Search for boosted $t\bar{t}H(H \rightarrow b\bar{b})$ with the ATLAS detector

Emma Winkels

Higgs Couplings 2018

UNIVERSITY OF SUSSEX

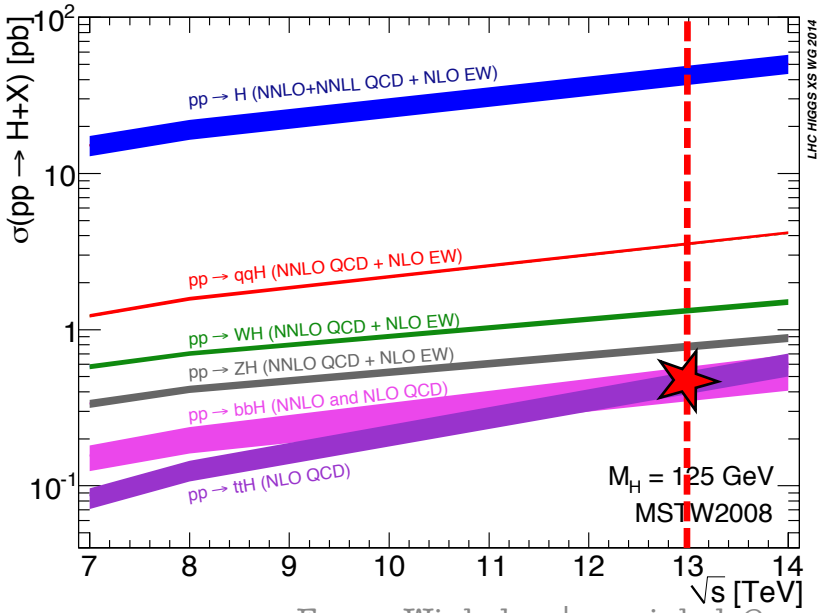
ATLAS EXPERIMENT

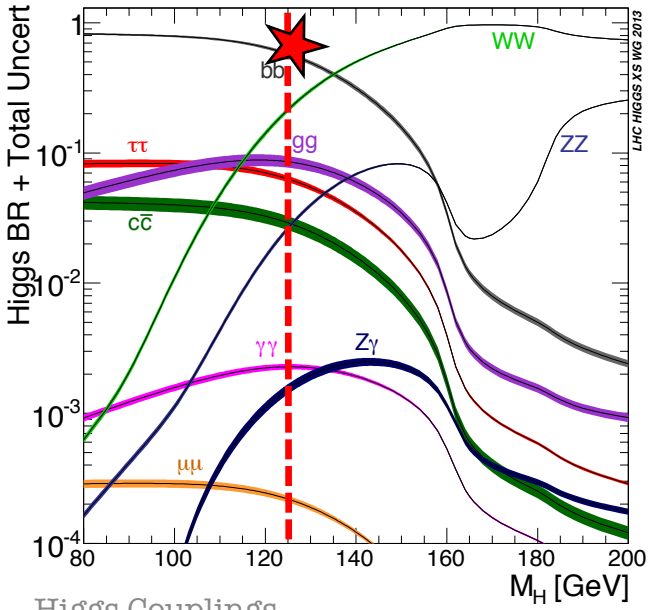
Introduction

- $t\bar{t}H$ is a rare Higgs boson production mode at the LHC
- First observation of $t\bar{t}H$ in ATLAS: Phys. Lett. B 784 (2018) 173
- Will discuss here the boosted $\mathrm{H} \rightarrow \mathrm{b\bar{b}}$ decay channel: Phys. Rev. D 97, 072016 (2018)

Source: LHC cross-section WG

$t\bar{t}H$ ~1% of total Higgs cross-section





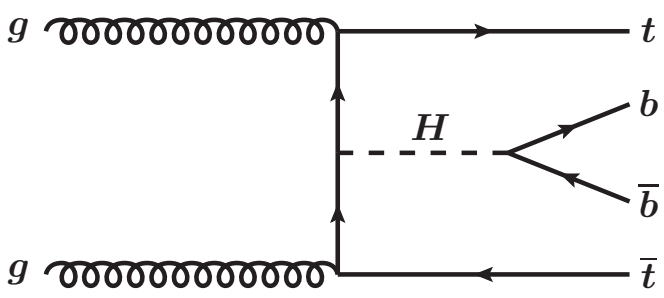
$H \rightarrow b\bar{b}$ decay BR ~58%

Motivation

All current Higgs measurements consistent with the Standard Model.

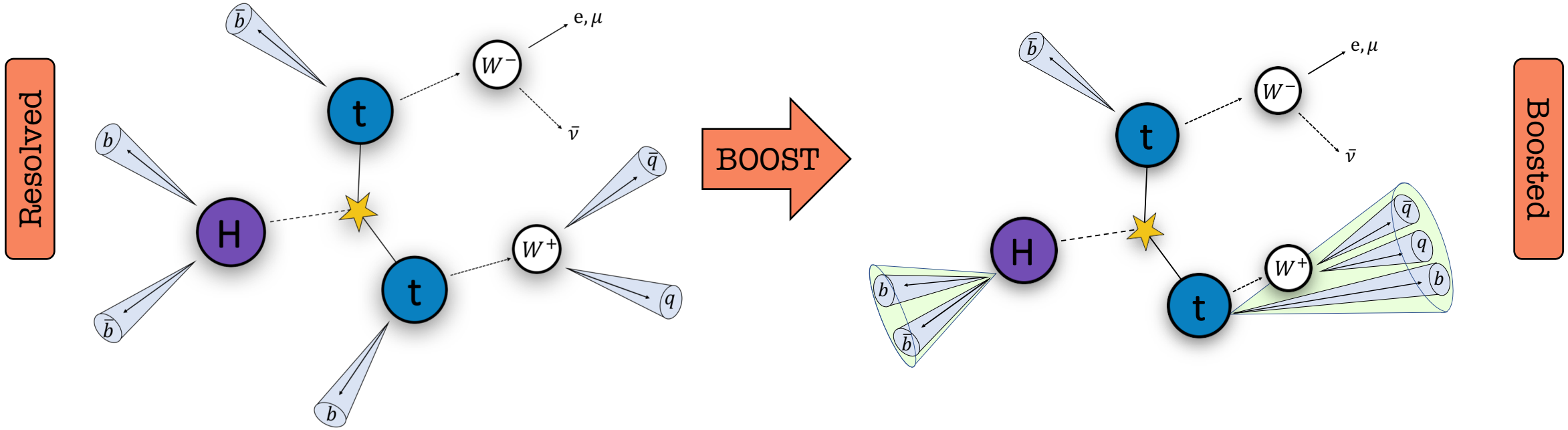
Any deviations in top Yukawa coupling would give indication for new physics.

- $t\bar{t}H$ is direct probe of the top Yukawa coupling whereas e.g. ggF only accesses it via loop effects.
- Boosted channel selects high-p_T events which leads to different kinematics and a simplified combinatorial background compared to the resolved channel.

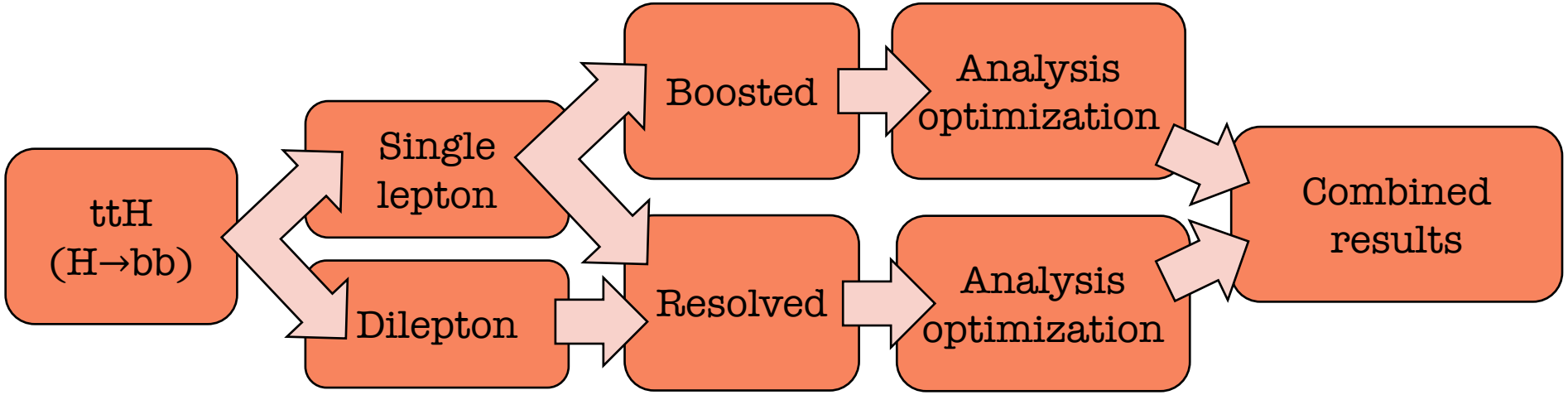


The boosted analysis

- Using $36\ fb^{-1}$ of ATLAS data from 2015-2016 at $\sqrt{s} = 13$ TeV
- Target events where both the Higgs and hadronic top are boosted



Analysis overview



The boosted region:

- Gets priority over the resolved single-lepton regions
- Is combined with the full resolved analysis in the final result

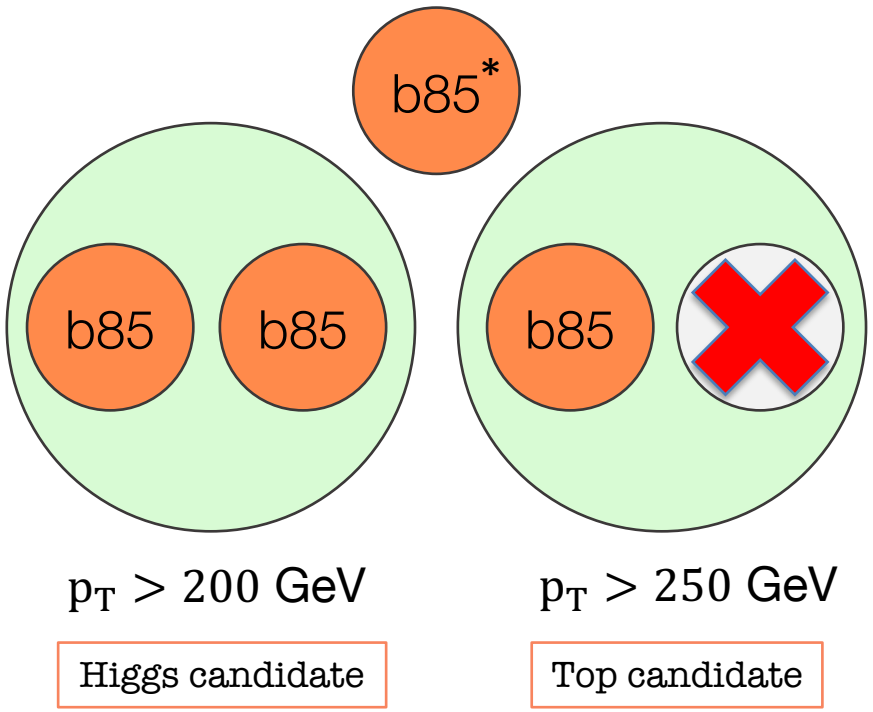
Boosted signal region

Preselection to select candidate events:

Higgs candidate: large reclustered jet with ≥ 2 *b*-tagged subjets

Top candidate: large reclustered jet with 1 *b*-tagged subjet and ≥ 1 *b*-veto

≥ 1 **additional *b*-jet** outside the top and Higgs candidates

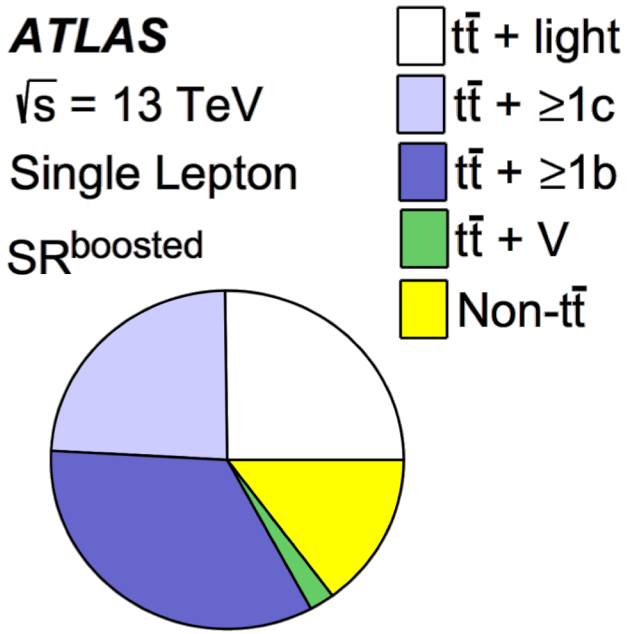


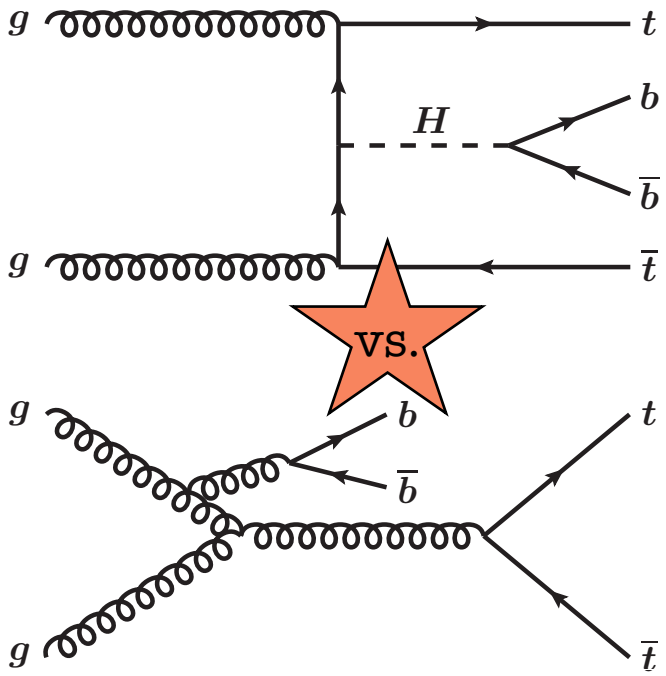
All jets are constructed with the anti-k_T algorithm.

Small jets have R=0.4 and large jets R=1.0.

$*$ The subjets are *b*-tagged at the loose (85%) working point.

Boosted signal region composition





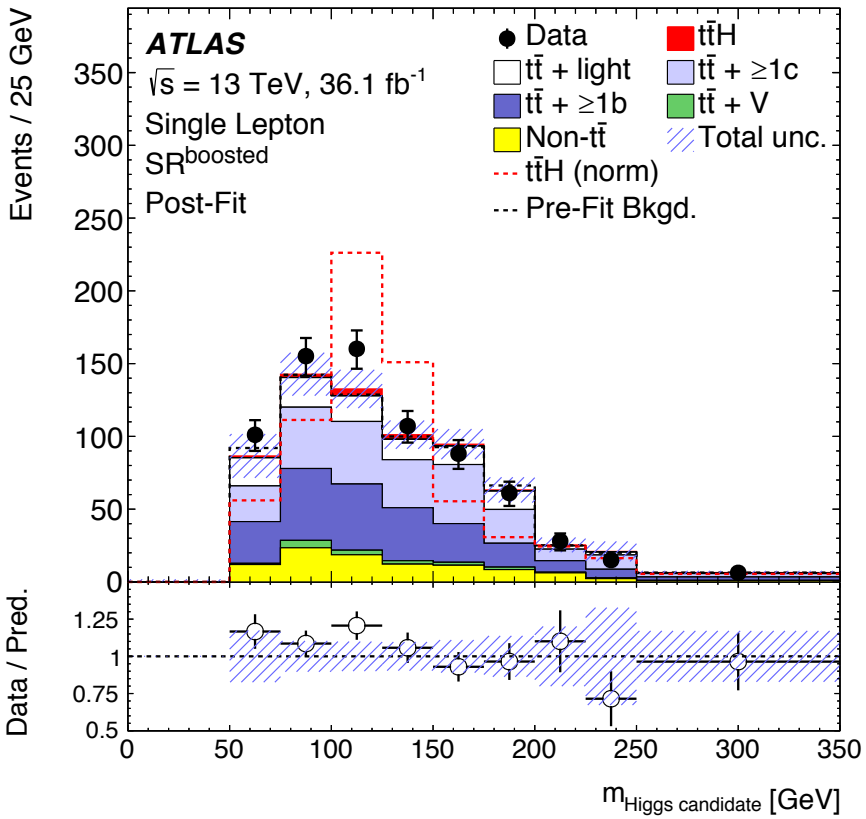
Low S/B ~ 2.4% and background dominated by $t\bar{t} + \geq 1b$ which is very difficult to distinguish from $t\bar{t}H$.

Use multivariate analysis to discriminate between signal and background events.

Multivariate analysis

- Use a Boosted Decision Tree (BDT) to separate signal from background events.
- Eight variables are chosen as input:
  - The Higgs candidate mass
  - The first splitting scale of the top candidate (jet substructure variable)
  - 4 angular variables
  - 2 *b*-tagging variables where all jets are assigned a score depending on their likelihood to be a *b*-jet

	None	Loose	Medium	Tight	Very tight
b-tagging efficiency	-	85%	77%	70%	60%
Score	1	2	3	4	5

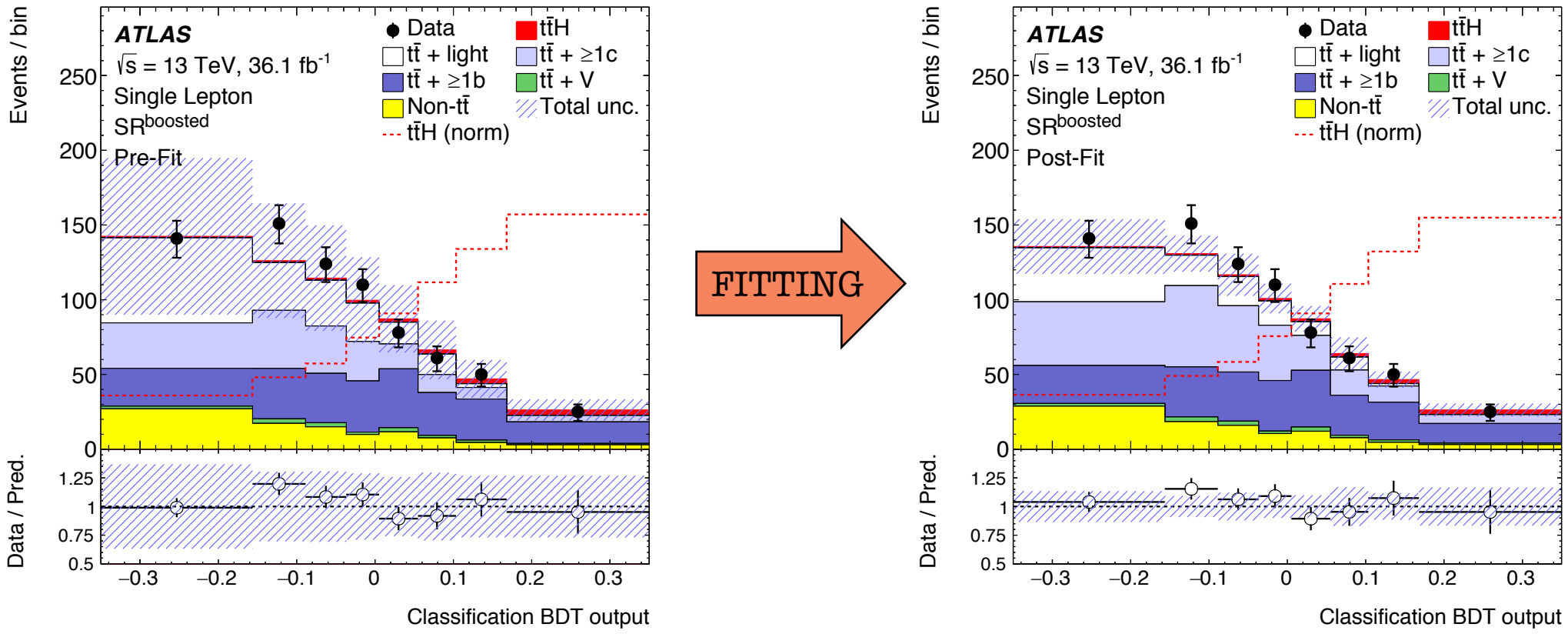


Statistical analysis

- Perform a binned profile likelihood fit simultaneously across all regions to test for presence of signal.
- The boosted region is fitted together with 8 resolved signal regions and 10 resolved control regions.
- In the signal regions, we fit to the BDT.
- Systematics included as nuisance parameters in the fit.
- Extract $\mu_{t\bar{t}H} = \sigma_{obs}^{t\bar{t}H} / \sigma_{SM}^{t\bar{t}H}$.

The control regions are used to constrain systematic uncertainties on backgrounds.

Boosted BDT



The uncertainty is reduced post-fit due to profiling of the nuisance parameters.
More information see p.23 of arXiv 1712.08895.

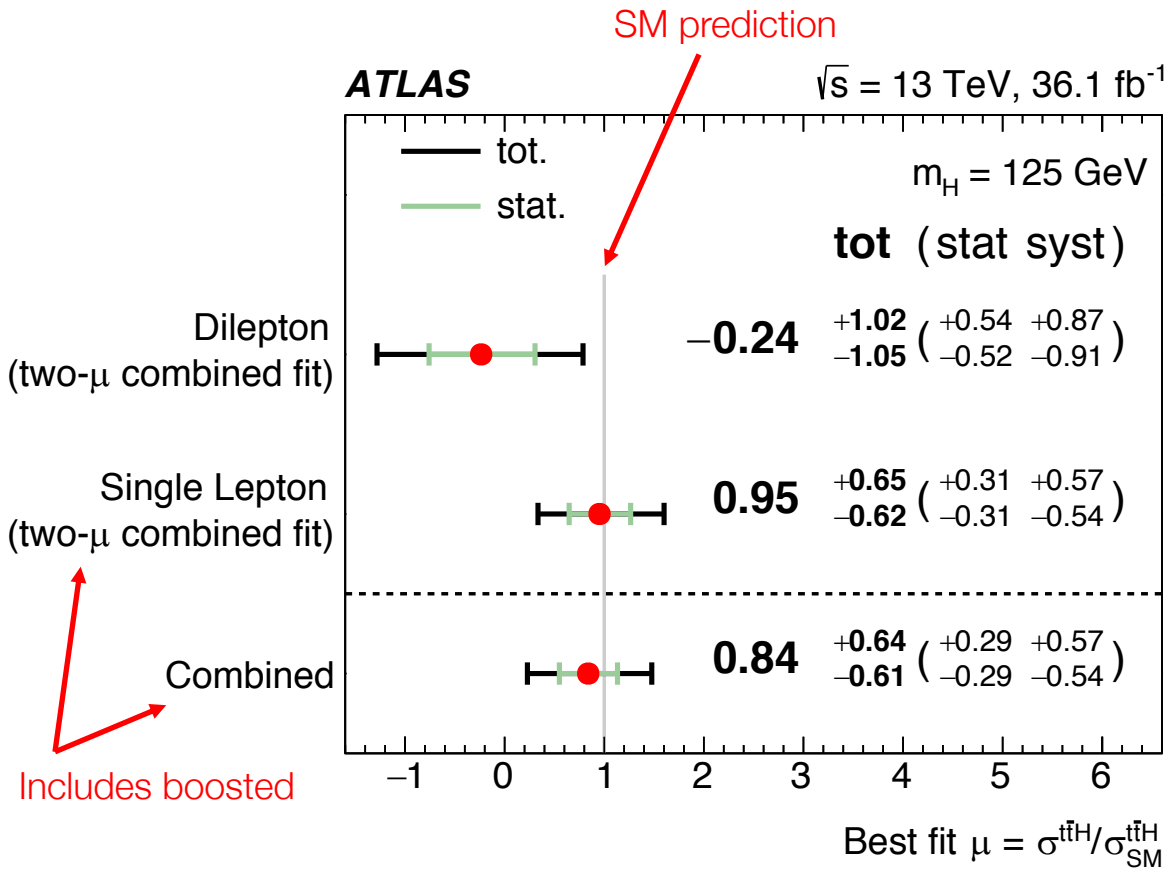
Uncertainties

The analysis is limited by systematics. Main source is modelling of $t\bar{t} + \geq 1b$.

We compare different Monte Carlo generators to assess uncertainties on matrix element, showering, and radiation variations.

Uncertainty source	$\Delta\mu$	
$t\bar{t} + \geq 1b$ modeling	+0.46	−0.46
Background-model stat. unc.	+0.29	−0.31
b-tagging efficiency and mis-tag rates	+0.16	−0.16
Jet energy scale and resolution	+0.14	−0.14
$t\bar{t}H$ modeling	+0.22	−0.05
$t\bar{t} + \geq 1c$ modeling	+0.09	−0.11
JVT, pileup modeling	+0.03	−0.05
Other background modeling	+0.08	−0.08
$t\bar{t}$ + light modeling	+0.06	−0.03
Luminosity	+0.03	−0.02
Light lepton (e, μ) id., isolation, trigger	+0.03	−0.04
Total systematic uncertainty	+0.57	−0.54
$t\bar{t} + \geq 1b$ normalization	+0.09	−0.10
$t\bar{t} + \geq 1c$ normalization	+0.02	−0.03
Intrinsic statistical uncertainty	+0.21	−0.20
Total statistical uncertainty	+0.29	−0.29
Total uncertainty	+0.64	−0.61

Results resolved+boosted



Best-fit value:

$\mu_{t\bar{t}H}$ = 0.84 ± 0.29(stat) $^{+0.57}_{-0.54}$ (syst)

- Excluded $\mu_{t\bar{t}H} > 2.0$ at 95% CL
- Signal seen with obs (exp) significance of $1.4\sigma(1.6\sigma)$

Currently the boosted signal region does not add significant sensitivity to the channel.

- More data → tighter SR selection
- Many ongoing studies for improvements

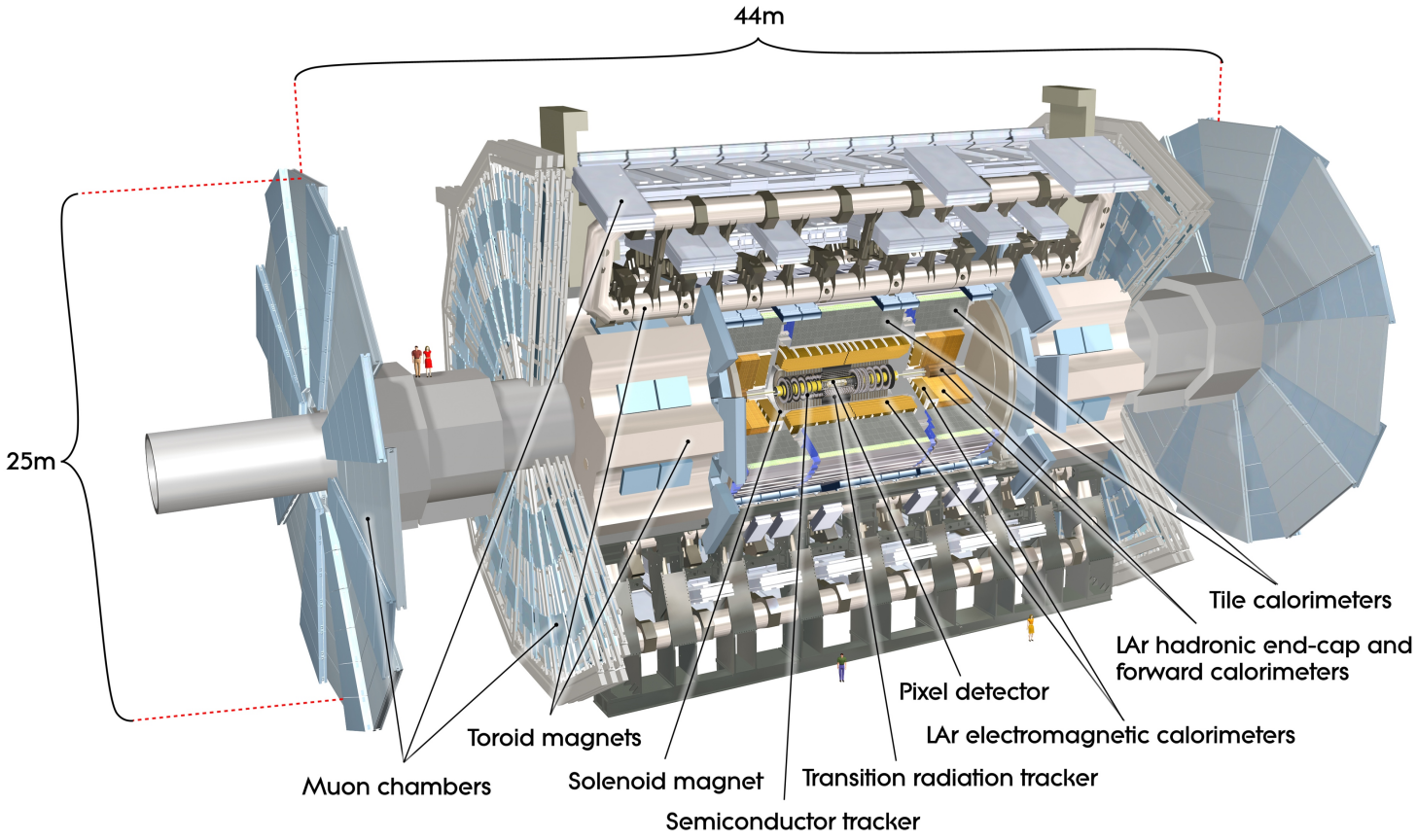
Conclusions

- Very challenging analysis due to dominant $t\bar{t} + \geq 1b$ background.
- Measured $\mu_{t\bar{t}H}$ consistent with the SM and background-only hypotheses.
- The boosted channel does not currently add significant sensitivity to the resolved analysis.
  - First time it is included in ATLAS; many studies on-going to improve the performance for full Run-2 analysis.
  - More data and better MC statistics will allow for tighter signal region.
- The analysis is systematics limited → need to rethink our approach to dealing with these uncertainties.

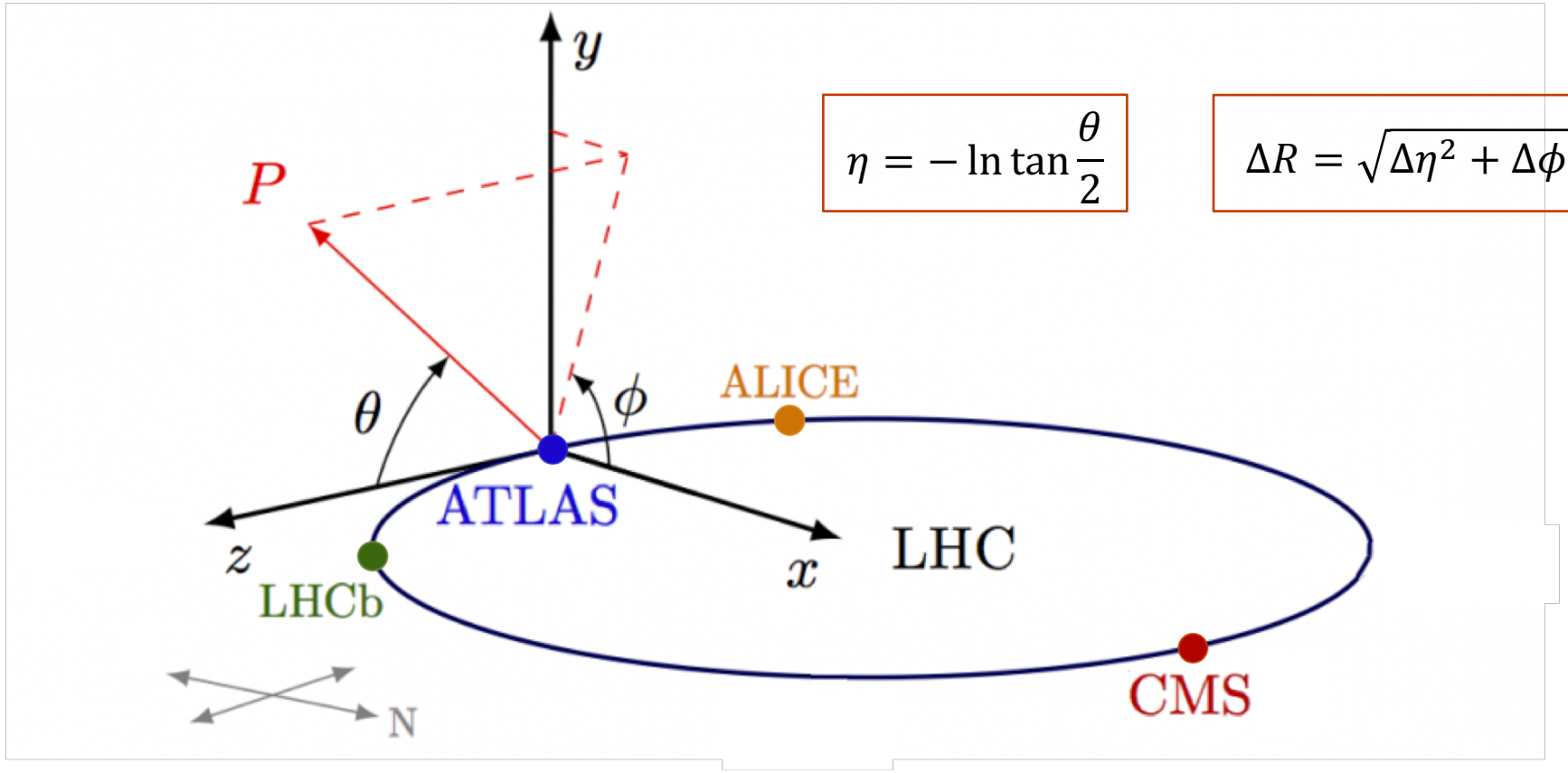
BACK-UP

The ATLAS detector

The ATLAS collaboration
>5000 members
38 countries
182 institutes



The ATLAS coordinate system



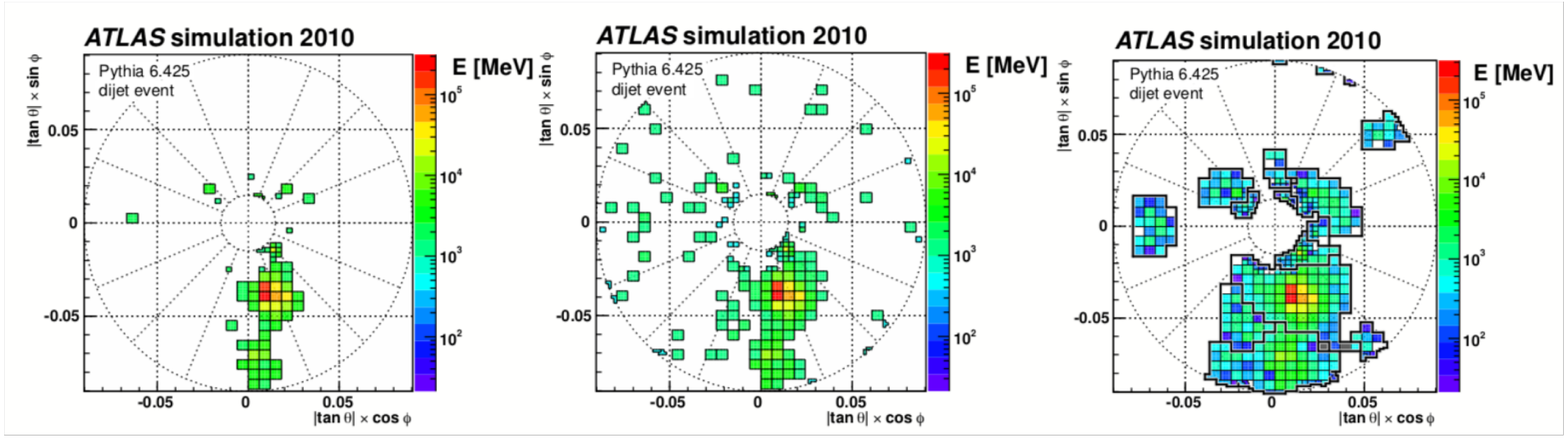
$$\eta = -\ln\tan\frac{\theta}{2}$$

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$$

Granularity of ATLAS calorimeter

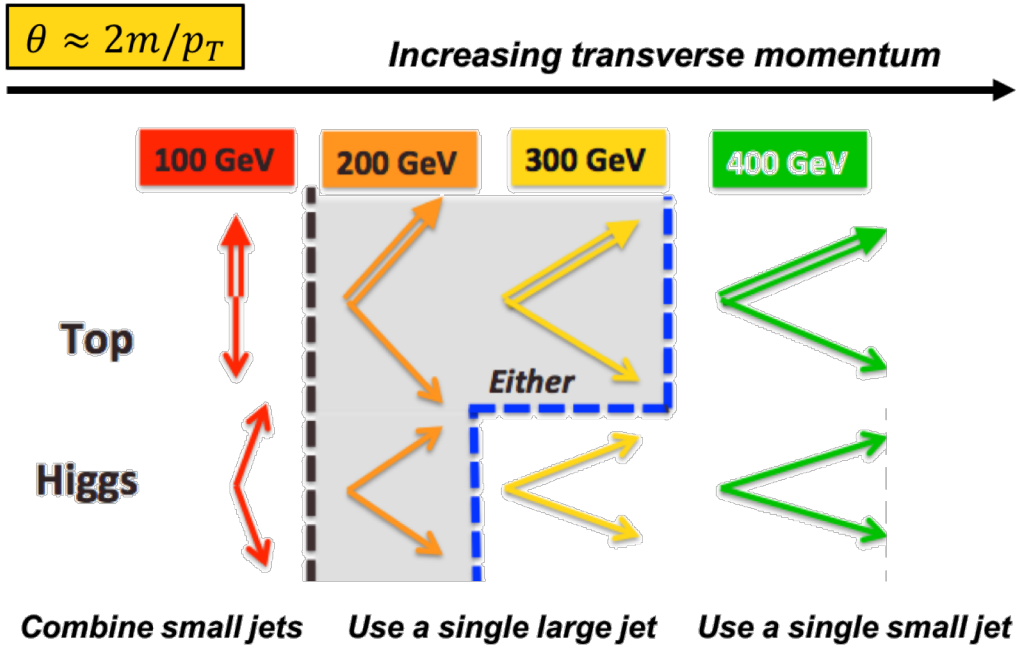
The hadronic calorimeter is coarser than the EM calorimeter. If we want to use all calorimeter layers we are limited by coarsest layer (R_{min}~0.2). But we can also ignore coarse layers and use only very fine layers.

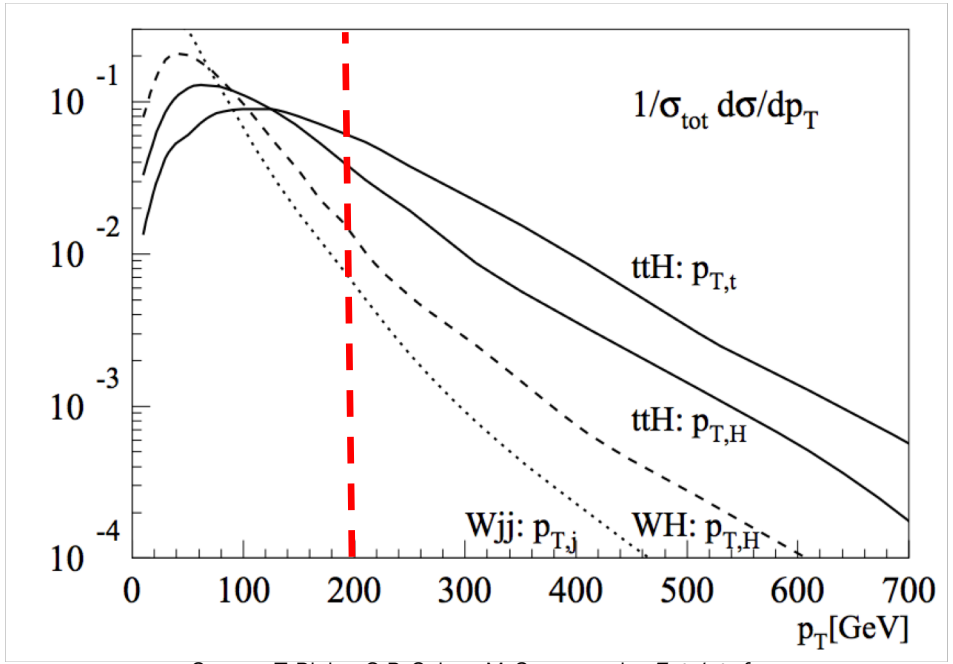
- EM granularity: $\Delta\eta \times \Delta\phi \approx 0.025 \times \pi/128$
- Hadronic granularity: $\Delta\eta \times \Delta\phi \approx 0.1 \times \pi/32$



arXiv 1603.02934

Boosted $t\bar{t}H$





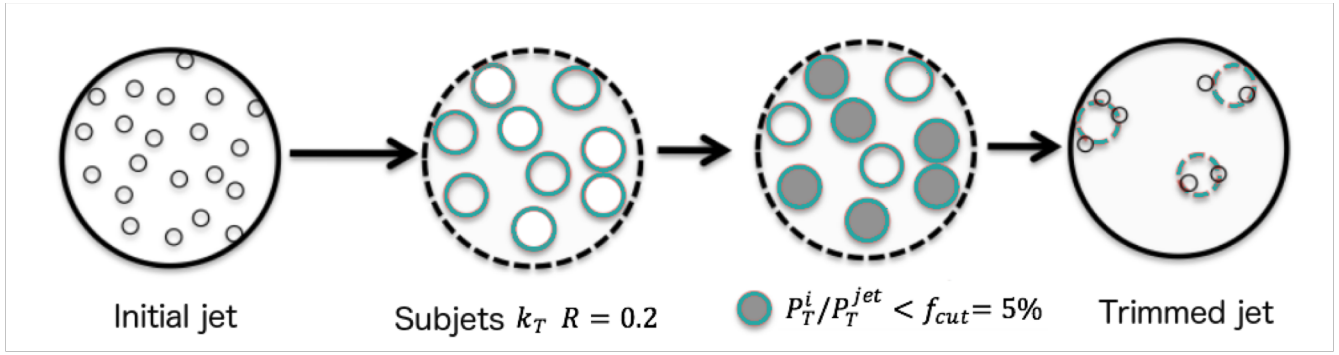
Source: T. Plehn, G.P. Salam, M. Spannowsky, *Fat Jets for a Light Higgs*, arXiv:0910.5472

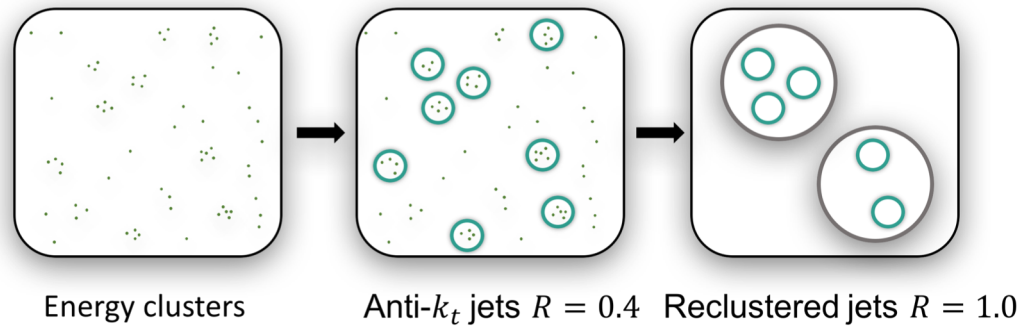
Jets *b*-tagging

- The analysis relies heavily on *b*-tagging: identifying jets containing *b*-hadrons (*b*-jets)
- Using a multivariate *b*-tagging algorithm optimised to select *b*-jets and separate them from *c*-jets, τ-jets, and other jets (light jets)
- MV2c10 uses number of variables to exploit the long lifetime of the b-hadron
- Four working points available from this algorithm

Working point	Efficiency for b-jets	Rejection factor of c-jets	Rejection factor of light-jets
Loose	85%	3.1	33
Medium	77%	6	134
Tight	70%	12	381
Very tight	60%	34	1538

Large jets





Trimming:

- Build large jets from energy clusters in calorimeter
- Extra set of systematics

Reclustering:

- Build large jets from fully calibrated small jets
- No extra systematics

Jet substructure variables

- Jet mass is calculated from the jet constituents (calorimeter energy deposits):

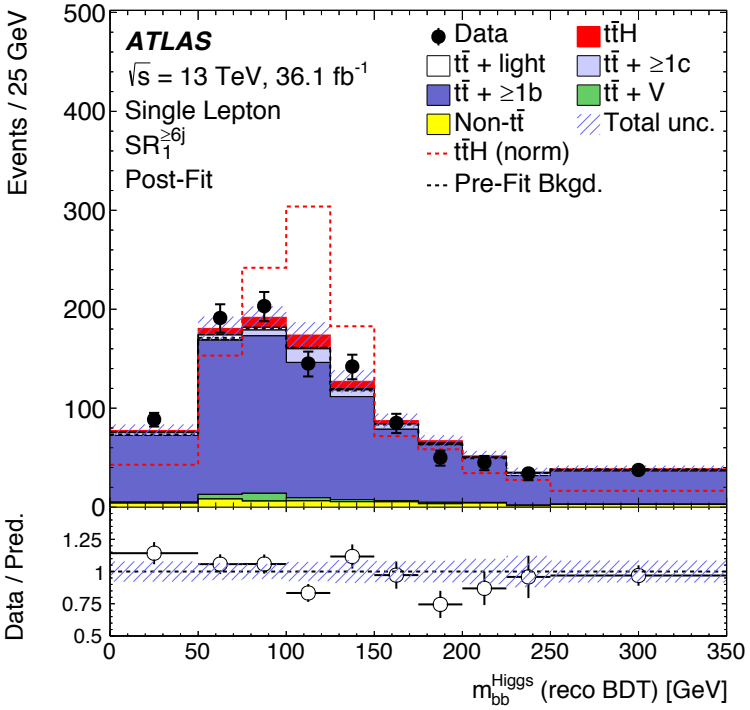
$$m^{calo} = \sqrt{\left(\sum_i E_i\right)^2 - \left(\sum_i \vec{p}_i\right)^2}$$

- First k_T splitting scale: $\sqrt{d_{12}} = \min(p_{T,i}, p_{T,j}) \times \frac{\Delta R_{i,j}}{R}$ where *i, j* are the two proto-jets combined at the final step of the k_T jet algorithm.

Higgs mass reconstruction

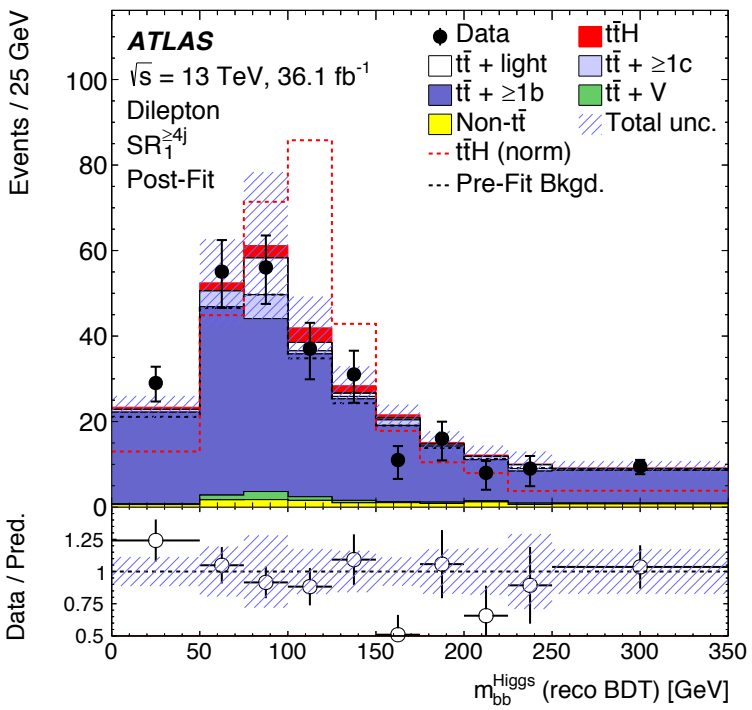
arXiv 1712.08895

Single-lepton



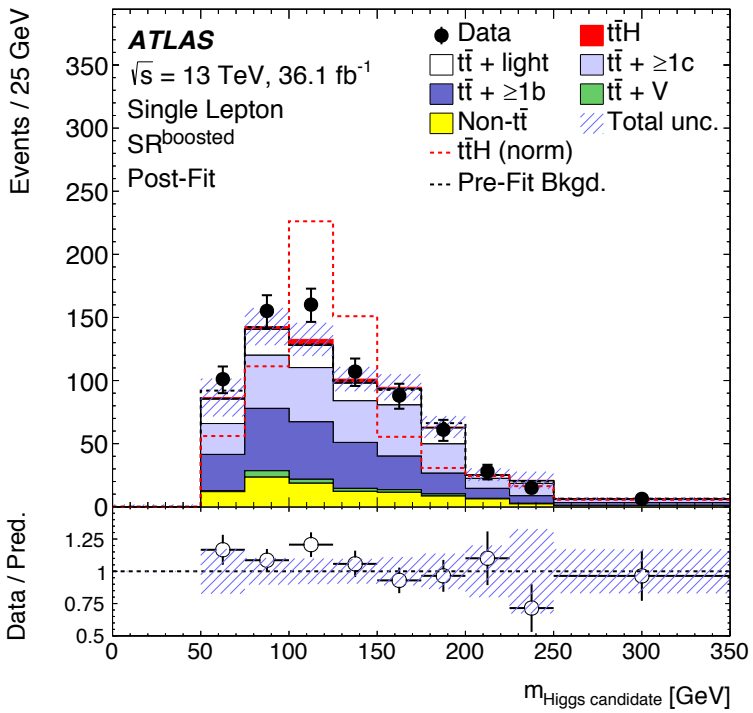
Correct Higgs candidate found in 48% of events

Dilepton



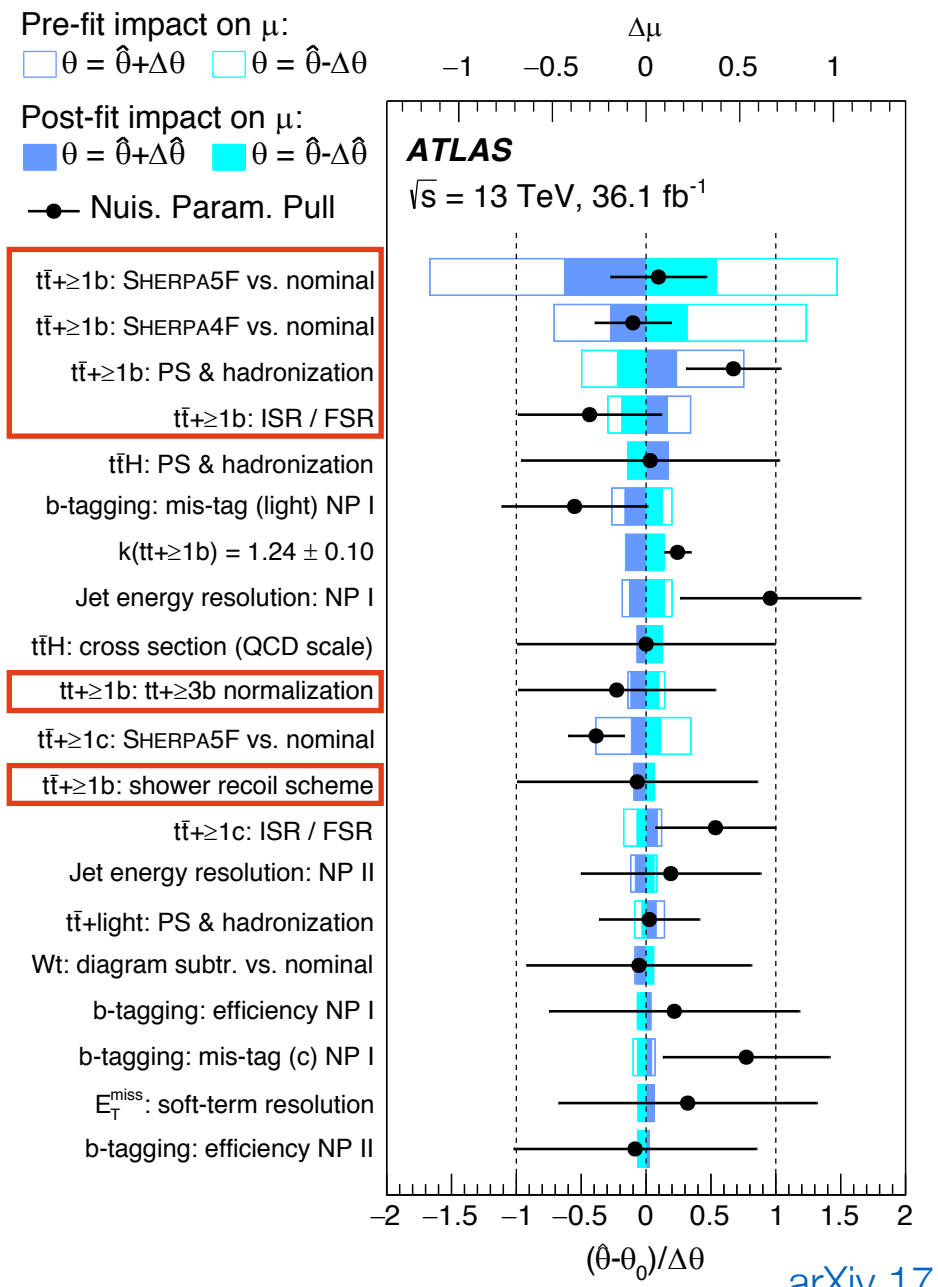
Correct Higgs candidate found in 49% of events

Boosted



Correct Higgs candidate found in 47% of events

Systematics ranking

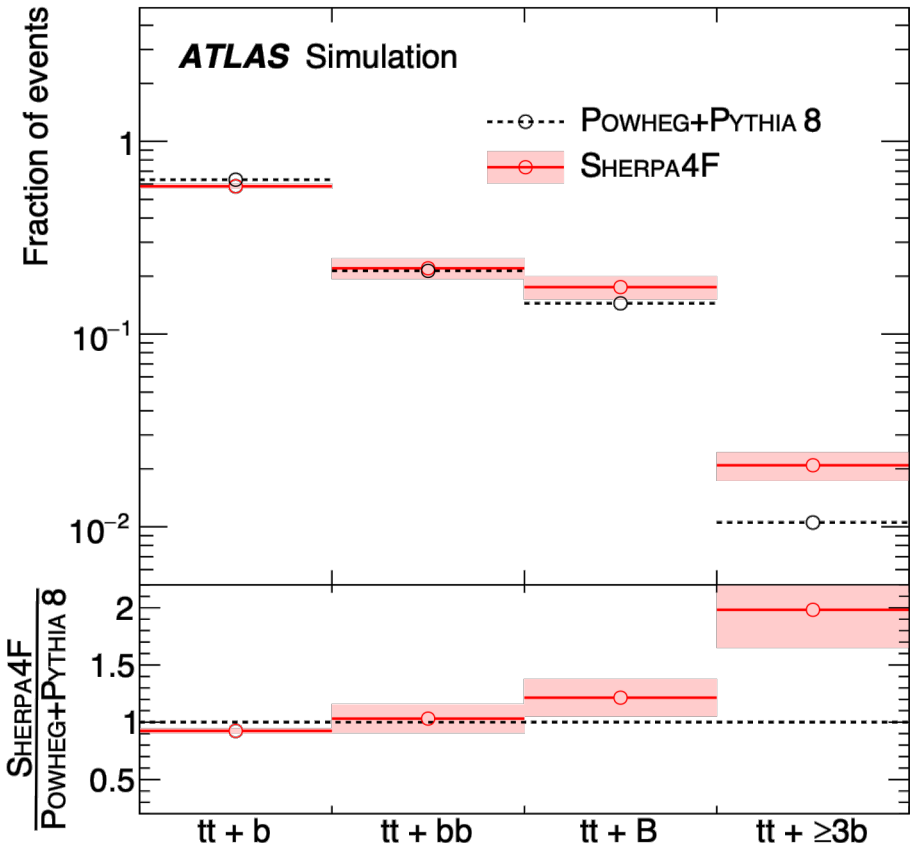


arXiv 1712.08895

$t\bar{t}$ + jets modelling

- Treatment of $t\bar{t}$ +jets is crucial in the analysis
- Nominal $t\bar{t}$ MC sample is Powheg+Pythia8, normalised to NNLO+NNLL cross section calculation
- Separate into $t\bar{t}$ + light, $t\bar{t}$ +≥ 1c and $t\bar{t}$ +≥ 1b
- Dividing $t\bar{t}$ +≥ 1b into 4 more categories, the fractions of which are reweighted to Sherpa sample which has NLO calculation and takes into account b-quark mass.
- Large number of systematics covering $t\bar{t}$ modelling

arXiv 1712.08895



$t\bar{t}$ + jets modelling systematics

All sources of systematic uncertainty for $t\bar{t}$ +jets modelling.

Systematic source	Description	$t\bar{t}$ categories
$t\bar{t}$ cross-section	Up or down by 6%	All, correlated
$k(t\bar{t} + \geq 1c)$	Free-floating $t\bar{t} + \geq 1c$ normalization	$t\bar{t} + \geq 1c$
$k(t\bar{t} + \geq 1b)$	Free-floating $t\bar{t} + \geq 1b$ normalization	$t\bar{t} + \geq 1b$
Sherpa5F vs. nominal	Related to the choice of NLO event generator	All, uncorrelated
PS & hadronization	Powheg+Herwig 7 vs. Powheg+Pythia 8	All, uncorrelated
ISR / FSR	Variations of μ_R, μ_F, h_{damp} and A14 Var3c parameters	All, uncorrelated
$t\bar{t} + \geq 1c$ ME vs. inclusive	MG5_aMC@NLO+Herwig++: ME prediction (3F) vs. incl. (5F)	$t\bar{t} + \geq 1c$
$t\bar{t} + \geq 1b$ Sherpa4F vs. nominal	Comparison of $t\bar{t} + b\bar{b}$ NLO (4F) vs. Powheg+Pythia 8 (5F)	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1b$ renorm. scale	Up or down by a factor of two	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1b$ resumm. scale	Vary μ_Q from $H_T/2$ to μ_{CMMPS}	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1b$ global scales	Set μ_Q, μ_R, and μ_F to μ_{CMMPS}	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1b$ shower recoil scheme	Alternative model scheme	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1b$ PDF (MSTW)	MSTW vs. CT10	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1b$ PDF (NNPDF)	NNPDF vs. CT10	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1b$ UE	Alternative set of tuned parameters for the underlying event	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1b$ MPI	Up or down by 50%	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 3b$ normalization	Up or down by 50%	$t\bar{t} + \geq 1b$

Non-$t\bar{t}$ modelling

- Nominal $t\bar{t}H$ signal model is MG5_aMC@NLO+Pyhtia8 (NLO)
- $t\bar{t}W, t\bar{t}$Z: MG5_aMC@NLO+Pythia8 (NLO)
- Single-top (Wt, s-channel, t-channel): Powheg+Pythia6 (NLO)
- w+jets, z+jets: Sherpa 2.2.1 (NLO <3 partons, LO <5 partons)
- $tHqb$: MG5_aMC@NLO+Pythia8 (LO)
- tWH: MG5_aMC@NLO+Herwig++ (NLO)
- $t\bar{t}t\bar{t}$, $t\bar{t}WW$: MG5_aMC@NLO+Pythia8 (LO)
- tZ: MG5_aMC@NLO+Pythia6 (LO)
- tZW: MG5_aMC@NLO+Pythia8 (NLO)