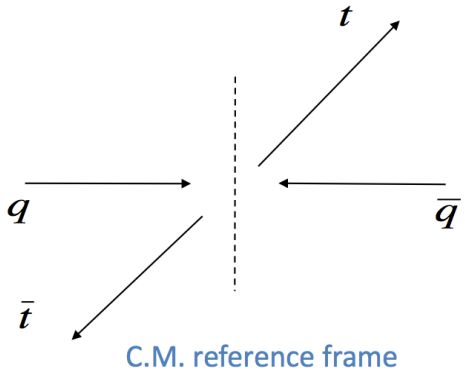
Top Quark Charge Asymmetry with full Run II at ATLAS using Fully Bayesian Unfolding

Clement Helsens
CERN, EP
On behalf of the ATLAS top A_C analysis team
LHCTop WG November 2019

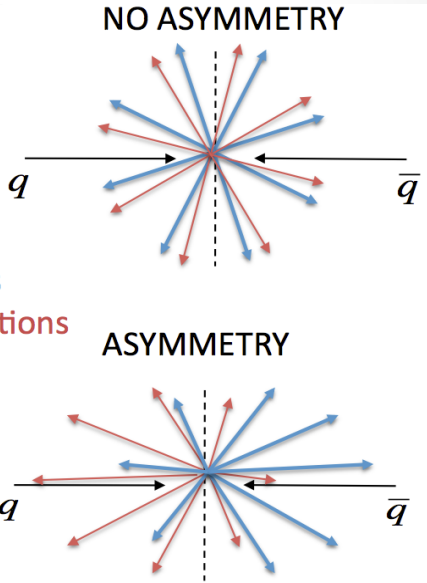
Top quark: pair asymmetry

What is the top quark pair charge asymmetry?

-> Measure anisotropy in (anti-)top production wrt incoming (anti-)quark direction



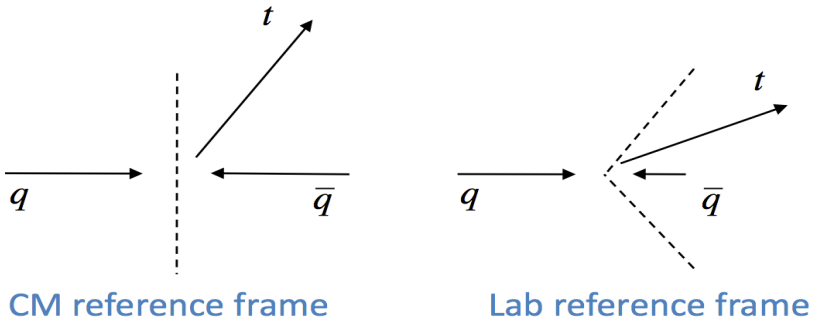
C.M. reference frame



- Top preferentially produced in the incoming quark direction
- Anti-top preferentially produced in the incoming anti-quark direction
- Requires asymmetric initial state: quark anti-quark annihilation

Asymmetry at the LHC

- More diluted: qqbar → ttbar only ~15/10% @ 8/13 TeV
- (anti)-top is preferentially produced along the direction of the incoming (anti)-quark



CM reference frame Lab reference frame

- Direction of the incoming quark not known in pp collisions

BUT

- Quarks more likely to be valence quarks (→larger fraction of momentum)
- Only sea anti-quarks available
- Top quarks more forward than anti-tops

LHC experimental challenge

- Small fraction of quark-antiquark annihilation
- Direction of the incoming quark unknown in pp collisions

Small A_C <1%

- Measured A_C distorted by resolution and acceptance.
  Need to extrapolate to parton-level asymmetry to compare with theory.

- Large systematic effects from parton-level extrapolation and from background modeling.

Measurements

- Perform a complete set of measurements of the charge asymmetry:

1. Inclusive measurement of A_C

2. Differential measurements:
   the dependence of A_C versus the kinematic of the top quark pair might provide useful information for understanding SM and BSM discrimination

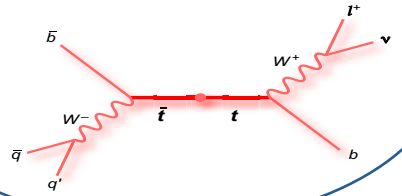
   1. A_C versus $\beta_{z,tt}$: sensitive to the enhancement of the asymmetry at high rapidity due to the increase of the qqbar->ttbar fraction

   2. A_C versus m_{tt}: different behavior for different BSM scenarios
      Also interested to be performed with a $\beta_{z,tt}$ cut to enhance qqbar->ttbar fraction

   3. Also A_C versus $p_{T,tt}$: more sensitive to asymmetry from ISR/FSR interference

Analysis flow

top pair events

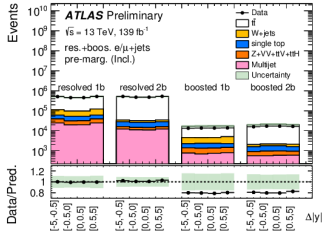


Kinematic reconstruction

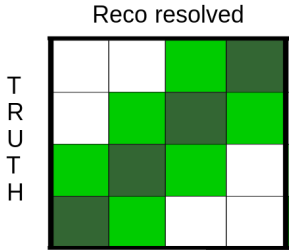
$\Delta|y|$, m_{tt}, $p_{T,tt}$, $\beta_{z,tt}$

$$L = \mathcal{B}(\widetilde{E}_{p,1}, \widetilde{E}_{p,2}|m_W, \Gamma_W) \cdot \mathcal{B}(\widetilde{E}_{lep}, \widetilde{E}_\nu|m_W, \Gamma_W) \cdot \mathcal{B}(\widetilde{E}_{p,1}, \widetilde{E}_{p,2}, \widetilde{E}_{p,3}|m_t, \Gamma_t) \cdot \mathcal{B}(\widetilde{E}_{lep}, \widetilde{E}_\nu, \widetilde{E}_{p,4}|m_t, \Gamma_t) \cdot \mathcal{W}(\hat{E}_x^{miss}|\widetilde{p}_{x,\nu}) \cdot \mathcal{W}(\hat{E}_y^{miss}|\widetilde{p}_{y,\nu}) \cdot \mathcal{W}(\hat{E}_{lep}|\widetilde{E}_{lep}) \cdot \prod_{i=1}^{4} \mathcal{W}(\hat{E}_{jet,i}|\widetilde{E}_{p,i}) \cdot P(b \text{ tag} \mid \text{quark}),$$

Background estimation



Unfolding to parton level Resolution, acceptance



Results

Unfolding

FBU Unfolding

arXiv:1201.4612
https://pypi.python.org/pypi/fbu
fbu-users@cern.ch

A_C in simulation (Powheg+Pythia)

Parton level	Δ|y|~0.58%
Reconstructed	Δ|y|~0.21%

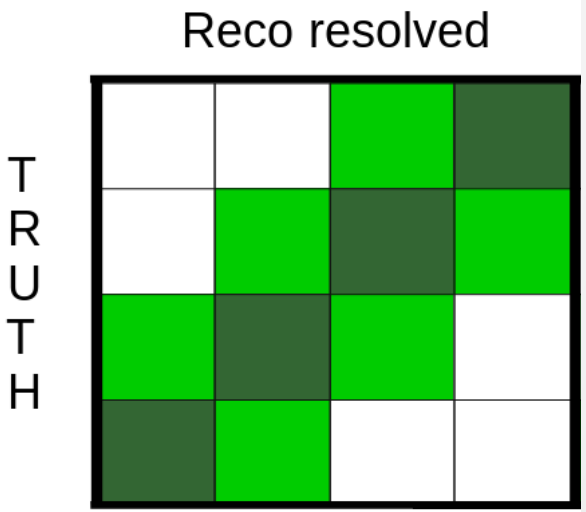
Estimates parton level A_C spectra to compare directly with theory

Reconstructed Δ|y| distribution

Map of event migration from truth to reco bins

Parton level Δ|y| distribution

$$R = M \cdot T$$



- Fully Bayesian Unfolding
  - No explicit matrix inversion
  - Prior choice determines bias
  - "automatic" handling of systematics and correlations

Fully Bayesian Unfolding

arXiv:1201.4612
https://pypi.python.org/pypi/fbu
fbu-users@cern.ch

T: Δ|y| distribution to estimate before resolution and acceptance effect

D: Δ|y| distribution measured in data distorted by resolution and acceptance effect

$$p(\mathbf{T}|\mathbf{D}) \propto L(\mathbf{D}|\mathbf{T}) \cdot \pi(\mathbf{T})$$

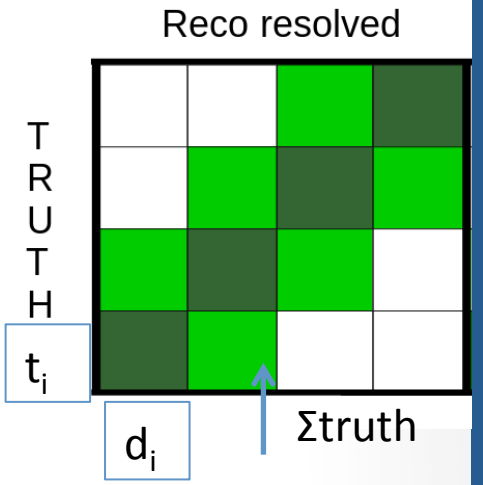
P(**T**|**D**): probability for **T** to be the correct truth-level spectrum given **D**

L(**D**|**T**): Likelihood of observing **D** given **T**. Compares the effect of acceptance and resolution on a given **T** with what is measured in data **D**

π(**T**): Prior probability of **T**. Encompass the prior knowledge or assumptions on the distribution **T**

The unfolding procedure estimates directly the probability density for A_C at the parton level

FBU likelihood



$$\mathcal{L}(\mathbf{D}|\mathbf{T}) = \prod_{i=1}^{n} Poisson(d_i, r_i(\mathbf{T}, \mathcal{M}))$$

$$r_i(\mathbf{T}, \mathcal{M}) = \sum_{j=0}^{n} \epsilon(t_j) p(r_i|t_j)$$

Migration matrix M

- Where d_i, r_i and t_i are the i-th bins of **D**, **R** and **T**
- $\varepsilon(t_i)$: acceptance for i-th bin
- $p(r_i|t_j)$: probability for an event in t_j to be reconstructed in r_i

FBU background

- In the real world **D** rarely corresponds to a pure signal sample. The background prediction **B** can be plugged in the Likelihood.

- The background prediction is included in the likelihood definition
  -> No background subtraction from data distribution before unfolding

$$\mathcal{L}(\mathbf{D}|\mathbf{T}) = \prod_{i=1}^{n} Poisson(d_i, r_i(\mathbf{T}, \mathcal{M}))$$

$$\mathcal{L}(\mathbf{D}|\mathbf{T}, \mathbf{B}) = \prod_{i=1}^{n} Poisson(d_i, r_i(\mathbf{T}, \mathcal{M}) + b_i)$$

The prior π(T)

- The π(**T**) encompass the prior knowledge or our assumptions on how the **T** parameters are distributed.
  - The simplest π(**T**) is a constant (‘flat') prior.
  - In general $\pi(\mathbf{T}) = e^{\alpha S(\mathbf{T})}$
- The choice of a prior that constrains the **T** parameters space corresponds to applying a regularization with strength.
  The curvature prior corresponds to a generalization of Tikhonov regularization
  - Curvature $$C(\mathbf{T}) = \sum_{t=2}^{n-1}\left[(T_{t+1} - T_t) - (T_t - T_{t-1})\right]^2$$
  - $S(\mathbf{T}) = \left| C(\mathbf{T}) - C(\tilde{\mathbf{T}}) \right|$ where $C(\tilde{\mathbf{T}})$ is a reference value

The prior π(T)

The charge asymmetry analysis is using a constant prior (no regularization)

FBU marginalization

- Bayesian inference framework→ marginalization of systematics
- The posterior probability is integrated over the nuisance parameters
- From the practical point of view, the algorithm has to sample a (**n**+**m**)-dimensional space where **m** is the number of nuisance parameters considered and **n** number of bins.
- The marginal probability is the projection of the sample on the **n**-dimensions.

$$\mathcal{L}(D|T) = \int \mathcal{L}(D|T,\theta) \cdot \mathcal{N}(\theta)\, d\theta$$

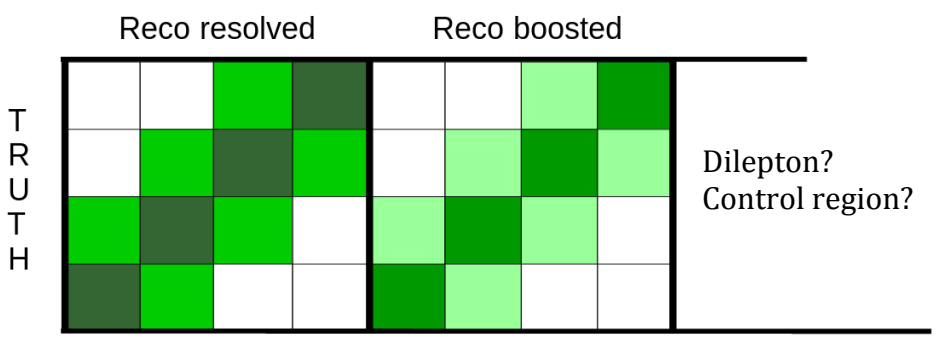
Nuisance parameter

Prior on NP

Integrate over NP

FBU Channel combination

- The flexibility of the formalism allows to combine channels as product of likelihoods
- The response matrix is extended in the reco-dimension
- Correlations of nuisance parameter handled
- Can even add (like in profile likelihood) control regions to learn about a background



$$\mathcal{L}\left(\{\boldsymbol{D}_1 \cdots \boldsymbol{D}_{N_{ch}}\}|\boldsymbol{T}\right) = \int \prod_{i=1}^{N_{ch}} \mathcal{L}\left(\boldsymbol{D}_i|\boldsymbol{T};\boldsymbol{\theta}\right) \cdot \mathcal{N}(\boldsymbol{\theta})\, d\boldsymbol{\theta}$$

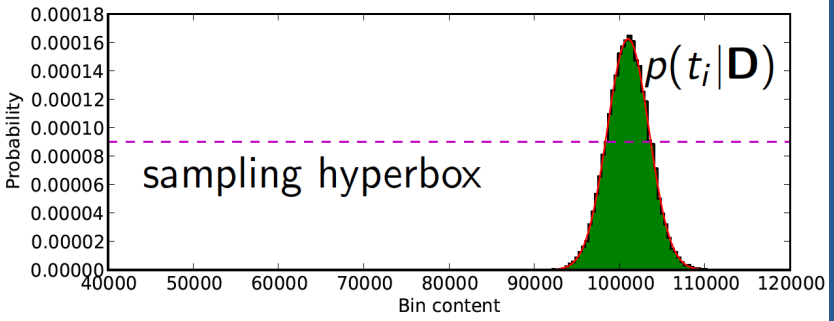
FBU in practice

- Ingredients:
  1. Input distributions $\{\mathbf{D_i}\}$ observed in data with $\mathbf{n_{reco}}$ bins for $\mathbf{n_{channels}}$
  2. Prediction for the background distributions $\mathbf{\{B_i\}}$
  3. $\mathbf{n_{truth}}\times(\mathbf{n_{reco}}\times\mathbf{n_{channels}})$ response matrix **M** derived from training signal sample
  4. Choice for $\pi(\mathbf{T})$
  5. Boundaries of $\mathbf{n_{truth}}$-dimensional space where $p(\mathbf{T}|\mathbf{D})$ has to be computed

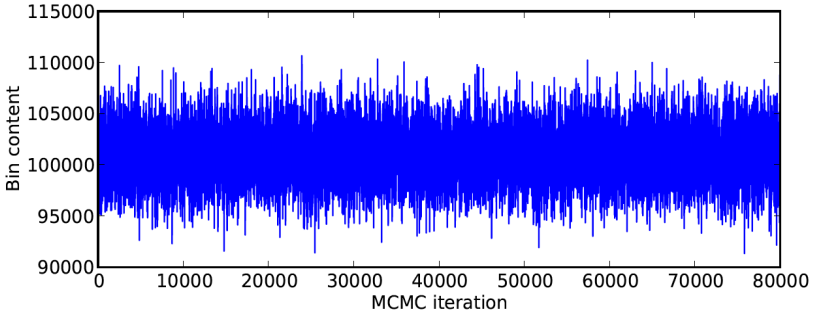
- Output:
  - $\mathbf{n_{truth}}$-dimensional probability density $p(\mathbf{T}|\mathbf{D})$
  - Measurement: posterior distribution of A_C calculated from the $\mathbf{n_{truth}}$-bin posteriors, mean and standard deviation of the posterior is used for results

FBU in practice

- The probability p(**T**|**D**) does not have a closed form and has to be estimated numerically.
- Rejection sampling is very inefficient
- Markov Chain Monte Carlo (MCMC) methods allow to perform the sampling with reasonable CPU-time.
  - But still the larger the number of dimensions the longer it takes…



trace of MCMC iterations



Sources of uncertainty and bias

- The source of uncertainties can be divided in three categories:
  - Background normalizations.
  - Object definitions (JES,b-tagging), modeling
  - Bias due to the choice of the binning, limited MC stat, regularization, response,…

- The first two are known nuisance parameters distributed with a certain prior (typically Gaussian)

- We are not interested in measuring the nuisance parameters ϑ, but we want to use the marginal likelihood to compute p(**T**|**D**).
- Off course we have posterior probability density for all NP

$$\mathcal{L}(\boldsymbol{D}|\boldsymbol{T}) = \int \mathcal{L}(\boldsymbol{D}|\boldsymbol{T},\boldsymbol{\theta}) \cdot \boldsymbol{\mathcal{N}}(\boldsymbol{\theta})\, d\boldsymbol{\theta}$$

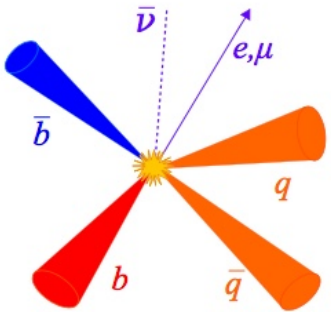
A_C @13TeV with 137fb^{-1} L+jets

ATLAS-CONF-2019-026

Event selection

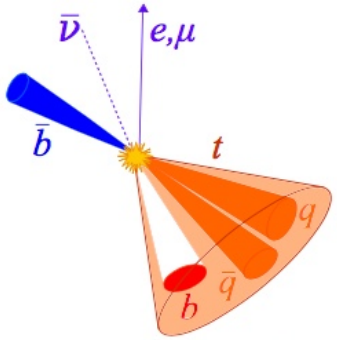
- L+jets common resolved and boosted selection
  - Exactly 1 isolated e/μ with p_T > 28 GeV
  - e+jets: E_T^{miss} > 30 GeV, M_T^W > 30 GeV; μ+jets: $M_T^W + E_T^{miss}$ > 60 GeV
  - ≥ 1 b-tagged small-R (R = 0.4) jet (MV2c10 - 77% eff. WP)

- Resolved
  - ≥ 4 small-R jets, p_T>25GeV
  - Veto boosted events
  - BDT discriminant requirements (≈ 52% eff.)



- Boosted
  - ≥ 1 large-R (R = 1.0) top-tagged jet with
  - p_T>350GeV and $|\eta| < 2$, opposite to lepton
  - ≥ 1 small-R jet close to lepton ($\Delta R(jet,l) < 1.5$)

  m(tt) > 500 GeV



Samples, Yields, Event reconstruction

- Powheg+Pythia8 tt signal sample
- MC-driven backgrounds
  - single top - Powheg+Pythia8
  - W+jets, Z+jets - Sherpa 2.2.1, diboson - Sherpa 2.2.1/2.2.
  - ttW, ttZ, ttH - aMCatNLO+Pythia8
- Data-driven fake/non-prompt lepton contribution - matrix method

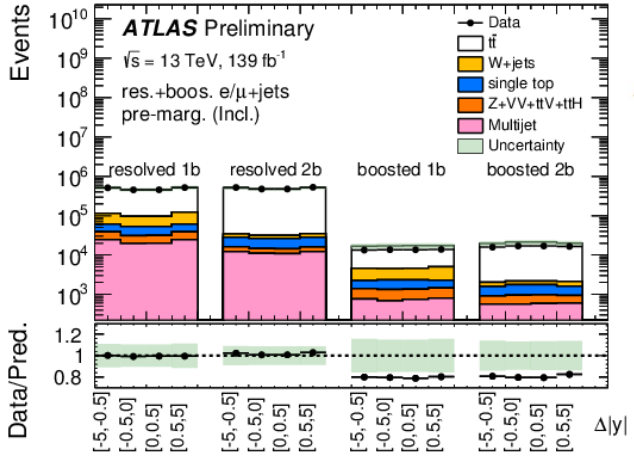
- BDT technique reconstruction in resolved topology (KLFitter lnL, m_t^{had}, m_W^{had}, b-tag information, angular distributions...)
- After optimisation: BDT discriminant > 0.3, accepting 52% signal and 27% background

Process:	Resolved		Boosted	
	1*b*-excl.	2*b*-incl.	1*b*-excl	2*b*-incl.
$t\bar{t}$	1520000±120000	1840000±150000	50000± 7000	74000±10000
Single top	89000± 12000	49000± 8000	3600± 1200	3000± 1200
W + jets	200000± 23000	23000± 14000	10000± 5000	1800± 1000
$Z + VV + t\bar{t}X$	52000± 28000	15000± 8000	2600± 1300	1400± 800
Multijet	90000± 40000	47000± 23000	3000± 1500	2300± 1200
Total Prediction	1950000±200000	1980000±160000	69000±11000	83000±11000
Data (139 fb^{-1})	1945037	2009526	54710	66582

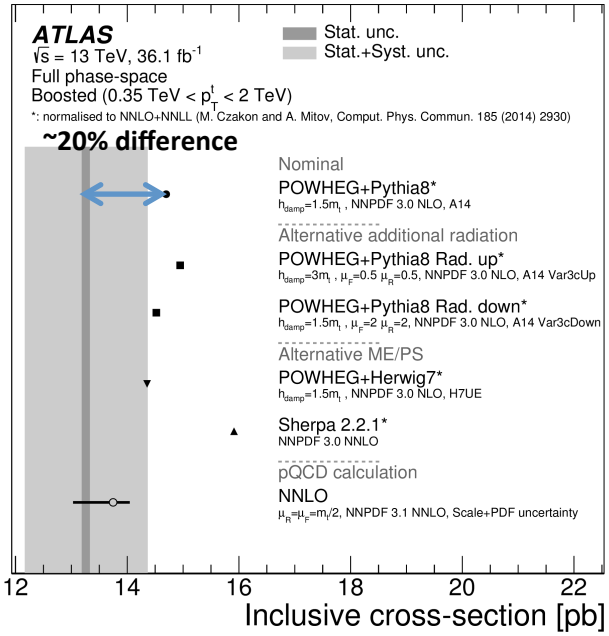
Systematics uncertainties

- Experimental unc.: lumi (1.7%), pileup, leptons, JVT, JES, JER, large jets, E_T^{miss}...
- Signal modelling - MU (aMCatNLO), PS (Powheg+Herwig7), ISR, FSR, top mass (172 GeV) (all resolved/ boosted and shape-only), PDFs
- Signal normalisation is a free parameter in FBU (parameter of interest in each bin)
- Backgrounds:
  - Single top - 5.3% norm., DR vs. DS + MU, PS, ISR, FSR
  - W+jets - norm. from scale variations (1/2 b), CKKW, QSF, μR, μF - all shape-only
  - ZVVttX - 50% norm.
  - Fakes - 50% norm. (resolved/boosted, 1/2 b), 2 parametrisations difference as a shape-only uncertainty
- All systematic uncertainties are embedded in FBU, except MC statistic in response matrix (evaluated with pseudo-experiments) and uncertainty due to unfolding bias (from linearity of the response) Both are added in quadrature

Unfolding - boosted normalisation



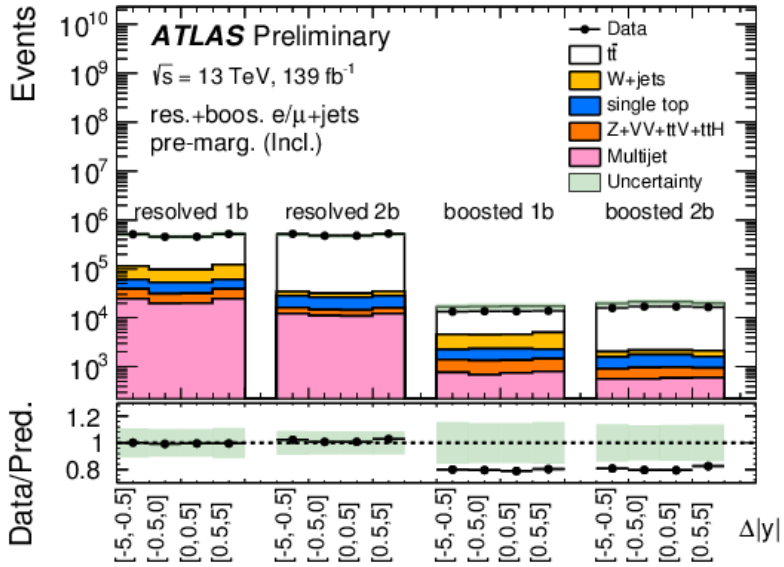
$$p\left(T|\{D_1 \cdots D_{N_{ch}}\}\right) = \int \prod_{i=1}^{N_{ch}} \mathcal{L}\left(D_i|R_i(T, K_{\text{boosted}}; \theta_s), B_i(\theta_s, \theta_b)\right) \mathcal{N}(\theta_s)\, \mathcal{N}(\theta_b)\, \pi(T)\, \pi(K_{\text{boosted}})\, d\theta_s\, d\theta_b,$$



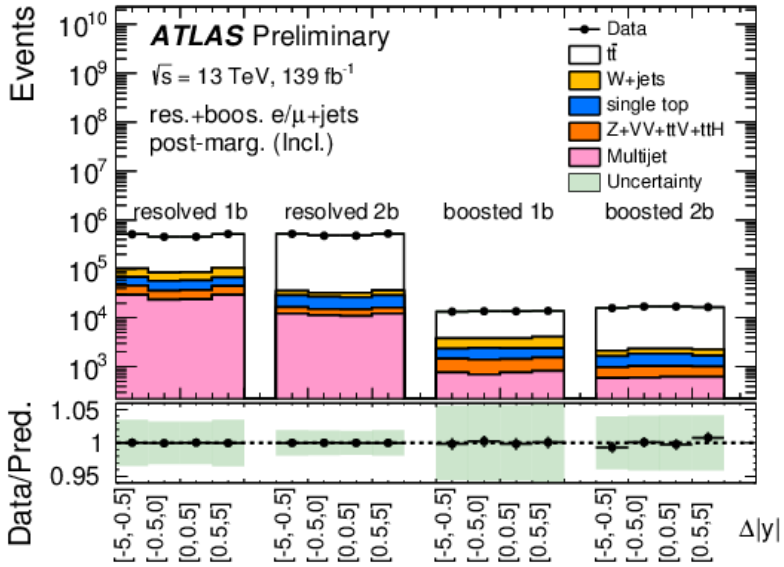
- All signal regions with 4 Δ|y| bins form one distribution in FBU
- Signal normalisation is a free parameter in each bin, however reco-level resolved + boosted regions normalisation is tied together by response matrix
- Introduce norm. uncertainty in boosted regions only with uniform prior $[0,2T_{MC}]$ – absorb boosted normalisation mismodelling

Pre/post-marginalisation Δ|y| inclusive

Pre-marginalisation:



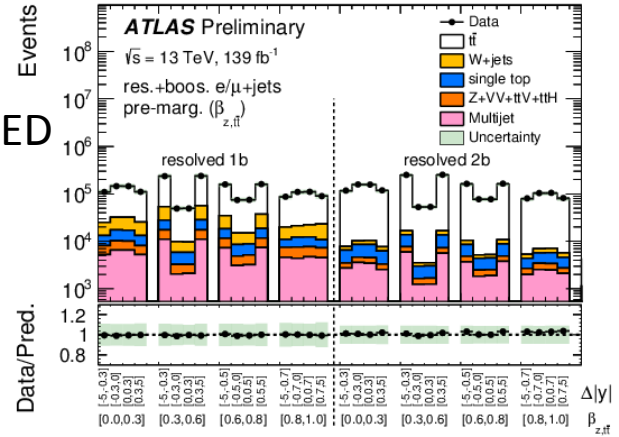
Post-marginalisation:

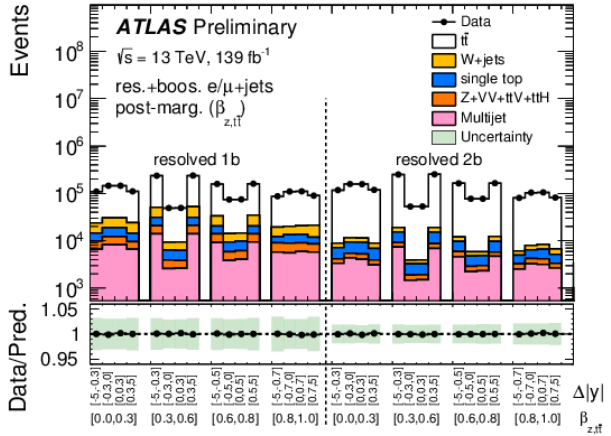


Note the change of the scale!

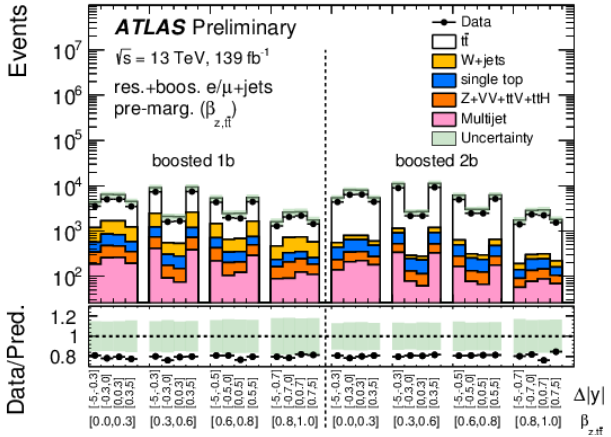
Pre/post-marginalisation Δ|y| $\beta_{z,t\bar{t}}$

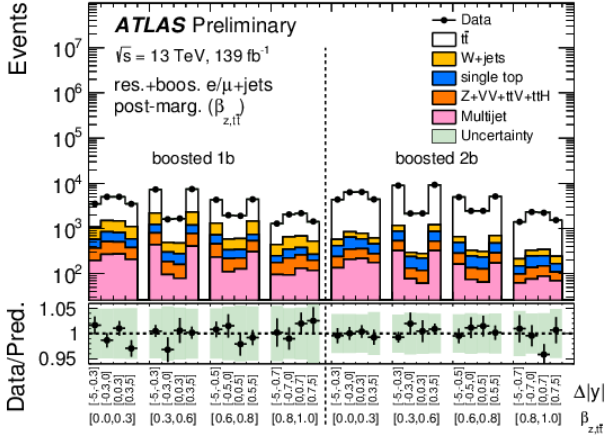
RESOLVED





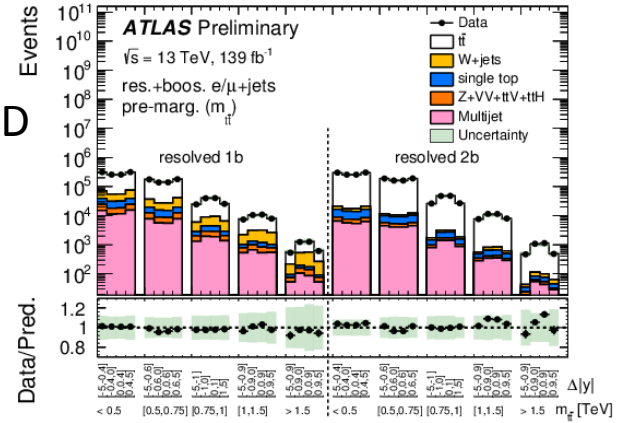
BOOSTED

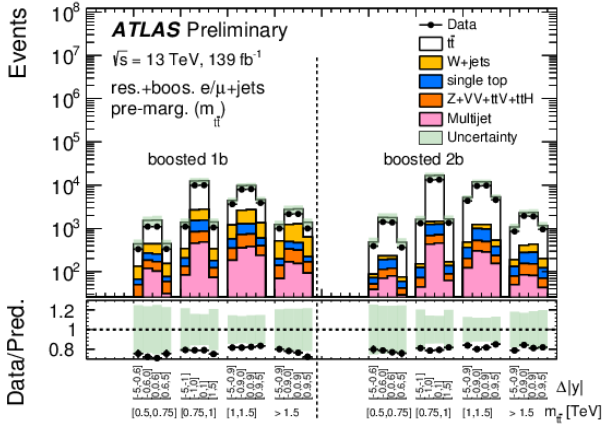




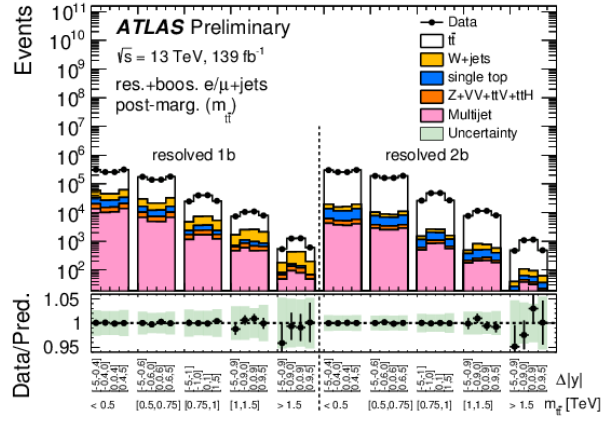
Pre/post-marginalisation Δ|y| m_{tt}

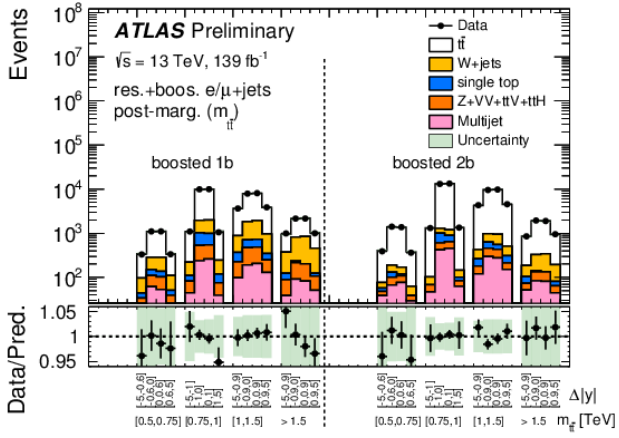
RESOLVED





BOOSTED

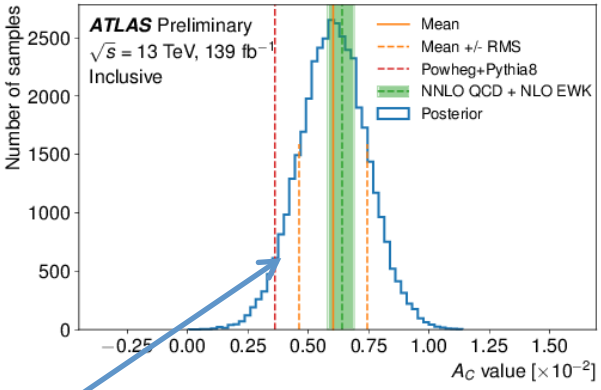


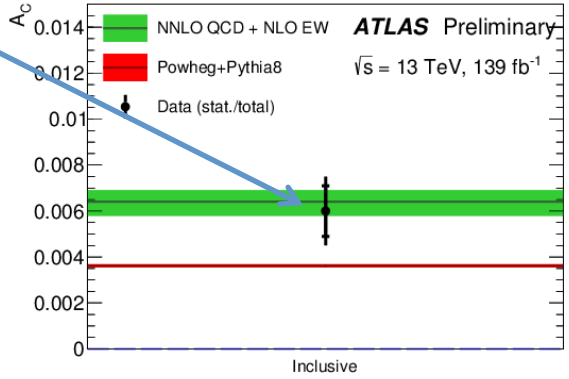


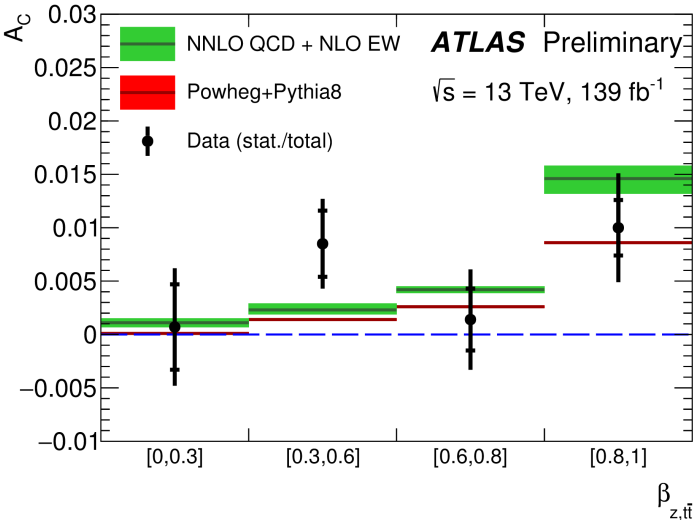
Results

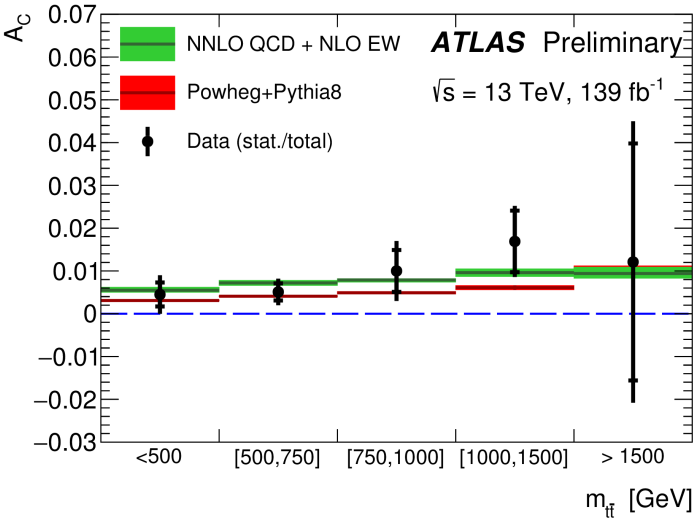
Presented at EPS for the first time!

Posterior distribution of A_C Measurement = Mean±RMS





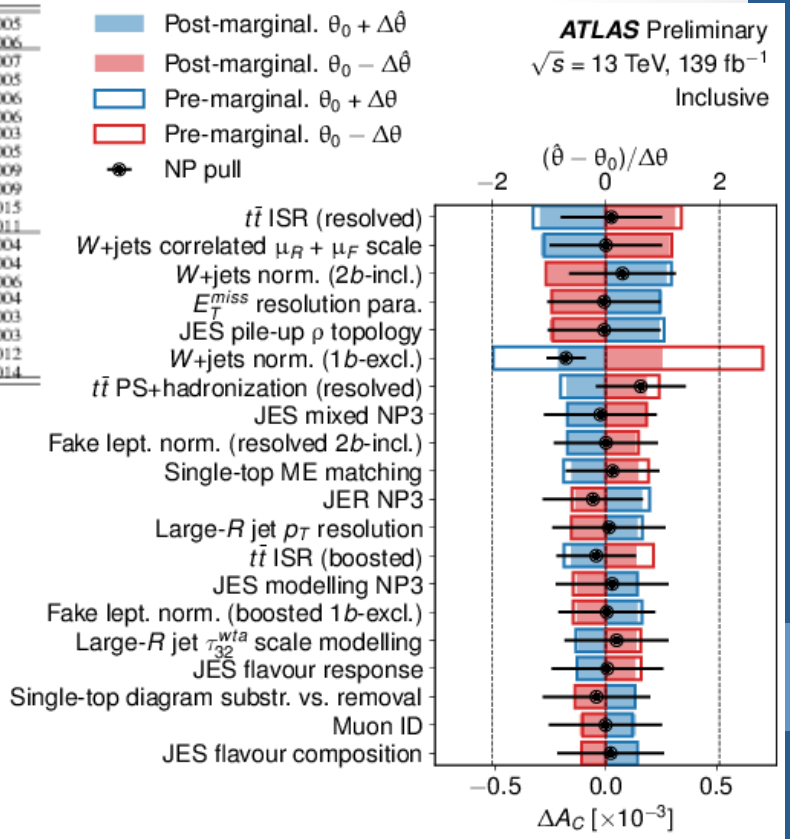




- Statistical uncertainties are mostly dominating
- Non-zero inclusive asymmetry is observed at 4σ

Results

		Data 139 fb^{-1}						SM prediction
		A_C	Stat.	Syst.	MC stat.	Bias	Total unc.	
	Inclusive	0.0060	0.0011	0.0009	0.0005	0.0001	0.0015	$0.0064^{+0.0005}_{-0.0006}$
$m_{t\bar{t}}$	< 500 GeV	0.0045	0.0028	0.0034	0.0013	0.0001	0.0045	$0.0055^{+0.0007}_{-0.0005}$
	500-750 GeV	0.0051	0.0020	0.0021	0.0009	<0.0001	0.0031	$0.0072^{+0.0006}_{-0.0006}$
	750-1000 GeV	0.0100	0.0049	0.0046	0.0021	0.0001	0.0070	$0.0079^{+0.0003}_{-0.0005}$
	1000-1500 GeV	0.0169	0.0072	0.0027	0.0029	0.0004	0.0083	$0.0096^{+0.0009}_{-0.0009}$
	> 1500 GeV	0.0121	0.0277	0.0150	0.0092	0.0005	0.0329	$0.0094^{+0.0015}_{-0.0011}$
$\beta_{z,t\bar{t}}$	0-0.3	0.0007	0.0040	0.0032	0.0020	0.0001	0.0055	$0.0011^{+0.0004}_{-0.0004}$
	0.3-0.6	0.0085	0.0031	0.0025	0.0013	0.0003	0.0042	$0.0023^{+0.0006}_{-0.0004}$
	0.6-0.8	0.0014	0.0029	0.0033	0.0015	0.0004	0.0047	$0.0042^{+0.0003}_{-0.0003}$
	0.8-1.0	0.0100	0.0026	0.0042	0.0013	0.0007	0.0051	$0.0146^{+0.0012}_{-0.0014}$



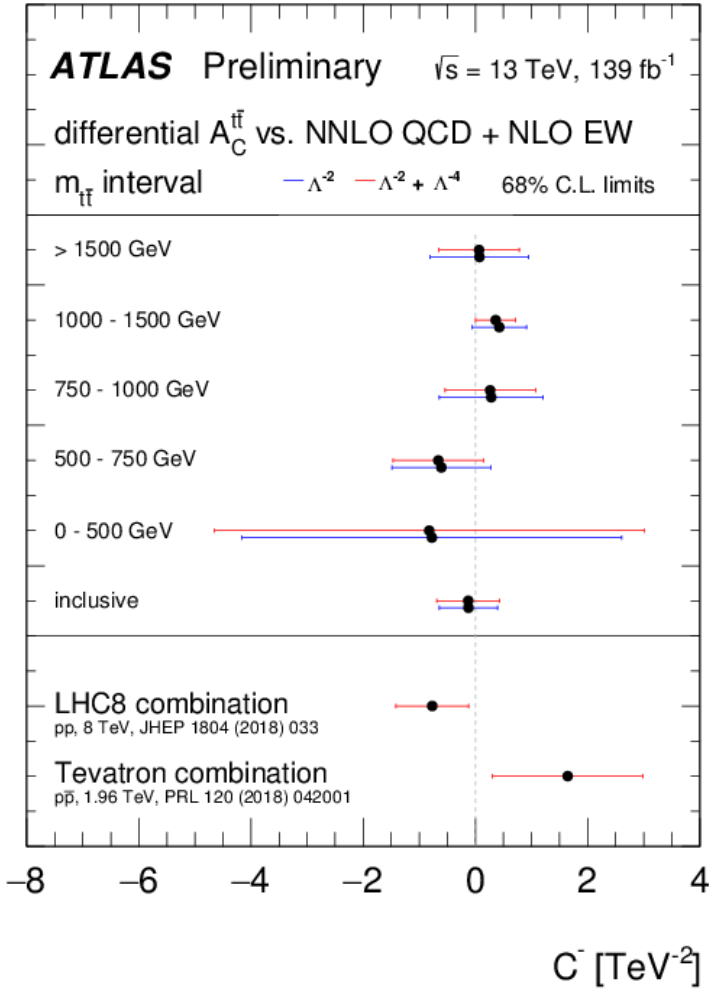
EFT interpretation

- BSM physics can be described in a model-independent way

$$\mathcal{L}_{eff} = \mathcal{L}_{SM} + \frac{1}{\Lambda^2}\sum_i C_i O_i + \mathcal{O}\left(\Lambda^{-4}\right)$$

- Using notation from Eur.Phys.J. C76 (2016) no.4, 200 the A_C measurements are sensitive to $C^- = (C_1 - C_2)$ difference in Wilson coefficients

- No signs of BSM physics observed, strong constraint wrt. Tevatron A_{FB} and LHC 8 TeV combination



Summary

1. After the 8TeV measurements we thought 13 TeV will be challenging due to further wash out of asymmetry from increased gg fraction, but we managed to improve on the analysis and measure for the first time an evidence of charge asymmetry in pp collisions

2. FBU unfolding allows to marginalise systematic uncertainties and mitigate their effect

3. I do not have fair comparison with other unfolding method, but I hope this is something that could be done

4. Detailed documentation with notebooks is being written to hopefully make this method used a bit more widely