MEASUREMENT OF THE TOP QUARK PAIR PRODUCTION CROSS SECTION IN PROTON-ANTIPROTON COLLISIONS AT $\sqrt{s} = 1.96$ TeV

HADRONIC TOP DECAYS WITH THE D0 DETECTOR

This work is part of the research program of the "Stichting voor Fundamenteel Onderzoek der Materie (FOM)", which is financially supported by the "Nederlandse organisatie voor Wetenschappelijke Onderzoek (NWO)".

The author was financially supported by the University of Twente and by the "Nationaal instituut voor subatomaire fysica (Nikhef)".

ISBN 978-90-6488-032-2

Cover design: Kees Huyser (www.huyser.net)

MEASUREMENT OF THE TOP QUARK PAIR PRODUCTION CROSS SECTION IN PROTON-ANTIPROTON COLLISIONS AT $\sqrt{s} = 1.96$ TeV

HADRONIC TOP DECAYS WITH THE D0 DETECTOR

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de rector magnificus,
prof. dr. W.H.M. Zijm,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
op vrijdag 16 januari 2009 om 15.00 uur

door

Jeroen Guido Hegeman
geboren op 28 februari 1978
te Utrecht

Dit proefschrift is goedgekeurd door:
Promotor: prof. dr. ing. B. van Eijk

Contents

Introduction

Of the six quarks in the standard model the top quark is by far the heaviest: 35 times more massive than its partner the bottom quark and more than 130 times heavier than the average of the other five quarks. Its correspondingly small decay width means it tends to decay before forming a bound state. Of all quarks, therefore, the top is the least affected by quark confinement, behaving almost as a free quark. Its large mass also makes the top quark a key player in the realm of the postulated Higgs boson, whose coupling strengths to particles are proportional to their masses. Precision measurements of particle masses for e.g. the top quark and the W boson can hereby provide indirect constraints on the Higgs boson mass.

Since in the standard model top quarks couple almost exclusively to bottom quarks ($t \to Wb$), top quark decays provide a window on the standard model through the direct measurement of the Cabibbo-Kobayashi-Maskawa quark mixing matrix element V_{tb}. In the same way any lack of top quark decays into W bosons could imply the existence of decay channels beyond the standard model, for example charged Higgs bosons as expected in two-doublet Higgs models: $t \to H^+b$.

Within the standard model top quark decays can be classified by the (lepton or quark) W boson decay products. Depending on the decay of each of the W bosons, $t\bar{t}$ pair decays can involve either no leptons at all, or one or two isolated leptons from direct $W \to e\overline{\nu}_e$ and $W \to \mu\overline{\nu}_\mu$ decays. Cascade decays like $b \to Wc \to e\overline{\nu}_e c$ can lead to additional non-isolated leptons. The fully hadronic decay channel, in which both Ws decay into a quark-antiquark pair, has the largest branching fraction of all $t\bar{t}$ decay channels and is the only kinematically complete (i.e. neutrino-less) channel. It lacks, however, the clear isolated lepton signature and is therefore hard to distinguish from the multi-jet QCD background. It is important to measure the cross section (or branching fraction) in each channel independently to fully verify the standard model.

Top quark pair production proceeds through the strong interaction, placing the scene for top quark physics at hadron colliders. This adds an additional challenge: the huge background from multi-jet QCD processes. At the Tevatron, for example, $t\bar{t}$

production is completely hidden in light $q\bar{q}$ pair production. The light (i.e. not bottom or top) quark pair production cross section is six orders of magnitude larger than that for $t\bar{t}$ production. Even including the full signature of hadronic $t\bar{t}$ decays, two b-jets and four additional jets, the QCD cross section for processes with similar signature is more than five times larger than for $t\bar{t}$ production. The presence of isolated leptons in the (semi)leptonic $t\bar{t}$ decay channels provides a clear characteristic to distinguish the $t\bar{t}$ signal from QCD background but introduces a multitude of W- and Z-related backgrounds.

In the absence of any leptons (from the W decays) one has to resort, in addition to using the above multi-jet signature, to signal-background separation based on jet properties (e.g. p_{T} and η) and inter-jet characteristics of $t\bar{t}$ decays, like energy- and rapidity differences. The level of complexity encountered in these multi-jet events is perhaps most clearly illustrated by the fact that no theoretical or Monte Carlo models exist that are able to accurately describe QCD multi-jet collider data. Another complication of the high jet multiplicity is that effects like jet reconstruction efficiency and energy calibration apply to all jets, multiplying the effects and making efficiency and calibration precision of extreme importance. Moreover, whereas in the case of (semi)leptonic $t\bar{t}$ decays energy/momentum calibration is mostly relevant to mass (as opposed to cross section) analyses, the heavy reliance on jet- and inter-jet characteristics makes the calibration of paramount importance for any analysis in the hadronic channel.

With the above points in mind the analysis described in this thesis sets out to measure the top-antitop quark pair production cross section at a center-of-mass energy of $\sqrt{s} = 1.96$ TeV in the fully hadronic decay channel. The analysis is performed on 1 fb^{-1} of Tevatron Run IIa data taken with the D0 detector between July 2002 and February 2006. A neural network is used to identify jets from b-quarks and a likelihood ratio method is used to separate signal from background. To increase efficiency several multi-jet triggers are combined (OR-ed), correcting for any efficiency overlaps between the different triggers. Special care was taken in the trigger selection to maintain the normalisation of the integrated luminosity. To avoid reliance on, possibly imperfect, Monte Carlo models for the modelling of the QCD background, the background was modelled using a dedicated data sample. The $t\bar{t}$ signal was modelled using the ALPGEN and PYTHIA Monte Carlo event generators. The generated signal sample was passed through the full, GEANT based, D0 detector simulation and reconstructed using the default D0 reconstruction software.

The measured cross section is

$$\sigma_{t\bar{t}} = 7.5 \pm 1.3 \text{ (stat)} \, ^{+0.8}_{-0.9} \text{ (syst)} \, \pm 0.5 \text{ (lumi)}$$

assuming a top quark mass of $m_t = 172.4$ GeV/c^2 (the current world average top mass measurement combination). The systematic uncertainty is dominated by the background modelling: $\Delta\sigma_{t\bar{t}}(\text{background modelling}) = ^{+0.48}_{-0.61}$ pb. Above cross section is in perfect agreement with the standard model prediction of 7.50 ± 0.58 pb.

This thesis starts with a brief overview of the standard model of particle physics (chapter 1). The focus lies on the key role of the top quark in the standard model and its connection to the Higgs sector. The basics of hadron-hadron interactions are discussed in the context of $t\bar{t}$ pair production at the Tevatron. Chapters 2 and 3 provide information on the experimental context of this thesis. Chapter 2 gives and overview of the Fermilab Tevatron collider and the D0 detector. Chapter 3 focuses on the reconstruction algorithms and the definitions of physics objects used in the D0 experiment. Uniformity and precision calibration of calorimeter- and jet energy in high luminosity hadron collider experiments is a challenging task, but of much importance especially for multi-jet analyses. Detailed understanding is required of the effects and interplay of many different contributions, originating not only from the particle interactions themselves but also related to detector effects and collider parameters. Chapter 4 is devoted to the description of the calorimeter- and jet energy calibration in the D0 experiment. Chapters 5 to 9 describe the $t\bar{t}$ cross section measurement. Chapters 5 and 6 discuss the samples, event selection and triggers used. One of the key ingredients of this analysis is the use of collider data to model the QCD background. The background modeling procedure is described in detail in chapter 7. Special attention is paid to validation of the background samples and to estimates of the systematic uncertainties related to the background model. The likelihood method used to separate signal from background is discussed in chapter 8. In chapter 9 the measured cross section and associated uncertainties are presented and the cross section is compared to the theoretical prediction based on the standard model.

Chapter 1

The standard model of particle physis

The nature of day-to-day objects around us is described by the laws of physics. The paths and interactions of billiard balls on the table follow the laws of classical mechanics. The orbits of planets, even of whole galaxies, are described by the general theory of relativity. The properties of bulk materials are the field of solid state physics and at the atomic level are governed by quantum mechanics. Delving one step deeper, inside the atomic nuclei lies the realm of subatomic physics. Modern sub-atomic physics, or *particle physics*, is described by the *standard model*: a relativistic quantum field theory describing all known fundamental interactions except gravity. A multitude of good literature on the standard model is available (see for example references [1, 2]). In this chapter only a brief overview is given, focusing on the as yet heaviest particle in the model: the top quark.

1.1 Matter particles and force carriers

All matter consists of atoms and molecules. Atoms in turn consist of a nucleus, made up of protons and neutrons, surrounded by one or more electrons. This level of description of matter and its behaviour is the domain of nuclear physics. The protons and neutrons themselves again consist of smaller particles called quarks. More specifically two kinds of quarks: *up* and *down*. Also, the electron has a partner: the neutrino. The neutrino, however, is invisible and its presence only detectable indirectly in for instance radioactive beta decay. Together these four particles form the first of three generations of matter particles in the standard model (see figure 1.1). The second and third generations contain heavier partners of the first generation particles. The second generation partner of the electron is the two-hundred-times-heavier muon.

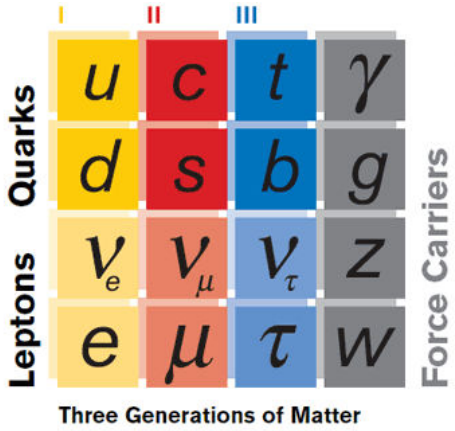


Figure 1.1: *The three generations of particles in the standard model (columns I, II, III) together with the force carriers (last column). Only particles from the first generation occur abundantly on earth.*

Their third generation partner is the tau, again seventeen times heavier than the muon. Like the electron, both muon and tau have an invisible neutrino partner. Together these particles are called the leptons. The heavier second generation partners of the up- and down quarks are called *charm* and *strange*. The third generation contains the heaviest quarks, the *top* and *bottom*. All matter around us is made up of particles from the lightest first generation. Neutrinos are generated in, for example, nuclear fusion processes inside the sun. Millions of neutrinos reach the earth every second together with a broad spectrum of electromagnetic radiation: photons. Since the neutrino interaction probability is extremely small almost all pass through the earth unnoticed. Cosmic rays like protons and light nuclei interact in the earth's atmosphere and generate (among other things) muons which in turn decay almost exclusively into electrons and neutrinos. The quarks from the second and third generations are rare and only produced in extreme cosmic environments and particle physics laboratories. The quark names are often written shorthand using just the first letter, e.g. t for the top quark. The leptons are abbreviated e (electron), μ (muon) and τ (tau lepton) and their neutrinos as ν with corresponding subscript.

Three separate families of *anti*particles exist, with exactly the same masses as their matter partners but with opposite quantum numbers like electric charge. Even rarer than most of the matter particles, antiparticles are signified with a 'bar' over their

particle name, e.g. $\bar{\mathrm{t}}$ and $\bar{\mathrm{b}}$, or with an explicitly mentioned opposite electric charge. The latter convention is most common for the leptons: $\mathrm{e}^+\mathrm{e}^-$ instead of $\mathrm{e}\bar{\mathrm{e}}$.

The standard model describes the properties and interactions of all particles, except for their behaviour under gravity. However, gravity is many orders of magnitude ($\sim 10^{40}$) weaker than the three fundamental forces incorporated in the standard model:

- The *electromagnetic force*, mediated by photons (γ). The best known of the fundamental forces and manifest all around us, the electromagnetic force encompasses visible light, radiant heat, electricity, etc.

- The *weak force*, carried by particles called the weak vector bosons, the $\mathrm{W}^{\pm}$ and Z^0. The weak force is the force behind radioactive beta decay and responsible for the decay of heavy quarks.

- The *strong nuclear force* (Quantum ChromoDynamics or QCD, keeping together the quarks inside protons and neutrons. Quarks do not occur freely, they are confined to bound states with balancing quantum numbers. The mediator particles of the strong force are called gluons. One of the special features of the strong interaction is asymptotic freedom: unlike in electroweak interactions, the strength of the force *in*creases with increasing distance. Prying apart quarks from a bound state builds up a force field between them. At some point the energy stored in the field is sufficient to be converted into new (anti)quarks, again forming bound states. In high energy particle collisions, where strongly interacting particles are forcibly separated, this process occurs repeatedly, leading to collimated bundles of particles called jets.

In over twenty years of experiments the standard model has been extensively tested and shown to accurately predict many quantities, often to astounding precision. Examples are the production- and decay rates of different particles and the masses of the W- and Z bosons. The best illustration of the level of internal consistency within the standard model is probably illustrated by the indirect limits on the top quark mass obtained by the LEP experiments. Quantum corrections to the Z boson propagator modify the Z decay width. Precision electroweak fits to the Z pole predicted the top quark mass to be within $m_{\mathrm{t}} = 173^{+13}_{-10}$ GeV/c^2 well before the top quark was experimentally discovered [3]. [1]

[1] Indirect limits on the top quark mass derived from precision electroweak fits have been available (at least) as far back as 1992. This specific result was published after the discovery of the top quark.

Successful as the standard model is, it is not complete. Apart from missing the connection to gravity, there are experimental observations it does not explain. For example, it does not *explain* how strongly particles interact. The strengths of the individual couplings have to be introduced by hand. Another 'missing link' of the standard model concerns the particle masses. Not only does it not explain the particle masses, they don't exist at all: standard model particles are massless. [2] In total some 29 parameters are required to operate the standard model.

1.2 The Higgs mechanism

The standard model is founded on mathematical symmetries hinting at conserved quantities. These gauge symmetries require all particle fields to be intrinsically massless. One way to introduce non-zero particle masses without breaking renormalisability is through spontaneous symmetry breaking by the *Higgs mechanism* [4, 5, 6]. [3] In it's simplest form, a complex Higgs double field fills the vacuum, acquiring a non-zero vacuum expectation value and breaking the $SU(2)_L \times U(1)_Y$ symmetry (which rules the electroweak force) locally to a residual $U(1)_{\mathrm{EM}}$ symmetry. The Higgs doublet introduces four new parameters. Three of these four degrees of freedom are transformed into Goldstone bosons which are absorbed into the W and Z bosons, making these bosons massive while leaving the photon massless. The fourth degree of freedom leads to a new massive scalar particle: the Higgs boson. Additional (gauge invariant) terms can be formed in the Lagrangian to generate the masses of the quarks and leptons. [4] The coupling strength between a particle and the Higgs field scales with the particle mass. One important aspect of the Higgs mechanism is that it introduces a relationship between the W and Z masses and the electroweak mixing angle θ_{W} which determines the relative strengths of the electromagnetic and weak interactions: $M_{\mathrm{W}}/M_{\mathrm{Z}} = \cos\theta_{\mathrm{W}}$. Experimental results show that this is indeed the case to within 1‰ [11]. Unfortunately, the theory does not predict the mass of the Higgs boson. Even though the vacuum expectation value $v = (\sqrt{2}G_{\mathrm{F}}^{-1/2})$ is fully fixed by the Fermi

[2] Explicit mass terms in the standard model Lagrangian would break local gauge invariance, leaving the model unrenormalisable. Unrenormalisable field theories have no predictive value.

[3] The term 'Higgs mechanism' does not do justice to the many other contributors. Brout and Englert at around the same time reached essentially the same conclusion [7], as did Guralnik, Hagen and Kibble [8]. Higgs, however, was the one to postulate a new, massive scalar particle.

[4] Experimental evidence has also shown that not all neutrinos can be massless [9, 10]. Inclusion of neutrino masses into the standard model requires additional changes beyond the Higgs mechanism.

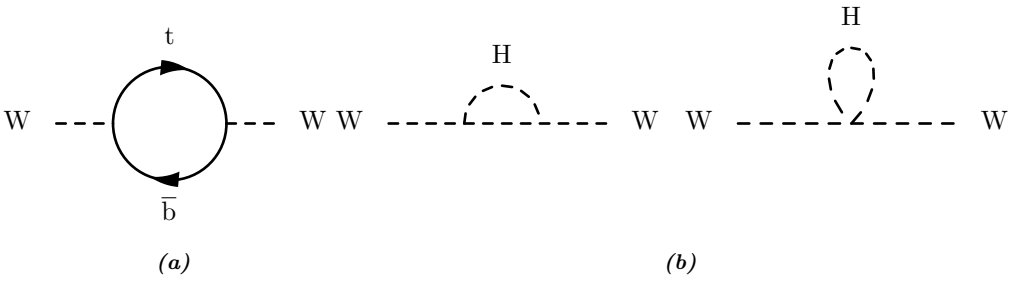


Figure 1.2: *Radiative corrections to the W boson mass due to virtual quark loops (a) and due to Higgs boson loops (b). Similar diagrams exist that modify the Z boson propagator.*

coupling constant G_{F} the Higgs boson mass $M_{\mathrm{H}} = \sqrt{2\lambda}v$ also depends on the unknown quartic Higgs self-coupling λ.

Direct searches for the Higgs boson have been performed at four LEP experiments. So far the famous boson remains elusive. The combined experimental result gives a lower bound on the Higgs boson mass of $M_{\mathrm{H}} \geq 114.4$ GeV/c^2 at 95% confidence level [12].

Precision measurements of standard model observables can give indirect information on the Higgs boson mass. Both the top quark and W boson masses are sensitive to the Higgs boson mass through radiative corrections (figure 1.2). Higgs boson loops modify the boson propagators and hence their masses. Similar loops containing top quarks also modify the vector boson masses (the lighter quarks also contribute but to a much less extent). Having measured the W and top quark masses, it's possible to determine the most likely value of the Higgs boson mass. Figure 1.3 illustrates this relationship between the W, t and H masses together with recent values of M_{W} and m_{t}.

Recent results combining measurements of the top quark mass and the W boson mass [13] favor low Higgs mass values. Figure 1.4 shows the χ^2 curve of a global fit to all [5] precision electroweak data as a function of the Higgs boson mass. One of the limiting factors in this global fit is the precision to which electroweak coupling strength α (specified at the scale of the Z mass: $\alpha(M_{\mathrm{Z}}^2)$) is known. The coupling strength changes as a function of the momentum scale due to vacuum polarisation loop corrections.

[5] The indirect measurement of the W boson mass based on neutrino-nucleon scattering by the NuTeV collaboration deviates $\approx 2.7\sigma$ from any other (direct or indirect) measurement. This measurement is not used in the global fit, but its effect is shown in figure 1.4 as the curve marked 'incl. low Q^2 data'.

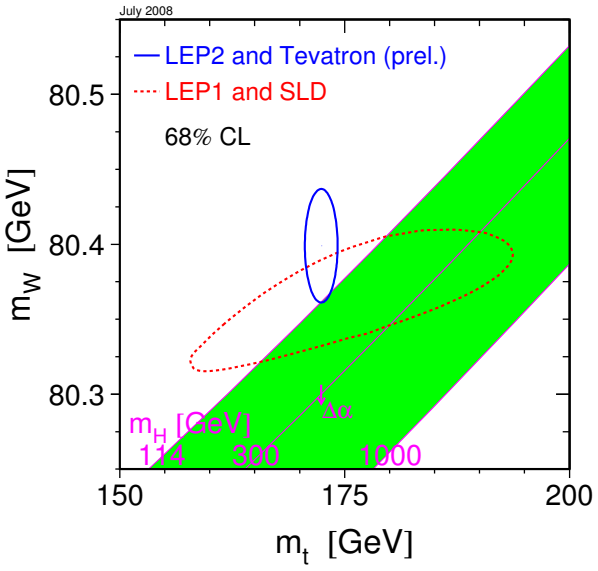


Figure 1.3: *Measured values of M_W and m_t combining the LEP1 experiments and SLD data (area enclosed by dashed line), and for the combination of the four LEP2 experiments with CDF and D0 (ellipse enclosed by solid line) [13]. Possible values of the Higgs boson mass are represented by slanted lines. The filled area spans the range from the lower LEP exclusion limit up to $M_H = 1$ TeV. The arrow marked $\Delta\alpha$ demonstrates the effect on the relation between the masses for a one-sigma variation of the electroweak coupling constant ($\alpha(M_Z)$ increases in the direction of the arrow).*

Contributions come from both lepton loops (known to third order with negligible uncertainty) and loops containing quarks. The contribution from top quarks is small and depends on m_t so it is evaluated inside the fit. The contribution from the other five quarks $\delta\alpha^{(5)}_\mathrm{had}$ is determined from data to be $\delta\alpha^{(5)}_\mathrm{had} = 0.02758 \pm 0.00035$ [14]. This result combines measurements from the BES [15] collaboration, as well as from CMD-2 [16] and KLOE [17]. An alternative approach is to determine $\delta\alpha^{(5)}_\mathrm{had}$ from theory (using minimal experimental input). This gives a value of $\delta\alpha^{(5)}_\mathrm{had} = 0.02749 \pm 0.00012$ [18]. The effect on the global electroweak fits is shown in figure 1.4 as an alternative curve. All remaining theoretical uncertainties are summarised in the shaded error band around the solid curve.

The preferred value for the Higgs boson mass (minimum of the χ^2 curve) is $M_\mathrm{H} = 84^{+34}_{-26}$ GeV/c^2 (68% confidence level, experimental uncertainty only). The one-sided

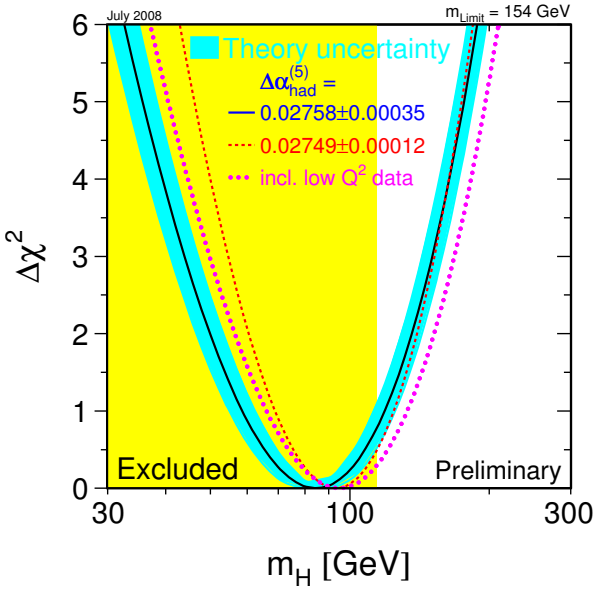


Figure 1.4: *$\Delta\chi^2$ Curves of fits to all high-Q^2 experimental data from the LEP, SLD, CDF and D0 experiments as a function of the Higgs boson mass assuming two values for the hadronic corrections to the electroweak coupling constant [13]. The minimum of the solid curve corresponds to a preferred Higgs boson mass of $M_H = 84$ GeV/c^2. The experimental uncertainty obtained from a $\Delta\chi^2 = 2.7$ step along the solid line is $^{+34}_{-26}$ GeV/c^2. The vertical shaded area on the left-hand side represents the LEP exclusion region.*

95% confidence level interval derived from the same curve ($\Delta\chi^2 = 2.7$) gives an upper limit of $M_{\mathrm{H}} < 154$ GeV/c^2. Combining this result with the LEP exclusion region increases this upper bound to $M_{\mathrm{H}} < 185$ GeV/c^2.

1.3 Top quark physics

Due to its large mass the top quark obtains a special role in the standard model, making it a prime candidate for searches for physics beyond the standard model, e.g. Higgs searches or searches for anomalous top quark decay modes.

Single top quarks are produced via the electroweak force, $\mathrm{t\bar{t}}$ pairs through the strong interaction, providing two independent windows on standard model top physics. The

single top channel allows for the direct measurement of the t → Wb CKM [19, 20] coupling V_{tb}.

This section focuses on standard model physics processes involving top quarks. [6] This is one of the areas of experimental particle physics producing new and improved measurements at an astounding rate. By no means is the representation here meant to provide a complete overview. For a detailed review of top quark physics see for example ref. [21].

1.3.1 Top quark properties

The top quark is the $Q = 2/3$, $T_3 = 1/2$ member of the third quark generation. With a mass of $m_{\mathrm{t}} \approx 170$ GeV/c^2 it is much heavier than its weak-isospin partner the bottom quark ($m_{\mathrm{b}} \approx 4.5$ GeV/c^2) and by far the heaviest of all quarks.

The standard model top quark almost exclusively decays into a bottom quark and a W boson; the t → Wb branching fraction is larger than 99.8% [11]. Decays into (a W boson and) a s or a Wd quark are suppressed with respect to the Wb channel by the square of the corresponding CKM matrix elements. The top quark decay width in the standard model (considering only the Wb decay channel and ignoring terms of the order of $\frac{m_{\mathrm{b}}^2}{m_{\mathrm{t}}^2}$, α_s^2 and $\frac{\alpha_s}{\pi}\frac{M_{\mathrm{W}}^2}{m_{\mathrm{t}}^2}$) is given by [22]:

$$\Gamma_{\mathrm{t}} = \frac{G_{\mathrm{F}} m_{\mathrm{t}}^3}{8\pi\sqrt{2}} \left(1 - \frac{M_{\mathrm{W}}^2}{m_{\mathrm{t}}^2}\right)^2 \left(1 + 2\frac{M_{\mathrm{W}}^2}{m_{\mathrm{t}}^2}\right) \left[1 - \frac{2\alpha_s}{3\pi}\left(\frac{2\pi^2}{3} - \frac{5}{2}\right)\right].$$

Using recent precision measurements of G_{F}, M_{W} and m_{t} [11] this gives an approximate decay width of ~ 1.3 GeV. The correspondingly short lifetime of $\mathrm{O}(10^{-24})$ s, compared to the time scale governing QCD processes, $1/\Lambda_{\mathrm{QCD}} = \mathrm{O}(10^{-23})$ s, means top quarks will predominantly decay before forming bound states. All information of the quantum numbers carried by the quarks is transferred to the decay products instead of being lost in the hadronisation process. Therefore, the decay of the top quark can be described by the decay of a 'free quark'.

Top quark decay channels

The experimental signatures of $\mathrm{t\bar{t}}$ events can be classified based on the decay products of the Ws. One third of all Ws decay into leptons, spread approximately evenly

[6] Whenever particles or processes are mentioned, the charge conjugate particles and/or processes are implied. E.g. the top decay t → Wb denotes both $\mathrm{t} \to \mathrm{W^+b}$ and $\mathrm{\bar{t}} \to \mathrm{W^-\bar{b}}$.

over electrons, muons and taus. The other two thirds decay hadronically, into quark-antiquark pairs. Of the hadronic decays almost half ($\approx 46\%$) of the decays go to $c\bar{s}$. The tau leptons decay into electrons, muons or hadrons. The first two channels contribute to the direct leptonic decays. Hadronic tau decays lead to narrow jets, which are hard to distinguish from jets originating from partons (quarks and gluons). Depending on the missing transverse energy and the tau identification efficiency, part of the hadronic tau decays will be absorbed in the hadronic W decay channels.

Of the $t\bar{t}$ decays, the di-lepton channel contains those events in which both Ws decay leptonically into either an electron or a muon ($\approx 6\%$). The cases in which only one of the Ws decays leptonically (into an electron or a muon) and the other one into two jets define the lepton+jets channel ($\approx 34\%$). The all-jets channel contains all events in which both Ws decay hadronically ($\approx 46\%$), leading to an experimental signature of six jets in the absence of any leptons. Subsequent leptonic decays in the cascade decay of the top quark predominantly result in additional leptons hidden inside the hadronic jets. (A more detailed division into $t\bar{t}$ decay channels is given in table 5.1 on page 104.)

Direct leptonic decays have the advantage of containing an isolated lepton which is both convenient to trigger on and useful in the selection of signal events. The disadvantage is the presence of the neutrino accompanying the lepton. This not only requires additional kinematic constraints to complete the events, it also introduces a dependence on the measured missing transverse energy. Missing E_{T} measurements depend on the calibration of all other physics objects like jets and EM clusters. This implicit definition makes $\not\!\!E_{\mathrm{T}}$ a complicated variable to calibrate. [7]

The experimental signature of the dominant all-jets channel (figure 1.5) is:

- At least six jets.
- Two of these jets are (in general high p_{T}) b-jets from the top quark decays.
- The four other quarks should in principle form two pairs, each with the W boson mass as invariant mass.
- Absence of any isolated leptons or missing transverse energy.

[7] There are several ways to experimentally define the missing energy. Even using the straightforward approach of summing over all calorimeter cells, $\not\!\!E_{\mathrm{T}}$ has to be corrected for the effects of zero-suppression and the presence of 'real' physics objects like jets, electromagnetic clusters and muons, turning it into an indirect measurement.

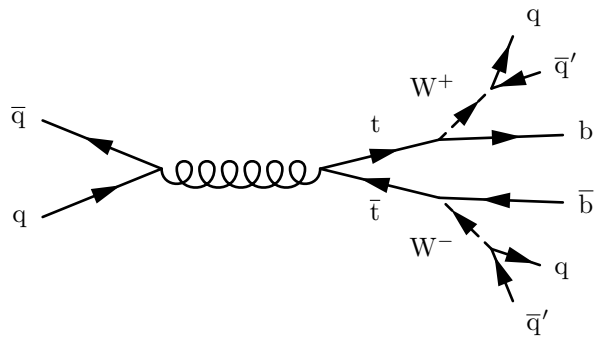


Figure 1.5: *Typical event signature of fully hadronic decays of top-antitop quark pairs: two high momentum b-jets in the presence of four additional jets from the hadronic W decays and in the absence of any leptons. Additional initial- or final-state radiation jets may be present.*

The absence of isolated leptons is a disadvantage from the trigger point of view. Although the high jet multiplicity makes it relatively easy to design high efficiency multi-jet triggers for this channel, it provides another challenge: avoiding too high trigger rates on QCD multi-jet events.

The multi-jet $t\bar{t}$ signal is extremely small compared to general QCD multi-jet production. The cross section for hard di-jet production (where additional radiative jets can increase in jet multiplicity) is more than six orders of magnitude larger. For $b\bar{b}$ di-jet production the cross section is already a thousand times smaller, while for six-jet events the QCD cross section is 'only' a hundred times larger than the expected $t\bar{t}$ cross section. Moreover, the cross section for QCD events with both a $b\bar{b}$ pair and four additional jets is approximately six times larger than the expected signal. This suggests a strong preselection of events on the presence of at least six jets, at least two of which are labelled as jets from b-quarks. Since approximately half of the W decays will produce c-quarks which are sometimes misidentified as b-quarks one should allow for the presence of more b-jets than the expected two from the $t\bar{t}$ signal.

W boson polarization

The $V - A$ nature of the $t \to Wb$ coupling in the standard model leads to a predominantly longitudinal polarisation of the W bosons: $\approx 70\%$ [23, 24, 25] for values of the top quark mass in the 170–175 GeV/c^2 range [26]. Any polarisation of the W boson directly affects the angular distributions of its decay products. Longitudinal polarisation results in a distribution proportional to $\sin^2\theta^*$, where θ^* is the angle

between the W boson's momentum in the top quark rest frame and the down-type quark momentum in the W boson rest frame. Transverse polarisation results in a $1 + \cos^2\theta^*$ behaviour. In the former case both decay products favour the direction perpendicular to that of the W while in the latter case the thrust axis tends to be aligned with the W direction. The relative strength of the longitudinal contribution leads to a more even distribution of energy between the two daughter jets of W bosons in hadronic top quark decays. This can be used as an additional characteristic of the experimental signature.

Top quark mass

Both the D0 and CDF collaborations have published direct measurements of the top quark mass in all decay channels. The highest precision contributions come from analyses in the lepton+jets channel. The di-lepton channel suffers from the very small branching fraction and kinematically from the presence of at least two neutrinos, leaving the system kinematically under-constrained. The all-jets channel, even though it has the highest branching fraction of all, lacks the presence of an isolated lepton, making it harder to separate signal from background. Correspondingly, mass measurements in the all-jets channel generally have lower precision than in the leptonic channels.

Traditionally, three main approaches are used in the lepton+jets mass analyses. The 'template method' uses an over-constrained kinematic fit testing the hypothesis that each event represents a $t\bar{t} \rightarrow$ lepton+jets signal. The W mass is used to constrain the neutrino momentum (up to an ambiguity in the sign of the component along the beam- or z-direction), leaving 24 possible parton-jet assignments. Using b-tagging information this number can be reduced to 12 (one b-tagged jet) or 4 (two b-tagged jets) possible solutions. For each event the solution with the best agreement (lowest χ^2) is chosen. The top quark mass is derived from the distribution of the fitted top quark mass.

In the 'matrix element method' [24] all possible parton-jet assignments are taken into account, each one weighted with the probability it represents a $t\bar{t}$ event. This probability is evaluated based on the leading-order matrix element approximation for $t\bar{t}$ production and decay, using transfer functions for the parton-to-jet transition modelling. Combining features from each of the above two methods, the 'ideogram method' [27] also considers all possible parton-jet assignments but events are weighted using the outcome of a kinematic fit to all 24 permutations and an event probability based

on a topological discriminant suppressing events which would show little separation between signal and background. The above two methods use the lepton for triggering and event selection but rely on the calorimeter information to reconstruct the jets and the top quark mass.

A more recently developed analysis method does not use any calorimeter information at all, relying solely on tracking information [28]. This method exploits the fact that the boost of the b-quark in the rest frame of the top quark depends on the ratio of the top and bottom quark masses: $\gamma_{\mathrm{b}} \approx 0.4 m_{\mathrm{t}}/m_{\mathrm{b}}$. This makes the transverse decay length of the bottom quark sensitive to m_{t}. This method requires very accurate reconstruction of secondary vertices and relatively low fake rates but avoids the multiplication of jet energy calibration uncertainties for four or more jets and provides an independent measurement of the top mass. (The CDF measurement based on this method shown in figure 1.6 is strongly statistics limited, explaining the large uncertainties.)

Events in the di-lepton channel are kinematically under-constrained and analyses typically rely on templates derived from Monte Carlo using an assumed top quark mass and integrating over the unknown quantities. The strong dependence on simulation, combined with the small branching fraction typically result in larger uncertainties in di-lepton analyses compared to those in the lepton+jets channel.

As discussed previously the absence of isolated leptons and the large QCD background prove challenging in the hadronic channel. Apart from that, the reconstruction of a top mass from six jets naturally broadens the mass peak due to combinatorics and stacking of jet energy calibration uncertainties.

Figure 1.6 summarises the Tevatron top quark mass measurements, and shows their combined world average determined by the Tevatron electroweak working group [29]. It should be noted that the top quark mass is known to a precision of less than one percent!

1.3.2 Top quark pair production

The production of top-antitop quark pairs proceeds via the strong interaction, requiring a hadron collider environment. In hadron-hadron collisions the interaction can be factorized into three parts: the low momentum interactions between the partons inside the hadrons, the hard parton-parton scatter and the subsequent hadronisation and decay of the collision products. The production cross section, containing the first two

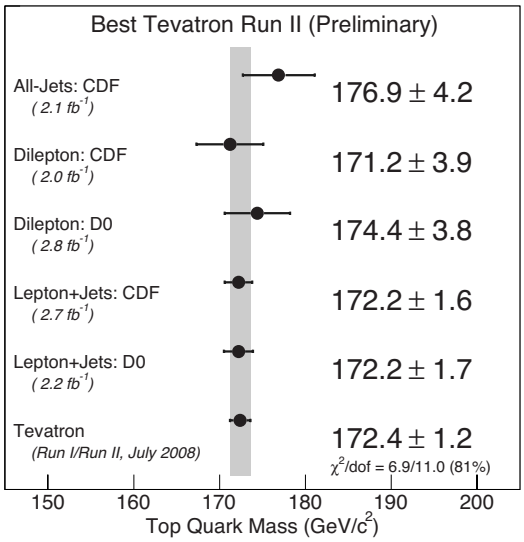


Figure 1.6: *Overview of the Tevatron top quark mass measurements and their combined world average value [29, 30].*

parts, can be written as [31]:

$$\sigma_{\mathrm{t\bar{t}}}(\sqrt{s}, m_{\mathrm{t}}) = \sum_{i,j} \int\int dx_i\, dx_j f_i(x_i, \mu_{\mathrm{F}}^2) f_j(x_j, \mu_{\mathrm{F}}^2) \times \hat{\sigma}_{\mathrm{t\bar{t}}}(\sqrt{s}, m_{\mathrm{t}}, x_i, x_j, \alpha_s(\mu_{\mathrm{R}}^2), \mu_{\mathrm{F}}^2)$$

where the indices i, j run over all incoming parton flavours, the $f_{i/j}(x_{i/j}, \mu_{\mathrm{F}}^2)$ are the parton density functions for parton flavours i/j inside the colliding hadrons, and $\hat{\sigma}_{\mathrm{t\bar{t}}}$ is the cross section for the hard scatter process of the two partons i and j into a $\mathrm{t\bar{t}}$ pair $ij \to \mathrm{t\bar{t}}$. The total cross section is the sum over all parton combinations i, j (quark-antiquark, gluon-gluon, etc.), integrated over all possible combinations of parton momenta x_i, x_j. The average distribution of the longitudinal momentum of each hadron over its constituents is specified by the parton distribution functions (PDFs) $f(x, \mu_{\mathrm{F}}^2)$. The PDFs parameterise the effects of the low momentum parton interactions inside the hadrons and represent probability density functions for finding a parton with momentum fraction x. The factorisation scale μ_{F} specifies the transition between the mass scales of the high momentum transfer hard scatter and the low momentum interactions. Over time several different parameterisations have been developed. One

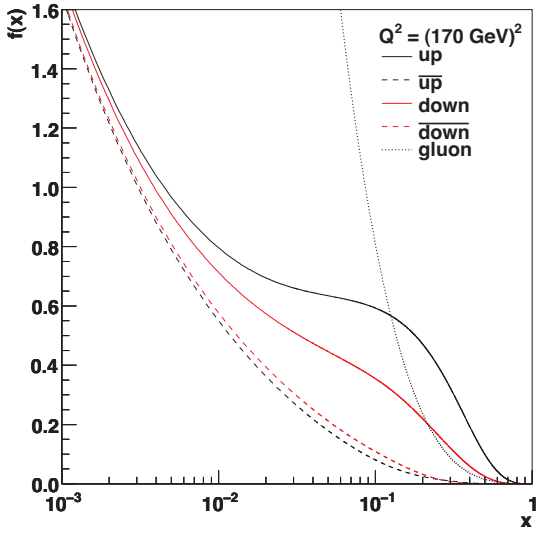


Figure 1.7: *Parton distribution functions according to the* CTEQ6.5 *parameterisation [33] (an NLO PDF based on an NLO approximation of the strong coupling constant α_s). The parton distribution functions cannot be determined theoretically, they are estimated from global fits of QCD calculations to particle collision data.*

well-known parameterisation is the one obtained by the CTEQ collaboration [32]. An example of the CTEQ6.5 PDF set [33] is shown in figure 1.7. This PDF set was chosen for its ability to describe the D0 jet data [34].

In quantum field theory each coupling between two or more particles is associated with a constant specifying the coupling strength. Each interaction can occur via an infinite number of different intermediate states. These subprocesses can be classified based on the number of loops in the associated Feynman diagrams: the order of the contribution. Loops can be interpreted as the temporary splitting of one of the intermediate particles into a particle-antiparticle pair which subsequently recombine. The contribution without loops is called the leading order (LO), the contribution containing one loop the next-to-leading order (NLO), etc. The integral over the particle momenta in the loop generally diverges. The divergences can be absorbed by redefining the particle masses and coupling constants in a process called renormalisation [35, 36]. The renormalisation procedure mathematically introduces an arbitrary cutoff mass scale μ_R. The renormalisation group equation states that for any physical quantity X its dependence on the renormalisation scale μ_R is exactly cancelled by the dependence

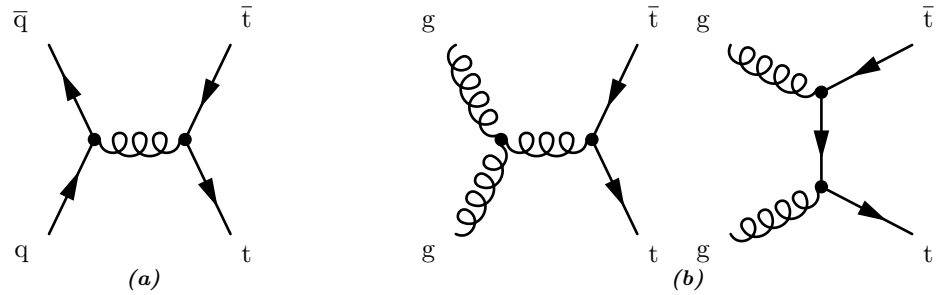


Figure 1.8: *Leading order ($O(\alpha_s^2)$) top-antitop pair production at the Tevatron: (a) quark-antiquark annihilation and (b) gluon fusion.*

of the coupling constants on μ_R [37, 38]. In the standard model all divergences can be absorbed this way, as has been shown by Veltman and 't Hooft [39]. This renormalisability is one of the key features of the standard model, without it, all predictive power beyond the leading order would be lost.

To calculate the hard interaction cross section perturbatively, the renormalisation scale μ_R has to be of the order of the hard interaction scale Q. For $t\bar{t}$ production $Q \sim m_t$. A customary choice for the factorisation and renormalisation scales is to take $\mu_F = \mu_R = m_t$ (similar considerations hold for the choice of both μ_F and μ_R; they are chosen to be equal for simplicity).

At the Tevatron, with $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV, the dominant top-antitop quark pair production mechanism is quark-antiquark annihilation, followed by gluon fusion. [8] At leading order in α_s these processes are shown in figure 1.8. [9] At next-to-leading order also flavour excitation and gluon splitting processes become important (figure 1.9). Since the quark-gluon initiated processes only contribute starting at next-to-leading order their contribution is small compared to the $q\bar{q}$ and gg contributions. At next-to-leading order in α_s the $q\bar{q}$ and gg initiated processes contribute approximately 85% and 15% respectively. The absolute $q\bar{q}$ contribution is well known, but due to the relatively large uncertainties on the gluon PDFs the gg contribution can vary from 11%–21% [40]. Assuming a top quark mass of $m_t = 172$ GeV/c^2 or $m_t = 173$ GeV/c^2, the standard model predicts a next-to-next-to-leading order [10] $t\bar{t}$ production cross section of 7.59 ± 0.58 pb or 7.37 ± 0.56 pb respectively [41]. Interpolated to the world average

[8] At the LHC ($\sqrt{s} = 14$ TeV), smaller values of x will suffice and $t\bar{t}$ production will be dominated by gluon fusion.
[9] Where Feynman diagrams are shown, the charge-conjugate and 'swapped' diagrams are also implied.
[10] More precisely: 'an approximate NNLO cross section which is exact to logarithmic accuracy' [41].

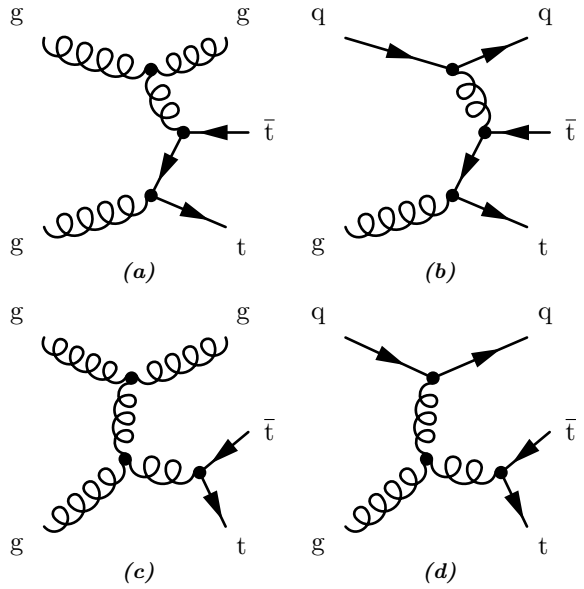


Figure 1.9: *Tree-level (i.e. loopless) next-to-leading order ($O(\alpha_s^3)$) contributions to $t\bar{t}$ production at the Tevatron. (a) and (b): Flavour excitation, (c) and (d): gluon splitting. The quark-gluon initiated processes, (b) and (d), only contribute starting at next-to-leading order and do not present a significant contribution to the overall $t\bar{t}$ production cross section.*

top mass of $m_t = 172.4$ GeV/c^2 this results in a predicted cross section of 7.50 ± 0.58 pb. This prediction is based on the CTEQ6.5M PDF set. The uncertainties on the above values contain contributions from PDF uncertainties and renormalisation/factorisation scale variations. Figure 1.10 shows the dependence of the cross section on the top quark mass.

The requirement that the momenta of the incoming partons provide enough energy to create a $t\bar{t}$ pair at rest places a lower limit on the possible momentum fractions: $x_1 x_2 \geq 4m_t^2/s$. Assuming a top quark mass of 170 GeV/c^2, at the Tevatron ($\sqrt{s} =$ 1.96 TeV) this leads to typical values of x of ≈ 0.17. For these value of x the PDF for the up valence quark dominates (see figure 1.7). This explains the relatively small gluon-gluon contribution to the $t\bar{t}$ cross section at the Tevatron. The rapid drop of the up and down PDFs above $x \approx 0.2$ leads to the strong mass dependence of the cross section shown in figure 1.10.

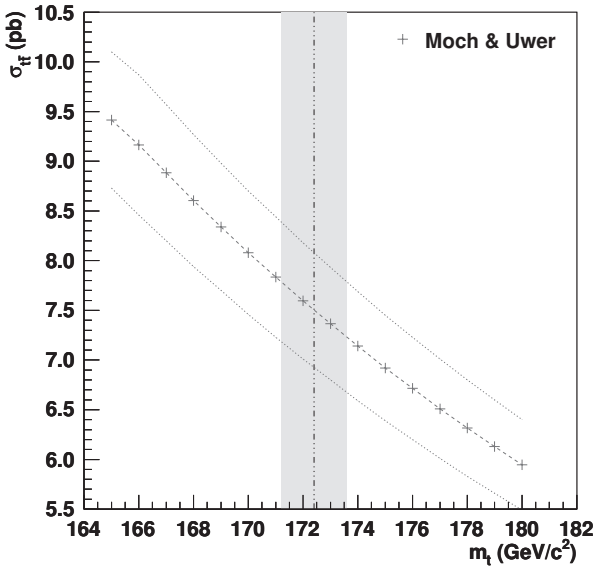


Figure 1.10: *Standard model prediction for top-antitop pair production at the Tevatron [41]. The dashed line with +-markers and the dotted lines show the central value and the uncertainty band respectively. The vertical, dash-dotted line and surrounding shaded area represent the current world average top mass from figure 1.6.*

1.3.3 Single top quark production

Figure 1.11 demonstrates single top quark production through the weak interaction. These processes allow direct measurement of the Cabibbo-Kobayashi-Maskawa (CKM) [20, 19] matrix element $|V_{\mathrm{tb}}|$. [11]

The expected cross sections are roughly a third of those for $t\bar{t}$ pair production. More importantly, single top events with hadronic W decays lead to four-jet final states which are completely buried in QCD background. The leptonic W decays have an expected rate comparable to that of the background. Here, however, the challenge is separating the $p\bar{p} \to t\bar{b} \to Wb\bar{b}$ signal from the $p\bar{p} \to Wb\bar{b}$ and $p\bar{p} \to t\bar{t} \to WbW\bar{b}$ backgrounds.

[11] A measurement of $|V_{\mathrm{tb}}|$ is also possible in the $t\bar{t}$ system through the ratio $BR(t \to Wb)/BR(t \to Wq) = |V_{\mathrm{tb}}|^2/(|V_{\mathrm{td}}|^2 + |V_{\mathrm{ts}}|^2 + |V_{\mathrm{tb}}|^2) = |V_{\mathrm{tb}}|^2$ (the denominator is one, assuming unitarity). However, this requires the assumption that there are no heavy fourth-generation quarks making $|V_{\mathrm{tb}}|^2/(|V_{\mathrm{td}}|^2 + |V_{\mathrm{ts}}|^2 + |V_{\mathrm{tb}}|^2) < 1$.

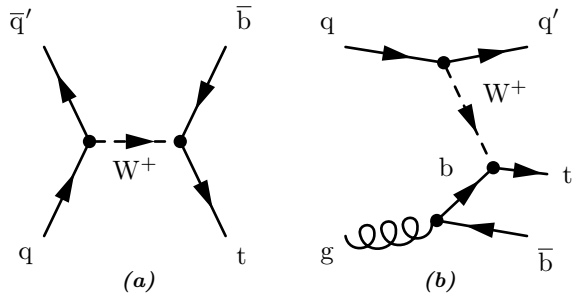


Figure 1.11: *The dominant contributions to single top quark production at the Tevatron: (a) s-channel and (b) t-channel.*

The search for a single-top signal was performed on 0.9 fb^{-1} of Run II data recorded between 2002 and 2005 and focused on final states containing (i) one high transverse momentum (p_{T}) lepton (electron with $p_{\text{T}} > 15$ GeV/c and $|\eta| < 1.1$ or muon with $p_{\text{T}} > 18$ GeV/c and $|\eta| < 2.0$), (ii) significant missing transverse energy ($15 < \not\!\!E_{\text{T}} < 200$ GeV/c) and (iii) a b-jet, all coming from the decay of the top quark. One or more additional jets are required, both to match the jets produced in association with the top quark and to allow for initial- and final-state radiation.

Events are triggered on a jet and a lepton, and selected requiring two, three or four jets. Jets are reconstructed using a cone jet algorithm with cone size $\mathcal{R}_{\text{cone}} = 0.5$, for details on D0 jet reconstruction please refer to section 3.3. The leading-p_{T} jet with $p_{\text{T}} > 25$ GeV/c and $|\eta| < 2.5$, the second-leading jet with $p_{\text{T}} > 20$ GeV/c and $|\eta| < 3.4$ and all subsequent jets with $p_{\text{T}} > 15$ GeV/c and $|\eta| < 3.4$.

A neural network b-jet tagging algorithm (see also section 3.3) is used to identify jets from b-quarks to enhance the signal content of the selected sample. Events in which the missing transverse energy is aligned with one of the selected objects are considered to be misreconstructed and removed.

The uncertainty on the background is dominated by the normalisation of the Monte Carlo background predictions to data and is approximately 18%. Since the expected uncertainty on the background event yield is larger than the expected number of single top quark events, a traditional counting experiment does not provide sufficient sensitivity. Instead a multivariate analysis technique using boosted decision trees [42, 43] is used to distinguish between signal and background. The decision tree method is a machine learning algorithm based on rooted binary trees which iteratively applies cuts to each event in a sample, resulting in a per-event signal probability or

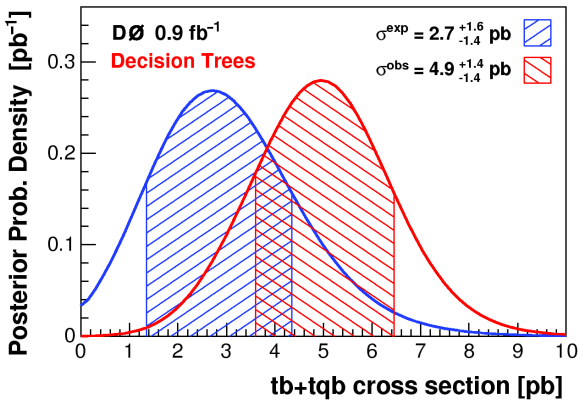


Figure 1.12: *The expected standard model and Bayesian posterior probability densities for the combined $s+t$-channel cross section analysis using boosted decision trees [45]. The standard model expectation was estimated using ensembles of pseudo-experiments.*

purity. The main difference with a traditional cut-based selection lies therein that subsequent cuts are still applied to events that have already failed one or more cuts. This classifies events in sets that pass and/or fail all possible combinations of cuts, assigning to each set an appropriate signal purity.

A Bayesian approach [44] is used to measure the single top quark production cross section. A binned likelihood is formed multiplying all possibilities (electron or muon, two, three or four jets, one or two b-jets) of the decision tree discriminant. A Poisson distribution is assumed for the signal counts and a flat non-negative prior for the signal cross sections. Systematic uncertainties are treated by integrating over Gaussian priors for each uncertainty. The posterior probability density is computed as a function of the assumed cross section. Figure 1.12 shows the expected and observed posterior densities for the decision tree analysis in the combined $s+t$ channel (figure 1.11).

At the end of 2006, the D0 collaboration presented first evidence for the production of single top quarks at the Tevatron collider with a combined $s+t$-channel cross section $\sigma(\mathrm{p\overline{p}} \to \mathrm{tb}+X) = 4.9 \pm 1.4$ pb [46].

At the Tevatron center-of-mass energy of $\sqrt{s} = 1.96$ TeV and assuming a top quark mass of $m_\mathrm{t} = 175$ GeV/c^2 the next-to-leading order (NLO) predictions for the production cross sections of the s and the t-channels are (1.98 ± 0.25) pb and (0.88 ± 0.11) pb respectively [47, 48].

In addition to the decision tree method described above, the search for single top events was also performed using two alternative multivariate techniques: a matrix element method similar to the approach described for the $t\bar{t}$ mass analyses (section 1.3.1), and a Bayesian neural network. The results are $4.8^{+1.6}_{-1.4}$ pb and $4.4^{+1.6}_{-1.4}$ pb respectively. Combination of all three measurements using a 'best linear unbiased estimator' (BLUE) method [49] gives a cross section of $\sigma = 4.7 \pm 1.3$ pb [45]; a 3.6-standard-deviation significance.

The D0 measurement of $\sigma = 4.7 \pm 1.3$ pb is in good agreement with the standard model expectation of 3.0 ± 1.3 pb (uncertainties estimated using ensemble test include effects of the methods).

Chapter 2

The Tevatron and the D0 detector

The field of experimental particle physics is the study of subatomic particles and their interactions. Since many of these particle do not occur freely in nature, the first step is to create them in a controlled laboratory environment. This makes use of Einstein's famous relation $E = mc^2$ by accelerating common particles like hydrogen nuclei to high energies and colliding them, converting the energy of these 'beam' particles into the masses of new particles.

This chapter takes a closer look at the Tevatron collider at Fermilab, which collides protons and antiprotons, and one of its detectors: D0, with which the data for this analysis was taken.

2.1 The Fermilab Tevatron collider

The Tevatron is a circular proton-antiproton collider located at Fermilab, near Chicago, in the USA. With a center-of-mass energy of $\sqrt{s} = 1.96$ TeV it is currently the highest-energy collider in the world. This section describes the Tevatron in some more detail, as well as the pre-accelerators supporting it.

2.1.1 A brief history

Commissioned in 1983, the Tevatron delivered a proton beam to several fixed target experiments, reaching a beam energy of 800 GeV/c in early 1984. It was the successor of the Main Ring, which up till then had been delivering a 400 GeV/c proton beam. In 1985 the Tevatron first ran in collider mode, colliding protons and antiprotons with a center-of-mass energy of $\sqrt{s} = 1.8$ TeV. At that time the 'Collider Detector at Fermilab' (CDF) [50] was the only experimental facility in the Tevatron. In 1992, at the start of Run I, CDF was joined by another detector: D0 [51]. In 1995 both the CDF and D0 collaborations announced the discovery of the top quark [52, 53], sought

for since the discovery of the bottom quark in 1977 [54]. Run I ends in 1996 and after extensive accelerator and detector upgrades a new data-taking period (Run II) starts in 2001. During the March 2006 shutdown the D0 detector is upgraded with improved tracking capabilities [55] and Level 1 calorimeter trigger electronics [56]. This shutdown marks the separation between Run IIa and Run IIb. The data for the analysis presented here was taken during Run IIa, between July 2002 and February 2006.

2.1.2 The accelerator chain

The Tevatron collider is preceded by a chain of pre-accelerators supplying beam particles to the Tevatron as well as to several fixed-target experiments (figure 2.1). The very first step in the Fermilab accelerator chain is the proton source. Hydrogen gas is fed into a Cockroft-Walton generator. There, electrons are added to the hydrogen atoms and the negative ions are accelerated to 750 keV before being injected into the linear accelerator or linac. The linac accelerates these ions further to an energy of 450 MeV by pushing them along on a high frequency electromagnetic wave. At the end of the linac the negative ions are bent into the booster, a circular accelerator (synchrotron) of 475 m circumference, and stripped of their electrons. The fact that new ions entering the booster have a negative charge, whereas the already present ions are positively charged makes it possible to 'wrap' each pulse of ions from the linac around the booster several times. The booster accelerates the, now positive, ions to 8 GeV/c. From here on the protons enter the main injector and can follow either one of two paths. They are either accelerated to 150 GeV/c and inserted into the Tevatron ring, or accelerated to 120 GeV/c for the fixed-target experiments or for antiproton production.

For the production of antiprotons, every 1.5 seconds part of the 120 GeV/c protons are focused onto a nickel alloy target. Interactions between the protons and the target material produce a plethora of secondary particles, including antiprotons. These secondary particles are focused using a lithium lens, a one-centimetre-diameter, ten centimetre long lithium rod pulsed with a high current to generate a strong magnetic field pointing radially inwards. Antiprotons around 8 GeV/c are selected and extracted using a pulsed magnet acting as charge-mass spectrometer. Approximately 20 suitable antiprotons are created for each million incoming protons. The incoming proton pulses from the main injector produce antiprotons in bunches. The debuncher accelerator spreads the bunches, producing a continuous particle beam while at the same time

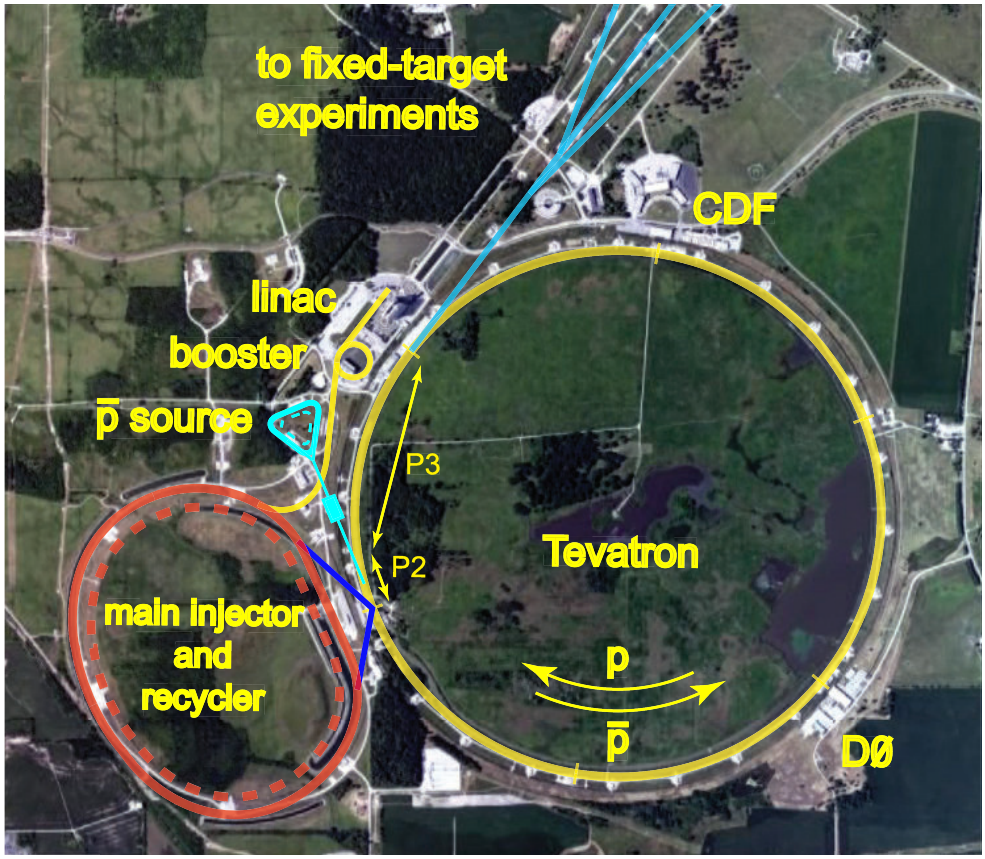


Figure 2.1: *Overview of the Fermilab accelerator chain. The 'triangular ring' houses the debuncher (outer ring) and the accumulator (inner ring). The target station is shown in the beamline between the Tevatron ring and the antiproton source. The P2 and P3 transfer lines connecting the Main Injector to the antiproton source and to the fixed-target switch-yard share the tunnel with the Tevatron (photo courtesy of Fermilab Visual Media Services).*

evening out the antiproton energies. This makes the antiproton beam easier to accept for the downstream accelerators. The debunching step takes approximately 100 milliseconds. The remaining time before the next proton pulse is used to 'cool' the antiprotons: their momenta perpendicular to the beam direction are reduced to create a narrower, more focused beam. After debunching and cooling successive series of antiprotons are 'stacked' in the accumulator. The accumulator can hold the antiproton beam for many hours or even days, until enough antiprotons are collected to be injected into the Tevatron to reach nominal luminosity.

Next, 150 GeV/c protons from the main injector are inserted into the Tevatron. Antiprotons from the accumulator are extracted into the main injector, accelerated to 150 GeV/c and injected in opposite direction into the Tevatron.

The Tevatron is the last, and most powerful, step in the Fermilab accelerator chain. It is a circular collider using helium-cooled superconducting dipole magnets at 4.3 K to bend the (anti)protons around the Tevatron ring. The Tevatron is not only the highest-energy collider in the world, it also has one of the world's largest cryogenics systems, with a total of more than 16000 hp of helium compressors. The dipoles employ a so-called 'warm-iron' design which keeps the iron magnet yoke outside the magnet cryostat. Driven by the limited available space in the main tunnel, in this design the yoke is separated substantially from the magnet coil, contributing only $\approx 10\%$ to the magnetic field. At the same time, however, it ensures the field strength depends almost linearly on the current, avoiding higher-order terms. This unlike 'cold-iron' magnets, which often contain current-dependent sextupole and decapole terms in the magnetic field [57]. After both protons and antiprotons have been inserted the beams are accelerated and cleaned to remove stray particles.

Finally, the beams are focused to collide in the interaction points. Particle collisions and beam interactions diminish the beam intensities. When enough new antiprotons are available the remains of the proton beam are dumped, the remaining antiprotons are recycled for the next fill, and the process starts again.

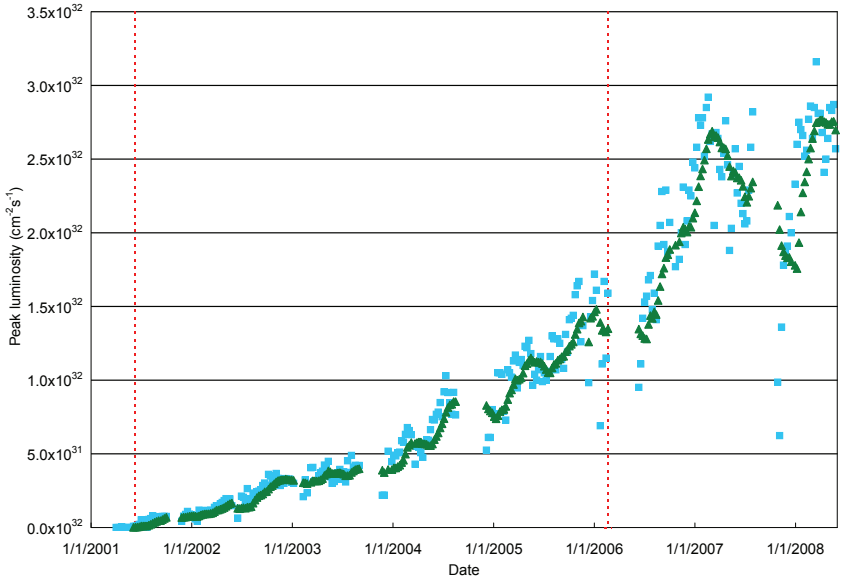
2.1.3 Tevatron performance: delivered luminosity

An important figure of merit for accelerator performance is the luminosity, describing the number of particles delivered per unit time and how well they are focused. The rate of physics events is directly related to the luminosity. The Tevatron luminosity is limited by the available amount of antiprotons. During the course of a store (the period the beams are kept colliding) the luminosity drops as the beams loose intensity.

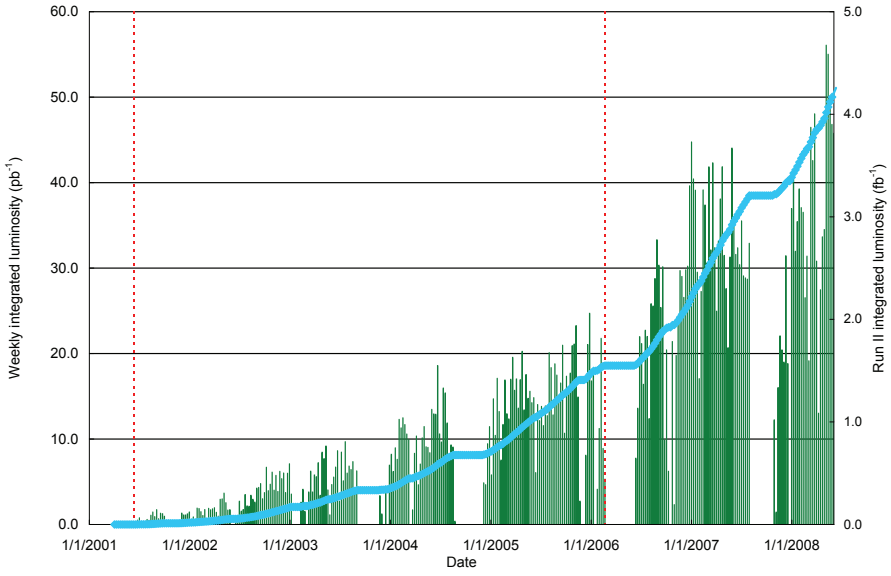
At high luminosity, at the beginning of a store, the luminosity decrease is dominated ($\approx$ 80% [58]) by beam depletion due to particle collisions. At lower luminosities dilution due to beam kinematics becomes increasingly important. Beam effects are dominated by broadening due to intra-beam scattering (especially for the proton beam, which is approximately ten times denser than the antiproton beam) and beam-beam interactions (apart from the two interaction points for the D0 and CDF experiments, the Run II Tevatron lattice contains 70 parasitic 'near interaction regions' [59]).

(Anti)protons in the Tevatron are grouped in bunches of $\approx$ 38 cm long, determined by the Tevatron radio-frequency (RF) acceleration system. A single turn around the Tevatron ring contains 1113 RF buckets with 18.8 ns separations between them. The buckets are grouped into 159 ticks of 7 buckets each. The 132 ns tick duration is the fundamental time unit for all Tevatron operations. Only the first bucket of each tick can contain a particle bunch. During normal collider operation there are 36 bunches of both protons and antiprotons in the Tevatron ring, grouped into three 'superbunches'. Within a superbunch the individual bunches are separated by two empty ticks (396 ns) while the superbunch trains are separated by abort gaps of 17 empty ticks ($\approx$ 2.5 µs). These gaps are required for the ramping time of the Tevatron abort system and are used by the experiments to synchronise electronics and take data using non-beam related (e.g. cosmic) triggers. An important side-effect of this bunch structure is that the two collider experiments, D0 and CDF, see different proton-antiproton bunch crossings. Any given proton bunch collides with different antiproton bunches at both experiments, leading to potentially different instantaneous luminosities between the two experiments (the determination of the luminosity at the D0 interaction point will be discussed in more detail in section 2.3.4).

After an arduous start of Run II, the performance of the Tevatron has been steadily improving. Figure 2.2a shows the development of the Tevatron peak instantaneous luminosity in Run II. Figure 2.2b shows the development of the delivered luminosity integrated over time, together with the amount of data recorded by the D0 experiment. The current Tevatron run (Run II) is expected to continue till 2009 and should result in 4 to 9 fb^{-1} of recorded luminosity [61] depending on the performance of the Tevatron and the efficiency of the CDF and D0 experiments.



(a)



(b)

Figure 2.2: *Development of Tevatron (a) peak luminosity and (b) integrated luminosity during Run II [60]. The dashed vertical lines indicate the data-taking period corresponding to the analysis in this thesis.*

2.2 Particle detection

Particle detectors can be divided into two basic groups: tracking detectors and calorimeters. Tracking detectors follow the trajectory of charged particles. [1] Combined with a magnetic field bending the particles, tracking detectors provide information on the particle momenta and charges. Calorimeters measure the energy deposited by particles traversing matter and provide important information for particle identification.

This section briefly discusses the basics of these two detector types and how each is implemented in the D0 detector. The analysis described in this thesis concerns a hadronic measurement: the relevant physics objects are calorimeter jets, tracking information is only used indirectly (e.g. in the identification of jets from b-quarks). Consequently, the main focus lies on calorimetric particle detection. Good sources of information on particle detection include references [62] and [63]. A good reference on the physics of calorimetry can be found in ref. [64]. A more concise overview is given in the Review of Particle Physics [11].

2.2.1 Tracking

Charged particles ionize the material they encounter. In the presence of an electric field the produced charges can be made to drift, inducing a detectable current. Depending on the design and technology used, tracking detectors provide 1D information (e.g. wire chambers and silicon strip detectors), 2D information (e.g. pixel detectors) or full 3D path information as in time projection chambers. Additional constraints like timing information can be used to extract 2D information from e.g. 1D hits along a wire. If a charged particle is made to traverse a series of detectors, it will leave a pattern of hits from which its path can be reconstructed. In the presence of a uniform magnetic field a charged particle will follow a curve with a radius proportional to its momentum and a direction determined by the sign of its charge. Apart from a very precise momentum measurement, this also allows for particle identification based on the ratio of energy to momentum E/p. Typical tracking detectors consist of many thin layers of active material, separated by as much empty space as possible to pose as little as possible disturbance of the particle paths. Among the main goals of tracking

[1] The average energy loss per unit distance dE/dx depends on a charged particle's speed. Simultaneous measurement of the momentum and dE/dx determination allows particle identification in tracking detectors based on determination of the particle mass. This approach works best in the non-relativistic energy range where energy loss is dominated by ionization. Good dE/dx resolution requires long tracks with many samplings to average out the large fluctuations on individual dE/dx measurement.

detectors are the reconstruction of tracks required for vertex reconstruction, and the matching with other (tracking) detectors. Precise information on the positions of vertices is paramount to locate the origin of the hard interaction, for the identification of heavy particle decays and for particle lifetime measurements. The innate thinness of tracking detectors allows for placement close to the interaction point, providing high precision position measurements.

2.2.2 Calorimetry

Calorimetry is the energy measurement of particles by stopping them, thereby absorbing all their energy, and generating a detectable signal proportional to the absorbed energy. While stopping particles allows for the measurement of their energies, characteristics of the energy depositions can be used in particle identification.

Unlike tracking, calorimetry is a destructive measurement. In tracking detectors the momentum resolution decreases with increasing energy due to straightening of the tracks. In contrast, the relative energy resolution in calorimetry improves with increasing energy. The calorimetric energy resolution is intrinsically limited by statistical fluctuation in the shower development which become less important at higher particle energies. In contrast to tracking detectors, in which as little material as possible is used to avoid disturbing the free path of the particles, calorimeters employ high density absorber materials to increase the stopping power. Sampling calorimeters consist of layers of high density absorber material interleaved with gaps containing active detector material to sample the developing energy depositions at various intervals.

Energy deposition

The dominant processes through which particles deposit energy in a calorimeter are scintillation and ionization. The electromagnetic interaction of particles with the Coulomb fields around the nuclei in the absorber material excites the matter constituents. When the excited atoms fall back scintillation (fluorescence) photons are emitted. This process is exploited in scintillation calorimeters in which the active medium is transparent (e.g. scintillating crystals) and photomultipliers are used to readout the generated light. In general a fraction of the emitted photons are lost in the absorber. Ionization occurs when electrons are knocked completely free from their nuclear orbits. The free electrons can be collected by applying an electric field,

causing them to drift, inducing a detectable current. To prevent electrons from being recaptured before reaching the electrodes the mean free electron path should be larger than the electrode separation. This leads to the choice of noble gases/liquids as the active material in many calorimeters. In the case of charged particles also bremsstrahlung due to deflection of incoming particles in the nuclear Coulomb fields plays a role and can lead to the emission of large numbers of photons. The energy loss from bremsstrahlung scales with the inverse of the square of a particle's mass and is in practice only relevant for electrons and very energetic muons.

The exact processes by which a particle deposits energy when traversing matter strongly depends on the type of particle, its energy and the density of the absorber material. Due to the different nature of the interactions involved, electrons and photons show significantly different behaviour than hadrons.

A particle entering a piece of material will initiate a cascade of interactions of decreasing energy. Secondary particles from these interactions will themselves undergo interactions, leading to a broadening 'shower' of particles in the material. The depth of the shower is governed by the interaction length (for nuclear interactions) and the radiation length (for electromagnetic interactions) of the material. Its width is determined by the material's Molière radius.

Electromagnetic showers

Energetic photons can split into e^+e^- pairs. This is the dominant process for photons above $E \approx 5$ MeV/c in uranium. Most photons of intermediate energies ($1 \lesssim E \lesssim 5$ GeV/c in uranium) will, upon entering a calorimeter, first undergo some Compton scatters, ionizing molecules by knocking electrons out of their bound states. When the photon energy has dropped below several hundred keV/c it is most likely to be captured by an atom which subsequently emits an electron (the photo-electric effect). For low energy photons also Rayleigh scattering plays a role: the deflection of photons by atomic electrons. Hereby photons do not loose any energy, only the spatial distribution of the deposited energy is influenced, broadening the showers.

When an electron with an energy of several GeV/c enters the calorimeter it will emit a shower of many thousands of bremsstrahlung photons. Energetic photons create electron-positron pairs, which in turn radiate bremsstrahlung photons themselves. In this way an electromagnetic shower develops in the absorber material. Once the average particle energy drops below the pair creation threshold, Compton scattering will take over as the dominant interaction process and the shower will stop growing.

The depth in the absorber at which this turnover occurs is called the electromagnetic shower maximum.

The shower maximum provides the best lateral depth at which to measure the shower position. Whereas low energy electrons are most likely to encounter ionized atoms in the bulk of the absorber and be captured, low energy positrons will annihilate with electrons, resulting in low energy photon pairs. This annihilation component makes electromagnetic shower development strongly dependent on the electron density in the absorber material.

Muons behave completely differently from their lighter counterparts, electrons (figure 2.3). Due to their large mass ($(m_\mu/m_\mathrm{e})^2 \approx 40000$) bremsstrahlung plays no significant role for all but very high energy muons. At very low energies the dominant muon energy loss process is due to collisions with the nuclei of the absorber material. For most of the relevant energy range muons only loose energy due to ionization. This effect decreases rapidly ($\sim 1/E^2$) with increasing muon energy. At the same time the maximum energy that can be transferred to a free electron in a single collision rises. The combination of these two effect results in a minimum in the energy loss curve. The energy corresponding to this turnover point is called the point of minimum ionization and such muons are called minimum ionizing particles (MIPs). Only at very high energies does bremsstrahlung come into play as an energy loss process. Figure 2.3 shows the contributions of the different energy loss processes as a function of muon energy for (anti)muons in a copper absorber. For most practical purposes muons can be considered MIPs, only depositing a few GeV/c in the traversal of a typical calorimeter.

Hadronic showers

While in EM showers all energy is eventually used to ionize the detector material, in hadronic showers some fraction of the incoming energy is always lost (dissipated as heat) in the absorber material and thus undetectable. Charged hadrons behave much like heavy muons, ionizing the detector material, until they encounter a nucleus. The strong interaction between an incoming hadron and a nucleus can change both the incoming hadron as well as the nucleus, leading to new hadrons (predominantly light hadrons like pions) and creating a hadronic shower. Since neutral hadrons cannot ionize the material, they only loose energy through strong nuclear interactions. Hadronic shower development is governed by strong interactions instead of by electromagnetic interactions. Due to the lower cross section for nuclear interactions hadronic showers

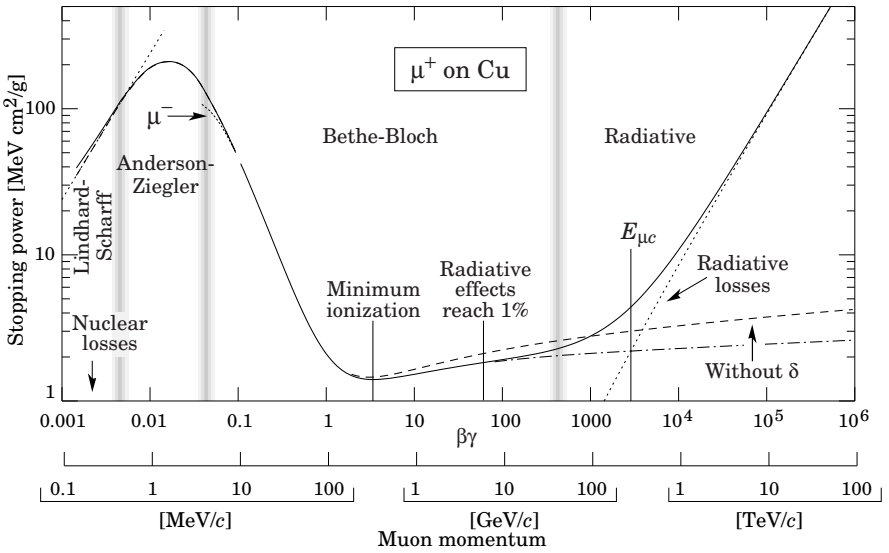


Figure 2.3: *Energy loss curve for muons impinging on copper showing the stopping power or average energy loss per unit length* $- < dE/dx >$ *[65]. Due to the high muon mass radiative losses (bremsstrahlung) only become important for muons with energies over* ≈ 300 *GeV/c.*

tend to develop deeper inside the material than EM showers. Some hadrons may decay electromagnetically (e.g. $\pi^0 \rightarrow \gamma\gamma$). This leads to the development of an electromagnetic component to hadronic showers. Since any neutral pions produced will most likely decay before initiating any nuclear interactions, the fraction of a shower's energy contained in the EM component increases with increasing shower depth. The EM shower fraction will tend to grow until the hadronic energy drops below the pion production threshold, implying that the higher the energy of the incoming hadron is, the larger the EM shower fraction. The time scale for the hadronic shower part is dominated by the thermalization of the neutrons produced in the nuclear interaction, much longer than the time scale for the electromagnetic part of the shower.

A side-effect of the hadron-nucleus collisions is spallation. When an incoming hadron collides quasi-freely with a nucleus, the target nucleons can initiate secondary collisions inside the nucleus leading to a cascade of fast nucleons. It is possible that in this step new hadrons are formed that, when energetic enough, can escape from the nucleus. The disturbed nucleus will emit any excess energy in the form of protons, neutrons and (to a lesser extent) α particles and photons. The nuclear binding energy

required to break up the nucleus (as well as the energy of any neutrinos created in possible radioactive decays) is undetectable. This presents an intrinsic limitation to the precision of hadronic, as compared to electromagnetic, calorimeters.

Another effect of the nuclear interactions is the creation of nuclei with long half-lives. In addition interactions of neutrons with the absorber material can activate the absorber. These effects lead to increased noise due to long-term radioactive processes uncorrelated with beam activity. [2]

Intrinsic non-linearities Whereas the energy response for EM showers is mainly influenced by instrumental effects, hadronic calorimeters suffer from two intrinsic, physical non-linearities. Due to the missing energy dissipated in hadronic showers the hadronic response is intrinsically lower than that for EM particles, commonly expressed as the ratio of pion to electron response: $\pi/e < 1$. Since the EM fraction increases with increasing energy the π/e ratio increases towards unity for high energy pions. The real difference in hadronic versus electromagnetic response is often denoted as e/h: the ratio of the average electron response to that of the purely hadronic part of an average pion shower.

2.3 The D0 detector at Fermilab

The D0 detector is a general purpose detector designed for the study of high energy proton-antiproton collisions. It is located at the D0 interaction point in the Fermilab Tevatron collider ring as shown in figure 2.1. Figure 2.4 shows a schematic view of the D0 detector. The z-axis is defined along the proton beam direction and the transverse (xy)-plane as the plane perpendicular to the beam direction. The azimuthal angle ϕ defines directions in the transverse plane. With unpolarised beams only differences in azimuthal angle are relevant. Many processes depend on the angle θ with respect to the incoming beam, and a common variable is the pseudorapidity η defined as $\eta \equiv 1/2 \tan(\theta/2)$. Zero pseudorapidity corresponds to the plane perpendicular to the beam-line through the interaction point. Closer to the beam direction, pseudorapidity grows towards positive (negative) infinity on the south (north) side of the detector.

[2] Because of its high density (depleted) uranium is a popular choice of absorber material. Fast neutrons activate the uranium-238 to uranium-239 (with a half life of slightly less than half an hour) which decays into neptunium: ${}^{238}_{92}\mathrm{U} + \mathrm{n} \rightarrow {}^{239}_{92}\mathrm{U} \rightarrow {}^{239}_{93}\mathrm{Np} + \mathrm{e}^- + \overline{\nu}_\mathrm{e}$. The neptunium in turn decays into plutonium and back to uranium(-235): ${}^{239}_{93}\mathrm{Np} \rightarrow {}^{239}_{94}\mathrm{Pu} + \mathrm{e}^- + \overline{\nu}_\mathrm{e}$, ${}^{239}_{94}\mathrm{Pu} \rightarrow {}^{235}_{92}\mathrm{U} + \alpha$. Uranium-235 follows the actinium series and eventually results in stable lead-207.

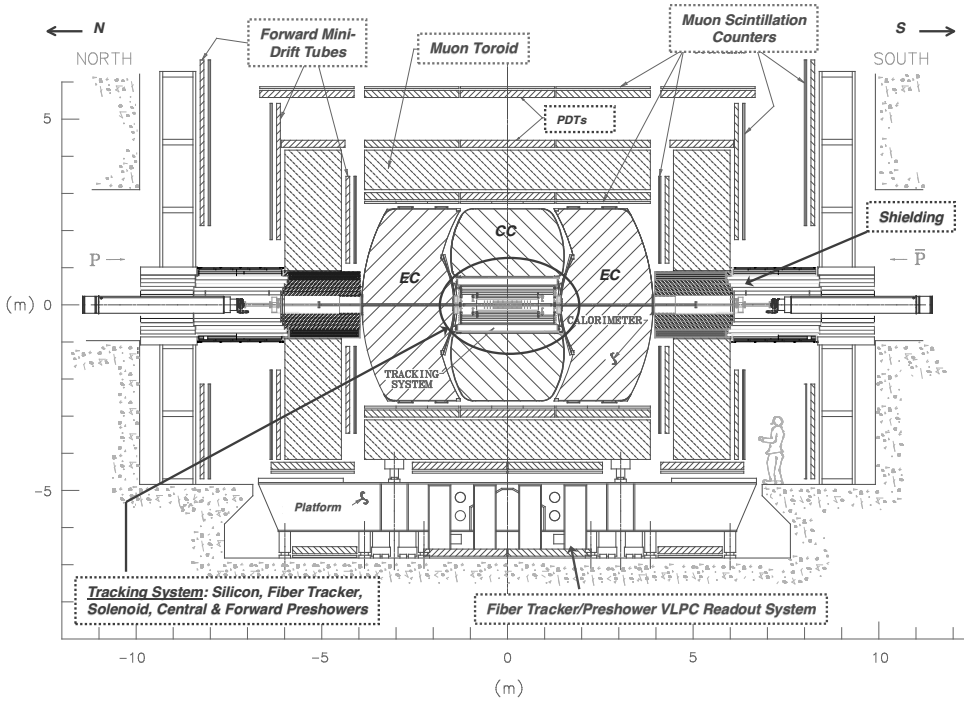


Figure 2.4: *Schematic view of the D0 detector [65]. Located inside the ellipse marked 'Tracking System' are the Silicon Microstrip Tracker (SMT), the Central Fiber Tracker (CFT), the preshowers and the solenoid. Surrounding the tracking system are the central and endcap calorimeters (CC and EC) and the muon system. Protons move north-to-south, antiprotons south-to-north. The proton direction defines the z-axis of the right-handed D0 coordinate system.*

Regions with low (high) $|\eta|$ are called central (forward). Real rapidity takes into account non-zero masses and is defined as $y \equiv 1/2 \ln((E + p_z)/(E - p_z))$.

This section describes the upgraded D0 detector as it was operating during Tevatron Run IIa. The design of the D0 detector follows the common cylindrical layout of many general purpose particle detectors. A precision tracking system surrounds the interaction region, operating in a solenoidal magnetic field. The central tracker is surrounded by a large, segmented calorimeter for energy measurement and particle identification, followed by a muon tracking system with a dedicated toroid. The following sections will discuss each of these subdetectors in detail, as well as the trigger system and the luminosity monitor. Focus lies on those detector parts most relevant to

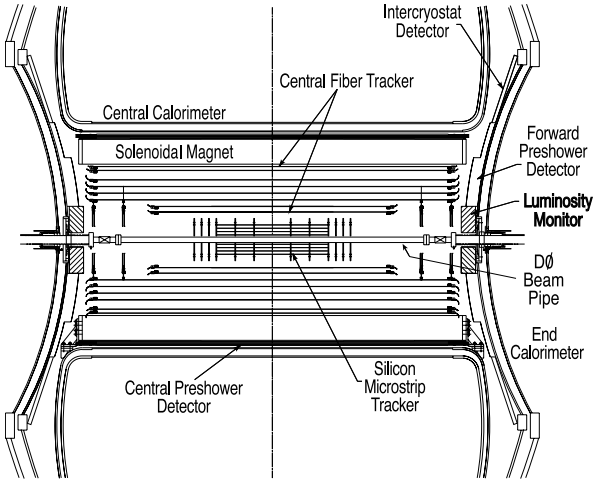


Figure 2.5: *Longitudinal cross section of the D0 central tracking system [65]. The Silicon Microstrip Tracker, centred around the beam pipe, is surrounded by the Central Fiber Tracker. Also shown are parts of the calorimeter system and the luminosity monitors.*

the analysis in this thesis. References [65] and [66] provide a more detailed description of the full D0 detector.

2.3.1 Central tracker

The central tracking detector consists of the Silicon Microstrip Tracker (SMT) closest to the beam pipe, the Central Fiber Tracker (CFT) surrounding the SMT, and a solenoid providing a 2 T magnetic field (figure 2.5).

Silicon Microstrip Tracker

The SMT provides both tracking and vertexing over almost the whole pseudorapidity coverage ($|\eta| \lesssim 3$) of the calorimeter and muon systems.

Due to the large extend of the D0 luminous region along the z-direction ($\sigma \approx 25$ cm) the SMT design combines both barrel-shaped and disk-shaped detectors in the central region, extended by larger disks in the forward regions (figure 2.6). In the central region the SMT consists of six barrel-shaped subdetectors, three on each side of the detector center. Each barrel consists of four concentric cylindrical layers of silicon detectors and is capped on the (high-$|z|$) outside with a disk-shaped subdetector called

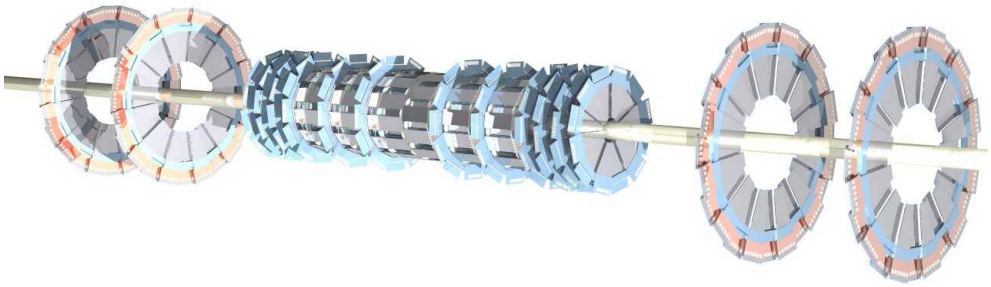


Figure 2.6: *Schematic view of the D0 Silicon Microstrip Tracker [65]. Shown in the middle are the six barrel subdetectors, separated and extended by twelve F-disks. On the outside are shown the four H-disks.*

an F-disk. On each of the outsides a unit of three additional F-disks is located. The most forward part of the SMT on each side consists of two large H-disks.

A detailed description of the design and geometry of the SMT can be found in ref. [67].

Central Fiber Tracker

Shown in figure 2.5, the Central Fiber Tracker (CFT) consists of eight concentric cylinders of 835 µm diameter scintillating fibers. The two innermost cylinders are 1.66 m long and fit between the two innermost SMT H-disks. The remaining six CFT cylinders are 2.52 m long and enclose the H-disks. The CFT provides tracking coverage up to $|\eta| \lesssim 1.7$.

Each CFT cylinder contains two doublets of fibers: one aligned along the beam direction and one at a 3° stereo angle to allow for the determination of hit positions along the z-direction. The scintillation light caused by charged particles traversing the fibers is transported through fiber waveguides and read out by photon counters residing in cryostats on the D0 readout platform underneath the detector.

The muon tracker

The largest of the D0 tracking subdetectors is the muon system, situated outside the calorimeters. Particles traversing the full calorimeter to be detected in the muon system will most likely be muons. Offline muon selection criteria reject the few fakes due to 'punch through' of hadrons. The muon system provides extensive coverage, up

to $|\eta| \approx 2$, interrupted only by the hole in the bottom region where the detector support structures leave no room for muon tracking chambers ($|\eta| < 1.25$, $4.25 < \phi < 5.15$).

The muon system consists of three layers of muon detectors: the A layer inside the toroid and the B and C layers outside. Each layer consists of three (B and C layers) or four (A layer) layers of rectangular drift tubes. For the Proportional Drift Tubes (PDTs) in the central muon system the typical drift time is ≈ 500 ns. In the forward muon system, consisting of Mini Drift Tubes (MDTs), the maximum drift time is ≈ 60 ns.

The central muon system includes two layers of scintillating material, inside the A layer and outside the C layer, that provide additional trigger information and aid in the rejection of out-of-time cosmic muons. Three layers of trigger scintillator counters are present in the forward muon system, segmented to match the CFT trigger sectors.

The preshower detectors

Intermediary between the tracking system and the calorimeter lie the preshower detectors. Preshower information aids in electron-photon separation and is used to improve spatial matching between central tracks and calorimeter energy clusters.

The preshower detectors consist of layers of interleaved triangular strips of scintillating material. Inside each strip a wavelength-shifting fiber similar to the CFT fibers is embedded. The preshower detectors share the waveguides, readout and most of the signal path with the CFT. The Central Preshower detector (CPS) covers up to $|\eta| \lesssim 1.3$ and sits between the solenoid and the central calorimeter. It consists of three fiber layers. The Forward Preshower detector (FPS) is attached to the endcap calorimeters and covers the range $1.5 \lesssim |\eta| \lesssim 2.5$. The FPS consists of two stacks of two fiber layers each, separated by a lead-stainless-steel absorber with a thickness of two radiation lengths. Seen from the interaction region, the inner two layers followed by the absorber cover $1.65 \lesssim |\eta| \lesssim 2.5$ while the outer layers cover $1.5 \lesssim |\eta| \lesssim 2.5$ (the region $1.5 \lesssim |\eta| \lesssim 1.65$ lies in the 'shadow' of the solenoid and does not require additional absorber material in order to initiate showers). The inner layers, or MIP layers are meant to provide track position measurements. The outer layers, or shower layers, are located behind (forward of) the absorber and are meant to help separate electromagnetic particles that will readily shower in the absorber, from heavier particles that will only leave a second MIP signal in the shower layers.

Calorimeter region		Absorber
CC	EM	3 mm uranium
	FH	6 mm uranium-niobium (2%) alloy
	CH	46.5 mm copper
EC	EM	4 mm uranium
	FH	6 mm uranium-niobium (2%) alloy
	CH	46.5 mm stainless steel

Table 2.1: *The different absorber materials used in the D0 calorimeter.*

Whereas the Central Preshower information is used extensively in the electron/photon identification, due to problems with the performance of the FPS electronics its presence is only relevant in the interpretation of calorimeter information.

2.3.2 The D0 calorimeter system

The D0 calorimeter is a sampling calorimeter using liquid argon as the active material, interspersed with plates of different absorbers. Showers develop in the dense absorber material and are sampled intermittently in the liquid argon gaps. To allow access to the inner detector the calorimeter has been designed as a central calorimeter (CC) which covers $|\eta| \lesssim 1.1$ and a mirror-pair of endcap calorimeters (ECs) extending coverage to $|\eta| \lesssim 4.2$ Both CC and ECs consist of an inner electromagnetic (EM) layer, followed by fine hadronic (FH) and coarse hadronic (CH) layers. The absorber material and thickness used depend on the calorimeter region (table 2.1).

Segmentation: cells and towers

The readout unit of the D0 calorimeter is a cell (figure 2.7): the combination of a liquid argon gap, an absorber plate and a readout board. The readout boards consist of pairs of G10 circuit board glued together and coated on the outside with high-resistivity carbon-loaded epoxy [3]. The absorber plates are grounded and a high voltage (typically

[3] Strictly speaking this is where the actual 'smallest readout units' are defined. The (copper-clad) inside of one of the G10 boards is milled into pads representing the desired segmentation pattern. These pads are the real smallest readout units but they are 'ganged together' in hardware to form calorimeter cells.

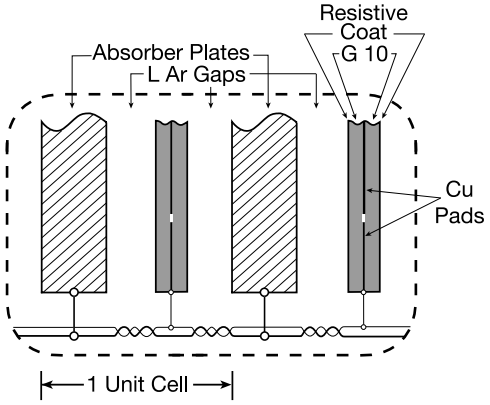


Figure 2.7: *Schematic view of a D0 calorimeter cell [66]. The absorber and readout plates are separated by* 2.3 *mm liquid-argon gaps, resulting in a typical drift time of approximately* 430 *ns.*

2 kV) is applied to the readout plates, thus creating a drift field for ionization charges in the 2.3 mm argon gap.

The bulk of the D0 calorimeter is segmented into calorimeter cells of approximately $\Delta\eta\times\Delta\phi = 0.1\times0.1$ in size. The third EM layer (starting at $\approx 4.0\ X_0$ and $\approx 6.8\ X_0$ thick in the CC), around the approximate electromagnetic shower maximum, is subdivided into cells of $\Delta\eta \times \Delta\phi = 0.05 \times 0.05$. The very forward cells in the ECs ($|\eta| \gtrsim 3.2$) have a fixed, coarser ϕ-segmentation: $\Delta\phi = 0.2$ while the η-size increases with increasing η. Cells at the same (η, ϕ) are grouped together into pseudo-projective calorimeter towers. The term 'pseudo-projective' expresses the fact that 'the *centres* of cells of increasing shower depth lie on rays projecting from the centre of the interaction region, but the cell boundaries are aligned perpendicular to the absorber plates' [66]. Figure 2.8 demonstrates this segmentation, together with several projective pseudorapidity lines. Please note that there is a small gap in the electromagnetic calorimeter between $1.2 < \eta < 1.3$.

Calorimeter electronics and trigger information

Calorimeter readout is achieved in three stages (figure 2.9). Charge preamplifiers mounted on the cryostats integrate the cell signals. The amplified signal is transported via twisted-pair cables to the BaseLine-subtraction and Storage (BLS) boards located underneath the cryostats. The BLS boards provide analog signal shaping and storage.

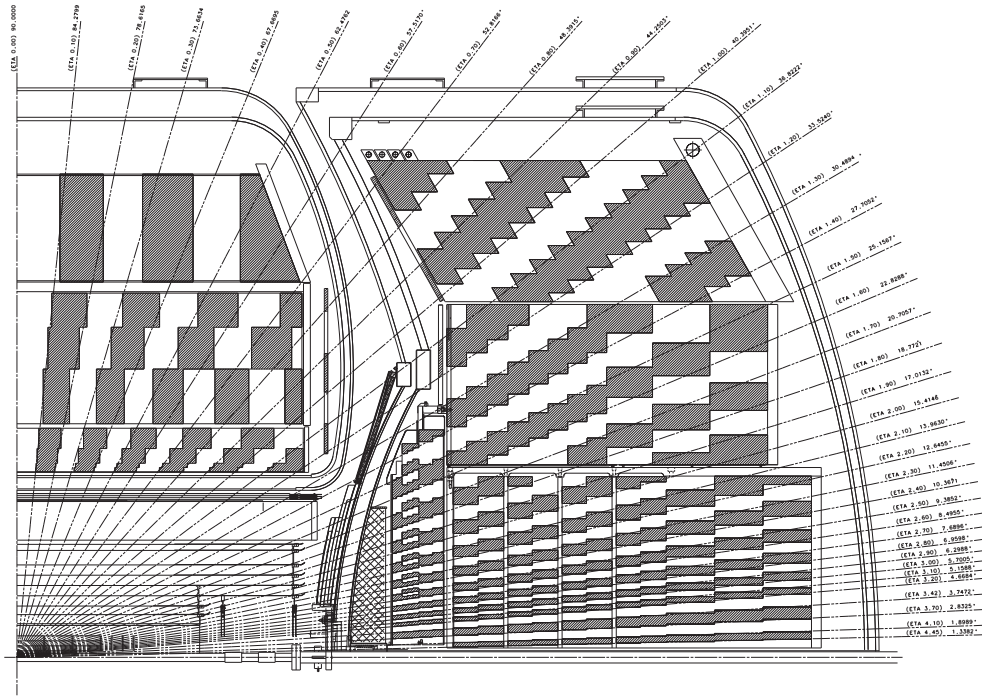


Figure 2.8: *Side-view of one quarter of the D0 calorimeter system [66]. The shading shows the segmentation into pseudo-projective towers.*

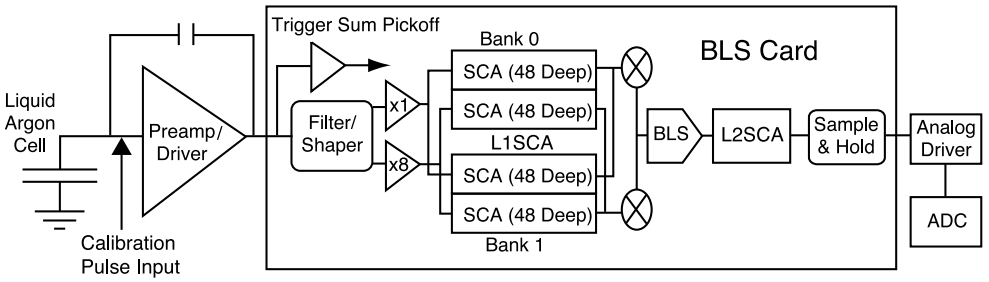


Figure 2.9: *Overview of the D0 calorimeter electronics [66].*

To suppress pile-up effects the amplified (integrated) signal is shaped into a narrower pulse. The rise time of the shaped signal is determined by the ≈ 430 ns drift time, the decay time is set by the RC time of the shaper: 250 ns. The resulting charge collection corresponds to approximately two-thirds of the charge read out from the calorimeter cell. The short integration time is mandatory to avoid subsequent bunch crossings to add pile-up to the interaction being read out. It represents one of the main challenges in Run II calorimeter and jet energy calibration. After signal shaping, two gain paths are available, maximising the readout dynamic range. The appropriate gain path for each channel is chosen automatically after baseline subtraction. Switched-capacitor arrays (SCAs) hold the signal for 4.2 µs awaiting a decision from the first level trigger (for details on the trigger system please refer to section 2.3.3). Upon receiving a Level 1 accept, the baseline value obtained from three times 132 ns before the current bunch crossing is subtracted and the data is held (for a maximum of 25 ms) in the Level 2 SCA. After receiving a Level 2 accept the signal is transferred to a sample-and-hold circuit and clocked out to the analog-to-digital converter.

The full, digitised calorimeter information is now available for use in the third level trigger. For the use of calorimeter information in the Level 1 and Level 2 trigger, special trigger towers are formed in the trigger electronics. The signal is extracted early in the electronics chain before shaping. Standard calorimeter towers are combined to create trigger towers of size $\Delta\eta \times \Delta\phi \approx 0.2 \times 0.2$. The larger trigger towers capture a large fraction of the showers, giving a better energy estimate for the trigger at the cost of the position resolution. The electromagnetic and fine hadronic parts are considered separately in the trigger. The Level 1 trigger signal is extracted before signal shaping. Due to its small sampling fraction and large capacitance, the coarse hadronic cells would contribute mostly noise.

Inter-cryostat detectors

The region between the central and endcap cryostats, $0.8 \lesssim |\eta| \lesssim 1.4$, represents a gap in the calorimeter coverage. In addition, a significant amount of uninstrumented material in the form of cables, support structures and the cryostat walls is present in this $|\eta|$ range. To mitigate the effects of this calorimeter inefficiency, extra calorimeter cells, called massless gaps, were added inside the cryostats. Furthermore, Inter-Cryostat Detectors (ICDs) were added in the gaps between the CC and EC cryostats. Each ICD consists of a ring of scintillating tiles, 0.1×0.1 in $\Delta\eta \times \Delta\phi$ and is read out using photomultiplier tubes, covering the range $1.1 < |\eta| < 1.4$. Operation and calibration

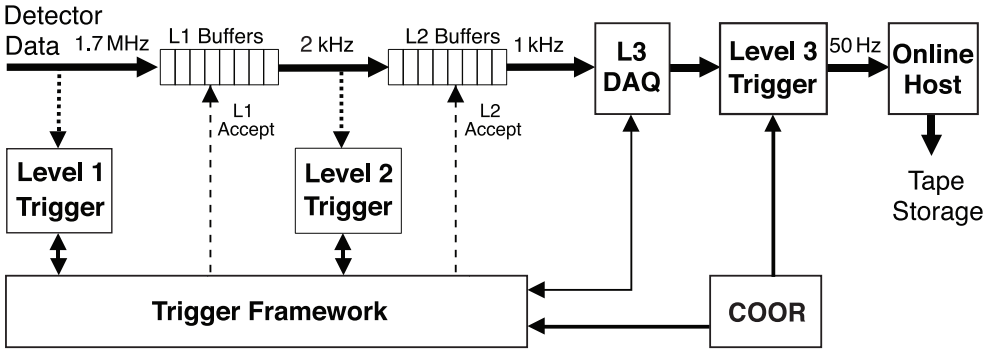


Figure 2.10: *Overview of the D0 trigger system [66]. The rates shown represent the design maxima.*

of the photomultipliers in the solenoid field, combined with the imperfect sampling in the inter-cryostat region (ICR), make the ICR a critical $|\eta|$ region in many physics analyses.

2.3.3 Trigger system

With a bunch crossing separation of 396 ns hard $p\overline{p}$ collisions occur in the Tevatron at an average rate of ≈ 1.7 MHz. The peak interaction rate within a bunch train is $1/(396\text{ ns}) = 2.5$ MHz. Taking into account the 17 empty bunches between successive trains shows the average interaction rate to be $2.5\text{ MHz} \times 36/53 = 1.7$ MHz. The D0 trigger system was designed to reduce the total collision rate to an acceptable rate for offline storage by selecting only events with interesting physics signatures. The specific trigger terms and assigned bandwidths are determined by the physics focus of the collaboration and have changed over time. Both practical and financial considerations make it impossible to study each and every hard interaction.

The D0 trigger system (figure 2.10) consists of three levels: Level 1, Level 2 and Level 3. Each of these levels selects events at a lower rate but in increasing detail. Events passing Level 3 selection are written to mass storage for full offline reconstruction and analysis.

The Level 1 trigger and the trigger framework

The first trigger level, Level 1, consists of dedicated hardware triggers based on a subset of all information from several subdetectors. L1Cal uses the EM and hadronic

trigger towers within $|\eta| < 3.2$ and the Level 1 central track trigger (L1CTT) and muon trigger (L1Muon) provide track trigger terms. The Level 1 output rate is ≈ 1.5 kHz.

The trigger framework (TFW) collects and combines trigger information from the individual Level 1 devices and makes the trigger decisions. It also provides a 'beam condition' for each of the 128 Level 1 trigger terms and takes care of trigger prescaling. Triggers with high accept rates are prescaled to fit within the available bandwidth. By randomly disabling the L1 trigger bit for x out of every n bunch crossings the rate is reduced by a factor $(n - x)/n$.

The Level 2 trigger

The Level 2 system combines the output from the Level 1 trigger with more detailed information from all subdetectors. It contains embedded processors on dedicated single-board computers (SBC's) for all subdetectors. At Level 2 rudimentary physics objects are built. A special global processor takes the outputs of these subdetector processors to check for correlations between trigger objects. Level 2 reduces the event rate to ≈ 700 Hz.

The Level 3 trigger

The third trigger layer, Level 3, is a fully programmable software trigger running on a processor farm. At Level 3 all information from the full detector is available and a simplified event reconstruction is performed. After input, output and unpacking roughly 200 ms per event is available for detailed analysis. Trigger lists and prescale settings are designed to provide a Level 3 accept rate around 50 Hz, dictated by the offline storage bandwidth. With falling luminosity, prescale sets are changed repeatedly to maintain this storage rate. Figure 2.11 shows the development of the D0 luminosity and trigger rates for a selected Tevatron run. As the luminosity drops, a new run is started with lower prescales to keep the L3 accept rate steady around 50 Hz.

2.3.4 Luminosity monitor

Many physics analyses require the precise determination of the absolute instantaneous and/or integrated luminosity. In absence of a crossing angle the instantaneous luminosity $\mathcal{L}$ is given by the number of particles that pass the luminous region per second

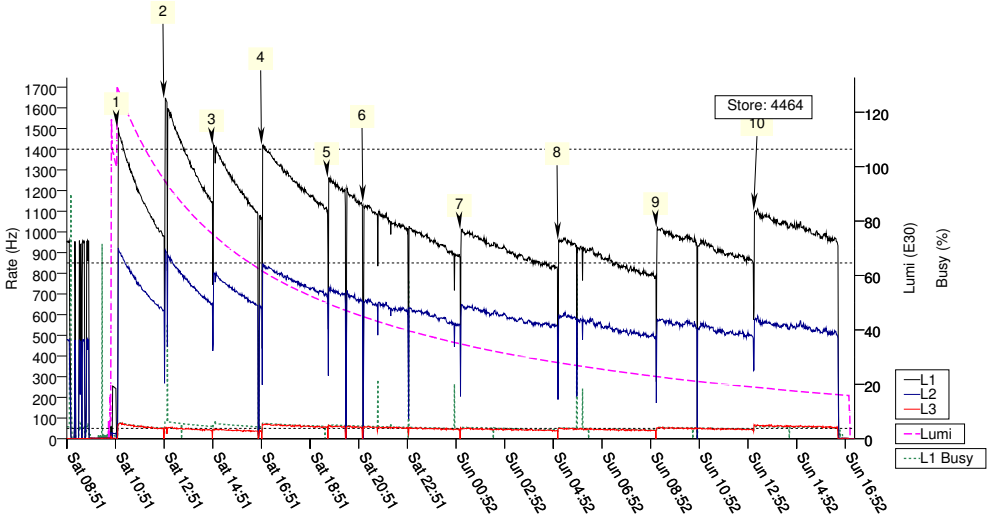


Figure 2.11: *Progression of D0 luminosity (dashed line) and trigger rates (solid lines) for Tevatron store number* 4464*, recorded October 22-23, 2005. The numbered arrows mark D0 run transitions, used to prescale and/or trigger settings to maximise physics output. The dotted line at the bottom showing occasional peaks represents the 'front-end busy' rate of the first trigger level. It shows a typical* $\approx 5\%$ *detector dead-time, slowly decreasing with decreasing luminosity. The double-peak structure just before the beginning of the first run shows the effects of the Tevatron beam cleaning. After initiation of collisions the beams are 'scraped' to remove the particle halos. Only after cleaning physics data-taking starts.*

divided by the transverse area of the interaction region:

$$\mathcal{L} = \frac{f B N_{\mathrm{p}} N_{\overline{\mathrm{p}}}}{2\pi \left(\sigma_{\mathrm{p}}^2 + \sigma_{\overline{\mathrm{p}}}^2\right)} F\left(\sigma_{\mathrm{l}}/\beta^{\star}\right).$$

Here f represents the Tevatron revolution frequency (47.713 kHz), $B = 36$ is the number of bunches per revolution, $N_{\mathrm{p}(\overline{\mathrm{p}})}$ the number of (anti)protons per bunch and $\sigma_{\mathrm{p}(\overline{\mathrm{p}})}$ the transverse spread of the (anti)proton bunches. F is a form factor depending on the longitudinal spread of the bunches σ_{l} and the value of the beta-function describing the accelerator tune at the interaction point $\beta^{\star}$. Determining the luminosity directly from the above equation requires detailed information about the beam dynamics that are not available to D0. An alternative technique is used,

relating the luminosity to a well-known cross section: the total inelastic cross section $\sigma_{\text{inelastic}}(1.96\text{ TeV}) = 60.7 \pm 2.4$ mb [68].

The D0 detector is equipped with a dedicated subdetector to measure the instantaneous Tevatron luminosity at the D0 interaction point. The luminosity monitor is also used to keep track of beam halo (radiated and/or stray particles around the beam) and to provide a fast estimate of the interaction z-positions for online monitoring.

The luminosity system consists of two circular arrays of 24 plastic scintillation counters each. These arrays are located on the outside of the endcap calorimeter cryostats and occupy the radial region between the beampipe and the Forward Preshower, covering the region $2.7 < |\eta| < 4.4$. The scintillators are read out using photomultipliers (PMTs) and are per twelve enclosed in light-tight enclosures. To avoid helium leaking from the Tevatron magnets to permeate and damage the PMTs these enclosures are constantly purged with dry nitrogen gas.

The luminosity is determined from the average numbers $\bar{N}_{\text{LM}}$ of inelastic $\text{p}\overline{\text{p}}$ interactions per beam crossing: $\mathcal{L} = \frac{f_{\text{beam}}\bar{N}_{\text{LM}}}{\sigma_{\text{LM}}}$ where f_{beam} is the beam crossing frequency and σ_{LM} represents the effective cross section (including acceptance and efficiency) for inelastic $\text{p}\overline{\text{p}}$ scattering in the D0 detector. The number of $\text{p}\overline{\text{p}}$ interactions per bunch crossing follows a Poisson distribution with mean μ. Depending on beam dynamics μ can vary strongly from crossing to crossing. To take this into account the instantaneous luminosities for all 36 different bunch crossings are measured independently.

The instantaneous luminosity is averaged over periods called luminosity blocks which are at most 60 s in duration to reduce the statistical uncertainty to $\ll 1\%$. Over the duration of one luminosity block the luminosity for each of the bunch crossings can be considered constant. Configuration changes always force the start of a new luminosity block. Luminosity blocks are marked with a unique luminosity block number or LBN.

Level 1 triggers may be correlated with the instantaneous luminosity. For example a high luminosity crossing is likely to trigger a high-p_{T} event, thus disabling the trigger for subsequent crossings due to readout dead time. For this reason it's important to determine the exact luminosity to which a given trigger is exposed. In D0 this is achieved by dividing triggers into exposure groups of triggers with common dead time and correcting the integrated luminosity for the fraction of ticks the exposure group was disabled. A separate correction is applied for dead time not related to the bunch-to-bunch luminosity fluctuations, like data-taking pauses and trigger prescales.

At many points in the trigger chain events can be lost due to e.g. readout problems, hardware failure or software problems. To keep track of the live time of the whole

trigger chain two special types of triggers are introduced: zero-bias (ZB), triggering on the presence of a bunch crossing, and minimum-bias (MB), triggering on a coincidence between the north and south luminosity counters signalling the presence of an inelastic collision. Neither of these trigger types can be rejected at any trigger level, but both are prescaled to an approximate accept rate of 1 Hz. Correcting for the difference in ZB and MB trigger rates between the first and third trigger levels, the recorded luminosity of a trigger n is defined as:

$$\mathcal{L}_{\text{recorded}}(n) = \frac{(N_{ZB} + N_{MB})_{\text{recorded}}}{(N_{ZB} + N_{MB})_{L1}} \mathcal{L}_{L1}(n)$$

where $\mathcal{L}_{L1}(n)$ represents the integrated luminosity to which L1 trigger n was exposed.

Chapter 3

Reconstruction and identification of physics objects

Physics objects like photons, electrons and jets are reconstructed from selected (sub)detector information. Subsequently, quality and identification criteria are applied to the reconstructed objects to reject fake or misreconstructed objects. Section 3.1 gives an overview of the algorithms used to reconstruct the primary (hard interaction) vertex and any secondary vertices used in the identification of jets from heavy quarks. Muons, used to identify b-jets, and electrons and photons, which are used in the jet energy calibration, are discussed in section 3.2. Jet reconstruction, as well as b-jet identification, is discussed in section 3.3 (jet energy calibration is the subject of chapter 4). Finally, section 3.4 briefly reviews the reconstruction of the missing transverse energy, one of the key ingredients for determining the jet energy scale.

Notation Jets are numbered in order of decreasing p_{T}, starting at one for the leading jet. This jet index is used as subscript for jet properties, e.g. $p_{\mathrm{T}\,1}$ for the leading jet transverse momentum. Jets from b-quarks (b-jets) and light jets are marked with subscripts b and l respectively. The notation in the axes labels of some figures deviates from the above to avoid 'sub-subscripts' and too small font sizes.

3.1 Tracks and vertices

Energy deposits from charged particles traversing the different layers of the central tracker are clustered across strips (fibers) in the SMT (CFT) to form hits. These hits are the input to the track reconstruction algorithms. The reconstructed tracks are used to reconstruct vertices.

3.1.1 Tracks

Due to the presence of the solenoidal magnetic field, charged particles follow helical trajectories through the detector. These helical tracks are parameterised using five parameters at the point of closest approach to the beam line. Two track parameters used in the following are 'distance of closest approach' (DCA), defined as the shortest distance between the track path and th primary vertex, and the 'impact parameter' DCA_{xy}, the DCA in the transverse plane.

Two independent tracking algorithms are used to reconstruct track candidates. The histogram track finder (HTF) [69] considers all possible two-layer hit-pairs to search for circular patterns in the transverse plane. Peaks in the two-dimensional distribution of circle-track parameters (radius and direction at the origin) imply the presence of track candidates. The alternative algorithm (AA) [70] tries to fit full (five-parameter) helical tracks to triplets of SMT hits (in three different layers). These trajectories are extrapolated to all other SMT layers and the CFT and nearby hits are added. If more than a single candidate hit is present in a given layer the track candidate is split to accommodate all possibilities. The resulting AA track candidates are used to find vertex candidates, after which the above procedure is repeated for 'CFT-only' tracks (tracks without any SMT hits) under the assumption that tracks originate from one of these vertex candidates. Both sets of track candidates are combined. Hits shared by multiple tracks are assigned to the longer track or (in case both tracks contain the same number of hits) to the track with the smaller χ^2, and tracks with too few hits or with bad fits are rejected. Finally, all tracks are refitted using a Kalman filtering algorithm [1] based on an 'interacting propagator' implementing the magnetic curvature and interactions with the detector material [71].

3.1.2 Primary vertices

Reconstruction of primary vertices (PVs) is implemented using an adaptive fitting technique [72]. To reduce the contributions of distant tracks, tracks are weighted according to their contribution to the χ^2 of the vertex fit. The three steps involved in PV reconstruction are:

[1] A Kalman filter can be seen as an non-iterative alternative to a χ^2 fit. Step-by-step, track candidates are extrapolated to the next surface. At each surface the extrapolated prediction is combined with the next hit and a 'filtered' track candidate is formed. After all hits are included, the resulting track parameters and uncertainties represent the optimal estimate of the true track parameters.

Track selection Only good tracks with at least two SMT hits and $p_{\mathrm{T}} > 0.5$ GeV/c are selected. Outside the SMT fiducial region ($z_{\mathrm{PV}} > 36$ cm for tracks with $\eta < 1$, increasing linearly to $z_{\mathrm{PV}} < 80$ cm for tracks with $\eta < 2$ for $z_{\mathrm{PV}} > 0$ and similarly for $z_{\mathrm{PV}} < 0$ [72]) this first requirement is dropped. Starting from the highest-p_{T} selected track, tracks are clustered along the z-direction. A track is added to the cluster if it is within $\Delta z < 2$ cm of the seed track.

Vertex reconstruction All tracks within a cluster are constrained to a single vertex using a Kalman filter technique, providing an estimate of the beam position and width. In the next step all tracks with an impact parameter significance less than five with respect to this beamspot position estimate are fitted into a common vertex using the adaptive fitting algorithm. Iteratively, all track candidates are Kalman-fitted into a vertex, weighting tracks based on their contributions to the vertex fit χ^2. Iteration stops when the weights have converged.

Vertex selection At a luminosity of $\mathcal{L} \approx 10^{32}$ cm^{-2}s^{-1} the number of interactions per bunch crossing follows a Poisson distribution with a mean of 3.5. One of these interactions should be the hard $\mathrm{p\overline{p}}$ scatter that fired the trigger, the others will most likely be minimum-bias collisions. Hard scatter vertices show a harder p_{T} spectrum for the associated tracks. For each primary vertex candidate the probability is determined that it corresponds to an elastic $\mathrm{p\overline{p}}$ interaction (i.e. is a minimum-bias vertex) and not to a hard-scatter event by comparing the $\log_{10}(p_{\mathrm{T}})$-distribution of the associated tracks to a predetermined template. The candidate with the lowest minimum bias probability is selected as 'the' primary vertex.

The primary vertex reconstruction efficiencies for QCD events as well as for the different $\mathrm{t\bar{t}}$ decay channels is approximately 100% and independent of the number of additional minimum bias vertices [73]. For QCD events the number of tracks associated with the PV averages between 35 and 40; for hadronic $\mathrm{t\bar{t}}$ decays between 45 and 50. The resolution for hadronic $\mathrm{t\bar{t}}$ events is in the transverse plane approximately 8 µm, and along the beam direction approximately 20 µm (neither distribution is a real Gaussian).

3.1.3 Secondary vertices

Apart from the primary vertex also secondary vertices (SVs) from the decay of long-lived particles like B-mesons, Λs and K^0_Ss need to be reconstructed. The main purpose of the reconstructed SVs is to identify jets originating from b-quarks.

The reconstruction and selection of secondary vertices follows these steps:

Track clustering Using a simple cone algorithm tracks around seed tracks with $p_T > 2$ GeV/c are clustered into 'track jets' with a minimum p_T of 5 GeV/c.

Vertex fitting In each track jet all potential vertices are fitted using a Kalman filter technique. First all two-track combinations of tracks in the track jet (ignoring tracks associated with the primary vertex) with a $\chi^2 < 10$ are generated. Then for each of these seeds any track is added that contributes less than 10 to the χ^2 of the vertex fit.

Vertex preselection To remove unphysical vertices and vertices from Λs and K^0_Ss each candidate vertex has to satisfy the following requirements. The opening angle α of the vertex (the angle between the vector sum of the associated tracks and the line connecting the secondary vertex with the primary vertex) has to be less than 0.1 to remove secondary vertices incompatible with the PV. The transverse decay length (the distance between the primary and secondary vertices) L_{xy} has to be less than 2.5 cm (to remove Λ etc. decays) and the longitudinal decay length L_z has to be less than 3 cm. The vertex quality is improved by requiring each vertex to contain at least one track with $p_T > 1.5$ GeV/c (2 GeV/c for vertices with only two associated tracks).

Final vertex selection The final step of the secondary vertex reconstruction takes care of duplicate track-vertex assignments. For each track jet the best vertex is selected (based on the χ^2 of the fit) and all vertices sharing at least one track with that vertex are removed. This step is repeated for all remaining vertices (in order of increasing χ^2).

For a more detailed description of the secondary vertex reconstruction algorithm, the choice of the parameter values and efficiency studies, see references [74, 75].

3.2 Leptons and photons

Even for the fully hadronic $t\bar{t}$ decay channel leptons and photons are important physics objects. Muons are used in the identification of b-jets. Photons are an important calibration tool for the jet energy scale (chapter 4).

3.2.1 Muons

Most muons leave only a minimal amount of (radiative) energy in the calorimeter, without initiating a shower. They pass through the calorimeter and are detected in the outermost subdetector: the muon spectrometer. In the analysis described in this thesis muons are only used indirectly. Events containing isolated muons are vetoed in order to ensure orthogonality of the data sample with respect to the $t\bar{t}$ cross section analyses in the leptonic decay channels. Muons inside jets are included in the jet energy scale corrections. Only a cursory overview of muon reconstruction is presented here. A more complete discussion can be found in ref. [76].

Muon candidates are reconstructed from tracks in both the central tracker and the muon tracker. Candidates reconstructed in the muon system alone are called 'local muons', matched to a track in the central tracker they become 'global muons'. Global muons have a much longer (combined) track and a correspondingly better p_{T} resolution.

Reconstructed muons are classified based on the muon layers they were detected in (the muon type) and additional quality criteria applied (the muon quality). In the analysis in this thesis only 'Medium NSeg3' muons within $|\eta| < 2$ are used.

NSeg3: The muon was reconstructed using track segments in both the muon system A- and BC-layers and matched to a global track.

Medium: The muon satisfies the following quality requirements:

- at least two wire hits in the A-layer,
- at least one scintillator hit in the A-layer,
- at least two wire hits in the BC-layer and
- at least one scintillator hit in the BC-layer. (For central muons with less than four BC-layer wire hits this requirement is dropped.)

An additional handle on the muon quality is the track matching between the central tracker and the muon system. The track quality cuts select the central tracks to be

considered for matching. No track matching requirements are applied to the muons for the jet energy scale corrections. The track matching for the orthogonality veto muons is required to satisfy:

- the χ^2 of the track fit should be less than four and
- for a track to be considered for matching, the distance of closest approach between the track and the primary vertex should be less than 0.02 cm (0.2 cm) for tracks with (without) SMT hits.

Muons used for jet energy scale corrections are required to have $p_{\mathrm{T}} < 500$ GeV/c and are otherwise rejected as mismeasured (this only removes grossly mismeasured muons and still requires special treatment to avoid disturbing the jet energy resolution, see also section 5.1.3). The veto muon selection uses the default lower p_{T} threshold of $p_{\mathrm{T}} \geq 15$ GeV/c.

The veto muons are also required to be isolated in the calorimeter and to be at least $\Delta\mathcal{R}(\eta, \phi) = 0.5$ away from any good jets. This rejects muons from heavy flavour decays in favour of muons from W decays.

To reject muons originating from cosmic ray showers a loose timing cut is applied. This timing cut requires all scintillator hits associated with a muon to originate from within a 10 ns time window around the bunch crossing time. The cut on the separation of the track and the primary vertex mentioned above also suppresses cosmic muons.

3.2.2 Electrons and photons

Both electrons and photons are initially reconstructed as electromagnetic clusters in the calorimeter. The subsequent identification criteria distinguish between electrons and photons by combining the information from the calorimeter, the Central Preshower and the tracker.

Electromagnetic clusters are reconstructed starting from seed towers (with $p_{\mathrm{T}} >$ 500 MeV/c) in the EM calorimeter and adding all neighbouring EM towers with $p_{\mathrm{T}} > 50$ MeV/c within $\Delta\mathcal{R}_{\mathrm{EM3}} < 0.3$ cm (CC) or within $\Delta\mathcal{R}_{\mathrm{EM3}} < 10$ cm in the third EM layer (EC). Clusters with $p_{\mathrm{T}} < 1$ GeV/c are rejected and to each remaining cluster are added all EM towers within $\Delta\mathcal{R} < 0.4$. The cluster centre is recalculated and this step is repeated until the cluster remains stable. The final cluster centre is defined as the energy-weighted mean of its cells in the third EM calorimeter layer.

3.2.3 Photon identification requirements

The jet energy scale analyses use photons to select clean calibration events. Most jet energy scale analyses use the so-called 'tight' photon selection. To qualify, an EM cluster has to satisfy the following criteria:

- The cluster is reconstructed with its axis either in the central region ($|\eta_{\mathrm{d}}| < 1.0$) or in the endcap regions ($1.5 < |\eta_{\mathrm{d}}| < 2.5$) and is in the detector fiducial regions (clusters reconstructed near inter-module boundaries are excluded).
- The fraction of the cluster energy deposited in the electromagnetic calorimeter (EMF) must be EMF > 0.96. This suppresses jets faking photons.
- To reduce sensitivity to photons from electron/positron conversions or otherwise associated with other physics objects, EM clusters that overlap with reconstructed tracks are rejected. The probability to have a spatially matched track must be $P(\chi^2_{\mathrm{spatial}}) < 0.1\%$, where $\chi^2_{\mathrm{spatial}}$ is the quadratic sum of the sigma-weighted track-cluster distances in ϕ and z:

$$\chi^2_{\mathrm{spatial}} = \left(\frac{\Delta\phi}{\sigma_\phi}\right)^2 + \left(\frac{\Delta z}{\sigma_z}\right)^2. \tag{3.1}$$

- The core cluster energy should be isolated: the energy in a ring with $0.2 < \Delta\mathcal{R} < 0.4$ around the axis, compared to the core energy within $\Delta\mathcal{R} < 0.2$ must be $E(0.2 < \Delta\mathcal{R} < 0.4)/E(\Delta\mathcal{R} < 0.2) < 0.07$.
- The squared cluster width in the third EM layer must be less than 14 cm^2 in the central region. In the endcaps a more complex, $|\eta_{\mathrm{d}}|$-dependent, cut is used.
- The scalar p_{T} sum of all tracks within a ring $0.05 < \Delta\mathcal{R} < 0.7$ around the cluster with $p_{\mathrm{T}} > 0.4$ GeV/c and a distance with respect to the primary vertex of $\Delta z < 2$ cm and $\mathrm{DCA}_{xy} < 1$ cm must be $\Sigma p_{\mathrm{T}} < 1$ GeV/c. This aids in the rejection of conversion photons, which lead to very narrow track geometries.

Use of the preshower information

Two additional variables are computed from Central Preshower (CPS) information. Both are based on the difference in azimuthal direction between the energy deposition in the preshower and the position of the EM cluster. These variables were designed to provide better separation between photons and EM-like jets.

Photons and electrons tend to result in very narrow energy depositions in the preshower, as opposed to jets. The first CPS variable weights the squared ϕ difference with the CPS strip energies to reject clusters associated with wide CPS energy depositions. This weighted difference has to be less than 0.003 GeV.

Many EM-like jets contain π^0s, most of which decay into two photons, leading to a multi-peak energy deposition in the preshower. The second CPS variable weights the squared ϕ difference by the squared strip energy to remove EM clusters associated with such multi-peak preshower signatures. This weighted difference has to be less than 0.0015 GeV^2.

3.3 Jets

Partons created in the hard-scatter, as well as the remnants of the beam particles are predominantly coloured objects and as such subject to strong interactions. Due to colour confinement collimated jets of partons are created, which hadronize into colourless particles. These particles, hadrons and mesons, are (in)directly detected. Jet reconstruction algorithms cluster energy depositions in the calorimeter to estimate the energy and direction of the original stable-particle jet. Most of the jets originating from the beam remnants lie very forward, outside the detector acceptance. Colour connections between the hard scatter participants and the beam remnants contribute to the 'underlying event', all physics not directly related to the hard interaction.

3.3.1 Noise suppression

Due to the presence of noise in the calorimeter, only those cells with a certain minimum amount of energy are considered in object reconstruction in order to avoid too large noise contributions. In the D0 calorimeter system several steps of zero-suppression are applied to remove noise.

Online, calorimeter readout electronics dynamically subtract the stored baseline signal from three clock ticks (396 ns) before the current interaction. Hot cells, cells that repeatedly contain large amounts of noise, mostly due to hardware failure, are also excluded online. In periods without beam, for each calorimeter channel the pedestal (the amount of noise present without interactions) is determined. The first step in offline zero-suppression is to remove all cells from the readout in which the absolute value of the energy after pedestal subtraction is less than 1.5 times the root-mean-square (RMS, or σ) of the pedestal.

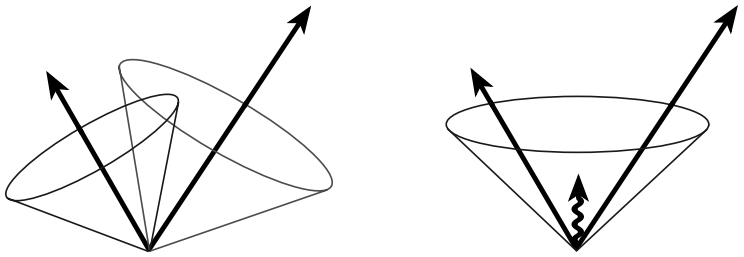


Figure 3.1: *Illustration of possible infrared sensitivity issues with jet algorithms [80]. The presence of soft radiation in between jets could lead these jets to become reconstructed as a single jet. In this case soft radiation would change the jet multiplicity.*

The T42 algorithm

To remove as much noise as possible from the calorimeter information before reconstruction D0 uses the T42 algorithm [77, 78]. All negative cells are removed; the large layer-to-layer differences in noise levels negate any noise cancelling effects when combining cells into towers. Apart from removing all cells with negative energies two levels of zero-suppression are applied, both based on the RMS value σ of the noise pedestal: a cell is suppressed if it has an energy of less than 2.5σ, or if it has no neighbours with energies of at least 4σ. Neighbouring cells in all three directions are considered, including corner cells; a typical cell has O(26) neighbours. The algorithm does not touch the first EM layer (since it has no 'inner' neighbours), the massless gaps and the inter-cryostat detector (since their geometries require special treatment). Depending on the type of data under consideration, the T42 algorithm removes 30–60% of cells in the event and improves reconstruction of all calorimeter objects [79].

3.3.2 Calorimeter jets

Ideally, a jet algorithm should accurately reconstruct the energies and directions of all jets in the event and, even more importantly, correctly reproduce the jet multiplicity, independently of event topology and detector geometry. In seed-based algorithms, which only search for jets around towers with a certain minimum amount of energy, two jets close together could either be reconstructed as two individual jets or, in the presence of soft (infrared) radiation between the jets, as a single jet (figure 3.1). Problems with collinear (radiation close to the direction of the original parton or collinear parton splitting) safety can occur if for example a hard parton in between

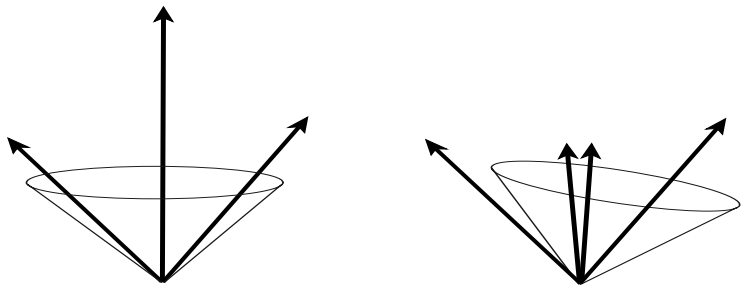


Figure 3.2: *Illustration of possible collinear sensitivity issues with jet algorithms [80]. In this case collinear splitting of the middle parton changes the initial seed to one of the other partons, resulting in a differently reconstructed jet. Whether or not the parton on the left-hand side is reconstructed as a jet depends on the parameters of the algorithm and on the parton energy.*

two softer partons splits such that the resulting partons have lower p_{T} than the two partons on the sides (figure 3.2). If the distance between the two outer partons is between $\mathcal{R}_{\mathrm{cone}}$ and $2\mathcal{R}_{\mathrm{cone}}$ the former will lead to a single jet around the middle parton, whereas the latter will lead to two individual jets, with energies depending on the ordering of the seeds. Especially when comparing to theoretical predictions infrared safety and collinear safety are important.

The D0 cone algorithm

Both partons created in hard interactions, and spectator partons from the (anti)proton remnants, are subject to the strong force and produce large numbers of particles in collimated jets. These particles in turn lead to energy depositions in many calorimeter cells. A jet reconstruction algorithm is employed to reconstruct jets from the energy depositions in the calorimeter.

D0 uses the 'Run II midpoint cone algorithm' [80, 81], an improved version of the simple fixed cone algorithm. Towers within a cone with radius $\Delta\mathcal{R}(y,\phi) = \sqrt{(y-y_{\mathrm{jet}})^2+(\phi-\phi_{\mathrm{jet}})^2} < \mathcal{R}_{\mathrm{cone}}$ are clustered together, [2] where $\mathcal{R}_{\mathrm{cone}}$ is the jet cone size. With the exception of some QCD analyses, almost all D0 analyses use jets with a cone size of $\mathcal{R}_{\mathrm{cone}} = 0.5$. Individual cells are treated as massless, and all cell positions are corrected for the position of the primary vertex. Massive towers are

[2] For efficiency reasons the preclustering step uses pseudorapidity η instead of real rapidity y when calculating $\Delta\mathcal{R}$. The final clustering uses rapidity.

reconstructed by summing all cells with the same (η, ϕ) [3] position containing enough energy to pass zero-suppression. Each tower with $p_{\mathrm{T}} > 0.5$ GeV/c is considered as a seed tower. The jet reconstruction algorithm operates on these towers. The first step groups towers within $\Delta\mathcal{R} < 0.3$ around each seed tower into preclusters, ignoring towers with $p_{\mathrm{T}} < 1$ MeV/c. Resulting preclusters with $p_{\mathrm{T}} < 1$ GeV/c or containing only a single tower are removed from the list. The next step loops over all preclusters P in order of decreasing p_{T}. If the closest precluster with respect to P is within $\Delta\mathcal{R} < \mathcal{R}_{\mathrm{cone}}/2$, P is considered a jet candidate. The four-vectors of all towers within $\Delta\mathcal{R}(y, \phi) < \mathcal{R}_{\mathrm{cone}}$ around the candidate are added (in the E-scheme):

$$
\begin{aligned}
p_{\mathrm{cand}} &= (E_{\mathrm{cand}}, \vec{p}_{\mathrm{cand}}) = \sum_{\mathrm{towers}} (E, \vec{p}) \\
y_{\mathrm{cand}} &= \frac{1}{2} \ln \left(\frac{E_{\mathrm{cand}} + p_{\mathrm{z,jet}}}{E_{\mathrm{cand}} - p_{\mathrm{z,jet}}} \right) \\
\phi_{\mathrm{cand}} &= \tan^{-1} \left(\frac{p_{\mathrm{y,jet}}}{p_{\mathrm{x,jet}}} \right)
\end{aligned}
$$

leading to a new candidate centroid $(y_{\mathrm{cand}}, \phi_{\mathrm{cand}})$. This step is repeated iteratively until the centroid is stable unless $p_{\mathrm{T}} < 3$ GeV/c, in which case the candidate is rejected. Stable candidates are added to a list of proto-jets unless they either contain 99% of the p_{T} of another proto-jet, or match another proto-jet within $\Delta\mathcal{R} < 0.005$. The third step repeats this procedure around each p_{T}-weighted average (y, ϕ) midpoint between any two proto-jets satisfying $\mathcal{R}_{\mathrm{cone}} < \Delta\mathcal{R} < 2\mathcal{R}_{\mathrm{cone}}$. This makes the algorithm safer to soft radiation [80]. The final step, splitting and merging, takes care of overlapping proto-jets. Again the list of proto-jets is processed in order of decreasing p_{T}. If a proto-jet contains towers also contained in any other proto-jet, the total p_{T} shared with its highest-p_{T} neighbour is determined. If the overlap contains more than 50% of the p_{T} of that neighbour, the two proto-jets are merged: the towers in the neighbour are added to the (higher-p_{T}) proto-jet, the neighbour is removed, and the centroid of the higher-p_{T} proto-jet is recalculated. If the overlap is less than 50%, the proto-jets are split: each tower in the overlap is assigned to the proto-jet closest in $\Delta\mathcal{R}$. After each step the list of proto-jets is reordered. This continues until no overlapping pairs are left. To get rid of physically meaningless jets, proto-jets with $p_{\mathrm{T}} < 6$ GeV/c are removed. The surviving proto-jets are the jet objects used in physics analyses.

[3] Strictly speaking, cells are combined based on their i_η and i_ϕ coordinates instead of their physical (η, ϕ) positions, the result is the same.

The same algorithm (with slight modifications to the proto-jet criteria) is also used in Monte Carlo to reconstruct particle jets from all stable particles instead of from calorimeter towers.

Two points bear special consideration. Like any lower cut-off, the 6 GeV/c minimum jet p_{T} introduces a bias by preferentially selecting jets that fluctuate up across the threshold and rejecting downward-fluctuated jets. More importantly, unlike the name of the algorithm suggests, the reconstructed jets are *not* circular. Jets that do not require splitting or merging can be considered 'circular on average' (an effect exploited by e.g. the jet energy scale offset correction), even though the per-jet energy depositions are not circular. Split and/or merged jets, however, represent many different geometries but are never circular (the offset correction also treats these jets as circular, for the simple reason that it's not possible with the current D0 software to properly handle these jets).

One of the advantages of cone jet algorithms is easier calibration compared to e.g. k_{T} algorithms, which lack the well-defined average jet geometry. However, at higher occupancies more and more jets will overlap, requiring special treatment and negating the advantages. With higher numbers of simultaneous interactions and higher jet multiplicities associated with the higher center-of-mass energy, jet reconstruction and calibration will be one of the challenges for the experiments at the Large Hadron Collider (LHC).

Jet detector coordinates Cells and thus towers are labelled by their (i_η, i_ϕ) coordinates. Reconstructed from calorimeter towers, jets have two sets of associated coordinates: the usual physics coordinates (η, ϕ) representing the real direction with respect to the primary vertex, and 'detector coordinates' $(\eta_{\mathrm{d}}, \phi_{\mathrm{d}})$ reflecting the position of the jet in the calorimeter. The latter are determined from the energy-weighted centre of the jet cluster and specifies which towers are contained in the jet. Conversion between the two jet coordinate systems involves not only the primary vertex position, but also the detector geometry and alignment. For all but very specific classes of events with central vertices and relatively central jets this conversion is impractical. Each pair of coordinates carries different, but important information. For example the jet response calibration depends on both η and η_{d}. The former describes the angle under which the jet traverses the solenoid, the latter explains which towers were involved in the reconstruction of the jet.

Jet quality requirements

To remove noisy jets, fake jets (e.g. photons misreconstructed as jets) and jets expected to show large reconstruction biases, a series of jet quality requirements is applied to each jet. Only jets passing these cuts are used in physics analyses.

The following jet quality criteria are used [82]:

- Jets tend to deposit energy somewhat evenly over the EM and hadronic layers if the calorimeter. The fraction of the jet energy deposited in the electromagnetic calorimeter (EMF) must be $0.05 < \text{EMF} < 0.95$. Jets with $|\eta_{\text{d}}| > 2.5$ (where the EM cells become larger and thus noisier) are required to satisfy $0.04 < \text{EMF} < 0.95$. This removes both noise jets (with suspiciously low EMF) and photons (with very high EMF) reconstructed as jets. [4]

- Due to the large sizes of the hadronic calorimeter cells especially the coarse hadronic cells are very noisy. To remove those jets dominated by coarse hadronic noise, the fraction of the jet energy deposited in the coarse hadronic calorimeter (CHF) must be:
  - $\text{CHF} < 0.44$ for central jets with $|\eta_{\text{d}}| < 0.8$,
  - $\text{CHF} < 0.46$ for jets in the endcap with $1.5 < |\eta_{\text{d}}| < 2.5$,
  - $\text{CHF} < 0.4$ otherwise.

  Jets in the $0.85 < |\eta_{\text{d}}| < 1.25$ range are dominated by the coarse hadronic calorimeter and are allowed to have $\text{CHF} < 0.6$ if 90% of the jet energy is contained in less than 20 cells.

- The jet has to be confirmed by the calorimeter Level 1 trigger information [83]. This is implemented as a cut on the ratio between the jet transverse momenta in the trigger system and in the precision readout:

$$L1_{\text{ratio}} \equiv \frac{p_{\text{T}}^{\text{L1 readout}}}{p_{\text{T}}^{\text{precision readout}}} \tag{3.3}$$

  where
  - $p_{\text{T}}^{\text{L1 readout}}$ is the sum of the scalar p_{T}s in a cone of $\mathcal{R} < 0.5$ around the jet using only the highest-energy L1 trigger towers in the event, and

[4] For jets in the gap in the EM calorimeter coverage ($1.2 < |\eta_{\text{d}}| < 1.3$, section 2.3.2) the minimum EMF cut is softened.

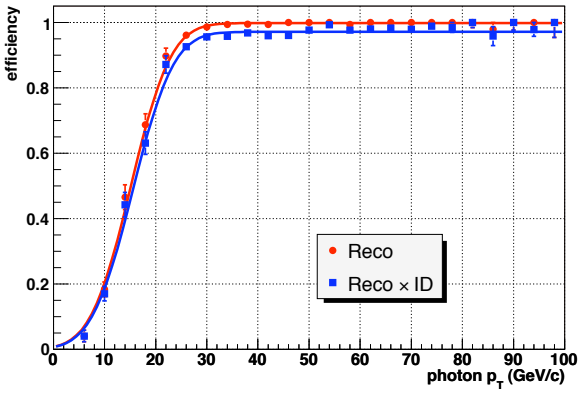


Figure 3.3: *Reconstruction efficiency and combined reconstruction and identification efficiency for central ($0.0 < |\eta| < 0.4$) $\mathcal{R}_{cone} = 0.5$ cone jets [84]. The efficiency was estimated in photon+jet events and is plotted as a function of the photon p_T.*

– $p_{\mathrm{T}}^{\text{precision readout}}$ is the vector p_{T} sum over all (reconstructed) towers assigned to the jet (but excluding the coarse hadronic layers and massless gaps).

Each jet must satisfy:

– $L1_{\mathrm{ratio}} > 0.5$, or

– $L1_{\mathrm{ratio}} > 0.35$ and $p_{\mathrm{T}} < 15$ GeV/c and $|\eta_{\mathrm{d}}| > 1.4$, or

– $L1_{\mathrm{ratio}} > 0.1$ and $p_{\mathrm{T}} < 15$ GeV/c and $|\eta_{\mathrm{d}}| > 3.0$, or

– $L1_{\mathrm{ratio}} > 0.2$ and $p_{\mathrm{T}} \geq 15$ GeV/c and $|\eta_{\mathrm{d}}| > 3.0$.

The purpose of this Level 1 confirmation is to remove jets dominated by noise in the precision readout electronics.

Differences in jet response between data and Monte Carlo, in combination with the 6 GeV/c reconstruction threshold introduce a relative data-to-MC bias. Jet energy scale corrections should remove the response differences. To remove the bias due to the reconstruction threshold, a higher offline cut is required. An offline cut of $p_{\mathrm{T}} > 15$ GeV/c is applied, which appears to be a safe threshold [85, 84].

For $\mathcal{R}_{\mathrm{cone}} = 0.5$ cone jets the approximate 50% efficiency point for the jet identification criteria lies around $p_{\mathrm{T}} > 15$ GeV/c [86]. As can be seen from figure 3.3,

at 15 GeV/c the combined jet reconstruction and ID efficiency [5] for central jets is approximately 50%, reaching full efficiency ($\approx 98\%$) around $p_{\mathrm{T}} \approx 30$ GeV/c [84].

Identifying jets from b-quarks

In many physics analysis, especially where concerning high jet multiplicities, it's advantageous to be able to identify the jets originating from b-quarks. The process of identifying b-jets is called b-tagging. b-Quarks hadronize into B-hadrons; since these hadrons live relatively long (approximately 10^{-12} s), they tend to cross a large distance between creation and decay. This results in a distinguishable decay vertex away from the primary vertex of the event. The presence of such a secondary vertex, as well as the fact that tracks from the decaying B-meson do not point back to the primary vertex, can be used to estimate the 'b-likeness' of jets. In D0 several different b-tagging algorithms have been developed. This analysis uses the latest addition: a neural network based b-tagger.

Taggability Track jets, clusters of tracks pointing back to the primary vertex, reconstructed with the same simple cone algorithm as used for secondary vertex reconstruction are an important b-tagging tool. To separate the efficiencies of the b-tagging algorithm from (e.g. track) reconstruction efficiencies, the notion of jet 'taggability' is introduced. Taggable jets fulfill all requirements to be operated on by b-tagging algorithms. For b-tagging efficiencies only taggable jets need to be considered. A jet is considered taggable if it is matched in $\Delta\mathcal{R} < 0.5$ with a track jet. For taggability track jets are reconstructed using slightly different parameters than in the case of secondary vertex reconstruction. Tracks around seed tracks with $p_{\mathrm{T}} > 1$ GeV/c are clustered into $\mathcal{R}_{\mathrm{cone}} = 0.5$ track jets. Tracks are required to have at least one SMT hit, $p_{\mathrm{T}} > 0.5$ GeV/c and DCA < 0.2 cm (0.4 cm) in the transverse plane (along the z-direction). Imperfections in Monte Carlo modelling of e.g. tracking efficiencies lead to differences in taggability between data and Monte Carlo. Explicit MC taggability corrections (see section 5.2.1) allow the use of common tagging efficiencies for both data and Monte Carlo.

[5] While the true efficiency can only be determined from Monte Carlo the tag-and-probe method, balancing an assumed jet against a well-known object (e.g. against a photon), allows the determination of an approximate efficiency consistently between both data and MC. This can then be used to correct the MC to match the data, providing the correct efficiency for cross section measurements.

The neural network b-tagger A neural network b-tagger has been developed [87, 88] that combines variables from three different b-taggers: the jet lifetime impact parameter (JLIP) tagger, the counting signed impact parameter (CSIP) tagger and the secondary vertex tagger (SVT).

The CSIP tagger counts the number of tracks in a jet with a large impact parameter significance with respect to the primary vertex. A high count means it is unlikely the tracks originated from the PV and the jet is tagged.

The JLIP tagger combines all track impact parameters to estimate the probability that all tracks in a jet originated from the primary vertex. Jets with low JLIP probabilities are tagged as b-jets.

The SVT uses tracks with large impact parameter significance to reconstruct secondary vertices. If a secondary vertex is found within $\Delta\mathcal{R}(\eta, \phi) < 0.5$ of a jet, the jet is b-tagged.

The neural network tagger (NN tagger) combines the following seven variables from the above taggers (if more than one secondary vertex is found within range of the jet, the neural network tagger selects the one with the highest impact parameter significance).

- The decay length significance of the secondary vertex selected by the NN tagger.
- The weighted combination of the impact parameter significances of all tracks in the jet (a CSIP variable).
- The JLIP probability that all tracks in the jet originated from the primary vertex.
- The χ^2-per-degree-of-freedom of a fit constraining all tracks in the jet to the selected secondary vertex.
- The number of tracks used by the SVT to reconstruct the selected secondary vertex.
- The mass of the selected secondary vertex.
- The number of secondary vertices found by the SVT within a cone of $\Delta\mathcal{R}(\eta, \phi) < 0.5$ around the jet.

To ensure all input taggers were operating in their design ranges some additional requirements are placed on the neural network inputs:

- the CSIP count has to exceed eight,
- the JLIP probability has to be less than 0.02 and
- the SVT decay length significance has to exceed 2.5.

The NN tagger was trained on $b\bar{b}$ and QCD di-jet MC events. The efficiency was measured in a data sample with jets containing muons and scaled to be applicable to inclusive jet samples using a MC-based correction factor. The efficiencies to tag b-, and c-jets, as well as the fake rate (to tag anything else but jets from heavy quarks) are parameterised as functions of jet pseudorapidity η and transverse momentum p_{T} (Tag Rate Functions, or TRFs). Figure 3.4 shows the Tag Rate Functions for the medium operating point (NN discriminant > 0.2) of the neural network tagger.

Twelve operating points have been defined as lower limits on the neural network discriminant output, ranging from NN > 0.1 (with an efficiency of $\approx 78\%$ and a fake rate of $\approx 11\%$) to NN > 0.925 (efficiency $\approx 38\%$, fake rate $\approx 1‰$).

A more detailed description of the training and structure of the neural network used can be found in ref. [87]. The definitions of the operating points and the relevant efficiency and fake rate parameterisations are documented in ref. [88].

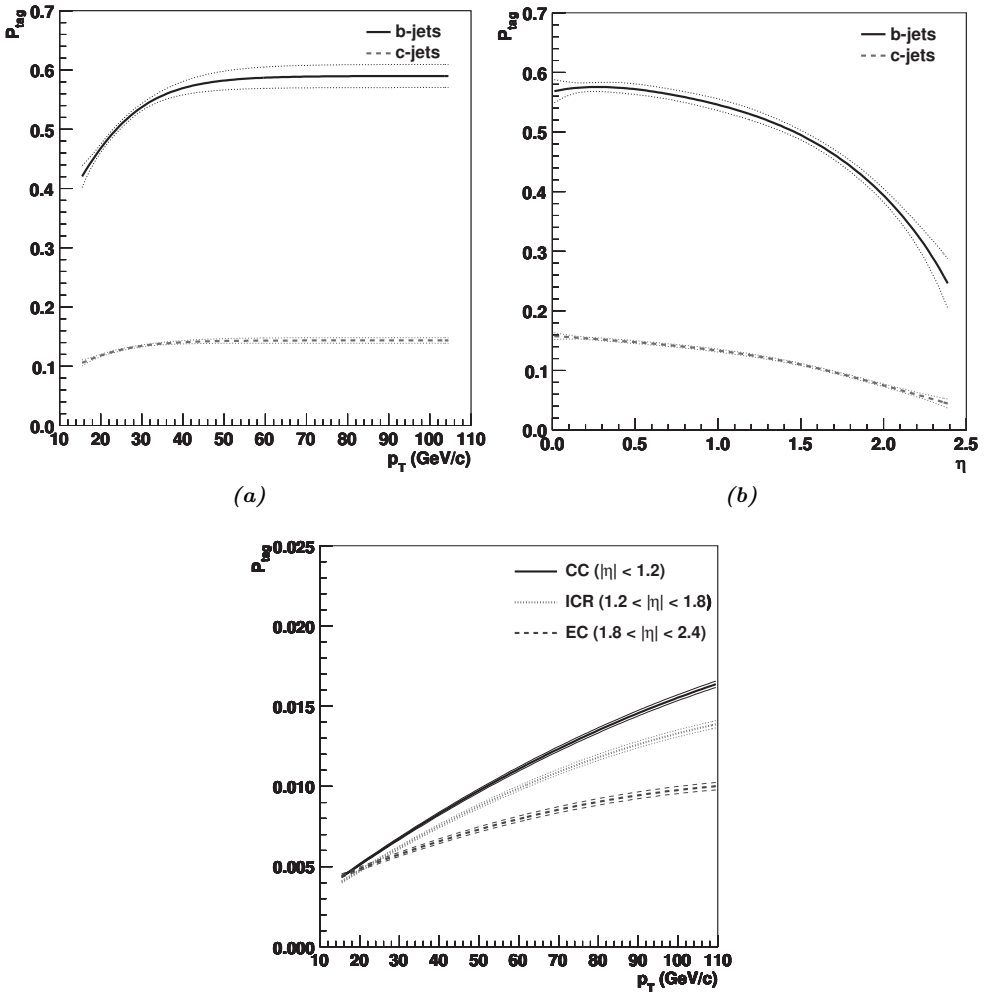
3.4 Missing transverse energy

Particles that go undetected, be it for real physics reasons (e. g. neutrinos) or due to detector acceptance or geometry, disturb the momentum balance in the transverse plane. The missing transverse energy ($\not\!\!E_{\mathrm{T}}$) is defined as this momentum imbalance:

$$(\not\!\!E_{\mathrm{T},x}, \not\!\!E_{\mathrm{T},y}) = -(\sum p_x, \sum p_y)$$

where the sum runs over all calorimeter cells, including the inter-cryostat detector and the massless gaps, that pass the T42 algorithm. The missing energy is calibrated indirectly by propagating the jet and EM calibrations. For the jet energy calibration the $\not\!\!E_{\mathrm{T}}$ is compensated for the fully corrected energy of the selected photons:

$$\not\!\!E^{\mathrm{corr}}_{\mathrm{T},x/y} = \not\!\!E_{\mathrm{T},x/y} - \sum_{\mathrm{photons}} (\not\!\!E^{\mathrm{corr}}_{\mathrm{T},x/y}(\mathrm{photon}) - \not\!\!E^{\mathrm{uncorr}}_{\mathrm{T},x/y}(\mathrm{photon})).$$



(a) *(b)*

(c)

Figure 3.4: *Tag Rate (efficiency) Functions for the $NN > 0.65$ operating point of the neural network b-tagger [88]. Shown are the efficiencies for b-, and c-jets (solid and dashed curves respectively) as a function of jet p_T (a) and η (b). Figure 3.4c shows the corresponding fake rate as a function of p_T, parameterised in three different η regions.*

3.5 Summarising

Both calorimeter and central tracker information is used to reconstruct electrons and photons. Electron-photon separation is based on calorimeter shower shape information as well as on the signal in the Central Preshower. Jets are reconstructed from calorimeter information alone, but their energy calibration is derived using photons (see chapter 4). Quality criteria suppress noisy and/or fake jets and ensure well-defined and clean jets.

Chapter 4

Calorimeter calibration and jet energy scale

Many physics analyses at hadron colliders involve multiple jets. This makes calibration of jet energies and momenta an important analysis input. Prime examples are the top quark mass, typically reconstructed from final states involving up to six jets, and searches for new particles or resonances, where the energy resolutions show up directly in the width of a possible signal peak. Of more indirect importance: the precise calibration of the calorimeter and reconstructed objects is the only way to improve the resolution on the missing transverse energy, important in searches for physics beyond the standard model.

This chapter describes the work done in D0 in Run II to calibrate the calorimeters and the jet energy scale (JES). In Run II, the availability of high-statistics calibration samples has led to a detailed understanding of jet development and detection in the D0 detector. This has been the work of many people over many years of which only a limited overview can be presented here. For an account of the calorimeter system performance during Run II, please refer to ref. [89]. More detailed information on the calorimeter calibration can be found in references [90] and [91]. A complete description of the jet energy scale methods and analyses is given in references [92, 93]. This chapter relies heavily on the latter references.

A shorter version of this chapter is to be published in the proceedings of the 'XIII International Conference on Calorimetry in High Energy Physics' [94].

4.1 Calorimeter calibration

With the upgrade from Run I to Run II the bunch crossing rate increased from 286 kHz to 2.5 MHz. The corresponding decrease in bunch separation from 3.5 µs to 396 ns,

together with the high instantaneous luminosities, result in a Poisson-distributed average number of up to five interactions per bunch crossing [95]. To reduce the effects of pile-up (energy depositions spilling over from previous bunch crossings) the charge integration time of the calorimeter preamplifiers has been reduced to 260 ns, compared to 2.2 µs in Run I. With a $\sim$ 430 ns drift time across the roughly 2.3 mm gaps in the calorimeter this makes the calorimeter readout much more sensitive to second-order effects like the mechanical precision of the calorimeter modules and assembly. Since hadronic showers tend to take longer to develop and deposit their energies (the latter governed by the long time scale of neutron thermalization) it also results in a strongly non-compensating calorimeter: $e/h > 1$.

4.1.1 Online calibration

The calorimeter electronics are calibrated using pulse charges injected into the preamplifiers (see figure 2.9 on page 43) to linearise the response and equalise the gain across all channels. This procedure is performed approximately every two weeks in periods without beam or after any hardware changes. The achieved linearity for the whole calorimeter system is better than 0.2%.

Between subsequent Tevatron stores the (noise) pedestal calibrations are performed. Pedestal and RMS values are determined for all cells and for both preamplifier gain paths. The history of the pedestal values shows stability within one ADC count for most channels. The results of the pedestal calibration are also used to suppress bad channels until repairs can be made.

4.1.2 Offline calibration

Offline ϕ and η intercalibrations make the calorimeter as uniform as possible and optimise the jet resolution.

The first step in the offline calorimeter calibration is the inter-ϕ calibration of the electromagnetic calorimeter using a specially collected EM data sample. Subsequently the EM calorimeter is made uniform across η using a $Z \rightarrow e^+e^-$ data sample. The inter-ϕ calibration for the hadronic part of the calorimeter is performed using a dedicated QCD sample. At this point the central calorimeter is calibrated to within $\approx 1\%$ and the endcaps to $\approx 2-3\%$. The final step in the calibration procedure is the η-calibration of the hadronic calorimeter. In this step the calorimeter is calibrated

to optimise the jet energy resolution for 45 GeV jets using a sample of back-to-back di-jet events.

The absolute size of energy fluctuations increases with increasing energy. However, higher-energy showers tend to develop later in the calorimeter, leading to higher sampling fractions in the hadronic (compared to the EM) calorimeter, with correspondingly smaller fluctuations. This implies that the energy resolution would benefit from relatively heavy weights for the hadronic calorimeter. The side-effect is that the jet response becomes non-linear. The energy-dependent response corrections in the jet energy scale correct for this non-linearity.

The calorimeter calibration, together with the jet energy scale, improves the jet p_{T} resolution by up to 15% for central ($|\eta| < 0.4$) jets from Higgs boson or top quark decays [90].

4.2 Jet energy scale

Typically, physics analyses depend on data-to-Monte Carlo comparisons of reconstructed physics objects. To allow for such comparisons the jet energy scale corrects (on average) the jet energy as measured in the calorimeter back to the jet energy as reconstructed at the stable-particle level.

To avoid as much as possible any dependence on the (possibly imperfect) detector simulation, especially considering the modified Run II detector geometry compared to the pre-Run I testbeam data, a data-driven calibration procedure is employed.

The jet energy scale correction consists of three subcorrections:

$$E_{\mathrm{jet}}^{\mathrm{particle}} = \frac{E_{\mathrm{jet}}^{\mathrm{measured}} - E_{\mathrm{O}}}{R_{\mathrm{jet}} S_{\mathrm{jet}}} \tag{4.1}$$

and corrects the energy as measured from the calorimeter cluster $E_{\mathrm{jet}}^{\mathrm{measured}}$ to that of the underlying particle jet $E_{\mathrm{jet}}^{\mathrm{particle}}$.

- The offset energy E_{O} corrects for energy in the jet that did not originate from the current hard $\mathrm{p\overline{p}}$ interaction. This includes contributions from electronics noise, noise due to radioactive decay in the calorimeter, left-over energy from previous interactions (pile-up) and spurious energy originating from multiple $\mathrm{p\overline{p}}$ interactions in the same bunch crossing.

- R_{jet} represents the energy response of the calorimeter to particle jets. Due to energy lost in material in front of the calorimeter, as well as the presence of uninstrumented detector regions, the response will be significantly less than unity.

- The out-of-cone showering correction S_{jet} corrects for the net flow of energy across the the jet cone boundary due to detector effects. E.g. particles curling into and out of the jet cone due to the magnetic field, and parts of the shower development in the calorimeter crossing the cone boundary.

It should be noted that the underlying event, consisting of the beam remnants and their connections with the hard scatter constituents, is an integral part of the physics process and neither can nor should be corrected for in the energy calibration procedure.

4.2.1 Definitions of the JES subcorrections

The true particle jet energy is defined as the sum of the energies of all particles in the particle jet:

$$E_{\text{jet}}^{\text{particle}} \equiv \sum_{i \in \text{particle jet}} E_i.$$

The jet energy, as measured inside the jet cone, contains contributions from particles present in the particle jet, from particles not present in the particle jet but leaking in due to detector effects, and from offset energy in the jet cone:

$$E_{\text{jet}}^{\text{measured}} \equiv \sum_{i \in \text{particle jet}} E_i^{\text{measured}} S_i + \sum_{i \notin \text{particle jet}} E_i^{\text{measured}} S_i + E_{\text{O}}.$$

Here E_i^{measured} is the visible energy deposited by particle i and S_i represents the fraction of that energy inside the jet cone. This leads to the definition of the 'true offset' energy:

$$E_{\text{O}} \equiv E_{\text{jet}}^{\text{measured}} - \sum_{i \in \text{particle jet}} E_i^{\text{measured}} S_i - \sum_{i \notin \text{particle jet}} E_i^{\text{measured}} S_i. \qquad (4.2a)$$

The 'true response' is defined as the ratio of the measured energy originating from particles in the particle jet to the energy of that particle jet:

$$R_{\text{jet}} \equiv \frac{\sum_{i \in \text{particle jet}} E_i^{\text{measured}}}{E_{\text{jet}}^{\text{particle}}}. \tag{4.2b}$$

Combining the above with the JES master equation, eq. (4.1), leads to the 'true showering' correction:

$$S_{\text{jet}} \equiv \frac{\sum_{i \in \text{particle jet}} E_i^{\text{measured}} S_i + \sum_{i \notin \text{particle jet}} E_i^{\text{measured}} S_i}{E_{i \in \text{particle jet}} E_i^{\text{measured}}}. \tag{4.2c}$$

S_{jet} corrects for the fact that not all energy depositions E_i^{measured} in eq. (4.2b) lie inside the jet cone.

Equations (4.2a)–(4.2c) define the true corrections. Since E_{O}, R_{jet} and S_{jet} are estimated from the data, they may suffer from a number of biases. The showering correction S_{jet} can be calibrated on Monte Carlo. For the response and the offset suitable bias corrections are determined from Monte Carlo. These bias corrections correct the average estimated corrections $\hat{E}_{\text{O}}$ and $\hat{R}_{\text{jet}}$ to the average true corrections E_{O} and R_{jet}. The equivalent of eq. (4.1) using the estimated corrections becomes:

$$E_{\text{jet}}^{\text{corrected}} = \frac{\left(E_{\text{jet}}^{\text{measured}} - \hat{E}_{\text{O}}\right) k_{\text{O}}}{\hat{R}_{\text{jet}} S_{\text{jet}} k_{\text{R}}}$$

where k_{O} and k_{R} represent the bias corrections to respectively the response and the offset corrections. In the rest of this text the 'hats' will be left off the estimated corrections where the meaning is clear from the context.

4.3 Sample selection

Simulation of complicated devices like the D0 detector down to percent-level accuracy is a highly non-trivial exercise. Experience has shown that D0 has not yet reached that level of perfection in the detector simulation. To avoid as much as possible any dependencies on both physics and detector simulation, the jet energy scale employs a data-driven analysis approach. All corrections are determined twice: once for Monte Carlo and once for collider data.

4.3.1 Data

Several data samples were used for the determination of the jet energy scale corrections. All data samples have been selected from the full Run IIa data set and correspond to approximately 1 fb^{-1}. Unless noted otherwise the standard data quality criteria (see section 5.1.1) were applied.

- A photon+jet sample collected using triggers requiring one or two isolated electromagnetic clusters above a given p_{T} is used as the main JES sample. The lowest p_{T} thresholds used is 6 GeV/c. This sample is used to measure the calorimeter response and to determine the showering corrections.

- A di-jet sample, together with the photon+jet sample, is used to determine the η-dependence of the calorimeter response. The di-jet sample reaches higher in jet transverse momentum and is mainly used for the high-p_{T} extrapolation.

- Zero-bias (ZB) and minimum-bias (MB) data samples were used to measure the offset energy.

4.3.2 Monte Carlo

To determine the jet energy scale for simulated events, photon+jet and di-jet Monte Carlo event samples were used. These samples were generated with version 6.323 of the PYTHIA [96] event generator using the CTEQ6L1 [97] parton distribution functions. The underlying event model in PYTHIA was set to 'Tune A' [98]. The photon+jet samples contain the $2 \to 2$ direct photon production processes $\text{p}\overline{\text{p}} \to \text{q}\gamma$ and $\text{p}\overline{\text{p}} \to \text{g}\gamma$ depicted in figure 4.1, the di-jet sample the $\text{p}\overline{\text{p}} \to \text{q}\overline{\text{q}}$, $\text{p}\overline{\text{p}} \to \text{qg}$ and $\text{p}\overline{\text{p}} \to \text{gg}$ QCD processes.

4.4 Offset subtraction

The purpose of the offset subtraction correction is to remove any energy that does not originate from the hard-scatter event. This includes contributions from noise, from pile-up, and from additional $\text{p}\overline{\text{p}}$ interactions within the current bunch crossing. In hadron-hadron collisions the hard scatter takes place between partons from the incoming beam particles, leaving behind the coloured remnants of those beam particles. Those remnants themselves are inherently part of the physics of the hard scatter event

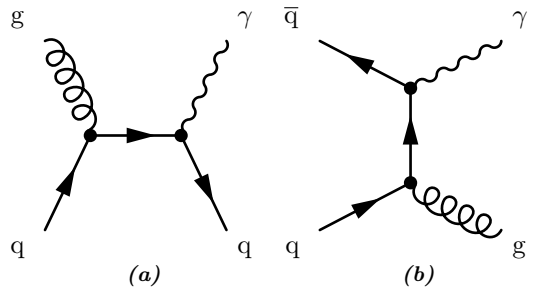


Figure 4.1: *Photon+jet production (at leading-order) through (a) Compton scattering and (b) quark-antiquark annihilation.*

and evolve into an 'underlying event'. Since this underlying event is related to the hard scatter, its energy is not subtracted in the offset correction.

The offset contribution is estimated from a combination of ZB and MB data. One important aspect of the offset subtraction regards the use of data *without* hard interactions, combined with the applied zero-suppression. Hard scatter events after zero-suppression are not simply the sum of the hard scatter energy plus the offset energy; the presence of offset energy on top of the hard scatter event makes parts of the event energy visible that would otherwise have been zero-suppressed. The offset subtraction thus not only removes purely the offset energy, but also the hard scatter energy that becomes visible due to the presence of the offset energy. This effect can only be studied in Monte Carlo and is explicitly corrected for.

The average offset energy $\hat{E}_{\mathrm{O}}$ is estimated for rings in i_η, summed over all towers along i_ϕ in that ring, and parameterised as a function of both the number of reconstructed primary vertices n_{PV} and the instantaneous luminosity L:

$$\hat{E}_{\mathrm{O}}^{\mathrm{ring}}(i_\eta, n_{\mathrm{PV}}, L) = \hat{E}_{\mathrm{NP}}^{\mathrm{ring}}(i_\eta, L) + \hat{E}_{\mathrm{MI}}^{\mathrm{ring}}(i_\eta, n_{\mathrm{PV}}, L).$$

Here $\hat{E}_{\mathrm{NP}}^{\mathrm{ring}}$ and $\hat{E}_{\mathrm{MI}}^{\mathrm{ring}}$ are the contribution from respectively noise and pile-up and from multiple interactions.

4.4.1 Noise and pile-up

The noise contribution to the offset contains both electronics noise in the precision readout and noise due to radioactive decays in the uranium absorbers in the calorimeter. Even though the online electronics calibration corrects the mean of the cell energy

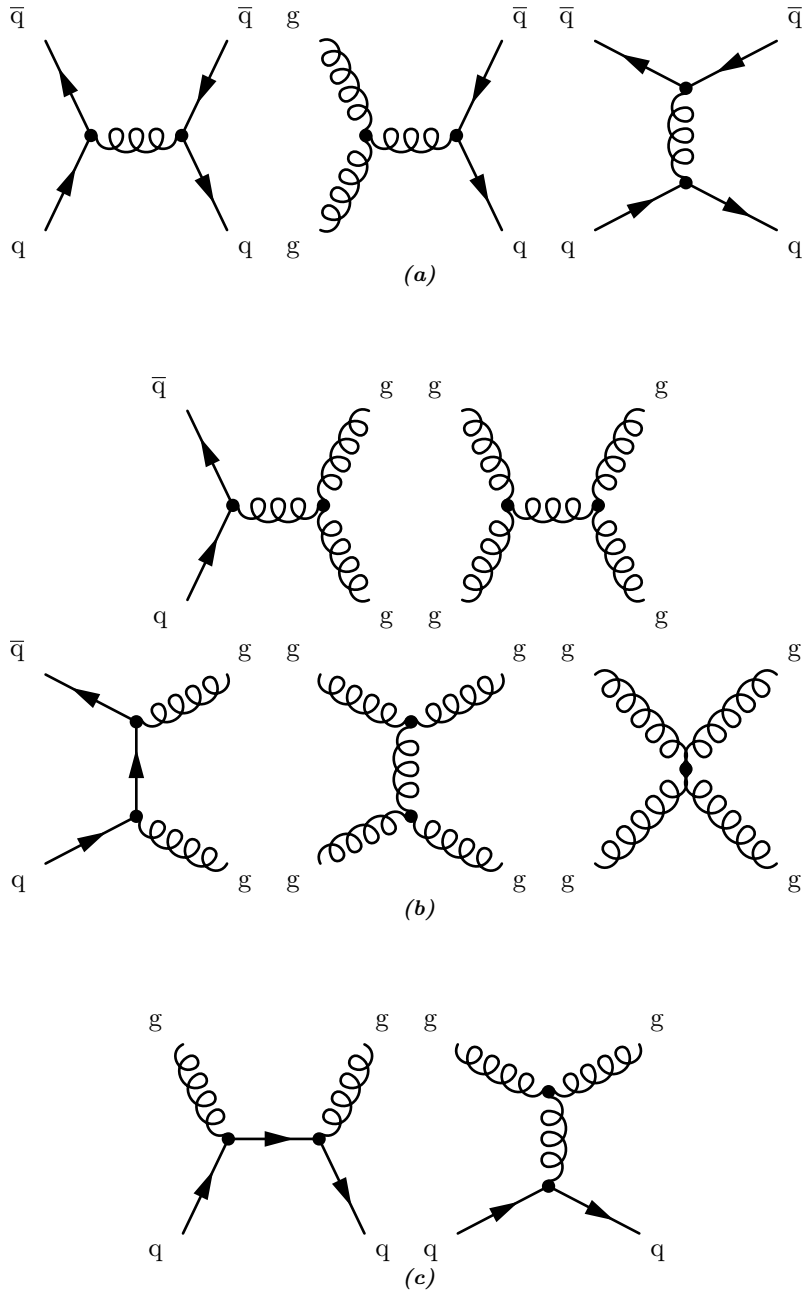


Figure 4.2: *Leading-order contributions to di-jet production resulting in two quark jets, (a), two gluon jets (b) or a quark and a gluon jet (c).*

distributions to zero, the inherent asymmetry of the noise energy distribution together with the effects of (symmetric) zero-suppression and the T42 algorithm leads to a net positive contribution from noise.

Since the 396 ns bunch crossing time in Run II is shorter than the decay time of the signal in the calorimeter electronics ($\approx$ 15 µs) the signal of the current bunch crossing is in general superposed on the remains of signals from previous bunch crossings. The amount of pile-up depends on the instantaneous luminosity of the previous bunch crossings, as well as on the positions of the currently colliding bunches in the superbunches.

The noise and pile-up contributions to the offset energy are estimated from a ZB data sample. By vetoing on any hits in the luminosity monitors as well as on the presence of any reconstructed primary vertices this sample can be depleted of interactions. The remaining events contain only noise and pile-up energy.

Figure 4.3 shows the average noise/pile-up energy densities for all rings in i_η. Since the tower widths depend on i_η the ring areas have been divided out to show the true noise/ pile-up increase for the forward regions. The peak structures around $8 < |i_\eta| < 15$ show the inter-cryostat regions. The low gain for the photomultipliers in the Inter-Cryostat Detectors requires large ADC-to-GeV conversion factors, strongly increasing the noise. As expected the pile-up contributions scale with increasing luminosity, especially in the forward regions of the calorimeter.

4.4.2 Multiple proton-antiproton interactions

The offset energy due to multiple $p\overline{p}$ interactions in a single bunch crossing can be estimated from minimum-bias data, assuming that each reconstructed primary vertex represents one $p\overline{p}$ interaction. Using MB events the average energy $\hat{E}_{\mathrm{MB}}$ per i_η ring is determined. The overall energy difference between minimum-bias events with n_{PV} reconstructed vertices and single-vertex events is considered to represent the energy originating from $N-1$ additional interactions:

$$\hat{E}^{\mathrm{ring}}_{\mathrm{MI}}(i_\eta, n_{\mathrm{PV}}, L) = \hat{E}^{\mathrm{ring}}_{\mathrm{MB}}(i_\eta, n_{\mathrm{PV}}, L) - \hat{E}^{\mathrm{ring}}_{\mathrm{MB}}(i_\eta, n_{\mathrm{PV}} = 1, L).$$

The dependence on the luminosity L reflects the $\approx 5\%$ spread of minimum-bias energy in events with the same vertex multiplicity.

The average offset energy due to multiple interactions is shown in figure 4.4. The dependence on the primary vertex multiplicity is almost perfectly linear. Linear fits

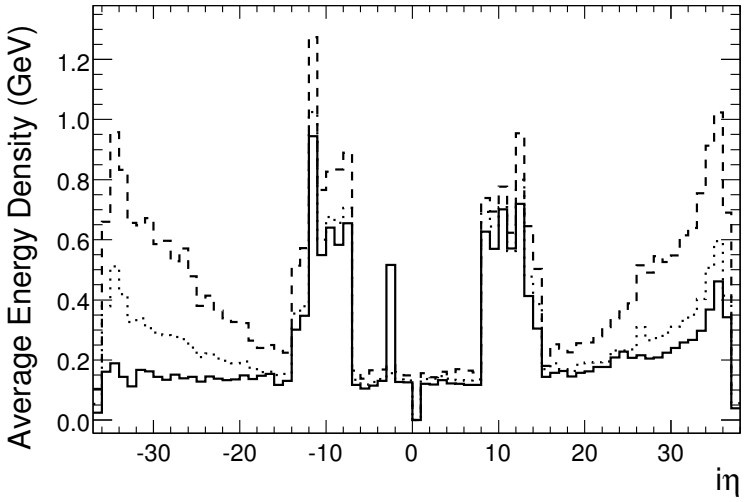


Figure 4.3: *The average energy density in the 'interaction depleted' zero-bias sample for different instantaneous luminosities: $L = 0.1 \times 10^{32}\ cm^2 s^{-1}$ (solid), $L = 0.5 \times 10^{32}\ cm^2 s^{-1}$ (dotted), $L = 1.2 \times 10^{32}\ cm^2 s^{-1}$ (dashed) [93]. The 'dip' between $i_\eta = 0$ and $i_\eta = 1$ is a binning effect showing the non-existent tower index $i_\eta = 0$. The forward-backward asymmetry in offset energy originates from the different proton-antiproton bunch sizes. The smaller antiproton bunches, travelling in negative z-direction, lead to less beam halo and beam-beam interactions, and correspondingly contribute less to the offset energy.*

are performed and the results used for the offset estimates. Figure 4.4b demonstrates that each additional interaction contributes equally to the offset energy (density).

4.4.3 Per-jet offset energy subtraction

To apply the offset correction to individual jets the per-ring parameterisation is converted to a per-jet offset energy estimate. Following the definition of jets as used in the cone jet algorithm used in D0, jets are approximated as occupying a circular area in (y, ϕ)-space with radius $\mathcal{R}_{\text{cone}}$ around the jet axis. The estimated offset energy inside the jet $\hat{E}_{\text{O}}$ is obtained from the sum over all towers in that circle: [1]

$$\hat{E}_{\text{O}} = \sum_{i_\eta \in \mathcal{R}_{\text{cone}}} \hat{E}_{\text{O}}^{\text{ring}}(i_\eta, n_{\text{PV}}, L) f^{\text{tower}}(i_\eta, \eta_{\text{jet}}^{\text{d}}),$$

[1] Jet detector pseudorapidity η_{d} is also written as η^{d} or η^{det}.

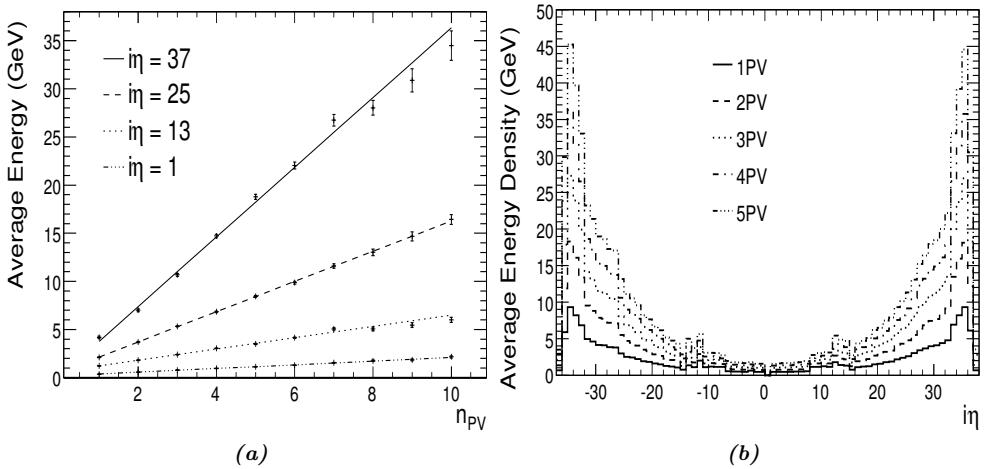


Figure 4.4: *(a) The average minimum-bias energy in various i_η rings as a function of the number of primary vertices n_{PV} (a) and as a function of i_η for various primary vertex multiplicities (b) [93].*

where f^{tower} represents the fraction of towers in the i_η ring that fall inside the jet cone. This depends solely on the detector geometry (through i_η) and on the position of the jet axis (through $\eta^{\text{d}}_{\text{jet}}$).

Figure 4.5 shows the total offset energy corrections for both $\mathcal{R}_{\text{cone}} = 0.5$ and $\mathcal{R}_{\text{cone}} = 0.7$ cone jets as a function the jet position $\eta^{\text{d}}_{\text{jet}}$.

4.4.4 Correcting for the zero-suppression bias

As already mentioned, the presence of offset energy underneath the energy from the hard scatter event can lead to additional energy from the hard scatter passing zero-suppression and becoming visible. This effect is particularly strong in regions with relatively little hard scatter energy, e.g. on the edges of jets, and can lead to significant deviations from the expected offset energy inside jets. Using a Monte Carlo sample of photon+jet events reconstructed with and without overlaid ZB data to simulate offset energy, the size of this effect can be estimated.

Back-to-back photon+jet events (ΔR(photon, jet) > 3) are selected and matched between the two samples. Only events in which the jets in the samples with and without overlay could be matched to within $\Delta R < \mathcal{R}_{\text{cone}}/2$ were used. The zero-suppression bias correction k_{R} can be computed from the average jet energy without

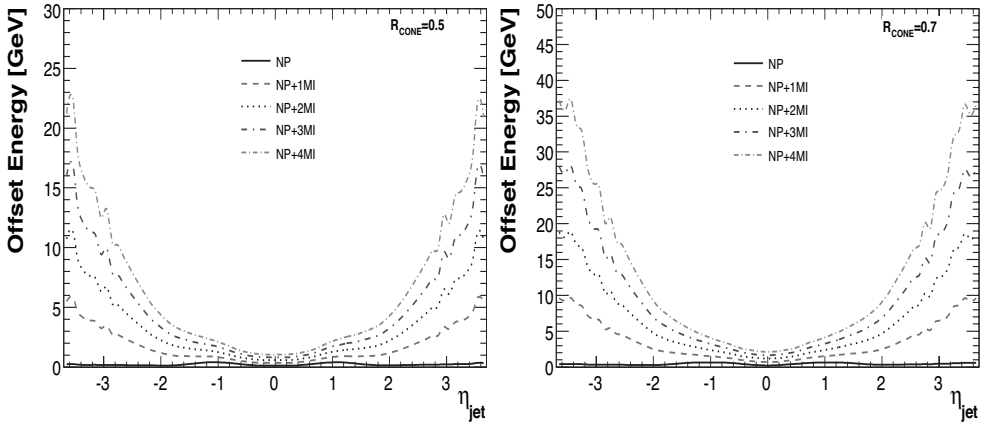


Figure 4.5: *Offset energy corrections for (a) $\mathcal{R}_{cone} = 0.5$ and (b) $\mathcal{R}_{cone} = 0.7$ cone jets as a function of the jet position η^d_{jet} [93].*

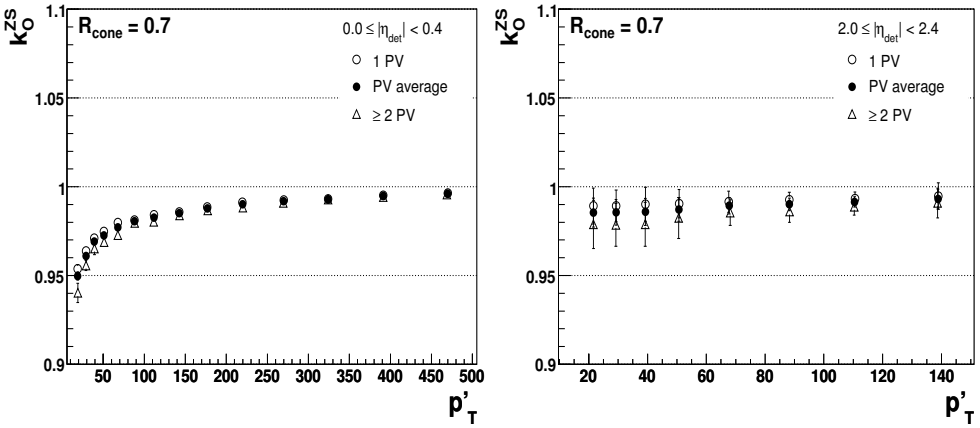
ZB overlay $E_{\text{jet}}^{\text{meas, no ZB}}$, and the average estimated jet energy after offset correction $E_{\text{jet}}^{\text{meas}} - \hat{E}_{\text{O}}$:

$$k_{\text{O}}^{\text{ZS}} = \frac{E_{\text{jet}}^{\text{meas, no ZB}}}{E_{\text{jet}}^{\text{meas}} - \hat{E}_{\text{O}}}.$$

The size of the bias correction (figure 4.6) is limited by the amount of offset energy present such that the (multiplicative) correction decreases with increasing jet energy. Zero-bias effects also lead to a dependence on the vertex multiplicity. The presence of additional interactions increases the amount of available offset energy underneath the jet. This vertex multiplicity dependence was not statistically significant enough to be parameterised; instead, the correction corresponding to the average multiplicity in the overlaid ZB sample ($\bar{n}_{\text{PV}} = 1.5$) was used and the observed dependence assigned as a systematic uncertainty. It should be noted that the offset bias correction, unlike the offset correction itself, depends on the jet energy.

4.4.5 Uncertainties

Due to the high available statistics in the ZB and MB data samples the uncertainty on the offset correction is completely dominated by any uncertainties on the bias



(a) *(b)*

Figure 4.6: *Bias corrections on the jet offset energy due to effects of zero-suppression [93]. The full circles correspond to the correction estimated for the average vertex multiplicity in the ZB sample. p'_T Represents the expected jet p_T based on the jet direction and the photon p_T, see also section 4.5.2.*

correction factor $k_{\mathrm{O}}^{\mathrm{ZS}}$. These uncertainties, shown in figure 4.7, cover the treatment of the vertex multiplicity dependence and the choice of the ΔR(jet, jet) matching cut between events with and without ZB overlay. The central value of the offset correction uses the average vertex multiplicity in the ZB sample. The difference between the $n_{\mathrm{PV}} = 1$ and $n_{\mathrm{PV}} \geq 2$ cases is assigned as a symmetric uncertainty. The effect of the ΔR cut is estimated by comparing the two cases $\Delta R < \mathcal{R}_{\mathrm{cone}}/2 \pm 0.1$. Again the difference is assigned as a systematic uncertainty.

4.5 Calorimeter response

The response measurements employ the 'missing transverse energy projection (MPF)' method [99]. The MPF method is based on the particle-level p_{T} balance of back-to-back tag-and-probe events. At detector level not the true transverse momenta are measured but their convolution with the detector response R: $p_{\mathrm{T}}^{\mathrm{measured}} = R p_{\mathrm{T}}^{\mathrm{particle}}$. Any difference between the tag and probe responses will present itself in the form of

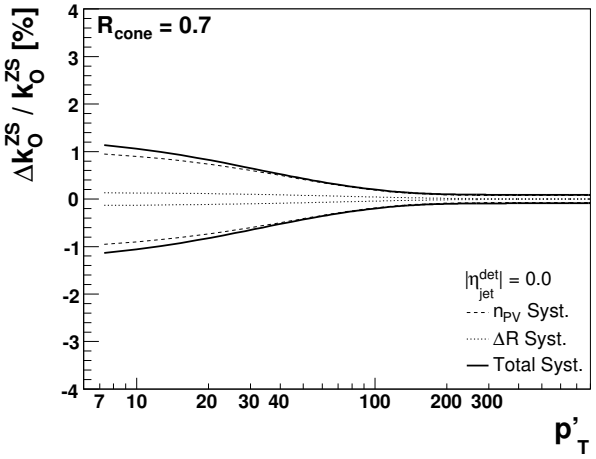


Figure 4.7: *Systematic uncertainties affecting the offset bias correction for very central jets [93]. The shape of the uncertainties does not change for forward jets, only the narrowing of the band moves towards higher p'_T values (from ≈ 100 GeV/c at $\eta^d_{jet} = 0$ to ≈ 200 GeV/c at $\eta^d_{jet} = 2$).*

missing transverse energy $\not\!\!E_{\mathrm{T}}$:

$$R_{\mathrm{tag}}\vec{p_{\mathrm{T}}}_{\mathrm{tag}} + R_{\mathrm{probe}}\vec{p_{\mathrm{T}}}_{\mathrm{probe}} = -\vec{\not\!\!E}_{\mathrm{T}}.$$

Assuming a perfectly calibrated tag object the probe response can be measured by projecting $\not\!\!E_{\mathrm{T}}$ onto the tag object in the transverse plane:

$$R_{\mathrm{probe}} = 1 + \frac{\vec{\not\!\!E}_{\mathrm{T}} \cdot \vec{p_{\mathrm{T}}}_{\mathrm{tag}}}{\vec{p_{\mathrm{T}}}^{\,2}_{\mathrm{tag}}}.$$

While at the particle level the probe consists of the full hadronic recoil against the tag, in the detector the probe consists of the reconstructed jet. By requiring only a single tag object and a single probe jet, tightly back-to-back ($\Delta R > 3.0$), the approximation can be made that the jet corresponds to the hadronic recoil, $R_{\mathrm{jet}} \approx R_{\mathrm{recoil}}$. The remaining bias due to the above assumption will be corrected for in the MPF bias correction (section 4.5.4).

For the JES response correction measurement both back-to-back photon+jet and di-jet events were used.

4.5.1 Photon energy scale

The electromagnetic energy scale is calibrated using a $Z \to e^+e^-$ sample, resulting in optimal calibration for electrons/positrons of $E \approx 45$ GeV. Since photons interact less with material than electrons this calibration is not directly applicable to photons. The difference between photon and electron energy scales can only be studied in Monte Carlo. A photon energy correction is available based on the default D0 detector simulation. Since then, the D0 W-mass group has developed an improved version of the detector simulation [100]. Apart from a better description of electromagnetic showers one of the main improvements is the addition of 'fudge' material in front of the calorimeter to improve the agreement with $Z \to e^+e^-$ data [101]. The preferred value for the amount of extra material is ≈ 0.28 radiation lengths. The photon energy scale difference between the improved simulation and the default was applied as additional correction (figure 4.8). The difference between two extreme amounts of fudge material was assigned as a systematic uncertainty on the photon energy scale.

The electron energy scale itself is well-known. However, the reliance of the photon energy scale on detector simulation (with the above mentioned imperfections) makes this one of the dominant sources of uncertainty on the final JES.

4.5.2 Absolute response

The absolute jet response in both data and MC is measured using the MPF method on photon+jet events. Only events with one or two good reconstructed primary vertices were selected. Each event is required to contain exactly one photon with $p_T > 7$ GeV/c within the central calorimeter $|\eta^d| < 1.0$. Photons in the data sample are required to have a p_T in the high efficiency region of the fired trigger, and to match within $\Delta R < 0.4$ with the L1 EM trigger cluster. Events with more than one reconstructed jet are rejected. [2] The jet has to be within $|\eta^d| < 0.4$ and back-to-back with the photon: $\Delta R > 3.0$. The jet centrality requirement was chosen such that even $\mathcal{R}_{cone} = 0.7$ cone jets are well contained within the central calorimeter. Any events with cosmic muon candidates were rejected. To further suppress cosmic events, as well as backgrounds from e.g. $Z/\gamma^* \to e^+e^-$ and $W(\to e\nu)$ + jet an additional upper limit on the missing transverse energy was applied. To avoid biasing the jet response this E_T threshold scaled with the photon p_T.

[2] Since many jets are also reconstructed as photons, an additional jet overlapping with the photon is allowed.

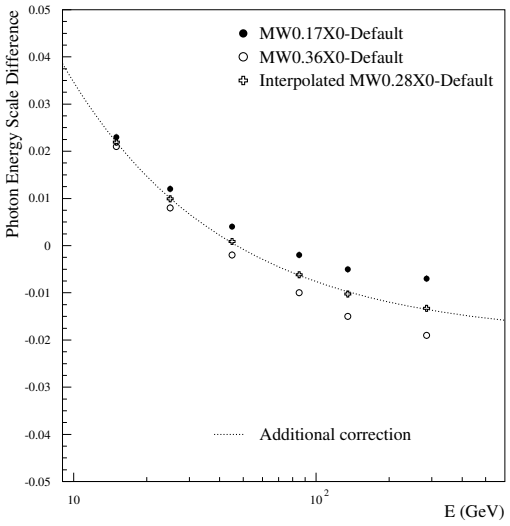


Figure 4.8: *Relative difference between the photon energy scale as determined using the default D0 detector simulation and the improved version used by the W-mass group [93]. The open and full circles represent the two extreme values of additional material. The plus markers show the interpolated correction corresponding to the most likely value of* $\approx 0.28\ X_0$*. The fit to these points (dotted line) is used as additional photon energy scale correction.*

Due to the steeply falling jet energy spectrum, upward fluctuations will dominate downward ones, resulting in a net positive bias. To suppress this effect [3] the response corrections are parameterised as a function of the expected jet energy E' (or transverse momentum p'_{T}) estimated from the photon transverse momentum $p_{\mathrm{T}\gamma}$ and the jet pseudorapidity η_{jet}:

$$E' \equiv p_{\mathrm{T}\gamma} \cosh\left(\eta_{\mathrm{jet}}\right), \tag{4.3a}$$

$$p'_{\mathrm{T}} \equiv \frac{E'}{\cosh\left(\eta_{\mathrm{jet}}^{\mathrm{det}}\right)}. \tag{4.3b}$$

[3] The remaining resolution bias is negligible (less than 1%) in the case of photon+jet events. For di-jet events it is of the order of 3–15% and is explicitly corrected for, see ref. [93].

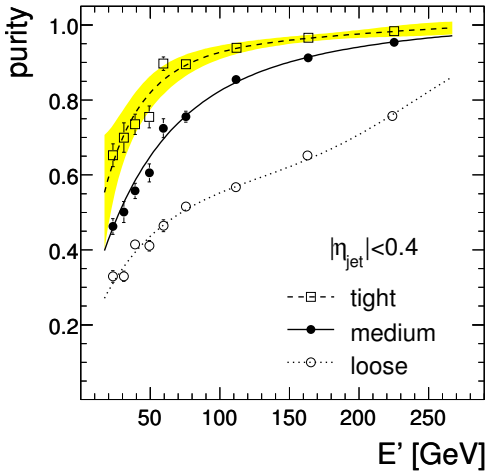


Figure 4.9: *Purity of the photon+jet sample in data for central jets [93]. The different curves correspond to three different sets of photon selection criteria: tight (dashed), medium (solid) and loose (dotted). The shaded band around the 'tight' curve shows the total uncertainty on the purity of the tight photon+jet sample.*

As described in section 3.2.3, the photon selection criteria include stringent quality cuts; nevertheless, the photon+jet sample contains a significant fraction of QCD di-jet contamination, typically due to jets with leading neutral pions. The different quark/gluon compositions of jets in photon+jet and di-jet events lead to different response corrections. Whereas photon+jet events at low p_{T} are dominated by quark jets, the gluon fraction increasing with increasing p_{T}, di-jet events show the opposite behaviour. Gluon jets tend to be wider, containing more particles with a correspondingly lower average particle energy. Both the tracker magnetic field as well as the presence of dead material lower the response of low-energy particles, resulting in an overall lower response for gluon jets compared to quark jets. In addition, since jets are wider than photons, the additional hadronic energy around the jets misidentified as photons reduces the $\not\!\!E_{\mathrm{T}}$ in the direction of the probe jet, increasing the measured MPF response in the contaminated photon+jet sample.

The photon purity was estimated for three different sets of photon selection criteria (figure 4.9) in both Monte Carlo and data (using a template-based method). Measured purities in data and MC were found to be in excellent agreement.

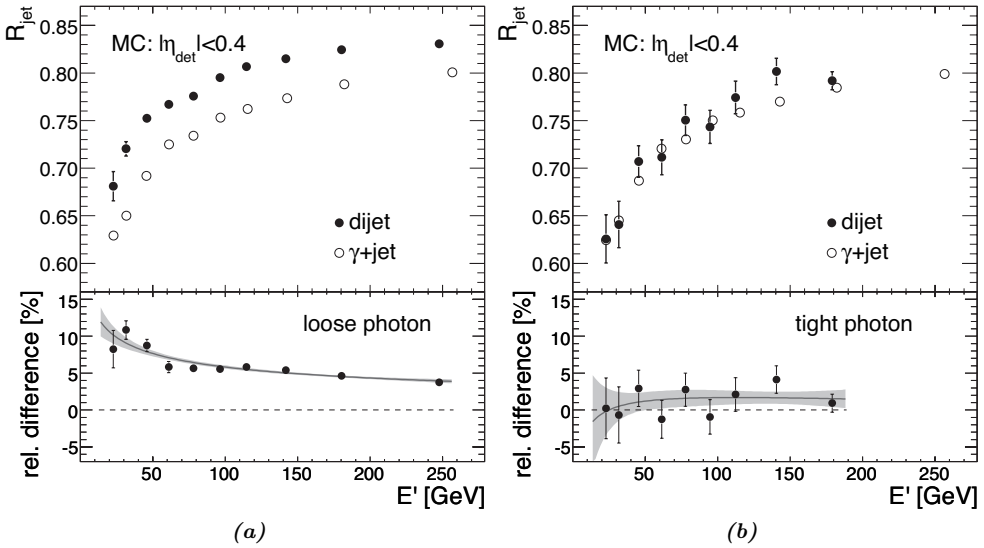


Figure 4.10: *Differences between the absolute jet responses estimated from photon+jet and di-jet events passing the photon+jet event selection [93], for the loose (a) and for the tight (b) photon selection criteria. For the tight photon selection criteria the effects of the di-jet contamination on the measured response is* $\approx 2\%$.

Figure 4.10 demonstrates the effect of di-jet contamination of the photon+jet sample on the measured response for both the loosest and the tightest photon selection criteria, confirming these expectations. Corrections for the di-jet contamination were studied in all three photon samples, bringing the responses back to the same value in all cases. The final response correction (figures 4.11 and 4.12) is based on the tight photon selection, giving the highest purity and the smallest purity correction.

Figure 4.11 shows the absolute response corrections fitted using a quadratic log fit:

$$R(E') = p_0 + p_1 \log\left(E'/E_0\right) + p_2 \log^2\left(E'/E_0\right) \tag{4.4}$$

where $E_0 = 100$ GeV. The uncertainties at high jet energies are dominated by the extrapolation of the fit results. To reduce these uncertainties a dedicated Monte Carlo study was performed in which an energy dependent scaling of the single pion response in the Monte Carlo was introduced to reproduce the jet response in data. This scaling was tuned to the data using a χ^2 fit. The resulting response curve (figure 4.12) has the

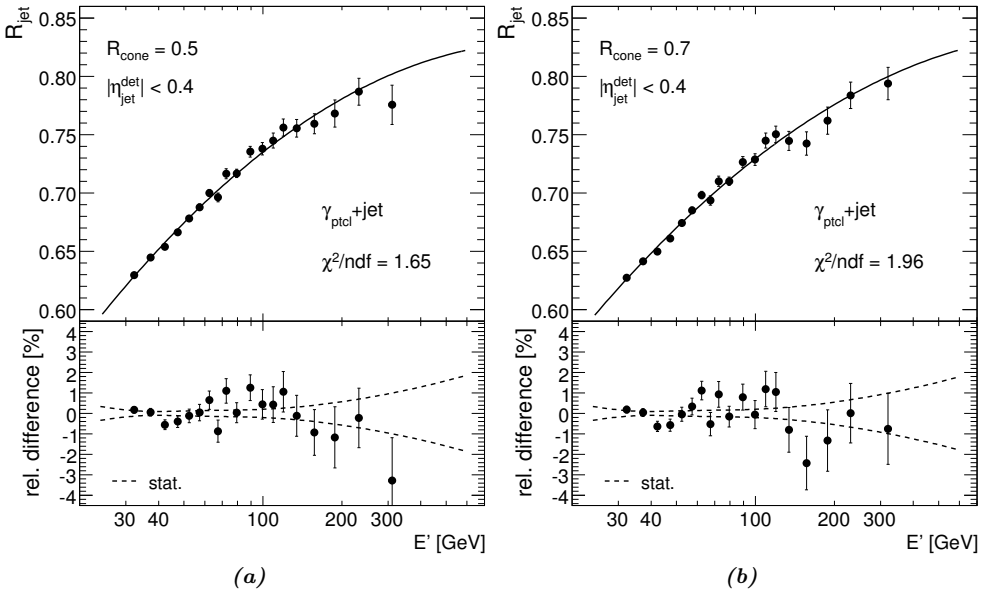


Figure 4.11: *Absolute response corrections (fitted using the quadratic log fit from eq. (4.4)) for (a) $\mathcal{R}_{cone} = 0.5$ and (b) $\mathcal{R}_{cone} = 0.7$ cone jets as a function of expected jet energy E' [92, 93].*

same central value as the fit in figure 4.11b but significantly smaller uncertainties. This shows that the 'tuned pion response' properly models the physics processes involved, e.g. the changing quark/gluon content in the photon+jet sample as a function of rising energy, and the fragmentation differences between quarks and gluons.

The size of the jet response corrections, $\approx 25\%$ at 100 GeV, is dominated by the limited integration time of the calorimeter preamplifiers.

4.5.3 η-Dependent response corrections

While the D0 calorimeter is fairly uniform within the central and endcap regions, the gaps between the cryostats are instrumented differently. Also, the amount of dead material depends on jet pseudorapidity. The relative response correction F_η calibrates forward jets with respect to central jets. Together with the absolute detector response this provides the response correction for forward jets.

The relative response is measured using the MPF method on both photon+jet and di-jet events. The tag photon (jet) is required to be within $|\eta| < 1.0$ ($|\eta| < 0.4$),

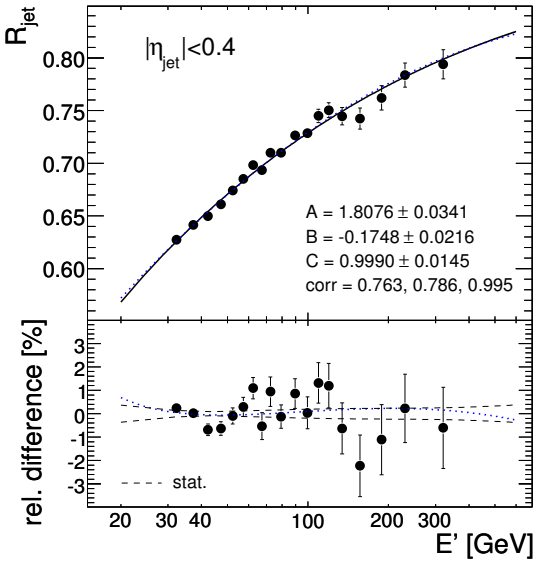
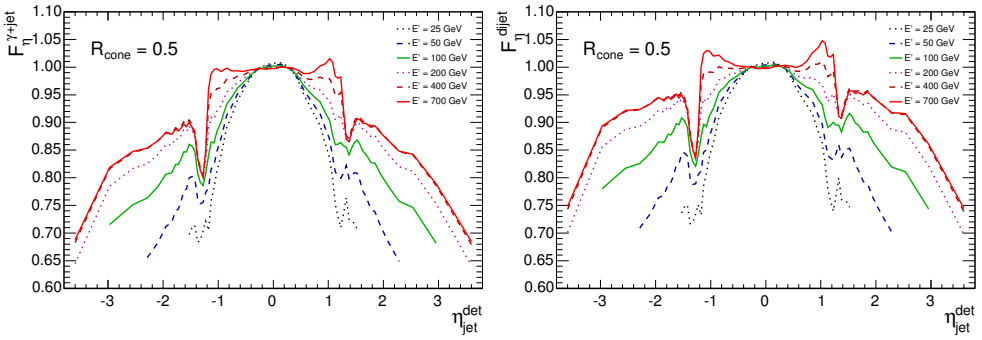


Figure 4.12: *Absolute response corrections for $\mathcal{R}_{cone} = 0.7$ cone jets as a function of expected jet energy E', fitted based on a Monte Carlo sample with tuned single pion response [93]. The bottom insets show the fit residuals (solid circles) and the assigned fit uncertainty (dashed lines). The dotted lines provide a comparison with the quadratic log fit result shown in figure 4.11b.*

the probe jet is unrestricted in pseudorapidity. Tag and probe are required to be back-to-back: $\Delta\phi > 3$. The di-jet sample reaches higher in jet transverse momentum and is used to determine the shape of the response corrections at high p_{T}. Figure 4.13 shows a summary of the relative response corrections for $\mathcal{R}_{\mathrm{cone}} = 0.5$ cone jets in both photon+jet and di-jet events for a wide range of jet energies. Since forward jets traverse more material before reaching the calorimeter, the relative quark-gluon jet differences becomes stronger for forward jets. The overall η-dependence decreases slightly with increasing jet energy.

4.5.4 MPF bias correction

Reconstructed photon clusters include less calorimeter cells than jets, making them less sensitive to zero-suppression effects. The additional offset energy inside the jet artificially increases the measured response. This effect counter-balances the zero-suppression bias on the offset discussed in section 4.4.4 and the bias correction k_{R} is



(a) *(b)*

Figure 4.13: *Relative response corrections F_η in data (a) for photon+jet and (b) for di-jet events [93]. In the central region the correction is unity by construction, the 'troughs' reflect the large corrections required for the Inter-Cryostat Detectors. The general trend of the curves going from central to forward regions is governed by forward jets traversing increasingly more material before reaching the calorimeter.*

estimated similarly in matched photon+jet samples reconstructed with and without ZB overlay.

Parameterisation of both bias corrections as a ratio k_O/k_R allows cancellation of many uncertainties, especially those related to the estimation of these corrections from Monte Carlo. The magnitude of this bias correction ratio is typically less than 1%, except for $|\eta_d| > 3$ where the remaining correction can be as large as 3%. The relative uncertainties on the ratio are dominated by the primary vertex multiplicity systematic described in section 4.4.4, and only exceed 1% below $p_T \approx 10$ GeV/c.

4.5.5 Response uncertainties

The uncertainties on the response correction, shown in figure 4.14, are dominated by the photon purity at low p_T, and the photon energy calibration, above $p_T \approx 30$ GeV/c. It should be noted that the response corrections for $\mathcal{R}_{cone} = 0.7$ cone jets are known to a precision of $\approx 1.3\%$ up to the highest jet energies. For $\mathcal{R}_{cone} = 0.5$ cone jets the additional step involving tuning of the single pion response was not performed, resulting in larger uncertainties (see also section 4.7). Different fragmentation models lead to different jet hadron spectra. This effect was studied by comparing the PYTHIA and HERWIG event generators. Uncertainties on (especially the gluon) parton distribution

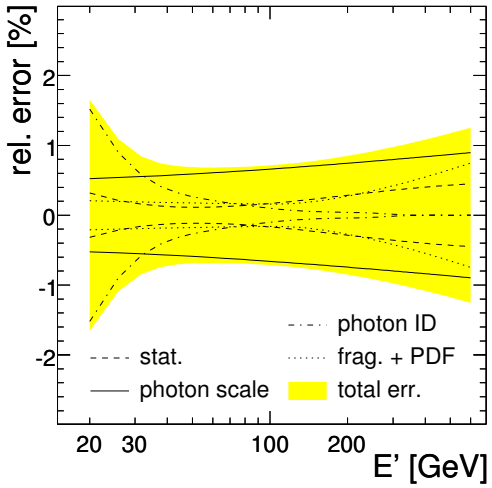


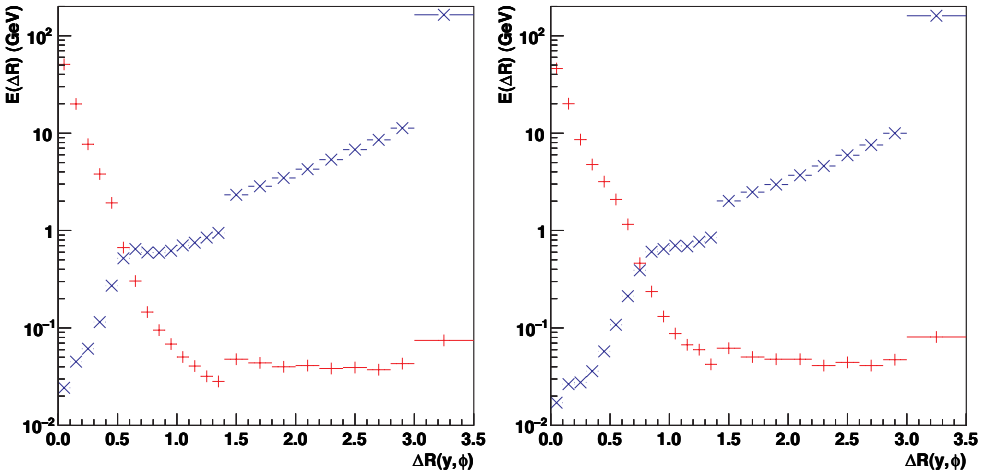
Figure 4.14: *Relative uncertainties on the response correction for $\mathcal{R}_{cone} = 0.7$ cone jets [93]. The uncertainties for $\mathcal{R}_{cone} = 0.5$ cone jets are almost identical, except for slightly larger statistical uncertainties and larger uncertainties on the high-energy extrapolation.*

functions result in an uncertainty on the jet parton flavour, estimated using the CTEQ6.1M PDF set, following the CTEQ group prescription [97].

4.6 Out-of-cone showering correction

The cone jet algorithm reconstructs the jet energy as the sum of all energy present within the jet cone. Due to detector effects, e.g. particles bending in the magnetic field or parts of the shower leaving the jet cone in the calorimeter, the energy inside the cone may differ from the true (particle-level) jet energy. The out-of-cone showering correction compensates for the net energy difference due to these detector showering effects [92, 102]. In this procedure it is crucial to understand the spatial distribution of energy deposited by an 'average jet'; therefore jet energy distributions are studied in Monte Carlo and compared to data.

Using back-to-back photon+jet events in Monte Carlo without any noise and/or multiple interactions the energy deposited by each particle can be tracked throughout the detector simulation. Selecting events with exactly one reconstructed offline jet matching a single particle-level jet, and distinguishing between particles assigned to the particle jet and other (non-jet) particles allows for the creation of jet and non-jet



(a) *(b)*

Figure 4.15: *Typical spatial jet (+) and non-jet (×) energy distributions as a function of the radial distance with respect to the jet axis $\Delta\mathcal{R}$. Central ($|\eta| < 0.4$), $100 < p'_T < 130$ GeV/c jets in photon+jet events; (a): $\mathcal{R}_{cone} = 0.5$ cone, (b): $\mathcal{R}_{cone} = 0.7$ cone jets. The discontinuities at $\Delta\mathcal{R} = 1.4$ and $\Delta\mathcal{R} = 3.0$ are caused by changes in bin size.*

energy deposition profiles as a function of the radial distance $\Delta\mathcal{R}$ away from the jet axis (an example is shown in figure 4.15). These profiles represent the average energy deposited in rings of given width at a distance $\Delta\mathcal{R}$ from the jet.

Even though it should be stressed that the treatment of jets as circular objects is incorrect on a per-jet basis, on average it can provide valuable insights into jet shapes and reconstruction effects, as demonstrated in figure 4.15:

- Particles in the particle jet deposit energy as far away as $\Delta\mathcal{R} > 3$ from the jet axis (beyond $\Delta\mathcal{R} = \pi$ the plateau quickly falls off, this is not clear from figure 4.15 due to the chosen binning and axis range). This shows the contribution of low-momentum particles curving around the detector in the solenoidal field. Similarly, all the way down to the core of the jet, energy is deposited by particles not part of the particle jet.

- The difference in shape between the jet profiles for $\mathcal{R}_{\text{cone}} = 0.5$ and $\mathcal{R}_{\text{cone}} = 0.7$ cone jets demonstrates that the same algorithm, applied with different cone sizes, reconstructs different jets, with different energies. Even if the 'same' jets are

reconstructed with both cone sizes the resulting jets differ in (average) shape and energy.

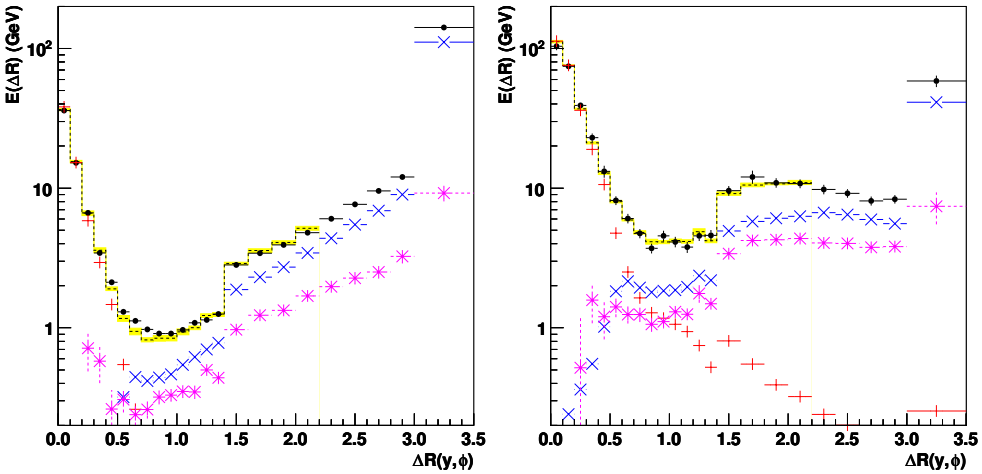
- The energy leaking into and out of the jet cone (non-jet and jet profiles respectively) shows a remarkable symmetry around $\Delta\mathcal{R} \approx \mathcal{R}_{\text{cone}}$.
- The increase in the non-jet profile with increasing $\Delta\mathcal{R}$ represents the increasing number of towers included in larger-radius annuli in combination with a uniform energy distribution. The photon is contained in the last bin, at $3.0 < \Delta\mathcal{R} < 3.5$.
- For both $\mathcal{R}_{\text{cone}} = 0.5$ and $\mathcal{R}_{\text{cone}} = 0.7$ a 'drop' in the profile (i.e. a step in the average energy) is visible just outside $\Delta\mathcal{R} = \mathcal{R}_{\text{cone}}$. This demonstrates an artifact of the cone jet reconstruction algorithm: the iterative procedure maximises the amount of energy inside the jet cone, correspondingly minimising the energy just outside the cone.

A template fit [4] to the data (after offset subtraction) is used to estimate the true energy distributions from particles inside/outside the jet. To avoid dependencies on the photon modelling, the fit is limited to $\Delta\mathcal{R} < 2.5$. The net showering correction is defined as the ratio between the jet energies given by the jet algorithm and by the fitted templates:

$$S = \frac{\int_0^{\mathcal{R}_{\text{cone}}} \left(E(\text{jet}) + E(\text{non-jet})\right)}{\int_0^{\infty} E(\text{jet})}$$

Figure 4.16 shows a pair of example fit result for $\mathcal{R}_{\text{cone}} = 0.5$ cone jets. Good agreement is observed over a wide kinematic range [93], especially considering these fits represent a comparison of the full spatial energy deposition in photon+jet events. There is, however, a clear trend in data-to-MC agreement: low p_{T} and forward jets show visibly less agreement than 'moderate-p_{T}' central jets [102]. This is most likely a combination of imperfections in the physics modelling of e.g. soft jets, and in the D0 detector simulation (e.g. detailed material descriptions become more important for forward jets). Not surprisingly the fit agreement is slightly worse around the jet edge, $\Delta\mathcal{R} \approx \mathcal{R}_{\text{cone}}$, as can be seen in figure 4.16a. This region is dominated by the interplay between energy depositions and zero-suppression. Ignoring this region around the jet edge in the fit does not significantly change the fit results. A uniform uncertainty was

[4] A χ^2 fit was used, taking into account the limited statistics in the energy profiles.



(a) *(b)*

Figure 4.16: *Fit of the jet (+) and non-jet (×) energy distributions to data, (a) for central ($|\eta| < 0.4$) and (b) for forward ($2.0 < |\eta| < 2.4$) $\mathcal{R}_{cone} = 0.5$ cone jets ($80 < p'_T < 100$ GeV/c). The full circles represent the data, the dashed line the sum of the fitted templates. The shaded band around the fit results reflects the statistical fit uncertainty. The offset energy profile (∗) is kept fixed in the fit. The strikingly different large-$\Delta\mathcal{R}$ behaviour in forward, compared to central, jets is due to the profile nearing the end of the detector in rapidity direction.*

adopted to account for the differences in fit agreement over the different pseudorapidity ranges.

One of the benefits of this fitting procedure is the lack of assumptions on where jets 'end', i.e. no ad hoc geometrical 'jet limits' are required. At the same time the fit represents a built-in validation of the procedure based on the agreement between the Monte Carlo and data energy deposition profiles. In Monte Carlo it is possible to determine the 'true' showering by tracking all energy depositions. This feature was exploited to calibrate the template-based showering approach for small biases (typically less than 0.5%).

This template method is sensitive to the relative quark-gluon content of the data sample through the shape of the energy profile. By separating the photon+jet Monte Carlo sample (based on the MC truth) into separate quark-jet and gluon-jet samples, the difference between the average quark- and gluonjet profile shapes can be studied.

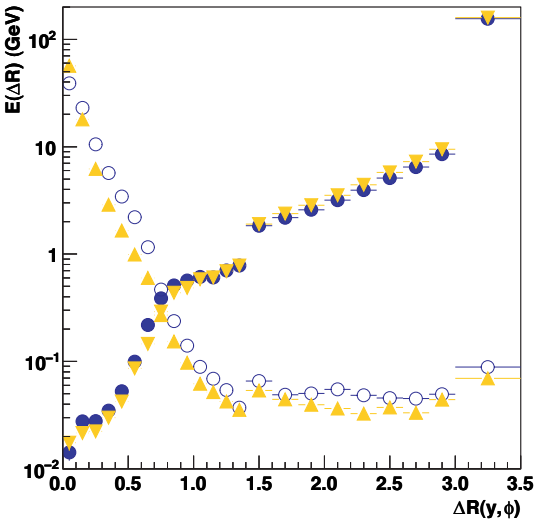
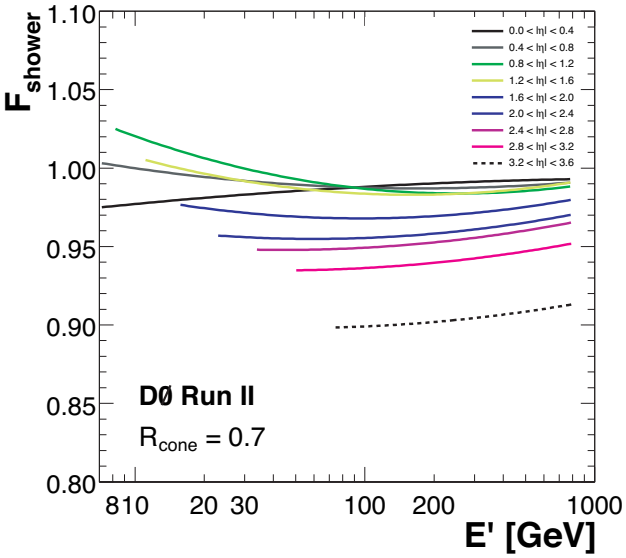


Figure 4.17: *The difference between quark (▲, ▼ markers) and gluon (○, • markers) jets in photon+jet events ($\mathcal{R}_{cone} = 0.7$ cone jets, $|\eta| < 0.4$, $100 < p'_T < 130$ GeV/c). The p'_T binning is based on the photon p_T (see eq. (4.3)), so the right-most bins, containing the photon, overlap. Quark jets show a steeper, narrower jet profile than gluon jets.*

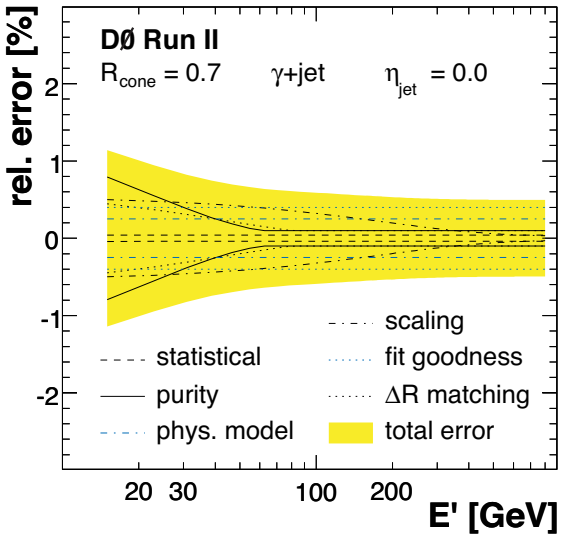
Figure 4.17 shows a typical comparison between quark- and gluonjet profiles for central $\mathcal{R}_{\text{cone}} = 0.7$ cone jets. Gluons experience twice as many colour connections as quarks do, leading to higher average particle multiplicities and lower average particle momenta. This results in wider gluon jets, with more low-momentum particles straying away from the jet axis.

Such shape differences are especially important in connection with the ΔR matching cut applied to relate reconstructed jets to particle-level jets. Tighter matching preferentially rejects wide jets, slightly distorting the jet profile shape. Inspection of the PYTHIA event record shows that predominantly gluon jets are removed. For consistency among all JES analyses a matching cut of $\Delta R < \mathcal{R}_{\text{cone}}/2$ was adopted. This matching cut was varied by ± 0.1 and the effect was included as a systematic uncertainty (marked 'ΔR matching' in figure 4.18b).

Figure 4.18a presents the out-of-cone showering corrections for $\mathcal{R}_{\text{cone}} = 0.7$ cone jets. Since the out-of-cone showering correction is dominated by edge effects the correction is slightly larger for $\mathcal{R}_{\text{cone}} = 0.5$ cone jets. Similarly the correction becomes more prominent for jets at larger pseudorapidities where the physical size of jets decreases



(a)



(b)

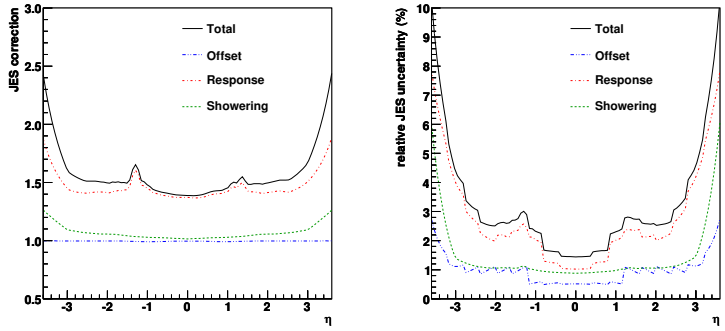
Figure 4.18: *Out-of-cone showering corrections (a) and the corresponding uncertainties (b) for central $\mathcal{R}_{cone} = 0.7$ cone jets in different pseudorapidity ranges. The dominant uncertainties are all related to the shape of the jet energy profile and its agreement with data. The uncertainty marked 'scaling' reflects the effect of down-scaling the single pion response in Monte Carlo to match the low-p_T jet response between data and Monte Carlo.*

(with respect to the calorimeter granularity). The dominant systematic uncertainties to the out-of-cone showering corrections (figure 4.18b) are all related to how well the true jet energy profiles are described by the MC templates. Impurities in the photon+jet sample modify the relative quark/gluon content which is translated directly into the shape of the jet profile. The same holds for the underlying physics models for fragmentation and hadronisation as well as for the underlying event modelling. These effects were studied in Monte Carlo to estimate corresponding systematic uncertainties.

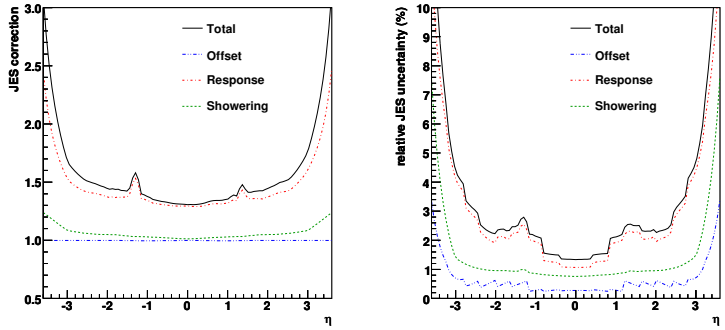
4.7 Combined JES corrections

Figures 4.19 and 4.20 show the combined jet energy scale corrections and corresponding uncertainties for $\mathcal{R}_{\text{cone}} = 0.5$ cone jets (originating from the detector centre) as functions of both jet pseudorapidity and transverse momentum. For different and/or more detailed views, please refer to references [92, 93]. Over a wide kinematic range the uncertainties are of the level of 1–2%. All subcorrections are know to the same level of precision. Due to the size of the response correction it becomes the dominant source of uncertainty on the combined jet energy scale. At low p_{T} the uncertainty is dominated by the purity of the photon+jet sample. At high-p_{T} the uncertainties are affected by decreasing statistics in the di-jet sample. Overall the limiting factors are the photon purity and the photon energy calibration.

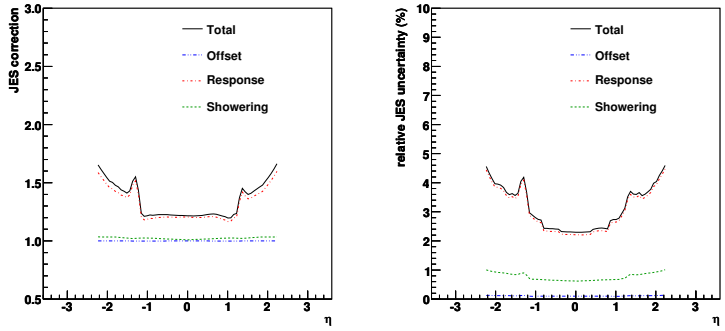
The high-precision jet energy calibration is directly visible in physics results. An example is the measurement of the inclusive jet cross section to a precision of $\approx$ 10% [103].



(a) $p_{\mathrm{T}} = 50$ GeV/c

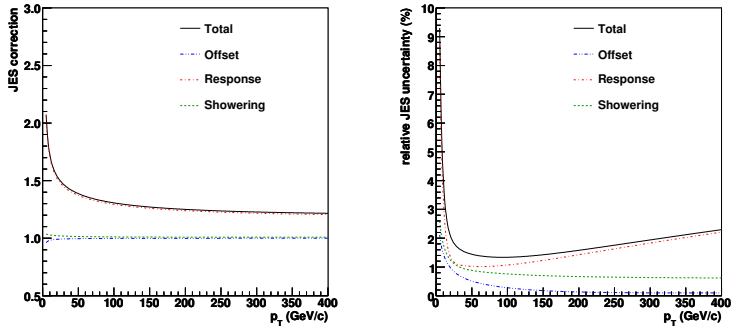


(b) $p_{\mathrm{T}} = 100$ GeV/c

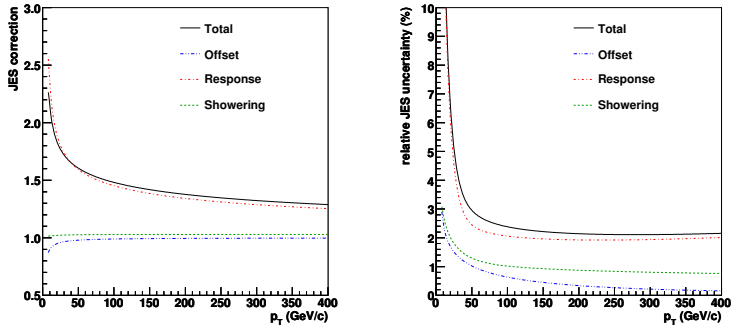


(c) $p_{\mathrm{T}} = 400$ GeV/c

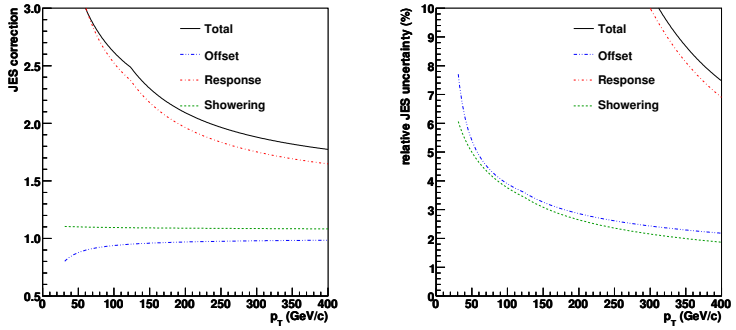
Figure 4.19: *Combined jet energy scale corrections (left-hand side) and uncertainties (right-hand side) for $\mathcal{R}_{cone} = 0.5$ cone jets as functions of jet pseudorapidity for different fixed jet p_Ts (all subcorrections are shown as multiplicative factors). The zero-suppression bias corrections have been folded into the offset correction, the response correction includes both absolute and relative response corrections, as well as the MPF bias correction.*



(a) $\eta = 0.0$



(b) $\eta = 1.0$



(c) $\eta = 2.5$

Figure 4.20: *Combined jet energy scale corrections (left-hand side) and uncertainties (right-hand side) for $\mathcal{R}_{cone} = 0.5$ cone jets as functions of jet p_T for different fixed values of jet η (all subcorrections are shown as multiplicative factors). The zero-suppression bias corrections have been folded into the offset correction, the response correction includes both absolute and relative response corrections, as well as the MPF bias correction.*

Chapter 5

Event samples and selection

This chapter describes the selection of the data (section 5.1) and Monte Carlo (section 5.2) samples used in the analysis. Special attention is paid to the discussion of calibration and MC correction factors. Selection cuts based on the hadronic $t\bar{t}$ event signature are applied to improve the signal-to-background ratio. The event selection and corresponding efficiency are introduced in section 5.3.

5.1 Dataset selection

The data sample covers almost [1] the full Run IIa data set, spanning the trigger list versions v8 through v14 and corresponding to approximately 1 fb^{-1} of recorded luminosity. The data was taken between July 2002 and February 2006.

5.1.1 Data quality requirements

Not all recorded data is suited for analysis. Problems with the performance of certain subdetectors or with the overall detector- and/or trigger configuration can be reason to exclude data from analysis. The data quality criteria remove this data. Data quality selection is applied based on run number, luminosity block number and even-by-event. Runs are the 'control room units' of data taking: periods generally ranging from two to four hours during which the detector and trigger configuration is kept unchanged. Bad runs reflect the presence of longer periods of impaired data taking. Runs are graded from the point of view of each of the subdetectors, allowing analysers to select only those with adequate data quality for all relevant subdetectors. Bad luminosity blocks are generally due to problems with the trigger or luminosity systems or to short periods

[1] Several months of data were taken in Run IIa while commissioning the detector and developing the trigger list. Trigger list version v8 contains the first trigger criteria with the full detector commissioned.

of deteriorated calorimeter performance. During event reconstruction, algorithms run to flag events showing signs of transient calorimeter issues like coherent noise [104]. Events in which any of the calorimeter quality flags were raised are rejected.

The data quality cuts remove events based on run number or luminosity block number. The corresponding integrated luminosity loss is taken into account in the luminosity calculation. Since luminosity information is not available on an event-by-event basis, the event based data quality is absorbed into the event selection (see section 5.4.1).

For this analysis the standard D0 data quality selection [105, 106] was applied. This ensures proper operation of all subdetectors, luminosity normalisability and only selects runs taken with official trigger lists.

5.1.2 b-Jet identification

The neural network b-jet tagger described in section 3.3.2 is used to identify b-jets. The NN > 0.65 operating point was chosen to suppress the fake rate to manageable levels; moreover the background modelling seems to perform better for samples selected with higher b-jet purities. See section 5.2.1 for a description of b-tagging in Monte Carlo events.

5.1.3 Jet energy scale

To calibrate all jet energies the final p17 jet energy scale including muon corrections was applied. The muon corrections correct for the jet energy 'missed' in the calorimeter and carried away by muons inside jets. High p_{T} muons, having nearly straight tracks, are often reconstructed with arbitrarily high transverse momenta. To prevent these mismeasured muons from deteriorating the jet resolution through the muon corrections, the size of the muonic jet energy corrections is 'capped' at a maximum of 60 GeV [107].

5.2 Monte Carlo samples

Monte Carlo samples for the $t\bar{t}$ signal were generated using the ALPGEN generator [108] v2.05 [2] for the hard interaction, followed by PYTHIA v6.319 [96] for fragmentation, shower development and hadronisation.

[2] The actual ALPGEN version used is 'v2.05_d0', which contains some fixes to the original v2.05 code. Further details can be found in references [109] and [110].

ALPGEN is a tree-level matrix element generator for multi-parton QCD processes, discarding virtual corrections. The connection of the hard jets from the matrix element to the softer jets originating from the (PYTHIA) parton shower is performed using Mangano's 'MLM matching' [111]. The matrix element calculation generates only hard partons, with $p_{\mathrm{T}} > 8$ GeV/c and a radial separation of $\Delta\mathcal{R} > 0.4$. After evaluation of the matrix element and shower development (but before hadronisation) the partons are clustered using a $\mathcal{R}_{\mathrm{cone}} = 0.4$ cone algorithm. Each matrix element parton is required to be matched within $\Delta\mathcal{R} < 0.7$ with a single jet, otherwise the event is rejected. Samples of $\mathrm{t\bar{t}}$ events with up to two additional light partons are generated. Event samples containing zero or one additional light parton are generated exclusively: only jets matching matrix element partons are accepted. The two-light-parton sample is generated exclusively: additional jets, not matched to hard partons, are allowed if they are softer than the matrix element partons. Subsequently, the individual parton multiplicity samples are combined (while accounting for the efficiency of the matching procedure).

For the underlying event modelling in PYTHIA the Tune A [112] parameter set was used. Tune A was obtained from CDF Run I data and still does a good job describing both the underlying event and the effects of multiple interactions in Run IIa [98].

Monte Carlo signal samples are available for several assumed values of the top quark mass, ranging from $m_{\mathrm{t}} = 165$ GeV/c^2 to $m_{\mathrm{t}} = 175$ GeV/c^2. All ALPGEN samples were generated for the all-jets, lepton+jets and di-lepton channels separately and combined according to the branching fractions shown in table 5.1.

The D0 detector simulation follows two steps. D0gstar [113, 114] is the GEANT [115] (v3.21.13) based D0 detector simulation. It contains the full detector geometry description including all material definitions and the magnetic field map. The second step is performed by D0sim [116, 117], which overlays ZB data events to simulate noise and pile-up effects, simulates detector noise and inefficiencies, and performs the digitisation of the simulated readout. The simulated signal is reconstructed using the default D0 reconstruction software also used for data.

5.2.1 b-Jet identification in Monte Carlo

In the presence of any differences in e.g. shower development between data and Monte Carlo, direct application of the b-tagging algorithm to Monte Carlo events could

Decay channel	Fraction (%)
all-jets	45.698
di-lepton (no tau)	6.452
di-tau	1.166
tau + e/μ	2.879
e/μ + jets	34.343
tau+jets	9.462

Table 5.1: *Branching fractions calculated from the individual particle branching fractions taken from the* PDG 2006 *[11]. Tau final states include only hadronically decaying tau leptons, leptonic tau decays are included in the respective leptonic channels.*

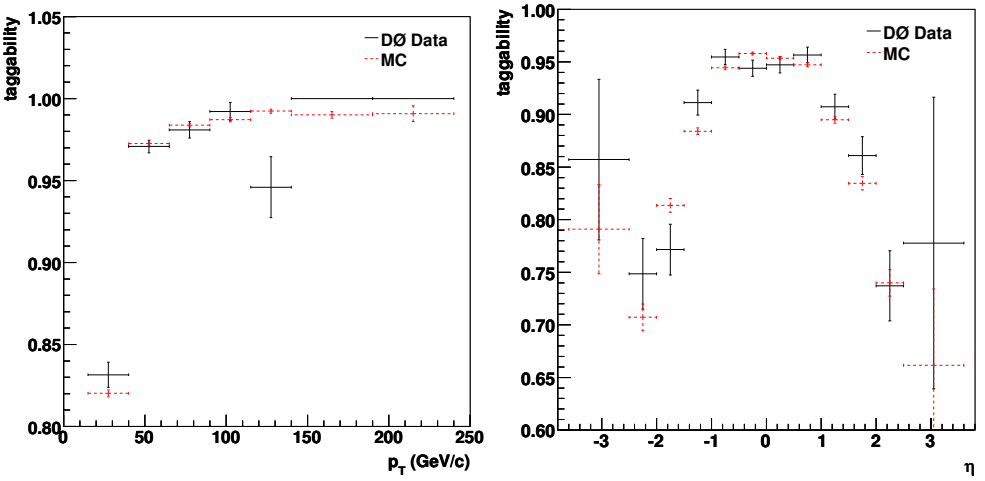
introduce biases. Instead of applying the tagger directly, taggability and b-tagging in Monte Carlo are applied following 'try-and-reject' methods.

Taggability correction

Jet taggability requires the spatial matching of a calorimeter jet with a track jet (section 3.3.2). Imperfections in the simulation of tracking and/or shower development can lead to taggability differences between data and Monte Carlo. This taggability difference was studied using samples with at least six jets within $|\eta_{\mathrm{d}}| < 2.5$, three with $p_{\mathrm{T}} > 40$ GeV/c the others with $p_{\mathrm{T}} > 15$ GeV/c. Since the flavour content of this data sample differs significantly from the (signal) Monte Carlo, b-jet selection criteria were applied to connaturalize both samples.

Figure 5.1 shows a comparison between data and Monte Carlo taggabilities based on the (tight) b-jet selection used in the analysis: $p_{\mathrm{T_b}} > 40$ GeV/c and NN > 0.65. Figure 5.2 shows the corresponding data-to-MC taggability ratio (▲ markers). Loosening the b-jet criteria, lowering the p_{T} cut from 40 to 15 GeV/c or the neural network tagger cut from 0.65 to 0.2, results in slightly lower relative taggabilities. The lowest relative taggability occurs for the loosest b-jet selection: $p_{\mathrm{T_b}} > 15$ GeV/c and NN > 0.2 (▼ markers in figure 5.2).

The overall differences are of the order of one percent up and down. This difference was treated by uniformly decreasing taggability in MC by 1.0% (i.e. randomly making one in a hundred taggable jets not taggable) and using that as a systematic 'down'



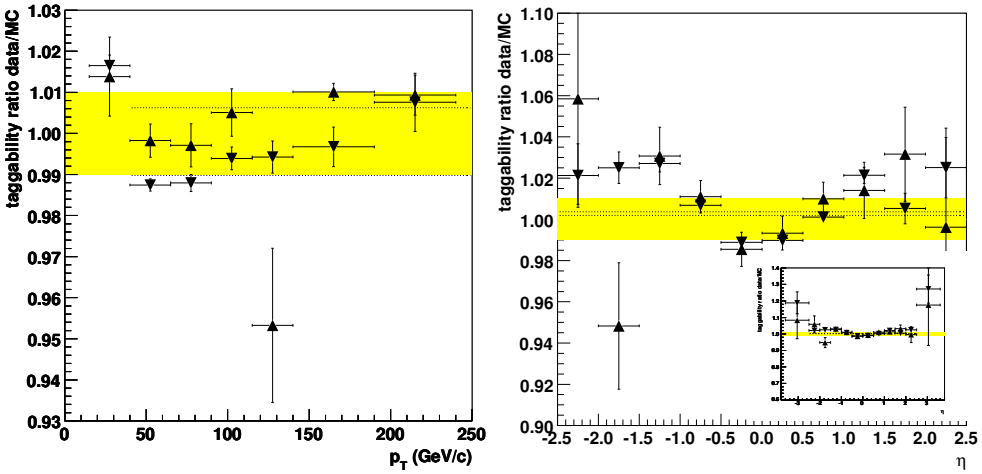
(a) vs. jet p_{T} ***(b)*** vs. jet η_{d}

Figure 5.1: *Taggability comparison between data and Monte Carlo, (a) as a function of jet transverse momentum and (b) as a function of jet pseudorapidity. Results shown represent the (tight) b-jet selection criteria of the analysis sample:* $p_{Tb} > 40$ *GeV/c and* $NN > 0.65$.

variation. The difference was symmetrised to represent the corresponding 'up' variation (see also section 9.2).

Monte Carlo b-tagging

The efficiencies for tagging b- and c-jets as well as the fake rate (jets from light quarks of taus) for the neural network tagger are described in ref. [88] and implemented as probability densities (tag rate functions or TRFs) parameterised as a function of jet transverse momentum and pseudorapidity. The functional form of the TRFs was obtained from Monte Carlo and scaled to match the data in a sample enriched in muonic jets (jets containing muons). The b-tagging in Monte Carlo was implemented using a Monte Carlo 'try-and-reject' method: for each taggable jet a random number is thrown and the TRF evaluated. If the tagging probability exceeds the random number the jet is tagged, otherwise it is left untagged.



(a) vs. jet p_{T} ***(b)*** vs. jet η_{d}

Figure 5.2: *Relative Monte Carlo-to-data taggability scale factor as a function of jet p_T (a) and as a function of jet pseudorapidity (b). ▲ Markers: tight b-jet criteria ($p_{T_b} > 40$ GeV/c, $NN > 0.65$) , ▼ markers: loose b-jet criteria ($p_{T_b} > 15$ GeV/c, $NN > 0.2$). The dashed lines represent (χ^2) fits to the ratios for the $p_T > 40$ GeV/c and $|\eta| < 2.5$ ranges. The inset in (b) shows the change in behaviour outside this η range. The shaded bands indicate the (1.00 ± 0.01) scale factor.*

5.2.2 Corrections applied to Monte Carlo

The Monte Carlo samples are corrected for several imperfections of the event generation and/or detector simulation:

- The MC samples were generated using CTEQ6L1 [118]. The central value of the Monte Carlo samples was reweighted to correspond to the more recent (NLO) CTEQ6.5 [33] parameterisation.
- The Bowler heavy quark fragmentation function [119] (describing the fractional energy a newly created hadron receives from its parent) used in PYTHIA was tuned [120] to reproduce the fractional B-hadron energy spectra [121, 122, 123] measured by the ALEPH [124], DELPHI [125] and OPAL [126] experiments. This parameter tuning is applied to the Monte Carlo by reweighting all B-hadrons. [3]

[3] This explicit reweighting only affects b-quark fragmentation. However, the tuned parameters were chosen such that the effect on c-quark fragmentation is minimal, maintaining consistency without explicit c-fragmentation reweighting.

The chosen parameterisation is consistent with all three LEP experiments but disagrees slightly with data [127] from the SLD experiment [128]. The effect of an alternative tuning to the SLD data is included as a systematic uncertainty on the measured cross section (section 9.2).

Jet energy scale and jet corrections

Apart from the above (technical) corrections the final p17 Monte Carlo jet energy scale was applied, including muon correction capping as for data. In addition Monte Carlo jet four-momenta were smeared to reproduce the relative jet-probe p_{T} imbalance in photon+jet and Z+jet events in data [129]. Additional smearing corresponding to the uncertainty on the jet p_{T} resolution, as well as the relative data-to-Monte Carlo jet energy scale, were included as systematic uncertainties on the measured cross section (see section 9.2).

Primary vertex position reweighting

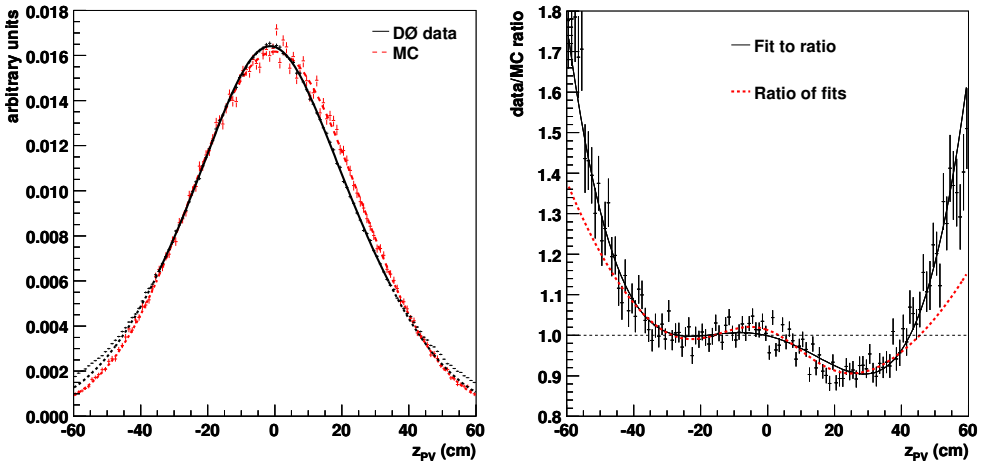
The position of the interactions along the beam (z) axis is determined by the $\beta^{\star}$ function describing the accelerator beam optics at the interaction point. Due to both machine (re)tunings and external influences (e.g. temperature changes) periodically small changes in beam optics occur. The interaction region in the simulation is kept fixed. To correct for any differences the distribution of the primary vertex z-position z_{PV} in Monte Carlo is reweighted to correspond to that in data within the selected range of $|z_{\mathrm{PV}}| < 35$ cm.

The required reweighting was estimated from events containing at least six jets with $p_{\mathrm{T}} > 15$ GeV/c within $|\eta_{\mathrm{d}}| < 2.5$. [4] To avoid a bias due to the b-tagging term in the third level trigger, only events firing the 4JT trigger were selected.

Figure 5.3a shows the z_{PV} distributions of this "6+15'' sample in data and MC. Both distributions are fitted with a sum of two Gaussians [5] (up to $|z_{\mathrm{PV}}| < 35$ cm), showing excellent fit agreement. The distribution in Monte Carlo is slightly shifted and

[4] Otherwise the same selection cuts were applied as described in section 5.3. Since the primary vertex distribution does not depend on the jet selection (as will be shown by the vertex acceptance fits in section 5.4.2), the p_{T} threshold was lowered to increase statistics.

[5] A better approximation of the shape of the interaction region should be the use of the real luminosity overlap integral $\frac{d\mathcal{L}(z)}{dz} \sim \frac{\exp^{-(z-z_0)^2/(2\sigma_z^2)}}{\sigma_x(z)\sigma_y(z)}$ where z_0 is the minimum of the β function along the z direction and the transverse beam widths depend on the beam emittance and on z through the transverse components of the β function. However, the details of this distribution are washed out by the combination of many random beam spot fluctuations over the course of Run IIa, reducing the shape to an approximate double Gaussian.



(a) *(b)*

Figure 5.3: *Primary vertex z-position distributions (a) for the '6+15' samples in data and MC, both scaled to unit area. Also shown are the fits to these distributions (solid and dashed lines respectively) and the fit prediction outside the $|z_{PV}| < 35$ cm fit range for data (dotted line). Figure 5.3b shows the ratio of data/MC distributions (markers), a fourth order polynomial fit to the ratio (solid line) and the ratio of the two fit results in the top figure (dotted line). The latter is used for the z_{PV} reweighting applied to the Monte Carlo.*

marginally narrower than the one in data. Figure 5.3b shows the ratio of data/MC for the z_{PV} distributions. The relative shift of the two distributions results in a 'double bump' structure. It was chosen to use the ratio of the data- and MC fit results to the z_{PV} distributions for the MC reweighting (dashed line in figure 5.3). A cross check using a fourth order polynomial fit to the ratio (solid line in figure 5.3b) gives compatible results.

It should be noted that the fit extrapolation (dashed line) outside the fit range in figure 5.3 lies underneath the data points. This discrepancy is caused by the slightly different track selection criteria outside the SMT fiducial region. For more details see section 5.4.2.

Luminosity reweighting

To simulate the effects of noise and pile-up, Monte Carlo samples are overlaid with zero-bias data. The ZB data samples contain predominantly low luminosity events. To achieve a representative luminosity distribution, the instantaneous luminosity distribution in Monte Carlo was reweighted to match that of the data sample. Figure 5.4 shows the luminosity profiles (recorded luminosity as a function of instantaneous luminosity) for the individual 4JT triggers used. (For details on the triggers please refer to chapter 6.) By taking a luminosity-weighted sum of the individual trigger profiles an inclusive luminosity profile can be formed. The inclusive luminosity profiles for the 3JT and 4JT trigger sets corresponding to the whole data taking period used are shown in figure 5.4b together with the inclusive Run IIa reference profile from the (always unprescaled) JT125 trigger. The small difference in shape between the inclusive 3JT and 4JT profiles is due to the fact that the 4JT triggers were relatively more exposed to low luminosity. Since the 4JT triggers represent most of the integrated luminosity for this analysis (see section 6.2) and are also more efficient than the 3JT triggers (see section 6.3) the inclusive 4JT luminosity profile is used to reweight the Monte Carlo samples.

5.3 Event selection

The typical event signature of at least six jets, among which two high p_{T} b-jets, in the absence of any isolated leptons is used to select predominantly $\mathrm{t\bar{t}}$-like events. Additional cuts are applied to reject ill-reconstructed events and to ensure orthogonality with other $\mathrm{t\bar{t}}$ cross section analyses, allowing for a combination of the results.

To ensure well-reconstructed events, events without reconstructed primary vertices are rejected. The first primary vertex (the one most likely to originate from a hard interaction) is required to have at least three associated tracks. The high track multiplicity in multi-jet events ($\approx$ 35–40 on average for the selected analysis sample) makes this cut almost 100% efficient.

To ensure orthogonality with respect to the (semi)leptonic $\mathrm{t\bar{t}}$ cross section analyses, events containing isolated leptons are rejected. [6]

The number of inelastic interactions per bunch crossing for the full Run IIa data set follows a Poisson distribution with mean of ≈ 2, slightly depending on the trigger. For

[6] Unfortunately, the information required to veto events selected by the tau+jets analysis was not available in our working files; no such veto was applied.

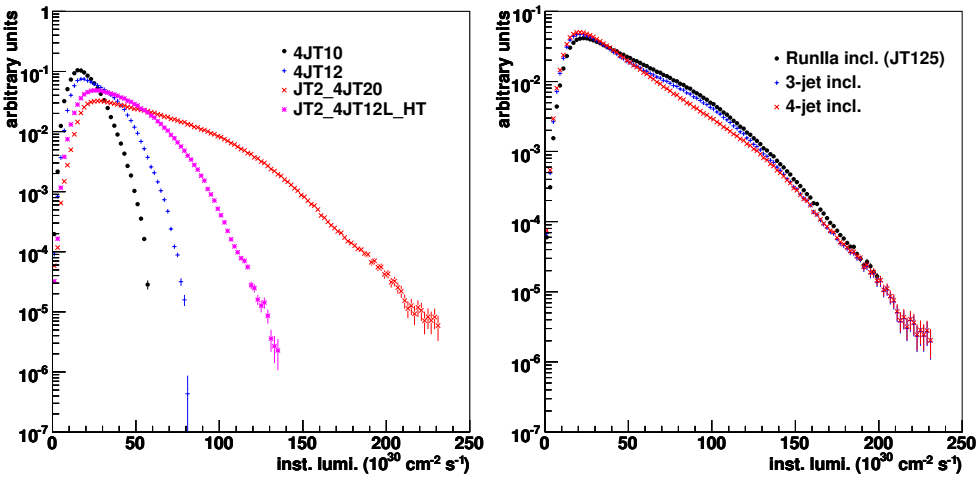


Figure 5.4: *Luminosity profiles (a) for the individual 4JT triggers and (b) for the combined 3JT and 4JT triggers. Triggers used in later run ranges are increasingly exposed to higher luminosities. All profiles are scaled to unit area (the legend in (a) is sorted in order of increasing run range).*

samples predominantly selecting high instantaneous luminosity events, like $t\bar{t}$ analyses, the average can go up as far as 5. To suppress the effects of multiple interactions, only taggable jets are considered. To a good approximation, taggable jets point back to the selected primary vertex. Ignoring untaggable jets separates jets from additional interactions from the hard scatter jets.

The first step of the real multi-jet event selection is the trigger requirement, selecting events that fired three- or four-jet triggers. The exact trigger requirement depends on the run range, see also section 6.1. To avoid biases due to vertex requirements present in (some of) the triggers, only events are selected in which the primary vertex lies within the fiducial region $|z_{\mathrm{PV}}| < 35$ cm.

Based on the hadronic $t\bar{t}$ decay signature, only events are allowed with at least six jets. To stay away from the low-E_{T} bias region and in the efficient regime of the jet identification, only jets with $p_{\mathrm{T}} > 15$ GeV/c are considered. Jets outside $|\eta_{\mathrm{d}}| < 2.5$ are also ignored. The large mass of the $t\bar{t}$ system compared to $\sqrt{s}$ leads to a fairly central distribution of the $t\bar{t}$ signal: a negligible fraction of signal exists with jet pseudorapidities beyond $|\eta| \approx 2.5$. The distribution of the QCD background is only

	Inclusive $t\bar{t}$ (%)		Hadronic $t\bar{t}$ (%)		Background (%)	
	inc.	cum.	inc.	cum.	inc.	cum.
Sample total	100.0	100.0	100.0	100.0		—
Vertex	83.9	83.9	83.8	83.8		—
Trigger	54.1	45.4	78.4	65.7		—
≥ 6 Jets with $p_{\mathrm{T}} > 15$ and $\lvert\eta_{\mathrm{d}}\rvert < 2.5$	39.8	18.1	51.5	33.8		—
Lepton veto	96.4	17.4	100.0	33.8		—
All jets taggable	65.1	11.3	68.2	23.1		—
≥ 2 Jets with NN > 0.65, $p_{\mathrm{T}} > 15$	35.8	4.1	36.0	8.3	100.0	100.0
≥ 1 Jet with NN > 0.65, $p_{\mathrm{T}} > 40$	95.4	3.9	95.4	7.9	75.2	75.2
≥ 2 Jets with NN > 0.65, $p_{\mathrm{T}} > 40$	65.1	2.5	65.1	5.1	28.9	21.7
One additional jet with $p_{\mathrm{T}} > 40$	97.4	2.5	97.6	5.0	70.5	15.3
Total		2.5		5.0		—

Table 5.2: *Incremental and cumulative selection efficiencies based on the $m_t = 170$ GeV/c^2* ALPGEN *Monte Carlo event sample.*

marginally wider, but the absolute scale is many times higher. The cut on η_{d} also confines jets to the well-understood 'central' region of the detector. Moreover, beyond $|\eta_{\mathrm{d}}| \approx 2.5$ the effects of pile-up increase, and the uncertainties on both the jet energy scale and the b-tagging increase.

One of the characteristics of $t\bar{t}$ events is the presence of two high momentum b-jets. This is implemented as a requirement of at least two jets b-tagged by the neural network tagger with NN > 0.65 and $p_{\mathrm{T}} > 40$ GeV/c. To avoid biases due to the trigger p_{T} thresholds the first three jets are required to satisfy $p_{\mathrm{T}} > 40$ GeV/c. This keeps all jets on the trigger turn-on plateau (after JES corrections).

After application of the above selection cuts the signal-to-background ratio is approximately one tenth.

5.3.1 Selection efficiency

The efficiencies for the above selection cuts are summarised in table 5.2 assuming a top quark mass of $m_{\mathrm{t}} = 170$ GeV/c^2.

The vertex quality criteria area easily satisfied in multi-jet events and highly efficient ($\approx 97\%$, see section 5.4.2). The vertex position cut is governed by beam dynamics and independent of the hard scatter physics, and as such independent of the decay channel. As expected the multi-jet trigger efficiency, as well as that of the 'at least six

jets' requirement, is slightly higher for the hadronic channel compared to inclusive $t\bar{t}$ decays. The high efficiency for the isolated lepton veto is presumably due to the busy multi-jet environment that leads to higher calorimeter energy cluster multiplicity. Somewhat surprisingly the taggability requirement cuts away more non-hadronic than hadronic signal events. This is probably due to the lower intrinsic jet multiplicity of the (semi)leptonic events. Since only events with six or more jets are selected, it is likely that the non-hadronic sample contains a larger fraction of events with additional pile-up than the hadronic sample, which already contains six jets in its simplest form. No hadronic versus non-hadronic efficiency differences are present in the jet p_{T} and tagging requirements. Those cuts are based on the generic $t\bar{t}$ decay signature of two hard b-jets (and requiring one additional hard jet); any $t\bar{t}$ event with six jets is likely to pass these cuts, independent of the decay mode.

The 5.0% hadronic efficiency combined with the 2.5% inclusive efficiency (and assuming the branching fractions shown in table 5.1) implies a 7.9% non-hadronic signal contamination. It is interesting to note that using the PYTHIA event generator, the inclusive and hadronic signal efficiencies are 2.1% and 4.3% respectively, resulting in an approximate 5.8% non-hadronic signal contamination. The main efficiency differences lie in the trigger and six-jet requirements (both higher for the ALPGEN sample), reflecting the effect of the harder p_{T} spectrum of the $t\bar{t}$ system in ALPGEN, bringing more jets inside the detector fiducial volume.

5.4 Global efficiencies

Apart from event selection efficiency two additional, global efficiency factors have to be taken into account:

- the event based data quality selection correcting for events lost due to transient calorimeter performance issues, and
- the primary vertex acceptance correction for the cut on z_{PV}.

5.4.1 Data quality inefficiency

The fractional data loss due to events removed based on the calorimeter event quality flags has been estimated in zero-bias data. The losses are stable over time, ranging from 2.2% to 3.2%. The luminosity-weighted average loss was estimated to be 2.7% [130].

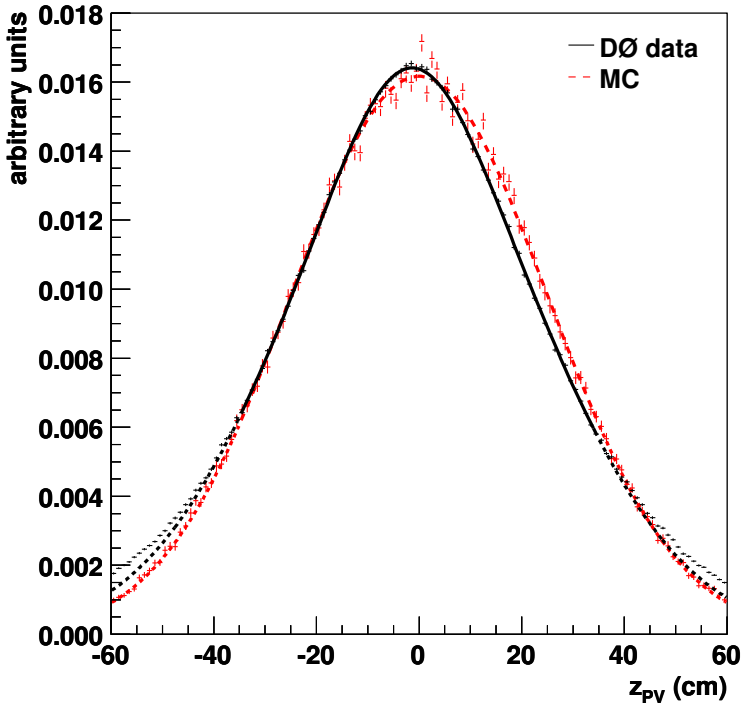


Figure 5.5: *The primary vertex z-position distribution (markers) in data (scaled to unit area) fitted with the sum of two Gaussians (solid line). The dotted line shows the fit extrapolation outside the* ± 35 *cm fit range.*

5.4.2 Primary vertex acceptance

The primary vertex acceptance corrects for the fraction of events missed due to the vertex position cut. Since the $|z_{\mathrm{PV}}| < 35$ cm cut is already contained in the selection efficiency, the acceptance is implemented as a relative efficiency correction.

The primary vertex reconstruction acceptance is estimated from the "6+15'' data sample also used for the vertex reweighting in Monte Carlo (section 5.2.2). The z_{PV} distribution for this sample is shown in figure 5.5, fitted within $|z_{\mathrm{PV}}| < 35$ cm with the sum of two Gaussians. The range is limited to well within the SMT fiducial range ($|z_{\mathrm{PV}}| \lesssim 38$ cm). For vertices outside that range slightly different track requirements are used (i.e. the requirement of a minimum number of SMT hits on a track is dropped) leading to a transition in vertex reconstruction efficiency around the fiducial boundary. The agreement between fit and data points in the fit range is excellent. Outside the fit region the data points lie above the fit extrapolation due to the looser track

requirements outside the SMT fiducial region that lead to an increase in both efficiency and fake rate. Integrating over the fit result inside and outside $|z_{\mathrm{PV}}| = 35$ cm shows a primary vertex acceptance of 81.1% with a fit uncertainty of 1.3% (absolute). A cross check, tightening the jet selection to require at least three jets with $p_{\mathrm{T}} > 40$ GeV/c shows this number is stable under variations of the jet p_{T} cuts [109]. The small differences observed between different event selections are added to the fit uncertainty, leading to an acceptance uncertainty of 2.0%.

Combined with the 83.9% efficiency of the vertex requirement in the event selection (table 5.2) this results in a 96.7% residual efficiency correction.

Chapter 6

Trigger description and integrated luminosity

In this chapter the triggers used in the data selection are presented. Trigger 'OR-ing' and the corresponding integrated luminosity are discussed. Special attention is paid to the multi-jet combinatorics involved in determining the trigger efficiencies, as well as to the non-jet trigger terms and their efficiencies.

6.1 Trigger requirements

Part of the event selection is the trigger requirement, selecting only events in which one or more relevant triggers fired. Two different sets of triggers are used: one requiring at least three jets (3JT) and one requiring at least four jets (4JT). Both trigger types contain additional terms to select predominantly high-mass events. These triggers were chosen to provide basic thresholds on the individual jet transverse momenta (p_{T}) as well as on the jet p_{T} sum (H_{T}), while maintaining high signal efficiency. Looser multi-jet triggers exist but are prescaled at high luminosities, making them less suited for small-signal analyses. Following developments in instantaneous luminosity, algorithm development and physics interest, the exact trigger definitions may change over time. Different run ranges have been defined grouping together periods with similar trigger conditions. Run ranges are labelled based on the version of the trigger lists used: v8, v9, v10, v11, v12, v13a/b and v14a/b. Combining ('OR-ing') two or more triggers can improve the trigger efficiency for certain classes of events. In the case of 'OR-ed' triggers, if at least one of the triggers fires, the event is accepted. Whether two triggers can be combined depends on their prescales. If at least one of the triggers is not prescaled, the triggers can be OR-ed. The integrated luminosity corresponding to the OR equals the luminosity to which the unprescaled trigger was

exposed. If both triggers were prescaled OR-ing is in general not possible. [1] One exception is the case in which both triggers shared the same Level 1 enable bit (which is also used for prescaling). In that case both triggers are exposed to exactly the same luminosity blocks and integrated luminosity. For each run range one 3JT and one 4JT trigger were selected. Where possible these triggers were OR-ed: either the 3JT or the 4JT trigger was required to fire. Only during the v8 run range, normalisation requires the exclusive use of the 4JT trigger. [2] Benefits of the trigger OR-ing include an increased integrated luminosity, increased trigger efficiency and less sensitivity to the uncertainties on the Level 3 b-tagging term (see section 6.3.3).

6.1.1 Trigger definitions

The first and second trigger levels use special approximations to physics objects like EM clusters and tracks. The third level trigger is based on partial event reconstruction; Level 3 trigger objects are similar to those after offline (full) reconstruction of the raw data. A brief description of the Level 3 trigger requirements is given below; the complete trigger and run range definitions can be found in ref. [109].

3JT

The 3JT trigger used in run ranges v8, v9 and v10 requires three L3 jets with $E_{\mathrm{T}} > 15$ GeV/c. From v10 on the 3JT L3 jet requirement was: three jets with $E_{\mathrm{T}} > 15$ GeV/c and two with $E_{\mathrm{T}} > 25$ GeV/c. From v10 on the 3JT triggers also required the L3 primary vertex to be within $|z_{\mathrm{PV}}| < 35$ cm. As of v11 the 3JT triggers contained an additional L3 term requiring at least one of the leading six jets to have a more-than-five-percent probability of being a b-jet.

4JT

In v8 the 4JT trigger required four jets with $E_{\mathrm{T}} > 10$ GeV/c, in v9 this requirement changed to: four jets with $E_{\mathrm{T}} > 10$ GeV/c and two jets with $E_{\mathrm{T}} > 20$ GeV/c. The 4JT L3 jet requirement in v12-v13 was: four jets with $E_{\mathrm{T}} > 12$ GeV/c, three with $E_{\mathrm{T}} > 15$ GeV/c, and two with $E_{\mathrm{T}} > 25$ GeV/c. In addition, in v13 the L3 H_{T} of all

[1] Two prescaled triggers can be OR-ed if the correlation between the prescales is known. For example: in two extreme cases two triggers with prescale one-half can each be exposed to completely orthogonal sets of luminosity blocks, or to the exact same blocks. This correlation information is not provided by the D0 trigger system.

[2] To achieve the OR-ing in run ranges v9-v14 five runs were removed from the data set. The resulting luminosity loss was less than 0.4 pb^{-1}.

Run range	$\mathcal{L}$ (pb^{-1})		
	3JT	4JT	Trigger used
v8	10.84	<u>24.54</u>	4JT
v9	23.93	<u>24.07</u>	OR
v10	6.63	<u>10.22</u>	OR
v11	<u>64.68</u>	<u>64.68</u>	OR
v12	199.40	<u>210.23</u>	OR
v13a	<u>51.72</u>	49.34	OR
v13b	<u>322.12</u>	257.39	OR
v14a	<u>189.83</u>	<u>189.83</u>	OR
v14b	<u>143.55</u>	<u>143.55</u>	OR
Total	1012.70	973.84	1040.96

Table 6.1: *The trigger choices and corresponding luminosities for the different run ranges [131]. The luminosity shown is the integrated recorded luminosity after data quality selection. 'OR' represents the OR of the 3JT and 4JT triggers. Underlined numbers represent the luminosity corresponding to the OR. Differences in exposed luminosity between the 3JT and 4JT trigger suites originates from differences in trigger prescaling.*

jets with $E_{\mathrm{T}} > 9$ GeV/c should be large than 120 GeV/c^2. During v14 the 4JT trigger required four L3 jets with $E_{\mathrm{T}} > 20$ GeV/c, and two with $E_{\mathrm{T}} > 25$ GeV/c.

6.2 Integrated luminosity

The integrated recorded luminosity, corrected for the data quality selection,[3] for all run ranges is shown in table 6.1. For the v8 run range only the 4JT trigger was used and the corresponding luminosity is shown. For all other run ranges both triggers were either prescaled identically, and thus exposed to the same integrated luminosity, or one of the triggers was unprescaled. In the latter case the luminosity is shown for the unprescaled trigger. The overall luminosity increase due to the trigger OR-ing compared to 4JT alone is $\approx 6\%$.

[3] All data quality cuts discussed in section 5.1.1 are accounted for in these luminosity numbers. The per-event data quality selection used to reject transient calorimeter issues is treated as a selection inefficiency and was discussed in section 5.3.

6.3 Trigger efficiencies

The trigger efficiency determination is based on single-object trigger turn-on curves.

These single-object turn-ons are determined from data by selecting orthogonal data samples based on independent triggers and counting the rate at which the trigger under study fires.

For each trigger level the single-object curves are combined into a single turn-on curve. The different trigger level turn-ons are multiplied to obtain the overall trigger efficiency. Care has to be taken to remove the efficiency overlap of the OR-ed triggers:

$$\epsilon(a \text{ OR } b) = \epsilon(a) + \epsilon(b) - \epsilon(a \text{ AND } b).$$

Effects of trigger changes over the different run ranges are taken into account by using the luminosity-weighted average efficiency:

$$\epsilon = \frac{\Sigma_i \mathcal{L}_i \epsilon_i}{\Sigma_i \mathcal{L}_i}$$

where the sum includes all run ranges i, ϵ_i is the trigger efficiency for run range i and the $\mathcal{L}_i$ are integrated luminosity values from table 6.1.

The overall trigger efficiencies are folded into the Monte Carlo samples as event weights.

6.3.1 Jet triggers: combinatorics

In combining the single object trigger turn-ons some care has to be taken to correctly handle the correlations between different jet multiplicity terms with different thresholds. For example a jet satisfying a 25 GeV/c threshold will also satisfy all terms with lower thresholds. Using a simplified notation $\textsc{jet}(m, x)$ for a generic multi-jet trigger term requiring m jets with $p_{\mathrm{T}} > x$ GeV/c any multi-jet trigger with M terms can be written as

$$\prod_{m=1}^{M} \textsc{jet}(m, p_{\mathrm{T}}(m)), \text{ with } p_{\mathrm{T}}(m+1) \geq p_{\mathrm{T}}(m).$$

The trigger probability for a multi-jet event can be written as 'one minus all possible ways not to fire the trigger':

$$P(\text{event}) = 1 - \sum_{j=0}^{M-1} P_j$$

where P_j represents the probability that j jets in the event satisfy the trigger. The trigger probability P_j for an event with N offline jets to fire a term requiring j jets can be written as the product of j jets satisfying the trigger and $N - j$ jets failing, properly summed over all permutations of j out of N jets:

$$P_j = \textsc{jet}(j, p_{\mathrm{T}}(j)) = \sum_{\text{permutations}} \left(\prod_{k=1}^{j} P(k; p_{\mathrm{T}}(j)) \prod_{l=j+1}^{N} 1 - P(l; p_{\mathrm{T}}(j)) \right),$$

with $P(k; p_{\mathrm{T}}(j))$ the probability that jet k satisfies the threshold $p_{\mathrm{T}}(j)$. P_0, the term requiring all jets in the event to fail the trigger, can be written as

$$\begin{aligned} P_0 &= 1 - P(\text{at least one jet satisfies}) \\ &= \prod_{i=1}^{N} \left(1 - P(i; p_{\mathrm{T}}(1))\right). \end{aligned}$$

Here $P(k; p_{\mathrm{T}}(1))$ represents the probability for jet k to satisfy the trigger term with the highest threshold. If any one of the jets would satisfy this threshold, it would also satisfy all other terms and thus satisfy the trigger. A general term P_i $(i > 0)$ contains i terms:

$$P_i = \sum_{j=1}^{i} P_{i,j}$$

with

$$\begin{aligned} P_{i,j} &= P_i^{i-j+1,p_{\mathrm{T}}(1);i-j+2,p_{\mathrm{T}}(2);\ldots;i-j+j,p_{\mathrm{T}}(j)} \\ &= \frac{1}{(i-j+1)!} \sum_{j\text{-jet permutations } j_1,\ldots j_j} P(j_1;p_{\mathrm{T}}(1)) \\ &\quad \times (P(j_2;p_{\mathrm{T}}(2)) - P(j_1;p_{\mathrm{T}}(1))) \\ &\quad \times \ldots \\ &\quad \times (P(j_j;p_{\mathrm{T}}(j)) - P(j_{j-1};p_{\mathrm{T}}(j-1))) \\ &\quad \times \prod_{k=1,k\neq j_1,\ldots,j_j}^{N} (1 - P(k;p_{\mathrm{T}}(j)))\,. \end{aligned}$$

Here $P(i;p_{\mathrm{T}}(j))$ is again the probability for jet i to pass threshold $p_{\mathrm{T}}(j)$. The sum runs over all permutations of j jets out of the N jets in the event. The terms containing the $j_1 \ldots j_j$ outside the product represent the probability of the j jets satisfying the trigger. To prevent double-counting between the different thresholds each of these terms consists of the probability difference of the jet satisfying thresholds i and $i+1$. The product represents the remaining $N-j$ jets failing the trigger. For a more detailed discussion see ref. [132].

6.3.2 Non-jet triggers: the Level 3 vertex term

During the v12-v14 run ranges the 3JT triggers contained a Level 3 term requiring the primary vertex to be within $|z_{\mathrm{PV}}| < 35$ cm. Figure 6.1 shows the trigger efficiencies for the Level 3 vertex terms in the 3JT triggers for these run ranges. Due to a problem in the Level 3 bookkeeping it was not possible to determine the efficiency for the vertex term directly from data for v13a. Since the v13a and v13b run ranges have very similar luminosity distributions, the v13b curve is used for v13a as well. The change in shape between v12/v13 and v14 is caused by a Level 3 algorithm change. The efficiency 'shoulders' outside $|z_{\mathrm{PV}}|$35 cm in figure 6.1 imply misreconstructed L3 vertices (assuming the offline reconstructed primary vertex is the correct one). To remove this effect an offline cut is applied to accept only events in which the offline reconstructed primary vertex lies within ±35 cm. The acceptance effect of this cut is treated as a global efficiency correction, see section 5.4.2.

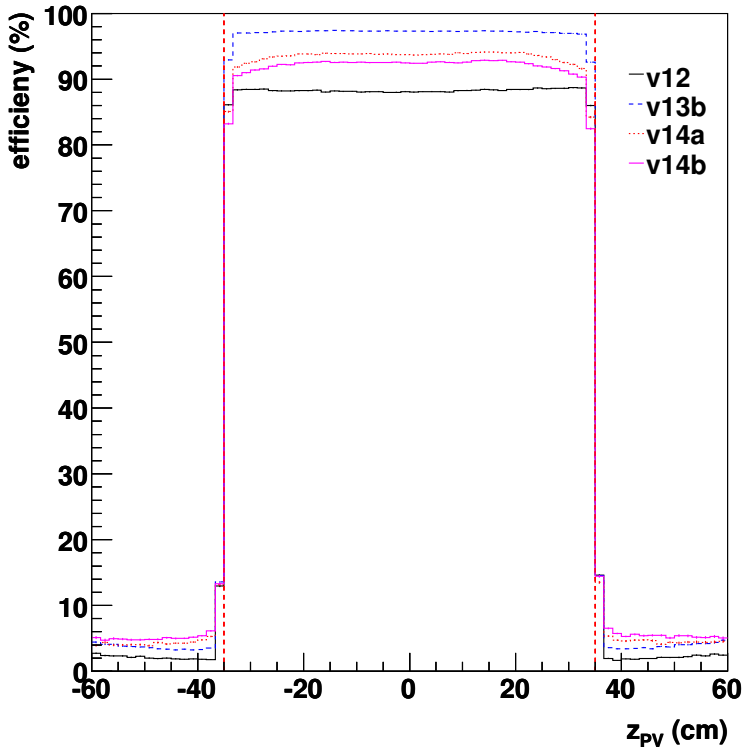


Figure 6.1: *Trigger efficiencies for the Level 3 vertex terms in the different 3JT triggers. Due to a problem in the Level 3 bookkeeping the exact turn-on for the v13a run range could not be determined. The v13b turn-on is used for v13a as well. The dashed vertical lines show the offline cut at* $|z_{PV}| < 35$ *cm.*

6.3.3 Non-jet triggers: the Level 3 b-tagging term

The 3JT triggers used for the v13 and v14 run ranges contain an impact parameter based b-tagging term at Level 3. The impact parameters for all jets with $E_{\mathrm{T}} > 10$ GeV are combined and events with a less-than-five-percent probability of containing at least one b-jet are rejected.

It is important to note that the efficiency of this L3 term is correlated with the offline selection. Offline at least two jets are required to be b-tagged by the neural network b-tagger with their neural network discriminants exceeding 0.65 (NN > 0.65). The b-tagging term in the trigger increases the signal fraction in the triggered data, meaning the efficiency has to be estimated taking into account the offline b-tagging

Run range	TRF(NN[2, 0.65])	P(IP)	P(NN[2, 0.65]|IP)	$\epsilon_{IP@L3}$
v13a	0.312	0.229128	0.121654	0.0893
v13b		0.269809	0.115691	0.1000
v14a		0.254121	0.132740	0.1081
v14b			0.136227	0.1110

Table 6.2: *The input variables to eq.* (6.1) *and the estimated efficiencies for the L3 b-tagging trigger terms.*

requirements. The efficiency of the L3 impact parameter term was estimated using Bayes's theorem:

$$\epsilon_{IP@L3} = P(\text{IP}|\text{NN}[2, 0.65]) = P(\text{IP}) * P(\text{NN}[2, 0.65]|\text{IP})/\text{TRF}(\text{NN}[2, 0.65]) \quad (6.1)$$

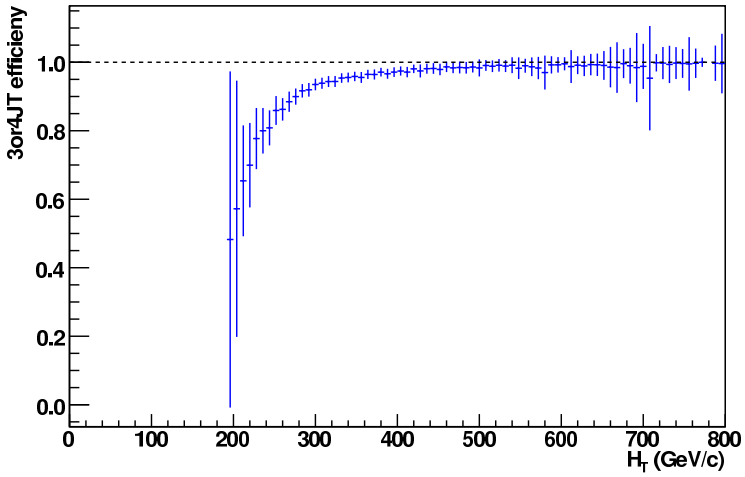
where P(IP) is the 'plain' efficiency of the IP term, 'NN[2, 0.65]' means 'at least two b-tagged jets with NN $>$ 0.65' and TRF(NN[2, 0.65]) is the efficiency of the NN b-tagger to tag at least two b-jets. The tagging efficiency for two or more tags, TRF(NN[2, 0.65]), was estimated by applying the TRFs to the signal MC before selection. The conditional efficiency of the tagging given that the IP term fired, P(NN[2, 0.65]|IP), was estimated in an orthogonal data sample containing only events that fired the 4JT trigger.

Due to lack of statistics to properly study all effects and correlations, these efficiencies were treated as constants (i.e. independent of η, p_{T}). This treatment introduces a significant uncertainty on the efficiency estimates. The trigger OR-ing, due to the large overlap between the three- and four-jet triggers, helps suppress the propagation of this uncertainty into the final result. Approximately 60% of events firing the 3JT trigger also fire the 4JT trigger, and can be treated as 4JT-triggered events, thus avoiding any uncertainties related to the 3JT triggers.

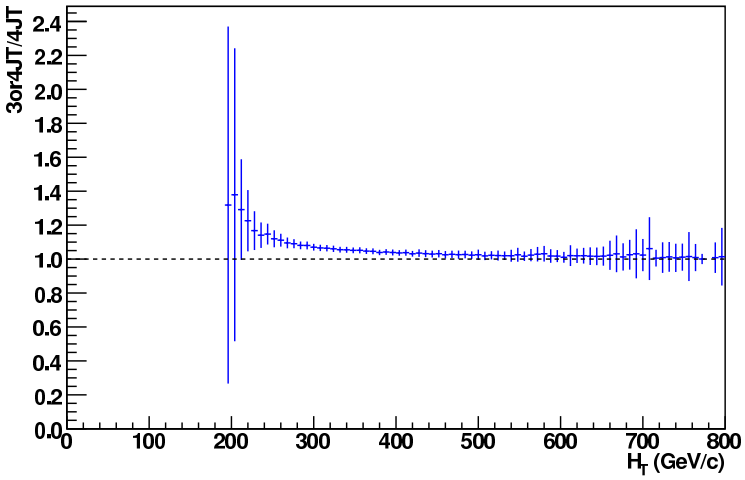
Table 6.2 summarises the input variables for eq. (6.1) and the resulting efficiencies for the L3 IP term. The L3 b-tagging terms are the main source of inefficiency of the 3JT triggers.

6.3.4 Combined trigger efficiencies

The overall, luminosity averaged trigger efficiency for the OR of the 3JT and 4JT triggers is shown in figure 6.2a. The OR of the two triggers accepts $\approx 75\%$ of the hadronic $t\bar{t}$ signal events. Figure 6.2b shows the relative increase in trigger efficiency of the OR-ed triggers with respect to 4JT alone. The increase is mainly visible at low H_T and originates from the L2 H_T terms in the 4JT triggers. The overall efficiency increase is $\approx 6\%$.



(a)



(b)

Figure 6.2: *(a) Luminosity-weighted average trigger efficiency vs. H_T and (b) the relative efficiency increase due to the trigger OR-ing as compared to 4JT alone. This increase is a result of the lower efficiency of the L2 H_T terms in the 4JT triggers and corresponds to $\approx 6\%$. The 3JT trigger efficiency is dominated by the IP@L3 terms and saturates well below 50%.*

Chapter 7

Background modelling

The complexity of multi-jet QCD events leads to challenges for event simulation. Matrix element calculations are able to accurately describe the physics behind the hard scattering processes. Hard multi-parton interactions, however, become increasingly hard to calculate for higher parton multiplicities, due to the rapidly increasing number of contributing Feynman diagrams. Parton shower methods describe the soft particle emissions from the hard scatter products but do not model hard, large-angle emissions well. 'Hybrid' generator combinations like ALPGEN followed by PYTHIA (see also section 5.2) generate the full (tree-level) final states including additional jets using matrix elements. Parton showers are then allowed to add soft jets, relying on a matching procedure to achieve a physically correct connection between the (hard) matrix element and (soft) parton shower phase space regions.

To circumvent the intricacies of QCD multi-jet event simulation, instead of relying on Monte Carlo, a data-driven background model is used. Starting from an event sample similar to the selected analysis sample, but even more strongly dominated by QCD multi-jet events, a dedicated background sample is generated.

The basic idea is to insert additional jets into four-jet and five-jet events, generating six-jet background events.

This is achieved by matching four-jet and five-jet events to events containing six or more jets [1] with similar phase space characteristics and donating all 'extra' jets to the event with lower jet multiplicity. This way artificial events with high jet multiplicity can be generated which are guaranteed not to contain any signal contributions.

In studying this background modelling procedure, the following points are important to consider:

- The starting sample used in the background generation procedure should contain a negligible amount of signal contamination. As the $t\bar{t}$ signal contains at least

[1] In this chapter, unless stated otherwise, 'six-jet' implies 'at least six jets'.

six jets, lowering the jet multiplicity should decrease the signal fraction. The cross section for QCD multi-jet production decreases strongly with increasing jet multiplicity (direct leptonic $t\bar{t}$ decays are still suppressed by the isolated lepton veto). Since the QCD background scales with approximately $\alpha_s \approx 0.1$, the QCD rate increases by roughly a factor ten. Even in the selected analysis sample the signal fraction is less than 10% (see chapter 9). The background generation procedure uses events with six or more jets with $p_T > 15\text{GeV}/c$. From the event selection efficiency in table 5.2 can be seen that 45% of those hadronic events are lost due to acceptance effects. Even if the full acceptance is recovered for four-jet events, the signal fraction cannot exceed 2%.

- The modelling procedure should not introduce any biases. This is most important when considering a (top quark) mass analysis. [2] It implies special care should be taken to verify that mass-related variables like H_T are modelled correctly.
- Any variable used in the multivariate analysis to distinguish between signal and background should be modelled correctly.

7.1 Background generation procedure

The background generation procedure is outlined in figure 7.1. The input consists of two event samples of different jet multiplicities:

The donor sample is a loosened version (see below) of the selected six-jet analysis sample of data candidate events. Jets from events in this sample will be donated to events with lower jet multiplicity in the acceptor sample.

The acceptor sample An additional data sample, selected exactly as the donor sample, but containing only five-jet (or four-jet) events. This sample contains a negligible amount of $t\bar{t}$ events and is the basis for the background sample. Two different acceptor samples were used: one containing four-jet- and one containing five-jet events.

The donor and acceptor samples use selection criteria very close to those used for the analysis sample (section 5.3). However, the two-step p_T threshold used in the analysis selection (15 and 40 GeV/c) strongly shapes the phase space occupied by the

[2] The same data sample, event selection an background model were also used to measure the top quark mass [133].

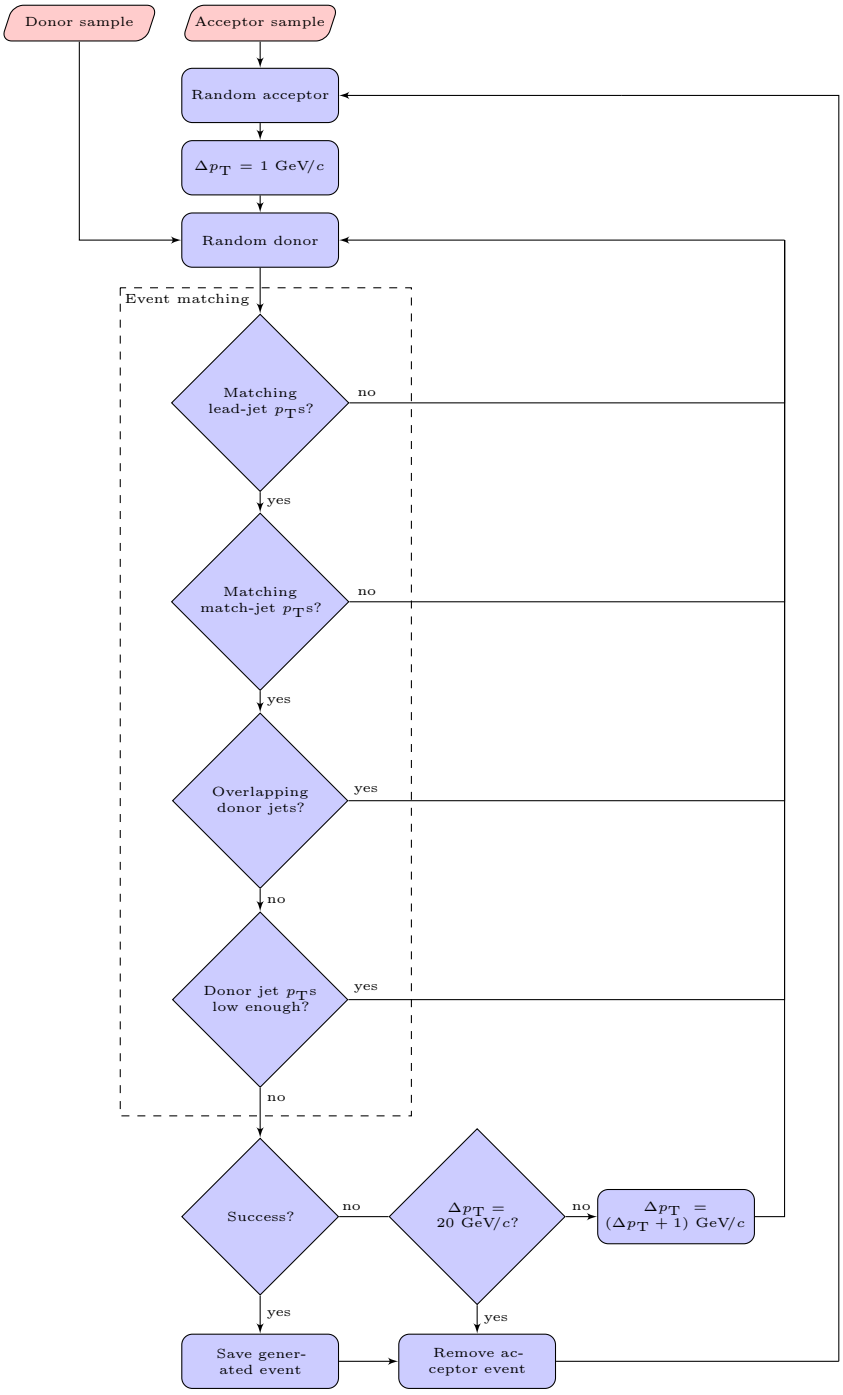


Figure 7.1: *The background generation procedure.*

selected sample. This manifests itself in artifacts around 40 GeV/c in the p_{T} spectra of the fourth and softer jets. To facilitate the phase space matching in the background generation all p_{T} cuts were lowered to 15 GeV/c. At the end of the procedure the background samples are tightened by increasing the p_{T} thresholds to require at least three jets with $p_{\mathrm{T}} > 40$ GeV/c.

In the matching procedure events from the acceptor sample are matched to randomly selected donor candidates. The goal of the matching is to combine events with similar phase space configurations. The phase space of an event is mostly determined by the hard interaction. The p_{T} of the leading jet is strongly correlated with the momentum transfer Q^2 of the hard scatter. This momentum transfer is of strong influence on the phase space characteristics of the event. Matching events with similar leading jet p_{T}s can be used to select events with similar Q^2 values. The QCD multi-jet sample is dominated by contributions from $2 \to 2$ and $2 \to 3$ processes. To more strongly select events with similar phase space characteristics an additional matching is required between the third acceptor jet and the fourth donor jet. The aim is to select events with similar characteristics between the softest acceptor jet and the (soft) jets to be donated. [3] These jets are referred to as 'match jets'.

The first step of the background generation matches events with similar transverse momenta of the match jets ($\Delta p_{\mathrm{T}} < 1$ GeV/c) and similar leading jet p_{T}s. First a leading jet p_{T} cut of $\Delta p_{\mathrm{T}} < 1$ GeV/c is tried, subsequently increased in steps of 1 GeV/c up to $(\Delta p_{\mathrm{T}})_{\mathrm{max}} = 20$ GeV/c in case no match is found. If no match is found the acceptor event is rejected. Matches which would result in events with unphysical configurations (i.e. any donor jet has a higher p_{T} than any acceptor jet, or any donor jet overlaps within $\Delta\mathcal{R} < 0.5$ with any acceptor jet [4]) are vetoed.

From matching donor-acceptor event pairs a background event is generated by inserting the softest jets from the donor event into the acceptor event. The fifth, sixth and softer jets from six-or-more-jet events are added to matched four-jet events, only the sixth and softer jets are added to a matched five-jet event.

Two versions of the background sample are generated: one starting from the four-jet acceptor sample (`4+2`) and one starting from the five-jet acceptor sample (`5+1`). Neither

[3] An alternative was tried in matching the fourth acceptor jet to the fourth donor jet. The agreement between the generated background and the data was significantly worse than for the third-to-fourth jet matching. This was understood as the fourth-to-fourth matching too strongly matching events with similar leading four jets, leaving too little room for the addition of two extra jets with smaller transverse momenta than the fourth jet.

[4] Due to jet merging (by the jet reconstruction algorithm) jets can be closer together than $\Delta\mathcal{R} = 0.5$. This value for the veto, however, reproduces the radial jet separation distribution in the selected sample.

sample clearly describes the data better/worse; therefore, the central background sample used in the analysis was created using equal contributions from each of the `4+2` and `5+1` samples. Due to the sizes of the acceptor samples the `4+2` sample is significantly larger than the `5+1` sample; however, for simplicity in estimating systematic uncertainties a fifty-fifty mix was adopted for the combined background sample.

The second step in the background generation applies two phase space shaping cuts. These cuts were tuned on the agreement of the generated background sample with the validation sample (see section 7.2):

$\not{E}_T > 5$ GeV The premise of this cut is that, since QCD multi-jet events are not expected to contain significant missing transverse energy, the presence of missing E_T implies the presence of a unreconstructed/unidentified jet. In this case the acceptor event really has a five-jet-like phase space and is especially suited for use in the background sample.

$\not{E}_T/H_{T4} < 0.1$ An upper limit on the missing E_T compared to the p_T sum of the leading four jets (H_{T4}) removes badly reconstructed events as well as events with leptons/neutrinos from semileptonic b-, and c-quark decays.

The distributions of most of the event shape variables are modelled well by the first step of the background model without shaping cuts. In that case, however, the p_T and H_T spectra tend to be too soft.

7.2 Background validation

The performance of the background model is studied in a sample with negligible signal contamination: the five-jet sample. To verify the background modelling a five-jet background validation sample (`4+1`) was generated starting from a four-jet acceptor sample and matching to a five-jet donor sample. [5] Figure 7.2 shows a comparison of the generated `4+1` sample to the original donor sample for several relevant variables (for a discussion of the variables used, please refer to section 8.1). Clearly the `4+1` background sample describes the original five-jet sample well. Overall, the event shape variables are modelled better than the single jet variables like p_T (it should be noted

[5] Again the donor and acceptor samples required respectively five and four jets with $p_T > 15$ GeV/c. The comparison shown is between the tightened samples requiring at least three jets with $p_T > 40$ GeV/c.

that part of the discrepancy in the distributions of the jet transverse momenta is enhanced by low statistics at high p_{T}). Not surprisingly, since the latter are more strongly correlated with the phase space configuration of the event.

The agreement between the generated five-jet background sample and the original five-jet donor sample was quantified using unbinned Kolmogorov-Smirnov tests [134] for a series of fifty-nine relevant variables including all jet p_{T}s, the p_{T} sums of all jets (H_{T}) and of the four leading jets (H_{T4}), the eigenvalues $\lambda_{1,2,3}$ of the momentum tensor, the centrality H_{T}/H, as well as several masses, jet rapidities and rapidity differences.

The Kolmogorov-Smirnov test is a goodness-of-fit test to estimate the probability that two observed sets of observations originate from the same underlying distribution. The test is based on the supremum of the difference between the cumulative distributions of the two sets of observations, which makes it sensitive to differences in both shape and position of the distributions.

For a large number of variables one would expect an approximately uniform distribution. Figure 7.3 shows the distribution of the Kolmogorov-Smirnov test results, demonstrating this distribution is indeed (within statistics) flat.

The background generation and validation procedure was also performed using looser b-id criteria, NN > 0.2. For that case the validation procedure showed significantly less agreement then for the NN > 0.65 case. This contributed to the choice of the 0.65 operating point for the b-tagging.

7.3 Systematic studies of the background modelling

The crucial piece of the background modelling is the matching and shaping of the generated phase space. Two individual systematic uncertainties, both related to the phase space occupied by the generated background sample, are considered. The effects of these systematic uncertainties on the generated background samples are discussed in the following sections, the influence of these systematic background variations on the measured cross section are summarised together with all other systematic uncertainties in section 9.2.

7.3.1 Phase space of the acceptor sample

The difference between the `4+2` and `5+1` background samples. This shows the sensitivity of the background model to the intrinsic phase space of the acceptor sample. Figures 7.4

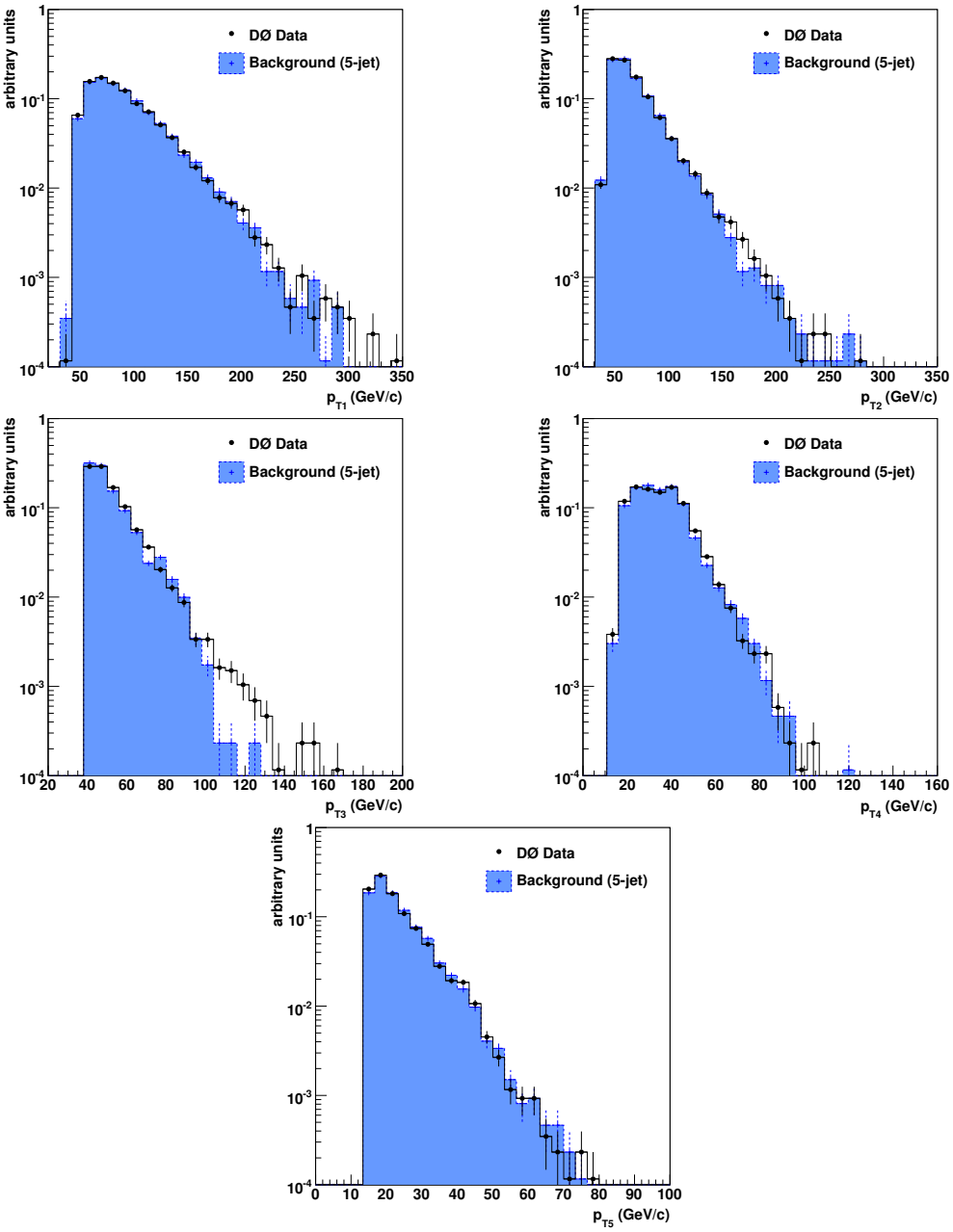


Figure 7.2: *Comparison between the 4+1 background validation sample and the original five-jet donor sample.*

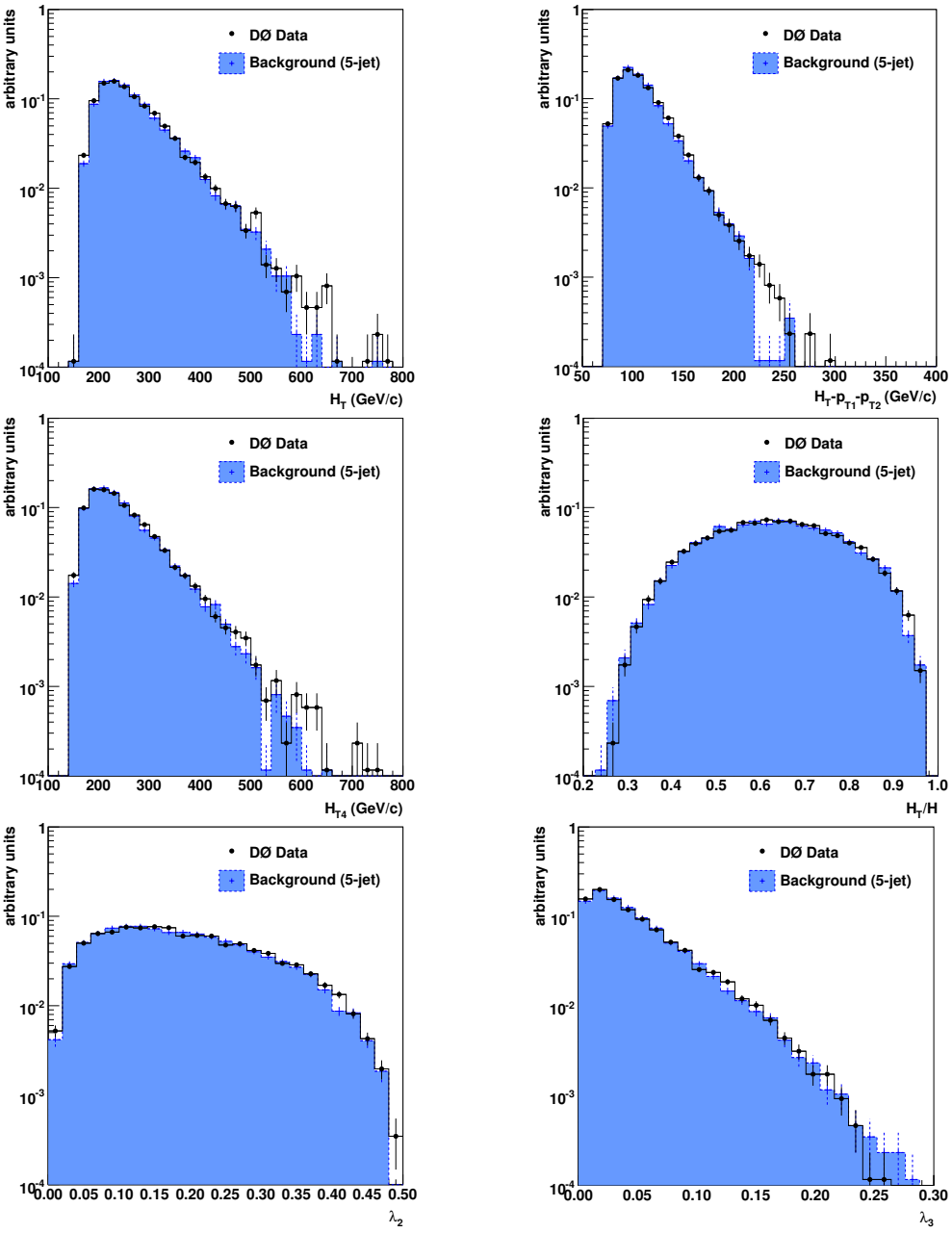


Figure 7.2: *(Cont.) Comparison between the 4+1 background validation sample and the original five-jet donor sample.*

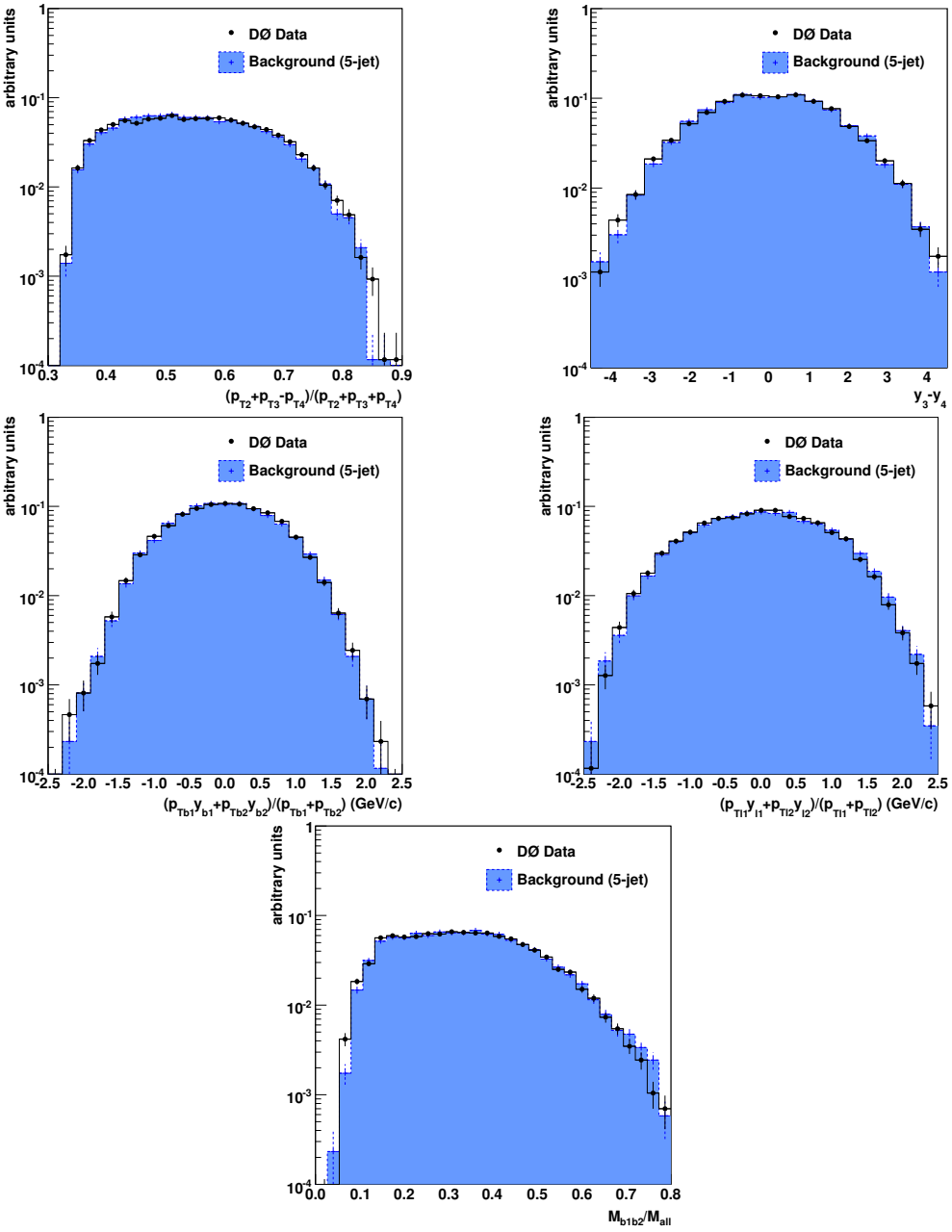


Figure 7.2: *(Cont.) Comparison between the 4+1 background validation sample and the original five-jet donor sample.*

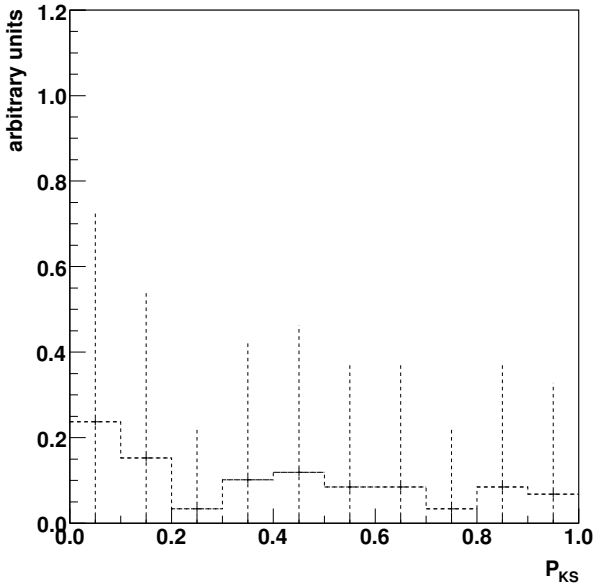


Figure 7.3: *Distribution of the Kolmogorov-Smirnov test results corresponding to figure 7.2.*

and 7.5 show a comparison between these two different background samples and the central (mixed) background sample. The two samples differ significantly in distribution shape for some of the test variables. Most of these differences are directly related to the intrinsic phase space differences between the four-jet and five-jet donor samples. This is also reflected by the fact that the largest differences occur for the jet transverse momenta: the individual jet momenta are more sensitive to phase space effects than the global event shape variables.

When considering the jet transverse momenta it's important to keep in mind the origin of the jets. For both the `4+2` and `5+1` samples the leading four jets originate from the acceptor event, but from four-jet and five-jet acceptor events respectively. The fifth jet in the `4+2` sample is the first donor jet, coming from a six-jet event, for `5+1` event the fifth jet is the softest acceptor jet. For both background samples the sixth and softer jets were part of the six-jet donor event.

The p_{T} distributions of the leading two jets are very similar (i.e. the same within statistics) in the `4+2` and `5+1` samples, demonstrating the dominance of $2 \to 2$ processes. The difference between the p_{T} distributions of the third jet is partly due to the Q^2 matching: four-jet events at a given Q^2 tend to have harder third jets than five-jet

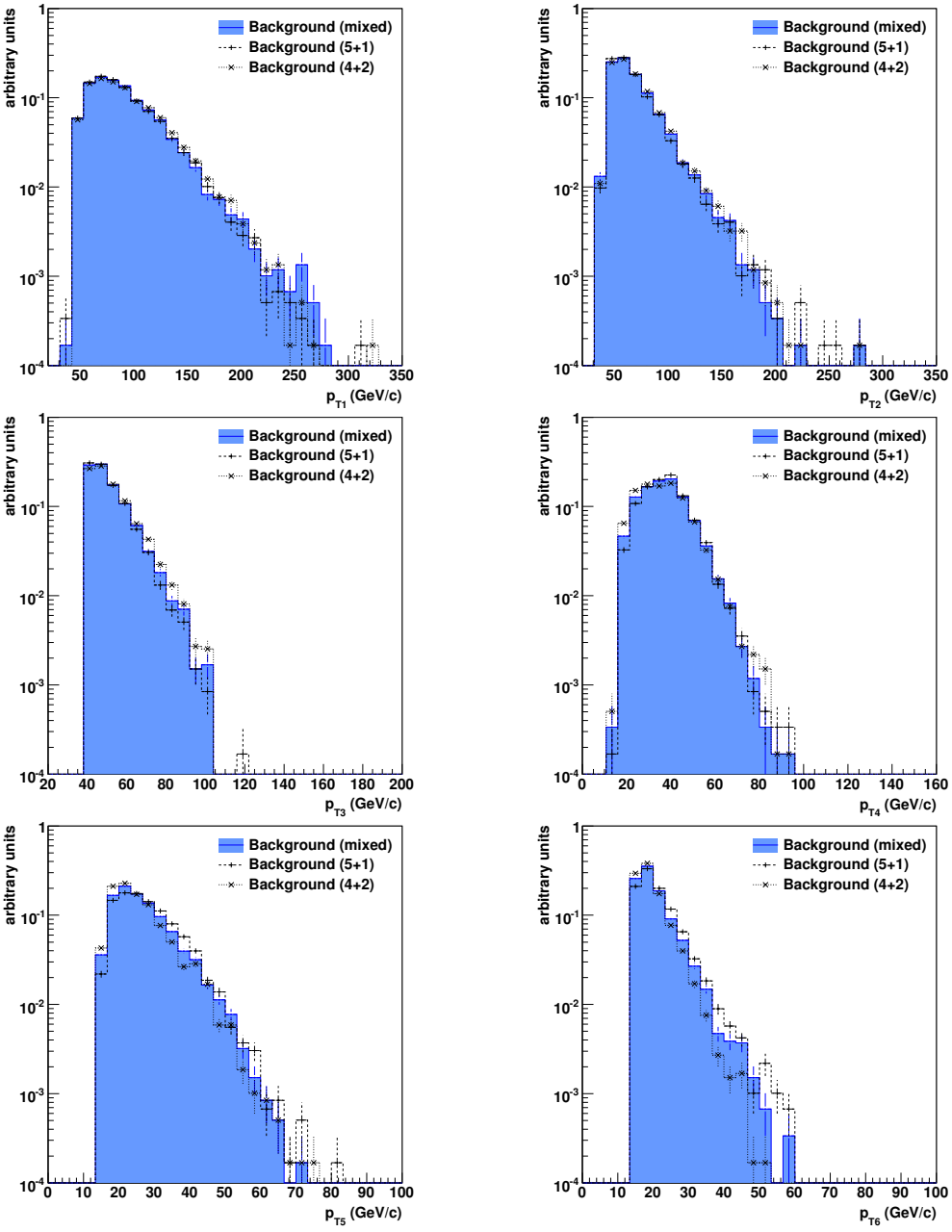


Figure 7.4: *Comparison between the 4+2 (× markers) and 5+1 (+ markers) background samples and the central background sample (solid fill) consisting of a fifty-fifty mix of the 4+2 and 5+1 samples.*

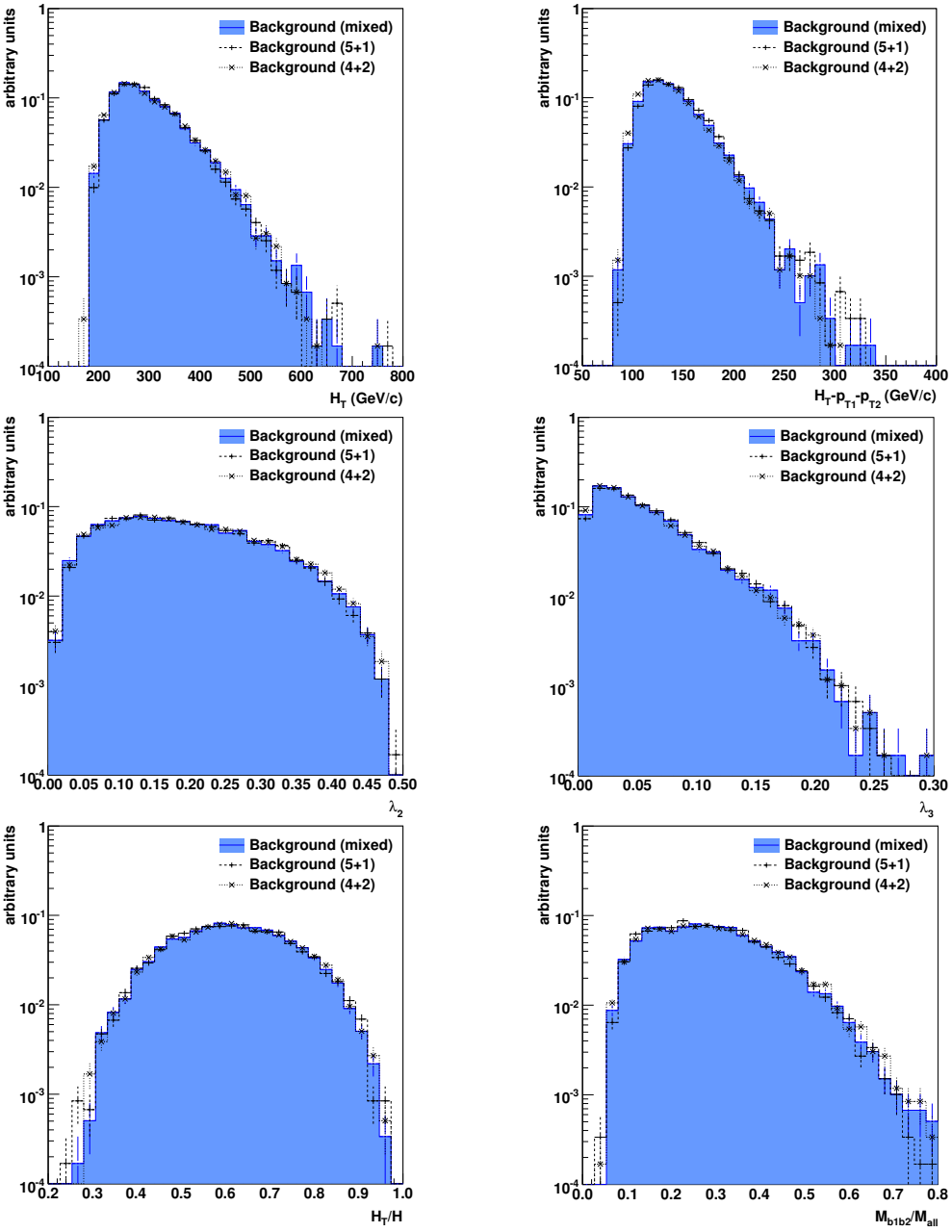


Figure 7.4: *(Cont.) Comparison between the 4+2 (× markers) and 5+1 (+ markers) background samples and the central background sample (solid fill) consisting of a fifty-fifty mix of the 4+2 and 5+1 samples.*

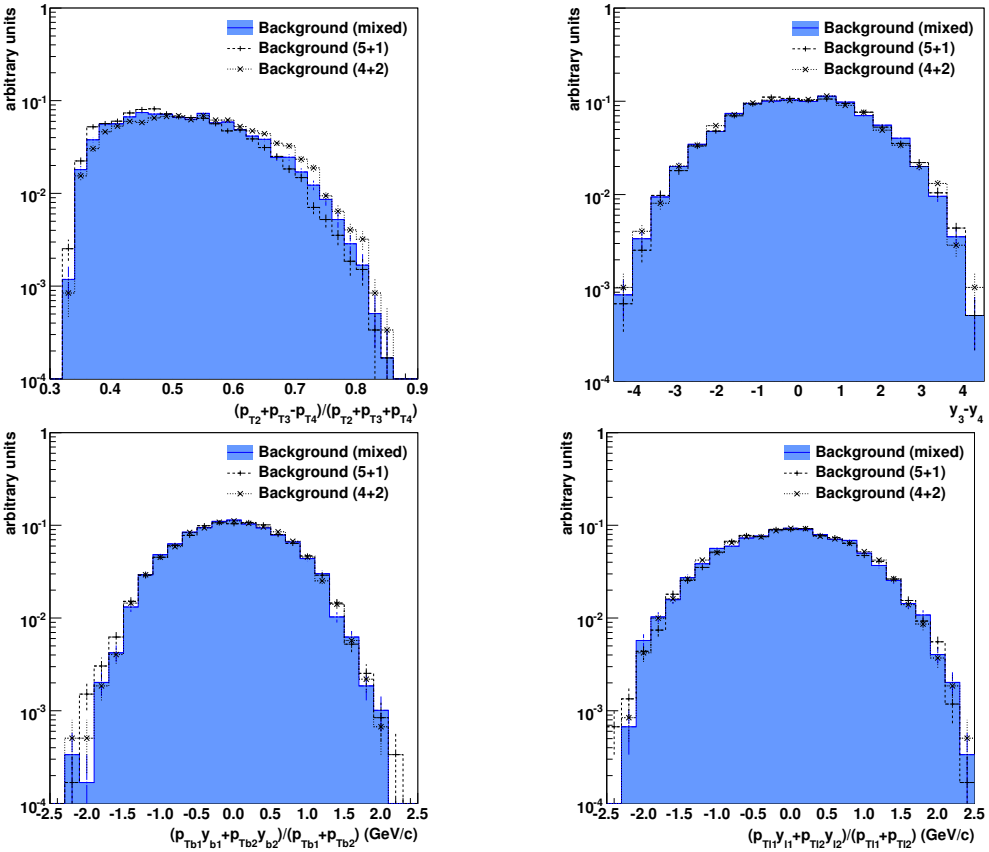


Figure 7.4: *(Cont.) Comparison between the 4+2 (× markers) and 5+1 (+ markers) background samples and the central background sample (solid fill) consisting of a fifty-fifty mix of the 4+2 and 5+1 samples.*

events with similar Q^2. A similar feature can be seen in the distribution of the fifth jet p_{T}: the fifth jet in five-jet events (the 5+1 histogram) tends to be harder than in six-jet events (the 4+2 histogram). The sixth jet in both background samples always originates from the donor sample; the differences between the sixth jet p_{T} spectra in the 4+2 and 5+1 samples directly reflect the phase space differences between six-jet events matched to four-jet and to five-jet events.

It is interesting to note the difference in shape of the $(p_{\mathrm{T}2}+p_{\mathrm{T}3}-p_{\mathrm{T}4})/(p_{\mathrm{T}2}+p_{\mathrm{T}3}+p_{\mathrm{T}4})$ variable distributions in figure 7.4. This variable is related to the asymmetry discussed in section 1.3.1. The observed difference in this case, however, is just a result

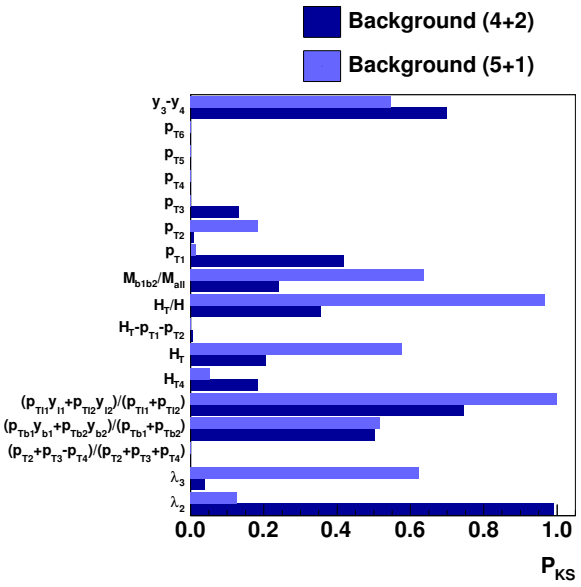


Figure 7.5: *Kolmogorov-Smirnov test results from comparing the 4+2 and 5+1 background samples to the central background sample. The disagreement for the jet p_Ts is emphasised by low statistics at high p_T. This explains why the agreement for H_T is still reasonable.*

of the differences in third jet p_{T} spectra. The second and fourth jets have very similar p_{T} distributions, but the softer third jet in the 5+1 sample moves this asymmetry variable towards lower values.

7.3.2 Variations of the shaping cuts

The effect of modifying the shaping cuts. The shaping cuts were optimised on the agreement between the 4+1 validation sample and the original five-jet donor sample. The lower bound on $\not{E}_{\mathrm{T}}$ opens up available phase space for the addition of jets and is an essential part of the background model. Varying the $\not{E}_{\mathrm{T}}/H_{\mathrm{T4}}$ cut allows for more or less phase space for additional jets. It was decided to keep the $\not{E}_{\mathrm{T}}$ cut fixed and only vary the $\not{E}_{\mathrm{T}}/H_{\mathrm{T4}}$ cut to reflect the effects of the choice of the shaping cuts. 'Up' and 'down' background samples were generated using $\not{E}_{\mathrm{T}}/H_{\mathrm{T4}} < 0.12$ and $\not{E}_{\mathrm{T}}/H_{\mathrm{T4}} < 0.08$ respectively. The effects are shown in figures 7.7 and 7.6. Clearly the upper shaping cut is not able to change the agreement between the 'up' and 'down' backgrounds and the central value significantly. This is most clear from the comparison of the

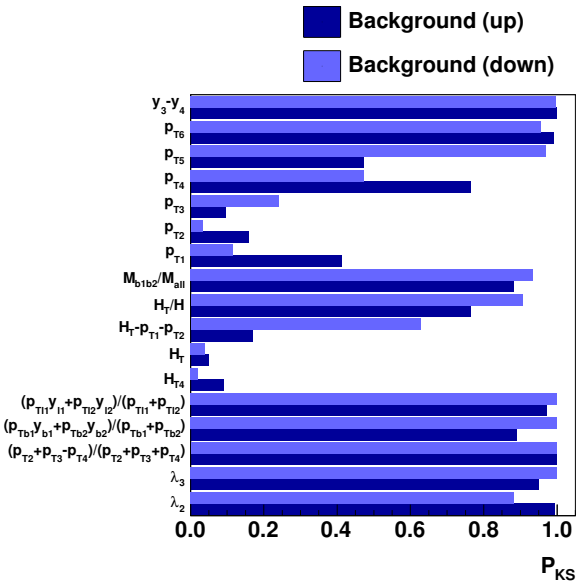


Figure 7.6: *Kolmogorov-Smirnov test results from comparing the 'up' and 'down' background samples to the central background sample.*

Kolmogorov-Smirnov tests between fluctuated and central background samples shown in figure 7.6.

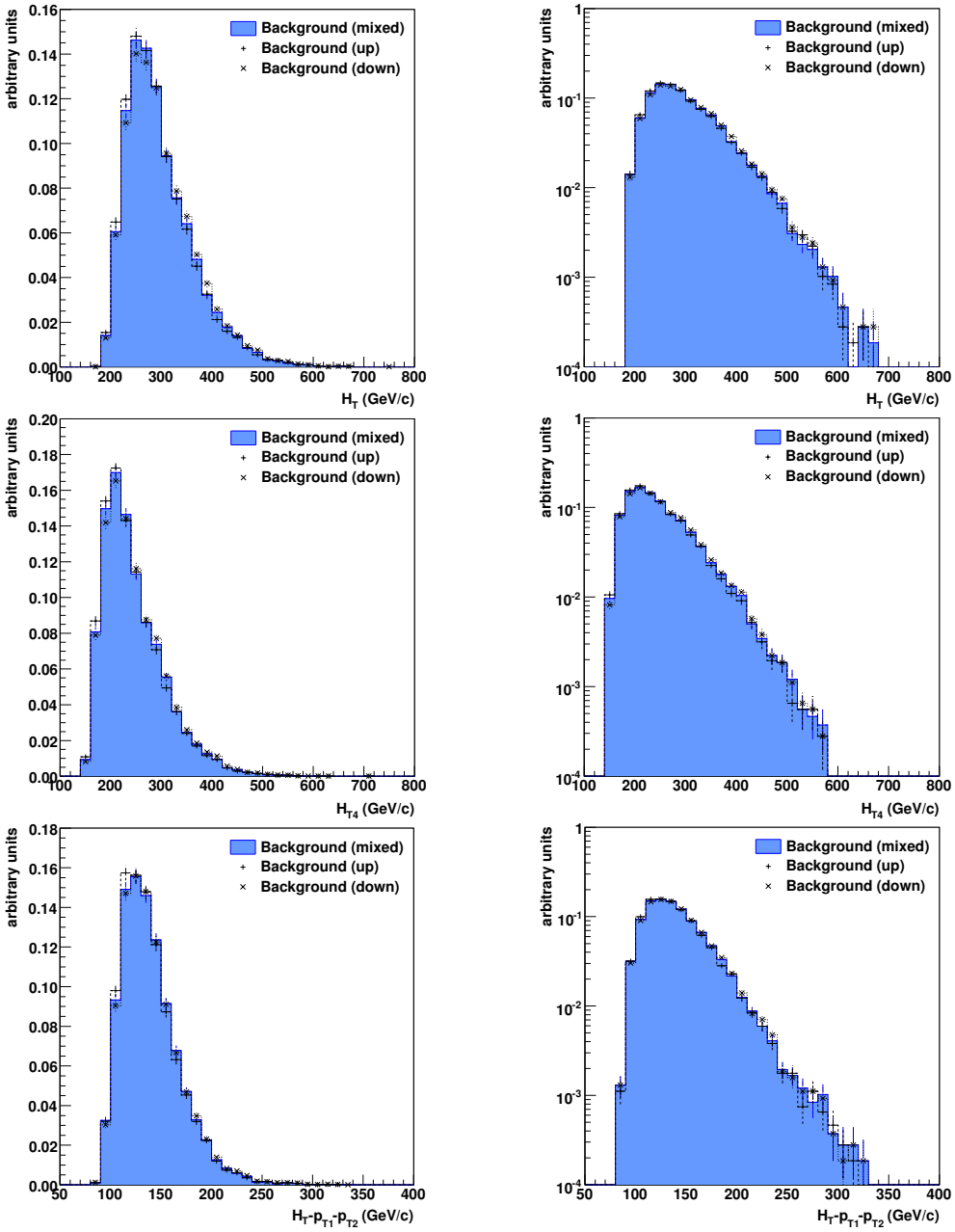


Figure 7.7: *Comparison between the 'up' (× markers) and 'down' (+ markers) background samples and the central background sample (solid fill). Left-hand side: linear scale, right-hand side: logarithmic scale.*

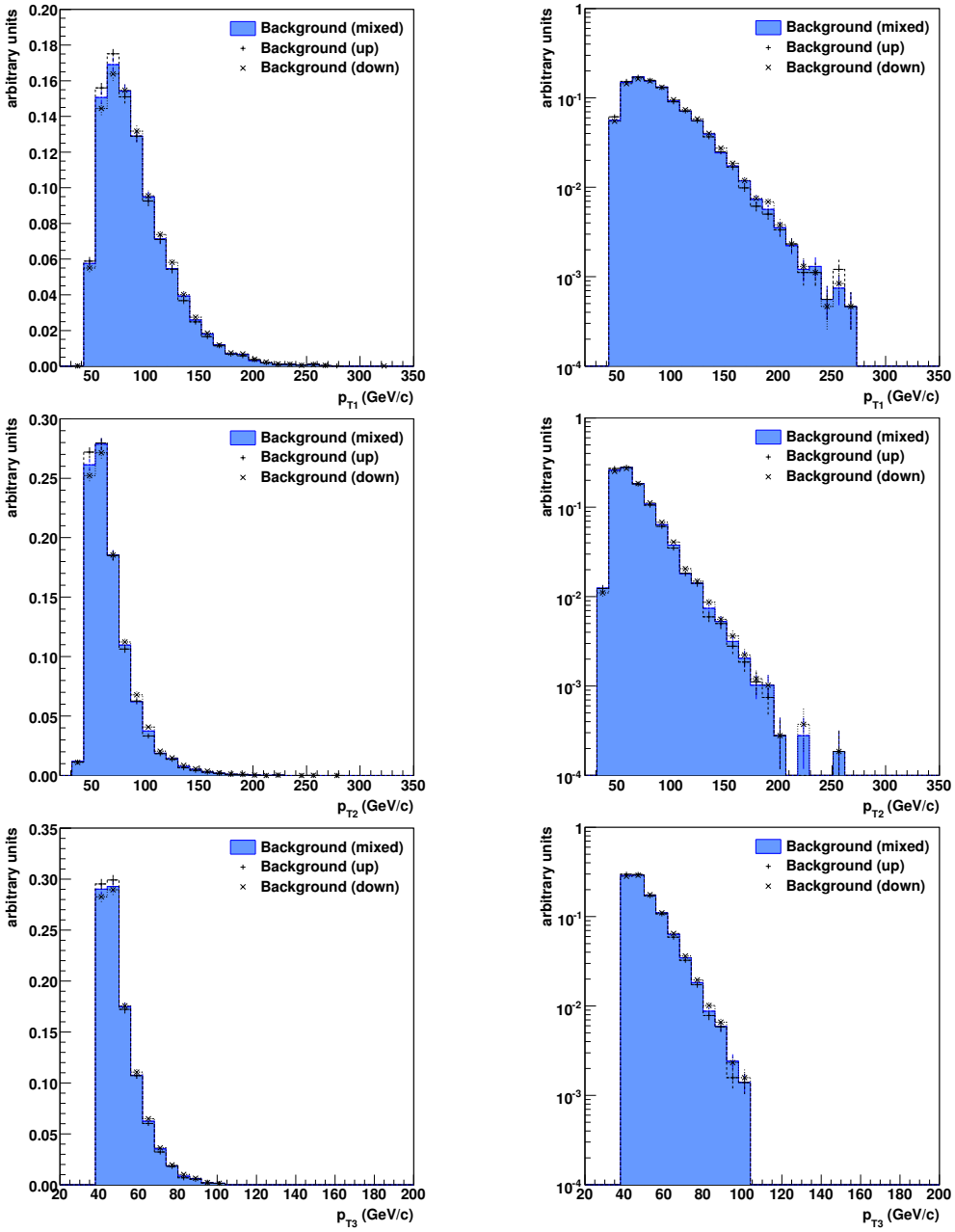


Figure 7.7: *(Cont.) Comparison between the 'up' (× markers) and 'down' (+ markers) background samples and the central background sample (solid fill). Left-hand side: linear scale, right-hand side: logarithmic scale.*

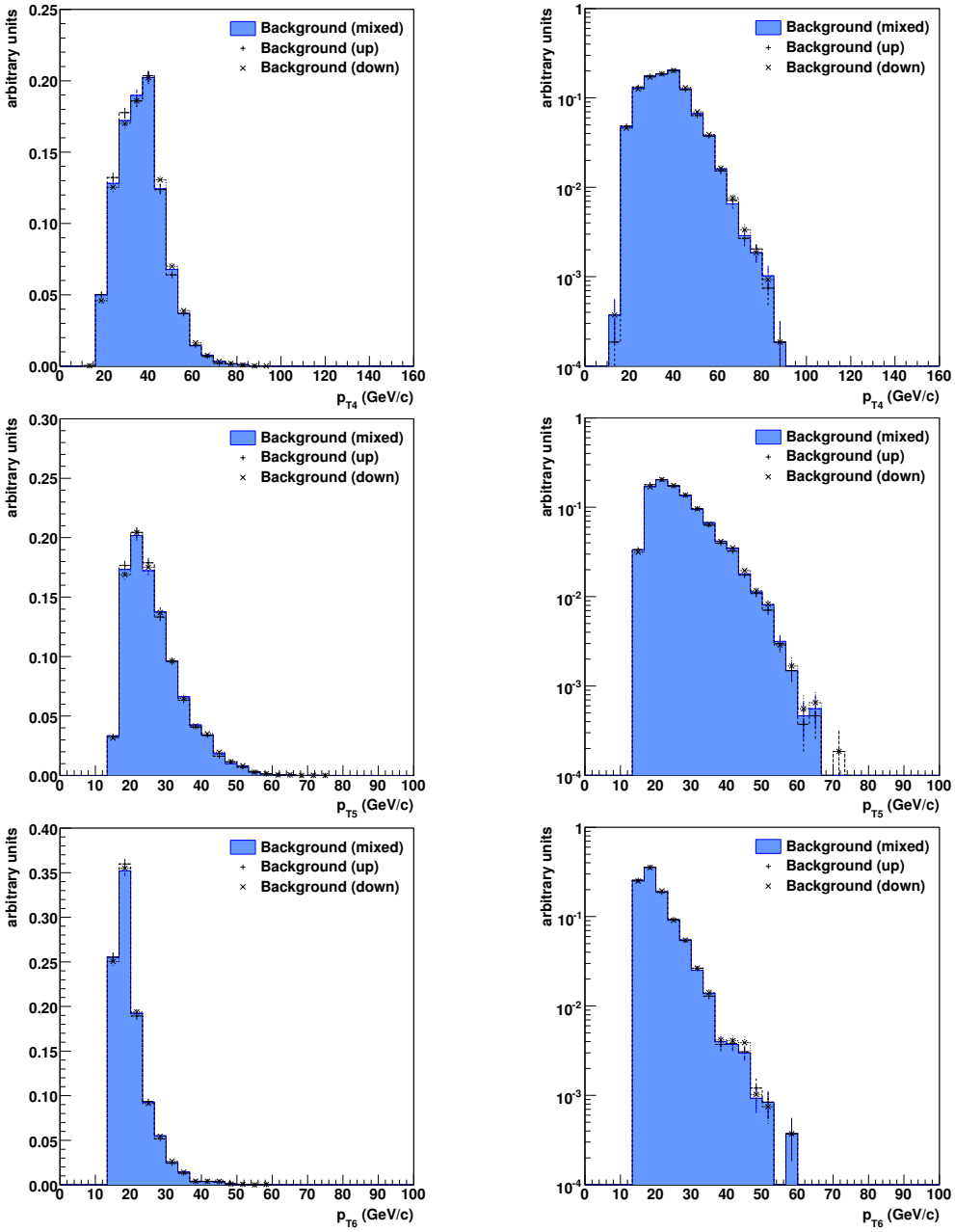


Figure 7.7: *(Cont.) Comparison between the 'up' (× markers) and 'down' (+ markers) background samples and the central background sample (solid fill). Left-hand side: linear scale, right-hand side: logarithmic scale.*

Chapter 8

Separating signal from background

After the selection described in chapter 5 the signal-to-background ratio in the selected data sample is approximately one to ten. The signal and background fractions in data were estimated using a template-based likelihood method (using version v3.8.11 of the TMVA [135] package).

The probability for an event in the measured data set $\vec{x}$ to be a signal event is given by:

$$P(S|\vec{x}) = \frac{P(\vec{x}|S)P(S)}{P(\vec{x}|S)P(S) + P(\vec{x}|B)P(B)}$$

where $P(\vec{x}|S)$ represents the probability to observe the measured $\vec{x}$ for signal events. $P(\vec{x}|S)P(S)$ represents exactly the contents of all histogram bins to which the $\vec{x}$ contribute, for the data under consideration.

In analogy, the likelihood ratio for an event i with observed set of characteristic variables $\vec{x}_i = \{x_k\}_i$ is defined as:

$$R(\vec{x}_i) = \frac{L(S;\vec{x}_i)}{L(S;\vec{x}_i) + L(B;\vec{x}_i)}$$

where L_S (L_B) is the likelihood function or joint probability density function for event i to represent signal (background). For independent variables x_k the likelihood function is simply the product of all individual probability densities: $L(S;\vec{x}) = \prod_k P_S(x_k)$. For convenience the ratio $R(\vec{x}_i)$ is written as

$$\begin{aligned} R(\vec{x}_i) &= \frac{\prod_k P_S(x_{k,i})}{\prod_k P_S(x_{k,i}) + \prod_k P_B(x_{k,i})} \\ &= \frac{\exp\left(\sum_k \ln\left(P_S(x_{k,i})/P_B(x_{k,i})\right)\right)}{\exp\left(\sum_k \ln\left(P_S(x_{k,i})/P_B(x_{k,i})\right)\right) + 1}. \end{aligned} \tag{8.1}$$

The ratio $R(\vec{x}_i)$ represents the normalised signal probability for event i, a measure of the likelihood the event is a signal event.

Using signal and background 'training' samples, characteristic distributions can be obtained for each of the input variables x_k for both types of events. To suppress sensitivity to statistical fluctuations in the 'training' samples the input distributions are smoothed and fitted with splines.

The values for $P_{S/B}(x_{k,i})$ in eq. (8.1) can now be evaluated from the signal and background variable templates and used to obtain templates for the likelihood output distribution R. These likelihood output templates combine the separating power of all input variables, and can be used in a fraction-fit to the likelihood output distribution obtained from data to estimate the signal content.

After an overview of these input variables (section 8.1), the likelihood output distributions will be discussed in section 8.2.

8.1 Likelihood input variables

All input variables were chosen based on the following criteria:

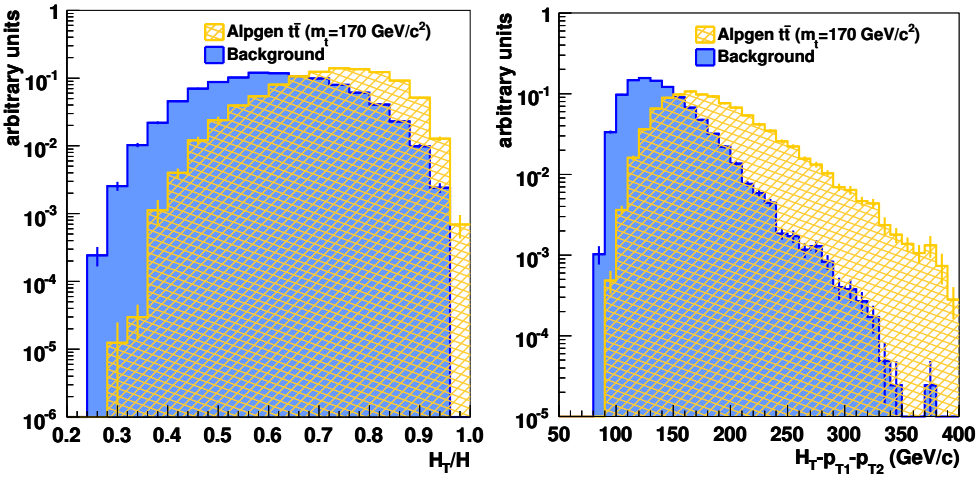
- they should contribute to separation between signal and background,
- they should be well-modelled in the background validation sample and
- they should introduce as little as possible (top) mass dependence into the likelihood.

The following nine variables were used:

H_T/H (Figure 8.1a)
The centrality is an event shape variable defined as the ratio of the sum of the jet transverse momenta to the sum of the jet energies:

$$\frac{H_{\mathrm{T}}}{H} \equiv \frac{\sum_{\mathrm{jets}} p_{\mathrm{T\,jet}}}{\sum_{\mathrm{jets}} E_{\mathrm{jet}}}.$$

It provides a handle on the mass of the particles produced in the hard scatter. High mass multi-jet events, like the $\mathrm{t\bar{t}}$ signal, are expected to be more central than general QCD multi-jet events.



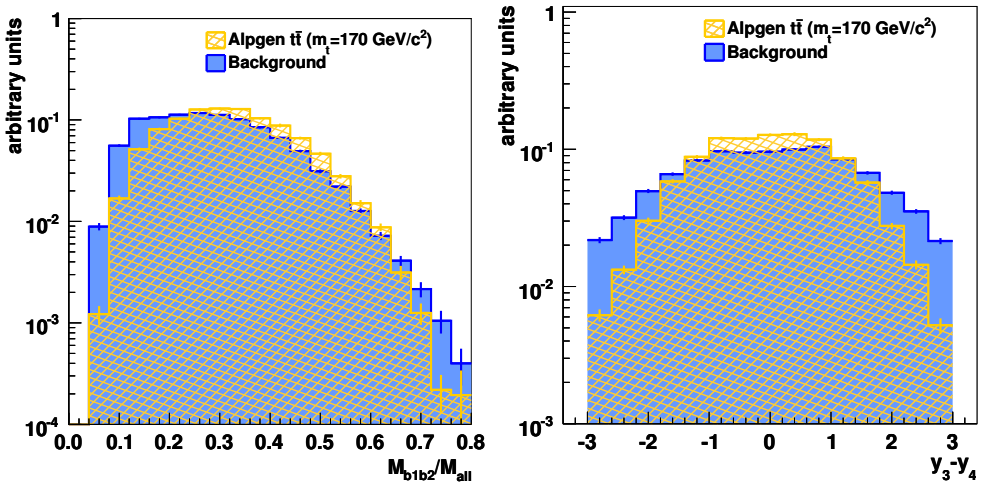
(a) *(b)*

Figure 8.1: *Distributions of the likelihood input variables for the $t\bar{t}$ signal (hashed) and for the multi-jet QCD background (solid fill).*

$H_T - p_{T1} - p_{T2}$ (Figure 8.1b)
The 'reduced H_T' is used as a compromise between H_T and introducing a mass dependence. H_T itself provides strong separation between signal and background, but is strongly correlated with the top quark mass. Subtracting the two leading jet p_Ts suppresses this mass dependence while retaining most of the separation.

M_{b1b2}/M_{all} (Figure 8.1c)
Unlike in $t\bar{t}$ decays, the QCD background contains a significant amount of $b\bar{b}$ pairs from gluon splitting. Those $b\bar{b}$ pairs are expected to be nearly collinear and have small invariant mass. This contribution is visible in the M_{b1b2}/M_{all} distribution as the additional contribution on the left-hand side of the background distribution compared to the signal. QCD $b\bar{b}$ events (with additional jet radiation to pass the six-jet selection) tend to contain most of the jet mass in the $b\bar{b}$ jet pair, while in $t\bar{t}$ decay the two additional jet pairs make up approximately twice the real W mass. This explains the difference between the background and signal distributions on the right-hand side.

$y_3 - y_4$ (Figure 8.1d)
In the case of hadronic $t\bar{t}$ decays, the two leading jets are often the b-jets from the top



(c) *(d)*

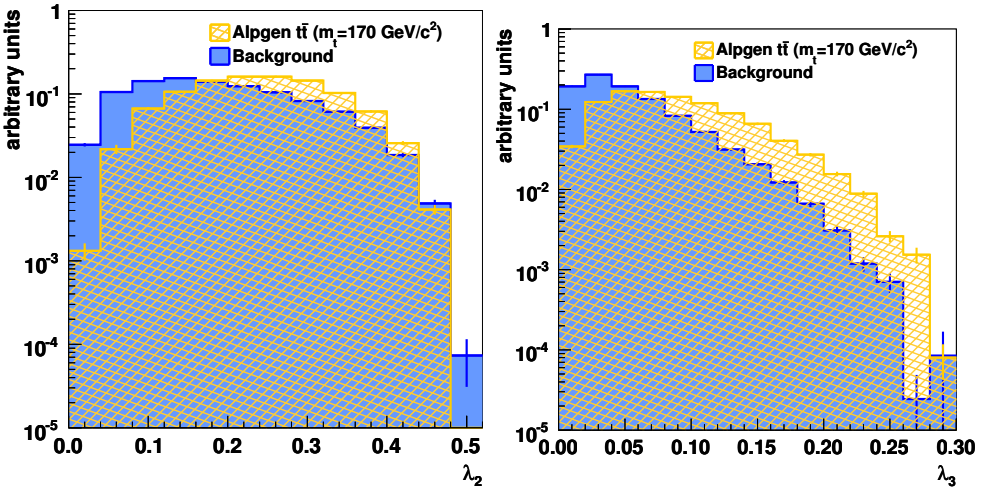
Figure 8.1: *(Cont.) Distributions of the likelihood input variables for the $t\bar{t}$ signal (hashed) and for the multi-jet QCD background (solid fill).*

quark decays. The third and fourth jets tend to form a quark-antiquark pair from the decay of the same W boson and to be closer together than is the case for random jet pairs. In the $t\bar{t}$ signal the rapidity difference between these jets provides an estimate of the average separation between the two jets from the leading W boson. The QCD background is dominated by di-jet events with additional radiative jets. For those events the third and fourth jets will generally be in opposite hemispheres. For the case of hard QCD multi-jet events all jets will be spread widely apart.

λ_2, λ_3 (Figures 8.1e and 8.1f.)
Both λ_2 and λ_3 are eigenvalues of the normalised squared momentum tensor [136]:

$$M^{\alpha,\beta} = \frac{\sum_i p_i^\alpha p_i^\beta}{\sum_i |\vec{p}_i|^2}$$

where the index i runs over all jets in the event and $\alpha, \beta = 1, 2, 3$ denote the three spatial components of the jet momenta. This tensor has three eigenvalues, defined (in decreasing order) as λ_1, λ_2 and λ_3. The sum of all three eigenvalues is one. The sum of λ_2 and λ_3 is a measure of the spherical symmetry of the event. The third eigenvalue of the squared momentum tensor, λ_3, is equal to the smallest sum of squared di-jet



(e) *(f)*

Figure 8.1: *(Cont.) Distributions of the likelihood input variables for the $t\bar{t}$ signal (hashed) and for the multi-jet QCD background (solid fill).*

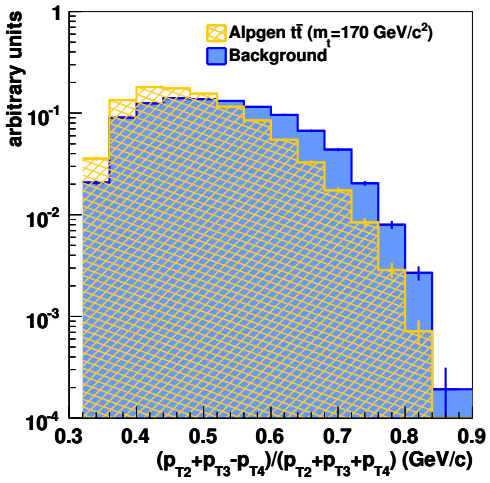
pair transverse momenta. λ_3 Determines the aplanarity of an event: a measure of how badly all jets fit in a single plane.

The QCD background is dominated by di-jet and three-jet events, often boosted, reducing their sphericity. Even though the $t\bar{t}$ signal events are basically di-jets, the large top pair mass prevents large boosts; in addition the two-step decay involving the (heavy) W bosons increases their sphericity. Compared to the QCD background hadronic $t\bar{t}$ events are (on average) more spherically symmetric and have correspondingly larger values of λ_2 and λ_3.

By introducing λ_2 and λ_3 as separate input variables, correlations between the input variables can be avoided.

$(p_{T2} + p_{T3} - p_{T4})/(p_{T2} + p_{T3} + p_{T4})$ (Figure 8.1g)

This variable is correlated with the energy asymmetry in quark-antiquark pairs from W decays where the W originates from a top quark, as described in section 1.3.1. In data, it is unknown exactly which jets originated from the W bosons. The two b-jets tend to be the leading jets. The third and fourth jets would then stem from the highest-p_T W boson. To account for the fact that the b-jets may not be the leading jets, but may be softer than one of the W's jets, the calculated asymmetry treats the



(g)

Figure 8.1: *(Cont.) Distributions of the likelihood input variables for the $t\bar{t}$ signal (hashed) and for the multi-jet QCD background (solid fill).*

second and third jets equally. Even though the effect is diluted by these assumptions, as well as by the use of the laboratory frame instead of the (anti)top quark rest frame, the expected tendency of the signal towards smaller asymmetries is still clearly present. The shape of the distribution for the QCD background is dominated by the relative contributions of the hard processes. For hard $2 \to 3$ or $2 \to 4$ processes the three leading jets are of approximately equal p_{T} and the value of the asymmetry lies around 1/3. For the dominant $2 \to 2$ processes, however, the p_{T} difference between the second and third jets is larger than between the leading two jets, pushing the asymmetry distribution towards higher values.

$(p_{\mathrm{T}b1}y_{b1} + p_{\mathrm{T}b2}y_{b2})/(p_{\mathrm{T}b1} + p_{\mathrm{T}b2})$ (Figure 8.1h)

The p_{T}-weighted rapidity sum of the leading two b-jets is related to the mass of the all-jet system. The large top quark mass prevents large boosts along the z-direction. This maintains relatively small jet rapidities while at the same time keeping the b-jets more back-to-back than is the case for QCD $b\bar{b}$ production. Both effects lead to a narrower distribution for the $t\bar{t}$ signal than for the QCD background.

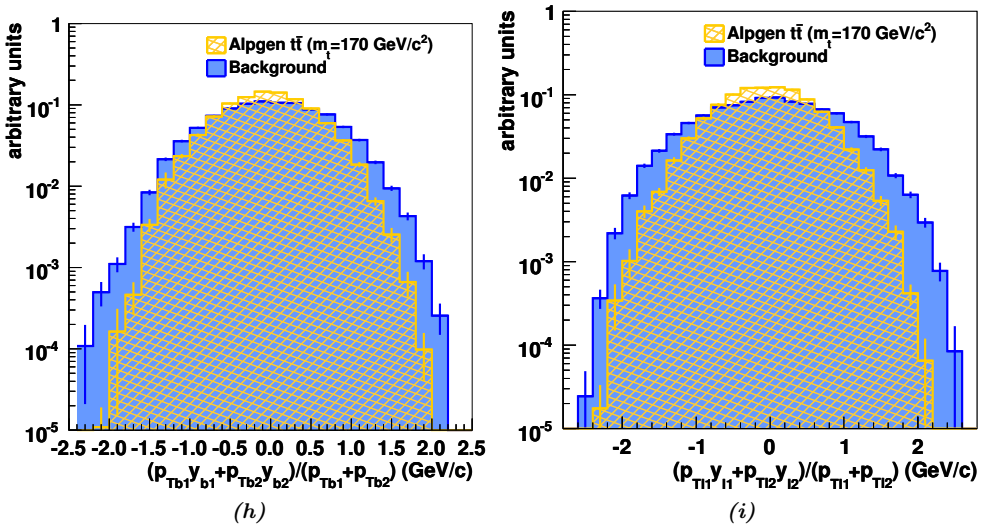


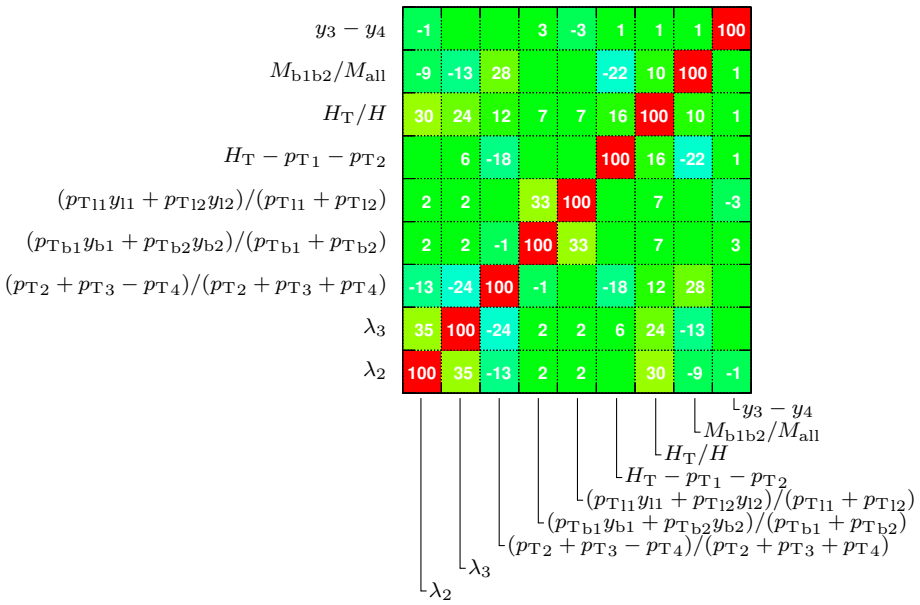
Figure 8.1: *(Cont.) Distributions of the likelihood input variables for the $t\bar{t}$ signal (hashed) and for the multi-jet QCD background (solid fill).*

$(p_{T\,l1}\,y_{l1} + p_{T\,l2}\,y_{l2})/(p_{T\,l1} + p_{T\,l2})$ (Figure 8.1i)
This variable represents the light-jet version of the previous parameter. The higher light-jet multiplicity broadens the distribution.

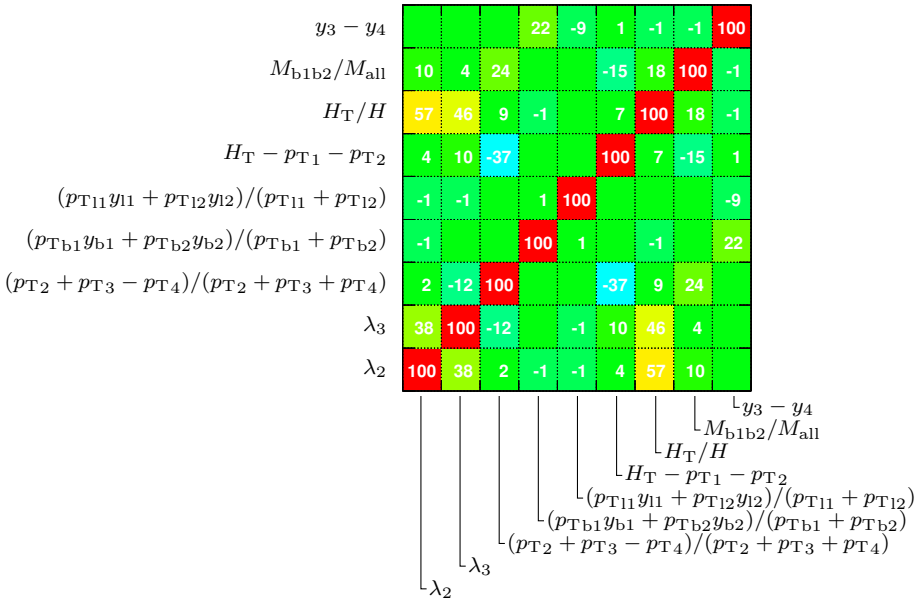
None of the above variables possess enough separating power by themselves to distinguish signal from background. Combination of all variables in a likelihood, however, does provide significant signal-background separation. As shown in figure 8.2, the individual input variables are only weakly correlated. While these correlations (as well as possible higher-order inter-variable dependencies) will lead to suboptimal separation, the likelihood output distributions for signal and background (section 8.2) show this to be an insignificant effect for the case at hand.

8.2 Likelihood 'training' and performance

The likelihood templates for the $t\bar{t}$ signal were determined from the $m_t = 170$ GeV/c^2 ALPGEN all-hadronic $t\bar{t}$ sample. The 50–50 mixed background sample from chapter 7 was used. The signal and background samples were randomly split into two halves each: one used to determine the templates for the likelihood analysis and one for use as a control sample to verify the likelihood performance.



(a) ALPGEN $\mathrm{t\bar{t}}$ signal with $m_{\mathrm{t}} = 170$ GeV/c^2



(b) Background

Figure 8.2: *Correlations between all likelihood input variables for the $\mathrm{t\bar{t}}$ signal (a) and for the QCD background (b).*

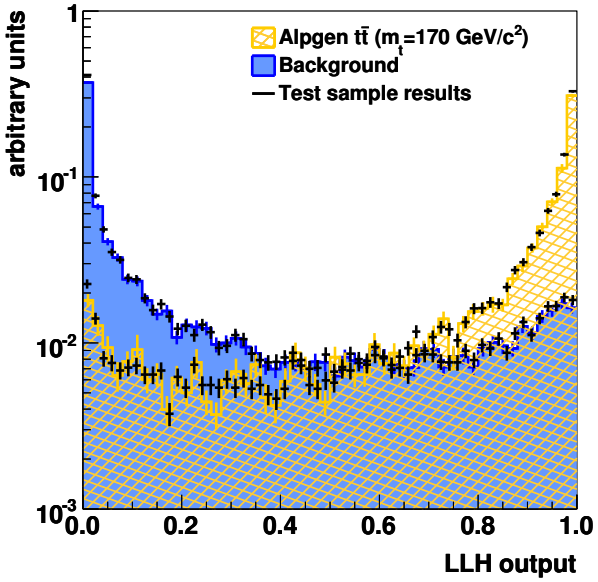


Figure 8.3: *Likelihood output distributions for signal and background together with an 'overtraining check'. Shown are the signal and background templates (hashed and solid fills respectively) and the likelihood output from the control samples (the second halves of the 'training' samples). The agreement between the templates and the control samples shows there is no sign of overtraining. Please note that the templates and the control results have a slightly different binning; the latter have been scaled to match the template integrals.*

The likelihood output distributions for signal and background are shown in figure 8.3. As expected, the signal probability peaks at one, the background probability at zero. The 'wrong-side peaks', for signal (at zero) and for background (at one), are due to the (weak) correlations between the input variables.

Figure 8.3 also shows an overtraining check performed on the used likelihood. Overtraining occurs when the 'training' samples used to determine the signal and background templates are too small to model all sample variations. In that case the performance of the likelihood will be much better on the training samples than on the control samples. The good agreement between the results from the control samples and the templates themselves shows overtraining is no issue in this case.

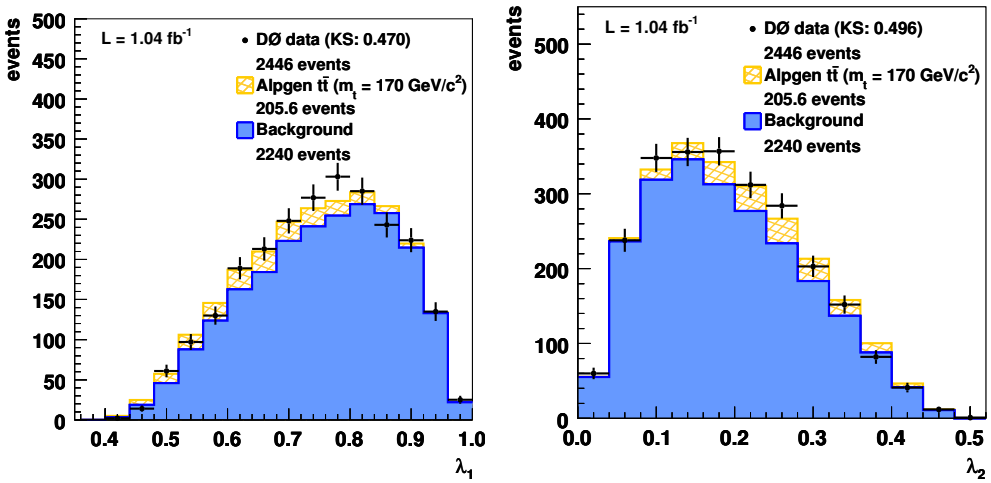


Figure 8.4: *Results from the likelihood fit compared to data. Shown are the background contribution (solid fill), the $t\bar{t}$ signal contribution (hashed) and the data (black markers).*

8.2.1 Likelihood control plots

To estimate the signal fraction the likelihood output distribution in data is fitted with a linear combination of the signal and background templates shown in figure 8.3. To verify the likelihood procedure, the resulting signal-background mix is compared to the data for a series of control variables, including:

- all likelihood input variables,
- the (second through ninth) Fox-Wolfram moments [137] representing different event shape measures,
- all (b-, and light-) jet p_{T}s and rapidities,
- the maximum rapidity difference ΔY_{max} between any two jets, and
- several other inter-jet rapidity differences, e.g. between the leading two b-jets, $y_{\mathrm{b1}} - y_{\mathrm{b2}}$, and between the leading b-, and light-jets, $y_{\mathrm{b1}} - y_{\mathrm{l1}}$.

Figure 8.4 shows this comparison for several of the control variables.

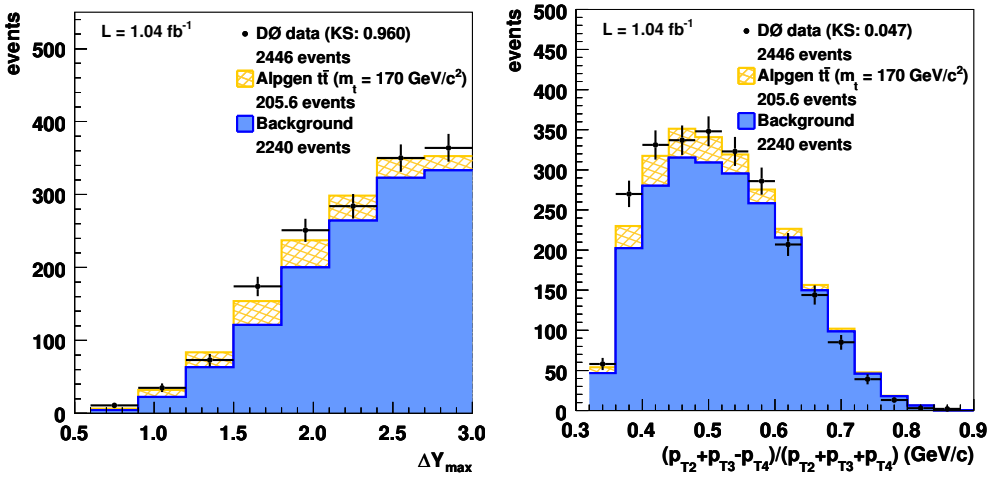


Figure 8.4: *(Cont.) Results from the likelihood fit compared to data. Shown are the background contribution (solid fill), the $t\bar{t}$ signal contribution (hashed) and the data (black markers).*

The agreement is quantified using a binned [1] Kolmogorov-Smirnov test, resulting in a flat test result distribution. The Kolmogorov-Smirnov test results are included in the legends (the 'KS' values). All other control plots, as well as additional linear/logarithmic versions of the plots shown, can be found in ref. [109]. Figure 8.4 demonstrates that the likelihood fit based on the selected signal and background samples adequately describes the data. The worse agreement for the jet transverse momenta is influenced by the imperfect jet p_{T} modelling in the background model discussed in chapter 7.

[1] For visual reasons some of the histograms in figure 8.4 have been rebinned. All Kolmogorov-Smirnov test were performed on the histograms before rebinning.

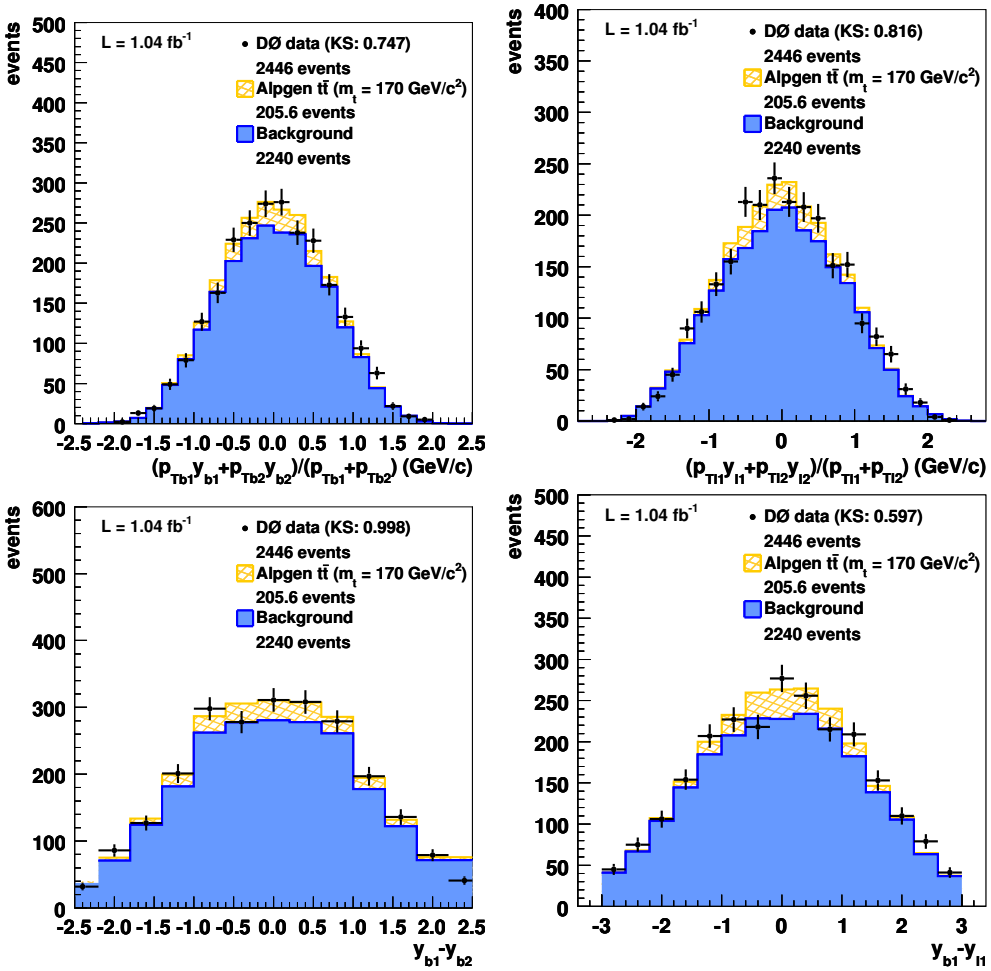


Figure 8.4: *(Cont.) Results from the likelihood fit compared to data. Shown are the background contribution (solid fill), the $t\bar{t}$ signal contribution (hashed) and the data (black markers).*

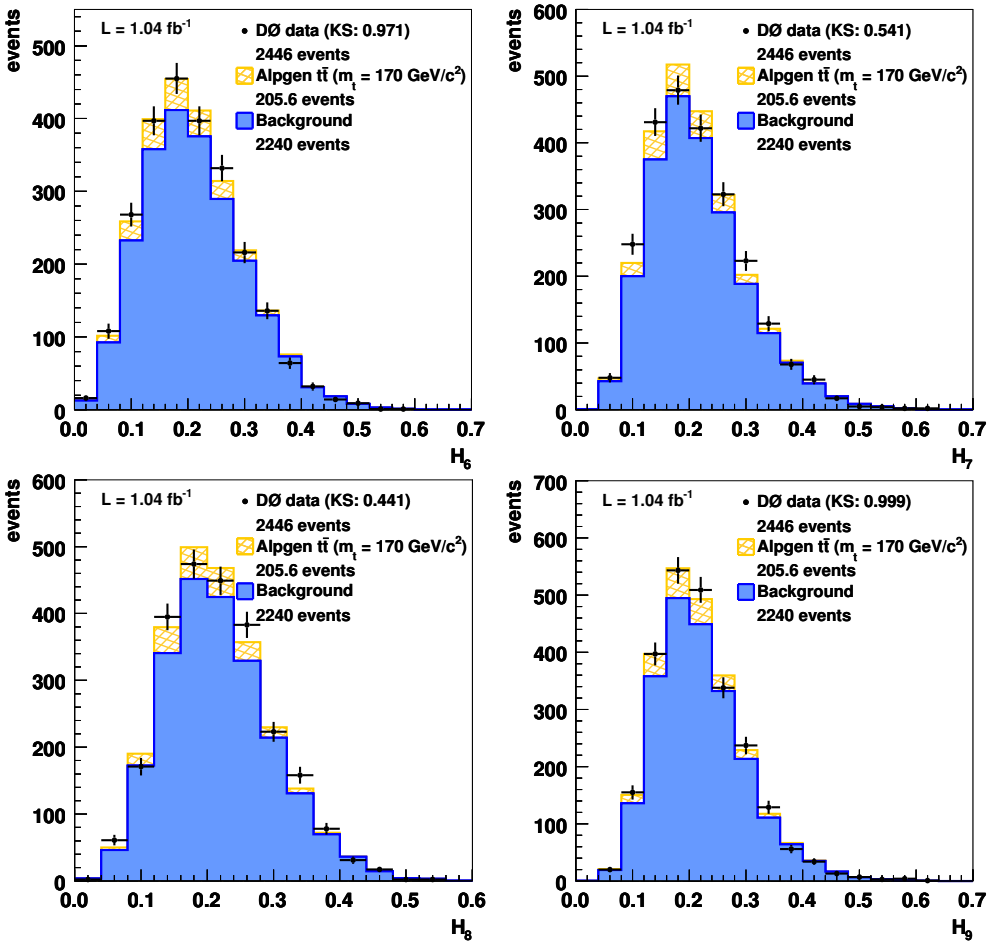


Figure 8.4: Cont.) Results from the likelihood fit compared to data. Shown are the background contribution (solid fill), the $t\bar{t}$ signal contribution (hashed) and the data (black markers).

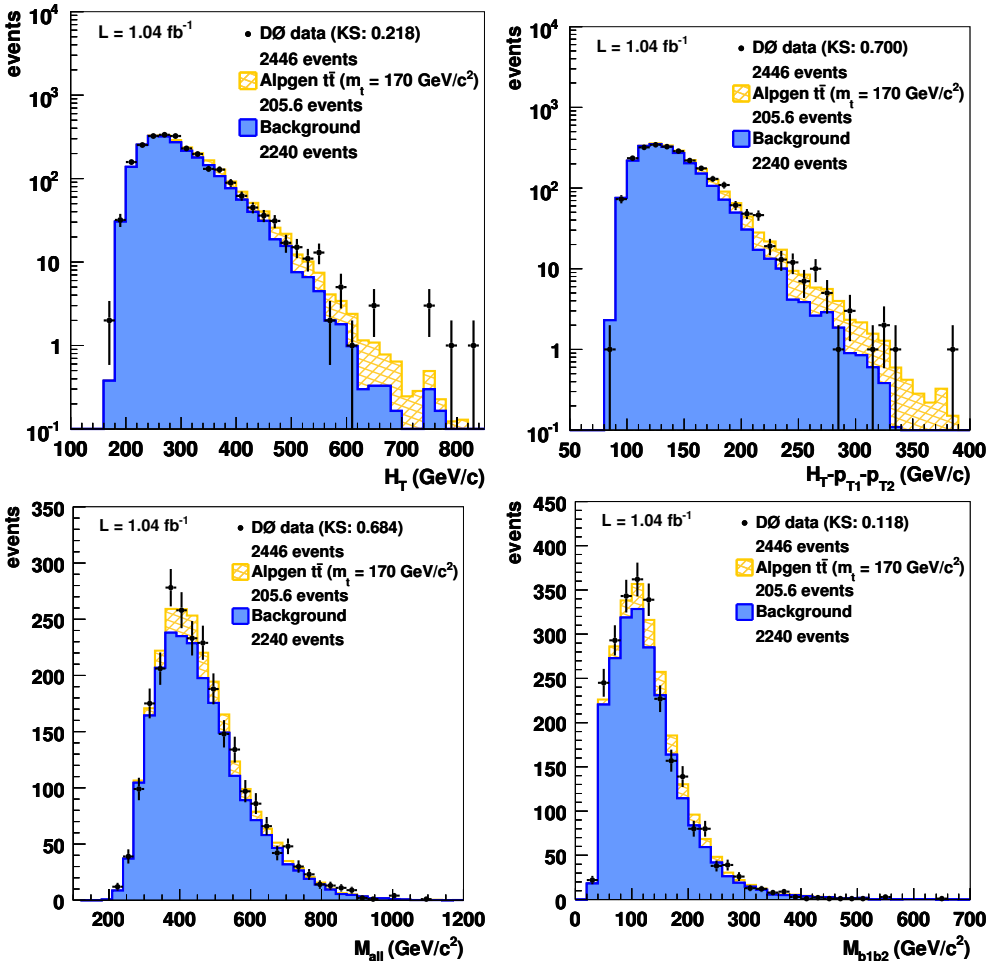


Figure 8.4: *(Cont.) Results from the likelihood fit compared to data. Shown are the background contribution (solid fill), the $t\bar{t}$ signal contribution (hashed) and the data (black markers).*

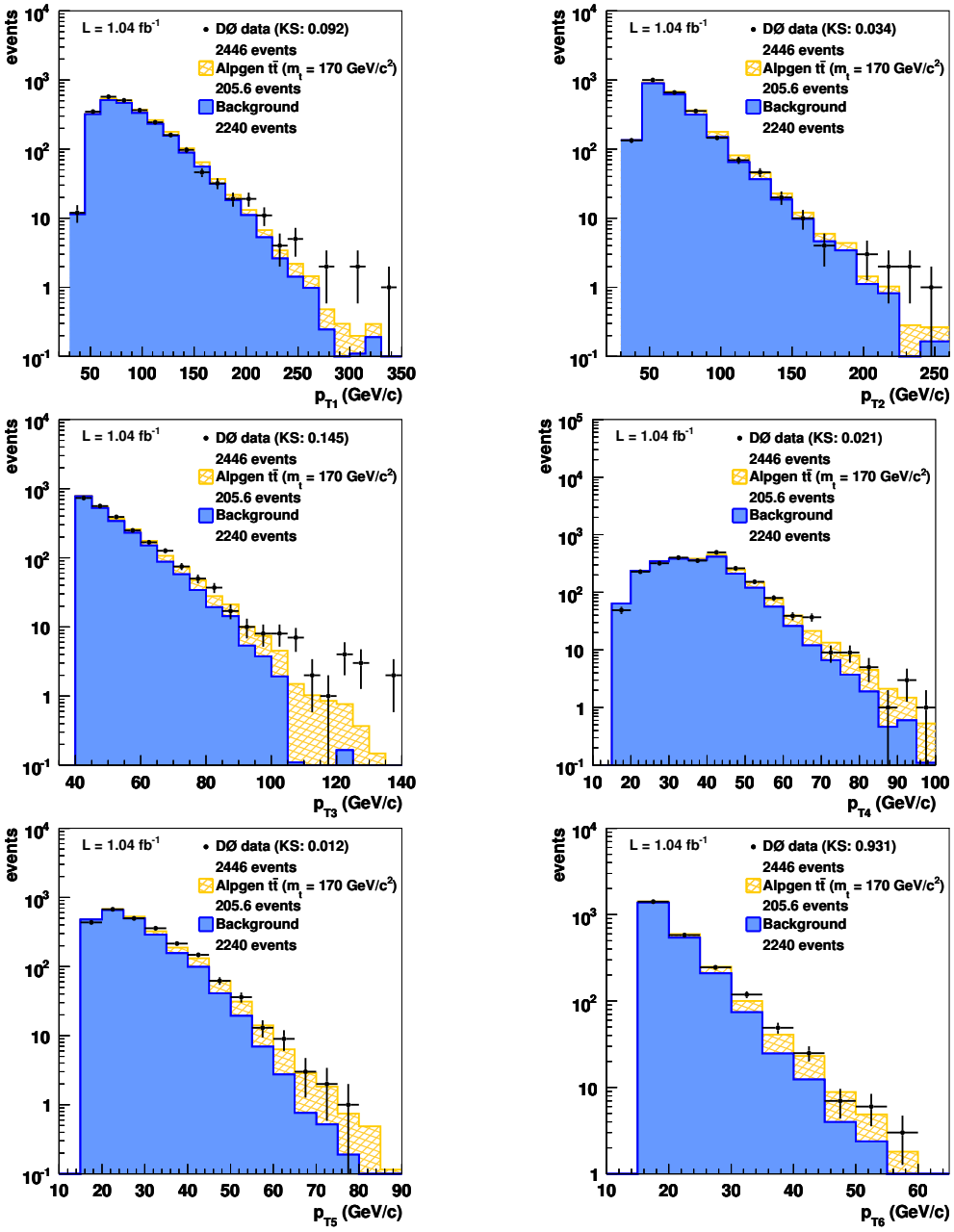


Figure 8.4: *(Cont.) Results from the likelihood fit compared to data. Shown are the background contribution (solid fill), the $t\bar{t}$ signal contribution (hashed) and the data (black markers).*

Chapter 9

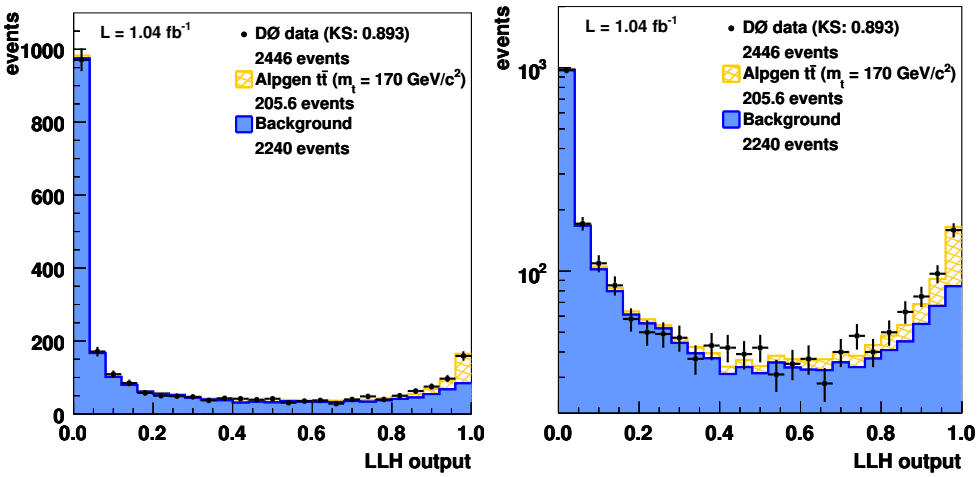
The measured cross section, systematic uncertainties and discussion

This chapter combines the results of the likelihood template fit described in chapter 8 with a discussion of relevant systematic uncertainties. In section 9.3.3, a retrospective overview is given of the full analysis, highlighting its major steps.

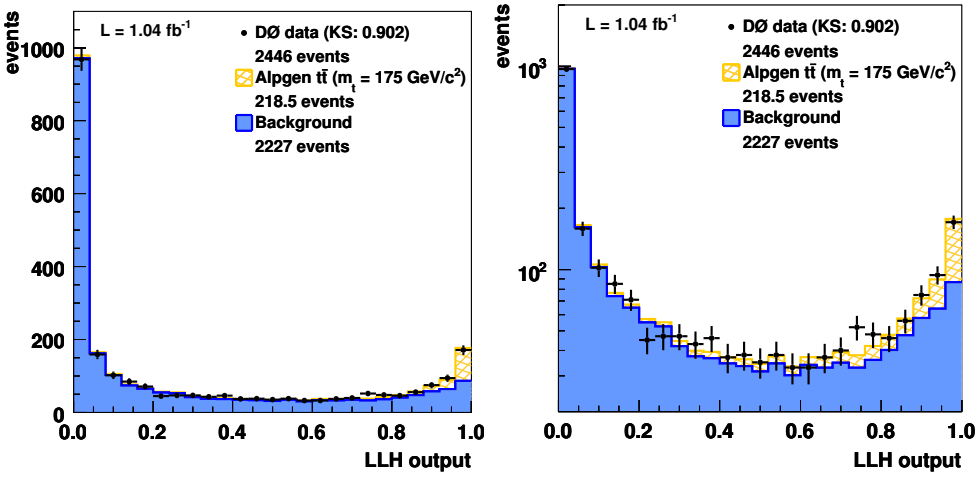
9.1 The measured cross section

The cross section measurement depends on the assumed top quark mass through the signal Monte Carlo used to derive the likelihood templates. Therefore, the cross section was estimated for two different top quark mass points, $m_{\mathrm{t}} = 170$ GeV/c^2 and $m_{\mathrm{t}} = 175$ GeV/c^2, and the results were interpolated to the current world average $m_{\mathrm{t}} = 172.4$ GeV/c^2 (see also figure 1.6). No significant mass dependence is present in the selection efficiencies. For inclusive (hadronic) $\mathrm{t\bar{t}}$ decays the selection efficiencies are 2.5% (5.0%) for $m_{\mathrm{t}} = 170$ GeV/c^2 and 2.7% (5.5%) for $m_{\mathrm{t}} = 175$ GeV/c^2. The corresponding non-hadronic contamination fractions are 7.9% and 6.2%. None of the likelihood input variables show any significant mass dependence. A comprehensive comparison can be found in ref. [109].

Figure 9.1 shows the direct results of fitting the signal and background likelihood templates to the data likelihood distributions. As was the case for the control variables in chapter 8, the fit results show good agreement with the data, as well as a good Kolmogorov-Smirnov probability (≈ 0.9). For $m_{\mathrm{t}} = 170$ GeV/c^2, the 205.6 signal events in a selected sample of 2446 events (8.4%), together with the 2.5% selection efficiency (section 5.3.1), the 97.3% (data quality, section 5.4.1) and 96.7% (vertex



(a)



(b)

Figure 9.1: *Likelihood output distribution for the selected data sample (black markers), together with the signal and background contributions (hashed orange and solid brown fill respectively) obtained from the fraction-fit described in chapter 8. (a): $m_t = 170\ GeV/c^2$, (b): $m_t = 175\ GeV/c^2$, left: linear scale, right: logarithmic scale.*

acceptance, section 5.4.2) global efficiency corrections, imply the presence of a total of $\approx$ 8700 signal events in the data sample before selection. Combined with the integrated luminosity of 1040.96 pb^{-1}, this leads to a production cross section of $\sigma_{t\bar{t}} = 8.6$ pb. For the $m_t = 175$ GeV/c^2 mass point there are 218.5 signal events (8.9%), the selection efficiency is 2.7%, and the global efficiencies are the same, giving an estimated cross section of 8.3 pb. Based on the number of signal events, the above approach assumes the presence of enough statistics that the event count can be described by a Gaussian (as opposed to a Poisson) distribution. A comparison with the full Poisson logarithmic likelihood ($-2 \ln Q_{\text{Poisson}}$) leads to the same cross section. Since the likelihood approach cannot take into account the effects of limited statistics on the template fit, the associated uncertainties (≈ 0.5 pb) are not used; the statistical uncertainty is estimated using ensemble tests (section 9.2.3).

9.2 Systematic uncertainties

The effects of systematic uncertainties and variations in input variables are studied using ensemble tests. All systematic uncertainties on the measured hadronic $t\bar{t}$ production cross section for both top quark mass points are summarised in table 9.1. In the following all sources of systematic uncertainties will be briefly discussed.

9.2.1 Signal model

Where possible, uncertainties related to aspects of the signal modelling in the Monte Carlo, e.g. b-fragmentation (section 9.2.6) and luminosity reweighting (section 9.2.7), are treated independently. An additional contribution to the MC signal modelling lies in the generation of additional radiative jets in the events. This uncertainty on the signal modelling was estimated by re-estimating the cross section using events with exactly six jets only. The difference between the two results, $\sigma_{t\bar{t}}$(six-or-more jets) = 8.6 pb (8.3 pb) and $\sigma_{t\bar{t}}$(six jets) = 7.7 pb (7.6 pb) for $m_t = 170$ GeV/c^2 (175 GeV/c^2), clearly overestimates any possible signal modelling uncertainty. The quoted uncertainty is obtained by symmetrising the difference around the measured cross section.

Source	Uncertainty (pb)			
	170 GeV/c^2		175 GeV/c^2	
Signal model	-0.44	$+0.44$	-0.37	$+0.37$
Background model	-1.90	$+1.92$	-1.73	$+1.75$
Signal statistics	-0.10	$+0.10$	-0.05	$+0.05$
Background statistics	-0.34	$+0.34$	-0.29	$+0.29$
Isolated lepton veto	-0.02	$+0.02$	-0.02	$+0.02$
PDF (CTEQ6.5M)	-0.22	$+0.18$	-0.23	$+0.19$
Jet ID	-0.03	$+0.01$	-0.03	$+0.03$
Jet energy scale	-0.53	$+0.53$	-0.41	$+0.41$
Jet energy resolution	-0.18	$+0.18$	-0.06	$+0.06$
b-Fragmentation	-0.03	$+0.01$	-0.01	$+0.01$
b-jet ID	-0.03	$+0.03$	-0.06	$+0.05$
b/light jet response	-0.01	$+0.01$	-0.01	$+0.01$
Trigger	-0.06	$+0.07$	-0.04	$+0.07$
Luminosity profile	-0.03	$+0.03$	-0.05	$+0.05$
Beam spot profile	-0.01	$+0.01$	-0.01	$+0.01$
Total	-2.07	$+2.09$	-1.86	$+1.87$

Table 9.1: *Systematic uncertainties on the $t\bar{t}$ cross section based on the* ALPGEN *signal Monte Carlo. The totals were obtained by adding the individual contributions in quadrature.*

9.2.2 Background model

As discussed in section 7.3, the systematic uncertainty on the background modelling contains two separate contributions:

- the differences with respect to the central value (based on the 50–50 'mixed' background sample) obtained by using the `4+2` and `5+1` background samples generated using different acceptor samples, and
- the effects of varying the $\not\!\!E_{\mathrm{T}}/H_{\mathrm{T}4}$ shaping cut.

The former contribution dominates, as demonstrated in table 9.2. Summed in quadrature these two uncertainties contribute $\Delta\sigma_{\mathrm{t\bar{t}}}(\text{background}) = ^{+1.92}_{-1.90}$ pb and $\Delta\sigma_{\mathrm{t\bar{t}}}(\text{background}) = ^{+1.75}_{-1.73}$ pb for $m_{\mathrm{t}} = 170$ GeV/c^2 and $m_{\mathrm{t}} = 175$ GeV/c^2 respectively.

Source	Uncertainty (pb)			
	170 GeV/c^2		175 GeV/c^2	
4+2-Only background		+1.89		+1.73
5+1-Only background	−1.77		−1.63	
'Down' background		+0.30		+0.25
'Up' background	−0.68		−0.58	
Total	−1.90	+1.92	−1.73	+1.75

Table 9.2: *Systematic uncertainties on the $t\bar{t}$ cross section due to the background modelling (based on the* ALPGEN *signal Monte Carlo). The two contributions were added in quadrature to form the total background modelling uncertainty.*

The background modelling dominates the uncertainty on the measured cross section. Judging from the relative sizes of the contributions due to the signal model (≈ 0.4 pb) and due to background modelling (≈ 1.8 pb), the data-based approach is capable of providing an adequate background model, but should benefit from a more thorough understanding of the differences between the different QCD multi-jet acceptor samples.

9.2.3 Signal and background statistics

The effects of finite statistics in the signal and background templates on the fit results were estimated by running ensembles of pseudo-experiments. The resulting root-mean-square values of the corresponding cross section distributions were assigned as symmetric uncertainties.

9.2.4 PDF uncertainties

After reweighting (see section 5.2.2), all Monte Carlo samples correspond to the central PDF of the CTEQ6.5M PDF set. This set also includes forty alternative PDFs representing up/down variations along the twenty eigenvectors of the parameter space. The modified tolerance approach [138] is used to estimate the effects of the individual PDF uncertainties on the measured cross section. This method properly estimates the uncertainties for physical observables X also in the case in which both up and down variations of a given PDF influence X in the same direction. Considering all N

(twenty for CTEQ6.5M) up and down variations the uncertainty on the observable X is given by:

$$\Delta X^{+}_{\max} = \sqrt{\sum_{i=1}^{N} \left[max\left(X_i^{+} - X_0, X_i^{-} - X_0, 0\right)\right]^2}$$

$$\Delta X^{-}_{\max} = \sqrt{\sum_{i=1}^{N} \left[max\left(X_0 - X_i^{+}, X_0 - X_i^{-}, 0\right)\right]^2}$$

where the index i runs over the different eigenvectors and X_i^{+} (X_i^{-}) represent the values of X using the up (down) variation of the i-th eigenvector.

9.2.5 Jet energy, resolution and ID

The standard uncertainties on the jet energy scale (described in section 4.7) are propagated through the analysis by shifting the JES corrections up and down by one sigma, resulting in $\Delta\sigma_{\mathrm{t\bar{t}}}(\mathrm{JES}) = {}^{+0.53}_{-0.53}$ pb and $\Delta\sigma_{\mathrm{t\bar{t}}}(\mathrm{JES}) = {}^{+0.41}_{-0.41}$ pb for $m_{\mathrm{t}} = 170$ GeV/c^2 and 175 GeV/c^2 respectively (symmetric by coincidence, not by construction).

Jet resolution effects are estimated by shifting all Monte Carlo jets (in p_{T}) by the relative data-to-MC jet energy scale [129]. This assumes that the full jet energy scale difference is due to ununderstood resolution and efficiency effects in the Monte Carlo.

The jet identification efficiency is almost 100% (see also section 3.3). Upward statistical fluctuations would exceed unity, so the systematics effects due to the jet ID efficiency use the statistical uncertainty on the efficiency for the downward variation, while for the upward variation the efficiency is set to 100%.

Of the above three points only the standard jet energy scale uncertainties apply to both data and Monte Carlo, the other two are MC-only effects.

9.2.6 b-Jet related uncertainties

The b-jet related uncertainties contain contributions from:

- b-fragmentation,
- overall jet taggability,

- b-jet identification, and
- b-jet energy calibration.

The b-fragmentation reweighting discussed in section 5.2.2 contains a single alternative parameterisation for use in systematic studies. This alternative reweighting is applied symmetrically by defining $w_{\text{down}} = w_{\text{central}} - (w_{\text{up}} - w_{\text{central}})$. The results are cross section uncertainties of $\Delta\sigma_{\text{t}\bar{\text{t}}}(\text{b-fragmentation}) = ^{+0.01}_{-0.03}$ pb for $m_{\text{t}} = 170$ GeV/c^2, and $\Delta\sigma_{\text{t}\bar{\text{t}}}(\text{b-fragmentation}) = ^{+0.01}_{-0.01}$ pb for $m_{\text{t}} = 175$ GeV/c^2.

As discussed in section 5.2.1, the taggability was varied up and down with 1.0% (absolute). The effects of the b-tagging efficiency were estimated by varying the TRFs (section 3.3.2, figure 3.4) in both data and Monte Carlo by their respective one-sigma uncertainties. The taggability and b-jet identification contributions are combined quadratically in table 9.1 as 'b-Jet ID'. The b-tagging contributions (TRF uncertainties and taggability) are the second largest systematic uncertainty on the measured cross section. The size of this effect is probably partly due to migration of b-jets into light-jets caused by the efficiency variations, changing the b-jet purity and undoubtedly affecting the background model as well.

The jet energy scale was determined on a sample dominated by light-jets. The systematic uncertainty due to differences in both kinematics and particle composition between b-, and light-jets, the relative b-to-light jet response is estimated by scaling the relative response for b-jets in Monte Carlo down to 0.982 and up to $1/0.982 = 1.018$, following the prescription in ref. [139].

9.2.7 Luminosity

The uncertainty on the total, integrated luminosity recorded by D0 during Run IIa is 6.1% [140], corresponding to $\Delta\sigma_{\text{t}\bar{\text{t}}} = 0.5$ pb. This contribution is not combined within the systematic uncertainty but quoted separately.

The systematic uncertainty on the reweighting of the overlaid zero-bias luminosity in Monte Carlo was estimated as the (symmetrised) full difference between performing the analysis using Monte Carlo samples with and without this luminosity reweighting. The result is a symmetric 0.05 pb uncertainty.

The difference between the luminosity profiles for the 3JT and 4JT triggers, as well as the difference between the 4JT profile (used in the reweighting) and the inclusive Run IIa profile are small. Considering the size of the effect of removing luminosity reweighting completely, these effects are negligible.

The contribution due to luminosity reweighting is marked '*luminosity profile*' in table 9.1.

9.2.8 Primary vertex position and reweighting

To estimate the effect of a possibly imperfect primary vertex distribution in the Monte Carlo ('Beam spot profile' in table 9.1) the analysis was repeated without applying the primary vertex position reweighting discussed in section 5.2.2. The full difference (0.01 pb) with the measured cross section was applied as a symmetric uncertainty.

9.2.9 Trigger-related uncertainty

All trigger efficiencies were varied within their one sigma uncertainties. The upward effect on the cross section is dominated by the Level 2 trigger uncertainty, ≈ 0.04 pb of the 0.07 pb, which in turn is dominated by the uncertainties on the Level 2 H_{T} terms. The downward effect is dominated by the uncertainties on the Level 3 trigger terms.

9.3 Discussion

Combining the statistical, systematic and luminosity uncertainties, the cross section results become:

$\boldsymbol{m_t = 170\ \mathrm{GeV}/c^2}$**:** $\sigma_{t\bar{t}} = 8.6\ \pm 1.3\ (\mathrm{stat})\ ^{+2.1}_{-2.1}\ (\mathrm{syst})\ \pm 0.5\ (\mathrm{lumi})$ pb,

$\boldsymbol{m_t = 175\ \mathrm{GeV}/c^2}$**:** $\sigma_{t\bar{t}} = 8.3\ \pm 1.1\ (\mathrm{stat})\ ^{+1.9}_{-1.9}\ (\mathrm{syst})\ \pm 0.5\ (\mathrm{lumi})$ pb.

Linear interpolation between the above measurements towards the world average top mass gives a cross section of:

$\boldsymbol{m_t = 172.4\ \mathrm{GeV}/c^2}$**:** $\sigma_{t\bar{t}} = 8.4\ \pm 1.2\ (\mathrm{stat})\ ^{+2.0}_{-2.0}\ (\mathrm{syst})\ \pm 0.5\ (\mathrm{lumi})$ pb,

based on an integrated luminosity of $\mathcal{L} = 1041\ \mathrm{pb}^{-1}$.

Previous D0 $t\bar{t}$ cross section measurements in the hadronic channel include:

$\sigma_{t\bar{t}} = 7.5\ ^{+3.1}_{-3.0}\ (\mathrm{stat})\ ^{+3.3}_{-2.0}\ (\mathrm{syst})\ \pm 0.5\ (\mathrm{lumi})$ pb, based on $\mathcal{L} = 163\ \mathrm{pb}^{-1}$ [141]

$\sigma_{t\bar{t}} = 4.5\ ^{+2.0}_{-1.9}\ (\mathrm{stat})\ ^{+2.4}_{-1.1}\ (\mathrm{syst})\ \pm 0.3\ (\mathrm{lumi})$ pb, based on $\mathcal{L} = 405\ \mathrm{pb}^{-1}$ [142]

(both of which assume $m_{\mathrm{t}} = 175$ GeV/c^2). The current result is well compatible with both. As is to be expected the statistical uncertainty slowly decreases with the larger available data sets. Luminosity uncertainties are direct fractions of the measured cross section, independent of decay channel and analysis techniques. Compared to the latter of the above results, the current systematic uncertainties have not changed significantly. The dominant contribution, however, is no longer the jet energy scale, but the new, data-driven background model. With a more detailed understanding of especially the phase space matching procedure in the background modelling the systematics should be reduced.

9.3.1 Tightening the selection

As described in section 5.3, this analysis was originally performed with an event selection requiring at least six jets with $p_{\mathrm{T}} > 15$ GeV/c and within $|\eta_{\mathrm{d}}| < 2.5$. At least three of these jets, among which the b-jets, are required to satisfy $p_{\mathrm{T}} > 40$ GeV/c (the so-called '3+40' sample). For cross-checks two alternative selections were used:

'2+40': loosening the p_{T} cut on the third jet to $p_{\mathrm{T}} > 15$ GeV/c, and

'4+40': tightening the p_{T} cut on the fourth jet to $p_{\mathrm{T}} > 40$ GeV/c.

For all three versions the signal selection efficiency is very high; the selected samples differ mainly in background content. Reducing the background suppresses the effects of imperfections of the background modelling on the measured cross section. Considering that the background-related uncertainties dominate the systematics (table 9.1), this effect outweighs the reduction in available statistics when switching from '3+40' to '4+40'.

Redoing the analysis after tightening the p_{T} cut on the fourth jet, the overall selection efficiencies change as shown in table 9.3.

Based on the '4+40' samples, the measured cross sections become:

$\boldsymbol{m_{\mathrm{t}} = 170}$ **GeV/**$\boldsymbol{c^2}$**:** $\sigma_{\mathrm{t\bar{t}}} = 7.9 \pm 1.3\ (\mathrm{stat})\ ^{+0.9}_{-0.9}\ (\mathrm{syst}) \pm 0.5\ (\mathrm{lumi})$ pb,

$\boldsymbol{m_{\mathrm{t}} = 172.4}$ **GeV/**$\boldsymbol{c^2}$**:** $\sigma_{\mathrm{t\bar{t}}} = 7.5 \pm 1.3\ (\mathrm{stat})\ ^{+0.8}_{-0.9}\ (\mathrm{syst}) \pm 0.5\ (\mathrm{lumi})$ pb,

$\boldsymbol{m_{\mathrm{t}} = 175}$ **GeV/**$\boldsymbol{c^2}$**:** $\sigma_{\mathrm{t\bar{t}}} = 7.1 \pm 1.2\ (\mathrm{stat})\ ^{+0.7}_{-0.8}\ (\mathrm{syst}) \pm 0.5\ (\mathrm{lumi})$ pb.

The observed cross section values obtained from the two selected samples are in good agreement. The statistical uncertainties remain the same, but the systematic

m_t (GeV/c^2)	170		175	
sample	'3+40'	'4+40'	'3+40'	'4+40'
Efficiency for inclusive $t\bar{t}$ (%)	2.5	2.0	2.7	2.2
Efficiency for hadronic $t\bar{t}$ (%)	5.0	4.1	5.5	4.5
Non-hadronic signal contamination (%)	7.9	5.7	6.2	5.9
Incremental signal efficiency (%)	80.6		81.2	
Incremental background efficiency (%)	37.6			

Table 9.3: *Summary of differences between the '3+40' and '4+40' samples. Please note that the 'non-hadronic contamination' is specified as a percentage of the final signal sample after selection. The 'incremental' efficiencies shown represent the '3+40'-to-'4+40' sub-selection.*

uncertainties are significantly reduced in the '4+40' results with respect to the original, '3+40', ones. This improvement is almost exclusively due to decreased background modelling uncertainties: $\Delta\sigma_{t\bar{t}}(\text{background}) = ^{+0.59}_{-0.70}$ pb for $m_t = 170$ GeV/c^2 and $\Delta\sigma_{t\bar{t}}(\text{background}) = ^{+0.36}_{-0.52}$ pb for $m_t = 175$ GeV/c^2 [109].

The above cross section value corresponding to $m_t = 172.4$ GeV/c^2 is adopted as *the* result of this analysis.

9.3.2 Comparison to the standard model prediction

Figure 9.2 shows the standard model prediction (presented in figure 1.10 on page 21) together with the measured cross section from this analysis. Each of the cross section measurements is in good agreement with the standard model prediction. The central result, $\sigma_{t\bar{t}} = 7.5 \pm 1.3\ (\text{stat})\ ^{+0.8}_{-0.9}\ (\text{syst}) \pm 0.5\ (\text{lumi})$ pb at $m_t = 172.4$ GeV/c^2 itself is in perfect agreement with the central theory curve.

9.3.3 Retrospective

The results presented in this thesis show that even in fully hadronic $t\bar{t}$ decays precision measurements are within reach. More importantly, this has been achieved without relying on Monte Carlo event generators to model the QCD background. However, to reach the precision required to allow more meaningful comparisons with the theoretical predictions, more work is needed.

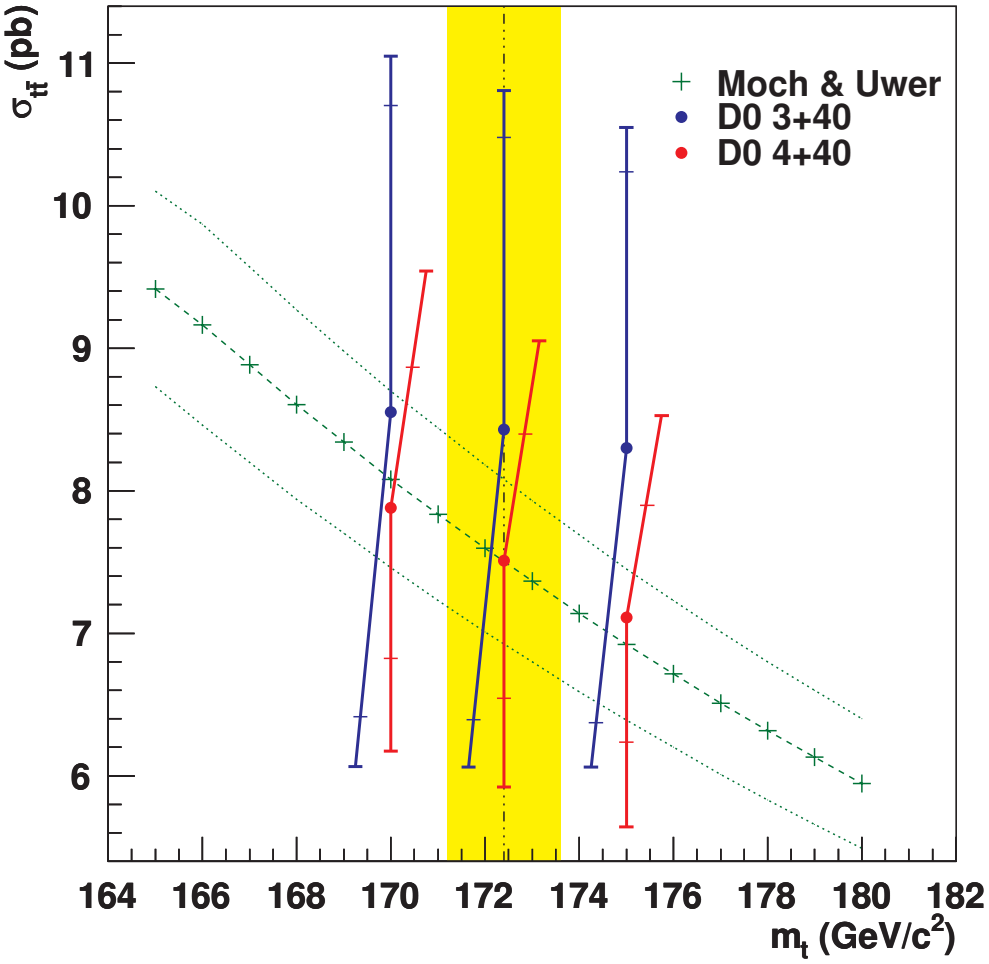


Figure 9.2: *The $t\bar{t}$ production cross section as measured in this analysis (• markers), together with the standard model predictions depicted in figure 1.10. Shown with the measured cross sections are the corresponding systematic uncertainties (including luminosity contributions, 'inner' error bars) and the total uncertainties ('outer' error bars). The vertical, shaded band represents the current world average top mass, $m_t = 172.4 \pm 1.2$ GeV/c^2, from figure 1.6.*

The addition of new data — approximately 4 fb^{-1} has currently been recorded by the D0 collaboration — will of course help. It is unlikely, however, that any analysis in the hadronic $\text{t}\bar{\text{t}}$ channel will ever be statistics-limited. A deeper understanding of both signal and background modelling is required. Uncertainties related to the signal model (as well as those related to b-tagging efficiencies) can probably most easily be reduced using a likelihood combination of measurements obtained in several different exclusive (b-)jet multiplicity bins. A more detailed understanding of the background model, e.g. the differences between the `4+2` and `5+1` background samples, will benefit from taking one step back, to MC generator level. Even though these generators may not be able to fully describe all QCD multi-jet characteristics in detail, they can help in understanding the (side-)effects of the intended phase space matching.

It is conceivable that the reconstruction of multi-jet environments, especially in the presence of multiple interactions, should benefit from jet algorithms that take into account the real geometric shape of jets on a per-jet basis, such as the k_{T} algorithm. On the other hand, such algorithms pose new challenges for calibration, and maintaining multiple algorithms may prove a significant burden for collaborations.

The disappointing efficiency of the Level 3 b-tagging terms, together with the difficulties involved in determining their efficiencies, means that unless these issues are solved at the trigger level, the 3JT triggers will not contribute significantly to hadronic $\text{t}\bar{\text{t}}$ data selection. With increasing data samples the small gains in efficiency and luminosity no longer outweigh the work required to combine the 3JT and 4JT trigger sets.

The huge progress in the D0 jet energy scale has shown that what at one point in time may seem to be insurmountable sources of uncertainty can be reduced to levels below statistical fluctuations.

Samenvatting

In 1869 wist Dmitri Mendeleev voor het eerst de natuurlijke elementen te rangschikken in het periodiek systeem. In de loop der tijd werden meer elementen ontdekt, diverse reeds nauwkeurig voorspeld door Mendeleev op grond van open plekken in zijn tabellen, en vonden hun plaats in het periodiek systeem. Tegelijkertijd ging men de wetmatigheid achter deze indeling steeds beter begrijpen.

De twintigste eeuw bracht de geboorte van de quantum mechanica. Na een tumultueuze start in de eerste jaren kwam het langzaam tot algemene acceptatie. Nog steeds kon (en kan) men discussieren over de grondbeginselen, maar verificatie van de voorspellingen in vele precisie experimenten maakte het tot een uiterst succesvolle theorie. Voor het eerst was het mogelijk het waterstof atoom in detail theoretisch te beschrijven. Quantum mechanica leidde Paul Dirac ertoe in 1928 het bestaan van antimaterie te postuleren. Pas vier jaar later zou het bestaan van het antipartner van het electron, het positron, door Carl Anderson experimenteel worden bevestigd (in kosmische straling).

Experimenten toonden het bestaan aan van steeds meer en andere deeltjes: pionen (π, in kosmische straling), lambdas (Λ, ook in kosmische straling), cascades (Ξ, in een bellenvat experiment op Brookhaven National Laboratory), enzovoort. Geïnspireerd door Mendeleev's periodiek systeem formuleerden Gell-Mann en Ne'eman in 1964 een indeling voor alle tot dan toe gevonden deeltjes, inclusief de protonen en neutronen die de kernen van alle atomen om ons heen vormen. Net als Mendeleev hielden zij nog open plaatsen over in hun systeem. Hiermee was voldaan aan één van de belangrijkste eisen voor een wetenschappelijke theorie: ze voorspelde het bestaan van (nog onontdekte) deeltjes om de gaten op te vullen. De classificatie van Gell-Mann en Ne'eman veronderstelde een onderliggende drievoudige symmetrie; de protonen en neutronen waren niet langer de kleinst mogelijke bouwstenen, er moesten minstens een drietal kleinere stenen zijn die de protonen en neutronen zelf vormden: de quarks. Tegen 1969 was het duidelijk dat hadronen, deeltjes zoals protonen en neutronen,

substructuur hebben en dat de bouwstenen hoogstwaarschijnlijk de gepostuleerde quarks waren.

Het fundament van de moderne deeltjesfysica is het *standaard model*. Dit model beschrijft de interacties tussen subatomaire deeltjes als quantum velden. Het bevat de quarks nodig voor de drievoudige symmetrie van Gell-Mann en Ne'eman maar ook een aantal zwaardere quarks zoals *bottom* (b) en *top* (t). Het massiefste quark, het top quark, heeft doordat het zo zwaar is een aantal unieke eigenschappen. Door de grote massa vervallen top quarks bijna altijd voordat gebonden toestanden gevormd kunnen worden; top quarks ondervinden bijna geen effecten van 'quark confinement' en benaderen 'vrije' quarks.

De speciale positie van het top quark in het standaard model maakt het nog belangrijker om nauwkeurig te verifiëren dat ook hier het standaard model de juiste voorspellingen doet. Uitbreidingen op het standaard model voorspellen het bestaan van een tot op heden ongezien deeltje, het Higgs boson. Door interacties met dit Higgs boson verkrijgen deeltjes massa's. Ook hier, alweer vanwege de grote massa, speelt het top quark een belangrijke rol. Quantum correcties relateren de massa's van onder andere het W boson en het top quark met de Higgs boson massa. Precisie metingen van de W en top massa geven indirecte informatie over het standaard model Higgs boson (zie ook figuur 1.3).

De Fermilab Tevatron versneller, gelegen nabij Chicago in de Verenigde Staten, is de enige versneller ter wereld waar momenteel top quarks geproduceerd worden. In het Tevatron worden protonen (waterstofkernen) en hun antideeltjes versneld en op elkaar gebotst. In deze botsingen wordt de kinetische energie van de (anti)protonen omgezet in massa van de botsings-producten. De meeste van deze deeltjes vervallen onmiddellijk weer maar door het bestuderen van de vervalsproducten is het mogelijk het botsings-proces te reconstrueren.

De D0 detector (zie figuur 2.4 op pagina 37) is één van de detectoren in de Tevatron versneller-ring. D0, speciaal gebouwd voor het bestuderen van proton-antiproton botsingen met hoge energie, bevat verschillende subdetectoren voor het volgen van electrisch geladen deeltjes en voor het reconstrueren van jets, gecollimeerde bundels van deeltjes afkomstig van 'quark fragmentatie'.

Top quarks worden voornamelijk geproduceerd in top-antitop paren. Standaard model top quarks vervallen nagenoeg altijd ($> 99.8\%$) in een bottom quark en een W boson ($\mathrm{t} \to \mathrm{bW}^+$ en $\bar{\mathrm{t}} \to \bar{\mathrm{b}}\mathrm{W}^-$). Afhankelijk van de vervalsproducten van de W bosons worden $\mathrm{t}\bar{\mathrm{t}}$ paar vervallen geclassificeerd als leptonisch (beide Ws vervallen naar

e, μ of τ), semileptonisch (één W vervalt naar e, μ of τ, de andere naar een quark-antiquark paar) of hadronisch (beide Ws vervallen naar een $q\bar{q}$ paar). Het laatste geval, hadronisch verval, vormt de grootste bijdrage: ongeveer 46%. In tegenstelling tot de (semi)leptonische vervallen bevat een hadronisch $t\bar{t}$ verval alleen jets, geen geïsoleerde leptonen (van W vervallen). Enerzijds maakt dit het uiterst moeilijk signaal van achtergrond te scheiden, anderzijds speelt slechts één enkel achtergrondsproces een rol: (non-$t\bar{t}$) QCD multi-jet productie. De werkzame doorsnede voor QCD multi-jet achtergrond ligt enkele ordes van grootte boven die voor het hadronische $t\bar{t}$ signaal. Eén van de cruciale aspecten van een analyse in het hadronische $t\bar{t}$ kanaal is het correct modelleren en begrijpen van deze achtergrond. Verder maakt de hoge jet multipliciteit multi-jet analyses zeer gevoelig voor de opeenstapeling van jet efficiency en calibratie effecten. De D0 jet energie calibratie heeft inmiddels een nauwkeurigheid bereikt van $\approx$ 2–3% over een groot gebied van jet momenta en pseudorapiditeit (zie voor een overzicht figuren 4.19 en 4.20 op pagina's 99–100). De nauwkeurigheid van de jet energie calibratie is niet langer de dominante factor in de onzekerheid voor veel analyses.

Dit proefschrift beschrijft een meting van de werkzame doorsnede voor top quark paar productie in het hadronische vervalskanaal. De data voor deze analyse werd verzameld tussen juli 2002 en februari 2006 met de D0 detector.

Op basis van de signatuur van een hadronisch $t\bar{t}$ paar verval zijn events (botsingen) met zes of meer jets geselecteerd. Minstens twee van de jets dienen gemerkt te zijn als jets afkomstig van b-quarks met grote transversale impuls ($p_{\mathrm{T}} > 40$ GeV/c). Voor het identificeren van b-jets is gebruik gemaakt van de nieuwste D0 'b-tagger' op basis van een neuraal netwerk.

De complexiteit van multi-jet QCD events komt direct terug in Monte Carlo simulatie. Harde jets (hoge p_{T}) worden goed beschreven door het matrix element voor de harde interactie, zachte jets door zogenaamde 'parton shower' evolutie. De faseruimte gebieden 'tussen' deze aanpakken in, zijn echter uiterst moeilijk correct te modelleren en vertrouwen op een 'matching' procedure. Het $t\bar{t}$ signaal is gemodelleerd met behulp van de ALPGEN en PYTHIA Monte Carlo event generatoren. ALPGEN is een multi-parton event generator op basis van 'boom-niveau' matrix elementen en verzorgt de harde interactie. PYTHIA's parton shower implementatie voegt zachte, radiatieve jets toe en simuleert fragmentatie, hadronisatie en verval van instabiele deeltjes. Detector simulatie en event reconstructie vindt plaats met behulp van de standaard D0 software.

Om onafhankelijk te zijn van mogelijke imperfecties van QCD multi-jet simulatie, is de achtergrond voor deze analyse volledig gebaseerd op collider data. Het achtergrond model maakt gebruik van een tweetal extra data samples, geselecteerd volgens dezelfde criteria als het zes-jet analyse sample, maar met slechts vier of vijf jets. De QCD productie daalt met de inverse van de sterke koppelingsconstante α_s, en de signaal bijdrage in het zes-jet sample bedraagt slechts $\approx 14\%$; deze vier-jet en vijf-jet samples worden volledig gedomineerd door QCD achtergrond events. Willekeurige vier-jet en vijf-jet events worden gepaard met 'echte' zes-jet events met vergelijkbare faseruimte. Door de 'extra' jets uit het zes-jet event over te nemen in de vier- of vijf-jet events, is het mogelijk om zes-jet events te creëren, gebaseerd op echte QCD events. Extra aandacht is besteedt aan validatie van de gegenereerde achtergrond events, op basis van een vergelijking tussen het vijf-jet data sample en een sample van vijf-jet events, gegenereerd door het combineren van vier- en vijf-jet events volgens bovenstaande procedure.

De fracties signaal en achtergrond in data zijn bepaald met behulp van een 'likelihood ratio' methode. De distributies voor een aantal karakteristieke variabelen zijn bepaald voor beide en gecombineerd in een likelihood functie. Een fit aan de data van een lineaire combinatie van de karakteristieke verdelingen van deze likelihood functie voor signaal en achtergrond geeft de meest waarschijnlijke relatieve bijdragen.

De systematische onzekerheden op de werkzame doorsnede worden gedomineerd door de signaal- en achtergrond model onzekerheden en door de onzekerheid op de efficiëntie van de b-jet identificatie (zie ook tabel 9.1).

De gemeten werkzame doorsnede is 7.5 ± 1.3 (stat) $^{+0.8}_{-0.9}$ (syst) ± 0.5 (lumi) pb (onder aanname van een top quark massa van $m_t = 172.4$ GeV/c^2, het huidige wereldgemiddelde van alle metingen), in uitstekende overeenstemming met de theoretische voorspelling op basis van het standaard model: 7.50 ± 0.58 pb (figuur 9.2).

Bibliography

References marked as 'D0 note' refer to internal notes only available within the D0 collaboration.

[1] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory* (Addison-Wesley Advanced Book Program, 1995).

[2] F. Halzen and A. D. Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics* (Wiley, 2001).

[3] ALEPH, Precision electroweak measurements on the Z resonance, Phys. Rept. **427**, 257 (2006).

[4] P. W. Higgs, Broken symmetries and the masses of gauge bosons, Phys. Rev. Lett. **13**, 508 (1964).

[5] P. W. Higgs, Broken symmetries, massless particles and gauge fields, Phys. Lett. **12**, 132 (1964).

[6] P. W. Higgs, Spontaneous symmetry breakdown without massless bosons, Phys. Rev. **145**, 1156 (1966).

[7] F. Englert and R. Brout, Broken symmetry and the mass of gauge vector mesons, Phys. Rev. Lett. **13**, 321 (1964).

[8] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, Global conservation laws and massless particles, Phys. Rev. Lett. **13**, 585 (1964).

[9] Super-Kamiokande, Y. Ashie *et al.*, Evidence for an oscillatory signature in atmospheric neutrino oscillation, Phys. Rev. Lett. **93**, 101801 (2004), `arXiv:hep-ex/0404034`.

[10] Super-Kamiokande, Y. Ashie *et al.*, A measurement of atmospheric neutrino oscillation parameters by Super-Kamiokande I, Phys. Rev. **D71**, 112005 (2005), `arXiv:hep-ex/0501064`.

[11] W.-M. Yao *et al.*, Review of Particle Physics, Journal of Physics G **33**, 1+ (2006).

[12] LEP Working Group for Higgs boson searches, R. Barate *et al.*, Search for the standard model Higgs boson at LEP, Phys. Lett. **B565**, 61 (2003).

[13] LEP Electroweak Working Group, The LEP electroweak working group web page, 2008, cited July 2008, `http://lepewwg.web.cern.ch/LEPEWWG/`.

[14] H. Burkhardt and B. Pietrzyk, Low energy hadronic contribution to the QED vacuum polarization, Phys. Rev. **D72**, 057501 (2005).

[15] BES experiments home page, `http://bes.ihep.ac.cn/`.

[16] CMD-2 experiment home page, `http://wwwcmd2.inp.nsk.su/`.

[17] KLOE experiment home page, `http://www.lnf.infn.it/kloe/Welcome.html`.

[18] J. F. de Troconiz and F. J. Yndurain, The hadronic contributions to the anomalous magnetic moment of the muon, Phys. Rev. **D71**, 073008 (2005).

[19] N. Cabibbo, Unitary symmetry and leptonic decays, Phys. Rev. Lett. **10**, 531 (1963).

[20] M. Kobayashi and T. Maskawa, CP violation in the renormalizable theory of weak interaction, Prog. Theor. Phys. **49**, 652 (1973).

[21] A. Quadt, Top quark physics at hadron colliders, Eur. Phys. J. **C48**, 835 (2006).

[22] M. Jezabek and J. H. Kuhn, QCD corrections to semileptonic decays of heavy quarks, Nucl. Phys. **B314**, 1 (1989).

[23] C. A. Nelson, B. T. Kress, M. Lopes, and T. P. McCauley, General tests for t –> W+ b couplings at hadron colliders, Phys. Rev. **D56**, 5928 (1997), `arXiv:hep-ph/9707211`.

[24] R. H. Dalitz and G. R. Goldstein, The Decay and polarization properties of the top quark, Phys. Rev. **D45**, 1531 (1992).

[25] G. L. Kane, G. A. Ladinsky, and C. P. Yuan, Using the top quark for testing standard model polarization and CP predictions, Phys. Rev. **D45**, 124 (1992).

[26] J. A. Aguilar-Saavedra, J. Carvalho, N. Castro, F. Veloso, and A. Onofre, Probing anomalous W t b couplings in top pair decays, Eur. Phys. J. **C50**, 519 (2007), `arXiv:hep-ph/0605190`.

[27] DELPHI, P. Abreu *et al.*, Measurement of the mass of the W boson using direct reconstruction at s**(1/2) = 183-GeV, Phys. Lett. **B462**, 410 (1999).

[28] CDF - Run II, A. Abulencia *et al.*, Measurement of the Top Quark Mass in $p\bar{p}$Collisions at $\sqrt{s} = 1.96$ TeV using the Decay Length Technique, Phys. Rev. **D75**, 071102 (2007), `arXiv:hep-ex/0612061`.

[29] Tevatron Electroweak Working Group, The Tevatron electroweak working group web page, cited July 2008, `http://tevewwg.fnal.gov/`.

[30] Tevatron Electroweak Working Group, and others, Combination of CDF and D0 Results on the Mass of the Top Quark, (2008), `arXiv:0808.1089 [hep-ex]`.

[31] J. C. Collins, D. E. Soper, and G. Sterman, Heavy Particle Production in High-Energy Hadron Collisions, Nucl. Phys. **B263**, 37 (1986).

[32] The CTEQ collaboration, The CTEQ collaboration home page, `http://www.phys.psu.edu/~cteq/`.

[33] W. K. Tung *et al.*, Heavy quark mass effects in deep inelastic scattering and global QCD analysis, JHEP **02**, 053 (2007).

[34] R. S. Thorne, A. D. Martin, W. J. Stirling, and R. G. Roberts, Update of MRST parton distributions, Acta Phys. Polon. **B33**, 2927 (2002), `arXiv:hep-ph/0207067`.

[35] G. 't Hooft, Renormalization of Massless Yang-Mills Fields, Nucl. Phys. **B33**, 173 (1971).

[36] G. t Hooft, RENORMALIZABLE LAGRANGIANS FOR MASSIVE YANG-MILLS FIELDS, Nucl. Phys. **B35**, 167 (1971).

[37] H. D. Politzer, RELIABLE PERTURBATIVE RESULTS FOR STRONG INTERACTIONS?, Phys. Rev. Lett. **30**, 1346 (1973).

[38] D. J. Gross and F. Wilczek, ULTRAVIOLET BEHAVIOR OF NON-ABELIAN GAUGE THEORIES, Phys. Rev. Lett. **30**, 1343 (1973).

[39] G. 't Hooft and M. J. G. Veltman, Regularization and Renormalization of Gauge Fields, Nucl. Phys. **B44**, 189 (1972).

[40] M. Cacciari, S. Frixione, M. L. Mangano, P. Nason, and G. Ridolfi, The t anti-t cross-section at 1.8-TeV and 1.96-TeV: A study of the systematics due to parton densities and scale dependence, JHEP **04**, 068 (2004).

[41] S. Moch and P. Uwer, Theoretical status and prospects for top-quark pair production at hadron colliders, Phys. Rev. **D78**, 034003 (2008), `arXiv:0804.1476 [hep-ph]`.

[42] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning* (Springer, 2003).

[43] B. P. Roe *et al.*, Boosted decision trees, an alternative to artificial neural networks, Nucl. Instrum. Meth. **A543**, 577 (2005), `arXiv:physics/0408124`.

[44] E. T. Jaynes and G. L. Bretthorst, *Probability Theory: The Logic of Science* (Cambridge University Press, 2003).

[45] D0, V. M. Abazov *et al.*, Evidence for production of single top quarks, (2008), `arXiv:0803.0739 [hep-ex]`.

[46] D0, V. M. Abazov *et al.*, Evidence for production of single top quarks and first direct measurement of $|V_{\rm tb}|$, Phys. Rev. Lett. **98**, 181802 (2007).

[47] B. W. Harris, E. Laenen, L. Phaf, Z. Sullivan, and S. Weinzierl, The fully differential single top quark cross section in next-to-leading order QCD, Phys. Rev. **D66**, 054024 (2002).

[48] Z. Sullivan, Understanding single-top-quark production and jets at hadron colliders, Phys. Rev. **D70**, 114012 (2004).

[49] R. J. Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*, The Manchester Physics Series, Reprint edition ed. (John Wiley & Sons, 1989).

[50] The CDF collaboration, `http://www-cdf.fnal.gov/`.

[51] The D0 collaboration, `http://www-d0.fnal.gov/`.

[52] CDF, F. Abe *et al.*, Observation of top quark production in $\bar{p}p$ collisions, Phys. Rev. Lett. **74**, 2626 (1995), `arXiv:hep-ex/9503002`.

[53] D0, S. Abachi *et al.*, Observation of the top quark, Phys. Rev. Lett. **74**, 2632 (1995), `arXiv:hep-ex/9503003`.

[54] S. W. Herb *et al.*, Observation of a dimuon resonance at 9.5-GeV in 400-GeV proton - nucleus collisions, Phys. Rev. Lett. **39**, 252 (1977).

[55] D0 Layer 0 silicon group, D0 Layer 0 Conceptual Design Report, D0 note 4415, 2004.

[56] M. Abolins *et al.*, Design and Implementation of the New D0 Level-1 Calorimeter Trigger, Nucl. Instrum. Meth. **A584**, 75 (2008), `arXiv:0709.3750 [physics.ins-det]`.

[57] P. Schmueser, Superconducting magnets for particle accelerators, Rept. Prog. Phys. **54**, 683 (1991).

[58] E. McCrory, V. Shiltsev, A. J. Slaughter, and A. Xiao, Fitting the luminosity decay in the Tevatron, in *Proceedings of the Particle Accelerator Conference (PAC 05)*, 2005, Talk given at the Particle Accelerator Conference (PAC 05), Knoxville, Tennessee, 16-20 May 2005.

[59] P. Bagley, Beam beam tune shifts for 36 bunch operation in the Tevatron, in *Proceedings of 5th European Particle Accelerator Conference (EPAC 96)*, pp. 1155–1157, 1996, Talk given at the 5th European Particle Accelerator Conference (EPAC 96), Sitges, Spain, 10-14 Jun 1996.

[60] Fermilab accelerator division luminosity summaries, `http://www-bdnew.fnal.gov/operations/lum/lum.html`.

[61] Department of Energy Review Committee, US Department of Energy, 2004, Department of Energy Review Committee Report on the Technical, Cost, Schedule, and Management Review of the Tevatron Run II Luminosity Upgrades.

[62] K. Grupen and B. Shwartz, *Particle Detectors*, Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology No. 26 (Cambridge University Press, 1996).

[63] D. Green, *The Physics of Particle Detectors*, Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology No. 12 (Cambridge University Press, 2005).

[64] R. Wigmans, *Calorimetry: Energy measurement in particle physics*, International series of monographs on physics No. 107 (Oxford science publications, 2000), Oxford, UK: Clarendon (2000) 726 p.

[65] D0, V. M. Abazov *et al.*, The upgraded D0 detector, Nucl. Instrum. Meth. **A565**, 463 (2006).

[66] D0, S. Abachi *et al.*, The D0 detector, Nucl. Instrum. Meth. **A338**, 185 (1994).

[67] D0 Upgrade Collaboration, D0 silicon tracker technical design report, D0 note 2169, 1994.

[68] S. Klimenko, J. Konigsberg, and T. M. Liss, Averaging of the inelastic cross sections measured by the CDF and the E811 experiments, FERMILAB-FN-0741.

[69] A. Khanov, HTF: histogramming method for finding tracks. the algorithm description, D0 note 3778, 2000.

[70] G. Borissov, Ordering a chaos or . . . technical details of AA tracking, Presentation in the 'all D0 meeting', 2003.

[71] H. Greenlee, The D0 Kalman track fit, D0 note 4303, 2003.

[72] A. Schwartzman and C. Tully, Primary vertex reconstruction by means of adaptive vertex fitting, D0 note 4918, 2005.

[73] A. Schwartzman, Y. Peters, and H. Greenlee, Adaptive primary vertex certification in p17, D0 note 5192, 2006.

[74] A. Schwartzman and M. Narain, Secondary vertex reconstruction using the Kalman filter, D0 note 3908, 2001.

[75] A. Schwartzman and M. Narian, Vertex fitting by means of the Kalman filter techinque, D0 note 3907, 2001.

[76] P. Calfayan *et al.*, Muon identification certification for p17 data, D0 note 5157, 2006.

[77] U. Bassler and G. Bernardi, Towards a coherent treatment of calorimetric energies: Missing transverse energy, jets, E.M. objects and the T42 algorithm, D0 note 4124, 2003.

[78] J.-R. Vlimant, U. Bassler, G. Bernardi, and S. Trincaz-Duvoid, Technical description of the T42 algorithm for the calorimeter noise suppression, D0 note 4146, 2003.

[79] G. Bernardi, E. Busato, and J.-R. Vlimant, Improvements from the T42 algorithm on calorimeter objects reconstruction, D0 note 4335, 2004.

[80] G. C. Blazey *et al.*, Run II jet physics, in *Proceedings of the Physics at RUN II: QCD and Weak Boson Physics Workshop*, pp. 47–77, 1999, hep-ex/0005012.

[81] E. Busato and B. Andrieu, Jet algorithms in D0 Run II software: Description and user's guide, D0 note 4457, 2004.

[82] A. Harel, Jet id optimization, D0 note 4919, 2005.

[83] A. Harel and R. Wagner, Improved L1 confirmation, D0 note 4932, 2005.

[84] B. Andrieu, A. Harel, H. Nogima, M. Rangel, and M. Voutilainen, Measuring Reconstruction*Jet-ID efficiencies using the tag and probe method in p17, D0 note 5250, 2006.

[85] N. Makovec and J.-F. Grivaz, Shifting, smearing and removing simulated jets, D0 note 4914, 2005.

[86] A. Harel, H. Nogima, M. Rangel, and M. Voutilainen, Combined JetID efficiency for p17, D0 note 5218, 2006.

[87] T. Scanlon, A neural network b-tagging tool, D0 note 4889, 2005.

[88] M. Anastasoaie, S. Robinson, and T. Scanlon, Performance of the NN b-tagging tool on p17 data, D0 note 5213, 2006.

[89] D. Duggan, The Performance and Operation of the D0 Calorimeter, To be published in proceedings of the XIII International Conference on Calorimetry in High Energy Physics (CALOR08), 26 - 30 May 2008, Pavia.

[90] K. Peters, Precision calibration of the D0 HCAL in run II, AIP Conf. Proc. **867**, 17 (2006).

[91] J. Kvita and K. Peters, Run II phi-intercalibration of the fine hadronic calorimeter, D0 note 5005, 2006.

[92] The D0 collaboration, Jet energy scale determination at D0 Run II, paper in preparation (to be submitted to Nucl. Instrum. Meth.).

[93] D0 Jet Energy Scale Group, Jet energy scale determination at D0 Run II (final p17 version), D0 note 5382, 2007.

[94] J. Hegeman, Jet Energy Scale at D0, To be published in proceedings of the XIII International Conference on Calorimetry in High Energy Physics (CALOR08), 26 - 30 May 2008, Pavia.

[95] CDF-II, R. Blair *et al.*, The CDF-II detector: Technical design report, FERMILAB-PUB-96-390-E.

[96] T. Sjostrand, L. Lonnblad, S. Mrenna, and P. Skands, PYTHIA 6.3: Physics and manual, (2003), `arXiv:hep-ph/0308153`.

[97] D. Stump *et al.*, Inclusive jet production, parton distributions, and the search for new physics, JHEP **10**, 046 (2003).

[98] CDF, R. Field and R. C. Group, PYTHIA Tune A, HERWIG, and JIMMY in Run 2 at CDF, (2005).

[99] CDF, F. Abe *et al.*, The Dijet angular distribution in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV, Phys. Rev. Lett. **69**, 2896 (1992).

[100] J. Stark, Modifications to the calorimeter simulation, Presentation at the 'MC summit', 2006.

[101] J. Stark, Fitting for the amount of required fudge material, Presentation at the D0 W mass meeting, 2007.

[102] J. Hegeman, Showering analysis for the final p17 jet energy scale, D0 note 5383, 2007.

[103] D0, V. M. Abazov *et al.*, Measurement of the inclusive jet cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$TeV, Accepted by PRL (2008), `arXiv:0802.2400 [hep-ex]`.

[104] L. Duflot, V. Shary, I. Torchiani, and R. Zitoun, cal_event_quality package, D0 note 4614, 2004.

[105] The D0 data quality group, The dq_defs package v2007-10-01.

[106] The D0 data quality group, The D0 data quality group web page, cited 20071220, `http://www-d0.fnal.gov/computing/data_quality/`.

[107] A. Harel, Capping the jes muon corrections, D0 note 5563, 2008.

[108] Alpgen generator home page, cited 2008-06-17, `http://mlm.home.cern.ch/mlm/alpgen/`.

[109] M. Begel *et al.*, Top-antitop pair production cross section in the hadronic channel, D0 note in preparation, 2008.

[110] D0 Alpgen page, cited 2008-06-17, `https://plone4.fnal.gov/P1/D0Wiki/simulation/alpgenreleases`.

[111] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, Matching matrix elements and shower evolution for top-quark production in hadronic collisions, JHEP **01**, 013 (2007).

[112] R. Field, Rick Field's Pythia tunes page, `http://www.phys.ufl.edu/~rfield/cdf/tunes/rdf_tunes.html`.

[113] D0gstar documentation web page, `http://www-d0.fnal.gov/computing/MonteCarlo/simulation/d0gstar.html`.

[114] Y. Fisyak and J. Womersley, D0gstar D0 GEANT Simulation o fthe Total Apparatus Response, D0 note 3191, 1997.

[115] M. Goossens, *GEANT Detector Description and Simulation Tool*, CERN Program Library Long Writeup No. W5013, October 1994 ed. (CERN, 1993).

[116] D0SIM User Manual, D0 note 407, 1986.

[117] D0SIM web page, `http://www-d0.fnal.gov/computing/MonteCarlo/simulation/d0sim.html`.

[118] D. Stump *et al.*, Inclusive jet production, parton distributions, and the search for new physics, JHEP **10**, 046 (2003), `arXiv:hep-ph/0303013`.

[119] M. G. Bowler, e+ e- Production of Heavy Quarks in the String Model, Zeit. Phys. **C11**, 169 (1981).

[120] Y. Peters, M. Begel, K. Hamacher, and D. Wicke, Reweighting of the fragmentation function for the D0 Monte Carlo, D0 note 5325, 2007.

[121] ALEPH, A. Heister *et al.*, Study of the fragmentation of b quarks into B mesons at the Z peak, Phys. Lett. **B512**, 30 (2001), `arXiv:hep-ex/0106051`.

[122] DELPHI, G. Barker *et al.*, A Study of the b-Quark Fragmentation Function with the DELPHI Detector at LEP I, DELPHI conference note 2002-069 CONF 603, 2002.

[123] OPAL, G. Abbiendi *et al.*, Inclusive analysis of the b quark fragmentation function in Z decays at LEP. ((B)), Eur. Phys. J. **C29**, 463 (2003), `arXiv:hep-ex/0210031`.

[124] ALEPH experiment home page, `http://aleph.web.cern.ch/aleph/`.

[125] DELPHI experiment home page, `http://delphiwww.cern.ch/`.

[126] OPAL experiment home page, `http://opal.web.cern.ch/Opal/`.

[127] SLD, K. Abe *et al.*, Precise measurement of the b-quark fragmentation function in Z0 boson decays, Phys. Rev. Lett. **84**, 4300 (2000), `arXiv:hep-ex/9912058`.

[128] SLD experiment home page, `http://www-sld.slac.stanford.edu/sldwww/sld.html`.

[129] C. Ochando and J.-F. Grivaz, SSR for p17, D0 note 5609, 2008.

[130] J. Hegeman, Event data quality efficiency for the p17 hadronic top cross section analyses, D0 note 5562, 2008.

[131] J. Hegeman, Luminosity determination and reweighting of Monte Carlo overlay luminosity for p17 hadronic top analyses, D0 note 5561, 2008.

[132] L. Christofek, R. Schwienhorst, M. Begel, and M. Kopal, Probability calculation for multi-object trigger configurations, D0 note 4882, 2005.

[133] D. W. K. Lam, *Measurement of Top Quark Mass in the All Hadronic Channel*, PhD thesis, Notre Dame University, 2008.

[134] I. Chakravarti, R. Laha, and J. Roy, *Handbook of methods of Applied Statistics Volume I* (Wiley, 1967).

[135] A. Hocker *et al.*, Tmva: Toolkit for multivariate data analysis, (2007).

[136] G. Hanson *et al.*, Evidence for Jet Structure in Hadron Production by e+ e- Annihilation, Phys. Rev. Lett. **35**, 1609 (1975).

[137] G. C. Fox and S. Wolfram, Event shapes in e+ e- annihilation, Nucl. Phys. **B149**, 413 (1979).

[138] D. Bourilkov, R. C. Group, and M. R. Whalley, LHAPDF: PDF use from the Tevatron to the LHC, (2006).

[139] A. Harel, An estimate of the data over MC, b over light jet response uncertainty for RunIIa JES, D0 note 5654, 2008.

[140] T. Andeen *et al.*, The D0 experiment's integrated luminosity for Tevatron Run IIa, D0 note 5398, 2007.

[141] F. Blekman, *Top quark pair production in proton antiproton collisions*, PhD thesis, Nikhef/University of Amsterdam, 2005.

[142] D0, V. M. Abazov *et al.*, Measurement of the $p\bar{p} \to t\bar{t}$ production cross section at $\sqrt{s} = 1.96$-TeV in the fully hadronic decay channel, Phys. Rev. **D76**, 072007 (2007), `arXiv:hep-ex/0612040`.

Acknowledgements

Dit proefschrift is niet zonder hulp tot stand gekomen. Velen hebben, op allerlei manieren, hun steentje bijgedragen. Een aantal van hen wil ik hierbij expliciet vermelden, zonder respect voor volgorde en ongetwijfeld niet compleet.

Allereerst wil ik mijn promotor Bob van Eijk bedanken. Niet veel promotors reizen hun studenten achterna naar het volgende continent. Zonder de vele inspirerende discussies (met en zonder koffie) was ik er nooit gekomen. Gelukkig had je altijd geduld om dingen uit te leggen en te bediscussiëren, ook als ik weer binnen kwam kleppen met de zoveelste (soms vermeende) leeuw of beer op het pad. Ik begrijp nu trouwens ook waarom schrijvers hun editor bedanken; tussen de eerste draft en deze uiteindelijke versie zit een wereld van verschil. Bob, bedankt. Ook wil ik bij deze Bob en Annette bedanken voor de gezelligheid (en Annettes kookkunst!) in Geneva (Illinois).

Sijbrand de Jong, onze Nederlandse D0-groepsleider, wil ik bedanken voor al zijn werk om onze D0-groep op gang te houden, en voor alle onverwachte natuurkunde lessen. Het was vaak een verrassing als je ineens op Fermilab stond, maar ik leerde dan altijd meteen iets nieuws, meestal over iets waar ik nog nooit zo over had nagedacht.

Going to the USA to work at Fermilab was quite a change. A different country, a car instead of a bike, buying milk by the gallon, etc. Fortunately there were a lot of people who quickly introduced me to life around Fermilab. I think back in particular to Mike Kirby (who let me borrow Bessie without even knowing me) and the rest of the Dutch D0 group over there: Pieter H (our link to the East China Inn), Miruna and Cristina, Pieter vdB (who also had a messed up schedule and so was often available for coffee and a chat at all possible hours), and Lucian. Axel deserves a big thank you, not only for being a good neighbour during my first winter in the Fermilab dorms, but also for answering all those questions I did not dare ask on ROOTTALK.

I owe a lot to Reiner Hauser and Paul Russo for software and debugging help when I manoeuvred myself into corners. I'm trying to remember at least half of what you taught me (and your mail addresses for the rest).

A lot of my time at Fermilab was spent in the D0 jet energy scale group, first led by Christophe and Markus, later by Aurelio and Alexander. I learned a lot there, albeit not always at my desired pace and time. I'm especially grateful to Aurelio for all the explanations and discussions about nitty-gritty physics details. Stubborn as I may be, I admit that I *did* learn a lot from you. Also a big thanks to Mikko and Jiri, who did not give up on my software when it balked, but sent me fixes and suggestions.

Many others at and around Fermilab deserve a big thank you. For answering questions, for sitting through and correcting my practice talks, and so on, and so on. A full list would dwarf the rest of this thesis.

I'm thinking back to good times spent with Doug, Maggy, and Ingo. To our visits to Chili's, to burning down trees, and to running around Nelson Lake after a day in the Outback. And Mike, Dennis and Gary: thank you for letting me join the group. I've never enjoyed BBQs and murky lakes this much!

Op het Nikhef (alleen de 'N' als hoofdletter!) verdient Frank Linde mijn grote dank. Hoewel mijn vierde jaar al ruimschoots verstreken was, gaf hij de moed niet op. Zijn hulp en financiële steun hielden mij lang genoeg op het droge om dit proefschrift af te maken.

Terug in Amsterdam waren de Pieters en ik de laatsten der D0-Mohikanen. We liepen redelijk synchroon en dus kwamen we min of meer tegelijkertijd terecht bij het uitzoekwerk voor onze detectorhoofdstukken en bij vragen als: 'Waarom doet LaTeX dat nou weer *zo?!*' Zonder al jullie vragen en antwoorden was dit boekje een stuk minder compleet geworden, en leven op Nikhef een stuk saaier. Niet dat de ATLAS groep (mét D0-friendly cubicle-kantoor) saai is hoor; ik zal de vrijdagmiddagborrels missen. Dit geldt ook voor de gang vol vraagbaken. Vooral Auke-Pieter, Wouter, Ivo, Marcel, 'Late lunch' Gossie, jullie deur stond altijd open voor vragen en discussies: thanks!

Both the analysis and I are indebted to Michael Begel. For pulling along the analysis, as well as for the basic ideas underlying the background model. At times it was pretty tricky keeping up with you and figuring out which files had gone where, especially with the eight-hour time difference. Some late nights on the phone taught me a lot of physics though. We will see each other at CERN. Mazzel tov!

Het is misschien een cliché maar niets is minder waar: zonder mijn ouders, papa Eef en mama Els, was ik nu niet geweest waar ik ben. Jullie leden mee als ik druk was (ben?), jullie kwamen me opzoeken in de States, behangden muren, verbouwden

keukens, niets was te gek. Zonder jullie steun woonde ik waarschijnlijk nog steeds op een zolderkamer totdat ik 'tijd had om iets te zoeken'.

Boogie en Bart wil ik bedanken voor alle garage sessies. Altijd gezellig, meestal later dan gepland, en een grote bron van inspiratie. (Volgens mij hebben we de meeste wereldproblemen al wel eens opgelost, maar ik ben het 's morgens vaak weer vergeten.)

Ere komt ook toe aan Kees Huyser. Dank voor de regelmatige injecties van praktisch denken, voor vele gezellige avonden in de kroeg, en (zeker niet in de laatste plaats) voor het 'klaar terwijl u wacht' omslagontwerp.

Tenslotte mijn dank aan Waterscouting Beatrix, voor de onuitputtelijke lijst met beschikbare goede daden. Ook al klonk ik af en toe niet zo, ik heb me er jaren lang goed vermaakt. Ik vond het ook fijn dat ik, na zo lang weg te zijn geweest, gewoon terug mocht komen in de staf en om te komen klussen. Op deze plek ook dank aan mijn 'kleine broertje' Maarten. Voor alle hulp links en rechts en voor het op orde houden van onze TS.

And last but not least: thank *you*, the reader, for making it this far!