Update on the Extraction of the Mass of the Top Quark in Multijet Final States - Likelihood vs. Bayesian Estimators

Eunil Won
University of Rochester
Rochester, NY 14627

October 22, 1996

DØ NOTE 2833

We report an optimization of an analysis to extract the mass of the top quark in multijet final states in data of RUN IB, corresponding to an integrated luminosity of 70 events/pb. Assuming that a small excess of events can be interpreted as arising from $t\bar{t}$ production according to Standard Model, we estimate the mass of the top quark as 179$\pm$37(statistical) + 14 - 10(systematic). In addition, we compare a Bayesian statistical method for the measurement of the mass of the top quark with the more conventional likelihood approach.

Contents

1 Overview

Based on the analysis of lepton+jets events at DØ[1], the uncertainty on the mass of the top quark was found to be approximately 20 GeV/c^2, statistical, and 22 GeV/c^2, systematic. More recently, great effort has been expended to reduce the uncertainty of the measurement in the lepton+jets channels by implementing a less stringent set of selection criteria on the data. Following a strategy similar to that used in the lepton+jets analysis, we have performed an optimization of the mass measurement for multijet final states.

In addition, this note describes a general model for estimating the mass of the top quark and its uncertainty based on Bayesian statistics. As an example, the analysis of multijet final states will be restated in the context of Bayesian statistics, and compared to the more conventional method of maximum likelihood.

2 Likelihood Fit

Independent of the $t\bar{t}$ decay channels, most mass analyses performed at DØ involve a final step in which an "unbinned" likelihood function of the following form[2] is maximized with respect to its parameters:

$$\begin{aligned} \mathcal{L}(m_t, n_s, n_b) &= \left[\frac{1}{\sqrt{2\pi}\sigma} e^{-(n_b - N_b)^2/2\sigma^2}\right] \left[\frac{(n_s + n_b)^N e^{-(n_s+n_b)}}{N!}\right] \\ & \left[\Pi_{i=1}^{N} \frac{n_s f_s(m_t, d_i) + n_b f_b(d_i)}{n_s + n_b}\right] \qquad (1) \end{aligned}$$

where N_b denotes the expected number of background events, N is the total number of observed events, σ is the systematic uncertainty in the estimate of the background, f_s and f_b are, respectively, the probability density functions for signal and background, and d_i are any kinematic variables sensitive to the mass of the top quark (e.g., scalar sum of the transverse momenta of the jets, the mass of three jets, etc.). The parameters m_t, n_s, and n_b denote, respectively, the top mass, the expected number of top-quark events, and the expected number of background events.

Because the uncertainty on the mass of the top quark is related directly to the sharpness of the log of the likelihood function near the maximum, it is important to choose selection criteria that provide as narrow a log-likelihood function as possible near that point. Taking the negative log of the likelihood function, we can write:

$$\begin{aligned} l(m_t, n_s, n_b) &= -log(\mathcal{L}(m_t, n_s, n_b)) \\ &= \sqrt{(}2\pi\sigma) + (n_b - N_b)^2/2\sigma^2 - N log(n_s + n_b) + (n_s + n_b) \\ &+ log(N!) - \sum_{i=1}^{N} \frac{n_s f_s(m_t, d_i) + n_b f_b(d_i)}{n_s + n_b}. \end{aligned} \quad (2)$$

Assuming that the negative of the log-likelihood function $l(m_t, n_s, n_b)$ has a local minimum at $(m_t^{min}, n_s^{min}, n_b^{min})$, and that it is approximately symmetric about the m_t, n_s and n_b axes at the minimum point, one could investigate the dependence on a change in the top mass by calculating the following quantity:

$$\begin{aligned} \delta l &= l(m_t + \delta m_t, n_s^{min}, n_b^{min}) - l(m_t, n_s^{min}, n_b^{min}) \\ &\approx \frac{\partial l(m_t, n_s^{min}, n_b^{min})}{\partial m_t} \delta m_t. \end{aligned} \quad (3)$$

Near the minimum, the quantity $\frac{\partial l(m_t, n_s, n_b)}{\partial m_t} \delta m_t$ can be rewritten as follows:

$$\frac{\partial l(m_t, n_s^{min}, n_b^{min})}{\partial m_t} \delta m_t = \sum_{i=1}^{N} \frac{-1}{f_s(m_t, d_i) + n_b^{min}/n_s^{min} f_b(d_i)} \frac{\partial f_s(m_t, d_i)}{\partial m_t}. \quad (4)$$

Consequently, the dependence on the top quark mass (and thereby its uncertainty) can be changed by changing n_s, n_b and the probability density function for signal $f_s(m_t, d_i)$. One can therefore draw the following conclusions:

1. Because the quantity $\frac{\partial l(m_t, n_s, n_b)}{\partial m_t} \delta m_t$ is *large* for a *narrow* log-likelihood function, to reach the smallest uncertainty on the mass of the top quark one should keep $\frac{\partial l(m_t, n_s, n_b)}{\partial m_t} \delta m_t$ as large as possible by adjusting the kinematic acceptance criteria.

2. Also, to reduce the uncertainty on the extracted mass, one should maximize the signal to background ratio to take advantage of the multiplicative term $1/(f_s + n_b/n_s f_b)$.

3. To minimize the uncertainty on the extracted mass, one should use the d_i that provide a signal probability function $f_s(m_t, d_i)$ that is as sensitive as possible to the input mass m_t.

4. Because n_s and n_b are constrained by the Poisson term in the likelihood function, f_s must be correlated with f_b. The impact of this correlation will be studied previously later.

An important matter is therefore how to establish optimal selection criteria that satisfy the above requirements.

2.1 Optimization

Since the first measurement of the mass of the top quark in DØ[1], the analysis has been re-optimized to reduce the uncertainty on the mass in the lepton+jets channel. However, because of a absence of high-p_T leptons and significant missing E_T, it is difficult to follow a similar strategy in the case of multijet final states. Instead, an extension of the grid search technique[4] has been pursued to establish selection criteria satisfying the requirements listed in the previous section. The specific procedure is described below.

After finding the optimal boundary[4] that maximizes the signal to background, one calculates the difference between the average H_T value for the signal and for background at any given point on that boundary and uses that to improve mass extraction, as follow. Figure 1 (a) shows the optimal boundary results from the full grid search. The grid points were sampled using data from Run Ia for background and $t\bar{t}$ events generated with the ISAJET program at m_t=180 GeV/c^2 for signal. The grid search was performed on H_T^3, *Centrality*, and *Aplanarity* only. For simplicity, the average value of H_T was chosen as the mass-dependent kinematic variable that enters into f_s and f_b. The difference in H_T between signal and background ($\triangle H_T$) is examined in Figs. 1 (b), (c) and (d). The selection criteria were then chosen to favor a signal probability density function $f_s(m_t, d)$ as different as possible from that of the background. This was achieved by choosing the points in Fig. 1 (b) and (c) that provide small H_T but large $\triangle H_T$, with the largest possible signal to background ratio that would still leave enough events to perform a mass analysis. Finally, the selection criteria were chosen to favor a signal probability density function f_s that would be as sensitive as possible

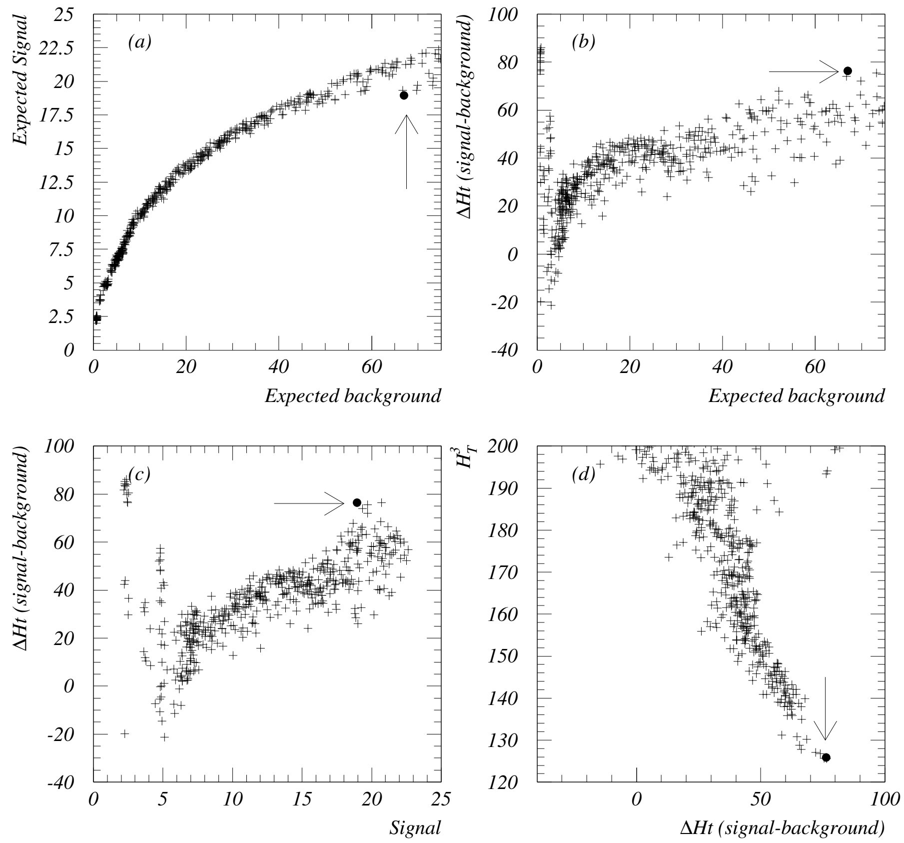

Figure 1: Results of the grid search. Arrows and the dark circles indicate the point chosen by the mass optimization procedure. (a) Expected $t\bar{t}$ signal vs background from a random grid search; only the points at the optimal boundary are shown here. (b) Difference between average H_T for signal and the background vs expected background. (c) Difference for average H_T vs expected signal. (d) H_T^3 vs difference of average H_T.

variable	Centrality	Aplanarity	H_T^3 (GeV)
cutoff	> 0.70955	> 0.055451	> 125.86

Table 1: Threshold selection criteria based on our "mass-sensitive" grid search.

to the input top mass m_t. [1] What is done is that H_T^3 is chosen to be as small as possible for any given set of selection criteria. To summarize, the optimization procedure involves:

- Obtaining on the optimal boundary so as to make $1/(1 + n_b/n_s)$ to be large.
- Requiring H_T^3 to be small so as to make $\partial f_s(m_t, d_i)/\partial m_t$ large.
- Requiring $\triangle H_T$ large so as to make f_s differ as much as possible from f_b.

Arrows in Fig. 1 (a), (b), (c), and (d) show the operating point chosen based on our qualitative optimization procedure, and Table 1 lists the selection criteria defined by the chosen point. Our strategy selects the criteria that tend to reduce the uncertainty on the measurement of the mass of the top quark.

2.2 Sensitivity

Once the cut-point has been chosen, the signal and background probability density functions, $f_s(m_t, d)$ and $f_b(d)$, can be constructed, and an estimate made of the uncertainty expected on the values of the estimated mass. The latter is obtained by generating an ensemble of events of assumed top quark mass and an ensemble of background events, extracting the value of the mass, by maximum likelihood, and examining the dispersion of the sampling distribution of these mass estimates.

For the present study, as the mass sensitive discriminator (d), we chose H_T, which is the sum of jet transverse energies—jets were defined as objects of 0.5 cone in η-ϕ space, with $E_T > 15$ GeV and $\mid \eta \mid < 2.0$. The main

[1] Ideally, one would calculate δl at every point in Fig. 1 (a) and then select the point that provides the largest δl. However, considering the available resources, that would be a challenging task to perform.

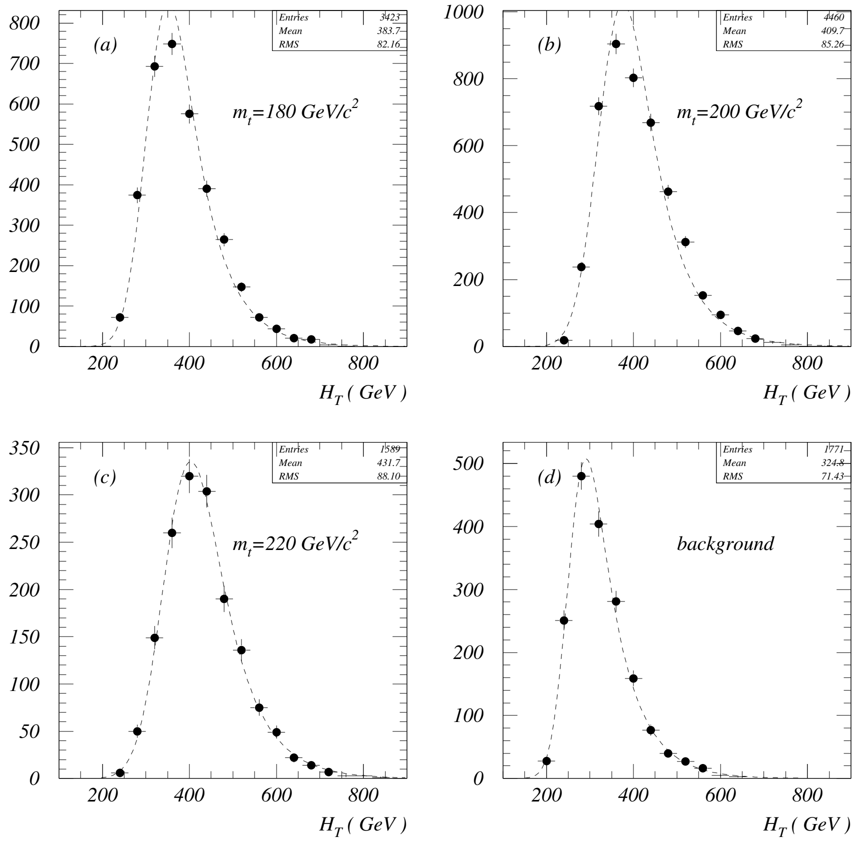

Figure 2: Distributions in H_T for signal(ISAJET) and background (data from RUN IB), after application of the selection criteria. Dashed curves are the fits to the distributions using the functional forms given in the text. Results are for: (a) m_t = 180 GeV/c^2, (b) m_t = 200 GeV/c^2, (c) m_t = 220 GeV/c^2, and (d) background.

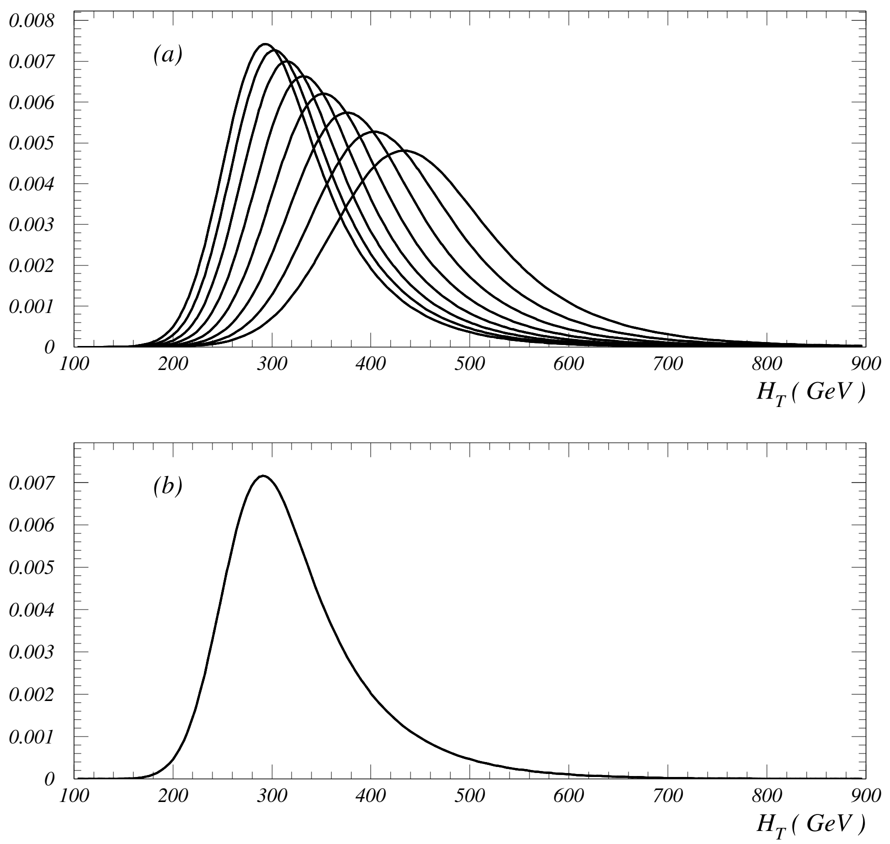

Figure 3: Probability density functions for signal and background after application of the selection criteria. (a) Signal probability density functions are for a top-quark mass 100 to 240 GeV/c^2 in mass intervals of $\triangle m_t$=20 GeV/c^2. (b) Probability density function for the background (same as Fig 2 d).

advantage of doing a mass analysis with the variable H_T, as opposed to the fitted mass value m_i, is that one avoids the combinatoric background introduced in the kinematic fitting procedure. However, there are also disadvantages: m_i should be the sharpest variable for extracting the mass of the top quark since, in principle, m_i is the variable that is most strongly correlated with the top mass. Also, he H_T analysis is more likely to mis-measure the mass of the top quark if $t\bar{t}$ production has a non-negligible contribution from non-standard sources (e.g., heavy $t\bar{t}$ resonances, as suggested by Hill and Parke[5], etc.).

The standard DØ jet-energy correction package (CAFIX 5.0) was applied to data prior to setting restrictions on jet E_T and η. Owing to the long tail of the H_T spectrum, in both background and signal, a function with a multiplicative exponential decay term was adopted as the functional form to model the probability density functions for background and signal. In particular, the probability function used to characterize the signal was:

$$f(m_t, d) = NG((d - p_1)/p_2)e^{-p_3 d} \tag{5}$$

where

$$\begin{aligned} p_1 &= a_1 + a_2 m_t + a_3 (m_t)^2 \\ p_2 &= a_4 + a_5 m_t + a_6 (m_t)^2 \\ p_3 &= a_7 + a_8 m_t \\ d &= H_T, \end{aligned}$$

N is a normalization constant, a_i are free parameters in the fit, and $G(x)$ is the frequency function, defined as

$$G(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\, dt.$$

The probability function for the background is then modeled using the same functional form as used for the signal, but without introducing m_t-dependent p_i. Figures 2 (a),(b), and (c) show the distributions in H_T for fixed m_t of the fit of the signal probability function(s) to ISAJET events of different values in m_t. Figure 2 (d) shows the results of the fit to the background (data from RUN IB). The quality of the fit will be discussed later when we turn to systematic uncertainties. Figure 3 (a) shows the probability density function $f_s(m_t, d)$ for different input top masses, divided into 20 GeV/c^2

mass intervals, and (b) shows, for comparison, the background probability density function $f_b(d)$ given in Fig. 2 (d). From Fig. 3 (a), it is clear that $f_s(m_t, d)$ changes faster at higher m_t, which indicates that the optimization procedure yields smaller uncertainty for higher m_t values. The background shape is similar to the signal for top mass at $m_t \approx 100\ GeV/c^2$. Note that in a previous analysis[3] the background shape was similar to the signal for $m_t \approx 160\ GeV/c^2$, which suggests that out procedure has led to an improvement of the current analysis with respect to the previous one.

The background shape was modeled using the data from Run Ib. Consequently, we expect it to contain some contribution from top production that may bias the shape of the background. To correct for this bias, we defined the following corrected background function:

$$f_b = \frac{1}{1-C}(f_b - Cf_s) \tag{6}$$

where C is the fraction of events in the background sample that can be attributed to signal ($t\bar{t}$ production) at any given mass m_t. The value of C was estimated from the value of the cross section for $t\bar{t}$ production given by the Standard Model.

The effect of the signal to background ratio on the uncertainty in the mass was discussed in the previous section. To study this effect in detail, we generated ensembles of 100 events assuming different signal to background ratios, and an input top quark mass of $m_t = 180\ GeV/c^2$. Figure 4 shows the sampling distributions of the maximum likelihood mass estimates for different signal to background ratios: 100/0, 70/30, 50/50, and 30/70. These plots confirm that our entire maximum likelihood procedure is sound. Table 2 lists the standard deviations (RMS) of the four sampling distributions. It is clear from these results that the uncertainty on the top quark mass is larger for smaller values of the signal to background ratio, beyond that expected from the $\sqrt{n_s}$ scaling rule.

To estimate the uncertainty expected in the mass extraction for the selection criteria given in Table 1, we generated ensembles of events assuming a realistic number of signal and background events. (The procedure for choosing the number of signal and background events will be discussed shortly.) Figures 5(a), (b), (c) and (d) show the distribution of mass values that minimize the negative log-likelihood function for ensembles generated with input top-quark masses of $m_t = 160, 180, 200$, and $220\ GeV/c^2$, respectively. The signal events were generated using the ISAJET event generator, and the background events were chosen at random from data of Run Ib.

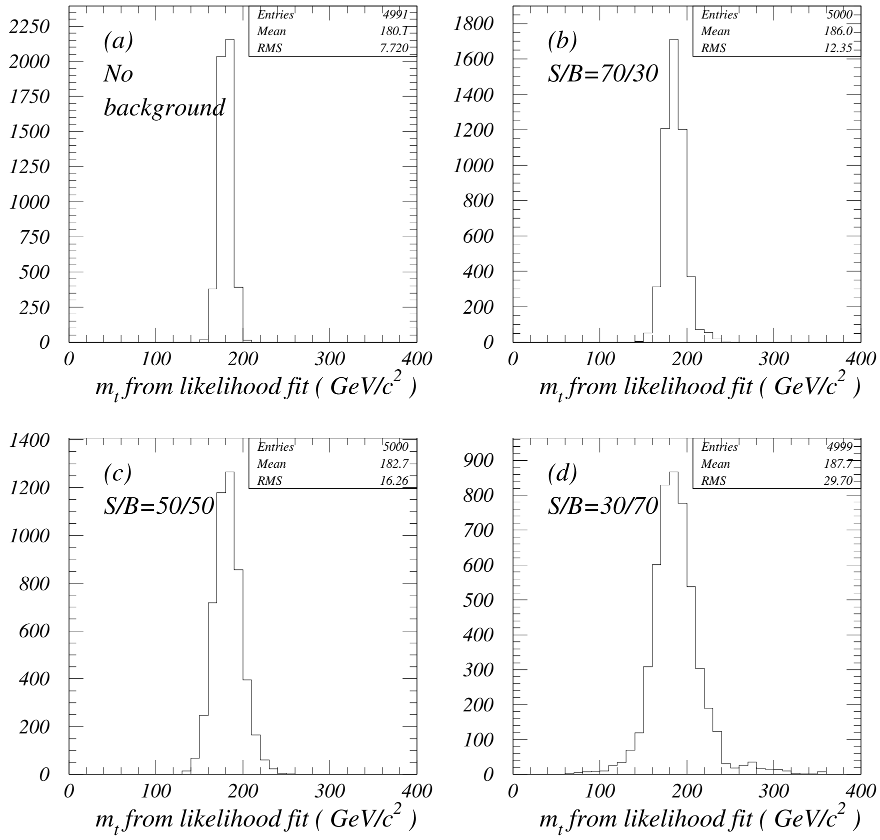

Figure 4: Results from ensemble tests of the likelihood fit. Signal events were chosen from ISAJET $t\bar{t}$ at $m_t = 180\ GeV/c^2$, and background events were from RUN IB data. (a) Signal to background ratio of S/B=100/0, (b) S/B=70/30, (c) S/B=50/50, and (d) S/B=30/70 were assumed in the analysis.

signal / background	100/0	70/30	50/50	30/70
RMS (GeV/c^2)	7.7	12.4	16.3	29.7
Expectation from $\sqrt{n}_s$ scaling	7.7	9.2	10.9	14.1

Table 2: Relation between the signal to background ratio and the expected uncertainty obtained from ensemble tests.

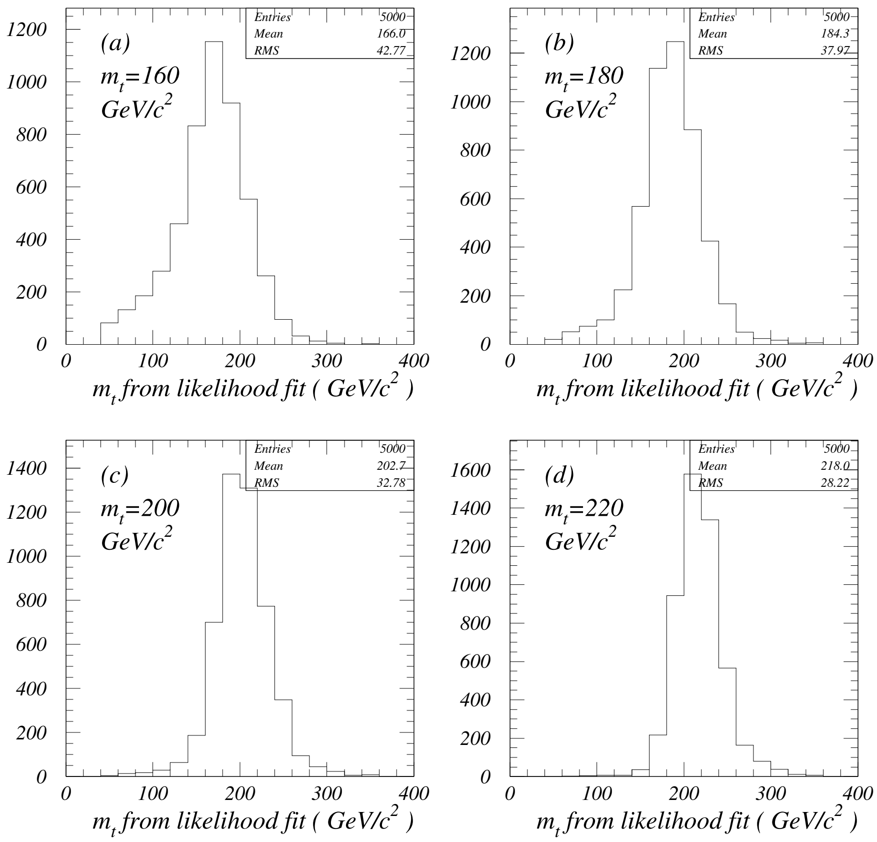

Figure 5: Results from likelihood fits to 5000 samples of events consisting of $N_{observed}$=72 and $N_{background}$=57 events, for: (a) m_t=160 GeV/c^2. (b) m_t=180 GeV/c^2. (c) m_t=200 GeV/c^2. (d) m_t=220 GeV/c^2.

input m_t (GeV/c^2)	160	180	200	220
RMS (GeV/c^2)	42.8	38.0	32.8	28.2

Table 3: Relation between mass of the top quark and the expected uncertainty from ensembles that have event samples comparable to the data from RUN IB.

	events
kinematic criteria	1771
with muon tagging	72
expected background	56.9
systematic uncertainty in background	11.4

Table 4: Result of the imposition of criteria on data from RUN IB.

All events were required to pass the kinematic criteria given in Table 1. The RMS values of the sampling distributions decrease with the top quark mass, presumably because both of the similarity in shape of the background and signal probability functions for small m_t and because of increase in variance of the H_T distributions. Table 3 lists the rms value of the fitted mass as a function of the assumed top quark mass. This will be compared in the next section to our result for data from RUN IB.

2.3 Likelihood Fit

The total integrated luminosity of 70 events/pb of RUN IB data was used in the current analysis. Events were required to pass the JET_MULTI trigger and filter. A detailed description of these requirements can be found elsewhere[4]. A total of 1771 events remained after imposing the selection criteria given in Table 1. The background was estimated after restricting the sample to events with at least one muon detected near any jet in a given event. This, so called "b-quark tagging", is described in detail in Ref [4]. Only 72 events remained after requiring b-quark tagging, and the remaining background was estimated to be 56.9 events. Table 4 summarizes these numbers. An "unbinned" likelihood fit was performed on the resulting data sample, which gave a top quark mass estimate of $m_t = 182.9\ GeV/c^2$. An *approximation* to the standard deviation of sampling distribution can be

	fit result
m_t	182.9 + 43.2 - 121.0 GeV/c^2
n_s	13.1 $\pm$ 13.8
n_b	58.3 $\pm$ 11.3

Table 5: Result of the likelihood fit on data from RUN IB.

obtained from the likelihood curve by the usual rule: a change in the value of -log($\mathcal{L}$) of 0.5. This gives $+43.2 - 121.0$ GeV/c^2. But from Table 3, the standard deviation expected on the basis of ensemble studies is 37.2 GeV/c^2 at an input top-quark mass of 182.9 GeV/c^2. Therefore, we conclude that in this case getting the error from the likelihood curve is unreliable and we should use the error obtained from the ensemble tests (less subject to fluctuations). Figure 6 shows the result from the likelihood fit, and the distribution of the 72 events in H_T. Table 5 summarizes the values of the parameters obtained in the likelihood fit where uncertainties were obtained by a change in -log($\mathcal{L}$) of $\pm$ 0.5. The large uncertainty on n_s indicates that the current analysis does not provide a significant excess of signal to support the existence of the top quark production in multijet final states.

2.4 Systematic Uncertainties

The following items comprise the major sources of systematic uncertainty that affect the extraction of the mass of the top quark in this analysis.

1. Shift of extracted mass from the input top mass due to a bias in the procedure: Figure 5 shows a shift in the extracted mass that depends on the input mass. The systematic shift is defined as the the difference between the input masses and the average of the values obtained in the likelihood fit. Table 6 gives this uncertainty as a function of the input mass. This dependence can be used to correct the final result.

2. Jet energy scale: Uncertainty on the mass of the top quark due to the uncertainty in the jet energy scale was estimated as follows. Assuming an uncertainty in energy of 5 %, ensemble tests were reprocessed using jet energies scaled up by 105 % and down by 95 %. The observed offsets depend on the input mass. Figures 7 (a), (b), (c), and (d) show the mass distributions resulting from the likelihood fit for different input

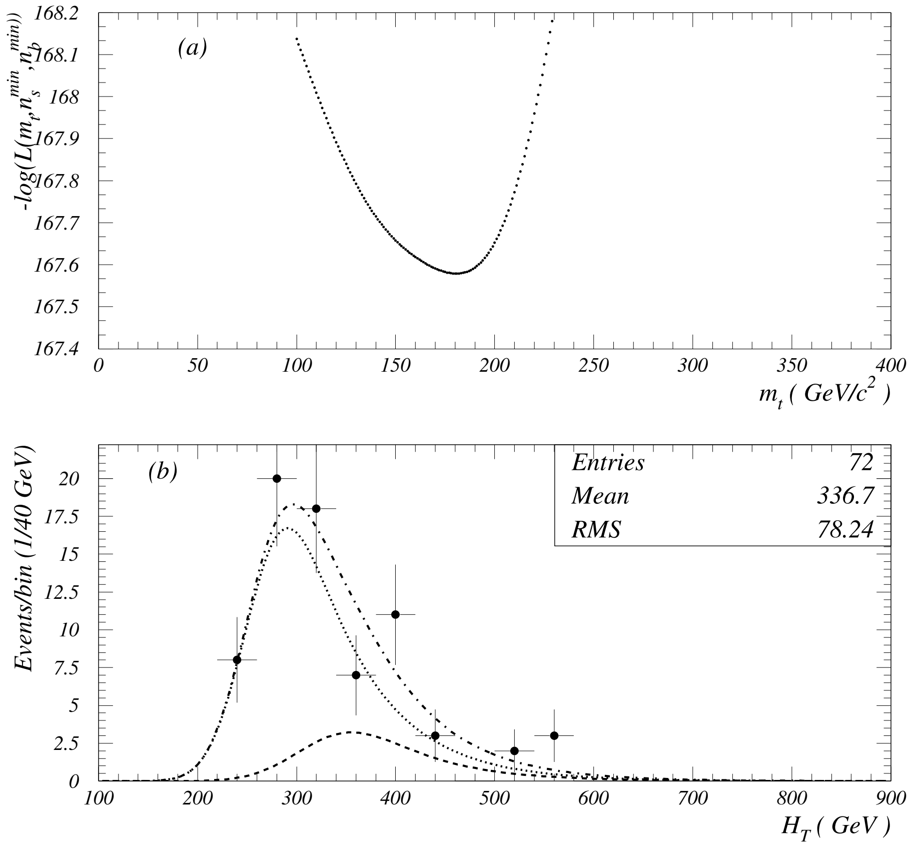

Figure 6: Results from the likelihood fit for: (a) -log(likelihood) as a function of m_t, for fixed n_s^{min} and n_b^{min}; (b) the H_T distribution for our sample of 72 events, with the three curves representing signal, background, and the sum of signal and background, obtained from the likelihood fit.

input m_t (GeV/c^2)	160	180	200	220
observed shift (GeV/c^2)	+6.0	+4.3	+2.7	-2.0

Table 6: Relation between the input top-quark mass and the expected shift due to a bias in the analysis procedure.

input m_t (GeV/c^2)	160	180	200	220
(+) uncertainty (GeV/c^2)	26.5	13.2	6.1	2 .8
(-) uncertainty (GeV/c^2)	6.2	-8.6	-10.5	-7.6

Table 7: Relation between the input top-quark mass and the expected uncertainty due to a $\pm$ 5 % uncertainty in the jet-energy scale.

input m_t (GeV/c^2)	160	180	200	220
uncertainty (GeV/c^2)	0.1	2.4	5.4	6.2

Table 8: Uncertainty in extracted m_t due to the parameterization of the probability density functions.

masses. The systematic uncertainty in the mass is taken as the mean of the offsets between the result of the input mass and for the energy-shifted distributions. Table 7 gives the uncertainty as a function of the input top-quark mass. For the case of lower input top masses, approximately 10 % of the likelihood fits were unsuccessful when the jet energy scale was reduced by 5 %, Such fits provided top-quark masses close to zero, and those results were not used in the calculation of mean values, and may underestimate the uncertainty.

3. Parameterization of the probability density functions: The uncertainty in the parameterization of f_s and f_b was estimated by redoing the fit with a different functional form. Gaussians were used to parametrize the f_s and f_b. Changes in the extracted mean values are taken as the uncertainty and are shown in the Table 8. The uncertainty in these functions is correlated with the shift in the extracted mass, but the degree of the correlation has not as yet been quantified.

4. Models of parton fragmentation : In this analysis, we used only the ISAJET event generator. Owing to the lack of events from other event generators available in this stage, we cannot estimate the systematic uncertainty due to different models of parton-fragmentation. However, we believe these to be as large as the uncertainty due to the jet energy scale, based on the recent studies in lepton+jets group.

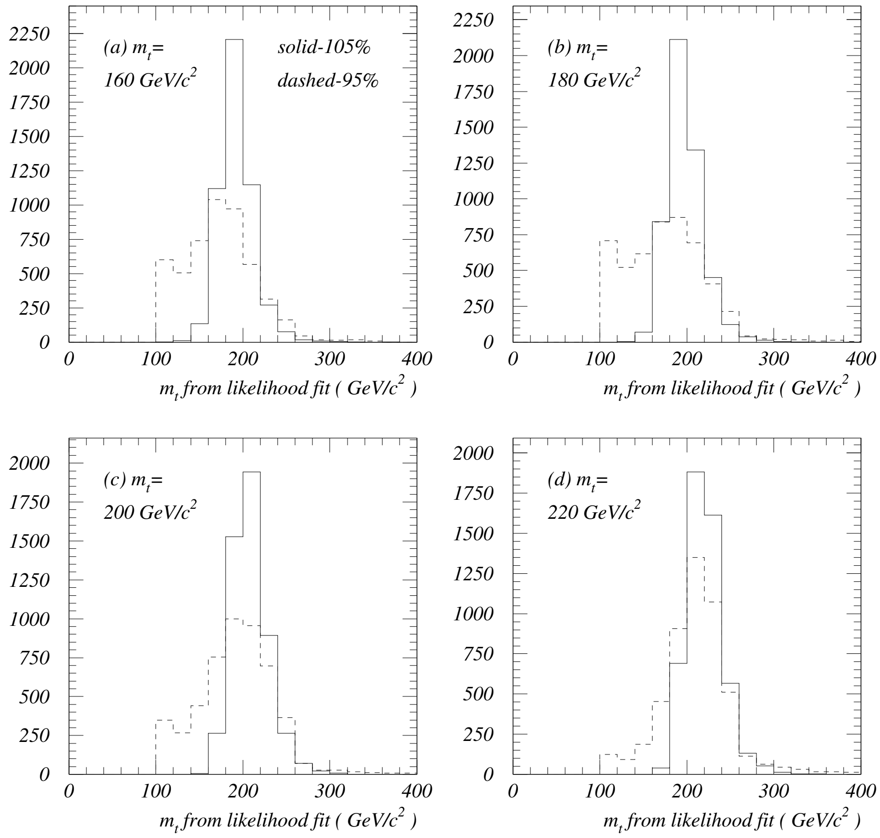

Figure 7: Results from studies of $N_{observed}$=72 and $N_{background}$=57 events, when jet energies were scaled up by 5 %(solid lines), and scaled down by 5 %(dashed lines) for; (a) m_t=160 GeV/c^2, (b) m_t=180 GeV/c^2, (c) m_t=200 GeV/c^2, and (d) m_t=220 GeV/c^2.

Note that the first two systematic uncertainties become small for higher top-quark masses, an effect similar to that observed in Fig. 5. Adding the systematic uncertainties in quadrature (for an input top-quark mass 180 GeV/c^2), we obtain a corrected mass of the top quark of 179$\pm$37(statistical) +14-10(systematic) GeV/c^2.

2.5 Consistency Check

To demonstrate the effectiveness of the optimization procedure explained in the previous section, we randomly chose four different reference points and redid the likelihood fit analysis. Figure 8 shows the four different new points in the same space as Fig. 1. The grid points are same as the one in Fig. 1 with the open circles indicating the new selections and the closed circles the point determined by the optimization procedure. Table 9 lists limiting values for the parameters, and the results of applying them to RUN IB data.

Set	Centrality	Aplanarity	H_T^3 (GeV)	$N_{observed}$	$N_{background}$
1	> 0.67165	> 0.051859	> 145.25	45	36.8
2	> 0.52689	> 0.033776	> 150.92	60	46.6
3	> 0.70147	> 0.035275	> 156.97	32	25.0
4	> 0.58524	> 0.022085	> 172.70	26	22.6

Table 9: Four randomly chosen threshold selection criteria along the optimal boundary, and the results of the choices. the results.

For each of the four new points, probability density functions were rebuilt and ensemble tests performed again. Figure 9 shows the result from ensemble tests. The RMS values of for all four sets are larger than the value of 38 GeV/c^2 one obtained with the optimization procedure. This indicates the veracity of the method. For Set 4, the likelihood fit failed frequently, and that produced a peak near a mass of 80 GeV/c^2.

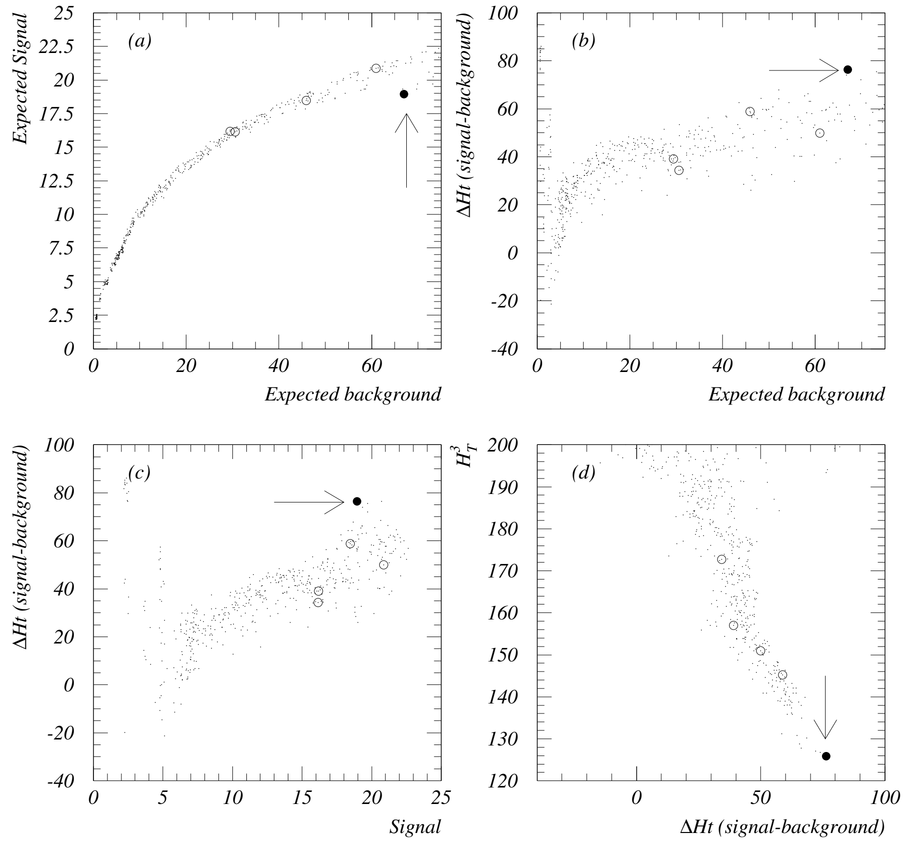

Figure 8: Results of the grid search. Arrows and the closed circles indicate the point chosen by the optimization procedure. Four open circles indicate the points chosen randomly. (a) Expected $t\bar{t}$ signal vs background from a random grid search, only the optimal boundary of the search is shown here. (b) Difference for average H_T between the signal and the background vs expected background. (c) Difference for average H_T vs expected signal. (d) H_T^3 vs difference for average H_T.

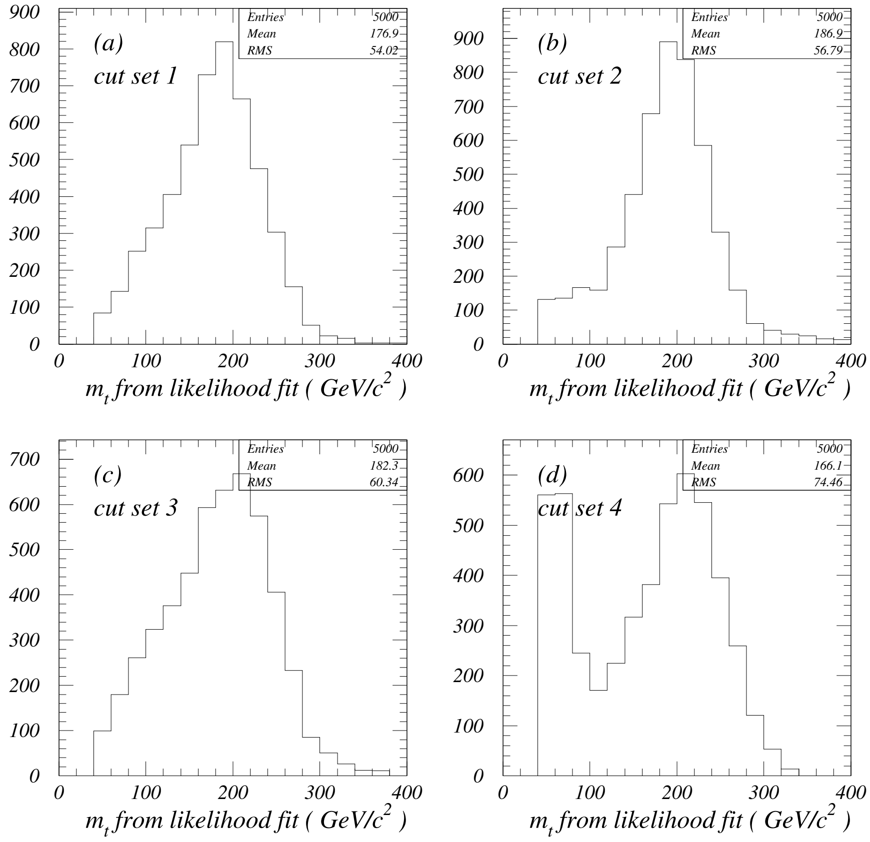

Figure 9: Results from ensemble tests of the likelihood fit. Signal events were chosen from ISAJET $t\bar{t}$ at $m_t = 180\ GeV/c^2$ and background events were from Run Ib data. Results are for: (a) Set 1, (b) Set 2, (c) Set 3, and (d) Set 4.

3 Bayesian Inference

In Bayesian inference, probability[6][7] is interpreted as

a measure of degree of belief in a proposition.

This differs from the frequentistic interpretation: the number of occurrences of a given outcome divided by the total number of trials. The Bayesian interpretation may not look superior to the frequentistic one at first, but together with the standard rules of probability theory, from which $Bayes'\ Theorem$ is a trivial but rigorous result, a powerful method of analysis emerges. In principle, we can ask questions such as: What is the probability of the specific hypothesis θ (for example, θ may represent that m_t is 200 GeV/c^2, n_s is 20 events, etc) given the facts D (for example, D describing that N events were observed, the H_T distribution of N events was known, and $N_{background}$ events were expected from background processes, etc }) conditioned on certain assumptions I that is the status of information available regardless of D. Bayes' Theorem for any hypothesis θ_j states:

$$P(\theta_j|D,I) = \frac{P(D|\theta_j,I)P(\theta_j|I)}{\int_{for\ all\ \theta_i} d\theta_i P(D|\theta_i,I)P(\theta_i|I)}, \tag{7}$$

where $P(A|B)$ denotes a probability of proposition A, given that proposition B is true.

1. $P(D|\theta_j, I)$ is the probability assigned to data D, sometimes called the *likelihood*, given that the hypothesis θ_i and I are true.

2. $P(\theta_j|I)$ is the *initial*, or a *priori* probability of θ_i, i.e., the probability of the hypothesis irrespective of the facts D.

3. $P(\theta_j|D,I)$ is called the a *posteriori* probability of hypothesis θ_i, given D and I are true. This is the probability that is needed to make inferences. It is sometimes called the Bayes probability.

If one wants to know the a posteriori probability for a particular hypothesis, θ_o, namely $P(\theta_o|D,I)$, then an integration over the remaining hypotheses θ_i provides the answer:

$$P(\theta_o|D,I) = \int_{i\neq o} d\theta_i\ P(\theta_i|D,I). \tag{8}$$

This is the general method in Bayesian inference theory used to eliminate "nuisance" parameters. These being parameters that are not of great interest. No such general procedure exists in conventional statistics.

The question of what is the best estimate of an hypothesis θ_o, given its a posteriori probability $P(\theta_o|D,I)$, is, in fact, an open issue because it depends on what is used as the estimator. Jaynes[6] discusses two choices for the best estimator (θ_o^*) of a hypothesis θ_o. One is a mean estimate:

$$\theta_o^* = \int d\theta_o \; \theta_o P(\theta_o|D,I). \tag{9}$$

The other is the *median* which is also called Laplace's "most advantageous" estimator. The posteriori mean estimate is that which minimizes the mean square error, and the median estimate is the estimate which minimizes the expected magnitude $\mid \theta - \theta^* \mid$, referred to as Laplace's criterion. There are many measures of uncertainty on the estimated parameter, of which the variance of the a posteriori distribution

$$\sigma_{\theta_o}^2 = \int d\theta_o \; (\theta_o - \theta_o^*)^2 P(\theta_o|D,I), \tag{10}$$

is perhaps the most common. However, a 68.3% confidence interval can also be computed and taken that as the uncertainty[8].

$$0.683 = \int_{\sigma_{low}}^{\sigma_{high}} d\theta_o \; P(\theta_o|D,I). \tag{11}$$

If the likelihood is symmetric, and the a priori probability is isotropic over the range of consideration, the mean and the median estimate, as well as the maximum likelihood estimate, agree with each other. However, the Bayesian approach provides information about the hypothesis over the entire range that is considered, and not only for the single value that maximizes the likelihood. Furthermore, while a properly normalized likelihood is a true probability in Bayesian inference, it has no such interpretation in conventional statistics. Thus, a Bayesian approach yields the best estimate of a parameter, as well as a statement about its consistency with the given hypothesis.

3.1 Choice of the Likelihood and the a Priori Probability

In the top-quark mass analysis, the desired quantity is not the likelihood, $P(D|m_t,I)$, but rather the a posteriori probability $P(m_t|D,I)$, that is, the posteriori probability of the top-quark mass being m_t, *given* D, the observations or experimental facts and I, the a priori assumptions. In the case that n_s and n_b are also of interest, Bayes' theorem provides:

$$P(m_t, n_s, n_b|D, I) = \frac{P(D|m_t, n_s, n_b, I)P(m_t, n_s, n_b|I)}{\int dm_t dn_s dn_b P(D|m_t, n_s, n_b, I)P(m_t, n_s, n_b|I)}, \quad (12)$$

and the a posteriori $P(m_t|D, I)$ is obtained by simply "marginalizing" with respect to n_s and n_b; that is, by performing integrations over n_s and n_b.

Choice of the Likelihood

For our likelihood we can use the following:

$$P(D|m_t, n_s, n_b, I) = \Pi_{i=1}^{N} \frac{e^{-n} n^{N_i}}{N_i!} \quad (13)$$

where

$$n(m_t, d_i, n_s, n_b) = n_s f_s(m_t, d_i) + n_b f_b(d_i),$$

and N_i is the number of events in the i_{th} bin. This is, in fact, a standard binned likelihood function. Another choice could be the likelihood used earlier in this note:

$$\begin{aligned} P(D|m_t, n_s, n_b, I) \quad = \quad & \left[\frac{(n_s + n_b)^N e^{-(n_s+n_b)}}{N!}\right] \\ & \left[\Pi_{i=1}^{N} \frac{n_s f_s(m_t, d_i) + n_b f_b(d_i)}{n_s + n_b}\right]. \end{aligned} \quad (14)$$

But, in general, such an unbinned likelihood function always has a term that is not a function of m_t. This can cause divergences when integrating over $P(m_t|D, I)$. that is, it is possible to have the constant term is sizeable enough that $\int dm_t m_t P(m_t|D, I)$ becomes not well defined.

Choice of the a Priori Probability

The a priori probability $P(m_t, n_s, n_b|I)$ must, somehow, express total ignorance of the values of m_t, n_s, and n_b. How best to do that is an unresolved issue. One choice is the *uniform prior*:

$$P(m_t, n_s, n_b|I) \times dm_t dn_s dn_b = constant \times dm_t dn_s dn_b. \quad (15)$$

Another choice, still expressing total ignorance, is *Jeffrey's a Priori*[6], which for the mass of the top quark can be written as:

$$P(m_t, n_s, n_b|I) = \frac{1}{m_t}, \tag{16}$$

One can argue that objects in nature are distributed as 1/x. For example, size of animals, size of molecules, masses of elementary particles (which is our present concern), size of stars, etc, seem to be distributed according to 1/x [9]. This assumption favors small m_t values. The number of background events N_b and the uncertainty σ on N_b, can be taken as a priori knowledge of n_b:

$$P(m_t, n_s, n_b|I) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(n_b - N_b)^2/2\sigma^2}. \tag{17}$$

We will test the impact of various priors in the following section.

3.2 Monte Carlo Tests

To compare the Bayesian and likelihood methods, the selection criteria of Sect. 2 were also used in this analysis. However, a signal to background ratio of 50 to 50 was chosen to test the efficacy of the Bayesian formulation. Using the unbinned likelihood given in Eq. 14, with an m_t-independent prior for m_t, and with a Gaussian prior for n_b, we calculated the posteriori probabilities. After the imposition of the selection criteria of Table 1, a sample of fifty background events and fifty signal were retained from data of RUN IB and from ISAJET $t\bar{t}$ production at m_t=180 GeV/c^2, respectively. Figure 10 (a) shows the posteriori probability, $P(m_t, n_s|D, I)$, as a function of the number of top events and of the mass of the top quark indicating the expected signal. Figure 10 (b) shows contours of equal probability. The distributions show a peak near the input values of m_t and n_s, again demonstrating the clear signal in n_s - m_t space. Figure 10 (c) shows the posteriori probability distribution $P(m_t|D, I)$. The mean estimate on m_t is 183.9 GeV/c^2, and, more importantly, the uncertainty on m_t is 12.6 GeV/c^2, which can be compared with the result from the likelihood fit of 16.3 GeV/c^2 in Fig.4(c). It can be argued that because the uncertainty from the likelihood fit was based on fits to each ensemble of many pseudo experiments, it is not fair to compare the two numbers. However, a fundamental notion in Bayesian statistics is that inferences should be based on what is actually observed, and not on what we could have observed. Therefore, regardless of the fairness of the comparison, the sampling distribution of Bayesian estimates is cannot alter

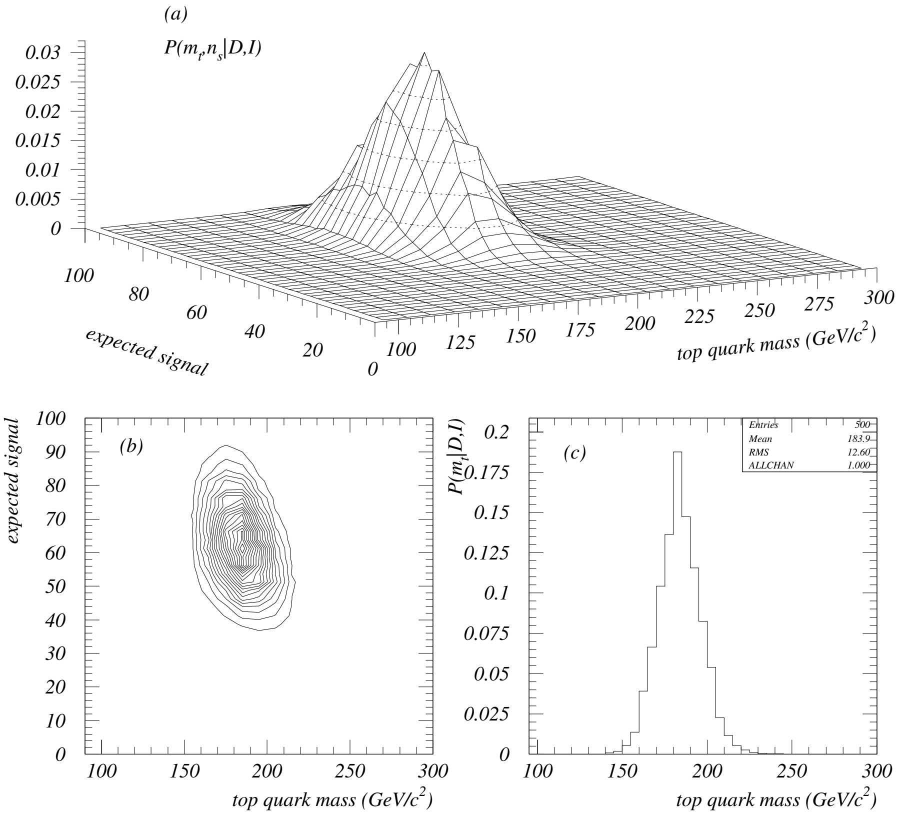

Figure 10: Posteriori probability distributions using an unbinned likelihood. A total of 50 signal events were chosen from ISAJET $t\bar{t}$ at $m_t = 180\ GeV/c^2$, and 50 background events from data of RUN IB. (a) $P(m_t, n_s|D, I)$ as a function of m_t and n_s. (b) Contour of equal probability of (a). (c) $P(m_t|D, I)$ as a function of m_t.

the inferences based on some given data set. This is because the posteriori probability, by construction, incorporates the likelihood (and therefore the actually observed data) and the assumed a priori probability, and, necessarily, describes all that we know about the hypothesis being tested. This is, precisely the importance of Bayes' theorem.

Since $P(m_t|D,I)$ is symmetric, and its shape is approximately Gaussian, one expects that the square root of the variance and the 68.3% confidence interval estimate will provide similar answers (as should the mean and the median estimates in this case). In general, however, these uncertainty measures can be different.

For our second choice of the likelihood, we used the binned likelihood function given in Eq. 13. The same signal to background ratio of 50 to 50 was used again for our study. We again assumed a Gaussian prior for n_b and an m_t-independent prior for m_t. As seen in Table 10, both the binned and unbinned likelihood provides the same uncertainty on the mass of the top quark of 12.6 GeV/c^2.

When we assume a $1/m_t$ form for the a priori on the mass of the top quark, with the unbinned likelihood the mean and the uncertainty on top quark mass become 175.0 GeV/c^2 and 13.1 GeV/c^2, respectively, and with the binned likelihood the mean and the uncertainty are 175.7 GeV/c^2 and 12.8 GeV/c^2, respectively. Figures 11 (a), (b), and (c) show the posteriori probabilities for the binned likelihoods. The enhancement of probability at $n_s \approx 70$, was observed only when the uniform was used for n_b, which indicates that one can introduce a bias in the measurement without imposing a Gaussian prior for n_b. However, it should be recognized that the concept of "bias" is not very important in Bayesian inference, because what matters is that the Bayesian estimates be "consistent", that is, converge to the true value of mass as more data are accumulated.

Table 10 summarizes the results using different combinations of likelihoods and priors. When $1/m_t$ is assumed for the prior, the mean value of m_t gets slightly lower than the case of the uniform prior, because $1/m_t$ favors lower values of top-quark mass. However, due to the fact that $P(m_t|D,I)$ is sharply peaked away from zero, assuming $1/m_t$ as the prior on the top quark mass did not change the mean and the uncertainty significantly. As expected, for cases when a Gaussian prior was assumed for n_b, the uncertainties on m_t were somewhat smaller.

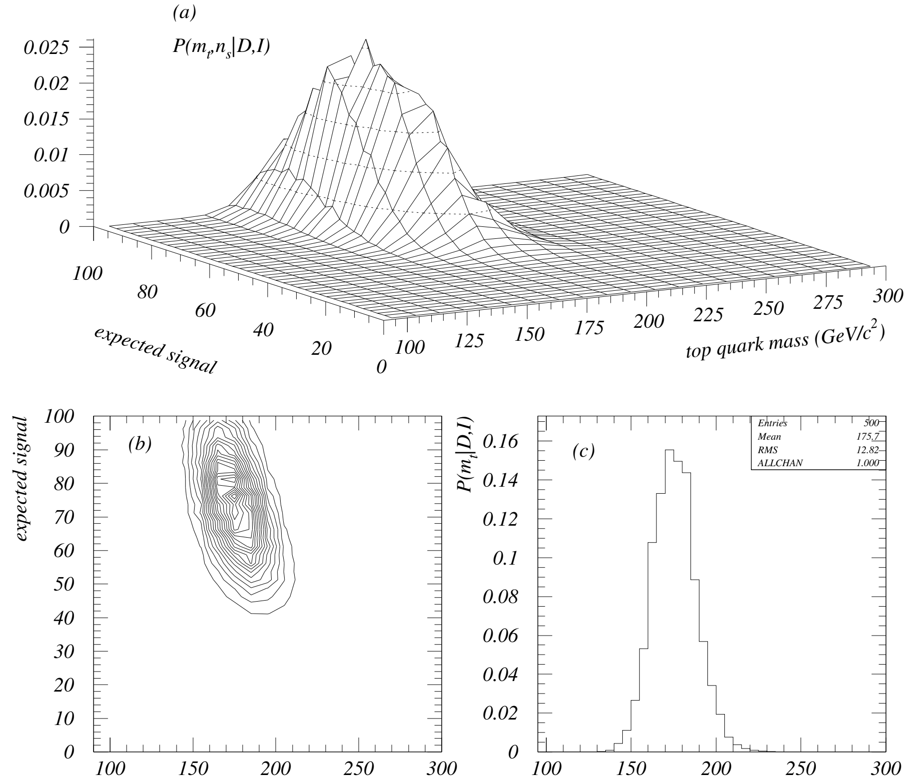

Figure 11: Posteriori probability distributions using a binned likelihood. A total of 50 signal events were chosen from ISAJET $t\bar{t}$ at $m_t = 180\,GeV/c^2$ and 50 background events from data of RUN IB. (a) $P(m_t, n_s|D, I)$ as a function of m_t and n_s. (b) Contours of equal probability in (a). (c) $P(m_t|D, I)$ as a function of m_t.

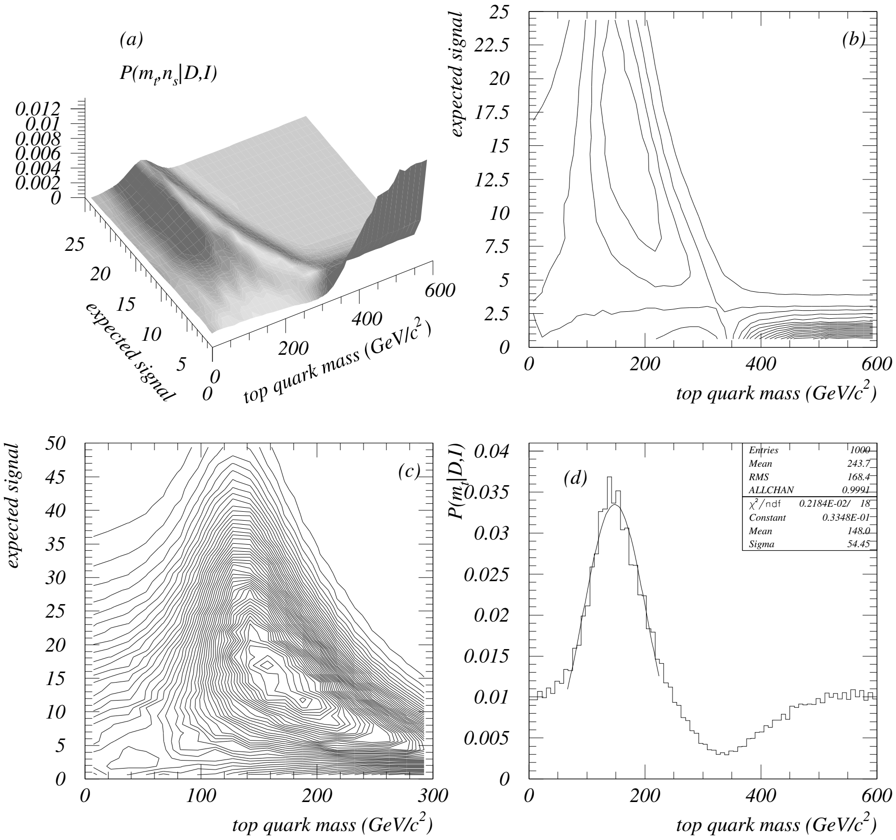

Figure 12: Posteriori probability distributions with unbinned likelihood, uniform prior on m_t, and Gaussian prior on n_b. The data are the 72 events from RUN IB analyzed previously using the standard likelihood procedure. (a) $P(m_t, n_s|D, I)$ as a function of m_t and n_s. (b) Contours of equal probability in (a). (c) Magnification of (b) in m_t. (d) $P(m_t|D, I)$ as a function of m_t.

3.3 Bayesian Extraction of the Mass of the Top Quark

Data for the total integrated luminosity of 70 events/pb of RUN IB data was used again to obtain the Bayesian estimation of the mass of the top quark. All the event selection criteria, including jet energy corrections, were done in exactly the same way as in the previous likelihood-fit analysis. Consequently, the results given in Table 4 are still valid for this analysis. To compare directly with the results of the likelihood fit, we take an unbinned likelihood, uniform prior for m_t, and a Gaussian prior for n_b. The choice of the posteriori probability $P(m_t, n_s|D, I)$ is shown in the Figs. 12 (a) and (b). There is a sharp rise at very high m_t, approximately at 550 GeV/c^2, where n_s is near zero. This is probably caused by the several events at high H_T values in Fig. 6 (b), because the peak appears only when n_s is near zero. However, we should note that above $m_t = 220$ GeV/c^2 the signal probability density function f_s is extrapolated from a two-dimensional fit at lower mass, and it may therefore not be valid for masses beyond $m_t = 220$ GeV/c^2. There is another rise at $m_t \approx 200$ GeV/c^2 that parallels the n_s axis that can be explained by the small excess of events near $H_T \approx 400$ GeV observed in Fig. 6 (b). Figure 12 (c) shows the same contour plot as 12(b) but with the restricted ranges of m_t to analyze contours in detail. There two peaks visible, one at $m_t \approx 160$ and one at 180 GeV/c^2. The peak near 180 GeV/c^2 is higher and can be interpreted as the one corresponding to the minimum in the likelihood fit. Figure 12 (d) shows the a posteriori probability, $P(m_t|D, I)$ as a function of m_t. A peak in the distribution is observed at 153.6, with an RMS uncertainty of 72.3 GeV/c^2. That is comparable with the result from the likelihood fit analysis. One difference here between the present treatment and the standard approach should be noted and that is that here the parameter n_b has been integrated out, whereas in the likelihood fit only one value of n_b, namely n_b^{min}, was considered in the extraction of the central value of m_t.

3.4 Systematic Uncertainty

The jet energy scale appears to be a primary source of systematic uncertainty in the extraction of the mass of the top-quark. In what follows, we treat this uncertainty as one more piece of prior information that is provided to us. Suppose that the fractional uncertainty on the jet energy is q, we can then write:

$$E_{jet}(q) = (1 + q)E_{jet}^{measured} \tag{18}$$

	mean (GeV/c^2)	uncertainty (GeV/c^2)
unbinned $P(D\|m_t, n_s, n_b, I)$ flat for n_b flat for m_t	176.0	13.5
unbinned $P(D\|m_t, n_s, n_b, I)$ Gaussian for n_b flat for m_t	183.9	12.6
unbinned $P(D\|m_t, n_s, n_b, I)$ flat for n_b $1/m_t$ for m_t	175.0	13.1
unbinned $P(D\|m_t, n_s, n_b, I)$ Gaussian for n_b $1/m_t$ for m_t	183.1	12.5
binned $P(D\|m_t, n_s, n_b, I)$ flat for n_b flat for m_t	176.6	12.9
binned $P(D\|m_t, n_s, n_b, I)$ Gaussian for n_b flat for m_t	184.3	12.6
binned $P(D\|m_t, n_s, n_b, I)$ flat for n_b $1/m_t$ for m_t	175.7	12.8
binned $P(D\|m_t, n_s, n_b, I)$ flat for n_b $1/m_t$ for m_t	183.5	12.6

Table 10: The mean and uncertainty on the top mass for different combinations of likelihoods and priors. An input mass of 180 GeV/c^2 and a signal to background ratio of 50/50, were assumed in these Monte Carlo studies of the Bayesian approach.

where $E_{jet}(q)$ denotes the rescaled jet energy and $E_{jet}^{measured}$ the measured jet energy, respectively. Because rescaling the jet energy changes the shape of signal and background as well as the distribution of events, f_s and f_b must therefore also depend on q:

$$f_s(m_t, d_i) \rightarrow f_s(m_t, d_i, q)$$
$$f_b(d_i) \rightarrow f_b(d_i, q)$$

The posteriori probability that takes proper account of the uncertainty in the jet energy scale is given by the following integral:

$$\begin{aligned} P(m_t|D,I) &= \int_{q_{low}}^{q_{high}} dq P(m_t|D,I) \\ &= \frac{1}{N} \int_{q_{low}}^{q_{high}} dq P(D|m_t,q,I) P(q|I) P(m_t|I), \end{aligned} \tag{19}$$

where

$$P(q|I) = \frac{1}{\sqrt{2\pi}\sigma_q} e^{-q^2/2\sigma_q^2}, \tag{20}$$

N is the normalization, and σ_q is the estimated uncertainty on the jet energy scale. While trivial in principle, this integral is nontrivial in practice because one has to rebuild f_s and f_b at every step of the integration over q. We generated two sets of 100,000 different values of q's(q_1 and q_2, to assign jet energy scales to f_s and f_b in a random manner.), assuming σ_q to be 5 %, and restricting fluctuation in q to within $\pm$ one σ_q of q=0. In each step of the integration over q, $f_s(q_1)$ and $f_b(q_2)$ were recalculated by repeating the fitting procedure explained in the Section 2.2. Figure 13 (a) shows the $f_b(q)$ function for the first hundred different points of q_2's. This stability shows that the fitting was done successfully on each step. We also scanned the results of the entire fits, and confirmed that χ^2/dof was always less than 1.2. Results for q-dependent f_s were also checked, and no evidence found for problems with the procedure. The contour plot of $P(m_t, n_s|D,I)$ in Fig. 13 (b) shows a slight shift of the distribution toward higher m_t values, but no significant changes in overall shape. The posteriori probability, $P(m_t|D,I)$ is shown in Fig. 13 (c) as the solid histogram. The dashed histogram is the same plot, but for the case when we do not consider uncertainty in the jet energy scale (equivalent to $P(q|I) = \delta(q-0)$), namely Fig. 12 (d). The mean value of m_t from a Gaussian fit to the peak shifted up by 5 GeV/c^2, and the standard deviation for the Gaussian fit increased from 54.5 to 72.3 GeV/c^2, for the q-dependent formulation.

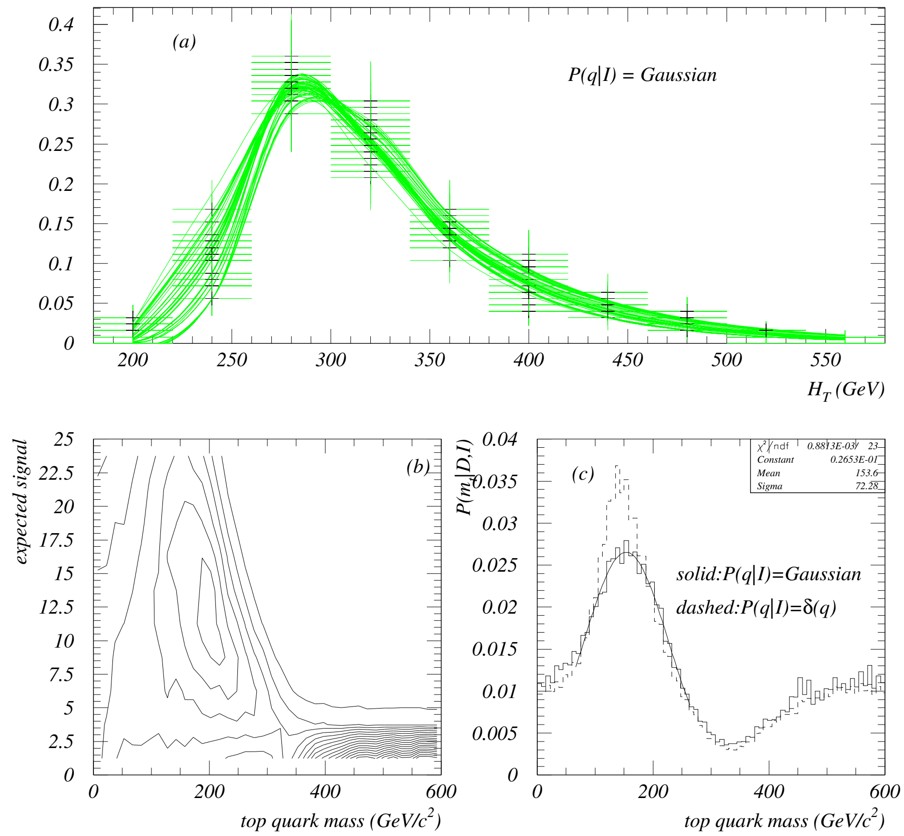

Figure 13: Posteriori probability distributions with unbinned likelihood, uniform prior on m_t, Gaussian prior on n_b, and Gaussian prior on the jet-energy scale. The standard 72 events from Run Ib data were used in the analysis. (a) First hundred f_b functions at hundred different q values. (b) Contour plot of $P(m_t, n_s|D, I)$. (c) Solid line shows $P(m_t|D, I)$ as a function of m_t for the analysis including the dependence on scale (q), and the dashed line corresponds to the result of Fig. 12 (d).

4 Summary

We have performed an optimization for the extraction of the mass of the top quark in multijet final states through extending the existing random-grid-search technique and through a detailed investigation of the likelihood function. We found that the level of confidence for the presence of the top quark is not strong enough at this stage of the analysis. If the excess of events relative to the estimated background is assumed to be from standard $t\bar{t}$ production, then the usual likelihood fit method provides a top quark mass of 179$\pm$ 37(statistical)+14-10(systematic) for 70 events/pb of RUN IB data.

We have also investigated Bayesian estimators for the mass of the top quark in multijet final states. When the posteriori probability is sharply peaked away from limiting values of the parameters, we found that both methods, likelihood and Bayesian, yielded comparable results. Taking the mode, that is, the peak value, of the posteriori probability as an estimator for the top quark mass, it gives an estimate of 150$\pm$79 GeV/c^2 (statistical+jet-energy systematic error) for the 70 events/pb data of RUN IB. However, at present level of statistics, the signal to background ratio is poor (just 15 to 57) and leads to a posteriori probability for data that is not sharply peaked. Consequently, we still have no clear evidence for a $t\bar{t}$ signal in the multihadronic channel. Besides anticipating more events from RUN 2, it would seem to be vital to improve the signal to background ratio in order to make any further advances in this channel.

5 Acknowledgements

I would like to thank Harrison Prosper providing many important points on Bayesian statistics, and Tom Ferbel for correcting all kinds of mistakes that I have made, critical reviews on Bayesian parts, and an excellent editorial assistance. I acknowledge the support of this research by the U.S. Department of Energy, the U.S. National Science Foundation, the Commissariat à L'Energie Atomique in France, the Ministry for Atomic Energy and the Ministry of Science and Technology Policy in Russia, CNPq in Brazil, the Departments of Atomic Energy and Science and Education in India, Colciencias in Colombia, CONACyT in Mexico,the Ministry of Education, Research Foundation and KOSEF in Korea, and the A. P. Sloan Foundation.

References

[1] S. Abachi, et al., "Observation of the Top Quark", Phys. Rev. Lett. **72** 2138, (1995).

[2] F. Abe, et al., Phys. Rev. **D** vol. **50** 2966,(1994).

[3] E. Won, DØ Note 2586, (1995).

[4] C. Stewart, C. Cretsinger, E. Won, DØ Note 2588a, (1995).

[5] C. Hill, S. Parke, Phys. Rev. **D49** 4454, (1995).

[6] E. T. Jaynes, "Probability Theory: The Logic of Science".

[7] G. D'Agostini, DESY 95-242, "Probability and Measurement Uncertainty in Physics - a Bayesian Primer" (1995).

[8] H. B. Prosper, U of DØ, "Bayesian Probability" (1995).

[9] F. James, "The Statistics of Very Small Samples - Bayesian and Classical Approaches" (1986), CERN Cours/Lecture Series(also DØ Note 1686).