First Determination of the Electric Charge of the Top Quark

PER HANSSON

KTH Engineering Sciences

Licentiate Thesis
Stockholm, Sweden 2006

KTH Engineering Sciences

Licentiate Thesis

First Determination of the Electric Charge of the Top Quark

Per Hansson

Particle and Astroparticle Physics, Department of Physics
Royal Institute of Technology, SE-106 91 Stockholm, Sweden

Stockholm, Sweden 2006

Cover illustration: View of a top quark pair event with an electron and four jets in the final state.
Image by DØ Collaboration.

Akademisk avhandling som med tillstånd av Kungliga Tekniska Högskolan i Stockholm framlägges till offentlig granskning för avläggande av filosofie licentiatexamen fredagen den 24 november 2006 14.00 i sal FB54, AlbaNova Universitets Center, KTH Partikel- och Astropartikelfysik, Roslagstullsbacken 21, Stockholm.
Avhandlingen försvaras på engelska.

ISBN 91-7178-493-4
TRITA-FYS 2006:69
ISSN 0280-316X
ISRN KTH/FYS/--06:69--SE

Printed by Universitetsservice US AB 2006

Abstract

In this thesis, the first determination of the electric charge of the top quark is presented using 370 pb^{-1} of data recorded by the DØ detector at the Fermilab Tevatron accelerator. $t\bar{t}$ events are selected with one isolated electron or muon and at least four jets out of which two are b-tagged by reconstruction of a secondary decay vertex (SVT). The method is based on the discrimination between b- and $\bar{b}$-quark jets using a jet charge algorithm applied to SVT-tagged jets. A method to calibrate the jet charge algorithm with data is developed. A constrained kinematic fit is performed to associate the W bosons to the correct b-quark jets in the event and extract the top quark electric charge. The data is in good agreement with the Standard Model top quark electric charge of $2e/3$. The scenario where the selected sample is solely composed of an exotic quark Q with charge $4e/3$ is excluded at 92% confidence level. Using a Bayesian approach, an upper limit on the fraction of exotic quarks $\rho < 0.80$ at 90% confidence level is obtained.

I de blindas rike, är den enögde kung.

Niccolò Machiavelli

Contents

Introduction

It is widely believed that the new particle discovered at Fermilab in 1995 [1] is the long-sought top quark. Its currently measured properties are consistent with the Standard Model (SM) expectations for the top quark, but many of its properties are still poorly known. In particular, the electric charge, which is a fundamental quantity characterizing a particle, has not yet been measured for this quark. It still remains not only to confirm that the discovered quark has charge $+2e/3$ and hence the expected SM quantum numbers, but also to measure the strength of its electromagnetic (EM) coupling to rule out anomalous contributions to its EM interactions. Indeed, one alternative interpretation has not yet been ruled out: that the new particle is a charge $-4e/3$ quark. In the published top quark analyses of the CDF and DO collaborations [2], the pairing of the b quarks and the W bosons in $p\bar{p} \to t\bar{t} \to W^+W^-b\bar{b}$ processes are not determined. As a result, there is a twofold ambiguity in the electric charge assignment of the "top quark". In addition to the SM assignment $t \to W^+b$, $t \to W^-b$ is also conceivable, in which case the "top quark" would actually be an exotic quark with charge $q = -4e/3$. The analysis presented in this thesis is not carried out within the framework of any extension to the SM. Nevertheless interpreting the particle found at Fermilab as a charge $-4e/3$ quark is consistent with current precision electroweak data. Current $Z \to \ell^+\ell^-$ and $Z \to b\bar{b}$ data can be fitted with a top quark of mass $m_t = 270$ GeV, provided that the right-handed b-quark mixes with the isospin $+1/2$ component of an exotic doublet of charge $-1e/3$ and $-4e/3$ quarks, $(Q1\ , Q4)_R$ [3]. If the top quark had a mass of $m_t = 270$ GeV, it would so far have escaped detection at the Fermilab Tevatron. The CDF collaboration has carried out a search for a heavy t'-quark using 760 pb^{-1} of data and excludes masses up to 258 GeV [4]. With data sets beyond 1fb^{-1} and combining DØ and CDF, the Tevatron will be capable of detecting t'-quarks with masses of 270 GeV and more. It should also be noted that a mass of 270 GeV merely corresponds to the best fit to SM precision electroweak data in these models and the mass of such a heavy fermion could still be above 300 GeV.

In this thesis, the first determination of the electric charge of the top quark using ~ 365 pb^{-1} of $p\bar{p}$ data collected with the DØ experiment is presented. The result of the measurement is described in the paper

DØ Collaboration, V. M. Abazov *et. al*, *"Experimental discrimination between charge* $2e/3$ *top quark and charge* $4e/3$ *exotic quark production scenarios"*, hep-ex/0608044, submitted to Phys. Rev. Lett.

The thesis is outlined as follows: chapter 1 gives an overview of the SM and the top quark. The DØ detector is described in chapter 2 and the object reconstruction is presented in chapter 3. The analysis to determine the top quark charge is described in chapter 4 followed by a conclusion and outlook in chapter 5.

Authors Contribution

In this thesis the result of my work at the DØ experiment at Fermilab between February 2004 and summer 2006 is presented. Arriving at Fermilab I quickly started working in the top quark group with a feasibility study to determine the possibility and the amount of data needed for a determination of the electric charge of the top quark. The top quark charge had not been measured before and was considered very difficult due to the low statistic sample of top quarks.

I have been responsible for the entire analysis from the first day. This analysis was developed through intense collaboration with Dr. Christophe Clément and Dr. David Milstead. At the start, most work went into studying various jet charge algorithms and their optimization as described in Sec. 4.3. In autumn 2004, I showed that a measurement should be possible with the data that was collected during this period and the work was accelerated towards forming a full analysis. During winter 2004 and spring 2005 most of my work went into defining and validating the jet charge calibration discussed in Sec. 4.4 and finding and studying various sources of systematic uncertainties. Based on the result of the top quark pair cross section, the top quark charge measurement was first presented as a preliminary result at the PANIC05 conference in October 2005. During winter 2005 and spring of 2006 I worked mostly on refining the data calibration method but also to develop the method of a simultaneous measurement of the fraction of exotic quarks in the sample. The result was finally submitted to Physical Review Letters for publication in the summer of 2006.

During 2005 I was involved in studies of the jet reconstruction efficiency and energy calibration, especially studying the out-of-cone radiation correction described in chapter 3. During spring 2006 the DØ detector was upgraded, extending the silicon vertex detector with an additional layer allowing for an improved tracking of charge particles. I was responsible for upgrading and developing the online software displaying the silicon tracking detector status.

Notation

As mentioned earlier, the particle discovered at Fermilab is widely believed to be the SM top quark. To this date, many of its parameters are poorly known. Until all its properties are determined with high precision, exotic scenarios (not included

in the SM) are not excluded and only measurements such as the one presented in this thesis can finally decide if the particle is the SM top quark or an exotic quark. This thesis has no preconceived opinion on the true nature of the particle discovered. From now on, the name "top" in this thesis is simply a notation chosen for consistency with other papers referenced. The "top" quark refers to the SM top quark only when specifically indicated or when a comparison of the exotic quark scenario with the SM scenario is carried out.

Chapter 1

The Standard Model Top Quark

Elementary particle physics research is the quest for understanding the smallest constituents of matter and their interactions. The SM is the theoretical framework used to describe the known elementary particles and their interactions. The current view is that all matter is made up of three kinds of particles: leptons, quarks and mediators. In the SM the particle matter consists of spin-1/2 quarks and leptons, which, down to a scale of around 10^{-18} m appear elementary [1]. There are six "flavors" of quarks and leptons arranged in three generations. There are four fundamental forces through which these elementary particles interact; gravity, electromagnetic, weak and the strong force. The electromagnetic- and weak force are manifestations of one single force, called the electroweak force, in the Glashow-Weinberg-Salam (GWS) model and the number of forces are then reduced to three. The SM is a quantum field theory (QFT) based on the symmetry group $SU(3) \times SU(2)_L \times U(1)_Y$; all particles are described as fields and forces between them are interpreted as being due to the exchange of mediator particles. These particles are known as gauge bosons, which are spin-1 particles[2]. The building blocks of the SM are summarized in Tab. 1.1. For a pedagogical introduction to elementary particle physics and the Standard Model, see e.g. Ref. [7]. The SM has been extremely successful and agrees with nearly all experimental data so far [5]. However, the SM is a not a complete theory of particle physics. For example, it does not incorporate gravity, nor can it account for so-called dark matter and energy. Many theoretical extensions of the SM have been postulated which predict the existence of hitherto-undiscovered fundamental particles including exotic quarks and leptons [8].

The top quark is the partner to the bottom quark in the weak isospin doublet in the SM. Unless otherwise specified, in this chapter the term "top quark" refers

[1] Elementary means here that they don't have any internal structure.

[2] The graviton is postulated to mediate the gravitational force and have spin-2 but is yet to be observed.

	Symbol	Name	Mass (MeV)	Charge (e)
Quarks (spin=1/2)	u	up	1.5 to 4	+2/3
	d	down	4 to 8	-1/3
	s	strange	80 to 130	-1/3
	c	charm	1150 to 1350	+2/3
	b	bottom	4100 to 4400	-1/3
	t	top	172.5 GeV	+2/3
Leptons (spin=1/2)	ν_e	electron neutrino	$<$ 3 eV	0
	e	electron	0.511	-1
	ν_μ	muon neutrino	$<$0.19	0
	μ	muon	105.7	-1
	ν_τ	tau neutrino	$<$18.2	0
	τ	tau	1777.0	-1
Gauge bosons (spin=1)	γ	photon	0	0
	g	gluon	0	0
	W	W	$(80.425\pm0.038)\times10^3$	1
	Z	Z	$(91.1876\pm0.0021)\times10^3$	0
Higgs boson (spin=0)	H	Higgs	$>$114 GeV	

Table 1.1. The SM particles [5]. The Higgs boson is yet to be observed, direct searches for the Higgs boson puts a lower limit of 114 GeV on its mass [6].

to the SM top quark, the discovery of which was announced by the DØ and CDF experiments around a decade ago [1]. This is in contrast to the notation discussed in the introductory chapter. The existence of the top quark was expected since the discovery of the bottom (b) quark in 1977 implied the existence of a further quark to complete the quark sector with a three generation structure. This chapter describes the current experimental status of the discovered quark and the notation is adapted to simplify the discussion. The first direct studies of the top quark were performed during Run I of the Tevatron at center-of-mass energy of $\sqrt{s} =$ 1.8 TeV and continued during Run II with higher-statistic samples. The Tevatron remains to date the only top factory. The top quark mass was predicted from precision electroweak measurements from LEP, SLD, NuTeV and $p\bar{p}$ colliders[9] before its discovery. Due to the limited number of top quarks observed so far its properties are less well experimentally determined than those of other known quarks. However, most existing results are consistent with the particle possessing the quantum numbers of the SM-top quark [5].

There are several reasons why the top quark is interesting in the framework of the SM and possible physics beyond it:

- The top quark production and decay properties are poorly known and provide important tests of the SM at the Tevatron.
- The short expected lifetime of the top quark implies that it is the only quark that will decay before it hadronizes.
- The top quark mass is an important parameter in precision electroweak fits and can thus constrain theoretical models of physics beyond the SM [10].
- The top quark may have special dynamics related to new particles beyond the SM due to its large coupling to the Higgs boson.
- The large mass of the top quark and the increasing production cross section at higher energies implies that top quark production will be one of the principal sources of background when searching for evidence of New Physics processes at the Large Hadron Collider, which will collide protons at $\sqrt{s} = 14$ TeV from 2008 onwards.

1.1 Production of the Top Quark

Evidence for the direct production of the top quark has been obtained by the DØ and CDF collaboration solely via the measurement of $t\bar{t}$ pair-production processes. The two main processes are $q\bar{q} \rightarrow t\bar{t}$ and $gg \rightarrow t\bar{t}$, as shown in Fig. 1.1. The quark and gluon content of the proton is described by so-called parton distribution functions. These describe the probability to find a gluon or a quark of a certain flavor carrying a fraction x of the proton's (or anti-proton's) momentum. The value of x required for production of top quarks decreases with increasing collision energy. At the Tevatron energy, the top quark pairs are produced approximately in

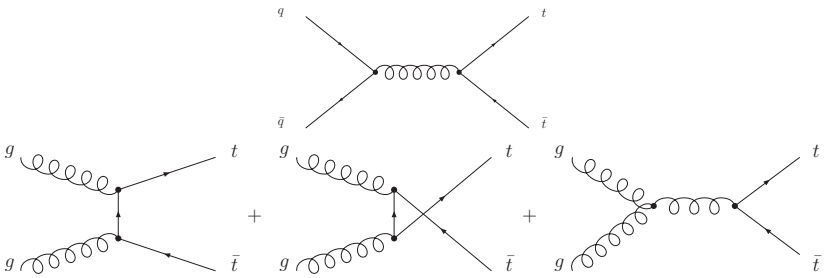

Figure 1.1. Lowest order Feynman diagrams for the top quark pair production at the Tevatron.

85% of the events by quark anti-quark fusion $q\bar{q} \rightarrow t\bar{t}$ and in 15% from gluon fusion $gg \rightarrow t\bar{t}$ [11]. The top quark is also produced singly via the weak interaction via the so called s- and t-channel. Discovering single top production is more experimentally challenging due to a less distinctive event signature and larger backgrounds. No experimental evidence for production of single top quarks has been found so far [12].

The total top quark pair and single production cross section in the SM at a center-of-mass energy of $\sqrt{s} = 1.96$ TeV is calculated to be ≈ 7 pb and ≈ 3 pb respectively [13].

1.2 Decay of the Top Quark

In the SM the top quark is predicted[3] to decay to a W^+ boson and a b-quark with a branching ratio of ~ 0.999 [5]. The large decay width (≈ 1.5 GeV) corresponds to a lifetime of around 5×10^{-25} s. This lifetime is shorter than the corresponding time for hadronization and thus no bound states with t or $\bar{t}$ exists [14].

1.3 Mass of the Top Quark

The top quark is heavier than any other elementary particle found so far. The mass of the top quark have been measured to the best relative precision of all the quarks. Combining the results from both experiments at the Tevatron the world-average top quark mass is 172.5 ± 2.3 GeV. More information on the techniques and results from the top quark mass analyzes can be found in [15]. The precision electroweak measurements from e.g. LEP can be used to make an indirect prediction of the top quark mass. The result, $179.4^{+12.1}_{-9.2}$ GeV, is consistent with the direct measurements.

[3] Assuming only three families and unitarity of the flavor mixing matrix (called Cabibbo-Kobayashi-Maskawa matrix or CKM in short) $|V_{tb}| \simeq 1$.

1.4 Experimental Tests of the Standard Model Top Quark Sector

To put the work in this thesis into perspective a summary of the world measurements in the top quark sector is given.

1.4.1 Top Quark Pair Production Cross Section

Both DØ and CDF have measured the $t\bar{t}$ production cross section. It is extracted by counting the number of observed events, estimating the number of background events and measuring the integrated luminosity (taking into account the acceptance). Any abnormal top quark decay such as $t \to H^+b$ can result in a lower cross section than predicted by the SM. A higher than expected cross section would hint at new unknown production mechanisms. One example can be found in Ref. [16].

So far all direct measurements of the $t\bar{t}$ production cross sections are in agreement with the SM prediction. Figure 1.2 shows the measured cross sections from the DØ collaboration. The full list of cross section measurements at the Tevatron can be found in [5].

1.4.2 Top Quark Decay Branching Ratio

As discussed above, within the SM the dominant decay mode for the top quark is $t \to W^+b$. The CKM matrix [18] element V_{tx} (with $x = b, s, d$) determines the coupling between the top quark and other flavors. The W^+d and W^+s decay modes are suppressed by the square of the mixing matrix elements. The predicted values of the mixing matrix can be tested by determining the ratio R of branching ratios $\mathcal{B}$ for the processes,

$$R = \frac{\mathcal{B}(t \to Wb)}{\mathcal{B}(t \to Wq)}. \tag{1.1}$$

The SM prediction[4] is $0.9980 < R < 0.9984$ at 90% confidence level and the current best measurement [19] is $R = 1.03^{+0.19}_{-0.17}$, in good agreement with the SM.

1.4.3 *W* Boson Helicity

New physics has been searched for in the dominant top quark decay vertex $t \to W^+b$ where the helicity of the W boson is sensitive to anomalous contributions from new physics beyond the SM. In the SM the right-handed fraction of W bosons is suppressed compared to the longitudinal fraction ($\sim 70\%$). By studying the angular distribution of the W boson decay products with respect to the top quark direction, DØ puts an upper limit of 0.23 on the fraction of right-handed W bosons at 95% confidence level [20]. Direct measurements of the longitudinal fraction give a value of $0.74^{+0.22}_{-0.34}$ [21] and 0.56 ± 0.31 [22].

[4] This ratio can be expressed in the elements of the mixing matrix elements as $R = \left(|V_{tb}|^2\right) / \left(|V_{tb}|^2 + |V_{ts}|^2 + |V_{td}|^2\right)$.

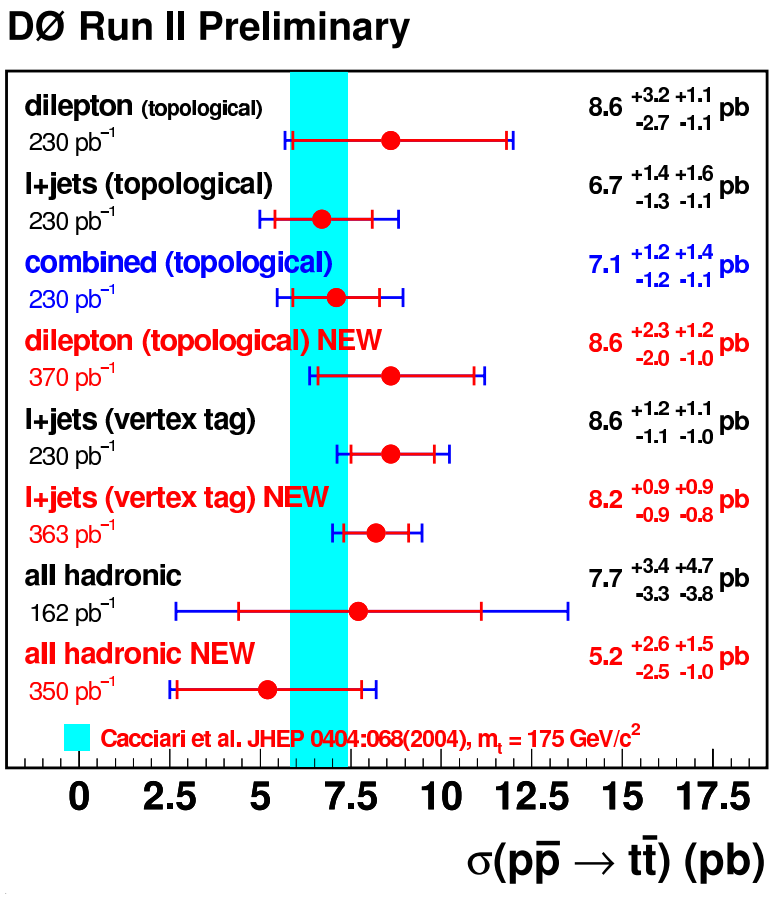

Figure 1.2. The $t\bar{t}$ production cross section measured by the DØ collaboration as of fall 2005. The figure contains both published and preliminary results [17]. The notation of the different measurements is explained in Sec. 3.1.1.

1.4.4 Resonances and Rare Decays

Due to its large mass there are various physics models [23; 24] beyond the SM in which the top quark plays a central role. In these models, a heavy particle decaying to $t\bar{t}$ can be produced with cross sections large enough to be visible at the Tevatron. The DØ and CDF collaborations have searched for $t\bar{t}$ production via an intermediate particle state by looking for narrow-width peaks in the spectrum of the invariant mass of $t\bar{t}$ events. DØ and CDF report no evidence for such an intermediate state and exclude masses of such a state up to 680 GeV [25].

1.4.5 Top Quark Spin Correlations

The top quarks in $t\bar{t}$ pairs produced from unpolarized incoming particles in $q\bar{q}$ annihilation are expected to be unpolarized. However, their spin is expected to be highly correlated with a higher fraction of events in which the spins are aligned rather than anti-aligned. DØ measured the spin correlation in a low statistics sample in Run I and found no deviation from the SM prediction[26].

1.4.6 The Standard Model Higgs Boson

A key concept of the SM is the so-called gauge invariance, which can be interpreted as a transformation of fields in such a way that they do not change (they are gauge invariant). It can be be shown that to keep a theory like QED gauge invariant an additional interaction must be introduced, i.e. the photon. From gauge invariance it can be shown that the SM predicts massless mediator bosons (and also massless fermions), while the $W^{\pm}$ and Z bosons are known to have a large mass. By postulating another field, the Higgs field, that all particles interact with, this problem is taken into account and all masses are consistent with the theory. This additional field implies the existence of the Higgs boson as the mediator of the field and is today the only undiscovered particle in the SM. For a good introduction to gauge theories and spontaneous symmetry breaking, see e.g. [7].

Although the couplings of the Higgs boson to other particles are predicted the mass of the Higgs boson is not. It can however be inferred from high precision measurements of the W boson mass where virtual loop corrections involving both the Higgs boson and the top quark contribute. The same principle used when predicting the top quark mass before its discovery. Figure 1.3 shows the dependence of the Higgs boson mass on the top quark and W boson masses. Although the uncertainty on the prediction of the Higgs boson mass is large, it is evident that the experiments imply a low-mass Higgs boson mass. The 95% lower confidence limit on the Higgs mass from direct searches is 114 GeV and the upper 95% confidence limit is 166 GeV (including the direct search lower limit increases the upper limit to 199 GeV) [27].

The search for evidence of the existence of the Higgs boson is currently one of the largest efforts in the particle physics community and will be addressed at the LHC.

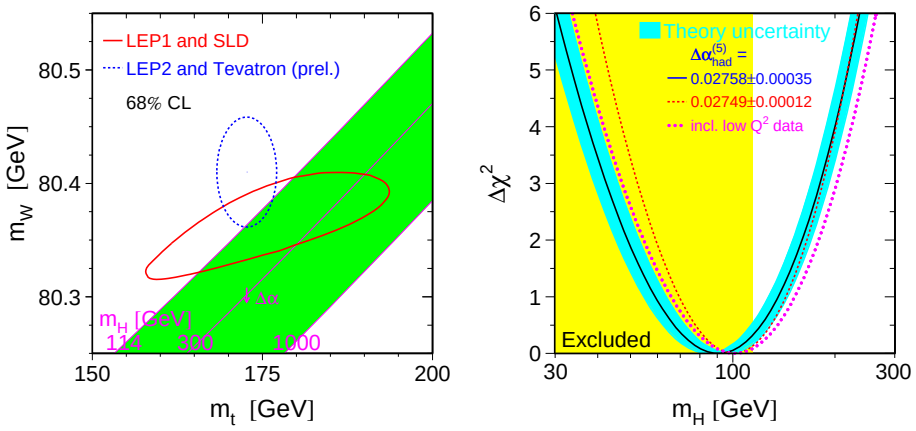

Figure 1.3. Constraints on the Higgs boson mass as a function of the W boson and the top quark mass (left) and the fit to the electroweak parameters as a function of the mass of the Higgs boson (right) [27].

Chapter 2

The DØ Detector

The DØ detector was proposed in 1983 to study proton anti-proton collisions at the Fermilab Tevatron accelerator. The purpose was to study a wide range of phenomena focusing on high-mass states and high-p_T processes. The DØ detector performed well during Run I of the Tevatron, which lasted from 1992 to 1996. Among many impressive results was the discovery of the long sought top quark and measurements of its mass. During Run I, the Tevatron operated with six bunches of protons and anti-protons with 3500 ns between each bunch-crossing. The center-of-mass energy was 1.8 TeV and the peak instantaneous luminosity was typically around $1 - 2 \times 10^{31}\text{cm}^{-2}\text{s}^{-1}$. The data recorded by the DØ experiment in Run I amounted to approximately 120 pb^{-1}. Following the completion of the Fermilab Main Injector and other substantial Tevatron upgrades, the DØ experiment was running again in 2001. In this phase, called Run II, the Tevatron is operated with 36 bunches of protons and anti-protons with 396 ns between each bunch-crossing and a center-of-mass energy of 1.96 TeV. The instantaneous luminosity increased by a factor of ten.

To take advantage of the increased luminosity and center-of-mass energy delivered by the Tevatron the DØ experiment was greatly upgraded during 1996-2001. Among the major upgrades it is important to note that the tracking system from Run I which lacked a magnetic field and suffered from radiation damage was replaced with a silicon microstrip tracker and a fiber tracking detector in a 2 T magnetic field. The detector consists of three major subsystems: the central tracking detectors, a uranium/liquid-argon calorimeter and a muon spectrometer. A side-view of the upgraded DØ detector is shown in Fig. 2.1. This chapter gives a brief description of the upgraded DØ detector and those components most pertinent to the analysis presented in this thesis. A more detailed description can be found in Ref. [28].

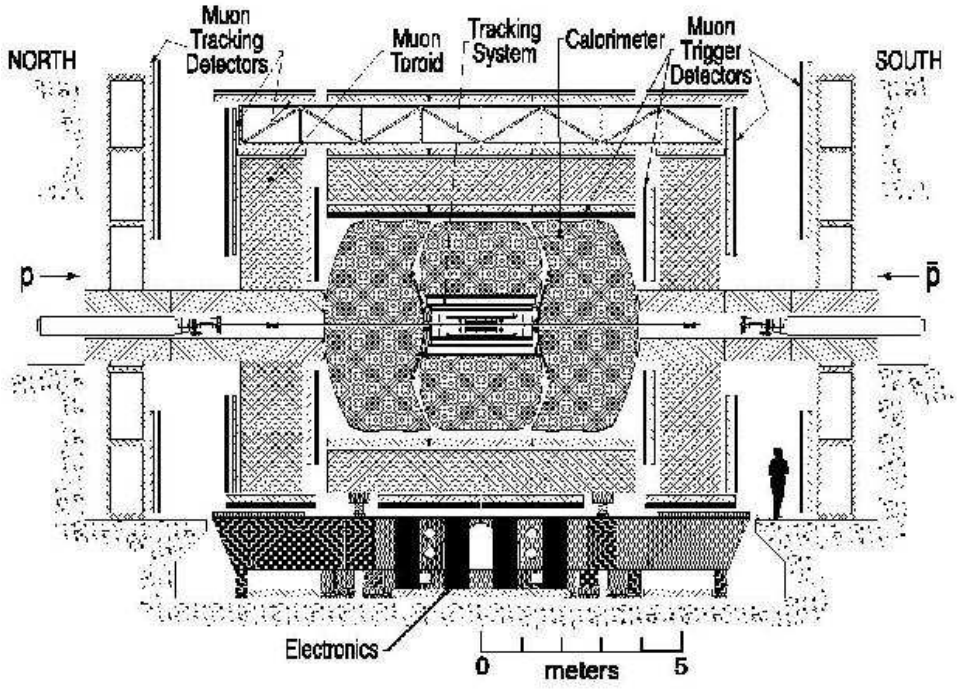

Figure 2.1. Diagram showing the upgraded DØ detector as seen from the outside of the Tevatron ring. The $+z$ axis is to the right, $+y$ is up and $+x$ is out of the page [28].

2.1 The DØ Coordinate System

In the detector description and the data analysis, the standard DØ collaboration coordinate system is used where the positive z-axis points in the direction of the proton beam, the positive x-axis points radially outward from the Tevatron center and the positive y-axis is pointing upwards. To specify a direction in the detector, the polar and azimuthal angles θ and ϕ can be used. Since the angle θ is not invariant under Lorentz transformations along the z-axis it is common to use the pseudorapidity $\eta = -ln(tan\frac{\theta}{2})$ instead. η approximates the true rapidity $\mathbf{y} = \frac{1}{2}ln(\frac{E+p_z}{E-p_z})$ in the kinematic region where the mass is negligible, i.e. when $E \approx p$. The separation between two objects labeled 1 and 2 can be expressed as the distance ΔR between them in the (η, ϕ) plane, defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. The term "forward" is commonly used to describe regions of the detector at large $|\eta|$. Since the initial momentum along the beam axis is unknown and some particles escape detection close to the beam axis the measured variables are in general quantities transverse to the beampipe direction, such as transverse momentum (p_T) or energy (E_T), and missing transverse energy, $\not\!\!E_T$, from neutrinos escaping the detector.

2.2 The Central Tracking System

The measurement of tracks of charged particles and the reconstruction of a production or decay vertex is an important part of experimental studies at collider experiments. A precisely determined primary interaction vertex allows accurate measurements of lepton p_T, jet E_T and $\not\!\!E_T$. Using the tracking information it is possible to identify jets containing decay products of a b-quark by finding tracks emanating from a secondary vertex which is displaced with respect to the primary interaction vertex. This is especially important for top quark physics were the dominant top quark decay is to a b-quark and a W boson. The central tracking system in DØ was completely replaced after Run I. The new system consists of two parts: The Silicon Microvertex Tracker (SMT) and the Central Fiber Tracker (CFT) enclosed in a magnetic field oriented along the beam axis. The 2 T magnetic field is provided by a 2.8 m long superconducting solenoid magnet with a radius of approximately 60 cm. Charged particles produced in the collision are bent around the field lines in a magnetic field of strength B. The radius r of the particle trajectory can be used to calculate the p_T through [29]:

$$p_T[\text{GeV}] = 0.3 \times r[\text{m}] \times B[\text{T}]. \tag{2.1}$$

Combined, the SMT and CFT locates the primary interaction vertex with a resolution of ≈ 35 μm along the beam direction. They provide an impact parameter[1] resolution of about 15 μm in the $r - \phi$ plane for particles with $p_T > 10$ GeV in the

[1] The impact parameter is defined as the distance of closest approach (d_{ca}) of the track to the primary vertex in the plane transverse to the beamline. The impact parameter significance is defined as $d_{ca}/\sigma_{d_{ca}}$, where $\sigma_{d_{ca}}$ is the uncertainty on d_{ca}.

central region [28]. In addition, they also provide information on track p_T to the trigger system for fast event decisions. A schematic view of the central tracking system is shown in Fig. 2.2.

2.2.1 The Silicon Microvertex Tracker

The Silicon Microvertex Tracker (SMT) is the innermost part of the DØ detector. Its purpose is to provide both high-quality vertex finding and high resolution tracking. Its design is primarily dictated by the accelerator environment. For example the length of the device is determined based on the length of the interaction region, $\sim$ 25 cm. Since the SMT has to cover a significant solid angle it is difficult to ensure that the detector planes are always perpendicular to the outgoing particle trajectories. Therefore, the SMT has a barrel design interspersed with discs in the central region, while the forward region consists primarily of disks. There are six barrels, each with four silicon readout layers. Each barrel is attached (at the high $|z|$ side) to a disk with wedge detectors. At the outside of the barrel-disk assembly three disks are mounted. In the forward region four larger disks provide tracking capabilities up to $|\eta| = 3$, see Fig. 2.3. Particles with low pseudorapidities are mainly measured by the barrels while particles with larger pseudorapidities are also measured by the disks.

There are several different types of silicon sensors. Both the disks and barrels uses a combination of single-sided and double-sided sensors depending on the location in the SMT (varying with both layer and $|z|$ for the barrel). The SMT has in total 912 readout modules, with 792,576 channels. Most of the sensors have a pitch of 50 μm and the hit resolution is approximately 10 μm (improving from the $1/\sqrt{12}$ dependence due to the pulse height information). The resolution in z-direction varies depending on the detector type in the various part of the SMT ranging from around 35 μm to 450 μm for 90° and 2° stereo angle detectors respectively. The p_T resolution for central tracks with $|\eta| < 2$ varies with momentum from $2 - 5\%$ at track momentum of around 1 GeV to $5 - 10\%$ for tracks with approximately 10 GeV momentum. The resolution degrades fast in the forward region up to 30% for tracks around 10 GeV at $|\eta| \approx 3$.

2.2.2 The Central Fiber Tracker

The Central Fiber Tracker (CFT) surrounds the SMT and covers the radial space from 20 to 52 cm from the center of the beampipe as shown in Fig. 2.2. The essential part of the CFT is the scintillating fiber system. Each fiber is 835 μm in diameter (including cladding which is approximately 50 μm thick) and oriented along the beam pipe in doublet layers on eight concentric cylinders. The innermost two cylinders are 1.66 m long and the outer six are 2.52 m long. The fibers in each doublet layer are separated by half the fiber diameter to achieve total coverage. Each cylinder supports one axial (oriented along the beam axis) doublet layer, see Fig. 2.4, and a second doublet layer oriented with $\pm 3°$ stereo angle. The scintillating fibers are arranged in a multiclad structure using polystyrene as core material

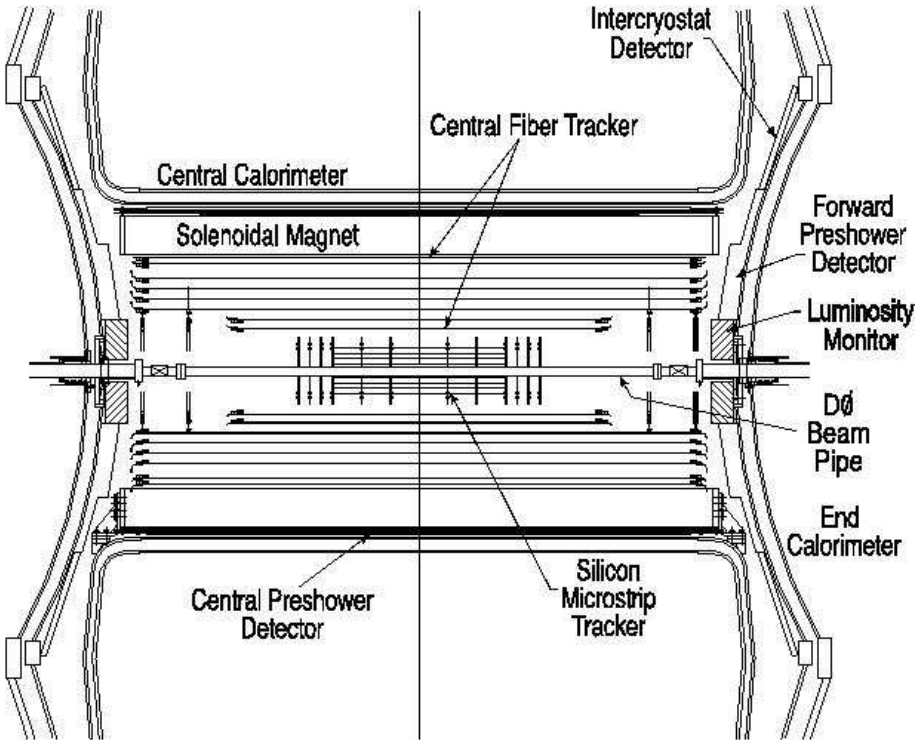

Figure 2.2. Schematic view of the central tracking system. The preshower detectors, calorimeter and luminosity monitors are also shown [28].

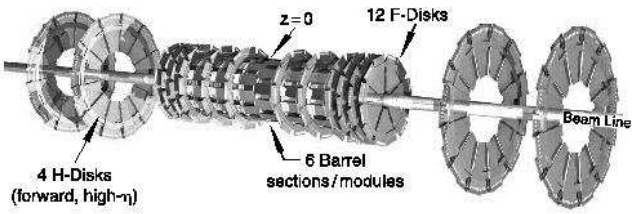

Figure 2.3. Design view of the Silicon Microvertex Tracker [28].

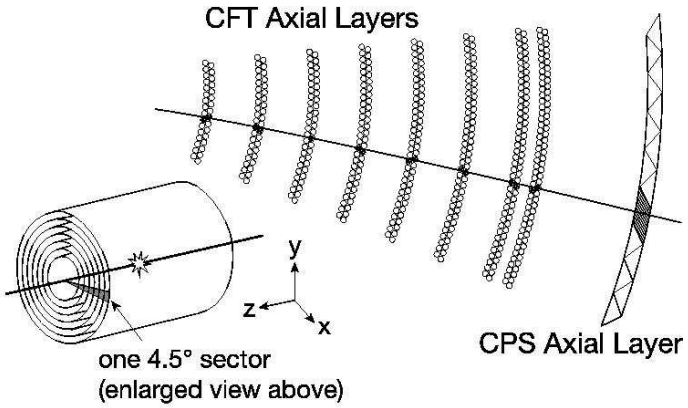

Figure 2.4. View of the axial layers of the CFT and the Central Preshower (CPS) detectors with a hypothetic track overlaid. Each of the axial doublet layer has an associated additional doublet layer with a stereo angle of $\pm 3^\circ$ not shown in this picture.

and paraterphenyl as the light-emitting material. To get the light out a second wavelength-shifter material is added and the light is transported via a clear fiber to the Visible Light Photon Counters (VLPC's) connected to one end of the fibers where the light is converted to an electric pulse and read out. The CFT has in total 76,800 channels of VLPC read out and the hit resolution is around 100 μm. Approximately 200 km of scintillating and 800 km of clear fiber is used in the CFT in total.

The CFT's axial layers are part of the fast Level 1 trigger which aid in finding the interesting collisions discussed in more detail in Sec 2.7.

The p_T resolution achieved combining SMT and CFT is studied using $Z \rightarrow \mu^+\mu^-$ events and resolutions of $\sigma/p_T^2 \approx 0.002$ have been obtained [30].

2.3 The Preshower Detectors

The preshower detectors provide an early energy sampling and good position measurement. The detectors are designed to help in electron identification and to correct for the energy lost in the upstream material (mainly the solenoid). The fast response also allows the preshower detectors to be part of the event trigger.

The design consists of two similar detectors, the Central Preshower Detector (CPS) and the Forward Preshower Detector (FPS). The CPS (FPS) consists of

three (two) layers of triangular strips of scintillating material interleaved to remove any gaps.

The central preshower detector (CPS) is placed in the 5 cm gap between the solenoid magnet and the central calorimeter covering $|\eta| < 1.3$ as shown in Fig. 2.2. Inside the CPS, a lead radiator about one radiation length, X_0, thick (corresponding to ≈ 0.55 cm) and 244 cm long is inserted. The solenoid and the lead radiator together comprise about two radiation lengths (the solenoid is $0.9X_0$ thick) for normal incident particles increasing to about four radiation lengths at maximum CPS coverage. Electrons and photons are converted into showers in the upstream material and this provides a discrimination between electrons or photons and pions, where the latter mostly passes through without showering.

The two (north and south) forward preshower detectors are mounted on the inner part of the end cap calorimeter (see Fig. 2.2) covering $1.5 < |\eta| < 2.5$. Each detector consists of a two radiation lengths thick stainless steel radiator sandwiched between two layers of scintillating strips. This design allows for position measurements as well as possible discrimination between electrons or photons and pions.

2.4 The Calorimeter

The calorimeter absorbs and measures particle energy and the position of the deposited energy. It consists of a central calorimeter (CC) and two (north and south) end cap calorimeters (EC), see Fig. 2.5. The Run II calorimeter is essentially the same calorimeter as in Run I but with upgraded electronics adapted to the new accelerator environment, i.e. the higher bunch crossing frequency. The CC covers a region up to $|\eta| < 1.0$ and the two end cap calorimeters extend the coverage to $|\eta| = 4$, as shown in Fig. 2.6. The CC and EC are constructed in three parts; the electromagnetic section (EM) closest to the beam pipe followed by the fine hadronic section (FH) and the coarse hadronic (CH) section. The active medium for all the calorimeters is liquid argon and the three calorimeters are enclosed in a cryostat at a temperature of approximately 80 K. Different locations have different absorber plates. The main absorber used in the EM calorimeter is nearly pure depleted uranium assembled into thin plates (≈ 3 mm) in both CC and EC. The fine hadronic section uses 6 mm thick uranium alloy plates (both in the CC and EC) and the coarse hadronic section uses 46.5 mm thick copper (stainless steel) plates in the CC (EC).

The readout cells of the calorimeter are arranged in sizes such that each cell covers $\Delta\phi \times \Delta\eta = 0.1 \times 0.1$, which is comparable to the transverse sizes of showers: $1-2$ cm for EM showers and about 10 cm for hadronic showers. Longitudinal depth segmentation is important when distinguishing between electrons or photons and hadrons. In the EM calorimeter there are four depth layers (in both EC and CC). The third layer is placed at the expected shower maximum and is twice as finely segmented in the lateral direction for increased spatial resolution. The amount of material (tracking, cryostats, solenoid, etc.) between the interaction region and the first active gap in the EM calorimter at amounts to approximately $4X_0$ in the

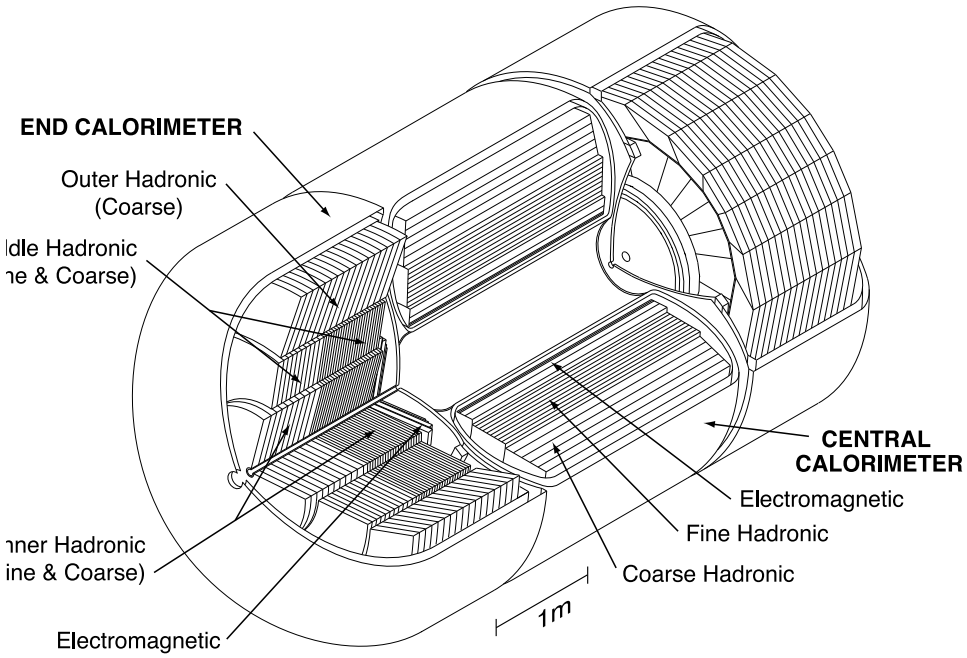

Figure 2.5. View of the central and two end cap calorimeters [28].

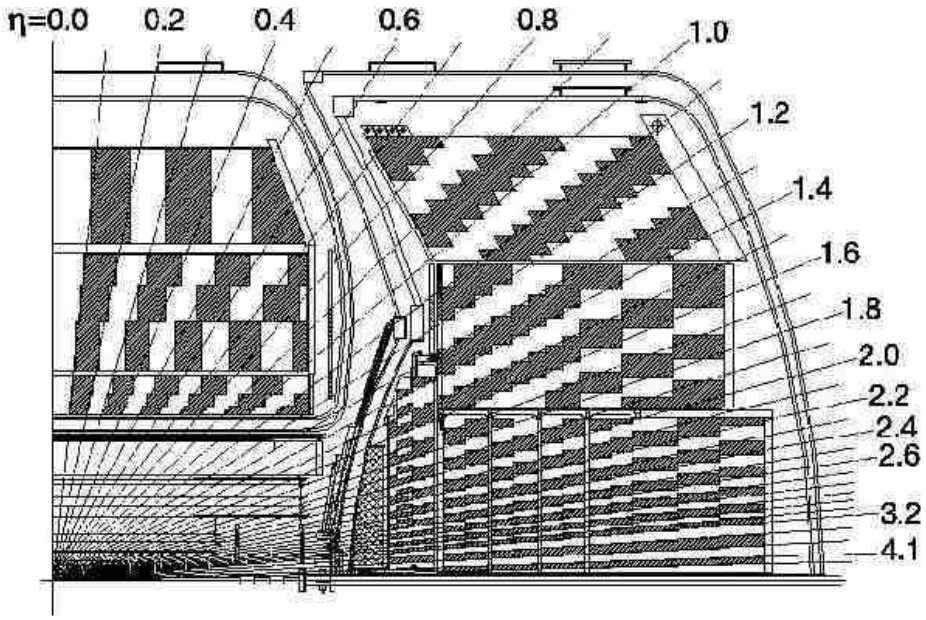

Figure 2.6. Schematic view of a portion of the DØ calorimeter showing the transverse and longitudinal segmentation pattern. The shaded areas corresponds to cells grouped together for readout. The lines indicate values of psuedorapidity, as measured with respect to the centre of the detector [28].

CC and $4.4X_0$ in the EC. The EM calorimeter contains uranium comparable to approximately 20 radiation lengths ($X_0^{\mathrm{U}} = 3.2$ mm) to capture the overwhelming part of the electromagnetic shower. As the nuclear interaction length is much larger than the radiation length ($\lambda_I^{\mathrm{U}} \approx 10.5$ cm $\approx 30X_0^{\mathrm{U}}$) the hadronic particles typically deposits most of its energy in the hadronic part of the calorimeter.

The calorimeter provides trigger information to all three trigger levels. The Level 1 and Level 2 triggers are based on analog sums of energy in special trigger towers.

The energy resolution of a sampling calorimeter can be parametrized by

$$\frac{\sigma(E)}{E} = \sqrt{C^2 + \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2}. \tag{2.2}$$

The parameter C is called the "constant term" and comes from calibration errors or other systematic effects, N is an energy independent "noise term" including contributions from uranium decays and electronic noise. The largest contribution comes from the "sampling term", S, which is the statistical error in the sampling procedure. For the Run II detector, preliminary studies shows a degradation of the calorimeter resolution from several sources e.g. worse noise characteristics of the detector electronics, shorter pulse shaping due to the increased bunch crossing frequency, large cell-to-cell miscalibrations and more upstream material from the new tracking system, degrading the sampling term. The jet energy resolution is described in Sec. 3.6.3.

2.4.1 The Inter-Cryostat Detector

Due to the fact that the calorimeter is contained in three separate cryostats it has incomplete coverage in the region $0.8 < |\eta| < 1.4$. Therefore, scintillation counters with a cell size matching the calorimeter as well as single cell structured scintillation counters are inserted in this region. These detectors allow for a sampling of the inter-cryostat region improving the energy resolution.

2.5 The Muon Spectrometer

Muons lose only a small fraction of their energy in the central tracking system and calorimeter. The DØ muon system [31] is located around the calorimeter and is used to trigger and to measure muon p_T and charge independently of the tracking system. An overview of the muon system is shown in Fig. 2.7. The system is divided into a central and forward detector. A 1.8 T magnetic field is supplied by a 109 cm thick iron toroid magnet, built in three sections to allow for easier access to the inner part of the detector. The central magnet is located at a radial distance of 318 cm from the beam line covering the region $|\eta| < 1.0$. The forward toroids are located at $454 < |z| < 610$ cm. The muon detectors consist of proportional drift tubes (PDT's), mini drift tubes (MDT's) and scintillation counters. The PDT's are

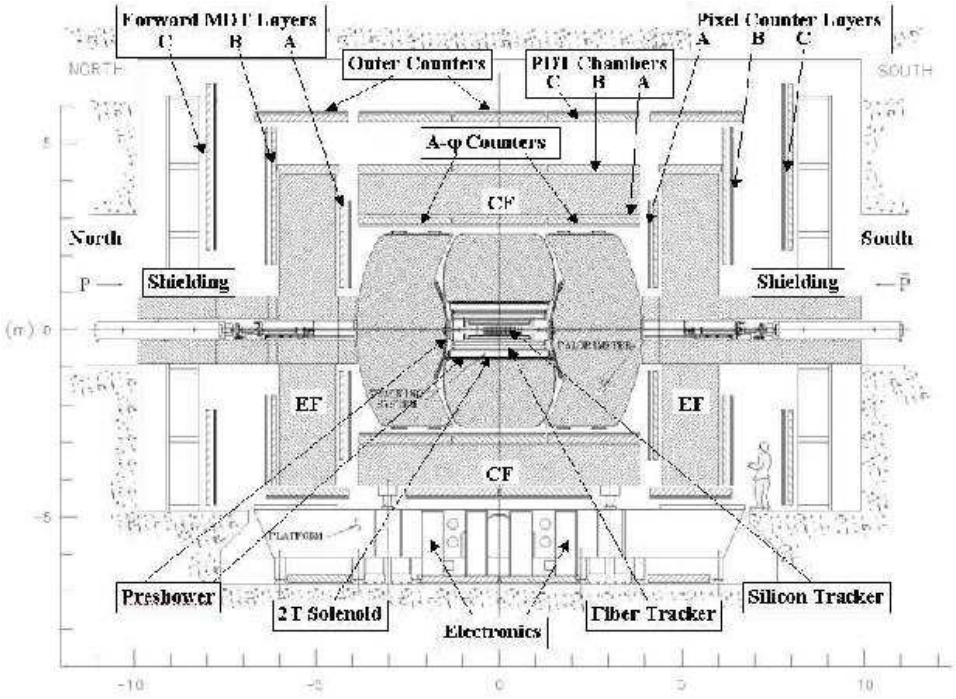

Figure 2.7. Schematic view of the muon system. CF and EF denote the toroid magnets [31].

rectangular volumes filled with gas and cover $|\eta| < 1.0$. A charged particle ionizes the gas and the electrons are amplified at the 50 μm thick anode wire. The Vernier cathode pads above and below the wire are segmented to provide information on the ionization position along the wire. The maximum electron drift velocity is 450 ns and gives a single wire resolution of around 1 mm in the radial direction of the wire for 10 cm wide drift cells. The MDTs extend the coverage up to $|\eta| < 2.0$ and consist of drift tubes with shorter electron drift times ($40 - 60$ ns) than the PDTs (the MDT cell width is 9.4 mm and the length ranges from 1 to 6 m). The radial resolution for single wires is ≈ 0.7 mm.

Both the central and forward drift chambers consist of three layers, A, B and C. The A layer is located inside the toroid magnet while the B and C layers are outside. Each layer also has a sheet of scintillating pixels (except layer B in the central region) used for triggering, cosmic muon (and other background) rejection and track reconstruction. The scintillator geometry is matched to the central fiber tracker trigger read out to provide matching of tracks from the central tracking system to the muon system at the first trigger level. The scintillation counters allow for triggering on muons with p_T down to 3 GeV (the A layer drift tubes and scintillation counters also allow for triggering on muons that do not penetrate the toroid magnet). The muon system drift tubes are shown in Fig. 2.8.

Directly below the DØ detector, the support structure and the readout electronics causes the muon system to have only partial coverage in this region. The forward C layer of scintillation detectors are shown in Fig. 2.9. The overall momentum resolution, including information from the silicon microvertex tracker and central fiber tracker, is defined by the central tracking system for muons with momentum up to approximately 100 GeV. The muon spectrometer improves the resolution only for very high energy muons.

2.6 Luminosity Monitoring

The number of observed events N_{class} for a certain class of process in a collider is given by,

$$N_{\text{class}} = \epsilon A \sigma_{\text{class}} \int \mathcal{L} dt, \tag{2.3}$$

where A is the acceptance, σ_{class} is the cross section for the process and ϵ the probability to record the event if it is within the acceptance region (i.e. the efficiency). For a cross section measurement e.g. $t\bar{t}$ production, the efficiency and acceptance for $t\bar{t}$ events (and background) is calculated and the only unknown in Eq. 2.3 is $\sigma_{\text{class}=t\bar{t}}$ and the proportionality factor called *instantaneous luminosity* $\mathcal{L}$. The luminosity is defined by the beam parameters of the Tevatron accelerator e.g. the number of protons and anti-protons in each bunch, the bunch crossing frequency, the lateral bunch size, the bunch overlap in the collision region etc.

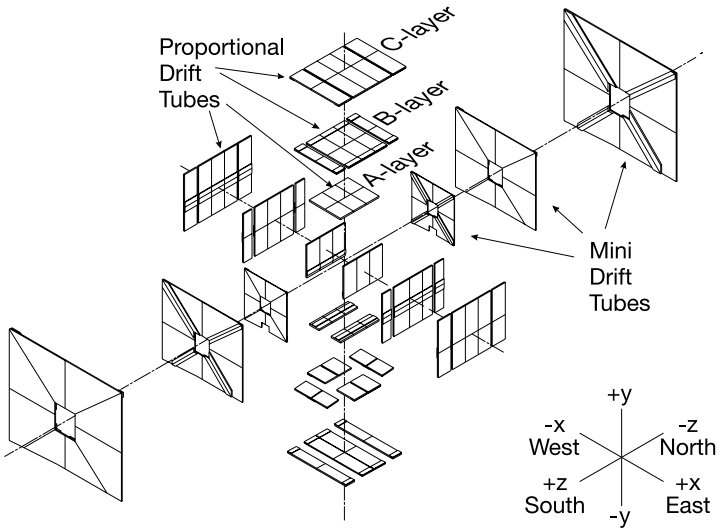

Figure 2.8. Exploded view of the wire chambers in the muon system [31].

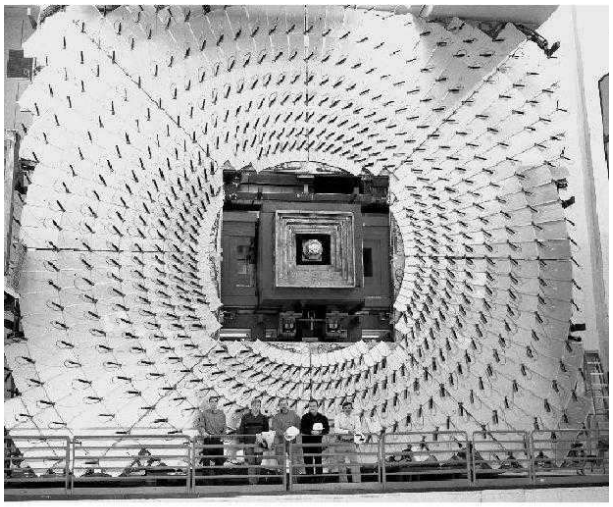

Figure 2.9. Photograph of the forward C layer scintillator counter [31].

From Eq. 2.3 the instantaneous luminosity can be calculated by counting the number of observed events for a process with known cross section. At DØ [32], the process used is the inelastic $p\bar{p}$ cross section $\sigma_{p\bar{p}}$, i.e.

$$\mathcal{L} = \frac{1}{\epsilon A \sigma_{p\bar{p}}} \frac{dN}{dt}. \tag{2.4}$$

The inelastic $p\bar{p}$ cross section has been measured by several experiments [33] to be $\sigma_{p\bar{p}} = 60.7 \pm 2.4$ mb. The detectors used for counting the interaction rate are the Luminosity Monitors (LM) consisting of two arrays of 24 plastic scintillators located at $z = \pm 140$ cm from the center of the detector covering the region $2.7 < |\eta| < 4.4$ as shown in Fig. 2.10. They are attached to the inner part of the cryostat housing the end cap calorimeters, see Fig. 2.2. In addition to measuring the luminosity, the LM is used to provide a fast measurement of the position of the primary interaction vertex, used by the fast Level 1 trigger. The time-of-flight difference between particles in the $p\bar{p}$ collision is calculated with a resolution of ≈ 0.3 ns and the resolution of the vertex position in the z-direction is < 10 cm [28].

2.7 The Trigger System

Bunch crossings at the Tevatron occur at 2.5MHz rate. This immense rate is needed as the overwhelming majority of $p\bar{p}$ encounters result in collisions of little interest.

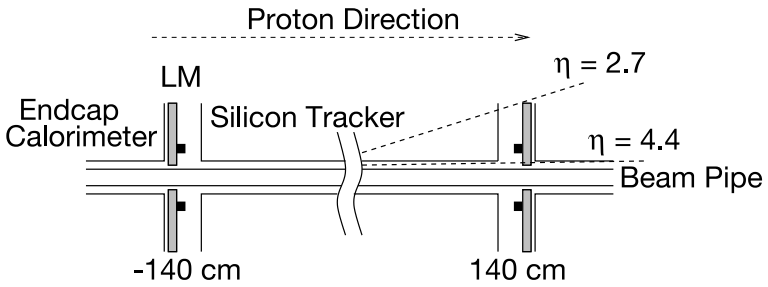

Figure 2.10. Schematic view of the location of the luminosity monitoring detectors.

The production of heavy objects like the top quark, W or Z bosons or collisions that could indicate the existence of New Physics processes occur at very low rate. The trigger system allows for a fast event-by-event decision on whether or not the collision was interesting and reduces the output rate to 50 Hz, more suitable for writing to disk.

The trigger is a 3-tiered system where each tier (Level 1, Level 2 and Level 3) investigates the event in larger detail than the preceding one and restricts the amount of data sent to the next level. An event can fail the trigger because: It did not fulfill the trigger requirements and was declared uninteresting, it was mistaken for an uninteresting event (trigger inefficiency) or the trigger system was busy processing previous events (dead time).

The Level 1 (L1) Trigger is built from specialized hardware investigating every event for interesting features. Pipelines mounted on the front-end electronic boards allows for an event decision in 4.2 μs. The L1 trigger receives input from several subdetectors: The calorimeter L1 trigger looks for patterns of large transverse energy deposits in special trigger towers[2], the Central Fiber Tracker L1 trigger looks for tracks exceeding predetermined thresholds in transverse momentum, the L1 muon trigger searches for muon candidates with a matched track from the CFT,

[2]Due to noise considerations not all trigger towers are used in the L1 calorimeter trigger.

an indication of the collision point is given by the luminosity monitors. The L1 reduces the rate from 2.5 MHz to about 2 kHz.

The level 2 (L2) trigger stage has an event decision time of approximately 100 μs and further analyzes the event by two stages upon a L1 trigger accept: A preprocessing stage that analyzes the data into simple physics objects e.g. track clusters, and a global stage that combines trigger information from different subdetectors e.g. matching tracks from the inner detector to electromagnetic clusters in the calorimeter. The output rate of the L2 trigger is about 1 kHz.

The last trigger level, Level 3 (L3), is a software trigger which reduces the event rate from to 50 Hz to allow for writing the interesting events to disk for later offline processing. The L3 trigger is fully programmable using algorithms based on complete physics objects which are as sophisticated as those available during the offline reconstruction phase. A L3 decision is based on full event information with complex variables such as spatial separation between jets and electrons, invariant masses of objects, displaced tracks from the primary vertex, etc.

Chapter 3

Event Reconstruction

The data in a single event collected from the DØ detector is the immediate detector response from nearly a million detector readout channels. To find evidence for the products of the collision and measure their properties these signals needs to be processed carefully. To reduce the huge amount of data from the experiment the information is handled by a chain of sophisticated software algorithms which create and define physics objects that represent the particles originating from the $p\bar{p}$ interaction. Each algorithm is designed to identify a particular object often based on the required efficiency and purity. The analysis presented here is based on the $t\bar{t}\rightarrow l\nu jjb\bar{b}$ final state which requires identification of the primary vertex, tracks, leptons (electrons and muons), jets and their flavors and missing transverse energy $\not\!\!E_T$.

This chapter describes the event signature of top quark pair production and the most important background processes. A short description of the identification and reconstruction of the different physics objects is also given.

3.1 Event Signatures

3.1.1 Experimental Signature of $t\bar{t}$ Production

Since the top quark decays almost exclusively through $t \rightarrow Wb$, the final state of the top quark pair production can be characterized by the decay of the two W bosons. The W boson decays leptonically via $W \rightarrow e\nu$, $W \rightarrow \mu\nu$ or $W \rightarrow \tau\nu$ with a branching fraction of $\approx 11\%$ each or to hadrons with $\approx 67\%$. The decay modes of the W bosons are reflected in the experimental search channels:

- **All jets channel**
  Both W bosons decay hadronically into $q\bar{q}$ pairs and the final state is characterized by two b-quark jets and at least four jets from the hadronization of the $q\bar{q}$ pairs. No significant $\not\!\!E_T$ is expected. This channel has the largest branching fraction but suffers from large multijet backgrounds.

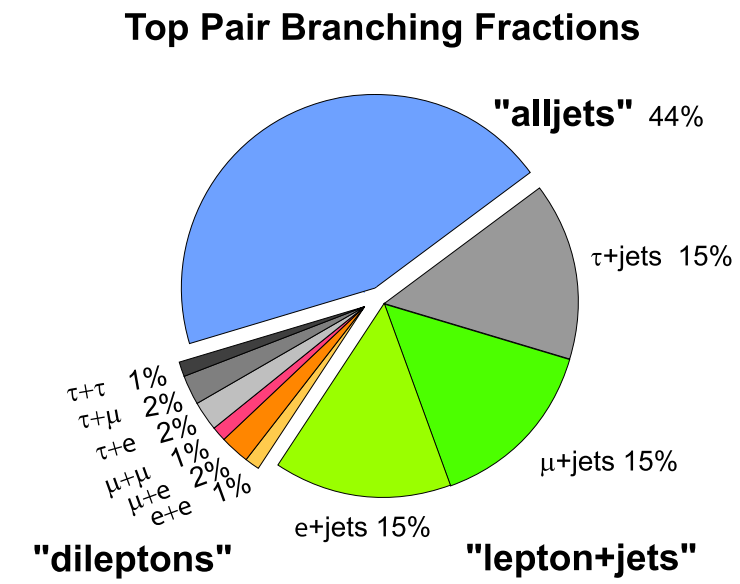

Figure 3.1. Schematic view of the characterisation of the top quark pair decay channels and their branching fractions [34].

- **lepton-plus-jets channels**
  One W boson decays hadronically and the other leptonically. The final state is characterized by two b-quark jets, at least two jets from the $q\bar{q}$ pair, one charged lepton and significant $\not\!\!E_T$ due to the neutrino from the leptonically decaying W boson. This decay chain provides a clean signature of a single isolated lepton with high transverse momentum and large $\not\!\!E_T$. Together with the large branching fraction this channel is most promising for measurements of top quark properties and is also the one used to determine the electric charge of the top quark in the the present analysis. This channel is also referred to as e+jets and μ+jets separately depending on the flavor of the charged high transverse momentum lepton or $t\bar{t} \rightarrow \ell$+jets collectively.

- **dilepton channels**
  Both W bosons decay leptonically. The final state is characterized by two b-quark jets, two charged leptons and large $\not\!\!E_T$. These channels has an excellent signal-to-background ratios but suffer from small branching fractions.

The top quark pair decay channels and their branching ratios are summarized in Fig. 3.1. Note that the top quark pair analysis in DØ includes the leptonically decaying τ in the ℓ+jets and dilepton channels since this gives a similar experimental signature.

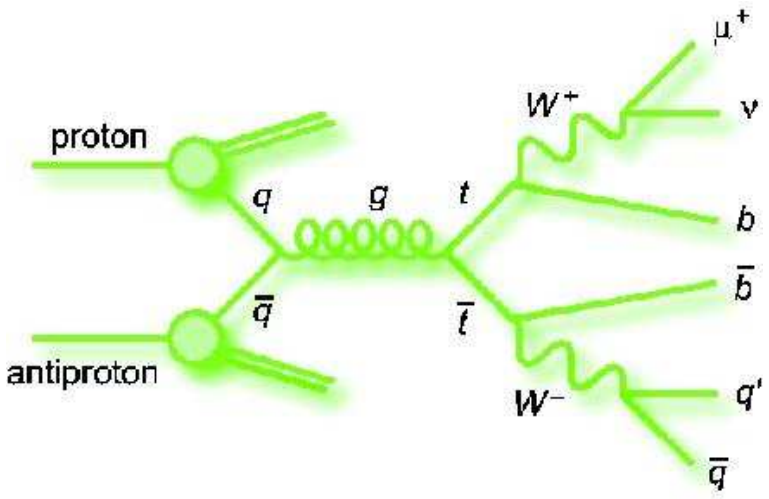

Figure 3.2. A sketch of a $t\bar{t} \rightarrow \mu$+jets event [34].

Figure 3.2 shows a schematical view of a μ+jets event. Additional jets can be produced in all channels due to initial (ISR) and final state radiation (FSR) discussed below.

In summary, the main feature of a $t\bar{t} \rightarrow \ell$+jets event is the presence of one charged isolated lepton produced centrally in the detector with high transverse momentum, a neutrino with comparable momentum giving rise to a significant $\not\!\!E_T$ and several jets.

3.1.2 Background Signature

At hadron colliders, QCD multijet production has a large cross section and is initially the largest background. This strong production of jets contain no genuinely isolated leptons nor missing transverse energy. However, these can be faked by instrumental effects. A jet fluctuating to a high electromagnetic content can fake an electron. In addition, semi-leptonic decay modes of b- and c-quarks can give rise to fake isolated leptons if the associated jet is not reconstructed, either due to a low energy deposition in the calorimeter or inefficiency of the jet reconstruction. In combination, fake missing transverse energy can arise due to unreconstructed jets, the neutrino from the heavy flavor decay or a mismeasurement of the lepton momentum. However, even an unreconstructed jet deposits a small amount of energy in the calorimeter and gives signal in the tracking detectors. Thus, a good handle to suppress this background is to require that the lepton is isolated from other activity in the detector.

The main source of W boson production at the Tevatron is due to quark antiquark fusion. In this process gluon radiation from the incoming quark lines (ISR)

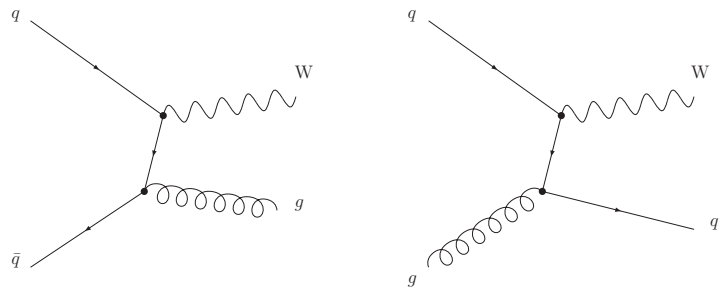

Figure 3.3. Examples of Feynman diagrams for the production of W bosons with one additional parton in the final state, called W+1jet.

can give the W boson transverse momentum and add more partons to the final state. Partons are a common used notation, remaining from the 70's, for the constituents of a hadron (quarks and gluons). Figure 3.3 shows examples of W production with an additional parton in the final state, called W+1jet[1]. Processes leading to even higher number of partons in the final state can be produced by gluon radiation of the quark and gluon lines in the initial or final state. Computation of W boson production in association with up to four partons in the final state has been performed to leading order [35] with different techniques to handle the immense amount of Feynman diagrams contributing to the process.

Due to the similar experimental signature as $t\bar{t} \rightarrow \ell$+jets events, the production of W (with subsequent leptonic decay) in association with four partons in the final state is the dominant background in this analysis after standard preselection of W boson candidates. Even though the objects in the final state are the same as in a $t\bar{t} \rightarrow \ell$+jets event, there are significant differences in several aspects of the event that can be exploited to separate $t\bar{t}$ events from this background. The analyzes measuring the $t\bar{t}$ production cross section utilize the fact that; (*i*) W bosons produced from the decay of top quarks have on average larger transverse momentum and are produced at lower $|\eta|$ (*ii*) the additional jets are mainly produced by gluon radiation, resulting in jet with lower transverse momenta and at higher $|\eta|$ [36]. Another way to discriminate between a $t\bar{t}$ and a W+jets event is to exploit the fact that a $t\bar{t}$ event has a higher fraction of b-quarks in the final state. The b-quarks in a $t\bar{t}$ event hadronize into B hadrons and the event is expected to have at least two heavy flavor jets originating from b-quarks (see Sec. 3.6.4). As discussed in more detail in Sec. 4.2, the dominant background in the $t\bar{t} \rightarrow \ell$+jets channel after the requirement of at least two jets identified as b-quark jets is the production of a W boson in association with four jets out of which two are b-quark jets (denoted as $Wb\bar{b}jj$).

[1]Here the additional partons in the final state are assumed to hadronize into jets.

3.2 Tracks

Charged particles traversing the inner detector deposit energy in the silicon layers of the SMT and produce scintillation light in the CFT[2]. The hits in the different inner detector layers together with the bending in the magnetic field allows for the reconstruction of the particle's trajectory. The track reconstruction algorithm groups together hits in different detectors into clusters which are then fitted to find a possible physical path of the particle [37].

3.3 Primary Vertex

In proton anti-proton collisions several interactions between the constituent partons are possible. The primary interaction vertex (PV) are the point were the hard scattering (high transverse momentum) interaction took place. The spread of the interaction point in the $(x-y)$ plane, transverse to the beam line, is small due to the transverse size of the Tevatron beam which is of the order of 30 μm. In z-direction, the spread of the PV position extends up to 60 cm following a Gaussian distribution with a width of approximately 25 cm. Finding the primary vertex (PV) is crucial for all b-tagging algorithms and in order to determine if a lepton originates from the PV.

The PV algorithm [38] used by DØ starts by fitting all reconstructed tracks to a common vertex and removes bad track fits until a predefined value of the goodness of fit is reached. The same procedure is repeated for the tracks that were removed until all tracks are assigned to a PV. There are two similar implementations of the PV algorithm ,*DØ reco* and *DØ root*, with the difference that *DØ root* has an additional step of clustering tracks in the z-direction and slightly tighter track selection criteria (the d_{ca} significance is required to be ≤ 3.0 compared to ≤ 5.0). In both algorithms, only tracks with $p_T > 0.5$ GeV are considered and at least two hits in the SMT detector. If more than one PV is found, the hard scatter vertex is selected by observing that hard scatter vertices have on average tracks with larger transverse momentum associated to it than minimum bias vertices [39].

The performance of the PV selection algorithms are comparable. There are on average 20 tracks in a generic QCD multijet event and the average PV reconstruction efficiency is 98%. This efficiency is about 100% in the central $|z|$ region of the SMT fiducial region ($|z| < 36$ *cm* for the barrel) and drops quickly outside of this region due to the requirement of at least two SMT hits for tracks forming the PV.

3.4 Muons

Muons are identified in the drift chambers and scintillation counters by matching hits in the layers on either side of the toroid magnet. The DØ muon group has

[2]In reality, tracks do not always have hits in all layers of the SMT and CFT.

established a set of standard muon identification criteria applied to the candidate muon [40]:

- At least two A layer wire hits,
- at least one A layer scintillator hit,
- at least two BC layer wire hits,
- at least one BC scintillator hit(except for central muons with less than four BC wire hits),
- to be inconsistent with a cosmic muon based on timing information from the scintillator hits.

A muon identified in the above fashion is the basis for the muon reconstruction. The superior track resolution of the central tracker (SMT and CFT) is used to improve the muon's momentum resolution. Therefore, in addition to the above criteria, a track consistent with originating from the PV is required to be spatially matched to the muon candidate.

Muon tracks with no hits in the SMT (which have been shown to have a worse fit and thus a worse resolution) are re-fitted constraining the muon track to the PV in order to improve their momentum resolution.

The muon momentum scale and resolution was determined by reconstructing the Z boson invariant mass peak in $Z \rightarrow \mu^+\mu^-$ events. Comparison of the invariant mass peak in data and simulation reveals a significantly better resolution in the simulation than in data as well as a shifted peak position. This is accounted for by smearing the reconstructed muon momenta in simulated events to match the resolution in data [41].

3.5 Electrons

The ability to identify and reconstruct high p_T electrons is essential for many analyzes, including top quark measurements, electroweak processes and searches for New Physics. Being charged particles, electrons deposit energy in the central tracking detectors before showering predominantly in the EM section of the calorimeter. The main backgrounds to reconstructed true electrons (so-called "fake" electrons) are:

- π^0 showers overlapping with a track from a charged particle,
- photons which convert to e^+e^- pairs,
- $\pi^\pm$ which undergo charge exchange in the detector material,
- fluctuations of hadronic showers.

At DØ electron identification involves three steps. First, electron candidates are searched for by looking for clusters in the EM calorimeter. Secondly, a track in the central tracking system that is spatially matched to the EM cluster is searched for and finally the electron has to pass a likelihood test based on shower shape variables. To handle all the sources of backgrounds while keeping the efficiency to reconstruct real electrons high, several variables are used:

- The fraction of energy deposited in the electromagnetic part of the calorimeter is required to be above 90% of the total deposited energy in the calorimeter inside the cone of $\Delta R < 0.2$.
- Electrons tend to be isolated from other activity in the calorimeter. Therefore, at most 15% of the energy of the cluster is allowed to be deposited in a hollow cone ($0.2 < \Delta R < 0.4$) around the electron's direction.
- The shower shape of candidate EM clusters is compared to the expected shape from electrons [42].
- A track is required to point to the EM cluster.
- A seven parameter likelihood is built that rejects background-like EM candidates [43].
- The electron candidate is required to be in the central calorimeter, since the fake electron background is not completely understood in the end cap calorimeters.

The electron momentum scale and resolution was studied by reconstructing the Z boson invariant mass peak in $Z \to e^+e^-$ events. The comparison of data with simulation further revealed a higher resolution in the simulation and a corresponding scale factor and smearing is applied to simulated electrons to reproduce the measured quantities. Detailed information on the selection criteria, electron momentum scale and resolution can be found in Ref. [44].

3.6 Jets

In the analysis presented in this thesis, jets form an essential ingredient in the event selection. Each event is required to have at least four jets out of which two are identified as b-quark jets. This section describes the identification and energy calibration of jets and explains the identification of jets originating from the hadronization of b-quarks.

3.6.1 Jet Identification

Jets are reconstructed based on finding calorimeter towers with an energy above a predefined threshold of $E_T > 0.5$ GeV which are further collected into clusters which

form candidate jets [45]. The jet cone size is $\Delta R < 0.5$ and the uncalibrated transverse energy reconstruction threshold is 8 GeV. Before any calorimeter physics object (jets, electrons, photon or $\not\!\!E_T$) are reconstructed a special calorimeter algorithm is applied to remove measured cell energies likely to arise from noise [46; 47; 48].

After jets have been reconstructed it is important to reject those that are poorly reconstructed or are electrons or photons mis-identified as jets. Therefore, several additional requirements based on the expected properties of jets are applied: The probability for a jet to deposit a large fraction of its energy in the coarse hadronic section of the calorimeter is low (and the coarse hadronic section is subject to a higher noise level due to its larger cell size) and the energy deposition in this section is required to be below 40%. Electrons and photons typically deposits all of their energy in the EM section and therefore the fraction of energy deposited in the EM part for jets is required to be between 5% and 95%. Jets reconstructed from few or single cells containing a large fraction of the jets total energy are likely to be fake jets due to noise in the cell. Therefore, jets are rejected if one cell contains above 90% of the total energy or the ratio of the cell with the highest energy to the next-to-highest is above 10. Since the Level 1 calorimeter trigger has an independent readout chain it has been proven to be a powerful handle to reject fake jets due to noise in the precision readout electronics. Therefore, the sum of energy in the Level 1 calorimeter trigger towers is required to be above 24-40% (12-20%) in the central and end cap calorimeter (inter-cryostat) regions depending on the fraction of total energy deposited in the coarse hadronic section. If a jet overlaps with EM candidates (which are also reconstructed as jets if their energy is above 8 GeV) an ambiguity appears about which energy scale that should be applied. The solution is to reject jets that overlap with electrons or photons within $\Delta R < 0.5$. The minimum transverse momentum after scale and corrections described below is required to be 15 GeV.

3.6.2 Jet Energy Scale

Quarks and gluons from the hard interaction hadronize into jets and deposits energy in the calorimeter as shown schematically in Fig. 3.4. The goal of the Jet Energy Scale (JES) is to correct the measured jet energy in the calorimeter back to the stable-particle energy before interacting with the calorimeter. Various effects cause the measured energy to be different from the particle jet e.g. the use of a sampling calorimeter, noise and dead-material.

The JES is derived using γ+jet events in a back-to-back configuration [50]. Using the fact that the electromagnetic energy scale is known to high precision[3] the JES can be extracted from the transverse momentum imbalance in such an event.

The JES is divided into different subcorrections applied in order of appearance to jets: The **offset correction** corrects for energy not part of the hard scatter

[3]The EM energy scale can be calibrated by calculating the invariant mass of electrons or photons in inclusive samples such as $Z \to e^+e^-$, $J/\Psi \to e^+e^-$ and $\pi^0 \to \gamma\gamma$.

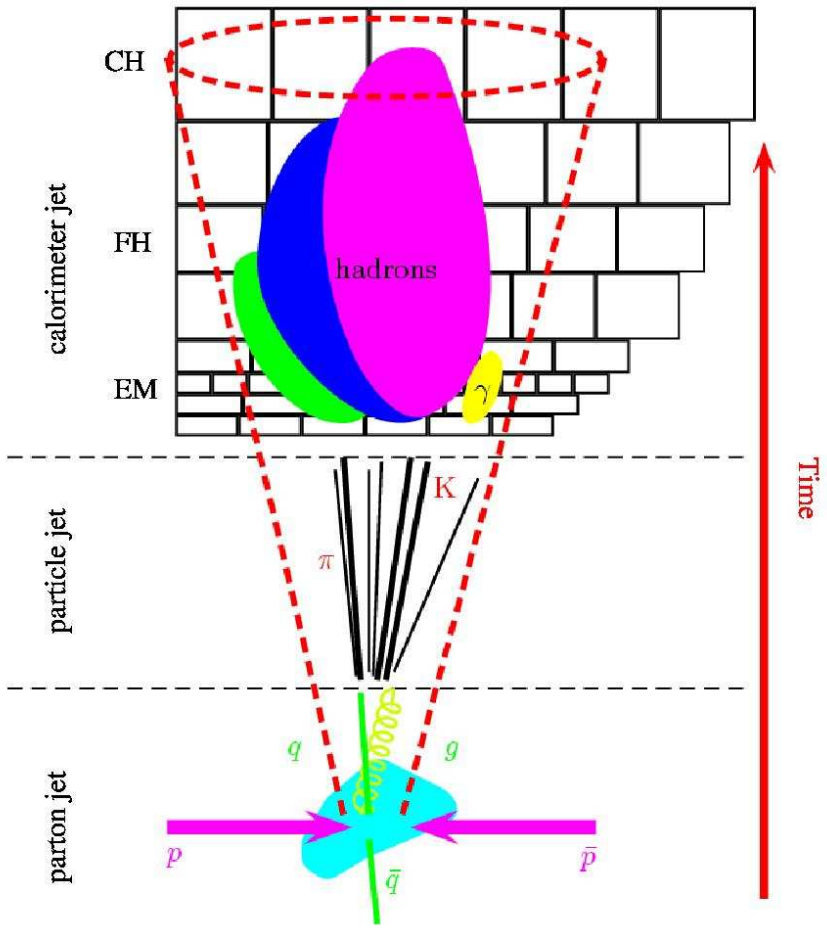

Figure 3.4. Schematic view of the process taking a parton from the hard scattering to energy deposited in the calorimeter [49].

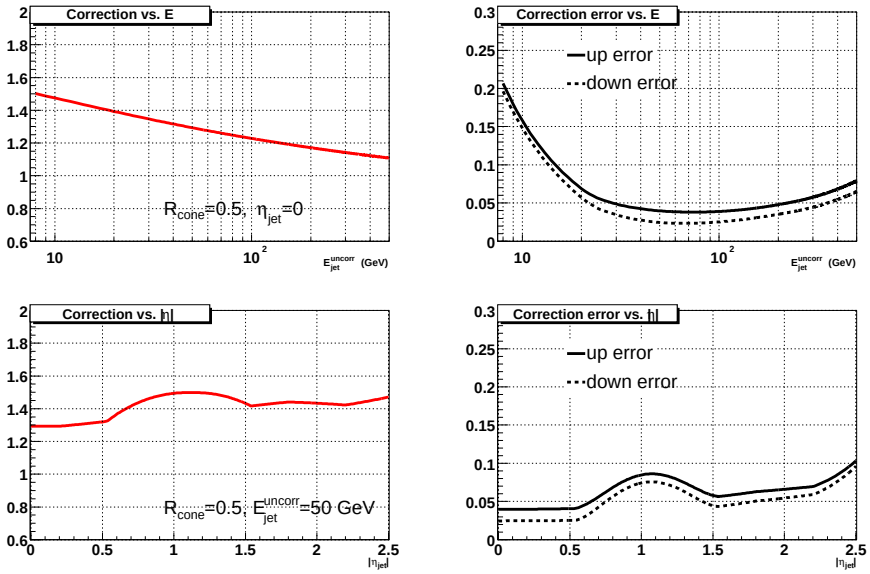

Figure 3.5. The JES corrections for data (left) and the error (right) as a function of the measured jet energy (top) and η (bottom). The "up" and "down" error are the combined undertainties from all the sub-corrections [49].

(detector and electronic noise, pile-up and energy from the underlying event, see Sec. 3.8 for a description of pile-up and underlying event). This luminosity dependent offset correction is calculated in data using events triggered by the luminosity monitors, signaling a possible inelastic $p\bar{p}$ collision (minimum-bias events). The **response correction** is a correction for the non-uniform response of parts of the detector (dominated by the ICD region) and an absolute response for a uniform treatment of jet energies derived from well defined γ+jet events in a back-to-back configuration. The last correction is the **showering correction** which attempts to correct for particles inside the jet that deposits their energy outside the jet cone (or the reverse process). This correction is extracted from jet profiles in γ+jet events.

The final result is a JES correction factor shown in Fig. 3.5 and 3.6 for data and simulation respectively.

3.6.3 Jet Energy Resolution

The Jet Energy Resolution (JER) can be determined by studying the p_T imbalance in dijet events [51]. The asymmetry variable $A = \left(p_T^{\text{jet}_1} - p_T^{\text{jet}_2}\right) / \left(p_T^{\text{jet}_1} + p_T^{\text{jet}_2}\right)$ is calculated and the jet resolution can be inferred from the width of the asymmetry

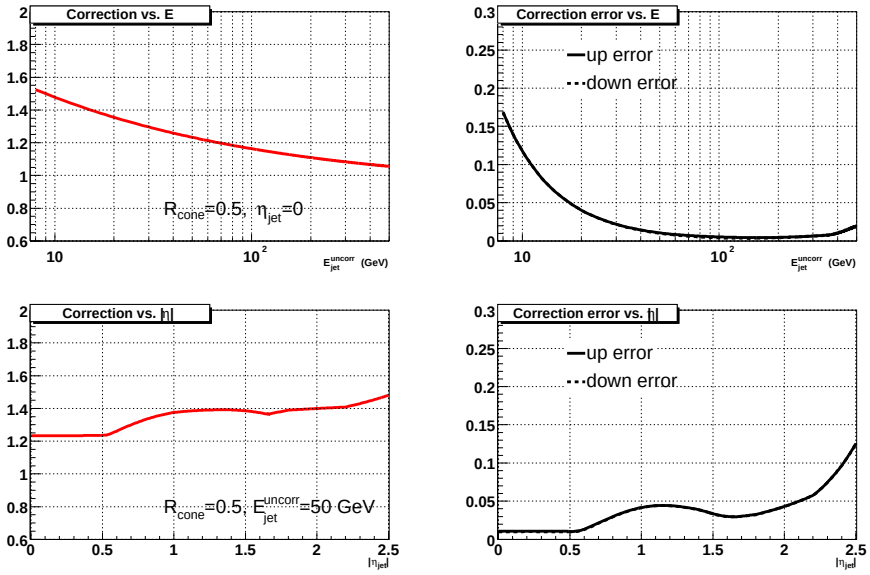

Figure 3.6. The JES corrections for simulated jets (left) and the error (right) as a function of the measured jet energy (top) and η (bottom). The "up" and "down" error are the combined undertainties from all the sub-corrections [49].

$\lvert\eta_{\text{det}}\rvert$ range	N	S	C
$0.0 < \lvert\eta_{\text{det}}\rvert < 0.5$	5.05	0.753	0.0893
$0.5 < \lvert\eta_{\text{det}}\rvert < 1.0$	0.0	1.20	0.0870
$1.0 < \lvert\eta_{\text{det}}\rvert < 1.5$	2.24	0.924	0.135
$1.5 < \lvert\eta_{\text{det}}\rvert < 2.0$	6.42	0.0	0.0974

Table 3.1. The JER parameters for data.

$\lvert\eta_{\text{det}}\rvert$ range	N	S	C
$0.0 < \lvert\eta_{\text{det}}\rvert < 0.5$	4.26	0.658	0.0436
$0.5 < \lvert\eta_{\text{det}}\rvert < 1.0$	4.61	0.621	0.0578
$1.0 < \lvert\eta_{\text{det}}\rvert < 1.5$	3.08	0.816	0.0729
$1.5 < \lvert\eta_{\text{det}}\rvert < 2.0$	4.83	0.0	0.0735

Table 3.2. The JER parameters for simulated jets.

σ_A as,

$$\frac{\sigma_{p_T}^{\text{jet}}}{p_T^{\text{jet}}} = \sqrt{2}\sigma_A. \tag{3.1}$$

At lower energies (< 50 GeV) $\gamma + jet$ events are used.[4]. The asymmetry variable A_{pj} is calculated as $A_{pj} = \left(p_T^{\gamma} - p_T^{\text{jet}}\right)/p_T^{\gamma}$ and the JER can be expressed as,

$$\frac{\sigma_{p_T}^{\text{jet}}}{p_T^{\text{jet}}} = \sigma_{A_{pj}} \times \frac{p_T^{\gamma}}{p_T^{\text{jet}}}. \tag{3.2}$$

The JER measured using the procedure above is then parametrized as (see Sec. 2.4),

$$\frac{\sigma_{p_T}^{\text{jet}}}{p_T^{\text{jet}}} = \sqrt{\frac{N^2}{p_T^2} + \frac{S^2}{p_T} + C^2}, \tag{3.3}$$

and the results for data and simulation are shown in Tab. 3.1 and Tab. 3.2 respectively. More details can be found in [51]. The resolution is clearly better in the simulation and simulated jets are therefore smeared to match the resolution in data.

3.6.4 *b*-Quark Jets

Jets can further be classified by their flavor. A jet originating from the hadronization of a gluon (g), u-, d- or s-quark or from the a c- or b-quark is referred to as a light jet or a heavy flavor jet respectively. There are two techniques to distinguishing between light and heavy flavor jets: **Soft Lepton Tagging** uses the

[4]This because the triggers used to select dijet events are inefficient for jet energies below 50 GeV.

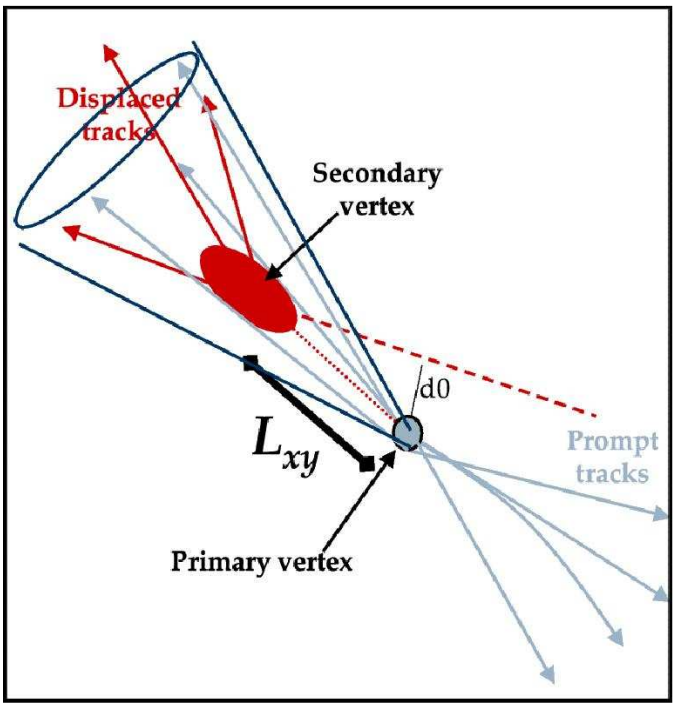

Figure 3.7. A quantitative description of a jet with a secondary vertex arising from the long lifetime of a B hadron. Several displaced tracks are fitted to the secondary vertex which can be used to identify b-quark jets. d_0 is the distance of closest approach of a track with respect to the primary vertex.

presence of a lepton within a jet as a signature for a semi-leptonic decay of a heavy flavor hadron. The branching fraction for a semi-leptonic ($\mu^{\pm}$ or $e^{\pm}$) B hadron decay is $\approx 10\%$ for each lepton mode [5]. **Lifetime Tagging** uses the the fact that hadrons containing a b-quark have a lifetime of approximately 1.6 ps. A B hadron originating from the primary vertex will therefore travel a significant distance from the primary vertex before decaying (the average flight length is around 3 mm for 40 GeV hadron). Approximately 70% of the B mesons with a decay length [52] greater than 1 mm have more than 2 displaced tracks from its decay products with an impact parameter significance above three [53]. The lifetime tagging algorithm therefore searches for tracks significantly displaced from the primary vertex as shown schematically in Fig. 3.7. In this analysis both ways to identify heavy flavor jets are used. A jet tagged by the lifetime tagging algorithm is called a "b-tagged jet" or simply a "tagged jet". A jet tagged by the soft lepton tagging algorithm is called a "μ-tagged jet". Due the difficulty of identifying an electron within a jet, the soft lepton tagging algorithm uses only the muonic semi-leptonic decay mode. Note that a b-tagged jet is not necessarily a jet originating from the hadronization of a b-quark or even a c-quark. The tagging technique may wrongly

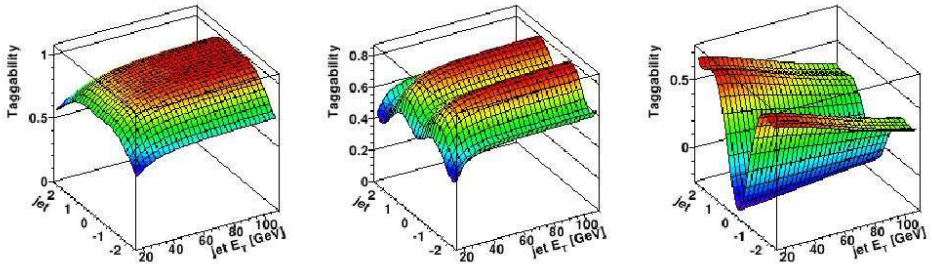

Figure 3.8. Taggability parameterizations as a function of jet E_T and η for the combined preselected data sample in three different regions of z_{PV}: $|z_{PV}| < 30$(left), $30 < |z_{PV}| < 45$(middle) and $45 < |z_{PV}| < 60$(right).

tag light jets as discussed below in more detail. In this thesis a jet from a x-quark ($x = b, c, s, u, d$) is called x-quark jet[5] or gluon jets for gluon initiated jets.

To separate the performance of the different algorithms from detector effects, such as calorimeter noise and tracking inefficiencies, the probability for a jet to be tagged using lifetime tagging is broken down into two components: *(i)* The probability for a jet to be taggable (also called *taggability*) and *(ii)* the probability for a taggable jet to be tagged (also called *tagging efficiency*). The DØ b-identification group [53] defines a taggable calorimeter jet as a jet reconstructed in the calorimeter matched to a track-based jet (track-jet) within $\Delta R < 0.5$. A track-jet is defined by the following track requirements: $p_T > 0.5$ GeV (where at least one track has $p_T > 1$ GeV), at least one hit in the central part of the SMT detector and matched to the primary vertex within 0.5 in $r - \phi$ space and within 2 cm in the z-direction.

The taggability is determined directly from data and parameterized as a function of the jet p_T and η. Due to the requirement of hits in the central region of the SMT the taggability is expected to have a large dependence on the z-position of the primary vertex. The taggability is shown in Fig. 3.8 for the preselected data sample (see Sec. 4.2 for the sample definition).

There are several ways of tagging a jet using lifetime tagging [53]. In this analysis the Secondary Vertex Tagger (SVT) algorithm is used. The SVT algorithm finds tracks with large impact parameters within the matched track-jets and reconstructs secondary vertices that are within $\Delta R < 0.5$ of the calorimeter jet direction. A calorimeter jet is tagged by the SVT algorithm (SVT-tagged) if it has at least one secondary vertex with decay length significance [52] greater than 7.0. More information on the exact definition of the SVT tagging algorithm can be found in Ref. [54] and on the selections used in this analysis in Ref. [55]. The simulation is unable to describe the details of the tracking and is thus overestimating the tagging

[5]Here the notation for a jet originating from e.g. a b or $\bar{b}$-quark is called b-quark jet collectively. In later sections the difference between a b or $\bar{b}$-quark jet is exploited and the notation will be obvious.

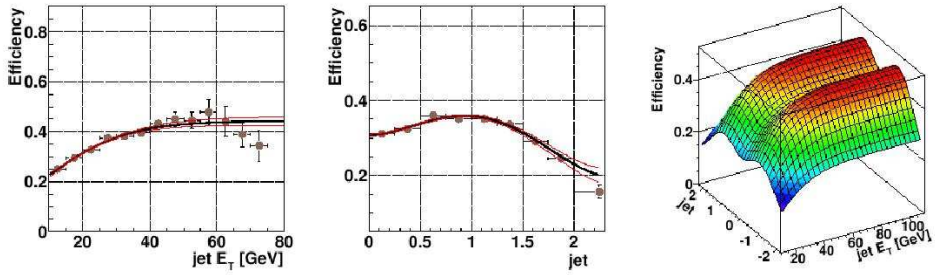

Figure 3.9. The semi-leptonic tagging efficiency in data as a function of jet p_T (left), η (middle) and the combined two-dimensional parametrization (right) [56].

efficiency for simulated jets. Therefore the tagging efficiency is measured on data with as little input as possible from simulation.

To determine the SVT tagging efficiency for b-quark jets in data, a sample enriched in b-quark jets is selected by requiring at least one μ-tagged jet in the event. The measured SVT tagging efficiency of jets in this sample is called "semi-leptonic" and is parametrized as a function of jet p_T and η as shown in Fig. 3.9. In order to find the tagging efficiency for inclusive b-quark jets in data a scale factor is derived. This scale factor is the ratio of the measured semi-leptonic tagging efficiency in data and the semi-leptonic tagging efficiency in simulated events. The tagging efficiency for inclusive jets in the simulation is then multiplied with the scale factor to find the tagging efficiency for inclusive b-quark jets in data. Assuming that the scale factor for b- and c-quark jets are the same, the c-quark jet tagging efficiency in data can be calculated in a similar way and varies between 7-12%. Due to the limited track resolution and/or mis-reconstructed tracks, light jets can be wrongly tagged. The probability for a light jet to be tagged ("mistag rate") is derived on a QCD data sample (which dominantly consists of light jets) and parametrized again as function of jet p_T and η. The mistag rate is of the order of 1%.

3.7 Missing Transverse Energy

In the final state of a $t\bar{t} \to \ell$+jets event the neutrino can only be detected by the measurement of the imbalance of momentum in the transverse plane. This imbalance is measured by the vector sum of all energy depositions in calorimeter cells. Cells in the coarse hadronic section of the calorimeter are excluded if they are not part of a reconstructed jet. This has been shown to improve the missing transverse energy resolution due to the higher noise level in those cells. The vector opposite to this total visible momentum is the raw missing transverse energy ($\not\!\!E_{T\,raw}$).

The calorimeter response to electromagnetic particles such as electrons and photons is different than for hadronic particles, and in particular jets. This difference

propagates directly to the $\not\!\!E_{T\,\mathrm{raw}}$ if the energy depositions are not calibrated correctly. All jets are corrected for the jet energy scale and the correction is also propagated to the $\not\!\!E_{T\,\mathrm{raw}}$. The same procedure is used for the EM calibration. After all corrections, the resulting missing transverse energy is called the "calorimeter missing transverse energy".

Muons are minimal ionizing particles and traverse the whole detector if the muon momentum exceeds a few GeV and their measured transverse momentum must be added to the missing transverse energy. The muon deposits only a small fraction of its total energy in the calorimeter which is estimated and subtracted from the missing transverse energy vector. This fully corrected missing transverse energy is denoted simply as $\not\!\!E_T$.

3.8 Monte Carlo Simulation

The detailed study of proton anti-proton collisions requires a detailed understanding of all aspects of the event. Monte Carlo simulation of such collisions are necessary to provide this knowledge.These simulations includes the hard scattering interaction, hadronization, detector response and digitization, allowing for a detailed comparison between simulated events and data.

The hard scatter interaction is modeled using ALPGEN 1.3.3 [35] that calculates the leading order matrix element. The set of parton distribution functions used is CTEQ5L [57].

The complexity of $p\bar{p}$ collisions is due to the fact that the colliding (anti-)protons are composite states of many partons. As a consequence, an interesting hard (high transverse momentum) interaction is accompanied by what is called the **underlying event**. It consists of the *beam-beam remnants*, which are what is left over after a parton has been knocked out of the initial incoming (anti-)proton, and *multiple parton interactions* where in addition to the hard scattering interaction there are one or more semi-hard interaction in the same event. A study of the transverse momentum distribution of charged particles has led to the so-called "Tune A" [59] tuning of the Pythia underlying event model parameters to better describe data.

In each bunch crossing of protons and anti-protons more than one hard scatter $p\bar{p}$ collision may occur. These **multiple interactions** are modeled by superimposing simulated events with minimum bias data. This data is collected by selecting real collider events with minimum activity in the detector i.e. not triggered by a high transverse momentum lepton, jet or missing transverse momentum. In addition $p\bar{p}$ collisions from consecutive bunch-crossings can sometimes be reconstructed in the same event. This process, called **Pile-up**, is taken into account in the simulation. At long distances, QCD becomes non-perturbative and in this domain the colored partons are transformed into colorless hadrons, a process called **fragmentation**[6]. There exist several phenomenological models describing this process,

[6]The fragmentation and decay of particles are often collectively called hadronization.

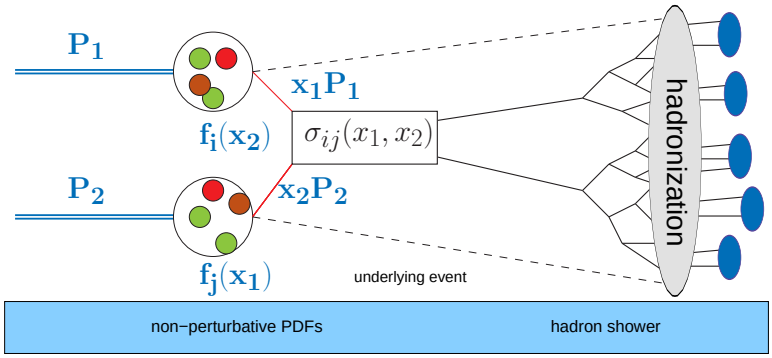

Figure 3.10. A schematic view of a $p\bar{p}$ collision.

PYTHIA uses the Lund string model [60] for fragmentation. Heavy quark fragmentation is an important aspect in this analysis and two models (BOWLER [61] and PETERSON [62] fragmentation) with alternative heavy quark fragmentation schemes are used as a cross-check in the analysis presented in this thesis. The decay of B mesons are handled by EVTGEN [63] and other decays are handled by PYTHIA. The important simulation of the **detector response** to charged and neutral particles is simulated using *DØ gstar* [64], a GEANT3 [65] model describing the material and geometry of the DØ detector. The simulated signal produced by the detector is digitized using the software package *DØ sim*. From this point onwards events from the simulattion and data can be treated in a uniform way. The DØ event reconstruction software packages (*DØ reco*) transform the detector signals into reconstructed physics objects such as electromagnetic clusters, muon candidates, tracks, etc. Finally, the *TopAnalyze* [66] program processes the reconstructed events further with algorithms and object identification selections specific for the DØ top quark working group and produces ROOT [67] files. Throughout the analysis presented in this thesis, ROOT is used as analysis tool.

3.8.1 Simulated Samples

The Monte Carlo samples used in the analysis are listed in Tab. 3.3. The samples are generated with parameters given in Ref. [68] using ALPGEN as primary generator. Due to the uncertainty in the modeling of initial state radiation in $t\bar{t}$ events, a $t\bar{t}j \rightarrow \ell$+jets sample with a jet in the initial state is produced to estimate this effect.

Process	Generator	PDF	Comment
$t\bar{t} \to \ell+$ jets	Alpgen+Pythia	CTEQ5L	
$t\bar{t}j \to \ell+$ jets	Alpgen+Pythia	CTEQ5L	
$t\bar{t} \to \ell\ell$	Alpgen+Pythia	CTEQ5L	
$t\bar{t} \to \ell+$ jets	Alpgen+Pythia	CTEQ5L	Bowler fragmentation [61]
$t\bar{t} \to \ell+$ jets	Alpgen+Pythia	CTEQ5L	Peterson fragmentation [62]
$Wbbjj$	Alpgen+Pythia	CTEQ5L	
$Z \to bb$	Pythia	CTEQ5L	
$Z \to b\bar{b} \to \mu X$	Pythia	CTEQ5L	
$Z \to c\bar{c}$	Pythia	CTEQ5L	
$c\bar{c}$	Pythia	CTEQ5L	

Table 3.3. The Monte Carlo samples used in the analysis.

Chapter 4

Determination of the Electric Charge of the Top Quark

4.1 Overview of the Method

The determination of the electric charge of the top quark proceeds in three steps. First a pure sample of $t\bar{t}$ events in the ℓ+jets channel is selected in data. This final state provides a sample with a signal-to-background ratio of ~ 1. The jet charge algorithm is sensitive to poorly simulated tracking, and is therefore only applied to jets tagged by the SVT algorithm. Only events with four or more reconstructed jets are considered, two of which are required to be SVT-tagged further increasing the purity of the sample (signal-to-background ratio ~ 11). Each of the selected $t\bar{t} \rightarrow \ell$+jets events have two "legs", one with a leptonically decaying W boson ($t \rightarrow Wb \rightarrow \ell\nu b$) and one with a hadronically decaying W boson ($t \rightarrow Wb \rightarrow qq^{'}b$), see Fig. 4.1.

The second step of the analysis consists of assigning the correct jets and leptons to the two "legs" of the event. To make this assignment the same constrained kinematic fit package as for measurements of the top quark mass is used [69]. The goal of the analysis presented in this thesis is to discriminate between two hypotheses: The SM top quark charge of $+2e/3$ or an exotic quark with charge $-4e/3$. In fact, the analysis is only sensitive to the modulus of the quark's charge ($|2e/3|$ or $|4e/3|$). This limitation does not lead to any loss of information since it is always assumed that charge is conserved and every event contains one quark and one anti-quark. It also allows for an equivalent treatment of the quark and anti-quark. In each $t\bar{t}$ event, the absolute value of the charge of the quark and anti-quark are computed, which are assumed to be the same. Thus, each $t\bar{t}$ candidate event has two observables Q which are built in the following way: The first combines the charged of the lepton from the W decay and the charge of the b-quark jet associated to the leptonic side of the event by the kinematic fit. The charge of the b-quark jet is computed using a jet charge algorithm and simply added to the lepton charge

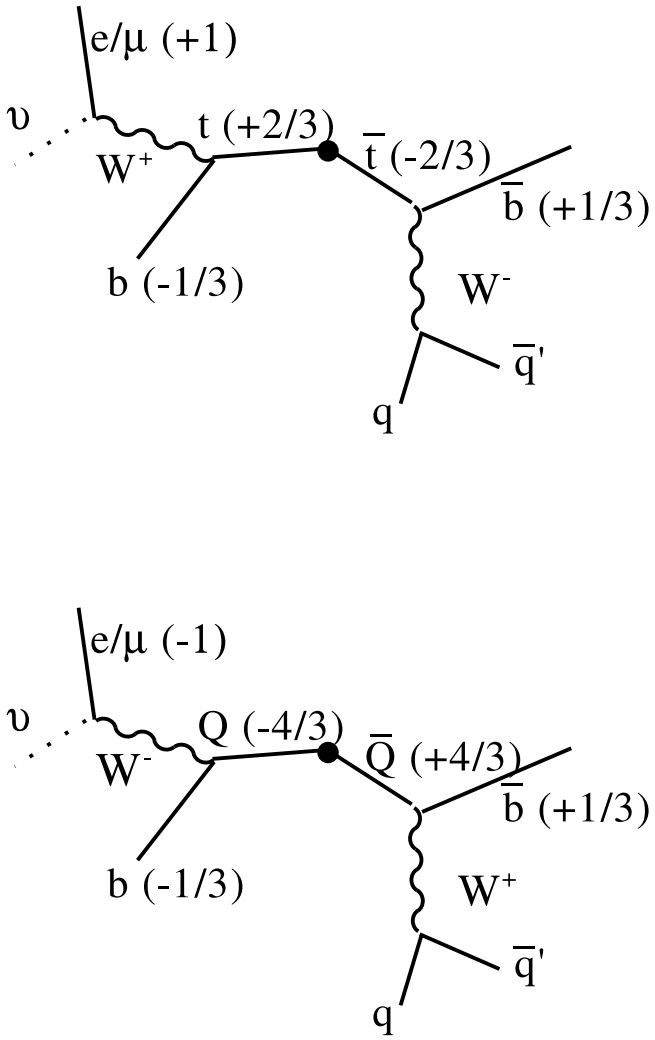

Figure 4.1. Illustration of the decay of the two different hypotheses, the SM top quark pair (top) and an exotic quark pair decay (below). Process where the lepton appears on the $\bar{t}$ or $\bar{Q}$ are also possible.

and the absolute value of this number is taken. The second observable is the charge of the second "leg" of the event. It is obtained by taking the charge of the second b-quark jet and subtracting the charge of the lepton. This procedure is almost equivalent to the doubling of the dataset size in terms of statistical sensitivity. Any mistake in the assignment of the lepton and the b-quark jet due to the kinematic fit will automatically propagate to both legs of the event, therefore the two charge measurements are not uncorrelated. Nevertheless, the measurement of the b-quark jet charge is applied twice since it improves the sensitivity of the measurement.

The jet charge distributions for b- and $\bar{b}$-quark jets are extracted from data and used to derive the expected distribution for the SM and the exotic scenarios. The distribution of Q observed in the selected $t\bar{t}$ data sample is then compared to the SM and exotic expectations for Q.

In the next section the $t\bar{t}$ event selection and the sample composition are presented. Section 4.3 is devoted to the description of the jet charge algorithm. A method is developed to derive the performance of the jet charge algorithm for SVT-tagged jets using data. This is presented in Sec. 4.4. Section 4.5 gives a detailed description of the method to discriminate between the $2e/3$ and $4e/3$ scenarios by combining the kinematic fit with the jet charge algorithm. The systematic uncertainties are reviewed in Sec. 4.6. The final result is presented in Sec. 4.7 together with an upper limit on the fraction of exotic quark pairs in the data set.

4.2 Signal Sample

As discussed earlier only a fraction of the collisions delivered by the Fermilab Tevatron is recorded. Furthermore, well known variables in the data are continuously monitored by detector experts and only the data marked by these experts as good is used. This analysis uses data collected by the DØ experiment in a period from June 2002 to August 2004. The total integrated luminosity amounts to approximately 363 pb^{-1} and 366 pb^{-1} in the μ+jets and e+jets channels respectively. The difference between the two channels is due to the different signal triggers used. In this section the selection of a data sample enriched in $t\bar{t} \to \ell$+jets events is described. The signal event signature consists of:

- One charged high transverse momentum lepton (electron or muon, either prompt or from a leptonically decaying τ),
- large missing transverse energy,
- two jets from the hadronization of the b-quarks from the top and anti-top quark decay,
- two jets from the hadronically decaying W boson.

The requirement of one high transverse momentum electron or muon and large missing transverse energy rejects most multijet backgrounds. Other physics processes, like W+jets, have the same signature.

There are three stages of the event selection: The first stage is to make sure that the interesting events fulfilling the $t\bar{t} \to \ell$+jets event signature are written to disk. Therefore a set of trigger requirements is defined. Secondly, a set of selection criteria is defined to select a sample enriched in events with isolated high p_T leptons (composed primarily of W+multijet and $t\bar{t}$ events). This second stage is referred to as the preselection and is mostly concerned with selections based on lepton requirements. In the third and last stage the $t\bar{t}$ events are separated from most of the backgrounds by requiring the presence of two b-tagged jets in the event.

4.2.1 Trigger Selection

The triggers used to record the signal sample require a lepton and at least one jet and are different for the e+jets and μ+jets channels[1].

The signal trigger for the e +jets channel requires at Level 1, at least one electromagnetic calorimeter (EM) tower with transverse energy E_T above 10 GeV and one additional calorimeter tower (EM+H) with E_T above 5 GeV. At Level 2 an EM candidate with electromagnetic fraction above 85% and E_T above 10 GeV is required. Level 3 requires an EM candidate with E_T above 15 GeV, passing a transverse shower shape criteria and one Level 3 jet with E_T above 15 GeV.

[1]The specific trigger requirements are divided into well defined trigger lists. These trigger lists change with time to accommodate detector and luminosity changes.

For the μ+jets channel the signal trigger requires a calorimeter trigger tower above 5 GeV at Level 1. At Level 2, the calorimeter trigger varies depending on the period when the data was collected and changes from no requirement to at least one Level 2 jet with E_T above 10 GeV. The Level 3 calorimeter trigger requirement is one jet candidate with E_T above 20 or 25 GeV depending on the period of data taking. The muon trigger uses information from both Level 1 and Level 2. At Level 1, a candidate is required two have a coincidence between at least two layers of scintillators and similar requirements for Level 2 with the additional requirement of hits in the drift tubes.

The probability for a $t\bar{t}$ event to pass all the trigger requirements is expressed in the $t\bar{t}$ trigger efficiency. The per muon, electron and jet probabilities to fire the trigger are derived on data. The $t\bar{t}$ trigger efficiency is then obtained by folding per lepton and per jet efficiencies with the η and E_T/p_T of the lepton and jets in simulated events. The $t\bar{t}$ trigger efficiency in the e+jets (μ+jets) channel is $92.82 \pm 0.08\%$ ($91.65 \pm 0.91\%$) in events with four or more reconstructed jets.

More detailed information on the specific trigger requirements used and the measurement of the trigger efficiency can be found in [41; 44; 70].

4.2.2 Preselection

Apart from requirements on the charged lepton the preselection in both channels are identical. The preselection in this analysis are the same as in Ref. [56] where a more detailed discussion can be found.

The common event preselection criteria for both μ+jets and e+jets channel are:

- At least four jets with $p_T > 15$ GeV and $|\eta| < 2.5$. All additional jets in the event are subject to the same p_T and η requirements,
- missing transverse energy $\not\!\!E_T > 20$ GeV,
- a primary vertex with at least three tracks fitted to it and a z coordinate z_{PV} within the fiducial volume of the SMT detector ($|z_{\mathrm{PV}}| \leq 60$ cm) ,
- the distance in the z-direction between the primary vertex and the lepton track has to be less than 1 cm.

Preselections specific to the μ+jets channel

The event preselection criteria specific for the μ+jets channel are:

- One muon with $p_T > 20$ GeV and $|\eta| < 2.0$.
- The muon is required to be separated ($\Delta R(\mu, \mathrm{jet}) > 0.5$) from reconstructed jets and also isolated from activity in the calorimeter by requiring that the scalar sum of E_T of calorimeter clusters in a hollow cone between $\Delta R = 0.1$ and $\Delta R = 0.4$ away from the muon is less than 8% of the muon p_T.

- The matched muon track is required to be isolated from other activity in the central tracker by requiring that the sum of p_T of all tracks inside a cone of $\Delta R = 0.5$ around the muon is less than 6% of the muon p_T.
- The missing transverse energy direction must be separated from the direction of the muon in ϕ. The missing transverse energy in QCD events passing the muon isolation requirements is found mostly in or opposite the muon direction. This can be explained by $b\bar{b}$ production where one or both B hadrons decay semi-leptonically and the jet is not reconstructed.
- Reject events with an electron with $p_T > 15$ GeV in the central or the end-cap calorimeter ensuring orthogonality between the $t\bar{t} \to \mu$+jets, $t\bar{t} \to e$+jets and $t\bar{t} \to e\mu$ channel.
- Reject events with a second muon with $p_T > 15$ GeV. This ensures orthogonality between the $t\bar{t} \to \mu$+jets channel and the $t\bar{t} \to \mu^+\mu^-$ channel and also rejects $Z \to \mu^+\mu^-$ events.

Preselections specific to the e+jets channel

The event preselections specific to the e+jets channel are:

- The presence of an electron with $p_T > 20$ GeV and $|\eta| < 1.1$.
- The missing transverse energy direction must be separated from the direction of the electron in ϕ in order to eliminate events in which a jet was misidentified as an electron.
- Reject events with a second electron with $p_T > 15$ GeV ensuring orthogonality between the $t\bar{t} \to e$+jets and $t\bar{t} \to e^+e^-$ channels. This also rejects $Z \to e^+e^-$ events.
- Reject events with a muon with $p_T > 15$ GeV thus ensuring orthogonality between $t\bar{t} \to e$+jets, $t\bar{t} \to e\mu$ and $t\bar{t} \to \mu$+jets channels.

Preselected Sample

The number of preselected events in the μ+jets (e+jets) channel are shown in Tab. 4.1 (4.2) together with the expected contribution of QCD and W+jets events. To estimate the fraction of QCD events a sample with a "looser" selection criteria is defined[2]. By calculating the probability for events that pass the "looser" selection to also pass the normal preselection requirement (for both QCD events and events with a true isolated lepton e.g. W+jets and $t\bar{t}$), the fraction of QCD events in the preselected sample can be estimated. More information on how to estimate the QCD background can be found in Ref. [36].

[2] In the μ+jets channel, the muon isolation is loosened by only requiring the muon to be isolated from jets in a cone of $\Delta R = 0.5$. In the e+jets channel, the requirement on the electron likelihood is dropped.

	1jet	2jets	3jets	≥4jets
$N^{presel}_{\mu+jets}$	10101	3863	933	231
$N^{presel}_{(W\mu)+jets}$	9726±103	3669±63	874±31	215±16
$N^{presel}_{QCD\mu+jets}$	375±12	195±7	60±3	16±2

Table 4.1. Number of preselected events in the μ+jets channel and expected contribution from QCD and W-like events categorized by the number of (exclusive) jets in the events, extracted from Ref. [56].

	1jet	2jets	3jets	≥4jets
N^{presel}_{e+jets}	12668	4587	1078	277
$N^{presel}_{(We)+jets}$	12130±173	4142±101	909±41	221±19
$N^{presel}_{QCDe+jets}$	538±127	445±72	169±22	56±7

Table 4.2. Number of preselected events in the e+jets channel and expected contribution from QCD and W-like events categorized by the number of (exclusive) jets in the events, extracted from Ref. [56].

The preselection efficiency for $t\bar{t} \to \ell$+jets events is calculated from simulated events and scale factors are applied to take into account significant data-to-simulation discrepancies. The efficiency to select $t\bar{t} \to \ell$+jets events is $13.4 \pm 1.8\%$ ($13.0 \pm 1.5\%$) in the μ+jets (e+jets) channel requiring at least four or more jets. Other backgrounds than W+jets and QCD have preselection efficiencies less than 1% except single top quark and $t\bar{t}$ dilepton backgrounds which have preselection efficiencies around 1.5%.

4.2.3 Final Event Selection

The majority of the backgrounds does not contain b-quark jets and thus the identification of b-quark jets can be used to preferentially select $t\bar{t}$ events while removing background events. This is used to obtain a pure sample of $t\bar{t} \to \ell$+jets events in the last stage of the event selection:

The event must contain at least two jets tagged by the SVT algorithm.

Sample Composition

The composition of the data sample after double-tagging (called signal sample) is determined in Ref. [56] starting from the preselected sample and the predicted number of W+jets and QCD events. This is further multiplied by the probability for an event to contain jets that are b-tagged. To obtain an estimate of the number and type of events after applying b-tagging on the preselected sample, the probability for

an event to have two or more b-tagged jets must be calculated. This event tagging probability depends strongly on the topology and jet flavors in the event. It is determined from Monte Carlo simulations, applying to each jet the per jet b-tagging efficiency derived from data (as described in Sec. 3.6.4). The QCD background contribution to the signal sample is estimated to be < 0.01 events. To estimate the W+jets contribution to the signal sample, the predicted number of W+jets events in the preselected sample is split into its expected flavor composition as predicted by ALPGEN. The W+jets background in the preselected sample is dominated by non-heavy flavor jets (81%) due to the low production cross section of a W boson in association with heavy flavor jets. The low probability to tag light jets leads to an effective rejection of $W+$ light jets in the signal sample (< 0.01 events). The only sizable W+jets background remaining is the production of a W boson in association with a $b\bar{b}$ pair which has an event tagging probability ($\approx 10\%$) on the same order as a $t\bar{t} \to \ell$+jets event ($\approx 15\%$).

Other physics backgrounds that contributes to the signal sample are diboson production ($WW \to \ell$+jets, $WZ \to \ell$+jets, $WZ \to jj\ell\bar{\ell}$, $ZZ \to \ell\bar{\ell}jj$), single top production and Z+jets($Z \to \tau\tau \to \ell$+jets and $Z \to \ell$+jets). The Z+jets background has a similar event signature as $t\bar{t}$ but is suppressed due to no physical source of $\not\!\!E_T$. For a given process i, the number of events N_i^{presel} in the preselected sample is determined by $N_i^{\text{presel}} = \sigma_i \epsilon_i^{\text{presel}} \mathcal{BR}_i \int \mathcal{L}\text{dt}$, where σ_i, $\epsilon_i^{\text{presel}}$, $\mathcal{BR}_i$ and $\int \mathcal{L}\text{dt}$ are, respectively, the cross section, the preselection efficiency, the branching fraction for the specific process and the integrated luminosity. The expected number of events in the signal sample is estimated by multiplying N_i^{presel} with the event tagging probability for that specific process. The only non-negligible backgrounds are the single top quark production and the $t\bar{t} \to$ dileptons processes.

After all selections, 21 data events remain. Table 4.3 and 4.4 summarizes the sample composition according to predicted signal and background contributions in the e+jets and μ+jets channel respectively. It should be noted that in the analysis presented in this thesis, only events with four or more jets are used.

	e+jets		
	2 jets	3 jets	$\geq$4 jets
W+light	0.015±0.002	0.012±0.001	< 0.01
$W(c\bar{c})$	0.021±0.003	0.014±0.001	< 0.01
$W(b\bar{b})$	0.29±0.04	0.12±0.01	0.03±0.01
Wc	0.037±0.002	0.015±0.001	< 0.01
$Wc\bar{c}$	0.36±0.02	0.15±0.01	0.05±0.01
$Wb\bar{b}$	4.77±0.17	1.46±0.10	0.49±0.06
W+jets	5.5±0.2	1.77±0.10	0.59±0.06
QCD	< 0.01	0.29±0.34	< 0.01
tb	0.96±0.01	0.52±0.01	0.16±0.01
$t\bar{t} \rightarrow ll$	1.76±0.02	0.92±0.02	0.20±0.01
diboson	0.47±0.02	0.06±0.01	< 0.01
$Z \rightarrow \tau^+\tau^-$	0.03±0.03	0.02±0.02	< 0.01
Background	8.3±0.2	3.58±0.37	0.76±0.36
Syst. uncert. (bkg)	+1.42-1.42	+0.49-0.51	+0.17-0.15
$t\bar{t} \rightarrow l$+jets	0.94±0.03	5.8±0.1	8.6±0.1
Sum pred.	9.2±0.2	9.3±0.4	9.4±0.4
Syst. uncert.	+1.49-1.46	+0.87-0.87	+1.22-1.42
Observed	11	7	**13**

Table 4.3. Summary of observed and predicted number of events with two SVT-tagged jets in the e+jets channel. The individual contributions from the various backgrounds processes are shown. Unless explicitly stated, uncertainties are statistical only.

	μ+jets		
	2 jets	3 jets	$\geq$4 jets
W+light	0.028±0.003	0.016±0.002	< 0.01
$W(c\bar{c})$	0.027±0.004	0.015±0.001	< 0.01
$W(b\bar{b})$	0.30±0.04	0.09±0.01	0.04±0.01
Wc	0.044±0.003	0.017±0.001	< 0.01
$Wc\bar{c}$	0.35±0.01	0.13±0.01	0.05±0.01
$Wb\bar{b}$	4.69±0.15	1.56±0.09	0.48±0.05
W+jets	5.4±0.2	1.83±0.09	0.59±0.05
QCD	0.02±0.14	0.03±0.10	< 0.01
tb	0.85±0.01	0.48±0.01	0.15±0.01
$t\bar{t} \to ll$	1.51±0.02	0.84±0.02	0.18±0.01
diboson	0.46±0.02	0.07±0.01	< 0.01
$Z \to \tau^+\tau^-$	< 0.01	< 0.01	< 0.01
Background	8.3±0.2	3.25±0.15	0.77±0.12
Syst. uncert. (bkg)	+1.40-1.39	+0.49-0.51	+0.15-0.16
$t\bar{t} \to \ell$+jets	0.76±0.02	5.1±0.1	8.8±0.1
Sum pred.	9.0±0.2	8.4±0.2	9.6±0.1
Syst. uncert.	+1.44-1.44	+0.92-0.94	+1.33-1.50
Observed	11	4	**8**

Table 4.4. Summary of observed and predicted number of events with two SVT-tagged jets in the μ+jets channel. The individual contributions from the various backgrounds processes are shown. Unless explicitly stated, uncertainties are statistical only.

4.3 Jet Charge Algorithms

In this section a method to use tracks inside jets to form a variable which is capable to discriminate between jets arising from the hadronization of b- and $\bar{b}$-quarks is presented.

4.3.1 Jet Charge Algorithm Definition

Various algorithms to discriminate between b- and $\bar{b}$-quark jets have been used in the past [71; 72; 73; 74; 75]. The idea is to use the particle tracks reconstructed by the central tracker, associate them with a reconstructed jet in the calorimeter and compute the collected charge of the jet. In the hadronization, the initial quark transverse momentum is shared in a multi-stage process between many particles. A common feature of the modeling of this process is that in most jets arising from the hadronization of a b- or $\bar{b}$-quark, the tracks with highest p_T are the decay products of the B hadron in the jet. A weighted sum of the charges of the tracks are therefore used in most algorithms. The weights are usually functions of the track momentum or its projection along a certain direction. In designing such a jet charge algorithm one needs to decide:

- What tracks are associated with the jet, within a cone in ΔR or in a whole hemisphere,
- the quality criteria of the tracks to be considered,
- what weight to give each track.

In this analysis, only tracks that fulfill the following criteria are considered:

1. Track $p_T > 0.5$ GeV,
2. distance between the track and jet axis must be less than 0.5 in ΔR,
3. a distance to the primary vertex in the z-direction of less than 1 cm unless the track is fitted to a secondary vertex.

At least four jets are present in the $t\bar{t} \to \ell$+jets events and it is therefore important to consider only tracks that are associated with the corresponding jet. Also, in this analysis the interest is to calculate only the jet charge for the jets originating from the b- and $\bar{b}$-quarks in the $t \to Wb$ decay. The SVT algorithm determines what jets should be considered and only tracks within a ΔR cone (studied below) of the jet axis are considered.

The performance of two algorithms is evaluated, algorithm I:

$$Q_{\text{jet}} = \frac{\sum_i q_i \cdot p_{T_i}^a}{\sum_i p_{T_i}^a}, \tag{4.1}$$

and algorithm II:

$$Q_{\text{jet}} = \frac{\sum_i q_i \cdot |p_{||_i}|^a}{\sum_i |p_{||_i}|^a}, \tag{4.2}$$

where the subscript i runs over all charged tracks passing the track quality cuts described above and that are located within a cone of ΔR from the jet axis. Each track has a charge q_i and a transverse momentum p_{T_i}, while $p_{||_i}$ represent the projection of the track momentum along the jet axis. The parameter a is an arbitrary number which is optimized from simulated $t\bar{t} \to \ell$+jets events. For $a = 0$, the weight given to each track is equal to one and hence, p_T independent, while $a = \infty$ is equivalent to considering solely the highest p_T track. Figure 4.2 shows the jet charge distribution for SVT-tagged b- and $\bar{b}$-quarks in simulated $t\bar{t} \to \ell$+jets events.

4.3.2 Optimization

A simple optimization of the jet charge algorithm parameters using b- and $\bar{b}$-quark jets in simulated $t\bar{t} \to \ell$+jets events is performed. In order to quantify the separation between b- and $\bar{b}$-quark jets, a variable called discriminating (or discriminant) power D was defined as:

$$D = \frac{|a_b - a_{\bar{b}}|}{\sqrt{V_b + V_{\bar{b}}}}, \tag{4.3}$$

where a_b, V_b and $a_{\bar{b}}, V_{\bar{b}}$ are the mean and variance of the jet charge distributions obtained for b- and $\bar{b}$-quark jets respectively.

The purpose of the algorithm is to discriminate between b- and $\bar{b}$-quark jets from the decay of top quarks in $t\bar{t} \to \ell$+jets events. The optimization is carried out using simulated $t\bar{t} \to \ell$+jets events. In the optimization, a reconstructed jet is labeled as a true b-quark ($\bar{b}$-quark) jet if a b-quark ($\bar{b}$-quark) at parton level in the Monte Carlo history is found within a cone of $\Delta R = 0.5$ with respect to the jet axis. If more than one heavy flavor (b or $\bar{b}$) parton was inside the cone, the closest one is used. Only SVT-tagged jets are considered. To avoid biases from the event topology, the events were required to have at least four reconstructed jets with $p_T > 15$ GeV and $|\eta| < 2.5$ (in the simulated $Z \to b\bar{b}$ events used for comparison below, exactly two jets were required).

In the optimization, the parameter a was varied between 0.4 and 1.6 in steps of 0.2 and the size of the jet cone was varied between 0.3 and 0.7 in steps of 0.1. Figure 4.3 shows the discriminating power for the algorithms I and II as function of a and ΔR. Little difference is found between the two algorithms and algorithm I is chosen for the remaining of this analysis. It is found that the highest discriminating power is obtained for $\Delta R = 0.5$ and $a = 0.6$.

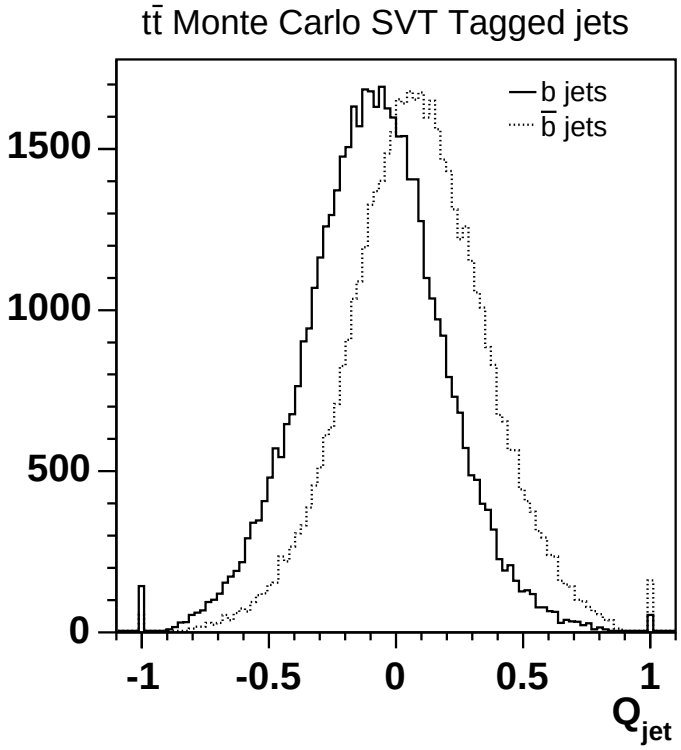

Figure 4.2. Jet charge distribution for SVT-tagged b and $\bar{b}$-quark jets in simulated $t\bar{t} \to \ell$+jets events using algorithm I with $a = 0.6$ and a jet cone size of $\Delta R = 0.5$.

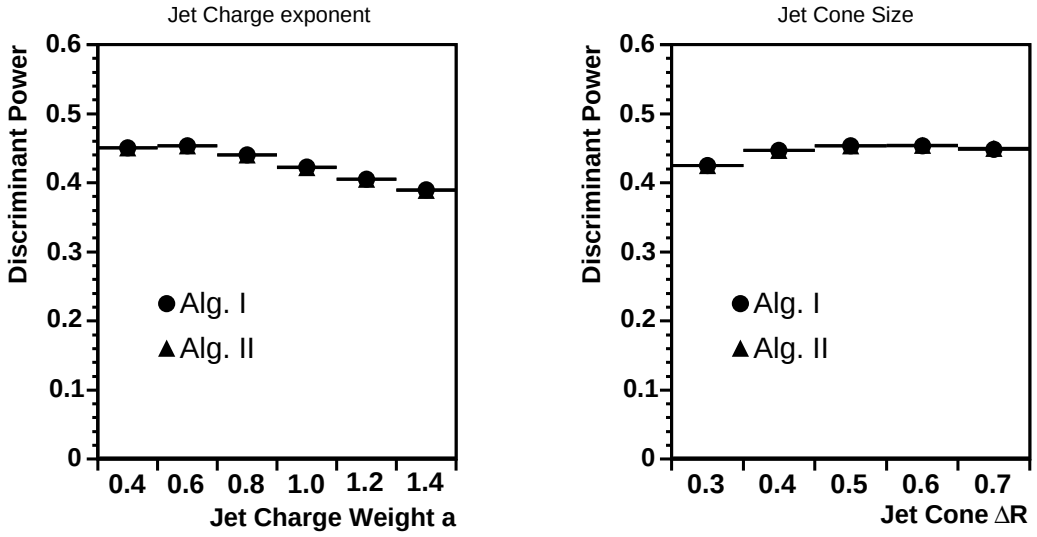

Figure 4.3. Comparison of the discriminating power for the jet charge algorithms I and II as function of a (left) and ΔR (right).

In general, tracks originating from the hadronization are expected to come from the primary vertex unless they come from the decay products of the B hadron and consequently displaced. Therefore, the maximum distance of the tracks to the primary vertex in the z-direction was studied. Tracks determined to originate from a displaced B hadron vertex are not included. In this study using simulated events, the discriminating power increases when loosening this requirement. This is expected to be a variable heavily affected by the detailed modeling of the tracking which is known to be poor. Therefore, a conservative requirement of < 0.1 cm is used to ensure that tracks only close to the primary vertex are considered.

In Fig. 4.4 the discriminating power is shown as function of the minimum requirement of the p_T of the tracks. This variable is also expected to exhibit the same critical dependence on the correct modeling of the tracking. Therefore, the minimum p_T requirement was chosen to be 0.5 GeV to ensure the quality of the tracks. The requirements on the minimum track p_T and maximum distance from the track to the primary vertex are kept in the rest of this analysis.

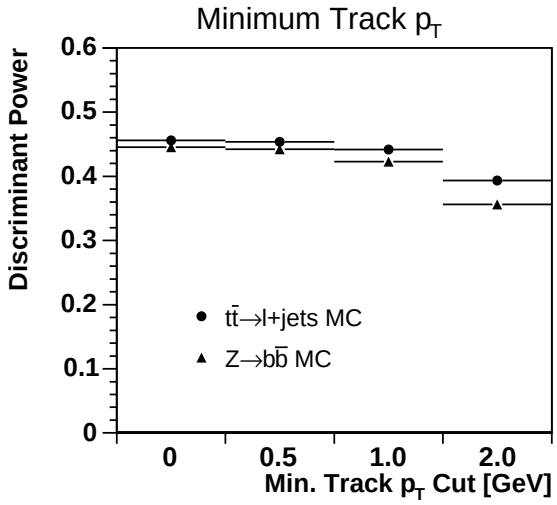

Figure 4.4. Discriminating power as a function of the minimum track p_T requirement.

4.4 Jet Charge Calibration on Data

The Monte Carlo description of the DØ tracking system includes the detailed geometry of the detector and a modeling of the electronic noise. Nevertheless, the Monte Carlo simulations are unable to describe fine details of track quality distributions such as χ^2, hit multiplicities and tracking efficiency within jets. Physics analyzes in DØ that use the tracking detectors for b-tagging determine the performance of the b-tagging algorithms on data, or with as little input from simulation as possible. A similar approach is chosen in this analysis.

The goal of the present section is to show that the jet charge distribution templates for SVT-tagged b-, $\bar{b}$-, c- and $\bar{c}$-quark jets can be extracted from data. The jet charge distributions are then used to derive the expected distributions of the charge observables in the SM top and exotic quark scenarios.

The difficulty is to find the true flavor of the particle initiating a jet in data. Therefore, a data sample enriched in $b\bar{b}$ pairs is selected. In such an event with exactly two jets back-to-back in azimuth, called dijet event, one of the jets is required to contain a muon from the semi-leptonic decay of a B hadron and is referred to as the *tag*-jet, the muon associated with the *tag*-jet is referred to as the *tag*-muon or simply tagging muon. Studies [76; 77; 78] have shown that the mechanism for $b\bar{b}$ production at the Tevatron depends heavily on the azimuthal distance between the jets, with large distances dominated by flavor creation. In such an event, the charge of the muon can be used to find the type of quark initiating the *tag*-jet and, consequently, also the other jet in the event (referred to as the *probe*-jet) as given in Tab. 4.5. The method above is called the *tag*-and-*probe* method.

Muon	Quark initiating the *tag*-jet	Quark initiating the *probe*-jet
μ^-	b	$\bar{b}$
μ^+	$\bar{b}$	b

Table 4.5. Charge of the muon in the *tag*-jet and the corresponding type of b-quark in the *tag*-jet and the *probe*-jet, here assuming no B mixing and only direct semi-leptonic b-decays.

The starting point are the dijet samples (defined below in Sec. 4.4.1) dominated by b- and $\bar{b}$-quark jets. The dijet samples (illustrated in Fig. 4.5) contain a jet with a μ- and a SVT-tag on one side and a SVT-tagged jet on the other side of the event. The goal is to find the jet charge templates for b- and c-quark jets in general, not necessarily μ-tagged jets and therefore the jet charge distribution is extracted from the *probe*-jet.

In the simplified case, when there is no B mixing [5], nor contamination of the sample by c-quark jets or light-jets, and all the tagging muons were coming from a direct B hadron decay, the charge of the tagging muon would reliably tell if the *tag*-jet was initiated by a b- or a $\bar{b}$-quark. According to this procedure, the two

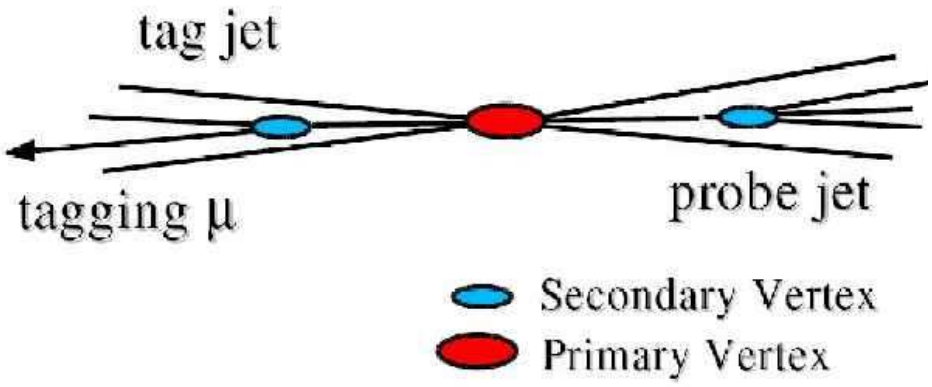

Figure 4.5. Illustration of the *tag*-and-*probe* method in the tight dijet sample (defined in Sec 4.4.1). The events contain exactly two SVT-tagged jets in a back-to-back configuration in azimuth, one of which is also μ-tagged.

distributions of jet charge for jets that are believed to be b-quark jets and those that are believed to be $\bar{b}$-quark jets can be extracted by plotting separately Q_{jet} for the μ^+ and μ^- samples. This ideal scenario is complicated by a number of issues in data. First the sample is not pure $b\bar{b}$ production, but contains a small fraction of $c\bar{c}$ production. Therefore the observed b- and $\bar{b}$-quark jet charge distributions from the *probe*-jet are a mixture of the jet charge for b-quark jets and for c-quark jets. The contamination by c-quark jets in the dijet samples is determined in Sec. 4.4.3.

The tagging muon can also arise from a decay which is not a direct B hadron decay. This sort of muon is referred to as a cascade muon. It can either arise from the decay of a D meson or lighter hadrons. In this case, the relation between the charge of the tagging muon and the type of b-quark in the *probe*-jet is not given by Tab. 4.5 anymore. B mixing on the side of the *tag*-jet can also destroy the correlation between the muon sign and the type of b-quark initiating the *probe*-jet. The B meson on the *probe*-jet side can also mix, but this does not matter, since in the end the data-derived b-quark jet charge distributions is applied to the b-quark from the decay of the top quarks in $t\bar{t}$ events, not the B meson. Another effect that can destroy the correlation between the muon sign and the type of b-quark initiating the *probe*-jet is an incorrectly measured the muon charge.

To illustrate the effect of the B mixing, cascade decay and c-contamination, Fig. 4.6 shows a comparison of the discriminating power of the jet charge distributions between:

i) Simulated $Z \to b\bar{b}$ events where the true b-type (b or $\bar{b}$) is extracted from the Monte Carlo truth to sort between the b- and $\bar{b}$-quark jet charge. Note that we look for the b- or $\bar{b}$-quark in the Monte Carlo history that is closest in ΔR to the jet axis.

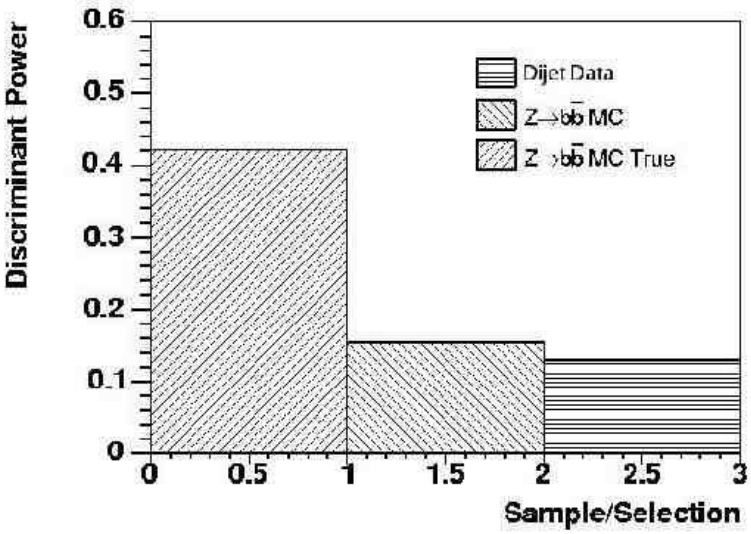

Figure 4.6. Comparison of the discriminating power using the *tag*-and-*probe* method in the tight dijet sample (rightmost bin) and in simulated $Z \to b\bar{b}$ events (middle bin). In the leftmost bin, the discriminating power for simulated $Z \to b\bar{b}$ events is shown using the true charge of the b-quark jet from the Monte Carlo history to sort the b- and $\bar{b}$-quark jets.

ii) simulated $Z \to b\bar{b}$ events using the *tag*-and-*probe* method,

iii) a dijet sample (defined in Sec.4.4.1) using the *tag*-and-*probe* method.

As expected the discriminating power in case *i*) is much better than in either case *ii*) and *iii*) which both suffer from cascade decays and B mixing. Data also suffers from contamination by $c\bar{c}$ events and from poorer tracking than the simulated events.

The amount of contamination from cascade decay and c-quark jets can be changed by requiring a high $p_{T,\mathrm{rel}}$, defined as the relative momentum of the tagging muon with respect to the jet axis, see Fig. 4.7. $p_{T,\mathrm{rel}}$ is expected to be larger for b-quark jets than from c-quark or light jets due to the larger b-quark mass. This effect is illustrated in Fig. 4.8. The goal is to extract the jet charge distributions of the *probe*-jet. The tagging muon only serves as the source to find the true type of quark initiating the jet. Thus, after taking into account all sources that affect the charge of the tagging muon (that can destroy the correlation between the charge and the type of quark) the extracted jet charge distributions from the *probe*-jet should be independent of the specific requirement on the minimum $p_{T,\mathrm{rel}}$ of the tagging muon. As a cross-check, the extraction of the jet charge distributions are therefore performed with different requirements of the minimum $p_{T,\mathrm{rel}}$ of the tagging muon. The different requirements are however later shown to give similar

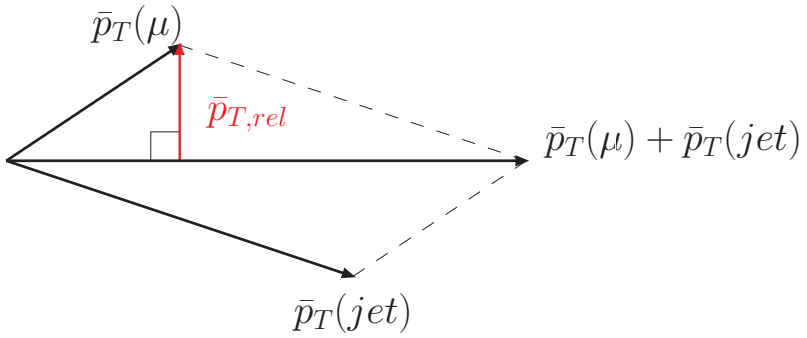

Figure 4.7. Definition of $p_{T,\mathrm{rel}}$ for a muon within a jet.

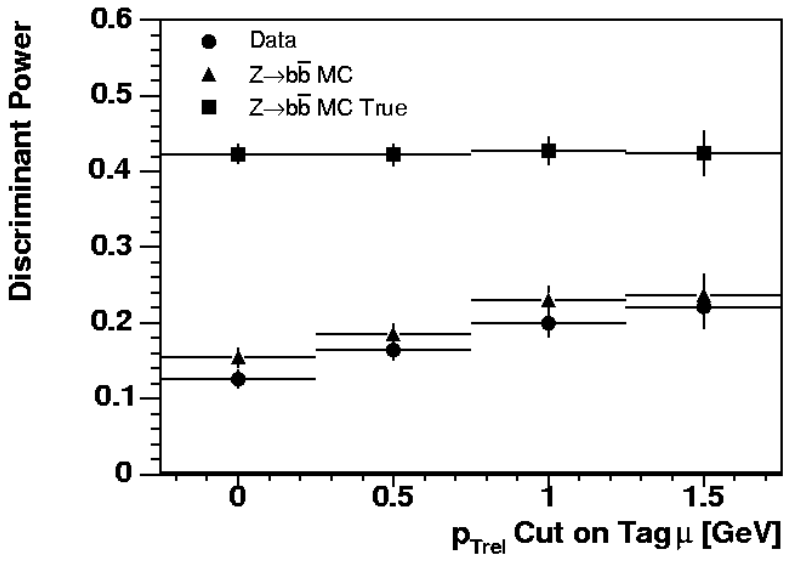

Figure 4.8. The effect of the requirement of a minimum $p_{T,\mathrm{rel}}$ of the tagging muon in the dijet sample and simulated $Z \to b\bar{b}$ events using the *tag*-and-*probe* method. For comparison the discriminating power in simulated $Z \to b\bar{b}$ events is shown when using the Monte Carlo truth to find the quark initiating the jet.

results. The final jet charge distributions are derived without a $p_{T,\mathrm{rel}}$ requirement on the tagging muon in order to minimize the statistical uncertainty.

4.4.1 Dijet Data Samples

To calibrate the jet charge algorithm using data a sample enriched in b-quark jets is selected as discussed above. The data sample is based on a generally selected sample requiring a muon matched to a jet. The following additional requirements are used to further enhance the sample in events with $b\bar{b}$ production:

- The event must have exactly two jets j_1 and j_2 with $p_T > 15$ GeV and $|\eta| < 2.5$.
- The azimuthal distance ($\Delta\phi$) between j_1 and j_2 is larger than 3.0.
- j_1 and j_2 must be b-tagged using the SVT algorithm,
- j_1 must be associated with a muon within $\Delta R(\mu, \mathrm{jet}) < 0.5$ from the jet axis.

In summary, this sample has exactly two jets in a back-to-back configuration in ϕ, one jet j_1 being both SVT-tagged and μ-tagged and the second jet j_2 being SVT-tagged as shown schematically in Fig. 4.5. This sample is referred to as the "**tight dijet sample**".

Similarly, the "**Loose Dijet sample**" is defined in the same way as the "tight dijet sample" apart from the requirement of the SVT-tag for j_1 which is removed. Thus, the "tight dijet sample" is a subset of the "loose dijet sample".

4.4.2 Extraction of Jet Charge Templates from Dijet Data

In the ideal case where the tagging muon comes from a direct B hadron decay without B mixing, the *tag*-and-*probe* method would in fact yield the true jet charge distributions. In practice when the tag muons are for example of positive sign the *probe*-jets are in majority from b-quark jets but mixed with a certain fraction of $\bar{b}$-, c-, and $\bar{c}$-quark jets.

The jet charge distributions of the *probe*-jet obtained in the tight dijet sample are denoted: $f_{\mu+}$ and $f_{\mu-}$ where the subscript $\mu\pm$ indicates the sign of the tagging muon. Similarly, f_b and $f_{\bar{b}}$ denotes the jet charge distributions for b and $\bar{b}$-quark jets, these are the distributions we want to extract from the data.

In absence of B mixing, if the fraction of c-quark jets were zero and if all tagging muons were coming from a direct B decay, then the jet charge distributions for b- and $\bar{b}$-quark jets would simply be given by: $f_b = f_{\mu+}$ and $f_{\bar{b}} = f_{\mu-}$.

In reality, care has to be taken to the processes which change the sign of the tagging muon. In the tight dijet sample x_{flip} is defined as the fraction of tagging muons which do not have the same sign as the parent B hadron due to cascade decay, or because it originates from a kaon or pion decay.

If this is the only process that affects the correlation between the sign of the tagging muon and the quark that initiated the *probe*-jet, $f_{\mu+}$ and $f_{\mu-}$ can be written as:

$$f_{\mu^+} = x_{\text{flip}} \times f_{\bar{b}} + x_{\text{noflip}} \times f_b \tag{4.4}$$
$$f_{\mu^-} = x_{\text{flip}} \times f_b + x_{\text{noflip}} \times f_{\bar{b}} \tag{4.5}$$

where x_{noflip} is simply $1 - x_{\text{flip}}$.

The tight dijet sample contains also a fraction of c- and $\bar{c}$-quark jets. Therefore the jet charge templates $f_{\mu+}$ and $f_{\mu-}$ contain also a fraction of the jet charge distributions f_c and $f_{\bar{c}}$ for c- and $\bar{c}$-quark jets respectively. The fraction of c-quark jets in the tight dijet sample is denoted x_c. Equation 4.5 can therefore be rewritten as

$$\begin{aligned} f_{\mu^+} &= (1 - x_c)\,(x_{\text{flip}} \times f_{\bar{b}} + x_{\text{noflip}} \times f_b) + x_c \times f_{\bar{c}} \\ f_{\mu^-} &= (1 - x_c)\,(x_{\text{flip}} \times f_b + x_{\text{noflip}} \times f_{\bar{b}}) + x_c \times f_c, \end{aligned} \tag{4.6}$$

Note that the fraction of c ($\bar{c}$) jets contributing to f_{μ^+} (f_{μ^-}) are neglected. These contributions are at most of the order of $x_c \times x_{\text{flip}}$ relative to the flipped muon contributions from direct B decays. In addition, the muons from D meson decays have significantly lower momenta and are less likely to pass the muon momentum requirement in the dijet samples. Therefore the component of c ($\bar{c}$) jets contributing to f_{μ^+} (f_{μ^-}) is further suppressed.

Equations 4.6 provides two equations with four unknowns (f_b, $f_{\bar{b}}$, f_c, $f_{\bar{c}}$). To be able to extract these four jet charge distributions another two equations are needed. This is achieved by using the loose dijet sample, which has a different fraction of c-quark jets, that we denote x'_c because of the relaxation of the SVT-tag on the *tag*-jet.

In the same fashion as for the tight dijet sample, f'_{μ^+} and f'_{μ^-} are the jet charge distributions observed for the *probe*-jet in the loose dijet data when the tag muon is positive or negative respectively. The equivalent of equations 4.6 in the loose dijet sample are:

$$\begin{aligned} f'_{\mu^+} &= (1 - x'_c)\left(x'_{\text{flip}} \times f_{\bar{b}} + x'_{\text{noflip}} \times f_b\right) + x'_c \times f_{\bar{c}} \\ f'_{\mu^-} &= (1 - x'_c)\left(x'_{\text{flip}} \times f_b + x'_{\text{noflip}} \times f_{\bar{b}}\right) + x'_c \times f_c, \end{aligned} \tag{4.7}$$

Using Eq. 4.7 and 4.6 the system of equation can be solved for f_b, $f_{\bar{b}}$, f_c and $f_{\bar{c}}$ provided that the fraction of c-quark jets and the fraction of times the measured tagging muon charge is changed with respect to the quark initiating the jet. The procedure to determine the fraction of c-quark jets in both samples is described in Sec. 4.4.3 and the calculation of fraction of events with changed muon charge sign x_{flip} and x'_{flip} is described in Sec. 4.4.4.

The solutions to Eq. 4.6 and are 4.7:

$$
\begin{aligned}
f_{\bar{b}} &= \frac{f'_{\mu^-}x_c - f'_{\mu^-}x_c x_{\text{flip}} - f'_{\mu^+}x_c x_{\text{flip}} - f_{\mu^-}x'_c + f_{\mu^-}x_{\text{flip}}x'_c + f_{\mu^+}x_{\text{flip}}x'_c}{(-1+2x_{\text{flip}})(x'_c - x_c)} \\
f_{b} &= \frac{f'_{\mu^+}x_c - f'_{\mu^-}x_c x_{\text{flip}} - f'_{\mu^+}x_c x_{\text{flip}} - f_{\mu^+}x'_c + f_{\mu^-}x_{\text{flip}}x'_c + f_{\mu^+}x_{\text{flip}}x'_c}{(-1+2x_{\text{flip}})(x'_c - x_c)} \\
f_{\bar{c}} &= \frac{-f_{\mu^+} + f'_{\mu^+} - f'_{\mu^+}x_c + f_{\mu^+}x'_c}{x'_c - x_c} \\
f_{c} &= \frac{-f_{\mu^-} + f'_{\mu^-} - f'_{\mu^-}x_c + f_{\mu^-}x'_c}{x'_c - x_c}
\end{aligned}
\tag{4.8}
$$

4.4.3 Fraction of *c*-Quark Jets in the Dijet Samples

The difference in mass between the b-quark, c-quark and light quarks implies that the momentum distribution of muons within jets from semi-leptonic decays of B-, D- and light mesons are different. This property is utilized to extract the respective fraction of b-quark, c-quark and light jets in the dijet samples which were used to extract the jet charge distributions for b- and c-quark jets above.

The expected $p_{T,\text{rel}}$ spectra from b- and c-quark jets found from Monte Carlo simulation are used to fit the observed $p_{T,\text{rel}}$ spectra in the dijet samples. The light flavor contribution is assumed to be negligible. This is also confirmed a posteriori by the fact that the fitted fraction of c-quark jets is small and that the light jet tagging efficiency is ≈ 15 times lower than than the c-quark tagging efficiency (see Sec. 3.6.4).

The expected b-quark jet $p_{T,\text{rel}}$ spectrum is determined using simulated $Z \to b\bar{b}$ events including cascade decays ($b \to (c \to) \mu$) and the expected c-quark jet $p_{T,\text{rel}}$ spectrum is determined from simulated $Z \to c\bar{c}$ events. $p_{T,\text{rel}}$ templates in three bins of muon p_T: 4 GeV$< p_{T\mu} <$ 8 GeV, 8 GeV$< p_{T\mu} <$ 10 GeV and $p_{T\mu} >$ 10 GeV and in three bins of jet p_T: 15 GeV$< p_T <$ 35 GeV, 35 GeV$< p_T <$ 55 GeV and $p_T >$ 55 GeV are constructed. The jet flavor in the Monte Carlo simulated events is determined by matching the direction of the reconstructed jet to the heaviest hadron flavor within a cone of $\Delta R < 0.5$. If there is more than one hadron found within the cone, the jet is considered to be a b-quark jet if the cone contains at least one B hadron. It is called a c-quark jet if there is at least one D meson in the cone and no B hadron. The fitted parameter is the fraction of c-quark jets among the *tag*-jets of the dijet samples.

The $p_{T,\text{rel}}$ template fits are shown in Fig. 4.9 and Fig. 4.10 for the loose and tight dijet samples respectively. The fitted fractions of b- and c-quark jets are presented in Tab. 4.6 and 4.7 respectively. The fit also takes into account the statistical uncertainty on the $p_{T,\text{rel}}$ spectra obtained from simulated events as well as the uncertainty from the observed spectrum.

The total fraction of of c- and b-quark jets in the dijet samples can be extracted by the sum of the measured fraction in each bin weighted by the event population

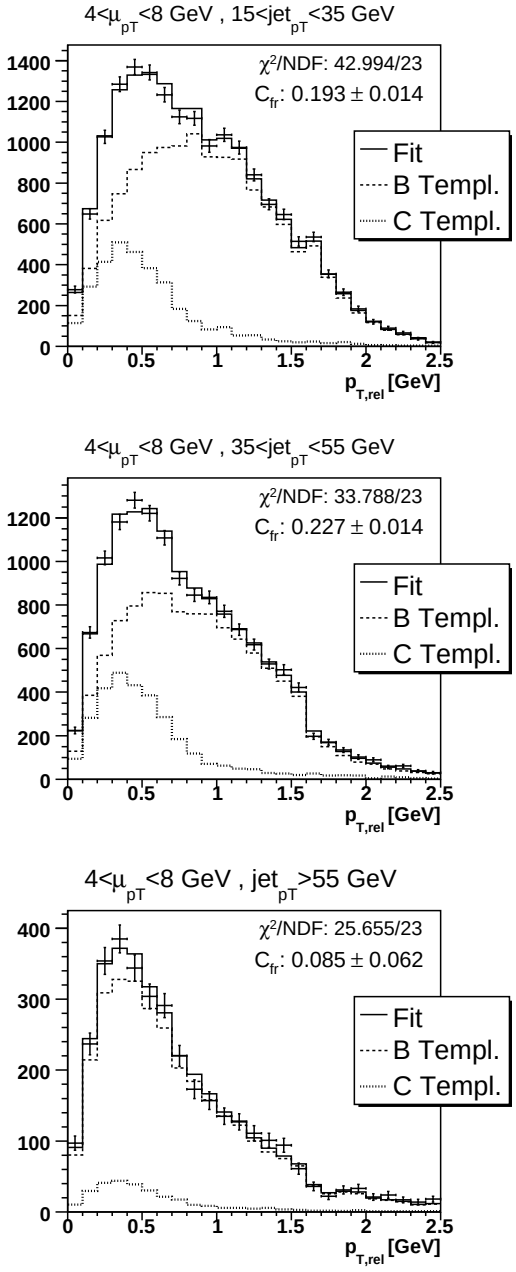

Figure 4.9. The $p_{T,\mathrm{rel}}$ template fit in the loose dijet sample for muon p_T between 4 GeV and 8 GeV and three bins of jet p_T. Similarly for the other two muon p_T bins.

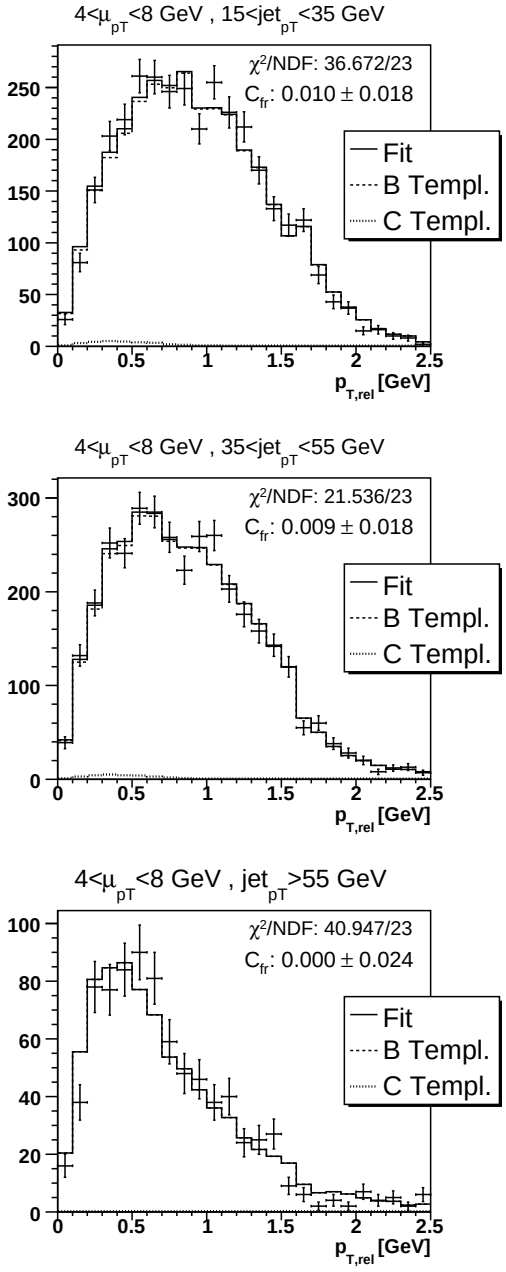

Figure 4.10. The $p_{T,\mathrm{rel}}$ template fit in the tight dijet sample for muon p_T between 4 GeV and 8 GeV and three bins of jet p_T. Similarly for the other two muon p_T bins.

Table 4.6. Result of the $p_{T,\mathrm{rel}}$ fit for the loose dijet sample for the nine different combinations of muon p_T and jet p_T.

Muon p_T	Jet p_T	Fitted c-fraction
$4 < p_{T\mu} < 8$ GeV	$15 < p_T < 35$ GeV	0.19 ± 0.01
$4 < p_{T\mu} < 8$ GeV	$35 < p_T < 55$ GeV	0.23 ± 0.01
$4 < p_{T\mu} < 8$ GeV	$p_T > 55$ GeV	0.09 ± 0.06
$8 < p_{T\mu} < 10$ GeV	$15 < p_T < 35$ GeV	0.15 ± 0.03
$8 < p_{T\mu} < 10$ GeV	$35 < p_T < 55$ GeV	0.24 ± 0.02
$8 < p_{T\mu} < 10$ GeV	$p_T > 55$ GeV	0.07 ± 0.06
$p_{T\mu} > 10$ GeV	$15 < p_T < 35$ GeV	0.18 ± 0.04
$p_{T\mu} > 10$ GeV	$35 < p_T < 55$ GeV	0.21 ± 0.02
$p_{T\mu} > 10$ GeV	$p_T > 55$ GeV	0.14 ± 0.03
Weighted sample average:		0.19 ± 0.02

Table 4.7. Result of the $p_{T,\mathrm{rel}}$ fit for the tight dijet sample for the nine different combinations of muon p_T and jet p_T.

Muon p_T	Jet p_T	Fitted c-fraction
$4 < p_{T\mu} < 8$ GeV	$15 < p_T < 35$ GeV	0.01 ± 0.02
$4 < p_{T\mu} < 8$ GeV	$35 < p_T < 55$ GeV	0.01 ± 0.02
$4 < p_{T\mu} < 8$ GeV	$p_T > 55$ GeV	0.00 ± 0.02
$8 < p_{T\mu} < 10$ GeV	$15 < p_T < 35$ GeV	0.01 ± 0.04
$8 < p_{T\mu} < 10$ GeV	$35 < p_T < 55$ GeV	0.02 ± 0.03
$8 < p_{T\mu} < 10$ GeV	$p_T > 55$ GeV	0.00 ± 0.03
$p_{T\mu} > 10$ GeV	$15 < p_T < 35$ GeV	0.02 ± 0.05
$p_{T\mu} > 10$ GeV	$35 < p_T < 55$ GeV	0.01 ± 0.02
$p_{T\mu} > 10$ GeV	$p_T > 55$ GeV	0.000 ± 0.004
Weighted sample average:		0.01 ± 0.02

Table 4.8. Fraction of times the tagging muon changes sign x_{flip} in four different bins of *tag*-jet p_T. The value of x_{flip} plugged into Eq. 4.8 is the weighted average to take into account the *tag*-jet p_T spectra of the dijet samples.

Jet p_T	x_{flip}
$15 < p_T < 25$ GeV	0.257 ± 0.064
$25 < p_T < 35$ GeV	0.290 ± 0.013
$35 < p_T < 45$ GeV	0.305 ± 0.009
$p_T > 45$ GeV	0.317 ± 0.009
Weighted sample average	0.303 ± 0.006

in each bin. As expected, removing the requirement of a SVT-tag on the *tag*-jet enhances the fraction of c-quark jets.

4.4.4 Determination of the Tagging Muon Charge Flip Fraction

The total fraction of times the tagging muon charge sign is "flipped" compared to the quark that initiated the jet in the tight (loose) dijet sample is given by the variable x_{flip} (x'_{flip}). The value of x_{flip} is measured from simulated $b\bar{b}$ events where the Monte Carlo truth information allows for a determination of the correlation between the tagging muon and the quark sign. The fraction of times the tagging muon changes sign depends on the p_T spectrum of the jets. At lower jet p_T, the probability that a cascade muon is found above the reconstruction threshold is lower and the fraction of cascade muons thus increases as a function of increasing jet p_T. The fraction of times the tagging muon changes sign determined from simulated events is therefore weighted to the p_T spectrum observed in the tight dijet sample. The result is shown in Tab 4.8. Any topological differences between the loose and tight dijet sample such as the jet p_T spectrum could lead to a difference between x_{flip} and x'_{flip}. Therefore, the *tag*-jet p_T and η spectra are compared (see Fig. 4.11) but no significant differences were found and $x_{\text{flip}} = x'_{\text{flip}}$ is thus used.

Cross-check of tagging muon charge flip

The fraction of times the tagging muon in the loose and tight dijet samples changes sign calculated above can be cross-checked with data by requiring that the *probe*-jet in the tight dijet sample contain a muon track (with the same quality as the tagging muon). If $N_{b\bar{b}}$ is the total number of events in this sample, N_{os} is the number of events with two opposite sign muons and N_{ss} the number of events with two same sign muons, the following relationships can be used to predict the number of N_{ss} and N_{os} in this sample given the flip fraction estimated above,

$$N_{b\bar{b}} = N_{ss} + N_{os} \tag{4.9}$$

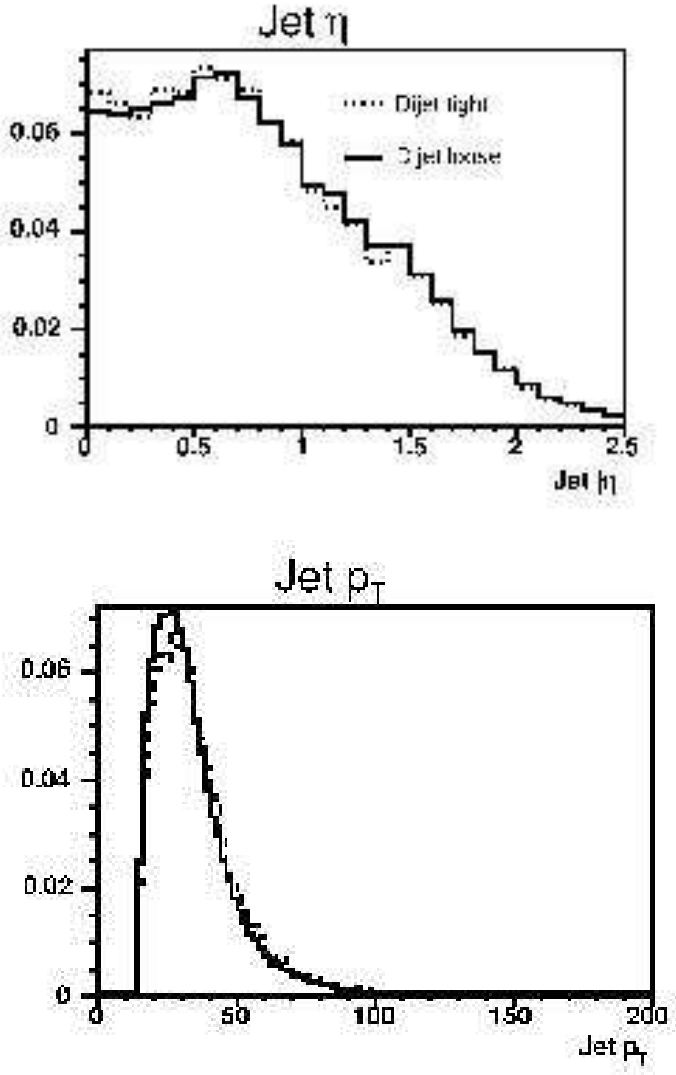

Figure 4.11. Comparison of jet η (top) and p_T (bottom) for *tag*-jets in the loose (solid) and tight (dashed) dijet sample.

$$N_{ss} = 2N_{b\bar{b}}x_{\text{flip}}(1 - x_{\text{flip}}). \quad (4.10)$$

The prediction gives $N_{ss} = 76 \pm 6$ and $N_{os} = 105 \pm 15$ while the observation is $N_{ss} = 79$ and $N_{os} = 103$ which is consistent within the statistical uncertainty. The caveat of this cross-check is that a rather broad range of x_{flip} is allowed before conflicting with the statistical uncertainties.

4.4.5 Correction for Kinematical Differences in the Signal and Dijet Samples

The jet charge templates were derived from the dijet data samples above. The aim is to extract the jet charge templates for b- and c-quark jets in $t\bar{t} \to \ell$+jets events. Figure 4.12 shows a comparison of the jet p_T and η for the *probe*-jet in the tight dijet sample and the SVT-tagged b-quark jets in simulated $t\bar{t} \to \ell$+jets events. The b-quark jet from the top quark decay has as expected harder p_T and more central η spectra. From the correlation between jet p_T and the number of tracks associated with the jet one can expect that the discriminating power depends on the jet p_T. A dependence on jet $|\eta|$ is also expected due to the geometry of the inner tracking detector where particles around $|\eta| = 1.0$ traverses more layers in the SMT (a similar behavior exists for the tracking efficiency in Fig. 3.9 in Sec. 3.6.4). The jet p_T and $|\eta|$ dependences of the jet charge algorithm are shown in Fig. 4.13. As the jet charge algorithm performance is improving with increasing jet p_T and mostly decreasing with increasing $|\eta|$ the conclusion is that the jet charge templates derived on the dijet samples are underestimating the jet charge algorithm discriminating power when applied to b-quark jets in $t\bar{t} \to \ell$+jets events.

This section describes how to correct for this by weighting simulated $t\bar{t}$ events to have the same b-quark jet p_T and η spectrum as the dijet samples. A correction function is then extracted based on the deviation of the weighted compared to the unweighted jet charge templates. This correction is subsequently applied to the jet charged templates derived from the dijet samples to extract the expected jet charge templates for b- and c-quark jets in $t\bar{t} \to \ell$+jets events.

Jet Kinematical Weighting

To fully take into account the kinematical differences for b-quark jets in the dijet samples and $t\bar{t} \to \ell$+jets both the difference in jet p_T and η and their correlations have to considered. This can be achieved by weighting the b-quark jet p_T spectrum of $t\bar{t} \to \ell$+jets to the *probe*-jet p_T spectrum in the tight dijet sample (this direction of weighting is the only possible due to the low statistics at high p_T in the dijet samples). Figure 4.14 shows the weight and the result is shown in Fig. 4.15. The weighting of the p_T spectrum does not take into account the differences in η spectrum between the two samples. This difference is treated in a similar fashion. The simulated $t\bar{t} \to \ell$+jets events are weighted again but this time with respect to the b-quark jet η spectrum. The weight as a function of η is shown in Fig. 4.16. After applying both weights the jet p_T and η spectrum of the different samples

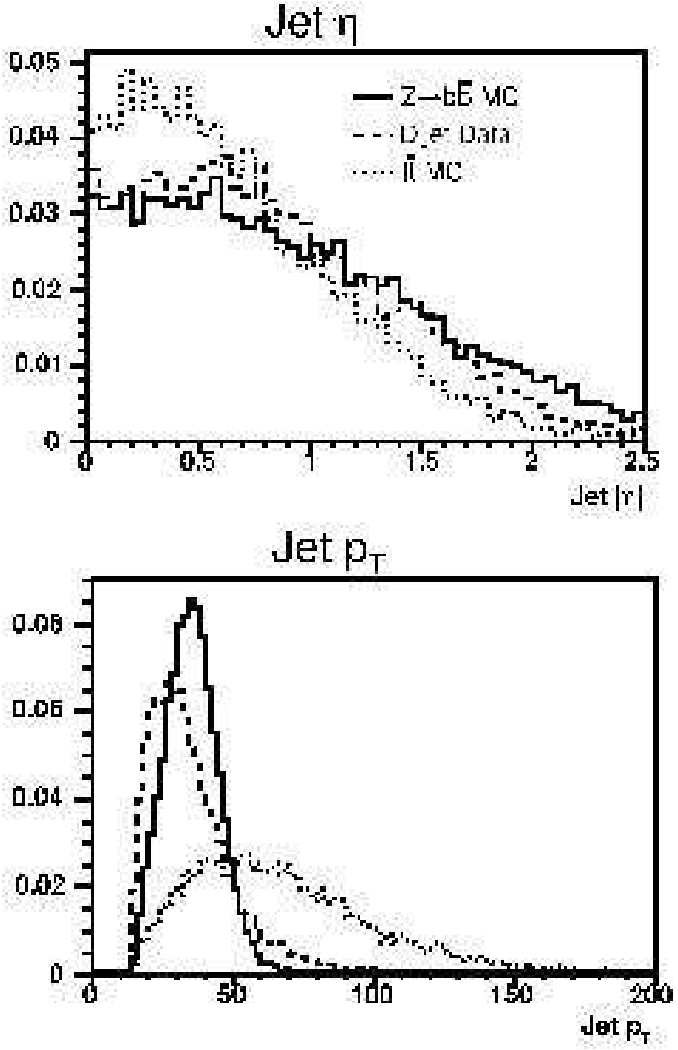

Figure 4.12. Jet η (top) and p_T (bottom) for the *probe*-jet in the tight dijet sample, in simulated $Z \to b\bar{b}$ events and for SVT-tagged jets in simulated $t\bar{t} \to \ell$+jets events.

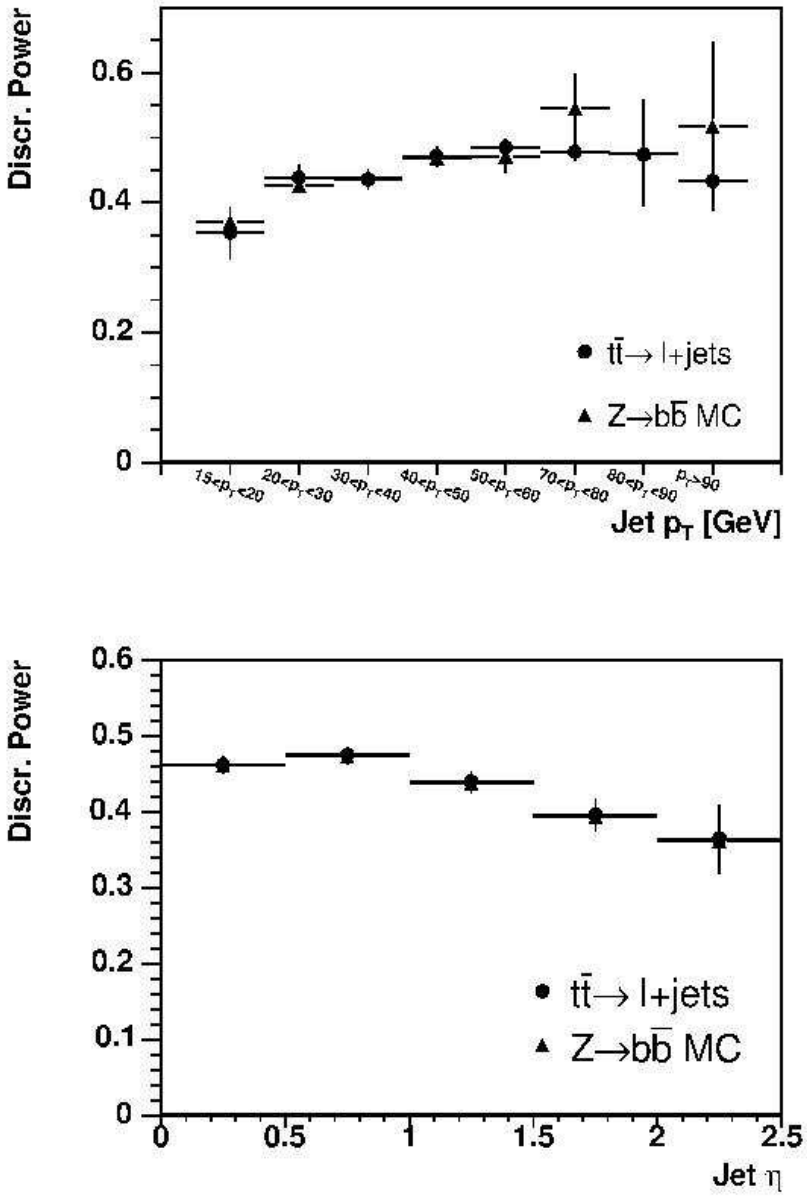

Figure 4.13. Discriminating power for SVT-tagged b-quark jets as a function of jet p_T (top) and η (bottom) for simulated $t\bar{t} \to \ell$+jets (circles) and $Z \to b\bar{b}$ events (triangles).

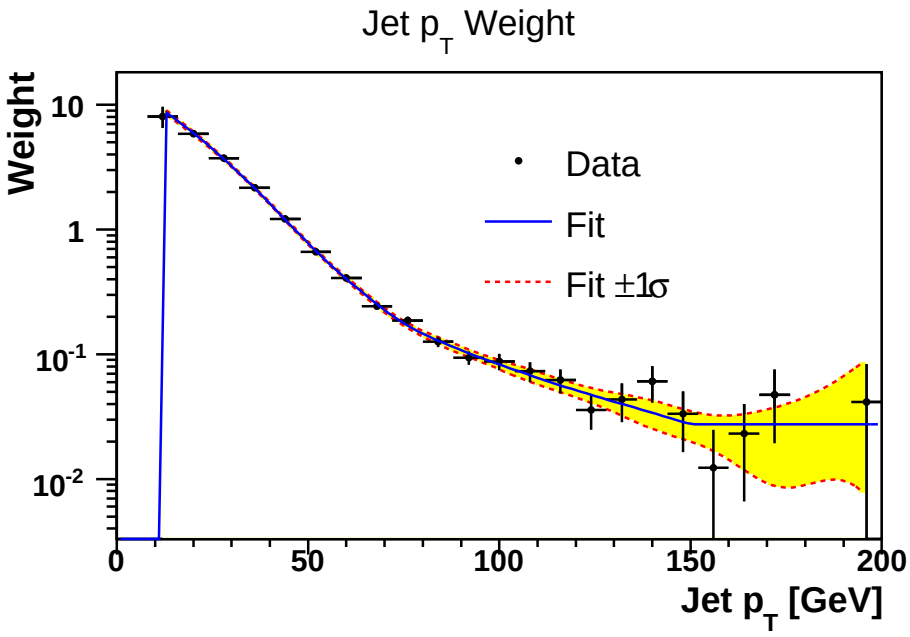

Figure 4.14. The weight applied to the b-quark jet p_T spectrum of simulated $t\bar{t} \to \ell$+jets events to obtain the same p_T spectrum as the *probe*-jet in the tight dijet sample.

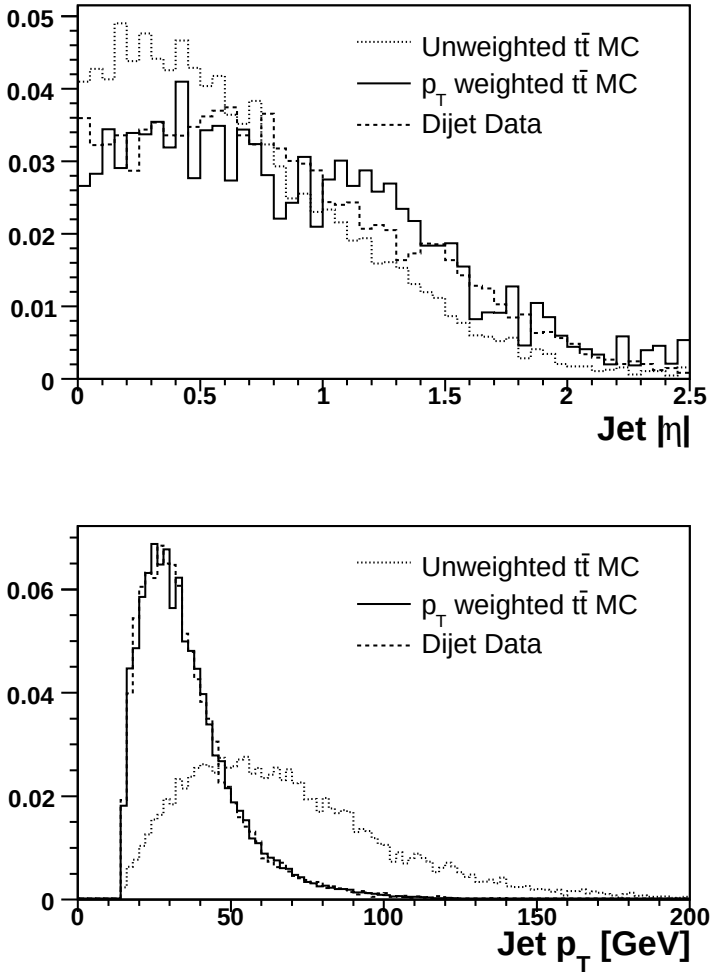

Figure 4.15. Comparison of jet η (top) and p_T(bottom) for SVT-tagged b-quark jets in simulated $t\bar{t} \to \ell$+jets events passing signal selections before and after weighting of the p_T spectrum. The p_T spectrum of *probe*-jets in the tight dijet sample is shown for comparison.

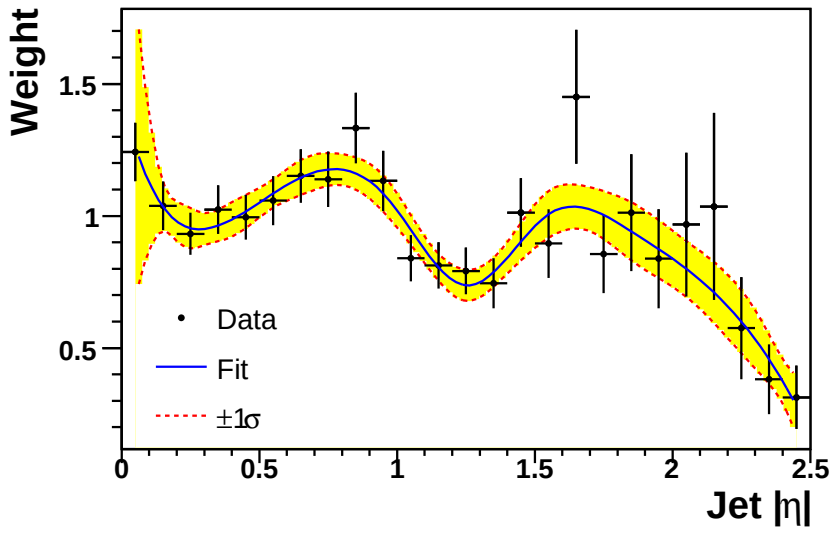

Figure 4.16. The η weight applied to the p_T weighted b-quark jets in simulated $t\bar{t} \to \ell$+jets events.

agree as can be seen in Fig. 4.17, confirming the validity of the assumption of uncorrelated p_T and η.

Derivation of the Kinematical Correction

The correction is extracted as the ratio between the b-quark jet charge templates in normal and re-weighted simulated events. The jet charge templates can be seen as probability densities to observe a certain jet charge given the true charge and type of quark. In the following, the jet charge templates are modeled as functions of the measured jet charge Q_{jet} defined in Eq. 4.1. Further, we denote:

- $f_b^{\text{data}}(Q_{\text{jet}})$ the b-quark jet charge template obtained directly from dijet data (as described in Sec. 4.4.2) before any correction,
- $f_b^{t\bar{t}}(Q_{\text{jet}})$ the b-quark jet charge template from simulated $t\bar{t} \to \ell$+jets events passing signal selections,
- $f_b^{t\bar{t},\omega}(Q_{\text{jet}})$ the b-quark jet charge template from simulated $t\bar{t} \to \ell$+jets events passing signal selections re-weighted as discussed in the previous section,
- $f_b(Q_{\text{jet}})$ the jet charge distribution from the tight dijet data, corrected to reproduce the jet charge distribution for jets with the same p_T and η as the simulated $t\bar{t}$ events.

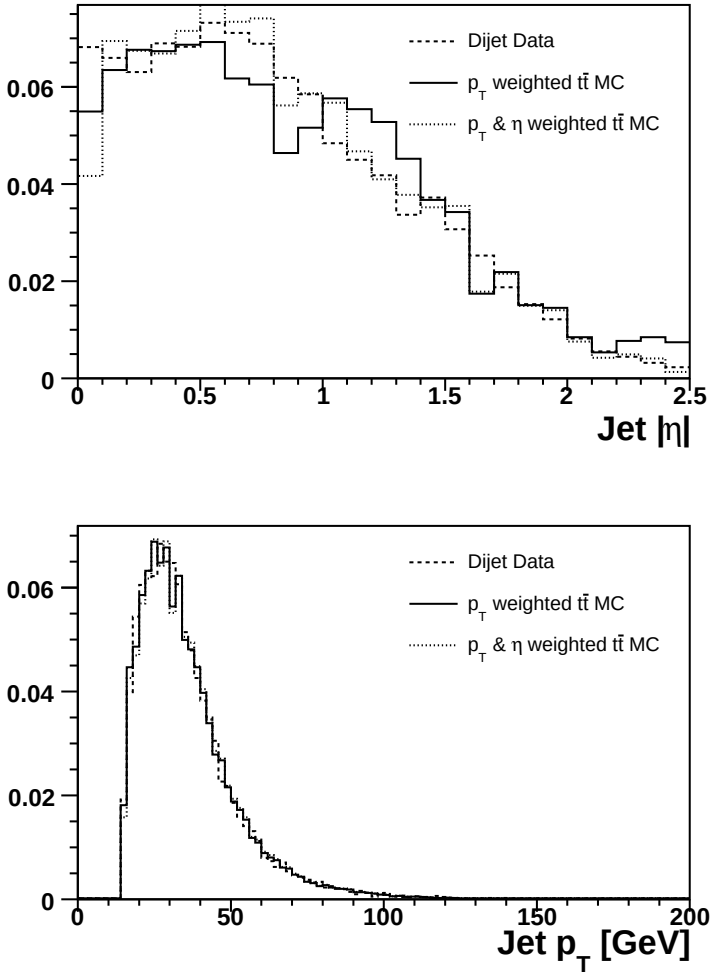

Figure 4.17. Comparison of jet η (top) and p_T(bottom) for b-quark jets in simulated $t\bar{t} \to \ell$+jets events passing signal selections before and after weighting. The *probe*-jet p_T and η spectrum in the tight dijet sample is shown for comparison.

The corrected distribution is obtained in the following way:

$$f_b(Q_{\rm jet}) = f_b^{\rm data}(Q_{\rm jet}) \cdot \frac{f_b^{t\bar{t}}(Q_{\rm jet})}{f_b^{t\bar{t}_\omega}(Q_{\rm jet})} \tag{4.11}$$

The correction for the $\bar{b}$-quark jet charge template is derived in a similar way. The correction functions are shown in Fig. 4.18 for both b- and $\bar{b}$-quark jets. As expected, the corrections decrease the width of the jet charge templates since a higher jet p_T give a larger probability to observe a higher number of tracks associated with the jet. The more tracks associated with the jet means less probability to observe $|Q_{\rm jet}|$ close to one. Figure 4.19 shows the discriminating power for b-quark jets before and after the kinematical correction.

4.4.6 Final Jet Charge Distributions Extracted from Data

After applying the correction for the kinematical differences in the dijet and signal samples the final b-, $\bar{b}$-, $c-$ and $\bar{c}$-quark jet charge templates shown in Fig. 4.20 are obtained. The jet charge templates are normalized to an area of one and can be seen as the probability density to measure a certain jet charge $Q_{\rm jet}$, given the type of quark (b, $\bar{b}$, c or $\bar{c}$) initiating the SVT-tagged jet. These templates are subsequently used to derive the expected charge templates for the SM top and the exotic quark.

Figure 4.21 shows the discriminating power of the final b-quark jet charge templates compared to the tight dijet sample after basic selections. The error band is the combined systematic uncertainty related to the extraction of the templates which is propagated to the final result. As required, the discriminating power of the extracted jet charge templates is independent of the minimum $p_{T,\rm rel}$ requirement on the tagging muon (remember that the final jet charge templates are derived with no requirement on the $p_{T,\rm rel}$ of the opposite tagging muon). As expected, the jet charge algorithm is less efficient for data than for Monte Carlo events where the Monte Carlo history is available to obtain the true type of quark initiating the jet.

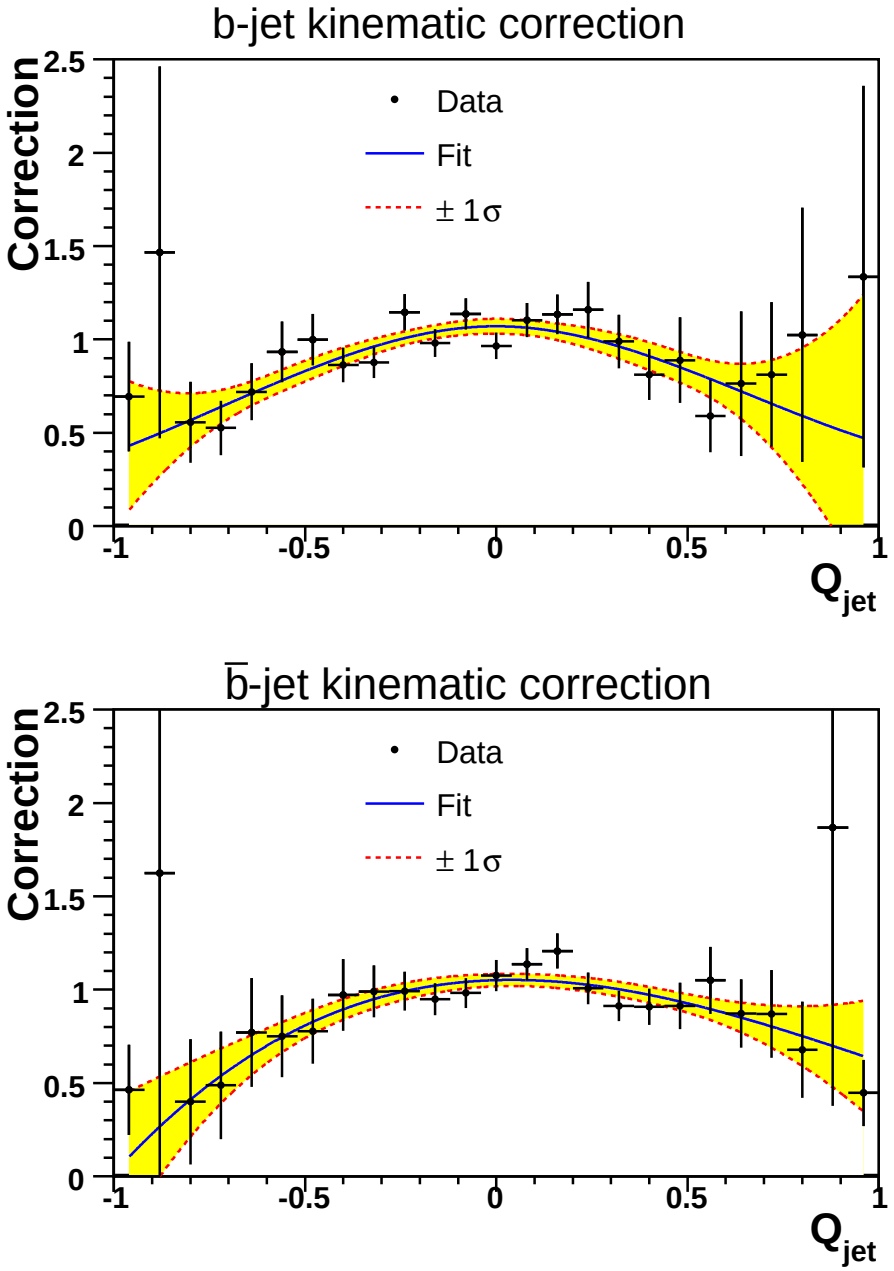

Figure 4.18. The kinematical correction function for b- (top) and $\bar{b}$-quark jets (bottom).

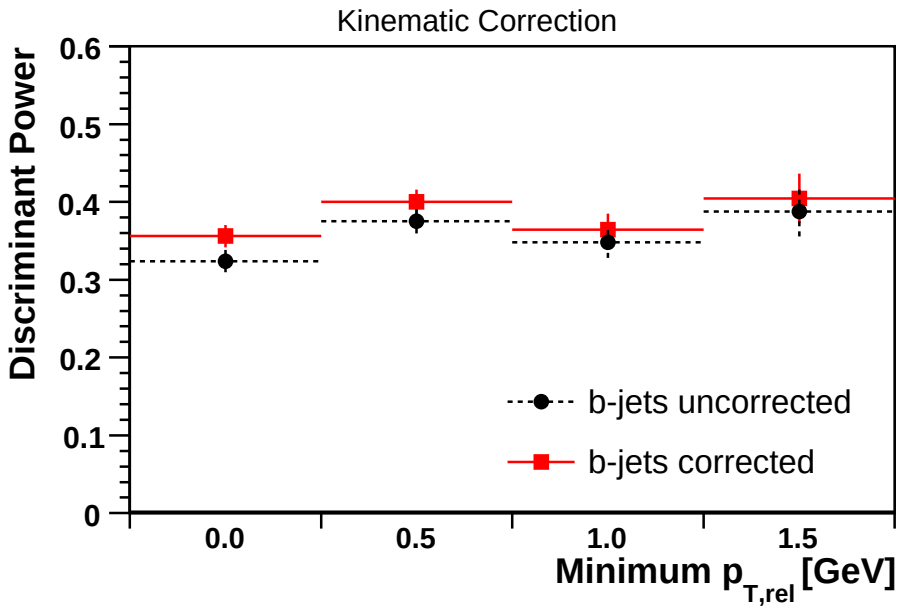

Figure 4.19. Discriminating power for b-quark jet charge templates before and after the kinematical correction is applied as a function of the $p_{T,\mathrm{rel}}$ cut. As expected the kinematical correction increases the discriminating power.

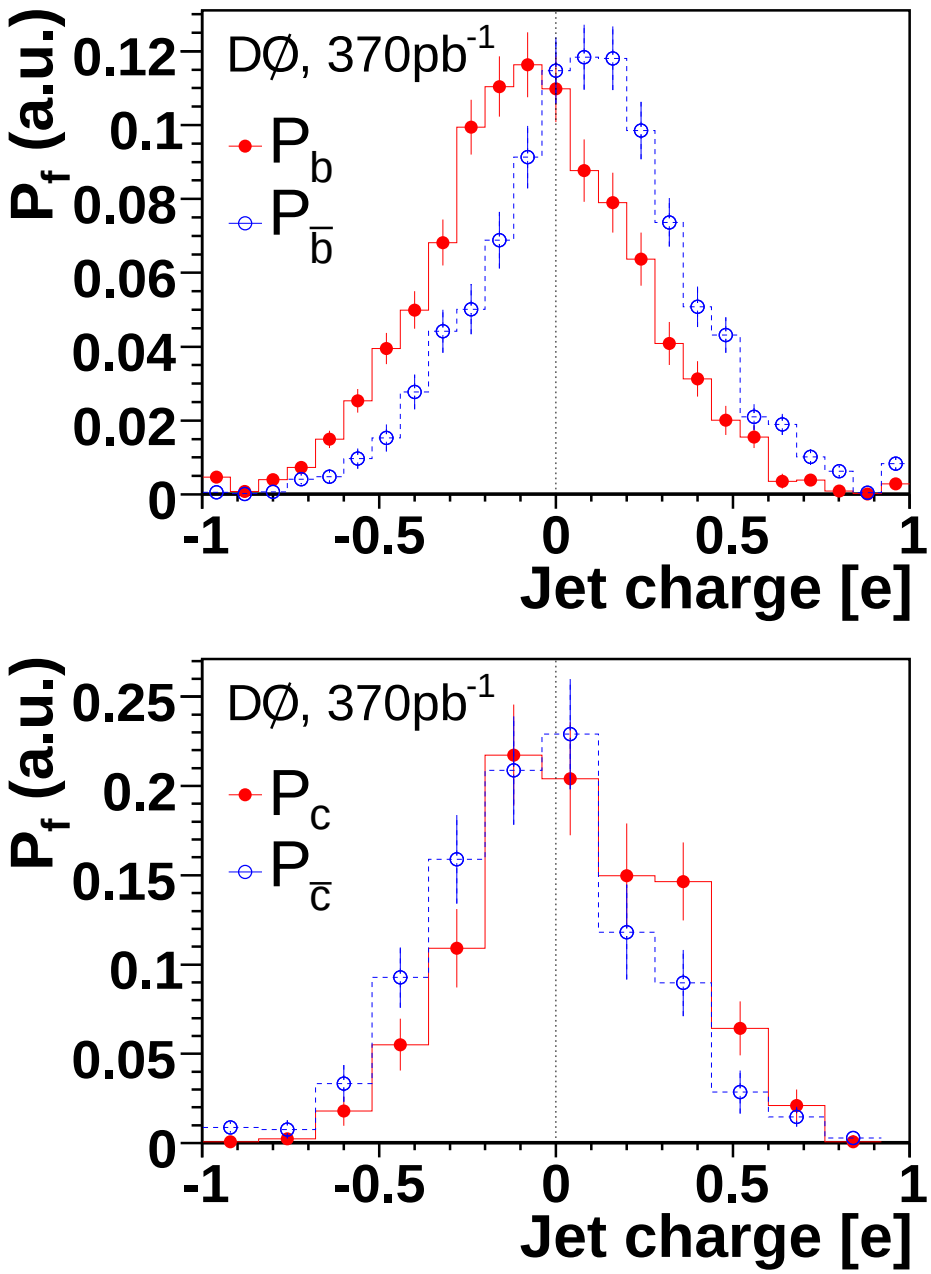

Figure 4.20. Jet charge templates for b- and $\bar{b}$-quark jets (top) and c- and $\bar{c}$-quark jets (bottom) extracted from data.

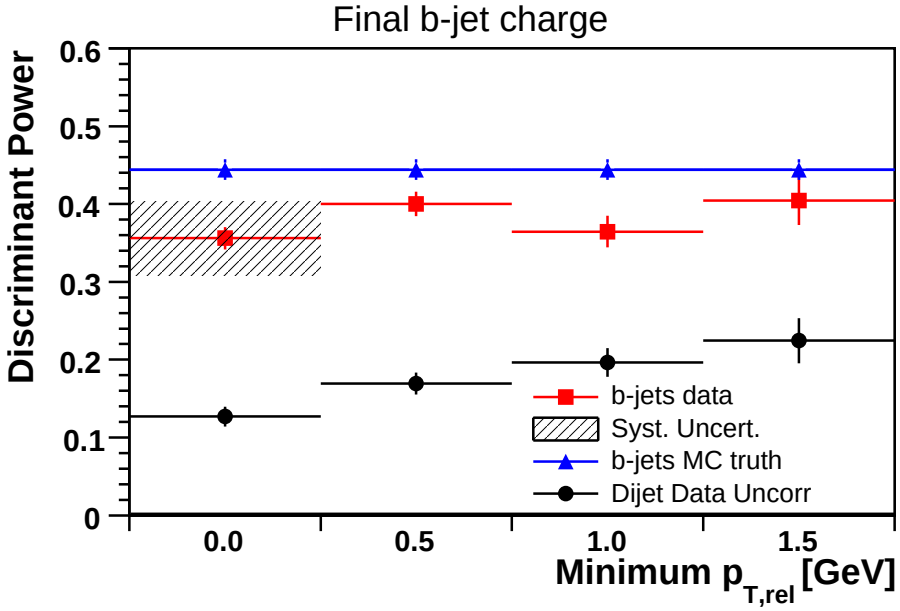

Figure 4.21. Discriminating power as function of the $p_{T,\mathrm{rel}}$ requirement for b-quark jets after all corrections and in the tight dijet sample after selection. The discriminating power in simulated events using the Monte Carlo history to find the true charge of the quark initiating the jet is shown for comparison.

4.5 Top Quark Charge Observables

In order to discriminate between the charge $2e/3$ SM top quark and the charge $4e/3$ exotic quark scenarios an event observable and an expectation of this observable for the two different hypotheses is needed. The charge $2e/3$ SM top quark and the charge $4e/3$ exotic quark scenarios will simply be refereed to as the SM and exotic scenarios respectively.

The observable is based on reconstructing the charge of the decay products of the top quark, assuming as mentioned earlier the decay $t \to bW^+$ (and similarly in the exotic charge scenario $Q \to bW^-$). Since there are two top quarks in each event, the top quark charge can be measured twice in each event.

One top quark charge is constructed as the sum of the charge of the charged isolated lepton (referred to as the charged lepton if not otherwise specified) and the jet charge of the b-quark jet from the same top quark. The second top quark charge in the event is constructed as the sum of the second b-quark jet charge *minus* the charge of the charged lepton. The two observables in each event are defined as:

$$\begin{aligned} Q_1 &= |q_\ell + q_b| \\ Q_2 &= |-q_\ell + q_B| \end{aligned} \tag{4.12}$$

where q_ℓ is the charge of the charged lepton, q_b is the charge of the b-quark jet on the leptonic leg of the event (defined in Sec. 4.1) and q_B is the charge of the b-quark jet in the hadronic leg of the event. The charges of the SVT-tagged jets q_b and q_B are obtained by applying the jet charge algorithm discussed earlier. The next section describes how the two b-quark jets are assigned to the two legs of the event.

4.5.1 Associating SVT-Tagged Jets to the Correct W Boson

Recall that the final state objects from the decay of $t\bar{t}$ pair in the ℓ+jets channel are two jets from the decay of the W boson and a b-quark jet from the hadronic side of the event and a b-quark jet together with a charged lepton and a neutrino from the leptonic side of the event. In order to compute the top quark charge observables, the two b-quarks in the event needs to be assigned to two reconstructed jets. This is done using a kinematic fitting algorithm initially developed to measure the top quark mass [79]. Below, the main features of the kinematic fit is described, more detailed information can be found in Ref. [69].

Kinematic Fit

The algorithm used performs a kinematic fit of top quark pair candidate events in the ℓ+jets topology. If the event contains the decay of a $t\bar{t}$ pair, with four jets (the issue regarding events with more than four jets is discussed below) in addition to the

charged lepton and neutrino, the three jets (two light jets from the W boson and a b-quark jet) forming the invariant mass of the "hadronic" top quark (the top quark on the hadronic side of the event) is expected to be equal to the invariant mass formed by the charged lepton, neutrino and remaining b-quark jet (the leptonic side of the event). In addition to this constraint, two jets (out of three) on the hadronic side of the event is expected to form the invariant mass of the W boson as is the charged lepton and the neutrino on the leptonic side of the event. Together with the overall energy-momentum conservation in the collision this over-constrains the problem.

The input variables to the kinematic fit are:

- The measured energy (or momentum) of the four jets and the charged lepton (3 components each) and their directions. The masses of the jets are fixed to zero except for the jets assigned as b-quark jets which are given a mass of 4.6 GeV.

- The x and y components of the measured missing transverse energy, $\not\!\!E_T$, represents the transverse momentum of the neutrino. Note that the neutrino momentum in the longitudinal direction (p_z^ν) cannot be inferred from momentum imbalance in a similar way due to the spectator quarks in the colliding $p\bar{p}$ pair which carries away a large fraction of the momentum in this direction which is mostly unmeasured close to the beam direction.

The conclusion is that there are 17 measured and one unmeasured variable (the neutrino momentum in the z direction) under subject to three constraints when reconstructing the top quark event,

$$m_W^{\text{leptonic side}} = 80.4 \text{ GeV}, \tag{4.13}$$

$$m_W^{\text{hadronic side}} = 80.4 \text{ GeV}, \tag{4.14}$$

$$m_t = m_{\bar{t}}, \tag{4.15}$$

which makes the problem twice over-constrained[3]. In this analysis, which is not concerned with measuring the mass of the top quark, the top quark mass itself is used as an additional constraint and fixed to 175 GeV (which is the top quark mass used in the generation of the simulated $t\bar{t}$ events).

If the correspondence between jets and partons were known, a kinematic fit would not be needed. In general, this is not known. Therefore, the algorithm tries all 12 possible permutations of jets and the "best" one can be chosen (there are 24 permutations for four jets but the exchange of the two jet from the W boson

[3]There are 13 particles in the event, 4 jets, two leptons, the $p\bar{p}$ pair, $W^\pm$, the $t\bar{t}$ pair and an additional pseudo-particle X to obtain total four-momentum balance which gives 52 (13×4) variables. The constraints come from the quarks, lepton and $\not\!\!E_T$ giving 23 $\left[5 \times 4 + 3\ (p_x^\nu, p_y^\nu), \text{massoftheneutrino}\right]$ constraints. The $p\bar{p}$ pair gives an additional 8 (2×4), four-momentum conservation in all 5 vertices gives 20, the known masses of $W^\pm$ gives 2 and $m_t = m_{\bar{t}}$ gives one constraint. This leads in total to 54 constraints for 52 variables and thus twice over-constrained.

leaves the fit unchanged). To quantify "best", a "χ^2" is defined which reflects to what degree it is possible to satisfy the given constraints for each jet permutation. The kinematic fit iteratively changes the measured and assigned kinematic variables based on their uncertainties and their impact on the χ^2. For a given jet permutation the variables are pulled until the χ^2 stops changing and the constraints are satisfied.

For events with more than four jets the additional jets are assumed to originate from initial (ISR) or final state radiation (FSR). The number of permutations grows fast when including extra jets in the fit (note that if the extra jet is assumed to be FSR it must be merged with another jet in the event and 5 jets give 140 permutations, 6 jets give 1020, etc.). As in the top mass analyses [79], this analysis uses only the four highest p_T jets in the fit. Other jets are assumed to arise from ISR [69].

Consider now $t\bar{t} \to \ell$+jets events where exactly two of the jets in the event are SVT-tagged. These b-tagged jets are denoted as j_{SVT1}, ..., j_{SVTn}, with $n = 2$ and the other jets j_1 ..., j_n, with $n \geq 2$. If j_b and j_B are denoted as the b-quark jets from the leptonic and hadronic leg of the event respectively, and j_{w1} and j_{w2} the two jets from the hadronic W decay, then each permutation considered by the kinematic fit associates one jet (j_{SVT1}, j_{SVT2}, j_1 and j_2) to each of the jets j_b, j_B, j_{w1} and j_{w2}. There are exactly 12 combinations, but only two for which the SVT-tagged jets are associated to the two jets j_b and j_B from the b-quarks of the leptonic and hadronic legs:

1. $j_{SVT1} \leftrightarrow j_b$ and $j_{SVT2} \leftrightarrow j_B$, or
2. $j_{SVT2} \leftrightarrow j_b$ and $j_{SVT1} \leftrightarrow j_B$.

For an event with more than two SVT-tagged jets the number of permutations for which the kinematic fit can associate SVT-tagged jets to j_b or j_B is multiplied. Note that only 1% of $t\bar{t}$ events in Monte Carlo passing preselection have three or more SVT-tagged jets.

> Only permutations for which j_b and j_B are associated to SVT-tagged jets are considered.

Sometimes the input kinematic variables are very far from the constraints and the fit may therefore fail to pull the variables enough to satisfy the constraints or it does not find a stable χ^2 and fails to converge. The kinematic fit fails to converge in $17.6 \pm 0.6\%$ of simulated $t\bar{t}$ events passing signal selections for any of the two permutations where the SVT-tagged jets are associated to j_b and j_B, see Tab. 4.9. One should note that it is not seldom that the W decay contains a c-quark jet, and the SVT-tagged jets do not correspond to 100% of actual b-quark jets, see Tab. 4.10. Finally, out of the permutations for which the kinematic fit assigns the SVT-tagged jet to j_b and j_B, only the permutation with the lowest χ^2 is used. Based on the Monte Carlo history, it is observed that the lowest χ^2 permutation is the correct one in $83.8 \pm 1.8\%$ of the cases. In the signal sample, 21 events are selected, out of

Number of permutations	Tagged $t\bar{t}$ MC (%)	Data Pred. tagged $t\bar{t}$ MC (%)
0	17.6 ± 0.6	17.7 ± 0.8
1	5.6 ± 0.3	6.1 ± 0.4
2	71.9 ± 1.4	73.0 ± 2.0
3	0.2 ± 0.1	0.2 ± 0.1
4	1.9 ± 0.2	1.2 ± 0.2
5	0.20 ± 0.05	0.03 ± 0.03
6	2.6 ± 0.2	1.7 ± 0.2

Table 4.9. Number of permutations where the kinematic fit converges for simulated $t\bar{t} \to \ell$+jets events. More than two permutations can arise in events where there are more than two SVT-tagged jets. In the rightmost column the SVT-tagged jets in the event was predicted by using the tagging efficiencies derived from data.

Jet flavor	Lowest χ^2 (%)	Double b-tagged events (%)	Data pred. b-tagged events (%)
no match	0.00 ± 0.00	3.40 ± 0.16	3.64 ± 0.24
u	0.37 ± 0.06	0.45 ± 0.06	0.33 ± 0.07
d	0.37 ± 0.06	0.47 ± 0.06	0.40 ± 0.08
s	0.33 ± 0.05	0.45 ± 0.06	0.27 ± 0.06
c	4.90 ± 0.22	7.16 ± 0.24	7.53 ± 0.35
b	94.03 ± 1.29	88.07 ± 1.10	87.84 ± 1.56

Table 4.10. Jet flavor for the SVT-tagged jets in simulated $t\bar{t} \to \ell$+jets events passing signal selections. The permutation giving the lowest χ^2 in the kinematic fit gives as expected a higher fraction of b-quarks matched to the jets compared to any permutation (middle and last column). Using the tagging efficiency derived from data gives a similar result as for using the explicit SVT-tags in the Monte Carlo events. The true jet flavor is found by matching the closest parton within a cone of $\Delta R < 0.5$ to the jet axis.

which 16 have at least one converged permutation where both SVT-tagged jets are assigned to j_b and j_B. This is statistically consistent with the prediction $17.6 \pm 0.6\%$ from simulated events.

4.5.2 Expected Charge Templates in the Standard Model and Exotic Scenarios

The expected charge distribution templates in both scenarios are derived from simulated $t\bar{t}$ events and the jet charge templates derived from data earlier.

Standard Model Charge Template

The estimated charge distribution of the SM top and the exotic quark depends on how well the b- and $\bar{b}$-quark jets can be identified but also the assignment of the correct b-quark jet to correct W boson in the event. Therefore, it is necessary that the kinematic fit is applied to the simulated events in an equivalent way as to data in order to extract the expected distributions.

The procedure to obtain the expected SM top quark charge distribution is explained below:

1. Apply signal selection criteria as defined in Sec. 4.2 on simulated $t\bar{t}$ events.
2. Fit the event using the constrained kinematic fit.
3. Select the lowest χ^2 combination for which the jets j_b and j_B are associated to SVT-tagged jets j_{SVT1} and j_{SVT2}.
4. Determine the true flavor of the jets j_b and j_B using jet-parton matching. The true flavor can be: b-, $\bar{b}$-, c- , $\bar{c}$- or light (u, d, s) jet.
5. The jet charges q_b and q_B is set to one randomly chosen value according to the probability density function $f(Q_{\text{jet}})$ derived from data depending on the flavor of the jet e.g. if the flavor is $\bar{b}$, then the probability density function $f_{\bar{b}}(Q_{\text{jet}})$. If it is a light jet, then a random value from the corresponding jet charge probability density function is used, derived from Monte Carlo[4].
6. Compute the two observables $Q_1 = |q_\ell + q_b|$ and $Q_2 = |-q_\ell + q_B|$.
7. Make an entry for Q_1 and another entry for Q_2 in a histogram to store the expected charge distribution for the SM top quark.

To decrease the statistical uncertainty due to the sampling of the jet charge distributions step 5, steps 6 and 7 are carried out 200 times per event. The predicted SM top quark charge distribution referred to as p_{sm} is shown in Fig. 4.22.

[4]It is very seldom we need to sample the light jet charge templates due to the low probability to tag a light jet ($\approx 1\%$), see Tab. 4.10

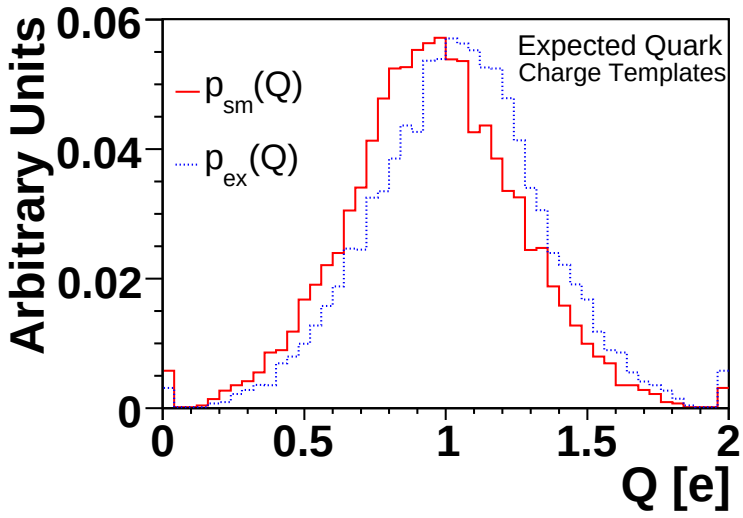

Figure 4.22. The predicted SM top ($p_{\rm sm}$) and exotic ($p_{\rm ex}$) quark charge templates including backgrounds.

Exotic Scenario Charge Template

The estimated charge distribution template for the exotic scenario is derived in a similar way as for the SM top quark. From Fig. 4.1 it can be seen that it can be obtained by simply replacing the above step 6 by the following

6 . The two expected observables are derived by permuting the jet charge of the SVT-tagged jet on the leptonic and hadronic leg of the event: $Q_1 = |q_\ell + q_B|$ and $Q_2 = |-q_\ell + q_b|$.

The expected charge template for the exotic scenario is also shown in Fig. 4.22. Note that with this procedure the charge template for the exotic scenario is a mirror distribution of the Standard Model top quark charge template.

4.5.3 Backgrounds

The signal sample contains in addition to $t\bar{t}$ events also a small fraction of background processes, mostly W boson production in association with a $b\bar{b}$ pair and two or more jets ($Wb\bar{b}jj$) but also some expected single top quark events. Details about the sample composition is described in Sec. 4.2. The observed charge in the signal sample is to be compared with a combination of the expected SM top quark and exotic quark charge templates described above and of the charge contribution from the background processes. Due to the large signal-to-background ratio in the signal sample ($\sim$ 11), the effect of adding the charge contribution from the background processes is small. The backgrounds considered are the $Wb\bar{b}jj$ and single top production while the other background processes are neglected. Two assumptions are

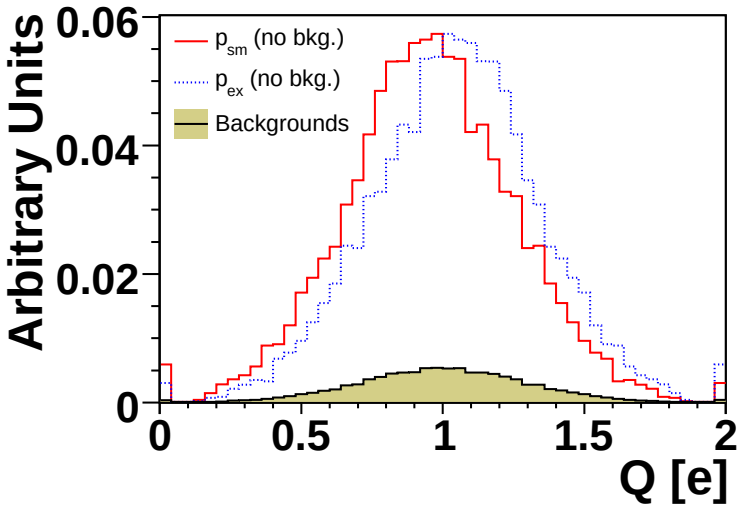

Figure 4.23. The combined background charge contribution compared to the SM top and exotic quark charge templates.

made: *i)* the charge of the isolated high p_T lepton is uncorrelated with the b-quark jets in $Wb\bar{b}jj$ events, *ii)* the b-quark jet charge in $Wb\bar{b}jj$ events are the same as for the b-quark jets in $t\bar{t}$ events. Due to the low expected contribution from the single top background (≈ 0.3 events) it is simply modeled by the same charge template as the $Wb\bar{b}jj$ background.

Figure 4.23 shows the background charge template contribution and the expected SM and exotic templates. The SM and exotic templates are normalized to an area of one and can be seen as the probability density functions to observe a certain charge Q in the SM and exotic scenarios. In the rest of this thesis we denote these probability density functions by $p_{\mathrm{sm}}(Q)$ and $p_{\mathrm{ex}}(Q)$ for the SM top quark and the exotic quark scenario respectively.

4.6 Systematic Uncertainties

In this section the sources of systematic uncertainties and the method to evaluate them are briefly described. A systematic uncertainty can affect the measurement in two ways: It can change the jet charge templates derived from data or it can affect the kinematic fit that assigns the SVT-tagged jets in the event to the correct W boson. The result in the end is an uncertainty on the SM top and exotic quark templates. Thus, for each systematic uncertainty, the SM top and exotic quark charge templates are re-derived taking into account the uncertainty considered. The result is a set of varied SM top and exotic quark templates for each systematic which is taken into account when extracting the final result.

Dependence on the Fragmentation Model

This analysis use as little input as possible from the Monte Carlo simulation in deriving the jet charge algorithm performance. For instance, the extraction of the correction function in Sec. 4.4.5, uses the ratio of Monte Carlo modeled variables as input. Nevertheless, the dependence of a different model for the fragmentation of b-quark jets has been investigated. As discussed in Sec. 3.8, the fragmentation of partons is modeled by the Lund string model. Figure 4.24 and 4.25 shows the difference in discriminating power between b- and $\bar{b}$-quark jet charge distributions compared to the alternative Bowler and Peterson fragmentation models respectively. No systematic uncertainty due to the fragmentation model needs to be accounted for since the b- and c-quark jet charge templates are derived from data and as noted above, the kinematic corrections are ratios of Monte Carlo distributions, resulting in a large cancellation of any existing difference among the fragmentation models.

Additional Jet Dependence

The jet charge templates are derived from dijet samples where exactly two reconstructed jets are required while $t\bar{t} \to \ell$+jets events have at least four jets. More activity in the event could affect the jet charge templates due to a higher number of tracks and possible overlap between tracks originating from different jets. This dependence is studied in Fig. 4.26 by comparing the discriminating power for b- and $\bar{b}$-quark jets in simulated $t\bar{t} \to \ell$+jets and $t\bar{t} \to \ell\ell$+jets events where only two jets are expected (additional jets can arise due to ISR and FSR). No sizable difference is found and thus no systematic uncertainty is considered.

Tagging Efficiency in Data and Monte Carlo

The b-tagging efficiencies are known to be different in simulated events compared to data. The method to determine the expected charge distributions for the SM top and exotic quark charge scenarios depends on the tagging efficiency for different jet flavors e.g. a SVT-tagged jet in a $t\bar{t}$ event is not always a b-quark jet but

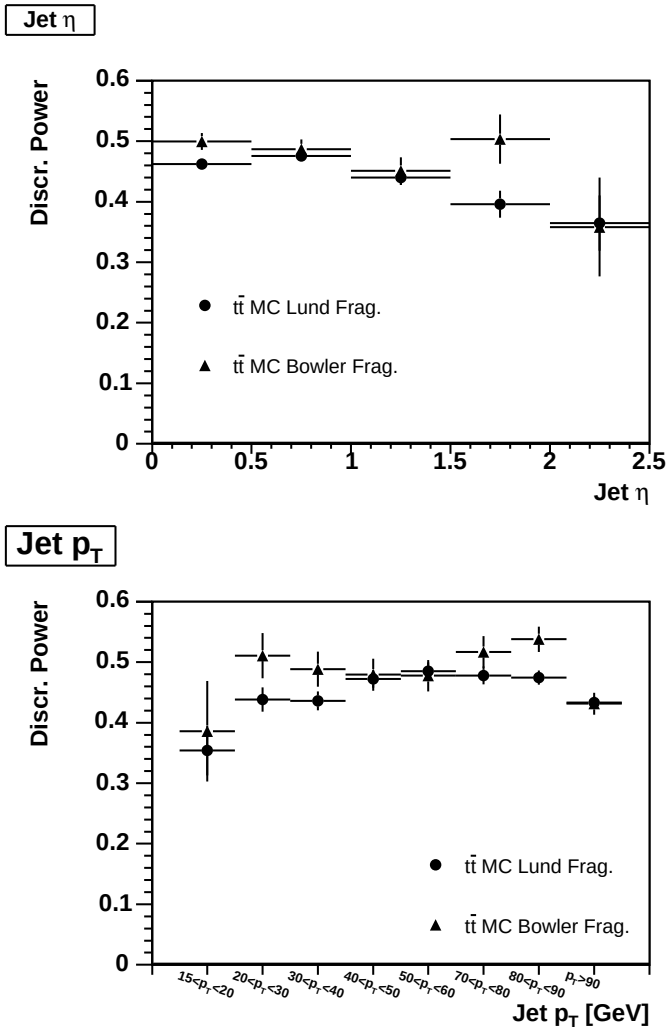

Figure 4.24. A comparison of the discriminating power in simulated $t\bar{t}$ events using the Lund string and the Bowler fragmentation model as functions of η (top) and p_T (bottom).

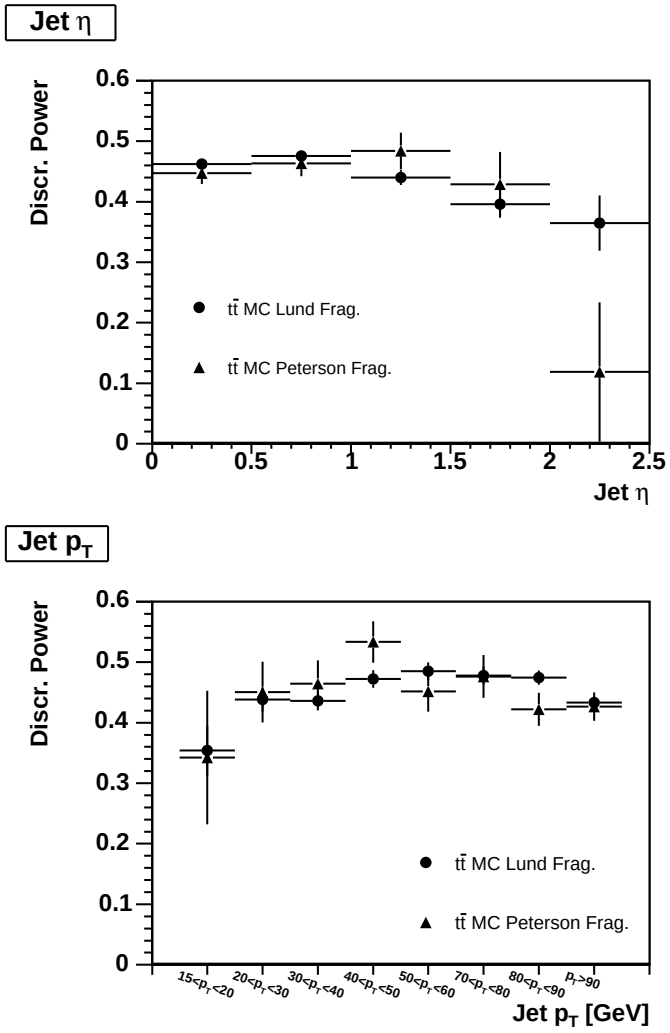

Figure 4.25. A comparison of the dicriminating power in simulated $t\bar{t}$ events for the Lund string and Peterson fragmentation model as functions of η (top) and p_T (bottom).

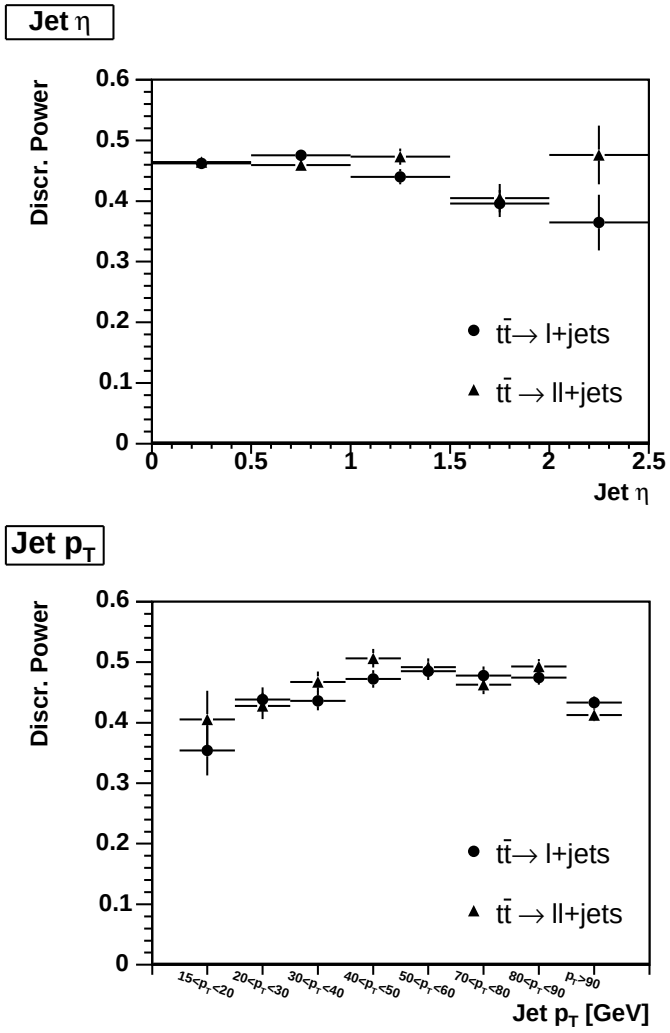

Figure 4.26. Comparison of the discriminating power between SVT-tagged b- and $\bar{b}$-quark jets in simulated $t\bar{t} \to \ell$+jets and $t\bar{t} \to \ell\ell$+jets events as functions of jet η (upper plot) and p_T (lower plot).

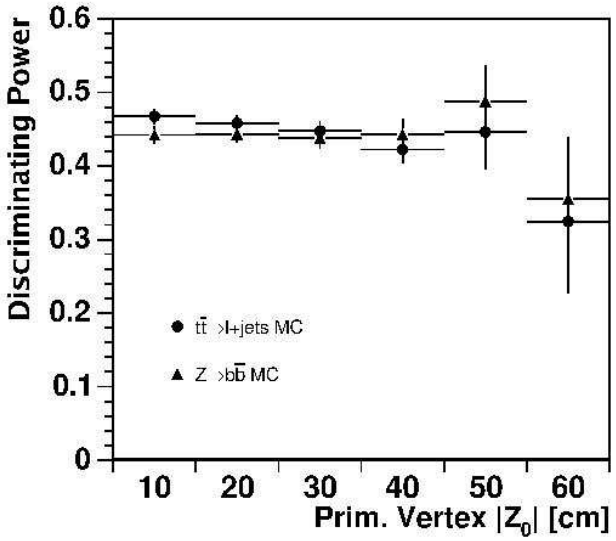

Figure 4.27. Discriminating power as a function of the PV z-position for SVT-tagged b- and $\bar{b}$-quark jets in simulated $t\bar{t} \to \ell$+jets and $Z \to b\bar{b}$ events.

sometimes a c-quark jet from the W boson decay. In addition the rate at which the various types of jets are SVT-tagged are different in data and simulated events. For this analysis, only the relative fraction of SVT-tagged jets of different flavors are important and not the absolute efficiencies.

To make sure that the procedure to find the flavor of the SVT-tagged jets in simulated $t\bar{t}$ events is correct, it is repeated but relying on the tagging efficiencies derived from data to predict which of the jets in the event that are SVT-tagged. The expected SM top and exotic quark charge templates were re-derived using this alternative procedure and the change was negligible.

Dependence on the Primary Vertex Position

The tracking efficiency and other tracking related variables depends on the geometry of the tracking detectors. Since samples with different PV distributions are used to extract the jet charge templates for jets in the signal sample, the discriminating power for b- and $\bar{b}$-quark jets was calculated with respect to the PV position. No significant dependence was found (see Fig. 4.27) which may be related to the fact that only SVT-tagged jets are considered. SVT-tagged jets are by definition associated with a set of well defined tracks (see Sec. 3.6.4).

Sample	SVT- & μ-tagged jets (%)
Dijet data	3.0 ± 0.2
$t\bar{t}$ MC	8.6 ± 0.2
Signal sample	11.9 ± 5.0
$Z \to b\bar{b}$ MC	3.2 ± 0.2

Table 4.11. Fraction of SVT-tagged jets that have a muon associated to it in various samples. The difference between the dijet data and simulated $t\bar{t}$ events (passing signal selections) is taken into account as a systematic uncertainty. For comparison, the fractions in the signal sample and simulated $Z \to b\bar{b}$ events are shown.

Fraction of μ-Tagged Jets in the Dijet and Signal Sample

The jet charge algorithm considers all tracks satisfying the standard quality cuts in Sec. 4.3.1. All jets are required to be tagged by the SVT algorithm but no veto is applied to reject μ-tagged jets. The muon track usually has significantly higher transverse momentum than other tracks in the same jet, and the jet charge distribution for b-quark jets containing a muon is significantly different. The jet charge templates are derived on a different sample than the signal sample and therefore a possible bias may exists due to different fraction of μ-tagged jets in the samples. Table 4.11 shows the fraction of the SVT-tagged jets that also has a muon associated to it. A systematic uncertainty due to this discrepancy is obtained by re-deriving the jet charge templates for b- and c-quark jets with a veto on any *probe*-jet that has a muon matched to it within a cone of $\Delta R < 0.5$.

Fraction of c-Quark Jets in the Dijet Samples

The uncertainty on the fraction of c-quark jets in the dijet samples originates from the uncertainty of the $p_{T,\mathrm{rel}}$ fit. The b- and c-quark jet charge distributions are re-derived varying the fraction of c-quark jets according to this uncertainty.

Fraction of Muon Charge Sign Change in the Dijet and Monte Carlo Samples

The amount of cascade decay passing the loose and tight dijet selections depends on the jet p_T. The simulated $Z \to b\bar{b}$ events and the dijet samples have similar but not equal jet η and p_T spectrum. Therefore, there is still the possibility that the fraction of cascade muons is different in the dijet samples and simulated $Z \to b\bar{b}$ events, that are used to extract the fraction of times the muon charge sign is different from the quark initiating the jet. This uncertainty is taken into account by re-deriving the jet charge templates by varying the fraction of times the muon charge changes sign, x_{flip} (calculated in Sec. 4.4.4), according to the uncertainty on the weighted average.

The measured charge of the muon can be different from the quark that initiated the jet simply because the muon track was poorly reconstructed. This effect exists

also in the simulated events and we observe that it affects about 1% of the muons inside jets in simulated $Z \to b\bar{b}$ events. The rate at which the muon charge is measured incorrectly might be different in data and simulation. This potential difference between data and simulation is taken into account by varying how often the charge is misidentified between 0% and 3%.

Statistical Uncertainty on the Kinematic Weighting

The procedure to weight the $t\bar{t}$ Monte Carlo events to reproduce the jet p_T and η spectrum of the dijet samples has a number of uncertainties associated with it. By using the $\pm 1\sigma$ band in the fits, the statistical uncertainty can be evaluated. There are two weights that are applied, the jet p_T and η weights, both with a statistical uncertainty associated with them. Conservatively, the uncertainties are added in quadrature and the jet charge templates are re-derived varying this uncertainty.

Statistical Uncertainty on the Kinematic Correction

There is an additional systematic uncertainty related to the kinematic correction which arises from the limited statistics used in the correction. This systematic is evaluated by using the $\pm 1\sigma$ bands on the kinematic correction fit shown in Fig. 4.18 and re-deriving the jet charge templates using these varied correction functions.

Statistical Uncertainty on the Jet Charge Templates

The extraction of the jet charge template in Sec. 4.4.2 leads to a statistical uncertainty on the jet charge distributions. This uncertainty is evaluated by fitting the jet charge distributions with Gaussian functions and varying the fitted parameters within their uncertainties to produce $\pm 1\sigma$ templates. The fit for b- and c-quark jet charge distributions are shown in Fig. 4.28 and 4.29 respectively. The validity of this procedure was estimated by allowing each bin of the jet charge distributions to vary according to a Gaussian centered on the bin content and a width according to the error of that bin. After varying all bins in the jet charge templates the discriminating power is calculated and put in a histogram. This procedure is repeated $10,000$ times and the resulting distribution is fitted with a Gaussian where the width is compared to the difference in discriminating power between the $\pm 1\sigma$ varied jet charge distributions. The varied jet charge templates have a discriminating power that agrees with the width of the Gaussian and the statistical uncertainty on the templates are thus reasonable.

Top Quark Mass Uncertainty

The constrained kinematic fit uses the top quark mass as an additional constraint at a value of 175 GeV. This value is chosen to be consistent with the mass used in the generation of the simulated events. Top quark mass is known only to a certain precision and constraining the mass in the fit to an alternative value might affect the fit performance. To take this into account, the SM top and exotic quark charge

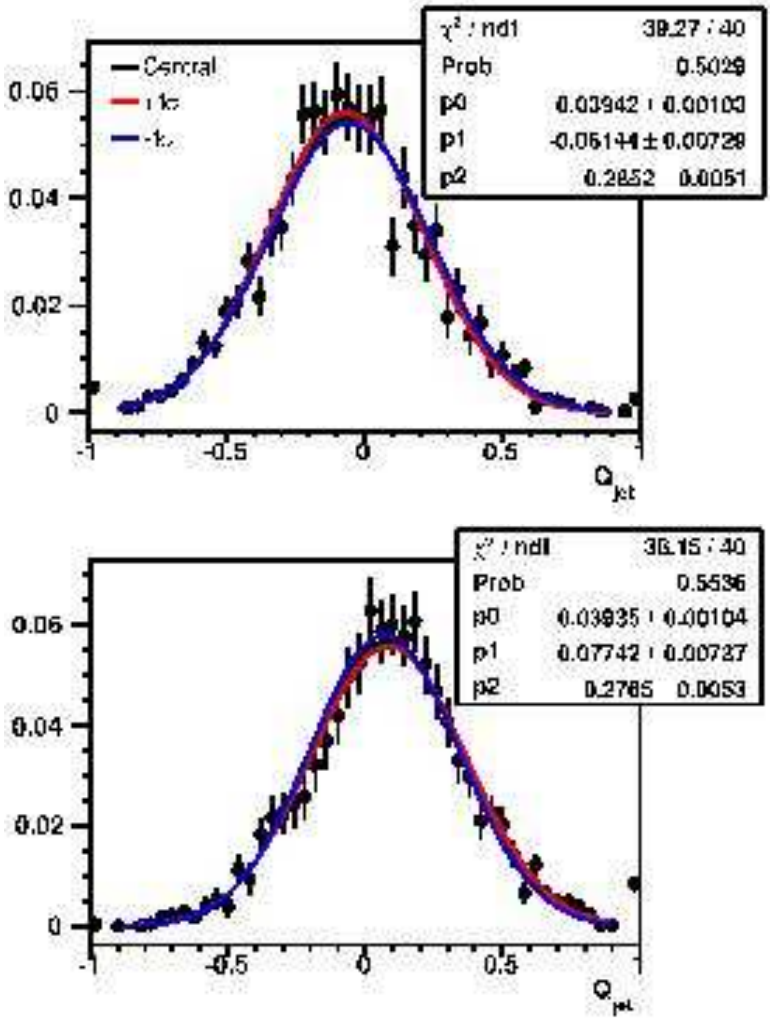

Figure 4.28. The b- and $\bar{b}$-quark jet charge distributions (top and bottom respectively) fitted with a Gaussian function. The $\pm 1\sigma$ jet charge templates are derived by varying the parameters of the Gaussian within their errors. The varied jet charge templates are calculated from the difference between the varied and central Gaussian.

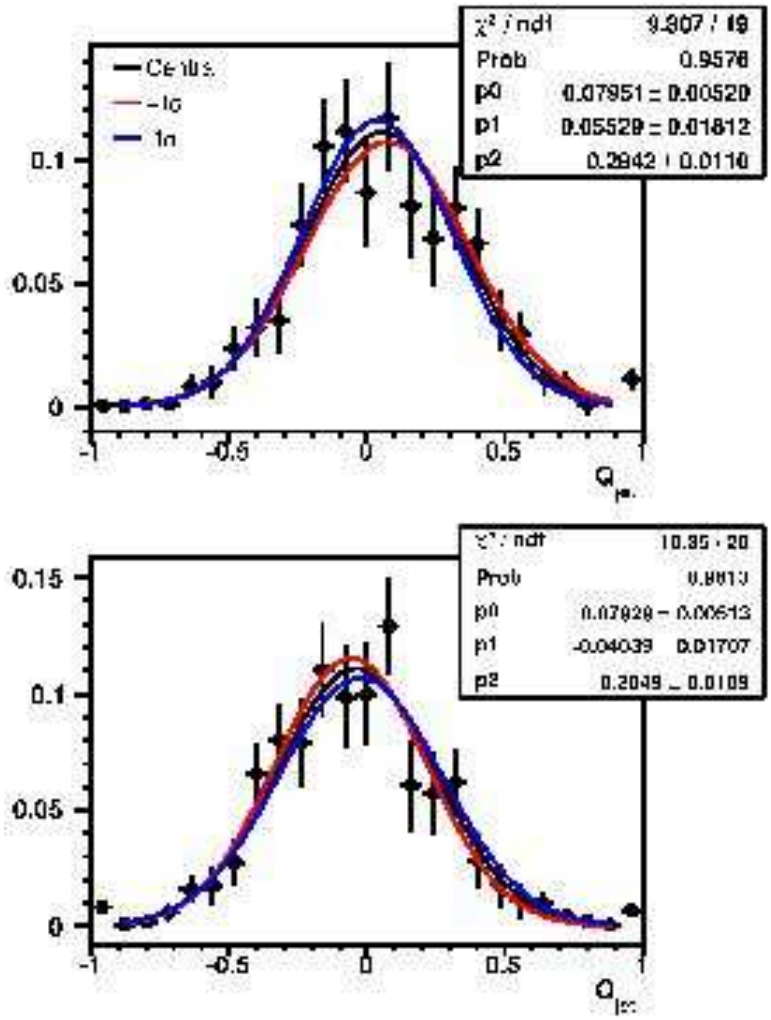

Figure 4.29. The c- and $\bar{c}$-quark jet charge distributions fitted with a Gaussian function (top and bottom respectively). The $\pm 1\sigma$ jet charge templates are derived by varying the parameters of the Gaussian within their errors. The varied jet charge templates are calculated from the difference of the varied to the central Gaussian.

templates are re-derived using simulated $t\bar{t}$ events generated with different masses, while keeping the same top mass constraint in the fit. Events generated with a top quark mass of 170 and 180 GeV are used to evaluate this systematic uncertainty. The current world average uncertainty in the top mass is $\approx$ 2.3 GeV [15] and therefore the SM top and exotic quark charge templates obtained from the 170 and 180 GeV samples are scaled corresponding to this uncertainty.

Jet Energy Scale

The jet energy scale correction is an attempt to correct the measured jet energies in the calorimeter back to the stable-particle level before interacting with the detector as described in Sec. 3.6.2 and is applied to all jets. The jet energy scale correction is different for jets in data and in simulated events as shown in Fig. 3.5 and 3.6. The uncertainty from the jet energy scale is found by adding the uncertainties from data and simulation in quadrature, conservatively treating them totally uncorrelated. This uncertainty changes the jet energies in the event and may consequently affect the kinematic fit. The SM top and exotic quark charge templates are re-derived using the varied jet energy scale correction.

Jet Energy Resolution

The observed p_T spectrum of jets in simulated events is smeared to match the resolution measured in data. The uncertainties on the parametrization of the jet energy introduces an uncertainty not taken into account by the jet energy scale. To account for this, the parametrization is varied according to the uncertainty and the SM top and exotic quark charge templates are re-derived.

Jet Reconstruction and Identification Efficiency

The efficiency to reconstruct a jet subject to all requirements in Sec. 3.6.1 is higher in simulated events than in data [41]. The discrepancy is most prominent in the low p_T region ($15 - 25$ GeV). Therefore, a p_T dependent data-to-simulation scale factor is derived and jets in simulated events are removed to reproduce the jet reconstruction efficiency observed in data. This scale factor is varied according to its uncertainty to take into account any effect on the kinematic fit and the SM top and exotic quark charge templates are re-derived.

Composition of the Signal Sample

The signal sample composition discussed in Sec. 4.2 has some uncertainty. Due to the large signal-to-background ratio this uncertainty has a small net effect on the analysis. It is evaluated by taking into account the statistical and systematic uncertainty when extracting the final result. A more detailed description is given in Sec 4.7.

$b\bar{b}$ Production Mechanism

The dominant $b\bar{b}$ production mechanism in the dijet samples is assumed to be flavor creation. Other processes contributing to the production of $b\bar{b}$ pairs (flavor excitation and gluon splitting) destroy the correlation between the sign of the tagging muon and the quark initiating the *probe*-jet. The fraction of flavor creation is sensitive to the azimuthal distance between the jets. Therefore, the jet charge templates are re-derived with the requirement of $\Delta\phi > 2.65$ instead of $\Delta\phi > 3.0$ on the azimuthal distance between the two jets in the dijet samples to take this uncertainty into account.

$t\bar{t}$ Signal Modeling

In the kinematic fit it is assumed that the four highest p_T jets in the event are the result of the hadronization of the partons from the $t\bar{t}$ decay. As discussed before, additional jets can arise from ISR and FSR. When $t\bar{t}$ events are produced in association with a jet, the additional jet can be misinterpreted as a decay product from the $t\bar{t}$ pair. To assess the uncertainty in the modeling of these effects, events has been generated using a dedicated simulation of the production of $t\bar{t}$ events together with an additional parton using ALPGEN. The fraction of such events is estimated not to be larger than 30% [80]. The SM top and exotic quark charge templates are re-derived using this sample and systematic uncertainty of 30% of the difference between this and the default simulated $t\bar{t}$ sample is quoted.

To estimate the effect of the individual sources of systematic uncertainties, the discriminating power between the varied SM top and exotic quark charge templates is calculated and shown in Tab. 4.12.

Systematic uncertainty	Variation
Fraction of c-quark jets	± 0.003
Fraction of muon charge sign change	± 0.007
Stat. uncert. on the kinematic weighting	± 0.003
Stat. uncert. on the kinematic correction	± 0.014
$b\bar{b}$ production mechanism	± 0.008
Stat. uncert. b-quark jet charge templates	± 0.023
Stat. uncert. c-quark jet charge templates	± 0.002
Fraction of μ-tagged jets	± 0.005
$t\bar{t}$ signal Modeling	± 0.007
Jet energy scale	± 0.003
Jet energy resolution	± 0.012
Jet reconstruction efficiency	± 0.009
Top quark mass uncertainty	± 0.002

Table 4.12. The maximum shift in discriminating power between the varied SM top and exotic quark charge templates for the individual sources of systematic uncertainty.

4.7 Results

The underlying models in this analysis are the two hypotheses of a SM top quark with a charge of $2e/3$ or an exotic quark with charge $4e/3$ referred to in the following as H_{sm} and H_{ex} respectively. A hypothesis test is performed in Sec. 4.7.1 to assess the validity of the different hypotheses using a likelihood ratio test. However, the hypothesis test only provides the confidence in rejecting any of the two discrete scenarios, it does not assess the probability of a mixture of quarks with charge $2e/3$ and $4e/3$ in the signal sample. Therefore, in Sec. 4.7.2 the fraction of quarks with the exotic charge $4e/3$ is estimated using the method of maximum likelihood.

4.7.1 Discrimination Between Charge 2e/3 Top Quark and Charge 4e/3 Exotic Quark Production Scenarios

In data a certain set of charges Q_1 and Q_2 are observed in the leptonic and hadronic leg of the ℓ+jets event respectively. There are 21 selected events in the signal sample. In 16 events, the kinematic fit converges in at least one of the permutations where the two jets j_b and j_B are associated to SVT-tagged jets. Each of the 16 events provides two observations of the charge, giving 16 observations of Q_1 and 16 observations of Q_2, i.e. a total of 32 charge observations. Figure 4.30 shows the binned and unbinned distribution of measured charges in the data overlaid with the expected SM top and exotic quark charge templates. As can be seen from the figures, the observed data prefers the SM hypothesis. Below, a procedure to quantify the discrimination between the two hypotheses is described.

Likelihood Ratio

Often when discriminating between two scenarios it is useful to define a test statistics. If the set of 32 measured values Q_1 and Q_2 are denoted by a vector $\mathbf{q}^{\mathrm{data}}$ then the test statistics, denoted Λ, can be a single number or a vector with fewer components than $\mathbf{q}^{\mathrm{data}}$. Λ is a function of the data and its value reflects the level of agreement between the data and the hypothesis. Under assumption of each hypothesis different Λ will be obtained for repeated experiments.

The goodness is quantified by the P-value, which is the probability to find Λ in the region of equal or lesser compatibility assuming a hypothesis H than the level of compatibility observed with the real data Λ^{data}. In this analysis, the test statistics Λ is the ratio of the likelihoods $\mathcal{L}$ for the two different hypotheses H_{sm} and H_{ex} given a vector of data $\mathbf{q}$

$$\Lambda(\mathbf{q}) = \frac{\mathcal{L}(\mathbf{q}|H_{\mathrm{sm}})}{\mathcal{L}(\mathbf{q}|H_{\mathrm{ex}})}, \tag{4.16}$$

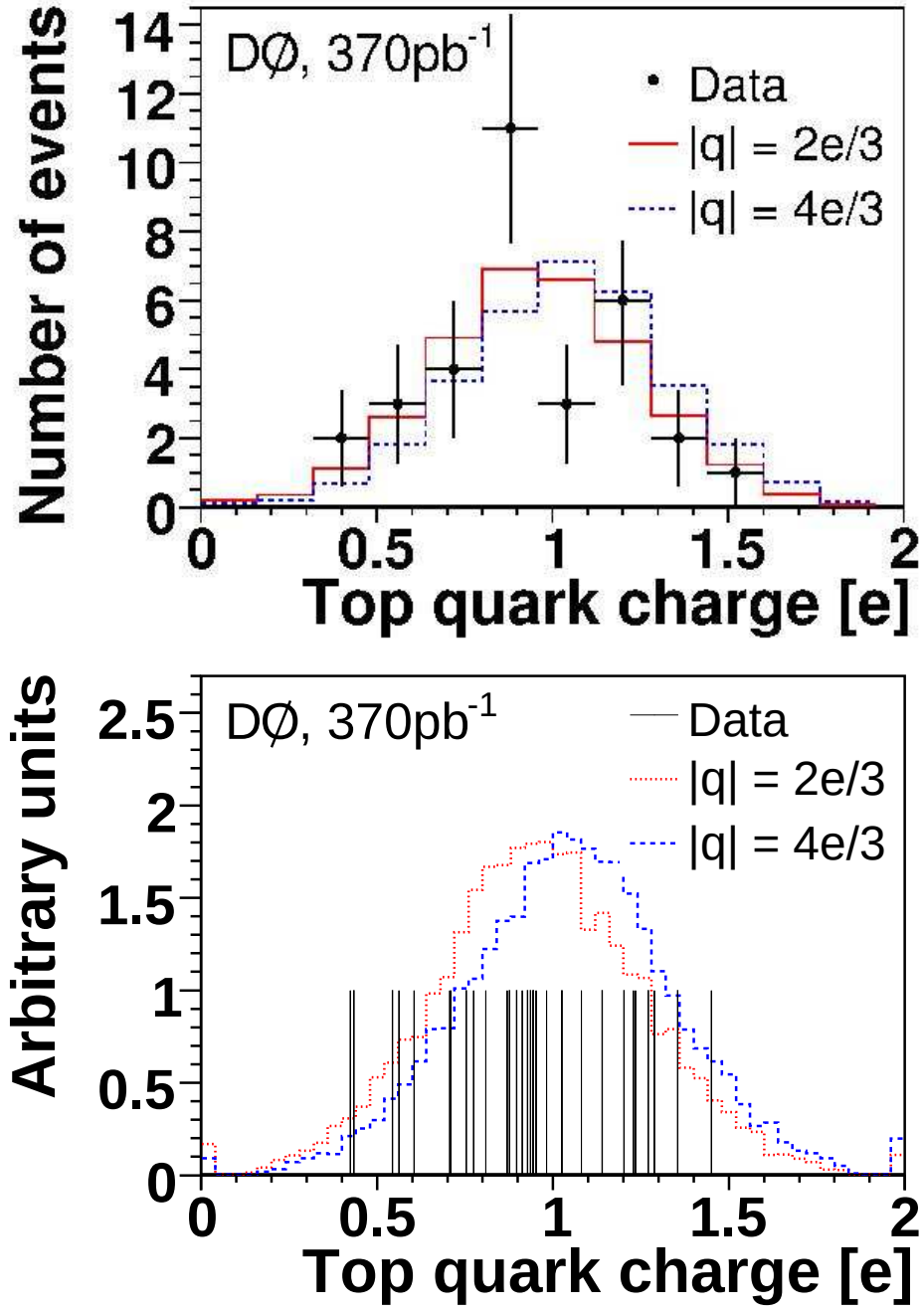

Figure 4.30. The 32 observed charges compared to the SM top and exotic quark charge templates for binned (top) and unbinned (bottom) data.

where the likelihood functions are defined as

$$\mathcal{L}(\mathbf{q}|H_{\mathrm{sm}}) = \prod_i p_{\mathrm{sm}}(q_i) \tag{4.17}$$

$$\mathcal{L}(\mathbf{q}|H_{\mathrm{ex}}) = \prod_i p_{\mathrm{ex}}(q_i). \tag{4.18}$$

Here the probability density functions (p.d.f.s) p_{sm} and p_{ex} are the SM top and exotic quark charge templates derived in Sec. 4.5.2 and the subscript i runs from $1, 2, .., 32$.

Under the assumption that the distribution of $\mathbf{q}$ follows a hypothesis H, the p.d.f. $g(\Lambda|H)$ will be determined for the test statistics. Here, the two hypotheses gives:

$$\Lambda^{\mathrm{sm}} \equiv g(\Lambda|H_{\mathrm{sm}}) \tag{4.19}$$

$$\Lambda^{\mathrm{ex}} \equiv g(\Lambda|H_{\mathrm{ex}}). \tag{4.20}$$

Given the definition of the test statistic, the P-value for the exotic scenario is

$$P = \int_{\Lambda^{\mathrm{data}}}^{\infty} \Lambda^{\mathrm{ex}}(\Lambda) d\Lambda. \tag{4.21}$$

The likelihood ratio obtained from Eq. 4.16 with the set $\mathbf{q}^{\mathrm{data}}$ of 32 observed charges in the signal sample is,

$$\Lambda^{\mathrm{data}} = 4.27$$

Generation of Pseudo-experiments

The P-value is not the probability of the hypothesis but rather the probability under the assumption of the hypothesis of obtaining data as incompatible with the hypothesis as the data observed. Thus, in order to calculate the P-value for the exotic scenario defined in Eq. 4.21, the p.d.f. Λ^{ex} has to be computed.

This can be determined by the generation of pseudo-experiments which in an idealized world would correspond to building a set of new experiments and measure the charge under the assumption of H_{ex}. The procedure is as follows:

1. *Signal and background fraction*
   The signal and background fractions are allowed to fluctuate according to the statistical and systematic uncertainties: The number of background events N_{bkg} is obtained from a Binomial distribution with a mean at the prediction given by a Gaussian distribution with a mean of 1.53 (the central prediction) and a standard deviation equal to the total uncertainty on the number of background events (see Fig. 4.31).

2. *Systematic uncertainties*
   The various sources of systematic uncertainties affect the template distributions and have to be taken into account. The method used here allows the

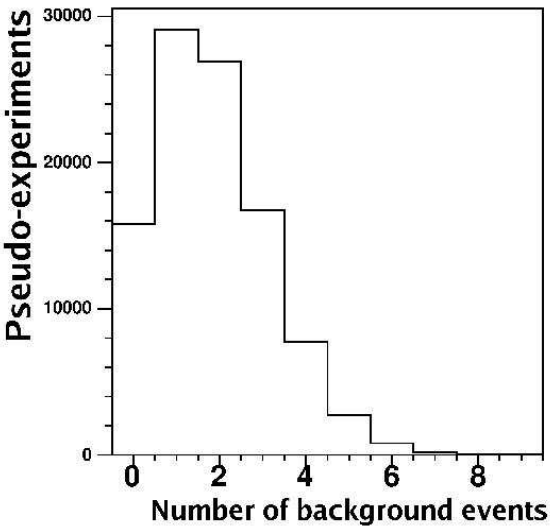

Figure 4.31. The distribution of the number of background events for each pseudo-experiment.

systematic uncertainties to change the default (or best guess) template. All systematic uncertainties j are assumed to follow Gaussian distributions with widths σ_j. Each systematic uncertainty is modeled by a free parameter ν_j called *nuisance parameter* [81]. If the default SM top quark template is characterized by a function $p_{\rm sm}(Q)$ and if $p^{j+}_{\rm sm}(Q)$ and $p^{j-}_{\rm sm}(Q)$ are the $\pm 1\sigma$ varied templates for the source of error j, then the SM top quark template (and similar for the exotic quark charge template) as a function of all systematic uncertainties can be written as:

$$p_{\rm sm}(Q,\nu_j) = p_{\rm sm}(Q) + \sum_j \nu_j (p^{j\pm}_{\rm sm}(Q) - p_{\rm sm}(Q)), \tag{4.22}$$

$$p_{\rm ex}(Q,\nu_j) = p_{\rm ex}(Q) + \sum_j \nu_j (p^{j\pm}_{\rm ex}(Q) - p_{\rm ex}(Q)), \tag{4.23}$$

and similar for the exotic template. ν_j follow Gaussian distributions with mean zero and standard deviation one.

3. *Pseudo-measurements*
   The number of pseudo-measurements has to be equal to the total number of observed charges and thus $p_{\rm ex}(Q)$ is randomly sampled $32 - N_{bkg}$ times and the background template (given in Sec. 4.5.3) N_{bkg} times to obtain a set of pseudo-measurements $\mathbf{q}^{\rm pseudo}$.

4. *Likelihood ratio*
   Given the set of pseudo-measurements $\mathbf{q}^{\rm pseudo}$, the likelihood ratio is calculated using the generated $p_{\rm sm}(Q,\nu_j)$ and $p_{\rm ex}(Q,\nu_j)$ to take into account the systematic uncertainties.

By repeating steps 1-4 $100,000$ times (each time with a new background fraction and new set of nuisance parameters ν), $\Lambda^{\rm ex}$ can be determined. In a similar way, the p.d.f. $\Lambda^{\rm sm}$ can be determined by sampling the $p_{\rm sm}(Q)$ charge template in step 2. Both $\Lambda^{\rm ex}$ and $\Lambda^{\rm sm}$ are shown in Fig. 4.32 together with the observed $\Lambda^{\rm data}$ in data.

Confidence Level

From Fig. 4.32 it is clear that the observed set of charges is more SM like than exotic like. The calculation of what confidence level α the observed charges excludes the hypothesis of 100% quarks with the exotic $4e/3$ charge in the signal sample is simply done by observing that

$$1 - \alpha = P. \tag{4.24}$$

The confidence level is therefore extracted by calculating the P-value as defined in Eq. 4.21

$$\int_{\Lambda^{\rm data}}^{\infty} \Lambda^{\rm ex}(\Lambda) d\Lambda = 0.078, \tag{4.25}$$

and the corresponding confidence level is $\alpha = 92.2\%$.

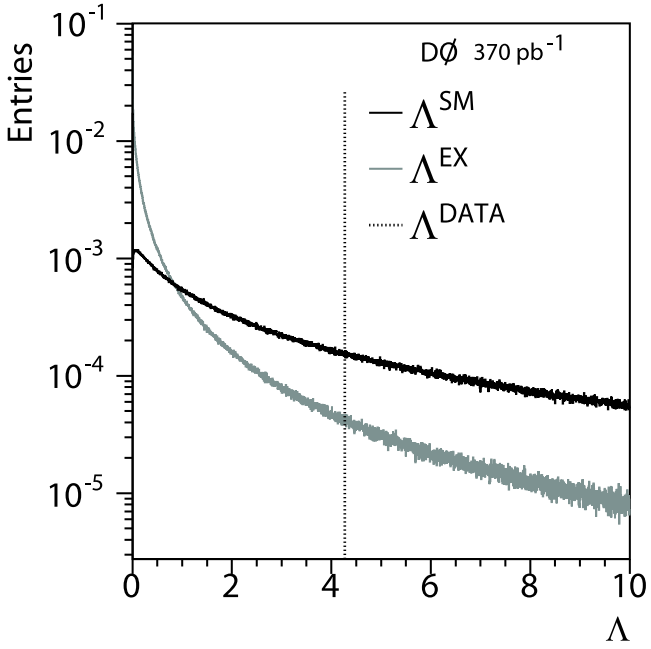

Figure 4.32. The distribution of $\Lambda^{\rm sm}$ and $\Lambda^{\rm ex}$ resulting from the pseudo-experiments together with observed value of $\Lambda^{\rm data}$.

Systematic uncertainty	α	α^{expected}	$\alpha^{\text{expected}}_{\text{single}}$
Stat. uncert.	95.82	95.25	95.25
Fraction of c-quark jets	95.75	95.23	95.24
Fraction of muon charge sign changes	95.66	95.21	95.26
Stat. uncert. on the kinematic weighting	95.56	95.21	95.25
Stat. uncert. on the kinematic correction	94.37	94.11	94.35
$b\bar{b}$ production mechanism	93.70	93.35	94.55
Stat. uncert. b-quark jet charge templates	93.47	93.22	95.18
Stat. uncert. c-quark jet charge templates	93.33	93.12	95.23
Fraction of μ-tagged jets	93.28	93.06	95.19
$t\bar{t}$ signal modeling	93.11	92.89	95.12
Jet energy scale	93.08	92.62	95.15
Jet energy resolution	92.82	92.26	95.16
Jet reconstruction efficiency	92.44	91.79	95.17
Top quark mass uncertainty	92.22	91.17	95.17

Table 4.13. Breakdown of the systematic uncertainties and their cumulative effect (α and α^{expected}) on the confidence level. $\alpha^{\text{expected}}_{\text{single}}$ is the effect on the confidence level allowing only one source of systematic uncertainty to enter the calculation.

The probability that the exotic hypothesis gives a set of charges as incompatible with the observed median of Λ^{sm} is also calculated. This can be thought of as an expected confidence level and is given by:

$$1 - \alpha^{\text{expected}} = \int_{\Lambda^{\text{sm}}_{\text{Median}}}^{\infty} \Lambda^{\text{ex}}(\Lambda) d\Lambda = 0.088 \tag{4.26}$$

which gives the expected confidence level $\alpha^{\text{expected}} = 91.2\%$ in good agreement with the observed 92.2%.

The effect of the systematic uncertainties on the confidence level can be seen in Tab. 4.13 where the cumulative and individual effect of including systematic uncertainties are shown.

4.7.2 Fraction of Charge $2e/3$ Top Quark and Charge $4e/3$ Exotic Quarks

The exclusion at 92% confidence level that the set of data $\mathbf{q}^{\text{data}}$ arose from the decay of a quark with the exotic $4e/3$ electric charge does not address the possibility that the set of data results from a mixture of a $2e/3$ and a $4e/3$ charge quark. In this section the fraction of exotic $4e/3$ charge quarks in the signal sample is estimated.

Maximum Likelihood Fit

To extract as estimate of the fraction of exotic quarks ρ the method of maximum likelihood is used. If $p_{\text{mix}}(\mathbf{q}^{\text{data}}; \rho)$ is the p.d.f. from where the measurements $\mathbf{q}^{\text{data}}$

arises, the estimated fraction of exotic quarks is the value of ρ that maximizes the likelihood function,

$$\mathcal{L}(\rho) = \prod_{i}^{32} p_{\mathrm{mix}}(q_i^{\mathrm{data}}|\rho). \tag{4.27}$$

The p.d.f.s $p_{\mathrm{mix}}(\mathbf{q}^{\mathrm{data}}|\rho)$ for a specific fraction of exotic quarks ρ is found by mixing the pure SM top quark charge template $p^{sm}(q)$ and the pure exotic quark charge template $p^{ex}(q)$ with the appropriate fraction.

$$p_{\mathrm{mix}}(q|\rho) = \rho \times p^{\mathrm{ex}}(q) + (1-\rho) \times p^{\mathrm{sm}}(q). \tag{4.28}$$

For each value of ρ (taken to be $0, 0.05, ..., 1.0$) the unbinned likelihood that the data $\mathbf{q}^{\mathrm{data}}$ is consistent with the sum of signal and background is computed as described in Eq. 4.27. The $-\log\mathcal{L}$ is minimized for the 32 observed charges in data and the result is a set of $-\log\mathcal{L}$ points versus ρ. The result can be fitted with a parabola to find the maximum likelihood estimator of ρ.

Interpretation of the Result to a Confidence Interval

For Gaussian errors, $-\log\mathcal{L}$ has a parabolic shape and the confidence region (the probability that this region contains the true value of ρ) is given by,

$$-\ln\mathcal{L}(\rho) \leq -\ln\mathcal{L}_{\mathrm{max}} + \Delta\ln\mathcal{L}, \tag{4.29}$$

and the 68% (90%) standard deviation confidence intervals are given by $\Delta\mathcal{L} = 1/2(2.71/2)$ [5].

The result is also reported using a Bayesian approach where the prior p.d.f. $\pi(\rho)$ characterizes the prior knowledge of the true value of ρ. Using Bayes theorem,

$$f(\rho|\mathbf{q}^{\mathrm{data}}) = \frac{\mathcal{L}(\mathbf{q}^{\mathrm{data}}|\rho)\pi(\rho)}{\int_{-\infty}^{\infty}\mathcal{L}(\mathbf{q}^{\mathrm{data}}|\rho')\pi(\rho')d\rho'}, \tag{4.30}$$

the Bayesian confidence interval can be determined by computing the interval $[\rho_{\mathrm{min}}, \rho_{\mathrm{max}}]$ which contain a given fraction $1-\alpha$ of the probability,

$$1-\alpha = \int_{\rho_{\mathrm{min}}}^{\rho_{\mathrm{max}}} f(\rho, |\mathbf{q}^{\mathrm{data}})d\mathbf{q}^{\mathrm{data}}. \tag{4.31}$$

In this analysis, a flat prior p.d.f. according to the physical region is used i.e. it is one in the interval $[0,1]$ and zero elsewhere. Using this, the calculation

of the Bayesian interval is given by finding the interval $[\rho_{\min}, \rho_{\max}]$ such that the probability of $1-\alpha$ is,

$$\frac{\int_{\rho_{ML}}^{\rho_{\max}} \mathcal{L}(\mathbf{q}^{\text{data}}|\rho)\ d\rho}{\int_0^1 \mathcal{L}(\mathbf{q}^{\text{data}}|\rho)\ d\rho} = \frac{\int_{\rho_{\min}}^{\rho_{ML}} \mathcal{L}(\mathbf{q}^{\text{data}}|\rho)\ d\rho}{\int_0^1 \mathcal{L}(\mathbf{q}^{\text{data}}|\rho)\ d\rho} = \frac{(1-\alpha)}{2}, \tag{4.32}$$

where ρ_{ML} is the fraction at the minimum of the parabola (i.e. the fitted fraction). If ρ_{ML} is outside the physical region a one-sided confidence interval is reported as,

$$\frac{\int_0^{\rho_{\max}} \mathcal{L}(\mathbf{q}^{\text{data}}|\rho)\ d\rho}{\int_0^1 \mathcal{L}(\mathbf{q}^{\text{data}}|\rho)\ d\rho} = 1-\alpha, \tag{4.33}$$

if $\rho_{ML} < 0$ or,

$$\frac{\int_{\rho_{\min}}^1 \mathcal{L}(\mathbf{q}^{\text{data}}|\rho)\ d\rho}{\int_0^1 \mathcal{L}(\mathbf{q}^{\text{data}}|\rho)\ d\rho} = 1-\alpha. \tag{4.34}$$

if $\rho_{ML} > 1$.

Validation of the Fit Procedure

Before applying the maximum likelihood fit to the selected signal sample the fit procedure is validated by performing pseudo-experiments where the input fraction of exotic quarks is known. Ideally, the fit returns the same fraction without any deviations together with information on the expected statistical sensitivity.

One pseudo-experiment is built by randomly sampling the p.d.f. $p_{\text{mix}}(q|\rho)$ for a specific exotic fraction ρ to obtain pseudo-data. In this procedure, the correct signal and background fractions are taken into account. A maximum likelihood fit is then performed with the same procedure as applied to the real data. By performing a large number of these pseudo-experiments for each input fraction $\rho = 0.0, 0.05, .., 1.0$ the performance of the fit can be evaluated. The fitted versus input fraction shown in Fig. 4.33 is consistent with offset zero and slope one. The pull and the width of the fitted pull functions are shown in Fig. 4.34.

For each pseudo-experiment, a confidence interval can be calculated. Figure 4.35 shows the expected confidence intervals (taken as the mean of the confidence intervals for all pseudo-experiments) versus input fraction for both methods described above. The Bayesian mean value is constrained by the prior assumption of $\rho \in [0, 1]$. The conclusion is that the expected width of the 68% confidence interval is approximately ± 0.6.

Systematic Uncertainties

Sources of systematic uncertainties affect the fitted fraction in two ways: they can either change the fraction of signal and background or they can change the templates $p_{\text{mix}}(q|\rho)$. The first is taken into account by allowing the predicted background fraction to vary according to its total uncertainty, constrained to the total

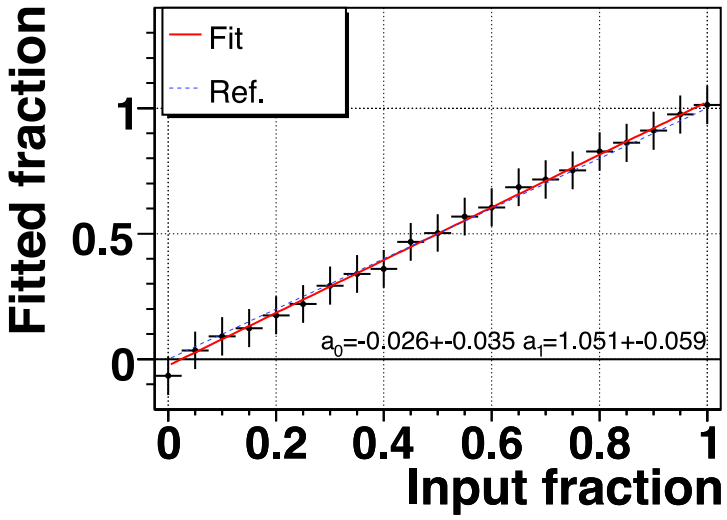

Figure 4.33. The fitted versus input fraction of exotic quarks in the pseudo-experiments.

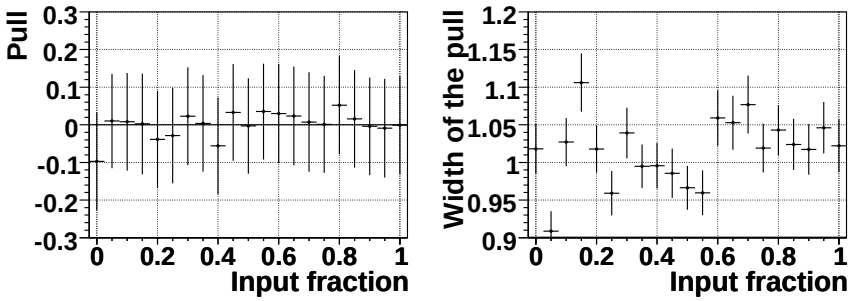

Figure 4.34. The pull (left) and the width of the fitted pull function (right) versus the input fraction of exotic quarks.

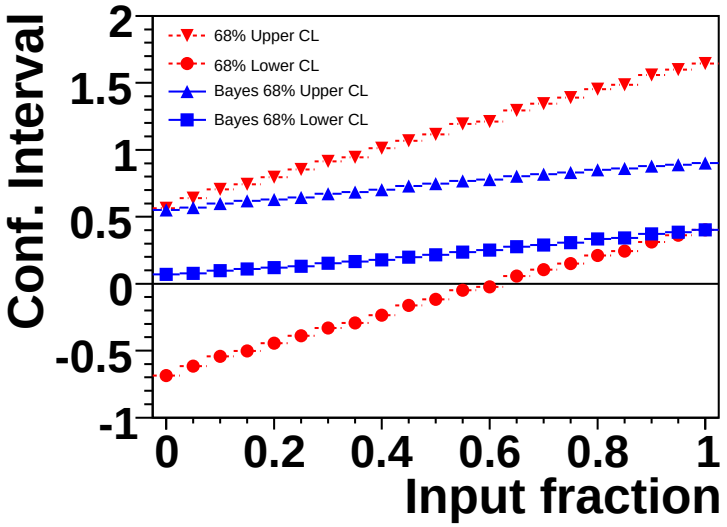

Figure 4.35. The average 68% confidence intervals for the normal (dashed) and Bayesian (solid) method. The red (blue) depicts the upper (lower) limit.

number of events observed in data. The effect of the other systematic uncertainties are estimated by performing pseudo-experiment using the default p.d.f. $p_{\text{mix}}(q|\rho)$ but with pseudo-data randomly sampled from the varied templates. The average observed shift in the fitted fraction for each systematic uncertainty (averaged over two input fractions) is added in quadrature to obtain the final shift from the systematic uncertainties. Table 4.14 shows the estimated systematic uncertainties.

Results from Data

Applying the maximum likelihood fit to the observed charges $\mathbf{q}^{\text{data}}$ results in the fit shown in Fig. 4.36. The maximum likelihood estimator that minimizes the $-\log\mathcal{L}$ function is $\rho = -0.13 \pm 0.66(\text{stat}) \pm 0.11(\text{syst})$, in good agreement with the SM expectation. The effect of including systematic uncertainties in the $-\log\mathcal{L}$ function is shown in Fig. 4.37. The systematic uncertainties are taken into account by convoluting the $-\log\mathcal{L}$ function with a Gaussian centered on the minimum $(-\log\mathcal{L}_{\text{max}})$ and a width equal to the total systematic uncertainty calculated above. The upper limit on the fraction of exotic quarks in the signal sample is 0.80 (0.52) at 90% (68%)C.L. The confidence intervals are shown in Tab. 4.15 with statistical uncertainty only and in Tab. 4.16 including systematic uncertainties.

Systematic uncertainty	$\Delta\rho$
Fraction of c-quark jets	± 0.023
Fraction of muon charge sign change	± 0.027
Stat. uncert. on the kinematic weighting	± 0.029
Stat. uncert. on the kinematic correction	± 0.027
$b\bar{b}$ production mechanism	± 0.045
Stat. uncert. b-quark jet charge templates	± 0.045
Stat. uncert. c-quark jet charge templates	± 0.022
Fraction of μ-tagged jets	± 0.031
$t\bar{t}$ signal Modeling	± 0.021
Jet energy scale	± 0.024
Jet energy resolution	± 0.035
Jet reconstruction efficiency	± 0.038
Top quark mass uncertainty	± 0.022
Total	± 0.111

Table 4.14. Summary of the observed shift in the fitted fraction for the systematic uncertainties.

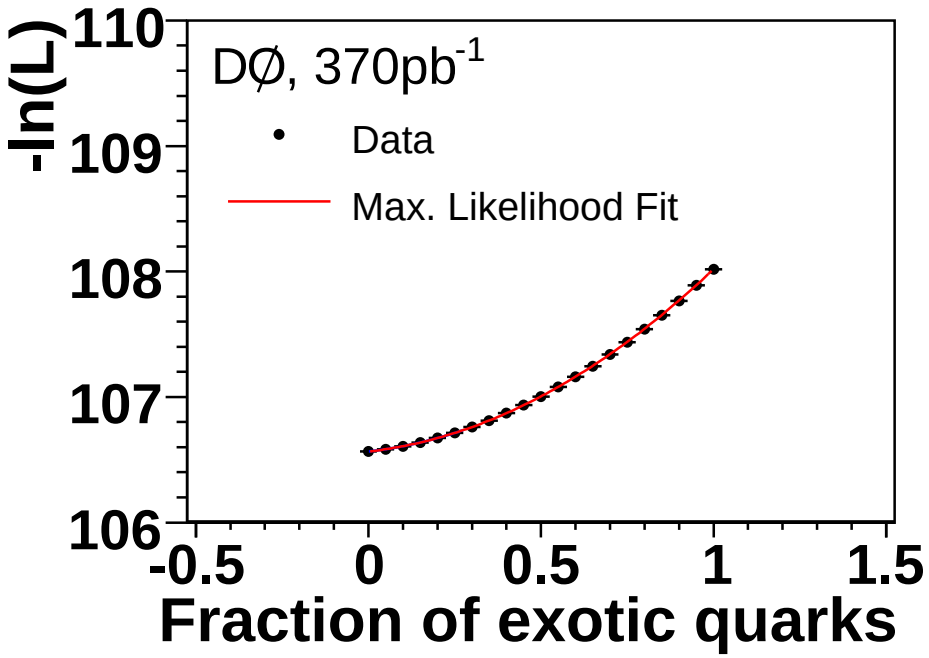

Figure 4.36. The maximum likelihood fit applied to the observed charges $\mathbf{q}^{\text{data}}$ in the signal sample.

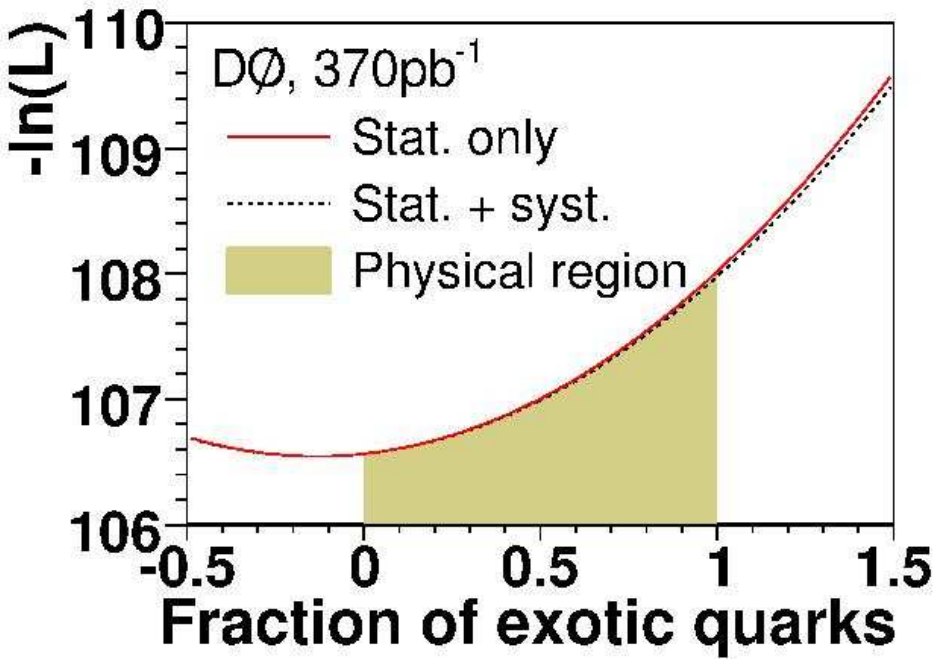

Figure 4.37. The fitted $-\log\mathcal{L}$ with and without systematic uncertainties.

Probability	ρ	ρ (Bayesian)
68%	$-0.79 < \rho < 0.53$	$0 < \rho < 0.52$
90%	$-1.22 < \rho < 0.95$	$0 < \rho < 0.79$

Table 4.15. Confidence intervals for statistical uncertainties only.

Probability	ρ	ρ (Bayesian)
68%	$-0.80 < \rho < 0.54$	$0 < \rho < 0.52$
90%	$-1.24 < \rho < 0.97$	$0 < \rho < 0.80$

Table 4.16. Confidence intervals including systematic uncertainties.

Chapter 5

Conclusion and Outlook

The first determination of the electric charge of the top quark is presented in this thesis. A data sample from $p\bar{p}$ collisions at $\sqrt{1.96}$ TeV recorded with the DØ detector at the Fermilab Tevatron collider corresponding to an integrated luminosity of 365 pb^{-1} was used. The top quark decays predominantly to bW^+ ($> 99.9\%$ [5]. The predicted electric charge of the top quark is $2e/3$ in the SM. The top quark electromagnetic coupling is however, not measured and since the correlation between the charge sign of the decay products, i.e. W^+ and b or W^- and $\bar{b}$, was not known an exotic quark with electric charge $-4e/3$ [3] was not excluded.

A pure sample of $t\bar{t}$ candidate events was selected in the ℓ+jets channel characterized by one high p_T charged lepton from the W boson decay and four jets out of which two are tagged using a b-tagging algorithm. A jet charge algorithm for b-quark jets based on the p_T weighted sum of charges of the tracks associated with the jet was developed to distinguish between b- and $\bar{b}$-quark jets. The algorithm was calibrated using dijet data dominated by $b\bar{b}$ events. Template distributions to discriminate between the SM top and exotic quark charge scenarios were obtained by reconstructing the charge of the decay products: The charge of the high p_T lepton and the b-tagged jets where the charge was extracted using the jet charge algorithm. A kinematic fit was performed to associate one of the b-quark jets to the charged lepton.

A likelihood ratio test to discriminate between the SM and the exotic charge hypothesis shows good agreement with the SM. The possibility that 100% of the selected events are exotic quarks with electric charge $4e/3$ is ruled out to 92% C.L. Nevertheless, the candidate heavy quarks could still be a mixture of two particles with the different charges. Using a Bayesian method to estimate the confidence intervals including both statistical and and systematic uncertainties, the fraction of exotic charges ρ in the selected sample is

$$0 < \rho < 0.52 \quad (68\% \text{ C.L.}),$$

$$0 < \rho < 0.80 \quad (90\% \text{ C.L.}).$$

The fraction of exotic quarks in the selected sample is consistent with zero, the SM prediction.

The recorded integrated luminosity at the Tevatron is presently above 1fb^{-1} which together with another year of data taking should allow for an exclusion of the exotic charge scenario close to 95%.

The Large Hadron Collider (LHC) at CERN is planned to start data taking during 2007. It will increase the luminosity with a factor of ~ 100 and the center-of-mass to 14 TeV. The top quark production cross-section is more than a factor of 100 larger and the top quark pair production rate will be approximately 8 million per year [82] allowing for a possible determination of the top quark electromagnetic coupling [83].

In general the top quark will play a major role in the physics program at LHC due to its the large mass, possibly playing a key role in electroweak symmetry breaking. The LHC is devoted to search for New Physics and any signature involving high p_T leptons, jets and $\not\!\!E_T$ are bound to have the top quark as a major background. In addition, the two jets from the hadronic decay of the W boson in a $t\bar{t} \to \ell$+jets event may be used to calibrate the jet energy scale by reconstructing the invariant mass of the W boson.

Acknowledgements

I begin with thanking my supervisor Bengt Lund-Jensen for his decision to give me this position at KTH. Although sometimes far away from the analysis work he has always made sure that I was taken care of, not only research- and financial-wise, but also in my everyday life as a "doktorand".

When I started at KTH I was sent to Fermilab to work in the top working group. Christophe Clément became my supervisor and teammate for this project and has guided me from square one. Thank you Christophe: for your brains, for teaching me physics, for explaining all the details, for your company during long nights at Fermilab, for the lunch excursions, for the smooth collaboration and for your friendship.

David Milstead has been of excellent help in my work and taking time to discuss and telling me how things really works. Thank you for getting me back on track when I was lost and for your cool english style. Thank you Sten Hellman for inspiring us to start this project.

I would like to thank the whole DØ collaboration for running the experiment. In the top working group Regina Demina, Aurelio Juste, Chris Tully, Erich Varnes and Martijn Mulders made numerous comments and suggestions concerning this analysis. Sergey Burdin helped with B-physics related issues. In particular, I would like to mention the excellent people in the Swedish Consortium who made me feel welcome from the start: Jonas Strandberg, thank you for helping me with DØ code and answering all my questions about b-tagging. Thank you for your attempts to help me understand baseball. Nils Gollub, thank you for the root-help, the many discussions and the fun times which hopefully continues in the coming years. Your proof-reading is much appreciated. Sara Strandberg always listened to my questions and helped me with everyday life at the lab. Thank you for lending me your car. I'm also very grateful to Barbro Åsman who not only helped me with my travel arrangements to Fermilab and the DØ administration but also

allowed me to drive her sports car (until Nils borrowed it and it broke down).

My good friends Paul DiTuro and Alejandro Daleo helped me to stay in shape at Fermilab and got me out from the lab occassionally. I will never forget the crazy nights in Chicago, not only (!) because of the "TVR"-scar. Christian Schwanenberger was the only one putting up with my ice hockey babble.

Thank you everybody in the particle physics group at KTH. Tore Ersmark, my roommate, has been my excellent computer support. Thank you Jens Lund for the laughs and (inappropriate) jokes and Sara Bergenius Gavler who kept organizing despite the workaholic crew. I would also like to thank Mark Pearce, Janina Östling, Jan Conrad, Erik Fredenberg, Karl-Johan Grahn, Silvio Orsi, Cecilia Marini Bettolo and Petter Hofvenberg for their nice company and lively discussions.

A big thanks goes to Christopher Engman who provided lodging and to Hanna Källman and Therese Nordlund who are always there for me. Svalan made sure I thought about other things than physics from time to time, thank you.

Without the full support and love of my parents, Kaj and Thory, this thesis would not have been possible. Thank you for raising me and allowing me to find my interests. I love you.

I also would like to thank my brother and best friend Björn for all the discussions, the good times, the trips and general "hets". You are the best.

Finally, without my wonderful Anna this thesis would have been a lot shorter. She has always supported me and accepted my absence many evenings and weekends. You mean the world to me!

Bibliography

[1] CDF Collaboration, F. Abe *et al.*, Phys. Rev. Lett. **74**, 2626 (1995); DØ Collaboration, S. Abachi *et al.*, Phys. Rev. Lett. **74**, 2632 (1995).

[2] P. C. Bhat, H. Prosper, and S. S. Snyder, Int. J. Mod. Phys. A **13**, 5113 (1998).

[3] D. Chang, W. FG. Chang, and E. Ma, Phys. Rev. D **59**, 091503 (1999); **61**, 037301 (2000); D. Choudhury, T. M. Tait, C. E. Wagner, Phys. Rev., D **65** (2002) 053002.

[4] CDF Collaboration, CDF Conference Note 8495 (2006).

[5] S. Eidelman *et al.*, Phys. Lett. B 592, 1 (2004); Particle Data Group web site: `http://pdg.lbl.gov`.

[6] G. Abbiendi *et al.*, Phys. Lett. B **565** 61-75 (2003).

[7] D. Griffiths, Introduction to Elementary Particles, Wiley, (1987).

[8] F. Wilczek, Int. J. Mod. Phys. **A13**, 863 (1997).

[9] LEP Collaboration, Phys. Lett. B **276** 247 (1992).

[10] G. Altarelli, M. Grunewald, Phys. Rep. 403-404 189-201 (2004).

[11] C. Cacciari *et al.*, JHEP 0404:068 (2004); N. Kidonakis, R. Vogt, Phys. Rev. D **68**, 114014 (2003).

[12] DØ Collaboration, V.M. Abazov *et al.*, Phys. Lett. B **622**:265-276 (2005); CDF Collaboration, D. Acosta *et al.*, Phys. Rev. D **71**, 012005 (2005).

[13] S. Cortese, R. Petronzio, Phys. Lett. B **253**, 494-498 (1991); D.A. Dicus, S.S.D. Willenbrock, Phys. Rev. D **34**, 155 (1986).

[14] I.I. Bigi *et al.*, Phys.Lett. **B** 181, 157 (1986).

[15] CDF and DØ Collaborations, hep-ex/0603039.

[16] Christopher T. Hill, Stephen J. Parke, Phys. Rev. D **49**, 4454-4462 (1994).

[17] DØ Collaboration, V.M. Abazov *et al.*, Top Physics Online Archive, available at: `http://www-d0.fnal.gov/Run2Physics/top/index.html`

[18] N. Cabibbo, Phys. Rev. Lett. **10**, 531-533 (1963); M. Kobayashi, T. Maskawa, Prog. Theor. Phys. **49**, 652-657 (1973).

[19] DØ Collaboration, V.M. Abazov *et al.*, Phys. Lett. **B** 639, 616 (2006).

[20] DØ Collaboration, V.M. Abazov *et al.*, hep-ex/0609045, submitted to Phys. Rev. Lett.(2006).

[21] CDF Collaboration, A. Abulencia *et al.*, Phys. Rev. **D** 73, 111103 (2006).

[22] DØ Collaboration, V.M. Abazov *et al.*, Phys. Lett. **B** 617, 1 (2005).

[23] C.T. Hill, Phys. Lett. **B266**, 419 (1991).

[24] C.T. Hill, Phys. Lett. **B345**, 483 (1995).

[25] CDF Collaboration, T. Affolder *et al.*, Phys. Rev. Lett. 85, 2062-2067 (2000); DØ Collaboration, V.M. Abazov *et al.*, Phys. Rev. Lett. 92, 221801 (2004); DØ Collaboration, V.M. Abazov *et al.*, DØ Conference Note 4880-CONF (2005).

[26] DØ Collaboration, B. Abbott *et al.*, Phys. Rev. Lett. 85, 256-261 (2000).

[27] The LEP Electroweak Working Group, hep-ex/0511027 (2005); Lep Electroweak Working Group web site: `http://lepewwg.web.cern.ch`

[28] DØ Collaboration, V.M. Abazov *et al.*, Nucl. Instrum. and Meth. A **565**, 463-537 (2006).

[29] K. Kleinknecht, "Detectors for particle radiation", Cambridge.

[30] R. Hopper, G. Landsberg, DØ Note 4230 (2003).

[31] DØ Collaboration, V.M. Abazov *et al.*, Nucl. Instrum. Meth. A **552**, 372 (2005).

[32] T. Edwards *et al.*, FERMILAB-TM-2278-E (2004).

[33] S. Klimenko, J. Konigsberg, T. M. Liss, FERMILAB-FN-0741 (2003).

[34] DØ top working group, web site: `http://www-d0.fnal.gov/Run2Physics/top/`.

[35] M. L. Mangano *et al.*, JHEP **07**, 001 (2003).

[36] DØ Collaboration, V.M. Abazov *et al.*, Phys. Lett. B 626, 45 (2005).

[37] H. Greenlee, DØ Note 4303 (2004).

[38] A. Garcia. Bellido, DØ Note 4320 (2004); M. Narain, F. Stichelbaut, DØ Note 3560 (2004).

[39] A. Schwartzmann, M. Narain, DØ Note 4042 (2002).

[40] C. Clément *et al.*, DØ Note 4350 (2004).

[41] C. Clement *et al.*, DØ Note 4667 (2004).

[42] D. Whiteson *et al*, DØ Note 4184 (2003).

[43] J. Kozminski *et al*, DØ Note 4449 (2004).

[44] C. Clement *et al.*, DØ Note 4662 (2004).

[45] G. Blazey *et al*, DØ Note 3750 (2000).

[46] U. Bassler, G. Bernardi, DØ Note 4124 (2002).

[47] J.R. Vlimant *et al*, DØ Note 4146 (2003).

[48] G. Bernardi, E. Busato, J.R. Vlimant, DØ Note 4335 (2004).

[49] DØ jet energy scale working group, web page: `http://www-d0.fnal.gov/phys_id/jes/d0_private/jes.html`.

[50] DØ Collaboration, B. Abbot *et al.*, Nucl. Instrum. Meth. A **424**, 352 (1999).

[51] M. Agelou *et al*, DØ Note 4775 (2005).

[52] Decay length L_{xy} is defined as the distance from the primary to the secondary vertex in the plane transverse to the beamline. Decay length significance is defined as $L_{xy}/\sigma_{L_{xy}}$, where $\sigma_{L_{xy}}$ is the uncertainty on L_{xy}.

[53] DØ *b*-identification Working Group webpage: `http://www-d0.fnal.gov/phys_id/bid/d0_private/bid.html`.

[54] D. Boline *et al*, DØ Note 4796 (2005); A. Schwartzman, M. Narain, DØ Note 4080 (2003); A. Schwartzman, M. Narain, DØ Note 4081 (2003).

[55] DØ Collaboration, V. Abazov *et al.*, Phys. Lett. B 626, 35 (2005).

[56] C. Clément *et al.*, on the *Measurement of the $t\bar{t}$ cross section in the ℓ+jets channel at sqrt(s) = 1.96 TeV using lifetime tagging*, see: `http://www-d0.fnal.gov/Run2Physics/top/private/Confs05/ljets_btag_note_v2_1.pdf`.

[57] J. Huston *et al.*, Eur.Phys.J. C12 375-392 (2000).

[58] T. Sjöstrand *et al.*, Comp. Phys. Commun. **135**, 238 (2001).

[59] CDF Collaboration, T. Affolder *et al.*, Phys. Rev. D **65**, 092002 (2002).

[60] B. Andersson, G. Gustafson, G. Ingelman and T. Sjöstrand, Phys. Rep. **97** 31 (1983); B. Andersson, "The Lund Model", Cambridge University Press (1998).

[61] M.G. Bowler, Z. Phys. C11, 169 (1981).

[62] C. Peterson *et al.*, Phys. Rev. D **27** 105 (1983).

[63] D. J. Lange, Nucl. Instrum. Meth. A **462**, 152 (2001).

[64] DØ Collaboration, see: `http://www-d0.fnal.gov/computing/MonteCarlo/simulation/d0gstar.html`.

[65] R. Brun, F. Carminati, CERN Program Library Long Writeup W5013 (unpublished) (1993).

[66] M. Klute, L. Phaf and D. Whiteson, DØ Note 4122 (2003).

[67] ROOT, "An Object-Oriented Data Analysis Framework", see: `http://root.cern.ch`.

[68] The DØ Top Working group, DØ Note 4419 (2004).

[69] DØ Collaboration, A. Abachi *et al.*, Phys.Rev.Lett. **79**, 1197 (1997); Scott Snyder, DØ Doctoral Thesis, State Univeristy of Stony Brook, FERMILAB-THESIS-1995-27 (1995).

[70] M. Agelou *et al.*, DØ Note 4512 (2004).

[71] ALEPH Collaboration, D. Buskulic *et al.*, Phys. Lett., B **356** (1995).

[72] CDF Collaboration, T. Affolder *et al*, Phys. Rev., D **61**, 072005 (2000); CDF Collaboration, R. Affolder *et al*, Phys. Rev., D **60**, 072003 (1999).

[73] DELPHI Collaboration, J. Abdallah *et al.*, hep-ex/0412004, CERN-PH-EP-2004-062.

[74] L3 Collaboration, Acciarri, M. *et al.*, Phys. Lett., B **439** 225 (1998).

[75] OPAL Collaboration, G. Abbiendi *et al.*, Phys. Lett., B **546** 29-47 (2002).

[76] R. D. Field, Phys. Rev. D **65**, 094006 (2002).

[77] CDF Collaboration, D. Acosta *et al.*, Phys. Rev. D **71**, 092001 (2005).

[78] Daniel Abraham Wijngaarden, PhD thesis , FERMILAB-THESIS-2005-14 (2005).

[79] DØ Collaboration, S. Abachi *et al.*, Phys. Rev. Lett. **79**, 1197 (1997).

[80] DØ Collaboration, V.M. Abazov *et al.*, submitted to Phys. Rev. **D**, hep-ex/0609053 (2006).

[81] E.T. Jaynes, "Probability Theory", Cambridge.

[82] Atlas Collaboration, ATLAS TDR 14, CERN/LHCC 99-14 (1999).

[83] U. Baur, Phys. Rev. D**64** 094019 (2001); U. Baur, A. Juste, L. H. Orr, D. Rainwater, *Probing electroweak top quark couplings at hadron and lepton colliders*, to appear in the proceedings of 8th DESY Workshop on Elementary Particle Theory: Loops and Legs in Quantum Field Theory, Eisenach, Germany, 23-28 Apr 2006, hep-ph/0606264, (2006).