Top/EWK group meeting 7/13/01
Ambra Gresele
(+ A. Castro)

All-hadronic $t\bar{t}$:
Preliminary study of top mass reconstruction

- Motivation
- 3C mass fit
- specific kinematic selection
- likelihood fit
- conclusions

Motivation: pros and cons

We want to develop the procedure for the measurement of the top mass in all-hadronic $t\bar{t}$ events. In fact, for such a measurement:

+ we do not need to know precisely the b-tag efficiency;
+ we do not need a precise evaluation of the background (we leave this to the likelihood fit);
+ we can extract the mass spectrum for the background from untagged events (as done in Run I, we expect no bias);

However:

– we need to have Run II jet corrections;

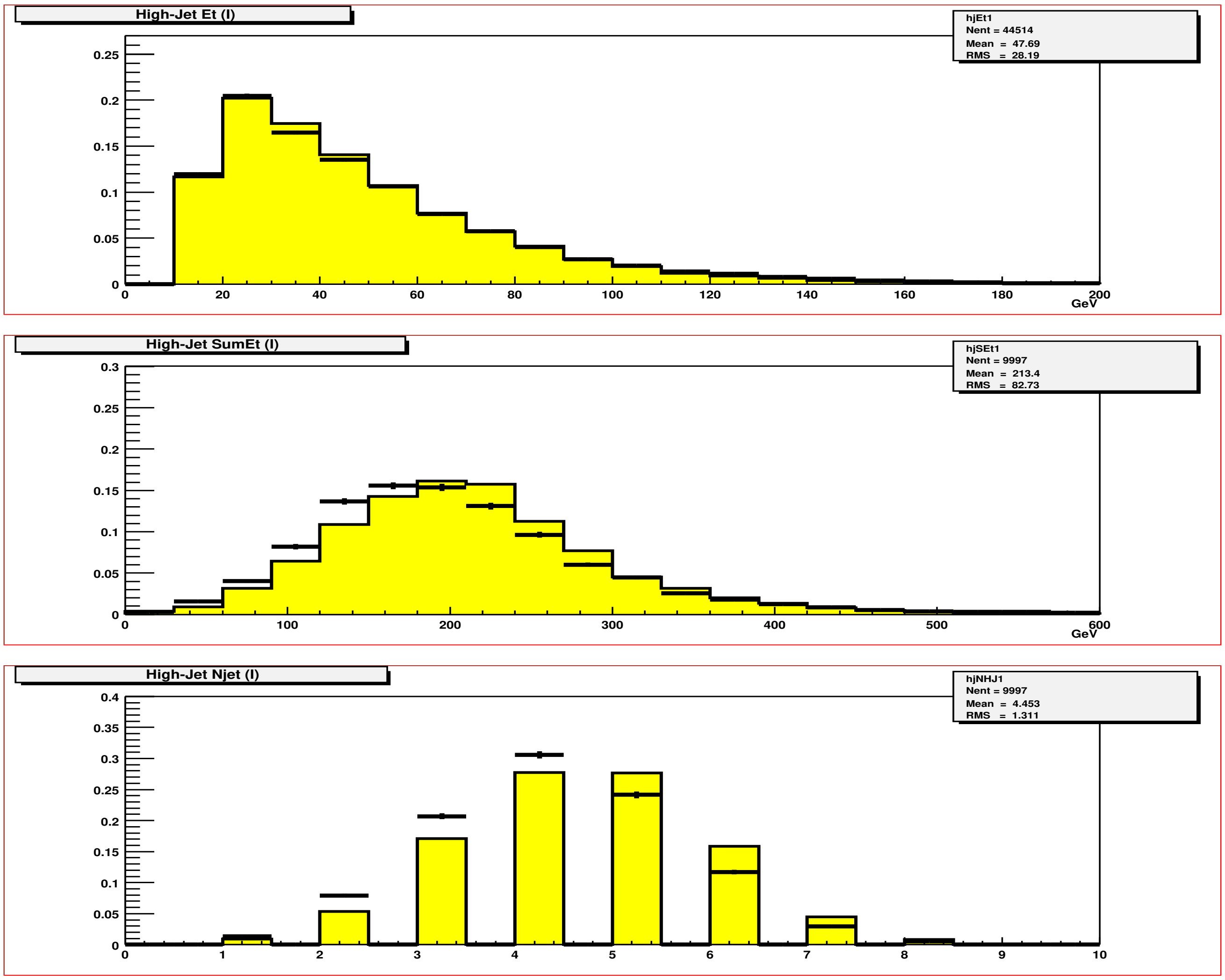
– we need to have Run II process-specific corrections.

$\boxed{t\bar{t}\ \textbf{Monte Carlo samples: I}\ vs\ \textbf{II}}$

$t\bar{t}$ events generated with **HERWIG** and simulated with **cdfSim** (3.15.0).

We have compared this simulator to **qfl** (events generated with $\sqrt{s} = 2.0$ TeV):

- E_T, $\sum E_T$ (and similar) spectra agree quite well;
- there is some discrepancy instead on N_{jet}.



Histogram: **qfl** (and Run I software)

Points: **cdfSim** (and Run II software)

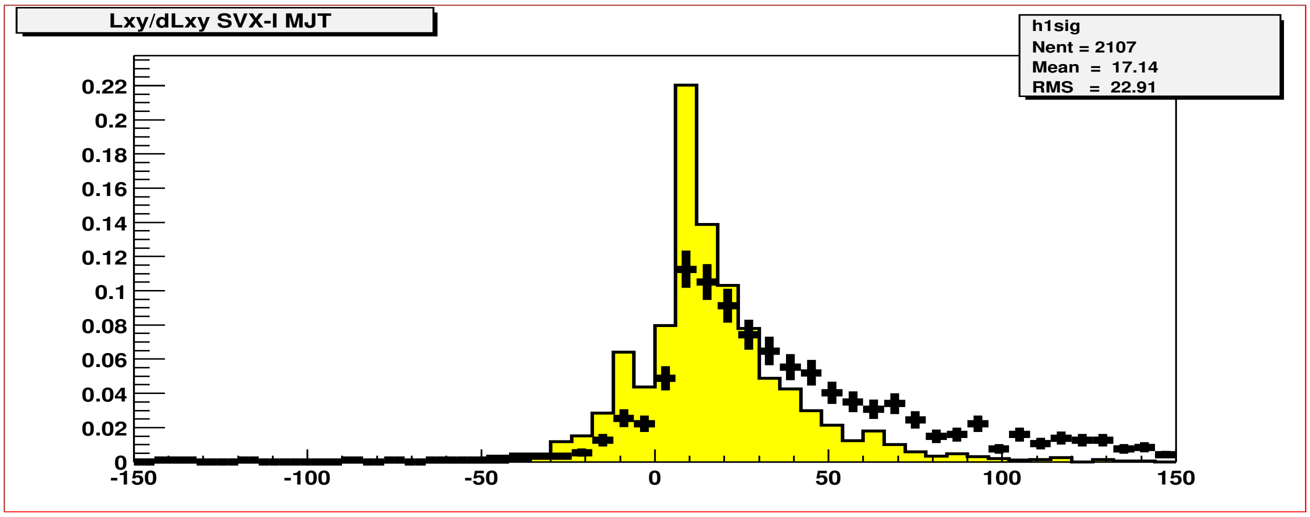
SecVtx b-tag

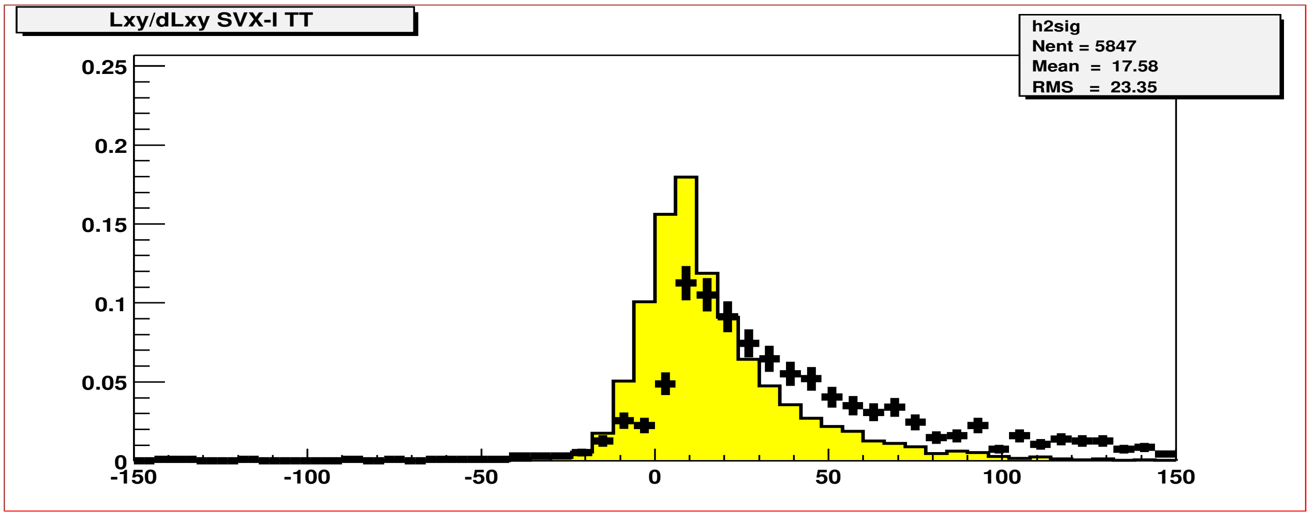
The (standard, 3.16.0) SecVtx-II seems to behave reasonably: we compared it to Secvtx-I for Run I MC $t\bar{t}$ events and for Run I multijet events.

SecVtx-II has:

- slightly larger values for L_{xy}
- slightly smaller values for δL_{xy}
- $L_{xy}/\delta L_{xy}$ shifted towards higher values, so the efficiency is higher: $\epsilon \approx 60\%$ instead of $\approx 40\%$.

We assume that the larger value of ϵ will be sustained by the estimates based on the data themselves.





Histogram: **SecVtx-I**, Points: **SecVtx-II**

Top: w.r.t. multijet events; Bottom: w.r.t. $t\bar{t}$ events

3C Mass fit with MINUIT

We want to reconstruct the kinematics of

$$p\bar{p} \to t\bar{t} \to W^{+}bW^{-}\bar{b} \to j_1 j_2 b j_3 j_4 \bar{b}$$

using the four-momenta of the 6 leading jets (using Run I jet corrections for now).

Measurements provide 3 constraints:

- $M_{j_1 j_2} = M_W$
- $M_{j_3 j_4} = M_W$
- $M_{j_1 j_2 b_1} = M_{j_3 j_4 b_2}$

We mimimize a 3C χ^2:

$$\chi^2 = \frac{(M_{jj} - M_W)^2}{\sigma^2_{M_W}} + \frac{(M_{jj} - M_W)^2}{\sigma^2_{M_W}} + \frac{(M_{jjb} - M_{jjb})^2}{\sigma^2_{M_t}}$$

($\sigma_{M_W} = 2.1$, $\sigma_{M_t} = 2.5$ GeV/c^2, as in PRD but not essential).

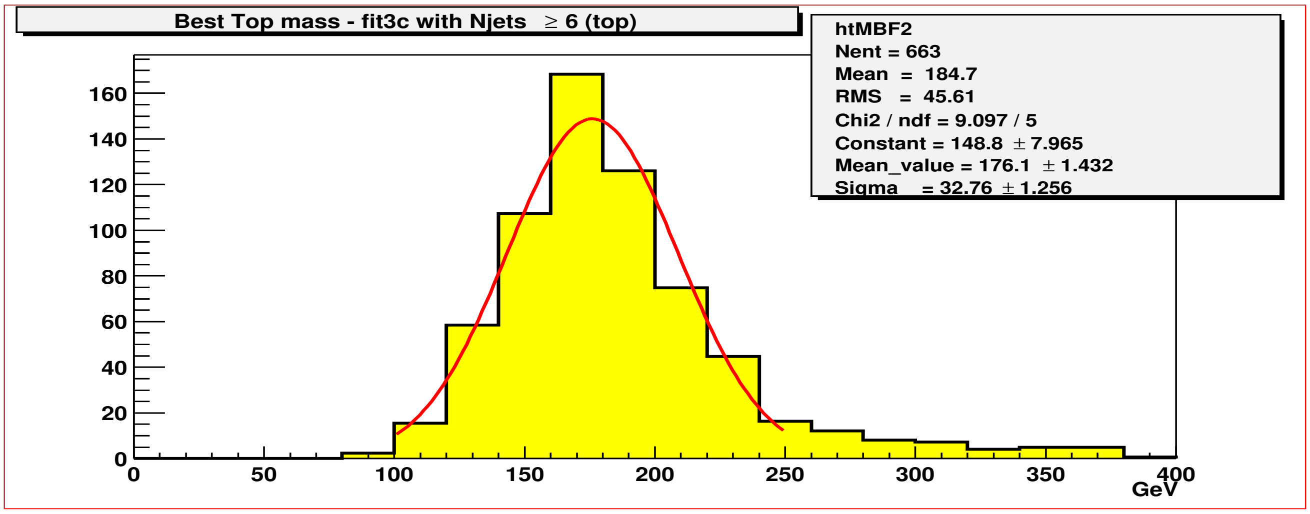
We require ≥ 1 (0) btag and this means (using the 6 leading jets) 30 (90) permutations and pick the lowest-χ^2 one.

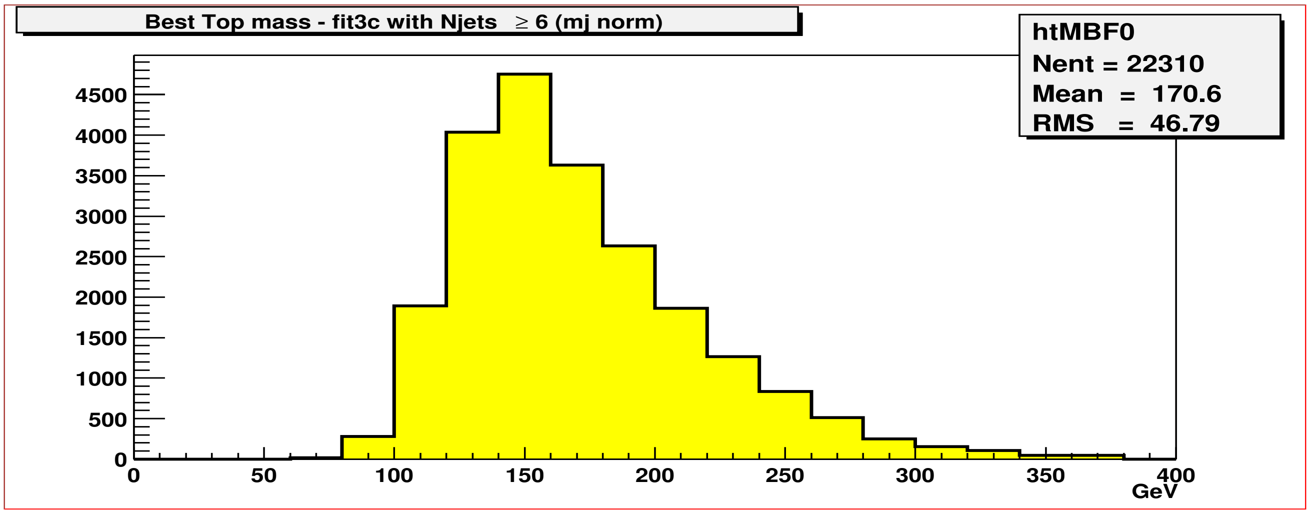
$\boxed{\textbf{Mass fitting}}$

For the time being, we use Run-I JTC96 corrections, and select events with $N_{jet} \geq 6$ ($E_T \geq 15$ GeV). We want to compare:

- the triplet mass for signal events, i.e. $t\bar{t}$ events having **at least one btag**
- the triplet mass for background events, i.e. multijet events (from Run I)

The peak in the background distribution is well separated from the peak of the signal:


Best Top mass - fit3c with Njets ≥ 6 (top)
htMBF2
Nent = 663
Mean = 184.7
RMS = 45.61
Chi2 / ndf = 9.097 / 5
Constant = 148.8 ± 7.965
Mean_value = 176.1 ± 1.432
Sigma = 32.76 ± 1.256
GeV


Best Top mass - fit3c with Njets ≥ 6 (mj norm)
htMBF0
Nent = 22310
Mean = 170.6
RMS = 46.79
GeV

$\boxed{\textbf{Kinematical selection}}$

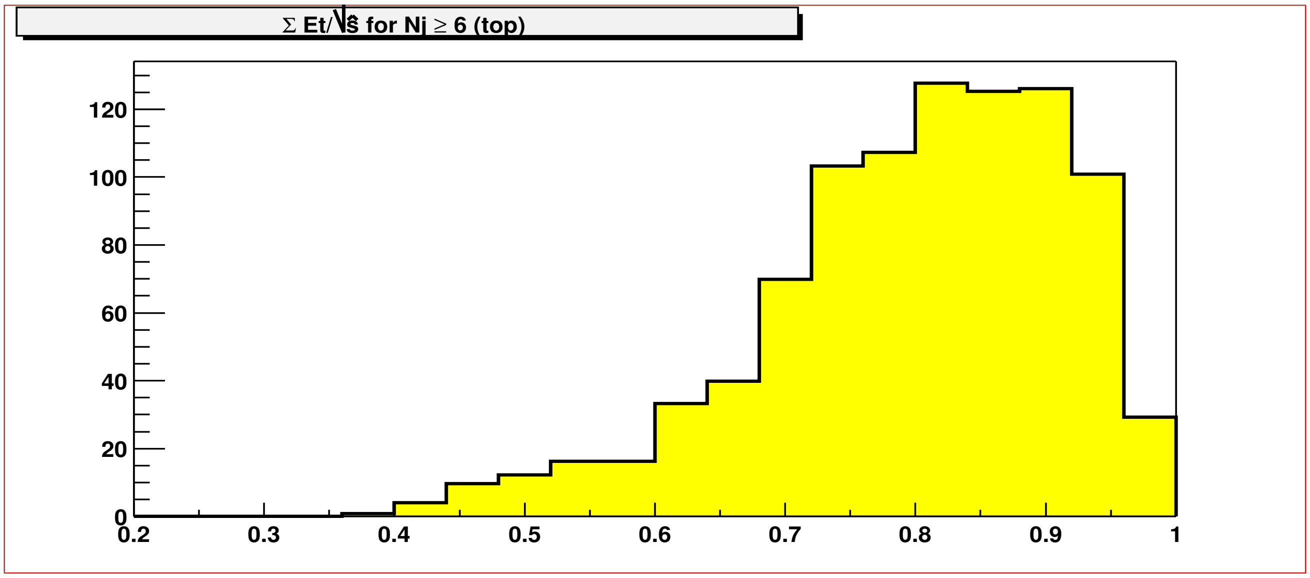
We want to develop a kinematical selection which keeps the two distributions well separated. For instance it is not appropriate to cut strongly on $\sum E_T$ or else the peak of the background spectrum moves where the signal is.

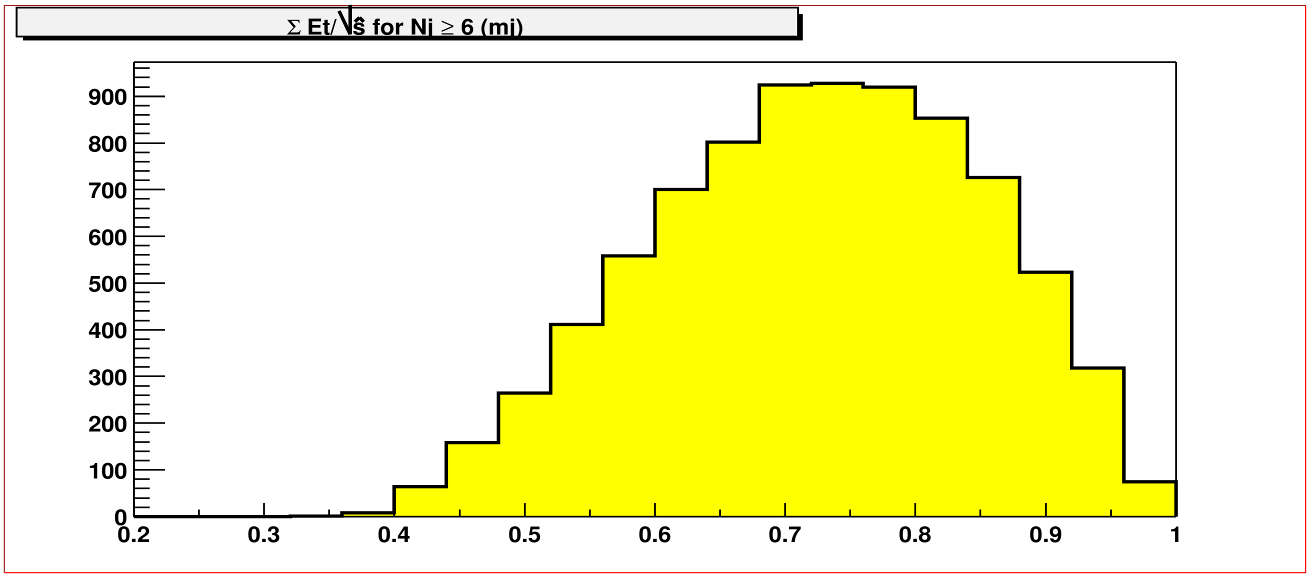
Appropriate variables are relative and direction quantities like:

- $\sum E_T/\sqrt{\hat{s}} \geq 0.75$
- $|\cos\theta^*| \leq 0.5$ (for the two leading jets)
- $\sum |P_z|/\sum E_T \leq 0.9$

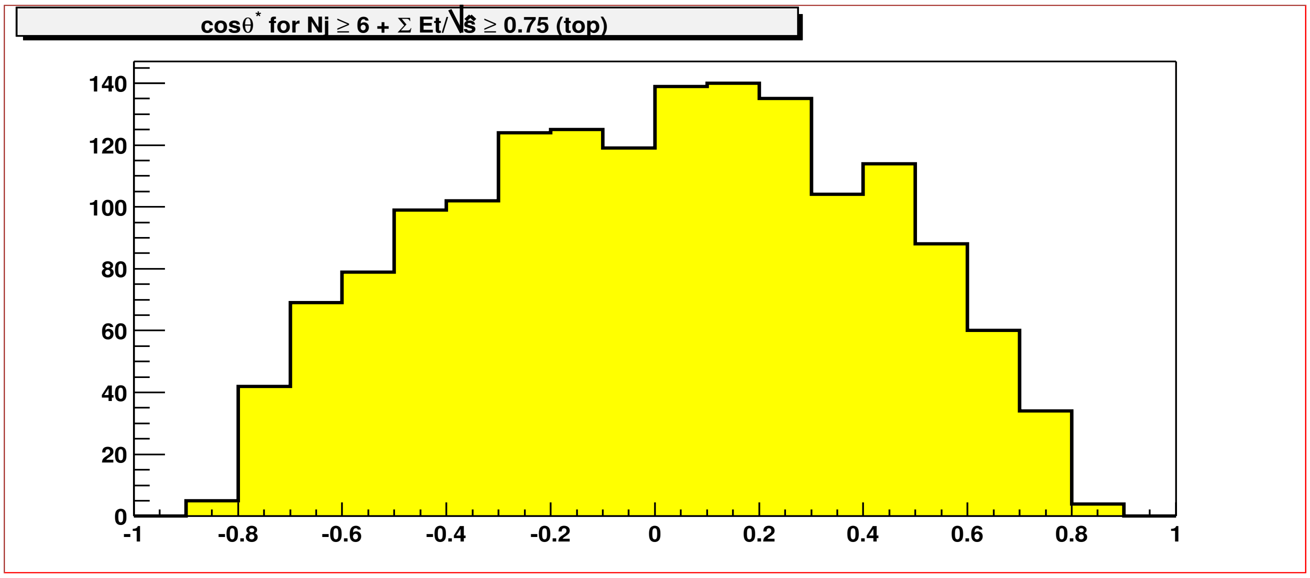
The cuts chosen provide a good background rejection: a gain of a factor of 3 in S/N, with $\epsilon \approx 30\%$.

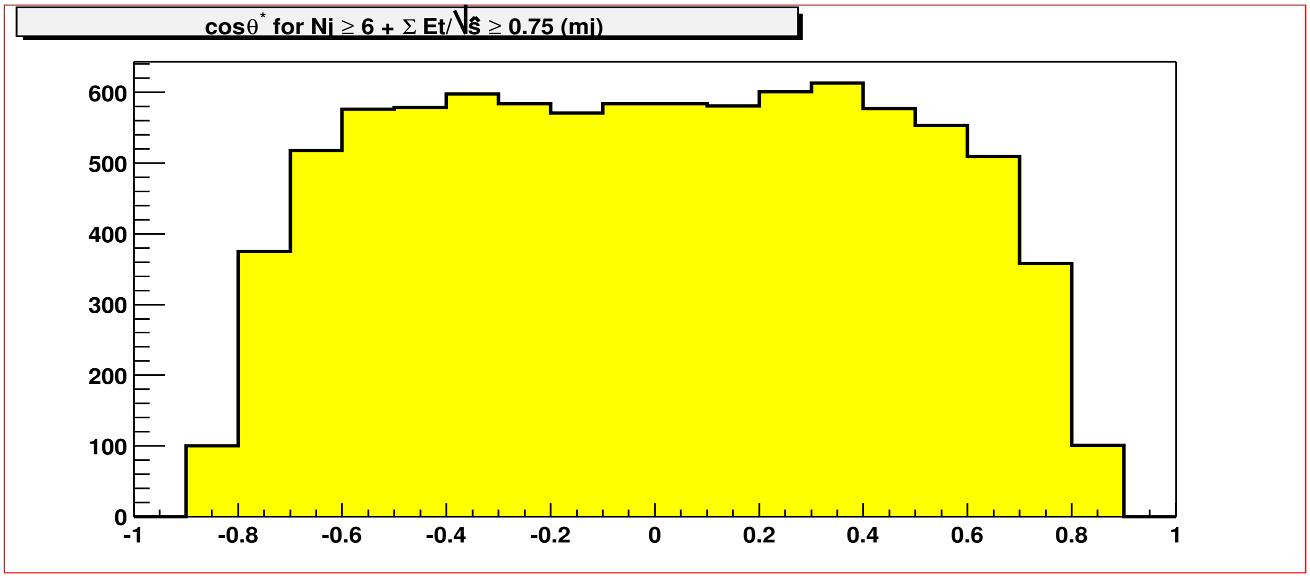
$$\boxed{\sum E_T/\sqrt{\hat{s}}}$$



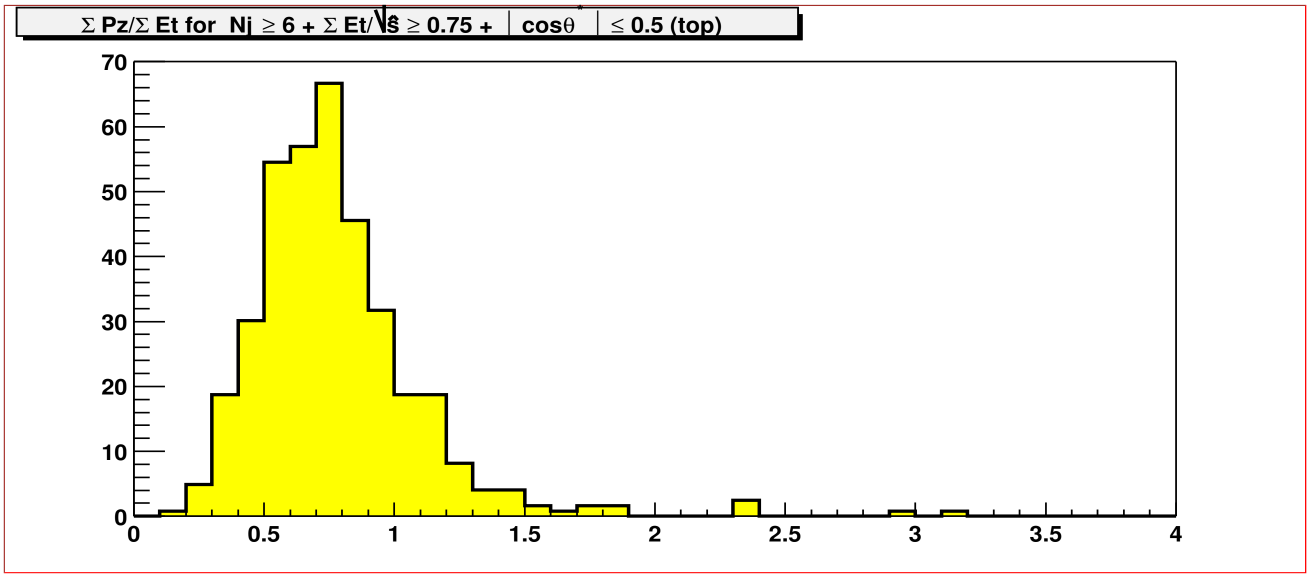


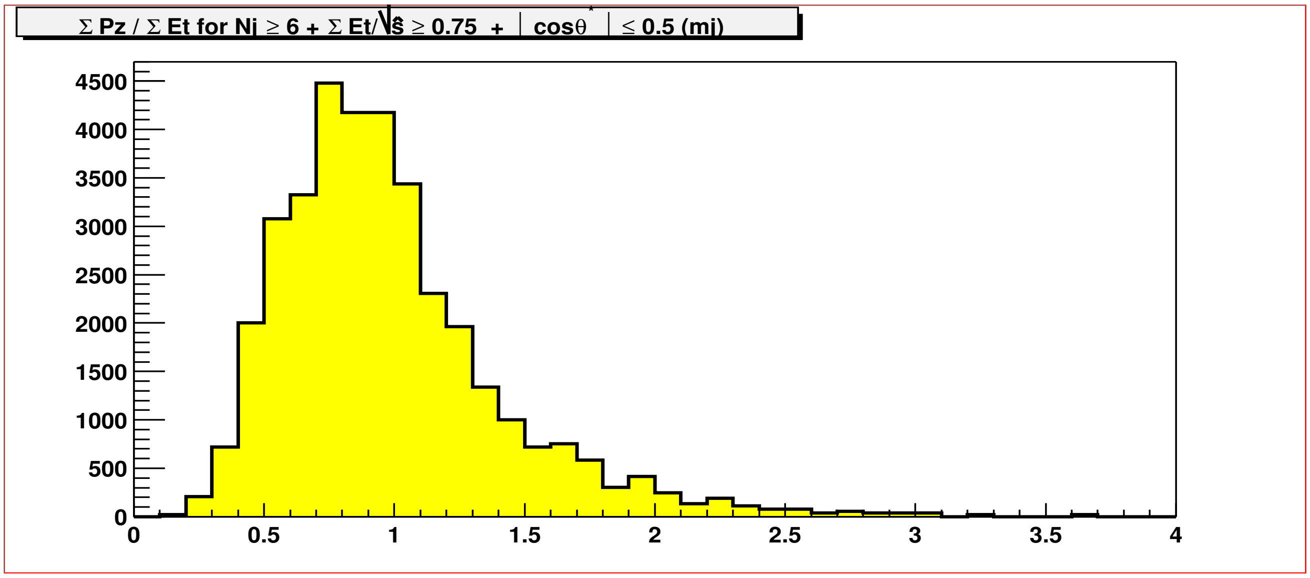
$$\boxed{\cos\theta^*_{1,2}}$$



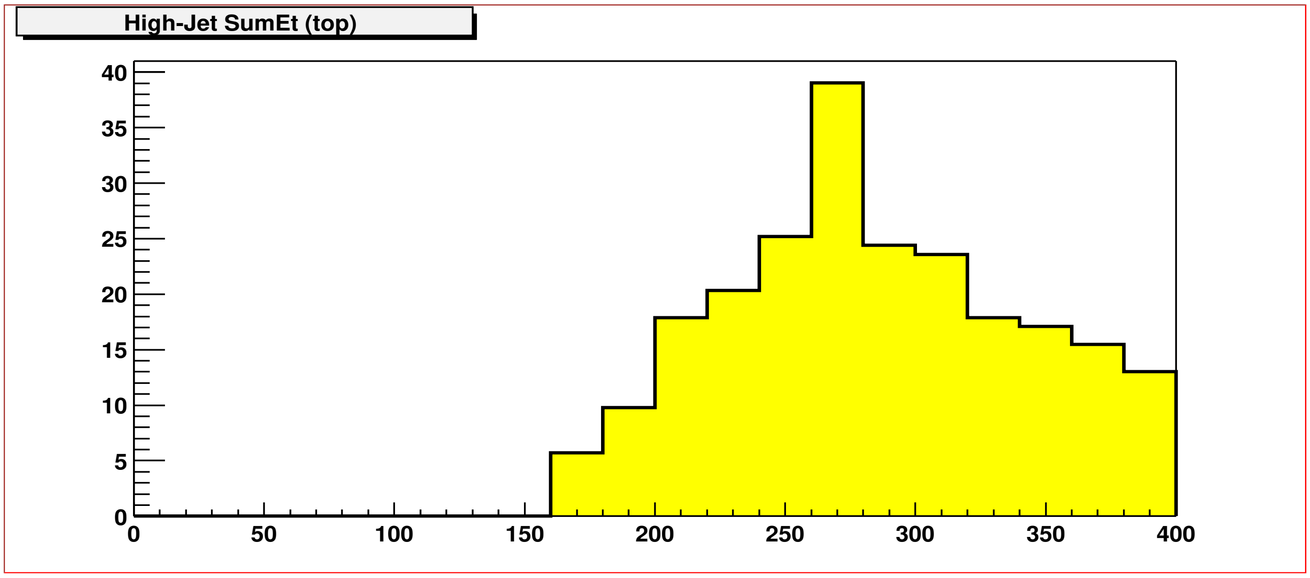


$$\boxed{\sum |P_z| / \sum E_T}$$

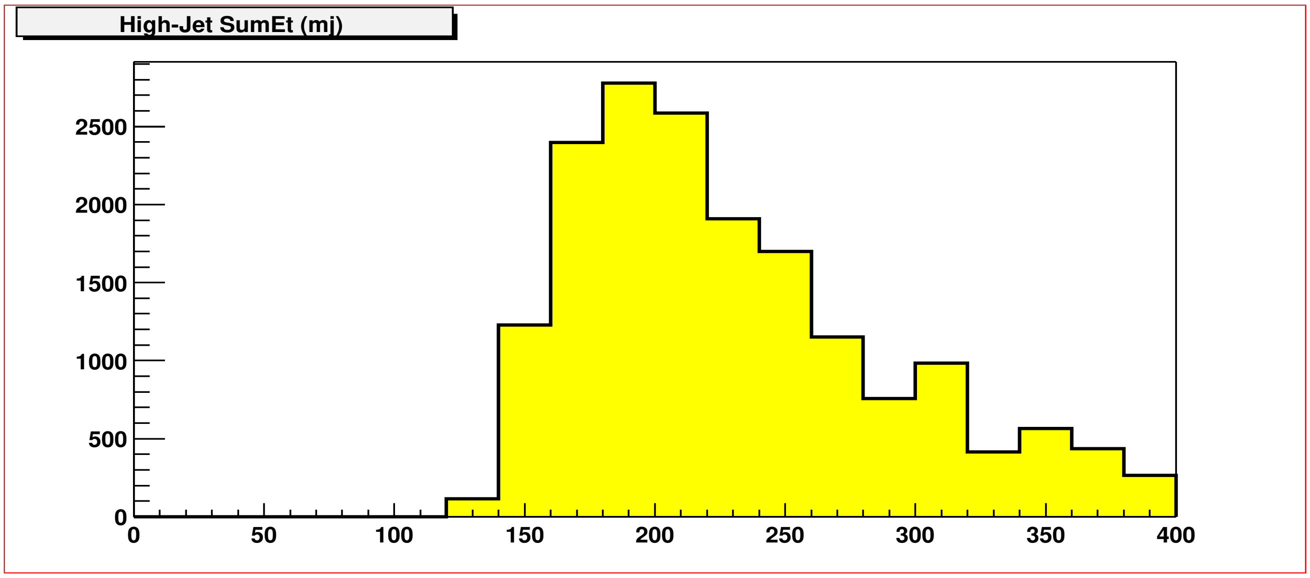




$$\boxed{\sum E_T}$$



High-Jet SumEt (top)
40
35
30
25
20
15
10
5
0
0
50
100
150
200
250
300
350
400



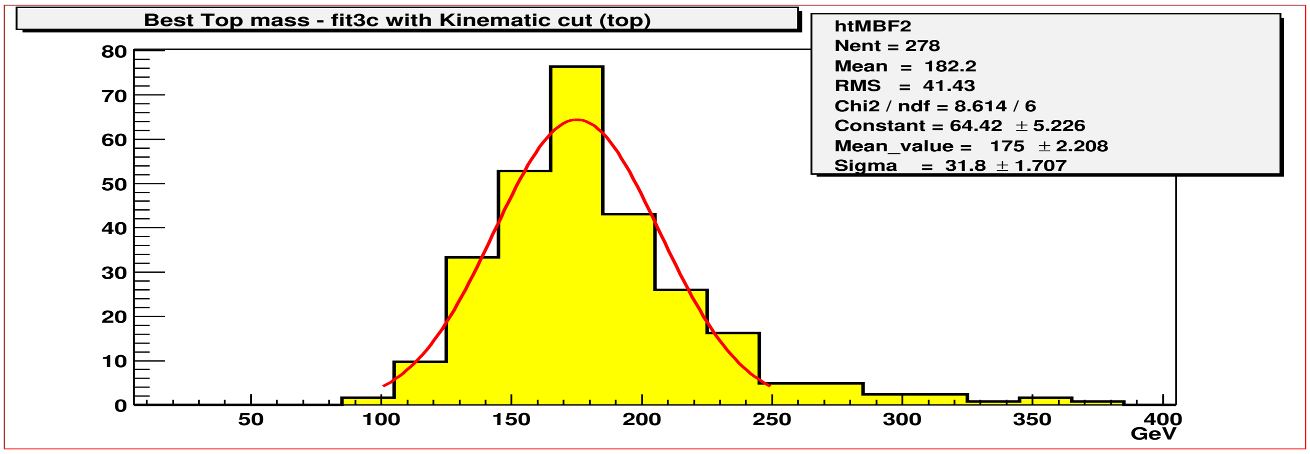
High-Jet SumEt (mj)
2500
2000
1500
1000
500
0
0
50
100
150
200
250
300
350
400

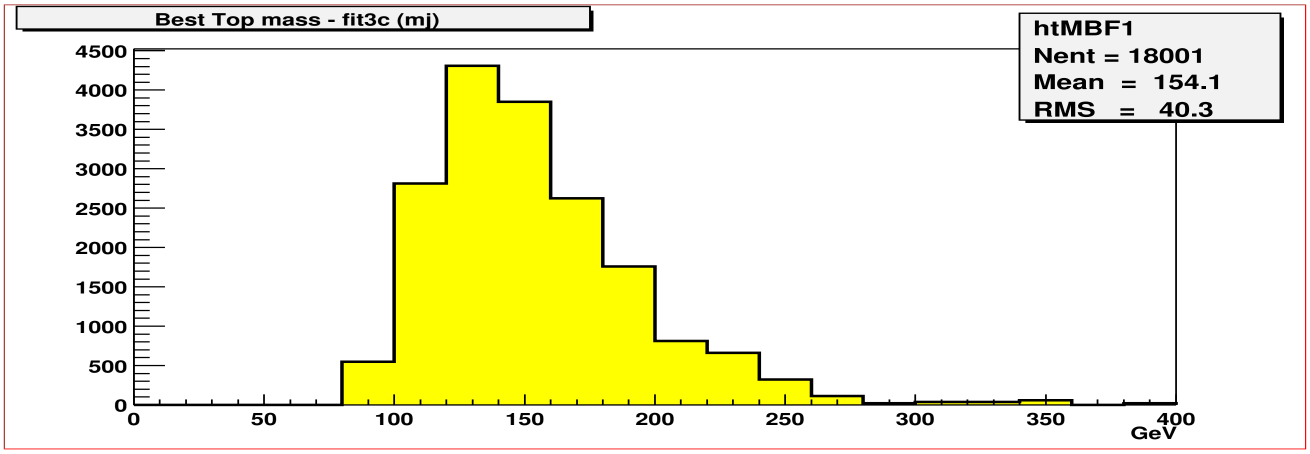
Mass fits after kinematical selection

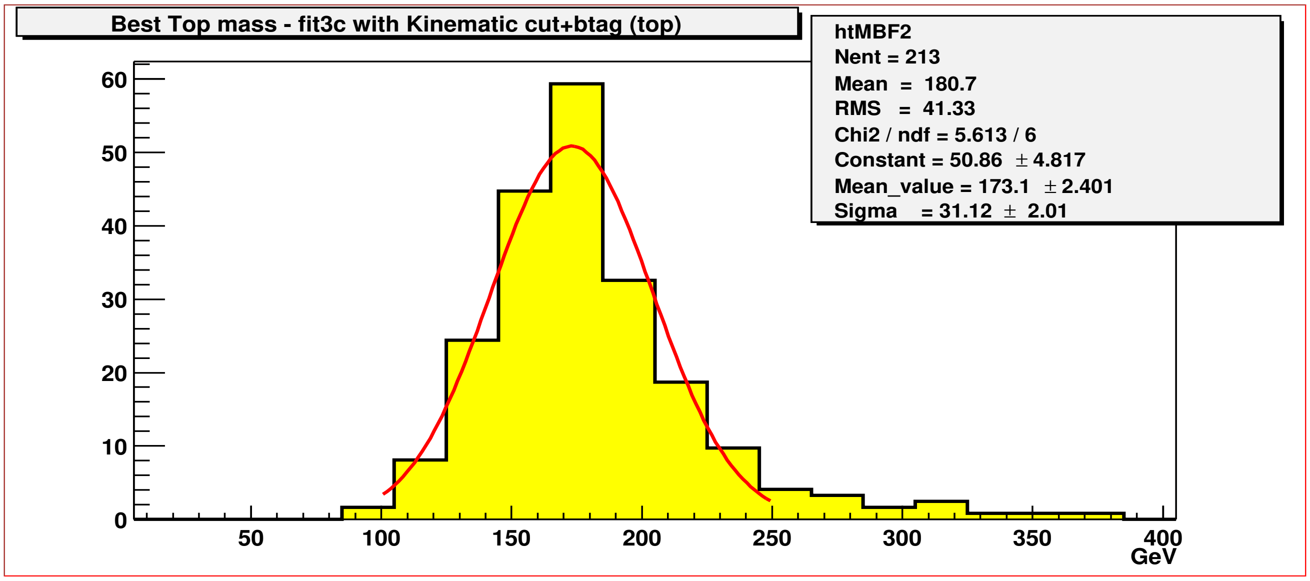
Very preliminary

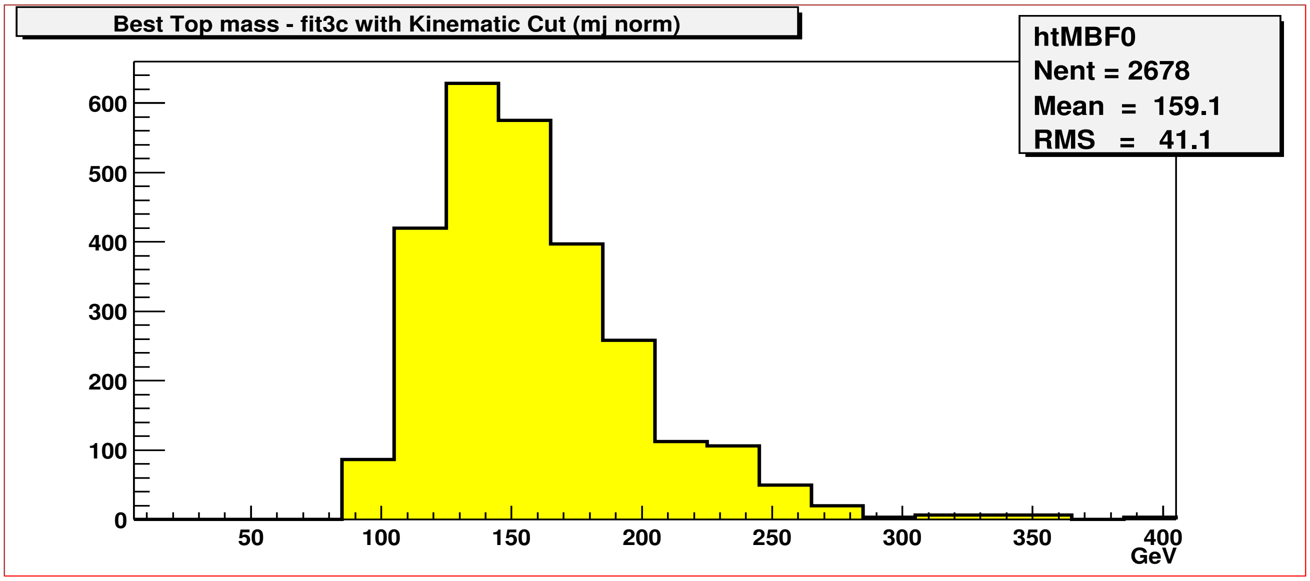
The mean value is 175 GeV/c^2 while $\sigma \approx 18\%$.

We assume to be able to obtain $\sigma \approx 15\%$ once the appropriate corrections are in place.






Best Top mass - fit3c with Kinematic cut+btag (top)
htMBF2
Nent = 213
Mean = 180.7
RMS = 41.33
Chi2 / ndf = 5.613 / 6
Constant = 50.86 ± 4.817
Mean_value = 173.1 ± 2.401
Sigma = 31.12 ± 2.01
60
50
40
30
20
10
0
50
100
150
200
250
300
350
400
GeV


Best Top mass - fit3c with Kinematic Cut (mj norm)
htMBF0
Nent = 2678
Mean = 159.1
RMS = 41.1
600
500
400
300
200
100
0
50
100
150
200
250
300
350
400
GeV

Comparison $t\bar{t}$ *vs* **multijet** events

We now normalize the events as appropriately as we can (1 fb^{-1}):

- for $t\bar{t}$, we assume $\sigma_{t\bar{t}} = 6.5$ pb, (i.e. $[\sigma_{HRW} + \sigma_{PYT}]/2$), and full trigger efficiency;
- for the multijet events, we use the Run-I sample, enhance it by 70% (expected due to higher $\sqrt{s}$, maybe too much!), apply a *kludge* b-tag, i.e. 3% fake rate for each jet in acceptance ($|\eta| \leq 1$)

We expect $S \approx 213$ evts, $N \approx 2678$, $S/N \approx 1/12$.

Can we extract a signal in such a situation?

Likelihood fit

We perform a likelihood fit to separate signal from background:

- we use the spectrum from untagged Run-I multijet events as template for the background;
- starting from Gaussian distributions centered at masses M_t between 160 and 190 (in steps of 5 GeV/c^2) and with $\sigma = 0.15 * M_t$ we generate histograms of 2000 events each and translate them into signal templates.
- we simulate a *data* spectrum summing 2678 events distributed as the background template and 213 events distributed as the template for $M_t = 175$

We maximize the usual likelihood:

$$L = \frac{1}{\sqrt{2\pi}\sigma_b} e^{-(n_b - N_b)^2/2\sigma_b^2} \frac{e^{-(n_s+n_b)}(n_s + n_b)^N}{N!}$$

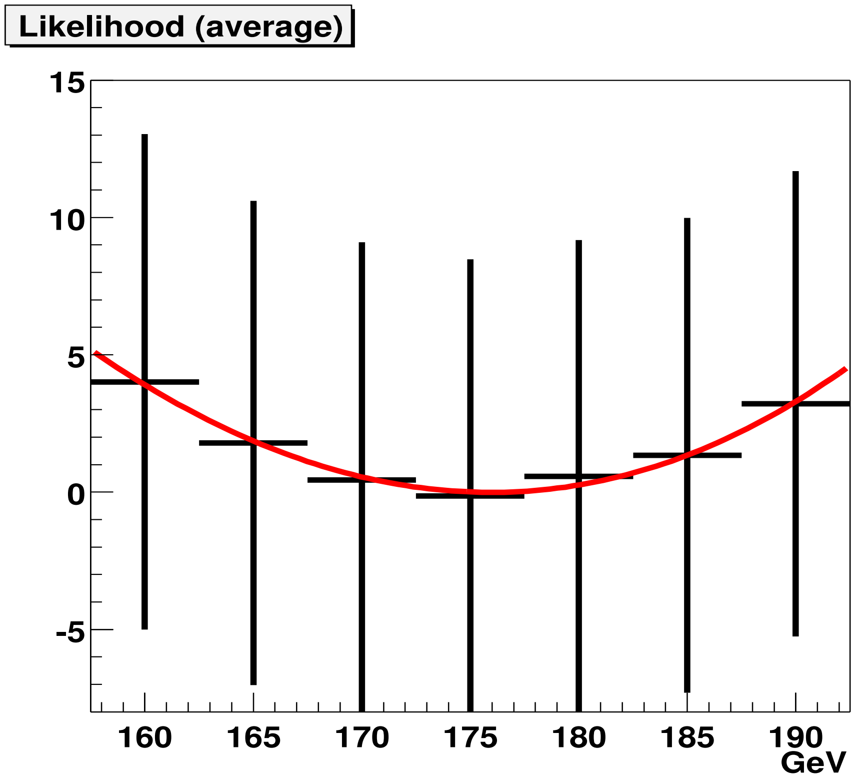
$$\times \prod_i^N \frac{n_b f_b(m_i) + n_s f_s(m_i, M_t)}{(n_b + n_s)}$$

where N_b is the background obtained from the kludge b-tag, while for σ_b we take $0.05 \cdot N_b$ (dominated by systematics on the fake rate). f_b and f_s are background and signal templates, n_b and n_s the amount of background and signal.

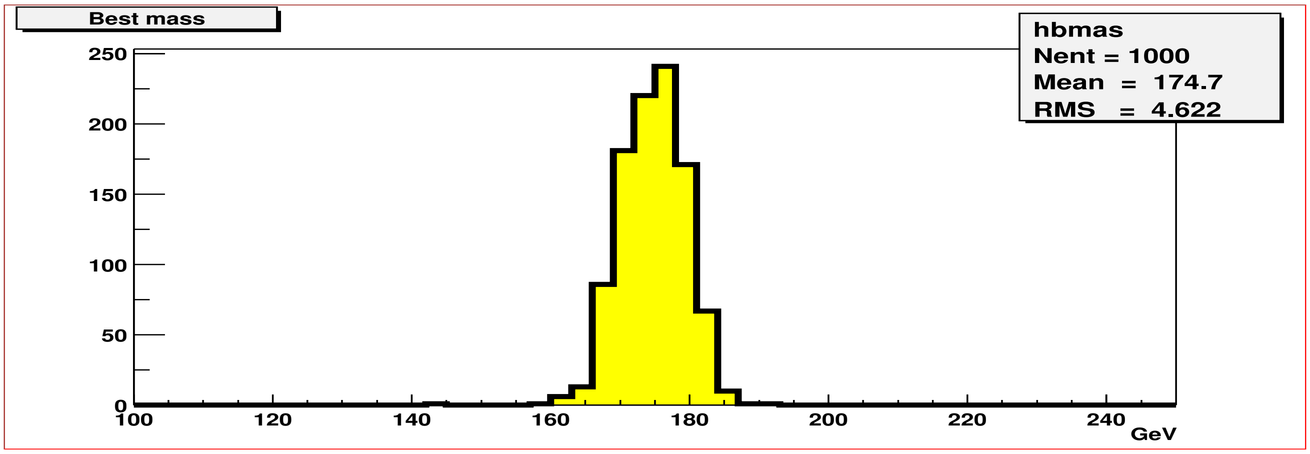
$\boxed{\textbf{Likelihood fit}}$

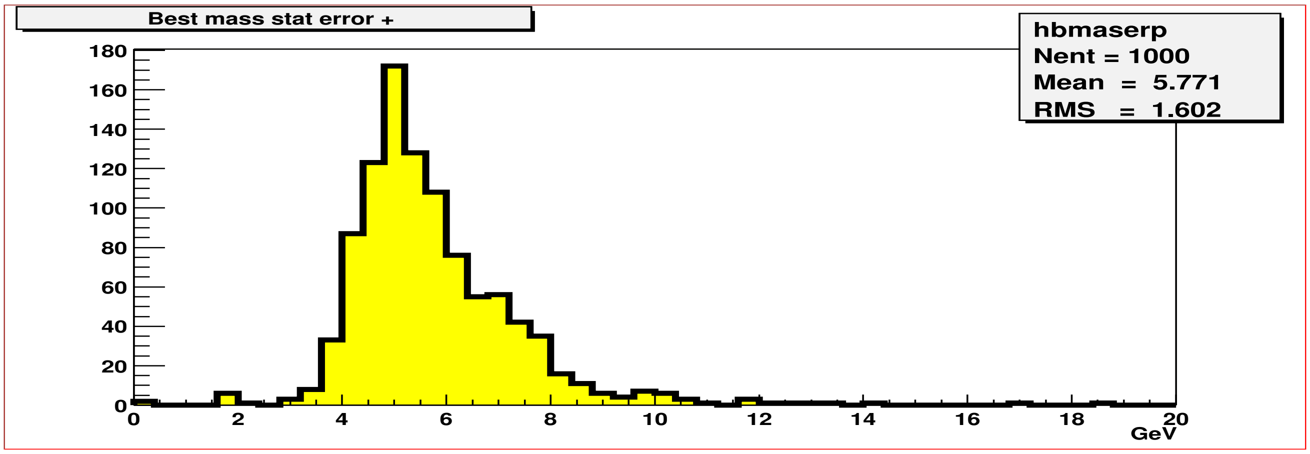
From the likelihood fit we obtain:

- $n_s = 226$ (w.r.t. 213 events)
- $n_b = 2665$ (w.r.t. 2678 events)
- best top mass = 174.7 GeV/c^2 (w.r.t. 175)



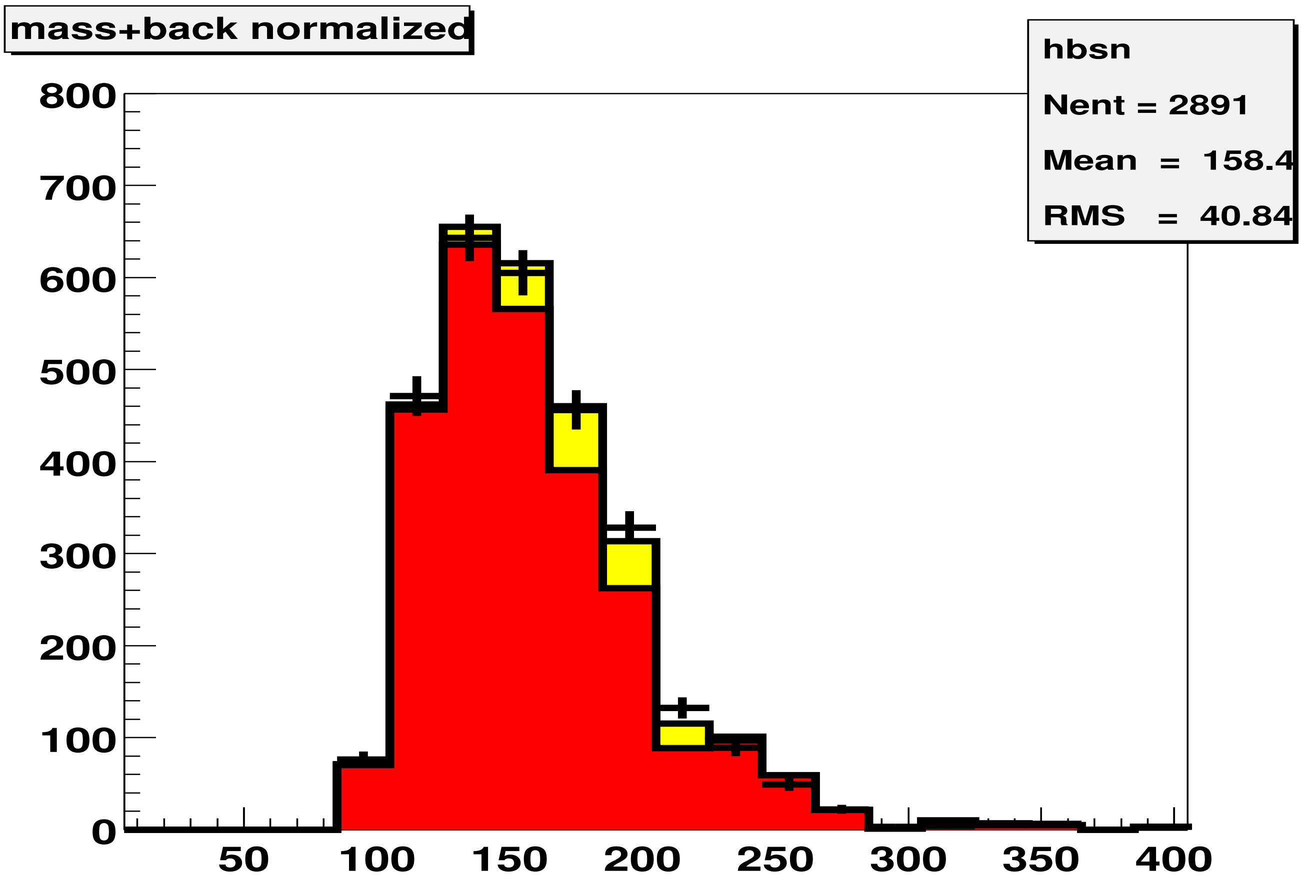
Best mass and its uncertainty





We consider the mean as best mass: $m_t^{best} = 174.7$, while the mean uncertainty amounts to ± 5.8.

“Simulated” data



Comparison between *simulated* data and the sum of signal and background.

$\boxed{\textbf{Simple checks}}$

- we have removed the $n_b \sim N_b$ constraint from the likelihood $\Rightarrow$ we get $m_t^{best} = 174.7$ (w.r.t. 175)

- we begin from different inputs: for $M_t = 170$ we get $m_t^{best} = 171.6$; for $M_t = 180$ we get $m_t^{best} = 178.8$

The background estimate is not essential and the fit behaves quite linearly w.r.t. the input mass (between 170 and 180 GeV/c^2).

Double tags

Using double tags we expect that:

- the mass shape for the signal doesn't change;
- the mass shape for the background doesn't change a lot;
- the ratio S/N will be better.

Because the *kludge* tag does not consider correlations, we prefer to wait for the new data for this analysis.

Conclusions

We have conducted a **very preliminary** study for top mass measurement in the all-hadronic $t\bar{t}$ channel.

Such a study, based on **cdfSim** $t\bar{t}$ samples and on an extrapolation of the Run-I multijet data allows to define a simple set of kinematic requirements which are suitable for the mass measurement.

Assuming a "reasonable" behavior for the 3C mass fit (i.e. $\langle m_{jjb} \rangle \approx M_t$, $\sigma_m = 0.15 * M_t$), a likelihood fit separates quite well signal from background ($S = 226$ w.r.t. an input of 213 events) providing a mass measurement close to the input mass ($m_t^{best} = (174.7 \pm 5.8)$ GeV/c^2 w.r.t. 175).

Run-II jet corrections and Run II multijet data are needed for more accurate estimates.