CDF/ANAL/TOP/CDFR/9117
Version 2.3
June 11, 2008

Search for Single Top Quark Production using the Matrix Element Technique

Florencia Canelli,[1] Peter Dong,[2] Bernd Stelzer,[2], Rainer Wallny [2]

[1] *Fermilab*
[2] *University of California, Los Angeles*

Abstract

We present an improved search for electroweak single top quark production using 2.2 fb^{-1} of CDF II data collected between 2002 and 2007. The analysis employes a matrix element technique which is used to calculate event probability densities for the signal and background hypothesis. The ratio of signal and background event probability densities is used as a discriminant variable which we will fit to the data. We search for a combined single top s- and t-channel signal as well as s- and t-channel separately. We measure a cross section of $2.3^{+0.8}_{-0.7}$ pb assuming a top quark mass of 175 GeV/c^2. This allows a measurement of $|V_{tb}| = 0.90^{+0.14 experiment}_{-0.15} \pm 0.07^{theory}$. and $|V_{tb}| > 0.61$ at 95% confidence. We use the MCLIMIT [20] to calculate the signal significance. The expected (median) p-value for this analysis is 0.0000050 which corresponds to a 4.4σ signal significance assuming single top quark production at the rate predicted by the Standard Model. The observed p-value is 0.00021, corresponding to a 3.5σ excess in data.

Contents

a

0 Changes since Blessing for Winter Conferences 2008

Since blessing for the Winter Conferences 2008 we performed a few (small) changes which are summarized below. The changes resulted in a shift in cross section on the order of 0.1 pb and a shift in sensitivity/significance of 0.1 standard deviations, as shown in the summary table below.

	Winter Conference 2008	Publication 2008 (2.2/fb)
Cross section	$2.2^{+0.8}_{-0.7}$ pb	$2.3^{+0.8}_{-0.7}$ pb
Expected p-value	4.5 sigma	4.4 sigma
Observed p-value	3.4 sigma	3.5 sigma

- Removed truncation of shape uncertainties for JES, ISR, FSR, and PDF as decided in the top group meeting (May 1st 2008), described in CDF9251.
- Systematic shape errors are smoothed using a 5 bin median smoothing procedure to remove statistical fluctuations in several low statistics systematic samples.
- The background estimation has changed slightly due to the following effects:
  - A small bug in the background estimate was fixed : The Method2 prediction had been counting the integer (jet.isTag) instead of (jet.isTag==1) which resulted in counting negative tags as well as positive tags (mostly affected the charm background)
  - New 2.2 fb^{-1} mistag matrix and asymmetry corrections are included in the background estimate (remove the previously used period 13 scale factor with additional uncertainty).
  - The JointPhysics scale factor class has been extended to period13 data (and not scaled from data period 12)
  - CMX Arches and CMX Miniskirt/Keystone contributions are accounted for in the lepton ID and trigger efficiency scale factors described in CDF9185.
- The determination of PDF errors are changed to be in sync with other top quark analyses at CDF.
- ISR and FSR rate and shape variations are treated as correlated uncertainties, since deviation from the default settings would influence both effects in the same way.

1 Introduction

Finding single top quark production is challenging since it is rarely produced ($\sigma_{singletop} \sim$ 2.9pb) in comparison with other processes with the same final state like W+jets and $t\bar{t}$. The signal to background ratio of the analysis is small, typically on the order of S/B$\sim$1/18 [1, 2]. This calls for a better discrimination of signal and background events which can be achieved by using more information to characterize each event.

This note describes a new analysis technique that attempts to make optimal use of information in the data. The implementation of the method is derived from a precision measurement of the top quark mass and W helicity in $t\bar{t}$ lepton+jets events using a matrix element analysis technique [3]. Although the main ingredients of the analysis are based on the calculation of event probabilities as in the mentioned analyses, this note presents a novel approach on how to apply these event probability densities to a search.

For the current iteration, our analysis was significantly improved. We have added several matrix element calculations for background sources (most notably a $t\bar{t}$ probability density, see section 2.6), added new "loose" muon types (as described in [4]) and extended the acceptance to the W+3 jet bin as well as improved transfer functions (see section 2.4.1).

2 Methodology

The method presented here relies on the evaluation of event probability densities for signal and background processes based on the Standard Model differential cross-section calculation.

The data for the first Tevatron Run II single top publication [5] was of the form of measured values of the kinematic variable H_T for each observed event. The shape of the H_T templates used in the likelihood method corresponds to the shape of the differential cross-section distributions $d\sigma/dH_T$ for signal and background. An improved discrimination between signal and background processes could be achieved by including all kinematical shapes or variables. That is by calculating the fully differential cross section on an event-by-event basis for the signal hypothesis and the background hypothesis to quantify how likely the event is to be either signal or background.

In general a differential cross-section is given by [6]:

$$d\sigma = \frac{(2\pi)^4 |M|^2}{4\sqrt{(q_1 \cdot q_2)^2 - m_{q_1}^2 m_{q_2}^2}} \, d\Phi_n(q_1 + q_2; p_1, .., p_n) \tag{1}$$

where $|M|$ is the Lorentz invariant matrix element; q_1, q_2 and m_{q_1}, m_{q_2} are the four momenta and masses of the incident particles; and $d\Phi_n$ is the n-body phase space given by [6]:

$$d\Phi_n(q_1 + q_2; p_1, .., p_n) = \delta^4(q_1 + q_2 - \sum_{i=1}^{n} p_i) \prod_{i=1}^{n} \frac{d^3 p_i}{(2\pi)^3 2E_i} \tag{2}$$

The CDF detector would be 'ideal' if we could measure all four momenta of the initial and final state particles very precisely. In this case we could use this formula without modification and normalize it to the total cross section to define the event probability:

$$P_{evt} \sim \frac{d\sigma}{\sigma}$$

However, several effects have to be considered: (1) the initial state interaction is initiated by partons inside the proton and antiproton, (2) neutrinos in the final state are not identified directly, and (3) the energy resolution of the detector can not be ignored. To address the first point, the differential cross section is folded over the parton distribution functions. To address the second and third points, we integrate over all particle momenta which we do not measure (e.g. p_z of the neutrino), or do not measure very well, due to resolution effects (e.g. jet energies). The integration reflects the fact that we want to sum over all possible particle variables (y) leading to the observed set of variables (x) measured with the CDF detector. The mapping between the particle variables (y) and the measured variables (x) is established with the transfer function, $W(y,x)$. After incorporating the effects mentioned above, the event probability takes the form:

$$P(x) = \frac{1}{\sigma} \int d\sigma(y) dq_1 dq_2 f(x_1) f(x_2) W(y,x) \tag{3}$$

where $d\sigma(y)$ is the differential cross section in terms of the particle variables; $f(x_i)$ are the PDFs, with x_i being the fraction of the proton momentum carried by the parton ($x_i = E_{q_i}/E_{beam}$); and $W(y,x)$ is the transfer function. Substituting Equation 1 and 2 into Equation 3, and considering a final state with four particles (n=4), the event probability becomes:

$$P(x) = \frac{1}{\sigma} \int 2\pi^4 |M|^2 \frac{f(x_1)}{|E_{q_1}|} \frac{f(x_2)}{|E_{q_2}|} W(y,x) d\Phi_4 dE_{q_1} dE_{q_2} \tag{4}$$

where the masses and transverse momenta of the initial partons are neglected (i.e. $\sqrt{(q_1 \cdot q_2)^2 - m_{q_1}^2 m_{q_2}^2} \simeq 2E_{q_1}E_{q_2}$).

As it will be explained in the next section, in this analysis we calculate event probabilities with the signal hypothesis for s-channel and t-channel and with background hypothesis with the major contributions.

We construct an event probability discriminant, EPD, using the signal and background probabilities to measure the single top content in data. The idea is to construct a distribution which is a superposition of the single-top and background event-probability-discriminant shapes and find the single-top and background normalizations such that *if the N events observed in data were chosen from respective single-top and background Poisson distributions having those means, the likelihood of observing these particular data would be maximized.*

Section 2.1 presents the event selection employed and the expected number of events. In Section 2.3 we will show the different matrix element used in the event

probabilities. In Section 2.4 we will derive the transfer function and make certain assumptions on the mapping between particles and measured objects. Section 2.5 will show the calculation of the phase space factor. Section 2.6 will add all these pieces into event probabilities. We explain the construction an event probability discriminant using the signal and background probabilities in Section 2.8 and the likelihood used to fit the data in Section 2.9. The rest of the note discusses the systematic uncertainties and the results on the data.

2.1 Event Selection and Expected Number of Events

Candidate events for this analysis are selected by requiring a W + 2 jet event topology where the W decays leptonically, $W \to e\nu_e$ and $W \to \mu\nu_\mu$. One or both of the two jets should be identified as a b-jet using the secondary vertex tag requirement. The detailed event selection and the estimate strategy has been performed and summarized in a separate CDF note [1, 2]. Table 1 lists the expected event yield. For the MC based background estimates [2] only covers uncertainty on the event detection efficiency.

Process	Number of Events in 2.2 fb^{-1}		Sample
	W + 2 jets	W + 3 jets	
s-channel signal	40.3 ± 5.8	13.1 ± 1.9	stop00
t-channel signal	60.8 ± 8.9	19.9 ± 2.6	stopm0
$Wb\bar{b}$	451.1 ± 136.0	138.0 ± 41.7	btopXp
$Wc\bar{c} + Wc$	372.5 ± 114.8	103.2 ± 31.8	ctopXw,stopwX
$Mistags$	337.1 ± 41.9	101.6 ± 12.8	ptopXw, utopXw
$non-W$	60.5 ± 24.2	21.0 ± 8.4	anti-electron / jet-electron / non-isolated data
WW	40.5 ± 4.5	14.3 ± 1.6	itopww
WZ	20.0 ± 1.6	5.8 ± 0.5	itopwz
ZZ	0.6 ± 0.1	0.3 ± 0.1	itopzz
$Z + jets$	25.5 ± 3.8	10.5 ± 1.5	ztopXY
$t\bar{t}$ dilepton	48.2 ± 6.9	40.9 ± 5.8	ttop75
$t\bar{t}$ non-dilepton	93.8 ± 13.4	286.9 ± 40.8	ttop75
Total signal	101.1 ± 14.7	30.0 ± 4.7	
Total prediction	1550.8 ± 256.6	753.7 ± 87.6	
Observed in data	1546	719	

Table 1: Number of expected single top and background events in 2.2/fb of CDF data passing all event selection requirements.

As mentioned in the previous section, in this analysis, we calculate event probability densities for the s-channel and t-channel single top processes, as well as for the $Wb\bar{b}$,

Wcj, Mistags (Wgg) and $t\bar{t}$ background processes. The diboson, Wcc and non-W events are assumed to be represented fairly well by the background probability density.

2.2 Non-triggered Muons

In this iteration of the analysis, we have included new non-triggered muons according to the work of [4]. The acceptance gain is equivalent to 50% of the CMUP contribution for the signal sample, while the background acceptance is increased by about 35%. The overall sensitivity gain is estimated to be 7%.

2.3 Matrix Element

We calculate the matrix element ($|M|^2$) for the event probability at leading order perturbation theory by using the HELAS (HELicity Amplitude Subroutines for Feynman Diagram Evaluations) package [7]. The correct subroutine for a given process are automatically generated by the MadGraph program [8]. We use different subroutines for calculating event probabilities for the s-channel, t-channel, $Wb\bar{b}$, Wcj, Wgg, and $t\bar{t}$ hypotheses in the two-jet bin. Figures 1 to 6 show the different Feynman diagrams used for each channel.

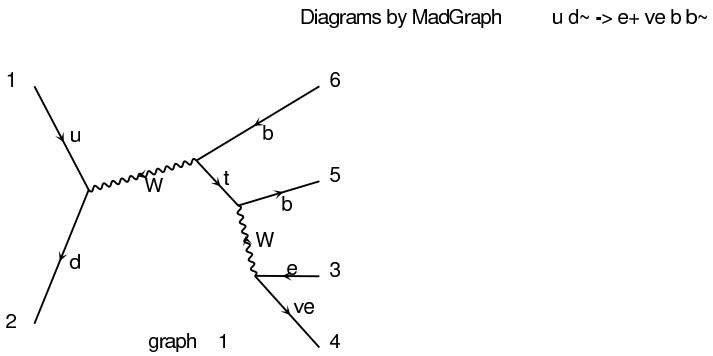


Figure 1: Leading order Feynman diagram for s-channel single top quark production and decay used in the calculation of the s-channel event probability in the two-jet bin.

In the three-jet bin, we calculate event probabilities for s-channel, t-channel, $WB\bar{b}$, and $t\bar{t}$, as shown in Figures 7 to 10.

2.4 Transfer Function

The transfer function, $W(y, x)$, provides the probability of measuring the set of observable variables (x) that correspond to the set of production variables (y). The set (y) represents all final state particle momenta at the particle level, while the set (x) represents the measured momenta (of the corresponding object) with the CDF detector. In the case of well-measured objects, $W(y, x)$ is taken as a δ-function (i.e. the measured momenta are used in the differential cross section calculation). When the

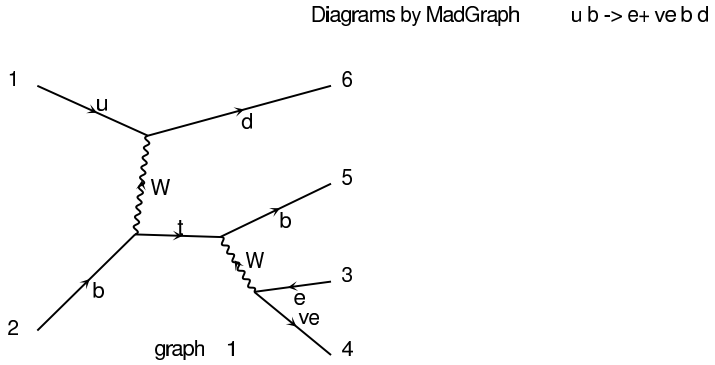


Figure 2: Leading order Feynman diagram for t-channel single top quark production and decay used in the calculation of the t-channel event probability in the two-jet bin.

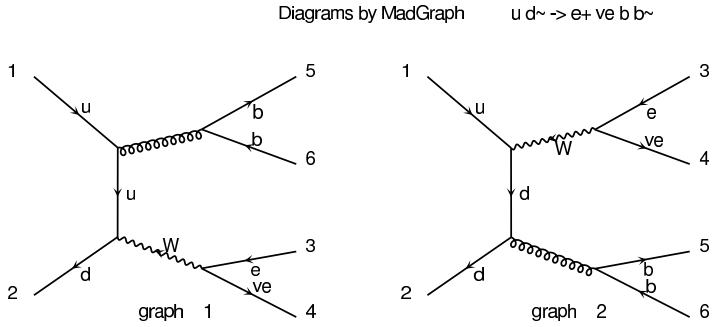


Figure 3: Leading order Feynman diagram for $Wb\bar{b}$ production and decay used in the calculation of the $Wb\bar{b}$ event probability in the two-jet bin.

detector resolution cannot be ignored, $W(y,x)$ is taken as a Gaussian-type function. For unmeasured quantities, like the momenta of the neutrino, the transfer function is unity (the transverse momenta of the neutrino, however, can be inferred from energy and momentum conservation).

Lepton momenta are well-measured with the CDF detector and we will assume δ-functions for them (first factor of Equation 5). The jet angular resolution of the calorimeter is also good (on the order of $\sigma_{\Delta_R} \simeq 0.07$) and we assume δ-functions for the transfer function of the jet directions (second factor of Equation 5). The resolution of the measured jet energies, however, is not negligible and the transfer function is derived in Section 2.4.1. Using these assumptions, $W(y,x)$ takes the following form for the four particle final state we consider in the single top search (lepton, neutrino and two jets):

$$W(y,x) = \delta^3(\vec{p_l}^{\,y} - \vec{p_l}^{\,x}) \prod_{i=1}^{2} \delta^2(\Omega_i^y - \Omega_i^x) \prod_{j=1}^{2} W_{jet}(E_{parton_j}, E_{jet_j}) \tag{5}$$

where $\vec{p_l}^{\,y}$ and $\vec{p_l}^{\,x}$ are the produced and measured lepton momenta, Ω_i^y and Ω_i^x are the produced quark and measured jet angles, and E_{parton_j} and E_{jet_j} are the produced quark and measured jet energies.

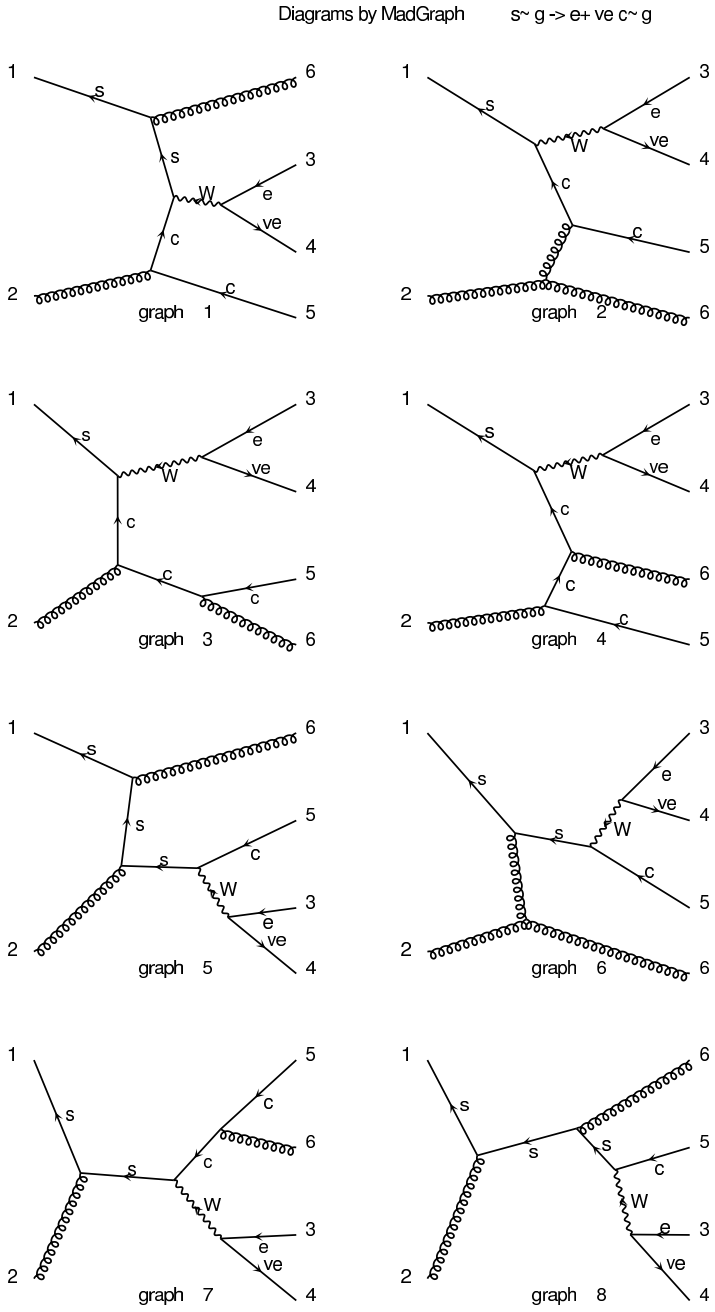


Figure 4: Leading order Feynman diagrams for Wcj production and decay used in the calculation of the Wcj event probability in the two-jet bin.

2.4.1 Jet-Parton Energies Transfer Functions

The transfer between parton and jet energies is determined by the transfer function $W_{jet}(E_{parton}, E_{jet})$. The standard CDF jet energy corrections correct the energies of jets in a way that the means of the corrected jet energies and the original parton energies

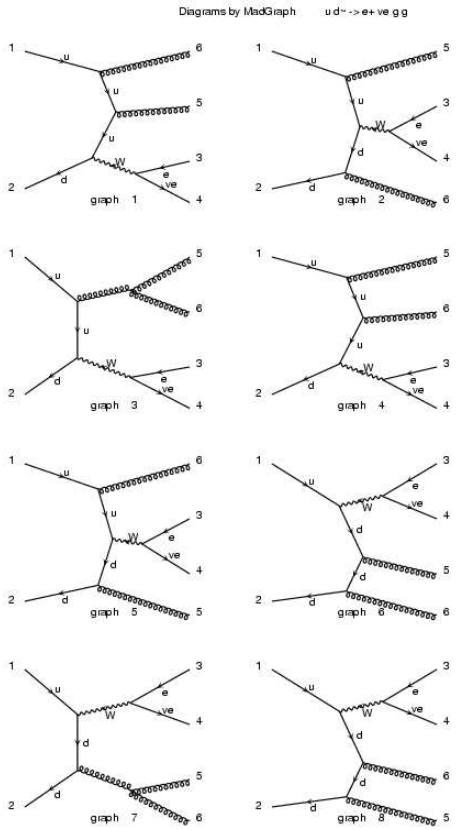


Figure 5: Leading order Feynman diagrams for Wgg production and decay used in the calculation of the Wgg event probability in the two-jet bin.

are equal. Such corrections, however, do not account for the shape of the difference in energies: the shape of the $\delta_E = (E_{parton} - E_{jet})$ distribution). This distribution is asymmetric and features a significant tail at positive δ_E, as shown in Figure 11.

We parameterize the δ_E distribution as a sum of two Gaussian functions: one to account for the sharp peak and one to account for the asymmetric tail:

$$W_{jet}(E_{parton}, E_{jet}) = \frac{1}{\sqrt{2\pi}(p_2 + p_3 p_5)}(exp\frac{-(\delta_E - p_1)^2}{2p_2^2} + p_3\ exp\frac{-(\delta_E - p_4)^2}{2p_5^2}) \quad (6)$$

where the parameters p_i have a linear dependence on E_{parton}, i.e.

$$p_i = a_i + b_i E_{parton}$$

A total of 10 parameters $(a_1, b_1, .., a_5, b_5)$ are therefore required to specify $W_{jet}(E_{parton}, E_{jet})$.

After specifying the transfer function, we can apply the general event probability of Equation 4 to the case of the single top analysis.

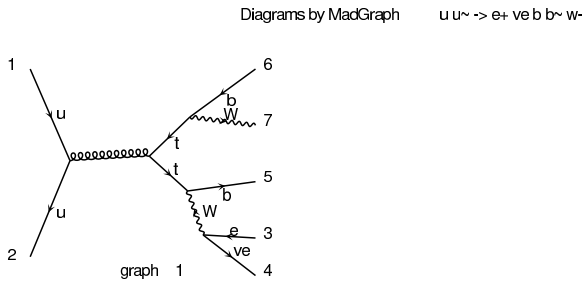


Figure 6: Leading order Feynman diagrams for $t\bar{t}$ production and decay used in the calculation of the $t\bar{t}$ event probability in the two-jet bin.

For the summer 2007 analysis we have implemented transfer functions with more information. We now use the energies stored in cone 0.7 jets when available. We also include different transfer functions for 3 different detector η regions.

Our event selection still requires 2 cone 0.4 jets. We use transfer functions obtained with 0.7 cone jets for events where we find only 2 tight cone 0.7 jets. In the s-channel MC this is 84% of the events. Tight cone 0.7 jets are defined in the same way as cone 0.4 (p_T >20 GeV and $|\eta|$ <2.8). The remaining 16% events are treated with cone 0.4 transfer functions.

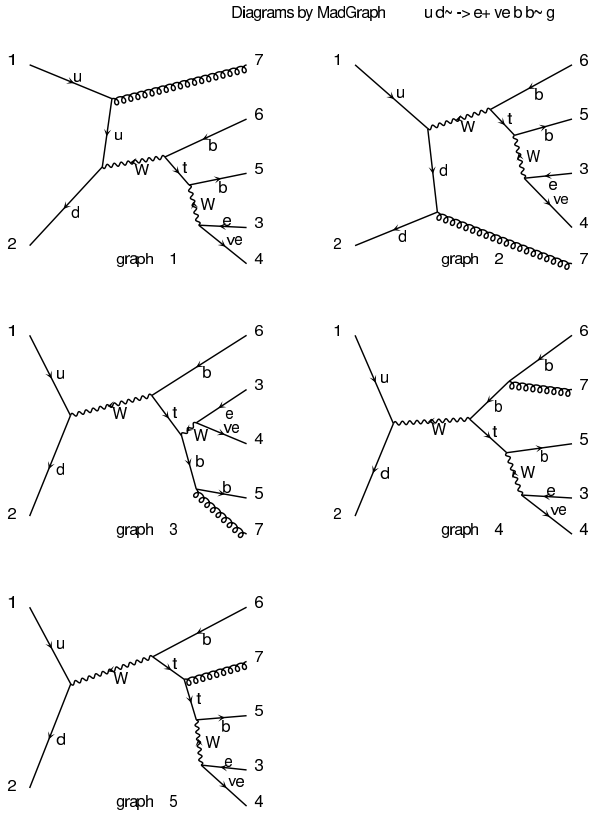


Figure 7: Leading order Feynman diagram for s-channel single top quark production and decay used in the calculation of the s-channel event probability in the three-jet bin.

A different transfer function is obtained for cone 0.4 and cone 0.7 jets in different $|\eta|$ regions defined as: 0-0.9, 0.9-1.2, and 1.2-2.8.

2.4.2 Parameters for $W_{jet}(E_{parton}, E_{jet})$

We determine the parameters of the transfer function $W_{jet}(E_{parton}, E_{jet})$ for the b-jet case using the s-channel Monte Carlo sample. For light jets we use the t-channel sample. These transfer functions are also applied to the background probabilities with b or light-jets. For the winter 2008 analysis we implemented transfer functions obtained

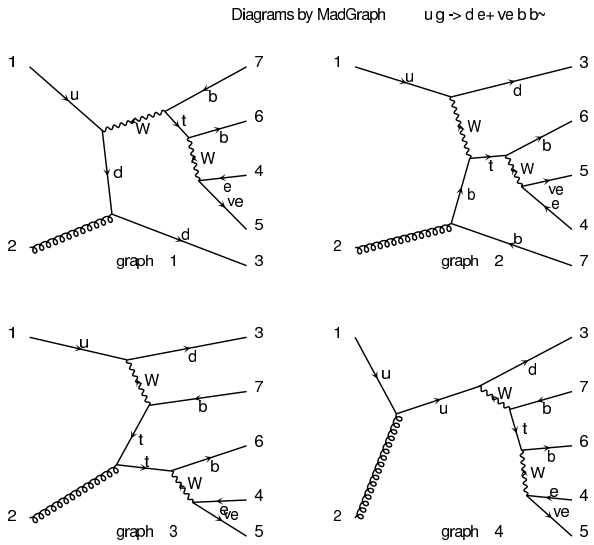


Figure 8: Leading order Feynman diagram for t-channel single top quark production and decay used in the calculation of the t-channel event probability in the three-jet bin.

from gluons which are later applied to all the probabilities with gluons in the final state. (This transfer functions were obtained only in one region of η and from the Wc process.) We apply all event selection requirements of the single top analysis and match reconstructed jets to the quarks. The quark is required to be aligned within a cone of $\Delta R < 0.4$ around the reconstructed jet-axis in order to be considered 'matched'. We correct the matched jet up to level 5 corrections and write out the jet energy together with the energy of the original particle. We do this for cases where we have exactly 2 jets of cone 0.7 size. In cases where there are less or more than 2 jets of cone 0.7 we use the cone 0.4 jets energies matched to the quark. As an example, the

distribution of measured jet energies for cone 0.4 versus the original parton energy of the b quark is shown in the upper plot of Fig. 12. The parameters of the transfer function are then derived by performing a maximum likelihood fit to these events. If $\mathrm{n}(E_{jet}, E_{parton})dE_{jet}dE_{parton}$ is the number of jets with jet energies between E_{jet} and $E_{jet}+dE_{jet}$, and particle energies between E_{parton} and $E_{parton}+dE_{parton}$ in this sample, then:

$$n(E_{jet}, E_{parton})dE_{jet}dE_{parton} = n(E_{parton})dE_{parton}W_{jet}(E_{parton}, E_{jet})dE_{jet} \qquad (7)$$

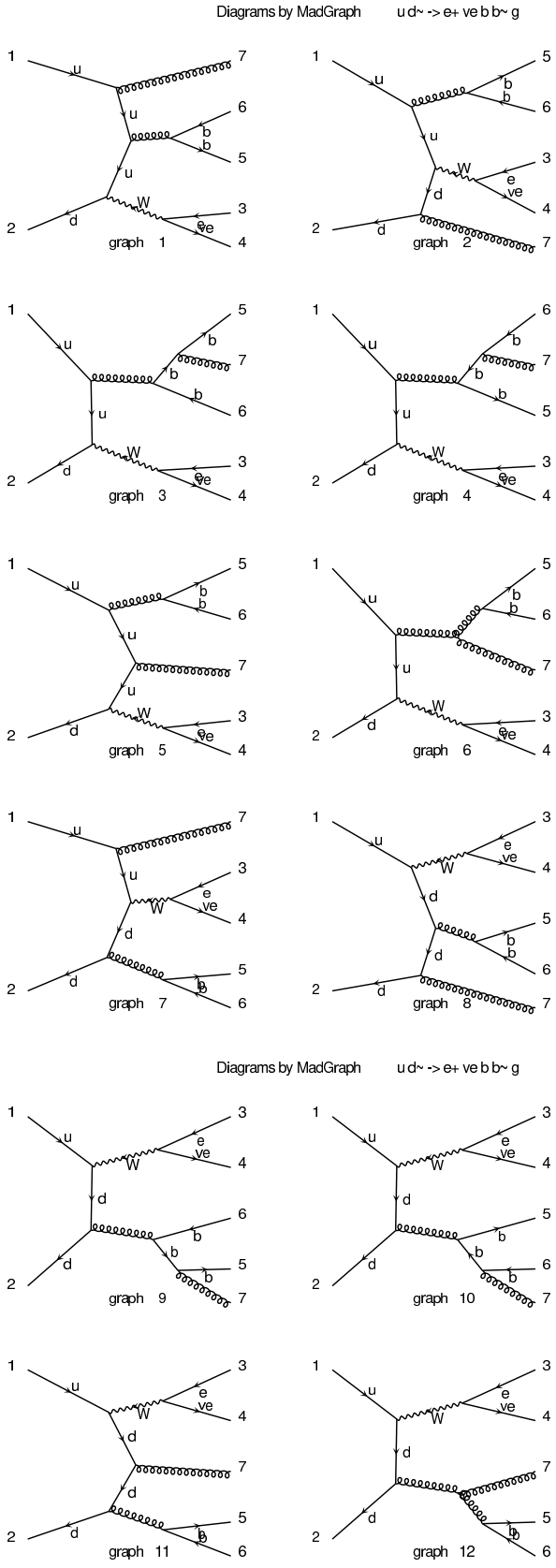
where $n(E_{parton})dE_{parton}$ is the number of particles with an energy between E_{parton} and $E_{parton} + dE_{parton}$. The parameters of $W_{jet}(E_{parton}, E_{jet})$ are determined such to maximize the agreement in Equation 7.

The parametrization for cone 0.4 is shown in the bottom plot of Fig. 12. The upper plot shows the two dimensional distribution of E_{parton} vs E_{jet} obtained from the Monte Carlo sample, while the bottom plot shows the prediction of this shape in a functional contour plot using the transfer function and the particle density $n(E_{parton})$. The performance of the parametrization is best shown in Fig. 13 and 14, which shows the $\delta E = (E_{parton} - E_{jet})$ distribution (histogram) compared to the prediction from the transfer function (solid line) for the central η region for cone 0.4 and 0.7, respectively.

We do similar tests in different η regions and for light jets and find similar agreement between Monte Carlo prediction and the parametrization.

Finally, we compare the previous transfer function with this new derived one. We build a discriminant (see later sections) and compare the efficiency for background and signal using 2000 events s-channel and $Wb\bar{b}$. Figure 15 shows an improvement of this new transfer function over the previous used one.

In a similar way we test the new transfer functions for gluons. Figure16 shows a good agreement between our prediction and the Monte Carlo. Figure 17 shows an improvement with respect of the previous discrimination by using this transfer function in the construction of the Wc probability.


Diagrams by MadGraph u d~ -> e+ ve b b~ g
graph 1
graph 2
graph 3
graph 4
graph 5
graph 6
graph 7
graph 8
Diagrams by MadGraph u d~ -> e+ ve b b~ g
graph 9
graph 10
graph 11
graph 12

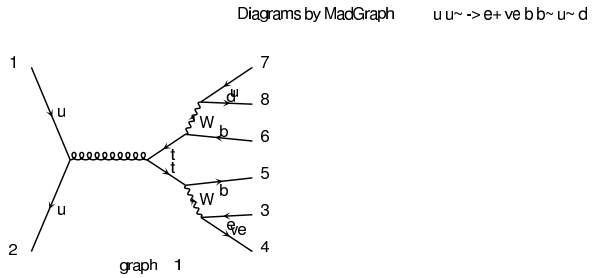


Figure 10: Leading order Feynman diagrams for $t\bar{t}$ production and decay used in the calculation of the $t\bar{t}$ event probability in the three-jet bin.

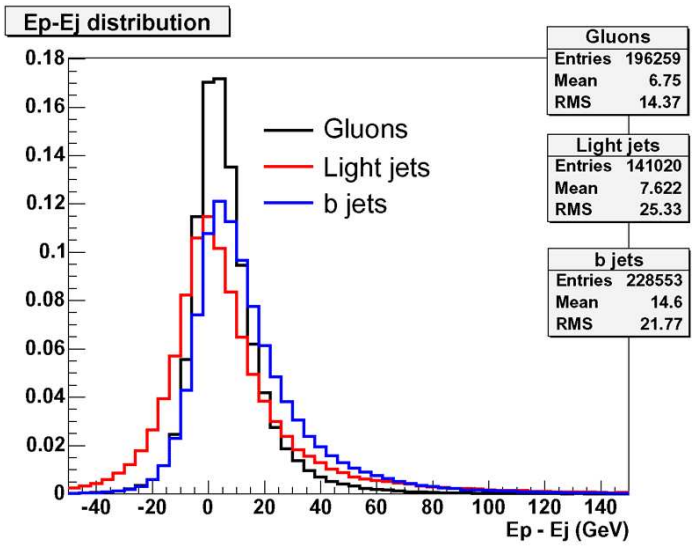


Figure 11: Distribution of $\delta_E = (E_{parton} - E_{jet})$ for matched jets to partons in s-channel (b-jets), t-channel (light-jets), Wc (gluon-jets) Monte Carlo events (passed through full detector simulation).

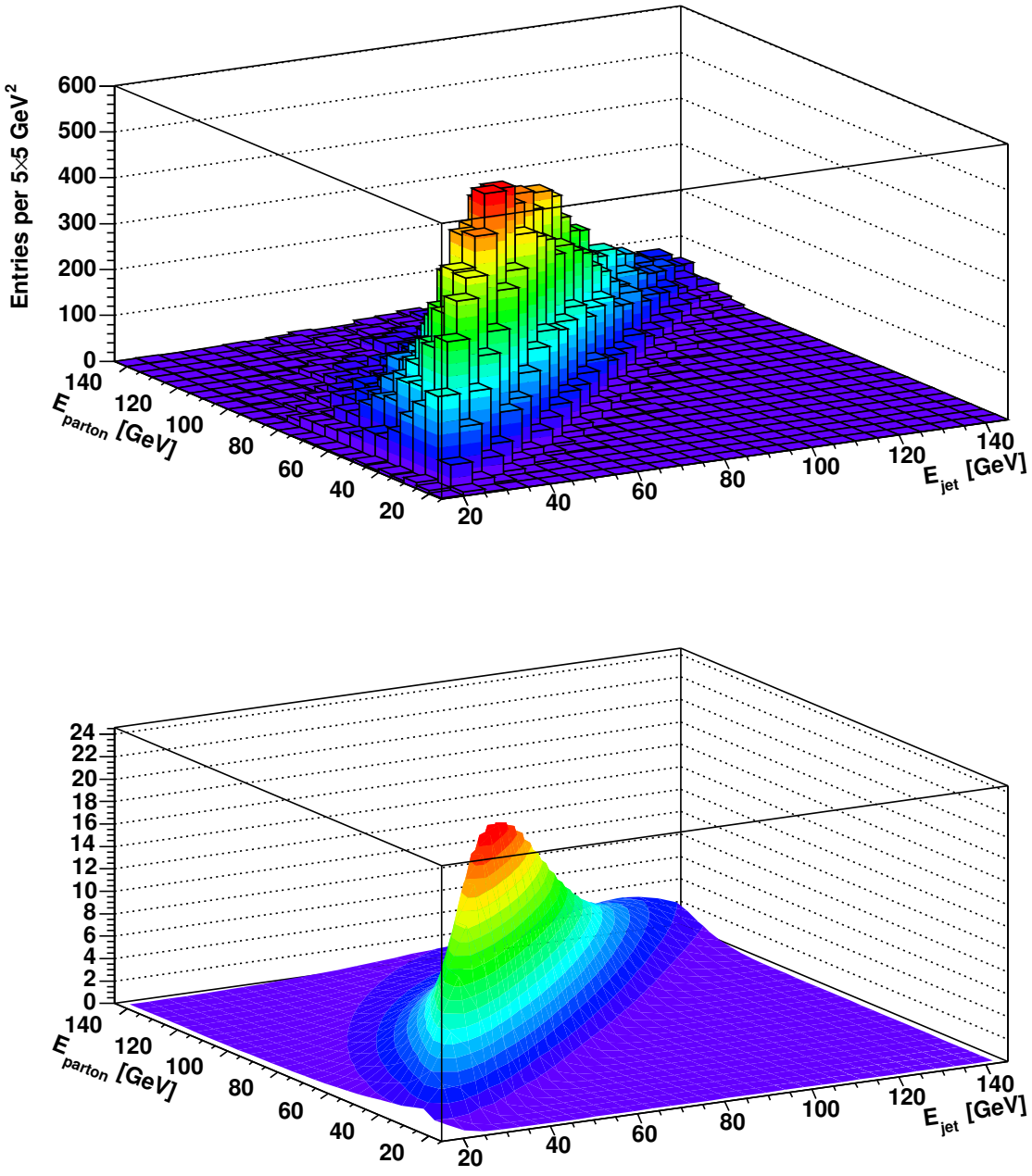


Figure 12: TOP: Lego plot of E_{parton} vs E_{jet} (passed through full GEANT detector simulation) for a sample of matched jets of cone 0.4 to partons in s-channel single top Monte Carlo events. BOTTOM: Functional form of E_{parton} vs E_{jet}, where E_{jet} is predicted using the transfer function $W_{jet}(E_{parton}, E_{jet})$ of Equation 6 and the particle density $n(E_{parton})$.

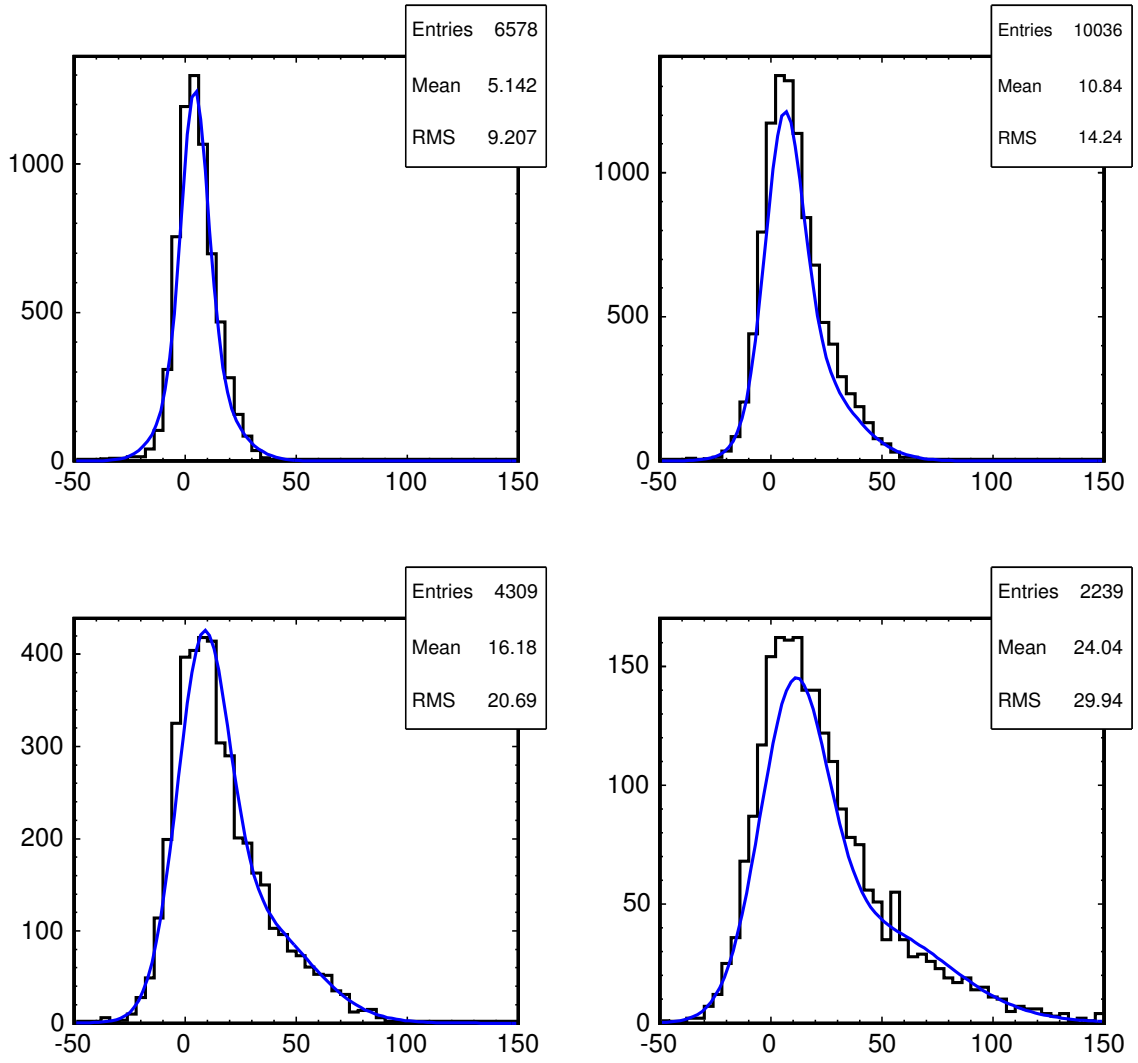


Figure 13: Distributions of $\delta E = (E_{parton} - E_{jet})$ for different ranges of parton energy of matched jets to partons (20-60, 60-100, 100-140, 140-240 GeV). The histograms are (s-channel) single top Monte Carlo events after full detector simulation and cone 0.4 jet (level 5) corrections. The solid line corresponds to the transfer function using the parameters of cone 0.4.

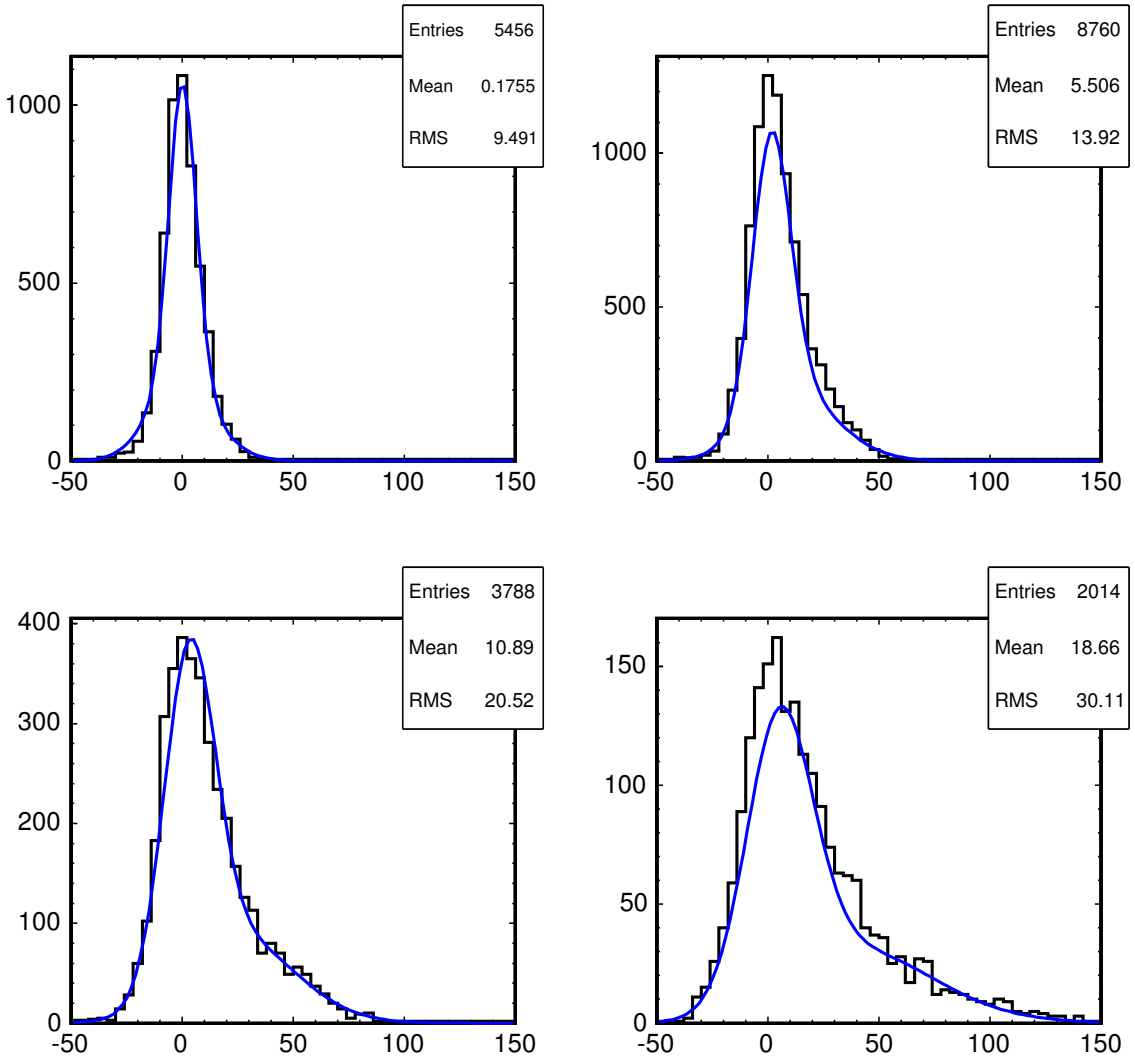


Figure 14: Distributions of $\delta E = (E_{parton} - E_{jet})$ for different ranges of parton energy of matched jets to partons (20-60, 60-100, 100-140, 140-240 GeV). The histograms are (s-channel) single top Monte Carlo events after full detector simulation and cone 0.7 jet (level 5) corrections. The solid line corresponds to the transfer function using the parameters of cone 0.7.

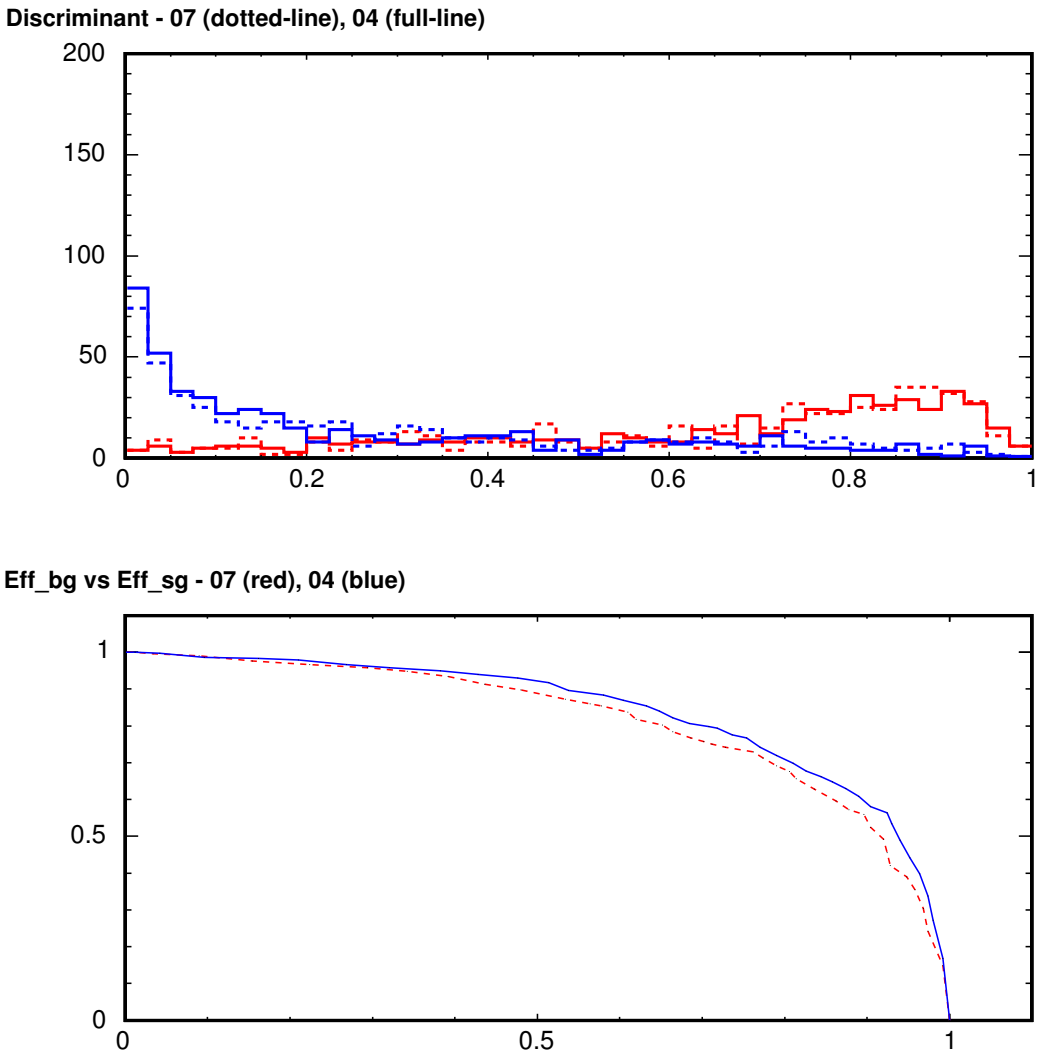


Figure 15: Comparison of the discriminant between previous and new TF.

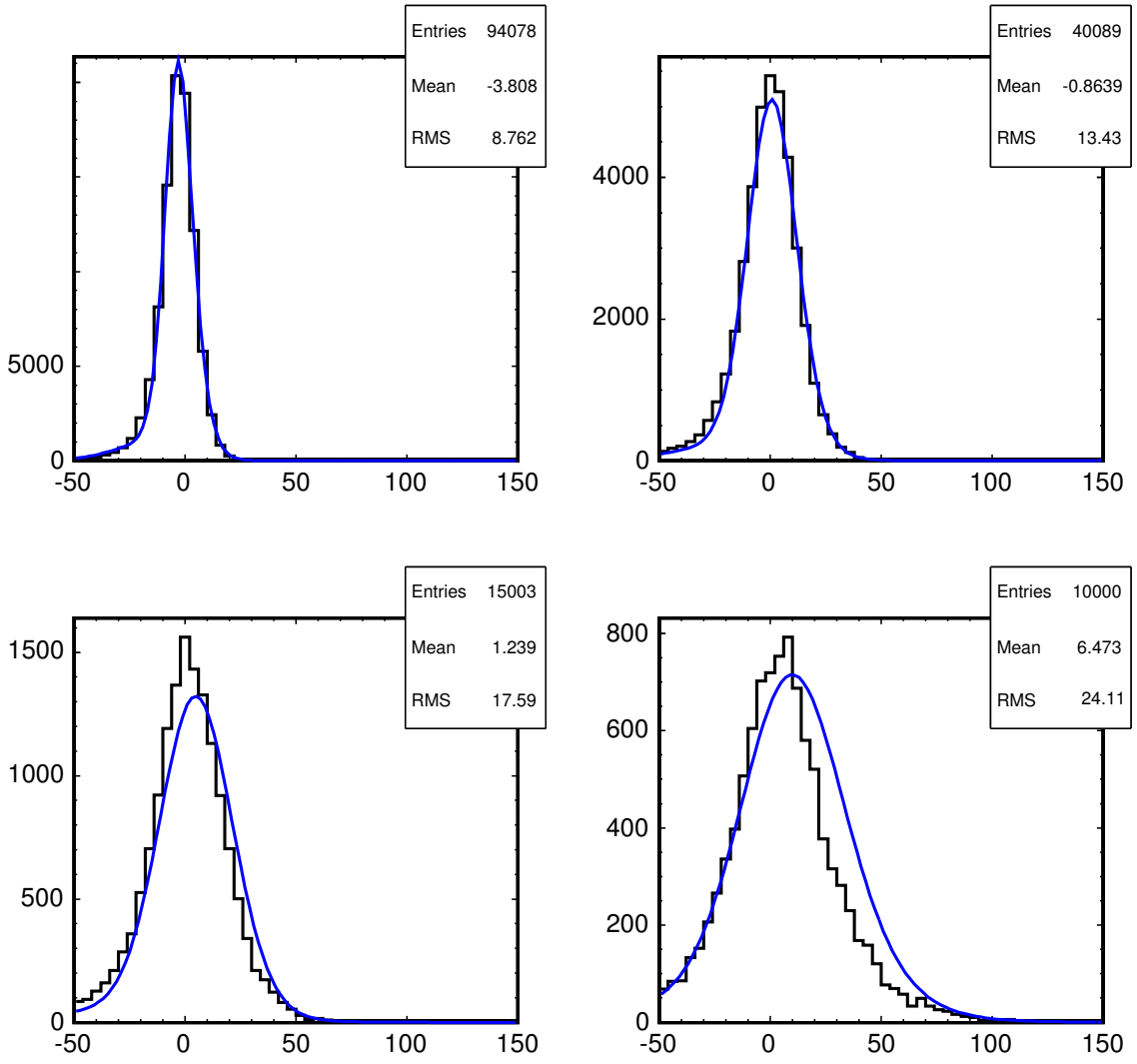


Figure 16: Distributions of $\delta E = (E_{parton} - E_{jet})$ for different ranges of parton energy of matched jets to gluons (20-60, 60-100, 100-140, 140-240 GeV). The histograms are Wc Monte Carlo events after full detector simulation and cone 0.7 jet (level 5) corrections. The solid line corresponds to the transfer function using the parameters of cone 0.7.

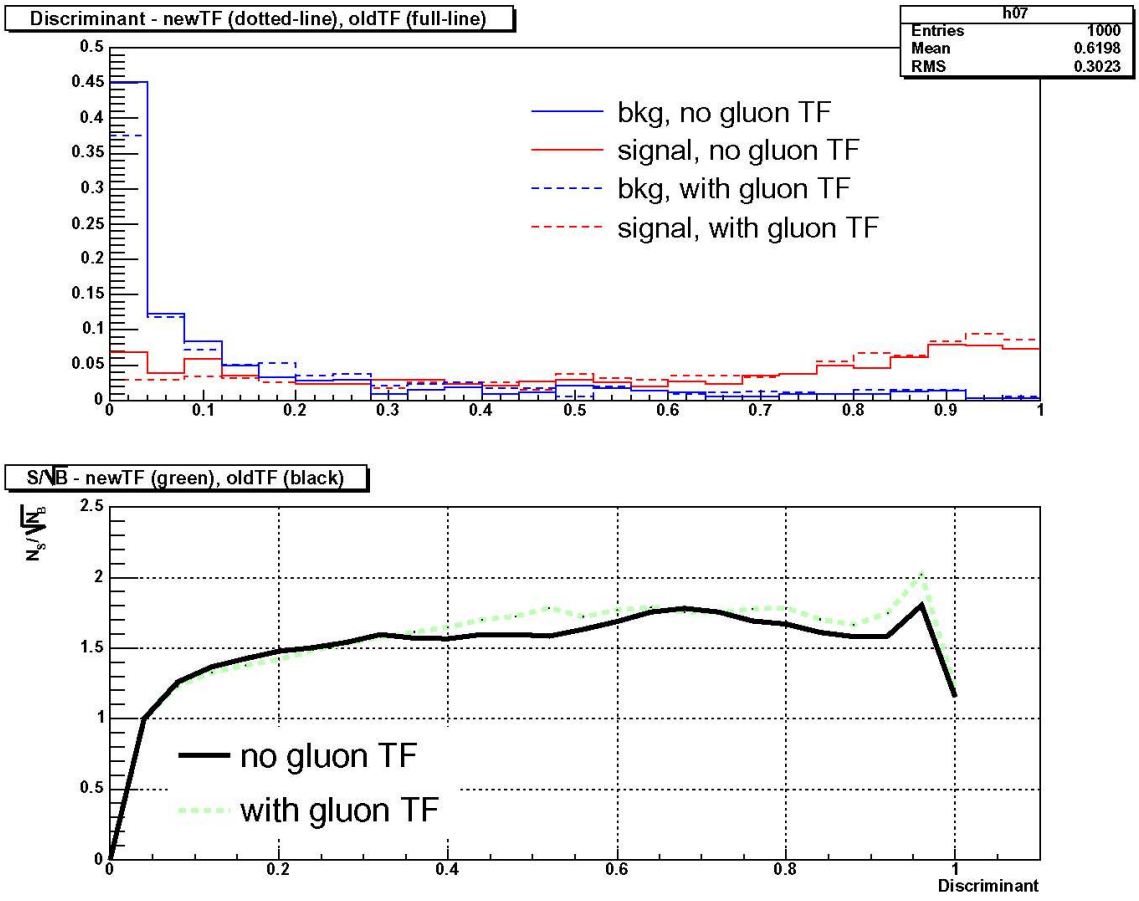


Figure 17: Comparison of the discriminant between previous and new TF.

2.5 Phase Space

The integration of the differential cross section has to be performed over 14 variables corresponding to the momentum vectors of the four final state particles (12 variables) and the longitudinal momenta of the initial state partons (2 variables). There are 11 δ-functions inside the integrals: four for total energy and momentum conservation (part of the phase space factor, see Equation 2) and seven for the transfer function (three for the lepton momentum vector and four for the jet angles, see Equation 5). The calculation of the event probability therefore involves a three dimensional integration. The integration is performed numerically over the absolute value of the quark momenta ($\rho_i = |\vec{p}_i|$) and the longitudinal momentum of the neutrino ($p_{\nu,z}$). The phase space for single top events is derived in Appendix A and has been expressed as a function of the variables $(\rho_1, \Omega_1, \rho_2, \Omega_2, \vec{p}_l, m_W^2)$. The result is:

$$\begin{aligned} d\Phi_4 &= \delta(E_{q_1} + E_{q_2} - \sum_{i=1}^{4} E_i)\ \delta(p_{q_1,z} + p_{q_2,z} - \sum_{i=1}^{4} p_{i,z}) \\ &\quad \times \frac{dm_W^2}{|2E_l \frac{p_{\nu,z}}{E_\nu} - 2p_{\nu,z}|} \frac{d^3\vec{p}_l}{(2\pi)^3 2E_l} \frac{1}{(2\pi)^3 2E_\nu} \prod_{i=1}^{2} \frac{\rho_i^2 d\rho_i d\Omega_i}{(2\pi)^3 2E_i} \end{aligned} \tag{8}$$

2.6 Event Probability Densities

Substituting the phase space factor (Equation 8) and the transfer function (Equation 5) into the expression for the event probability (Equation 4), we obtain:

$$\begin{aligned} P(x) = \frac{1}{\sigma} \int 2\pi^4 |M|^2 \frac{f(x_1)}{|E_{q_1}|} \frac{f(x_2)}{|E_{q_2}|} \delta^3(\vec{p}_l{}^{\,y} - \vec{p}_l{}^{\,x}) \prod_{i=1}^{2} \delta^2(\Omega_i^y - \Omega_i^x) \prod_{j=1}^{2} W_{jet}(E_{jet}, E_{parton}) \Phi_4 \\ \times \delta(E_{q_1} + E_{q_2} - \sum_{i=1}^{4} E_i) \delta(p_{q_1,z} + p_{q_2,z} - \sum_{i=1}^{4} p_{i,z}) d^3\vec{p}_l dm_W^2 dE_{q_1} dE_{q_2} \prod_{i=1}^{2} d\rho_i d\Omega_i \end{aligned} \tag{9}$$

where,

$$\Phi_4 = \frac{2\pi^4}{(2\pi)^9} \frac{1}{2E_l} \frac{1}{2E_\nu} \frac{\rho_1^2}{2E_1} \frac{\rho_2^2}{2E_2} \frac{1}{|2E_l \frac{p_{\nu,z}}{E_\nu} - 2p_{l,z}|}$$

The integration over E_{q_1} and E_{q_2} eliminates the two δ-functions in the second line of Equation 9. The integration over the lepton momenta and the quark solid angles eliminate the δ-functions in the first line of Equation 9 associated with $W(y, x)$. The final event probability takes the form:

$$P(x) = \frac{1}{\sigma} \int d\rho_1 d\rho_2 dm_W^2 \sum_{comb,\nu} |M|^2 \frac{f(q_1)}{|q_1|} \frac{f(q_2)}{|q_2|} \Phi_4 \prod_{i=1}^{2} W_{jet}(E_{parton,i}, E_{jet,i}) \tag{10}$$

For events with two b-jets (like s-channel single top), we don't know which jet came from the top quark decay so we have to calculate the probability for both possible jet-parton assignments. The total probability is summed over both combinations. We also try both combinations when there are two tags.

For three-jet events, we have the same basic configuration with one more jet. For s-channel, t-channel, and $Wb\bar{b}$ diagrams this is a radiated gluon. This requires us to perform four integrals instead of three. We assume that the tag information is correct and try all combinations consistent with the tag.

The $t\bar{t}$ matrix element is a special case, because its final state is not the same as single top and the events only filter in when final state particles go undetected. For the two-jet case, we use a diagram in which one final-state W boson is undetected while the other decays leptonically. Then we integrate over all three components of the momentum of the missing particle. In the three-jet case, we assume a lepton+jets diagram wherein one light jet is undetected, and we integrate over the momentum of the missing jet. This requires six integrals in the two-jet case and seven in the three-jet case.

2.7 Numerical Integration

We perform the three-dimensional integrations with a C++ version of the CERN library function DADMUL[13]. This algorithm is a fully deterministic adaptive quadrature technique that works well for up to three integrations. However, it becomes prohibitively slow for more than three integrals, so we moved to a Monte Carlo integration technique. We use the CERN library algorithm DIVONNE as implemented and improved in the CUBA integration package[14]. This is a Monte-Carlo-based integration based on stratified sampling and aided by methods from numberical optimization. It gives consistent answers with adaptive quadrature for three-integral matrix elements and gives answers that are self-consistent to at least five significant figures.

We ask each algorithm to converge to 1% estimated error in ten million function calls. Calculating all probabilities (seven for the two-jet bin and four for the three-jet bin) takes from five to ten minutes per event, most of which time is used by the $t\bar{t}$ matrix element calculation.

2.8 Event Probability Discriminant

The event probability density makes use of all measured quantities[1] to specify each event. This should provide good discrimination between signal and background. It uses both possible jet combinations in the event so that the right jet-parton association is always included. The expression for the event probability of Equation 10 can be applied to the s-channel and t-channel single top process as well as to the major backgrounds in the analysis, namely $Wb\bar{b}$, $Wc\bar{c}$ and Wcj. Only the matrix element has to be changed to represent the desired process.

We use the event probability densities as ingredients to build an event probability discriminant, i.e. a distribution which separates signal from background which we can use to fit the data. Figure 18 shows three different possible discriminants. The first

[1]The only exception is the unclustered energy, the energy deposited in the calorimeter from all sources except from leptons and jets.

of the three just plots the correlation of the s-channel probability vs $Wb\bar{b}$ background for s-channel and $Wb\bar{b}$ Monte Carlo events. The s-channel single top events (red) are separated fairly well[2] from the background (blue). Since the separation of signal and background is more enhanced along the $y = x$ line, a one dimensional projection of this separation is the ratio of signal and background probability, $EPD = P_s/P_b$ shown in the middle. Perhaps the most intuitive discriminant is the ratio of signal over signal + background probability $EPD = P_s/(P_s + P_b)$ (shown on the right). This discriminant is close to zero for ratios dominated by P_b and close to unity for ratios dominated by P_s. This is the discriminant we will use in this analysis.

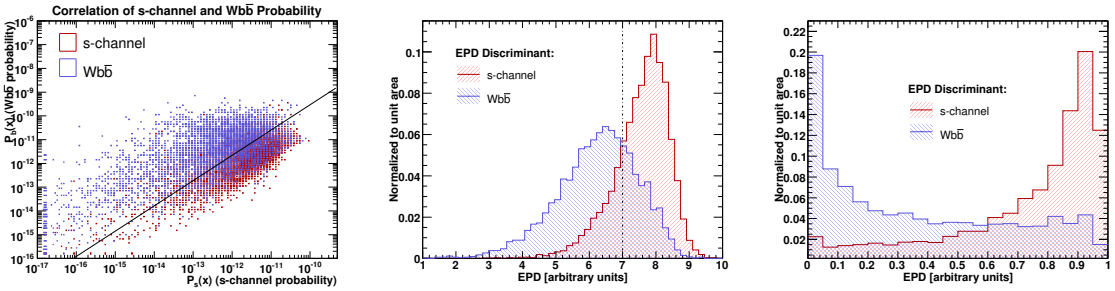


Figure 18: Several event-probability-discriminants (EPD). Correlation of the s-channel probability vs $Wb\bar{b}$ background (left); Ratio of signal and background probability (middle) and ratio of signal over signal + background probability (right).

2.8.1 Including Secondary Vertex Information

Several of the sizable backgrounds in the single top analysis listed in Table 1 don't actually have a b-quark in the final state, but are falsely identified as such. This happens either because a light quark jet is falsely identified to have a displaced secondary vertex (mistags) due to tracking resolution or because real heavy charm quark decays happen to have a sufficiently long life-time to be tagged by the secondary vertex tagger. Therefore, it would be useful to have better separation of b-quark jets from charm or light quark jets. The matrix element based event-probabilities use all event kinematic information to characterize signal and background but can not distinguish b quark jets from charm or light quark flavor jets. On the other hand, the Karlsruhe Neural Net b-tagger uses secondary vertex information, like the secondary vertex mass, the number of tracks, the decay length of the b quark and 22 more variables to distinguish b-quark jets from charm or light quark flavor jets. [9].

Figure 19 shows the neural network b-tagger output for b jets and non-b jets background (left). Both distributions are very well separated. In this analysis we can include the neural network b-jet information by using the network output as a b-jet probability in our discriminant. This is motivated by the right plot of Fig. 19

[2]The s-channel separation from background is in fact harder than for the t-channel due to less kinematically distinct features.

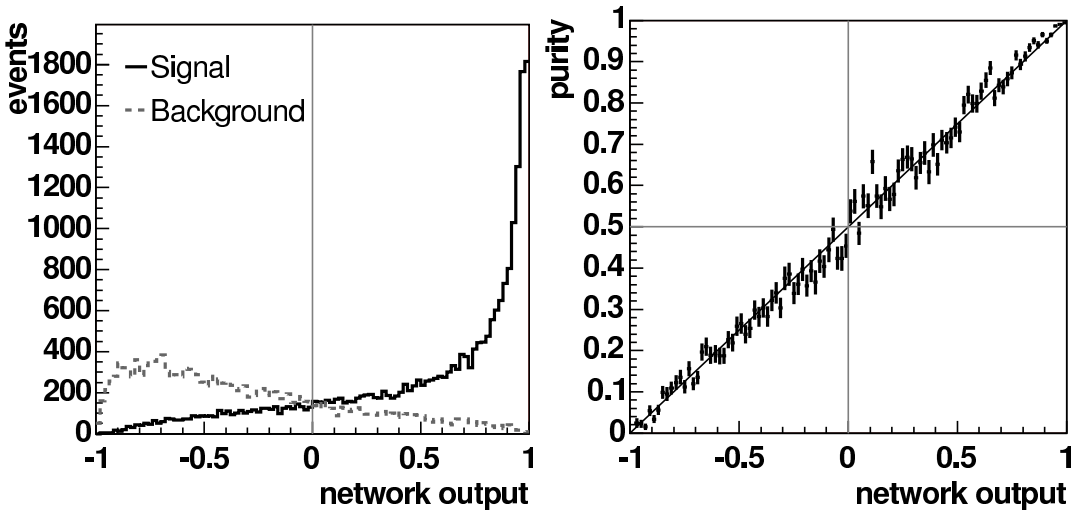


Figure 19: Neural network b-tagger output for b jets and non b jets background (left). Neural Network b-tag output vs purity (right) [9].

which shows the network output versus b-jet purity. We define the b-jet probability as $b = 0.5 \cdot (network\ output + 1)$. Since single top always has at least one b quark in the final state we re-write the event probability discriminant as:

$$EPD = \frac{b \cdot P_{singletop}}{b\,(P_{singletop} + P_{Wb\bar{b}} + P_{t\bar{t}}) + (1-b)\,(P_{Wc\bar{c}} + P_{Wcj} + P_{Mistag})} \tag{11}$$

In addition, each probability can receive a normalization constant to adjust the relative weight for that probability. We tune these constants in Monte Carlo to achieve the greatest expected sensitivity, while making sure every bin is sufficiently populated in Monte Carlo statistics. We also create separate discriminants for the single-tag and double-tag bins. This gives us the ability to tune the discriminants differently in the two tag bins: t-channel dominates in the single-tag bin, but s-channel is much greater in the double-tag bin. In the double-tag bin, we multiply the b-jet probabilities of both jets as a weight.

There is a small additional gain from weighting the two possible t-channel diagrams separately. One case has an u quark and a b quark in the initial state while the other has a $\bar{u}$ quark and a b. The matrix elements are identical, but the PDFs are different, so we gain sensitivity from treating them separately. We also gain some sensitivity from the Wg + jet matrix element for mistags, which is identical to the Wc + jet matrix element with different PDFs, transfer functions, and quark masses. The weights we use are shown in Table 2.

2.8.2 Distributions of the Event Probability Discriminant

In principle, we could define separate discriminants for s-channel and t-channel and search for them separately. In this note, however, we will focus on the combined search

	2 jet		3 jet	
	Single-tag	Double-tag	Single-tag	Double-tag
s-channel	3	5	5	4
t-channel (u quark)	0.023	0.02	4	3.96
t-channel ($\bar{d}$ quark)	0.605	0.04	0.044	0.04
$Wb\bar{b}$	1.513	0.917	1.1	0.5
$Wc\bar{c}$	0.045	0.0033	1	0.5
Wc+jet	0.25	0.017	—	—
Wg+jet (Mistag)	0.005	0.0033	—	—
Wgg (Mistag)	0.005	0.0067	—	—
$t\bar{t}$	0.0001	0.00003	0.0045	0.0045

Table 2: Coefficients used to weight the various event probability densities in the discriminant. Mistag and Wc+jet diagrams are not calculated in the three-jet discriminant because they are less significant backgrounds.

only, which uses only the two discriminants described above. As can be seen from Figure 20 the separation between signal and background is better for the t-channel in the single-tag case but better for the s-channel in the double-tag case. Separation is better in the two-jet bin than the three-jet bin.

2.9 The Likelihood Function

The likelihood function, $\mathcal{L}$, is a function of the unknown Poisson means for signal and background and is defined such that it expresses the joint probability of observing the N data events at their respective values of the event probability discriminant. The values of the Poisson means at which $\mathcal{L}$ achieves its maximum, corresponds to the most probable estimate for the true signal and background content in the data sample.

We perform a binned likelihood fit to the event probability discriminant. We choose five free parameters in the fit. The single top, W+*bottom* jets (i.e. $Wb\bar{b}$,WZ,ZZ, Z+jets, non-W), W+*charm* jets (i.e. $Wc\bar{c}$, Wcj, WW,), mistags and $t\bar{t}$ normalization (i.e. cross-sections). We group the templates into W+*bottom* jets , W+*charm* jets and mistags in the likelihood fit since the contributing histograms look similar and they are expected to behave similar under variations of the NN b-tagger used. The histograms which contribute to the single top, W+*bottom*, W+*charm* and Mistag templates are shown in Figure 25.

To make it easier to compare the different fit parameters, we define the fit parameter as $\beta_j = \sigma_j^{Fit}/\sigma_j^{SM}$ where β_j is unity when the fit result corresponds to the expected number of events obtained from the independent *Method 2* signal/background estimate:

$$\mathcal{L} = \prod_{j=2}^{5} G_j(\beta_j;\sigma_j) \prod_{k=1}^{B} \frac{e^{-\mu_k} \cdot \mu_k^{n_k}}{n_k!} \tag{12}$$

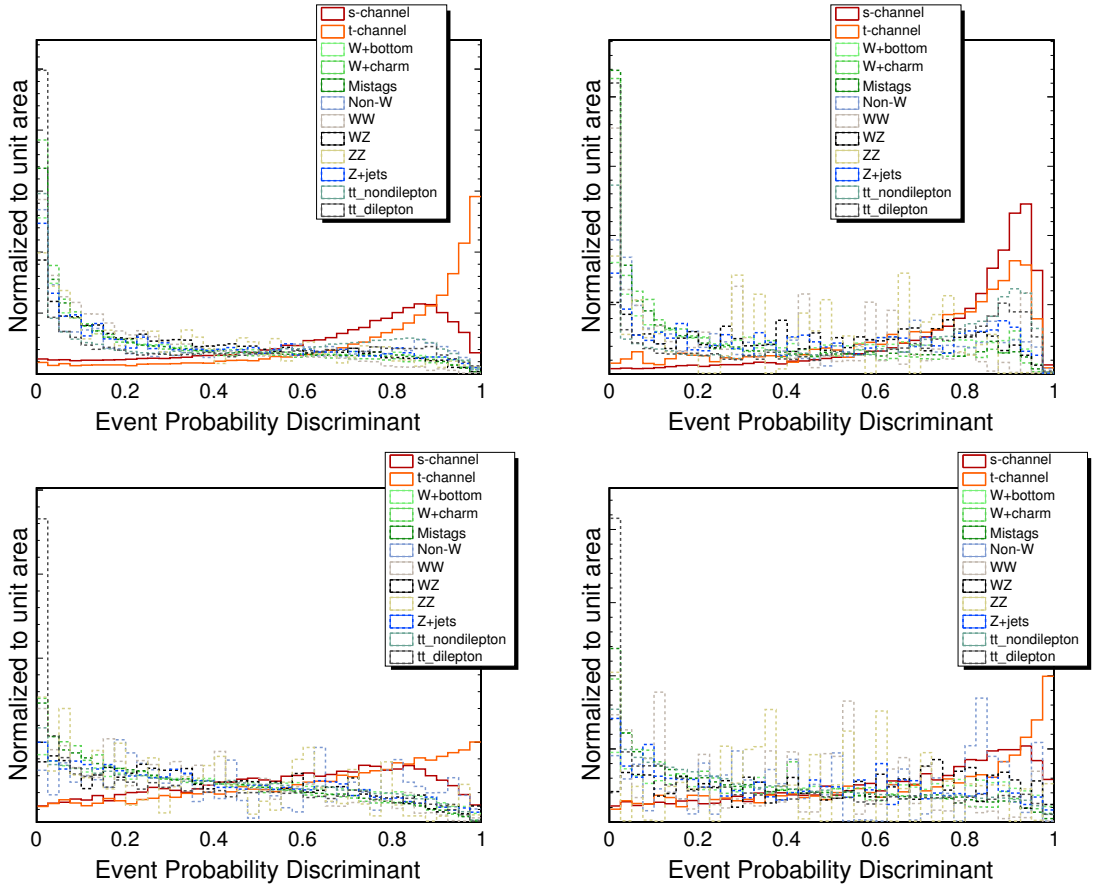


Figure 20: Discriminants for all signal and background processes. The top two plots show the two-jet bin while the bottom plots show the three-jet bin. Single-tag discriminants are on the left side, while double-tag discriminants are on the right. All histograms are normalized to unit area.

The Gaussian constraints to the backgrounds are given by:

$$G_j(\beta_j;\sigma_j) = \frac{1}{\sqrt{2\,\pi\cdot\sigma_j^2}}\exp\left[-\frac{1}{2}\cdot\left(\frac{\beta_j-1.0}{\sigma_j}\right)^2\right] \tag{13}$$

$$\mu_k = \beta_{\text{single top}}\cdot T_{jk}+\beta_{\text{W+bottom jets}}\cdot T_{jk}+\beta_{\text{W+charm jets}}\cdot T_{jk}+\beta_{\text{mistags}}\cdot T_{jk}+\beta_{\text{t}\bar{\text{t}}}\cdot T_{jk} \tag{14}$$

The index k runs over the bins of the fitted histogram. The template histograms are normalized to the predicted number of events as shown in Table 3. This means, $\sum_{k=1}^{B} T_j k = N_j^{pred}$. Figure 21 shows the templates used in the likelihood.

In addition, the prediction in each bin needs an additional Gaussian uncertainty due to the limitations of Monte Carlo statistics. Each bin is allowed to fluctuate according to the total uncertainty in that bin, which is the sum in quadrature of the weight of

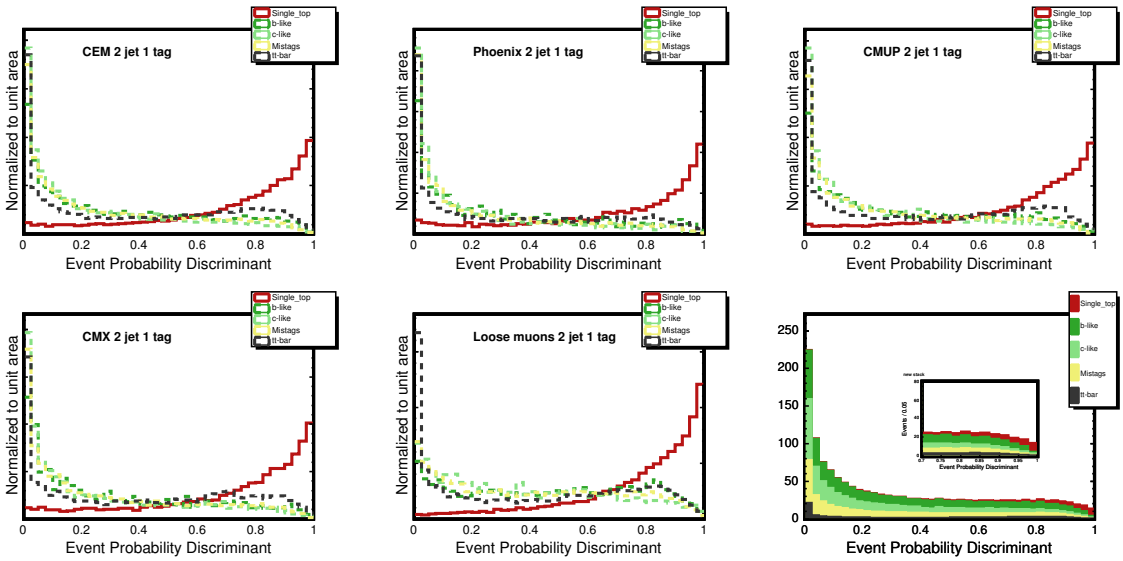


Figure 21: Templates of EPD for each dataset used in the likelihood function. Each template is composed of processes combined and weighted according to the Standard Model prediction. These templates are for the two-jet single-tag bin.

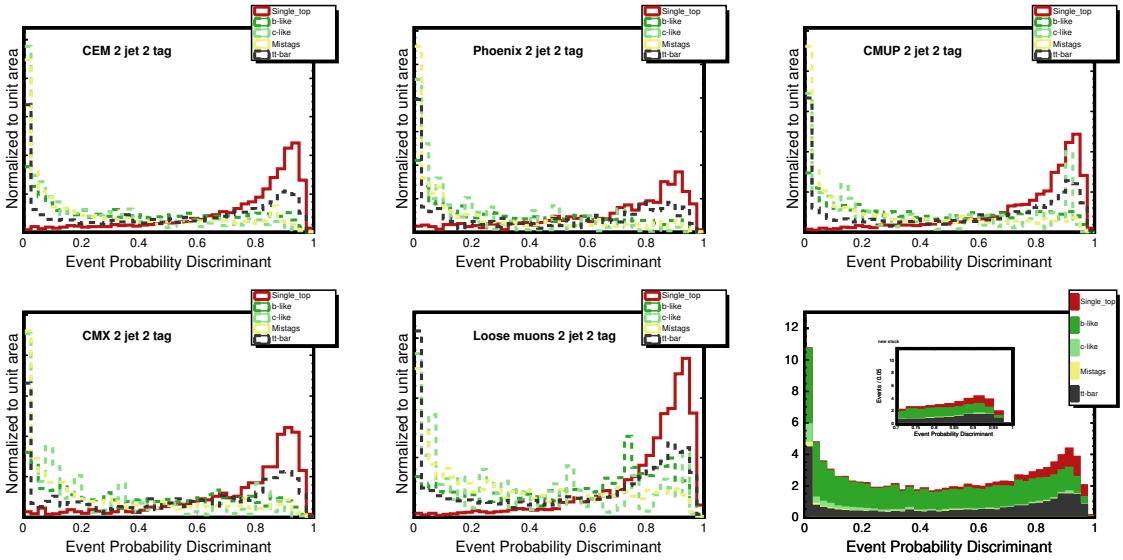


Figure 22: EPD templates for the two-jet double-tag bin.

each event. This prevents us from overestimating our sensitivity due to a fluctuation in Monte Carlo.

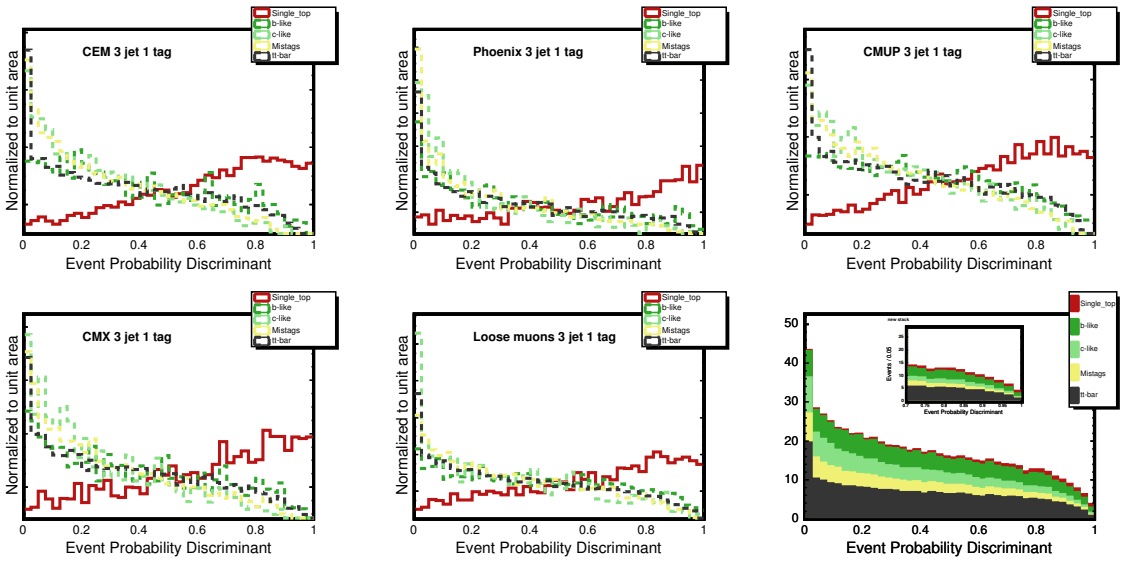


Figure 23: EPD templates for the three-jet single-tag bin.

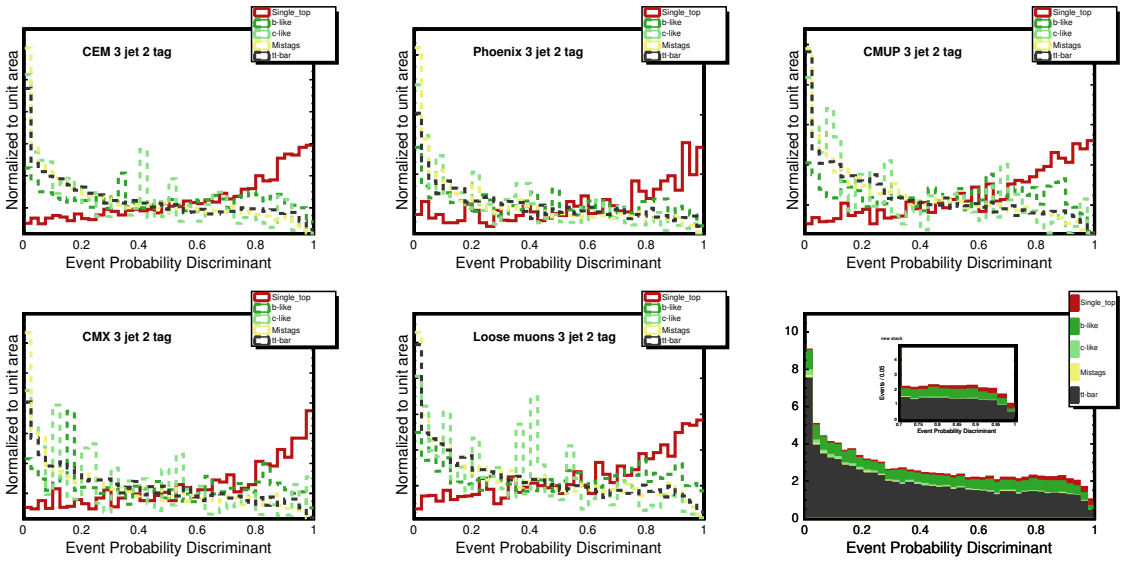


Figure 24: EPD templates for the three-jet double-tag bin.

3 Incorporating Systematic Uncertainties

Systematic uncertainties can bias the outcome of this analysis and have to be incorporated into the result. We address systematic uncertainty from several different sources: (1) jet energy scale (2) initial state radiation (3) final state radiation (4) parton distribution functions (5) the event generator, the uncertainty in the event detection efficiency and luminosity (6) neural network jet flavor separator uncertainty, (7) ALPGEN Monte Carlo factorization and renormalization scale uncertainty (8) uncertainty

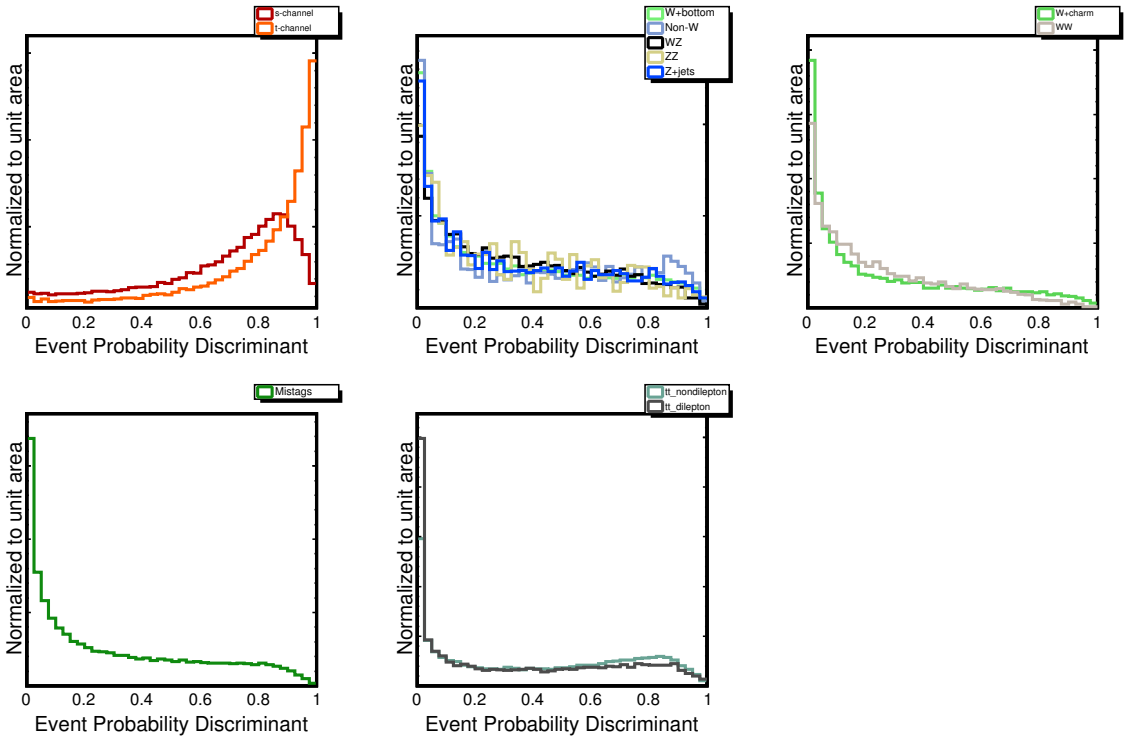


Figure 25: EPD distributions for processes that contribute to the single top (top left), b-like (top middle), c-like (top right), mistag (bottom left) and $t\bar{t}$ templates (bottom middle). All histograms are normalized to unit area. When they are combined to make the template histograms, the correct relative normalization is used. These histograms show the two-jet single-tag case.

(j)	Process	Number of events expected in 2.2/fb	Gaussian constraint Δ_j
(1)	$N^{pred}_{single\ top}$	135.7±19.7	N/A
(2)	$N^{pred}_{bottom-like}$	757.2±186.9	24.7%
(3)	$N^{pred}_{charm-like}$	569.2±157.6	27.7%
(4)	N^{pred}_{Mistag}	448.0±74.1	16.5%
(5)	$N^{pred}_{t\bar{t}}$	493.3±70.3	14.3%

Table 3: Number of expected single top and background events in 2.2fb^{-1} of CDF data used to normalize the likelihood fitter templates and Gaussian constraints.

on the mistag model, (9) uncertainty on the non-W flavor composition, and (10) uncertainty on the modeling of the Monte Carlo simulation, as estimated from the most discrepant shapes in the control variables.

Systematic uncertainties can influence both the expected event yield (normalization) and the shape of the discriminant distribution.

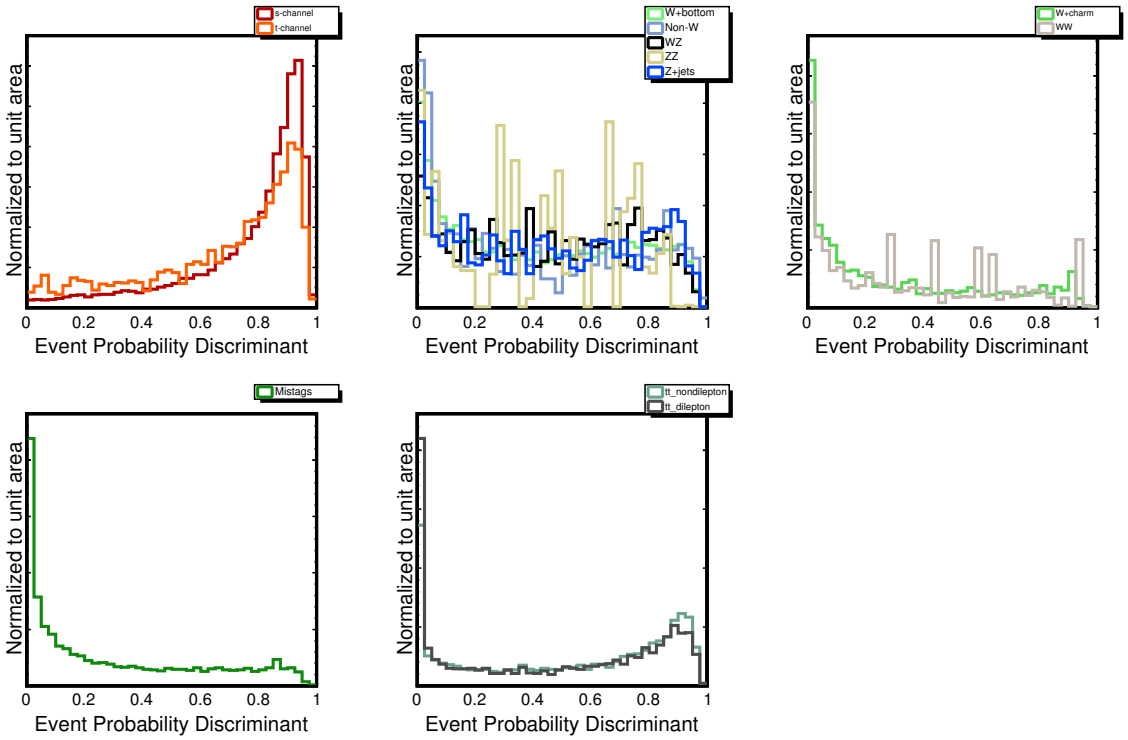


Figure 26: EPD distributions for the two-jet double-tag case.

Normalization uncertainties are estimated by recalculating the acceptance using Monte Carlo samples altered due to a specific systematic effect. The single top normalization uncertainty is the difference between the systematically shifted acceptance and the default one and are shown in Table 4.

The effect of the uncertainty in the jet energy scale is evaluated by applying jet-energy corrections that describe $\pm 1\sigma$ variations to the default correction factor. Systematic uncertainties due to the modeling of ISR and FSR are obtained from dedicated Monte Carlo samples where the strength of ISR/FSR was increased and decreased in the parton showering to represent $\pm 1\sigma$ variations [11]. To evaluate the uncertainty associated with the specific choice of parton distribution functions, we use the recommendation from the joint physics group and vary the 40 independent eigenvectors of the CTEQ parton distribution functions and compare to the MRST PDFs. We quadratically sum the uncertainty from the CTEQ and MRST PDF uncertainty if the difference between the CTEQ and MRST PDFs is larger than the CTEQ uncertainty.

The effect of event generator, event detection, b-tag scale factor, and luminosity uncertainty is determined from the background estimate (for the signal template only; the background templates have these numbers included in their Gaussian constraints). The neural network jet flavor separator has a systematically shifted outputs that we employ as systematic uncertainty. ALPGEN Monte Carlo systematic uncertainties in shape are estimated from dedicated samples—rate uncertainties are already included

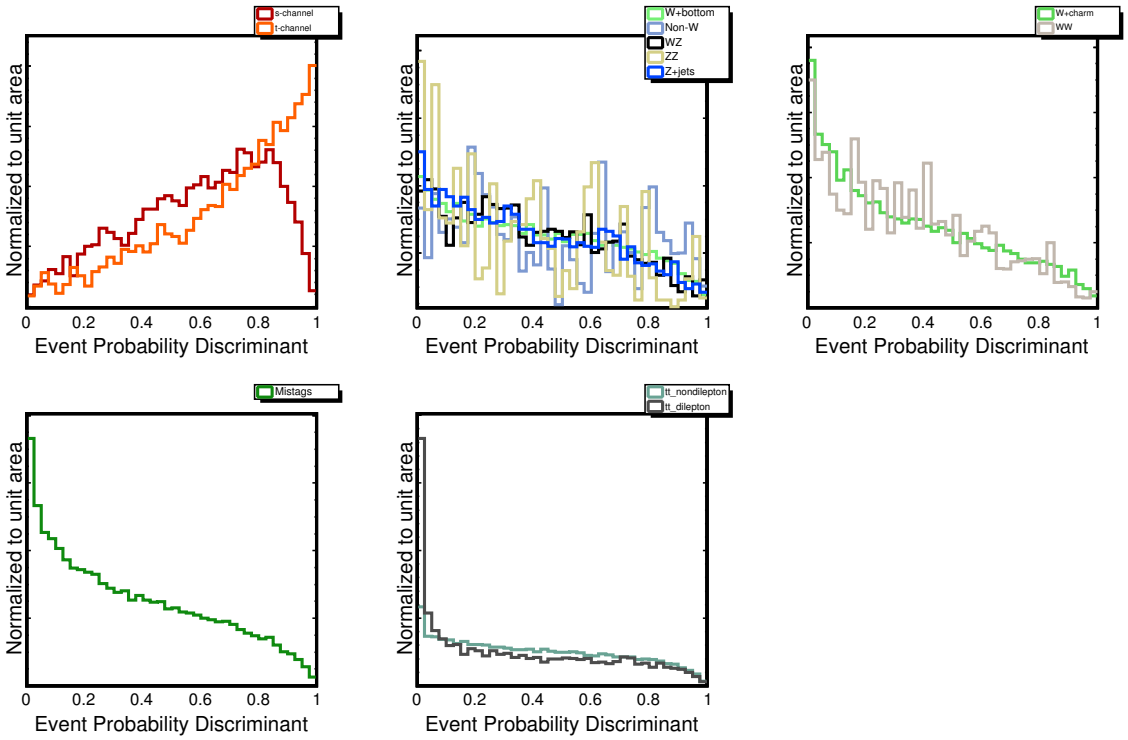


Figure 27: EPD distributions for the three-jet single-tag case.

in the background normalization uncertainties.

The uncertainty on the mistag model shape is determined by weighting pretag data with the mistag matrix [19]. The non-W flavor composition is determined from studies made using the neural net b-tagger to estimate the flavor composition of the non-W samples. The uncertainty is estimated by comparing the default flavor ratios (45% b, 40% c, 15% light) with an "extreme" estimate (60% b, 30% c, 10% light).

In order to account for possible mis-modeling in Monte Carlo, we assign a systematic to variables that look sufficiently discrepant to suspect a mis-modeling. We do this by re-weighting the templates using the distributions of that variable in the sideband of events which have at least one taggable jet but no tagged jets. Two variables (Figure 29 were chosen for this systematic: the pseudo-rapidity of the second jet, which shows an excess in data in the far forward region, asymmetrically on the east side; and the quantity ΔR between the two jets, in which the Monte Carlo comes close but does not match the data perfectly.

For all backgrounds the normalization uncertainties are represented by the uncertainty on the predicted number of background events, obtained from Method2 and are incorporated in the analysis as Gaussian constraints in the likelihood fit: $G(\beta_j|1,\Delta_j)$.

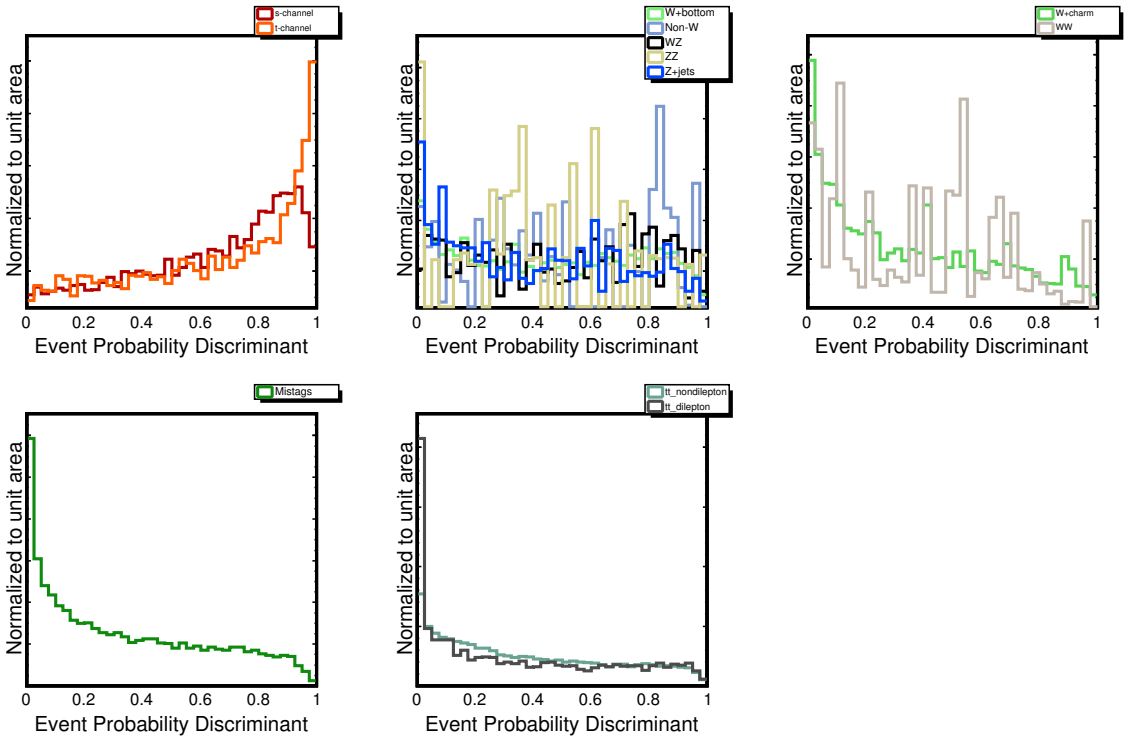


Figure 28: EPD distributions for the three-jet double-tag case.

The individual Δ_j were listed in Table 3.

$$\mathcal{L}(\beta_1,\ldots,\beta_5;\delta_1,\ldots,\delta_{11}) = \underbrace{\prod_{k=1}^{B}\frac{e^{-\mu_k}\cdot\mu_k^{n_k}}{n_k!}}_{Poisson\ term}\cdot\underbrace{\prod_{j=2}^{5}G(\beta_j|1,\Delta_j)}_{Gauss\ constraints}\cdot\underbrace{\prod_{i=1}^{11}G(\delta_i,0,1)}_{Systematics} \tag{15}$$

$$where,\ \mu_k = \sum_{j=1}^{5}\beta_j \cdot \underbrace{\left\{\prod_{i=1}^{11}\left[1+|\delta_i|\cdot(\epsilon_{ji+}H(\delta_i)+\epsilon_{ji-}H(-\delta_i))\right]\right\}}_{Normalization\ Uncertainty} \tag{16}$$

$$\underbrace{\cdot\,\alpha_{jk}}_{Shape\ P.} \cdot \underbrace{\left\{\prod_{i=1}^{11}\left(1+|\delta_i|\cdot(\kappa_{jik+}H(\delta_i)+\kappa_{jik-}H(-\delta_i))\right)\right\}}_{Shape\ Uncertainty} \tag{17}$$

All systematic normalization and shape uncertainties are incorporated in the analysis into the likelihood as nuisance parameters, conform with a fully Bayesian treatment [15]. We take the correlation between normalization and shape uncertainties for a given source into account [16]. The relative strength of a systematic effect due to the source i is parameterized by the nuisance parameter δ_i in the likelihood function, constrained to a unit-width Gaussian (last term in Equation 15). The $\pm1\sigma$ changes

in the normalization of process j due to the i^{th} source of systematic uncertainty are denoted by $\epsilon_{ji}+$ and $\epsilon_{ji}-$ (see Equation part 16). The $\pm 1\sigma$ changes in bin k of the H_T templates for process j due to the i^{th} source of systematic uncertainty are quantified by κ_{jik+} and κ_{jik-} (see Equation part 17). $H(\delta_i)$ represents the Heaviside function, defined as $H(\delta_i) = 1$ for $\delta_i > 0$ and $H(\delta_i) = 0$ for $\delta_i < 0$. The Heaviside function is used to separate positive and negative systematic shifts (for which we have different normalization and shape uncertainties). The variable δ_i appears in both the term for the normalization (Equation 16) and the shape uncertainty (Equation 17), which is how correlations between both effects are taken into account.

We marginalizing the likelihood function by integrating $\mathcal{L}(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \delta_1, ..\delta_{11})$ over the nuisance parameters $\beta_2, \beta_3, \beta_4, \beta_5, \delta_1, ..\delta_{11}$ for many possible values of the single top cross-section β_1 from [0..5]. The resulting reduced likelihood $\mathcal{L}(\beta_1)$ is a function of the single top cross-section β_1 only.

We list all systematic acceptance changes due to systematic uncertainties (aside from the uncertainty in the background estimate, which includes luminosity, generator, b-tagging, and heavy flavor uncertainties) in Table 4 and Table 5.

Systematic	Process	1 tag	2 tag
JES	Single top	-0.6% / $+0.3\%$	$+2.2\%$ / -1.1%
JES	b-like	$+6.0\%$ / -6.6%	$+9.7\%$ / -9.9%
JES	c-like	$+6.1\%$ / -5.6%	$+10.0\%$ / -10.0%
JES	$t\bar{t}$	$+9.9\%$ / -9.4%	$+8.5\%$ / -7.6%
ISR	Single top	$+1.9\%$ / $+2.1\%$	$+0.3\%$ / $+6.6\%$
ISR	$t\bar{t}$	-2.6% / -7.1%	$+0.6\%$ / -9.4%
FSR	Single top	$+4.8\%$ / -0.7%	$+7.5\%$ / $+0.8\%$
FSR	$t\bar{t}$	-5.1% / -2.6%	-8.0% / -1.7%
PDF	Single top	-3.0% / $+3.0\%$	-2.0% / $+2.0\%$
PDF	$t\bar{t}$	-1.8% / $+1.8\%$	-1.7% / $+1.7\%$
m_{top}	Single top	$+1.7\%$ / -1.1%	$+1.0\%$ / -0.6%
m_{top}	$t\bar{t}$	-3.0% / $+1.4\%$	$+0.5\%$ / $+3.0\%$

Table 4: Rate systematic uncertainties used in this analysis for the $W + 2$ jet bin . The m_{top} samples are used not as a systematic but as a different mass point for the measurement.

Figures 30 through 38 show the shifts in the distributions caused by shifts of one standard deviation in the different sources of uncertainty.

Systematic	Process	1 tag	2 tag
JES	Single top	-9.4% / $+10.3\%$	-8.3% / $+4.2\%$
JES	b-like	$+6.8\%$ / -6.3%	$+9.5\%$ / -10.0%
JES	c-like	$+4.7\%$ / -5.6%	$+15.4\%$ / -11.4%
JES	$t\bar{t}$	$+4.6\%$ / -5.1%	$+5.5\%$ / -5.1%
ISR	Single top	-3.3% / -4.8%	$+5.8\%$ / -5.0%
ISR	$t\bar{t}$	-0.6% / -4.5%	-0.5% / -6.6%
FSR	Single top	-3.3% / -3.8%	$+2.2\%$ / -2.3%
FSR	$t\bar{t}$	-3.4% / -2.2%	-3.4% / -2.7%
PDF	Single top	-2.6% / $+2.6\%$	-1.9% / $+1.9\%$
PDF	$t\bar{t}$	-1.8% / $+1.8\%$	-1.7% / $+1.7\%$
m_{top}	Single top	$+3.2\%$ / -2.7%	$+2.7\%$ / -2.6%
m_{top}	$t\bar{t}$	-0.7% / $+0.8\%$	-0.6% / -1.0%

Table 5: Rate systematic uncertainties used in this analysis for the $W + 3$ jet bin. The m_{top} samples are used not as a systematic but as a different mass point for the measurement.

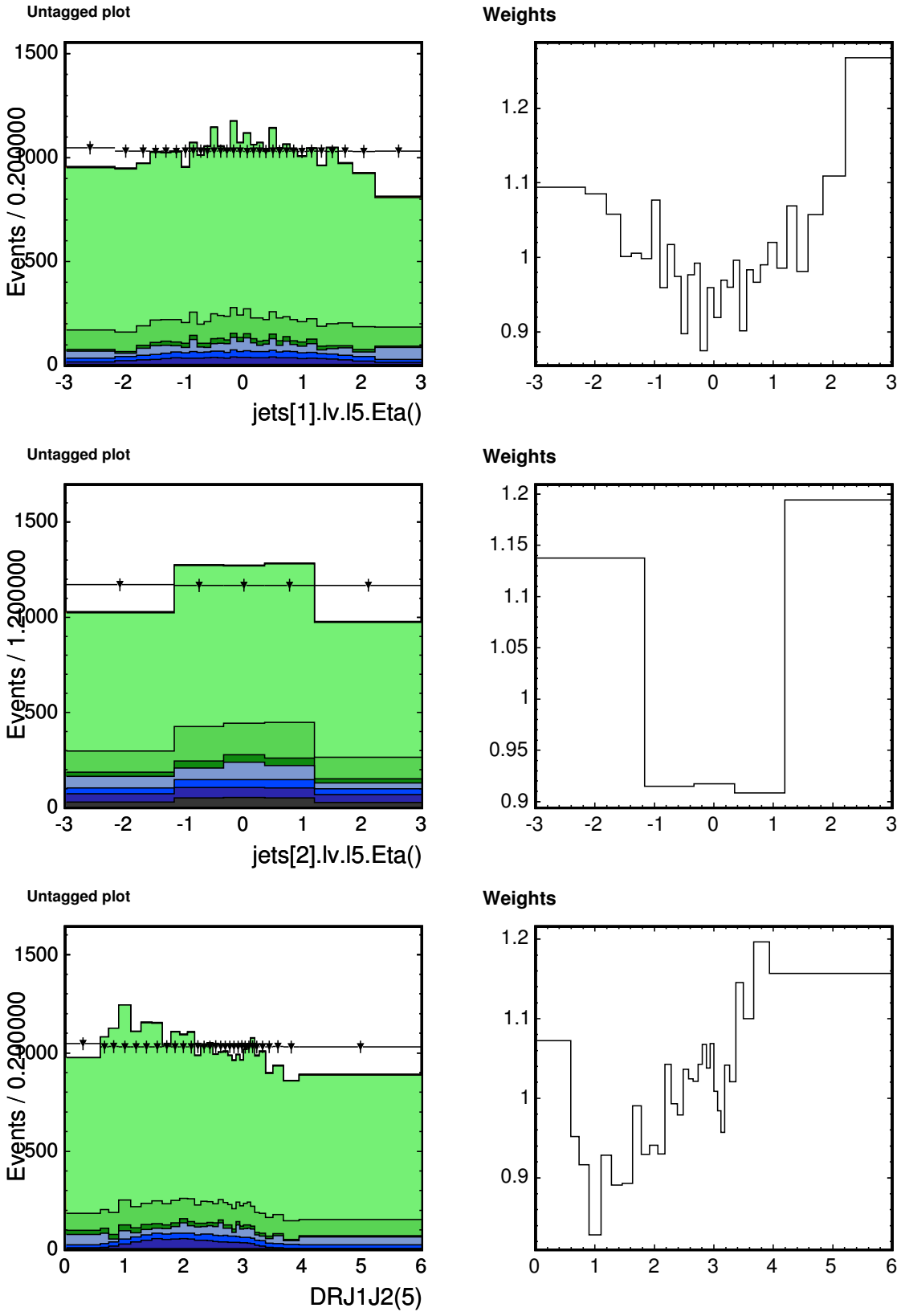


Figure 29: Distributions which suggest a mis-modeling in Monte Carlo, binned so as to be flat in data (to reduce the effect of statistical uncertainties) and the resulting weights assigned: pseudorapidity of the second jet in the two-jet bin, pseudorapidity of the third jet in the three-jet bin, and ΔR between the jets in the two-jet bin.

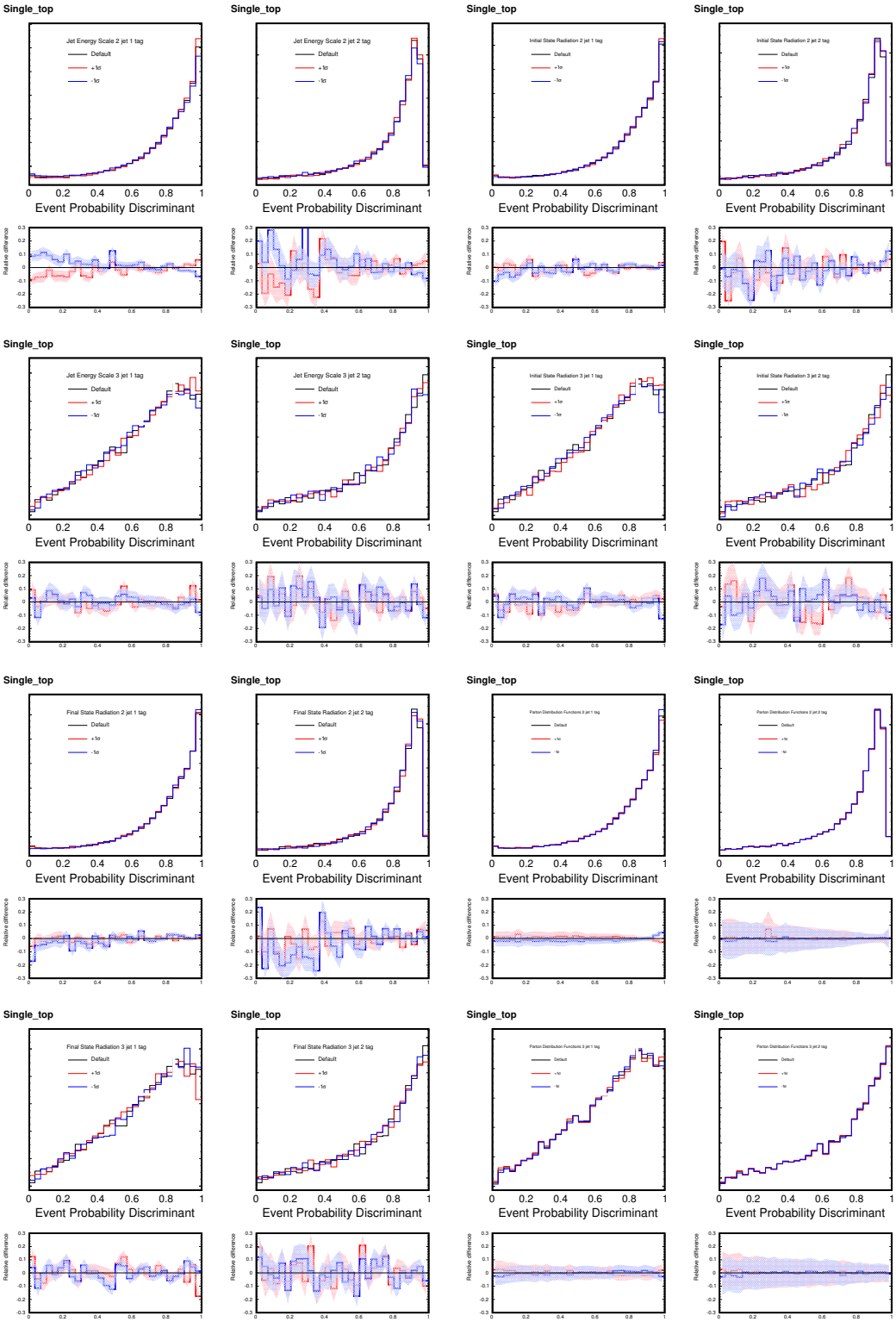


Figure 30: Shape systematics for the single top template.

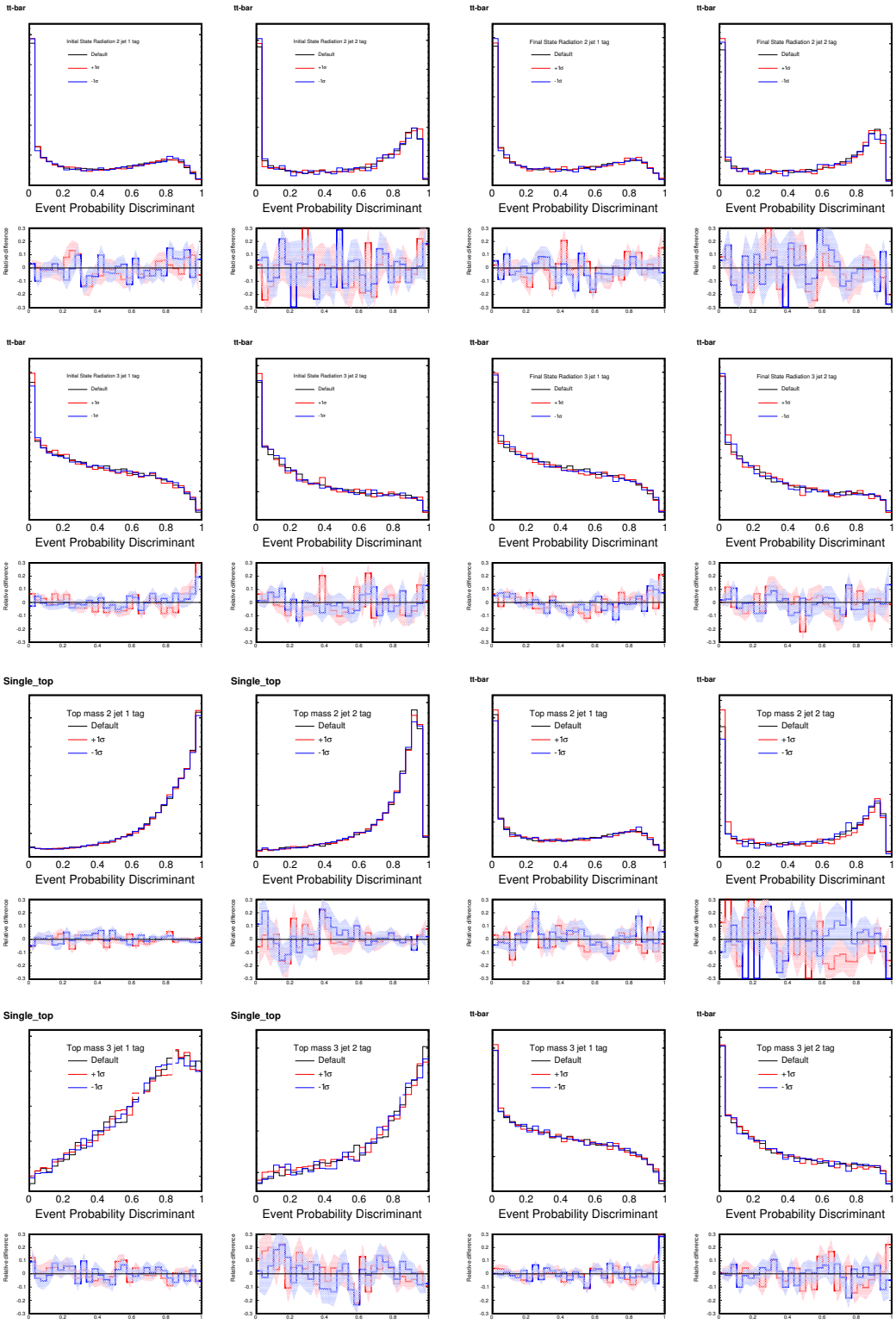


Figure 31: Shape systematics for initial and final state radiation and top mass for the $t\bar{t}$ template. The top mass samples (170 and 180 GeV/c^2) are not used as a source of systematic uncertainty, but are used to illustrate the change in the shape of the distribution for different top masses.

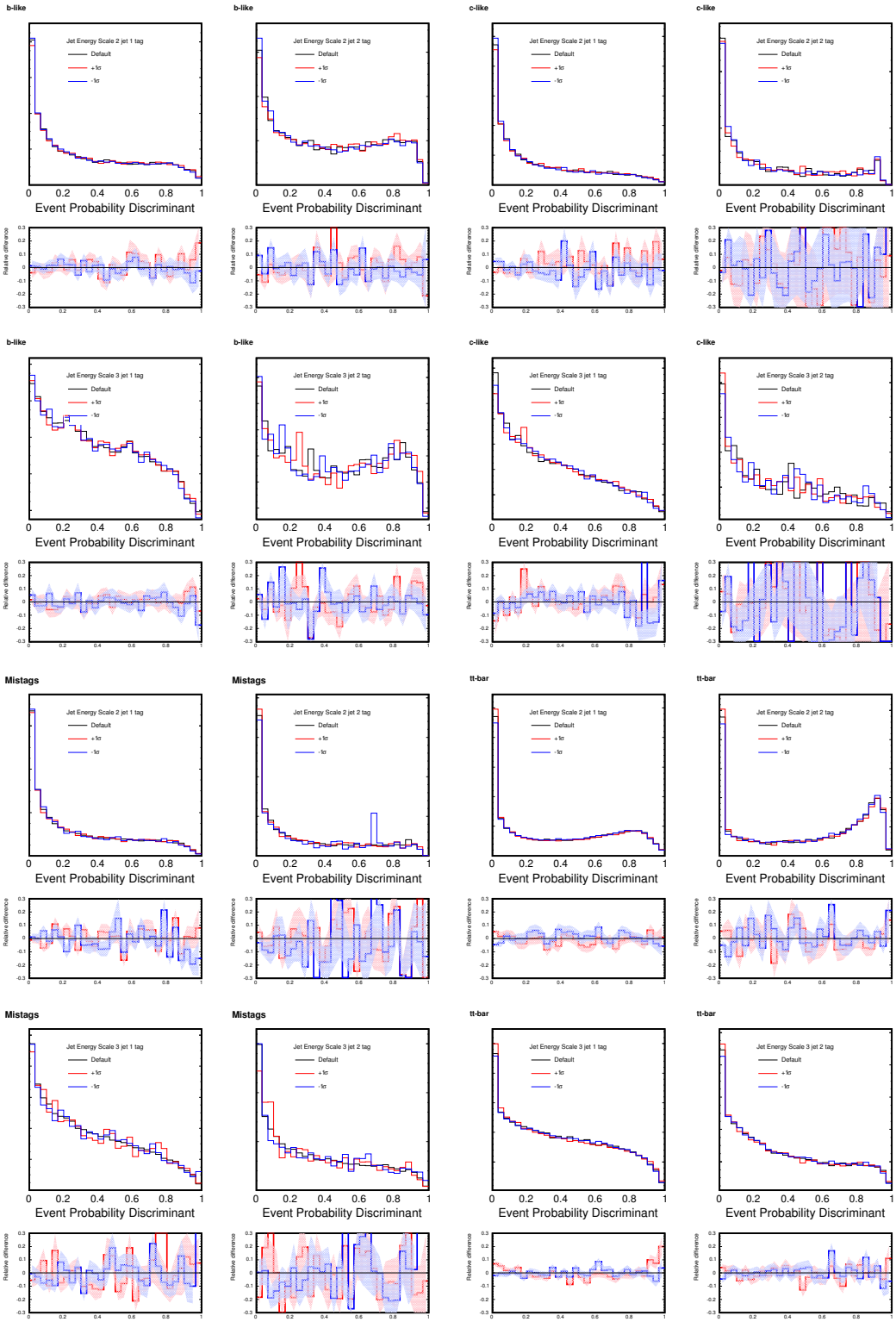


Figure 32: JES systematic uncertainty evaluated in background events.

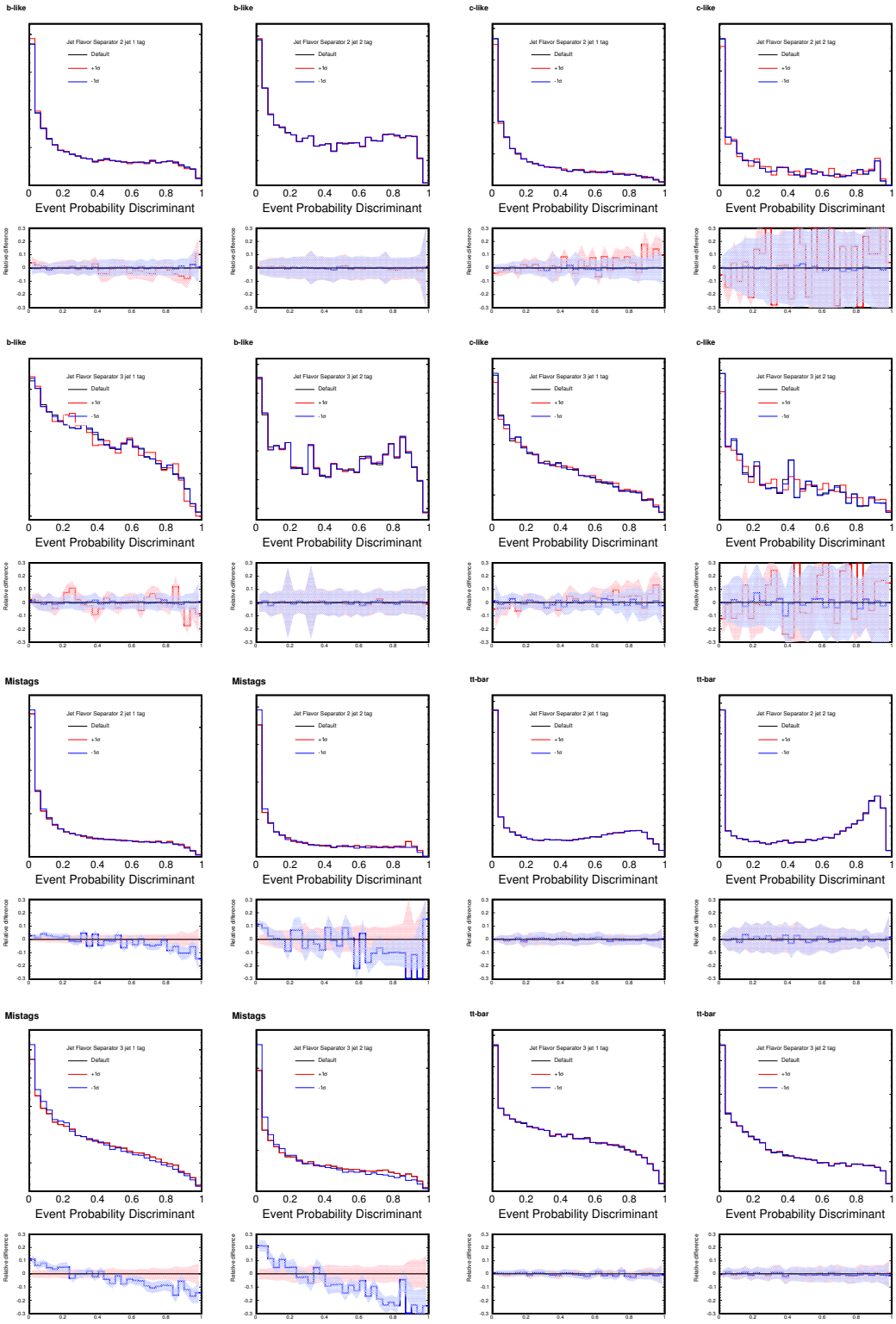


Figure 33: Neural net jet flavor separator systematic uncertainty evaluated in background events.

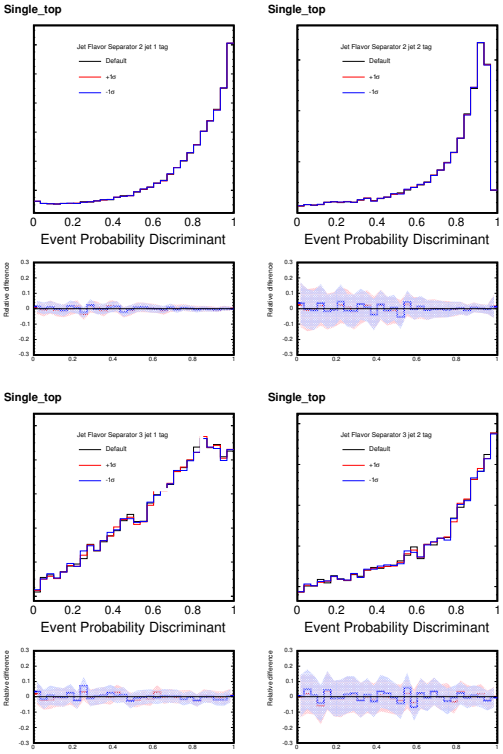


Figure 34: Jet flavor separator systematics uncertainty evaluated on the single-top template.

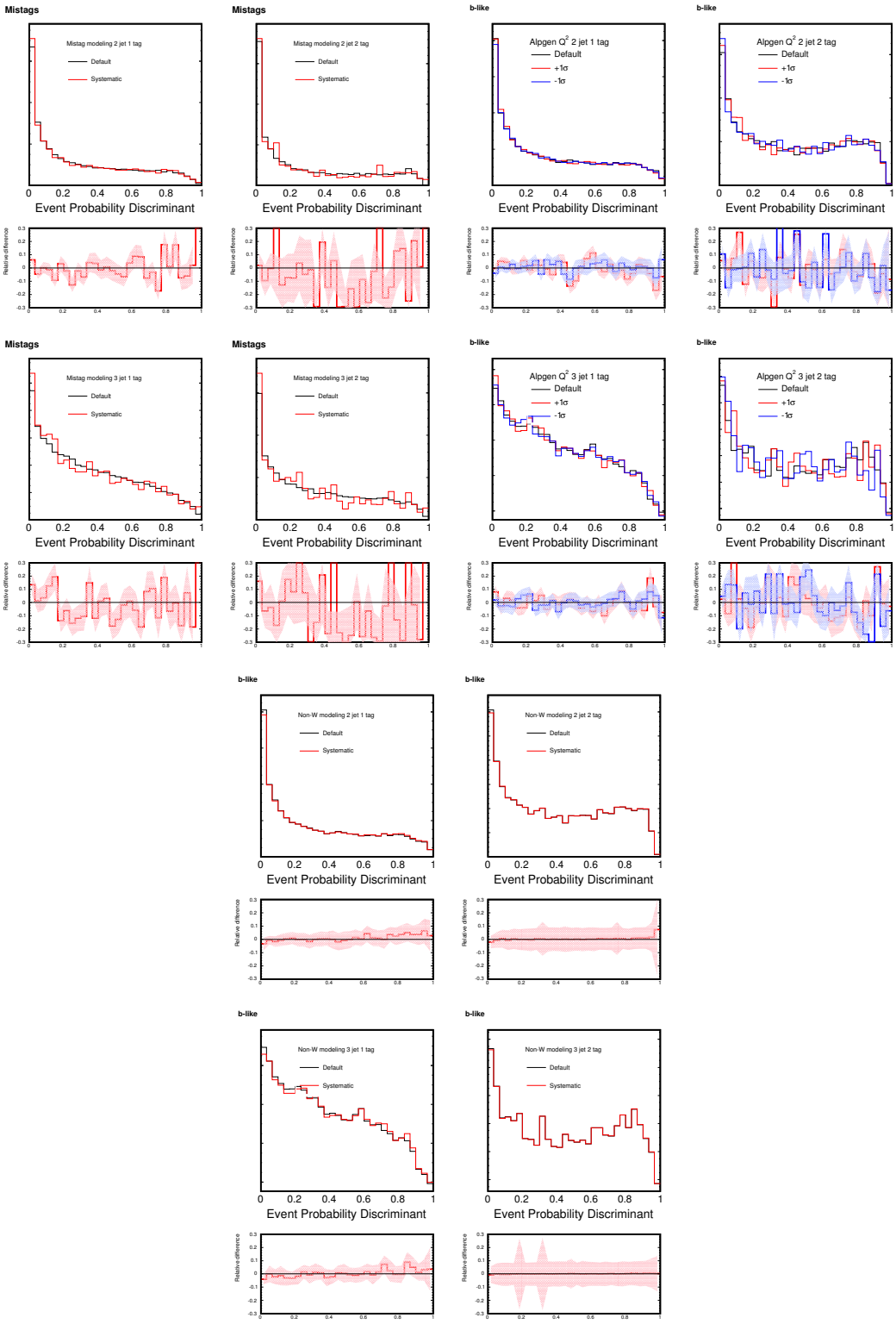


Figure 35: Background systematic uncertainties, including mistag model uncertainty, ALPGEN factorization and renormalization scale uncertainty, and non-W uncertainty.

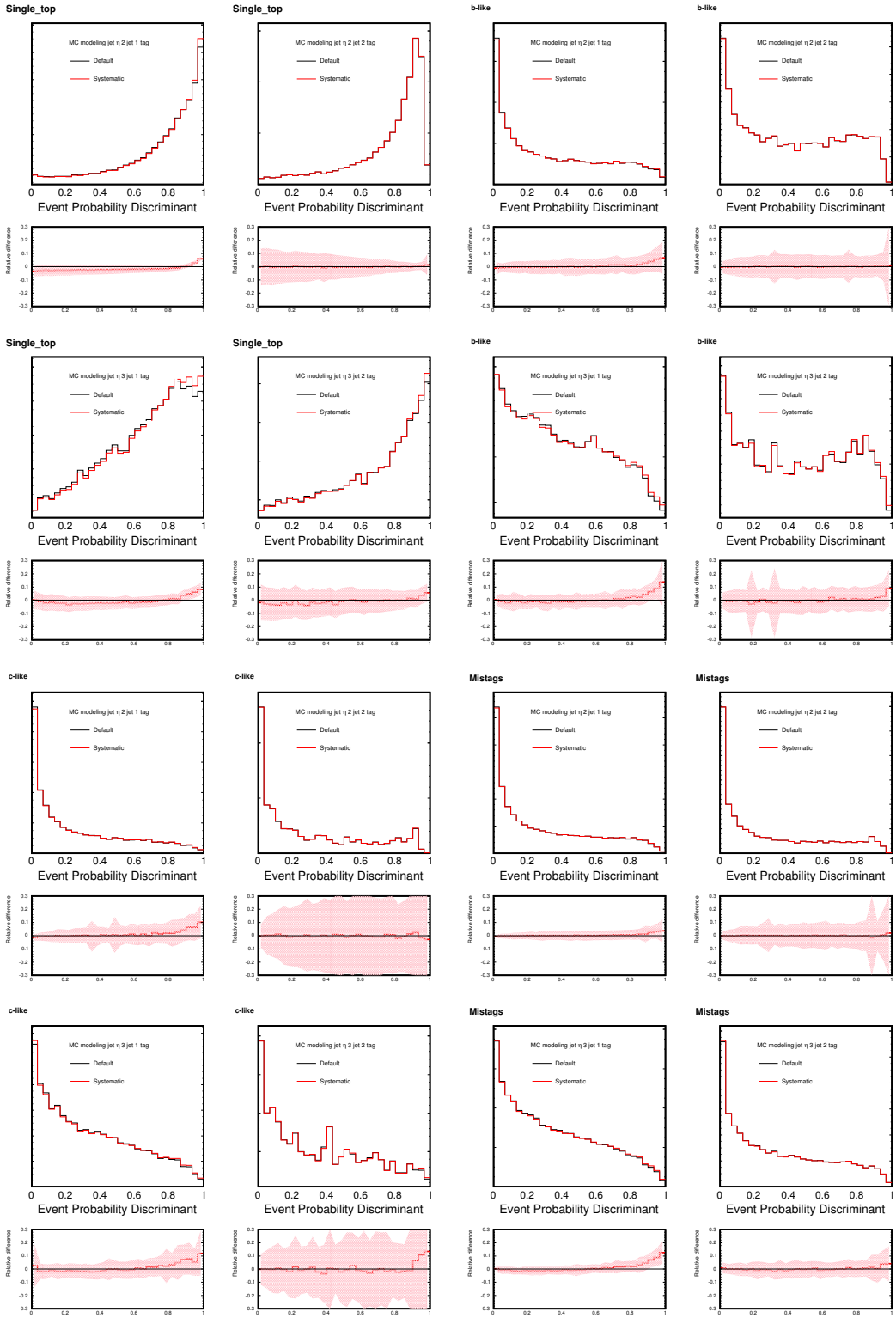


Figure 36: Systematic uncertainty due to mis-modeling of last jet pseudo-rapidity.

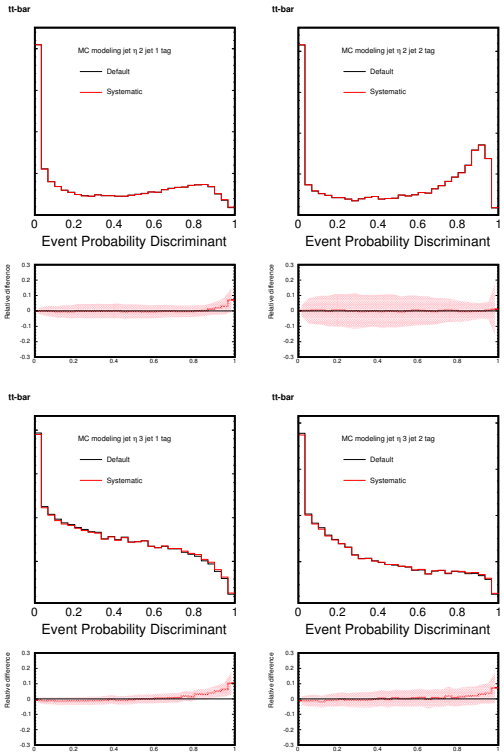


Figure 37: Systematic uncertainty due to mis-modeling of last jet pseudo-rapidity.

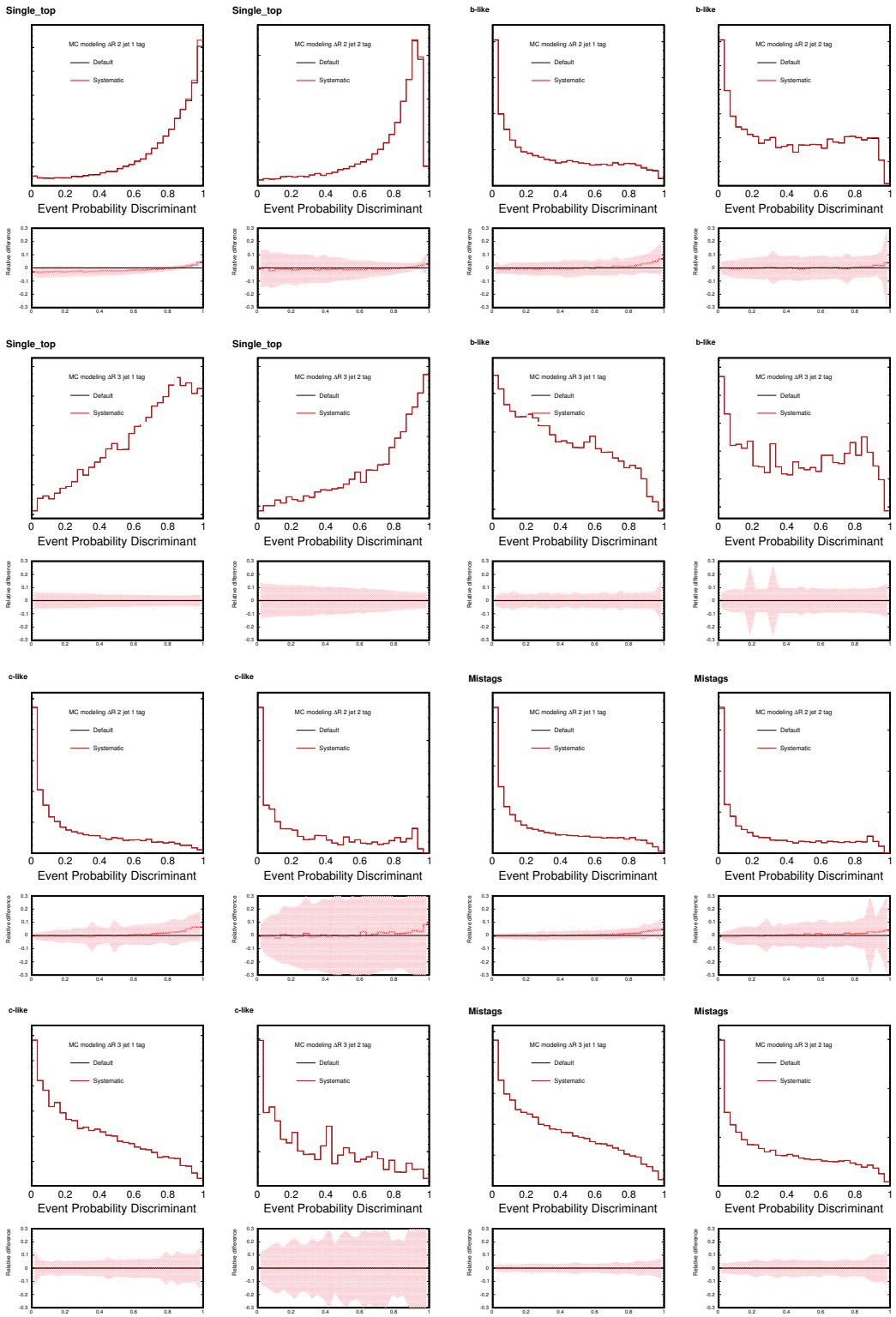


Figure 38: Systematic uncertainty due to mis-modeling of ΔR between the two jets.

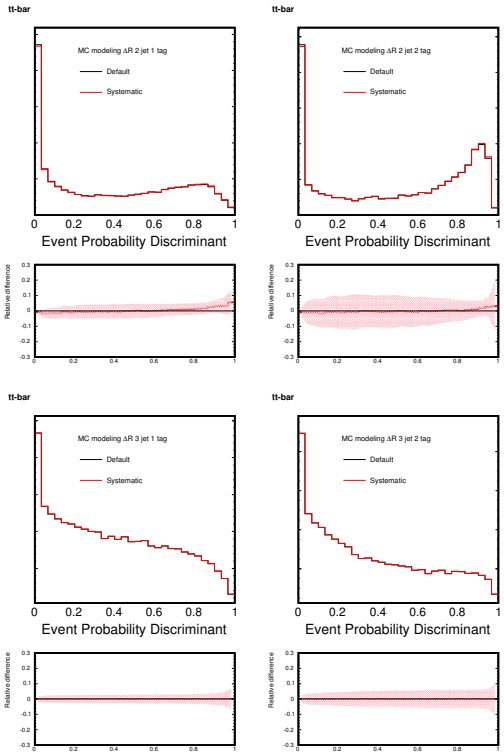


Figure 39: Systematic uncertainty due to mis-modeling of ΔR between the two jets.

3.0.1 Testing the Maximum Likelihood Method

To test the maximum likelihood method, we apply it to many pseudo-experiments. Each pseudo-experiment consists of a sample of simulated data similar to the one we would expect to obtain from repeated CDF II experiments. The distribution of fitted parameters can then be statistically examined.

Each pseudo-experiment consists of an ensemble of EPD values randomly generated according to the expected EPD distribution (template) for signal and background. Each pseudo-experiment is composed of $N_{single\ top}$, $N_{W+b\ jets}$, $N_{W+c\ jets}$, N_{Mistag} and $N_{t\bar{t}}$ Monte Carlo events, respectively. We first modify the normalization of the templates by allowing all systematic uncertainties (including background normalization uncertainty) to fluctuate as a Gaussian with a width determined by their uncertainties. In the case of asymmetric errors, we take a different Gaussian for positive and negative values of the nuisance parameter, causing a discontinuity at the central value. Having arrived at the template normalizations for a given pseudo-experiment, we compute the particular values of $N_{single\ top}$, $N_{W+b\ jets}$, $N_{W+c\ jets}$, N_{Mistag} and $N_{t\bar{t}}$ by numerically Poisson fluctuating the signal and background predictions summarized in Table 3.

Before we draw values from the templates, we first distort the templates by the shape systematic uncertainties. Each systematic parameter associated with a shape uncertainty causes each template to change shape, so we distort templates accordingly before drawing events from them.

Once the systematic uncertainties have been properly accounted for, we randomly draw $N_{single\ top}$, $N_{W+b\ jets}$, $N_{W+c\ jets}$, N_{Mistag} and $N_{t\bar{t}}$ events from the Monte Carlo templates of Figure 21. The obtained dataset represents one pseudo-experiment. The ensemble of EPD values are, of course, different for each pseudo-experiment.

We test the machinery with a linearity check. This involves taking different input values of the single top cross section, ranging from zero to four times the Standard Model expectation, and checking that the median of many pseudo-experiments returns the fit value that we put in. (We use the median because the Bayesian prior at zero prevents fitted cross sections from getting any lower, thus biasing the mean of results that go close to zero.) We marginalize all nuisance parameters in each pseudo-experiment to produce a cross section and plot the median at each value. The resulting plot is shown in Figure 41. The pull distribution shows that the fitter is unbiased, as shown in Figure **??**.

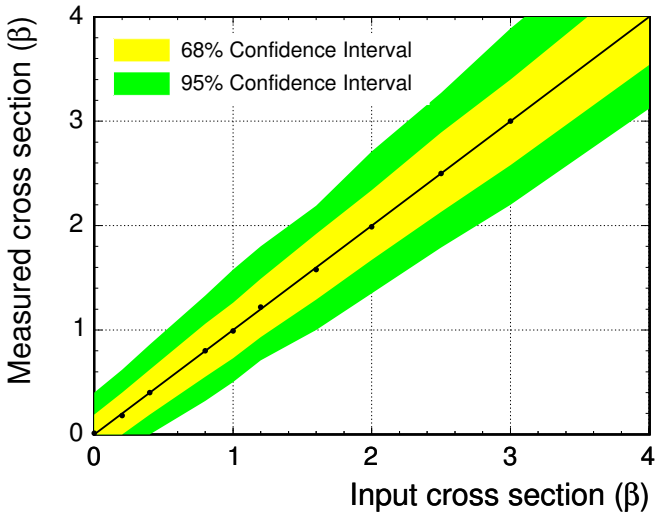


Figure 40: A check of the linearity of the likelihood and marginalization machinery in pseudo-experiments generated at different values of the single top cross section. The line is a diagonal line representing the ideal values and the markers indicate the median values of the pseudo-experiments at each point. The yellow band covers 68% of the pseudo-experiments and the green band covers 95%.

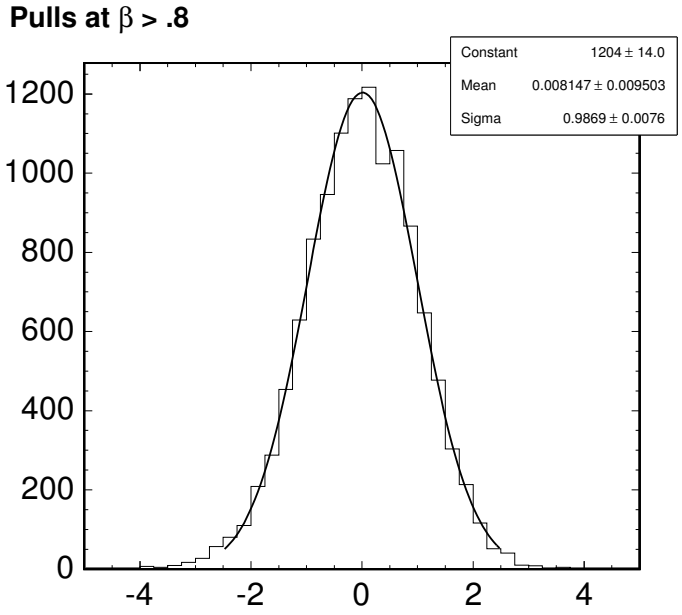


Figure 41: The pulls from the linearity check. We add the pulls for all different values of beta greater than 0.8. Because of the cutoff in the Bayesian prior, we cannot plot unbiased pulls near a beta of zero.

3.0.2 Multiple Fit Regions

To increase the sensitivity of the analysis, we create separate templates for the untriggered muons, which have a substantially different background composition from the rest of the subdetectors. We create templates for single and double tags separately, and for the two-jet and three-jet bins separately. We fit for one common cross section across all eight different regions when we perform the final fit. This allows us to include information from

3.0.3 Template Binning

We performed a test in each of our twenty channels to determine the optimal binning. Once the effect of Monte Carlo statistics was considered, we saw little change in sensitivity with the number of bins, and we settled on 40 bins, which gave a 1% improvement over the 30-bin templates.

3.1 Cross Check of the Method

We have seen in previous sections how the event-probabilities are defined and calculated. In this section we cross-check whether the MC prediction of the event-probability-discriminant

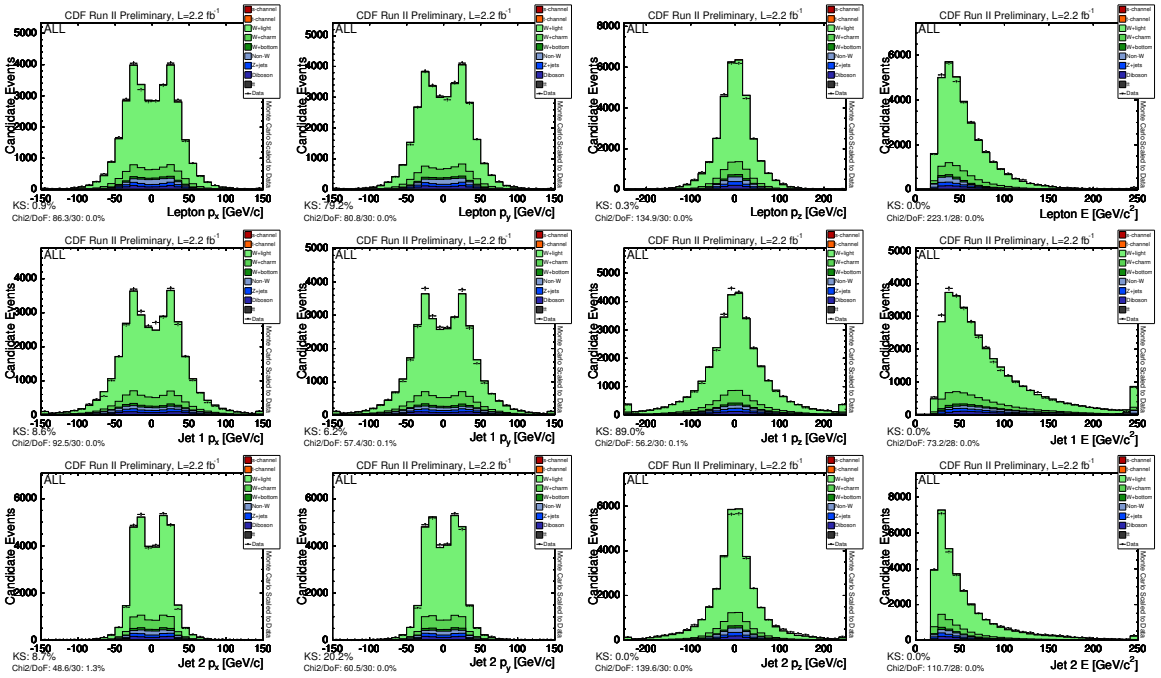


Figure 42: Comparison of the event probability input 4-vectors for untagged W + 2 jet control data (with at least one taggable jet) compared to the Monte Carlo prediction.

represents the data well without looking at the single top candidate events. We choose the "taggable but not tagged" side-band data. That is, we select W+2 jets events according to our nominal event selection and require that at least one jet is taggable but that neither are tagged by the SECVTX algorithm. This event selection is orthogonal to the single top signal region while it still represents a very similar kinematic event topology. Another advantage is that this sample has very little contribution from top (<0.5%). The measured quantities which are input in the probabilities are the same as in the single top analysis, namely the measured four-vectors of the lepton as well as the measured four-vectors of the leading and second leading jet. Figures 42 and 43 shows data/Monte Carlo comparisons for the measured four-vector components of the lepton and the two jets. Figures 44 and 45 shows the transverse momentum and pseudo-rapidity distribution for the same reconstructed objects. We use ALPGEN W + 2 jet Monte Carlo and W + 2 jet data events. The Monte Carlo distributions are normalized to the data. We find generally good agreement between data and Monte Carlo prediction. We also plot the distribution of the event probability discriminant shown in Figure 46. The agreement is good, which assures us that the event probability discriminant in Monte Carlo is well represented by data.

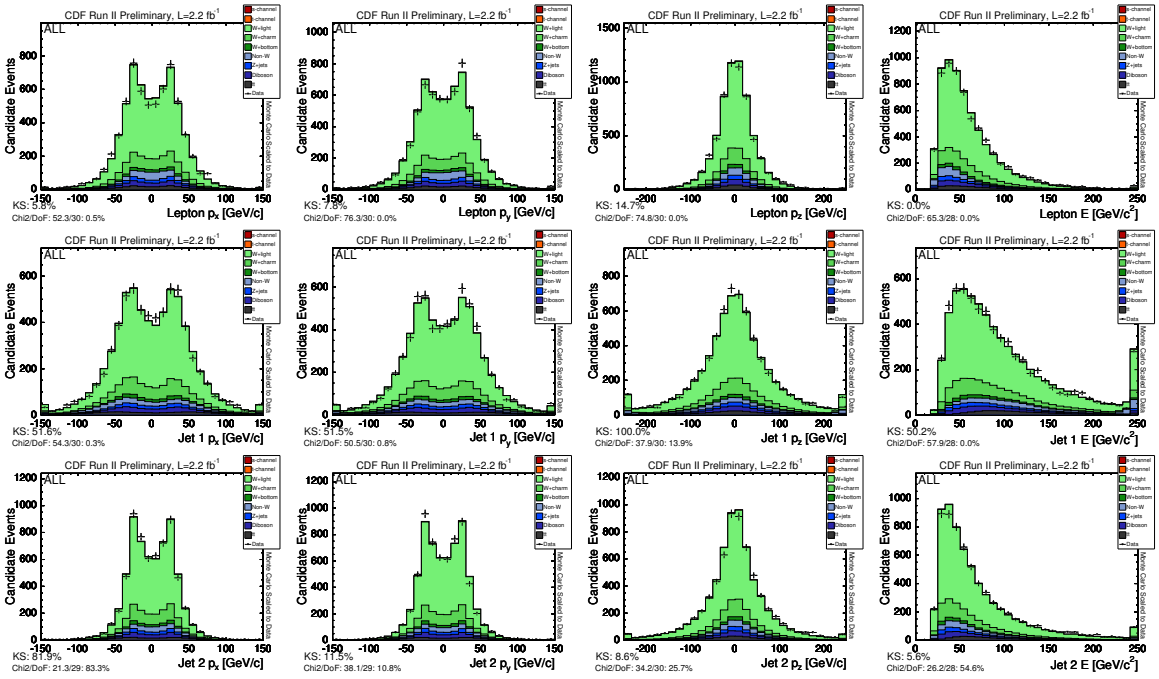


Figure 43: Comparison of the event probability input 4-vectors for untagged W + 3 jet control data (with at least one taggable jet) compared to the Monte Carlo prediction.

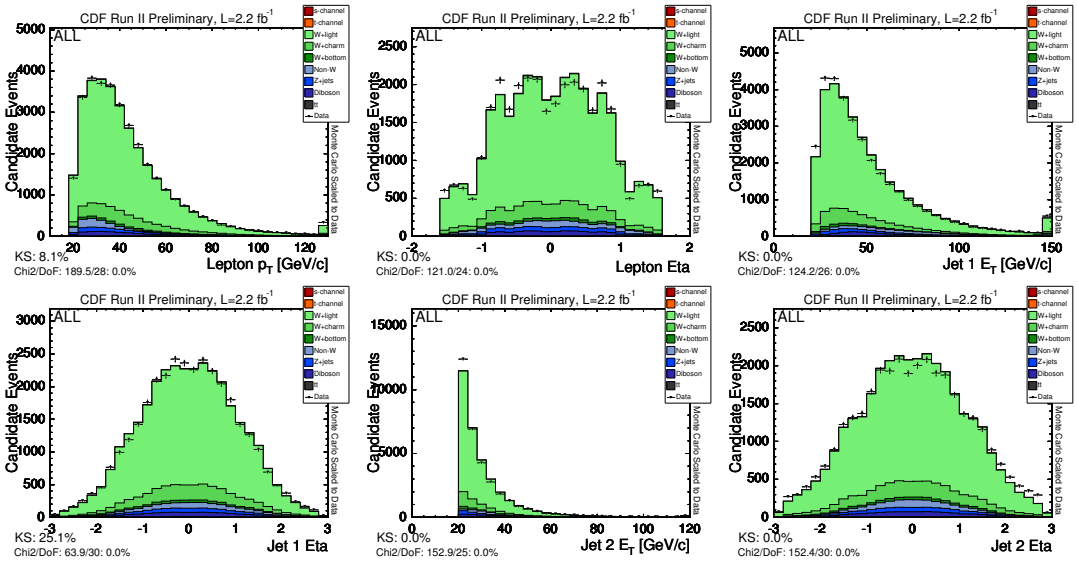


Figure 44: Comparison of the event probability input p_T and pseudo-rapidity distribution for untagged W + 2 jet control data (with at least one taggable jet) compared to the Monte Carlo prediction.

We also cross-checked the shape of our $t\bar{t}$ template by looking in the $t\bar{t}$-enriched data tagged lepton + 4 jets sample. Figure 47 shows the comparison. The lepton

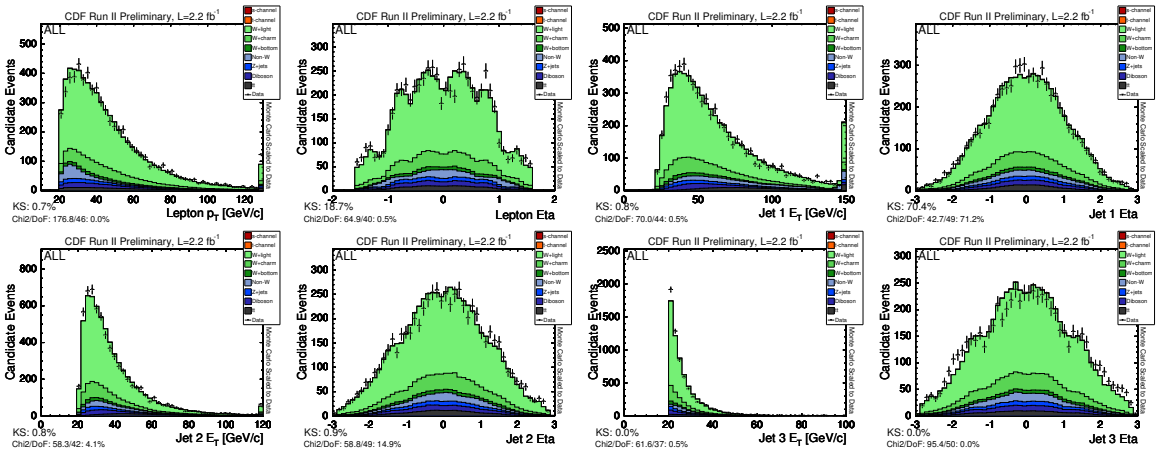


Figure 45: Comparison of the event probability input p_T and pseudo-rapidity distribution for untagged W + 3 jet control data (with at least one taggable jet) compared to the Monte Carlo prediction.

+ 4 jets sample has an expected $t\bar{t}$ contribution of about 75%. In these plots, the Monte Carlo distributions are normalized to the data. Within statistics, we find good agreement in the data and Monte Carlo shapes.

After performing cross-checks in data sidebands, we take a look at the tagged W+2 jets data. Figures 48 and 49 show all six input variables. We compare data to Monte Carlo distributions normalized to the *Method 2* prediction for the tagged W+2 jets data, i.e. the single top signal region. For shape comparison, the stacked Monte Carlo distributions are scaled to the data using an overall scale factor of 1.15. Within uncertainty, we find good agreement between data and Monte Carlo.

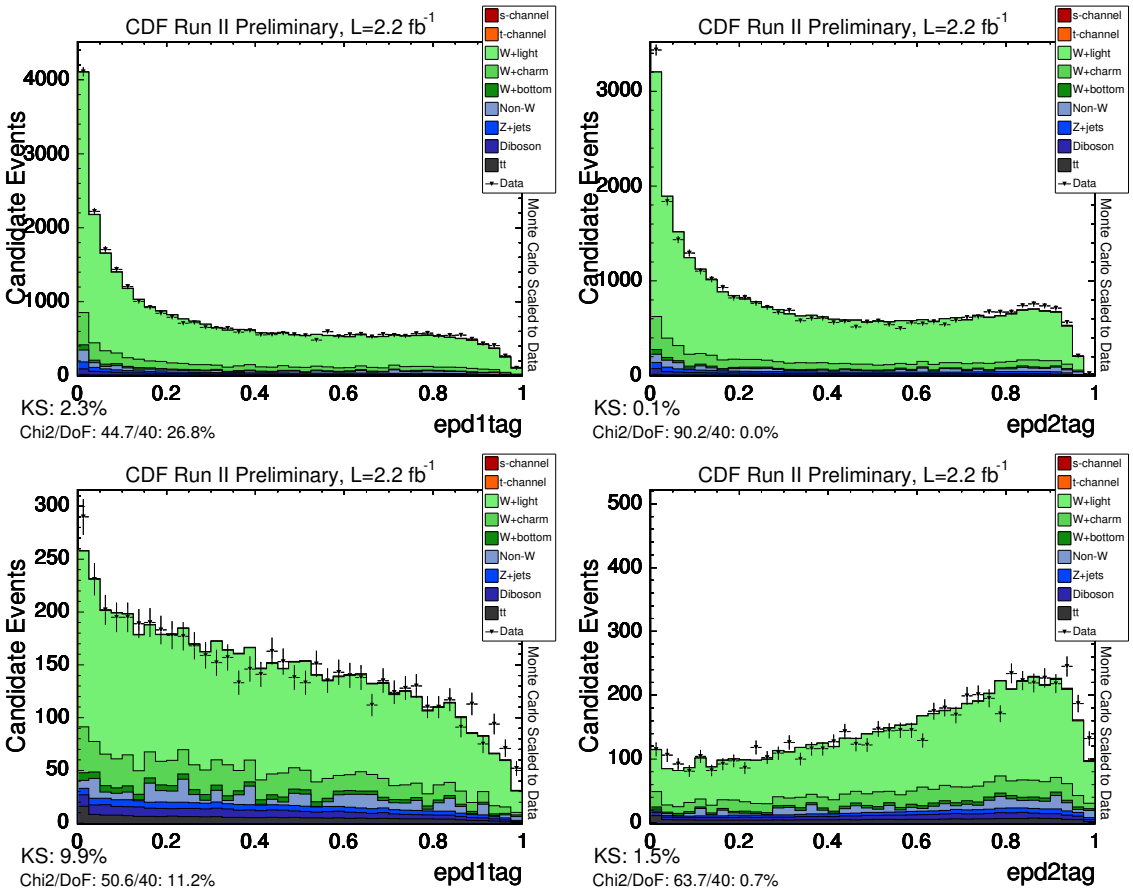


Figure 46: Comparison of the event probability discriminant for $W+$ 2 jet Monte Carlo and $W+$ 3 jet data in the 'taggable but untagged' control sample. The error bars on the data points are Gaussian errors.

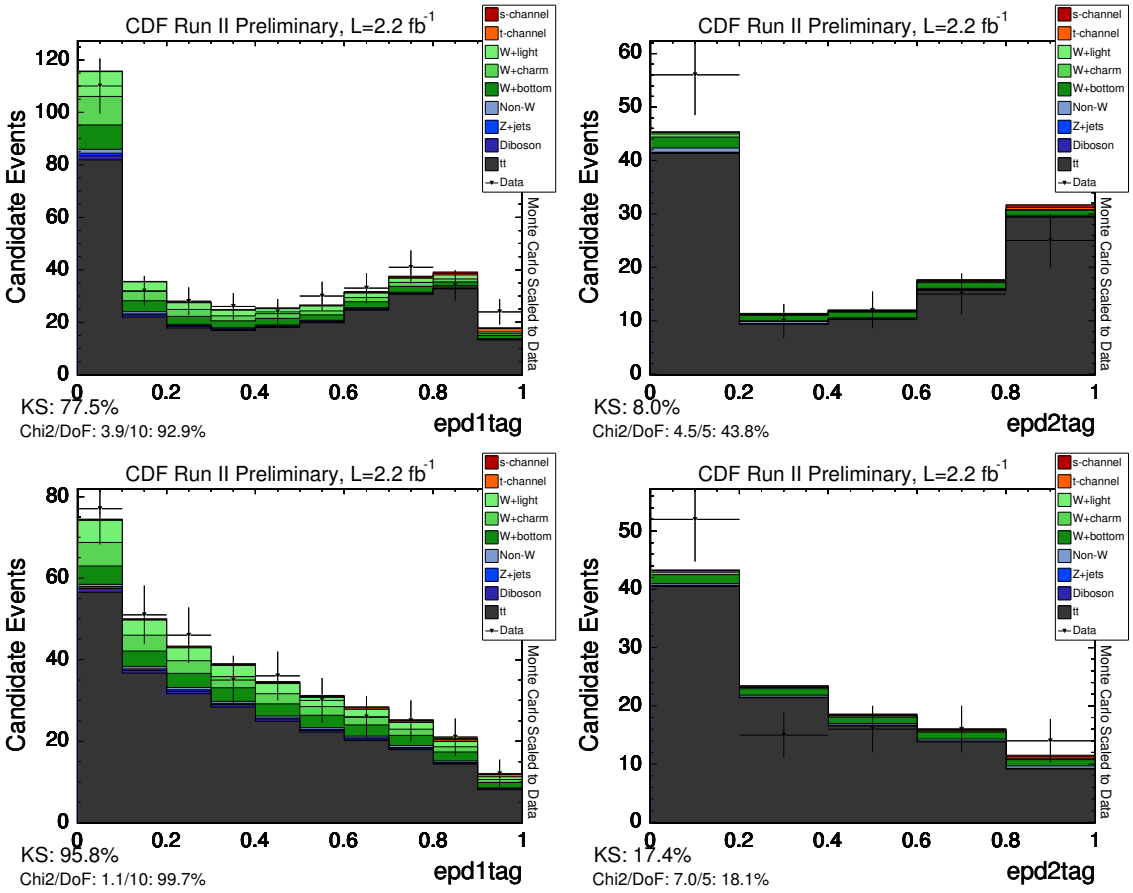


Figure 47: Comparison of the event probability discriminant for lepton + four jets data and Monte Carlo, using the leading two jets in the matrix element calculation. Only the data and the dominant $t\bar{t}$ sample were processed; the other backgrounds use the two-jet shapes. The left-hand plots are the single tagged events and the right-hand plots are double-tagged. The top plots are in the two-jet bin and the bottom plots are in the three-jet bin.

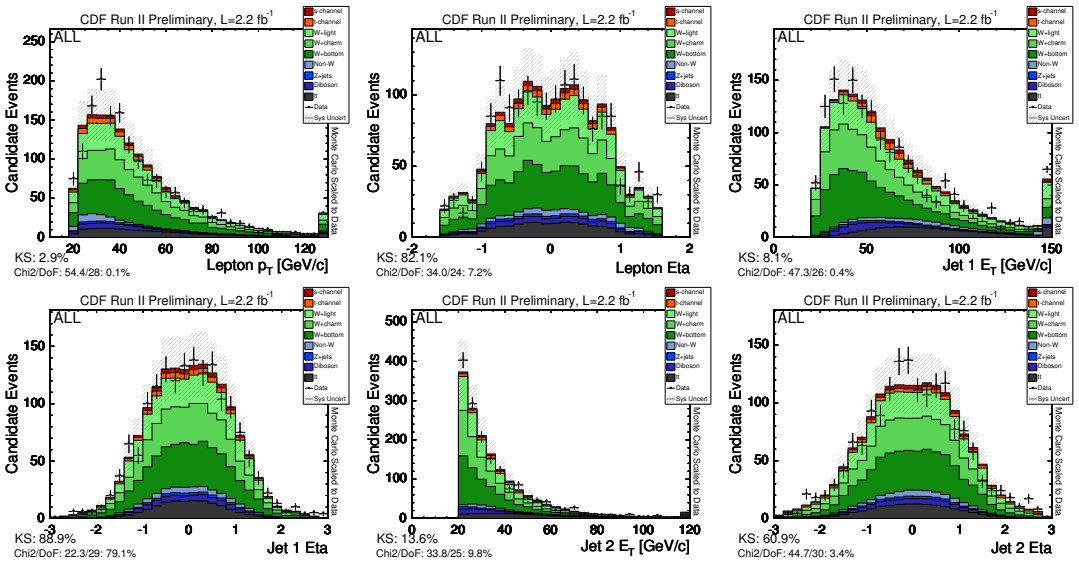


Figure 48: Comparison of the event probability input variables for Monte Carlo prediction and data in the W+2 jet signal region.

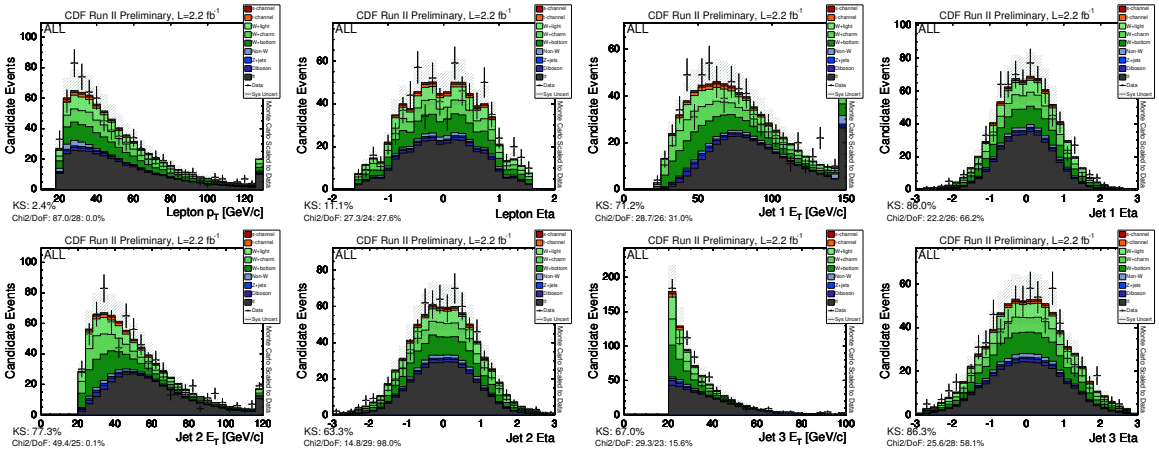


Figure 49: Comparison of the event probability input variables for Monte Carlo prediction and data in the W+3 jet signal region.

4 Result with CDF II Data

We apply the analysis to 2.2 fb^{-1} of CDF Run II data. We compare the EPD distribution of 2260 candidate events with the sum of predicted single top and background templates.

In order to extract the most probable single top content in the data we perform the maximum likelihood method described in Section 2.9. The posterior p.d.f is obtained by using Bayes' theorem:

$$p(\beta_1|data) = \frac{\mathcal{L}^*(data|\beta_1)\pi(\beta_1)}{\int \mathcal{L}^*(data|\beta_1')\pi(\beta_1')d\beta_1'}$$

where $\mathcal{L}^*(data|\beta_1)$ is the reduced likelihood and $\pi(\beta_1)$ is the prior p.d.f. for β_1. We adopt a flat prior, $\pi(\beta_1) = H(\beta_1)$, in this analysis, with H being the Heaviside step function.

The most probable value (MPV) corresponds to the most likely combined single top production cross section given the data. The uncertainty corresponds to the range of highest posterior probability density which covers 68.27% [6].

We perform marginalization using the likelihood function of Equation 15 with all systematic rate and shape uncertainties included in the likelihood function. The most probable value for the single top cross section is obtained at $2.3^{+0.8}_{-0.7}$ pb. The posterior probability density is shown in Figure 51.

It is useful to check that our background constraints are not biasing our answer by performing an unconstrained fit, in which the Gaussian constraints on the background normalizations are removed. Unfortunately, such a fit has no meaning in a marginalization method, so to use the information we must perform a fit, profiling the nuisance parameters. This can only be done reliably when the nuisance parameters are smooth and symmetric, so we perform the fit with only the background constraints and without the additional systematic uncertainties. The resulting fit gives a value of $2.6^{+0.7}_{-0.7}$ pb, and the resulting templates are shown in Figure 52.

If we remove the constraints and perform the fit again, we get a fit value of $2.7^{+0.8}_{-0.7}$. The resulting templates are shown in Figure 53. This indicates that our normalization constraints are not biasing our answer.

5 Expected Sensitivity and Hypothesis Test

We interpret the result using the CLs/CLb method developed at LEP [20]. We compare our data against two models, one asserting that the data is due to background processes only (b) and one which includes Standard Model single top production in addition to the background processes (s+b). We propagate all systematic uncertainties in our statistical method as described in the next section 3. Using the test statistic $Q = \frac{\mathcal{L}(data)|s+b}{\mathcal{L}(data)|b}$ we compute the probability (p-value) that the background only (b) model fluctuated equal or up to the observed value Q_{obs} in the data (observed p value) and

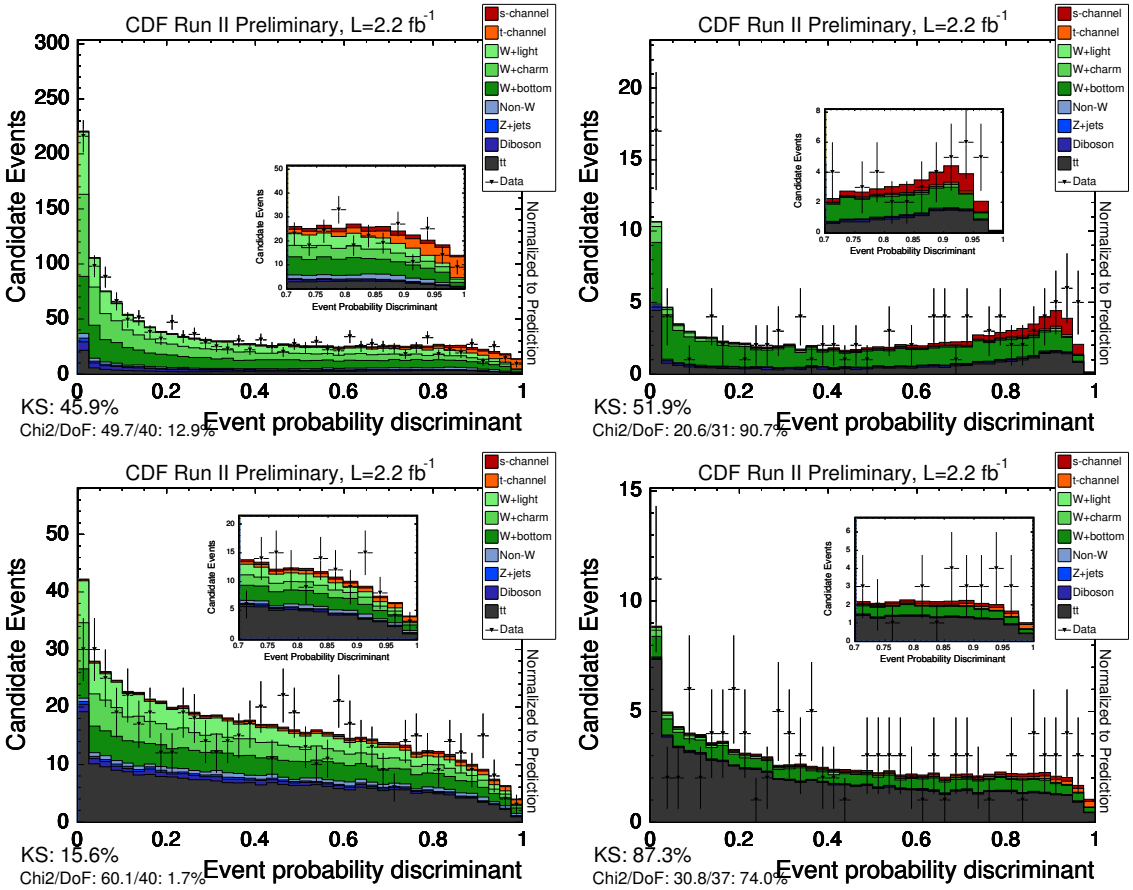


Figure 50: CDF data compared to Monte Carlo prediction for signal and background. The top plots show the two-jet bin while the bottom plot shows the three-jet bin; the left-hand plots show the single-tagged events while the right-hand plots show the double-tagged events.

to the median Q value of signal+background (s+b) pseudo-experiments (expected p-value). Figure 54 shows the distribution of the test statistics for pseudo-experiments performed for (b) and (s+b). We expect a p-value of 3×10^{-4} % (4.5σ). and observe a p-value of 0.03% (3.4σ) in the data.

6 Measurement of $|V_{tb}|$

We can measure the CKM matrix element $|V_{tb}|$ by using the fact that the cross section of single-top processes is proportional to $|V_{tb}|^2$ (due to the single tb electroweak vertex in the Feynman diagram). Therefore, we need only take the square root of our measured cross section divided by the expected Standard Model cross section. This measurement assumes $|V_{td}|^2 + |V_{ts}|^2 << |V_{tb}|^2$ and a Standard Model V-A and CP-conserving Wtb vertex.

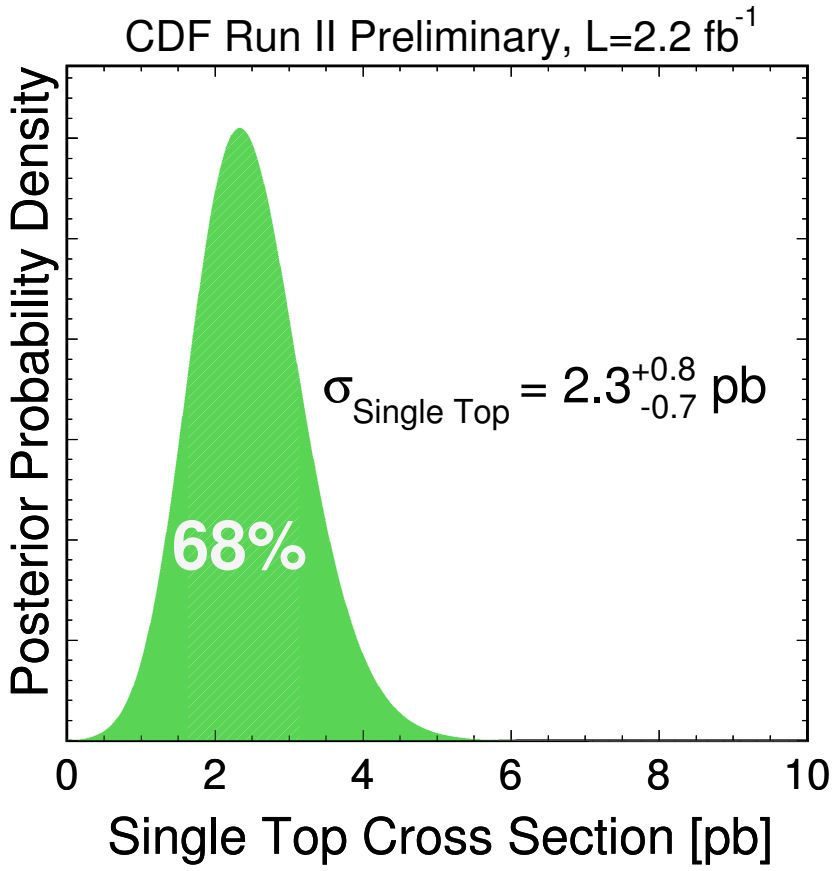


Figure 51: Marginalized likelihood of the single top cross section using 2.2/fb of data. The error band shows the 68% uncertainty (all systematics included) on the measurement.

We account for theory uncertainty due to the dependence of the single top cross-section on the top quark mass, the factorization and renormalization scales, parton distribution functions and the value of α_s as discussed in [22]. We measure: $|V_{tb}| = 0.90^{+0.14 experiment}_{-0.15} \pm 0.07^{theory}$.

We can also calculate a lower limit on $|V_{tb}|$ by integrating our posterior probability density in $|V_{tb}|^2$. Using a flat prior for $|V_{tb}|^2$ from 0 to 1 we obtain $|V_{tb}| > .64$ at 95% confidence (see Figure 55).

7 Conclusions

We have used the matrix element analysis technique in a direct search for electroweak single top quark production. Our search was done simultaneously for s-channel and t-channel single top production. To extract the most probable single top content in data, we apply a maximum likelihood technique. All sources of systematic rate and

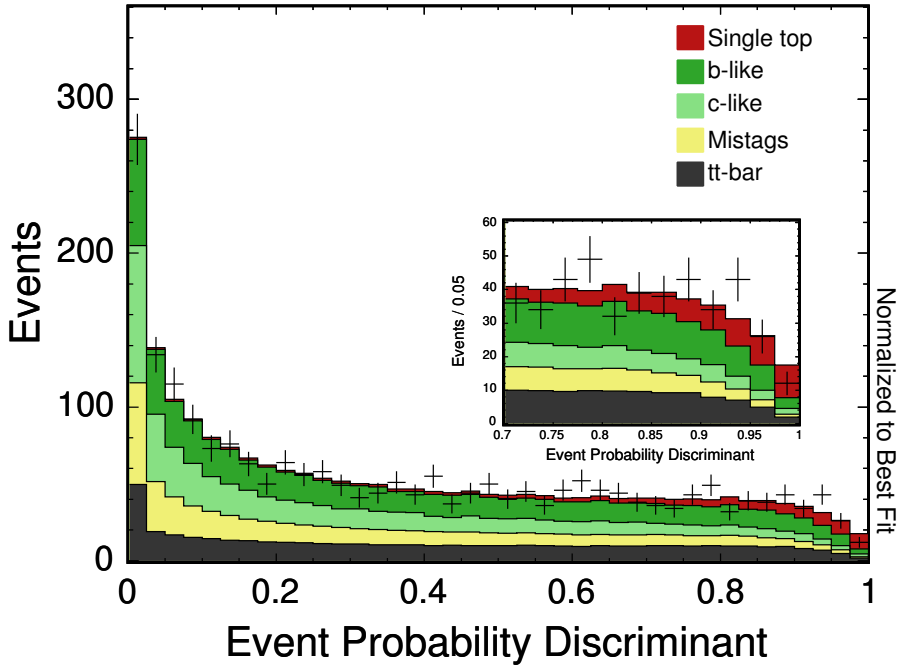


Figure 52: Matrix element EPD fit with only background normalization uncertainties constrained to their predicted values.

shape uncertainty are included in the likelihood function. We have analyzed 2.2 $fb^{-1}$of CDF Run II data and measure a combined s- and t-channel single top production cross section of:

$$\sigma_{singletop} = 2.3^{+0.8}_{-0.7} \text{ pb (all systematics included)}$$

assuming a top quark mass of 175 GeV/c^2. This allows us to measure the value of $|V_{tb}| = 0.90^{+0.14 experiment}_{-0.15} \pm 0.07^{theory}$ and set a limit of $|V_{tb}| > 0.64$ at 95% confidence. We use the MCLIMIT program [20] to calculate the signal significance. The observed p-value in $2.2 fb^{-1}$ of CDF data is 0.00021. The expected (median) p-value in pseudo-experiments is 0.0000050.

Acknowledgments

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural Sciences and Engineering Research Council of Canada; the National Science Council of the Republic of China; the Swiss National Science Foundation; the A.P. Sloan Foundation; the Bundesministerium fuer Bildung und Forschung, Germany; the

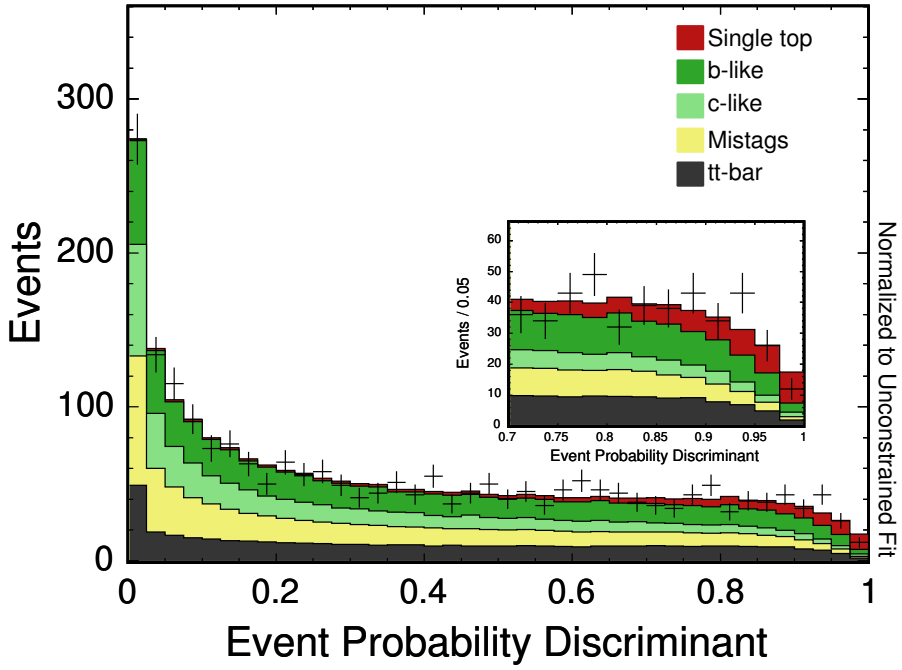


Figure 53: Matrix element EPD fit with only background normalization uncertainties and no constraints applied. This uses no information from the background estimate and shows that the background estimate is not biasing the fit value.

Korean Science and Engineering Foundation and the Korean Research Foundation; the Particle Physics and Astronomy Research Council and the Royal Society, UK; the Russian Foundation for Basic Research; the Comision Interministerial de Ciencia y Tecnologia, Spain; and in part by the European Community's Human Potential Programme under contract HPRN-CT-20002, Probe for New Physics.

We would also like to thank Tom Junk (UIUC) for very very many very very useful discussions.

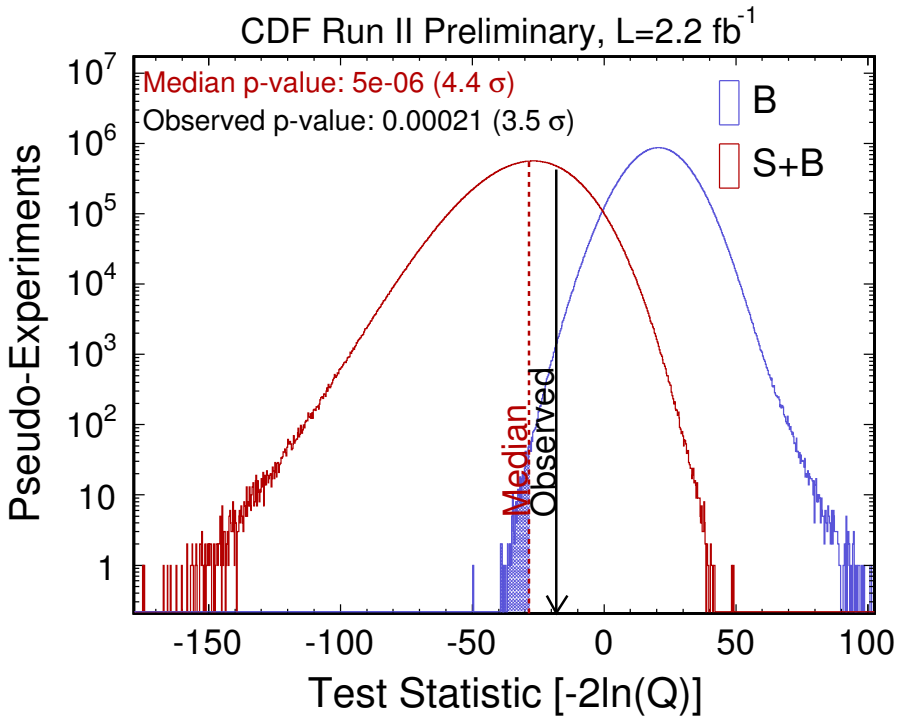


Figure 54: Distribution of the test statistics $Q = \frac{\mathcal{L}(data)|s+b}{\mathcal{L}(data)|b}$ for the single top + background hypothesis (s+b) and the background only hypothesis (b) in 25 million pseudo-experiments. The dashed line shows the median of the (s+b) hypothesis which is used to determine the expected p-value.

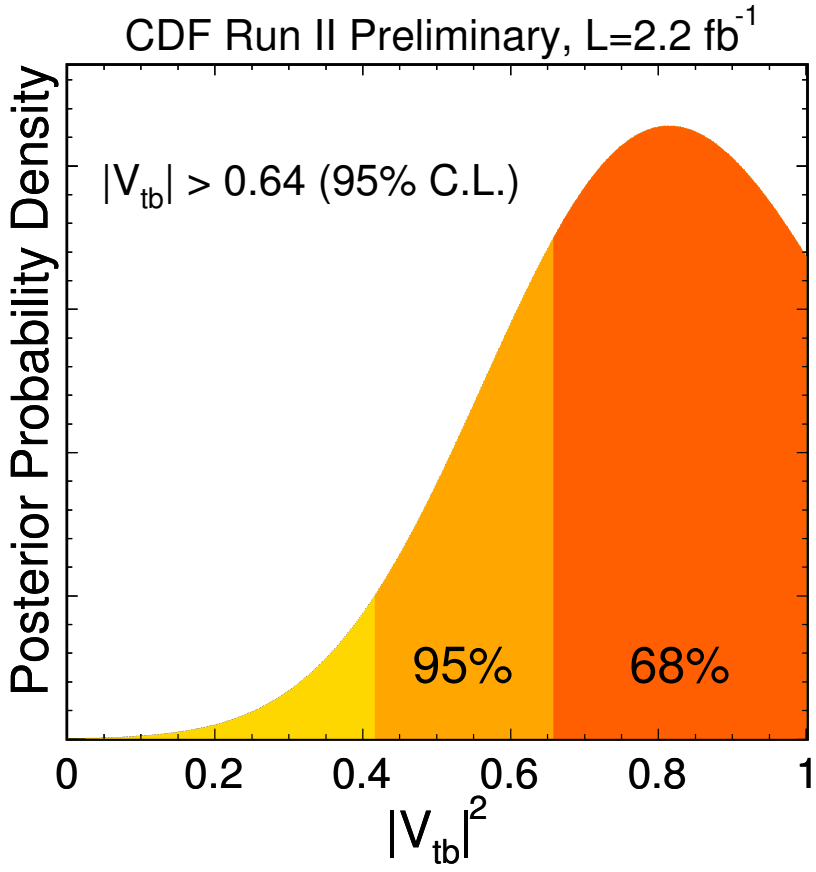


Figure 55: Likelihood scan to set a lower bound on $|V_{tb}|$.

References

[1] CDF Single Top Group *Data based background estimate for Summer 2006 Single-Top search*, CDF 8292 (2006)

[2] CDF Single Top Group *Event detection efficiency for single-top events and MC based background estimate for Summer 2006*, CDF 8286 (2006)

[3] V. M. Abazov *et al.*, *A precision measurement of the mass of the top quark*, Nature **429**, 638 - 642 (2004); M.F. Canelli, *Helicity of the W Boson in Single-Lepton $t\bar{t}$ events*, Ph.D thesis, University of Rochester (2003)

[4] F. Canelli, B. Casal Larana, C. Group, A. Ruiz, B. Stelzer, R. Vilar *Increasing Muon Acceptance with MET Plus Jet Triggers*, CDF Note 8489 (2006)

[5] CDF Collaboration, D. Acosta *et al.*, *Search for electroweak single-top-quark production in $p\bar{p}$ collisions at $\sqrt{s}$=1.96 TeV*, Phys. Rev. D **71**, 012005 (2005)

[6] Particle Data Group, *The Review of Particle Physics*, http://pdg.lbl.gov (2004)

[7] E. Murayama, I. Watanabe and K. Hagiwara, *HELAS: HELicity Amplitude Subroutines for Feynman Diagram Evaluations*, KEK Report 91-11 (1992)

[8] T. Stelzer and W. F. Long, *HELAS: Automatic Generation of Tree Level Helicity Amplitudes*, Phys. Commun. **81**, 357-371 (1994)

[9] S. Richter *et al.*, *A Neural Network b Tagger for Single-Top Analyses*, CDF 7816 (2005)

[10] L. Demortier L. Lyons, *Everything you always wanted to know about pulls*, CDF Note 5776 (2002)

[11] Y.-K. Kim, U.-K. Yang, *Initial state gluon radiation studies on Drell-Yan data for top-pair production in hadron Collider*, CDF Note 6804 (2003)

[12] F. James, *MINUIT, Function Minimization and Error Analysis Reference Manual*, CERN Program Library Long Writeup, D506 (1994)

[13] A.C. Genz and A.A. Malik, *Remarks on algorithm 006: An adaptive algorithm for numerical integration over an N-dimensional rectangular region*, J. Comput. Appl. Math. 6 (1980) 295-302.
A. van Doren and L. de Ridder, *An adaptive algorithm for numerical integration over an n-dimensional cube*, J.Comput. Appl. Math. 2 (1976) 207-217.

[14] T. Hahn, *CUBA—A library for multidimensional numberical integration*, Comput. Phys. Commun. 168 (2005) 78.

[15] L. Demortier, *Bayesian treatments of Systematic Uncertainties*, Proceedings of Advanced Statistical Techniques in Particle Physics, Grey College, Durham, 18 - 22 March 2002,
http://www.ippp.dur.ac.uk/Workshops/02/statistics/proceedings.shtml

[16] C. Ciobanu, T. Junk, T. Müller, P. Savard, B. Stelzer, W. Wagner, T. Walter, *Likelihood Function for Single Top Search with 162 pb^{-1}*, CDF Note 7106 (2004)

[17] Sarah Budd, Matthias Buehler, Catalin Ciobanu, Peter Dong, Thomas Junk, Jan Lueck, Thomas Muller, Svenja Richter, Bernd Stelzer, Wolfgang Wagner, Rainer Wallny *Mistag Model Used for Single-Top Summer 2006 Analyses*, CDF Note 8490 (2006)

[18] Sarah Budd, Matthias Buehler, Catalin Ciobanu, Peter Dong, Thomas Junk, Jan Lueck, Thomas Muller, Svenja Richter, Bernd Stelzer, Wolfgang Wagner, Rainer Wallny *Estimation and Modeling of Non-W Background for Single-Top Searches*, CDF Note 8489 (2006)

[19] Guimaraes da Costa, Sherman *Study of SecVtx Fake Rates with 1/fb*, CDF Note 8263 (2006) *Loose and Tight SecVtx Tag Matrices with 1/fb*, CDF Note 8264 (2006)

[20] L. Read, J.Phys G **28**, 2693 (2002) and T. Junk, Nucl. Instum. Meth. **434**, 435 (1999). See also P. Bock *et al.* (The LEP Collaborations), CERN-EP-98-046 (1998) and CERN-EP-2000-055 (2000).

[21] L. Demortier, L. Lyons, *Everything You Always Wanted to Know about Pulls*, CDF Note 5776 (2001) 4, 6 4, 6 4 4, 7 4 4, 58, 67, 68 7 7 26, 27 33 25 25 35 35 34 1, 58, 61, 68 68

[22] B.W. Harris *et. al.*, *Phys. Rev.* D **66**, 054024 (2002)
Z. Sullivan *Phys. Rev.* D **70**, 114012 (2004). 60

A Calculation of Phase Space for Single Top Events

For s-channel single top events[3], $qq' \to t\bar{b} \to b\bar{b}l\nu$, the phase space factor can be written as:

$$d\Phi_4(q_1+q_2;p_1,p_2,p_3,p_4) = \delta^4(q_1+q_2-\sum_{i=1}^{4} p_i)\prod_{i=1}^{4}\frac{d^3\vec{p}_i}{(2\pi)^3 2E_i} \tag{18}$$

where q_1 and q_2 are the four momenta of the initial quarks; p_1, p_2 are the four momenta of the b and $\bar{b}$ quark, respectively; and p_3, p_4 are the four momenta of the lepton and neutrino, respectively. It is convenient to change variables from momenta $(\vec{p}_1,\vec{p}_2,\vec{p}_3,\vec{p}_4)$ to $(\rho_1,\Omega_1,\rho_2,\Omega_2,\vec{p}_l,\vec{p}_{\nu,T},m_W)$ where $\rho_i = |\vec{p}_i|$ is the absolute momentum of the quarks, Ω_i are the quark solid angles, and m_W is the mass of the W boson. One way to perform the transformation is to use the recursive character of the phase space [6]:

$$\begin{aligned}\prod_{i=1}^{4}\frac{d^3\vec{p}_i}{(2\pi)^3 2E_i} &= \delta(p_W^2-m_W^2)dm_W^2\prod_{i=1}^{4}\frac{d^3\vec{p}_i}{(2\pi)^3 2E_i} \\ &= \frac{dm_W^2}{|\frac{\partial p_W^2}{\partial p_{\nu,z}}|}\frac{d^3\vec{p}_l}{(2\pi)^3 2E_l}\frac{d\vec{p}_{\nu,T}}{(2\pi)^3 2E_\nu}\prod_{i=1}^{2}\frac{\rho_i^2 d\rho_i d\Omega_i}{(2\pi)^3 2E_i}\end{aligned} \tag{19}$$

In the last step, the δ-function was integrated with respect to $p_{\nu,z}$ using:

$$\int f(x)\delta[g(x)]dx = \frac{f(a)}{|g'(a)|} \quad ,at\ g(a)=0 \tag{20}$$

The partial derivative becomes (neglecting neutrino and lepton masses):

$$\begin{aligned}\frac{\partial p_W^2}{\partial p_{\nu,z}} &= \frac{\partial(p_l+p_\nu)^2}{\partial p_{\nu,z}} \\ &= \frac{\partial}{\partial p_{\nu,z}}(m_l^2+m_\nu^2+2E_lE_\nu-2p_{l,z}p_{\nu,z}-2\vec{p}_{l,T}\vec{p}_{\nu,T}) \\ &= 2E_l\frac{p_{\nu,z}}{E_\nu}-2p_{\nu,z}\end{aligned} \tag{21}$$

Finally, we substitute Equation 19 into Equation 18 and integrate two δ-functions with respect to the transverse momentum of the neutrino $\vec{p}_{\nu,T}$. The remaining two δ-functions are integrated with respect to the initial quark's longitudinal momentum and energy in the event probability. The expression for the phase space for single top events is:

$$\begin{aligned}d\Phi_4 &= \delta(E_{q_1}+E_{q_2}-\sum_{i=1}^{4}E_i)\ \delta(p_{q_1,z}+p_{q_2,z}-\sum_{i=1}^{4}p_{i,z}) \\ &\times\frac{dm_W^2}{|2E_l\frac{p_{\nu,z}}{E_\nu}-2p_{\nu,z}|}\frac{d^3\vec{p}_l}{(2\pi)^3 2E_l}\frac{1}{(2\pi)^3 2E_\nu}\prod_{i=1}^{2}\frac{\rho_i^2 d\rho_i d\Omega_i}{(2\pi)^3 2E_i}\end{aligned} \tag{22}$$

[3] and similar for other W + 2 jet topologies

B Comparison of profiling and marginalizing likelihoods

The Bayesian prescription for dealing with nuisance parameters (systematic uncertainties, in this case) is to integrate them out, a procedure often referred to as "marginalization" [6]. However, it is often more convenient to fit the nuisance parameters instead, usually referred to "profiling.". Profiling has two major advantages: it is faster to run and simpler to implement (using the well-understood MINUIT package), and it gives best-fit values for not only the parameter of interest (cross section) but also for all the nuisance parameters, which provides interesting extra information.

However, profiling is only an approximation to the proper Bayesian method, and its validity as an approximation relies upon certain criteria. In particular, it requires the shape of the reduced likelihood function for all nuisance parameters to be smooth with single peaks, because it uses the best fit value and error in place of a full integration. A reduced likelihood shape, that has two sharp peaks will not be treated properly by the profiling approximation.

Because of its advantages in implementation and speed, we used profiling in past analyses and planned to use it in this one. However, in this analysis, we found that several of our reduced likelihood distributions had a double-peak structure: the most severe case we found came from the the jet energy scale nuisance parameter (Figure 56). Most likely, the choice of asymmetric rate and shape uncertainties are the cause of the observed double peak structures in the distribution of the reduced likelihood functions for some nuisance parameters. We concluded that the profile likelihood method was not a proper approximation of the full Bayesian method in this particular case and decided to switch to a fully marginalization method as provided by MCLIMIT [20].

We tested the coverage of the profiling method by fitting ten thousand pseudo-experiments thrown in the way discussed in section 3.0.1. We only use the systematic uncertainties from the background normalizations in this test because we know that asymmetric errors are untrustworthy in the profiling method. We plot the central fit values, fit error, and pull ((true value - fit value) / fit error) for each parameter, following the prescription for asymmetric errors and Gaussian constraints given in [21]. The results are plotted in Figure 57 and show that we have coverage in this method.

We verified that marginalization and profiling yielded the same result when only symmetric systematic uncertainties were present. These uncertainties are Gaussian by construction and thus should come out the same for both methods. Adding the jet energy scale systematic caused both answers to shift, but by different amounts, as expected, because profiling responds incorrectly to the systematic uncertainty.

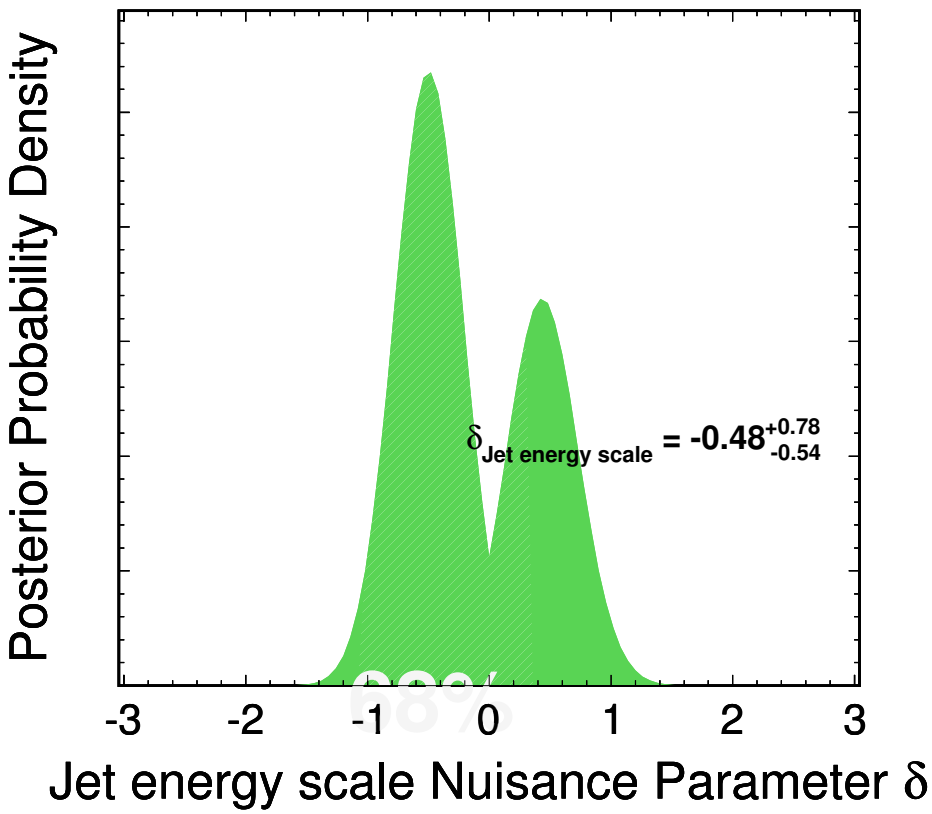


Figure 56: A profiled likelihood scan of the jet energy scale nuisance parameter. The double-peak structure indicates that profiling will be unable to properly handle this parameter and motivates the switch to marginalization.

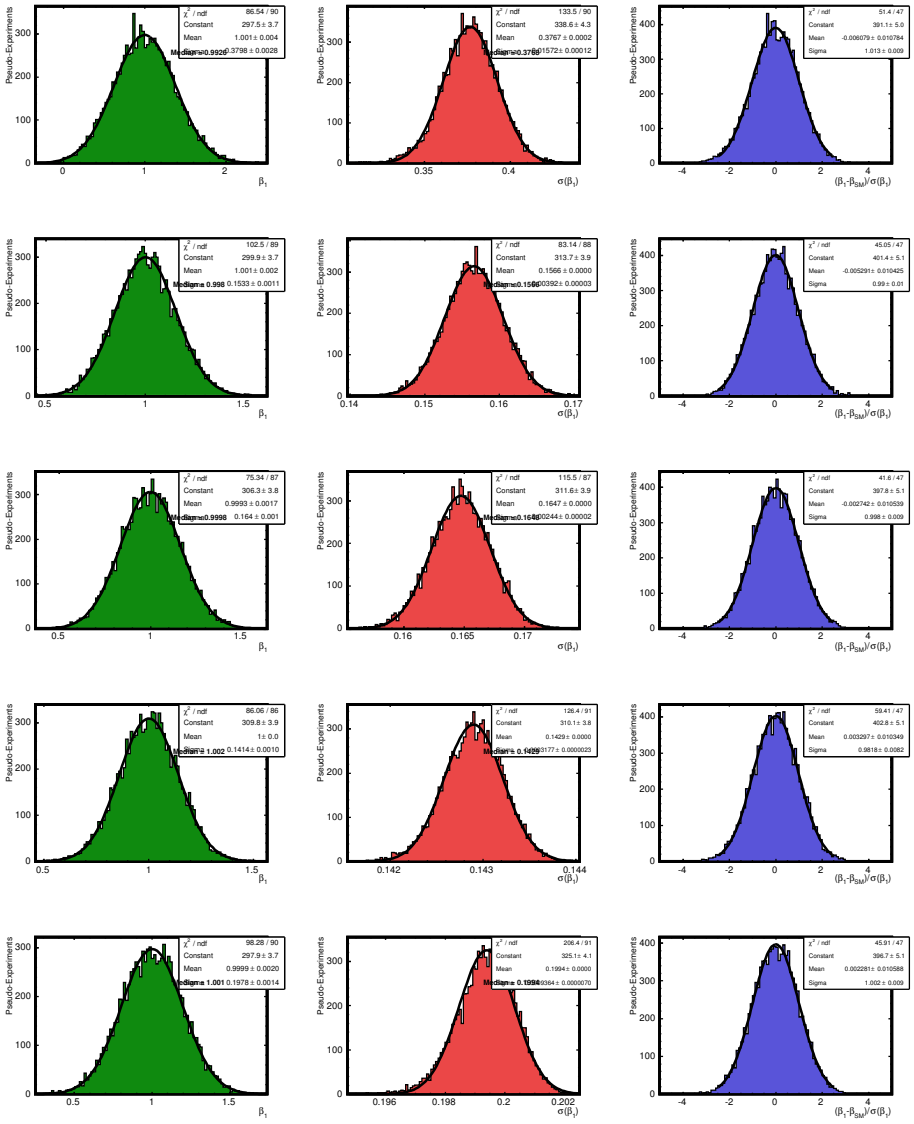


Figure 57: A profiled likelihood scan of the jet energy scale nuisance parameter. The double-peak structure indicates that profiling will be unable to properly handle this parameter and motivates the switch to marginalization.