Search for the Standard Model Higgs boson produced in association with top quarks in the H->bb final state with ATLAS

arXiv.1712.08895

Alessandro Calandri
CPPM-Aix Marseille Université

IPHC Séminaire Hautes Energies
Strasbourg, March 16th 2018

Outline of the talk

- ✓ Search for the Standard Model ttH→bb
  - results released by ATLAS on 2015+2016 Run 2 data https://arxiv.org/abs/1712.08895 (accepted for publication in Phys. Rev. D)
  - quick detour on b-tagging algorithm and performance optimization for the physics analysis (ATLAS-PHYS-PUB-2017-013)
- ✓ Recap on the ATLAS results on 2015+2016 data on ttH→WW/ZZ/ττ ("multilepton" final state) and final ttH combination, https://arxiv.org/abs/1712.08891 (accepted for publication in Phys. Rev. D)
- ✓ Wrapping-up and conclusions

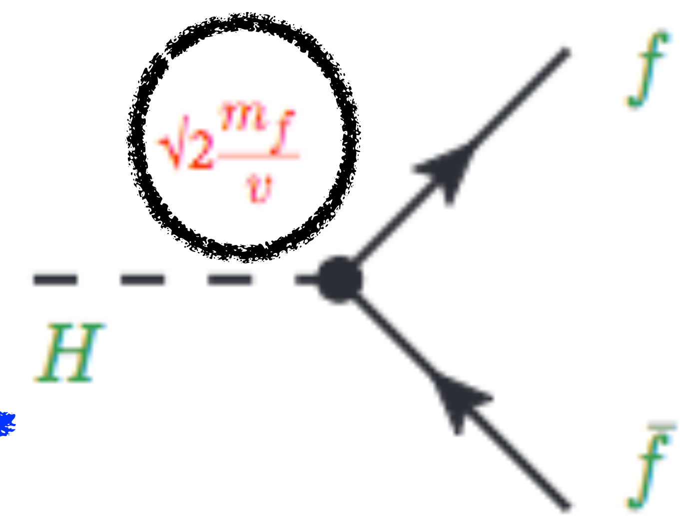
Yukawa couplings at the LHC

✓ Higgs boson Standard Model (SM) coupling to fermion proportional to $(\sqrt{2}m_f)/v$

- Yukawa's interactions in the SM → electroweak symmetry breaking mechanism: creation of fermion masses
- any significant deviation from the expected value would hint to New Physics

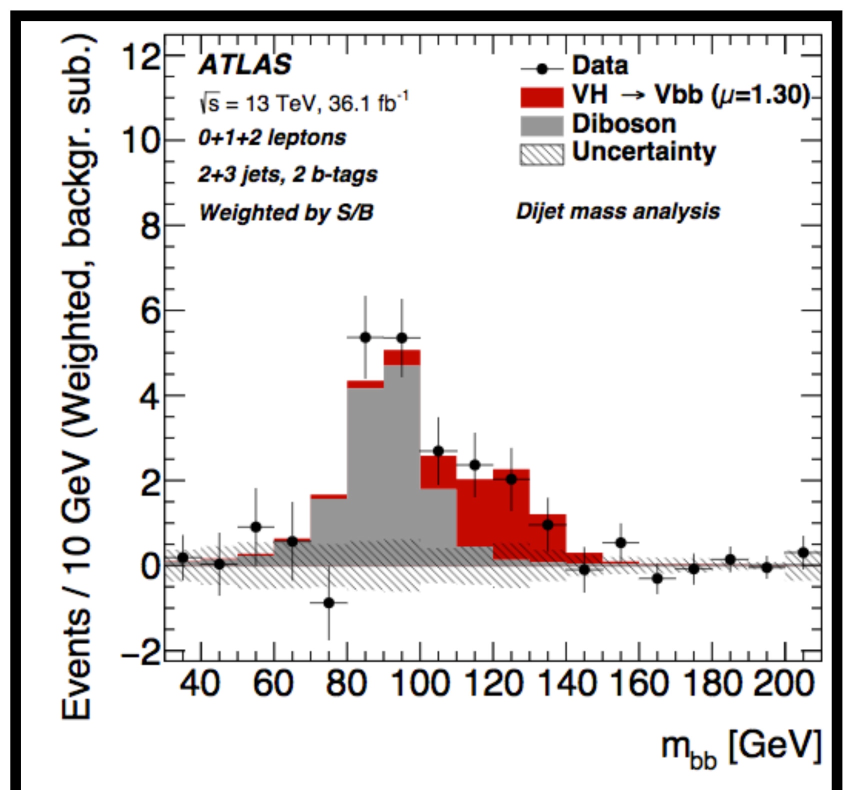
✓ Yukawa coupling ~ $m^{fermion}$

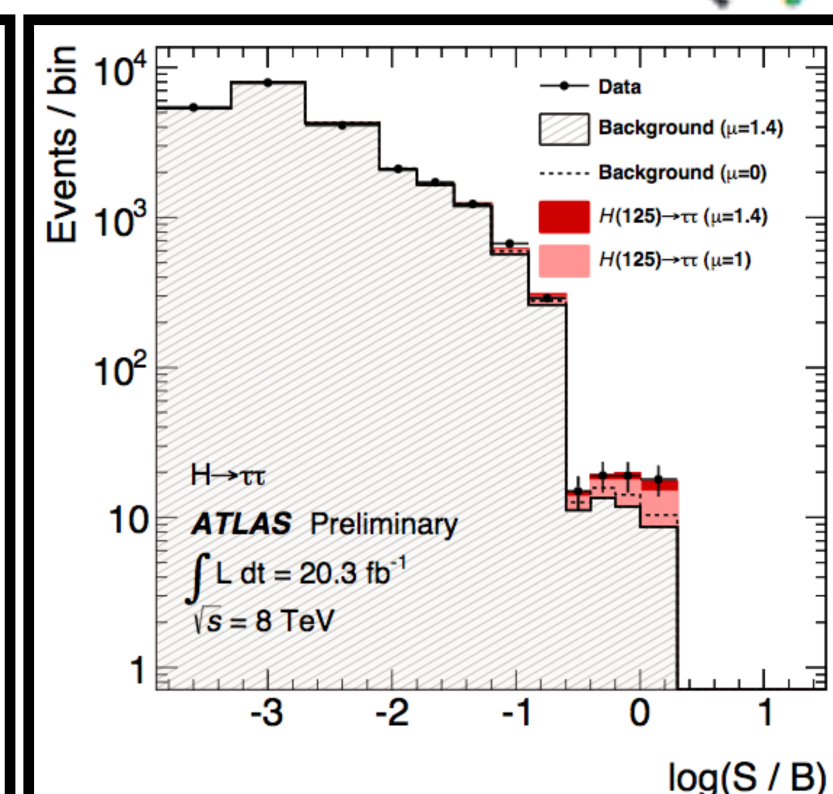
- largest Yukawa coupling (λ~1) for top as heaviest fermion



✓ Some Yukawa couplings already observed in Run 1 and Run 2

- H→ττ: tau Yukawa-couplings observed in Run 1 (4.1σ obs., 3.2σ exp.)
- H→bb: evidence of b-quark Yukawa couplings with Run 2 data (2015+2016)

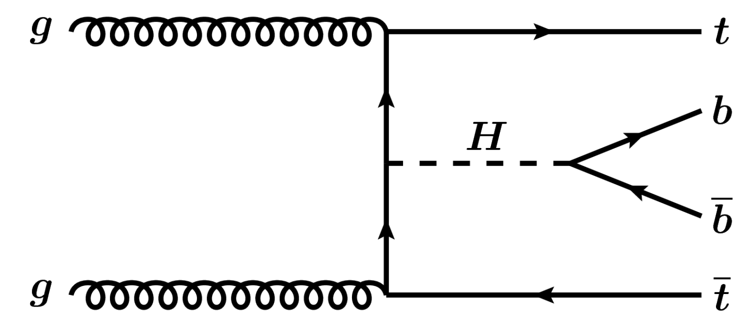


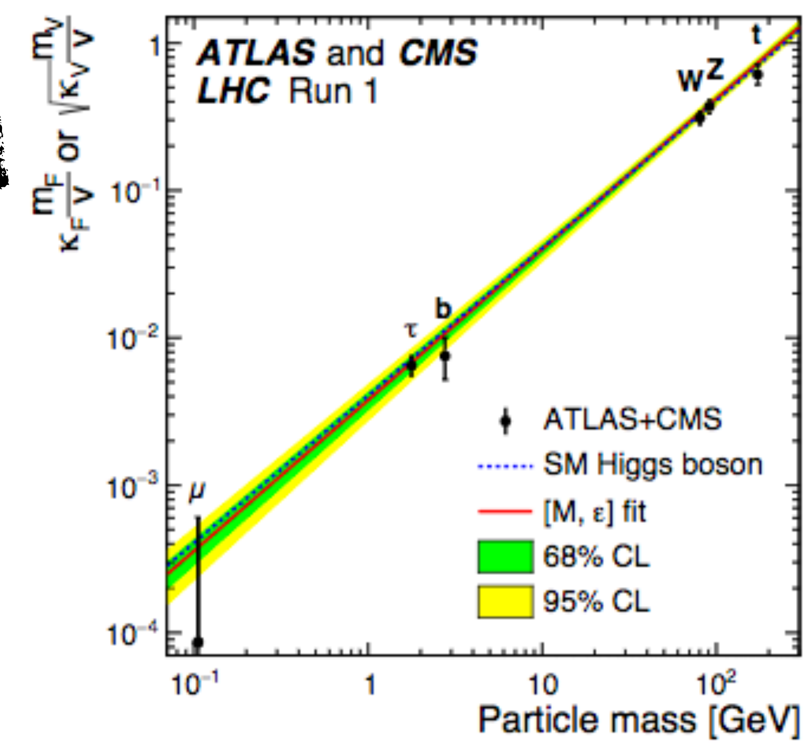


Top-quark Yukawa couplings at the LHC

✓ ttH provides a test to the measurement the top-quark Yukawa couplings (tree-level)

- cross-section proportional to λ^2
- drives perturbative calculations of SM properties for high energy scale





✓ Top-Yukawa couplings also extracted indirectly from main production mechanisms, gluon-gluon fusion and $H \rightarrow \gamma\gamma$ decay state

- Higgs couples preferentially with top and bottom quarks
- assuming no BSM contributions in the top loop

✓ Constraints on top-quark Yukawa couplings already provided with Run 1 measurements

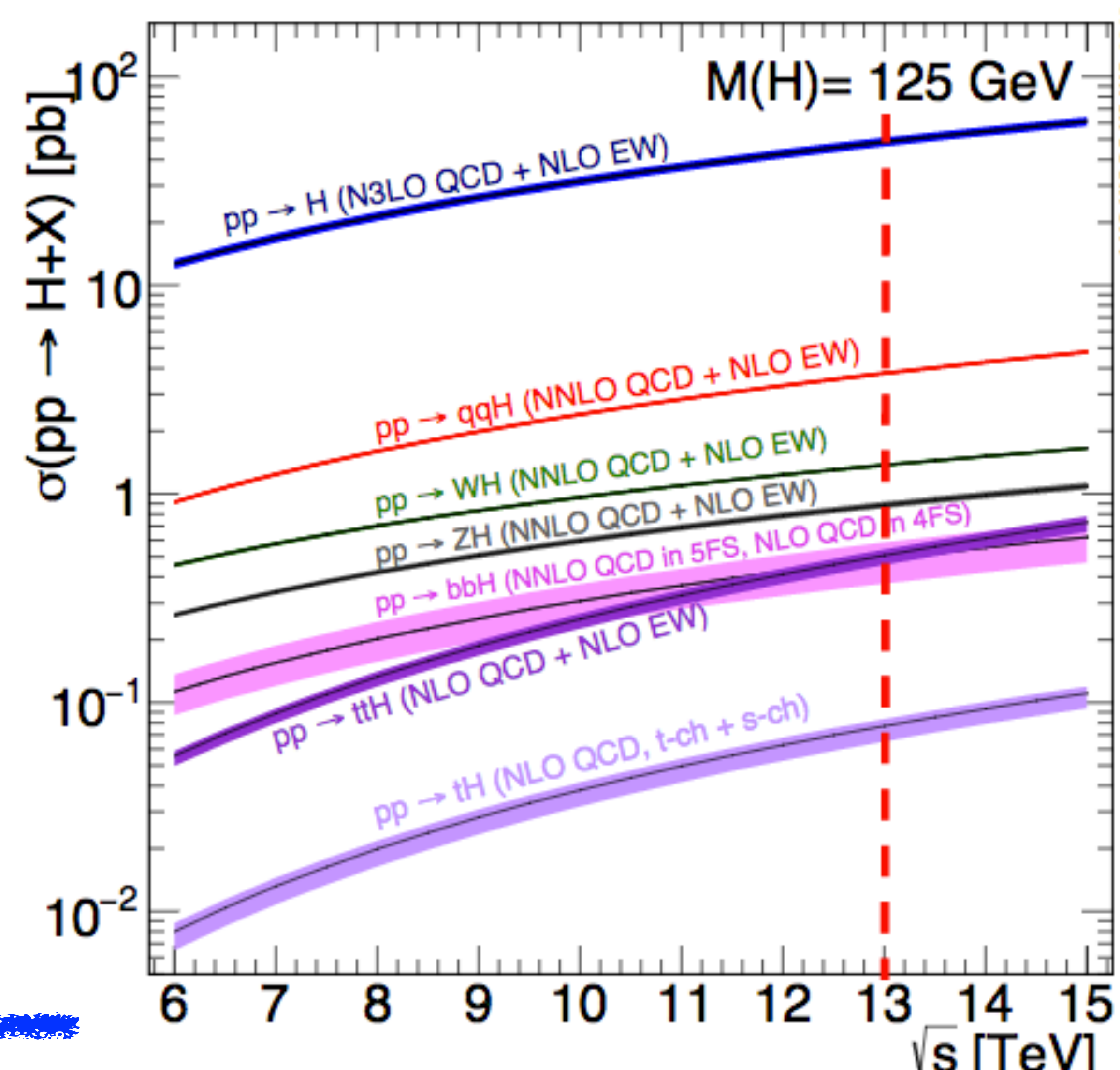
- ATLAS/CMS combination available

	ATLAS +CMS	ATLAS	CMS
Kt	0.87±0.15	0.98±0.20	0.77±0.20
μ (ttH)	2.3±0.7	1.9±0.8	2.9±1.0

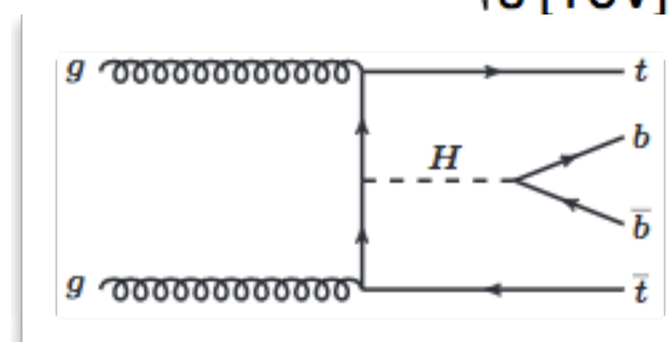
ttH production at LHC

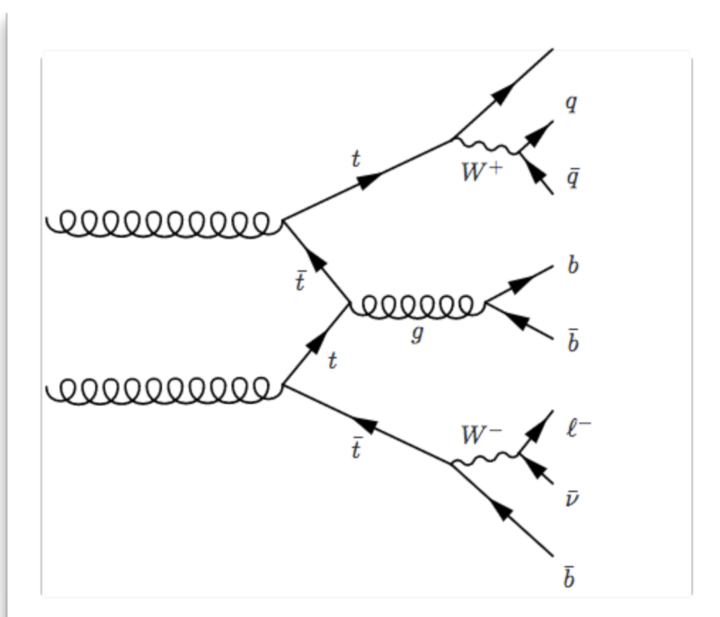
✓ ttH → very small production cross-secton at $\sqrt{s}$=13 TeV

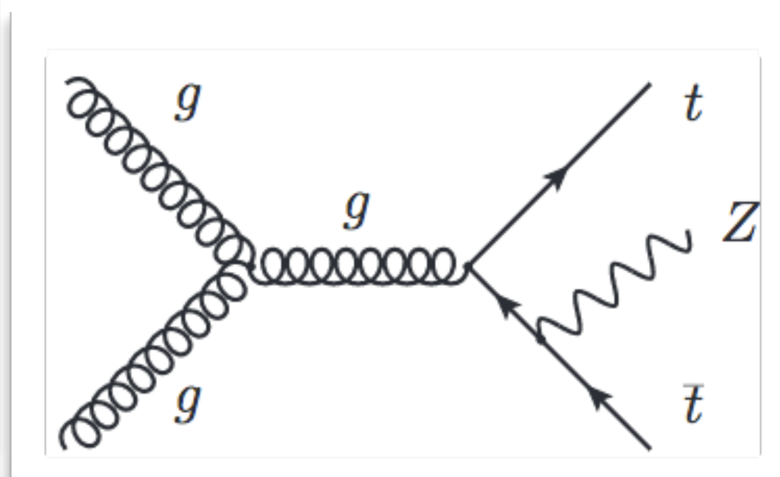
- σ=507 pb - 2% of the inclusive Higgs production at the LHC (pp→H+X)
- theoretical uncertainties on ttH signal cross section from QCD scales and choice of PDF (6% and 4% respectively)
- typical final states with jets, b-jets and lepton → large object multiplicity and challenging reconstruction techniques

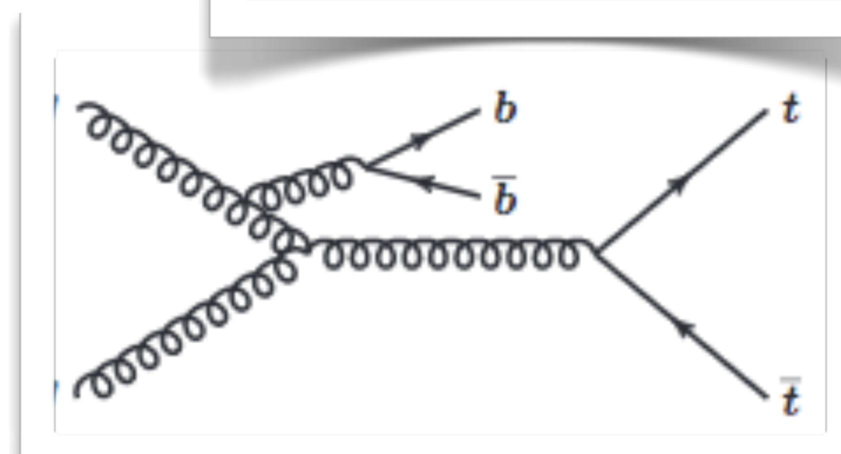


✓ Dominant background components to ttH signal process with very large yields and theoretical uncertainties ($\sigma^{tt+jets}$ =830 pb, σ^{tt+V} =0.8 pb)



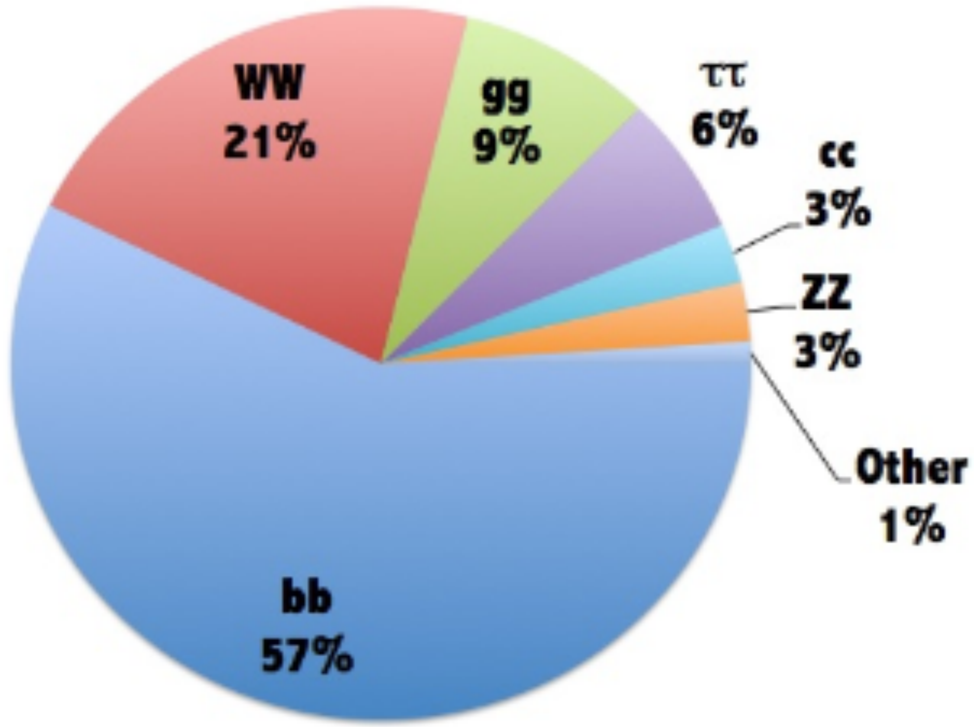




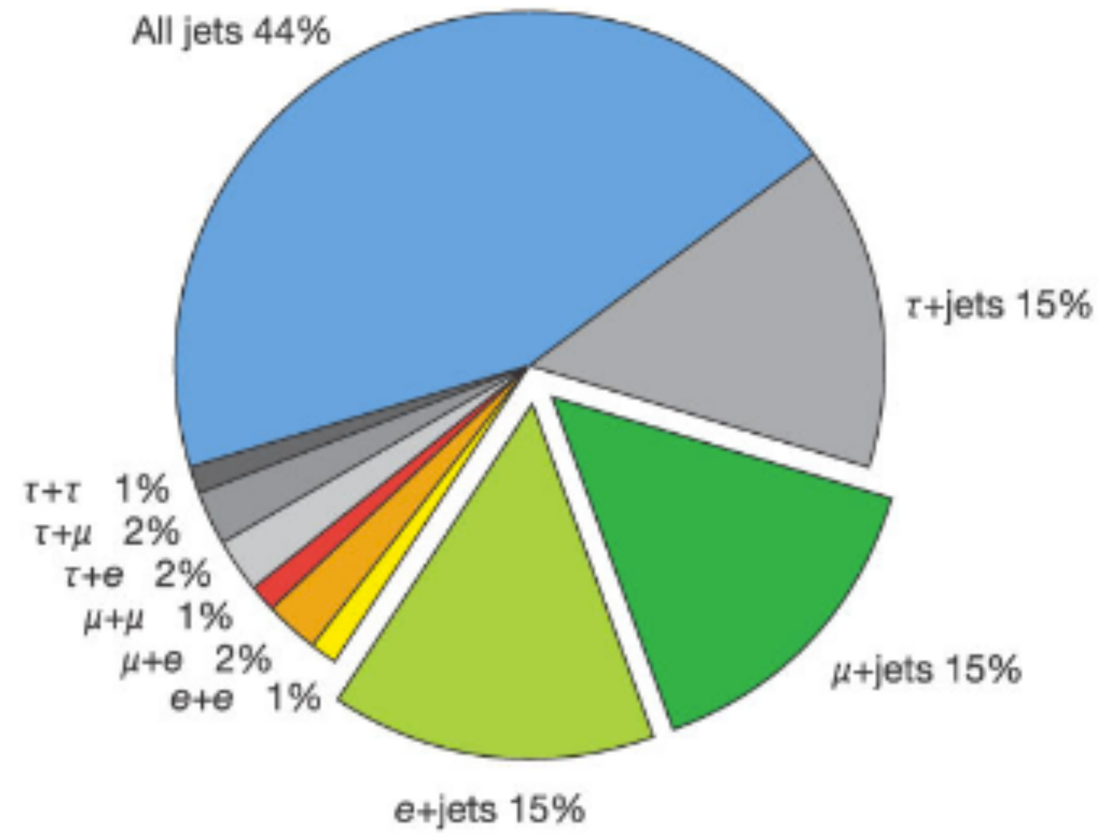


ttH channels and experimental signatures

Higgs boson decay modes for m_H =125 GeV



Decay modes of tt decay system

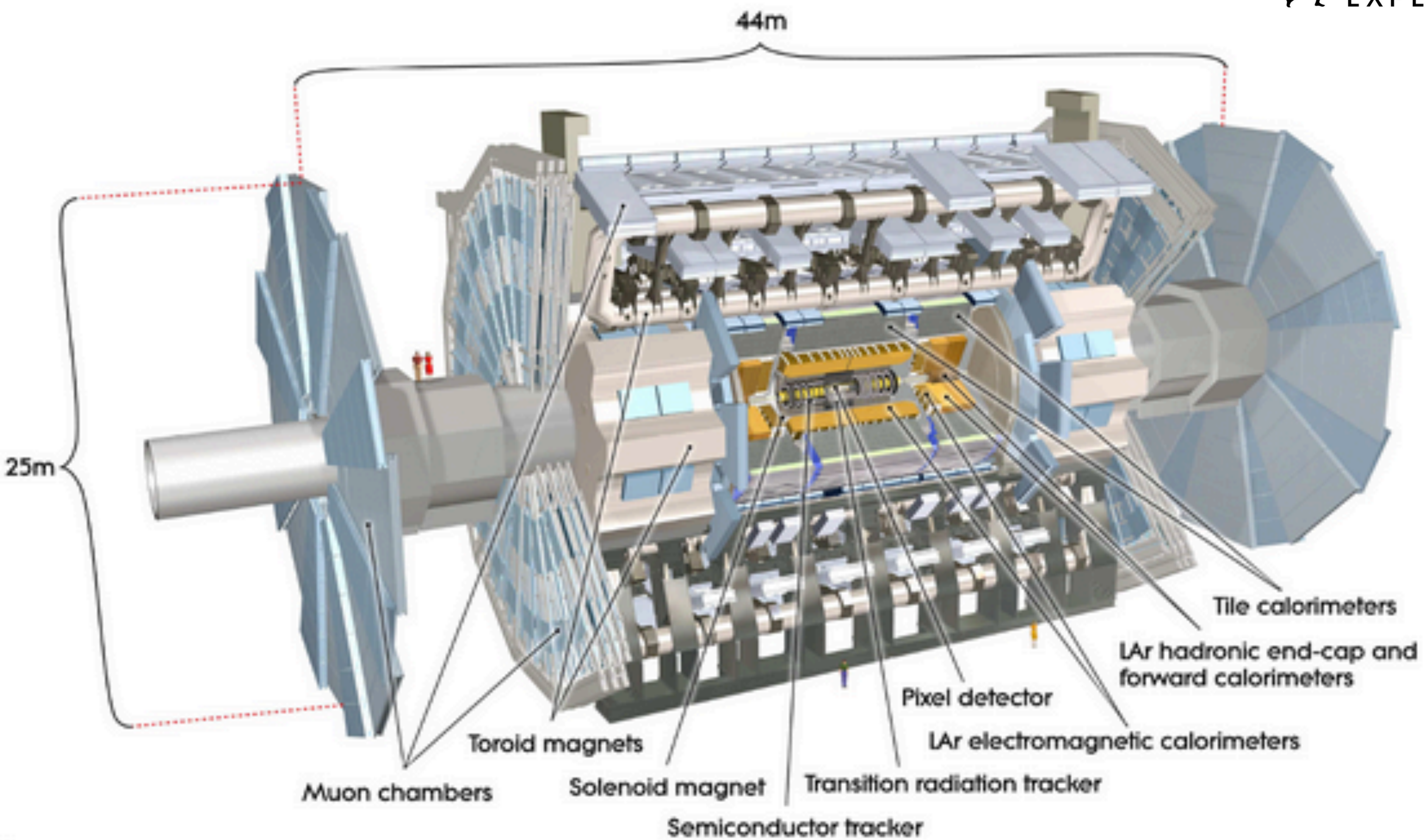


✓ ATLAS Run 2 ttH measurements performed by ATLAS with 2015+2016 data

Channel	H decay	tt decay	Journal paper
ttH (H→bb)	H→bb	1,2 leptons (e,μ)	arXiv:1712.08895 (PRD)
ttH (H→multi-lepton)	H→WW*/ZZ*/ττ	1,2 leptons (e,μ,τ)	arXiv:1712.08891 (PRD)
H→ZZ*→4l	H→ZZ*→4l	0,2 leptons (e,μ)	ATLAS-CONF-2017-043
H→γγ	H→γγ	0,2 leptons (e,μ)	ATLAS-CONF-2017-045
ttH combination			arXiv:1712.08895 (PRD)

✓ Focus on ATLAS ttH (H→bb) results in this seminar and quick wrap-up on ttH combination

The ATLAS experiment

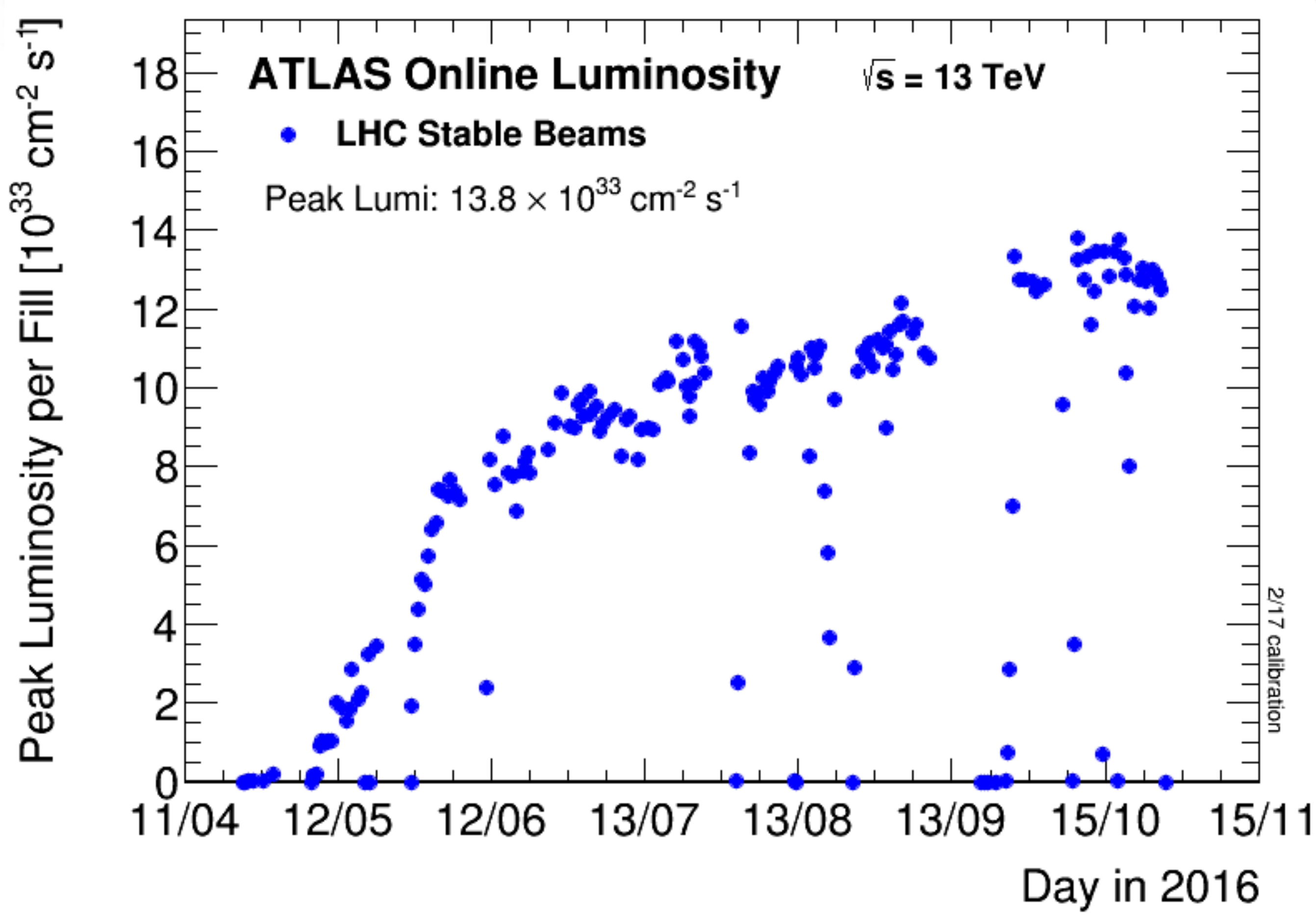


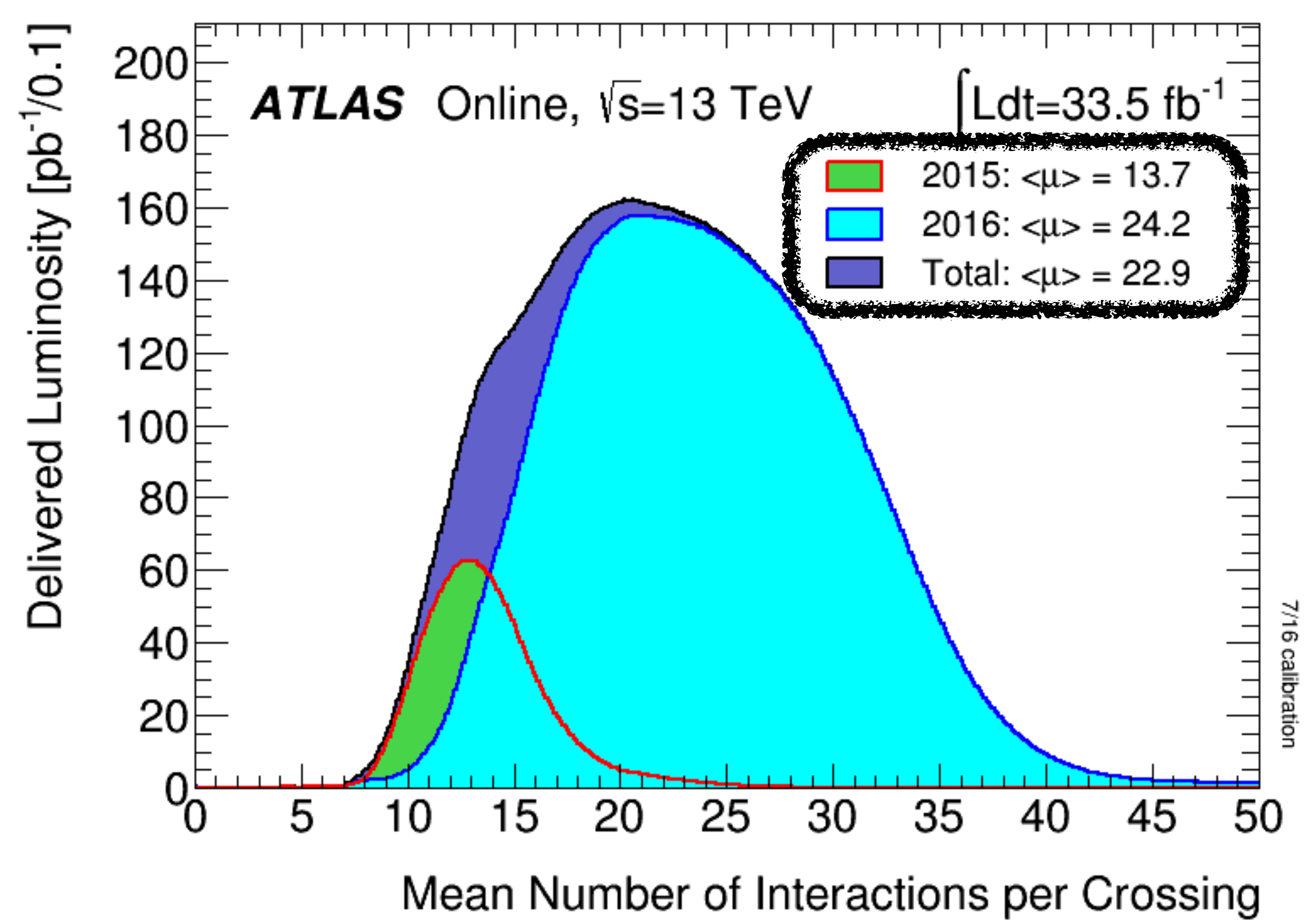
✓ Fundamental to fully exploit identification and reconstruction of final state particles, jets, b-jets, and leptons

- excellent understanding and optimization of performance of various sub-detectors (inner tracker, calorimeters and muon spectrometer)

Data-taking

- ✓ Excellent performance of the LHC accelerator and the ATLAS experiment for 2015-2016 data-taking
- ➡ Approximately 36.5 fb^{-1} pp data collected by ATLAS in 2015-2016 after data quality requirements
  - data recording efficiency over 93% in 2015 and 2016
  - mean number of interactions per bunch-crossing ⟨μ⟩ ~ 25
- ✓ A lot more data collected in pp collisions in 2017 (40 fb^{-1}) with higher peak luminosity and larger μ
- ✓ ttH Run 2 legacy result will target full 2015-2018 data statistics (~100 fb^{-1})
  - expected by late 2018/beginning 2019

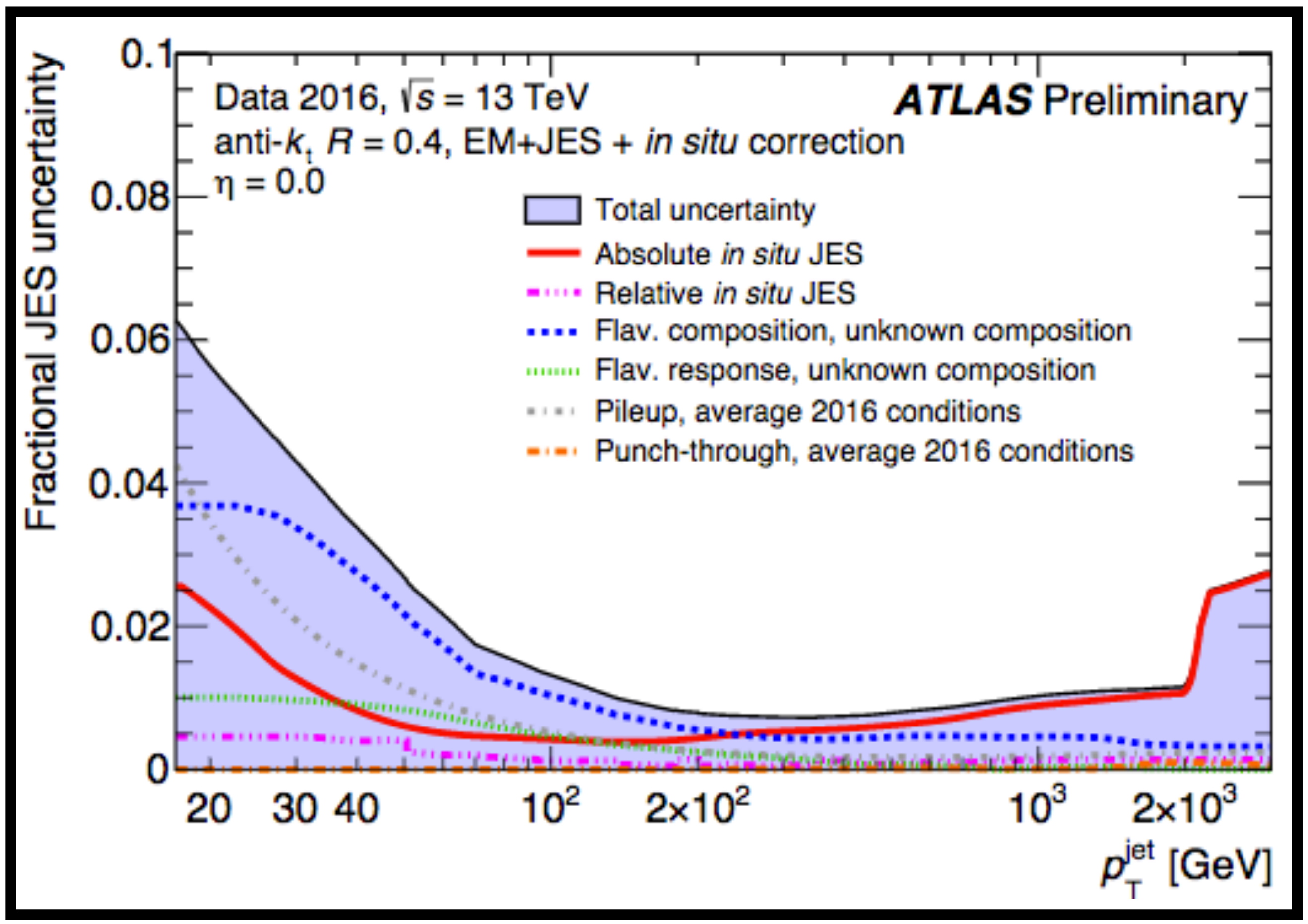




Performance of jets and leptons

✓ Small uncertainty on jet energy response and calibration

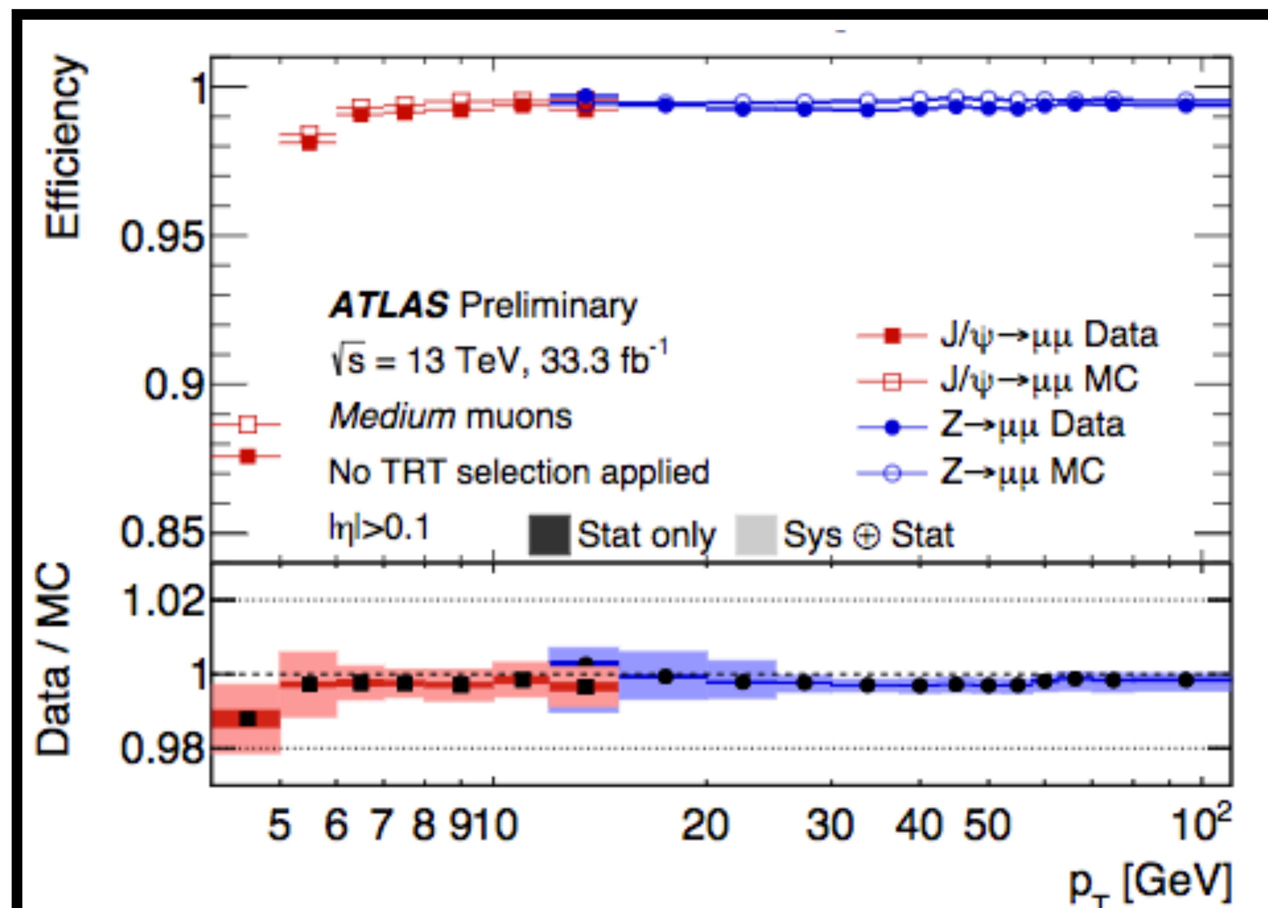
- jet energy scale and jet energy resolution uncertainties below 1% for pt>150 GeV and approximately 5-7% in the low pt region (pt<60 GeV)



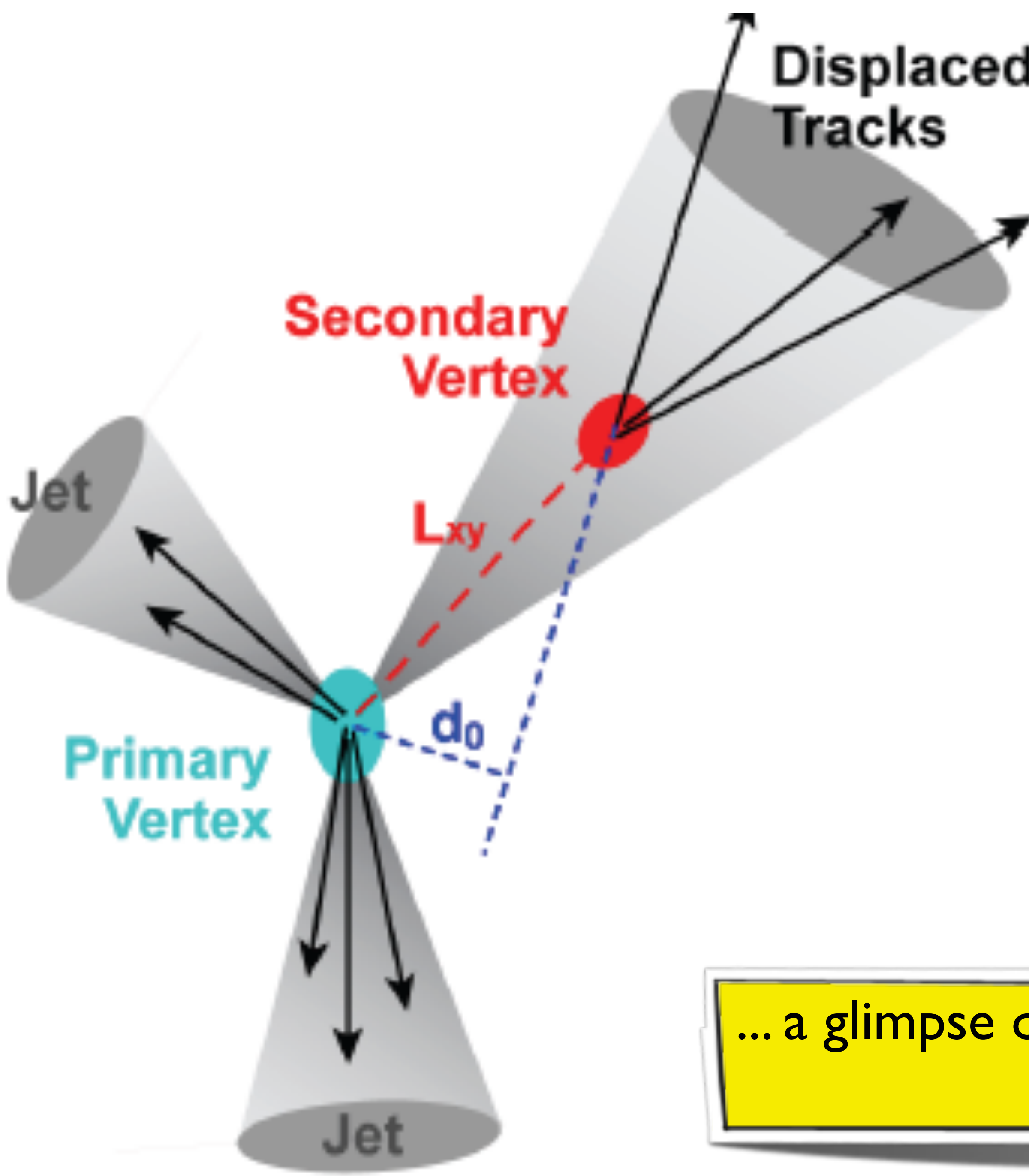
Very small experimental uncertainties arising from jet calibration (scales and resolution) affecting the measurement

✓ Lepton and muon performance

- very high lepton reconstruction and efficiency and negligible uncertainty
  - $Z \to ll$, $J/\psi \to ll$
- robust energy calibration for electrons (Z, J/ψ) and muons
- pile-up dependence of calibration found to be small



A quick dive into the b-tagging realm



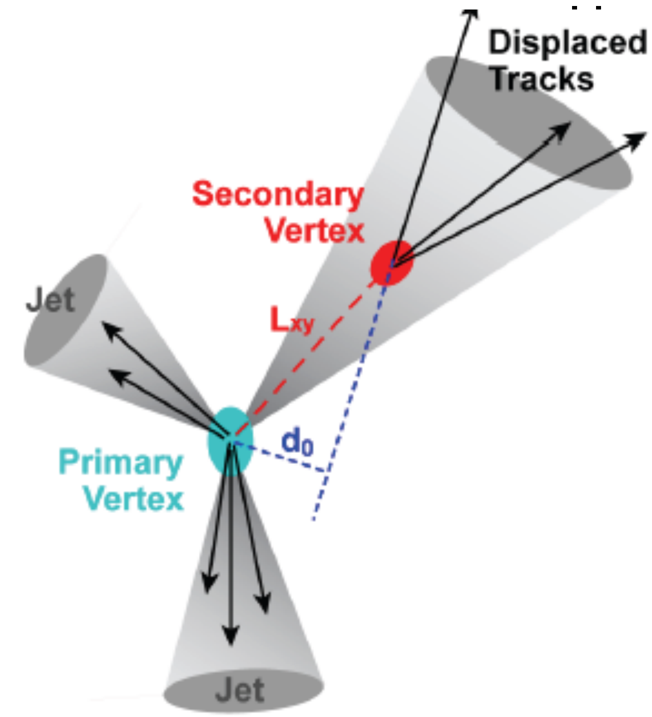
... a glimpse on b-tagging algorithm developments and performance...

ATL-PHYS-PUB-2016-012

ATL-PHYS-PUB-2017-003

ATL-PHYS-PUB-2017-012

B-tagging chain



Improvements of the existing tagger available

New features

IP based (IP2D;IP3D)

Inclusive secondary vertex (SV1)

Topological secondary & tertiary vertices (JetFitter)

NN based on track parameters (IPRNN)

Semi-leptonic decays to muons (SMT)

BDT (MV2)

a new NN (DL1)

Charm tagging

ttbar sample

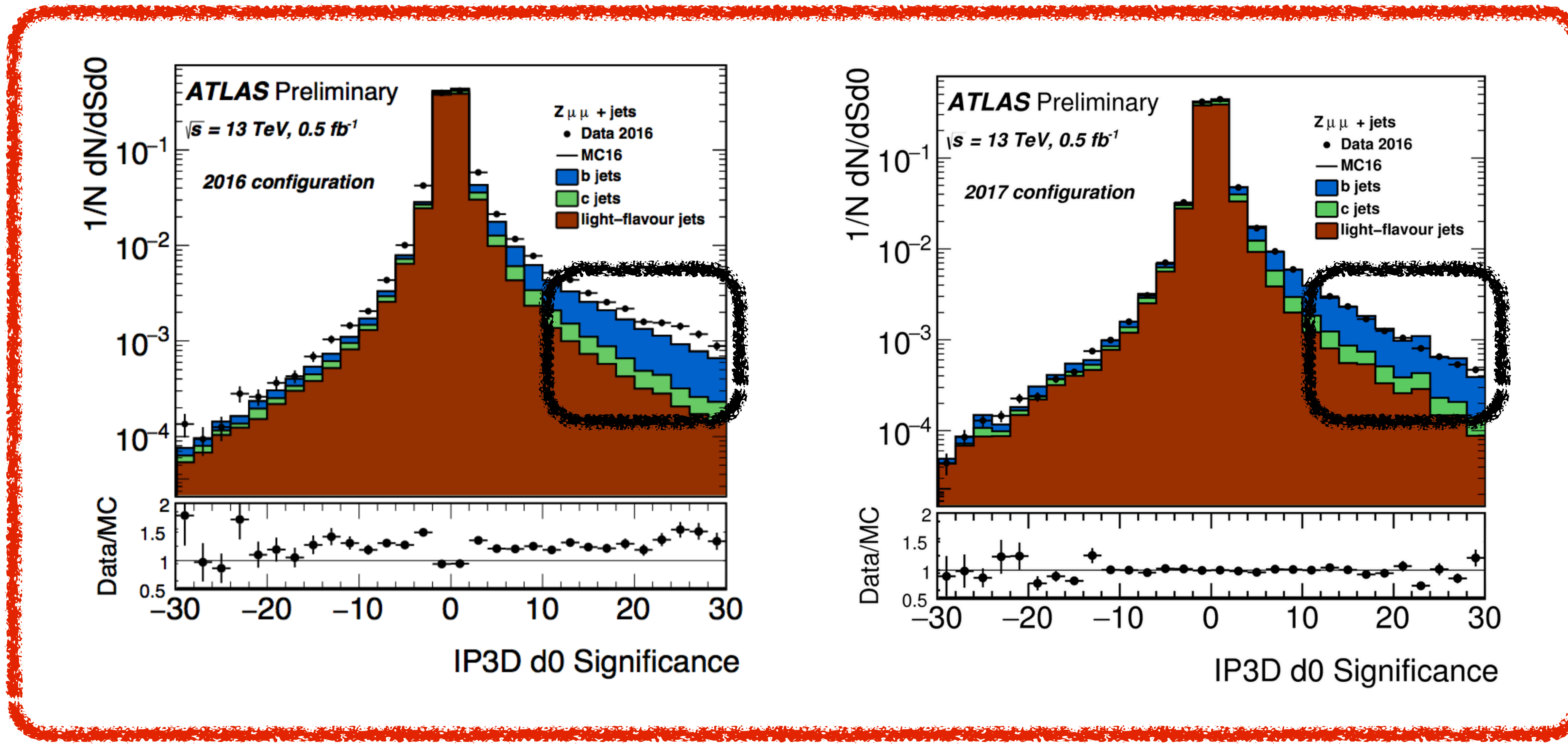
custom made Z' sample

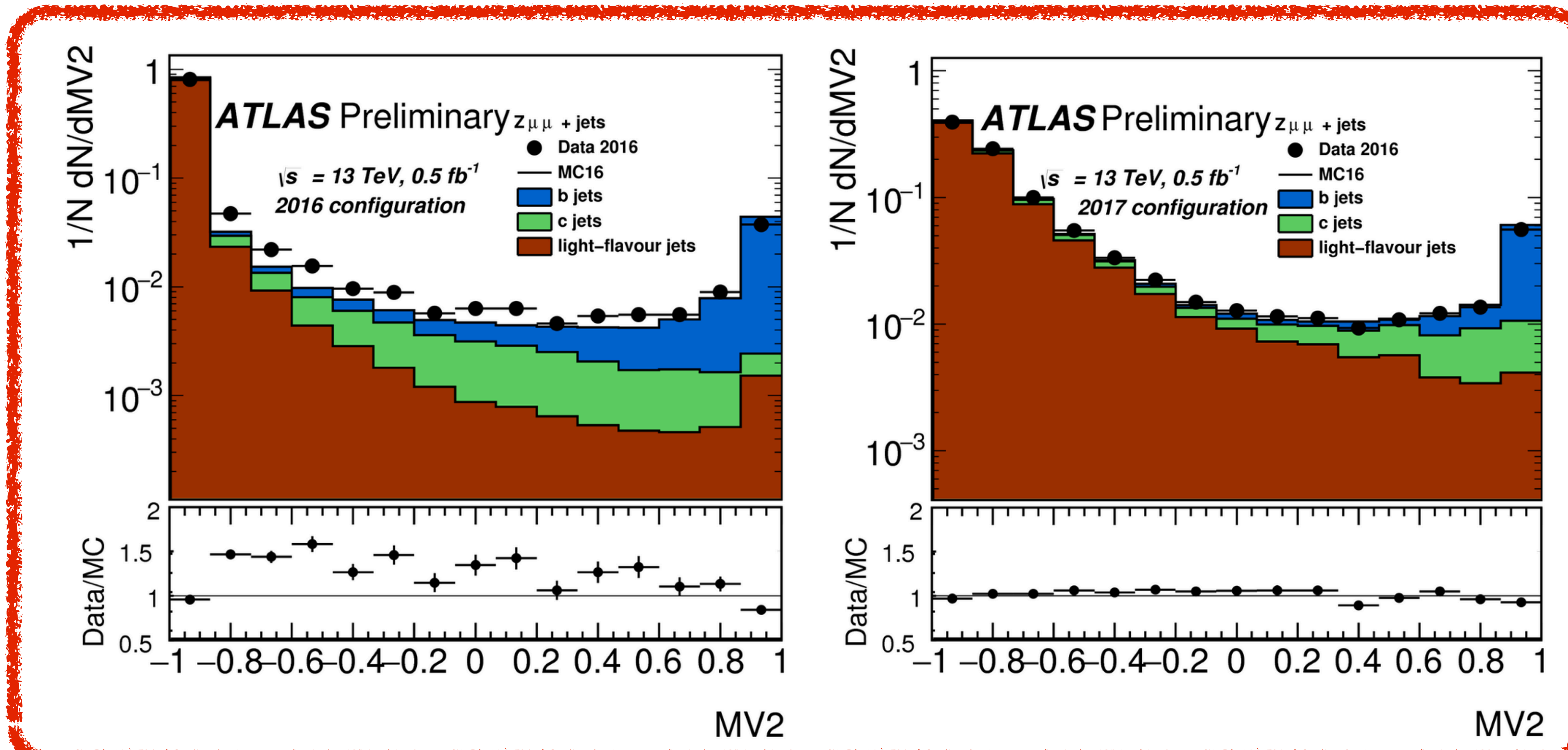
hybrid sample

Tracking

✓ Digitization model - uniform charge deposition along trajectory of ionizing particle → more realistic charge distribution exploited by the Bichsel model (point-like interactions)

- track-hit residual resolution smeared
- broader IP distributions in MC - better data/MC agreement
- studied on eμ enriched data sample and Zμμ +jets
- improvements in input d0 distributions and in the final MV2c10 tagger





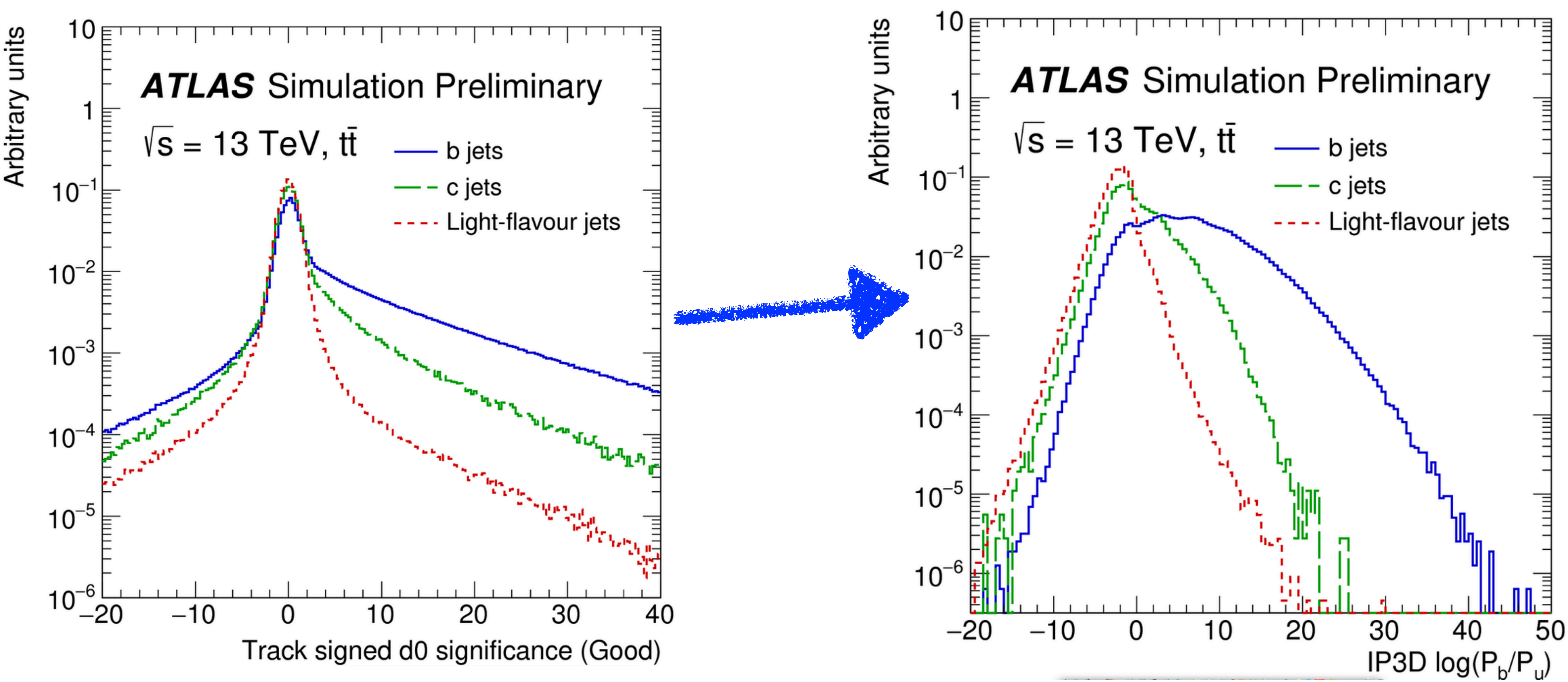
IPTag - track categorization

- Impact-parameter-based taggers make use of transverse and longitudinal impact parameter significance to separate b/c- and light-flavour hypotheses
- Tracks are required to have pt>1 GeV, |d0|<1mm, |z0 · sinθ|<1.5mm, 7 or more silicon hits, with at most 2 silicon holes, at most one of which is in the pixel detector
- Track categories are defined by silicon hit patterns on a reconstructed track such that quality criteria are used to build different templates for u/b/light d0/z0 significances

#	Category	Fractional contribution [%] *b*-jets	*c*-jets	light-jets
0	No hits in first two layers; expected hit in IBL and b-layer	1.9	2.0	1.9
1	No hits in first two layers; expected hit in IBL and no expected hit in b-layer	0.1	0.1	0.1
2	No hits in first two layers; no expected hit in IBL and expected hit in b-layer	0.04	0.04	0.04
3	No hits in first two layers; no expected hit in IBL and b-layer	0.03	0.03	0.03
4	No hit in IBL; expexcted hit in IBL	2.4	2.3	2.1
5	No hit in IBL; no expected hit in IBL	1.0	1.0	0.9
6	No hit in b-layer; expected hit in b-layer	0.5	0.5	0.5
7	No hit in b-layer; no expected hit in b-layer	2.4	2.4	2.2
8	Shared hit in both IBL and b-layer	0.01	0.01	0.03
9	At least one shared pixel hits	2.0	1.7	1.5
10	Two or more shared SCT hits	3.2	3.0	2.7
11	Split hits in both IBL and b-layer	1.0	0.87	0.6
12	Split pixel hit	1.8	1.4	0.9
13	Good	83.6	84.8	86.4

IPTag - outputs and results

- ✓ Log likelihood ratio computed as the sum of per-track contributions starting from p_b and p_u templates for b and light flavour jet hypotheses respectively
    - Assuming no correlations within the various components contributing to the sum of all tracks
- ✓ In addition to the LLR separating b vs light, LLR functions are also computed to separate b- vs c- and c- vs light-flavour jets



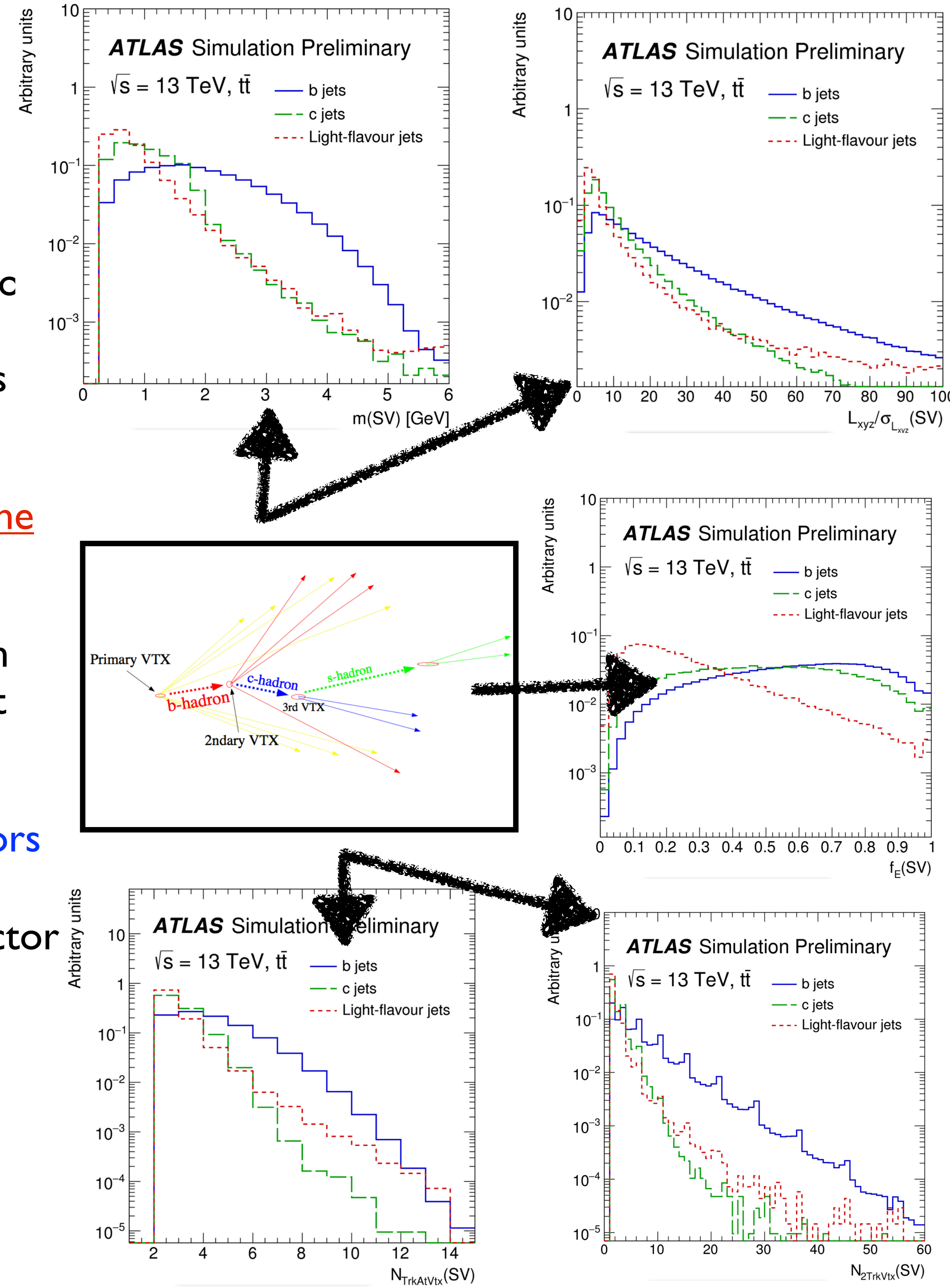
Secondary vertex reconstruction

✓ Look for secondary vertices to identify b-jets

- reconstructing two-track vertices
- tracks are rejected if SV likely to originate frc the decay of a long-lived particle (Ks, Λ), photon conversions or hadronic interactions with material

✓ Extra track-requirements are used to improve the performance for the 2017 LHC run

- tracks ordered in pt - at most 25 tracks with largest pt (against fragmentation at high b-jet pt)
- minimal number of hits in the silicon detectors increased by one for tracks with |η|>1.5 (improved track quality and amount of detector material mitigated)
- tracks with low Sd0 and high Sz0 removed (impact of pileup leading to fake vertices reduced)



The Soft Muon Tagger

- ✓ Muons from semileptonic b-decays within jets
  - Background sources in light-jets that produce a muon candidates are: prompt muons from W randomly matched to light-jets, muons from decay in flight of light hadrons, punch-through muons
- ✓ Input variables separating those muons from those of b/c-hadron decays: ptRel, ΔR, d0
- ✓ Additional input variables defining the quality of the muon track
- ➡ Information included in BDT as input to MV2

$$\text{MIS} = \frac{p_{ID} - p_{MS}^{extr}}{\sigma_{E_{loss}}}$$

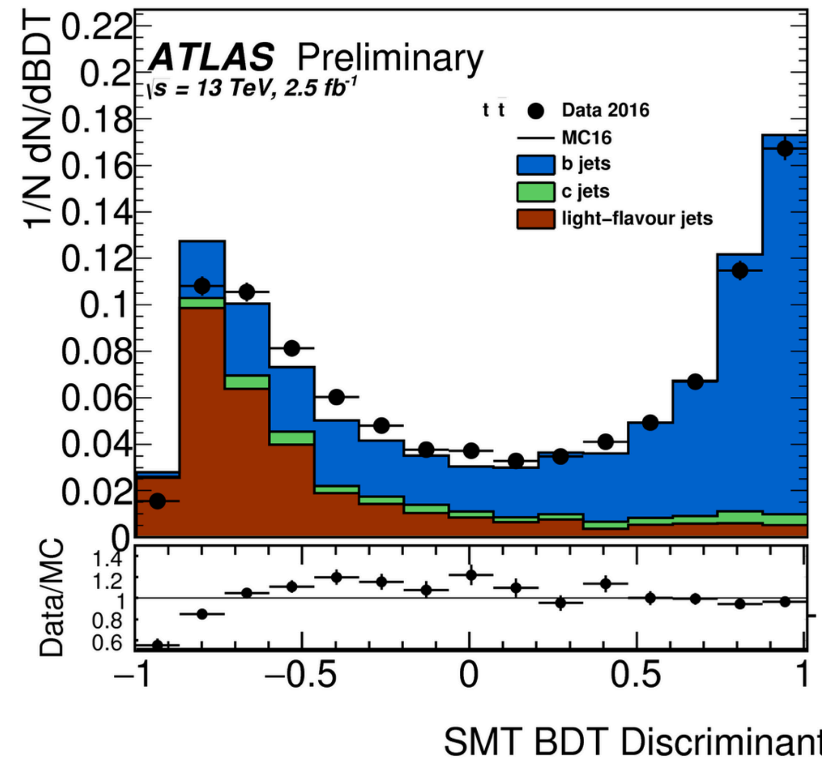
$$\Delta R(\mu, \text{jet})$$

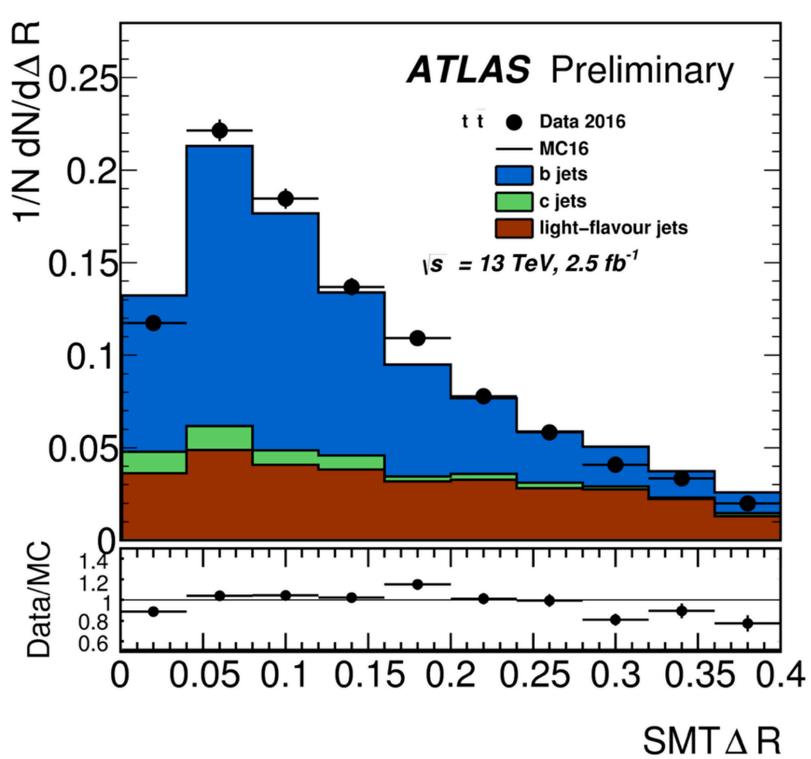
$$p_T^{rel}$$

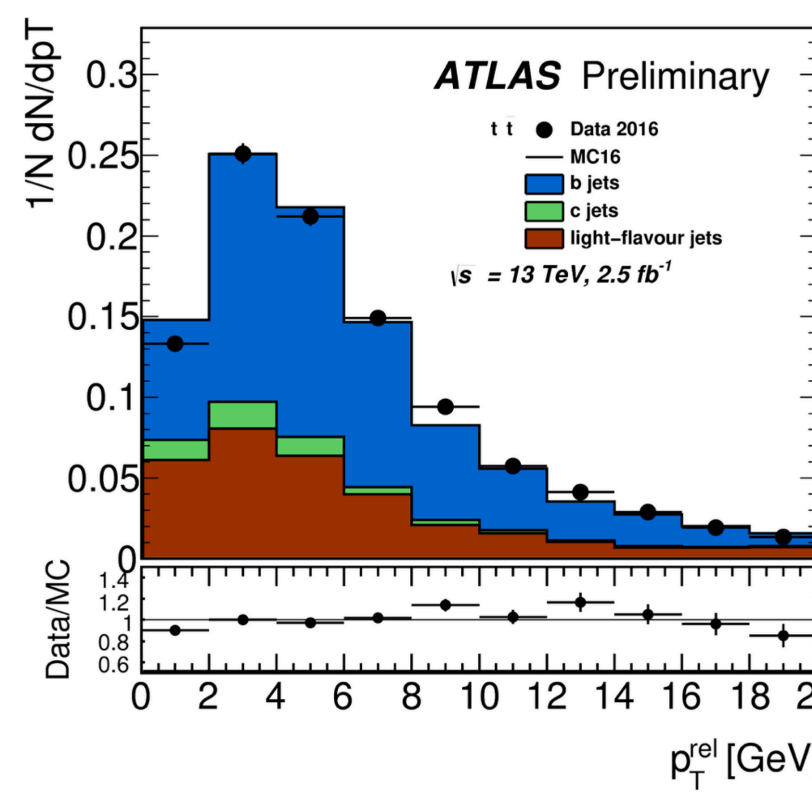
$$\text{SNS} = q \times \sum_i \frac{\Delta\phi^i_{scatt}}{\sigma_{\Delta\phi^i_{scatt}}}$$

$$q \text{ over } p \text{ ratio} = \frac{(q/p)_{ID}}{(q/p)_{MS}}$$

$$d_0^{PV}$$

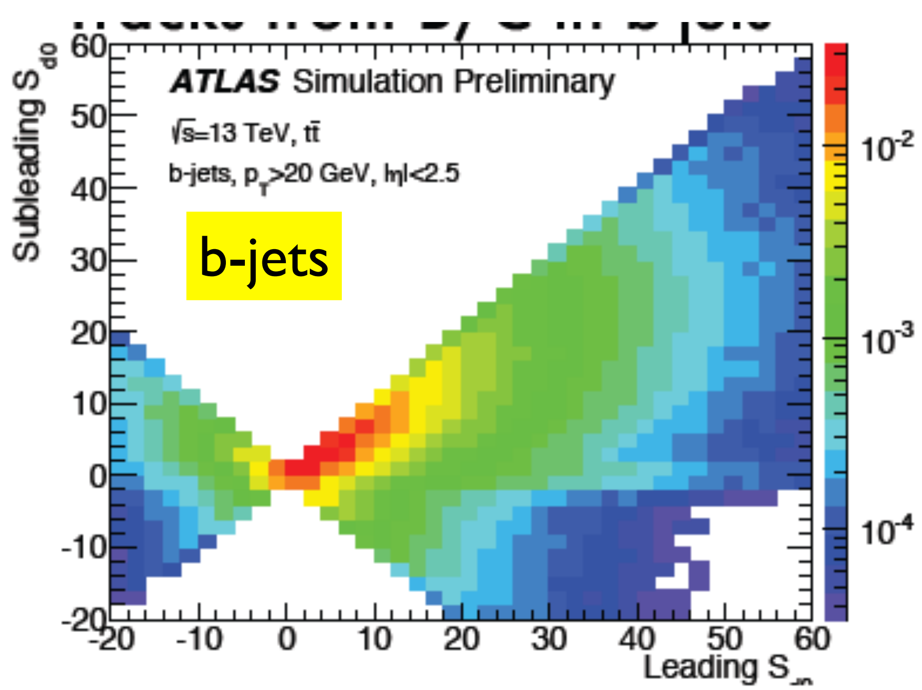


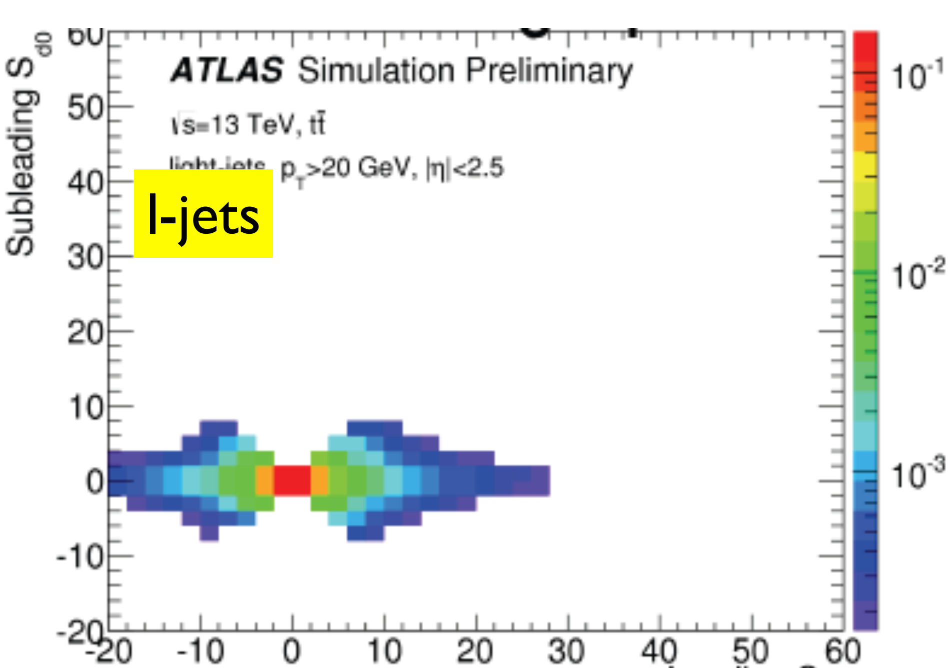




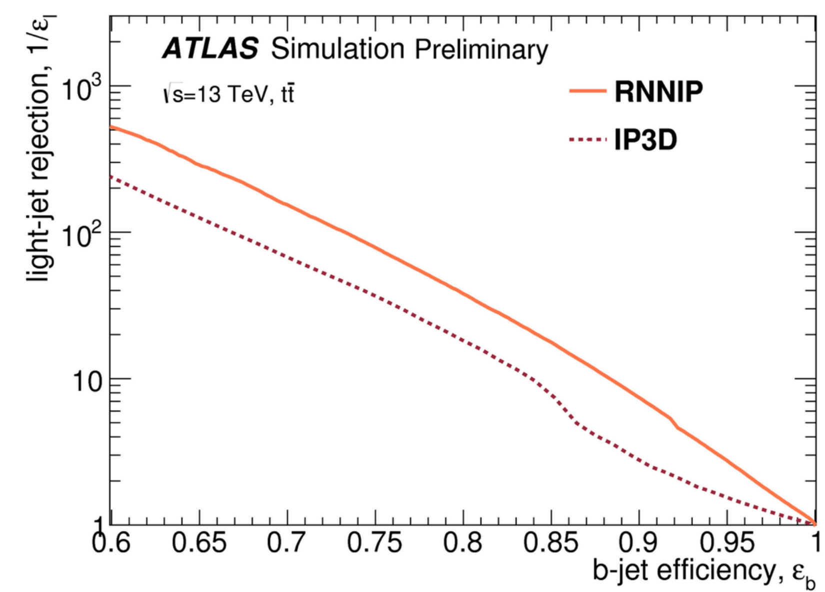
Impact-parameter taggers with recurrent NN

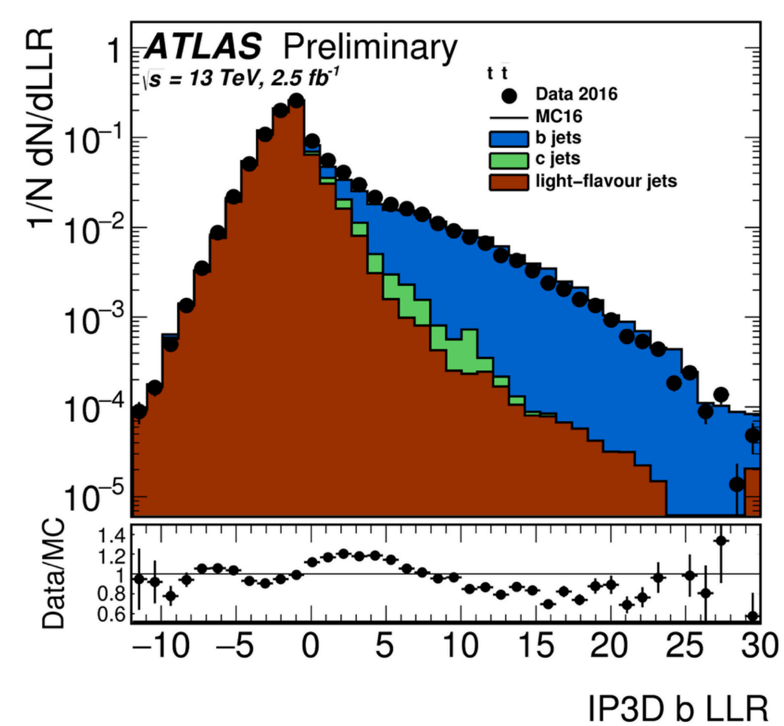
- ✓ Correlation between tracks associated to jets exploited with modern NN techniques (Recurrent Neural Network tagger)
- ✓ IP3D → properties of tracks are treated as independent and the template PDF's in different hit categories are built neglecting track-to-track correlations - complementarity with IPTag investigated
- ✓ Sequential dependencies between discriminating variables used for full characterization of properties of b-jets





Track Variable
S_{d_0}
S_{z_0}
Category
p_T^{frac}
ΔR(track, jet)

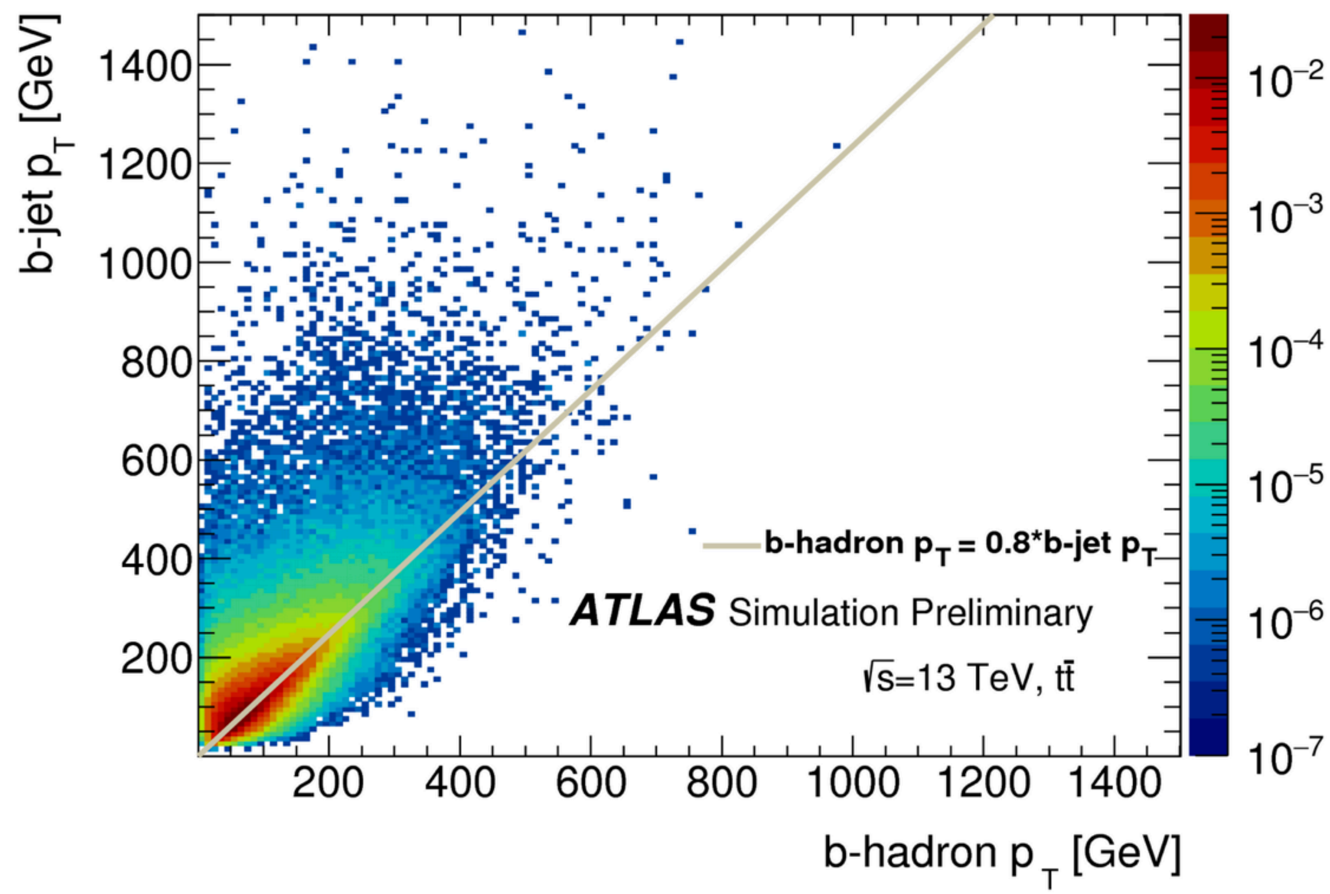


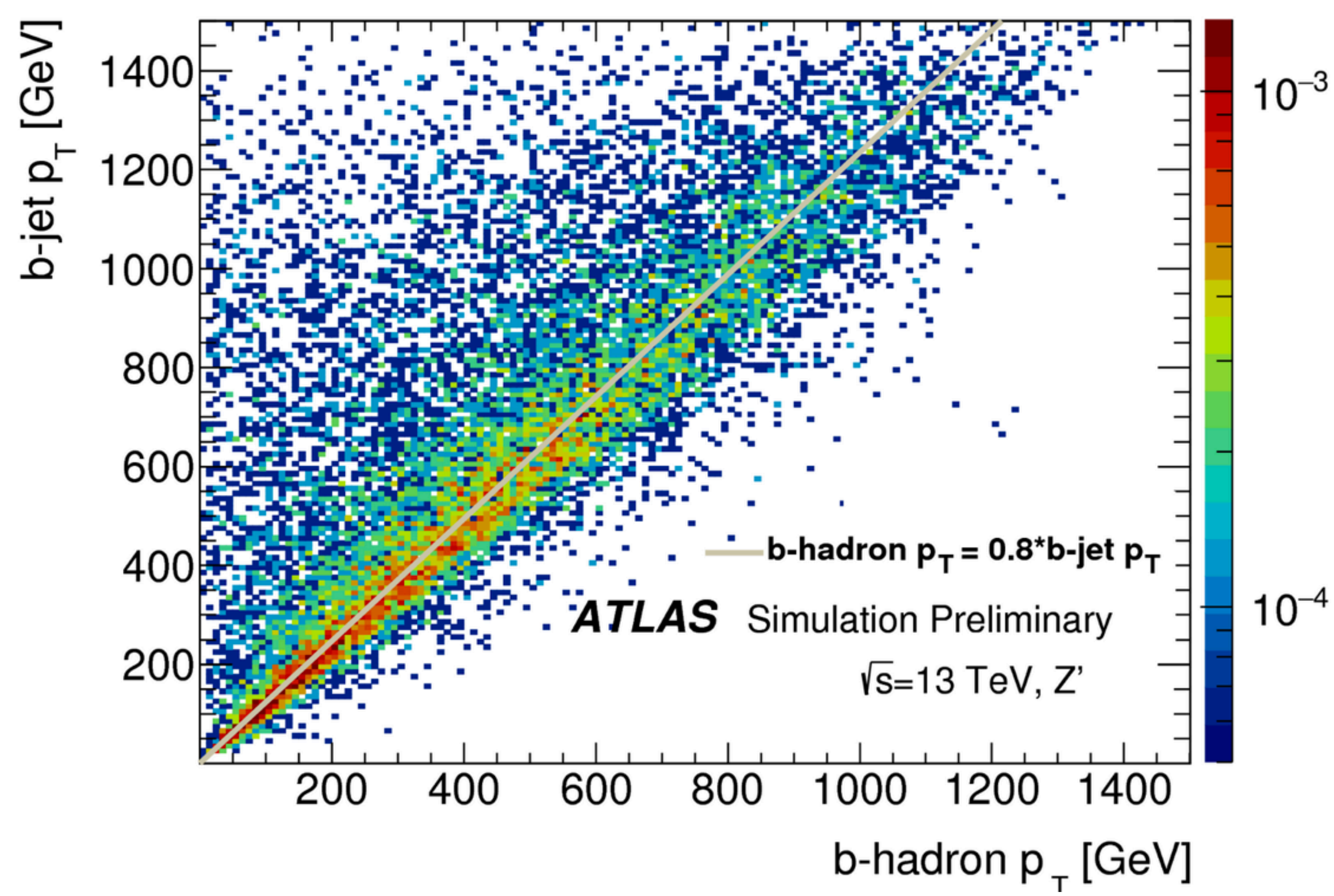


Algorithm training samples

✓ Studied b-hadron pt vs b-jet pt correlation in ttbar and broad Z' sample

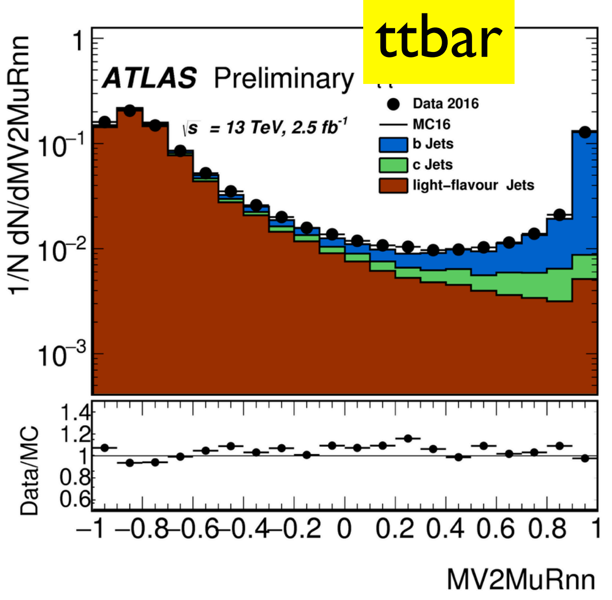
- ttbar sample looses correlation above mT, while Z' fully characterizes the high pt phase space

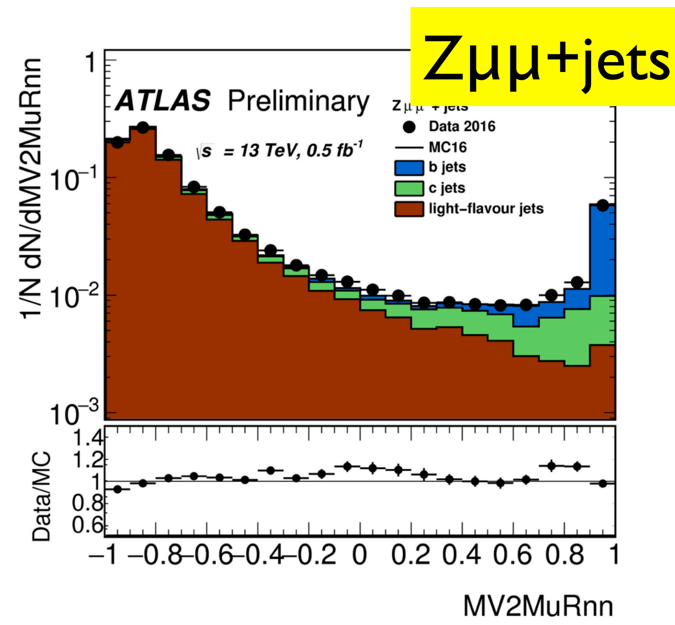


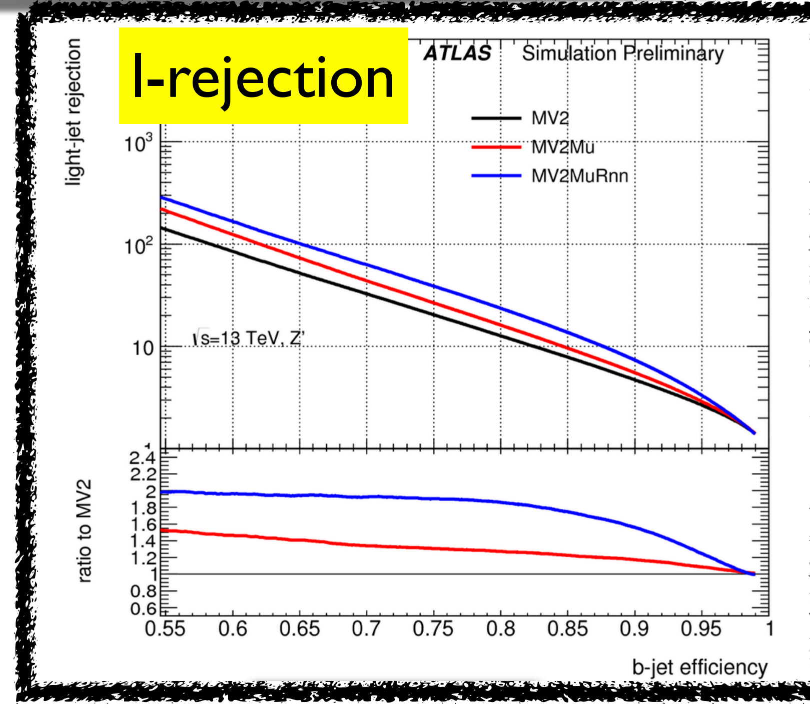


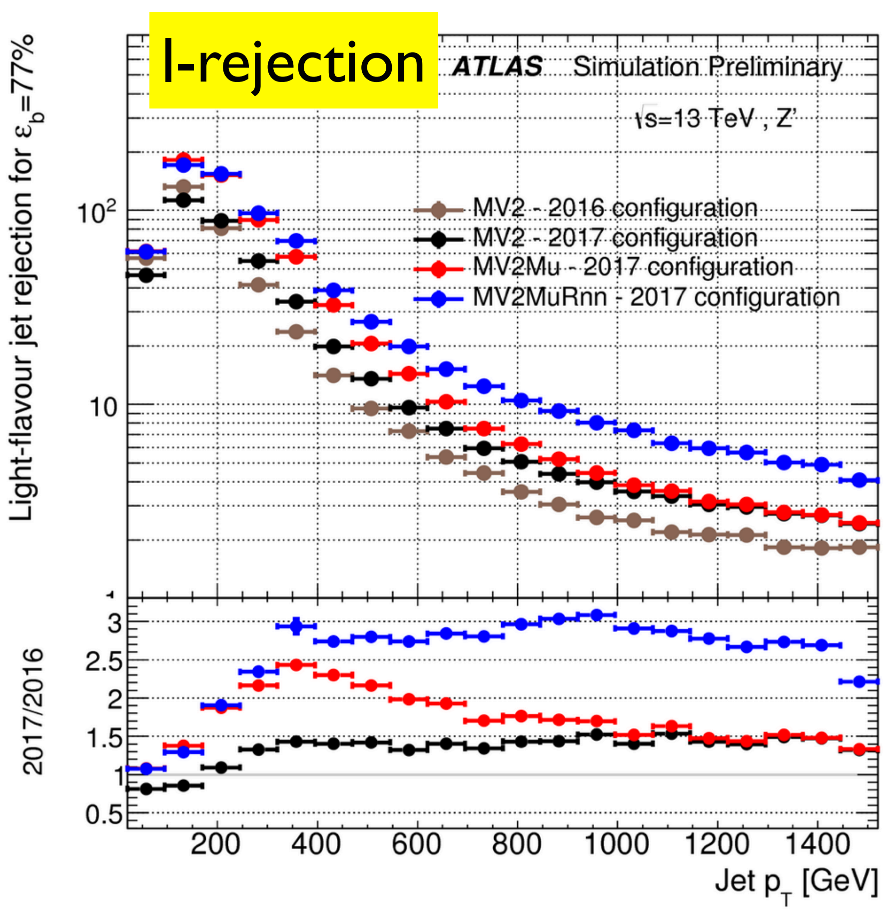
Algorithm performance

➡ MV2 - std MV2 inputs as in r20.7, MV2Mu - std MV2 inputs + SMT, MV2MuRnn - std MV2 inputs + SMT+ RNN

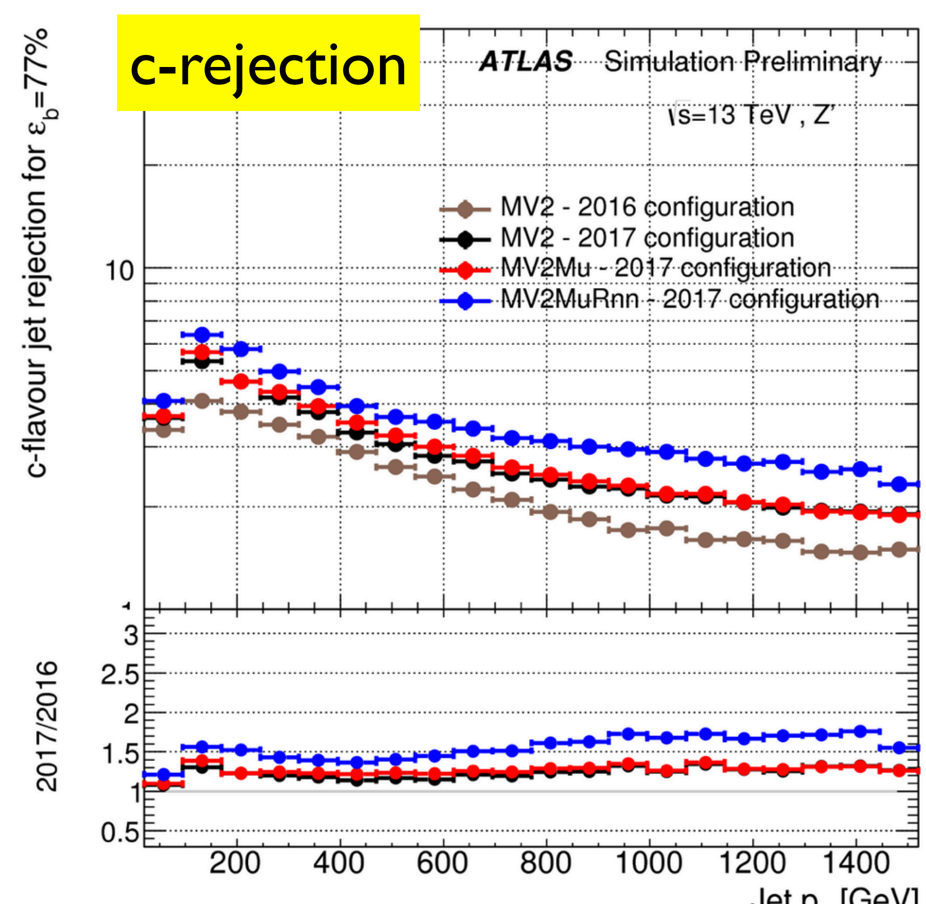






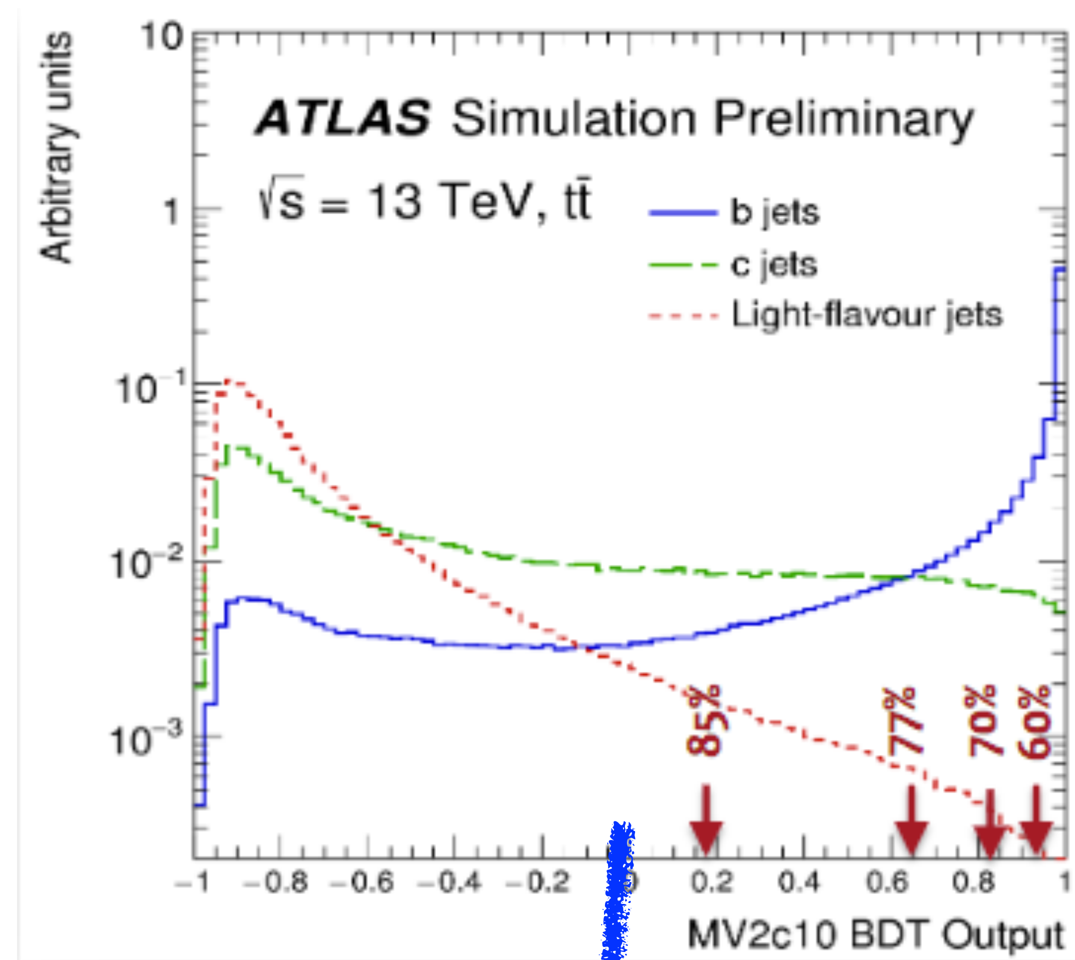


Improvements on the full pt spectrum from various low-level tagger contributions (SMT at low-medium pt, RNN at high pt)



How b-tagging is used

- ✓ b-jet identification exploited with MV2c10 multivariate discriminant
  - b vs c and b vs light separation
  - input variables accounting for track impact parameters, displaced secondary vertices, decay chain of B-hadrons and kinematics properties of the final state
  - 5 b-tagging response according to b-jet efficiency
- ✓ Analysis categories based on MV2 response for jets being tagged
- ✓ Correction factors to MV2 response extracted on data (calibration) for three jet flavours (b-, c- and light-flavour jets)

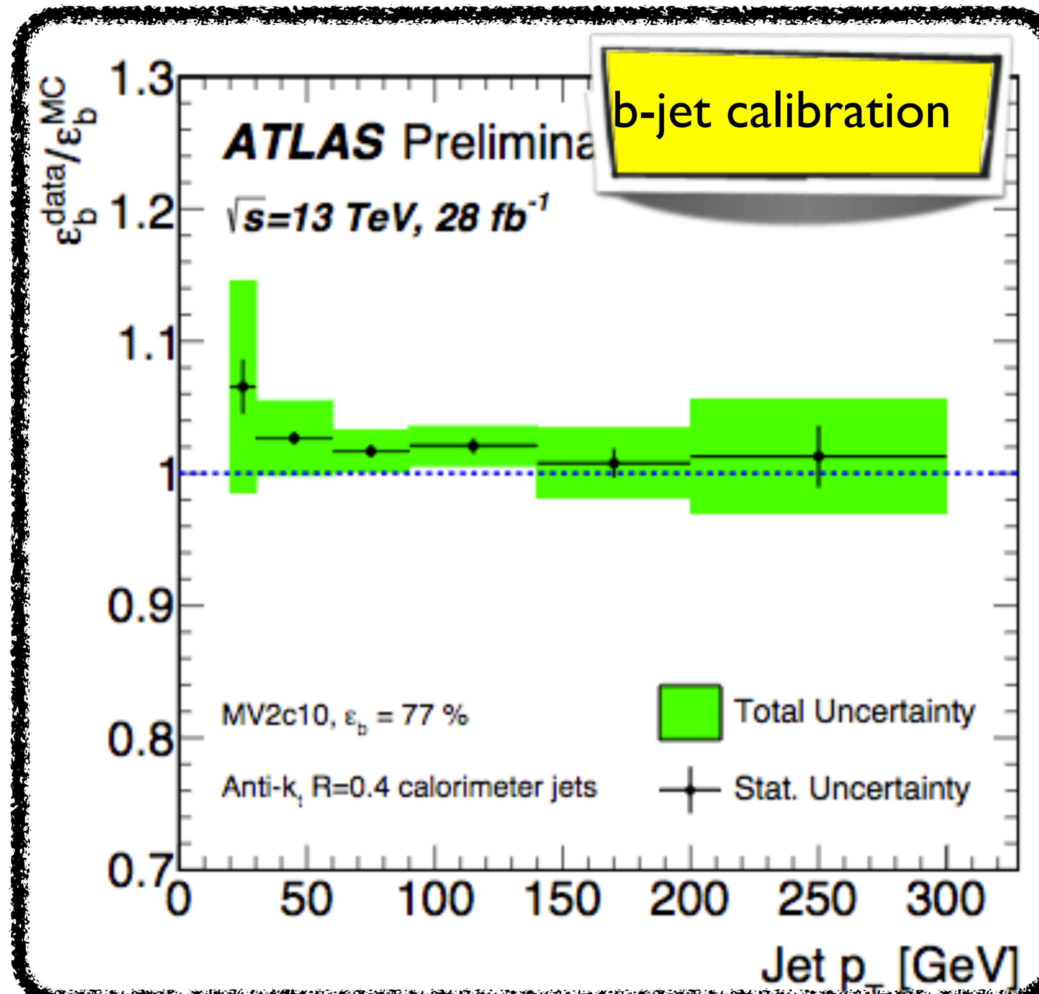


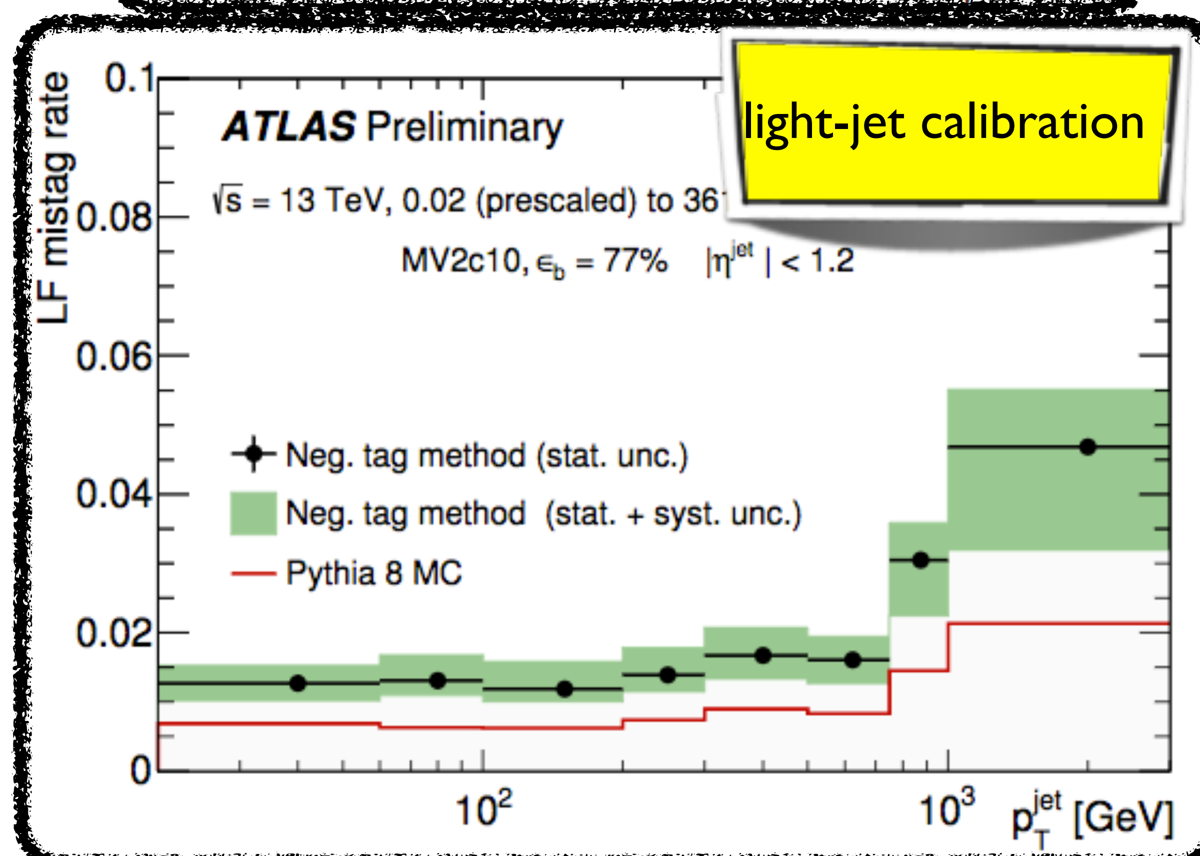
	None	Loose	Medium	Tight	Very tight
score	1	2	3	4	5
$\varepsilon(b)$	100%	85%	77%	70%	60%

b-tagging calibration

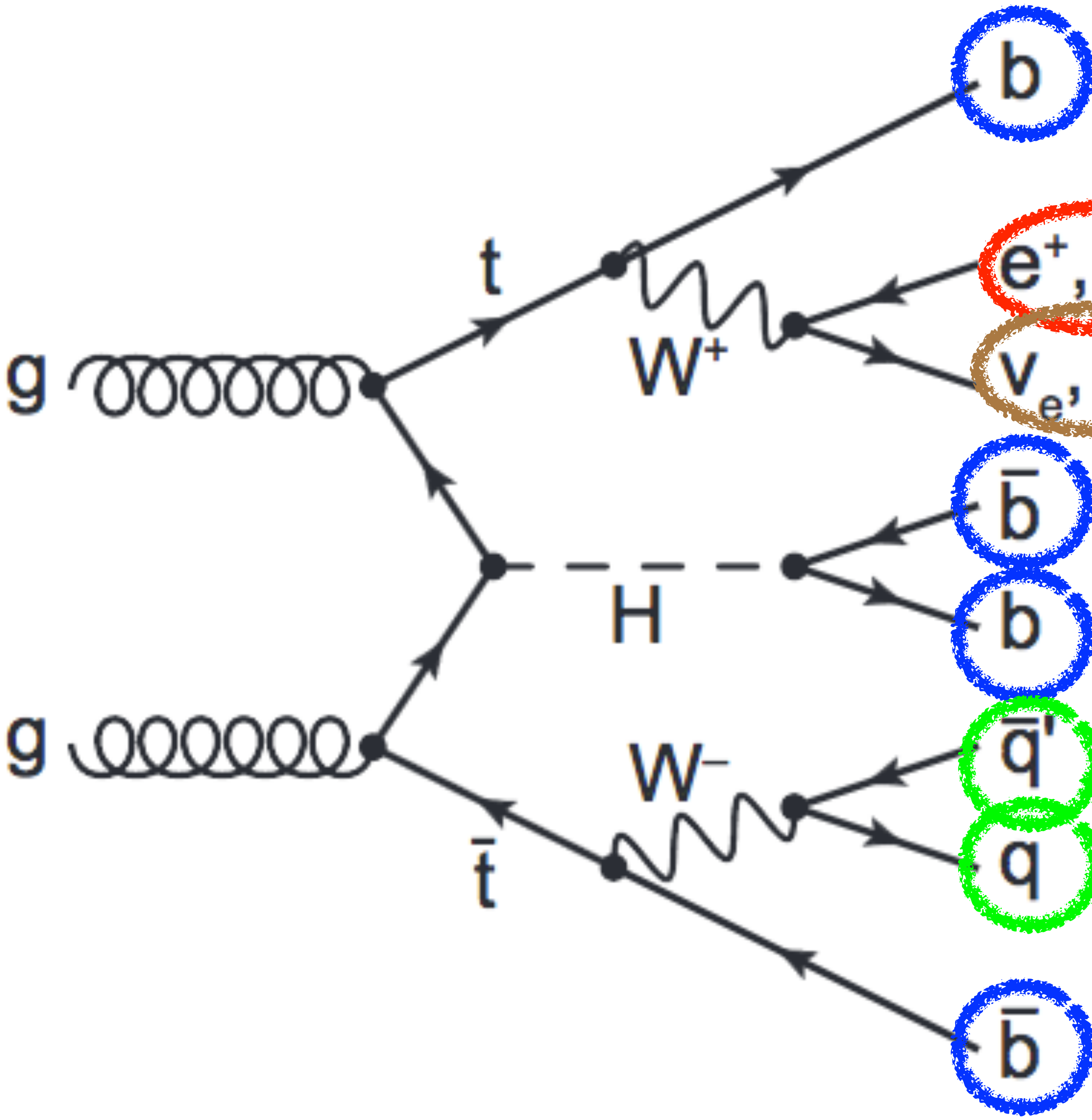
✓ Calibration from data → scale factors (SF) extracted as b-jet efficiency in data and compared to efficiency in MC

- b-jet calibration
  - dileptonic ttbar selection for tag&probe and PDF methods → 3/8% uncertainty on SF
- c-jet calibration
  - W+c calibration (triggering on the presence of μ from semileptonic c-decay) → 8-22% uncertainty on SF
- light-jet calibration
  - light-flavour mistag rate extracted with negative tag method in dijet events → 15-50% uncertainty
- Calibration for 4 b-tagging operation points (60%, 70%, 77%, 85% b-jet efficiency) available
  - used to define analysis regions (see later...)





ttH(H->bb) - let's have a closer look



☑ 4 b-jets - fundamental b-jet identification and reconstruction as two b-jets come from the Higgs signal decay (large combinatorics background)

☑ Leptons (1, 2 leptons according to tt system decay) - very useful to trigger events (unlike the full-hadronic final state)

☑ Neutrinos from W decay

☑ Additional non-b jets

✓ Final state reconstruction is very important to achieve a good measurement

- in light of the challenging experimental signature, easy to misidentify/loose particles → analysis split in categories

ttH(H->bb) - analysis channels and selection

Lepton preselection

✓ 1-lepton final state

- 1 e/μ (p_T>27 GeV), ≥5 jets (p_T>25 GeV)

✓ 2-leptons final state

- 2 opposite-sign e/μ (p_T>27 GeV), ≥3 jets (p_T>25 GeV), Z-mass window veto

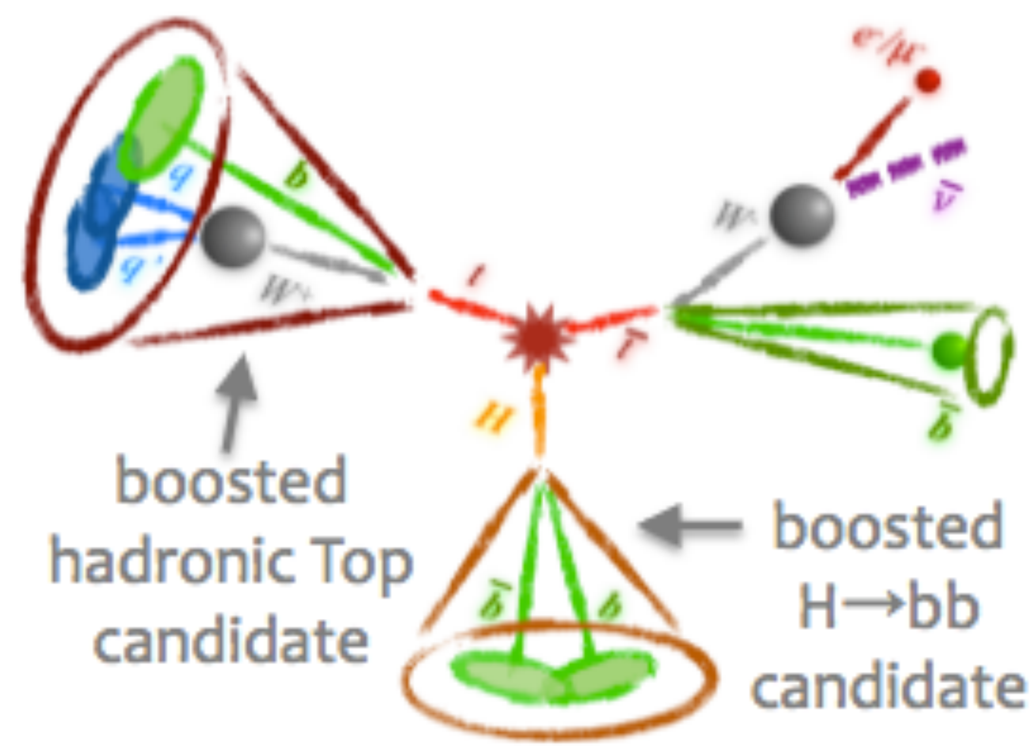
Requirements on b-jets

b-tagging requirements	Resolved (low p_T)
Single lepton final state	≥2 very tight b-tags or ≥3 medium b-tags
Dilepton final state	≥2 medium b-tags

Boosted category in 1-lepton

✓ High pt subcategory of 1-lepton channel

- Higgs boson and hadronically decay top quark produced with high transverse momentum (boosted)
- large radius jets (R=1.0) formed by reclustering R=0.4 calorimeter jets
- 1 loose b-tag outside large-jets



tt+jets production

✓ Dominant background - tt+jets production

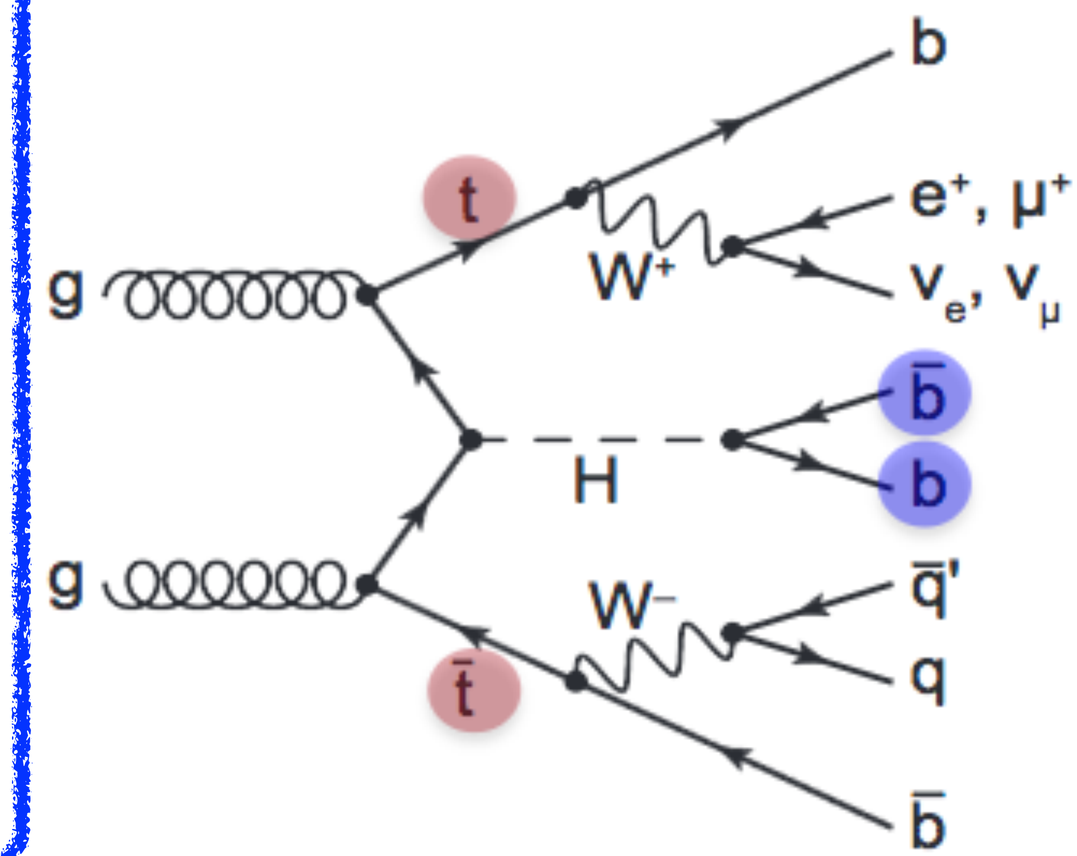
- large yields and challenging modeling in Monte Carlo simulation

✓ Matrix element generator - Powheg

- NNLO+NNLL cross section
- NLO generator with 5 flavour (5F) scheme (massless b-quark)
- PDF extracted for 5 flavours

✓ Parton shower and hadronization - Pythia8

- most updated tunes to data (ATLAS-PHYS-2016-020)

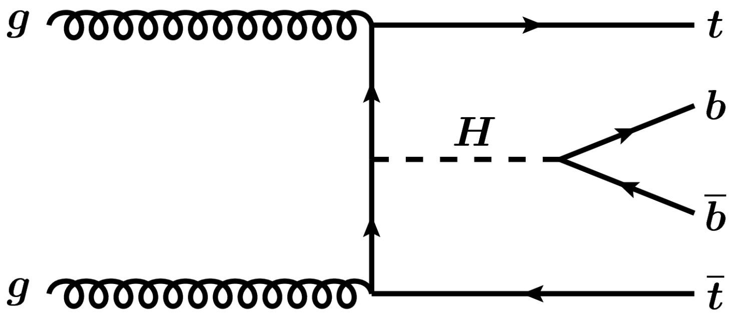


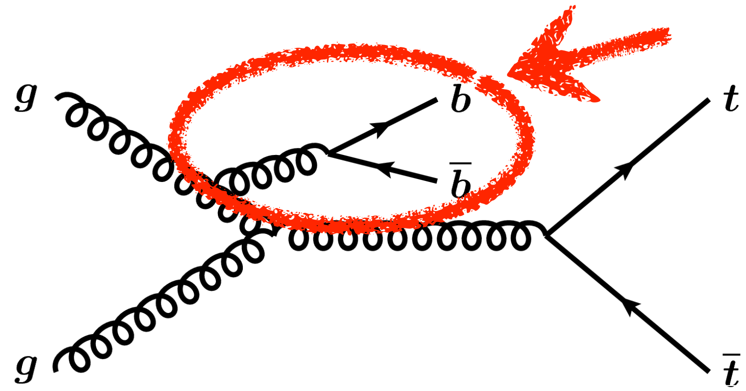
✓ tt+heavy flavour sample split in 3 sub-components based on the flavour of additional jets

- tt+≥1b, tt+≥1c, tt+light; dominant component in the measurement is tt+≥1b (next slide)

✓ Systematic uncertainties on tt+jets → comparison to alternative samples with different Monte Carlo matrix elements and generators, parton shower or radiation schemes

Main background - tt+≥1b production





★ very large σ(ttbb) compared to signal

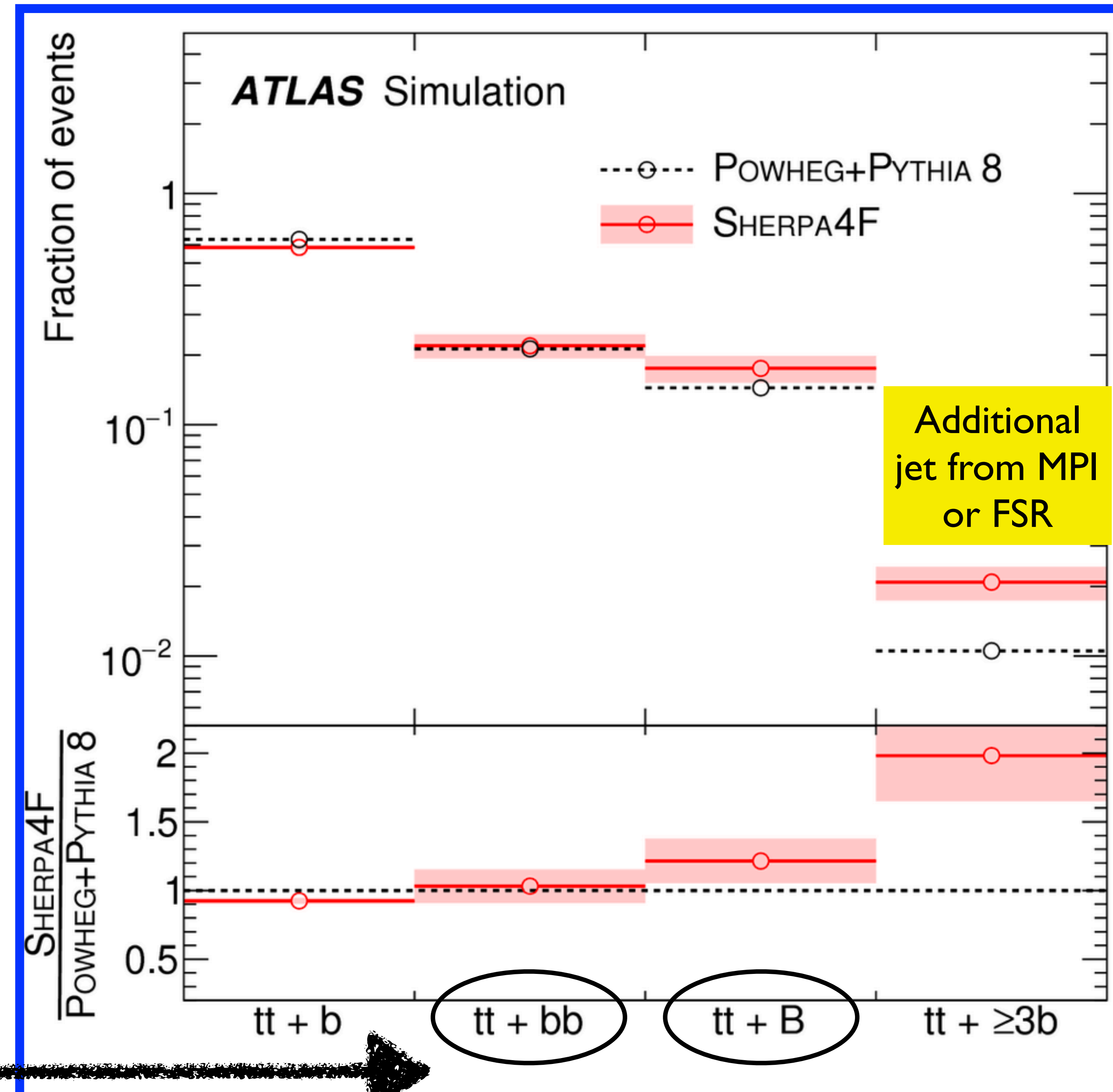
★ Same signature of tt(H→bb) signal

★ No constrain in data and little theoretical knowledge/MC modeling

✓ tt+≥1b split in subcategories according to hadron multiplicity (tt+b, tt+bb, tt+B, tt+3b)

✓ Each event fraction in subcategories reweighted to predictions extracted from Sherpa +OpenLoops to improve modeling

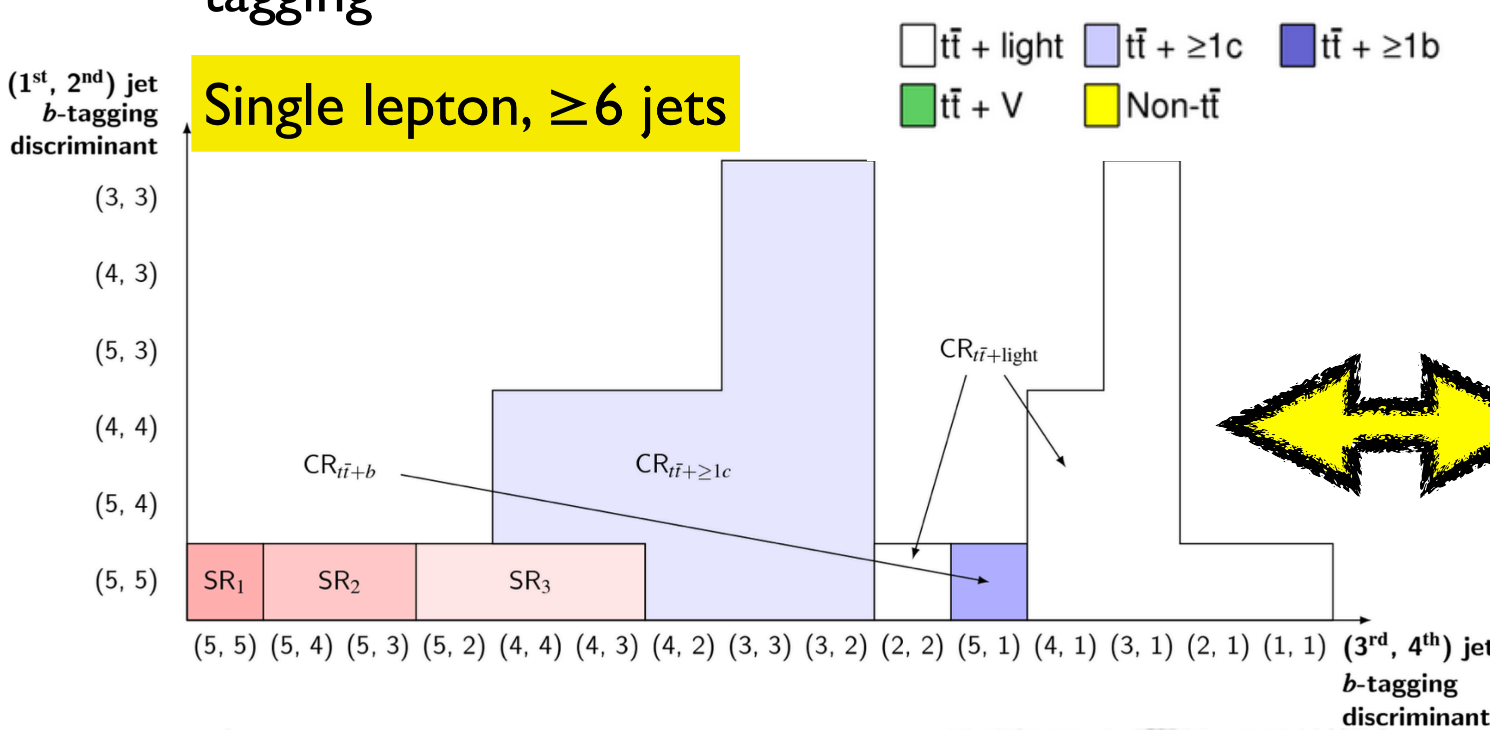
- nominal Powheg+Pythia8 → gluon splitting to bb from parton shower
- Sherpa+OpenLoops→matrix-element for tt +bb production with massive b-quarks
  - g →bb from matrix element improves characterization

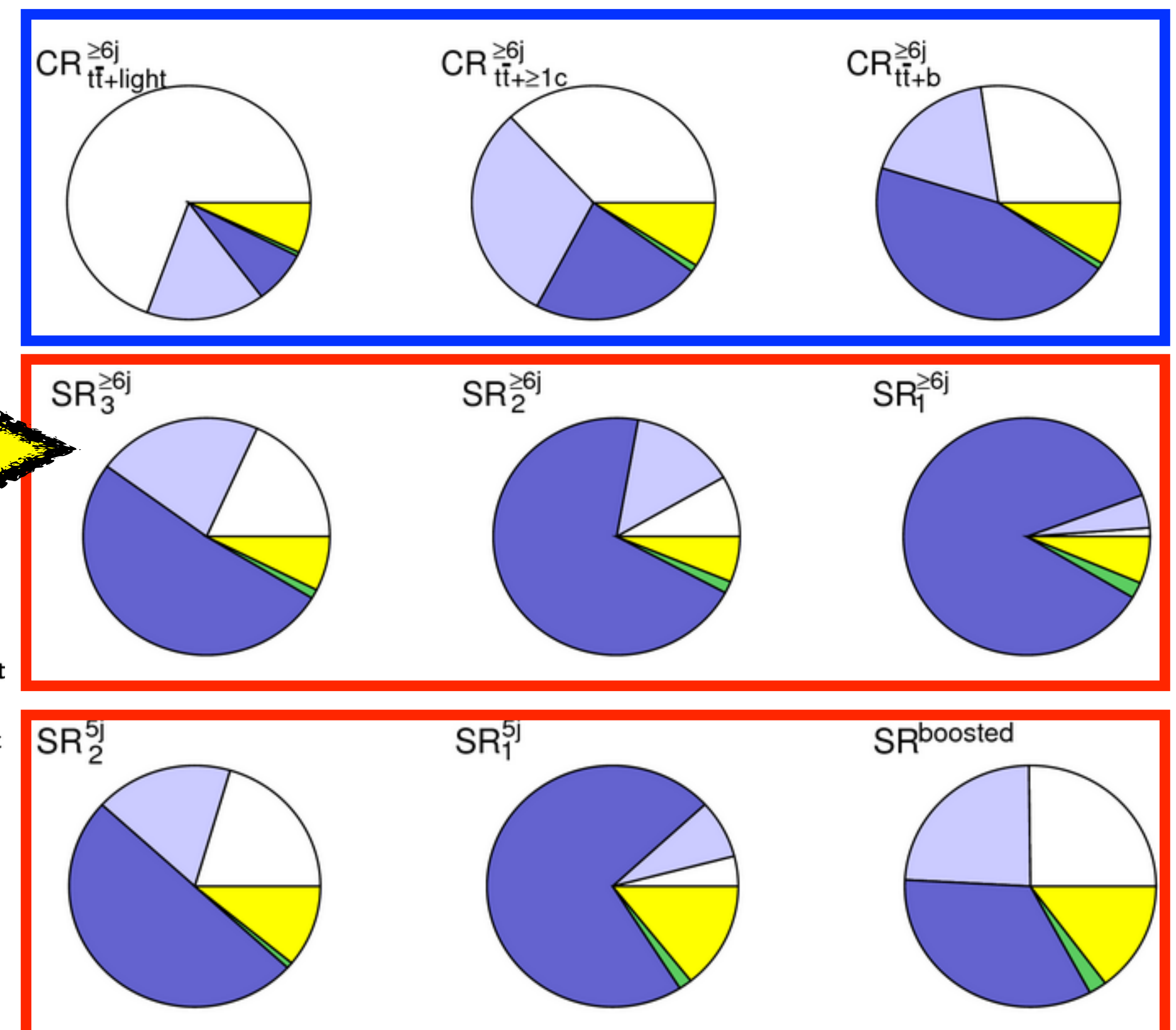


Signal and control region - single lepton

✓ Requirements on b-tagging discriminants for jets in the event defined to split phase-space and create signal and control region (≥5 jets and ≥6 jets)

- control regions (CR) enriched in reducible background
- signal region (SR) enriched in signal and reducible background (tt+≥1b)
- signal purity in ultra-pure signal region: 1.6-5.3%
- highest purity regions in single lepton ≥6j with 4b very tight b-tags
- control region dominated in tt+≥1c and tt+light and created by loosening requirements on b-tagging



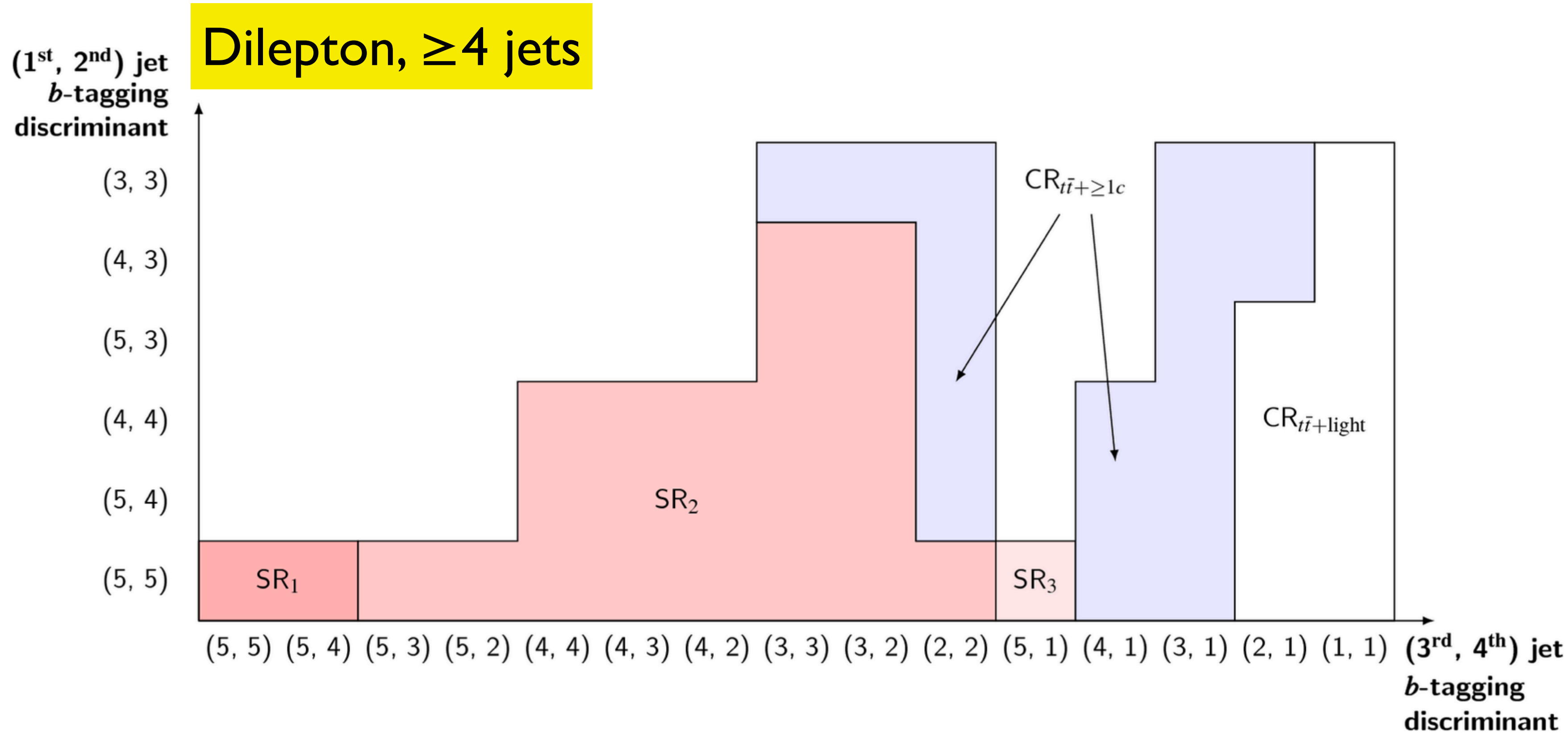


★ Constrain background uncertainties and measure normalization of background components (ttb, ttc) in CR, extract signal component in SR

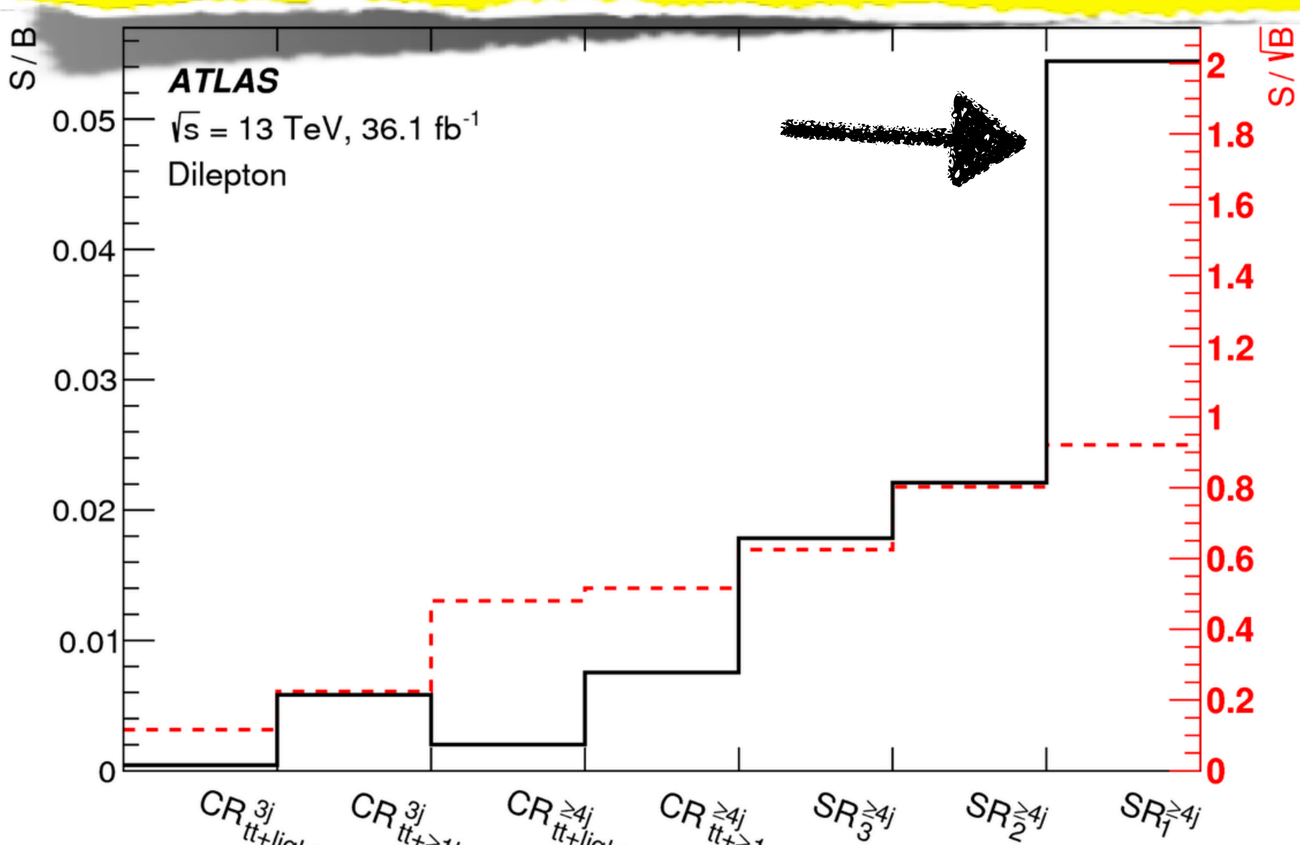
Signal and control region - dilepton

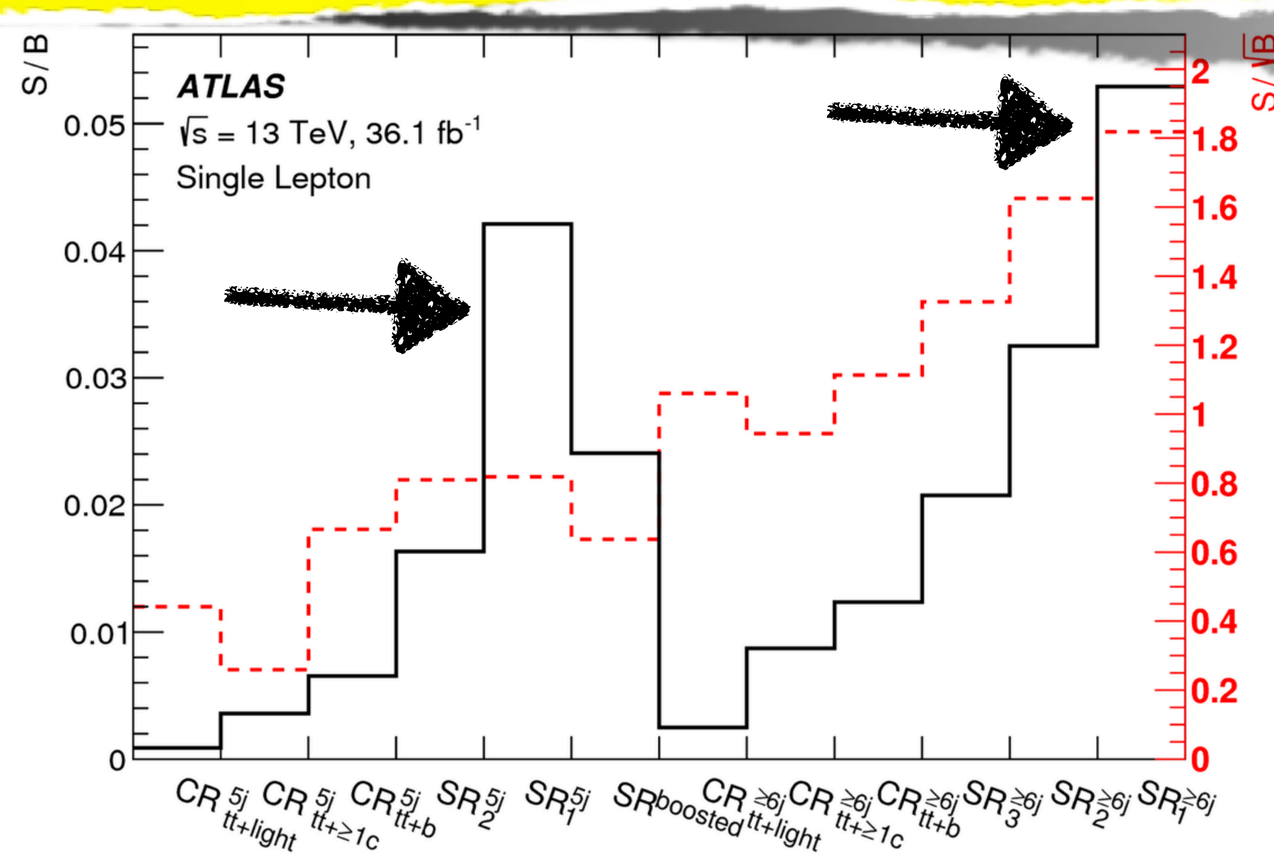
✓ Similar approach in dilepton final - signal and control regions to separate ttH and tt+≥1b from the tt+≥1c and tt+light components

- SR for ≥4 jets and highest purity in 3 very-tight b-tags+1 tight/very-tight b-tags
- CRs dominated by tt+≥1c and tt+light background



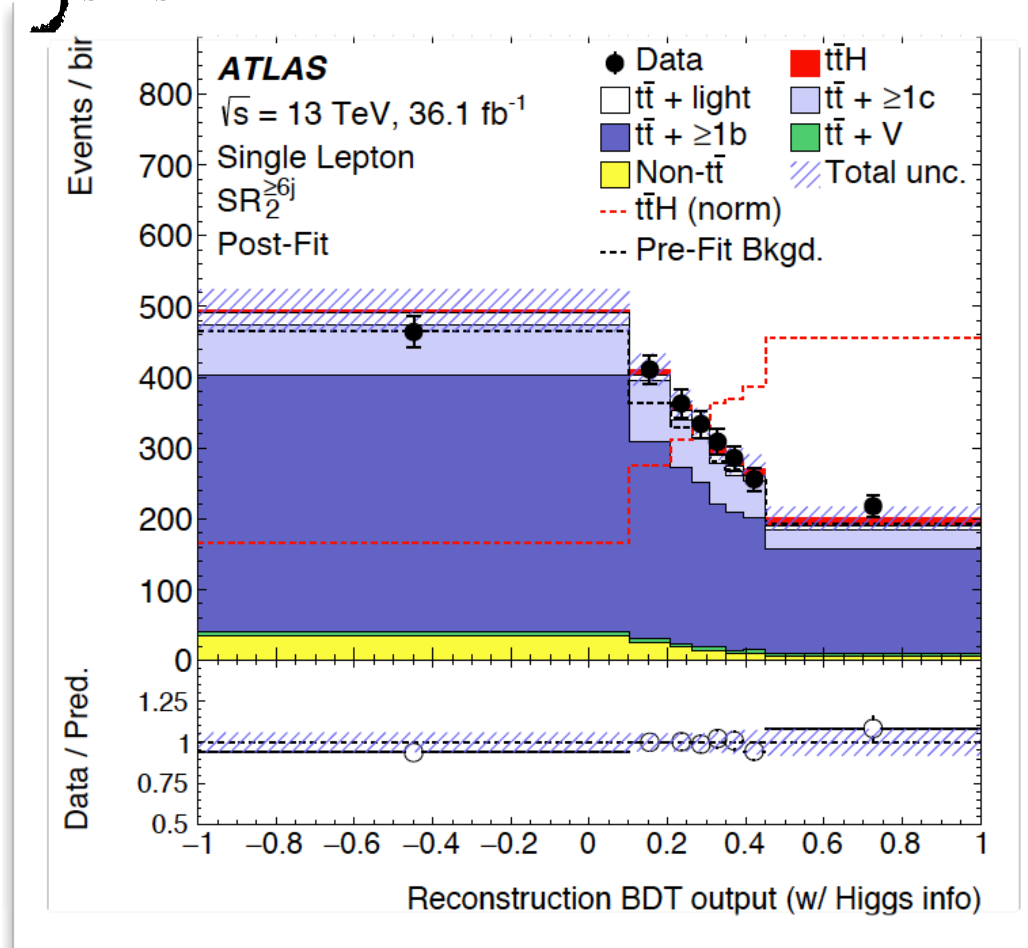
★ Largest sensitivity for high jet multiplicity (≥4 in DL, ≥6 in SL) with very tight b-tagging requirements

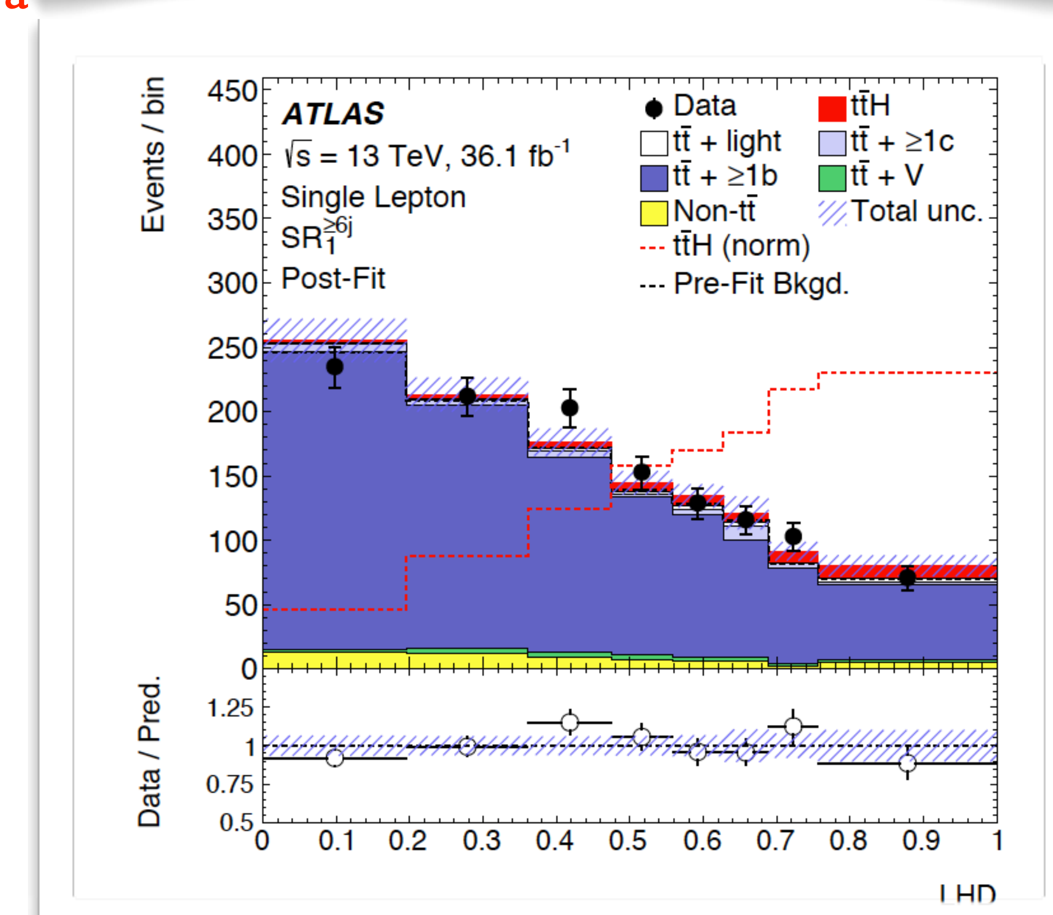




Multivariate analysis

- Reconstruction BDT [resolved, SR]
  - aiming at reconstructing the ttHbb system and reducing combinatorics background
  - Higgs boson correctly reconstructed in 50% (35%) of the cases with (without) Higgs kinematics in BDT training
- Likelihood discriminant [single lepton resolved, SR]
  - kinematic input variables (invariant mass, angles) in likelihood discriminant
- Matrix element method, MEM [single lepton resolved, 6j ultra-pure SR]
  - matrix-element method for best separation in most sensitive signal region
- Classification BDT
  - main signal/background separation algorithm
  - general kinematics of the final state, b-tagging variables
  - reconstruction BDT, likelihood discriminant and matrix element method included (if present)

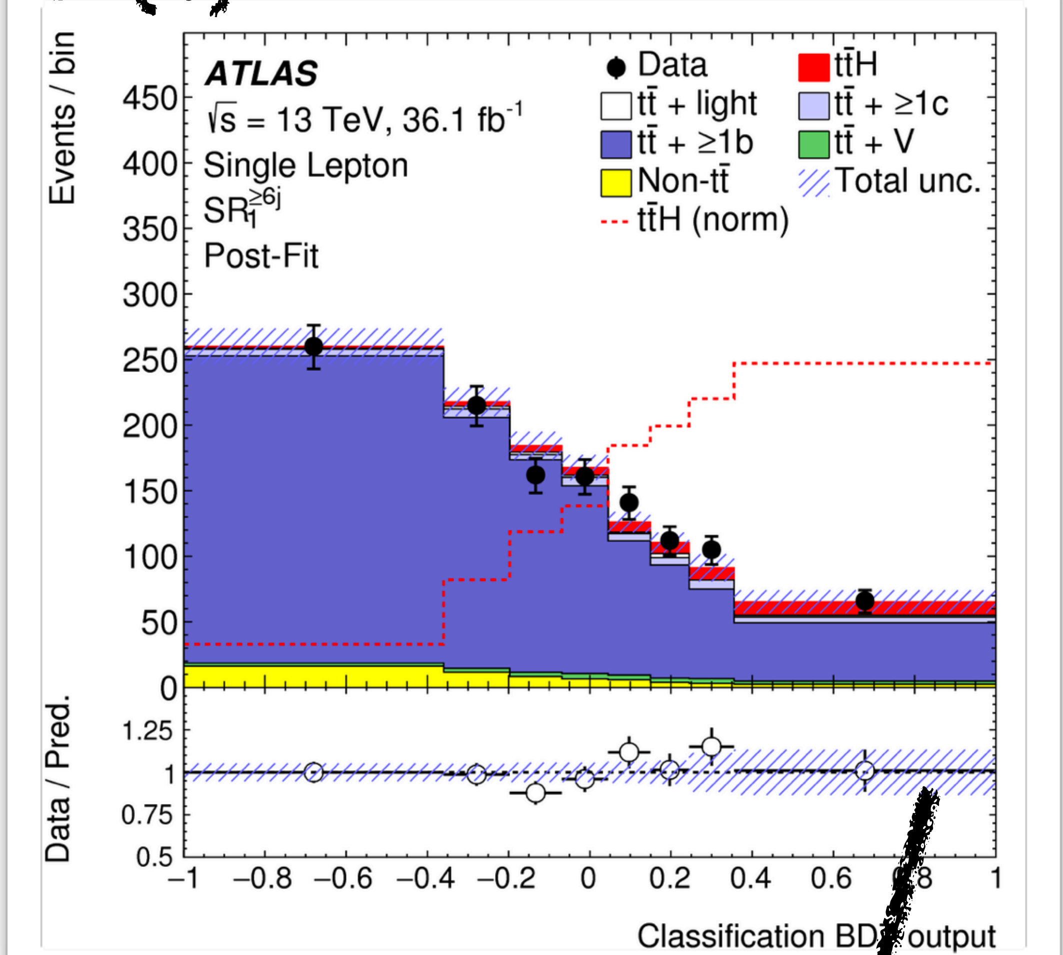


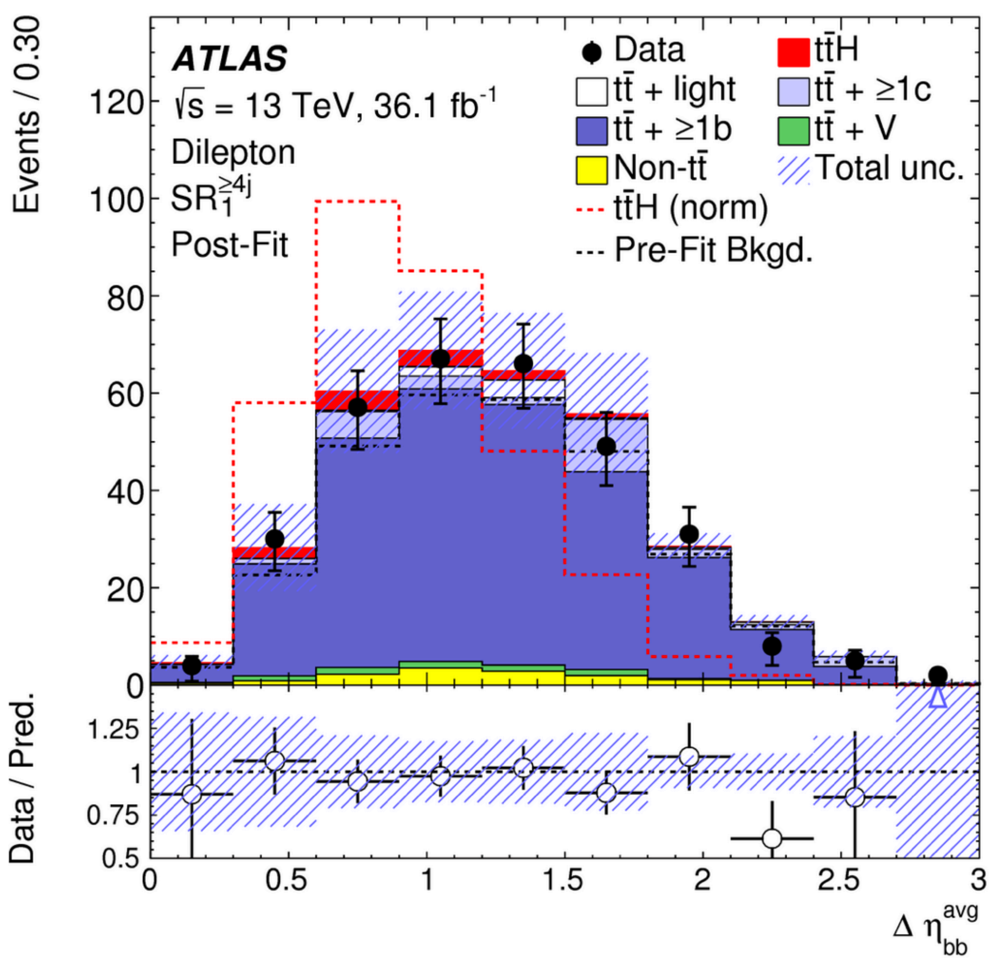


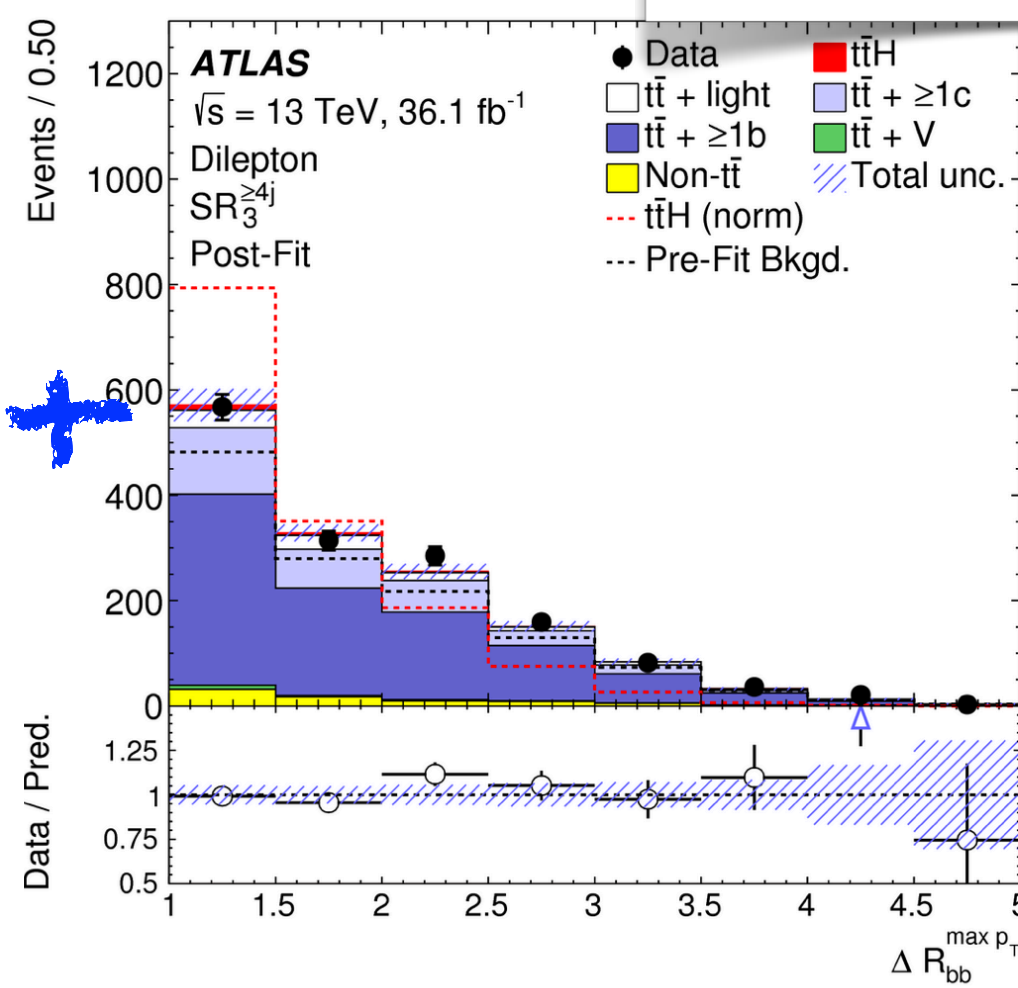
Multivariate analysis (2)

➡ Classification BDT incorporates kinematics inputs and response functions from multivariate discriminants in various signal regions (reconstruction BDT, likelihood discriminant, MEM)

- binary separation between ttH signal and ttbar background
- data/MC modeling of input variables considered
  - kinematic variables and b-tagging output response
- included in the fit for single-lepton and dilepton







★ Largest signal sensitivity driven by upmost bin in the classification BDT

Finally...fitting

✓ Simultaneous profile likelihood fit to signal and control regions

- signal regions (6 in SL and 3 in DL) → shape of classification BDT discriminant
- control regions (6 in SL and 4 in DL) → $H_T = \sum p_T$ for tt+≥1c control region in SL and 1-bin in all other control regions

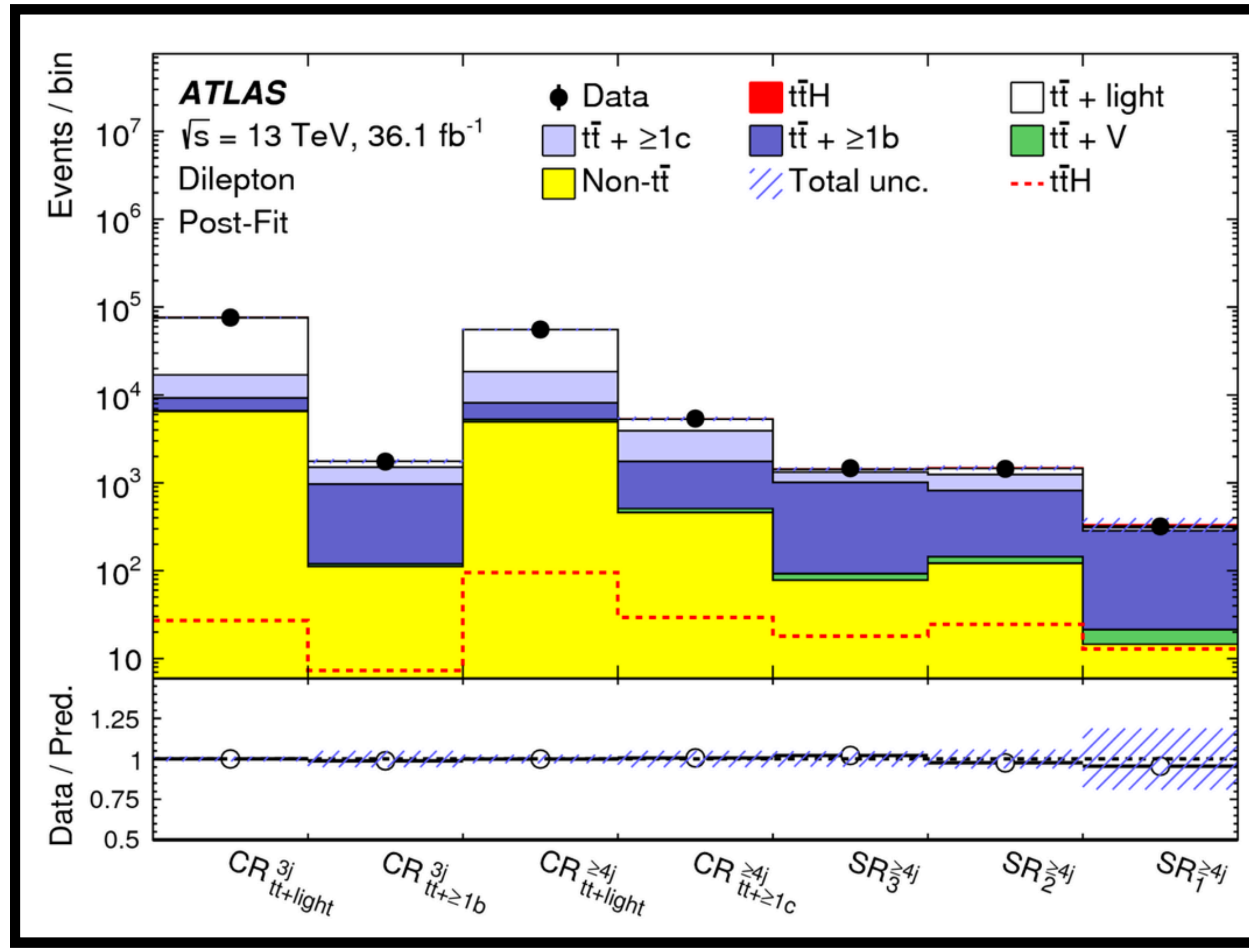
✓ Fitting benchmark parameters:

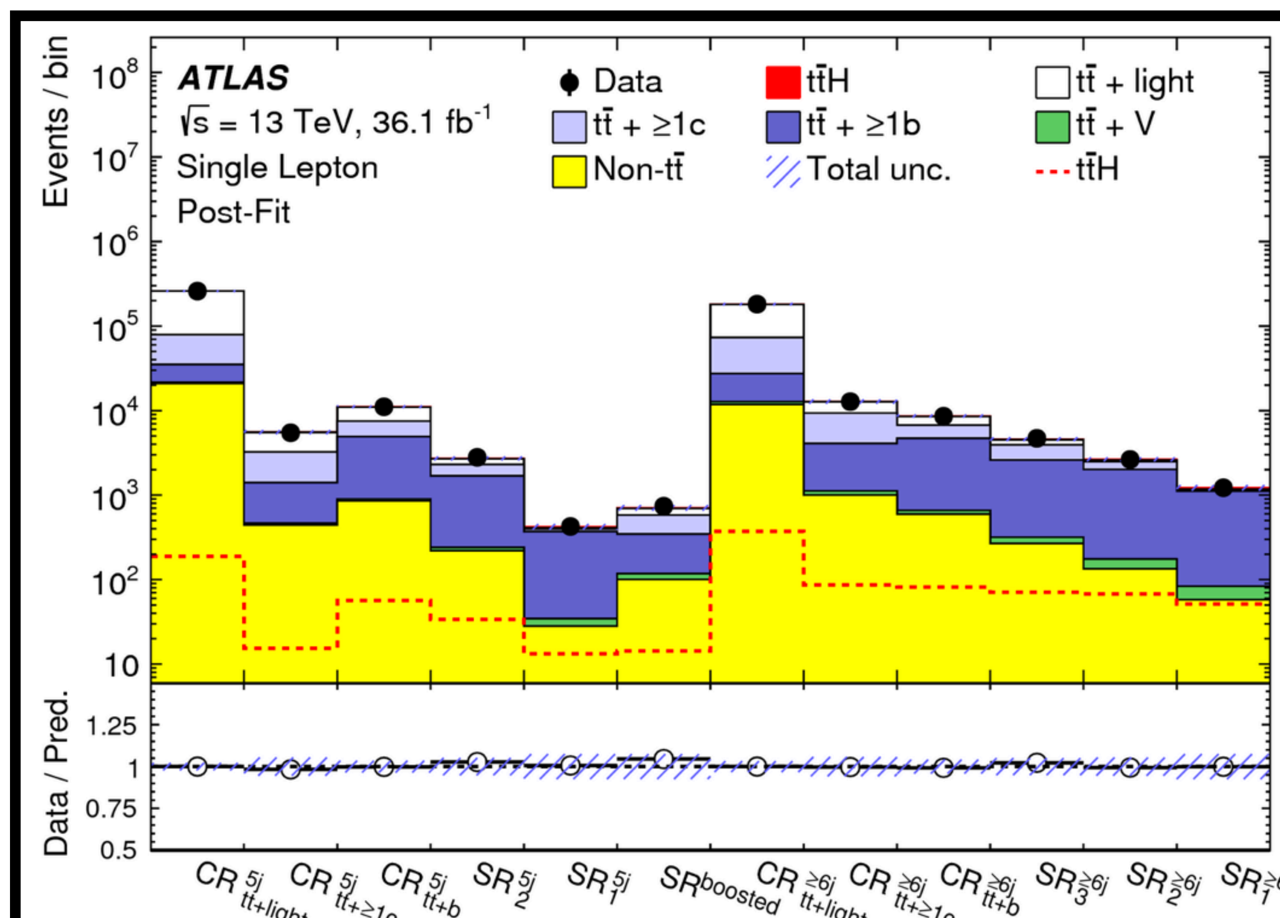
- ttH signal strength, $\mu_{ttH} = \sigma_{ttH}/\sigma^{SM}_{ttH}$
- tt+≥1b and tt+≥1c normalizations left free floating in the fit

✓ Nuisance parameters from systematic uncertainties included in the fit model

➡ Excellent data/MC prediction agreement in post-fit yields

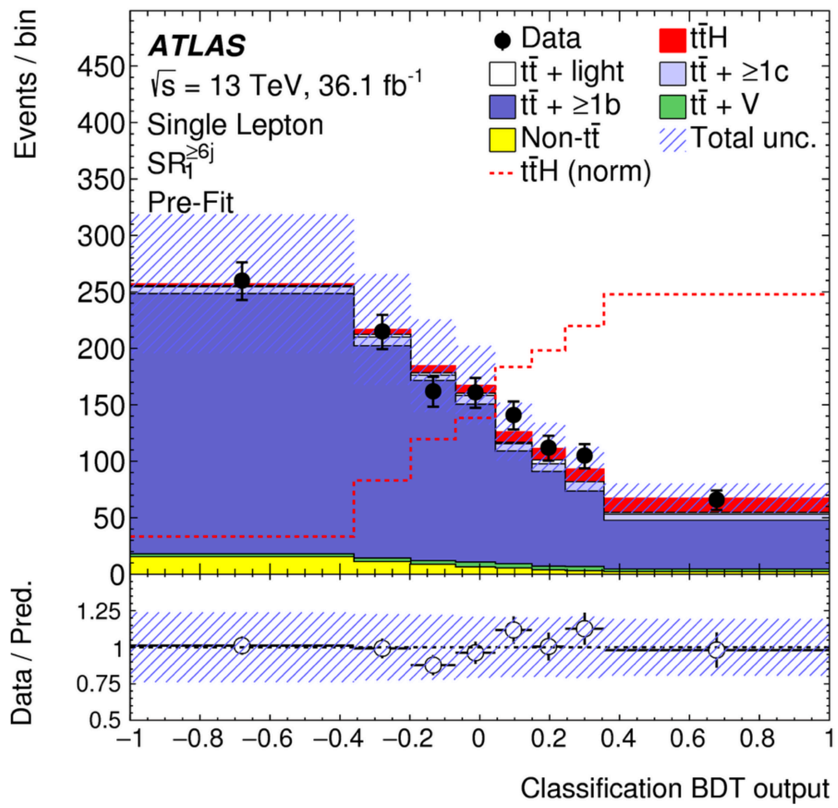
- remaining differences covered by the total uncertainties

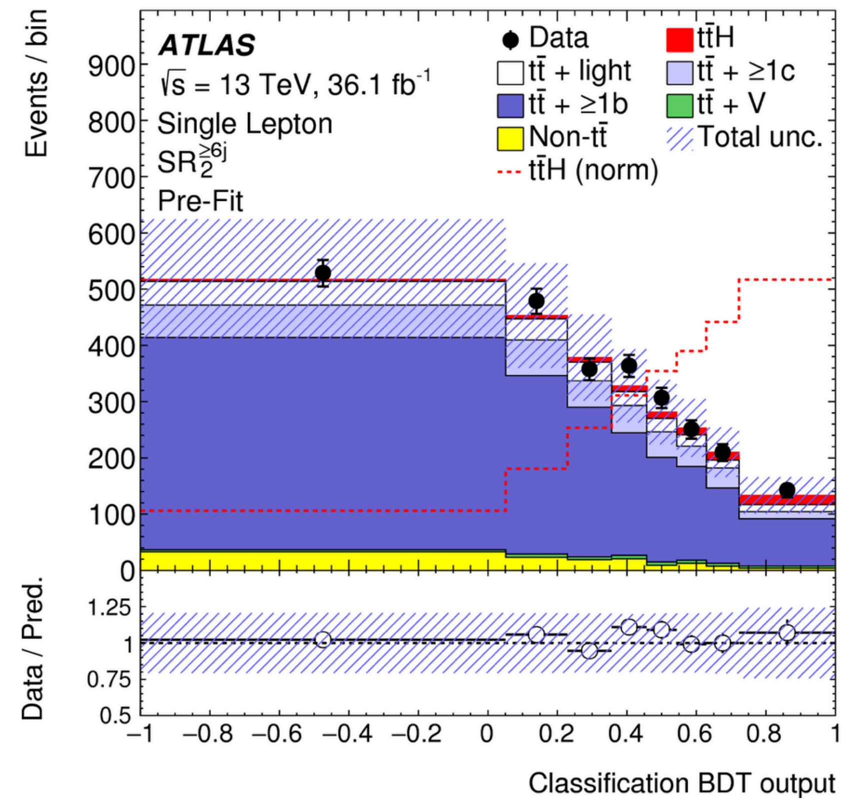


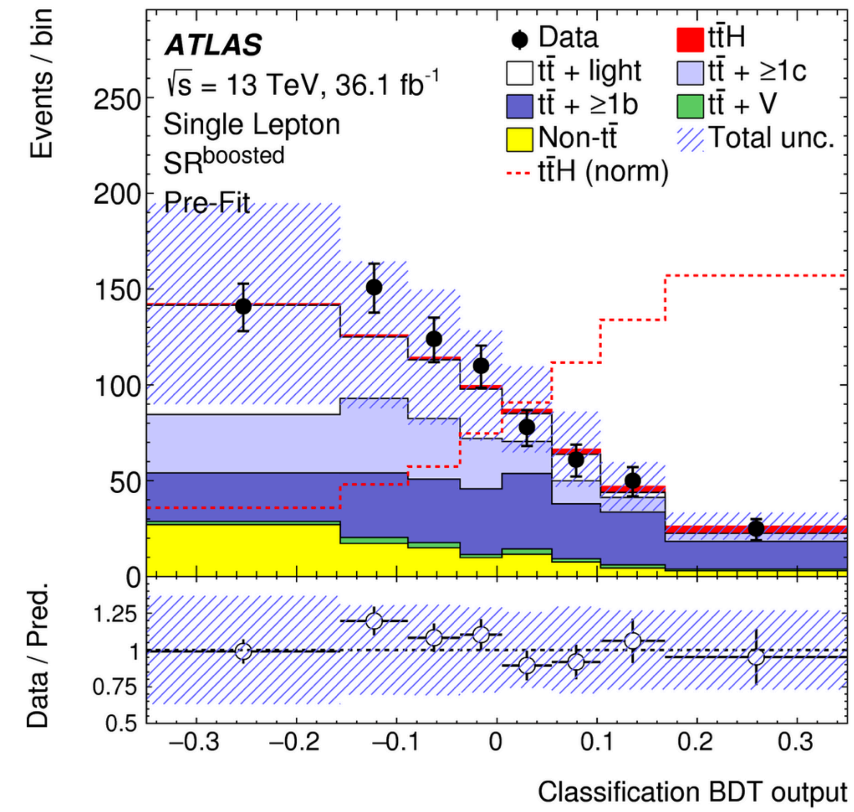


Pre/post fit distributions in single lepton

Pre-fit





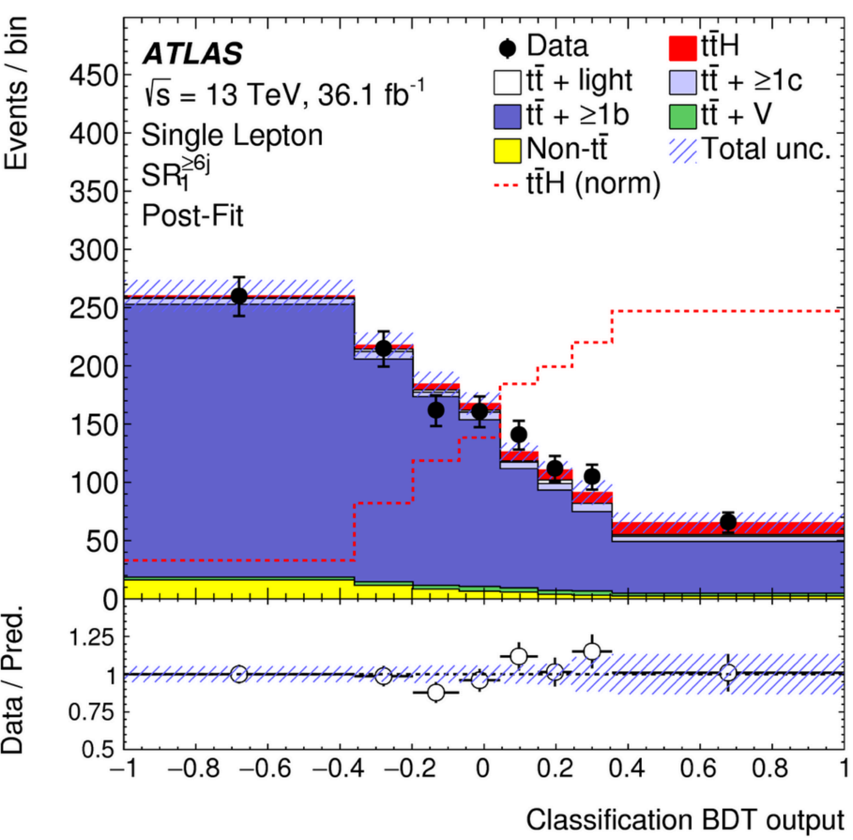


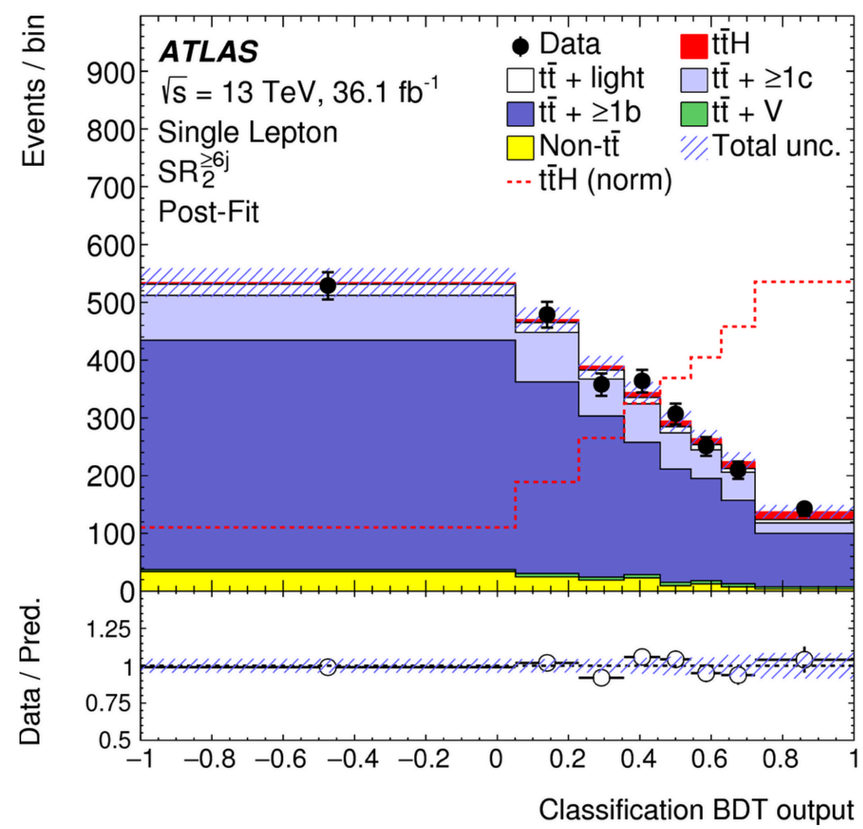
SR1

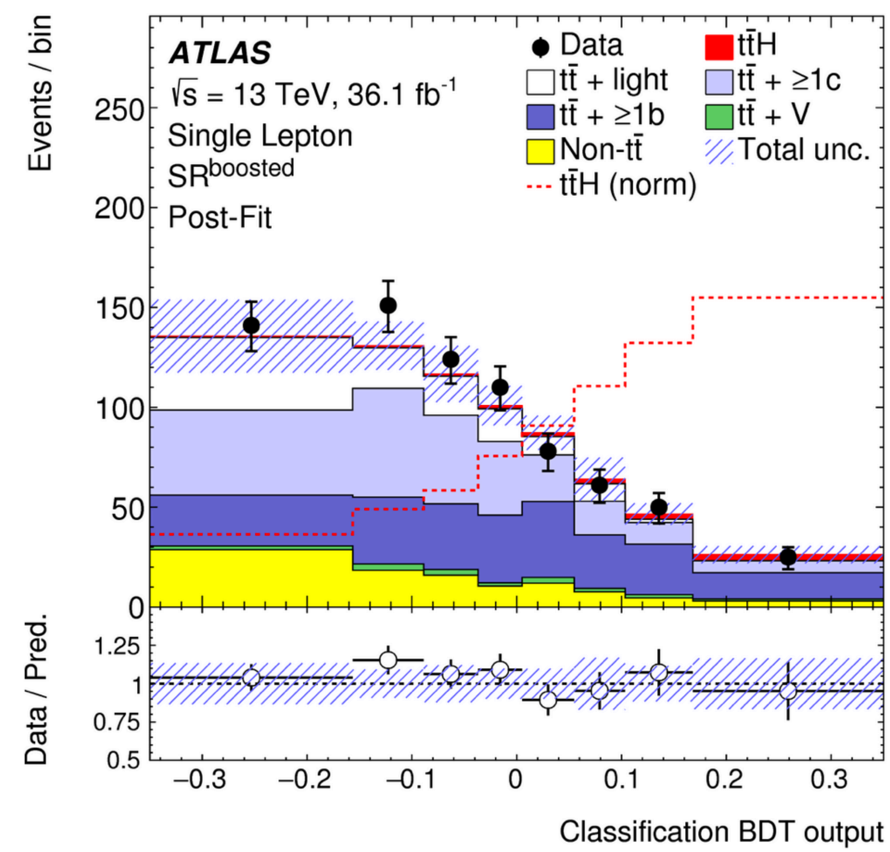
SR2

SR boosted

Post-fit

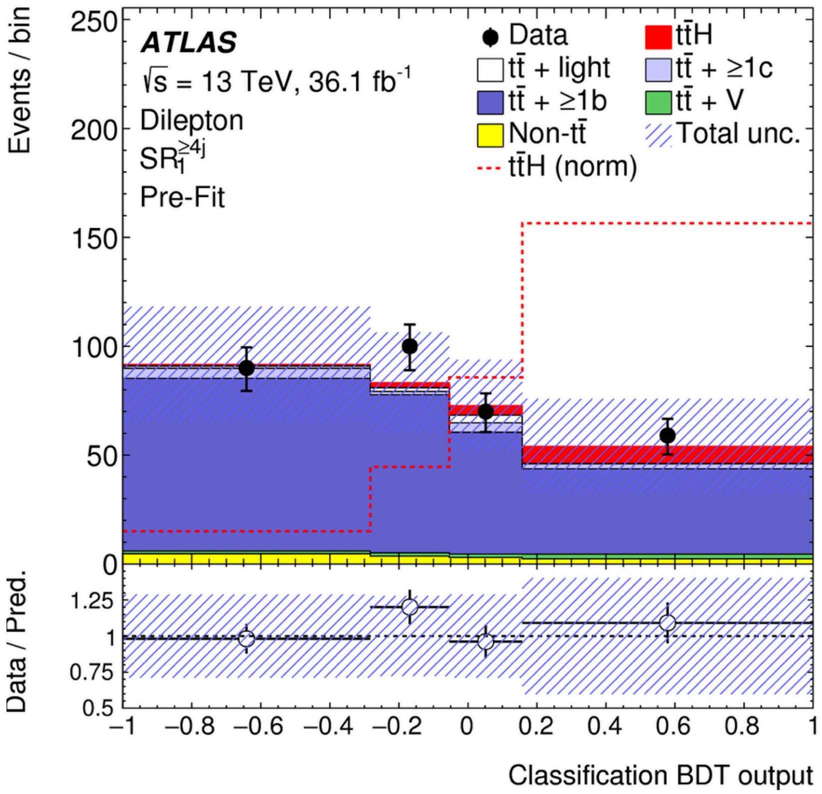


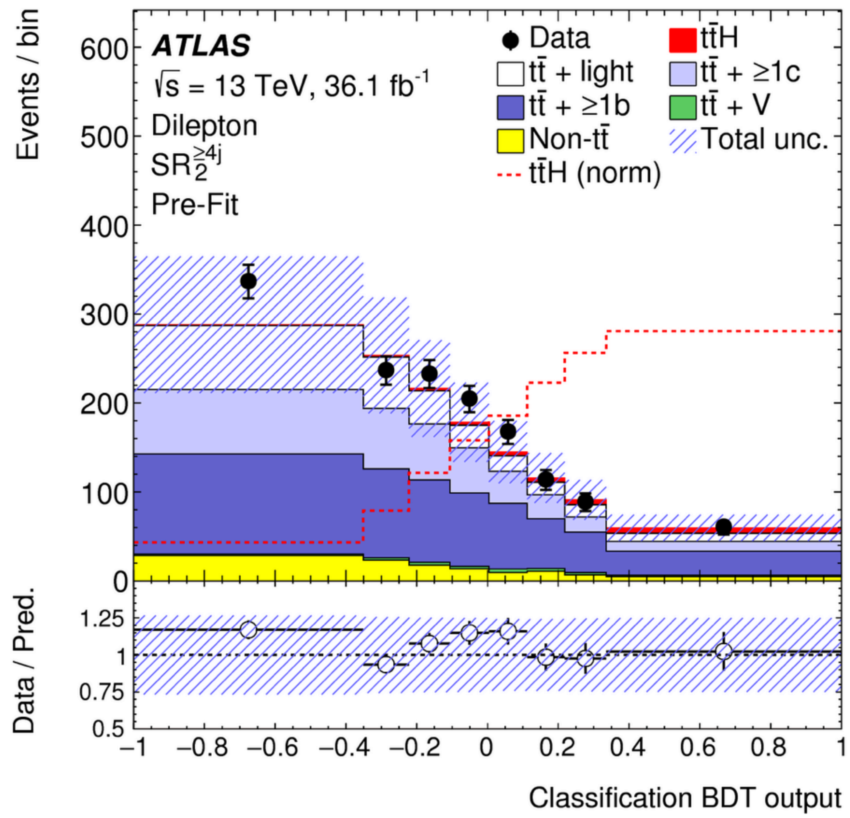


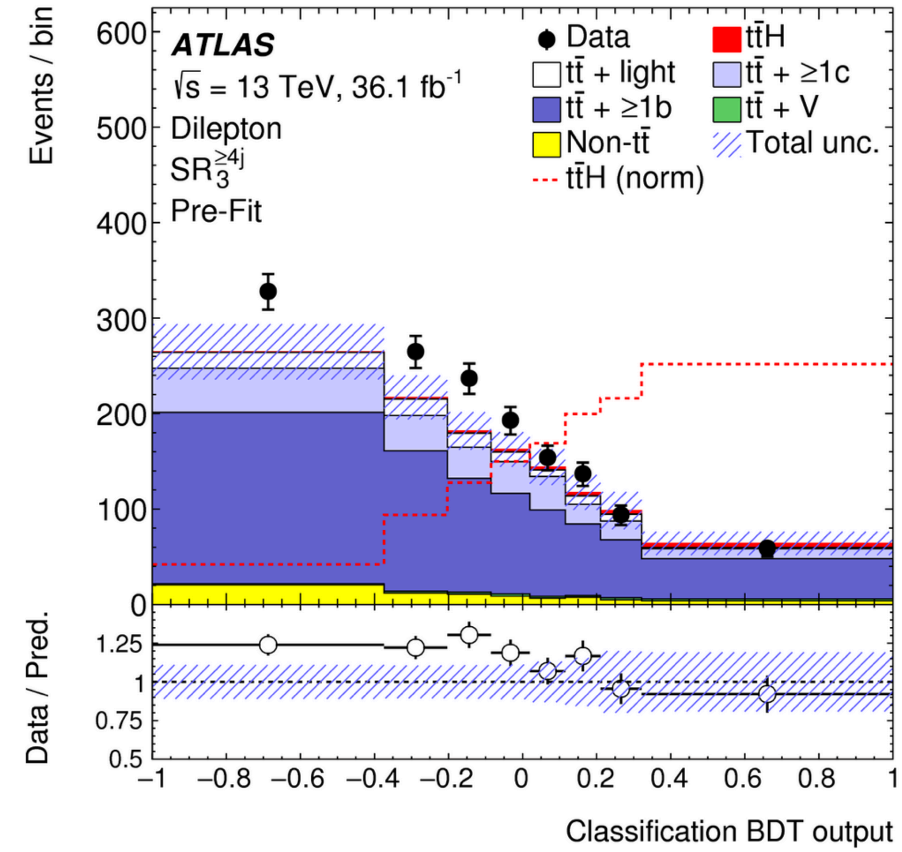


Pre/post fit distributions in dilepton lepton

Pre-fit

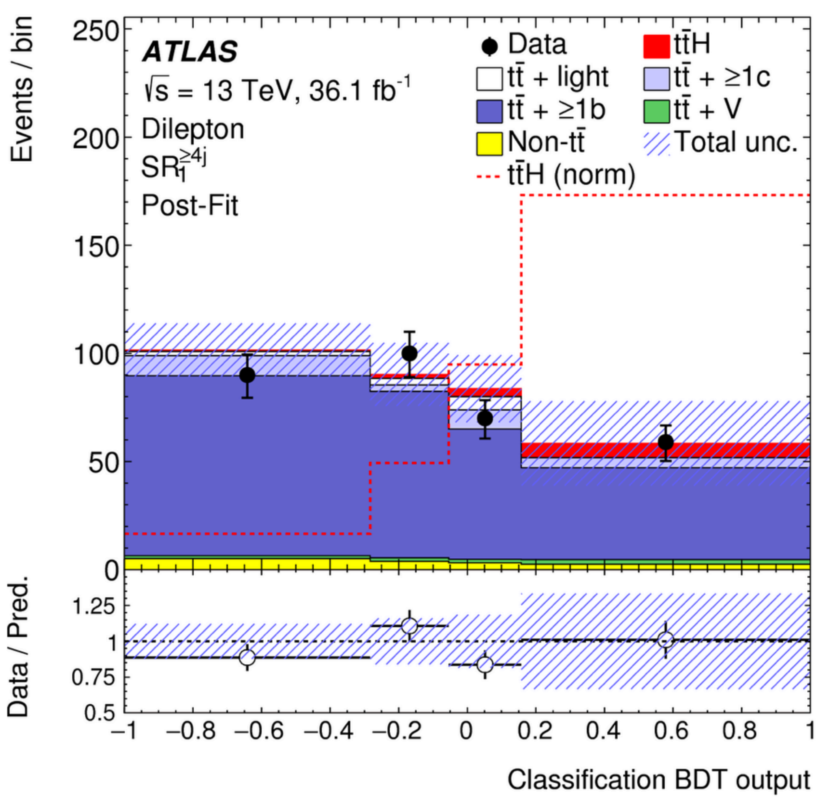


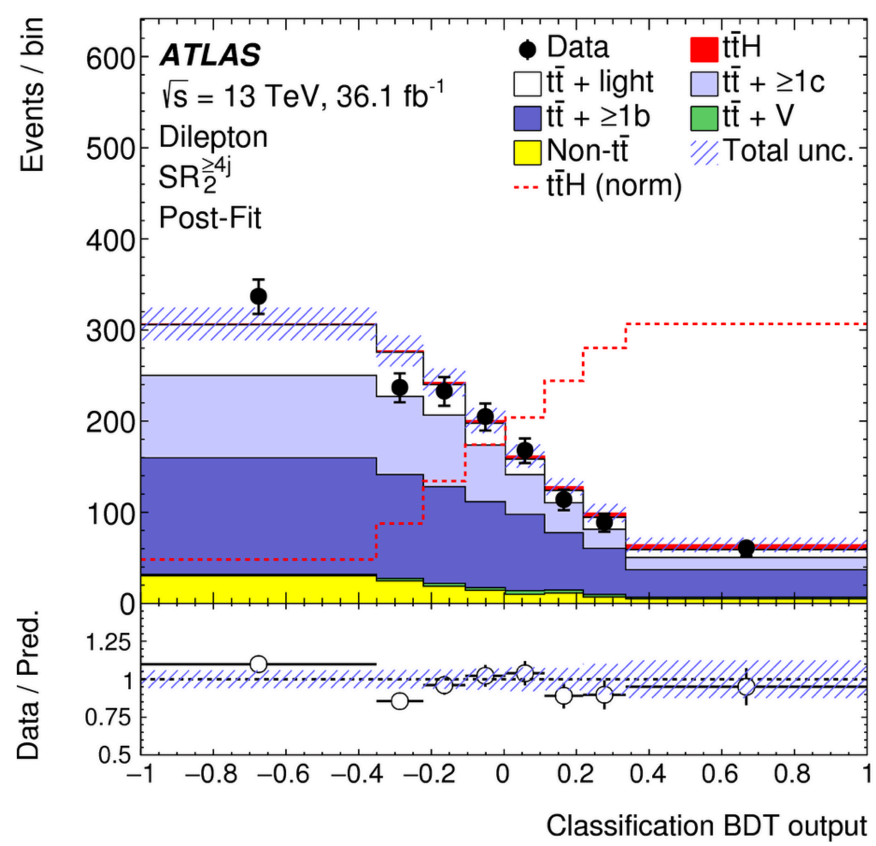


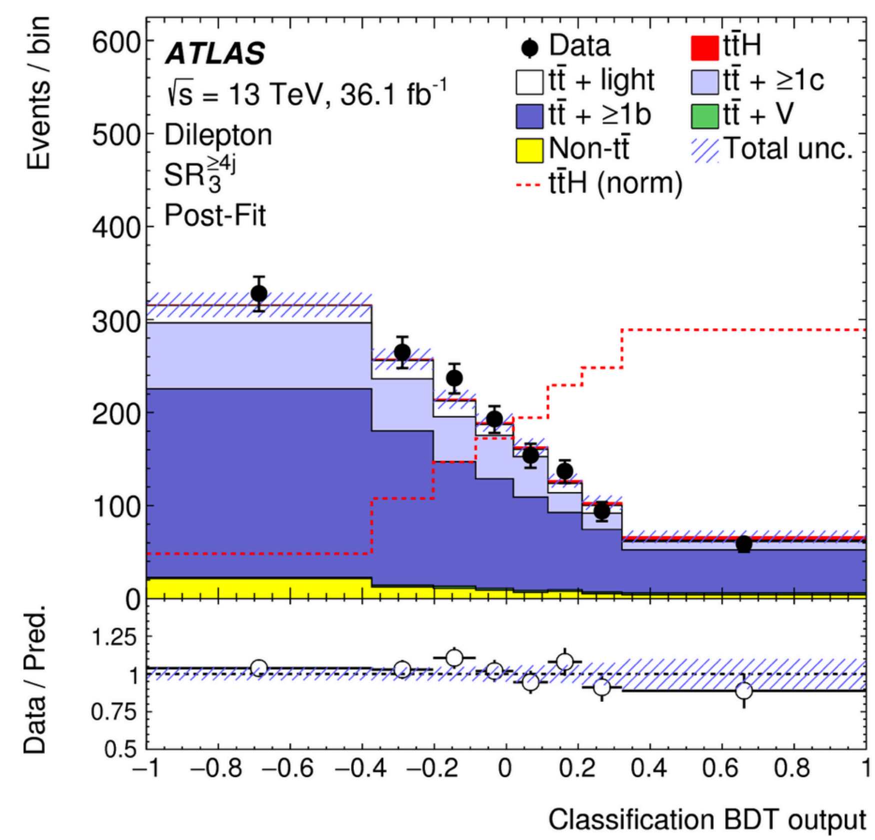


SR1 SR2 SR3

Post-fit







Systematic uncertainties

Uncertainty source	$\Delta\mu$	
$tt + > 1b$ 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

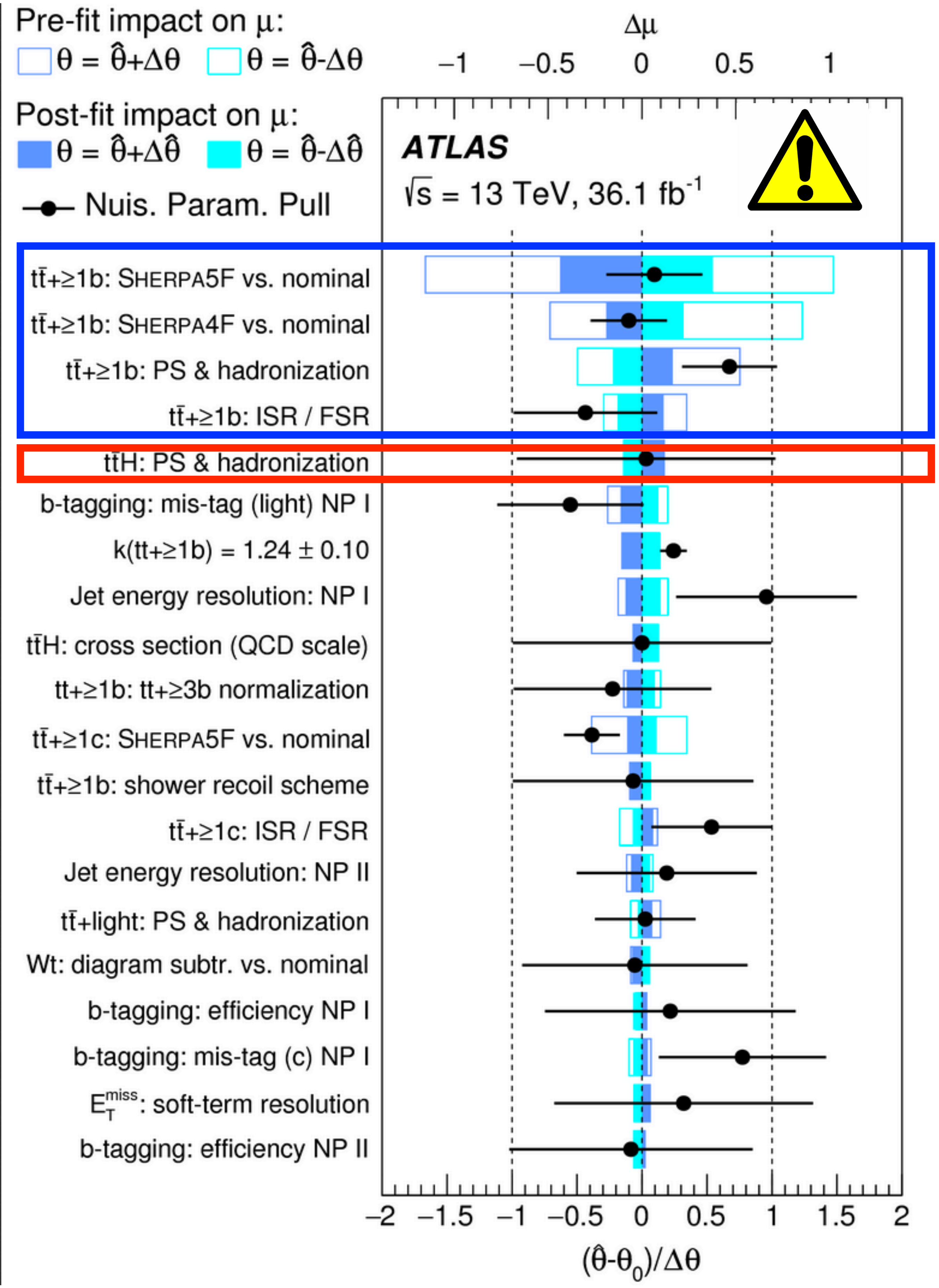
✓ Analysis is largely systematics-limited (~62% total uncertainty on the ttH signal strength)

- main source is tt+≥1b modeling
- large contributions on available Monte Carlo statistics
  - mostly relevant for the largest systematics uncertainties (tt+≥1b)
- experimental uncertainties contributing less, b-tagging and jet energy scale/resolution

✓ Work ongoing to reduce the dominant tt+HF uncertainty

- data-driven approaches to estimate tt+HF component
- SM g→bb cross section measurement

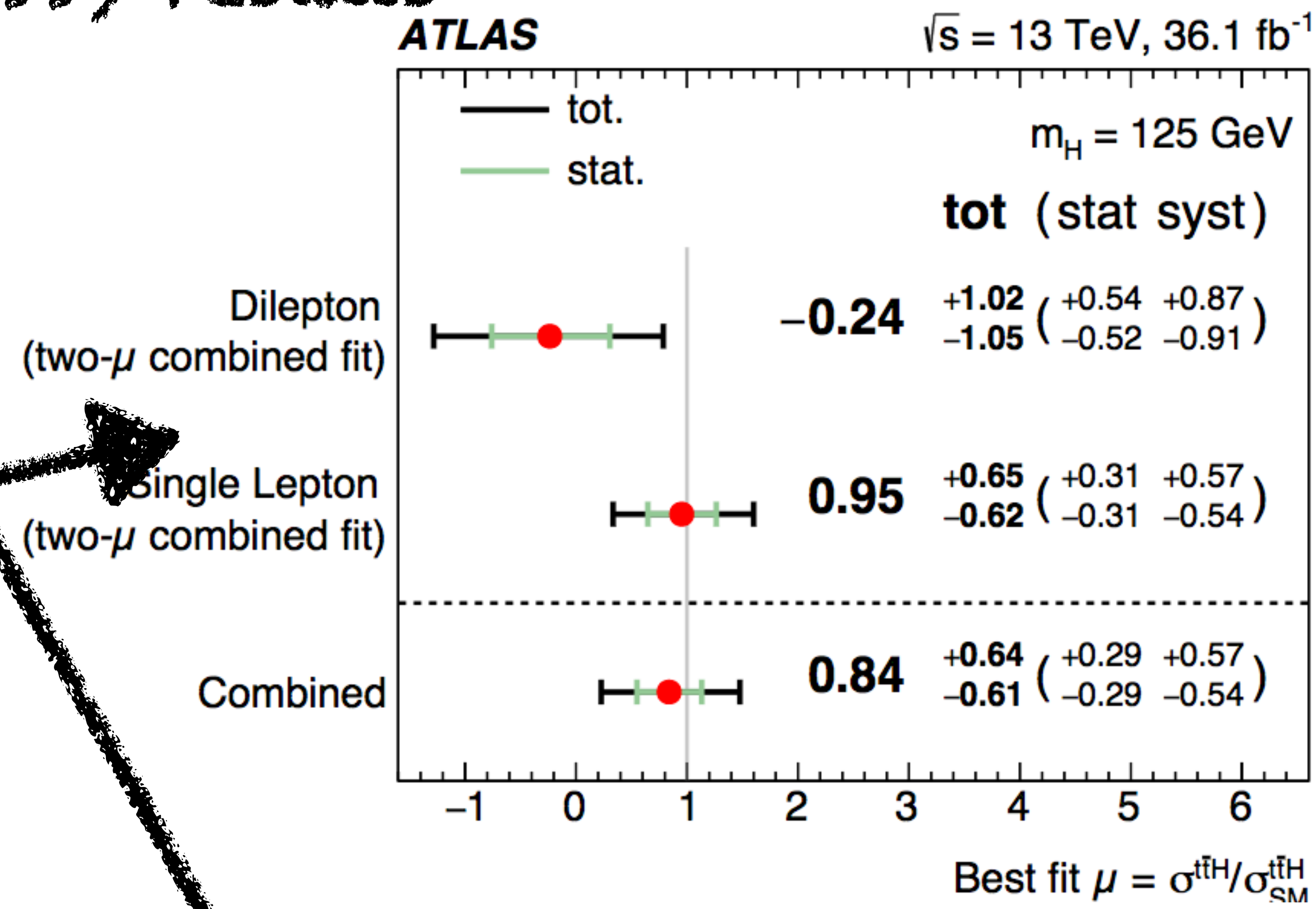
Systematic uncertainties (2)

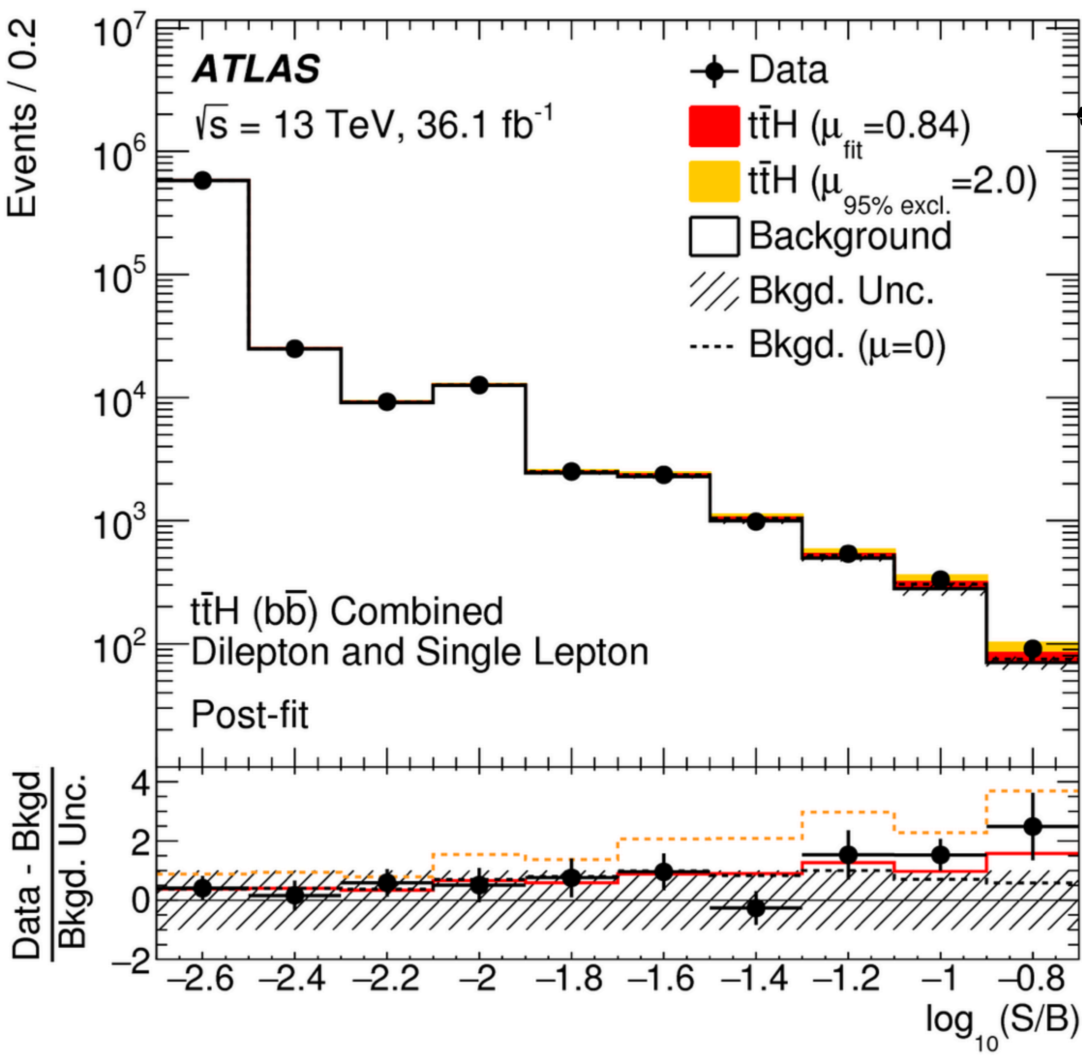


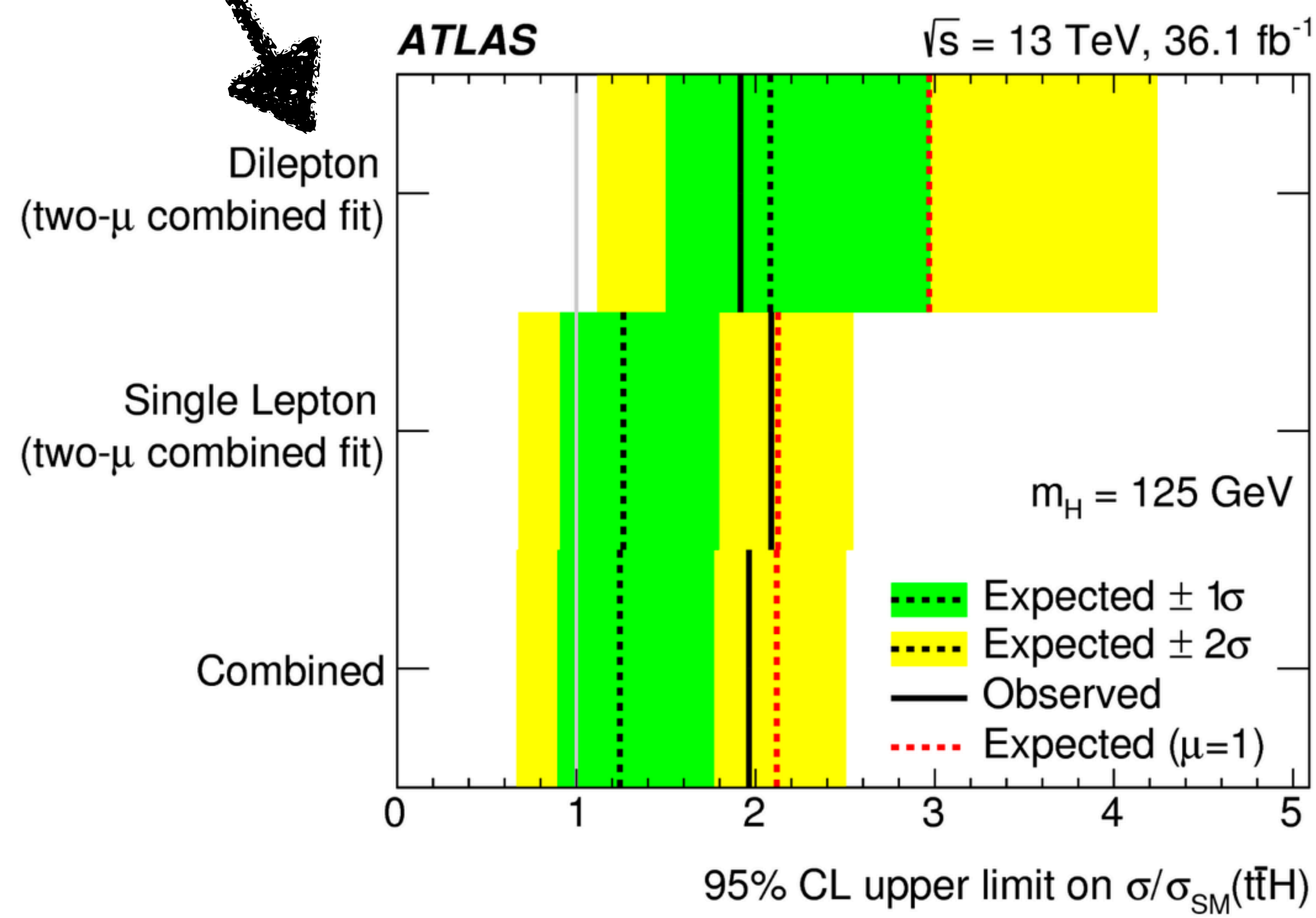
- ✓ Leading source of systematic uncertainties from tt+≥1b modelling (two-point systematics)
    - comparisons of various generators wrt nominal sample (Sherpa5F and Sherpa4F vs nominal)
    - characterization of parton shower, hadronization modeling and ISR/FSR components in Powheg+Pythia8
- ✓ Relatively large impact on ttH signal uncertainty (parton shower and hadronization model)
- ✓ Dominant experimental systematics on b-tagging relatively small
- ✓ Some constraints of major nuisance parameters testify larger variations than those observed in data
- ✓ Minor "pulls" of uncertainties from nominal values due to imperfect data/MC modeling

ttH(H->bb) results

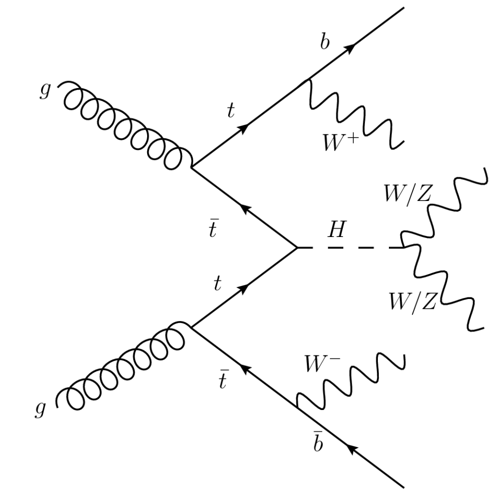
- ✓ Signal strength μ_{ttH} =0.84±0.61/0.64
  - error on signal strength dominated by systematics uncertainties
- ✓ ttH signal significance (against the background-only hypothesis): 1.4σ (expected 1.6σ)
- ✓ Exclusion limit at 95 % CL: μ_{ttH}>2.1

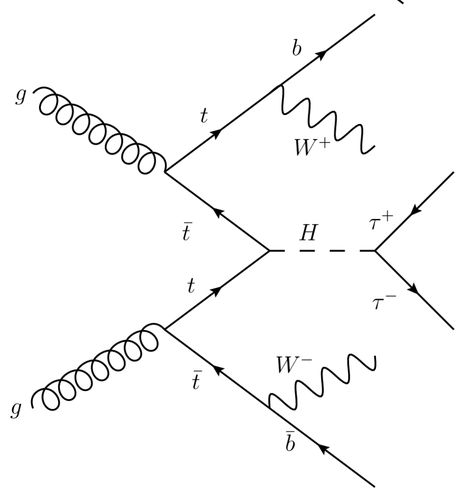






ttH(H->ZZ*, WW*, τ τ) results

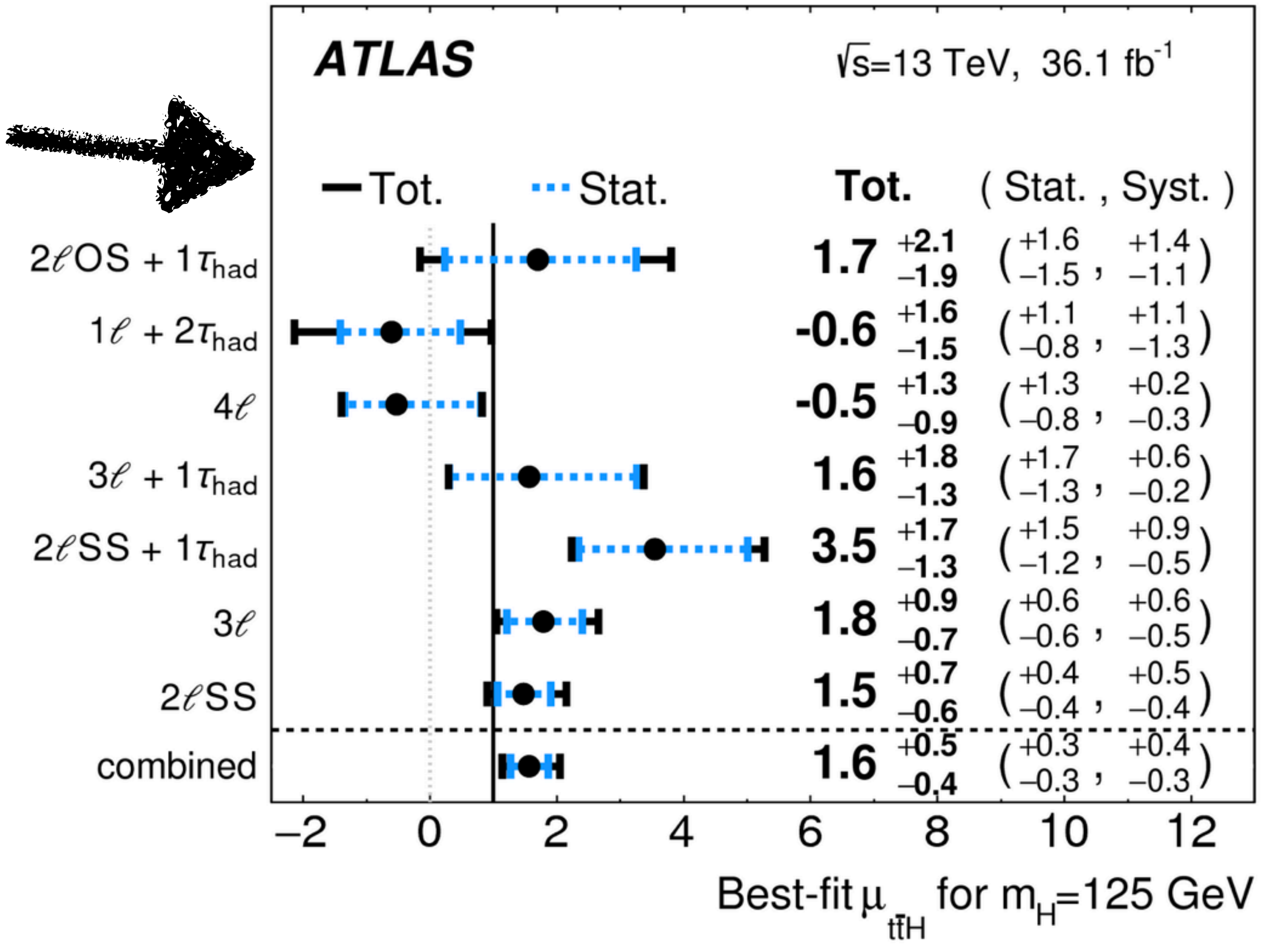


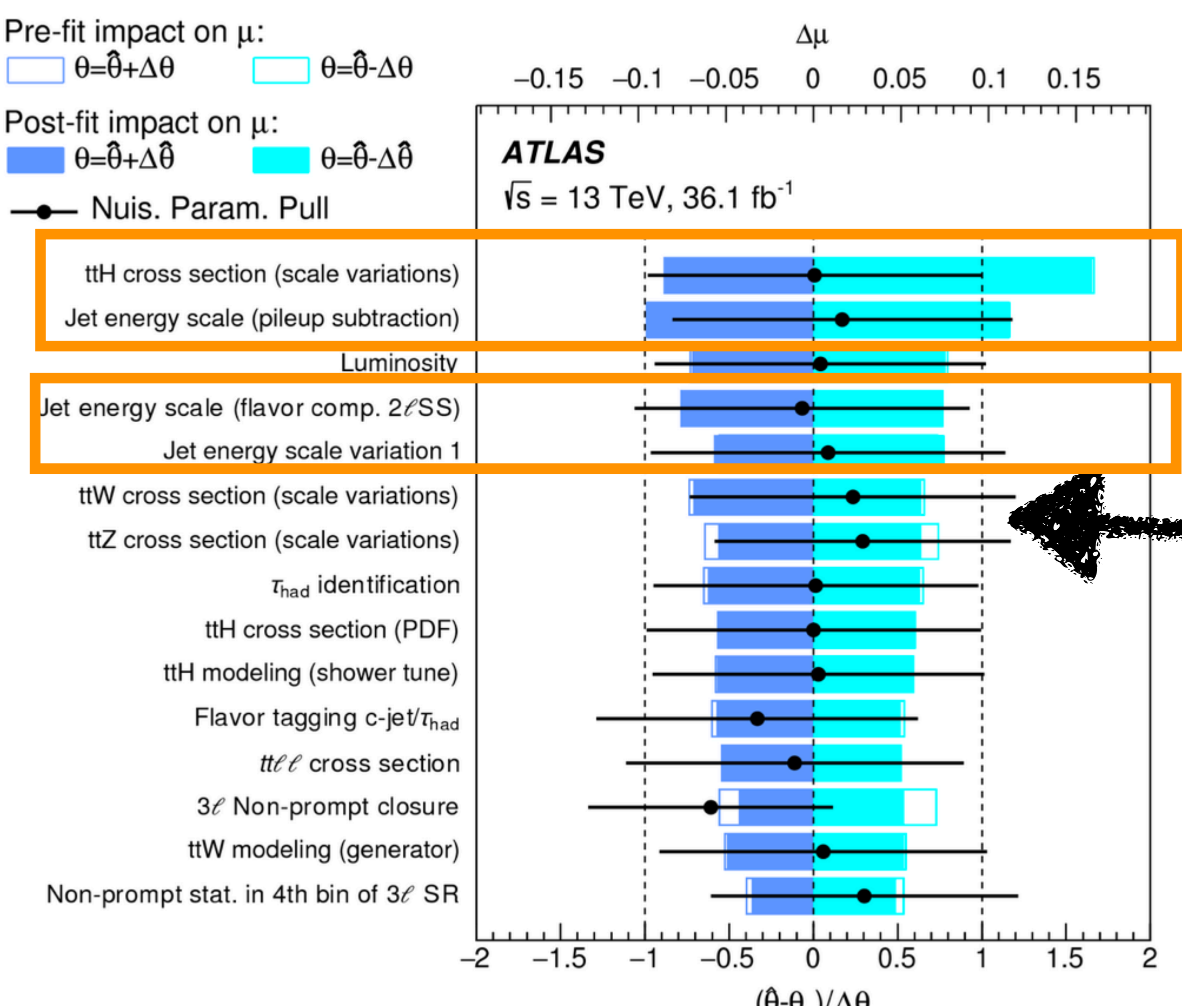


✔ Signal strength μ_{ttH} =1.6±0.5/0.4

✔ ttH signal significance in the multilpeton final state: 4.1σ (expected 2.8σ)

✔ Good compatibility among channels (34%)





✔ Measurement is dominated by systematics uncertainties

- dominant uncertainty related to data statistics (45% of total uncertainty)
- ttH modeling
- experimental uncertainties (jet energy scale, resolutions, b-tagging)
- non-prompt lepton estimate

ttH combination: results

✓ Measurements combined:

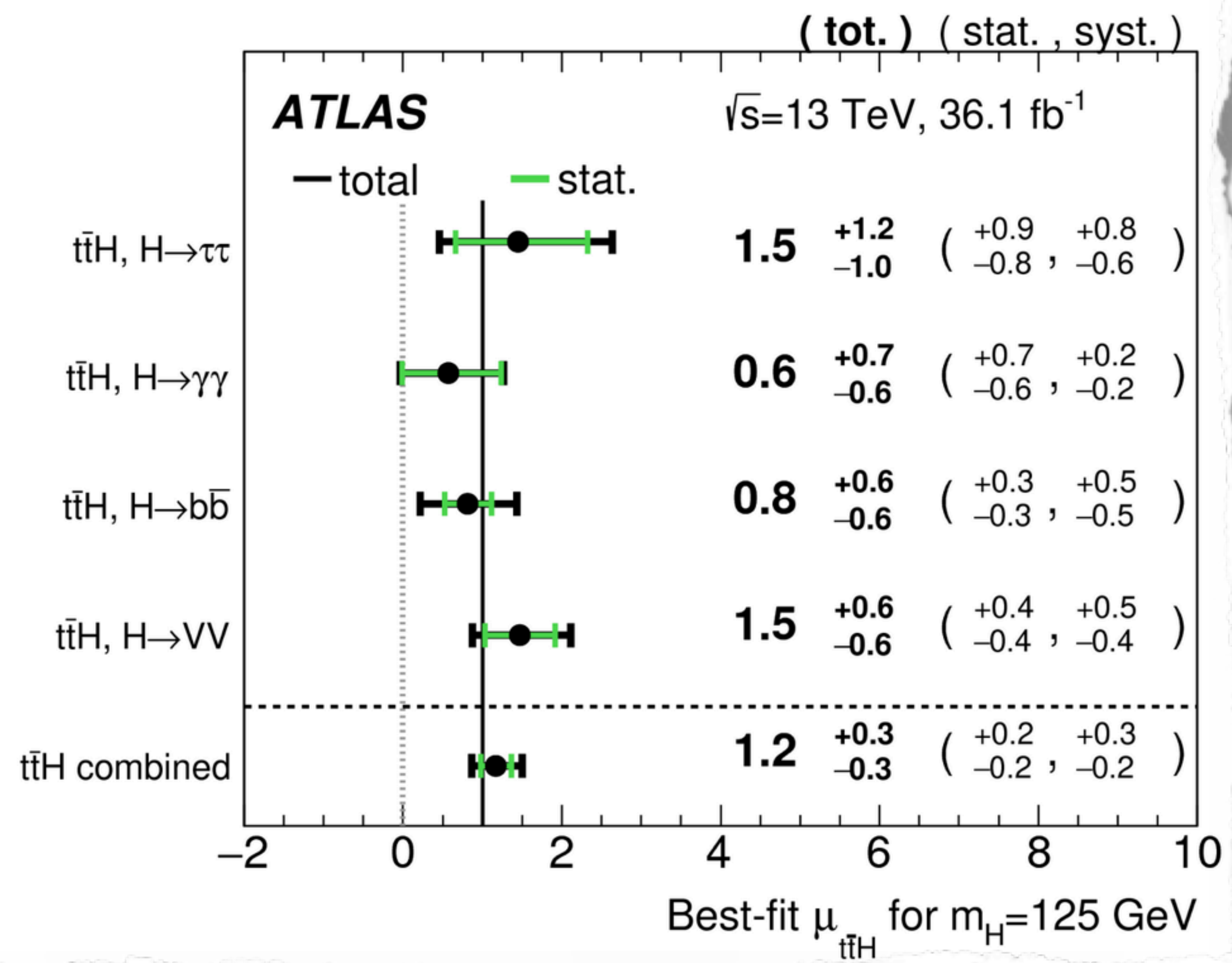
- ttH(H→bb)
- ttH(H→Multileptons)
- H→γγ and H→ZZ*→4l (ttH-enriched categories)
- very good compatibility among channels (38%)

✓ Best-fit results (combination)

- μ_{ttH}=1.17±0.19(stat)±0.27/0.23 (sys)
- σ_{ttH}=590±155 pb (sys)

✓ Evidence of ttH process in combination

- 3.8σ expected (μ_{SM}=1),
- 4.2σ observed ($\mu_{best-fit}$=1.17)



Channel	Best-fit μ Observed	Best-fit μ Expected	Significance Observed	Significance Expected
Multilepton	$1.6\ ^{+0.5}_{-0.4}$	$1.0\ ^{+0.4}_{-0.4}$	4.1σ	2.8σ
$H \to b\bar{b}$	$0.8\ ^{+0.6}_{-0.6}$	$1.0\ ^{+0.6}_{-0.6}$	1.4σ	1.6σ
$H \to \gamma\gamma$	$0.6\ ^{+0.7}_{-0.6}$	$1.0\ ^{+0.8}_{-0.6}$	0.9σ	1.7σ
$H \to 4\ell$	< 1.9	$1.0\ ^{+3.2}_{-1.0}$	—	0.6σ
Combined	$1.2\ ^{+0.3}_{-0.3}$	$1.0\ ^{+0.3}_{-0.3}$	4.2σ	3.8σ

ttH combination: uncertainties

✓ Combined ttH measurement is systematics-limited (H→γγ, H→ZZ still limited by data statistics)

- most of the uncertainty comes from tt+HF modeling in H→bb (20% of the total uncertainty) and ttH signal modeling for H→bb and H→Multilepton (9% of the total uncertainty)
- simulation statistics is still an issue for both channels
- experimental uncertainties are mostly dominated by jet energy scale and jet-flavour tagging

Uncertainty Source	$\Delta\mu$	
$t\bar{t}$ modeling in $H \to b\bar{b}$ analysis	+0.15	−0.14
$t\bar{t}H$ modeling (cross section)	+0.13	−0.06
Non-prompt light-lepton and fake τ_{had} estimates	+0.09	−0.09
Simulation statistics	+0.08	−0.08
Jet energy scale and resolution	+0.08	−0.07
$t\bar{t}V$ modeling	+0.07	−0.07
$t\bar{t}H$ modeling (acceptance)	+0.07	−0.04
Other non-Higgs boson backgrounds	+0.06	−0.05
Other experimental uncertainties	+0.05	−0.05
Luminosity	+0.05	−0.04
Jet flavor tagging	+0.03	−0.02
Modeling of other Higgs boson production modes	+0.01	−0.01
Total systematic uncertainty	+0.27	−0.23
Statistical uncertainty	+0.19	−0.19
Total uncertainty	+0.34	−0.30

✦ Large reduction on tt+HF uncertainty in H→bb final state needed to achieve better precision!

Wrapping-up and conclusions

Wrapping-up and conclusions

✓ Search for ttH(H→bb) in ATLAS with 2015+2016 Run 2 data at LHC

- observed signal significance 1.4σ
- measurement is dominated by the tt+HF (especially tt+≥1b) uncertainty

✓ b-tagging is a fundamental tool to achieve excellent performance in reconstruction and identification of b-jets in the physics analysis

- several improvements and new features contribute in large gain in background rejection and signal efficiency

✓ Evidence of the ttH process (4.2σ observed) when combining ttH(H→bb), ttH(H→Multilepton), H→γγ, H→ZZ (ttH-enriched categories) □□□

Wrapping-up and conclusions

✓ Search for ttH($H \rightarrow bb$) in ATLAS with 2015+2016 Run 2 data at LHC

- observed signal significance 1.4σ
- measurement is dominated by the tt+HF (especially tt+≥1b) uncertainty

✓ b-tagging is a fundamental tool to achieve excellent performance in reconstruction and identification of b-jets in the physics analysis

- several improvements and new features contribute in large gain in background rejection and signal efficiency

✓ Evidence of the ttH process (4.2σ observed) when combining ttH($H \rightarrow bb$), ttH($H \rightarrow$ Multilepton), $H \rightarrow \gamma\gamma$, $H \rightarrow ZZ$ (ttH-enriched categories)

...looking ahead for the ttH observation with the final 2015-2018 Run 2 data!

Wrapping-up and conclusions

✓ Search for ttH(H→bb) in ATLAS with 2015+2016 Run 2 data at LHC

- observed signal significance 1.4σ
- measurement is dominated by the tt+HF (especially tt+≥1b) uncertainty

✓ b-tagging is a fundamental tool to achieve excellent performance in reconstruction and identification of b-jets in the physics analysis

- several improvements and new features contribute in large gain in background rejection and signal efficiency

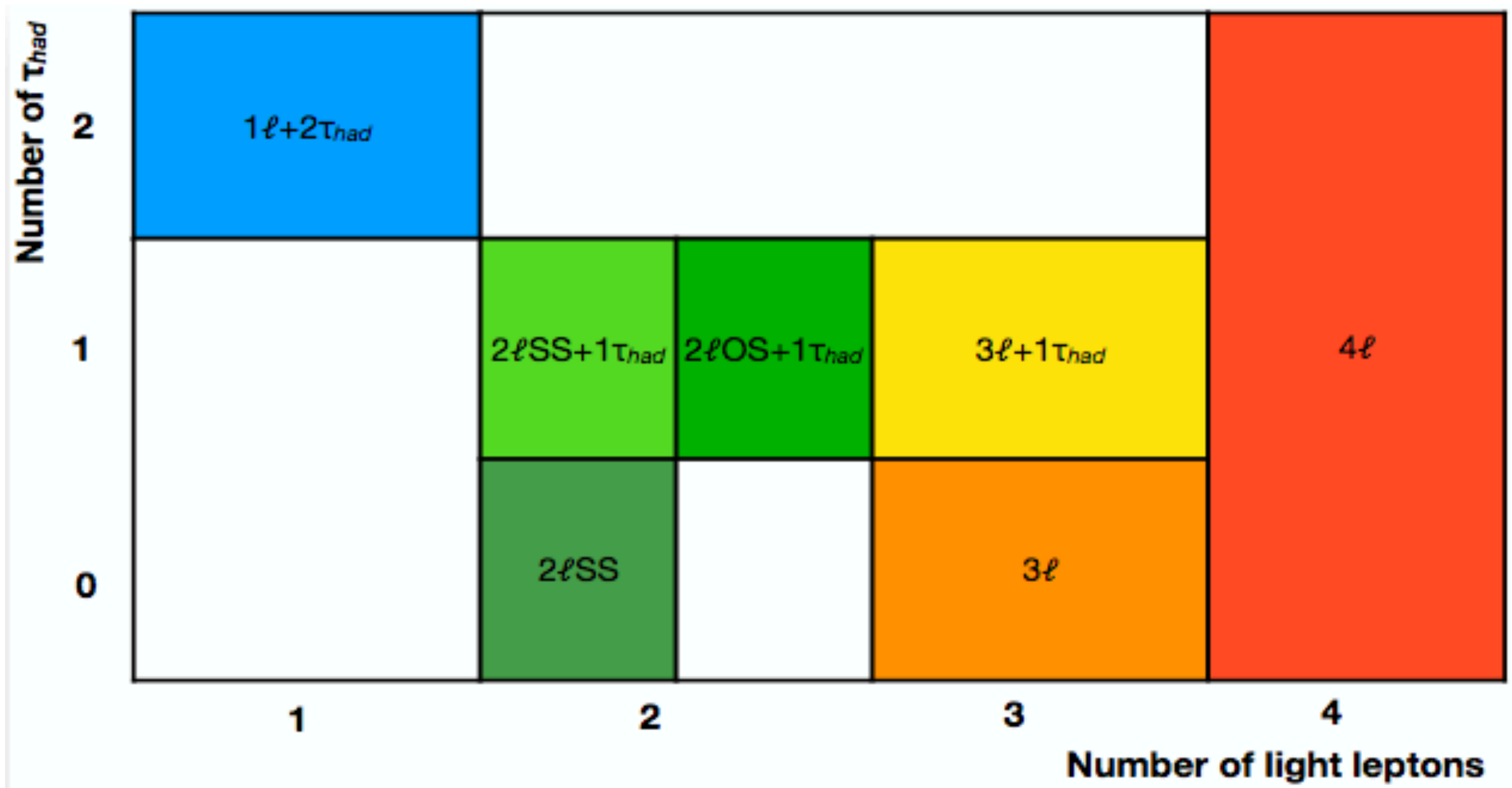
✓ Evidence of the ttH process (4.2σ observed) when combining ttH(H→bb), ttH(H→Multilepton), H→γγ, H→ZZ (ttH-enriched categories)

...looking ahead for the ttH observation with the final 2015-2018 Run 2 data!

Thank you for your attention!

Additional slides

$t\bar{t}H(H\to ZZ^*, WW^*, \tau\tau)$ – strategy



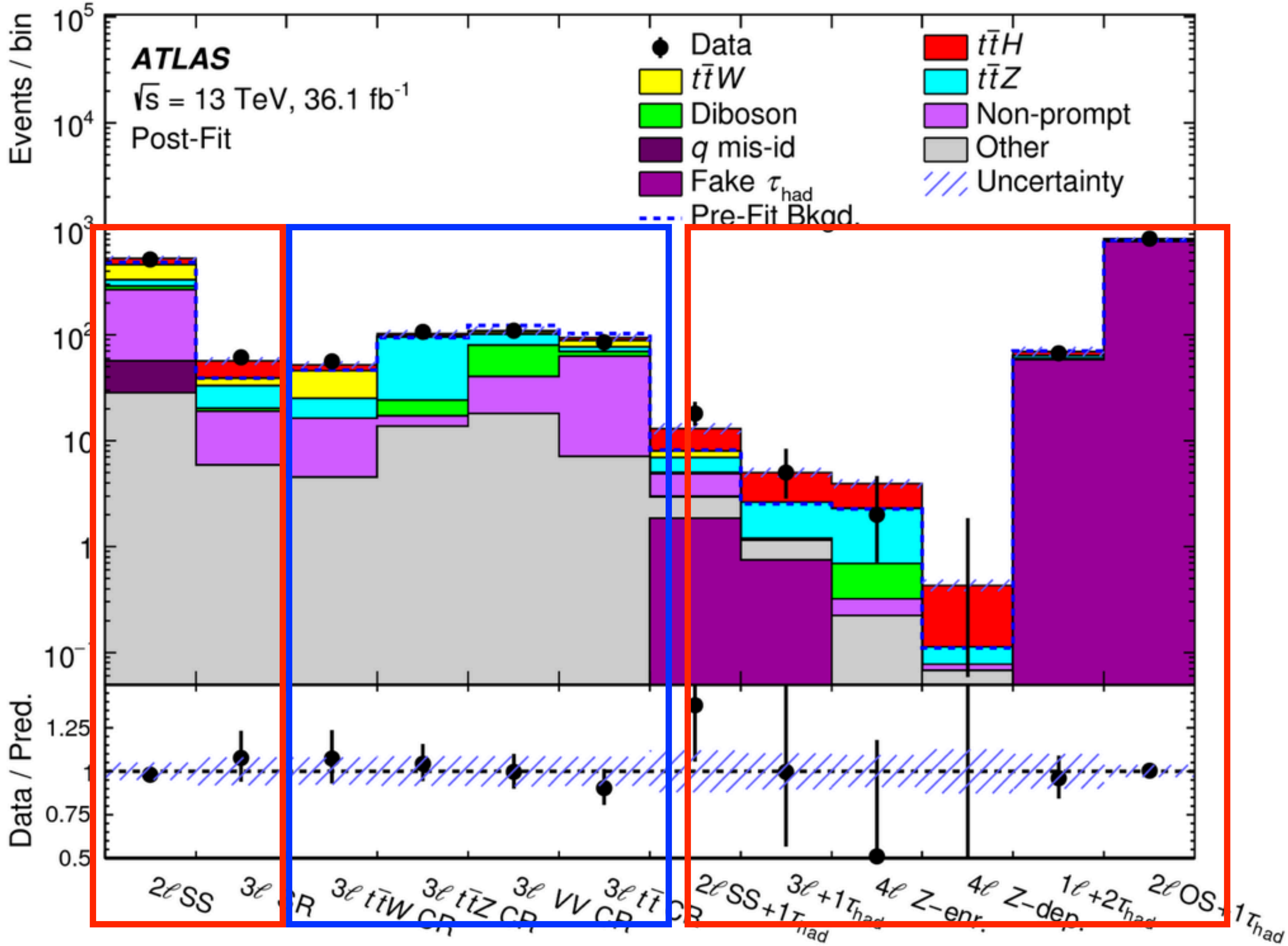
- L — loose
- $L^\dagger$ — L isolated
- L^* — $L^\dagger$ with non-prompt clean.
- M — medium
- T — tight
- T^* — very tight

	2ℓSS	3ℓ	4ℓ	$1\ell+2\tau_{\mathrm{had}}$	$2\ell\mathrm{SS}+1\tau_{\mathrm{had}}$	$2\ell\mathrm{OS}+1\tau_{\mathrm{had}}$	$3\ell+1\tau_{\mathrm{had}}$
Light lepton	2T*	1L*, 2T*	2L, 2T	1T	2T*	$2L^\dagger$	$1L^\dagger$, 2T
τ_{had}	0M	0M	–	1T, 1M	1M	1M	1M
N_{jets}, $N_{b-\mathrm{jets}}$	≥ 4, $=1,2$	≥ 2, ≥ 1	≥ 2, ≥ 1	≥ 3, ≥ 1	≥ 4, ≥ 1	≥ 3, ≥ 1	≥ 2, ≥ 1

ttH(H->ZZ*, WW*, $\tau\tau$) - backgrounds

- ✓ Prompt-leptons or T-jets estimated from MC
  - irreducible: ttW, ttZ and diboson
- ✓ Electron charge misidentification
  - data-driven estimate from misidentification rate in $Z\rightarrow e^+e^-$ vs $Z\rightarrow e^+e^+/Z\rightarrow e^-e^-$
- ✓ Fake or non-prompt light leptons
  - semileptonic b-hadron decays and photon conversions
  - data-driven estimation
- ✓ Fake hadronic taus
  - light-flavour jets and electron misidentified as taus
  - data-driven estimation in CR; extrapolation to SR
- ✓ New important reconstruction techniques
  - lepton reconstruction
  - BDT to mitigate charge misidentification
  - BDT to mitigate non-prompt e/μ

ttH(H->ZZ*, WW*, τ τ) - fits

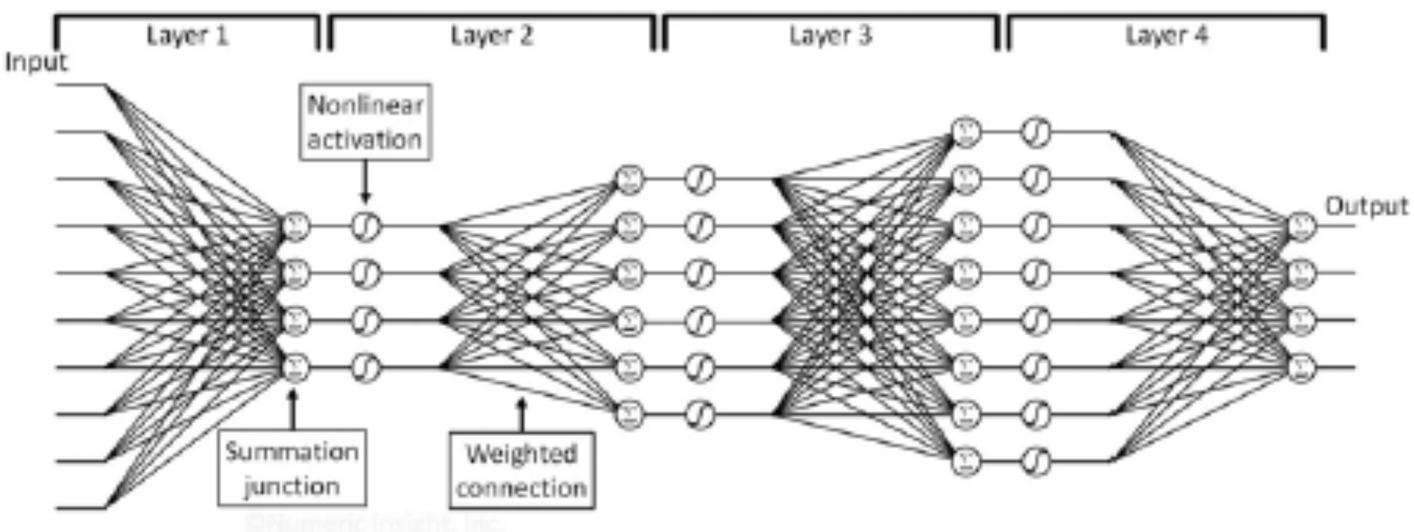


✓ 8 signal regions and 4 control regions treated with BDT shape or 1-bin (BDT trained against dominant background of a given region)

Uncertainty Source	$\Delta\mu$	
$t\bar{t}H$ modeling (cross section)	$+0.20$	-0.09
Jet energy scale and resolution	$+0.18$	-0.15
Non-prompt light-lepton estimates	$+0.15$	-0.13
Jet flavor tagging and τ_{had} identification	$+0.11$	-0.09
$t\bar{t}W$ modeling	$+0.10$	-0.09
$t\bar{t}Z$ modeling	$+0.08$	-0.07
Other background modeling	$+0.08$	-0.07
Luminosity	$+0.08$	-0.06
$t\bar{t}H$ modeling (acceptance)	$+0.08$	-0.04
Fake τ_{had} estimates	$+0.07$	-0.07
Other experimental uncertainties	$+0.05$	-0.04
Simulation sample size	$+0.04$	-0.04
Charge misassignment	$+0.01$	-0.01
Total systematic uncertainty	$+0.39$	-0.30

Deep Learning

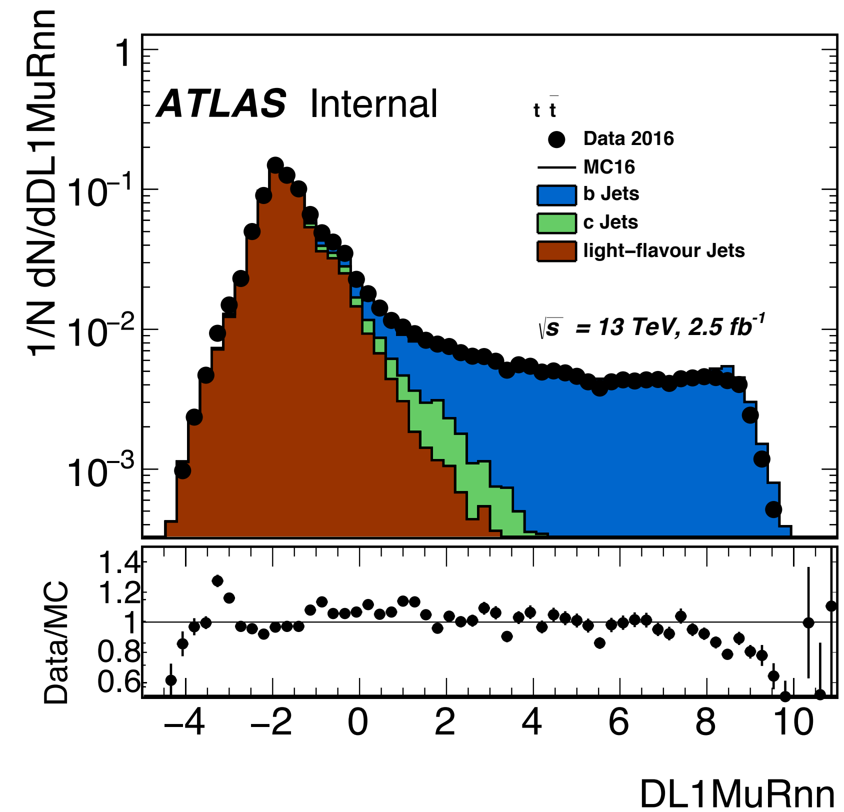
- Exploits the advantage of multivariate techniques with multiple output nodes - exploited for b-c tagging

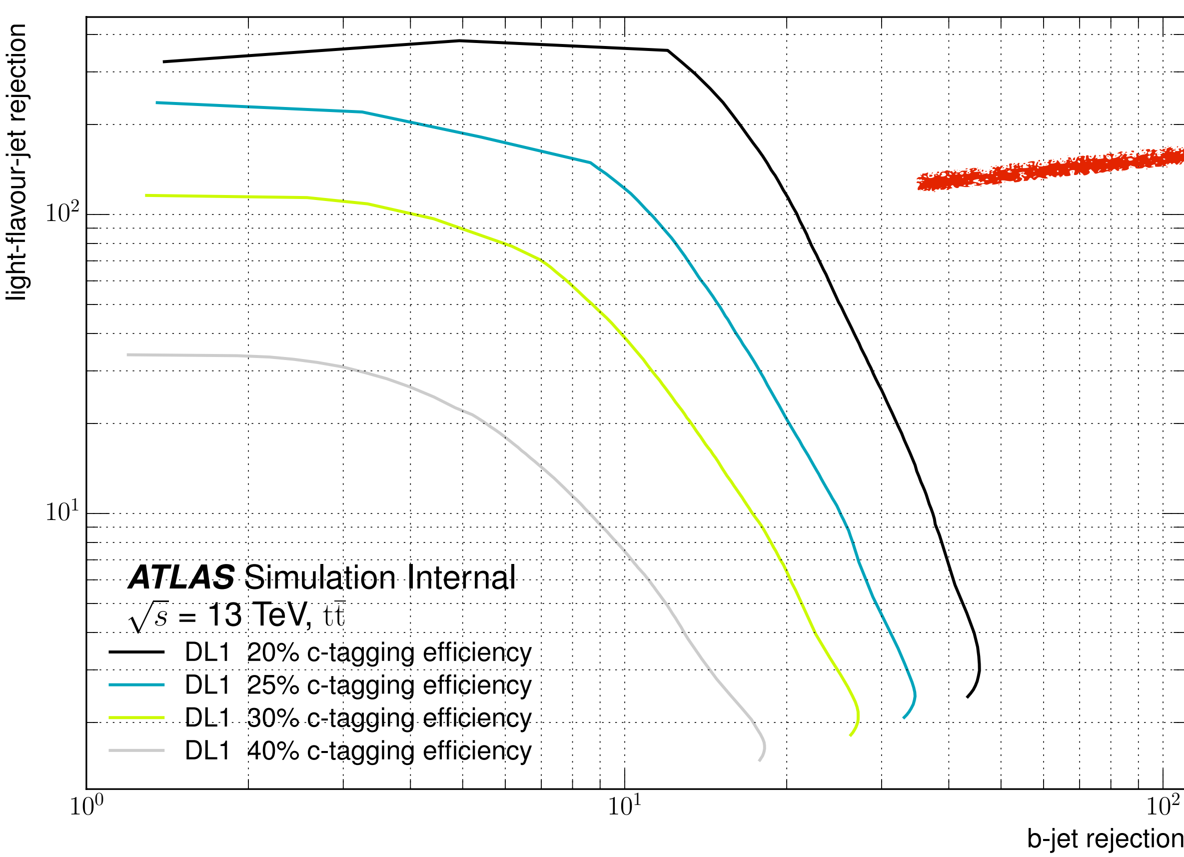


p_{light}

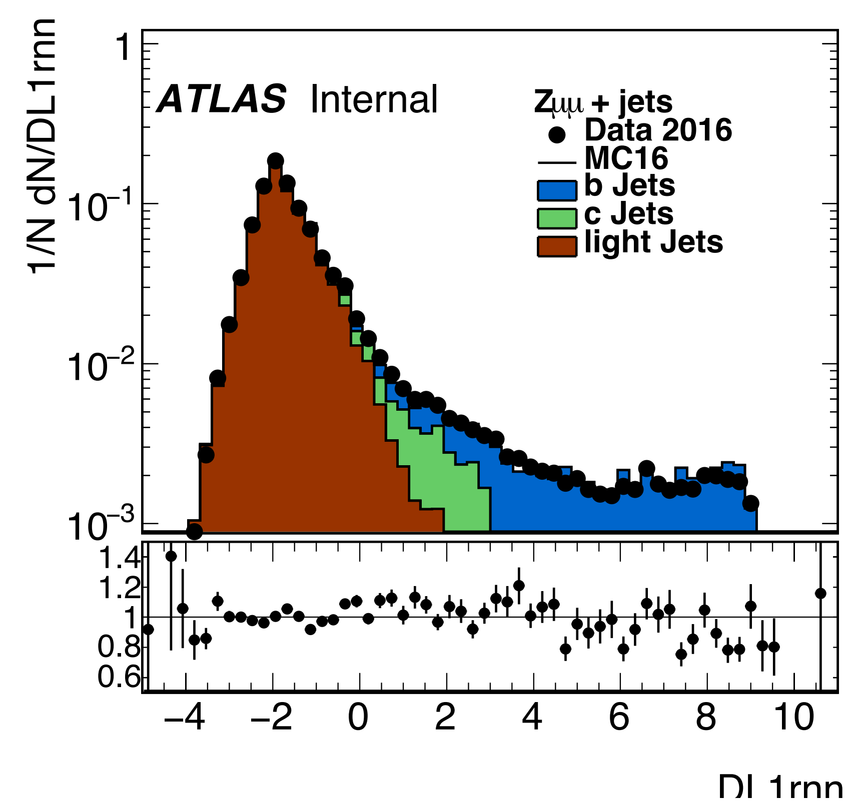
p_{charm}

p_{bottom}



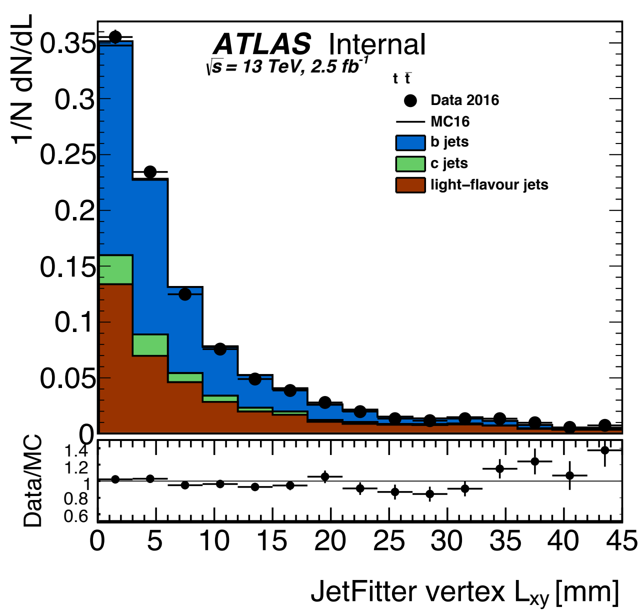


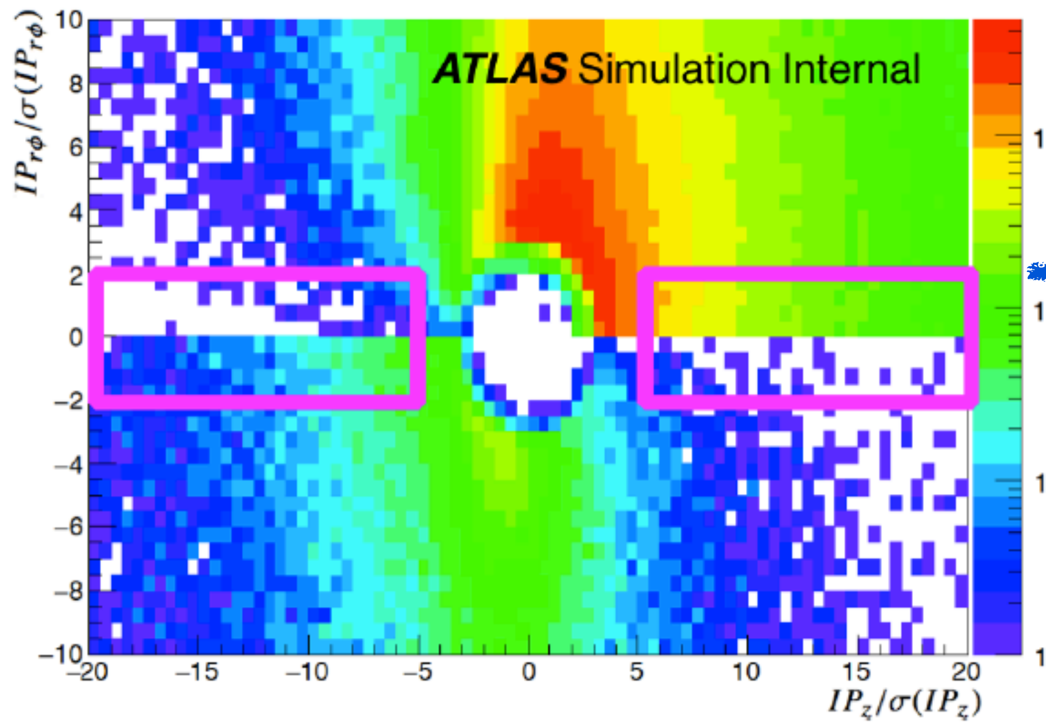
For a given training, c/l rejection can be optimized by varying the c-fraction at the evaluation stage

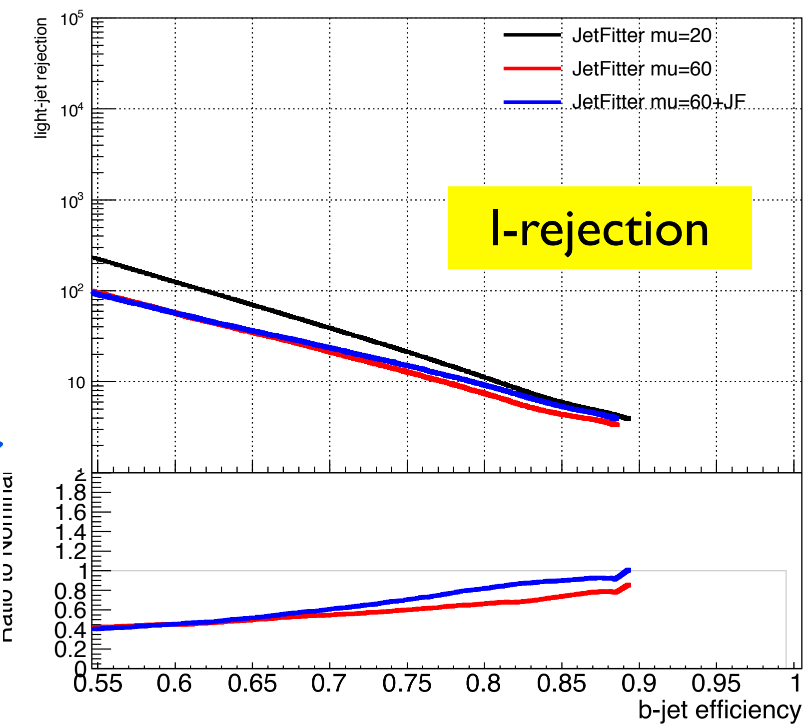


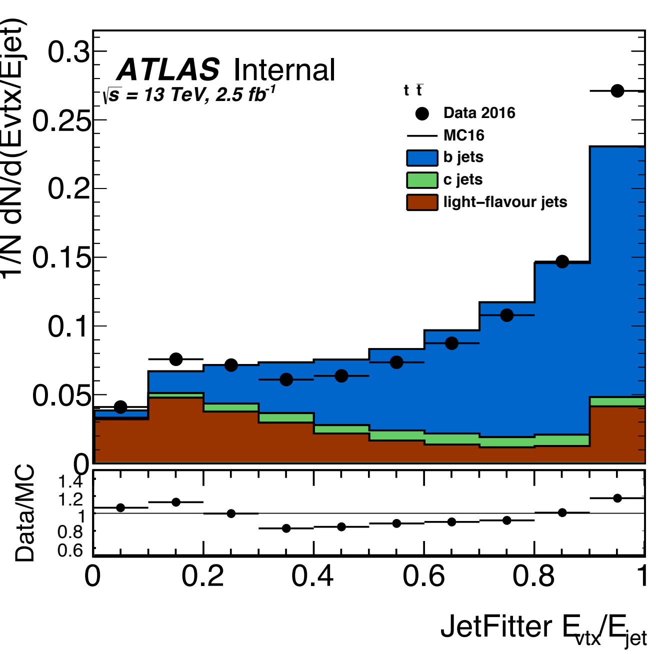
JetFitter

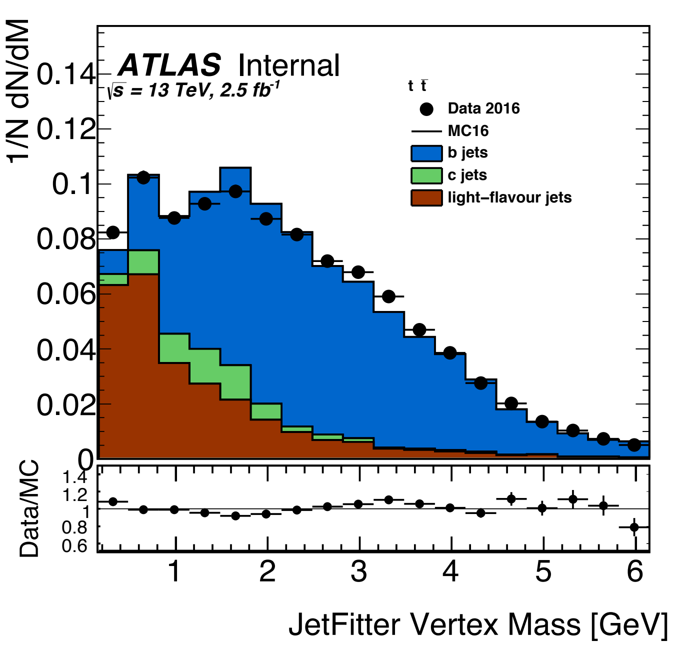
- Rejection of tracks from pile-up by applying a box cut on Sd0 vs Sz0 of JF tracks
  - large mitigation of performance degradation in high pile up environment
  - studies with dedicated high pile-up samples with μ=60, 80

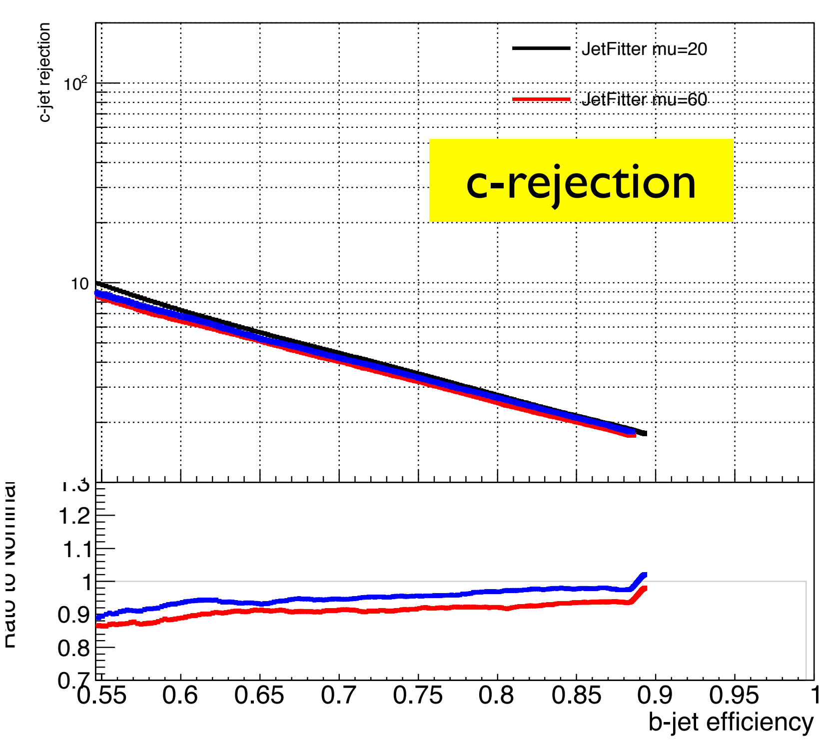










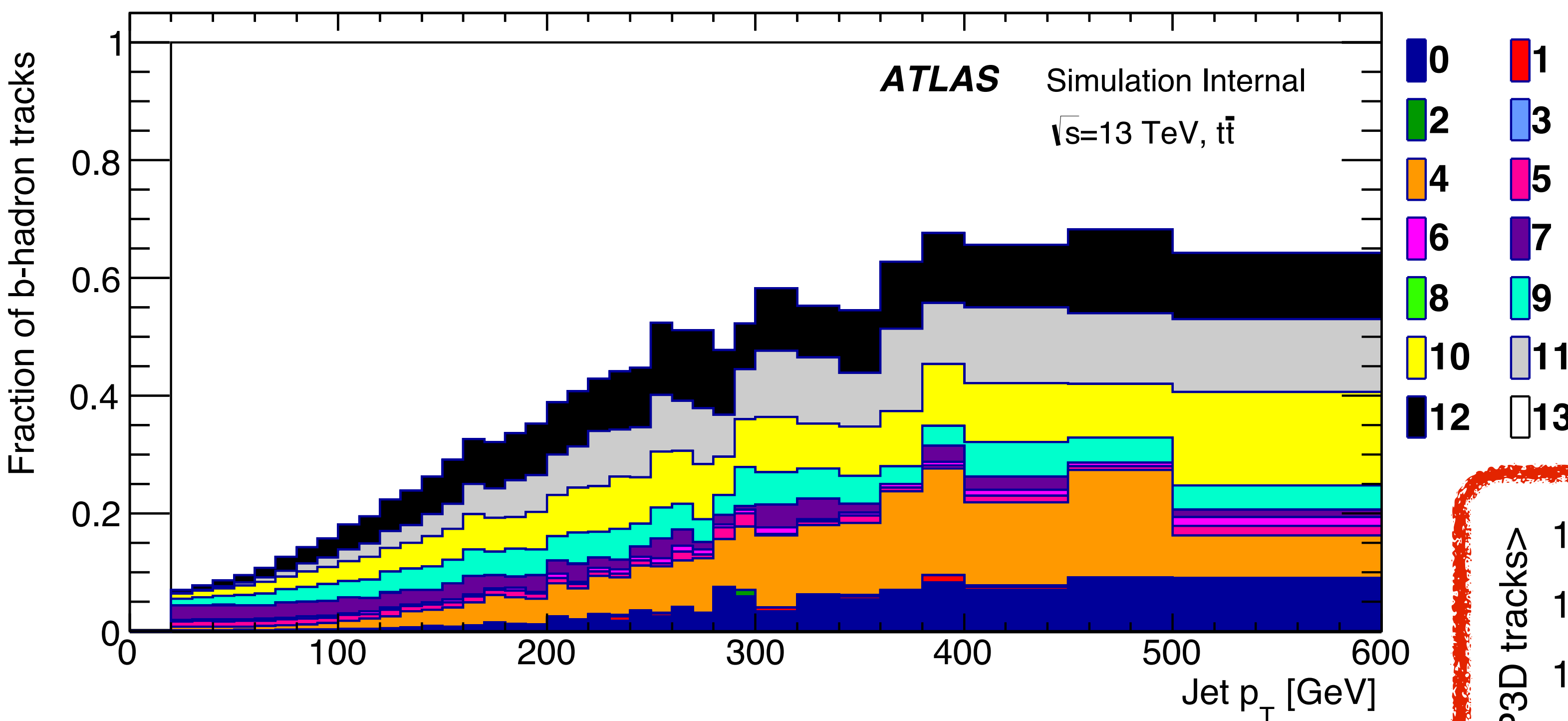


IPTag (4) - main changes compared to rel20.1

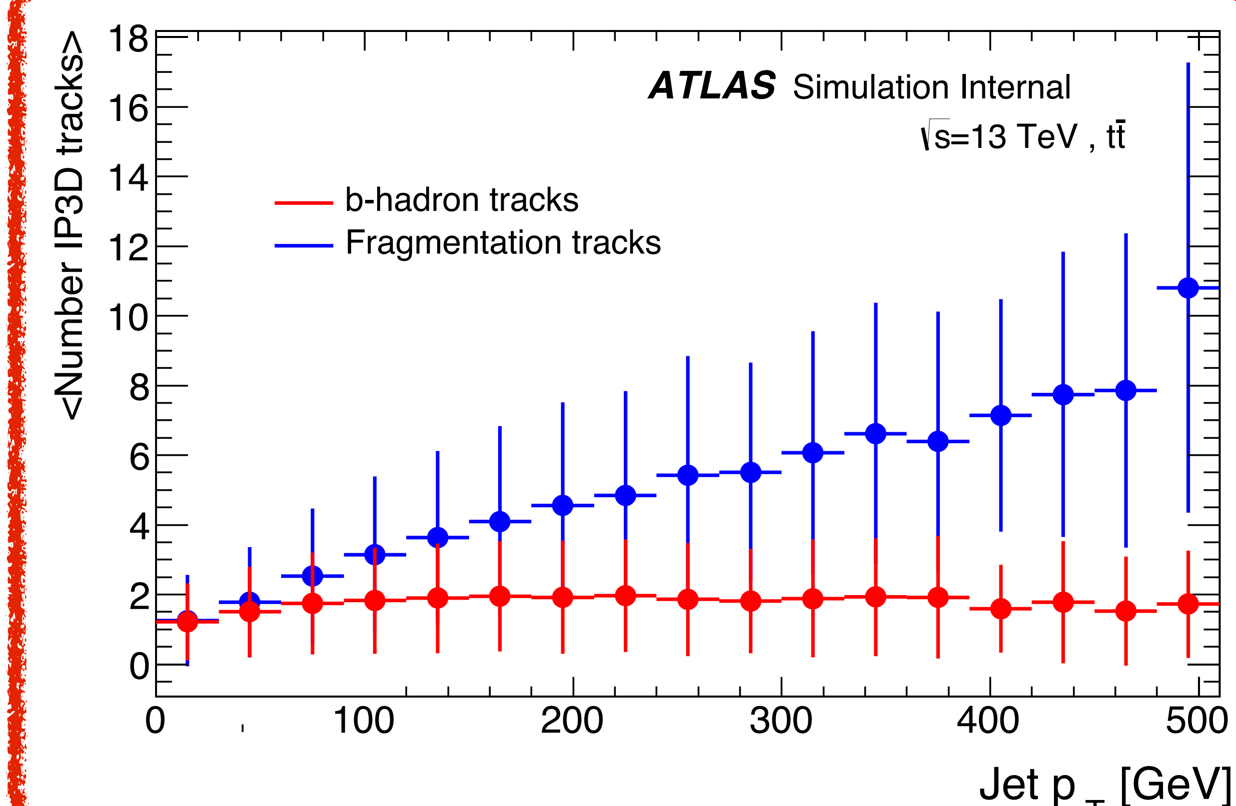
1. Requirement on the number of pixel hits relaxed from at least 2 to at least 1 (no inefficiencies in the high b-jet pt region)

2. Ignoring tracks from conversions, Ks, Λ and material interactions (SV output) → sizable gain in performance achieved (15% on light rejection @77%b-jet efficiency)

3. Reference histograms produced with a mixture of Z'→ttbar and ttbar for categories with no hits in IBL and b-layer (0 and 1)



- Average number of tracks selected for the IP algorithm as a function of jet pt
  - jet-fragmentation tracks dominant in medium-high b-jet pt region compared to b-hadron tracks

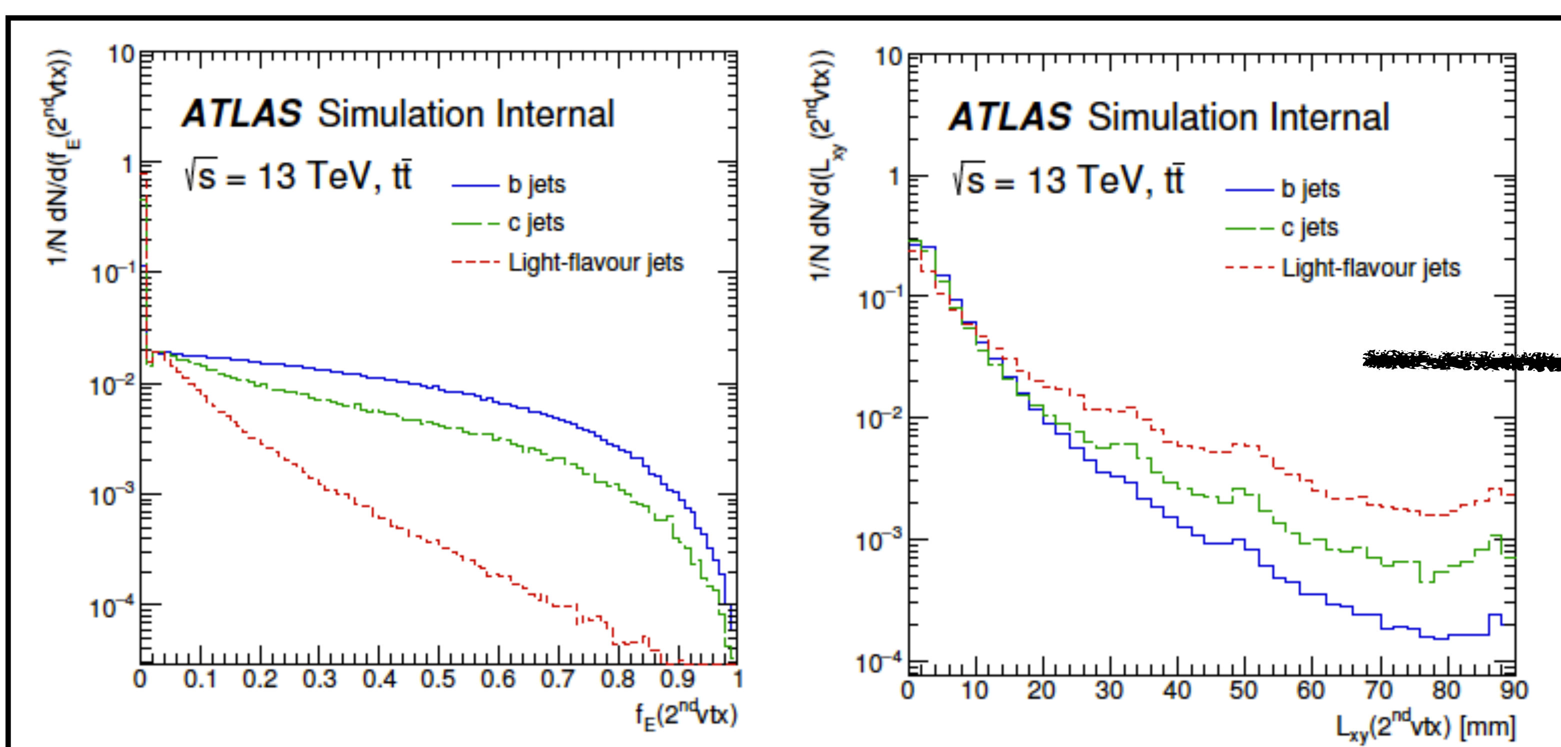


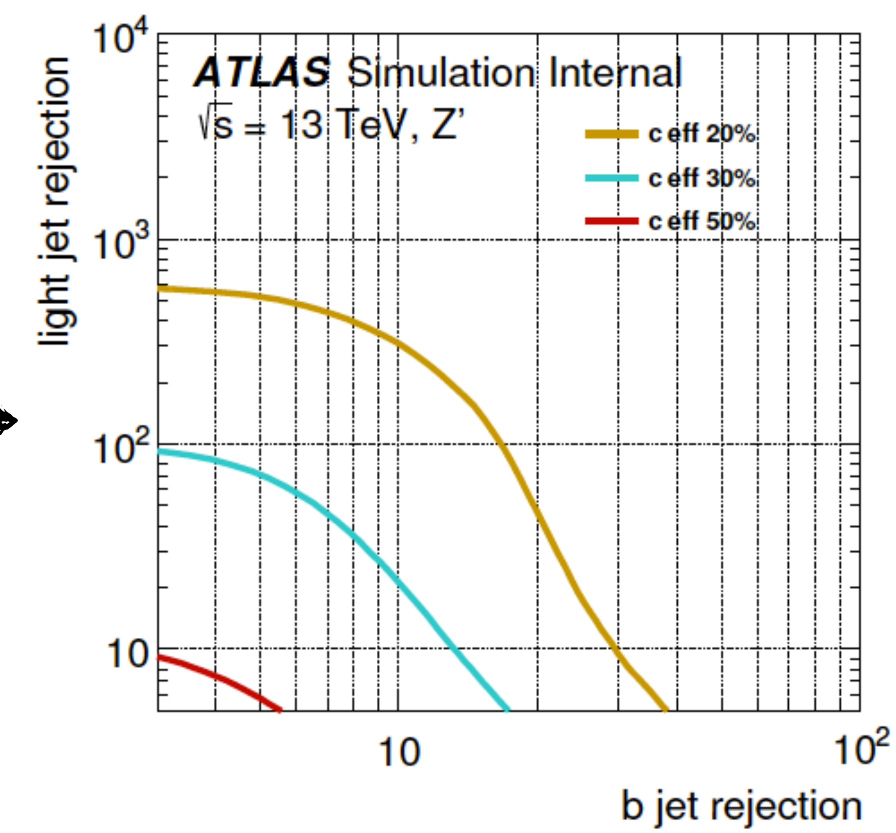
c-tagging

- Discrimination of c from b/light is very important for several physics studies

- Discrimination exploited by the topology and the kinematics of the displaced vertex reconstructed JetFitter - two taggers provided, MV2c100 (b/c discrimination), MV2cl100 (b/l discrimination)

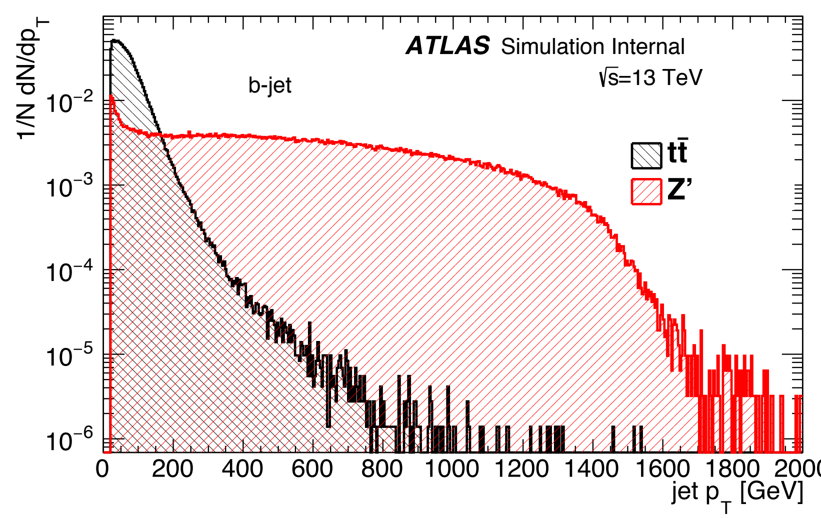
Variable Name	Description
L_{xyz}	Three-dimensional displacement of secondary vertex from the primary vertex
L_{xy}	Transverse displacement of the secondary vertex
$Y_{trk}^{min}, Y_{trk}^{max}, Y_{trk}^{avg}$	Min, Max and Avg. track rapidity of tracks in jet
$Y_{trk}^{min}, Y_{trk}^{max}, Y_{trk}^{avg}$ (2^{nd} vtx)	Min, Max and Avg. track rapidity of tracks at secondary vertex
m	Invariant mass of tracks associated to secondary vertex
E	Energy of charged tracks associated to secondary vertex
f_E	Energy fraction of charged tracks (from all tracks in the jet) associated to secondary vertex
N_{trk}	Number of tracks associated to the secondary vertex

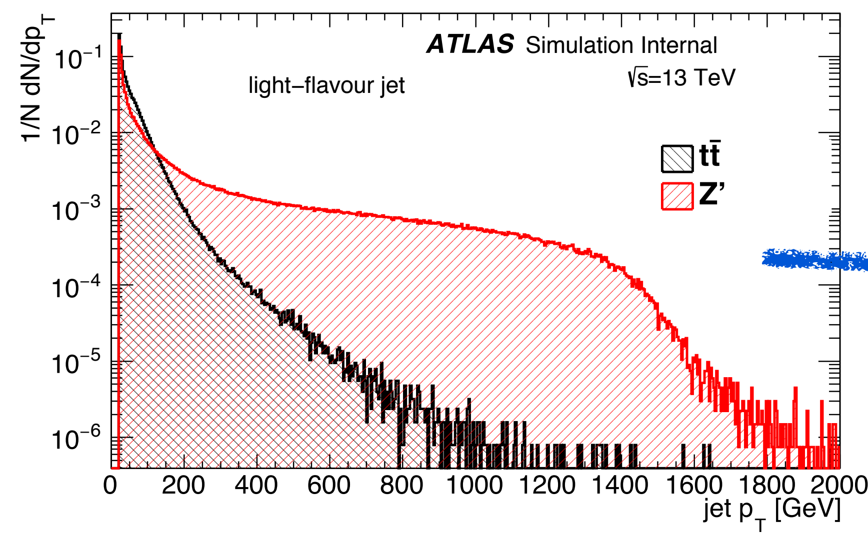


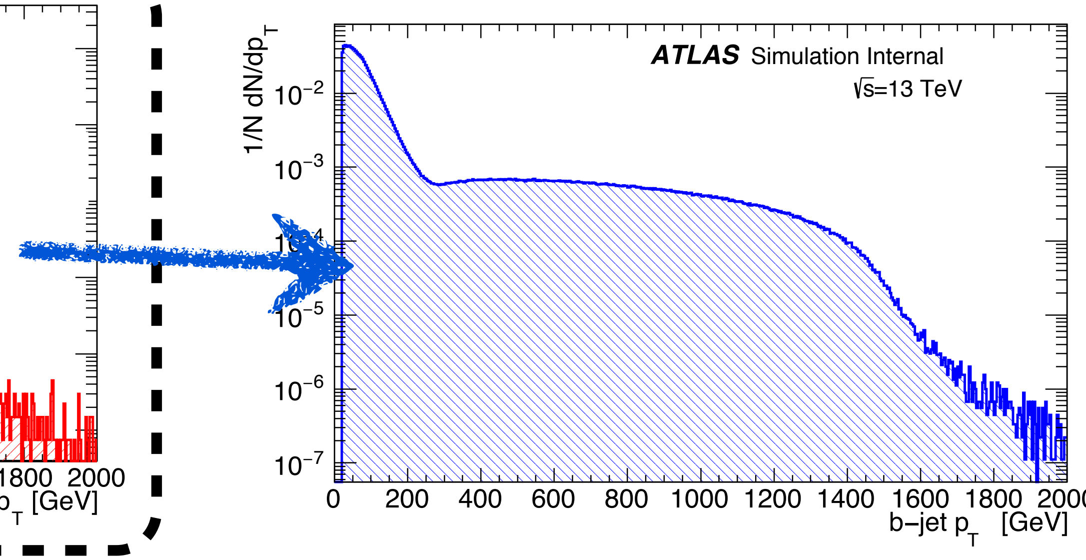


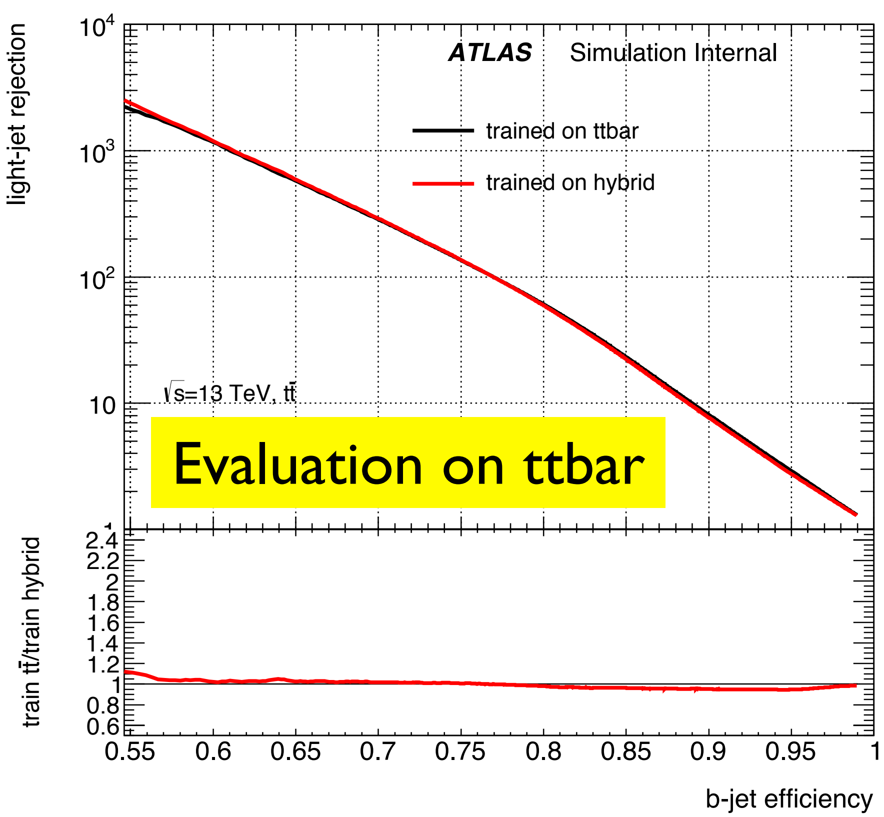
Algorithm training samples (2)

- New hybrid sample used for training of high level tagger algorithms
  - similar algorithm performance at low pt but significantly larger rejections at high pt

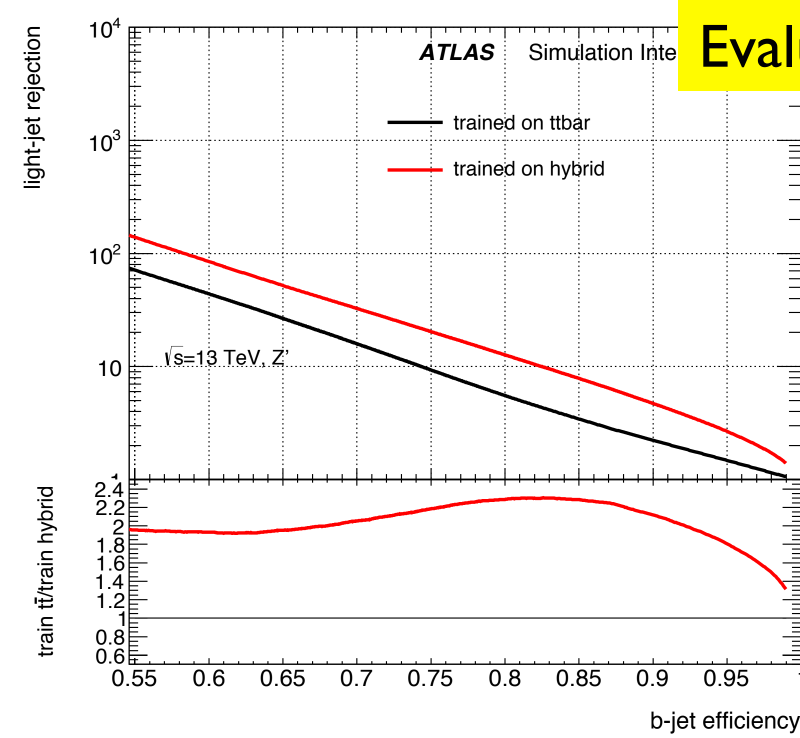








Evaluation on ttbar

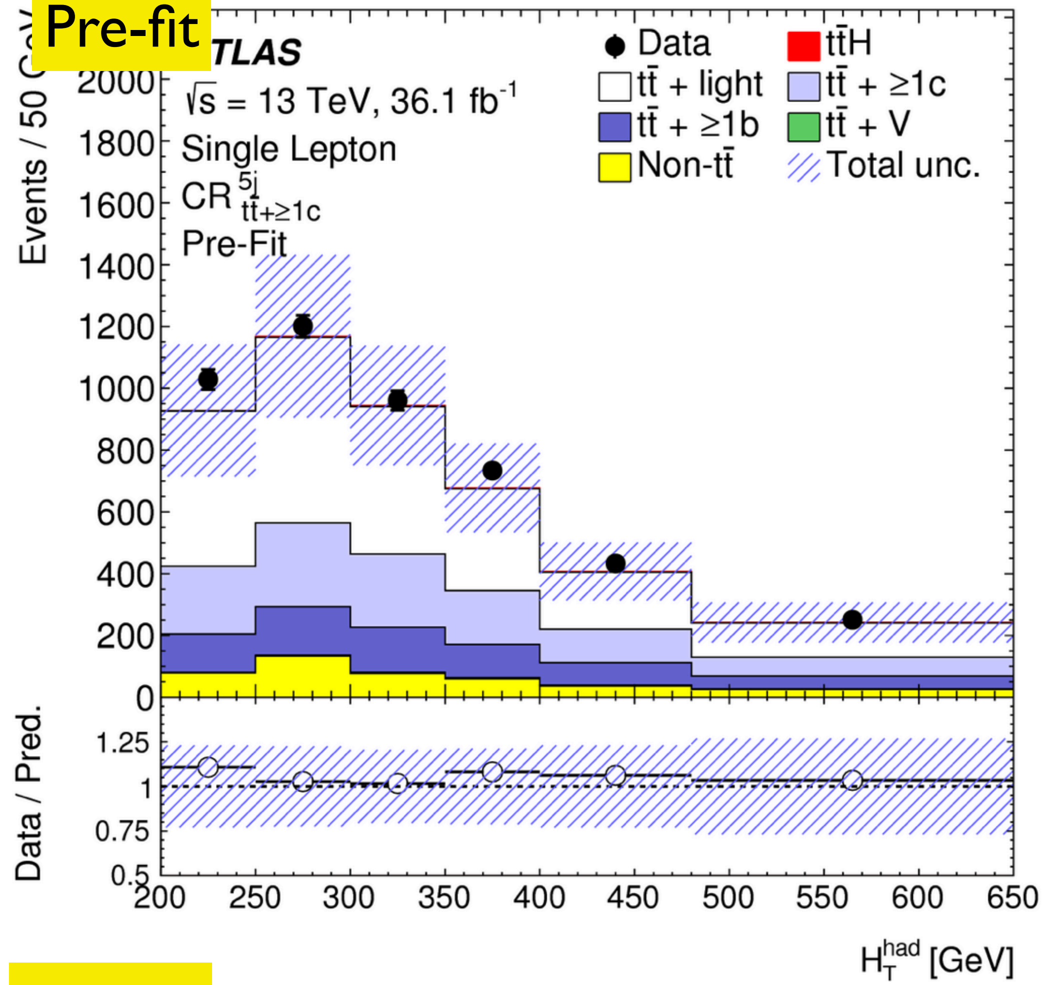


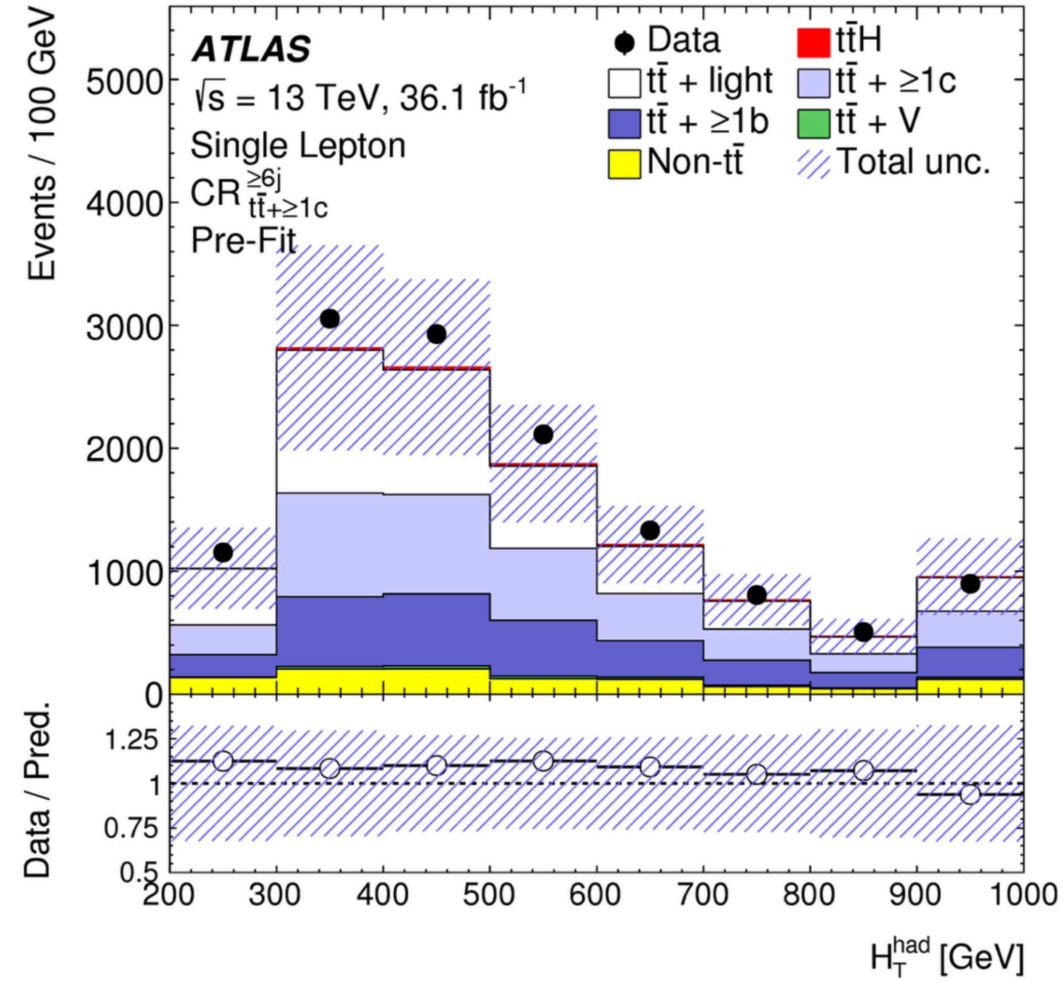
Evaluation on Z'

Large boost in performance in the high-pt phase space (factor 2 at 77 % b-eff WP)
Similar performance when evaluating on ttbar

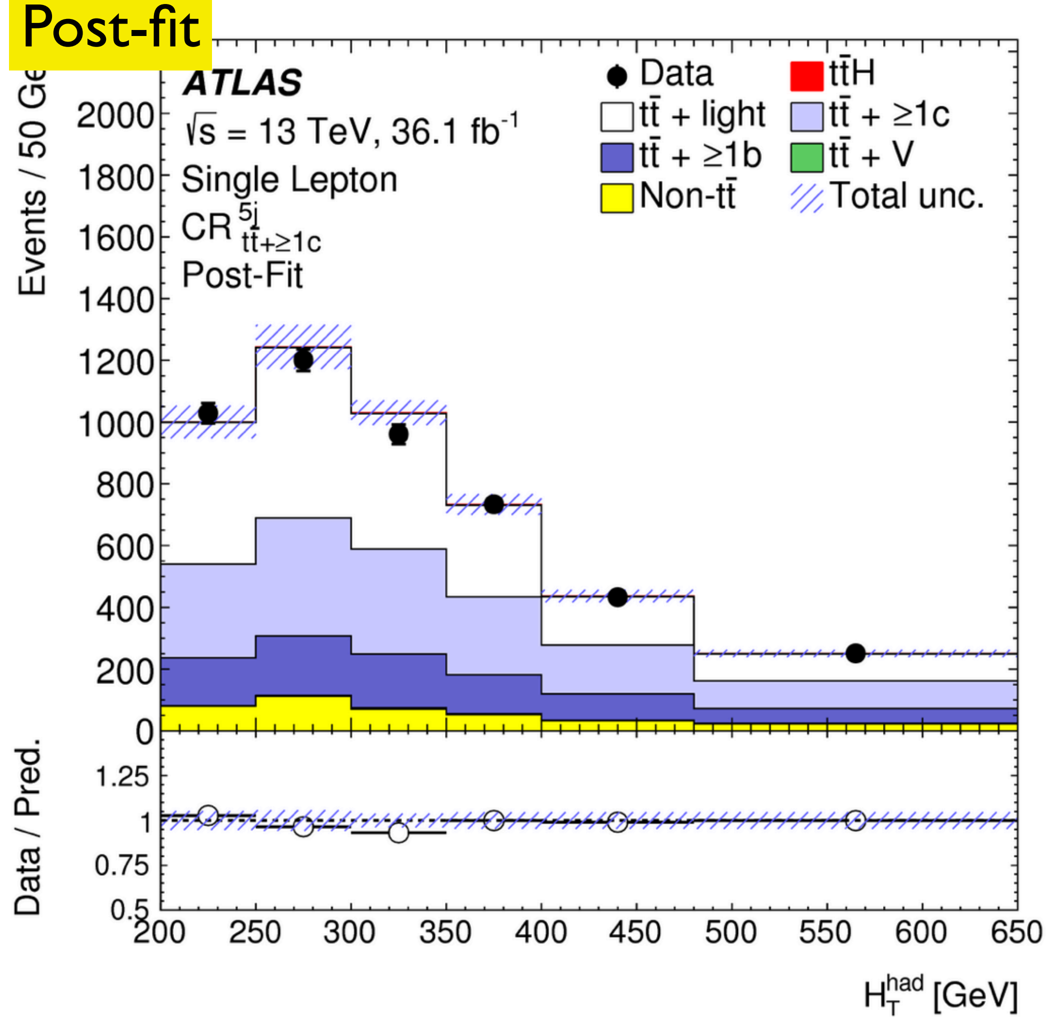
Ht fit

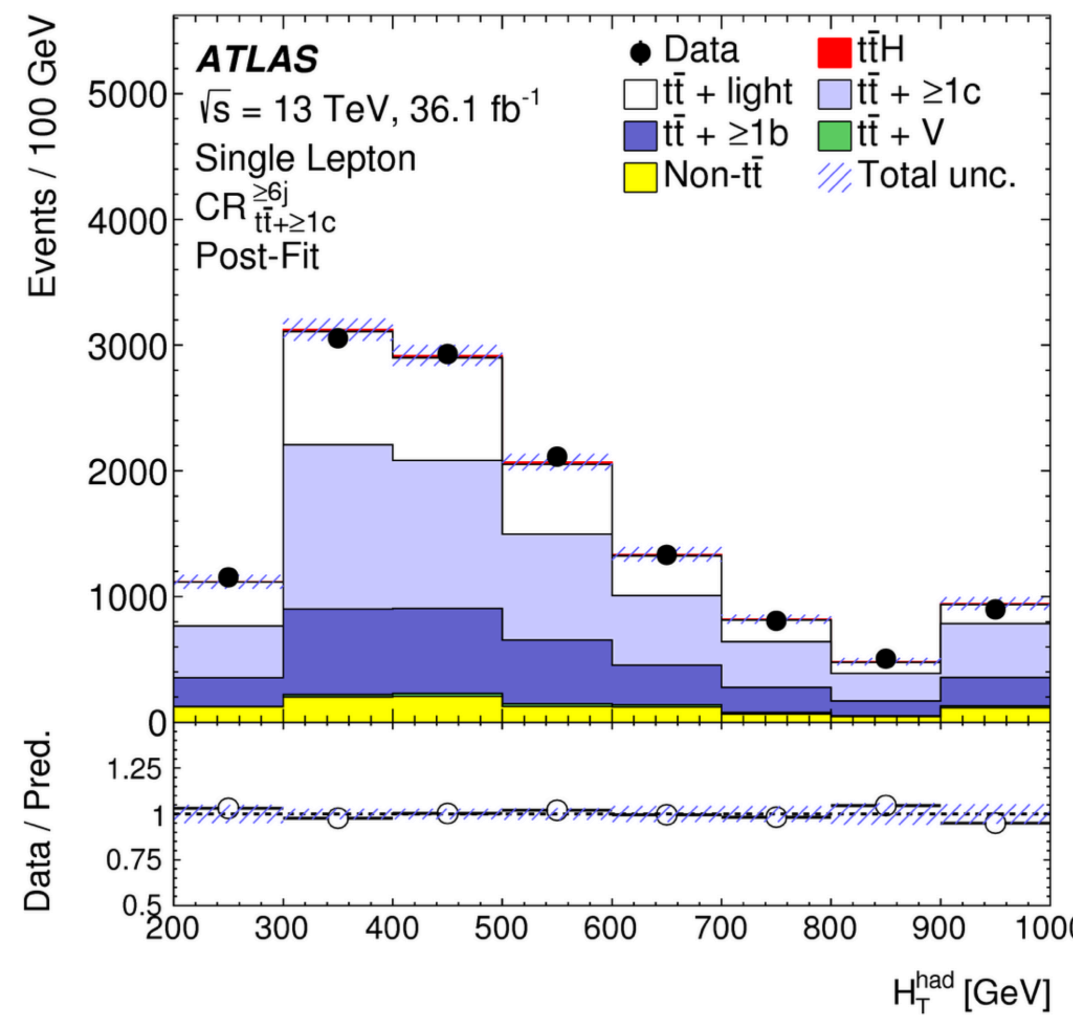
Pre-fit



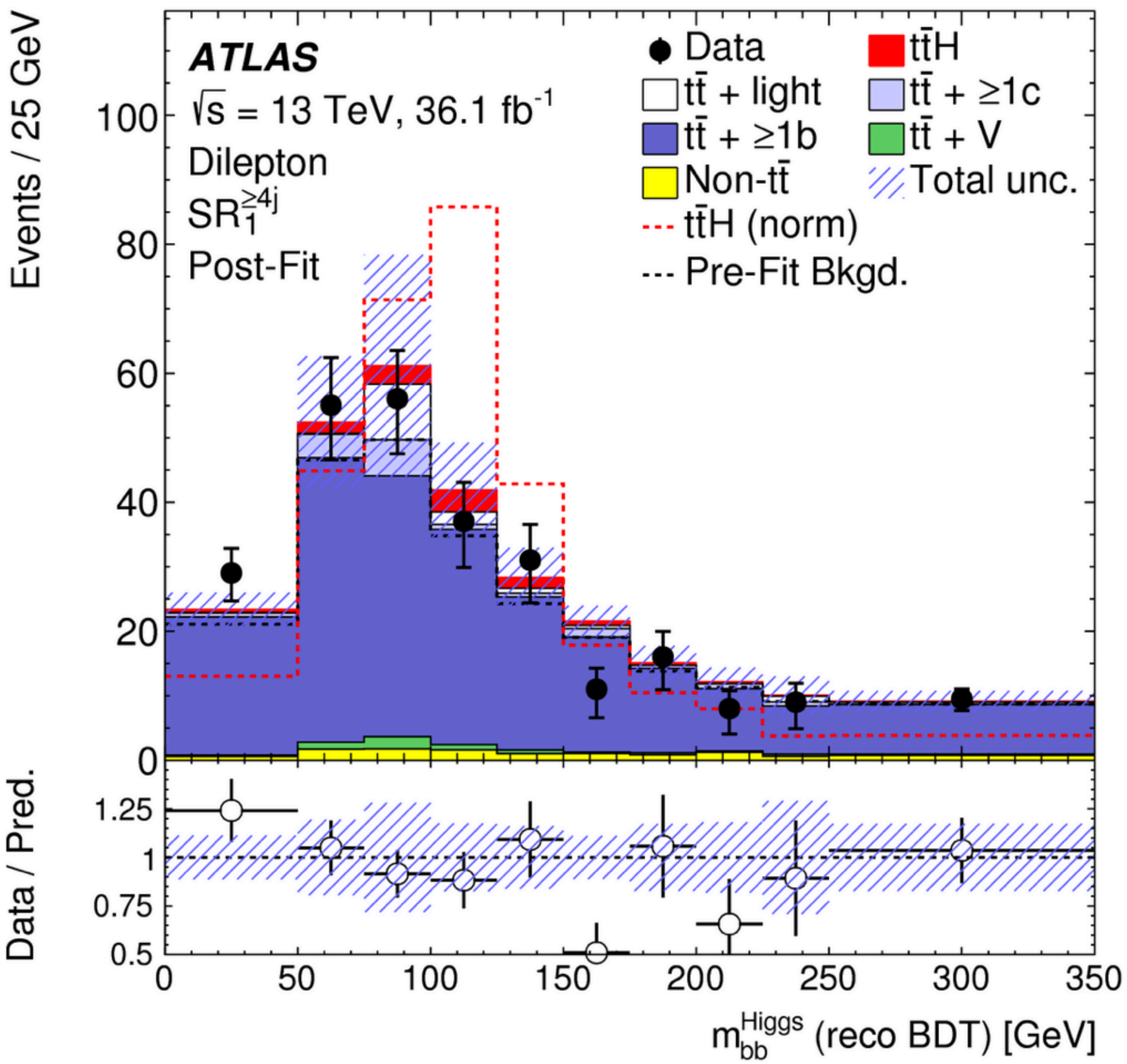


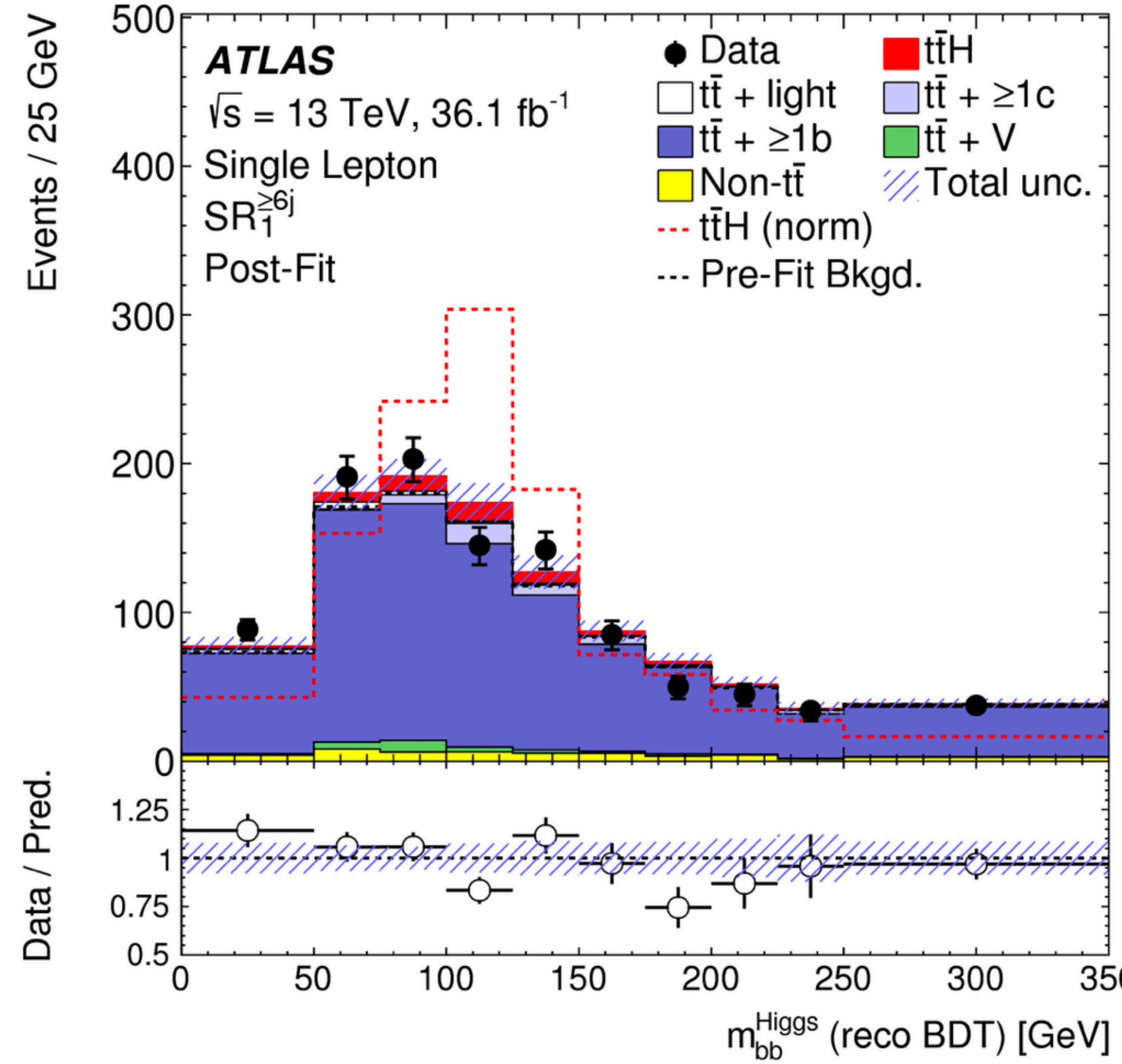
Post-fit





mbb


ATLAS
√s = 13 TeV, 36.1 fb-1
Dilepton
SR1≥4j
Post-Fit
Data
tt̄H
tt̄ + light
tt̄ + ≥1c
tt̄ + ≥1b
tt̄ + V
Non-tt̄
Total unc.
tt̄H (norm)
Pre-Fit Bkgd.
Events / 25 GeV
Data / Pred.
m_bb^Higgs (reco BDT) [GeV]


ATLAS
√s = 13 TeV, 36.1 fb-1
Single Lepton
SR1≥6j
Post-Fit
Data
tt̄H
tt̄ + light
tt̄ + ≥1c
tt̄ + ≥1b
tt̄ + V
Non-tt̄
Total unc.
tt̄H (norm)
Pre-Fit Bkgd.
Events / 25 GeV
Data / Pred.
m_bb^Higgs (reco BDT) [GeV]

Systematic sources

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 $H_{\mathrm{T}}/2$ 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%	$t\bar{t}+\geq 1b$