Top Mass with Multivariate Templates (Blessed)

J.Freeman, A.Galtieri, J.Lys, P.A.Movilla Fernández, I.Volobouev
LBNL

Outline

- Strategy
- Event Selection
- Signal subsamples
- Templates
- Likelihood
- Results
- Systematics
- Summary

Strategy

- Focus on $t\bar{t} \to bq\bar{q}\bar{b}l\nu$ channel with 1 or 2 SVX tags.
- Reduce jet energy systematic by calibrating the global jet energy scale (JES) with the W mass.
- Reduce statistical uncertainty by estimating the probability to pick correct jet-to-quark assignment ("permutation").
- Improve signal/background separation by using other kinematic variables in addition to the reconstructed invariant top mass.
- Kernel Density Estimation is employed to make multivariate templates.

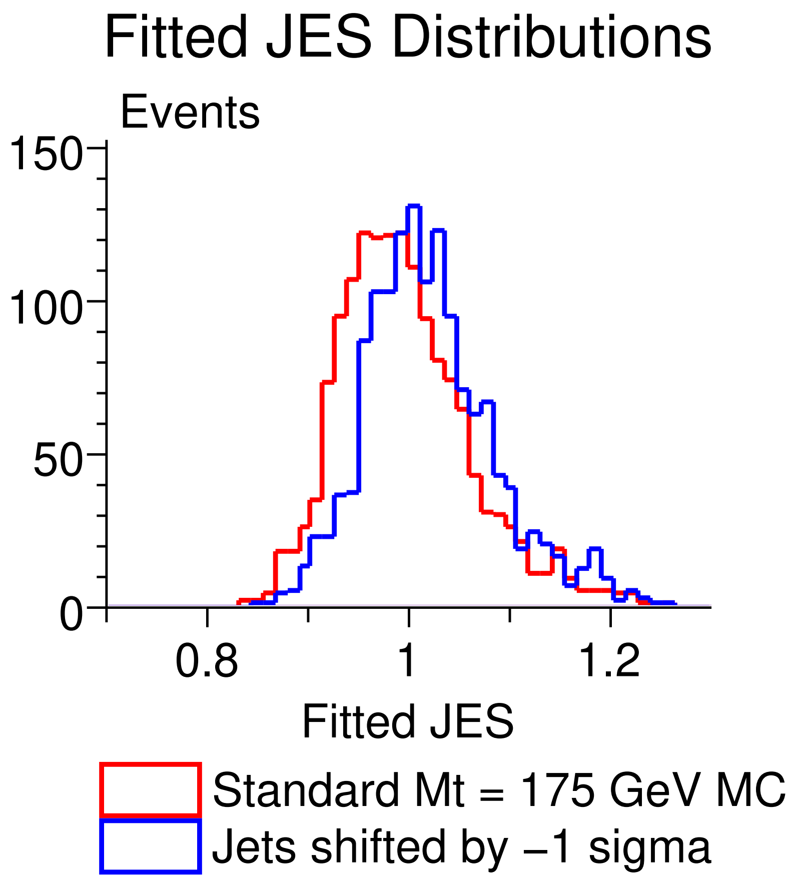
More information:

CDF Note 6970

`http://www-cdf.lbl.gov/~pmf/topmass/`

Kinematic Fitter

- JES is included into W mass kinematic fit using a Gaussian constraint (tunable parameter).
- All jet energies of an event are scaled with JES (permutation-by-permutation).
- For the correct permtation, scale shifts due to W mass constraint compensate on average systematic shifts ($\rightarrow$ plot).



- Apply LBL top specific corrections on top on generic level 7 corrections (i.e. only b quarks corrected):
  $\sim 10\%$ smaller statistical errors than JF corrections.

Data Set

Standard lepton+jets cuts applied plus additional citeria:

- At least ≥ 3.5 jets.
- Good silicon runs.
- Runs 163463, 163474 (wrong beam line) removed
- Run 164844 (wrong luminosity measurement) accepted
- Tridents removed
- At least one jet permutation which agree with SVX tagging information
- χ^2 of fit < 10
- Phoenix electrons not reconstructed and therefore not used in dilepton veto

$\Downarrow$

33 candidate events (≥ 3.5 jets)

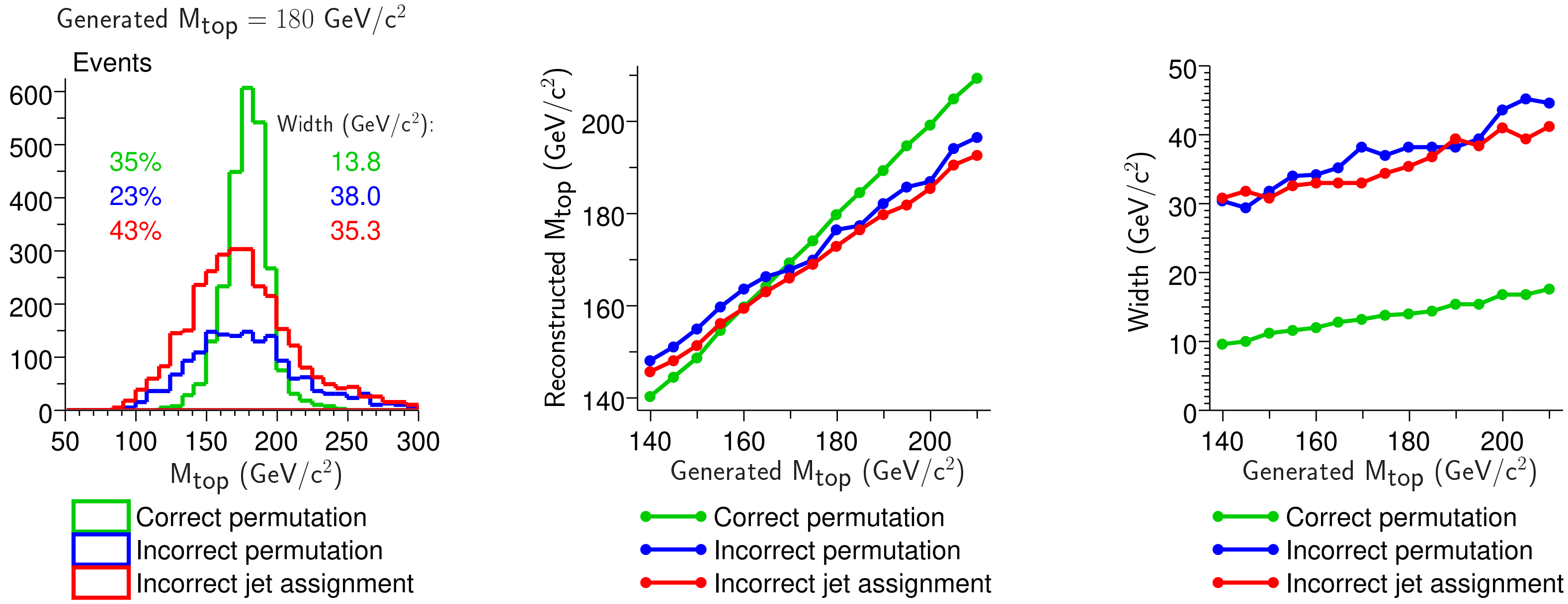
23 candidate events (≥ 4 jets)

(for JES constraint = 0.07)

Signal Subsamples

Three signal templates for the $t\bar{t} \to bq\bar{q}\bar{b}l\nu$ process:

- Good permuation samples GP (correct jet to quark assignment)
- Bad permutation samples BP (jets assigned to quarks but swapped)
- Incorrect jets IJ (at least one jet not assigned to a quark)



GP has much better mass resolution.

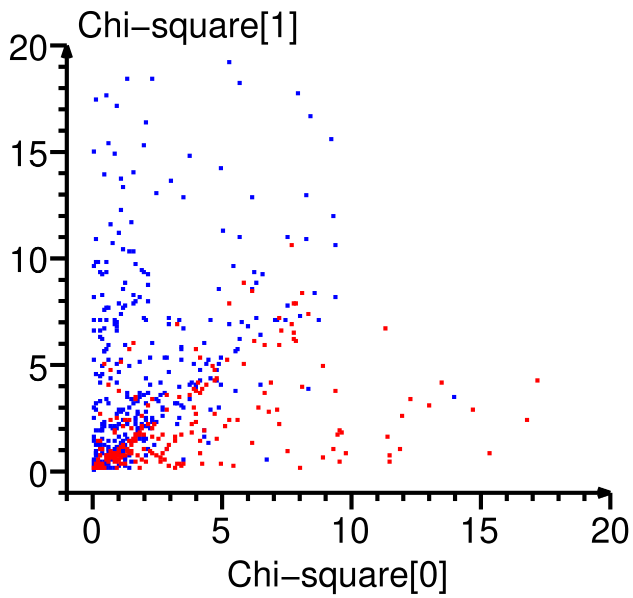
GP slope M_{rec} vs. M_{gen} is 1.

Signal Subsamples (2)

- Prediction of correct permutation probability p_{GP} inspired by a "diffusion model":

$$p_{\mathsf{GP}} = \frac{a_b}{\sum_{i=1}^{12} a_i \exp\left(-\frac{\chi_i^2-\chi_b^2}{w_e(\chi_i^2+\chi_b^2)}\right)}$$

$$w_e(y) = \exp(b_e + c_e y + d_e y^2)$$



- Diffusion constants a_i, b_e, c_e, d_e determined from MC (for a given permutation error type e)

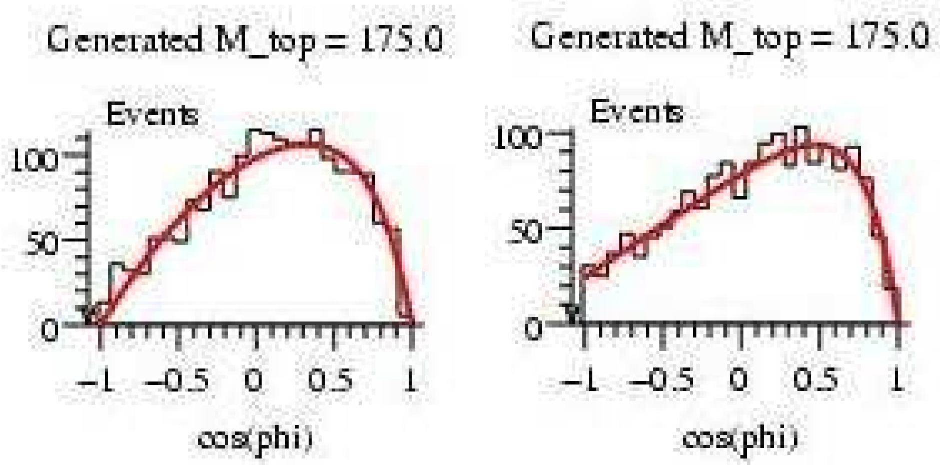
- Include additional kinematical information X with weak M_{top} dependence. Include a GP/$\overline{\mathsf{GP}}$ discriminator $\kappa = \frac{p_{X|\mathsf{GP}}}{p_{X|\overline{\mathsf{GP}}}}$ into p_{GP} definition:

$$p_{\mathsf{GP|X}} = \frac{\kappa p_{\mathsf{GP}}}{\kappa p_{\mathsf{GP}} + (1 - p_{\mathsf{GP}})}$$

Signal Subsamples (3)

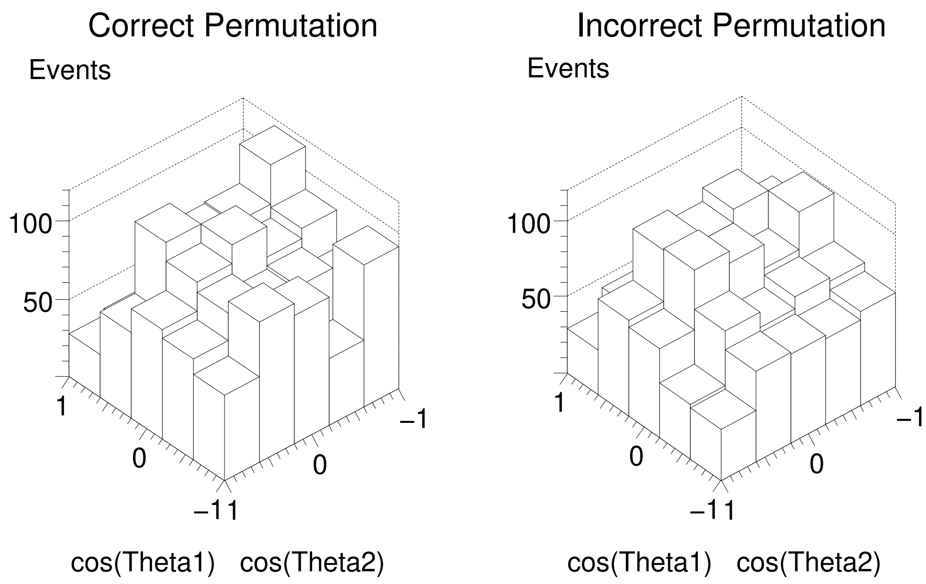
- Cosine of the angle ϕ between l and b_l in the rest frame of the leptonic W

$$\kappa_\varphi = A(1 - e^{-\frac{\cos(\varphi)-B}{C}})$$



- $t\bar{t}$ spin correlations $sc = \cos(\theta_1)\cos(\theta_2)$: θ_1: Angle between l and t spin quantization axis in t rest frame, θ_2: Angle between q and $\bar{t}$ spin quantization axis in $\bar{t}$ rest frame (use quark that makes smaller angle with hadronic b in parent W rest frame).

$$\kappa_{sc} = D(1 + E\cos(\theta_1)\cos(\theta_2))$$



Templates

- 2-dimensional templates based on reconstructed invariant mass and further kinematic variables.
- Use Kernel Density Estimation (KDE) as a nonparametric multivariate density reconstruction (see John's and Igor's talks for details).
- Templates are made for three types of signals (Herwig).

  GP, BP, IJ
- Templates are made for three types of backgrounds:
  - mistag: obtained using mistag matrix on $W4j$ Alpgen MC
  - non-W QCD: from data using $Iso > 0.2$, $\not\!\!E_{\mathrm{T}} > 20$ GeV events
  - $Wb\bar{b}$: from Alpgen MC (used for all other sources of background)

Signal/Background Discrimination

- Use histogram divergence measures to get an idea how useful a variable may be in separating signal from background:

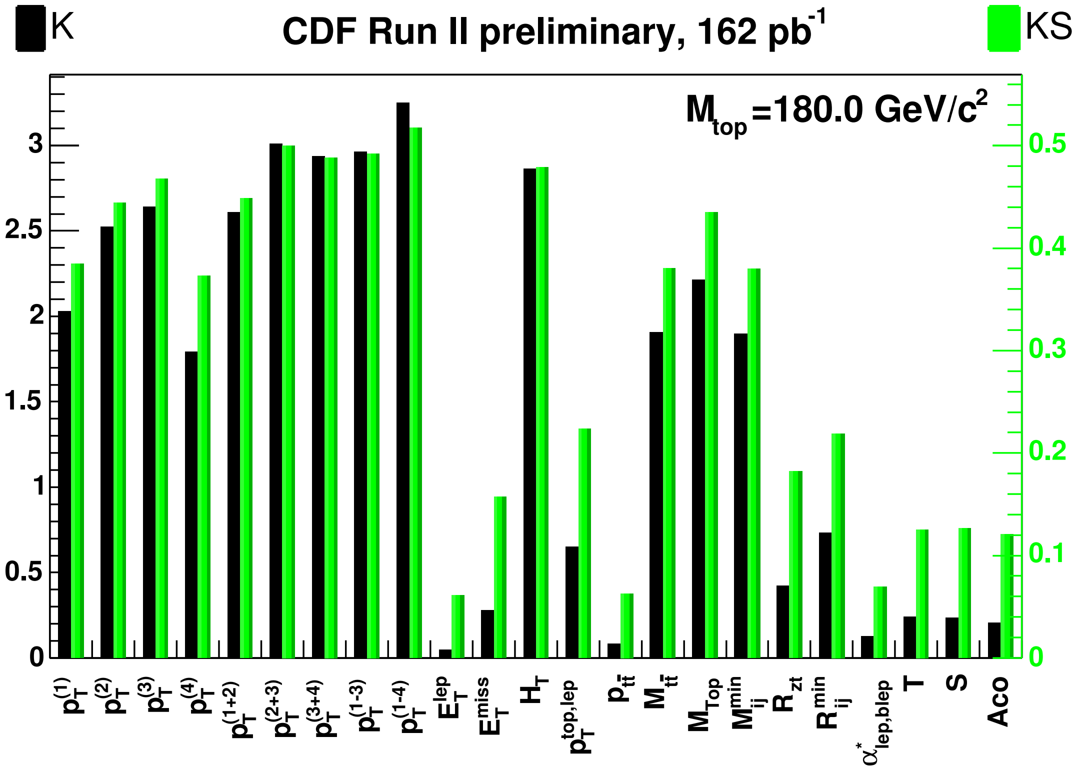
$$d(p_1, p_2) = \int \mathrm{d}x\ p_1(x) \times \log\left(p_1(x)/[\alpha \cdot p_2(x) + (1-\alpha)\cdot p_1(x)]\right)$$
$$d_s(p_1, p_2) = d(p_1, p_2) + d(p_2, p_1)$$

e.g: $\alpha = 0.5$: "K" divergence

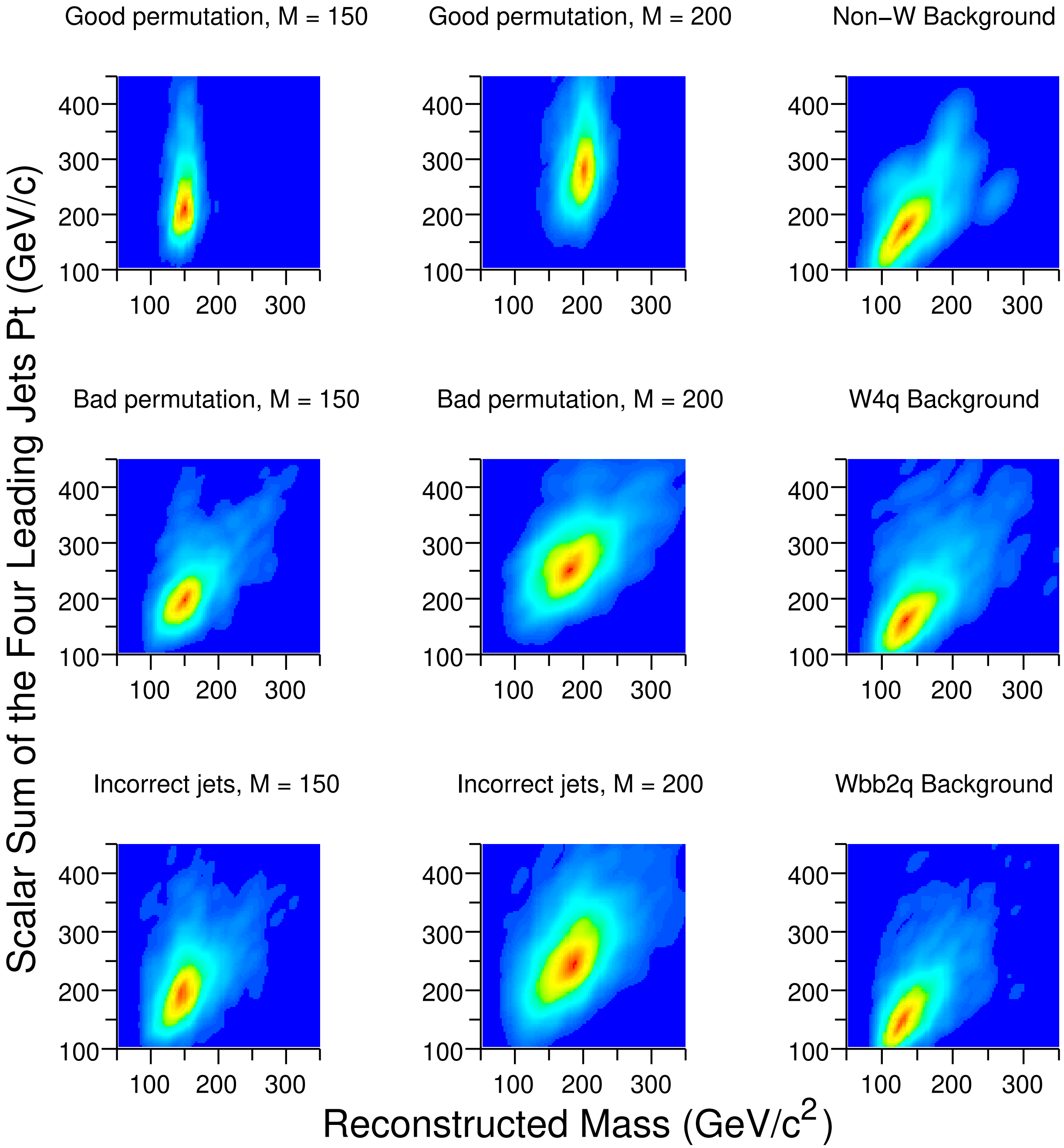
- "KS" distance

- Best discriminator: $p_T{}^{(1-4)}$ sum p_T of four leading jets.

- M_{top} only average performer



Templates (2)


Good permutation, M = 150
Good permutation, M = 200
Non−W Background
Bad permutation, M = 150
Bad permutation, M = 200
W4q Background
Incorrect jets, M = 150
Incorrect jets, M = 200
Wbb2q Background
Scalar Sum of the Four Leading Jets Pt (GeV/c)
Reconstructed Mass (GeV/c2)
100
200
300
400

Likelihood

$$L(m_t) = \Omega(f_b, N) \prod_{i=1}^{N} (f_b P_b(m_i, x_i) + (1 - f_b) P_s(m_i, x_i, m_t))$$

$$P_s(m, x, m_t) = (p_{\mathsf{GP}} S_{0,m_t}(m, x) + (1 - p_{\mathsf{GP}}) S_{1,m_t}(m, x))\, p_{\mathsf{CJ}} + (1 - p_{\mathsf{CJ}}) S_{2,m_t}(m, x)$$

$$P_b(m, x) = \sum_{\text{bg types}} a_j B_j(m, x), \qquad \sum_{\text{bg types}} a_j = 1$$

N is the number of observed events.

m_i is the top mass of event i. x_i symbolizes all other template variables.

P_s and P_b are the signal and background densities.

f_b is the background fraction.

$\Omega(f_b, N)$ is the background constraint. For now, $\Omega(f_b, N) = 1$.

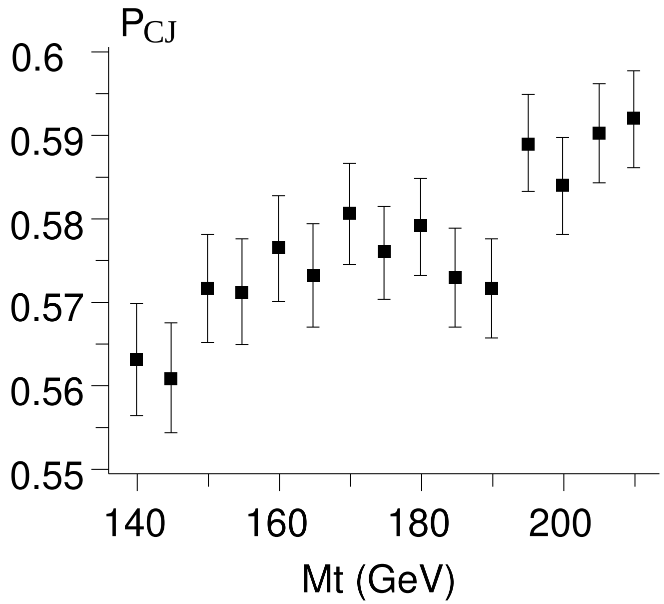
B_j templates for the different backgrounds with composition coefficients a_j.

S_{0,m_t}, S_{1,m_t}, S_{2,m_t} are the three signal templates.

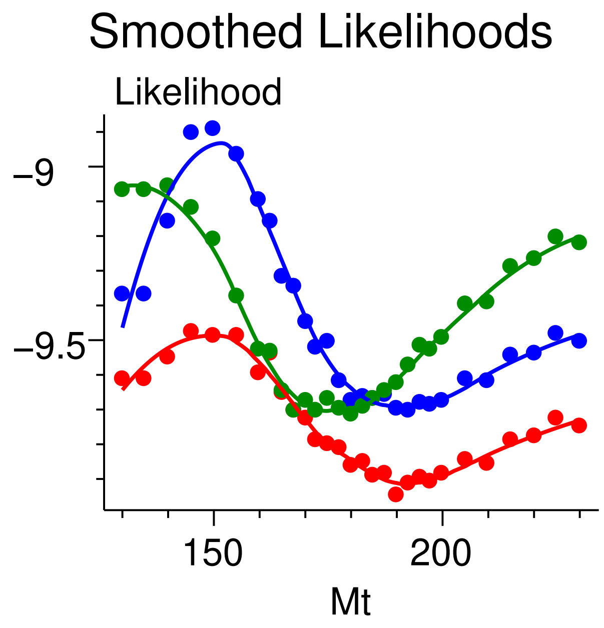
p_{GP}, p_{CJ} are probabilities for GP and CJ=GP+BP.

Likelihood (2)

- Likelihood normalization is exact only if p_{GP}, p_{CJ} are independet from m_t, x.
- Only weak dependence on top mass observed.



- Likelihood is discrete in top mass m_t. Use local polynomial regression to interpolate and smooth likelihood curves.
- Interpolation is performed for each event separately, resulting log likelihood curves are then added to obtain the likelihood of whole sample.



Background

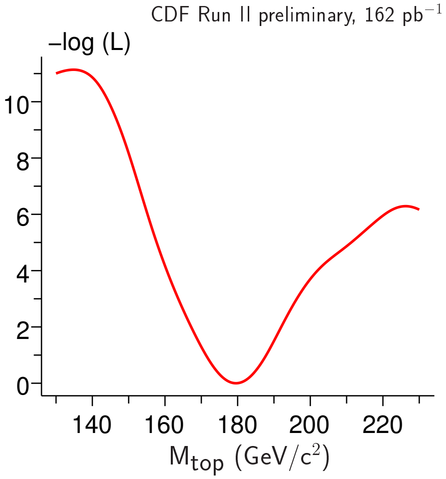
- Background composition follows $t\bar{t}$ cross section measurement and is fixed.
- Overall signal/background ratio allowed to float freely.
- Background scaling to 3.5 jets is done from 3 jets assuming the same 3.5/3 jets ratio. Efficiencies of additional cuts are folded in.

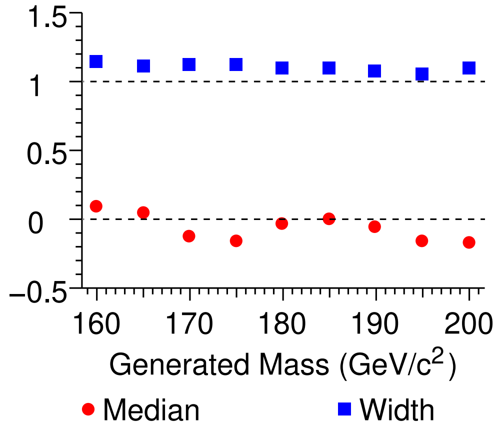
Background Summary

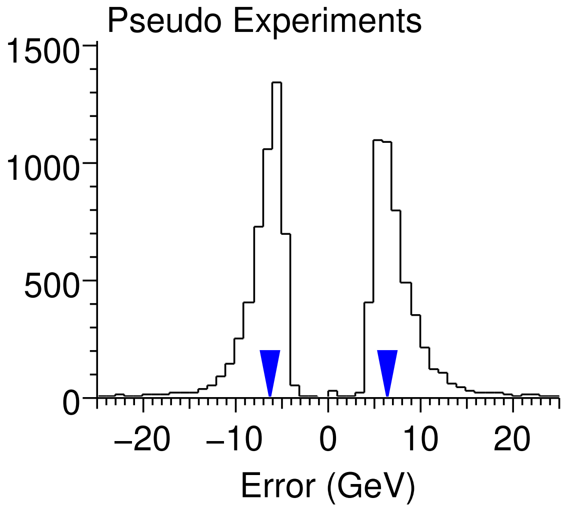
Background	W+3.5jets	W+$\geq$4jets	Total
mistags	0.90±0.15	0.85±0.13	1.75±0.28
W $b\bar{b}$, W $c\bar{c}$	0.83±0.29	0.60±0.22	1.43±0.51
Wc,WW/WZ etc.	0.41±0.13	0.17±0.07	0.58±0.20
Single top	0.21±0.04	0.09±0.02	0.30±0.06
Sum of above 3	1.45±0.34	0.86±0.24	2.31±0.58
non-W (QCD)	1.24±0.22	1.32±0.25	2.56±0.47
Total Backg	3.59±0.43	3.03±0.37	6.62±0.80
Events Observed	11	29	40
Have a mass fit	10	23	33

Results

Fit to data using M_t and $p_T^{(1-4)}$ (JES constraint 0.07):





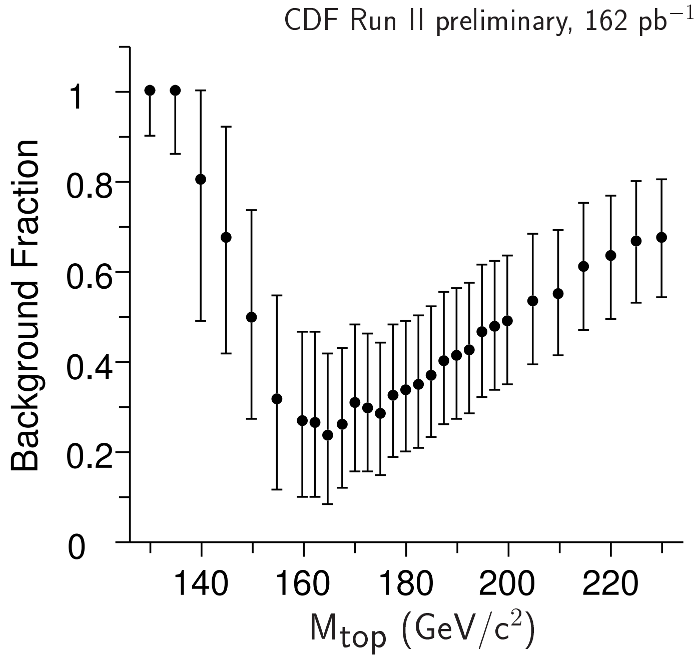


- From PE:

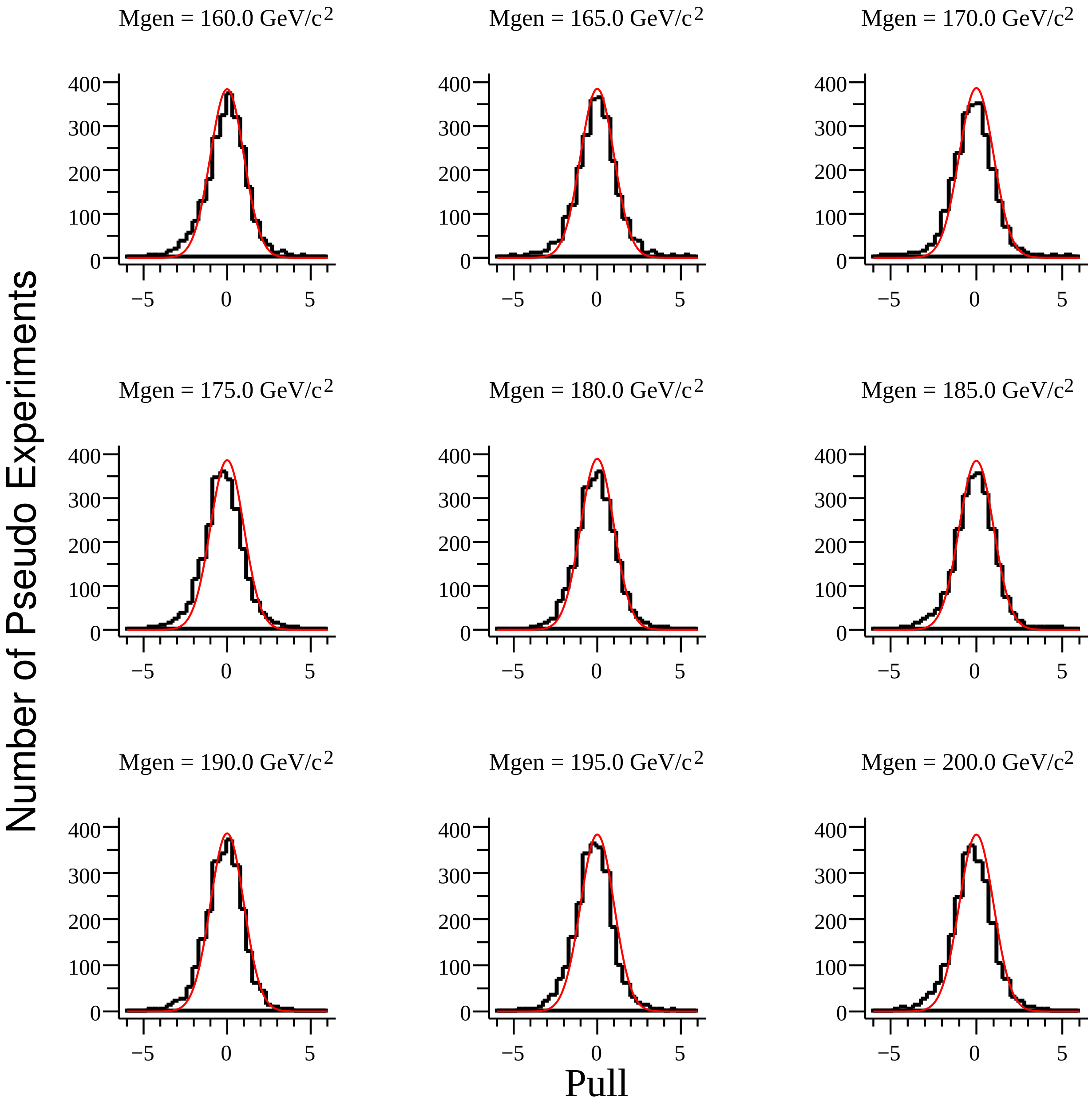
  Pulls ok, average pull width is 1.10
- Top mass + corrected statistical error:

  $m_t = 179.6^{+6.4}_{-6.3}$
- Most probable background fraction:

  $f_b = 0.34 \pm 0.14$

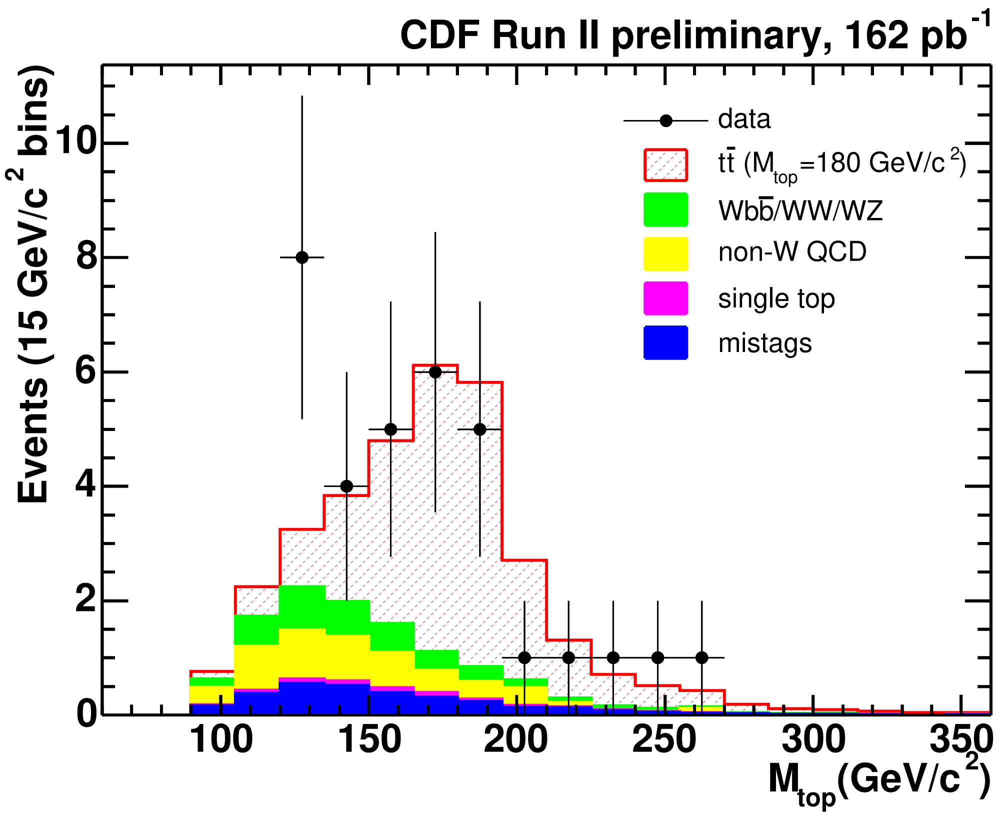


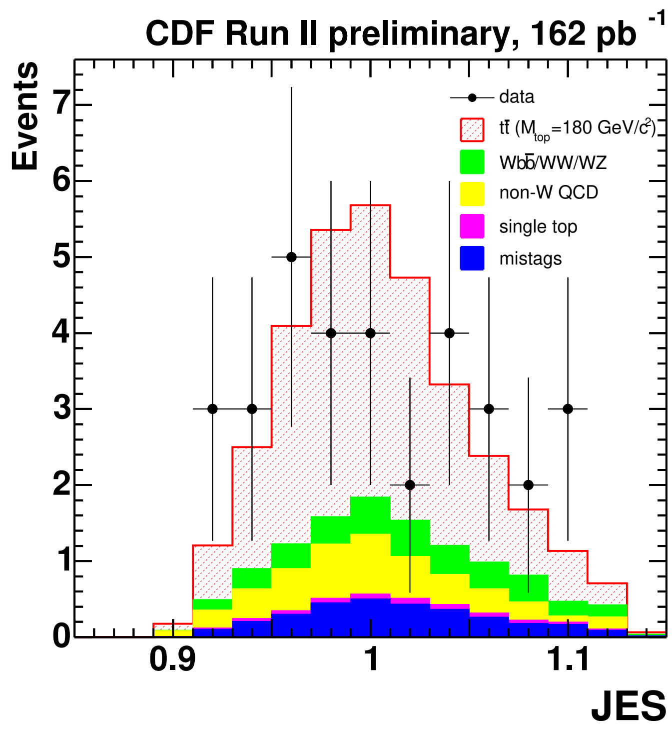
Pulls


Mgen = 160.0 GeV/c2
Mgen = 165.0 GeV/c2
Mgen = 170.0 GeV/c2
Mgen = 175.0 GeV/c2
Mgen = 180.0 GeV/c2
Mgen = 185.0 GeV/c2
Mgen = 190.0 GeV/c2
Mgen = 195.0 GeV/c2
Mgen = 200.0 GeV/c2
Number of Pseudo Experiments
Pull
0
100
200
300
400
−5
0
5

Data vs. Signal + Background

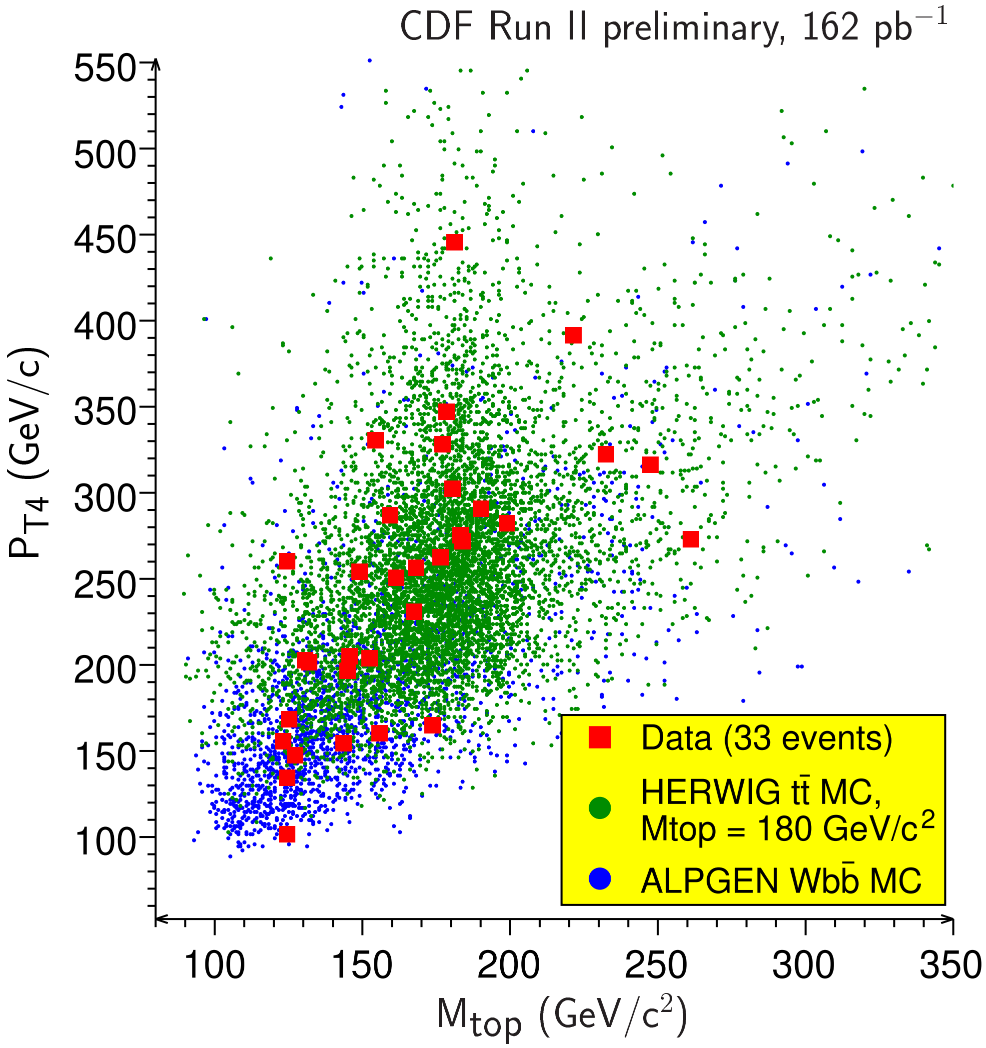
Fit to data using M_t and $p_T^{(1-4)}$ (JES constraint 0.07):



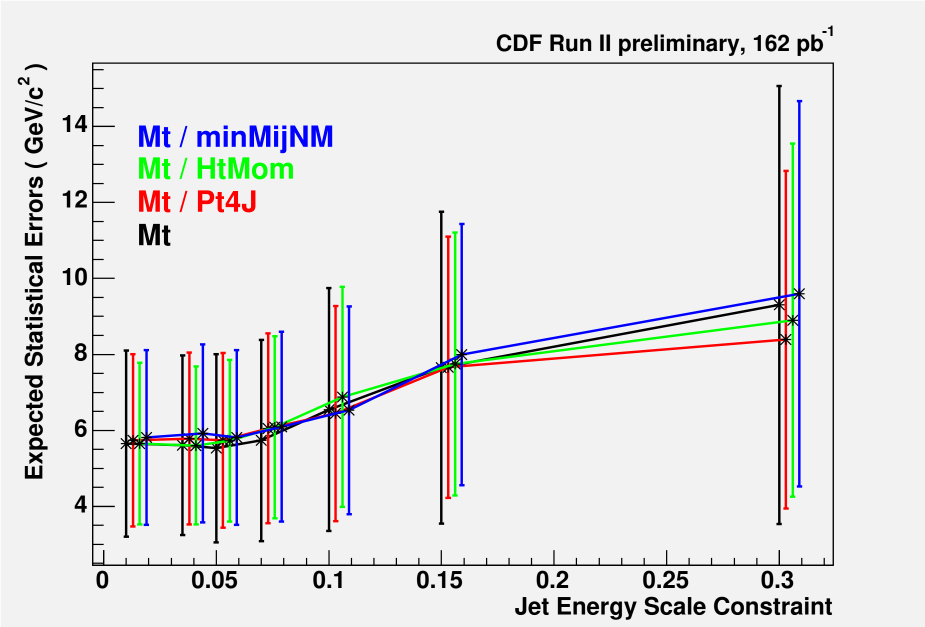


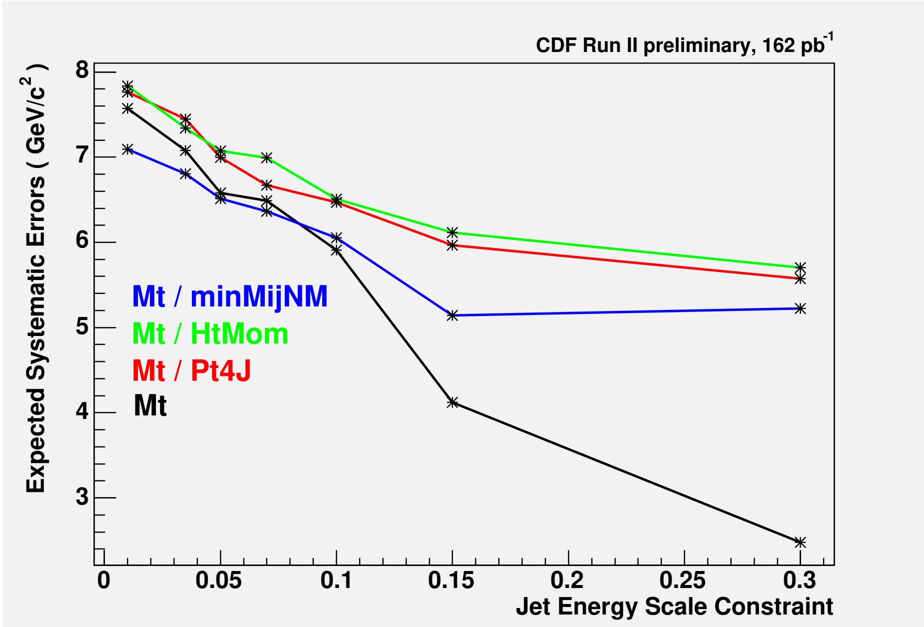
Data vs. Signal + Background (2)

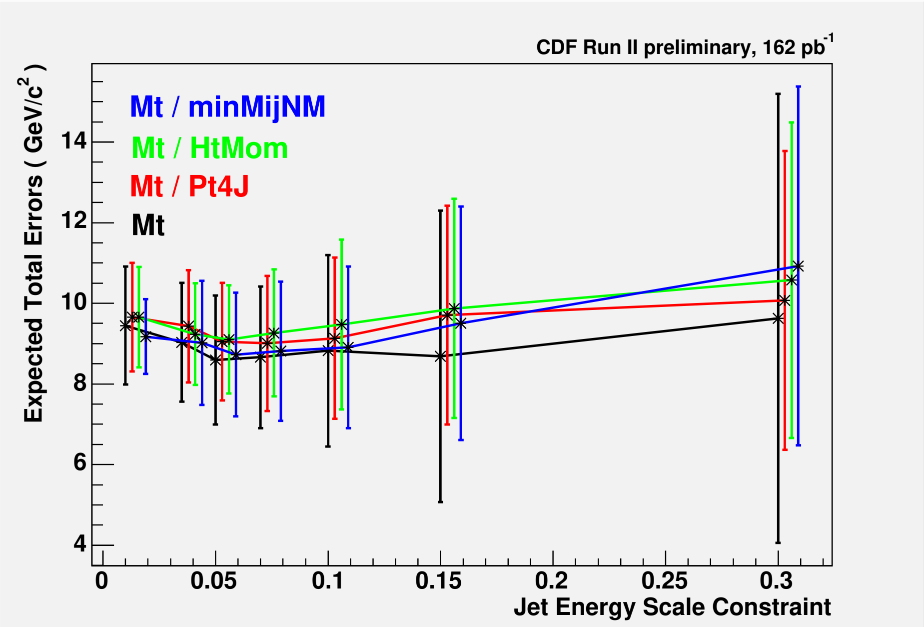
Fit to data using M_t and $p_T^{(1-4)}$ (JES constraint 0.07):



Statistical + Systematic Errors







- M_t and $p_T^{(1-4)}$: optimal JES constraint value is 0.07.

Systematics

- Jet energy systematics according to Jet Corrections group (CDF Note 6565)

$$\Delta M_{\text{top}} = \pm 6.7 \text{ GeV/c}^2$$

- Generator dependence:

	$M_{\text{top}}(\text{GeV/c}^2)$	ΔM_{top} (GeV/c^2)
Pythia tune A (default)	174.5 ± 0.2	
Herwig (default)	174.4 ± 0.3	-0.1
Herwig (no spin correlations)	174.7 ± 0.2	$+0.2$
Pythia Run I	173.9 ± 0.2	-0.6

Systematics (ISR/FSR)

- ISR:

	$M_{\text{top}}(\text{GeV}/c^2)$	ΔM_{top} (GeV/c^2)
Pythia tune B	174.9 $\pm$ 0.3	$+0.4$
Pythia no ISR	174.0 $\pm$ 0.3	-0.5
Pythia less ISR	174.7 $\pm$ 0.3	$+0.2$
Pythia more ISR	174.5 $\pm$ 0.3	$+0.0$

- FSR:

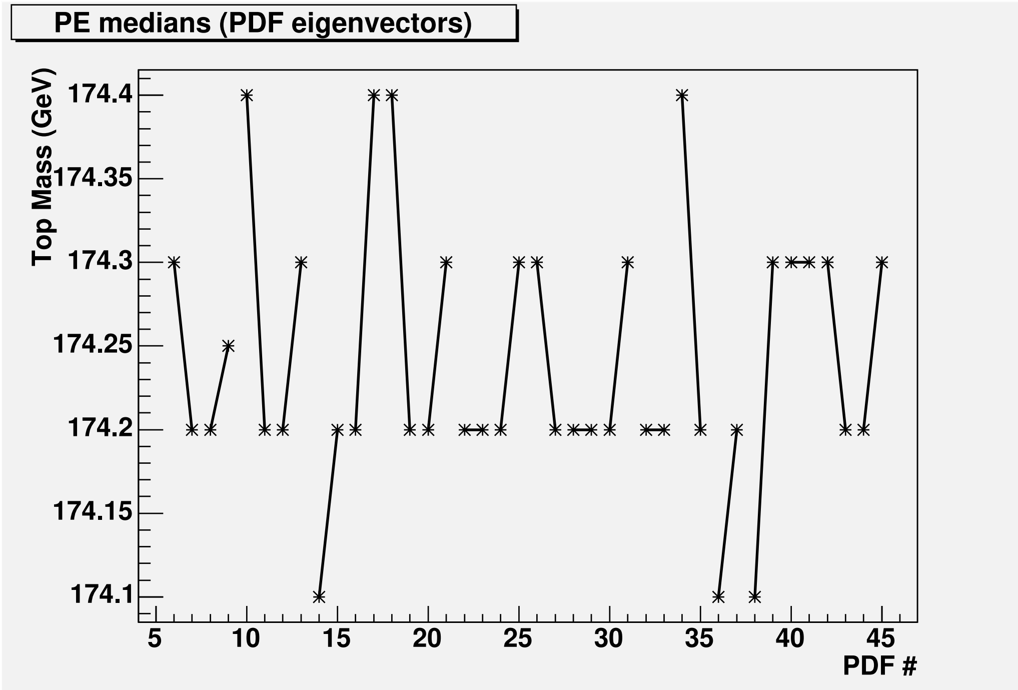
	$M_{\text{top}}(\text{GeV}/c^2)$	ΔM_{top} (GeV/c^2)
Pythia less FSR	174.9 $\pm$ 0.3	$+0.4$
Pythia more FSR	175.1 $\pm$ 0.3	$+0.6$
Pythia no hard gluon from b quark	173.5 $\pm$ 0.3	-1.0

Systematics (PDF)

- PDF:

	M_{top}(GeV/c^2)	ΔM_{top} (GeV/c^2)
MRST72	174.7 $\pm$ 0.3	$+0.2$
MRST75	174.8 $\pm$ 0.3	$+0.3$
MRST72-MRST75	—	-0.1
Variation of PDF eigenvectors	—	$+0.3$

20 CTEQ6M eigenvectors: Combine $(M_+ - M_-)/2$ values in quadrature:



Systematics (Background Shape)

- Variation in background fractions:

Background Template	Shift Direction (σ)	ΔM_{top} (GeV/c^2)
$Wb\bar{b}$	$+1$	$+0.1$
	-1	$+0.0$
mistags	$+1$	$+0.0$
	-1	$+0.2$
QCD	$+1$	$+0.2$
	-1	$+0.1$

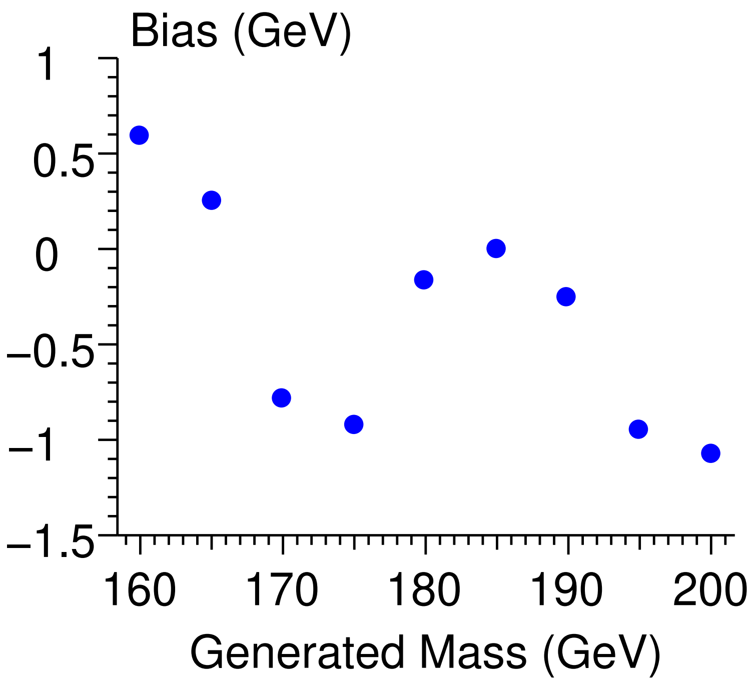
- Variation due to variations in Q^2 used to simulate $Wb\bar{b}$:

Q^2	M_{top} (GeV/c^2)
$(2M_W)^2$	174.1 ± 0.3
M_W^2	173.9 ± 0.3
$(M_W/2)^2$	174.0 ± 0.3
$M_W^2 + (p_{\mathrm{T}}^W)^2)$	173.9 ± 0.3

Systematics (Tagging + Fitting)

- We don't use the scale factor since we handle one and two tags in a similar way. Only difference is p_{GP}.
  Perform PE in which the probability to pick up an event with double SVX tag was reduced by factors 0.74, 0.81, and 0.88, according to b tag scale factor 0.81 ± 0.07 (CDF 6904).
  $\Rightarrow$ largest shift $\Delta M_{\mathrm{top}} = +0.3\mathrm{GeV}/c^2$

- Add RMS of bias:
  $\Rightarrow \Delta M_{\mathrm{top}} = \pm 0.7\mathrm{GeV}/c^2$



Systematics (Summary)

Use largest systematic shift from each category:

Systematic	ΔM_{top} (GeV/c^2)
Jet Energy	6.7
Generators	0.2
ISR	0.2
FSR	0.6
PDF	0.6
Background Shape	0.4
b Tagging	0.3
Fitting Procedure	0.7
Total	6.8

Summary

- Top mass from multivariate templates (blessed):

$$M_{\text{top}} = 179.6^{+6.4}_{-6.3}(\text{stat.}) \pm 6.8(\text{syst.})\ \text{GeV}/c^2$$

- Most probable background fraction:

$$f_b = 0.34 \pm 0.14$$

Our M_{top} result is consistent with two further blessed CDF II results:

- Dynamical Likelihood Method:
  $M_{\text{top}} = 177.8^{+4.5}_{-5.0}(\text{stat.}) \pm 6.2(\text{syst.})\text{GeV}/c^2$

- Template Method:
  $M_{\text{top}} = 174.9^{+7.1}_{-7.7}(\text{stat.}) \pm 6.5(\text{syst.})\text{GeV}/c^2$