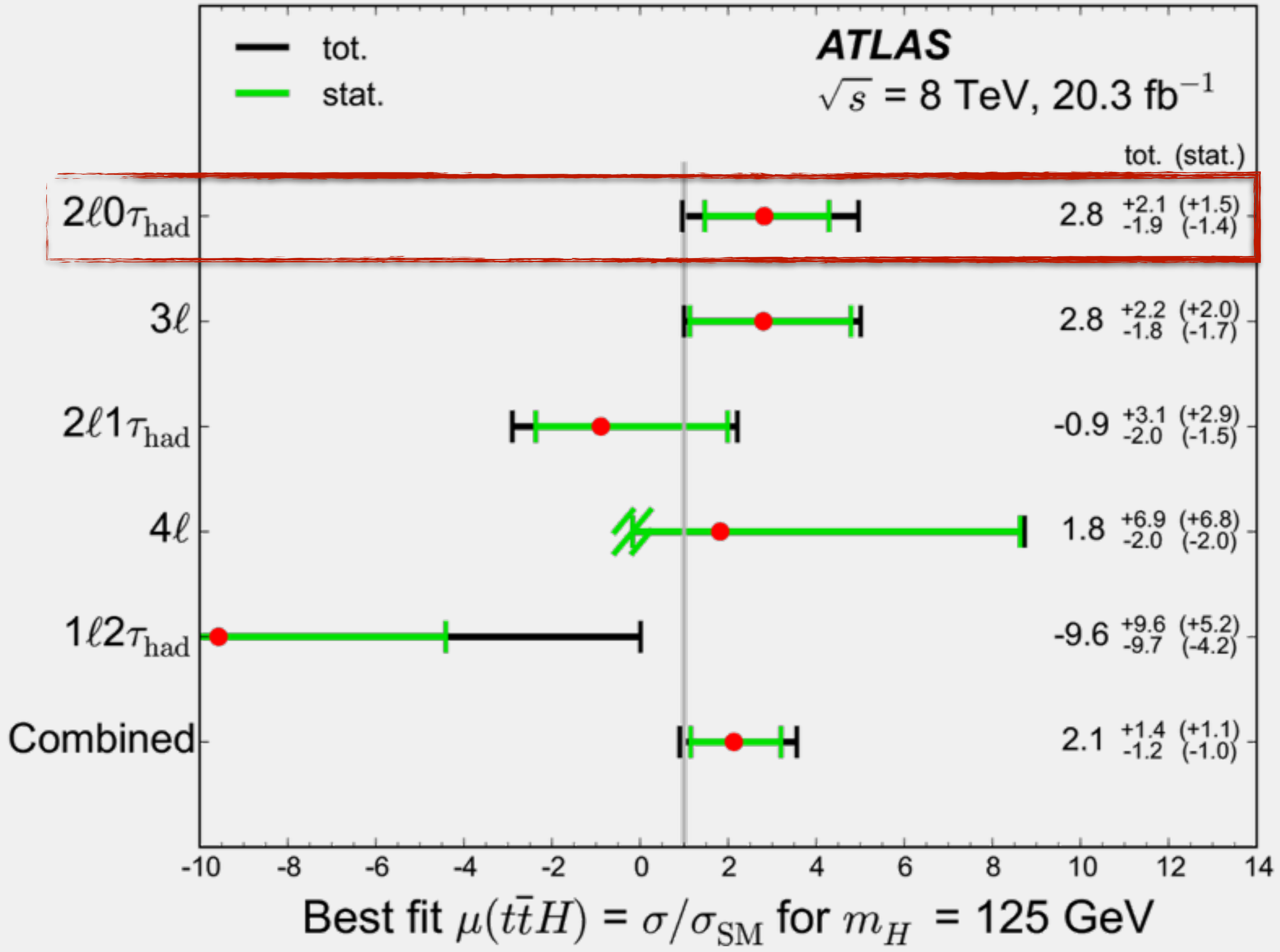
$$\begin{cases} \mu_{ttH} = \dfrac{\sigma(pp \longrightarrow ttH)}{\sigma_{SM}(pp \longrightarrow ttH)} = 2.1^{+1.4}_{-1.2} \\ \mu_{ttH} < 4.7 \text{ with 95\% CI} \end{cases}$$

Consistent with the Standard Model, but limited by statistics.

Run 2 : 100 fb^{-1} data, $\sqrt{s} = 13$ TeV

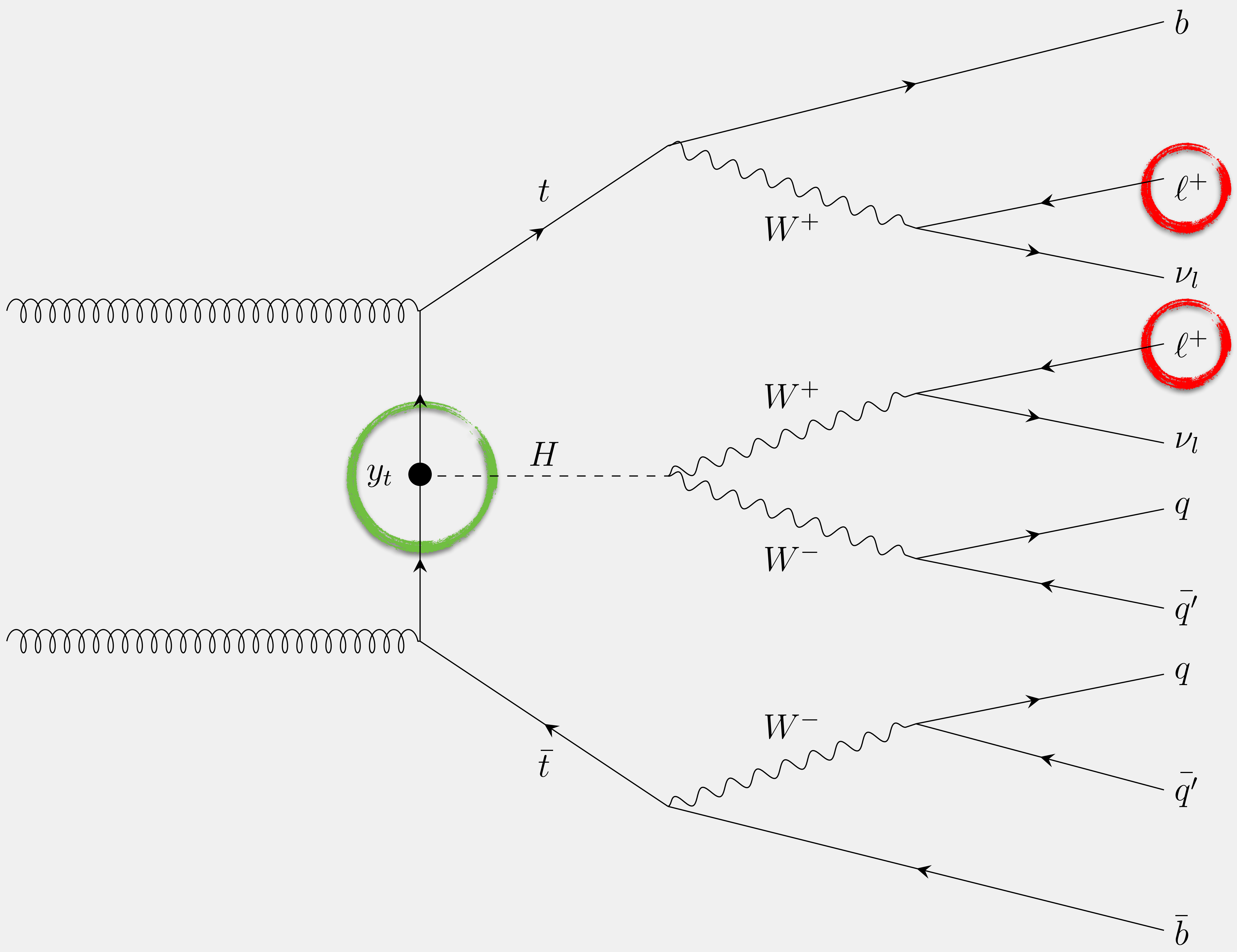
μ measurement possible **for the first time with the Run 2 of LHC.**
Great opportunity to get a **precise measurement**.

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



ttH in same-sign leptons channel

ttH signal



H ⟶ WW (22%), also H ⟶ ττ (6.3%) and H ⟶ ZZ (2.6%)

W ⟶ lν (21%), τ ⟶ l$\nu\nu$(35.2%), Z ⟶ ll (6.7%) for l = e, μ

Caracteristics

- **Three channels** : ee, eμ, μμ
- **6 jets** including **2 b-jets**
- **Missing transverse energy** due to neutrinos

A clear signature

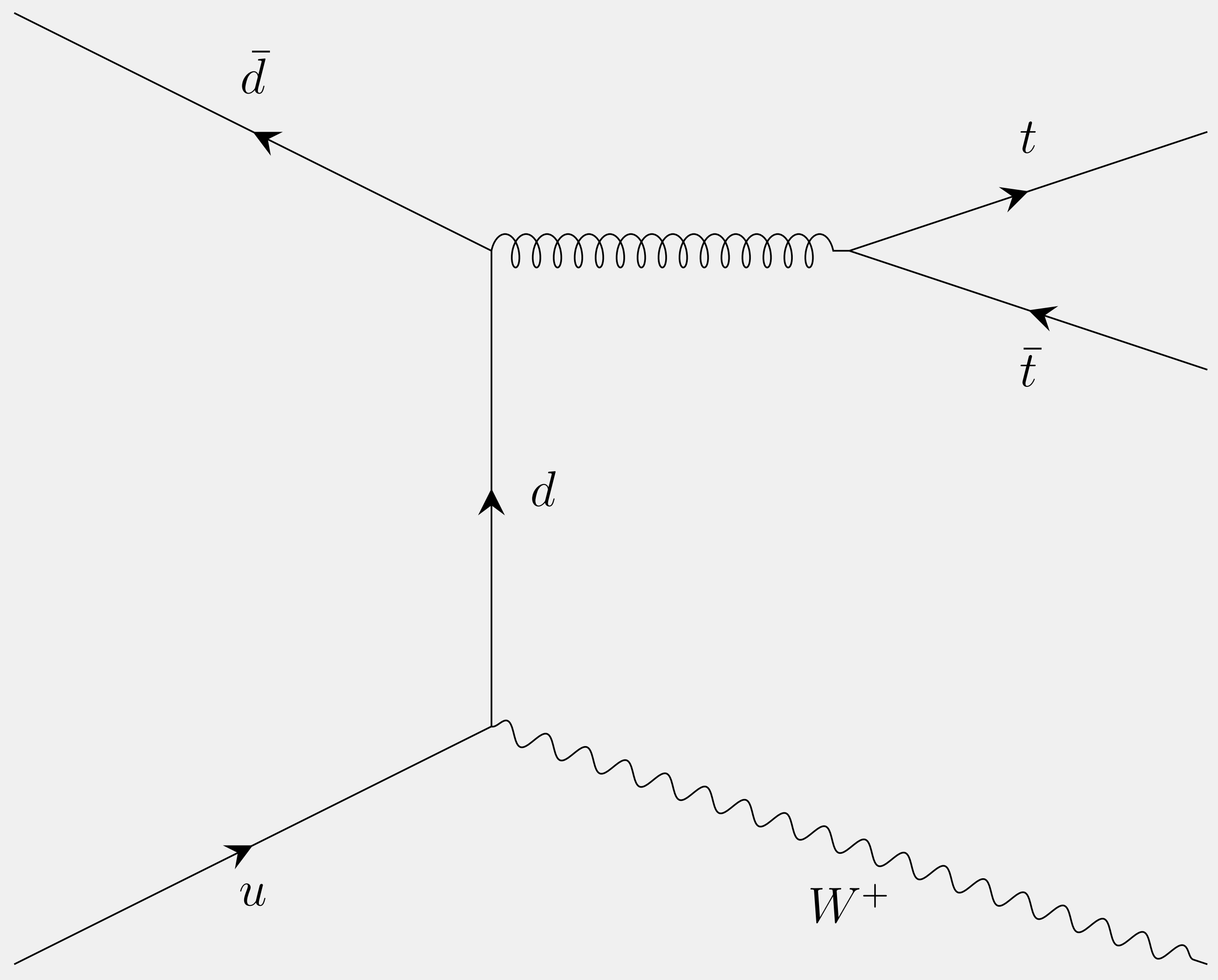
- Standard Model background with a two same-sign leptons final state **low but irreducible**
- Fake leptons background (mainly from heavy flavour decays) and leptons charge-flip **dominant but reducible**.

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

$\sigma \times 3,9$

ttH in same-sign leptons channel

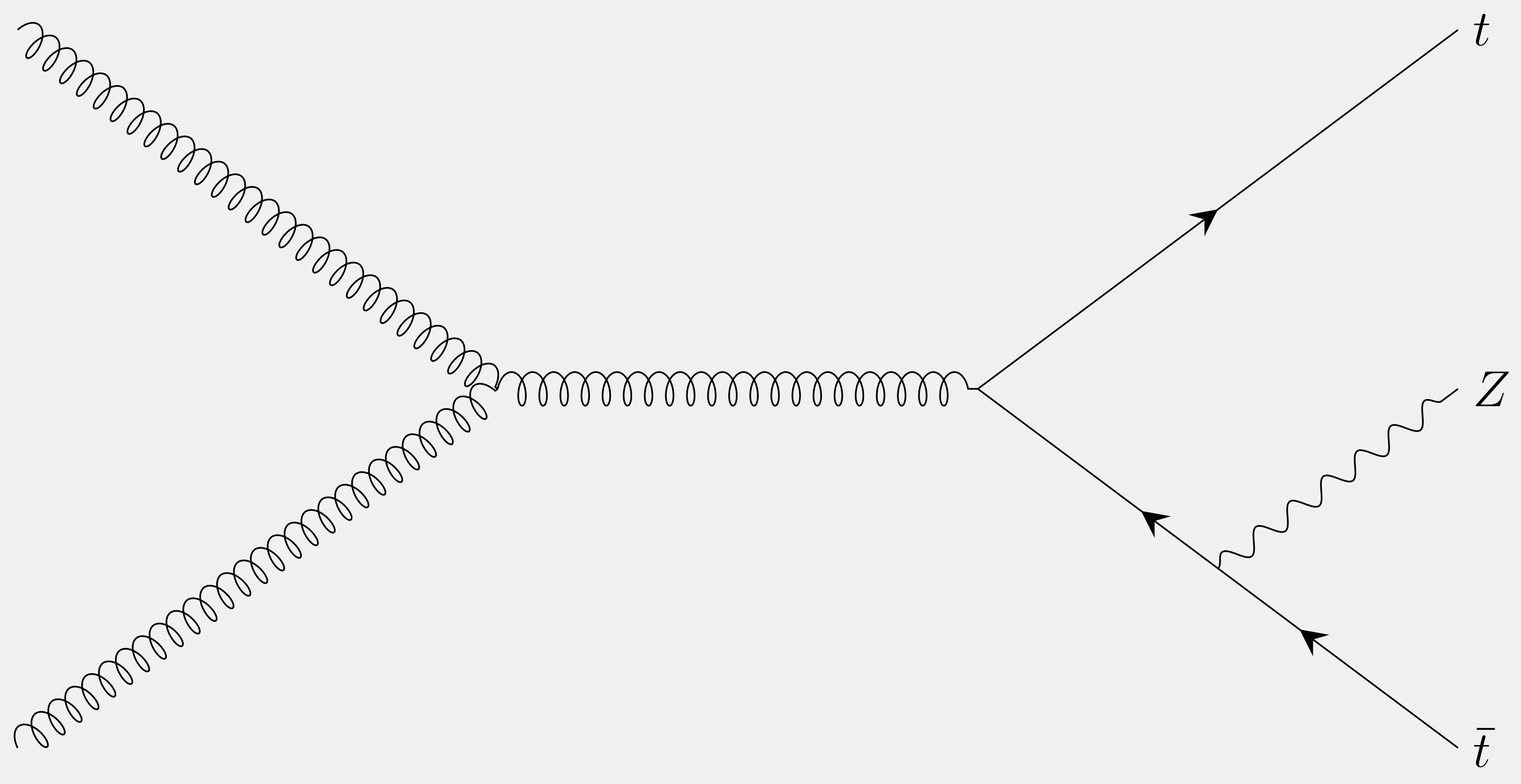
Backgrounds



ttW + 0, 1, 2, jets

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

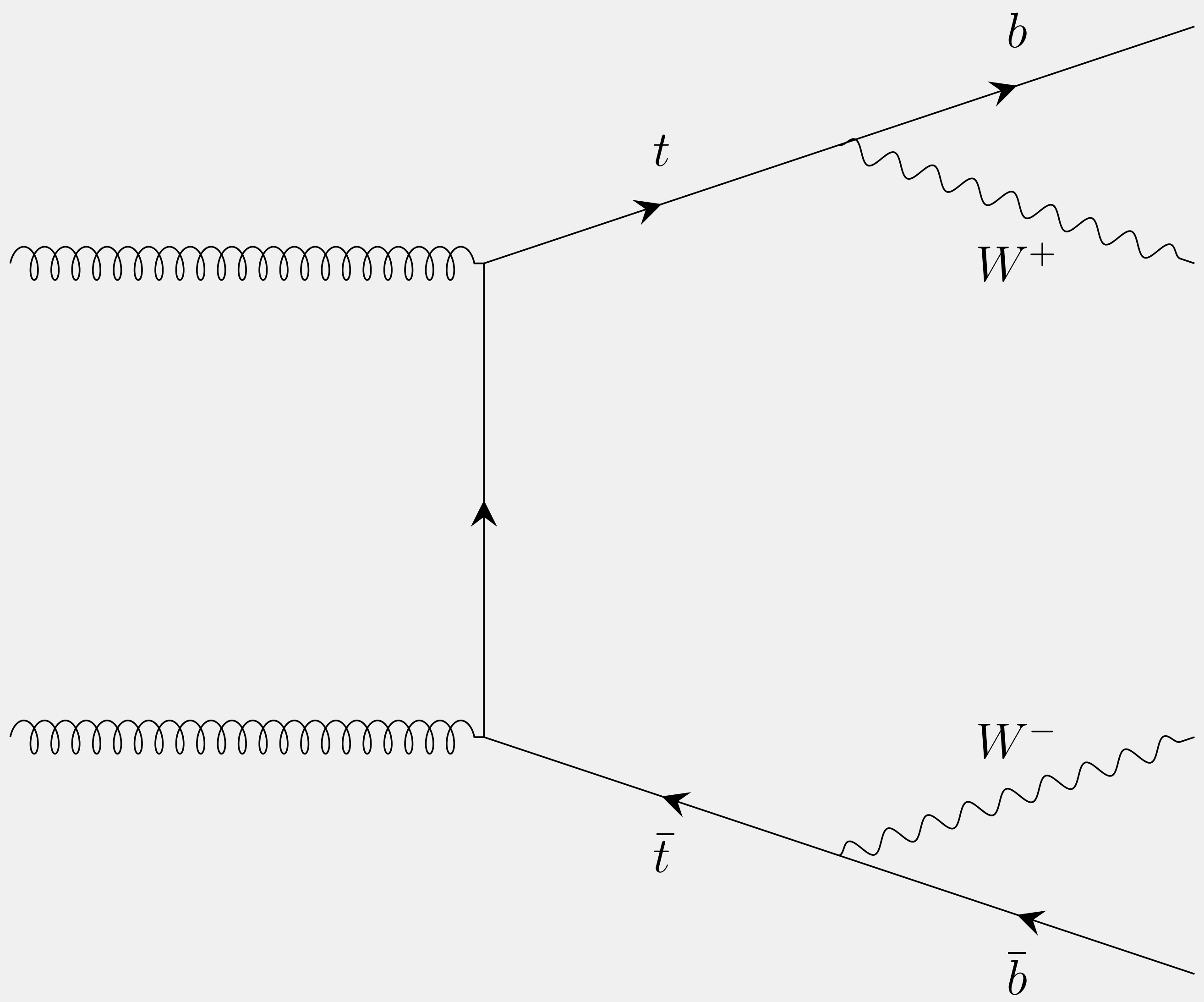
$\sigma \times 2,9$



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.**

$\sigma \times 2,8$



$t\bar{t}$

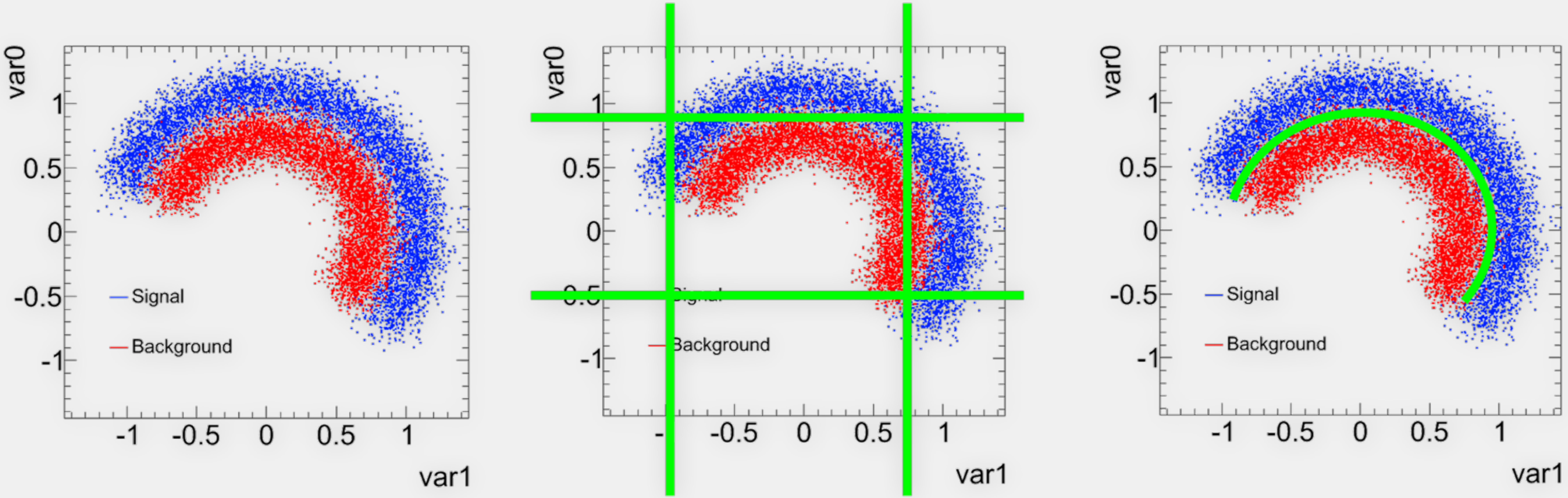
Dominating background (but reducible) with **additional jets** and **non-prompt leptons.**

$\frac{\sigma(t\bar{t})}{\sigma(t\bar{t}H)}$ ~ 2000 (1500) for $\sqrt{s}$ = 8 TeV (13 TeV)

$\sigma \times 3,3$

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).