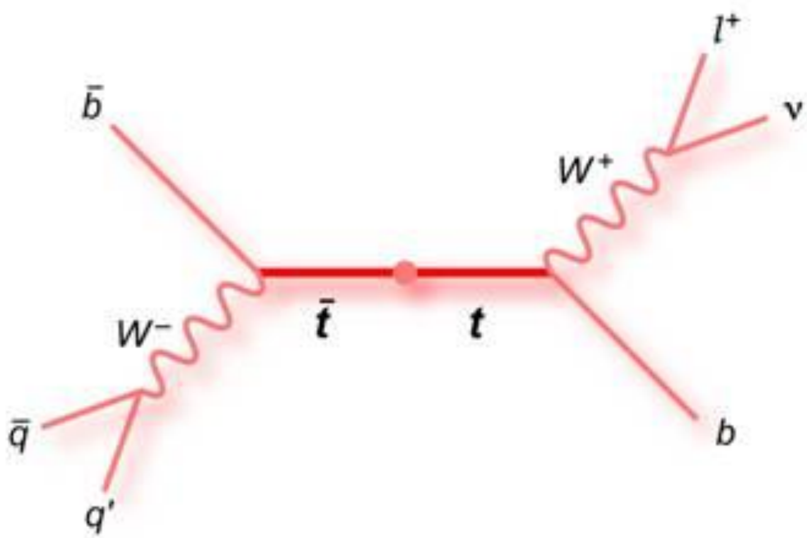
Determination of A^t_{FB} using boosted top tagging at 1.4TeV

Alasdair Winter
University of Birmingham
CLICdp 22 Feb 2018

Overview

- Aims:
  - identify semileptonic $t\bar{t}$ decays at 1.4TeV
  - examine prospects for determining A_{FB}^{t} through measurement of the top angular distribution
- Boosted topology makes conventional top tagging techniques a challenge
  - b-tagging alone no longer viable!
- Approach is to use the concept of fat jets and look at jet substructure



Topics for today

- Correcting bugs:
  - Signal definition
  - S' determination

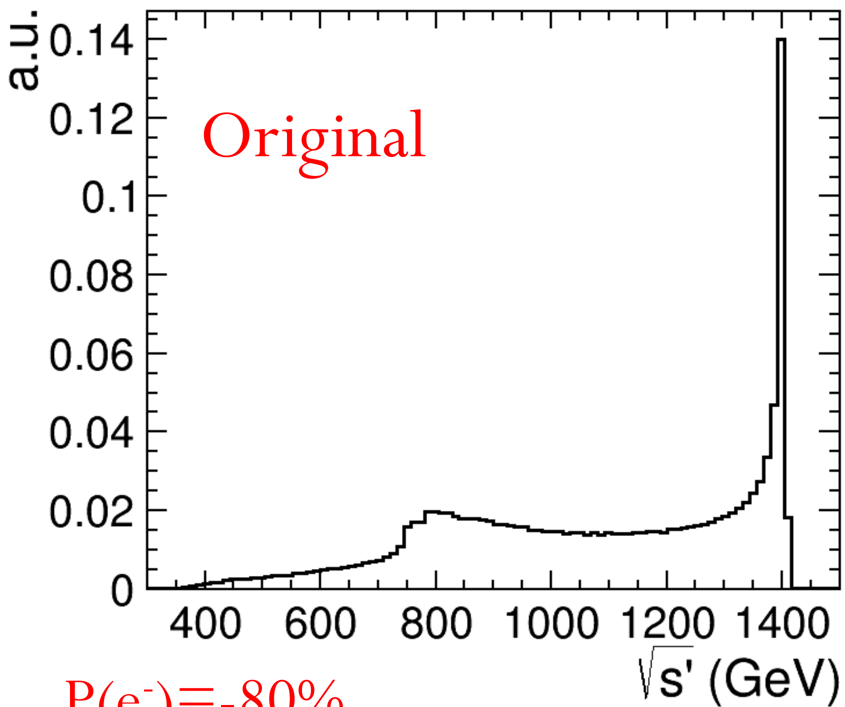
- First look at +80% polarization results

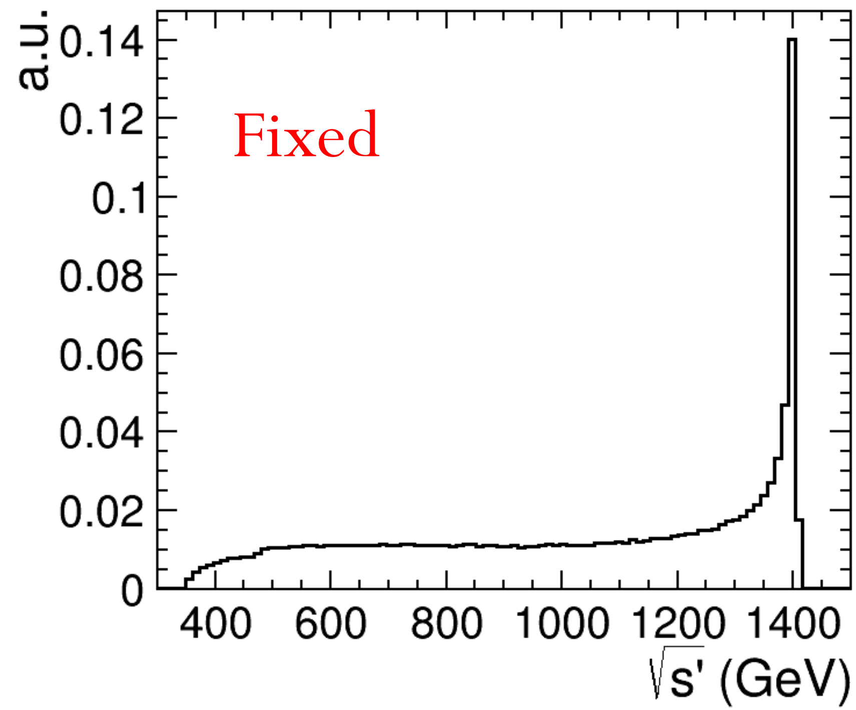
- Testing fitting techniques

Bug Fixing

- Two important bugs found- hadn't been spotted before as they are most significant at lower S', until now the two analysis only compared results at E>1200GeV
- S prime determination: bug in code meant that truth level s prime was being defined as the sum of the energies of the electron positron pair rather than the invariant mass of the pair- fixed now!
- Signal definition: algorithm for searching for ttbar pairs in qqqqlv sample was found to stop before trying all possible fermion combinations- resulted in true ttbar events being wrongly assigned to the single top sample- fixed!

Bugs Fixing





$P(e^-)=-80\%$

Energy (GeV)	Old σ (fb)	Corrected σ (fb)
400-900	12.0	16.6
900-1200	13.2	11.0
>1200	18.4	18.4

Overall, signal cross section is now larger than previously thought → can expect better precision! ☺

Extracting A_{FB} and Statistical Uncertainty

1. Split signal events into two samples- A and B
2. Level signal samples to all correspond to the same luminosity
3. Evaluate signal efficiencies post event selection using sample B
4. Subtract background & apply efficiency correction to sample A
   - Assume no uncertainty on efficiency or background subtraction as they can be modelled to arbitrary precision with enough MC
   - Fractional uncertainty on each bin $= \frac{\sqrt{S+B}}{S}$
5. Repeat process with samples inverted
6. Combine resulting Cosθ distribution
7. Fit to: $\frac{d\sigma}{dCos\theta} = \frac{3}{8}(1 + Cos^2\theta)\sigma_U + \frac{3}{4}(1 - Cos^2\theta)\sigma_L + A_{FB}\mathrm{Cos}\theta\ \sigma_{Tot}$
8. Scale uncertainty from fit to the nominal luminosity of $750\mathrm{fb}^{-1}$

Fit results

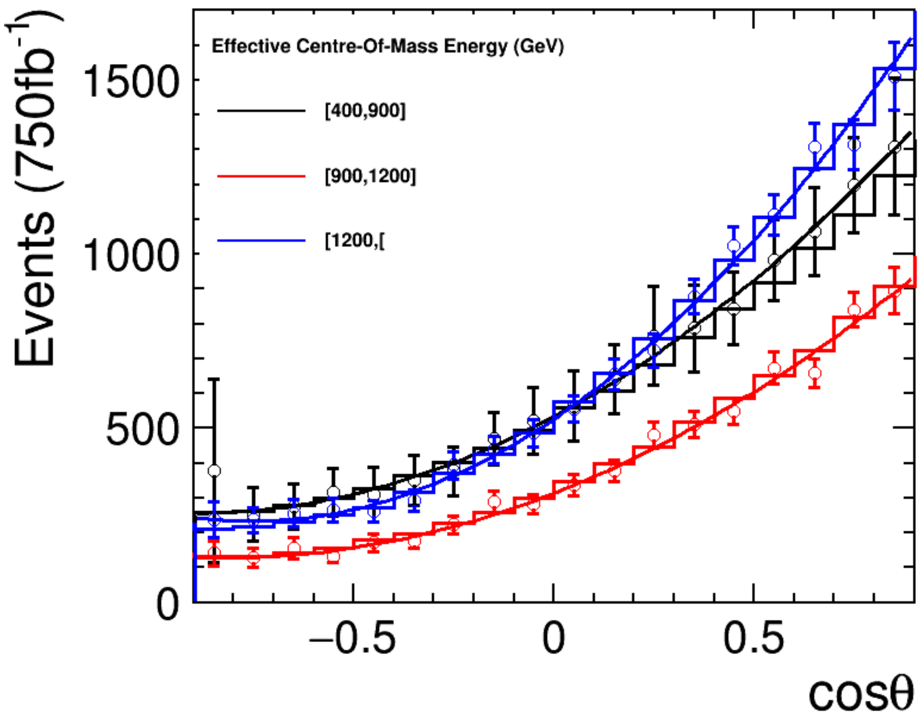
Polarization = -80%

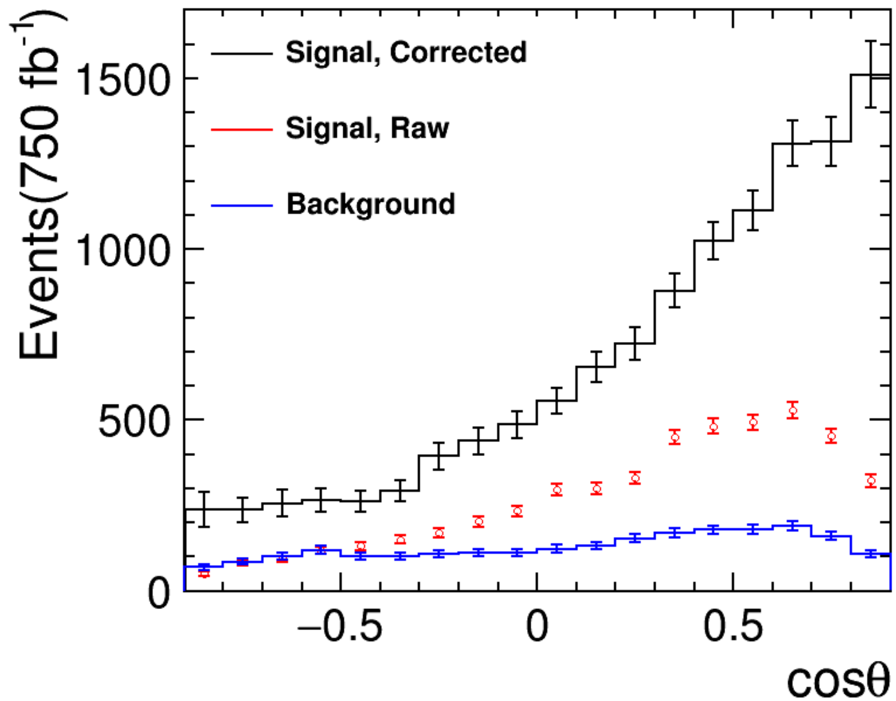
Energy (GeV)	A_{FB} (True)	A_{FB} (Reco)	σ (True, fb)	σ (Reco, fb)
>1200	0.563	0.562 +/- 0.016	18.4	18.4 +/- 1.0
900-1200	0.547	0.547 +/- 0.019	11.0	11.0 +/- 0.7
400-900	0.457	0.456 +/- 0.038	16.6	16.6 +/- 2.1

Polarization = +80%

Energy (GeV)	A_{FB} (True)	A_{FB} (Reco)	σ (True, fb)	σ (Reco, fb)
>1200	0.621	0.619 +/- 0.020	9.8	9.9 +/- 0.7
900-1200	0.605	0.597 +/- 0.026	5.9	5.9 +/- 0.5
400-900	0.525	0.512 +/- 0.050	8.6	8.7 +/-1.7

Fit Results- P(e^-)=-80%





- Example plots for the upcoming top paper
- LHS: histogram=generator level, points and fit = final reconstruction
- RHS: All data is at reconstructed level, E>1200GeV

Fitting techniques

- Fractional uncertainty on cross sections observed to be rather large
- Try to compare to just measuring significance, $\frac{\sqrt{S+B}}{S}$
- To rule out effects from background subtraction, signal efficiency, detector acceptance, look at generator level info

Energy (GeV)	σ_{Fit} (fb)	Relative Err. From Fit (%)	$\frac{\sqrt{S+B}}{S}$ (%)
>1200	18.4	1.79	0.85
900-1200	11.0	2.33	1.10
400-900	16.6	1.94	0.90

- Fit clearly not providing best possible performance!

Fitting techniques

- Current approach uses a three parameter fit:

$$\frac{d\sigma}{dCos\theta} = \frac{3}{8}(1 + Cos^2\theta)\sigma_U + \frac{3}{4}(1 - Cos^2\theta)\sigma_L + A_{FB}\text{Cos}\theta\ \sigma_{Tot}$$

- Where $\sigma_U, \sigma_L, \sigma_{Tot}$ are the unpolarised, longitudinally polarised and total cross sections, where $\sigma_{Tot} = \sigma_U + \sigma_L$
- Could see improvement in uncertainty if less parameters needed
- Try switching to alternative fit method:

$$\frac{d\sigma}{dCos\theta} = \sigma_{Tot}(1 + Cos^2\theta + \frac{8}{3}A_{FB}Cos\theta)$$

- Equivalent under the assumption $\sigma_U \gg \sigma_L$
- Currently see $\sigma_U \sim 35 \times \sigma_L$ for E>1200GeV

Fitting techniques- generator level

- Compare results from 3 parameter fit, 2 parameter fit and counting total events at generator level

Energy (GeV)	$\sigma_{3\ \text{Par Fit}}$ (fb)	Rel. Err. (%)	$\sigma_{2\ \text{Par Fit}}$ (fb)	Rel. Err. (%)	σ_{Counting} (fb)	Rel. Err. (%)
>1200	18.42	1.79	18.41	0.85	18.42	0.85
900-1200	11.02	2.33	11.01	1.10	11.01	1.10
400-900	16.57	1.94	16.44	0.90	16.56	0.90

Energy (GeV)	$Afb_{3\ \text{Par Fit}}$	Rel. Err. (%)	$Afb_{2\ \text{Par Fit}}$	Rel. Err. (%)	$\frac{Forward-Back}{Total}$	Rel. Err. (%)
>1200	0.563	1.11	0.570	0.97	0.564	1.25
900-1200	0.547	1.49	0.557	1.31	0.547	1.68
400-900	0.457	1.55	0.485	1.35	0.457	1.75

- 2D fit shows bias in central values
  - More prominent for lower energy ($\sigma_U \gg \sigma_L$ assumption breaks down)
- 3D fit best for A_{FB}, not as good for cross section

Fitting techniques- Reco. Level

- Calculating the cross section from counting total events

Energy (GeV)	$\sigma_{3\ Par\ Fit}$ (fb)	Rel. Err. (%)	$\sigma_{Counting}$ (fb)	Rel. Err. (%)	sqrt(S+B)/S (%)
>1200	18.44	5.16	18.41	1.73	1.73
900-1200	11.03	6.14	11.01	2.09	2.09
400-900	16.59	12.74	16.56	4.09	4.09

- Results agree with truth level info (by construction)
- Vast improvement in cross section uncertainty

Final Results

- Optimal results come from using a 3D fit to extract A_{FB} but integrating the total distribution to calculate the cross section

Polarization = -80%

Energy (GeV)	A_{FB} (True)	A_{FB} (Reco)	σ (True, fb)	σ (Reco, fb)
>1200	0.563	0.562 +/- 0.016	18.4	18.4 +/- 0.3
900-1200	0.547	0.547 +/- 0.019	11.0	11.0 +/- 0.2
400-900	0.457	0.456 +/- 0.038	16.6	16.6 +/- 0.7

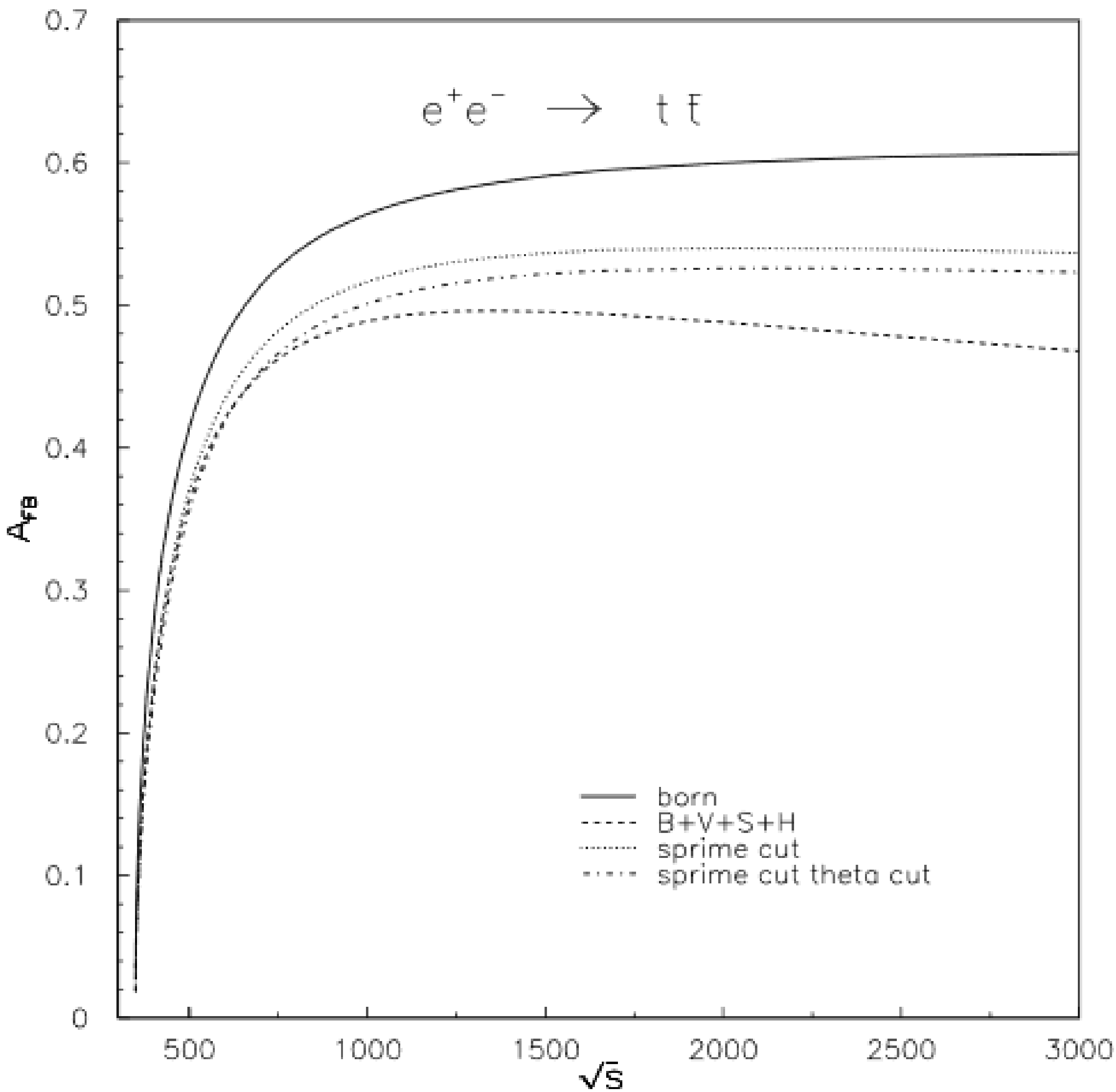
Polarization = +80%

Energy (GeV)	A_{FB} (True)	A_{FB} (Reco)	σ (True, fb)	σ (Reco, fb)
>1200	0.621	0.619 +/- 0.020	9.8	9.8 +/- 0.2
900-1200	0.605	0.597 +/- 0.026	5.9	5.9 +/- 0.2
400-900	0.525	0.512 +/- 0.050	8.6	8.6 +/- 0.5

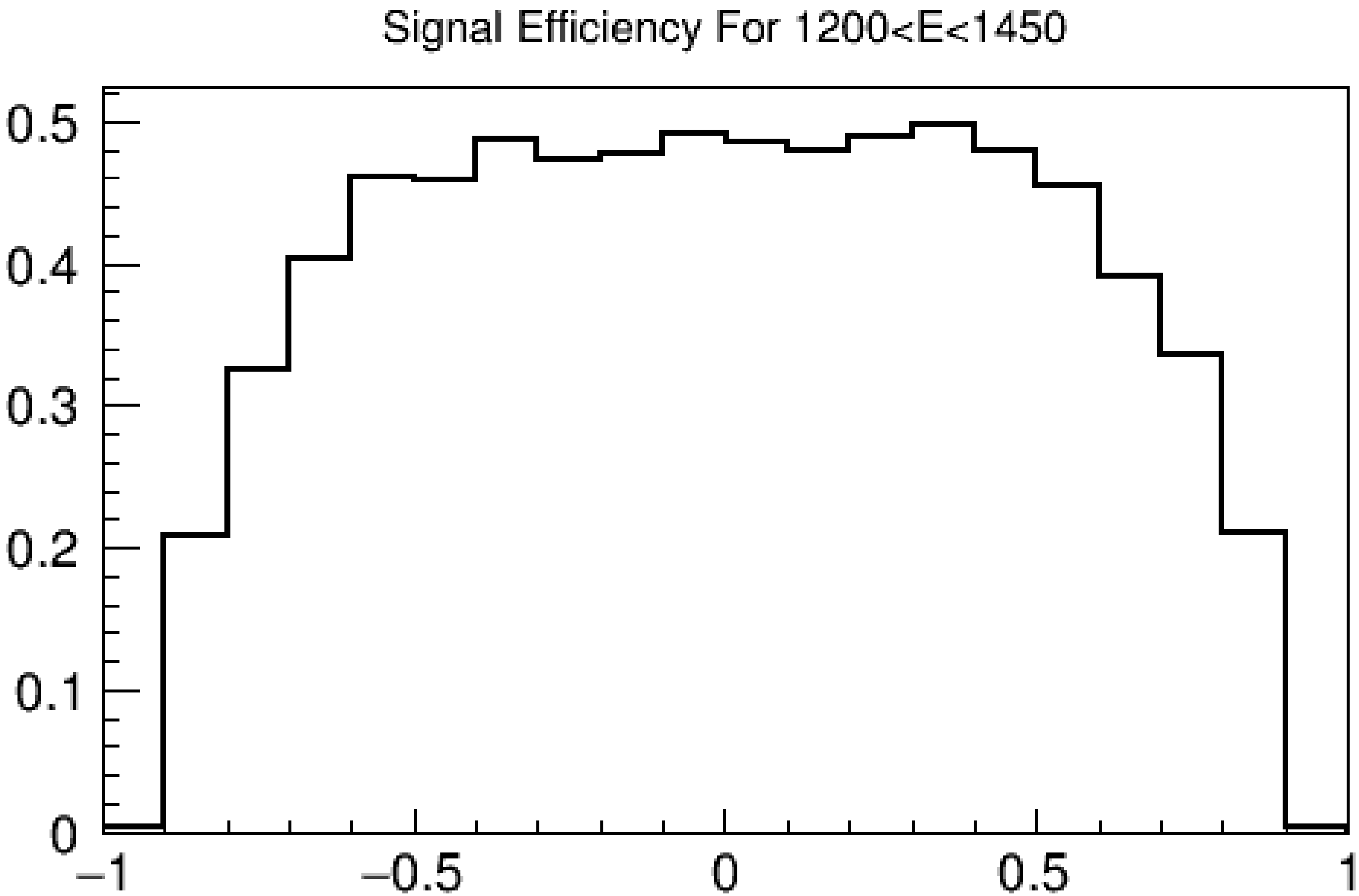
Conclusions

- Bugs found in signal definition- now fixed ☺
- Results for +80% polarization now included in analysis
- Studies into the performance of the fit suggest it works well for measuring A_{FB}, however integrating the total events works better for cross section
- Systematics need re-evaluated following bug fixes
  - Code already in place for much of this

Backup Slides



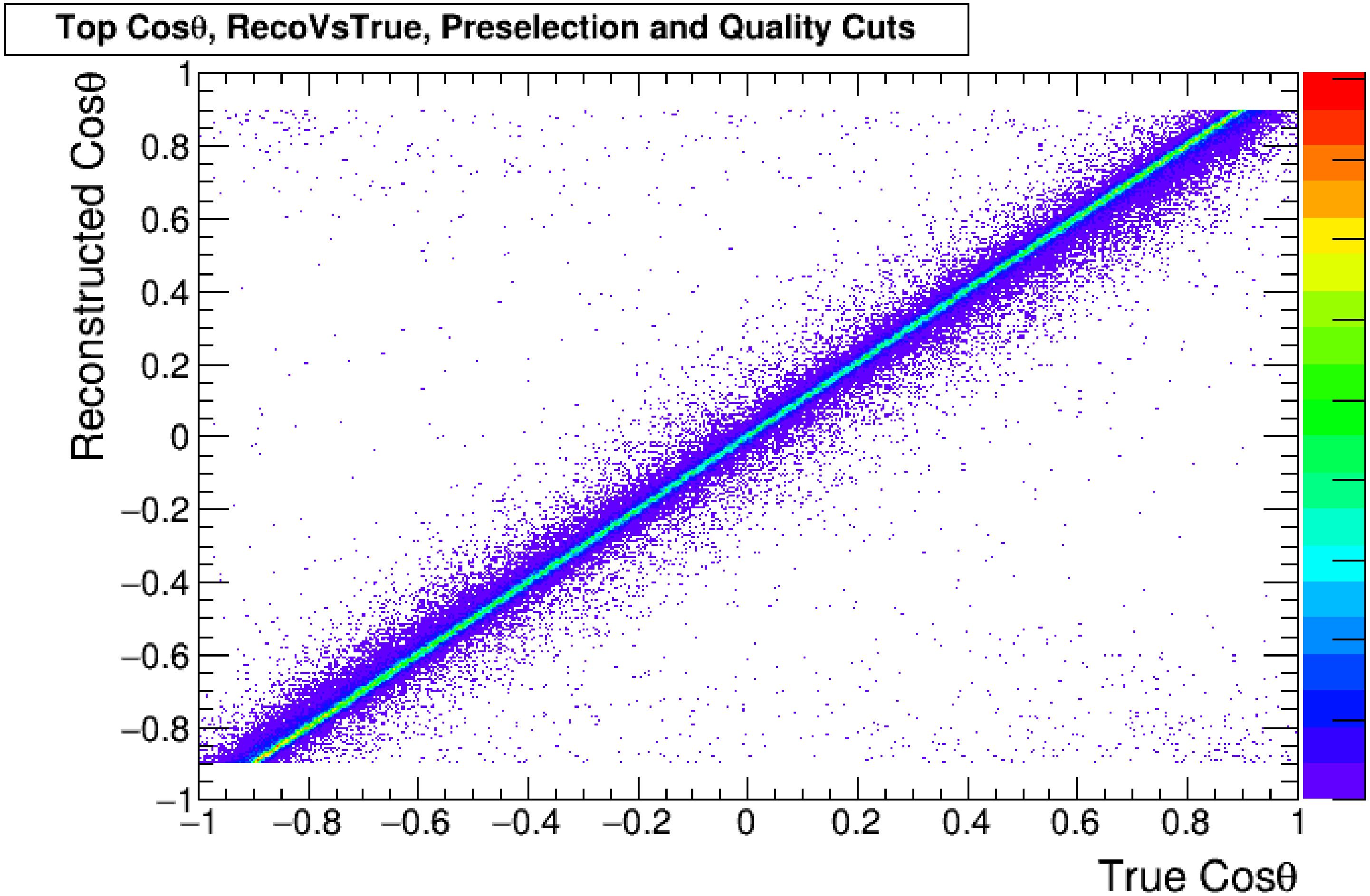
J. Fleisher et al, 2003, https://arxiv.org/pdf/hep-ph/0302259.pdf


Signal Efficiency For 1200<E<1450
0.5
0.4
0.3
0.2
0.1
0
−1
−0.5
0
0.5
1

Quality Cuts

- No Cuts: Clear problems seen in off diagonal regions

Quality Cuts



- Quality + Preselection cuts: Clear improvement seen, only diagonal elements remain

Quality Cuts

- Preselection Cuts (pre-existing to remove background events):
  - Visible Pt>200 GeV
  - Hadronic Top Energy>100 GeV
  - Leptonic B Jet Pt>20 GeV
  - -log(Y23)<7 && -log(Y34)<9
  - abs(Top Cosθ)<0.9
- Quality Cuts:
  - Hadronic Top Mass>100 GeV
  - Hadronic Top Pt>100 GeV
  - Leptonic B Jet Mass<100 GeV
  - 0.2<Collinearity of highest and next highest energy subjets<0.8
  - -log(Y23)>3
  - Pz Constraint from fit<100 GeV
- Currently use same cuts across full energy range
  - Need to tweak this slightly as some variables are energy dependent…

Variables currently used to train BDT

- Visible Energy and Pt
- Hadronic Fat Jet Energy and Pt
- Leptonic Fat Jet Mass
- Leptonic 1SubJettiness, 1SubJettiness/2SubJettiness
- Relative angle of the 3 subjets within hadronic fat jet
- Isolated lepton Energy, Pt and total momentum
- N Lepton candidates with E>30GeV
- Angular separation between lepton and hadronic fat jet
- -log(Y23)
- Major thrust
- Leptonic Top Energy
- Highest and next to highest btags