TOP-12-016: Measurement of the top polarization in the dilepton final state

D. Barge, C. Campagnari, D. Kovalskyi, V. Krutelyov (UCSB)
W. Andrews, G. Cerati, D. Evans, F. Golf, I. MacNeill,
S. Padhi, Y. Tu, F. Wuerthwein, A. Yagil, J. Yoo (UCSD)
L. Bauerdick, K. Burkett, I. Fisk, Y. Gao, O. Gutsche, B. Hooberman,
S. Jindariani, J. Linacre, V. Martinez Outschoorn (FNAL)

Pre-Approval
8th June 2012

Documentation

CMS PAS TOP-12-016

DRAFT
CMS Physics Analysis Summary

The content of this note is intended for CMS internal use and distribution only

2012/06/01
Head Id: 126609
Archive Id: 127027P
Archive Date: 2012/05/31
Archive Tag: trunk

Measurement of the top polarization in the dilepton channel

The CMS Collaboration

DRAFT

Abstract

A measurement of top quark polarization in $t\bar{t} \to \ell^+\ell^-$ events is performed in a data sample corresponding to a total integrated luminosity of 5.0 fb^{-1} collected by the CMS experiment in pp collisions at a centre-of-mass energy of 7 TeV at the LHC. In view of a significant excess reported in the top forward backward asymmetry at the Tevatron, the measured value is compared to the standard model expectation, and is found to be in good agreement.

Available on CMS information server

CMS AN -2012/190

The Compact Muon Solenoid Experiment

Analysis Note

The content of this note is intended for CMS internal use and distribution only

17 May 2012 (v3, 01 June 2012)

Measurements of top-quark pair asymmetries in the dilepton final state at $\sqrt{s} = 7$ TeV

D. Barge, C. Campagnari, A. George, D. Kovalskyi, V. Krutelyov

University of California, Santa Barbara

W. Andrews, G. Cerati, D. Evans, F. Golf, I. Macneill, S. Padhi, Y. Tu, V. Wellke, F. Würthwein, A. Yagil, J. Yoo

University of California, San Diego

L. Bauerdick, K. Burkett, I. Fisk, Y. Gao, O. Gutsche, B. Hooberman, S. Jindariani, J. Linacre, V. Martinez Outschoorn

Fermi National Accelerator Laboratory, Batavia, Illinois

Abstract

Measurements of several dilepton asymmetries in $t\bar{t} \to \ell^+\ell^-$ events are performed in a data sample corresponding to a total integrated luminosity of 5.0 fb^{-1} collected by the CMS experiment in pp collisions at a centre-of-mass energy of 7 TeV at the LHC. The observables include the lepton charge and the top charge asymmetry, the lepton azimuthal asymmetry, as well as the top polarization and spin correlation. In view of a more significant excess reported in related observables at the Tevatron for high $t\bar{t}$ system mass, the results are also given for $t\bar{t}$ system mass above 450 GeV. The measured values of these observables are found in agreement with their standard model expectations.

- http://cms.cern.ch/iCMS/analysisadmin/cadi?ancode=TOP-12-016

Introduction

- We want to measure the top polarization in the dilepton final state
- top decays before hadronization can wash out polarization
- charged lepton is best spin analyzer
  - $$\mathcal{P}_n = \frac{N(\cos\theta_{\ell,n} > 0) - N(\cos\theta_{\ell,n} < 0))}{N(\cos\theta_{\ell,n} > 0) + N(\cos\theta_{\ell,n} < 0))}$$
  - measured in the helicity basis (angle $\theta_{\ell,n}$ of lepton measured in parent top's rest frame, relative to direction of the top in the ttbar CM)
- Any significant difference from the SM expectation could be a signal of NP
- The work is inspired by this theory paper: http://arxiv.org/abs/1105.3743 by D. Krohn, T. Liu, J. Shelton, L.T. Wang

Analysis Strategy

- Use baseline event selections (with slight changes) from our search for heavy top-like quark pair analysis (EXO-11-050)
  - purpose of this selection is to reject events other than ttbar
  - EXO-11-050 is submitted to PLB
- Datasets: DoubleElectron, DoubleMu, MuEG collected by high p_T dilepton triggers
- Summer11 MC
- We measure the top polarization and differential cross-section in $\cos\theta_{\ell,n}$ at parton level after background subtraction and unfolding
- We also look at 2 signal regions where NP is expected to be more prominent

Event selection (preselection)

- Event cleaning: if >= 10 tracks; at least 25% purity; at least 1 good DA vertex (not isFake, ndf > 4, rho < 2 cm, z < 24 cm)
- 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)
  - ΔR > 0.4 from all leptons passing analysis selection
  - ≥1b tags: CSVM
- MET > 30 GeV
- Z veto: 76<m_{ll}<106 GeV veto (for SF leptons)
- m_{ll}>12 GeV to veto low mass resonances (SF leptons)

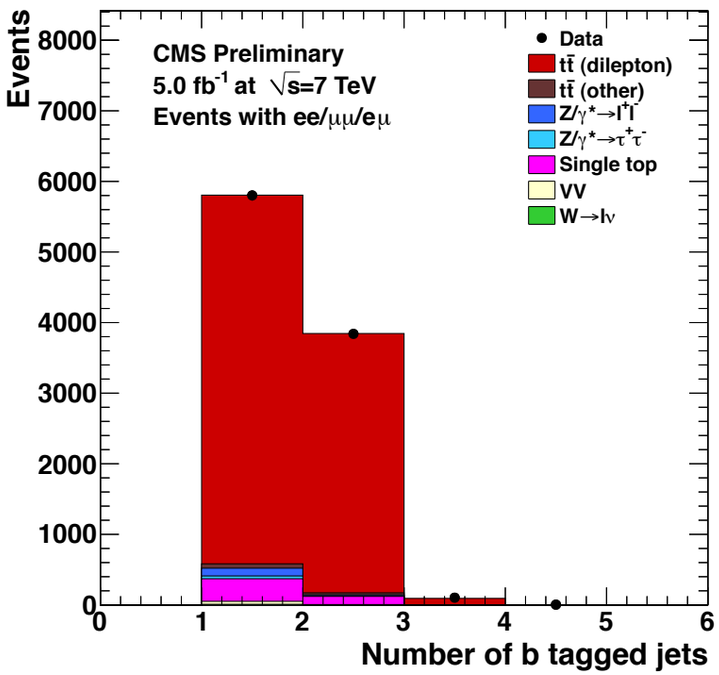
Preselection Yields (5.0 fb^{-1})

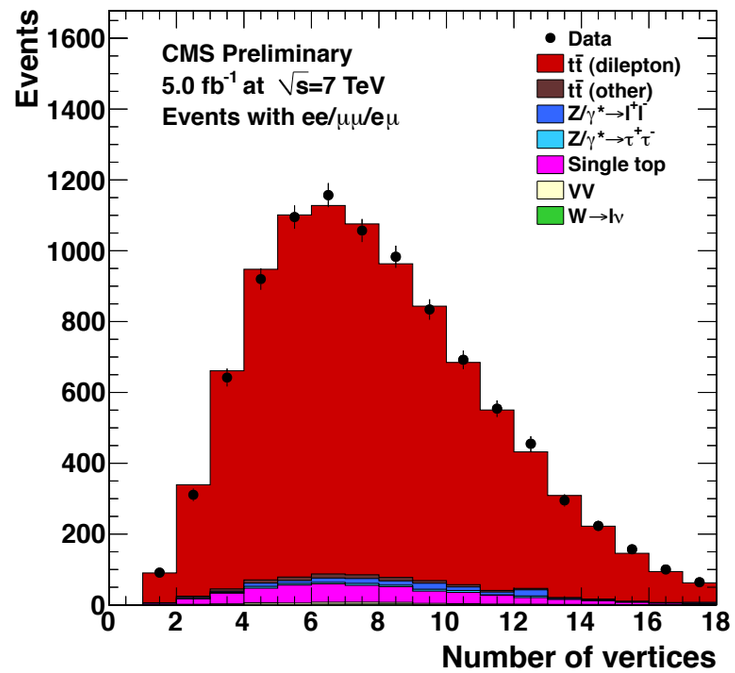
Sample	ee	$\mu\mu$	eμ	all
$t\bar{t} \to \ell^+\ell^-$	1791.7 ± 4.4	2127.3 ± 4.7	5069.4 ± 7.3	8988.5 ± 9.7
$t\bar{t} \to$ other	32.5 ± 2.9	4.8 ± 1.1	53.3 ± 3.6	90.7 ± 4.8
W + jets	< 1.9	4.7 ± 3.3	4.7 ± 3.4	9.4 ± 4.7
DY$\to$ ee	52.3 ± 5.8	< 0.6	< 0.6	52.3 ± 5.8
DY$\to \mu\mu$	< 0.6	72.8 ± 6.5	1.6 ± 0.9	74.4 ± 6.5
DY$\to \tau\tau$	17.6 ± 3.3	8.7 ± 2.2	18.7 ± 3.2	45.0 ± 5.1
Di-boson	10.6 ± 0.5	13.0 ± 0.5	24.0 ± 0.7	47.6 ± 1.0
Single top	84.9 ± 2.3	101.2 ± 2.4	252.1 ± 3.9	438.2 ± 5.1
Total (simulation)	1989.6 ± 8.8	2332.6 ± 9.3	5423.8 ± 10.3	9746.0 ± 16.4
Data	1961	2373	5412	9746

Uncertainties are statistical only

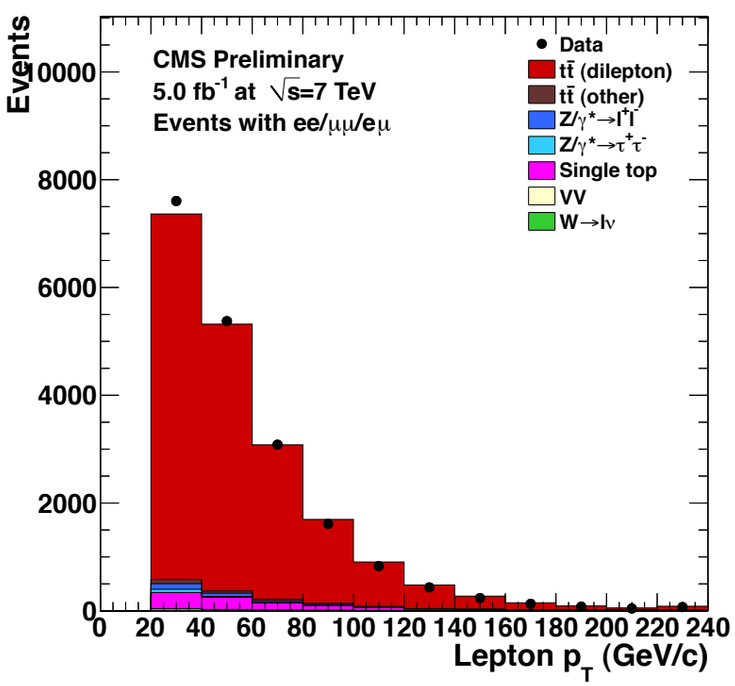
- MC events are weighted to match trigger efficiency, b tagging efficiency, and number of vertices distribution in data
- We use powheg-pythia for the $t\bar{t} \to \ell^+\ell^-$ component
  - normalized so that total MC yield matches data (corresponds to inclusive ttbar xsec of 167.7 pb)
  - $t\bar{t} \to \ell^+\ell^-$ contributes 92% of the total yield
- Comparison plots on next slide

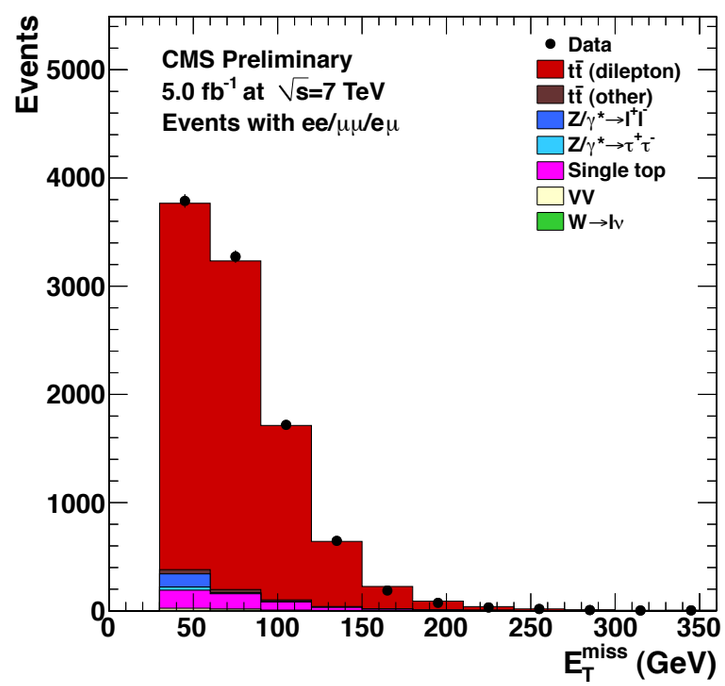
Data-MC comparison

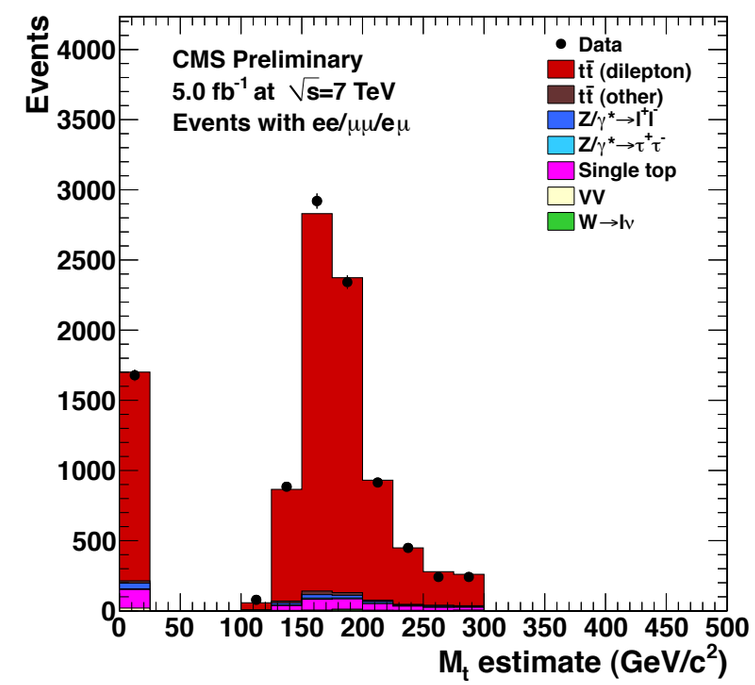




- Vertex reweighting has been applied to MC
- Selected plots: #btag jets, #vertices, lepton p_T, MET and reconstructed M_t
- Data and MC agreement is reasonable





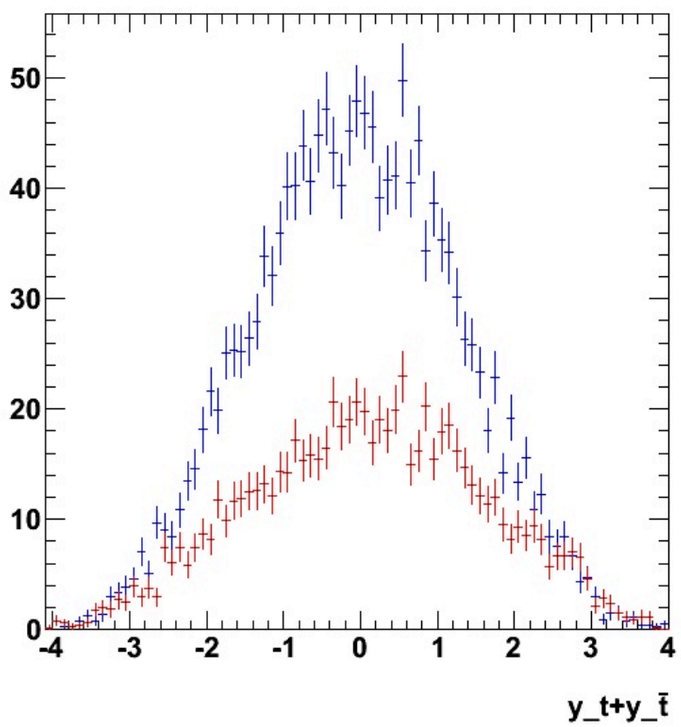


Reconstruction of Top kinematics

- 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
- Each events is reconstructed using a range of possible M_t values between 100-300 GeV in 1 GeV steps.
  - M_t value with the maximum averaged weight over possible solutions is taken
  - ttbar kinematics taken from solution with largest weight
- Events with no solutions are discarded (~17%)

Signal Regions

- Signal Region I: M_{tt} > 450 GeV
  - NP contribution expected to be enhanced at high M_{tt}
- Signal Region II: M_{tt} > 450 GeV and $|y_t+y_{tbar}|>2$
  - NP signal expected only in qqbar -> ttbar component
  - the gluon PDFs fall more rapidly at large x than the quark PDFs so gg -> ttbar tends to be more central than qqbar -> ttbar.



Red: from quark annihilation
Blue: from gluon fusion

gg reduced to similar level as qqbar when $|y_t+y_{tbar}|>2$

Background estimation

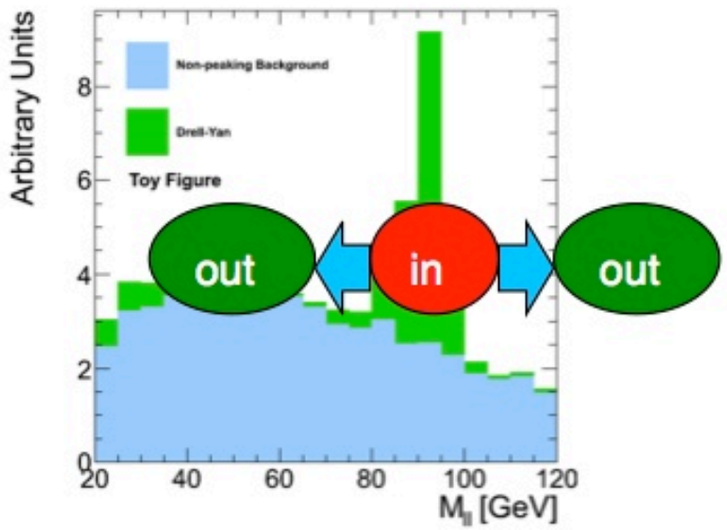
- We use the MC from the previous slides to estimate the background
- We make cross-checks for the DY and fake components using data-driven methods, and find good agreement
- We then assign an appropriate background normalization systematic

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

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

same-flavor events in Z peak

- Estimate for pre-selection region: 142.4 ± 15.0 events
  - consistent with MC prediction of 126.7 ± 8.7 events
- Estimate for Signal Region I: 47.6 ± 10.6 events
  - consistent with MC prediction of 39.9 ± 4.8 events
- Estimate for Signal Region II: 10.8 ± 6.0 events
  - consistent with MC prediction of 9.5 ± 2.4 events

Data-driven BG estimates: Fakes

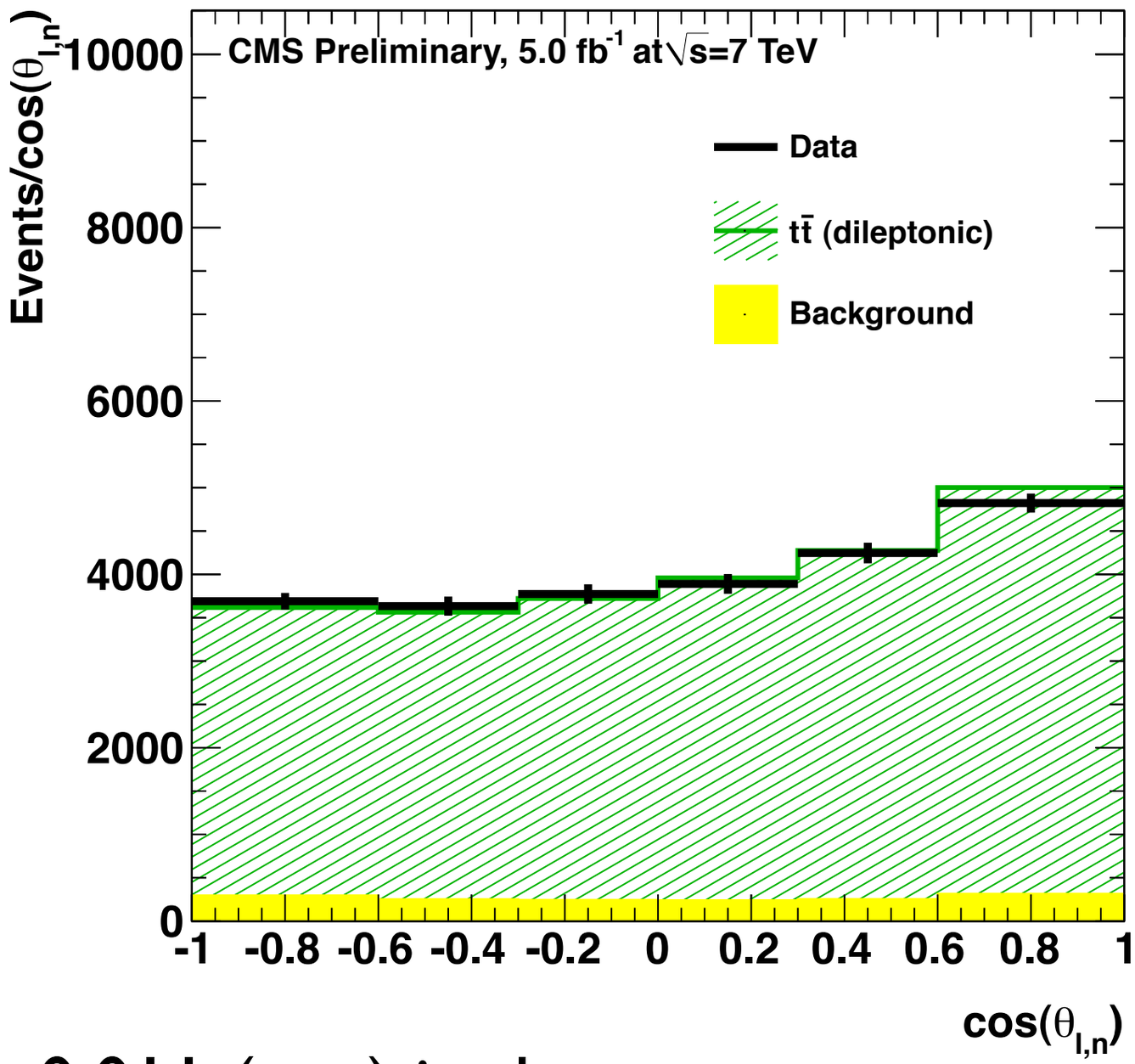
- Estimate contribution from fake leptons using the data-driven tight-to-loose method described in CMS AN-2010/257
  - measure tight-to-loose fake rates as a function of lepton P_T and eta
  - estimate number of fakes in data based on number of fakeable object (FOs). Weight each lepton+FO event by:
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

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

$$w_i = \frac{\epsilon_{\rm fake}(p_{\rm Ti}, \eta_i)}{1 - \epsilon_{\rm fake}(p_{\rm Ti}, \eta_i)}$$

- Estimate for pre-selection region: 138^{+281}_{-138} events
  - consistent with MC prediction 100.1 ± 6.7 events
- Estimate for Signal Region I: $41.7^{+108.8}_{-41.7}$ events
  - consistent with MC prediction of 47.1 ± 5.4 events
- Estimate for signal region: $6.6^{+16.4}_{-6.6}$ events
  - consistent with MC prediction of 8.2 ± 2.5 events

Reco level asymmetries: preselection

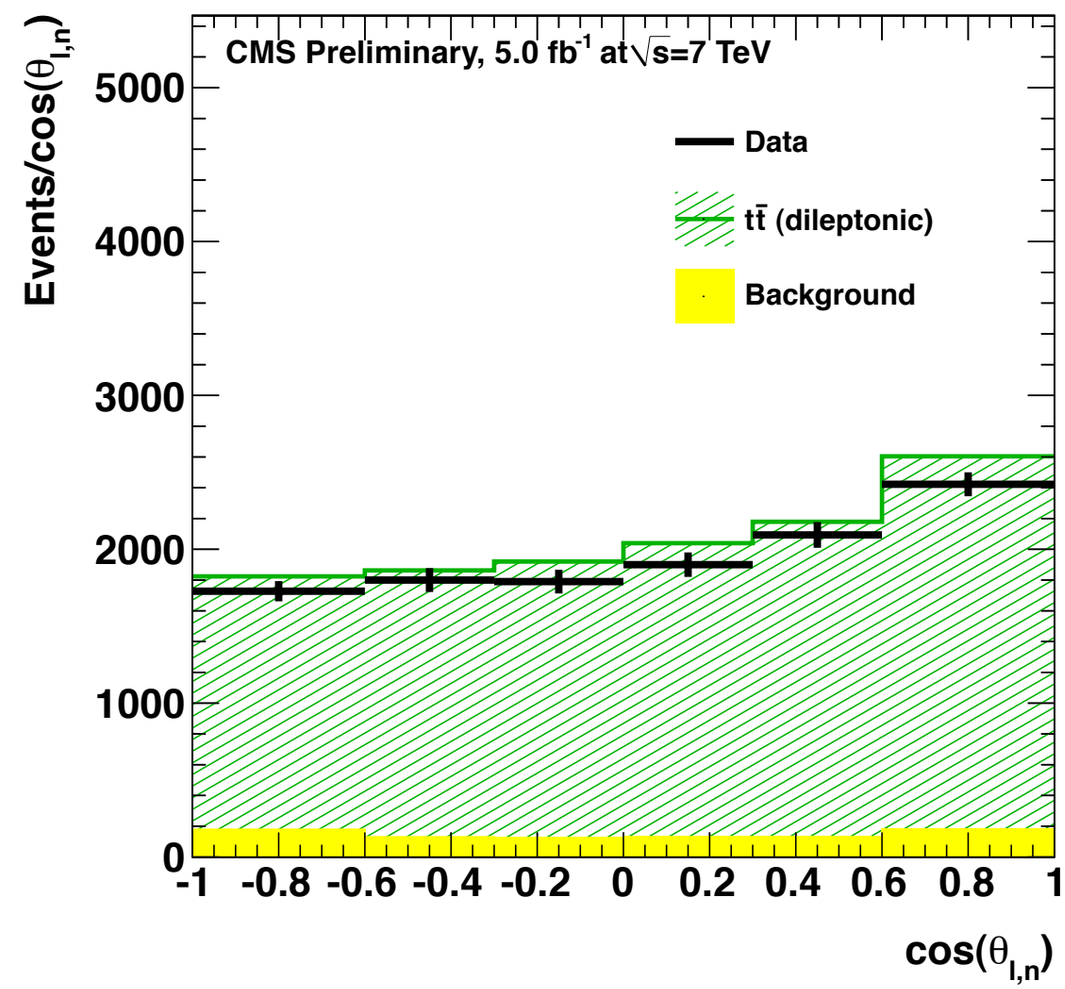


- P_n = 0.083 ± 0.011 (stat) in data
- P_n = 0.103 ± 0.002 (stat) in MC (sum of ttbar and background)

Reco level asymmetries: Signal Region I

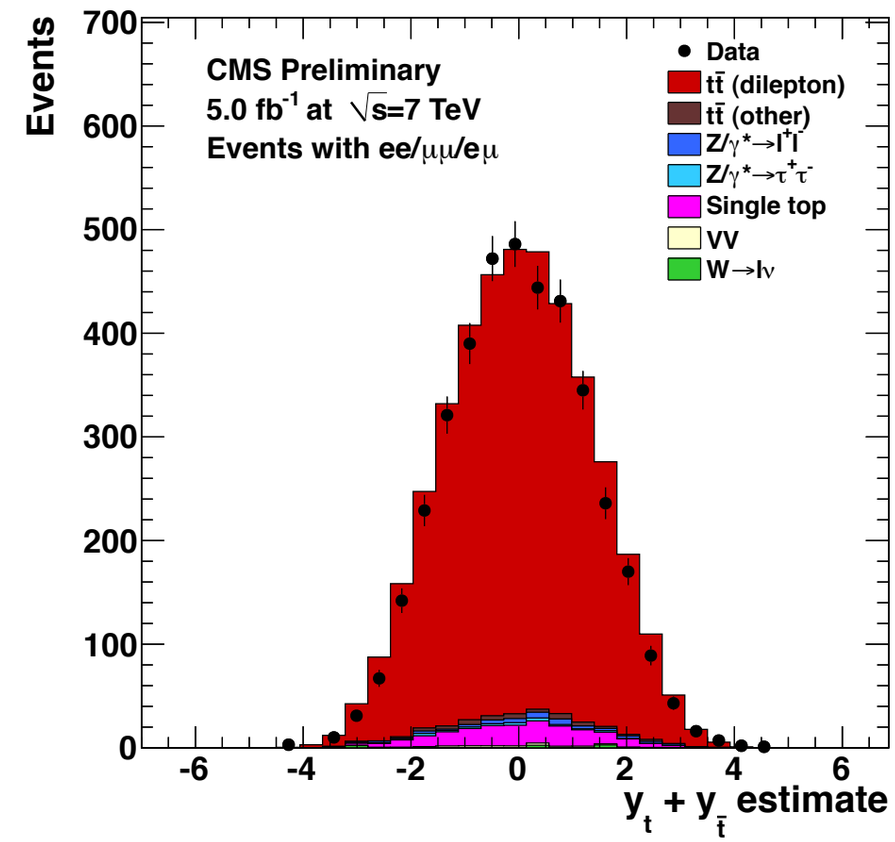
Sample	ee	$\mu\mu$	eμ	all
$t\bar{t} \to \ell^+\ell^-$	777.6 ± 2.9	921.4 ± 3.1	2143.0 ± 4.8	3842.0 ± 6.4
$t\bar{t} \to$ other	14.6 ± 2.0	1.5 ± 0.6	23.4 ± 2.4	39.6 ± 3.2
W + jets	0.0 ± 0.0	4.7 ± 3.3	2.8 ± 2.8	7.5 ± 4.4
DY$\to$ ee	13.8 ± 2.9	0.0 ± 0.0	0.0 ± 0.0	13.8 ± 2.9
DY$\to \mu\mu$	0.0 ± 0.0	25.6 ± 3.8	0.5 ± 0.5	26.1 ± 3.9
DY$\to \tau\tau$	7.4 ± 2.2	2.6 ± 1.2	8.0 ± 2.1	18.0 ± 3.3
Di-boson	3.9 ± 0.3	4.7 ± 0.3	9.7 ± 0.5	18.3 ± 0.6
Single top	32.8 ± 1.4	41.9 ± 1.6	101.3 ± 2.5	176.0 ± 3.3
Total (simulation)	850.2 ± 5.3	1002.4 ± 6.3	2288.7 ± 6.9	4141.4 ± 10.7
Data	801	970	2164	3935

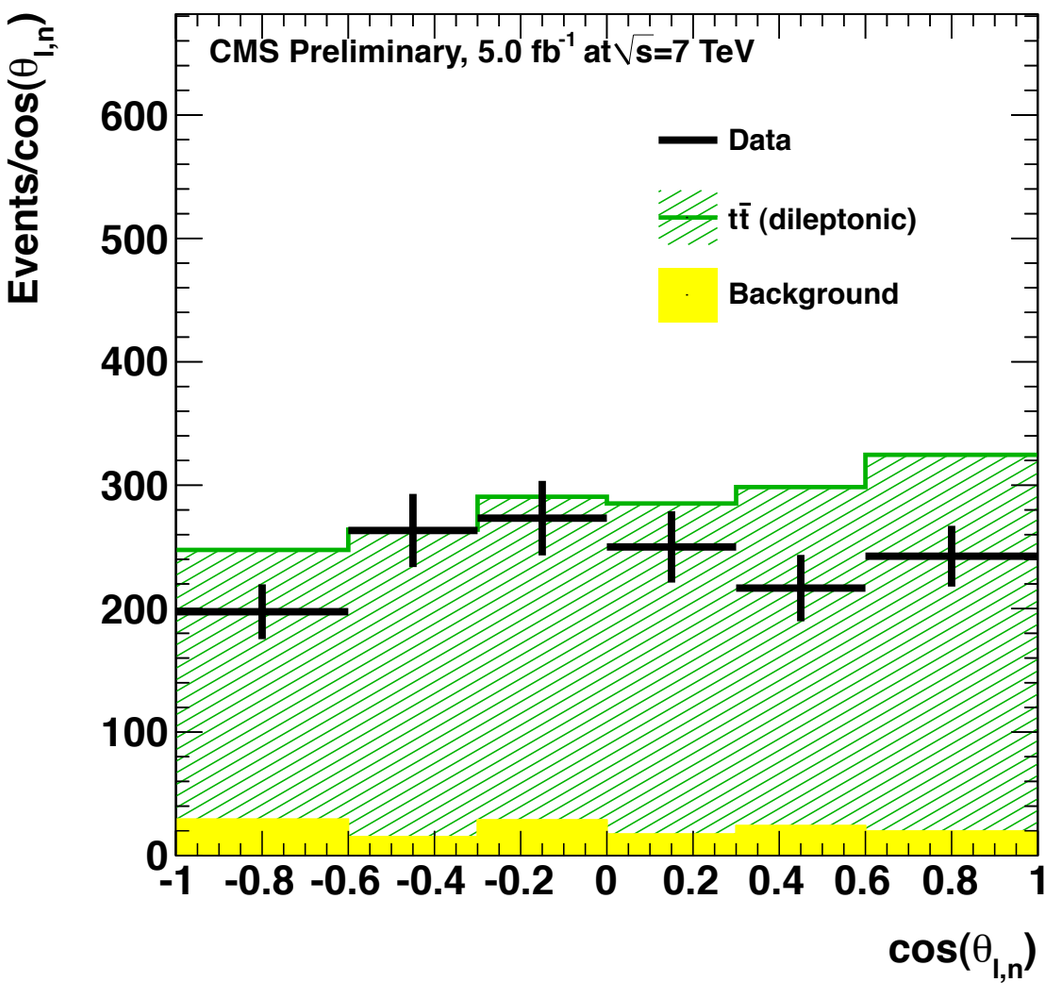
- $P_n = 0.101 \pm 0.016$ (stat) in data
- $P_n = 0.106 \pm 0.003$ (stat) in MC



Reco level asymmetries: Signal Region II

Sample	ee	$\mu\mu$	eμ	all
$t\bar{t} \to \ell^+\ell^-$	104.9 ± 1.1	124.1 ± 1.1	292.2 ± 1.8	521.2 ± 2.3
$t\bar{t} \to$ other	2.8 ± 0.9	0.4 ± 0.3	2.8 ± 0.8	6.0 ± 1.2
W + jets	0.0 ± 0.0	2.2 ± 2.2	0.0 ± 0.0	2.2 ± 2.2
DY$\to$ ee	2.5 ± 1.3	0.0 ± 0.0	0.0 ± 0.0	2.5 ± 1.3
DY$\to \mu\mu$	0.0 ± 0.0	6.5 ± 1.9	0.5 ± 0.5	7.0 ± 2.0
DY$\to \tau\tau$	0.0 ± 0.0	1.0 ± 0.7	0.9 ± 0.7	1.9 ± 1.0
Di-boson	0.7 ± 0.1	0.5 ± 0.1	1.4 ± 0.2	2.5 ± 0.2
Single top	3.7 ± 0.5	5.9 ± 0.6	13.4 ± 0.9	23.0 ± 1.2
Total (simulation)	114.5 ± 1.9	140.7 ± 3.3	311.2 ± 2.3	566.3 ± 4.5
Data	103	116	258	477





- MC total yield 19% more than data
  - $|y_t+y_{tbar}|$ more peaked in data
  - normalization does not affect value of polarization
- $P_n = -0.006 \pm 0.046$ (stat) in data
- $P_n = 0.069 \pm 0.008$ (stat) in MC
- consistent within large uncertainty

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

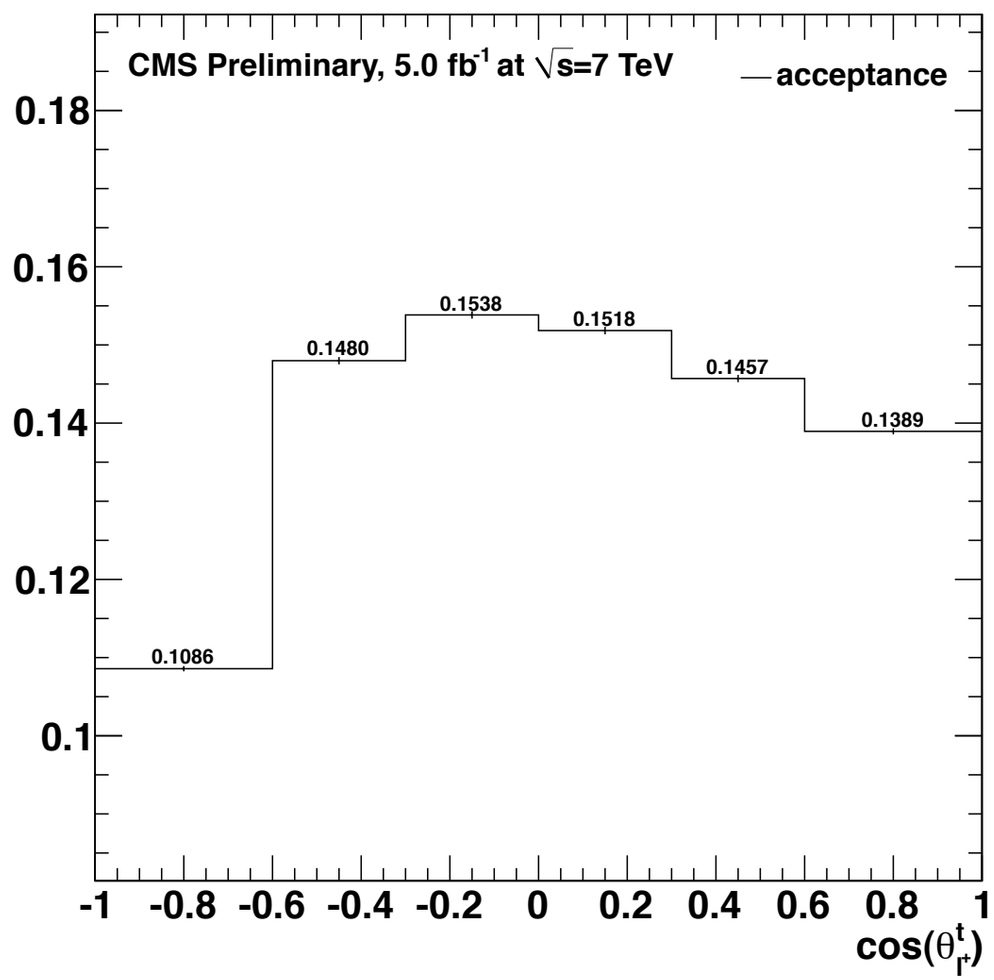
- We use regularized unfolding based on Singular Value Decomposition (SVD)
  - implemented in ROOT compatible package RooUnfold
  - SVD approach to data unfolding (Hocker and Kartvelishvili hep-ph/9509307)

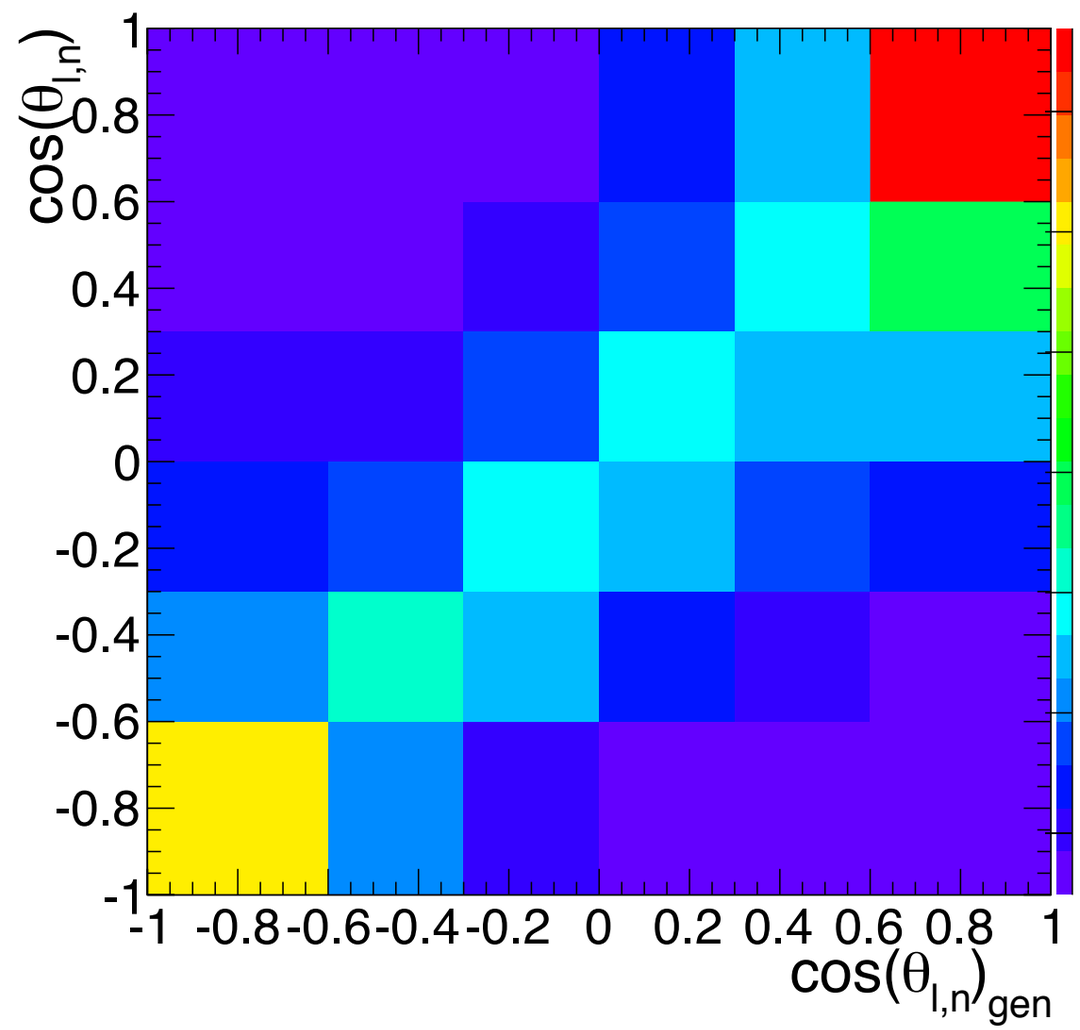
Unfolding II

- 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:





Systematics

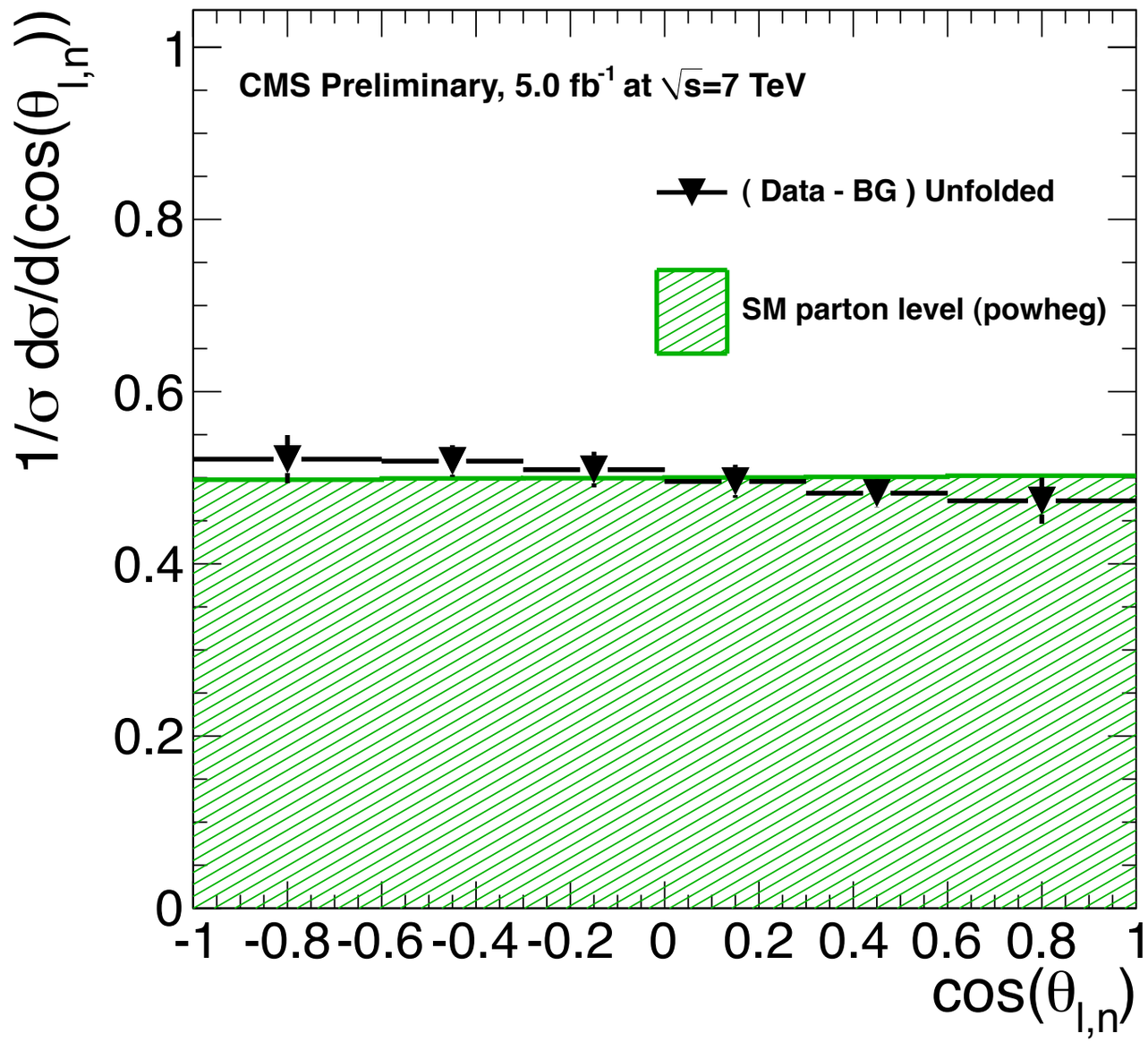
- Systematics are evaluated on the unfolded result

Table 6: Systematic uncertainties.

JES	BG	modeling	unfolding	top mass	$b-$tagging	Trigg(lep ID)	PU	Total
0.043	0.009	0.014	0.009	0.019	0.001	0.000	0.002	0.050

- JES is by far the dominant systematic
  - calculated assuming a 7.5% uncertainty on the hadronic energy scale (after L1FastL2L3 correction)
  - this directly affects the shape of the $\cos\theta_{\ell,n}$ distribution
- For BG we vary normalizations by 50 or 100%
- Most other systematics assessed by varying the model used to calculate the unfolding
  - PU and b-tagging and trig/lep ID eff from reweighting powheg MC
  - modelling, matching, top mass use MC@NLO and madgraph MC
- Small unfolding bias also gives a systematic

Unfolded results



- $P_n = -0.035 \pm 0.028 \pm 0.050$ in data
- $P_n = 0.003 \pm 0.0004$ in MC (parton level, no cuts)
- Result in data is consistent with the SM

Summary

- Performed an analysis measuring top polarization
  - NP responsible for Tevatron A_{fb} could cause deviation from SM
- Backgrounds predicted by MC, but checked using data-driven methods and conservative normalization systematics calculated
- Measure P_n at parton level using unfolding technique (extensively validated)
  - $P_n = -0.035 \pm 0.028 \pm 0.050$
  - consistent with SM expectation of ~0
- Also compare $\cos\theta_{\ell,n}$ distribution between data and MC at reco level in 2 signal regions
  - no significant deviation observed

Backup

Previous presentations by Yanjun Tu, Jacob Linacre and Sergo Jindariani in top property group meeting:

https://indico.cern.ch/getFile.py/access?contribId=1&resId=0&materialId=slides&confId=180584

http://indico.cern.ch/getFile.py/access?contribId=8&sessionId=0&resId=0&materialId=slides&confId=180655

http://indico.cern.ch/getFile.py/access?contribId=1&resId=0&materialId=slides&confId=187624

https://indico.cern.ch/getFile.py/access?contribId=2&resId=0&materialId=slides&confId=190610