New Perspectives Conference 2023

Measurement of the top quark-antiquark pair charge asymmetry in events with highly boosted top quarks in lepton+jets channel in proton-proton collisions at 13 TeV with the CMS detector

Beren Ozek, Cecilia E. Gerber, Ricardo Escobar, Titas Roy

University of Illinois at Chicago

FERMILAB-SLIDES-23-145-CMS

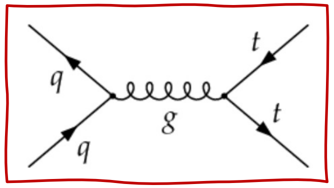
Outline

- Introduction
- Charge Asymmetry
- Analysis Techniques
- Event Selection
- Event Reconstruction
- Unfolding Method
- Results

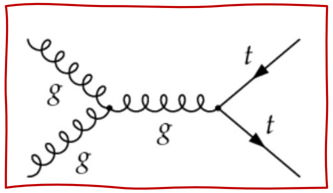
FERMILAB-SLIDES-23-145-CMS

Introduction

- Top quark has a **unique role** in the **Standard Model** and **BSM**.
  - Being the **heaviest known** elementary particle.
  - Decays **before hadronization** due to its short lifetime, allowing its properties to be measured precisely through its decay products.



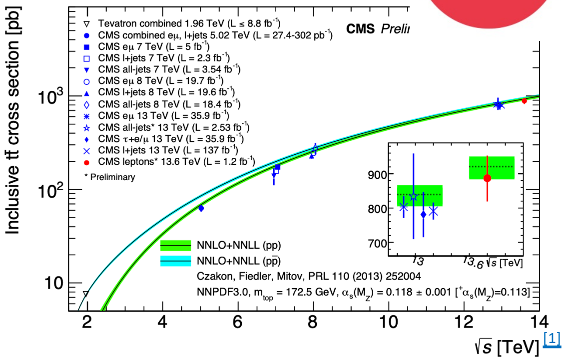
Tevatron vs LHC
Proton-antiproton vs proton-proton
85% qqbar initial state vs 90% gluon-gluon initial state
15% gluon-gluon initial state vs 10% qqbar initial state



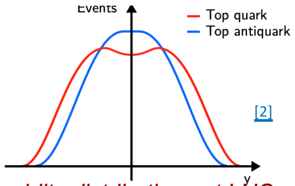
- At leading order, both are symmetric due to charge conjugation.

- At **higher orders asymmetries from quark-antiquark and quark-gluon initial states**.
  - This causes top quarks to be produced in the direction of the incoming quark which affect the overall distribution of top quarks

- The LHC is a proton-proton collider dominated by gluon fusion $t\bar{t}$ production with a symmetric nature, meaning no strict forward direction is **defined**.
  - The rapidity distribution is symmetric in forward-backward, but **asymmetric in central-peripheral (forward-central asymmetry)**

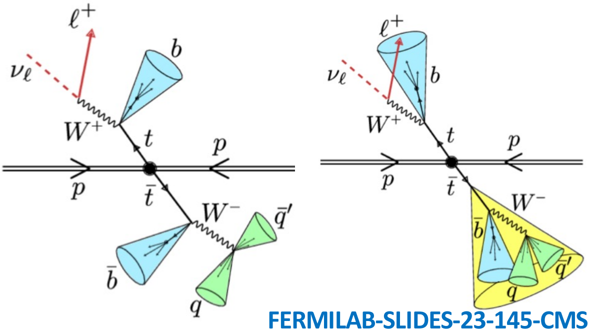


[1]



[2]

Rapidity distributions at LHC central-peripheral asymmetry



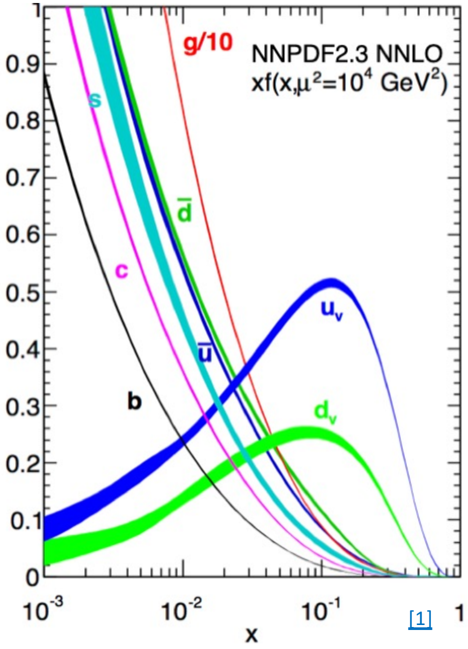
[1] First measurement of the top quark pair production cross section in proton-proton collisions at √s = 13.6 TeV, [2] arXiv:1706.00428

Motivation: Charge Asymmetry

- **Charge asymmetry is larger in boosted events** since at high momentum transfer, the relative contribution of valence quarks increases
  - The **$t\bar{t}$ charge asymmetry** is studied by analyzing differences in absolute rapidities.

- Δy is **Lorentz invariant** wrt boosts on the z-axis, so it is a suitable variable to study asymmetry

- [2]Recent paper from our group, this measurement is the first one to use CMS data at **$\sqrt{s}$=13 TeV, optimizing the reconstruction of highly Lorentz-boosted ttbar with an invariant mass above 750GeV.**



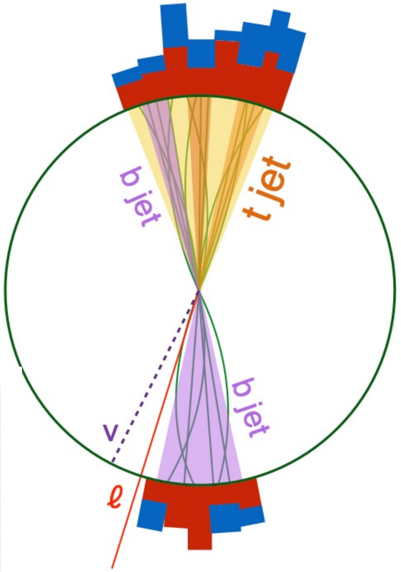
[1]

$$\boxed{\Delta|y_{t\bar{t}}| := |y_t| - |y_{\bar{t}}|}$$

$$\boxed{A_C = \frac{N(\Delta|y| > 0 - N(\Delta|y| < 0)}{N(\Delta|y| > 0 + N(\Delta|y| < 0)}}$$

[1] arXiv:1509.02358, [2] arXiv:2208.02751

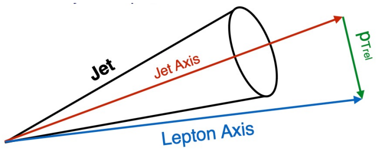
Techniques & **Event Selection** in Semileptonic channel (*μ/e+jets*)

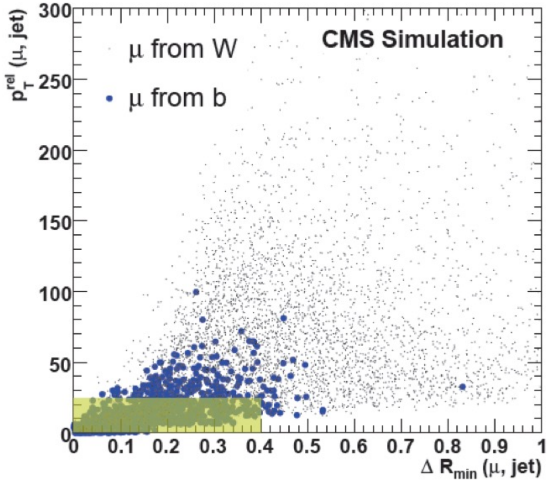


- Dedicated techniques:
    - Hadronically decaying top quarks: “top tagging” and “W-tagging” techniques
    - Leptonically decaying top quarks: high pt, no isolation cut for leptons
- The goal is to select **$t\bar{t}$ semileptonic events**
    - Main backgrounds: ST, WJets, Diboson, DYJets, QCD

Event Selection:

- **No isolation** requirement and **high-pT** for leptons
- At least two (AK4) narrow jet with one of them has to be b-tagged
- **2D cut** to control QCD multijet background**:**
    - **Δ*Rmin*(l, jet) > 0.4**
    - *p*T,rel(l, jet) > 25 GeV
- AK8 Jets with pT > 200 GeV
- MET > 50 GeV



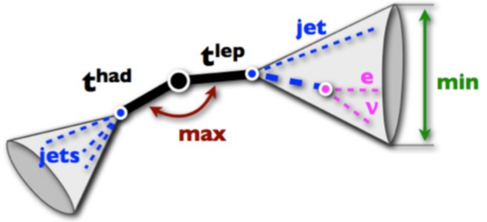


FERMILAB-SLIDES-23-145-CMS

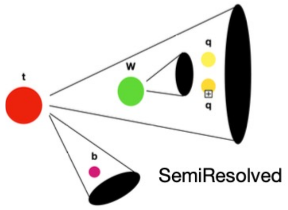
Event Reconstruction in the Semileptonic channel (*μ / e + jets)*

Event Reconstruction:

- Top quark pair events are reconstructed by assigning the four vectors of the final decay products to either the *leptonic* (t_{lep}) or *hadronic* (t_{had}) top leg

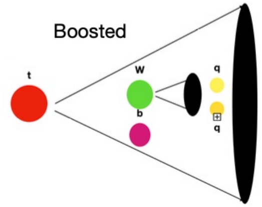


- Events are separated into three topologies for assigning the jets:
  - **Merged**: Top-tagged jets assigned to t_{had} AND jets with ΔR > 0.8 from t_{had} assigned to $t_{lep.}$ 1 Top-tagged & 0 WTagged
  - **Semi-resolved**: W-tagged jets assigned to t_{had} AND jets with ΔR > 0.8 from W-tag assigned to t_{lep} or t_{lep} or neither. 0 Top-tagged & 1 W-tagged
  - **Resolved**: 0 Top tagged jets & 0 W tagged jets.



SemiResolved

Boosted



- All combinations are tested, but only the one satisfying minimum **χ**2 is kept-as reconstructed top masses are expected to be close to the true top quark mass.

$$\chi^2 = \chi^2_{lep} + \chi^2_{had} = [\frac{M_{lep} - \bar{M}_{lep}}{\sigma_{M_{lep}}}]^2 + [\frac{M_{had} - \bar{M}_{had}}{\sigma_{M_{had}}}]^2$$

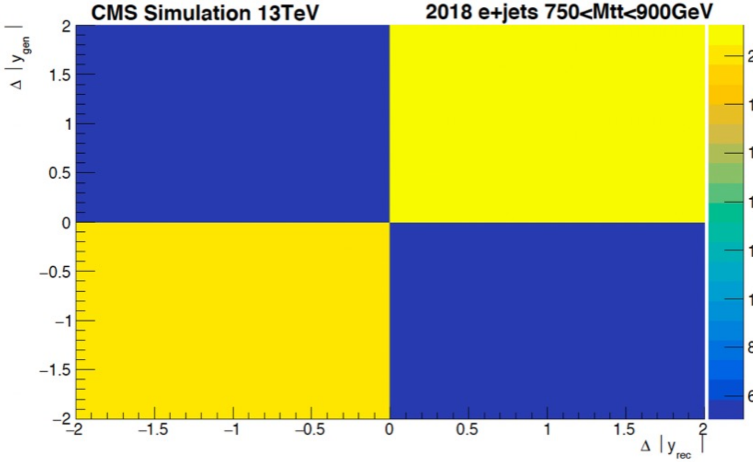
FERMILAB-SLIDES-23-145-CMS

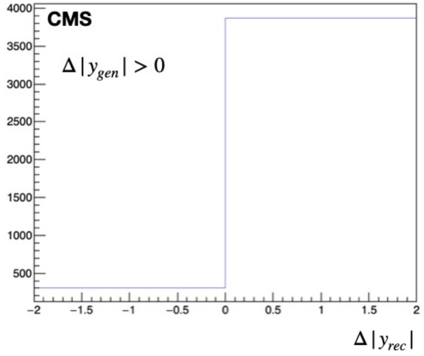
Unfolding and Combine Tool

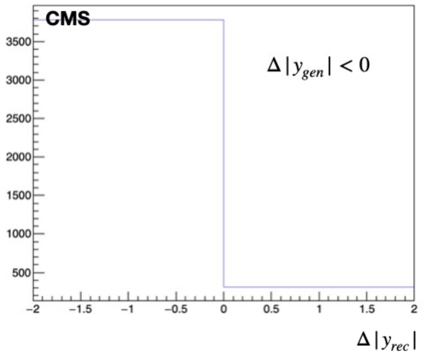
- **Unfolding method** is used **to remove smearing in the reconstructed data.**
    - Due to the **finite resolution of the particle detectors**, the spectrum of events **are smeared with respect to the true one**.
- **Higgs Combine tool,** which uses the **maximum likelihood** approach **is used.**
    - Estimates the best-fit values by maximizing the likelihood function, including **signal strengths and nuisance parameters**.

$$\mathcal{L}_k = \prod_{j=1}^{N_{\text{reco}}} P\left(n_j; \sum_{i=1}^{N_{\text{gen}}} A_{ji}(\vec{\delta}_u)\mu_i(\vec{\delta}_u) + b_j(\vec{\delta}_u)\right) N(\vec{\delta}_u)$$

- Taking **the projection of the response matrix** allows to quantify the **contributions of truth-level bins to the reconstructed-level bin**.
- We define the structure of the likelihood function with a **datacard**
    - Input of the Combine tool.
    - Each datacard represents reconstructed bins
    - 12 datacards
        - 2 lepton flavors- Electrons and muons
        - 2 mass regions - 750 GeV < Mtt < 900GeV, Mtt > 900GeV
        - 3 years - (16,17,18)







FERMILAB-SLIDES-23-145-CMS

Results and Summary

- Measuring charge asymmetry through a maximum likelihood fit that unfolds the full-phase space and accounts for all uncertainties in the Higgs Combine tool.

$$A_C = \frac{r_{\text{pos}} N^{\text{gen}}_{\text{pos}} - r_{\text{neg}} N^{\text{gen}}_{\text{neg}}}{r_{\text{pos}} N^{\text{gen}}_{\text{pos}} + r_{\text{neg}} N^{\text{gen}}_{\text{neg}}},$$

- We use **signal strength parameters (**r_{pos} and r_{neg}) in the maximum likelihood fit **to scale** the reconstructed distribution back to the unfolded full-phase space region.

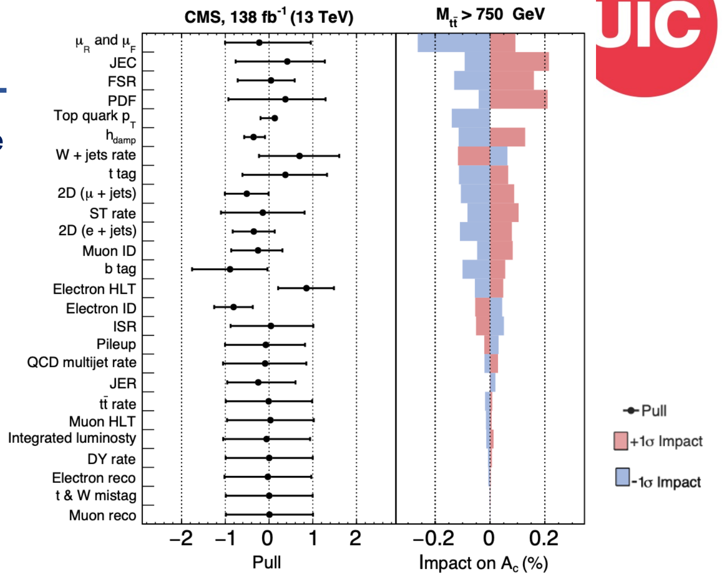
- Unfolded *Ac* in the full phase space presented **in different mass regions** with the combination of μ + jets and e + jets channels.

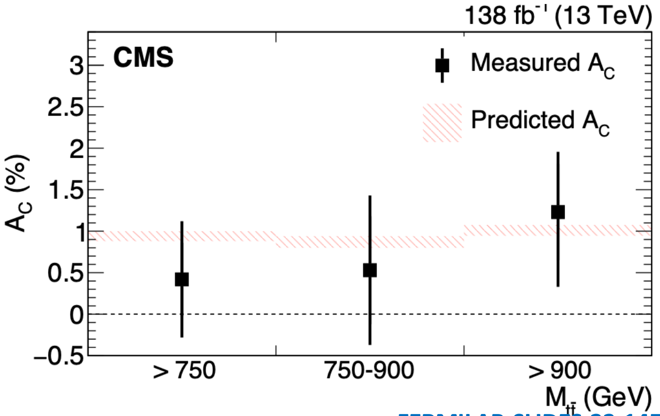
$$A_C^{full} = (0.42^{+0.64}_{-0.69})\%$$

- Theoretical prediction at next-to-next-to-leading order in QCD perturbation theory with next-to-leading-order electroweak corrections is $(0.94^{+0.05}_{-0.07})\%$. [1]
  - There is good agreement.

Work in progress:

- Adding isolated lower pT leptons to the analysis.
- We will be able to have more bins and go lower in mass by adding more Run2 data
- The result will be used as input for global EFT interpretation





FERMILAB-SLIDES-23-145-CMS

[1] M. Czakon et al., "Top-quark charge asymmetry at the LHC and Tevatron through NNLO QCD and NLO EW", *Phys. Rev. D* **98** (2018) 014003

Thank you!

Backup

Event Selection in the Semileptonic channel (*μ / e + jets*)

- Event passes one of the single muon trigger paths
- One signal muon with $pT^{\mu} > 55$GeV and $|\eta^{\mu}| < 2.4$
- **No isolation requirement**
- High-pT muons, there is a **2D cut with $\Delta R_{min}(l, jet) > 0.4$** & $p_{T,rel}(l, jet) > 25$ GeV
- At least one AK4 jet with pT > 50GeV and $|\eta| < 2.5$
- A second AK4 jet with pT > 30GeV and $|\eta| < 2.5$
- At least one AK4 jet with pT > 30GeV and $|\eta| < 2.5$ has to be b-tagged
- MET > 50 GeV

- Event passes one of the single electron trigger paths
- Exactly one electron, either with $35 < pT^{e} < 120$ GeV or with $pT^{e} > 120$ GeV, "mva- based electron ID wp80" and $|\eta e| < 2.5$ SC
- **No isolation requirement**
- High-pT electrons, there is a **2D cut with $\Delta R_{min}(l, jet) > 0.4$** & $p_{T,rel}(l, jet) > 25$ GeV
- At least one AK4 jet with pT > 50 GeV and $|\eta| < 2.5$
- A second AK4 jet with pT > 30 GeV and $|\eta| < 2.5$
- At least one AK4 jet with pT > 30 GeV and $|\eta| < 2.5$ has to be b-tagged
- MET > 50GeV

Top tagging & W tagging

- Top tagging
  - The t-tagging algorithm used in this analysis is based on the combination of a cut on the jet **soft-drop mass** and the **N-subjecttiness ratio $\tau32$.** An AK8PUPPI jet candidate is said to be t-tagged if it passes the following selection:
    - 105 GeV < Msd < 210 GeV, where Msd is the ungroomed mass of the AK8 PUPPI jet, given by the soft-drop algorithm.
    - $T32 \equiv \tau3/\tau2 < 0.65$, where Tn is the N-subjectness of the ungroomed AK8 PUPPI jet.

- W -tagging
  - The w-tagging algorithm used in this analysis is based on the combination of a cut on the jet soft-drop **mass** and the N-subjectiness ratio $\tau21$. An AK8PUPPI jet candidate is said to be W-tagged if it passes the following selection:
    - 65 GeV < Msd < 105 GeV, where Msd is the ungroomed mass of the AK8 PUPPI jet, given by the soft-drop algorithm.
    - $\tau21 \equiv \tau2/\tau1 < 0.45$, where Tn is the N-subjecttiness of the ungroomed AK8 PUPPI jet.

Reconstruction of $t\bar{t}$

For assigning the jets, we have three-phase spaces

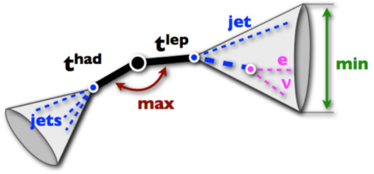
  - Boosted: the t -tagged jet is taken as th and only AK4 jets with ∆R > 0.8 from th are considered candidates for tl.
  - Semi-Resolved: the W -tagged jet is assigned to th and a list of all possible assignments of AK4 jets with ∆R > 0.8 from the W -tag jet is constructed, i.e. each jet is assigned to either tl, th, or neither of the two.
  - Resolved: a list of all possible assignments of AK4 jets is constructed, i.e. each jet is assigned to either tl, th, or neither of the two.

- In all these three cases, for each jet assignment hypothesis, the th and tl 4-momenta is given by the sum of the 4-momenta of the corresponding assigned reconstructed objects. Hypotheses with no jets assigned to either th or tl are not retained.

- Finally, only one tt hypothesis is selected in each event. The selection criterium is based on the fact that the reconstructed top quark masses are expected to be close to the true top quark mass. This is implemented in a χ2 discriminator given by

$$\chi^2 = \left[\frac{M_{lep} - \bar{M}_{lep}}{\sigma_{M_{lep}}}\right]^2 + \left[\frac{M_{had} - \bar{M}_{had}}{\sigma_{M_{had}}}\right]^2$$

Mlep and *Mhad* are the invariant masses of the reconstructed leptonic and hadronic top quarks, respectively.



boosted high momentum phase space

- In events with a *t* -tagged or *W* -w-tagged jet, *Mhad* is constructed from the groomed (softdrop) mass of the AK8 jet; this is done because, compared to the ungroomed mass, the groomed mass is much less dependent on the jet *p*T and it provides greater discrimination power against backgrounds. For each event, the hypothesis with the smallest *χ*2 value is chosen.

Reconstruction of the top quark pair events

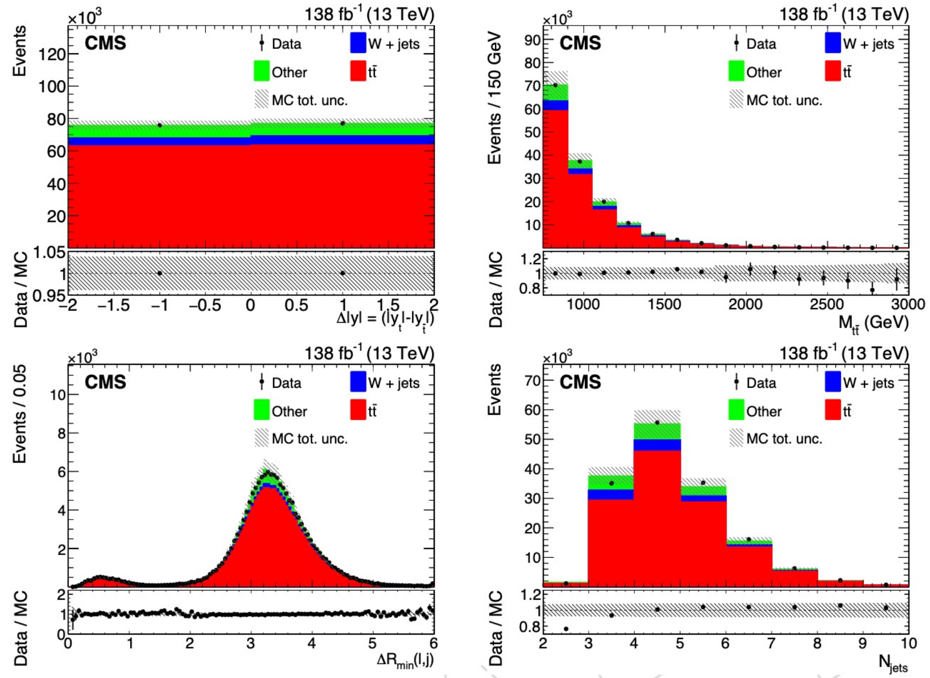


Figure 1: Comparison between data and MC simulation for kinematic distributions based on events in the candidate sample (described in Section 4): $\Delta|y|$ (upper left), reconstructed $M_{t\bar{t}}$ (upper right), distance between the lepton and the closest AK4 jet $\Delta R_{\min}(\ell, j)$ (lower left), and the number of AK4 jets (lower right). The vertical bars on the points show the statistical uncertainty in the data. The shaded bands represent the total uncertainty in the MC predictions

Likelihood Function

$$\mathcal{L}_k = \prod_{j=1}^{N_{\text{reco}}} P\left(n_j; \sum_{i=1}^{N_{\text{gen}}} A_{ji}(\vec{\delta}_u)\mu_i(\vec{\delta}_u) + b_j(\vec{\delta}_u)\right) N(\vec{\delta}_u)$$

- $P(n;\mu)$ represents the Poisson probability of observing n events when μ are expected.
- The indexes i and j run over the number of bins at generator level (N_{gen}) and reconstruction level (N_{reco}), respectively. In this analysis, we use two bins ($N_{\text{reco}} = N_{\text{gen}} = 2$) corresponding to the positive (bin 1) and negative (bin 2) difference between the absolute value of the top quark and antiquark rapidities $\Delta|y| = |y_t| - |y_{\bar{t}}|$
- A_{ji} is the response matrix, which gives the probability for an event reconstructed in bin j to have been produced in bin i. It is implemented by including the relevant number of reconstructed and generated simulated $\mathrm{t\bar{t}}$ events for each entry, which are subject to the effects of the nuisance parameters. This implementation allows the matrix to account for effects from detector resolution (smearing) as well as detector acceptance and efficiency.
- $\mu_1 = r_{\text{pos}} N_{\text{pos}}^{\text{gen}}$ and $\mu_2 = r_{\text{neg}} N_{\text{neg}}^{\text{gen}}$, where r_{pos} and r_{neg} are the signal strengths multiplying the number of signal events at generator level with $M_{\mathrm{t\bar{t}}}^{\text{gen}} > 750\,\text{GeV}$ in which the value of the value of $\Delta|y|^{\text{gen}}$ is positive ($N_{\text{pos}}^{\text{gen}}$) or negative ($N_{\text{neg}}^{\text{gen}}$), respectively.
- n_j corresponds to the number of data events in bin j.
- b_j represents the number of background events predicted in bin j.
- $N(\vec{\delta}_u)$ are the priors for the nuisance parameters with normalization uncertainties assigned a log-normal distribution and all other uncertainties a normal distribution.

Charge Asymmetry Calculation

$$A_C = \frac{N_{unf}(\Delta|y_{gen}| > 0) - N_{unf}(\Delta|y_{gen}| < 0)}{N_{unf}(\Delta|y_{gen}| > 0) + N_{unf}(\Delta|y_{gen}| < 0)}$$

$$N_{unf}(\Delta|y_{gen}| > 0) = r_{pos} \times \frac{N_{truth}(\Delta|y| > 0)}{\alpha\epsilon^{pos}}$$

$$N_{unf}(\Delta|y_{gen}| < 0) = r_{neg} \times \frac{N_{truth}(\Delta|y| < 0)}{\alpha\epsilon^{neg}}$$

acceptance $\alpha\epsilon$ measured at generator level, corrects back from the fiducial phase space of a given channel to the full phase space :

$$A_C^{full} = \frac{\alpha\epsilon^{neg} \times r_{pos} \times N_{truth}(\Delta|y| > 0) - \alpha\epsilon^{pos} \times r_{neg} \times N_{truth}(\Delta|y| < 0)}{\alpha\epsilon^{neg} \times r_{pos} \times N_{truth}(\Delta|y| > 0) + \alpha\epsilon^{pos} \times r_{neg} \times N_{truth}(\Delta|y| < 0)}$$

$$r_{pos} = \frac{\alpha\epsilon^{pos}}{\alpha\epsilon^{neg}} r_{neg} \times \frac{N_{truth}(\Delta|y| < 0)}{N_{truth}(\Delta|y| > 0)} \times \frac{1 + A_C^{full}}{1 - A_C^{full}}$$

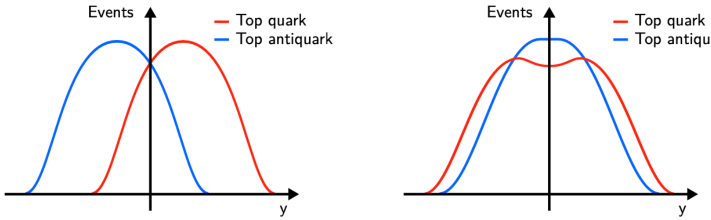
- r_{neg}: the signal strength which scales the contribution of the events with a $\Delta|y_{gen}| < 0$,
- A_C^{unf}: which gives us the value of the unfolded charge asymmetry A_C in the full phase space and its uncertainty for each of the 3 mass ranges.

Forward-Backward Asymmetry in top pair production at the Tevatron

- Investigation of the charge asymmetry in heavy quark production was performed at the Tevatron accelerator by CDF and D0 experiments.
- Tevatron was a very suitable collider for studying $t\bar{t}$ charge asymmetry due to the dominant $q\bar{q} \to t\bar{t}$ production channel.
- Asymmetry, defined as:

$$A_{\mathrm{FB}} = \frac{N(\Delta y > 0) - N(\Delta y < 0)}{N(\Delta y > 0) + N(\Delta y < 0)},$$

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$$



Rapidity distributions at Tevatron and LHC respectively
Left: forward-backward symmetry
Right:central-peripheral asymmetry

Δy is a suitable variable to study the asymmetry because it is Lorentz invariant with respect to boosts along the z-axis.

Δy is defined as:

$$\Delta y = y_t - y_{\bar{t}}$$