DØ Note 5287
November 28, 2006
Version 1.2

Search for Single Top Quark Production Using The Matrix Element Analysis Technique in 1 fb^{-1} of Data

E. Aguiló,[2] P. Baringer[10] A. Bean,[10] C. Belanger-Champagne,[11] J.A. Benitez,[12] E.E. Boos,[13] R. Brock,[12] V. Bunichev, [13] K. Chan,[2] L. Christofek,[4] Y. Coadou,[17] L.V. Dudko,[13] M. Erdmann,[1] T. Gadfort,[18] A. García-Bellido,[18] C. Gerber,[9] D. Gillberg,[17] G. Gutierrez,[7] P. Gutierrez,[14] A.P. Heinson,[5] U. Heintz,[3] S. Herrin,[16] S. Jabeen,[3] S. Jain,[14] A. Juste,[7] S. Kappler,[1] D. Kau,[8] G. Kertzscher,[11] M. Kirsch,[1] L. Li,[5] J. Mitrevski,[6] R. Moore,[2] M. Narain,[3,4] D. O'Neil,[17] M. Pangilinan,[3] J. Parsons,[6] M. Perfilov,[13] C. Potter,[11] H.B. Prosper,[8] R. Schwienhorst,[12] E. Shabalina,[9] J. Steggemann,[1] T. Tim,[9] C. Tully,[15] M. Vetterli,[17] B. Vachon,[11] G. Watts,[18] M. Weber[7]

[1] *RWTH Aachen University*
[2] *University of Alberta*
[3] *Boston University*
[4] *Brown University*
[5] *University of California, Riverside*
[6] *Columbia University*
[7] *Fermi National Accelerator Laboratory*
[8] *Florida State University*
[9] *University of Illinois, Chicago*
[10] *University of Kansas*
[11] *McGill University*
[12] *Michigan State University*
[13] *Moscow State University*
[14] *University of Oklahoma*
[15] *Princeton University*
[16] *Rice University*
[17] *Simon Fraser University*
[18] *University of Washington*

This note describes the application of the matrix element analysis technique to the search for single top quark production at DØ using nearly 1 fb^{-1} of Run II data. From a comparison of the matrix element discriminants between data and the background model, we measure the single top quark production cross section:

$$\sigma\left(p\bar{p} \to tb + tqb + X\right) = 4.6^{+1.8}_{-1.5} \text{ pb}$$

Contents

1. INTRODUCTION

The search for single top quark production at the Tevatron has turned out to be a significantly more difficult task than anybody anticipated. Despite the large production cross section predicted by the standard model (about half of that for top quark pairs) and the distinct event signature involving a leptonic W boson decay and two or more jets (at least one being a b jet), this top quark production mode has not yet been observed. The main reason is that signal events are overwhelmed by a background from W+jets orders of magnitude larger. At jet multiplicities above two, even $t\bar{t}$ becomes a significant background. Both backgrounds can mimic the signal event characteristics in such a way that no cut on a single distribution can be used to effectively remove the background while preserving a large enough signal fraction. However, the combined information from several discriminant variables in a multivariate approach can potentially achieve enough sensitivity to unambiguously establish a signal. By now this is widely acknowledged as a requirement for any viable search for single top production, and all ongoing analyses at DØ are of a multivariate nature.

Methods such as neural networks or decision trees are being succesfully used at DØ to build efficient event discriminants to separate signal from background. These methods rely on identifying an optimal (and usually very large) set of kinematic and topological variables that collectively encode most of the available information to classify signal and background events. These methods are "learning machines" since they adjust to minimize the misclassification rate in samples of pseudo-data containing pure signal and background events that are presented to them. In this note we describe the so-called "Matrix Element-based Single Top Search" which, as will be explained below, is intrinsically different in approach.

First of all, the matrix element (ME) method attempts to maximize the sensitivity to the signal by making exhaustive use of the available kinematic information in the event, as contained in the four-momenta for the reconstructed objects. Information regarding b tagging has also been incorporated in the current version of the analysis. This is the complete set of variables considered in this analysis. Then, starting from these four-momenta, this method uses the matrix elements for the different signal and background processes to numerically compute the event probability density for each hypothesis (signal and background). This method has been developed and successfully applied at DØ in the past for parameter estimation such as the top quark mass [1, 2] or the longitudinal W boson helicity fraction in top quark decays [3]. In this analysis, the computed event probability densities are used to build an optimal event discriminant between signal and background, representing the first application of this method to a search at DØ.

The analysis described in this note is based on the selected "lepton+jets" data samples described in Ref. [4]. The current analysis is restricted to two-jet and three-jet events, with at least one of the jets b tagged.

This note is organized as follows. Section 2 presents an overview of the method, from the calculation of the event probability density functions, to the construction of the different event discriminants. Section 3 contains comparisons between data and the background model for the different event discriminant distributions from control and signal samples. Section 4 presents results of the analysis run on a number of Monte Carlo ensembles. Section 5 is devoted to a discussion of the expected performance, both in terms of significance and cross section measurement. Section 6 presents the measured results from data, followed by Section 8 which shows characteristics of the events selected according to their values of the signal discriminants. Finally, Section 9 is devoted to a summary and conclusions.

2. SEPARATION OF SIGNAL FROM BACKGROUND USING MATRIX ELEMENTS

2.1. Method Overview

The search for single top quark production critically depends on being able to discriminate between signal-like and background-like events or, in other words, to perform the optimal test of hypothesis H_0: signal (S), against hypothesis H_1: background (B). In this case, given an event characterized by a set of variables $\vec{x}$, the most powerful test of hypothesis is given by the Neyman-Pearson test, which involves a cut on the ratio:

$$\ell(\vec{x}, H_0, H_1) = \frac{P(\vec{x}|H_0)}{P(\vec{x}|H_1)} = \frac{P_S(\vec{x})}{P_B(\vec{x})}, \tag{1}$$

where $P_S(\vec{x})$ and $P_B(\vec{x})$ stand, respectively, for the signal and background probability density functions of $\vec{x}$. The power of such a test is also maximal if the set of variables $\vec{x}$ is complete, in the sense that they contain all information that is relevant to discriminate between signal and background. Such optimal test of hypothesis can also be performed on the a-posteriori Bayesian probability (or Bayes posterior) of "S given $\vec{x}$ " defined as:

$$P(S|\vec{x}) = \frac{P_S(\vec{x})}{P_S(\vec{x}) + P_B(\vec{x})} = \frac{\ell(\vec{x}, S, B)}{\ell(\vec{x}, S, B) + 1}, \tag{2}$$

as it can trivially be expressed in terms of $\ell(\vec{x}, S, B)$.

Many classifiers used in high energy physics, e.g., neural networks, attempt to approximate $P(S|\vec{x})$, achieving in many cases very competitive performance. In such methods, significant effort is devoted at the beginning to selecting the optimal set of variables that have good discriminating power between signal and background. This is a time-consuming task, resulting in many instances in a (not necessarily complete) set of very sophisticated kinematic and/or topological variables, for which the neural network must try to learn all correlations in the multi-dimensional space.

The matrix-element-based single top quark search attempts to make use of all the available kinematic information in the event. Therefore, $\vec{x}$ represents the set of reconstructed four-momenta for all selected final state objects in the event. For instance, in the case of ℓ+2jets events, $\vec{x} = (\vec{p}_\ell, \vec{p}_{j1}, \vec{p}_{j2})$. The reconstructed $\not\!\!E_T$ is not explicitly used, as it results from imposing momentum conservation in the transverse plane, and thus it does not represent an independent observable. The method is general enough that information related to fragmentation or b tagging, can in principle also be incorporated. In fact, information about the latter is already being used in the current analysis. In order to achieve optimal signal-to-background discrimination, given $\vec{x}$, an event discriminant is defined as:

$$D_S(\vec{x}) = P(S|\vec{x}) = \frac{P_S(\vec{x})}{P_S(\vec{x}) + P_B(\vec{x})} \tag{3}$$

where the signal hypothesis S can be: s-channel tb, t-channel tqb, or inclusive tb+tqb single top quark production. The signal and background probability density functions as a function of $\vec{x}$ are computed numerically based on the normalized differential cross section for signal and background processes, respectively. Since the differential cross section for the process of interest is proportional to the matrix element squared, we call this method the "Matrix Element (ME) Method." We shall refer to Eq. 3 as the "ME discriminant."

The ME discriminants are evaluated for data events, as well as Monte Carlo events for the different signal and background processes. Templates for the ME discriminant distributions corresponding to the background model are formed and compared to the same distribution in data, and a binned likelihood as a function of the signal cross section(s) is computed. A Bayesian method is then applied to compute the Bayes posterior for the "signal cross section(s) given the data" which, in case of an excess over the background expectation, can be used to estimate the single top production cross section(s).

2.2. Calculation of the Event Probability Density Functions

The ME event probability density function is defined as the properly normalized differential cross section for an event characterized by the reconstructed four-momenta $\vec{x}$:

$$P(\vec{x}) = \frac{1}{\sigma} \times \frac{\partial \sigma}{\partial \vec{x}}. \tag{4}$$

These quantities are explained in the following three sections.

2.2.1. Differential Cross Section

The differential cross section, $d\sigma(\vec{x})$ is given in Eq. 5; it is defined as the phase space integration of the differential cross section at the parton level, convoluted with a transfer function relating the reconstructed objects to the original partons.

$$d\sigma(\vec{x}) = \sum_{i,j} \int d\vec{y} \left[f_i(q_1, Q^2) dq_1 \times f_j(q_2, Q^2) dq_2 \times \frac{\partial \sigma_{hs,ij}(\vec{y})}{\partial \vec{y}} \times W(\vec{x}, \vec{y}) \times \Theta_{\text{Parton}}(\vec{y}) \right] \tag{5}$$

where

- $\sum_{i,j}$ is a sum of initial parton flavors in the hard scatter collision. For example, an s-channel collision can occur via $u\bar{d}$, $c\bar{s}$, $d\bar{u}$, or $s\bar{c}$ annihilation.
- $f_i(q, Q^2)$ is the parton distribution function for parton i carrying momentum q, evaluated at the factorization scale Q^2. For W+jets processes we use $Q^2 = M_W^2 + \sum_{jets}(M_i^2 + P_{i,T}^2)$ whereas for single top processes we use $Q^2 = m_t^2$. This analysis uses CTEQ6 [7] leading-order parton distribution functions via LHAPDF [8]
- $\partial \sigma_{hs,ij}(\vec{y})/\partial \vec{y}$ is the differential cross section for the hard scatter collision. This quantity is proportional to the square of the leading order matrix element [9] as given by:

$$d\sigma_{hs} = \frac{(2\pi)^4}{4\sqrt{(q_1 q_2)^2 - m_1^2 m_2^2}} |\mathcal{M}|^2 d\Phi_n(\vec{y}) \tag{6}$$

  where the first term is the flux factor, the second term is the matrix element squared, and the third term is the n-body phase space factor, with $n = 4(5)$ for two-jet (three-jet) events. Matrix elements in this analysis were obtained from the Madgraph [10] leading-order matrix-element generator.

- $W(\vec{x}, \vec{y})$ is called the transfer function, which represents the conditional probability of the observed state in the detector ($\vec{x}$) given the original partons ($\vec{y}$). The transfer functions are determined using Monte Carlo where the final-state flavor and four-vector

is known. A transfer function is determined for each jet flavor and for several regions of the calorimeter. The functions used in this analysis are the same ones derived and being used in the top mass analysis [11]. An example plot of the distribution of the difference between the jet energy and its parent parton energy, for three types of jets (light, soft-muon-vetoed b-jet and soft-muon-tagged b jet), is shown in Fig. 1.

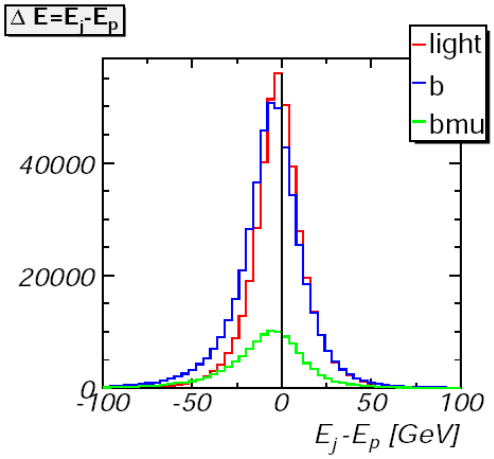

FIG. 1: Energy difference between a jet and its matched parton for three types of jets for all eta regions and all jet energies.

The electron transfer function is described as a Gaussian distribution with sigma dependent on the electron energy and pseudorapidity [12]. The muon transfer functions were determined for muons with and without an SMT hit and are parametrized as a function of $1/p_T$. The functions used in this analysis are again the same as the ones used in the top mass analysis [13, 14].

- $\Theta_{\rm Parton}(\vec{y})$ represents the parton level cuts applied in order to avoid singularities in the matrix element evaluation. These cuts are looser than the corresponding cuts at the reconstructed level. All cross sections were calculated with the following parton level cuts:
  - Parton isolation: $\Delta{\rm R}(q_i,q_j) > 0.5$
  - Minimum parton P_T: $P_T(q_i) > 6$ GeV
  - Maximum parton pseudorapidity: $|\eta(q_i)| < 3.5$
  - No cuts applied to the lepton or neutrino
- $\int d\vec{y} dq_1 dq_2$ is an integration over the matrix element phase space. The matrix element phase space for two parton final state events is defined by the 14 independent spatial degrees of freedom for the lepton, neutrino, and two partons as shown in Eq. 7

$$d\vec{y}_2 = dq_1 dq_2 d|p|_\ell d\Omega_\ell d|p|_\nu d\Omega_\nu d|p|_{q1} d\Omega_{q1} d|p|_{q2} d\Omega_{q2} \tag{7}$$

Events with three partons in the final state have 17 independent degrees of freedom and has a phase space defined in Eq. 8.

$$d\vec{y}_3 = dq_1 dq_2 d|p|_\ell d\Omega_\ell d|p|_\nu d\Omega_\nu d|p|_{q1} d\Omega_{q1} d|p|_{q2} d\Omega_{q2} d|p|_{q3} d\Omega_{q3} \tag{8}$$

Four (six) degrees of freedom are removed from the integration for two (three) parton event by assuming equal angles for partons and jets. This approximation has been

verified in Monte Carlo. Two more degrees of freedom are removed by assuming well measured lepton angles. Four more degrees of freedom are removed from the integration by energy-momentum conservation, leaving four integration variables. The final integration phase space is then transformed to suit the matrix element being integrated. W+jets matrix element integrations use the phase space defined in Eqs. 9 and single top matrix element integrations use the phase space in Eq. 10. The phase space for events with three jets is shown in Eq. 11 and 12.

$$d\vec{y}_{\mathrm{W+jets-2jets}} = dr_W d|p_{q1}|d|p_{q2}|dp_z^{system} \tag{9}$$

$$d\vec{y}_{\mathrm{singletop-2jets}} = dr_{\mathrm{top}} dr_W d|p_{q2}|dp_z^{system} \tag{10}$$

$$d\vec{y}_{\mathrm{W+jets-3jets}} = dr_W d|p_{q1}|d|p_{q2}|d|p_{q3}|dp_z^{system} \tag{11}$$

$$d\vec{y}_{\mathrm{singletop-3jets}} = dr_{\mathrm{top}} dr_W d|p_{q2}|d|p_{q3}|dp_z^{system} \tag{12}$$

where dr_W and dr_{top} are used to uniformly sample a Breit-Wigner distribution from the square of the invariant mass distribution for W boson and top quark production, respectively. This choice of sampling minimizes integration time because the matrix elements for these processes are negligable in regions where the invariant mass of the lepton and neutrino is far from the W mass and similarly for top events where the mass of the lepton, neutrino, and b-quark is far from the top mass. More information regarding the choice of variables and the multidimensional integration can be found in Appendix A. The multidimensional integrals were calculated using the GNU [5] Scientific Library's version of the VEGAS Monte Carlo integration technique [6].

2.2.2. Matrix Element Processes

For events with exactly two jets, we consider a total of five matrix elements for 2→4 processes (top quark and W bosons are treated off-shell): s-channel single top (ud→tb), t-channel single top (ub→td), Wbb production (ud→$Wb\bar{b}$), Wcg production (sg→Wcg), and Wgg production (ud→Wgg). "tqb" is used sometimes when referring to t-channel events because the main diagram that describes these events is the 2→5 diagram with "tqb" in the final state. This analysis uses a 2→4 diagram with "tq" as the final state, as seen in Fig 2, because events with two jets require at most two quarks or gluons in the final state. The three W+jets matrix elements were chosen by running the Madgraph Monte Carlo generator and selecting those processes which gave the largest cross section after an overall b-tagging efficiency was applied. The leading order diagrams for these channels (displayed as 2→3 processes for simplicity) are shown in Fig. 2.

For events with three jets, we consider a total of three matrix elements: s-channel single top (ud→tbg), t-channel single top (ug→tbd), and Wbb production (ud→$Wb\bar{b}g$). The number of W+jets matrix elements considered has been limited due to the significant time required to compute the differential cross section in Eq. 5. As will be discussed in Section 2.3.1, this can only result in a reduced sensitivity but does not affect the validity of the analysis. The leading order diagrams for these channels (displayed as 2→3, 4 processes for simplicity) are shown in Fig. 3.

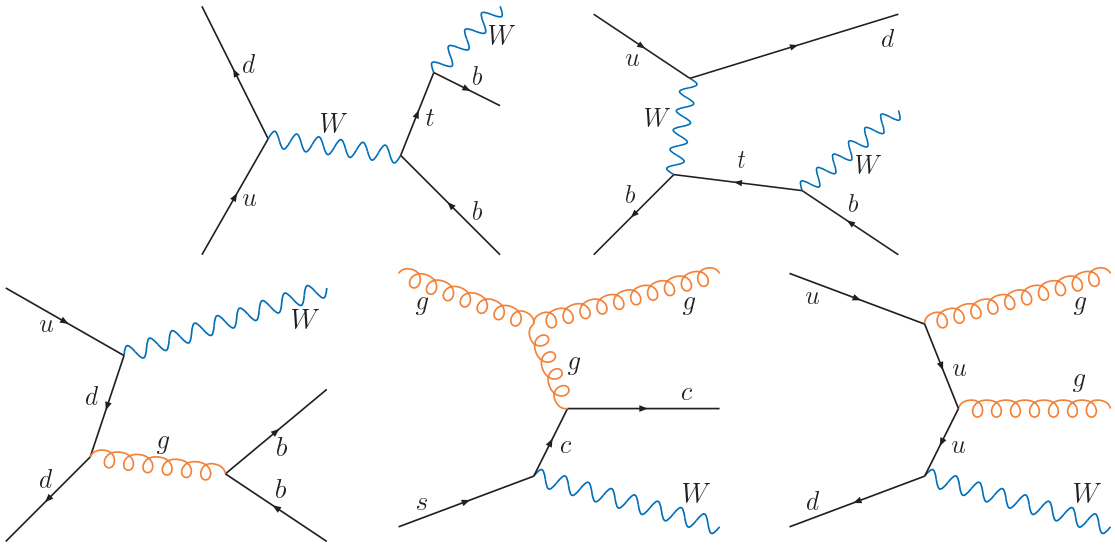

FIG. 2: Representative Feynman diagrams corresponding to the leading-order matrix elements used for event probability calculation for events with exactly two jets. Upper row, signals: $ud \to tb$, $ub \to td$; lower row, backgrounds: $ud \to Wbb$, $sg \to Wcg$, $ud \to Wgg$.

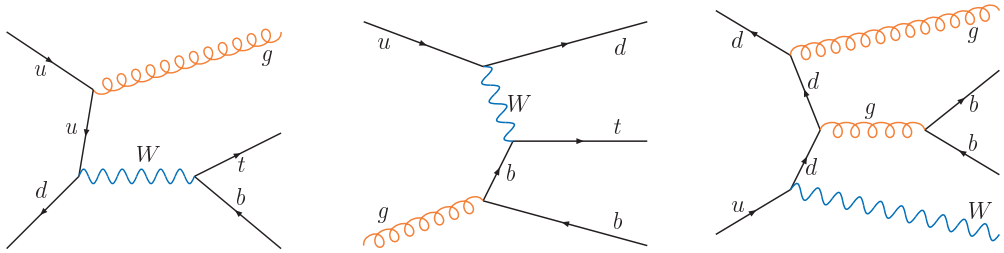

FIG. 3: Representative Feynman diagrams corresponding to the leading-order matrix elements used for event probability calculation for events with exactly three jets. Left two plots: signals, $ud \to tbg$, $ug \to tbd$; right plot: background,$ud \to Wbbg$.

2.2.3. Cross Section

The event probability density function defined in Eq. 4 requires a normalization constant to retain the probability density interpretation. The normalization constant σ is defined as the detector level phase space integration $\int d\vec{x}$ of the differential cross section defined in Eq. 5.

$$\sigma = \sum_{i,j} \int d\vec{x} d\vec{y} \left[\frac{\partial \sigma_{i,j}(\vec{y})}{\partial \vec{y}} \times W(\vec{x}, \vec{y}) \times \Theta_{\text{cuts}}(\vec{x}) \right] \tag{13}$$

Because all the events for which the probability density will be computed have had selection cuts applied, it is necessary to consider this in the calculation of the cross section. The term $\Theta_{\text{cuts}}(\vec{x})$ is included in the calculation to simulate the selection cuts. All cross sections were calculated with the following selection cuts:

- Lepton $P_T > 15$ GeV
- Electron (muon) $|\eta| < 1.1(2.0)$
- Missing $E_T > 15$ GeV

- Leading jet $P_T > 25$ GeV
- Leading jet $|\eta| < 2.5$
- Second jet $P_T > 20$ GeV
- Second jet $|\eta| < 3.5$
- Third jet $P_T > 20$ GeV (if three-jet event)
- Third jet $|\eta| < 3.5$ (if three-jet event)

The selection cuts shown above are slighty different from the canonical single top cuts. These cuts are included in the normalization calculation to approximate the relative acceptance difference between signal and background events. The cross sections computed for each signal and background process for two- and three-jet events are summarized in Table 1. In all instances, the statistical uncertainty from the Monte Carlo integration is below 1%.

	Cross Section × Branching Fraction [fb]							
	2-jet events				3-jet events			
	1 tag		2 tags		1 tag		2 tags	
	Electron	Muon	Electron	Muon	Electron	Muon	Electron	Muon
Signals								
$tb(g)$	8.07	10.4	6.90	8.90	6.02	7.64	5.22	6.66
$tq(b)$	19.6	26.8	0.27	0.38	6.34	8.56	5.40	7.40
Backgrounds								
$Wbb(g)$	29.5	41.9	24.6	34.7	6.34	8.56	5.40	7.40
Wcg	36.4	54.0	0.33	0.61				
Wgg	52.3	74.5	0.33	0.47				

TABLE 1: Cross section times branching fraction for each analysis channel.

2.2.4. Treatment of Combinatorial Background

The event probability density shown in Eq. 4 assumes a known assignment between a jet and parton from the matrix element. In practice, this assignment is not a-priori known and we must sum over all possible assignments. In general, the event differential cross section is modified as shown in Eq. 14:

$$\begin{aligned} d\sigma(\ell, j_1, j_2) &= \alpha_{j_1 \to p_1} \alpha_{j_2 \to p_2} d\sigma(\ell, j_1 \to p_1, j_2 \to p_2) + \\ &+ \alpha_{j_2 \to p_1} \alpha_{j_1 \to p_2} d\sigma(\ell, j_2 \to p_1, j_1 \to p_2) \end{aligned} \tag{14}$$

where the α parameters relate to the probability of the jet-parton match. If there is no a-priori knowledge of the correct assignment, these quantities can be made equal. Thereby no preference is given to either assignment.

This analysis uses information from the neural network b tagger [15] to weight the different jet-parton combinations depending on whether a given jet is tagged or not and which parton flavor is being assigned to it when summing over the combinatorial background (see Eq. 14). Therefore the α weights are related to the jet tag-rate functions for the different hypothesized jet flavors (b, c and light), as shown in Table 2. The different jet tag-rate functions for each flavor have been provided by the B-ID group parametrized in terms of jet P_T and η: $\varepsilon_{\text{flavor}}(j) = \varepsilon_{\text{flavor}}(P_T(j), \eta(j))$.

Jet-Parton Matching Weight Factors		
Parton flavor	b tagged	Not tagged
b	ε_b	$1-\varepsilon_b$
c	ε_c	$1-\varepsilon_c$
light	ε_l	$1-\varepsilon_l$

TABLE 2: Weights for the event differential cross section depending on the b-tagging status of the jet and jet-parton assignment.

For example, let us assume we have a given two-jet event with j_1 tagged and j_2 not tagged, and we are trying to evaluate the probability density function for the Wcg hypothesis, summing over the two possible jet-parton assignments. In this case, the differential cross section would be given by:

$$\begin{aligned} d\sigma_{Wcg}(\ell, j_1, j_2) &= \varepsilon_c(j_1)(1-\varepsilon_l(j_2))d\sigma_{Wcg}(\ell, j_1{\to}c, j_2{\to}g) + \\ &+ \varepsilon_l(j_2))(1-\varepsilon_c(j_1))d\sigma_{Wcg}(\ell, j_2{\to}c, j_1{\to}g). \end{aligned} \quad (15)$$

This differential cross section must then be integrated over the reconstructed lepton and jets four-momenta in order to compute the event probability density function.

2.3. Single Top Discriminant

2.3.1. Discriminant Definition

As discussed in Section 2.1, the signal and background probability density functions are used to build an event discriminant, defined in Eq. 3. In particular, it is possible to define discriminants specific to each of the single top production modes, as well as inclusive for single top production, assuming the SM ratio of s- and t-channel cross sections.

The channel-specific discriminants are defined as:

$$D_{tb(tqb)}(\vec{x}) = \frac{P_{tb(tqb)}(\vec{x})}{P_{tb(tqb)}(\vec{x}) + P_B(\vec{x})}. \quad (16)$$

where tqb here generically refers to the t-channel process, whether based on the tq matrix elements (two-jet events) or the tqb matrix elements (three-jet events).

For two-jet events, the background probability density function $P_B(\vec{x})$ is given by:

$$P_B(\vec{x}) = C_{Wbb}P_{Wbb}(\vec{x}) + C_{Wcg}P_{Wcg}(\vec{x}) + C_{Wgg}P_{Wgg}(\vec{x}), \quad (17)$$

where C_{Wbb}, C_{Wcg}, and C_{Wgg} are, in principle, the relative fractions of each background in the data. However, as discussed below, these background fractions are optimized for the analysis.

The background fractions for the two-jet discriminant were found by a grid search of each point (C_{Wbb}, C_{Wcg}, C_{Wgg}) and calculating the expected Bayes ratio defined as the maximum of the two-dimensional posterior divided by the value of the posterior at zero signal cross section. This procedure was used by each analysis in the single top group to optimize the final discriminant variable. The values of the background fractions are summarized in Table 3.

For the three-jet analysis, the background probability density function $P_B(\vec{x})$ is given by:

$$P_B(\vec{x}) = P_{Wbbg}(\vec{x}), \quad (18)$$

Optimized Background Fractions				
	1 tag		2 tags	
	Electron	Muon	Electron	Muon
C_{Wbb}	0.20	0.40	0.67	1
C_{Wcg}	0.40	0.40	0	0
C_{Wgg}	0.40	0.20	0.33	0

TABLE 3: Background fractions chosen for each analysis channel in two-jet events.

which represents a further step in simplification relative to the two-jet analysis.

At this point, a remark is in order. As can be seen, the background probability density function does not contain multijet or $t\bar{t}$ matrix elements. Even for the case of W+jets, some simplifications are made regarding which processes to consider and the relative contributions of each of them to the total background probability density. It should be stressed that the result of such approximations can only be reduced discrimination. The background model still contains all processes and the event discriminant is computed for all of them to build the templates that will be compared to data.

2.3.2. One-Dimensional Discriminants

This section contains overlayed plots of the one-dimensional (1D) tb and tq discriminants for signal and background events. The events in the plots come from the combined e,μ / 1,2 tags / 2,3 jets channel. Figures 4 and 5 show good discrimination between signal and W+jets or multijets backgrounds. However, the discrimination is poorer between signal and $t\bar{t}\to\ell\ell$ and $t\bar{t}\to\ell$+jets events as shown in Fig. 6. The lack of discrimination power for $t\bar{t}$ events is due to the fact that the current version of the analysis does not yet include a $t\bar{t}$ probability density function in the definition of the discriminant. The $t\bar{t}$ matrix element takes much longer to integrate because there are six partons in the final state while there are four in the single top and W+jets matrix elements. Adding a $t\bar{t}$ matrix element is envisioned as a future improvement for this analysis.

2.3.3. Two-Dimensional Discriminants

This analysis uses a two-dimensional (2D) discriminant computed from the 1D tb and tq discriminants discussed in the previous section. The 2D discriminant is more powerful than either 1D projection because it selects events with both tb and tq characteristics, which helps to further reduce the W+jets and $t\bar{t}$ background which may have either characteristic but not necessarily both.

Figures 8 and 9 show the 2D discriminants for single top quark signals and for all the backgrounds. The plots are normalized to unit volume.

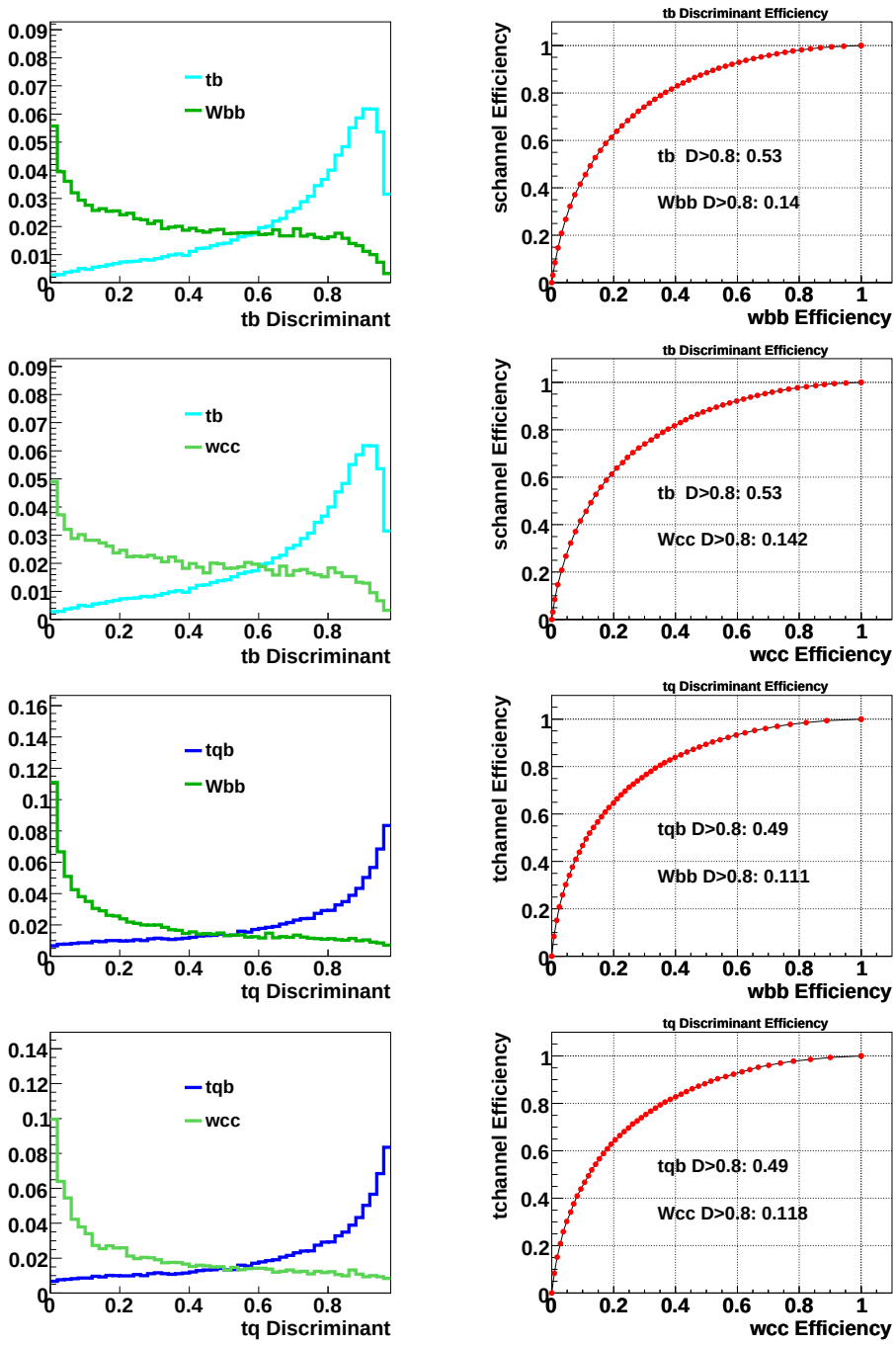

FIG. 4: Discriminant plots and efficiency curves for: first row, tb vs. Wbb, second row, tb vs. Wcc, third row, tq vs. Wbb, and fourth row, tq vs. Wcc. The numbers in the efficiency curves (right column) represent the fraction of signal or background the remains after a discriminant cut of 0.8.

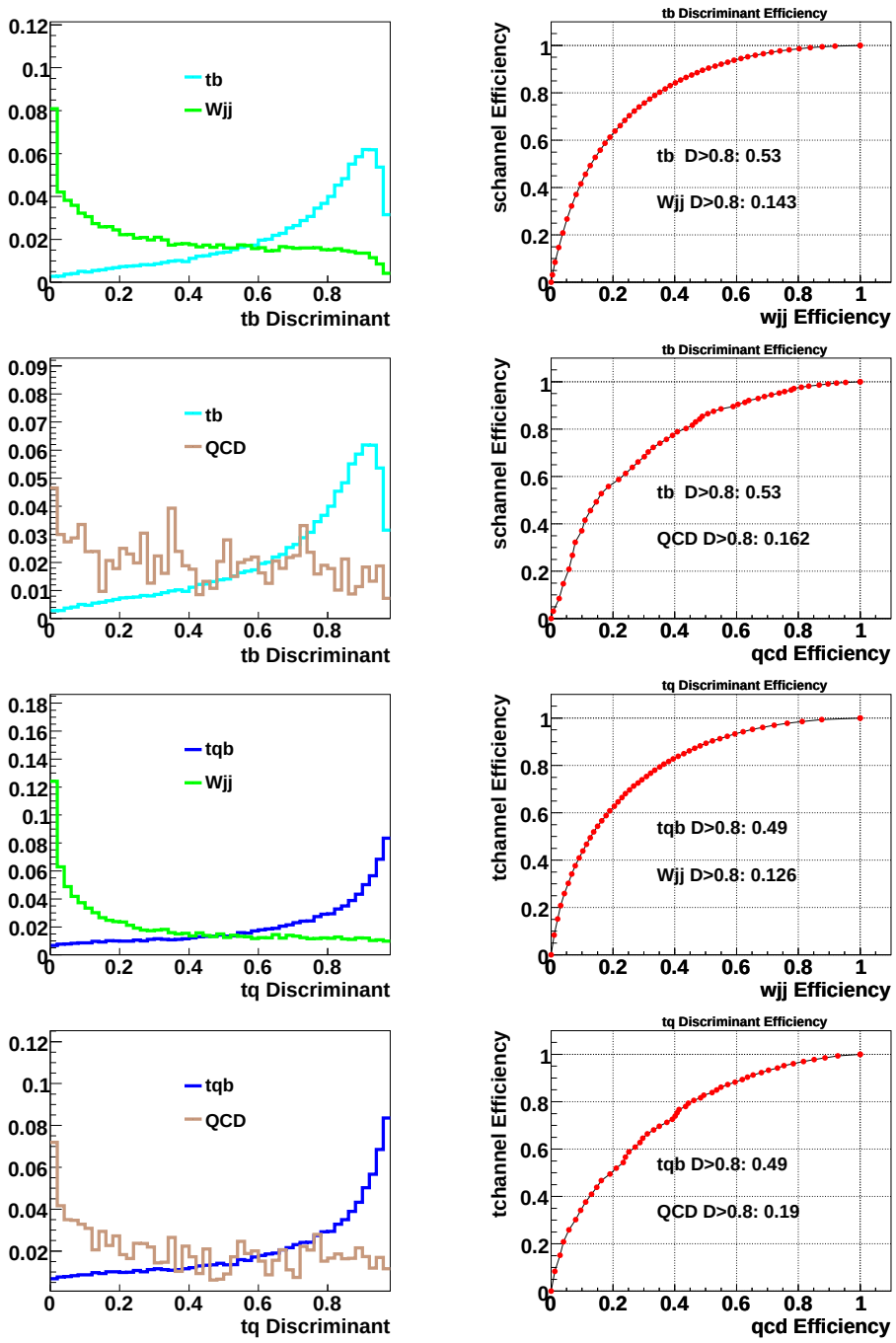

FIG. 5: Discriminant plots and efficiency curves for: first row, tb vs. Wjj, second row, tb vs. multijets, third row, tq vs. Wjj, and fourth row, tq vs. multijets. The numbers in the efficiency curves (right column) represent the fraction of signal or background the remains after a discriminant cut of 0.8.

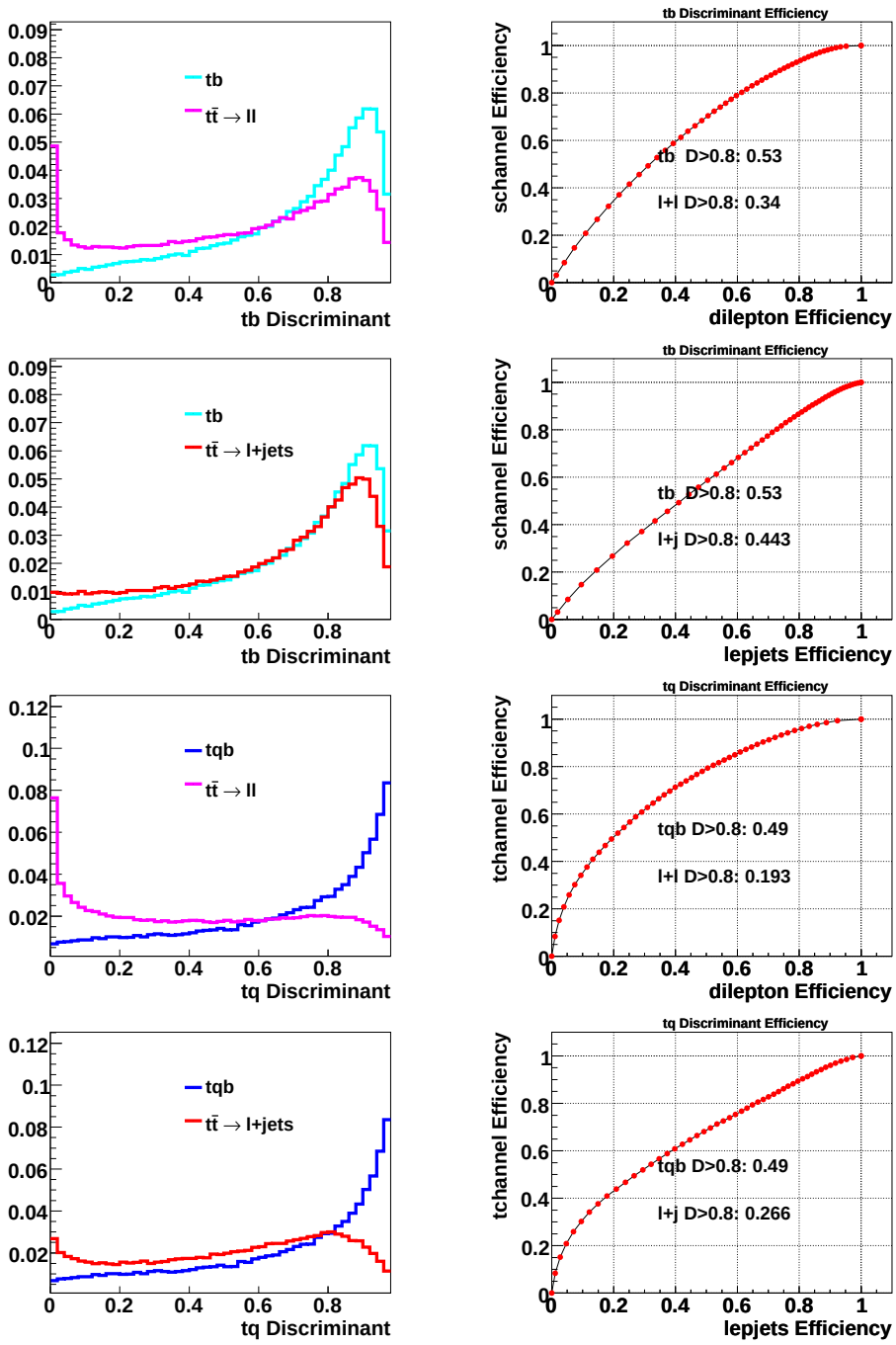

FIG. 6: Discriminant plots and efficiency curves for: first row, tb vs. $t\bar{t}\to\ell\ell$, second row, tb vs. $t\bar{t}\to\ell$+jets, third row, tq vs. $t\bar{t}\to\ell\ell$, and fourth row, tq vs. $t\bar{t}\to\ell$+jets. The numbers in the efficiency curves (right column) represent the fraction of signal or background the remains after a discriminant cut of 0.8.

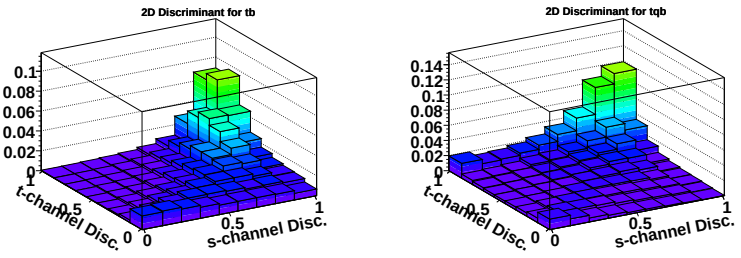

FIG. 7: 2D-discriminant templates for: left, tb, and right, tqb Monte Carlo events.

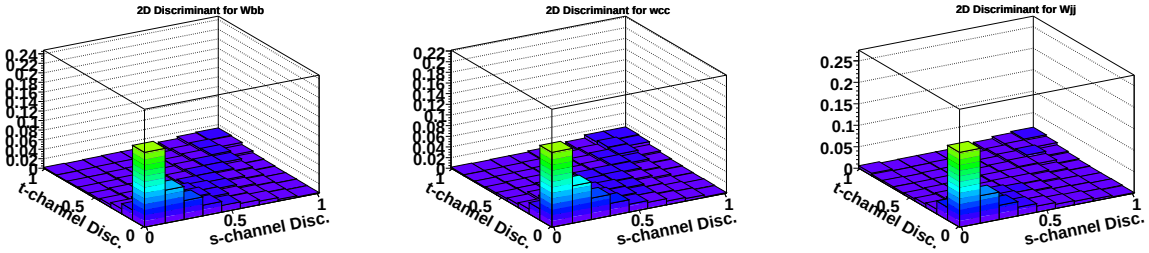

FIG. 8: 2D-discriminant templates for: left, Wbb, middle, Wcc, and right, Wjj Monte Carlo events.

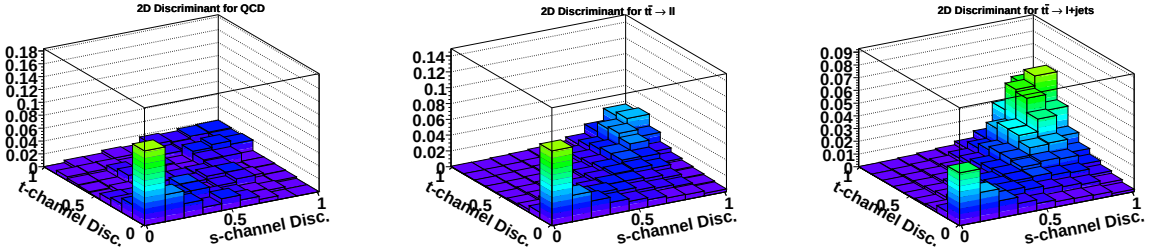

FIG. 9: 2D-discriminant templates for: left, multijets events, middle, $t\bar{t}\rightarrow\ell\ell$, and right, $t\bar{t}\rightarrow\ell$+jets Monte Carlo events.

3. CROSS-CHECK SAMPLES

A crucial step in the single top search is to establish that the background model is appropriate while minimizing looking into the search region. For this purpose, two background-dominated control samples are defined, and a comparison between the 1D discriminants in data and the background model is performed.

These two control samples are selected by applying the standard event selection, and requiring in addition $H_T < 175$ GeV and $H_T > 300$ GeV, respectively. Two-jet events are largely dominated by W+jets events, and thus it is the modeling of this background that is mainly addressed by these cross checks. We refer to these control samples as "soft W+jets" and "hard W+jets" respectively. In the case of three-jet events, the "hard W+jets" sample also contains a significant fraction of $t\bar{t}$.

The "soft W+jets" sample selects low momentum W+jets and multijets events and almost no top-quark events. Figures 10 and 11 compare the tb and tq discriminants between data and the background model for events with two and three jets respectively. Good agreement is seen between data and expectation.

The "hard W+jets" sample selects mainly $t\bar{t}$ and high momentum W+jets events. Figures 12 and 13 compare the tb and tq discriminants between data and the background model for events with two and three jets. Good agreement is again found between data and expectation.

Most of the signal, and thus the problematic W+jets background, has H_T between these cuts. Therefore, by confirming that the observed discriminant distribution is well reproduced by the background model for the softest and hardest W+jets events, we gain confidence that the W+jets background in the signal region is also well modeled.

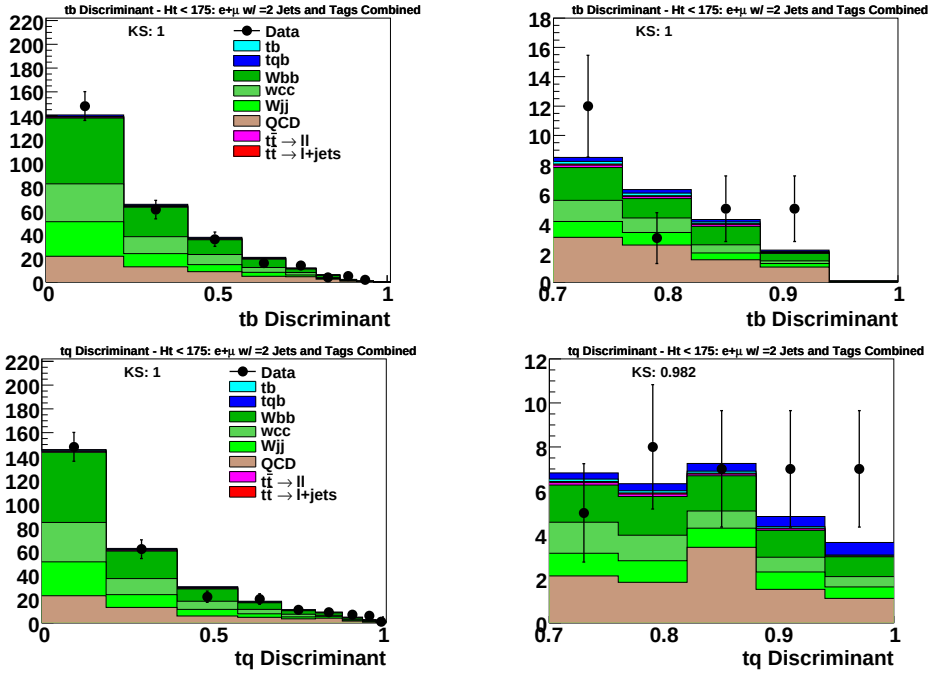

FIG. 10: "Soft W+jets" cross-check plots in two-jet events for the tb discriminant (upper row) and the tq discriminant (lower row). The left column shows the full discriminant region while the right column shows the high discriminant region above 0.7.

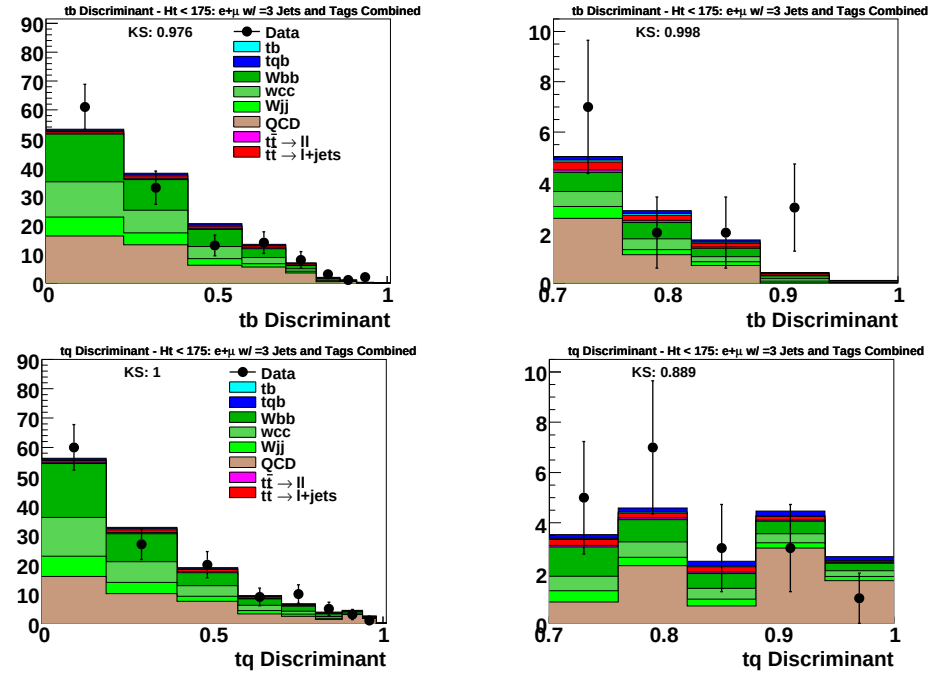

FIG. 11: "Soft W+jets" cross-check plots in three-jet events for the tb discriminant (upper row) and the tq discriminant (lower row). The left column shows the full discriminant region while the right column shows the high discriminant region above 0.7.

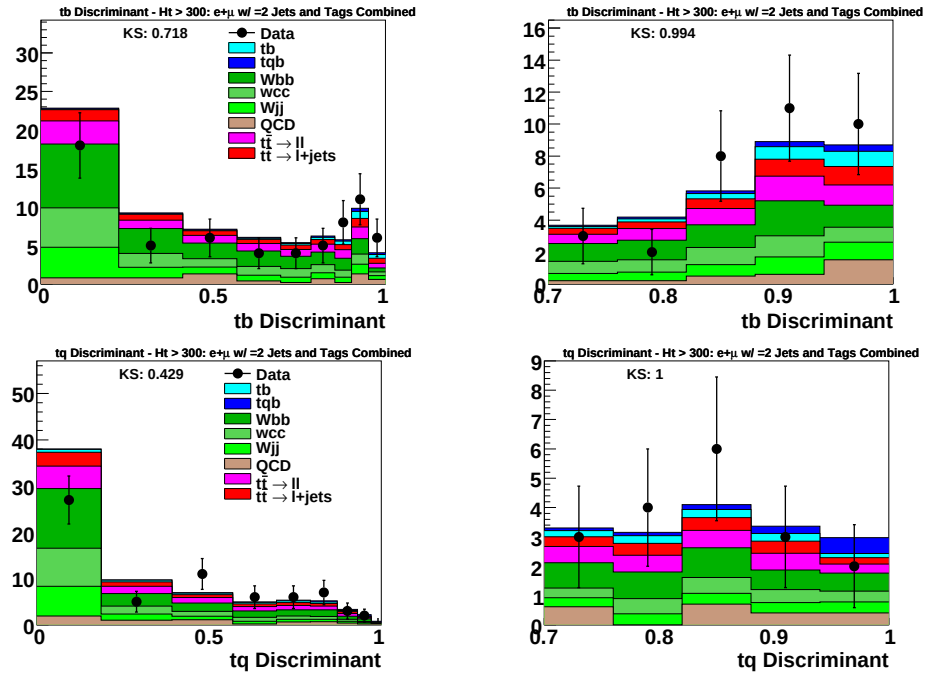

FIG. 12: "Hard W+jets" cross-check plots in two-jet events for the tb discriminant (upper row) and the tq discriminant (lower row). The left column shows the full discriminant region while the right column shows the high discriminant region above 0.7.

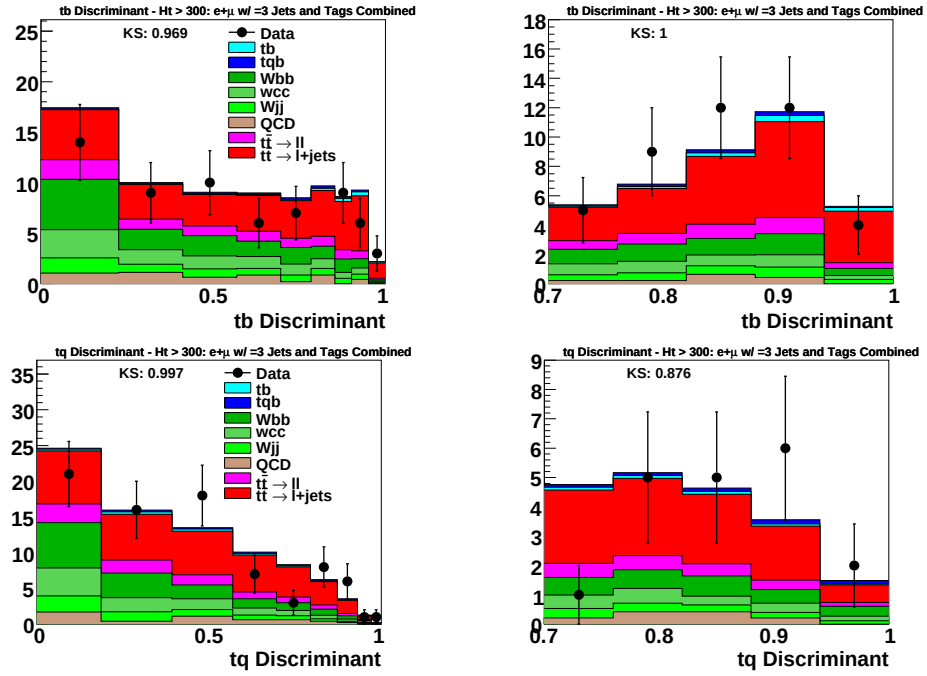

FIG. 13: "Hard W+jets" cross-check plots in three-jet events for the tb discriminant (upper row) and the tq discriminant (lower row). The left column shows the full discriminant region while the right column shows the high discriminant region above 0.7.

4. ENSEMBLE TESTS

Several ensembles of pseudo-data events were produced from the background model to test the performance of the analysis. These ensembles include:

1. 400 fake data samples with SM single top content ($\sigma_{tb+tqb} = 2.86$ pb).
2. Four sets of 100 samples each (named A, B, C and D), with unknown (to the analyzers) single top content, but with SM ratios between the tb and tqb cross sections.
3. Two sets of 200 samples with the SM ratio between the tb and tqb cross sections, and with a total single top cross section of $\sigma_{tb+tqb} = 4.5$ pb and $\sigma_{tb+tqb} = 4.7$ pb.

The full analysis chain was run using each pseudo-dataset as if it were the real dataset. The results are presented in the following subsections.

4.1. Ensemble Tests with SM Signal

Figure 14 shows the distribution of estimated cross sections from one ensemble generated with the SM input signal cross sections for the electron, muon, and e+μ channel. The ensemble contains 400 samples of both 2-jet and 3-jet events. The input cross section in each case is 2.86 pb. The most probable value of the distribution is 2.75 pb while the mean is 3.18 pb.

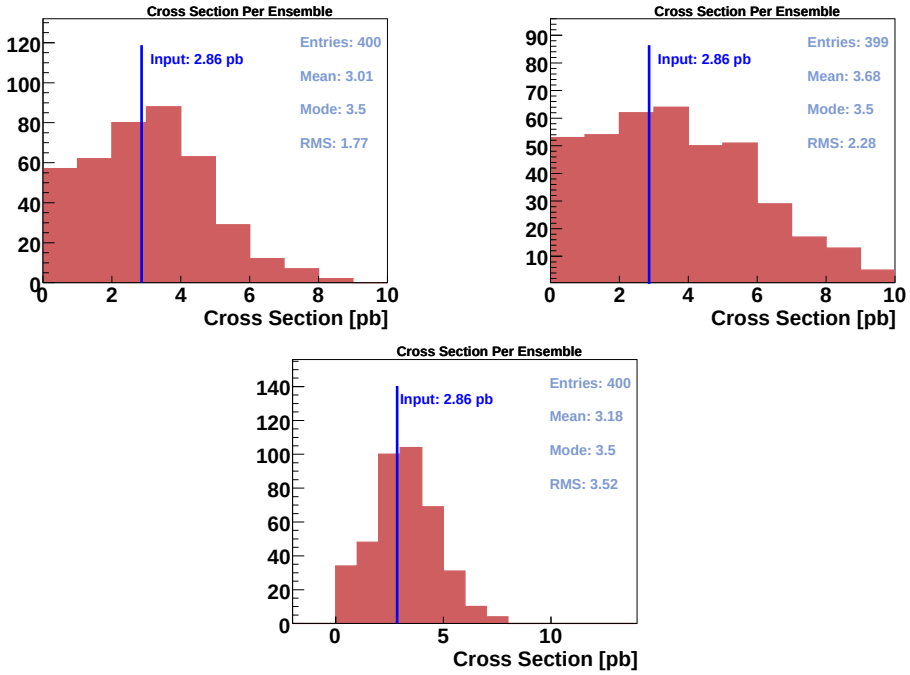

FIG. 14: Results of standard model ensemble test for the electron channel (upper left plot), the muon channel (upper right plot), and electrons and muons combined (lower plot).

4.2. Ensemble Tests with Non-SM Signal

Five more ensembles were generated with a non-SM tb+tqb cross section, but SM tb:tqb cross section ratio ($\sigma_s/\sigma_t = 0.44$). The results are shown in Fig. 15.

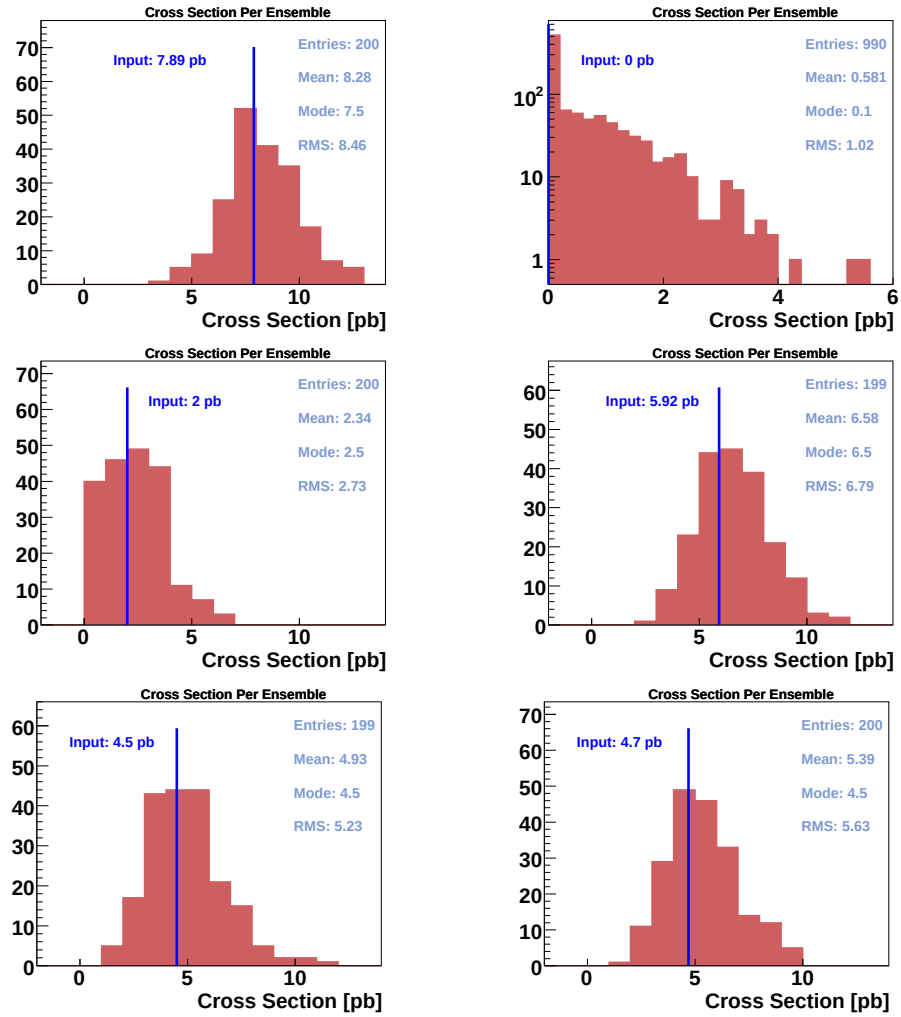

FIG. 15: Results using the ensembles with non-SM cross section but SM tb:tqb ratio. The upper row contains ensembles A and B, the middle row shows ensembles C and D, the bottom row shows the results from two ensembles with input cross sections close to our measured cross section of 4.6 pb.

Our measured values of the cross sections taken from the means of the ensembles are:

- A: 3.00 × SM (true value = 2.76 × SM)
- B: 0.20 × SM (true value = 0)
- C: 0.86 × SM (true value = 0.70 × SM)
- D: 2.38 × SM (true value = 2.06 × SM)

We show these measured cross sections versus the input values in Fig. 16 as a calibration check. We fit a straight line through points greater than 1 pb where it shows good agreement with the measurements. Below that, the measurement is biased to produce a value greater than input because the results are constrained to be positive.

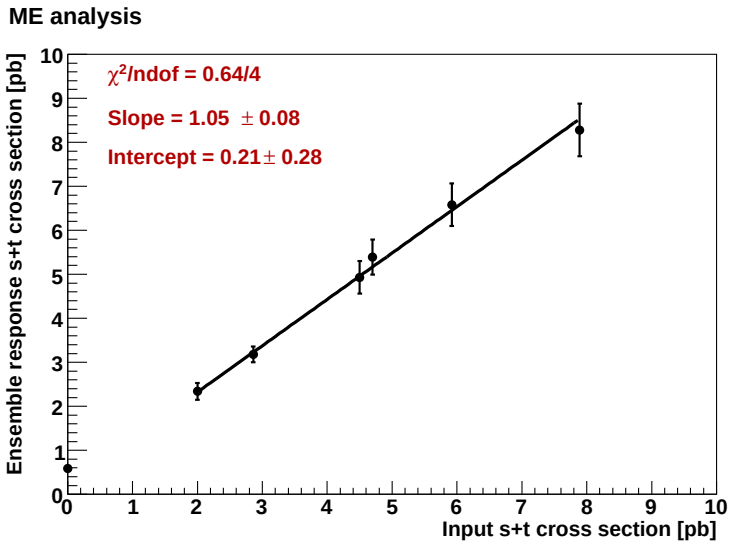

FIG. 16: Measured signal cross section versus input cross section in the ensemble tests. The ensemble response is obtained from the mean of the distributions in Figs. 14 and 15.

5. EXPECTED RESULTS

In this section we present results regarding the expected performance of the analysis. To obtain these results the number of observed events has been set equal to the expected signal, according to the standard model prediction, plus the expected background.

Figures. 17 and 18 show the resulting tb+tqb posterior for the combined e+$\mu \geq 1$ b-tag channel in two-jet and three-jet events. Figure 19 shows the tb+tqb posterior for the combination of all channels. The left figures correspond to the case of only statistical uncertainties considered, whereas the right figures also include systematic uncertainties.

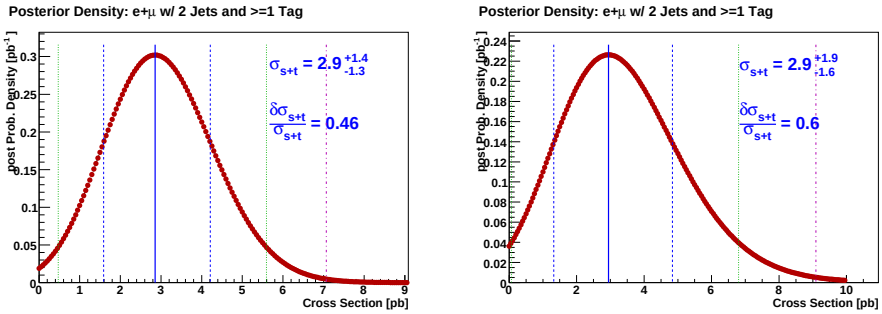

FIG. 17: Expected 1D posterior plots for the combined e+$\mu \geq 1$ b-tag channel in two-jet events, with statistical uncertainties only (left plot) and including also systematic uncertainties (right plot).

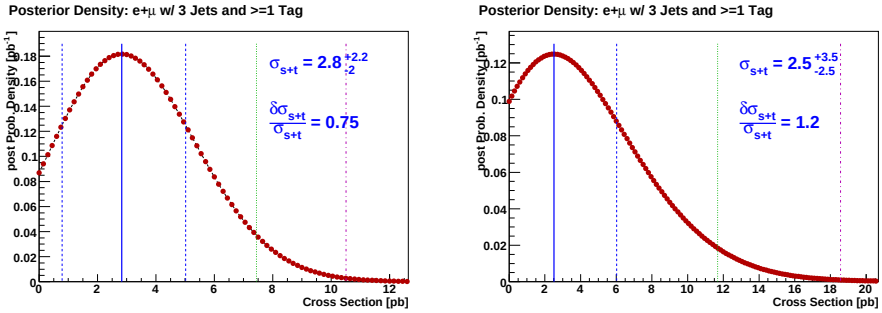

FIG. 18: Expected 1D posterior plots for the combined e+$\mu \geq 1$ b-tag channel in three-jet events, with statistical uncertainties only (left plot) and including also systematic uncertainties (right plot).

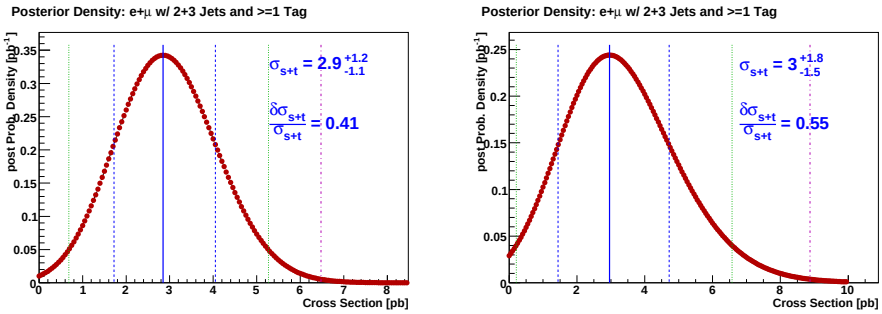

FIG. 19: Expected 1D posterior plots for the combination of all channels, with statistical uncertainties only (left plot) and including also systematic uncertainties (right plot).

5.1. Expected Signal Significance

We calculate the significance of the expected results using the *tb+tqb* posteriors presented in the previous section. The Bayes ratio (see Appendix 10 in Ref. [4]) is defined as the peak value of the posterior divided by the posterior at zero signal cross section. The larger the Bayes ratio, the larger the expected significance of the result. This measure has been used to optimize the analysis in the individual channels. Results are shown in Table 4.

Expected Bayes Ratios for the *tb+tqb* Signal							
	1,2tags + 2,3jets		e,μ + 2,3jets		e,μ + 1,2tags		All
	e-chan	μ-chan	1 tag	2 tags	2 jets	3 jets	channels
Statistics only	8.1	4.1	17.8	1.9	15.9	2.1	33.3
With systematics	3.8	2.6	6.3	1.4	6.2	1.3	**8.4**

TABLE 4: Expected Bayes ratios, without and with systematic uncertainties, for many combinations of analysis channels. The best value from all channels combined, with systematics, is shown in bold type.

5.2. Expected Cross Section

Table 5 shows the expected cross sections for various combinations of analysis channels. The expected result for each combination is consistent with the standard model cross section. Table 6 summarizes the relative uncertainty on the expected *tb+tqb* cross section measurement, defined as half the width of the *tb+tqb* posterior, divided by the cross section value at the posterior peak.

Expected *tb+tqb* Cross Section							
	1,2tags + 2,3jets		e,μ + 2,3jets		e,μ + 1,2tags		All
	e-chan	μ-chan	1 tag	2 tags	2 jets	3 jets	channels
Statistics only	$2.8^{+1.5}_{-1.4}$	$2.8^{+1.8}_{-1.7}$	$2.9^{+1.3}_{-1.2}$	$2.8^{+2.5}_{-2.2}$	$2.9^{+1.4}_{-1.3}$	$2.8^{+2.2}_{-2.1}$	$2.9^{+1.2}_{-1.1}$
With systematics	$3.0^{+2.2}_{-1.8}$	$3.1^{+2.5}_{-2.1}$	$2.9^{+1.8}_{-1.6}$	$2.7^{+3.4}_{-2.7}$	$2.9^{+1.9}_{-1.6}$	$2.5^{+3.5}_{-2.5}$	$\mathbf{3.0^{+1.8}_{-1.5}}$

TABLE 5: Expected *tb+tqb* cross sections, without and with systematic uncertainties, for many combinations of the analysis channels. The final expected result of this analysis are shown in the lower right hand corner in bold type.

Relative Uncertainties on the Expected *tb+tqb* Cross Section							
	1,2tags + 2,3jets		e,μ + 2,3jets		e,μ + 1,2tags		All
	e-chan	μ-chan	1 tag	2 tags	2 jets	3 jets	channels
Statistics only	52%	60%	45%	83%	46%	75%	41%
With systematics	67%	75%	59%	115%	60%	121%	**55**%

TABLE 6: Relative uncertainties on the expected *tb+tqb* cross section, without and with systematic uncertainties, for many combinations of the analysis channels. The best value from all channels combined, with systematics, is shown in bold type.

6. OBSERVED RESULTS

This section presents the matrix element results obtained using the $\sim$1 fb^{-1} data set.

Figures 20 and 21 show the tb and tqb discriminants for the combined e,μ / 1,2 tags events for two-jet and three-jet events where the data distributions may be compared to the background model. The SM prediction for single top quark production has been added to the background sum in the plots. The individual channel plots for the 1D discriminants are shown in Appendix B.

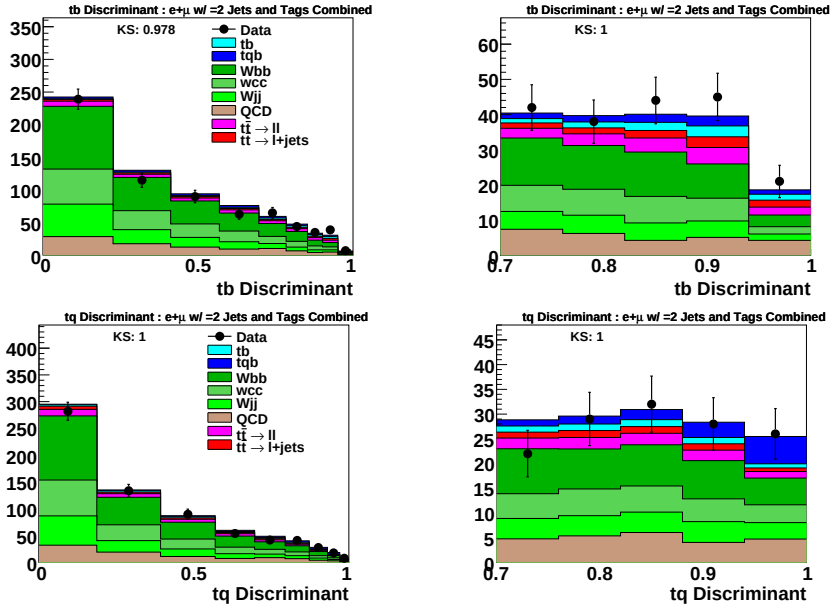

FIG. 20: Discriminant plots for the e+μ channel with two jets and $\geq$ 1 b tag. Upper row: tb discriminant; lower row: tq discriminant. Left column: full output range; right column: close-up of the high end of the distributions.

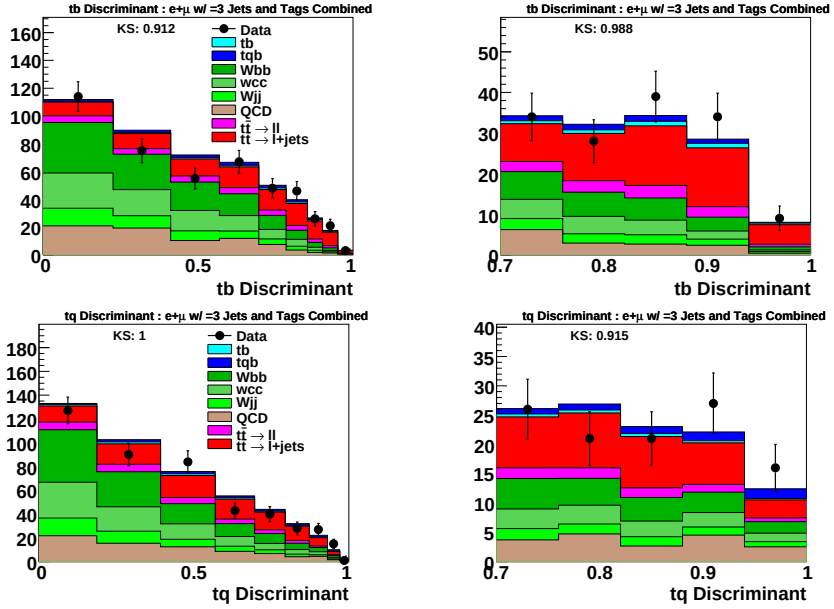

FIG. 21: Discriminant plots for the e+μ channel with three jets and $\geq$ 1 b tag. Upper row: tb discriminant; lower row: tq discriminant. Left column: full output range; right column: close-up of the high end of the distributions.

Figure 22 shows the 2D discriminant distribution in data for all analysis channels (e,μ / 1,2 tags), separately for two-jet and three-jet events. This is the typ of distribution we use for the single top quark cross section measurement, but we actually use each different analysis channel separately.

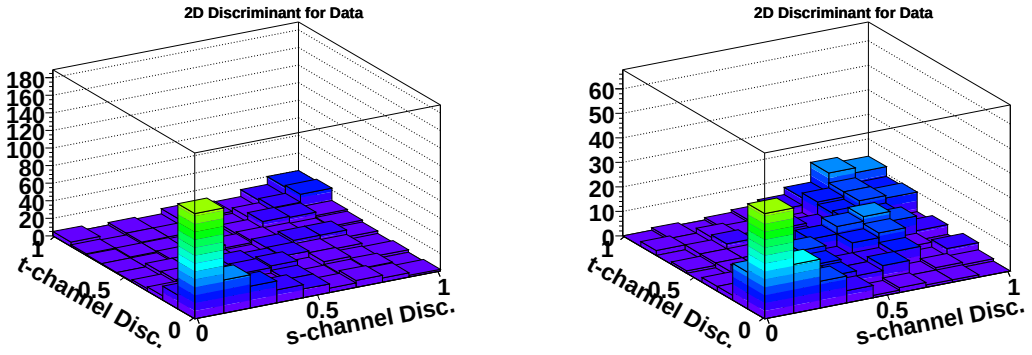

FIG. 22: 2D discriminant distribution in data (e,μ / 1,2 tags combined) for two-jet (left) and three-jet (right) events.

6.1. Measured Cross Section

Figure 23 shows the observed tb+tqb posterior without and with systematic uncertainties for all channels combined.

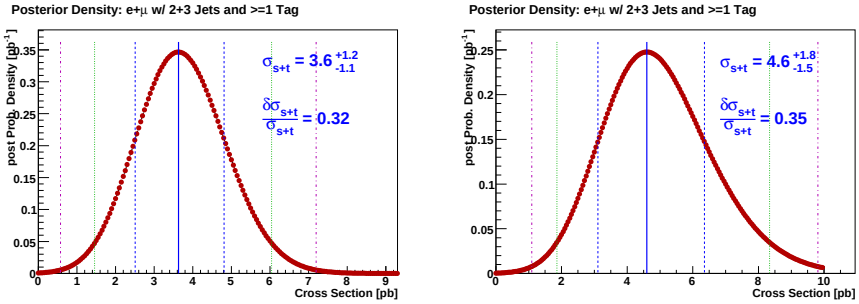

FIG. 23: Measured 1D posterior plots for the combined e+$\mu \geq 1$ b-tag channel with statistical uncertainties only (left plot) and with systematic uncertainties as well (right plot).

Table 7 shows the measured cross sections from various combinations of analysis channels. The averaged relative uncertainties on the measured cross sections are shown in Table 8.

Measured tb+tqb Cross Section							
	1,2tags + 2,3jets		e,μ + 2,3jets		e,μ + 1,2tags		All
	e-chan	μ-chan	1 tag	2 tags	2 jets	3 jets	channels
Statistics only	$3.0^{+1.5}_{-1.4}$	$4.5^{+1.8}_{-1.7}$	$2.8^{+1.2}_{-1.2}$	$7.9^{+3.3}_{-3.0}$	$3.5^{+1.4}_{-1.3}$	$3.9^{+2.3}_{-2.2}$	$3.6^{+1.2}_{-1.1}$
With systematics	$3.1^{+2.2}_{-1.8}$	$7.4^{+3.0}_{-2.5}$	$4.5^{+2.0}_{-1.7}$	$6.8^{+4.7}_{-3.8}$	$4.7^{+2.0}_{-1.7}$	$4.9^{+3.7}_{-3.1}$	$\mathbf{4.6^{+1.8}_{-1.5}}$

TABLE 7: Measured tb+tqb cross sections, without and with systematic uncertainties, for many combinations of the analysis channels. The final result of this analysis is shown in the lower right hand corner in bold type.

Relative Uncertainties on the Measured tb+tqb Cross Section							
	1,2tags + 2,3jets		e,μ + 2,3jets		e,μ + 1,2tags		All
	e-chan	μ-chan	1 tag	2 tags	2 jets	3 jets	channels
Statistics only	50%	39%	44%	40%	37%	57%	32%
With systematics	64%	38%	41%	62%	39%	70%	**35**%

TABLE 8: Relative uncertainties on the measured tb+tqb cross section, without and with systematic uncertainties, for many combinations of the analysis channels. The best value from all channels combined, with systematics, is shown in bold type.

The combined result with full systematics is

$$\sigma\left(p\bar{p} \rightarrow tb + tqb + X\right) = 4.6^{+1.8}_{-1.5} \text{ pb}.$$

A breakdown of the uncertainties on the tb+tqb cross section measurement is given in Table 9. The systematic uncertainties were calculated using an ensemble containing 200 datasets generated with an input single top cross section of 4.7 pb. The cross section of each dataset was measured and the average posterior width (average of upper and lower 1σ uncertainties) was calculated over all datasets for each source of systematic uncertainty independently. The systematic uncertainty for each source was estimated by subtracting in quadrature from the average posterior width obtained with a particular source of systematic, the average posterior width without systematic uncertainties. The total expected systematic uncertainty is estimated by adding in quadrature all the individual expected systematic uncertainties. The statistical uncertainty of the measurement is estimated by subtracting in quadrature the total expected systematic uncertainty from the actual total uncertainty.

Contributions to the Cross Section Uncertainty	
Systematics components	
Luminosity	0.69 pb
$t\bar{t}$ cross section	0.74 pb
Matrix method	0.84 pb
Trigger	0.48 pb
Primary vertex	0.31 pb
Lepton ID	0.50 pb
Jet ID	0.18 pb
Jet fragmentation	0.63 pb
Jet energy scale	0.57 pb
Tag-rate functions	0.60 pb
Combined systematics	$+1.34$ -1.02 pb
Statistics	$+1.19$ -1.13 pb
Total uncertainty	$+1.79$ -1.50 pb

TABLE 9: Contribution of each systematic uncertainty to the total systematic uncertainty on the tb+tqb cross section.

Figure 24 shows the cross sections measured for combined tb+tqb production in each independent analysis channel, and the combined result, taken from the 1-d posterior density distribution measurements.

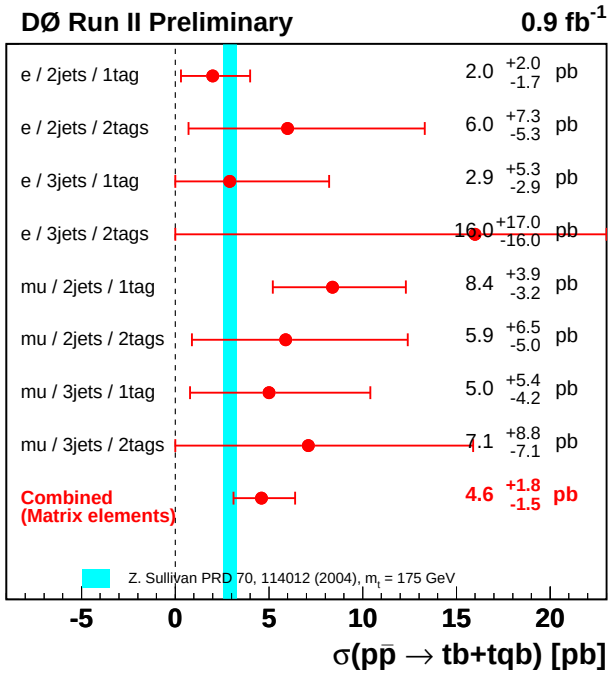

FIG. 24: Summary plot of the measured single top quark cross sections showing the individual measurements and their combination.

7. SIGNAL SIGNIFICANCE

We calculate the probability that the background alone could fluctuate up to or above the measured cross section. This is known as the p-value, see Appendix 10 in Ref. [4]. Figure 25 shows the distribution of cross sections for the zero-signal ensemble with full systematics included. From this distribution, we calculate a p-value of 0.22%, which corresponds to a 2.9σ fluctuation from zero.

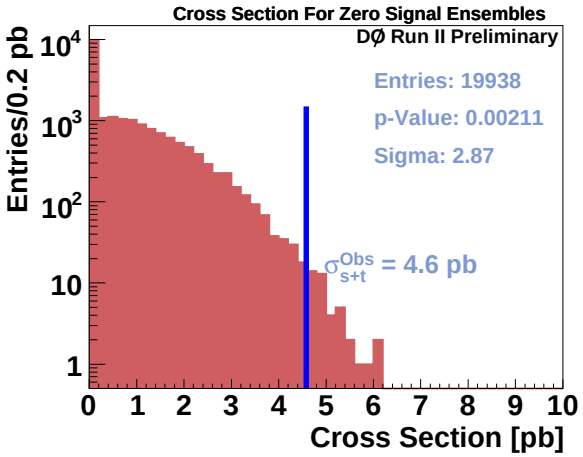

FIG. 25: Distribution of cross sections from a zero-signal ensemble with full systematics included.

8. EVENT CHARACTERISTICS

After the matrix element discriminant has been calculated, it is possible to place a cut on this value to select events in data and Monte Carlo to see if they are consistent with single top quark production. For this section, an event is considered very single top quark like if both the s-channel and t-channel discriminants are greater than 0.7. Similarly, an event is considered background like if both discriminants are less than 0.4. Figure 26 shows the invariant mass of the lepton, neutrino, and tagged jet before and after the discriminant cut, and Fig. 27 shows the lepton-charge times pseudorapidity of the untagged jet.

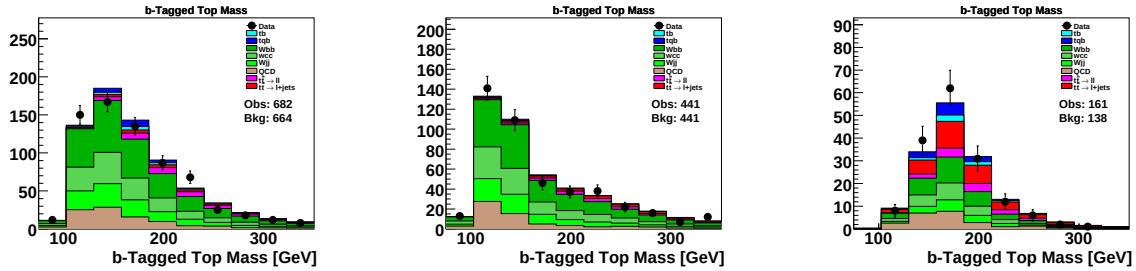

FIG. 26: Invariant mass of the lepton, neutrino, and tagged jet for all events (left plot), for events with $D < 0.4$ (middle plot), and events with $D > 0.7$ (right plot).

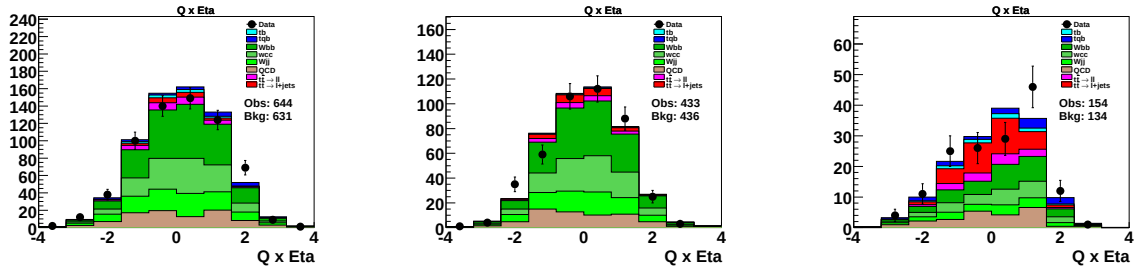

FIG. 27: Lepton charge times the pseudorapidity of the untagged jet for all events (left plot), for events with $D < 0.4$ (middle plot), and events with $D > 0.7$ (right plot). The number of observed events is different from the b-tagged top mass plot because this variable is only defined for events with at least one untagged jet.

9. SUMMARY AND CONCLUSIONS

We have used leading order matrix elements to separate expected single top quark signals from background. This analysis used lepton+jets events selected from nearly 1 fb^{-1} of Run II data. The two-dimensional discriminant output from the analysis is used to measure the single top cross section. We obtain the following result:

$$\sigma\left(p\bar{p} \rightarrow tb + tqb + X\right) = 4.6^{+1.8}_{-1.5} \text{ pb}$$

This result has a p-value of 0.22%, corresponding to a 2.9σ Gaussian equivalent significance, meaning it is highly unlikely to be an upward fluctuation of the background that has produced this measured cross section.

APPENDIX A: PROBABILITY CALCULATION

1. Differential Cross Section at the Parton Level

The matrix element analysis technique reconstructs each event to the final state four-vectors to evaluate the signal and background leading order matrix element. The following sections derive the signal and background probabilities starting from the final state at the parton level and then relating these objects to the physical quantities measured in the detector. The following also assumes a lepton, neutrino, and two quarks in the final state.

The probability density for a process to occur at a hadron-hadron collider is given as an integral of the hard-scatter differential cross section over all possible ways of producing the process from the quarks and gluons inside the hadron. This probability density, shown below, is a convolution of the hard scatter differential cross section with a parton distribution function for each of the two partons from the hadrons with an integral over all possible momentum fractions x_i, x_j from each initial parton.

$$\mathcal{P}(\vec{y}) = \frac{1}{\sigma}\sum_{i,j}\int f_i(q_1, Q^2)dq_1 \times f_j(q_2, Q^2)dq_2 \times d\sigma_{hs,ij}(\vec{y}) \tag{A1}$$

where the normalization constant σ is defined as integral of the differential cross section over the initial- and final-state phase spac:

$$\sigma = \int\sum_{i,j}\int f_i(q_1, Q^2)dq_1 \times f_j(q_2, Q^2)dq_2 \times \frac{\partial\sigma_{hs,ij}(\vec{y})}{\partial\vec{y}}d\vec{y} \tag{A2}$$

and finally, the hard-scatter differential cross section is defined as the product of the final state phase space factor, the square of the matrix element amplitude and an overall flux factor:

$$d\sigma_{hs} = \frac{(2\pi)^4}{4}\frac{|\mathcal{M}|^2}{\sqrt{(q_1q_2)^2 - m_1^2m_2^2}}\frac{d^3p_1}{(2\pi)^3 2E_1}\frac{d^3p_2}{(2\pi)^3 2E_2}\frac{d^3p_\ell}{(2\pi)^3 2E_\ell}\frac{d^3p_\nu}{(2\pi)^3 2E_\nu}\delta^4(q_1q_2; p_1, p_2, p_\ell, p_\nu) \tag{A3}$$

2. Evaluating the Hard Scatter Differential Cross Section

The following section evaluates the differential cross section shown in Eq. A2 given a set of inital and final state four-vectors.

The first assumption made is that all collisions occur along the beam axis with no net transverse momentum. This means the initial state four vectors can be written as

$$q_1 = (E_{beam}x_1, 0, 0, E_{beam}x_1) \tag{A4}$$

$$q_2 = (E_{beam}x_2, 0, 0, -E_{beam}x_2) \tag{A5}$$

The next assumption is that all particle masses are known and are negligible compared to their energies and thus can be ignored for this calculation. The flux factor (shown below) in the hard scatter cross section can now be written in terms in the two momentum fractions of the incoming partons:

$$\frac{1}{\sqrt{(q_1q_2)^2 - m_1^2m_2^2}} \rightarrow \frac{1}{\sqrt{(q_1q_2)^2}} \rightarrow \frac{1}{2E_{beam}x_1x_2} \tag{A6}$$

For the remainder of the note, the following notation will be used to distinquish quarks, leptons, and neutrinos: p_ℓ is the momemtum of the lepton, $p_{1,2}$ is the momentum of the first and second final state partons, and p_ν is the neutrino momentum. Because the phase space is written in terms of rectangular coordinates, the next step towards the final differential cross section equation is to redefine the phase space factors in terms of spherical coordiniates. This is done for all final state particles except the neutrino for reasons that will be clear later in the document.

$$d\Phi_4 = \frac{d^3p_1}{(2\pi)^3 2E_1}\frac{d^3p_2}{(2\pi)^3 2E_2}\frac{d^3p_\ell}{(2\pi)^3 2E_\ell}\frac{d^3p_\nu}{(2\pi)^3 2E_\nu}\delta^4 \rightarrow \tag{A7}$$

$$\frac{1}{16(2\pi)^{12}}\frac{|p_1|^2 d|p_1|d\Omega_1}{E_1}\frac{|p_2|^2 d|p_2|d\Omega_2}{E_2}\frac{|p_\ell|^2 d|p_\ell|d\Omega_\ell}{E_\ell}\frac{d^x_\nu d^y_\nu d^z_\nu}{E_\nu}\delta^4 \tag{A8}$$

To summarize, the full hard scatter differential cross section is now

$$d\sigma_{hs} = \frac{1}{128(2\pi)^8 E_{beam}}\frac{|\mathcal{M}|^2}{2x_1x_2}\frac{|p_1|^2 d|p_1|d\Omega_1}{E_1}\frac{|p_2|^2 d|p_2|d\Omega_2}{E_2}\frac{|p_\ell|^2 d|p_\ell|d\Omega_\ell}{E_\ell}\frac{d^x_\nu d^y_\nu d^z_\nu}{E_\nu}\delta^4 \tag{A9}$$

3. Evaluating the Hadron-Hadron Differential Cross Section

The next step to writing the full hadron-hadron differential cross section is to rewrite Eq. A9 such that any integration over the phase space will remove the four-dimensional delta function required for energy and momentum conservation. The delta function is currently written such that it will vanish only over integrations of total p_x, p_y, p_z, and E. Because there was an original assumption of no net transverse momentum in the collision, the total p_x and p_y can be solved for the neutrino transverse momentum.

$$\sum_i P^x_i = p^x_1 + p^x_2 + p^x_\ell + p^x_\nu = 0 \rightarrow p^x_\nu = -p^x_1 - p^x_2 - p^x_\ell \tag{A10}$$

$$\sum_i P^y_i = p^y_1 + p^y_2 + p^y_\ell + p^y_\nu = 0 \rightarrow p^y_\nu = -p^y_1 - p^y_2 - p^y_\ell \tag{A11}$$

The total p_z and requirement can be rewritten in terms of the intial parton's momemtum fraction and the other final state partons' z momenta and thus solve for the neutrino p_z:

$$\sum_i P^z_i = p^z_1 + p^z_2 + p^z_\ell + p^z_\nu - E_{beam}x_1 + E_{beam}x_2 = 0 \rightarrow$$
$$p^z_\nu = -E_{beam}(x_1 - x_2) - p^z_1 - p^z_2 - p^z_\ell \tag{A12}$$

Finally, the total energy delta function implies the following:

$$E_{beam}x_1 + E_{beam}x_2 = E_1 + E_2 + E_\ell + E_\nu \tag{A13}$$

At this point, is it useful to rewrite the full differential cross section at the parton level:

$$\begin{aligned} d\sigma(\vec{y}) = \sum_{i,j}\int f_i(x_1,Q^2)dx_1 \times f_j(x_2,Q^2)dx_2 \times \frac{1}{128(2\pi)^8 E_{beam}}\frac{|\mathcal{M}|^2}{2x_1x_2} \times \\ \frac{|p_1|^2 d|p_1|d\Omega_1}{E_1}\frac{|p_2|^2 d|p_2|d\Omega_2}{E_2}\frac{|p_\ell|^2 d|p_\ell|d\Omega_\ell}{E_\ell}\frac{d^x_\nu d^y_\nu d^z_\nu}{E_\nu} \times \\ \delta(p^x_\nu + p^x_1 + p^x_2 + p^x_\ell) \times \\ \delta(p^y_\nu + p^y_1 + p^y_2 + p^y_\ell) \times \\ \delta(p^z_\nu + E_{beam}(x_1 - x_2) + p^z_1 + p^z_2 + p^z_\ell) \times \\ \delta(E_{beam}x_1 + E_{beam}x_2 - E_1 - E_2 - E_\ell - E_\nu) \end{aligned} \tag{A14}$$

The next step is to rewrite the integrational variables, x_1 and x_2, in terms of the total energy and total p_z:

$$x_1 = \frac{E_{tot} + p^z_{tot}}{2E_{beam}} \tag{A15}$$

$$x_2 = \frac{E_{tot} - p^z_{tot}}{2E_{beam}} \tag{A16}$$

Now, the integration over x_1 and x_2 can be rewritten in terms of E_{tot} and p_z:

$$dx_1 dx_2 = \frac{1}{J(x_1, x_2; E_{tot}, p^z_{tot})} dE_{tot} dp^z_{tot} \tag{A17}$$

$$J(x_1, x_2; E_{tot}, p^z_{tot}) = 2E^2_{beam} \tag{A18}$$

At this point the integration over the total energy and p_z will constrain the two incoming partons' momentum fractions through Eq. A15 and A16.

The full differential cross section at the parton level can now be written as

$$\begin{aligned} d\sigma(\vec{y}) = \sum_{i,j} \int f_i(x_1, Q^2) \times f_j(x_2, Q^2) \times \frac{1}{128(2\pi)^8 E_{beam}} \frac{|\mathcal{M}|^2}{2x_1 x_2} \times \\ \frac{|p_1|^2 d|p_1| d\Omega_1}{E_1} \frac{|p_2|^2 d|p_2| d\Omega_2}{E_2} \frac{|p_\ell|^2 d|p_\ell| d\Omega_\ell}{E_\ell} \frac{d^x_\nu d^y_\nu d^z_\nu}{E_\nu} \times \\ \int \frac{1}{2E^2_{beam}} dp^z_{tot} \end{aligned} \tag{A19}$$

where the implicit integration over the four dimensional delta function yields the following formulas for the neutrino four vector and the incoming partons' momentum fraction in terms of the remaining differential variables.

$$p^x_\nu = -p^x_1 - p^x_2 - p^x_\ell \tag{A20}$$

$$p^y_\nu = -p^y_1 - p^y_2 - p^y_\ell \tag{A21}$$

$$p^z_\nu = -p^z_{tot} - p^z_1 - p^z_2 - p^z_\ell \tag{A22}$$

$$x_1 = \frac{E_1 + E_2 + E_\ell + E_\nu + p^z_{tot}}{2E_{beam}} \tag{A23}$$

$$x_2 = \frac{E_1 + E_2 + E_\ell + E_\nu - p^z_{tot}}{2E_{beam}} \tag{A24}$$

4. Relating Reconstructed Objects to Partons

The previous sections have calculated the differential cross section for a hadron-hadron collision producing a lepton, neutrino, and two partons in the final state. These particles are not exactly what is measured in the detector and thus it is necessary to relate quantities. To do this, the differential cross section is convoluted with a function, $W(\vec{x}, \vec{y})$, which is the probability of producing a final state, $\vec{y}$, and observed state, $\vec{x}$, in the detector. The resulting differential cross section is then integrated over the final state phase space, $d\vec{y}$:

$$\frac{\partial \sigma'(\vec{x})}{\partial \vec{x}} = \int \frac{\partial \sigma(\vec{y})}{\partial \vec{y}} W(\vec{x}, \vec{y}) d\vec{y} \tag{A25}$$

where the function $W(\vec{x}, \vec{y})$ is assumed to be factorizable for each measured object:

$$W(\vec{x}, \vec{y}) = \prod_i W_i(\vec{x_i}, \vec{y_i}) \tag{A26}$$

a. Jets

The transfer function for jets measured in the calorimter is assumed to only be a function of the relative energy difference between the two objects and all angles are assumed to be well measured:

$$W_{jet}(\vec{x}_{jet}, \vec{y}_{parton}) = W(E_{jet} - E_{parton}) \times \delta(\Omega_{jet} - \Omega_{parton}) \tag{A27}$$

where $W(E_{jet} - E_{parton})$ is parametrized using the following functional form:

$$W(E_{jet} - E_{parton}) = \frac{e^{-\frac{(E_{jet}-E_{parton}-p_1)^2}{2p_2^2}} + p_3 e^{-\frac{(E_{jet}-E_{parton}-p_4)^2}{2p_2^5}}}{2\pi(p_2 + p_3 p_5)} \tag{A28}$$

where $p_i = \alpha_i + \beta_i \times E_{parton}$. The five α and five β parameters are determined by minimizing a likelihood formed by measuring the parton energy in Monte Carlo and the matched jet energy also in Monte Carlo. The parameters used for this analysis were determined in several regions of the calorimeter to account for the resolution differences in the detector.

b. Electrons

The transfer function for electrons [12] is assumed to be solely a function of the reconstructed energy of the electron, E_{reco}, the parton energy of the electron, E_{parton}, and θ, the production angle with respect to the beam axis:

$$W_{electron}(\vec{x}_{reco}, \vec{y}_{parton}) = W(E_{reco}, E_{parton}, \theta) \times \delta(\Omega_{reco} - \Omega_{parton}) \tag{A29}$$

where $W(E_{reco}, E_{parton}, \theta)$ is parametrized using the following functional form:

$$\begin{aligned}
W(E_{reco}, E_{parton}, \theta) &= \frac{1}{2\pi\sigma}\exp[-\frac{(E_{reco} - E_{center})^2}{2\sigma^2}] && \text{(A30)}\\
E_{center} &= 1.0002 E_{parton} + 0.324\, GeV/c^2 && \text{(A31)}\\
\sigma &= 0.028 E_{center} \oplus \text{Sampling}(E_{center}, \eta) E_{center} \oplus 0.4 && \text{(A32)}\\
\text{Sampling}(E, \theta) &= \left[\frac{0.164}{\sqrt{E}} + \frac{0.122}{E}\right] \exp\left[\frac{\text{p1}(E)}{\sin\theta} - \text{p1}(E)\right] && \text{(A33)}\\
\text{p1}(E) &= 1.35193 - \frac{2.09564}{E} - \frac{6.98578}{E^2}. && \text{(A34)}
\end{aligned}$$

c. Muons

The transfer function for muons [13] is assumed to be a function of

$$\Delta\left(\frac{q}{p_t}\right) = \left(\frac{q}{p_t}\right)_{reco} - \left(\frac{q}{p_t}\right)_{parton} \tag{A35}$$

and of η_{CFT},

$$W_{muon}(\vec{x}_{reco}, \vec{y}_{parton}) = W\left(\Delta\left(\frac{q}{p_t}\right), \eta_{\text{CFT}}\right) \times \delta(\Omega_{reco} - \Omega_{parton}) \tag{A36}$$

where $W\left(\Delta\left(\frac{q}{p_t}\right), \eta_{\rm CFT}\right)$ is parametrized using a single Gaussian:

$$W\left(\Delta\left(\frac{q}{p_t}\right), \eta_{\rm CFT}\right) = \frac{1}{2\pi\sigma}\exp\left\{-\frac{\left[\Delta\left(\frac{q}{p_t}\right)\right]^2}{2\sigma^2}\right\} \tag{A37}$$

$$\sigma = \begin{cases} \sigma_o & : \quad |\eta_{\rm CFT}| \leq \eta_o \\ \sqrt{\sigma_o^2 + [c(|\eta_{\rm CFT}| - \eta_o)]^2} & : \quad |\eta_{\rm CFT}| > \eta_o \end{cases} \tag{A38}$$

There are three fitted parameters in the above equations: σ_o, c, and η_o, each of which is actually fitted by two sub-parameters:

$$par = par(0) + par(1) \cdot 1/p_t. \tag{A39}$$

Furthermore, these parameters are derived for four classes of events: those that were from before or after the 2004 shutdown, when the magnetic field strength changed, and in each run range, those that have an SMT hit and those that do not.

As a simplification, we assume $q_{reco} = q_{parton}$, that is, we do not consider charge misidentification

5. Full Differential Cross Section and Normalization

The full differential cross section at the detector object level can now be written as

$$\frac{\partial\sigma^{'}(\vec{x})}{\partial\vec{x}} = \int dp_{tot}^z dq_1 dq_2 dp_\ell \sum_{i,j} f_i(q_1, Q^2) \times f_j(q_2, Q^2)$$
$$\times \frac{1}{256(2\pi)^8 E_{beam}^3} \frac{|\mathcal{M}|^2}{2x_1x_2} \times \frac{p_1^2}{E_1}\frac{p_2^2}{E_2}\frac{p_\ell^2}{E_\ell}\frac{1}{E_\nu} \times W_{Lepton} W_{Jet1} W_{Jet2} \tag{A40}$$

The final step to evaluating the probability density is to properly normalize the differential cross section in Eq. A40. This is done by integration of the differential cross section over all possible states in the detector. Since the event selection cuts will change the number events due to acceptance losses, this must be accounted for in the overall normalization (cross section) calculation. The total cross section is then written as

$$\sigma = \int \frac{\partial\sigma^{'}(\vec{x})}{\partial\vec{x}} d\vec{x} = \int d\vec{x} dp_{tot}^z dq_1 dq_2 dp_\ell \sum_{i,j} f_i(q_1, Q^2) \times f_j(q_2, Q^2)$$
$$\times \frac{1}{256(2\pi)^8 E_{beam}^3} \frac{|\mathcal{M}|^2}{2x_1x_2} \times \frac{p_1^2}{E_1}\frac{p_2^2}{E_2}\frac{p_\ell^2}{E_\ell}\frac{1}{E_\nu} \times W_{Lepton} W_{Jet1} W_{Jet2} \times \Theta_{\rm Cuts}(\vec{x}) \tag{A41}$$

APPENDIX B: DISCRIMINANT OUTPUT PLOTS

MATRIX ELEMENT OUTPUTS FOR THE ELECTRON CHANNEL WITH TWO JETS

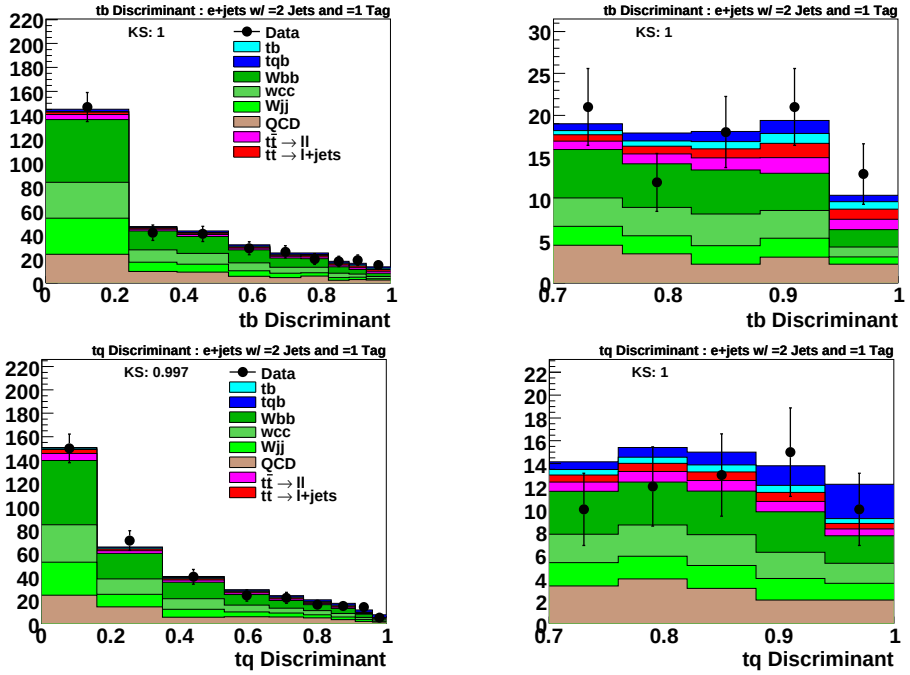

FIG. 28: Discriminant plots for the electron channel with one b tag. Upper row: tb discriminant, lower row: tq discriminant. Left column, full discriminant range, right column, close-up of the high end of the distribution.

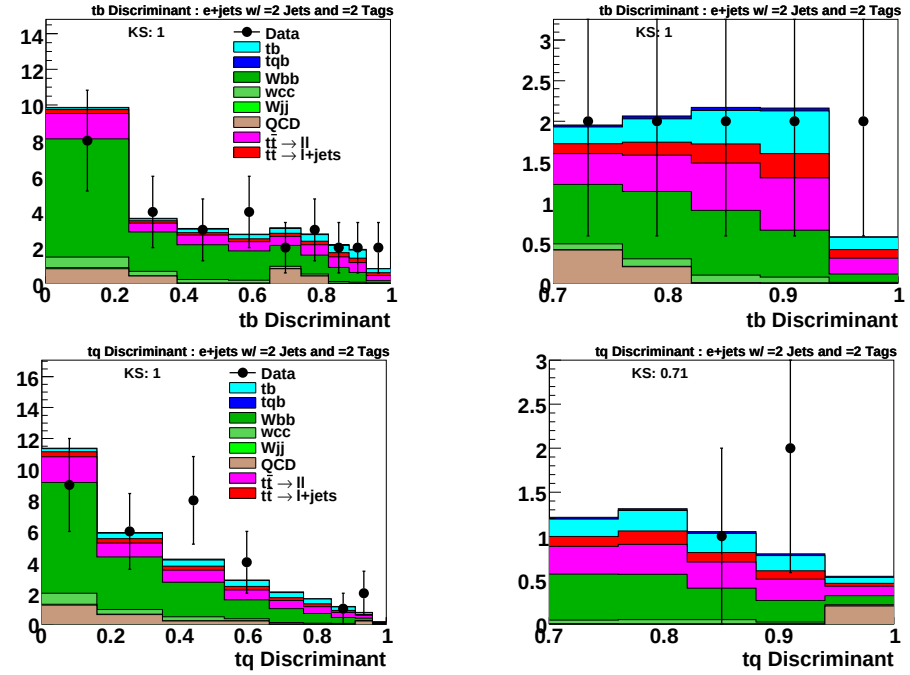

FIG. 29: Discriminant plots for the electron channel with two b tags. The plot layout is the same as in Fig. 28.

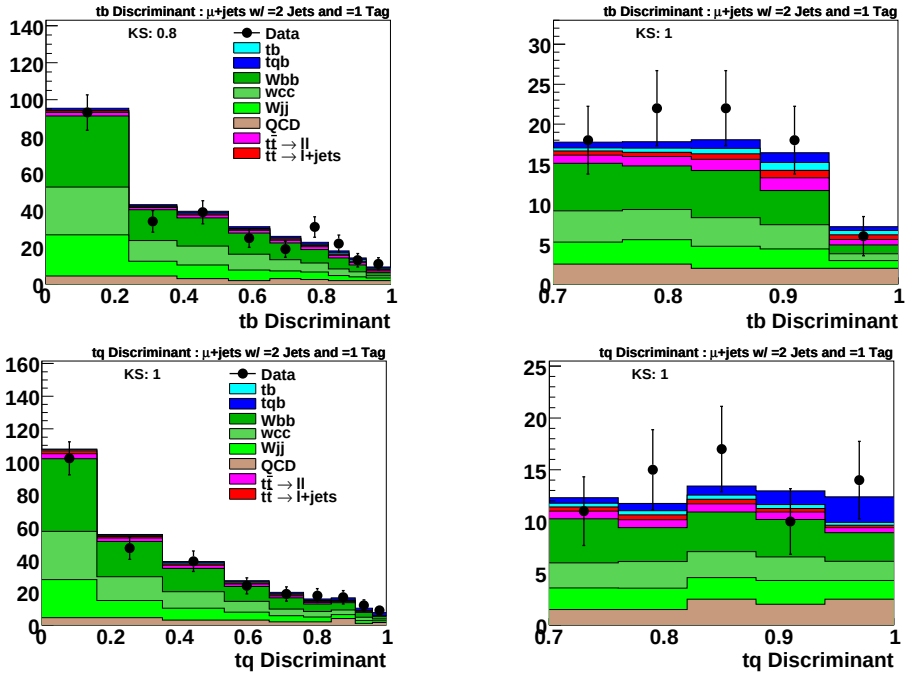

FIG. 30: Discriminant plots for the muon channel with one b tag. Upper row: tb discriminant, lower row: tq discriminant. Left column, full discriminant range, right column, close-up of the high end of the distribution.

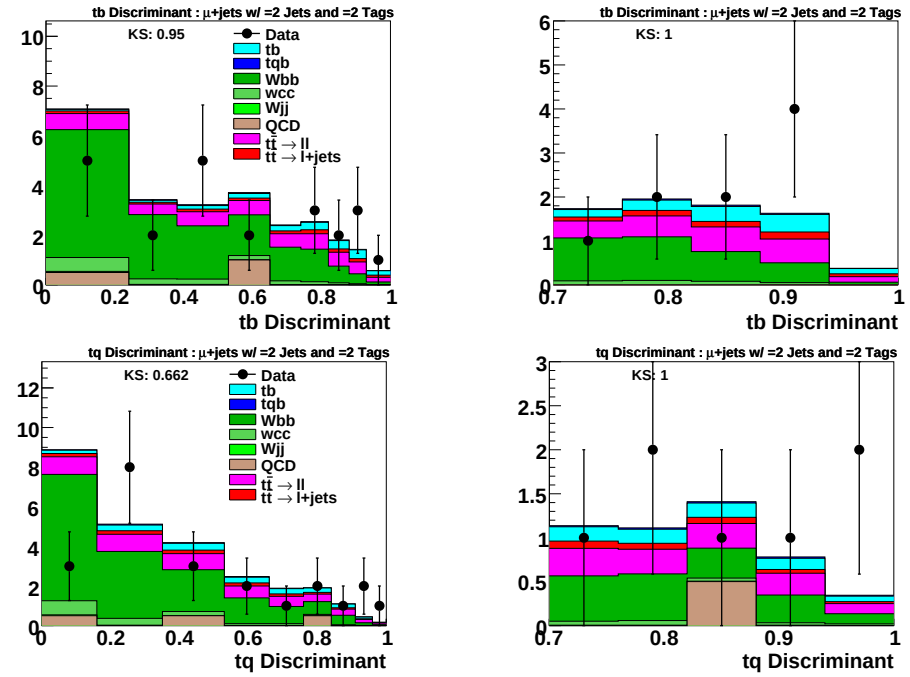

FIG. 31: Discriminant plots for the muon channel with two b tags. The plot layout is the same as in Fig. 28.

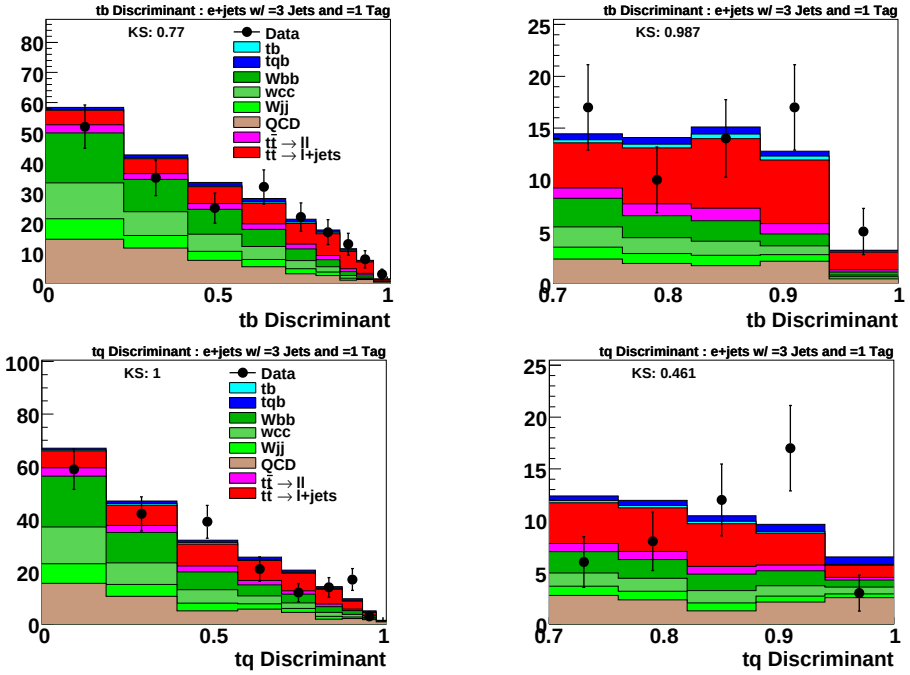

FIG. 32: Discriminant plots for the electron channel with one b tag. Upper row: tb discriminant, lower row: tq discriminant. Left column, full discriminant range, right column, close-up of the high end of the distribution.

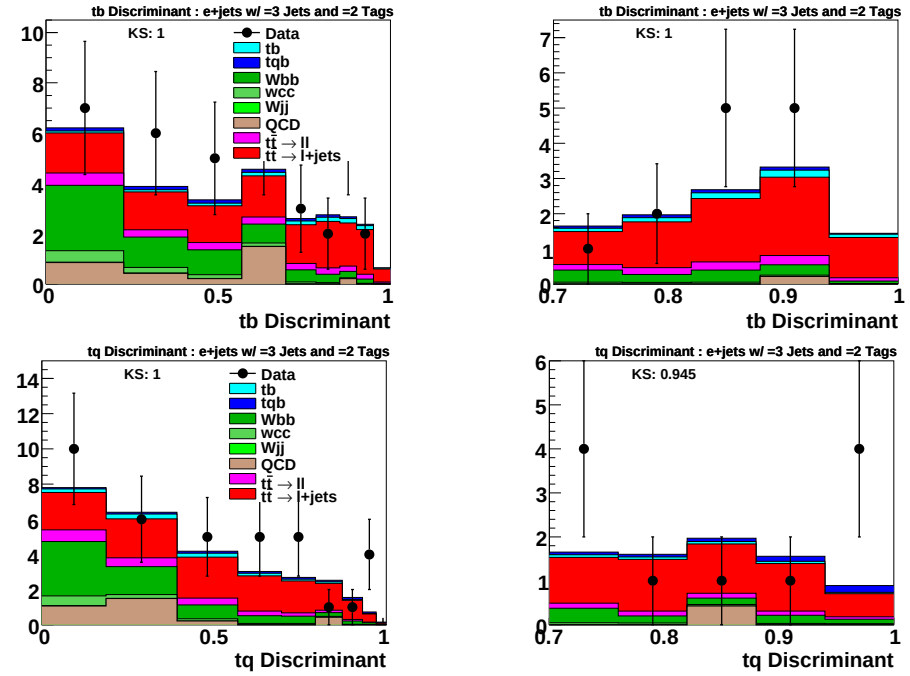

FIG. 33: Discriminant plots for the electron channel with two b tags. The plot layout is the same as in Fig. 28.

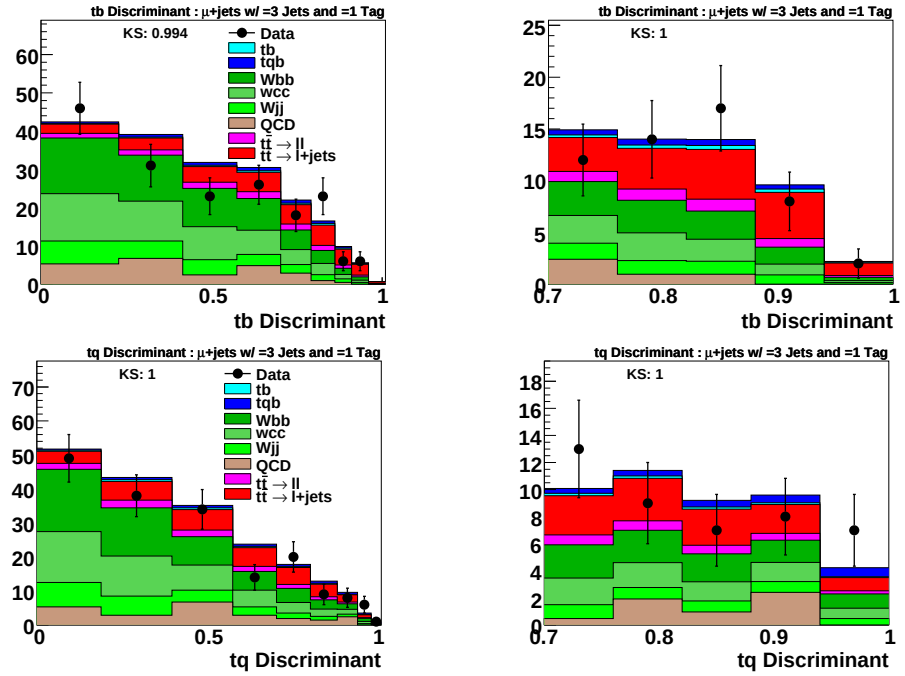

FIG. 34: Discriminant plots for the muon channel with one b tag. Upper row: tb discriminant, lower row: tq discriminant. Left column, full discriminant range, right column, close-up of the high end of the distribution.

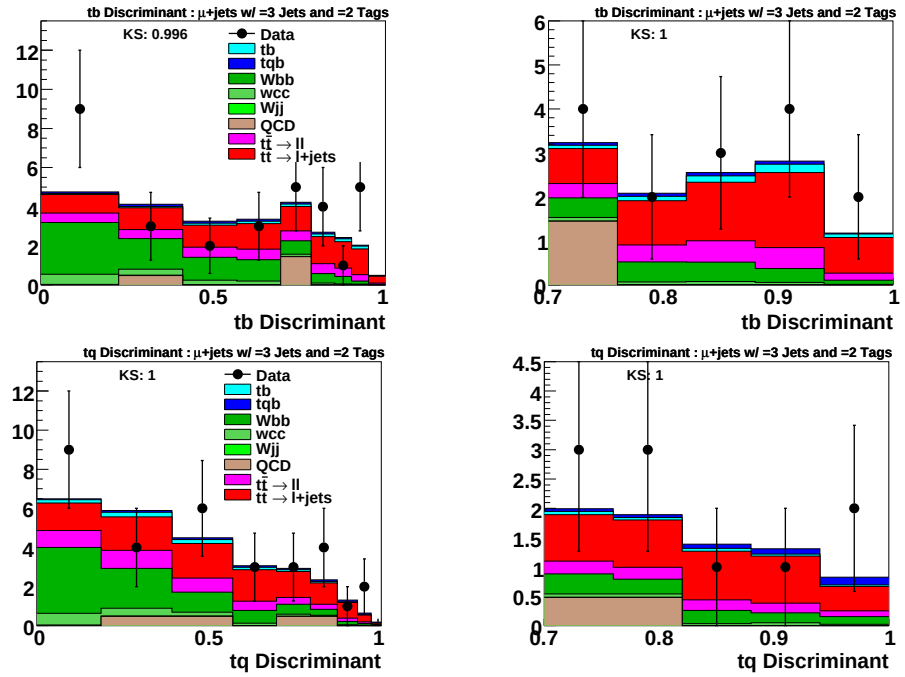

FIG. 35: Discriminant plots for the muon channel with two b tags. The plot layout is the same as in Fig. 28.

APPENDIX C: INTEGRATION VARIABLE REMAPPING

1. Introduction

This section will layout the jacobian needed for the $10 \rightarrow 10$ remapping of variables for the parton level cross section. The base variables used are shown below.

- p_3: Absolute momentum of the lepton
- p_5: Absolute momentum of the first quark
- p_6: Absolute momentum of the second quark
- p_{tot}^z: Total p_z of the system
- $cos(\theta_3)$: Cosine(θ) of the lepton
- ϕ_3: ϕ of the lepton
- $cos(\theta_5)$: Cosine(θ) of the first quark
- ϕ_5: ϕ of the first quark
- $cos(\theta_6)$: Cosine(θ) of the second quark
- ϕ_6: ϕ of the second quark

Other variables that are useful for the integration are

- m_{34}: Mass of the lepton and neutrino (W mass)
- m_{345}: Mass of the lepton, neutrino, and first quark (top mass)
- m_{56}: Mass of the first and second quark ($b\bar{b}$ mass)

Since some of these variables are sharp peaks (W and top masses), it is much better to sample from the expected distribution rather than make requirements of the invariant masses. The W and top masses are expected to follow a Breit-Wigner distribution shown below.

$$\sigma(M_{34}) = \frac{1}{\pi}\left[\frac{\gamma}{(M_{34} - M_W)^2 + \gamma^2}\right] \tag{C1}$$

where M_{34} is the mass of the lepton and neutrino. Similarly, the top mass has the following expected distribution.

$$\sigma(M_{345}) = \frac{1}{\pi}\left[\frac{\gamma}{(M_{345} - M_{top})^2 + \Gamma^2}\right] \tag{C2}$$

where M_{345} is the mass of the lepton, neutrino, and first quark.

2. Sampling from a Breit-Wigner mass distribution

Sampling from a Breit-Wigner distribution is done by selecting a random point between 0 and 1 from the cumulative distribution function of the BW function. The cumulative distribution function, of cdf, for the Breit-Wigner distribution is shown below.

$$\int \sigma(m; m_0; \Gamma) = F(m; m_0, \Gamma) = \frac{1}{\pi}\tan^{-1}\left[\frac{m - m_0}{\Gamma}\right] + \frac{1}{2} \tag{C3}$$

The value of F is taken as a random number between 0 and 1. After selecting a value of F, the next step is to solve for m. As a function of F, defined as u for the following, the mass is

$$m = m_0 + \Gamma \tan\left[\pi(u - \frac{1}{2})\right] \tag{C4}$$

3. Sampling from a Breit-Wigner S_{cm} distribution

In the previous example, a distribution was sampled using a random number uniformly distributed from 0 to 1. In, this example, a new random number is used that is uniformaly samples from 0 to 1, but the maximum and minimum values of the variable are taken into account in the Jacobian.

The distribution of the variable S_{cm} is the following

$$s = m_0^2 + m_0\Gamma \tan\left[m_0\Gamma r\right] \tag{C5}$$

where r is defined in terms of the random variable, u, that is uniformaly distributed between 0 and 1.

$$r = (r_{max} - r_{min}) \times u + r_{min} \tag{C6}$$

where r_{max} and r_{min} are defined in terms of the variable s_{cm}.

$$r = \frac{1}{m_0\Gamma} \tan\left[\frac{s - m_0^2}{m_0\Gamma}\right] \tag{C7}$$

$$r_{min} = \frac{1}{m_0\Gamma} \tan\left[\frac{s_{min} - m_0^2}{m_0\Gamma}\right] \tag{C8}$$

$$r_{max} = \frac{1}{m_0\Gamma} \tan\left[\frac{s_{max} - m_0^2}{m_0\Gamma}\right] \tag{C9}$$

4. Jacobian for random sampling of a Breit-Wigner distribution around the W mass squared, s_{34}

The first case to consider is the Breit-Wigner sampling around the W mass S_{34} distribution and replace the integration variable, p_3 or the lepton momentum.

$$|J(p_3, u)| = \left|\frac{\partial p_3}{\partial u}\right| \tag{C10}$$

Because u is redined in terms of the variable r, we can rewrite C10 in terms of r instead of u.

$$|J(p_3, u)| = \left|\frac{\partial p_3}{\partial u}\right| = \left|\frac{\partial p_3}{\partial r} \times \frac{\partial r}{\partial u}\right| \tag{C11}$$

And since the variable r is sampling the S_{34} distribution it makes sense to define the Jacobian in terms of this variable instead of p_3.

$$|J(p_3,u)| = \left|\frac{\partial p_3}{\partial r} \times \frac{\partial r}{\partial u}\right| = \left|\frac{\partial p_3}{\partial s_{34}} \times \frac{\partial s_{34}}{\partial r} \times \frac{\partial r}{\partial u}\right| = \left|\frac{\frac{\partial s_{34}}{\partial r} \times \frac{\partial r}{\partial u}}{\frac{\partial s_{34}}{\partial p_3}}\right| \tag{C12}$$

Equation C12 has three components: $\frac{\partial s_{34}}{\partial r}$, $\frac{\partial r}{\partial u}$, and $\frac{\partial s_{34}}{\partial p_3}$. From equation C6, the partial derivative of r with respect to u is

$$\frac{\partial r}{\partial u} = r_{max} - r_{min} = \Delta r \tag{C13}$$

Next, the partial of s_{34} with respect to r can be determined from equation C5.

$$\frac{\partial s_{34}}{\partial r} = (m_W\Gamma_W)^2 \sec^2\left[m_W\Gamma_W r\right] \tag{C14}$$

Inserting the value of r(s) as defined in equation C7, equation C14 can be re-written as

$$\frac{\partial s_{34}}{\partial r} = (m_W\Gamma_W)^2 \sec^2\left[m_W\Gamma_W r\right] = (m_W\Gamma_W)^2 \sec^2\left[\arctan\left[\frac{s_{34}-m_W^2}{m_W\Gamma_W}\right]\right] \tag{C15}$$

Equation C15 is solved by defining a right triangle where

$$\begin{aligned}\tan(\theta) &= \frac{s_{34}-m_W^2}{m_W\Gamma_W} \\ \cos(\theta) &= \frac{1}{\sqrt{1+\left[\frac{s_{34}-m_W^2}{m_W\Gamma_W}\right]^2}}\end{aligned} \tag{C16}$$

Using these definitions, equation C15 is finally defined as

$$\frac{\partial s_{34}}{\partial r} = (m_W\Gamma_W)^2 \sec^2\left[\arctan\left[\frac{s_{34}-m_W^2}{m_W\Gamma_W}\right]\right] = (m_W\Gamma_W)^2 + (s_{34}-m_W^2)^2 \tag{C17}$$

Finally, we need the partial derivative of s_{34} with repect to p_3. First, we define s_{34}

$$s_{34} = m_3^2 + m_4^2 + 2E_3E_4 - 2p_3^xp_4^x - 2p_3^yp_4^y - 2p_3^zp_4^z \tag{C18}$$

Since the neutino four-vector is defined in terms of all the other particles in the event, we need to rewrite equation C18 to expose all the dependences on p_3. For the following, it is assumed that the lepton and neutrino are massless meaning $E_3 = p_3$.

$$\begin{aligned}s_{34} = {} & 2p_3\sqrt{(-p_3^x-p_5^x-p_6^x)^2 + (-p_3^y-p_5^y-p_6^y)^2 + (p_{tot}^z-p_3^z-p_5^z-p_6^z)^2} \\ & -2p_3^x(-p_3^x-p_5^x-p_6^x) - 2p_3^y(-p_3^y-p_5^y-p_6^y) - 2p_3^z(p_{tot}^z-p_3^z-p_5^z-p_6^z)\end{aligned} \tag{C19}$$

After combining like terms, we can evaluate the partial derivative of s_{34} with respect to p_3 that yields the relatively simple formula

$$\frac{\partial s_{34}}{\partial p_3} = 2(p_3 + p_4)(1 - \hat{p_3} \cdot \hat{p_4}) \tag{C20}$$

Finally, we can rewrite the Jacobian defined in C12 as

$$|J(p_3, u)| = \left| \frac{\frac{\partial s_{34}}{\partial r} \times \frac{\partial r}{\partial u}}{\frac{\partial s_{34}}{\partial p_3}} \right| = \frac{\Delta R \times [(m_W \Gamma_W)^2 + (s_{34} - m_W^2)^2]}{2(p_3 + p_4)(1 - \hat{p_3} \cdot \hat{p_4})} \tag{C21}$$

In some cases, it is also common to replace the first quark momentum integration with the Breit-Wigner sampling variable. In that case, we need to evaluate

$$|J(p_5, u)| = \left| \frac{\frac{\partial s_{34}}{\partial r} \times \frac{\partial r}{\partial u}}{\frac{\partial s_{34}}{\partial p_5}} \right| \tag{C22}$$

For this substitution, we only need to evaluate the partial derivative of s_{34} with respect to p_5. Assuming a massless quark, the result is

$$\frac{\partial s_{34}}{\partial p_5} = 2p_3(\hat{p_3} \cdot \hat{p_5} - \hat{p_4} \cdot \hat{p_5}) \tag{C23}$$

Combining equation C23 with C22 yields

$$|J(p_5, u)| = \left| \frac{\frac{\partial s_{34}}{\partial r} \times \frac{\partial r}{\partial u}}{\frac{\partial s_{34}}{\partial p_5}} \right| = \frac{\Delta R \times [(m_W \Gamma_W)^2 + (s_{34} - m_W^2)^2]}{2p_3(\hat{p_3} \cdot \hat{p_5} - \hat{p_4} \cdot \hat{p_5})} \tag{C24}$$

5. Jacobian for random sampling of a Breit-Wigner distribution around the top mass squared, s_{345}

The next case to consider is the Breit-Wigner sampling around the top mass squared S_{345} distribution and replace the integration variable, p_3 or the lepton momentum. As before, we need need to calculate the following

$$|J(p_3, u)| = \left| \frac{\partial p_3}{\partial r} \times \frac{\partial r}{\partial u} \right| = \left| \frac{\partial p_3}{\partial s_{345}} \times \frac{\partial s_{345}}{\partial r} \times \frac{\partial r}{\partial u} \right| = \left| \frac{\frac{\partial s_{345}}{\partial r} \times \frac{\partial r}{\partial u}}{\frac{\partial s_{345}}{\partial p_3}} \right| \tag{C25}$$

Equation C25 has three components: $\frac{\partial s_{345}}{\partial r}$, $\frac{\partial r}{\partial u}$, and $\frac{\partial s_{345}}{\partial p_3}$. We know $\frac{\partial r}{\partial u}$ from equation C13 where r_{min} and r_{max} are defined by the s_{345} system instead of the s_{34} system. We also know $\frac{\partial s_{345}}{\partial r}$ from C17 where we replace s_{34} with s_{345}.

$$\frac{\partial s_{345}}{\partial r} = (m_t \Gamma_t)^2 + (s_{345} - m_t^2)^2 \tag{C26}$$

We do need the partial derivative of s_{345} with repect to p_3. First, we define s_{345}

$$s_{345} = m_3^2 + m_4^2 + m_5^2 + 2E_3E_4 + 2E_3E_5 + 2E_4E_5 - 2p_3^x p_4^x - 2p_3^x p_5^x - 2p_4^x p_5^x - 2p_3^y p_4^y - 2p_3^y p_5^y - 2p_4^y p_5^y - 2p_3^z p_4^z - 2p_3^z p_5^z - 2p_4^z p_5^z \tag{C27}$$

As before, the neutino four-vector is defined in terms of all the other particles in the event so we need to rewrite equation C27 to expose all the dependences on p_3. For the following, it is assumed that the lepton, neutrino, and quark are massless meaning $E_3 = p_3$.

$$\begin{aligned} s_{345} = {} & 2p_3\sqrt{(-p_3^x - p_5^x - p_6^x)^2 + (-p_3^y - p_5^y - p_6^y)^2 + (p_{tot}^z - p_3^z - p_5^z - p_6^z)^2} + \\ & 2p_3p_5 + 2\sqrt{(-p_3^x - p_5^x - p_6^x)^2 + (-p_3^y - p_5^y - p_6^y)^2 + (p_{tot}^z - p_3^z - p_5^z - p_6^z)^2}p_5 - \\ & 2p_3^x(-p_3^x - p_5^x - p_6^x) - 2p_3^x p_5^x - 2(-p_3^x - p_5^x - p_6^x)p_5^x - \\ & 2p_3^y(-p_3^y - p_5^y - p_6^y) - 2p_3^y p_5^y - 2(-p_3^y - p_5^y - p_6^y)p_5^y - \\ & 2p_3^z(-p_3^z - p_5^z - p_6^z) - 2p_3^z p_5^z - 2(-p_3^z - p_5^z - p_6^z)p_5^z \end{aligned} \tag{C28}$$

After combining like terms, we can evaluate the partial derivative of s_{345} with respect to p_3 as

$$\frac{\partial s_{345}}{\partial p_3} = 2(p_3 + p_4 + p_5)(1 - \hat{p_3} \cdot \hat{p_4}) \tag{C29}$$

Finally, we can rewrite the Jacobian defined in C25 as

$$|J(p_3, u)| = \left| \frac{\frac{\partial s_{345}}{\partial r} \times \frac{\partial r}{\partial u}}{\frac{\partial s_{345}}{\partial p_3}} \right| = \frac{\Delta R \times [(m_t\Gamma_t)^2 + (s_{345} - m_t^2)^2]}{2(p_3 + p_4 + p_5)(1 - \hat{p_3} \cdot \hat{p_4})} \tag{C30}$$

Instead of replacing the lepton momentum integration variable, it is also common to replace the first quark momentum integration variable, p_5. In that case, we need to evaluate the following Jacobian.

$$|J(p_5, u)| = \left| \frac{\partial p_5}{\partial r} \times \frac{\partial r}{\partial u} \right| = \left| \frac{\partial p_5}{\partial s_{345}} \times \frac{\partial s_{345}}{\partial r} \times \frac{\partial r}{\partial u} \right| = \left| \frac{\frac{\partial s_{345}}{\partial r} \times \frac{\partial r}{\partial u}}{\frac{\partial s_{345}}{\partial p_5}} \right| \tag{C31}$$

The only difference is that partial derivative of s_{345} with respect to p_5 instead of p_3. However, since s_{345} is invariant under a change of p_3 and p_5 the partial derivatives must be equal. Thus,

$$|J(p_5, u)| = \left| \frac{\frac{\partial s_{345}}{\partial r} \times \frac{\partial r}{\partial u}}{\frac{\partial s_{345}}{\partial p_5}} \right| = \frac{\Delta R \times [(m_t\Gamma_t)^2 + (s_{345} - m_t^2)^2]}{2(p_3 + p_4 + p_5)(1 - \hat{p_4} \cdot \hat{p_5})} \tag{C32}$$

6. Jacobian for random sampling of two Breit-Wigner distributions around the top mass squared, s_{345} and W mass squared, s_{34}

The next situation is to sample from a Breit-Wigner around the top mass squared and the W mass squared, or s_{345} and s_{34}. It is common to replace the lepton momentum and first

quark momentum integration variables with the two new variables. Since we are replacing two variables, we need to evaluate the following Jacobian

$$|J(p_3, p_5; u_1, u_2)| = \begin{vmatrix} \frac{\partial p_3}{\partial u_1} & \frac{\partial p_3}{\partial u_2} \\ \frac{\partial p_5}{\partial u_1} & \frac{\partial p_5}{\partial u_2} \end{vmatrix} \tag{C33}$$

where u_1 and u_2 are the sampling variables around the top mass squared and W mass squared, respectively.

We have already computed the partial derivatives for each of these cases in the previous two sections, thus the result is

$$|J(p_3, p_5; u_1, u_2)| = \begin{vmatrix} \frac{\partial p_3}{\partial u_1} & \frac{\partial p_3}{\partial u_2} \\ \frac{\partial p_5}{\partial u_1} & \frac{\partial p_5}{\partial u_2} \end{vmatrix} = \begin{vmatrix} \frac{\frac{\partial s_{345}}{\partial r}\times\frac{\partial r}{\partial u}}{\frac{\partial s_{345}}{\partial p_3}} & \frac{\frac{\partial s_{345}}{\partial r}\times\frac{\partial r}{\partial u}}{\frac{\partial s_{345}}{\partial p_5}} \\ \frac{\frac{\partial s_{34}}{\partial r}\times\frac{\partial r}{\partial u}}{\frac{\partial s_{34}}{\partial p_3}} & \frac{\frac{\partial s_{34}}{\partial r}\times\frac{\partial r}{\partial u}}{\frac{\partial s_{34}}{\partial p_5}} \end{vmatrix} =$$

$$\begin{vmatrix} \frac{\Delta R_{345}\times\left[(m_t\Gamma_t)^2+(s_{345}-m_t^2)^2\right]}{2(p_3+p_4+p_5)(1-\hat{p_3}\cdot\hat{p_4})} & \frac{\Delta R_{34}\times\left[(m_W\Gamma_W)^2+(s_{34}-m_W^2)^2\right]}{2(p_3+p_4)(1-\hat{p_3}\cdot\hat{p_4})} \\ \frac{\Delta R_{345}\times\left[(m_t\Gamma_t)^2+(s_{345}-m_t^2)^2\right]}{2(p_3+p_4+p_5)(1-\hat{p_4}\cdot\hat{p_5})} & \frac{\Delta R_{34}\times\left[(m_W\Gamma_W)^2+(s_{34}-m_0^2)^2\right]}{2p_3(\hat{p_3}\cdot\hat{p_5}-\hat{p_4}\cdot\hat{p_5})} \end{vmatrix} \tag{C34}$$

7. Sampling from a Polynomial S_{cm} distribution

*** This is where I am taking a function from Aurelio and I can't seem to derive it on my own ***

The distribution of the variable S_{cm} according to a polynomial power distribution is

$$s = m_0^2 + [(1-\alpha)r]^{\frac{-1}{\alpha-1}} \tag{C35}$$

where r is defined in terms of the random variable, u, that is uniformaly distributed between 0 and 1.

$$r = (r_{max} - r_{min}) \times u + r_{min} \tag{C36}$$

where r_{max} and r_{min} are defined in terms of the variable s_{cm}.

$$r = \frac{1}{1-\alpha} \times \left[s - m_0^2\right]^{1-\alpha} \tag{C37}$$

$$r_{min} = \frac{1}{1-\alpha} \times \left[s_{min} - m_0^2\right]^{1-\alpha} \tag{C38}$$

$$r_{max} = \frac{1}{1-\alpha} \times \left[s_{max} - m_0^2\right]^{1-\alpha} \tag{C39}$$

where alpha can not equal 1.

8. Jacobian for random sampling of a polynomial distribution starting at m_{pole}

The first case to consider is sampling around a falling polynomial distribution for the mass squared of two quarks in the event, s_{56}. We need to define the Jacobian with respect p_5 or

p_6. Since s_{56} is invariant under an interchange of particle 5 and 6, the Jacobian will be the same for each momentum integration. The following assume p_5 will be replaced with the variable, u, which is sampled from a polynomial distribution.

$$|J(p_5, u)| = \left|\frac{\partial p_5}{\partial u}\right| \tag{C40}$$

Because u is redined in terms of the variable r, we can rewrite C40 in terms of r instead of u.

$$|J(p_5, u)| = \left|\frac{\partial p_5}{\partial u}\right| = \left|\frac{\partial p_5}{\partial r} \times \frac{\partial r}{\partial u}\right| \tag{C41}$$

And since the variable r is sampling the S_{56} distribution it makes sense to define the Jacobian in terms of this variable instead of p_5.

$$|J(p_5, u)| = \left|\frac{\partial p_5}{\partial r} \times \frac{\partial r}{\partial u}\right| = \left|\frac{\partial p_5}{\partial s_{56}} \times \frac{\partial s_{56}}{\partial r} \times \frac{\partial r}{\partial u}\right| = \left|\frac{\frac{\partial s_{56}}{\partial r} \times \frac{\partial r}{\partial u}}{\frac{\partial s_{56}}{\partial p_5}}\right| \tag{C42}$$

Equation C42 has three components: $\frac{\partial s_{56}}{\partial r}$, $\frac{\partial r}{\partial u}$, and $\frac{\partial s_{56}}{\partial p_5}$. From equation C36, the partial derivative of r with respect to u is

$$\frac{\partial r}{\partial u} = r_{max} - r_{min} = \Delta R_{56} \tag{C43}$$

Next, the partial of s_{56} with respect to r can be determined from equation C35.

$$\frac{\partial s_{56}}{\partial r} = [r(1-\alpha)]^{\frac{\alpha}{1-\alpha}} \tag{C44}$$

Inserting the value of r(s) as defined in equation C37, equation C44 can be re-written as

$$\frac{\partial s_{56}}{\partial r} = \left[s_{56} - m_0^2\right]^{\alpha} \tag{C45}$$

Finally, we need the partial derivative of s_{56} with repect to p_5. First, we define s_{56}

$$s_{56} = m_5^2 + m_6^2 + 2E_5E_6 - 2p_5^x p_6^x - 2p_5^y p_6^y - 2p_5^z p_6^z \tag{C46}$$

we can evaluate the partial derivative of s_{56} with respect to p_5.

$$\frac{\partial s_{56}}{\partial p_5} = 2p_6(1 - \hat{p_5} \cdot \hat{p_6}) \tag{C47}$$

Finally, we can rewrite the Jacobian defined in C42 as

$$|J(p_5,u)| = \left| \frac{\frac{\partial s_{56}}{\partial r} \times \frac{\partial r}{\partial u}}{\frac{\partial s_{56}}{\partial p_5}} \right| = \frac{\Delta R_{56} \times [s_{56} - m_0^2]^\alpha}{2p_6(1 - \hat{p_5} \cdot \hat{p_6})} \tag{C48}$$

and

$$|J(p_6,u)| = \left| \frac{\frac{\partial s_{56}}{\partial r} \times \frac{\partial r}{\partial u}}{\frac{\partial s_{56}}{\partial p_6}} \right| = \frac{\Delta R_{56} \times [s_{56} - m_0^2]^\alpha}{2p_5(1 - \hat{p_5} \cdot \hat{p_6})} \tag{C49}$$

[1] V.M. Abazov *et al.*,(DØ Collaboration), "A Precision Measurement of the Mass of the Top Quark," Nature **429**, 638 (2004).
[2] V.M. Abazov *et al.*, (DØ Collaboration), "Measurement of the Top Quark Mass in the Lepton+Jets Final State with the Matrix Element Method," hep-ex/0609053.
[3] V.M. Abazov *et al.*, (DØ Collaboration), "Helicity of the W Boson in Lepton+Jets $t\bar{t}$ Events," Phys. Lett. **B 617**, 1 (2005).
[4] The Single Top Working Group, "Search for Single Top Quark Production in 1 fb^{-1} of Data," DØ Note 5285 (2006).
[5] The GNU Scientific Library, `http://www.gnu.org/software/gsl/`
[6] G.P. Lepage, "VEGAS: An Adaptive Multidimensional Integration Program," Cornell Laboratory of Nuclear Sciences report CLNS-80/447, (1980).
[7] J. Pumplin *et al.*, "Parton Distributions and the Strong Coupling: CTEQ6AB PDFs," JHEP **0602**, 032 (2006).
[8] The Les Houches Accord PDF Interface, `http://hepforge.cedar.ac.uk/lhapdf/`
[9] W.-M. Yao *et al.*, "The Review of Particle Physics," J. Phys. **G 33**, 1 (2006).
[10] F. Maltoni and T. Stelzer, "MadEvent: Automatic Event Generation with MadGraph," JHEP **0302**, 027 (2003).
[11] P. Shiefferdecker and M. Wang, "Jet Transfer Functions Derived From p17 $t\bar{t}$ Monte Carlo," DØ Note 5136, (2006).
[12] Private communication with L. Wang. Parametrization found in the file `/work/cole-clued0/leiwang/wz_epmcs/p170303_sampling/wz_epmcs/src/pmcsana.cpp`
[13] P. Haefner and F. Fiedler, "Determination of the Muon Transfer Function for Top Mass Measurements," DØ Note 4818, (2005).
[14] P. Haefner and F. Fiedler, "Muon Transfer Function Parameters for p17 MC," DØ Note 5214, (2006).
[15] M. Anastasoaie, S. Robinson, and T. Scanlon, "Performance of the NN b-Tagging Tool on p17 Data," DØ Note 5213, (2006).