



# Searches for CP symmetry violation in the top quark sector with CMS at the LHC, and metrology measurements for the tracker Endcap upgrade at the High Luminosity LHC

---

PhD day - 23/04/2026

**Enzo Fillaudeau**

Advisor : **Nicolas Chanon**

CMS group



---

## Outline :

1. Analysis of CMS data : search for CP violation using single top quark production in the t-channel.
2. Upgrade of the CMS tracker for the HL-LHC : metrological analysis of mechanical structures (dees).
3. Phenomenological study: reinterpretation of LHC data for the search for CPT violation.

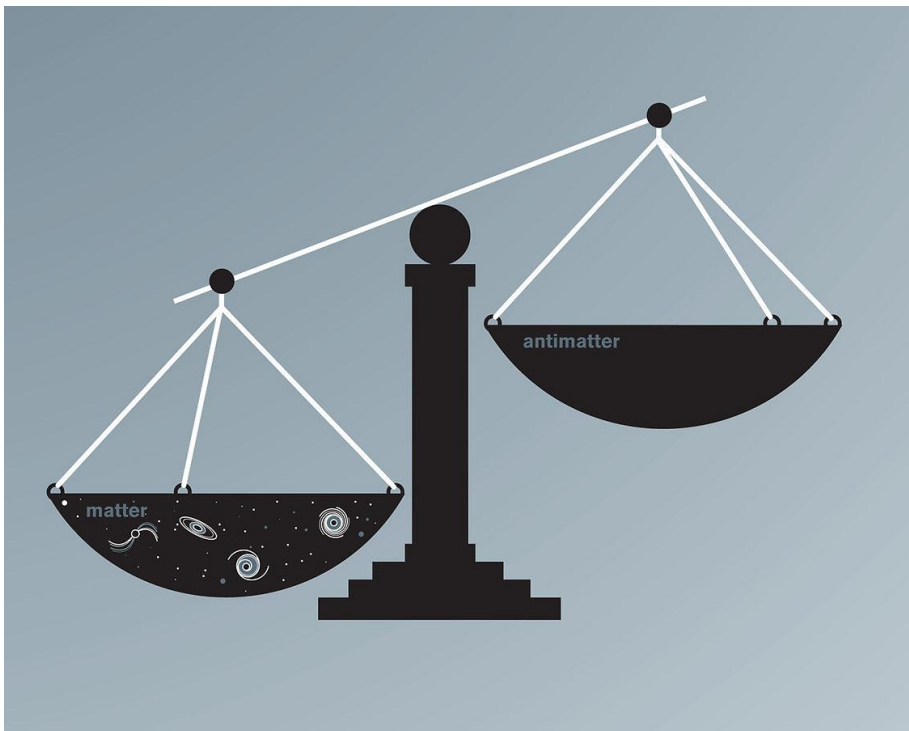


Outline :

- 1. Analysis of CMS data : search for CP violation using single top quark in the t-channel.**
2. Upgrade of the CMS tracker for the HL-LHC : metrological analysis of mechanical structures (dees).
3. Phenomenological study: reinterpretation of LHC data for the search for CPT violation.

# Motivation : Matter-antimatter asymmetry

Asymmetry of the baryon-number in the Universe

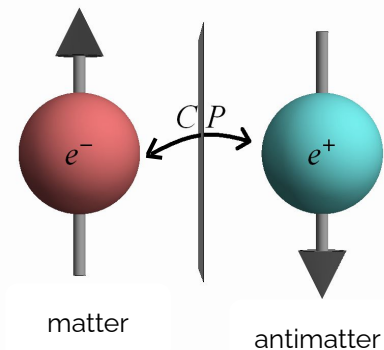


SM predictions :  $\eta_{SM} = \frac{n_B - n_{\bar{B}}}{n_\gamma} \propto 10^{-20}$

Observations :  $\eta_{obs} = \frac{n_B - n_{\bar{B}}}{n_\gamma} \propto 10^{-10}$

$$\Rightarrow \frac{\eta_{SM}}{\eta_{obs}} \propto 10^{-10}$$

**Discrepancy between the SM prediction and observations**



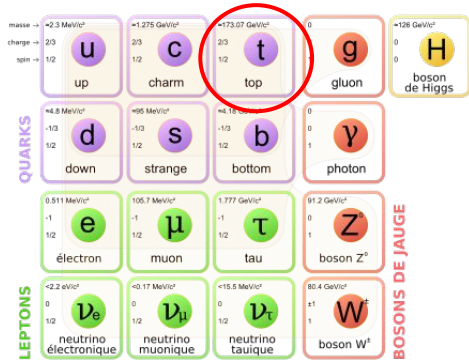
C : Charge conjugation  $Q \rightarrow -Q$

P : Parity transformation  $\vec{r} \rightarrow -\vec{r}$

**This asymmetry can be explained by the existence of sources of CP or CPT violation during baryogenesis.**

# Context : Top quark and CMS detector at the LHC

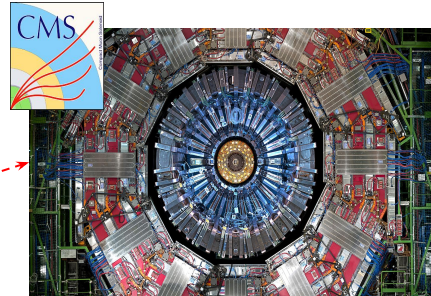
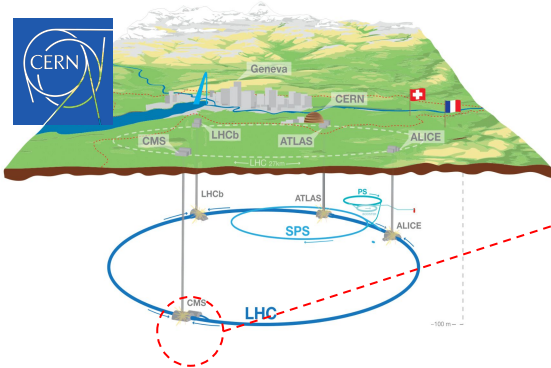
## The top quark as a probe for CP violation searches



### Top quark :

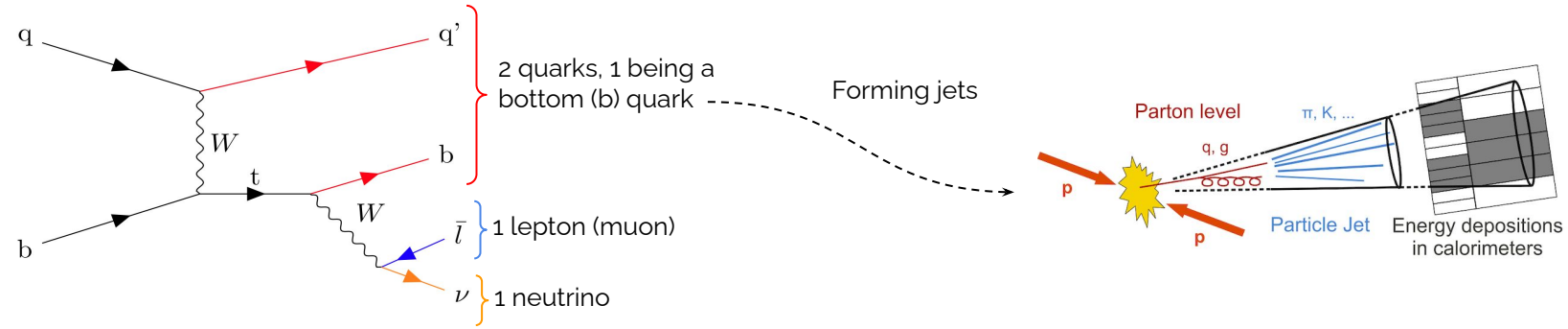
- Heaviest elementary particle
- Mass close to potential new physics energy scale
- Ideal candidate for search of Beyond the Standard Model physics (CP violation)

## Analysis carried out in the context of LHC Run 2 with the CMS detector

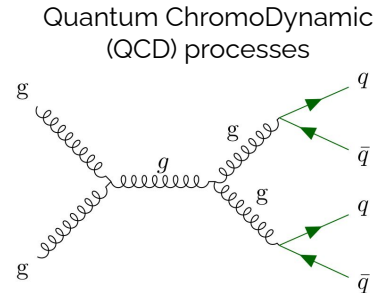
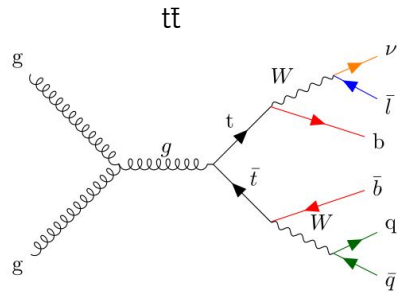
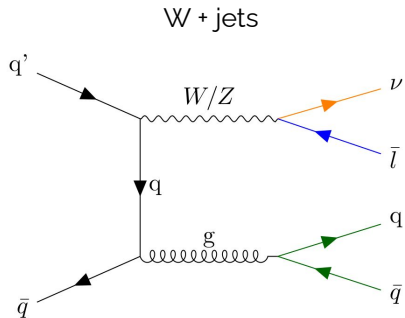


**Top quarks are produced in large numbers at the LHC : about 100 million generated during Run 2 (2015–2018).**

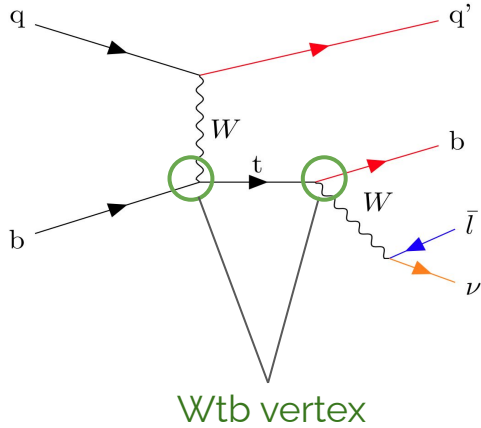
**Search for CP violation** through single top quark production in the t-channel and their subsequent decay into W and b.



**Background processes** can mimic signal in the case where jets or leptons are misidentified or not reconstructed :



# Goal of the analysis on the search of CP violation



Signal process dynamic

Analysis in the framework of *Effective Field Theory (EFT)*

$$\mathcal{L}_{eff}^{(6)} = \mathcal{L}_{SM} + \sum_i \frac{C_i^{(6)}}{\Lambda_i^2} O_i^{(6)} + h.c.$$

Non-Hermitian Operators involving two quarks and a boson

$$O_{uW}^{(6)} = (\bar{q}\sigma^{\mu\nu}\tau^I t)\tilde{\varphi}W_{\mu\nu}^I$$

$$O_{dW}^{(6)} = (\bar{q}\sigma^{\mu\nu}\tau^I b)\varphi W_{\mu\nu}^I$$

$$O_{\varphi ud}^{(6)} = (\tilde{\varphi}^\dagger iD_\mu\varphi)(\bar{u}\gamma^\mu d)$$

Wilson Coefficients

$$c_{tW} + ic_{tW}^I$$

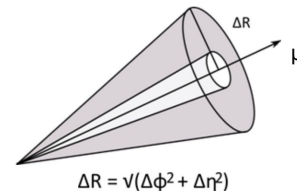
$$c_{bW} + ic_{bW}^I$$

$$c_{\varphi tb} + ic_{\varphi tb}^I$$

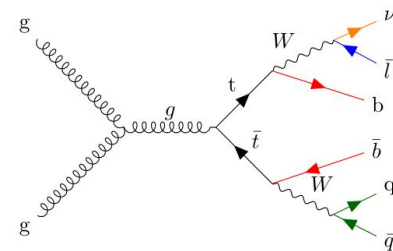
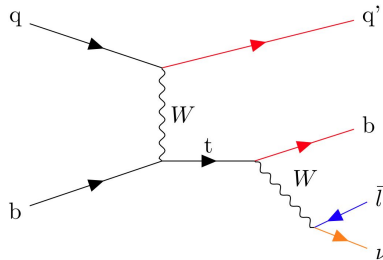
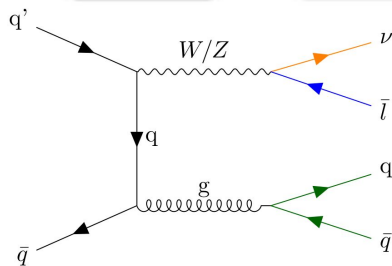
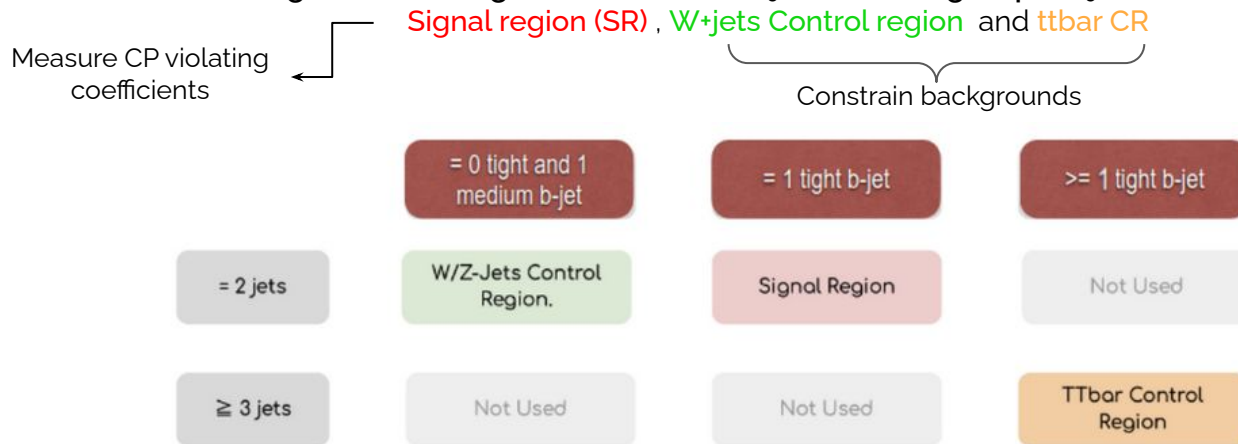
- EFT impacting both the production and decay of top quark
- This vertex can be modified by CP violation

The goal of this analysis is to measure the imaginary part of the 3 Wilson coefficients

- We require events with **exactly 1 isolated muon** (criterion on the energy deposit in a cone with opening  $\Delta R$  around the muon)



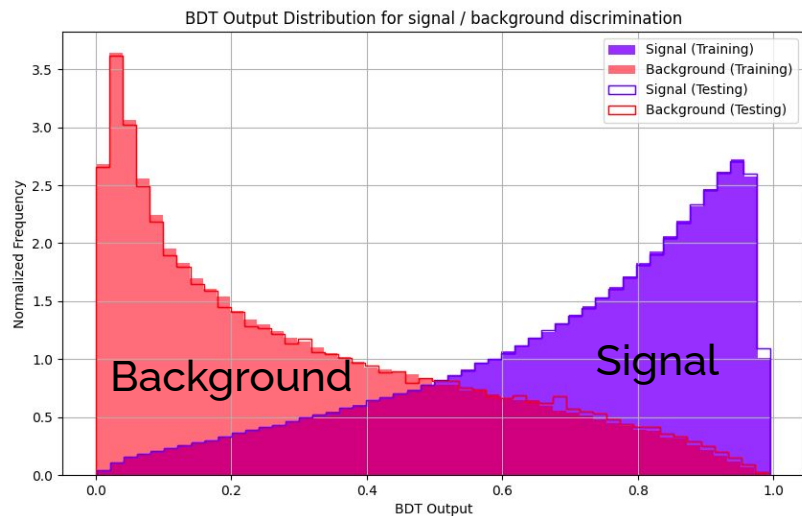
Events are divided into three regions according to their number of jets, including b-quark jets (identified using a b-tagging algorithm):



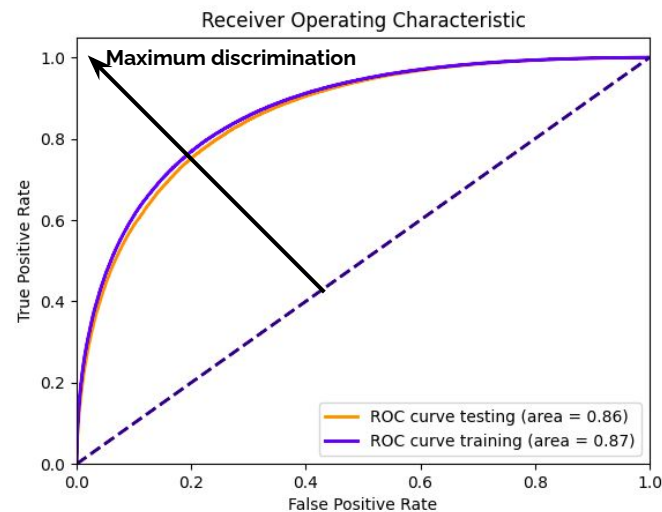
*The Signal Region has a large contribution from other background processes*

- The strategy is to optimize the signal extraction by using a **Boosted Decision Tree** (BDT) to discriminate **single top t-channel (signal)** from **W+jets and QCD processes**.
- The BDT model is trained with the simulated events in the SR using **distribution of discriminating variables**, e.g. the angular separation between the 2 jets

The output is a powerful discriminating distribution :



The performance of the model is assessed with a ROC curve :



Similar performance for the training and testing datasets indicates that the model is **not** overtrained

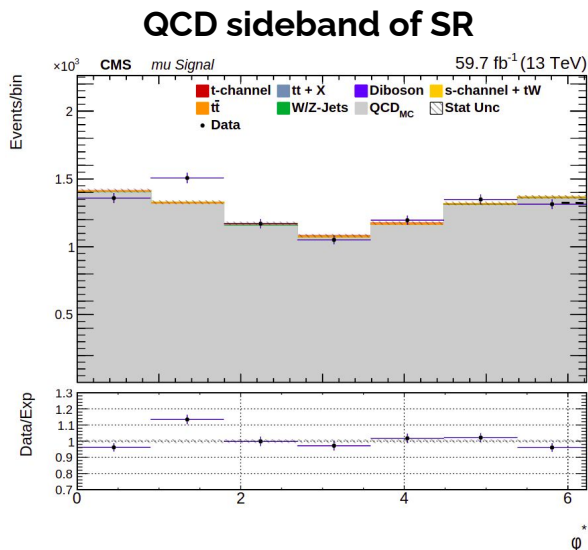
QCD processes have a large impact on this analysis but :

- ➔ insufficient simulated MC events available
- ➔ MC QCD is badly modeled

A **QCD enriched region** (sideband region) is defined for each signal and control region to estimate QCD from data by :

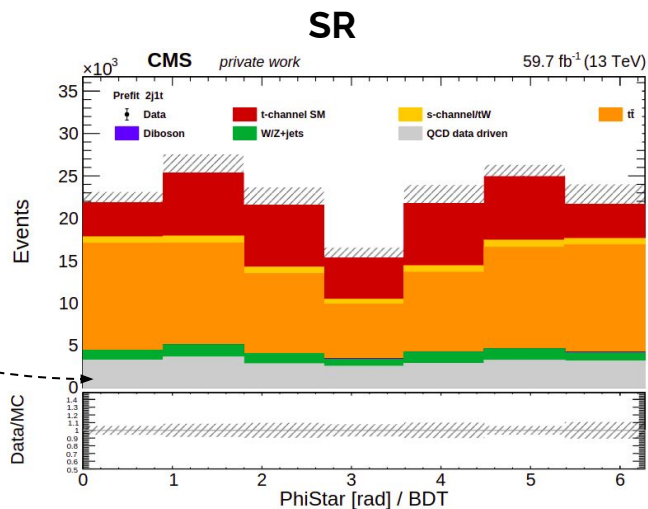
- Inverting the isolation criterion on the muon (a non-isolated muon, i.e. surrounded by significant hadronic activity, is typical of QCD processes)
- Removing the BDT score requirement in the sideband of the SR (to avoid depleting the QCD sideband and losing statistics)

*An example with  $\phi^*$  (variable discriminating SM and EFT) :*



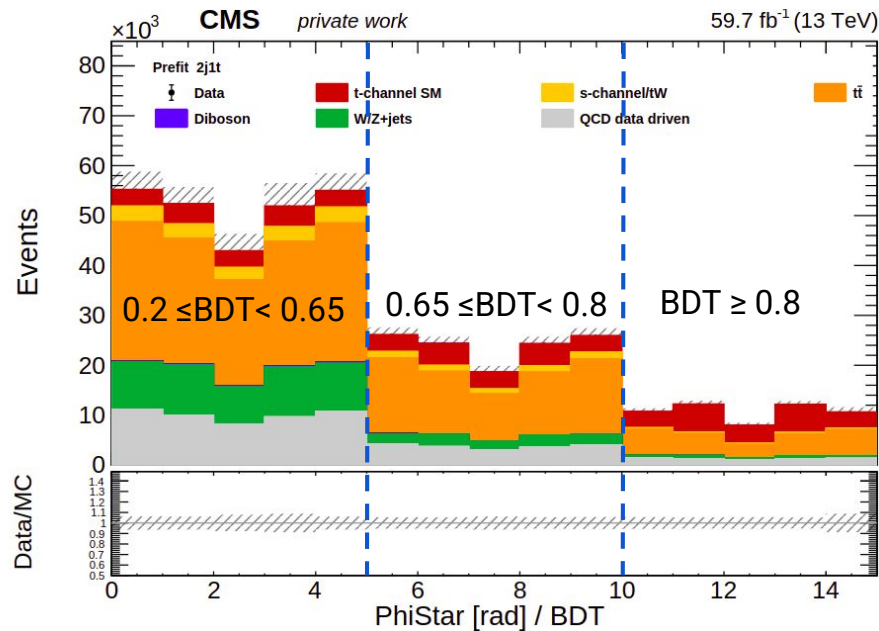
Shape of data in the QCD sideband is extracted ...

... to estimate the QCD in the SR

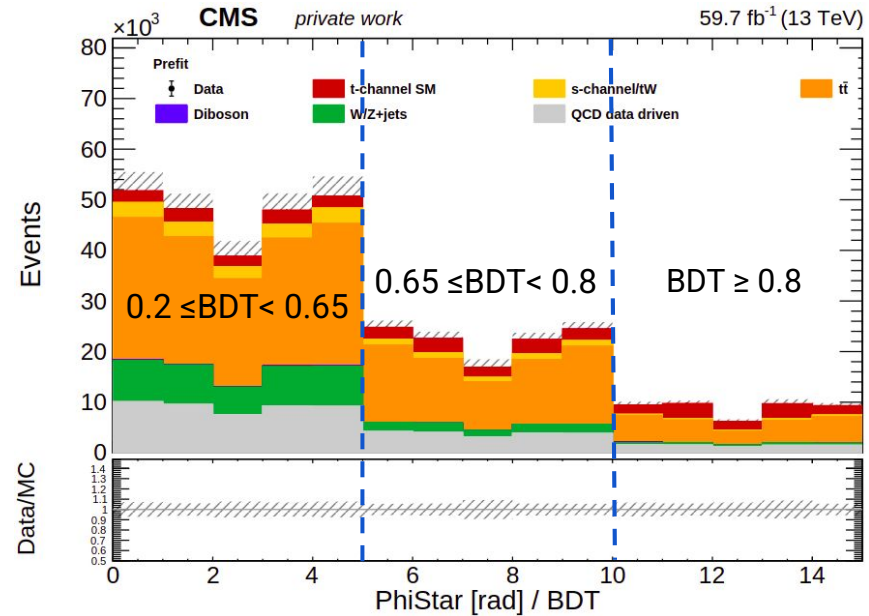


# Fit strategy - Signal region

- ❖ The SR is separated into a **Top** and **Antitop** channel, determined from the lepton charge
- ❖ The  $\varphi^*$  distribution in bins of the BDT is used to measure the CP violating EFT coefficients



**Top channel**

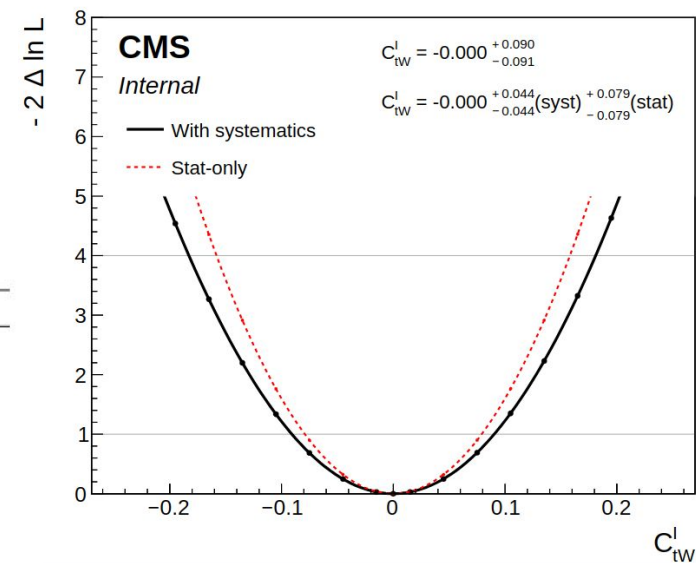


**Antitop channel**

- ❖ Fit performed using the **likelihood ratio method** with the BDT vs  $\phi^*$  distribution in the SR
- ❖ The SRs are fitted **simultaneously** with the CRs
- ❖ We obtain **expected sensitivity** of the analysis on the measurement of the coefficients

Source	$C_{tW}^I$	$C_{bW}^I$
Best fit	-0.00	+0.00
<b>Total unc.</b>	<b>+0.09 / - 0.09</b>	<b>+1.08 / - 1.13</b>
Stat. unc.	+0.08 / - 0.08	+0.44 / - 0.44
MC stat. unc.	+0.04 / - 0.04	+0.43 / - 0.48
Other syst. unc.	+0.01 / - 0.02	+0.88 / - 0.92

Uncertainties determined with a likelihood scan



- Sensitivity dominated by **statistics** for  $C_{tW}^I$  and **other systematics** for  $C_{bW}^I$
- Reduce the data statistical part by improving the signal/background discrimination
- Reduce the systematics on  $C_{bW}^I$  by changing the discriminant observable

---

## Outline :

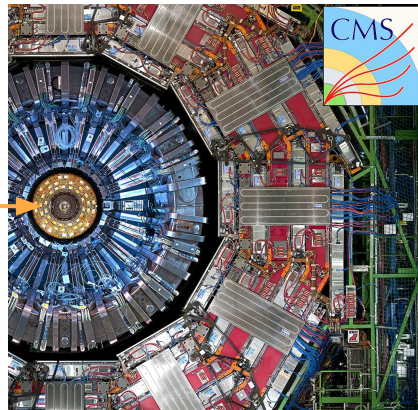
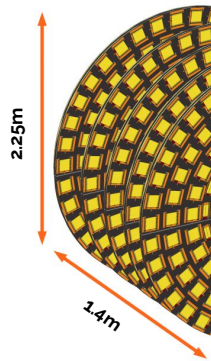
1. Analysis of CMS data : search for CP violation using single top quark in the t-channel.
2. **Upgrade of the CMS tracker for the HL-LHC : metrological analysis of mechanical structures (dees).**
3. Phenomenological study: reinterpretation of LHC data for the search for CPT violation.

# The new CMS tracker for HL-LHC

TEDD = Tracker Endcap Double-Discs, future CMS tracker for the HL-LHC

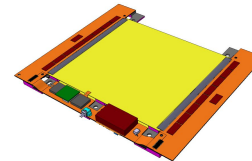
## The CMS tracker needs an upgrade to :

- sustain the high level of radiation
- trigger data acquisition on events with high momentum tracks

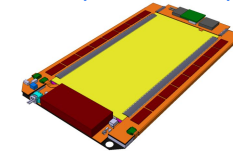


CMS detector at CERN

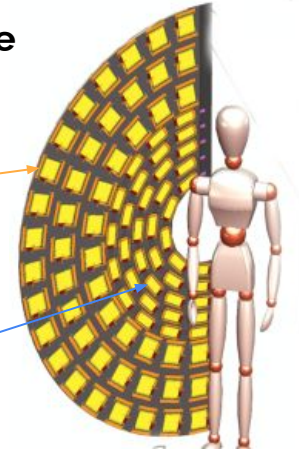
2S = strip - strip



PS = pixel - strip

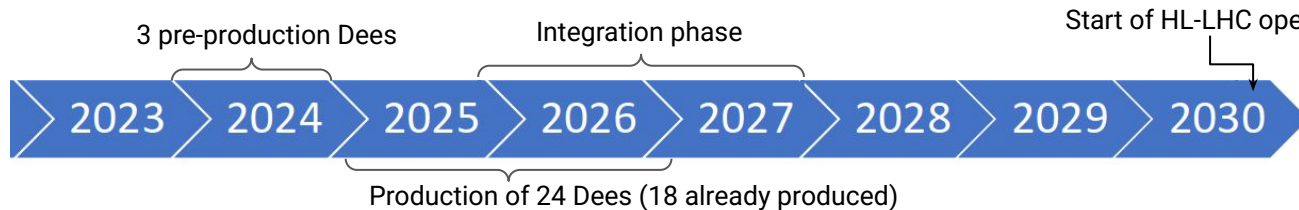


Dee



Main mechanical structure of the TEDD holding the detection modules

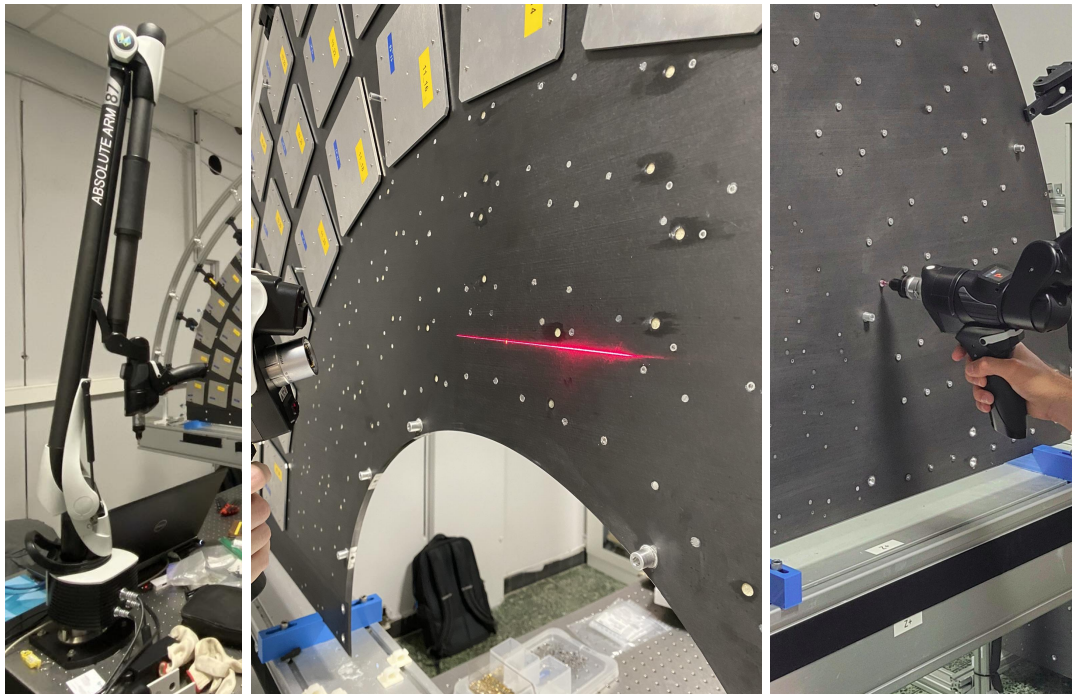
CMS group at IP2i is responsible of the **construction** and **qualification** of 24 Dees



Metrology is a step of the qualification of the Dees : the **flatness** of the structure and the **displacement of the inserts** are verified

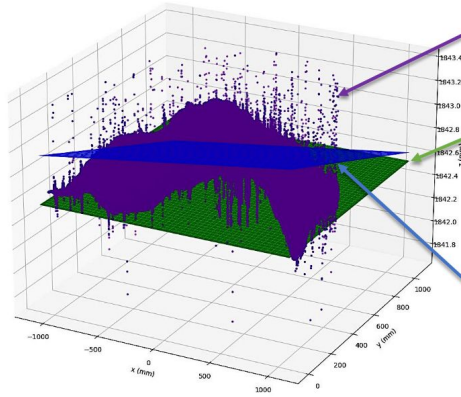
Metrology performed with a **Hexagon Arm** equipped with :

- a laser head AS1 (precision is 0.04 mm) to scan the Dee's flatness on its surface ( $N \sim 10^6$  points in 3D)
- a spherical ruby tip ( $\varnothing$  6 mm) for precise position measurement of the inserts ( $\pm 0.03$  mm)



# Dee flatness measurement

Excessive deformation of a Dee could prevent the proper assembly of a double-disk.

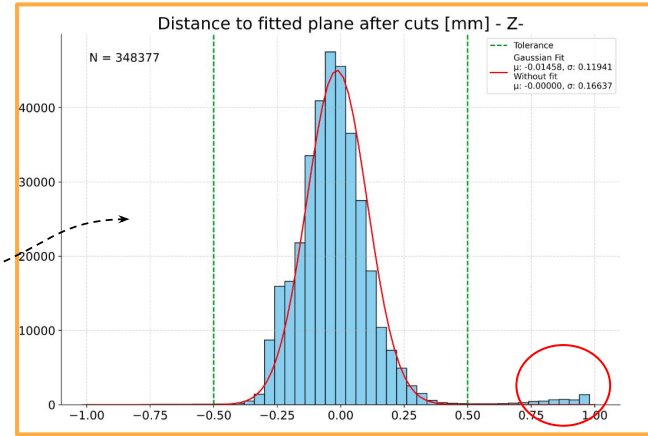


**High-density scanning:** used a laser head mounted on a hexagon arm to collect ~1 million points per scan.

**Nominal-plane method (default):** compute each point's distance to the design plane—but misalignments can bias results.

**Fitted-plane method (New):** fit the best-match plane through the measured data, then calculate point-to-fitted-plane distances.

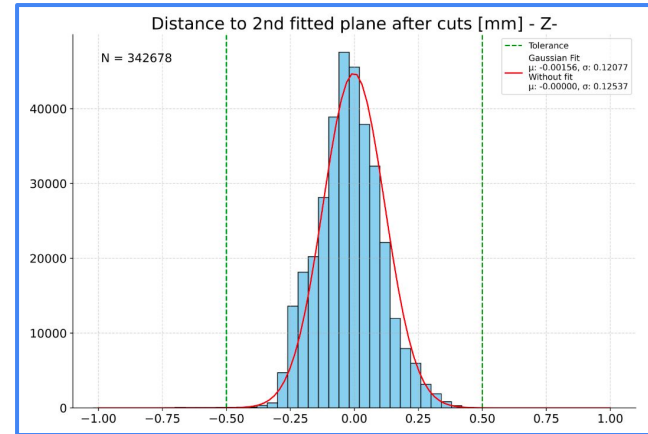
**Benefit:** removes global tilt, isolating true local deviations for an unbiased flatness metric.



Tail of the distribution outside of the specifications

→ Due to the detection of inserts on the Dee's surface but non relevant for the flatness measurement

**Cleaning of the data** is performed by applying a Z-cut relative to fitted plane for outlier filtering

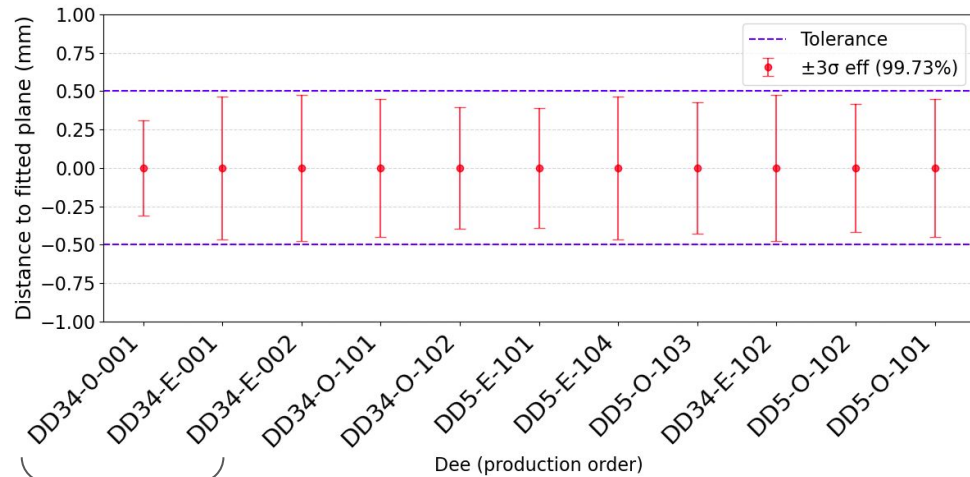


# Summary of the metrology measurement over production

Throughout the entire duration of the thesis : 24 Dees to analyze in total (16 observables for each of the 2 sides of a Dee).

- ➔ I worked on a way to present a large number of results in a compact form
- ➔ Automatic filling of files with Dee data during the execution of the analysis framework

*Example with the summary of Dee flatness*



Specifications : Nearest distance between measured points and fitted planes comprised between **±0.5 mm**

Effective sigma : smallest interval containing  $3\sigma$  of the population

- ➔ Stable flatness of the Dees throughout production
- ➔ All Dee's flatness are within specifications

pre-production Dees

## Outline :

1. Analysis of CMS data : search for CP violation using single top quark in the t-channel.
2. Upgrade of the CMS tracker for the HL-LHC : metrological analysis of mechanical structures (dees).
3. **Phenomenological study: reinterpretation of LHC data for the search for CPT violation.**

# Motivation and context of the phenomenological study

Use of a **CPT-violating** model within the **EFT formalism** : the **Standard Model Extension** (SME) :

- Motivated by frameworks such as string theory or loop quantum gravity
- Inclusion of all CPT-violating operators in the Standard Model Lagrangian

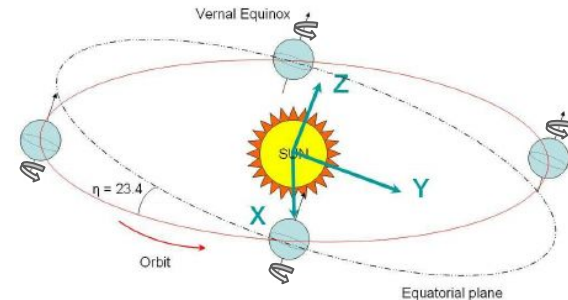
Part of the SME Lagrangian governing odd CPT symmetry :  $\mathcal{L}^{CPT-} \supset (1 - \gamma^5) \underbrace{b_\mu \bar{t} \gamma^\mu t}_{\text{CPT violating coefficient}} + \dots$

CPT violating coefficient  $\equiv$  **constant vector** in an inertial reference frame

Reference frame used to present the results : Sun-centered frame (**SCF**)

Measurement frame : Laboratory frame (**CMS detector**)

- The **rotation of CMS around the Earth's axis** induces a **modulation of the top quark production cross section** as a function of **sidereal time** ( $T \sim 23\text{h } 56\text{min}$ )

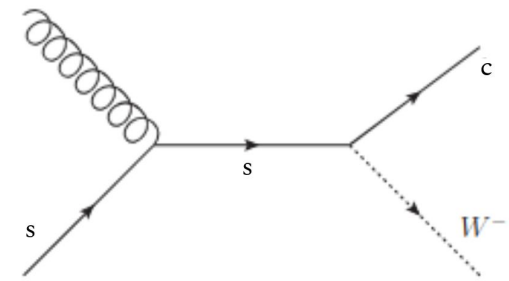
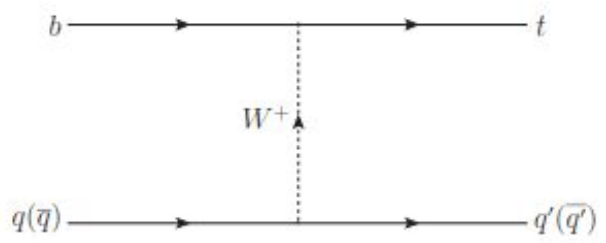


Temporal modulation function :  $f_{SME}(t) = b_\mu \underbrace{R^\mu}_v \underbrace{\frac{\delta P^\mu}{P}}_{\text{CPT-violating corrections}}$

*Time-dependent rotation matrix relating the SCF and the CMS reference frame*

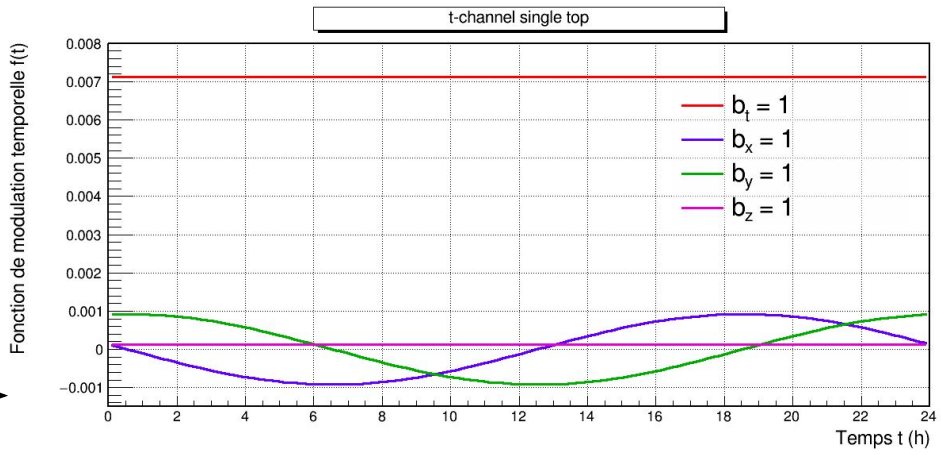
*Ratio of the process cross section including CPT-violating corrections to the Standard Model case*

We aim to evaluate the sensitivity that could be achieved on the measurement of  $b_\mu$  at the LHC using two processes : **single top in the t-channel** and **W+c**.



- Selections are applied at parton level on generated events for both processes
- Ratio of the process cross section is computed for each event
- Rotation matrix linking CMS and SCF is calculated
- Temporal modulation is computed

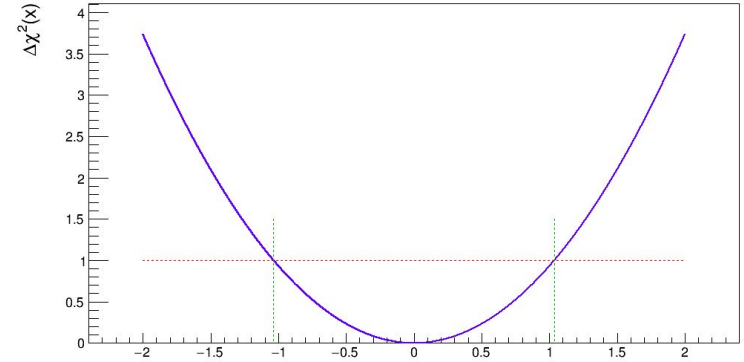
$$f_{SME}(t) = b_\mu R^\mu \nu \frac{\delta P^\mu}{P}$$



# Expected sensitivity on CPT violating coefficients measurement

- Simplified estimate of the statistical and systematic uncertainty on the expected sensitivity to the CPT-violation coefficient with LHC Run 2, using the  $\chi^2$  method

Scan du Chi2 pour  $b_t$  avec top solitaire



Top related coefficient

Charm related coefficient

	My study employing <b>single top</b> process 95% C.L. <b>expected</b>	[arXiv:2405.12162] employing top-antitop mass difference in $t\bar{t}$ event 95% C.L. <b>observed</b>
$b_T$ (GeV)	[-2.72, 2.72]	[-0.13, 0.29]
$b_X$ (GeV)	[-22.52, 22.52]	[-0.8, 0.8]
$b_Y$ (GeV)	[-22.52, 22.52]	[-0.8, 0.8]
$b_Z$ (GeV)	[-116.72, 116.72]	[-4.6, 4.6]

My study employing <b>W+c</b> process 95% C.L. <b>expected</b>
[-0.028, 0.028]
[-13.98, 13.98]
[-13.98, 13.98]
[-31.62, 31.62]

- **CMS analysis : CP violation with single top t-channel**

- Use a machine learning technique for lepton identification
- Add the measurement of the third coefficient :  $C_{\varphi tb}^I$
- Start the pre-approval process of the analysis

- **Tracker Upgrade**

- Dee production well ongoing, still 15 Dees to analyze

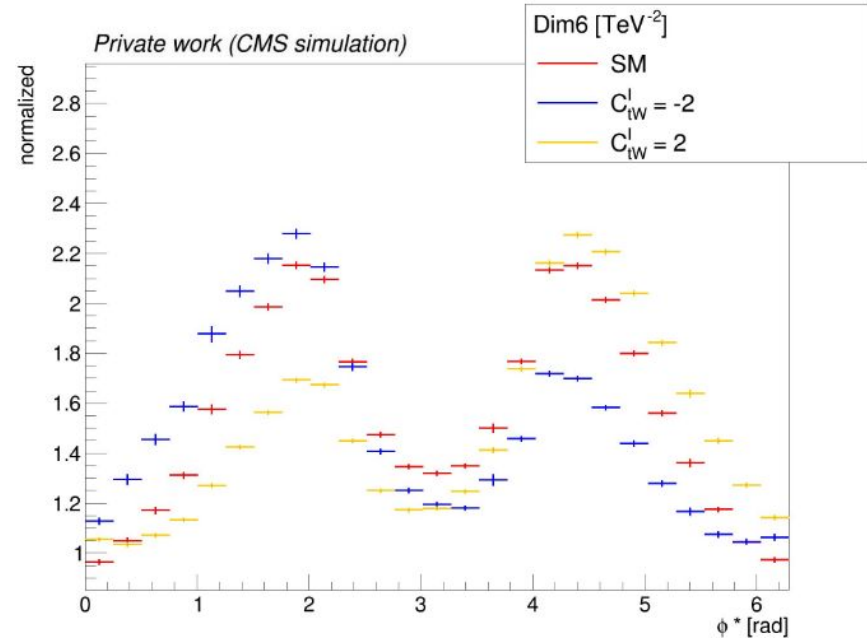
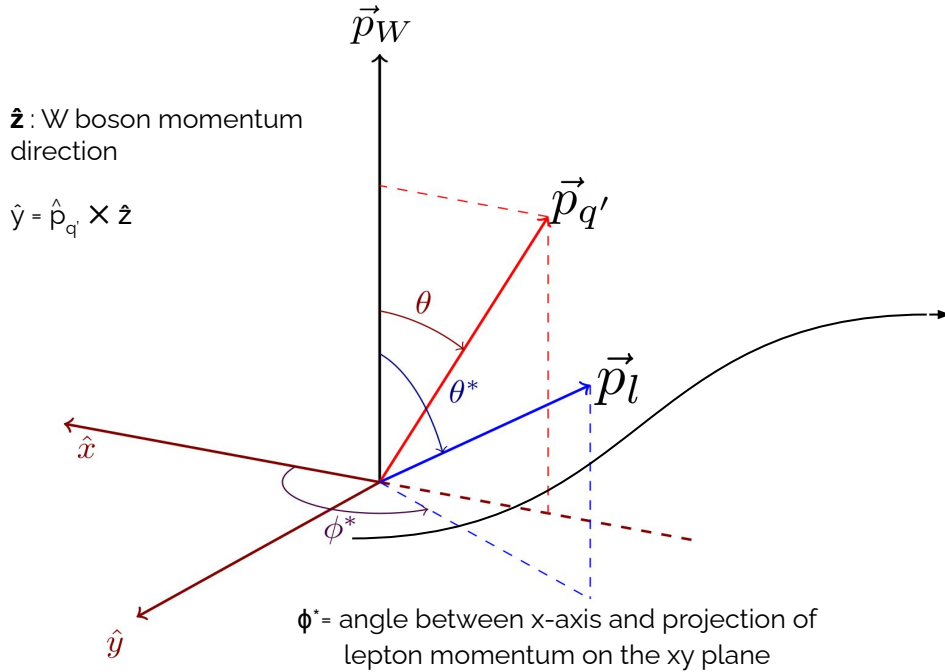
- **CPT phenomenology study**

- Simulated how quarks and gluons hadronize (turn into detectable particles) using a MC generator for my samples of simulated events
- Working on estimating sensitivity with event selection applied on hadronized events

**Thank you for your attention :)**

## Top quark rest frame

Reference frame used in ATLAS 8 TeV [arXiv:1707.05393]



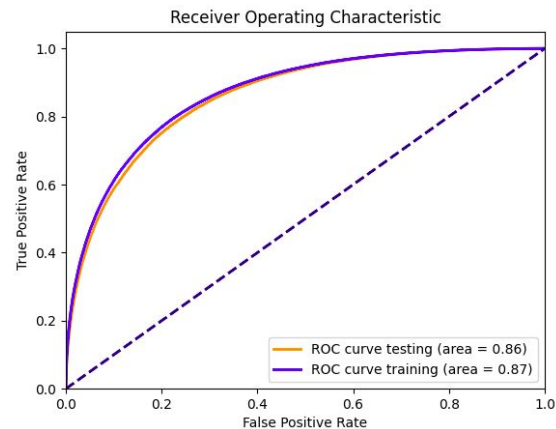
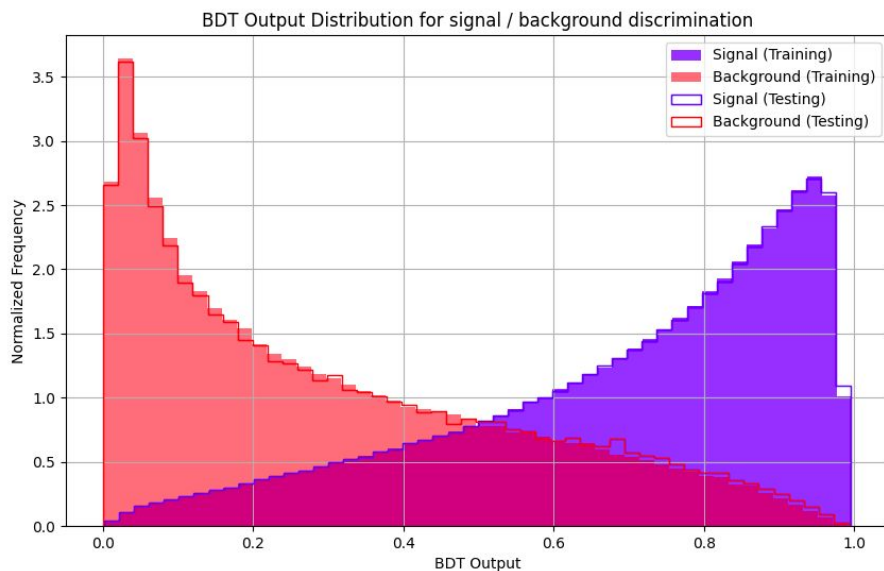
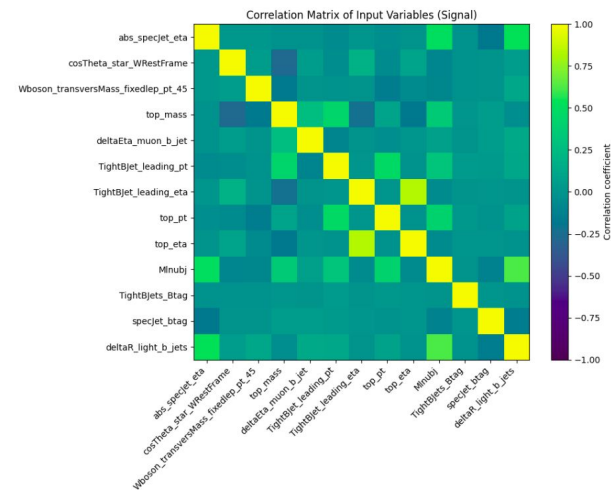
The shape of the distribution varies depending on the value of the EFT coefficient

The value of EFT coefficients can be extracted using such angular distributions

# Analysis strategy - MVA in the **Signal Region**

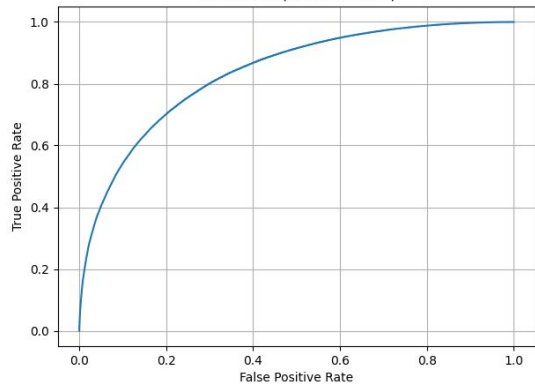
Trained a **new BDT** in the SR to discriminate **single top (ST) t-channel** from **W+jets and QCD processes** (previous BDT was signal vs. all backgrounds)

**Input variables for the BDT** :  $\cos(\theta^*)$ ,  $\eta$  of the spectator jet, spec jet btagging score,  $M_T W^{fix}$ , top  $p_T$ , top  $\eta$ ,  $M_{top}$ ,  $\Delta\eta(\mu, \text{b-jet})$ ,  $\Delta R(\text{spec jet}, \text{b-jet})$ , b-jet  $p_T$ , b-jet  $\eta$ , b-jet btagging score, lepton  $p_T$ ,  $M_{lvbj}$

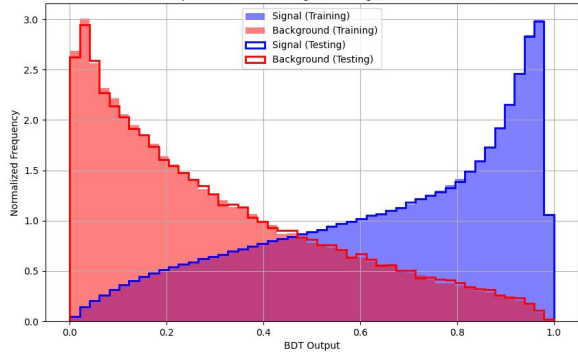


## W+jets

ROC Curve (AUC = 0.8354)

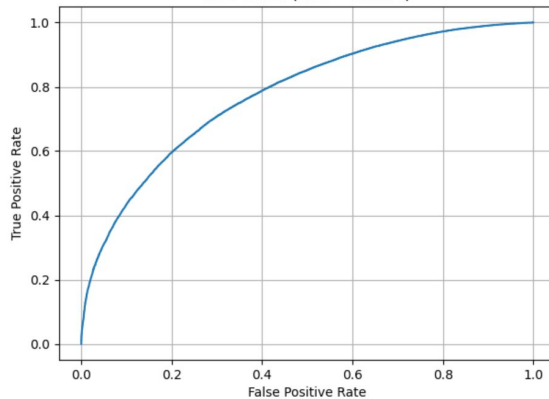


BDT Output Distribution for signal / background discrimination

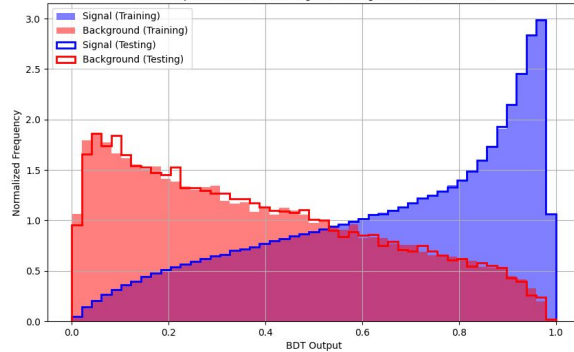


## QCD

ROC Curve (AUC = 0.7774)

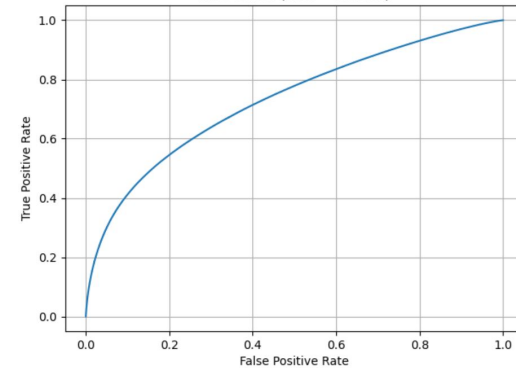


BDT Output Distribution for signal / background discrimination

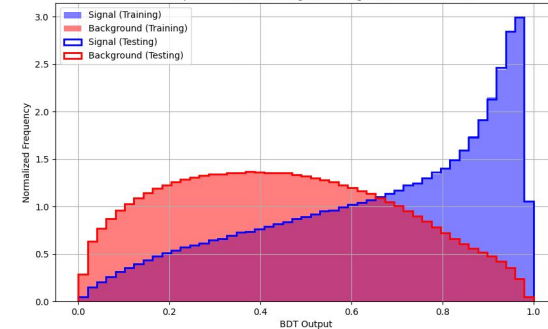


## t $\bar{t}$

ROC Curve (AUC = 0.7292)



BDT Output Distribution for signal / background discrimination

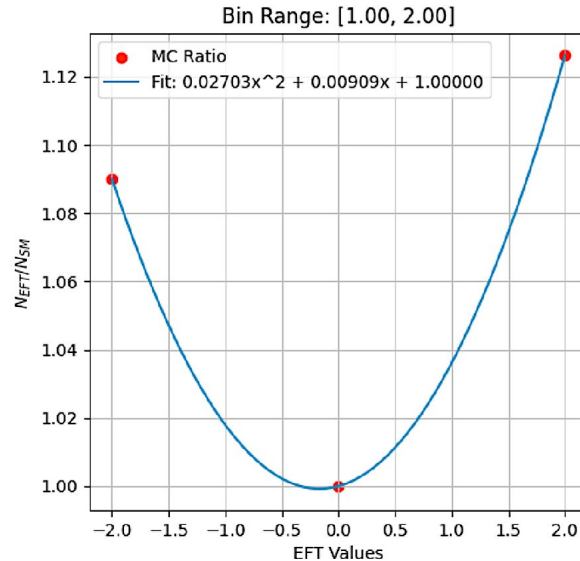
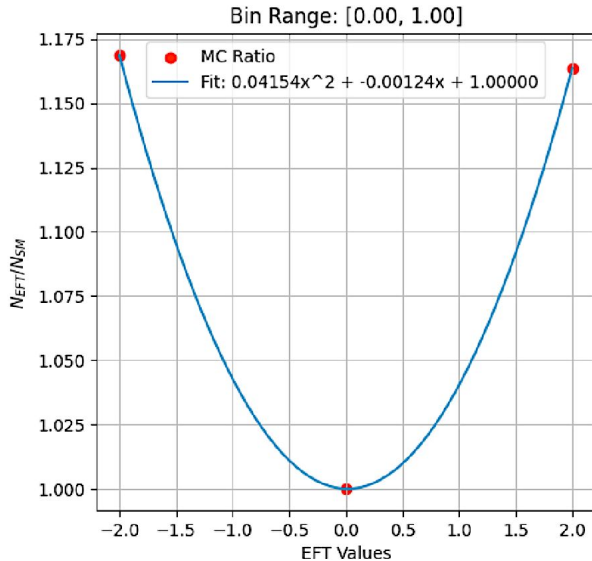


Evaluating the expected statistical sensitivity by fitting the SM+EFT model to the Asimov (SM) dataset using a profile likelihood statistical method (CMS Combine)

Using AnalyticAnomalousCoupling plugging to model the EFT template :

<https://github.com/amassiro/AnalyticAnomalousCoupling>

This framework provides a Combine-based model for EFT fits overcoming the issues of negative templates that may arise from interference terms



- Fitting polynomial of order 2 in each bin of our observable
- Using the three parameters of the fit to produce the templates

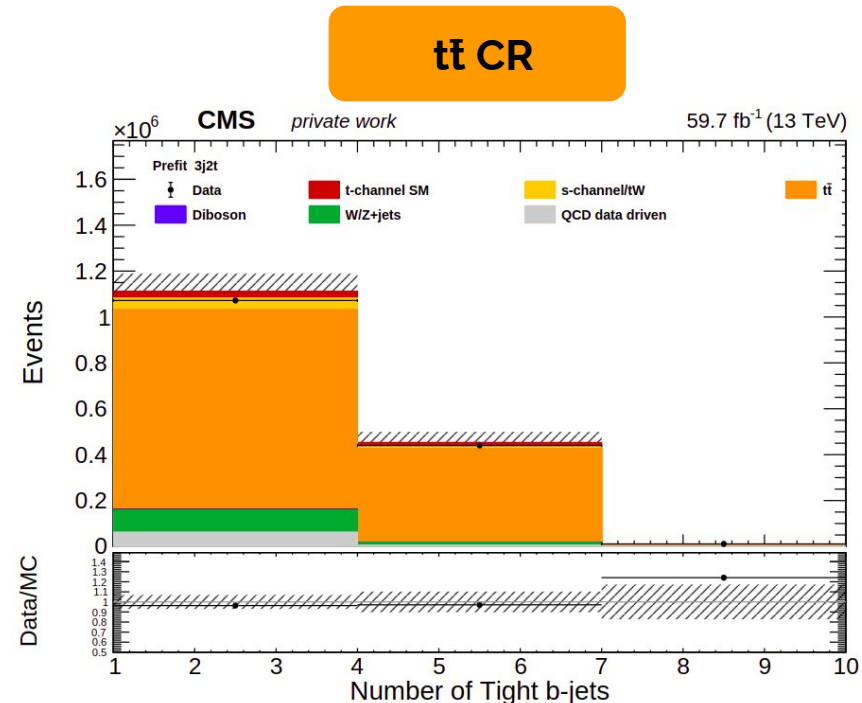
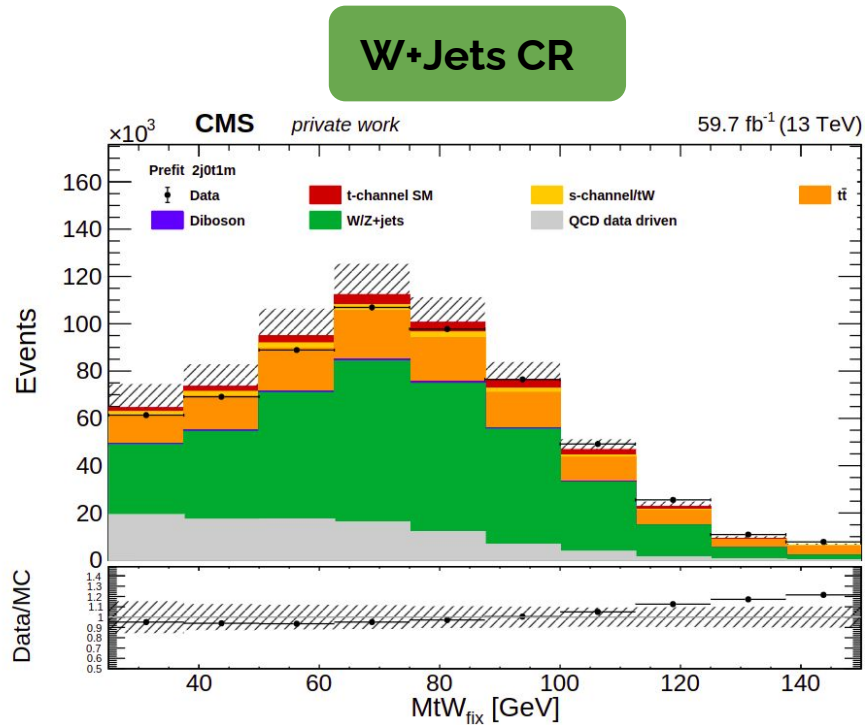
$$\text{Sm} = |\mathcal{A}_{SM}|^2,$$

$$\text{Lin}_\alpha = 2\Re(\mathcal{A}_{SM}\mathcal{A}_{Q_\alpha}^\dagger)$$

$$\text{Quad}_\alpha = |\mathcal{A}_{Q_\alpha}|^2$$

# Fit strategy - Control regions

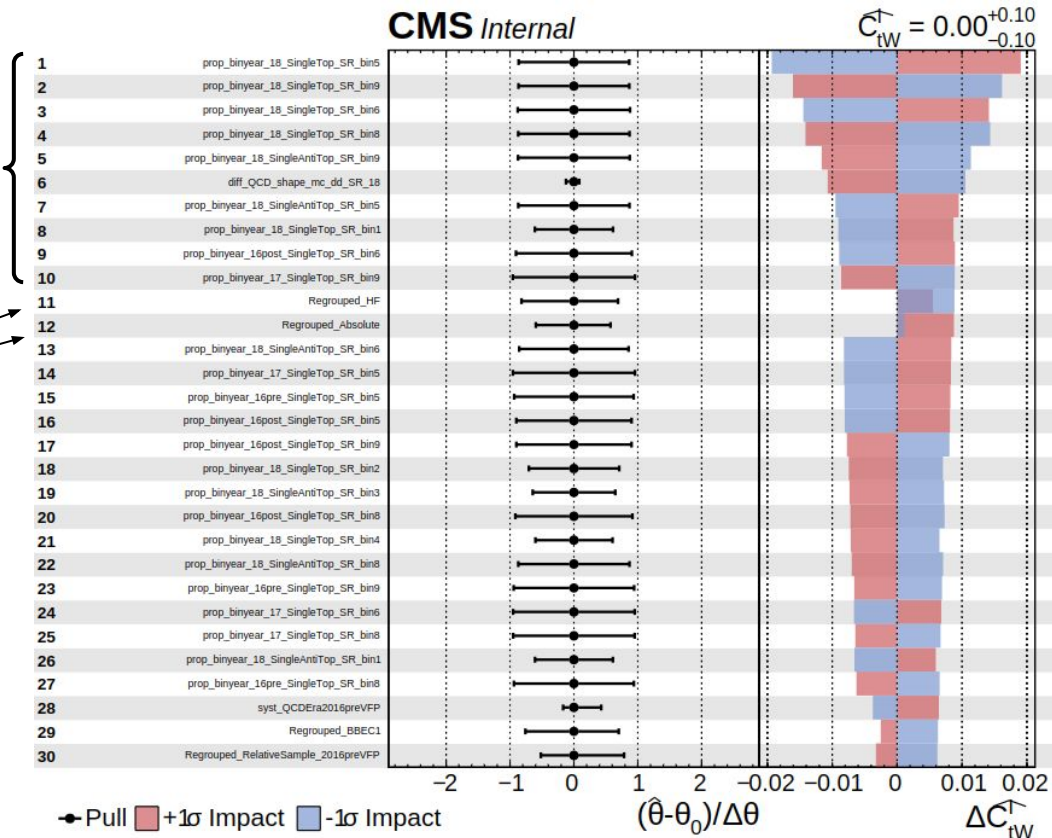
Variables in the CRs are chosen because they **separate the main background process** in the control region **from other processes**, allowing its **contribution to be accurately estimated**.



These distributions will be used in the final fit to **reduce the uncertainty on the main background normalization**

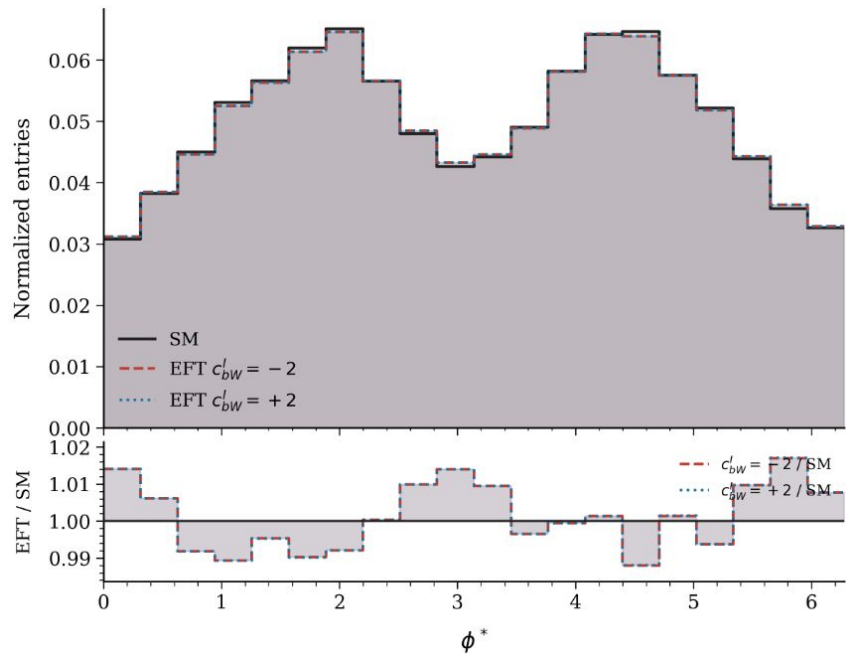
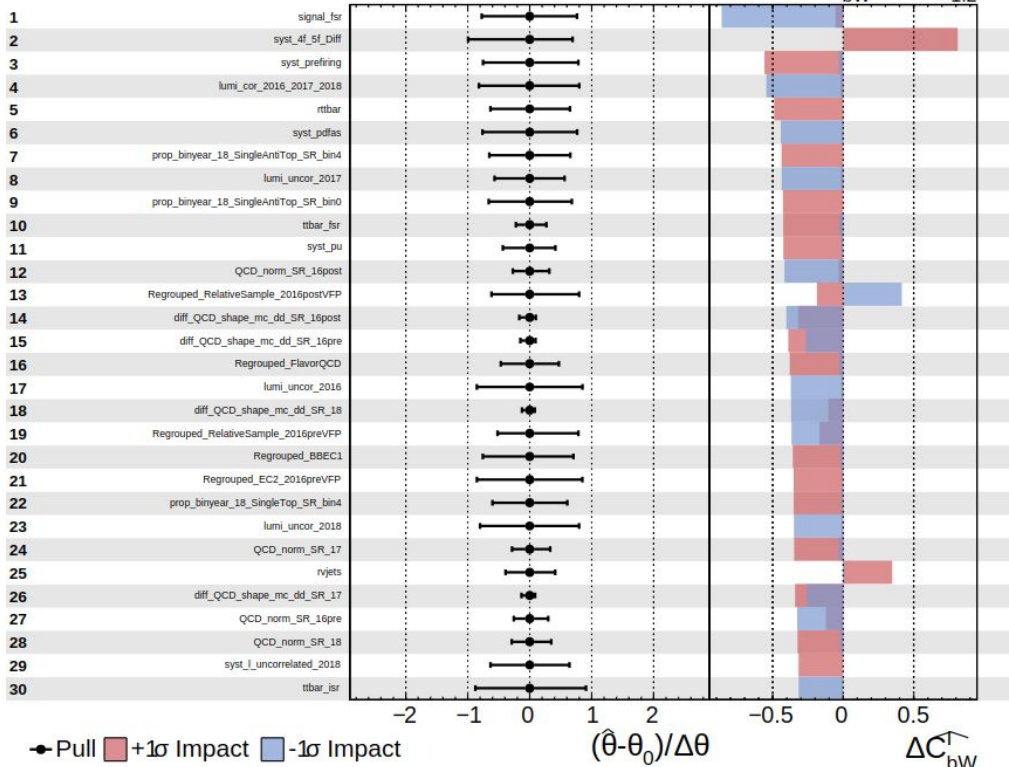
MC statistical uncertainties remain among the highest-ranked nuisance parameters, but their overall impact is significantly reduced compared to the previous strategy

Need to move to full set of JEC uncertainties (reduced one is one-sided)



CMS Internal

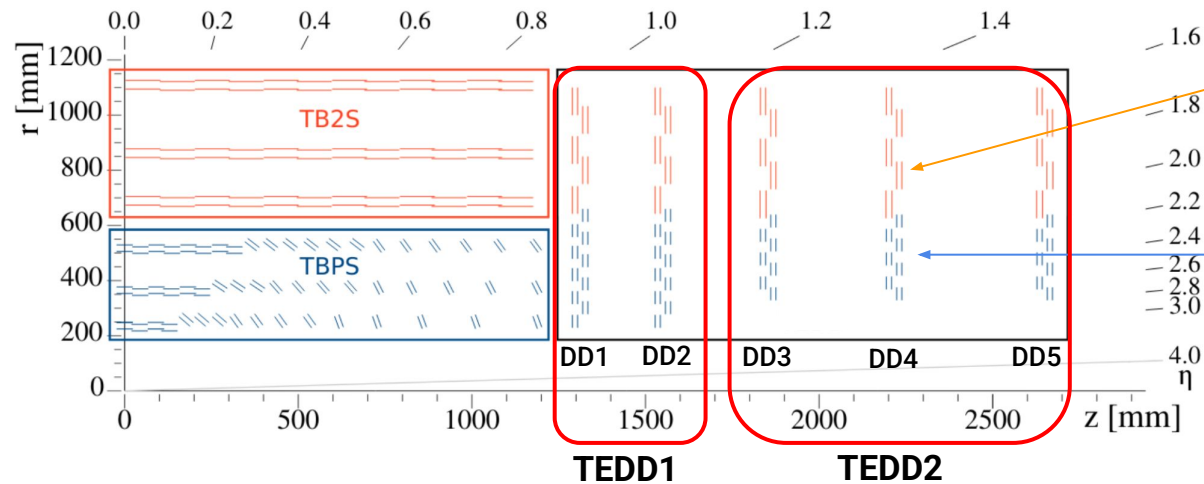
$C_{bW}^I = 0.0^{+1.1}_{-1.2}$



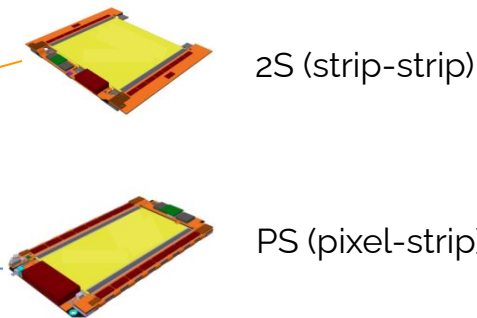
Many nuisances are one sided because  $C_{bW}^I = +2$  and  $C_{bW}^I = -2$  effect goes in the same direction.

# Tracker Endcap Double-Disks (TEDD)

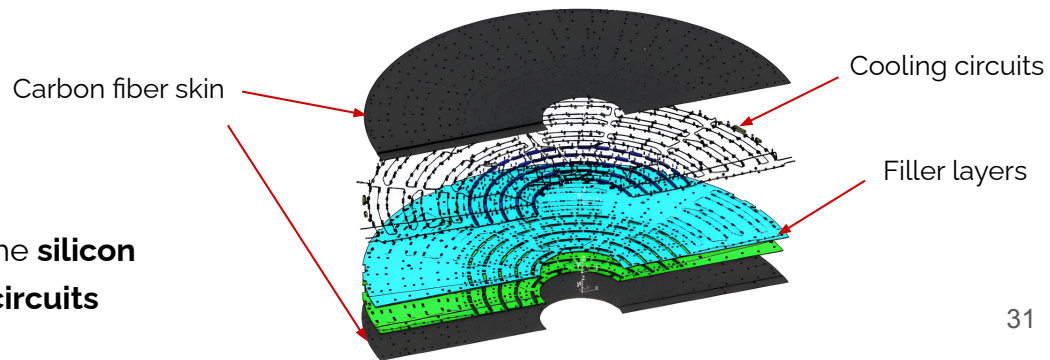
Endcap part of the future outer tracker of the CMS



Detection modules :

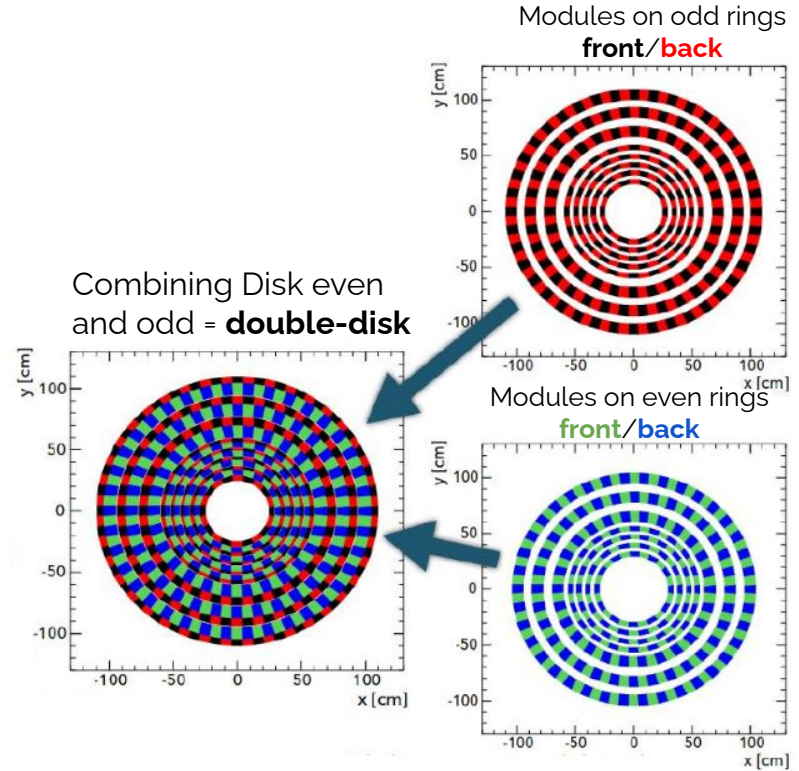


- Each TEDD is composed of 5 double-disks
  - TEDD1 : DD1 & DD2  $\Rightarrow$  DESY
  - **TEDD2 : DD3, DD4 & DD5  $\Rightarrow$  IP2i**
- A disk is made of 2 Dees
- **Dee** : elementary **mechanical structure** holding the **silicon detection modules** and embedding the **cooling circuits**



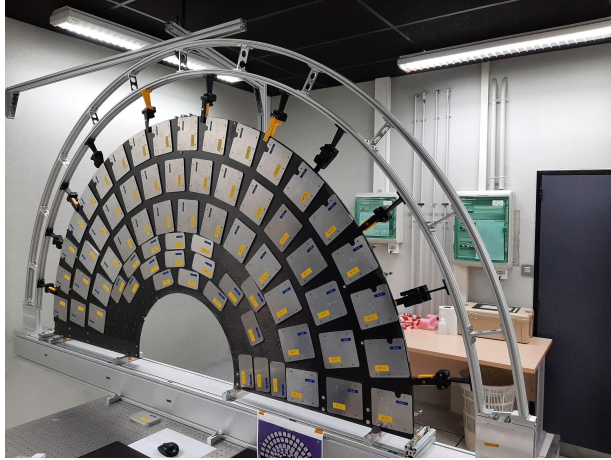
Modules will be arranged in rings on both sides of the Disks :

- ❖ each Disk provides **complete  $\phi$  coverage**
- ❖ combining the odd and even Disks ensures **full radial coverage**

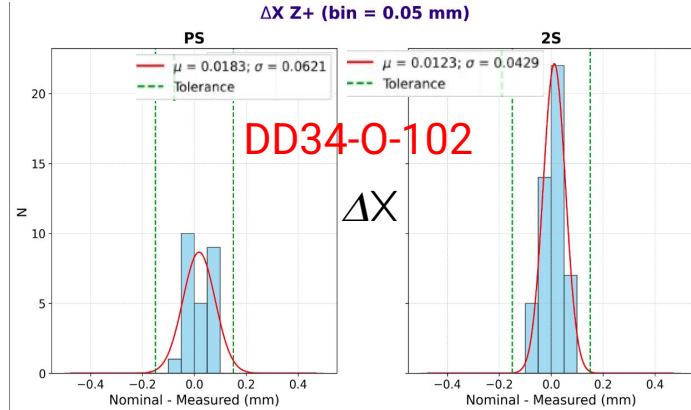
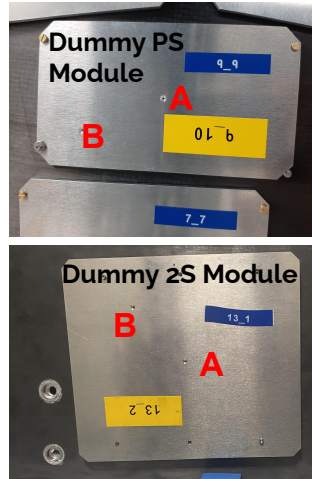


# Displacement of PS and 2S modules

The X and Y displacement of the 2S and PS modules is verified in metrology to ensure the modules can be properly integrated



Conic shaped reference points on the PS and 2S modules



**Measurements:** Dummy modules with precision holes are mounted onto the Dee. The XY coordinates of conic shaped A and B points are measured.

- $\Delta X$  = Displacement on the X axis
  - $\Delta Y$  = Displacement on the Y axis
  - $\Delta \theta$  = Rotation of the module
- } Determined using difference between nominal and measured position of A point
- } Determined using angular difference between nominal and measured AB line

# Data cleaning DD35-O-102

