Measuring the top quark Yukawa coupling using $t\bar{t}$ dilepton events at CMS

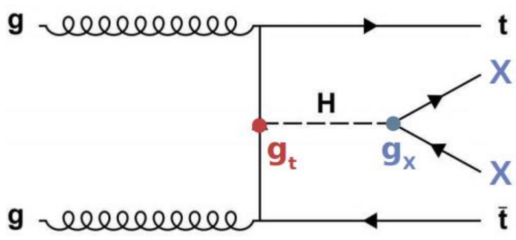
(CMS-TOP-19-008, arXiv:2009.07123)

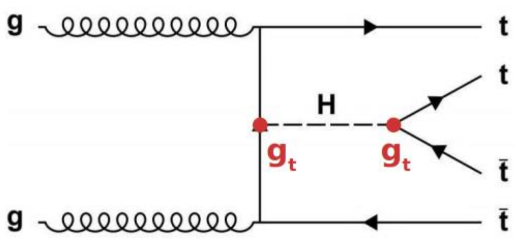
Presented by Evan Ranken

On behalf of the CMS Collaboration

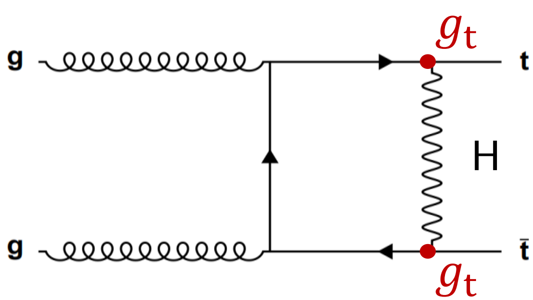
The top-Higgs Yukawa coupling

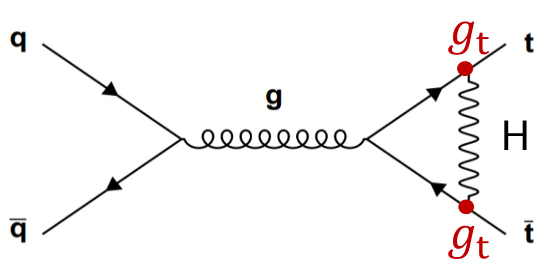
- This coupling enters directly in, for instance, $t\bar{t}H$ production:





- But it can also appear in $t\bar{t}$ production via **virtual Higgs exchange**:





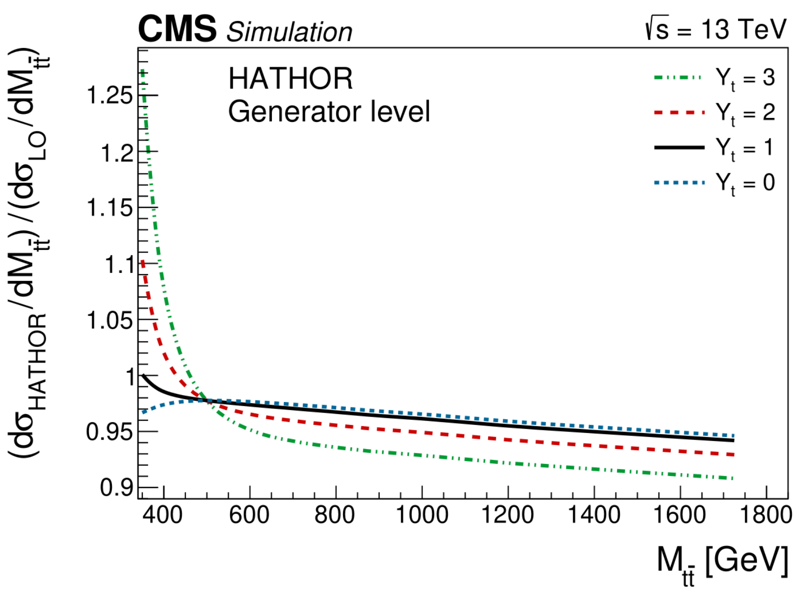
- EW diagrams begin to enter noticeably into $t\bar{t}$ production at $\mathcal{O}(\alpha_S^2\alpha)$ —or in the case of Higgs exchange, more precisely at $\mathcal{O}(\alpha_S^2 g_t^2)$

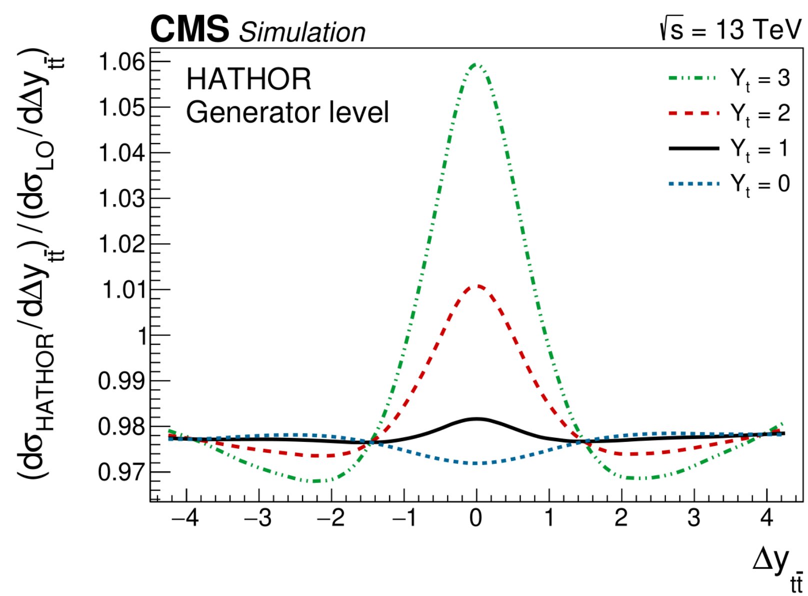
Electroweak corrections to $t\bar{t}$ production

- These EW diagrams have a minimal effect on the overall cross section $\sigma_{t\bar{t}}$, but can have noticeable shape effects on differential distributions

- We generate a multiplicative correction $R_{\text{EW}}(M_{t\bar{t}}, \Delta y_{t\bar{t}})$ as a function of the **invariant mass** $M_{t\bar{t}}$ and the **difference in top quark rapidity,** $\Delta y_{t\bar{t}}$

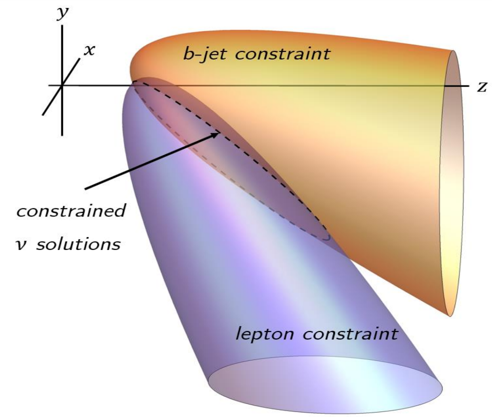
- We parametrize the R_{EW} as a function of $Y_t = g_t \,/\, g_t^{SM}$, and perform a profile likelihood fit on the parameter Y_t





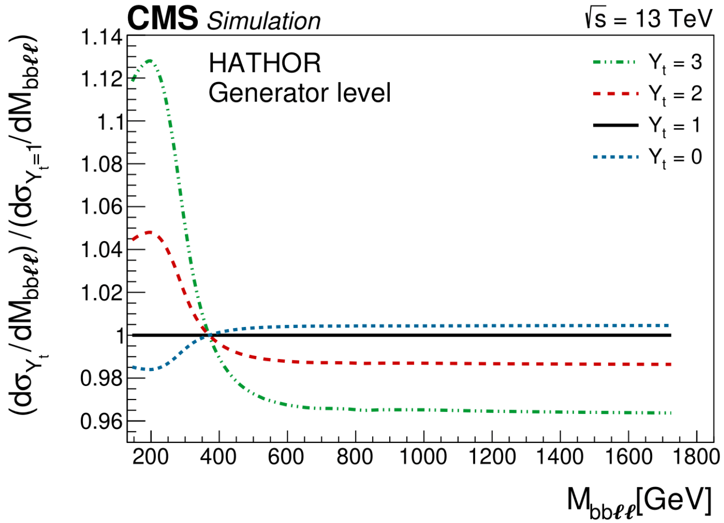
Event reconstruction & kinematic variables

- Although reconstruction of dilepton-channel top quark kinematics is theoretically possible to good approximation, it is hindered by multiple effects:
  - Highly sensitive to measured p_T^{miss}
  - Deviations from on-shell behavior

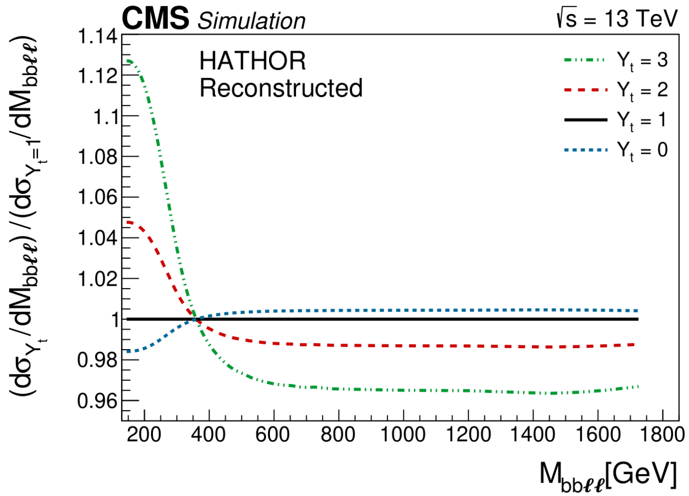


- We find that good sensitivity can be achieved using "proxy" kinematic variables: →

$$M_{bb\ell\ell} = M(b + \bar{b} + \ell + \bar{\ell})$$
$$|\Delta y_{b\ell b\ell}| = |y(b + \bar{\ell}) - y(\bar{b} + \ell)|$$

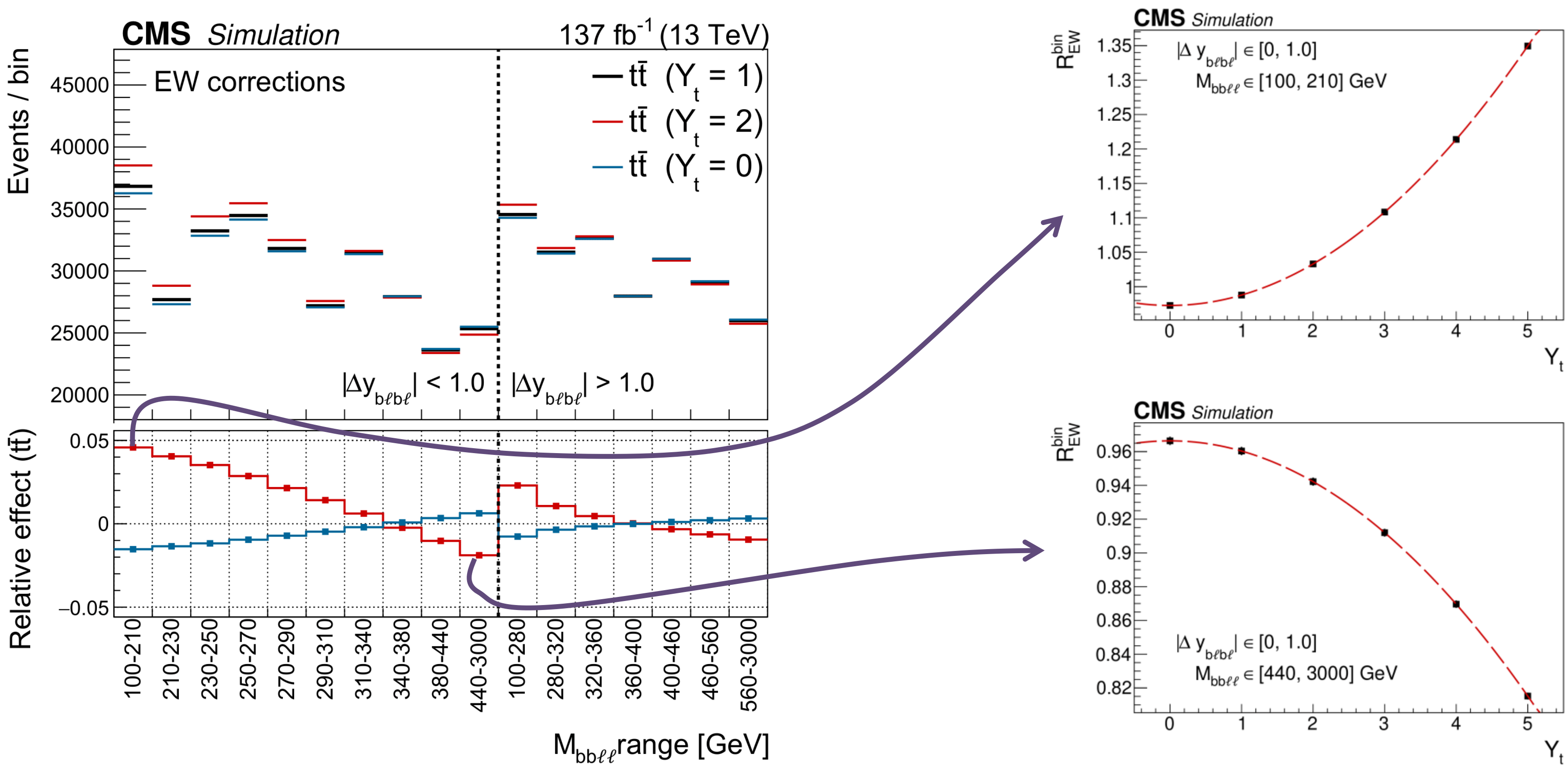


Sensitivity is preserved → by reconstruction



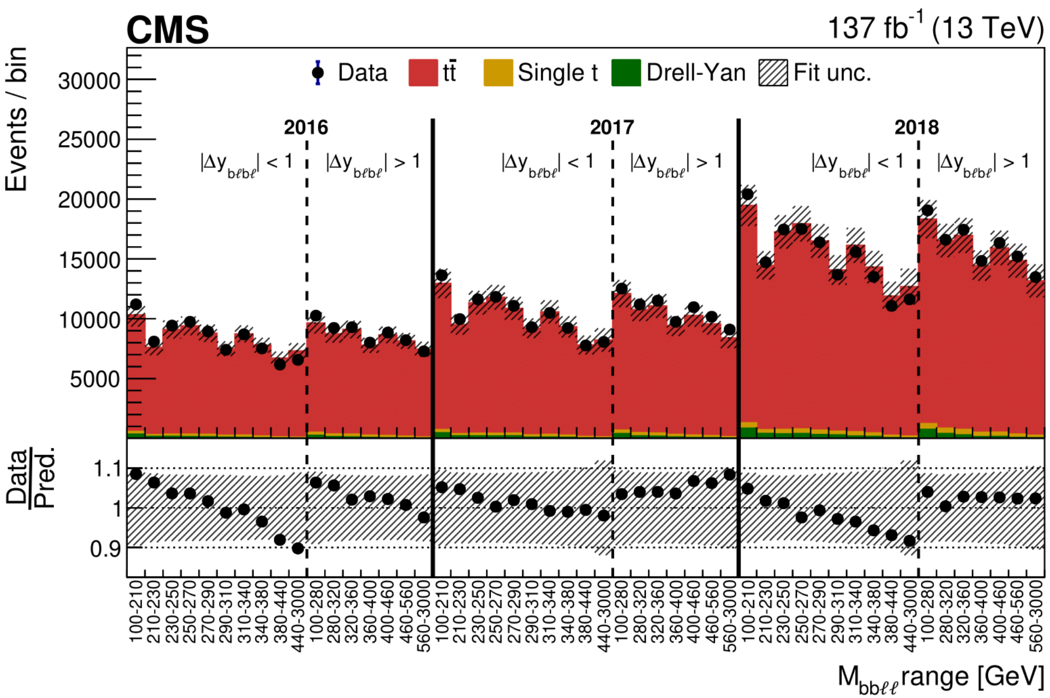
Effect of top Yukawa on binned data

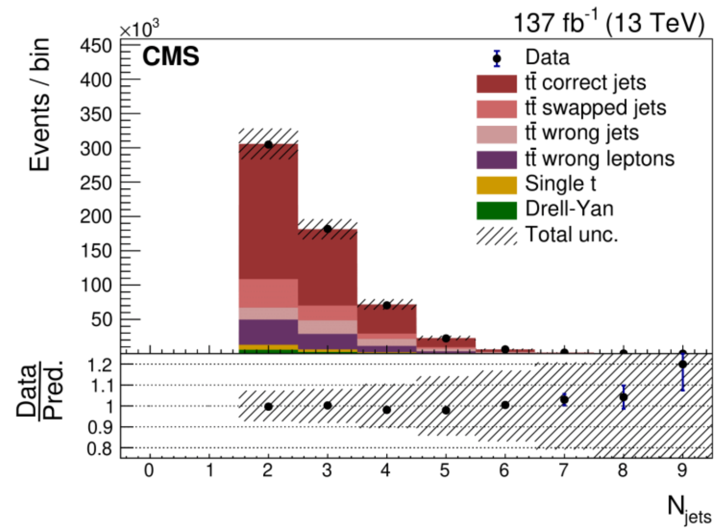
- We employ a coarse binning in $\Delta y_{b\ell b\ell}$ and a finer binning in the main sensitive variable, $M_{\mathrm{bb}\ell\ell}$. Each bin has a yield $\propto Y_t^2$
- The shape effect of the EW corrections in this binning is shown below

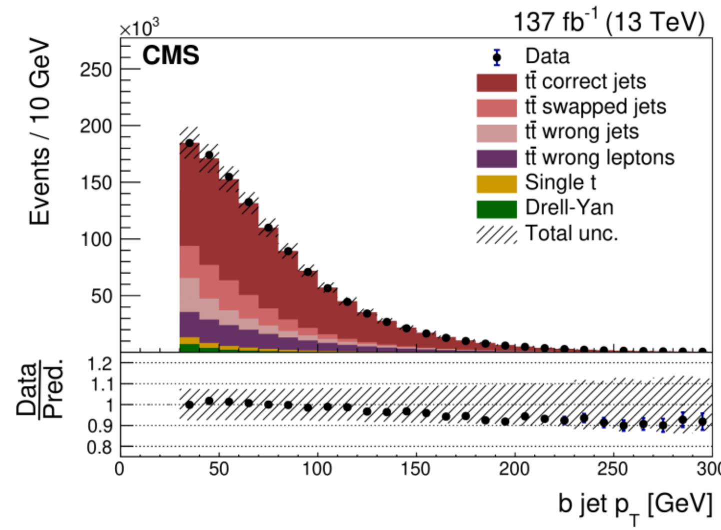


Pre-fit agreement

- Before fitting the parameter Y_t, we compare agreement between data and MC simulation across the full Run 2 period, both in the final binning and in other kinematic variables

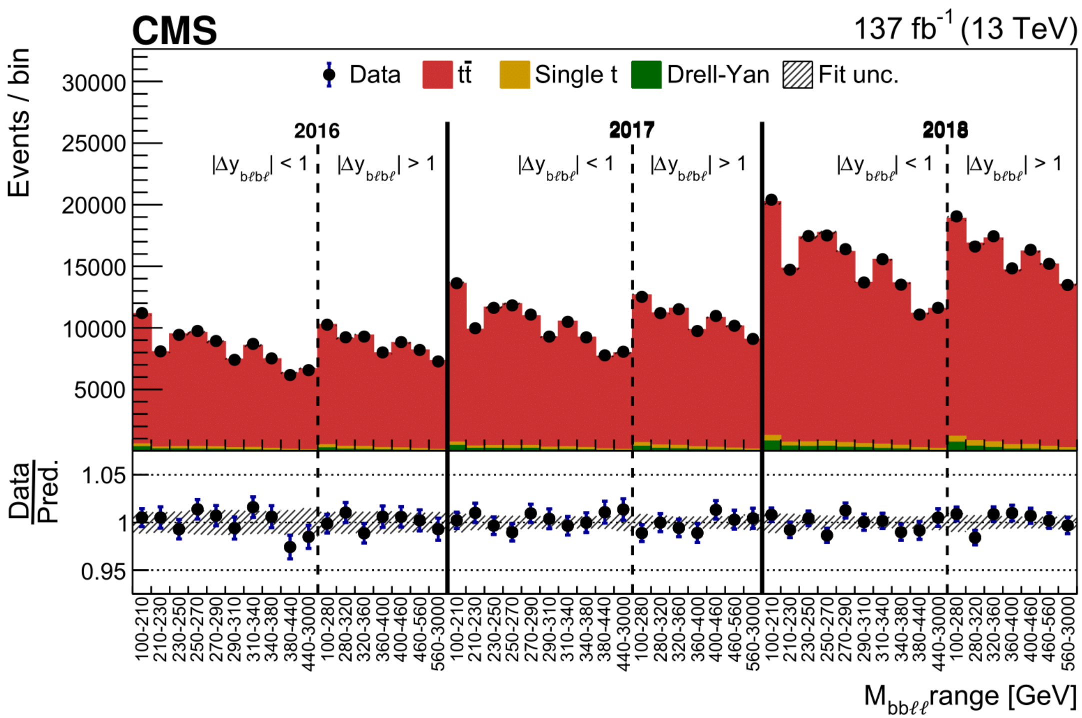


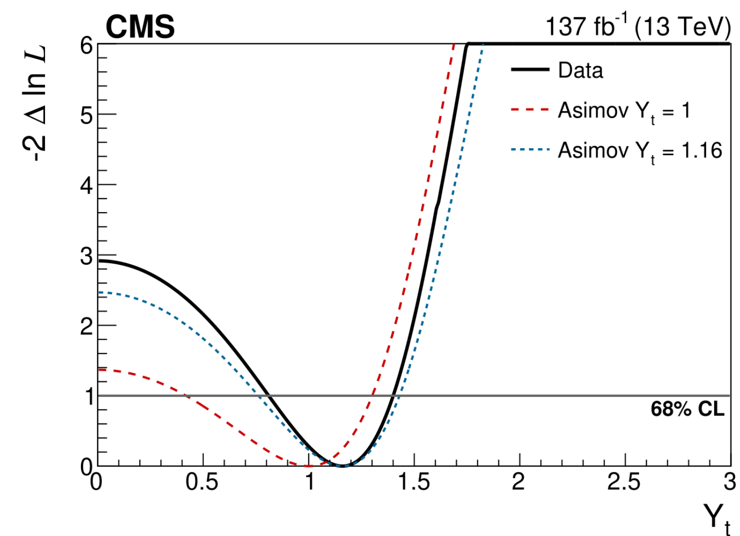




Measurement outcome

- The profile likelihood scan yields $Y_t = 1.16^{+0.07}_{-0.08}(\text{stat})^{+0.17}_{-0.27}(\text{syst})$
- This achieves better precision than a similar CMS measurement in the lepton+jets channel, using 2016 data to obtain $Y_t = 1.07^{+0.34}_{-0.43}$ [*PRD* **100** 072007 (2019)]
- We also bound $Y_t < 1.54$ at approximately 95% confidence, similar the most recent CMS "four top" search result of $Y_t < 1.7$ [*EPJC* **80** 75 (2020)]





Conclusions

- We report a measured value of $Y_t = 1.16^{+0.24}_{-0.35}$
- This measurement uses a **novel approach** to determine the top quark Yukawa coupling, requiring **minimal assumptions** about other parameters
- It provides a **complementary result** to the recent paper in the lepton+jets channel using 2016 data [*PRD* **100** 072007 (2019)] with some differences:
  - Lower backgrounds
  - Full CMS Run 2 dataset (2016-2018, 137 fb^{-1})
  - Partial kinematic reconstruction
- The measurement is **limited by modelling uncertainties**, and we could see more precision added to similar measurements as modelling improves

Backup

Comparison to other measurements

- The most precise CMS measurement of the top-Higgs Yukawa coupling comes from a combination fit using the κ-framework, yielding

$$Y_t = 0.98 \pm 0.14$$ [*EPJC* **79** 421 (2019)]

  - However, that measurement uses a combination of several Higgs-related processes and folds in assumptions about other standard model parameters

- A more closely related measurement is that of four-top production. The most recent CMS four-top search yielded [*EPJC* **80** 75 (2020)]

$$Y_t < 1.7 \text{ @ 95\% confidence}$$

  - ATLAS also recently released a preliminary four top search, but they do not quote a Yukawa limit (arXiv:2007.14858)

- CMS-TOP-19-008: $Y_t = 1.16^{+0.24}_{-0.35}$, $Y_t < 1.54$ @ 95% C.L.

Dominant uncertainties

- Due to the complicated nature of modern event simulation and CMS data-taking from 2016-2018, we have over 50 independent sources of uncertainty in the fit

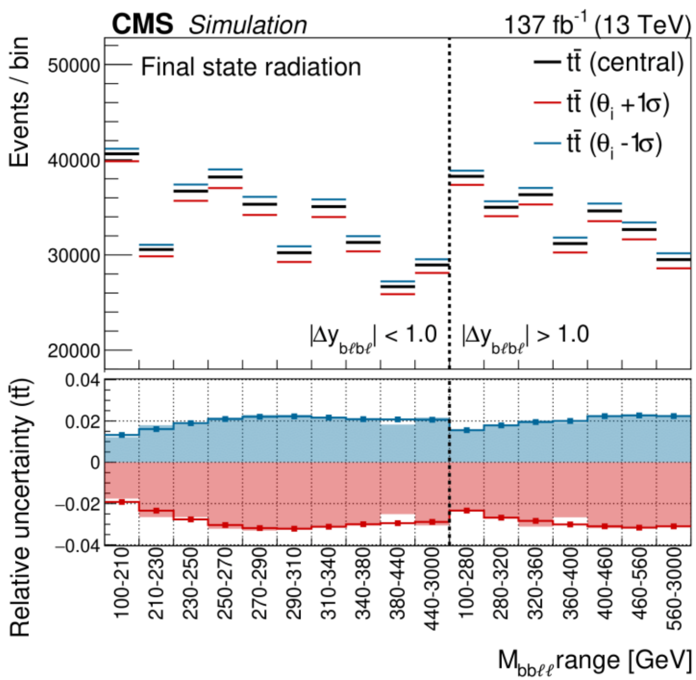
- The measurement is found to be limited primarily by *modelling uncertainties*, with the dominant few being:

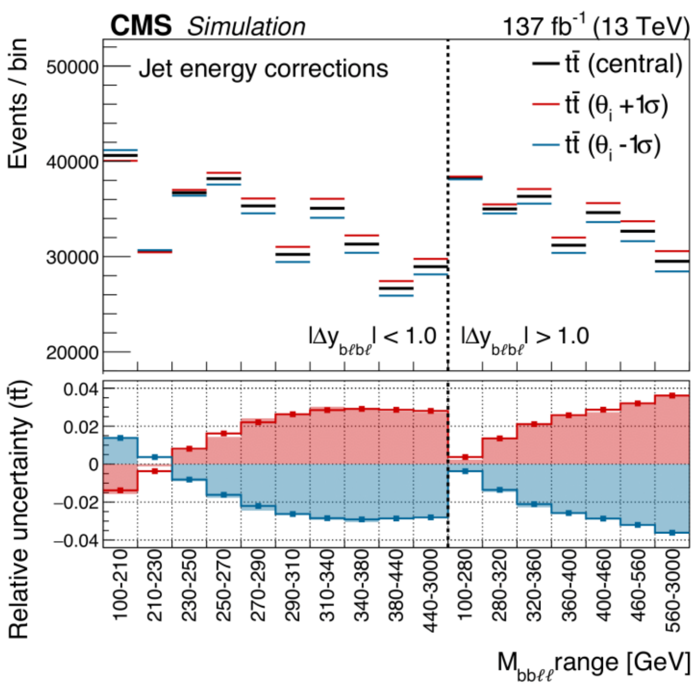
  - **EW correction uncertainty**
    (covering contributions order $\alpha_S^3\alpha$ terms)
  - **Renormalization and factorization scale uncertainties**
    (covering higher order QCD terms)
  - **Parton showering scale uncertainties**
    (uncertainty on modelling of initial- and final-state radiation)

- The only dominant *experimental uncertainty* comes from the calibration of the CMS detector's jet energy response as a function of jet η and p_{T}:
  - **jet energy scale corrections**

Handling of uncertainties

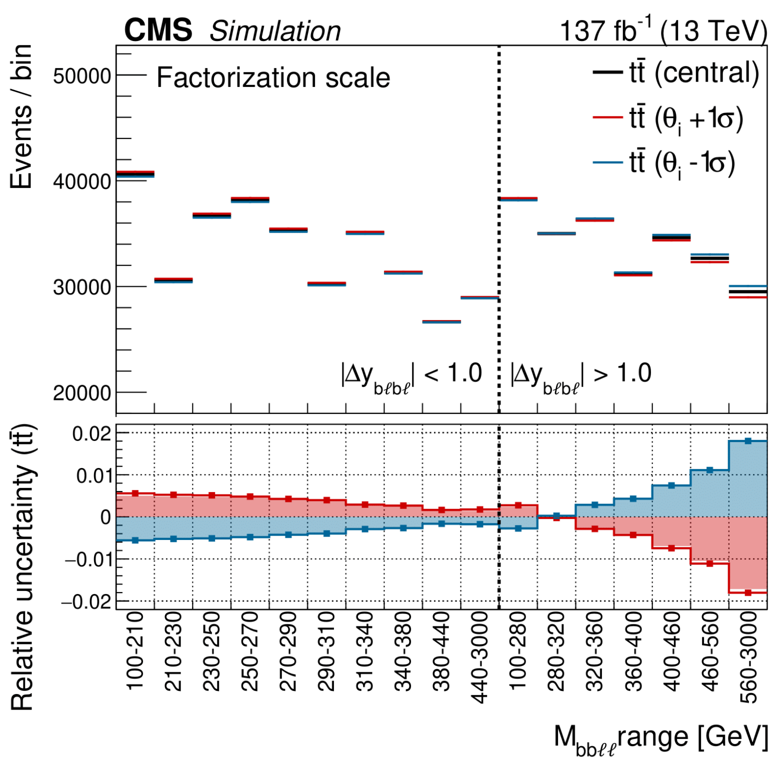
- Uncertainties are included on signal and background normalization, as well as total integrated luminosity of events observed at CMS
- Other sources of systematic uncertainty are included as shape effects via *vertical template morphing*, with sample templates shown below

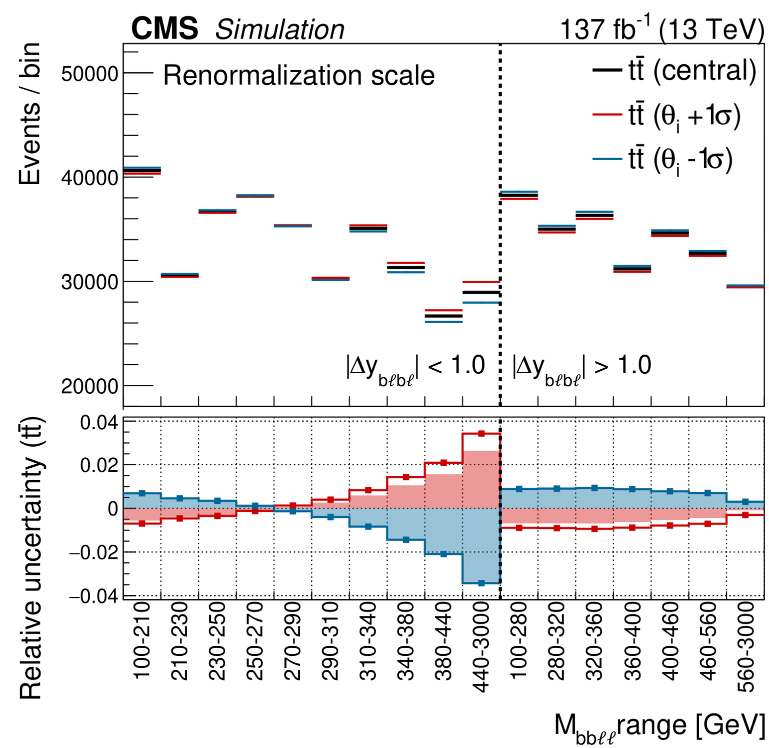




Dominant uncertainties

- Renormalization and factorization scale are each varied up and down by a factor of 2 in the POWHEG simulation
- The overall normalization effect is removed, as a separate uncertainty is applied on the $t\bar{t}$ cross section calculation (performed at NNLO)

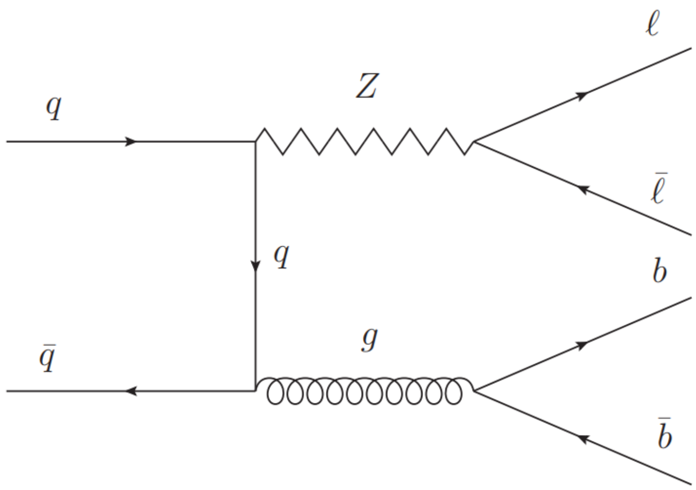




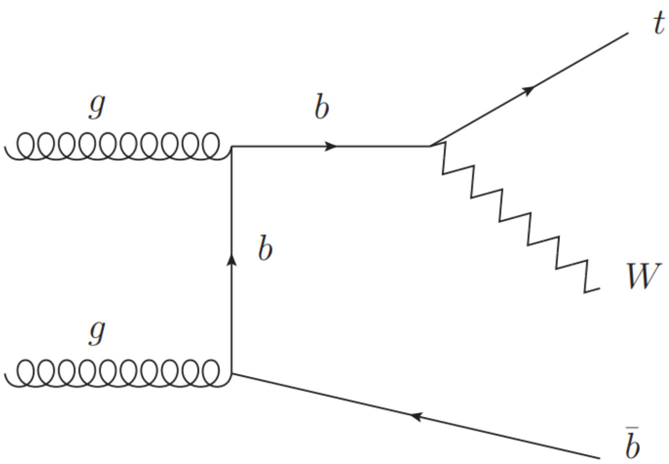
Signal and background modeling

- **Signal:** $t\bar{t}$ modeled with Powheg+Pythia, reweighted via *HATHOR*
- **Backgrounds**
  - Drell—Yann Z boson decays ($\sim$ 2% of final sample)
  - Single top production ($\sim$ 2% of final sample)
  - Diboson decays modeled with Pythia (< 0.5% of final sample)

Drell-Yan background event:



Single top background event:



Statistical setup and measurement

- The measurement is performed via a maximum likelihood fit, with uncertainty assessed by a profile likelihood scan. We define the likelihood

$$\mathscr{L} = \left[\prod_{\mathrm{bin} \in (M_{\mathrm{b}\ell}, |\Delta y|_{bl})} \mathscr{L}_{\mathrm{bin}} \right] \times p(\phi) \times \prod_i p(\theta_i) \; ,$$

$$\mathscr{L}_{\mathrm{bin}} = \mathrm{Poisson} \left[n^{\mathrm{bin}}_{\mathrm{obs}} \middle| s^{\mathrm{bin}}(\{\theta_i\}) \times R^{\mathrm{bin}}_{\mathrm{EW}}(Y_{\mathrm{t}}, \phi) + b^{\mathrm{bin}}(\{\theta_i\}) \right],$$

where

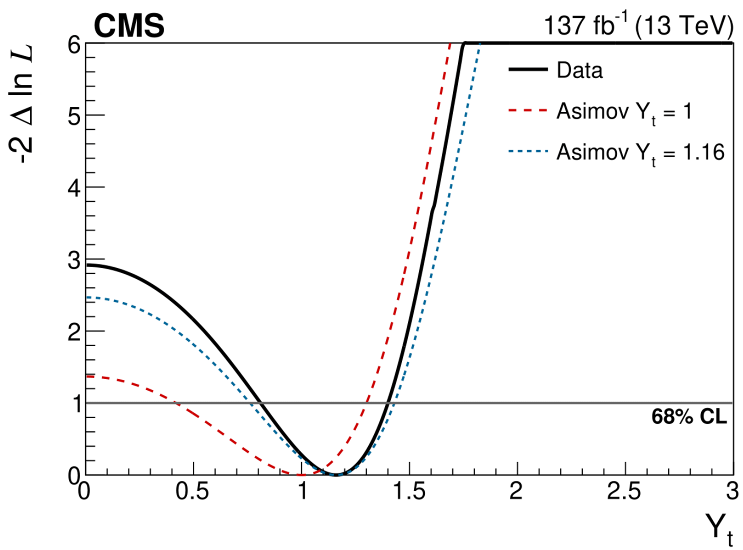
- $n^{\mathrm{bin}}_{\mathrm{obs}}$ = total observed bin count
- s^{bin} = expected signal (all $\mathrm{t\bar{t}}$ events)
- b^{bin} = expected background
- $\{\theta_i\}$, ϕ = nuisance parameters
- $p(\theta_i)$, $p(\phi)$ = penalty terms (Gaussian or log-normal pdfs)

and $R^{\mathrm{bin}}_{\mathrm{EW}}(Y_{\mathrm{t}}, \phi)$ is the EW correction multiplier, including uncertainty modulated by ϕ

Comparison to lepton+jets measurement

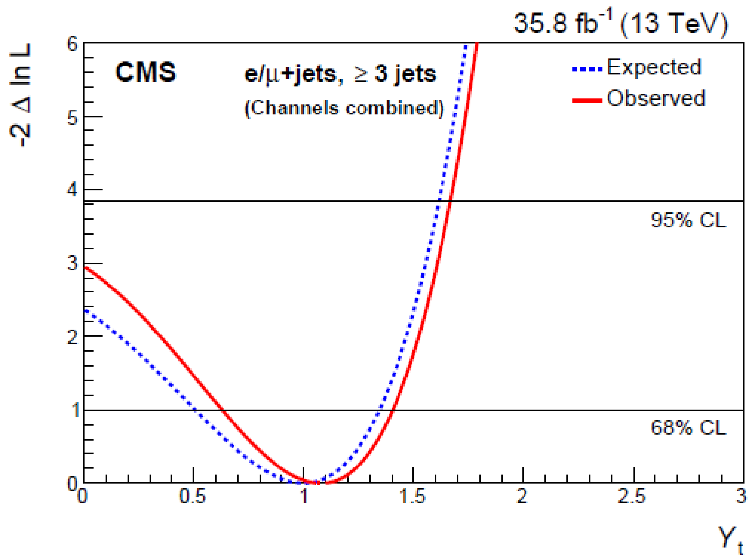
- Our group performed a similar measurement in the lepton+jets channel using 2016 data only, published in 2019 (arXiv:1907.01590). We compare the results here

this measurement:



$$Y_t = 1.16^{+0.24}_{-0.36}$$

Lepton+jets, 2016 data:



$$Y_t = 1.07^{+0.34}_{-0.43}$$

Event counts after selection

Source	2016 (36 fb^{-1})	2017 (41 fb^{-1})	2018 (60 fb^{-1})	All (137 fb^{-1})	% MC
$\mathrm{t\bar{t}}$	140,830±130	170,550±100	259,620±150	571,010±220	96.2%
Drell-Yan	1920±50	2690±80	4960±130	9840±170	1.7%
Single t	3020±30	3520±20	5830±30	12,370±50	2.1%
Diboson	140±10	150±10	250±20	540±20	0.1%
Total	145,940±150	177,400±120	270,660±200	593,760±280	100%
Data	144,817	178,088	264,791	587,696	98.9%

Trigger & lepton selection

- Events are selected using **single-lepton triggers** in the CMS trigger system

- Online selection
  - **muons:** single isolated muon with $p_{\mathrm{T}} > 24/27/24$ GeV in 2016/17/18
  - **electrons:** single isolated electron with $p_{\mathrm{T}} > 27/27/32$ GeV in 2016/17/18

- Offline Selection
  - **Leading lepton**: $p_{\mathrm{T}} > 30$ GeV*, isolated
    *electrons in 2018 require $p_{\mathrm{T}} > 34$ GeV to match trigger
  - **Second lepton:** $p_{\mathrm{T}} > 20$ GeV, isolated, opposite charge
  - **Veto:** additional isolated leptons with $p_{\mathrm{T}} > 15$ GeV

Jet selection and background removal

- We use the **DeepCSV** algorithm to identify b jets
- This algorithm has 3 "b tagging" working points (WP): *loose, medium, tight* (in order of decreasing sensitivity, increasing specificity)
- **Jet selection:**
  - $p_{\mathrm{T}} > 30$ GeV
  - Require at least 2 jets passing *loose* WP
  - Consider events with >2 b-tagged jets if exactly 2 satisfy a higher WP
- **Additional selection requirements**
  to remove Drell-Yann background in the ee and $\mu\mu$ channels, we exclude events with:
  - 81 GeV $< M_{ll} <$ 101 GeV **(Z-mass cut)**
  - $M_{ll} < 50$ GeV (low-mass Drell-Yann cut)
  - $p_{\mathrm{T}}^{\mathrm{miss}} < 30$ GeV (missing p_{T} cut)

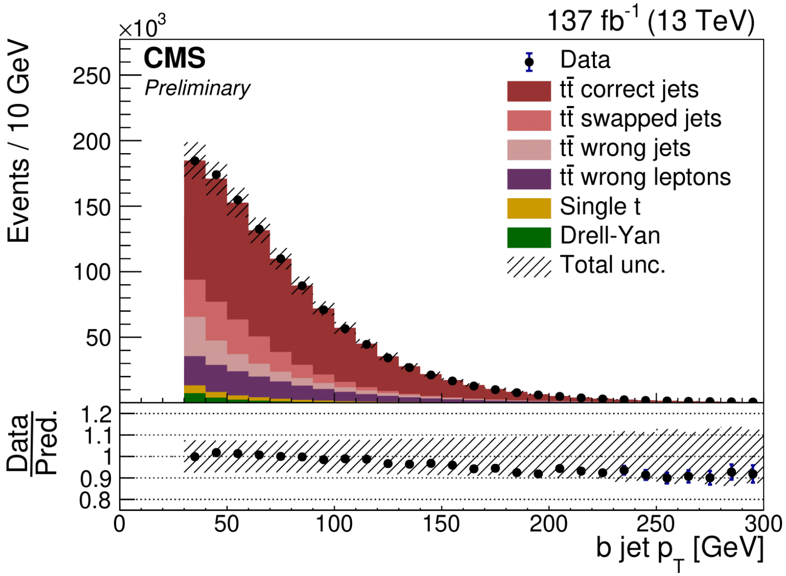
All systematic uncertainty sources

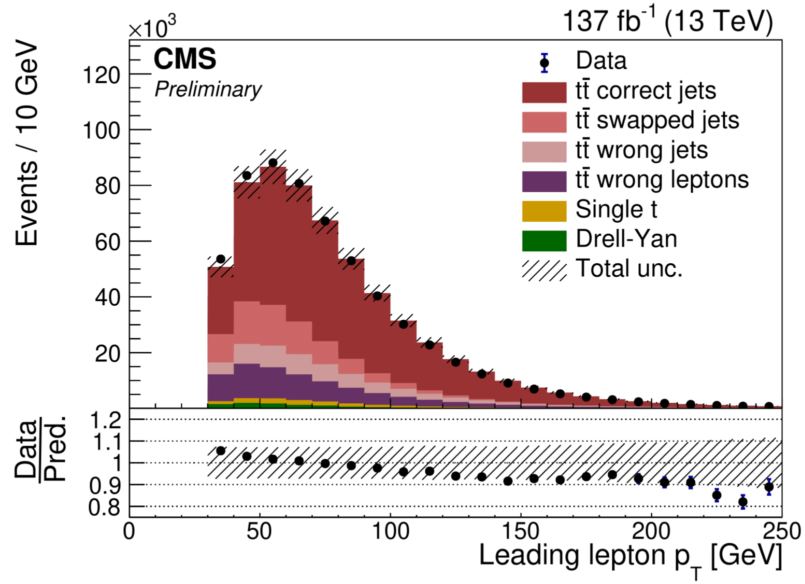
Correlation between years:	16-17	16-18	17-18
EW correction uncertainty	100%	100%	100%
background normalization	100%	100%	100%
renormalization & factorization scales	100%	100%	100%
b fragmentation, b decay	100%	100%	100%
pdf uncertainties	0%	0%	100%
ISR, FSR	50%	50%	100%
top mass *	100%	100%	100%
hdamp, MC tune *	0%	0%	0%
btag SFs, lepton SFs ^	50%	50%	50%
lumi^	30%	30%	30%
pileup	100%	100%	100%
JER	0%	0%	0%
JES (26 components)	**Various**		

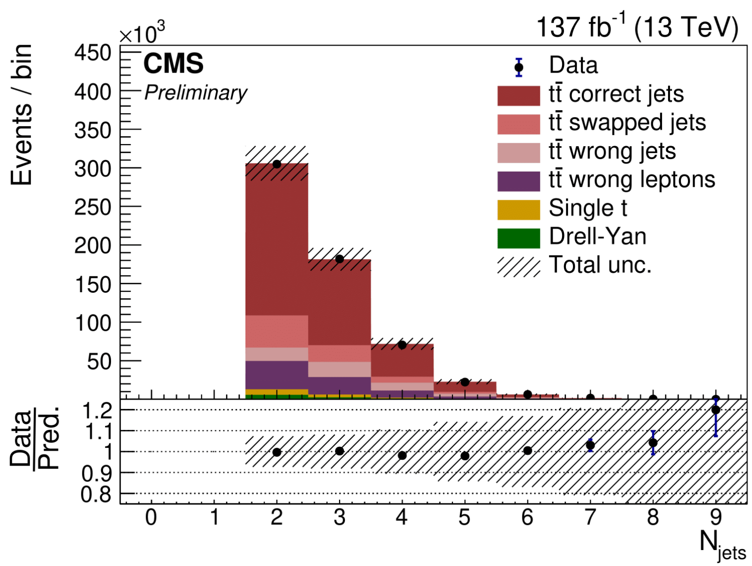
* Indicates sources with significant statistical noise.

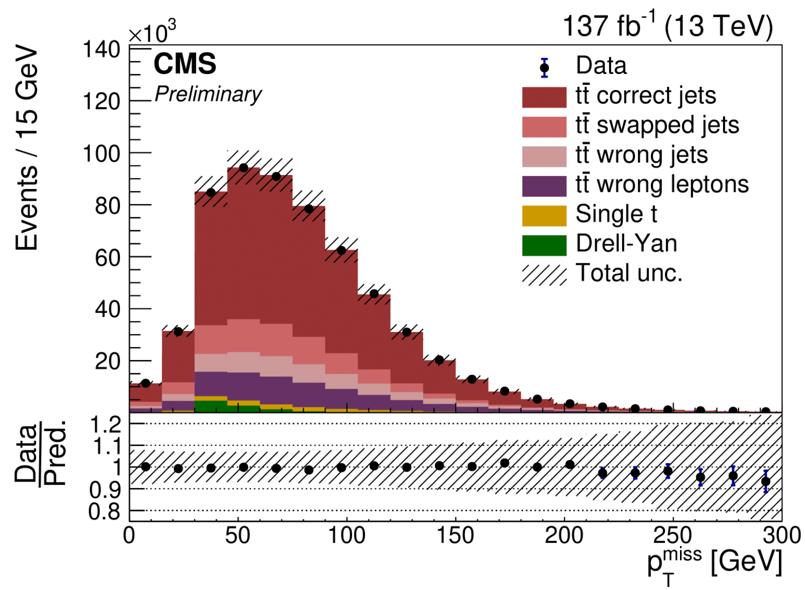
^ Indicates sources whose correlations were modified for simpler implementation.

Agreement between data and simulation





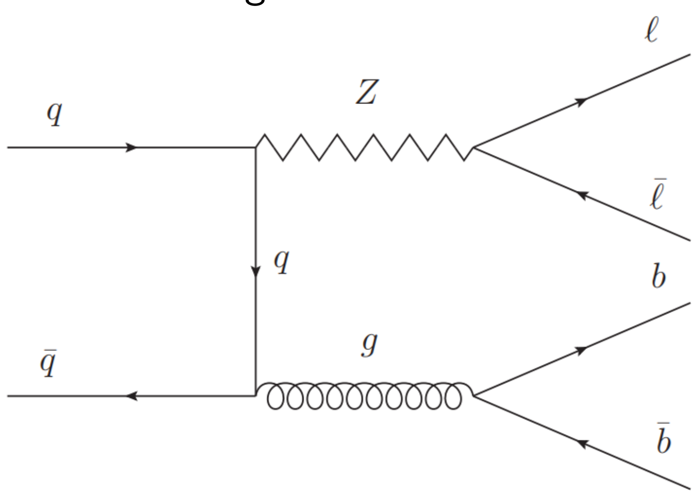




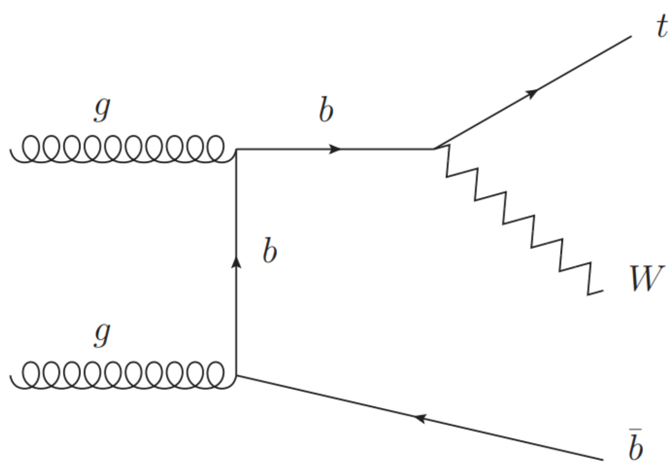
Signal and background modeling

- **Signal:** $t\bar{t}$ modeled with PowhegV2+Pythia8, reweighted via *HATHOR*
- **Backgrounds**
  - Single top production modeled with PowhegV2+Pythia8 ($\sim 2\%$ of final sample)
  - Drell-Yann decays modeled with Madgraph at LO ($\sim 2\%$ of final sample)
  - Diboson decays modeled with Pythia ($< 0.5\%$ of final sample)

Drell-Yan background event:



Single top background event:



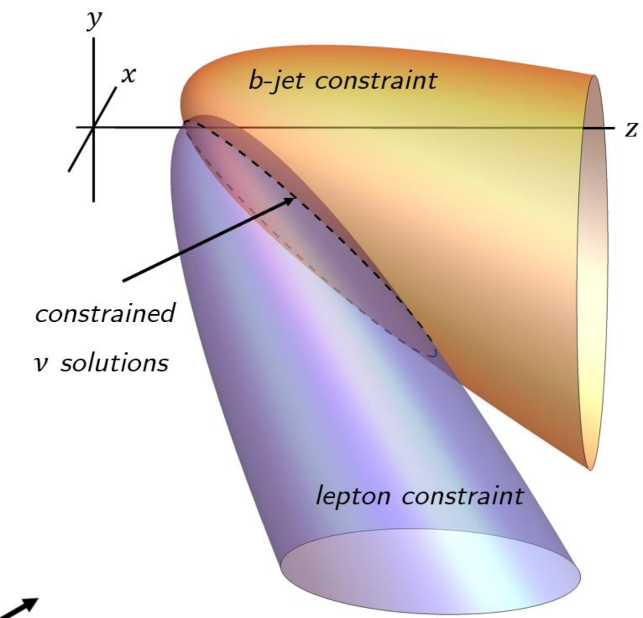
Kinematics of dilepton decays

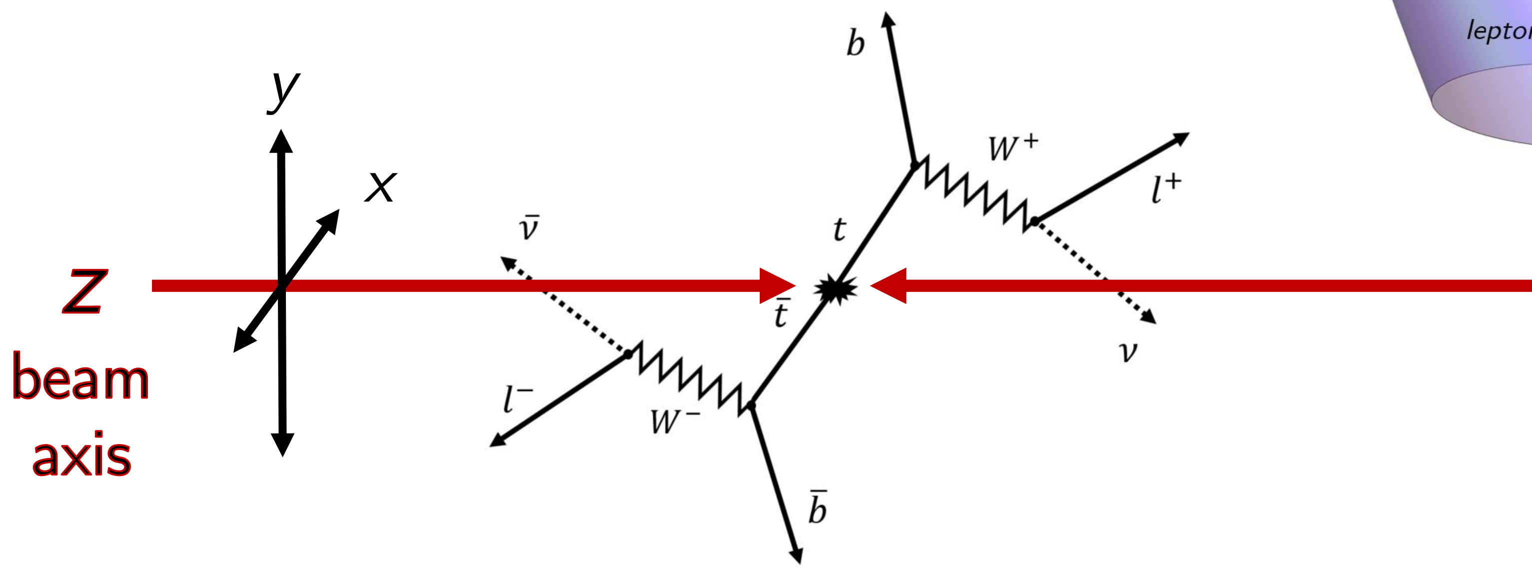
- Conservation of energy E and momentum $\boldsymbol{p}$ dictate that on-shell particles must obey the relationships

$$E_t = E_b + E_W = E_b + E_l + E_\nu$$
$$\boldsymbol{p}_t = \boldsymbol{p}_b + \boldsymbol{p}_W = \boldsymbol{p}_b + \boldsymbol{p}_l + \boldsymbol{p}_\nu$$

- While we also have the constraint on *transverse momentum* $\boldsymbol{p}_{\mathrm{T}}$:

$$\sum_{\text{all}} \boldsymbol{p}_{\mathrm{T}} = 0 = \left(\sum_{\text{observed}} \boldsymbol{p}_{\mathrm{T}} \right) + \boldsymbol{p}_{\mathrm{T}}^{\text{miss}}$$





Electroweak contribution formalism

- A generic observable $\Sigma^{t\bar{t}}$ (such as $\frac{d\sigma}{dM_{t\bar{t}}}$) is expanded in a series as

$$\Sigma^{t\bar{t}}(\alpha_s, \alpha) = \sum_{m+n\geq 2} \alpha_s^m \alpha^n \Sigma_{m,n}$$

- In this analysis the following contributions are of interest:

$$\begin{aligned} \Sigma_{\text{LO QCD}} &= \alpha_s^2 \Sigma_{2,0} \\ \Sigma_{\text{NLO QCD}} &= \alpha_s^3 \Sigma_{3,0} \\ \Sigma_{\text{NLO EW}} &= \alpha_s^2 \alpha \Sigma_{2,1} \\ \Sigma_{\text{extra EW}} &= \alpha_s^3 \alpha \Sigma_{3,1} \end{aligned}$$

- A standard EW corrections K-factor can be defined as:

$$K_{\text{EW}}^{\text{NLO}} = \frac{\Sigma_{\text{LO QCD}} + \Sigma_{\text{NLO EW}}}{\Sigma_{\text{LO QCD}}}$$

- Sometimes, $\mathcal{O}(\alpha_s^3 \alpha)$ terms are included up to one loop in the definition of "NLO EW" contributions

Electroweak contribution formalism

- We take a *multiplicative* approach to adding EW corrections:

$$\Sigma_{\mathrm{POWHEG}\times\mathrm{EW}} = K_{\mathrm{EW}}^{\mathrm{NLO}} \times \Sigma_{\mathrm{POWHEG}}$$

as opposed to a strictly *additive* approach:

$$\Sigma_{\mathrm{POWHEG+EW}} = \Sigma_{\mathrm{POWHEG}} + \Sigma_{\mathrm{NLO\ EW}}$$

- This multiplicative approach approximates the inclusion of order $\alpha_S^3\alpha$ terms. This approximation works particularly well at high values of s, where dominant EW corrections take the form of Sudakov logarithms.

- Near the production threshold, there is significant uncertainty in this method due to $\mathcal{O}(\alpha_S^3\alpha)$ contributions. To account for this, in the sensitive region where expected bin yield is increasing as a function of Y_{t}, the difference between multiplicative and additive approaches is taken as an uncertainty, suggested by HATHOR authors.

- Even using a generator that includes EW diagrams up to order $\mathcal{O}(\alpha_s^3\alpha)$ at one loop will not avoid this uncertainty (due to 2-loop contributions)

Implementation of EW uncertainty

- The full expression for the EW multiplier R_{EW} used in the fit, including the associated uncertainty, is

$$R_{\mathrm{EW}}^{\mathrm{bin}}(Y_{\mathrm{t}}, \phi) = (1 + \delta_{\mathrm{EW}}^{\mathrm{bin}}(Y_{\mathrm{t}})) \times (1 + \delta_{\mathrm{QCD}}^{\mathrm{bin}} \delta_{\mathrm{EW}}^{\mathrm{bin}}(Y_{\mathrm{t}}))^{\phi}.$$

  Nominally, $R_{\mathrm{EW}}^{\mathrm{bin}} = 1 + \delta_{\mathrm{EW}}^{\mathrm{bin}}$, but here $\delta_{\mathrm{QCD}}^{\mathrm{bin}} \delta_{\mathrm{EW}}^{\mathrm{bin}}$ represents the cross-term arising from the difference in multiplicative and additive approaches. Specifically,

$$\delta_{\mathrm{EW}} = \quad (K_{\mathrm{EW}}^{\mathrm{NLO}}(Y_{\mathrm{t}}) - 1) \longrightarrow \quad \delta_{\mathrm{EW}}^{\mathrm{bin}} = \frac{n_{\mathrm{HATHOR}}^{\mathrm{bin}} - n_{\mathrm{LO}}^{\mathrm{bin}}}{n_{\mathrm{LO}}^{\mathrm{bin}}}$$

$$\delta_{\mathrm{QCD}} = \frac{\Sigma_{\mathrm{POWHEG}} - \Sigma_{\mathrm{LO\ QCD}}}{\Sigma_{\mathrm{POWHEG}}} \longrightarrow \quad \delta_{\mathrm{QCD}}^{\mathrm{bin}} = \frac{n_{\mathrm{POWHEG}}^{\mathrm{bin}} - n_{\mathrm{LO}}^{\mathrm{bin}}}{n_{\mathrm{POWHEG}}^{\mathrm{bin}}}$$

- The uncertainty term is included only in the sensitive region where virtual Higgs exchange enhances $\mathrm{t\bar{t}}$ production.