UNIVERSITÄT BONN

Physikalisches Institut

Measuring the $t\bar{t}$ Production Cross Section in the Electron-Plus-Jets Channel in Proton-Antiproton Collisions at $\sqrt{s} = 1.96$ TeV with the DØ Detector at the Tevatron – A Monte Carlo Study –

von
Su-Jung Park

Abstract: The measurement of the $t\bar{t}$ production cross section at $\sqrt{s} = 1.96$ TeV using the final state with an electron and jets is studied with Monte Carlo event samples. All methods used in the real data analysis to measure efficiencies and to estimate the background contributions are examined. The studies focus on measuring the electron reconstruction efficiencies as well as on improving the electron identification and background suppression. With a generated input cross section of 7 pb the following result is obtained:

$$\sigma_{t\bar{t}} = (7 \pm 1.63 \text{ (stat)}^{+0.94}_{-1.14} \text{ (sys)}) \text{ pb.}$$

Post address:
Nußallee 12
D-53115 Bonn
Germany

BONN-IB-2004-05
Bonn University
März 2004

UNIVERSITÄT BONN

Physikalisches Institut

**Measuring the $t\bar{t}$ Production Cross Section
in the Electron-Plus-Jets Channel
in Proton-Antiproton Collisions at $\sqrt{s} = 1.96$ TeV
with the DØ Detector at the Tevatron
– A Monte Carlo Study –**

von
Su-Jung Park

Dieser Forschungsbericht wurde als Diplomarbeit von der mathematisch-naturwissenschaftlichen Fakultät der Universität Bonn angenommen.

Angenommen am:	09. Februar 2004
Referent:	Prof. Dr. N. Wermes
Korreferent:	Prof. Dr. E. Hilger

Contents

1	Introduction	1
2	Theory	3
2.1	The Standard Model	3
2.2	The Top Quark	4
2.2.1	Virtual Top Quark	4
2.2.2	Production and Decay	6
2.2.3	Sources of Background	11
3	Experimental Setup	15
3.1	The Tevatron Collider	15
3.2	The DØ Detector	16
3.2.1	Tracking System	17
3.2.2	Calorimeter System	20
3.2.3	Trigger System	22
3.3	Monte Carlo Simulation	24
4	Physics Objects	27
4.1	Electrons	27
4.1.1	Electron Reconstruction	28
4.1.2	Efficiency in Monte Carlo and Data	30
4.1.3	Sources of Background	38
4.1.4	Electron Likelihood Studies	40
4.1.5	Electron Likelihood in p14	52
4.2	Jets	55
4.2.1	Jet Reconstruction	55
4.2.2	Jet Reconstruction Efficiency	56
4.2.3	Jet Energy Scale	57
4.3	Missing Transverse Energy	57
5	Cross Section Analysis	59
5.1	Preselection	59
5.1.1	Loose Preselection	60
5.1.2	Tight Preselection	61
5.1.3	Trigger Efficiencies	61
5.1.4	Summary	64

5.2	QCD Background and Matrix Method	64
5.3	W+Jets Background and Berends Scaling	67
5.3.1	W+Njets Event Selection	67
5.3.2	Normalization	68
5.3.3	Trigger Bias	70
5.3.4	Berends Scaling	72
5.4	Topological Cuts	73
5.5	$t\bar{t}$ Cross Section Results	76
5.6	Systematics	78
5.6.1	Jet Energy Scale	78
5.6.2	Monte Carlo Modeling	79
5.6.3	Correlation of Samples	79
5.6.4	Electron Identification Efficiencies	81
5.6.5	Summary	83
6	Conclusion and Outlook	85
A	Monte Carlo Samples	87
	Bibliography	89
	Acknowledgements	93

1 Introduction

Elementary Particle Physics engages itself in the search for the fundamental building blocks of matter and their interactions with each other. In the beginning of its scientific history stands the discovery of the electron by J. J. Thomson in 1897. Gravity and the electromagnetic force were the known interactions of matter and with the discoveries of the proton and the neutron, thus finding all the constituents of atoms, a very simple model seemed to emerge. But starting in the 1930s all kinds of new particles appeared and new models involving new interactions demanded even more particles, bringing about antiparticles, neutrinos and many mesons and baryons. The most complete and successful theory to describe all observations so far is the Standard Model of Particle Physics, deriving three of the four fundamental forces from the principle of local gauge invariance, categorizing the elementary particles into gauge bosons, three families of quarks and leptons and the Higgs boson. The Standard Model has successfully made quantitative predictions for interaction rates and couplings as well as predicted the existence of, at that time, undiscovered elementary particles such as the τ neutrino.

The discovery of the top quark at Fermi National Accelerator Laboratory (Fermilab) by the CDF and DØ collaborations in 1995 [1, 2, 3] was another impressive confirmation of the Standard Model. The measured mass is in perfect agreement with the expectations from electroweak fits. Shortly after the discovery of the top quark, Run I (1992-1995) of the Fermilab Tevatron was completed and was followed by an upgrade of the accelerator and the detectors. Since the Tevatron is the highest-energy collider in the world with a center-of-mass energy of 1.8 TeV in Run I and 1.96 TeV in Run II and in fact the only collider where the production of the top quark is possible, further studies of the top quark had to wait for the start of Run II in the spring of 2001. With improved detectors and an integrated luminosity exceeding that of the entire Run I, more precise measurements on the top quark cross section, mass and spin can be made now. But further studies are not only desirable for completing the knowledge about the last of the discovered quarks of the Standard Model. The top quark being by far the heaviest of the elementary particles might be helpful in the understanding of the generation of mass and the Yukawa coupling and in reaching beyond the Standard Model.

At a center-of-mass energy of 1.96 TeV the top quark can be produced either as a pair with its antiquark partner by the strong interaction or just by itself via the weak interaction. This analysis focuses on the top quark pair production $t\bar{t}$. The top quark decays to almost 100% to a W boson and a b quark and the subsequent decay of the W boson defines the

different decay channels. The channel examined in this analysis is the electron-plus-jets ($e+jets$) channel $t\bar{t} \rightarrow W^+bW^-\bar{b} \rightarrow e^\pm\nu_e q\bar{q}b\bar{b}$, where one W boson decays to an electron and an electron neutrino and the other W boson decays to quarks that result in jets. Figure 1.1 depicts this process and shows the typical event signature, an isolated electron with a high transverse momentum p_T , large missing transverse energy $\cancel{E}_T$ due to the neutrino, and four or more jets. The correct identification of the electron will therefore play an important role in reducing the background comprised of QCD multijet and W +multijet events.

Figure 1.1: Top quark decay in the electron+jets channel.

The studies done for this thesis concentrate on the improvement of the electron identification, its efficiency determination and the measurement of the top quark pair production cross section in the controlled environment of Monte Carlo generated data. Chapter 2 gives an introduction to the Standard Model and the physics of the top quark and Chapter 3 describes the experimental environment and the process of simulating Monte Carlo events. Chapter 4 discusses the identification of the physics objects done at DØ, concentrating on the electron. The main analysis is presented in Chapter 5, documenting all steps made to separate the $t\bar{t}$ signal from the QCD and W +multijet background and to measure the $t\bar{t}$ cross section in a Monte Carlo sample. The systematic uncertainties of this analysis are discussed in detail. Chapter 6 summarizes the studies made and gives an outlook on additional possible improvements.

2 Theory

The Standard Model describes the properties of elementary particles and their interactions with one another. It predicted the existence of the top quark, which was discovered only in 1995 at the Tevatron. This chapter gives an introduction to the Standard Model and the theory of the top quark.

2.1 The Standard Model

Currently all matter is believed to consist of three kinds of elementary particles, leptons, quarks, and the force mediators. The leptons can be classified into three families, the electron and its neutrino, the muon and its neutrino, and the tau and its neutrino:

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix} \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}$$

For each lepton there is also an antilepton, so that there are twelve leptons in total.

Quarks have six flavors, and they can be arranged into three families of doublets, the down quark and up quark, the strange quark and charm quark, and the bottom quark and top quark:

$$\begin{pmatrix} u \\ d \end{pmatrix} \quad \begin{pmatrix} c \\ s \end{pmatrix} \quad \begin{pmatrix} t \\ b \end{pmatrix}$$

Each quark has an antiquark partner, and quarks additionally come in three colors, so that there are 36 quarks in total.

The Standard Model describes three of the four fundamental forces. Since gravity is much too weak in the realm of elementary particles, it is not included. Each of the forces has a mediator, the photon for the electromagnetic force, the two $W^\pm$ bosons and the Z boson for the weak force and eight gluons for the strong force. These are twelve mediators, and in addition at least one Higgs boson is expected in the Standard Model, which makes a total of 61 elementary particles. A summary of the fundamental forces, their mediators and the theory describing them is shown in Table 2.1. A more extensive discussion of the Standard Model can be found in [4] and [5].

Force	Mediator	Theory
Strong	8 gluons	Chromodynamics
Electromagnetic	photon	Electrodynamics
Weak	$W^\pm$ and Z	Flavordynamics
Gravitation	graviton	

Table 2.1: The four fundamental forces.

2.2 The Top Quark

There are several reasons why it is especially interesting to study the top quark more closely. Because of its extremely high mass it is speculated that the top quark might play a special role in nature, but even without exotic properties it is worthwhile to be studied. Since the top quark is the most recently observed quark, its properties such as spin, isospin, mass, charge, its couplings and its production mechanism are not very well known yet. Besides from being interesting by itself, the top quark also has implications on other analyses. The top quark mass for example is needed for electroweak precision analyses and the Higgs boson couples most strongly to the top quark. Furthermore top quark events constitute the background to new physics events, so they must be better understood to be able to discover new physics.

2.2.1 Virtual Top Quark

In the electroweak theory all quantities at tree level can be calculated by knowing only three parameters [6]. One possible set is the fine-structure constant α , the Fermi coupling constant G_F and the mass of the Z boson M_Z , which are all known very precisely [7]:

$$\begin{aligned}\alpha &= \frac{1}{137.03599976(50)} \\ G_F &= 1.16639(1) \times 10^{-5} \text{ GeV}^{-2} \\ M_Z &= 91.1876(21) \text{ GeV}\end{aligned}$$

The uncertainty is given in parentheses.

So for example the W boson mass at tree level is given by:

$$\begin{aligned}M_W^2 &= \frac{1}{2}M_Z^2 \left(1 + \sqrt{1 - \frac{4\pi\alpha}{\sqrt{2}G_F M_Z^2}} \right) \\ &= \frac{\frac{\pi\alpha}{\sqrt{2}G_F}}{s_W^2} \quad \text{with} \quad s_W^2 := 1 - \frac{M_W^2}{M_Z^2}.\end{aligned}\tag{2.1}$$

Taking into account one-loop diagrams this changes to

$$M_W^2 = \frac{\frac{\pi\alpha}{\sqrt{2}G_F}}{s_W^2(1 - \Delta r)}, \quad (2.2)$$

where Δr depends on the type of loop. Loops involving the top quark like those shown in Figure 2.1 contribute to the W and Z masses in the following way:

$$(\Delta r)_{top} \approx -\frac{3G_F m_t^2}{8\sqrt{2}\pi^2} \frac{1}{t_W^2} \quad \text{with} \quad t_W^2 := \tan^2 \theta_W. \quad (2.3)$$

It can be seen that this correction depends quadratically on the top quark mass. Similar one-loop diagrams involving the Higgs boson are shown in Figure 2.2 and contribute as well:

$$(\Delta r)_{Higgs} \approx \frac{11G_F M_Z^2 c_W^2}{24\sqrt{2}\pi^2} \ln \frac{m_h^2}{M_Z^2} \quad \text{with} \quad c_W^2 := \cos^2 \theta_W. \quad (2.4)$$

But this correction only depends logarithmically on the Higgs boson mass. Thus it can be concluded that the one-loop correction to the W boson mass is much more sensitive to the top quark mass.

Figure 2.1: Virtual top-quark loops.

Figure 2.2: Virtual Higgs-boson loops.

In addition to correcting the prediction of the W boson mass with the formulae given, a precision measurement of the top quark mass and the W boson mass can be used to predict the Higgs boson mass. Figure 2.3 shows the W boson mass versus the top quark mass with corresponding Higgs boson masses. Direct and indirect measurements of the W boson mass and the top quark mass are in good agreement, implying that a light Higgs boson is favored by the electroweak data.

Similarly the top quark mass itself can be determined by fits of electroweak data assuming different Higgs boson masses. This was done even before the top quark was discovered.

Figure 2.3: m_W vs. m_t with lines of constant Higgs mass. The dashed ellipse is the 68% confidence level outline of direct measurements ($p\bar{p}$ colliders and LEP2), the solid line is the 68% confidence level outline of fits to electroweak precision measurements (LEP1 + SLD data) [8].

The χ^2 of the fits indicating the favored top quark mass to be around 120-170 GeV/ c^2 , can be seen in Figure 2.4 and the evolution over the years from 1989 to 2003 including the direct measurements can be followed in Figure 2.5.

2.2.2 Production and Decay

At the Tevatron the top quark is produced either in pair production via the strong interaction or as single top via the charged-current weak interaction. It decays predominantly to a W boson and a b quark.

Top Quark Production via the Strong Interaction

The primary source of top quarks at the Tevatron is the top quark pair production. Its subprocesses quark-antiquark annihilation and gluon fusion are shown in lowest order in Figure 2.6. At a center-of-mass energy $\sqrt{s} = 1.96$ TeV the quark-antiquark annihilation dominates with 85% to 15% gluon fusion, whereas at a center-of-mass energy of $\sqrt{s} = 14$ TeV, which will be reached at the LHC, gluon fusion dominates with 90% to only 10% contribution of the quark-antiquark annihilation. This dependency on the center-of-mass

Figure 2.4: χ^2 vs. m_t from electroweak fits. Left: fits made to LEP data, right: fits to LEP, $p\bar{p}$ collider and neutrino experiment data. [9].

Figure 2.5: Top quark mass measurement versus time. The open diamonds represent fits from precision electroweak data. The triangles show the CDF measurements, the inverted triangles the DØ measurements of the top quark mass. The squares are the world-average of these direct measurements. The dashed line shows the lower bound on the top quark mass from $p\bar{p}$ colliders, the solid line the bound from e^+e^- colliders. [10].

Figure 2.6: Lowest order diagrams for $t\bar{t}$ pair production. Top: quark-antiquark annihilation, bottom: gluon fusion.

energy $\sqrt{s}$ can be explained by considering the parton model of the proton. Figure 2.7 shows the $t\bar{t}$ pair production by the collision of a proton with four-momentum P_1 and an antiproton with four-momentum P_2 , where $P_1 \cdot P_2$ is given by the center-of-mass energy $\sqrt{s}$ of the collision:

$$s := (P_1 + P_2)^2 \approx 2P_1 \cdot P_2. \quad (2.5)$$

Figure 2.7: Parton model description of the $t\bar{t}$ pair production.

The protons each consist of quarks, antiquarks and gluons, and one parton of the proton will collide with one parton of the antiproton. They each carry only a fraction of the proton's four-momentum, x_1P_1 or x_2P_2 , respectively. The center-of-mass energy of this subprocess has to fulfill $\sqrt{\hat{s}} \geq 2 \cdot m_t$ to produce a $t\bar{t}$ pair, giving

$$\hat{s} := (x_1P_1 + x_2P_2)^2 \approx 2x_1x_2P_1 \cdot P_2 \approx x_1x_2s \geq 4m_t^2. \quad (2.6)$$

To get a feeling for the magnitude of x for different colliders, the assumption $x_1 \approx x_2 =: x$ can be made, from which follows:

$$x \approx \frac{2m_t}{\sqrt{s}}. \quad (2.7)$$

This assumption gives $x \approx 0.18$ for the Tevatron and $x \approx 0.025$ for the LHC. The parton distribution functions shown in Figure 2.8 give the probability $f(x)dx$ of finding a certain parton species with a given momentum fraction between x and $x + dx$ when probing the proton at a scale Q^2 . The relevant scale for $t\bar{t}$ production is $Q^2 = (175 \text{ GeV})^2$, and the corresponding distributions show that for $x \approx 0.18$ at the Tevatron, the up quark distribution is still larger than the gluon distribution, but at smaller values for x the gluon distribution dominates.

Figure 2.8: Parton distribution functions as functions of the momentum fraction x , note that the gluon function is scaled down by a factor of 10. Left: small Q^2 , right: $Q^2 = m_{top}^2$, relevant for $t\bar{t}$ production [11].

The most recent calculation of the $t\bar{t}$ pair production cross section assuming a top mass of 175 GeV at a center-of-mass energy of 1.96 TeV including next-to-next-to-leading order (NNLO) soft-gluon corrections, next-to-next-to-next-to-leading-logarithm (NNNLL) corrections and some virtual terms is found in [12] and predicts:

$$\begin{aligned}\sigma^{p\bar{p} \rightarrow t\bar{t}}(m_t = 175 \text{ GeV}, \sqrt{s} = 1.8 \text{ TeV}) &= (5.24 \pm 0.31) \text{ pb, and} \\ \sigma^{p\bar{p} \rightarrow t\bar{t}}(m_t = 175 \text{ GeV}, \sqrt{s} = 1.96 \text{ TeV}) &= (6.77 \pm 0.42) \text{ pb.}\end{aligned}$$

The DØ measurement of Run I [13] gives the following cross section:

$$\sigma^{t\bar{t}}(m_t = 172.1 \text{ GeV}, \sqrt{s} = 1.8 \text{ TeV}) = (5.69 \pm 1.21 \text{ (stat)} \pm 1.04 \text{ (sys)}) \text{ pb.}$$

The CDF cross section of Run I [14] is:

$$\sigma^{t\bar{t}}(m_t = 175 \text{ GeV}, \sqrt{s} = 1.8 \text{ TeV}) = (6.5_{-1.4}^{+1.7}) \text{ pb.}$$

The most recent preliminary result of DØ in Run II as presented at the Lepton Photon Conference 2003 is [15]:

$$\sigma^{t\bar{t}}(\sqrt{s} = 1.96 \text{ TeV}) = (8.1_{-2.0}^{+2.2} \text{ (stat)}_{-1.3}^{+1.6} \text{ (sys)} \pm 0.8 \text{ (Lumi)}) \text{ pb.}$$

Top Quark Production via the Weak Interaction

The top quark can also be produced by the weak interaction as single top. The lowest order diagrams of the subprocesses, namely s-channel, t-channel and Wt associated production, are shown in Figure 2.9. Single top production has not been observed yet, but

Figure 2.9: Lowest order diagrams for single top production. Left: s-channel, middle: t-channel, right: Wt associated production.

an observation of the s-channel and t-channel is expected for the Tevatron Run II. The cross sections are proportional to $|V_{tb}|^2$, which allows the first direct measurement of the CKM matrix element V_{tb} .

Decay of the Top Quark

The top quark has an extremely short lifetime of only $\approx 4.6 \cdot 10^{-25}$ seconds and subsequently decays via the charged-current weak interaction. Since in the Standard Model $|V_{tb}| \simeq 1$, the preferred decay products are a W boson and a b quark, with a predicted branching fraction of $BR(t \rightarrow Wb) > 0.998$. The further decay modes of the two W bosons determine the $t\bar{t}$ decay channel. Both the leptonic and the hadronic decays of the W bosons are possible, their branching ratios are listed in Table 2.2. The resulting pos-

Decay mode	Branching ratio BR
$W^+ \rightarrow q\bar{q}'$	$(67.96 \pm 0.35) \%$
$W^+ \rightarrow \tau^+ \nu_\tau$	$(10.74 \pm 0.27) \%$
$W^+ \rightarrow e^+ \nu_e$	$(10.72 \pm 0.16) \%$
$W^+ \rightarrow \mu^+ \nu_\mu$	$(10.57 \pm 0.22) \%$

Table 2.2: Branching ratios of the W^+ boson decay modes [7].

sible $t\bar{t}$ decay channels can be categorized into the all-hadronic channel, the lepton+jets channels, the di-lepton channels, and the τ -channels, which are difficult to analyze because of the τ decay into another lepton or hadrons. All channels and their theoretical branching ratios on tree level are illustrated in Figure 2.10.

Figure 2.10: Pie chart of the $t\bar{t}$ decay channels. The branching ratios are the theoretical predictions on tree level.

This analysis addresses the $e+jets$ channel, where one of the W bosons decays hadronically into quarks and the second one leptonically into an electron and an electron neutrino. The exact branching ratio of the $e+jets$ channel $BR(t\bar{t} \rightarrow e+jets)$ can be calculated as follows:

$$\begin{aligned}
 BR(t\bar{t} \rightarrow e+jets) &= BR(t\bar{t} \rightarrow W^+bW^-\bar{b} \rightarrow e^\pm \bar{\nu}_e^{(-)} q\bar{q}'b\bar{b}) \\
 &= 2 \cdot BR(t\bar{t} \rightarrow W^+bW^-\bar{b} \rightarrow e\nu_e q\bar{q}'b\bar{b}) \\
 &= 2 \cdot (BR(t \rightarrow Wb))^2 \cdot BR(W \rightarrow e\nu_e) \cdot BR(W \rightarrow q\bar{q}') \\
 &= (BR(t \rightarrow Wb))^2 \cdot 2 \cdot (10.72 \pm 0.16)\% \cdot (67.96 \pm 0.35)\% \\
 &= (BR(t \rightarrow Wb))^2 \cdot (14.57 \pm 0.23)\%, \tag{2.8}
 \end{aligned}$$

With $BR(t \rightarrow Wb) \simeq 100\%$ this gives a branching ratio of $BR(t\bar{t} \rightarrow e+jets) = (14.57 \pm 0.23)\%$. The signature of a $t\bar{t} \rightarrow e+jets$ event will be one isolated high p_T electron and large missing transverse energy due to the neutrino from the leptonic W boson decay, two light jets from the hadronic W boson decay and two jets from the hadronization of the b quarks. The event display of a candidate for a $t\bar{t} \rightarrow e+jets$ event, which was actually recorded in the data of Run #168562 as Event #3591065, is shown in Figure 2.11.

2.2.3 Sources of Background

The background to $t\bar{t} \rightarrow e+jets$ events consists of QCD multijet events and $W+jet$ events.

Figure 2.11: A $t\bar{t} \rightarrow e + jets$ candidate event in the x-y view (Run #168562, Event #3591065).

QCD Background

QCD multijet events can resemble $t\bar{t}$ events, if an electron is faked by a pion or a semileptonic heavy flavor decay and the missing transverse energy $\cancel{E}_T$ is mismeasured. How this type of background can be suppressed by a reliable identification of electrons will be discussed in Sections 4.1 and 5.2.

W+Jet Background

The dominant background to $t\bar{t}$ events, however, is W+multijet production, where a W boson is produced via the weak interaction and jets arise from initial state gluon radiation or a quark. Figures 2.12, 2.13 and 2.14 show some possible processes. The suppression of the W+jet background in this analysis is the topic of Sections 5.3 and 5.4.

Figure 2.12: Feynman diagram of a W+0jet event.

Figure 2.13: Feynman diagrams of W+1jet events.

Figure 2.14: Feynman diagrams of W+2jets events.

3 Experimental Setup

The DØ experiment is one of the two experiments at Fermilab's Tevatron collider, which is currently the highest-energy collider in the world. The Tevatron has a radius of 1 km and collides beams of protons and anti-protons at a center-of-mass energy of 1.96 TeV. Fermilab itself is located in Batavia, Illinois, not far west of Chicago. The first data-taking period lasted from 1992 to 1995 (Run I) and during that time the CDF and DØ detectors collected $\sim 125 \text{ pb}^{-1}$ of data each. After an upgrade of the accelerator and the detectors Run II started in March 2001 and by 2009 up to $4.5\text{-}9 \text{ fb}^{-1}$ of data is expected.

This chapter introduces the accelerator and the detector components of the DØ detector. Because the analysis presented in this thesis is based on simulated data, the Monte Carlo generators are also discussed.

3.1 The Tevatron Collider

The Tevatron is the last stage of the accelerator chain, which starts with accelerating hydrogen ions in a Cockroft-Walton accelerator, followed by a linear accelerator. After stripping off the electrons the protons are accelerated to 8 GeV in the Booster synchrotron ring and then to 150 GeV in the Main Injector, where they are also arranged in bunches. Antiprotons are produced by directing some of the proton bunches from the Main Injector at a nickel/copper target. The antiprotons are accumulated and accelerated to 8 GeV before being returned to the Main Injector, where they are also accelerated to 150 GeV. The Tevatron finally accelerates the proton and the antiproton bunches from 150 GeV to 980 GeV each. Both protons and anti-protons are arranged into 36 bunches with a 396 ns bunch spacing, the 36 bunches are again divided into three super-bunches with $2 \mu\text{s}$ between them. Figure 3.1 shows the whole accelerator chain. A more detailed description can be found in [16].

Figure 3.1: The accelerator chain at Fermilab.

3.2 The DØ Detector

Figure 3.2 shows a longitudinal section of the Run II DØ detector, which is around 20 m long and 13 m high. The detector is a multi-purpose detector consisting of several subsystems, which ensure precise measurement of electrons, muons and jets [17]. From the inside going outwards there are the tracking system, the calorimeter system and the muon system. Besides those there are also luminosity monitors and forward proton detectors in the forward region as well as a three-level trigger system. The upgrade for Run II included the complete tracking system with its solenoid, preshower detectors, forward muon chambers, an improved shielding and a faster data acquisition system [18]. For this analysis the tracking system, the calorimeter and the trigger system are especially interesting and will be discussed in detail.

The coordinate system at DØ is a right-handed system with the x-coordinate pointing towards the center of the ring, the y-coordinate upwards and the z-coordinate in the direction of the proton beam. A more useful and common set of coordinates is (r, φ, η) with:

$$r = \sqrt{x^2 + y^2}, \quad (3.1)$$

$$\phi = \tan \frac{y}{x}, \quad \text{and} \quad (3.2)$$

$$\eta = -\ln \tan \frac{\theta}{2}, \quad (3.3)$$

Figure 3.2: Longitudinal section of the DØ detector.

where θ is the angle to the z -axis and η is called the pseudorapidity. Usually the coordinates are given relative to the primary vertex of an event, which in the case of η is then called “physics η ”. Another possibility are the detector coordinates, which are relative to the center of the detector, which will likewise give the “detector η ”.

3.2.1 Tracking System

Due to the solenoid with its magnetic field pointing along the z -direction charged particles will be deflected. By measuring the radius of curvature r of the track in the $r - \varphi$ plane, the transverse momentum $p_T = \sqrt{p_x^2 + p_y^2}$ of the charged particle can be determined:

$$r[\text{m}] = \frac{p_T[\text{GeV}]}{0.3 \cdot B[\text{T}]} \quad (3.4)$$

Measuring the track direction in the $r - z$ plane as well gives the complete three-dimensional information of the track.

Figure 3.3 shows a longitudinal section of one quarter of the inner detector with the tracking system and the preshower detectors.

Figure 3.3: Longitudinal section of the DØ tracking and preshower systems.

Silicon Vertex Detector

The Silicon Vertex Detector or Silicon Microstrip Tracker (SMT) is the innermost of the detector components [18]. An important purpose is to reconstruct secondary vertices from hadrons containing bottom quarks to be able to identify jets which originate from bottom quarks and to distinguish them from jets which originate from light quarks. Another purpose is the distinction from electrons to pions. The detector is 2.40 m long and covers a large range of pseudorapidity up to $|\eta| \approx 3$. Figure 3.4 shows the composition of the Silicon Vertex Detector of six barrels, twelve F-Disks in between and at the end of the barrel segment and four H-Disks in the forward region. Figure 3.5 shows the cross section of a barrel detector.

Figure 3.4: 3-D view of the Silicon Vertex Detector.

Figure 3.5: Cross section of a Silicon Vertex Detector barrel with its layer structure.

This combination of barrels and disks is necessary because the interaction point is extended in z -direction ($\sigma_z = 25$ cm) and makes it difficult to have the tracks always hit the detector surface perpendicularly. The barrel detectors primarily measure the $r - \varphi$ coordinate, whereas the disks also measure the $r - z$ coordinate. The single-hit resolution is very high with $\approx 10 \mu\text{m}$ and small stereo angles provide unambiguous pattern recognition. In total the Silicon Vertex Detector has 793,000 readout channels.

Scintillating Fiber Tracker

The Scintillating Fiber Tracker or Central Fiber Tracker (CFT) is 2.60 m long and can be found at a radial distance of 20-52 cm from the beam as shown in Figure 3.3 [18]. It covers the pseudorapidity up to $|\eta| \approx 1.6$. In total there are 74,000 scintillating fibers in eight doublet layers, where every second doublet layer has an additional doublet layer with a small stereo angle. This provides a spatial hit resolution of about $100 \mu\text{m}$. The fibers have a diameter of $835 \mu\text{m}$ and scintillate around 530 nm, which is the yellow-green part of the visible spectrum. 11 m long waveguides conduct the light to the Visible Light Photon Counter (VLPC), where the signal is read out.

Solenoid

The magnetic field causing the curvature of tracks is provided by the 2 Tesla superconducting solenoid magnet enclosing the tracking system. It is 2.80 m long, has a mean radius of 60 cm and its operating temperature is 10 K.

3.2.2 Calorimeter System

The calorimeter system consists of preshower detectors, the liquid argon calorimeter and the intercristal detector and identifies and measures the energy of electrons, photons and jets. Muons being minimum ionizing particles only leave a very small fraction of their energy in the calorimeter and have to be measured by the muon system. Neutrinos deposit no energy at all in the detector, but their presence in an event can be inferred by the imbalance of the energy in the $r - \varphi$ plane after correcting for the muons. This is called the missing transverse energy $\cancel{E}_T$.

Preshower Detectors

Figure 3.3 shows the preshower detectors, which are designed to enhance electron identification. The Central Preshower Detector (CPS) is found right behind the solenoid at a radius of about 72 cm and covers the region up to a pseudorapidity of $|\eta| = 1.2$. Its purpose is to correct the electromagnetic energy for effects in the solenoid and besides measuring the energy it can also serve as a tracker because of its good spatial resolution of 7 mm. The CPS has three layers of scintillating fibers, which are read out by wavelength-shifting (WLS) fiber readouts in 7,680 channels. The signal is conducted by 10 m long fibers to the VLPCs. The Forward Preshower Detectors (FPS) are mounted right on the end calorimeter cryostats and cover the pseudorapidity range $1.4 < |\eta| < 2.5$. There are two layers of scintillators with a lead absorber in between.

Calorimeter

Because of its excellent performance the DØ liquid argon calorimeter itself has not changed since Run I [19], only the front-end electronics was upgraded. Figure 3.6 shows the longitudinal section of a quarter of the calorimeter. There are three parts, the central calorimeter (CC) and two end calorimeters (EC). Together they cover a pseudorapidity range up to $|\eta| = 4.2$. Each calorimeter part is longitudinally divided into electromagnetic and hadronic sections, the central calorimeter into three of them, from the inside to the outside, the electromagnetic (CCEM), the fine hadronic (CCFH) and the coarse hadronic (CCCH) section. The end cap calorimeters are divided into four sections, the electromagnetic (ECEM) section, the inner hadronic (ECIH) covering $2.0 < |\eta| < 4.2$, the middle hadronic (ECMH) covering $1.5 < |\eta| < 2.0$ and the outer hadronic (ECOH) $1.1 < |\eta| < 1.5$ section. These sections are longitudinally subdivided into cylindrical floors, the electromagnetic sections into EM1 to EM4 and the hadronic sections into FH1 to FH3 and CH going from the inside to the outside. In η and φ the calorimeter is divided into towers of $\Delta\eta \times \Delta\varphi = 0.1 \times 2\pi/64 \simeq 0.1 \times 0.1$, only in the cells of the EM3 layer the granularity is doubled in both η and φ .

Figure 3.6: Longitudinal section of a quarter of the DØ calorimeter.

Figure 3.7: Unit cell in the ECEM of the liquid argon DØ calorimeter [19].

Figure 3.7 shows a unit cell of the calorimeter consisting of a depleted uranium absorber plate, where the showers develop, a liquid argon gap, where the ionization charges are generated, a signal board made of copper, where the charges are collected, followed by another liquid argon gap.

In principle the energy in the calorimeter is measured by counting electron-ion pairs in the liquid argon, where the number n of them is proportional to the deposited energy E . Because the counting gives a Poisson error of $\sqrt{n}$, the relative resolution of the energy measurement should be:

$$\frac{\Delta E}{E} \propto \frac{\Delta n}{n} = \frac{1}{\sqrt{n}} \propto \frac{1}{\sqrt{E}}. \quad (3.5)$$

Including other contributions the exact formula is:

$$\left(\frac{\Delta E}{E}\right)^2 = C^2 + \frac{S^2}{E} + \frac{N^2}{E^2}, \quad (3.6)$$

where the contributions C , S and N come from calibration errors, sampling fluctuations and noise contributions, respectively. Test-beam measurements with electrons and pions [17] gave the results listed in Tabel 3.1.

Particle	C	S	N
e	0.014	0.135 $\sqrt{\text{GeV}}$	0.14 GeV
π	0.032	0.41 $\sqrt{\text{GeV}}$	1.3 GeV

Table 3.1: Test-beam results for calorimeter resolution.

The ratio of the electromagnetic to hadronic response of the DØ calorimeter is 1.02-1.09, which makes it an almost compensating calorimeter.

Intercryostat Detector

Between the central calorimeter and the endcap calorimeter there is a gap due to the cryostats. Because of this gap the region $1.1 < |\eta| < 1.4$ has no coverage by the electromagnetic part of the calorimeter, which is made up for by the Intercryostat Detector (ICD) in exactly this region. The ICDs consist of single layers of scintillating tiles and are mounted on the end cap calorimeters as can be seen in Figures 3.3 and 3.6.

3.2.3 Trigger System

At the Tevatron there is a bunch crossing every 396 ns, resulting in a data rate of 2.5 MHz, but only 50 Hz of data can be stored [20]. The purpose of the trigger system is to recognize

interesting events and initiate the storage for those. The DØ trigger achieves that with its three levels, the elements are shown in Figure 3.8.

Figure 3.8: The DØ trigger elements.

Level 1

On Level 1 (L1) the trigger decision is entirely based on hardware. The calorimeter towers, which are defined to be $\Delta\eta \times \Delta\varphi = 0.2 \times 0.2$ on L1, have to exceed certain thresholds in transverse energy and the hit patterns found in the preshower detectors, the CFT and the muon chambers have to match tracks with a transverse momentum above a preprogrammed value. Together they provide electron triggering up to $|\eta| = 2.5$ and muon triggering up to $|\eta| = 2.0$. The L1 trigger operates deadtimeless and makes a decision in 4.2 μ s, passing on 10 kHz to Level 2.

Level 2

The Level 2 (L2) trigger consists of two stages. In the preprocessor stage the information from the subsystems is collected separately to form objects like a track or an energy cluster. The subsystems included are the calorimeter, the preshower detectors, the CFT and the muon chambers. A Level 2 SMT trigger is being commissioned. All the information is combined in the global processor stage to form physics objects like an electron candidate and based on this information the trigger decision is made. L2 can have a deadtime of 5% at the highest data rates, requires 100 μ s at the most and reduces the data rate to 1 kHz.

Level 3

On Level 3 (L3) a simplified reconstruction of the entire event is made on approximately 100 farm nodes. It further reduces the data rate to the desired 50 Hz, which can then be written to tape for offline analysis. The average event size is 250 kBytes.

3.3 Monte Carlo Simulation

This section briefly discusses the simulation of Monte Carlo events and introduces the main analysis software tool of DØ top quark analyses.

The output of event generators should have the same average behavior and the same fluctuations as real data. In nature the fluctuations arise from quantum mechanics, whereas in the generators Monte Carlo techniques are used. The steps of simulating an event and the problems involved are the following:

1. The parton interaction is calculated in perturbative QCD. This requires taking into account the parton distribution functions of the incoming proton and anti-proton, calculating the matrix element and the phase space. Because of the complexity of the perturbative calculations, only a limited number of higher orders can be considered, usually up to one-loop corrections. DØ at this time uses the parton distribution functions CTEQ5 [11]. For the calculation of the matrix elements, ALPGEN [21] is used.
2. Initial and final state radiation is added by parton showering. This also includes bremsstrahlung of electrons. Parton showering uses approximations as a simplified approach, and experience has shown that the resulting jets are often too soft. For this reason sometimes the higher order matrix elements are calculated separately to obtain a more correct description as will be seen later with W+multijet events in this analysis. At DØ parton showering is done by JETSET, which is now part of PYTHIA [22].
3. There are two types of additional interactions possible in the event. If one parton of the proton or anti-proton interacts with another parton, the rest of the proton and anti-proton will still be color-connected to the main interaction. These beam remnants also contribute to the event, but their influence is not precisely known. This is called the soft underlying event, which does not depend on the luminosity. The second type of additional interactions are the multiple proton interactions, meaning that another proton anti-proton pair from the bunches collides. This occurs more often the higher the luminosity is.
4. Quarks and gluons by themselves carry color charge and have to be transformed into jets of colorless hadrons. The hadronization can be divided into fragmentation

and decays. Especially the fragmentation is not understood very well and so this part is based on models, which vary in the different generators. These include string fragmentation (Lund model) and independent fragmentation by PYTHIA and cluster fragmentation by HERWIG [23]. DØ uses the string fragmentation model by JETSET, which is part of PYTHIA.

5. After producing the events the detector response has to be simulated as well, taking into account the geometry and the material of the detector. This is done by *DØgstar*, which is based on GEANT [24]. To digitize the data it is processed by *DØsim* and the event is reconstructed by *DØreco*.

The main analysis tool used in $t\bar{t}$ analyses at DØ is TopAnalyze [25, 26], a software-package which processes data in Data Summary Tape files (DST) or thumbnails and produces root-tuples, DSTs or thumbnails. It extracts the information about the reconstructed objects in the event, but it is also possible to redo parts of the reconstruction. A list of the Monte Carlo samples used for the studies in this thesis can be found in Appendix A.

4 Physics Objects

This chapter will address all the physics objects important for the $t\bar{t} \rightarrow e + jets$ analysis, namely electrons, jets and missing transverse energy $\cancel{E}_T$. Their reconstruction and reconstruction efficiencies will be discussed. The studies done concentrated on the electron reconstruction, therefore this part will be presented in more detail.

The reconstruction software packages exist in different versions with new versions being released every few months. For the Lepton Photon Conference 2003 analyses version p13 was used. The current version is p14. The entire data of $D\bar{O}$, recorded between summer 2002 and summer 2003 is being reprocessed with p14, so that a coherent data set of $\sim 210 \text{ pb}^{-1}$ is available for analyses to be presented at the winter conferences 2004.

4.1 Electrons

Isolated electrons with a high transverse momentum p_T play a vital role in this analysis. Identifying them with a high efficiency while suppressing the background is the goal of the electron reconstruction.

The information for reconstructing electrons is obtained from three subdetectors, the calorimeter, the tracking detectors, and the central preshower detector. The latter was only recently planned to be introduced into the analyses, but after showing some problems it has not been implemented yet. First, the calorimeter information is used to define an EM cluster, forming an electron candidate. These candidates are then confirmed by additional information from the calorimeter, the tracking system and the preshower detector.

The reconstruction steps will be discussed in this section, beginning with a description of the electron reconstruction as it was done before the reconstruction version p13. Studies on the signal efficiencies and the methods to determine them in Monte Carlo and data were made. QCD multijet events are a major background for electrons. The sources of fake electrons in this background were examined in Monte Carlo. This section concentrates on the discussion of the electron likelihood, introducing the first version and studying several possibilities for improvement in Monte Carlo. It concludes with the presentation of the current version of the electron likelihood which is used for the p14 data. The studies

presented in this thesis form an integral part of the improvements resulting in the new p14 electron likelihood which is now used as the standard electron identification throughout the DØ collaboration.

4.1.1 Electron Reconstruction

The first step of the electron reconstruction uses calorimeter information only. In the CC EM clusters are defined to be a set of adjacent towers around the tower with the highest energy content in a cone of $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} < 0.2$. In the EC EM clusters are a set of adjacent cells with a transverse distance of less than 10 cm from an initial cell with the highest energy content.

Actual EM clusters are expected to deposit a large fraction of their energy in the electromagnetic layers of the calorimeter. Therefore electron candidates are required to have an EM fraction

$$f_{EM} := \frac{E_{EM}}{E_{tot}} > 0.9. \quad (4.1)$$

Figure 4.1 shows the EM fraction distribution of reconstructed electrons in a $Z \rightarrow e^+e^-$ Monte Carlo sample.

Figure 4.1: EM fraction f_{EM} of reconstructed electrons in a $Z \rightarrow e^+e^-$ Monte Carlo sample.

By definition f_{EM} should always be less than or equal to 1, but as can be seen in Figure 4.1 there seem to be electrons with less total energy than electromagnetic energy. This effect is a consequence of noise in the calorimeter. Even without any collisions the calorimeter electronics reads a signal of a certain distribution, which can be measured when the collider is turned off. The mean value of the distribution called the noise pedestal is determined and subtracted from the measured signal. But because of the distribution there can be cells with less noise than the average and subtracting the mean value will

result in a negative energy content of this cell. If this happens the total energy measured in the cluster can be less than the electromagnetic energy. A new algorithm called t42 developed for the next analyses will remove this feature by ignoring cells with supposedly negative energy content.

The shower development of the candidate cluster throughout the calorimeter layers is also taken into account by comparing it to the shower development of test beam electrons, electrons from W boson events and $Z \rightarrow e^+e^-$ events. A χ^2 , which is also called H-matrix, with eight degrees of freedom is attributed to the cluster and in all reconstruction versions before p14, including the one used for the following study, it was required to be

$$H - matrix < 20. \quad (4.2)$$

The H-matrix distribution of reconstructed electrons in a $Z \rightarrow e^+e^-$ Monte Carlo sample is shown in Figure 4.2.

Figure 4.2: H-matrix of reconstructed electrons in a $Z \rightarrow e^+e^-$ Monte Carlo sample.

Additionally, the cluster is required to be isolated in (η, φ) space, meaning that the total energy in a cone of 0.4 around the candidate without the EM energy contained in the 0.2 cone must be less than 15% of the EM energy in the 0.2 cone:

$$f_{iso} := \frac{E_{tot}(\Delta R < 0.4) - E_{EM}(\Delta R < 0.2)}{E_{EM}(\Delta R < 0.2)} < 0.15. \quad (4.3)$$

In Figure 4.3 the isolation distribution of reconstructed electrons in a $Z \rightarrow e^+e^-$ Monte Carlo sample is shown. The observed structure is a consequence of the calorimeter geometry.

Electron candidates, which fulfill all of the above requirements, are called loose electrons. Further confirmation is needed to suppress the overwhelming QCD background and is obtained from requiring a track match or an electron likelihood cut.

Figure 4.3: Isolation f_{iso} of reconstructed electrons in a $Z \rightarrow e^+e^-$ Monte Carlo sample.

Before the reconstruction version p13 only tracking information was used to confirm a loose electron. It was demanded that the candidate had a track match in (η, φ) space and transverse energy measured in the calorimeter with the transverse momentum of the measured track, meaning

$$P(\chi_{trackmatch}^2) > 0.01 \quad \text{with} \quad (4.4)$$

$$\chi_{trackmatch}^2 = \left(\frac{\Delta\varphi}{\sigma_\varphi} \right)^2 + \left(\frac{\Delta z}{\sigma_z} \right)^2 + \left(\frac{E_T/p_T - 1}{\sigma_{E_T/p_T}} \right)^2. \quad (4.5)$$

$\Delta\varphi$ (or Δz) is the difference between the φ (z)-coordinate of the cluster in the EM3 layer of the calorimeter and the φ (z)-coordinate of the track impact in the EM3 layer. E_T/p_T is the ratio of the calorimeter transverse energy and the transverse momentum of the track. σ_φ , σ_z and σ_{E_T/p_T} are the root mean squares (RMS) of the three variable distributions, respectively. $P(\chi_{trackmatch}^2)$ is the χ^2 -probability for a track match in φ , z , and E_T/p_T .

Starting with p13 this technique was replaced by a likelihood-based confirmation, which exploits more of the available information. It will be discussed later in this section.

4.1.2 Efficiency in Monte Carlo and Data

In this part the electron identification efficiency will be discussed. For Monte Carlo events there are two methods to determine this efficiency. The Monte Carlo method takes advantage of the knowledge what really happened in a given event. The data method is developed to measure the electron identification efficiency from $Z \rightarrow e^+e^-$ data. Since Monte Carlo simulations do not describe the physics or the detector performance perfectly, efficiencies for analyses are determined in data, wherever possible, in order to achieve the best precision possible. The difficulty here is the selection of a pure signal sample, in this

case a sample which contains actual electrons. For high p_T electrons a $Z \rightarrow e^+e^-$ sample is used, because the reconstruction of one tight electron, one loose electron candidate and the requirement of both to form an invariant mass compatible with the Z-mass provides a pure sample of electrons as can be seen in Figure 4.4.

Figure 4.4: Invariant Z mass distribution of loose electron pairs in a data sample. Left: events with at least one tight electron, right: all events, the background is clearly noticable [27].

The presented Monte Carlo studies are made to develop ways to learn about electrons in the $t\bar{t}$ sample by studying them in a $Z \rightarrow e^+e^-$ sample. The studies concentrate on two questions. $Z \rightarrow e^+e^-$ events have a different topology than $t\bar{t} \rightarrow e + jets$ events, which results in different electron identification efficiencies. How different they are, the reasons for that and ways to take those differences into account is the first question, which will be addressed by comparing the efficiencies obtained by the Monte Carlo method in the two different samples. The second question being addressed is the comparison of the two methods examining the efficiencies for the $Z \rightarrow e^+e^-$ sample.

For the efficiency studies p13.05 Monte Carlo samples were used. They consisted of 21,000 $t\bar{t} \rightarrow e + jets$ events and 101,000 $Z \rightarrow e^+e^-$ events. In addition to the usual electron candidate requirements as described in Section 4.1.1 a $p_T > 15$ GeV cut was applied, since the $t\bar{t}$ cross section measurement uses high p_T electrons.

Monte Carlo Method The Monte Carlo method simply calculates the ratio of the number of found tight electrons in the sample to the number of generated electrons of the sample:

$$\varepsilon_{MC} = \frac{\text{number of tight electrons}}{\text{number of generated electrons}} = \frac{N_{tight}}{N_{gen}}, \quad (4.6)$$

where ε_{MC} depends on p_T , η , and φ due to geometrical effects in the detector and resolution effects. The Monte Carlo method to measure efficiencies can be used in both

$t\bar{t} \rightarrow e + jets$ and $Z \rightarrow e^+e^-$ samples, but of course only with Monte Carlo information. The error on the efficiency is given by the binomial error:

$$\Delta(\varepsilon_{MC}) = \sqrt{\frac{\varepsilon \cdot (1 - \varepsilon)}{N_{gen}}} \quad (4.7)$$

Comparison of Monte Carlo Method Efficiencies in the Two Samples A direct comparison of the efficiencies as a function of p_T and as a function of ϕ obtained by the Monte Carlo method in CC and EC in the two different samples is shown in Figure 4.5. The figure shows that the efficiency in the $Z \rightarrow e^+e^-$ sample is higher than in the $t\bar{t} \rightarrow e + jets$ sample.

Figure 4.5: Electron identification efficiencies measured with Monte Carlo method in a $t\bar{t} \rightarrow e + jets$ (gray circles) and $Z \rightarrow e^+e^-$ sample (black squares) without any adjustments. Top left: as function of the generated p_T in CC, top right: as function of the generated p_T in EC, bottom left: as function of ϕ in CC with $p_T > 15$ GeV, bottom right: as function of ϕ in EC with $p_T > 15$ GeV.

There are two possible reasons for the observed differences in the measured efficiencies. The first reason is the different topology of the two samples, electrons from the Z decay

being more isolated than the ones from the top quark decay. The second reason are the different p_T and η distributions of the electrons in the two samples. When looking at the efficiencies as functions of p_T , the efficiencies are integrated over η , so that a different η distribution will change the efficiency in a certain p_T region.

In the following, two adjustments will be developed to account for the differences of the measured efficiencies in the $t\bar{t}$ and the $Z \rightarrow e^+e^-$ sample. In a first step the isolation of electrons will be required in both samples. In a second step the differences in the p_T distributions will be removed by reweighting the p_T distribution of the electrons in the $Z \rightarrow e^+e^-$ sample to the one of the electrons in the $t\bar{t}$ sample.

Figure 4.6: ΔR distance of the electron candidate to the next jet in the $t\bar{t} \rightarrow e + jets$ and $Z \rightarrow e^+e^-$ samples. Top: CC, bottom: EC.

Figure 4.6 compares the ΔR distance of the electron candidate to the nearest jet in the central calorimeter region in the $t\bar{t} \rightarrow e + jets$ sample to the $Z \rightarrow e^+e^-$ sample. It can readily be seen that the electron candidates from the Z decay are more isolated and that most $Z \rightarrow e^+e^-$ events do not contain any jets at all. The distribution looks very much the same in the endcap region. The peak at $\Delta R = 0$ and the dip at $\Delta R = 0.5$ are consequences of the reconstruction algorithm assigning calorimeter clusters both to an electron candidate and to a jet candidate, but then removing jets which are close ($\Delta R = 0.5$) to an electron candidate with $p_T > 15$ GeV. Now, if the electron candidate had less than $p_T = 15$ GeV the jet is not removed although the calorimeter cluster was probably due to an electron.

To compensate for the lesser isolation of the $t\bar{t}$ electrons, a cut was introduced in the Monte Carlo efficiency study to exclude all events with non-isolated electron candidates from the $t\bar{t}$ sample. The definition of a non-isolated electron was chosen to be an electron with a jet closer than $\Delta R = 0.9$, because the isolation criterium for an electron candidate requires a jet to stay further away than $\Delta R = 0.4$ and the radius of a jet cone is $R = 0.5$. Checking back with the ΔR distribution in Figure 4.6, not much statistics is lost by making this cut and no bias is introduced in the efficiency determination.

Now the agreement in the two samples is much better as can be seen in Figure 4.7. In fact in the central calorimeter region the two samples give the same efficiency, whereas in the endcap region they still show major differences. These will be taken care of by the next step.

Figure 4.7: Electron identification efficiencies measured with Monte Carlo method and ΔR cut in a $t\bar{t} \rightarrow e + \text{jets}$ (gray circles) and $Z \rightarrow e^+e^-$ sample (black squares). Top left: as function of p_T in CC, top right: as function of p_T in EC, bottom left: as function of ϕ in CC, bottom right: as function of ϕ in EC.

Figure 4.8 compares the p_T , η and ϕ -distribution of the electron candidate in the $t\bar{t} \rightarrow e + \text{jets}$ sample to the $Z \rightarrow e^+e^-$ sample. The p_T -distribution is obviously very different, especially in the sensitive region between 20 GeV and 60 GeV, where the efficiency

Figure 4.8: p_T , η and ϕ of the electron candidate for the $t\bar{t} \rightarrow e + \text{jets}$ and $Z \rightarrow e^+e^-$ sample.

changes rapidly with p_T . The η -distribution of the electrons from $t\bar{t}$ is narrower than the distribution of the electrons from Z decay. The ϕ -distributions are flat in both samples as expected. The most striking difference in the distributions therefore shows up in the p_T -distribution and is taken into account by weighting the p_T -distribution of the generated electrons from $Z \rightarrow e^+e^-$ to the distribution of the generated electrons from $t\bar{t} \rightarrow e + \text{jets}$. This is done for CC and EC separately to also account for some of the η dependency. The weighting will change the calculation of the binomial error to:

$$\Delta(\varepsilon_{MC,weighted}) = \sqrt{\frac{\varepsilon \cdot (1 - \varepsilon)}{N_{eff}}} \quad \text{with} \quad N_{eff} = \frac{(\sum_j g_j)^2}{\sum_j (g_j)^2} \quad , \quad (4.8)$$

g_j being the weight of event j in a given bin.

Figure 4.9: Electron identification efficiencies measured with the Monte Carlo method, ΔR cut and weighted p_T -distribution in a $t\bar{t} \rightarrow e + jets$ (dark gray circles) and in a $Z \rightarrow e^+e^-$ sample (unweighted as black squares, weighed as light triangles). Top left: as function of p_T in CC, top right: as function of p_T in EC, bottom left: as function of ϕ in CC, bottom right: as function of ϕ in EC.

The efficiency was recalculated with the new ΔR cut and with weighted p_T -distributions, the result is shown in Figure 4.9. Of course the efficiency as a function of p_T is not affected by the weighting of the p_T -distribution. But in the ϕ -dependent plots it can clearly be seen that the agreement between the two samples is now achieved. Further improvement could be made by weighting the η -distribution as well, but this was not pursued, because it is not expected to improve the already achieved agreement very much.

Data Method The data method of determining the electron identification efficiency in the $Z \rightarrow e^+e^-$ sample works as follows [27]. For each event two loose electron candidates are required. One of them, the tag electron “1”, also has to be tight. The invariant mass of the two electrons is calculated and if it lies within the Z mass window, the test electron “2” is most likely a real electron as well. It is then asked if that electron is also found as a tight electron. The ratio of the number of events with test electron “2” found as tight to

the number of events with test electron “2” found as loose gives the tight reconstruction efficiency. This data method was examined in a $Z \rightarrow e^+e^-$ Monte Carlo sample for this study.

Figure 4.10: Invariant Z mass distribution of loose electron pairs in a data sample. Left: events with no tight electron, middle: events with exactly one tight electron, right: events with exactly two tight electrons [27].

Figure 4.11: Invariant Z mass distribution of loose electron pairs in a $Z \rightarrow e^+e^-$ Monte Carlo sample. Left: events with no tight electron, middle: events with exactly one tight electron, right: events with exactly two tight electrons.

Figures 4.10 and 4.11 show the invariant mass distribution of electron pairs in events with 0, 1, or 2 tight electrons in a data sample and in the $Z \rightarrow e^+e^-$ Monte Carlo sample, respectively. This tight reconstruction efficiency, however, is not the same efficiency calculated with the Monte Carlo method, because it only asks what fraction of loose electrons will also be found as tight. It is in fact the tracking and matching efficiency. To compare the data method with the Monte Carlo method, the efficiency of finding loose electrons has to be multiplied to the described efficiency. In the data the efficiency for finding loose electrons is essentially the efficiency of the trigger system and can be measured as described in [28] and [29]. Here it was determined in a Monte Carlo method to be the ratio

of the number of loose electrons to the number of generated electrons and the complete electron identification efficiency determined with the data method reads:

$$\begin{aligned}
\varepsilon_{data} &= \frac{\text{number of events with test electron 2 tight}}{\text{number of events with test electron 2 loose}} \cdot \frac{\text{number of loose electrons}}{\text{number of generated electrons}} \\
&= \frac{N_{2tight}}{N_{2loose}} \cdot \frac{N_{loose}}{N_{gen}} \\
&= \varepsilon_2 \cdot \varepsilon_{loose}
\end{aligned} \tag{4.9}$$

The error on the efficiencies is again given by the binomial errors for weighted efficiencies as in Equation 4.8, combining the two by error propagation:

$$\Delta(\varepsilon_{data}) = \sqrt{\Delta(\varepsilon_2) \cdot \varepsilon_{loose} + \varepsilon_2 \cdot \Delta(\varepsilon_{loose})} \tag{4.10}$$

Comparison of the Two Methods in the $Z \rightarrow e^+e^-$ Sample The electron identification efficiency was determined by the data method in the $Z \rightarrow e^+e^-$ Monte Carlo sample. The minimal distance to the next jet cut and the weighting of the p_T distribution was applied. The result compared to the efficiencies obtained with the Monte Carlo method is shown in Figure 4.12. The figure also compares the results of the two methods in the $Z \rightarrow e^+e^-$ Monte Carlo sample to the one obtained in the $t\bar{t}$ sample. The three different methods in the two samples agree reasonably well, which can be seen in particular in the plot showing the efficiency versus φ . Especially the Monte Carlo and the data method in the $Z \rightarrow e^+e^-$ sample are consistent, whereas there is still a small difference to the $t\bar{t}$ sample in the endcap region. This remaining discrepancy is due to differences between the η distribution of the electrons in the $Z \rightarrow e^+e^-$ sample to the one of the electrons in the $t\bar{t}$ sample combined with differences in the η -dependant efficiency and should vanish when weighting in η as well. Nevertheless it can be concluded that the data method does allow to measure the correct efficiency and that the $Z \rightarrow e^+e^-$ sample can be used to determine the electron identification efficiencies for a $t\bar{t} \rightarrow e + jets$ analysis, provided it is reweighted at least in p_T .

4.1.3 Sources of Background

Before proceeding to the new likelihood-based confirmation of electron candidates the QCD multijet background which fakes good reconstructed electrons will be examined. It is believed to consist of three main sources.

1. Leading π^0 s that decay into photons, which convert to e^+e^- pairs or have a track faked by a charged particle.
2. Charged pions that undergo charge exchange in the detector material.
3. Fluctuations in the shower shape.

Figure 4.12: Electron identification efficiencies measured with two different methods in a $t\bar{t} \rightarrow e + jets$ (dark gray circles) and $Z \rightarrow e^+e^-$ sample (weighted with MC method as black squares, weighted with data method as light gray triangles). Top left: as function of p_T in CC, top right: as function of p_T in EC, bottom left: as function of ϕ in CC, bottom right: as function of ϕ in EC.

To confirm these assumptions a QCD Monte Carlo sample was examined. It contained 104,000 p13 QCD multijet Monte Carlo events with p_T cuts of 20, 40, or 80 GeV. Requiring a loose electron (EM fraction > 0.9 , isolation < 0.15 , H-matrix < 20) with $p_T > 15$ GeV gave 879 loose electron candidates. If additionally a track match with a χ^2 -probability > 0.01 was required, 38 tight electron candidates remained.

Of those 38 tight electron fakes, 19 were found to be single π^0 or η , 12 were several π^0 or η , three were charged pions, one was a photon that was radiated by a quark and the remaining three were non-isolated electrons from meson decays.

Instead the tight requirement can be based on a likelihood combining information from several detector components. In reconstruction version p13 used for the Lepton Photon conference 2003 analyses such a likelihood was introduced [30] and is described in more detail in Section 4.1.4. Using this likelihood with a cut at 0.4 and a spatial track match with χ^2 -probability > 0.01 as tight requirement, only five tight electron candidates survive.

Two are faked by a single π^0 or η , two by several π^0 or η , and one by a charged pion. See Table 4.1 for a summary.

Therefore it can be concluded that the QCD multijet background for electrons is mainly made up of neutral pions.

Requirements	total	by π^0/η	by $x \cdot (\pi^0/\eta)$	by $\pi^\pm$	by $u \rightarrow \gamma \dots$	by meson decays
loose electron ($p_T > 15$ GeV)	879					
loose electron ($p_T > 15$ GeV) + track match	38	19	12	3	1	3
loose electron ($p_T > 15$ GeV) + likelihood + spatial track match	5	2	2	1	0	0

Table 4.1: Number of fake electrons in QCD multijet events.

4.1.4 Electron Likelihood Studies

Although the track match confirmation does perform very well, additional information to suppress background is available. This information lies in the shape of the distributions of certain variables and can be used in a likelihood-based confirmation. The electron likelihood was introduced for the reconstruction version p13 and is described in detail in [30]. The most important facts will be summarized. This part concentrates on the Monte Carlo studies done to improve the electron likelihood. A discussion of some of these studies and the final p14 electron likelihood are also documented in [31].

Electron Likelihood in p13

The p13 electron likelihood uses six quantities:

1. χ^2 -probability of a spatial track match with

$$\chi_{spatial}^2 = \left(\frac{\Delta\varphi}{\sigma_\varphi} \right)^2 + \left(\frac{\Delta z}{\sigma_z} \right)^2. \quad (4.11)$$

$\Delta\varphi$ (or Δz) is the difference between the φ (z) -coordinate of the cluster in the EM3 layer of the calorimeter and the φ (z) -coordinate of the track impact in the EM3 layer. σ_φ and σ_z are the root mean squares (RMS) of the two variable distributions.

2. E_T/p_T as described in 4.1.1.
3. H-matrix as described in 4.1.1.
4. EM-fraction as defined in 4.1.
5. Distance of closest approach (DCA). The DCA is the shortest distance of the track to the primary vertex in the r - ϕ -plane.
6. Track isolation. This variable is defined to be the ΔR -distance to the second closest track, assuming that the closest track belongs to the electron candidate.

Some of these variables require a track candidate, so an additional requirement for loose electrons is made. The electron candidate has to have a track candidate that satisfies

$$|\Delta\varphi_{EM,track}| < 0.05, \quad |\Delta\eta_{EM,track}| < 0.05. \quad (4.12)$$

The reference distributions in [30] were made with data samples. For the signal $Z \rightarrow e^+e^-$ events were selected, for the background EM+jet events dominated by QCD di-jet and γ +jet events.

To differentiate between signal-like and background-like electron candidates a likelihood discriminant is calculated.

$$D(\mathbf{x}) = \frac{P_{sig}(\mathbf{x})}{P_{sig}(\mathbf{x}) + P_{bkg}(\mathbf{x})} \quad \text{with} \quad P_j(\mathbf{x}) = \prod_i P_j(x_i), \quad (4.13)$$

where $P_j(x_i)$ is the probability of the electron candidate to be consistent with sample j when looking at variable x_i . The closer $D(\mathbf{x})$ is to 1, the more signal-like is the candidate, the closer it is to 0, the more background-like is the candidate. This approach assumes the different variables to be uncorrelated.

In [30] it was shown that this likelihood performs much better than the track match confirmation. In the central calorimeter region the requirement of $D(\mathbf{x}) > 0.4$ gives 84% signal efficiency with 25% background efficiency, in the endcap calorimeter region that is 88% signal efficiency with only 10% background efficiency.

Electron Likelihood in Monte Carlo

In order to reproduce the likelihood studies of [30] in Monte Carlo the fake electron statistics in the available QCD multijet Monte Carlo samples is too low. But knowing

from the previous study that in QCD events electrons are mainly faked by pions, single pion samples can be used as the representative background sample. 84,000 single π^0 events and 173,000 single $\pi^\pm$ events were generated in p14.03 using the DØ MCSingle package d0_mcpg_gen. Together they serve as the background sample, mixed in the ratio of 1:2, because according to strong isospin symmetry they are produced in that ratio. For the signal sample 20,100 single electrons were generated also in p14.03. As can be seen in Figure 4.13, all samples were generated with flat p_T , η and ϕ distributions and the same amount of negatively and positively charged particles.

Figure 4.13: p_T , η , ϕ and charge distribution of generated single particles.

The preselection was kept as similar as possible to the one made in data, requiring one EM cluster with $p_T > 20$ GeV, EM fraction > 0.9 , isolation < 0.15 , and H-matrix < 20 with an associated track candidate with $\Delta\phi < 0.05$ and $\Delta\eta < 0.05$. Table 4.2 shows the number of events passing the preselection, separate for the central calorimeter region (CC), meaning $|\text{detector } \eta| < 1.1$, and the endcap calorimeter region (EC) for $1.5 < |\text{detector } \eta| < 2.5$. It can be seen that electrons faked by π^0 s dominate the background.

Of the six input variables used for the likelihood, two do not make much sense in the single particle samples, one being the distance of closest approach to the reconstructed primary vertex, because no vertex can be reconstructed with only one track, and the distance to

Sample	# of generated events	# of events passing in CC	# of events passing in EC
e	20,100	7,551	6,139
π^0	84,000	3,500	4,101
$\pi^\pm$	173,000	198	38

Table 4.2: Number of single particle Monte Carlo events passing the preselection.

the second closest track, because, by construction, in most of these events there is only one track. The latter variable should not have been used anyway. The distance to the second closest track, and especially the distribution of this variable, depends on the topology of the event and gives not only information about the quality of the electron. Electrons from $t\bar{t} \rightarrow e + jets$ events will have a different distribution in that quantity than electrons from $Z \rightarrow e^+e^-$ events, but still be electrons of the same quality. A replacement for this variable making use of the track information will be discussed in the next section.

The four remaining input variable distributions are shown in Figure 4.14 for the CC, and in Figure 4.15 for the EC.

Figure 4.14: The four likelihood input variables in CC. Top left: natural logarithm of the χ^2 -probability of the spatial track match ($\log(P(\chi^2_{spatial}))$), top right: E_T of the calorimeter cluster / p_T of the track, bottom left: H-matrix, bottom right: EM fraction.

Figure 4.15: The four likelihood input variables in EC. Top left: natural logarithm of the χ^2 -probability of the spatial track match ($\log(P(\chi^2_{spatial}))$), top right: E_T of the calorimeter cluster / p_T of the track, bottom left: H-matrix, bottom right: EM fraction.

Calculating the likelihood with those input variables one gets efficiencies in signal and background for different likelihood cuts as shown in Figure 4.16.

Figure 4.16: Likelihood cut efficiencies in signal and background with four input variables $\log(P(\chi^2_{spatial}))$, E_T/p_T , H-matrix and EM fraction.

Number of Tracks in the Vicinity of the Electron

The track information can be utilized to address two different questions. One is identifying fake electrons by photon conversions, which can be done by considering the number of tracks in a small cone around the electron candidate track. The second question is recognizing fake electrons that are part of or very close to a jet by examining the number of tracks in a jet-size cone around the electron candidate track.

Number of Tracks in a Small Cone Knowing that the background is dominated by π^0 s, it seems highly desirable to suppress fake electrons by photon conversions. For neutral pions first decaying into photons and those converting to e^+e^- pairs, one would expect to observe two tracks, whereas a real electron should just show one track. In the single particle Monte Carlo samples it is not necessary to restrict the examination of the number of tracks to a cone, because all additional tracks in the event will come from photon conversions. Therefore the single π^0 sample can instead be used to determine the optimal size of the small cone, which will be done in the next subsection.

Figure 4.17 shows the distribution of the new fifth input variable to the likelihood, the number of tracks in a single particle Monte Carlo event. Figure 4.18 then shows the efficiencies for different likelihood cuts comparing the likelihood with four input variables to the new one with five input variables. The plot shows a significant improvement and so the Monte Carlo studies recommend the introduction of this new variable.

Figure 4.17: The fifth likelihood input variable, the number of tracks per single particle MC event. Left: CC, right: EC.

Size of the Small Cone In more complicated samples there are many more tracks in a single event, which makes it necessary to introduce a small cone in which the number of tracks can be counted. The optimal cone size is examined in the single π^0 Monte Carlo sample. Figure 4.19 shows the distribution of the ΔR distance of the tracks in a single particle Monte Carlo event to the electron candidate track. Additionally, Figure 4.20

Figure 4.18: Likelihood cut efficiencies in signal and background.

shows the number of tracks collected in a 0.05 cone around the candidate track compared to the number of all tracks in the event, Figure 4.21 the same for a 0.1 cone. The plots suggest a 0.1 cone to collect all conversion tracks.

Figure 4.19: ΔR of the tracks in the event to the electron candidate track in the single electron and single π^0 samples in CC and EC. If there are no other tracks than the candidate track, or if the track is more than 0.2 away, an entry is made at 0.2.

Figure 4.20: Number of tracks in single π^0 events, comparing the number of all tracks in the event to the number of tracks in a 0.05 cone around the electron candidate track.

Figure 4.21: Number of tracks in single π^0 events, comparing the number of all tracks in the event to the number of tracks in a 0.1 cone around the electron candidate track.

Number of Tracks in a Large Cone Another track-related variable, which should give a good separation of real isolated electrons to fake electrons from QCD multijet events, is the number of tracks in a larger cone of 0.4-0.5, the size of a jet. This should be a good indication if the electron candidate was part of or close to a jet and help suppressing charged pions. However, it cannot be studied in the single particle Monte Carlo samples. Instead the proposed variable, a result of this MC study, is investigated in data events in [31] and shows indeed strong separation power.

p_T -dependent Input Variables

Some of the input variable distributions are p_T -dependent. As an example the change of the H-matrix distribution in the CC for four different p_T regions is shown in Figure 4.22. The shapes and the peak positions of the H-matrix distributions change, especially in

Figure 4.22: H-matrix distributions in the CC for different p_T regions, signal in light gray, background in dark gray. Top left: $20 \text{ GeV} \leq p_T < 50 \text{ GeV}$, top right: $50 \text{ GeV} \leq p_T < 80 \text{ GeV}$, bottom left: $80 \text{ GeV} \leq p_T < 110 \text{ GeV}$, bottom right: $110 \text{ GeV} \leq p_T < 140 \text{ GeV}$.

the background. For the p_T region between 20 GeV and 50 GeV there is a much better separation between signal and background than for the p_T integrated distribution. The

region is especially interesting for a $t\bar{t} \rightarrow e + jets$ analysis, because most of the electrons in $t\bar{t}$ events will have their transverse momentum in that region.

Using p_T -dependent input histograms did not enhance the performance, however, as can be seen in Figure 4.23. This is counterintuitive and in a first examination it turns out that some of the other input variable distributions cancel the better separation of signal and background in the H-matrix distributions. This effect should be studied more closely.

Figure 4.23: Performance of the likelihood using p_T -integrated input histograms (gray circles) compared to using p_T -dependant input histograms (black squares).

η -dependent Input Variables

The η -dependence of the input variable distributions was also examined. Again the H-matrix is shown as an example in Figure 4.24. The shape of the distribution does change, but it changes in the same way in signal and background, so that no additional separation is gained.

Three-Class-Likelihood

The electron likelihood described in 4.1.4 is a two-class-likelihood, as it only differentiates between the two classes signal and background. Since the studies of the QCD background (ref. 4.1.3) revealed that there are two main sources of fake electrons, namely π^0 s and $\pi^\pm$ s, which differ from real electrons in quite different characteristics, it seems to make sense to treat them separately. One way to do that is using a three-class-likelihood that discerns between the two different background classes, which will be explored in this part. The

Figure 4.24: H-matrix distributions for different η regions, signal in light gray, background in dark gray. Top left: $|\eta| < 0.5$, top right: $0.5 \leq |\eta| < 1.0$, middle left: $1.0 \leq |\eta| < 1.5$, middle right: $1.5 \leq |\eta| < 2.0$, bottom left: $2.0 \leq |\eta| < 2.5$.

difference of the two background distributions in comparison to the signal distributions for the central calorimeter region is shown in Figure 4.25.

Figure 4.25: The five likelihood input variables in CC for the three classes separately. Top left: logarithm of the χ^2 -probability of the spatial track match ($\log(P(\chi^2_{\text{spatial}}))$), top right: E_T of the calorimeter cluster / p_T of the track, middle left: H-matrix, middle right: EM fraction, bottom left: number of tracks per single particle MC event.

For the three-class-likelihood Equation 4.13 changes to the following:

$$D(\mathbf{x}) = \frac{P_{sig}(\mathbf{x})}{P_{sig}(\mathbf{x}) + P_{bkg1}(\mathbf{x}) + P_{bkg2}(\mathbf{x})} \quad \text{with} \quad P_j(\mathbf{x}) = \prod_i P_j(x_i), \quad (4.14)$$

The performance of three-class-likelihood does not seem to exceed the one of the simple two-class likelihood as can be seen in Figure 4.26.

Figure 4.26: Performance of the two-class-likelihood (gray circles) compared to the performance of the three-class-likelihood (black squares).

If however the same calculation is repeated on a pure $\pi^\pm$ background, the three-class-likelihood shows a huge advantage (see Figure 4.27). The three-class-likelihood could therefore play an important role in suppressing the charged pion background, but as former studies have shown the neutral pions make up the majority of the background, so that the charged pion rejection is not as important. The difficulty in data would also be to select a charged pion sample.

4.1.5 Electron Likelihood in p14

For the p14 electron identification a new version of the electron likelihood was developed, including all results and conclusions of this Monte Carlo study. In this part the most important facts will be summarized, for more detail see [31].

The H-matrix cut for an electron candidate was loosened from 20 to 75 to gain more signal efficiency.

The p14 likelihood uses the following seven input quantities:

Figure 4.27: Performance of the two-class-likelihood (gray circles) compared to the performance of the three-class-likelihood (black squares) on a pure $\pi^\pm$ background.

1. χ^2 -probability of a spatial track match as defined in 4.11.
2. E_T/p_T .
3. H-matrix.
4. EM-fraction.
5. Distance of closest approach (DCA) as described in 4.1.4.
6. Number of tracks in a $\Delta R < 0.05$ cone including the track candidate.
7. $\sum p_T$ of all tracks in a $\Delta R < 0.4$ cone excluding the track candidate.

In the original concept for the new likelihood it was planned to include CPS information for the very first time in an eighth variable for CC only:

$$CPS_{stripmax} = \frac{\text{Number of CPS strips}}{\text{EM energy}} \quad (4.15)$$

Due to a bug in the packing of CPS data, the energy information of the strips cannot be used, which is why the number of strips is used instead. Unfortunately, the following Monte Carlo study discovered that this CPS variable is topology-dependent.

The distributions of the eight input variables in the CC can be seen in Figure 4.28, which compares the distribution of the variables in a $t\bar{t}$ Monte Carlo sample to the distributions in a p_T -weighted $Z \rightarrow e^+e^-$ Monte Carlo sample. While the first seven distributions agree very well, the CPS variable distribution is shifted. This prompted the development of a new CPS variable, but because of another bug in the reconstruction, neither of the CPS variables could be used for the new electron likelihood in the end. The reference

Figure 4.28: The eight likelihood input variables in CC. The light gray distributions are from the $t\bar{t}$ Monte Carlo sample, the dark gray are from the weighted $Z \rightarrow e^+e^-$ Monte Carlo sample. Top from left to right: EM fraction, H-matrix, E_T/p_T , distance of closest approach in cm. Bottom from left to right: $\log(P(\chi^2_{\text{spatial}}))$, number of tracks in a $\Delta R < 0.05$ cone, $\sum p_T$ of all tracks in a $\Delta R < 0.4$ cone, number of CPS strips/EM energy.

distributions of the seven remaining input variables are again determined in $Z \rightarrow e^+e^-$ data or EM+jet data, respectively.

This new version of the electron likelihood performs even better than the first version in addition to being less topology-dependent. It was shown that a discriminant cut at 0.75 in CC and 0.8 in EC gives a signal efficiency of 89% to 12% background efficiency in the CC and 86% signal efficiency to 6% background efficiency in the EC. However, at the time when the main analysis of this thesis was made, these cuts were not optimized yet, which is the reason why this analysis will use a likelihood discriminant cut of 0.8 for both CC and EC.

4.2 Jets

Reconstructing jets, the reconstruction efficiency and the jet energy scale to correct back to the particle level energy are the subjects of this section.

4.2.1 Jet Reconstruction

For $t\bar{t}$ analyses jets are reconstructed using the “improved legacy cone algorithm” [32] with a cone size of $\Delta R = 0.5$. The resulting clusters undergo the following calorimeter cuts in order to remove different kinds of fake jets:

- $0.05 < f_{EM} < 0.95$ with EM fraction f_{EM} as defined in Equation 4.1. This cut is applied to remove electromagnetic particles from the list of jets.
- Fraction of energy deposited in the coarse hadronic calorimeter $CHF < 0.4$. Because there is a high amount of noise in the coarse hadronic section of the calorimeter, this cut is applied to remove noise jets.
- Ratio of the highest to the next-to-highest transverse energy cell $HotF < 10$. This prevents jets being clustered from hot cells.
- Number of towers containing 90% of the jet energy $n_{90} > 1$. This cut removes jets clustered from a single hot tower.
- Good jets used for jet multiplicity distributions and the calculation of topological variables are only those with $E_T > 15$ GeV and $|\eta| < 2.5$.

In the data an additional cut is made involving the energy of the jet on trigger level 1, because the electronics of the L1 trigger does not produce much noise. But on L1 the energy in the coarse hadronic (CH) section of the calorimeter is not included and the energy is not jet-energy-scale corrected yet, so that the cut for CC and EC reads:

$$\frac{\text{L1 } E_T}{\text{offline } E_T \text{ w/o CH and JES}} > 0.4.$$

For the ICD region not all towers are used yet, so that the performance is worse and the cut is loosened to 0.2.

Because EM particles (electrons and photons) with a transverse energy greater than 8 GeV are also reconstructed as cone jets and the correct energy scales have to be applied, all jets are categorized into three different objects:

- high p_T electrons: $p_T > 15$ GeV, EM fraction, H-matrix, isolation and track match as described in Section 4.1.1.

- high p_T “photons” (probably π^0 s): same cuts as electrons, but without track match.
- jets: 0.5 cone with above jet quality cuts, $\Delta R > 0.5$ to next electron or photon.

Electrons are corrected by the EM scale. “Photons” not matched to jets are also corrected by the EM scale. If they are matched to jets, they are corrected by the jet energy scale, just like all remaining jets which are not matched to an electron. After correcting these jet objects, the missing transverse energy $\cancel{E}_T$ has to be corrected as well.

4.2.2 Jet Reconstruction Efficiency

The jet reconstruction efficiency is determined in data. Dijet events, which satisfy the following requirements, are used. One jet, the tag jet, has to pass the calorimeter cuts described above (caljet) and be confirmed by the level 1 trigger and the presence of tracks. The probe jet has to be found as a trackjet back to back in φ , meaning $\Delta\varphi > 3.0$. The efficiency of finding the trackjet also as a caljet in a cone of $\Delta R < 0.5$ around the trackjet gives the reconstruction efficiency. Figure 4.29 shows the reconstruction efficiency including all the ID cuts in a p14 sample as light gray crosses with continuous error bars [33]. To be unbiased in flavor γ +jet events can also be used to determine the jet reconstruction efficiency.

Figure 4.29: Reconstruction efficiency for jets including all ID cuts in a p14 sample (light gray crosses with continuous error bars) [33].

4.2.3 Jet Energy Scale

Due to non-linearities, dead material, noise and showering effects in the calorimeter the measured energy of a jet is not equal to the particle level energy. This is corrected by the jet energy scale:

$$E_{corr} = \frac{E_{meas} - E_{offset}}{R \cdot S} \quad (4.16)$$

where E_{offset} is an energy offset caused by electronics and uranium noise, energy pile-up from previous crossings, additional $p\bar{p}$ interactions and underlying events. R is the response of the calorimeter to the jet and S is the shower leakage, that is the fraction of energy outside the 0.5 cone.

The jet energy scale is studied in γ +jet events in Monte Carlo as well as in data. Because Monte Carlo does not simulate the detector response perfectly, a different jet energy scale is applied in Monte Carlo.

4.3 Missing Transverse Energy

Neutrinos cannot be detected directly, instead their presence in an event can only be inferred by 3-momentum-conservation and missing measured energy. Because the Tevatron is a hadron collider, the energy z-component of the event is unknown, so that the calculation is limited to the transverse energy. To calculate the missing transverse energy $\cancel{E}_T$, the transverse energy of all cells greater than 100 MeV are added up vectorially except those in the coarse hadronic calorimeter if they are not part of a good jet because of the high level of noise in that part of the calorimeter. The vector opposite to the resulting vector is considered to be the $\cancel{E}_T$, assuming 3-momentum-conservation.

The following corrections must be applied to the raw $\cancel{E}_T$:

- Jet energy scale corrections for all good jets in the event.
- Electron scale corrections for all EM objects in the event.
- A muon being a minimum ionizing particle deposits only very little energy in the calorimeter. The $\cancel{E}_T$ has to be corrected with the measured momentum of a matched muon track taking into account the expected deposition of energy in the calorimeter from detector simulations.

5 Cross Section Analysis

In this chapter the cross section analysis will be carried out in Monte Carlo samples. All steps of the data analysis to determine $\sigma_{t\bar{t}}$ will be reproduced with a few necessary adaptations to the Monte Carlo samples.

In the electron-plus-jets channel the characteristic topology of a $t\bar{t}$ event is one isolated high transverse momentum electron, large missing transverse energy and four or more high transverse momentum jets. There are two main background sources which can provide a similar signature as discussed in Section 2.2.3:

1. QCD multijet events with a fake electron and mismeasured $\cancel{E}_T$. This background can be reduced by an efficient suppression of fake electrons as done in the preselection in 5.1, and is evaluated by the matrix method as will be described in 5.2.
2. W+multijet production where the W boson decays into an electron and an electron neutrino. The jets originate mainly from initial and final state radiation. This background can first be evaluated by the Berends scaling discussed in 5.3, and further reduced by the topological cuts introduced in 5.4.

This analysis follows the electron-plus-jets analysis [34] for the Lepton Photon Conference 2003, which was based on p13 data, with a few improvements developed for p14 analyses [35].

5.1 Preselection

A preselection based on the event signature is made to reject background that is grossly different from the $t\bar{t}$ signal. This section describes the cuts and their efficiencies for the signal as well as the efficiency to trigger jets and electrons.

5.1.1 Loose Preselection

The loose preselection applies cuts on the electron candidate, the missing transverse energy and the reconstructed primary vertex, which will be described in detail below. Taking into account all those cuts the overall efficiency for the loose preselection is:

$$\varepsilon_{loose} = (51.2 \pm 0.4_{stat}) \, \%.$$

Electron Candidate

A loose electron with $p_T > 20$ GeV in CC or EC is required. The selection efficiency is determined in the $t\bar{t} \rightarrow e + jets$ Monte Carlo sample and is found to be:

$$\varepsilon_{loosee} = (64.2 \pm 0.4_{stat}) \, \%.$$

The statistical error is a binomial error as in Equation 4.7. In data this efficiency is determined using the data method in a $Z \rightarrow e^+e^-$ sample, which requires reweighting as demonstrated in 4.1.2. The systematic uncertainty due to using the efficiency derived from the same $t\bar{t}$ Monte Carlo sample will be discussed later in 5.6.4.

To prevent an overlap with the analysis in the di-electron channel a second electron veto is applied. This cut has an efficiency of:

$$\varepsilon_{seceveto} = (99.2 \pm 0.1_{stat}) \, \%.$$

This efficiency is determined using $t\bar{t}$ Monte Carlo events even for the real data analysis.

Missing Transverse Energy

The cuts on the missing transverse energy are $\cancel{E}_T > 20$ GeV and $\Delta\varphi(\cancel{E}_T, e) > 0.5$. The latter cut insures that the electron is not faked or its p_T mismeasured or that it is part of a not reconstructed jet, i.e. isolated. In this analysis as well as for the data analysis their efficiency was determined using $t\bar{t}$ Monte Carlo and found to be:

$$\varepsilon_{MET} = (82.1 \pm 0.4_{stat}) \, \%.$$

Primary Vertex

The primary vertex is required to be reconstructed with at least 3 tracks and its z-coordinate $|z_{PV}| < 60$ cm. The efficiency is again determined using $t\bar{t}$ Monte Carlo:

$$\varepsilon_{PV} = (98.1 \pm 0.2_{stat}) \, \%.$$

5.1.2 Tight Preselection

In addition to all loose preselection requirements the tight preselection demands that the electron found passes the likelihood cut at 0.8. The efficiency was determined in $t\bar{t}$ Monte Carlo events:

$$\varepsilon_{tight} = (89.5 \pm 0.3_{stat}) \text{ \%}.$$

Again this efficiency can also be determined using $Z \rightarrow e^+e^-$ data, which will be explored in 5.6.4.

Because of the expected event signature with at least four jets the analysis will concentrate on those events. The fraction of $t\bar{t}$ Monte Carlo events with four or more reconstructed jets is:

$$\varepsilon_{4jets} = (58.6 \pm 0.6_{stat}) \text{ \%}.$$

5.1.3 Trigger Efficiencies

Although in Monte Carlo there are no triggers involved, some of the triggers have direct influence on the analysis, which is the reason for this section. The trigger used in the e+jets analysis for the Lepton Photon Conference 2003 is *EM15_2JT15*, which is part of trigger list versions 9-11. Basically it means that for each event one EM object with $p_T > 15$ GeV and 2 jets with $p_T > 15$ GeV each are required. Details will be discussed in the following.

Electron Trigger

The electron trigger consists of *CEM(1,10)* on level 1, *EM(.85,10.)* on level 2 and *ELE_SHT(1,15.)* on level 3, which means that on level 1 a cluster with $p_T > 10$ GeV is required, on level 2 in addition an EM fraction > 0.85 and on level 3 $p_T > 15$ GeV and the shower shape of the electron candidate is also considered. The turn-on curve for the electron trigger *EM15_2JT15* on all levels together is shown in Figure 5.1. It can be seen that for electrons with $p_T > 20$ GeV, which are the ones used in the analysis, the efficiency is essentially in saturation at 93%. This number could be multiplied to the total efficiency in this Monte Carlo analysis, but in the calculation of the $t\bar{t}$ cross section a division by the total efficiency is involved, so that this number will cancel. For this reason the electron trigger efficiency is not further included in this Monte Carlo analysis.

Jet Trigger

The jet trigger used on level 1 is *CJT(2,5)*, meaning that two calorimeter towers with $p_T > 5$ GeV are required. The level 2 trigger is *2JET(10.)*, requiring two jets with

Figure 5.1: EM trigger efficiency of the *EM15_2JT15* trigger on L1+L2+L3 in CC [34].

$p_T > 10$ GeV. On Level 3 *JET(2,15.,3.)* demands two jets with $p_T > 15$ GeV and $|\eta| < 3.0$. Only those calorimeter cells are used for reconstruction which show a signal that is above the noise pedestal distribution by 2.5σ . This is also called “applying a zero suppression limit of 2.5σ ”.

The jet triggers have influence on the analysis, because events with more jets will have a higher probability to be triggered. This effect is called trigger bias and will be simulated for this analysis. Because the electron will fulfill the requirements of the jet trigger it is sufficient if only one other jet also fires the jet trigger. Since the parametrization of the electron+jets channel trigger turn-on curves were not available, the *CJT(1,5)* on level 1 and the *JET(1,20.,3.)* on level 3 were used instead. Their efficiency turn-on curves are determined in data as described in Section 4.2.2 and are shown in Figures 5.2 and 5.3.

To simulate the jet triggers on level 1 and 3, each event is assigned a certain weight reflecting the probability of being triggered. This weight depends on the jets’ p_T and η , but in this analysis only the p_T -dependence was considered, because the parametrization in η was not available. For one event with N jets with their respective $p_{T,i}$ on, for example, trigger level 1 the weight of this event is:

$$w_{L1} = 1 - \prod_{i=1}^N (1 - \varepsilon_i(p_{T,i})). \quad (5.1)$$

Taking L1 and L3 triggers together the total weight by the jet triggers for each event is given by:

$$w = w_{L1} \cdot w_{L3}. \quad (5.2)$$

The jet trigger efficiency on the complete sample then is:

$$\varepsilon_{trig} = \frac{\text{weighed number of events with weight } w}{\text{unweighted number of events}}. \quad (5.3)$$

Figure 5.2: $CJT(1,5)$ trigger efficiency turn-on curve as a function of jet E_T [36].

Figure 5.3: $JET(1,20.,3.)$ trigger efficiency turn-on curve as a function of jet E_T [36].

The error on this efficiency is systematic only, and is determined in the following way. The jet trigger weight for each event is recalculated, but this time with p_T of each jet being 1 GeV off, thus simulating a mismeasurement of p_T :

$$w_{\pm} = 1 - \prod_{i=1}^N (1 - \varepsilon_i(p_{T,i} \pm 1 \text{ GeV})). \quad (5.4)$$

The corresponding jet trigger efficiencies on the whole sample are also recalculated:

$$\varepsilon_{\pm} = \frac{\text{weighed number of events with weight } w_{\pm}}{\text{unweighted number of events}}, \quad (5.5)$$

giving the systematic error:

$$\Delta_{sys}(\varepsilon_{trig}) = \frac{\varepsilon_+ - \varepsilon_-}{2}. \quad (5.6)$$

The result for $t\bar{t}$ signal events with four or more jets which pass the tight preselection is:

$$\varepsilon_{trig} = (99.1 \pm 0.1_{sys}) \text{ \%}.$$

For the analysis the same efficiency is also needed for W+4jets events that survive the tight preselection:

$$\varepsilon_{trig-W4jets} = (92.5 \pm 1.0_{sys}) \text{ \%}.$$

5.1.4 Summary

A summary of all the cuts and efficiencies in the $t\bar{t}$ signal sample is given in Table 5.1. The total efficiency of the selection in this analysis is $(14.3 \pm 0.3) \text{ \%}$, which is a considerable increase in efficiency compared to the signal efficiency of $(5.5 \pm 0.3_{stat}) \text{ \%}$ in the Lepton Photon Conference 2003 analysis [34]. The trigger efficiency and the efficiency of the topological cuts, which will be explained in Section 5.4, in the W+4jets sample is listed in Table 5.2.

5.2 QCD Background and Matrix Method

As already mentioned the statistics of fake electrons in the available QCD Monte Carlo sample is too low, so that for this analysis the step of removing the QCD background is only discussed in theory.

To evaluate the sum of W+jets and $t\bar{t}$ events ($N^{W+t\bar{t}}$) and to separate them from QCD multijet events faking the electron, the number of events after the loose preselection (N_{loose}) and after the tight preselection (N_{tight}) are used. The latter is a subsample of the first, the criterion being the likelihood cut on the electron at 0.8. With ε_{sig} being

$\varepsilon_{t\bar{t}}$	Selection Cut	Efficiency
ε_{loosee}	Loose electron in CC or EC with $p_T > 20$ GeV	$(64.2 \pm 0.4) \%$
$\varepsilon_{sece veto}$	Second electron veto	$(99.2 \pm 0.1) \%$
ε_{MET}	$\cancel{E}_T > 20$ GeV, $\Delta\varphi(\cancel{E}_T, e) > 0.5$	$(82.1 \pm 0.4) \%$
ε_{PV}	Primary vertex with ≥ 3 tracks and $ z < 60$ cm	$(98.1 \pm 0.2) \%$
ε_{tight}	Electron likelihood > 0.8	$(89.5 \pm 0.3) \%$
ε_{4jets}	$N_{jet} \geq 4$	$(58.6 \pm 0.6) \%$
ε_{trig}	Trigger efficiency for $N_{jet} \geq 4$ (sys. error)	$(99.1 \pm 0.1) \%$
ε_{topo}	Topological cuts	$(53.6 \pm 1.1) \%$
ε_{total}	Total	$(14.3 \pm 0.3) \%$

Table 5.1: Summary of the efficiencies for the $t\bar{t}$ signal sample with their statistical errors. The topological cuts will be discussed in Section 5.4.

$\varepsilon_{W+4jets}$	Selection Cut	Efficiency
$\varepsilon_{trig-W4jets}$	Trigger efficiency for $N_{jet} \geq 4$ (sys. error)	$(92.5 \pm 1.0) \%$
$\varepsilon_{topo-W4jets}$	Topological cuts (stat. error)	$(12.2 \pm 3.0) \%$

Table 5.2: Summary of the efficiencies for the W+4jets sample with their errors. The topological cuts will be discussed in Section 5.4.

the efficiency of the likelihood cut on real electrons in a W+jets and $t\bar{t}$ sample and ε_{QCD} being the same efficiency on fake electrons in a QCD sample the following set of equations holds:

$$\begin{aligned}
 N_{loose} &= N^{W+t\bar{t}} + N^{QCD} \\
 N_{tight} &= \varepsilon_{sig} \cdot N^{W+t\bar{t}} + \varepsilon_{QCD} \cdot N^{QCD}.
 \end{aligned} \tag{5.7}$$

This set of equations can also be written using a matrix, which is the reason why this evaluation of the QCD background is called “matrix method”. Solving the equations gives:

$$\begin{aligned}
 N^{W+t\bar{t}} &= \frac{N_{tight} - \varepsilon_{QCD} \cdot N_{loose}}{\varepsilon_{sig} - \varepsilon_{QCD}} \quad \text{and} \\
 N^{QCD} &= \frac{\varepsilon_{sig} \cdot N_{loose} - N_{tight}}{\varepsilon_{sig} - \varepsilon_{QCD}}.
 \end{aligned} \tag{5.8}$$

The efficiencies of the likelihood cut are determined as follows. For ε_{sig} a $Z \rightarrow e^+e^-$ data sample is used as described in Section 4.1.2. This includes either reweighting the p_T and η -distributions of the $Z \rightarrow e^+e^-$ sample or applying a correction factor

$$\frac{\varepsilon_{t\bar{t}}^{MC}}{\varepsilon_{Z \rightarrow ee}^{MC}}.$$

To determine ε_{QCD} a data sample enriched in QCD events by not applying the $\cancel{E}_T$ cuts of the preselection is used. The efficiency of the likelihood cut is then examined as a function of $\cancel{E}_T$ and the jet multiplicity. In the Lepton Photon Conference 2003 analysis it was found that ε_{QCD} is independent of the jet multiplicity and also flat in the low $\cancel{E}_T$ region, so that one consistent value could be used.

The number of W +jets and $t\bar{t}$ events is determined for each jet multiplicity bin. As a demonstration of how well the method works, Figure 5.4 shows the distribution of the transverse W boson mass in $t\bar{t} \rightarrow e + jets$ data after the tight preselection for different jet multiplicities.

Figure 5.4: Transverse W boson mass in different jet multiplicity bins for a data sample after the tight preselection. The QCD background as predicted by the Matrix Method is shown as shaded histogram [34].

5.3 W+Jets Background and Berends Scaling

After removing the QCD background the remaining W+jets background has to be separated from the $t\bar{t}$ signal. The method to achieve this is based on Berend's empirical scaling law, which says that since jets in electroweak W+jet events are essentially due to initial state radiation of gluons, the probability for one additional jet is equal to a constant α (in leading order this constant α is the strong coupling constant α_S):

$$\frac{\sigma(W + (N + 1)jets)}{\sigma(W + Njets)} = \alpha, \quad (5.9)$$

where N is the inclusive jet multiplicity. For the number of W+jet and $t\bar{t}$ events with i or more jets $N_i^{W+t\bar{t}}$ this gives:

$$N_i^{W+t\bar{t}} = \alpha^{i-1} \cdot N_1^W + f_i^{t\bar{t}} \cdot N^{t\bar{t}}, \quad (5.10)$$

where N_1^W is the number of W+jet events with one or more jets, $f_i^{t\bar{t}}$ is the fraction of $t\bar{t}$ events with i or more jets, and $N^{t\bar{t}}$ is the total number of $t\bar{t}$ events.

In this Monte Carlo analysis there are two steps to be made, which are not required in the data analysis. These two steps are the selection and normalization of the W+jet Monte Carlo sample to obtain a correct inclusive jet multiplicity distribution in the Monte Carlo W+jet events. The reasons and the procedure will be discussed before describing the influence of the trigger bias and the Berends scaling itself.

5.3.1 W+Njets Event Selection

The generators used for producing the W+jet Monte Carlo samples are ALPGEN for the calculation of the matrix element and the phase space and PYTHIA for the initial and final state radiation and parton showering. The easiest way to obtain a consistent W+Njet background would be to take a W+1jet sample and have PYTHIA take care of generating different number of jets via radiation and parton showering. There are two reasons why this cannot be done. The physical reason is that the cross section of W+Njet events decreases exponentially by the strong coupling constant α with each additional jet, so that the statistics in the W+4jet bin would be much too low. The technical reason is that the parton shower jets generated by PYTHIA tend to be too soft in p_T . Instead, separate W+1jet, W+2jet, W+3jet and W+4jet events were generated by ALPGEN, meaning that for each jet multiplicity an explicit matrix element calculation was made. These files need to be combined avoiding double counting of events, which could occur since jets can either originate from partons from the matrix element calculation or from parton showering. To be able to normalize them correctly, it must be verified that a W+Njet event has exactly N partons (or $\geq N$ for N=4) generated and N jets (or $\geq N$ for N=4) reconstructed.

Partons with the following cuts were generated by ALPGEN:

- Transverse momentum of parton $p_T > 12$ GeV.
- Pseudo-rapidity $|\eta| < 2.7$.
- Distance between two generated partons in (φ, η) -plane $\Delta R > 0.4$.

So the first step has to be checking if exactly N partons (or $\geq N$ for N=4), which pass the cuts, were actually produced at the same parton level vertex. The second step is verifying that exactly N good jets (or $\geq N$ for N=4) were reconstructed, a good jet being a jet with transverse momentum $p_T > 15$ GeV and pseudo-rapidity $|\eta| < 2.5$. In addition, the jets have to be matched to the partons, requiring a combination of parton-jet-pairs with each parton having a different jet closer than $\Delta R \leq 0.5$. Only after this jet-parton matching the loose and tight preselections are applied. A summary of the remaining number of events after the described cuts is given in Table 5.3.

Sample	number of events	parton cut	jet cut	jet-parton matching	loose preselection	tight preselection
$t\bar{t}$	15,156	-	-	-	7,764	6,947
W+1jet	20,599	19,634	10,851	9,475	6,006	5,406
W+2jets	20,485	17,772	7,260	5,396	3,341	2,975
W+3jets	18,756	17,570	5,373	3,203	1,877	1,678
W+ ≥ 4 jets	19,363	18,231	4,664	2,256	1,182	1,046

Table 5.3: Summary of the number of events passing selection cuts for the different Monte Carlo samples used in this analysis.

5.3.2 Normalization

Both the $t\bar{t}$ and W+Njet samples have to be normalized with respect to each other as well as to a certain integrated luminosity. Because for the spring conferences 2004 an integrated luminosity of 210 pb^{-1} of analyzed data is expected, the total normalization in this analysis will conform to this amount of data.

For the $t\bar{t}$ sample the weight can easily be calculated by:

$$w_{t\bar{t}} = \frac{\int \mathcal{L} dt \cdot BR \cdot \sigma_{t\bar{t}}}{N_{t\bar{t}}}, \quad (5.11)$$

where $\int \mathcal{L} dt = 210 \text{ pb}^{-1}$ is the integrated luminosity, $BR = (14.57 \pm 0.23) \%$ (see Section 2.2.2) the branching ratio of $t\bar{t} \rightarrow e + jets$, $\sigma_{t\bar{t}}$ the cross section of $t\bar{t}$ production and $N_{t\bar{t}} = 15,156$ the number of $t\bar{t}$ events before all preselections in the given sample. For the cross section $\sigma_{t\bar{t}} = 7 \text{ pb}$ was initially chosen. All values and the resulting weight are summarized in Table 5.4.

$\int \mathcal{L} dt$	210 pb ⁻¹
BR	14.57 %
$\sigma_{t\bar{t}}$	7 pb
$N_{t\bar{t}}$	15,156
$w_{t\bar{t}}$	0.01413

Table 5.4: Values used for the normalization of the $t\bar{t}$ sample and the final weight for an integrated luminosity of $\int \mathcal{L} dt = 210 \text{ pb}^{-1}$.

The normalization in the W+Njet samples could also be done by considering their cross sections, but the applied parton cuts make it difficult to assign the correct cross sections. Instead, the W+jet events, that pass the tight preselection, are normalized to the Berends scaling fit performed in data in [34], which were acquired with 92 pb⁻¹. The fit function as given in Equation 5.10 reads:

$$N_i^{W+t\bar{t}} = \alpha^{i-1} \cdot N_1^W + f_i^{t\bar{t}} \cdot N^{t\bar{t}},$$

where N_1^W is the number of W+jet events with one or more jets, $f_i^{t\bar{t}}$ is the fraction of $t\bar{t}$ events with i or more jets, and $N^{t\bar{t}}$ is the total number of $t\bar{t}$ events. The $t\bar{t}$ event fraction $f_i^{t\bar{t}}$ is determined in $t\bar{t}$ Monte Carlo in this analysis as well as in the data analysis. The fit to the unbiased inclusive jet multiplicity distribution in data gave $N_1^W = 9011 \pm 790$ W+jet events with one or more jets and $\alpha = 0.168 \pm 0.019$ [34]. With N_i^W being the number of events with i or more jets in the fitted data, N_j the number of tight events in the W+jjet sample and w_j the corresponding weight for that sample, the following equations hold:

$$\begin{aligned}
 N_1^W &= w_1 N_1 + w_2 N_2 + w_3 N_3 + w_4 N_4 = w_1 N_1 + \alpha N_1^W \\
 N_2^W = \alpha N_1^W &= w_2 N_2 + w_3 N_3 + w_4 N_4 = w_2 N_2 + \alpha^2 N_1^W \\
 N_3^W = \alpha^2 N_1^W &= w_3 N_3 + w_4 N_4 = w_3 N_3 + \alpha^3 N_1^W \\
 N_4^W = \alpha^3 N_1^W &= w_4 N_4 \quad .
 \end{aligned} \tag{5.12}$$

Solving this system of equations gives the weights:

$$\begin{aligned}
 w_1 &= (1 - \alpha) \cdot \frac{N_1^W}{N_1} \\
 w_2 &= (1 - \alpha) \cdot \alpha \frac{N_1^W}{N_2} \\
 w_3 &= (1 - \alpha) \cdot \alpha^2 \frac{N_1^W}{N_3} \\
 w_4 &= \alpha^3 \frac{N_1^W}{N_4} .
 \end{aligned} \tag{5.13}$$

These weights have to be multiplied by $(\int \mathcal{L} dt / 92 \text{ pb}^{-1})$ to correspond to the desired luminosity of $\int \mathcal{L} dt = 210 \text{ pb}^{-1}$. The values used for the normalization of the W+jet samples and their final weights are summarized in Table 5.5.

N_1^W	9011
α	0.168
N_1	5,406
N_2	2,975
N_3	1,678
N_4	1,046
$\int \mathcal{L} dt$	210 pb ⁻¹
w_1	3.166
w_2	0.9664
w_3	0.2878
w_4	0.09324

Table 5.5: Values used for the normalization of the W+jet samples and their final weights for an integrated luminosity of $\int \mathcal{L} dt = 210 \text{ pb}^{-1}$.

The obtained normalized $t\bar{t}$ and W+jet samples can now be filled into the inclusive jet multiplicity distribution, which is shown in Figure 5.5.

Figure 5.5: Inclusive jet multiplicity distribution for W+jets samples only and for W+jets samples plus the $t\bar{t}$ sample.

5.3.3 Trigger Bias

In principal the just acquired jet multiplicity distribution is exactly the one that should be fitted with Formula 5.10, but in the data analysis there is one more step to be made, which will be discussed in this section. This step involves the trigger bias by the jet triggers. To simulate the distribution that would be acquired in data, weights according to the trigger efficiency in each event are calculated as described in Section 5.1.3 and the resulting inclusive jet multiplicity distribution is shown in Figure 5.6. In this biased

Figure 5.6: Inclusive jet multiplicity distribution for W +jets samples only and for W +jets samples plus the $t\bar{t}$ sample with trigger bias.

distribution each bin i with a content of $\tilde{N}_i$ is assigned the Poisson error of $\sqrt{\tilde{N}_i}$, because this is the distribution that is observed in the data.

The trigger bias for each bin is evaluated by comparing the unbiased distribution to the biased one:

$$\varepsilon_{trigbias,i} = \frac{\tilde{N}_i}{N_i}, \quad (5.14)$$

with $\tilde{N}_i$ being the number of events in the i -th bin of the biased distribution and N_i the number of events in the i -th bin of the unbiased distribution. The error of the trigger bias is given by the binomial error:

$$\Delta\varepsilon_{trigbias,i} = \sqrt{\frac{\varepsilon_{trigbias,i} \cdot (1 - \varepsilon_{trigbias,i})}{N_i}}, \quad (5.15)$$

and depends on the integrated luminosity, since the unbiased distribution was normalized to it. The trigger bias for each bin is listed in Table 5.6. The table also lists the fraction of $t\bar{t}$ events compared to the total number of $t\bar{t}$ events in each bin, which were determined in Monte Carlo. Those will be needed for the Berends scaling later. In the data analysis

bin	trigger bias $\varepsilon_{trigbias,i}$	top fraction $f_i^{t\bar{t}}$
≥ 1	$(40.8 \pm 0.3) \%$	$(100.0 \pm 0.0) \%$
≥ 2	$(70.4 \pm 0.8) \%$	$(99.3 \pm 0.1) \%$
≥ 3	$(87.4 \pm 1.3) \%$	$(91.3 \pm 0.3) \%$
≥ 4	$(95.0 \pm 1.8) \%$	$(58.6 \pm 0.6) \%$

Table 5.6: Trigger bias and top fraction for each inclusive jet multiplicity bin.

the trigger bias correction is evaluated by comparing the distribution to the one obtained

by an unbiased trigger, namely the single-electron trigger, whereas the top fraction is also determined in a Monte Carlo sample.

Using the calculated trigger bias for each bin, the biased distribution is corrected back to the unbiased distribution, and the error on the i -th bin is simply obtained by error propagation of the trigger bias error and the error on the bin of the biased distribution. By construction this method gives the same unbiased distribution back, and the systematic error caused by this correlation of samples will be discussed in Section 5.6.3.

5.3.4 Berends Scaling

After simulating the trigger bias and correcting the distribution back to an unbiased one, the Berends scaling can now be carried out by fitting Formula 5.10

$$N_i^{W+t\bar{t}} = \alpha^{i-1} \cdot N_1^W + f_i^{t\bar{t}} \cdot N^{t\bar{t}},$$

where N_1^W is the number of W+jet events with one or more jets, $f_i^{t\bar{t}}$ is the fraction of $t\bar{t}$ events with i or more jets, and $N^{t\bar{t}}$ is the total number of $t\bar{t}$ events. The fit is shown in Figure 5.7. The results of the fit for α , N_1^W and $N^{t\bar{t}}$ are listed in Table 5.7. It also lists

Figure 5.7: Berends scaling. The hatched histogram shows the inclusive jet multiplicity distribution without the $t\bar{t}$ sample, the horizontal lines show the distribution including the $t\bar{t}$ sample with the fitted function going through.

the expected number of W+jets events in the fourth bin:

$$N_4^W = \alpha^3 \cdot N_1^W. \quad (5.16)$$

α	0.168 ± 0.005
N_1^W	$20,568.1 \pm 277.6$
$N^{t\bar{t}}$	98.2 ± 29.2
N_4^W	97.5 ± 8.5

Table 5.7: Results of the Berends scaling fit.

Because the inclusive jet multiplicity distribution was normalized to the result of the fit performed in data, the Berends scaling is fulfilled perfectly and the errors of the fit parameters are most likely underestimated. This will also be explored in the discussion of the systematic uncertainties due to the correlation of samples in Section 5.6.3. In the data analysis $N^{t\bar{t}}$ has such a huge error that the number does not have any significance.

5.4 Topological Cuts

This section describes the topological selection based on the event topology which suppresses the background even more. In the $e + jets$ analysis the following topological cuts developed in Run I are applied:

- The absolute value of the pseudo-rapidity of the reconstructed leptonic W boson

$$|\eta_W| < 2,$$

because electroweak W bosons are expected to be more in the forward region due to their preferred left-handed polarization. To reconstruct the leptonic W boson fully, the z -component of the neutrino momentum $p_z(\nu_e)$ has to be determined which is done under the assumption that the neutrino and the electron come from the decay of a particle with mass M_W , and that of the two possible results for $p_z(\nu_e)$ the one with the lesser absolute value is the correct one.

- The sum of the transverse momentum of the electron and the missing transverse energy

$$p_T(e) + \cancel{E}_T > 60 \text{ GeV},$$

which accounts for the W bosons having more transverse energy in $t\bar{t}$ events, as they are products of the top quark decay.

- For the following topological variables only good jets are considered, a good jet being a jet with $p_T > 15 \text{ GeV}$ and $|\eta| < 2.5$. Because the jets in the electroweak W +jet events originating from initial state radiation are expected to be very soft, the cut requires the sum of all good jet p_T to be:

$$H_T := \sum p_T(\text{good jet}) > 180 \text{ GeV}.$$

- Because of the exceptionally high mass of the top quarks produced, the phase space is limited and the top quarks are almost at rest. This results in very spherical events, while the background events are more planar because their jets originate from gluon radiation. A measure for the flatness of an event is the aplanarity $\mathcal{A}$, which is defined to be $\frac{3}{2}$ times the smallest eigenvalue of the normalized momentum tensor $\mathcal{M}$ of all good jets and the leptonic W:

$$\mathcal{M}_{ij} = \frac{\sum_o p_i^o p_j^o}{\sum_o |\vec{p}^o|^2}, \quad (5.17)$$

with $\vec{p}^o$ being the momentum vector of an object o and i and j the indices for cartesian coordinates. Large values of $\mathcal{A}$ mean more spherical events and small values more planar events, so the required cut is:

$$\mathcal{A} > 0.065.$$

The distributions of the topological variables in $t\bar{t}$ and W+jet events with four or more jets after the tight preselection and the cuts applied can be seen in Figure 5.8. The cut on $p_T(e) + \cancel{E}_T$ does not seem to make much sense at first, but it has to be considered that the QCD background is not included here. The efficiency of the topological cuts in this analysis as well as in the data analysis is determined in Monte Carlo samples, the errors being binomial errors of weighted events as for example in Equation 4.8. The results were already listed in Table 5.1 and 5.2, and are again summarized in Table 5.8.

Sample	ε_{topo}
$t\bar{t}$	$(53.6 \pm 1.1) \%$
W+4jets	$(12.2 \pm 3.0) \%$

Table 5.8: Efficiency of the topological cuts for the $t\bar{t}$ and the W+4jets sample.

Table 5.9 gives a summary of the number of events with four or more reconstructed jets at the last few stages of the analysis. The first column describes the stage, the second one gives the number of observed events $N_{obs} = N_4^{W+t\bar{t}}$ and the third one the number of expected W+jet background events $N_{bkg} = N_4^W$, which is determined as described in the following. The number of expected W+jet events in the fourth bin of the unbiased distribution is given by the results of the Berends scaling:

$$N_4^W = \alpha^3 \cdot N_1^W. \quad (5.18)$$

Applying the trigger bias $\varepsilon_{trig_W4jets}$ from Table 5.2 gives the number of expected events in the biased distribution:

$$\tilde{N}_4^W = \varepsilon_{trig_W4jets} \cdot N_4^W, \quad (5.19)$$

Finally the efficiency of the topological cut on W+4jet events from Table 5.8 is used:

$$N_{bkg} = \varepsilon_{topo_W4jets} \cdot \tilde{N}_4^W. \quad (5.20)$$

Figure 5.8: Topological variables in $t\bar{t}$ and $W + \text{jet}$ events after tight preselection. The cuts are shown as vertical lines.

Stage	N_{obs}	N_{bkg}
Unbiased distribution	155.1 ± 12.5	97.5 ± 8.5
Biased distribution	147.3 ± 12.1	90.3 ± 7.9
After topological cuts	41.6 ± 6.4	11.0 ± 2.9

Table 5.9: Number of events with four or more reconstructed jets at different stages of the analysis.

5.5 $t\bar{t}$ Cross Section Results

Having gathered all necessary information at this point, the cross section can be calculated. The $t\bar{t}$ cross section is given by:

$$\sigma_{t\bar{t}} = \frac{N_{obs} - N_{bkg}}{\int \mathcal{L} dt \cdot BR \cdot \varepsilon_{tot}}, \quad (5.21)$$

where N_{obs} is the number of observed events after topological cuts, N_{bkg} the number of expected W+jet background events after topological cuts, $\int \mathcal{L} dt$ the integrated luminosity, BR the branching fraction to the electron+jet final state and ε_{tot} is the total efficiency including the detector acceptance and trigger bias and described in Section 5.1.

The statistical error of the cross section is given by the errors on the number of observed and expected background events and the error on the total efficiency:

$$\Delta_{stat}(\sigma_{t\bar{t}}) = \sqrt{\left(\frac{\Delta(N_{obs})}{\int \mathcal{L} dt \cdot BR \cdot \varepsilon_{tot}}\right)^2 + \left(\frac{\Delta(N_{bkg})}{\int \mathcal{L} dt \cdot BR \cdot \varepsilon_{tot}}\right)^2 + \left(\frac{\Delta(\varepsilon_{tot}) \cdot (N_{obs} - N_{bkg})}{\int \mathcal{L} dt \cdot BR \cdot \varepsilon_{tot}}\right)^2}. \quad (5.22)$$

The luminosity measurement has an error of 10% which also contributes to the error of the cross section:

$$\begin{aligned} \Delta_{Lumi}(\sigma_{t\bar{t}}) &= \frac{\Delta(\int \mathcal{L} dt) \cdot (N_{obs} - N_{bkg})}{(\int \mathcal{L} dt)^2 \cdot BR \cdot \varepsilon_{tot}} \\ &= \frac{0.1 \cdot \int \mathcal{L} dt \cdot (N_{obs} - N_{bkg})}{(\int \mathcal{L} dt)^2 \cdot BR \cdot \varepsilon_{tot}} \\ &= 0.1 \cdot \sigma_{t\bar{t}}. \end{aligned} \quad (5.23)$$

Systematic uncertainties will be discussed in the following Section 5.6. Without the systematic error the Monte Carlo analysis with a generated cross section of $\sigma_{t\bar{t},gen} = 7$ pb gives the following result:

$$\sigma_{t\bar{t}} = (7.00 \pm 1.63_{stat} \pm 0.70_{Lumi}) \text{ pb.}$$

Generating 13 different cross sections for $t\bar{t}$ production between 4 pb and 10 pb in 0.5 pb steps gives the results shown in Table 5.10 and Figure 5.9. The exact values obtained for the cross sections are expected because of the full correlation of the samples used for the efficiency determination and the cross section analysis itself. The effect of using uncorrelated samples instead is examined in the systematics in 5.6.3.

The measured cross sections show no tendency to systematically deviate from the generated value in either direction, which shows that this analysis is not biased.

$\sigma_{t\bar{t},gen}$ [pb]	$\sigma_{t\bar{t}}$ [pb]	$\Delta_{stat}(\sigma_{t\bar{t}})$ [pb]
4	4.00	1.39
4.5	4.50	1.43
5	5.00	1.47
5.5	5.50	1.51
6	6.00	1.55
6.5	6.50	1.59
7	7.00	1.63
7.5	7.50	1.66
8	8.00	1.70
8.5	8.50	1.73
9	9.00	1.77
9.5	9.50	1.80
10	10.00	1.83

Table 5.10: Generated and measured $t\bar{t}$ cross sections for 13 different generated cross sections with their statistical errors. Exact values are expected because of the full correlation.

Figure 5.9: Measured $t\bar{t}$ cross section versus generated $t\bar{t}$ cross section. Errors shown are statistical only.

5.6 Systematics

The systematic uncertainty of the cross section is dominated by the uncertainty of the jet energy scale. Further sources include the Monte Carlo sample modeling, the correlation of the samples used for the determination of the efficiencies and the analysis and using electron identification efficiencies acquired in the $t\bar{t}$ sample rather than an independent $Z \rightarrow e^+e^-$ sample. In this section each of these sources will be discussed and their impact on the cross section will be measured with a generated $t\bar{t}$ cross section of 7 pb and an integrated luminosity of 210 pb⁻¹.

5.6.1 Jet Energy Scale

To determine the effect of the jet energy scale uncertainty, the jet energy scale is varied by

$$\pm\Delta\text{JES} = \sqrt{(\Delta_{sys}\text{JES})^2 + (\Delta_{stat}\text{JES})^2}$$

and all connected variables are recalculated. The efficiencies and the triggerbias are determined in the varied samples. Then the complete analysis is repeated on the unvaried samples and the results are summarized in Table 5.11. For a generated cross section of

Jet energy scale:	$-\Delta\text{JES}$	central value	$+\Delta\text{JES}$
<i>$t\bar{t}$ efficiencies</i>			
Loose preselection	$(51.2 \pm 0.4) \%$	$(51.2 \pm 0.4) \%$	$(51.2 \pm 0.4) \%$
Tight preselection	$(89.4 \pm 0.3) \%$	$(89.5 \pm 0.3) \%$	$(89.5 \pm 0.3) \%$
Top fraction in 4th bin	$(56.0 \pm 0.6) \%$	$(58.6 \pm 0.6) \%$	$(60.8 \pm 0.6) \%$
Trigger eff. in 4th bin	$(98.9 \pm 0.1) \%$	$(99.1 \pm 0.1) \%$	$(99.2 \pm 0.1) \%$
Topological cuts	$(50.9 \pm 1.1) \%$	$(53.6 \pm 1.1) \%$	$(55.1 \pm 1.0) \%$
<i>$W+4jets$ efficiencies</i>			
Trigger efficiency	$(91.9 \pm 1.0) \%$	$(92.5 \pm 1.0) \%$	$(93.3 \pm 0.9) \%$
Topological cuts	$(12.2 \pm 3.3) \%$	$(12.2 \pm 3.0) \%$	$(12.3 \pm 2.8) \%$
<i>Trigger bias</i>			
≥ 1	$(39.3 \pm 0.3) \%$	$(40.8 \pm 0.3) \%$	$(42.3 \pm 0.3) \%$
≥ 2	$(68.7 \pm 0.8) \%$	$(70.4 \pm 0.8) \%$	$(72.1 \pm 0.8) \%$
≥ 3	$(86.2 \pm 1.3) \%$	$(87.4 \pm 1.3) \%$	$(88.4 \pm 1.2) \%$
≥ 4	$(94.4 \pm 1.9) \%$	$(95.0 \pm 1.8) \%$	$(95.5 \pm 1.6) \%$
$\sigma_{t\bar{t}}$	$(7.79 \pm 1.82) \text{ pb}$	$(7.00 \pm 1.63) \text{ pb}$	$(6.51 \pm 1.75) \text{ pb}$

Table 5.11: Efficiencies and cross section results using the variation of the jet energy scale with statistical errors with a generated $t\bar{t}$ cross section of 7 pb.

7 pb the jet energy scale uncertainty gives a systematic error on the measured cross section of:

$$\Delta_{JES}(\sigma_{t\bar{t}}) = {}^{+0.79}_{-0.49} \text{ pb.}$$

5.6.2 Monte Carlo Modeling

In Section 5.3.1 W+jet events were selected based on the number of partons that pass certain cuts, the number of good jets and the requirement of jet parton matching. A variation of the conditions affects the trigger efficiency and the topological cut efficiency of the W+jet events and consequently the number of expected background events. To evaluate the impact on the cross section measurement, the cuts are varied and the obtained efficiencies replace the old ones in the calculation of the expected background events in Equations 5.19 and 5.20.

First the required parton cuts are varied. One possibility is to relax the p_T cut from 12 GeV to 8 GeV, another one to drop the requirement of a minimal distance of $\Delta R = 0.4$ between the partons. But both changes do not have much influence on the efficiencies, because the partons were generated with these cuts and loosening them will not gain any more or any different events.

Changing the required number of good jets is not an option, because it is needed to select correct events. So the last parameter to be varied is the definition of the jet-parton matching. By default it was required that each parton can be uniquely matched to a jet that is closer than $\Delta R = 0.5$. This condition is varied in both directions, once not requiring any jet-parton matching at all, and then requiring a stricter jet-parton matching with a matching distance of $\Delta R = 0.1$. The first change brings the efficiencies down, which results in less expected background events and thus in a larger $t\bar{t}$ cross section. The latter change does the opposite, resulting in a smaller $t\bar{t}$ cross section. This can be explained by recalling that a stricter jet-parton matching will select jets with a higher p_T , as can be seen in Figure 5.10, which shows the p_T of all good jets and H_T , the sum of all good jet p_T in one event.

The efficiencies obtained by the variation of the jet-parton matching definition and the resulting $t\bar{t}$ cross sections are summarized in Table 5.12. The systematic uncertainty on the cross section resulting from the Monte Carlo modeling for a generated cross section of 7 pb is therefore:

$$\Delta_{MC}(\sigma_{t\bar{t}}) = {}^{+0.46}_{-0.69} \text{ pb.}$$

5.6.3 Correlation of Samples

In this Monte Carlo analysis the efficiencies were determined in the same samples that were also used for the analysis itself due to limited Monte Carlo statistics. In a more

Figure 5.10: The upper plot shows the p_T of all good jets in all events for different jet-parton matching definitions. The lower plot shows H_T for different jet-parton definitions.

Jet parton matching:	no	$\Delta R = 0.5$	$\Delta R = 0.1$
<i>W+4jets efficiencies</i>			
Trigger efficiency	$(90.3 \pm 1.2) \%$	$(92.5 \pm 1.0) \%$	$(96.8 \pm 0.4) \%$
Topological cuts	$(10.2 \pm 2.3) \%$	$(12.2 \pm 3.0) \%$	$(14.9 \pm 9.2) \%$
$\sigma_{t\bar{t}}$	$(7.46 \pm 1.57) \text{ pb}$	$(7.00 \pm 1.63) \text{ pb}$	$(6.31 \pm 2.50) \text{ pb}$

Table 5.12: Efficiencies and cross section results using different jet-parton-matching in the W+jet samples with their statistical errors with a generated $t\bar{t}$ cross section of 7 pb.

correct approach the efficiencies should be measured in independent samples. This is simulated by splitting the samples into two each and determining the efficiencies in one half, while using the other half for the analysis. Because now only half of the statistics is available, the statistical error on the cross section increases. The efficiencies and the result for the $t\bar{t}$ cross section are shown in Table 5.13. Because it is arbitrary which half

Samples:	uncorrelated	correlated
<i>$t\bar{t}$ efficiencies</i>		
Loose preselection	$(51.6 \pm 0.6) \%$	$(51.2 \pm 0.4) \%$
Tight preselection	$(88.9 \pm 0.5) \%$	$(89.5 \pm 0.3) \%$
Top fraction in 4th bin	$(58.8 \pm 0.8) \%$	$(58.6 \pm 0.6) \%$
Trigger eff. in 4th bin	$(99.1 \pm 0.1) \%$	$(99.1 \pm 0.1) \%$
Topological cuts	$(54.1 \pm 1.5) \%$	$(53.6 \pm 1.1) \%$
<i>$W+4jets$ efficiencies</i>		
Trigger efficiency	$(92.6 \pm 1.0) \%$	$(92.5 \pm 1.0) \%$
Topological cuts	$(11.1 \pm 4.3) \%$	$(12.2 \pm 3.0) \%$
<i>Trigger bias</i>		
≥ 1	$(41.0 \pm 0.3) \%$	$(40.8 \pm 0.3) \%$
≥ 2	$(70.4 \pm 0.8) \%$	$(70.4 \pm 0.8) \%$
≥ 3	$(87.7 \pm 1.3) \%$	$(87.4 \pm 1.3) \%$
≥ 4	$(95.0 \pm 1.8) \%$	$(95.0 \pm 1.8) \%$
$\sigma_{t\bar{t}}$	$(7.22 \pm 1.75) \text{ pb}$	$(7.00 \pm 1.63) \text{ pb}$

Table 5.13: Efficiencies and cross section results using uncorrelated and correlated samples with their statistical errors with a generated $t\bar{t}$ cross section of 7 pb.

of the samples is used for determining the efficiencies and which half for being analyzed, the resulting systematic error on the cross section is symmetric and, for a generated cross section of 7 pb, amounts to:

$$\Delta_{corr}(\sigma_{t\bar{t}}) = \pm 0.22 \text{ pb.}$$

5.6.4 Electron Identification Efficiencies

As shown by the studies made in Section 4.1.2, the electron identification efficiencies determined in a $t\bar{t}$ sample with the Monte Carlo method and in a $Z \rightarrow e^+e^-$ sample with the Monte Carlo method as well as with the data method essentially give the same results. For that reason the electron identification efficiencies in the preselection of this analysis were determined in the $t\bar{t}$ sample directly using the Monte Carlo method, although the data analysis uses the data method on a $Z \rightarrow e^+e^-$ data sample. This section will now explore the influence of using the $Z \rightarrow e^+e^-$ sample instead.

In Section 4.1.2 two adjustments were developed to better compare the $t\bar{t}$ and $Z \rightarrow e^+e^-$ sample. It is required that ΔR between the candidate electron and the next jet is greater than 0.9 and the p_T -distribution of the generated electrons in the $Z \rightarrow e^+e^-$ sample is also weighted to the one of the generated electrons from the $t\bar{t}$ sample. Because the cross section analysis does not apply a ΔR -cut to jets, the only adjustment used is the weighting of the p_T distribution.

Four weighted p_T distributions of electrons from the $Z \rightarrow e^+e^-$ sample are obtained. First, the generated electrons, second, loose electrons with $p_T > 20$ GeV in CC or EC, third, loose electrons with $p_T > 20$ GeV in CC or EC with the event passing the remaining loose preselection cuts, and last, tight electrons with $p_T > 20$ GeV in CC or EC with the event passing the remaining loose preselection cuts. The remaining loose preselection cuts actually only mean the quality cuts on the primary vertex, because the $\cancel{E}_T$ cuts and the second electron veto do not apply to the $Z \rightarrow e^+e^-$ sample.

The efficiency for finding the loose electron is then the integral of the second distribution divided by the integral of the first:

$$\varepsilon_{loosee-Z} = \frac{\int p_T(\text{loose electrons, } p_T > 20 \text{ GeV, CC or EC})}{\int p_T(\text{generated electrons})}, \quad (5.24)$$

and the efficiency of the complete loose preselection is acquired by multiplying this efficiency to the efficiency of the second electron veto, the $\cancel{E}_T$ cuts and the primary vertex cuts in the $t\bar{t}$ sample.

The efficiency for the tight preselection is the integral of the fourth distribution divided by the third:

$$\varepsilon_{tight-Z} = \frac{\int p_T(\text{tight electrons, } p_T > 20 \text{ GeV, CC or EC, PV cuts})}{\int p_T(\text{loose electrons, } p_T > 20 \text{ GeV, CC or EC, PV cuts})}. \quad (5.25)$$

The errors on these efficiencies are given by the binomial error considering the weighting as in Equation 4.8.

Table 5.14 lists the preselection efficiencies and the resulting cross sections. It can be seen

EM-ID efficiencies from:	$Z \rightarrow e^+e^-$	$t\bar{t}$
<i>$t\bar{t}$ efficiencies</i>		
Loose preselection	$(66.7 \pm 0.2) \%$	$(51.2 \pm 0.4) \%$
Tight preselection	$(96.1 \pm 0.1) \%$	$(89.5 \pm 0.3) \%$
$\sigma_{t\bar{t}}$	$(6.27 \pm 1.46) \text{ pb}$	$(7.00 \pm 1.63) \text{ pb}$

Table 5.14: Efficiencies and cross section results using $Z \rightarrow e^+e^-$ electron identification efficiencies with their statistical errors with a generated $t\bar{t}$ cross section of 7 pb.

that the electron identification efficiencies are much higher in the $Z \rightarrow e^+e^-$ sample. This was already observed and examined before and is due to the better isolation of electrons in $Z \rightarrow e^+e^-$ events. Therefore the recommended ΔR -cut to jets from Section 4.1.2 should be introduced into the analysis or taken into account by introducing a scale factor

$$\frac{\varepsilon_{t\bar{t}}^{MC}}{\varepsilon_{Z \rightarrow ee}^{MC}}.$$

The determination of the efficiencies in the $Z \rightarrow e^+e^-$ sample gives a systematic error in only one direction, because these efficiencies will always be higher than in a $t\bar{t}$ sample. The systematic error for a generated cross section of 7 pb is:

$$\Delta_Z(\sigma_{t\bar{t}}) = -0.73 \text{ pb.}$$

5.6.5 Summary

The total systematic error on the cross section is calculated as follows:

$$\Delta_{sys}(\sigma_{t\bar{t}}) = \sqrt{\Delta_{JES}^2 + \Delta_{MC}^2 + \Delta_{corr}^2 + \Delta_Z^2}, \quad (5.26)$$

which gives for a generated $t\bar{t}$ cross section of 7 pb:

$$\Delta_{sys}(\sigma_{t\bar{t}}) = {}^{+0.94}_{-1.14} \text{ pb.}$$

The total error on the cross section is then

$$\Delta_{tot}(\sigma_{t\bar{t}}) = \sqrt{\Delta_{stat}^2 + \Delta_{sys}^2}. \quad (5.27)$$

Table 5.15 gives a summary of the systematic uncertainties for a range of generated $t\bar{t}$ cross sections and Figure 5.11 shows the measured cross sections with their statistical and total error.

generated $\sigma_{t\bar{t}}$ [pb]	Δ_{JES} [pb]	Δ_{MC} [pb]	Δ_{corr} [pb]	Δ_Z [pb]	Δ_{sys} [pb]
4	+0.49 -0.28	+0.46 -0.69	± 0.32	-0.42	+0.74 -0.91
4.5	+0.54 -0.31	+0.46 -0.69	± 0.31	-0.47	+0.77 -0.94
5	+0.59 -0.35	+0.46 -0.69	± 0.29	-0.52	+0.80 -0.98
5.5	+0.64 -0.39	+0.46 -0.69	± 0.27	-0.57	+0.83 -1.01
6	+0.69 -0.42	+0.46 -0.69	± 0.26	-0.62	+0.86 -1.05
6.5	+0.74 -0.46	+0.46 -0.69	± 0.24	-0.67	+0.90 -1.10
7	+0.79 -0.49	+0.46 -0.69	± 0.22	-0.73	+0.94 -1.14
7.5	+0.84 -0.53	+0.46 -0.69	± 0.20	-0.78	+0.97 -1.19
8	+0.89 -0.56	+0.46 -0.69	± 0.19	-0.83	+1.01 -1.23
8.5	+0.94 -0.60	+0.46 -0.69	± 0.17	-0.88	+1.05 -1.28
9	+0.99 -0.63	+0.46 -0.69	± 0.15	-0.93	+1.10 -1.33
9.5	+1.03 -0.67	+0.46 -0.69	± 0.13	-0.98	+1.14 -1.38
10	+1.08 -0.71	+0.46 -0.69	± 0.12	-1.04	+1.18 -1.44

Table 5.15: Summary of systematic uncertainties for different sources and generated cross sections.

Figure 5.11: Measured $t\bar{t}$ cross section versus generated $t\bar{t}$ cross section. Inner error bars denote the statistical error, outer bars the total error.

6 Conclusion and Outlook

This study was performed to examine various aspects of the $t\bar{t}$ cross section analysis in the electron-plus-jets channel. It concentrated on comparing methods for electron reconstruction efficiency determination and improving the electron likelihood. The main analysis reproduced the cross section analysis as carried out in data with adaptations to the Monte Carlo samples.

It was shown that it is possible to learn about electrons from $t\bar{t}$ events by studying $Z \rightarrow e^+e^-$ events, when taking into account the differences in the topology, especially the isolation, and by weighting the p_T distributions. Both data method and Monte Carlo method used for the determination of electron identification efficiencies agree very well. This validates using the data method in $Z \rightarrow e^+e^-$ data to determine efficiencies for a $t\bar{t}$ analysis. Further studies could examine how to remove the remaining differences, for example by weighting the η distribution as well.

The Monte Carlo studies of fake electrons in QCD confirmed the composition of the QCD background to electrons mainly being neutral pions. This permitted to contribute to the new $D\bar{O}$ electron likelihood by developing a new input variable tuned to neutral pions, the number of tracks in the direct vicinity of the electron to identify photon conversions and by proposing another variable, the number of tracks in a wider neighborhood to probe the presence of a jet. Furthermore, the Monte Carlo comparison of the proposed CPS variable distribution in $t\bar{t} \rightarrow e + jets$ and $Z \rightarrow e^+e^-$ samples prevented the implementation of a topology-dependent input variable into the new electron likelihood. Additional results for the electron likelihood include the fact that a three-class-likelihood is only helpful in separating charged pions, which, however, do not play a major role in faking electrons and that η -dependant input variables do not enhance the separation power. The p_T -dependant distributions of some input variables showed promising separation between signal and background, but a first examination did not translate this into an improvement of the likelihood performance. This should be studied further by for example only using the p_T -dependence of the most promising input variables.

In the main analysis measuring the $t\bar{t}$ cross section using Monte Carlo samples was studied. All steps necessary for the data analysis were reproduced and additional modifications to adapt the analysis to Monte Carlo samples were carried out. The preselection requirements target the typical event signature of a $t\bar{t} \rightarrow e + jets$ event, one isolated high- p_T electron, large $\cancel{E}_T$ and four or more jets. The overwhelming QCD background is estimated

by the matrix method, which benefits from an accurate electron identification efficiency measurement and the W +multijet background is evaluated by the Berends scaling. Further cuts based on the expected $t\bar{t}$ event topology are applied. Having measured the efficiencies of the cuts in the signal sample, the cross section and the statistical error can then be calculated. The sources for systematic uncertainties were scrutinized, quantifying the influence of the jet energy scale error, the Monte Carlo modeling process, the correlation of the efficiency samples and the analysis samples and the impact of using electron identification efficiencies determined in a $Z \rightarrow e^+e^-$ sample instead of the $t\bar{t}$ sample. The final result obtained with a $t\bar{t}$ input cross section of 7 pb yields

$$\sigma_{t\bar{t}} = (7 \pm 1.63 \text{ (stat)}^{+0.94}_{-1.14} \text{ (sys)}) \text{ pb},$$

where the exact value is obtained by definition and the errors are statistical and systematic, respectively. The measurement was repeated for a range of generated cross sections. The results do not show a bias of any kind and validate the methods of the analysis. The main improvement made compared to the analysis done for the Lepton Photon Conference 2003 is the implementation of the new electron likelihood which gains signal efficiency while suppressing the QCD background even more. This allows to include the EC region, which was formerly too contaminated by QCD background, into the analysis for the first time. Another contribution of this analysis is the suggestion of the reweighting of $Z \rightarrow e^+e^-$ distributions or applying a scale factor not only in the determination of the preselection efficiency but also in the signal efficiency for the matrix method. Further optimization of cuts can and will be done in the data.

A Monte Carlo Samples

Sample	Reconstruction Version	Release Version	top_analyze Version	Number of Events
<i>Efficiency Studies (Section 4.1.2)</i>				
$t\bar{t} \rightarrow e + jets$	p13.05	p13.09.00	v00-02-00	21,000
$Z \rightarrow e^+e^-$	p13.05	p13.09.00	v00-02-00	101,000
<i>Background Studies (Section 4.1.3)</i>				
QCD	p13.05	p14.03.00	v00-04-06	71,000
QCD	p13.08	p14.03.00	v00-04-06	33,000
<i>Likelihood Studies (Section 4.1.4)</i>				
Single electron	p14.03	p14.03.00	v00-04-06	20,100
Single π^0	p14.03	p14.03.00	v00-04-06	84,000
Single $\pi^\pm$	p14.03	p14.03.00	v00-04-06	173,000
<i>Cross Section Measurement (Chapter 5)</i>				
$t\bar{t} \rightarrow e + jets$	p14.02	p14.05.01	v00-04-14	15,156
$W + 1jet$	p14.02	p14.05.01	v00-04-14	20,599
$W + 2jets$	p14.02	p14.05.01	v00-04-14	20,485
$W + 3jets$	p14.02	p14.05.01	v00-04-14	18,756
$W + 4jets$	p14.02	p14.05.01	v00-04-14	19,363
$Z \rightarrow e^+e^-$	p14.03	p14.05.01	v00-04-14	247,000

Bibliography

- [1] F. Abe *et al.* (CDF Collaboration), *Observation of the Top Quark in $p\bar{p}$ Collisions with the Collider Detector at Fermilab*, Phys. Rev. Lett. **74**, 2626 (1995), hep-ex/9503002.
- [2] S. Abachi *et al.* (DØ Collaboration), *Observation of the Top Quark*, Phys. Rev. Lett. **74**, 2632 (1995), hep-ex/9503003.
- [3] F. Abe *et al.* (CDF Collaboration), *Study of $t\bar{t}$ Production in $p\bar{p}$ Collisions Using Total Transverse Energy*, Phys. Rev. Lett. **75**, 3997 (1995), hep-ex/9506006.
- [4] D. Griffiths, *Introduction to Elementary Particles*, Wiley (1987).
- [5] F. Halzen, A. Martin, *Quarks and Leptons*, Wiley (1984).
- [6] S. Willenbrock, *The Standard Model and the Top Quark*, hep-ph/0211067 (2002).
- [7] K. Hagiwara *et al.* (Particle Data Group Collaboration), *Review of Particle Physics*, Phys. Rev. D **66**, 010001 (2002).
- [8] The LEP Collaborations ALEPH, DELPHI, L3 and OPAL, The LEP Electroweak Working Group and the SLD Heavy Flavour Group, *A Combination of Preliminary Electroweak Measurements and Constraints on the Standard Model*, CERN-EP/2003-02, hep-ex/0312023 (2003).
- [9] The LEP Collaborations ALEPH, DELPHI, L3 and OPAL, *Electroweak parameters of the Z^0 resonance and the Standard Model*, Phys. Lett. B **276**, 247 (1992).
- [10] S. Willenbrock, *Hadron Colliders, the Standard Model and Beyond*, hep-ph/0212032 (2002).
- [11] H. L. Lai *et al.* (CTEQ Collaboration), *Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distributions*, Eur. Phys. J. C **12**, 375, (2000), hep-ph/9903282.

- [12] N. Kidonakis, R. Vogt, *Next-to-next-to-leading order soft-gluon corrections in top quark hadroproduction*, Phys. Rev. D **68**, 114014 (2003), hep-ph/0308222.
- [13] V. M. Abazov *et al.* (DØ Collaboration), *$t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV*, Phys. Rev. D **67**, 012004 (2003), hep-ex/0205019.
- [14] T. Affolder *et al.* (CDF Collaboration), *Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV*, Phys. Rev. D **64**, 032002 (2001), Erratum-ibid. D **67**, 119901 (2003), hep-ex/0101036.
- [15] The DØ Collaboration, *Update of the Measurement of the $t\bar{t}$ cross section at $\sqrt{s} = 1.96$ TeV*, FERMILAB-CONF-03-248-E (2003).
- [16] Fermilab Beams Division, *Run II Handbook*, Fermilab Beams Division Run II webpage, <http://www-bd.fnal.gov/runII/index.html>.
- [17] S. Abachi *et al.* (The DØ Collaboration), *The DØ Detector*, NIM, A338, 185 (1994), FERMILAB-PUB-93-179-E (1993).
- [18] S. Abachi *et al.* (The DØ Collaboration), *The DØ Upgrade: The Detector and Its Physics*, FERMILAB-PUB-96-357-E (1996).
- [19] S. Abachi *et al.* (The DØ Collaboration), *Beam Tests of the DØ Uranium Liquid Argon End Calorimeters*, FERMILAB-PUB-92-162 (1992), DØ Note 1434 (1992).
- [20] G. C. Blazey, *The DØ Run II Trigger*, FERMILAB-CONF-97-395-E (1997).
- [21] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, A. D. Polosa, *ALP-GEN, A Generator For Hard Multiparton Processes in Hadronic Collisions.*, JHEP **0307**, 001 (2003), hep-ph/0206293.
- [22] T. Sjöstrand, L. Lönnblad, S. Mrenna, P. Skands, *PYTHIA 6.2: Physics and Manual*, hep-ph/0108264, LU TP 01-21 (2002).
- [23] G. Marchesini, B. R. Webber, G. Abbiendi, I. G. Knowles, M. H. Seymour, L. Stanco, *HERWIG: A Monte Carlo Event Generator for simulating hadron emission reactions with interfering gluons. Version 5.1 - April 1991*, Comp. Phys. Comm. **67**, 465 (1992).
- [24] R. Brun, R. Hagelberg, M. Hansroul, J. C. Lassalle, *GEANT: Simulation Program For Particle Physics Experiments. User Guide And Reference Manual.*, CERN-DD-78-2-REV (1978).
- [25] M. Klute, L. Phaf, D. Whiteson, *TopAnalyze - A Framework Analyze Package for Top Group Analyses*, DØ Note 4122 (2003).

- [26] S. Anderson, E. Busato, M. Klute, S. Kulik, J. Strandberg, *top_analyze documentation*, webpage, http://www-d0.fnal.gov/Run2Physics/top/d0-private/wg/top_analyze/top_analyze_documentation.html
- [27] S. Anderson *et al.* (DØ Top Production Group), *Measurement of the $t\bar{t}$ cross section at $\sqrt{s} = 1.96$ TeV*, DØ Note 4116 (2003).
- [28] M. Klute, A. Quadt, *Measurements of Level 1 Trigger Efficiencies from DØ Data*, DØ Note 3949 (2002).
- [29] T. Golling, *Tool to determine L1 CEM(n,x) Trigger Efficiencies for a given data set*, DØ Note 4041 (2002).
- [30] D. Whiteson, L. Phaf, *Electron Likelihood*, DØ Note 4184 (2003).
- [31] J. Kozminski, R. Kehoe, H. Weerts, S.-J. Park, A. Quadt, J. Gardner, S. Jabeen, *Electron Likelihood in $p14$* , DØ Note in progress (2004).
- [32] G. Blazey *et al.*, *Run II Jet Physics*, DØ Note 3750 (2000).
- [33] F. Blekman, private communication (2003).
- [34] L. Phaf, I. Iashvili, *Measurement of the $t\bar{t}$ cross section at $\sqrt{s} = 1.96$ TeV in the e +jets channel*, DØ Note 4194 (2003).
- [35] The Top Physics Working Group of the DØ Collaboration, *Measurement of the $t\bar{t}$ cross section at $\sqrt{s} = 1.96$ TeV*, DØ Note in progress (2004).
- [36] M. Klute, *A Measurement of the $t\bar{t}$ Production Cross Section in Proton-Antiproton Collisions at $\sqrt{s} = 1.96$ TeV with the DØ Detector at the Tevatron using Final States with a Muon and Jets*, Ph.D. thesis, Bonn University (2004).

Acknowledgements

There are a lot of people who contributed to the successful completion of this thesis and my studies in general, and I am indebted to all of them.

First of all I want to thank my advisor Prof. Norbert Wermes, who by giving me the opportunity of writing this thesis and spending part of my time at DØ, made this thesis possible and also supported me with valuable advice.

My special thank goes to Dr. Arnulf Quadt, who always took the time to advise me and to encourage me. He helped me in many ways and thoroughly proofread this thesis.

I want to thank all the people in the group of Prof. Wermes and Prof. Kobel for their help and the supportive atmosphere they created, especially the people of the DØ Bonn group, Dr. Markus Klute, Markus Warsinsky, Tobias Golling, Jörg Meyer and Kevin Kröninger. The members of the DØ Top Physics Group have also been very helpful and I am especially grateful to Dr. Volker Büscher and Dr. Marumi Kado for their advice.

My studies in its extent and fullness would not have been possible without the generous financial support of the Studienstiftung des deutschen Volkes and the Deutscher Akademischer Austauschdienst.

I am also very grateful to my friends who have accompanied and supported me throughout my whole course of studies, Mi-Hwa Kong, Rene Hamburger, Katrin Fasen, Lars Pesch, and Mayra Fornet-Ponse.

Finally, I want to express my gratitude to those who have always believed in me, my parents, my godparents, Su-In, Su-Nam, and Ben:

고맙습니다.