2011

observation of t-channel electroweak top quark production

Nathan Edward Triplett
Iowa State University

Recommended Citation

Triplett, Nathan Edward, "observation of t-channel electroweak top quark production" (2011). *Graduate Theses and Dissertations*. 11969.
http://lib.dr.iastate.edu/etd/11969

Observation of t-channel electroweak top quark production

by

Nathan Triplett

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

Major: High Energy Physics

Program of Study Committee:
James Cochran, Major Professor
Soeren Prell
John Lajoie
Kerry Whisnant
Matthew Stanley

Iowa State University
Ames, Iowa
2011

Table of Contents

List of Tables

List of Figures

Acknowledgments

I would like to thank my advisor, Jim Cochran, for his guidance throughout my graduate career. Jim is also responsible for introducing me to the field of high energy physics. Thank you, Jim, for all of your help over the last seven years.

I would like to thank the members of my committee for reading and providing suggestions on my thesis: Soeren Prell, John Lajoie, Kerry Wisnant, and Matthew Stanley.

I would also like to thank the members of the single top group at DØ: Victor, for the long discussions about statistics; Liang, for keeping us focused on the big picture; Mark, for collaborating on the BDT analysis with me; Jyoti, Yun-Tse, and Weigang, for all your work on the selection, systematics, and other MVA analysis; Reinhard, for getting me connected with the DØ single top group in the first place; and Cecilia, Ann, and Aran, for the experience you brought to the group. It has been a pleasure working with all of you on this analysis.

I would also like to thank the many friends and family members who helped proofread this thesis. You made the final review process go much smoother.

Finally, I would like to thank my parents for their love and support throughout all of the years I have been in school. Thank you for always encouraging me in my scientific endeavors.

Abstract

The top quark is the heaviest known fundamental particle, with a mass of $172.0\ ^{+0.9}_{-1.3}$GeV [1]. This is nearly twice the mass of the second heaviest known particle, the Z boson, and roughly the mass of a gold atom. Because of its unusually large mass, studying the top quark may provide insight into the Higgs mechanism and other beyond the standard model physics.

Only two accelerators in the world are powerful enough to produce top quarks. The Tevatron, which first accelerated protons in 1983, has produced almost 400,000 top quarks, roughly half at each of its two detectors: DØ and CDF. The LHC is a much newer accelerator which currently has accumulated about 0.5% as much data as the Tevatron. However, when running at full luminosity, the LHC is capable of producing a top quark about once every second and will quickly surpass the Tevatron as the leading producer of top quarks. This analysis uses data from the DØ detector at the Tevatron, which are described in chapter 3.

Top quarks are produced most often in pairs of top and anti-top quarks through an interaction of the strong force. This production mode was first observed in 1995 at the Tevatron. However, top quarks can also be produced though an electroweak interaction, which produces just one top quark. This production mode was first observed at the Tevatron in 2008 [2]. Single top quark production can occur in different channels. In this analysis, a measurement of the cross section of the t-channel production mode is performed. This measurement uses 5.4 fb^{-1} of data and uses the technique of boosted decision trees in order to separate signal from background events. The t-channel cross

section is measured to be:

$$\sigma(p\bar{p} \rightarrow tqb + X) = 3.03\ ^{+0.78}_{-0.66}\ \mathrm{pb} \tag{0.0.1}$$

Additional cross section measurements were also performed for the s-channel as well as the $s + t$-channel. The measurement of each one of these three cross sections was repeated three times using different techniques, and all three methods were combined into a "super-method" which achieves the best performance. The details of these additional measurements are shown in appendix A.

Chapter 1. Introduction

All processes observed in nature can be described as interactions of four known forces: gravitational, electromagnetic, strong, and weak. The interactions of fundamental particles with the last three of these forces comprise the theory known as the standard model. While the standard model has been highly successful, it is only believed to be only an effective theory of a yet to be discovered more fundamental theory. Through the study of very massive particles produced in high energy accelerators, it is hoped that hints of this new physics will appear.

The top quark, being the most massive of all fundamental particles, is a natural candidate for study in this respect. While top quark pair production has been observed for 15 years, the observation of electroweak top quark production has only recently been achieved. Moreover, the observation of single top quark production was performed for both $s-$ and $t-$channel production modes combined. However, the rates of production in each channel can be affected differently by various beyond the standard model physics theories. This analysis is optimized to study the cross section of the t-channel single top quark cross section alone.

Chapter 2 contains an overview of the standard model, single top quark production, and the production of the various background contributions. Chapter 3 provides a description of the accelerator complex used to accelerate and collide protons and antiprotons, as well as a description of the DØ detector. Chapter 4 describes the process by which electronic signals from the detector are reconstructed into physics objects. The following two chapters discuss the data collected, the Monte Carlo simulation, and the criteria used to select events from both of these datasets. Chapter 7 contains a list of all systematic uncertainties considered in the measurement of the cross section.

After the event selection, advanced statistical techniques are employed to separate signal events from the background. The method employed is that of Boosted Decision Trees (BDTs), which is described in chapter 8. Chapter 9 describes how the output from the BDTs are used to perform the measurement of the cross section and the associated uncertainties.

Chapter 2. Theory

2.1 Standard model

2.1.1 Introduction

In 1864, Maxwell first put forth the idea that light was related to the electromagnetic force [3]. Over the next century, physics has not only discovered that the photon is the particle which mediates the electromagnetic force, but has also discovered two additional forces, the strong and weak forces, and the particles which mediate them. The Standard Model is the theory which describes how these force mediating particles interact with all other known fundamental particles. Experimental particle physics is the field of study concerned with measuring the many predictions of this theory. To date, very few experimental measurements can not be explained by the Standard Model, the most important of which are: gravity, dark matter, dark energy, and the non-zero neutrino mass.

2.1.2 Particles

The fundamental particles described by the standard model can be split into two subsets. Particles which carry a half integer spin are classified as fermions and can be further subdivided into quarks and leptons. The other subset consists of particles with integer spin and these are classified as bosons. These bosons serve as the mediator particles of the electromagnetic, weak, and strong forces. Gravity, though outside the scope of the standard model, is theorized to be mediated by a boson with spin 2.

There are three versions of each quark and lepton type, which differ only by their mass. The standard model classifies each of these mass groups into families. In each family there are two quarks and two leptons. One quark contains a charge of +2/3 and the other contains a charge of -1/3 and two leptons. The first lepton carries a charge of +1, while the second lepton is neutrally charged and is referred to as a neutrino. Neutrinos carry a mass that is many orders of magnitude less than the other fundamental particles, yet is non-zero. Due to carrying no electromagnetic charge, neutrinos only interact through the weak force and are therefore hard to detect. For each particle, there is also a corresponding anti-particle. These anti-particles have the same mass but oppositely signed charge. An anti-top quark, for instance, has a charge of -2/3.

Table 2.1.2 shows each of the fundamental fermions along with their respective properties.

	Charge	Family 1		Family 2		Family 3	
		Name	Mass	Name	Mass	Name	Mass
u-type quark	+2/3	up	7.5	charm	1100	top	173000
d-type quark	-1/3	down	4.2	strange	150	bottom	4200
lepton	-1	electron	0.51	muon	105	tau	1784
neutrino	0	ν_e	<0.0002	ν_μ	<0.19	ν_τ	<18.2

Table 2.1 Fermions in each of the three families and its associated properties [1]. Masses quoted in MeV.

2.1.3 Interactions

Fermions interact by exchanging spin-1 bosons between themselves. The three different forces each correspond to different bosons. The electromagnetic force is mediated by an exchange of massless photons. Exchanges of gluons, which are also massless, mediate the strong force, while the weak force is mediated by the exchange of either W^+, W^- or Z^0 bosons. Figure 2.1 shows all of the particles and their possible interactions.

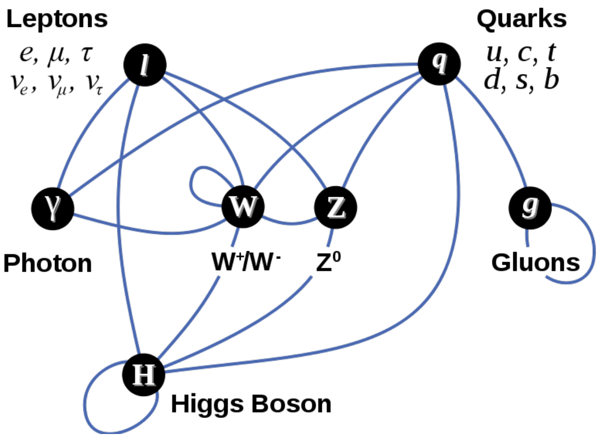


Figure 2.1 Each fundamental particle type. Blue lines represent particles that interact with each other at tree level.

The mediators of the weak force, the W and Z Bosons, are of special note as they have masses of 80.4 GeV and 91.2 GeV respectively [1]. If electroweak symmetry was an exact symmetry, the mediators of the weak and electromagnetic forces (the photon and the W/Z bosons, respectively) would all be massless. Since this is not the case, the electroweak symmetry must be a broken symmetry.

In the standard model, this symmetry breaking is achieved via the Higgs Mechanism. However, this theory implies the existence of one more massive boson, the Higgs Boson. Physicists are still trying to observe this particle.

2.2 Top quarks

The top quark is the heaviest of the six quarks, and the last one to have been discovered. Since each family of fermions contains two quarks, when the b-quark was discovered in 1977 with a charge of -1/3, the standard model required a partner quark with charge of +2/3, the top quark. This quark was finally discovered in 1995 by the DØ and CDF collaborations [4].

The top quark is the up-type quark in the third family of fermions and shares similarities to the up quark and charm quark. All three have a charge of +2/3, spin of 1/2, and interact via the strong, weak, and electromagnetic forces. The top quark, however, has a mass of 172 GeV, which is roughly 40 times heavier than the next heaviest quark and about 50,000 times heavier than the up quark.

This large mass implies a very short lifetime of ~5×10^{-25} s before it decays into a b-quark and W boson. All other free quarks undergo hadronization on a timescale of about 3×10^{-24} s. The top quark is unique in that it decays before this can happen, allowing us to measure the properties of the free quark.

2.3 Top quark pair production

In the Tevatron, protons collide with anti-protons at 1.96 GeV. Top quark pairs are produced in these collisions though several different production modes, as illustrated in Figure 2.2 . These production modes are split according to the initial states. Quark-antiquark ($q\bar{q}$) production is shown on the right of Figure 2.2, while the various gluon-gluon fusion production modes are shown on the left and center. At the Tevatron $q\bar{q}$ production dominates gluon-gluon fusion, producing about 85% of all $t\bar{t}$ pairs. DØ has measured the $t\bar{t}$ cross section to be $7.78\ ^{+0.77}_{-0.64}$pb at a center of mass energy of 1.96 GeV [5].

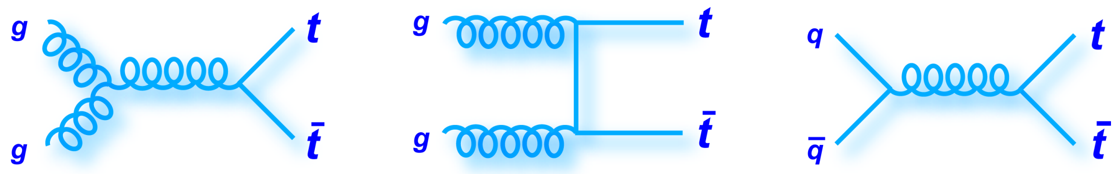


Figure 2.2 Top quark pair production modes. Gluon-gluon fusion is shown on the left and center. Quark-antiquark annihilation is shown on the right.

Top quarks have only been observed to decay into a b-quark and W boson. The b-quark undergoes hadronization and can be measured as a b-tagged jet. The W boson either decays into a pair of lighter quarks or into a charged lepton and neutrino. At first order, the W has a 1/9 chance of decaying leptonically into each of the three types of charged lepton with its associated neutrino and a 2/3 chance of decaying hadronically into a pair of light quarks which appear as light jets. Higher order corrections change these rates slightly. Table 2.2 shows the experimentally measured branching ratios.

Decay mode	Branching Ratio
$W \to e\,\nu_e$	10.75%
$W \to \mu\,\nu_\mu$	10.57%
$W \to \tau\nu_\tau$	11.25%
$W \to$ hadrons	67.60%

Table 2.2 Experimentally measured branching ratios for the W boson [1].

A pair of top quarks will decay into two b-quarks and two W bosons. Top quark pair events are categorized by the decay of the two W bosons as shown in Figure 2.3.

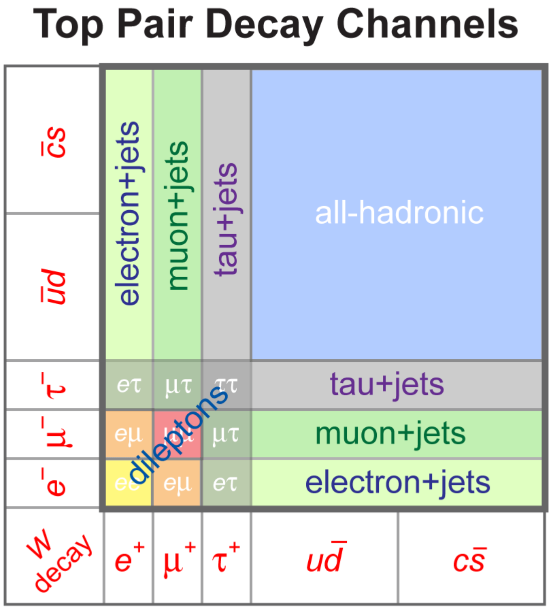


Figure 2.3 Top quark pair decay channels. Each top quark decays into a W boson and b-quark. The decay modes of the first W boson are on the horizontal axis and the second are on the vertical axis.

The di-lepton case occurs when both W's decay into electrons, muons, or taus. This occurs about one out of nine times. The measurement of this decay mode is complicated by the presence two neutrinos which only appear as an asymmetry in the energy deposition. Two neutrinos become statistically difficult to dis-entangle, and measuring the total energy of an event is a challenge. However, this mode is still important because the backgrounds for this channel are significantly smaller than for the other decay channels.

The all-jets case occurs when both W's decay hadronically and this happens about four out of nine times. This channel suffers from very large backgrounds as well as difficulty in correctly matching the light jets to the appropriate top quarks.

The lepton+jets case occurs when one of the W's decays into a lepton and the other decays hadronically. This also occurs about four out of nine times. The lepton+jets channel is the easiest channel to measure top quark properties. This is because there is exactly one lepton, one neutrino, a pair of b-tagged jets, and a pair of light quark jets coming from the top decays resulting in a clean signal with manageable backgrounds.

The top quark discovery was performed with both di-lepton and lepton + jets events [4]. As is discussed in the following section, the single top quark analysis only considers the case where the W Boson decays leptonically, however, all three top quark pair decay modes are important backgrounds to this process.

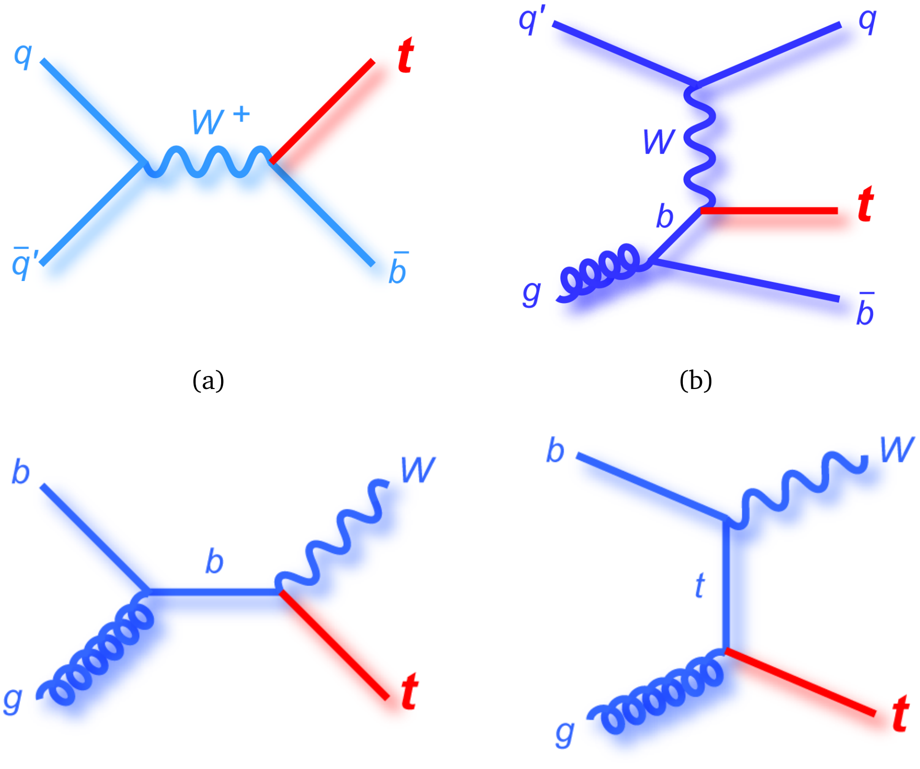
2.4 Electroweak top quarks

While most top quarks produced at the Tevatron are produced in pairs via the strong force, top quarks can also be produced via the electroweak interaction. Top quarks produced in this way are often referred to as single top quarks as only one top quark is produced.

The cross section for single top quark production is smaller than that of top pair production at the Tevatron. Additionally, single top quark events have a much larger background due to their lower jet multiplicity. These two factors made the search for the single top quark far more difficult than that of top quarks produced via the strong force. The DØ and CDF collaborations first announced the observation of the single top quark in 2009 [2]

2.4.1 Single top quark production

Single top quarks are produced via three distinct interactions, or channels: the s-channel, the t-channel, and the Wt-channel, as shown in Figure 2.4.



(a)

(b)

(c)

Figure 2.4 Single top quark production modes: s-channel (a), t-channel (b), Wt-channel (c).

2.4.1.1 s-channel

In the s-channel production mode, an up-type quark(anti-quark) interacts with a down-type anti-quark(quark) resulting in a top(anti-top) quark and anti-b(b) quark though the exchange of a time-like W boson. Figure 2.5 shows the theoretical Next Next to Leading Order (NNLO) cross section for this process at the Tevatron as a function of top quark mass. This analysis assumes a top mass of 172.5 GeV which results in a cross section of 1.04 pb [6]. As the initial state of this process is two quarks, it is highly suppressed at the LHC where gluon initial states dominate. Due to this, the Tevatron will have the best chance of observing s-channel single top quarks until the LHC has been running for many years.

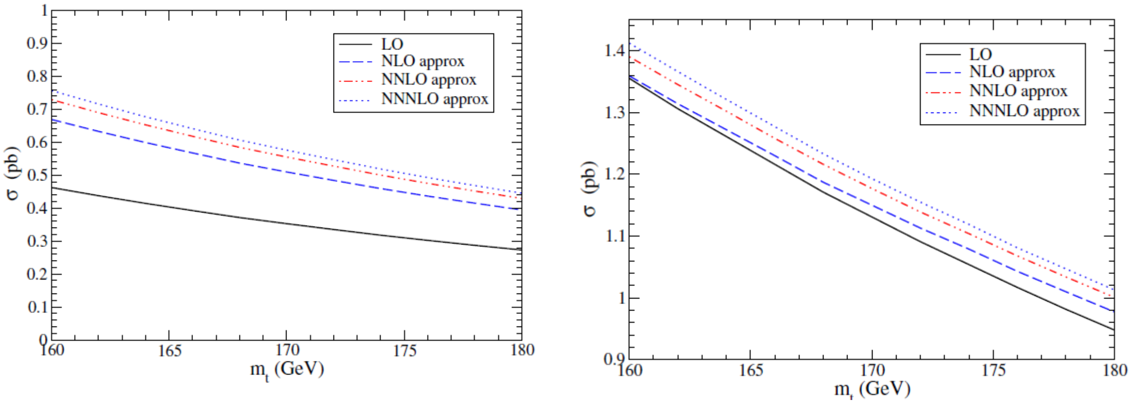


Figure 2.5 Single top quark cross section for s-channel (left) and t-channel (right) as a function of top quark mass [6]. The plots show the cross sections for top quark production. The cross section for anti-top quark production is identical at the Tevatron. Cross section values quoted throughout the text are for top and anti-top quark production combined.

2.4.1.2 t-channel

The t-channel production mode has the largest single top cross section at both the Tevatron and the LHC. In this production mode, a b-quark exchanges a space-like W boson with a light quark resulting in a top quark and another light quark. Due to the rarity of finding a b-quark in the sea of a proton, the leading order contribution comes when a gluon decays into a b and $\bar{b}$ quark pair and this b-quark exchanges the W boson with the light quark. Figures 2.5 show the theoretical NNLO cross section for this process. Assuming a top quark mass of 172.5 GeV, the cross section is predicted to be 2.26 pb at the Tevatron [6]. The extra light quark in this channel often recoils softly and ends up in the forward region of the detector. This forward light quark jet provides a unique handle for identifying the various single top channels.

2.4.1.3 Associated Wt production

The final single top production mode, associated Wt production, occurs when an initial gluon and b-type quark exchange a b-quark or top quark. This results in a final state containing a W boson and top quark as shown at the bottom of Figure 2.4. Due to the initial state gluon and b-quark, this mode is highly suppressed at the Tevatron and has a cross section of 0.28 pb [6]. At the LHC this production mode becomes important with a cross section of 66.5 pb at a center of mass energy of 14 TeV[7] (this compares to an s-channel cross section of ~11 pb and a t-channel cross section of ~247 pb [8]). Due to the small cross section of this channel at the Tevatron, this process is assumed to be negligible and is ignored for this analysis.

2.4.2 Motivation to measure single top

In addition to the interest in observing single top quark production in the s- and t-channels individually, single top quarks provide a window that can be used to view several predictions of Beyond the Standard Model (BSM) physics. A few of these are described below.

2.4.2.1 $\mathrm{V_{tb}}$ measurement

The Cabbibo-Kobayashi-Maskawa (CKM) matrix contains the strength of the flavor changing weak interactions between quarks. Assuming three generations of quarks and the unitarity of the matrix, the following equation holds [1]:

$$|V_{ub}|^2 + |V_{cb}|^2 + |V_{tb}|^2 = 1 \tag{2.4.1}$$

The values of V_{ub} and V_{cb} have been measured with high precision, however the value of V_{tb} has a considerably larger uncertainty and is traditionally calculated by using equation 2.4.1

Single top quark production involves the Wtb vertex which provides a factor of:

$$\frac{-ig_w}{2\sqrt{2}} V_{tb} \gamma^{\mu} (1 - \gamma^5) \tag{2.4.2}$$

to the single top production matrix element. This means that the single top quark cross section is proportional to $|V_{tb}|^2$ which allows for a direct measurement of $|V_{tb}|$ without the assumption of the unitarity of the CKM matrix.

2.4.2.2 Polarization

As top quarks decay without hadronizing, the properties of the top quark are transmitted through to its decay products. In the standard model, the Wtb vertex is left-handed which results in polarized single top quark production.

This polarization affects the angular distributions of those objects which decay from the top quark. In this analysis, these angular correlations are exploited to help identify single top events. However, with the large number of events expected at the LHC, the shape of these angular distributions will be measurable. This will allow an examination into the structure of the Wtb vertex.

2.4.2.3 Other new physics

In addition to introducing a change in the angular distributions of single top decay products, several Beyond the Standard Model physics processes can affect the cross sections of the various single top quark production channels.

Additional heavy charged bosons, such as a W' or charged Higgs will increase the s-channel cross section. Flavor Changing Neutral Current (FCNC) interactions will serve to increase the cross section of the t-channel production. Many other BSM theories predict changes in the single top quark cross sections, thus making an accurate measurement of both s- and t-channel production rates an important check for these theories.

2.4.3 Single top quark signature

At the Tevatron roughly 1 out of every 10^{10} events contains a single top quark, creating a significant challenge in discovering these events. However, single top quark events have some unique kinematic features which allow the background size to be reduced.

Single top quarks decay in the same way as pair produced top quarks: into a W boson and b-quark. As shown in table 2.2, the W boson can decay into light quarks or a charged lepton and neutrino. For this study, only the cases where the W boson decays into an electron or muon and its associated neutrino are considered.

The W boson decay results in a final state lepton, a large amount of missing energy, and a b-quark. Single top events also have an additional b-quark in the final state and the t-channel has a forward light quark. All quarks (except top quarks) appear as tightly packed clusters of particles in the detector known as jets. Since these interactions are quark-quark or quark-gluon, the remaining quarks in the proton are scattered and may produce other jets in the event.

Due to the high mass of the top quark, it is produced nearly at rest in the center of mass reference frame. Since its decay products are much lighter than the top quark, they become highly boosted and have no highly preferred direction. This results in many of the final state particles and jets having large amounts of momentum in the direction transverse to the beamline (p_T). Therefore, candidate single top events are events which contain a high p_T electron or muon, 1-2 high p_T b-tagged jets, 1-3 high p_T light jets, and a large missing transverse energy ($\not\!\!E_T$).

2.5 Background production

In order to measure the cross section of single top quarks, it is necessary to count the number of these events that have been produced. The previous section provides a strategy for selecting top quark events, however, many other processes can produce events which resemble single top quark events. These processes are referred to as background processes, and it is important to understand how likely each background is to contribute to single top like events.

2.5.1 Top pairs ($t\bar{t}$)

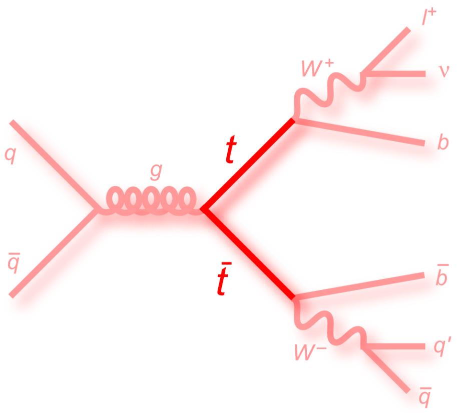


Figure 2.6 Diagram showing leading order top quark pair production and decay.

One of the largest backgrounds to single top events is top pair production, as described in Section 2.2. Both top quarks decay into W bosons and b-quarks. Only the case where one or both of the W bosons decay leptonically are considered as the number of all hadronic events passing the lepton and $\not{E}_T$ cuts is negligible. This process is the main background for events with high jet multiplicity.

In the case where one W boson decays into a lepton and neutrino and the other decays hadronically, this event resembles a single top event with slightly higher jet multiplicity. Additionally, jets which are close to each other can be difficult to distinguish from a single large jet, causing the event to fake a lower jet multiplicity.

In the case where both W bosons decay leptonically, the final state products are two b-tagged jets, two leptons, and a large $\not{E}_T$. If one of the leptons is mistakenly identified as a jet, this event mimics a single top event and passes the selection criteria.

2.5.2 W + jets

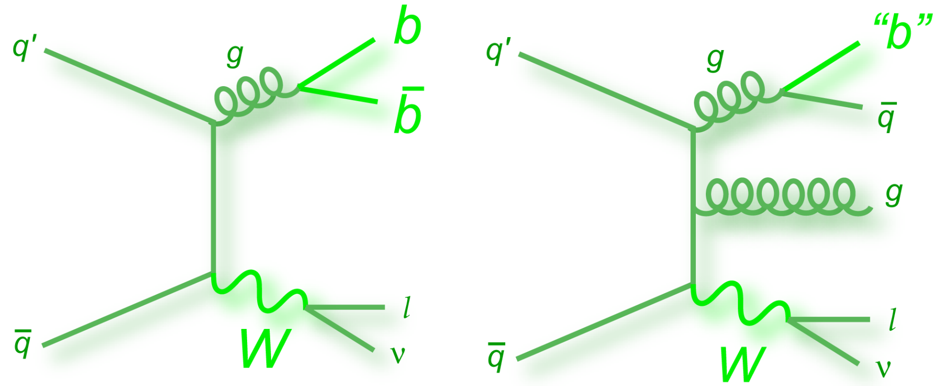


Figure 2.7 Some diagrams for leading order W + jets production modes. $W{+}bb$ is shown on the left, and W light jets is shown on the right.

Figure 2.7 shows the Feynman diagram for W Boson plus jets production. The leptonic decay of the W Boson produces a high p_T lepton and a large $\not{E}_T$. In order to pass the section cuts, one or two of the jets must be tagged as containing a b-jet. This can occur when some of the jets actually contain b-quarks, as in Figure 2.7 on the left, or when the algorithm identifying b-jets miss-classifies one of the jets, as shown in Figure 2.7 on the right. This process is the main background for events with low jet multiplicity.

2.5.3 Z + jets

Figure 2.8 shows the production of a Z boson plus jets. If the Z boson decays into a pair of electrons or muons and one of those leptons is not measured by the detector

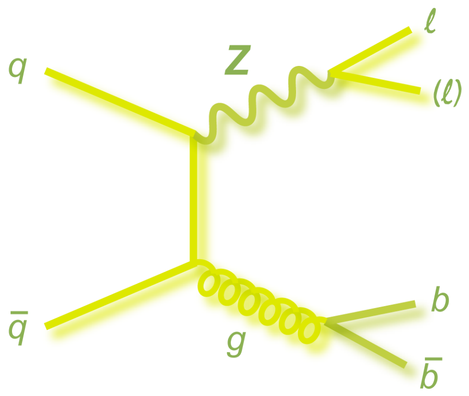


Figure 2.8 Example diagram for leading order Z + jets production.

it will appear as missing energy. Along with the other lepton this satisfies the first half of the event selection criteria. Similar to the W + jets background, if one of the jets contains a b-quark or is misidentified as containing one, the event mimics a single top event and will pass the event selection. For a Z + jets event to pass the selection, however, both of these unlikely events are required to occur at the same time (not constructing a lepton and identifying a b-quark). As a result, this background is less important than the $t\bar{t}$ and W + jets backgrounds.

2.5.4 QCD multijets

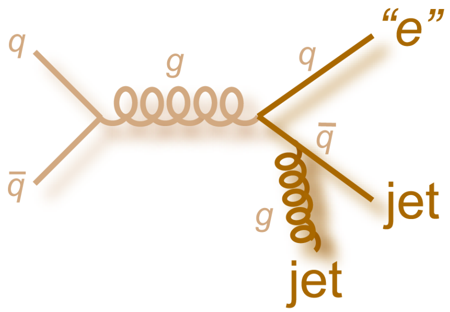


Figure 2.9 Example diagram for QCD multijets production

Figure 2.9 shows multijet production (sometimes referred to as QCD production).

This process creates events containing jets exclusively. In order to pass the event selection, one of the jets needs to be misidentified as a lepton, some of the energy in the jets needs to be missed, and one of the jets needs to be identified as containing a b-quark. While it is very rare for all three of these processes to occur at the same time, the cross section of this process is many orders of magnitude larger than that of the other backgrounds. As a result, enough of these rare occurrences pass the event selection to make a measurable contribution.

2.5.5 Diboson

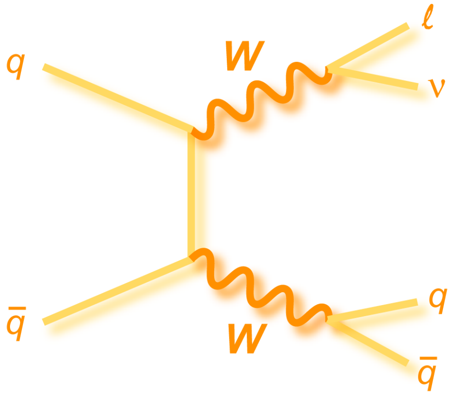


Figure 2.10 One example of diboson production is shown in this WW production diagram

Figure 2.10 shows a diagram for WW production. ZZ and WZ production can occur in a similar way. These three processes together are referred to as diboson production. When one of the bosons decays leptonically, one decays hadronically, and some of the jets are classified as containing b-quarks, the events will pass the event selection. The cross sections for these processes are not as large as those containing just a single W or Z boson, therefore this background is much smaller and less important than the others.

Chapter 3. Accelerator & Detector

3.1 Accelerators

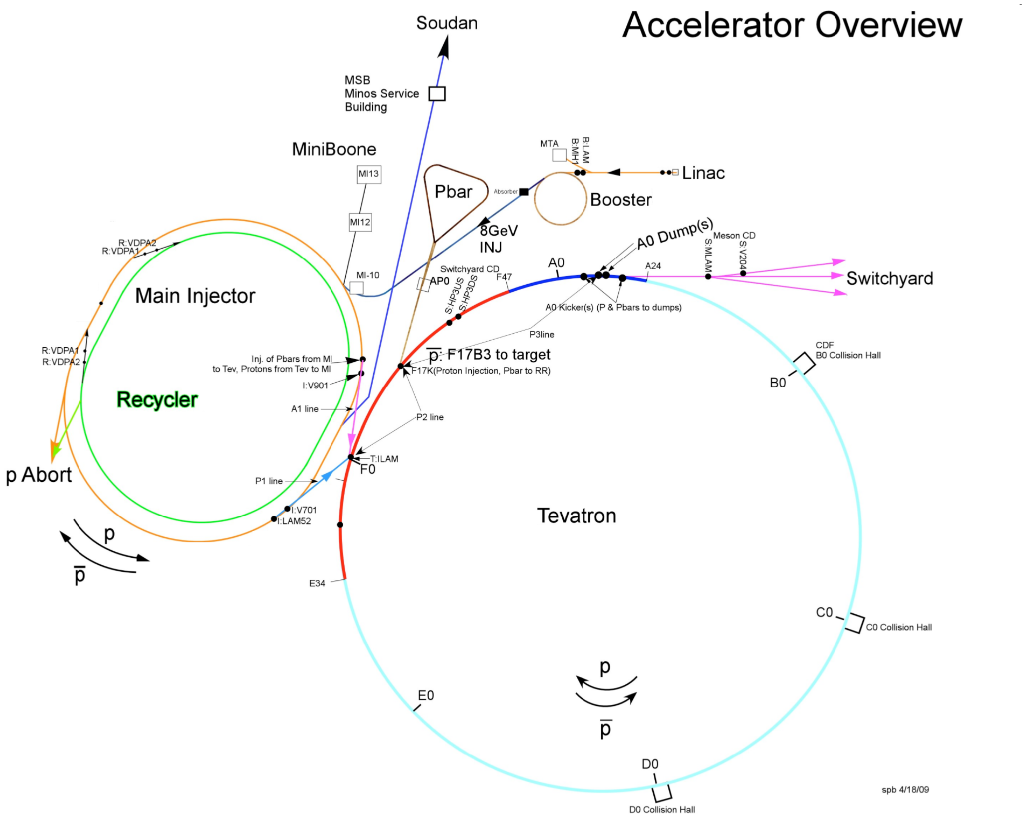


Figure 3.1 Schematic showing all of the accelerators at Fermilab [9].

The study of heavy quarks and leptons requires a high luminosity, high energy accelerator to act as a source of these particles. At the Fermi National Accelerator Laboratory, the Tevatron collides beams of highly energetic protons and anti-protons at a center of mass energy of 1.96 TeV. In these collisions all of the known quarks and leptons can be produced. The collisions take place at two points in the Tevatron which each have large detectors built around them. These detectors, CDF and DØ, perform measurements of the charge, energy, momentum, and position of each particle created in these collisions.

In order to produce and accelerate beams of protons and anti-protons to near the speed of light, a chain of five accelerators, and two storage rings are interconnected. Hydrogen ions are first generated and accelerated though the preaccelerator and then through the linear accelerator. After leaving the linear accelerator, the extra electrons are stripped away as the protons enter the first synchrotron, Booster. After being accelerated to 8 GeV in Booster, the protons are transfered to the Main Injector, a 150 GeV synchrotron. The main injector can then transfer the protons to the Tevatron or into the anti-proton production facility. The following sections briefly discuss each of the parts of the entire accelerator complex.

3.1.1 Preaccelerator

Fermilab has two 750 kV electrostatic preaccelerators which provide a beam of H- ions to the accelerator complex. The preaccelerators each consist of a H- source, a 750 kV potential provided by a Crockoft-Walton generator [10], and an accelerating column. Two preaccelerators are used to provide redundancy, and are named H- and I- respectively.

In order to produce H- ions, hydrogen gas is injected into the proton source where it experiences a large electromagnetic field and forms a dense plasma as shown in Figure 3.2. As energetic particles strike the cathode, there is a chance H- ions will be released from the surface. These ions are extracted through an 18 keV electric potential resulting in a beam current of 50 mA. This beam is then bent through a magnetic field in order to select H- ions and eliminate any other particles that are contaminating the beam. The next accelerator in the chain, the Linac, requires 15 bunches of H- ions per

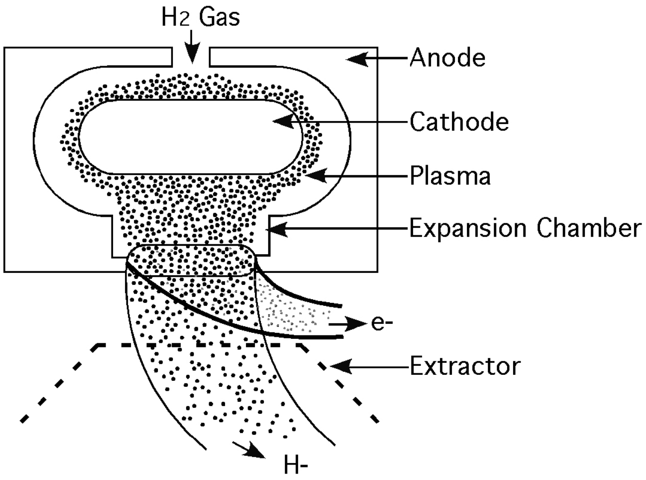


Figure 3.2 H Ion source. Hydrogen molecules are added (top), form a dense plasma, and strike the anode. They then become H- Ions (center) and are extracted through into the preaccelerator (bottom) [10].

second, so the proton source is pulsed at 15 Hz with each pulse lasting 80 μs.

The ion source is located in an electrically isolated dome that is held at -750 kV. This electric potential is produced using a Crockoft-Walton generator. The Crockoft-Walton generator is an AC-DC voltage multiplier which takes a 75 kV AC source and transforms it into 750 kV DC voltage. This is done by charging capacitors in parallel and discharging them in series. A network of diodes between the capacitors (shown in Figure 3.3) facilitate this resonating charging/discharging mode. The ion source in its dome, along with the Crockoft-Walton generator, is pictured below.

18 keV H- ions are accelerated from their starting potential at -750 kV to ground through an accelerating column. The column is constructed of seven electrically isolated disks which are held at roughly equal potential steps. This arrangement of disks provides a reasonably uniform electric field for the entire length of the accelerating column. After exiting the accelerating column, the beam passes through a chopper which cuts the beam into a bunch with a length of 40 μs. From the chopper the beam is passed through several focusing magnets before reaching the Linac.

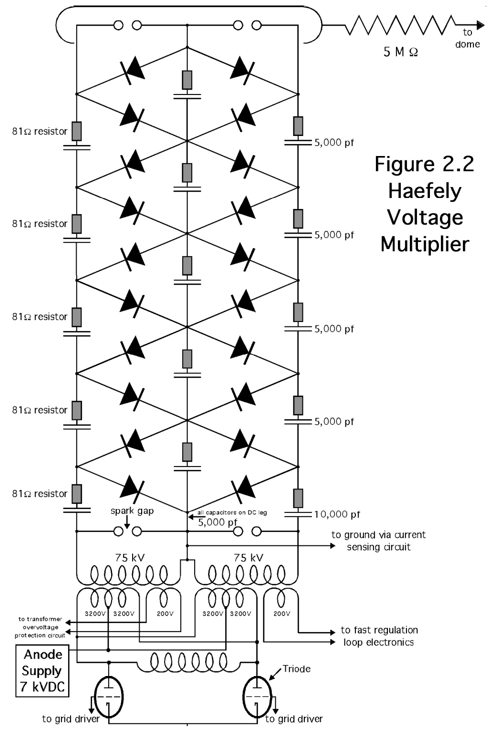


Figure 3.3 Crockoft-Walton circuit diagram. Used to transform 75 kV AC into 750 kV DC [10].

3.1.2 Linac

The Linac [10] is a linear accelerator which takes a 750 keV H- beam an accelerates it to 400 MeV before injecting it into Booster. The first stage of the Linac consists of a 79 m long Alvarez drift tube broken into five tanks, and it accelerates the beam to 116 MeV. This is followed by a 67 m long side-coupled Linac which accelerates the beam to the full 400 MeV through 7 modules.

Each tank in the Alvarez Linac consists of 23 to 59 individual drift tubes suspended in the center of the tank. A hole is bored through the center of each tube that allows the beam to pass through it. The drift tubes are arranged with a gap between them as shown in Figure 3.5. While the beam passes through the drift tube it is shielded from any external electric field; however, while passing through the gap the ions are able to be accelerated by the field.

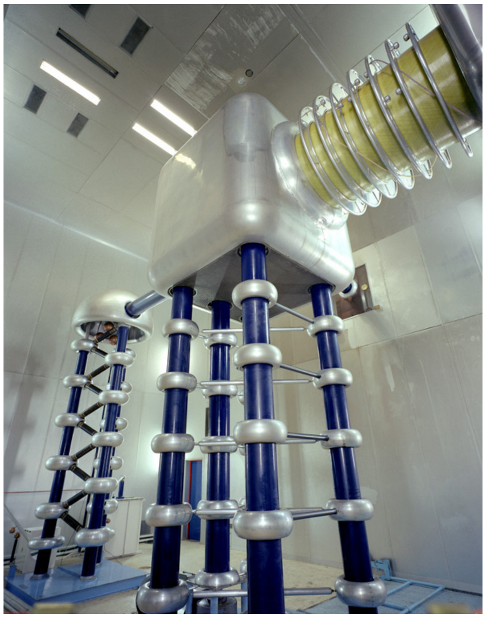


Figure 3.4 The Crockoft-Walton is visible in the background, and it is connected to the large cube shaped dome housing the H- source. The accelerating column is visible at the top right of the image [10].

Each of these tanks resonates at 201.24 MHz and is powered by a five MW RF source. The RF is phased such that the electrical field component is in the forward direction while each ion bunch is located in the gap between drift tubes. The electric field oscillates in the reverse direction while the beam is shielded from its force. The net effect is to provide a small amount of acceleration each time an ion bunch passes through a gap between drift tubes. Figure 3.6 provides a simplified sketch of this process.

The accelerating process is complicated by the fact that the bunches have non-infinitesimal longitudinal length. In this case, the front of the bunch will enter and exit the accelerating gap before the back portion of the bunch and see the RF field at a slightly different phase. If the RF phase is chosen so that the E field is largest when the bunch is centered in the accelerating gap, ions in the center of the bunch will receive more acceleration than those in the front or back of the bunch. This will cause the ions in the front to migrate towards the center but also cause the ions in the back of the

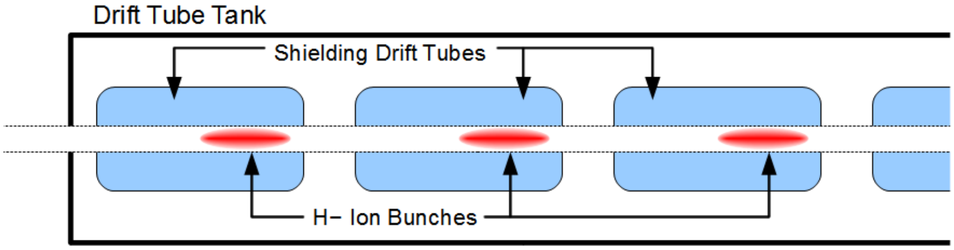


Figure 3.5 Simplified diagram of a drift tube accelerator. Ion bunches (shown in red) are shielded from the external electric fields while inside the drift tubes (light blue).

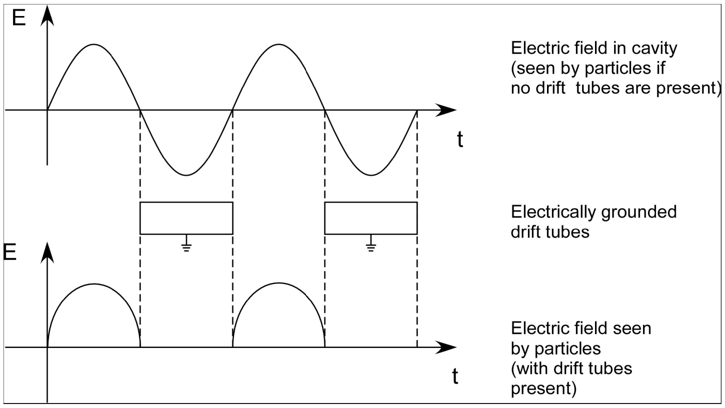


Figure 3.6 An alternating electric field is applied in each tank (top). The drift tubes shield the ions from this field while it points backwards. The overall field felt by the ions is always forward (bottom) [9].

bunch to drop further and further behind and eventually be lost. In order to minimize these losses the RF is phased at -32° from center. With this new phase, those ions at the back of the bunch receive the largest acceleration while those at the front receive the smallest resulting in stable operation as shown in Figure 3.7.

The side coupled Linac operates on the same principal as the drift tube Linac. In order to achieve higher acceleration per meter, however, the cells are modified to be more efficient. Unlike the drift tube Linac which consists of five large tanks each with dozens of small drift tubes, in the side coupled Linac each of the seven modules contains many smaller resonating cavities. These cavities are coupled such that each cavity

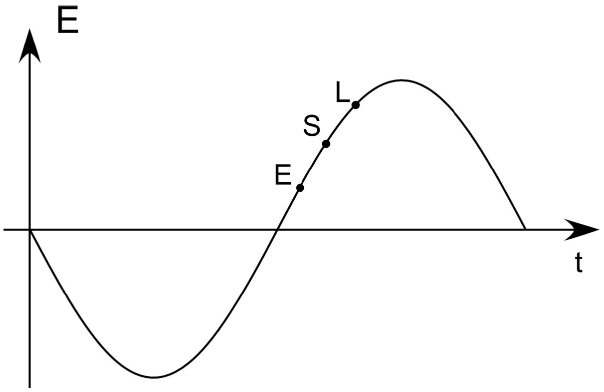


Figure 3.7 By phasing the electric field -32°,protons arriving early receive less acceleration than those arriving on time, and those that arrive late receive more. This has the effect of reducing the longitudinal bunch size [9].

is 180° out of phase with the adjacent cavities. An ion bunch enters one cavity and feels a forward accelerating field. This field reverses as the bunch reaches the next cavity where it again feels a forward accelerating field. As Figure 3.8 shows, the ion bunches are spaced such that there is a bunch in every fourth RF bucket. With this spacing all bunches will only feel a forward acceleration. For the side coupled Linac the RF power is supplied at 805 MHz by seven 10 MW klystrons.

3.1.3 Booster

After exiting the Linac, the H- ions are stripped of their electrons and injected into Booster, an 8 GeV synchrotron [11]. This ring is constructed of 96 dipole/quadrupole magnets and has a diameter of 151 meters. At a rate of 15 Hz, Booster is filled with 84 bunches of protons, accelerates them to 8 GeV, and extracts them to the Main Injector.

During injection into Booster, the accelerating RF cavities are tuned to be out of phase with each other such that there is no accelerating effect on the protons in the ring. The Linac supplies H- ions which are passed through a carbon foil located in a strong magnetic field. The foil strips off the electrons in most of the protons, and the magnetic field steers those protons which still retain an electron into a beam dump.

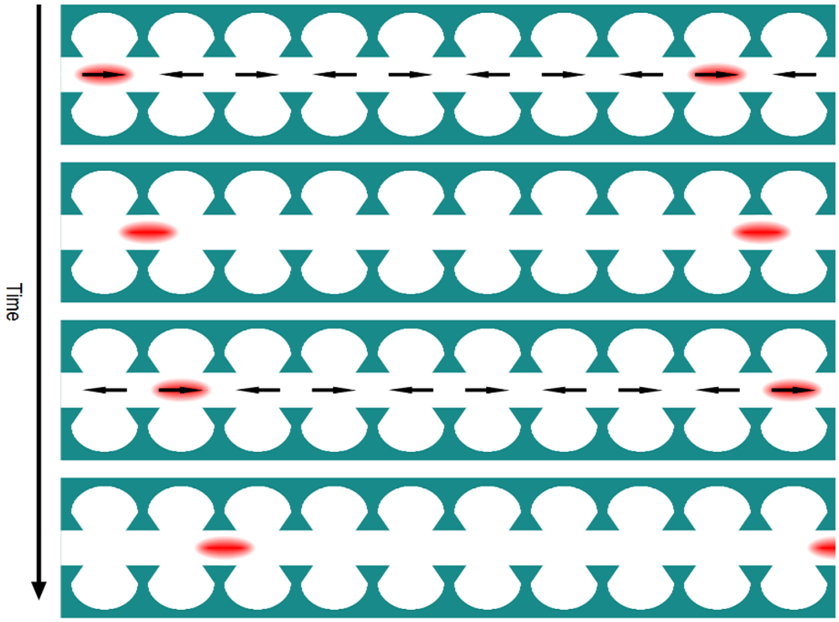


Figure 3.8 Diagram showing the progress of an ion bunch (red) though the side coupled Linac. Arrows show the direction of the electric field at the given time. Adjacent cells have fields pointing in opposite directions, and the field direction switches as the ion bunches pass between cells.

After the first protons have traveled a complete circuit of the ring, they are passed back through the carbon foil at the same time as new ions are introduced, as shown in figure 3.9. This has the result of merging the previous bunch with new ions, doubling its size. At injection energies, a proton will circle the Booster ring every 2.22 μs. The Linac provides bunches which are 40 μs long, allowing up to 18 bunches from the Linac to be merged into each other. However, the best performance typically comes when six sets of bunches are merged. These merged bunches are spaced around the Booster ring resulting in 84 new bunches, and each bunch is filled with roughly 3×10^{12} protons. At this point the magnetic field around the carbon foil is turned off so that the proton bunches no longer pass through it.

Over the next 100-200 μs, the 17 RF stations are brought into phase providing net acceleration to the proton bunches. The accelerating process takes about 33 ms during which the magnet currents must be ramped to full power while the RF frequency varies

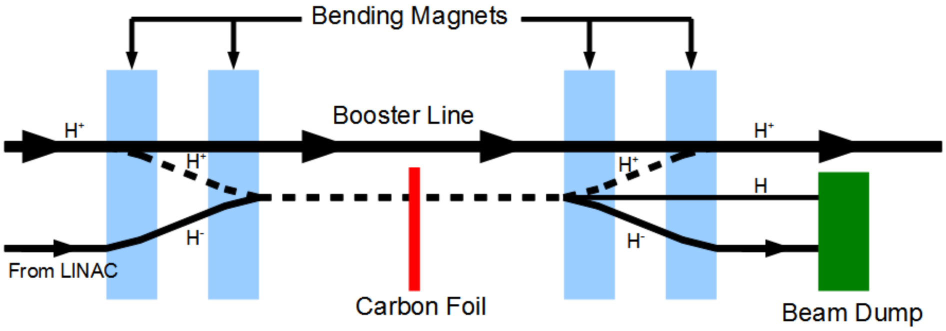


Figure 3.9 Diagram showing the path of ions in Booster and the path of ions during injection (dotted line) while the bending magnets are on. This causes both the H^+ and H^- ions to pass through the carbon foil.

from 37.9 MHz to 52.8 MHz. Once the beam is at full energy, fast "kicker" magnets are pulsed for roughly 1.6 μs, causing the beam to be extracted from Booster. The rise time for these kicker magnets is such that one of the 84 bunches is lost during the transition while the other 83 are passed to the Main Injector.

3.1.4 Main Injector

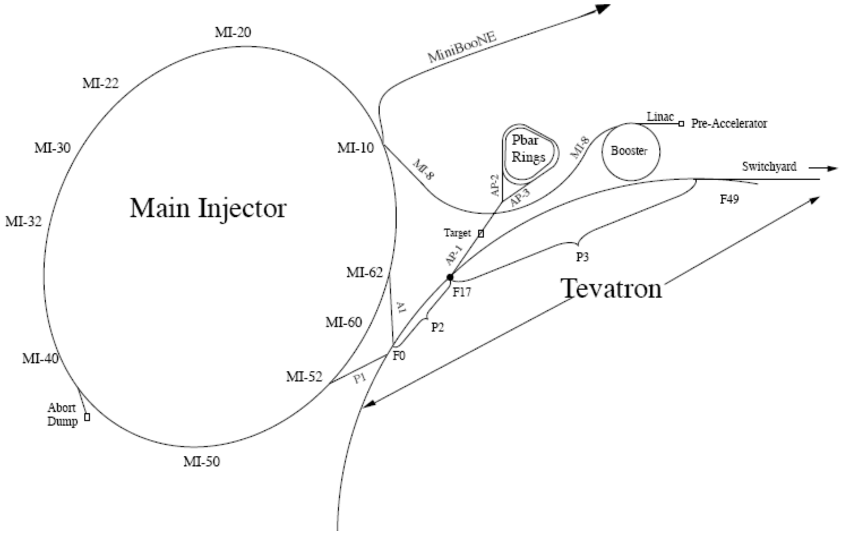


Figure 3.10 Location of the Main Injector relative to the rest of the accelerator complex. [12]

The Main Injector (shown in Figure 3.10) is a synchrotron with a circumference of over three km which accelerates protons or anti-protons from 8 GeV to 120-150 GeV depending on its mode of operation [12]. It takes about two seconds for the Main Injector to accelerate a batch of protons to its maximum energy before passing them off to one of several other experiments. In order to achieve this fast rise time, the dipole magnets in the Main Injector are constructed with a very small number of turns of wire. Therefore, to produce the required magnetic field, the wires must be capable of carrying a current of up to 9400 A. In addition to these dipoles, the Main Injector has numerous quadrupole, sextupoles, and octupoles designed to focus the beam.

While the Main Injector operates in several different modes for numerous experiments at Fermilab, two of these modes are important to the operation of the Tevatron. The first of these modes is $\bar{p}$ production mode. In this mode Booster produces two batches of 83 bunches which are stacked together forming 83 large bunches. These bunches are then accelerated to 120 GeV over the course of two seconds and then transferred to the $\bar{p}$ source where they are used to create anti-protons.

The second relevant mode of the Main Injector is shot setup mode. During shot setup the Main Injector fills the Tevatron with 24 bunches of protons and 24 bunches of anti-protons at an energy of 150 GeV. To fill the Tevatron with protons, Booster produces a batch containing seven proton bunches which are injected into the Main Injector at 8 GeV. These are accelerated to 150 GeV and then "coalesced" (merged) into a single large bunch. This large bunch is then injected into the Tevatron. After 36 bunches of protons have been injected, the main injector loads four bunches of anti-protons and accelerates them to 150 GeV before injecting them into the Tevatron. This process is repeated 9 times resulting in 36 bunches of protons and anti-protons circulating in the Tevatron.

3.1.5 Anti-proton production and storage

Anti-proton production is the limiting factor in achieving high luminosity at the Tevatron. In order to maximize the number of anti-protons delivered to the Tevatron, a network of accelerating and storage rings are employed. First 120 GeV protons are collided into a target where a small number of ~8 GeV anti-protons are produced. These are captured and successively stored in three rings where they are cooled. Finally the anti-protons are transferred to the Main Injector where they are accelerated to 150 GeV for injection into to the Tevatron [13, 9].

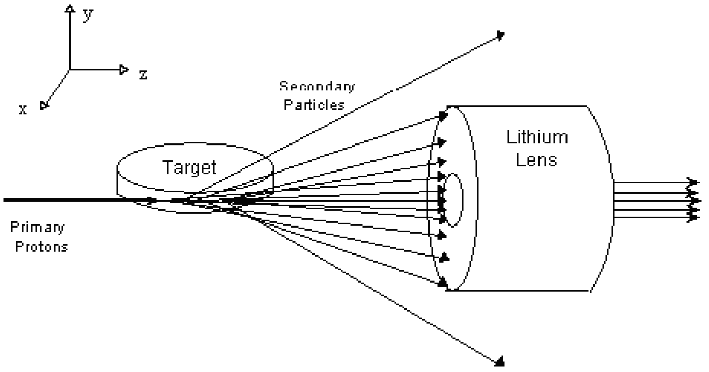


Figure 3.11 Anti-proton target and lithium lens [14]. The lens focuses the anti-protons by pulsing a 650 kA current inside of it.

In order to produce anti-protons, a bunch of 120 GeV protons are steered into a cylinder shaped target made of Inconel (a nickel-Iron alloy) every 2.2 seconds. Each bunch of protons contains ~8×10^{12} protons resulting in ~8×10^{7} anti-protons being produced. The newly produced anti-protons spray out from the target in a cone and are focused by a lithium lens as shown in Figure 3.11. The lens works by pulsing a large current through the cylinder of lithium which produces a magnetic field growing linearly with the radius of the lens. This has the effect of collimating the anti-proton beam. The 15 cm long lens has a 1 cm radius and operates at a maximum current of 650 kA, producing a magnetic field of 1 kT/m [13].

After being produced, the anti-protons are transferred to the debuncher storage ring. The debuncher rotates each bunch in phase space, reducing the momentum spread of the bunch while increasing its longitudinal spread. After this, the RF voltage is reduced over a period of milliseconds, debunching the beam. As a batch of anti-protons are created every 2.4 seconds, there are over two seconds of time remaining before the anti-protons need to be transferred to the accumulator. This time is devoted to reducing the transverse momentum of the beam using stochastic cooling.

Before the next batch of anti-protons can be produced, the current batch is transferred to another storage ring, the accumulator, which shares the same tunnel as the debuncher. To produce an anti-proton beam with the highest density of particles, the beams must be cooled further in the transverse direction. While the debuncher has two seconds to devote to this process, the accumulator has several hours to further cool the beam.

Much like the debuncher, the accumulator uses stochastic cooling. In this process, a set of pickups measure the transverse location of the bunch at one location along the ring. This signal is sent to the other side of the ring where a kicker provides a small adjustment to lower the transverse momentum of the bunch.

Once a significantly large number of anti-protons has been stored in the accumulator, they are transferred to the Recycler storage ring. This is achieved by first rebunching the beam in the accumulator in order to provide several empty spaces between bunches. Once the beam is bunched, a kicker magnet extracts the beam and transfers it into the Recycler using the space between bunches to allow the kicker magnets to rise to full power without losing any anti-protons in the process.

Located in the same tunnel as the main injector, the Recycler is a long term storage ring for anti-protons [15]. The Recycler was originally designed to accept any remaining anti-protons from the Tevatron after a store was over. However, this proved unfeasible and this mode of operation is not used. Instead, the Recycler accepts a batch of anti-protons from the Accumulator once they are sufficiently cooled and continues to cool them further while waiting for the next batch to be added from the Accumulator. Typically the Recycler will be filled with 5-10 batches of anti-protons before delivering them to the Main Injector in preparation for shot setup.

3.1.6 Tevatron

The largest accelerator at Fermilab, the Tevatron, accelerates particles to 980 GeV [16]. The Tevatron is a synchrotron with a radius of 1 km constructed from superconducting magnets. Proton and anti-proton bunches are injected into the same beam pipe and circulated in opposite directions. These bunches are focused and passed through each other at the DØ and CDF detectors.

The Tevatron consists of six 60° arcs (sectors) connected by short straight sections. The DØ and CDF detectors sit in two of these straight sections while others house the RF accelerating cavities, beam dump, and transfer lines to the Main Injector. In each of the six sectors, there are 17 cells of dipole and quadrupole magnets. Each cell consists of a set of magnets in the following configuration: a focusing quadrupole, four dipoles, a defocussing quadrupole, and four more dipoles. Figure 3.12 shows a cross section of a dipole magnet (left) and quadrupole magnet (right). As these magnets are superconducting, they are operated at ~4 K in a bath of liquid helium and provide a magnetic field of 4.2 T.

The ultimate goal of operating high energy accelerators is to produce rare and heavy particles in proton anti-proton collisions. One key performance metric of the accelerator is the instantaneous luminosity which is proportional to the number of proton anti-proton interactions. At the start of a store, the proton and anti-proton bunches are tiny and contain many particles, resulting in an instantaneous luminosity of 300-400 μb^{-1}/s. Over the many hour lifetime of the store, the luminosity decays as the particle bunches become less dense. This occurs as some of the particles are annihilated in collisions, but also due to a slow growth in the size of the bunches. Finally, the store is

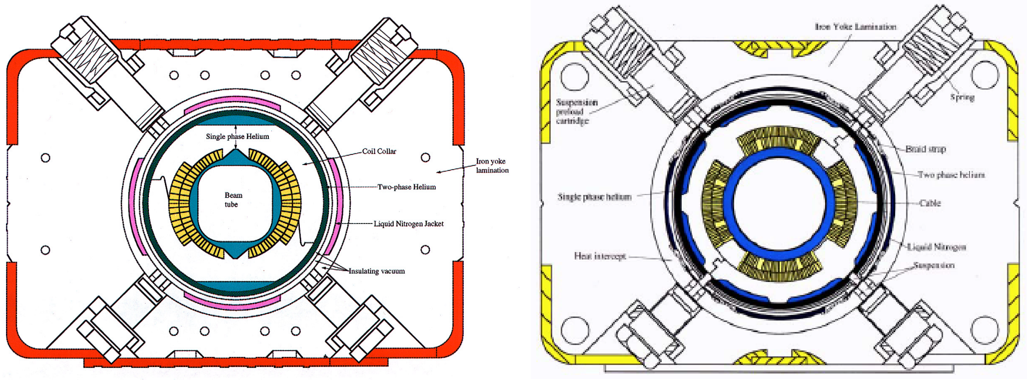


Figure 3.12 Cross section of a dipole magnet (left) and quadrupole magnet (right) [16]

terminated and the Tevatron is refilled with protons and anti-protons.

3.2 DØ detector

The proton and anti-proton beams collide at the D0 and B0 points of the Tevatron ring. Located at these points are the DØ and CDF detectors which attempt to measure the energy, momentum, and identity of particles produced in these collisions. The DØ detector was successfully commissioned in 1992 and upgraded to its current configuration in 2001 [17]. It consists of four major components: the central tracking system, the preshower detector, the calorimeter, and the muon detector. Figure 3.13 shows a schematic of the detector and the relative location of each of these subsystems.

The central tracking and muon detectors are primarily used to record the paths of charged particles and calculate the momentum of the particles while the calorimeter measures the energy of the particles. As different particle types have different responses to each detector, particles can be identified with a high degree of accuracy. Figure 3.14 is a cartoon schematic demonstrating how this process works.

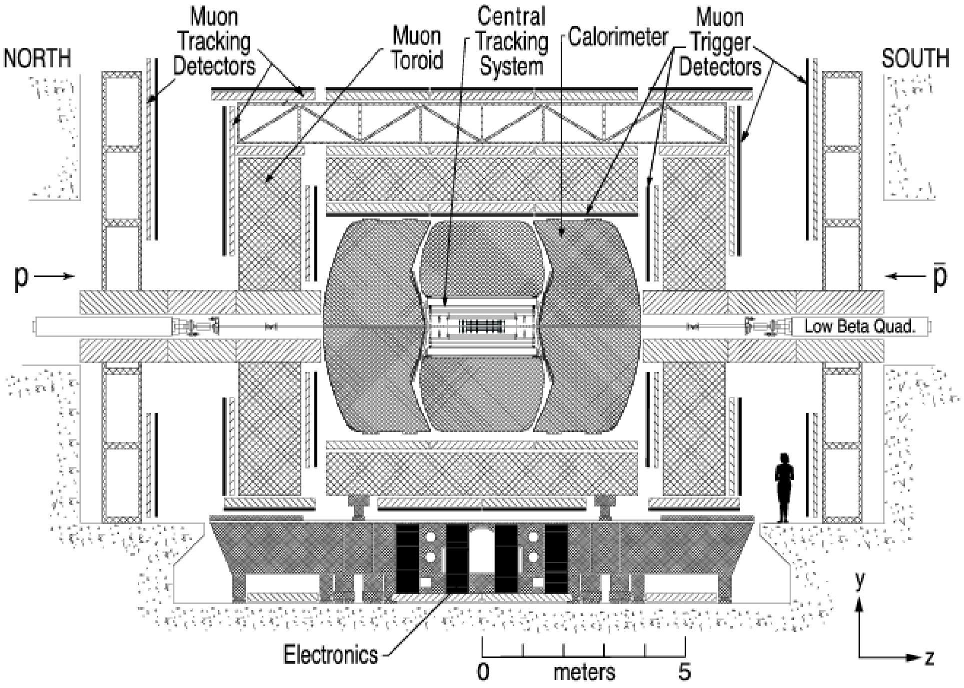


Figure 3.13 Schematic of the DØ Detector [17].

3.2.1 DØ coordinate system

The DØ coordinate system is a right-handed, Cartesian coordinate system with the origin at the interaction point. An alternate coordinate system with the origin at the geometrical center of the detector is also occasionally used, and it will be referred to as the detector coordinate system (e.g. η_{det}). The x-axis points outward from the center of the ring, the y-axis points straight up, and the z-axis points along the beam pipe in the direction of proton travel. As the incoming particles are exclusively traveling along the z-axis, the x-y plane is referred to as the transverse plane. This leads to a natural cylindrical coordinate where $r = \sqrt{x^2 + y^2}$ and ϕ is the azimuthal angle.

The rapidity for a particle is defined as:

$$y = \frac{1}{2} ln \frac{E + P_z}{E - P_z} \tag{3.2.1}$$

This angular definition is useful as Δy is a Lorentz invariant quantity. Assuming the mass of a particle is much smaller than its energy, rapidity can be approximated by

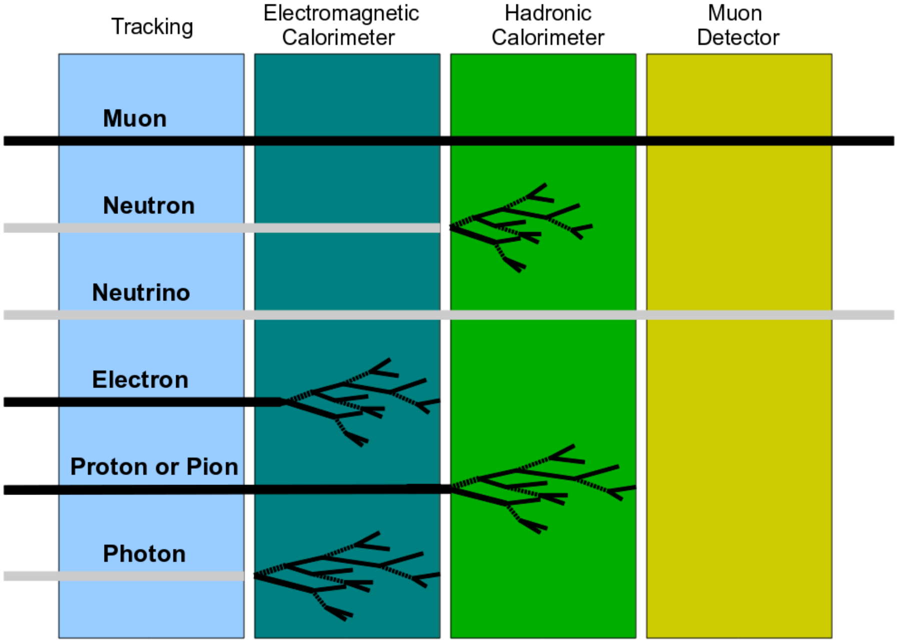


Figure 3.14 Each type of particle interacts differently through the various parts of the detector. Black lines indicate particles leaving energy deposits, grey lines indicate particles passing through without interacting.

pseudo-rapidity, η, which is defined as:

$$\eta = -ln(\tan\frac{\theta}{2}) \tag{3.2.2}$$

which is a purely geometric quantity. A Lorentz invariant measure of the distance between particles, ΔR, is often used in the analysis and is defined as:

$$\Delta R = \sqrt{\eta^2 + \phi^2} \tag{3.2.3}$$

3.2.2 Central tracking

The central tracking at DØ is composed of two sub-detectors: the Central Fiber Tracker (CFT) and the Silicon Micostrip Tracker (SMT). Both of these are located in the center of a 2 T solenoidal magnetic field. Figure 3.15 shows the locations of the tracking components.

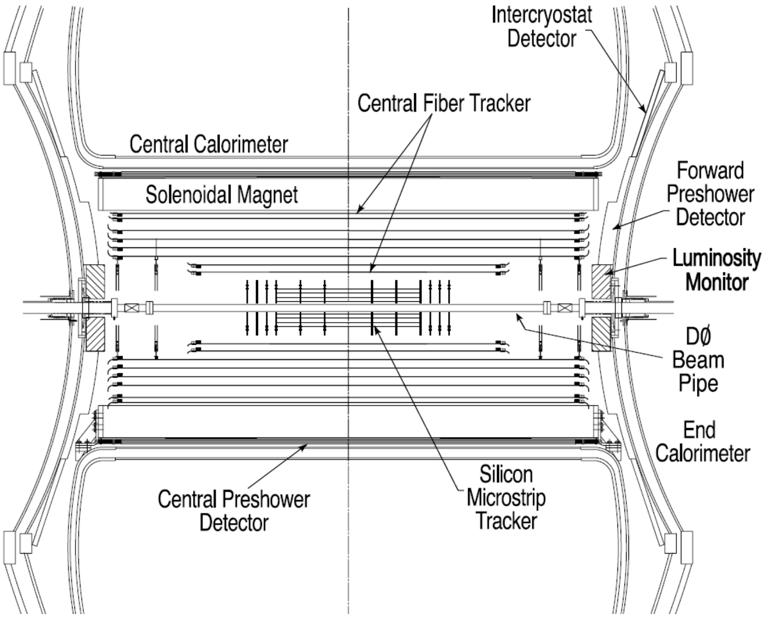


Figure 3.15 Schematic showing the DØ inner tracker in relationship to the solenoidal magnet and calorimeter [17].

The goal of the tracking detectors is to measure the location of charged particles passing through them. These tracks are extrapolated back in order to find the interaction point (primary vertex). Since the tracking detectors are located in a strong magnetic field, the paths of charged particles will be curved. The radius of this curve is used used to measure the transverse momentum of the particle.

3.2.2.1 Silicon microstrip tracker (SMT)

The SMT is constructed from silicon sensors which are arranged into cylindrical barrels around the beam pipe and disks perpendicular to the beam pipe. Figure 3.16 shows the arrangement of barrels and disks with respect to the beam pipe. Twelve smaller disks are inserted into the barrel region while four larger disks are located at higher η. This arrangement was chosen to provide good three dimensional reconstruction of all vertices, despite an interaction region covering 25 cm along the z-axis.

Each silicon sensor is a reversed biased np silicon wafer. When a charged particle

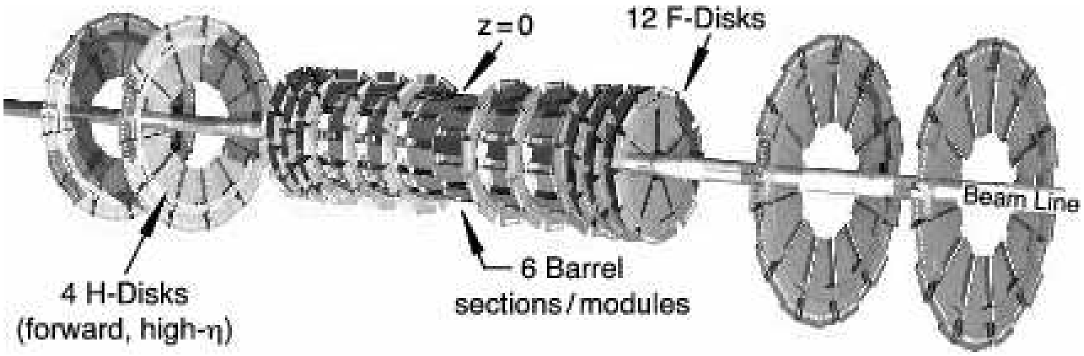


Figure 3.16 The Silicon Microstrip Tracker is constructed out of 6 barrel models interleaved with 12 disks. An additional 4 disks are located further from the interaction point [17].

passes through the silicon, it creates electron-hole pairs. Due to the large electric field in the wafer, these pairs quickly drift to the edge of the wafer where they read out as a current by the electronics.

3.2.2.2 Central fiber tracker (CFT)

The CFT is constructed from eight concentric cylinders of scintillating fibers occupying the region between $r = 20$ cm and $r = 52$ cm. Each cylinder contains two doublet layers, one oriented along the beam pipe and the second $\pm 3^\circ$ from the beam pipe. These stereo fibers allow resolution along the z direction leading to an overall resolution of about 100 μm. The scintillating fibers are connected to clear fiber waveguides that are read out by visible light photon counter cassettes capable of detecting single photons.

3.2.2.3 Solenoidal magnet

Surrounding the central tracker is a 2 T solenoid magnet which has a fairly uniform field over the entire tracking volume and is oriented in the z-direction as shown in Figure 3.17. The main purpose of the magnetic field is to allow the measurement of a particle's transverse momentum by bending their path as they pass through the magnetic field. The magnet is constructed from superconducting Cu:NbTi cables each carrying a current of 4749 A.

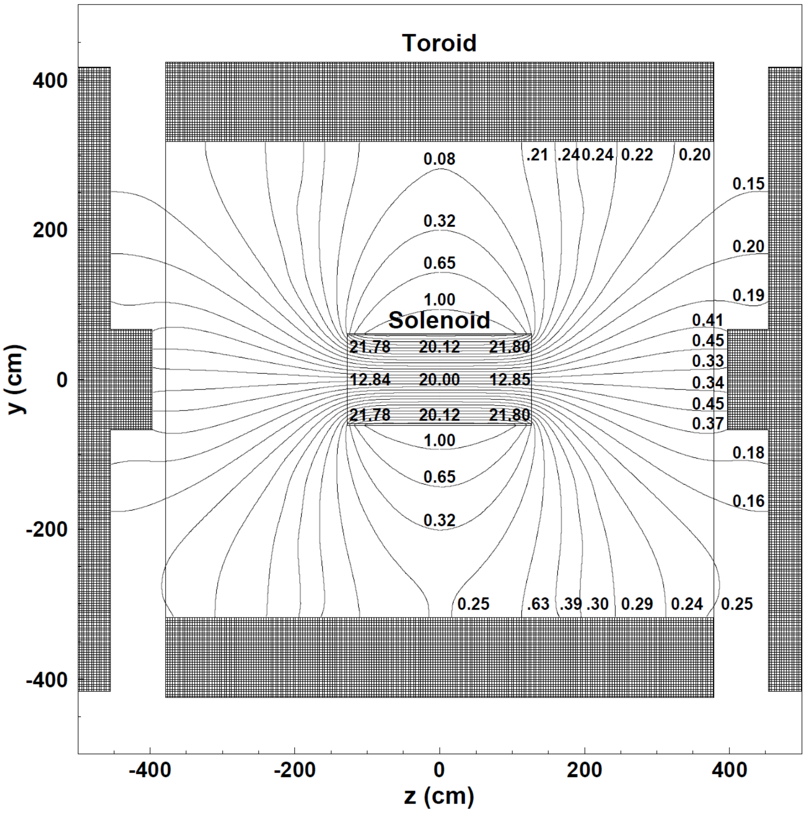


Figure 3.17 The DØ solenoidal magnetic field in kG [17].

3.2.3 Preshower detectors

The preshower detector is located between the solenoid magnet and calorimeter and acts as a tracking detector and calorimeter. This allows better matching between tracks and calorimeter clusters and also aids in electron/photon identification. In addition, the preshower detector is used in the offline reconstruction to correct the energy of calorimeter clusters for any energy losses in the tracking detector.

The preshower detector consists of a central preshower detector (CPS) and a forward preshower detector (FPS) which are visible in figure 3.15. The preshower detectors are constructed from strips of scintillating material connected to fiber optic cable which is read out by visible light photon counter cassettes similarly to the CFT. A one radiation length (X_0) thick lead and steel radiator is located in front of some of the layers of scintillators. Before encountering the radiator, photons leave no tracks in the detector and charged particles leave a single track. After passing through the radiator, electrons and photons start to shower and deposit clusters of energy, while heavier particles continue to leave a single track. The CPS is constructed with a radiator followed by three layers of scintillators while the FPS is constructed with two layers of scintillators followed by the radiator and then two additional layers of scintillators.

3.2.4 Calorimeter

The DØ calorimeter is composed of a central calorimeter covering the region up to $\eta = 1$ and two endcap calorimeters which cover the region up to $\eta = 4$. All three calorimeters perform the task of identifying and measuring the energy of photons, electrons, and jets. As shown in Figure 3.18, each of these calorimeters is further subdivided into an electromagnetic, coarse hadronic, and fine hadronic calorimeter.

All three calorimeters are constructed by alternating active cells with absorber plates. The absorber plates cause the particles passing through to start showering as demonstrated in Figure 3.19. This occurs due the choice of a dense material for the absorber plates, which increase the chance of a particle interacting with the absorber material. The showers of particles leaving the absorber plates then enter the active region where the energy measurement is made. The different subsections of the calorimeters use different absorber materials which assist in particle identification. The electromagnetic calorimeter uses 3-4 mm thick plates of depleted uranium, the fine hadronic sections are made from 6 mm plates of uranium-niobium, and the coarse hadronic region uses plates of 46.5 mm thick copper (central) or stainless steel (endcap).

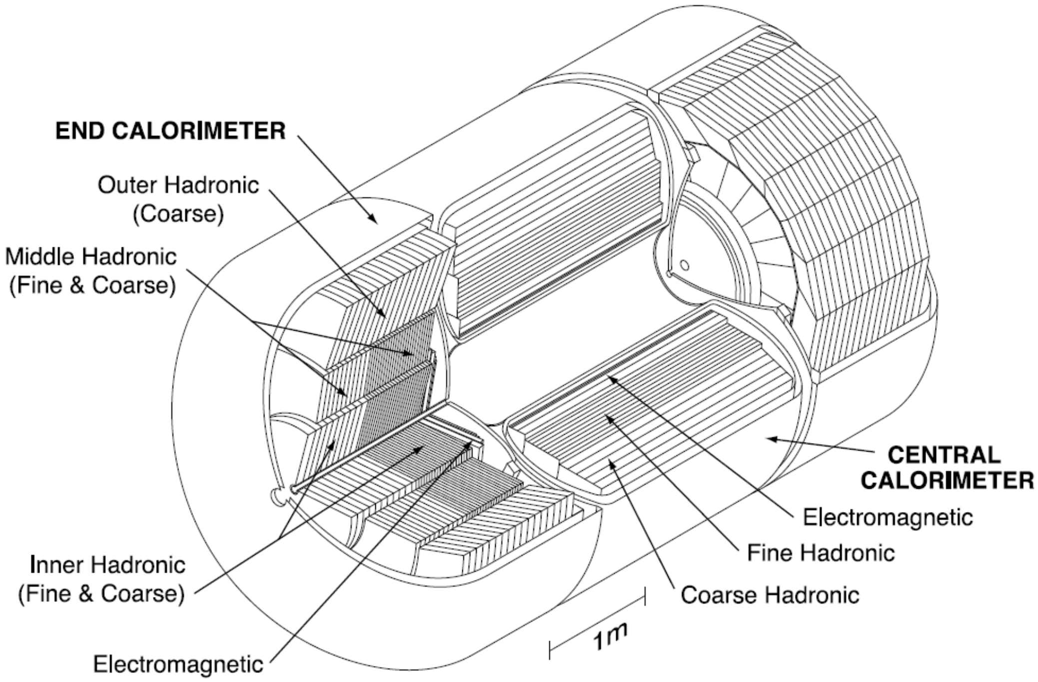
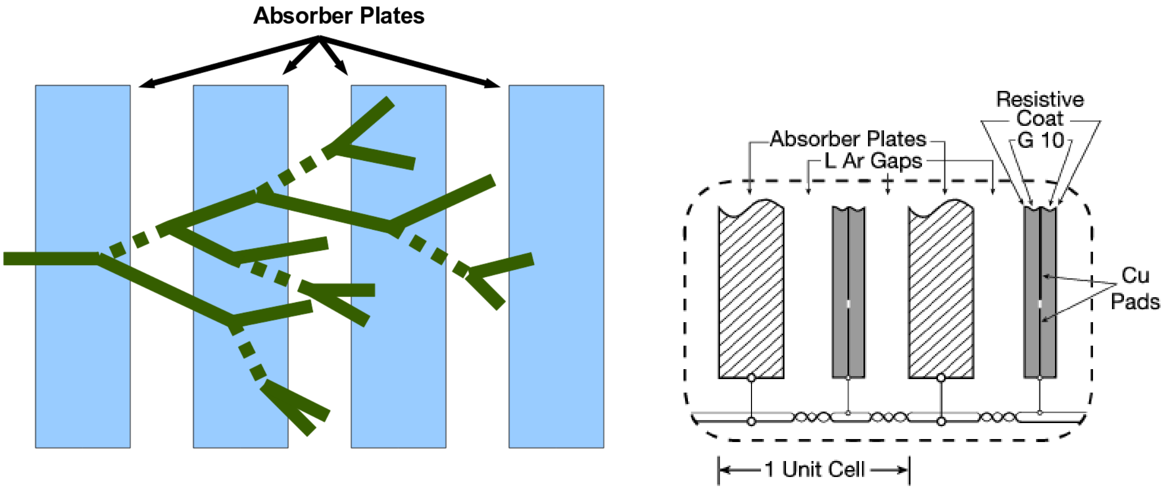


Figure 3.18 The DØ calorimeter is divided into a barrel region and two forward regions. Each of these three regions contains an electromagnetic, fine hadronic, and coarse hadronic calorimeter [17].

The active regions in all subsections of the calorimeter are filled with liquid argon and held at a temperature of approximately 90 K. A 2 kV voltage is applied across the liquid argon between the absorber plates and the pads (Figure 3.19). The showers of particles exiting the absorber plates ionize the liquid argon, and these ions drift towards the pads where they can be read out as a current. The incident particle's energy is proportional to the number of secondary particles produced in the absorber plate, which is in turn proportional to the current measured.

The cells are aligned in "pseudo-projective" readout towers as shown in Figure 3.20. Each tower consists of all cells which lie on a ray projected from the center of the detector. This allows the total energy from a single incident particle to be clustered into a cone (jet) which is treated as a single object in the analysis.



(a) Particle shower formation

(b) Calorimeter readout cell [17]

Figure 3.19 A particle traversing the calorimeter will shower as it passes through the absorber plates (left). The energy from this shower is measured by the readout pads (right).

3.2.5 Muon system

As muons pass through the calorimeter without depositing much energy, an additional layer of detectors is used to identify them and measure their momentum. The muon system consists of a toroidal magnet, three layers of drift tubes, and scintillators covering the region up to $\eta = 2$ [17, 18]. These operate in a similar manner as the inner tracker in that they measure the position of muons as they pass through each layer. The magnetic field bends the muons, allowing for the calculation of each particle's momentum.

The toroidal magnets are located between the first and second layers of detectors and operate at approximately 1.8 T with a current of 1500 A. Figure 3.21 shows the magnetic field in both the endcap and central regions. While the momentum of charged particles, including muons, is primarily measured using the inner tracker, the measurement of a muon's momentum with a standalone detector allows a low p_T cutoff in the Level 1 trigger, reducing the trigger rates. Additionally, the muon detector pro-

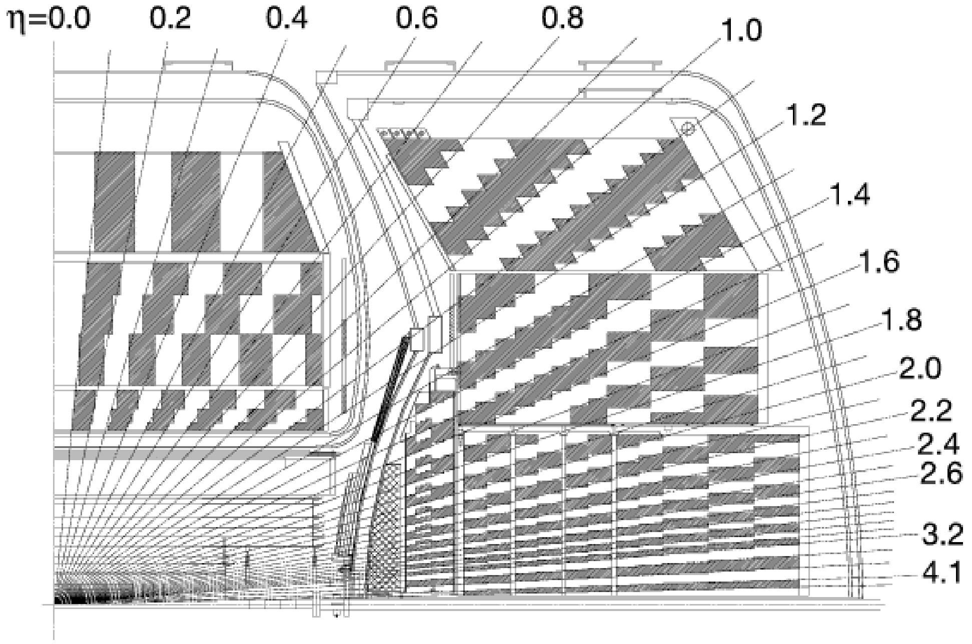


Figure 3.20 Readout cells in the calorimeter are arranged into "towers". These towers are roughly divided by rays originating at the center of the detector [17].

vides cleaner track matching with the central tracker while increasing the momentum resolution for muons.

Each of the three layers in the muon system contain both drift tubes and scintillators. The drift tubes make a measurement of the muon's position and momentum, while the scintillators provide timing information. Figure 3.22 shows the locations of the drift tubes and scintillators. In the central region, large proportional drift tubes (PDTs) are used while in the forward region mini drift tubes (MDTs) are used. Both of these consist of a wire held at a positive voltage between two walls held at a negative voltage, all within a gas mixture. Muons passing through the chamber ionize the gas and the freed electrons drift to the wire where they are read out as a current.

The PDTs, being larger, need to measure the location of the muons along the length of the wire. This is achieved by connecting two wires together at one end and reading them out at the other end. The time difference between the two signals provides

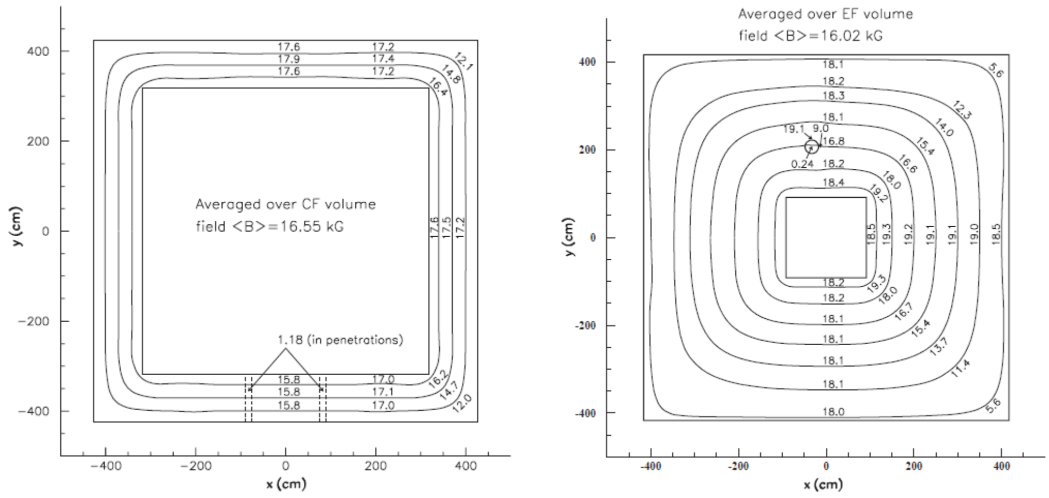


Figure 3.21 DØ toroidal magnetic field in kG [18].

a spatial resolution of between 10 cm and 50 cm depending on the location of the muon. The MDTs are oriented along the ϕ direction and do not make a measurement in this dimension. Instead, end cap scintillators are oriented to measure the location of muons in the ϕ direction. There are three layers of scintillators in the endcaps for this in addition to two layers in the central muon system. Light produced in the scintillators when a muon passes through is amplified by the use of photomultiplier tubes. The time resolution of the scintillators is about 2 ns and is sufficient to associate the hits in the drift tubes to a specific bunch crossing.

3.2.6 Luminosity monitor

In order to make an accurate measurement of the luminosity, $\mathcal{L}$, a standalone detector is used to measure $p\bar{p}$ collision rate at the interaction point. The luminosity is calculated using the formula $\mathcal{L} = \frac{f\bar{N}_{LM}}{\sigma_{LM}}$ where f is the beam crossing frequency, $\bar{N}_{LM}$ is the average number of inelastic collisions per bunch crossing, and σ_{LM} is the effective cross section for inelastic collisions. Since multiple interactions are hard to measure, $\bar{N}_{LM}$ is measured indirectly by counting the number of beam crossings which

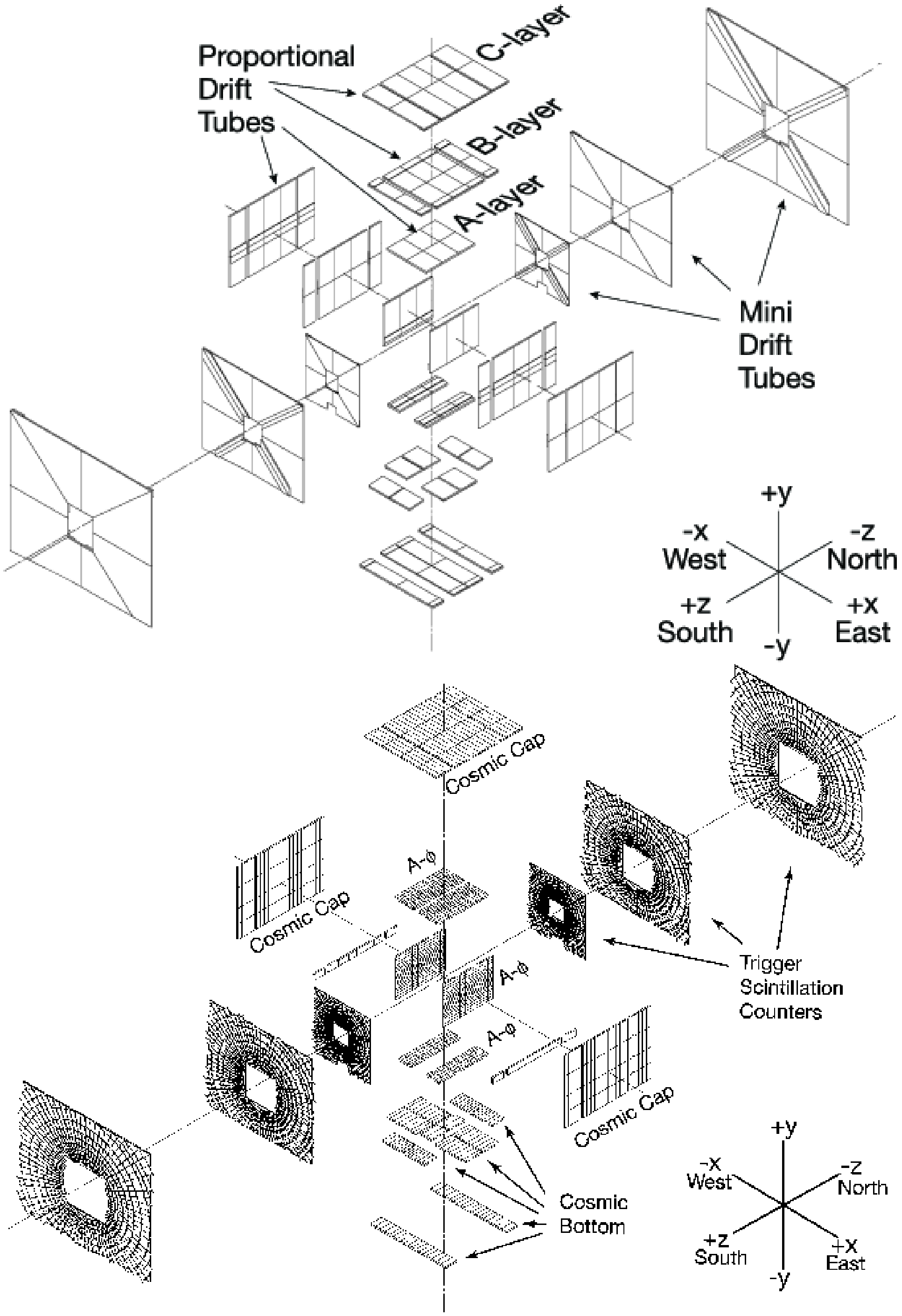


Figure 3.22 Location of muon detector components: drift tubes (top) and scintillators (bottom) [17].

contain no collisions and using Poisson statistics to infer $\bar{N}_{LM}$. The uncertainty on the luminosity is estimated to be 6.1%.

The luminosity monitor is constructed out of two discs located at $z = \pm 140$ cm. Each disk contains twenty-four wedge shaped plastic scintillation counters with a photomultiplier tube attached to each one. These disks cover the range of $\eta = 2.7$ to $\eta = 4.4$. Figure 3.23 shows schematic drawings of the luminosity monitor.

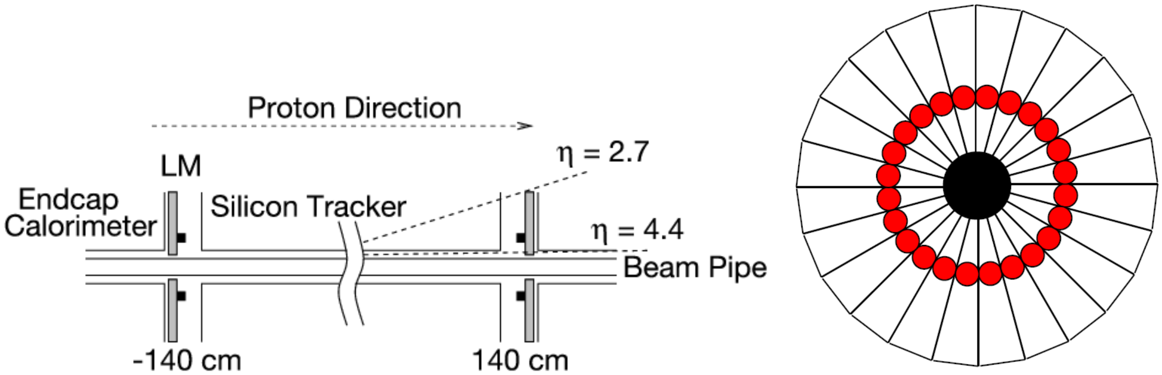


Figure 3.23 Luminosity monitor. Alignment with respect to the beam pipe (left) and schematic showing one disk (right). Wedges are scintilators, red circles are photo-multipliers, and the black circle is the beam pipe. [17].

3.2.7 Triggering

The size of each event at DØ is approximately 0.25 MB and the rate of collisions is 1.7 MHz. If every event were to be saved it would result in over 400 GB of data recorded every second. In order to keep the rate of data written to tape below 100 MB/s most events need to be discarded in real time. As most of the collisions do not produce interesting physics, the goal is to throw these away while keeping all of the events which are interesting. The trigger system achieves this by using three filters run in sequence. Each filter (Trigger Level) looks at a smaller number of events and, therefore, has more time to make a decision about keeping or rejecting an event. Figure 3.24 shows an overview of the data flow through all of the trigger levels.

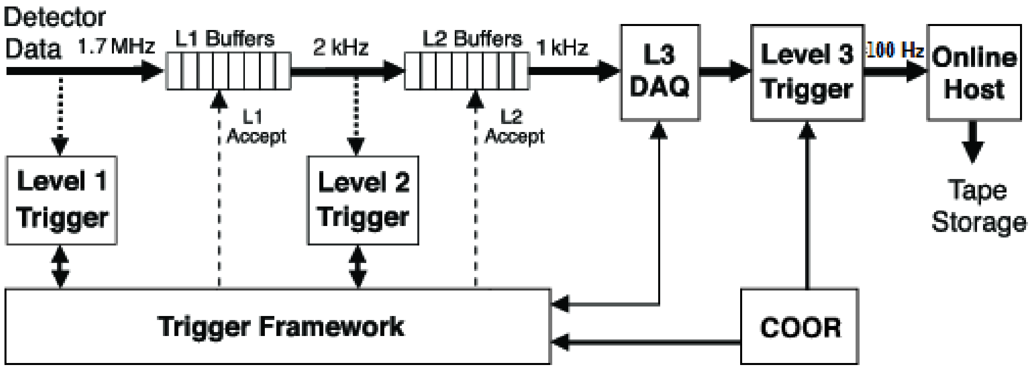


Figure 3.24 Schematic showing data flow and rates through all trigger levels [17].

The Level 1 Trigger reads every event from the detector at a rate of 1.7 MHz and accepts events at a maximum rate of 2 kHz. In order to make decisions on the required time scale of 3.5 μs, only the fastest parts of the detector are used as shown in Figure 3.25.

The Level 2 trigger accepts events at a rate of 1 kHz, and has about 100 μs to make a decision. It includes information from the SMT, better resolution from the muon detectors and calorimeters, and correlations between detectors.

The Level 3 trigger consists of a farm of Linux computers running a fast reconstruction on each event. It can then make decisions based on reconstructed particles such as electrons, tracks, and b-tagging information. Decisions take an average of one second to complete and the accept rate is kept at an average of 100 Hz.

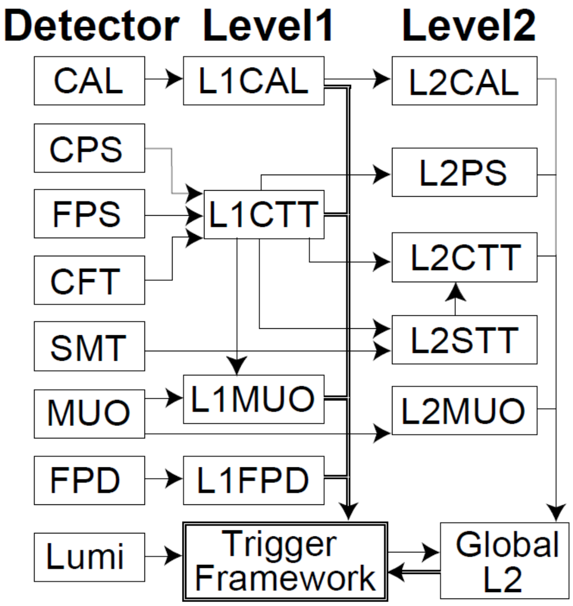


Figure 3.25 The level 1 trigger can only use the fastest detector components. The level 2 trigger is also able to read out some of the slower subsystems [17].

Chapter 4. Reconstruction

The previous chapter provided details on how each section of the DØ detector responds to particles passing through it. This information, however, needs to be translated from clusters of energy and hit locations into candidate objects like electrons, muons, and jets through a process called reconstruction. Tracks must be constructed from hits in the SMT, CFT, and muon detectors. These tracks must then be matched to energy deposits in the calorimeter, and from this the physics objects are identified. In this analysis, electrons, muons, jets, b-jets (jets which contain a b-quark), and missing transverse energy ($\not\!\!E_T$) are considered. Finally, quality cuts are applied to these objects to reduce the misidentification rates. These steps are described in the the sections which follow.

4.1 Tracks

As each charged particle passes though the SMT and CFT detectors it deposits small amounts of energy. This energy is reconstructed into hits in each layer corresponding to the location of the particle. The trajectory of the particle can then be determined by extrapolating a track from these hits. This process is complicated by the strong magnetic field in the tracking volume which curves the tracks, and by the thousands of hits present in a single event.

DØ uses two tracking algorithms to find tracks, and the results from both algorithms are then combined to form a final set of tracks. The first algorithm, histogram track finding (HTF), uses track curvature and the direction of the track at the origin [19]. The tracking also uses the Alternate Algorithm (AA), which extrapolates tracks from the SMT into the CFT [20].

The HTF first takes every possible pair of hits in the tracker, and then calculates the curvature (p) and direction (ϕ) of the track which would pass through each pair of hits and the interaction point. A histogram in (p,ϕ) space is filled from all of the possible combinations. A given track will have a specific direction and curvature, and all pairs of points on the track will share these values. Pairs of points not originating from the same track will fill the histogram uniformly. Tracks are then identified as peaks in the histogram. The details of the algorithm are discussed in [19].

The alternative algorithm (AA) starts by fitting a track to three hits on successive layers in the SMT and requiring the track to pass near the beam location. Additionally, the fit is required to have a $\chi^2 < 16$. This track is then extrapolated to the next layer of the detector and any hits at that location are added to the track, provided that χ^2 remains less than 16. This process is continued throughout the entire tracking volume. The list of tracks is ordered by the number of hits and χ^2. To reduce duplication, tracks are accepted in order as long as they contain few hits already in accepted tracks. Further details of the alternative algorithm can be found in [20]. The tracks from the two methods are then provided to the global track reconstruction. This algorithm refits the tracks and smooths them using the Kalman method to produce the final list of tracks. This algorithm is explained in [21].

4.2 Primary vertices

The primary vertex is the location of the hard collision between the $p\bar{p}$ pair, and is the origin of the particles produced in this collision. At DØ the location of the primary vertex is close to the center of the detector in the (x,y)-plane, but can vary over 1 meter in the z-direction. The main difficulty in finding the location of the primary vertex is to determine which tracks originate from it and which originate from secondary collisions and decays of heavy quarks.

The primary vertex is found by using the adaptive primary vertex algorithm [22]. First, using all of the tracks, the location of the vertex in the (x,y)-plane is determined by a χ^2 fit. Tracks with at least 2 SMT hits and a $p_T > 0.5$ GeV are clustered into regions 2 cm long in the z direction. Each cluster of tracks is fit to a primary vertex, after which those with the worst χ^2 are dropped until the overall fit has a χ^2 per degree of freedom less than 10. A primary vertex is then calculated by giving larger weight to those clusters with the better χ^2 values. The χ^2 fit is recalculated and a new primary vertex is found. This is iterated upon until it becomes stable.

If several primary vertices are found in a single event, the p_T distributions of the tracks are used to choose which will be used as the primary vertex. The vertices originating from a minimum bias scatter will have low p_T tracks. Each track is assigned the probability that it originated from a minimum bias interaction and the overall probability that a vertex came from a minimum bias collision can be calculated. The vertex with the lowest probability of being a minimum bias vertex is considered to be the primary vertex.

4.3 Calorimeter preprocessing

As described in section 3.2.4, the calorimeter is composed of many individual cells. A particle traveling through the calorimeter will deposit its energy into several neighboring cells as it showers. In order to reconstruct the energy of the particle, these cells are grouped into a cluster and their energies added together.

Due to electronic noise and hardware problems, a cell can become "hot" and be read out as having large amounts of energy, even though no particles passed through it. It is important, therefore, to suppress hot cells before forming calorimeter clusters. First, the calorimeter is read out while the accelerator is off in order to obtain a baseline for electronic noise. The energy width of each cell, σ_{cell}, is then measured. Hot cells are usually suppressed during data taking to avoid overloading the trigger. During the reconstruction, two additional noise suppression algorithms are used.

The first of these is the New Anomalous Deposits Algorithm (NADA [23] which takes advantage of the fact that if energy is deposited in one cell, it will also be deposited in some neighboring cells. Therefore, any cell which records an energy of larger than 1 GeV, has the energy of all 26 of its neighboring cells (in a $3 \times 3 \times 3$ box) summed together. If this energy is less than 100 MeV, the central cell is considered to be hot and it's energy is set equal to 1 MeV.

After the NADA algorithm, the T42 algorithm [24] is applied. This algorithm sets to zero the energy of any cell which has measured an energy of less than $2.5\sigma_{\text{cell}}$. Then the algorithm sets to zero any cells with an energy between $2.5\sigma_{\text{cell}}$ and $4.0\sigma_{\text{cell}}$, unless one of it's 26 neighboring cells has an energy above $4.0\sigma_{\text{cell}}$.

These two algorithms have the effect of removing isolated deposits of energy while accepting small deposits of energy near signal regions. The details of the clustering of calorimeter cells into objects is described in detail in sections 4.4 and 4.5. Finally, all remaining cells are combined into towers of cells, with each tower pointing back to the geometrical center of the detector.

4.4 Jets

Quarks and gluons produced from the hard scatter shower into a dense cluster of many particles called a jet. The total momentum of all of the particles in the jet is close to that of the initial quark or gluon. Jet reconstruction is the process of determining which cells in the calorimeter belong to a given jet. It is important when clustering the jet that the kinematic properties of the original quark or gluon match as closely as possible to those of the reconstructed jet. This requires that the jet algorithm must be both infrared safe and collinear safe [25].

Infrared safe jet algorithms are those which are not affected by an additional soft radiation. Figure 4.1 shows a typical example of how an additional soft radiation could cause two separate jets to be merged. When an quark or gluon splits into two pieces, both traveling in roughly the same direction carrying equal energy, this is collinear radiation. To be collinear safe, a jet algorithm should only use objects which are unaffected by collinear radiation. An example (as shown in Figure 4.1) of a non-collinear safe object would be the largest energy deposit in a cell.

The algorithm used in this analysis is the Run II Improved Legacy Cone Algorithm [25, 26]. First, calorimeter towers are combined into pre-clusters using a simple cone algorithm. This works by taking a p_T ordered list of calorimeter towers with $p_T > 500$ MeV and constructing a cone of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.3$ around the first of these. All towers inside this cone with $p_T > 1$ MeV are added to this pre-cluster and removed from the list of towers. Further cones are built around the next towers on the list, in order, until the list is empty.

Next, clusters are built out of the pre-clusters. By using pre-clusters instead of directly using towers, the algorithm has improved collinear safety. The pre-clusters are p_T sorted and a cone of $\Delta R = 0.5$ is constructed around the pre-cluster. Any pre-clusters in the cone are added to the jet. The center of the jet and its cone are re-calculated and the list of pre-clusters in the jet is updated. This process is iterated until stable. All pre-clusters are looped over, and jets are constructed around any which have $\Delta R > 0.25$ from any existing jet. To make the algorithm infrared safe, jets are also seeded using the midpoint between every set of pre-clusters as the starting point.

Finally, overlapping jets are merged or split to form the final list of jets. If two jets share pre-clusters containing more than 50% of one jet's p_T, the lower p_T jet is merged

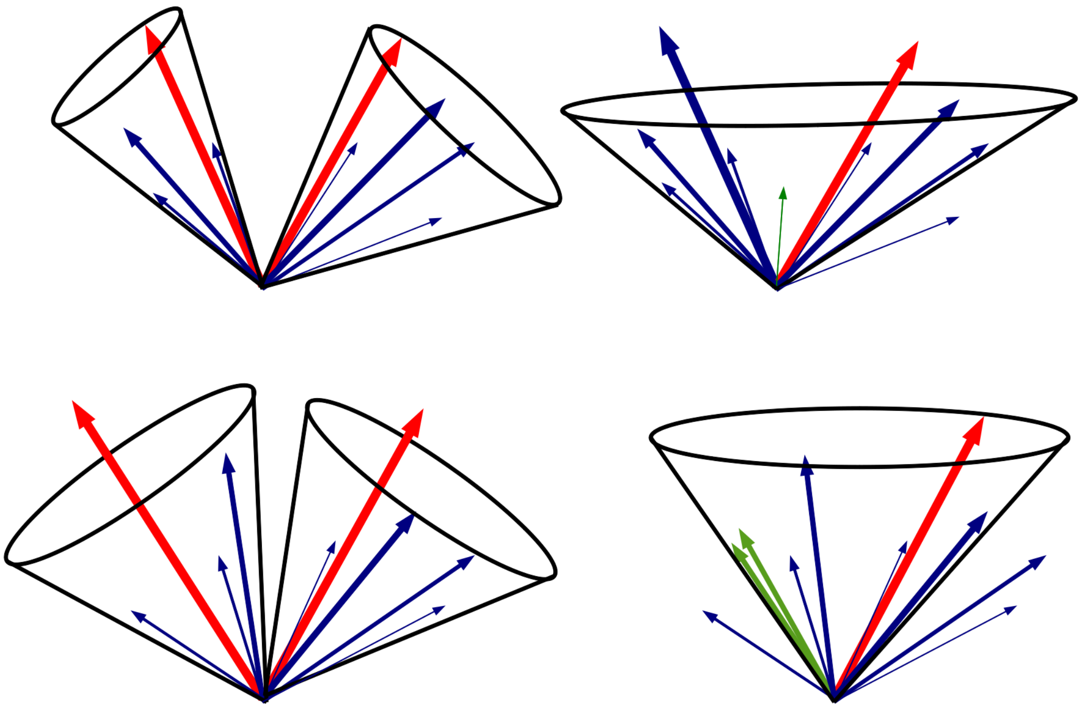


Figure 4.1 Non-IR safe jet reconstruction before (top left) and after (top right) additional infrared radiation shown in green. Non-collinear safe jet reconstruction before (bottom left) and after (bottom right) additional collinear radiation. Arrows show energy deposits. Red shows energy deposits used to seed jet reconstruction. Green shows the change due to IR or collinear radiation.

into the larger. Otherwise, shared towers are each assigned to the jet which is closer in ΔR.

Jets in the Monte Carlo simulation need to be corrected as they have different identification efficiency, an overly optimistic energy resolution, and a slightly different calorimeter response. These are corrected by using the "Jet Shifting, Smearing, and Removal algorithm" (JSSR) as described in [27]. The algorithm is derived using Z events containing exactly one jet and applied to all samples except the top quark pair and single top quark samples. Jets in this analysis are accepted if they pass the Jet-ID selection criteria described in [28], have a p_T > 15 GeV, and have an $|\eta|$ < 3.4. Jets which both have these corrections applied and pass the quality cuts are classified as "Good Jets".

4.5 Electrons

Electrons, being charged particles, leave tracks as they pass though the inner detector and also leave small clusters of energy in the electromagnetic calorimeter. The energy in the EM calorimeter is clustered using a simple cone algorithm similar to that used for the jets, but with a cone size of $R = 0.2$. In order to increase the efficiency of electron ID while keeping the misidentification rates small, several quality metrics are defined for electron candidates [29].

- **Electromagnetic fraction:** The ratio of the energy in the EM calorimeter towers to the energy in the entire calorimeter towers is calculated for towers in the electron cone. For real electrons, this value should be close to 1.
- **Isolation:** A cone with $R = 0.4$ is constructed around the electron's normal cone of radius of 0.2 and the ratio of energy in the small cone to energy in the large cone is calculated. As electrons produce small showers, this value should again be close to 1 for real electrons.
- **H-matrix:** Seven variables describing the shape of the electromagnetic shower are used as inputs into a 7×7 covariant matrix. From this covariant matrix a χ^2 value can be constructed comparing the shape of the electromagnetic shower to that of a typical shower [30].
- **Track match probability:** The χ^2 is calculated for the closest track to the electron cone and this χ^2 is then converted into the probability that the track is associated with the given electromagnetic cluster.

- $\mathcal{L}_{\mathrm{EM}}$: A likelihood is constructed from seven variables giving the probability that an electron object is a real electron [31]. The variables used are: Track match probability, H-matrix, electromagnetic fraction, (cluster E_T)/(track p_T), distance of closest approach of the track to the primary vertex, number of tracks in a cone of size R=0.05 around the track, sum of p_T for all other tracks in a cone of size R=0.4.

These quality variables are used in this analysis to define three categories of electrons with increasingly strict requirements [32]. Additionally, all electrons are required to be in the central calorimeter with $|\eta_{\mathrm{det}}| < 1.1$.

- **Ultra-loose electron:** Requires an EM fraction > 0.9, H-matrix χ^2 < 50, isolation < 0.2, and p_T > 15 GeV. No matching track is required. This electron definition is used for the QCD multijet background in order to increase the acceptance.

- **Loose isolated electron:** In addition to the requirements for an Ultra-loose electron, the isolation criteria is tightened to 0.15, There must be a track matched to the cluster with non-zero probability which has p_T > 5 GeV and Δz (primary vertex,track) < 1 cm. This electron definition is used to veto events with more than one isolated lepton.

- **Tight isolated electron:** In addition to the requirements for loose isolated electrons, tight isolated electrons are required to have $\mathcal{L}_{\mathrm{EM}} > 0.85$. This is the standard electron definition used throughout the analysis unless otherwise specified.

4.6 Muons

Muons leave tracks through both the central tracker and the layers of the muon system. In order to calculate the complete track through the detector these must be matched. First, tracks are constructed for each of the three layers in the muon detector.

The tracks in the outer two layers (BC) are then connected as there is no magnetic field present here. The tracks in the inner muon layer (A) are connected to those in layers B and C through the toroidal magnetic field. Finally, this track is propagated to the central tracker where it is matched to a track. Two quality levels are defined in this analysis for the muon reconstruction [33] [34].

- **Loose isolated muon:** Loose muons are required to have at least two wire hits in layer A, at least two in layer BC, at least one scintillator hit in layer A, and at least one in layer BC. To veto against cosmic ray muons, a timing requirement is added requiring the scintillator hits to be within 10ns of the expected time for muons from the collisions to reach the scintillators.

  Muons are also required to pass *medium* track quality definitions. This requires a track match with the central tracker of χ^2/dof $<$ 4 and distance of closest approach to the primary vertex $<$ 0.2 cm for a SFT track or $<$ 0.02 for a SMT track.

  Loose isolated muons are also required to have a p_T $>$ 4 GeV and must be at least ΔR>0.5 away from a jet. Loose isolated muons are used for vetoing events containing more than one isolated lepton.

- **Tight isolated muon:** In addition to the requirements for loose isolated muons, the p_T requirement is increased to 15 GeV and must pass the "TopScaledLoose" requirement. This means that the sum of the momenta of all other tracks in a cone of R $<$ 0.5 around the muon and the energy deposited in a cone of 0.1 $<$ R $<$ 0.4 around the muon must both be less than 20% of the muon p_T. The tight isolated muon definition is used as the standard muon definition throughout this analysis unless otherwise specified.

4.7 Missing energy ($\not\!\!E_T$)

All of the single top events considered in this analysis have a neutrino in the final state. As neutrinos pass through the detector without interacting, their presence can only be inferred. Prior to the proton anti-proton collision, the momentum in the transverse plane is zero. By conservation of momentum, the vector sum of the transverse momenta of all of the final state particles must also be zero. Therefore, the momentum of the neutrino is equal to the negative of this sum, denoted as $\not\!\!E_T$.

The vector sum of all cells in the electromagnetic and fine hadronic calorimeters is computed along with those cells in the coarse hadronic calorimeter which are contained in a good jet. As muons only deposit small amounts of energy into the calorimeter, the muon momentum (minus the energy loss in the calorimeter) must be added to this in order to calculate the total transverse energy sum. Finally, energy corrections which were applied to reconstructed objects such as jets are also accounted for in this sum [32].

4.8 b-jets

Single top events contain two b-quarks in the final state, both of which hadronize and are reconstructed as jets. Successfully identifying jets which originate from b-quarks significantly helps to reduce the background rate and is known as b-tagging. This process utilizes the relatively long lifetime of b-hadrons which allows them to travel several millimeters before decaying. Tracks from these jets can be identified by the fact that they come from a displaced secondary vertex.

The first requirement in identifying a b-jet is taggability. This is a set of quality cuts that ensure the tracking information is sufficiently well understood to be used for the tagging algorithm. Jets are considered taggable if they contain at least two tracks with $\Delta R < 0.5$ of a calorimeter jet, $\Delta R < 0.5$ between themselves, at least one SMT hit each, and at least one track having $p_T > 1$ GeV.

In order to quantify the likelihood that a jet originated from a b-quark, the Neural Network (NN) b-tagging tool is used [35]. The NN takes seven different input parameters and provides a smooth likelihood function as an output. The variables used, in order of separating power are: Decay length significance of the secondary vertex (SV), weighted combination of the tracks' impact parameter significances, probability that the jet originates from the primary vertex, χ^2/dof of the SV, number of tracks pointing to the SV, invariant mass of the tracks in the SV, and number of SV's in the jet. Information related to soft lepton tagging is specifically not included to the NN. This allows an independent tagging algorithm to be used for validation studies.

The NN output is divided into several working points, two of which are used in this analysis. Events that contain exactly two taggable jets passing the OLDLOOSE working point (NNoutput > 0.5) are considered to be "Two Tag" events. Events which do not have two jets passing the OLDLOOSE working point but have one taggable jet passing the TIGHT working point (NNoutput > 0.775) are considered to be "One Tag" events. Events with zero or more than two tagged jets are not used in this analysis.

Chapter 5. Data and Monte Carlo Samples

The single top cross section, σ_{signal}, is the rate of production of single top quarks at the Tevatron. It can be calculated from the formula:

$$N_{\text{signal}} = \mathcal{L}\sigma_{\text{signal}} \tag{5.0.1}$$

where $\mathcal{L}$ is the measured luminosity and N_{signal} is the total number of signal events. However, since not all single top events end up being recorded in the final sample, there is an additional efficiency, ϵ_{signal}, equal to the number of signal events passing the event selection over the total number of events. Additionally, a mixture of background events contaminate the final sample, and must be accounted for. The final number of measured events is the sum of all signal and background processes giving:

$$N_{\text{measured}} = \mathcal{L}(\epsilon_{\text{signal}}\sigma_{\text{signal}} + \sum_{i}^{N} \epsilon_i \sigma_i). \tag{5.0.2}$$

Where N is the number of background processes, and ϵ_i and σ_i are the efficiencies and cross sections for each background. The goal of this analysis is to measure σ_{signal} by measuring each other term and solving for σ_{signal}. N is measured by counting the number of events in the sample, and $\mathcal{L}$ is measured by the luminosity detector. The efficiencies for events to be included in the final data sample are modeled using Monte Carlo computer simulations. A separate Monte Carlo data set is used for each signal and background process as described in detail below. The cross sections of the background processes are taken from theory or other experiments.

5.1 Data sample

The data sample used in this analysis consists of data collected in two run periods. Run IIa consists of data taken between August 2002 and February 2006(Run numbers 151831-215670). Run IIb data was taken from June 2006 until June 2009 and has increasingly higher instantaneous luminosity which required the detector and trigger to be upgraded to their current forms, as described in the previous chapter. Table 5.1 shows the luminosity, run period, and trigger version used for data in this analysis [32]

Data sample	Trigger Version	Integrated Luminosity (pb^{-1})
Run IIa	v8-v14	1078.81
Run IIb	v15a	535.44
	v15b	688.02
	v15c	397.31
	v16	2661.90
Total		5360.48

Table 5.1 Integrated luminosities for each data set used in this analysis [32]

This analysis uses the DØ "Mega-OR" trigger which accepts events from all reasonable triggers. Combined into the "Mega-OR" trigger are 814 triggers for Run IIa electrons, 413 triggers for Run IIa muons, 647 triggers for Run IIb electrons, and 490 triggers for Run IIb muons. Using a large number of triggers, signal events which may have failed to pass one trigger requirement are usually accepted by a different trigger. This results in trigger efficiencies of 100% after taking into account the systematic uncertainties (for events otherwise passing our event selection). Therefore, no trigger efficiency correction factors are used as shown in appendix 1 of [32].

5.2 MC samples

Monte Carlo simulated events are necessary in order to model the selection efficiencies of each process and to train the Multi Variate Analysis, which will be discussed in chapter 8. For each background and signal, events are generated, decayed, and the detector response to the events is modeled. Finally, the events are reconstructed and the standard selection is applied to them. For each sample, some of the events are produced and reconstructed to simulate Run IIa events and the rest to simulate RunIIb. Run IIa reconstruction is known as p17 while Run IIb is known as p20. The two run periods have different instantaneous luminosity, trigger versions, detector configuration, and detector responses (due to radiation damage).

In this analysis, single top events are generated using CompHEP-SINGLETOP [36]. This event generator produces events with kinematic distributions that match next to leading order (NLO) calculations. For these events the top mass was assumed to be 172.5 GeV, and the CTEQ6M Parton Distribution Function (PDF) was used [37]. PYTHIA [38] is used to add underlying events, initial state radiation, and final state radiation. TAUOLA [39] is used to model the decay of tau leptons, and EVTGEN [40] is used to model the decay of b-hadrons.

The W+jets, Z+jets, and $t\bar{t}$ samples were generated with ALPGEN [41] version 2.11. The $t\bar{t}$ samples also use a top mass of 172.5 GeV. The parameters used in the generation of these events are discussed at [42]. Table 5.2 shows the number of generated Monte Carlo Events of each type.

5.3 QCD multijet backgrounds

QCD multijet events are selected by the analysis when they fake an isolated lepton, a large amount of $\not\!\!E_T$, and the heavy flavor requirement. While the chance of faking all of these is extremely low, the QCD cross section is sufficiently large that this background is still significant. Modeling this background in MC would require a prohibitively large number of events, most of which would be rejected due to the various cuts. Therefore, the QCD background is modeled using data.

Event Type	Cross Section [pb]	Branching Fraction	No. of p17 Events	No. of p20 Events
	The Monte Carlo Event Sets			
Signals				
$tb \to \ell$+jets	$1.04^{+0.04}_{-0.04}$	0.3240 ± 0.0032	0.6M	0.5M
$tqb \to \ell$+jets	$2.26^{+0.12}_{-0.12}$	0.3240 ± 0.0032	0.6M	0.5M
Signal total	$\mathbf{3.30^{+0.16}_{-0.16}}$	$\mathbf{0.3240 \pm 0.0032}$	**1.2M**	**1.0M**
Backgrounds				
$t\bar{t} \to \ell$+jets	$7.46^{+0.48}_{-0.67}$	0.4380 ± 0.0044	1.5M	1.3M
$t\bar{t} \to \ell\ell$	$7.46^{+0.48}_{-0.67}$	0.1050 ± 0.0010	1.5M	1.3M
Top pairs total	$\mathbf{7.46^{+0.48}_{-0.67}}$	$\mathbf{0.5430 \pm 0.0054}$	**3.0M**	**1.6M**
$Wb\bar{b} \to \ell\nu bb$	90.5	0.3240 ± 0.0032	2.7M	3.0M
$Wc\bar{c} \to \ell\nu cc$	260	0.3240 ± 0.0032	2.7M	3.0M
$Wjj \to \ell\nu jj$	$23,831$	0.3240 ± 0.0032	55M	97M
W+jets total	$\mathbf{24,182}$	$\mathbf{0.3240 \pm 0.0032}$	**60.4M**	**103M**
$Zb\bar{b} \to \ell\ell bb$	38.7	0.03366 ± 0.00002	0.7M	0.7M
$Zc\bar{c} \to \ell\ell cc$	106	0.03366 ± 0.00002	0.7M	0.7M
$Zjj \to \ell\ell jj$	$7,032$	0.03366 ± 0.00002	14M	4.0M
Z+jets total	$\mathbf{7,177}$	$\mathbf{0.03366 \pm 0.00002}$	**15.4M**	**5.4M**
$WW \to$ anything	11.6 ± 0.4	1.0 ± 0.0	2.0M	0.7M
$WZ \to$ anything	3.25 ± 0.11	1.0 ± 0.0	1.0M	0.6M
$ZZ \to$ anything	1.33 ± 0.04	1.0 ± 0.0	1.0M	0.5M
Diboson total	$\mathbf{16.2 \pm 0.6}$	$\mathbf{1.0 \pm 0.0}$	**4.0M**	**1.8M**

Table 5.2 The cross sections, branching fractions, and initial numbers of events in the Monte Carlo event samples. The symbol ℓ stands for electron plus muon plus tau decays. A branching fraction of 1.0 implies all known decays processes are included [32].

In order for an event to be accepted as a QCD background event, it must first pass the normal event selection requirements described in chapter 6, with the exception of the lepton requirement. For the electron channel, the reconstructed electron is not required to be matched to a track. The likelihood cut is then reversed, accepting electrons with $\mathcal{L}_{\text{EM}} < 0.85$, which results in an independent sample. For the muon channel, all of the standard lepton cuts are applied except the isolation criteria, which is reversed, accepting only events failing the tight muon isolation.

Events in the QCD sample are extremely likely to contain a fake lepton since all events have a lepton passing the loose isolation cuts but failing the tight isolation cuts. A scale factor is then calculated and applied to these events to correct for the probability that a fake lepton will pass the isolation requirements. This scale factor, as well as others applied to this sample are described in detail in the next chapter.

5.4 Top mass dependent samples

This analysis assumes a top mass of 172.5 GeV in both $t\bar{t}$ and single top samples. Due to the uncertainty of the measurement of the top quark mass, two additional samples are generated with top quark masses of 170.0 and 175.0 GeV. When these samples are merged with the nominal top mass samples, they are referred to as the "top mass merged" samples.

Merging the three samples together has the effect of tripling the yield of the new sample as there are now three times as many events. Each of the three samples is assigned a scale factor such that the yield of the merged sample is the same as the yield of the nominal mass sample. This still leaves two additional degrees of freedom which correspond to the relative amounts of each individual sample in the merged sample. These are chosen such that they minimize the variance of the yield of the merged sample [32].

This merged sample is used when training the Boosted Decision Trees to make them less sensitive to variations in top mass. When measuring the cross section, the unmerged top mass samples are used. Section 9 describes the measurement performed on the nominal mass sample, and Appendix C repeats the measurement on the top mass samples to calculate the final sensitivity to the top quark mass.

Chapter 6. Event Selection

6.1 General strategy

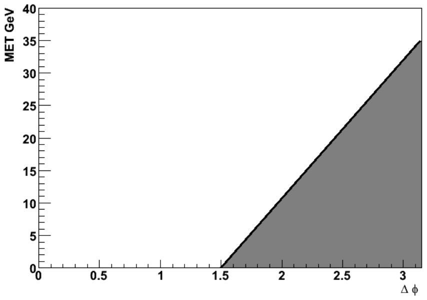
The event selection strategy for this analysis is to apply a set of loose cuts that selects events with a topology corresponding to a W boson decaying into a lepton. This requires events to have an isolated lepton and large $\not{E}_T$. The dataset is then divided into three sub-samples. Each sub-sample will contain events with exactly 2, 3, or 4 high-p_T, good jets. At this point the selected events are dominated by W + jet events with the $t\bar{t}$ contribution increasing in the larger jet multiplicity events. In order to achieve acceptable agreement between the data and the standard model MC events, an additional set of cuts are used which remove low p_T events. Once these cuts have been applied, the sample is known as the "pre-tag selection". Several correction factors are then applied to the Monte Carlo as described in Section 6.4. Finally, b-tagging requirements are added and the scale factors related to tagged jets are applied. The sample is further divided into events containing exactly one tagged jet and events containing exactly two tagged jets.

6.2 Pre-tag selection

Selection Criteria for both electron and muon channels:

- Good data quality (data only)
- Instantaneous luminosity > 0

- Pass Mega-OR trigger requirement
- Good primary vertex: $|z_{\mathrm{PV}}| < 60$ cm and at least 3 tracks pointing to it
- Exactly two, three, or four jets with $p_T > 15$ GeV and $|\eta_{\mathrm{det}}| < 3.4$. In Run IIb, jets must be vertex confirmed. Vertex confirmation requires that if a jet contains tracks, those tracks pointing back to the primary vertex must contain at least 85% of the energy of the jet.
- At least 1 jet must have $p_T > 25$ GeV
- In Run IIb, events with exactly three jets must have at least two jets with $p_T > 20$ GeV
- In Run IIb, events with exactly four jets must have at least two with $p_T > 25$ GeV
- Triangle cut of $|\Delta\phi(\mathrm{leading jet}, \not\!\!E_T)|$ vs $\not\!\!E_T$ with endpoints at $|\Delta\phi|$ from 1.5 to π rad with $\not\!\!E_T = 0$ GeV and from 0 to 35 GeV when $|\Delta\phi| = \pi$ rad



- $\not\!\!E_T > 20$ GeV for events with exactly two jets
- $\not\!\!E_T > 25$ GeV for events with three or four jets
- $\not\!\!E_T < 200$ GeV
- Scalar sum of the transverse energy (H_T) for all jets must be greater than 50 (55) GeV for events with exactly two jets in Run IIa (Run IIb)

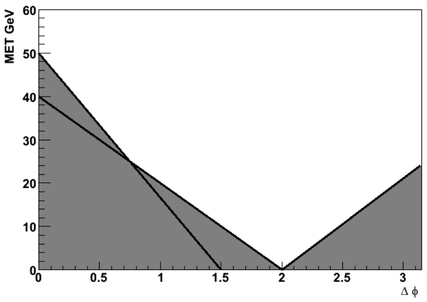
- H_T for all jets must be greater than 75 (80) GeV for events with exactly three jets in Run IIa (Run IIb)
- H_T for all jets must be greater than 100 (110) GeV for events with exactly four jets in Run IIa (Run IIb)
- Triangle cut of "second leading jet p_T" vs "H_T over all jets" with endpoints at second jet p_T from 0 to 27.5 with $H_T = 0$ GeV and H_T from 0 to 165 GeV with the second jet $p_T = 0$ GeV

- H_T over all jets, lepton, and $\not\!\!E_T > 120$ GeV for events with exactly two jets
- H_T over all jets, lepton, and $\not\!\!E_T > 140$ GeV for events with exactly three jets
- H_T over all jets, lepton, and $\not\!\!E_T > 160$ GeV for events with exactly four jets

Additional criteria for electron channel only:

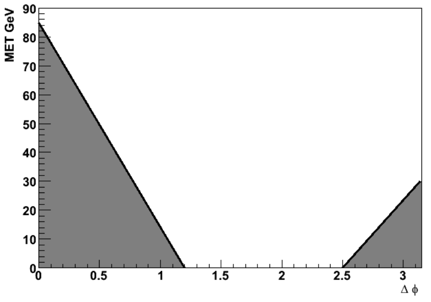
- Exactly one tight electron with $|\eta_{\rm det}|$, 1.1 and $p_T > 15(20)$ GeV for events with two (three or four) jets
- No additional loose electron with $p_T > 15$ GeV
- No tight isolated muon with $p_T > 15$ GeV and $|\eta_{\rm det}| < 2.0$
- Electron originating from the primary vertex within 1 cm in the z direction
- Three triangle cuts with $|\Delta\phi(e, \not\!\!E_T)|$ vs $\not\!\!E_T$

  - $|\Delta\phi|$ from 2 to 0 rad with $\not\!\!E_T$ = 0 GeV and $\not\!\!E_T$ from 0 to 40 GeV with $|\Delta\phi|$ = 0 rad
  - $|\Delta\phi|$ from 1.5 to 0 rad with $\not\!\!E_T$ = 0 GeV and $\not\!\!E_T$ from 0 to 50 GeV with $|\Delta\phi|$ = 0 rad
  - $|\Delta\phi|$ from 2 to π rad with $\not\!\!E_T$ = 0 GeV and $\not\!\!E_T$ from 0 to 24 GeV with $|\Delta\phi|$ = π rad



Additional criteria for muon channel only:

- Exactly one tight muon with p_T > 15 GeV and $|\eta_{\rm det}|$ < 2.0
- No additional loose muons with p_T > 4 GeV
- No loose electron with p_T > 15 GeV and $|\eta_{\rm det}|$ < 2.5
- Muon originating from the primary vertex within 1 cm in the z direction
- Two triangle cuts with $|\Delta\phi(\mu, \not\!\!E_T)|$ vs $\not\!\!E_T$
  - $|\Delta\phi|$ from 1.2 to 0 rad with $\not\!\!E_T$ = 0 GeV and $\not\!\!E_T$ from 0 to 85 GeV with $|\Delta\phi|$ = 0 rad
  - $|\Delta\phi|$ from 2.5 to π rad with $\not\!\!E_T$ = 0 GeV and $\not\!\!E_T$ from 0 to 30 GeV with $|\Delta\phi|$ = π rad



- Triangle cuts on muon track curvature significance vs. $|\Delta\phi(\mu, \not{E}_T)|$. $|\text{TrackCurvSig}| = |\frac{q/P_T}{\sigma(1/P_T)}|$, where q and p_T are the charge and transverse momentum of the charged track associated with the muon. These cuts are designed to reject poorly measured muons.
    - $|\Delta\phi|$ from 0.875π to π rad when $|\text{TrackCurvSig}| = 0$, and $|\text{TrackCurvSig}|$ from 0 to 4(6) when $|\Delta\phi| = \pi$ rad for Run IIa (Run IIb) period
    - $|\Delta\phi|$ from 2 to π rad when $|\text{TrackCurvSig}| = 0$, and $|\text{TrackCurvSig}|$ from 0 to 2(3) when $|\Delta\phi| = \pi$ rad for Run IIa (Run IIb) period
- Leading jet p_T > 30 GeV if this jet is located at 1.0 < $|\eta_{\text{det}}|$ < 1.5 (the region covered by the ICD)
- Second leading jet p_T > 25 GeV if this jet is located at 1.0 < $|\eta_{\text{det}}|$ < 1.5 for events in Run IIb with exactly three jets.

6.3 W + jets and QCD multijets normalization

All background process cross sections except W + jets and QCD multijets are derived from theory calculations [32]. These two cross sections, however, are fit to the data with the equation:

$$N = \lambda_{\text{wjets}} N_{\text{wjets}} + \lambda_{\text{mulitjets}} N_{\text{multijets}} \tag{6.3.1}$$

N, here, is the event yield in data minus the event yield of all non-W and non-multijet process after the pre-tag event selection, λ_{wjets} and $\lambda_{\text{multijets}}$ set the relative amount of W + jets vs. QCD multijets, and N_{wjets} and $N_{\text{multijets}}$ are the yields for the W + jets and QCD multijets samples. In order to calculate the size of λ_{wjets} and $\lambda_{\text{multijets}}$ three distributions which are sensitive to these variables are used: lepton p_T, $\not{E}_T$, and the transverse mass of the reconstructed W: $M_T(W)$. All values of λ_{wjets} between 0 and 4.0 are tried, and the corresponding value of $\lambda_{\text{multijets}}$ is calculated from equation 6.3.1 with the constraint that it can not be negative. The range of 0-4 is chosen as it was found to be sufficiently large to cover the range of values. The shapes of the three variables in the data are compared to the MC in each case and the values of λ_{wjets} and $\lambda_{\text{multijets}}$ are chosen to be the values which give the best overall agreement. The fit is performed for each jet multiplicity, lepton type, and run period. The derived values are shown in table 6.1.

	Normalization Scales							
	λ_{wjets}				$\lambda_{\text{multijets}}$			
	Run IIa		Run IIb		Run IIa		Run IIb	
	e	μ	e	μ	e	μ	e	μ
2 jets	1.071	1.042	0.971	0.954	0.444	0.014	0.388	0.034
3 jets	1.326	1.288	1.082	1.105	0.351	0.025	0.343	0.045
4 jets	1.371	1.553	1.070	1.294	0.309	0.033	0.383	0.024

Table 6.1 W+Jets and multijet scale factors [32]

6.4 Monte Carlo background modeling

In order to correctly model the MC background distributions after event selection, several selection efficiencies need to be corrected. This is achieved by measuring the efficiencies in the data and applying these correction factors to the MC. The following list describes the scale factors used in this analysis.

- **Primary vertex position reweighting:** The MC is reweighted so that the distribution of the primary vertex z location matches data [43].

- **Instantaneous luminosity reweighting:** The MC instantaneous luminosity distributions are generated by using zero-bias event overlays. As zero bias data is collected at different periods from the data, the instantaneous luminosity distribution of the MC is corrected by applying a scale factor to events so that these distributions agree.

- Z p_T **reweighing:** In the Z + jets samples, the ALPGEN Z p_T spectrum does not match the data. A scale factor depending on the true Z p_T and jet multiplicity is used to make the sample agree with Next Leading Order calculations [44].

- **EM ID efficiency correction factor:** the electron identification efficiency is slightly different in data than the MC. A scale factor is derived by looking at data and MC $Z \to ee$ events. This factor is applied to the tight electrons. Ultra-loose electrons have a scale factor consistent with one [32, 45, 46]

- **Muon ID and isolation efficiency correction factor:** The muon identification, track match, and isolation efficiencies are measured in $Z \to \mu\mu$ events in both data and MC. This process is described in detail in reference [47, 48]. The latest muon ID scale factors have been measured with an η dependence which was not included in the Muon ID scale factors used in this analysis. In order to add this information, muon events are reweighted in η_{det} to include the same η dependence in our correction. This is described in appendix 5 of [32].

- **Jet shifting, smearing, and removal (JSSR):** As described in chapter 3, jets in the MC have a higher efficiency and energy resolution than those in the data. In order to correct for this, the standard DØ JSSR algorithm is used to change the energy of the MC jets to match the data [25]. As recommended by the DØ JSSR group, this algorithm is used for the W + jets, Z + jets, and diboson samples while leaving it off for the top pair and single top samples.

- **QCD electron channel correction:** The electron channel QCD multijet sample is derived by inverting the electron isolation requirement. There is a potential kinematic bias in this sample due to the fact the fake electrons in this sample may not behave exactly the same as real electrons in the data[49, 50, 51]. This can by removed by applying a weight of $\frac{\epsilon_{\text{fake}-\text{e}}}{1-\epsilon_{\text{fake}-\text{e}}}$ to each event. $\epsilon_{\text{fake}-\text{e}}$ is defined as the efficiency for a fake electron, which passes the loose isolation requirements, to also pass the tight isolation requirement. This is measured in a selection of *Ultra-loose* electrons orthogonal to our QCD events by requiring $\not\!\!E_T < 10$ GeV. As shown in Figure 6.1, $\epsilon_{\text{fake}-\text{e}}$ has a dependence in lepton p_T. The QCD weight is therefore applied as a function of lepton p_T.

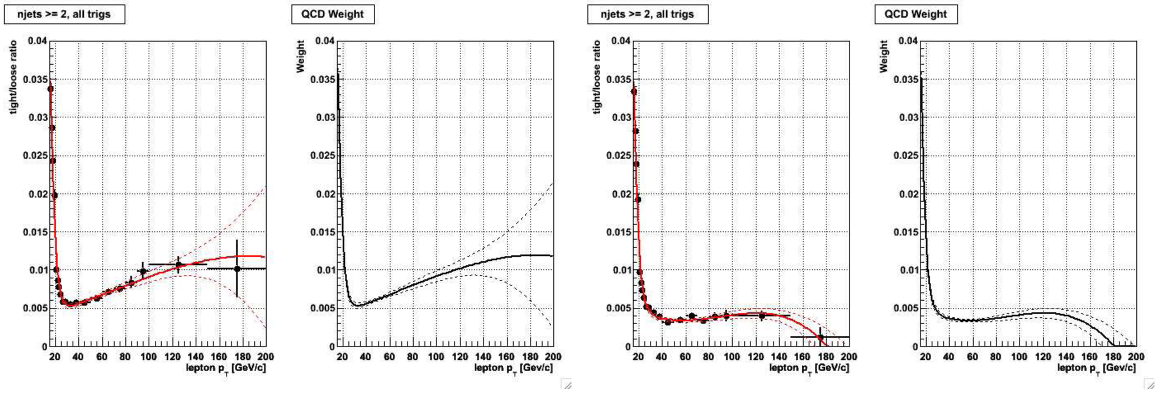


Figure 6.1 QCD correction factors. Far left plot shows the measured ratio of tight to loose leptons ($\epsilon_{\text{fake}-\text{e}}$). Left middle shows the ratio $\frac{\epsilon_{\text{fake}-\text{e}}}{1-\epsilon_{\text{fake}-\text{e}}}$. Right middle and far right show the same plots after subtracting real W and Z boson contamination [32].

The increasing fake rate at high lepton p_T is due to signal contamination from W + jets and Z + jets events which pass the low $\not\!\!E_T$ requirement. To achieve a more accurate measurement of $\epsilon_{\text{fake}-\text{e}}$ in the QCD sample, the signal contamination rate is measured in the MC for events with $\not\!\!E_T < 10$ GeV. This rate is subtracted from the data, resulting in the measurement shown in figure 6.1 on the right.

- W **and** Z **+ jets eta reweighing:** The W + jets and Z + jets Monte Carlo samples are generated using ALPGEN. While good at modeling these processes in general, ALPGEN does produce events with some differences when compared to the data, specifically in the jet distributions at large η. These discrepancies likely originate from the leading log approximation in the generator, biases introduced by the zero bias overlays, and in the difficulty of correctly modeling some detector regions such as the ICR. To correct these distributions, they are measured in the data and in the MC, after applying all event selection criteria except for b-tagging. All MC, except for Z + jets and W + jets, is subtracted from the data sample in order to estimate the W + jets and Z + jets kinematic shape and yield the data. This is compared to the Z + jets plus W + jets samples. The affected distributions are then reweighted with the requirement that the overall normalization of these samples must be held constant. The variables are reweighted sequentially in the following order: $\eta(\mathrm{Jet}_1)$, $\eta(\mathrm{Jet}_2)$, $\Delta R(\mathrm{Jet}_1, \mathrm{Jet}_2)$, $\eta(\mathrm{Jet}_3)$, $\eta(\mathrm{Jet}_4)$, where the jets are ordered in p_T and the three(four) jet reweightings are only applied where there are more than two(three) events. The details of this reweighting are described in Appendix 5 of [32].

- **Taggability and** b**-tagging corrections:** The rates for taggability, and tagging are slightly higher in the MC than the data. These are measured and corrected as described in section 6.5.

- W **+ jets heavy flavor correction:** The ratio of $W + bb$, $W + cc$, and $W + jj$, (where j is a light jet) cross sections have large uncertainties. Scale factors can be measured by comparing the samples with zero, one, and two tags. This measurement is described in detail in Appendix 7 of [32]. The result of that measurement is that the heavy flavor scale factor is consistent with one. Therefore no additional correction is needed.

6.5 b-tagging and tagging scale factors

As described in section 4.8, to detect events with a b-quark in the final state, the DØ b-ID group developed a NN algorithm which identifies jets originating from b-hadrons. Events in this analysis must contain exactly one or two b-tagged jets. Events are considered to have two tagged jets if exactly two jets have a b-tag NNoutput > 0.5. Events which do not pass the two tag criteria are considered to be one tagged events if there is exactly one jet with a b-tag NNoutput > 0.775. Additionally, for a jet to be considered tagged, it must first pass a set of quality cuts known as taggability.

One exception to the normal tagging algorithm is with the W + light jet Monte Carlo sample. Since, by definition, this sample only consists of light quark jets, events are selected by the tagging algorithm only if they fake a b-tag. As this rate is small, the large majority of this sample is discarded, resulting in low statistics. In order to increase the statistics after tagging, the method of random tagging is used in this sample [52]. This algorithm assigns every jet a 50% probability to be tagged and then adjusts the weights of these events to match the expected yields after tagging.

One complication arises with the standard tagging method because data and MC have different selection efficiencies for tagging and taggability. The MC does not have a perfect description of the detector material geometry, and therefore tends to be over-optimistic in tagging efficiencies. Both tagging and taggabilty rates are measured in the data and the MC is weighted to match these rates.

6.5.1 Taggability scale factor

Taggability efficiency is defined as the ratio of the number of jets passing the taggability requirement to the total number of jets. This efficiency is parametrized in jet

p_T, jet η, and the primary vertex z coordinate (PV_z) for the event. Taggability requires jets to have well defined tracks. Therefore the taggability efficiencies are expected to be lower for jets in regions outside of the tracking volume. Events with low p_T, high η, or with a primary vertex far from the origin of the detector will suffer from lower taggability rates.

Taggability is calculated separately for the data and the Monte Carlo. A per-jet scale factor, $T(\eta, p_T, PV_z)$ is defined to be equal to the ratio of taggability in data to that of MC. A scale factor, per event, can by constructed by taking the product of scale factors for taggable and untaggable jets as follows:

$$SF_T(p_T, \eta, PV_z) = \frac{T^{data}(p_T, \eta)}{T^{MC}(p_T, \eta)} \tag{6.5.1}$$

$$SF_{\bar{T}}(p_T, \eta, PV_z) = \frac{1 - T^{data}(p_T, \eta)}{1 - T^{MC}(p_T, \eta)} \tag{6.5.2}$$

$$\prod_{i=1}^{N_T} SF_T(p_{T,i}, \eta_i, PV_{z,i}) \times \prod_{i=1}^{N_{\bar{T}}} SF_{\bar{T}}(p_{T,i}, \eta_i, PV_{z,i}) \tag{6.5.3}$$

Here, N_T is the number of taggable jets and $N_{\bar{T}}$ is the number of untaggable jets per event. The scale factor for untaggable jets is required to preserve the overall normalization of the MC. Equation 6.5.3 assumes there is no correlation between taggability of the jets, however a weak correlation does exist. In order to compensate for some of this correlation, an additional scale factor is calculated as:

$$SF(N_T, N_{\bar{T}}) = \frac{E^{data}(N_T, N_{\bar{T}})}{E^{MC}(N_T, N_{\bar{T}})} \tag{6.5.4}$$

where E^{data} and E^{MC} are the number of events in the data and MC samples after applying equation 6.5.3 as a function of the multiplicity of taggable and untaggable jets. This corrects for the fact than an event with many jets can have a different taggability scale factor than an event with few jets. Combining this scale factor with equation 6.5.3 yields:

$$SF(N_T, N_{\bar{T}}) \times \prod_{i=1}^{N_T} SF_T(p_{T,i}, \eta_i, PV_{z,i}) \times \prod_{i=1}^{N_{\bar{T}}} SF_{\bar{T}}(p_{T,i}, \eta_i, PV_{z,i}) \tag{6.5.5}$$

which is applied to every MC event [32].

6.5.2 Tagging scale factor

The efficiencies for tagging b-jets in MC differs from data in the same way as taggability. In order to correct this, the b-ID group measured the efficiency for a jet to be tagged in both data and MC [53]. This efficiency is parametrized in jet η and jet p_T for each working point of the NN tagger. The scale factor for any jet is then calculated as:

$$SF_{\text{tag}}(p_T, \eta, NN_{\text{out}}) = \frac{\epsilon^{\text{data}}(p_T, \eta, NN_{\text{out}})}{\epsilon^{MC}(p_T, \eta, NN_{\text{out}})} \tag{6.5.6}$$

where ϵ is the tagging efficiency in data or MC. An event scale factor is constructed by taking the product of these scale factors over all taggable jets.

$$\prod_{i}^{N_T} SF_{\text{tag}}(p_{T,i}, \eta_i, NN_{\text{out},i}) \tag{6.5.7}$$

6.6 Event yields

The total number of weighted MC events, or MC yield, exactly matches the number of data events in the pre-tag sample. The distribution of events is shown in table 6.2

After the sample is tagged, it is divided into events containing exactly one and exactly two b-tagged jets. Table 6.3 and 6.4 show the event distributions after performing tagging and applying the tagging related corrections.

Pretagged event yields with channels combined for 5.4 fb^{-1} of data

	Electron+Muon,Run IIa+Run IIb			
	2 jets	3 jets	4 jets	All Channels
Signal				
tb	41.00 ± 0.17	18.00 ± 0.11	5.20 ± 0.06	64.00 ± 0.22
tqb	82.00 ± 0.34	39.00 ± 0.24	13.00 ± 0.14	134.00 ± 0.44
$tb+tqb$	124.00 ± 0.38	57.00 ± 0.26	18 ± 0.16	199.00 ± 0.49
Backgrounds				
$t\bar{t}\to ll$	96.00 ± 0.32	73.00 ± 0.27	24.00 ± 0.14	193.00 ± 0.44
$t\bar{t}\to l$+jets	49.00 ± 0.47	221.00 ± 0.99	316.0 ± 1.2	586.0 ± 1.6
$Wb\bar{b}$	878.0 ± 5.3	313.0 ± 3.2	86.0 ± 1.6	1,278.0 ± 6.4
$Wc\bar{c}$	2,006 ± 11	717.0 ± 6.2	196.0 ± 3.2	2,920 ± 13
Wcj	2,106 ± 13	506.0 ± 6.9	95.0 ± 3.0	2,706 ± 15
Wjj	24,993 ± 46	5,902 ± 23	1,227 ± 11	32,123 ± 53
$Zb\bar{b}$	63.00 ± 0.80	22.00 ± 0.47	5.90 ± 0.24	91.00 ± 0.96
$Zc\bar{c}$	134.0 ± 1.6	46.00 ± 0.93	12.00 ± 0.52	192.0 ± 1.9
Zjj	1,544.0 ± 6.8	350.0 ± 2.9	74. ± 1.4	1,967.0 ± 7.5
Dibosons	679.0 ± 2.0	196.0 ± 1.1	44.00 ± 0.54	919.0 ± 2.4
Multijets	1,025.0 ± 2.9	366.0 ± 2.1	159.0 ± 1.7	1,549.0 ± 3.9
Background Sum	33,572 ± 50	8,711 ± 26	2,240 ± 12	44,523 ± 58
Background + Signal	33,696 ± 50	8,768 ± 26	2,258 ± 12	44,722 ± 58
Data	33,696 ± 184	8,768 ± 94	2,258 ± 48	44,722 ± 211

Table 6.2 Pretagged event yields with statistical uncertainty for each jet multiplicity and for all analysis channels combined [32].

Once tagged event yields with channels combined for 5.4 fb^{-1} of data

	Electron+Muon, Run IIa+Run IIb			
	2 jets	3 jets	4 jets	All Channels
Signal				
tb	66.00 ± 0.46	27.00 ± 0.30	7.70 ± 0.17	100.00 ± 0.57
tqb	133.00 ± 0.78	58.00 ± 0.55	18.00 ± 0.32	209.0 ± 1.0
tb+tqb	198.00 ± 0.91	85.00 ± 0.63	26.00 ± 0.36	309.0 ± 1.2
Backgrounds				
$t\bar{t}\to ll$	168.00 ± 0.84	121.00 ± 0.75	37.00 ± 0.38	326.0 ± 1.2
$t\bar{t}\to l$+jets	144.0 ± 1.6	424.0 ± 2.7	462.0 ± 2.8	1,030.0 ± 4.2
$Wb\bar{b}$	989.0 ± 9.4	315.0 ± 4.6	84.0 ± 2.2	1,388 ± 11
$Wc\bar{c}$	554.0 ± 9.5	208.0 ± 4.9	56. ± 2.3	818 ± 11
Wcj	612 ± 15	137.0 ± 6.9	24.0 ± 2.2	774 ± 16
Wjj	1,066.0 ± 4.1	301. ± 2.2	74.0 ±1.1	1,441.0 ± 4.8
$Zb\bar{b}$	78.0 ± 1.6	28.0 ± 1.2	7.30 ± 0.53	113.0 ± 2.1
$Zc\bar{c}$	39.0 ± 1.4	17.0 ± 1.2	4.50 ± 0.51	60.0 ± 1.9
Zjj	82.0 ± 5.7	31.0 ± 3.6	5.3 ± 1.0	118.0 ± 6.8
Dibosons	158.0 ± 3.1	48.0 ± 1.7	12.0 ± 1.1	218.0 ± 3.7
Multijets	258.0 ± 2.2	117.0 ± 1.9	37.00 ± 0.83	412.0 ± 3.1
Background Sum	4,150 ± 22	1,745 ± 11	804.0 ± 5.2	6,699 ± 25
Background + Signal	4,348 ± 22	1,830 ± 11	830.0 ± 5.3	7,008 ± 25
Data	4,284 ± 65	1,772 ± 42	851 ± 29	6,907 ± 83

Table 6.3 Yields for one tag events with statistical uncertainty for each jet multiplicity and for all analysis channels combined [32].

Twice tagged event yields with channels combined for 5.4 fb^{-1} of data

	Electron+Muon, Run IIa+Run IIb			
	2 jets	3 jets	4 jets	All Channels
Signal				
tb	38.00 ± 0.31	17.00 ± 0.21	5.20 ± 0.12	60.00 ± 0.39
tqb	7.80 ± 0.21	14.00 ± 0.27	7.30 ± 0.19	29.00 ± 0.39
tb+tqb	46.00 ± 0.37	31.00 ± 0.34	13.00 ± 0.22	90.00 ± 0.55
Backgrounds				
$t\bar{t}\to ll$	90.00 ± 0.56	83.0 ± 0.52	27.00 ± 0.28	199.00 ± 0.82
$t\bar{t}\to l$+jets	31.00 ± 0.80	203.0 ± 1.7	334.0 ± 2.1	568.0 ± 2.8
$Wb\bar{b}$	252.0 ± 4.5	90.0 ± 2.6	26.0 ± 1.2	367.0 ± 5.3
$Wc\bar{c}$	41.0 ± 2.3	25.0 ± 1.8	11.0 ± 1.2	77.0 ± 3.2
Wcj	16.00 ± 0.65	7.30 ± 0.55	1.60 ± 0.16	24.00 ± 0.87
Wjj	30.00 ± 0.35	17.00 ± 0.28	6.00 ± 0.15	53.00 ± 0.47
$Zb\bar{b}$	16.00 ± 0.64	8.50 ± 0.49	2.60 ± 0.33	27.00 ± 0.87
$Zc\bar{c}$	3.20 ± 0.39	1.40 ± 0.27	0.56 ± 0.15	5.10 ± 0.50
Zjj	2.9 ± 1.3	1.80 ± 0.81	0.64 ± 0.28	5.4 ± 1.6
Dibosons	20.00 ± 0.81	6.60 ± 0.49	2.00 ± 0.27	29.00 ± 0.98
Multijets	19.00 ± 0.62	13.00 ± 0.64	6.30 ± 0.33	39.00 ± 0.95
Background Sum	520.0 ± 5.5	456.0 ± 3.9	418.0 ± 2.8	1,394.0 ± 7.3
Background + Signal	566.0 ± 5.5	487.0 ± 3.9	431.0 ± 2.8	1,484.0 ± 7.3
Data	597 ± 24	535 ± 23	432 ± 21	1,564 ± 40

Table 6.4 Yields for two tag events with statistical uncertainty for each jet multiplicity and for all analysis channels combined [32].

6.7 Data vs background model comparison

After the event selection, correction factors, and b-tagging, the data and MC background model agree quite well. Figure 6.3 shows a selection of plots comparing these. A legend showing the color scheme used throughout this paper is given in figure 6.2.

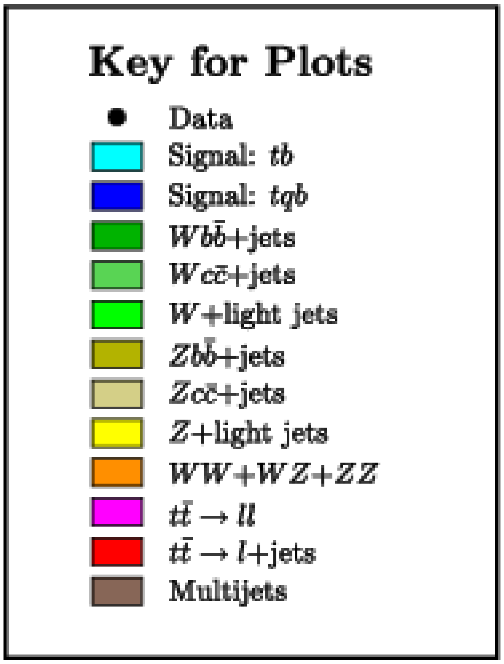


Figure 6.2 Legend for all data mc comparison plots used in this analysis.

6.8 Sample splitting

Once all of the event selection criteria and all correction factors have been applied, the entire data and MC sample is divided into six subsamples. This division is performed by splitting the sample by the number of jets and tagged jets. All events contain either exactly one or two tagged jets and either two, three, or four jets, creating six samples. Figure 6.4 shows that the distribution of MC signal and background differs significantly over these channels. As multivariate techniques will be employed

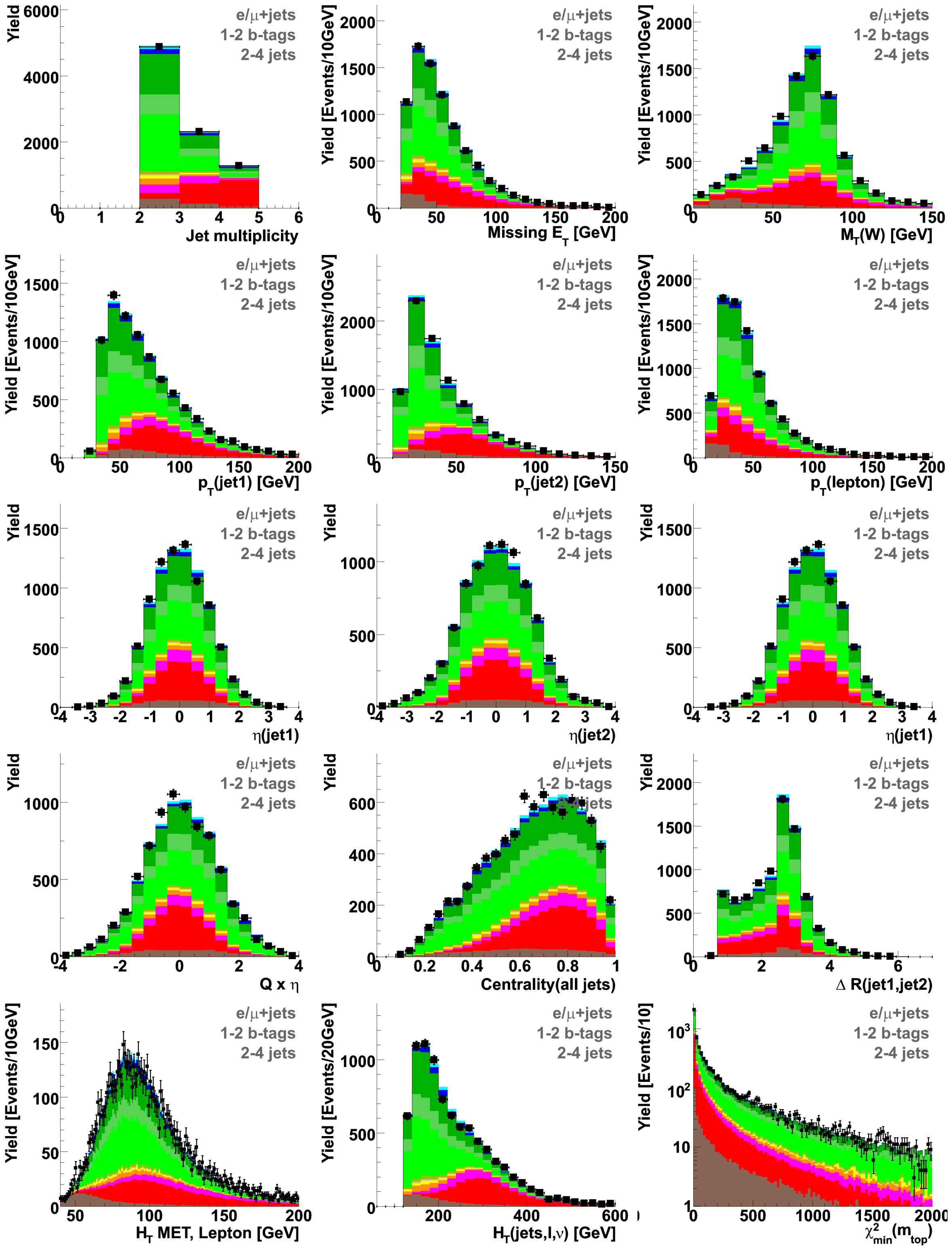


Figure 6.3 Data vs background model comparison across several important variables

to discriminate between signal and background events, having the sample split will allow better performance of the multivariate methods.

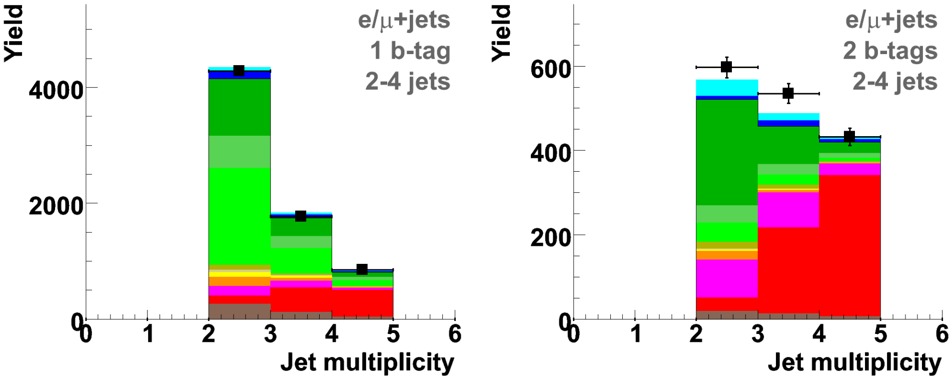


Figure 6.4 Jet multiplicity for one tagged events (left) and two tagged events (right).

In previous analysis at DØ, the samples were also split by lepton type and run period (Run IIa vs Run IIb). In this analysis it was found that the samples split in this way were sufficiently similar. Due to this, the increase in statistics from not splitting over run period and lepton type was more beneficial than the gain due to any differences in these samples.

Chapter 7. Systematic Uncertainties

7.1 Overview

In this analysis, two classes of systematic uncertainties are considered, those only affecting the normalization of various samples, and those which change the shapes of some kinematic distributions. For systematics which do not produce a change in shape, a one sigma fluctuation is recorded for each systematic variation, sample type, and analysis channel. These tables can be found in Appendix B and are used to construct the multi-variate discriminants with the correct ratios of each background process for the systematics. For each systematic affecting the shape of variables used by the multi-variate methods, two parallel samples are generated representing a one sigma fluctuation up and down in that systematic uncertainty. These are used to reproduce multi-variate discriminant outputs, thereby transforming the change of shape from the input parameters into the final discriminant. The systematic versions of the discriminant are used to calculate the total systematic uncertainty.

Table 7.1 shows the relative systematic uncertainty for each of the systematics considered in this analysis.

7.2 Normalization systematics

- **Integrated luminosity:** A 6.1% uncertainty on the luminosity is applied to the signal, $t\bar{t}$, Z + jets, and diboson samples [32].

Relative Systematic Uncertainties	
Normalization Only	
Integrated luminosity	6.1%
$t\bar{t}$ cross section	9.0%
Z + jets cross section	3.3%
Diboson cross sections	7.0%
Branching fractions	1.5%
Parton distribution functions (signal acceptances only)	2.0%
Triggers	5.0%
Instantaneous luminosity reweighting	1.0%
Primary vertex selection	1.4%
Color reconnection	1.0%
b/light jet response	(0.3-1.0)%
Electron identification	(2.8-3.8)%
Muon identification	2.1%
Jet fragmentation and higher order effects	(0.7-7.0)%
Initial-and final-state radiation	(0.8-10.9)%
b-jet fragmentation	2.0%
Taggability	(3.1-21.5)%
W + jets heavy-flavor correction	12.0%
Z + jets heavy-flavor correction	12.0%
W + jets normalization to data	1.8%
Multijets normalization to data	(30-40)%
MC and multijets statistics	(0.2-16)%
Normalization and Shape	
Vertex confirmation	(0.1-9.6)%
Jet energy scale	(0.3-14.6)%
Jet energy resolution	(0.2-11.6)%
Jet reconstruction and identification	(0.04-3.7)%
Angular correction	0.3%
b-tagging, single-tagged	(4.3-14.0)%
b-tagging, double-tagged	(5.8-11.2)%

Table 7.1 An overview of the relative systematic uncertainties used in this analysis. For systematics with both a normalization and shape component, only the normalization uncertainty is quoted in this table. The uncertainties are calculated in each of the six sub-samples, and the ranges show the variation of a given systematic uncertainty across these samples[32].

- **Theory cross sections:** Uncertainty on the single top and $t\bar{t}$ samples are calculated in [54] [55]. The uncertainty on s-channel is $\pm 5.3\%$, $t\bar{t}$ is $+6.4\% - 9.0\%$. Diboson cross section uncertainties are presented in [56], and the uncertainty on the total diboson sample is 3.3%. 3.3% is also used as the uncertainty on the Z + jets cross section as there no direct measurement of this background and its contribution to the sample is small.

- **Branching fractions:** The uncertainty on the branching fractions for W boson decay is 1.5% [1].

- **Parton distribution functions (PDF):** The effect on the signal acceptance by a change of the parton distribution functions is assigned to be 2%. This is estimated by considering 40 different CTEQ PDF sets and measuring the change in signal acceptance. The uncertainty is calculated by taking

  $$\sigma_{PDF} = \sqrt{\sum_{i=1}^{40} (A_{\text{nominal}} - A_i)^2}$$

  where A_i is the acceptance from one PDF set and A_{nominal} is the acceptance from the PDF set used throughout the analysis. The PDF uncertainty for the $t\bar{t}$, Z + jets, and diboson backgrounds are included in the uncertainty of the cross section [32].

- **Trigger efficiency:** The uncertainty on the trigger efficiency is taken to be 5% in all channels. This measured efficiency is always equal to or smaller than this value in each channel [32].

- **Instantaneous luminosity reweighting:** The uncertainty on the reweighting of the instantaneous luminosity distributions is 1.0%.

- **Primary vertex modeling and selection:** The uncertainty on the reweighting of the primary vertex z distribution is negligible The difference in primary vertex selection efficiency between data and MC is 1.4% [43, 57].

- **Electron reconstruction and identification efficiency:** The uncertainty on the electron scale factor is measured to be 4.8% [58].

- **Muon reconstruction and identification efficiency:** The uncertainty on the muon scale factor is taken from [48]. An additional 1% uncertainty is added in quadrature to account for the η reweighting described in Section 6.4. The resulting uncertainty is 2.1%

- **Jet fragmentation and higher-order effects:** The acceptance of $t\bar{t}$ events generated with ALPGEN+HERWIG and MC@NLO+HERWIG was compared with what was used in this analysis (ALPGEN+PYTHIA). The difference in acceptance is used to estimate the systematic uncertainty from both jet fragmentation and higher order effects. The uncertainty is dependent on jet multiplicity and ranges from 0.7% to 7.8%. Only the jet fragmentation systematic effect is applied to the Z + jets, single top, and diboson samples [32]. The uncertainty on higher order effects added is applied to the $t\bar{t}$ sample.

- **Initial-state and final-state radiation (ISR/FSR):** The uncertainty for ISR/FSR is estimated by taking $t\bar{t}$ samples generated with different ISR/FSR settings and comparing the acceptance to the nominal settings. The overall measured uncertainty varies between 0.8% and 10.9% depending on the sample and jet multiplicity [32].

- **Color reconnection:** The uncertainty due to color reconnection is measured to be 1% and is found by taking the difference between PYTHIA TuneACPro and TuneAPro [59].

- **b-jet fragmentation:** The uncertainty on the b-jet fragmentation was measured in [60] and was found to be 2%

- **W + jets and QCD multijets normalization:** The uncertainty from the W + jets and QCD multijets normalization, as described in section 6.3, is estimated at 30-40% for the QCD multijets sample and 1.8% for the W+jets sample [32].

- **W + jets heavy-flavor scale factor correction:** The uncertainty on the λ_{HF} scale factor is 12% and the uncertainty on the Wbb vs Wcc scale factor is 5%. The details of the calculation are shown in appendix 7 of [32].

- **Z + jets heavy flavor scale factor correction:** The uncertainty on the heavy-flavor scale factor for Zbb and Zcc is taken from NLO calculations. The uncertainty on λ_{HF} for Z + jets is considered to be the same as measured in the W + jets sample (12%).

- **Relative b/light jet response:** The uncertainty on the difference in inclusive jet response for b-hadrons and normal jets is calculated in [61] and a difference of 1.8% is measured.

- **Taggability efficiency correction factor:** Taggability efficiencies are calculated with one sigma up and down variations. Contributions come from the statistical uncertainty on the measurement and on the η, p_T, PV_z parametrization. These are then used to measure the uncertainty on the selection efficiency of each sample ranging from 3.5-17%.

- **Sample statistics:** The finite size of the data and the MC samples are taken into account in each bin of the final multi-variate discriminant output. Uncertainties quoted as "Statistical Only" consider the statistical uncertainty on the data only. Quoted uncertainties that include systematic effects, also include the statistical uncertainty of the MC samples.

7.3 Shape changing systematics

- **Vertex confirmation:** The systematic effect for the efficiency correction for MC vertex confirmation is estimated by creating two samples, one with a one sigma higher correction, and one with a one sigma lower correction. In addition to the shape changing effects, the normalization changes between 0.1% and 9.6%.

- **Jet energy scale (JES):** Two additional samples are created with the JES correction raised and lowered by one sigma in order to estimate the shape change uncertainty. The normalization uncertainty is between 0.3% and 14.6% depending on the sample.

- **Jet energy resolution:** The JER correction is fluctuated up and down by one sigma to create a pair of JER samples which are used to estimate the shape changing uncertainty. The normalization of the samples changes from between 0.2% and 11.6% for the various samples.

- **Jet reconstruction and identification:** The correction factor to the jet ID is turned off to create the JetID systematic sample. This sample is then symmeterized by taking half the difference, in the MVA output, for the nominal and the systematic. This is then added or subtracted it to the nominal in order to create symmetric JetID_{up} and $\text{JetID}_{\text{down}}$ samples. The normalization uncertainty of this systematic ranges from 0.04% to 3.7%.

- V **+ jets angular corrections:** The angular corrections are defined such that they preserve normalization in the pretag sample. After tagging, the uncertainty on the normalization is 0.3%. Two additional samples are also considered, with the angular corrections fluctuated up and down by one sigma. These samples have different shapes in the angular variables and are therefore used to determine the associated shape changing systematics.

- **b-tagging efficiency correction factor:** The b-tagging correction factor for MC is fluctuated up and down by one sigma to create the b-tagging shape changing systematic samples. A normalization difference of 4.3-15% is measured in the single tag sample and from 5.8-11.2% in the two tag sample.

Chapter 8. Boosted Decision Tree Analysis

8.1 Multivariate analysis techniques

After the event selection has been performed, the signal to background ratio is approximately 1:20. With this signal to background ratio, the uncertainty on the background prediction is much larger than the signal excess; therefore, a standard cross section measurement is not feasible.

The standard approach to this problem is to add additional cuts into the event selection in order to increase the signal to background ratio until the uncertainty on the background prediction is sufficiently small. The drawback of this technique is that for each additional cut that is applied, the amount of signal decreases as well as the amount of background. In the case of the single top analysis, no set of cuts was found that provided a signal to background ratio sufficient to make a good cross section measurement.

In order to overcome this problem, a multivariate analysis (MVA) technique is used. While there are many different implementations of MVAs, all of them attempt to calculate the probability that an unknown event with properties $\vec{x}$ is a signal event, $Pr(S|\vec{x})$. It does this by using the probability distribution functions for signal and background events, $f(\vec{x}|S)$ and $f(\vec{x}|b)$ [62]. The probability distribution functions are determined from the Monte Carlo distributions, and each MVA uses a different technique to estimate $Pr(S|\vec{x})$ from these distributions.

The process of constructing the function $Pr(S|\vec{x})$ is known as training. In this analysis, one quarter of the total MC is used to train the MVA. A separate MVA is trained for jet-tag multiplicity resulting in six MVAs being trained: one-tag two-jet, one-tag three-jet, one-tag four-jet, two-tag two-jet, two-tag three-jet, two-tag four-jet. During the training phase, the MVA is told which events are signal and which events are background.

The trained MVA can then be given events without the information if it is signal or background, and it will provide the likelihood that a given event is a signal event. This likelihood is referred to as the "MVA output" throughout this analysis. Monte Carlo events, which are independent from those used to train the MVA, are each assigned an MVA output along with all of the data events to create the MVA distributions shown in Section 8.5.

8.2 Boosted decision trees (BDT)

For this thesis, the MVA method chosen was that of Boosted Decision Trees (BDTs). A decision tree (DT) is a method of classifying events as either signal or background by dividing a sample into many subsamples (nodes) based on a set of sequential decisions. This is done by splitting the sample into two nodes based on a cut, and then further splitting each of those nodes based on other cuts. This process is continued recursively until there are insufficient statistics to continue. The cuts are chosen such that the any given final node contains nearly all signal or nearly all background events. The process of training a DT consists of finding the set of cuts which best separate the signal and background. This process is described in detail in section 8.2.3.

The performance of a poorly performing MVA can often be improved by "boosting" it. In the boosting process, instead of training just a single MVA, many MVAs are trained, each on a slightly different set of events. The output from each tree in the set is then averaged together to form the final output. In this analysis a standard decision tree is boosted by training thirty copies of the tree using the Ada Boost method [63] which is described in more detail in section 8.2.4.

8.2.1 Sample preparation

The first step in training the BDT is to create two samples of Monte Carlo events, one containing signal events, one containing background events. As mentioned in section 8.1, one quarter of the total Monte Carlo samples are used for the training process. The second quarter of the events are used for training the combination MVA which is described in section A.3. The last half of the events are used when measuring the cross section.

If the same events were used for both the training and the cross section measurement, a bias could be introduced. This bias comes from a MVA being unable to differentiate between statistical fluctuations and real kinematic shape differences in the training sample. The net effect is that an ensemble of events which look exactly like the training sample will have slightly better separation between signal and background than an independent set of events. As the data is independent from the training sample, the MC that is used in the cross section measurement must also be independent from the training sample in order to have the same separation as the data.

Once the Monte Carlo training sample has been selected, the overall signal yield is normalized to the background signal yield. When boosting a DT, the first step is to re-weight some of the events. After this re-weighting, the overall signal yield is again normalized to the background signal yield. Throughout the process of training the DT, events referred to as "weighted" are those events which have the boosting weight and normalization applied to them. When training the first DT, the boosting weight is set to 1 for all events and only the normalization weight is applied.

8.2.2 Training

A decision tree is constructed by applying a cut dividing the sample into two subsamples: those which pass the cut and those which fail. Each of these subsamples is then further subdivided by applying cuts on each of them. The distinguishing feature of a decision tree is that each cut is chosen to maximize the signal purity in events passing the cut and minimize it in events failing the cut. Therefore, different cuts will be chosen for each node at a given level in the tree. Figure 8.1 shows a hypothetical DT with a set of cuts applied to it. This process is continued recursively until there are less than 100 events in a node, or until no cuts improve the signal to background ratios in the nodes. When these conditions are met, the node is considered to be a leaf.

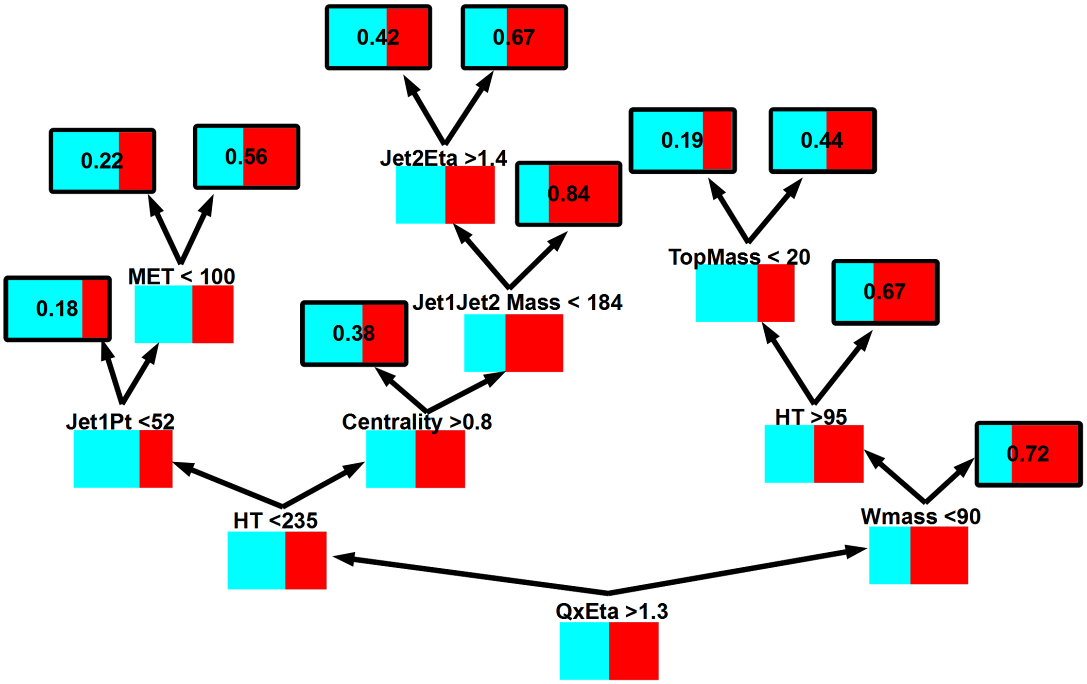


Figure 8.1 Example Decision Tree. Each red/blue box is a node. Relative amounts of signal and background in each node are shown in red and blue respectively. Leaves are outlined in black with the purity denoted on each one.

Each leaf is then assigned an output value based on its purity equal to

$$p = \frac{s_j}{s_j + b_j} \tag{8.2.1}$$

where s_j and b_j are signal and background yields in leaf j. When using the DT to classify unknown events, this purity is the output value used for all events ending in this leaf.

8.2.3 Node splitting

The process of choosing the best cut to split a given node is done by picking the cut causing the largest drop in node impurity. The impurity for any given node, i, is defined by the Gini Index [64]

$$i = \frac{sb}{s + b} \tag{8.2.2}$$

where s and b are the weighted number of signal and background events in the node. This measure of impurity is zero for nodes containing all signal or all background. It is maximal for nodes with equal amounts of signal and background, and symmetric in s and b. The change in impurity for any node splitting Δi is given by

$$\Delta i = i - (i_p - i_f) = \frac{sb}{s + b} - \frac{s_p b_p}{s_p + b_p} - \frac{s_f b_f}{s_f + b_f} \tag{8.2.3}$$

where the indexes f and p denote the node which fails and passes the cut respectively. Δi is 0 for a split which has the same signal to background ratios in both nodes and maximal when one node contains all signal events and the other contains all background events.

8.2.4 Boosting

Boosting can be used in order to improve the performance of any MVA, especially those which are weakly performing. This is achieved at the expense of additional complexity and computation time. In order to boost a DT collection of many DTs are trained, but each is trained on events with different weights. This results in a "forest" of decision trees which are all slightly different from each other. An unknown event is evaluated by using the average output of each of the DTs. In order to achieve optimal performance from the collection of trees, future trees must be trained such that they complement the trees already trained. In this analysis, this is achieved by using adaptive boosting, also known as Ada Boost [63].

After each tree is trained, events are examined to see if they are correctly classified as signal or background. Events which are misclassified are weighted by e^{α} (as defined in equation 8.2.4) and a new tree is trained on these re-weighted events. This process is iterated 30 times in this analysis to create a BDT with 30 boosts.

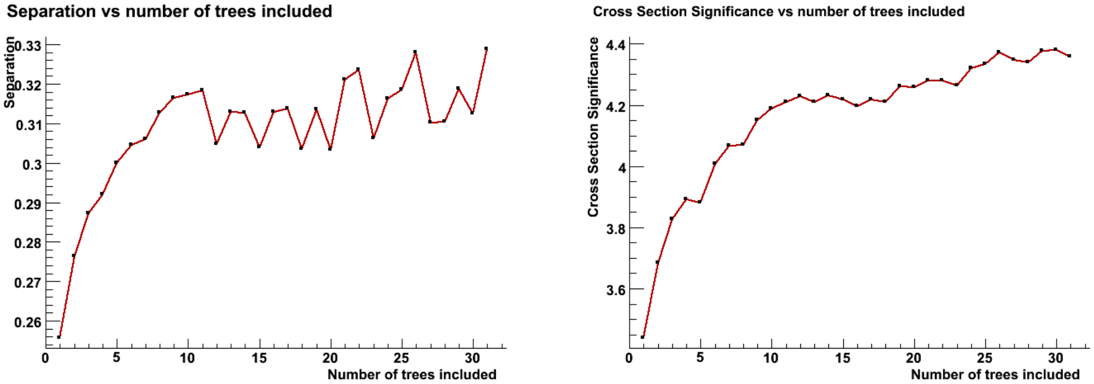
Signal events are considered to be misclassified if they have DT output < 0.5, and background events are considered misclassified if they have a DT output > 0.5. For each tree the value ϵ is calculated as the fraction of (weighted) events which are misclassified. The value α is then calculated as follows

$$\alpha = \beta \times ln \frac{1-\epsilon}{\epsilon} \tag{8.2.4}$$

where β is the Ada Boost parameter. Before training the next tree in the set, misclassified events are re-weighted by the factor e^{α} (which is larger than 1). After all 30 trees are trained, the final BDT output is given by

$$BDT_{\text{output}} = \frac{1}{\sum_{n=0}^{30} \alpha_n} \sum_{n=0}^{30} \alpha_n DT_{\text{output},n} \tag{8.2.5}$$

Figure 8.2 shows the performance as a function of the number of boosts used for one channel. Performance is calculated as separation (left) and cross section expected significance (right). Separation is defined as $\sum_{i=0}^{N} \frac{(s_i - b_i)^2}{s_i + b_i}$ where there are N bins and s_i and b_i are the signal and background yields in each bin respectively. Cross section expected significance is defined as $\sum_{i=0}^{N} \frac{s_i^2}{s_i + b_i}$. The performance increases as additional boosts are used until it reaches a maximum. This occurs at around 10-20 boosts, depending on the channel. The choice of 30 boosts therefore provides high performance while avoiding unnecessary complexity.


Separation vs number of trees included

Cross Section Significance vs number of trees included

Figure 8.2 Example BDT performance as function of number of boosts using the metric of separation (left) and cross section expected significance (right)

8.2.5 BDT parameter settings

The BDT implementation used in this analysis is the `classifier` package found in the DØ CVS repository. This package allows several BDT parameters to be tuned for the training step.

- **Number of boosts:** The number of trees used in the boosting phase of the training. For this analysis 30 trees are used, as discussed in the previous section.
- **Ada boost parameter:** Determines how much weight to give to misclassified events. Used in equation 8.2.4. A value of 0.2 is used in this analysis.
- **Minimal leaf size:** If a leaf has fewer than this number of events, it will not be sub-divided further. A value of 100 is used here.
- **Impurity measure:** This analysis uses the Gini impurity metric, which is shown in equation 8.2.2.
- **Bagging:** An alternative method of boosting. It is turned off for this analysis.

- **Pruning:** An additional method which reduces the number of leaves in a tree. No pruning was used in this analysis.

A study was performed to optimize these parameters for the single top observation analysis which can be seen in [64]. This study showed that the performance of the method was fairly insensitive to changes in all parameters except for the number of boosts. Therefore, for this analysis, the number of boosts was studied and optimized while parameters were set to the same value as in the observation analysis.

8.3 BDT variable selection

One of the most important steps in training a BDT is choosing the set of variables which are used as inputs to the method. If too few variables are used, there will be insufficient information to perform a good separation. The performance of a BDT only degrades slightly if too many input variables are used. Adding additional variables does increase the complexity of the analysis; therefore, the total number of variables should be kept under control. Figure 8.3 shows the performance of a BDT as a function of the number of input variables used. For this analysis 50 input variables were used. The same 50 variables are also used for each of the six jet and tag channels in order to minimize the complexity of the method.

The first step in choosing a set of input variables is rejecting those variables with poor agreement between data and MC background. This is done by comparing the data and the MC distributions for each variable using the ROOT implementation of the Kolmogorov-Smirnov (KS) test.

The KS test compares the two distributions and gives an output between 0 and 1. If the two distributions are sampled from the same underlying probability distribution function (i.e. the variable is well modeled), the KS values will be distributed uniformly. If the two distributions are not sampled from the same underlying distribution, the KS

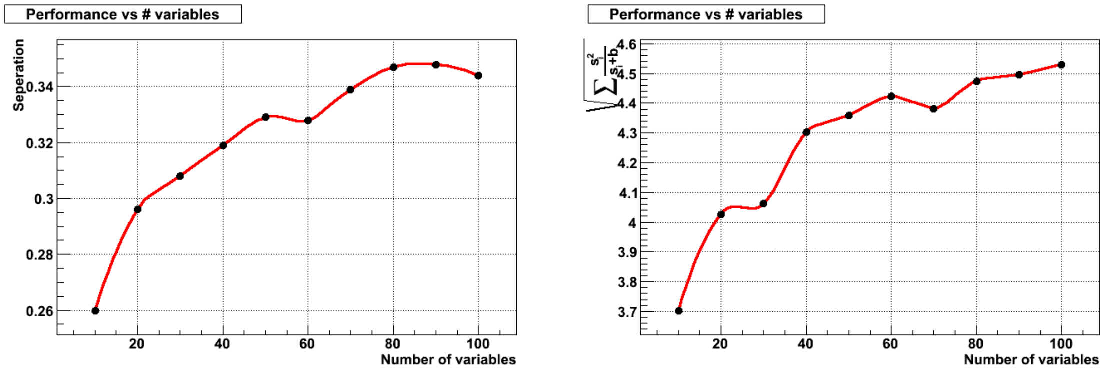


Figure 8.3 Example BDT performance as function of number of input variables using the metric of separation (left) and cross section expected significance (right)

distribution will be strongly biased towards 0. If a cut is placed to accept only variables with a KS > 0.10, then it is expected to remove 10% of the well modeled variables and much more than 10% of the poorly modeled variables.

As each variable in this analysis has six distributions (one for each jet and tag channel), each variable has six KS values. In order to keep a variable for inclusion in this analysis, the average of these six KS values must be greater than 0.30. This cut will reject 4.5% of well modeled variables. If the KS value is below 0.01 in any of the six channels, the variable is also rejected. This cut will reject about 5.9% of all well modeled variables. These two cuts combined will reject 9.5% of well modeled variables and the large majority of poorly modeled variables.

After applying these cuts, 136 variables remain. All of these variables are used to train six "trial" BDTs. Each of the trial BDTs contains 30 DTs, all of which list the importance factor for each variable. This factor is calculated for each variable by taking the sum of each Δi (from equation 8.2.3) over all node splits involving that variable. The importance of each variable across all 180 lists is averaged, and the variables are ranked according to their average values. The top 50 variables on this list are kept and used to train the final six BDTs for the analysis.

These 50 variables are described in the list below, and Figures 8.6-8.9 show the data/MC comparison for each of these variables for the 2-4 jet 1-2 tag sample.

The naming convention of the variables is as follows. Jets are ordered in p_T such that "jet1" is the jet with the largest p_T, "jet2" has the second largest and so forth. In the samples with exactly one tag, the "leading tagged jet" and "most b-like jet" both refer to the only jet passing the b-tagging criteria. In the samples with exactly two tags, the "leading tagged jet" is the jet passing the b-tagging criteria which has the largest p_T, the "second tagged jet" is the other tagged jet, and the "most b-Like jet" is the jet with a higher b-tag NN output. The "leading light quark jet" and "second light quark jet" are the untagged jets ranked in p_T order. The "best jet" refers to the jet which, when combined with the lepton and neutrino, forms an invariant mass closest to the top quark mass.

Several of the variables depend on properties of a reconstructed top quark or W boson. The W boson decays into an lepton and neutrino; however, the neutrino p_z can not be measured. In order to reconstruct the neutrino p_z the mass of the W boson mass (80.43 GeV) and the lepton four momentum are used to constrain the neutrino p_z to two values. The smaller of these two values is the default choice. If the larger solution is used it will be denoted as S2. In order to reconstruct the top quark, the four vectors of the lepton, neutrino, and a jet are combined. For each choice of jet, there is a corresponding reconstructed top quark.

Scatter plots were produced for every pair of input variables for both data and MC. These 5000 plots were compared by eye and show a high level of agreement. The correlation coefficient for data and MC was calculated from each of these plots. Figure 8.4 shows the difference in the data and MC correlation coefficients for each pair of input variables. This again demonstrates that the MC background model is a good representation of the data.

- **Object kinematic variables:**

  - **LeadingLightQuarkJetE:** Energy of leading light quark jet
  - **MostBLikeJetDetEta:** η_{det} of most b-like jet
  - **LeptonPt:** p_T of lepton
  - **BestJetBTagNN:** NNoutput of best jet
  - **LeptonDetPhi:** ϕ_{det} of lepton
  - **BestJetPt:** p_T of best jet
  - **NuPz S1:** p_z of neutrino
  - **MostBLikeJetEtaW:** Width in η of most b-like jet
  - **NotBest1EtaW:** Width in η of leading jet that isn't best jet
  - **NotBest1PhiW:** Width in ϕ of leading jet that isn't best jet
  - **NotBest2PhiW:** Width in ϕ of second leading jet that isn't best jet
- **Event variables:**
  - **Centrality AllJets:** $\frac{\sum H_T}{\sum E}$
  - **PVz:** Primary Vertex z position
  - **METPhi:** ϕ direction of $\not{E}_T$
  - **METPt:** p_T of $\not{E}_T$
  - **IsRunIIb:** RunIIa vs RunIIb
- **Charge of the lepton time eta:**
  - **QTimesEta:** Charge of the lepton times η of leading light jet
  - **NotBest1QTimesEta:** Charge of the lepton times η of the leading not best jet
  - **Jet1QTimesEta:** Charge of the lepton times η of the leading jet

  - **MostBLikeJetQTimesEta:** Charge of the lepton times η of the most b-like jet
  - **QTimesLeptonEta:** Charge of the lepton times η of the lepton
- **Reconstructed objects:**
  - **Pt Jet1Jet2:** Transverse momentum of the object constructed from the leading two jets
  - **InvariantMass Jet1Jet2:** Invariant mass of the object constructed from the leading two jets
  - **InvariantMass LightQuarkJets1 2:** Invariant mass of the object constructed from the two leading light jets
  - **InvariantMass AllJets:** Invariant mass of the object constructed from all jets
  - **InvariantMass AllJets MinusBTaggedJet:** Invariant mass of the object constructed from all jets except the leading tagged jet
  - **InvariantMass AllJets MinusBestJet:** Invariant mass of the object constructed from all jets except the best jet
  - **HT AllJetsLeptonMET:** H_T sum of all jets, the lepton, and the $\not\!\!E_T$
  - **HT AllJets MinusBestJet:** H_T sum of all jets except the best jet
  - **HT METlep:** H_T sum of the lepton and $\not\!\!E_T$
- **Angles between objects:**
  - **DeltaRmin AllJets:** Smallest ΔR between any two jets
  - **DeltaPhiLeptonMET:** $\Delta\phi$ between the lepton and $\not\!\!E_T$
  - **DeltaPhiJet2MET:** $\Delta\phi$ between jet2 and the $\not\!\!E_T$

  - **MostBLikeJetLeptonDeltaPhi:** $\Delta\phi$ between lepton and most b-like jet
  - **Cos LeptonBTaggedTopFrame BTaggedTopCMFrame:** Cosine between lepton (in tagged top frame) and the tagged top (in the cm frame) where the top quark is constructed from leading tagged jet and W
  - **Cos LeptonBestTopFrame BestTopCMFrame:** Cosine between lepton (in best top frame) and the best top (in the cm frame)
  - **Cos LightQuarkJetLepton BTaggedTop:** Cosine between leading light quark jet and the lepton in the tagged top frame
  - **Cos BTaggedJetLepton BTaggedTop:** Cosine between leading tagged jet and the lepton in the tagged top frame
  - **Cos LeptonQZ BestTop:** Cosine between the lepton and the (z axis)*(lepton charge) in the best top frame
- **Reconstructed** W
  - **Sphericity AllJetsW:** Sphericity of all jets plus the W
  - **Aplanarity AllJetsW:** Aplanarity of all jets plus the W
  - **WPt:** p_T of the reconstructed W
  - **WTransverseMass:** Transverse Mass of the reconstructed W: $\sqrt{m^2 + p_x^2 + p_y^2}$
- **Reconstructed Top:**
  - **TopMassMinChiSqr:** Minimum value of (constructed Top mass - theory top mass)^2 for all possible reconstructed top quarks
  - **ChiTopMass:** Reconstructed top mass which has the closest mass to MC
  - **TopMassMinSig:** Lowest significance of the top mass for reconstructed top quarks

- **Jet1TopMass S2:** Invariant mass of the top quark reconstructed from the leading jet and the W that is constructed using the larger ν_{pz} value
- **MostBLikeJetTopMass:** Invariant mass of the top quark reconstructed with the most b-like jet
- **MostBLikeJetTopMass S2:** Invariant mass of the top quark reconstructed with the most b-like jet and the W that is constructed with the larger ν_{pz} value
- **LeadingBTaggedJetTopMass:** Invariant mass of the top quark which is constructed with the leading b-tagged jet

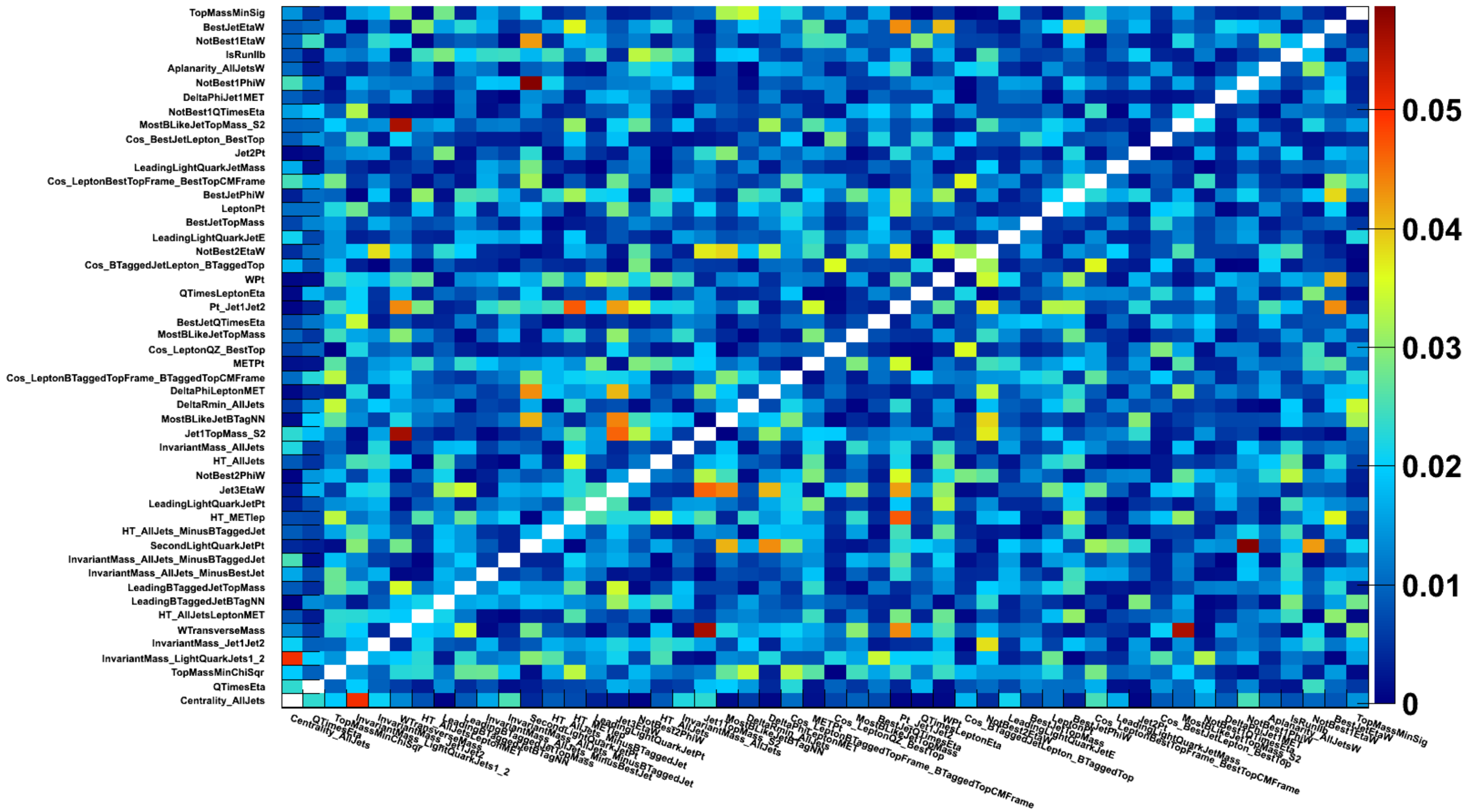


Figure 8.4 Correlation coefficient differences between data and MC. Calculated by considering every pair of kinematic variables in data and in MC. For each pair, the correlation coefficient was calculated in the data and MC. The absolute value of the difference is shown in the figure.

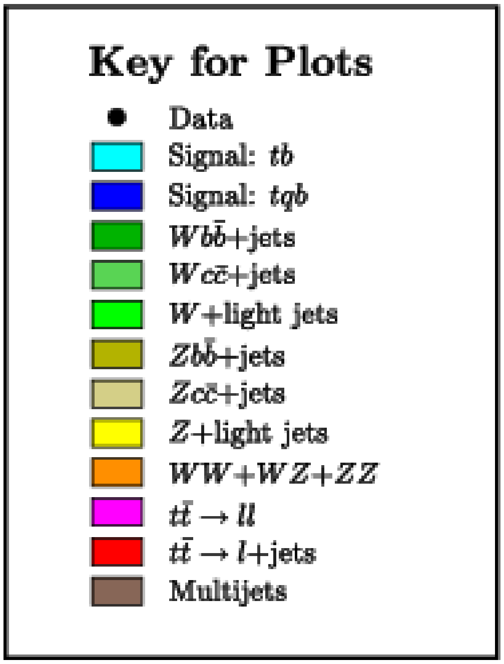


Figure 8.5 Legend for all data mc comparison plots used in this analysis.

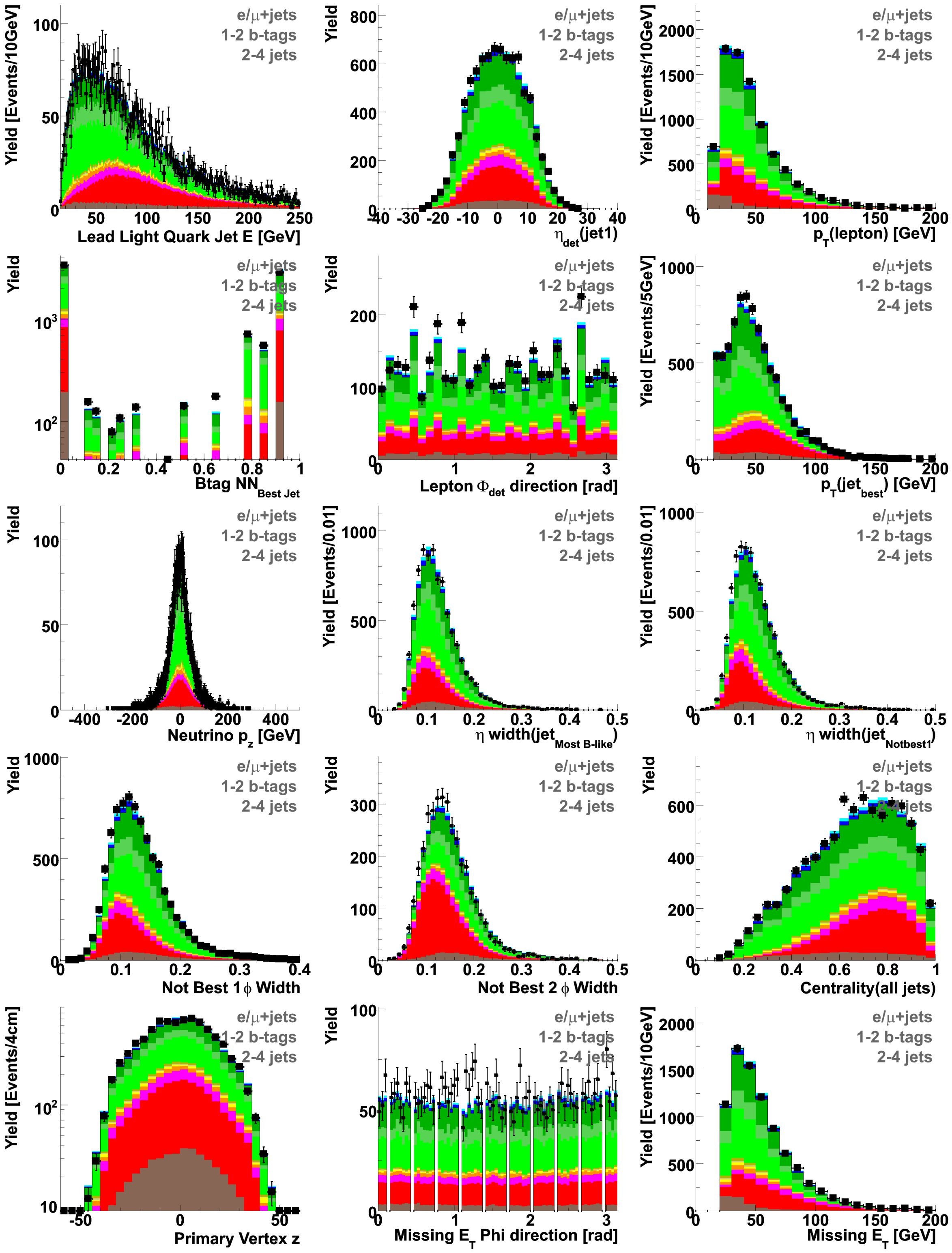


Figure 8.6 Data vs Background Model Comparison

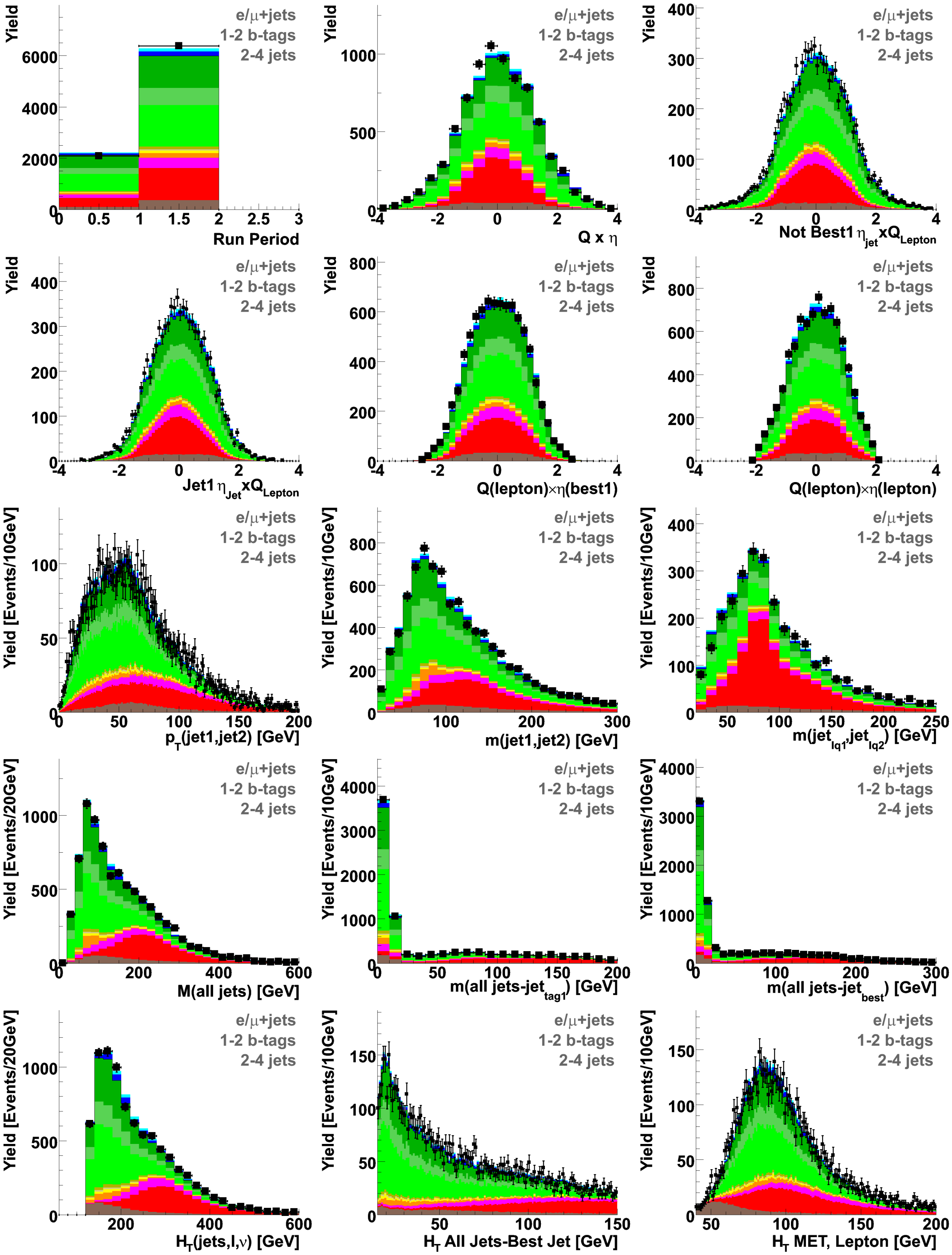


Figure 8.7 Data vs Background Model Comparison

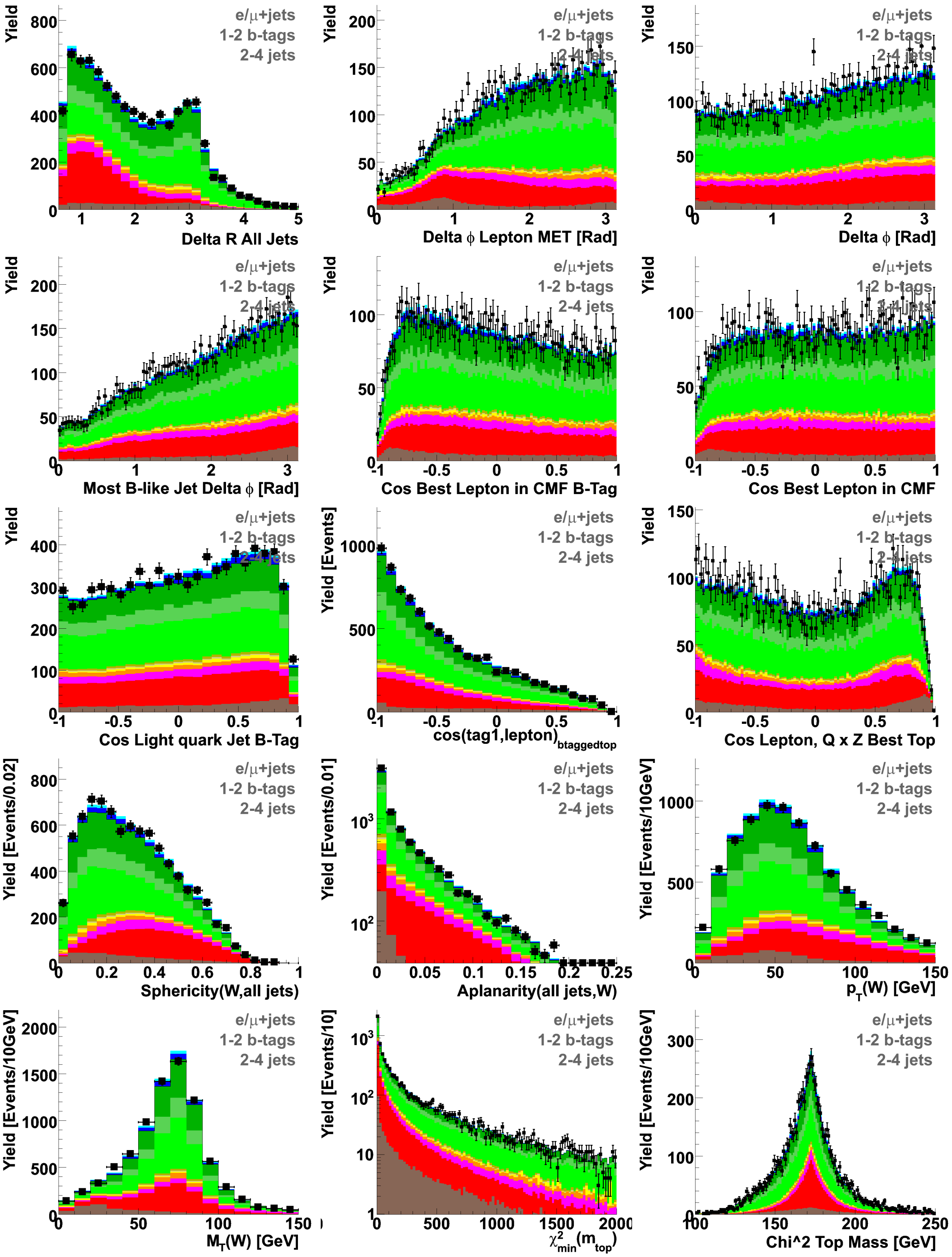


Figure 8.8 Data vs Background Model Comparison

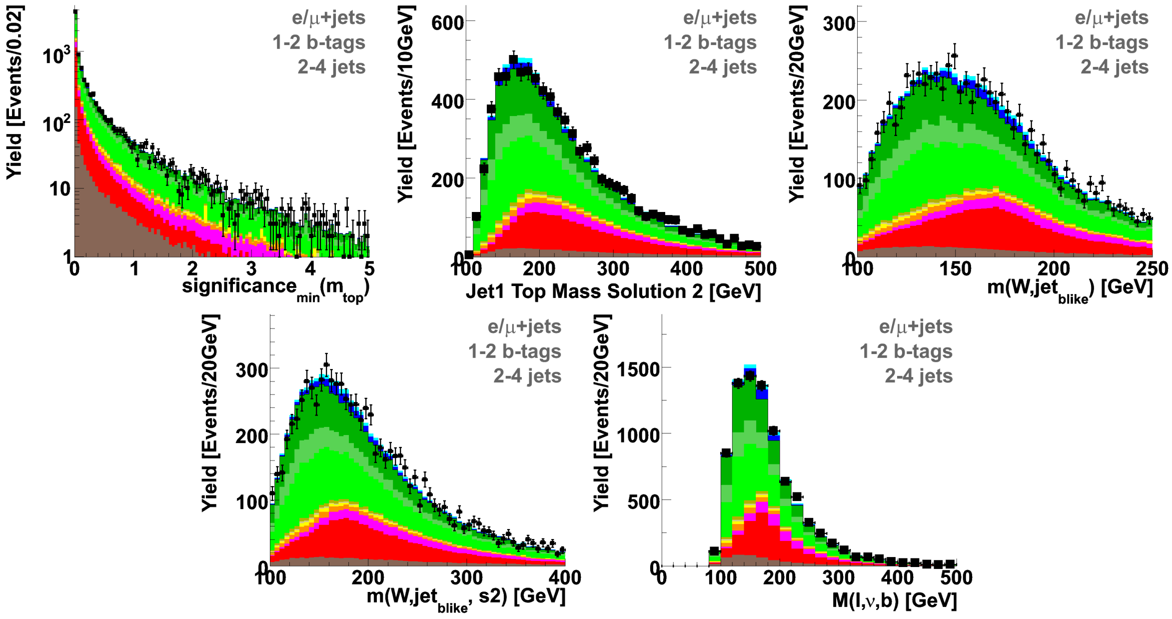


Figure 8.9 Data vs Background Model Comparison

8.4 Binning transformation

While boosting has the effect of increasing the separation of the signal and background distributions, it also causes both distributions to peak closer to 0.5 while leaving the regions close to 0 and 1 relatively empty. This effect is due to the fact that the BDT output is an average over 30 DTs. As these 30 inputs will be distributed over the 0 to 1 range, the average of them is likely to be close to 0.5 and unlikely to be close to 0 or 1. This will cause problems when attempting to evaluate the uncertainty on the expected backgrounds since some bins may have no background events. In order to overcome this problem, the BDT outputs are transformed to give a background a shape that approximates $1/x$. The transformation is then constrained so that when put into a histogram with 25 bins, each bin will contain at least 64 effective background events.

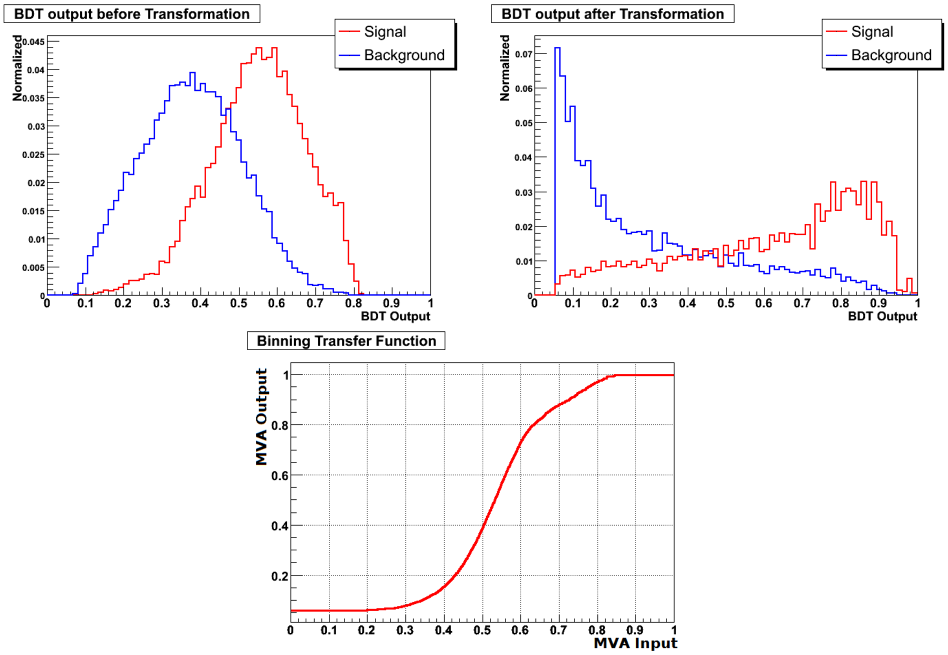


Figure 8.10 BDT output for signal and background before (top left) and after (top right) the transformation function is applied. The transformation function, $T(x)$ (bottom).

The transformation function, $T(x)$, shown in Figure 8.10 is defined such that the background follows a k/x curve in the region between 0 and 0.8, is linear from the region 0.8 to 0.95, and constrained to have at least 64 effective events in each bin above 0.95. The value of k is chosen such that the area under the curve is equal to 1. The slope and intercept of the linear fit are chosen such that the overall function is continuous. Figure 8.10 shows the shapes of the discriminant before and after the transformation is applied.

Once the transformation function is defined (Figure 8.10 bottom), it can be applied to the BDT output. This is done by taking

$$BDT_{\text{transformed output}} = T(\text{BDT}_{\text{output}})$$

for each event. This transformation is equivalent to using variable bin widths in the original histogram that are much wider close to 0 and 1 and thinner close to 0.5 [32]. The transformed output is now guaranteed to contain some background events in every bin which contains signal events which keeps the signal to background ratio finite in all bins and avoids any potential divide-by-zero problems.

8.5 BDT output shapes

After the transformation, the outputs of the BDT for the six channels are shown in Figure 8.11.

8.6 BDT cross checks

After training the six BDTs using one fourth of the total MC data sample, an additional validation step is performed by creating two cross check samples. These samples are formed by using the second fourth of the MC sample, but with additional cuts applied in order to make each of the two cross check samples background dominated.

The $t\bar{t}$ crosscheck sample is formed by taking events with exactly four jets and an H_T > 300 GeV. The high jet multiplicity and large amount of transverse energy in the event results in a sample that is almost entirely dominated by $t\bar{t}$ events. The W + jets cross check sample is formed by considering events with exactly one tagged and one untagged jet and with $H_T < 175$ GeV. The cuts in jet multiplicity and H_T ensure that the two cross check samples contain none of the same events. However, as the cross check samples are both subsets of the normal dataset, all data events in the cross check samples are present in the nominal sample.

The MVA outputs for the two cross check samples are shown in Figure 8.12. The data and MC background models agree well with each other for both of the two main backgrounds in the analysis. This inspires some confidence that the MC not only agrees for each input variable, but that any important higher order correlations are also well modeled in the MC.

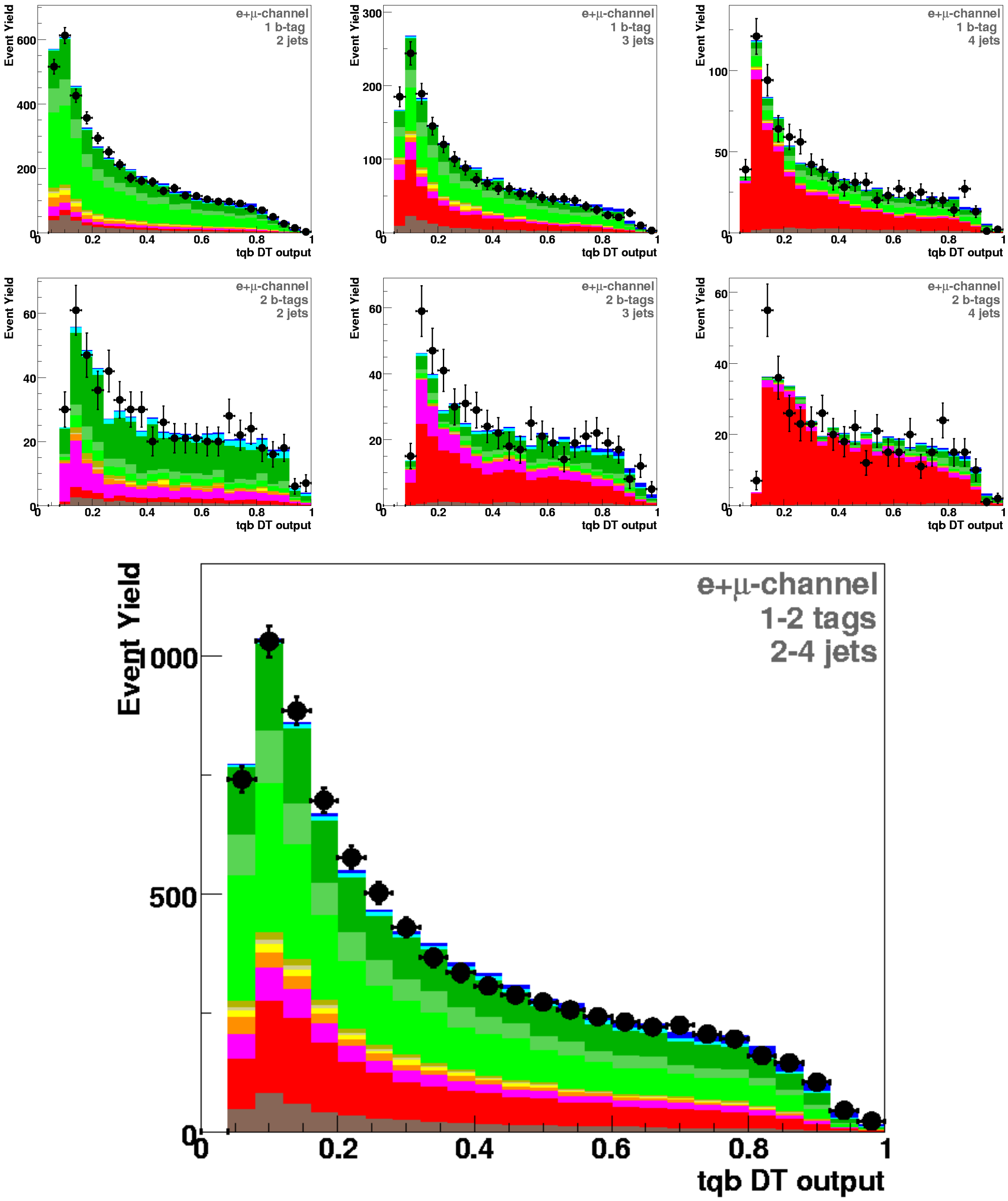


Figure 8.11 Transformed BDT outputs for each jet and tag multiplicity. Top row shows outputs for one tag events. Middle row shows outputs for two tag events. Bottom plot is the sum over all six channels.

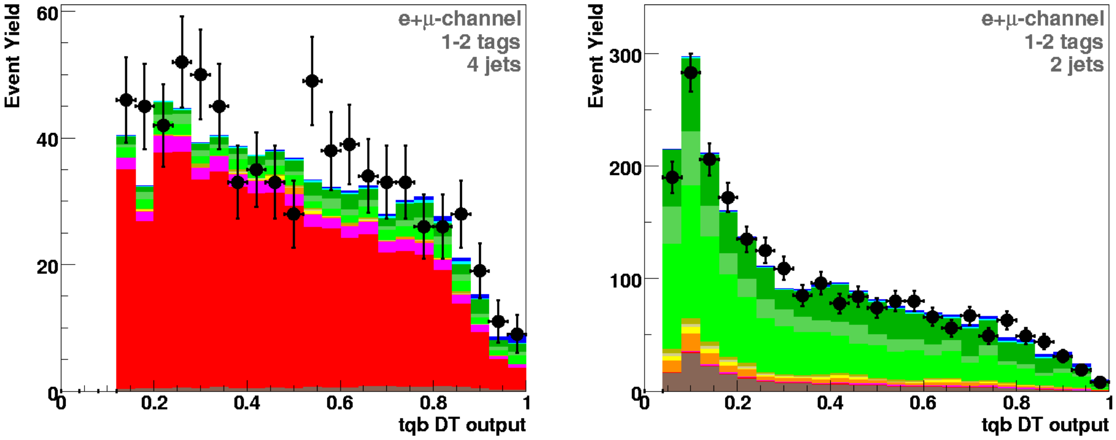


Figure 8.12 $t\bar{t}$ cross check sample on the left and W + jets cross check sample on the right

Chapter 9. Cross Section Measurement

9.1 Bayesian analysis techniques

In this analysis, the cross section measurement and its uncertainties are calculated using the `top_statistics` package [65]. While this package will generate both Bayesian and Frequentist confidence intervals, only the Bayesian methods are used for this measurement. The `top_statistics` package also provides several methods to validate the measured result as described in sections 9.3 and 9.2. For this analysis, the standard model s-channel single top cross section is used and is considered a background process.

9.1.1 Cross section measurement

The cross section measurement uses the 6 BDT outputs shown in figure 8.11. Each of these histograms contains 25 bins, and the entire 150 bins are treated as independent measurements with correlated uncertainties. The likelihood to observe D events in a bin with a mean number of events d is given by the Poisson distribution:

$$p(D|d) = \frac{e^{-d}d^{D}}{\Gamma(D+1)} \tag{9.1.1}$$

where Γ is the gamma function and $\Gamma(D+1) = D!$ for integer values of D. When observing data, D is always an integer; however, in section 9.2 the MC is used to estimate the data and D can take on non-integer values.

The mean number of events is estimated from the MC yields such that:

$$d = \alpha\mathcal{L}\sigma + \sum_{i=1}^{N} b_i \equiv a\sigma + \sum_{i=1}^{N} b_i \tag{9.1.2}$$

where α is the signal acceptance, $\mathcal{L}$ is the luminosity, and a and b_i are the number of signal and background events expected to pass the event selection cuts.

The 150 bins in the histograms can be treated as a distribution of observations, and the total likelihood is the product of the likelihoods for each individual bin [65].

$$L(\vec{D}|\vec{d}) \equiv L(\vec{D}|\sigma, \vec{a}, \vec{b}) = \prod_{j=1}^{150} L(D_j|d_j) \tag{9.1.3}$$

In this equation, $\vec{D}$, $\vec{d}$, $\vec{a}$, and $\vec{b}$ are the 150 dimensional vectors of each respective variable over the 150 histogram bins. This this is possible because the probability to observe a given number of events in one channel is independent of the yields in the other channels. Bayes theorem [66] can then be applied to compute the posterior probability density $p(\sigma, \vec{a}, \vec{b}\,|\vec{D})$. This can be integrated with respect to $\vec{a}$ and $\vec{b}$ to find the posterior density for the signal cross section given the values of $\vec{D}$ [65].

$$p(\sigma|\vec{D}) = \frac{1}{\mathcal{N}} \int\int L(\vec{D}|\sigma, \vec{a}, \vec{b})\pi(\sigma, \vec{a}, \vec{b}) d\vec{a} d\vec{b} \tag{9.1.4}$$

$\mathcal{N}$ is a normalization constant obtained by setting $\int_0^{\sigma_{\max}} p(\sigma|D)d\sigma = 1$. The function $\pi(\sigma, \vec{a}, \vec{b})$ is the prior probability density. The assumption is then made that the prior knowledge of $\vec{a}$ and $\vec{b}$ are independent of σ and the prior probability density can be factorized.

$$\pi(\sigma, \vec{a}, \vec{b}) = \pi(\vec{a}, \vec{b}|\sigma) = \pi(\vec{a}, \vec{b})\pi(\sigma) \tag{9.1.5}$$

The prior in σ is assumed to be flat over the range from 0 to $\sigma_{\max}$. This implies maximal ignorance of a preferred value and is a common convention. The posterior probability density is then

$$p(\sigma|\vec{D}) = \frac{1}{\mathcal{N}\sigma_{\max}} \int\int L(\vec{D}|\sigma, \vec{a}, \vec{b})\pi(\vec{a}, \vec{b}) d\vec{a} d\vec{b} \tag{9.1.6}$$

This equation can be solved numerically resulting in a posterior probability density curve. The cross section measurement is taken as the mode of the curve and the 68% interval around the mode is taken as the estimate of the uncertainty as demonstrated in Figure 9.1. It has been shown that this interval can be interpreted as an approximation of a Frequentist one sigma interval [65].

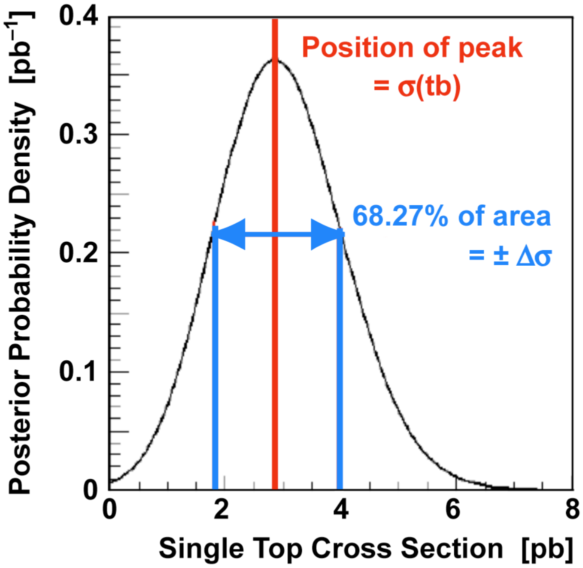


Figure 9.1 Example of posterior probability density distribution. Peak position is the measured value, width of the posterior is the measured uncertainty [65].

9.1.2 Treatment of systematic uncertainties

Two different types of systematic uncertainties are considered in this analysis: those which only change the normalization of any given sample and those which change the shape of the MVA output for a given sample. In both cases, the systematic uncertainties are included as part of the prior probability density and used in the numerical calculation of the posterior probability density. This calculation is performed by sampling 1M points from the prior density. This yields:

$$p(\sigma|\vec{D}) = \frac{1}{\mathcal{N}\sigma_{\text{max}}N_{\text{samples}}} \sum_{k=1}^{N_{\text{samples}}} L(\vec{D}|\sigma, \vec{a_k}, \vec{b_k}) \tag{9.1.7}$$

where $\vec{a_k}$ and $\vec{b_k}$ include the systematic uncertainties [65].

For normalization uncertainties, $\vec{a_k}$ and $\vec{b_k}$ are sampled from the MC as Gaussians with means set to the expected yields and widths corresponding to the normalization uncertainties.

For systematics which change the shape of the distribution, the sampling is done by measuring the differences between the nominal and the σ_+ and σ_- distributions on a bin by bin basis. The MC is sampled as a Gaussian with different positive and negative widths, and the mean is set equal to the expected yield on the nominal sample. Figure 9.2 illustrates this process.

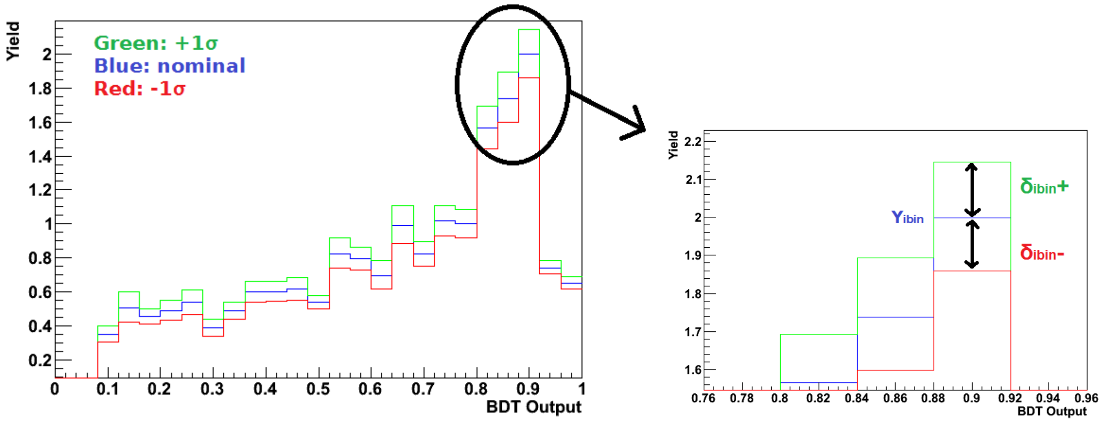


Figure 9.2 Treatment of shape changing systematic uncertainties in `top_statistics`. Bin by bin difference in the nominal and systematic samples is used to to define the Gaussian used for sampling the distributions [65].

9.2 Expected result

Once a cross section measurement is performed on the actual data, any further changes in the analysis method may bias the measurement. For example, if trying to decide between two values of a given parameter, the temptation is to use the one which gives the best observed result, even if it would not otherwise be the optimum value.

To avoid this source of potential bias, one first performs the cross section measurement using the MVA outputs containing only the MC values and not the data. One substitutes the distributions of the data with the total MC distributions. This value obtained by this method is known as the "expected result". By using the expected result, the analysis can be safely tuned to maximize the significance of the measurement. In addition, several validation steps are performed at this stage.

First, many choices of an upper bound on the prior (σ_{max}) were tried. The measurement of the cross section was insensitive to the choice of a prior as long as it covered a range sufficiently larger than the expected value. The chosen value for σ_{max} was 20 pb. In a similar manner, the number of numerical integration steps was varied. According to the results, as long as the number of steps was sufficiently high, the measurement was very stable. A value was chosen for the number of integration steps which was approximately five times larger than the minimum needed for a stable result.

A linearity test and expected significance test were then performed using the expected results. These are described in sections 9.3 and 9.5.

Using the final choice of parameters, the expected cross section measurement was performed as reflected in the posterior probability curve in figure 9.3. The measurement of the t-channel cross section is expected to be $2.40^{+0.71}_{-0.66}$ pb.

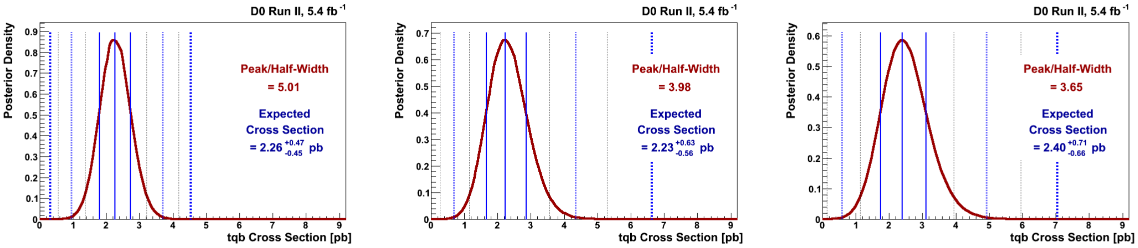


Figure 9.3 Expected results: no systematics(left), flat systematics (center), full systematics (right)

9.3 Linearity test

In order to determine if there is any bias in the cross section measurement, a linearity test is performed. For the linearity test, ensembles are generated with the nominal background yields but use several different signal yields. The signal cross section is measured in each ensemble and compared to the cross section used to generate each ensemble.

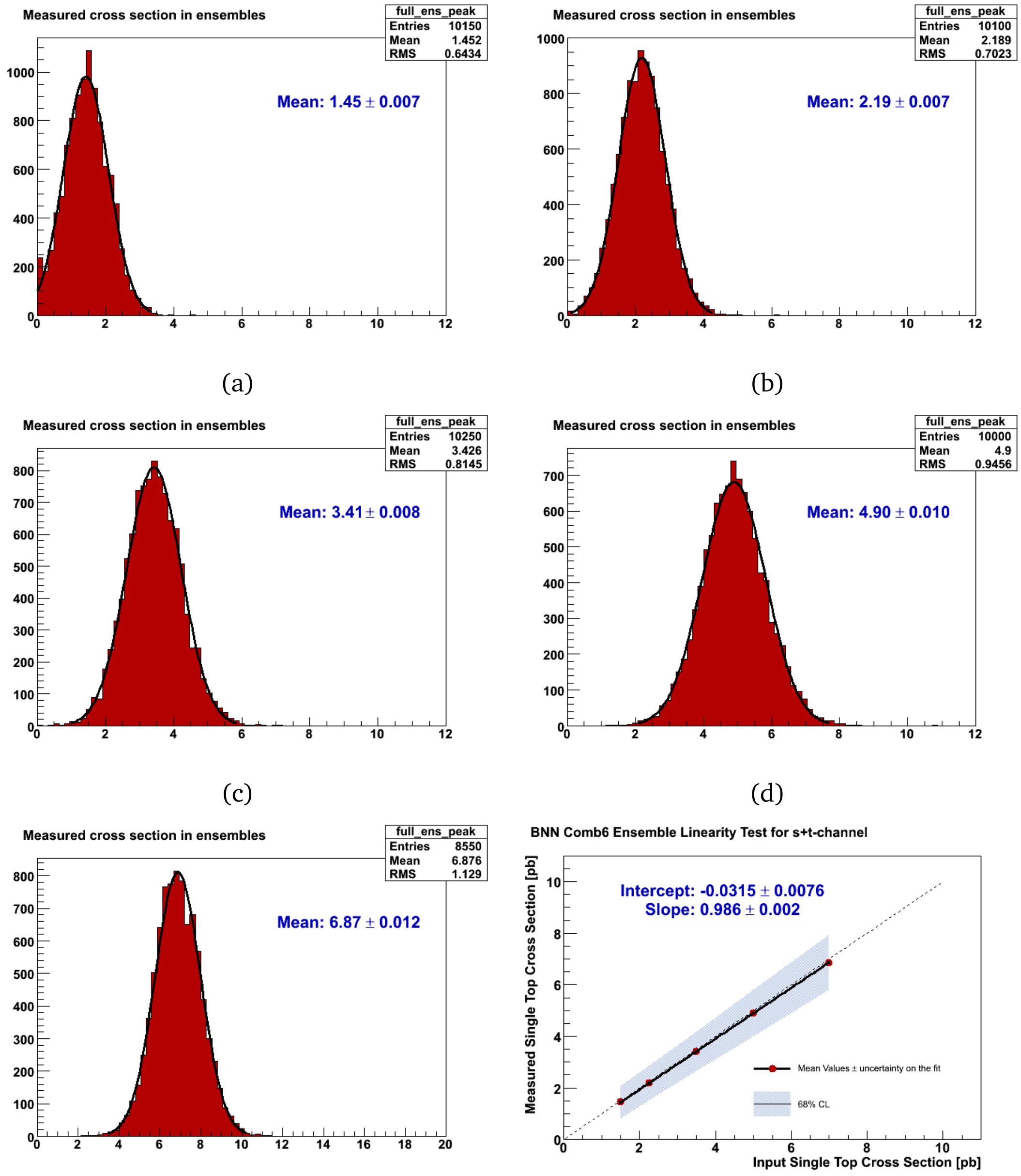
Each ensemble is formed by making 200 measurements of the cross section with the specified signal:background ratios. All systematic uncertainties are taken into account by sampling events as described in Section 9.1.2. Ensembles are generated with the following cross sections: 1.5 pb, 2.26 pb (nominal), 3.5 pb, 5 pb, and 7 pb.

Figure 9.4 shows the cross section distributions for each of the five ensembles, as well as the comparison of the input and measured signal cross sections. A Gaussian was fit to each of the five distributions. The mean of the Gaussian corresponds to the measured cross section and the width of the Gaussian corresponds to the uncertainty of the measurement.

9.4 Observed Results

After the validation steps have been finished, the measurement of the cross section can be performed using the actual data set. Figure 9.5 shows the BDT output histogram for all six channels summed together across the entire range of output values as well as zoomed into the signal rich region. This combined BDT output is provided only to understand the shapes of the distributions. The actual measurement uses the six histograms individually.

The cross section measurement can be performed in many different modes. First, the cross section can be calculated using only statistical uncertainties (no systematics). When calculating the cross section using systematic uncertainties, both shape changing



(a) (b)

(c) (d)

(e) (f)

Figure 9.4 Ensemble results for each of the five signal:background ratios (a-e), and correlation between input and measured signal cross section (f)

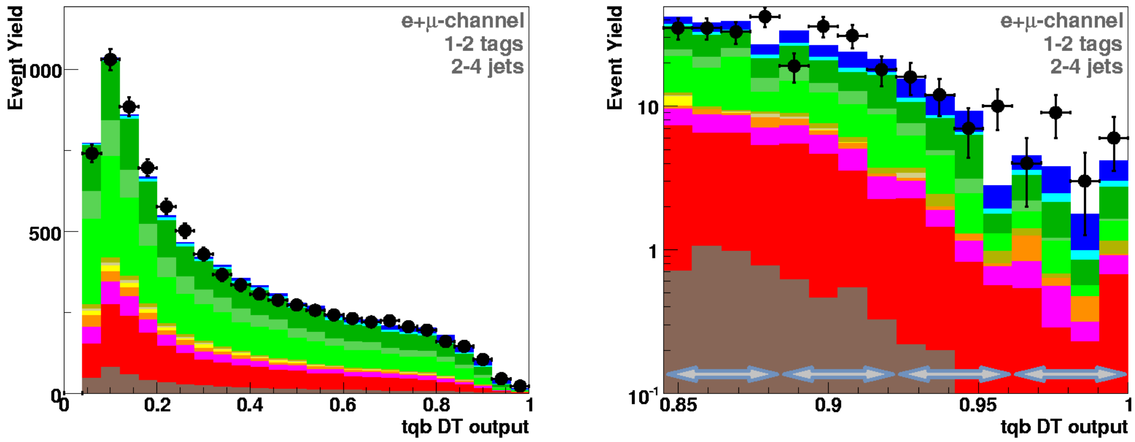


Figure 9.5 Final BDT(left). Final BDT zoomed into the signal-like region, each bin divided into 4 bins only for plotting purposes. Blue arrows show binning used for the cross section measurement(right)

and normalization uncertainties are considered (full systematics). Lastly, the cross section measurement is performed on each of the six channels individually as well as on the entire sample. Table 9.1 summarizes all of these measurements and Figure 9.6 displays the posterior probability density curves for the full systematics measurements.

Channel	No Systematics (pb)	Full Systematics (pb)
One Tag, Two Jets	$2.16^{+0.56}_{-0.53}$	$2.96^{+1.05}_{-0.95}$
One Tag, Three Jets	$2.19^{+0.92}_{-0.87}$	$3.66^{+1.53}_{-1.27}$
One Tag, Four Jets	$2.00^{+1.75}_{-1.64}$	$0.67^{+2.67}_{-0.67}$
Two Tags, Two Jets	$7.09^{+4.95}_{-4.71}$	$2.67^{+6.50}_{-2.67}$
Two Tags, Three Jets	$5.53^{+6.53}_{-5.53}$	$4.04^{+2.36}_{-2.02}$
Two Tags, Four Jets	$0.16^{+3.36}_{-0.16}$	$2.58^{+3.38}_{-2.58}$
One-Two Tags, Two-Four Jets	$2.33^{+0.45}_{-0.44}$	$3.03^{+0.78}_{-0.66}$

Table 9.1 Observed cross section measurement results.

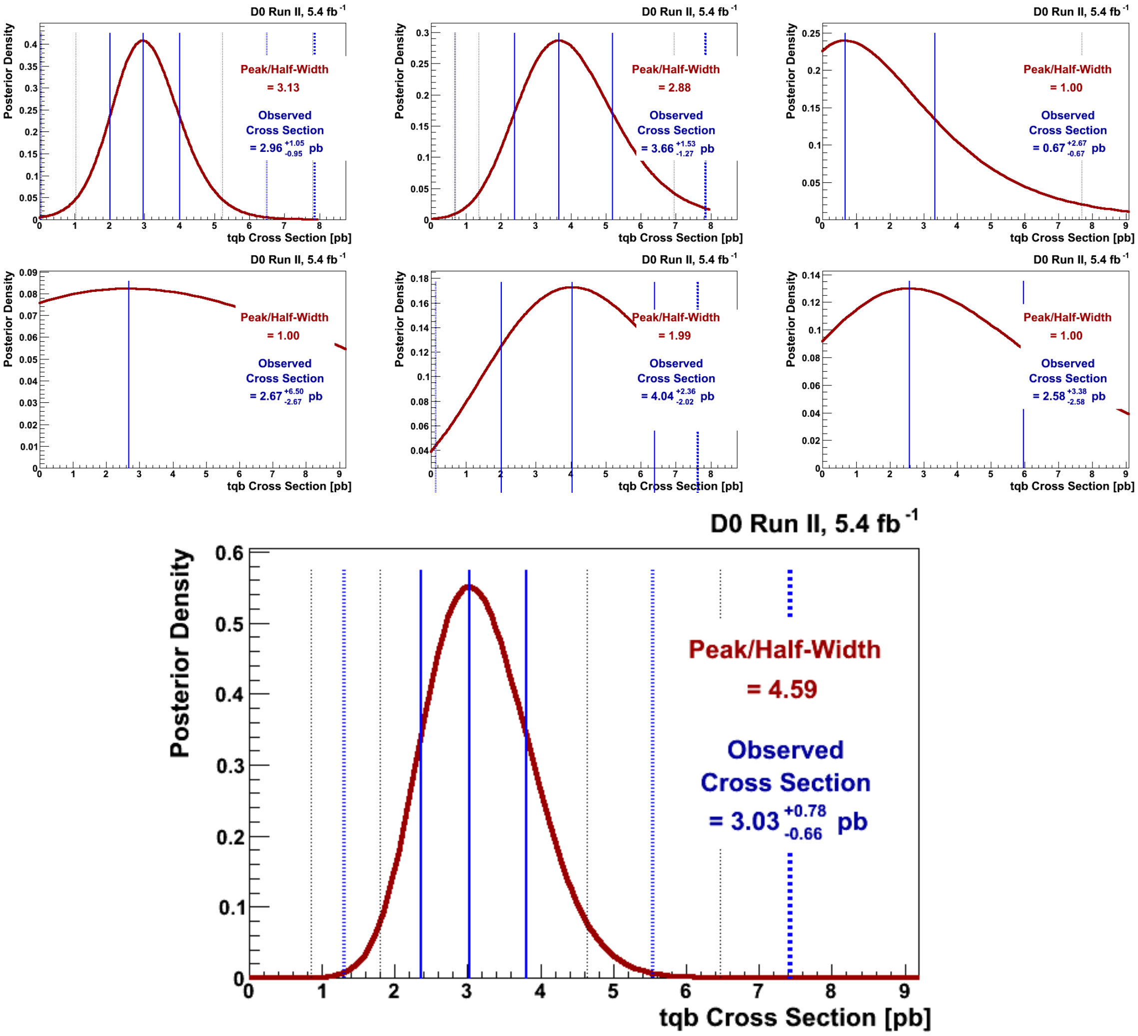


Figure 9.6 Observed posterior probability density curves with full systematics. Top row shows one tag events. Middle row shows two tag events. Bottom plot shows all events combined.

9.5 Signal significance

The significance of a measurement is one minus the probability making this measurement assuming null hypothesis is true. The null hypothesis, in this for the single top cross section measurement, is the hypothesis that there are no actual single top events in the dataset. The probability of measuring the observed cross section (or larger), assuming this hypothesis, is referred to as the "p-value" and denoted by the symbol α. The significance is usually quoted in units of standard deviations, σ, and can be tranformed from α through the equation:

$$\sigma = \sqrt{2}\,\mathrm{erf}^{-1}(1 - 2\alpha).$$

9.5.1 Asymptotic approximation of the log-likelihood ratio

The significance of the cross section measurement is calculated using the asymptotic approximation of the log-likelihood ratio (AALLR) [67]. This differs from the single top observation significance calculation which used a large number of ensembles. To repeat the previous significance calculation for all of the measurements made with the 5.4 fb^{-1} dataset would require at least a month of computing time using the DØ computing cluster. The asymptotic approach, however, requires roughly one hour using a single computer to calculate each significance. As it has been demonstrated that the two approaches are consistent with each other [62], the time saving outweighs the benefit of using the same approach as in previous analyses.

To calculate a p-value using the AALLR method the log-likelihood ratio must first be defined as:

$$Q = -2\, ln \left(\frac{L(1)}{L(0)} \right) \tag{9.5.1}$$

where $L(\mu)$ is defined to be the likelihood, including all systematics, for a cross section σ, where μ is defined as σ/σ_0 and σ_0 is the theoretical cross section [67]. Equation 9.5.1

is taking the ratio of the likelihood of the standard model single top cross section to the likelihood of a single top cross section of 0.

If the LLR is evaluated for a set of events with a given $\mu = \mu'$, the LLR will be distributed as a Gaussian with mean $\alpha^2(1 - 2\mu')$ and variance $4\alpha^2$ [68]. Where α is given by:

$$\alpha = \frac{\sqrt{-2\ ln\left(\frac{L(\mu)}{L(\hat{\mu})}\right)}}{|\mu - \mu'|} \tag{9.5.2}$$

where $\hat{\mu}$ is the maximum likelihood value. In the asymptotic limit, α is independent of the value of μ in equation 9.5.2, and it can be estimated in the case where $\mu' = 0$ and $\mu = \mu_{\text{expected}}$. In this case the probability density, $\rho(Q)$, is Gaussian with mean α^2 and variance $4\alpha^2$. This is the green curve shown in Figure 9.7. The calculation of the p-value can then be performed using the equation:

$$p = \int_{-\infty}^{Q_{\text{obs}}} \rho(Q)dQ = \Phi\left(\frac{Q_{\text{obs}} - \alpha^2}{2\alpha}\right) \tag{9.5.3}$$

where Φ is the cumulative distribution function of the normal distribution.

9.5.2 Measured significance

Figure 9.7 shows the expected and observed significance of the t-channel cross section measurement using a BDT. The green curve shows the probability that the background only hypothesis will yield a given log-likelihood ratio. The blue curve shows the probability that the background plus standard model signal will yield a given log-likelihood ratio. Both curves are normalized to unit area. The p-value at a specific log-likelihood point is then calculated by integrating the area under the green (background only) curve to the left of that point on the x-axis. In order to calculate a significance, one point on this axis must then be chosen. For the expected significance, the point chosen is the mode of the blue (signal plus background) curve which is represented by the dashed line. The gray band corresponds to the one sigma uncertainty on this value. In order to calculate the observed significance, the log-likelihood ratio value of the observed result is calculated. This is point is shown by the large black arrow. The expected significance is 4.1σ and the observed significance is 5.5σ. The large difference between the expected and observed significance is due to the large upwards fluctuation in the observed cross section which increased from the expected value of 2.40 pb to the observed value of 3.03 pb.

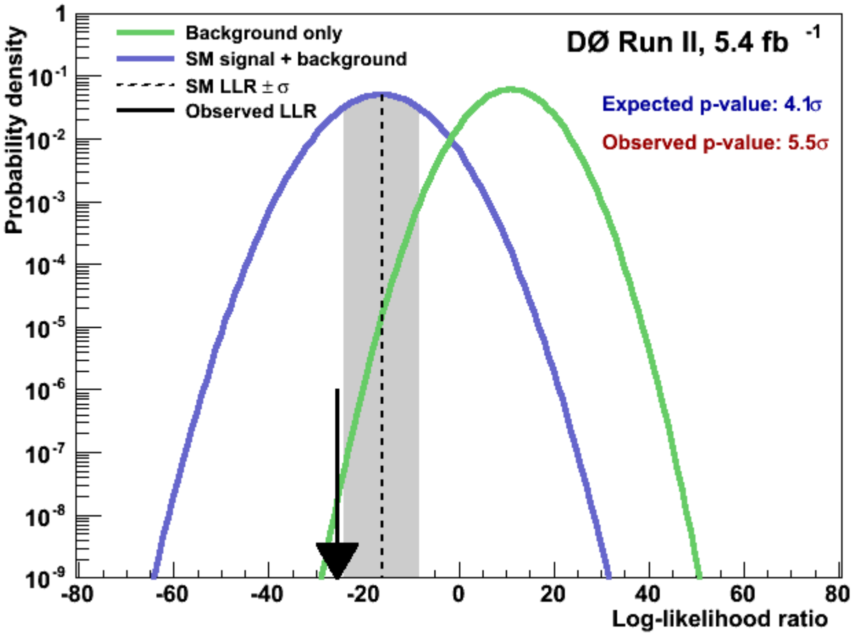


Figure 9.7 Significance of the t-channel cross section measurement with BDTs. The green curve is the probability density of the background only hypothesis, and the blue curve is the probability density of the standard model hypothesis. Both blue and green curves are normalized to unit area. The large arrow represents the observed value. The observed p-value is calculated by taking the integral of area under the green curve to the left of the arrow. The expected p-value is calculated by taking the integral of the area under the green curve to the left of the dashed line (at the mode of the blue curve).

Chapter 10. Summary

This thesis presents a measurement of the t-channel single top quark cross section. The sample contained 5.4 fb^{-1} of data and was divided into six exclusive subsamples by jet and tag multiplicity. After training six BDTs and employing Bayesian methods to calculate the cross section over all six channels, the final result is determined to be:

$$\sigma(p\bar{p} \rightarrow tqb + X) = 3.03^{+0.78}_{-0.66}\ \mathrm{pb} \tag{10.0.1}$$

This corresponds to a significance of of 5.5σ. The standard model prediction for this cross section is 2.26 ± 0.12 pb.

Several additional measurements of both the s- and t-channel cross sections were performed using this and multivariate techniques. The results were combined with this analysis to provide a final set of measurements. These results are shown in Appendix A

Appendix A. Other Analyses

A.1 s-channel

An additional cross section measurement was performed by considering s-channel single top as signal and t-channel single top as a background process. The input variable list for the BDT were re-determined and six new BDTs are trained which are used to perform the measurement of the s-channel cross section. Figure A.1 shows the posterior probability curve for this measurement. The final cross section is measured to be $0.68^{+0.41}_{-0.39}$ pb. This compares to the standard model prediction of 1.04 ± 0.04 pb.

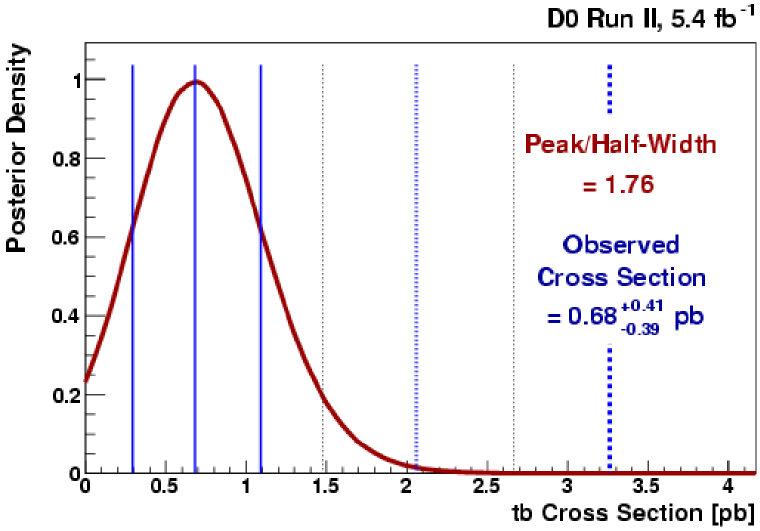


Figure A.1 Posterior probability density for the s-channel, calculated using BDTs.

A.2 $s + t$-channel

Another measurement can be performed by considering both s-channel and t-channel single top as a signals. The s- and t-channel cross sections are assumed to have the ratio predicted by the standard model, however the overall normalization is calculated from the data. The input variables for the BDT are again recalculated for this channel, a new set of BDTs are trained, and a new cross section measurement is performed. Figure A.2 shows the posterior probability curve for this measurement. The final cross section is measured to be $3.01^{+0.80}_{-0.75}$ pb. This compares to the standard model prediction of 3.30 ± 0.13 pb. While this result is smaller than the t-channel result alone, the two measurements are still statistically compatible owing to the large uncertainties. This can occur due to different event portions of phase space being utilized for the t-channel and $s + t$-channel discriminators.

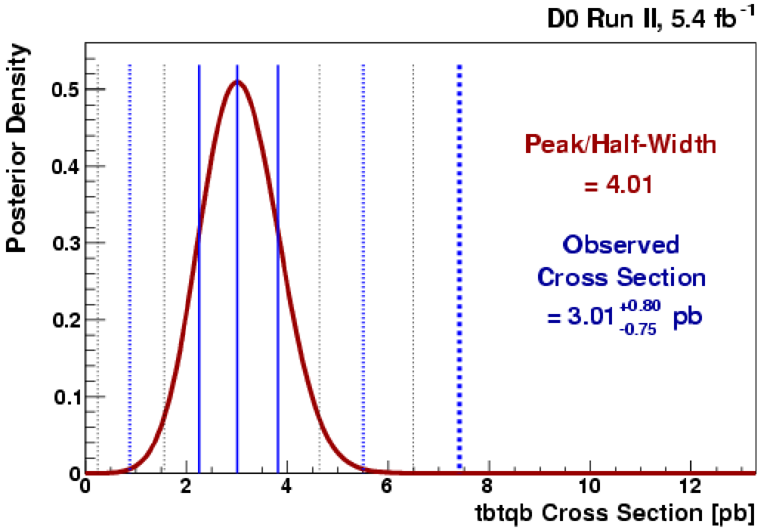


Figure A.2 Posterior probability density for $s + t$-channel using BDTs.

A.3 Combination analysis

In order to increase the significance of the cross section measurement, better separation of signal and background samples is needed. To this end, two more multivariate methods were tried: Bayesian Neural Networks (BNN), and NeuroEvolution of Augmenting Topologies (NEAT). The BNN method forms a neural network with 20 hidden nodes and is described in detail in [69]. NEAT employs a genetic algorithm which is used to design the topology of a neural network [70].

Both of these methods perform very slightly worse than the BDTs in most cases; however, if the methods are not entirely correlated with each other, additional performance can be gained by combining the information from all three methods. There are two approaches to looking into the correlations between methods.

In the first approach, the transformed output for each of the three methods for a given event (DT output, BNN output, NEAT output) can be plotted against each other. Figure A.3 shows three scatter plots formed by taking all combination of pairs between the three values. As all three methods use the binning transformation, they all have very similar shapes for the background distributions. This allows the direct comparison of the outputs, as the distributions are background dominated.

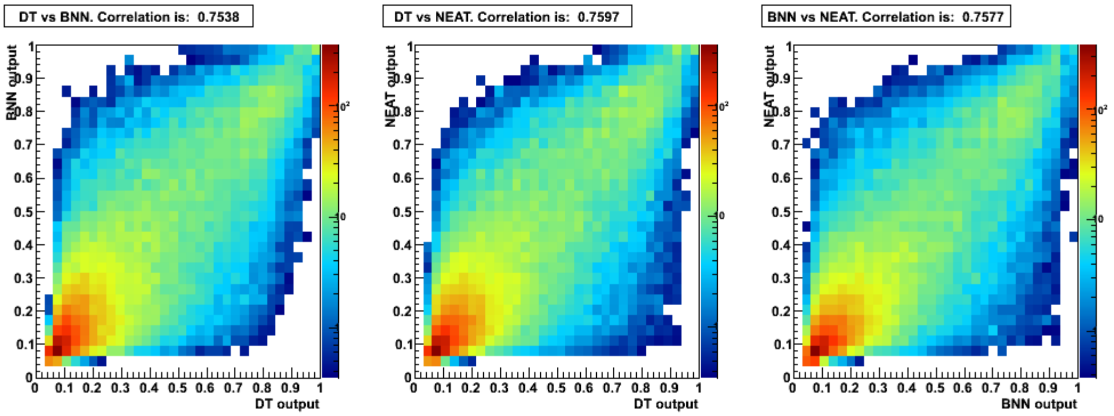


Figure A.3 Correlations between MVA outputs for each combination of methods: BDT vs BNN (left), BDT vs NEAT (center), BNN vs NEAT (right).

The second method of measuring the correlation between methods is to create 5,000 subsets of the MC, known as ensembles. Each ensemble is chosen by sampling events in such a way as to include correlations between systematic uncertainties and correlations between methods. The full procedure for this can be found in [65]. Once each ensemble is generated, the cross section, with full systematic uncertainties, is calculated with each of the three methods. Scatter plots showing the correlation between cross section measurements are shown in Figure A.4.

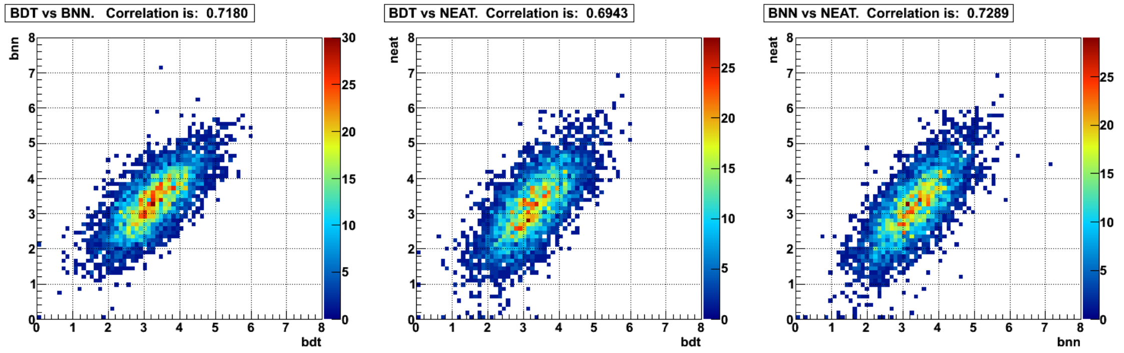


Figure A.4 Correlations between cross sections measured in each ensemble data set: BDT vs BNN (left), BDT vs NEAT (center), BNN vs NEAT (right).

As both approaches show, while there is correlation between the methods, it is not a perfect correlation. It is expected that performance can be gained by using all of the information from the three MVAs together.

The three methods are combined by using the BDT, BNN, and NEAT outputs as the input variables for a very simple Bayesian neural network [62]. This combination BNN is trained on a dataset independent from both the MC used to train the individual methods and the MC used to make the cross section measurement. The performance gain in the combination BNN can be seen in figure A.5 by comparing the signal efficiency vs background rejection for each of the individual methods and the combination method.

Four different combination BNNs are constructed. The first uses the BNN, BDT,

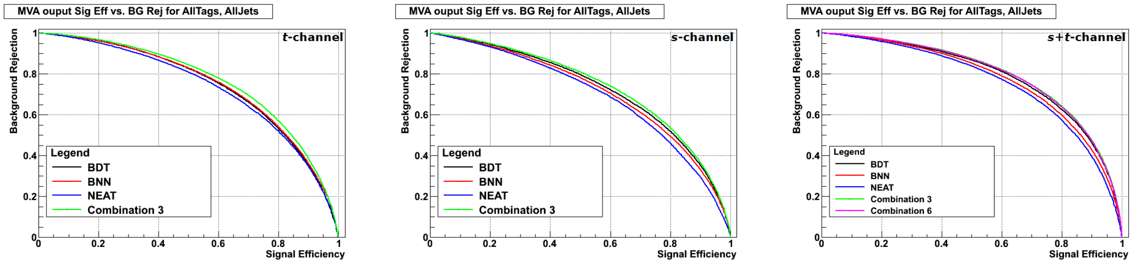


Figure A.5 Signal Efficiency vs Background Rejection for BDT, BNN, NEAT in the following channels: t-channel (left), s-channel (center), and s+t-channel (right). Better performing methods have a smaller area above the curve.

and NEAT methods trained using t-channel as signal. This is the t-channel BNN combination. An s-channel and $s + t$-channel combination are created in similar ways by using the MVAs trained on s- and $s + t$-channel respectively. The fourth combination is for $s + t$-channel and known as BNN combination-6 as it uses 6 variables: BDT, BNN, and NEAT trained on t channel and BDT, BNN, and NEAT trained on s channel. As figure A.5 shows, BNN combination-6 slightly outperforms the 3 variable $s + t$ combination method.

The posterior probability density curves for each of the four combinations are shown in figure A.6 and table A.1 shows the cross section measurement for each of the four methods.

Method	Cross Section (pb)
t-channel BNN combination	$2.86^{+0.69}_{-0.63}$
s-channel BNN combination	$0.68^{+0.38}_{-0.35}$
$s + t$-channel BNN combination	$3.02^{+0.68}_{-0.68}$
$s + t$-channel BNN combination-6	$3.43^{+0.73}_{-0.74}$

Table A.1 Observed cross section measurements for s-, t-, and $s + t$-channels. Errors shown are with systematic plus statistical.

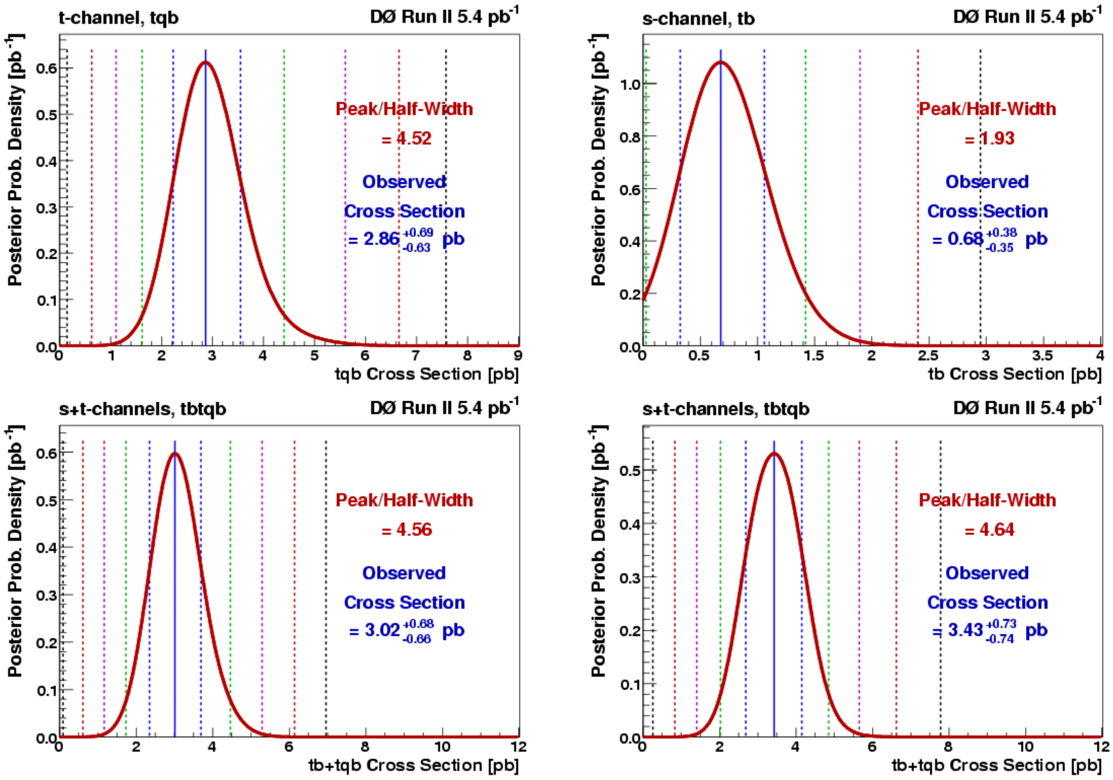


Figure A.6 Observed posterior probability density curves for the BNN combinations: t-channel (top left), s-channel (top right), $s+t$-channel 3 input combination (bottom left), $s+t$-channel combination-6 (bottom right)

A.4 Two dimensional measurements

The standard method of measuring the s- or t-channel cross section, treats the other process as a background occurring with standard model cross section. A second approach can be used in which a two dimensional posterior is constructed over the s- and t-channel cross sections simultaneously. In this method, no theoretical cross section is assumed for either single top channel. The 2D posterior created using this approach for the BNNComb6 discriminator is shown in figure A.7.

This two dimensional posterior probability density can then be integrated over in one dimension in order to create a standard 1D posterior density curve in the other dimension. Then the 2D plot can be integrated in the other dimension to create the remaining 1D posterior probability density. In this way, both the s- and t-channel cross sections can be measured simultaneously without assuming the standard model ratio of their cross sections. Figure A.7 shows this 2D posterior probability density created by using BNN combination-6 along with the two 1D posterior probability densities derived from it.

Using this method, the t-channel cross section is measured to be $2.32^{+0.68}_{-0.62}$ pb and the s-channel cross section is measured to be $1.05^{+0.48}_{-0.45}$ pb. The most significant measurement for the t-channel cross section comes from applying the 2D measurement method to the BNN combination discriminator which was trained on t-channel events. The cross section measurement made using this method is $2.90^{+0.59}_{-0.59}$ pb.

A.5 Combined significance

The asymptotic approximation of the log likelihood ratio is used to calculate the expected and observed significance for the BNN combination cross section measurements. Results denoted as "1D" are calculated using the standard one dimensional posterior probability density shown in sections A.1-A.3. Results denoted as "2D" are calculated by using a two dimensional posterior probability density as explained in section A.4.

The data shows that the $s + t$-channel cross section combined is very close to the expected results. The s-channel fluctuated downwards while the t-channel fluctuated upwards, both by about one standard deviation. The fact that the measured s channel cross section was smaller than expected, results in a slightly larger significance for t-channel when the s-channel cross section is not assumed to be the standard model value. However, the t-channel BNN combination method does find a greater than 5σ significance for t-channel the 2D case and a significance of exactly 5σ in the 1D case. Figure A.8 shows plots of the significance calculation for t-channel in the 1D and 2D methods. More details on these measurements can be found in [62]. This represents the first 5σ measurement of the t-channel single top cross section.

Method	Expected significance	Observed significance	Observed cross section (pb)
t-channel BNN combination-t (2D)	4.6σ	5.5σ	$2.90^{+0.59}_{-0.59}$
t-channel BNN combination-6 (2D)	4.4σ	4.1σ	$2.32^{+0.68}_{-0.62}$
t-channel BNN combination-t (1D)	4.3σ	5.0σ	$2.86^{+0.69}_{-0.63}$
s-channel BNN combination-s (1D)	2.7σ	1.8σ	$0.68^{+0.38}_{-0.35}$
s-channel BNN combination-6 (2D)	2.1σ	2.4σ	$1.05^{+0.48}_{-0.45}$
$s+t$-channel BNN combination-6 (1D)	5.5σ	5.4σ	$3.43^{+0.73}_{-0.74}$

Table A.2 Expected and observed significances made with the various BNN combination methods along with corresponding cross section measurements. Results with the best expected significances are in bold.

A.6 Summary

While all of the cross section measurements presented in the previous sections are slightly different, they are consistent within the statistical plus systematic uncertainties. In order to avoid biasing the result, the final choice of cross section measurement method was determined for each channel by taking the result with the largest expected significance. This is the result that should have the most sensitivity independent of statistical fluctuations in the data. The final cross section measurements used are the 2D t-channel measurement made with the BNN combination trained with t-channel as signal, the 1D s-channel measurement made with the BNN combination trained with s-channel as signal, and the 1D $s+t$-channel measurement made with the BNN combination-6 discriminator. The results are:

$$\sigma(p\bar{p} \to tqb + X) = 2.90\ ^{+0.59}_{-0.59}\ \text{pb}$$

$$\sigma(p\bar{p} \to tb + X) = 0.68\ ^{+0.38}_{-0.35}\ \text{pb}$$

$$\sigma(p\bar{p} \to tqb + tb + X) = 3.43\ ^{+0.73}_{-0.74}\ \text{pb}.$$

This represents the first 5σ measurement of t-channel single top, and the most accurate measurements of both s-channel and $s+t$-channel single top.

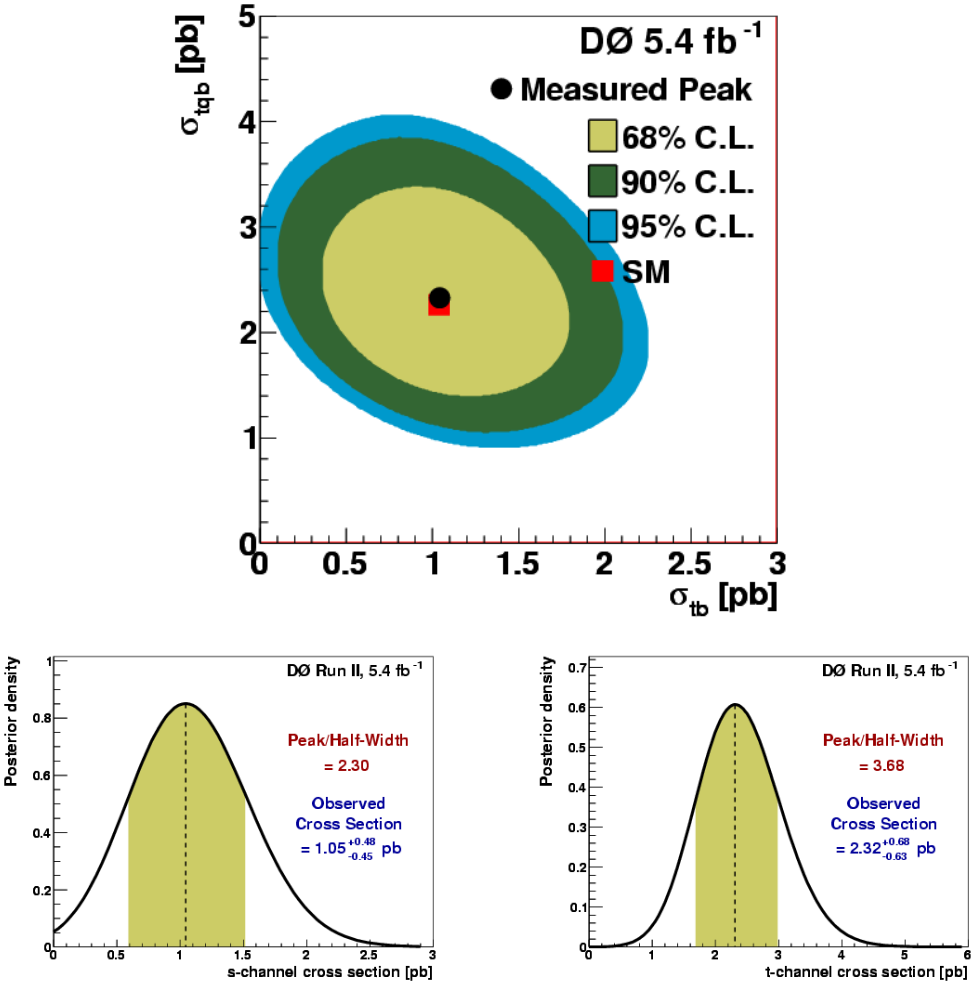


Figure A.7 Two dimensional posterior probability density calculated with BNN combination-6 (top). One dimensional s-channel posterior probability density calculated by integrating over t-channel (left), and one dimensional t-channel posterior probability density calculated by integrating over s-channel (right).

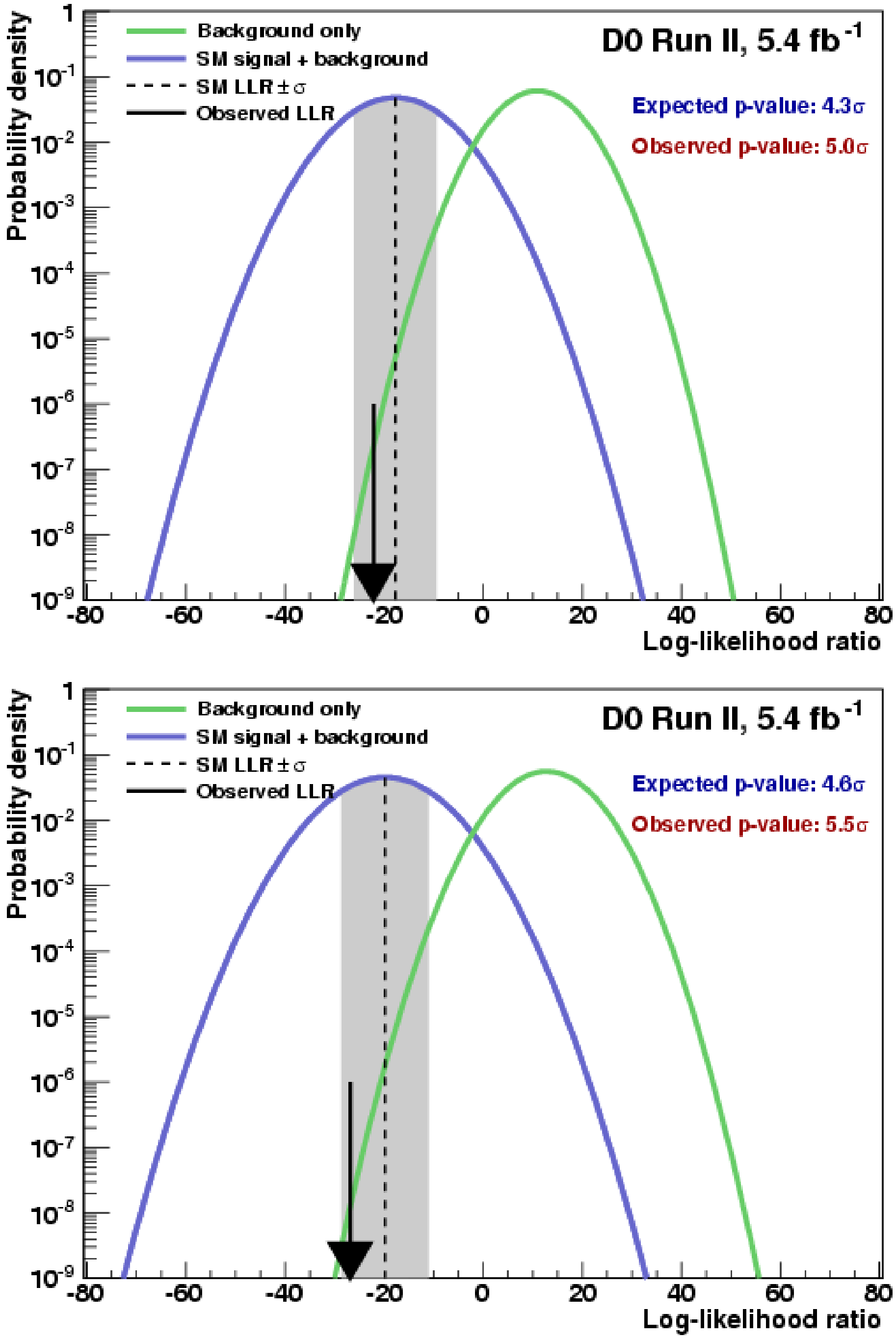


Figure A.8 Significance of the t-channel single top quark cross section measurement using the BNNcomb discriminator. The 1D method (top) yields a significance of 5.0σ, and the 2D method (bottom) yields a 5.5σ significance.

Appendix B. Systematic Uncertainty Tables

The systematic uncertainties presented in chapter 7 are given as ranges across all channels and samples. For the actual cross section measurements, the systematics are considered individually for each of these. The following tables provide the full list of all uncertainties used in this analysis.

A minus sign indicates that the uncertainty is treated as anti-correlated with other systematics with the same name. Specifically, the QCD and W+Jets IKS systematics are treated as anti-correlated with each other and the ISF/FSR systematics are treated as anti-correlated across the different jet multiplicities. These anti-correlations result from additional constraints placed on these systematics which require the total number of events to be conserved over the entire analysis. Due to these constraints, if a systematic sample removes events from one channel or sample, those events are added to the other to preserve normalization. An upward fluctuation in one sample will, therefore, be anti-correlated with the other sample which will see a downwards fluctuation.

UNCERTAINTIES FOR CHANNELS WITH TWO JETS

SINGLE TAG

						Percentage Errors							
	$t\bar{t}ll$	$t\bar{t}lj$	Wbb	Wcc	Wlp	Zbb	Zcc	Zlp	dibosons	multijet	tb	tqb	$tb + tqb$
Luminosity	6.1	6.1	—	—	—	6.1	6.1	6.1	6.1	—	6.1	6.1	6.1
Xsect.	9.0	9.0	—	—	—	3.3	3.3	3.3	7.0	—	3.8	5.3	4.8
Branching frac.	1.5	1.5	—	—	—	—	—	—	—	—	1.5	1.5	1.5
PDF	—	—	—	—	—	—	—	—	—	—	2.0	2.0	2.0
Triggers	5.0	5.0	—	—	—	5.0	5.0	5.0	5.0	—	5.0	5.0	5.0
Lumi. rewtg.	1.0	1.0	—	—	—	1.0	1.0	1.0	1.0	—	1.0	1.0	1.0
Prim. vertex	1.4	1.4	—	—	—	1.4	1.4	1.4	1.4	—	1.4	1.4	1.4
Color Reconnection	1.0	1.0	—	—	—	—	—	—	—	—	1.0	1.0	1.0
b/light Jet Response	−0.4	−0.4	—	—	—	1.0	—	—	—	—	0.5	0.5	0.5
Lepton ID	2.9	2.9	—	—	—	2.4	2.4	2.6	2.7	—	2.8	2.8	2.8
Jet frag. and higher order	1.6	1.6	—	—	—	—	—	—	—	—	0.7	0.7	0.7
ISR/FSR	4.3	4.3	—	—	—	—	—	—	—	—	0.8	0.8	0.8
b-jet frag.	2.0	2.0	—	—	—	2.0	—	—	—	—	2.0	2.0	2.0
Taggability	7.1	5.6	5.6	5.1	6.2	6.2	6.4	5.8	5.7	—	6.6	5.9	6.1
λ_{HF}^{W}	—	—	12.0	12.0	—	—	—	—	—	—	—	—	—
λ_{HF}^{Z}	—	—	—	—	—	12.0	12.0	—	—	—	—	—	—
IKS	—	—	1.8	1.8	1.8	—	—	—	—	−34.6	—	—	—

Table B.1 Uncertainties requiring exactly one tag and two jets.

DOUBLE TAG

						Percentage Errors							
	$t\bar{t}ll$	$t\bar{t}lj$	Wbb	Wcc	Wlp	Zbb	Zcc	Zlp	dibosons	multijet	tb	tqb	$tb + tqb$
Luminosity	6.1	6.1	—	—	—	6.1	6.1	6.1	6.1	—	6.1	6.1	6.1
Xsect.	9.0	9.0	—	—	—	3.3	3.3	3.3	7.0	—	3.8	5.3	4.8
Branching frac.	1.5	1.5	—	—	—	—	—	—	—	—	1.5	1.5	1.5
PDF	—	—	—	—	—	—	—	—	—	—	2.0	2.0	2.0
Triggers	5.0	5.0	—	—	—	5.0	5.0	5.0	5.0	—	5.0	5.0	5.0
Lumi. rewtg.	1.0	1.0	—	—	—	1.0	1.0	1.0	1.0	—	1.0	1.0	1.0
Prim. vertex	1.4	1.4	—	—	—	1.4	1.4	1.4	1.4	—	1.4	1.4	1.4
Color Reconnection	1.0	1.0	—	—	—	—	—	—	—	—	1.0	1.0	1.0
b/light Jet Response	−0.4	−0.4	—	—	—	1.0	—	—	—	—	0.5	0.5	0.5
Lepton ID	2.9	2.8	—	—	—	2.3	2.4	3.2	2.7	—	2.8	2.8	2.8
Jet frag. and higher order	1.6	1.6	—	—	—	—	—	—	—	—	0.7	0.7	0.7
ISR/FSR	4.3	4.3	—	—	—	—	—	—	—	—	0.8	0.8	0.8
b-jet frag.	2.0	2.0	—	—	—	2.0	—	—	—	—	2.0	2.0	2.0
Taggability	3.5	3.2	3.8	3.7	4.4	3.8	3.4	3.6	3.4	—	3.5	3.9	3.6
λ_{HF}^{W}	—	—	12.0	12.0	—	—	—	—	—	—	—	—	—
λ_{HF}^{Z}	—	—	—	—	—	12.0	12.0	—	—	—	—	—	—
IKS	—	—	1.8	1.8	1.8	—	—	—	—	−34.9	—	—	—

Table B.2 Uncertainties requiring exactly two tags and two jets.

UNCERTAINTIES FOR CHANNELS WITH THREE JETS

SINGLE TAG

							Percentage Errors						
	$t\bar{t}ll$	$t\bar{t}lj$	Wbb	Wcc	Wlp	Zbb	Zcc	Zlp	dibosons	multijet	tb	tqb	$tb+tqb$
Luminosity	6.1	6.1	—	—	—	6.1	6.1	6.1	6.1	—	6.1	6.1	6.1
Xsect.	9.0	9.0	—	—	—	3.3	3.3	3.3	7.0	—	3.8	5.3	4.8
Branching frac.	1.5	1.5	—	—	—	—	—	—	—	—	1.5	1.5	1.5
PDF	—	—	—	—	—	—	—	—	—	—	2.0	2.0	2.0
Triggers	5.0	5.0	—	—	—	5.0	5.0	5.0	5.0	—	5.0	5.0	5.0
Lumi. rewtg.	1.0	1.0	—	—	—	1.0	1.0	1.0	1.0	—	1.0	1.0	1.0
Prim. vertex	1.4	1.4	—	—	—	1.4	1.4	1.4	1.4	—	1.4	1.4	1.4
Color Reconnection	1.0	1.0	—	—	—	—	—	—	—	—	1.0	1.0	1.0
b/light Jet Response	0.0	0.0	—	—	—	0.6	—	—	—	—	0.8	0.8	0.8
Lepton ID	2.8	2.8	—	—	—	2.5	2.5	2.7	2.7	—	2.7	2.8	2.8
Jet frag. and higher order	1.7	1.7	—	—	—	—	—	—	—	—	3.7	3.7	3.7
ISR/FSR	2.2	2.2	—	—	—	—	—	—	—	—	−5.9	−5.9	−5.9
b-jet frag.	2.0	2.0	—	—	—	2.0	—	—	—	—	2.0	2.0	2.0
Taggability	13.3	8.6	7.9	8.0	8.8	9.4	7.7	14.8	8.7	—	9.3	8.6	8.9
$\lambda_{\mathrm{HF}}^{W}$	—	—	12.0	12.0	—	—	—	—	—	—	—	—	—
$\lambda_{\mathrm{HF}}^{Z}$	—	—	—	—	—	12.0	12.0	—	—	—	—	—	—
IKS	—	—	1.8	1.8	1.8	—	—	—	—	−36.2	—	—	—

Table B.3 Uncertainties requiring exactly one tag and three jets.

DOUBLE TAG

							Percentage Errors						
	$t\bar{t}ll$	$t\bar{t}lj$	Wbb	Wcc	Wlp	Zbb	Zcc	Zlp	dibosons	multijet	tb	tqb	$tb+tqb$
Luminosity	6.1	6.1	—	—	—	6.1	6.1	6.1	6.1	—	6.1	6.1	6.1
Xsect.	9.0	9.0	—	—	—	3.3	3.3	3.3	7.0	—	3.8	5.3	4.8
Branching frac.	1.5	1.5	—	—	—	—	—	—	—	—	1.5	1.5	1.5
PDF	—	—	—	—	—	—	—	—	—	—	2.0	2.0	2.0
Triggers	5.0	5.0	—	—	—	5.0	5.0	5.0	5.0	—	5.0	5.0	5.0
Lumi. rewtg.	1.0	1.0	—	—	—	1.0	1.0	1.0	1.0	—	1.0	1.0	1.0
Prim. vertex	1.4	1.4	—	—	—	1.4	1.4	1.4	1.4	—	1.4	1.4	1.4
Color Reconnection	1.0	1.0	—	—	—	—	—	—	—	—	1.0	1.0	1.0
b/light Jet Response	0.0	0.0	—	—	—	0.6	—	—	—	—	0.8	0.8	0.8
Lepton ID	2.8	2.8	—	—	—	2.5	2.4	2.2	2.7	—	2.8	2.8	2.8
Jet frag. and higher order	1.7	1.7	—	—	—	—	—	—	—	—	3.7	3.7	3.7
ISR/FSR	2.2	2.2	—	—	—	—	—	—	—	—	−5.9	−5.9	−5.9
b-jet frag.	2.0	2.0	—	—	—	2.0	—	—	—	—	2.0	2.0	2.0
Taggability	9.3	5.9	6.8	6.6	6.7	6.9	7.2	6.3	8.4	—	6.3	6.4	6.4
$\lambda_{\mathrm{HF}}^{W}$	—	—	12.0	12.0	—	—	—	—	—	—	—	—	—
$\lambda_{\mathrm{HF}}^{Z}$	—	—	—	—	—	12.0	12.0	—	—	—	—	—	—
IKS	—	—	1.8	1.8	1.8	—	—	—	—	−36.0	—	—	—

Table B.4 Uncertainties requiring exactly two tags and three jets.

UNCERTAINTIES FOR CHANNELS WITH FOUR JETS

SINGLE TAG

								Percentage Errors					
	$t\bar{t}ll$	$t\bar{t}lj$	Wbb	Wcc	Wlp	Zbb	Zcc	Zlp	dibosons	multijet	tb	tqb	$tb+tqb$
Luminosity	6.1	6.1	—	—	—	6.1	6.1	6.1	6.1	—	6.1	6.1	6.1
Xsect.	9.0	9.0	—	—	—	3.3	3.3	3.3	7.0	—	3.8	5.3	4.8
Branching frac.	1.5	1.5	—	—	—	—	—	—	—	—	1.5	1.5	1.5
PDF	—	—	—	—	—	—	—	—	—	—	2.0	2.0	2.0
Triggers	5.0	5.0	—	—	—	5.0	5.0	5.0	5.0	—	5.0	5.0	5.0
Lumi. rewtg.	1.0	1.0	—	—	—	1.0	1.0	1.0	1.0	—	1.0	1.0	1.0
Prim. vertex	1.4	1.4	—	—	—	1.4	1.4	1.4	1.4	—	1.4	1.4	1.4
Color Reconnection	1.0	1.0	—	—	—	—	—	—	—	—	1.0	1.0	1.0
b/light Jet Response	0.3	0.3	—	—	—	0.8	—	—	—	—	0.9	0.9	0.9
Lepton ID	2.8	2.8	—	—	—	2.6	2.5	2.4	2.7	—	2.7	2.8	2.8
Jet frag. and higher order	-7.0	-7.0	—	—	—	—	—	—	—	—	4.7	4.7	4.7
ISR/FSR	0.8	0.8	—	—	—	—	—	—	—	—	-10.9	-10.9	-10.9
b-jet frag.	2.0	2.0	—	—	—	2.0	—	—	—	—	2.0	2.0	2.0
Taggability	15.9	11.8	14.9	9.2	14.4	10.6	13.4	10.7	9.5	—	12.6	12.1	12.2
$\lambda^{W}_{\mathrm{HF}}$	—	—	12.0	12.0	—	—	—	—	—	—	—	—	—
$\lambda^{Z}_{\mathrm{HF}}$	—	—	—	—	—	12.0	12.0	—	—	—	—	—	—
IKS	—	—	1.8	1.8	1.8	—	—	—	—	-33.8	—	—	—

Table B.5 Uncertainties requiring exactly one tag and four jets.

DOUBLE TAG

								Percentage Errors					
	$t\bar{t}ll$	$t\bar{t}lj$	Wbb	Wcc	Wlp	Zbb	Zcc	Zlp	dibosons	multijet	tb	tqb	$tb+tqb$
Luminosity	6.1	6.1	—	—	—	6.1	6.1	6.1	6.1	—	6.1	6.1	6.1
Xsect.	9.0	9.0	—	—	—	3.3	3.3	3.3	7.0	—	3.8	5.3	4.8
Branching frac.	1.5	1.5	—	—	—	—	—	—	—	—	1.5	1.5	1.5
PDF	—	—	—	—	—	—	—	—	—	—	2.0	2.0	2.0
Triggers	5.0	5.0	—	—	—	5.0	5.0	5.0	5.0	—	5.0	5.0	5.0
Lumi. rewtg.	1.0	1.0	—	—	—	1.0	1.0	1.0	1.0	—	1.0	1.0	1.0
Prim. vertex	1.4	1.4	—	—	—	1.4	1.4	1.4	1.4	—	1.4	1.4	1.4
Color Reconnection	1.0	1.0	—	—	—	—	—	—	—	—	1.0	1.0	1.0
b/light Jet Response	0.3	0.3	—	—	—	0.8	—	—	—	—	0.9	0.9	0.9
Lepton ID	2.8	2.8	—	—	—	2.6	2.6	2.4	2.7	—	2.8	2.8	2.8
Jet frag. and higher order	-7.0	-7.0	—	—	—	—	—	—	—	—	4.7	4.7	4.7
ISR/FSR	0.8	0.8	—	—	—	—	—	—	—	—	-10.9	-10.9	-10.9
b-jet frag.	2.0	2.0	—	—	—	2.0	—	—	—	—	2.0	2.0	2.0
Taggability	11.7	8.2	9.8	6.4	8.8	9.7	18.4	21.1	7.7	—	8.6	14.1	11.7
$\lambda^{W}_{\mathrm{HF}}$	—	—	12.0	12.0	—	—	—	—	—	—	—	—	—
$\lambda^{Z}_{\mathrm{HF}}$	—	—	—	—	—	12.0	12.0	—	—	—	—	—	—
IKS	—	—	1.8	1.8	1.8	—	—	—	—	-34.3	—	—	—

Table B.6 Uncertainties requiring exactly two tags and four jets.

Appendix C. Top Mass Dependence

The mass of the top quark is only known experimentally to about 2 GeV, and this analysis assumes a top quark mass of 172.5 GeV. To account for the uncertainty in the top mass, the analysis is repeated with samples containing top quarks of mass 170.0 and 175.0 GeV respectively. These samples are the "top mass down" and "top mass up" samples. An additional sample is created containing a mixture of events from all three samples and is referred to as the "top mass merged sample".

The Boosted Decision Trees in chapter 8 were trained on the top mass merged sample in order to minimize the top mass dependence of the cross section measurement. The cross section was then measured using the nominal top mass samples as described in chapter 9. In addition the cross section measurement was performed on the top mass up and top mass down samples to measure the effect of the top mass uncertainty on the final cross section.

The main difference between the top mass samples is the change in the theoretical SM cross sections for both single top and $t\bar{t}$ samples. The SM cross section predictions for s and t channel are shown for each mass point in table C.1 along with the observed values measured using the BNN combination-6 2D measurements. The s+t channel measurements were made using the BNN combination 1D method.

Figure C.1 shows the expected and measured cross sections with each top mass sample.

Channel	170.0 GeV	172.5 GeV	175.0 GeV
s channel (theory)	$1.12^{+0.04}_{-0.04}$	$1.04^{+0.04}_{-0.04}$	$0.98^{+0.04}_{-0.04}$
s channel (observed)	$1.36^{+0.65}_{-0.59}$	$1.05^{+0.48}_{-0.45}$	$0.73^{+0.44}_{-0.41}$
t channel (theory)	$2.34^{+0.12}_{-0.12}$	$2.26^{+0.12}_{-0.12}$	$2.16^{+0.12}_{-0.12}$
t channel (observed)	$2.27^{+0.66}_{-0.61}$	$2.32^{+0.68}_{-0.63}$	$1.91^{+0.68}_{-0.60}$
s+t channel (theory)	$3.46^{+0.13}_{-0.13}$	$3.30^{+0.13}_{-0.13}$	$3.14^{+0.13}_{-0.13}$
s+t channel (observed)	$3.70^{+0.78}_{-0.80}$	$3.43^{+0.73}_{-0.74}$	$2.56^{+0.69}_{-0.61}$

Table C.1 Measured mass dependent cross section for each channel.

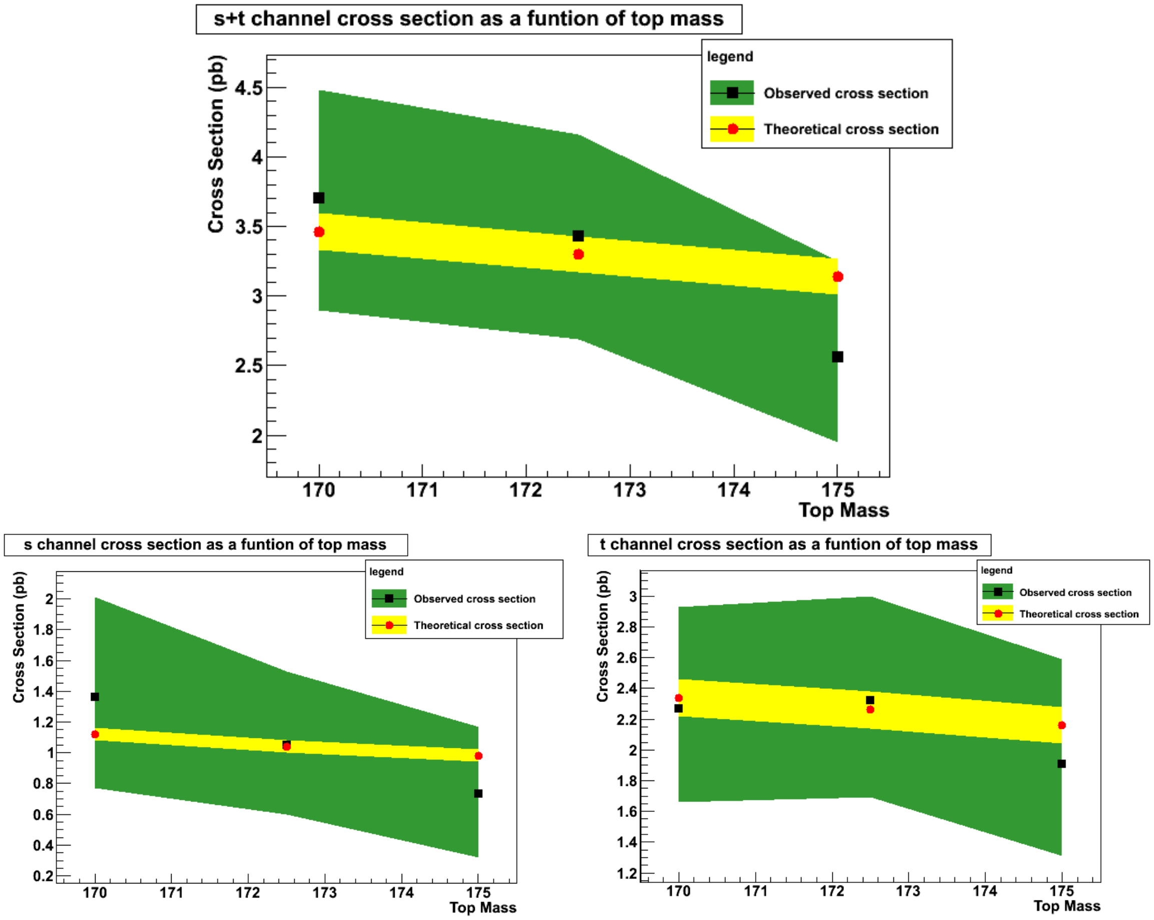


Figure C.1 Cross section dependence on top quark mass. Black points with green band are the measured values and uncertainties; red points with yellow band are theoretical values and uncertainties: s+t channel (top), s channel (bottom left), t channel (bottom right).

Bibliography

[1] K. Nakamura *et al.* (Particle Data Group), *"The Review of Particle Physics,"* J. Phys. G **37**, 075021 (2010).

[2] V.M. Abazov *et al.* (DØ Collaboration), *"Observation of Single Top Quark Production,"* Phys. Rev. Lett. **103**, 092001 (2009).

[3] J. Maxwell, *"A Dynamical Theory of the Electromagnetic Field,"* Philosophical Transactions of the Royal Society of London. Vol. 155, 459-512 (1865).

[4] S. Abachi *et al.* (DØ Collaboration), *"Observation of the Top Quark",* Phys. Rev. Lett. **74**, 2632 (1995).

[5] V. Avazov *et al.* (DØ Collaboration), *"Measurement of the top quark pair production cross section in the lepton+jets channel in proton-antiproton collisions at* $\sqrt{s}$*=1.96 TeV,"* FERMILAB-PUB-10/544-E, Submitted to Phys. Rev. D (2010).

[6] N. Kidonakis, *"Single top quark production at the Fermilab Tevatron: Threshold resummation and finite-order soft gluon corrections,"* Phys. Rev. D. **74**, 114012 (2006).

[7] J. Campbell and F. Tramontano, *"Next-to-leading order corrections to Wt production and decay,"* Nucl. Phys (2005) B726.

[8] Z. Sullivan, *"Understanding single-top-quark production and jets at hadron colliders,"* Phys. Rev. D **70**, 114012 (2004).

[9] C. Gattuso *et al.*, *"Accelerator Concepts Rookie Book,"* version 3.6 (2010).
`http://www-bdnew.fnal.gov/operations/rookie_books/rbooks/Concepts_v3.6.pdf`

[10] *"Linac Rookie Book,"* version 2.1 (2006).
`http://www-bdnew.fnal.gov/operations/rookie_books/rbooks/LINAC_v2.pdf`

[11] *"Booster Rookie Book,"* version 4.1 (2009).
`http://www-bdnew.fnal.gov/operations/rookie_books/rbooks/Booster_V4.1.pdf`

[12] *"Main Injector Rookie Book,"* version 1.1 (2003).
`http://www-bdnew.fnal.gov/operations/rookie_books/rbooks/Main_Injector_1.1.pdf`

[13] *"Antiproton Source Rookie Book,"* version 2.1 (2010).
`http://www-bdnew.fnal.gov/operations/rookie_books/rbooks/Pbar_2.1.pdf`

[14] J. Thompson, *"Introduction to Colliding Beams at Fermilab,"* FERMILAB-TM-1909 (1994).

[15] *"Recycler Rookie Book,"* version 1.42 (2010).
`http://www-bdnew.fnal.gov/operations/rookie_books/rbooks/Recycler_RB_v1.42.pdf`

[16] *"Tevatron Rookie Book,"* version 2.3 (2009).
`http://www-bdnew.fnal.gov/operations/rookie_books/rbooks/Tevatron_2.3.pdf`

[17] V.M. Abazov *et al.*, *"The upgraded DØ detector,"* Nucl. Instrum. Meth., **A565** 463-537 (2006).

[18] V.M. Abazov *et al.*, *"The muon system of the Run II DØ detector,"* Nucl. Instrum. Meth., **A552** 372-398 (2005)

[19] A. Khanov, *"HTF: histogramming method of finding tracks. The algorithm description,"* DØ Note 3778, (2000).

[20] D. Adams, *"Finding Tracks,"* DØ Note 2958, (1998).

[21] H. Greenlee, *"The DØ Kalman track fit,"* DØ Note 4303, (2004).

[22] A. Schwartzman, C. Tully, *"Primary Vertex Reconstruction by Means of Adaptive Vertex Fitting,"* DØ Note 4918, (2005).

[23] B. Olivier, U. Bassler, G. Bernardi, *"NADA: A New Event by Event Hot Cell Killer,"* DØ Note 3687, (2000).

[24] J. Vilimant *et al.*, *"Technical Description of the T42 Algorithm for Calorimeter Noise Suppression,"* DØ Note 4146, (2003).

[25] G. Blazey, *"Run II Jet Physics,"* DØ Note 3750, (2000).

[26] E. Bustato, B. Andreui, *"Jet Algorithms in the DØ Run II Software: Description and User's Guide,"* DØ Note 4457, (2004).

[27] *C. Ochando and J.-F. Grivaz, "SSR for p17,"* DØ Note 5609, (2008).

[28] A. Harel, *"Jet ID Optimization,"* DØ Note 4919, (2006).

[29] X. Bu, G. Chen, M. Takahashi, *"Electron Preselection Efficiency study for Run II data,"* DØ Note 6009, (2009).

[30] M. Narain, *"Electron identification in the DØ detector,"* Fermilab Meeting: DPF 92, Batavia, FERMILAB-CONF-93-054-E. (1992).

[31] J. Kozminski *et al.*, *"The Electron Likelihood in p14,"* DØ Note 4449, (2003).

[32] S. Bala *et al.*, *"Single Top Quark Production in 5.4 fb^{-1} of Data – Signal and Background Modeling and Event Selection,"* DØ Note 6099, (2011).

[33] O. Brandt *et al.*, *"Muon Identification Certification for the Summer 2009 Extended Dataset (Run11b-1 and -2),"* DØ Note 6025, (2010).

[34] P. Calfayn *et al.*, *"Muon Identification Certification for p17 data,"* DØ Note 5157, (2006).

[35] T. Scanlon, *"Development and optimization on MC of a Neural Network tool to identify b-jets,"* DØ Note 4889, (2005).

[36] E.E. Boos *et al.*, *"Method for Simulating Electroweak Top-Quark Production Events in the NLO Approximation: SingleTop Generator,"* Phys. Atom. Nucl. **69**, 1317 (2006). We use version 4.2p1.

[37] J. Pumplin *et al.*, *"New Generation of Parton Distributions with Uncertainties from Global QCD Analysis,"* J. High Energy Phys. **0207**, 012 (2002). Used versions CTEQ6M (signals) and CTEQ6L1 (backgrounds).

[38] T. Sjöstrand, S. Mrenna, and P. Skands, *"PYTHIA 6.4 Physics and Manual,"* J. High Energy Phys. **0608**, 026 (2006). Used version 6.409.

[39] S. Jadach *et al.*, *"The Tau Decay Library TAUOLA: version 2.4,"* Comput. Phys. Commun. **76**, 361 (1993). Used version 2.5.

[40] D.J. Lange, *"The EvtGen Particle Decay Simulation Package,"* Nucl. Instrum. Meth. **A 462**, 152 (2001). Used a DØ version similar to version 00-14-05.

[41] M.L. Mangano *et al.*, *"ALPGEN, a Generator for Hard Multiparton Processes in Hadronic Collisions,"* J. High Energy Phys. **0307**, 001 (2003). Used ALPGEN version 2.11.

[42] The Single top Working Group, *"Single Top Quark Production in 2.3 fb^{-1} of Data – Signal and Background Modeling and Event Selection,"* DØ Note 5810, (2011).

[43] H. Schellman, *"The Longitudinal Shape of the Luminous Region at DØ,"* DØ Note 5142, (2006).

[44] B.. Tiller, T. Nunneman, *"Measurement of the differential Z0-boson production cross-section as a function of transverse momentum,"* DØ Note 4660, (2004).

[45] J. Hays *et al.*, *"Single Electron Efficiencies in p17 Data and Monte-Carlo using p18.05.00 d0correct,"* DØ Note 5105, (2006).

[46] O. Atramentov *et al.*, *"Electron and Photon Identification with p20 data,"* DØ Note 5761, (2008). NOTE: Used the "re-certified EMID scale factors for p17, Winter and Summer 2009 datasets" provided by the EMID group. `https://plone4.fnal.gov/P1/D0Wiki/object-id/emid/emcert/Redo_Winter2009Summer2009/description_redo_winter09summer09`

[47] T. Gadfort *et al.*, *"Muon Identification Certification for p17 Data,"* DØ Note 5157, (2006).

[48] O. Brandt, S. Ch, M. Cookc, M. Eads, D. Hedin, A. Santos, B. Tuchming, Y. Yatsunenko, S. Youn, *"Muon Identification Certification for the Summer 2009 Extended Dataset (Run IIb-1 and -2),"* DØ Note 6109, (2010).

[49] C.E. Gerber, A. Juste, G.J. Otero y Garzon, A. Tanasijczuk, *"Studies on the Instrumental Background in electron+jets+MET Final State Selections using p17 Monte Carlo,"* DØ Note 5955, (2009).

[50] E. Barberis *et. al.*, *"The Matrix Method and its Error Calculation,"* DØ Note 4564, (2004).

[51] R. Schwienhorst and E. Perez, *"Uncertainty Calculation for Combined W+jets and QCD Backgrounds in Analyses with W+Jets Signatures,"* DØ Note 4597, (2004).

[52] J. BackusMayes, *"b-Tagging Studies in $ZH \rightarrow \mu\mu bb$"* Presentation in b-ID Meeting DØ Collaboration Week 12 May, (2010).

[53] T. Gadfort, A. Hass, D. Johnston, D.Lincoln, T.Scanlon, S. Schlobohm, *"Performance of the DØ NN b-tagging Tool on p20 Data,"* DØ Note 5554, (2007).

[54] N. Kidonakis, *"Single Top Quark Production at the Fermilab Tevatron: Threshold Resummation and Finite-Order Soft Gluon Corrections,"* Phys. Rev. D **74**, 114012 (2006).

[55] S. Moch and P. Uwer, *"Theoretical status and prospects for top-quark pair production at hadron colliders,"* Phys. Rev. D **78**, 034003 (2008).

[56] F. Deliot, C. Deterre, S. Shary, *"Z/γ^* cross section computation at NNLO using the FEWZ code,"* DØ Note 6050, (2010).

[57] H. Schellman, *"Run IIb Longitudinal Beam Shape,"* DØ Note 5540, (2007).

[58] O. Atramentov, D. Bandurin, X. Bu, B. Calpas, E. Carrera, D. Duggan, A. Ferapontov, M. Takahashi, T. Uzbyakova, H. Yin, *"Electron and Photon Identification with p20 data,"* DØ Note 5761, (2008). Updated numbers were taken from `https://plone4.fnal.gov/P1/D0Wiki/object-id/emid/emcert/Moriond2009/uncertainty_moriond09`

[59] F. Deliot, A. Grohsjean, C. Schwanenberger, E. Shabalina, M. Wang, D. Wicke, Z. Ye, *"Systematic Uncertainties in Top Quark Measurements,"* DØ Note 6024, (2010).

[60] Y. Peters, M. Begel, K. Hamacher, and D. Wicke, *"Reweighting of the Fragmentation Function for the DØ Monte Carlo,"* DØ Note 5325, (2007).

[61] A. Harel, *"An estimate of the data over MC, b over light jet response uncertainty for RunIIa JES,"* DØ Note 5654, (2008).

[62] S. Bala *et al.*, *"Measurement of the Single Top Production Cross Section in 5.4 fb*$^{-1}$ *of Data,"* DØ Note 6128, (2011).

[63] Y. Freund, R.E. Schapire, *"Experiments with a New Boosting Algorithm,"* in proceedings of ICML 148-156 (1996).

[64] E. Aguilo *et al.*, *"Observation of Single Top Quark Production in 2.3 fb*$^{-1}$ *of Data Using Boosted Decision Trees,"* DØ Note 5811, (2008).

[65] S. Jain, H. Prosper, and R. Schwienhorst, *"Statistical methods Implemented in the Package* `top_statistics`*,"* DØ Note 5817, (2008).

[66] I. Bertram *et al.*, *"A Recipe for the Construction of Confidence Limits,"* FERMILAB-TM-2104 (2000).

[67] G. Jain, K. Crammer, E. Gross, O. Vitells, *"Asymptotic formulae for likelihood-based tests of new physics,"* arXiv:1007.1727v2 (2010).

[68] T. Aaltonen *et al.* (CDF Collaboration), *"First Observation of Electroweak Single Top Quark Production"* Phys. Rev. Lett. 103, 092002 (2009).

[69] R. Neal, *"Bayesian Learning of Neural Networks,"* Springer-Verlag, New York, (1996).

[70] K. Stanley, R. Miikkulainen, *"Evolving Neural Networks through Augmenting Topologies,"* Evolutionary Computation 10(2): 99-127 (2002).