Probing the CP nature of the top-Higgs Yukawa coupling in $t\bar{t}H$ and tH events at ATLAS

Zak Lawrence

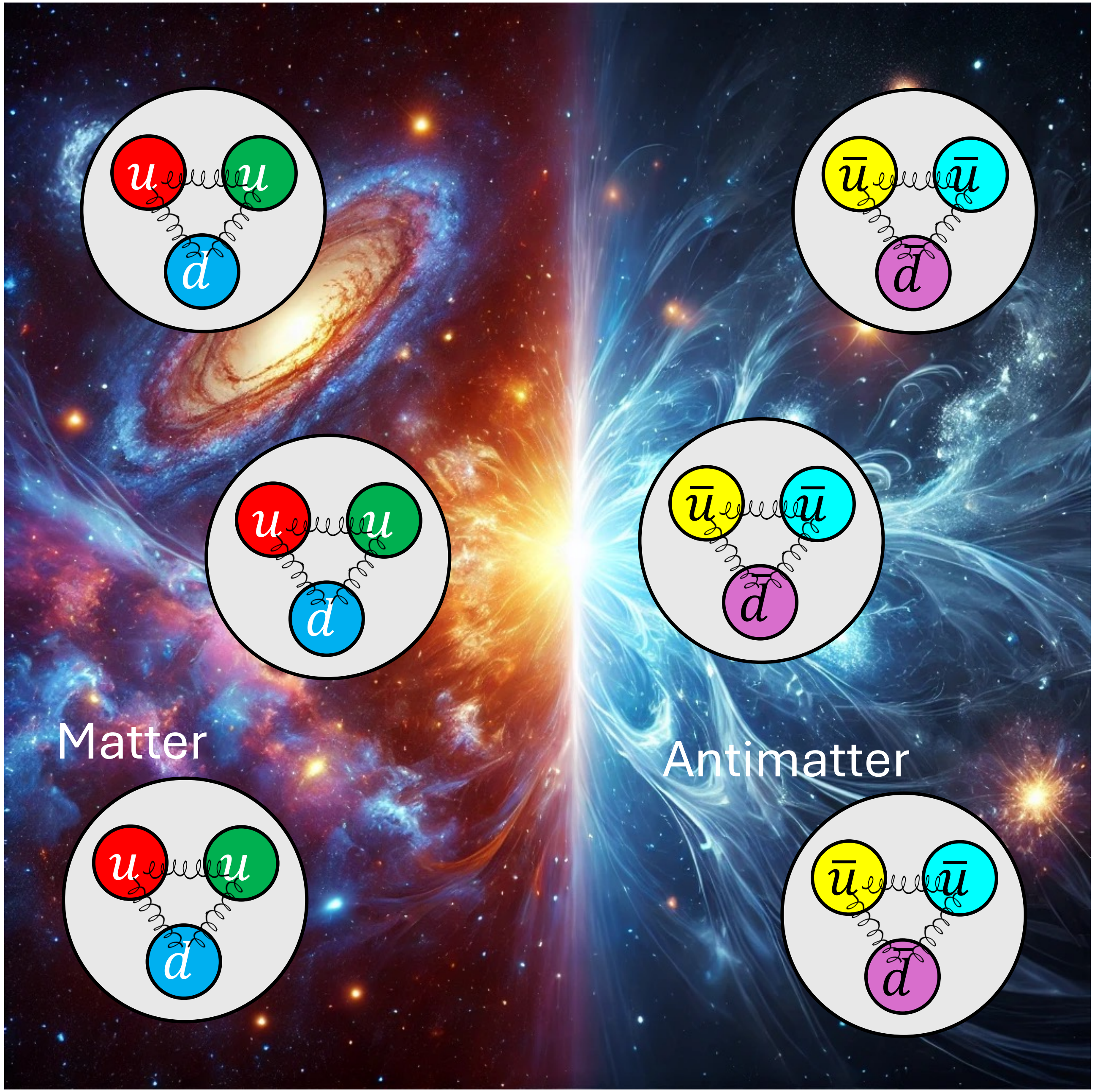
11th February 2025

What we don’t see

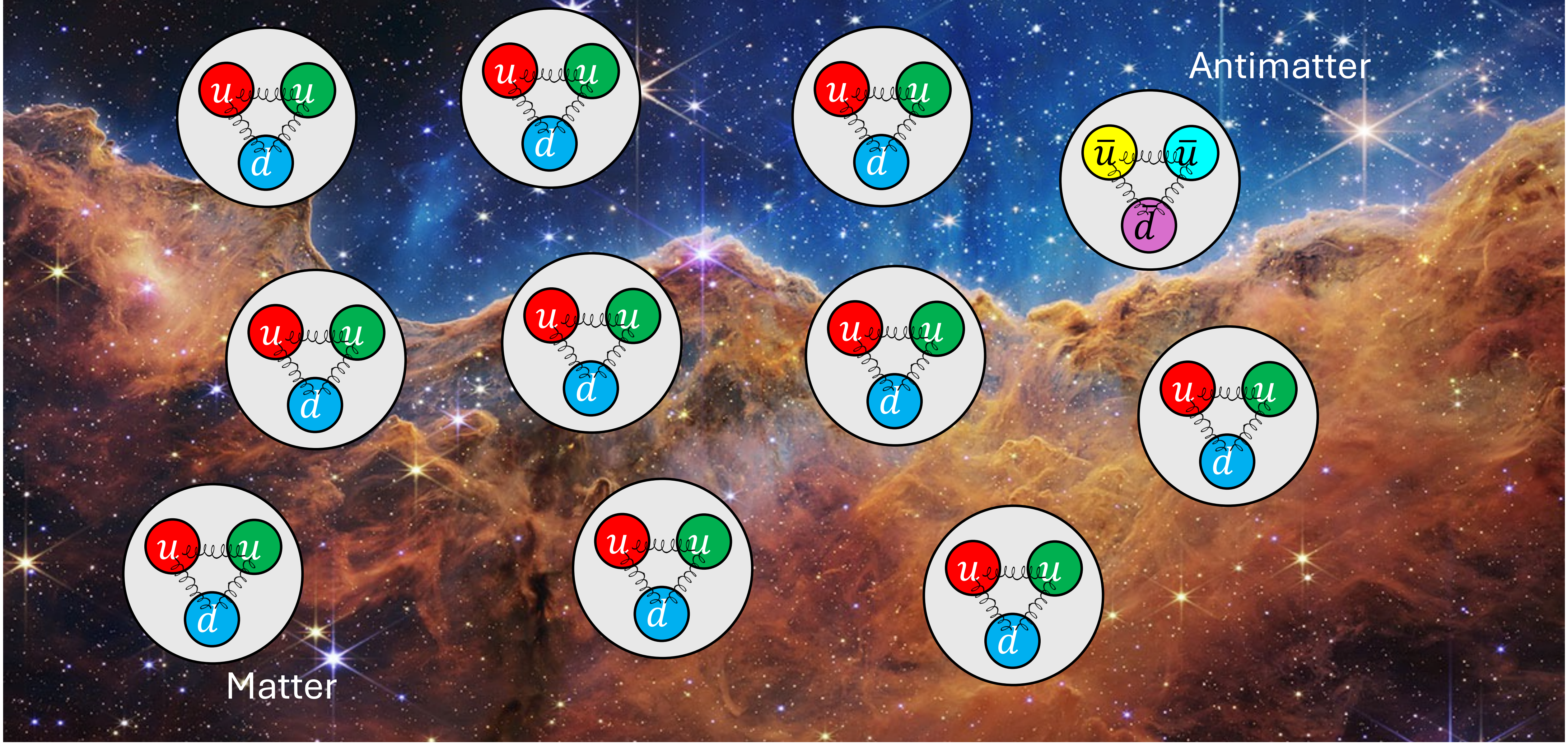
→We expect that when the universe formed there should have been an equal amount of both matter and antimatter created.

→This would have resulted in the annihilation of both and in turn an empty universe devoid of both matter and antimatter.

→Clearly this isn’t what we see..



What we See



Why search for CP violation?

The Sakharov conditions:

→To explain this difference between expectations and observations three necessary conditions must be satisfied:

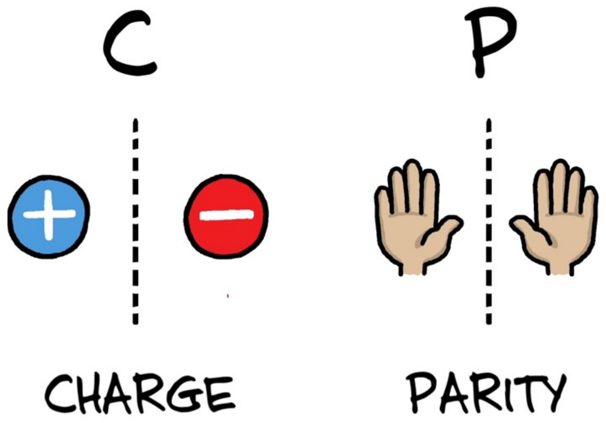
- →The violation of Baryon number.
- →*C*-symmetry and *CP*-symmetry violation.
- →Interactions out of thermal equilibrium.

C and P symmetries

→Charge and Parity are important symmetries of the SM.

→We have observed *C,P* and *CP* violation in the weak interaction

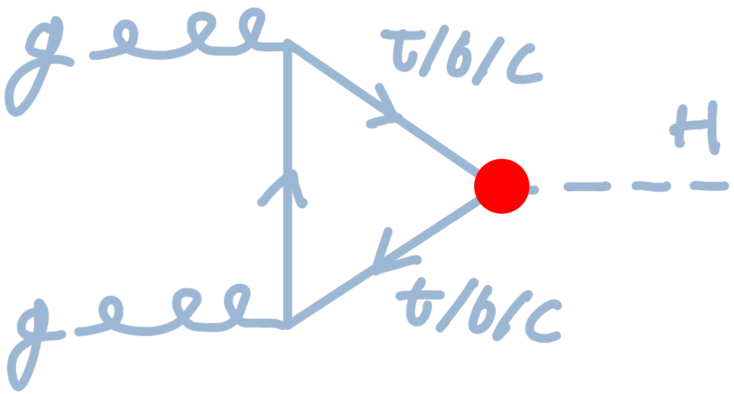
→**There is not enough *CP* violation to match the observed matter dominance.**



Yukawa Interactions

Yukawa couplings

$$\mathcal{L}_{\text{Yukawa}} = -\sum_{ij}^{3}[y_{ij}^{\ell}\bar{L}_L^j\Phi\bar{e}_R^i] - \sum_{ij}^{3}[y_{ij}^{u}\bar{\mathcal{Q}}_L^j\tilde{\Phi}u_R^i + y_{ij}^{d}\bar{\mathcal{Q}}_L^j\Phi d_R^i] + \text{H.c.},$$



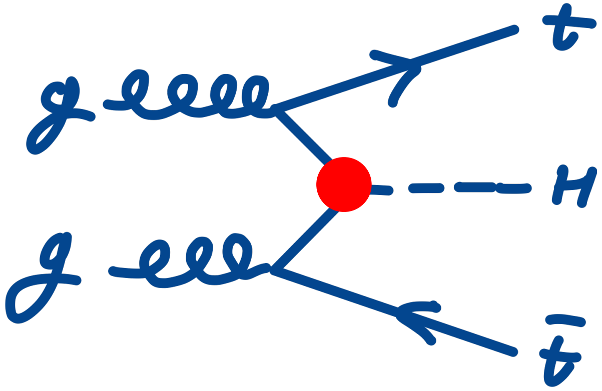
→The Yukawa couplings provide mass to the fermions in the SM → Measurement of the couplings is important to verify the predictions of the SM.

→The strength, y_{ij}^{k}, is proportional to the mass of the fermion.

→ Top quarks are the heaviest particles in the SM and so have the largest coupling, $y_t \approx 1$.

→ $t\bar{t}H$ provides a direct probe of the top-Higgs coupling at tree level.

→Allows us to search for CP violation in Yukawa interactions.



CP Violation in top Yukawa Couplings

***CP* parametrisation**

$$\mathcal{L}'_{t\bar{t}H} = -\kappa'_t y_t \phi \bar{\psi}_t (\cos\alpha + i\gamma_5 \sin\alpha)\psi_t$$

CP-even *CP*-odd

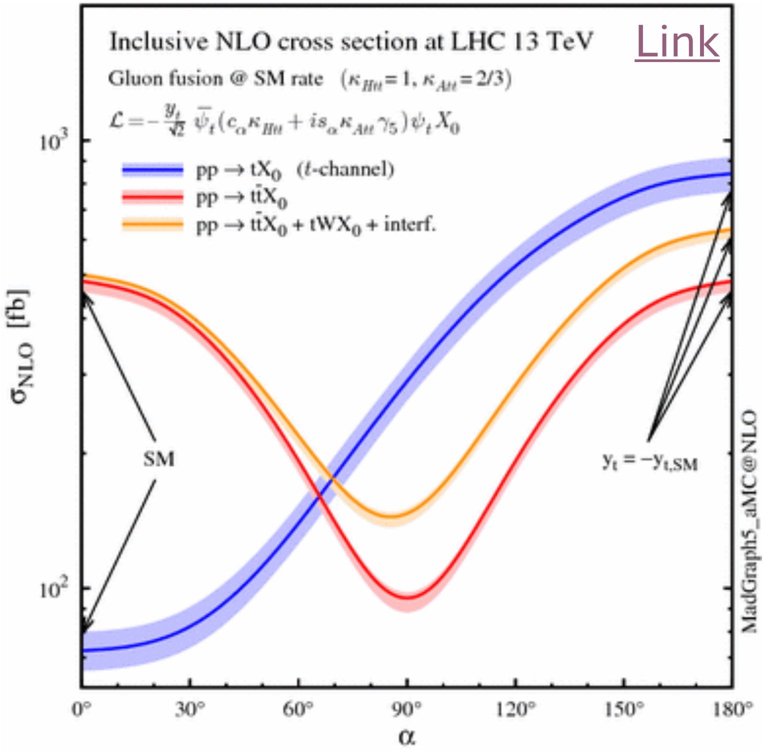
→Within the SM the Higgs is even under *CP* inversion.

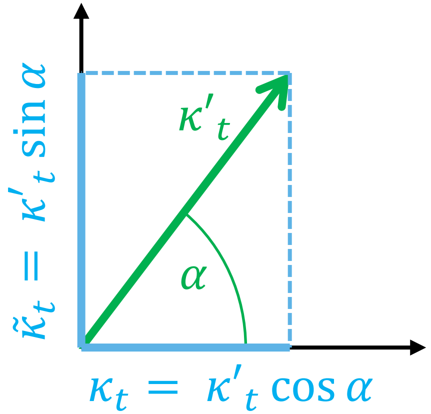
→*CP* violating terms can be added to the interaction and parametrised in terms of a coupling strength modifier κ'_t and a mixing angle α.

→Results in modifications to the observed cross section of the process.

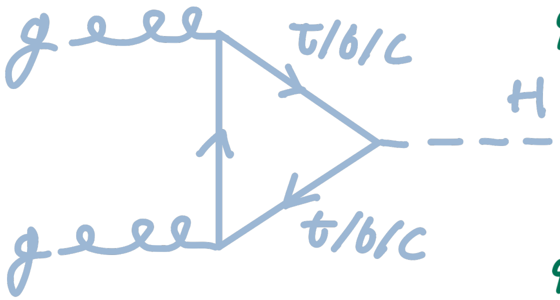
→From $t\bar{t}H$ alone we are unable to distinguish between y_t and $-y_t$, but sensitivity can be gained through contributions from tH.

→Previous results from $t\bar{t}H(H \rightarrow \gamma\gamma)$ have excluded a pure CP-odd coupling at 99.98% CL.

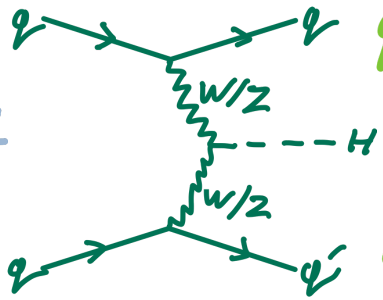




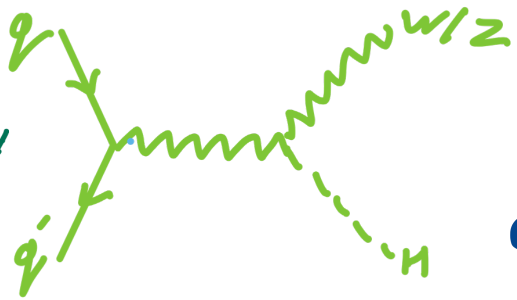
Higgs Production at the LHC



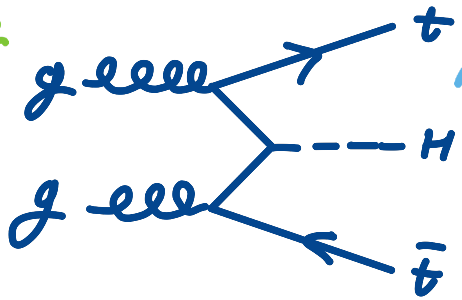
Gluon Fusion
ggF (~87%)



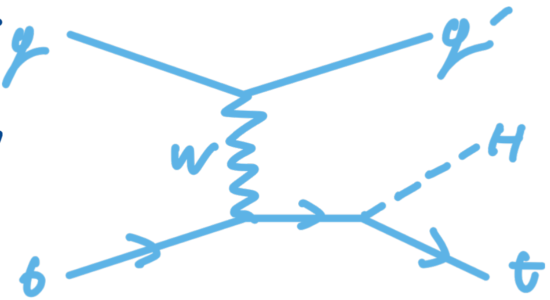
Vector boson
fusion VBF (~7%)



Associated
Production (~4%)



ttH/bbH (~2%)



Single Top
tH(~0.05%)

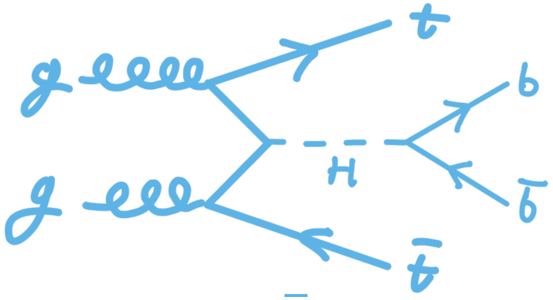
Decreasing Cross Section

→The Higgs is predominantly produced via the gluon fusion channel at the LHC.

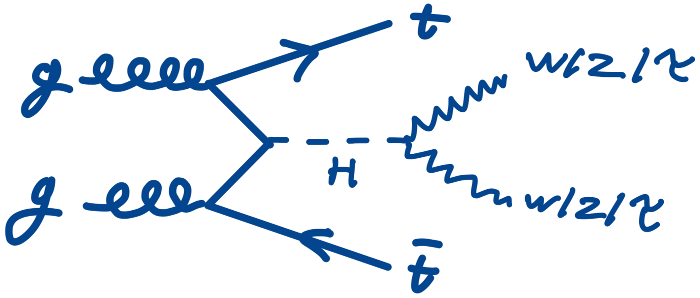
→$t\bar{t}H$ events make up a small fraction of the total Higgs production.

$$\frac{\sigma(pp \to t\bar{t}H)}{\sigma(pp \to H)} \approx 1\%$$

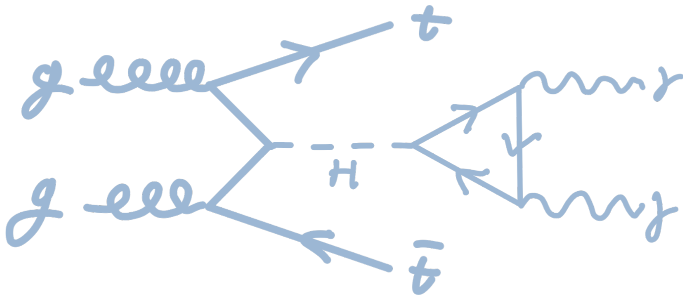
The $t\bar{t}H$ process



$H \to b\bar{b}$



Multilepton



$H \to \gamma\gamma$

Decreasing Branching Ratio

→The $t\bar{t}H$ process was first observed in 2018 by ATLAS and CMS.

→The $H \to b\bar{b}$ channel offers the largest branching ratio at 58% although comes with the challenge of a large $t\bar{t} + jets$ background.

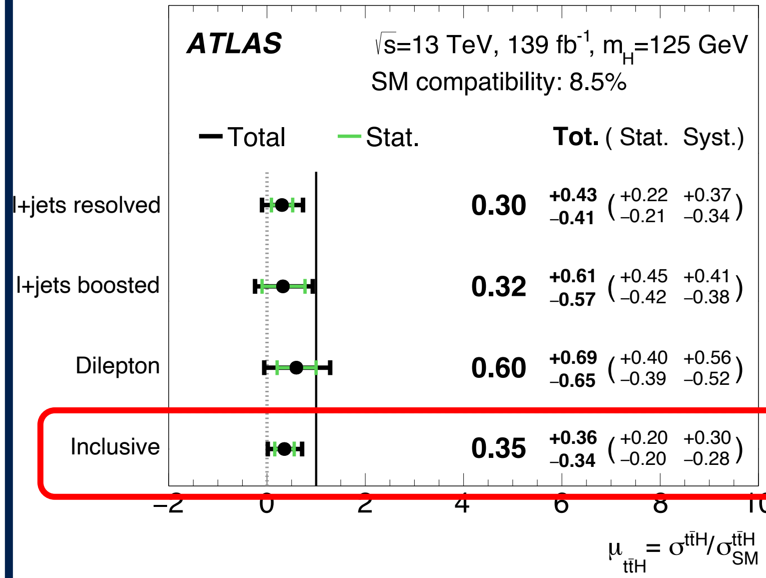
→Multilepton offers lower backgrounds and a clear signature while having a smaller branching ratio and requiring careful handling of fake lepton background.

→The $H \to \gamma\gamma$ channel can give a precise measurement of the top-Higgs coupling however contributions from non-SM particles to the loops could compensate for deviations in y_t from its SM value.

Experimental status: $t\bar{t}H(bb)$ Run 2 results

ATLAS (2022)

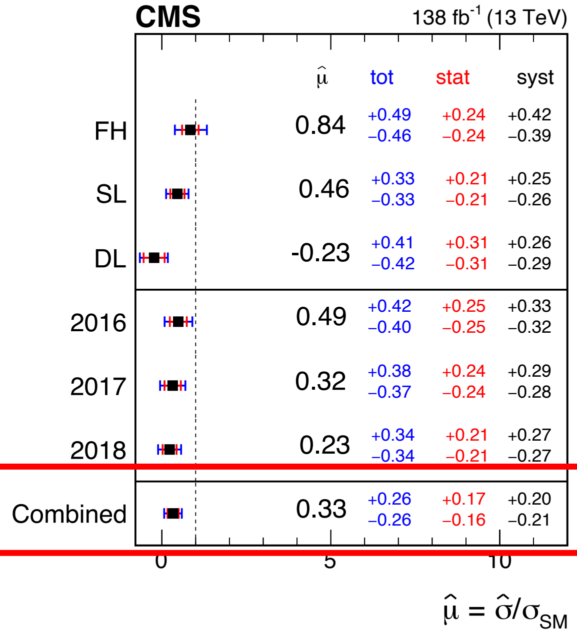
JHEP 06 (2022) 097



- → Results give inclusive $t\bar{t}H$ signal strength and a STXS measurement.
- → Offshoot analysis measures *CP* properties Phys. Lett. B 849 (2024) 138469 **(this talk)**

CMS (2024)

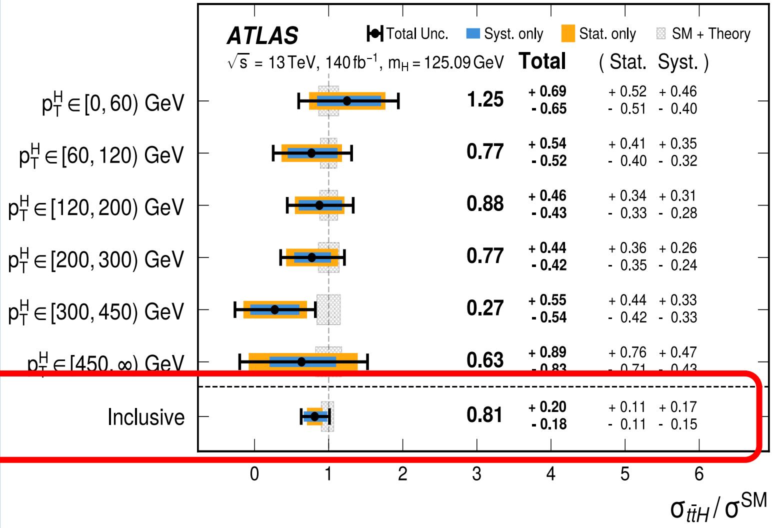
arXiv:2407.10896(accepted by JHEP)



- → Inclusive $t\bar{t}H$ signal strength and STXS.
- → Limit on tH signal strength.
- → *CP* measurement.

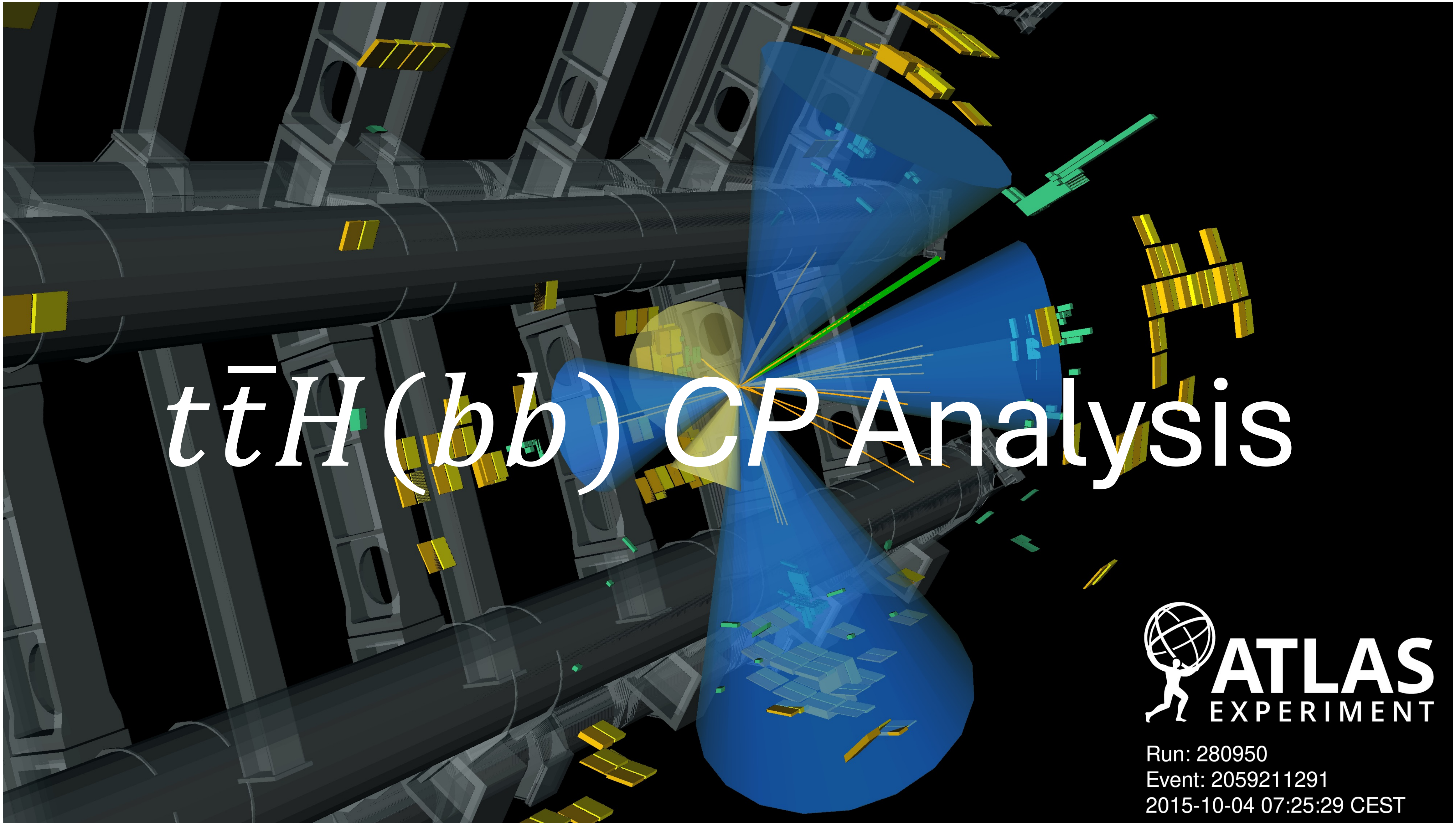
ATLAS Legacy (2024)

arXiv:2407.10904(submitted to EPJC)



- → Inclusive $t\bar{t}H$ signal strength and STXS.
- → Result includes updated object definitions and improvements to analysis strategy.
  - → For a full overview of the differences see this talk

$t\bar{t}H(bb)$ CP Analysis



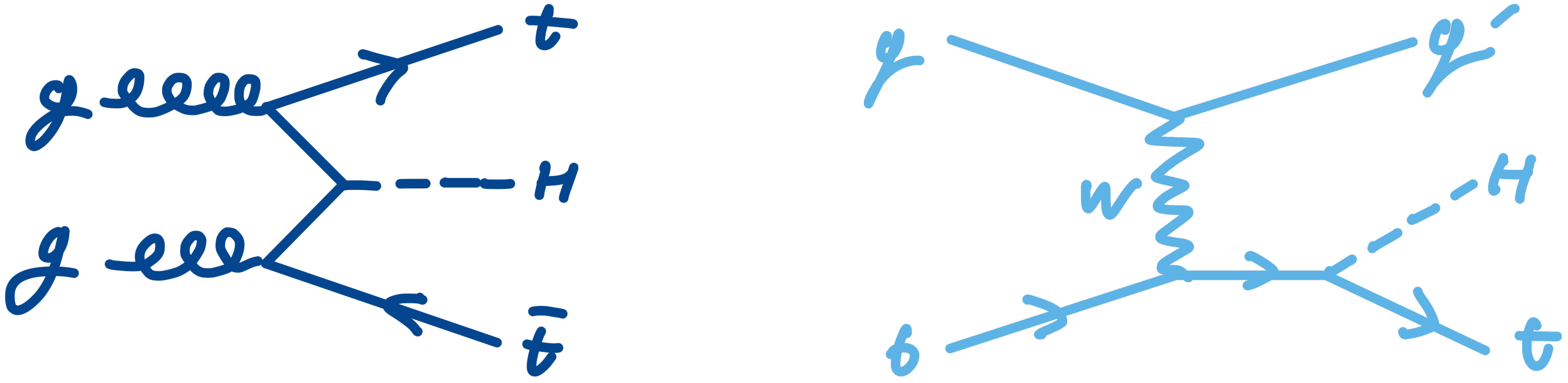
Analysis Overview

→The analysis targets events $t\bar{t}H$ events with the Higgs decaying to a pair of b-quarks.

→Makes use of data collected by ATLAS during Run 2 of the LHC.

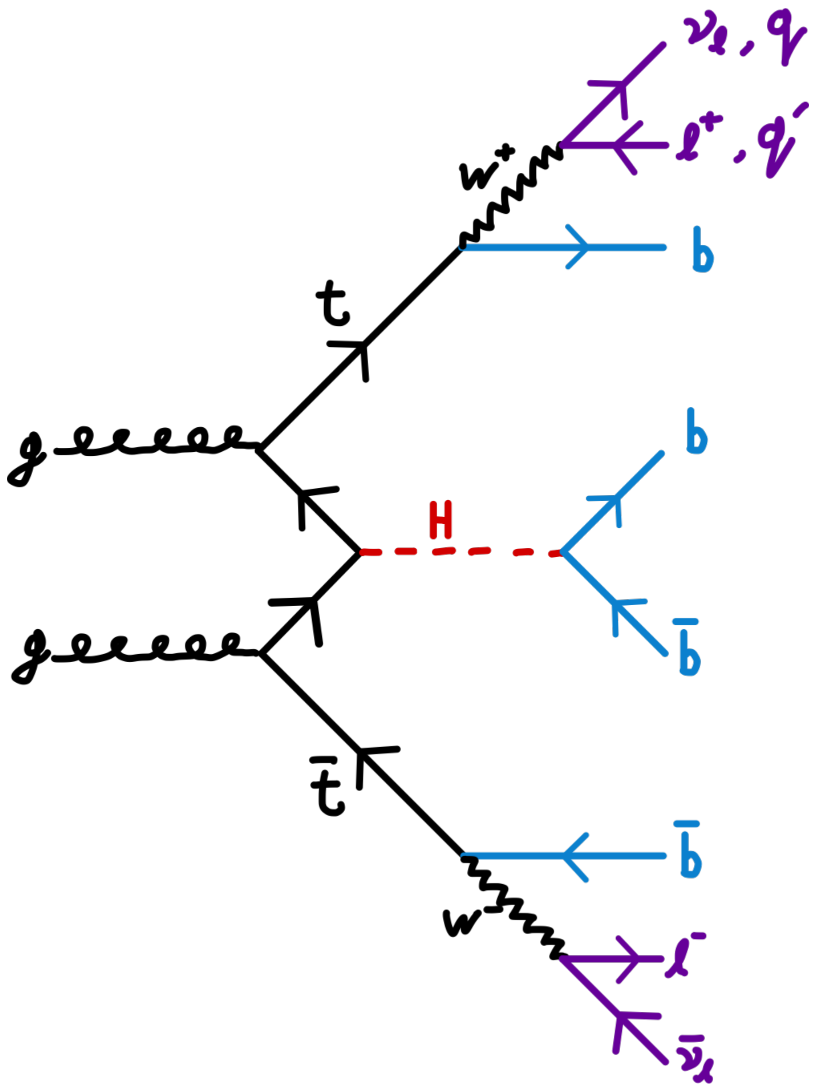
→Represents the first measurement of the CP properties of top-Higgs coupling in the $H \to b\bar{b}$ decay channel.

→Analysis strategy is based on Run 2 STXS analysis, with modifications made to probe CP properties of top-Higgs coupling.



Analysis targets the $t\bar{t}H$ and tH processes as signal.

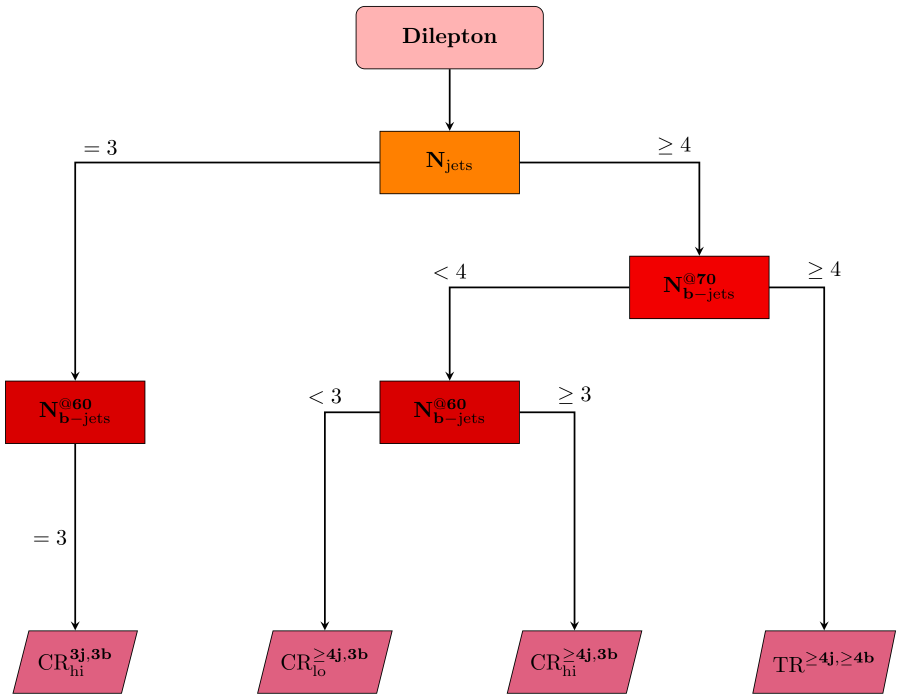
Analysis Overview

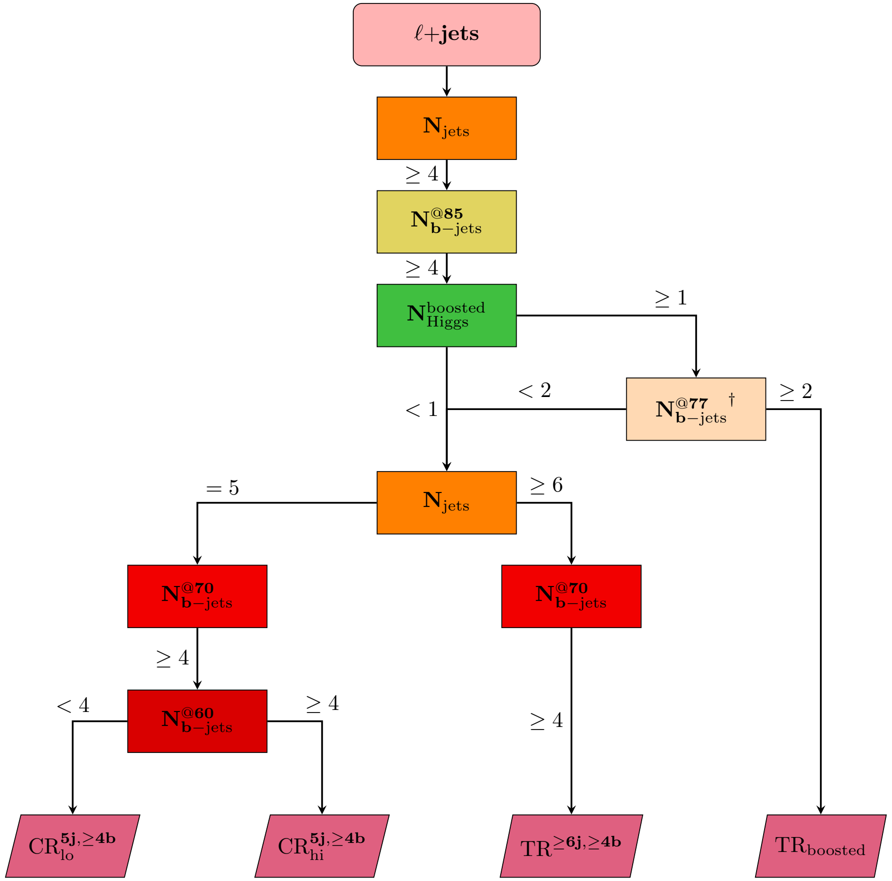


- The analysis splits events into three different analysis channels based on the decay of the $t\bar{t}$ system.
  - Only electrons and muons are considered, events with τ candidates are vetoed.
- Dilepton channel:
  - Both top quarks decay leptonically.
  - 4 *b*-tagged jets.
  - 2 charged leptons.
- Single lepton (l+jets) channel:
  - One top decays leptonically, the other hadronically.
  - Expect 6 jets, including 4 *b*-tagged jets.
  - 1 charged lepton
- Boosted Channel:
  - Is a subcategory of the l+jets channel where the two b-quarks are sufficiently boosted to be reconstructed as a single large R jet.

Event Classification

- Events selected for dilepton, and l+jets channels are further split with selections based on the number of jets and b-tagged jets.
- Selections define control regions and training regions

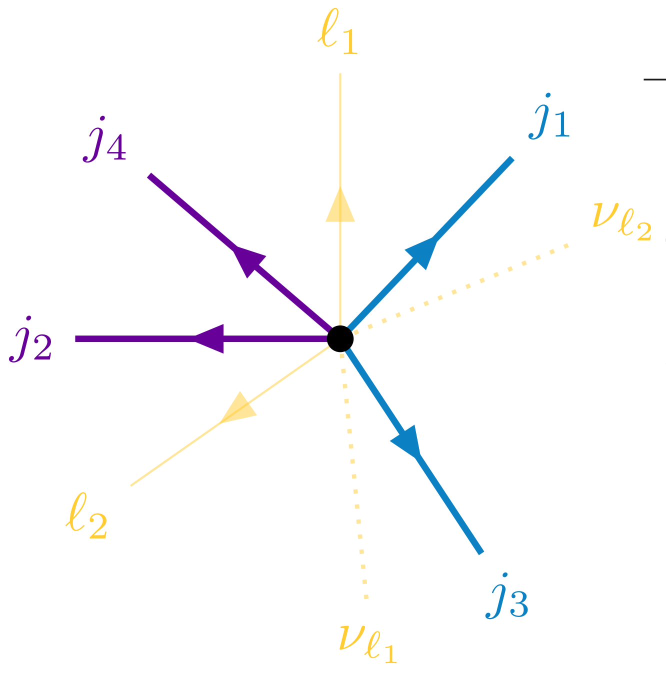




- **Control regions (CRs)** have a lower fraction of signal events and are defined to allow the fit to constrain the background model.
- **Training regions (TRs)** are used for the training of MVAs used in the definition of the signal regions.
  - Analysis utilises two BDT: One for reconstruction of $t\bar{t}H$ system and one for separation of signa from background.
  - Both BDTs originally developed for Run 2 STXS analysis.

Reconstruction BDT

- A Reconstruction BDT was trained to assign jets to parent particles.
  - Within the dilepton channel there a 12 possible jet parent particle assignments.

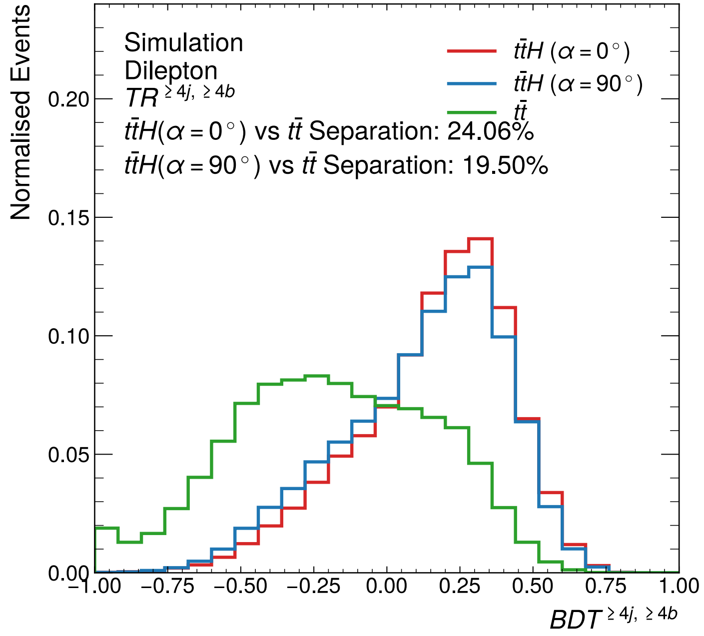


	j_1	j_2	j_3	j_4
1	t	$\bar{t}$	H	H
2	$\bar{t}$	t	H	H
3	$\bar{t}$	H	H	t
4	t	H	H	$\bar{t}$
5	H	H	t	$\bar{t}$
6	H	H	$\bar{t}$	t
7	H	t	$\bar{t}$	H
8	H	$\bar{t}$	t	H
9	t	H	$\bar{t}$	H
10	$\bar{t}$	H	t	H
11	H	t	H	$\bar{t}$
12	H	$\bar{t}$	H	t

Classification BDT

- A BDT was trained to separate $t\bar{t}H$ signal from backgrounds.
  - Was trained with CP-even $t\bar{t}H$ signal and $t\bar{t} + \geq 1b$ background.
  - Separate trainings performed in each of the training regions.
  - Achieves a separation around of 24% for CP-even and 19% for CP-odd events in the Dilepton channel.

$$\langle S^2 \rangle = \frac{1}{2} \sum_{i=1}^{N} \frac{(s_i - b_i)^2}{(s_i + b_i)},$$



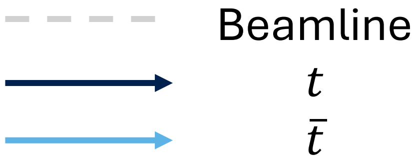
CP Sensitive Observables

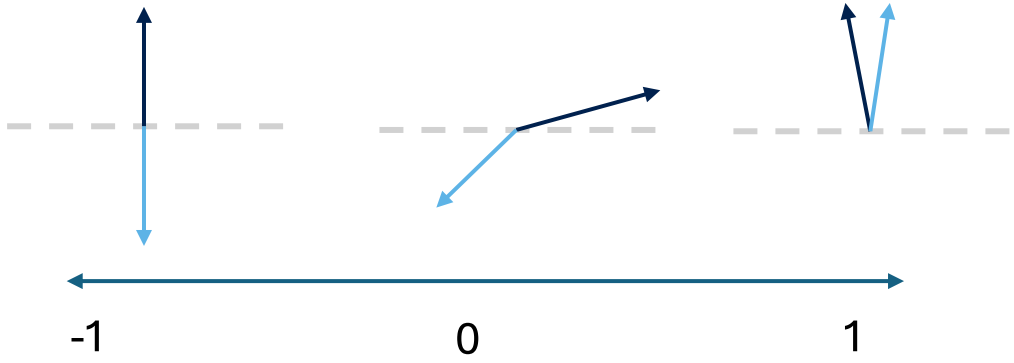
- Analysis achieves sensitivity to CP properties of the top-Higgs Yukawa coupling by using observables sensitive to the differences in kinematics between CP-even and CP-odd $t\bar{t}H$ events.
- Phenomenological studies have proposed a variety of different observables to use.
- Observables generally rely on being able to fully reconstruct the $t\bar{t}H$ system.
- It's possible to evaluate observables in a wide variety of rest frames.

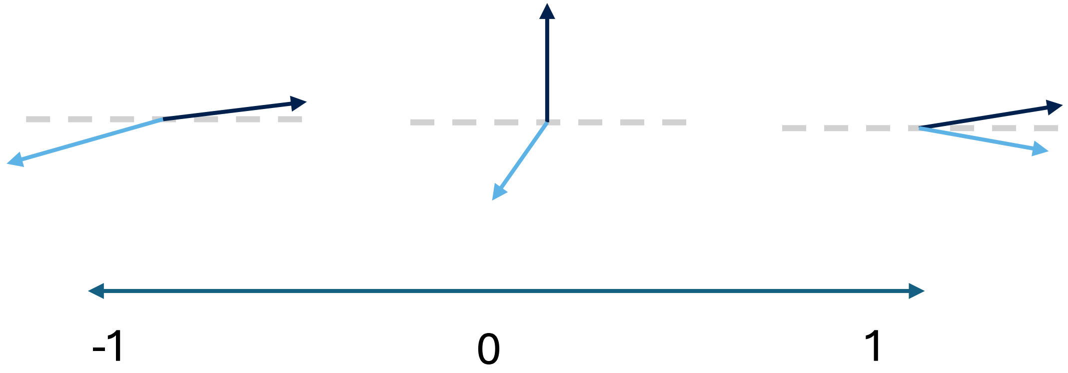
Examples of observables

Observable	Definition	Reference
P_T^H	$-$	
$\Delta\eta_{t\bar{t}}$	$\lvert\eta_t - \eta_{\bar{t}}\rvert$	
$\Delta\phi_{t\bar{t}}$	$\lvert\phi_t - \phi_{\bar{t}}\rvert$	
$m_{t\bar{t}H}$	$(p_t + p_{\bar{t}} + p_H)^2$	
$m_{t\bar{t}}$	$(p_t + p_{\bar{t}})^2$	
b_1	$\frac{(\vec{p}_t \times \vec{n})\cdot(\vec{p}_{\bar{t}} \times \vec{n})}{p_T^t p_T^{\bar{t}}}$	Link
b_2	$\frac{(\vec{p}_t \times \vec{n})\cdot(\vec{p}_{\bar{t}} \times \vec{n})}{\lvert\vec{p}_t\rvert\lvert\vec{p}_{\bar{t}}\rvert}$	
b_3	$\frac{p_t^x p_{\bar{t}}^x}{p_T^t p_T^{\bar{t}}}$	
b_4	$\frac{p_t^z p_{\bar{t}}^z}{\lvert\vec{p}_t\rvert\lvert\vec{p}_{\bar{t}}\rvert}$	

CP Sensitive Observables





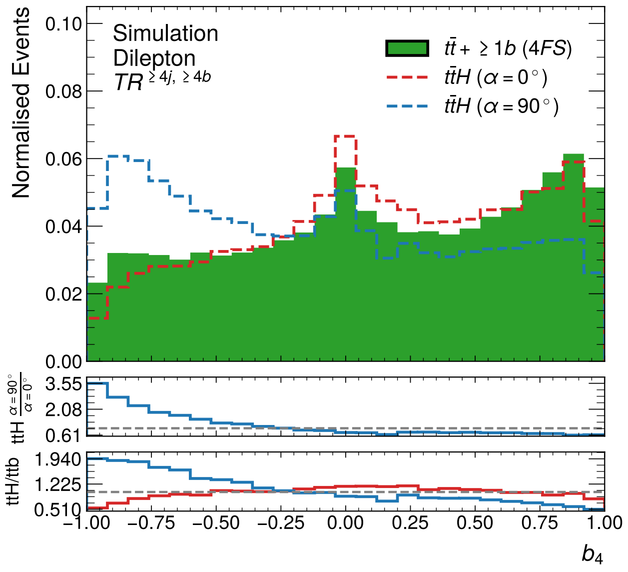


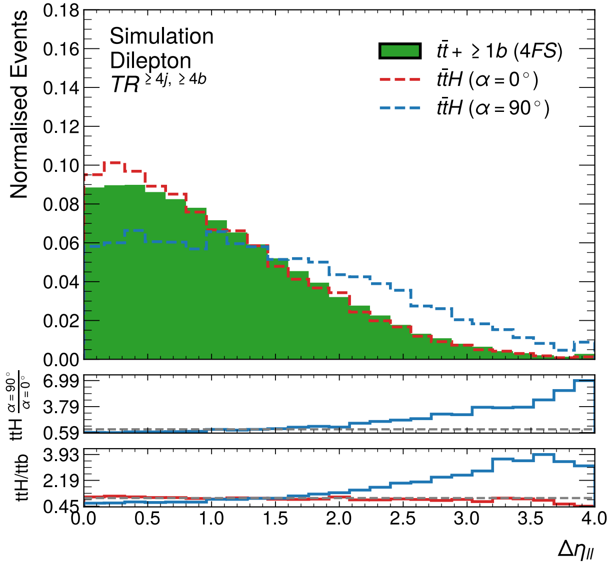
$$b_2 = \frac{(\vec{p}_t \times \vec{n}) \cdot (\vec{p}_{\bar{t}} \times \vec{n})}{|\vec{p}_t||\vec{p}_{\bar{t}}|}$$

$$b_4 = \frac{p_t^z p_{\bar{t}}^z}{|\vec{p}_t||\vec{p}_{\bar{t}}|}$$

- The observables b_1 - b_4 are all distributed between -1 and 1.
- Diagrams attempts to show roughly what angular separation are encapsulated by observables used in the analysis.
  - Just a rough sketch for visualisation purposes.

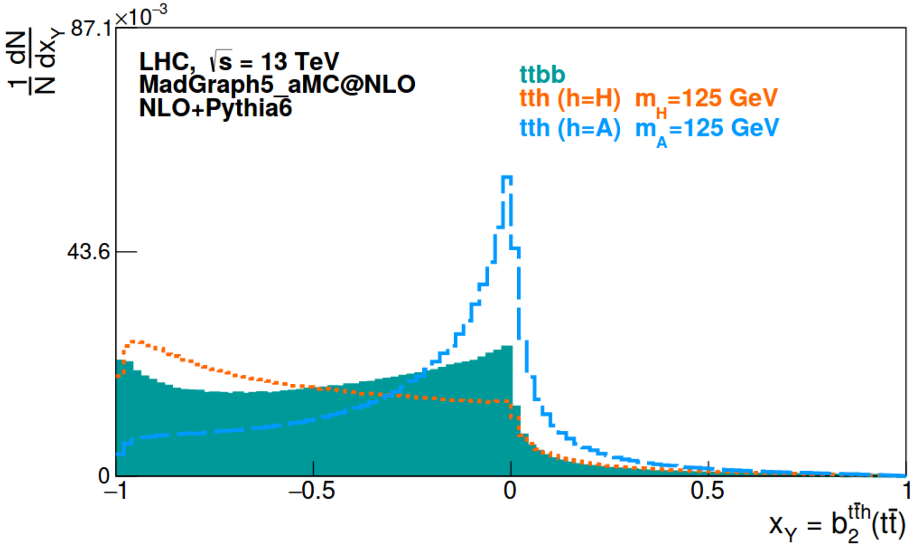
CP Sensitive Observables | Dilepton





- Within the dilepton channel two different observables are used.
  - Due to the reconstruction of events failing for a fraction of events.
- Fully reconstructed events make use of the b_4 observable:
  - There is an enhancement in top quarks travelling in opposite directions and closer to the beamline for *CP*-odd $t\bar{t}H$.
- For events that can't be fully reconstructed make use of $\Delta\eta_{ll}$:
  - Acts a proxy for the pseudorapidity difference between top quarks, which is larger for *CP*-odd $t\bar{t}H$.

CP Sensitive Observables | l+jets



- → In the l+jets channel the $b_2^{t\bar{t}H}$, which is evaluated in the $t\bar{t}H$ rest frame.
  - → Relies on the smaller azimuthal separation of top quarks and larger fraction of longitudinal momentum for CP-odd $t\bar{t}H$.
- → Within the boosted channel the output of the classification BDT is used.
  - → *CP*-odd $t\bar{t}H$ and tH is significantly enhanced compared to *CP*-even in the the boosted region.

Analysis Regions

Channel (TR)	Final SRs and CRs	Classification BDT selection	Fitted observable
Dilepton ($TR^{\geq 4j, \geq 4b}$)	$CR_{\text{no-reco}}^{\geq 4j, \geq 4b}$	–	$\Delta\eta_{\ell\ell}$
	$CR^{\geq 4j, \geq 4b}$	$BDT^{\geq 4j, \geq 4b} \in [-1, -0.086)$	b_4
	$SR_1^{\geq 4j, \geq 4b}$	$BDT^{\geq 4j, \geq 4b} \in [-0.086, 0.186)$	b_4
	$SR_2^{\geq 4j, \geq 4b}$	$BDT^{\geq 4j, \geq 4b} \in [0.186, 1]$	b_4
ℓ+jets ($TR^{\geq 6j, \geq 4b}$)	$CR_1^{\geq 6j, \geq 4b}$	$BDT^{\geq 6j, \geq 4b} \in [-1, -0.128)$	b_2
	$CR_2^{\geq 6j, \geq 4b}$	$BDT^{\geq 6j, \geq 4b} \in [-0.128, 0.249)$	b_2
	$SR^{\geq 6j, \geq 4b}$	$BDT^{\geq 6j, \geq 4b} \in [0.249, 1]$	b_2
ℓ+jets (TR_{boosted})	SR_{boosted}	$BDT^{\text{boosted}} \in [-0.05, 1]$	BDT^{boosted}

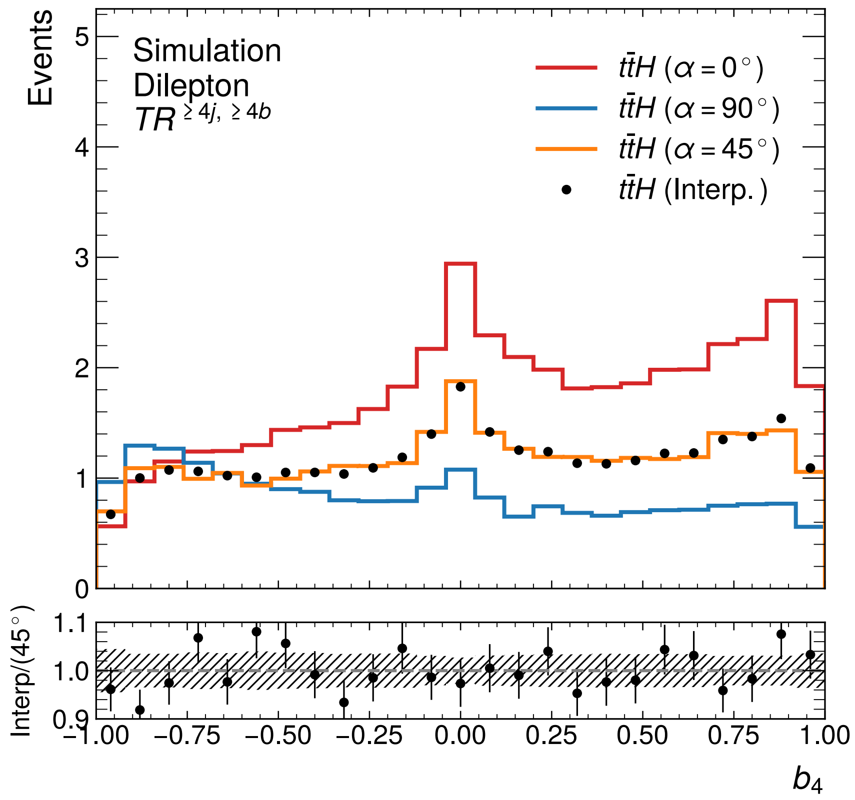
→ From the training regions (TRs) a further eight regions are defined based on cut on the classification BDT.

→ Regions with a signal to background ratio of >7.7% are classified as signal regions.

Signal Modelling

$t\bar{t}H$ Modelling: $N(\kappa_t', \alpha) = (\kappa_t' \cos\alpha)^2 \cdot N_{CP-even} + (\kappa_t' \sin\alpha)^2 \cdot N_{CP-odd}$

- →The $t\bar{t}H$ signal is modelled by interpolating between two MadGraph(NLO) samples, generated under a CP-even and CP-odd hypothesis.
- →The effects of interference are neglected
  - →Verified by comparing to maximally mixed sample.



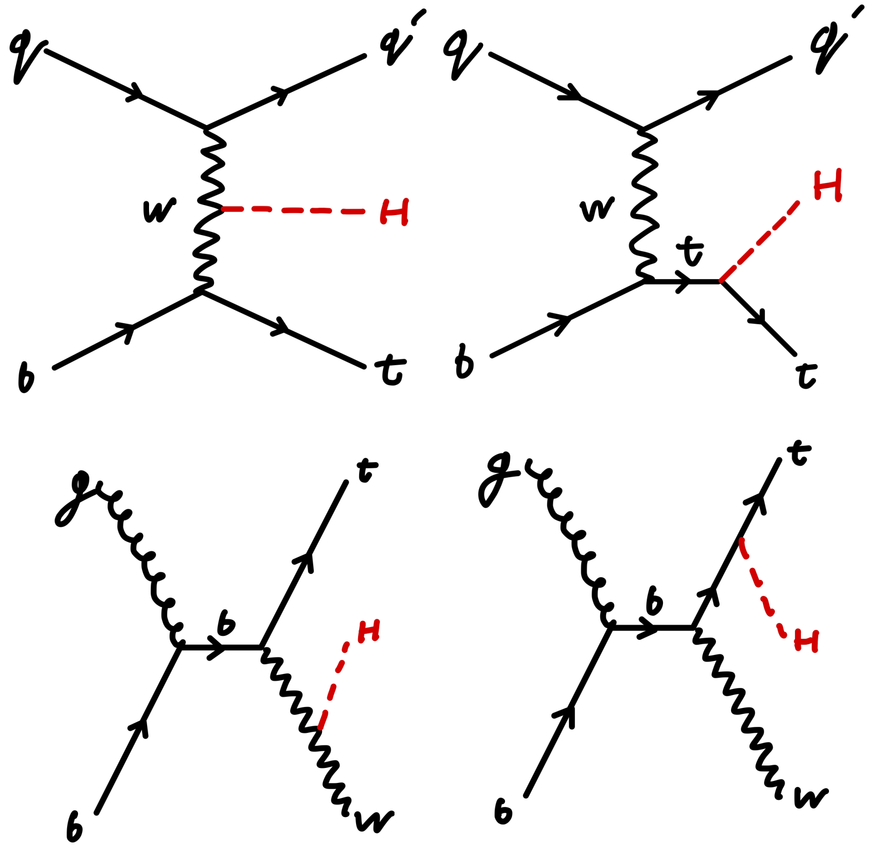
Signal Modelling

tH Modelling: $\frac{N(\kappa_t', \alpha)}{N_{CP-even}} = A \cdot (\kappa_t' C_\alpha)^2 + B \cdot (\kappa_t' S_\alpha)^2 + C \cdot (\kappa_t' C_\alpha) + D \cdot (\kappa_t' S_\alpha) + E \cdot (\kappa_t' C_\alpha) \cdot (\kappa_t' S_\alpha) + F$

→ Modelling of tH signal is complicated by the interference between diagrams where the Higgs couples to the top and those where it couples to a $W^\pm$.

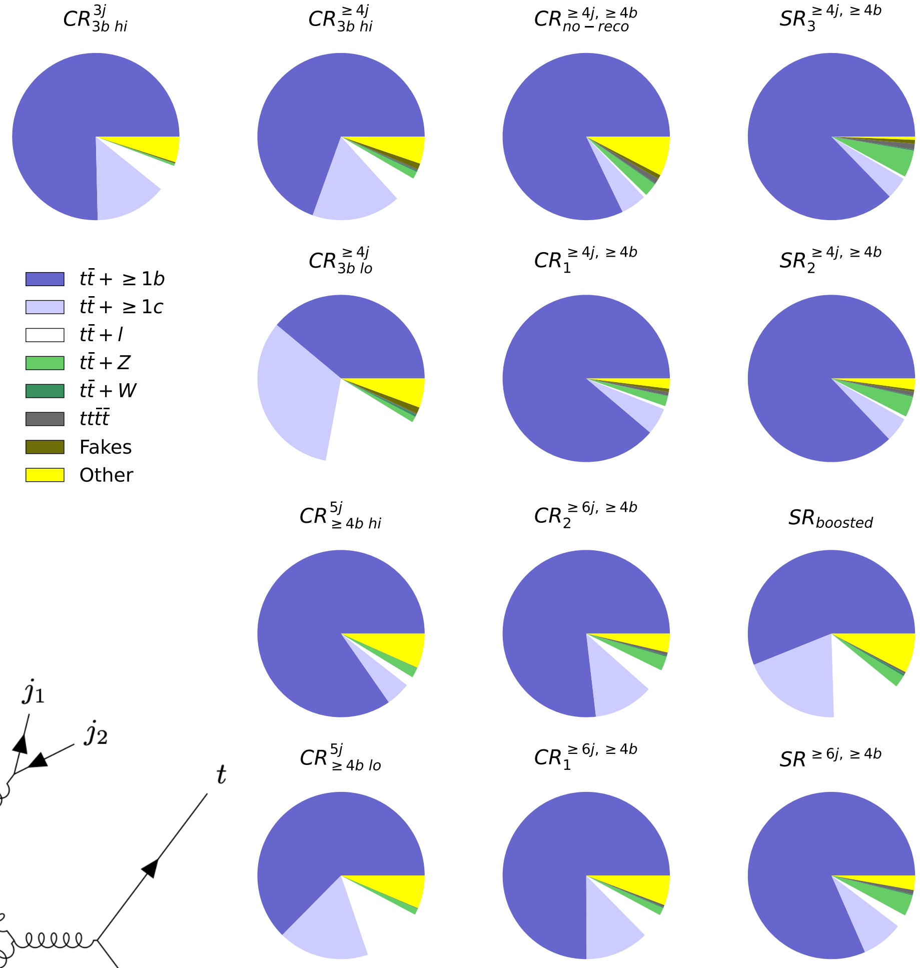
→ The values of the coefficients *A-F* are determined by performing a fit to 11 different MC samples produced with various values of κ_t' and α.

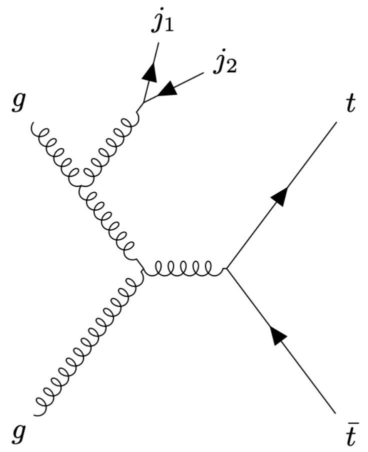
  → An independent fit is performed in each bin of each region in the analysis.



$t\bar{t}$ + jets Background

- The analysis is heavily impacted by the presence of the the $t\bar{t}$ +jets background.
  - Contributions from other backgrounds make a minor contribution to most regions.
- The background is divided into components based on the flavour of the additional jets.
  - $t\bar{t} + \geq 1b$: events where there is at least one true b-jet.
  - $t\bar{t} + \geq 1c$: events with no b-jets but at least one c-jet.
  - $t\bar{t}$ +light: any other events
- Modelling of the $t\bar{t} + \geq 1b$ component is done in the 4 flavour scheme, while the others are modelled in the 5 flavour scheme.





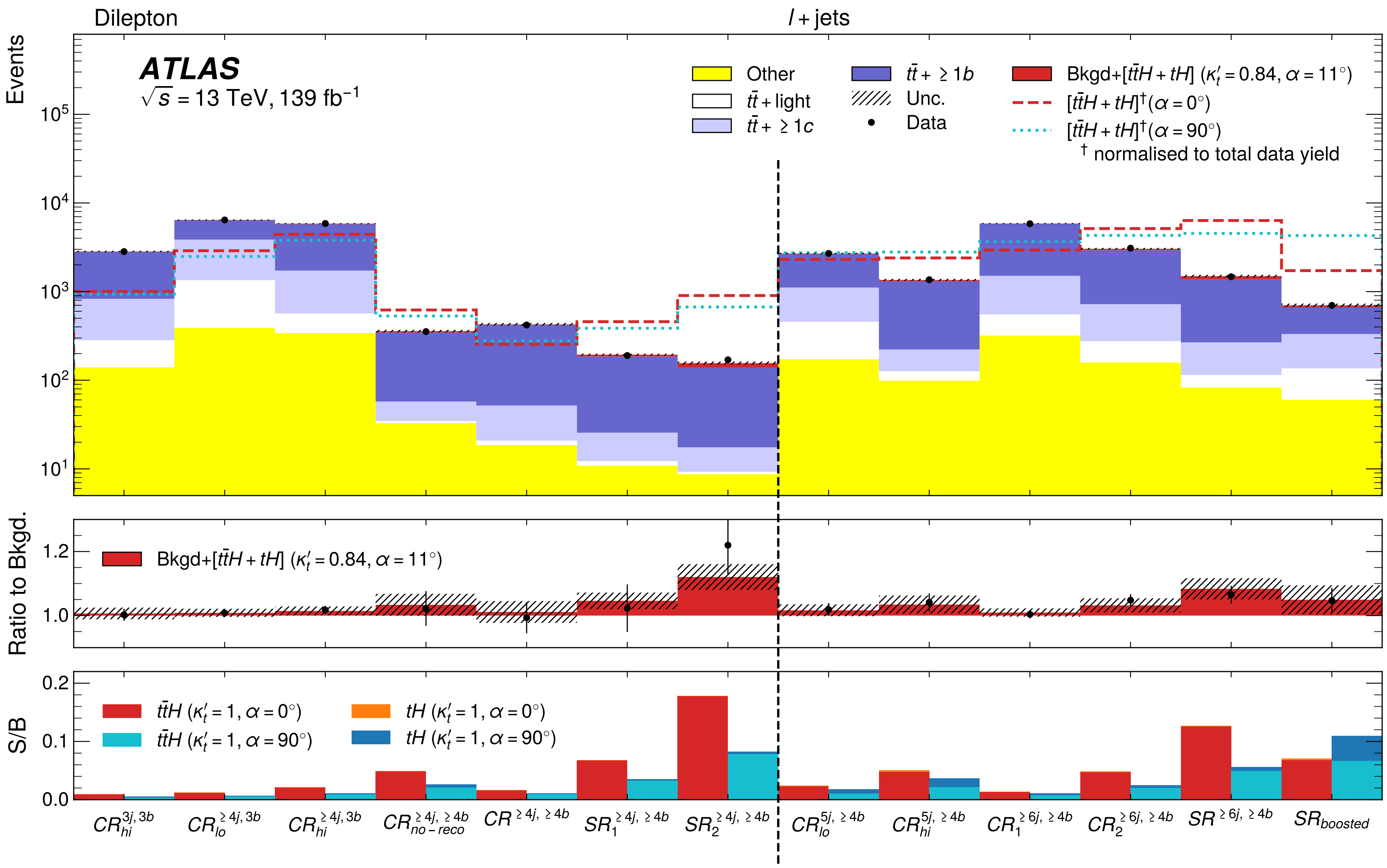
Systematic uncertainties

- Systematic uncertainties shown here apply just to $t\bar{t}$ +jets background.

Uncertainty source	Description		Components
$t\bar{t}$ cross-section	Up or down by 6%		$t\bar{t}$+light
$t\bar{t}+\geq 1b$ normalisation	Free-floating		$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1c$ normalisation	Up or down by 100%		$t\bar{t}+\geq 1c$
ISR	Varying $\mu_{\mathrm{R}}^{\mathrm{ISR}}$ (PS), μ_{R}&μ_{F} (ME)	in POWHEG-BOX+PYTHIA 8 $t\bar{t}+b\bar{b}$(4FS)	$t\bar{t}+\geq 1b$
		in POWHEG-BOX+PYTHIA 8 $t\bar{t}$(5FS)	$t\bar{t}+\geq 1c$, $t\bar{t}$+light
FSR	Varying $\mu_{\mathrm{R}}^{\mathrm{FSR}}$ (PS)	POWHEG-BOX+PYTHIA 8 $t\bar{t}+b\bar{b}$(4FS)	$t\bar{t}+\geq 1b$
		in POWHEG-BOX+PYTHIA 8 $t\bar{t}$(5FS)	$t\bar{t}+\geq 1c$, $t\bar{t}$+light
NLO matching	MADGRAPH5_aMC@NLO+PYTHIA 8 $t\bar{t}$(5FS)	vs. POWHEG-BOX+PYTHIA 8 $t\bar{t}$(5FS)	All
PS & hadronisation	POWHEG-BOX+HERWIG 8 $t\bar{t}$(5FS)	vs. POWHEG-BOX+PYTHIA 8 $t\bar{t}$(5FS)	All
Flavour scheme	POWHEG-BOX+PYTHIA 8 $t\bar{t}$(5FS)	vs. POWHEG-BOX+PYTHIA 8 $t\bar{t}+b\bar{b}$(4FS)	$t\bar{t}+\geq 1b$
$t\bar{t}+\geq 1b$ fractions	Variation of the relative fractions of $t\bar{t}+\geq 2b$ and $t\bar{t}+1b/B$		$t\bar{t}+\geq 1b$
$p_T^{b\bar{b}}$ shape	Shape mismodelling measured from data		$t\bar{t}+\geq 1b$

- Emphasis is placed on understanding the uncertainties applied to $t\bar{t}$ + jets with 13 nuisance parameters applied to the $t\bar{t}$ +$\geq 1b$ component alone.
- This analysis additionally includes uncertainty on the choice of flavour scheme compared to the Run 2 STXS analysis.
  - Results of analysis were found to be particularly sensitive to this effect.

Fit to data | overall



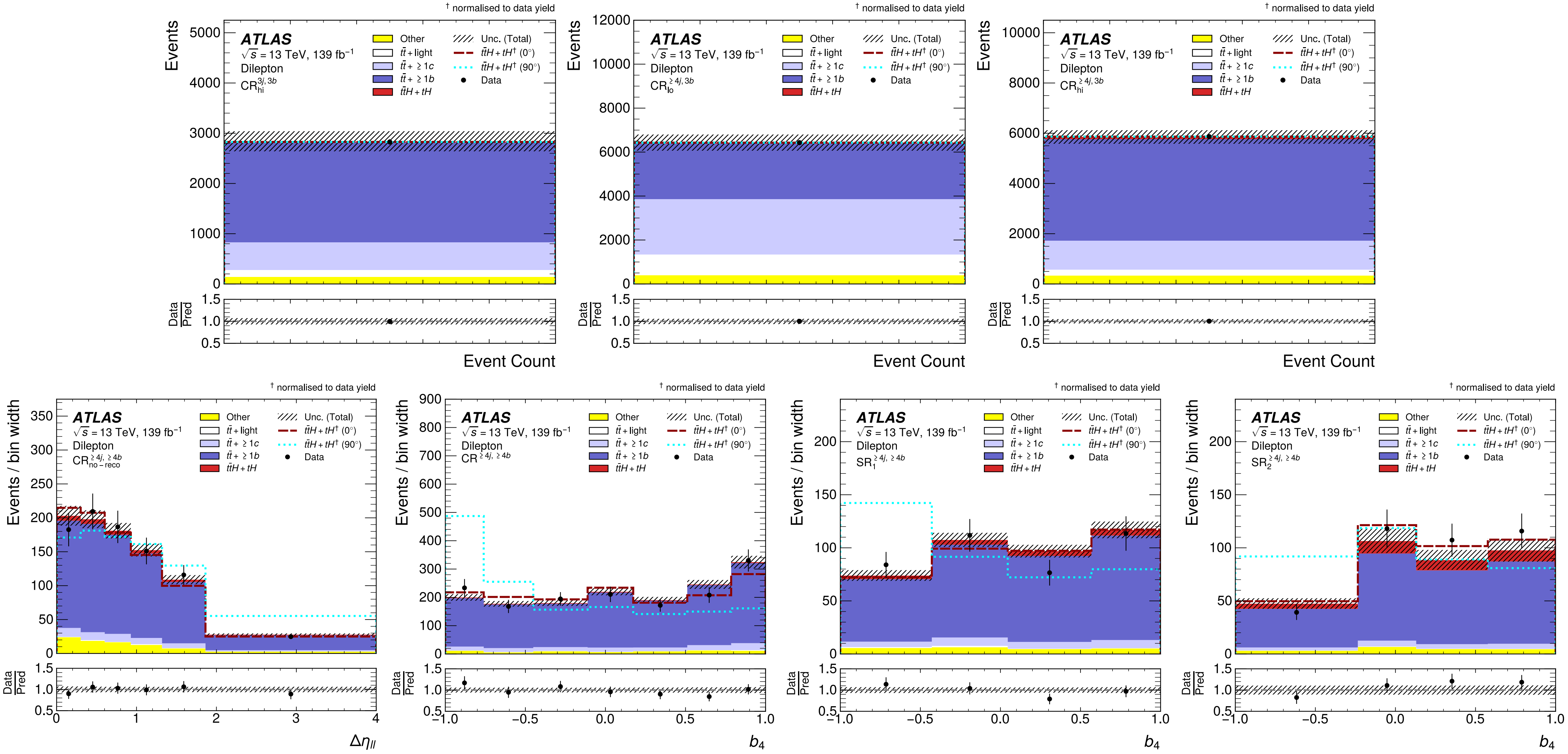
→Results obtained from profile likelihood fit to data.

→Overall there is good post-fit agreement across all regions.

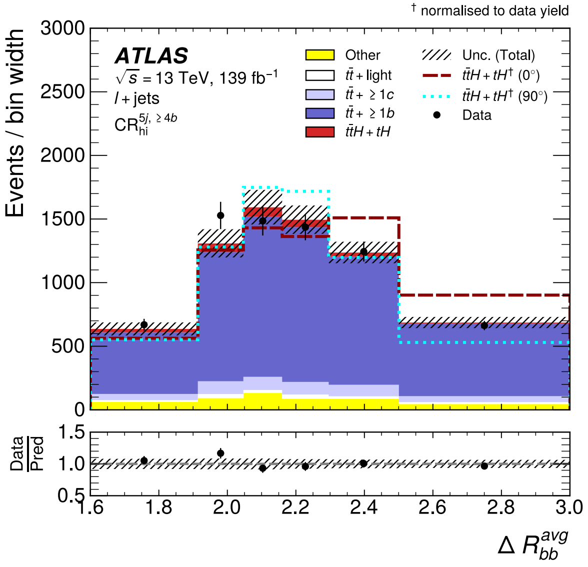
→Significance of signal over background of 1.3σ.

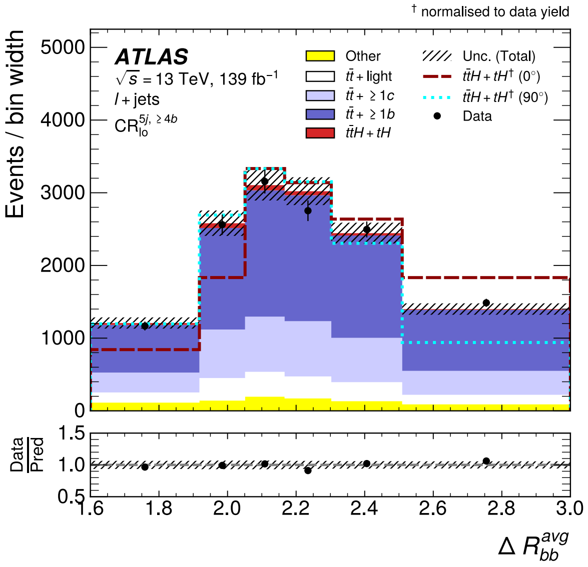
→Goodness of fit evaluated by comparing to a saturated model give a probability of 80% for compatibility to the data.

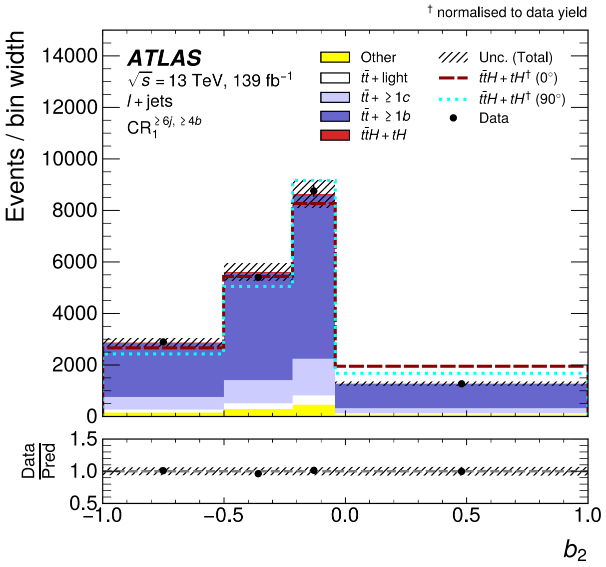
Fit to data | Dilepton

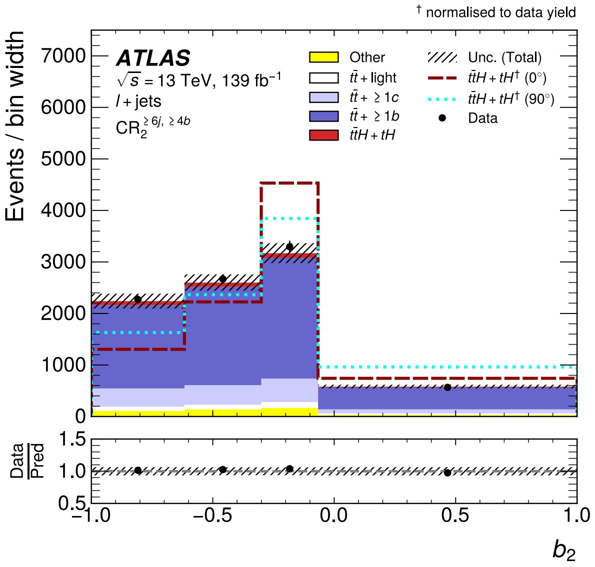


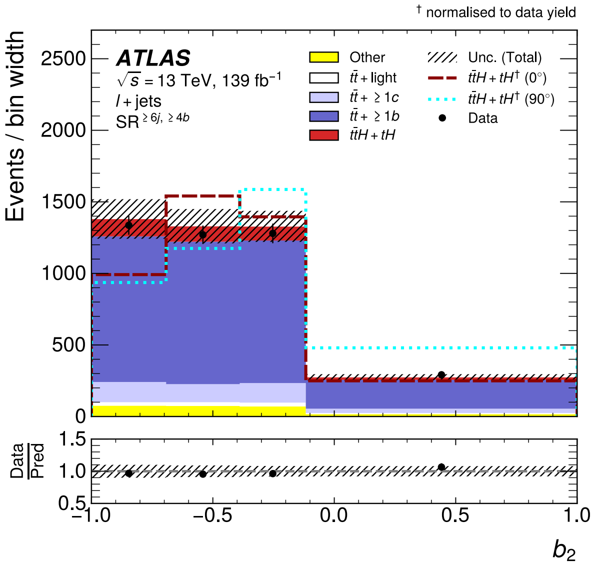
Fit to data | l+jets

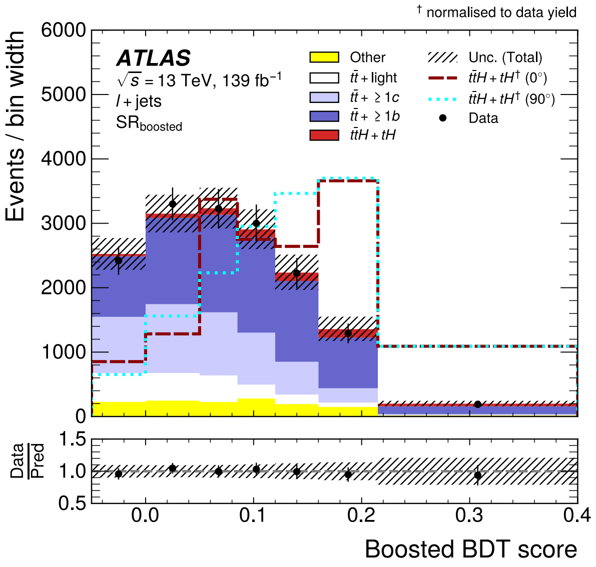




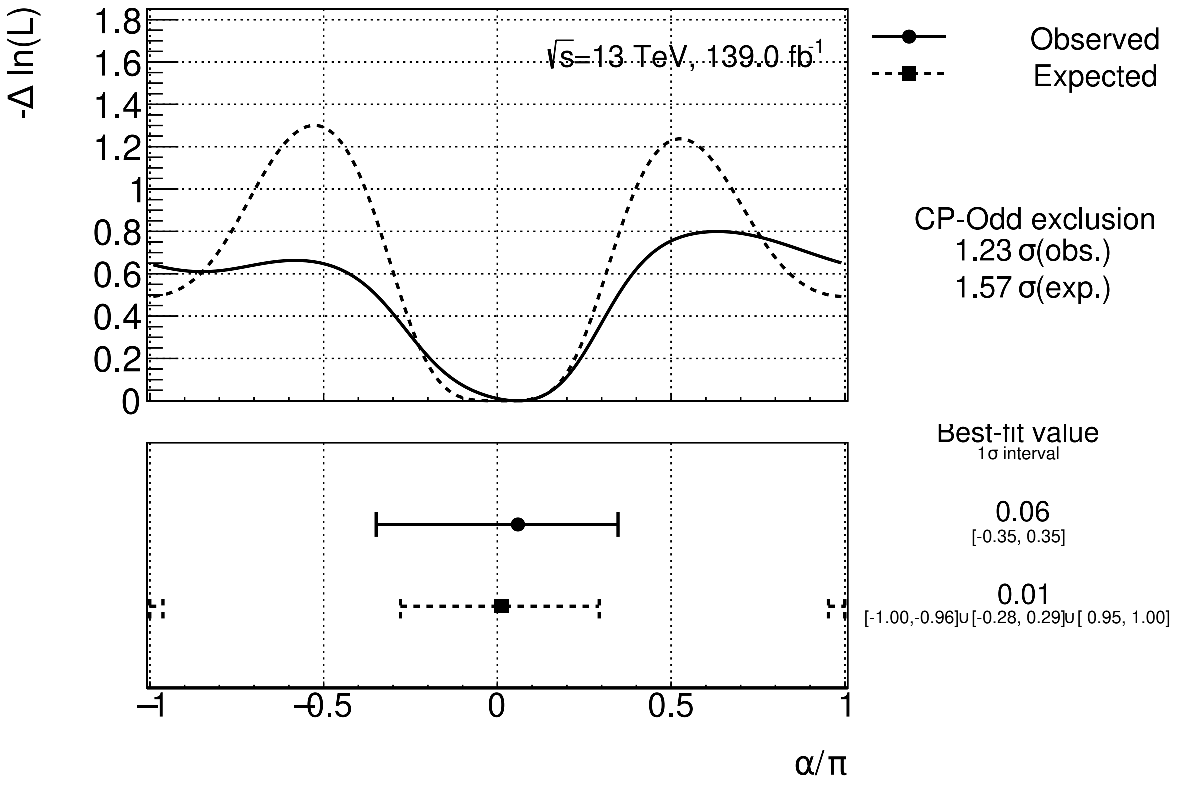


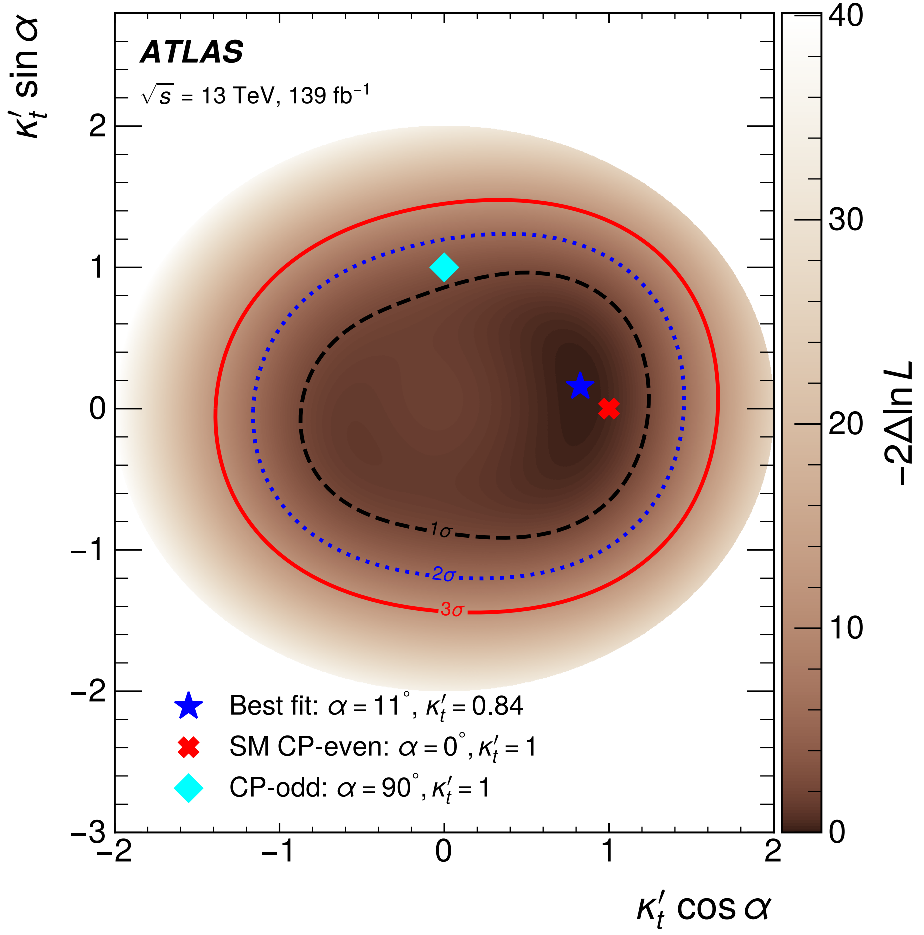






Results



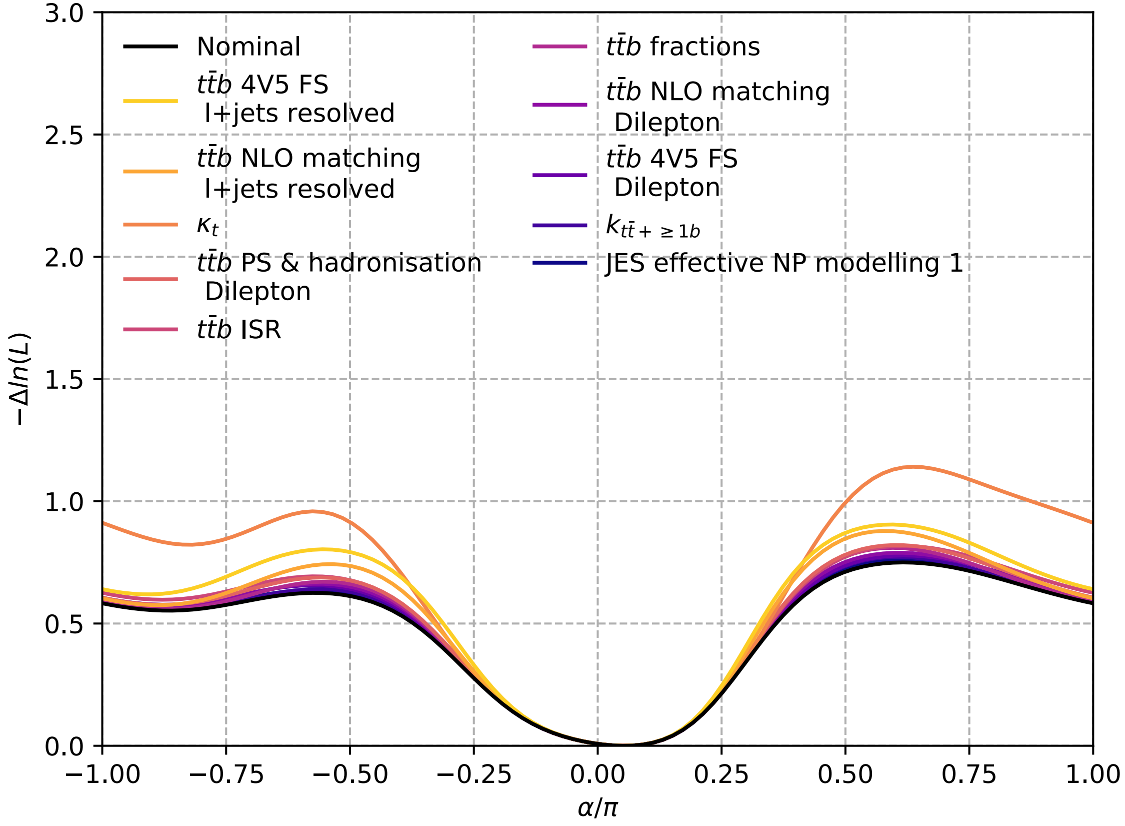


→**Observed**(expected) CP-odd exclusion significance of **1.23σ** (1.57σ)

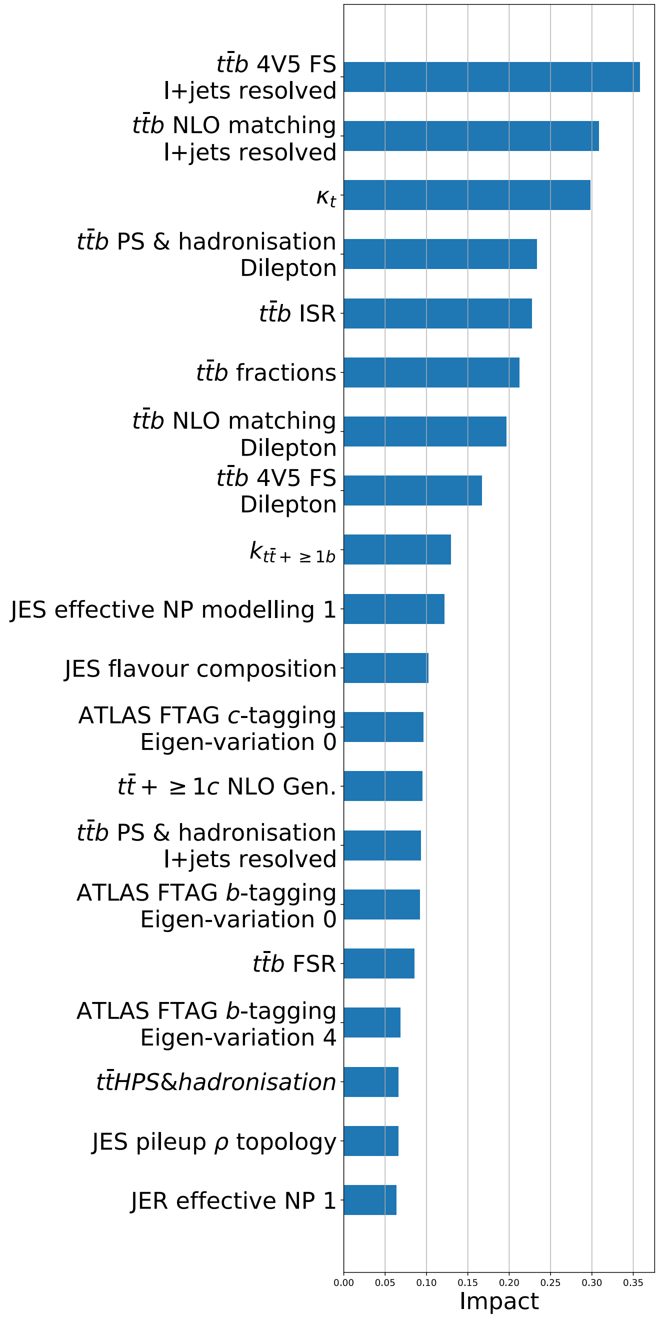
→Best fit value of $^{\alpha}/_{\pi} = 0.06^{+0.30}_{-0.44}$ → $\boldsymbol{\alpha = 11^{\circ}{}^{+55^{\circ}}_{-77^{\circ}}}$, $\boldsymbol{\kappa'_t = 0.84^{+0.30}_{-0.46}}$

Impact of Systematic Uncertainties

- Contribution of systematics to uncertainty is measured via the change in the $\pm 1\sigma$ exclusion region when fixing the systematic to its best-fit value.



$$\text{Impact} = \sqrt{\sigma_\alpha^2 - \sigma_\alpha'^2}$$



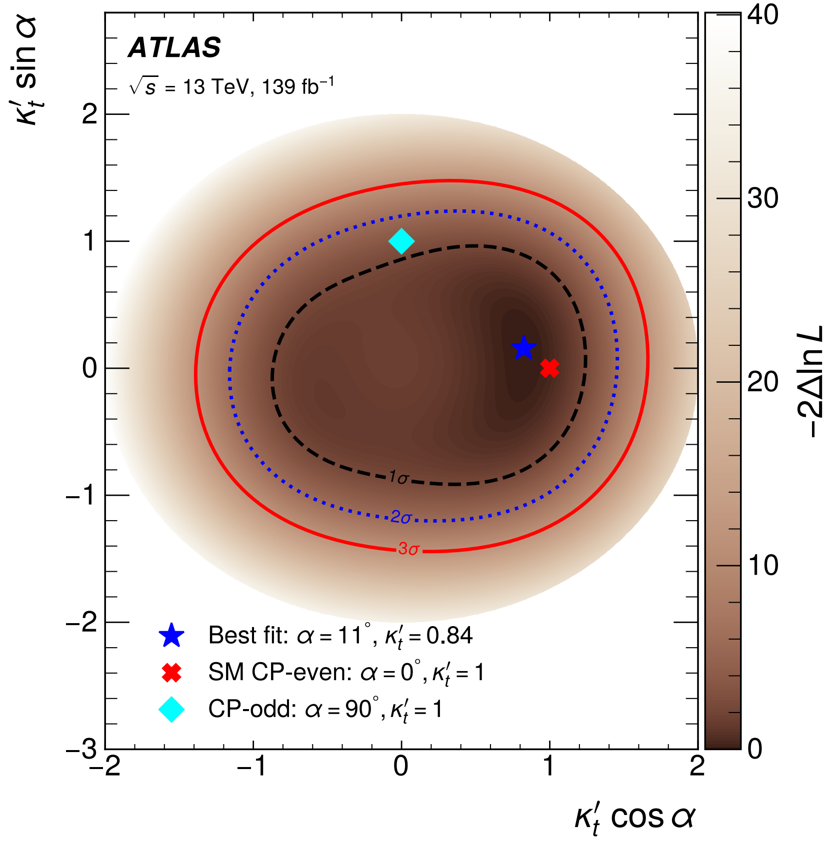
Post-fit uncertainties

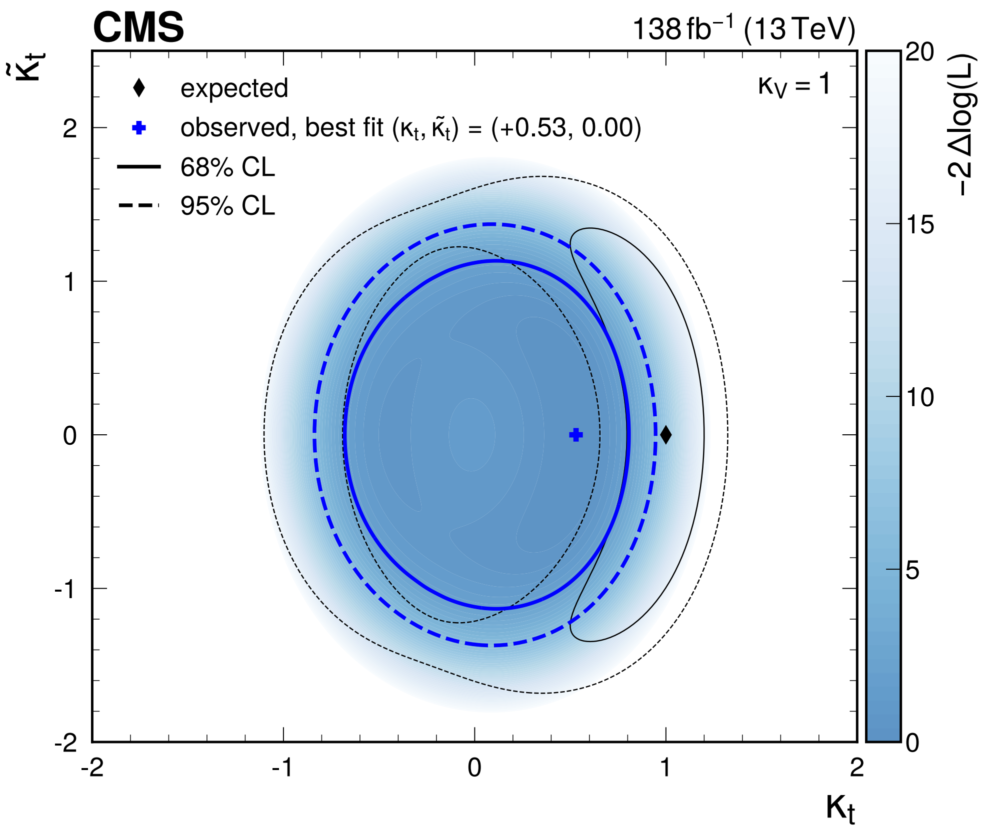
- Post fit uncertainty on α is dominated by contributions from systematic uncertainties.
- A substantial amount of that arises from the $t\bar{t}$ +jets background, which contribute $^{+37^\circ}_{-51^\circ}$ to the overall uncertainty.
  - Analysis would greatly benefit from an improvement in the modelling of this background.
- Statistical uncertainty is also a significant factor to limiting the total precision of the measurement.

Uncertainty source	$\Delta\alpha[^\circ]$	
Process modelling		
$t\bar{t}H$ modelling	+8.8	-14
$t\bar{t}+\geq 1b$ modelling		
$t\bar{t}+\geq 1b$ 4V5 FS	+23	-37
$t\bar{t}+\geq 1b$ NLO matching	+22	-33
$t\bar{t}+\geq 1b$ fractions	+14	-21
$t\bar{t}+\geq 1b$ FSR	+5.2	-9.9
$t\bar{t}+\geq 1b$ PS & hadronisation	+16	-24
$t\bar{t}+\geq 1b$ $p_T^{b\bar{b}}$ shape	+5.4	-4.6
$t\bar{t}+\geq 1b$ ISR	+14	-24
$t\bar{t}+\geq 1c$ modelling	+6.6	-11
$t\bar{t}$+light modelling	+2.5	-4.7
b-tagging efficiency and mis-tag rates		
b-tagging efficiency	+8.7	-15
c-mis-tag rates	+6.7	-11
l-mis-tag rates	+2.3	-2.7
Jet energy scale and resolution		
b-jet energy scale	+1.6	-3.8
Jet energy scale (flavour)	+7.8	-11
Jet energy scale (pileup)	+5.2	-7.9
Jet energy scale (remaining)	+8.1	-13
Jet energy resolution	+5.7	-9.3
Luminosity	+0	-0
Other sources	+4.9	-8
Total systematic uncertainty	+41	-54
$t\bar{t}+\geq 1b$ normalisation	+8.2	-13
κ_t'	+17	-33
Total statistical uncertainty	+32	-49
Total uncertainty	+52	-73

Comparison to recent CMS results

arXiv:2407.10896(accepted by JHEP)



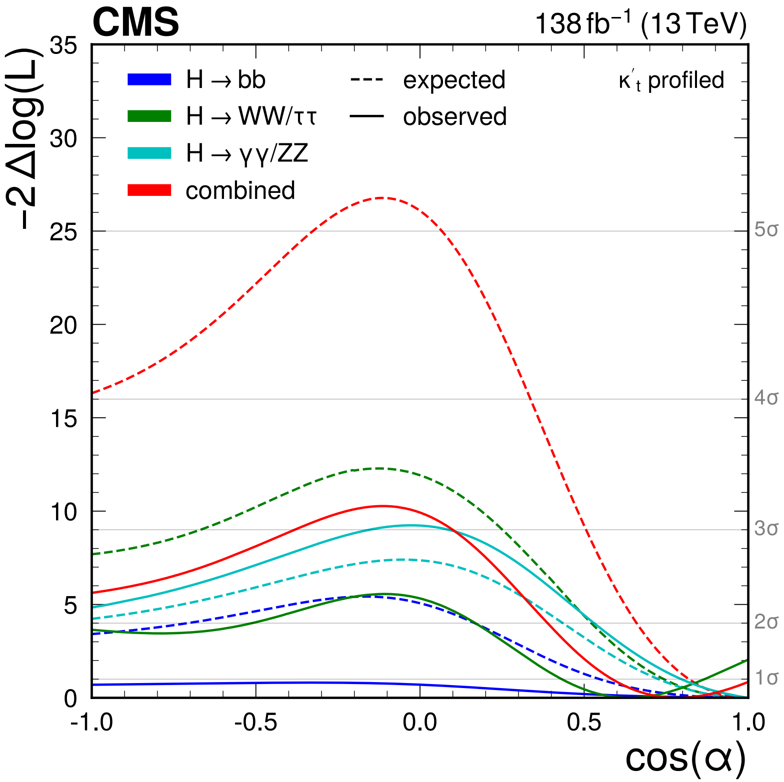


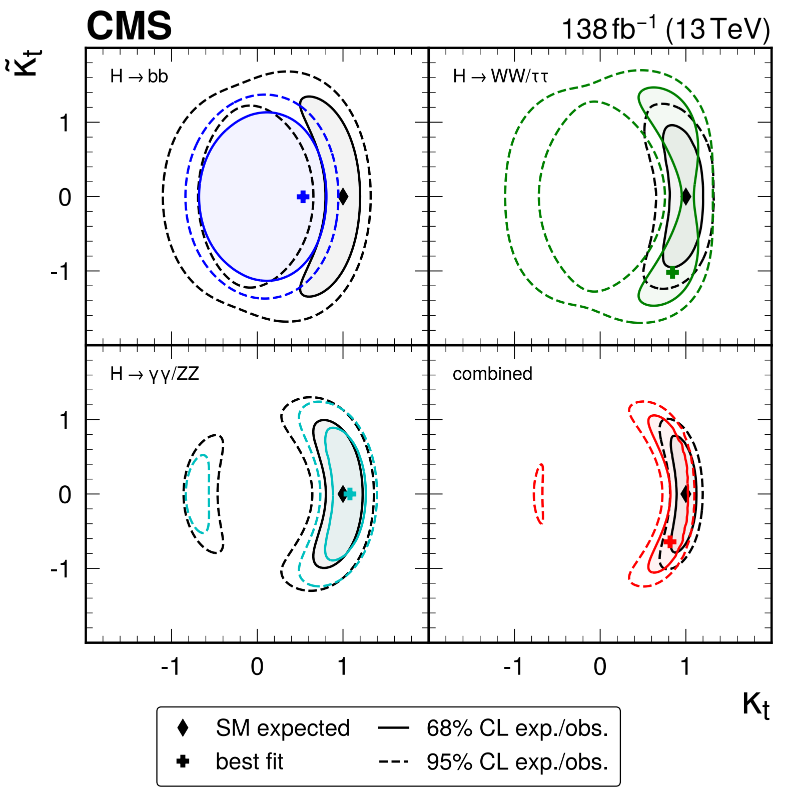
→Recent CMS results present a similar picture to ATLAS results, with results disfavouring a CP-odd coupling at about 1σ.

Recent CMS results

arXiv:2407.10896(accepted by JHEP)

- Recent results from CMS also provide combined limits on the mixing angle α
  - Able to exclude a purely CP-odd coupling at more than 3σ significance.
  - Maximally mixed scenario is still not excluded.





Summary

→ Additional sources of CP-violation are needed to explain the observed matter-antimatter asymmetry.

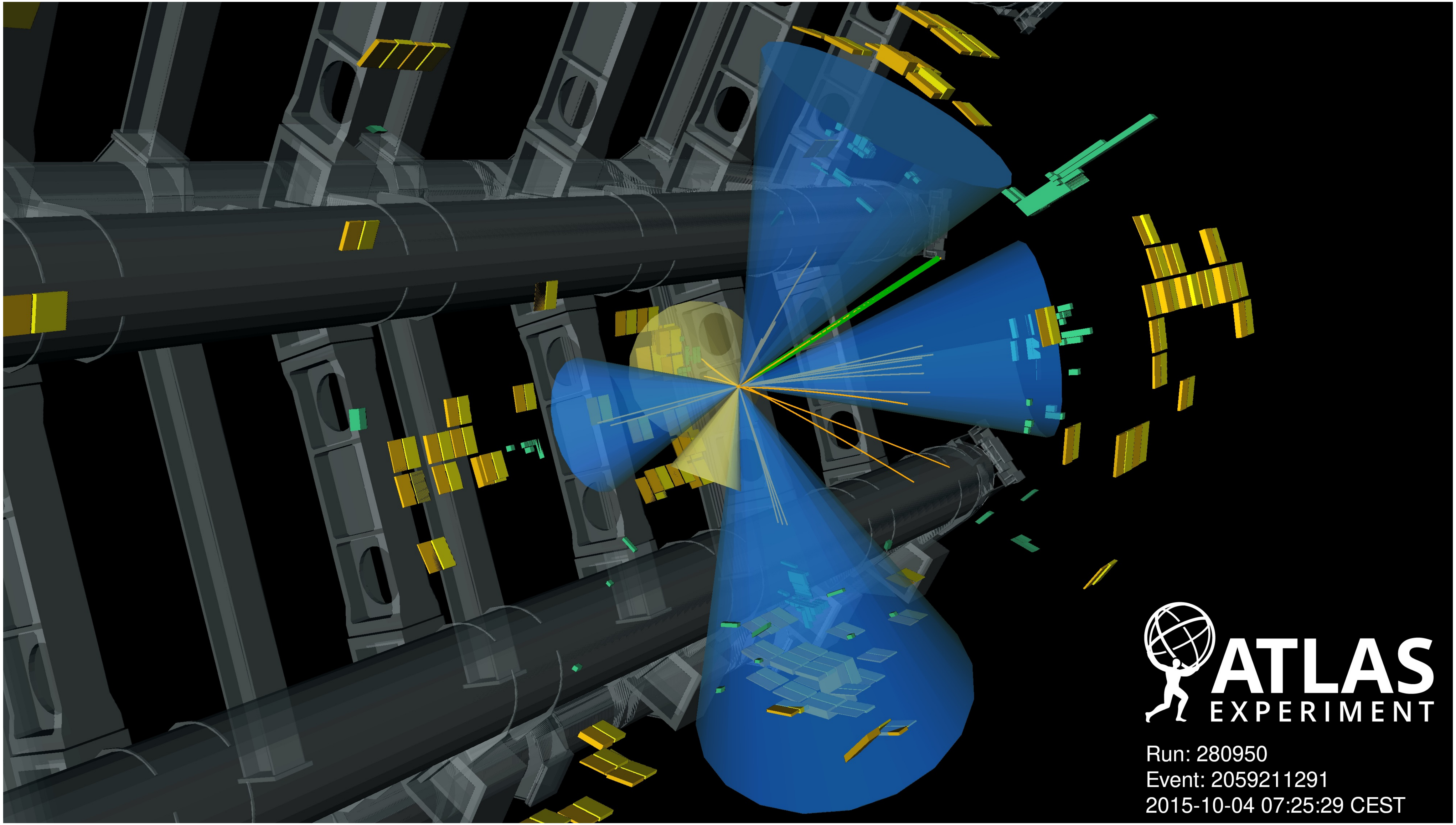
→ The Higgs sector provides a promising area to search for potential new sources.

  → Direct access to the top-Higgs coupling can be gained through probing the $t\bar{t}H$ and tH processes.

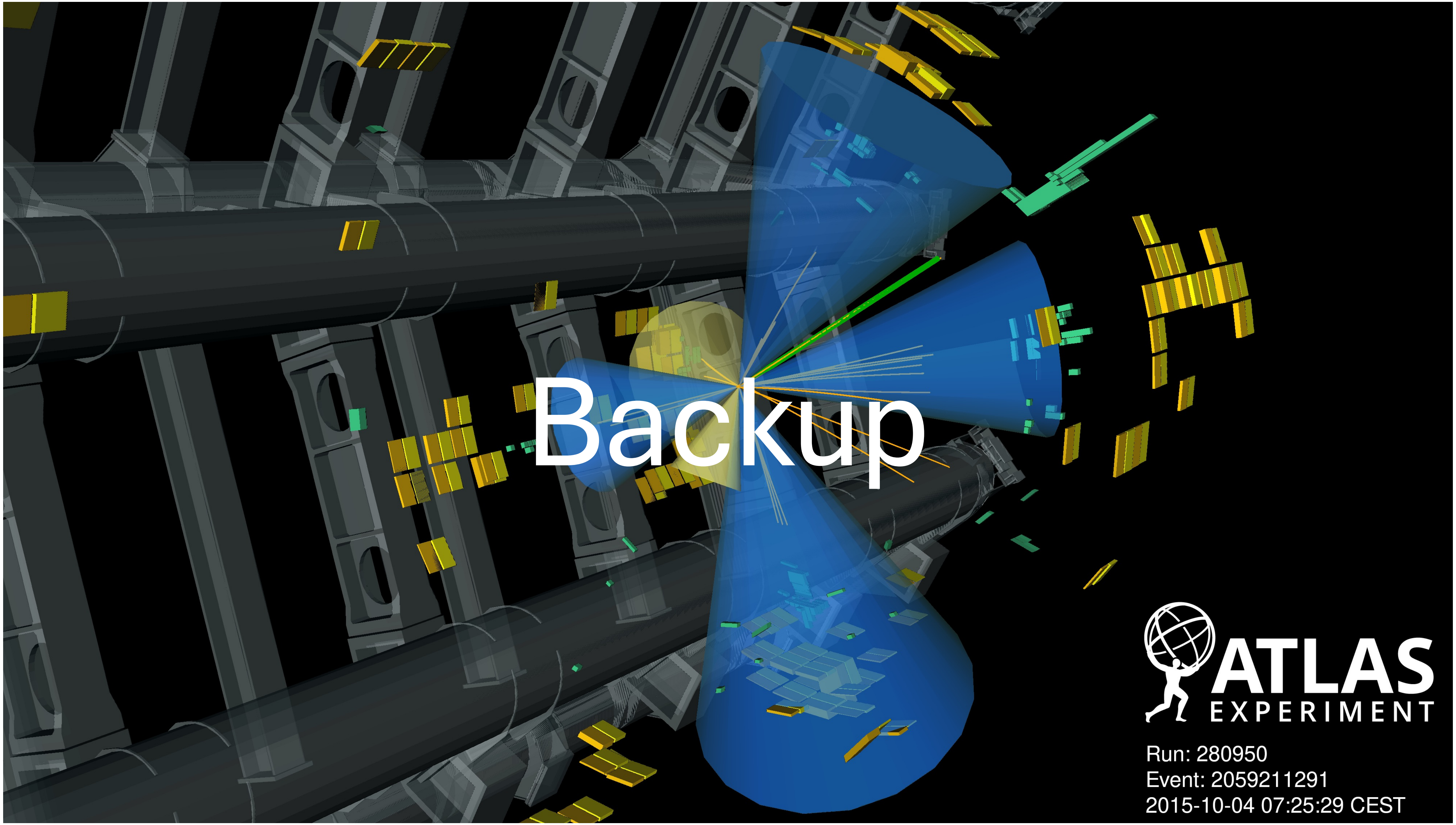
→ The $H \to b\bar{b}$ decay mode provides an interesting but challenging decay channel to study the effects of CP-violation.

  → Improved understanding of the $t\bar{t}$ + jets background would greatly benefit the sensitivity of the measurement.

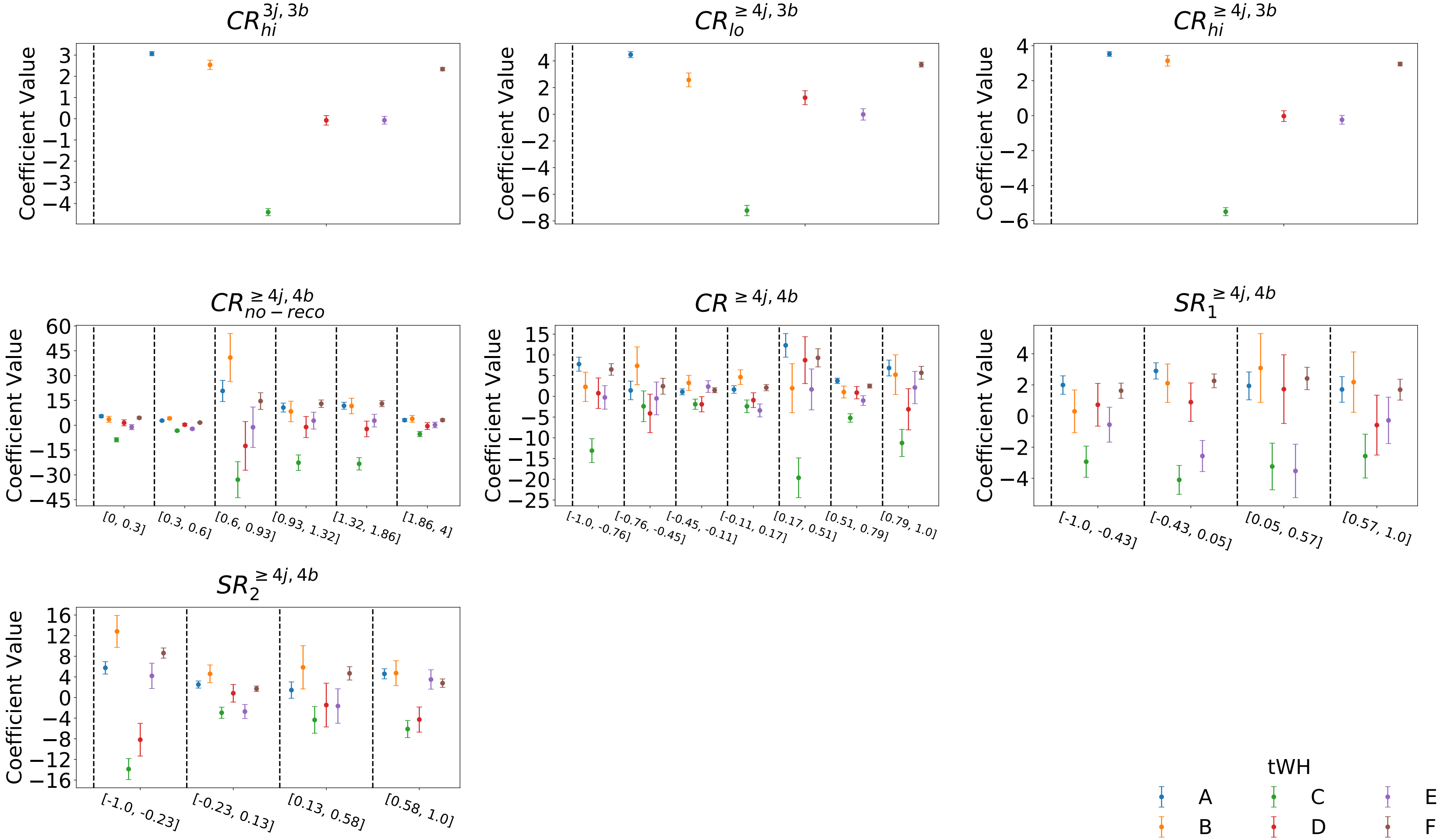
→ While a purely *CP*-odd coupling has been excluded by other results there is still the possibility of mixed states that would introduce a new source of *CP* violation.


ATLAS
EXPERIMENT
Run: 280950
Event: 2059211291
2015-10-04 07:25:29 CEST

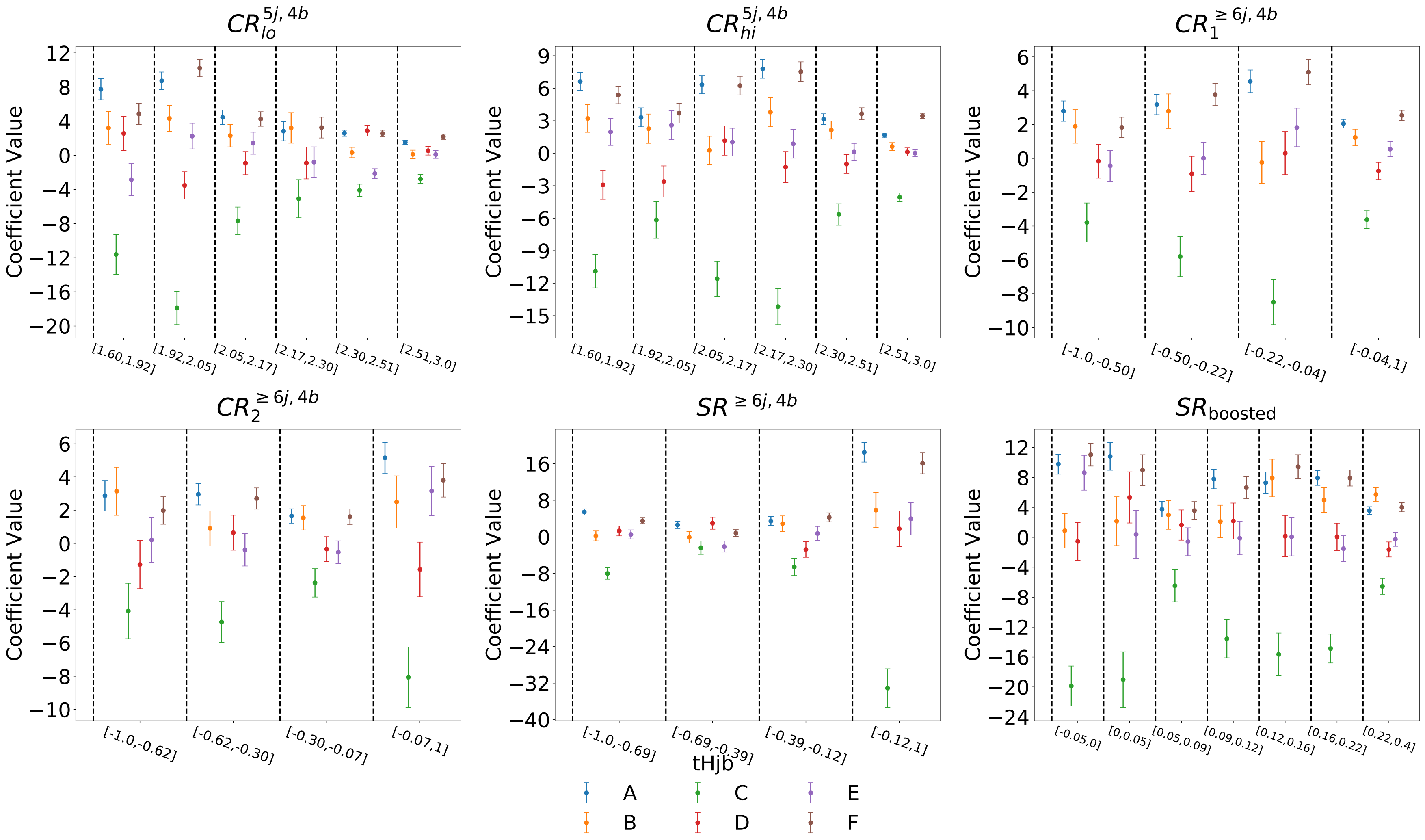
Backup


ATLAS
EXPERIMENT
Run: 280950
Event: 2059211291
2015-10-04 07:25:29 CEST

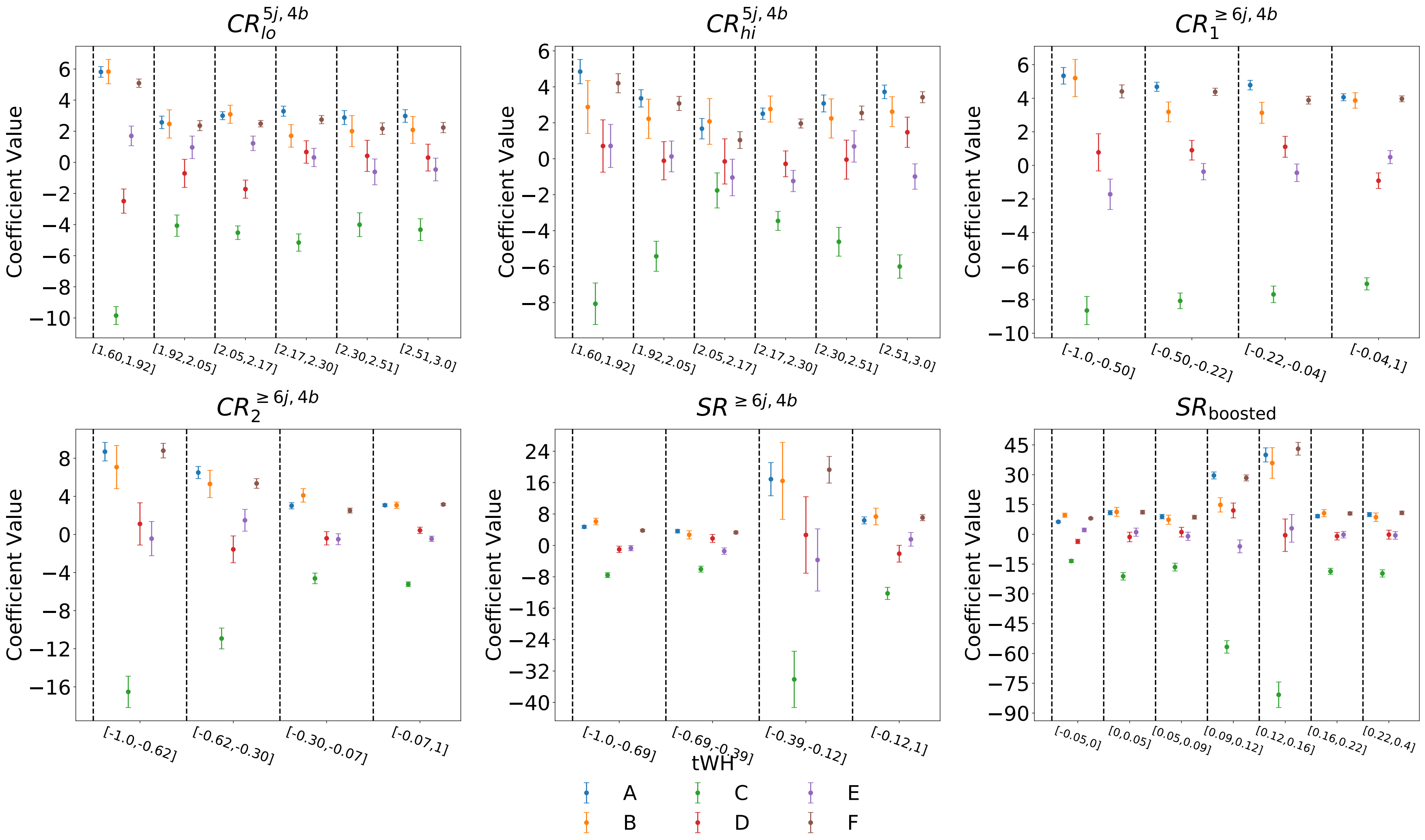
tWH Parametrisation Dilepton



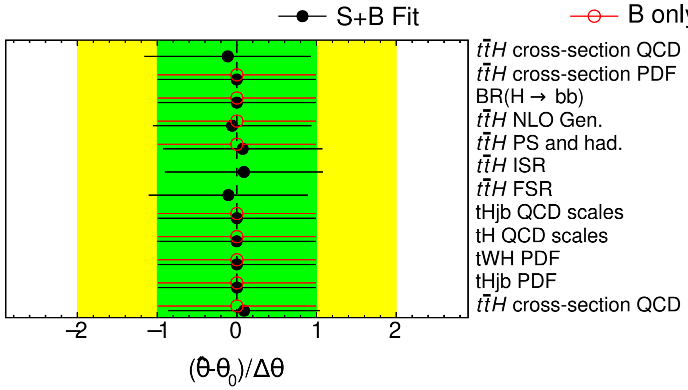
$tHjb$ Parametrisation l+jets

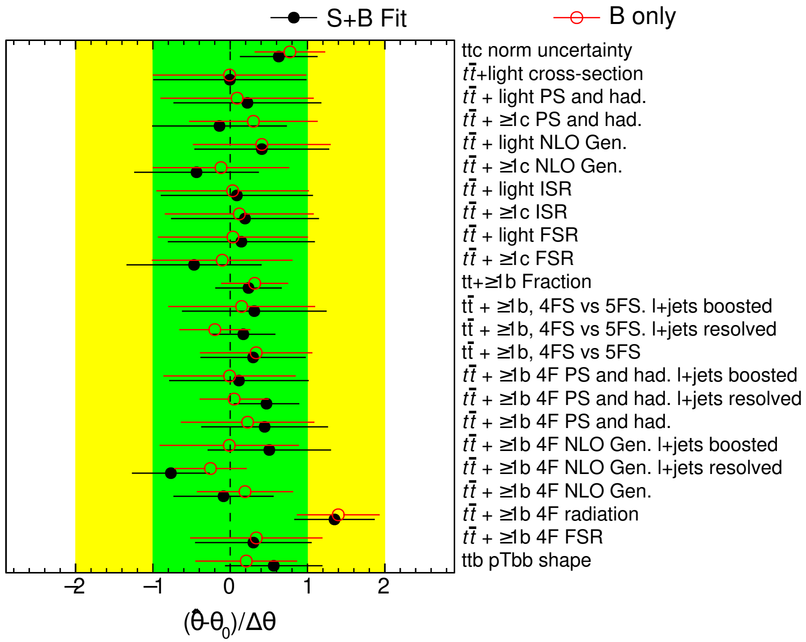


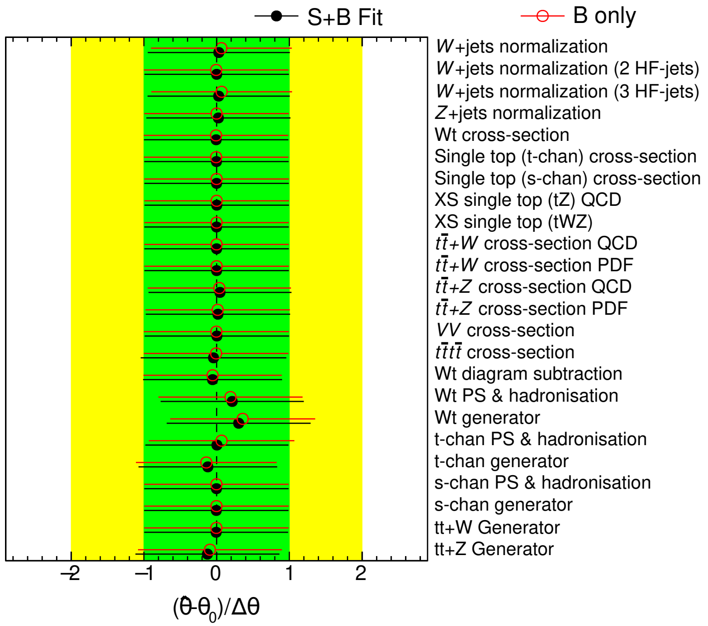
tWH Parametrisation l+jets

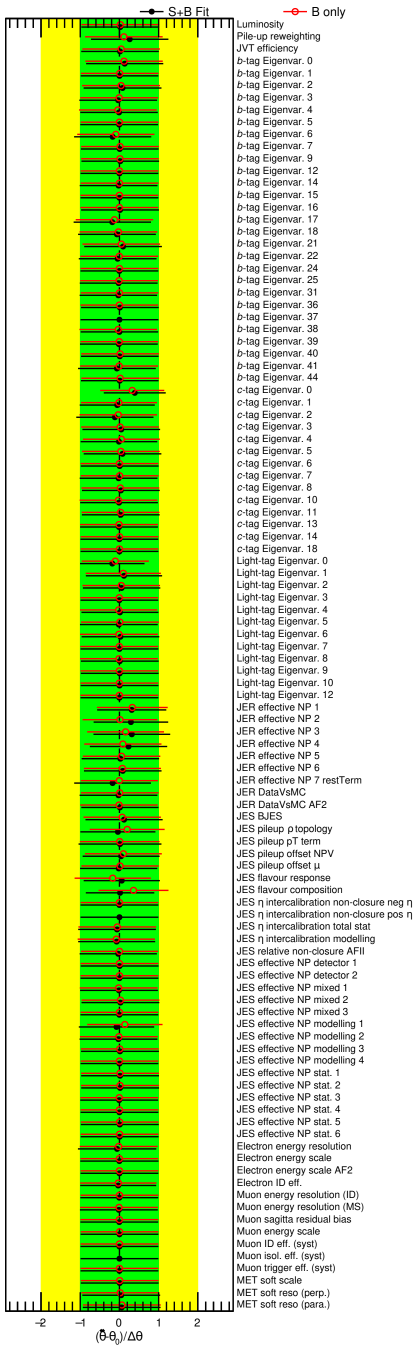


Nuisance parameter pulls

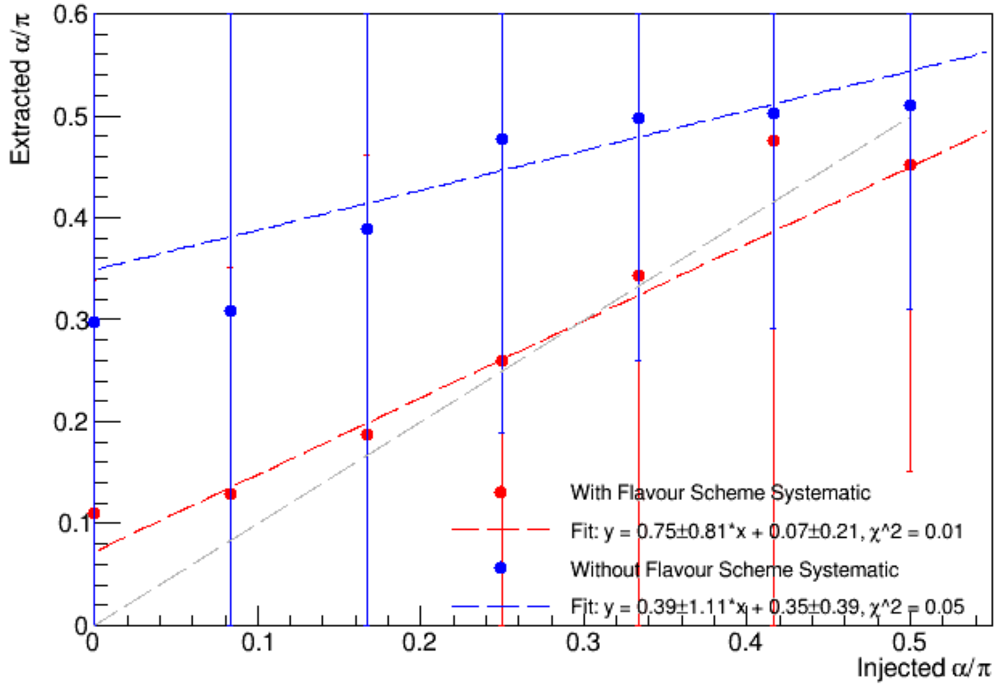








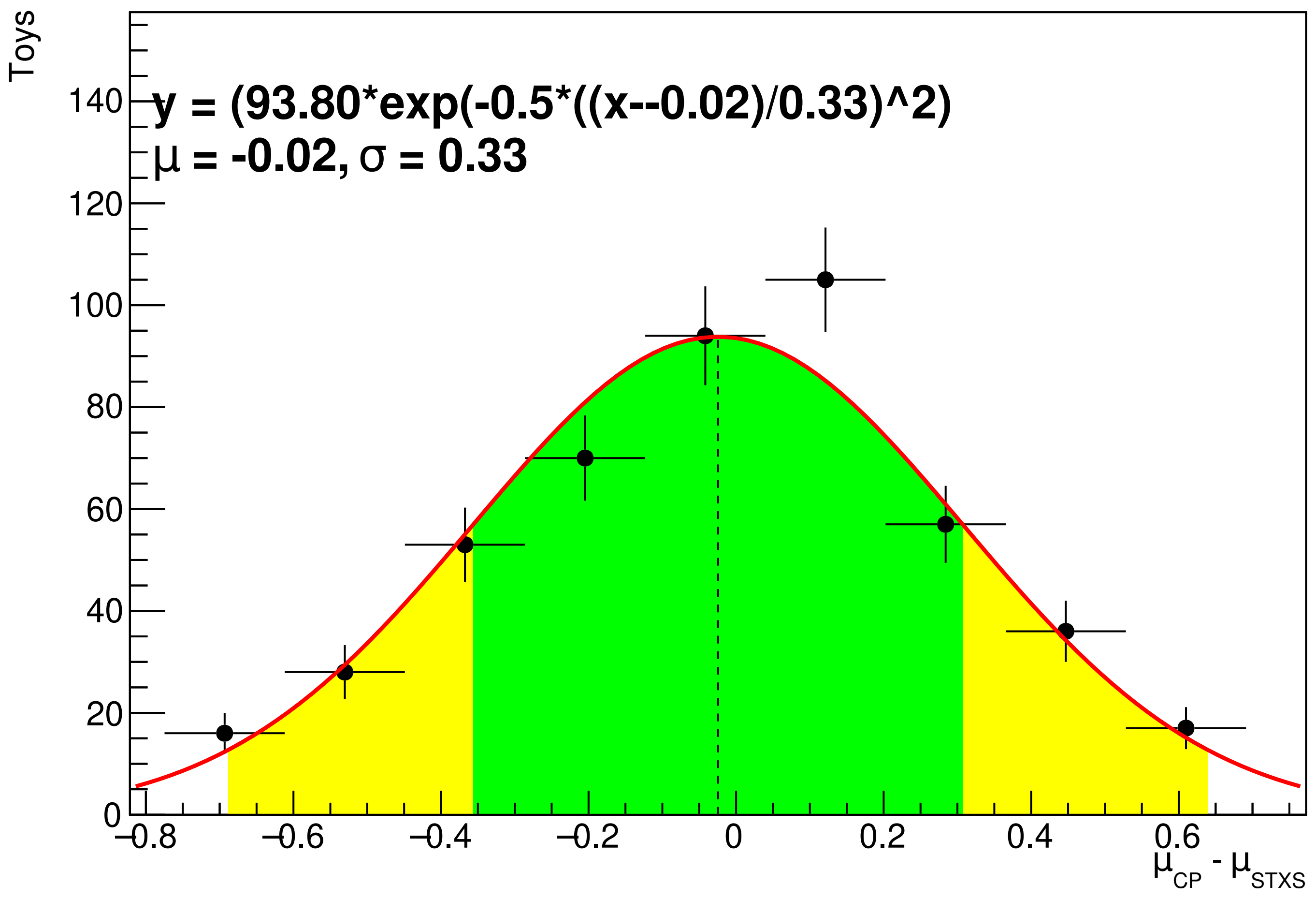
Flavour scheme uncertainty



- Fits to pseudo data with the $t\bar{t} + \geq 1b$ component of the background replaced by a 5 flavour scheme sample performed.
- Various value of the mixing angle α were injected to test is the model could recover this.
- Without the systematic a significant bias towards a CP-odd signal was obtained.
- Including the uncertainty was found to ameliorate this issue.

Compatibility of Results

- The compatibility of the results to those obtained from the previous ATLAS measurement was evaluated from a bootstrap study.
- Difference is compatible at 1σ.



CP Observables

- Distribution of CP observables when given a pair of random vectors uniformly distributed over a sphere.