7 TeV "AFB" analyses update: Test of top pT reweighting (TOP-13-003 and TOP-12-010)

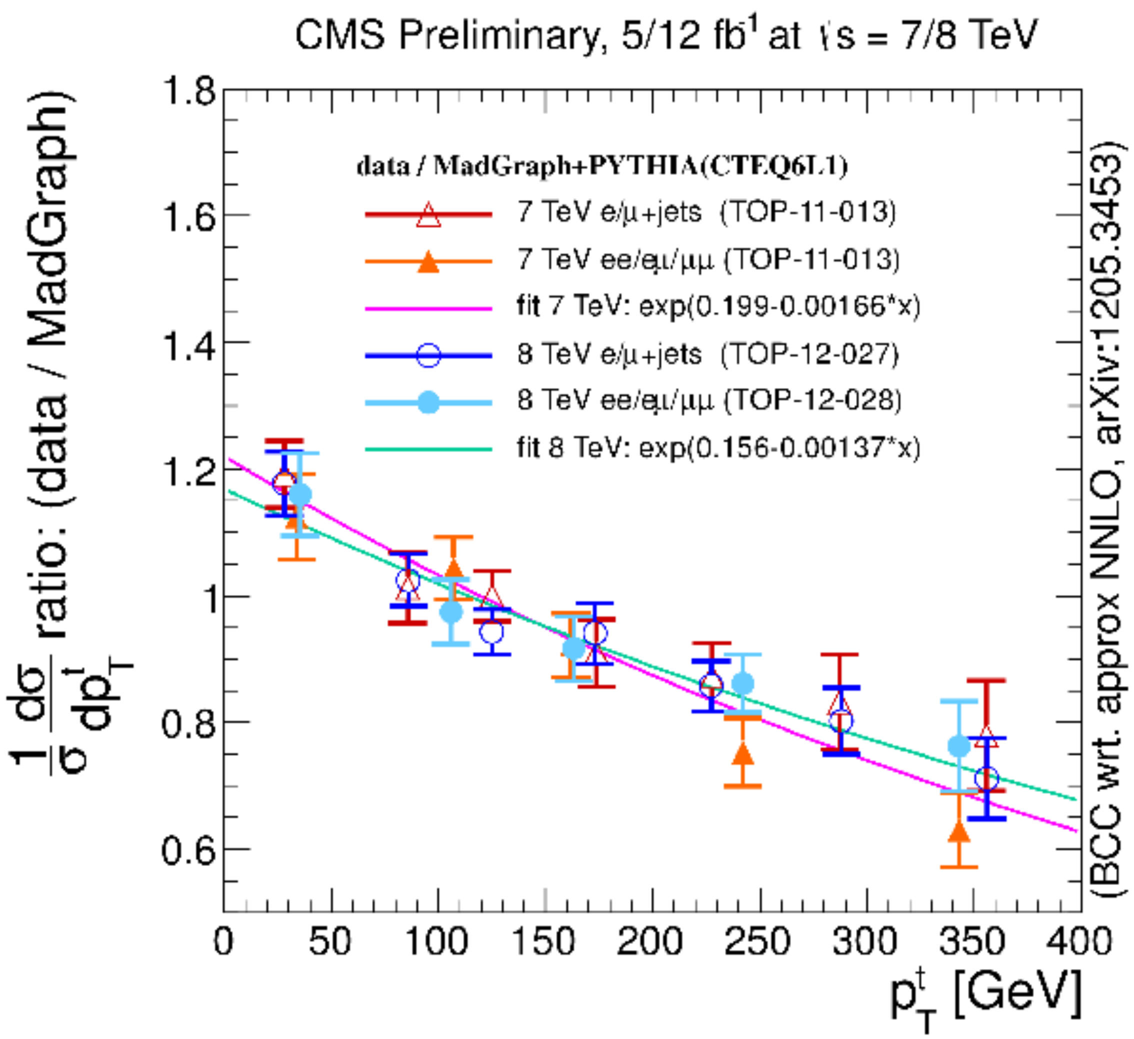
Jacob Linacre (FNAL)
for UCSD/UCSB/FNAL group

Top properties meeting
5^{th} July 2013

Introduction

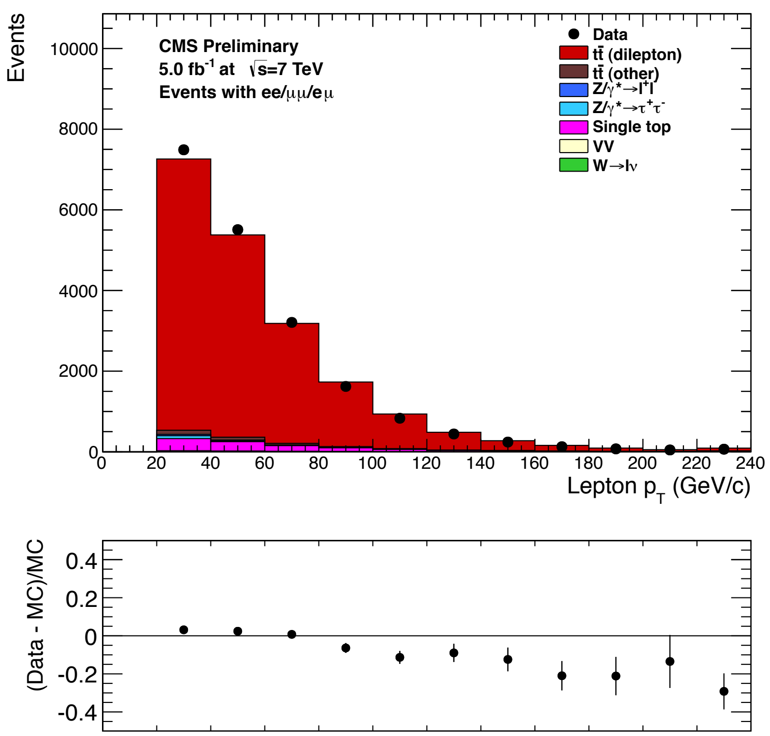
- We measure the top charge asymmetry, polarization and variables related to the spin correlation in the dilepton final state
- Top and lepton charge asymmtries: $$A_{lepC} = \frac{N(|\eta_{l^+}| > |\eta_{l^-}|) - N(|\eta_{l^+}| < |\eta_{l^-}|)}{N(|\eta_{l^+}| > |\eta_{l^-}|) + N(|\eta_{l^+}| < |\eta_{l^-}|)}$$
- Top polarization $$P_n = \frac{N(\cos(\theta_l^+) > 0) - N(\cos(\theta_l^+) < 0)}{N(\cos(\theta_l^+) > 0) + N(\cos(\theta_l^+) < 0)}$$
  - measured in the helicity basis
- Two spin correlation variables:
  - Direct (from the correlation between the + and - lepton directions)
    - $$A_{c1c2} = \frac{N(\cos(\theta_l^+) \times \cos(\theta_l^-) > 0) - N(\cos(\theta_l^+) \times \cos(\theta_l^-) < 0)}{N(\cos(\theta_l^+) \times \cos(\theta_l^-) > 0) + N(\cos(\theta_l^+) \times \cos(\theta_l^-) < 0)}$$
  - Indirect (lepton azimuthal asymmetry discriminates between correlated and uncorrelated $t\bar{t}$) - note, this is a purely leptonic variable (lab frame)
    - $$A_{\Delta\phi} = \frac{N(\Delta\phi_{l^+l^-} < \pi/2) - N(\Delta\phi_{l^+l^-} > \pi/2)}{N(\Delta\phi_{l^+l^-} < \pi/2) + N(\Delta\phi_{l^+l^-} > \pi/2)}$$

- We're using the top p_T reweighting mentioned in the news slides
  - we use the 7 TeV fitted function
- The plots on the following slides compare data and MC after our event selection, before and after top p_T reweighting
  - comparison is made for 3 different MC samples (madgraph, powheg-tauola, and MC@NLO)
  - MC is normalised to the data (shape comparison only)

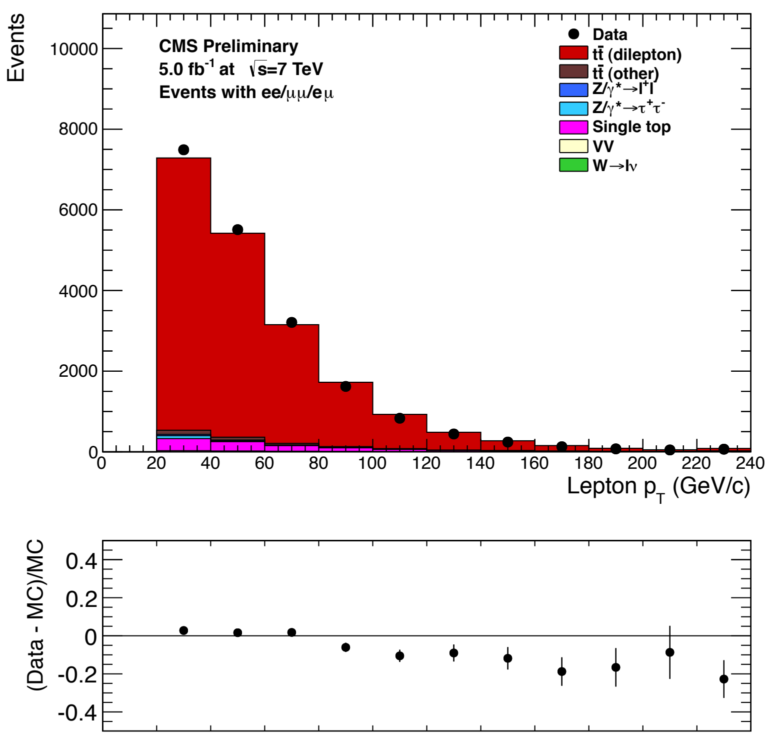


- Flick between each pair of slides to see the effect of the top p_T reweighting on the 3 different MCs

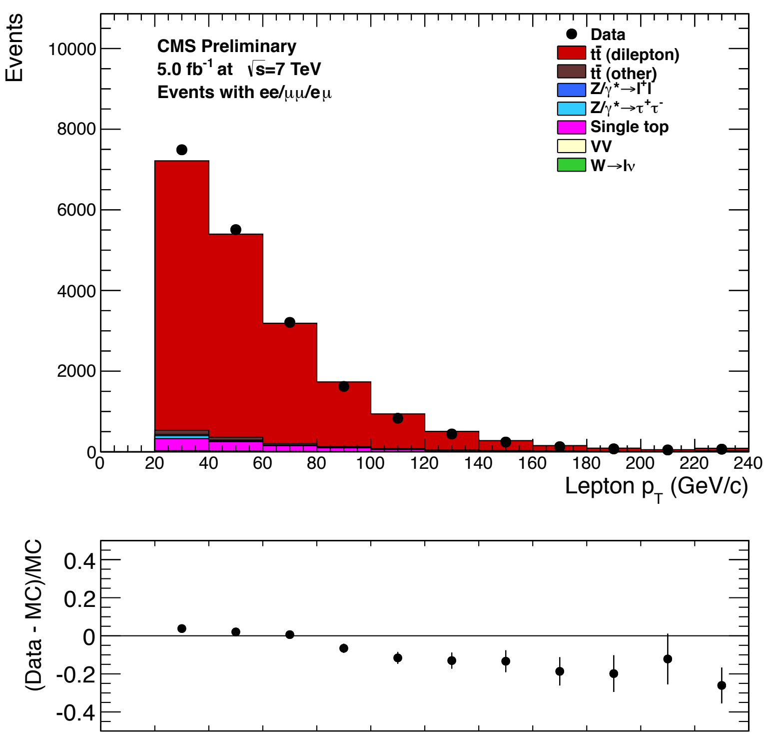
madgraph



powheg-tauola

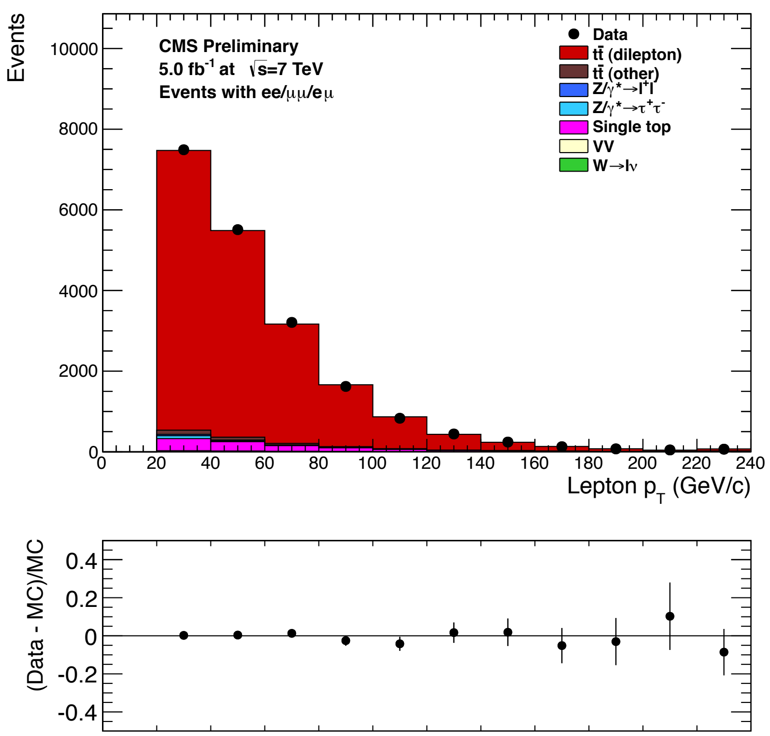


mc@nlo

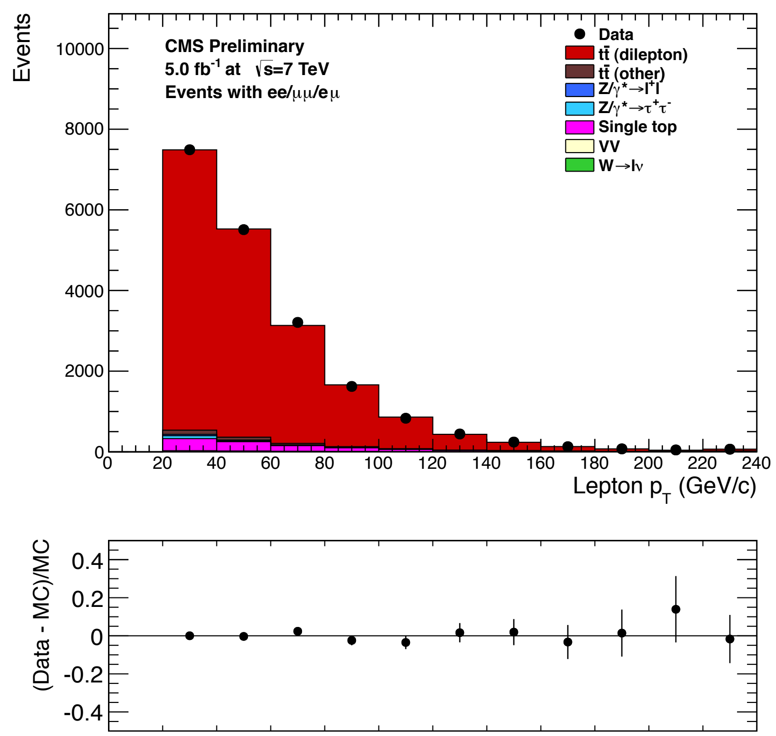


Lepton p_T, after reweighting

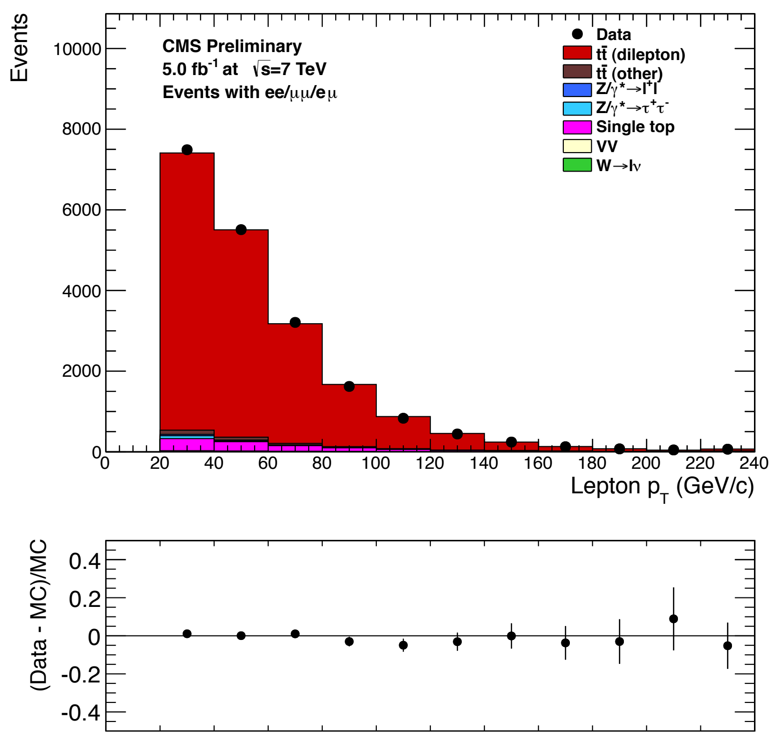
madgraph



powheg-tauola

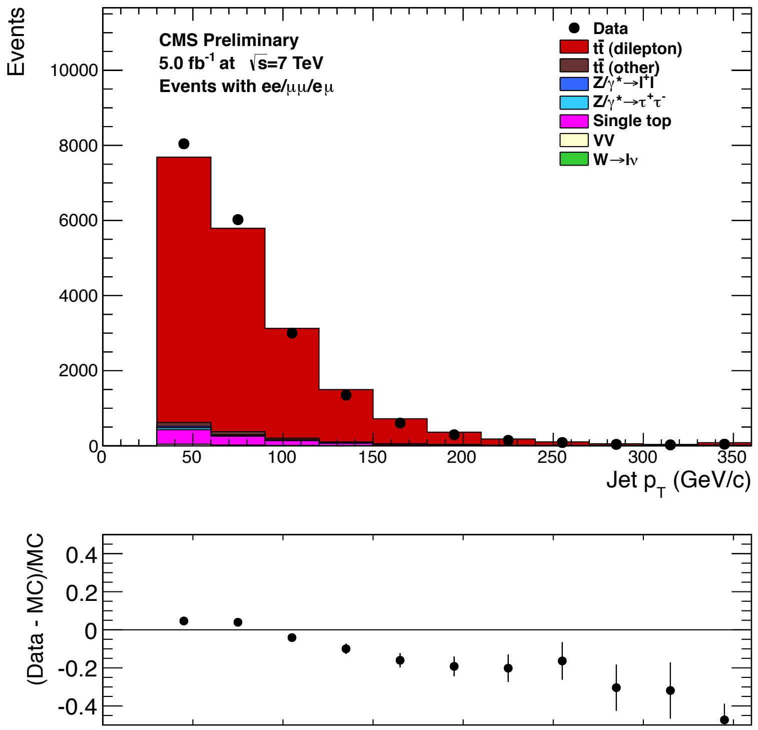


mc@nlo

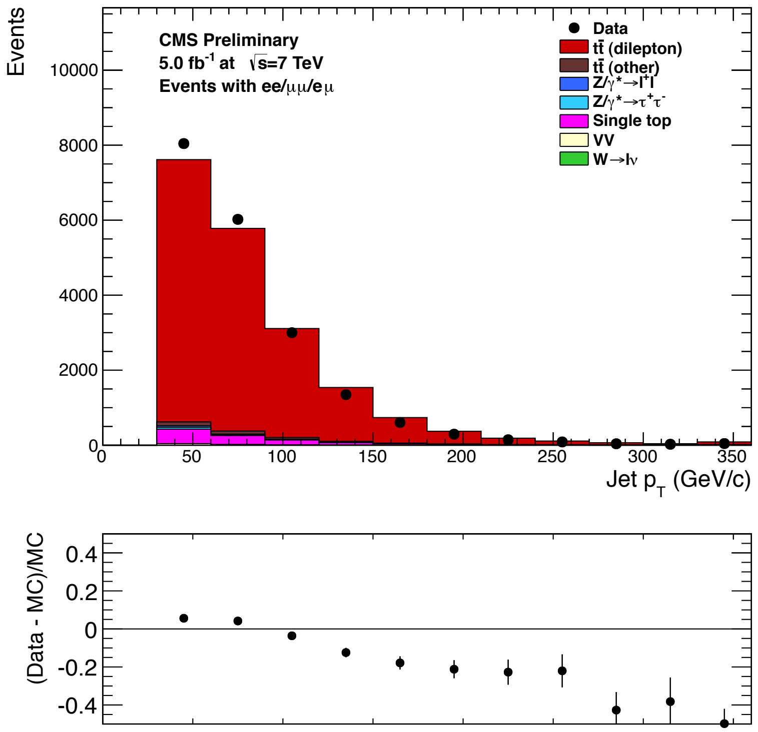


Jet p_T, before reweighting

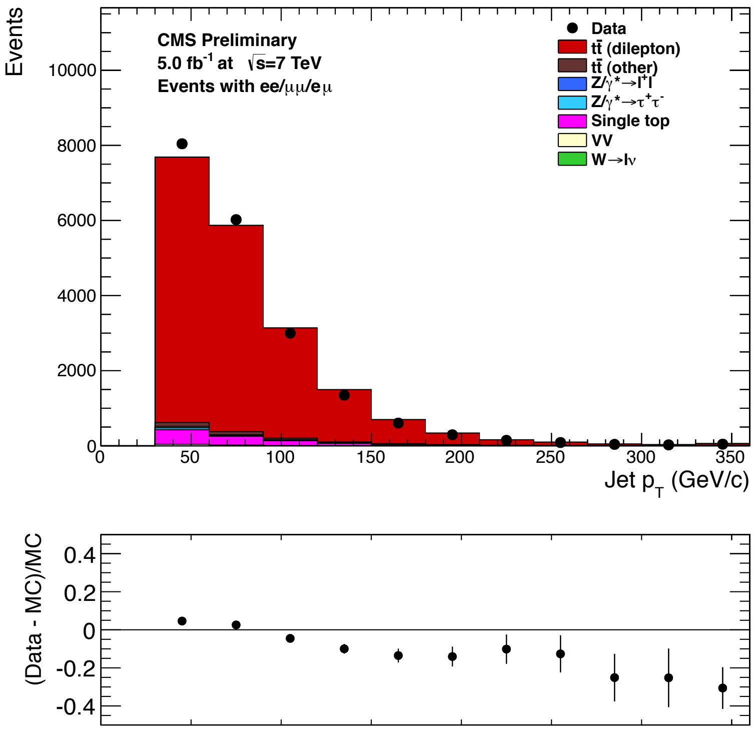
madgraph



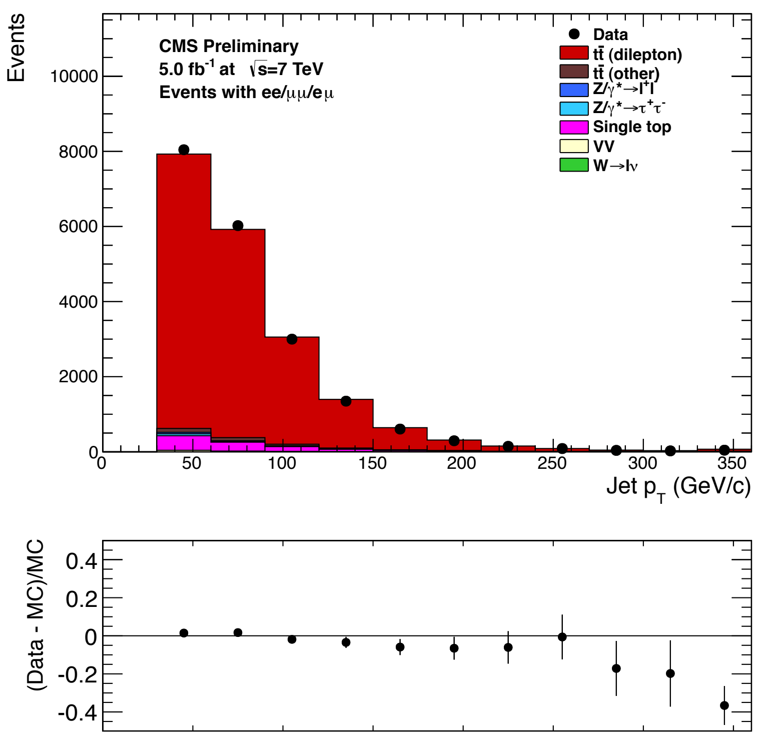
powheg-tauola



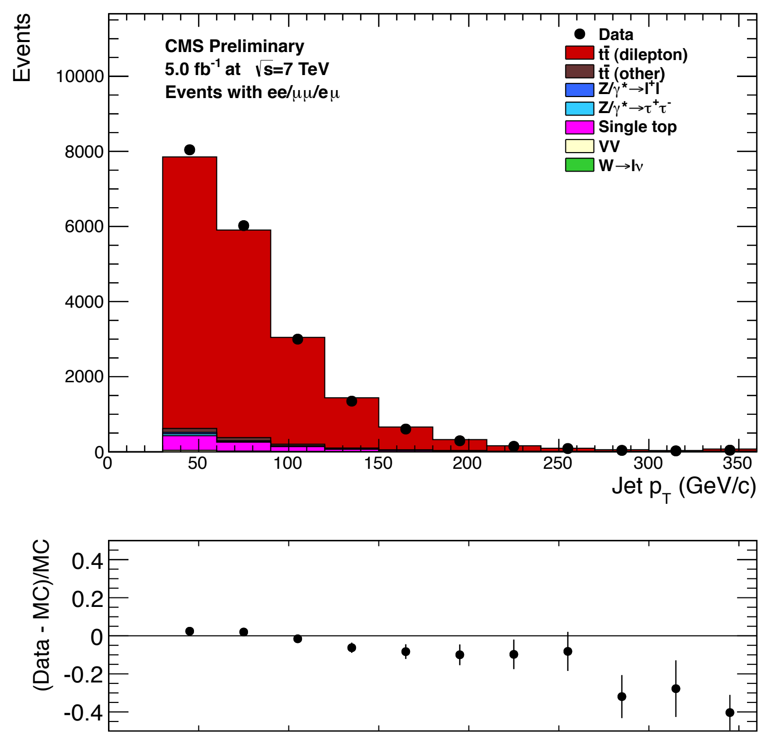
mc@nlo



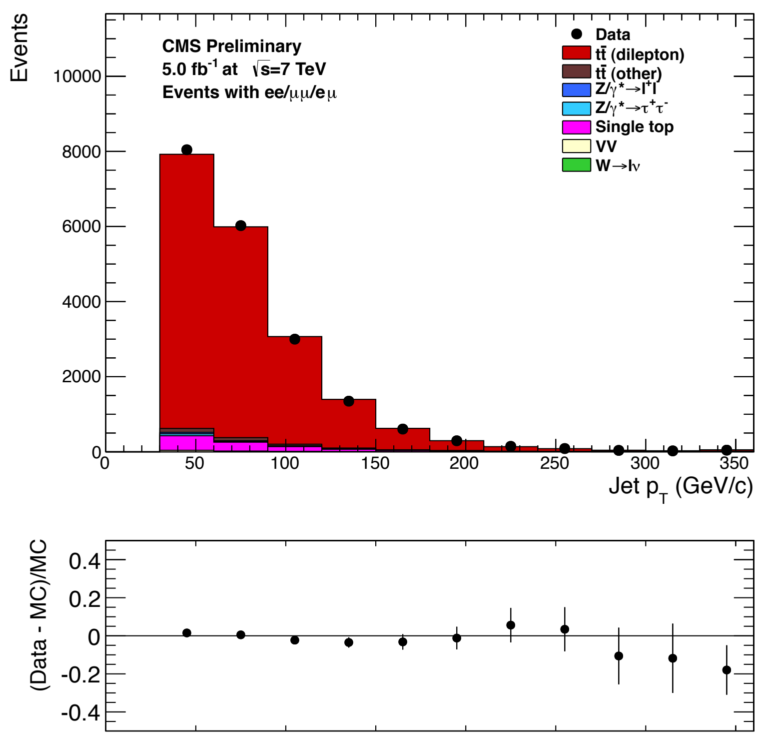
madgraph



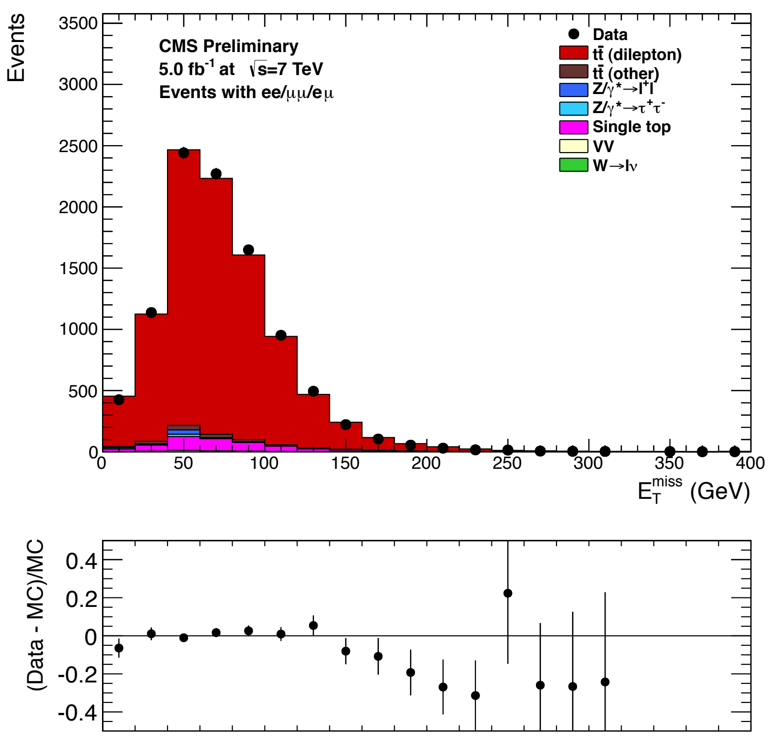
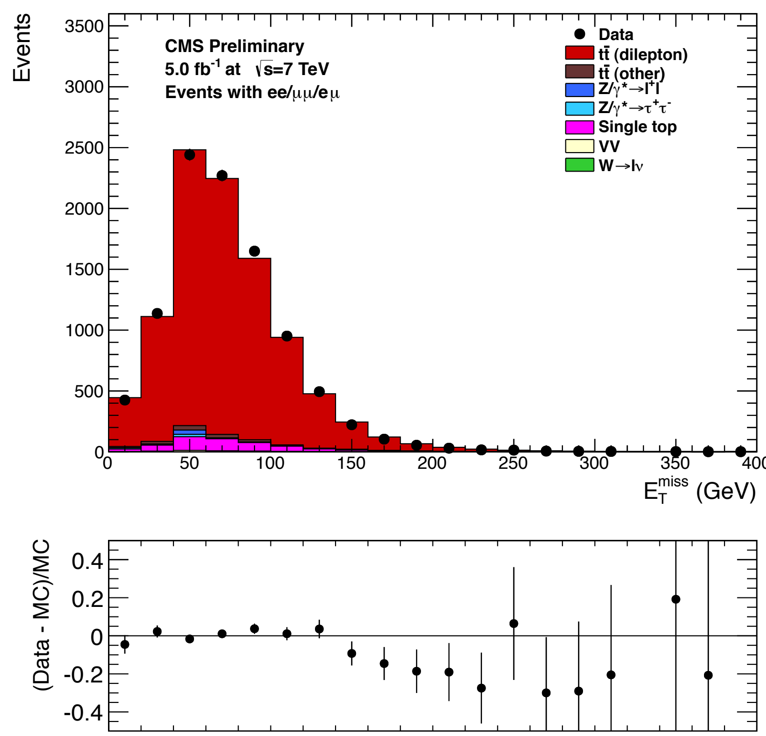
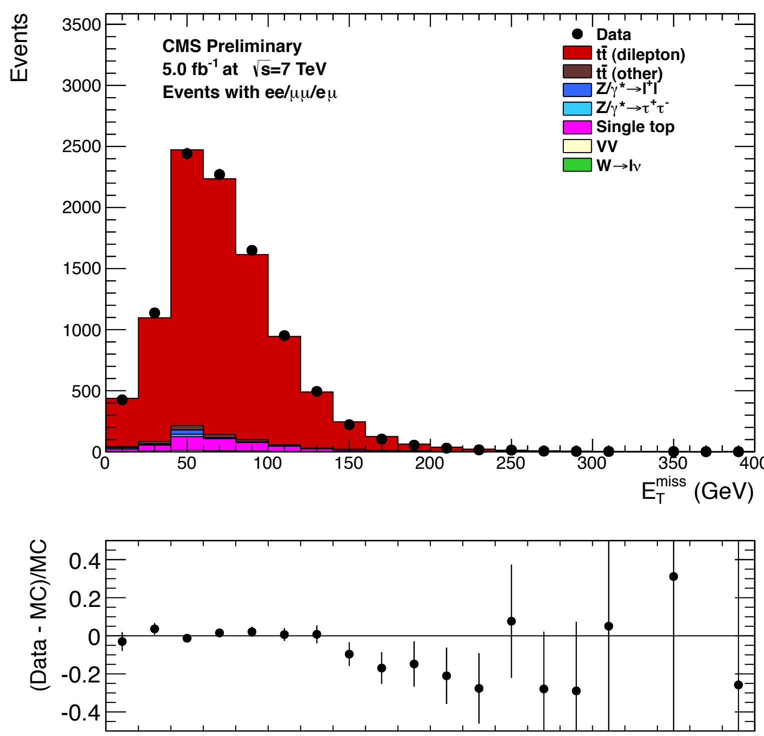
powheg-tauola



mc@nlo

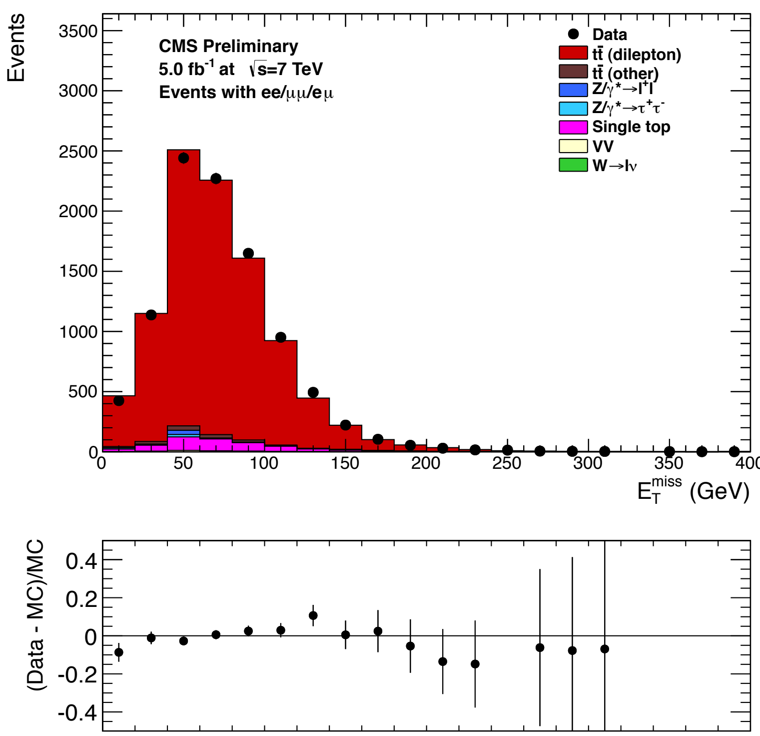


MET, before reweighting

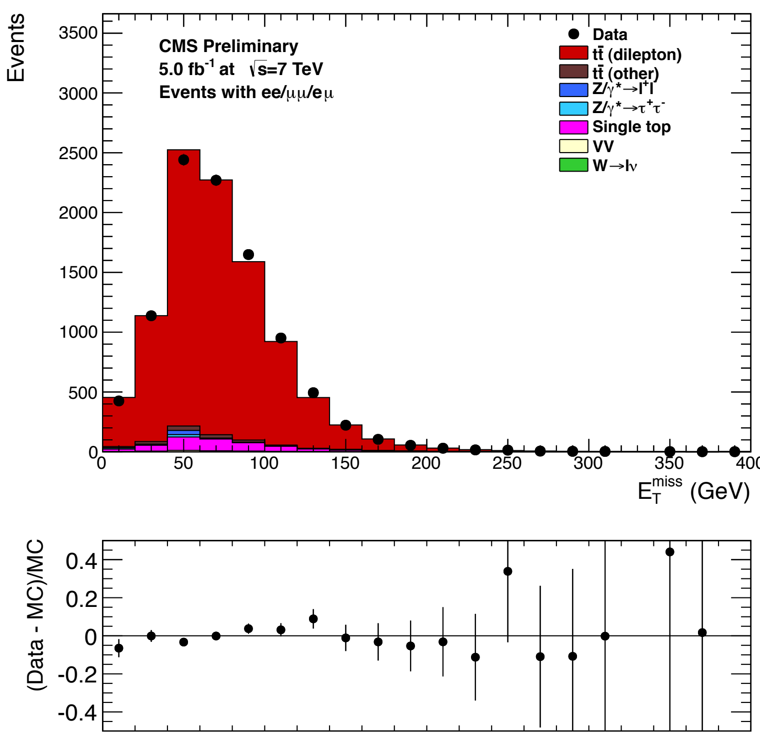
madgraph	powheg-tauola	mc@nlo
		

MET, after reweighting

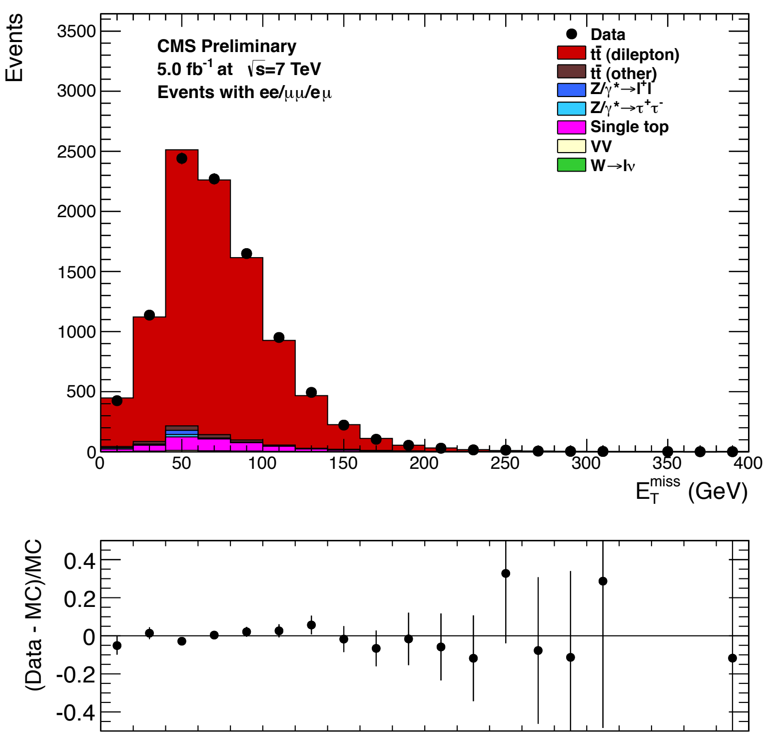
madgraph



powheg-tauola

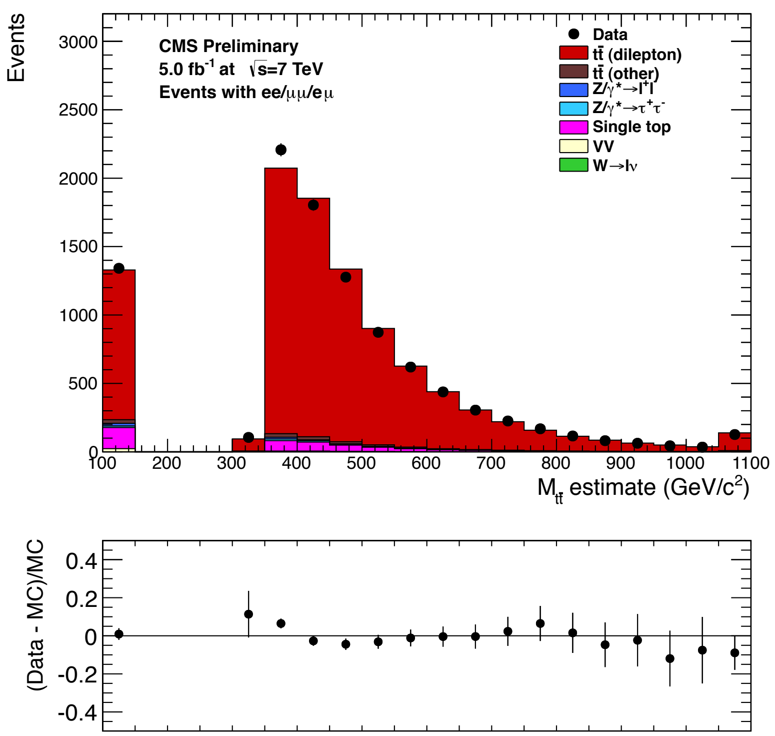


mc@nlo

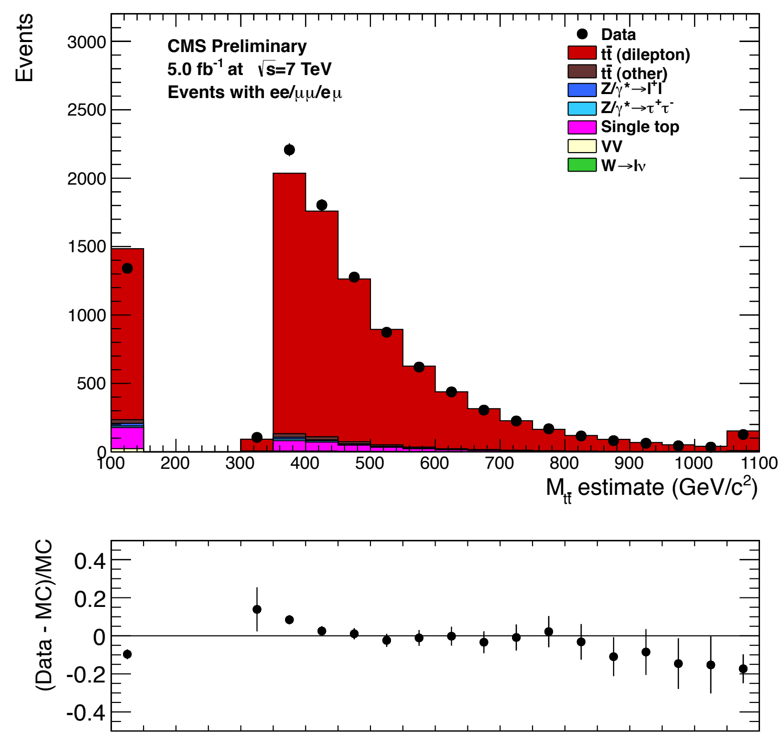


M_{ttbar}, before reweighting

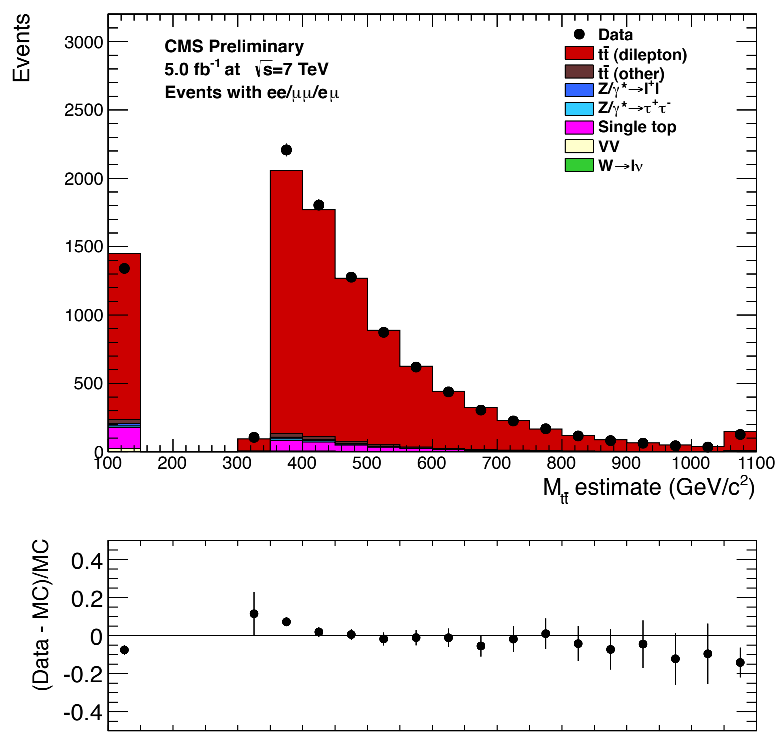
madgraph



powheg-tauola

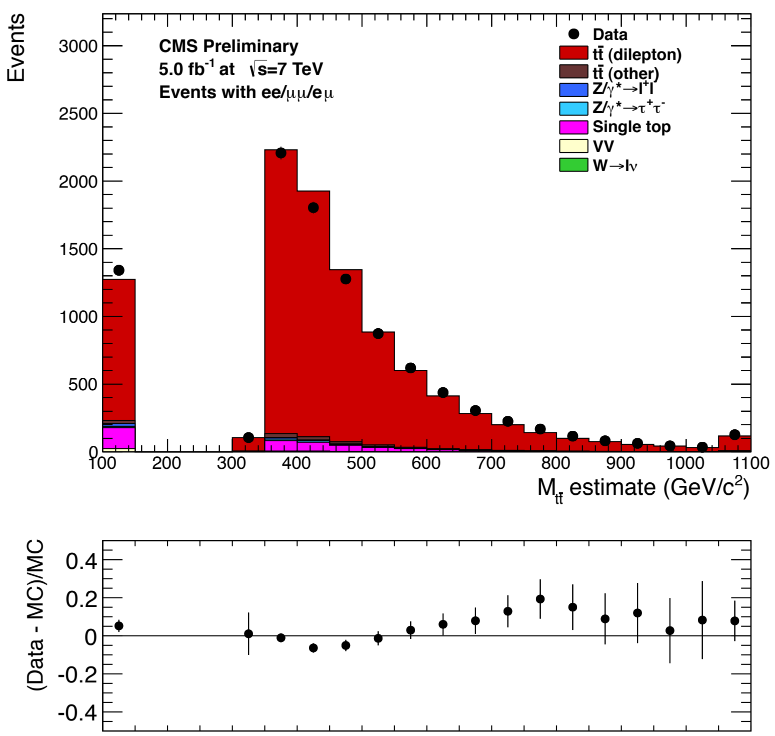


mc@nlo

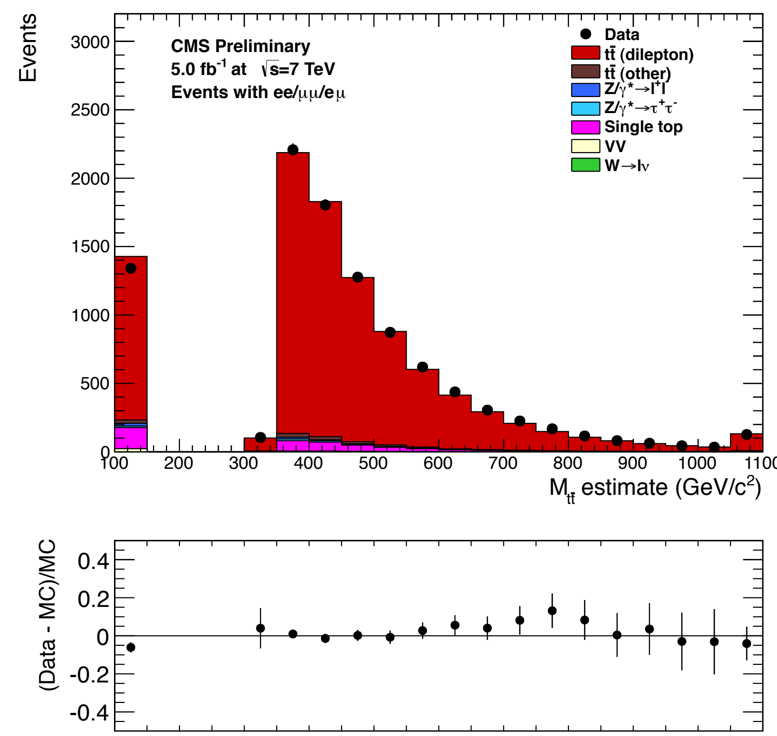


M_{ttbar}, after reweighting

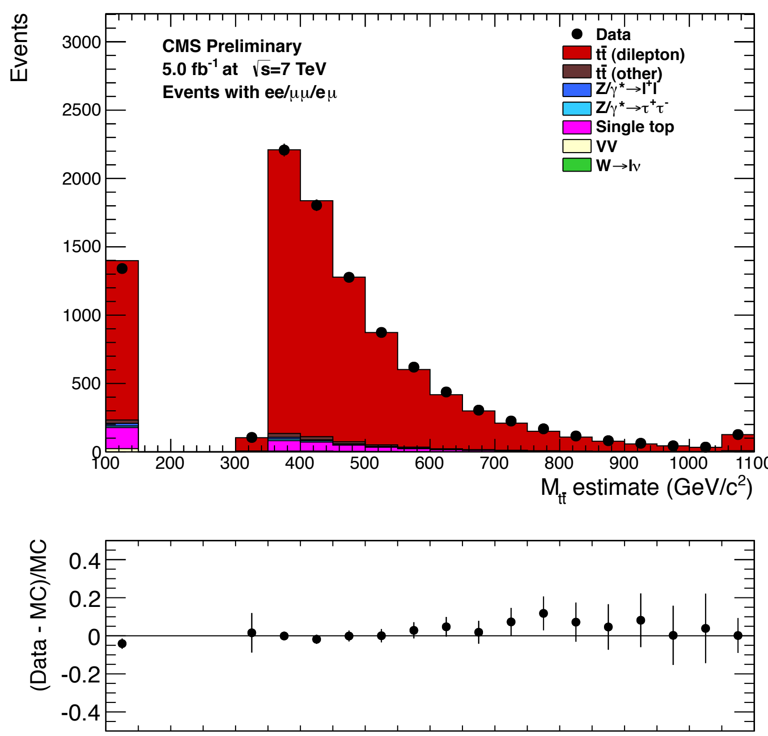
madgraph



powheg-tauola

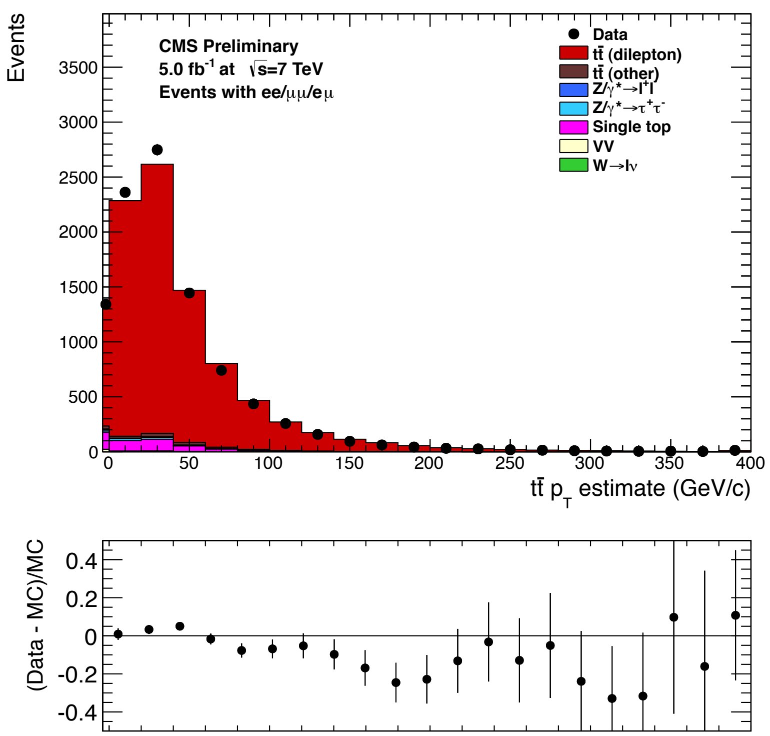


mc@nlo

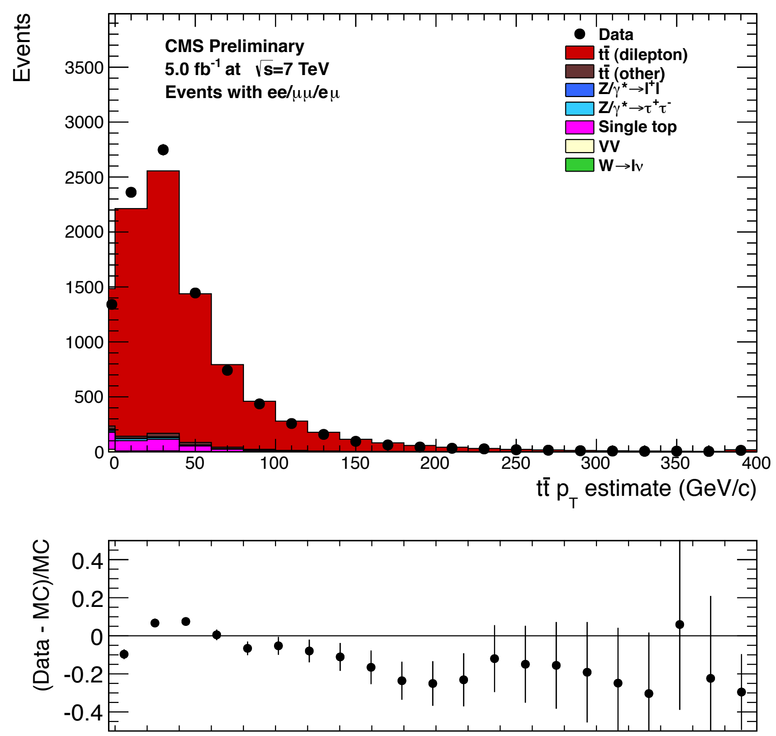


ttbar p_T, before reweighting

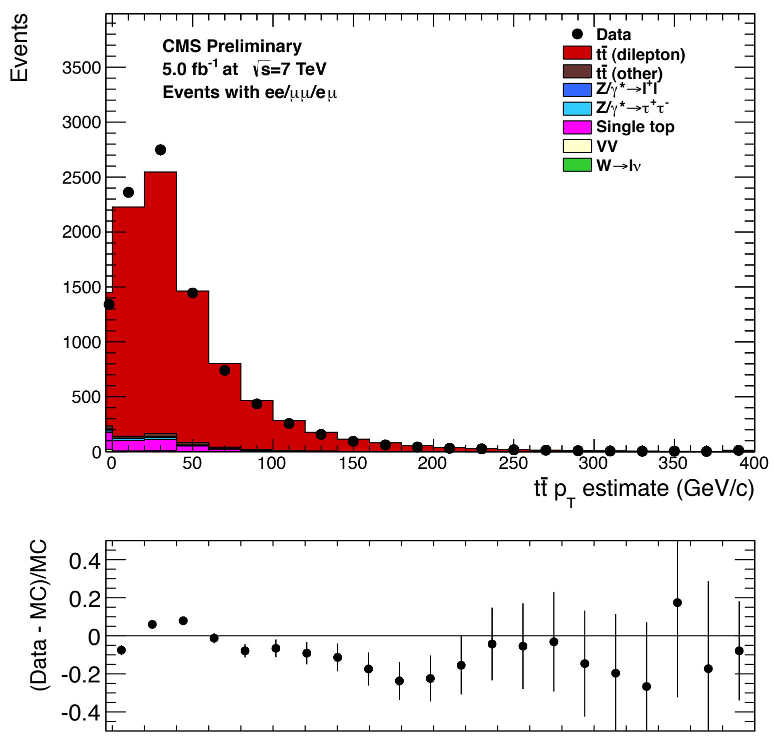
madgraph



powheg-tauola

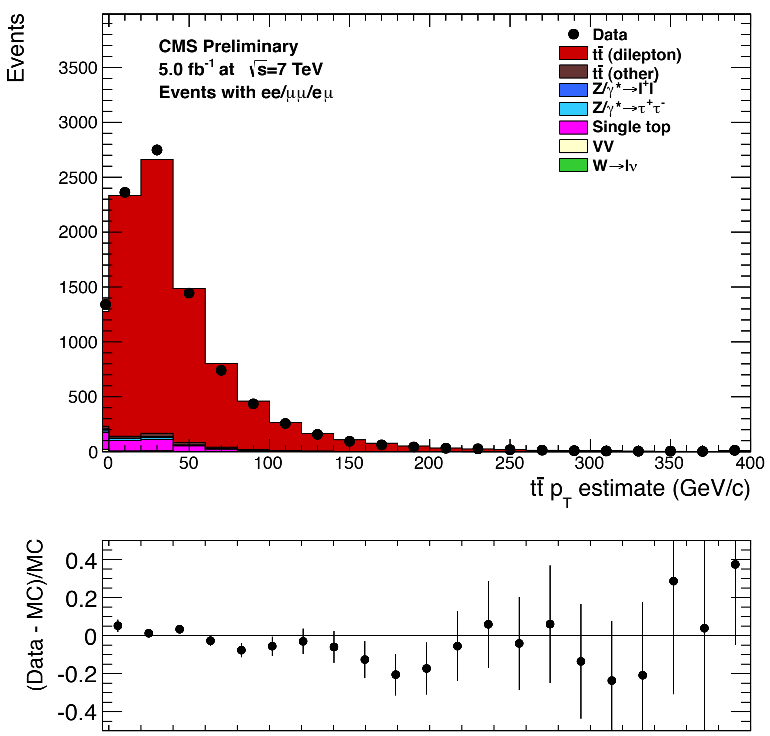


mc@nlo

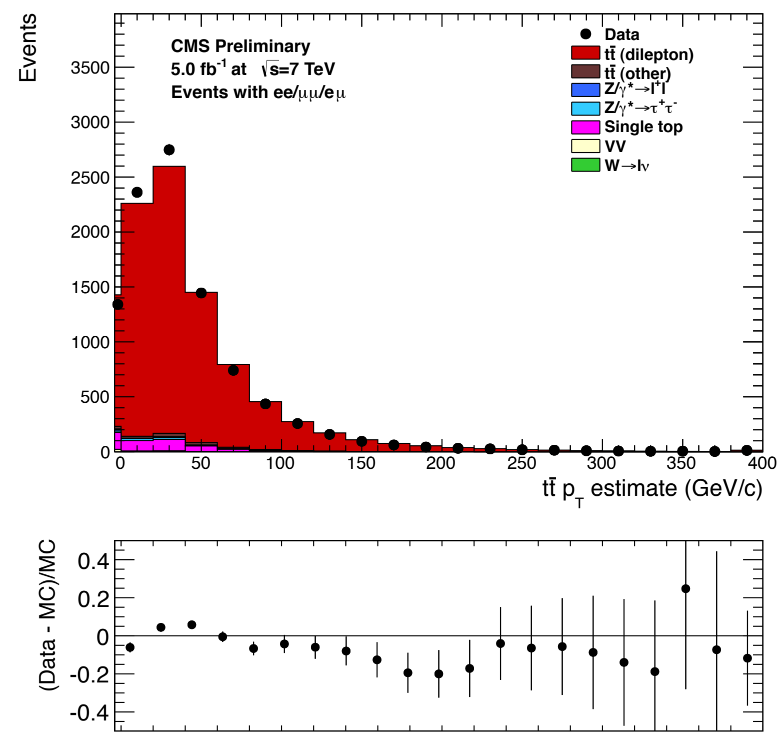


ttbar pT, after reweighting

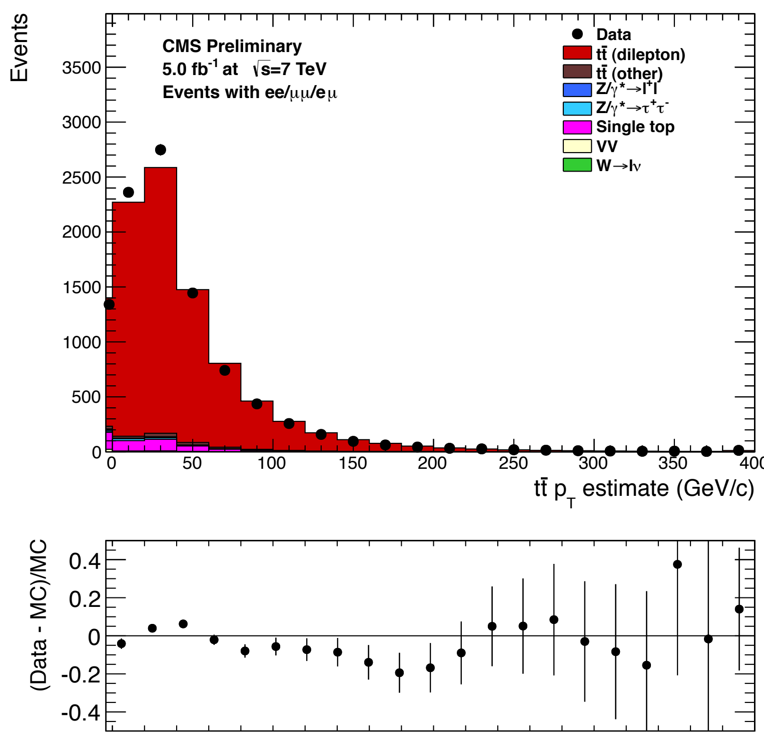
madgraph



powheg-tauola



mc@nlo



Conclusion on p_T reweighting

- Top p_T reweighting consistently improves data modeling of all three MC samples
  - of particular importance to our analysis, the reweighting significantly improves the MC modeling of $M_{ttbar,}$ which we are using in 2D unfolding to measure A_{FB} vs M_{ttbar} etc.
    - if we were not to use top p_T reweighting, we would have to consider an ad-hoc M_{ttbar} reweighting, which unlike the top p_T reweighting would be applied at reco level and thus be difficult to define in the fully inclusive phase space
  - also improves $p_{T,ttbar}$, which we are currently commissioning for 2D unfolding

Top p_T reweighting systematic

- It's clear the top p_T reweighting improves the MC modeling of the data in general
- Therefore seems sensible to use it even if the question of what it is correcting for is still open (e.g. is it a NNLO QCD correction?)
  - Kidonakis's approx NNLO differential xsec in top p_T agrees well with the data in both 7 TeV and 8 TeV, in both lepton+jets and dilepton, but apparently others have been unable to reproduce his results
  - good summary in these slides: https://indico.cern.ch/getFile.py/access?contribId=1&resId=0&materialId=slides&confId=254297
- The answer to the above question tells us how to estimate our systematic uncertainty on the top p_T reweighting:
  - known effect: only the stat uncertainty on the p_T reweighting function matters => take uncertainty directly from uncertainties on fitted function
  - unknown effect: no guarantee the reweighting improves all data/MC distributions (specifically, those we are trying to measure) => assign conservative systematic of 100% of the difference between nominal and reweighted (it can't be more wrong than this). For now we must take this systematic, but this can be changed at any time until the paper is finished.

Consequences for our measurements

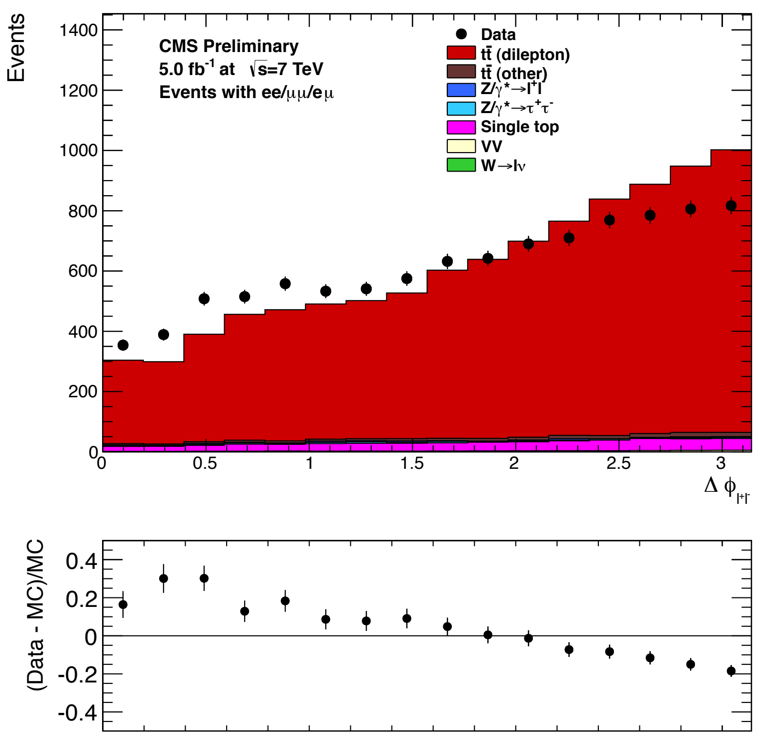
- Even the conservative systematic of 100% of the difference between nominal and reweighted is negligible for 4 of the 5 variables we are measuring. The total systematic uncertainty is only appreciably affected for lepton azimuthal asymmetry (Δphi(l^+,l^-)).
- We measure A[Δphi(l+,l-)] = 0.113 ± 0.012 (stat) (using top p_T reweighting)
- Total systematic uncertainty on A[Δphi(l+,l-)]:
  - excluding top p_T reweighting systematic: 0.006
  - with 50% top p_T reweighting systematic: 0.008 (50% is a conservative guess for the uncertainty from the fit alone. Here the top p_T syst alone is equal to 0.006, i.e. roughly the total of all other systs)
  - with 100% top p_T reweighting systematic: 0.013 (here the top p_T syst is 0.012 and completely dominant, although still only the same size as the stat uncertainty)
- Note the PAS result had a very conservative total systematic uncertainty (A[Δphi(l+,l-)] = 0.097 ± 0.015 ± 0.036), so the result is significantly improved even with the 100% top p_T reweighting systematic
  - even the most conservative top p_T reweighting systematic gives a reduction by a factor of 3 in the total systematic uncertainty

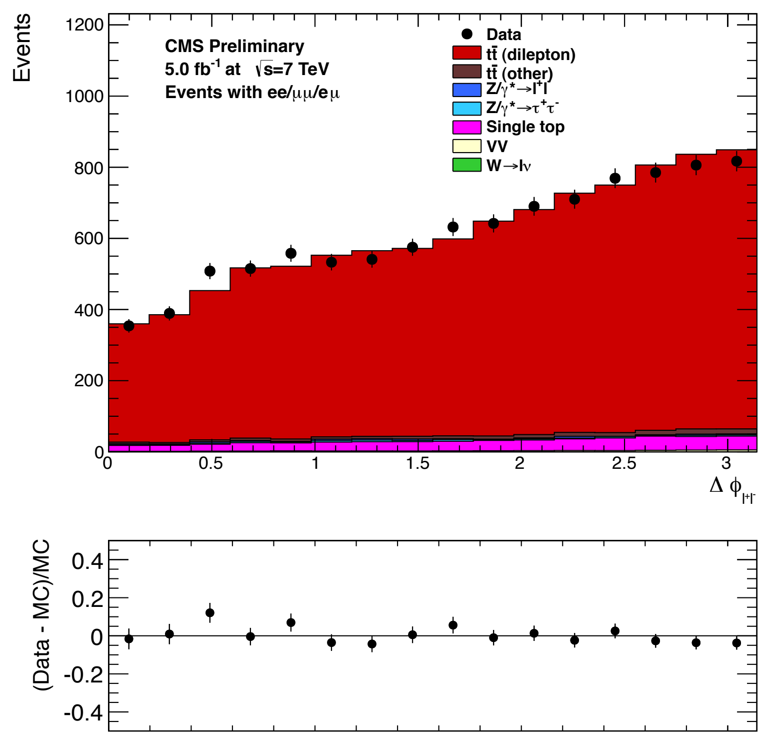
Next steps

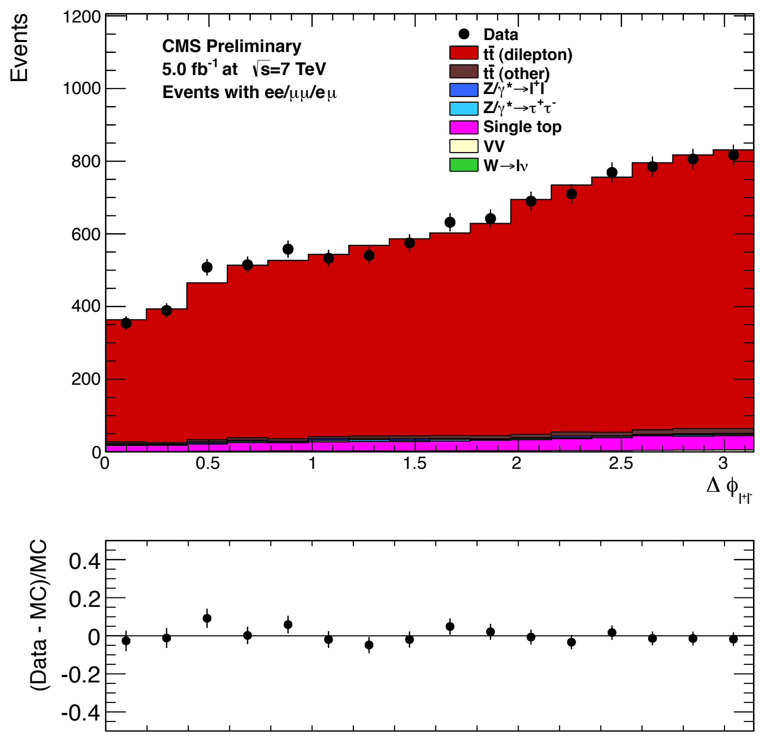
- Currently we have to view the p_T reweighting as an "unknown effect", i.e. set 100% systematics
- However, if we can get a theory prediction for Δphi at NNLO vs NLO (or some estimate of the correction), and the effect is comparable to the one we observe from p_T reweighting, we can argue for the smaller systematics
- Δphi plots before/after reweighting are on the next two slides
  - lower p_T tops have a greater degree of spin correlation, so reweighting has a significant effect
  - of course we can't use these distributions to test MC modeling because this is what we are trying to measure

Δphi(l^+,l^-), before reweighting

madgraph*	powheg-tauola	mc@nlo



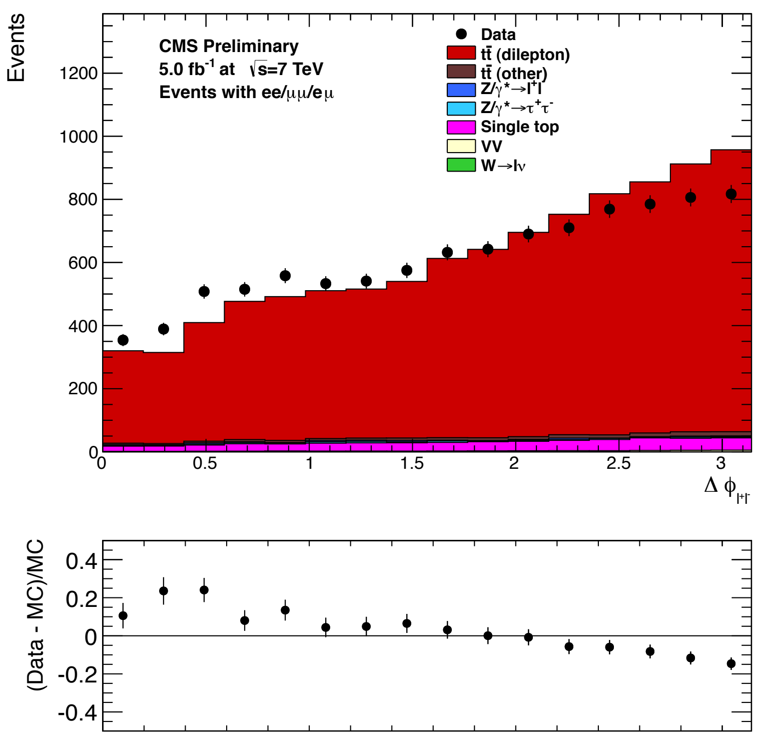




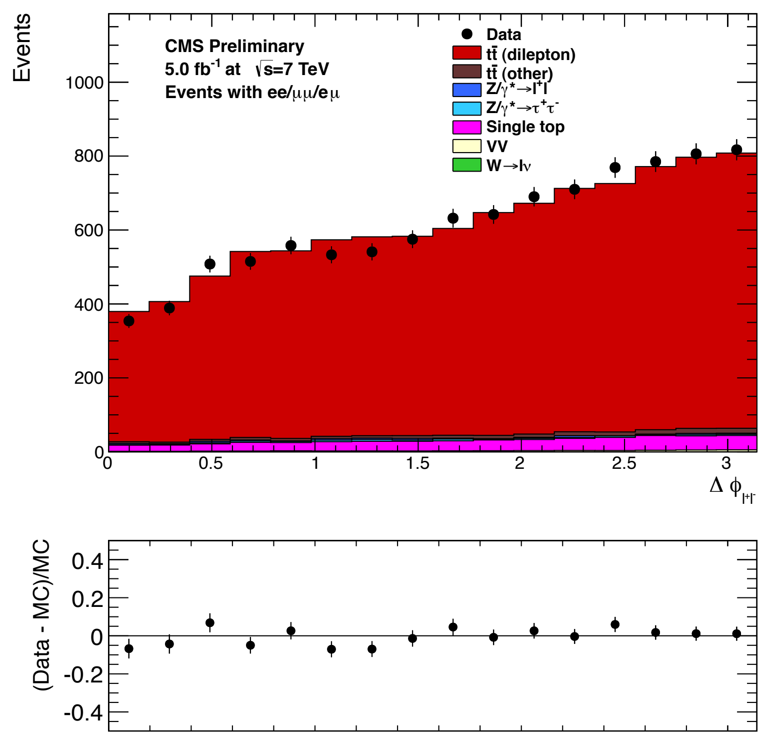
*** madgraph sample has no spin correlation**

Δphi(l$^+$,l$^-$), after reweighting

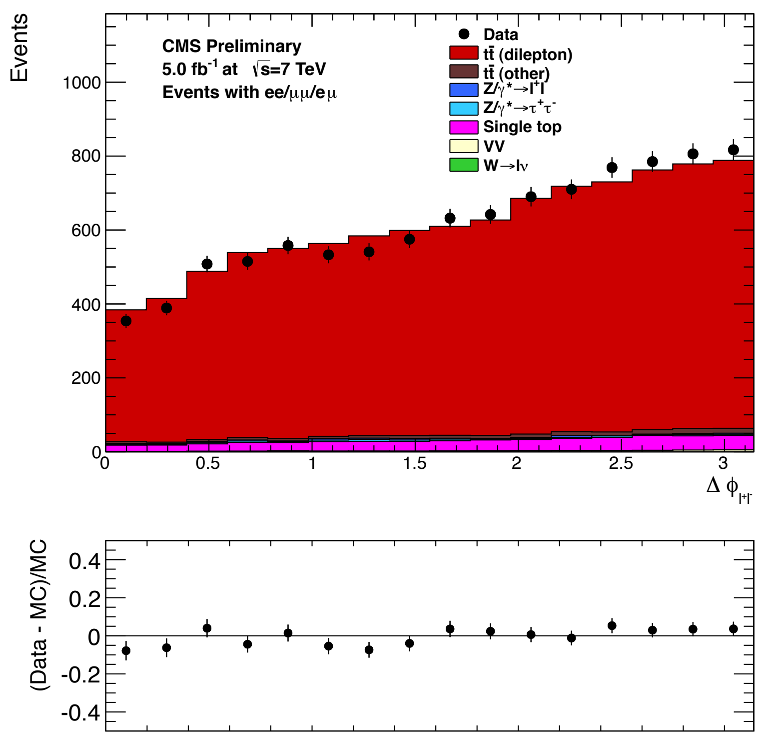
madgraph



powheg-tauola



mc@nlo



*** madgraph sample has no spin correlation**

General status update

- Data driven background predictions complete
- Systematics complete except for PDF (which is almost finished)
  - when this is done the final results for the papers will be complete
- We're currently updating the paper drafts (TOP-13-003 and TOP-12-010)

Backup

Event selection

- **Selection designed to reject events other than ttbar**
- Dilepton triggers: dimuon, dielectron or electron-muon
- 2 opposite sign isolated leptons: $p_T > 20$ GeV, $|eta| < 2.5$ (2.4) for e (μ)
- ≥ 2 pf jets with $p_T > 30$ GeV, $|eta| < 2.5$
  - loose pfjet ID (L1FastL2L3 corrected)
  - $\Delta R > 0.4$ from all leptons passing analysis selection
  - ≥ 1b tags: CSVM
- MET > 40 GeV (ee and μμ channels only)
- Z veto: $76 < m_{ll} < 106$ GeV veto (for SF leptons)
- $m_{ll} > 20$ GeV to veto low mass resonances (SF leptons)

Event Samples

- TTJets_TuneZ2_7TeV-madgraph-tauola_Summer11-PU_S4_START42_V11-v1 , 154 pb
- TTTo2L2Nu2B_7TeV-powheg-pythia6_Summer11-PU_S4_START42_V11-v1 , 16.2 pb
- /TT_TuneZ2_7TeV-mcatnlo/Fall11-PU_S6_START42_V14B-v1/AODSIM , 154 pb
- T_TuneZ2_tW-channel_7TeV-madgraph_Summer11-PU_S4_START42_V11-v1 , 7.87 pb
- T_TuneZ2_t-channel_7TeV-madgraph_Summer11-PU_S4_START42_V11-v1 , 41.92 pb
- T_TuneZ2_s-channel_7TeV-madgraph_Summer11-PU_S4_START42_V11-v1 , 3.19 pb
- Tbar_TuneZ2_tW-channel_7TeV-madgraph_Summer11-PU_S4_START42_V11-v1 , 7.87 pb
- Tbar_TuneZ2_t-channel_7TeV-madgraph_Summer11-PU_S4_START42_V11-v1 , 22.65 pb
- Tbar_TuneZ2_s-channel_7TeV-madgraph_Summer11-PU_S4_START42_V11-v1 , 1.44 pb
- WJetsToLNu_TuneZ2_7TeV-madgraph-tauola_Summer11-PU_S4_START42_V11-v1 , 31314 pb
- DYJetsToLL_TuneD6T_M-50_7TeV-madgraph-tauola_Summer11-PU_S4_START42_V11-v1 , 3048 pb
- DYToEE_M-20_CT10_TuneZ2_7TeV-powheg-pythia_Summer11-PU_S4_START42_V11-v1 , 1666 pb
- DYToMuMu_M-20_CT10_TuneZ2_7TeV-powheg-pythia_Summer11-PU_S4_START42_V11-v1 , 1666 pb
- DYToTauTau_M-20_CT10_TuneZ2_7TeV-powheg-pythia-tauola_Summer11-PU_S4_START42_V11-v1 , 1666 pb
- DYToEE_M-10To20_TuneZ2_7TeV-pythia6_Summer11-PU_S4_START42_V11-v1 , 3319.61 pb
- DYToMuMu_M-10To20_TuneZ2_7TeV-pythia6_Summer11-PU_S4_START42_V11-v1 , 3319.61 pb
- DYToTauTau_M-10To20_CT10_TuneZ2_7TeV-powheg-pythia-tauola_Summer11-PU_S4_START42_V11-v2 , 3319.61 pb
- WWJetsTo2L2Nu_TuneZ2_7TeV-madgraph-tauola_ummer11-PU_S4_START42_V11-v1, 4.783 pb
- WZJetsTo2L2Q_TuneZ2_7TeV-madgraph-tauola_Summer11-PU_S4_START42_V11-v1, 1.786 pb
- WZJetsTo3LNu_TuneZ2_7TeV-madgraph-tauola_Summer11-PU_S4_START42_V11-v1, 0.856 pb
- ZZJetsTo2L2Nu_TuneZ2_7TeV-madgraph-tauola_Summer11-PU_S4_START42_V11-v1, 0.30 pb
- ZZJetsTo2L2Q_TuneZ2_7TeV-madgraph-tauola_Summer11-PU_S4_START42_V11-v1, 1.0 pb
- ZZJetsTo4L_TuneZ2_7TeV-madgraph-tauola/_Summer11-PU_S4_START42_V11-v1, 0.076 pb
- /Wprime_SM_400_Madgraph_v2/yanjuntu-Wprime_SM_400_Madgraph_v2-f3d3f52ad6235ba5a3ccb05162c152b9/USER
- /Wprime_ttbar_600_Madgraph/yanjuntu-Wprime_ttbar_600_Madgraph-f3d3f52ad6235ba5a3ccb05162c152b9/USER
- AxigluonR_2TeV_ttbar_MadGraph_sergo-AxigluonR_2TeV_ttbar_MadGraph

- Data: May10th rereco + Prompt v4 + Aug05th rereco + Prompt v6 + 2011B Data (5.0 fb^{-1})

Triggers

- Double Electron
  - HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_CaloIsoVL
  - HLT_Ele17_CaloIdT_TrkIdVL_CaloIsoVL_TrkIsoVL_Ele8_CaloIdT_TrkIdVL_CaloIsoVL_TrkIsoVL
  - HLT_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL
- Double Muon
  - HLT_DoubleMu7
  - HLT_Mu13_Mu7
  - HLT_Mu13_Mu8
  - HLT_Mu17_Mu8
- Electron Muon
  - HLT_Mu17_Ele8_CaloIdL
  - HLT_Mu8_Ele17_CaloIdL
  - HLT_Mu17_Ele8_CaloIdT_CaloIsoVL
  - HLT_Mu8_Ele17_CaloIdT_CaloIsoVL

For the high p_T dilepton triggers, the efficiencies listed in Table 1, Table 2, Table 3 and Table 4 are applied to ee, $\mu\mu$ and $e\mu$ Monte Carlo Events. Details of the measurement of the trigger efficiencies are described in [12].

Table 1: The efficiency of the leading leg requirement for the double electron trigger, averaged over the full 2011 data.

Measurement	$0.0 \leq \|\eta\| < 1.5$	$1.5 \leq \|\eta\| < 2.5$
$20 \leq p_T \leq 30$	0.9849 ± 0.0003	0.9774 ± 0.0007
$p_T > 30$	0.9928 ± 0.0001	0.9938 ± 0.0001

Table 2: The efficiency of the trailing leg requirement for the double electron trigger, averaged over the full 2011 data.

Measurement	$0.0 \leq \|\eta\| < 1.5$	$1.5 \leq \|\eta\| < 2.5$
$20 \leq p_T \leq 30$	0.9923 ± 0.0002	0.9953 ± 0.0003
$p_T > 30$	0.9948 ± 0.0001	0.9956 ± 0.0001

Table 3: The efficiency of the leading leg requirement for the double muon trigger, averaged over the full 2011 data.

Measurement	$0.0 \leq \|\eta\| < 0.8$	$0.8 \leq \|\eta\| < 1.2$	$1.2 \leq \|\eta\| < 2.1$	$2.1 \leq \|\eta\| < 2.4$
$20 \leq p_T \leq 30$	0.9648 ± 0.0007	0.9516 ± 0.0013	0.9480 ± 0.0009	0.8757 ± 0.0026
$p_T > 30$	0.9666 ± 0.0003	0.9521 ± 0.0005	0.9485 ± 0.0004	0.8772 ± 0.0012

Table 4: The efficiency of the trailing leg requirement for the double muon trigger, averaged over the full 2011 data.

Measurement	$0.0 \leq \|\eta\| < 0.8$	$0.8 \leq \|\eta\| < 1.2$	$1.2 \leq \|\eta\| < 2.1$	$2.1 \leq \|\eta\| < 2.4$
$20 \leq p_T \leq 30$	0.9655 ± 0.0007	0.9535 ± 0.0013	0.9558 ± 0.0009	0.9031 ± 0.0023
$p_T > 30$	0.9670 ± 0.0003	0.9537 ± 0.0005	0.9530 ± 0.0004	0.8992 ± 0.0011

Lepton selections

Electron selection

- p_T > 20 GeV; |eta| < 2.5
- VBTF90 (cuts tightened to match CaloId+TrkIdVL HLT requirements)
- d0 (PV) < 0.04 cm, dz (PV) < 1 cm --calculated w.r.t. 1st good DA PV
- no muon $\Delta R < 0.1$
- <=1 miss hits, |dist| < 0.02 cm and < 0.02, CMS AN-2009-159
- Veto electrons with a supercluster in the transition region (1.44 <|eta| < 1.56)
- iso/p_T < 0.15 (EB pedestal subtraction 1 GeV, no fastjet correction)
- ecaliso/p_T < 0.2

Muon selection

- p_T > 20 GeV; |eta| < 2.4
- global and tracker muon
- $chi^2/ndf < 10$
- nValidHits > 10 -- to be updated to frac of validHits
- valid StandAloneHits > 0
- d0 (PV) < 0.02 cm, dz (PV) < 1 cm --calculated w.r.t. 1st good DA PV
- (p_T)/p_T < 0.1
- iso/p_T < 0.15 (no fastjet correction)

Preselection Yields (5.0 fb^{-1})

Sample	ee	$\mu\mu$	$e\mu$	all
ttdil	1535.60 ± 9.82	1813.86 ± 10.31	5747.85 ± 18.69	9097.31 ± 23.50
ttotr	39.74 ± 1.63	4.06 ± 0.46	93.09 ± 2.41	136.88 ± 2.94
wjets	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
DYee	16.85 ± 3.28	0.00 ± 0.00	0.00 ± 0.00	16.85 ± 3.28
DYmm	0.00 ± 0.00	22.96 ± 3.66	3.80 ± 1.60	26.76 ± 3.99
DYtautau	13.35 ± 2.92	6.59 ± 1.94	31.22 ± 4.21	51.16 ± 5.48
VV	8.27 ± 0.44	10.20 ± 0.47	27.90 ± 0.81	46.37 ± 1.03
tw	72.54 ± 2.11	86.77 ± 2.23	289.37 ± 4.20	448.68 ± 5.20
Total MC	1686.35 ± 11.10	1944.43 ± 11.35	6193.23 ± 19.84	9824.00 ± 25.41
Data	1631.00 ± 40.39	1964.00 ± 44.32	6229.00 ± 78.92	9824.00 ± 99.12

Uncertainties are statistical only

- MC events are weighted to match trigger efficiency, b tagging efficiency, and number of vertices distribution in data
- We use MC@NLO for the $t\bar{t}$ component
  - normalized so that total MC yield matches data
  - $t\bar{t} \to \ell^+\ell^-$ contributes 92% of the total yield
- Comparison plots on next slide

Background estimation

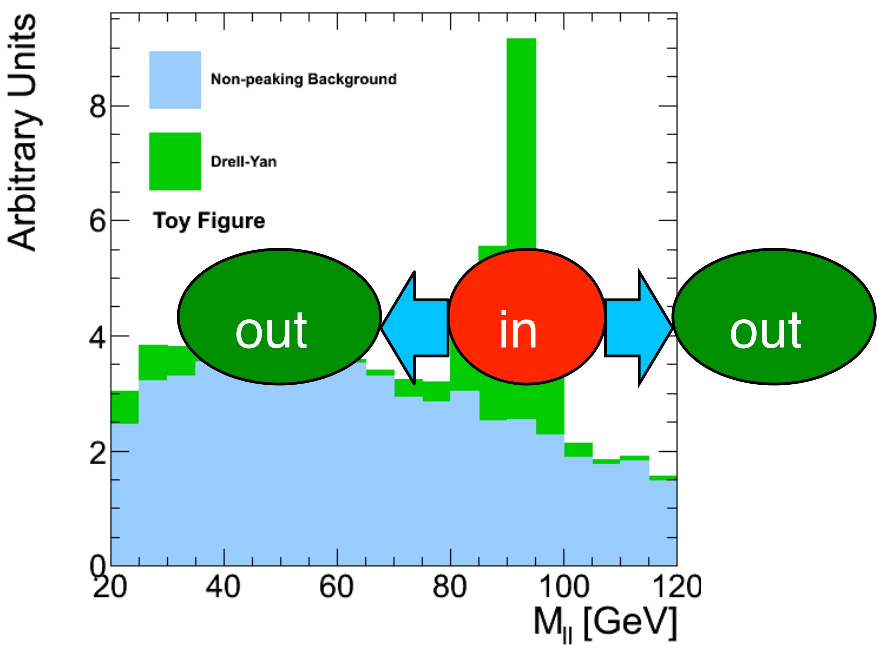
- We use the MC from the previous slides to estimate the background
- We make cross-checks for the DY and fake components using data-driven methods, and find reasonable agreement
- We then assign an appropriate background normalization systematic

Data-driven BG estimates: DY

- Estimate ee and μμ Drell-Yan using the method in CMS AN-2009-023:

  $R_{out/in}$ method

- Use data in Z peak to predict DY yields in the signal region by propagating via the MC ratio out/in-peak



Z-peak to signal region ratio from MC, verified in data

opposite-flavor events in Z peak

$$N_{\rm out}^{\ell\ell,{\rm exp}} = R_{\rm out/in}^{\ell\ell}\left(N_{\rm in}^{\ell\ell} - N_{\rm in}^{\rm non-Z}\right)$$

same-flavor events in Z peak

- Estimate for pre-selection region: 45.6 ± 6.8 (stat+syst) events
  - consistent with MC prediction of 39.8 ± 4.9 events

Data-driven BG estimates: Fakes

- Estimate contribution from fake leptons using the data-driven tight-to-loose method described in CMS AN-2010/257
  - measure tight-to-loose fake rates as a function of lepton P_T and eta
  - estimate number of fakes in data based on number of fakeable object (FOs). Weight each lepton+FO event by:
  - use MC to account for signal contamination in the FO sample
  - fake background primarily from ttbar- decaying to lepton+jets

$$\varepsilon_{fake}(p_T,\eta) = \frac{N_{pass\ tight}(p_T,\eta)}{N_{loose}(p_T,\eta)}$$

$$w_i = \frac{\epsilon_{\text{fake}}(p_{\text{T}i}, \eta_i)}{1 - \epsilon_{\text{fake}}(p_{\text{T}i}, \eta_i)}$$

- Estimate for pre-selection region: 232^{+294}_{-232} (stat+syst) events
  - consistent with MC prediction 146 ± 8 events

Reconstruction of Top kinematics

- Each $t\bar{t} \to \ell^+\ell^-$ event has 2 neutrinos.
- also ambiguity in combining b-jets and leptons from same top
- It is a challenge to reconstruct top mass
- We use the analytical matrix weighting technique (AMWT) described in http://arxiv.org/abs/arXiv:1105.5661

Unfolding I

- Selection cuts and detector response are modelled by the acceptance (A) and smearing (S) matrices

- Given a true binned distribution x_i we observe b_k in our detector (after background subtraction):

$$b_k = S_{kj} A_{ji} x_i$$

Inversion:

$$x = A^{-1} S^{-1} b$$

S – migration matrix, A – acceptance matrix.

A is diagonal, S has off-diagonal elements due to migration from one bin to another

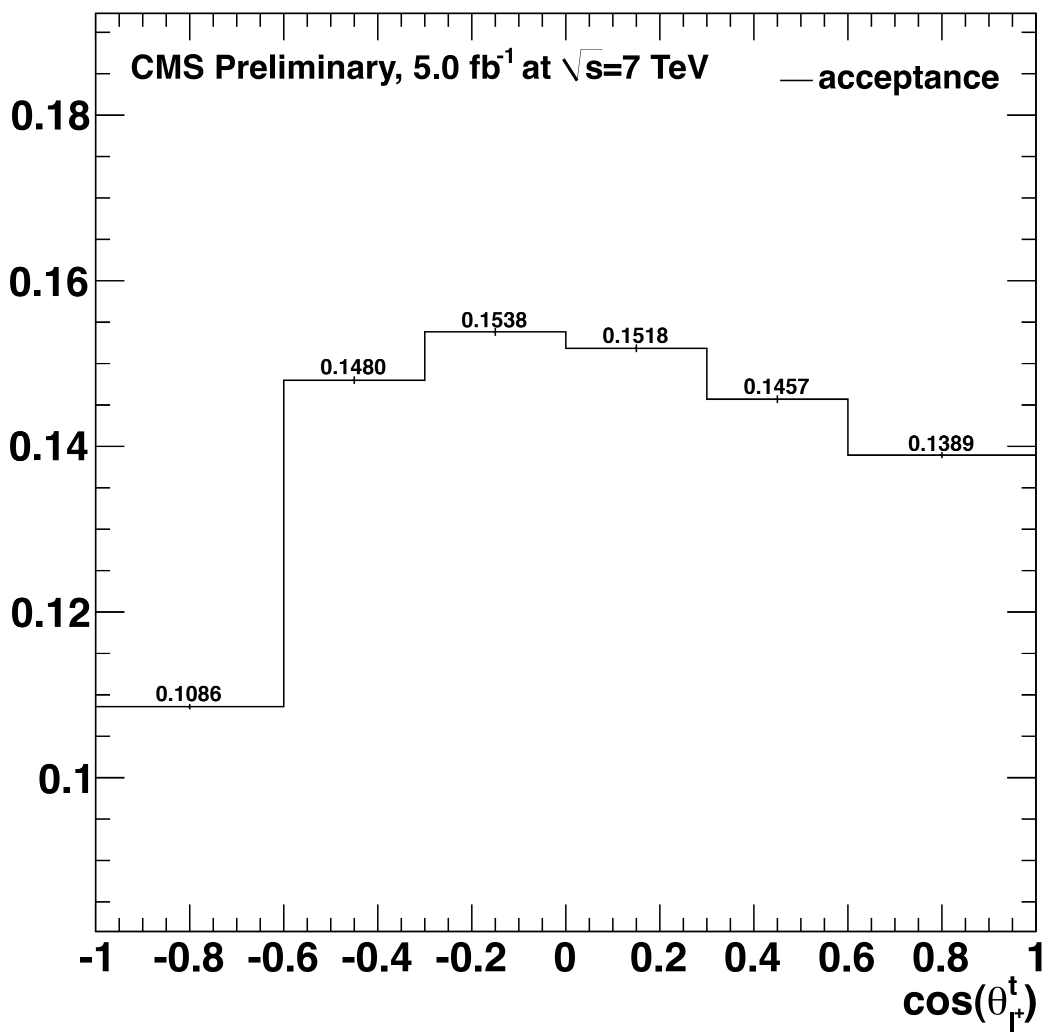
- We use regularized unfolding based on Singular Value Decomposition (SVD)
  - implemented in ROOT compatible package RooUnfold
  - SVD approach to data unfolding (Hocker and Kartvelishvili hep-ph/9509307)

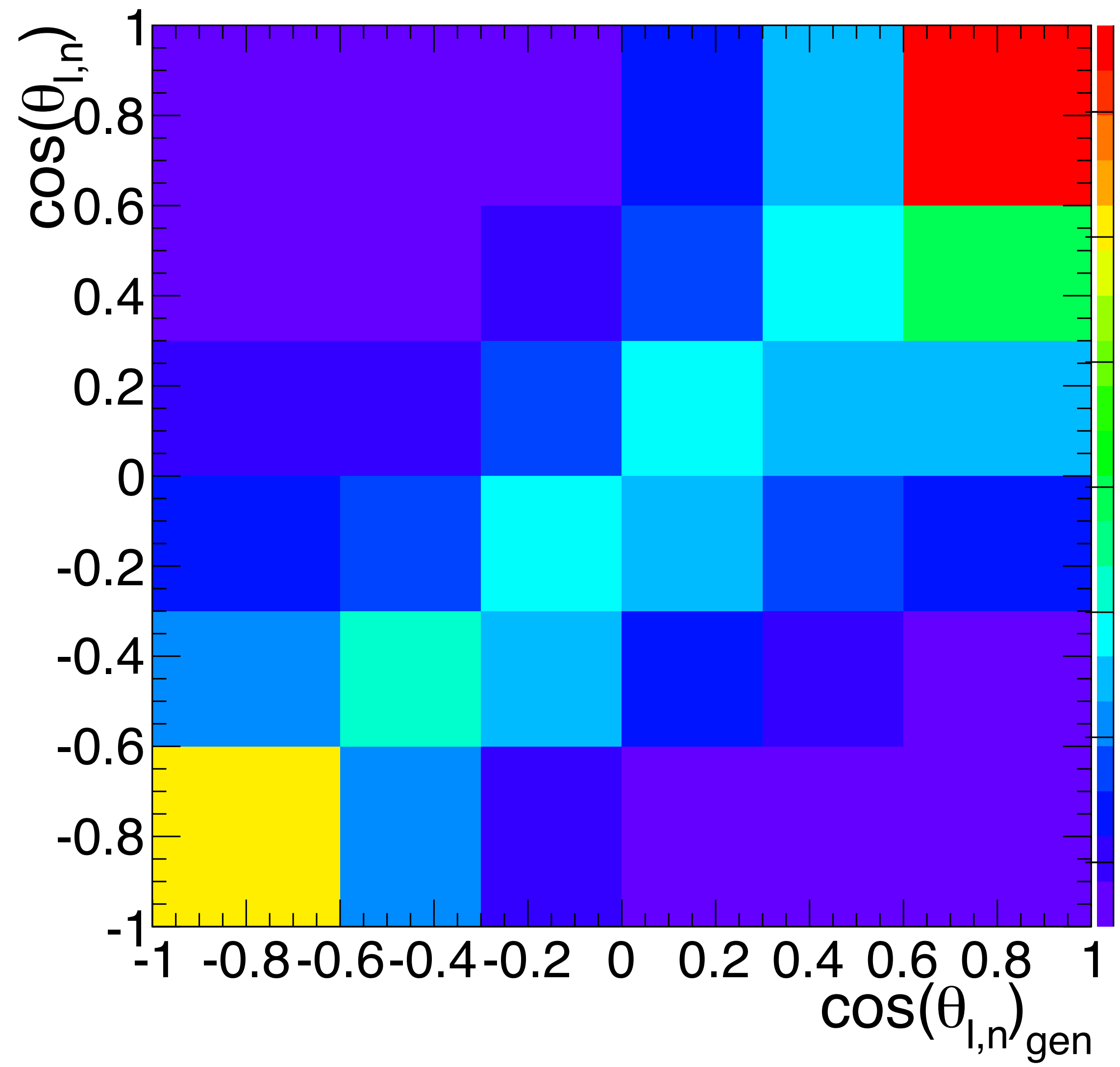
Unfolding II

- Performed extensive tests using pseudo-experiments to ensure proper performance of the unfolding algorithm
- We use 6 bins for unfolding:

B1	B2	B3	B4	B5	B6
[-1.0,-0.6]	[-0.6,-0.3]	[-0.3,-0.0]	[0.0, 0.3]	[0.3, 0.6]	[0.6, 1.0]

- Acceptance matrix and smearing matrix bins:



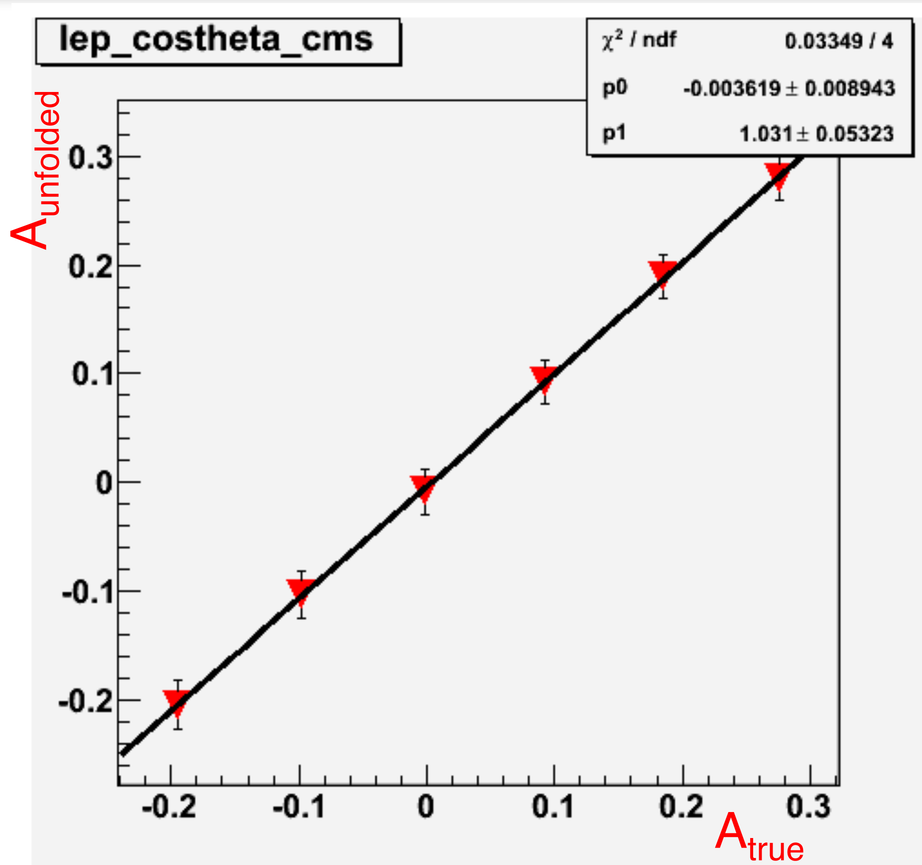


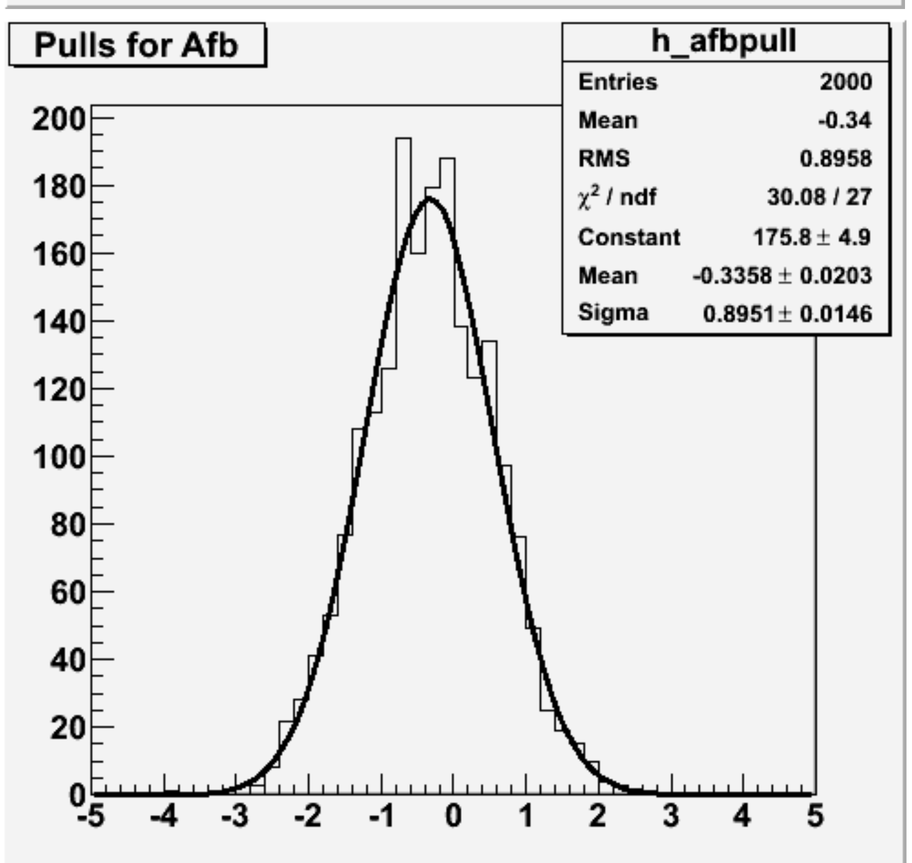
Unfolding: linearity check I

- start from ttbar in the pre-selection region
- most of our variables have no asymmetry for top
- introduce artificial asymmetry by reweighting events based on generator level quantity, for example:
    - if we are measuring Afb($|\eta_{l+}| - |\eta_{l-}|$) then reweight events as:

      weight=1+K(($|\eta_{l+}| - |\eta_{l-}|$))
    - vary K from -0.5 to 0.5 with 0.2 steps
    - covers much larger Afb range than expected from new physics

- Generate pseudo-experiments by fluctuating reweighted distribution, unfold every time
- 2000 pseudo-experiments
- Compare average to the true value





True	Measured	Unfolded
-0.19 +-0.011	-0.036+-0.011	-0.20+-0.022
-0.097+-0.011	0.02+-0.012	-0.10+-0.021
-0.002+-0.011	0.076+-0.011	-0.008+-0.021
0.092+-0.011	0.13+-0.011	0.092+-0.021
0.18+-0.011	0.18+-0.011	0.19+-0.02
0.28+-0.011	0.24+-0.011	0.28+-0.02

- Small Bias in the mean: assign systematic uncertainty
- Slight over-estimation of the uncertainty (we don't correct this)