Between even and odd: probing the CP nature of the interaction between the Higgs boson and the top quark

LIP seminar
November 23rd 2023

Ana Luísa Carvalho

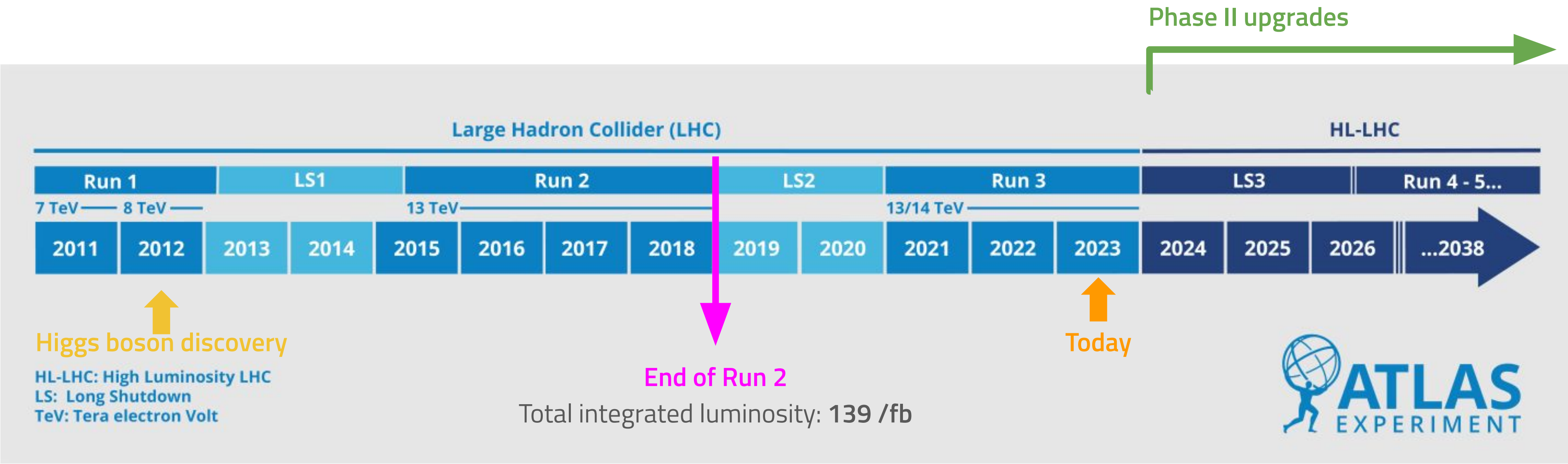
State of the art at a glance

- **2012:** Discovery of the Higgs boson in the ZZ, WW and $\gamma\gamma$ decay channels
- **2018:** First observation of Higgs boson production in association with a top quark pair (ttH) using a combination of decay channels
- **2020:** First measurement of the charge-parity properties of the interaction between the Higgs boson and the top quark in ttH production with Higgs decaying to photons (H$\rightarrow\gamma\gamma$)

This talk

- Particle physics
- Experimental analysis of proton-proton collisions
- Properties of the Standard Model particles
- Higgs boson
- Interaction between Higgs boson and top quark (ttH and tH)
- First measurements of differential cross-section and charge-parity properties of the ttH coupling in H→bb

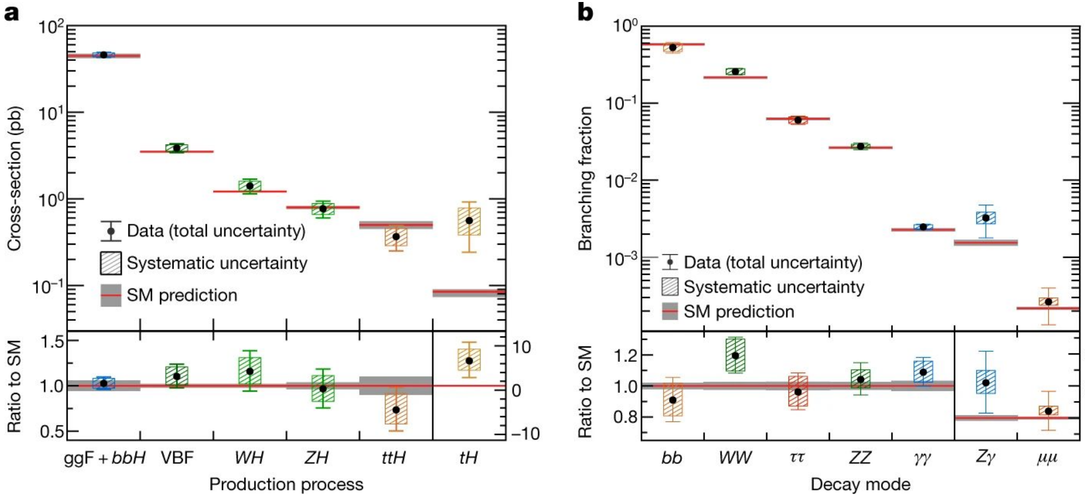
Timeline of the LHC

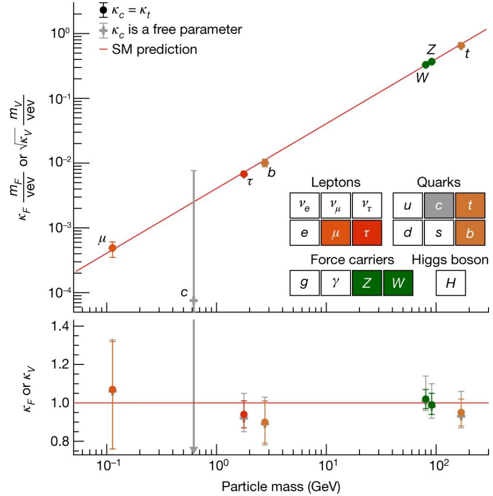


Status of Higgs boson measurements

[Nature 607, 52 (2022)]

- Cross sections and branching ratios measured using the full Run 2 dataset for (almost) all production processes and decay channels
- Simplified Template Cross-Section (STXS) framework extensively used





tH production and H→Zγ and H→$\mu\mu$ missing

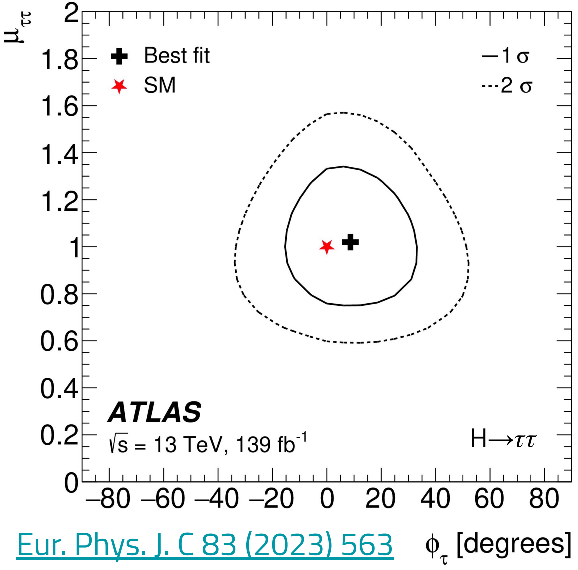
Couplings proportional to mass

Even more detailed measurements

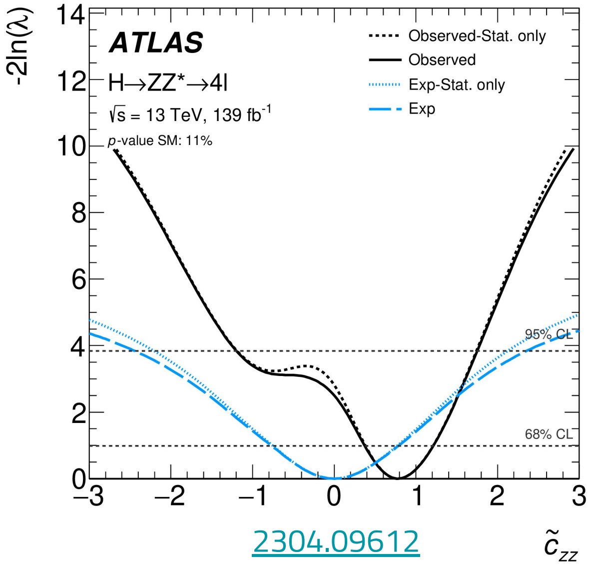
- In many channels, there is enough sensitivity to start probing the structure of the Higgs couplings

Fermions ⇒ Potential new physics effects modify the couplings at tree-level ⇒ CP-odd exclusion ~3σ

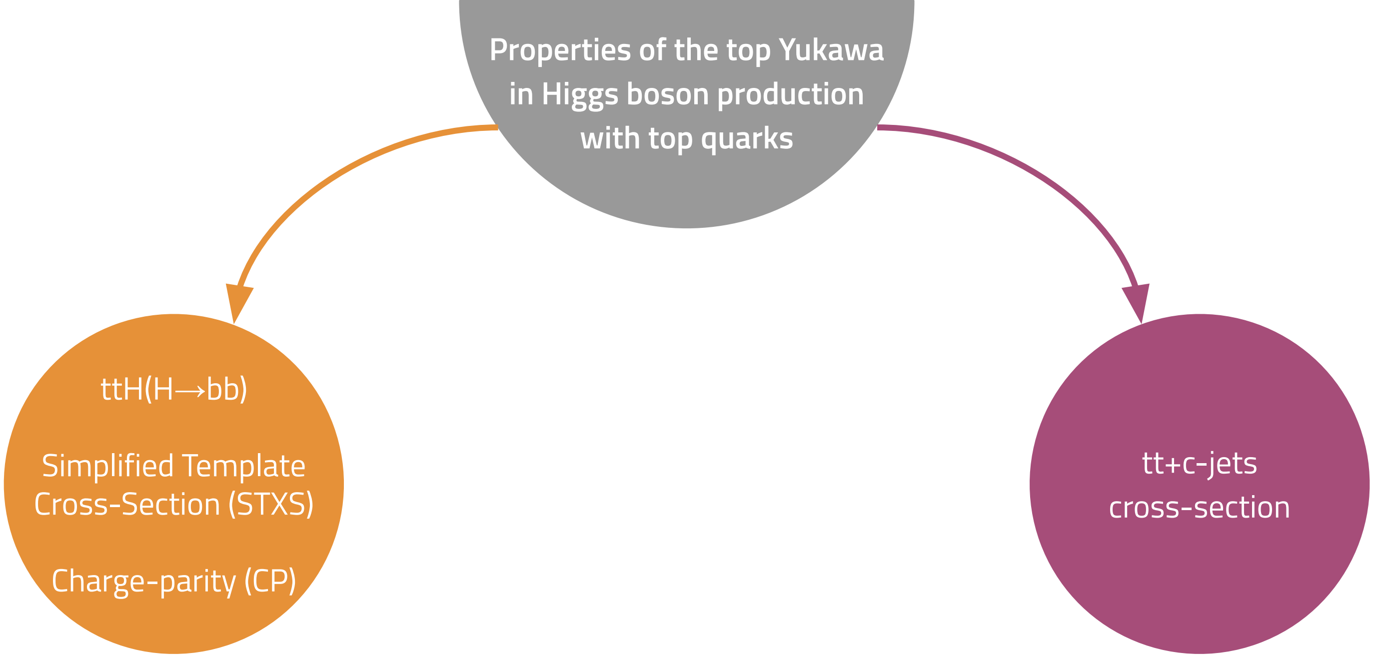
Bosons ⇒ Impact of new physics effects comes from dimension >4 operators ⇒ Set limits on EFT coefficients



Eur. Phys. J. C 83 (2023) 563



2304.09612



First measurements of Higgs boson production in association with top quarks and decaying to b quarks

First cross-section measurement of top quark pair production in association with c-tagged jets

Properties of the top Yukawa in Higgs boson production with top quarks

ttH(H→bb)

Simplified Template Cross-Section (STXS)

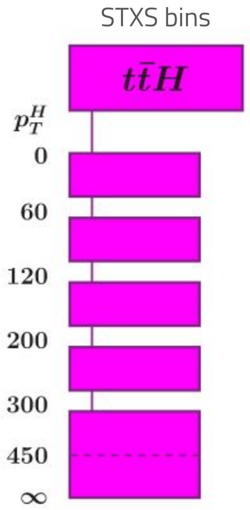
Charge-parity (CP)

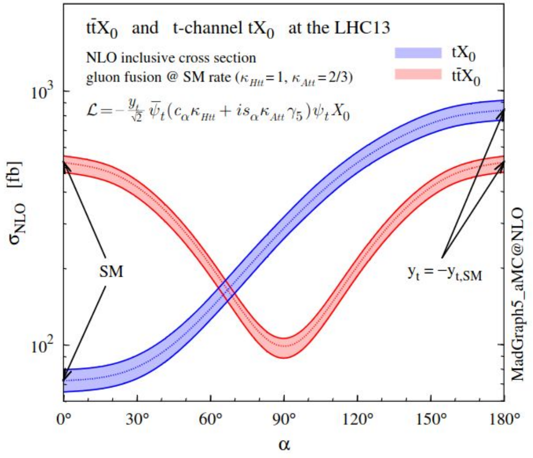
1. Motivation and overview
2. Extracting CP information
3. Event topology
4. Backgrounds
5. Event selection & strategy
6. Results
7. Systematic uncertainties

tt+c-jets cross-section

Motivation and overview

- Study the **charge-parity** properties of the top Yukawa coupling
- Necessary detour: **measure the ttH cross-section with H→bb**
  - Full Run 2 dataset ⇒ Simplified Template Cross-Section (STXS) measurement
  - Stepping stone to build CP analysis
- **Two separate analyses** with a lot in common: samples, object reconstruction, event selection, statistical analysis, treatment of systematic uncertainties
  - Key differences: signal parameterization and **treatment of tH**





- Treated as signal in CP analysis and as background in STXS analysis
- tH has very small cross-section in the SM (not observed yet) but:
  - Large increase for CP-mixed and CP-odd scenarios
  - Sensitivity to sign of the coupling

Extracting CP information

$$\kappa_t \equiv \kappa_t' \cos\alpha$$
$$\tilde{\kappa}_t \equiv \kappa_t' \sin\alpha$$

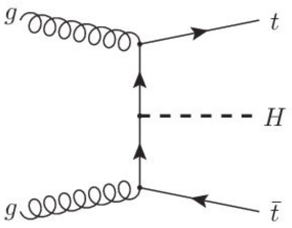
Higgs characterization model (EFT)

$$\mathcal{L} = -\frac{m_t}{v}\left\{\bar{\psi}_t \kappa_t \left[\cos(\alpha) + i\sin(\alpha)\gamma_5\right]\psi_t\right\} H$$

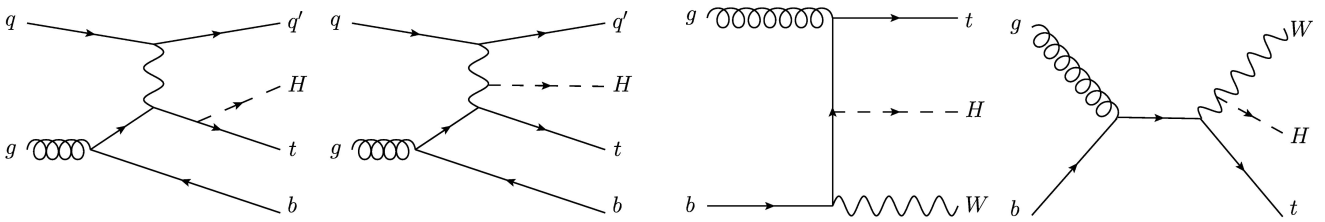
CP-even (SM)

CP-odd

Parameterize yield in each analysis bin as a function of the CP parameters



$$\frac{N(\kappa_t, \tilde{\kappa}_t)}{N_{\mathrm{SM}}} = A\kappa_t^2 + B\tilde{\kappa}_t^{\,2}$$



Pure t-H CP-even/odd

t-H CP-even/odd interference

$$\frac{N(\kappa_t, \tilde{\kappa}_t)}{N_{\mathrm{SM}}} = A\kappa_t^2 + B\tilde{\kappa}_t^{\,2} + C\kappa_t + D\tilde{\kappa}_t + E\kappa_t\tilde{\kappa}_t + F$$

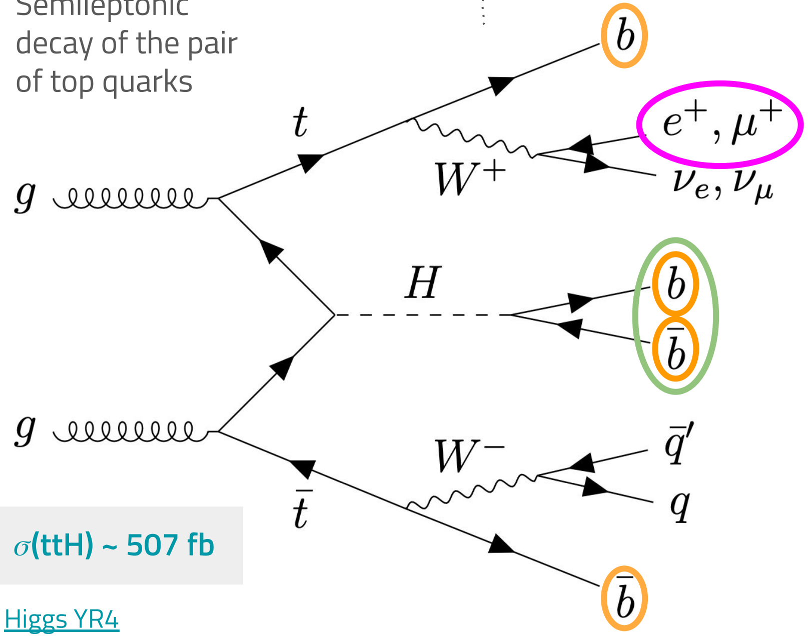
t-H and W-H interference

Pure W-H

Event topology

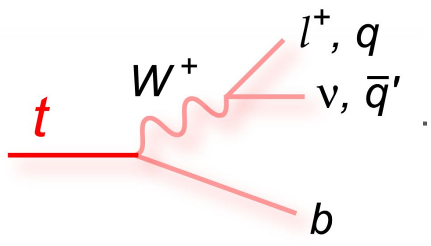
Semileptonic decay of the pair of top quarks

Simulated in the 5FS using Madgraph (CP) or Powheg (STXS) and Pythia8



σ(ttH) ~ 507 fb

Higgs YR4



Both W bosons decay leptonically
Dileptonic channel

≥4 jets, **≥4 b-jets** (2 from Higgs)
=2 charged leptons

Only one W boson decays leptonically
Semileptonic channel

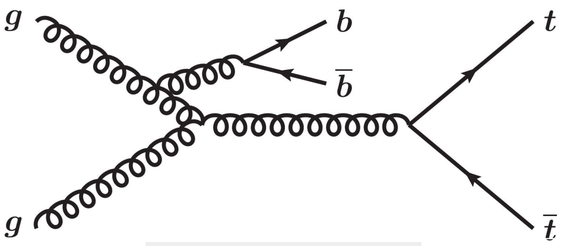
≥6 jets, **≥4 b-jets** (2 from Higgs)
=1 charged lepton

"Boosted" regime: Higgs boson p_T>300 GeV ⇒ Decay products can be reconstructed as a single jet

Background processes

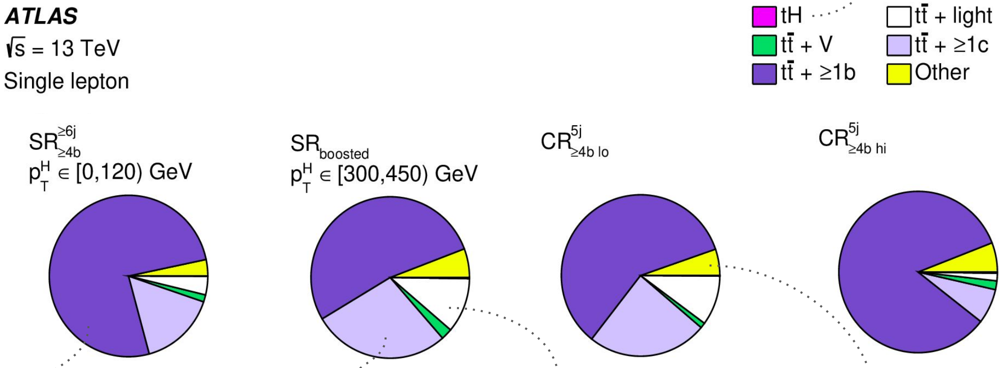
- Dominant background: ttbar production with additional b-tagged jets (tt+≥1b)
- Contribution from signal and background estimated from MC simulation
- tt+≥1b cross-section measured directly from data in the fit regions

Simulated as two independent processes using Madgraph+Pythia8: **tHjb** (4FS) and **tWH** (5FS)



σ(ttbb) ~ 2000 fb

Higgs YR4



tt+≥1b: ttbb sample 4FS Powheg + Pythia8

tt+≥1c, tt+light: ttbar inclusive sample 5FS Powheg + Pythia8

Other: WW/ZZ, single-top, W/Z+jets

Multivariate methods

- **Reconstruction BDT:** assign jets to truth-level partons
- **Reconstruction boosted neural network:** assign to each large-R jet the probability that it originates from QCD, top or Higgs boson production
    - Higgs if P(H)>0.6

Allow to reconstruct high-level variables used in the training of the classification BDTs

- **Classification BDT:** separate ttH CP-even from backgrounds
    - Used as final discriminant in signal regions in STXS analysis
    - Used to split events into regions increasingly richer in signal in CP analysis
    - Trained separately in dilepton, single-lepton resolved and boosted, in inclusive regions split based on the number of jets (≥4j, ≥6j, boosted selection)

Event selection and analysis strategy

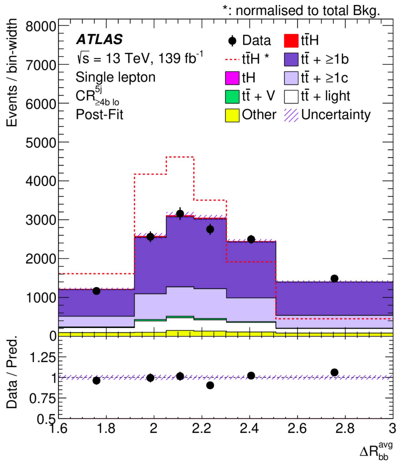
- Events are first split into channels (dilepton and single-lepton) and assigned to orthogonal signal and control regions based on **number of jets and b-jets**
- Signal regions are further subdivided in different ways for the STXS and CP measurements
  - **STXS:** bins of reconstructed Higgs candidate p_T
  - **CP:** bins of classification BDT

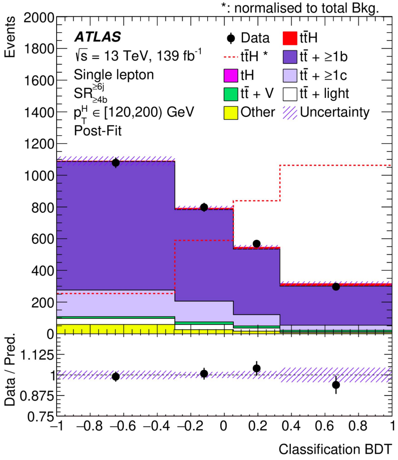
Region		Dilepton				Single-lepton			
		$SR^{\geq 4j}_{\geq 4b@70}$	$CR^{\geq 4j}_{3b@60}$	$CR^{\geq 4j}_{3b@70}$	$CR^{3j}_{3b@60}$	$SR^{\geq 6j}_{\geq 4b@70}$	$CR^{5j}_{\geq 4b@60}$	$CR^{5j}_{\geq 4b@70}$	Boosted
N_{jets}			≥ 4		$= 3$	≥ 6	$= 5$		≥ 4
$N_{b\text{-tag}}$	@85%			–			≥ 4		
	@77%			–			–		$\geq 2^{\dagger}$
	@70%	≥ 4		$= 3$			≥ 4		–
	@60%	–	$= 3$	< 3	$= 3$	–	≥ 4	< 4	–
$N_{\text{boosted cand.}}$				–			0		≥ 1

Common between STXS and CP analyses

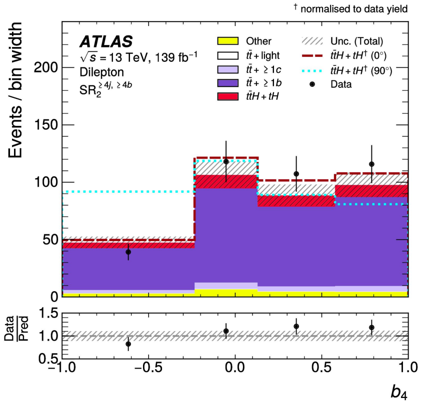
Event selection and analysis strategy

- Simultaneous maximum likelihood fit to all signal and control regions
- Control regions: fit **event yield** or average ΔR_{bb} to constrain the shape of the backgrounds
- Signal regions
  - **STXS:** fit output discriminant of classification BDT
  - **CP:** fit CP-sensitive variable





Signal vs backgrounds



ttH+tH CP-even vs CP-odd

$$b_4 = \frac{p_1^z p_2^z}{|\vec{p}_1||\vec{p}_2|}$$

Dilepton

$$b_2 = \frac{(\vec{p}_1 \times \hat{n}) \cdot (\vec{p}_2 \times \hat{n})}{|\vec{p}_1||\vec{p}_2|}$$

Single-lepton

$\hat{n}$ = z-axis

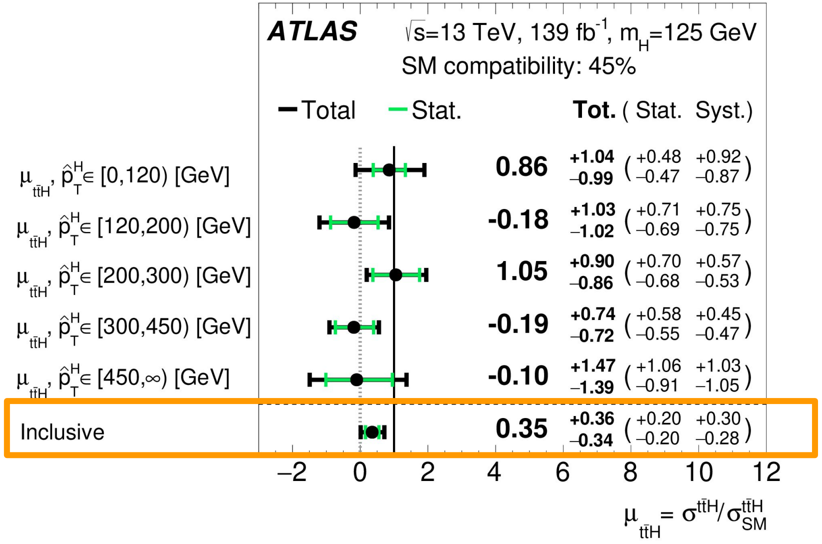
Phys. Rev. Lett. 76, 4468

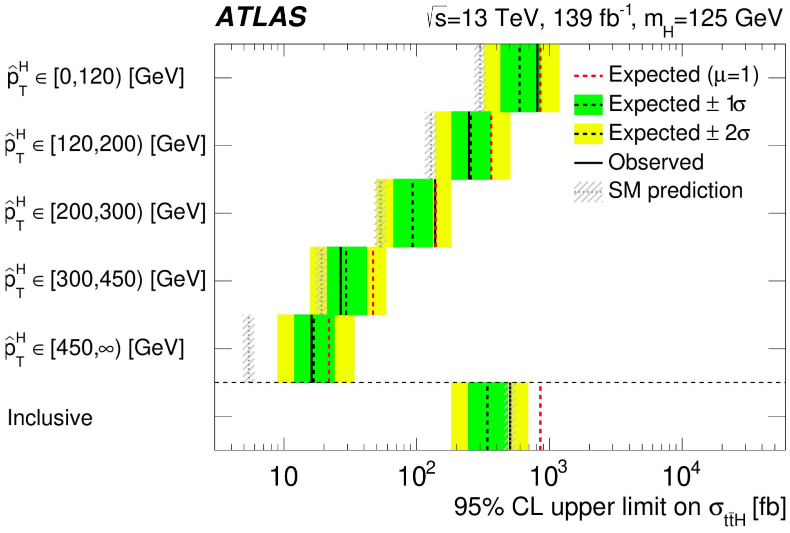
arXiv:1909.00490

Cross section measurement results

JHEP 06 (2022) 97

Statistical significance: 1.0σ (2.7σ) observed (expected)





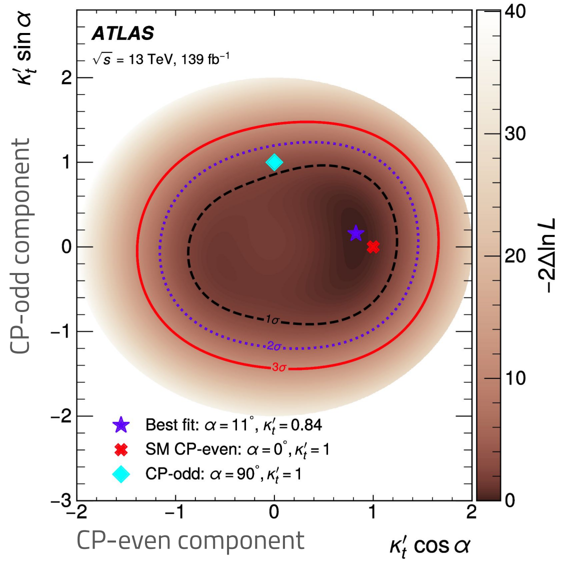
tt+≥1b normalization factor: $k(t\bar{t}+\geq 1b) = 1.28 \pm 0.08$

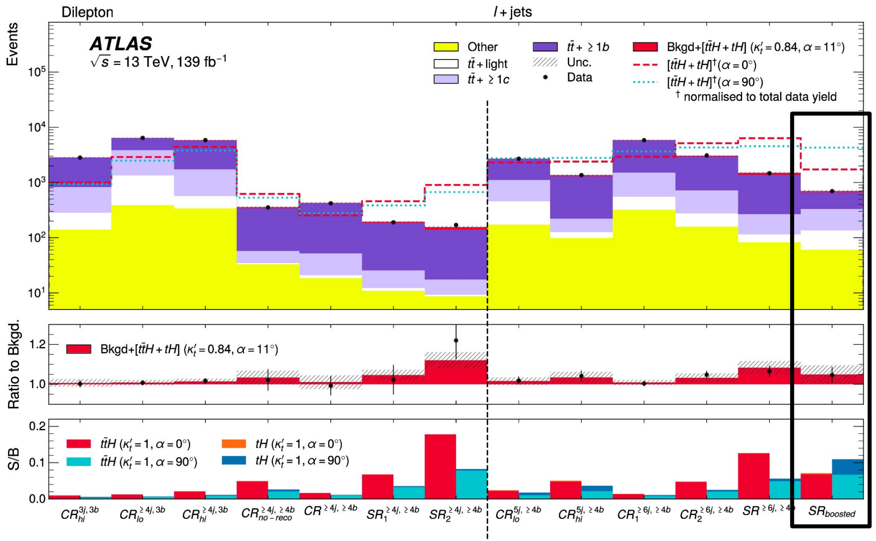
Charge-parity results

2303.05974

Best fit value of CP-mixing angle: $\alpha = 11^{\circ}{}^{+56^{\circ}}_{-77^{\circ}}$

CP-odd exclusion significance: 1.2σ observed

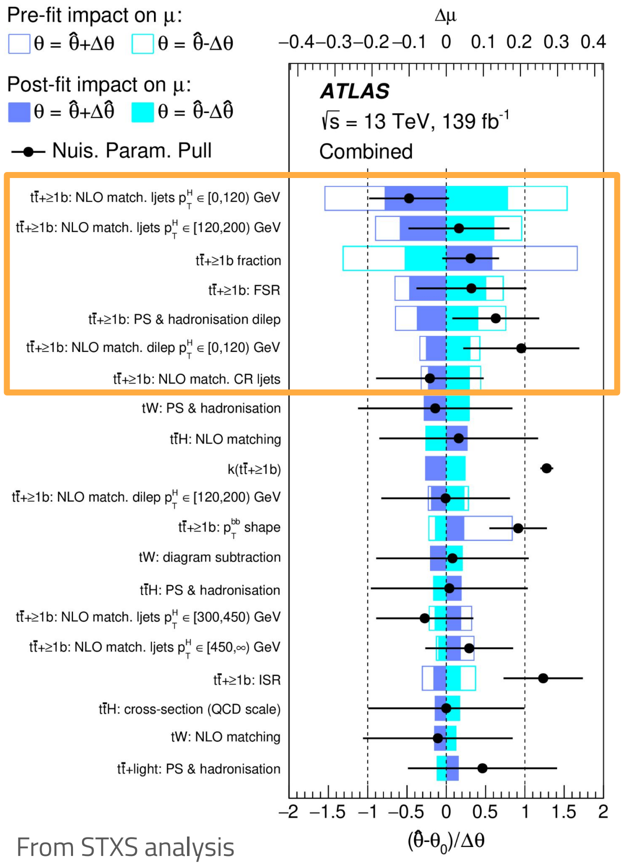




tt+≥1b normalization factor: $k(t\bar{t}+\geq 1b) = 1.30^{+0.09}_{-0.08}$

- Increased CP-odd acceptance in boosted region
- tH CP odd cross section shows large contribution

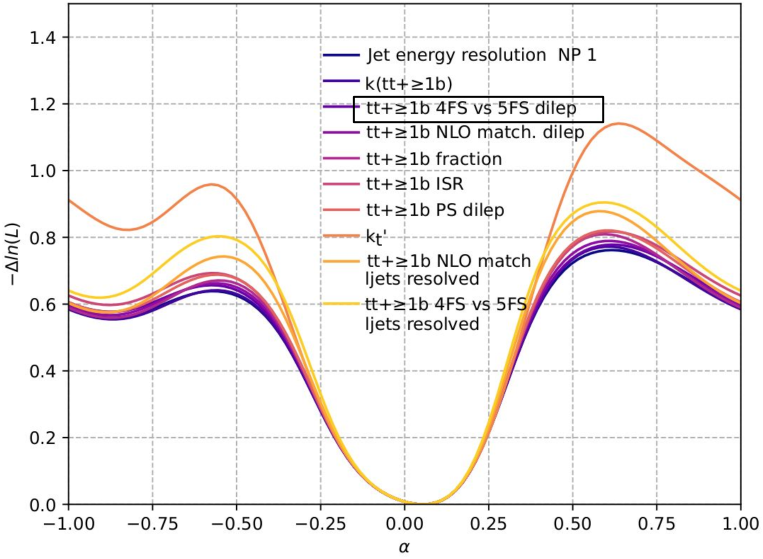
Systematic uncertainties



From STXS analysis

- Analyses dominated by uncertainties in the **modeling of tt+≥1b background**
- Experimental uncertainties are subdominant
  - Largest contribution from b-tagging efficiency and c- and light-jets mistag rates

Systematic uncertainties in the CP measurement



- Additional systematic comparing two tt+≥1b prediction with four or five quark flavors in the PDFs (4 vs 5 flavor scheme)
  - Shape not captured by any of the other modeling uncertainties
- Alternative method developed in CP analysis
  - Directly use the likelihood scans

First measurements of Higgs boson production in association with top quarks and decaying to b quarks

- First ttH simplified template cross-section measurement in H⟶bb
    - Largely impacted by the (poor) modeling of the dominant tt+≥1b background
    - Low sensitivity overall but important contribution to high transverse momentum regime, where other decay channels lack statistical power

- First measurement of charge-parity properties of the interaction between Higgs boson and top quark in ttH production with H⟶bb
    - Introduction of novel and powerful CP-sensitive angular variables
    - Confirms very important contribution of single-top Higgs production for CP studies as well as of the boosted regime

Properties of the top Yukawa in Higgs boson production with top quarks

ttH($H \rightarrow bb$)

Simplified Template Cross-Section (STXS)

Charge-parity (CP)

tt+c-jets cross-section

1. Motivation and overview
2. Identification of HF jets
3. Event selection & strategy
4. Results

Motivation and overview

- Important background in ttH analysis in hadronic final states
- Especially since the ongoing ttH(H→bb) analysis uses a looser pre-selection
- Important measurement for tuning MC generators
- Recently done by CMS, not measured before in ATLAS

Measure **tt+1c** and **tt+≥2c** signal strengths and ratios to inclusive tt+jets production in the **fiducial** and **full generator** phase spaces

- Dilepton channel allows for a better separation between tt+light and tt+≥1c than single-lepton channel, but it has reduced statistical power
- Dedicated c-tagger employed to improve sensitivity
- Combination with single-lepton channel ongoing and documented internally

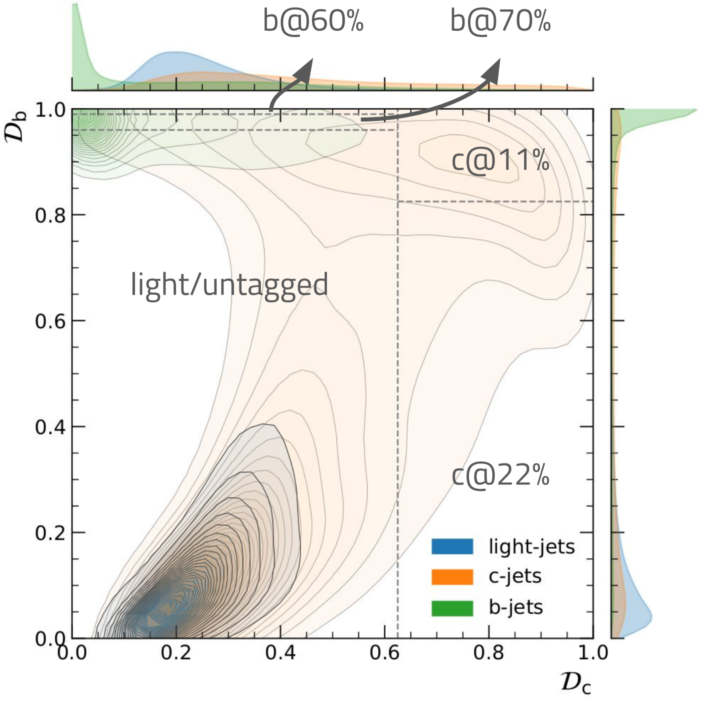
Identification of heavy flavor jets

- ATLAS standard b-tagging algorithm ⇒ Multivariate algorithm that outputs the probability of a jet being a b-, c- or light-jet
- Based on these probabilities, discriminants for b- and c-jets are constructed
- In the 2-dimensional plot defined by these discriminants, 5 bins are defined and calibrated

	b efficiency	c rejection	light rejection	$\mathcal{D}_b$	$\mathcal{D}_c$
$b@60\%$	60.3%	37.1	2320	≥ 0.990	< 0.625
$b@70\%$	70.0%	12.2	573	≥ 0.963	< 0.625
	c efficiency	b rejection	light rejection	$\mathcal{D}_b$	$\mathcal{D}_c$
$c@11\%$	11.3%	28.7	1051	≥ 0.825	≥ 0.625
$c@22\%$	22.4%	18.9	104	–	≥ 0.625

Contours from ttbar inclusive MC sample

$$\mathcal{D}_b = \log \frac{p_b}{f_c \cdot p_c + (1 - f_c) \cdot p_l}$$

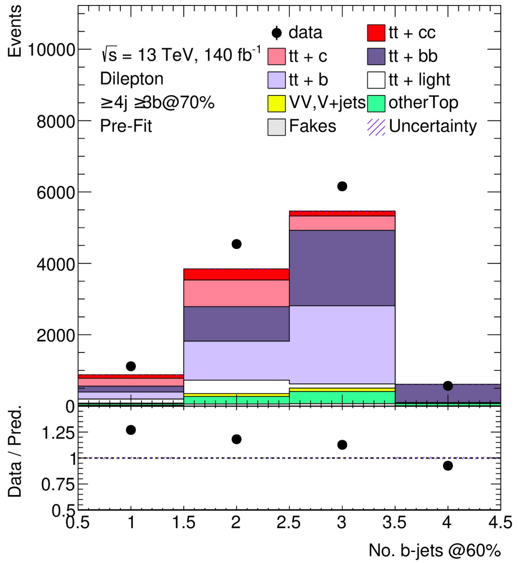


$$\mathcal{D}_c = \log \frac{p_c}{f_b \cdot p_b + (1 - f_b) \cdot p_l}$$

Event selection and analysis strategy

- Events selected with single-lepton triggers and required to have exactly two leptons
- Pre-selection: ≥3 jets, ≥2 b-jets @70% WP
- Divide phase space into orthogonal regions based on number of jets, b- and c-jets

Jets	Region	No. jets	b@70%	b@60%	c@22%	c@11%	
3 jets	$CR_1^{2\ell 3j}$	3	2	—	0	—	tt+light CR
3 jets	$CR_2^{2\ell 3j}$	3	3	3	—	—	tt+≥1b CRs
3 jets	$CR_3^{2\ell 3j}$	3	3	<3	—	—	tt+≥1b CRs
3 jets	$SR_{loose}^{2\ell 3j}$	3	2	—	1	—	SR
≥4 jets	$CR_1^{2\ell \geq 4j}$	≥ 4	2	—	0	—	tt+light CR
≥4 jets	$CR_2^{2\ell \geq 4j}$	≥ 4	≥ 3	≥ 3	—	—	tt+≥1b CRs
≥4 jets	$CR_3^{2\ell \geq 4j}$	≥ 4	≥ 3	<3	—	—	tt+≥1b CRs
≥4 jets	$SR_{loose}^{2\ell \geq 4j}$	≥ 4	2	—	1	—	SRs
≥4 jets	$SR_{tight}^{2\ell \geq 4j}$	≥ 4	2	—	≥ 2	—	SRs



No systematics included in error band

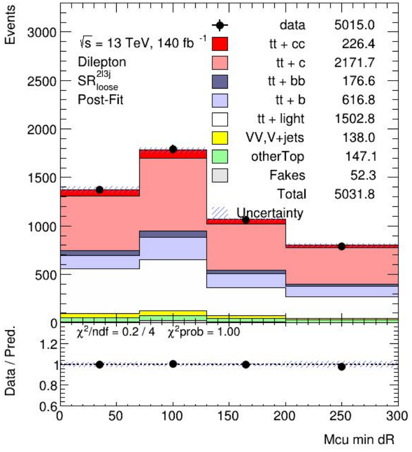
Results

- Simultaneous maximum likelihood fit in all regions
  - CRs: number of events
  - 3j SR: invariant mass between leading c-jet and closest jet
  - ≥4j SRs: number of jets

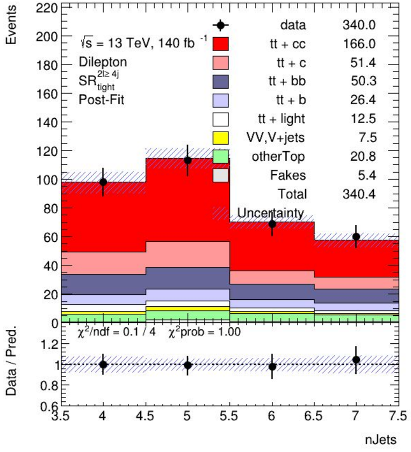
Chosen for consistency with single-lepton channel and to reduce tensions in statistical model while keeping high sensitivity

tt+1c enriched

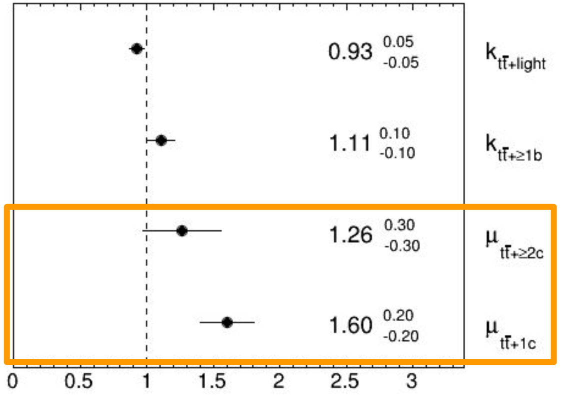
tt+≥2c enriched



Invariant mass between leading c-jet and closest jet in ΔR



Number of jets



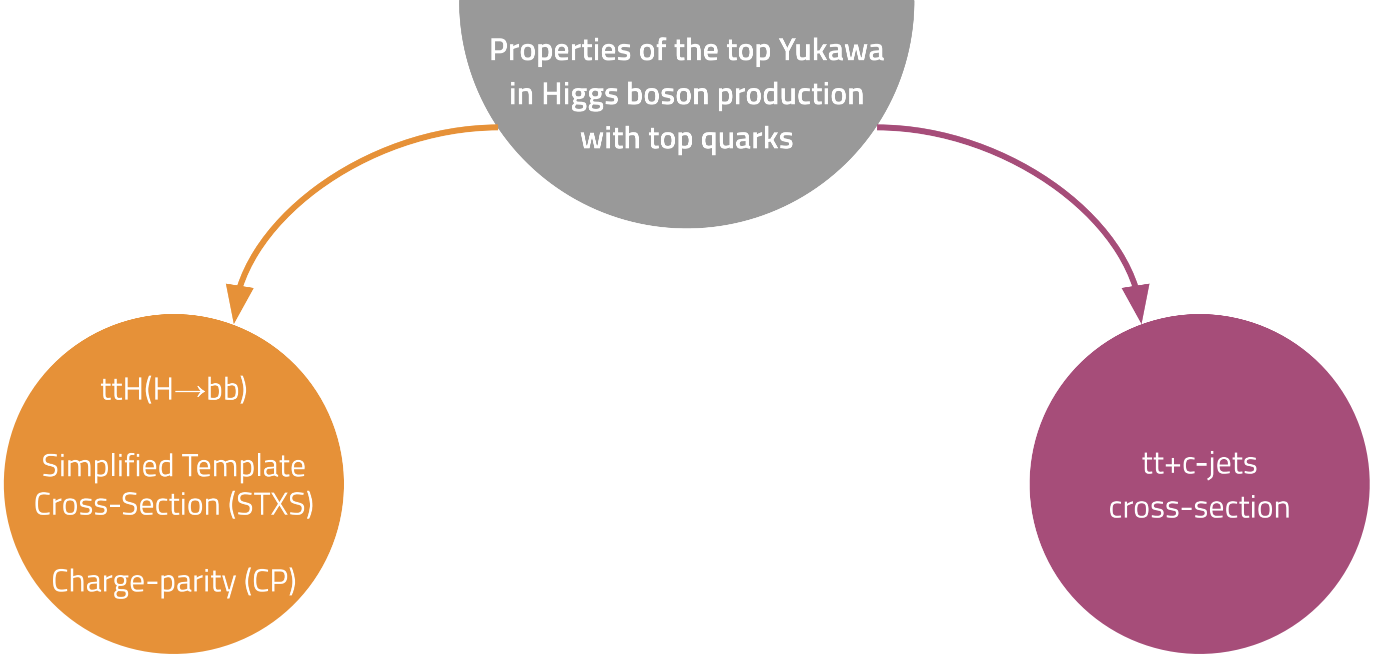
Background normalization factors

Parameters of interest (signal strengths)

Tension with respect to prediction:
~1σ for tt+≥2c and ~3σ for tt+1c

First cross-section measurement of top quark pair production in association with c-tagged jets

- Important measurement for the further development of MC event generators in ever more challenging corners of the phase space
- Top quark production in association with c-tagged jets is becoming a more important background as current ttH analysis are moving towards looser pre-selections



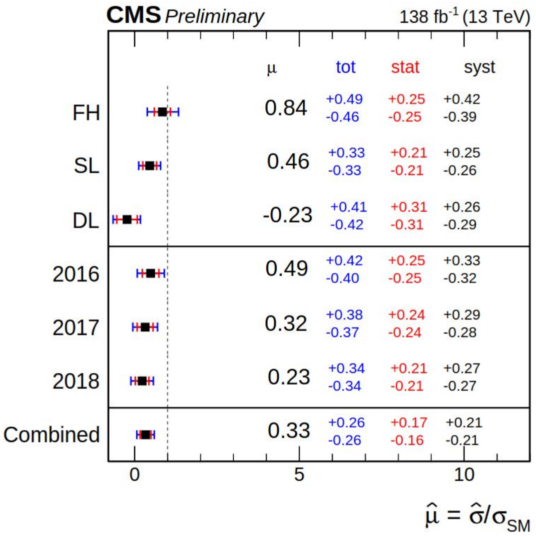
What does the (near) future hold?

Legacy ttH(H⟶bb) analysis

- Legacy ⇒ Reference results of the best ATLAS can do with the full Run 2 dataset
- Benefits from improvements in jet reconstruction and heavy flavor identification
- Fully consistent model of dominant tt+≥1b background using 4FS samples
- Looser pre-selection and more modern multiclass neural networks

What will it bring?

- Sensitivity to the lowest STXS bin (0-60 GeV) not probed in previous analysis
- Increased sensitivity inclusively and in all STXS bins
- First results with a fully consistent four-flavor scheme ttbb background model



Further insight into the CP structure

- **Higgs combination** in the EFT framework with the goal of setting limits on CP-odd operators

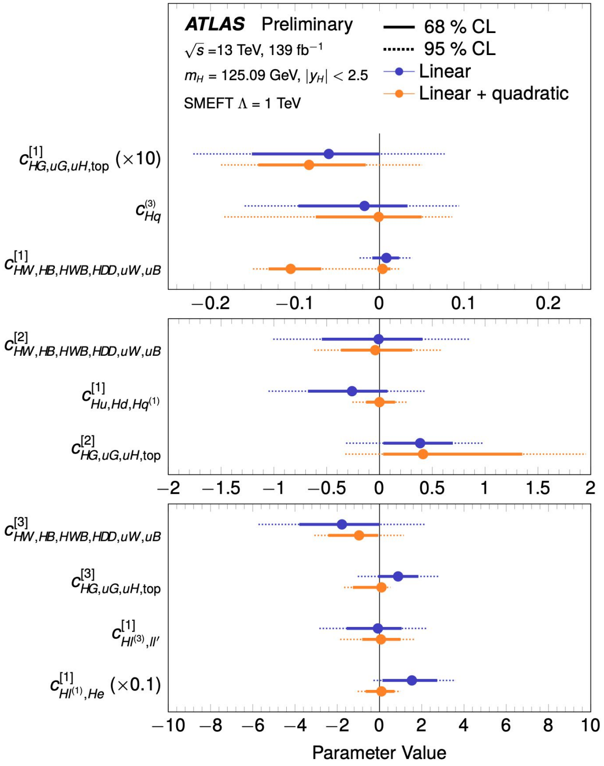
$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{c_i^{(5)}}{\Lambda_i}\mathcal{O}_i^{(5)} + \sum_i \frac{c_i^{(6)}}{\Lambda_i^2}\mathcal{O}_i^{(6)} + \sum_i \frac{c_i^{(7)}}{\Lambda_i^3}\mathcal{O}_i^{(7)} + \sum_i \frac{c_i^{(8)}}{\Lambda_i^4}\mathcal{O}_i^{(8)} + \ldots$$

Wilson coefficients
= free parameters of the theory

operators from SM fields with higher mass dimension
(Lorentz invariance, gauge invariance, locality)

From Brian Moser's talk at Higgs 2020

- Continue to **optimize CP analysis**: STXS in bins of angular variables, boosted regime, pure CP-odd observables (vanish for CP even), analysis dedicated to tH production, ...



Latest limits on CP-even operators from Higgs combination

Conclusions

- Associated production of Higgs boson with top quarks is a very challenging channel
- The $H \rightarrow bb$ decay channel suffers from a large combinatorial background due to the large number of (b-)jets in the final state
- Due to irreducible ttbb contribution, need to rely on Monte Carlo simulation to estimate the background
- Severely impacted by large modeling uncertainties
- Crucial to measure dominant ttbar production with additional heavy flavor jets (ttbb and ttcc)
- The near future will bring interesting results in $ttH(H \rightarrow bb)$ and in charge-parity studies

Thank you

Backup

Top Yukawa coupling in ttH(H→bb)

Object reconstruction

- Full Run 2 dataset corresponding to 139 /fb
- Data collected with single-lepton triggers
  - Low pT threshold and lepton isolation requirement
  - High pT threshold, looser identification and without isolation requirement

Object	Reconstruction	pT	$\|\eta\|$	Identification	Consistency with PV
Electrons/ Muons	ID tracks matched to calorimeter clusters/ tracks (full or MS-only)	>10 GeV	<2.47 (excluding transition region) /<2.5	Medium/ Loose	$\|z_0 \sin(\theta)\|<5$ mm $d_0/\sigma(d_0)<5/3$
Small-R jets	Anti-k_T R=0.4 Inputs: noise-suppressed topo clusters	>25 GeV	<2.5	---	JVT Medium WP
Large-R jets (reclustered jets)	Anti-k_T R=1.0 Inputs: constituents of R=0.4 jets	>200 GeV (M>50 GeV)	<2.5	---	---

MC samples

- Contribution from signal and background estimated from MC simulation
- Cross-section of dominant tt+≥1b background measured directly from data in the fit regions

Sample	Matrix element generator	Parton shower and hadronization	Normalization	Flavor scheme (FS)
ttH	PowhegBox (STXS) Madgraph 5 (CP)	Pythia 8	NLO QCD+EW	5
tHjb	Madgraph 5		—	4
tWH			—	5
ttbb	PowhegBoxRes + OpenLoops		—	4
Inclusive tt+jets	PowhegBox		NNLO+NNLL QCD	5

Four- and five-flavor schemes

- Four-flavor scheme: massive b-quarks in the matrix element calculation but PDFs with only 4 quark flavors (u,d,c,s)
  - Full kinematics of the b-quark correctly taken into account at LO
  - Possibly large logs are not resummed ⇒ shown to be quite small for a large class of LHC processes
- Five-flavor scheme: massless b-quarks but PDFs with five quark flavors (u,d,s,c,b)
  - For very large values of Q^2, the logs might become large and spoil the convergence of the fixed order perturbative expansion
  - Considering m_b as a small parameter, the initial state logs can be resummed into the b-quark PDFs and the final state ones into the fragmentation functions
  - Simplifies the matrix element calculation because it reduces the number of final state particles
  - Requires the introduction of phase space cuts that lead to configurations with one unresolved b-quark to no be properly taken into account

1203.6393

$$\log \frac{Q^2}{m_b^2}$$

Reconstruction BDT | Training

- Signal: all correct assignments of jets to truth-level partons, identified by truth matching
- Background: all other assignments of jets to partons
- Only b-jets can be assigned to b-partons to reduce number of possible permutations
- Trained separately in ≥6j (single-lepton) and ≥4j (dilepton) regions (with ≥4b)
- For each permutation, the values of the input variables (next slide) are calculated
    - Including kinematic properties of intermediate particles that become available

Reconstruction BDT | Input variables

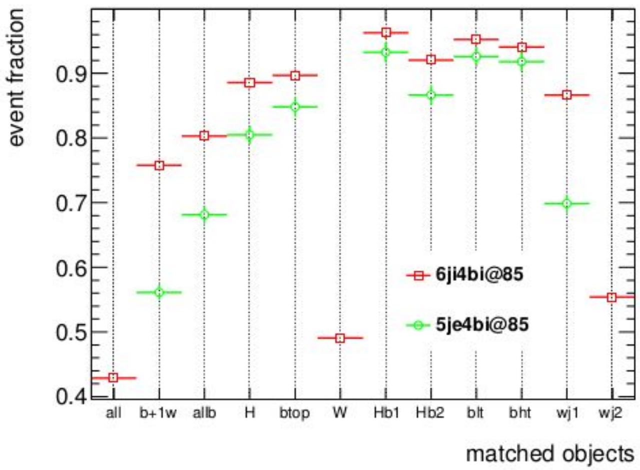
Variables	BDT w/ Higgs info.	BDT w/o Higgs info.
Topological information from $t\bar{t}$		
Mass of top	✓	✓
Mass of anti-top	✓	✓
Mass difference between top and anti-top	✓	✓
$\Delta R(\ell, b)$ from top	✓	✓
$\Delta R(\ell, b)$ from anti-top	✓	✓
$\lvert\Delta R(\ell, b)$ from top - $\Delta R(\ell, b)$ from anti-top$\rvert$	–	✓
$\Delta R(b$ from top, b from anti-top$)$	✓	–
$\Delta\phi(b$ from top, b from anti-top$)$	–	✓
p_T b from top	–	✓
p_T b from anti-top	–	✓
Min. $\Delta\eta(\ell, b$ from top or anti-top$)$	–	✓
Topological information from the Higgs-boson candidate		
Max. $\Delta R($Higgs, b from top or anti-top$)$	✓	–
Mass of Higgs	✓	–
$\Delta R($Higgs,$t\bar{t})$	✓	–
$\Delta R(b_1$ from Higgs, b_2 from Higgs$)$	✓	–

Table 17: List of input variables for the reconstruction BDTs in the dilepton channel. The top and anti-top candidates are built from one lepton and one b-jet. The lepton charge defines the top or anti-top candidates. Topological information from the Higgs-boson candidate is only used for the BDT that includes Higgs-boson candidate information.

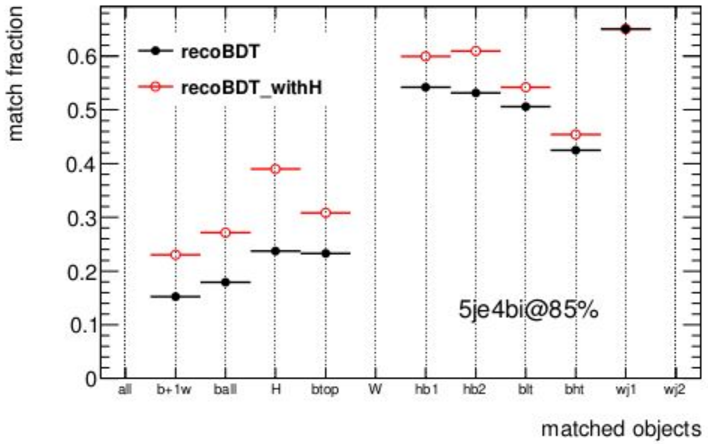
Variables	BDT w/ Higgs info.	BDT w/o Higgs info.
Topological information from $t\bar{t}$		
Mass of top_{lep}	✓	✓
Mass of top_{had}	✓	✓
Mass of W_{had}	✓	✓
Mass of W_{had} and b from top_{lep}	✓	✓
Mass of W_{lep} and b from top_{had}	✓	✓
$\Delta R(W_{\text{had}}, b$ from $\text{top}_{\text{had}})$	✓	✓
$\Delta R(W_{\text{had}}, b$ from $\text{top}_{\text{lep}})$	✓	✓
$\Delta R(\ell, b$ from $\text{top}_{\text{lep}})$	✓	✓
$\Delta R(\ell, b$ from $\text{top}_{\text{had}})$	✓	✓
$\Delta R(b$ from top_{lep}, b from $\text{top}_{\text{had}})$	✓	✓
$\Delta R(q_1$ from W_{had}, q_2 from $W_{\text{had}})$	✓	✓
$\Delta R(b$ from t_{had}, q_1 from $W_{\text{had}})$	✓	✓
$\Delta R(b$ from t_{had}, q_2 from $W_{\text{had}})$	✓	✓
Min. $\Delta R(b$ from top_{had}, q_i from $W_{\text{had}})$	✓	✓
$\Delta R($lep, b from $\text{top}_{\text{lep}}) -$ min. $\Delta R(b$ from top_{had}, q_i from $W_{\text{had}})$	✓	✓
Topological information from the Higgs-boson candidate		
Mass of Higgs	✓	–
Mass of Higgs and q_1 from W_{had}	✓	–
$\Delta R(b_1$ from Higgs, b_2 from Higgs$)$	✓	–
$\Delta R(b_1$ from Higgs, lepton$)$	✓	–

Table 18: List of input variables for the reconstruction BDTs in the single-lepton resolved channel. The subscript had (lep) indicates the hadronically (leptonically) decaying W-boson or the corresponding top-quark candidates. The symbol b_i refers to b-tagged jets from the Higgs-boson candidate decay sorted by p_{T}. The symbol q_i refers to jets from the W-boson candidate decay sorted by p_{T}. Topological information from the Higgs-boson candidate is only used for the BDT that includes Higgs-boson candidate information.

Reconstruction BDT | Performance

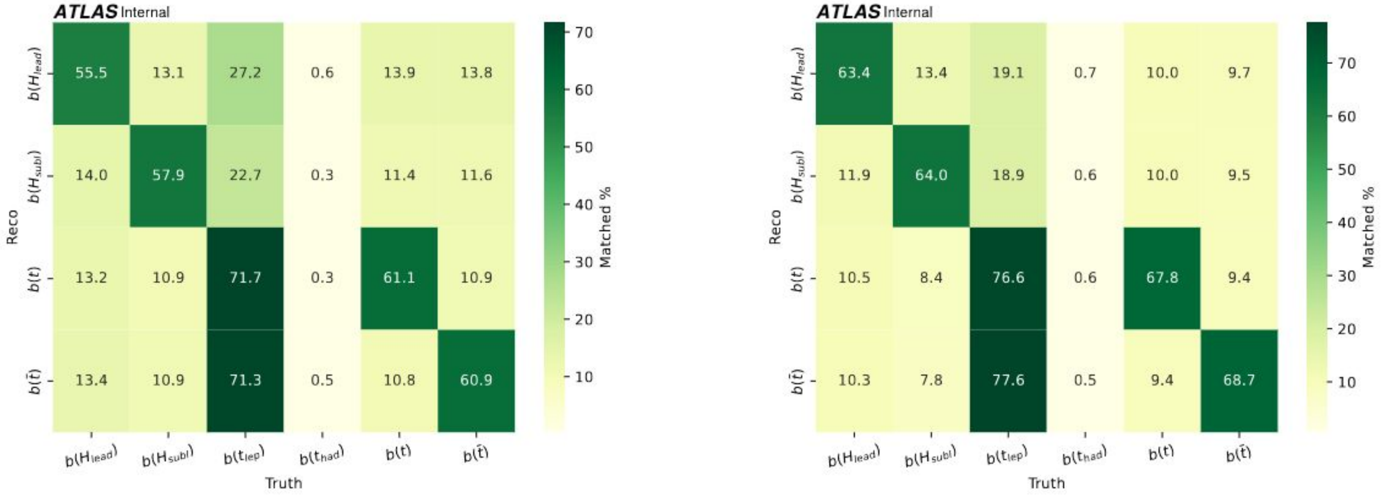


Fraction of truth matched objects



Fraction of correctly matched objects from reconstruction BDT

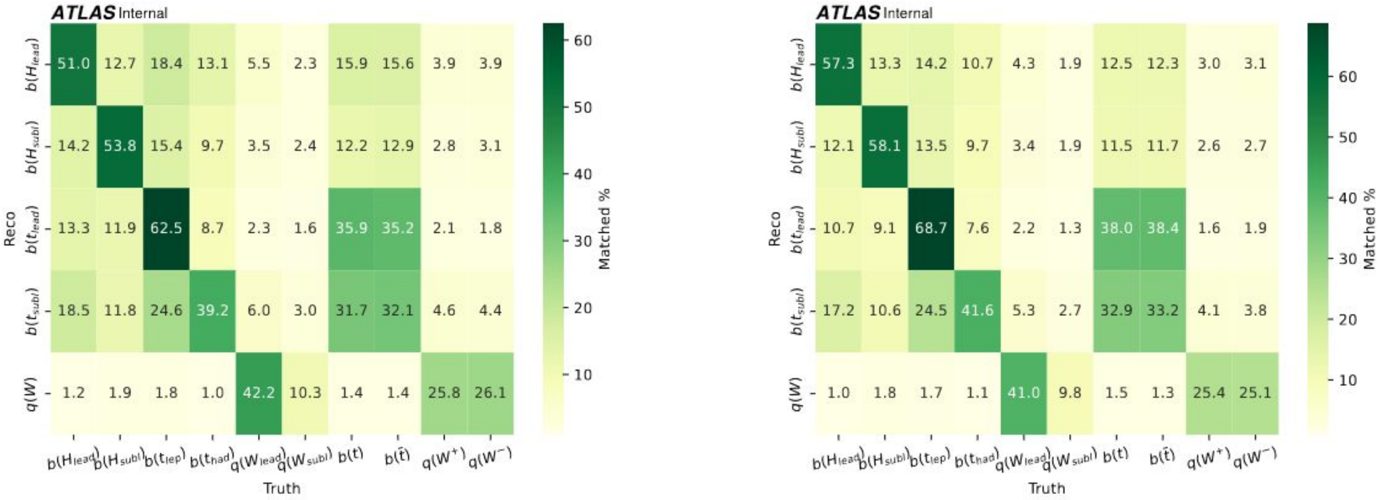
Reconstruction BDT | CP sensitivity



(a) CP even (b) CP odd

Figure 1: Confusion matrix of parton assignment with the reconstruction BDT for dilepton events of CP even (left) and CP odd (right) $t\bar{t}H$ events. The numbers indicate the fraction of reconstructed objects matched to various truth partons. Note for dilepton events the partons from the leptonically decaying top can be from either top, hence the matched % exceeds 100%.

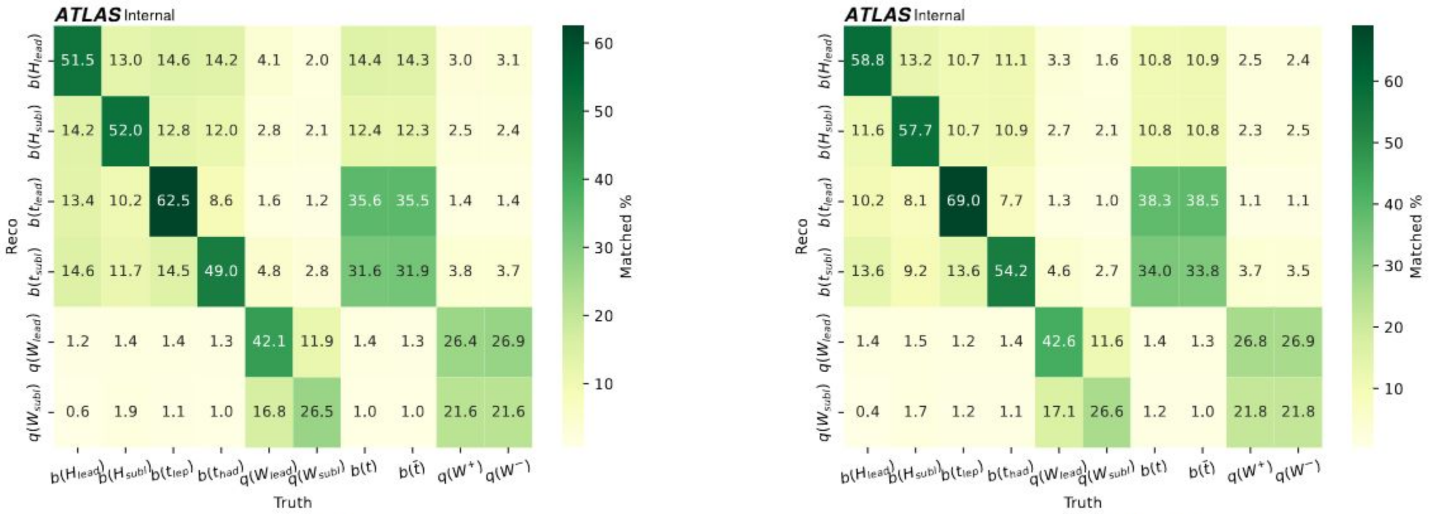
Reconstruction BDT | CP sensitivity



(a) CP even (b) CP odd

Figure 2: Confusion matrix of parton assignment with the reconstruction BDT for ℓ+ jets 5j events of CP even (left) and CP odd (right) $t\bar{t}H$ events. The numbers indicate the fraction of reconstructed objects matched to various truth partons. Here the W refers to one of the jets originating from the W (the reconstruction of W is incomplete for 5j events).

Reconstruction BDT | CP sensitivity

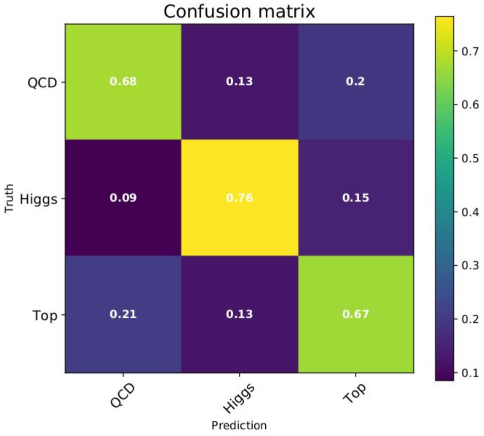


(a) CP even (b) CP odd

Figure 3: Confusion matrix of parton assignment with the reconstruction BDT for ℓ+ jets 6j events of CP even (left) and CP odd (right) $t\bar{t}H$ events. The numbers indicate the fraction of reconstructed objects matched to various truth partons.

Boosted neural network | Training and performance

- Multiclass neural network to distinguish between top, Higgs and QCD jets
- Trained on jet-by-jet basis using ttH PP8
- Higgs and top categories are defined by truth matching of jets to partons
-



Variable	Description
m^{RCjet}	mass of reclustered jet
$\sqrt{d_{12}}$	first splitting scale
$\sqrt{d_{23}}$	second splitting scale
Q_W	minimum invariant mass of constituent pairs
$n_{\text{constituents}}$	number of constituents in the RC jet
$p_{\text{T}}^{\text{const}_1}$	p_{T} of the constituent leading in p_{T}
$p_{\text{T}}^{\text{const}_2}$	p_{T} of the constituent sub-leading in p_{T}
$\text{mv2}^{\text{const}_1}$	PC MV2c10 score of the constituent leading in p_{T}
$\text{mv2}^{\text{const}_2}$	PC MV2c10 score of the constituent sub-leading in p_{T}
$\Delta R(\text{const1}, \text{const2})$	angular separation between leading and sub-leading constituents in p_{T}
$m^{b\text{-jets}}$	invariant mass of all b-tagged constituents
$m^{\text{light-jets}}$	invariant mass of all untagged constituents
mv2_{min}	minimum constituent PC MV2c10 score
mv2_{max}	maximum constituent PC MV2c10 score
$\Delta R(\text{consts})_{\text{max}}$	maximum angular separation between two constituents
$\Delta R(\text{consts})_{\text{min}}$	minimum angular separation between two constituents
mv2^{rest}	PC MV2c10 score of all constituents except the leading and sub-leading in p_{T}

Table 4: List of variables included in the DNN training. The substructure variables $\sqrt{d_{12}}$, $\sqrt{d_{23}}$ [107] and Q_W are calculated using the constituents information of the RC jets.

Classification BDT | Training

- Signal: ttH PP8
- Background: ttbb PP8 4FS and ttbar inclusive PP8 5FS
- Trained in inclusive regions with ≥6j and ≥4j, with ≥4b

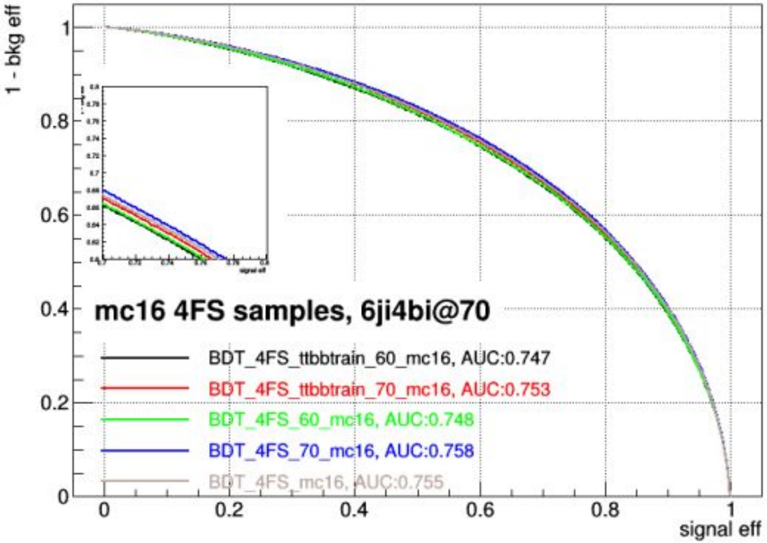
Variable	Definition	
General kinematic variables		
m_{bb}^{min}	Minimum invariant mass of a b-tagged jet pair	✓
$m_{bb}^{\text{min}\ \Delta R}$	Invariant mass of the b-tagged jet pair with minimum ΔR	✓
$m_{jj}^{\text{max}\ p_{\text{T}}}$	Invariant mass of the jet pair with maximum p_{T}	✓
$m_{bb}^{\text{max}\ p_{\text{T}}}$	Invariant mass of the b-tagged jet pair with maximum p_{T}	✓
$\Delta\eta_{bb}^{\text{avg}}$	Average $\Delta\eta$ for all b-tagged jet pairs	✓
$N_{bb}^{\text{Higgs 30}}$	Number of b-tagged jet pairs with invariant mass within 30 GeV of the Higgs-boson mass	✓
Variables from reconstruction BDT		
BDT outputs	Output of the reco. BDT w/ Higgs info. for the combination selected by the reco. BDTs w/ or w/o Higgs info.	✓**
m_{bb}^{Higgs}	Higgs candidate mass	✓
$\Delta R_{H,t\bar{t}}$	ΔR between Higgs candidate and $t\bar{t}$ candidate system	✓*
$\Delta R_{H,\ell}^{\text{min}}$	Minimum ΔR between Higgs candidate and lepton	✓
$\Delta R_{H,b}^{\text{min}}$	Minimum ΔR between Higgs candidate and b-jet from top	✓

Table 20: Variables used in the classification BDTs in the dilepton signal regions. For variables depending on b-tagged jets, only jets b-tagged using the 70% working point are considered. For variables from the reconstruction BDT, those with a * are from the BDT using Higgs-boson information, those with no * are from the BDT without Higgs-boson information while for those with a ** both versions are used.

Variable	Definition	
General kinematic variables		
$\Delta R_{bb}^{\text{avg}}$	Average ΔR for all b-tagged jet pairs	✓
$\Delta R_{bb}^{\text{max}\ p_{\text{T}}}$	ΔR between the two b-tagged jets with the largest vector sum p_{T}	✓
$\Delta\eta_{jj}^{\text{max}}$	Maximum $\Delta\eta$ between any two jets	✓
$m_{bb}^{\text{min}\ \Delta R}$	Mass of the combination of two b-tagged jets with the smallest ΔR	✓
$N_{bb}^{\text{Higgs 30}}$	Number of b-tagged jet pairs with invariant mass within 30 GeV of the Higgs-boson mass	✓
Aplanarity	$1.5\lambda_2$, where λ_2 is the second eigenvalue of the momentum tensor [129] built with all jets	✓
H_1	Second Fox–Wolfram moment computed using all jets and the lepton	✓
Variables from reconstruction BDT		
BDT output	Output of the reconstruction BDT	✓*
m_{bb}^{Higgs}	Higgs candidate mass	✓
$m_{H,b_{\text{lep top}}}$	Mass of Higgs candidate and b-jet from leptonic top candidate	✓
$\Delta R_{bb}^{\text{Higgs}}$	ΔR between b-jets from the Higgs candidate	✓
$\Delta R_{H,t\bar{t}}$	ΔR between Higgs candidate and $t\bar{t}$ candidate system	✓*
$\Delta R_{H,\text{lep top}}$	ΔR between Higgs candidate and leptonic top candidate	✓
Variables from likelihood calculations		
LHD	Likelihood discriminant	✓
Variables from b-tagging		
$w_{b\text{-tag}}^{\text{Higgs}}$	Sum of b-tagging discriminants of jets from best Higgs candidate from the reconstruction BDT	✓
B_{jet}^{3}	3rd largest jet b-tagging discriminant	✓
B_{jet}^{4}	4th largest jet b-tagging discriminant	✓
B_{jet}^{5}	5th largest jet b-tagging discriminant	✓

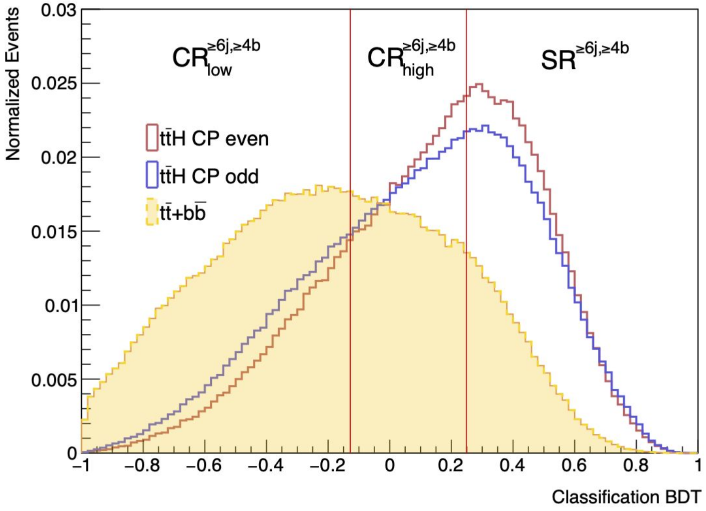
Table 21: Input variables to the classification BDTs in the single-lepton signal regions. For variables depending on b-tagged jets, jets are sorted by their pseudo-continuous b-tag score, and by their p_{T} when they have the same pseudo-continuous b-tag score. For variables from the reconstruction BDT, those with a * are from the BDT using Higgs-boson information, those with no * are from the BDT without Higgs-boson information. Variables B_{jet}^{3} and B_{jet}^{4} carry no discriminating information in the $\text{SR}_{\geq 4b\ \text{hi}}^{\geq 6j}$ region.

Classification BDT | Performance



Classification BDT | How does it look for CP-odd?

- Trained to separated ttH CP-even from ttb background
- Applied to ttH CP-odd as well ⇒ Performance is very similar to CP-even



Boosted classification BDT | Training

- Signal: ttH PP8, truth Higgs pT>300 GeV
- Backgrounds: all
-

Variable	Description
m_{Higgs}	Higgs candidate mass
$p_{\text{T}}{}^{\text{Higgs}}$	Higgs candidate transverse momentum
$\eta^{\text{lep}}_{\text{Higgs}}$	η of the Higgs candidate relative to the lepton
$\text{P(H)}_{\text{Higgs}}$	DNN Higgs probability for the Higgs candidate
m_{hadTop}	hadronic top candidate mass
$p_{\text{T}}{}^{\text{hadTop}}$	hadronic top candidate transverse momentum
$\eta^{\text{lep}}_{\text{hadTop}}$	η of the hadronic top candidate relative to the lepton
$\text{PCB}^{\text{jet}_i}_{\text{hadTop}}$	PCB score of the i^{th} jet associated to the hadronic top
m_{lepTop}	leptonic top candidate mass
$p_{\text{T}}{}^{\text{lepTop}}$	leptonic top candidate transverse momentum
$\text{PCB}^{\text{jet}}_{\text{lepTop}}$	PCB score of the jet associated to the leptonic top
n_{jets}	small-R jets multiplicity
$\Delta R(\text{Higgs, hadTop})$	ΔR between the Higgs and the hadronic top candidates
$\Delta R(\text{Higgs, lepTop})$	ΔR between the Higgs and the leptonic top candidates
$\Delta R(\text{hadTop, lepTop})$	ΔR between the hadronic top and the leptonic top candidates
$p_{\text{T}}{}^{t\bar{t}H}$	transverse momentum of the $t\bar{t}H$ system
$p_{\text{T}}{}^{t\bar{t}}$	transverse momentum of the $t\bar{t}$ system
PCB^{sum}	PCB score sum of the jets associated to the Higgs, hadronic and leptonic top
$\text{PCB}^{\text{add jet}}$	PCB score of the additional jet in the event

Table 22: Input variables to the classification BDTs in the boosted single-lepton signal region. For variables depending on b-tagged jets, jets are sorted by their pseudo-continuous b-tag (PCB) score, and by their p_{T} when they have the same b-tag score. Moreover, the i index goes from zero to two.

Statistical analysis

$$\mathcal{L}(\mu, \boldsymbol{\theta}) = \prod_{i}^{\text{bins}} \mathcal{P}\left(N_i | \lambda\right) \times \prod_{j}^{\text{NPs}} \mathcal{F}_j\left(\tilde{\theta}_j | \theta_j\right)$$

- Gaussian: systematic uncertainties
- Poisson: uncertainties associated with the finite size of the MC samples

$$\mathcal{P}(N|\lambda) = \frac{\exp^{-\lambda} \lambda^N}{N!}$$

$$\lambda = \mu \cdot S_i(\boldsymbol{\theta}) + \sum_{b}^{\text{bkgs}} k_b \cdot B_{bi}(\boldsymbol{\theta})$$

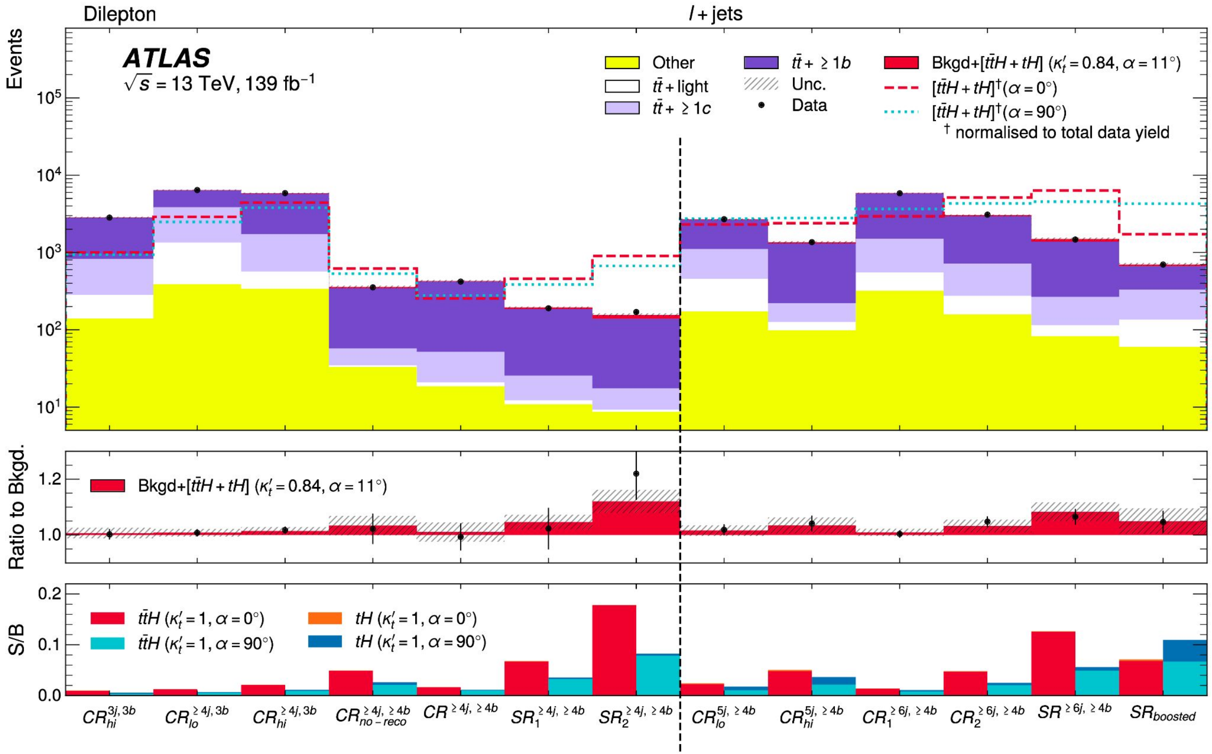
$$S(\boldsymbol{\theta}) = S_0 \times \prod_{k}^{NPs} \nu(\theta_k), \qquad B(\boldsymbol{\theta}) = B_0 \times \prod_{k}^{NPs} \nu(\theta_k)$$

$$\nu_i^{\text{shape}}(\theta) = 1 + \epsilon_i \times \theta, \qquad \nu^{\text{norm}}(\theta) = (1 + \epsilon)^{\theta}$$

Statistical analysis

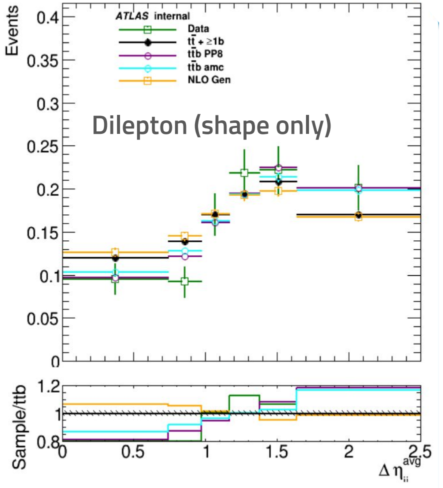
- To each systematic uncertainty a nuisance parameter (θ) is assigned, with a corresponding constraint term in the likelihood function
- For most sources of systematic uncertainties the nuisance parameters are correlated, meaning that the same parameter is applied in all channels and regions
  - This is the default procedure because it allows for the constraints on background parameters (that come mostly from the control regions) to be transferred to the signal regions
  - In some cases, systematic uncertainties can be decorrelated, meaning that independent nuisance parameters are assigned to different channel and/or regions
  - This provides more flexibility to the fit model but usually increases the uncertainty and can lead to overfitting

CP analysis | ttH and tH contribution



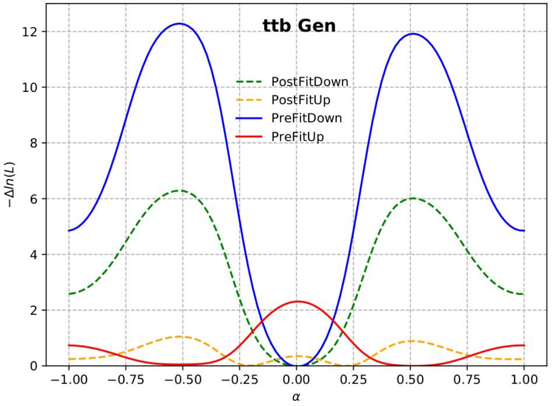
CP analysis | Systematic uncertainties

- **4FS vs 5FS systematic** uncertainty included in CP analysis because shape difference between these two samples is not captured by any of the other modeling systematics, in particular the **NLO matching**
-



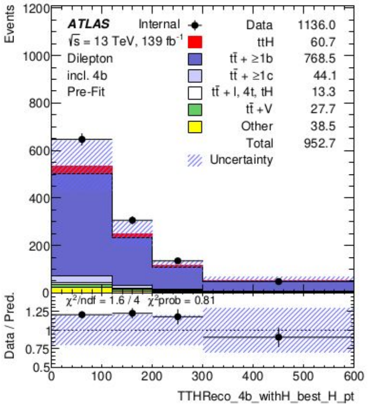
CP analysis | Alternative ranking method

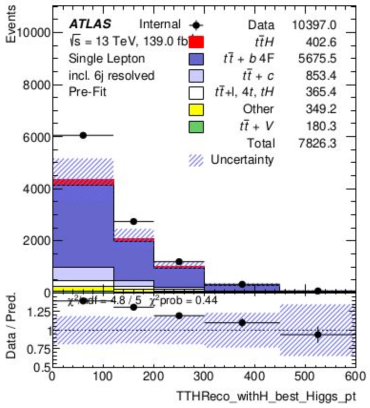
- Default ranking method in TRExFitter:
    - Fix each NP to its pre- and post-fit up and down variations and re-run fit
    - For each of the four fits, calculate the difference in mu with respect to the nominal fit
    - These differences determine the size of the blue bars in the ranking plot
- In the CP analysis, the likelihood function does not have a parabolic shape around the minimum, as it is the case for mu
- By fixing some NPs to its up/down variations a double minimum structure appears in the likelihood for one of the variations and not the other, leading to one–sided variations



Pseudodata tests injecting Higgs pT mismodeling

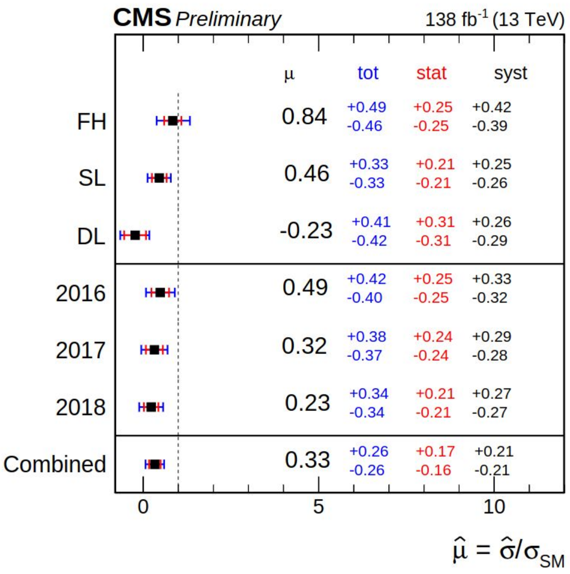
- A dedicated systematic is used in the STXS analysis to correct the known modeling of the Higgs boson transverse momentum
- Derived directly in the inclusive analysis regions to correct the mismodeling and then applied to the fit (debatable)
    - Shown to be pulled to 1σ and thus to correct the mismodeling post-fit
- Also included in the CP analysis, even though the sensitivity to the Higgs boson pT is much smaller
- By creating a pseudo dataset with the mismodeling injected and fitting with the model that includes this systematic, shown to be pulled to ~0.5σ





Latest ttH(H→bb) CMS analysis

- HIG-19-011-pas
- First CMS result using ttbb 4FS as nominal prediction
- Also sees a downward fluctuation (of similar size to ATLAS) in mu in the dileptonic and semileptonic channels (not so pronounced in the hadronic channel)
- ttb normalization factor: 1.19±0.13



Cross section of tt+c-jets

Motivation and overview

- The legacy ttH(H→bb) STXS measurement is ongoing
- Looser pre-selection ⇒ Larger contribution from ttbar production in association with c-jets (tt+≥1c)
  - Normalization and shapes of tt+≥1c known to be mismodelled
  - Very little theoretical and experimental exploration of this background component
  - This ttbar subcomponent has not been measured before by the ATLAS Collaboration

Measure **tt+1c** and **tt+≥2c** signal strengths and ratios to inclusive tt+jets production in the **fiducial** and **full generator** phase spaces

- Dilepton channel allows for a better separation between tt+light and tt+≥1c than single-lepton channel, but it has reduced statistical power
- Dedicated c-tagger employed to improve sensitivity
- Combination with single-lepton channel ongoing and documented internally

Fiducial vs full phase space fits

- The difference is the way in which the modeling systematics are normalized: to remove normalization differences between nominal and modeling variations in the fiducial or full generator phase space at truth level

Example from
[arXiv:1712.06857]

$$L(\beta^{t\bar{t}}, \beta^{W_1}, \beta^{W_{2,3}}, \delta_{b\text{-tag}}, \delta_{\mathrm{JES}}) = \prod_{k=1}^{M} \frac{e^{-\mu_k} \cdot \mu_k^{n_k}}{n_k!} \cdot G(\delta_{b\text{-tag}}; 0, 1)$$

with

$$\mu_k = \beta^s \cdot \nu_s \cdot \alpha_k^s + \sum_{j=1}^{2} \beta^{W_j} \cdot \nu_{W_j} \cdot \alpha_k^{W_j} + \sum_{b=1}^{4} \nu_b \cdot \alpha_k^b,$$

$$\beta^s = \beta^{t\bar{t}} \left\{ 1 + \sum_{i=1}^{2} |\delta_i| \cdot (H(\delta_i) \cdot \epsilon_{i+} + H(-\delta_i) \cdot \epsilon_{i-}) \right\},$$

$$\alpha_k^s = \alpha_k^{t\bar{t}} \sum_{i=1}^{2} |\delta_i| \cdot \{(\alpha_{ki}^{+} - \alpha_k) \cdot H(\delta_i) + (\alpha_{ki}^{-} - \alpha_k) \cdot H(-\delta_i)\}.$$