B-tagging Calibration and Observation of Higgs Boson Decays to a pair of bottom quarks with the ATLAS Detector

李昌樵 (TDLI / SJTU)

Aug 16th, 2021

第十一届晨光杯青年优秀论文评选报告

About Me

- Education:
  - 2008.9 - 2012.6 Bachelor in Physics, USTC[1], Hefei (China).
  - 2012.6 - 2018.11 Co-tutorship Ph.D. in Particle Physics,
    USTC and LPNHE[2] / SU[3], Paris (France).
    Supported by the CSC[2] from 2016 to 2018.

    Thesis is supervised by Prof. LIU Yanwen and Dr. MARCHIORI Giovanni

- Current and Previous Positions:
  As a commitment to CSC, I started the domestic service since I graduated.
  - 2019.1 - 2020.5 Postdoctoral Researcher, USTC, Hefei (China).
  - 2020.6 - present Postdoctoral Researcher, SJTU-TDLI, Shanghai (China).

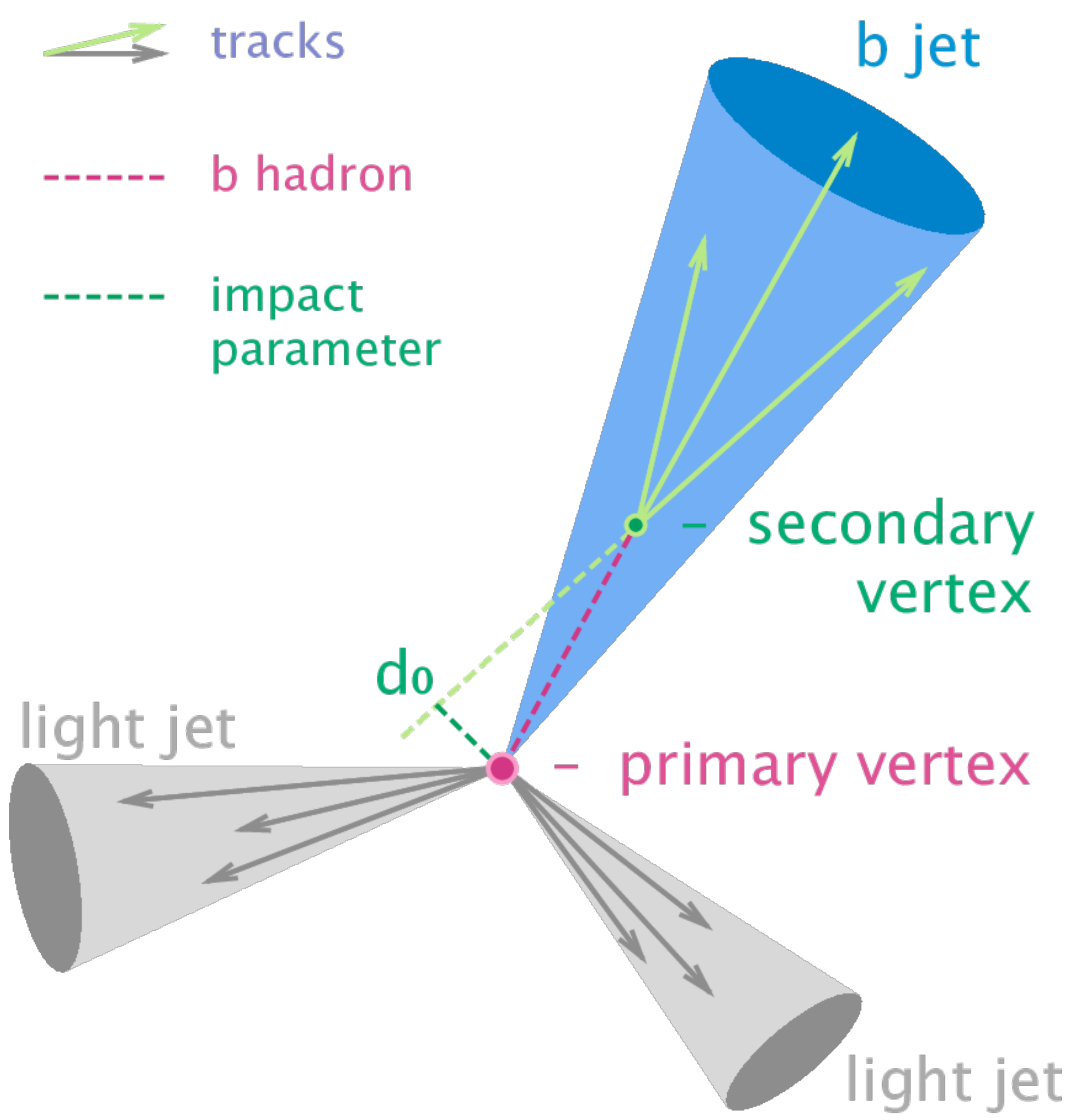
1.University of Science and Technology of China
2.Laboratoire de Physique Nucléaire et de Hautes Energies
3.Sorbonne Universités
4.China Scholarship Council

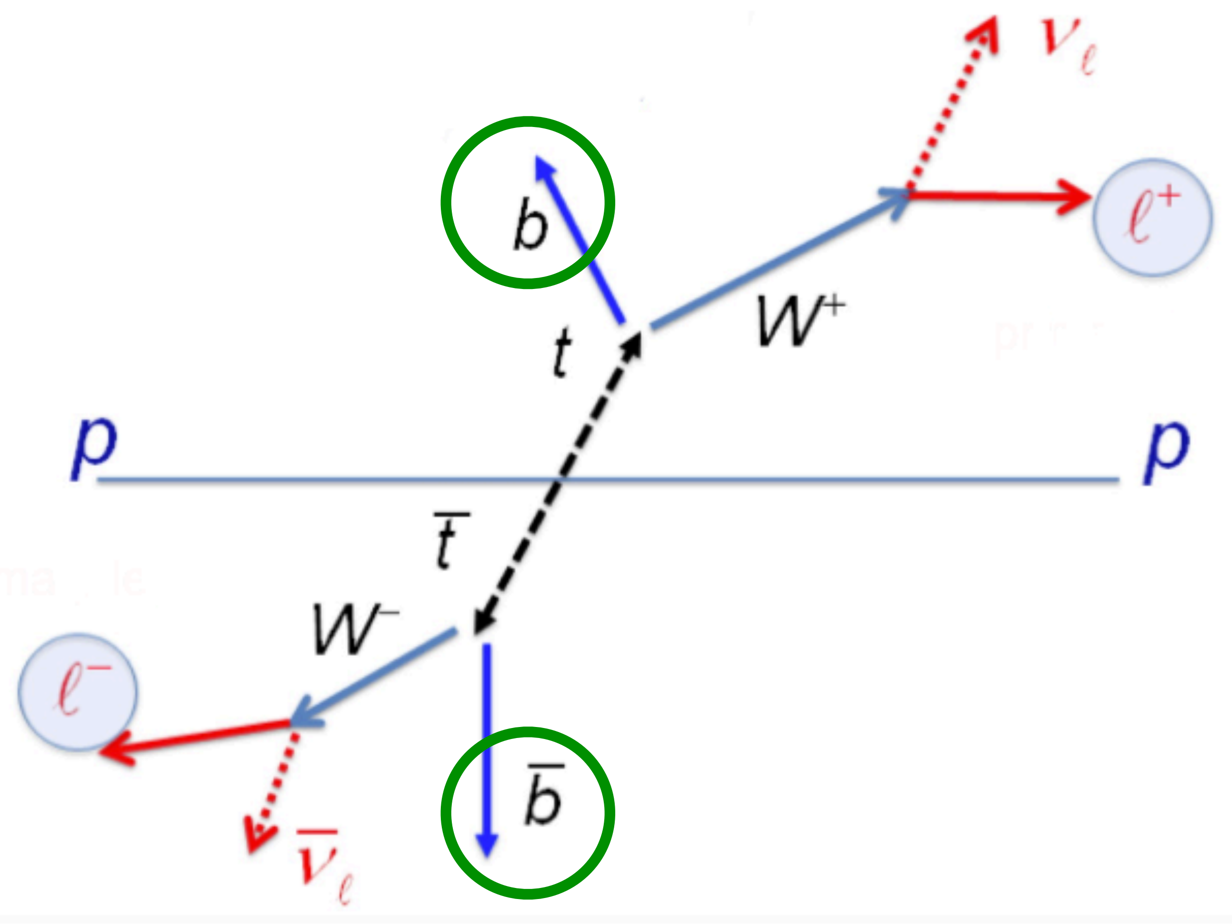
During my Ph.D.

- Involved into two related topics:
  - Identification of the b-jets and Search for the $H \rightarrow b\bar{b}$ decay in VH production mode
- *B-tagging calibration with 2015-16 data JHEP 08 (2018) 89*
  - The paper combines the results of *Tag-and-Probe* method (by our team, USTC-Sorbonne) and *PDF* method (by UCL team)
  - My qualification task (QT), one of the internal note editors
- *VHbb analysis with 2015-16 data JHEP 12 (2017) 024*
  - a.k.a *Evidence paper*, mainly contributed to the statistical study and sample production
  - Presented the results at La Thuile 2018
- *VHbb analysis with 2015-17 data Phys. Lett. B 786 (2018) 59*
  - a.k.a *Observation paper*, mainly contributed to the statistical study by providing the guidance to the younger students and contributed to European Update of Particle Physics Strategy (HL-LHC extrapolation for VHbb)
- *VHbb differential XS measurement JHEP 05 (2019) 141*
  - Dominant contribution including truth categorization, evaluation of the systematics etc.
  - One of the internal note editors and Presented the analysis at the paper approval meeting (PAM)

B-tagging calibration

- Identification of the jets originated from b quarks by exploiting long lifetime of b-hadrons
- MV2 tagger: Boosted Decisions Tree (BDT), combining the output of basic algorithms
- Calibrations of the MV2 tagger
  - PDF method (conventional)
  - Tag-and-Probe method (my QT)
    - Dileptonic ($e^{\pm}\mu^{\mp}$) ttbar events: rich b-jets, high S/B
    - Exactly 2 jets for the Tag-and-Probe:
      - Tag jets: pass 85% tagging WP of *MV2c10*
      - Probe jet: the other jet in the pair High purity
    - The efficiency is measured with probe jets





B-tagging calibration: Results

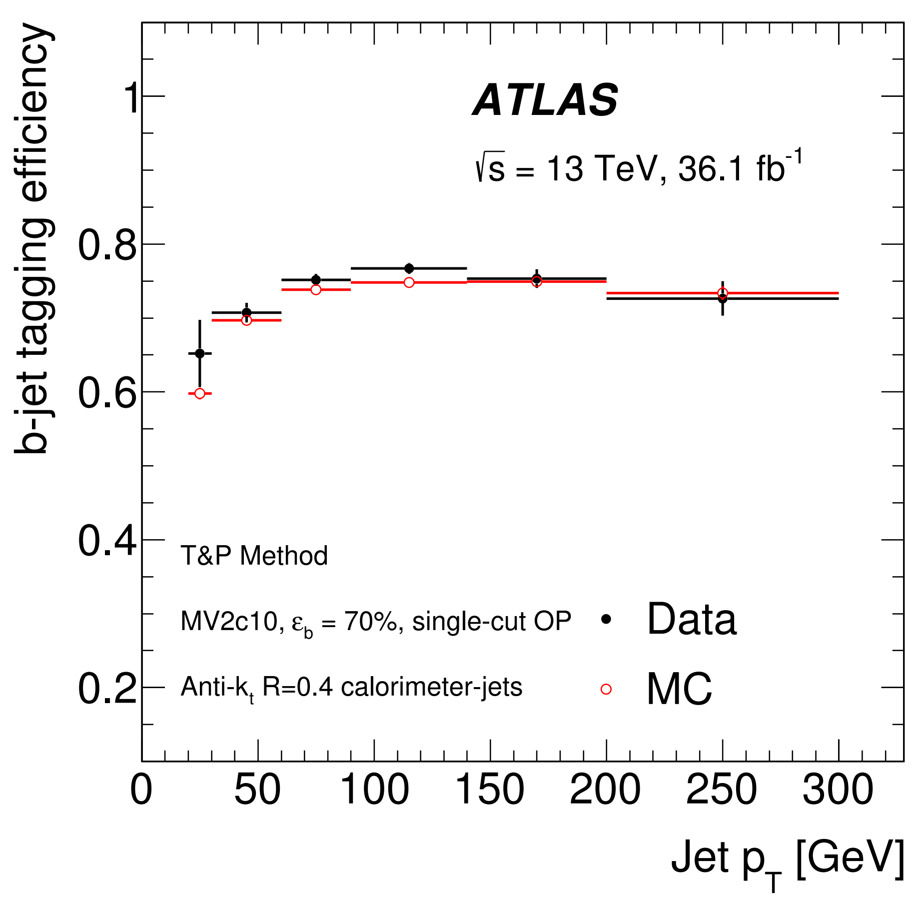
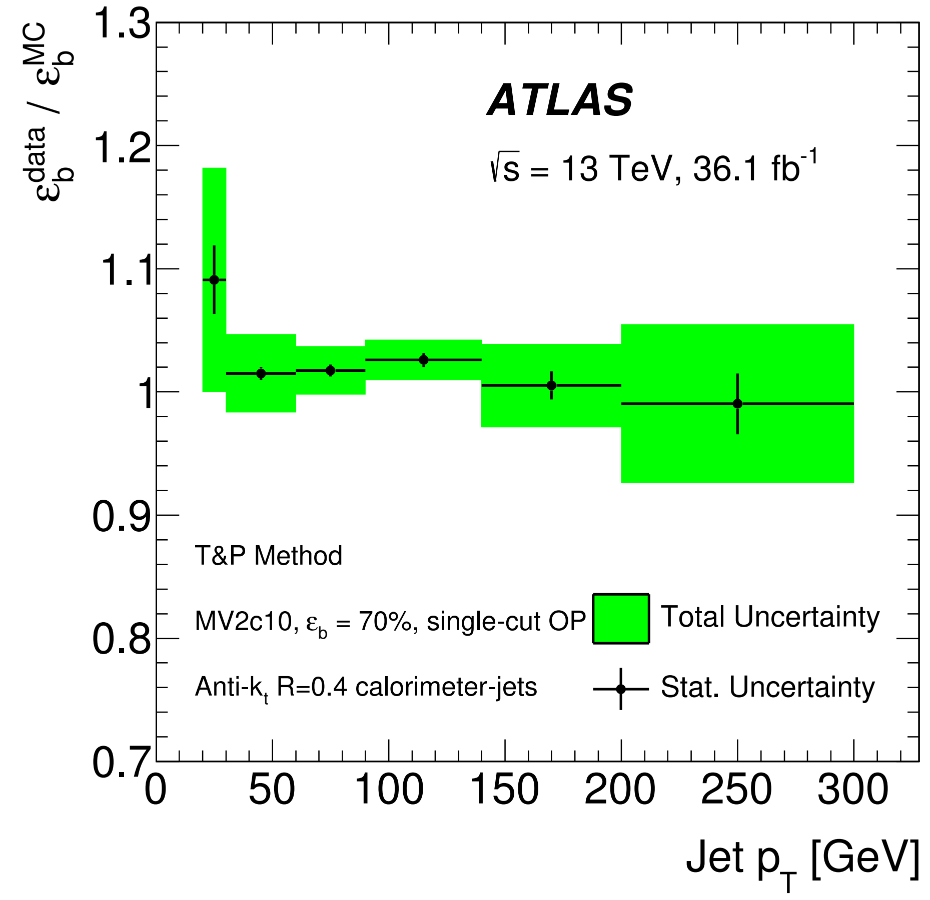
- The MVA approach introduced, a BDT trained to suppress background
- The theory systematics significantly reduced by the cut on BDT score
- Efficiencies are measured vs jet p_T
- Data over MC efficiency scale factors (SF) are derived, to correct ATLAS simulations
- Results:
  - Systematic uncertainties are dominant; they are ~2% for medium-pT jets from $H \rightarrow b\bar{b}$
  - SFs consistent with unity within uncertainties
- The precise calibration supports us for more accurate physics results

Published for SISSA by Springer

Received: May 7, 2018
Revised: July 23, 2018
Accepted: August 6, 2018
Published: August 16, 2018

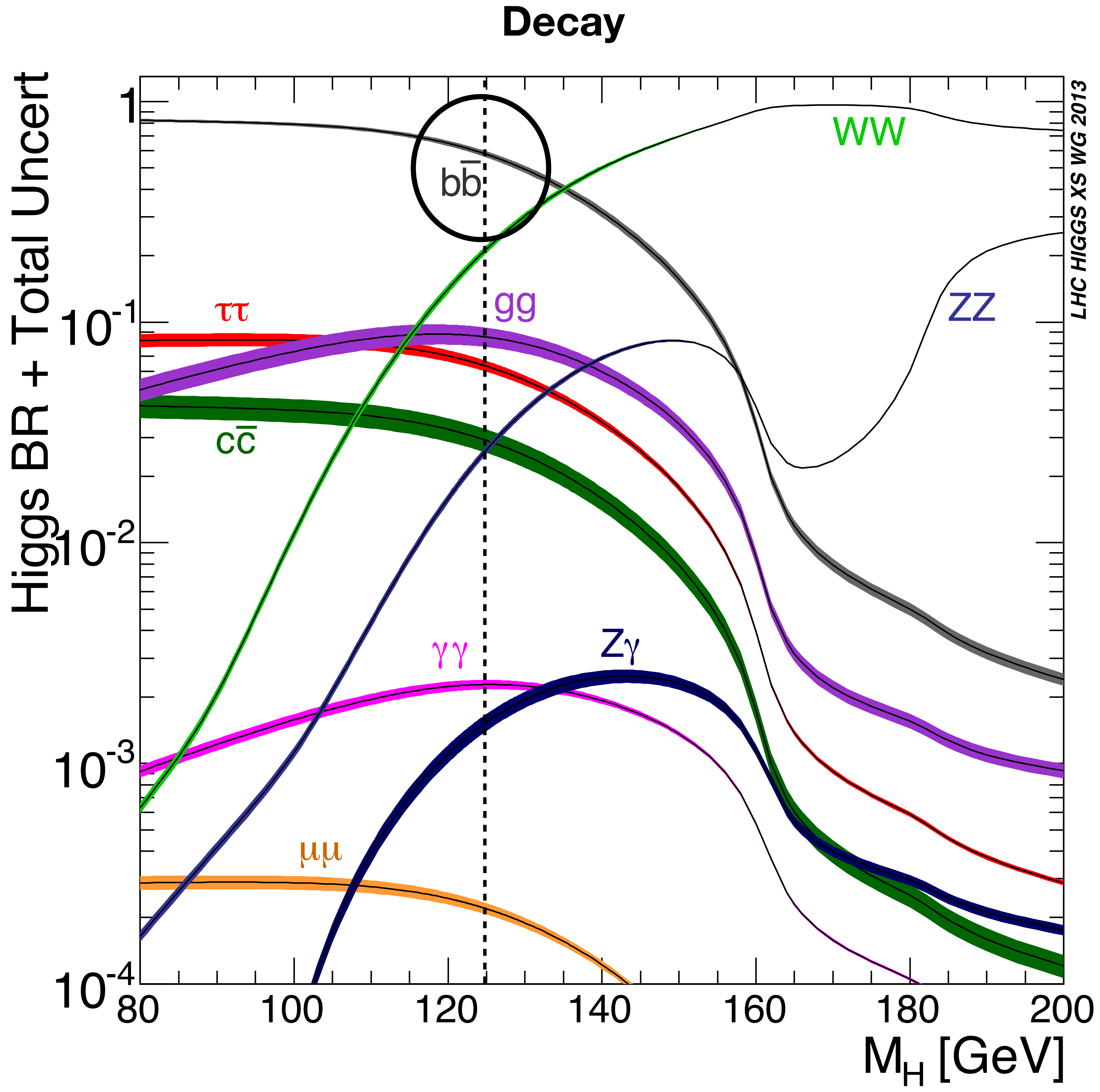
JHEP08(2018)089

Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV

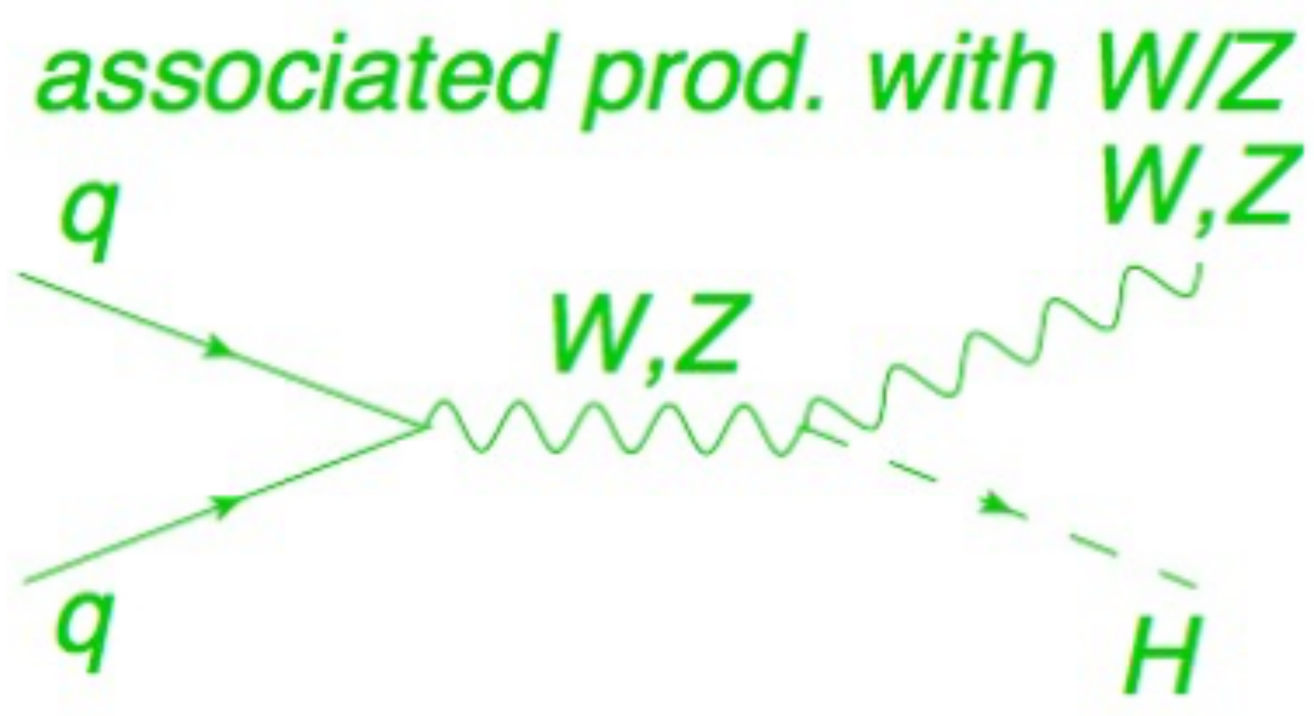
Efficiencies	Scale factors
	

Search for the $H \to b\bar{b}$ decay in VH

- $H \to b\bar{b}$ decay:
  - Highest branching ratio: Br=58%
  - Tests Higgs Yukawa coupling to fermions
  - Br($H \to b\bar{b}$) constrains invisible Higgs decays



- Associated VH production:



  - One of the four main Higgs production modes at LHC
    (at 13 TeV: σ_{tot} ~ 56 pb, σ_{VH} ~ 2.2 pb)
  - Leptonic signature
    - Better triggering
    - Better MJ suppression
  - Most sensitive mode for $H \to b\bar{b}$ at the LHC

Analysis Strategy

- H (→ $b\bar{b}$) recoiling against V (→leptons)

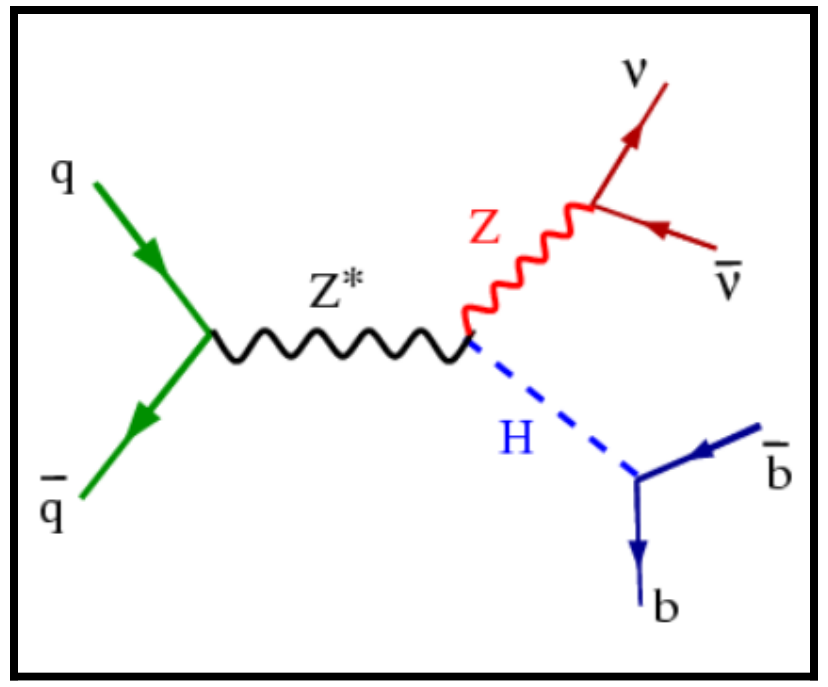
H → $b\bar{b}$

- 2 high-p_T b-jets, not from pile-up, b-tagged
- Kinematic properties consistent with VH production, e.g. m_{bb}~125 GeV

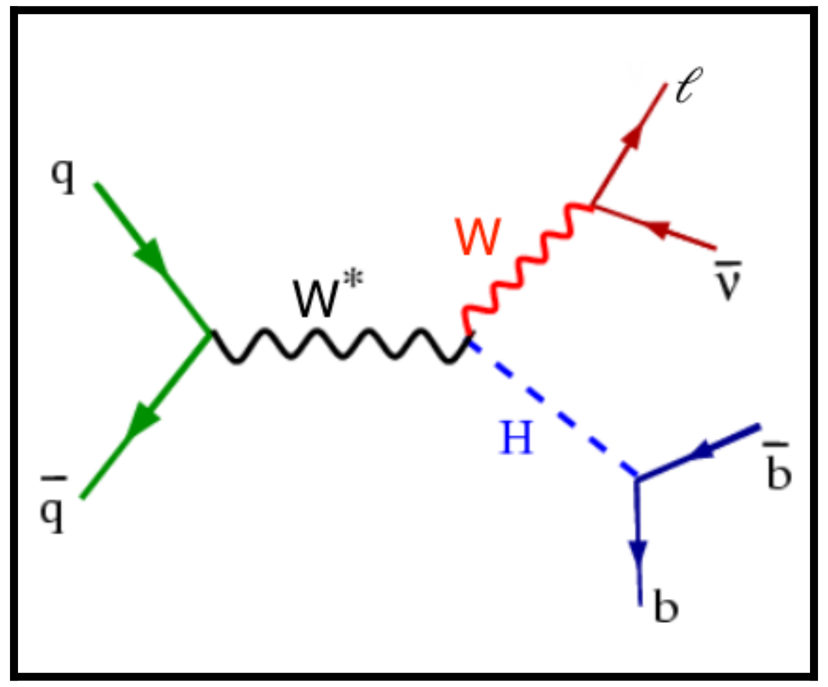
V → leptons

- 1 or 2 isolated charged leptons (W→lν, Z→ll) and/or large MET[1] (Z→νν, W→lν)
  - Also useful for triggering purposes
  - Z→ll: same flavour, m_{ll}~m_Z
- Channels denoted by the number of reconstructed charged leptons (e or μ)

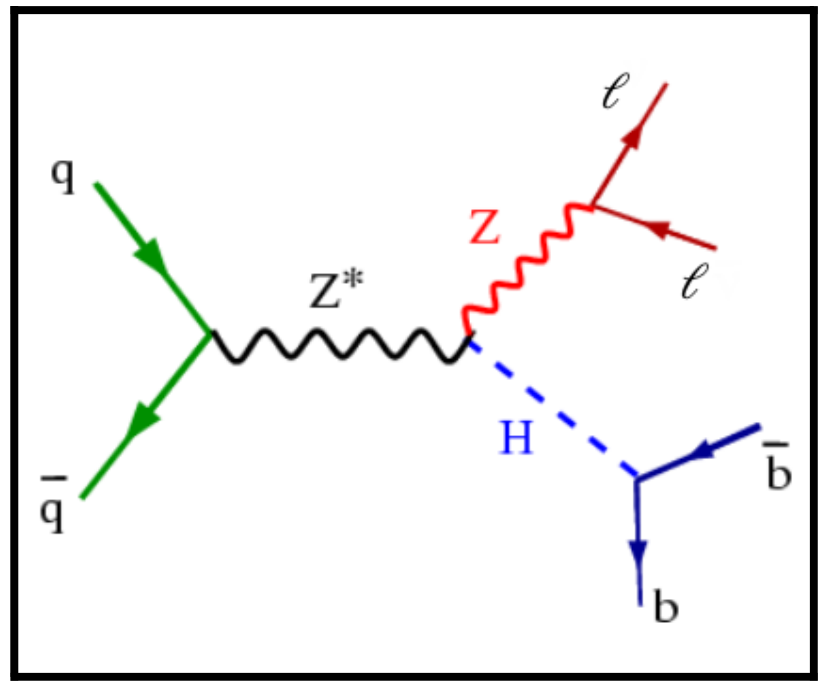
0-lepton



1-lepton



2-lepton

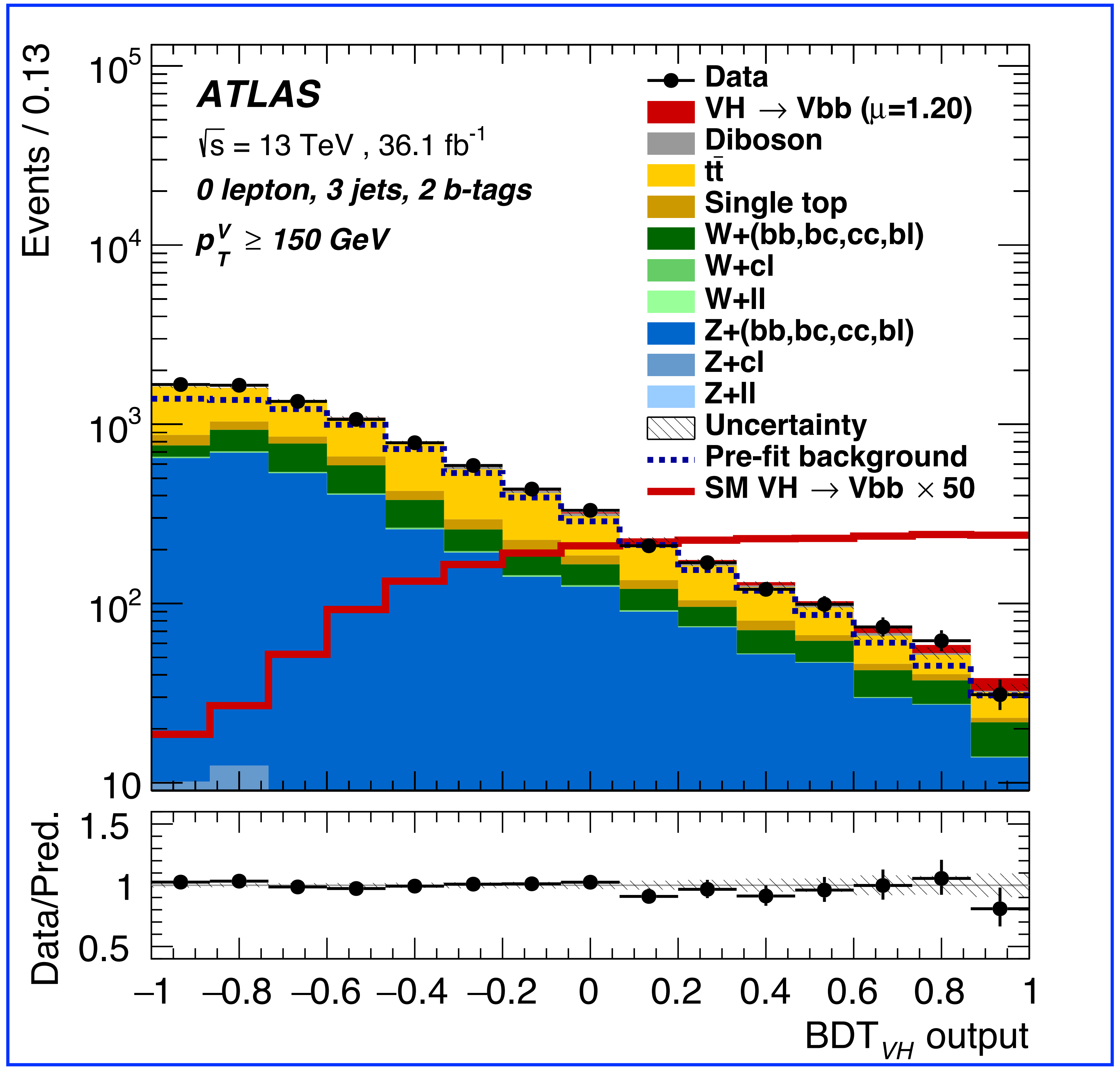


1. Transverse missing energy

Event Categorisation

- Event categories with different S/B to increase sensitivity: split with p_T^V and Njet in each lepton channel
- Main discriminant variables: m_{bb}, p_T^V, and ΔR_{bb} (Combined into a Boosted Decision Tree with other var.)
- Separate training for lepton/pTV/Njet regions
- Combined Likelihood built across channels and multiple analysis regions

Channel	SR/CR	Categories			
		$75\ \text{GeV} < p_T^V < 150\ \text{GeV}$		$p_T^V > 150\ \text{GeV}$	
		2 jets	3 jets	2 jets	3 jets
0-lepton	SR	-	-	BDT	BDT
1-lepton	SR	-	-	BDT	BDT
2-lepton	SR	BDT	BDT	BDT	BDT
1-lepton	W + HF CR	-	-	Yield	Yield
2-lepton	$e\mu$ CR	m_{bb}	m_{bb}	Yield	m_{bb}



Evidence of H→b$\bar{b}$

- Express measurement in terms of signal strength

$$\mu = \frac{\sigma \times BR}{(\sigma \times BR)_{SM}}$$

- Results (36.1 fb^{-1}):
  - Evidence of H→b$\bar{b}$ at 3.5 σ (3.0 σ exp.)
  - $\mu = 1.20^{+0.42}_{-0.36}$
  - Dominant uncertainties : signal modeling, MC statistics, b-tagging

- Run1+Run2 (36.1 fb^{-1}) Combination:
  - Evidence of H→b$\bar{b}$ at 3.6 σ (4.0 σ exp.)
  - Combined $\mu = 0.90^{+0.28}_{-0.26}$

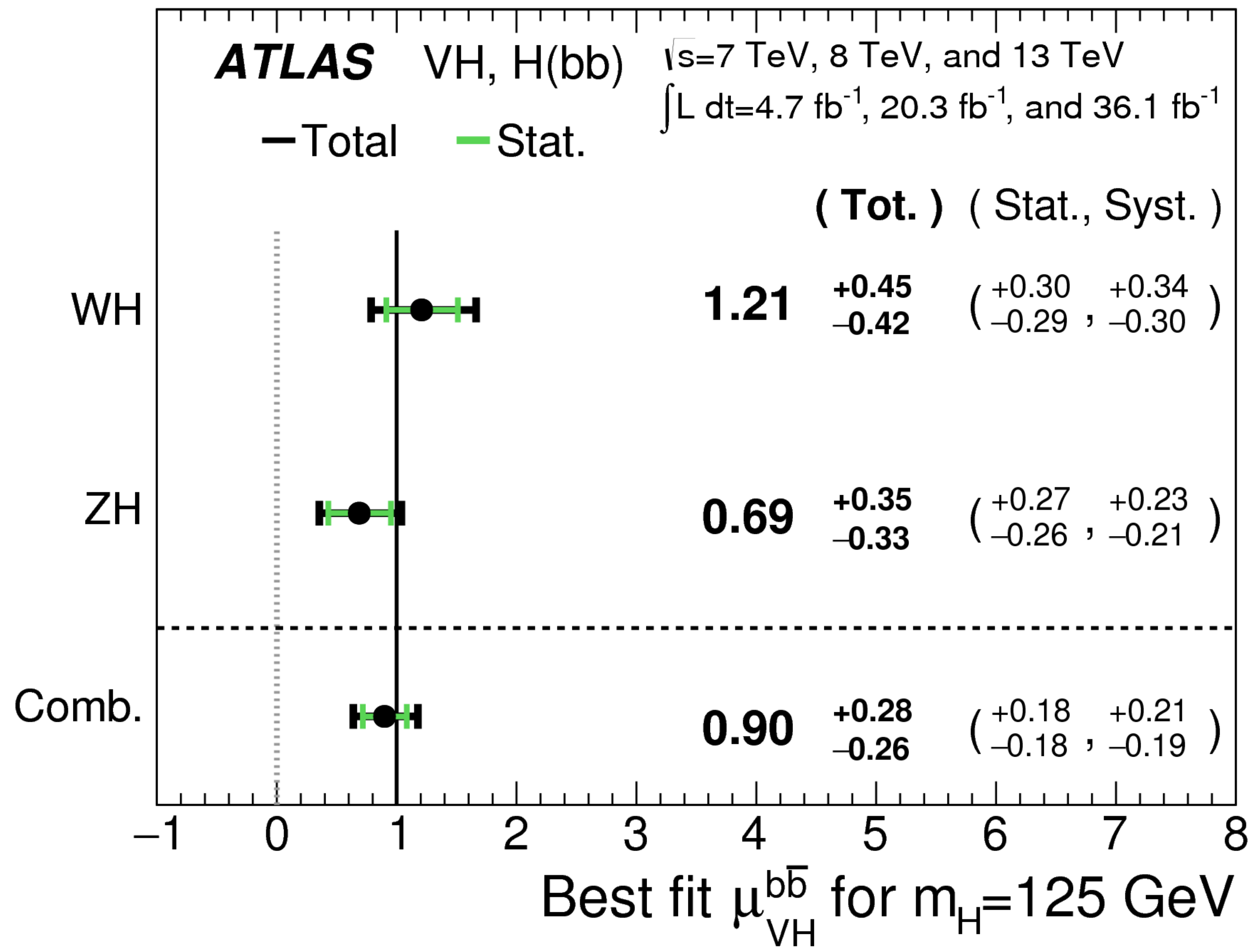
JHEP PUBLISHED FOR SISSA BY SPRINGER

RECEIVED: *August 11, 2017*
ACCEPTED: *October 31, 2017*
PUBLISHED: *December 6, 2017*

JHEP12(2017)024

data 2015-2016: 36 fb^{-1}

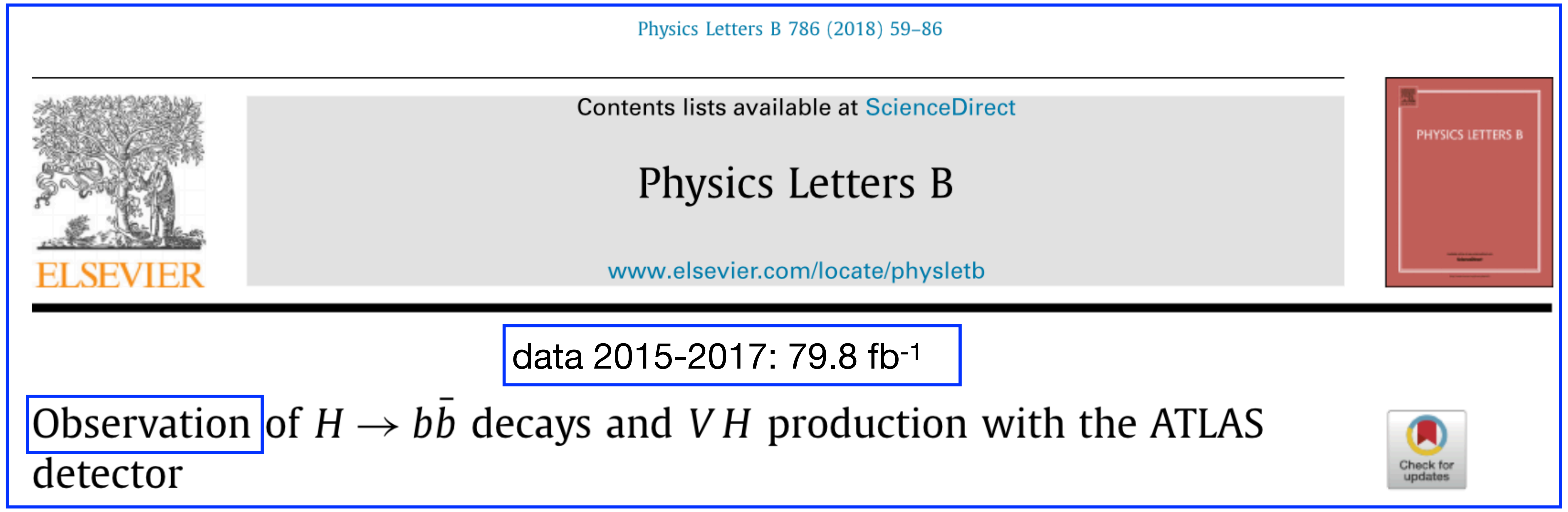
Evidence for the $H \to b\bar{b}$ decay with the ATLAS detector



VHbb Analysis (79.8 fb^{-1})

- Main updates from “evidence”:
  - More data: 79.8 fb^{-1} vs. 36.1 fb^{-1}
  - More MC statistics (filters)
  - Better evaluation of systematic uncertainties
  - Total error reduced by 34%
- Results (79.8 fb^{-1}):
  - Significance: 4.9σ (4.3 σ exp.)
  - $\mu = 1.16^{+0.27}_{-0.25}$
- Combination with Run-1 Analysis:
  - Significance: 4.9σ (5.1 σ exp.)

$$\mu_{\mathrm{WH}} = 1.08^{+0.24}_{-0.23}(\mathrm{stat.})^{+0.29}_{-0.27}(\mathrm{syst.})$$
$$\mu_{\mathrm{ZH}} = 0.92^{+0.21}_{-0.20}(\mathrm{stat.})^{+0.19}_{-0.17}(\mathrm{syst.})$$
$$\mu_{\mathrm{VH}} = 0.98^{+0.14}_{-0.14}(\mathrm{stat.})^{+0.17}_{-0.16}(\mathrm{syst.})$$


Physics Letters B 786 (2018) 59–86

Contents lists available at ScienceDirect

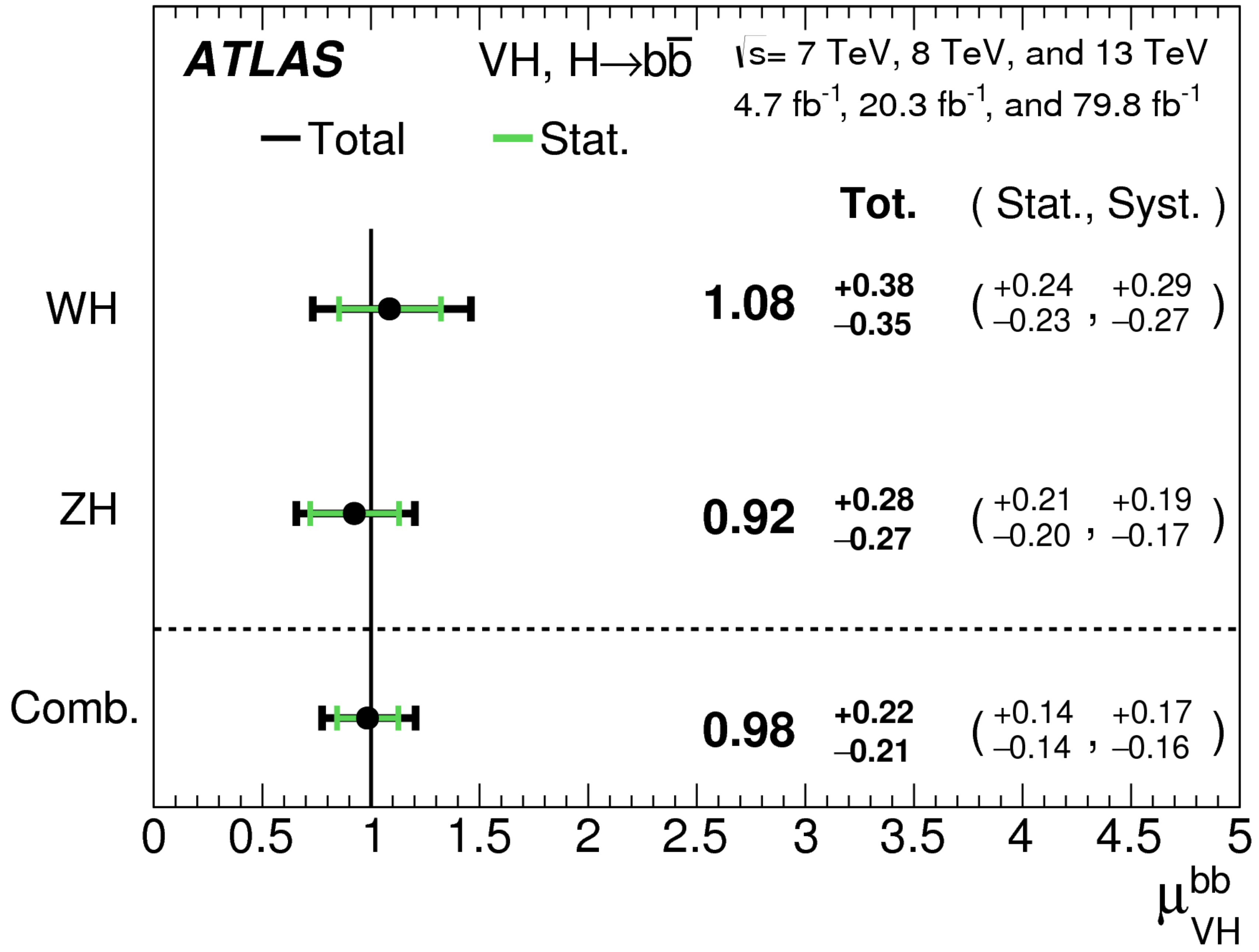
Physics Letters B

www.elsevier.com/locate/physletb

ELSEVIER

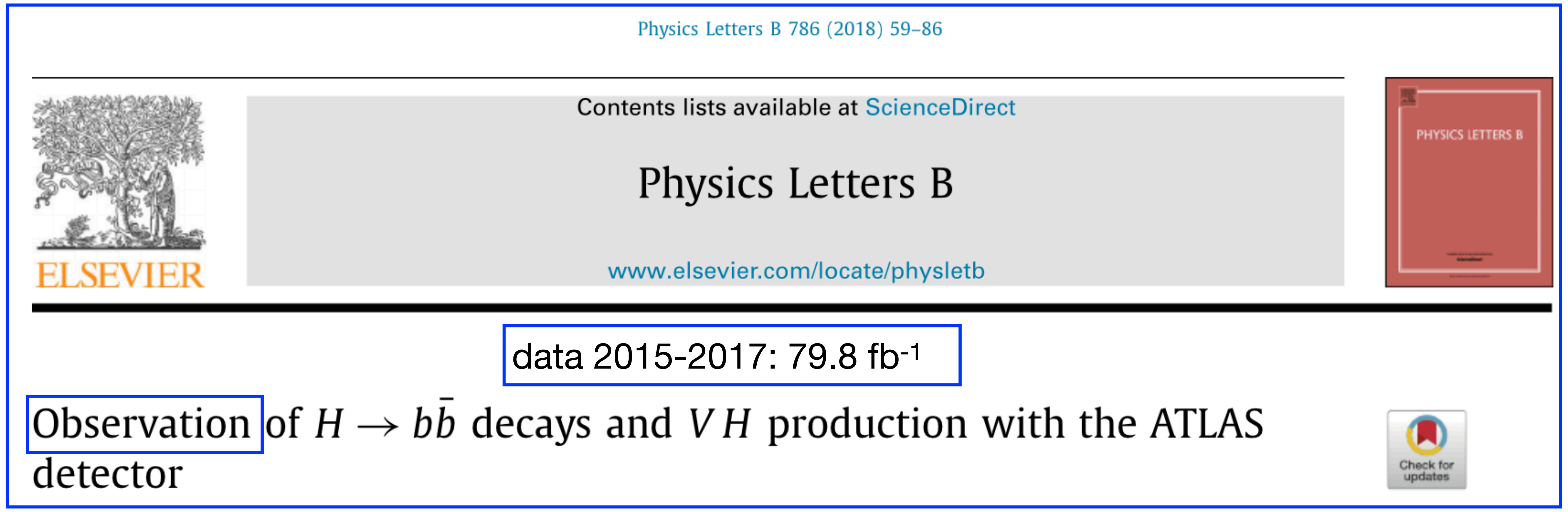
data 2015-2017: 79.8 fb^{-1}

Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector



Observation of $H \to b\bar{b}$ and VH production

- Main updates from "evidence":
  - More data: 79.8 fb^{-1} vs. 36.1 fb^{-1}
  - More MC statistics (filters)
  - Better evaluation of systematic uncertainties
  - Total error reduced by 34%


Physics Letters B 786 (2018) 59–86

Contents lists available at ScienceDirect

Physics Letters B

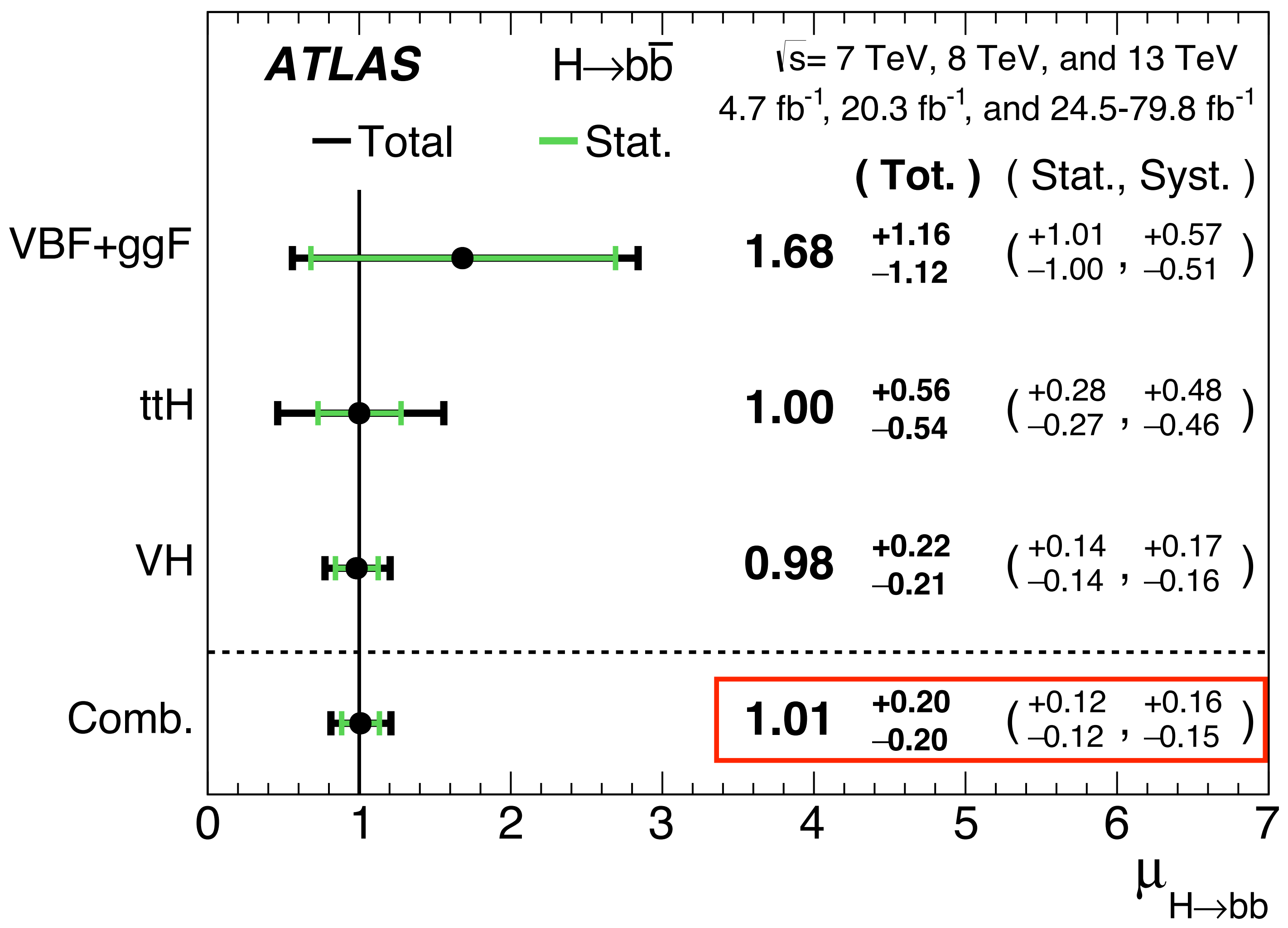
www.elsevier.com/locate/physletb

ELSEVIER

data 2015-2017: 79.8 fb^{-1}

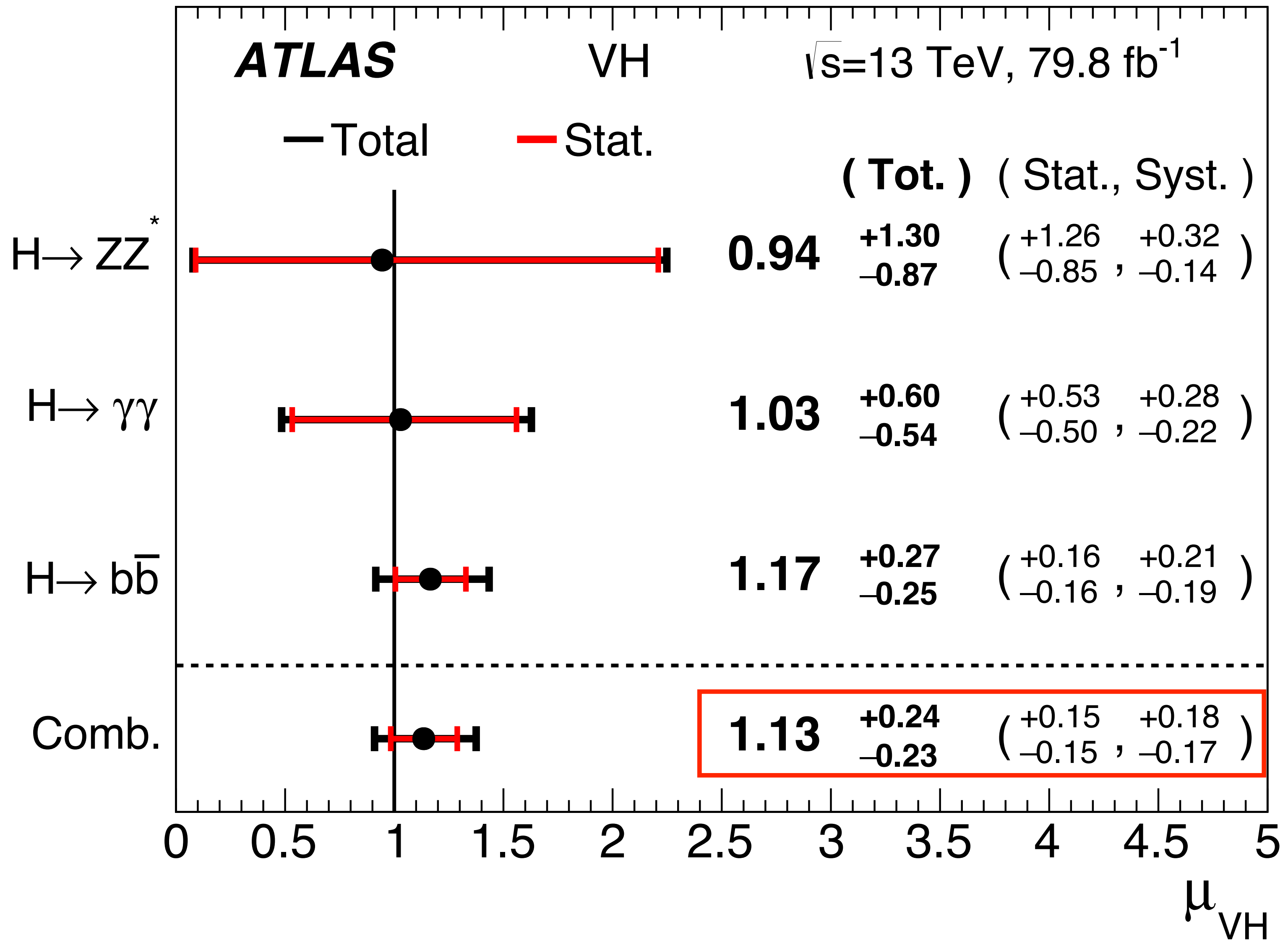
Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector

- Combination with other production modes or decay modes:



- With VBF(+ggF), ttH
- Exp. significance = 5.5σ
- Obs. significance = 5.4σ

$H \to b\bar{b}$ observed



- With $H \to \gamma\gamma$, $H \to ZZ$
- Exp. significance = 4.8σ
- Obs. significance = 5.3σ

VH observed

More detailed measurement

- Measurement of XS as the function of p_T^V
  - Regions sensitive to new physics (higher p_T^V) isolated

WH, $150 < p_T^W < 250$ GeV
WH, $p_T^W > 250$ GeV
ZH, $75 < p_T^Z < 150$ GeV
ZH, $150 < p_T^Z < 250$ GeV
ZH, $p_T^Z > 250$ GeV

4 dimension-6 operators

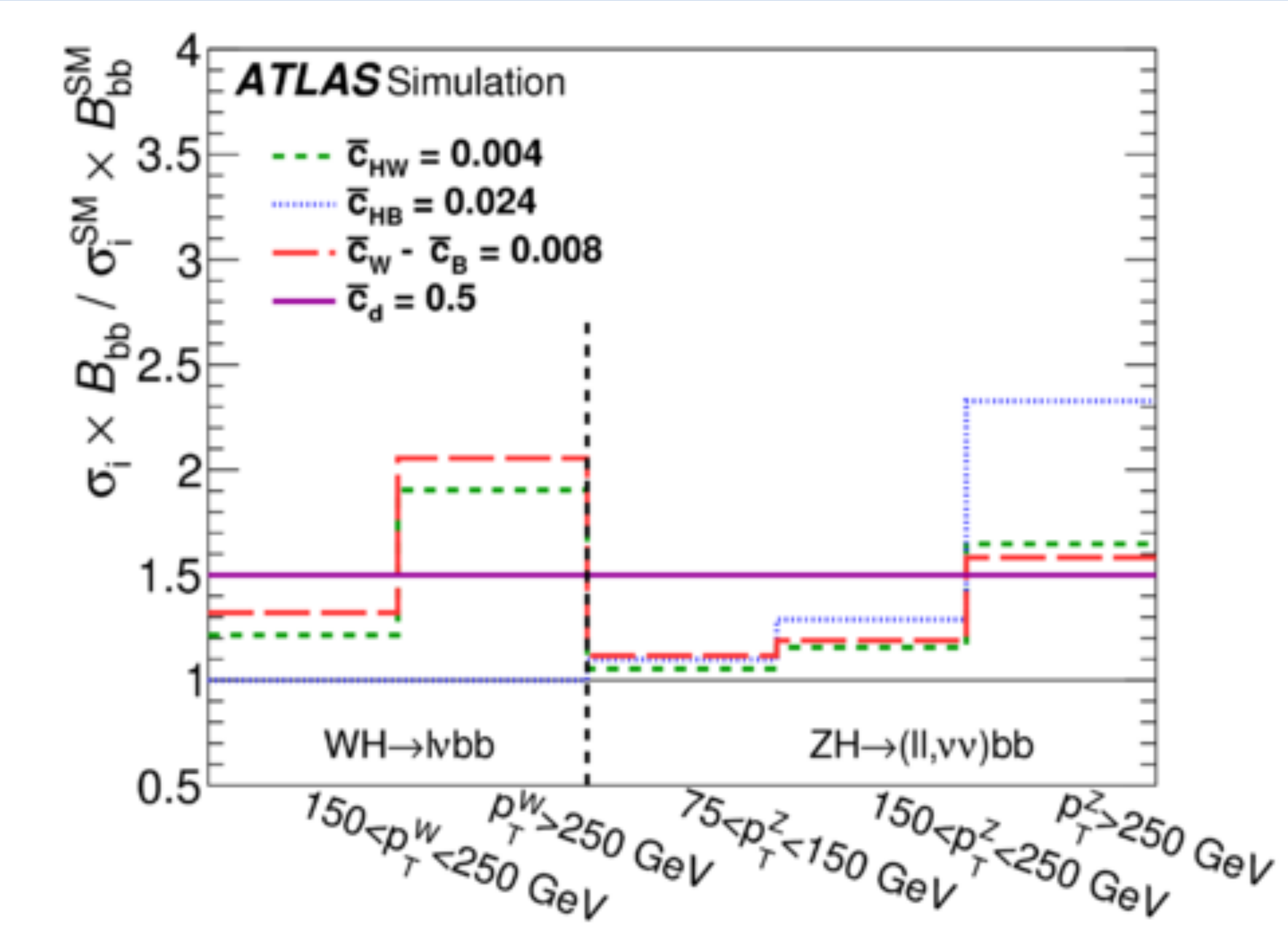
$\mathcal{O}_W$, $\mathcal{O}_B$, $\mathcal{O}_{HW}$, and $\mathcal{O}_{HB}$

coefficients recast into

$$\bar{c}_{HW} = \frac{m_W^2}{g}\frac{c_{HW}}{\Lambda^2},\ \bar{c}_{HB} = \frac{m_W^2}{g'}\frac{c_{HB}}{\Lambda^2},$$

$$\bar{c}_W = \frac{m_W^2}{g}\frac{c_W}{\Lambda^2},\ \bar{c}_B = \frac{m_W^2}{g'}\frac{c_B}{\Lambda^2}$$

the sum $\bar{c}_W + \bar{c}_B$ is already strongly constrained

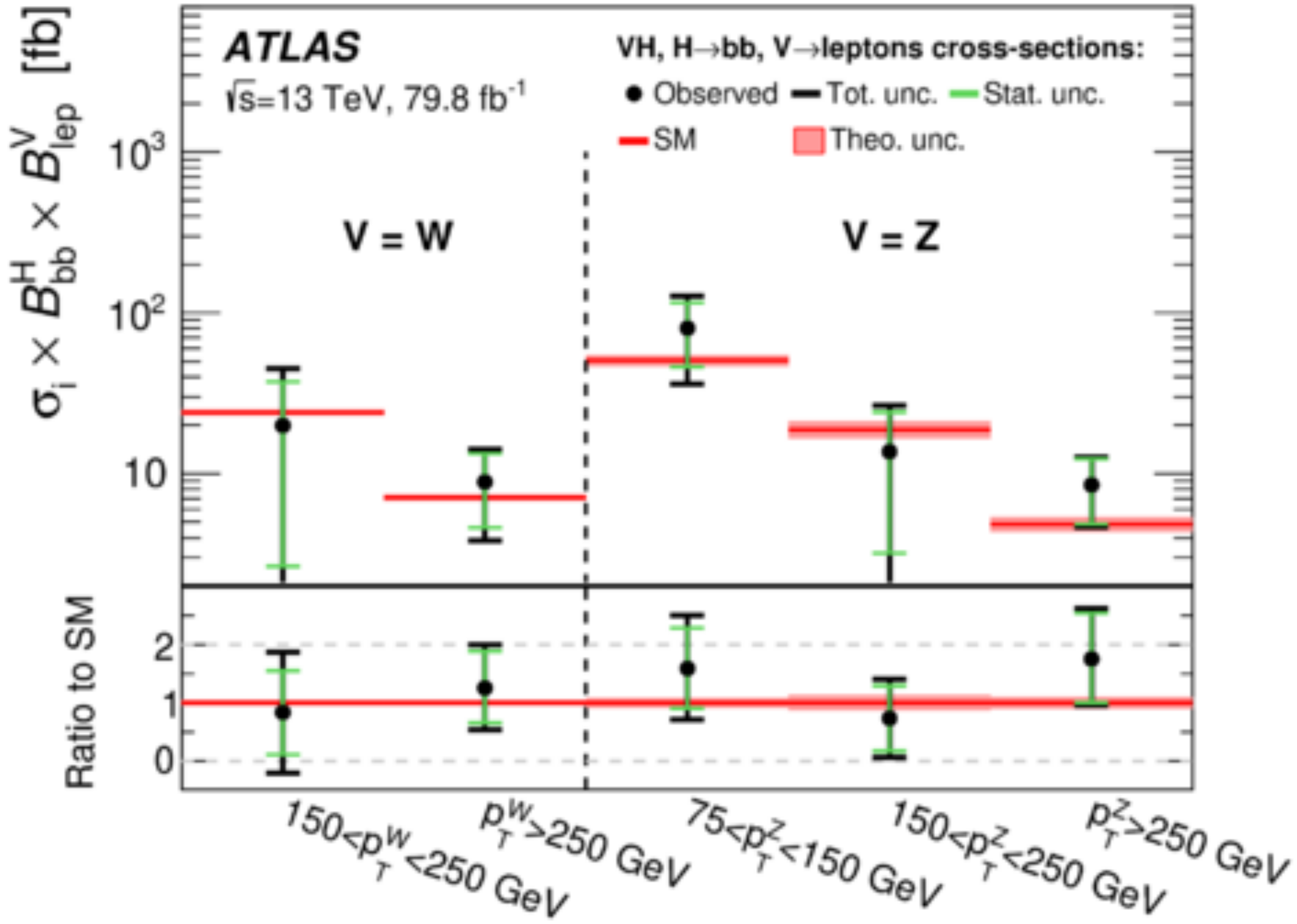


More detailed measurement

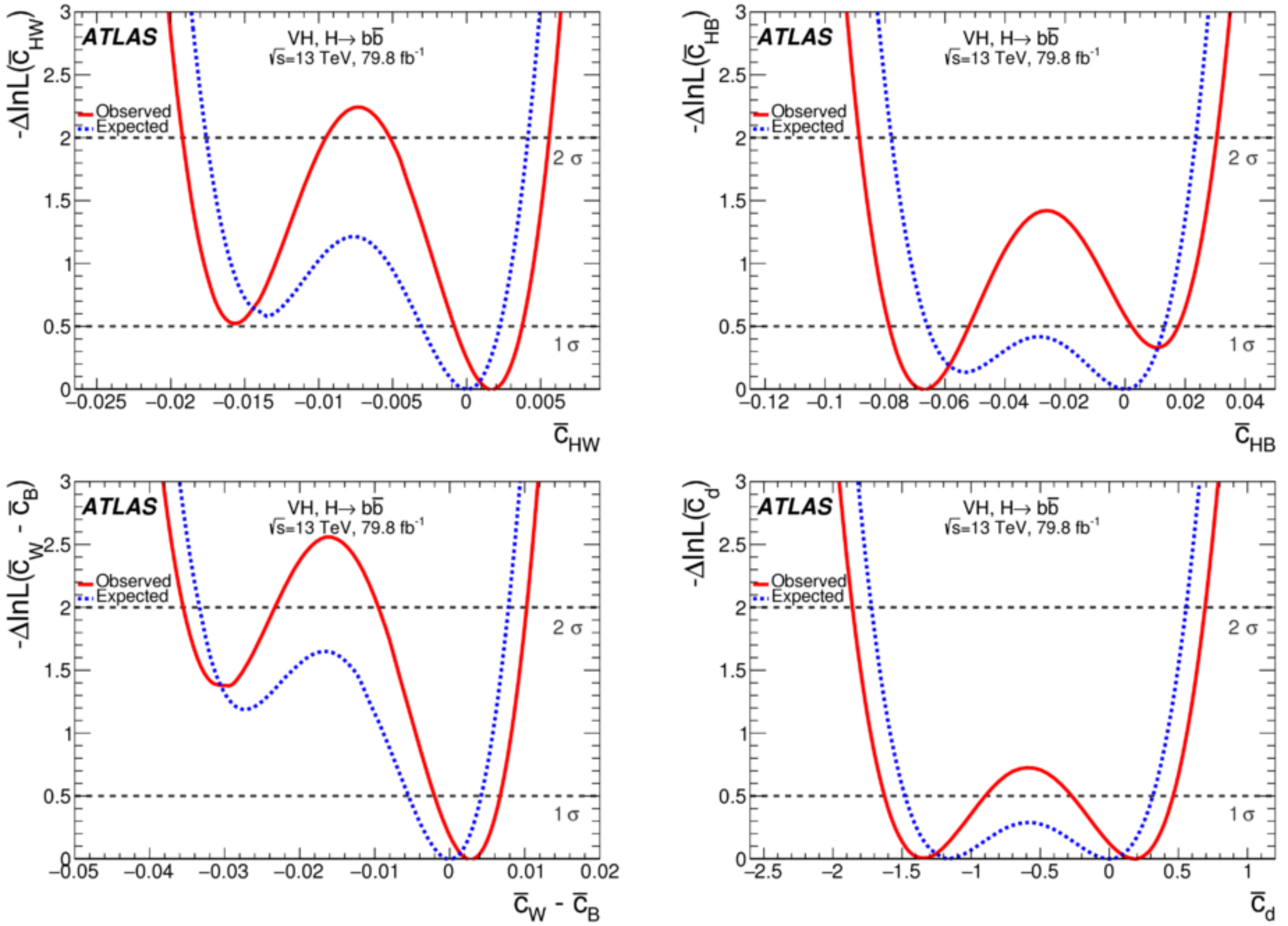
- Measurement of XS as the function of p_T^V
  - Regions sensitive to new physics (higher p_T^V) isolated
- Results:
  - Good agreement between data and SM prediction

JHEP
PUBLISHED FOR SISSA BY SPRINGER
RECEIVED: March 13, 2019
REVISED: May 7, 2019
ACCEPTED: May 13, 2019
PUBLISHED: May 23, 2019
JHEP05(2019)141

Measurement of VH, H → $b\bar{b}$ production as a function of the vector-boson transverse momentum in 13 TeV pp collisions with the ATLAS detector



Constraints set on coefficients of the new physics operators



Summary

- During my Ph.D., I made the contributions to the b-tagging calibration and VHbb analysis for the ATLAS experiments.
- The discovery of the VH and $H \rightarrow b\bar{b}$ obtained through a tight cooperation of the physicists all over the world.
- Chinese teams made indispensable contributions:
  - Nanjing University, Shandong University, SJTU/TDLI, Sun Yat-Sen University, Academia Sinica, USTC, IHEP etc.
- VHbb development after my graduation:
  - The analysis is updated with more statistics. (Full Run2 data)
  - The measurement is performed in new approach. (Boosted)
  - More detailed measurements. (Higher p_T^V bin)
  - More advanced tagging algorithm. ($X \rightarrow bb$ NN tagger developed for boosted)

Grateful to have this chance to communicate with you

on this interesting scientific topic!

Backup

Data and simulated event samples

- data: 36 fb^{-1} collected in 2015 and 2016
- simulated samples of ttbar and main backgrounds (single-top, Z+jets, dibosons)
- Nominal MC samples used to measure central values

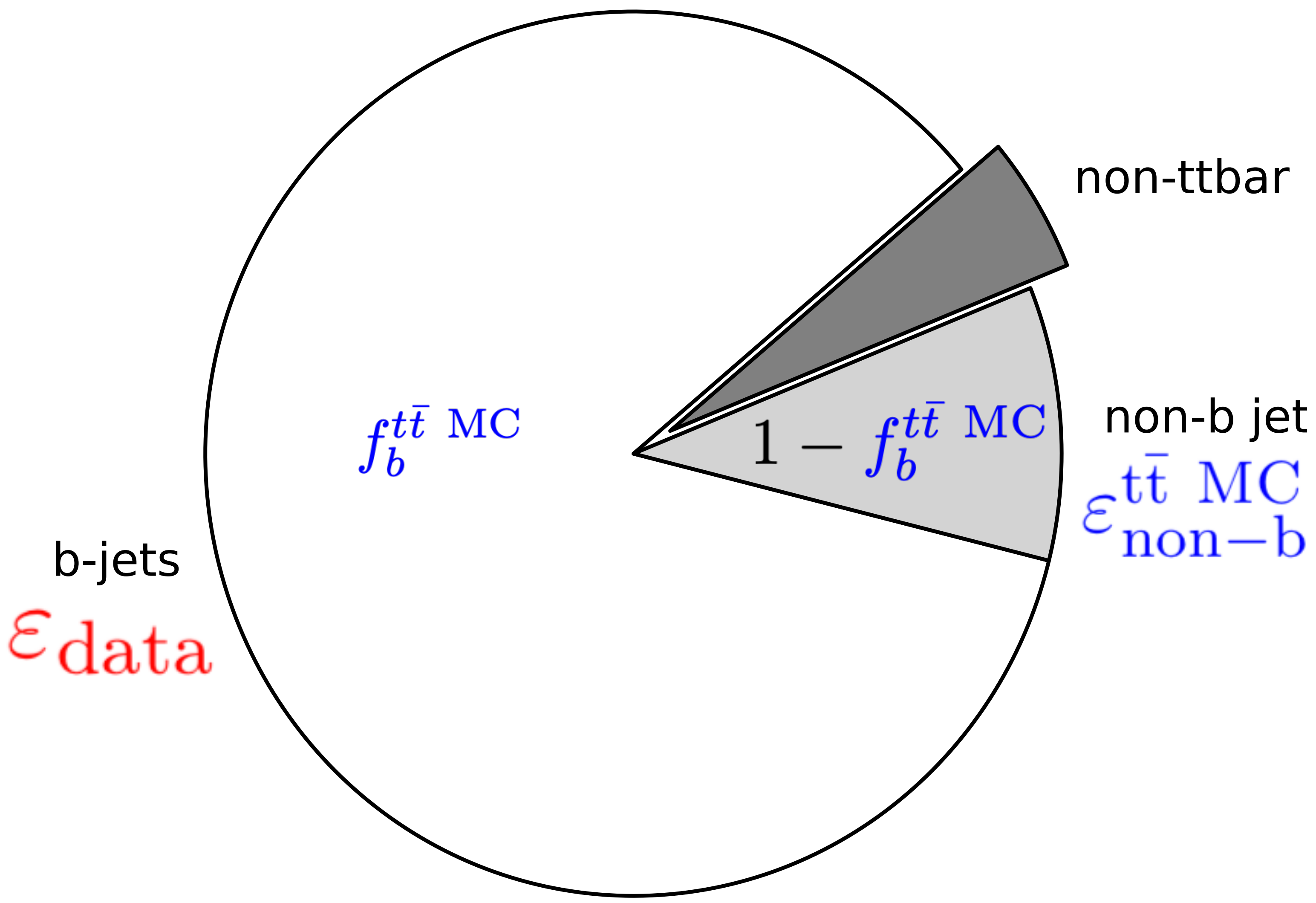
		Events having tag jets
	diboson	291 ± 9
	singletop	5331 ± 39
	Z+jets	570 ± 44
~90%	$t\bar{t}$	66807 ± 157
	fake leptons	853 ± 29
	total MC	73853 ± 170
	data	72792

- Alternative samples used to evaluate modelling systematics

Process	Matrix element generator	PDF Set, Tune	Hadronisation/ Fragmentation	Order in pQCD of Inclusive σ
Nominal Generators				
$t\bar{t}$	POWHEG-BOX v2	CT10, PERUGIA2012	PYTHIA 6.428	NNLO+NNLL
Single top (Wt)	POWHEG-BOX v1	CT10, PERUGIA2012	PYTHIA 6.428	NNLO [84]
Z + jets	MG5_AMC@NLO 2.2.2 [85]	NNPDF23LO [86], A14 [87]	PYTHIA 8.186	NNLO [88]
Diboson	SHERPA 2.1.1	CT10	SHERPA 2.1.1	NLO

tt̄	Hard Scatter Generation	aMC@NLO+Herwig++
	Fragmentation Hadronization	Powheg+Pythia6
	PDF central value	aMC@NLO+Herwig++ with CT10 PDFs aMC@NLO+Herwig++ with PDF4LHC15 100 PDFs
	Additional Radiation	radlo: hdamp = 2 x m_{top}, scale=0.5 radlo: hdamp = 1 x m_{top}, scale=2.0
single top (Wt)	Hard Scatter Generation	aMC@NLO+Herwig++
	Fragmentation Hadronization	Powheg+Pythia6
	Additional Radiation	radlo: hdamp = 2 x m_{top}, scale=0.5 radlo: hdamp = 1 x m_{top}, scale=2.0
	ttbar diagram overlap	Powheg+Pythia6 Diagram Removal Scheme (nominal) Powheg+Pythia6 Diagram Subtract Scheme
Z+jets	Har Scatter Generation	Powheg+Pythia8

MV2c10 Tagging Efficiency Measurement



- Contaminations subtracted relying on the simulation
  - Poor modelling: $\varepsilon^{\mathrm{t\bar{t}\ MC}}_{\mathrm{non-b}}$
  - Large b-jet purity is crucial for the measurement precision

1.Subtract non-ttbar events

$$\varepsilon^{\mathrm{Uncorr}}_{\mathrm{data}} = \frac{N^{\mathrm{pass}}_{\mathrm{data}} - N^{\mathrm{pass}}_{\mathrm{non-t\bar{t}}}}{N_{\mathrm{data}} - N_{\mathrm{non-t\bar{t}}}}$$

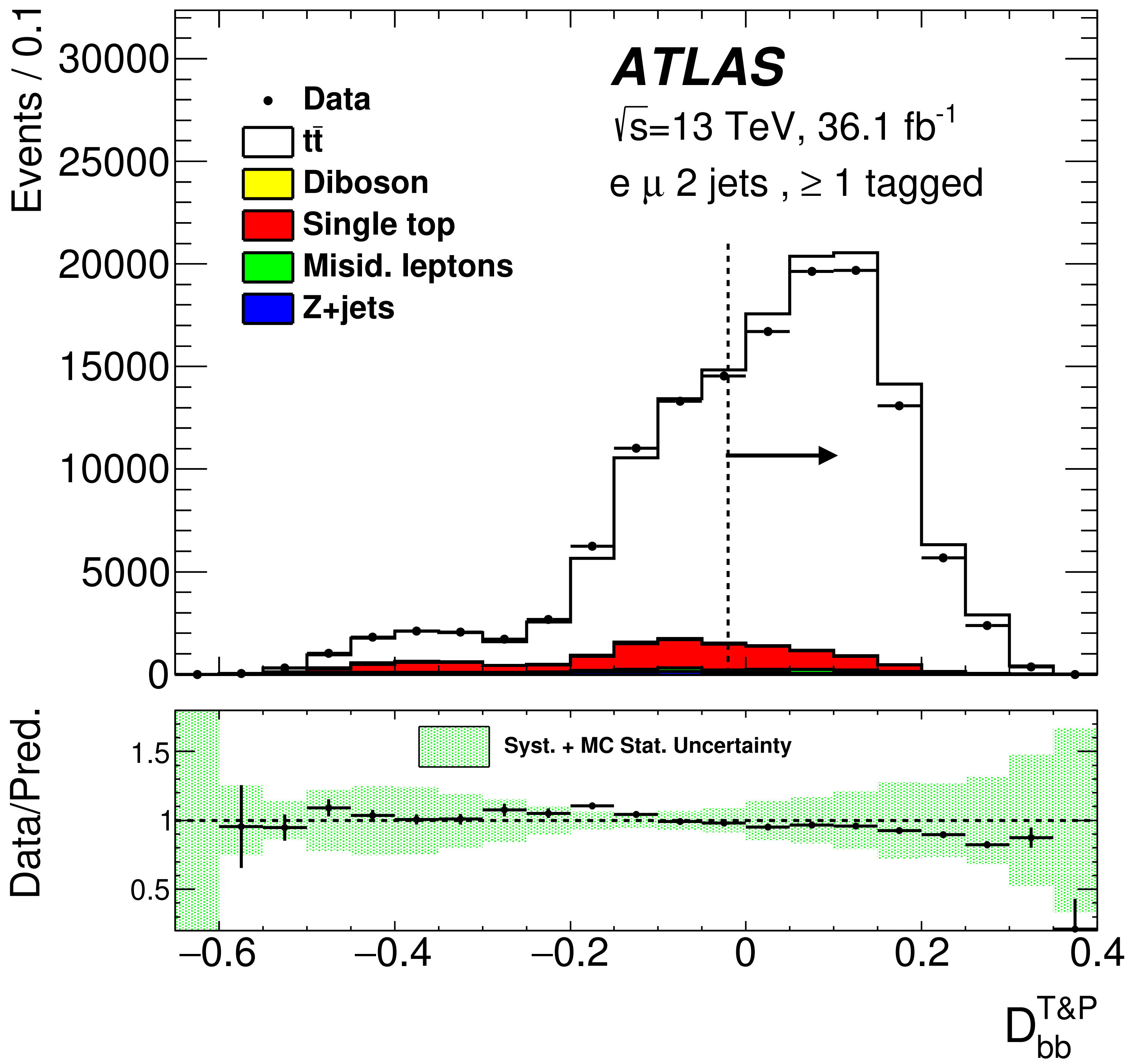
2.Subtract non-b jets

$$\varepsilon^{\mathrm{Uncorr}}_{\mathrm{data}} = f_{\mathrm{b}}^{\mathrm{t\bar{t}\ MC}} \times \varepsilon_{\mathrm{data}} + (1 - f_{\mathrm{b}}^{\mathrm{t\bar{t}\ MC}}) \times \varepsilon^{\mathrm{t\bar{t}\ MC}}_{\mathrm{non-b}}$$

$$\varepsilon_{\mathrm{data}} = \frac{\varepsilon^{\mathrm{Uncorr}}_{\mathrm{data}} - (1 - f_{\mathrm{b}}^{\mathrm{t\bar{t}\ MC}}) \times \varepsilon^{\mathrm{t\bar{t}\ MC}}_{\mathrm{non-b}}}{f_{\mathrm{b}}^{\mathrm{t\bar{t}\ MC}}}$$

Purity BDT: increase the b-jet purity

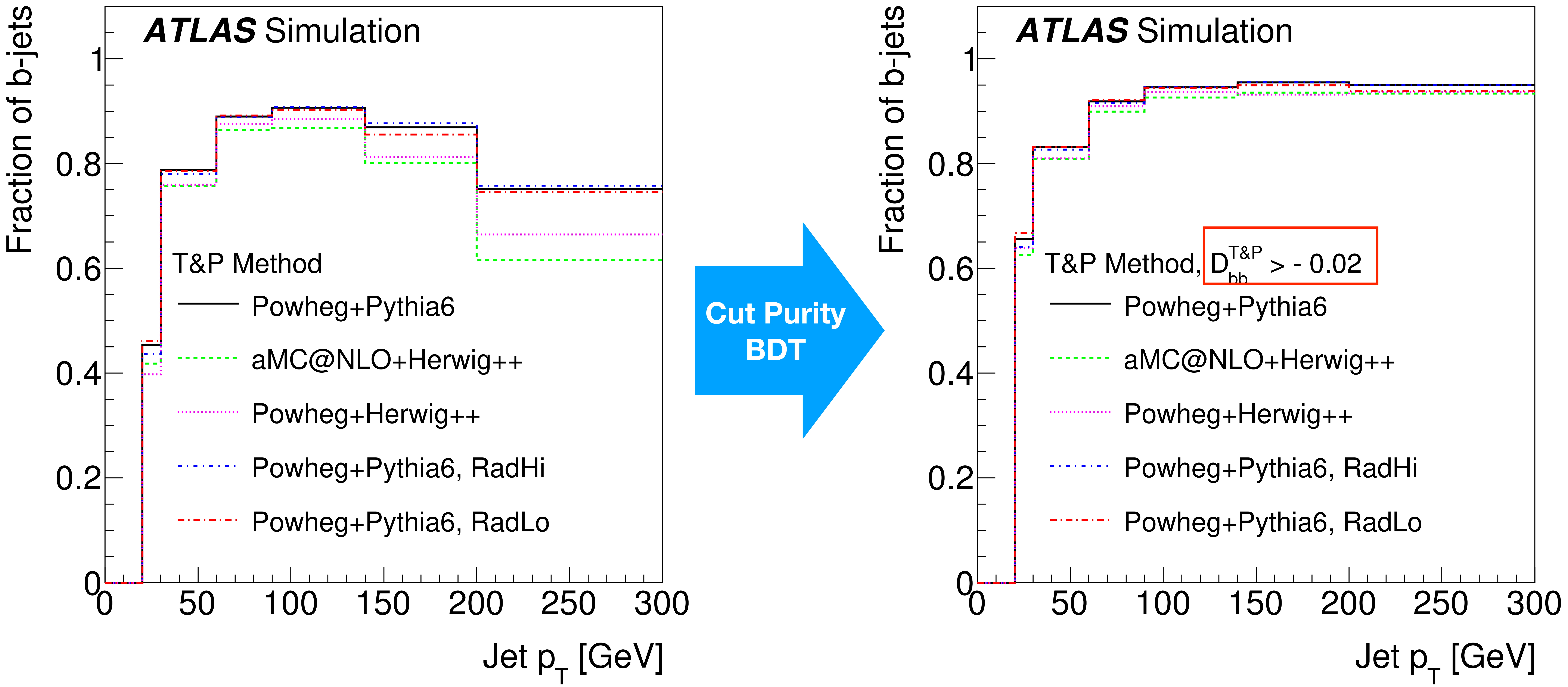
- Larger b-jet purity could be obtained with a larger fraction of events with 2-jet pair as (b,b) flavours
- An event-level "**Purity BDT**" is trained to discriminate (b,b) pair in $t\bar{t}$ from other 2-jet combinations in $t\bar{t}$ and single top events, and used to suppress the latter
- 7 variables used as inputs for the training



Variable	Meaning
$p_{\mathrm{T}}^{j_0}$	p_{T} of the leading-p_{T}jet
$p_{\mathrm{T}}^{j_1}$	p_{T} of the subleading-p_{T}jet
$p_{\mathrm{T}}^{l_0}$	p_{T} of the leading-p_{T}lepton
$E_{\mathrm{T}}^{\mathrm{miss}}$	Missing transverse momentum
nFjets	Number of forward jets
minR(l_0,j)	Minimal ΔR separation between the leading lepton to all jets
m_{lj}	Minimum of the two averages of invariant masses of a jet and a lepton for the two combinations, symbolically, $\min((m_{l0j0}+m_{l1j1})/2,(m_{l0j1}+m_{l1j0})/2)$

Optimisation of the Purity BDT requirement

- Cut on the purity BDT can increase the b-jet purity (reducing syst. uncertainties) at the cost of smaller number of probes (increasing stat. uncertainties)



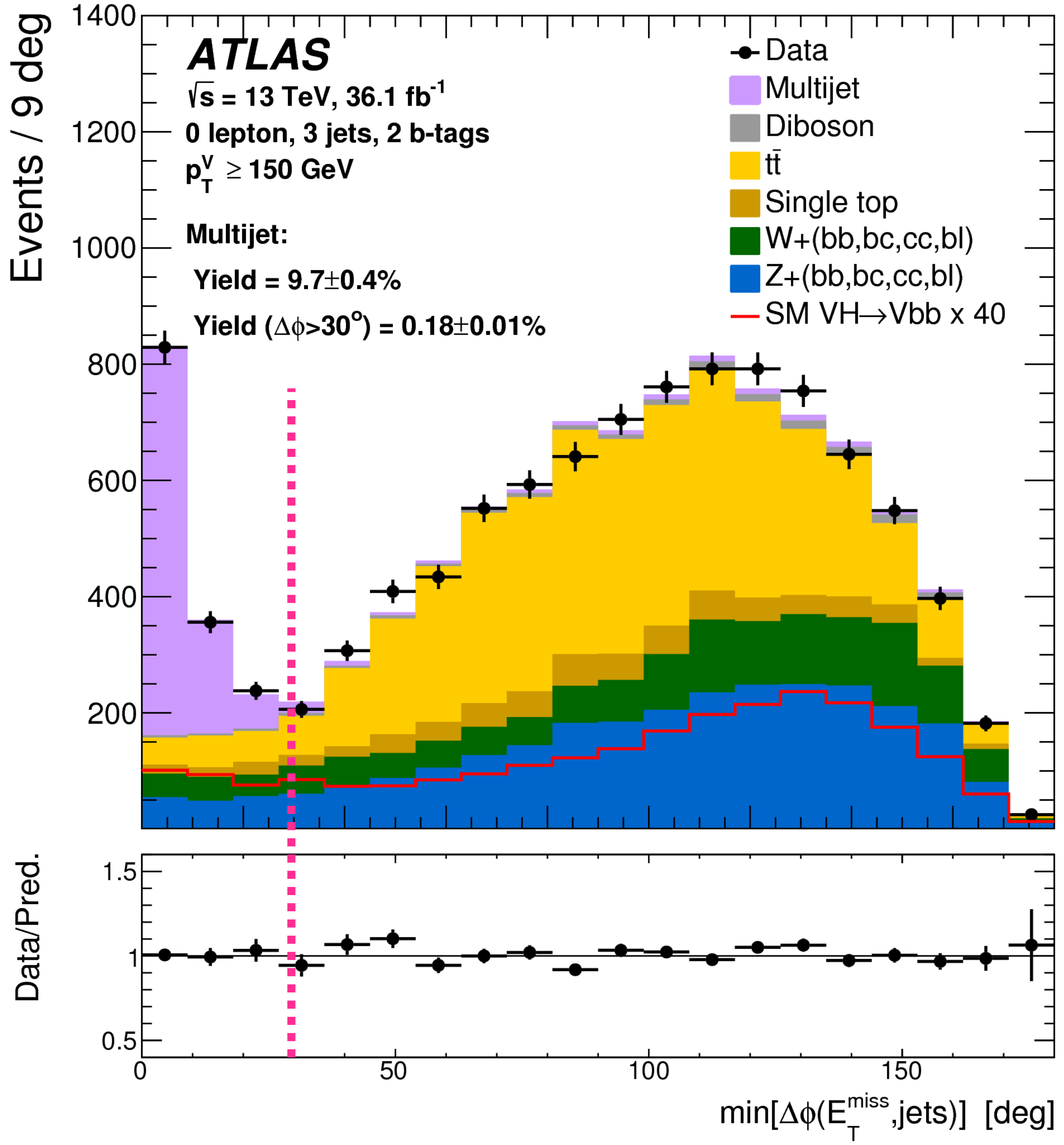
- Chosen cut minimises total (stat+syst) uncertainty

- Optimisation favours cut that improves the b-jet purity quite significantly for low-p_T and high-p_T jets

Specific selection vs lepton channel

- Additional selection criteria to suppress processes hard to model and estimate: QCD multi-jet
  - Take the 0-lepton as an example

Variable	Selection
MET	>150 GeV
$H_T = \Sigma\, p_T$ jets	>120 GeV for 2-jet events >150 GeV for 3-jet events
$\Delta\phi(E_T^{miss}, p_T^{miss})$	< 90°
$\Delta\phi(b,b)$	< 140°
$\Delta\phi(E_T^{miss}, bb)$	> 120°
$\min[\Delta\phi(E_T^{miss}, jet)]$	> 20° for 2-jet events > 30° for 3-jet events



- H_T cut to avoid trigger turn-on mis-modelling
- Angular cuts to reject QCD multi-jet

The Fit Model

- Combined Likelihood fit is built across channels and multiple analysis regions

Channel	SR/CR	Categories			
		$75\,\text{GeV} < p_T^V < 150\,\text{GeV}$		$p_T^V > 150\,\text{GeV}$	
		2 jets	3 jets	2 jets	3 jets
0-lepton	SR	—	—	BDT	BDT
1-lepton	SR	—	—	BDT	BDT
2-lepton	SR	BDT	BDT	BDT	BDT
1-lepton	W + HF CR	—	—	Yield	Yield
2-lepton	$e\mu$ CR	m_{bb}	m_{bb}	Yield	m_{bb}

- Each bin contributes with a Poisson term

$$\mathcal{L}(\mu,\boldsymbol{\theta}) = \left[\prod_{i\in\text{bins}} \text{Pois}\left(n^i \middle| \mu\nu_s^i(\boldsymbol{\theta}) + \nu_b^i(\boldsymbol{\theta})\right)\right]$$

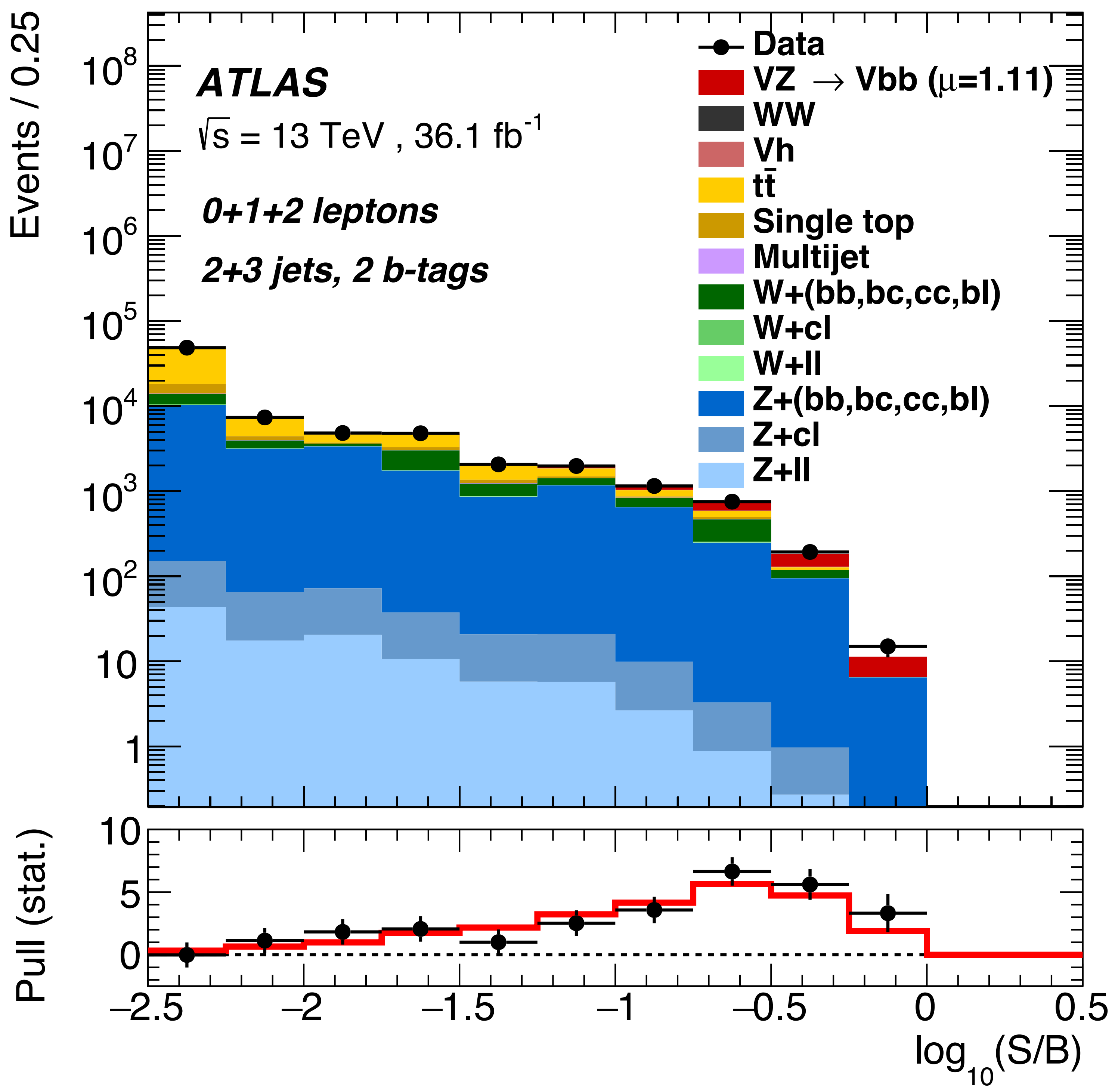
Parameter of interest

$$\mu = \frac{\sigma \cdot \text{BR}}{\sigma_{\text{SM}} \cdot \text{BR}_{\text{SM}}}$$

Nuisance parameters (NPs) $\boldsymbol{\theta}$:

- Uncertainties from performance:
  - Lepton / Jet / MET / b-tagging
- Parametrized shapes and relative normalisations across regions

Cross-checks: results with 36 fb^{-1}

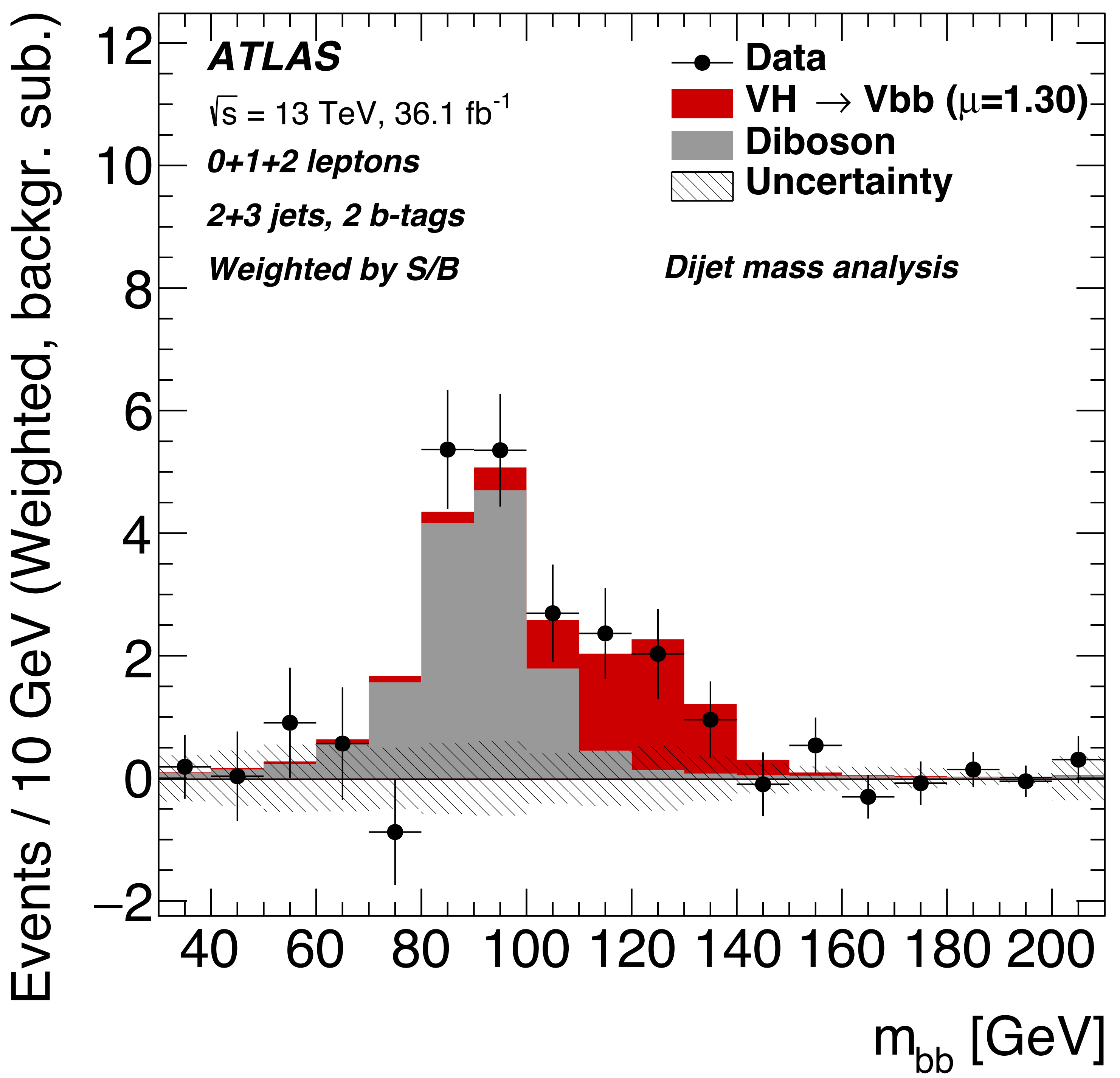


- Signal strength

$$\mu_{VZ} = 1.11^{+0.12}_{-0.11}(\text{stat.})^{+0.22}_{-0.19}(\text{syst.})$$

- Expected significance: 5.3σ
- Observed significance: 5.8σ

VZ, Z → bb̄



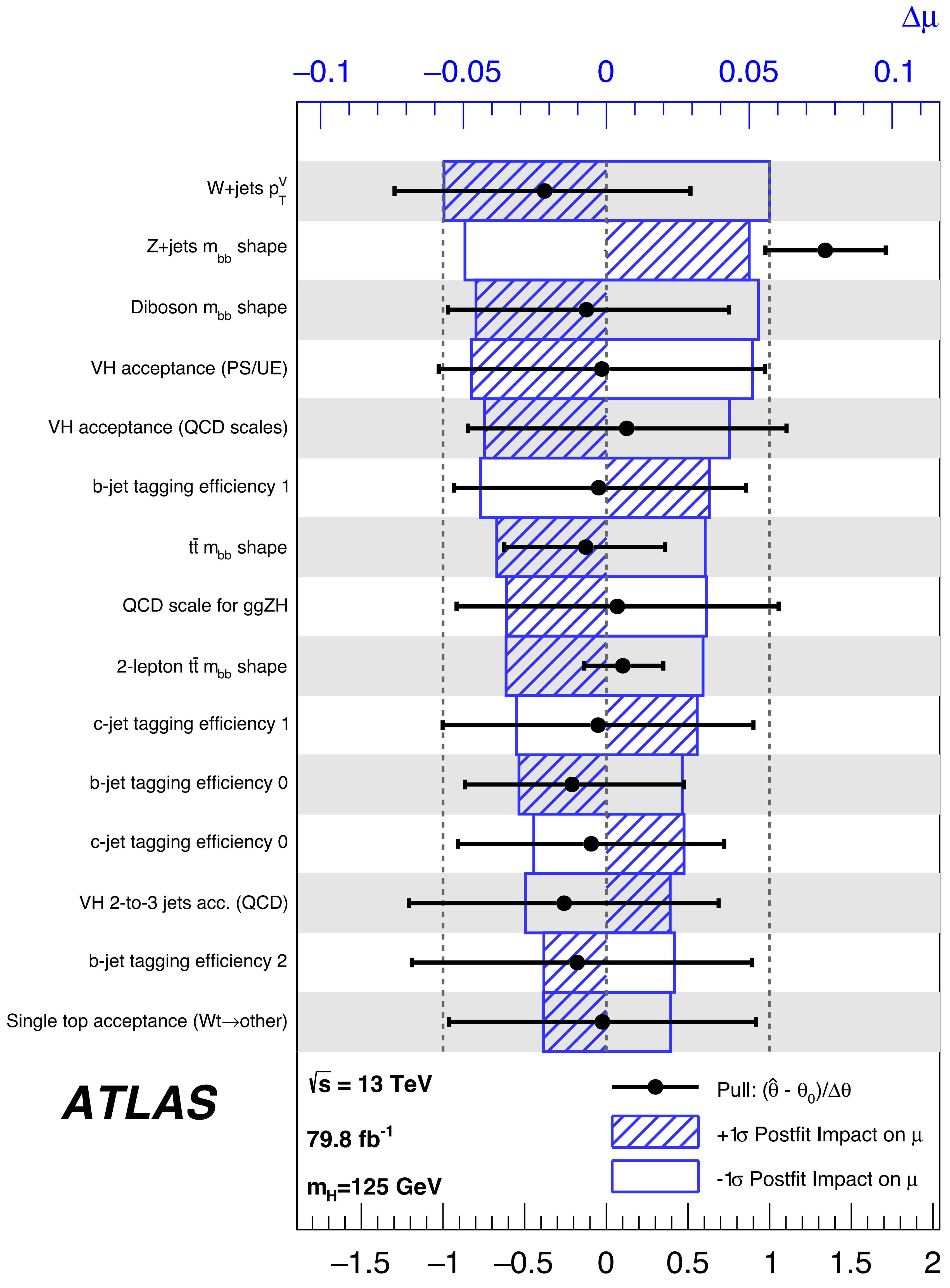
- Signal strength

$$\mu = 1.30^{+0.28}_{-0.27}(\text{stat.})^{+0.37}_{-0.29}(\text{syst.})$$

- Expected significance: 2.8σ
- Observed significance: 3.5σ

VH, H → bb̄
Dijet mass

Impact of systematic uncertainties on μ_{VH}



Source of uncertainty		σ_μ
Total		0.259
Statistical		0.161
Systematic		0.203
Experimental uncertainties		
Jets		0.035
E_T^{miss}		0.014
Leptons		0.009
b-tagging	b-jets	0.061
	c-jets	0.042
	light jets	0.009
	extrapolation	0.008
Pile-up		0.007
Luminosity		0.023
Theoretical and modelling uncertainties		
Signal		0.094
Floating normalisations		0.035
Z + jets		0.055
W + jets		0.060
$t\bar{t}$		0.050
Single top quark		0.028
Diboson		0.054
Multi-jet		0.005
MC statistical		0.070

- Dominant effects:
  - Signal Modelling
  - Background Modelling
    - W+jet
    - Single top Wt
    - Z+jets
    - $t\bar{t}$
  - b-tagging calibration
  - Limited Monte Carlo statistics

Citation (not precise)

Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV — [PDF] arxiv.org

Atlas Collaboration - arXiv preprint arXiv:1805.01845, 2018 - arxiv.org

Page 1. EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN) JHEP 08 (2018) 89 DOI: 10.1007/JHEP08(2018)089 CERN-EP-2018-047 5th September 2018 Measurements of b-jet tagging efficiency with the ATLAS detector using t ̄t events at √ s = 13 TeV …

☆ 99 被引用次数：433 相关文章 所有 18 个版本 »

Evidence for the $H \to b\bar{b}$ decay with the ATLAS detector — [PDF] arxiv.org

Atlas Collaboration - arXiv preprint arXiv:1708.03299, 2017 - arxiv.org

… Submitted to: JHEP … arXiv:1708.03299v2 [hep-ex] 11 Aug 2017 … was established only via the observation of the decay into τ-leptons through the combination of ATLAS and CMS Run 1 results [11], and, more recently, through the combination of CMS Run 1 and Run 2 results [12] …

☆ 99 被引用次数：269 相关文章 所有 19 个版本 »

Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector — [PDF] arxiv.org

Atlas Collaboration - arXiv preprint arXiv:1808.08238, 2018 - arxiv.org

Page 1. EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN) Phys. Lett. B 786 (2018) 59 DOI: 10.1016/j.physletb.2018.09.013 CERN-EP-2018-215 2nd October 2018 Observation of H → b ̄ b decays and VH production with the ATLAS detector …

☆ 99 被引用次数：360 相关文章 所有 29 个版本 »

Measurement of VH, $H \to b\bar{b}$ production as a function of the vector-boson transverse momentum in 13 TeV pp collisions with the ATLAS detector — [PDF] arxiv.org

ATLAS collaboration - arXiv preprint arXiv:1903.04618, 2019 - arxiv.org

Page 1. EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN) JHEP 05 (2019) 141 DOI: 10.1007/JHEP05(2019)141 CERN-EP-2019-019 13th November 2019 Measurement of VH, H → b ̄ b production as a function of the vector-boson transverse momentum …

☆ 99 被引用次数：56 相关文章 所有 56 个版本 »