CDF/ANAL/CDF/PUBLIC/5920
Version 3.0

A Summary of the Treatment of Systematics in the CDF Run I Top Quark Production Cross Section Measurement

Pekka K. Sinervo

University of Toronto

16 April 2002

1 Introduction

The CDF collaboration has published a final measurement of the top quark production cross section using data collected during 1992–1995 (Run I) [1]. This measurement is basically a counting experiment, where one relates the production cross section, $\sigma_{t\bar{t}}$, to the number of candidate events, $N_{t\bar{t}}$, using the formula

$$\sigma_{t\bar{t}} = \frac{N_{t\bar{t}} - N_{bkg}}{\epsilon L}, \tag{1}$$

where N_{bkg} is the number of background events in the candidate sample, ϵ is the efficiency for identifying a $t\bar{t}$ event as a member of the candidate sample, and L is the integrated luminosity for the experiment.

The "random" uncertainties associated with this measurement come from the Poisson fluctuations in $N_{t\bar{t}}$. This note summarizes the treatment of uncertainties in this measurement in a context where a number of channels were combined together to give a final Run I result.

The primary reference to the measurement discussed here [3] provides a very nice capsule overview of how the final result was determined. Details on the cross section measurements in the individual channels are provided in a series of notes referenced therein.

2 Definition of Uncertainties

We conventionally divide sources of uncertainty into two classes, "random" or "statistical" uncertainties, and "systematic" uncertainties. Random uncertainties are those that would normally scale by the total size of the data sample, typically with a $1/\sqrt{N}$ dependence. Systematic uncertainties can be defined in various ways [2], but for this measurement are those that affect a parameter or procedure used in extracting the result from the data.

Statistical uncertainties arise from the random fluctuations in the data sample. In the case of an observation where one counts the number of events, the Poisson fluctuations in the statistic of relevance, namely the number of observed events, would be the source of the statistical uncertainty. In more complex examples, it is typical to define a likelihood function that describes the data and use the shape of the likelihood function to determine a statistical uncertainty.

Systematic uncertainties, on the other hand, are the result of uncertainties in various parameters or inputs into the measurement that are required to interpret the results, or, more generally, uncertainties in the experimental conditions or the theoretical model used to interpret the data. These uncertainties can be characterized in a number of ways. There are in principle some sources of uncertainty where the effects scale with the total number of events, but are considered by convention to be systematic uncertainties. There are on the other hand, those sources of uncertainty that do not have any dependence on the size of the data sample. There are at two least reasons why we would wish to make the distinction between these two types of systematic uncertainty:

1. The uncertainties that depend on the event sample size will scale with the total integrated luminosity of the sample, and will therefore be reduced in future running. The other class of uncertainties will not be reduced unless steps are taken to improve the measurement.

2. The two classes of uncertainty are likely to create different correlations between our measurement and measurements made by a different experiment or different technique. This difference should be taken into account when combining two or more measurements.

3 The Total Cross Section Measurement

The CDF collaboration detected pair production of top quarks by selecting events in which there was evidence of the semileptonic or hadronic decay of two top quarks. Events in which both top quarks decay semileptonically, e.g., $t \to b \ell \nu_\ell$ where ℓ denotes either a muon or electron, are, by far the cleanest since the backgrounds associated with a dilepton signal are relatively low. It is, however, a small sample due to the approximately 10% semileptonic branching fraction for a single top quark. The largest sample of events arises in the case where both top quarks decay hadronically, e.g. $t \to b q \bar{q}'$, but this sample also is severely contaminated with background arising from QCD multi-jet production. The sample arising from the case where one top quark decays semileptonically and the other hadronically produces an event signature consisting of a single lepton candidate, missing transverse energy from the neutrino and in principle 4 quark jets. This "lepton+jets" channel is intermediate in size between the dilepton and all hadronic channels, and has moderate background contamination (most of the background arises from $W + b\bar{b} + X$ production).

Thus, the strategy used in the Run I cross section measurement was to select samples of top quark candidates for these three channels, estimate the backgrounds in each sample, and then correct the estimated number of signal events for acceptance and efficiency effects to derive a cross section measurement in each channel. The final step was to combine the measurements from the different channels into a single measurement.

Channel	Total Events	Background Rate	Aceptance $\times$ Efficiency
Dileptons	9	2.4 ± 0.5	0.0074 ± 0.0008
Lepton+Jets (SVX tags)	29	8.0 ± 1.0	0.039 ± 0.006
Lepton+Jets (SLT tags)	25	13.2 ± 1.2	0.012 ± 0.002
All Hadronic (kinematic)	42.8 ± 17.9	N/A	0.055 ± 0.012
All Hadronic (double tag)	36.7 ± 13.7	N/A	0.045 ± 0.014

Table 1: The data for the Run I top quark cross section measurement. The values for the "All Hadronic" channels represent the signal sample sizes after background subtraction.

The primary data for this measurement are summarized in Table 1. The method used to combine these data was to perform a maximum likelihood fit, where the likelihood was parametrized as a function of $\sigma_{t\bar{t}}$.

4 Random Uncertainties

The random uncertainties in this measurement arise from the total number of candidate events, namely 29 observed SVX-tagged events, 25 SLT-tagged events, 9 dilepton events and the observed rate of all-hadronic events above background. The two lepton+jets samples have an overlap of 7 events but this is ignored (as the effect has been shown to be less than 10% in the overall uncertainty of the cross section derived from these two measurements). There are two samples of hadronic candidate events with large backgrounds. The background-subtracted numbers of events for these two samples are 42.8 ± 17.9 and 36.7 ± 13.7, which are assumed to be correlated Gaussian statistics. The correlation in the two all-hadronic event samples was significant and was modelled by including a correlation coefficient $\rho = 0.34$ in the joint probability density for the number of observed all-hadronic events in the sample

$$G_{had}(N_{had1}, N_{had2}; \rho, \mu_{had1}, \mu_{had2}), \tag{2}$$

where μ_{had1} and μ_{had2} are the expected mean number of observed events (and are functions of the $\sigma_{t\bar{t}}$). Note that this latter term does not include any background contributions as the event rates N_{had1} and N_{had2} are both numbers of signal events above background. The form of this probability density is shown in Fig. 1.

The uncertainties associated with the observed event rates are incorporated into the final result through the use of a maximum likelihood fit. In effect, the numbers of dilepton and lepton+jet events are treated as Poisson statistics. In order to combine the "all hadronic" events, the likelihood function includes the two-dimensional Gaussian probability distribution G_{had}. The means μ_{had1} and μ_{had2} of this Gaussian distribution would be given by the total top quark cross section times its branching fraction and acceptances into the two all-hadronic channels. The natural logarithm of this likelihood function is shown in Fig. 2, where we fix each of the other parameters in the calculation to their nominal value and just vary $\sigma_{t\bar{t}}$.

In order to include the uncertainties in the measurement, the overall form of the likeli-

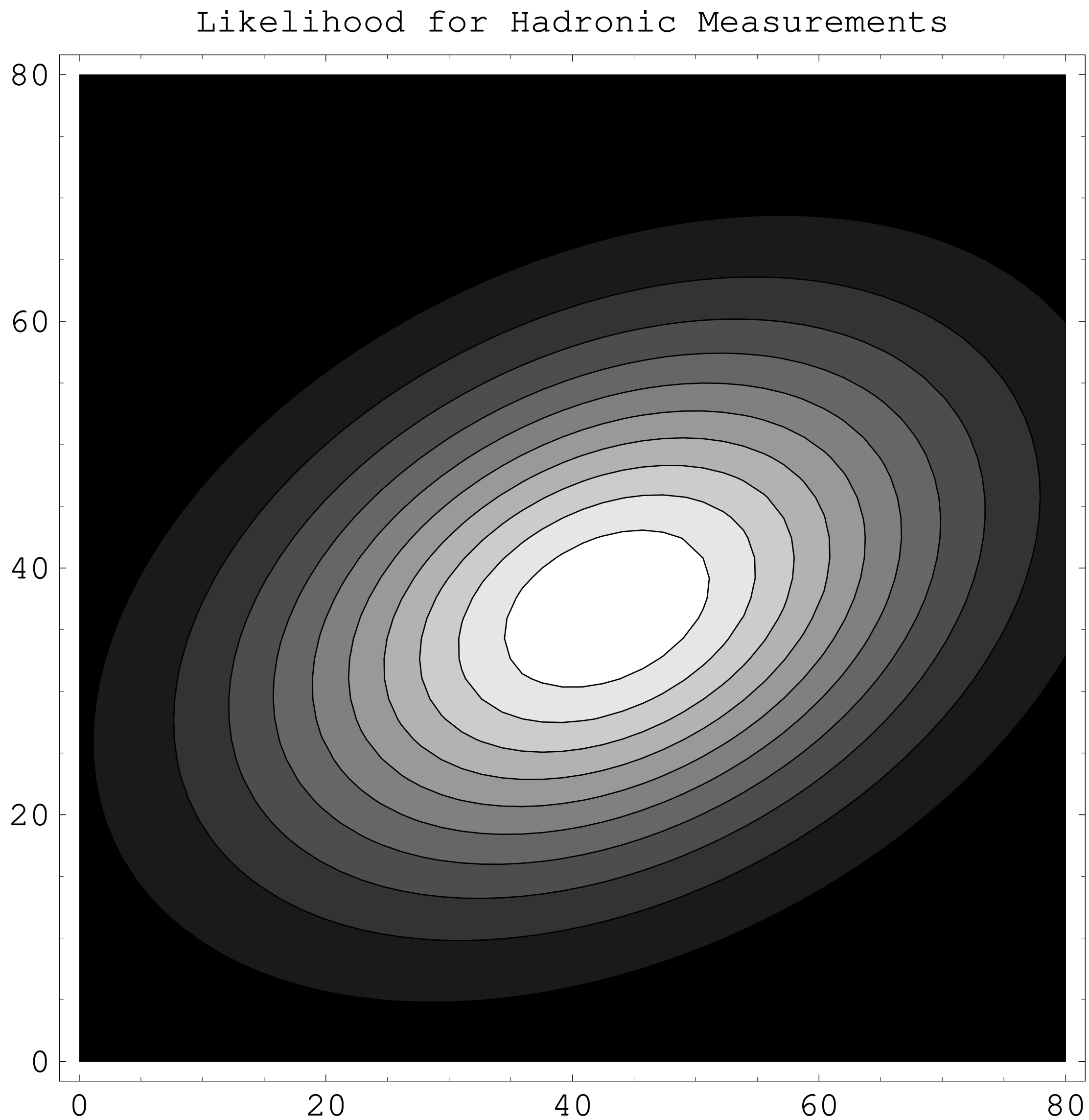

Figure 1: The form of the likelihood distribution for the two correlated hadronic channel measurements. The vertical and horizontal axis are μ_{had1} and μ_{had2}, respectively. The contours are in units of 0.1 in likelihood.

hood function is augmented from that described above to be

$$L = \left[\prod_{i}^{svx;slt;dil} P(N_i; \mu_i(\sigma_{t\bar{t}}; \ldots)) \right] \times$$

$$G_{had}(N_{had1}; N_{had2}; \rho; \mu_{had1}; \mu_{had2}) \times \prod_{j=1}^{16} G(X_j; \bar{X}_j; \sigma_j); \qquad (3)$$

where the first two factors represent the statistical uncertainties (the Poisson probability distributions for the observed number of events and the Gaussian probability distribution describing the uncertainties in the two all-hadronic rates). The last factor represents the 16

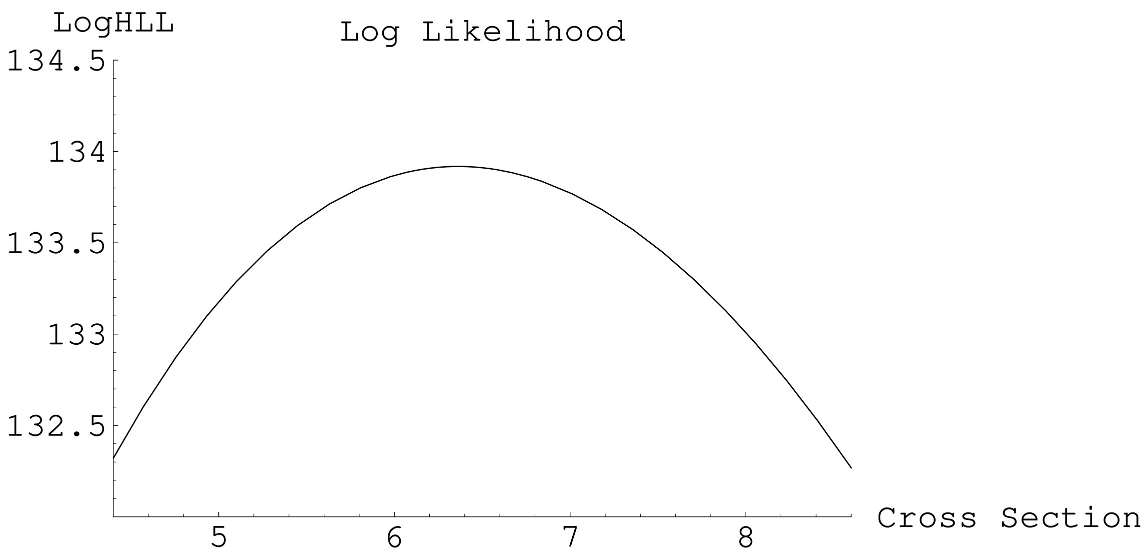

Figure 2: The dependence of the logarithm of the likelihood on $\sigma_{t\bar{t}}$. The value for each "nuisance parameter" has been set to its nominal value.

sources of systematic uncertainty, discussed below, where G is a one-dimensional Gaussian distribution with the form

$$G(X; \hat{X}, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(X - \hat{X})^2}{2\sigma^2}\right), \tag{4}$$

where X is the fitted deviation of the parameter suffering the systematic uncertainty, $\hat{X}$ is estimate of the parameter determined from independent studies and σ is its associated uncertainty.

5 Systematic Uncertainties

The philosophy chosen in this analysis is to estimate the individual effects that created further uncertainty in the resulting top quark cross section. These effects ranged from detector acceptance, trigger efficiency and lepton identification efficiency to uncertainties in the modelling of the top quark production process (which then affect the measurement by modifying the expected top quark acceptance and efficiency). Overall, there are 16 different sources of systematic effects, parametrized by X_j, $j = 1, 2, \cdots, 16$ with uncertainties ranging in relative size from about 4% to about 70%. Typically, each of these have been estimated from studies based on either data or Monte Carlo samples (often both), with the goal of identifying a range that corresponds to a 68% confidence level interval. The very largest uncertainty (70%) is associated with a relatively small component of a background source and so does not contribute significantly to the overall uncertainty in $\sigma_{t\bar{t}}$. The rates of expected events N_i are now functions of the X_j as well as $\sigma_{t\bar{t}}$.

Although the sources of these additional uncertainties are quite different, they are incorporated into the $\sigma_{t\bar t}$ measurement by treating them as Gaussian statistics $\{X_j\}$ with normal distributions characterized by the uncertainty σ_j assigned to the central value $\tilde{X}_j$ of each parameter. With this assumption, the systematic uncertainties can be included in the likelihood function, as illustrated in Eq. 3. The cost of this is that the likelihood is now no longer just a function of the parameter of interest ($\sigma_{t\bar t}$), but also of 16 "nuisance" parameters.

To proceed from the likelihood to a measured value and uncertainty for the parameter of interest, one has to adopt a statistical framework to express the result. This analysis implicitly assumes a Bayesian approach where one uses the likelihood and an assumption for a prior for $\sigma_{t\bar t}$ to obtain a posterior probability. Since the likelihood incorporates the systematic uncertainties as "nuisance parameters", a standard technique to factor them out is the "profile likelihood" technique. The log of the likelihood function is maximized using MINUIT with respect to the $\{X_j\}$ for each value of $\sigma_{t\bar t}$ to obtain a function of only $\sigma_{t\bar t}$. One then assumes a uniform prior for $\sigma_{t\bar t}$ and multiplies the profile likelihood with this prior to obtain a posterior probability. The uncertainty in the cross section is then defined to be the one-half unit change in the log-likelihood of this posterior probability from its maximum value. The result is

$$\sigma_{t\bar t} = 6.55^{+1.68}_{-1.41}\ \mathrm{pb}, \tag{5}$$

where the uncertainty now includes both statistical and systematic effects. Note that this is approximately 0.15 larger than the maximum likelihood estimate for $\sigma_{t\bar t}$ that one would "read off" of Fig. 2. This is due to the fact that the systematic uncertainties have different effects on the five event rates that are being constrained to a common mean.

6 Discussion

The choice to include in the likelihood function all of the systematic uncertainties is one that has been employed in many other analyses within CDF [4]. It has the convenient effect of burying in the likelihood both statistical and systematic uncertainties and providing a formula for combining them. Clearly, this procedure rests on the assumption that one can describe uncertainties arising from the systematic effects with Gaussian probability distributions (with a large enough number of independent systematic effects, one can invoke the mean limit theorem, but the sensitivity of the result to the form of the probability distributions for the nuisance parameters should be checked in each analysis).

Of perhaps more immediate interest is the interpretation of this likelihood function for those wishing to use it to place confidence intervals on the $t\bar t$ production cross section, as was done in Eq. 5. In what was actually done, a Bayesian approach was used once a profile likelihood had been obtained, thereby factoring out the nuisance parameters. An alternate Bayesian approach [5] would have been to define the likelihood without the product of the Gaussian systematic terms, i.e. as

$$L_{alt} = \left[\prod_{i}^{svx,slt,dil} P(N_i; \mu_i(\sigma_{t\bar t}, \ldots)) \right] \times G_{had}(N_{had1}, N_{had2}; \sigma, \mu_{had1}, \mu_{had2}). \tag{6}$$

One now could assume priors for each of the parameters. A natural choice would be use a Gaussian distribution for the prior for each of the systematic uncertainties, and a prior for $\sigma_{t\bar{t}}$ that is uniform from $\sigma_{t\bar{t}} = 0$ to infinity.[1] The product of L_{alt} with these prior distributions results in a 17-dimensional posterior density that one can then use to set credibility intervals. A typical approach would be to integrate out the nuisance parameters, resulting in a reduced posterior density that is a function of $\sigma_{t\bar{t}}$ only, and that can then be used to set Bayesian credibility intervals for $\sigma_{t\bar{t}}$.

A frequentist statistician could use the likelihood to determine the best measurement of $\sigma_{t\bar{t}}$. However, the most robust technique to define a confidence interval given the complexity of the likelihood function would be to perform a Monte Carlo calculation. She or he would have to define the appropriate ensemble of repetitions of the measurement, taking into account changes in procedure and assumptions that result in the variations in the parameters that introduce the "systematic uncertainties". To the extent that each parameter does have a single true value and the experiment provides different estimates of its value when repeated (as modelled by the Gaussian distribution function incorporated into the likelihood function), a standard frequentist confidence interval could be determined. However, this approach could be computationally very prohibitive. Practical approaches to address this calculation are available [6].

This interpretation is not possible in principle in the context where the systematic uncertainty arises from theoretical considerations that we are unable to constrain from the data. In this case, an alternate approach would be to not include such theoretical effects as systematic uncertainties, but instead to characterize the sensitivity of the final result (and its uncertainty) on variations in the theoretical assumptions. This would more clearly identify such assumptions and allow different experiments to more effectively combine their measurements using a common framework.

References

[1] Measurement of the $t\bar{t}$ cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV, T. Affolder et al., Phys. Rev. D 64, 032002 (2001).

[2] The CDF Statistics web page section on systematic uncertainties gives several operational definitions and practical examples.

[3] The New Combined CDF $t\bar{t}$ Production Cross Section for Run 1, CDF Note 5043, The Top Group (June 1999).

[4] Measurement of the CP Violation Parameter $\sin(2\beta)$ in $B^0\overline{B^0} \to J/\psi K$ Decays, F. Abe et al., Phys. Rev. Lett 81, 5513-5518 (1998).

Measurement of b quark fragmentation fractions in the production of strange and light B mesons in proton anti-proton collisions at $\sqrt{s} = 1.8$ TeV, F. Abe et al., Phys. Rev. D. 60, 092005 (1999).

[1] Although this prior is "improper" and difficult to justify for an objective Bayesian, it is nonetheless a choice that reflects our unwillingness to assume anything about the parameter we are specifically performing the experiment to measure. Note, however, that this choice of prior is dependent on the metric used. For example, if we had chosen a prior uniform in $\sigma^2_{t\bar{t}}$, we would obtain different credibility intervals.

Measurement of b Quark fragmentation fractions in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV, T. Affolder et al., Phys. Rev. Lett. 84, 1663-1668 (2000).

[5] L. Demortier, Bayesian Treatment of Systematic Uncertainties, CDF Note CDF/PUB/CDF/PUBLIC/5928, in preparation.

[6] G. Punzi, in preparation.