Dissertation ETH Nr. 16561

Measurement of b-quark jet shapes at CDF

A dissertation submitted to the

Swiss Federal Institute of Technology Zürich

for the degree of

Doctor of Natural Sciences

presented by

Alison Lister

MPhys, The University of Oxford

born 18th June 1980

citizen of Pully, Vaud and Great Britain

Prof. G. Dissertori examiner
Prof. T. Gehrmann co-examiner

- 2006 -

Contents

Table of Contents iii

Abstract v

Résumé vii

Introduction 1

1 Theoretical Background 3
- 1.1 The Standard Model of Particle Physics . . . 3
  - 1.1.1 QCD . . . 6
- 1.2 From Partons to Jets . . . 7
- 1.3 Jet Shapes . . . 11

2 The CDF Experiment 17
- 2.1 The Tevatron . . . 17
- 2.2 The CDF Detector . . . 20
  - 2.2.1 Inner Tracking . . . 21
  - 2.2.2 Calorimeters . . . 22
  - 2.2.3 Trigger System . . . 23

3 Jet Algorithms 25
- 3.1 Cone Algorithms . . . 26

4 Secondary Vertex Tagging at CDF 31
- 4.1 The SecVtx Algorithm . . . 31
- 4.2 Applying SecVtx to Different Jet Algorithms . . . 33
  - 4.2.1 Algorithm Comparison Using a Sub-Cone of Size 0.4 . . . 34
  - 4.2.2 Effect of Changing the Sub-Cone Size from 0.4 to 0.7 . . . 52
  - 4.2.3 Conclusions on the SecVtx Algorithm Study . . . 59

5 Event Selection 61
- 5.1 Datasets . . . 61
- 5.2 Monte Carlo Samples . . . 62
- 5.3 Event Selection . . . 63
- 5.4 Average p_T Correction . . . 64
- 5.5 p_T Thresholds . . . 64

6 b-quark Jet Shapes 67
- 6.1 Unfolding Method . . . 68
- 6.2 Raw Shapes . . . 69
- 6.3 Purity . . . 75

6.4 Biases Due to SecVtx Tagging . . . 77
6.5 Hadron Level Corrections . . . 77
6.6 Hadron Level b-quark Jet Shapes . . . 81

7 Systematic Studies **83**
7.1 Extraction of the Statistical Error on the Purity . . . 84
7.2 Estimate of the c-fraction in Data . . . 84
7.3 Estimate of the Error on $f_{1\mathrm{b}}$ and $f_{1\mathrm{c}}$ . . . 89
7.4 Jet Energy Scale . . . 90
7.5 Missing E_T Significance . . . 90
7.6 Primary Vertex z-position . . . 90
7.7 Inclusive vs. nonb-jet Shapes . . . 91
7.8 Jet Shapes Using tracks . . . 93
7.9 Jet Shapes with Tower $p_T > 0.5\,\mathrm{GeV}$ . . . 97
7.10 MC Dependence of the Unfolding Parameters . . . 101
7.10.1 Purity . . . 101
7.10.2 Biases Due to SecVtx Tagging . . . 103
7.10.3 Hadron Level Corrections to the Jet Shapes . . . 105
7.10.4 Hadron Level b-jet Shapes Using the Herwig MC for the Unfolding Parameters . . . 106
7.11 MC Modelling of the SecVtx Performance . . . 112
7.12 Total Systematic Errors . . . 116
7.13 Changing the Rapidity Cut to $0.1 \leq |Y| \leq 0.7$ . . . 118

8 Results **123**
8.1 MC Reconstruction Checks . . . 123
8.2 Integrated b-jet Shapes . . . 134
8.3 Variation with p_T . . . 137
8.4 Dependence on the Single b-quark Jet Fraction . . . 138

9 Summary and Conclusions **145**

A Differential b-quark Jet Shapes **147**

B CMS ECAL Detector Control System **149**

Bibliography **157**

Acknowledgements **159**

Abstract

The main topic of this thesis is the measurement of b-quark jet shapes at CDF. CDF is an experiment located at Fermilab, in the United States, which studies proton-antiproton collisions at a centre of mass energy of 1.96 TeV. To reach this energy, the particles are accelerated using the Tevatron accelerator which is currently the highest energy collider in operation. The data used for this analysis were taken between February 2002 and September 2004 and represent an integrated luminosity of about $300\,\mathrm{pb}^{-1}$. This is the first time that b-quark jet shapes have been measured at hadron colliders.

The basis of this measurement lies in the possibility of enhancing the b-quark jet content of jet samples by requiring the jets to be identified as having a displaced vertex inside the jet cone. Such jets are called tagged. This enhances the b-quark jet fraction from about 5% before tagging to $20 - 40\%$ after tagging, depending on the transverse momentum of the jets. I verified that it is possible to apply this secondary vertex tagging algorithm to different cone jet algorithms (MidPoint and JetClu) and different cone sizes (0.4 and 0.7). I found that the performance of the algorithm does not change significantly, as long as the sub-cone inside which tracks are considered for the tagging is kept at the default value of 0.4.

Because the b-quark purity of the jets is still relatively low, it is necessary to extract the shapes of b-quark jets in a statistical manner from the jet shapes both before and after tagging. The other parameters that enter into the unfolding equation used to extract the b-quark jet shapes are the b-jet purities, the biases due to the tagging requirement both for b- and nonb-jets and the hadron level corrections. The last of these terms corrects the measured b-jet shapes back to the shapes expected at hadron level which makes comparisons with theoretical models and other experimental results possible.

This measurement shows that, despite relatively large systematic uncertainties, the measured b-quark jet shapes are significantly different from those expected from the so-called Pythia Tune A Monte Carlo simulation, the most widely used Leading Order Monte Carlo model at CDF. This difference can be mostly attributed to the fact that the fraction of b-quark jets that originate from flavour creation (where a single b-quark is expected inside the same jet cone) over those that originate from gluon splitting (where two b-quarks are expected to be inside the same jet cone) is slightly different in the Pythia Tune A Monte Carlo predictions than in data. This measurement can help in the tuning of the fraction of gluon splitting to flavour creation b-quark jets in the Monte Carlo simulation. This tuning is particularly important for the extrapolation up to LHC energies where many searches will involve b-quark jets.

During the first year of my thesis work, I worked on the implementation of a prototype detector control system for the electromagnetic calorimeter which is being built for the CMS experiment at CERN. The prototype which I implemented was used to monitor and control the high voltage, low voltage, cooling and precision temperature monitoring systems during the summer 2003 test-beam. This was one of the first, almost complete, systems implemented and used by an LHC experiment for test-beam monitoring.

Résumé

Le sujet principal de cette thèse est la mesure des formes de jets ("jet shapes") de quarks b en utilisant des données de CDF. CDF est une expérience qui se trouve au Fermilab, aux Etats Unis, qui étudie des collisions de protons avec des antiprotons à une énergie de centre de masse de 1.96 TeV. Pour atteindre cette énergie, les particules sont accélélées par le Tevatron qui est actuellement l'accélérateur doté de la plus haute énergie. Les données utilisées pour cette analyse ont été saisies entre février 2002 et septembre 2004. Cela représente une luminosité integrée d'environ $300\,\mathrm{pb}^{-1}$. Pour la première fois, les formes de jets de quarks b ont été mesurées à un accélérateur hadronique.

La base de cette mesure est la possibilité d'augmenter la fraction des jets de quarks b dans les données en exigeant que les jets soient identifiées comme ayant un "vertex" déplacé à l'intérieur du cone du jet; les jets sont alors dits "taggés". Cela augmente la fraction des jets provenant de quarks b de $\sim 5\%$ avant tout tagging à $20 - 40\%$, dépendant de la quantité de mouvement transverse, après le tagging. J'ai fait une étude pour vérifier s'il était possible d'appliquer cet algorithme pour tagger les vertex secondaires sur différents algorithmes coniques de jets (MidPoint et JetClu) et sur des différentes tailles de cone (0.4 et 0.7). Il en résulte que la performance de l'algorithme SecVtx ne change pas de manière significative, pour autant que le cone à l'intérieur duquel se trouvent les traces considérés pour le tagging soit maintenu à sa taille originale de 0.4.

Vu que la pureté des jets de quarks b est encore relativement faible, il est nécessaire d'extraire les formes des jets de quarks b d'une manière statistique à partir des formes des jets avant et après le tagging. Plusieurs autres paramètres entrent dans l'équation utilisée afin d'extraire la forme des jets de quarks b. Ces paramètres sont la pureté, les corrections pour l'influence dû au tagging sur les jets de quarks b et les jets de quarks non-b et les corrections hadroniques. Le dernier de ces paramètres est nécessaire pour obtenir une mesure qui est aussi indépendante que possible des particularités du détecteur. Ce terme corrige les formes des jets de quarks b mesurés jusqu'au niveau hadronique et rend plus facile les comparaisons à des modèles théoriques ainsi qu'à d'autre expériences.

Cette mesure montre que, malgré des incertitudes systématiques assez larges, les formes des jets de quarks b mesurés diffèrent de manière significative des formes prédites par le Leading Order (LO) Monte Carlo le plus répandu à CDF, qui s'appelle Pythia Tune A. Cette différence peut être en majeure partie attribuée au fait que la somme des jets de quarks b qui viennent du "flavour creation" (où un seul quark b est attendu à l'intérieur de chaque cone de jet) divisé par ceux qui viennent du "gluon splitting" (où deux quarks b sont attendus dans le même cone de jet) est différente dans les données que dans le Monte Carlo Pythia Tune A. Cette mesure peut aider à cerner la fraction des jets de quarks b qui proviennent du gluon splitting ou du flavour creation dans les simulations Monte Carlo. Ceci est particulièrement important pour pouvoir extrapoler les résultats de mesures sur les jets de quarks b aux énergies du futur LHC où beaucoup de recherches de nouvelles particules utiliseront des jets de quarks b.

Pendant la première année de ma thèse, j'ai travaillé sur l'implémentation d'un protoype du Detector Control System (DCS) pour le calorimètre électromagnetique qui est en train d'être construit pour l'expérience CMS au CERN. Le prototype que j'ai implémenté a été utilisé pendant le faisceau test de l'été 2003 pour la surveillance et le contôle des systèmes de haute tension, de basse tension, du circuit de refroidissement et pour les mesures de précision de la température. C'était un des premiers systèmes, presque complets, mis au point et utilisés par une expérience du LHC pour un système de surveillance de faisceau test.

Introduction

The main topic of my thesis is the measurement of b-quark jet shapes using $\sim 300\,\mathrm{pb}^{-1}$ of data from the Collider Detector at Fermilab - CDF - experiment. It is the first time that such a measurement has been carried out at a hadron collider.

The layout of this thesis is as follows. Chapter 1 briefly describes the Standard Model of particle physics. This is currently the best model for describing the interactions of elementary particles. Particular emphasis is put on the theory of the strong interaction, Quantum Chromodynamics (QCD). It is important to understand what models are used by the Monte Carlo (MC) generators to relate what can be directly calculated within the framework of the Standard Model, the hard-scatter partons, to what is observed in our detectors, the final state hadrons. A brief description of the basis for such models can be found in section 1.2. The final section of this chapter defines the notion of jet shapes and b-quark jets, also called b-jets.

Chapter 2 describes the experimental setup. It starts by describing the accelerator complex at Fermilab which produces anti-protons and accelerates protons and antiprotons, in various stages, culminating in the final acceleration in the Tevatron accelerator. The two beams are collided at the location of the two experiments: CDF and D0. The CDF detector is described in section 2.2. Particular emphasis is put on the silicon vertex detectors and the calorimetic systems which are of particular importance for my analysis.

Follows, in chapter 3, a description of the different jet algorithms used at CDF. This discussion concentrates on cone algorithms of which the algorithm used for this analysis, MidPoint, is part. The emphasis is put on the differences between the Run I algorithm, JetClu, and the new, improved Run II algorithm, MidPoint.

The next chapter, chapter 4, presents the secondary vertex tagging algorithm which is used to enhance the heavy flavour content of the samples used for this analysis. Inclusive samples contain only about 4% of b-quark jets, not sufficient to carry out any detailed measurements of b-quark jet properties. By tagging the jets this fraction is increased to $\sim 30\%$. This chapter also presents a study carried out by a colleague from the University of Geneva and myself, which can also be found in CDF internal note number 7045. The study determines that it is possible to use the current secondary vertex tagging algorithm, SecVtx, with all the jet cone algorithms used at CDF, irrespective of the cone size, as long as the cone which is used by the algorithm stays at its default value.

The remaining chapters of this thesis present the analysis which I carried out between October 2003 and March 2006. This analysis has been written up in two CDF internal notes [1] and [2]. This first note presents an analysis methodology which was not used for the final measurement because of the large systematic uncertainties. It presents an attempt to measure separately the jet shapes for b-jets that contain a single b-quark inside the jet cone and for b-jets that contain two (or more) b-quarks inside the same jet cone. The second note describes the methodology presented here. It contains, in addition to the material presented in this thesis,

an appendix on the comparison between data and Pythia Tune A MC of some general quantities.

Chapter 5 presents the different datasets used for this analysis along with the Monte Carlo samples. Section 5.3 presents the additional cuts applied to the jets as well as the corrections applied to account for the effects of the detector on the transverse momentum of the jets. Finally the binning used for this analysis is explained.

Chapter 6 presents the unfolding method used to extract the b-quark jet shapes from the measured, raw, jet shapes for tagged and inclusive jets. This must be done statistically as the tagged samples contain only $\sim 30\%$ of b-quark jets. The different parameters of the unfolding method are presented in sections 6.3 through 6.5. The parameters needed to extract the b-quark jet shapes are the raw inclusive and tagged shapes, the b-quark jet fraction of the samples and bias factors to account for the effect of the tagging on b- and nonb-jets separately. Finally, correction factors are applied to correct the b-quark jet shapes back to hadron level. Section 6.6 presents the hadron level b-quark jet shapes, showing only the statistical errors on the measurement.

Chapter 7 details the systematic error studies carried out for this analysis. The main sources of systematic uncertainties turn out to be due to the uncertainty on the fraction of b-quark jets that contain only a single b-quark jet inside the jet cone and similarly the fraction of c-quark jets, part of the background, that contain only a single c-quark inside the jet cone. Another main source of systematic uncertainty is the difference between the hadron level b-quark jet shapes measured from raw shapes using tracks and using calorimeter towers.

The final results for the integrated b-quark jet shapes are presented in chapter 8. The results are compared to Pythia Tune A Monte Carlo predictions for both the inclusive and the b-quark jet shapes. The agreement between the data and the Pythia Tune A MC is not very good for b-quark jet shapes. Different representations of these results are shown. In particular the results are compared to the Pythia Tune A Monte Carlo predictions for jets containing a single b-quark inside the jet cone with those containing two of them. The former are mainly from flavour creation whereas the latter come mainly from gluon splitting events. A fit to the data of the single and double b-quark jet shapes allows the extraction of the most probable single b-quark fraction. This fraction turns out to be $10-20\%$ lower than predicted by the Pythia Tune A MC.

Chapter 9 wraps up the main part of this thesis by drawing some conclusions about this analysis. Appendix A presents a brief comment about the reasons why it was not possible to measure the differential b-quark jet shapes for this analysis.

I spent the first year of my thesis developing a detector control system for the ECAL detector for the CMS experiment at the LHC. I wrote applications to control the high voltage, low voltage, cooling system and also a 2003 test-beam specific application to map the temperature profile inside a supermodule. This work is presented in appendix B.

Chapter 1

Theoretical Background

What is the world made of? This is one of the fundamental questions that both scientists and philosophers have been trying to answer for centuries. The idea that there must be some intrinsic, fundamental constituents of matter which can't be broken down into smaller pieces is nothing new. It is only over the last century, with the advent of modern physics and modern technologies, that much progress has been made in the understanding of the basic constituents of matter. Particle physics is the science that explores these basic constituents and how they interact. It is not certain that what is now believed to be the fundamental constituents of matter really are the smallest building blocks of the universe. Yet no experimental evidence has so far led us to believe otherwise. We know that our current model is not the end of the story but it is at least a very good approximation to what has been observed up to now.

Currently, the model that best explains the fundamental particles and their interactions, as we see them, is called the Standard Model of particle physics (SM) and is discussed below. The particular aspects of this model which are needed to understand the scope of this analysis are expanded in the following sections.

1.1 The Standard Model of Particle Physics

As far as experiments have shown, the world is made up from only a small set of particles. These particles can be split into three categories. The first two categories are spin $\frac{1}{2}$ particles, fermions. These are shown in the top table of figure 1.1. These are on the one hand the quarks which are the basic constituents of particles such as protons and neutrons. On the other hand there are the leptons, among which are to be found the electrons and neutrinos. The final category contains the set of bosons, spin 1 particles, which are the mediators of three of the four fundamental forces known in nature. These are shown in the middle table of figure 1.1 One final particle is needed in the Standard Model. This is the Higgs boson. It has not yet been discovered but could explain the origin of the masses of the different particles [3].

Quarks can be arranged into three groups, called generations, of increasing mass, where each generation consists of a quark with electric charge $+\frac{2}{3}$ and one with charge $-\frac{1}{3}$. The quarks in the lowest mass generation are the basic constituents of protons and neutrons. These are the up and down quarks. The second generation contains the charm and strange quarks and the final generation consists of the top and bottom quarks. The existence of the heaviest of these quarks, the top quark, was postulated many years before it was discovered at the Tevatron in 1995 [5]. It was first postulated in 1977, the year the bottom quark was discovered.

Leptons may also be arranged into three generations of increasing mass. Each negatively

FERMIONS

matter constituents
spin = 1/2, 3/2, 5/2, ...

Leptons spin = 1/2

Flavor	Mass GeV/c^2	Electric charge
ν_e electron neutrino	$<1\times10^{-8}$	0
e electron	0.000511	−1
ν_μ muon neutrino	<0.0002	0
μ muon	0.106	−1
ν_τ tau neutrino	<0.02	0
τ tau	1.7771	−1

Quarks spin = 1/2

Flavor	Approx. Mass GeV/c^2	Electric charge
u up	0.003	2/3
d down	0.006	−1/3
c charm	1.3	2/3
s strange	0.1	−1/3
t top	175	2/3
b bottom	4.3	−1/3

BOSONS

force carriers
spin = 0, 1, 2, ...

Unified Electroweak spin = 1

Name	Mass GeV/c^2	Electric charge
γ photon	0	0
W^-	80.4	−1
W^+	80.4	+1
Z^0	91.187	0

Strong (color) spin = 1

Name	Mass GeV/c^2	Electric charge
g gluon	0	0

PROPERTIES OF THE INTERACTIONS

Property \ Interaction	Gravitational	Weak (Electroweak)	Electromagnetic (Electroweak)	Strong: Fundamental	Strong: Residual
Acts on:	Mass – Energy	Flavor	Electric Charge	Color Charge	See Residual Strong Interaction Note
Particles experiencing:	All	Quarks, Leptons	Electrically charged	Quarks, Gluons	Hadrons
Particles mediating:	Graviton (not yet observed)	W^+ W^- Z^0	γ	Gluons	Mesons
Strength relative to electromag for two u quarks at: 10^{-18} m	10^{-41}	0.8	1	25	Not applicable to quarks
Strength relative to electromag for two u quarks at: 3×10^{-17} m	10^{-41}	10^{-4}	1	60	
for two protons in nucleus	10^{-36}	10^{-7}	1	Not applicable to hadrons	20

FIGURE 1.1: The Standard Model of particle physics in a nutshell! The first table gives a description of the matter components, the fermions. The second describes the force carriers, the vector bosons. The last one summarises the different interactions [4]. *The residual strong interaction is the strong interaction of colour-neutral protons and neutrons when they form nuclei. It is due to a residual strong interaction between their colour-charged constituents, like for example the Van der Waals force in water which is a residual electric force.*

charged lepton is paired up with a neutral neutrino. The three pairs are, again in order of increasing mass: the electron and electron neutrino, the muon and muon neutrino and the tau and tau neutrino. Electrons are the only stable, i.e. non-decaying, charged leptons and, along with the up and down quarks that form the nuclei, they form all the atoms found on earth.

The final category of particles contains the mediators of forces and interactions. The four fundamental forces of nature are: gravity along with the electromagnetic, the weak and the strong forces. Gravity is negligible with respect to the strengths of the other forces, given the high energy reached by particle accelerators and the small masses of the fundamental particles. Gravity is not included in the Standard Model of particle physics [1].

The electromagnetic force is mediated by massless photons that couple to charged particles. Photons themselves carry no electric charge so they cannot couple to each other.

The weak force is mediated by two charged (W^+ and W^-) and one neutral (Z^0) vector bosons. The strength of the weak coupling between particles is determined by the weak isospin. The weak force does not couple to the mass eigenstates but to the weak isospin eigenstates of the fermions. The mass and the weak isospin eigenstates are not identical for quarks; they are related to each other by a matrix called the CKM matrix (after Cabibbo, Kobayashi and Maskawa). Fermions carry weak isospin, $T = 1/2$ with the third component $T_z = \pm 1/2$. The vector bosons each carry one unit of isospin, $T = 1$, with third component $T_z = 1, -1, 0$ for the W^+, W^- and Z^0, respectively.

The strongest of all forces is the strong force, also known as the colour force. It describes the interaction between quarks and is mediated by gluons. Gluons each carry two units of colour charge and can couple to each other. The strong force will be discussed in more detail in the next section. The bottom table of figure 1.1 shows the force mediators and what the forces act on.

As mentioned previously, there is still one fundamental building block of the Standard Model that is missing, the Higgs boson. The basic idea is that a scalar field, the Higgs field, couples to the different fundamental particles to give them mass after spontaneous symmetry breaking. The Higgs potential is thought to be a "mexican hat" potential: a rotationally symmetric potential which has a local maximum at zero and a rotationally symmetric minimum (symmetric around ϕ say). This minimum is called the vacuum expectation value. The vacuum state does not have the same symmetry an the interaction potential. The symmetry is thus spontaneously broken. The interaction between the Higgs field and the other particles is calculable within the SM but the mass of the Higgs boson is not predicted [6].

We now know there must be physics beyond the Standard Model. There is a substantial number of experiments which have shown results which cannot be integrated into the Standard Model. One example worth mentioning is that neutrinos (ν_e, ν_μ and ν_τ) have been found to have non-zero masses. In the SM the neutrinos are massless particles.

Another reason to look at theories beyond the Standard Model is to attempt to unify not only the weak and the electromagnetic forces, as has already been done, but also the strong force and maybe even gravity. One such theory is called Supersymmetry. It postulates that there must be a symmetry between the fermions and bosons. A supersymmetry transformation turns a bosonic state into a fermionic one and vice-versa. The operator Q which generates such tranformations must be an anti-comutating spinor where

$$Q|\text{Boson}>=|\text{Fermion}> \;\; ; \;\; Q|\text{Fermion}>=|\text{Boson}> . \tag{1.1}$$

[1] But gravity is the dominant force when considering cosmological observations.

Q has to be a fermionic operator, it must therefore carry spin angular momentum 1/2 so it is clear that supersymmetry must be a space-time symmetry [7].

The Standard Model is thought to be an effective theory, valid up to a particular energy scale, usually called Λ, so it is important to determine what this scale is and what are the scales at which the different symmetry breaking mechanisms occur.

It is to answer many of these questions as well as to test the validity of the current models to a very high precision that we carry out experiments using large colliders and experimental detectors (see chapter 2).

1.1.1 QCD

Quantum Chromodynamics is the theory of the strong interaction [8] [9]. It is the theory which describes how particles that carry a colour charge interact. Quarks carry one unit of colour charge, antiquarks carry anticolour and gluons each carry both a colour and an anticolour. There are three colour charges which are traditionally called red (r), green (g) and blue (b). The colour quantum number is never observed in nature so all interactions must be invariant under colour interchange. This discrete symmetry under colour interchange is a special unitary group symmetry in three dimensions, SU(3) symmetry. One of the consequences of this group symmetry is the self-coupling of the mediators of the strong force, i.e. the self-coupling of gluons. We would expect $3^2 = 9$ combinations of colour-anticolour pairs for the gluons but one of the eigenstates ($\frac{1}{\sqrt{3}}(r\bar{r} + g\bar{g} + b\bar{b})$) is a totally symmetric colour singlet which carries no net colour charge so it is decoupled. The remaining eight colour eigenstates of the gluons are

$$r\bar{b}, r\bar{g}, b\bar{g}, b\bar{r}, g\bar{r}, g\bar{b}, \frac{1}{\sqrt{2}}(r\bar{r} - b\bar{b}), \frac{1}{\sqrt{6}}(r\bar{r} + b\bar{b} - 2g\bar{g}) \tag{1.2}$$

The fact that the colour quantum number is not an observable (unlike the electric charge) implies that quarks can never be found as free particles but are confined to colour neutral bound states called hadrons. Hadrons can contain either two ($q\bar{q}$) or three (qqq) quarks whose total colour charge must be zero. The former are called mesons, the latter baryons. Other possibilities such as qq or $qq\bar{q}\bar{q}$ are not bound, i.e. the potential is found to be repulsive.

The colour charge of the strong interaction is analogous to the electric charge in electromagnetic interactions in that both forces are mediated by a massless vector boson. However, the photon itself does not carry any electric charge whereas the gluon is (colour) charged. This difference, which makes the theory non-Abelian, i.e. it allows the self-coupling of gluons, turns out to be crucial in the understanding of the features of quark interactions at short distances. Since gluons are massless particles, one might expect the QCD potential to have a similar $1/r$ form to the Coulomb potential between elementary charges. In fact the quark-antiquark potential is often taken to be of the form

$$V_s = -\frac{3}{4}\frac{\alpha_s}{r} + kr \tag{1.3}$$

where α_s is the strong coupling constant, of the order of 1 [2]. However, in very high q^2 collisions [3], $\alpha_s \sim 0.1$ and single gluon exchange is a good approximation. The first term, dominant at small separations, arises from this single gluon exchange. The second, linear, term is associated with the quark confinement at large separations. Because of the linear term, attempts to free a quark from a hadron simply result in the creation of a new $q\bar{q}$ pair. Several models exist to explain this process. One of these models is that the force lines of the colour field can be imagined to

[2] This is to be compared with the electroweak $\alpha \sim \frac{1}{137}$

[3] q^2 is the momentum transfer squared, i.e. a measure of how much momentum is exchanged in the interaction. It is defined as $q = p_i - p_f$ where p_i and p_f are the initial and final state total momentum, respectively.

be pulled together by the gluon-gluon interactions so that they form a tube. Pulling apart this tube, the stored energy (kr) eventually reaches the point where it is more favourable, energetically, to create a new $q\bar{q}$ pair resulting in two short tubes instead of one long one. From the sizes and masses of the hadrons, we expect $k \sim 1\,\mathrm{GeV/fm}$. This corresponds to a force between the quarks of 14 tons weight! The process by which new $q\bar{q}$ pairs are produced is called fragmentation. It is an important factor in the understanding of hadronic events, as described in section 3.

Formally, QCD interactions can be expressed, in a similar way to QED, by a Lagrangian. The QCD Lagrangian has the form

$$\mathcal{L}_{\mathrm{QCD}} = -\frac{1}{4}F^{\mu\nu,a}F^a_{\mu\nu} + \bar{q}(i\partial\!\!\!/ - m_q)q - g_s\bar{q}\gamma^\mu T^a q A^a_\mu. \tag{1.4}$$

The first term is the kinetic term of the gluon field, A^a_ν, expressed in terms of the gluon field tensor, $F^a_{\mu\nu}$, which is expressed as

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - g_s f^{abc} A^b_\mu A^c_\nu. \tag{1.5}$$

The last term of equation 1.5, which is non-Abelian, is required by the gauge invariance constraint under local SU(3) transformations of the gluon field. The second term of equation 1.4 is the free Dirac field equation for the quarks whereas the third term is the colour interaction current term between quarks and gluons.

In analogy to QED, the strength of QCD interactions can be expressed as $\alpha_s = \frac{g_s}{4\pi}$. This coupling constant is dependent on Q^2 [4]. The evolution of α_s with Q^2 is dictated by the renormalisation of quantum corrections and gives an evolution equation of the form

$$\mu^2\frac{\partial\alpha_s}{\partial\mu^2} = \beta(\alpha_s) \tag{1.6}$$

where $\beta(\alpha_s)$ is a convergent series as long as $\frac{\alpha_s}{4\pi} < 1$. At leading order, the value of α_s at some scale Q^2 can be expressed in terms of the $\alpha_s(Q_0^2)$ at another scale Q_0^2 as

$$\frac{1}{\alpha_s(Q^2)} = \frac{1}{\alpha_s(Q_0^2)} + \frac{\beta_0}{4\pi}\log\frac{Q^2}{Q_0^2} \tag{1.7}$$

The specific running of α_s with energy is typical of a non-Abelian theory: the coupling decreases with increasing q^2. In fact, at asymptotically large q^2, the coupling tends to zero, i.e. the quarks behave as free particles. This is known as asymptotic freedom and allows the inter-quark interactions inside the hadrons to be neglected at hadron colliders when dealing with the primary hard interactions.

In the following section, the perturbative process whereby the hard-scattered partons (quarks or gluons) decay to a series of final state partons is discussed along with the non-perturbative process which turns the partons into the hadrons that are observed in particle physics detectors.

1.2 From Partons to Jets

One of the main problems of many QCD analyses is to try and understand what is seen in the detectors (part (6) of figure 1.2), namely tracks and energy deposits in the calorimeters, in terms of what is produced in the primary interactions (part (1) of figure 1.2). There are of course effects due to the reconstruction algorithms as well effects due to the interaction of the particles

[4] Q^2 is defined as $-q^2$.

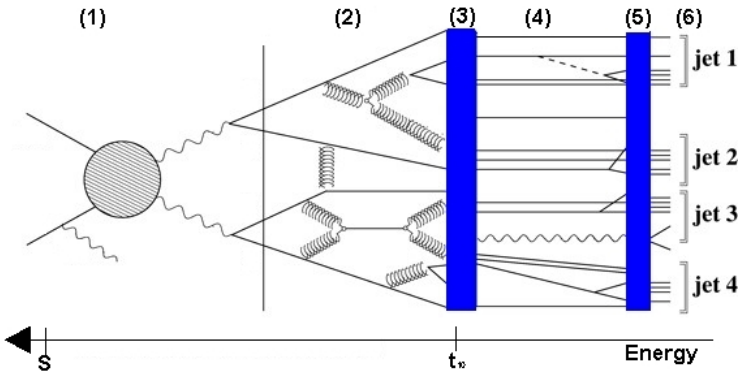

FIGURE 1.2: Schematic drawing of an interaction involving quarks. First the hard-scattering occurs which could include some initial and final state radiation (1). In this case a decay of two photons (wavy lines) to two pairs of quarks (straight lines) is shown. Then the fragmentation process occurs (2): many gluons (spring like lines) can be emitted which can in turn decay to quark-antiquark pairs. This process occurs until the energy scale of each parton is close to t_0. The non-perturbative hadronisation process then occurs (3) which turns the partons into hadrons. The latter can decay into a series of final state particles (4) which are measured by the detector (5). Finally objects such as jets are reconstructed from what is observed in the detector (6). In this case four jets would be observed.

with the detectors (parts (4) to (6) of figure 1.2). The effect of the reconstruction algorithms of particle showers, called jets, will be discussed in chapter 3. Detector effects should of course be accounted for in any analysis but lets concentrate on the physics part of this process, namely fragmentation and hadronisation (parts (2) and (3) of figure 1.2). The process whereby the initial partons create many more partons is known as the fragmentation process. This process is illustrated in the part (2) of figure 1.2. A quark can emit a gluon or, with the help of a nearby quark, create a new quark-antiquark pair. A gluon can emit another gluon or decay into a quark-antiquark pair. This process continues until the virtual masses squared of the partons are of the order of the infrared cut-of scale, t_0, taken to be of the order of $1\,\mathrm{GeV}^2$ [5]. After this, the low-momentum-transfer, long-distance regime is entered, in which non-perturbative effects become important. The most important of these effects is hadronisation which converts the partons into the observed hadrons (part (3) of figure 1.2). Most models of these phenomena, when implemented in Monte Carlo simulations, assume that the two processes, fragmentation and hadronisation, occur in sequence and can thus be treated independently [8].

Fragmentation models focus on the calculation of approximate results of parton showering where the enhanced terms, such as soft or collinear parton emission from the primary parton, are taken into account for all orders of the perturbative expansion in terms dependent on α_s. The parton shower models represent an approximate perturbative treatment of QCD dynamics at scales larger than t_0. The implementation of fragmentation models in Monte Carlo simulations mostly use Sudakov form factors, with or without the requirement of angular ordering. These form factors determine the probability of evolving from scale t to the shower cut-off scale, t_0, without branching. Each parton species has its own form factor, $\Delta(t)$. The form factors are expressed as

$$\Delta(t) = \exp\left\{-\int_{t_0}^{t}\frac{dt'}{t'}\int_{\epsilon(t')}^{1-\epsilon(t')} dz\frac{\alpha_s(z,t')}{2\pi}\hat{P}_{qq}(z)\right\} \tag{1.8}$$

[5] In the limit of massless partons, the virtual mass squared of a parton is simply its momentum squared

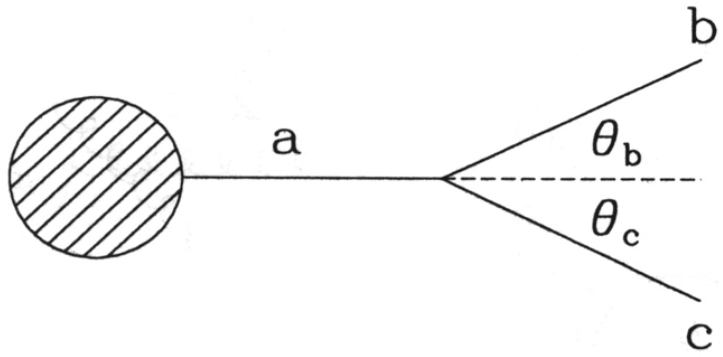

FIGURE 1.3: Schematic drawing of a parton a branching into two partons b and c. If parton a is a quark, partons b and c must be a quark and gluon or vice-versa. If parton a is a gluon, partons b and c can be both gluons or both quarks. The rest of the interaction is ignored in this diagram. This picture is a reproduction of a figure from [8].

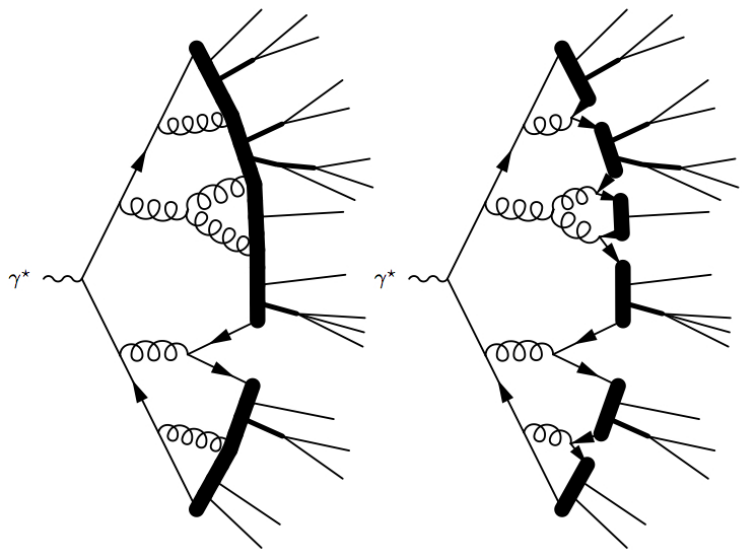

FIGURE 1.4: Schematic drawing of the hadronisation process whereby partons become a final set of hadrons. The left hand diagram represents the string model whereas the right hand diagram represents the cluster model. In both cases, the decay represented is that of a photon (wiggly line) decaying to a quark-antiquark pair (straight lines) which then fragments to other partons including gluons (cork-screwed lines). This figure is reproduced from [9].

for the collinear approximation, i.e. not requiring angular ordering of the shower, and

$$\tilde{\Delta}(\tilde{t}) = \exp\left\{ -\int_{4\,t_0}^{\tilde{t}} \frac{dt'}{t'} \int_{\sqrt{\frac{t_0}{t'}}}^{1-\sqrt{\frac{t_0}{t'}}} dz \frac{\alpha_s(z^2(1-z)^2 t')}{2\pi} \hat{P}_{qq}(z) \right\} \tag{1.9}$$

for the soft approximation where angular ordering appears. These equations hold as long as the virtuality of the outgoing partons (b and c of figure 1.3) is significantly smaller than the vitualities of the incident parton (a)

$$p_b^2, p_c^2 \ll p_a^2 = t \tag{1.10}$$

t is the scale of the interaction, often set, as above, to the momentum squared of the original parton. t_0 is the parton shower cut-off scale, i.e. the scale at which partons hadronise. z is the energy fraction of the outgoing parton with respect to the incoming one

$$z = E_b/E_a \tag{1.11}$$

$\hat{P}_{qq}(z)$ is called the unregularised splitting function and is related to the corresponding Altarelli-Parisi kernel [10]. $\epsilon(t)$ is a cut-off which constrains the emission to be visible (i.e. not soft). $\alpha_s(z, t')$ is the value of the strong coupling constant at energy fraction z and scale t'.

In principle, the shower cut-off, t_0, is an arbitrary parameter, uncorrelated with the hadronisation process but when t_0 is increased, the parton shower is terminated earlier and there are fewer partons to hadronise but with higher virtualities [6] that must be capable of producing more hadrons. Thus the hadronisation models should have a parameter t_0 whose effect cancels out when the parton shower model and the hadronisation model are combined. In practice t_0 is tuned such that the final number of hadrons agrees with experimental observation. The favoured value of t_0 tends to be a few times Λ_{QCD}^2. Λ_{QCD} is the momentum scale at which α_s diverges.

Many different hadronisation models exist, the main ones being string models and cluster models. The string model is described here as this is the model used in the production of most of the Monte Carlo samples used in this thesis (Pythia) [11]. In the Lund string model, the partons lose energy to the colour field which collapses into a string-like configuration between them [12]. The string has a uniform energy per unit length, consistent with the linear term of the QCD potential of equation 1.3. The string breaks up into hadron-sized pieces through spontaneous $q\bar{q}$ pair production. This is illustrated in figure 1.4, left, where the incoming partons form strings which then give rise to the hadrons. Recent results from CDF show that, at hadron colliders, the final number of partons is approximately half the number of charged hadrons [13]. The actual result quotes

$$\frac{N_{\pm\mathrm{hadrons}}}{N_{\mathrm{partons}}} = 0.56 \pm 0.1 \tag{1.12}$$

The outcome of the previous discussion is that for any process involving outgoing partons, a large number of hadrons are seen in the detectors because of the fragmentation and hadronisation processes. Due to the higher probability of emitting partons with small angles with respect to the initial parton, most of the final hadrons will be concentrated in a direction close to that of the initial parton. This shower of final state hadrons is what is called a jet.

At hadron colliders the situation is complicated by a number of different phenomena. Unlike at lepton colliders, the total centre of mass energy of the hard interaction, $\hat{s}$, is not known for

[6] The virtuality is defined as how far off the mass shell the particles are. For the case of massless particles this is simply the momentum squared, p^2, of the particle.

each interaction. This is schematically illustrated in figure 1.5. Only one parton (a and b in figure 1.5) from each of the hadrons (h_1 and h_2) participate in the hard-scattering. The partons carry an unknown fraction of the total hadron's energy (f_1 and f_2). The fractional momentum carried by the partons is described by Parton Distribution Functions (PDFs). These are discussed in reference [9]. The remaining partons inside the colliding hadrons participate only minimally in the interaction.

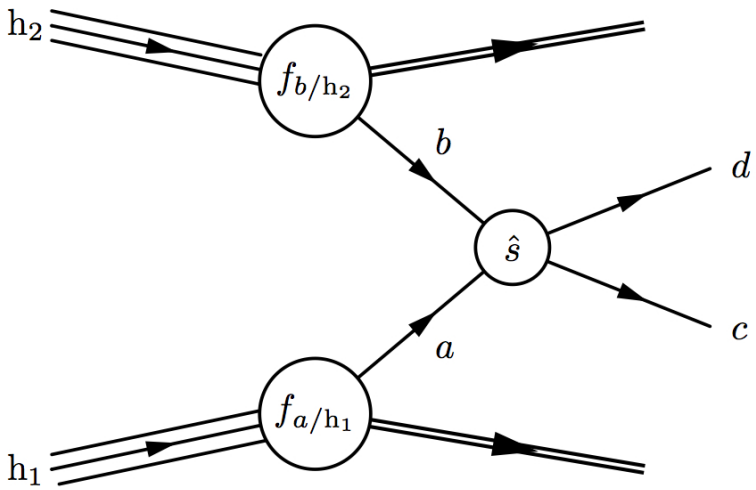

FIGURE 1.5: Schematic drawing of the hard-scattering component of a hadron-hadron collision. Two initial hadrons (h_1 and h_2) are collided but only one parton from each of the hadrons (a and b) participate in the hard interaction. The centre of mass energy of the interaction $\hat{s}$ is therefore unknown. This leads to the fact that the total energy of the outgoing partons (c and d) is also unknown. This figure is reproduced from [9].

Moreover, the presence of a significant amount of soft interactions between the colliding hadrons (beam-beam remnant interactions), possible multiple parton interactions and gluon radiation from the initial partons before the hard interaction (initial state radiation) complicate things even further. These additional processes are collectively known as the underlying event. The underlying event, along with the finite detector resolution, leads to the fact that it is not possible to unambiguously determine, even in Monte Carlo simulation, which hardrons came from which initial parton, i.e. which hadrons belong to which parton jet. Jet algorithms are used to group the final state hadrons into jets whose properties are as close to the initial parton ones as possible. Some common jet algorithms with their benefits and limitations will be discussed later, in chapter 3. It is sufficient at this stage to mention that in this thesis a cone algorithm is used. A jet is therefore defined by a direction, an energy and a cone size that corresponds to the jet opening angle.

The next section describes the particular property of jets that is studied in this thesis, namely jet shapes.

1.3 Jet Shapes

The internal structure of jets is dictated principally by multi-gluon emission from a primary outgoing parton. This process is described by fragmentation models. Multi-gluon emission involves higher order QCD processes which are hard to calculate. Different models are thus used to implement the fragmentation process in the simulation. The implementations must often be

tuned to reproduce the experimental data as accurately as possible. On top of this, a good understanding of the hadronisation process is needed in order to compare simulation results at parton level with what is observed in the detector. Moreover, the underlying event, an important component of any hadronic collision, plays a non-negligible role in the internal structure of jets. A relevant quantity for studying the internal structure of jets is the jet shape as will be discussed shortly but first a definition of what is meant by jet shapes is necessary.

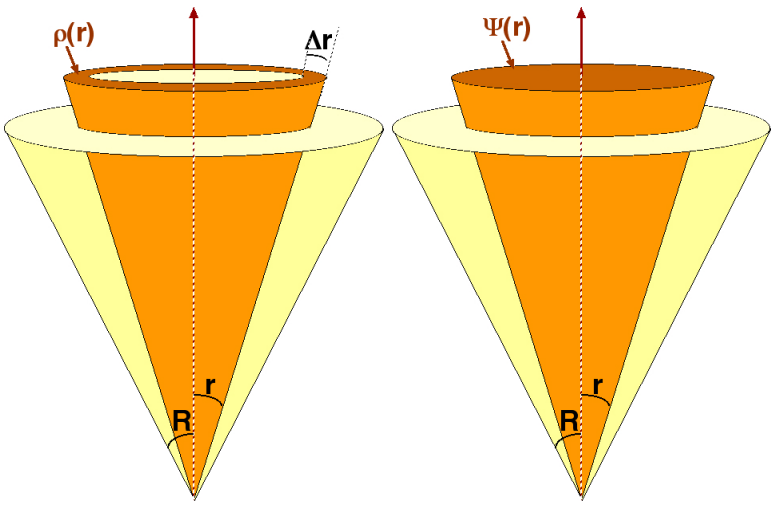

FIGURE 1.6: Schematic drawing of the differential (left) and integrated (right) jet shapes.

Jet shapes look at the fractional transverse momentum (p_T) distribution inside the jets as a function of the distance away from the jet axis. This can be expressed either as a differential or integrated quantity. The differential jet shape is the rate of change of p_T with increasing distance away from the jet axis, whereas the integrated shape is the fractional p_T inside a cone around the jet axis as illustrated in figure 1.6. This quantity can be computed at parton or hadron level in the Monte Carlo (MC) simulations and by using either the tracks or the calorimetric towers at detector level. Formally, the distance away from the jet axis is defined as

$$r = \sqrt{(\Delta\phi)^2 + (\Delta Y)^2} \tag{1.13}$$

where $\Delta\phi$ and ΔY are the angular distances in the (ϕ,Y)-plane between the objects and the jet direction. Y is the rapidity which is given by

$$Y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right), \tag{1.14}$$

and ϕ is the direction in the plane orthogonal to the beam-direction relative to the vertical. The integrated shape is expressed as

$$\Psi(r) = \left\langle \frac{p_T(0,r)}{p_T(0,R)} \right\rangle \tag{1.15}$$

where R is the jet cone radius (c.f. chapter 3) and $p_T(0,r)$ is the sum of the transverse momentum of all objects inside a sub-cone of radius r around the jet axis. Similarly, the differential jet shape is defined as

$$\rho(r) = \frac{\partial\Psi(r)}{\partial r} = \left\langle \lim_{\Delta r\to 0}\left(\frac{p_T(0,r+\Delta r) - p_T(0,r)}{p_T(0,R)\Delta r}\right)\right\rangle \tag{1.16}$$

The integrated shapes are normalised such that $\Psi(r = R) = 1$, i.e. the fractional transverse momentum of the objects inside a cone of radius equal to the jet cone radius around the jet axis is unity. Similarly, the differential shapes are normalised such that $\int_0^R \rho(r)dr = 1$. This requirement again comes from the fact that the fractional transverse momentum of the objects inside the jet cone must be equal to unity.

Jet shapes are sensitive to whether the initial hard-scattered parton was a quark or a gluon. The flavour inclusive jet shapes have been measured at CDF and show that the ratio between the quark- and gluon-jet production cross sections is well reproduced by the Pythia Tune A Monte Carlo simulation [7] [15].

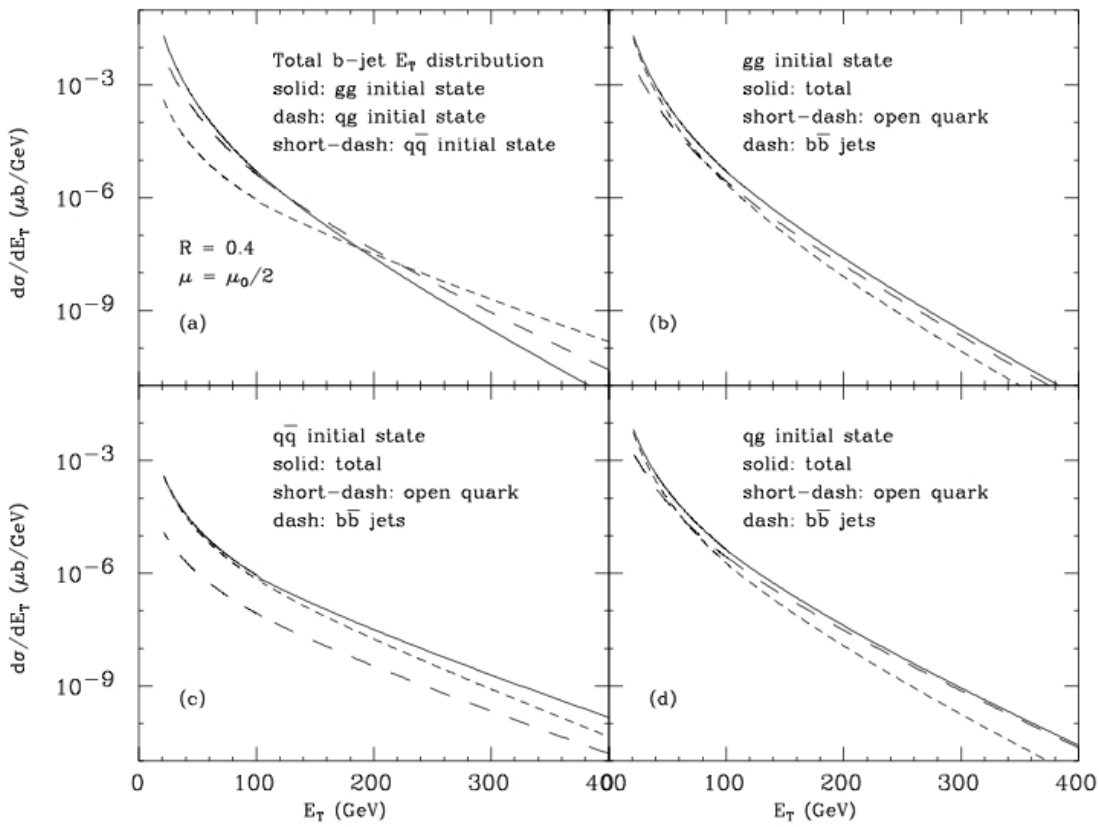

FIGURE 1.7: Differential b-jet cross section for different intial state configurations for a renormalisation/factorisation scale $\mu = \mu_0/2$ and jet cone radius of 0.4 (a). Differential components of the production process: $gg \rightarrow$ b-jet (b), $q\bar{q} \rightarrow$ b-jet (c) and $qg \rightarrow$ b-jet (d).

It is also expected, but has never been measured at CDF, that jet shapes are sensitive to the flavour of the quark jets. In particular it is interesting to look at the shape of b-quark jets where the difference with respect to the inclusive shape is expected to be maximal. Moreover, the shapes of b-quark jets are expected to be sensitive to the relative fraction of gluon splitting and flavour creation events. In the former case, the b and the $\bar{b}$ quarks are expected to be most of the time inside the same jet [16], leading to significantly broader jet shapes than for the latter case. The fraction of gluon splitting events is an important parameter for the tuning of any Monte Carlo simulation. The b-jet cross section p_T dependence is sensitive to the relative fraction of gluon splitting to flavour creation b-jets, which is directly linked to the relative fractions of initial

[7] The parameters of Pythia Tune A were tuned to the CDF Run I underlying event. [14]

state production processes which are gg, qg or $q\bar{q}$ [16]. Figure 1.7, reproduced from [16], shows in the top left hand plot the relative contributions of the different initial state configurations to the differential b-jet cross section. The other three plots (b)-(d) show the differential b-jet cross section for each of the three initial state configurations along with the relative contribution due to jets which contain two b-quarks (called $b\bar{b}$ jets in these plots). These plots were made for a scale $\mu = \mu_0/2$ where $\mu_0 = \sqrt{p_T^2 + m_{\mathrm{b}}^2}$ and for a jet cone algorithm with a cone of size 0.4. It is interesting to note that even though the probability that a gluon jet will split into a $b\bar{b}$ pair grows with jet energy, the fraction of primary gluons in the final state becomes smaller. One of the aims of this analysis is to check if the fraction of b-jets originating from gluon splitting, as well as it's evolution with p_T, is well described in the Monte Carlo models.

It is very important to have a good understanding of b-quark jets because they appear in a number of searches for new physics both at the Tevatron and at future accelerators such as the Large Hadron Collider (LHC). The tuning of the Monte Carlo simulations at the Tevatron is very important for any extrapolation to the LHC energies. Such extrapolations provide a good basis for both qualitative and quantitative tests of the future sensitivity of many new physics channels at the LHC.

In Monte Carlo simulations, b-quark jets are defined as jets which have at least one b-quark inside the jet cone. The measurement presented in this thesis therefore combines jets originating from flavour creation with those where the b-quarks come from gluon splitting. The condition for a jet to be a b-quark jet can be expressed by the condition

$$\Delta R_{\mathrm{b-quark} \rightarrow \mathrm{jet\,axis}} \leq R_{\mathrm{jet\,cone}} \tag{1.17}$$

As mentioned above, the shapes of jets containing one or two b-quarks inside the jet cone are expected to be significantly different. Figure 1.8 shows the hadron level predictions using Pythia Tune A for the integrated b-quark jet shapes in four different p_T bins (see chapter 5 for the definition of the binning). Also shown in these plots are the predictions for inclusive jets as well as single and double b-quark jets. By measuring the shape of the b-quark jets and comparing them to the Monte Carlo predictions, it is possible to verify if the relative fraction of flavour creation versus gluon splitting jets is correctly implemented in the simulation, at least for heavy flavoured jets.

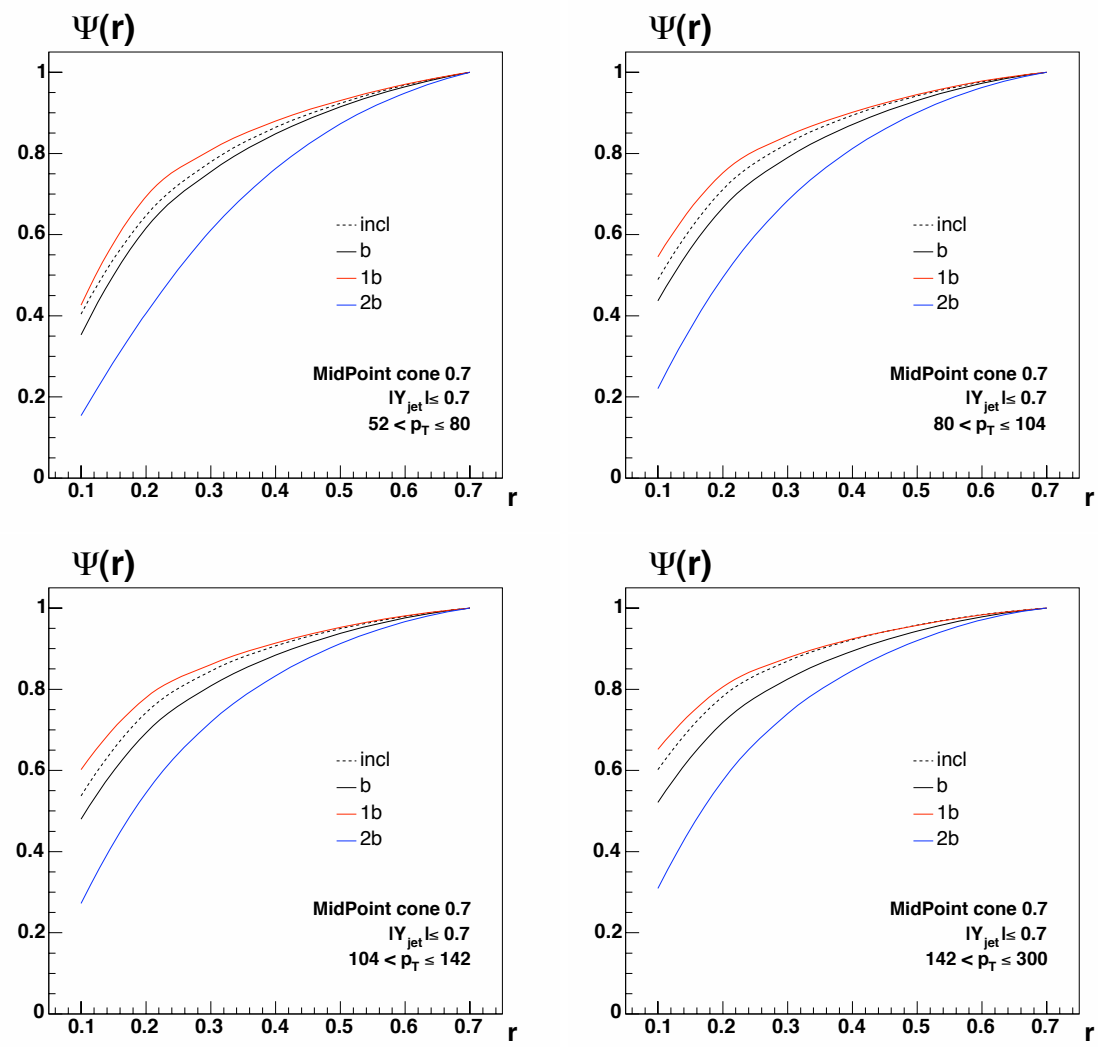

FIGURE 1.8: Hadron level predictions for the integrated jet shapes for b-quark jets and inclusive jets. Also shown are the predictions for single and double b-quark jets.

Chapter 2

The CDF Experiment

The CDF - Collider Detector at Fermilab - experiment is located at one of the two collision points around the Tevatron accelerator at Fermilab. The other Tevatron experiment is called D0. Both CDF and D0 are what are commonly called general purpose detectors.

The results of this analysis were obtained by analysing about $\sim 300\,\mathrm{pb}^{-1}$ of CDF Run II data collected from February 2002 through September 2004 [1].

2.1 The Tevatron

The Tevatron is a synchrotron ring, whose diameter is 2 km. It is designed to accelerate protons and antiprotons using super-conducting magnets with a magnetic field of 4.2 Tesla. Particles are accelerated up to an energy of 1.96 TeV [17]. In each direction, 3 trains of 12 bunches of particles circulate. The particles are made to collide at the two specific points around the ring where the detectors are located. Each bunch of protons and antiprotons contains approximately $2 \cdot 10^{11}$ and $2 \cdot 10^{10}$ particles, respectively. Bunches of particles collide every 396 ns at the centre of the detectors.

The Tevatron is the last accelerator in a large and complex system which not only accelerates the particles to the desired energy but also produces the antiprotons. An overview of the accelerator complex is shown in figure 2.1.

Protons are produced by ionising hydrogen; a dense plasma of H^+ ions is created which collides with a cesium coated cathode, occasionally picking up two electrons from the surface. These H^- ions are separated away and accelerated up to an energy of 750 keV by first a Crockroft-Walton accelerator followed by a Linac which brings the particles to 400 MeV. A thin film of graphite is then used to strip off the electrons before the protons are sent to the Booster, a small 150 m radius synchrotron, where they are accelerated to 8 GeV and divided into bunches. The Main Injector takes these protons and accelerates them further to 150 GeV before they are sent into the Tevatron for their final acceleration to 980 GeV.

To produce the antiprotons, a beam of 120 GeV protons is sent from the Main Injector onto a composite target. The antiprotons are then separated off and collimated into an 8 GeV beam using a lithium lens. The beam is then de-bunched and sent into the Accumulator where it is

[1] The amount of data collected is measured in integrated luminosity, $\int \mathcal{L}(t)dt$. $\mathcal{L}$ is the instantaneous luminosity which is defined as $\mathcal{L} = \frac{n_b^p \cdot n_b^{\bar{p}} \cdot R}{A_{\mathrm{beam}}}$. $n_b^{p,\bar{p}}$ is the number of particles per bunch in the proton and antiproton beams, respectively. R is the bunch crossing rate which is $1/396\,\mathrm{ns}^{-1}$. A_{bean} is the beam-beam cross section.

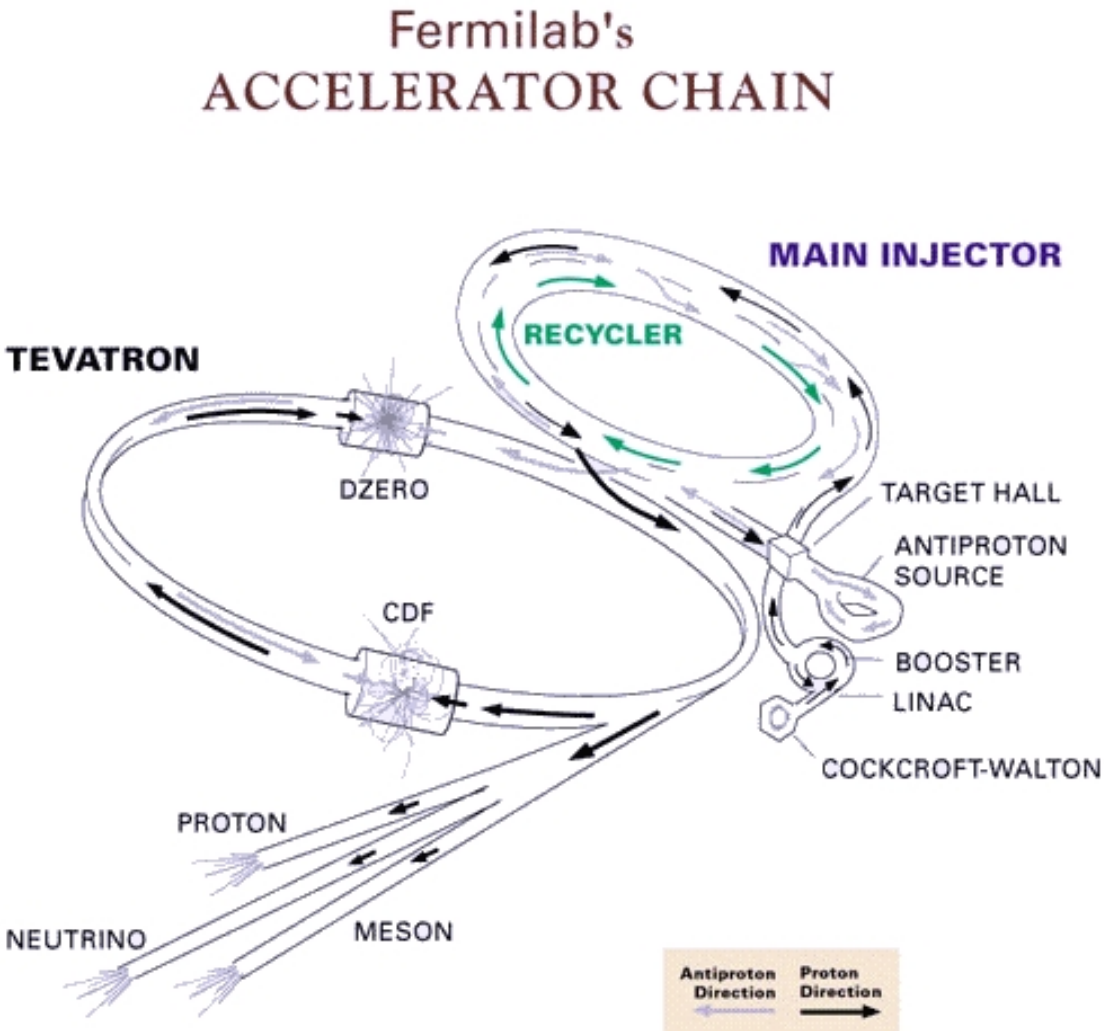

FIGURE 2.1: Schematic view of Fermilab's accelerator complex showing the location of the two experiments as well as the main accelerator facilities.

stochastically cooled to reduce its energy dispersion. When the number of anti-protons is large enough, they are sent into the Recycler ring where further cooling is achieved using stochastic cooling and also recently by using electron cooling. The antiprotons are then sent back via the Main Injector to the Tevatron where they are also accelerated to 980 GeV.

The total integrated luminosity measured at CDF is shown in figure 2.2 from the beginning of Run II which started in April 2001. Also shown is the total integrated luminosity written to tape, i.e. excluding all the detector dead-times. The peak instantaneous luminosity keeps increasing and at the time of writing the record was $\sim 1.70\,10^{32}\,\mathrm{cm}^{-2}$ (January 2006). Figure 2.3 shows the evolution of the instantaneous luminosity since the beginning of Run II.

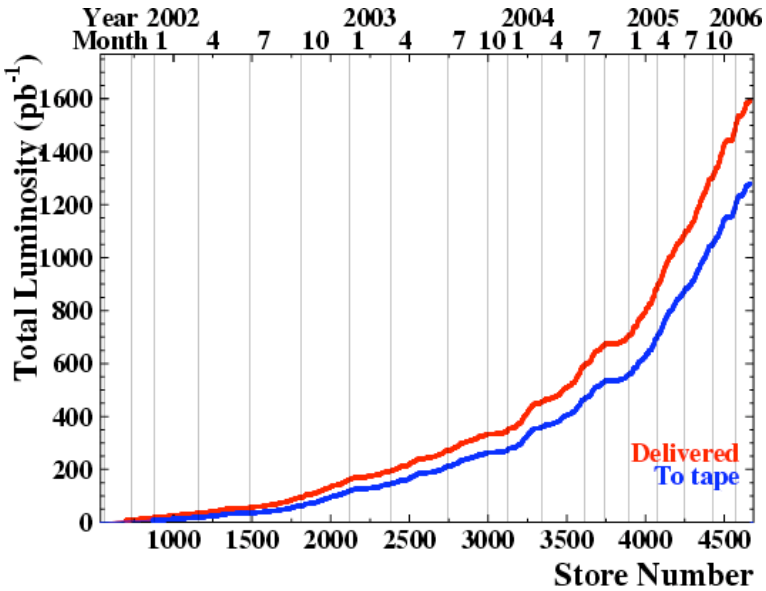

FIGURE 2.2: Total integrated luminosity during Run II operation as measured by CDF (red curve). The blue curve shows the luminosity which was written to tape, i.e. excluding all the detector dead-times.

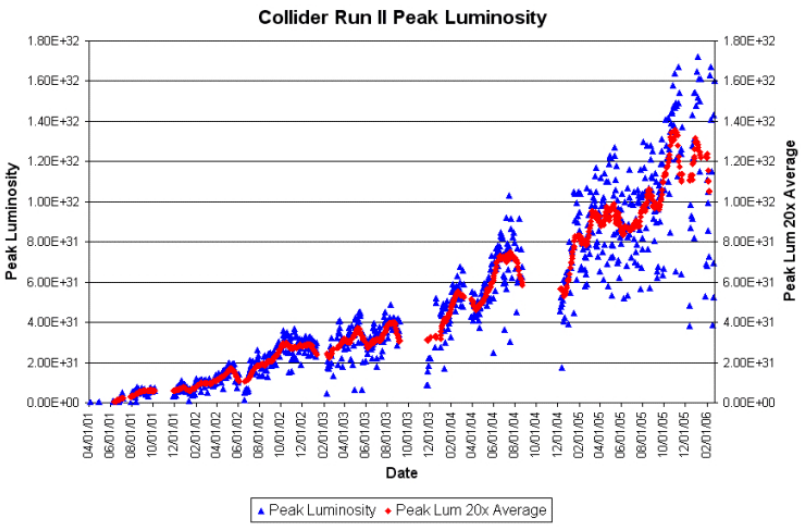

FIGURE 2.3: Instantaneous luminosity as measured by the Fermilab accelerator division.

2.2 The CDF Detector

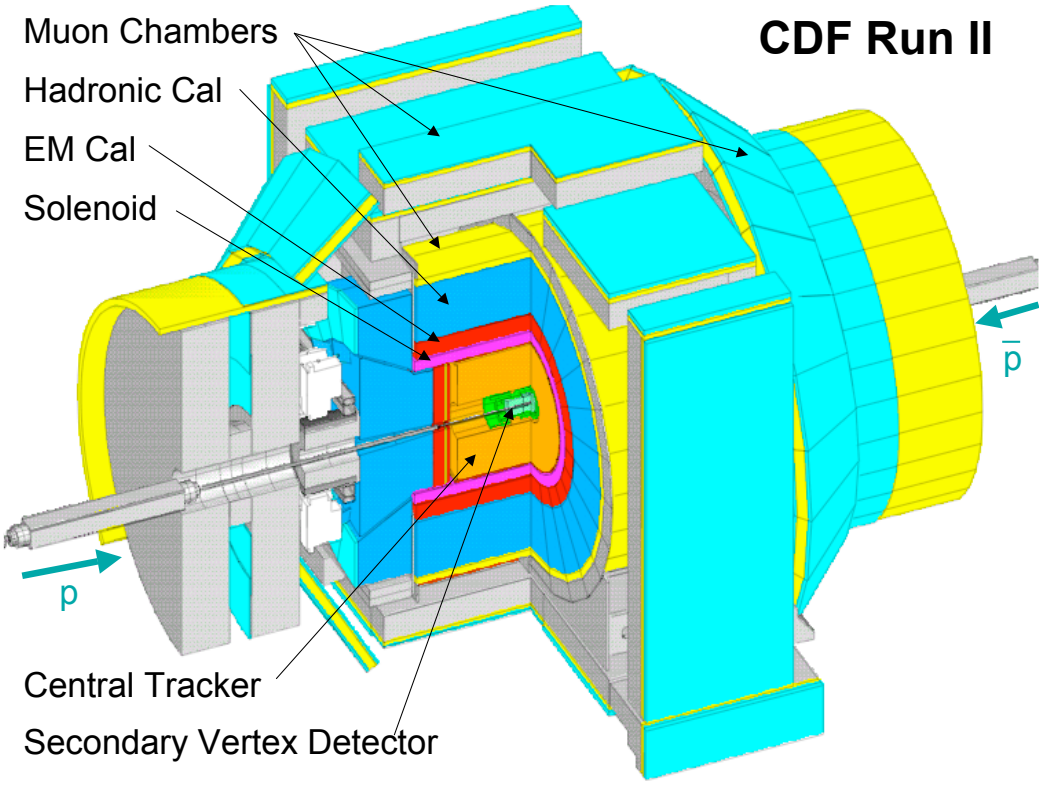

FIGURE 2.4: Schematic drawing of the CDF detector showing the different sub-detectors.

CDF is rotationally symmetric around the beam direction as well as forward-backward symmetric with respect to the most probable beam interaction point. The precision charged particle tracking and the fast calorimetry, as well as the fine grained muon detection were optimised in size and performance to obtain as precise knowledge of the properties of the particles generated in the collisions as possible.

Run II operation started in April 2001, nearly five years after the end of the first run period [18]. The resolution of many of the parts of the detector was significantly increased between these two running periods with both upgraded software and hardware components [19]. The detector had to be able to cope with the higher number of bunch crossings. The centre of mass energy was also slightly increased. During Run I, which started in 1992 and ended in 1996, the bunch crossing rate was 3.5 μs, the centre of mass energy 1.8 TeV.

The CDF detector, a schematic of which is shown in figure 2.4, consists of three distinct spatial regions: the barrel region, where the sub-detectors are concentric around the beam-pipe and of increasing radius, and the two plug regions, where the sub-detectors are parallel to each other in a plane orthogonal to the beam-pipe and of increasing distance away from the most probable interaction point.

The CDF sub-detectors can be split into three different categories: the inner tracking, the calorimeters and the muon chambers. The latter will not be discussed here as they are not used for this analysis. It suffices to say that the muon chambers are tracking devices which detect

the presence and position of the muons. Muons are not stopped by the calorimeters.

2.2.1 Inner Tracking

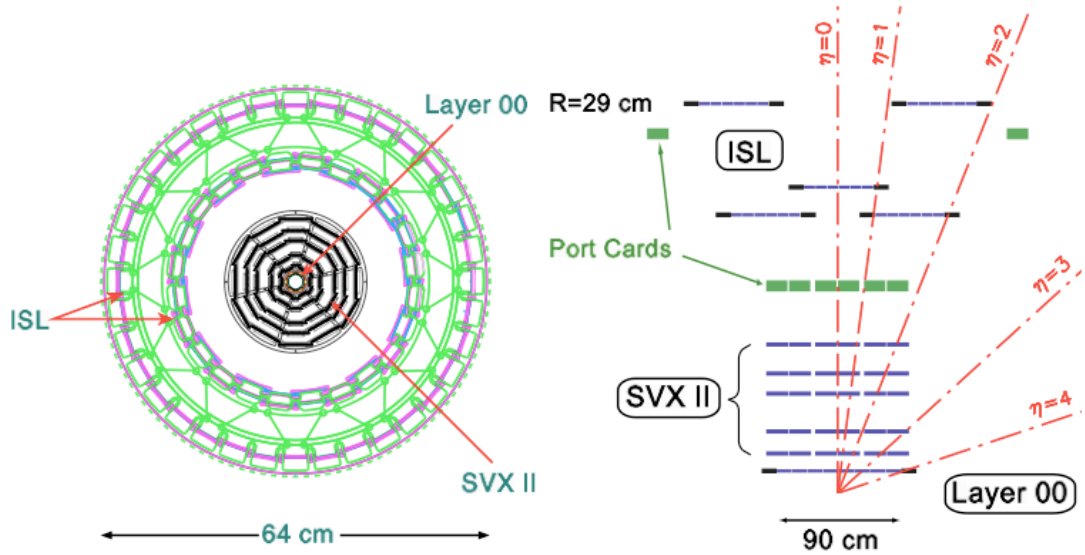

FIGURE 2.5: $r-\phi$ and $r-z$ perspectives of the silicon detectors, showing all the sub-components. Note the compression of the z-scale.

Closest to the beam-pipe are the silicon vertex detectors which greatly enhance the ability to reconstruct vertices that are displaced with respect to the primary interaction. Such vertices, that come from the decay of heavy hadronic states, are called secondary vertices. The silicon detectors are shown in figure 2.5 in both the $r-\phi$ and $r-z$ perspectives. The $722'432$ readout channels of the silicon strip sensors that make up these sub-detectors allow very precise measurements of the position of charged particles close to the beam-pipe. This sub-detector is new for Run II. The silicon tracking system is composed of three different sub-systems. The inner-most one, Layer00, consists of 48 single layered ladders of silicon detectors mounted directly onto the beam-pipe. Then comes the double-sided Silicon Vertex detector (SVXII). This sub-detecctor combines 360 double-sided ladders in a layout of six 15 cm axial sections of twelve 30° ϕ slices and five radial layers. The segmentation in ϕ allows the processing of the silicon data to reconstruct displaced vertices already at Level 2 (see later section on triggers). Finally, outside the SVXII, comes the Intermediate Silicon Layer (ISL). This large radius detector is made up of 296 double-sided ladders. It has a total surface area of $\sim 3.5\,\mathrm{m}^2$ and covers a pseudo-rapidity region up to $|\eta| = 2$ [2]. The silicon layer closest to the beam-pipe is a mere 1.6 cm from the centre of the beam-pipe, while the outermost layer is 28 cm away. The silicon system, taken as a whole, gives 3D information about the location of the tracks. The resolution in ϕ is smaller than $30\,\mu$m and smaller than $60\,\mu$m in the z-direction (along the beam-pipe).

Outside the silicon detectors is the Central Outer Tracker (COT). The purpose of this detector is to provide an accurate position reconstruction of all charged particles that go through the detector. It is an open cell drift chamber with a maximum drift time of ~ 100 ns. It is made of 96 layers of parallel arrays of wires. The operation principle of a wire chamber is that when a

[2] The pseudo-rapidity, η, is the rapidity in the limit where the momentum of the particles is much higher than their masses ($p \ll m$). It is defined as $\eta = -\ln(\tan(\theta/2))$ where θ is the angle to the beam-pipe direction.

charged particle goes through the gas, it ionises some of the particles that are then attracted to the nearest charged wires, thus creating a measurable electric signal on these wires from which the spatial position as well as some timing information can be extracted. The hit resolution for the COT is smaller than 180 μm. It also provides some dE/dx information which can be used for particle identification.

The whole tracking system is inside a super-conducting solenoid magnet which provides an axial magnetic field of 1.4 Tesla. The bending of the trajectories of charged particles in the magnetic field gives information about their sign and momentum. The design momentum resolution of the tracking system combining the silicon detectors and the COT is $\delta p_T/p_T^2 = 0.1\%$ for tracks with $p_T > 4\,\text{GeV}$ and $\delta p_T/p_T^2 = 0.6\%$ for tracks with $p_T \sim 0.5\,\text{GeV}$. Between the COT and the solenoid is the new scintillator based Time Of Flight (TOF) detector that is not used for this analysis.

The extrapolation of the tracks, combining information from all the tracking systems, back to distances smaller than the beam-pipe radius, allows the reconstruction of secondary vertices. This secondary vertex reconstruction is what makes it possible to detect some hadronic states containing heavy flavour quarks. Heavy-flavour hadrons decay into lighter particles most of the time before they leave the beam-pipe. Tracks reconstructed inside the detector are extrapolated back to their most probable location inside the beam-pipe. Most of the tracks will have originated from the primary interaction, the primary vertex. Yet some of the tracks might intersect a short distance away from it, leading to believe that they originated from the decay of a heavy-flavoured particle. The secondary vertex spatial precision at CDF is currently approximately 40 μm, including the 30 μm contribution from the accuracy of the knowledge of the beam-line position [20]. The secondary vertex reconstruction algorithms will be discussed in detail in chapter 4.

2.2.2 Calorimeters

The principle of calorimetry at collider detectors is that particles entering the calorimeters interact with the material, depositing their energy into it. The calorimeters use heavy absorber material, the passive layers, interleaved with scintillating material, the active layers. By measuring the energy deposited in the sensitive regions of the detector by the showers initiated by these particles, it is possible to calculate the energy of the incident particle. Unlike the tracking system, calorimeters are sensitive to neutral particles so the energy deposited in the calorimeters, if they were 100% hermetic, would be the total energy of the interaction. To get the best possible energy resolution, it is important not only to make the detectors as hermetic as possible but also to limit the energy lost by the particles, through interactions with the material, before they reach the calorimeters. This is the reason why particular care is needed in the development and characterisation of the tracking detectors and their support structures.

The CDF calorimeters are separated into two distinct sub-detectors. The inner one, which uses lead as an absorber material, called the electromagnetic calorimeter, absorbs the energy from most electromagnetic particles as well as pions. The outer one, which uses iron as an absorber material, the hadronic calorimeter, stops all other strongly interacting particles, mostly the hadrons. There are both barrel ($|\eta| < 1.1$) and plug ($1.1 \leq |\eta| < 3.6$) calorimeters. Their concepts are essentially the same but the plug calorimeters need to be more resistant to radiation as the particle flux is much larger at small angles with respect to the beam-pipe. Both electromagnetic and hadronic calorimeters are segmented into towers in directions approximately aligned with respect to the centre of the detector. The segmentation in ϕ and η is 15° and 0.11, respectively, for the barrel region and $7.5^o \rightarrow 15^o$ and $0.1 \rightarrow 0.6$, respectively, for the plug

regions. This segmentation allows for position reconstruction of the energy deposits in these sub-detectors. The resolution obtained in a test-beam (for single electrons and single pions) are $\frac{\Delta E}{E} = \frac{16\%}{\sqrt{E}} \oplus 1\%$ for the central electromagnetic calorimeter and $\frac{\Delta E}{E} = \frac{80\%}{\sqrt{E}} \oplus 5\%$ for hadrons. The energy resolution of the calorimeters is very important for jet reconstruction (see chapter 3).

In the central region there are also two position detectors: the central pre-radiator gas chamber (CPR) just before the electromagnetic calorimeter and the shower maximum detector (CES) inside it. Neither of these detectors is used in this analysis.

2.2.3 Trigger System

The total interaction rate is very high, too high to be able to write all the information from every collision to tape. On top of this, most interactions do not have any interesting physics in them. In order to extract only the potentially interesting events and to keep the rate of events written to tape at a manageable level, a sophisticated trigger system is required. At CDF, the maximum rate for recording events to tape is 75 Hz while the Tevatron collision rate is 2.5 MHz. The trigger is organised into three levels.

The Level 1 (L1) trigger uses basic information from the outer tracker, individual calorimeter towers and muon chambers and reconstructs rough tracks. This trigger has almost no dead-time. It possesses a 42 stage pipeline and can make a trigger decision every 132 ns with a total latency time of 5544 ns.

At Level 2 (L2), the trigger adds information to the L1 objects and does limited event reconstruction. At L2 the events are processed in programmable processors and the event rate is reduced to ~ 300 Hz.

Level 3 (L3) takes events that pass L2 and processes them using an almost complete event reconstruction, using a network of parallel processors to reduce the rate to ~ 50 Hz.

At the time of writing, the trigger tables have been re-written to be able to make the best use of the high instantaneous luminosities achieved by the Tevatron. The QCD jet triggers used for this analysis and discussed in more detail in section 5.1, have not been significantly affected by this (other than by the use of different pre-scales [3]).

[3] Pre-scales are used to limit the number of events read out. Only one in every so many events which pass the trigger cuts is read out and written to tape. Pre-scales are usually constant values but there also exist dynamic pre-scales which fix the event read-out rate to a given value, say 1 Hz. The pre-scale thus depends on the luminosity.

Chapter 3

Jet Algorithms

A very important part of many QCD analyses is the ability to reconstruct jets. Some particles, such as charged leptons, or photons, go through the detector without decaying or emitting any other particles. This results in a very clear signature in the detector: a single track in the tracker or no track, for the cases of charged and neutral particles, respectively, and a well-localised energy deposit in the calorimeters. Hadronic particles, on the other hand, are more complicated and complex to study. Because of the nature of the strong interaction, a single hard interaction involving a quark or a gluon will result in a large number of hadrons seen in the detector, as previously discussed in chapter 1.2. This evolution from a single particle to a large number of final state particles is governed by QCD processes, in particular by fragmentation processes (radiation of other particles from the primary one) as well as hadronisation (the process by which quarks and gluons are bound into hadrons). The signature of a hadronic interaction in our detector is a large number of tracks seemingly close together in (real or phase) space with an energy deposition in the calorimeters spread over a similar area as the tracks. This shower of observed particles is what is referred to as a jet [21].

At hadron colliders the definition of jets is not straightforward. Because of the possibility of having gluon-mediated interactions, which by far dominate the interaction cross section, there is a possibility of having many jets in the event. This can lead to the shower of particles from one parton overlapping with those from another. There are also many more processes happening, when protons and anti-protons collide, than just the hard interaction. These other processes are what is referred to as the underlying event [14]. All of these effects result in a large number of tracks and calorimeter energy deposits which must be disentangled into objects such as jets or isolated particles. This disentangling is done using jet algorithms.

There are two main classes of jet algorithms which have been studied and used so far at CDF. The first class of algorithms, the most widely used, are called cone algorithms. They look at the energy deposited in the towers of the calorimeter and define all particles within a physical cone of fixed size to belong to the same jet. Where the different algorithms differ is in the method used to find the final cones, in the size of the cones, or in the way of dealing with cones which overlap. This is the class of algorithms which is discussed below. The second class of algorithms, called the k_T algorithms, define jets in terms of towers which are not close to each other from a spatial point of view but are close to each other in k_T [22][23]. k_T is proportional to the relative transverse momentum between two particles. This class of algorithms will not be discussed further but more information about this algorithm can be found in the references mentioned above.

3.1 Cone Algorithms

Figure 3.1 (top left) shows a schematic picture of a jet containing tracks and calorimeter energy deposits. Calorimeter deposits are represented as red towers, tracks are red lines and the jet is represented by everything which is inside the yellow cone.

All current cone algorithms are based on the so-called "Snowmass Algorithm" which defines both the stability conditions and the properties of the jets [21]. The stability condition is usually defined as the requirement that the change of the jet properties from small variations in the location of the centre of the jet should be minimal.

The properties of a cone algorithm jet are

- Cone size, R (usually 0.4 or 0.7 but sometimes 1.0)
- Energy (E) and momentum (p)
- Direction (Y, ϕ)

The cone size defines the spatial extent (in rapidity, Y, and ϕ) of the jet i.e. the region covered by towers considered to be inside the jet. To be more precise the cone size is defined as

$$R = \sqrt{\Delta\phi^2 + \Delta Y^2} \tag{3.1}$$

which is the angular separation in (ϕ, Y)-space between the jet direction and the outer edge of the cone.

Each cone algorithm defines differently the relationship between the kinematic quantities of the jet and those of each of the towers which belong to a jet. For example MidPoint, the algorithm most used at CDF by the QCD group, defines the jet quantities by

$$P^{\text{jet}} = \sum P^{\text{tow}} \tag{3.2}$$

where P^{jet} is the 4-vector momentum (p_x,p_y,p_z,E) of the jet and P^{tow} is that of the towers [24]. The sum is over all towers inside the jet. This defines the energy and direction of the jet.

JetClu, another cone algorithm used at CDF, mostly for Run I, defines the jet energy as

$$E_T^{\text{jet}} = \sum E_T^{\text{tow}} \tag{3.3}$$

and the jet direction as

$$\eta^{\text{jet}} = \frac{\sum \eta^{\text{tow}} E_T^{\text{tow}}}{E_T^{\text{jet}}} \tag{3.4}$$

and

$$\phi^{\text{jet}} = \frac{\sum \phi^{\text{tow}} E_T^{\text{tow}}}{E_T^{\text{jet}}} \tag{3.5}$$

where η is the pseudo-rapidity.

In an ideal world, to find the final jets, each tower of each event should be considered as a potential centre of a jet. The algorithm would iterate over each possible jet until the local minimum of the jet potential would be found. The minimum of the potential is the point where the change in the kinematic jet quantities (4-vector moment for MidPoint) used by the algorithm, when moving the centre of the jet in space, is the smallest. These local minima would then be

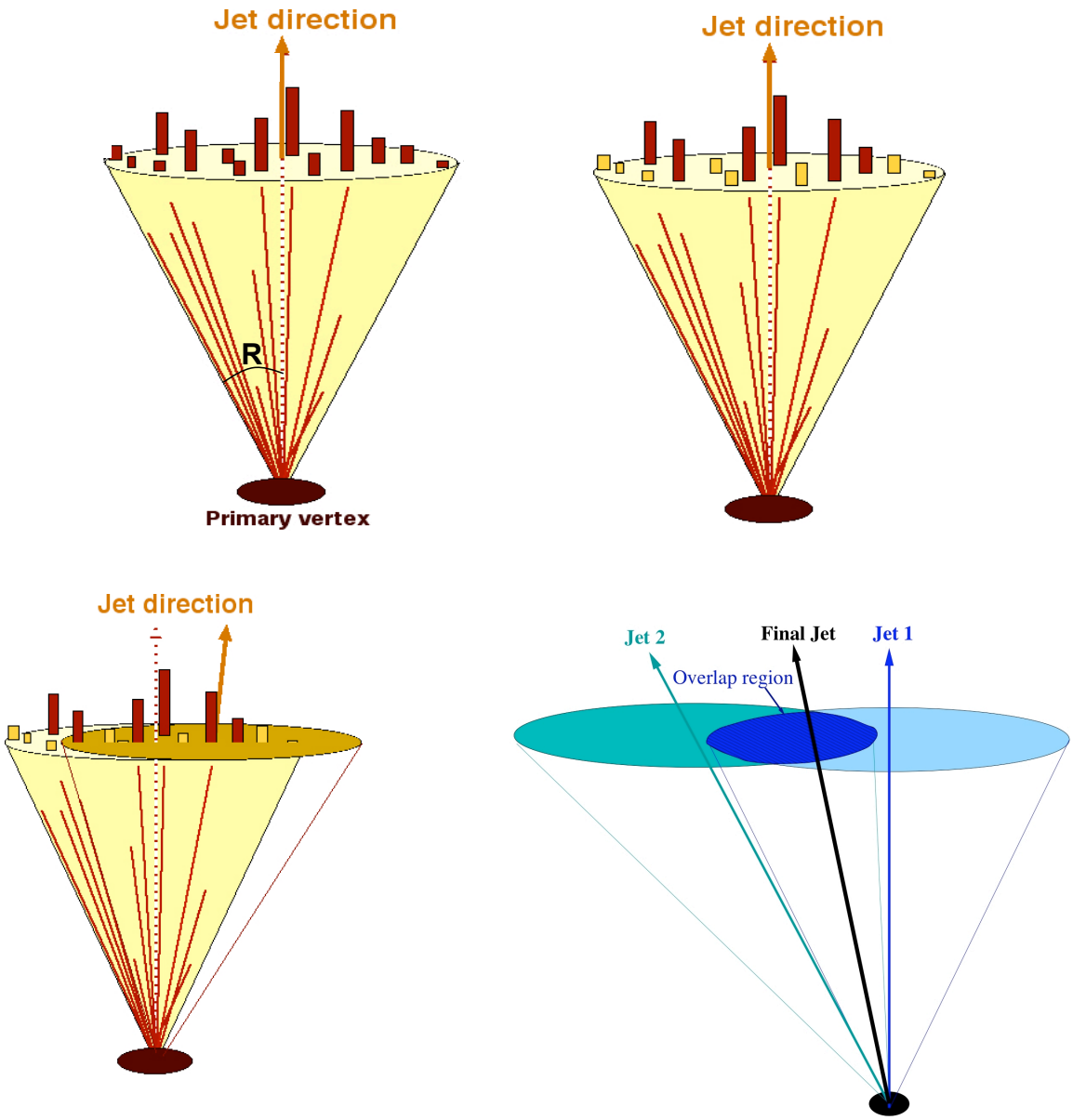

FIGURE 3.1: Schematic drawing of a cone algorithm jet showing the tracks and towers inside the jet cone (top left) and illustration of the seed mechanism (top right), ratcheting process (bottom left) and splitting/merging procedure when the jets are merged (bottom right).

the centres of the jets.

In the real world there are a number of compromises which must be made in order to cope with limited computing power and high trigger decision rates. The main compromises will be discussed below. These are

- the use of seed towers
- the notion of ratcheting (only present in JetClu)
- the merging/splitting procedure when two jets are closer than twice the cone size from each other

It is in these compromises, as well as in the choice of the size of the jet cone, that the different cone algorithms differ. Table 3.1 summarises all the different properties and compromises of the two cone algorithms used at CDF: JetClu and MidPoint.

Property	MidPoint	JetClu
Jet quantities defined by	$P^{\text{jet}} = \sum P^{\text{tow}}$	$E_T^{\text{jet}} = \sum E_T^{\text{tow}}$ $\eta^{\text{jet}} = \frac{\sum \eta^{\text{tow}} E_T^{\text{tow}}}{E_T^{\text{jet}}}$ $\phi^{\text{jet}} = \frac{\sum \phi^{\text{tow}} E_T^{\text{tow}}}{E_T^{\text{jet}}}$
Cone sizes used	0.4, 0.7 or 1.0	
Seed towers (GeV)	1.0	
Ratcheting	none	present
f_{merge}	75% (50% old)	75%
Iterative merging	yes	no
Infrared and collinear safe	yes	no

TABLE 3.1: Comparison between the JetClu and MidPoint cone algorithms.

Seed towers are defined as towers with a higher threshold (usually set to 1 GeV) than the threshold to be included in a jet (usually set to 0.1 GeV). Only the seed towers are considered as potential centres for the jets as illustrated in figure 3.1 (top right). This greatly reduces the time needed to find stable cones. There are a number of problems with the use of seed towers, some of which are discussed below.

Ratcheting is a compromise used in the JetClu algorithm only. It constrains all seed towers initially inside a jet to remain inside a jet even if they are outside the final jet cone. This leads to jets which can have strange shapes and high amounts of transverse energy outside the cone radius, see figure 3.1 (bottom left). There is no reasonable way to simulate the role of ratcheting in theoretical calculations, since its role depends in detail on the level of soft radiation in the event as well as on the details of the calorimeter. MidPoint, on the other hand, is infrared and collinear safe [24]. Infrared safe means that the jet configuration is stable with respect to the emission of a soft parton. Collinear safe means that the jet configuration is stable with respect to the emission of a collinear parton.

The final compromise mentioned here is the merging/splitting procedure. It is possible, because of the use of seed towers, that, in the final state, two or more jets can be separated by less than twice their cone size, see figure 3.1 (bottom right). In this case there are a number of towers which are inside both jet cones, something which would have no physical meaning. In both cone algorithms described here, this problem is dealt with by introducing a merging/splitting fraction, f_{merge}. If the fraction of energy of a jet which overlaps with another jet is larger than

this fraction f_{merge}, the jets are merged together into a single jet. All the towers in the initial jets then belong to the new jet and the kinematic quantities of the jet are re-calculated from this new set of towers. If this overlap fraction is smaller than f_{merge}, the towers in the overlap region are individually associated with the jet whose centre is closest and again the kinematic quantities are re-calculated. This fraction f_{merge} can be set to different values, the default for both JetClu and MidPoint is 75%. Another difference between these jet algorithms is the number of times this merging/splitting procedure may occur. For JetClu the number of iterations is limited to one, no more than two jets may be merged together, leading to a maximum distance between the outermost towers and the jet axis of just less than twice the cone size. In the case of MidPoint this procedure is iterative and starts with the highest transverse momentum jet. This can lead to four or five jets being merged together into a single jet which then covers a large fraction of the detector (with the final ΔR sometimes as large as 2.5). The f_{merge} fraction is chosen to be relatively large (used to be 50% and is now 75%) to minimise this effect.

The choice of optimum cone size is not unique but depends on the process under investigation. If the main use of jets in the analysis is simply to count their number above a certain transverse energy/momentum threshold then a cone size of 0.4 could be the best choice; this would be the case for many searches for new particles. Or if one is interested in a jet veto, i.e. only events without jets are of interest, then again a cone of 0.4 would be most appropriate. If one is interested in getting the highest possible accuracy in terms of energy/momentum reconstruction of the initial partons, such as in measurements of jet cross sections, or when investigating internal properties of jets, such as jet shapes, then a larger cone size is most often used. This ensures the least amount of energy is left out from the jet which initially came from the parton whose properties are investigated. The cone should not be too large either so as not to bring in too much of the underlying event. For this reason most QCD studies use a cone size of 0.7 but some use a cone of 1.0.

It should be noted that jet algorithms can be applied at detector level either to tracks or to calorimeter towers. For Monte Carlo simulations, algorithms can also be run on the final state particles or even the partons. The algorithms remain the same for all the different types of objects considered. For simplicity the term towers was used but it is important to remember that algorithms and jets are not only limited to these objects. Towers are the default objects currently used by all CDF analyses for the reconstruction of jets.

The JetClu algorithm has so far been the preferred jet reconstruction algorithm at CDF Run II principally because of the desire for backward compatibility with Run I results. The Run II workshop proceedings suggest the use of the MidPoint algorithm for both the CDF and D0 experiments at Fermilab for the reasons mentioned above and in the following paragraph. These are described in more detail in reference [25].

To illustrate one of the problems with jet algorithms and JetClu in particular, we consider the situation where there are two localised clusters of towers (or two tracks) that are separated by just less than twice the cone size [24]. This situation is illustrated in the left hand diagram of figure 3.2. Seeds would be placed at each of these clusters and stable cones would be found centred approximately on each of the clusters, leading to two distinct jets. At parton level, this topology could come from one parton with high momentum which splits into two roughly equal energy partons, with a fairly large angular separation, before hadronisation occurs, in which case the ideal jet solution would be one where both these clusters are included in a single jet centred roughly halfway between them. The MidPoint algorithm partially solves this problem by artificially placing a seed tower exactly half way between all of the seed towers (hence the name MidPoint). This is shown on the right hand diagram of figure 3.2. In fact it turns out that

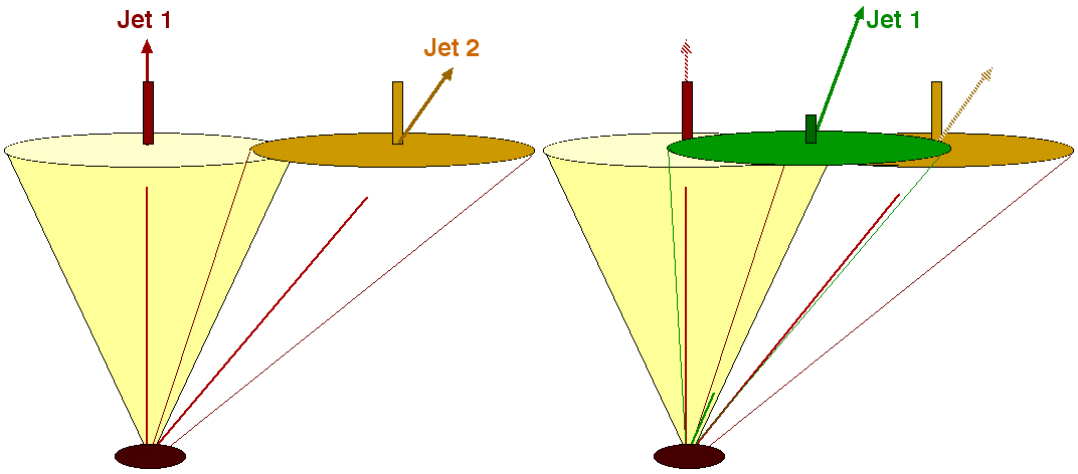

FIGURE 3.2: Schematic drawing of the jet reconstruction of a two parton event, where the partons are separated by just less than twice the cone radius. In JetClu the event would be reconstructed as two jets (left) whereas in MidPoint the use of the extra seed leads to one single jet (right). JetClu is not Infrared safe because the emission of a soft parton between the two hard partons (right) yields a different jet configuration: the same as the default MidPoint configuration. MidPoint on the other hand is infrared safe.

in practise it is only necessary to place a midpoint seed between seeds separated by less than twice the cone radius. In MidPoint, the stable configuration would be one in which the centre of the jet is roughly half-way between the two partons. Now let's add a very soft gluon emitted somewhere between these two partons and see what happens. The configuration is now one with three seeds centred on each of the partons. JetClu will reconstruct this configuration as one jet, centred half-way between the two high p_T partons, just as MidPoint did for the case without the soft gluon. MidPoint will reconstruct exactly the same as what it reconstructs for the leading order process. From this, one can see that JetClu is not infrared safe where as MidPoint is. In fact there is one small subtlety which is that experimentally, the algorithm most likely doesn't see the soft gluon; the p_T of the soft gluon is most probably much smaller than the threshold used by the algorithm for towers to be included in the jet. So in that case JetClu would still reconstruct two jets as the third calorimetric cluster will be very small. The jets will still be reconstructed differently at hadron and calorimeter level. Of course this still doesn't answer the question of which of the configurations, two jets centred each on one of the hard partons or a single jet centred half-way between the two, is the "correct answer". There is no correct answer because there is no way to know, experimentally, if these partons should be considered as a single gluon jet or as two parton jets! What is important is that the same results are obtained both at parton and hadron level. It is also important to use infrared safe algorithms so that fixed order perturbative QCD calculations yield finite results.

Chapter 4

Secondary Vertex Tagging at CDF

There are a number of different ways to enhance the heavy flavour content of data samples at CDF. One of these methods is called the soft lepton tagger. It identifies low momentum electrons or muons inside jets coming preferentially from semi-leptonic decays of B-hadrons [1]. The limitations of this method are mainly due to the low semi-leptonic branching ratio which is of the order of 10% [2] and the wrong identification of leptons (lepton fakes). This tagger is of limited use for this analysis because of the low branching ratio.

The tagging algorithm used for the analysis presented in this thesis exploits the long life-time of the B-hadrons, which is of the order of 1.5 ps (which corresponds to a proper decay length, $c\tau$, of 450 μm). Considering the large relativistic boost of the events, B-hadrons travel a few millimetres before decaying. The decay products of the B-hadron originate from the location of the decay, the secondary vertex, and not from the primary vertex like all other particles in the event. The tagging algorithm is thus based on the reconstruction of secondary vertices inside jets. It is called SecVtx. The main limitation of this tagging method is the fact that hadrons that contain a c-quark can have decay lengths, in the detector, similar to those of B-hadrons. Another limitation is the finite resolution of both the primary and secondary vertex locations. The way secondary vertices are reconstructed using SecVtx is discussed below, followed by a description of a study carried out by a colleague from the University of Geneva and myself which shows that it is possible to apply this algorithm to jets reconstructed using any of the different jet cone algorithms currently used at CDF and with any cone size.

4.1 The SecVtx Algorithm

The SecVtx algorithm is based on the reconstruction of secondary vertices inside a jet as illustrated in figure 4.1. It relies on a cut on the significance of the 2-dimensional projection along the jet axis of the distance between the primary and secondary vertices. To reconstruct a secondary vertex, tracks which satisfy

$$\Delta R_{\mathrm{sub-cone}} = \sqrt{(\eta_{\mathrm{track}} - \eta_{\mathrm{jet}})^2 + (\phi_{\mathrm{tracks}} - \phi_{\mathrm{jet}})^2} \leq 0.4 \tag{4.1}$$

are considered. This cone of size 0.4 inside which tracks are taken is generally referred to as the sub-cone size. The sub-cone size is kept at 0.4 despite the fact that jets can be reconstructed with cones of size 0.4, 0.7 or 1.0. This can be justified by the fact that the heavy flavour hadron will most of the time be close to the jet axis and it's decay products should be close to the direction of the initial hadron. Using softer tracks further away from the jet axis might

[1] A B-hadron is a hadron which contains a b-quark.
[2] This is the fraction of B-hadron decays where the W-boson decays to a lepton and a neutrino.

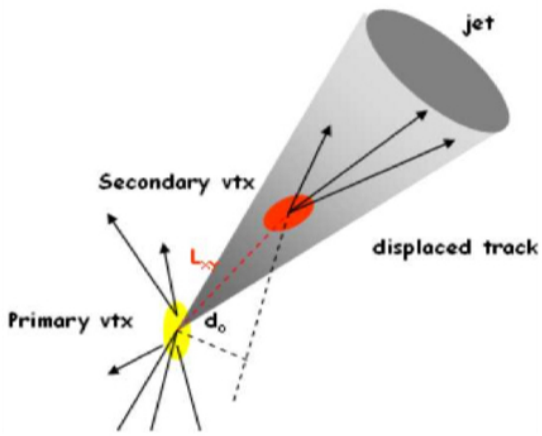

FIGURE 4.1: Schematic representation of a secondary vertex inside a jet.

also lead to an increase in fake secondary vertices. This choice is further justified in section 4.2.2.

In order to reduce the number of tracks which are not well measured, a number of track quality cuts are applied. These are [26]

- $r - \phi$ hits in at least 3 SVX layers
- χ^2 per degree of freedom less than 8.0
- $p_T > 0.5\,\mathrm{GeV}$
- d_0 with respect to the primary vertex $< 0.15\,\mathrm{cm}$
- $\delta(Z_0)$ with respect to the primary vertex $< 2\,\mathrm{cm}$
- K_s and Λ removal
- Conversion track removal

d_0 is the distance of closest approach of the track with respect to the primary vertex, as illustrated in figure 4.1. $\delta(Z_0)$ is the distance along the z-axis between the z-projection of the point of closest approach and the z-position of the primary vertex. The χ^2 of a track is the sum of the squares of the deviations between the location of the silicon hits and the fitted track. The number of degrees of freedom is the number of parameters used when fitting the track to the hits in the silicon. The cuts on the d_0 and the $\delta(Z_0)$ are applied to reduce the effects of interactions in the detector material. Tracks which pass these basic selection criteria are labelled as being *good tracks*.

Having selected good tracks, an attempt at reconstructing a secondary vertex is made using a set of criteria shown in table 4.1. The first set of iterations are called Pass 1. If at the end of the Pass 1 iterations no secondary vertex is found, Pass 2 is attempted with slightly different selection criteria. After ordering tracks according to their quality [3], a secondary vertex is reconstructed, both for Pass 1 and Pass 2, by starting with the best displaced track, attaching the next track and trying to reconstruct a secondary vertex. If the χ^2 is smaller than a predefined

[3] The quality of a track in this case is based on its number of hits in the silicon vertex detector, its p_T and its d_0 significance.

seed vertex χ^2 (50 for Pass 1, no limit for Pass 2) we have a secondary vertex seed. Next all tracks whose d_0 significance with respect to the seed vertex are smaller than the predefined attachment significance (4.0 for Pass 1, no limit for Pass 2) are attached to the secondary vertex. If any track contributes to the total vertex χ^2 more than the track prune χ^2 (45 and 30 for Pass 1 and Pass 2, respectively), it is removed. The final vertex requirements are that the total vertex χ^2 is smaller than a predefined limit (50 for both Pass 1 and Pass 2), that there must be at least one track in the vertex with p_T larger than predefined limit (1.0 and 1.5 for Pass 1 and Pass 2, respectively) and that the $L_{\rm xy}$ significance must be above 7.5 for both Pass 1 and Pass 2. The $L_{\rm xy}$ is the projection onto the jet axis of the 2-dimensional spatial separation between primary and secondary vertices. The $L_{\rm xy}$ significance is defined as $\frac{L_{\rm xy}}{\sigma_{L_{\rm xy}}}$, where $\sigma_{L_{\rm xy}}$ is the error on the $L_{\rm xy}$ position. If all these requirements are fulfilled, the jet is declared to be tagged. If the $L_{\rm xy}$ is negative, as illustrated in the right of figure 4.2, the jet is called a negatively tagged jet. The fraction of negatively tagged jets gives an indication of the number of fake positively tagged secondary vertices. A positively tagged jet is schematically shown in figure 4.2 (left).

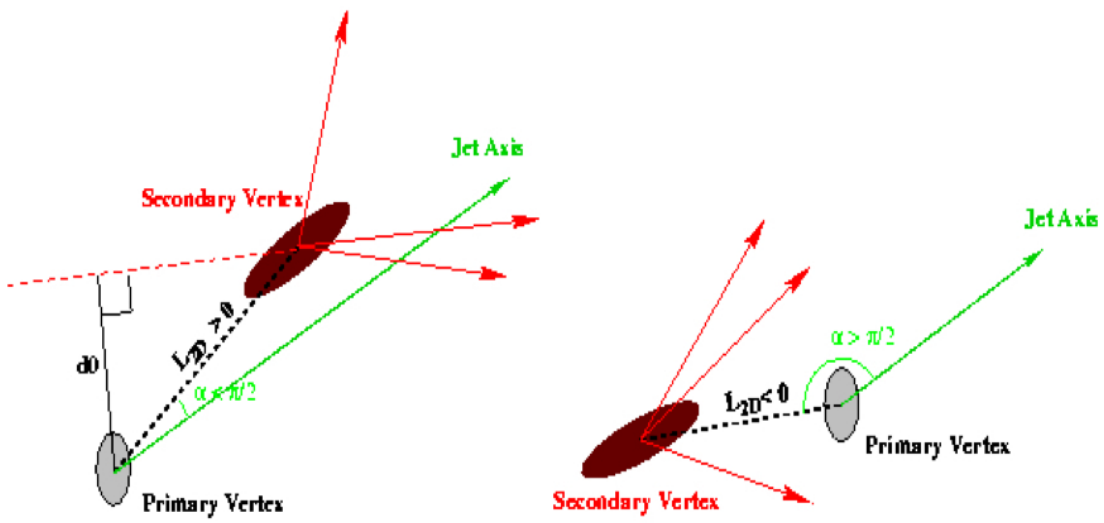

FIGURE 4.2: Schematic representation of a jet positively and negatively tagged by SecVtx. $L_{\rm 2D}$ is the same as the $L_{\rm xy}$ in the text.

All the cut parameters both for track quality requirements and secondary vertex cuts were optimised for the current analysis version [26]. The optimisation was done in such a way as to maximise the efficiency while keeping a reasonably low fraction of negative tags.

It should also be noted that there exists another SecVtx tagging algorithm called the loose tagger which has a higher efficiency but at the cost of a lower purity. Because the purity of the sample is already a limiting factor for the analysis presented in this thesis, the tight tagger is used.

4.2 Applying SecVtx to Different Jet Algorithms

The JetClu algorithm has so far been the preferred jet reconstruction algorithm at CDF Run II principally because of the need for backward compatibility with Run I results. The Run II workshop proceedings suggest the use of the MidPoint algorithm for both the CDF and D0

	Pass 1	Pass 2
No. SVX r-ϕ layers	3	
Track χ^2/ nDOF	8.0	
Track p_T (GeV)	0.5	1.0
Track d_0 significance	2.0	3.5
Track $\delta(Z_0)$ (cm)	2.0	
Track d_0 cm	0.15	
Attachment significance	4.0	
Highest track p_T	1.0	1.5
Track prune χ^2	45	30
Vertex fit χ^2	50	
Seed Vertex χ^2	50	
Vertex L_{xy} significance	7.5	

TABLE 4.1: Optimised SecVtx parameters for Pass1 and Pass2 tracks.

experiments at Fermilab for a number of reasons [27]. However, much work has been put into the understanding of the secondary vertex tagging using the JetClu algorithm with a cone of size 0.4. The study presented in this section was carried out jointly with a colleague from the University of Geneva in order to justify the use of the SecVtx algorithm for analyses involving MidPoint jets. This study can also be found in [28].

In the next sub-section the effect of changing algorithm and cone size but keeping the sub-cone size always at 0.4 is investigated. In sub-section 4.2.2 the effect of changing the sub-cone size from 0.4 to 0.7 is studied. No significant changes are expected while keeping the sub-cone size constant but when the sub-cone size is varied, the observed quantities are expected to be quite different. The final sub-section draws some general conclusions about the applicability of the SecVtx tagging algorithm using the MidPoint algorithm instead of JetClu, as well as larger jet cone sizes and different merge fractions.

4.2.1 Algorithm Comparison Using a Sub-Cone of Size 0.4

Some quantities relevant to the performance of SecVtx were identified and plotted for the five following cases

- JetClu cone of size 0.4
- JetClu cone of size 0.7
- MidPoint cone of size 0.4 with merge fraction 50%
- MidPoint cone of size 0.7 with merge fraction 50%
- MidPoint cone of size 0.7 with merge fraction 75%

All these cases use a sub-cone of size 0.4.

The Jet50 dataset is used for these comparisons (see section 5.1). The runs are required to pass the good run criteria with the silicon vertex detector functioning , while individual events must pass the following selection cuts:

- Missing E_T significance ≤ 5.0 (see section 5.3). This cut rejects most of the background from cosmic rays

- The total transverse energy (E_T) of the event ≤ 2 TeV
- At least one good vertex in the event with $|Z_{\text{vtx}}| \leq 60\,\text{cm}$, to ensure a good secondary vertex reconstruction.

A total of 5'627'855 events passed these selection cuts.

There are of course some unavoidable intrinsic differences between the JetClu and MidPoint algorithms, such as the fact that JetClu uses the Lorentz-boost invariant quantities transverse energy and pseudo-rapidity, whereas MidPoint uses transverse momentum and rapidity, as mentioned in chapter 3. All comparisons as a function of E_T (p_T) or η (Y) are thus affected by this difference. This effect is relatively small except for a few plots where explicitly mentioned. A detailed study of this effect is beyond the scope of this study. The JetClu quantities E_T and η were chosen for the plots.

Comparison of Some General Quantities

Table 4.2 shows, for the different jet algorithms, cone sizes and merge fractions

- the total number of jets in the events which pass the selection cuts above
- the number of jets which are considered by the SecVtx algorithm
- the number of positively tagged jets
- the number of negatively tagged jets

The total number of tags varies between the different scenarios considered as does the total tag rate, defined here as the ratio between the number of positively tagged jets and the number of jets considered by the SecVtx algorithm. The number of jets considered by SecVtx was chosen instead of the total number of jets which pass the trigger cuts because this reduces the effect of the bias due to the different trigger efficiencies of the different algorithms [29]. The tag rates are shown in table 4.3 along with the percentage of jets which are considered by the SecVtx algorithm.

Algorithm	# jets	# jets in SecVtx	# pos tag jets	# neg tag jets
JetClu 0.4	38'333'713	15'061'161	279'739	77'431
JetClu 0.7	35'088'905	16'150'490	271'489	79'187
MidPoint 0.4 $f_{\text{merge}} = 50\%$	35'1816'89	15'369'116	279'064	78'981
MidPoint 0.7 $f_{\text{merge}} = 50\%$	33'484'417	16'661'198	260'116	76'964
MidPoint 0.7 $f_{\text{merge}} = 75\%$	39'183'065	16'675'557	268'979	78'784

TABLE 4.2: Total number of jets, number of jets which are considered by the SecVtx algorithm, number of positively tagged and negatively tagged jets for the five algorithms considered

For all the plots in this section which are divided into two parts, the lower parts represent the ratio of the distributions for the different jet algorithms over the distribution for JetClu cone of size 0.4.

Figure 4.3 shows the number of jets per event. Figure 4.3a, the total jet multiplicity in the event, shows, as expected, a difference between the different algorithms, cone sizes and merge fractions. Since JetClu sometimes reconstructs two clusters of towers, closer than twice the cone size, as two jets when MidPoint reconstructs them as a single jet, we expect a higher jet

Algorithm	% jets in SecVtx	% + tag rate	% - tag rate
JetClu 0.4	39.29	1.86	0.51
JetClu 0.7	46.03	1.68	0.49
MidPoint 0.4 $f_{\text{merge}} = 50\%$	43.67	1.82	0.51
MidPoint 0.7 $f_{\text{merge}} = 50\%$	49.76	1.56	0.46
MidPoint 0.7 $f_{\text{merge}} = 75\%$	42.56	1.61	0.47

TABLE 4.3: Percentage of jets considered by the SecVtx algorithm as well as the positive and negative tag rates

multiplicity for JetClu than for MidPoint (if the merging fraction were the same).

There are fewer jets per event with the lower f_{merge} fraction than with the higher one, for a given cone size and algorithm, because the jets are more likely to be merged together; only 50% of the energy needs to be overlapping as opposed to 75% for the higher merge fraction. Because of geometrical/spatial constraints, fewer jets are expected per event when considering a cone of size 0.7 instead of 0.4, which is what is observed. Nevertheless, it is necessary to point out that, although we expect the contrary, the number of MidPoint jets with a cone size of 0.7 with merging fraction 75% is 15% higher than MidPoint with f_{merge} of 50% and 7% higher than the number of JetClu jets. Table 4.4 shows that the high jet multiplicity of MidPoint $f_{\text{merge}} = 75\%$ is mostly due to soft jets (below 15 or 20 GeV). This is a feature that should be better understood when optimising the jet algorithm. For the events that are considered in this study, the differences between the algorithms are removed due to the fact that the SecVtx cuts on the jets are tight.

Algorithm	# jets (all)	$E_T > 15\,\text{GeV}$	$E_T > 20\,\text{GeV}$	$E_T > 30\,\text{GeV}$	$E_T > 50\,\text{GeV}$
JC07	35'088'905	14'315'300	12'783'220	10'885'300	6'956'728
MP07 75%	39'183'065	14'544'120	12'821'140	10'845'760	6'697'447
MP07 50%	33'484'417	15'121'810	13'291'400	11'232'530	7'269'577

TABLE 4.4: Total number of jets for JetClu and MidPoint algorithms (cone size 0.7) for different E_T cuts.

Looking at the E_T distribution of all jets, of the jets considered by SecVtx and of the tagged jets (figure 4.4), there is no significant difference between the algorithms for the same cone size. As larger cones contain more towers, the distribution is shifted slightly upwards in energy relative to the distributions using cones of size 0.4. One can also see the effect of the selection cuts of SecVtx and the tagging selection which remove most of the low E_T jets.

Next, secondary vertex quantities are examined for all tagged jets. Distributions of the L_{xy} for both positively and negatively tagged jets (figures 4.5a and b, respectively), of the pseudo-$c\tau$ [4] (fig. 4.5c), of the L_3 [5] (fig. 4.5d) as well as of the secondary vertex masses for positively and negatively tagged jets (fig. 4.6a and b, respectively) all show no significant differences between algorithms, cone sizes and merge fractions.

The tag rates are shown in figure 4.7 as a function of E_T and ϕ of the tagged jets, as well as a function of the number of good tracks in the tagged jets and the z-component of the primary vertex location. Figure 4.8a shows the tag rate vs. η of the tagged jets. Although the total tag

[4] pseudo-$c\tau = \frac{L_{xy} M_{\text{vtx}}}{p_{T_{\text{vtx}}}}$

[5] L_3 is the 3-dimensional distance between the primary and secondary vertices

rate is not the same for all jet algorithms, the shapes of the distributions are the same as can be seen in figure 4.8c. Figure 4.8b shows the same distribution as 4.8a but with an additional cut on the E_T of the jets at 50 GeV. JetClu cone size 0.4 has the highest tag rate when considering all jets but has the lowest rate when considering only jets above 50 GeV. This shows that the differences between the tag rates is most probably due to the differences in behaviour of the algorithms at low E_T.

In figure 4.9a, the excess tag rate as a function of E_T is shown. The excess tag rate is defined as the difference between the number of positively tagged and negatively tagged jets over the total number of jets considered by SecVtx. This excess gives a rough estimate of the heavy flavour content of the jets. The distribution of the fraction of tracks in the sub-cone of size 0.4 around the jet axis that are used for the tag shows no significant difference between the five different cases.

The final general quantities shown are quantities related to the tracks which make up the secondary vertices. First the number of tracks are compared at different stages of the SecVtx algorithm. Figure 4.10 shows the number of good tracks, Pass 1 and Pass 2 tracks for all jets considered by SecVtx as well as the number of tracks used in the tag for positively and negatively tagged jets. The angular separation between the jet axis and the tracks used in the tag along with the ϕ distribution of these tracks and their d_0 distance from the primary vertex are shown in figure 4.11. The transverse momentum, p_T, distribution of these tracks is also shown in this figure. These are control variables to make sure the tracks are distributed inside the jet in the same way for all cases considered, and that SecVtx is picking up mostly the same tracks to make the tags, independently of which algorithm is used. In the ΔR distribution of the tracks, there is a difference between the different cone sizes which is most likely due to the differences in the E_T distribution of the tagged jets. The distributions for JetClu and MidPoint with the new merge fraction and the same cone sizes are the same. No convincing explanation for the differences in the ΔR distributions has been found yet. The number of $r - \phi$ hits in the SVX detector as well as the number of hits in the COT detector are also plotted (figure 4.12).

The plots in this section show that there are sometimes differences between the algorithms but either these differences are small with respect to the precision of the SecVtx algorithm (of the order of a few percent), or these differences can be explained by the intrinsic differences between the algorithms. It is evident that the use of the new merging fraction with MidPoint significantly reduces the differences with respect to the behaviour of the JetClu algorithm. In particular the tag rates are more similar, with only a 4% difference in the absolute tag rate for cone size 0.7, instead of a 7% difference seen with the old merge fraction.

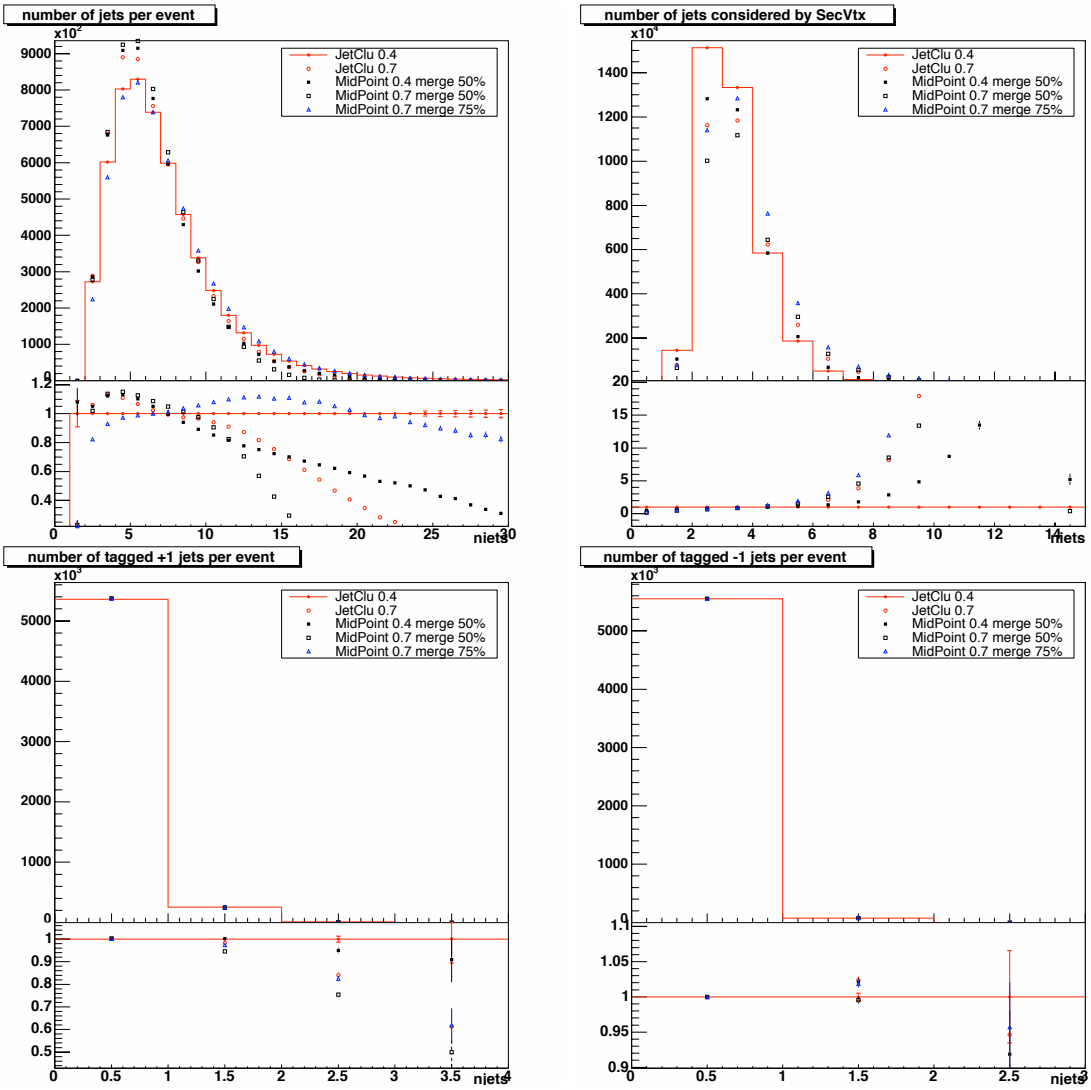

FIGURE 4.3: Number of jets (a), jets considered by the SecVtx algorithm (b), positively tagged jets (c), negatively tagged jets (d) per event.

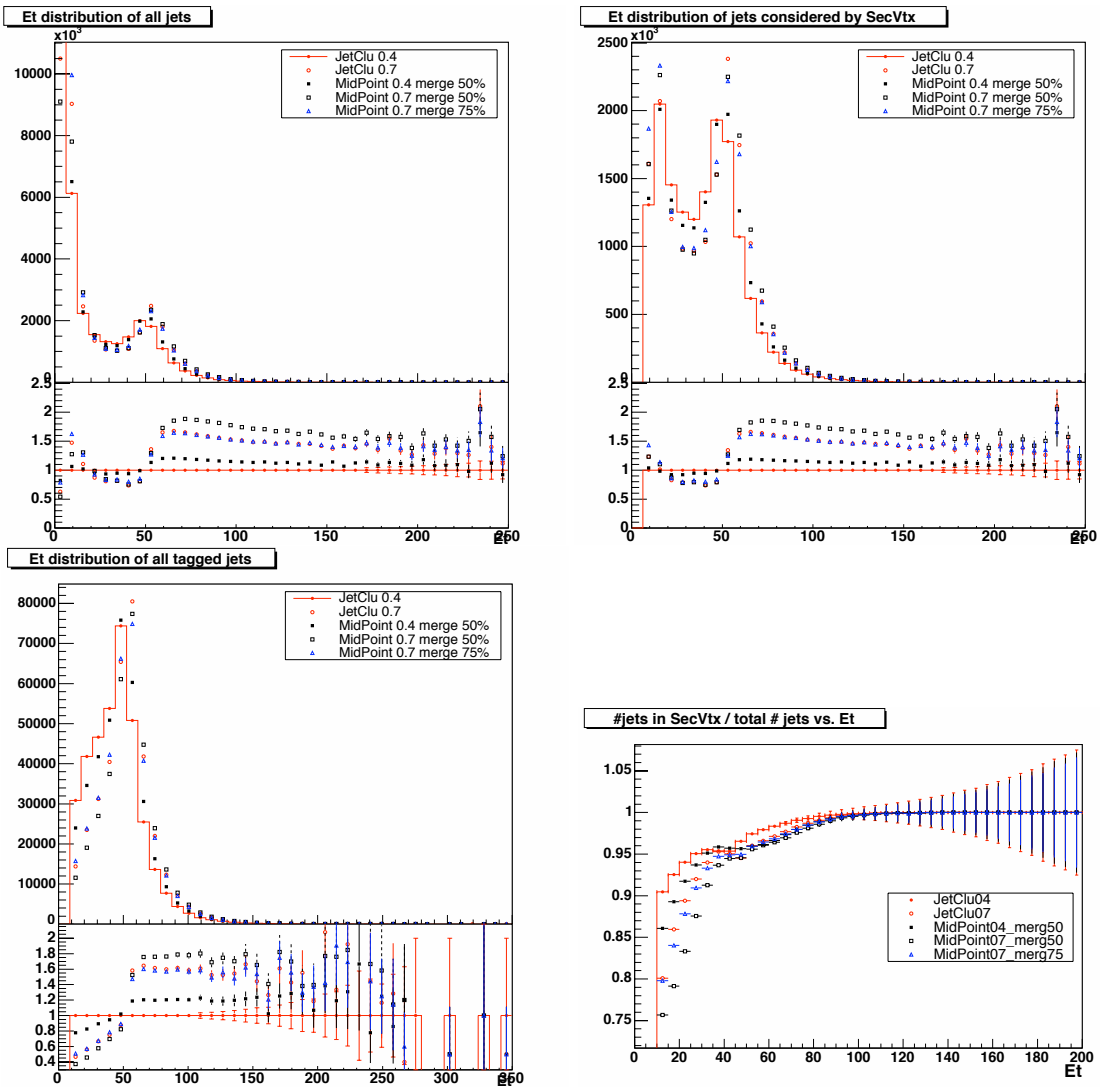

FIGURE 4.4: E_T distribution of all jets (a), of jets considered by SecVtx (b), of tagged (+1 or -1) jets (c) and fraction of jets which are considered by SecVtx vs. E_T (d).

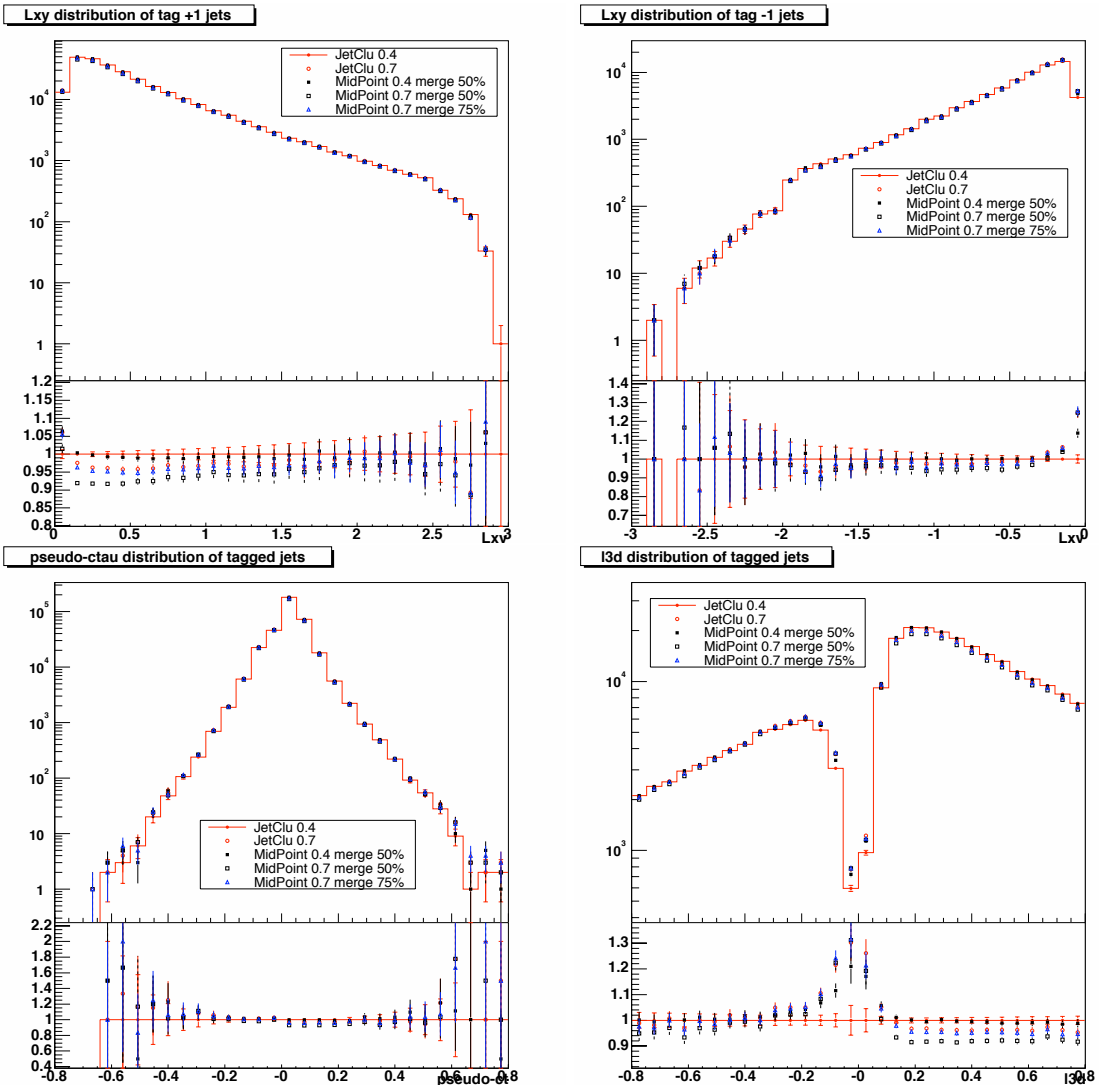

FIGURE 4.5: L_{xy} distribution for positively (a) and negatively (b) tagged secondary vertices, pseudo-cτ (c) and L_{3D} (d) distributions of all tagged secondary vertices.

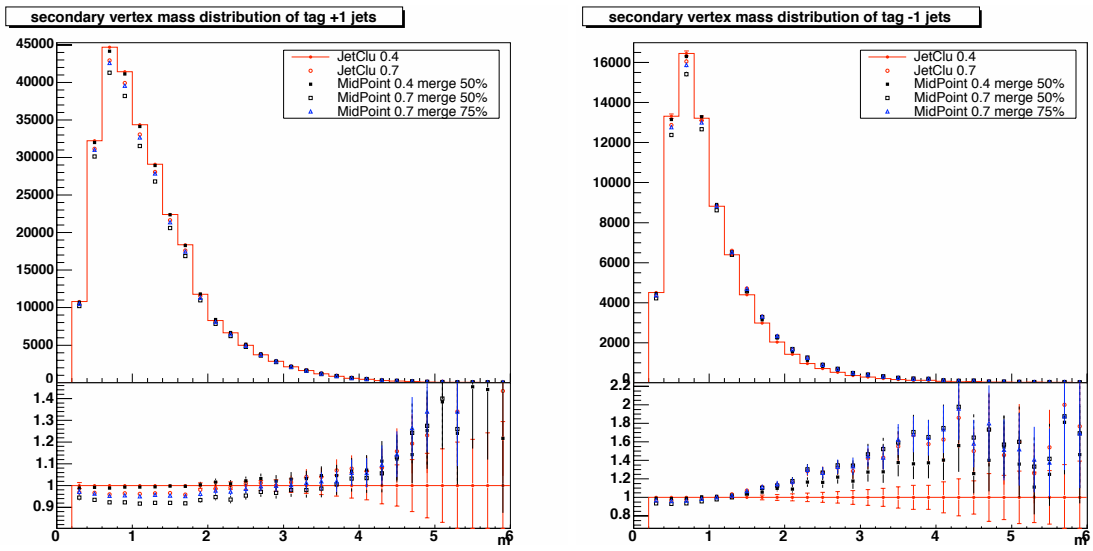

FIGURE 4.6: Secondary vertex mass distribution for positively (a) and negatively tagged (b) jets.

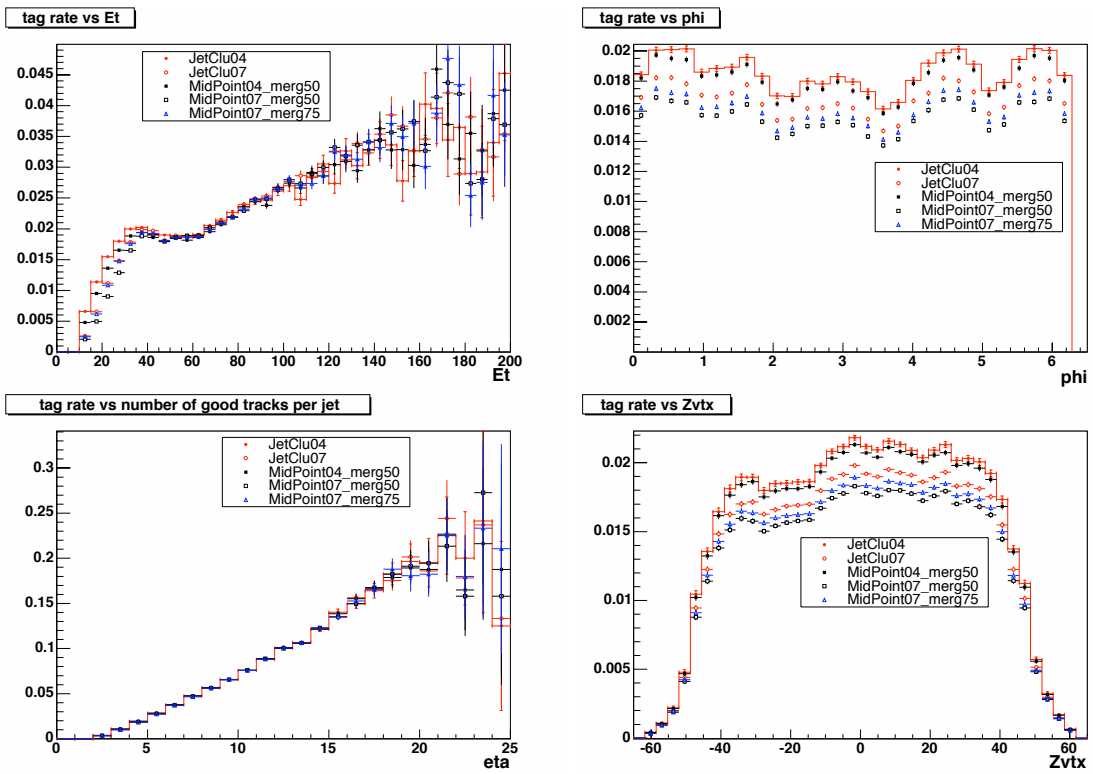

FIGURE 4.7: Fraction of positively tagged jets vs. E_T (a), ϕ (b), number of good tracks (c), Z_{vtx} (d).

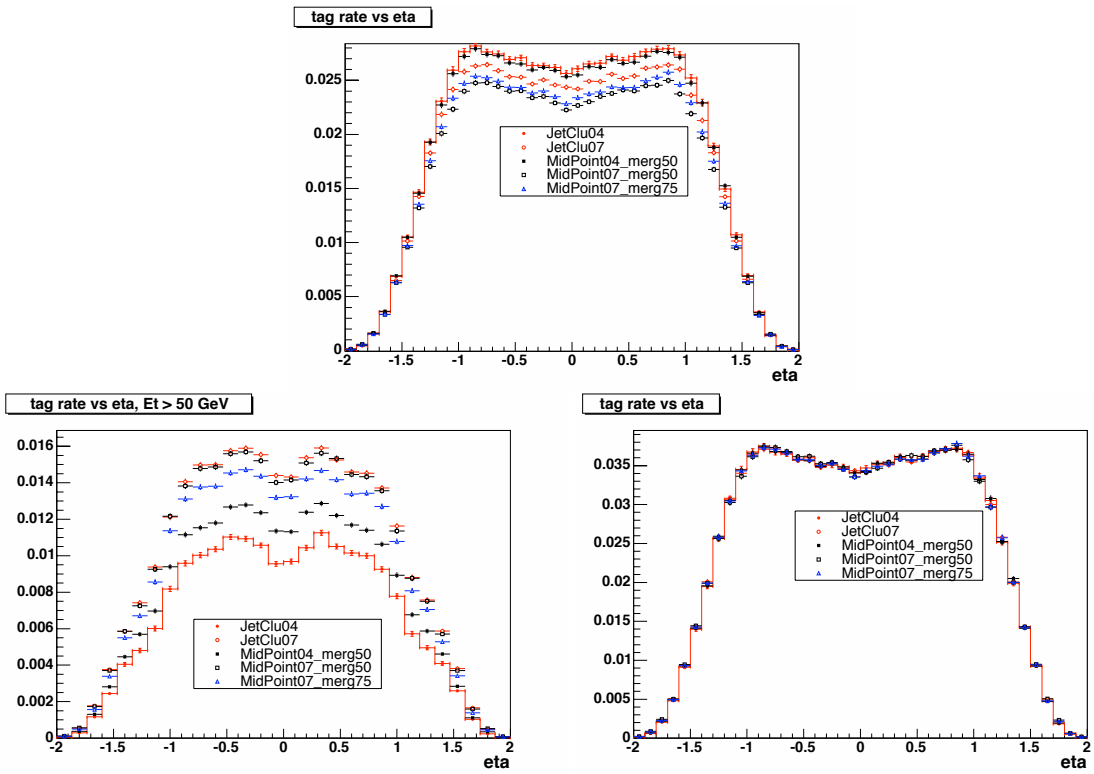

FIGURE 4.8: Fraction of positively tagged jets vs. η (a), vs. η where $E_T > 50\,\mathrm{GeV}$ (b) and vs. η where the rates are each normalised to unit integral (c).

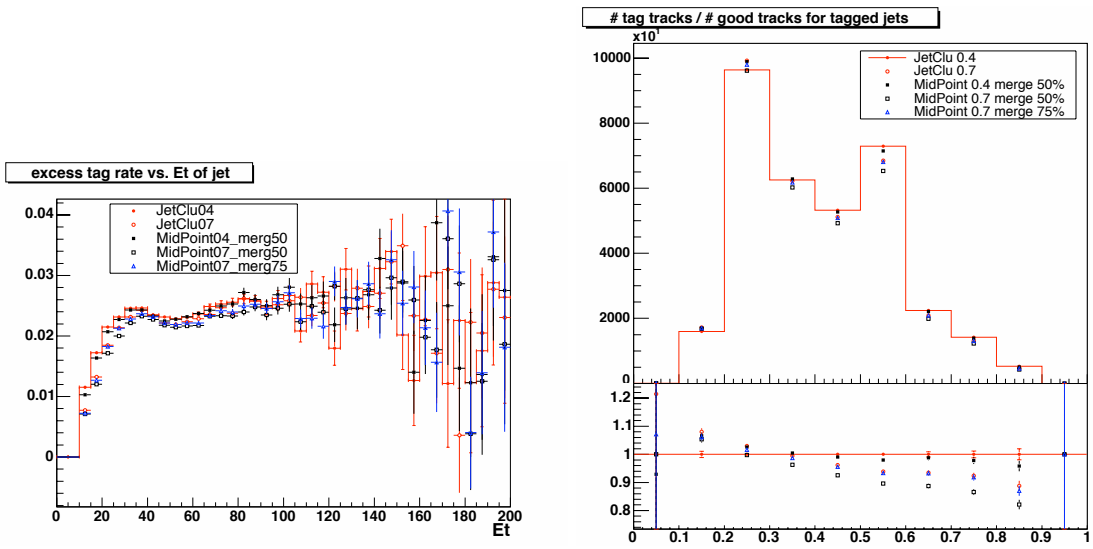

FIGURE 4.9: Excess tag rate: difference between the number of positively and negatively tag jets over the total number of jets considered by SecVtx vs. E_T (a) and ratio of number of tag tracks over number of good tracks in tagged jets (b).

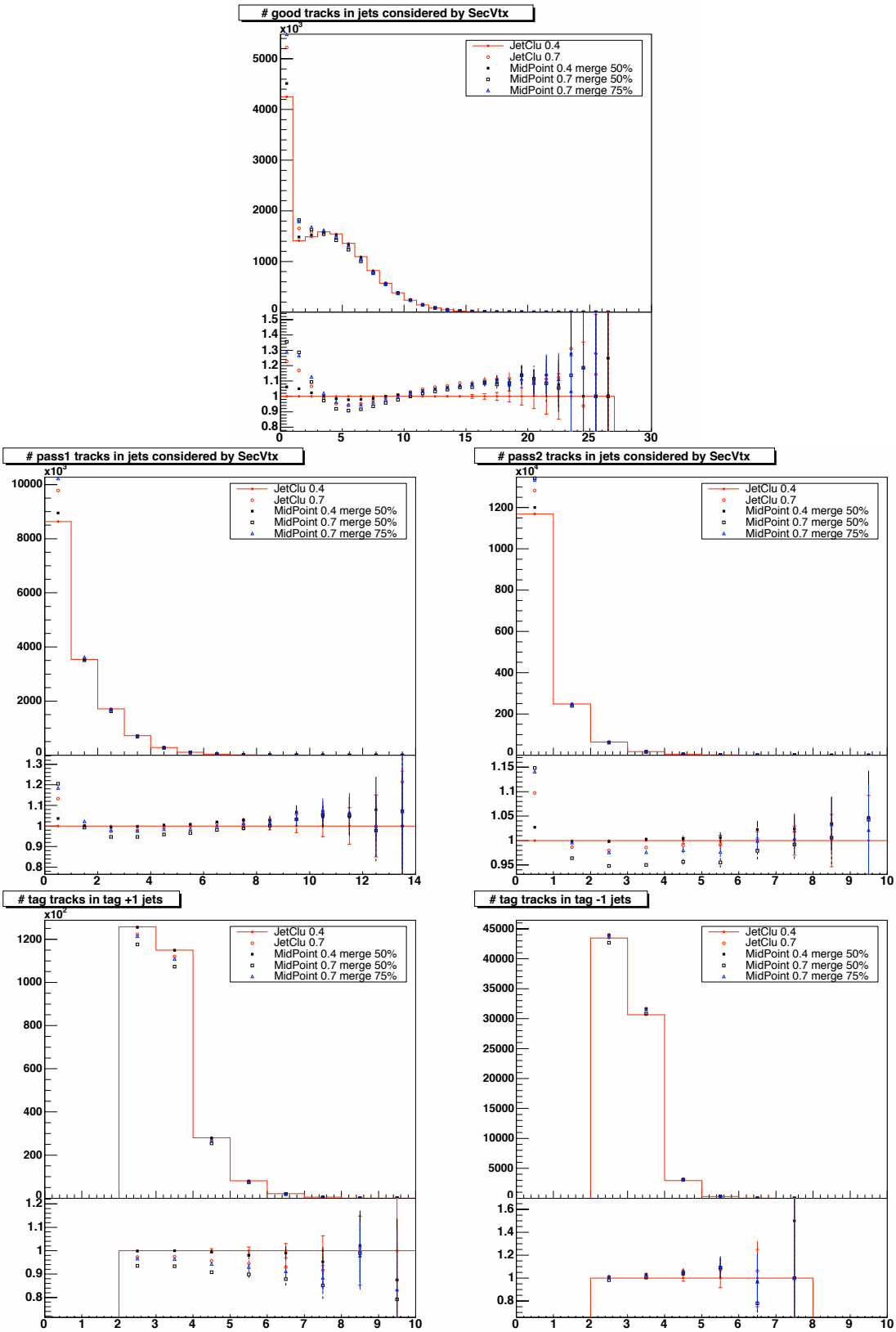

FIGURE 4.10: Number of good tracks (a), Pass 1 tracks (b), Pass 2 tracks (c), number of tracks used by SecVtx for positively tagged (d) and negatively tagged (e) jets.

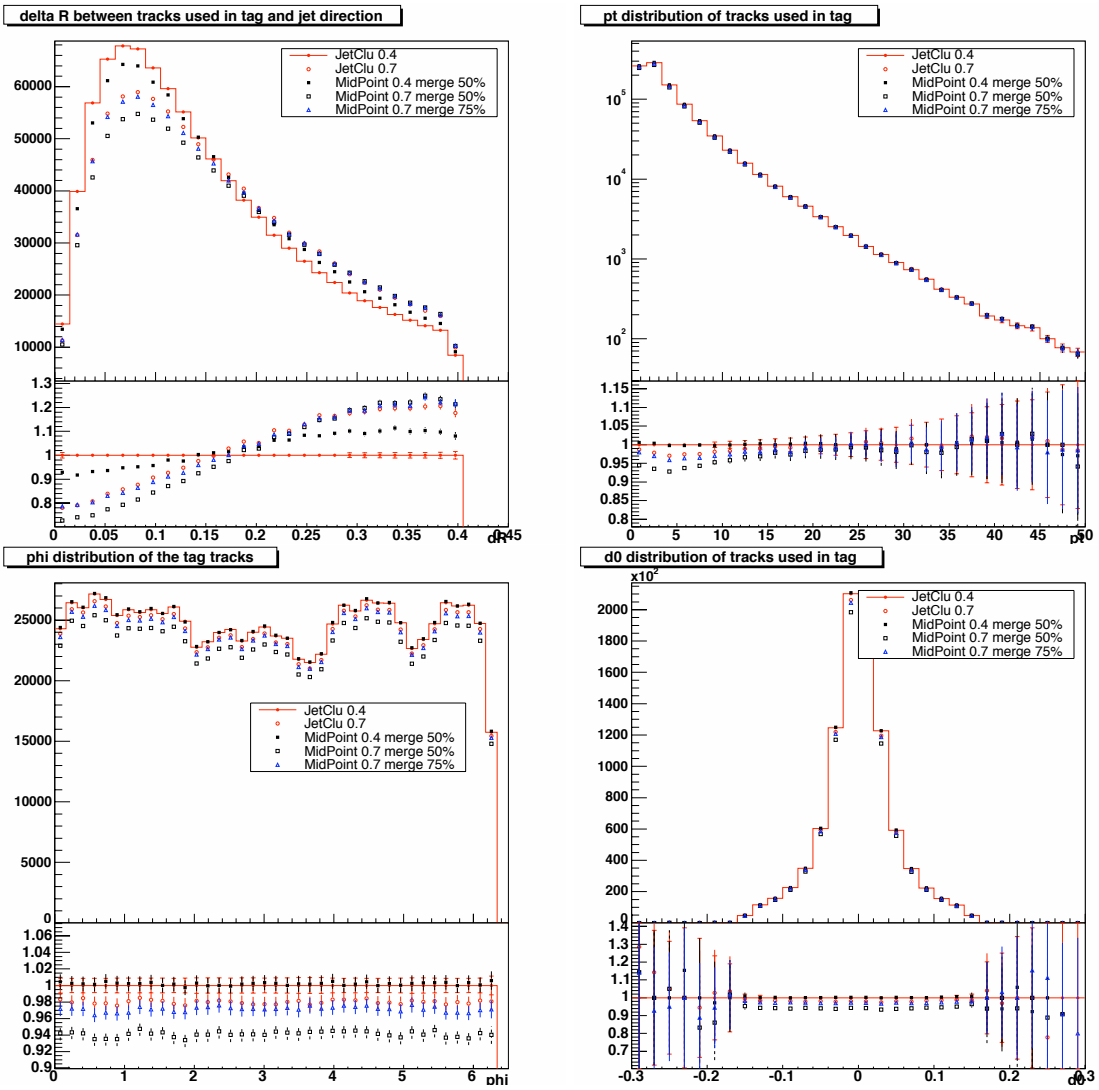

FIGURE 4.11: ΔR separation between jet direction and tracks used in the tag (a), p_T (b), ϕ (c) and d_0 (d) distribution of these tracks.

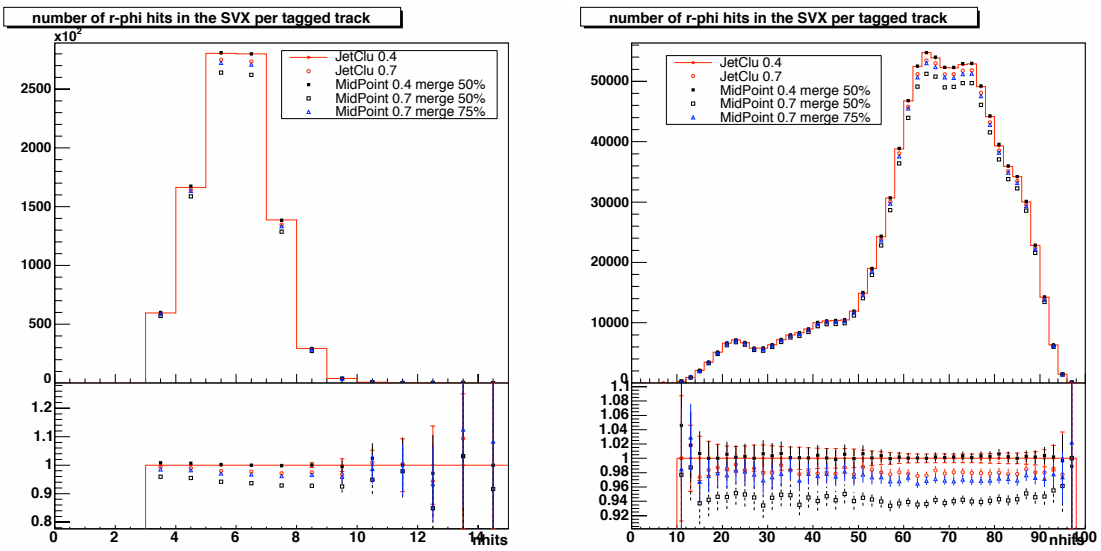

FIGURE 4.12: Number of $r-\phi$ hits in SVX per tag track (a) and number of hits in the COT per tag track (b).

Event by Event Comparisons Between Jets

The general quantities related to SecVtx presented in the previous section show no significant difference between algorithm or cone size, yet it is still important to investigate more closely the differences by looking at event by event comparisons between the algorithms. The two algorithms are compared for both cone sizes of 0.4 and 0.7, still with a sub-cone of size 0.4. In the case of jets with a cone size of 0.7, both merge fractions, $f_{\text{merge}} = 0.5$ and 0.75, are considered, the latter being the official recommendation for analyses submitted to the Winter 2005 conferences. In a first step all events are considered in which there is at least one tag in each of the two algorithms.

Taking the tagged JetClu jets as reference, the angle between the JetClu jet and the closest, in ϕ, tagged MidPoint jet was computed (solid lines in figure 4.13, left column). It was found that in most cases these two jets are very close together, with only a very small fraction of jet pairs being further apart than the cone size of the jets. The rise in the $\Delta\phi$ distribution around π is assumed to come from $b\bar{b}$ or $c\bar{c}$ events where, for example, both jets are tagged in JetClu but only one is tagged in MidPoint or vice-versa. In order to remove this event topology, that could be misleading for further interpretations, we consider only events with one tag per event (dashed line in figure 4.13). The rise around π disappears and the tagged jets clearly correspond to the same reconstructed calorimetric object. As a confirmation of this, the difference in the L_{xy} of the secondary vertices between the two algorithms is plotted in the right hand column of figure 4.13.

Figure 4.14 shows, for cone sizes 0.4 and 0.7 (0.7 cones with both $f_{\text{merge}} = 0.75$ and $f_{\text{merge}} = 0.5$), the L_{xy} of the jets reconstructed using JetClu (horizontal axis) versus the L_{xy} of jets reconstructed using MidPoint jets (vertical axis) for all jets with at least one tag (positive or negative) in at least one of the algorithms; on the right, the projections are shown[6]. It can clearly be seen that most jets are tagged using both algorithms and have approximately equal values of L_{xy}. Moreover, the probability of having a jet tagged in one algorithm but not in the other is similar for both algorithms as can be seen from the symmetric distribution of the lines at -7, which is set to be the value of L_{xy} when no tagged jet is found. Over 95% of the events are tagged in both algorithms for cones of size 0.4 and cones of size 0.7 with $f_{\text{merge}} = 0.75$. For the case of cone size 0.7 and $f_{\text{merge}} = 0.5$, about 6.6% of the total number of jets tagged using JetClu are not tagged using MidPoint, making the tagging probability more asymmetric. This is caused by the lower merging fraction as explained below.

A closer look is taken at the cases where the jet is tagged using one algorithm but not when using the other. First, taking jets tagged using JetClu but whose corresponding MidPoint jet is not tagged, the ΔR separation of the pair of jets is shown in figure 4.15 in the left column. The right column shows plots of the opposite case, when the jets are tagged using MidPoint but there is no corresponding tag when using JetClu. In all cases, the small ΔR separation shows that, in general, the "object" reconstructed as a jet is the same using both algorithms. The fact that it is tagged in one case but not in the other is interpreted by random tagging inefficiencies or because of small differences in the jet directions. For example, one or more tracks used for the secondary vertex reconstruction might move from being just inside the sub-cone to being just outside it. A further quantity studied is

$$\Delta = \frac{(E_T^{\text{MP}} - E_T^{\text{JC}})}{E_T^{\text{MP}}} \tag{4.2}$$

where the MidPoint jet is tagged and the closest JetClu jet is considered, or vice-versa. Figure 4.16 shows this quantity Δ for all combinations. The most straightforward interpretation for

[6]On these projections, the "hole" around 0 is due to the SecVtx algorithm cut on L_{xy} significance ($L_{xy}/\sigma_{L_{xy}}$) used to select jets with secondary vertices.

the results is obtained for jets of cone size 0.4, when the jet is tagged using JetClu but not using MidPoint (left) and vice-versa (right): two peaks are clearly visible. The first peak was already discussed above. The second peak, centred around 0.5, is due to events where a single MidPoint jet is reconstructed as two jets, sharing the energy roughly equally between them, when using JetClu. This is due to the difference in the seeding mechanisms between the two algorithms and in the different merging and splitting procedures, as explained in chapter 3. The dependence on the latter is visible also when comparing the two merging fractions for jets of cone size 0.7.

This is further justified by the fact that there sometimes is a second jet, close to the initial one, that can be combined with the other jet before the Δ comparison is made. Figure 4.17 shows the fractional transverse energy considering events with at least one jet tagged when using MidPoint but no tags when considering JetClu jets, for jet cones of size 0.4. Also shown is the Δ comparison when considering the E_T of the tagged MidPoint jet and the sum of the E_T of the two closest JetClu jets. In most of the cases, the two JetClu jets appear to correspond to the same calorimetric object, that is reconstructed as an unique jet using MidPoint. The second peak at 0.5 decreases significantly when comparing the MidPoint jet to the sum of the two closest JetClu jets.

The splitting of one jet into two when using JetClu is less obvious for cone sizes of 0.7, because of the fact that we are considering only tracks within a sub-cone of size 0.4, significantly smaller than the jet cone size.

The conclusions of this part of the study are that the probability that a jet is tagged when it is reconstructed using MidPoint but not when using JetClu is very close to the probability that a jet is tagging when it is reconstructed using JetClu but not when using MidPoint. Moreover, the fraction of jets tagged using one algorithm but not using the other represents fewer than 5% of the total number of tagged jets.

Scale Factor Data/Monte Carlo for b-tagging Efficiency

The present study has been performed using data from an inclusive jet sample, where the true heavy flavour content is unknown. For this reason it is not possible to make a direct comparison between the b-tagging efficiency values for the different jet algorithms and jet cone sizes. Nevertheless, a careful study has been performed in reference [29] using the electron sample and the methodologies developed by the high p_T b-tagging group at CDF. The conclusion is that the same scale factor [7] as obtained with JetClu cone size 0.4 can be adopted for MidPoint cone size 0.7 jets (f_{merge} =0.75), together with the related systematic uncertainties.

[7] The scale factor is the factor by which the tagging efficiency measured in the Pythia Tune A MC simulation must be multiplied in order to obtain the tagging efficiency in data.

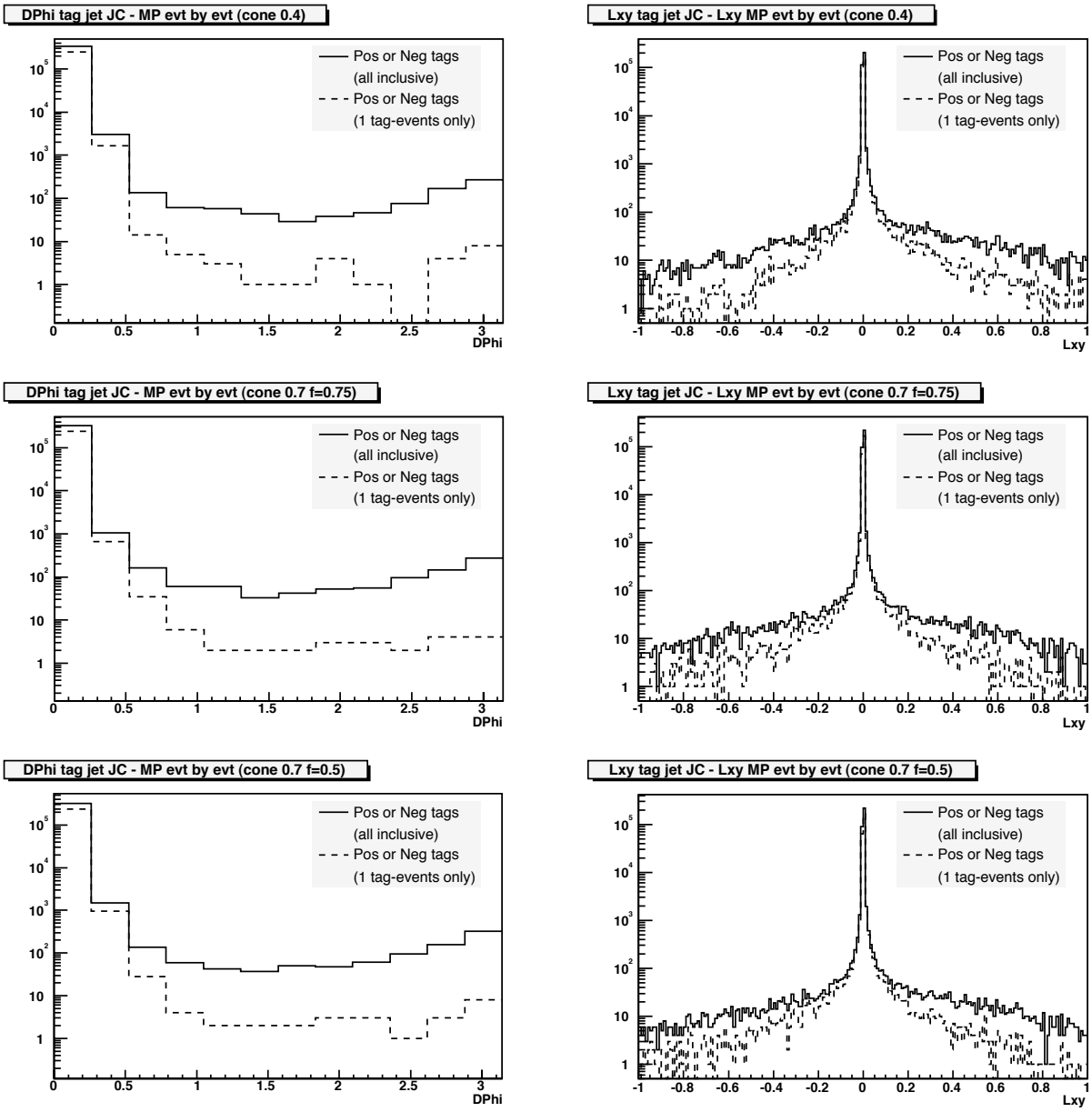

FIGURE 4.13: $\Delta\phi$ (left column) and ΔL_{xy} (right column) distributions between a tagged JetClu jet and the nearest tagged MidPoint jet. The solid line corresponds to all tagged jets, while the dashed line is for events with only one tagged jet for each case. Both positively and negatively tagged jets are considered.

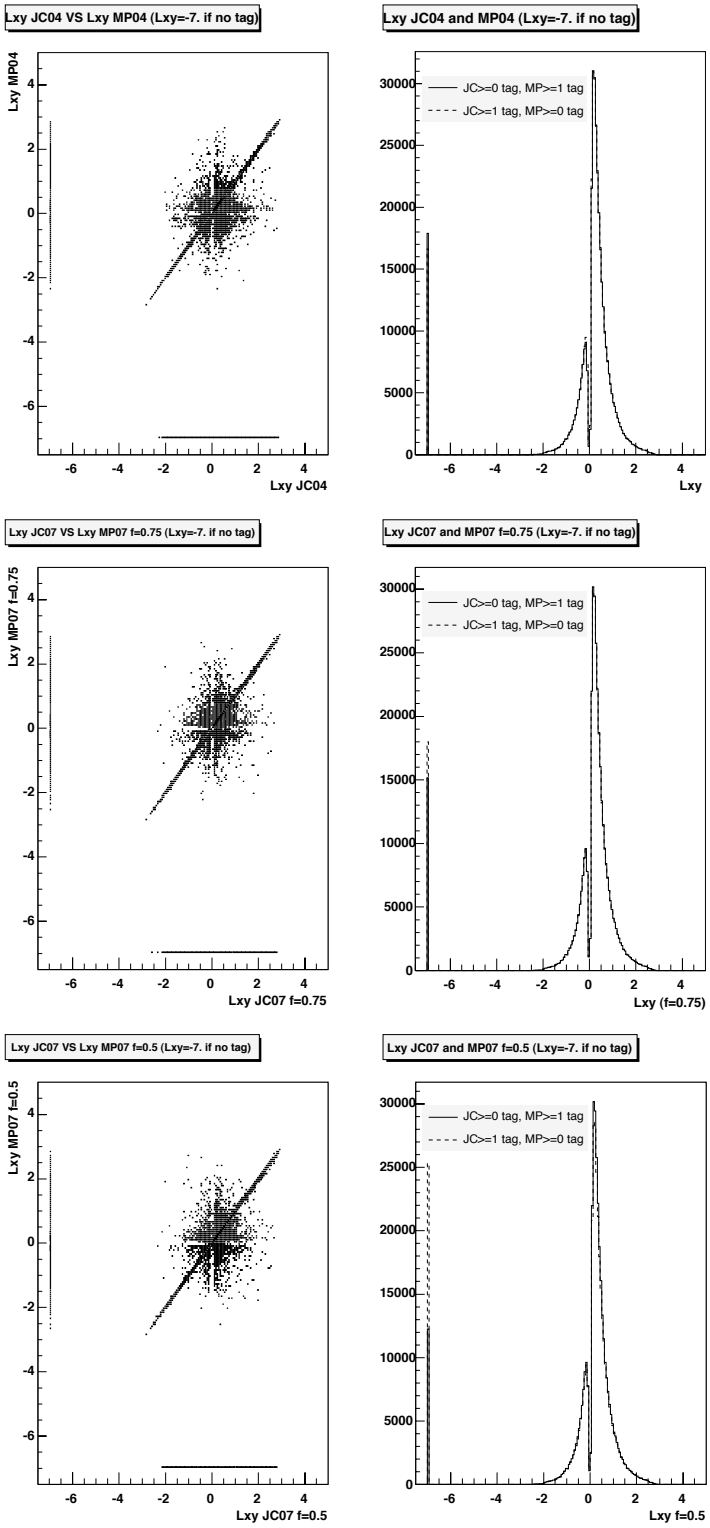

FIGURE 4.14: L_{xy} distribution of the secondary vertices of the jets plotting JetClu versus MidPoint for cone size 0.4 (top), cone size 0.7, $f_{\text{merge}} = 0.75$ (middle) and cone size 0.7, $f_{\text{merge}} = 0.5$ (bottom). L_{xy} is fixed at -7 when no reconstructed secondary vertex is found. The left hand plots show the scatter plots of MidPoint versus JetClu whereas the right hand plots show the distributions for the cases where there are no tags in JetClu but tags in MidPoint and vice-versa.

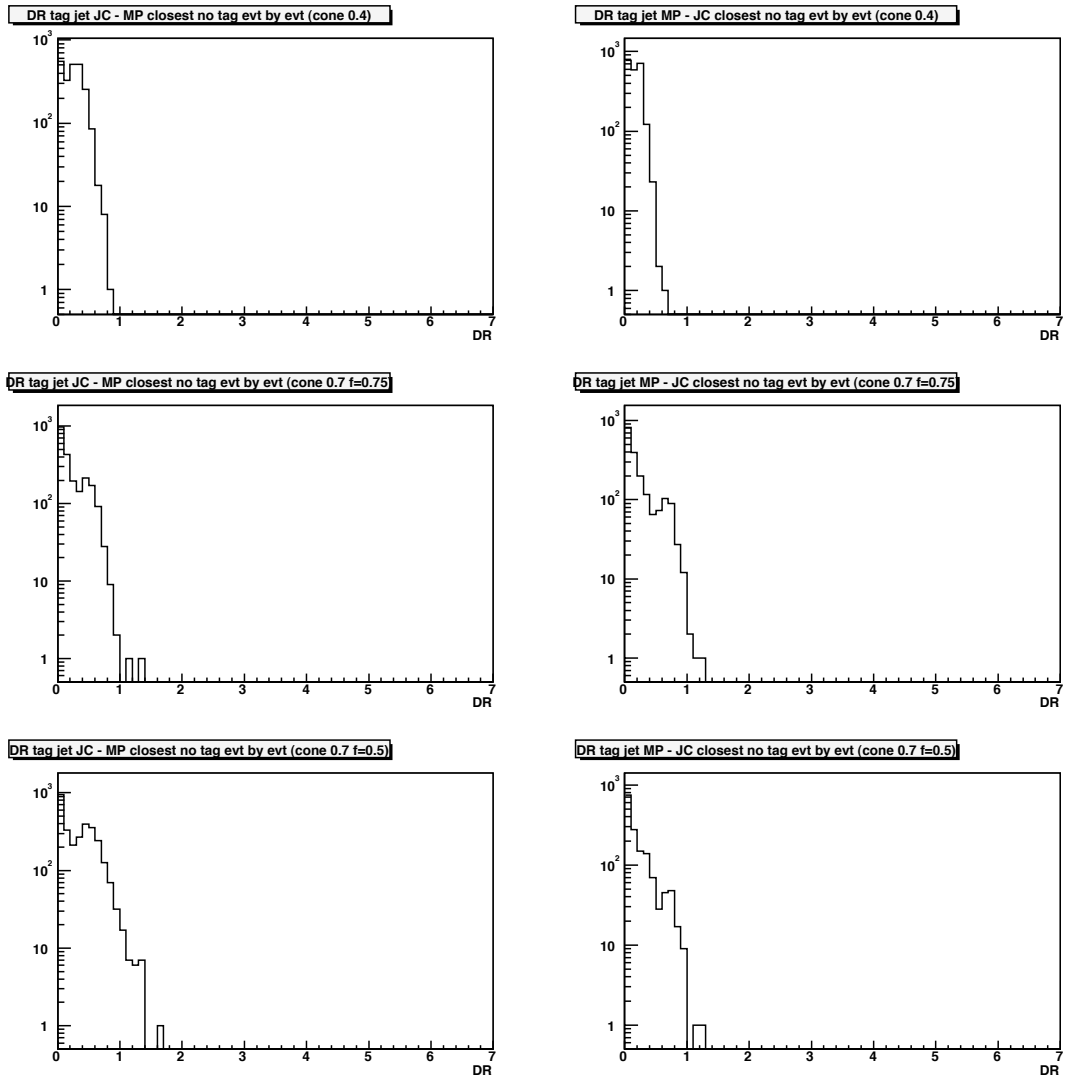

FIGURE 4.15: ΔR separation for jet pairs defined with a jet tagged using JetClu but not when using MidPoint (column on the left) and vice-versa (column on the right) for cone size 0.4 (top), cone size 0.7, $f_{\text{merge}} = 0.75$ (middle) and cone size 0.7, $f_{\text{merge}} = 0.5$ (bottom).

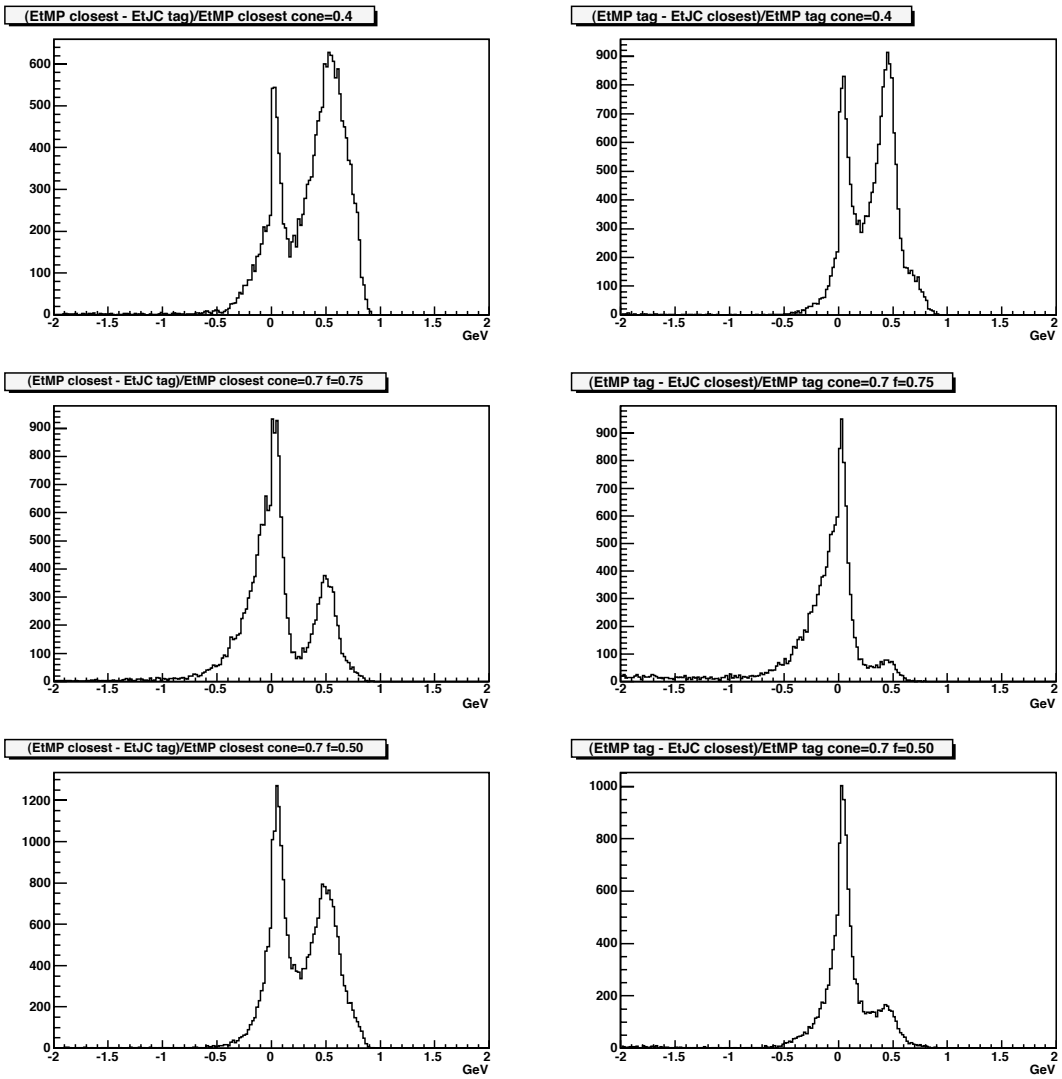

FIGURE 4.16: Fractional E_T difference (Δ) between pairs of jets tagged when using JetClu but not when using MidPoint and vice-versa for cone size 0.4 (top), cone size 0.7, $f_{\text{merge}} = 0.75$ (middle) and cone size 0.7, $f_{\text{merge}} = 0.5$ (bottom).

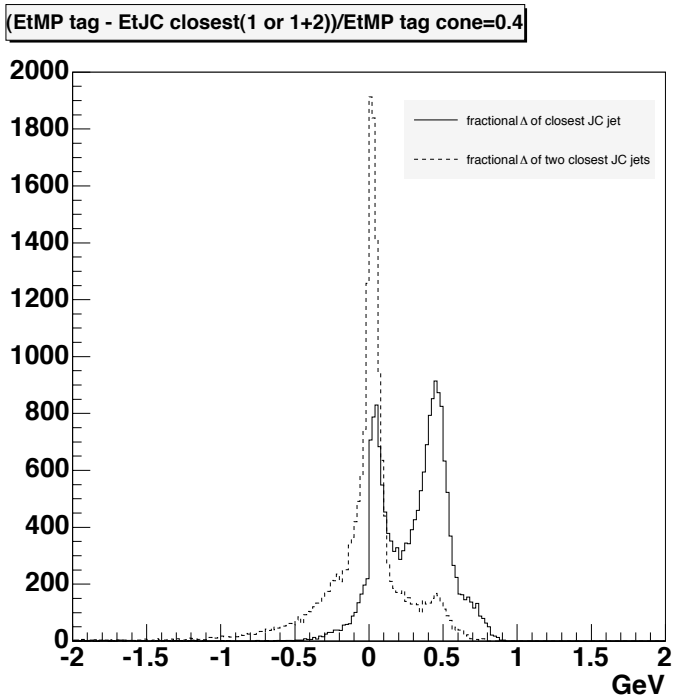

FIGURE 4.17: Fractional E_T difference (Δ) considering cone size 0.4 jets tagged in MidPoint but not in JetClu (solid line). Shown as a dashed line is the Δ computed from the tagged MidPoint jet and the sum of the E_T of the two closest JetClu jets: $E_T^{JC}=E_{T\,1}^{JC}+E_{T\,2}^{JC}$.

4.2.2 Effect of Changing the Sub-Cone Size from 0.4 to 0.7

The SecVtx algorithm has been optimised for jets with a cone of size 0.4. Although using jets of cone size 0.7, many analyses consider only tracks inside a sub-cone of size 0.4 to reconstruct secondary vertices. The motivation is mainly related to the fact that, on average, tracks used for secondary vertex tagging of heavy flavour jets are close to the jet axis. Tracks on the edge of the jet cone are most often soft. Including them into the SecVtx algorithm might lead to the reconstruction of fake secondary vertices. This would thus increase the mis-tagging rate. On the other hand, it might be interesting to evaluate the effects of changing the sub-cone from a size of 0.4 to a size of 0.7 and estimate possible advantages.

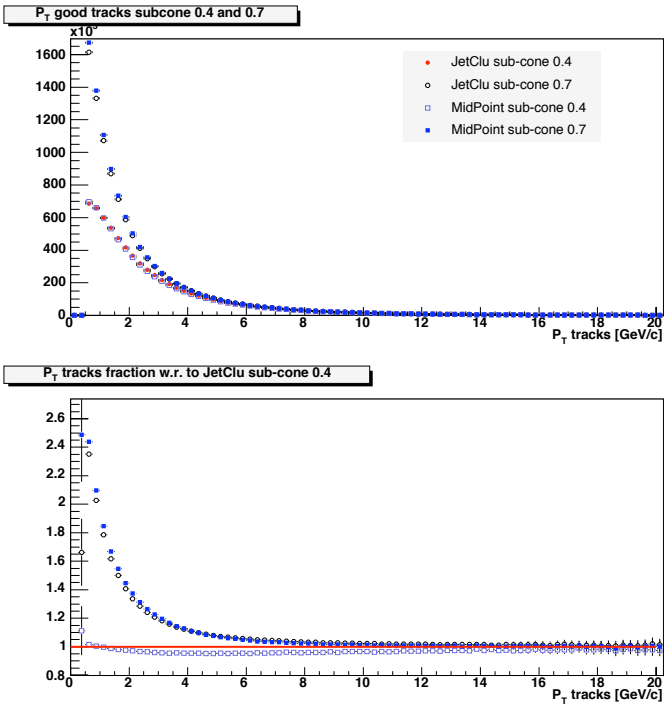

FIGURE 4.18: Transverse momentum of good tracks inside a sub-cone of size 0.4 and 0.7 for jets with a cone size of 0.7 (JetClu and MidPoint) with $E_T < 10\,\text{GeV}$ and $|\eta| < 2.4$. The lower plot shows the ratio with respect to the JetClu cone size 0.4.

The effect of using a larger sub-cone should be more pronounced for softer, lighter, jets which is why the Jet20 inclusive sample is used here (integrated luminosity about 150 pb^{-1}).

The same event selection is used as the one presented at the beginning of section 4.2.1, changing the missing E_T significance cut to 3.5. This study was carried out using an older version of the offline software, 5.3.1. It was not deemed necessary to update this part of the study because the conclusions below do not depend on the re-optimisation of the SecVtx algorithm performed when passing from version 5.3.1 to version 5.3.3_nt of the offline reconstruction software. The MidPoint algorithm implemented here has still the old merging fraction of 50%.

The number of tagged jet increases by about 40% when extending the region considered by SecVtx to the full jet cone; on the other hand, the number of negatively tagged jets increases by about 90% (see table 4.5).

Figure 4.19 shows a comparison of the positively and negatively tagged jet fractions between jets of sub-cone sizes 0.7 and 0.4, as a function of the transverse energy. The maximum increase in tag rate is found, as expected, for soft jets, below 30 GeV but the negative tag rate increases

Algorithm	sub-cone	N tag+ jets	N tag- jets
JetClu	0.4	140'099	29'488
JetClu	0.7	194'894	55'222
MidPoint	0.4	135'120	28'909
MidPoint	0.7	196'506	56'828

TABLE 4.5: Number of tagged jets (positive and negative) for MidPoint and JetClu algorithms for jets that have a sub-cone size of 0.4 and a sub-cone size of 0.7.

by approximately of 80%, independent of the jet energy. Figures 4.20 show the secondary vertex mass, the L_{xy} and the pseudo-cτ distributions. From the L_{xy} distributions it is quite clear that for both negatively and positively tagged jets, the increase in the number of tagged jets comes from jets with small L_{xy}, in agreement with the observation that the additional tracks used by the vertex reconstruction have on average a lower momentum.

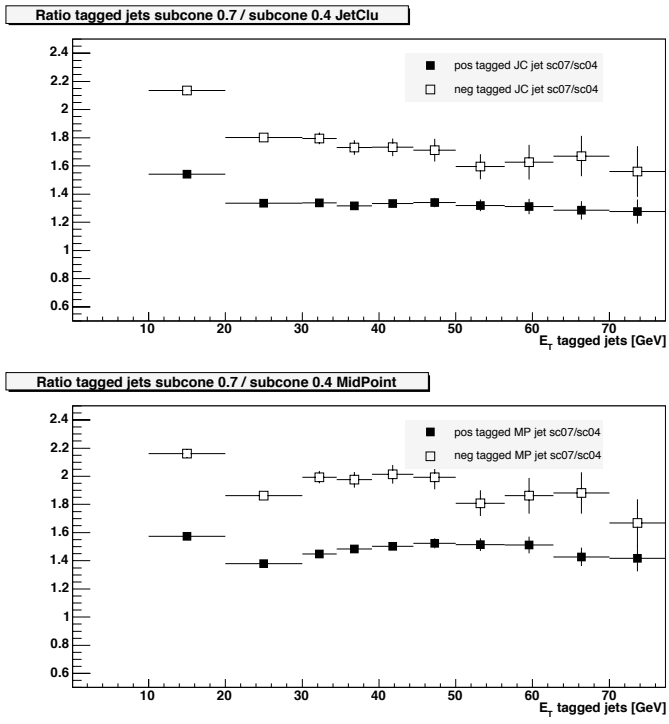

FIGURE 4.19: Positively and negatively tagged jet fraction distribution, $N^{07}_{tag}/N^{04}_{tag}$, as a function of jet transverse energy for JetClu jets (top) and MidPoint $f_{merge} = 0.5$ (bottom).

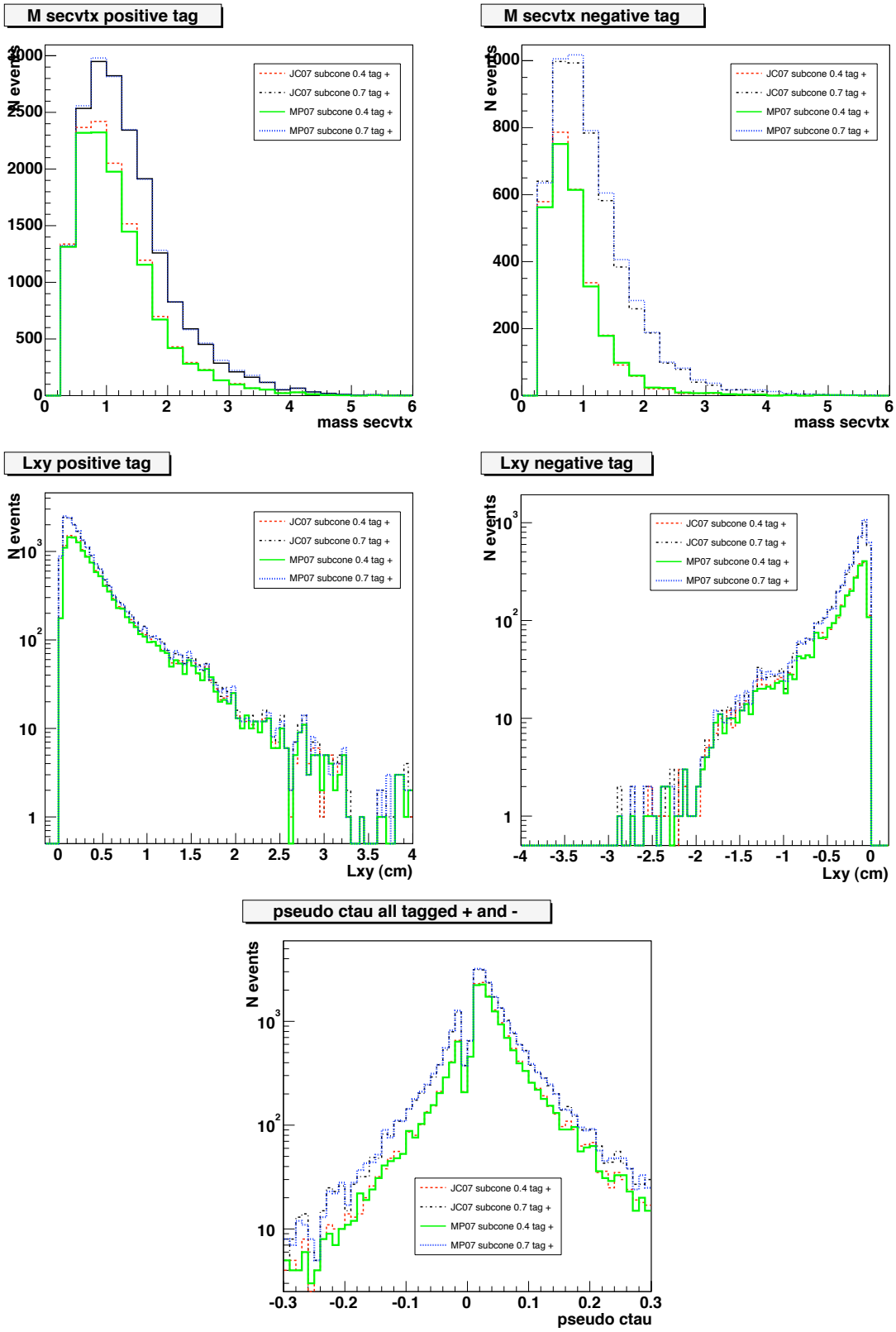

FIGURE 4.20: Secondary vertex mass for positively and negatively tagged jets (top left and right); L_{xy} distribution for positively and negatively tagged jets (middle left and right); pseudo-cτ distribution (bottom) showing the sum of the positively and negatively tagged jets.

Monte Carlo Comparisons

In inclusive jet samples, the fraction of heavy flavour jets is unknown. It is therefore not possible to derive the real effect of a larger sub-cone on the b-tagging efficiency and on the purity, defined as the fraction of tagged jets that are truly b-jets. For this purpose, a Pythia Tune A Monte Carlo sample has been used (1.2 M events of Pt18 dijet Pythia Tune A MC, see section 5.2), where it is possible to define the heavy flavour content at the truth level. The same analysis cuts described previously in this chapter have been used, and only JetClu jets are considered, since no differences are expected between the different jet algorithms. Finally, being interested in a relative number and not in the absolute value of the tagging efficiency and b-jet fraction, no E_T cut has been applied to the jets to take into account the threshold effects due to the hard cut on the transverse momentum of the initial hard-scattered partons in the Monte Carlo generation.

The total number of jets within the central rapidity range $|\eta| < 2.4$ is 1'906'577. Table 4.6 presents the results for both sub-cones sizes 0.4 and 0.7 for the JetClu algorithm for heavy flavour jets, while table 4.7 shows the average tagging efficiency and the tagged jets fraction for b-, c- and light- (and gluon-) jets.

sub-cone	N^b_{jets}	N^c_{jets}	N_{tag+}	N_{tag-}	N^b_{tag+}	N^b_{tag-}	N^c_{tag+}	N^c_{tag-}
0.4	37'545	106'773	12'078	1'196	7'202	117	3'397	147
0.7	37'545	106'773	16'265	2'157	9'660	167	4'019	226

TABLE 4.6: Number of b- (c-)jets for $|\eta| < 2.4$, number of tagged jets and tagged b-(c-)jets (positive and negative) for the JetClu algorithm with sub-cone sizes 0.4 and 0.7 from MC simulation.

sub-cone	$\epsilon^b_{tag} \pm$ stat.err.	$f^{truth}_b \pm$ sta.err.
0.4	0.192±0.002	0.602±0.009
0.7	0.257±0.003	0.593±0.008
sub-cone	ϵ^c_{tag}	f^{truth}_c
0.4	0.0321±0.006	0.281±0.005
0.7	0.0382±0.006	0.247±0.004
sub-cone	ϵ^{light}_{tag}	f^{truth}_{light}
0.4	$(8.4\pm0.2)\times10^{-4}$	0.122±0.003
0.7	$(1.47\pm0.03)\times10^{-3}$	0.159±0.003

ϵ^b_{tag} diff	f^{truth}_b diff
+34%	-1.7%
ϵ^c_{tag} diff	f^{truth}_c diff
+18.3%	-12.2%
ϵ^{light}_{tag} diff	f^{truth}_{light} diff
+74.8%	+30.0%

TABLE 4.7: Tagging efficiency and the true fraction of tagged b-, c- and light-jets. Light jets are all jets not containing heavy flavour, i.e gluon jets or *uds* quark jets.

The number of tagged b-jets - and consequently the b-tagging efficiency - increases on average by 35%, although the largest difference is for jets with $E_T < 30$ GeV. Figure 4.21 shows the ratio for b (c, light)-jets of the number of tagged jets for a sub-cone size of 0.7 over that for a sub-cone size of 0.4 ($N^{b/c/light\ 07}_{tag}/N^{b/c/light\ 04}_{tag}$) as a function of jet E_T. The increase in tagging efficiency for c-jets is lower, while the probability of mis-tagging is almost doubled, especially at low E_T. For comparison, the tag rate ratios for sub-cone size 0.7 over sub-cone size 0.4 for heavy flavour and light jets are shown in figure 4.22 for higher E_T jets. A dijet Pythia Tune A Monte Carlo sample with a minimum hard-scattered energy of 90 GeV is used. The increase in the b-tagging efficiency is below 10% for jets with $E_T > 50$ GeV, while the mistag probability, for sub-cone size of 0.7, is almost 50% higher than for sub-cone sizes of 0.4.

The fraction of b-tagged jets with respect to the overall number of tagged jets is compatible, within the statistical error, for the two sub-cone sizes. On the other hand, the fraction of tagged

c- (light-)jets decreases (increases) by 15 to 30% when increasing the sub-cone size from 0.4 to 0.7 . Figure 4.23 shows the tagged jet fraction as a function of jet E_T.

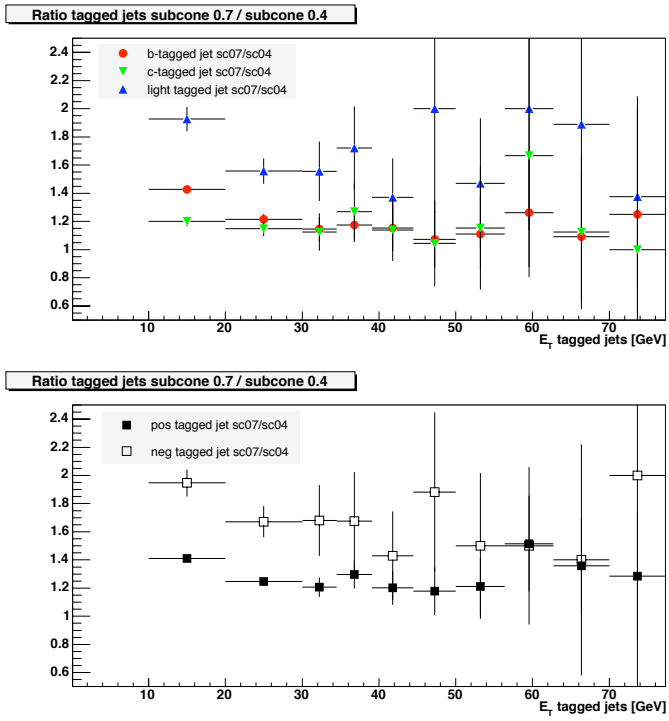

FIGURE 4.21: $N_{tag}^{b/c/light\ 07}/N_{tag}^{b/c/light\ 04}$ ratio as a function of tagged jet E_T (top). Ratio of the number of positively and negatively tagged jets with sub-cone sizes 0.7 and 0.4 (bottom).

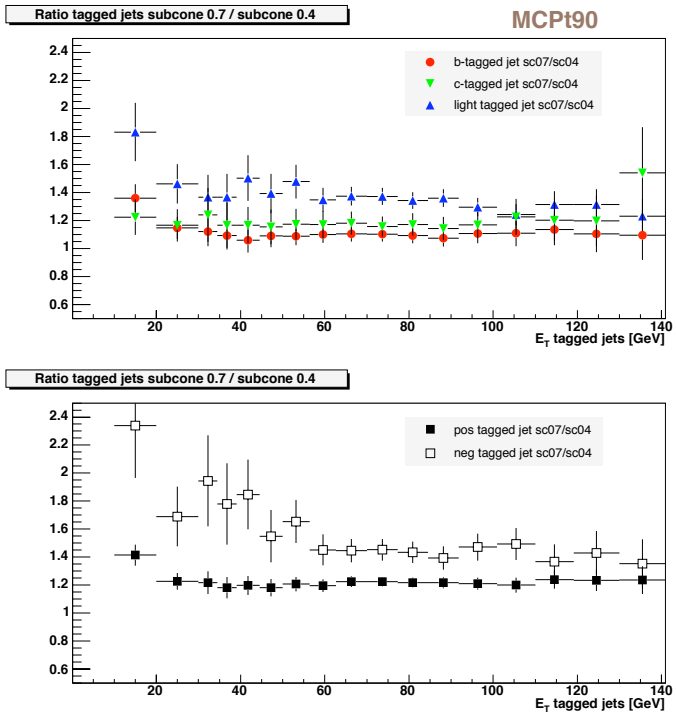

FIGURE 4.22: $N_{tag}^{b/c/light\ 07}/N_{tag}^{b/c/light\ 04}$ ratio as a function of tagged jet E_T, for Pt90 MC (top). Ratio of the number of positively and negatively tagged jets with sub-cone sizes of 0.7 and 0.4 (bottom).

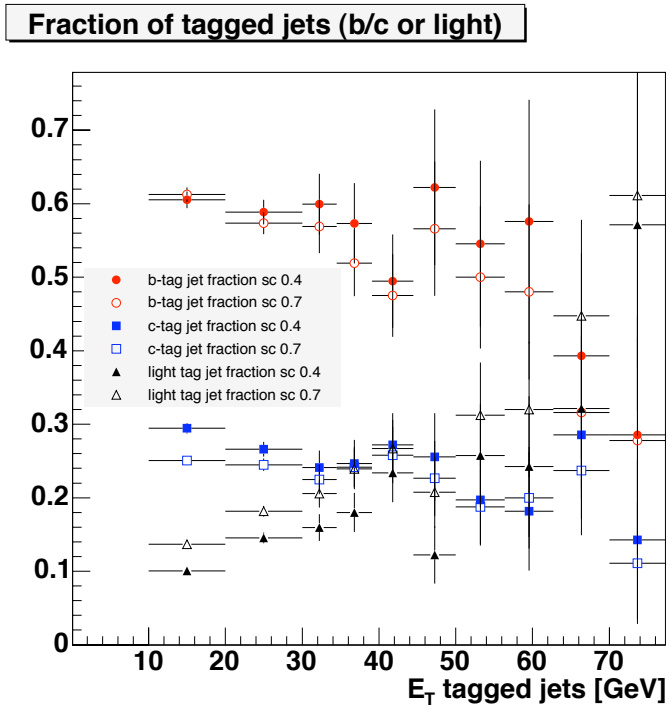

FIGURE 4.23: b, c and light tagged jets fraction as a function of tagged jet E_T for sub-cone size 0.4 (full marker) and 0.7 (empty marker).

These results show that the use of a wider cone opening angle for tracks used to reconstruct secondary vertices could be advantageous from a statistical point of view. Nevertheless, it would be necessary to carry out a more detailed analysis on the SecVtx algorithm in order to evaluate the impact on data of a higher non-heavy flavour jets tagging probability.

A re-optimisation of the tagging algorithm is beyond the purpose of this study. Nevertheless, a preliminary study has also been performed on data. Although it is unknown a priori, the true number of b-jets before tagging is independent of the SecVtx algorithm used. The following equation can thus be written

$$N_{\mathrm{b-jets}} \rightarrow N^{04}_{+\mathrm{tag}} \frac{f^{04}_{\mathrm{btag}}}{\epsilon^{04}_{\mathrm{btag}}} = N^{07}_{+\mathrm{tag}} \frac{f^{07}_{\mathrm{btag}}}{\epsilon^{07}_{\mathrm{btag}}} \times K \tag{4.3}$$

where K takes into account the difference in the scale factors between the two sub-cone sizes. Using the mass of the secondary vertices as a discriminating quantity, the b-jet content after tagging can be extracted (see section 6.3). An overall fit that considers all tagged jets is performed and mass templates for b- and nonb-jets are used. The resulting b-jet fraction is compatible, within the errors, for the two sub-cone sizes (see figure 4.24):

- 0.410±0.010 for sub-cone size 0.4
- 0.420±0.009 for sub-cone size 0.7

The values for the K-factor show that the difference between scale factors for sub-cone sizes 0.4 and 0.7 is of the order of 7%.

We conclude that, especially for soft jets with a transverse energy below 30 GeV, the use of an extended sub-cone size for secondary vertex reconstruction can be motivated by statistical gain. Nevertheless, due to the increase in the mis-tagging probability, this implies the need for

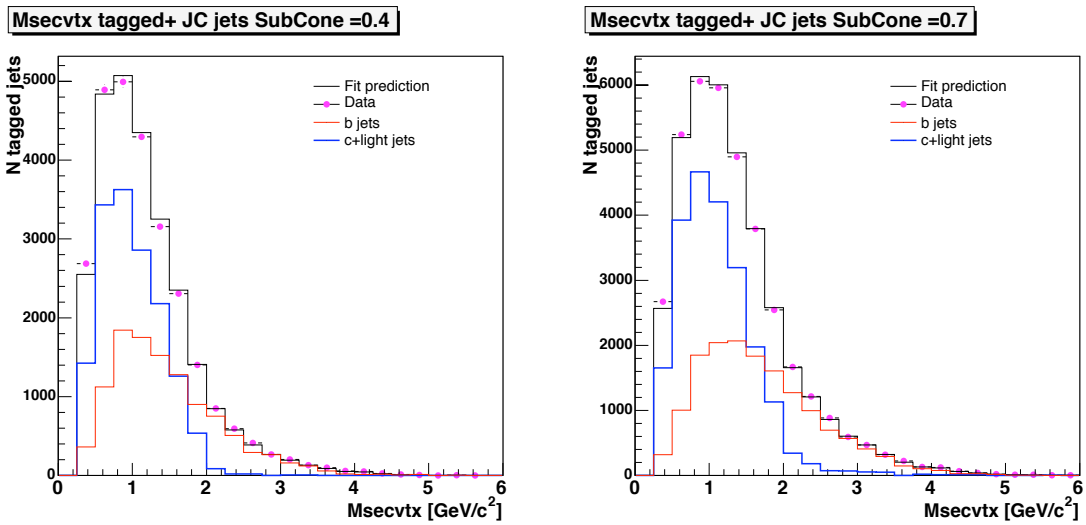

FIGURE 4.24: Mass of secondary vertex for sub-cone sizes 0.4 and 0.7, for data and Monte Carlo simulation.

a more careful study of the algorithm, especially concerning the scale factor and the related systematic uncertainties.

4.2.3 Conclusions on the SecVtx Algorithm Study

In this chapter, several aspects for the performance of the SecVtx algorithm were tested using different jet algorithms, cone sizes, sub-cone sizes and merging/splitting fractions. As long as the sub-cone size (the size of the cone in which the tracks which make up the secondary vertex can be found) is kept constant, the SecVtx algorithm and optimisation need not be modified. It it thus possible to use the SecVtx algorithm with MidPoint or JetClu, with a jet cone of sizes 0.4 or 0.7 and in the case of MidPoint with the old merge fraction, 50%, or the new one, 75%, without these changes affecting the performance of the secondary vertex tagger. The new merge fraction seems to agree better with JetClu than the older one. The merge fractions are now the same for both algorithms. This study was carried out using inclusive jet data, evaluating the tagging rates as function of global quantities. Previous studies performed using electron samples (data and Monte Carlo) that compare default JetClu cone size 0.4 jets with MidPoint cone size 0.7 jets ($f_{\mathrm{merge}} = 75\%$), have demonstrated that the b-tagging efficiencies and scale factors are consistent.

The effect of changing from a sub-cone of size 0.4 to one of 0.7 for the use with jet cones of size 0.7 has been investigated. It has been found that, although the tag rate increases, so does the mis-tag rate, leading to no obvious net gain in performance. Nevertheless, the use of a wider cone for tracks entering the secondary vertex algorithm leads to significantly higher statistics for soft jets (below 30 GeV). More studies would be needed to optimise the SecVtx algorithm and to carefully treat the larger mis-tag rate of the jets.

Chapter 5

Event Selection

In this chapter the description of the different datasets and MC samples is presented along with the event selection. First, in section 5.1 a description of the triggers is given for the different datasets used. Next, the different Monte Carlo samples used for this analysis will be described and in section 5.3 the selection cuts will be discussed. Section 5.4 discusses the average jet correction which corrects the p_T of the jets for detector and resolution effects.

5.1 Datasets

A sample of inclusive jet events is selected from $\sim 300\,\mathrm{pb}^{-1}$ of CDF Run II data collected from February 2002 through to September 2004. Four different datasets are used. These are called Jet20, Jet50, Jet70 and Jet100. The trigger paths for these datasets are similar, differing only by the cut thresholds and the nominal pre-scales. They are summarised in table 5.1. The basic idea of these triggers is to select events which have at least one jet which passes the trigger thresholds. There are no requirements on anything other than the jets.

At the trigger level 1, events are selected that have at least one calorimeter tower with the transverse component of the energy deposit above a nominal threshold (5 or 10 GeV). These trigger paths are called ST05 and ST10, respectively, where ST stands for single tower. A cut on the transverse energy of calorimetric clusters is then applied at level 2 (cut at 15, 40, 60 and 90 GeV for the Jet20, Jet50, Jet70 and Jet100 trigger paths, respectively). These triggers are called CL15, CL40, CL60 and CL90, where CL stands for cluster. A calorimetric cluster is defined as follows: starting from the seed towers, all towers whose p_T is above a certain threshold and that touch the seed tower are attached to the cluster. Then all towers, again passing the p_T cut, that touch the cluster are attached to the cluster. This iteration continues until the there are no more neighbouring calorimeter towers which have energy deposits above the p_T threshold. These are the stable calorimetric clusters.

Finally, at level 3, full jets are reconstructed using the JetClu cone algorithm with a cone of size 0.7 (see section 3). The jet algorithm is the same as the offline JetClu algorithm but at trigger level the primary vertex of the event is assumed to be at the centre of the detector, i.e. at (0,0,0) in the (Y, ϕ, z) coordinates of the detector. Cuts are applied on the transverse energy of the jets (20, 50, 70 and 100 GeV for the different datasets). The level 3 triggers for the different datasets are called J20, J50, J70 and J100, where J stands for jets.

Dataset	Level 1	Level 2	Level 3
Jet20	ST05	CL15	J20
Jet50	ST05	CL40	J50
Jet70	ST10	CL60	J70
Jet100	ST10	CL90	J100

TABLE 5.1: Summary of the trigger tables for the jet datasets.

5.2 Monte Carlo Samples

A number of different Monte Carlo (MC) samples are used both in the unfolding procedure of this analysis, presented in section 6, and as a final comparison with the data. The samples used are leading order MC simulations of the QCD $2 \rightarrow 2$ process that include initial and final state gluon radiation through parton showering.

The main samples used for the unfolding procedure of this analysis are generated using Pythia Tune A [1]. They use CTEQ5L PDFs and the default Lund fragmentation model [30][12]. Different samples were generated with different requirements on the minimum transverse momentum of the hard scattering; these are: Pt18, Pt40, Pt60, Pt90, Pt120, Pt150 and Pt200, where the number is the applied hard-scattering p_T cut. A p_T cut is needed on the hard-scattered process to avoid the effect of the divergence of the matrix element as $p_T \rightarrow 0$. The different cuts are used in order to have high statistics over a wide range of transverse momentum.

The fraction of b-jets in the inclusive jet samples is only of the order of 4%. In order to increase the b-jet statistics, samples were generated with the same transverse momentum cuts as above but which have an additional filter (HEPG filter) which selects only events which have at least one b-quark in the event. These samples are referred to as the b-filtered samples.

For comparison purposes and for systematic studies, samples generated using Herwig [31], with the same transverse momentum cuts as above, are also used. Only inclusive samples are available, not b-filtered ones.

To investigate the influence of the PDFs on the unfolding procedure as well as on the final shapes, a sample is generated using Pythia Tune A but changing the PDFs to CTEQ6L [32]. This sample was generated using a transverse momentum cut at 60 GeV and applying the additional b-quark filter to enhance the b-jet content.

Further samples are generated using Pythia Tune A but changing the fragmentation model used for heavy quarks to the Peterson model with the ϵ_b parameter set to 0.006 [33]. In the Peterson model, the fragmentation function for heavy quarks is described by

$$D_{\mathrm{Q}}^{\mathrm{b,c}}(z) = \frac{N}{z(1 - \frac{1}{z} - \frac{\epsilon_{\mathrm{b,c}}}{(1-z)})}. \tag{5.1}$$

The most likely value for the ϵ_b parameter is 0.006 ± 0.002 at the Tevatron [34]. Two samples of this type were generated with transverse momentum cuts at 18 and 120 GeV and with the additional b-quark filter. Table 5.2 shows a summary of all the MC samples used.

[1] The different parameters of Pythia are tuned to the CDF Run I underlying event.

Pythia Tune A inclusive	Pythia Tune A HEPG b-filter	Herwig inclusive	Pythia Tune A CTEQ6L HEPG b-filter	Pythia Tune A Peterson frag. HEPG b-filter
Pt18	Pt18	Pt18		Pt18
Pt40	Pt40	Pt40		
Pt60	Pt60	Pt60	Pt60	
Pt90	Pt90	Pt90		
Pt120	Pt120	Pt120		Pt120
Pt150	Pt150	Pt150		
Pt200	Pt200	Pt200		

TABLE 5.2: Summary of the MC samples used.

5.3 Event Selection

In this analysis, jets are reconstructed using the MidPoint cone algorithm with a cone size of 0.7 and a splitting/merging fraction of 75% (see chapter 3). Events are selected which have at least one jet in the central rapidity region ($|Y_{\text{jet}}| \leq 0.7$).

For this analysis, it is necessary to ensure that, when the data were taken, not only the calorimeters were functioning properly but also that the silicon detectors were on and functioning properly. At CDF, each run is assigned something like a "quality stamp". This has quality bits for each potential type of analysis requirement. In this analysis, the requirement is that the silicon detector and the QCD bits are set [2].

A cut is applied on the total missing E_T significance of the event to remove a large fraction of the cosmic background [3]. It must be below a certain threshold that is dependent on the dataset used. These thresholds are: 3.5, 5.0, 6.0 and 7.0, respectively, for Jet20, Jet50, Jet70 and Jet100, the same as those used for most other QCD analyses [29][35].

This analysis considers all jets which are in the central rapidity region where the secondary vertex tagging algorithm is the best understood. A cut on $|Y_{\text{jet}}| \leq 0.7$ is thus applied.

To ensure good secondary vertex reconstruction, a cut is required on the z-component of the primary vertex which must be within 50 cm of the centre of the detector. To remove any potential effects due to multiple parton interactions on the final jets, all events with multiple primary vertices were removed.

As mentioned previously, the b-jet content of the jet samples is relatively low, below 4%. It would therefore be impossible to make any precision measurement of b-jet properties with such a low purity. It is thus necessary to require that the jets be positively tagged by the SecVtx tight tagger described in chapter 4. This increases the b-jet fraction to $20 - 40\%$ for high and low p_T jets, respectively, as shown in section 6.3. The requirement that the jets be tagged introduces a bias in the selected sample. Nevertheless, this can be accounted for in this analysis, as will be shown in chapter 6. It should be noted that the sub-cone inside which tracks are considered by the SecVtx algorithm is maintained at the default value of 0.4 [28].

[2] Version 7 of the good run list is used.

[3] The missing E_T significance is defined as $\frac{E_T^{\text{mis}}}{\sqrt{\sum E_T}}$, where E_T^{mis} and $\sum E_T$ are the missing and total transverse energy of the event, respectively.

The following list summarises all the cuts applied to the jets in the samples, after requiring the runs to pass the quality cuts

- One (and only one) primary vertex with $|Z_{\text{vtx}}| \leq 50$ cm,
- Missing E_T significance larger than a given threshold (dependent on the dataset),
- $|Y_{\text{jet}}| \leq 0.7$,
- SecVtx tight tag on jets (for the tagged sample only).

5.4 Average p_T Correction

Jets, therefore also jet shapes, are computed at the level of calorimeter towers but it is important to obtain results that are detector independent. Jet properties are affected by the detector response. The detector resolution is not perfect and is not necessarily the same over the whole ϕ and η region.

For this reason it is necessary to correct the p_T of the jets for all detector effects. This correction is called the hadron level correction. It corrects the p_T of the jets back to what they are expected to have been before the particles entered the detector. This correction factor is calculated in this analysis by matching, in MC, hadron level and calorimeter jets based on an angular selection in the $Y - \phi$ plane as detailed in [29]. For each tagged calorimeter jet in the range $|Y_{\text{jet}}| < 0.7$, the angular difference, ΔR, between that jet and each of the hadron level jets in the event is computed. The hadron level jet associated to the calorimetric jet is taken to be the one with the smallest ΔR (as long as this difference is smaller than 0.7, the jet cone radius). A scatter plot of calorimetric jet p_T versus hadron level jet p_T was produced for the whole p_T range covered by the jet datasets and is reproduced in figure 5.1. The scatter plot is fitted to a fourth order polynomial. The fit results give a correction function

$$p_T^{\text{corr}} = 0.5 + 1.24\,p_T - 0.0012\,p_T^2 + 4 \cdot 10^{-6}\,p_T^3 - 5 \cdot 10^{-9}\,p_T^4 \tag{5.2}$$

where p_T is the measured calorimetric transverse momentum and p_T^{corr} is the corrected transverse momentum of the jet. This correction function is shown on the right hand side of figure 5.1. The corrections are of the order of 20% at low p_T, down to 10% at higher p_T. The last two terms of this correction function are only important for very large p_T jets.

Except where explicitly mentioned, the binning for this analysis and for all plots shown is done in corrected p_T, from now simply referred to as p_T, to enable a better comparison with other experimental measurements or theoretical models.

5.5 p_T Thresholds

Because of its nature, the trigger is not fully efficient at the trigger threshold and the behaviour close to the trigger threshold might not be well understood. A similar effect is seen in the MC samples where there is a turn-on effect.

It is therefore necessary to consider each dataset only in the p_T region where the trigger effects are negligible. This is assumed to be the case when the trigger efficiency is above 99%. The dataset thresholds are thus taken to be the same as those used in [29]. The trigger efficiency curves as a function of raw, calorimeter level, p_T are shown in figure 5.2. These plots are

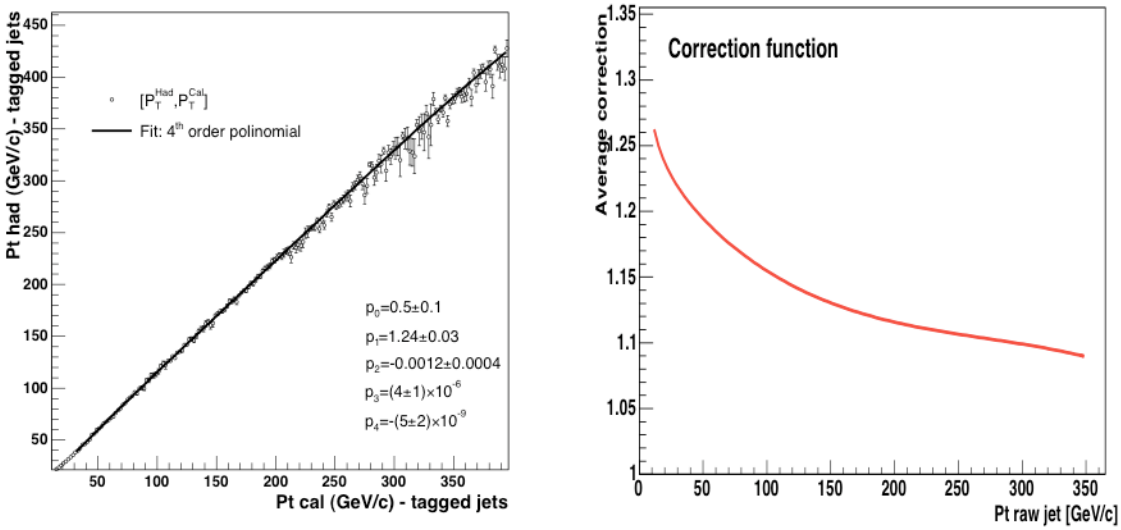

FIGURE 5.1: Scatter plot of the hadron level vs calorimetric jet p_T for a wide range of p_T (left). The correction function applied is shown in the right hand plot. These plots are taken from [29].

Dataset	p_T^{raw} limit [GeV/c]	p_T^{corr} limit [GeV/c]
Jet20	43	52
Jet50	67	80
Jet70	89	104
Jet100	123	142

TABLE 5.3: p_T threshold for the different datasets in raw and corrected p_T.

taken from [29]. Table 5.3 shows the thresholds in raw and corrected p_T for the different datasets.

In MC, there is a similar problem to that of the trigger efficiency. It is due to the cut on the minimum p_T of the hard-scattered partons. The jet energy will not be the same as the energy of the hard-scattered partons; it is therefore necessary to make sure that the jets are not biased by this effect. The final binning of this analysis is more conservative than the one used for the b-jet cross section measurement [29]. Only one MC sample is used for each data bin. The ranges in p_T of the four final bins, along with the datasets and MC samples used, are shown in table 5.4. Also shown in this table are the total number of inclusive and tagged jets in data that pass all the selection cuts.

p_T limits [GeV/c]	Dataset	MC sample	N_{jets}	$N_{\text{tagged jets}}$
52-80	Jet20	Pt18	161'524	4'677
80-104	Jet50	Pt40	354'922	13'367
104-142	Jet70	Pt60	134'907	5'874
142-300	Jet100	Pt90	377'650	18'673

TABLE 5.4: Final binning in jet p_T, showing the datasets and MC samples used for each p_T bin. Also shown are the total number of jets and the total number of tagged jets in data which pass all the selection cuts.

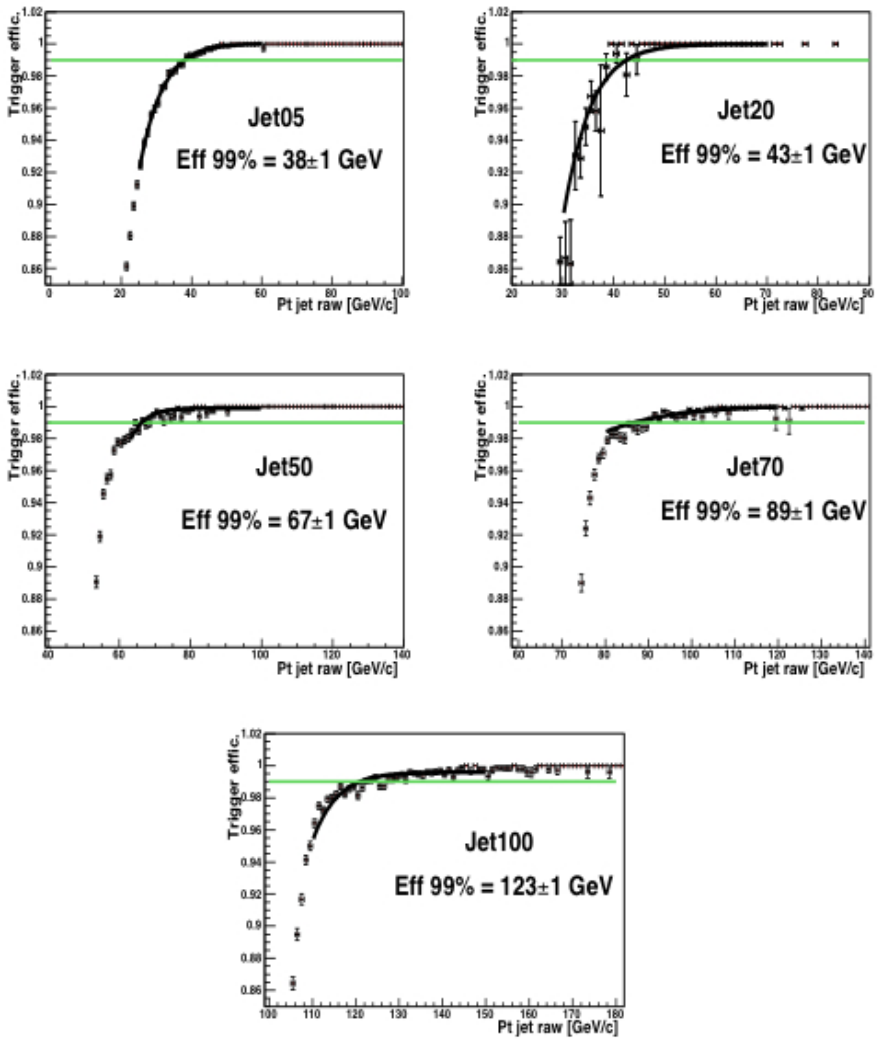

FIGURE 5.2: Trigger efficiency as a function of raw jet p_T for the different jet datasets, showing as a horizontal green line the 99% threshold used. The figures are taken from [29].

Chapter 6

b-quark Jet Shapes

The subject of this analysis is the study of b-quark jet shapes. As described in section 1.3, a b-jet is defined as a jet which contains at least one b-quark. These jets are mainly flavour creation jets (for the cases where there is one b-quark inside the jet) and jets from gluon splitting processes where the gluon splits into a $b\bar{b}$ pair. For the latter case, the two b-quarks are most of the time expected to be inside the same jet cone, as was shown in section 1.3. The fraction of b-quark jets from gluon splitting has not been measured at CDF Run II. It is important measure this fraction experimentally in order to check if the MC models describe the interplay between the initial state gluon production, where the gluon splits into a $b\bar{b}$ pair, and the initial state heavy flavour production. In section 7.3, a plot comparing the Pythia Tune A predictions to a next-to-leading order calculation shows that the gluon splitting rate is most likely underestimated in Pythia Tune A. The shapes of b-quark jets, in MC, show a significant dependence on the relative fraction of b-quark jets containing one or two b-quarks inside the same jet cone as was shown towards the end of chapter 1. The measurement of b-quark jet shapes is thus a good test to verify the modelling of the gluon splitting process in MC. This is particularly useful for extrapolations of such processes to LHC energies, where a good understanding of b-jets is necessary for many searches for new particles.

In chapter 1.3 jet shapes were introduced. In practise, the average jet shapes are measured, where the average is taken over all jets in the samples. Equation 1.15 for the integrated shapes becomes

$$\Psi(r) = \frac{1}{N_{\text{jets}}} \sum_{\text{jets}} \frac{p_T(0,r)}{p_T(0,R)}. \tag{6.1}$$

For the differential shape, equation 1.16 becomes

$$\rho(r - \Delta r/2, r + \Delta r/2) = \frac{1}{N_{\text{jets}}} \sum_{\text{jets}} \frac{p_T(0, r + \Delta r/2) - p_T(0, r + \Delta r/2)}{p_T(0,R)\,\Delta r}, \tag{6.2}$$

where Δr is the bin size in r. It is not possible to obtain, in data, even with good tagging, a sample which consists of only b-jets. It is not even possible to obtain a sample containing only heavy flavoured jets. For this reason, it is necessary to extract the jet shapes of b-quark jets from a background of light-, gluon- and c-jets. This can only be done statistically.

For the remainder of this thesis, the term "raw" will be used for quantities measured at the level of the detector, using calorimetric or track information, before any correction factors are applied. The terms "corrected" or "hadron level" are used when the effects of the detector have been taken into account or, in MC, when the quantities are calculated directly using the final set of particles, after the fragmentation and hadronisation processes. Hadron level or corrected

quantities are therefore detector independent.

In the next section, the unfolding procedure is presented and all its different parameters are explained. The raw shapes for both inclusive and tagged jets are then shown. Section 6.3 details the method used to calculate the purity of the sample. Section 6.4 describres the bias corrections used to correct for the effect of applying the SecVtx tight tagger. In section 6.5 the hadron level corrections to the jet shapes are shown. These are the correction factors applied to the b-jet shapes in order to correct the shapes for detector effects. Finally, section 6.6 presents the hadron level b-quark jet shapes showing only the statistical errors. Chapter 7 presents the systematic studies carried out for this analysis. Chapter 8 presents the final results for the b-quark jet shapes.

6.1 Unfolding Method

As mentioned above, the sample of tagged jets used for this analysis does not contain only b-quark jets but also a background of jets that don't contain any b-quarks; these are called nonb-jets. The measured shape, after tagging, will thus be a combination of the b-jets and nonb-jets which can be written as

$$\Psi_{\mathrm{meas}}(r) = p_{\mathrm{b}}\Psi^{\mathrm{b}}_{\mathrm{meas}}(r) + (1 - p_{\mathrm{b}})\Psi^{\mathrm{nonb}}_{\mathrm{meas}}(r), \tag{6.3}$$

where p_{b} is the fraction of jets which contain at least one b-quark; this is called the purity. The superscripts b and nonb represent b- and nonb-jets, respectively, and the shapes are the raw, measured shapes after tagging. Although the jet algorithm is only run at calorimeter level, the jet shapes can be measured using either the calorimetric towers or the tracks that are inside the jet cone.

The fact that a SecVtx tagged jet sample is used most probably biases the measured shapes. These biases could be due to the fact that the SecVtx algorithm requires jets with cleaner, better defined tracks which could lead to cleaner jet shapes with fewer soft particles. This bias is most likely different for b-jets than for nonb-jets. A bias term, which depends on the distance from the jet axis, r, must thus be added to correct for this effect in order to extract the unbiased detector level shapes for b- and nonb-jets. These bias terms, $b^{\mathrm{b}}(r)$ and $b^{\mathrm{nonb}}(r)$ for b- and nonb-jets respectively, are defined such that

$$\Psi^{\mathrm{b}}_{\mathrm{meas}}(r) = b^{\mathrm{b}}(r)\Psi^{\mathrm{b}}_{\mathrm{det}}(r), \tag{6.4}$$

and

$$\Psi^{\mathrm{nonb}}_{\mathrm{meas}}(r) = b^{\mathrm{nonb}}(r)\Psi^{\mathrm{nonb}}_{\mathrm{det}}(r), \tag{6.5}$$

where the $\Psi_{\mathrm{det}}(r)$ terms represent the detector level shapes, before any tagging requirements. Combing equations 6.3 through 6.5 gives for the measured shape, after tagging,

$$\Psi_{\mathrm{meas}}(r) = p_{\mathrm{b}}b^{\mathrm{b}}(r)\Psi^{\mathrm{b}}_{\mathrm{det}}(r) + (1 - p_{\mathrm{b}})b^{\mathrm{nonb}}(r)\Psi^{\mathrm{nonb}}_{\mathrm{det}}(r). \tag{6.6}$$

This equation can be re-written, extracting the b-jet shape, as

$$\Psi^{\mathrm{b}}_{\mathrm{det}}(r) = \frac{\Psi_{\mathrm{meas}}(r) - (1 - p_{\mathrm{b}})b^{\mathrm{nonb}}(r)\Psi^{\mathrm{nonb}}_{\mathrm{det}}(r)}{p_{\mathrm{b}}b^{\mathrm{b}}(r)}. \tag{6.7}$$

It is important that no measured quantity is influenced by the particular experimental setup. This is particularly important when comparing results with theoretical models or to other experimental results. For this reason, it is necessary to correct the b-jet shapes back to hadron

level, i.e. to remove all influence on the measurement of the tracker or calorimeters. This is in part done by an average correction to the p_T of the jets, discussed in section 5.4. It is also important to correct the actual jet shapes back to hadron level. This removes, among other things, the influence of the magnetic field and the calorimeter response on the jet shapes. Correction factor, $C_{\mathrm{had}}(r)$, are thus applied to extract the b-jet shapes at hadron level. Equation 6.7 can be written in it's final form as

$$\Psi^{\mathrm{b}}_{\mathrm{had}}(r) = C_{\mathrm{had}}(r)\frac{\Psi_{\mathrm{meas}}(r) - (1-p_{\mathrm{b}})b^{\mathrm{nonb}}(r)\Psi^{\mathrm{nonb}}_{\mathrm{det}}(r)}{p_{\mathrm{b}}b^{\mathrm{b}}(r)} \tag{6.8}$$

This last equation is the equation used to unfold the hadron level b-jet shapes from the raw tagged and inclusive jet shapes. Appendix A contains a discussion into the reasons why it was not possible, in this analysis, to measure the differential b-quark jet shapes using a similar unfolding method.

It is apparent from the above discussion that in order to measure the shapes of b-quark jets it is necessary to measure a number of other parameters. These will be discussed individually in the following sections. The different terms are

- $\Psi_{\mathrm{meas}}(r)$: the measured shapes in data, after tagging
- $\Psi^{\mathrm{nonb}}_{\mathrm{det}}(r)$: the detector level shapes of nonb-jets, without any tagging
- p_{b}: the b-jet purities, after tagging
- $b^{\mathrm{b}}(r)$: the biases on the b-jets due to the tagging requirement
- $b^{\mathrm{nonb}}(r)$: the biases on the nonb-jets due to the tagging requirement
- $C_{\mathrm{had}}(r)$: the hadron level corrections to the b-jet shapes

6.2 Raw Shapes

It is necessary for this analysis to measure, in data, the shapes of the tagged jets as well as the shapes of the nonb-jets. The shapes are measured at calorimeter level. The shapes measured using tracks are used in an independent measurement of the hadron level b-quark jet shapes, to get an idea of the systematic effect linked to the response of the calorimeters to low momentum particles.

The tagged shapes are straightforward to measure as they are the average shapes for all the tagged jets in the samples, measured at calorimeter level. Figure 6.1 shows the integrated jet shapes for tagged jets, as measured in data (black points) and compared to the Pythia Tune A MC predictions for each of the four p_T bins. The data and Pythia Tune A MC jet shapes are not expected to be the same because of the different fractions of b-quark jets in data and MC, as will be shown in section 6.3.

Given the very low fraction of b-jets in inclusive jet production, less than 4%, it is possible to approximate the nonb-jet shapes to those of the inclusive jet shapes, before any tagging requirements. The assumption that

$$\Psi^{\mathrm{nonb}}_{\mathrm{det}}(r) \approx \Psi^{\mathrm{incl}}_{\mathrm{det}}(r) \tag{6.9}$$

is used. The difference between these shapes, in Pythia Tune A MC, is negligible as can be seen from figure 6.2 and 6.3. The latter shows the ratio of the nonb over the inclusive shapes, which is

compatible with unity with a maximum deviation of less than 0.5%. No systematic uncertainty is therefore related to the use of this approximation. Figure 6.4 shows the integrated shapes for inclusive jets, as measured in data (black points) and compared with the Pythia Tune A MC predictions for each of the four p_T bins. The agreement between data and Pythia Tune A MC is not perfect, especially for the lowest p_T bin and the first bins in r. No reasonable explanation for this difference has been found so far. The effect of the difference in the inclusive jet shapes between data and Pythia Tune A MC on the hadron level b-quark jet shapes will be further investigated in chapter 8. Figure 6.5 shows, as black open squares, the raw inclusive shapes and as red triangles the raw tagged shapes. It is the difference between the inclusive and the tagged jet shapes that is needed in order to be able to expect a reasonable precision in the final measurement.

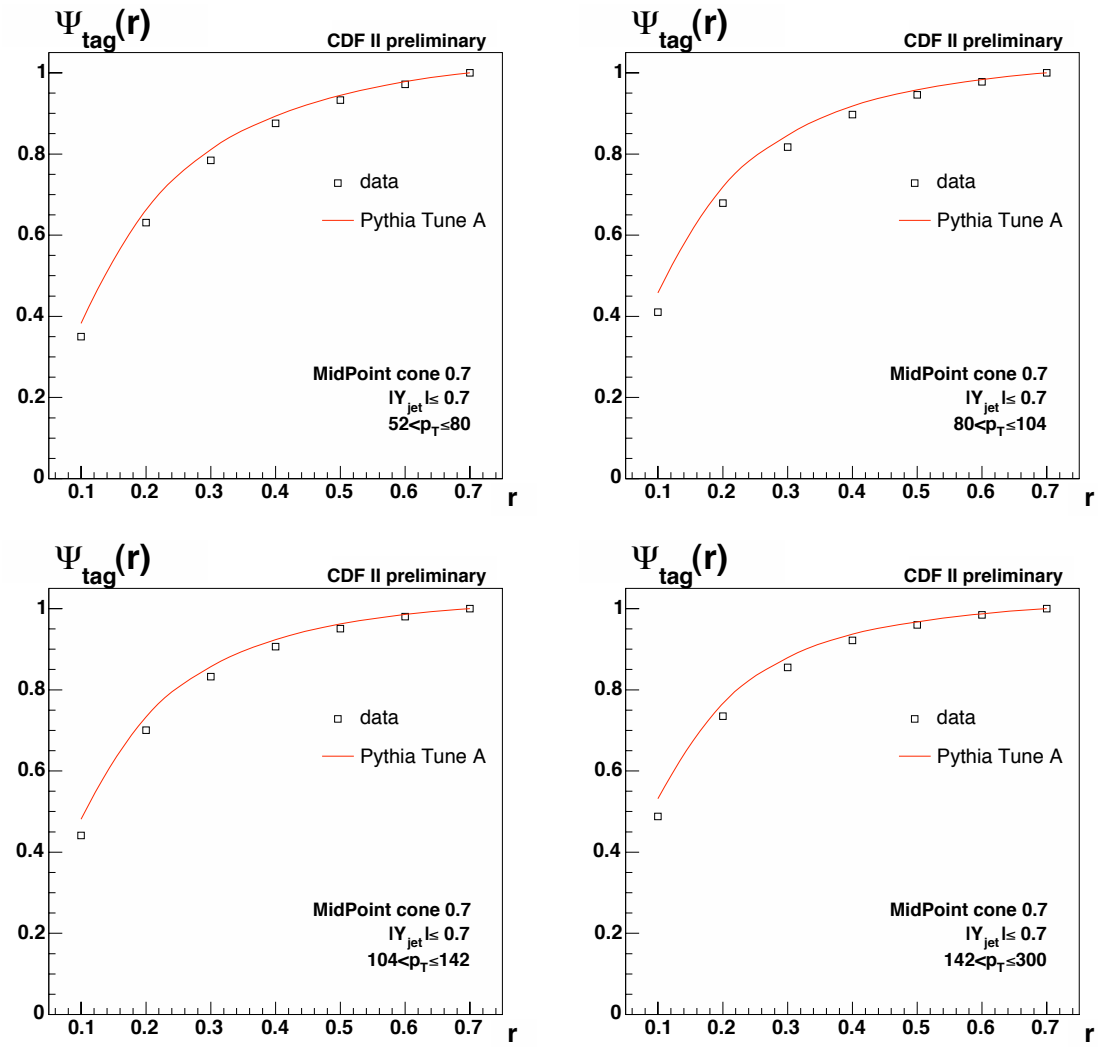

FIGURE 6.1: Raw integrated shapes for tagged jets. The Pythia Tune A MC predictions (red line) are compared to the data (black points). Only the statistical errors are shown. These are smaller than the data points in these plots.

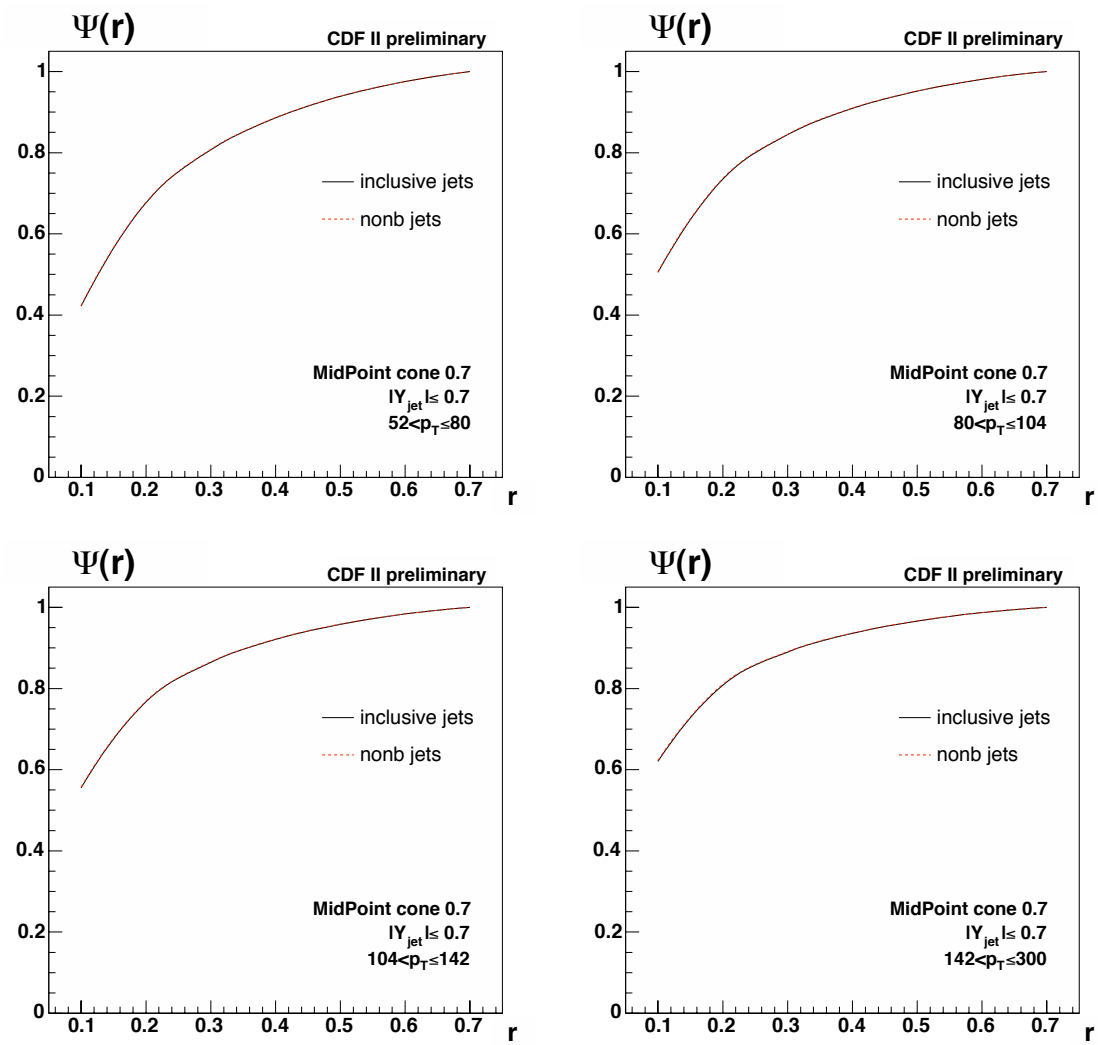

FIGURE 6.2: Raw integrated shapes, in Pythia Tune A, for inclusive (black full line) and nonb-jets (red dashed line).

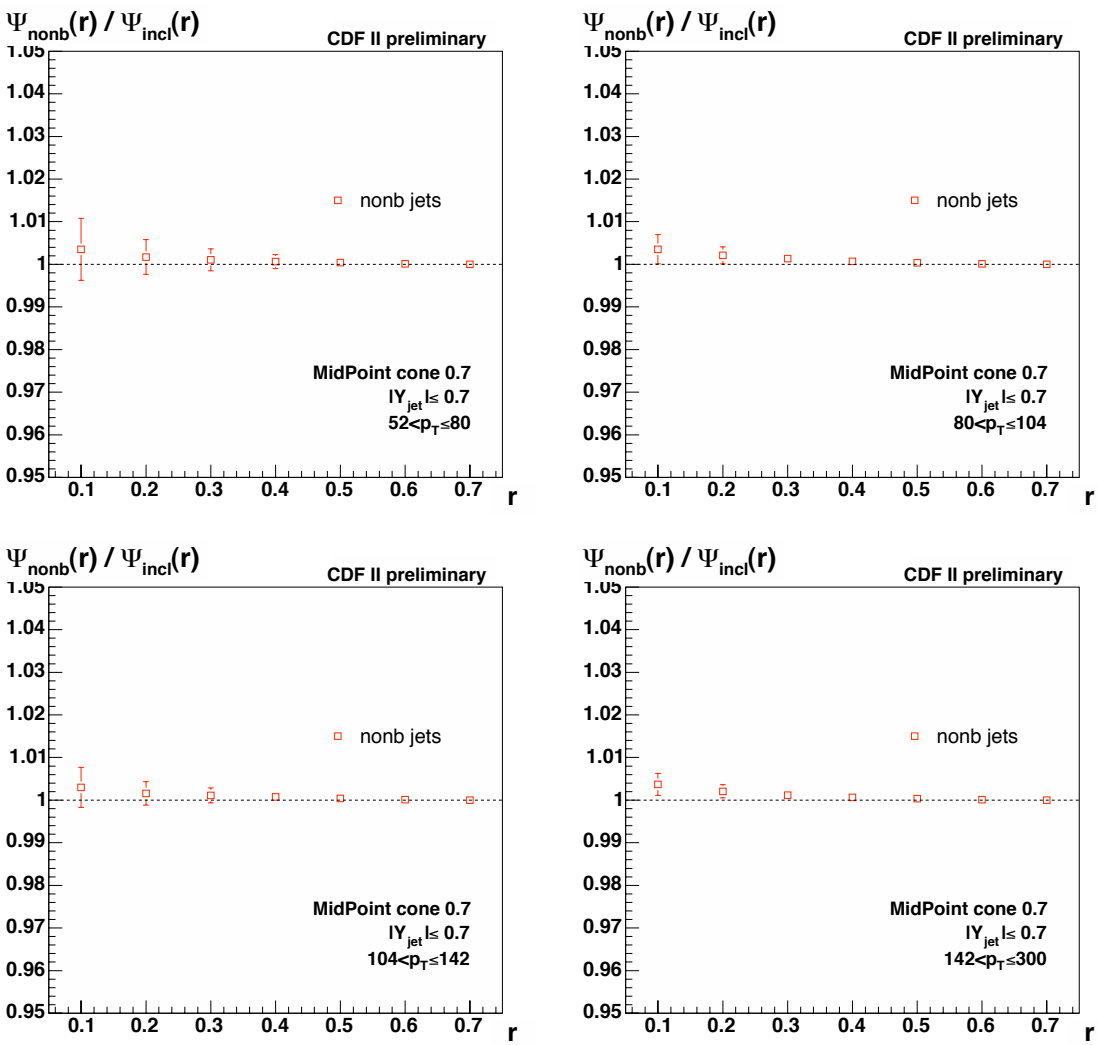

FIGURE 6.3: Ratio of the raw nonb-jet integrated shapes, in Pythia Tune A, over the inclusive integrated shapes (red points). Only the statistical errors are shown.

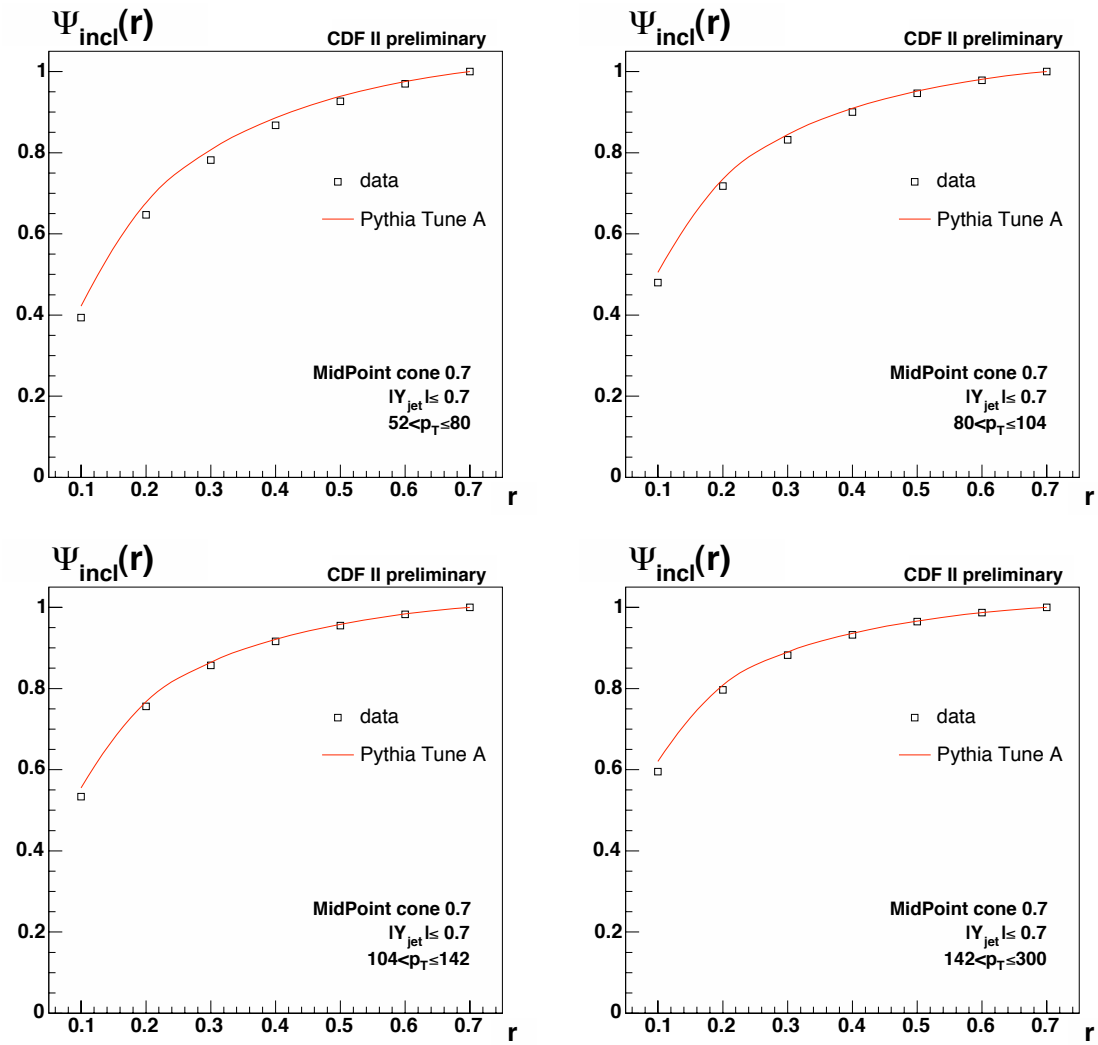

FIGURE 6.4: Raw integrated shapes for inclusive jets. The Pythia Tune A MC predictions (red line) are compared to the data (black points). Only the statistical errors are shown.

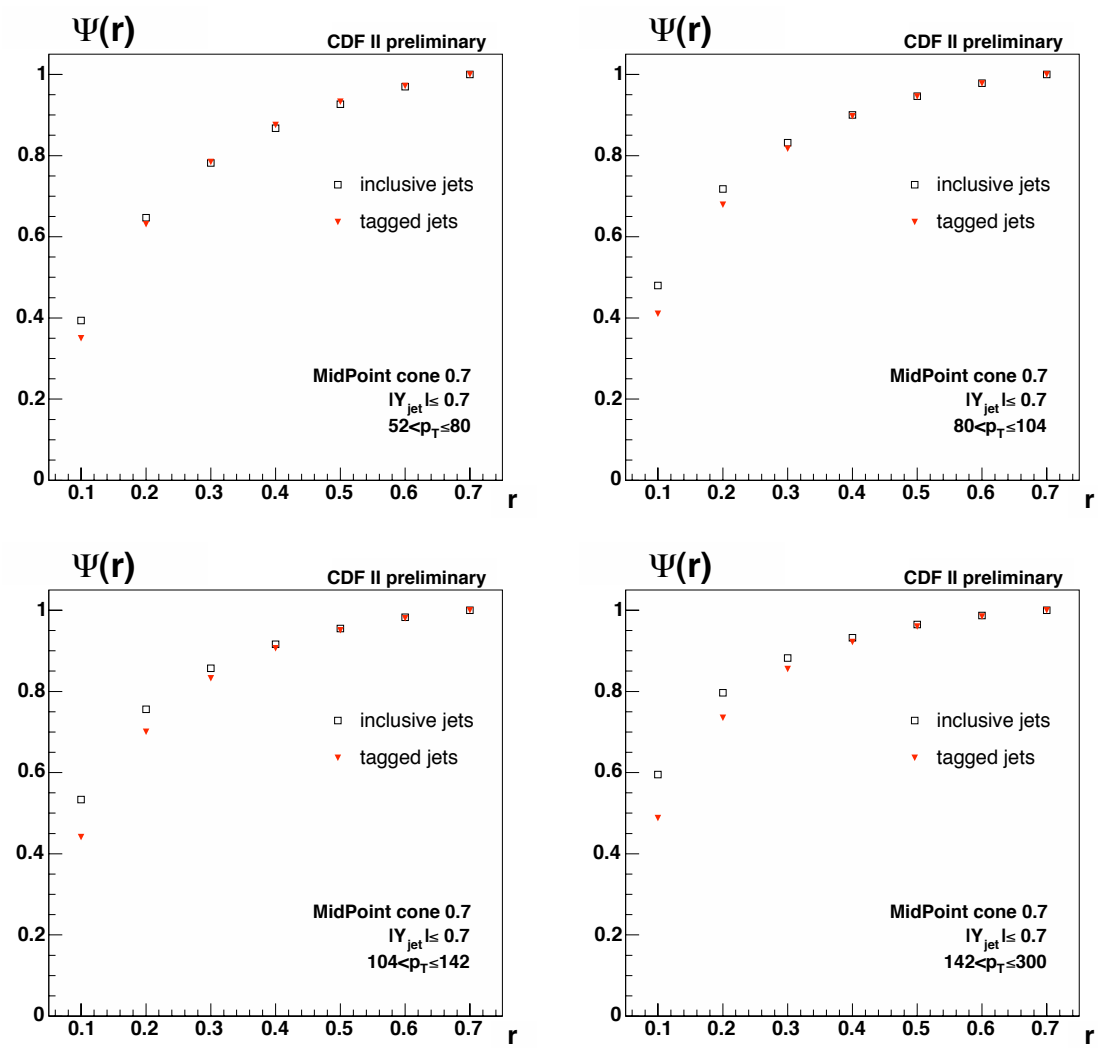

FIGURE 6.5: Raw integrated shapes for inclusive (black open squares) and tagged jets (red full triangles).

6.3 Purity

The fraction of tagged jets which are b-jets, the purity, is extracted from a fit to the secondary vertex mass distributions for b- and nonb-jets [29]. It is not possible to reconstruct the full hadron invariant mass due to the presence of neutral particles which are not detected in the tracking detectors and also due to the detector resolution. Nevertheless, the distribution of the invariant mass of the tracks used to find the secondary vertex, referred to as the secondary vertex mass, is significantly different for heavy flavoured jets and for light flavoured or gluon jets and is also different for b-quark jets and c-quark jets. Using the MC samples, distributions of the secondary vertex masses for tagged jets are obtained for each p_T bin, separately for b- and nonb-jets. These distributions, known as templates, are renormalised to unit area. The errors on each bin of the templates correspond to the inverse of the square root of the number of entries in that bin. Templates are obtained from the Pythia Tune A MC samples for the b-jets from the b-filtered samples and for the nonb-jets from the inclusive samples. These templates are shown in figure 6.6. The measured distribution in data of the secondary vertex masses is fitted to the b- and nonb-templates, using a binned χ^2 minimisation method [1], to find the most likely fraction of jets that are b-jets. The stability of the fits is tested with respect to changes in the fit range and changes in the number of bins used to cover this fit range. The stability is found to be very good. Figure 6.7 shows the distributions in data along with the fitted distributions for each p_T bin. This shows that the fit describes the data well. Figure 6.8 shows the fit results for p_{b} as a function of the p_T of the jets along with the values in MC. The errors shown in this figure are the statistical errors on the purity (see section 7.1). Table 6.1 shows the results for the purity in both data and Pythia Tune A MC for all the bins.

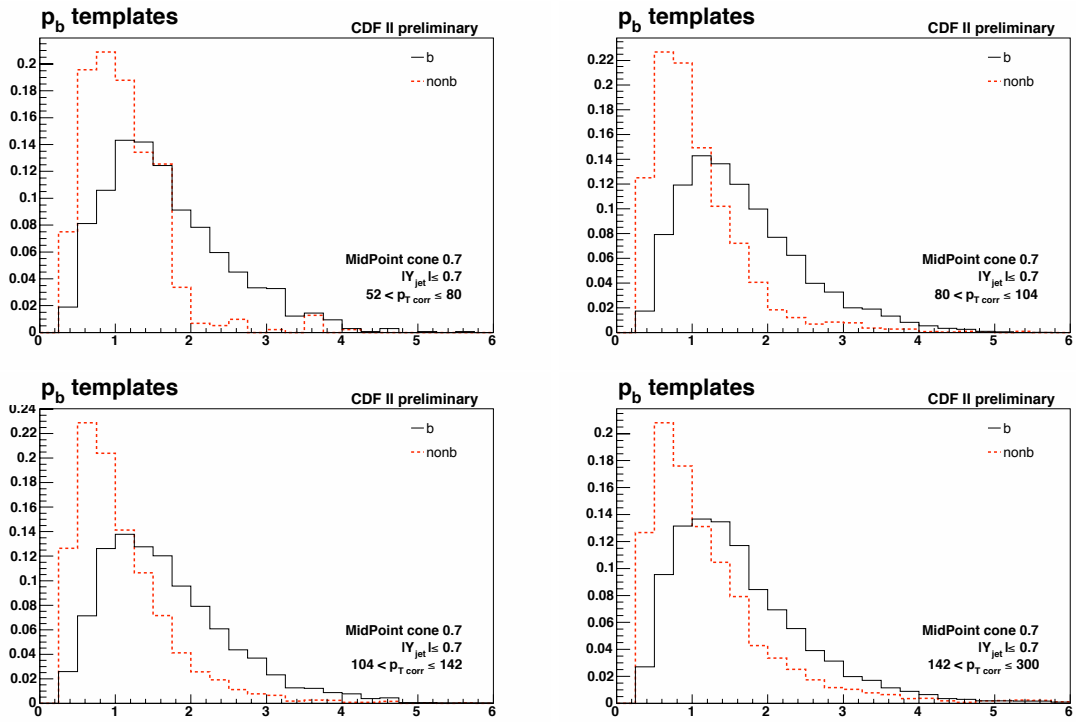

FIGURE 6.6: Normalised secondary vertex mass distributions for b- (black full line) and nonb-jets (red dashed line) for all the p_T bins.

[1] This is done using the ROOT routine TFractionFitter

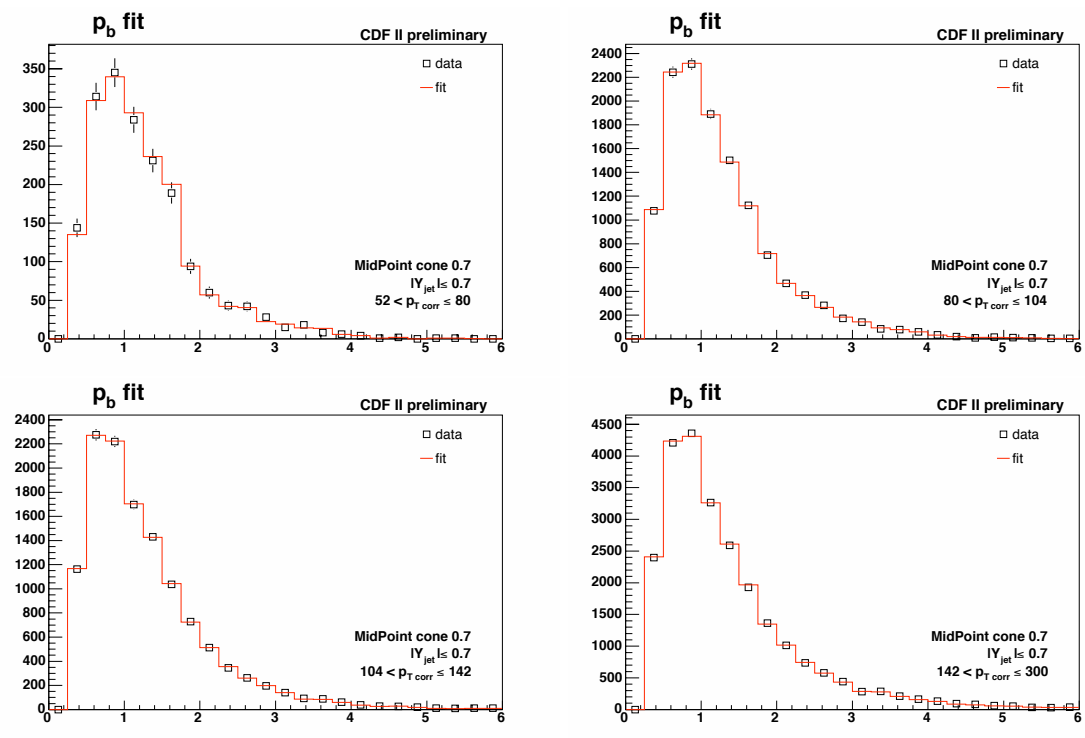

FIGURE 6.7: Secondary vertex mass distribution in data (black points) compared to the fitted distribution (red line) for all the p_T bins.

p_T range	p_{b}	p_{b} MC
52-80	0.320 ± 0.018	0.519
80-104	0.338 ± 0.010	0.448
104-142	0.300 ± 0.010	0.368
142-300	0.242 ± 0.009	0.304

TABLE 6.1: b-jet purity in data, as obtained from the fit results for each p_T bin, shown together with the Pythia Tune A MC values. The errors quoted correspond to the statistical errors only.

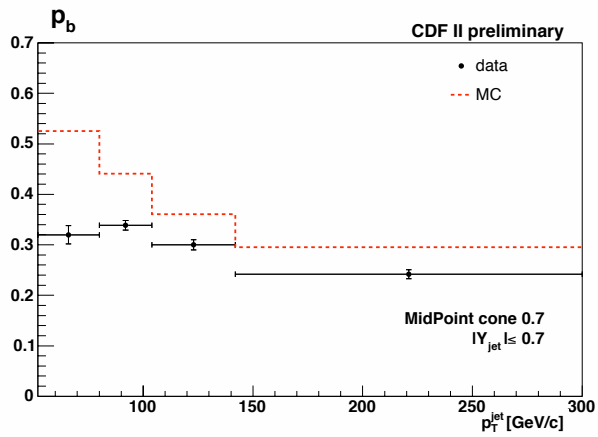

FIGURE 6.8: b-jet purity in data (points) as a function of jet p_T, shown along with the values obtained from MC (dashed red line). The error bars indicate the statistical errors only.

6.4 Biases Due to SecVtx Tagging

The requirement that the jets must be tagged by the SecVtx tight tagger introduces a bias in the measured shapes. These biases must be corrected for. They are different for each p_T bin, each bin in r and different for b- and nonb-jets. The bias terms are defined as the ratios, as obtained from MC, between the tagged and the inclusive jet shapes for b- and nonb-jets separately. They are thus defined as

$$b_{\mathrm{b}}(r) = \frac{\Psi^{\mathrm{tag}}_{\mathrm{b\,MC}}(r)}{\Psi^{\mathrm{incl}}_{\mathrm{b\,MC}}(r)} \tag{6.10}$$

for b-jets and

$$b_{\mathrm{nonb}}(r) = \frac{\Psi^{\mathrm{tag}}_{\mathrm{nonb\,MC}}(r)}{\Psi^{\mathrm{incl}}_{\mathrm{nonb\,MC}}(r)} \tag{6.11}$$

for nonb-jets. A bias term larger than unity for the low r bins implies that the shapes are narrower after tagging than before. Similarly, a bias term smaller than unity for the low r bins implies that the shapes get wider after the tagging requirement is applied.

Figure 6.9 shows the bias corrections, as a function of r, for b-jets (black open squares) for all the p_T bins considered. Only the errors due to the MC statistics are shown. Also shown, as red lines, are the biases for single and double b-quark jets (called 1b- and 2b-jets). The maximum bias for b-jets is of the order of 7%. The fact that the biases due to tagging on b-jets are sometimes somewhat smaller than both the biases due to tagging on the single or the double b-quark jet shapes can be understood by the fact that there are two competing effects. The first is that the tagging on b-quark jets leads to narrower jets, the tagging algorithm selects preferentially events where the heavy flavoured quarks are close to the jet axis. The second is that the tagging efficiencies are not the same for single and double b-quark jets. It is expected, from MC studies, that is is more likely to tag a double b-quark jet than a single b-quark jet. Thus the fraction of b-quark jets which have two b-quarks inside the same jet cone is significantly higher after tagging than before tagging. This is shown in figure 6.10. Double b-quark jet shapes being broader on average than single b-quark jets, the b-quark jets after tagging are expected to get broader under this effect. These two competing effects: selection of narrower jets by the algorithm but selection of a larger fraction of broader double b-quark jet shapes, tend to cancel each other out to a certain extent, leading to smaller biases.

Figure 6.11 shows the bias corrections, as a function of r, for nonb-jets (black open squares) for all the p_T bins considered. Only the errors due to MC statistics are shown. Also shown are the biases for c- and light+gluon-jets (as red lines). The maximum bias for nonb-jets is of the order of 18%. It is immediately apparent from this plot that the errors on the tagging bias corrections for nonb-jets due to the MC statistics are relatively large. These errors could only be reduced by a considerable increase in the MC statistics for each p_T bin, something which was not possible within the framework of this thesis.

6.5 Hadron Level Corrections

In order to compare the results to any future theoretical models or other experimental results, it is important to measure the jet shapes at hadron level. It is therefore necessary to correct for the detector effects. This is in part done by correcting the p_T of the jets, as described in section 5.4. On top of this, an additional correction is needed to correct the jet shapes.

The correction factors are obtained in a similar way to the one used in reference [15]. Correction factors are computed for each bin in r and each bin in p_T, from MC. The b-quark jet

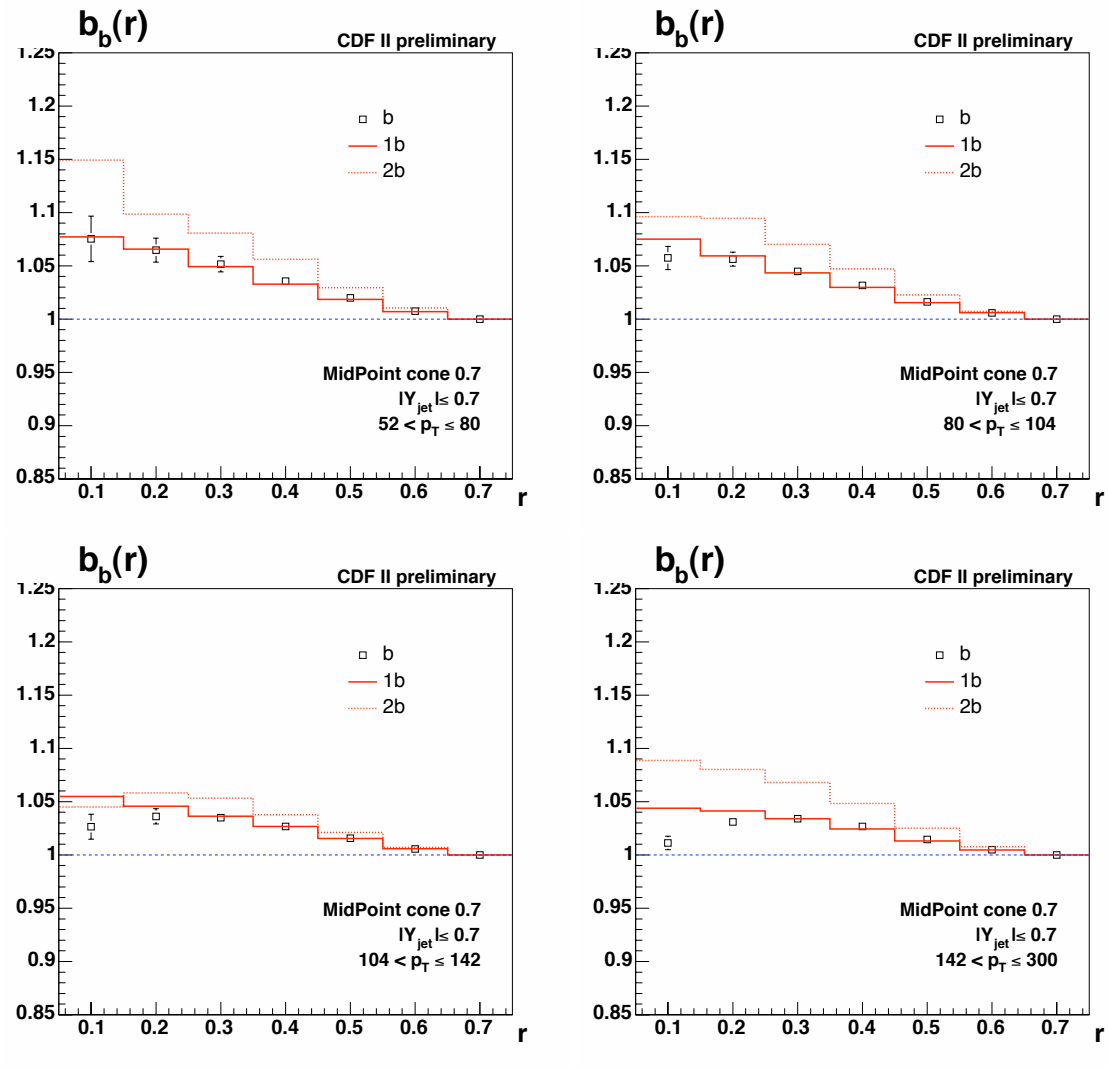

FIGURE 6.9: Bias due to tagging on b-jets (open black squares) shown alongside the bias for 1b- and 2b-jets (full and dotted red lines, respectively). The errors shown are the errors due to the MC statistics.

shapes in MC are computed at detector level and at hadron level. The correction factors are defined as

$$C^{\text{had}}(r) = \frac{\Psi^{\text{had}}_{\text{MC}}(r)}{\Psi^{\text{det}}_{\text{MC}}(r)}, \tag{6.12}$$

where $\Psi^{\text{det}}_{\text{MC}}(r)$ are the shapes computed at calorimeter level or using tracks, and $\Psi^{\text{had}}_{\text{MC}}(r)$ are the MC truth shapes computed considering all stable particles, after the fragmentation and hadronisation processes. Figure 6.12 shows the correction factors for each p_T bin. The correction factors are of the order of 4% at most.

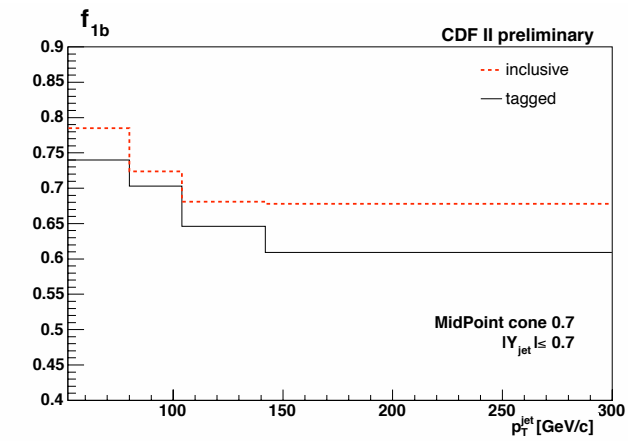

FIGURE 6.10: Fraction of b-jets containing a single b-quark inside the jet cone as predicted by Pythia Tune A MC. The f_{1b} fraction after tagging (black line) is compared to the fraction before tagging (red dashed line).

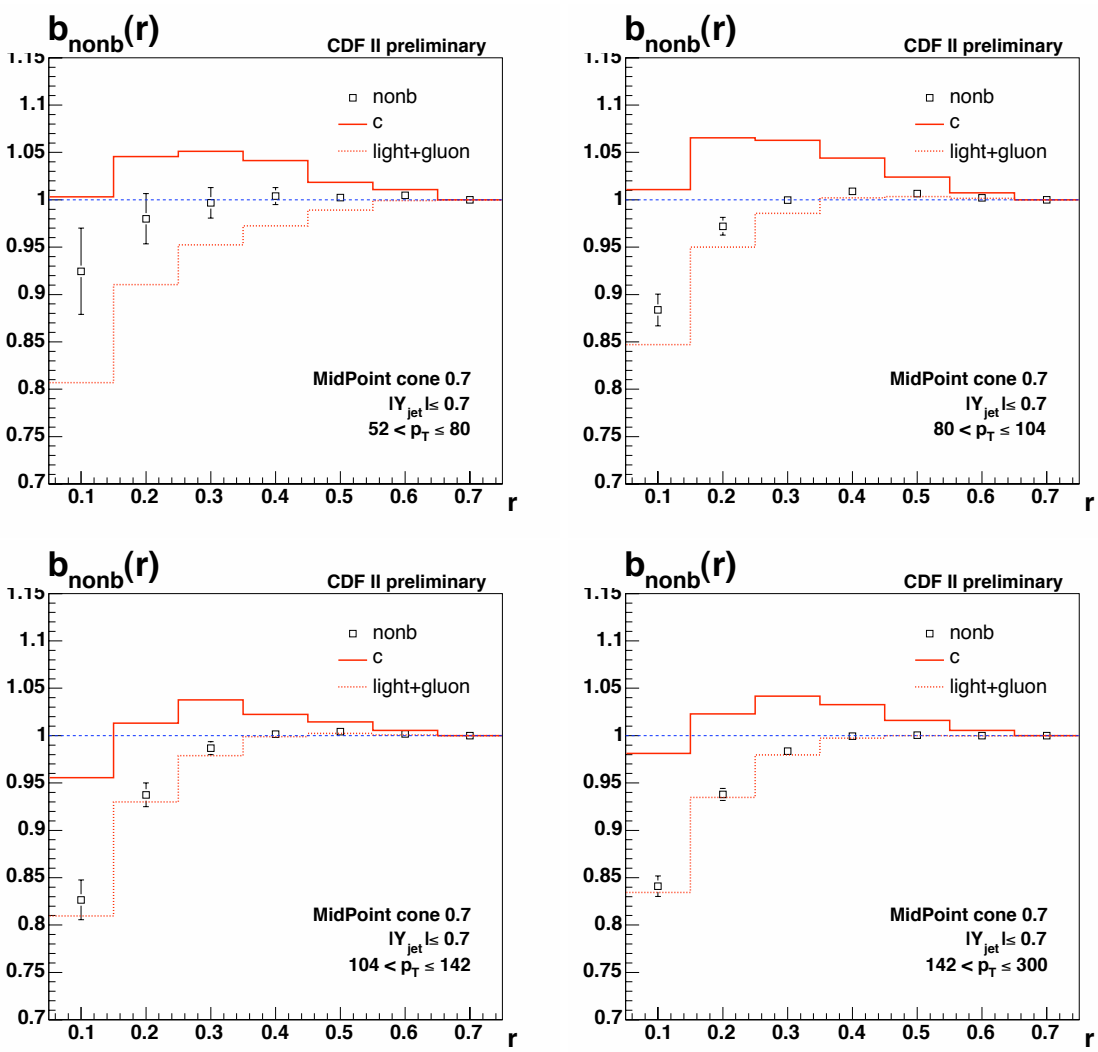

FIGURE 6.11: Bias due to tagging on nonb-jets (black open squares) shown alongside the bias for c- (full red line) and light+gluon-jets (dotted red line). The errors shown are the errors due to the MC statistics.

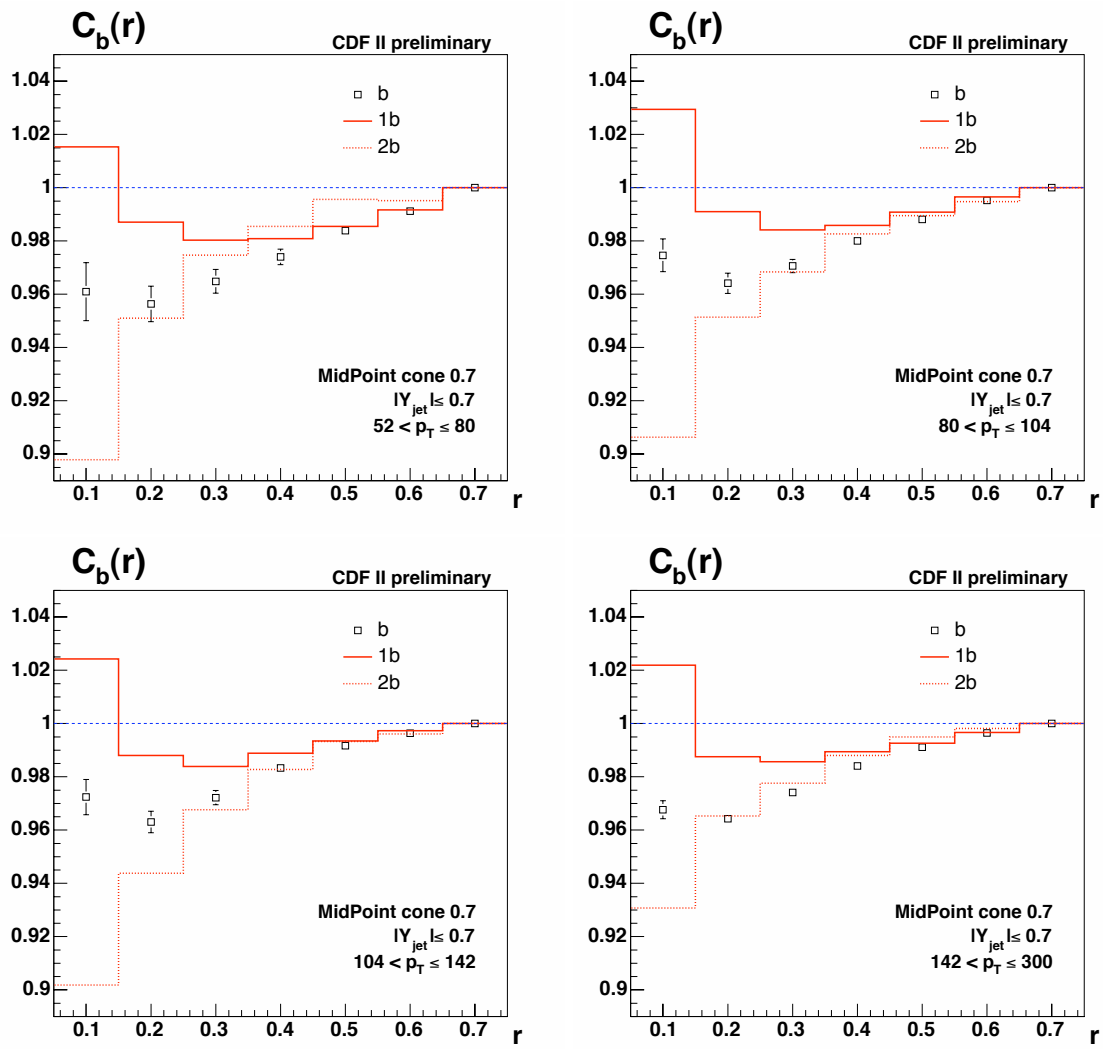

FIGURE 6.12: Hadron level corrections to the b-jet shapes. The errors shown are the MC statistical errors.

6.6 Hadron Level b-quark Jet Shapes

Figure 6.13 shows the hadron level b-quark jet shapes obtained from the unfolding procedure discussed in the previous sections. Only the statistical errors, which are most of the time smaller than the points, are shown at this point. The final results will be shown in chapter 8. The data are compared to the Pythia Tune A MC hadron level predictions for inclusive and b-quark jet shapes.

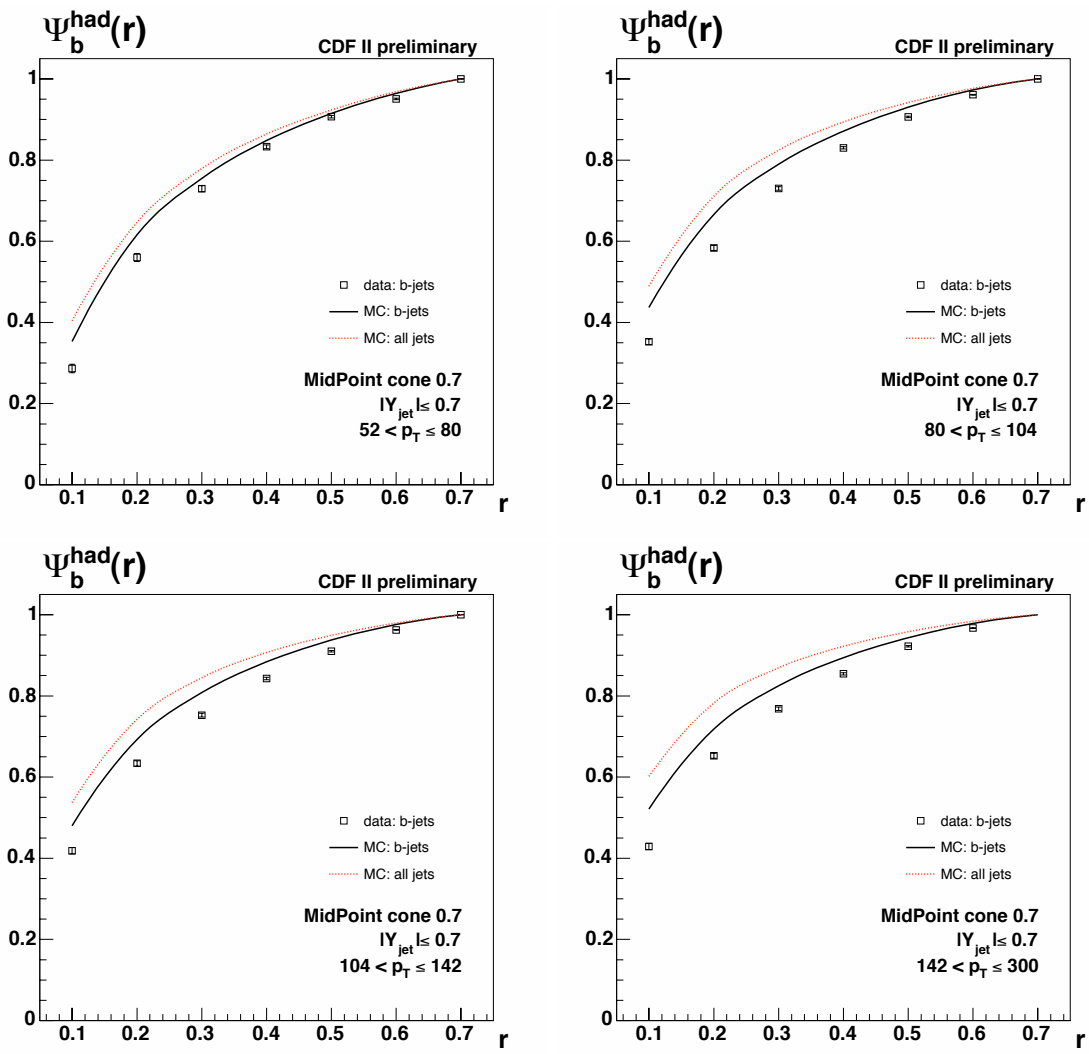

FIGURE 6.13: Hadron level b-quark jet shapes for all p_T bins considered (black open squares). Only the statistical errors are shown. Also shown are the Pythia Tune A MC predictions for inclusive (red line) and b-quark jet shapes (black line).

Chapter 7

Systematic Studies

In this chapter, the systematic studies carried out for this analysis are presented. First the various uncertainties are evaluated in order to determine what variation of each parameter is needed. The hadron level b-quark jet shapes are calculated for each parameter variation and the differences between the default hadron level b-quark jet shapes and those computed using the varied parameters are taken as the systematic uncertainties on the measurements due to that effect. Except where explicitly mentioned, the uncertainties on the different fractions discussed below are quoted as absolute errors, not relative ones.

First, it is necessary to extract the statistical error on the purity from the total error quoted by the fitter. The error quoted by the secondary vertex mass fitter is a combination of the statistical errors of the data and of the MC templates. This method is presented in the next section.

The relative fraction of c-jets in the nonb-jets is not guaranteed to be accurate in the Pythia Tune A MC. It is necessary to find a method to estimate how good the MC description of this fraction is. This is done in section 7.2. This fraction has an effect on the SecVtx mass templates for nonb-jets and on the biases on nonb-jets due to the tagging requirement.

A number of the parameters used in the unfolding method are sensitive to the relative fraction of b-jets which have only one b-quark inside the jet cone. This parameter is called here $f_{1\mathrm{b}}$ and is defined as

$$f_{1\mathrm{b}} = \frac{N^{1\mathrm{b}}_{\mathrm{jets}}}{N^{1\mathrm{b}}_{\mathrm{jets}} + N^{2\mathrm{b}}_{\mathrm{jets}}}, \tag{7.1}$$

where $N^{1\mathrm{b}}_{\mathrm{jets}}$ and $N^{2\mathrm{b}}_{\mathrm{jets}}$ are the number of b-jets which contain one and two b-quarks, respectively. Section 7.3 presents a previous study comparing the Pythia Tune A results to NLO calculations. This allows a 20% uncertainty on this $f_{1\mathrm{b}}$ fraction to be set.

There are a certain number of standard systematic studies which are carried out for most QCD analyses [15] [29]. These are the variations of the jet energy scale, the primary vertex location and the missing E_T significance.

The systematic uncertainty related to the differences between the hadron level b-quark jet shapes calculated using tracks instead of calorimeter towers is presented in section 7.8. This systematic study is carried out to verify the detector response to low momentum particles. Section 7.9 looks into the differences in the hadron level b-quark jet shapes obtained from shapes measured at calorimeter level but using only towers with $p_T > 0.5\,\mathrm{GeV}$ and those obtained using the default cut at 0.1 GeV.

As mentioned in section 5.2, different samples were generated in order to investigate the effect on the unfolding method of using a specific MC sample with a specific parameter set. These studies are presented in section 7.10.

Section 7.11 presents a study to investigate the dependence on the MC modelling of the SecVtx performance. Some distributions, related to the SecVtx algorithm, such as the number of Pass 1 tracks, are not well described by the Pythia Tune A MC. The distribution of the number of Pass 1 tracks in Pythia Tune A MC is re-weighted to agree with the data. The different biases and hadron level corrections used for the unfolding of the b-quark jet shapes are re-computed using the re-weighted Pythia Tune A MC and are found to be within the errors due to the MC statistics from the default unfolding paramters.

The final systematic uncertainties are shown in section 7.12 for each bin in r and each bin in p_T for each of the effects studied in this chapter. The dominant sources of systematic uncertainties are also discussed.

The final section of this chapter presents a study, which is not included as a source of systematic error, on the effect of changing the jet rapidity cut from $|Y| \leq 0.7$ to $0.1 \leq |Y| \leq 0.7$. The purpose of this study is to check if the previously published inclusive results [15], computed using the latter rapidity cuts, can be directly compared to the results of this analysis. It turns out that removing the central rapidity region does not affect the b-jet shapes or the inclusive shapes.

7.1 Extraction of the Statistical Error on the Purity

Before being able to investigate systematic uncertainties on the purity, it is important to extract the contribution of the statistical error to the total error on the purity quoted by the fitting algorithm. The fitting algorithm returns a total error which takes into account both the statistical errors of the templates and of the data. In order to extract the errors due to the statistics of the data, the statistics of the templates are artificially increased until the contribution to the total error of the fit is negligible. The error contribution due to the statistics of the templates goes roughly as $1/\sqrt{N}$; by increasing artificially the statistics of the Monte Carlo templates, one is effectively reducing the error due to the templates to zero. This method is similar to that carried out in [29]. In practise this is done by scaling the template histograms for b- and nonb-jets by a factor of 10^3, which reduces the effective error of the templates by a factor of $\sqrt{10^3}$. This factor is motivated by figure 7.1, taken from [29], which shows for one particular bin in p_T the total error of the fit as a function of the factor by which the templates are scaled. The blue line shows the asymptotic error. This is taken as the statistical error. This shows that a scaling of 10^3 is more than sufficient to consider that the error quoted by the fitter is the statistical error.

The error on the fit when using the scaled histograms is taken as the statistical error, $\Delta p_{\mathrm{b}}^{\mathrm{stat}}$. The error due to the MC statistics is thus given by

$$\Delta p_{\mathrm{b}}^{\mathrm{MC\,stat}} = \sqrt{(\Delta p_{\mathrm{b}}^{\mathrm{fit}})^2 - (\Delta p_{\mathrm{b}}^{\mathrm{stat}})^2} \tag{7.2}$$

where $p_{\mathrm{b}}^{\mathrm{fit}}$ is the total error quoted by the fitting algorithm.

7.2 Estimate of the c-fraction in Data

An important source of systematic uncertainty is the fraction of nonb-jets which are c-jets. This will have an effect on the secondary vertex mass templates for nonb-jets which are used for the

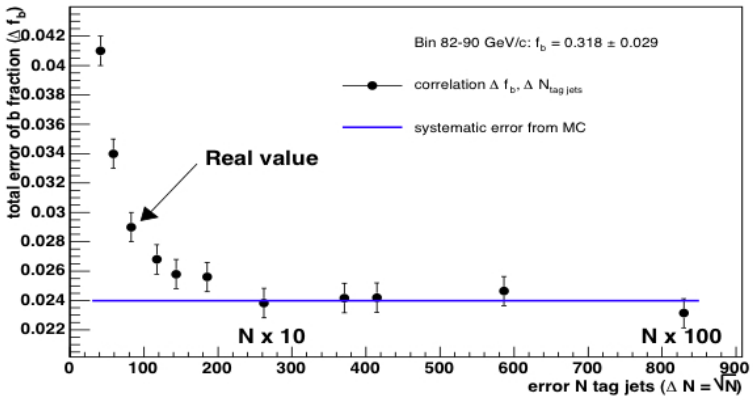

FIGURE 7.1: Error on the purity of the sample returned by the fitting algorithm as a function of the artificial scaling applied to the MC templates.

measurement of the purity of the sample. It will also have an effect on the biases due to tagging on the nonb-jets. Previous studies quote a conservative error of $20 - 30\%$ (relative error) [29]. It is important to see if this error can be reduced. This was done by carrying out a secondary vertex mass fit using three different templates instead of the standard two, as was presented in section 6.3. Different templates are obtained for b-jets, c-jets and light+gluon-jets.

The limitation of this method is that the statistics for the c- and light+gluon-jet templates is too low for the fits to converge. In order to overcome this statistical limitation, it is possible to use higher p_T MC samples without biasing the outcome of the fit. The higher p_T samples have a larger number of events in the p_T bins of interest because the cut on the hard-scattered p_T is higher. The assumption that the use of the higher p_T MC samples does not bias the outcome of the secondary vertex mass fit must be checked before results for the c-fraction can be obtained. As shown in section 5.2, the MC samples used for the four p_T bins of this analysis are: Pt18, Pt40, Pt60 and Pt90. In this section these are referred to as the default samples. The higher p_T MC samples would thus be: Pt40, Pt60, Pt90 and Pt120, respectively. The comparisons in this section are not applicable to the lowest p_T bin (which uses Pt18 by default) because the p_T distribution of the Pt40 sample is not correct, due to turn on effects, up to about 60 GeV. The p_T distributions for Pt18 and Pt40 are therefore not the same over the whole range of the bin. For the other three p_T bins, a set of comparisons are made which show that the use of the higher p_T samples do not bias the fit results. The three tests carried out are the following

- The b- and nonb-jet templates for the default and higher p_T samples are within errors of each other as shown in figure 7.2.
- The results of the fit of the data to the standard b- and nonb-jet templates using the default and the higher p_T samples are within 4% of each other, well within the total error on the fit. The results of the fit are shown for both cases in table 7.1. The difference between the two fitted fractions is also shown in this table.
- The fit of the data to the b-, c- and light+gluon-templates using the higher p_T samples give results for the b-jet fraction that are within 2% of the results using the default, two-template fit. The results are shown in table 7.2.

These tests all show that the use of the higher p_T MC samples for the templates does not bias in any way the results of the fits. The values for the c-purity can therefore be trusted.

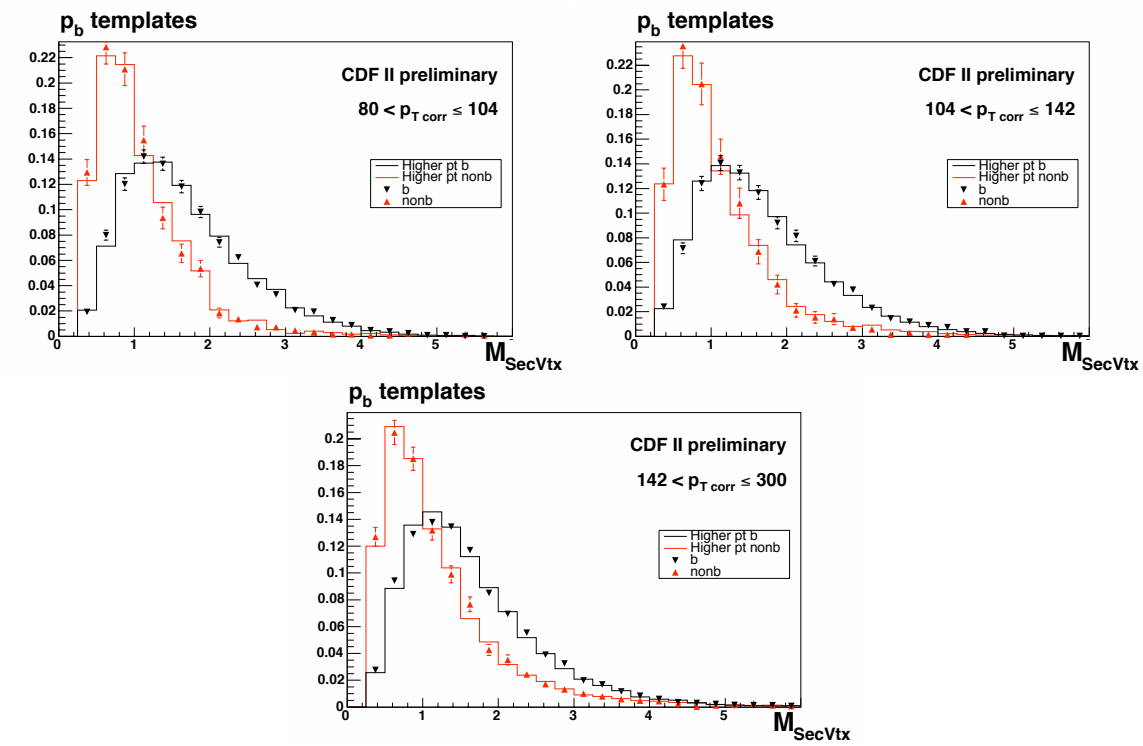

FIGURE 7.2: Secondary vertex mass distributions for b- (black) and nonb-jets (red) for all the p_T bins. The standard MC samples, shown as points with the MC statistical errors, are compatible with the higher p_T samples (lines).

p_T range	$p_\mathrm{b} \pm \Delta p_\mathrm{b}$	$p_\mathrm{b} \pm \Delta p_\mathrm{b}$ higher p_T templates	Δp_b
80-104	0.339±0.023	0.299±0.016	0.040
104-142	0.300±0.025	0.275±0.015	0.025
142-300	0.242±0.026	0.242±0.015	0.000

TABLE 7.1: Results for the b-jet purity, p_b, using the standard MC templates and the ones obtained using the higher p_T templates. The errors shown are the errors returned by the fitter which include the statistical errors and the errors due to the MC statistics. Also shown is the difference between these two values for each bin.

p_T range	$p_\mathrm{b} \pm \Delta p_\mathrm{b}$	$p_\mathrm{c} \pm \Delta p_\mathrm{c}$	$f_\mathrm{c} \pm \Delta f_\mathrm{c}$	f_c in MC	Δf_c	Δp_b
80-104	0.327±0.022	0.307±0.047	0.456±0.071	0.418	0.038	0.012
104-142	0.284±0.018	0.245±0.039	0.342±0.055	0.309	0.033	0.016
142-300	0.240±0.023	0.190±0.044	0.250±0.060	0.254	0.004	0.002

TABLE 7.2: Results for p_b, p_c and f_c in data from the fits using the higher p_T templates. The f_c fraction in MC is also shown along with the difference between the fit and the MC values. The differences between the p_b values obtained with the standard fit and the 3-parameter fit are shown in the last column. The errors shown are the errors returned by the fitter which include the statistical error and the error due to the MC statistics.

The templates for b-, c- and light+gluon-jets using the higher p_T MC samples are shown in figure 7.3 for the three highest p_T bins. The distribution in data, along with the fit to the three templates is shown in figure 7.4. This shows that the fit is very good. The c-content of

the nonb-jets, f_c, is defined as

$$f_c = \frac{N^c_{\text{jets}}}{N^{\text{nonb}}_{\text{jets}}} = \frac{N^c_{\text{jets}}}{N^{\text{tot}}_{\text{jets}} - N^b_{\text{jets}}} = \frac{p_c}{1 - p_b}, \tag{7.3}$$

where p_b and p_c are the fitted b- and c-fractions and N^x_{jets} are the number of jets of flavour x in the sample. Figure 7.5 shows the results for f_c as a function of p_T for the three highest p_T bins. This shows that that the c-content of the nonb-jet fraction is well described in MC. The fit gives results, shown in table 7.2, that are within 4% of the MC values. For this analysis, the MC truth values for f_c are used because of the fact that it is not possible to extract from data the c-content for the lowest p_T bin. In order to account for the 4% difference in the fitted fractions, along with the uncertainty on the fit, about 3%, a systematic error of 5% is considered for the fraction of nonb-jets which are c-jets.

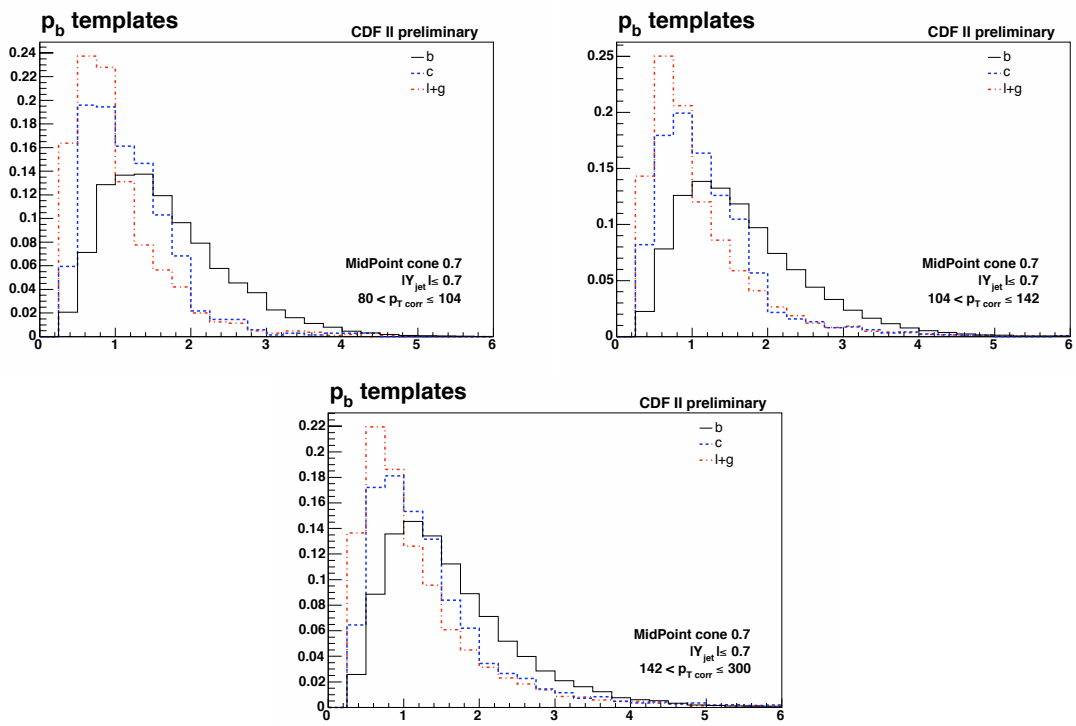

FIGURE 7.3: Secondary vertex mass distributions for b- (black solid line), c- (blue dashed line) and light+gluon-jets (red dotted line) for all the p_T bins.

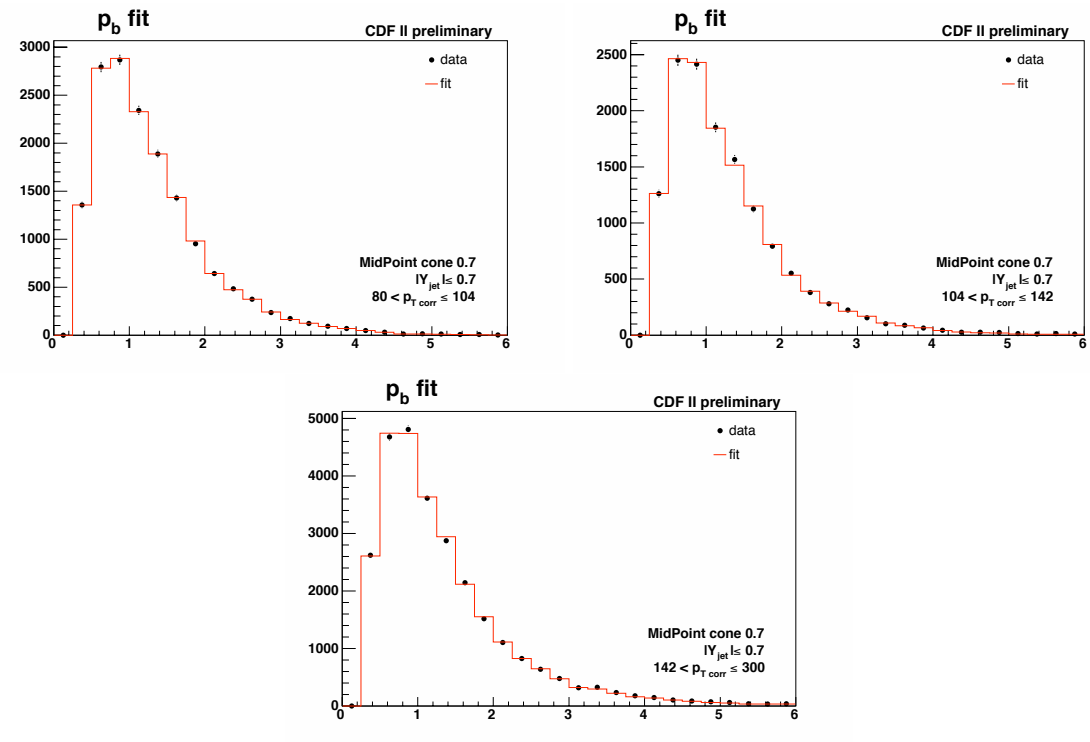

FIGURE 7.4: Secondary vertex mass distribution in data (black points), showing the statistical errors, compared to the fitted distribution using the three templates (red line) for the three highest p_T bins.

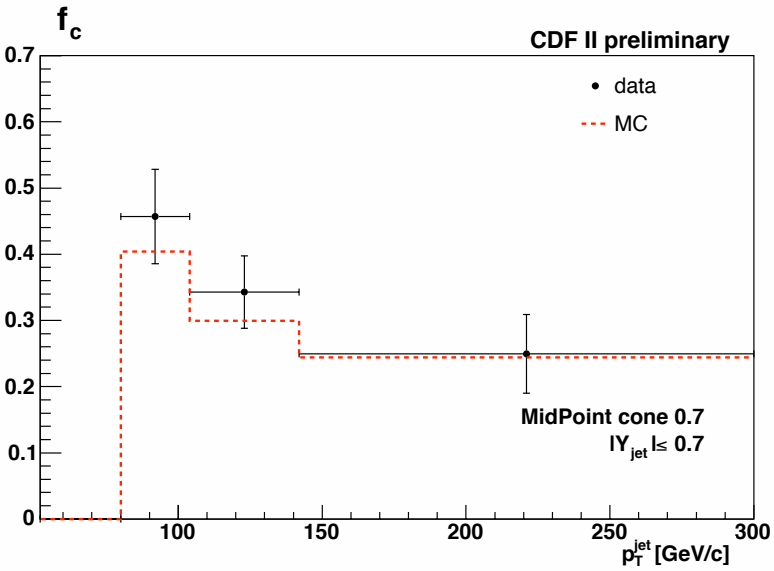

FIGURE 7.5: Fraction of nonb-jets which are c-jets as a function of jet p_T (black points) shown along with the values obtained from MC (red dashed line) which are used in this analysis. The errors shown are the errors returned by the fitter which include the statistical errors and the errors due to the MC statistics.

7.3 Estimate of the Error on f_{1b} and f_{1c}

Jets that contain two b-quarks inside the same jet cone are mainly jets which originate from a gluon splitting into a $b\bar{b}$ pair. Such jets are expected to be significantly wider than jets from flavour creation, containing most of the time only a single b-quark inside the cone. The fraction of gluon splitting events has not been measured to a high precision at hadron colliders. The total fraction of inclusive jets at CDF that are gluon jets has been shown, by the inclusive jet shape analysis, to agree well with the data [15]. One of the aims of this analysis is to determine how well the fraction of gluon splitting events in b-jet production is modelled by the MC.

Many of the distributions used for the unfolding, such as the secondary vertex mass templates, the bias corrections and the hadron level corrections, are expected to be different for jets that contain one or two b-quarks. The parameters used in the unfolding are therefore somewhat sensitive to the fraction of b-jets that have a single b-quark, f_{1b}. In a similar way, the fraction of c-jets with one c-quark inside the jets, f_{1c}, also has an effect on the analysis procedure. It is thus important to obtain an estimate for the uncertainty of these fractions in order to include this effect as a systematic error.

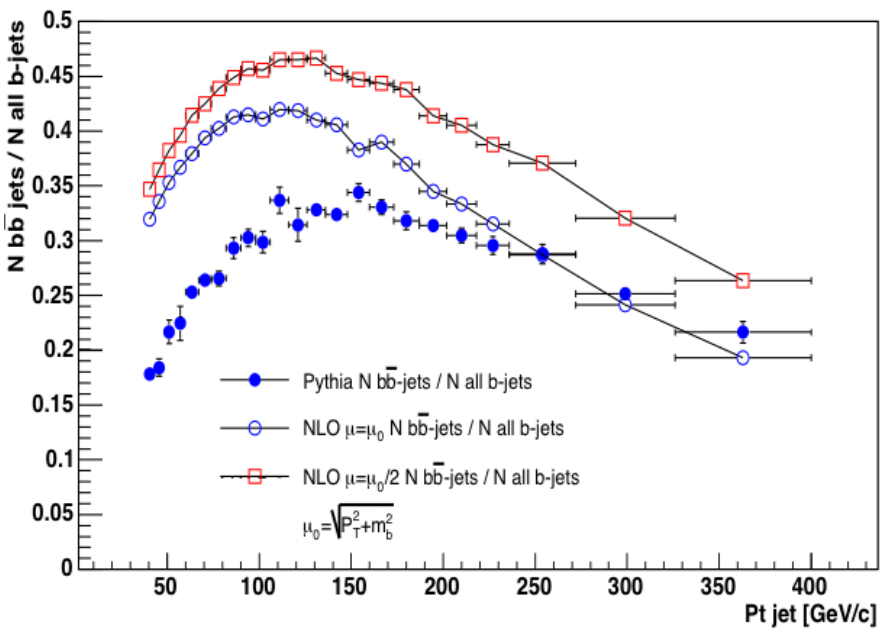

FIGURE 7.6: Fraction of b-jets containing two b-quarks inside the jet cone. Pythia Tune A is compared to NLO calculations using both a scale of $\mu = \mu_0$ and $\mu = \mu_0/2$. This plot is taken from [35].

Reference [35] shows a comparison of the fraction of jets containing two b-quarks, which is equal to $1 - f_{1b}$, between Pythia Tune A, where the jet algorithm is run at hadron level, and Next to Leading Order (NLO) calculations. This comparison is reproduced in figure 7.6. The renormalisation and factorisation scale choice of the NLO calculation affects the $b\bar{b}$ fraction. The maximum difference with respect to the Pythia Tune A values is obtained with a scale of $\mu = \mu_0/2$, where $\mu_0 = \sqrt{p_T^2 + m_b^2}$. The maximum difference is of the order of 17% (absolute). The plotted fractions are for inclusive b-jets, before any tagging is applied. It is reasonable to expect that the tagging efficiency for double b-quark jets is higher than that for single b-quark jets. This can be verified by comparing, in Pythia Tune A MC, the f_{1b} fraction before and after tagging, as was already shown in figure 6.10. The double b-quark jet fraction increases by about 10% after the tagging requirement is applied. There is no reason to believe that the tagging

fractions should increase differently for the Pythia Tune A and the NLO predictions but the final systematic uncertainty considered is taken to be slightly higher than the simple difference between the Pythia Tune A LO and NLO predictions. The total systematic uncertainty on the f_{1b} fraction is taken to be 20% (absolute). The error on the fraction of c-quark jets which have a single c-quark inside the jet cone should be similar and is therefore also taken to be 20%. It turns out that when fitting the final hadron level b-jet shapes to an unknown f_{1b} fraction, the results of the fit are within this 20% uncertainty for all the bins, as will be shown in section 8.4.

7.4 Jet Energy Scale

A systematic error must be associated to the uncertainty of the jet corrections as well as the uncertainty on the detector level jets. This systematic error is a measure of how well the jet energy corrections, presented in section 5.4, correct the measured jet p_T back to the hadron level one. The systematic uncertainty on the jet energy scale quoted for the official jet energy corrections is 3% (relative error) [36]. Another study specifically calculated the jet corrections for b-jets and found results that were in good agreement with the inclusive jet corrections [29]. The comparison between the official jet corrections and the b-jet specific corrections is shown in figure 7.7 and 7.8 for inclusive and tagged jets, respectively. Also shown in this figure are the corrected p_T distributions using the same correction factors as the inclusive jet shape analysis [15]. This shows that the official jet correction are very similar to the b-jet specific corrections and to the inclusive jet shape analysis corrections. The systematic error on the jet energy scale should combine both the error quoted for the official jet corrections and the uncertainty on the b-jet fragmentation which is 0.6%. Adding these two effects in quadrature leads to a total systematic uncertainty on the jet energy scale of 3% (relative error). For this analysis, the effect on the final hadron level b-quark jet shapes is evaluated by shifting the p_T of the bins by $\pm 3\%$ and re-computing the shapes.

7.5 Missing E_T Significance

The systematic effect due to the choice of the missing E_T significance cut is taken to be the difference in the hadron level b-quark jets between the default shapes and those computed using the standard CDF variation of $\pm 15\%$ (relative) [29] [15]. The cut on the missing E_T significance removes a large fraction of the cosmic events but the shapes are not expected to be sensitive to the values of the cuts because the number of events which are removed by this cut is very small. This systematic effect is thus expected to be small for all the parameters.

7.6 Primary Vertex z-position

The cut on the location of the primary vertex is varied by ± 5 cm around the event selection cut at 50 cm and the effect on the hadron level b-quark jet shapes is investigated. Again this effect is expected to be small because the corrections to the jets and secondary vertices due to the location of the primary vertex are well understood in this region.

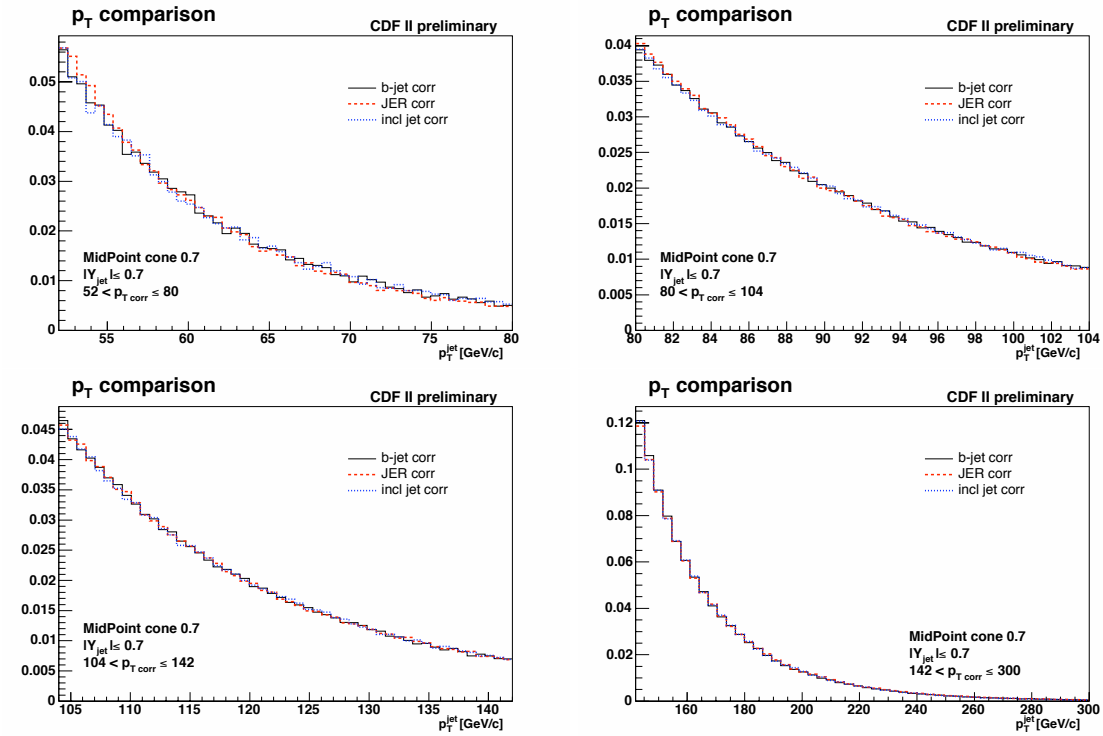

FIGURE 7.7: Comparison in data of the corrected p_T distributions of all jets for each dataset obtained with the official jet corrections (red dashed line), the b-jet specific corrections (black line) and the corrections used in the inclusive jet shape analysis (blue dotted line). The distributions have been normalised to unit area.

7.7 Inclusive vs. nonb-jet Shapes

As was shown in section 6.2, the use of the inclusive jet shapes instead of the nonb-jet shapes in the unfolding equation does not introduce any significant systematic error. The differences between the raw inclusive and nonb-jet shapes, as obtained from the Pythia Tune A MC, are found to be smaller than the statistical errors.

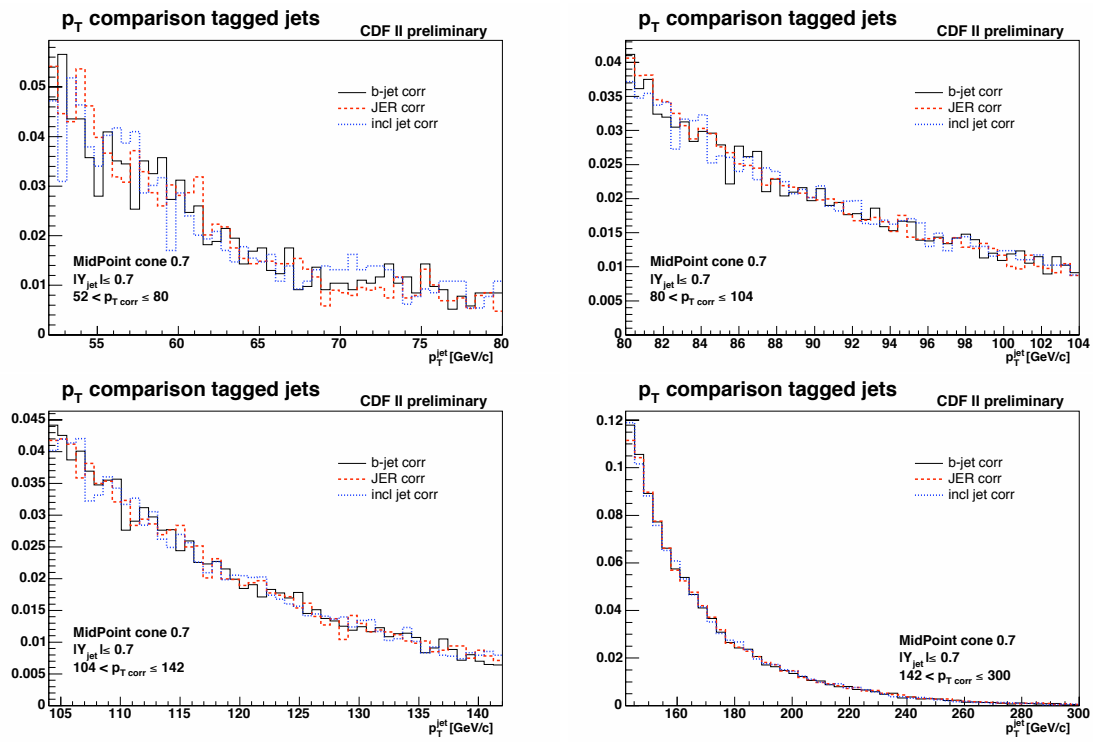

FIGURE 7.8: Comparison in data of the corrected p_T distributions of tagged jets for each dataset obtained with the official jet corrections (black line), the b-jet specific corrections (red dashed line) and the corrections used in the inclusive jet shape analysis (blue dotted line).

7.8 Hadron Level b-jet Shapes Obtained Using Tracks

The hadron level b-jet shapes obtained starting from the raw calorimetric shapes are compared to those obtained starting with tracks. The jet algorithm is still run using calorimeter towers. To be included inside a jet, the tracks must pass certain cuts

- $\Delta R_{\text{track}\to\text{jet}} \leq 0.7$, to ensure the tracks are inside the jet cone
- $0.5 < p_T^{\text{track}} < 100\,\text{GeV}$, to ensure the tracks are well reconstructed
- track $|Z_0| < 2\,\text{cm}$ with respect to the primary vertex, to ensure the tracks come from the same primary vertex as the jet

These cuts are the same as used for the inclusive jet shape analysis [15].

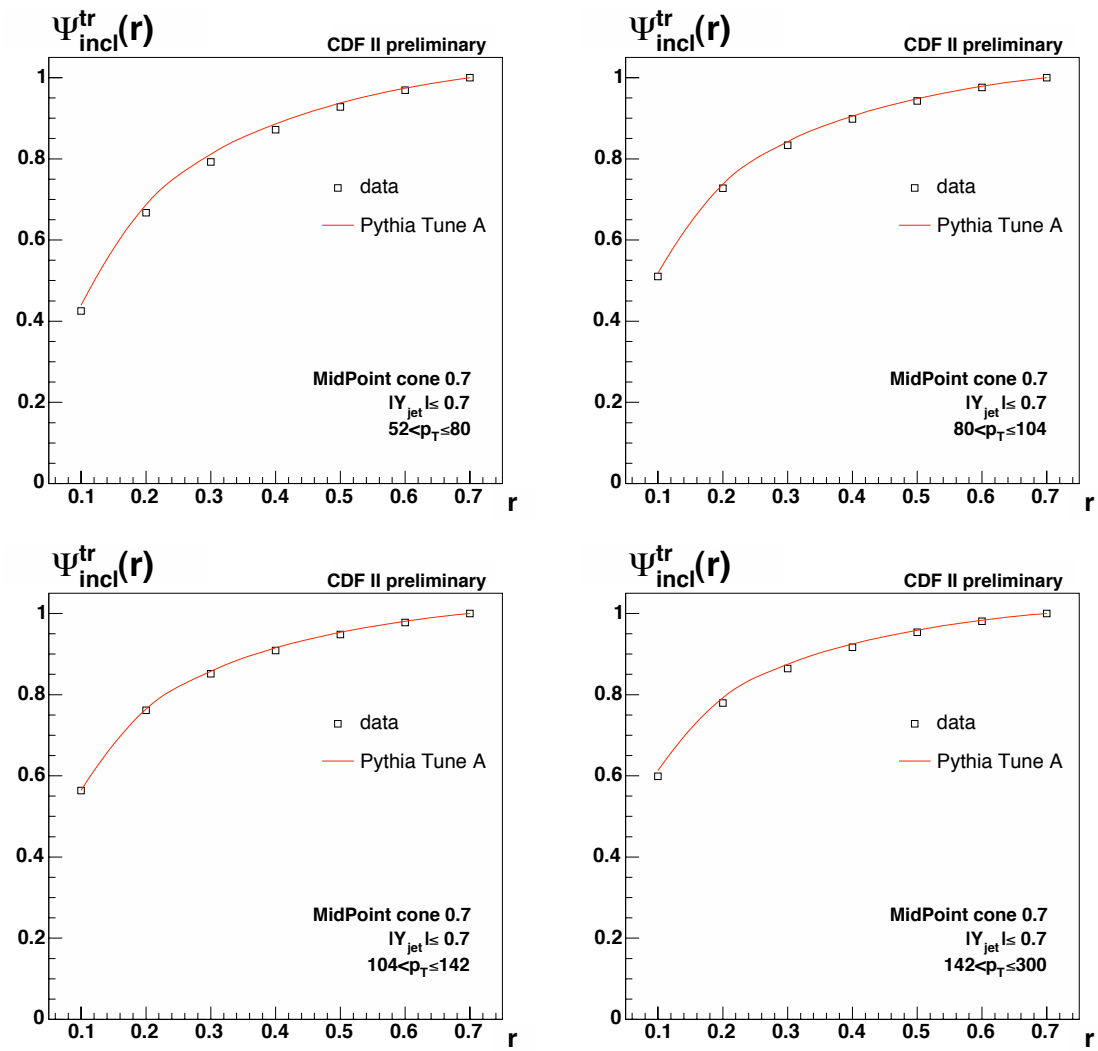

FIGURE 7.9: Raw integrated shapes for inclusive jets measured using tracks. The Pythia Tune A MC predictions (red line) are compared to the data (black points). Only the statistical errors are shown.

Figure 7.9 shows a comparison between data and the Pythia Tune A MC of the raw inclusive shapes measured using tracks. Figure 7.10 shows the same comparison for tagged jets. As for the case of calorimetric jet shapes, the agreement between data and MC is not perfect. The data and MC shapes for the tagged jets are expected to be slightly different due to the different

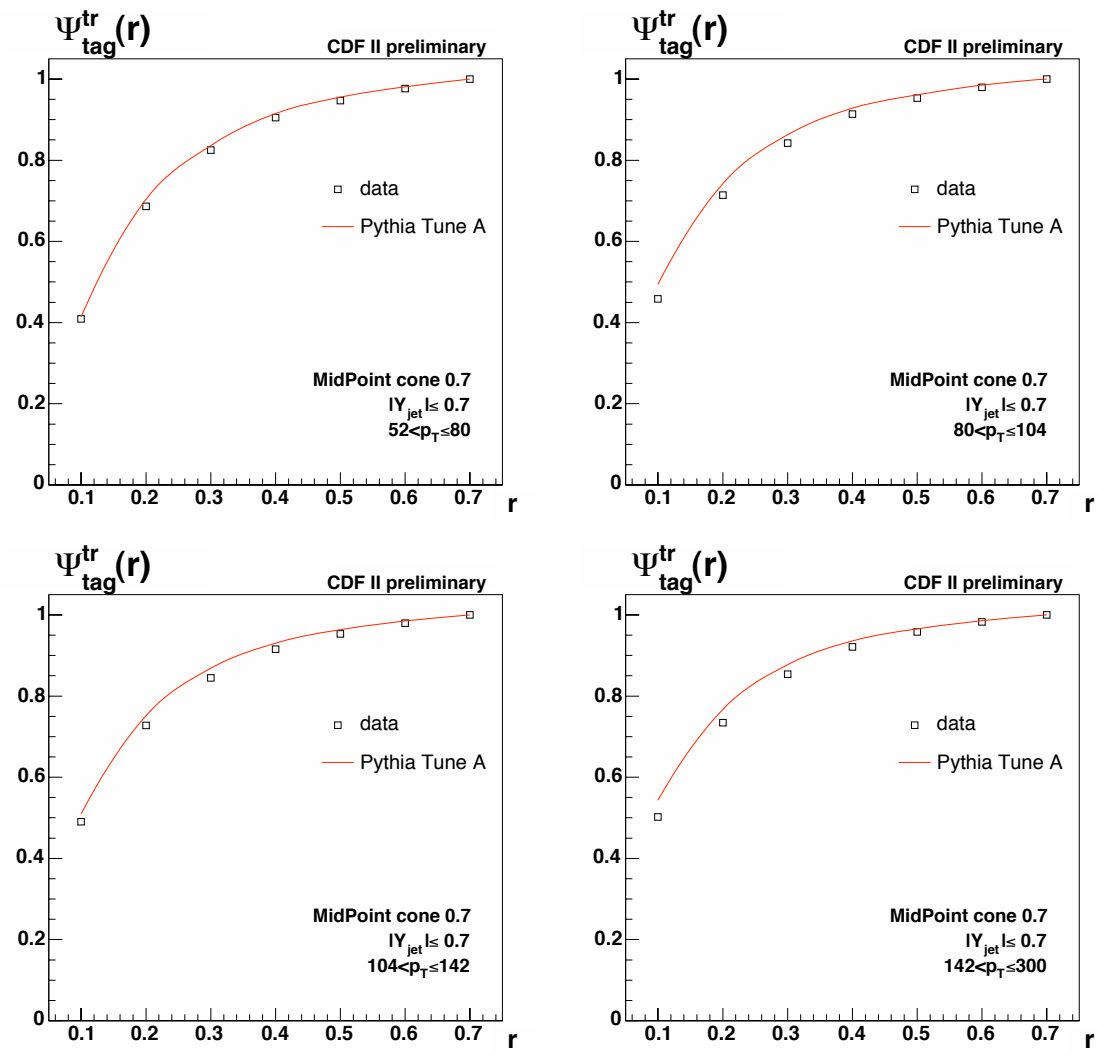

FIGURE 7.10: Raw integrated shapes for tagged jets measured using tracks. The Pythia Tune A MC predictions (red line) are compared to the data (black points). Only the statistical errors are shown.

fractions of b-jets.

The bias corrections to correct for the tagging requirement and the hadron level corrections are re-computed for the shapes measured using the tracks. The hadron level b-quark jet shapes are shown in figure 7.11. The differences between the final hadron level b-jet shapes calculated using the default towers and using tracks are shown in figure 7.12. These differences are included as systematic errors. The largest difference between the hadron level b-quark jet shapes reconstructed from tracks and calorimeter towers is obtained for the lowest p_T bin for which the difference is of the order of 0.05.

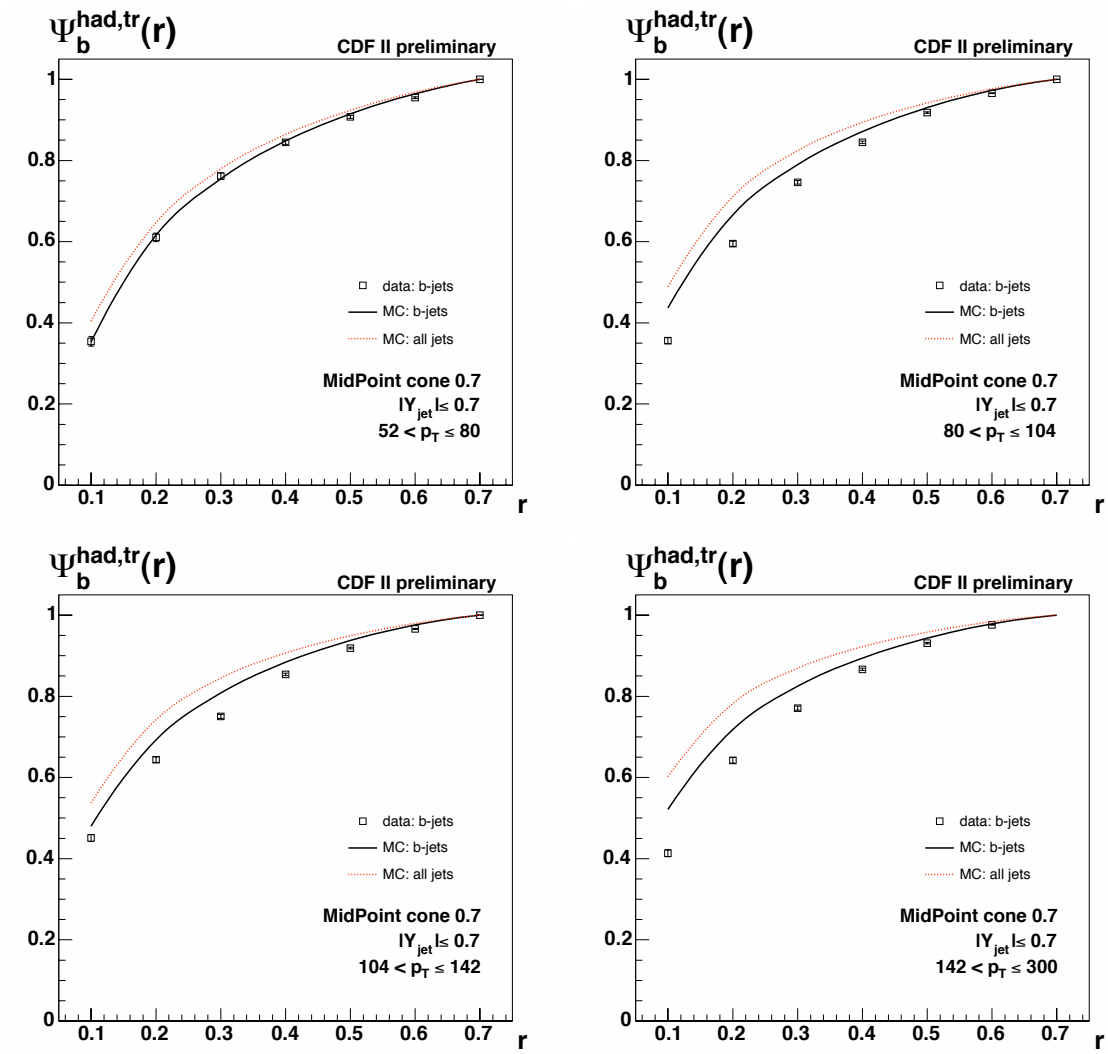

FIGURE 7.11: Hadron level integrated b-jet shapes for each of the p_T bins considered, starting from tracks. The results are shown as black points, only the statistical errors are shown. The MC predictions for inclusive jets and for b-jets are shown as red and black curves, respectively.

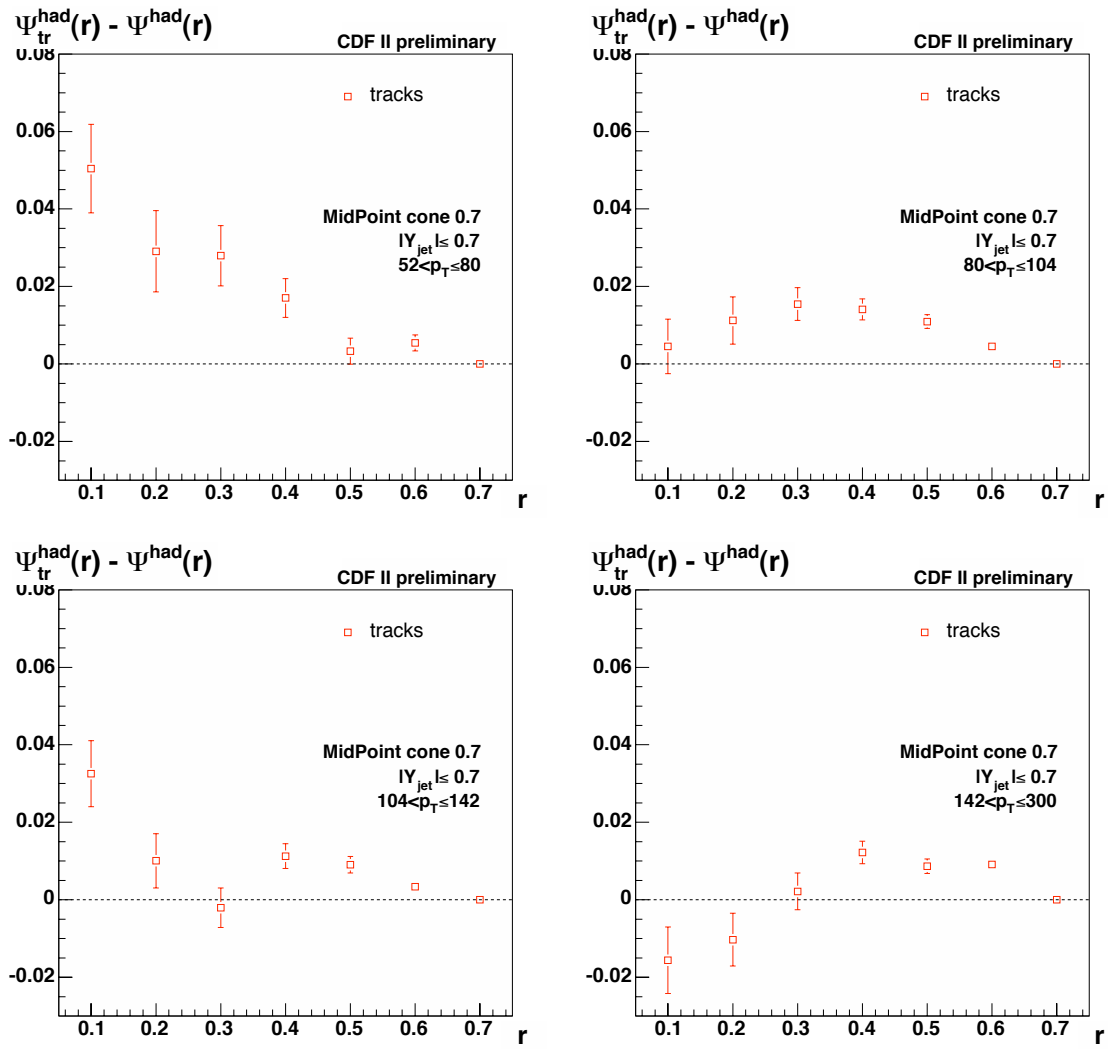

FIGURE 7.12: Difference between the hadron level integrated b-jet shapes starting from tracks and starting from calorimeter towers (red points). Only the statistical errors of the shapes reconstructed using tracks are shown.

7.9 Hadron Level b-jet Shapes Obtained Using Calorimeter Towers with $p_T > 0.5$ GeV

The hadron level b-jet shapes obtained starting from the raw calorimetric shapes with the default cut on the p_T of the towers at 0.1 GeV are compared to those obtained with a cut on the p_T of the towers at 0.5 GeV. This study is carried out in order to investigate how well the CDF detector simulation models low transverse momentum calorimeter energy deposits.

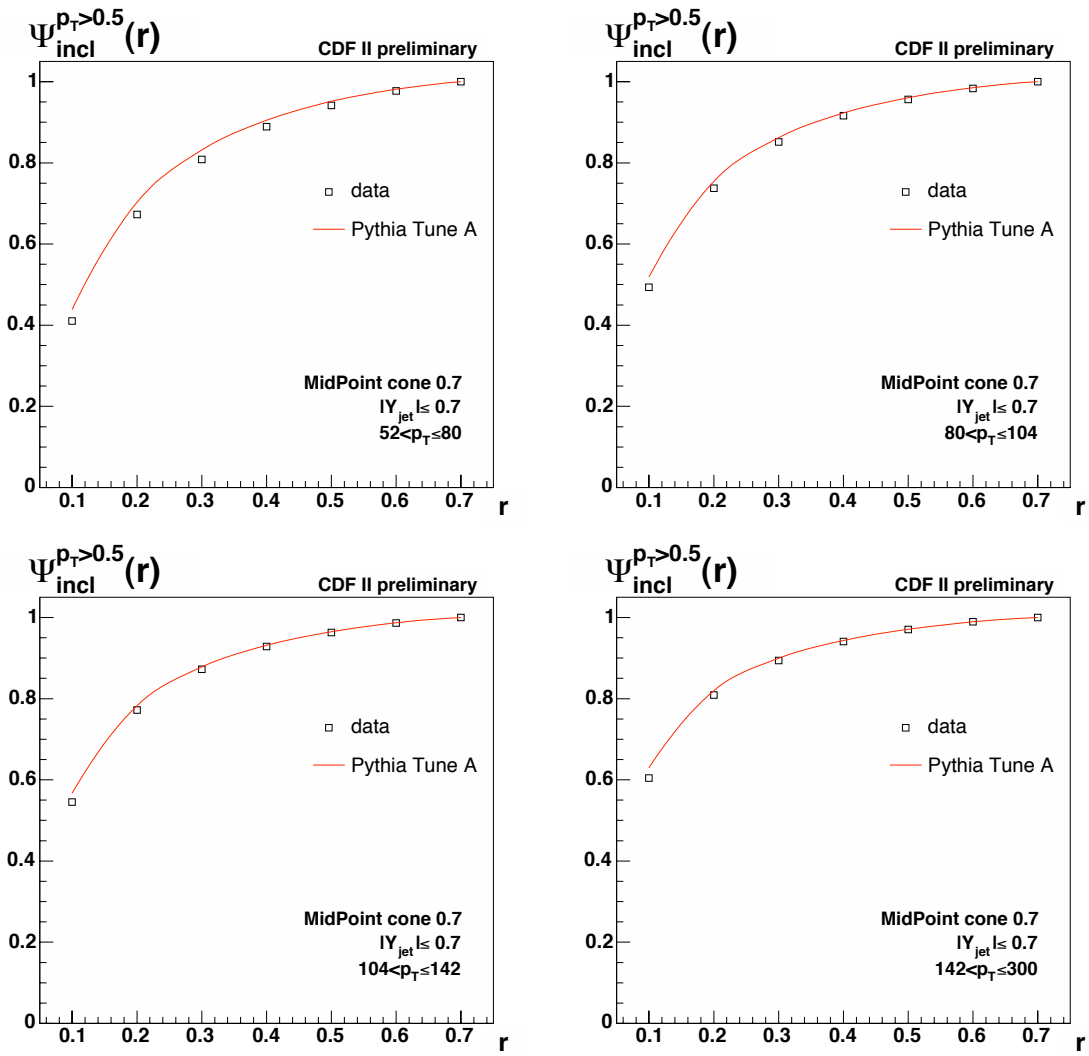

FIGURE 7.13: Raw integrated shapes for inclusive jets measured using calorimeter towers with a cut on the p_T of the towers at 0.5 GeV. The Pythia Tune A MC predictions (red line) are compared to the data (black points). Only the statistical errors are shown.

Figure 7.13 shows a comparison between data and the Pythia Tune A MC of the raw inclusive shapes measured with the additional p_T cut. Figure 7.14 shows the same comparison for tagged jets. As for the default case, the agreement between data and MC is not perfect. The data and MC shapes for the tagged jets are expected to be different because of the different fractions of b-jets in data and MC.

The bias corrections to correct for the tagging requirement and the hadron level corrections are re-computed for the raw shapes measured with this additional p_T cut on the towers. The

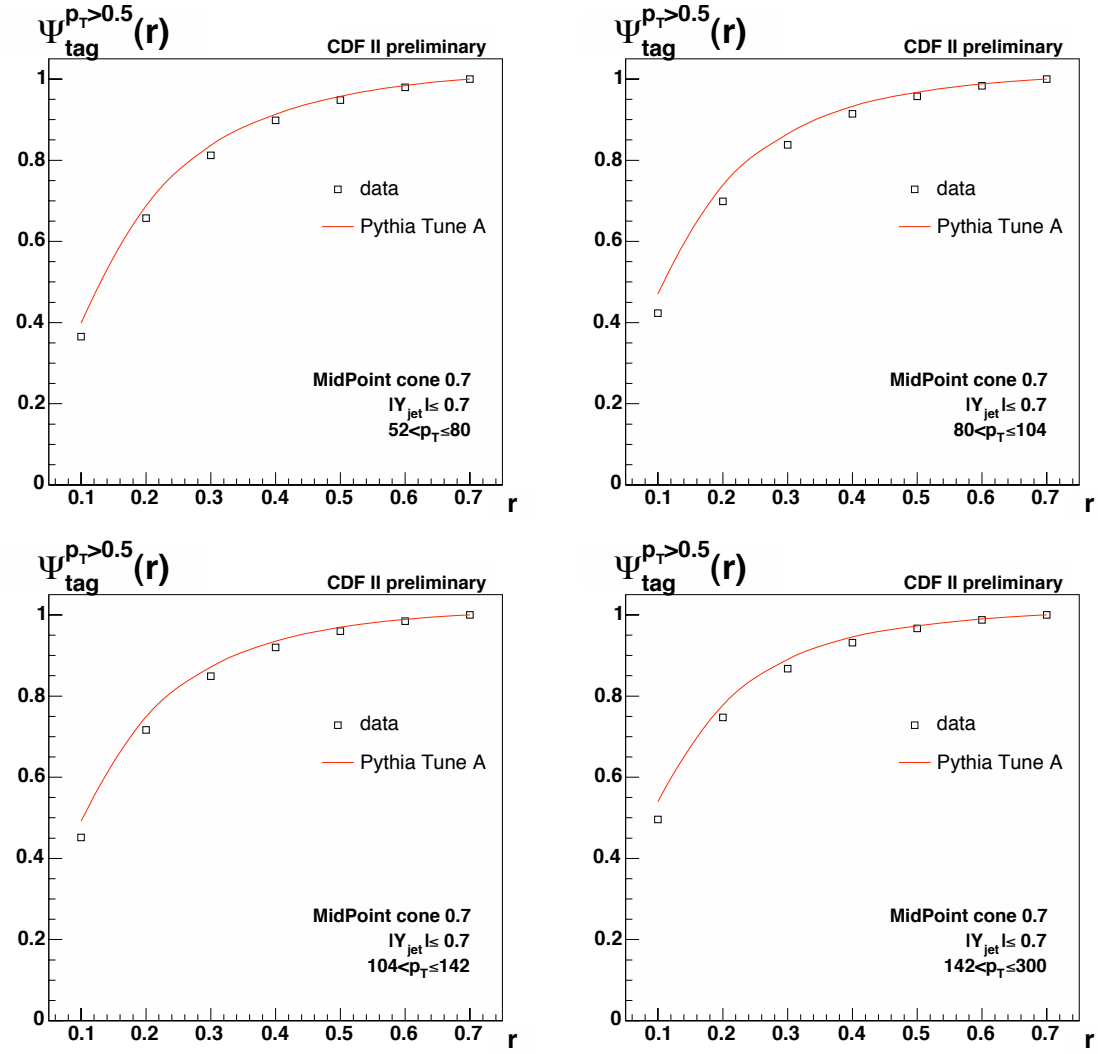

FIGURE 7.14: Raw integrated shapes for tagged jets measured using calorimeter towers with cut on the p_T of the towers at 0.5 GeV. The Pythia Tune A MC predictions (red line) are compared to the data (black points). Only the statistical errors are shown.

hadron level b-jet shapes measured using the additional p_T cut on the towers are shown in figure 7.15. The differences between the hadron level b-jet shapes calculated using the default cut on the tower p_T and the cut at 0.5 GeV are shown in figure 7.16. This difference is included as a source of systematic error. It is found to be of similar size to the statistical errors and therefore not a dominant source of systematic uncertainty. The largest difference with respect to the default cut on the calorimeter tower p_T is about 0.005.

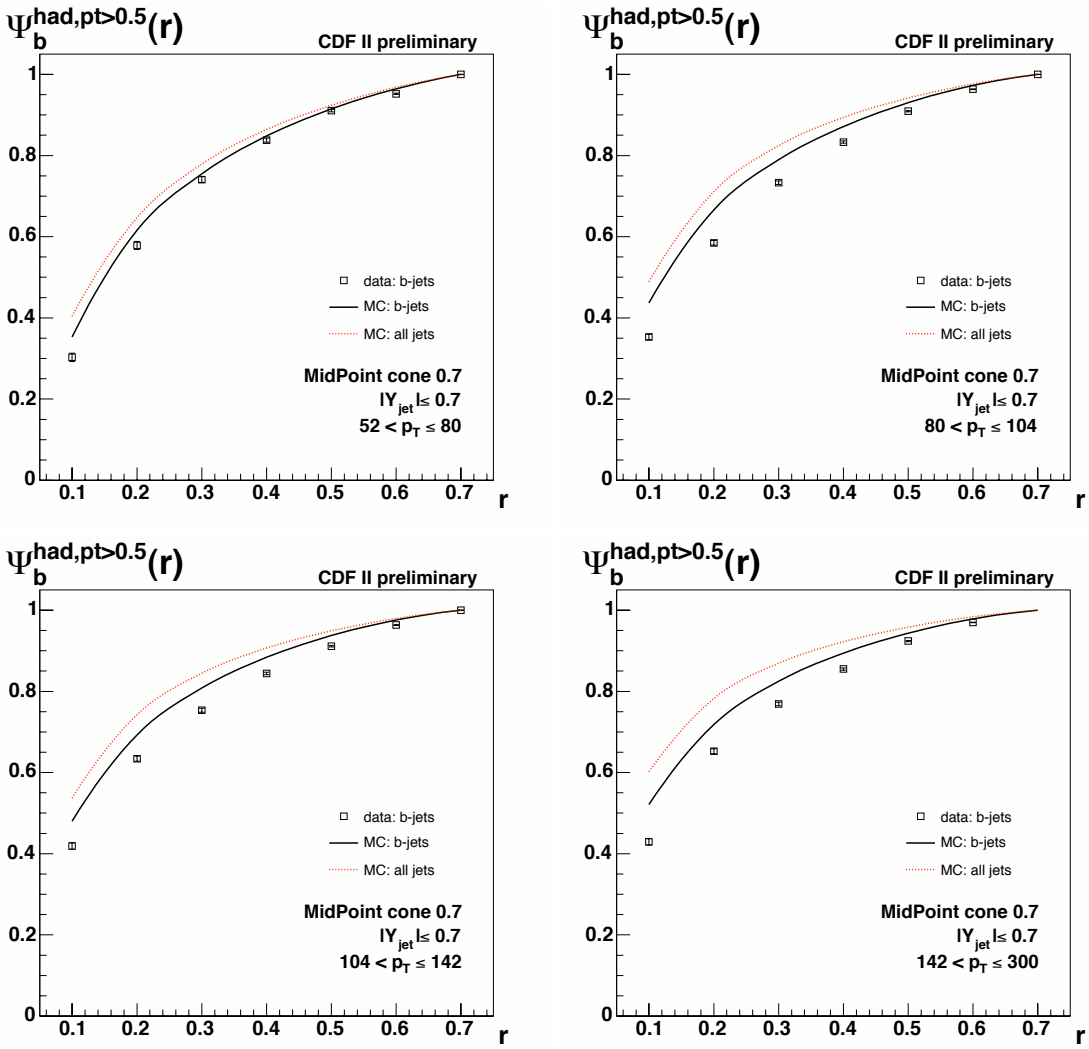

FIGURE 7.15: Hadron level integrated b-jet shapes for each of the p_T bins considered, starting from calorimeter towers with p_T larger than 0.5 GeV. The results are shown as black points, only the statistical errors are shown. The MC predictions for inclusive jets and for b-jets are shown as red and black curves, respectively.

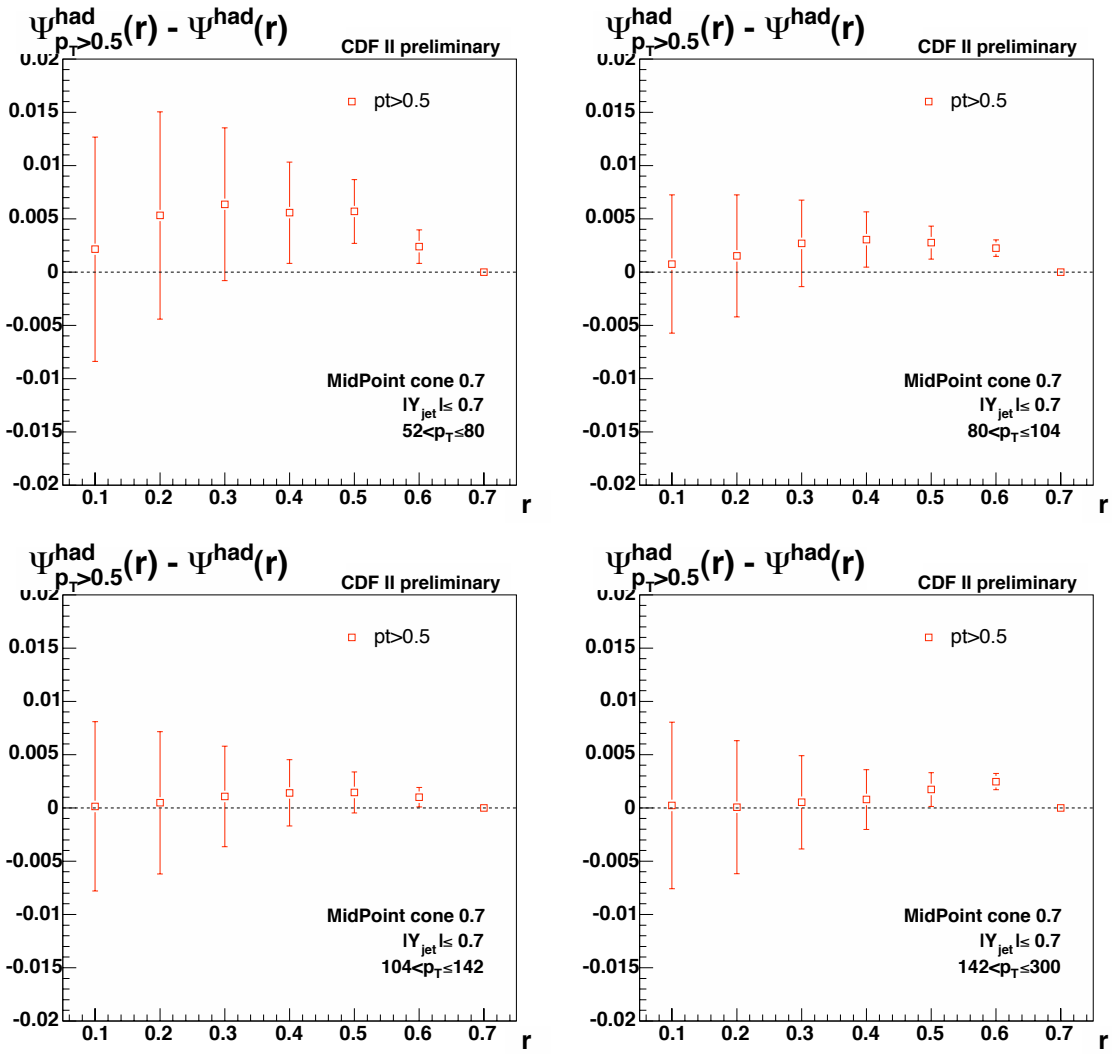

FIGURE 7.16: Difference between the hadron level integrated b-jet shapes starting from calorimeter towers with p_T larger than 0.5 GeV and the default cut at 0.1 GeV. The results are shown as black points, only the statistical errors of the shapes from towers with $p_T > 0.5$ GeV are shown.

7.10 MC Dependence of the Unfolding Parameters

On top of the effects mentioned in the previous sections, the MC dependence of the parameters of the unfolding procedure is investigated. These studies are detailed in the next sections.

7.10.1 Purity

The difference in the templates obtained with different MC samples is investigated. Because the statistics of the Herwig MC samples is not very large, it was necessary to carry out the same procedure as detailed in section 7.2. The higher p_T Herwig nonb-templates were used. As for the c-fraction extraction, the lowest p_T bin cannot be considered because of the different p_T distributions. The Herwig templates agree within statistical errors with those from Pythia Tune A. Figure 7.17 shows the comparison between the templates using Pythia Tune A (solid lines) and Herwig (points). The fit results using the higher p_T Herwig samples are within the quoted fit errors of the Pythia Tune A results, as can be seen in table 7.3.

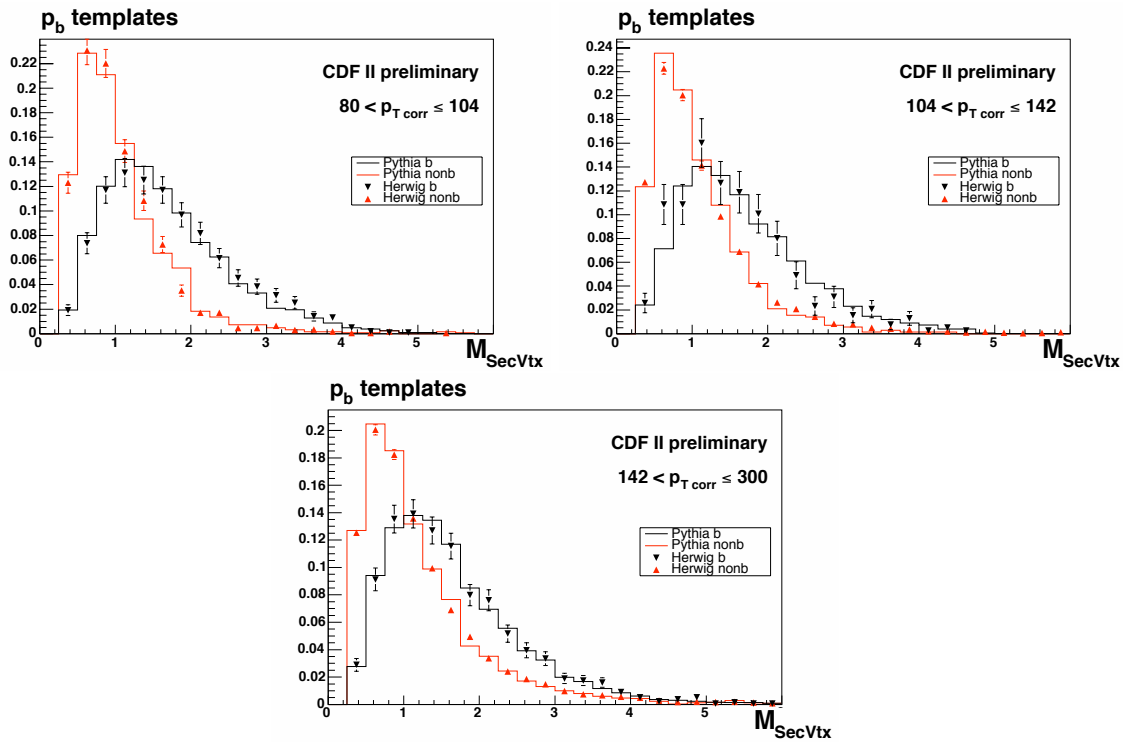

FIGURE 7.17: Secondary vertex mass distributions for b- (black) and nonb-jets (red) for all the p_T bins. The full lines represent the Pythia Tune A MC templates whereas the points represent the Herwig templates using the higher p_T MC samples.

A similar procedure can be applied to investigate the effect of the fragmentation function and PDFs on the secondary vertex mass templates. Because only b-filtered samples are available for the Pythia Tune A MC samples with the CTEQ6L PDFs as well as those with the Peterson fragmentation function, the comparison is only made for the b-jet templates. The b-jet templates for the default Pythia Tune A MC are shown in figures 7.18 and 7.19 along with the b-jet template obtained using Pythia Tune A with CTEQ6L PDFs for the former and with the Peterson fragmentation functions for the latter. The statistics for the comparison to Pythia Tune A with the Peterson fragmentation for the lowest bin are not very high. This shows that

	Herwig		
p_T range	p_b	fit error	Δp_b
80-104	0.342	0.024	0.019
104-142	0.327	0.032	0.027
142-300	0.253	0.020	0.011

TABLE 7.3: Fitted purities using the Herwig templates. The total fit errors quoted by the fitting algorithm are shown as well as the differences between the fit results and those obtained with Pythia Tune A.

at least for b-jets, where we expect the difference to be maximal, the templates are very similar.

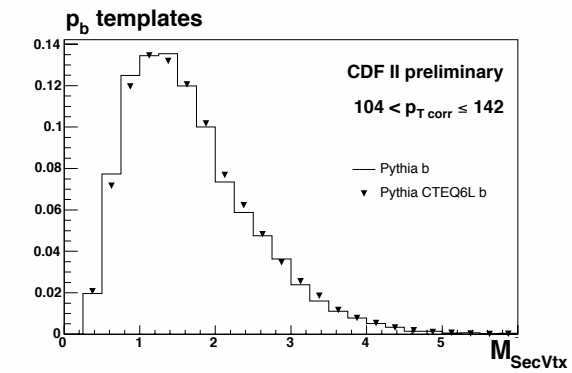

FIGURE 7.18: Secondary vertex mass distributions for b-jets. The line represents the default Pythia Tune A templates, whereas the points represent the Pythia Tune A predictions using the CTEQ6L PDFs.

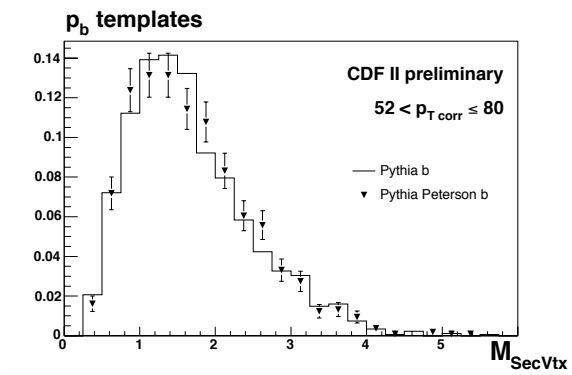

FIGURE 7.19: Secondary vertex mass distributions for b-jets. The line represents the default Pythia Tune A templates whereas the points represent the Pythia Tune A predictions using the Peterson fragmentation function.

There is not enough statistics in the samples produced with different PDFs or a different fragmentation function to carry out the secondary vertex mass fit. It is therefore not possible to compare fitted purities to verify that these samples do not change the outcome of the fit. Nevertheless it is possible to see, by comparing the templates for b-quark jets to the default Pythia Tune A MC ones, that the differences between the fitted purities should not be large.

7.10.2 Biases Due to SecVtx Tagging

The bias corrections are re-computed for all the different MC samples available: Pythia Tune A, Herwig, Pythia Tune A with CTEQ6L PDFs, Pythia Tune A with the Peterson fragmentation model. The biases obtained are compared to the default Pythia Tune A ones.

Biases Due to the Tagging on b-jets

Figure 7.20 shows the comparison of the tagging biases on b-jets between Pythia Tune A (red line) and Herwig (black full points). The biases obtained using Herwig are compatible with those obtained using Pythia Tune A. The errors on the biases for Herwig are relatively large due to the limited b-jet statistics.

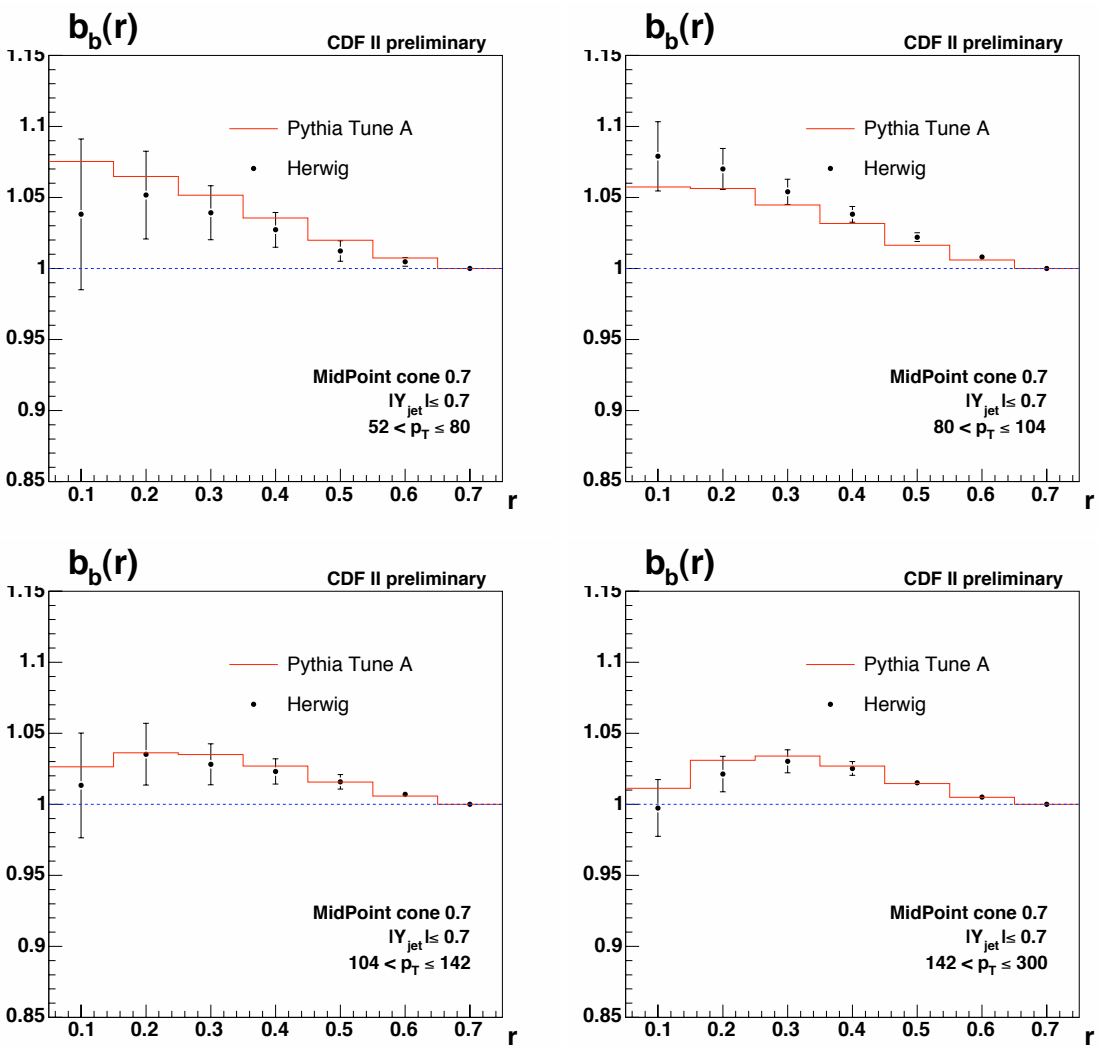

FIGURE 7.20: Comparison between Herwig (black full points) and Pythia Tune A (red line) MC of the biases due to the tagging on b-jets. The errors shown are the MC statistical errors.

Figure 7.21 shows a comparison of the tagging biases on b-jets between the default Pythia Tune A MC and Pythia Tune A with the Peterson fragmentation model. The difference in the bias for the first bin in r is relatively large but all other bins are similar to the default Pythia Tune A biases.

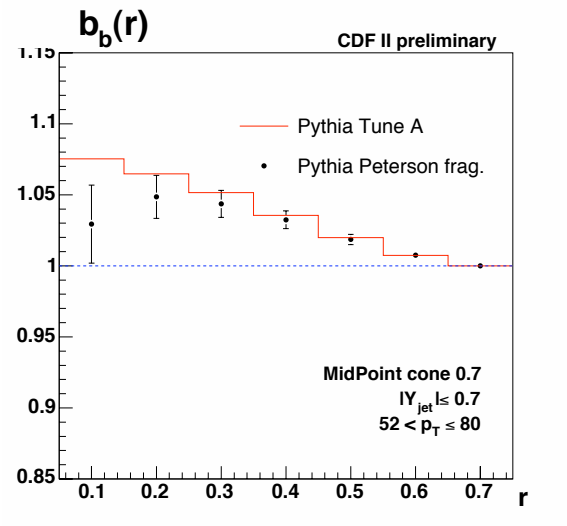

FIGURE 7.21: Comparison of the biases due to the tagging on b-jets between the default Pythia Tune A MC (red line) and Pythia Tune A with the Peterson fragmentation model (black full points). The errors shown are the MC statistical errors.

Figure 7.22 shows a comparison of the tagging biases to b-jets between the default Pythia Tune A MC and Pythia Tune A with CTEQ6L PDFs. The biases obtained for Pythia Tune A with CTEQ6L PDFs are similar the default Pythia Tune A biases.

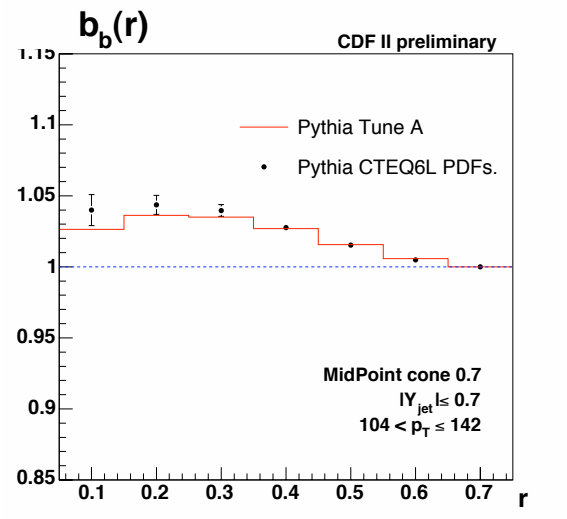

FIGURE 7.22: Comparison of the biases due to the tagging on b-jets between the default Pythia Tune A MC (red line) and Pythia Tune A with CTEQ6L PDFs (black full points). The errors shown are the MC statistical errors.

Biases Due to the Tagging on nonb-jets

The tagging biases on nonb-jets obtained using Herwig are compatible with those obtained using Pythia Tune A. The errors on the biases are relatively large due to the limited statistics in the tagged MC samples. Figure 7.23 shows the comparison between Pythia Tune A (red line) and Herwig (black full points). It is not possible to make a comparison between the default Pythia

Tune A and Pythia Tune A with Peterson fragmentation or CTEQ6L PDFs because there is not enough statistics for tagged nonb-jets.

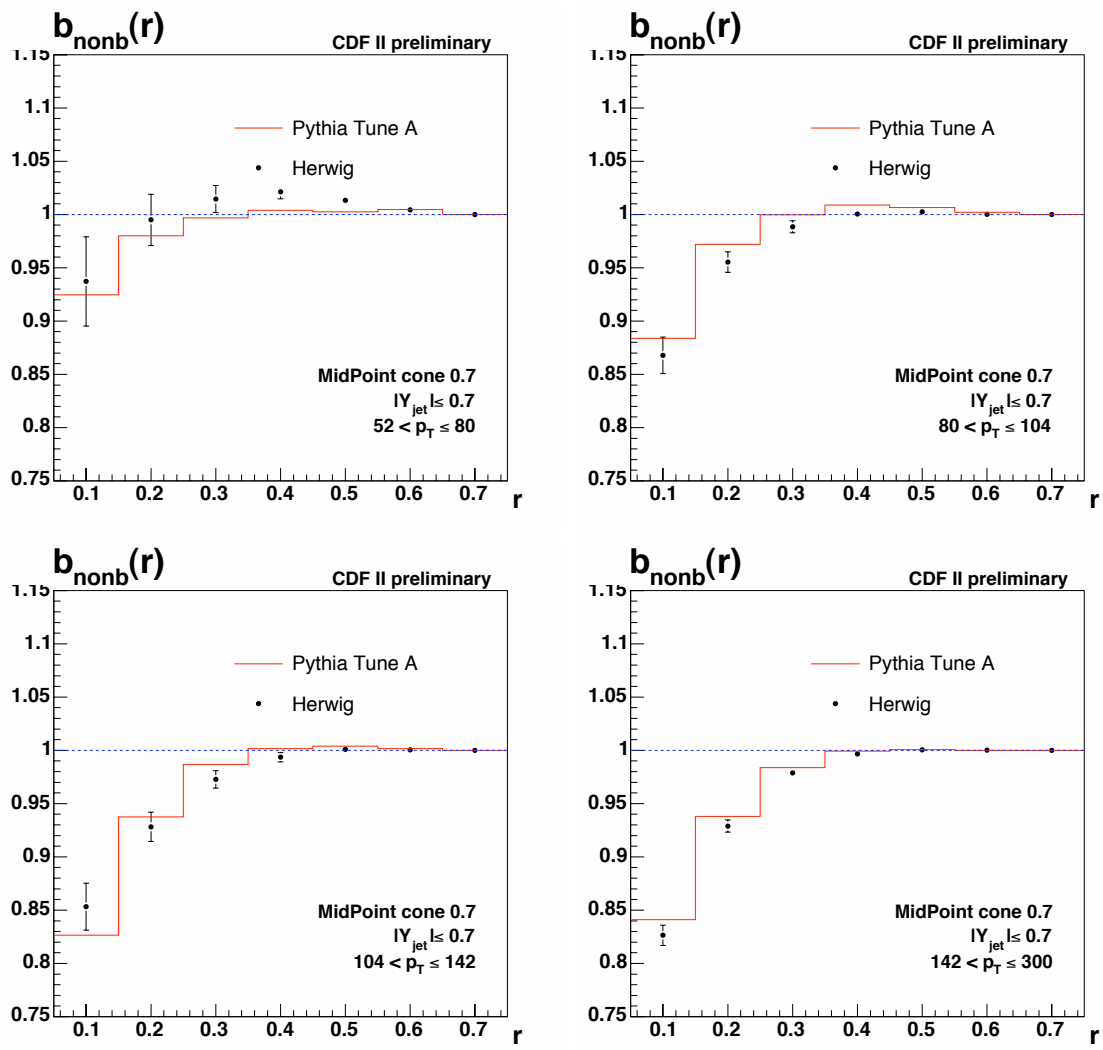

FIGURE 7.23: Comparison between Herwig (black full points) and Pythia Tune A (red line) MC of the biases due to the tagging on b-jets. The errors shown are the MC statistical errors.

7.10.3 Hadron Level Corrections to the Jet Shapes

The hadron level corrections are computed for all the different MC samples used. Figure 7.24 shows the comparison between the hadron level corrections obtained using Herwig and Pythia Tune A. The correction factors are not in good agreement for the 2nd and 4th jet p_T bins. The statistics for the 1st bin does not allow any conclusions to be drawn. The hadron level b-jet shapes are shown in figure 7.25. The shapes in Herwig are slightly wider than in Pythia Tune A. Figure 7.26 shows the same comparison for detector level b-quark jet shapes where the agreement between Herwig and Pythia Tune A is reasonably good. It is not possible to draw any conclusions about any systematic effect due to the MC dependence of the hadron level corrections from these comparisons because of the low statistics of the Herwig samples. Section 7.10.4 compares the final hadron level b-quark jet shapes unfolded using the Pythia Tune A MC to those obtained using the Herwig MC. This was only possible for the three highest p_T bins

because of the lack of statistics for the lowest p_T bin.

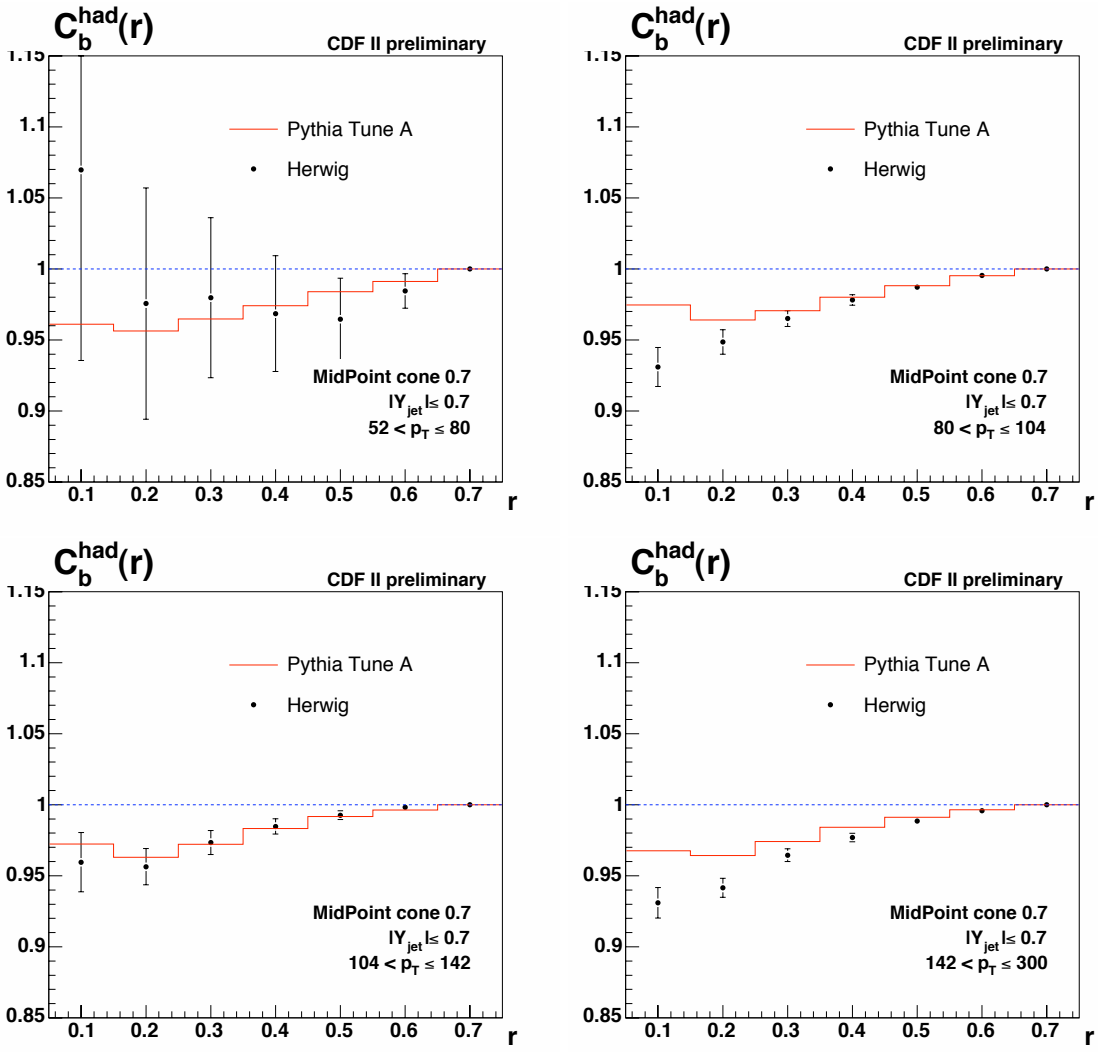

FIGURE 7.24: Comparison between Herwig (black full points) and Pythia Tune A (red line) MC of the hadron level corrections to b-jets. The errors shown are the MC statistical errors.

It is also possible to compare the default hadron level corrections to those obtained with Pythia Tune A using the Peterson fragmentation model. This is shown in figure 7.27. The hadron level corrections are compatible.

Finally, a comparison is made between the hadron level corrections to Pythia Tune A using the CTEQ6L PDFs and the default Pythia Tune A. This is shown in figure 7.28. The hadron level corrections are very similar for these two cases.

7.10.4 Hadron Level b-jet Shapes Using the Herwig MC for the Unfolding Parameters

Despite the fact that the statistics of the Herwig MC samples are not very large, it is possible to carry out the whole of the unfolding procedure using the tagging biases, hadron level corrections and secondary vertex mass templates from the Herwig MC samples instead of from the Pythia

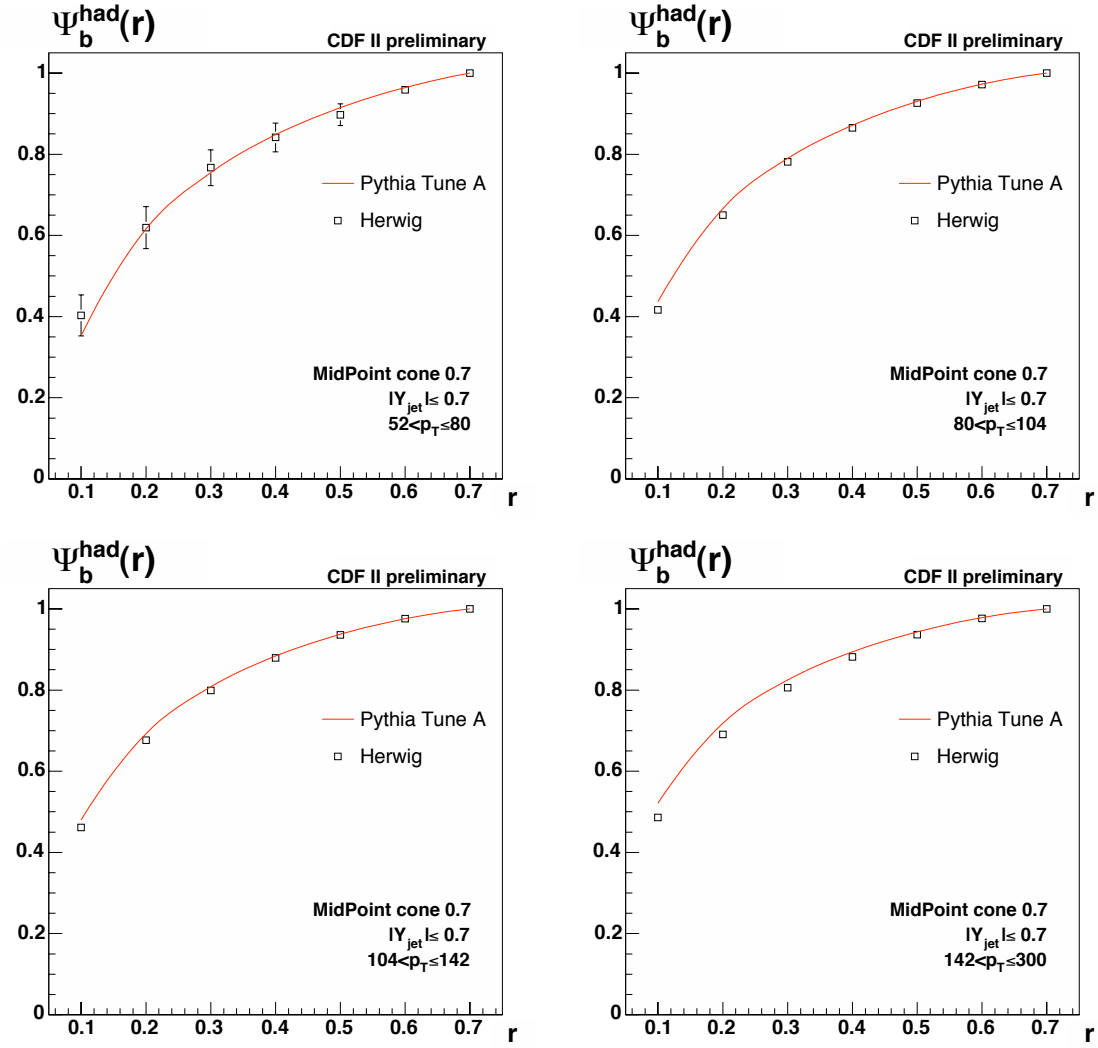

FIGURE 7.25: Comparison between Herwig (black full points) and Pythia Tune A MC (red line) of the hadron level b-jet shapes. The errors shown are the MC statistical errors.

Tune A samples. However, this was not possible for the lowest jet p_T bin because of the lack of statistics in that bin.

Figure 7.29 shows the hadron level b-quark jet shapes in data using the Herwig MC samples for the unfolding. Figure 7.30 shows the differences between the hadron level b-quark jet shapes using Herwig for the unfolding and using the default Pythia Tune A MC for the unfolding. The results are within the total systematic errors. Due to the lack of statistics in the lowest p_T bin, this effect can not be included as a systematic uncertainty. It is possible to conclude from this study that the hadron level b-quark jet shapes are not particularly sensitive to the use of a particular choice of MC for the unfolding parameters. In order to make a better comparison, more Herwig MC samples would have to be generated, particularly b-filtered samples, which was beyond the scope of this analysis.

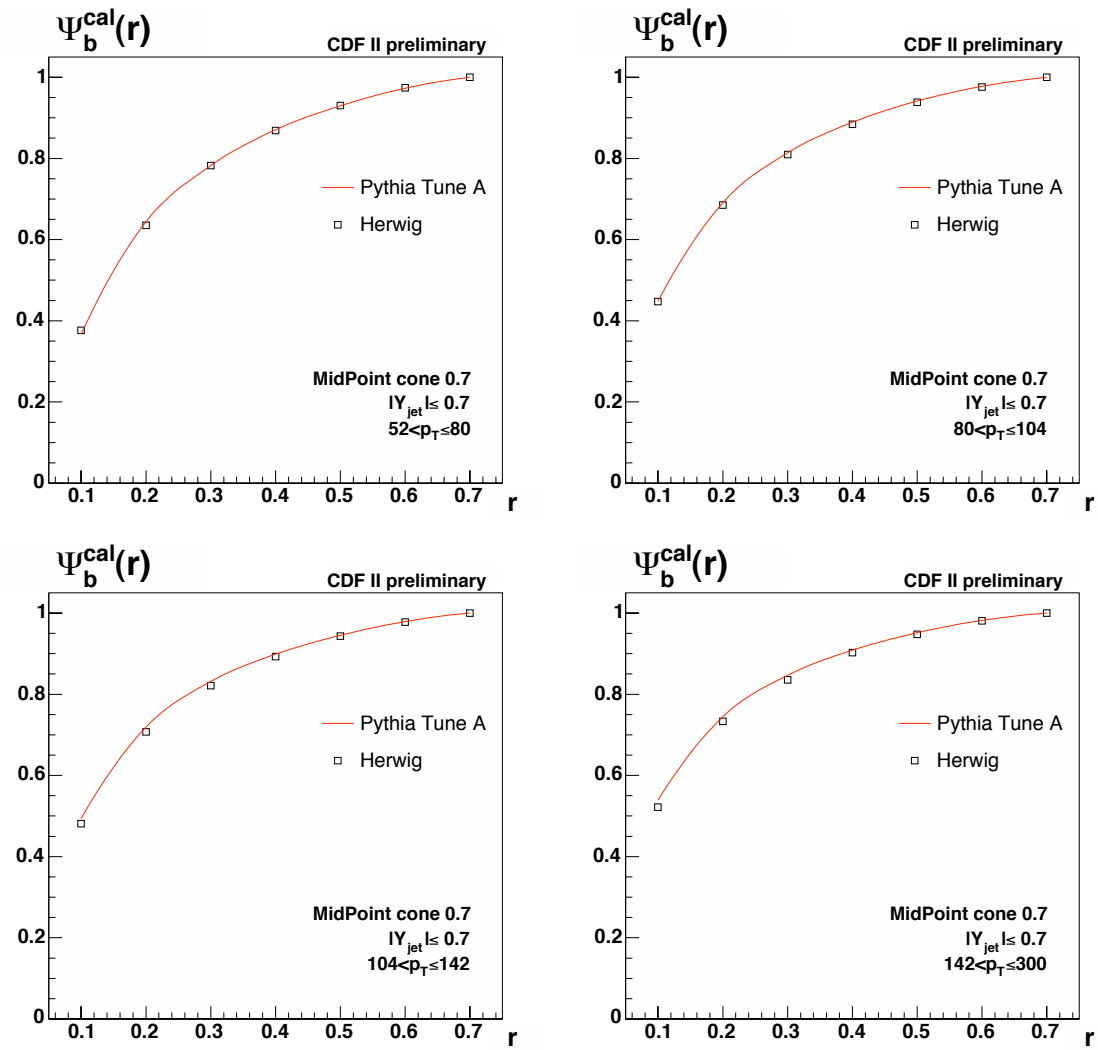

FIGURE 7.26: Comparison between Herwig (black full points) and Pythia Tune A MC (red line) of the detector level b-jet shapes. The errors shown are the MC statistical errors.

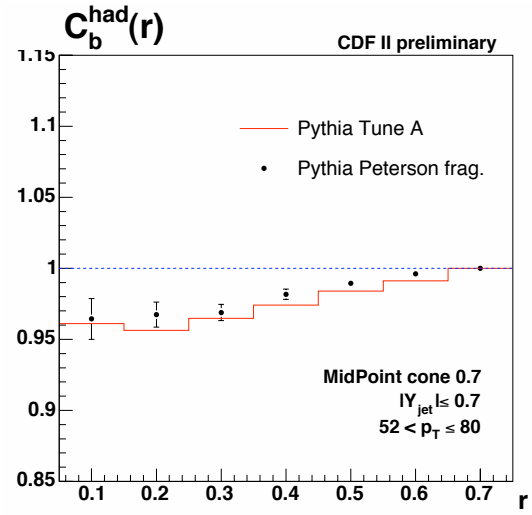

FIGURE 7.27: Comparison between Pythia Tune A using the Peterson fragmentation model (black full points) and the default Pythia Tune A MC (red line) of the hadron level corrections to b-jets. The errors shown are the MC statistical errors.

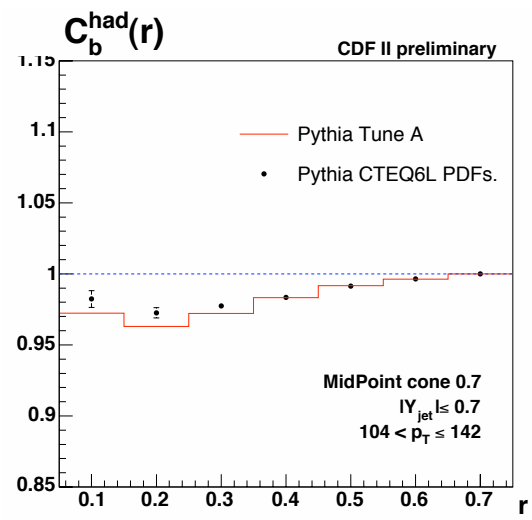

FIGURE 7.28: Comparison between Pythia Tune A using the CTEQ6L PDFs (black full points) and the default Pythia Tune A MC (red line) of the hadron level corrections to b-jets. The errors shown are the MC statistical errors.

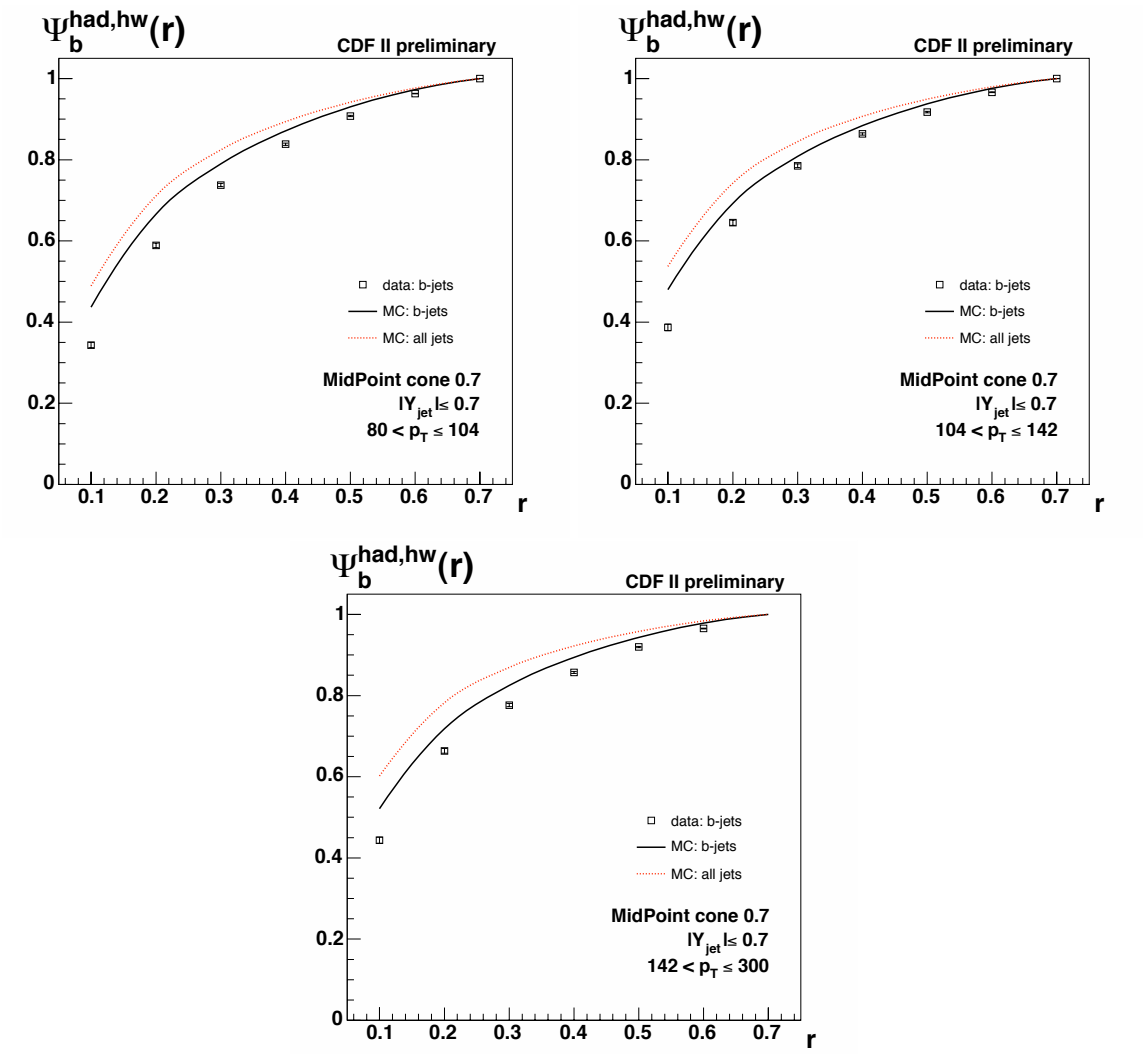

FIGURE 7.29: Hadron level b-quark jet shapes. Data (shown as black points) using the unfolding parameters obtained from the Herwig MC samples is compared to the Pythia Tune A MC predictions (red line). Only the statistical errors are shown.

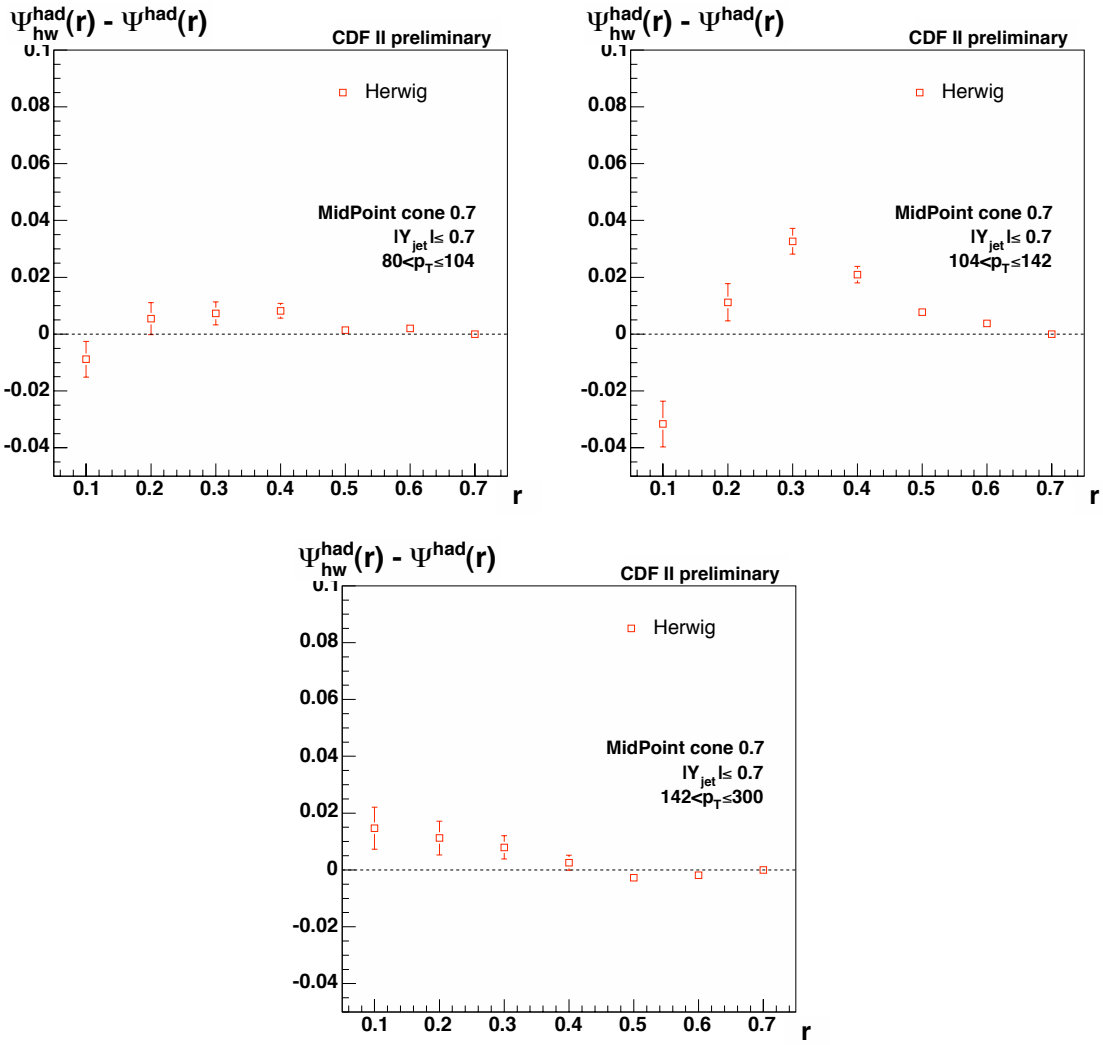

FIGURE 7.30: Difference between hadron level b-quark jet shapes using the unfolding parameters obtained from the Herwig MC samples and the ones obtained from Pythia Tune A. Only the statistical errors when using the Herwig unfolding factors are shown.

7.11 Dependence on the MC Modelling of the SecVtx Performance

Despite huge progress in the past two years in the understanding of the SecVtx performance and in particular in the differences in performance between data and MC, there are still a number of parameters, relevant to SecVtx, which do not agree perfectly between data and MC. In particular the number of Pass 1 tracks (see section 4 for the definition) does not agree perfectly between data and MC, as can be seen in figure 7.31.

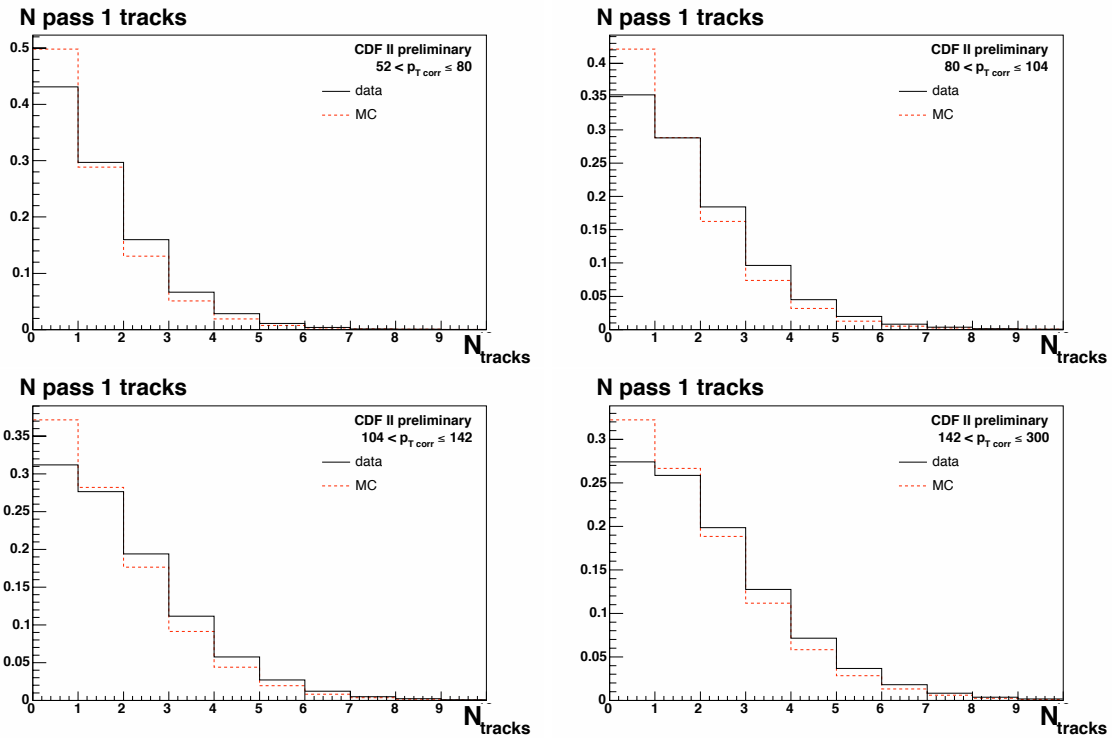

FIGURE 7.31: Plot of the distribution of the number of Pass 1 tracks inside the jet for all jets.

In order to investigate if this discrepancy between the data and the MC affects any of the unfolding parameters, the Pythia Tune A MC samples are re-weighted in such a way that the distributions of the number of Pass 1 tracks agree between data and MC. The weight function is defined for each value of the number of tracks as

$$W(n) = \frac{f_{\text{data}}(n)}{f_{\text{MC}}(n)} \tag{7.4}$$

where $f_{\text{data/MC}}(n)$ is defined as the fraction of inclusive jets which pass the selection cuts that have n Pass 1 tracks for data and MC, respectively. These functions are normalised to unit area. Having defined the weight function for each of the four p_T bins, the jet shapes in MC are re-weighted according to the number of Pass 1 tracks in the jet. The biases due to tagging on b-jets (figure 7.32) and on nonb-jets (figure 7.33) are re-computed using the new weighted shapes. The hadron level corrections are also re-computed (figure 7.34). These biases and hadron level corrections are compared to the default values obtained for the unweighted shapes. The biases are always found to be within the statistical errors of the default biases. The hadron level corrections are also compatible, within statistical errors, to those obtained using the default MC shapes. There is therefore no systematic uncertainty associated to the slight differences with

respect to data of the modelling of the SecVtx performance in Pythia Tune A MC.

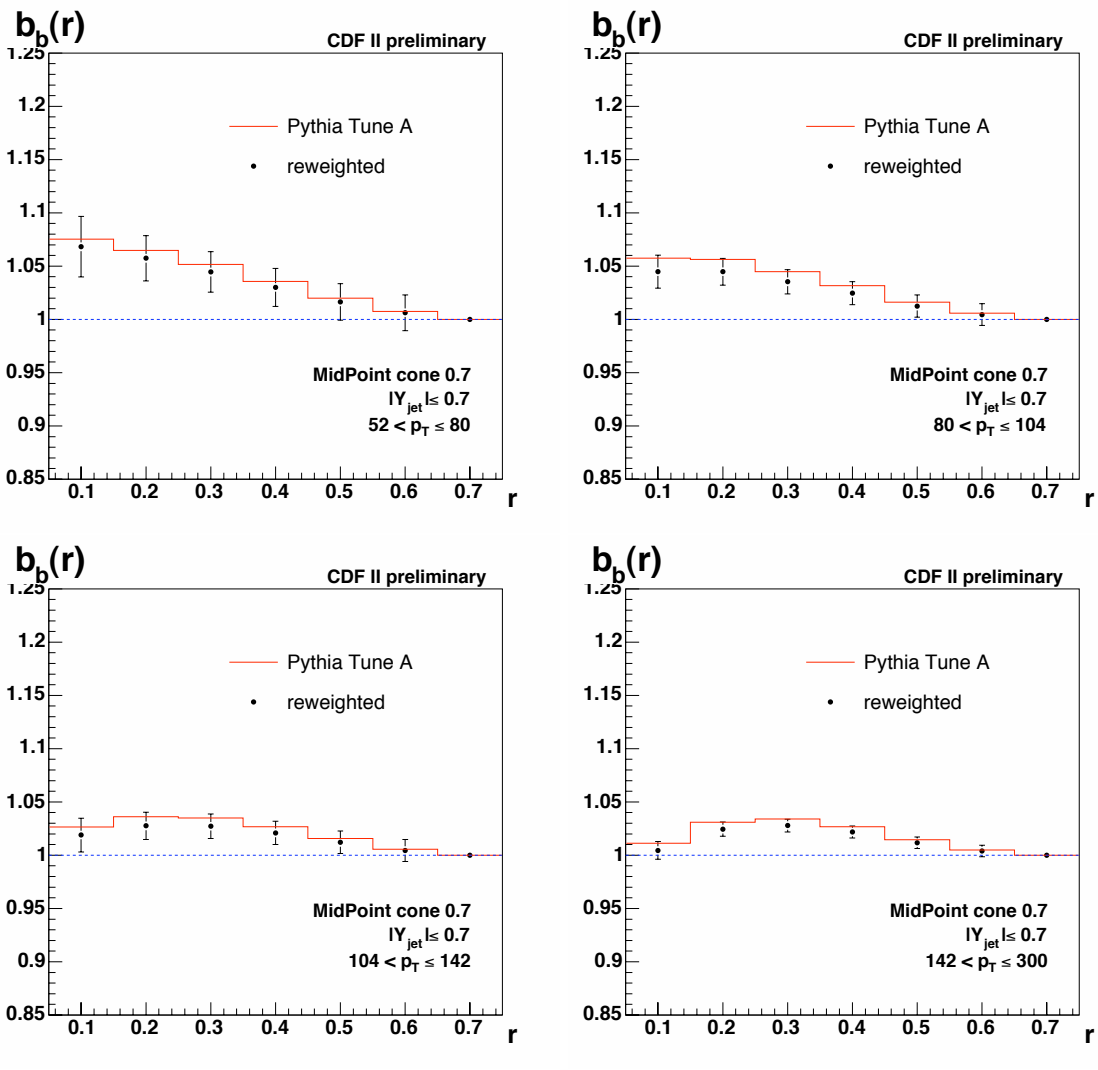

FIGURE 7.32: Comparison of the tagging biases on b-jets between the default Pythia Tune A MC shapes (red line) and the reweighted ones (black points). Only the MC statistical errors are shown.

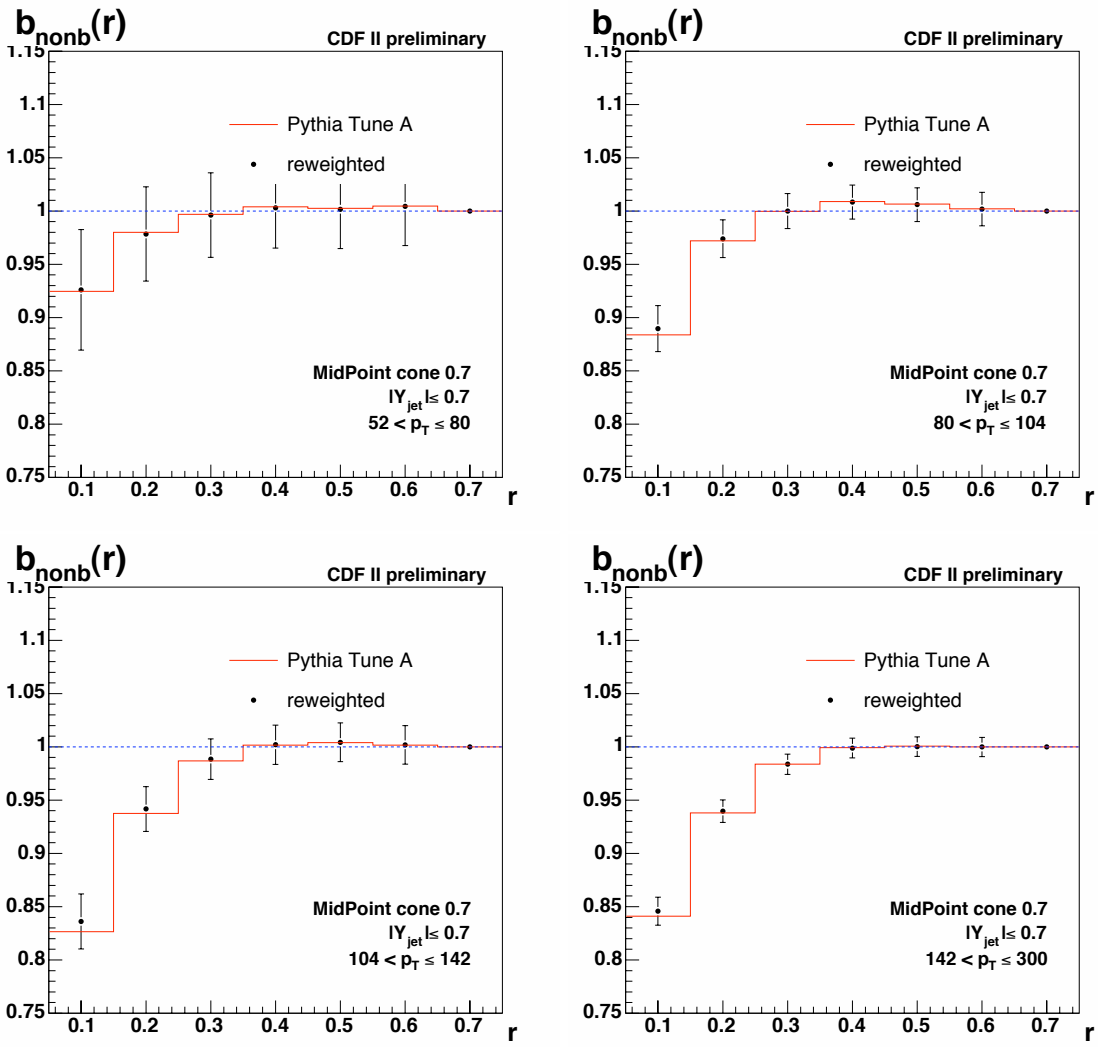

FIGURE 7.33: Comparison of the tagging biases on nonb-jets between the default Pythia Tune A MC shapes (red lines) and the reweighted ones (black points). Only the MC statistical errors are shown.

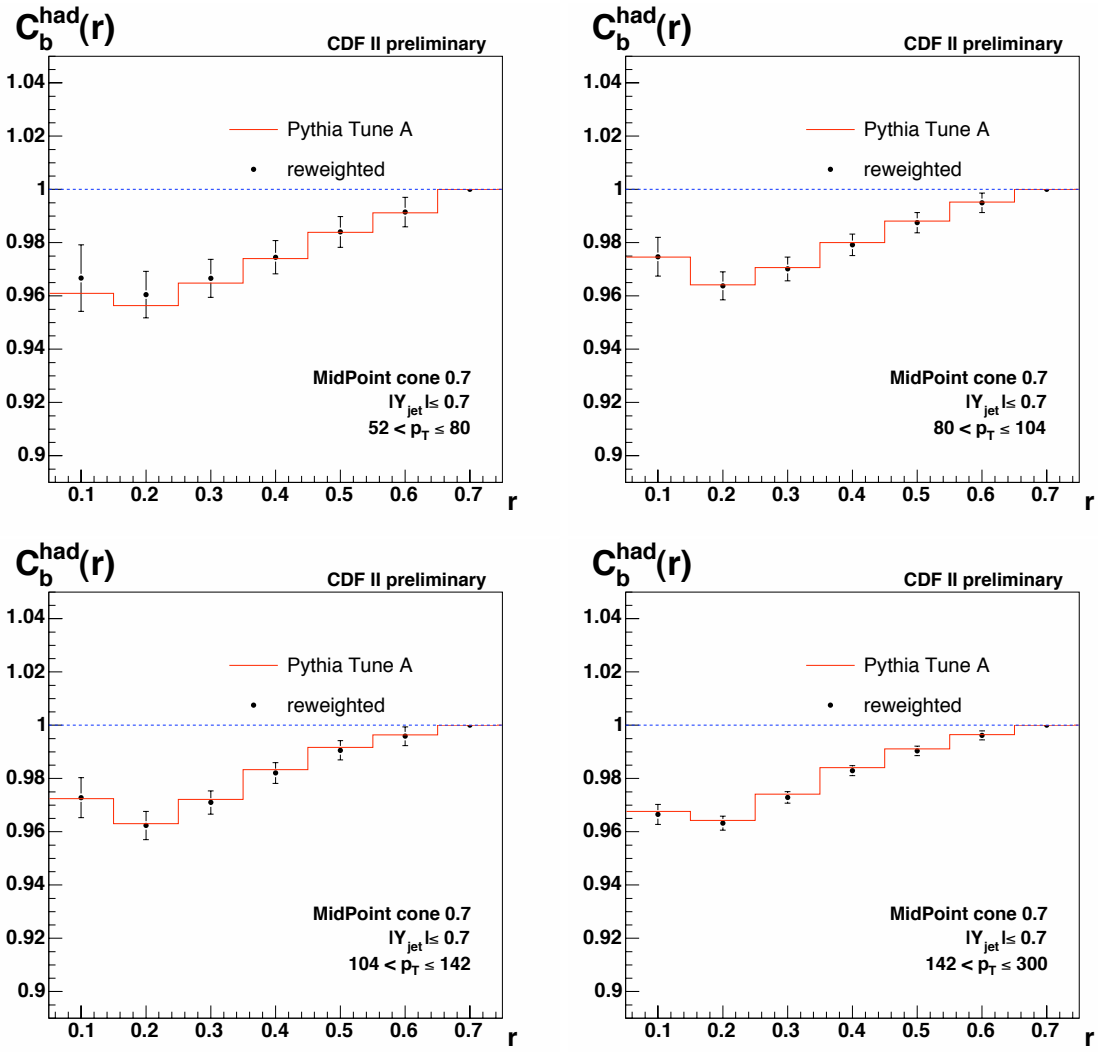

FIGURE 7.34: Comparison of the b-jet hadron level corrections between the default Pythia Tune A MC shapes (red lines) and the reweighted ones (black points). Only the MC statistical errors are shown.

7.12 Total Systematic Errors

The hadron level b-jet shapes are computed from equation 6.7 for the default case as well as for each of the systematic variations mentioned in the previous sections. The systematic variations considered for this analysis are

- jet p_T bins $\pm 3\%$
- missing E_T significance cut $\pm 15\%$
- $|Z_{\mathrm{vtx}}|$ cut $\pm 5\,\mathrm{cm}$
- $f_{1\mathrm{b}} \pm 20\%$ (absolute)
- $f_{1\mathrm{c}} \pm 20\%$ (absolute)
- $f_{\mathrm{c}} \pm 5\%$ (absolute)
- jet shapes measured using tracks
- jet shapes measured using calorimeter towers with $p_T > 0.5\,\mathrm{GeV}$

The total, statistical and systematic uncertainties are shown in table 7.4 for each p_T bin and r bin. Also shown are the various contributions from the individual effects.

The dominant sources of systematic errors vary as a function of the p_T bin. These are

- $f_{1\mathrm{c}}$ variation,
- $f_{1\mathrm{b}}$ variation,
- the difference between the b-jet shapes reconstructed from tracks and from calorimeter towers,
- MC statistics.

Despite the fact that the limited amount of statistics in the MC samples leads to a significantly large errors on the hadron level b-quark jets, particularly for the nonb-jet tagging biases and the nonb-jet secondary vertex mass templates, this effect is never the largest systematic one. Reducing the error due to the MC statistics will not significantly reduce the total systematic error.

The systematic uncertainties were also computed in an alternative manner [2]. The total systematic and statistical errors were computed for each of the parameters individually. These errors were then combined to get the total systematic and statistical errors on the hadron level b-quark jet shapes. In this different approach, the systematic uncertainty calculation is assuming no correlations between the different parameters but is at risk of double-counting some of the effects. The total systematic errors obtained using this method were found to be very similar to the ones obtained using the default method. They turn out to be very slightly higher, by a maximum of 15%, than the ones obtained using the approach detailed above.

52 − 80 GeV												
				Individual systematic errors								
r bin	tot	stat tot	syst tot	MC stats	$p_T \pm 3\%$	$\lvert Z_{\mathrm{vtx}}\rvert \pm 5\,\mathrm{cm}$	$\not{E}_T$ sig $\pm 15\%$	$f_{1\mathrm{b}} \pm 20\%$	$f_{1\mathrm{c}} \pm 20\%$	$f_{\mathrm{c}} \pm 5\%$	tracks	tower $p_T > 0.5\,\mathrm{GeV}$
0.1	0.072	0.011	0.071	0.035	0.011	0.001	0.000	0.011	0.032	0.008	0.050	0.002
0.2	0.055	0.010	0.055	0.034	0.001	0.002	0.000	0.014	0.027	0.007	0.029	0.005
0.3	0.045	0.007	0.044	0.025	0.006	0.001	0.001	0.011	0.018	0.005	0.028	0.006
0.4	0.027	0.005	0.027	0.016	0.002	0.001	0.000	0.007	0.008	0.004	0.017	0.006
0.5	0.014	0.003	0.013	0.010	0.000	0.000	0.000	0.003	0.004	0.002	0.003	0.006
0.6	0.008	0.002	0.007	0.004	0.001	0.000	0.000	0.001	0.000	0.001	0.005	0.002
80 − 104 GeV												
				Individual systematic errors								
r bin	tot	stat tot	syst tot	MC stats	$p_T \pm 3\%$	$\lvert Z_{\mathrm{vtx}}\rvert \pm 5\,\mathrm{cm}$	$\not{E}_T$ sig $\pm 15\%$	$f_{1\mathrm{b}} \pm 20\%$	$f_{1\mathrm{c}} \pm 20\%$	$f_{\mathrm{c}} \pm 5\%$	tracks	tower $p_T > 0.5\,\mathrm{GeV}$
0.1	0.040	0.006	0.040	0.015	0.007	0.001	0.000	0.012	0.033	0.006	0.005	0.001
0.2	0.031	0.006	0.030	0.013	0.012	0.002	0.000	0.010	0.019	0.004	0.011	0.002
0.3	0.024	0.004	0.024	0.009	0.009	0.001	0.000	0.008	0.010	0.003	0.015	0.003
0.4	0.017	0.003	0.017	0.006	0.003	0.000	0.000	0.005	0.005	0.002	0.014	0.003
0.5	0.013	0.002	0.013	0.003	0.003	0.000	0.000	0.002	0.003	0.001	0.011	0.003
0.6	0.006	0.001	0.006	0.002	0.002	0.000	0.000	0.001	0.001	0.000	0.005	0.002
104 − 142 GeV												
				Individual systematic errors								
r bin	tot	stat tot	syst tot	MC stats	$p_T \pm 3\%$	$\lvert Z_{\mathrm{vtx}}\rvert \pm 5\,\mathrm{cm}$	$\not{E}_T$ sig $\pm 15\%$	$f_{1\mathrm{b}} \pm 20\%$	$f_{1\mathrm{c}} \pm 20\%$	$f_{\mathrm{c}} \pm 5\%$	tracks	tower $p_T > 0.5\,\mathrm{GeV}$
0.1	0.055	0.008	0.054	0.025	0.014	0.003	0.001	0.011	0.030	0.005	0.033	0.000
0.2	0.030	0.007	0.029	0.021	0.000	0.001	0.001	0.009	0.015	0.003	0.010	0.000
0.3	0.018	0.005	0.018	0.014	0.004	0.000	0.000	0.007	0.006	0.003	0.002	0.001
0.4	0.016	0.003	0.016	0.009	0.003	0.001	0.000	0.005	0.002	0.002	0.011	0.001
0.5	0.011	0.002	0.011	0.005	0.001	0.000	0.000	0.002	0.001	0.001	0.009	0.001
0.6	0.005	0.001	0.005	0.002	0.001	0.000	0.000	0.001	0.000	0.001	0.003	0.001
142 − 300 GeV												
				Individual systematic errors								
r bin	tot	stat tot	syst tot	MC stats	$p_T \pm 3\%$	$\lvert Z_{\mathrm{vtx}}\rvert \pm 5\,\mathrm{cm}$	$\not{E}_T$ sig $\pm 15\%$	$f_{1\mathrm{b}} \pm 20\%$	$f_{1\mathrm{c}} \pm 20\%$	$f_{\mathrm{c}} \pm 5\%$	tracks	tower $p_T > 0.5\,\mathrm{GeV}$
0.1	0.036	0.008	0.035	0.020	0.007	0.000	0.001	0.014	0.018	0.003	0.016	0.000
0.2	0.026	0.006	0.025	0.016	0.009	0.001	0.001	0.011	0.007	0.002	0.010	0.000
0.3	0.016	0.004	0.016	0.011	0.007	0.000	0.000	0.008	0.002	0.004	0.002	0.001
0.4	0.016	0.003	0.016	0.007	0.002	0.000	0.000	0.006	0.001	0.003	0.012	0.001
0.5	0.010	0.002	0.010	0.004	0.000	0.000	0.000	0.002	0.000	0.001	0.009	0.002
0.6	0.010	0.001	0.010	0.002	0.000	0.000	0.000	0.001	0.000	0.001	0.009	0.002

TABLE 7.4: Total, statistical and systematic errors on the final b-jet shapes along with the contributions from each parameter. The last bins in r always have by definition zero error so they are not displayed.

7.13 Changing the Rapidity Cut to $0.1 \leq |Y| \leq 0.7$

In order to compare the hadron level b-jet shapes to the previously published inclusive jet shapes, it is important to see if the different cut requirements on the rapidity of the jets have an effect on the measured shapes. Figure 7.35 shows a comparison, in data, between the inclusive shapes using a rapidity cut on the jet of $|Y| \leq 0.7$ (black curve) and those measured removing the central rapidity region, i.e. requiring $0.1 \leq |Y| \leq 0.7$ (red curve). Figure 7.36 shows the ratio of the shapes removing the central rapidity region over the shapes including the central rapidity region. Figure 7.37 shows a comparison between the shapes for tagged jets using the two different rapidity cuts. Figure 7.38 shows the ratio of the shapes removing the central rapidity region over the shapes including the central rapidity region. Both these sets of figures show that the measured shapes are very similar for both cases; they are within less than 1% of eachother. This allows the published inclusive jet shape results to be plotted on the same figures as the b-jet shape results.

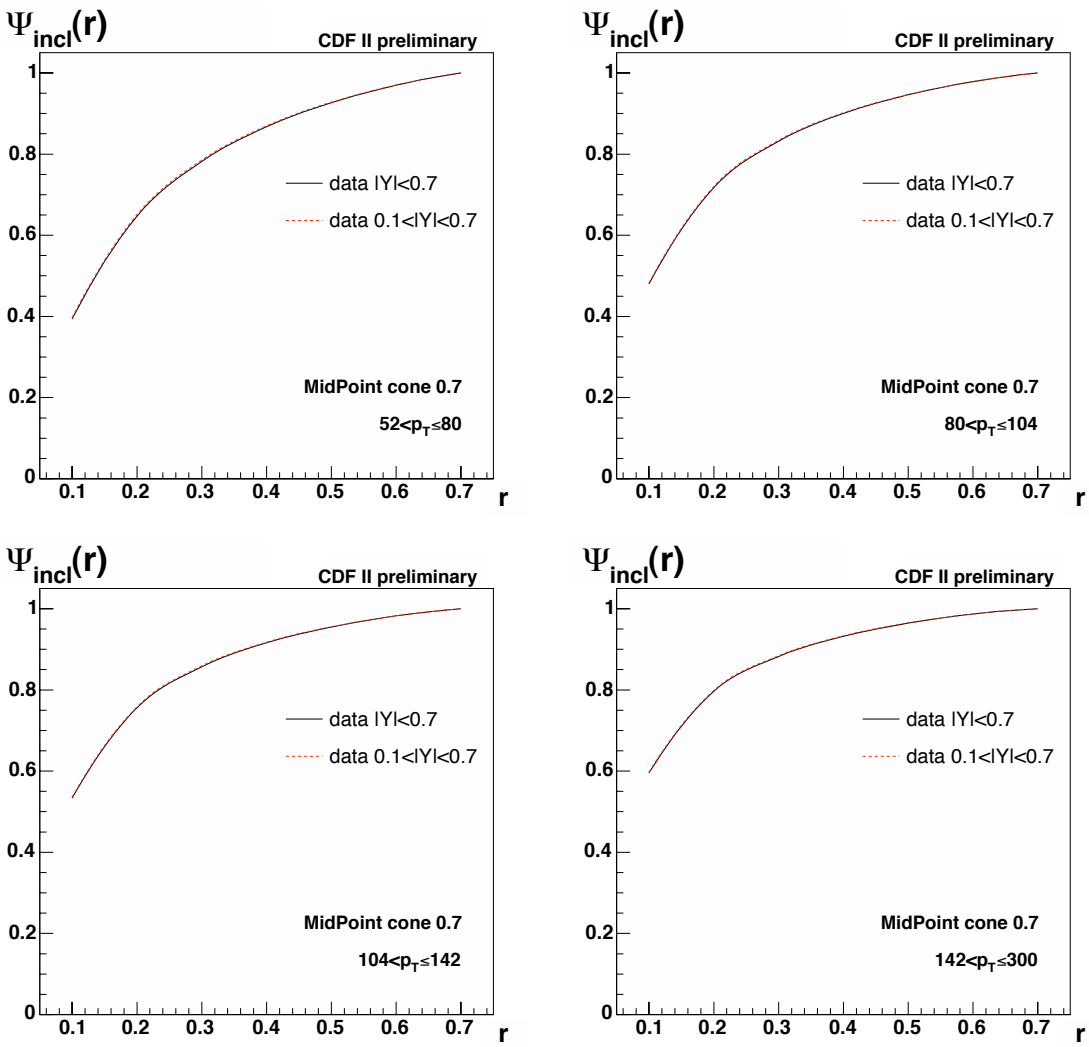

FIGURE 7.35: Comparison in data between the inclusive shapes with the rapidity cut $|Y| \leq 0.7$ (black curve) and the cuts used for the measurement of the inclusive jet shapes, $0.1 \leq |Y| \leq 0.7$ (red curve).

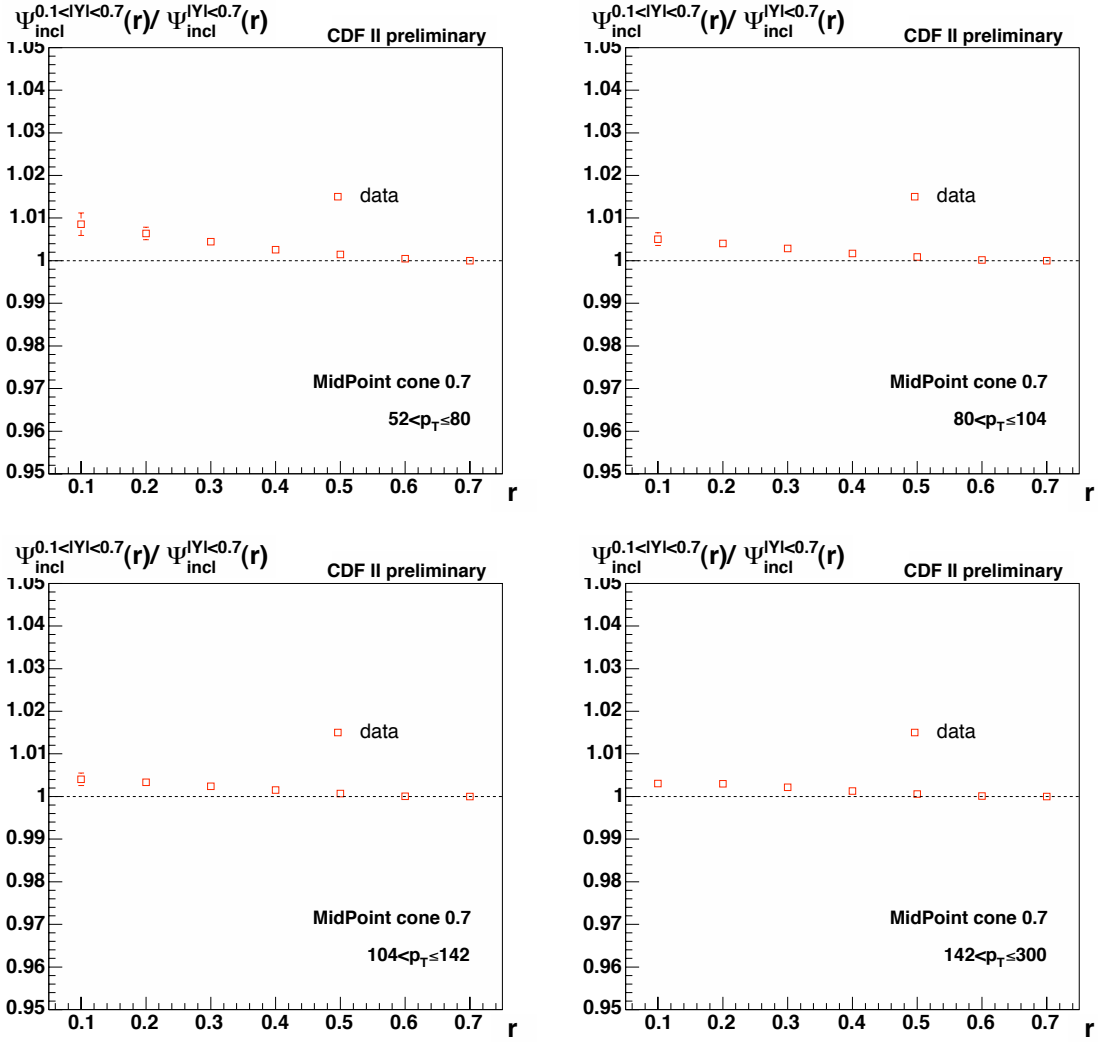

FIGURE 7.36: Ratio in data between the inclusive shapes with the rapidity cut $0.1 \leq |Y| \leq 0.7$ over the shapes with $|Y| \leq 0.7$ (red points). Only the statistical errors of the measured shapes with $0.1 \leq |Y| \leq 0.7$ are shown.

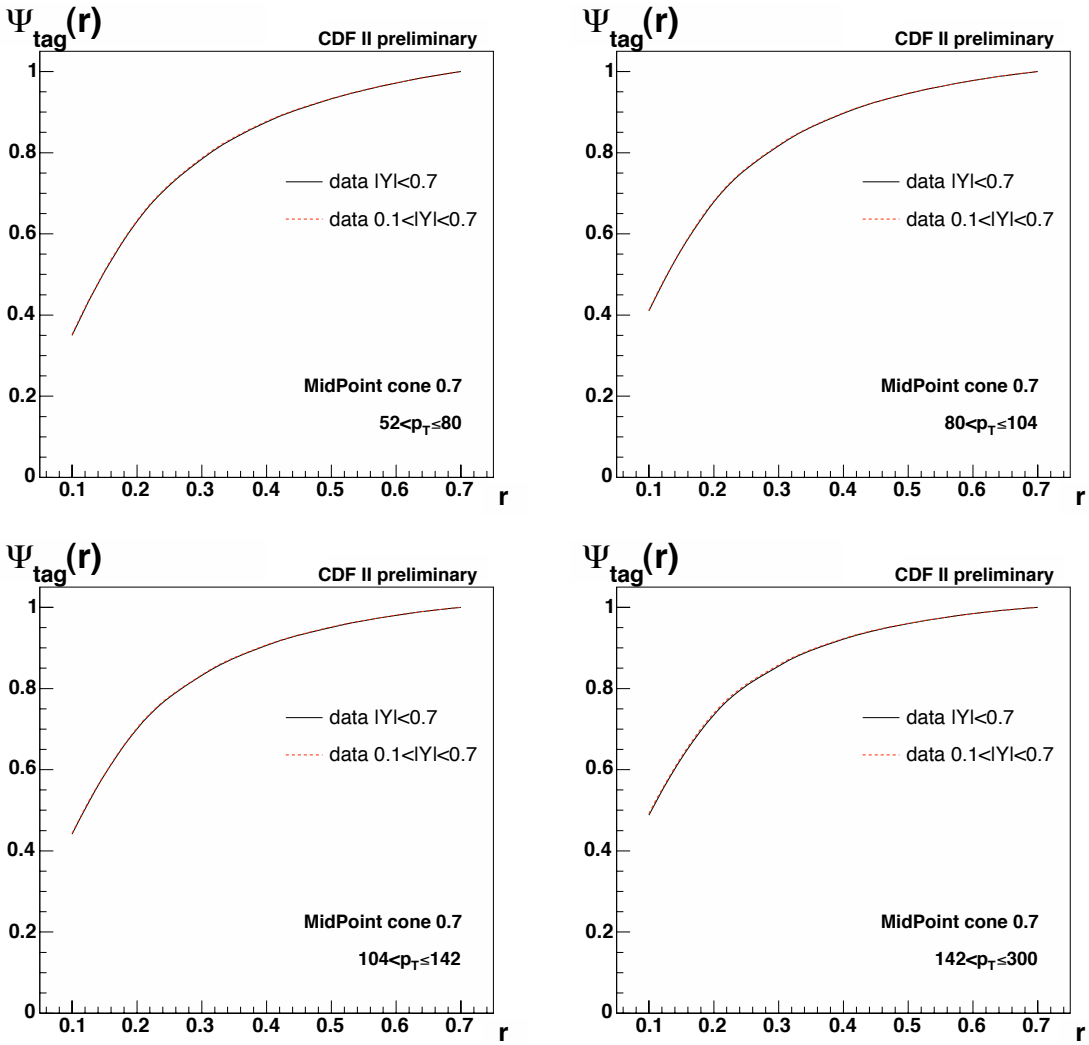

FIGURE 7.37: Comparison in data between tagged jet shapes with the rapidity cut $|Y| \leq 0.7$ (black curve) and the cuts used for the measurement of the inclusive jet shapes, $0.1 \leq |Y| \leq 0.7$ (red curve).

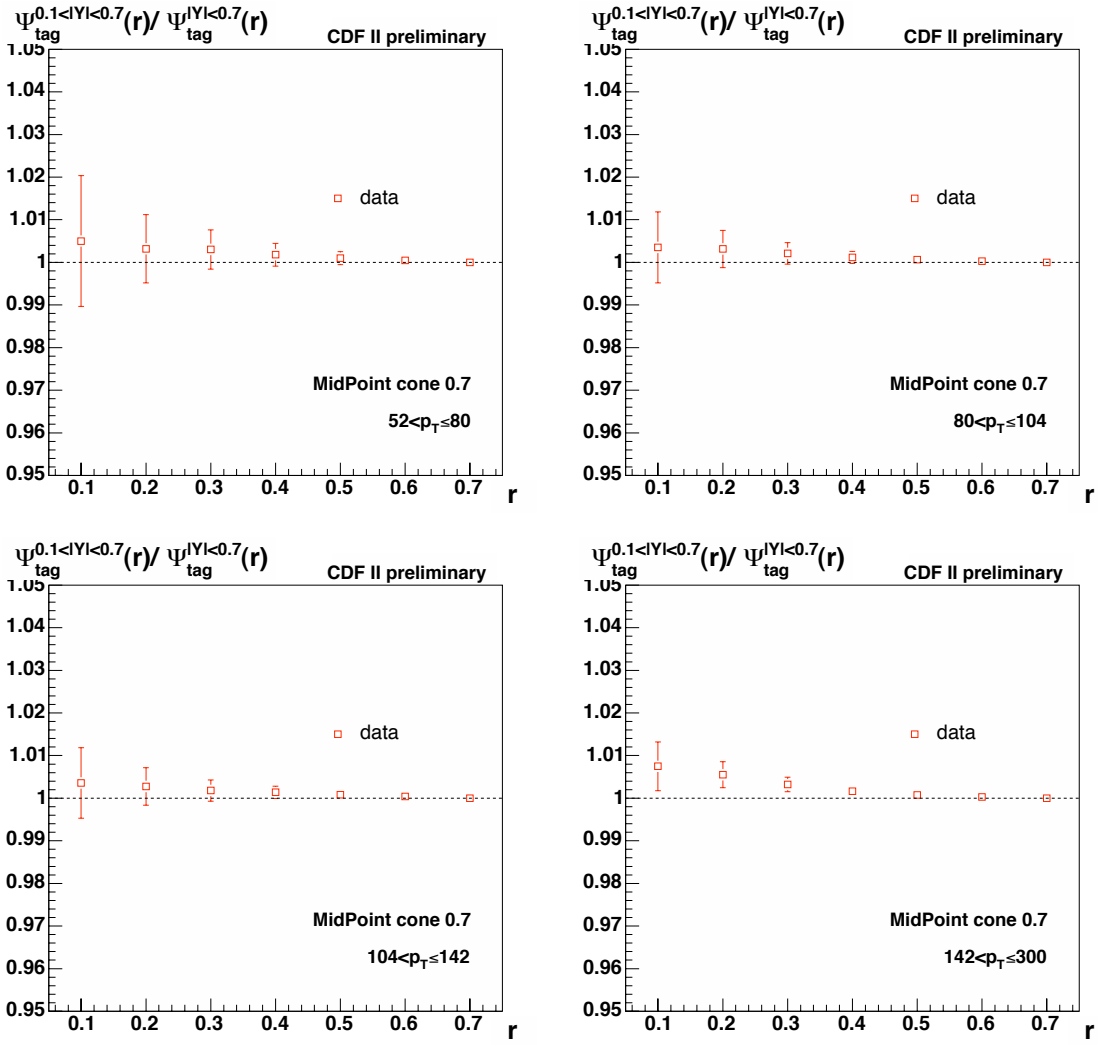

FIGURE 7.38: Ratio in data between the tagged jet shapes with the rapidity cut $0.1 \leq |Y| \leq 0.7$ over the shapes with $|Y| \leq 0.7$ (red points). Only the statistical errors of the measured shapes with $0.1 \leq |Y| \leq 0.7$ are shown.

Chapter 8

Results

Combining the measurements of all the different terms of equation 6.7, the b-jet shapes can be calculated. Section 8.1 shows that it is possible, starting from the Pythia Tune A MC predictions for the inclusive and tagged jet shapes, to reconstruct the detector level and hadron level b-quark jet shapes predicted by the same MC. This is done for shapes measured using calorimeter towers and for those measured using tracks. Section 8.2 presents the final results for the integrated jet shapes, including all the systematic errors. Then, in section 8.3, the same results are shown but this time the fractional energy outside a fixed cone radius is plotted as a function of the p_T of the jet. These results are compared to the inclusive jet shape results and the Pythia Tune A predictions for inclusive and b-jets. Also shown is a comparison with the shapes expected from Pythia Tune A for single b-quark jets and double b-quark jets. Finally, section 8.4 shows the results of a fit to the data in order to determine the most likely $f_{1\mathrm{b}}$ fraction. This shows that the Pythia Tune A MC predictions are between 10% and 20% higher (in absolute terms) than the values found in data.

8.1 MC Reconstruction Checks

In this section, the whole analysis is carried out starting from the raw calorimeter and track shapes as measured when taking Pythia Tune A MC predictions instead of the data. This is to test that the reconstruction method actually returns the correct detector level and hadron level shapes.

Figure 8.1 shows the comparison between the detector level b-quark jet shapes and the reconstructed shapes starting from inclusive and tagged jets. In this plot, the "true" MC purity is used instead of the fraction obtained from the secondary vertex mass fit to the inclusive MC distribution. Moreover, all the tagging bias terms, for b- and for nonb-jets, are taken from the b-filtered Pythia Tune A MC. The inclusive shapes are still taken from the inclusive Pythia Tune A MC but the tagged shapes are taken from the b-filtered Pythia Tune A MC. This means that the whole procedure, except for the values of the inclusive shapes, is taken from the same MC samples as used for the comparison. Figure 8.2 shows the same plots but starting from the shapes measured using tracks. Figures 8.3 and 8.4 show the shapes unfolded back to hadron level compared to the hadron level b-jet shapes when starting from calorimeter towers and tracks, respectively. All these sets of plots show basically perfect agreement between data and MC [1].

Figures 8.5, 8.7 and 8.9 show the comparison between the unfolded Pythia Tune A MC and the hadron level predictions of Pythia Tune A when starting from raw calorimetric towers, tracks

[1] The very small differences can be attributed to the rounding of the values for each individual parameter to floats.

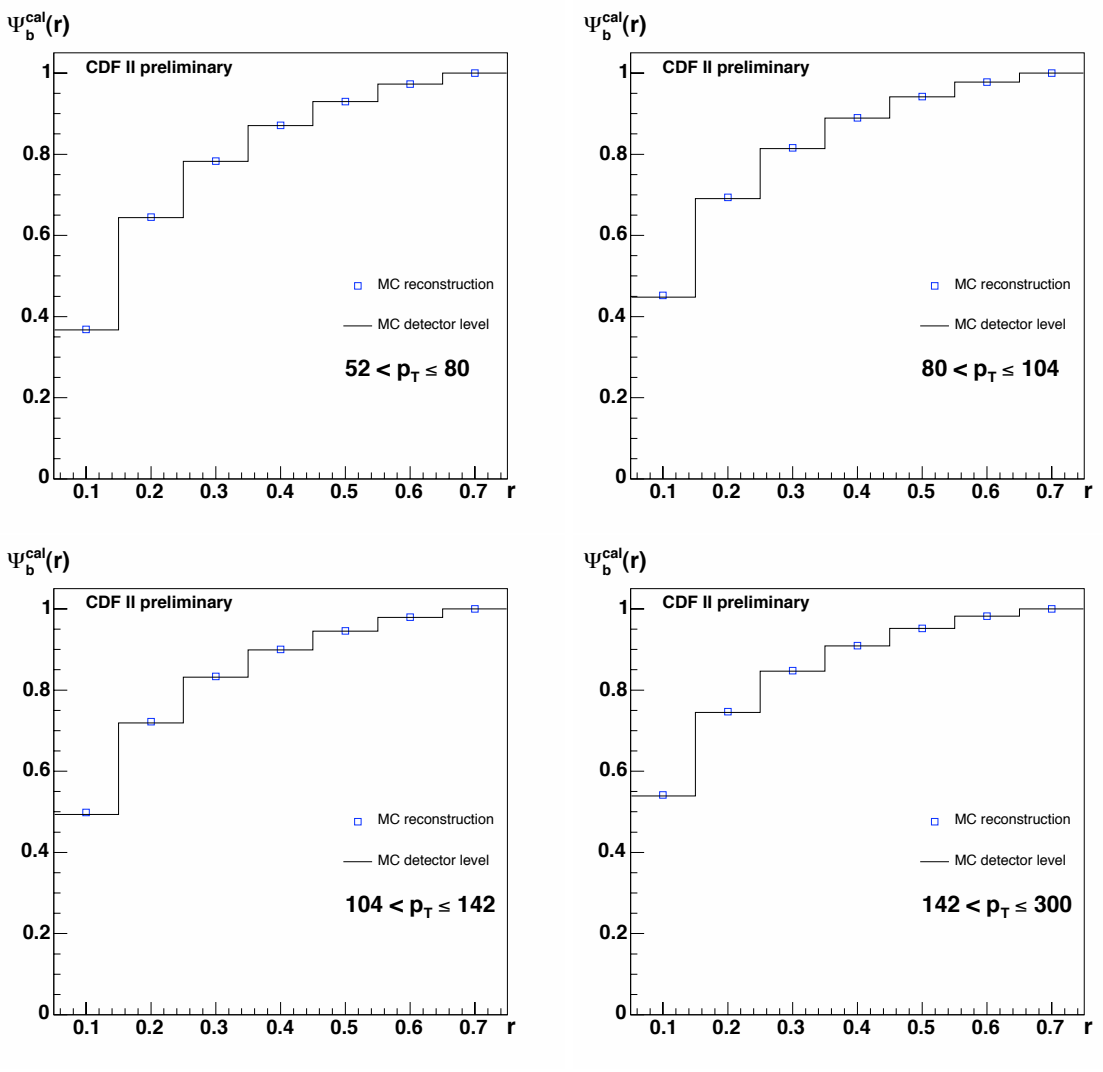

FIGURE 8.1: Detector level integrated b-jet shapes using calorimeter towers for each of the p_T bins considered. The reconstructed detector level MC shapes are shown as blue points. The Pythia Tune A MC predictions for b-jets at calorimeter level are shown as black lines.

and calorimeter towers with $p_T > 0.5\,\text{GeV}$, respectively. The differences with respect to the previous plots is that the parameters are taken from the relevant Pythia Tune A MC samples and not all from the b-filtered samples. The purities are also taken from the secondary vertex mass fit results when fitting the templates to the MC distributions for tagged jets. Figures 8.6, 8.8 and 8.10 show the ratio of the reconstructed over the "true" hadron level b-quark jet shapes for the shapes reconstructed using the calorimeter towers, the tracks and the calorimeter towers with $p_T > 0.5\,\text{GeV}$, respectively. The shapes are in good agreement with the hadron level predictions.

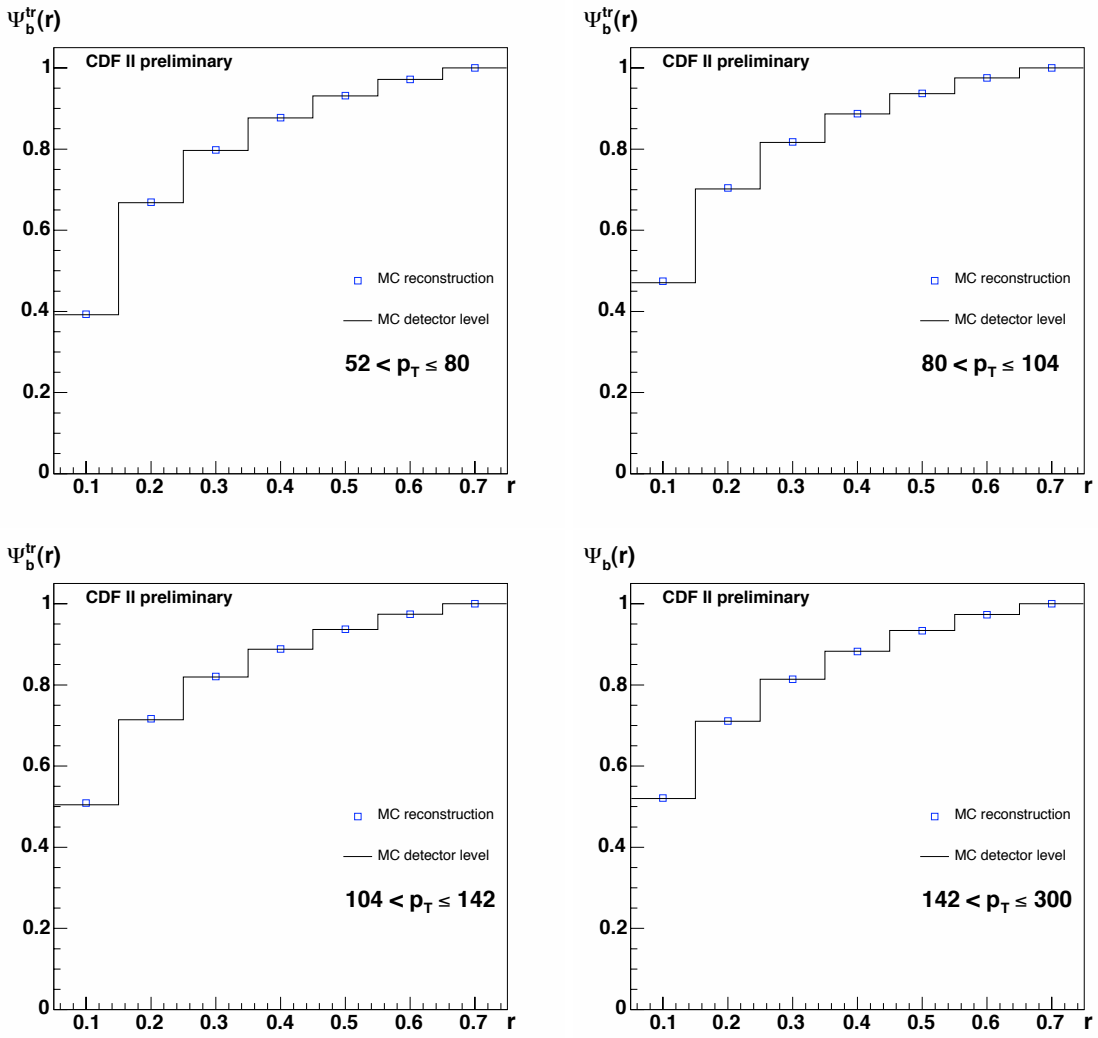

FIGURE 8.2: Detector level integrated b-jet shapes using tracks for each of the p_T bins considered. The reconstructed detector level MC shapes are shown as blue points. The MC predictions for b-jets at track level are shown as black lines.

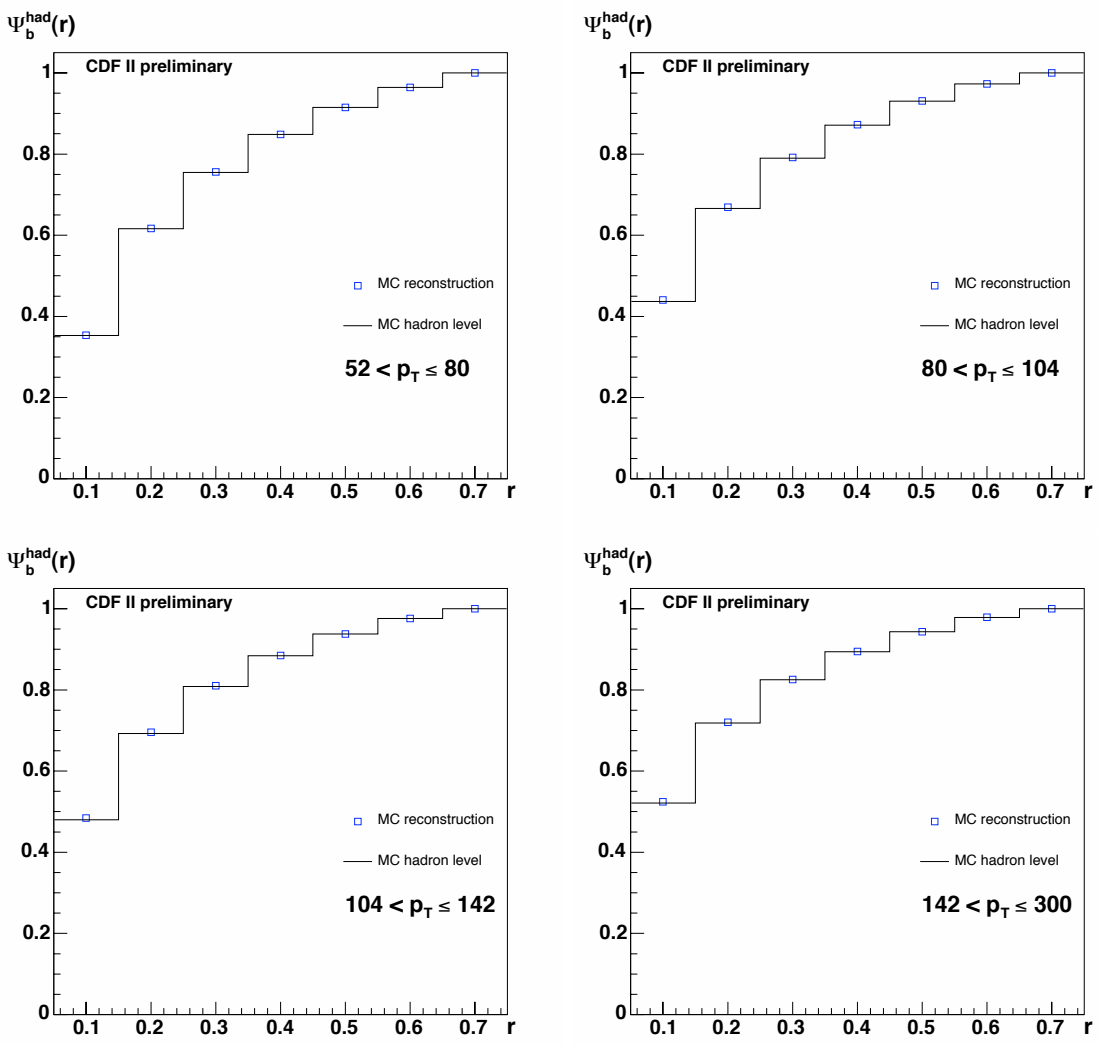

FIGURE 8.3: Hadron level integrated b-jet shapes for each of the p_T bins considered. The reconstructed hadron level MC shapes are shown as black points. The MC predictions for b-jets at hadron level are shown as black lines.

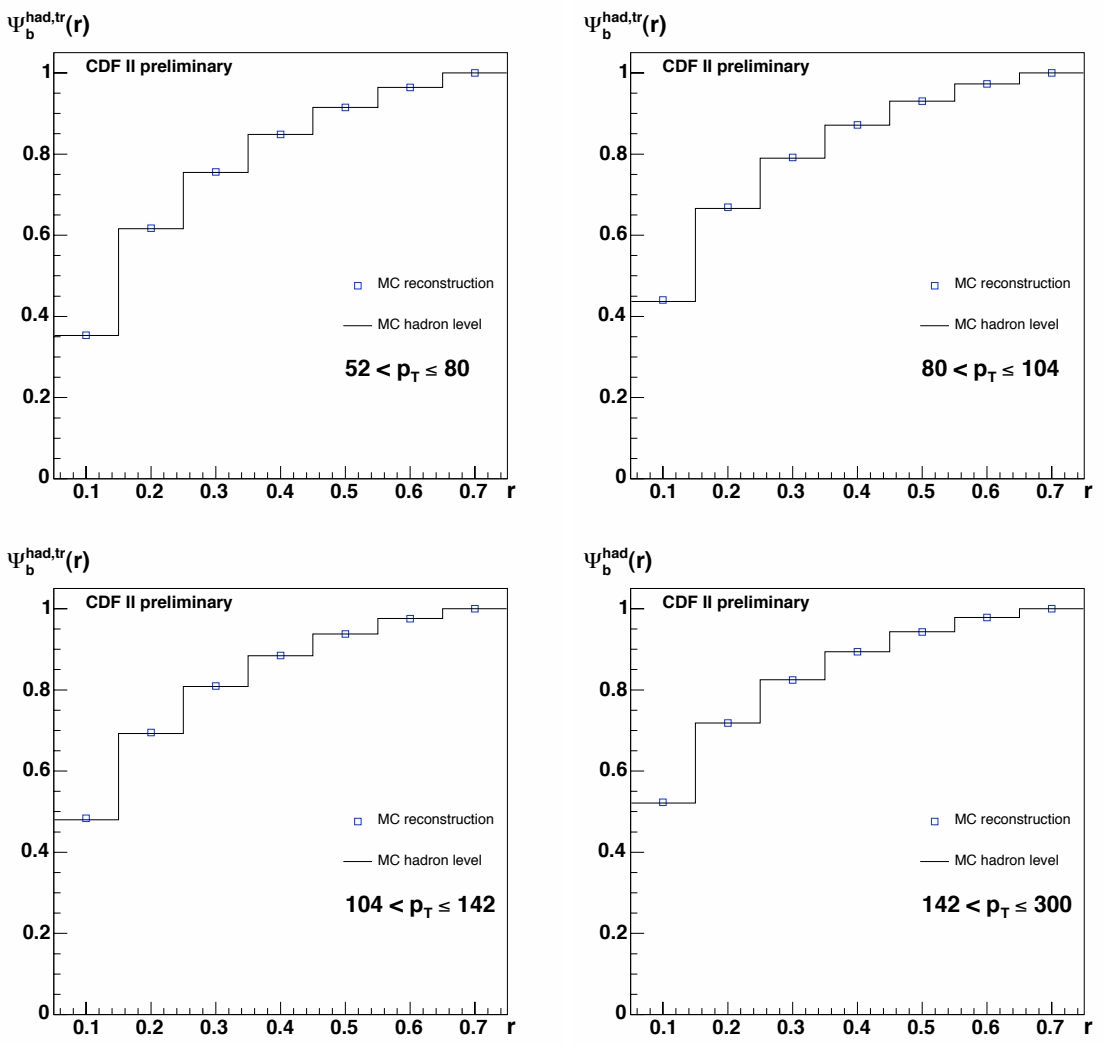

FIGURE 8.4: Hadron level integrated b-jet shapes for each of the p_T bins considered. The reconstructed hadron level MC shapes, starting from tracks, are shown as black points. The MC predictions for b-jets at hadron level are shown as black lines.

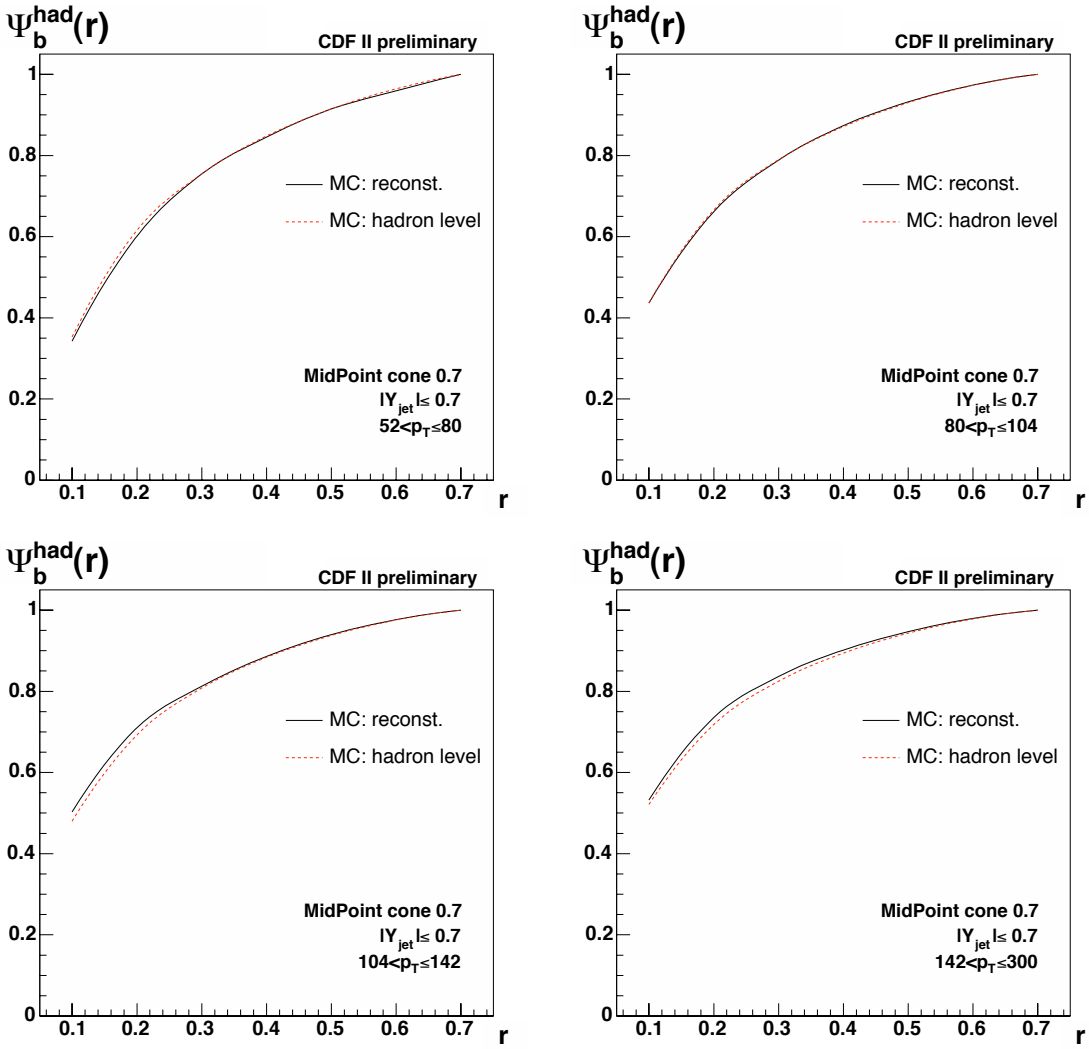

FIGURE 8.5: Hadron level integrated b-jet shapes in Pythia Tune A MC for each of the p_T bins considered. The reconstructed hadron level MC shapes are shown as black lines. The MC predictions for b-jets at hadron level are shown as red dashed curves.

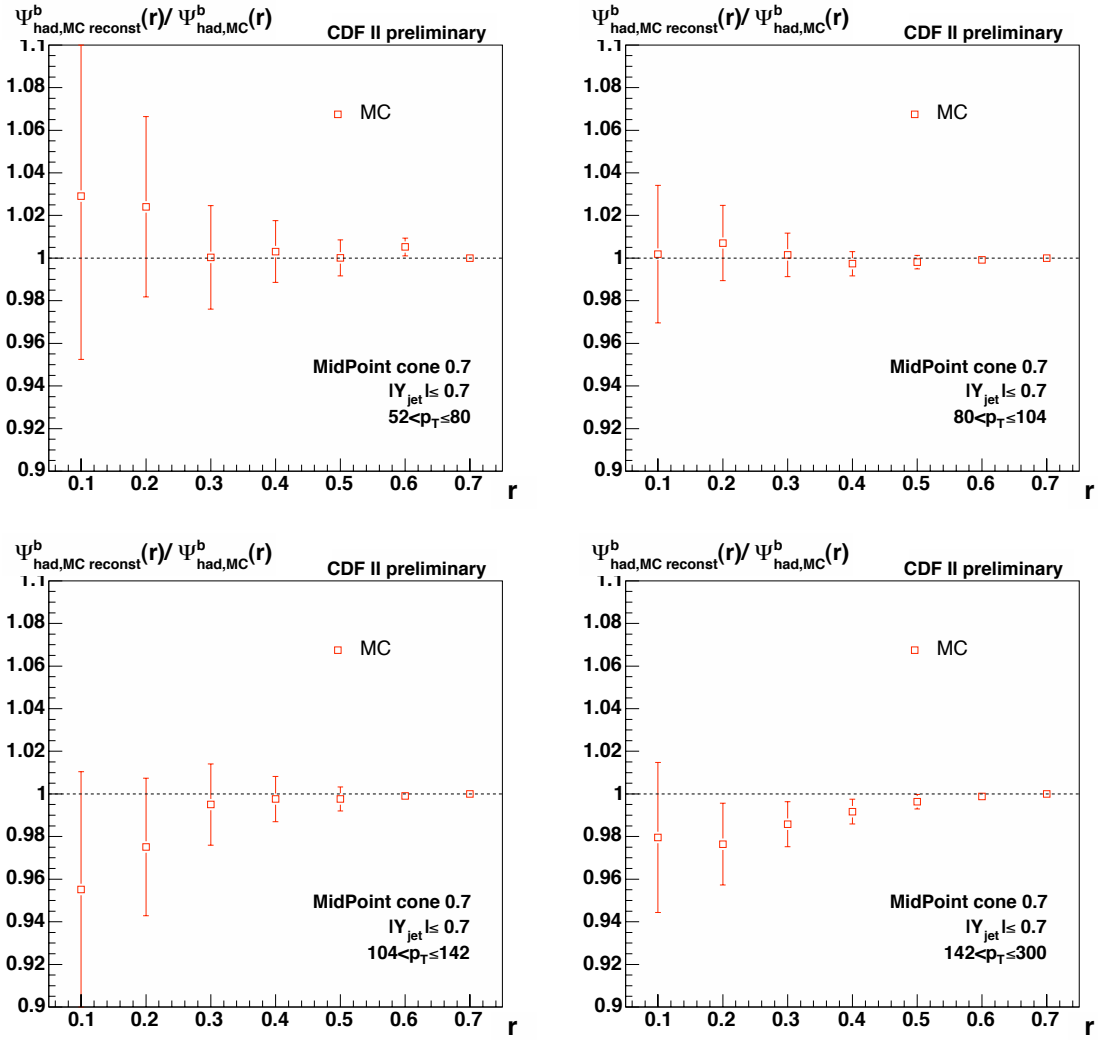

FIGURE 8.6: Ratio of the reconstructed hadron level Pythia Tune A MC shapes over the "true" hadron level Pythia Tune A MC shapes. The errors shown are only the statistical errors of the reconstructed shapes.

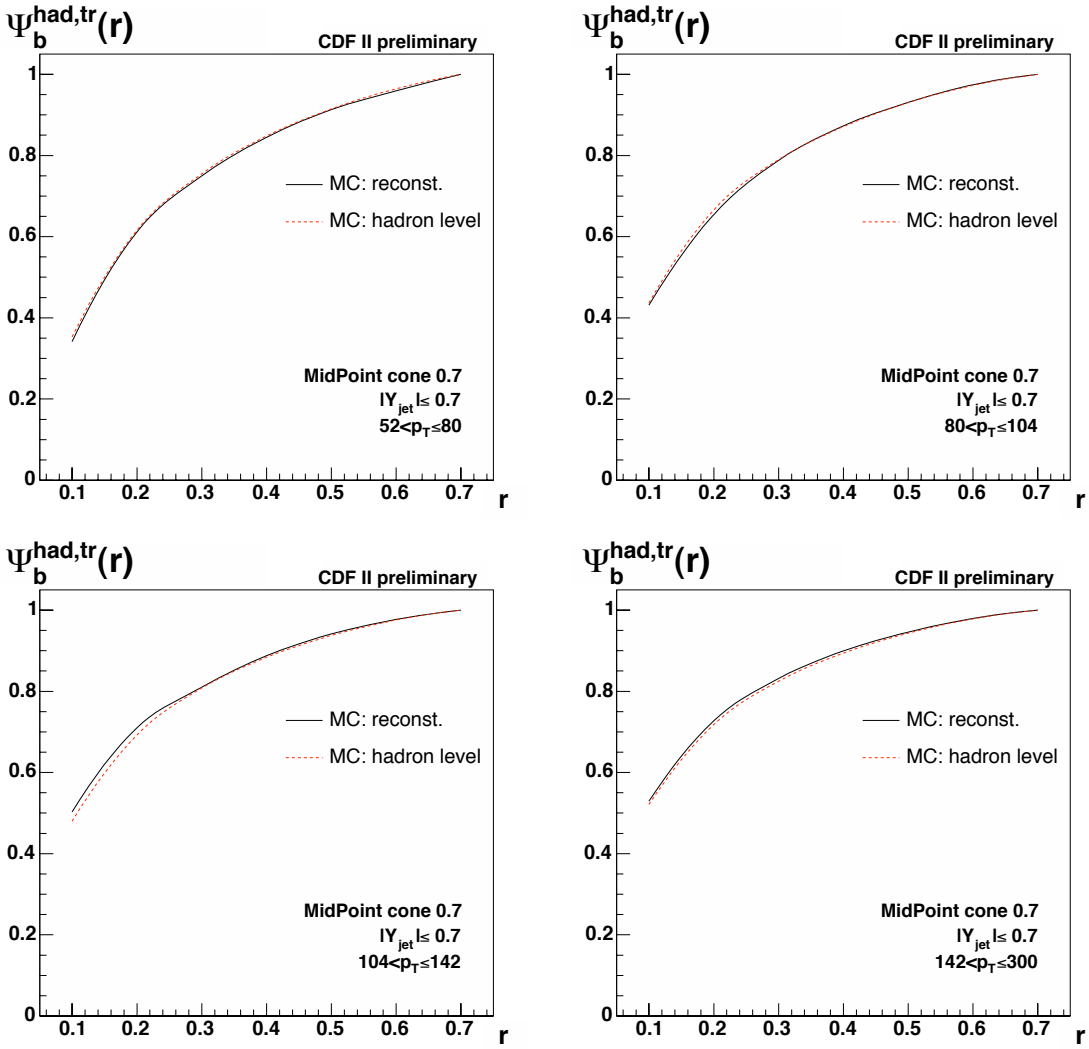

FIGURE 8.7: Hadron level integrated b-jet shapes in Pythia Tune A MC for each of the p_T bins considered. The reconstructed hadron level MC shapes, starting from tracks, are shown as black lines. The MC predictions for b-jets at hadron level are shown as red dashed curves.

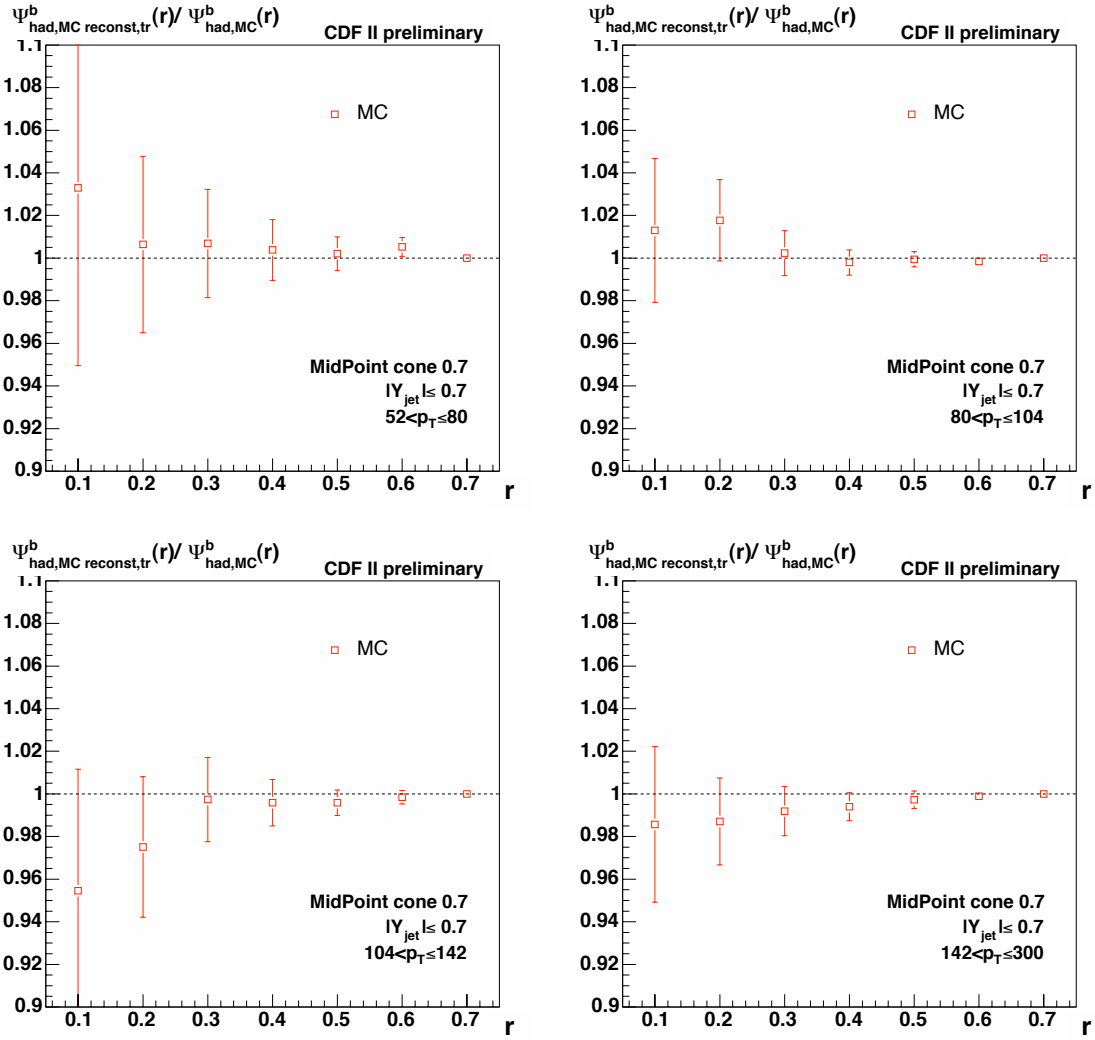

FIGURE 8.8: Ratio of the reconstructed hadron level Pythia Tune A MC shapes, starting from raw track shapes, over the "true" hadron level Pythia Tune A MC shapes. The errors shown are only the statistical errors of the reconstructed shapes.

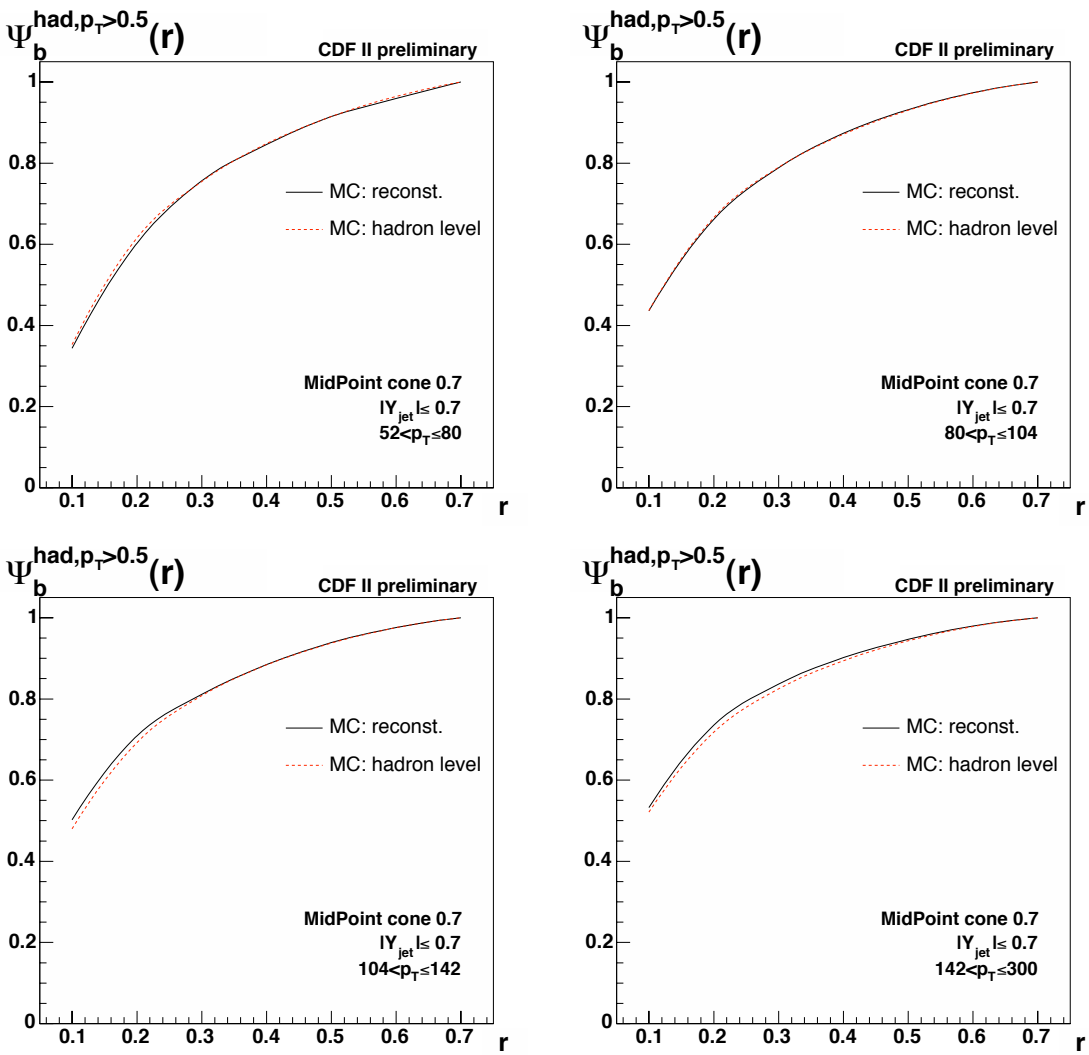

FIGURE 8.9: Hadron level integrated b-jet shapes in Pythia Tune A MC for each of the p_T bins considered. The reconstructed hadron level MC shapes, starting from calorimeter towers with p_T larger than 0.5 GeV, are shown as black lines. The MC predictions for b-jets at hadron level are shown as red dashed curves.

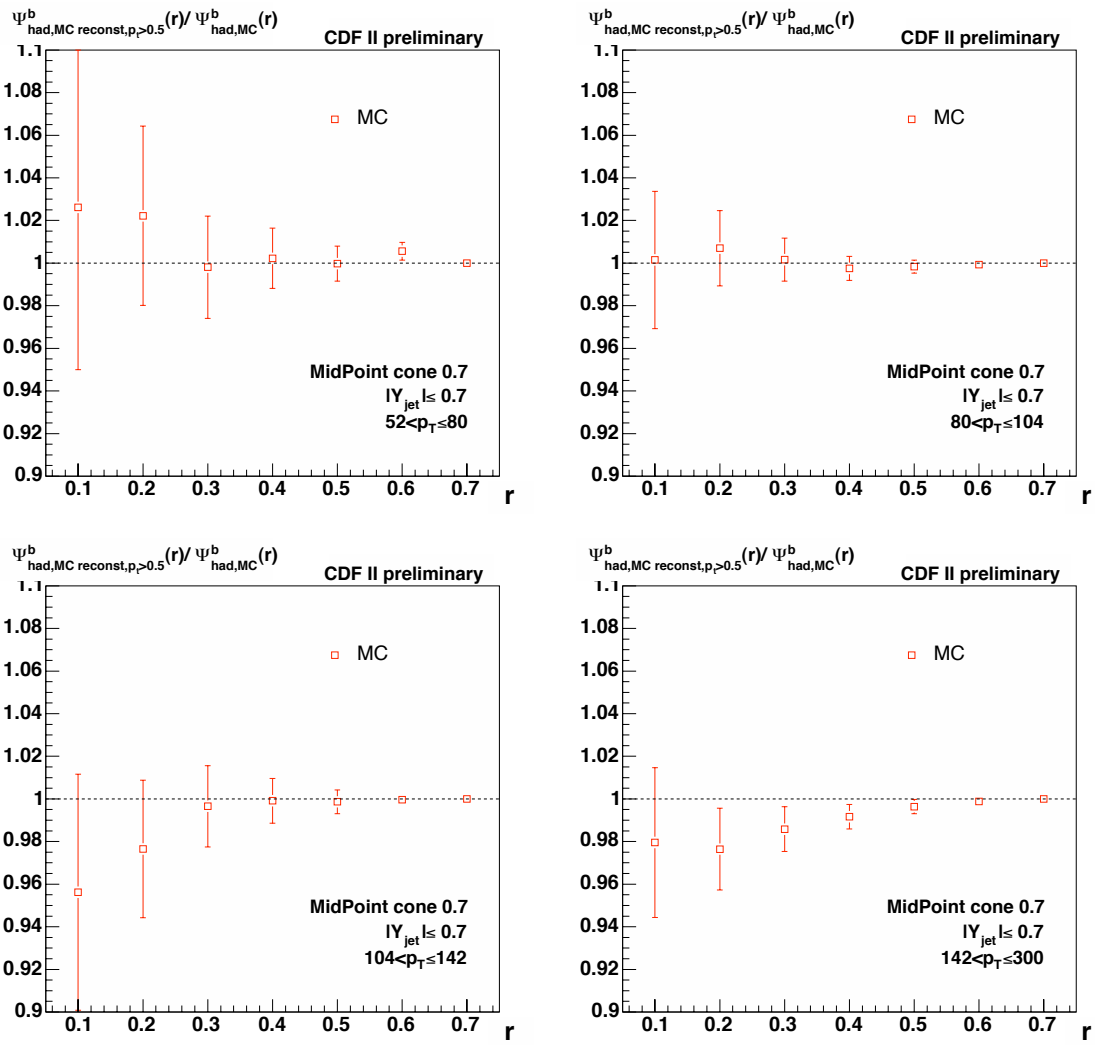

FIGURE 8.10: Ratio of the reconstructed hadron level Pythia Tune A MC shapes, starting from raw shapes reconstructed using calorimeter towers with $p_T > 0.5\,\mathrm{GeV}$, over the "true" hadron level Pythia Tune A MC shapes. The errors shown are only the statistical errors of the reconstructed shapes.

8.2 Integrated b-jet Shapes

The hadron level integrated b-jet shapes are shown in figure 8.11, where both the statistical and the systematic errors are shown. The statistical errors are very small with respect to the systematic error, smaller than the points on the plots. Also shown in this plot are the MC Pythia Tune A predictions for the inclusive and the b-jet shapes. There are significant deviations from the MC predictions of the b-jet shapes for all but the first p_T bin.

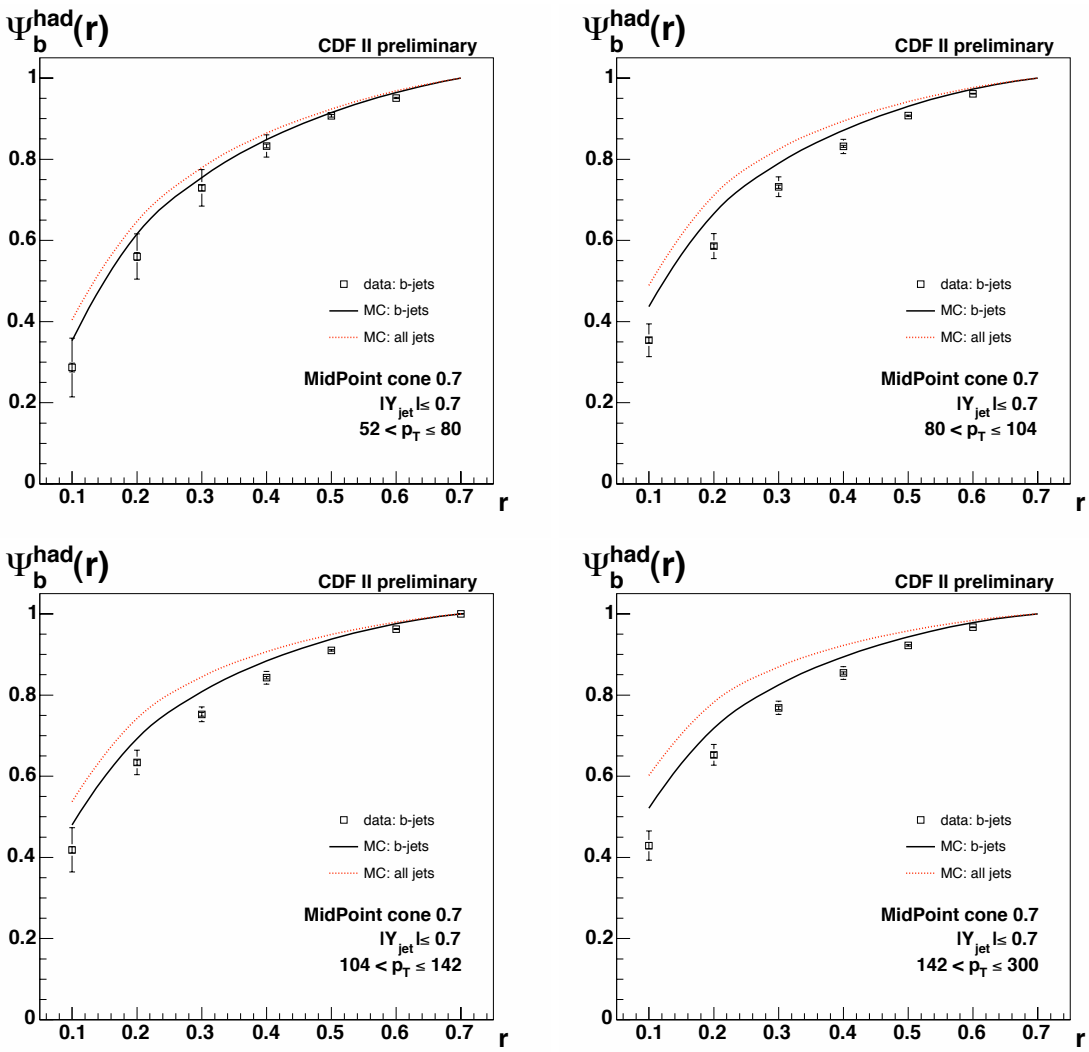

FIGURE 8.11: Hadron level integrated b-jet shapes for each of the p_T bins considered. The results are shown as black points where the error bars represent the total and statistical errors. The statistical errors are not visible as they are smaller than the points. The MC predictions for inclusive jets and for b-jets are shown as red and black curves, respectively.

Figure 8.12 shows the reconstructed hadron level b-quark jet shapes, where the inclusive jet shapes are taken from the Pythia Tune A MC predictions, whereas the tagged shapes are still measured in data. Only the statistical errors are shown. Figure 8.13 shows the ratio between the hadron level b-quark jet shapes reconstructed using the inclusive shapes from MC and the default hadron level b-quark jet shapes. Only the statistical errors of the numerator are shown on the points. The yellow bands show the total systematic errors of the hadron level b-quark jet shapes. This shows that the final results are somewhat sensitive to the small differences

in the measured inclusive jet shapes between the data and the Pythia Tune A MC. These differences are most of the time within the total systematic uncertainties on the final measurement.

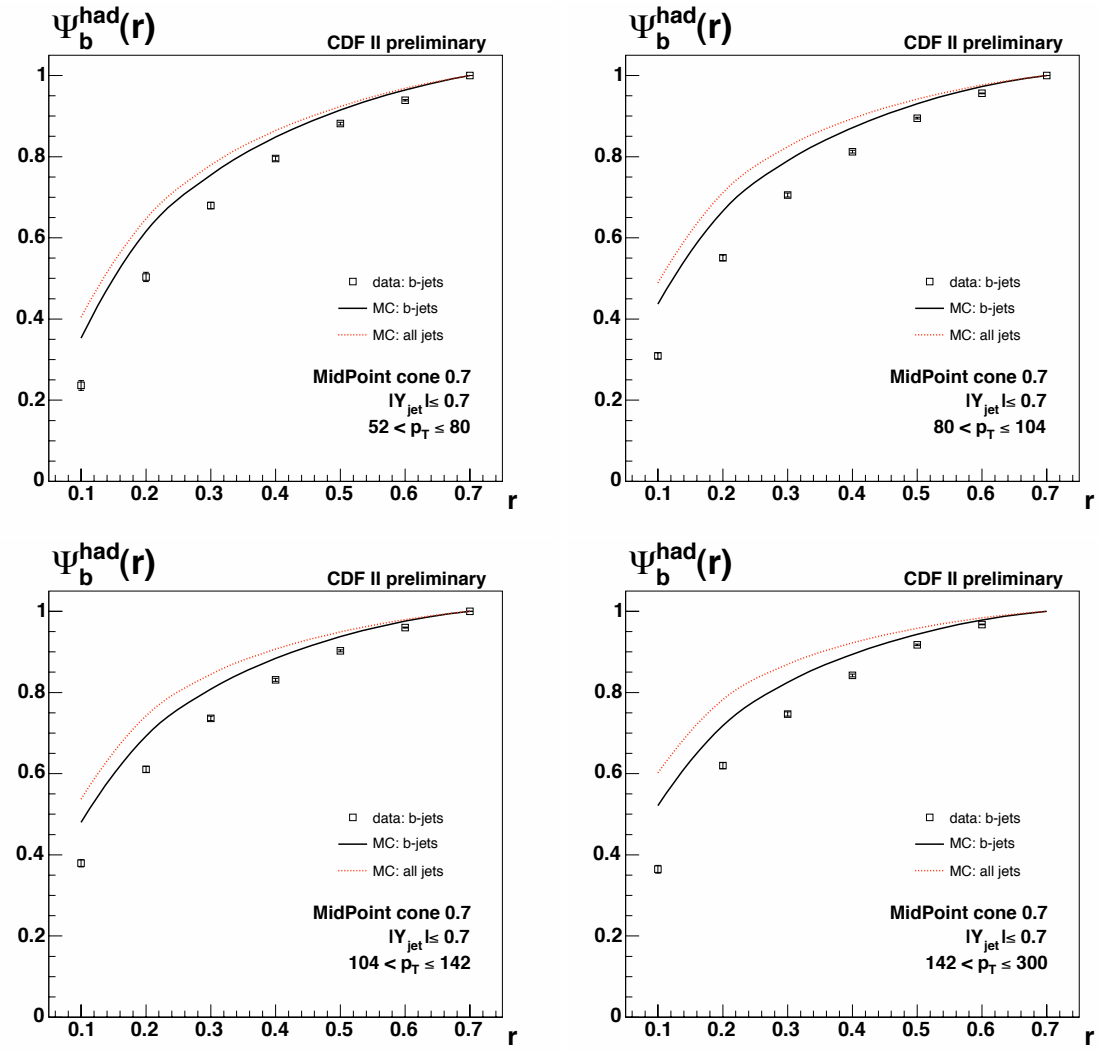

FIGURE 8.12: Hadron level integrated jet shapes in data but using the raw inclusive shapes obtained from Pythia Tune A MC. The results are shown as black points. Only the statistical errors are shown. The MC predictions for inclusive jets and for b-jets are shown as red and black curves, respectively.

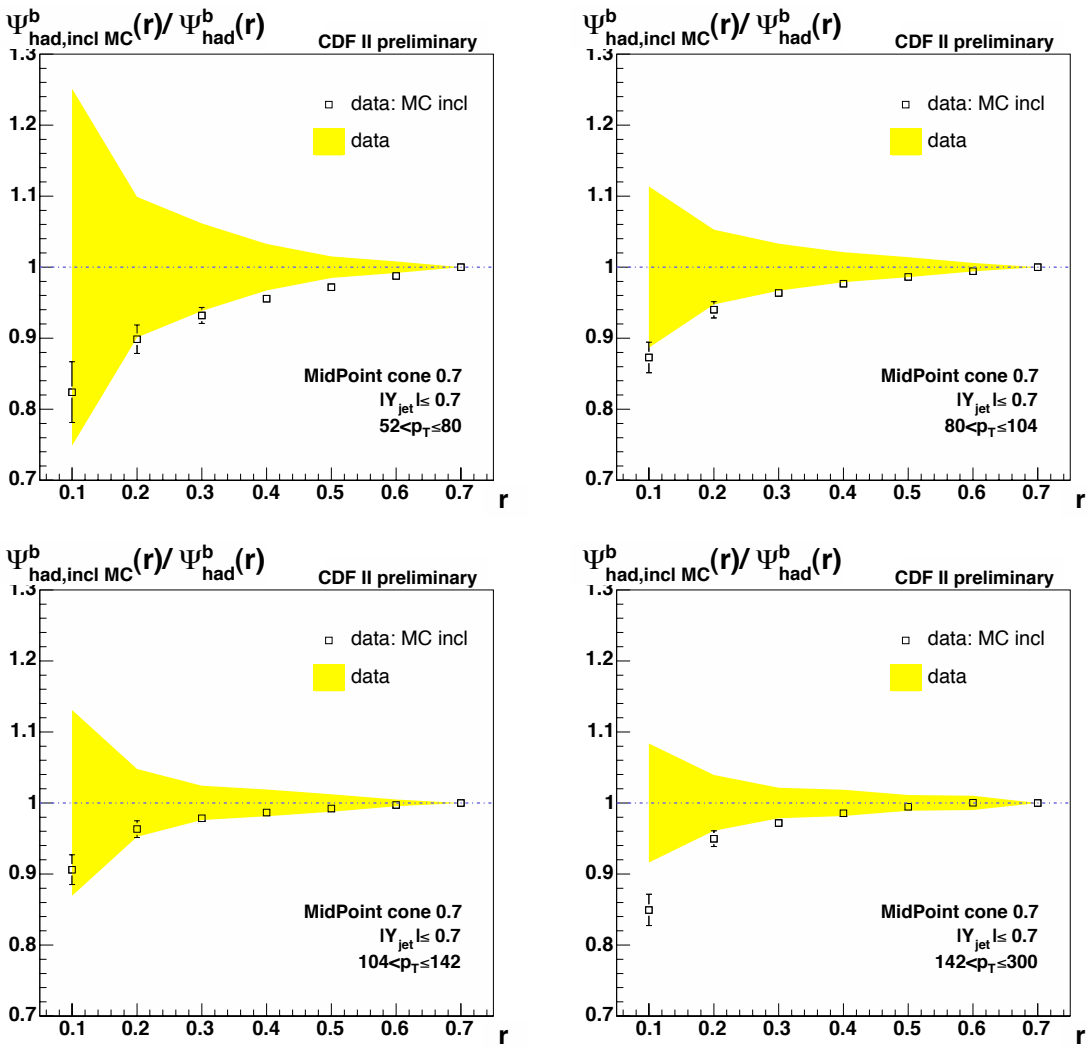

FIGURE 8.13: Ratio of the hadron level b-quark jet shapes reconstructed using the raw inclusive shapes taken from Pythia Tune A MC over those reconstructed from the raw inclusive shapes measured in data. The ratios are shown as black points, where the errors are the statistical errors on the numerator only. The yellow shaded areas represent the total errors on the hadron level b-quark jet shapes.

8.3 Variation with p_T

Another way of looking at these results is to plot the fractional p_T outside a cone of fixed radius r as a function of the p_T of the jets. This gives an idea of the change in width of the jets as the energies of the jets increase. Jets of a particular flavour are expected to become narrower as the p_T increases, mainly due to the running of the strong coupling constant, α_s. There is also a small effect due to the boost of the jets.

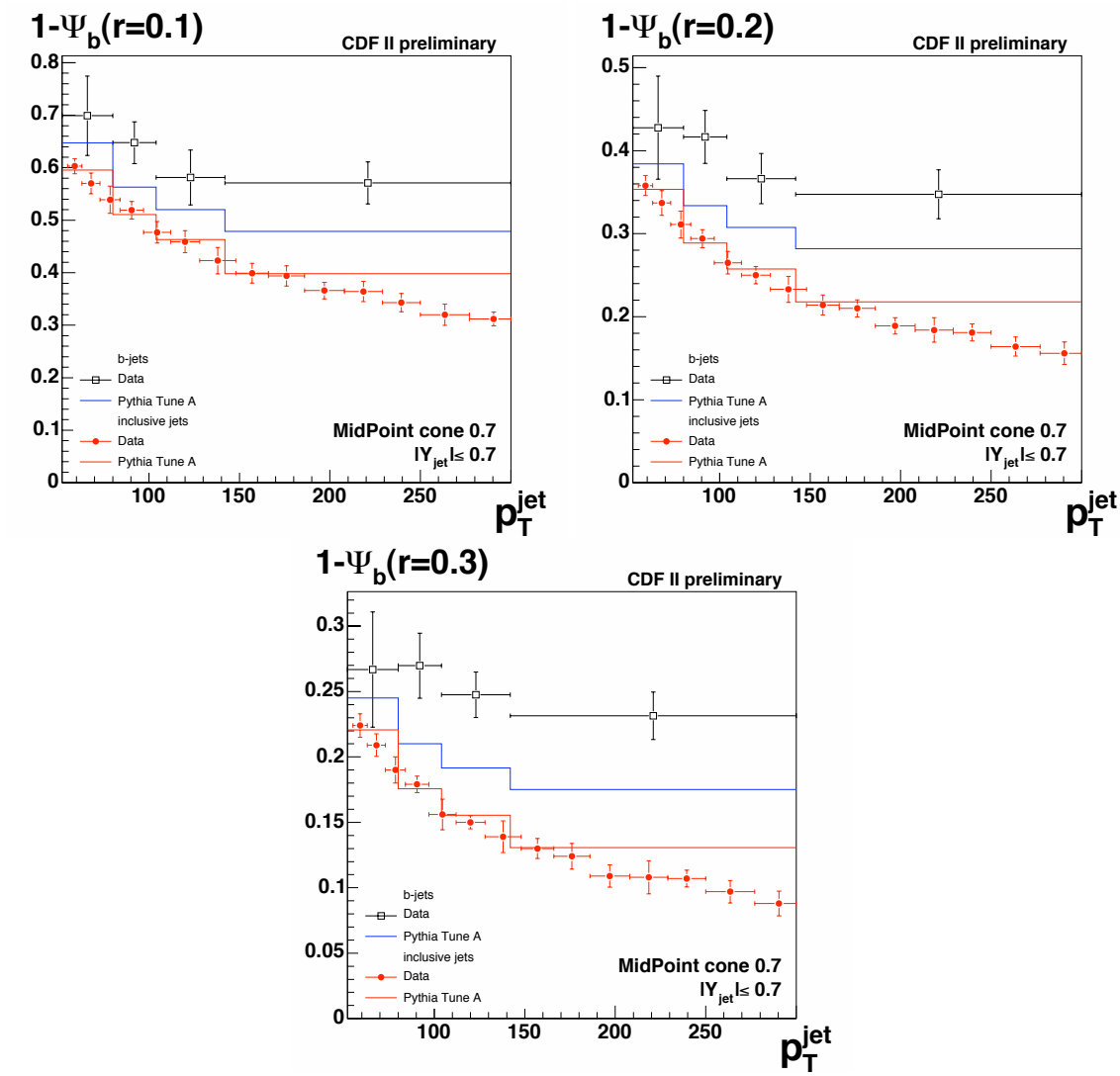

FIGURE 8.14: Fractional p_T outside a cone of radius $r = 0.1$, 0.2 and 0.3 around the jet axis as a function of the p_T of the jet. The results for b-quark jet shapes are shown as black points and the Pythia Tune A MC predictions are shown as blue lines. Also shown are the previously published inclusive jet shape results (red points) and the Pythia Tune A MC predictions (red lines) for inclusive jet shapes. The errors bars on the data points represent the total errors.

Figure 8.14 shows the evolution with jet p_T of the fractional p_T outside a cone of fixed radius (0.1, 0.2 and 0.3 are shown). Also shown on these plots are the predictions from the Pythia Tune A MC for both the b-jets and the inclusive jets. The previously published inclusive jet shapes results are shown on these plots as red points [15]. The latter agree very well with the MC predictions. The b-quark jet shapes measured in data do not agree with the Pythia Tune A MC predictions. The values shown represent the average value over the whole p_T range of the

bin. Because there are many more jets in the low p_T region of each bin, the average shapes for each bin are going to be dominated by the lower p_T jets. For this reason, when comparing the inclusive results and predictions, the two are expected to agree for the lower p_T regions of the bins.

8.4 Dependence on the Single b-quark Jet Fraction

The main reason why the Pythia Tune A MC predictions do not match the data very well is thought to be because the fraction of b-jets which come from gluon splitting is badly reproduced in Leading order MC. Figure 8.15 shows the same results as figure 8.14 but comparing the results to the Pythia Tune A MC predictions for b-quark jets as well as jets containing a single b-quark and jets with two b-quarks inside the same jet cone.

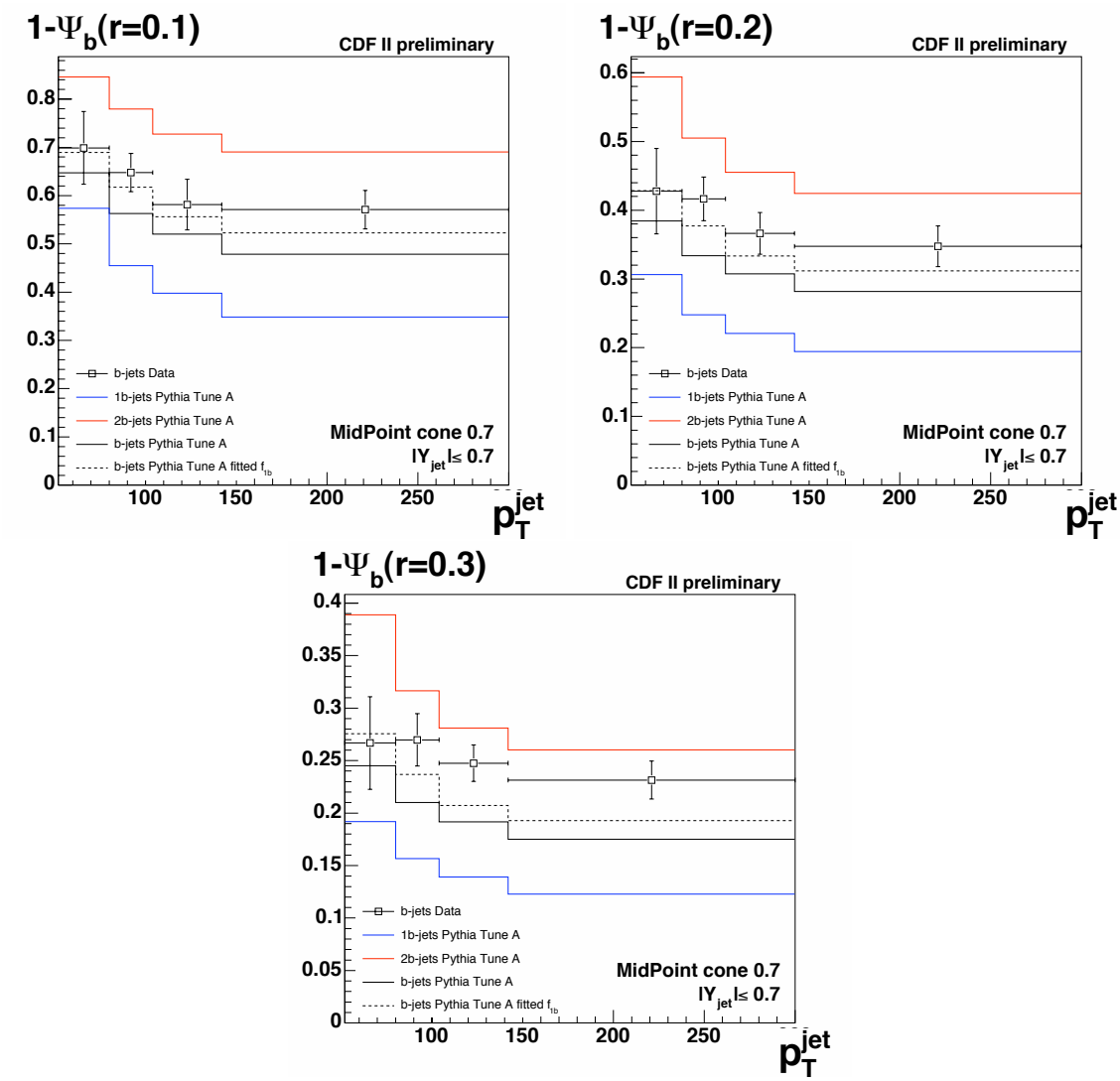

FIGURE 8.15: Fractional p_T outside a cone of radius $r = 0.1$, 0.2 and 0.3 around the jet axis as a function of the p_T of the jet. The results are shown as black points. The vertical error bars represent the total errors. The Pythia Tune A MC predictions for single and double b-quark jet shapes are shown as blue and red lines, respectively. Also shown are the Pythia Tune A MC b-quark jet shapes obtained from the fitted $f_{1\mathrm{b}}$ fractions (black dotted lines) and using the default $f_{1\mathrm{b}}$ fractions (black full lines).

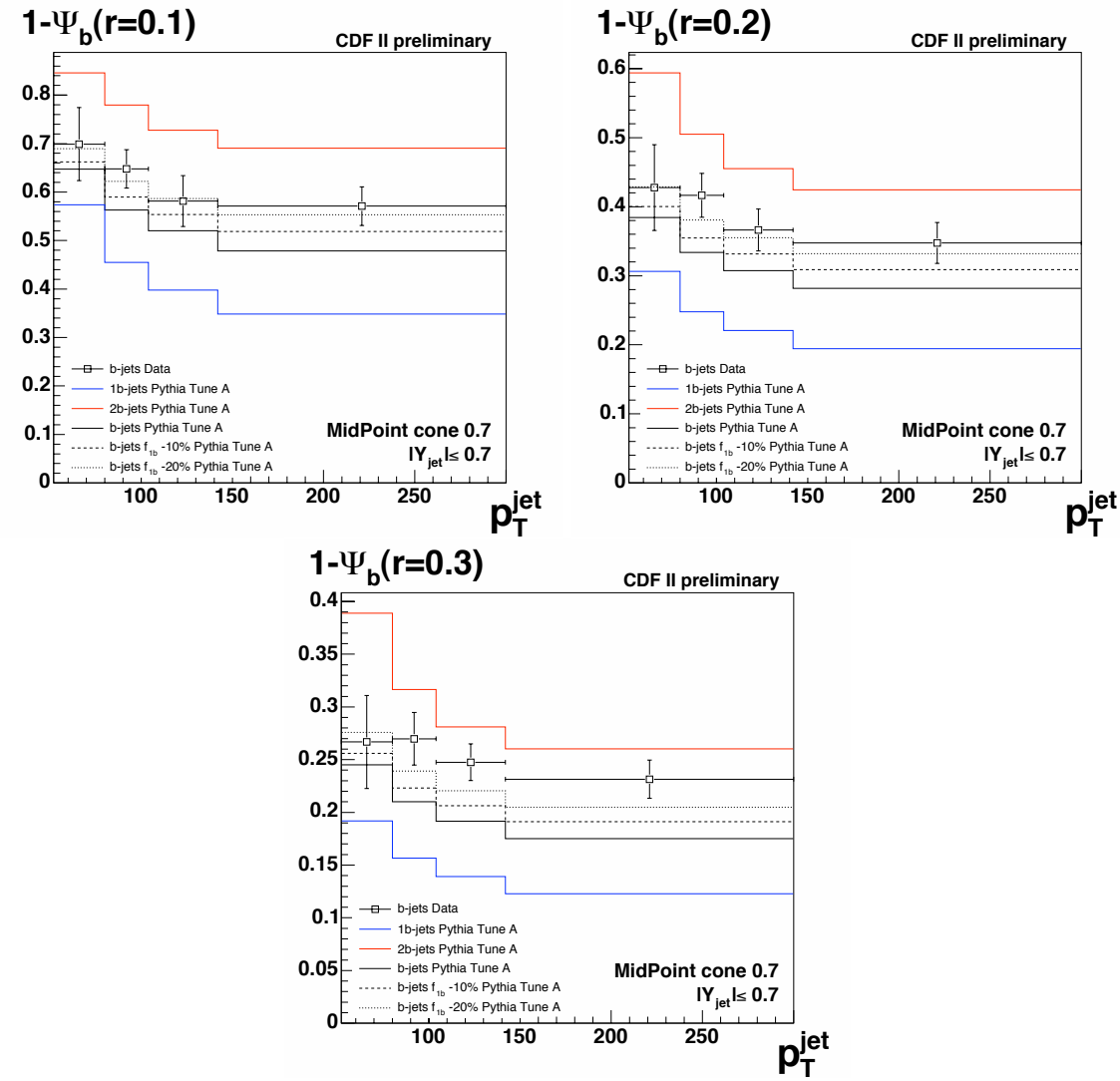

FIGURE 8.16: Fractional p_T outside a cone of radius $r = 0.1$, 0.2 and 0.3 around the jet axis as a function of the p_T of the jet. The results are shown as black points. The Pythia Tune A MC predictions for single and double b-quark jet shapes are shown as blue and red lines, respectively. Also shown are the Pythia Tune A MC predictions (black full lines) for b-quark jet shapes using the default $f_{1\mathrm{b}}$ fraction as well as using an $f_{1\mathrm{b}}$ fraction 10% and 20% lower, in absolute terms (black dashed and dotted lines, respectively).

p_T range	$f_{1\mathrm{b}}$	$f_{1\mathrm{b}}$ MC	$\Delta f_{1\mathrm{b}}$
52-80	0.575	0.774	-0.199
80-104	0.498	0.684	-0.186
104-142	0.519	0.627	-0.108
142-300	0.490	0.602	-0.112

TABLE 8.1: Fitted fraction of single b-quark jets in data, alongside the Pythia Tune A MC predictions and the difference between these two values.

It is possible to extract from data, for each p_T bin, the most probable $f_{1\mathrm{b}}$ fraction. This is done using the same binned χ^2 minimisation method as the one used for the SecVtx mass fit. The b-jet shapes obtained in data are fitted to the Pythia Tune A MC predictions for the single and double b-quark jets. The results obtained from the fit for each bin are shown in table 8.1. The fit errors are very large because of the large errors on the data and the fact that

the correlations between the bins have not been taken into account. For this reason it is not possible, at this stage, to state that these values truly constrain the f_{1b} fraction but they are an indication of the most probable value for this f_{1b} fraction. Also shown in that table are the f_{1b} fractions for each bin as predicted by Pythia Tune A and the difference between the predicted and fitted values. The difference is always within the 20% systematic uncertainty used for this analysis. It appears that the Pythia Tune A Leading Order MC systematically underestimates the gluon splitting rate in b-quark jet production. Figure 8.16 shows the variation with p_T of the fractional transverse momentum of the b-quark jets outside a cone of fixed radius compared to the values predicted by Pythia Tune A MC for b-quark jets, single b-quark jets, double b-quark jets as well as the predictions for b-quark jets if the f_{1b} fraction is decreased by 10% and 20% (absolute) with respect to the default Pythia Tune A MC values.

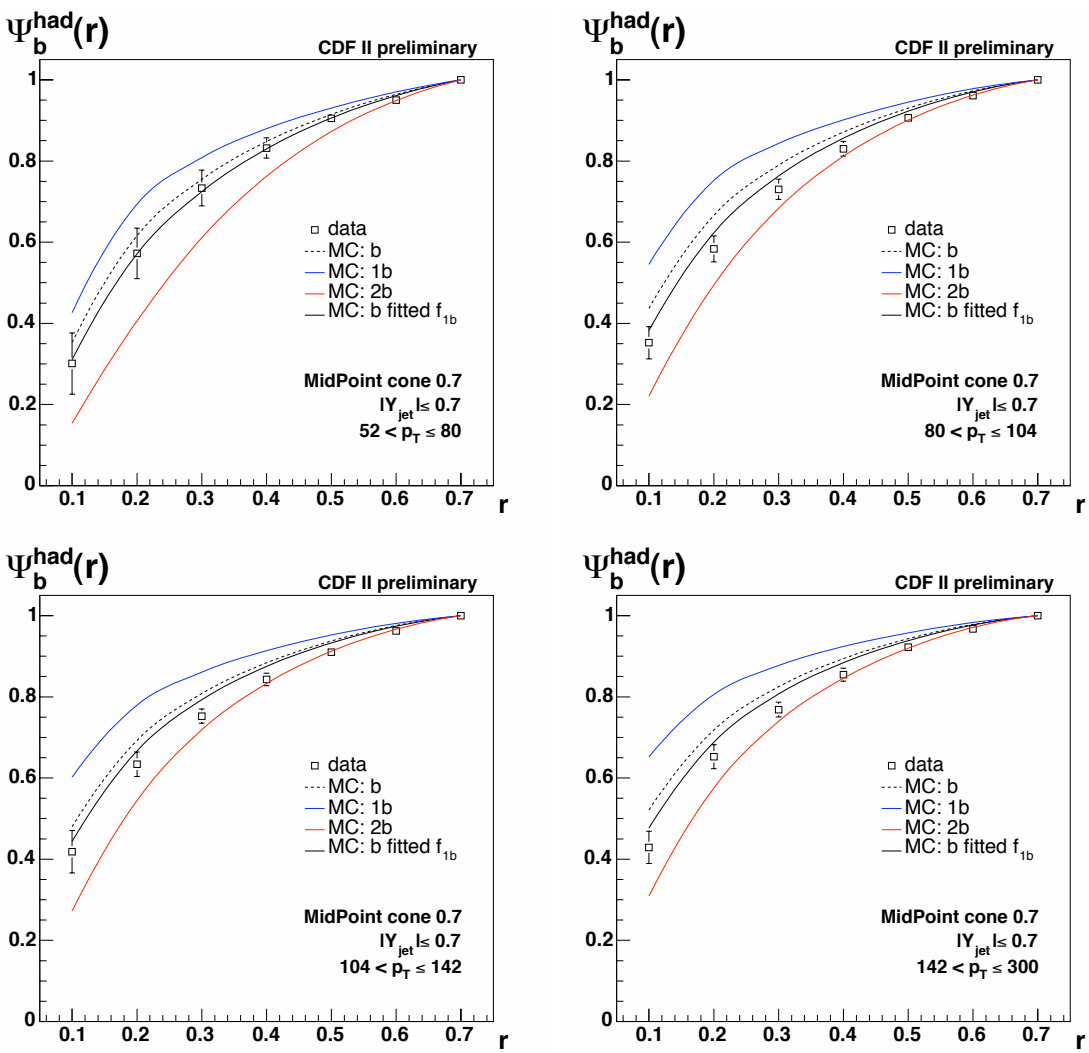

FIGURE 8.17: Hadron level integrated jet shapes for b-quark jets in data (black points). The total errors are shown. The Pythia Tune A MC predictions using the default and the fitted f_{1b} fractions are also shown (dotted and full black lines, respectively) along with the predictions for single and double b-quark jet shapes (red and blue lines, respectively).

Figure 8.17 shows a comparison of the hadron level integrated shapes of b-quark jets in data (black points) alongside the predictions from Pythia Tune A for b-quark jets, using the MC values for the f_{1b} fraction (dotted black line) and the fitted values for the f_{1b} fractions (full

black line). Also shown are the predictions for single and double b-quark jets (red and blue full lines, respectively). These plots show that the agreement between data and Pythia Tune A MC with the fitted f_{1b} fractions is better than with the default fraction. Figure 8.18 shows the ratio of the predicted Pythia Tune A MC jet shapes over the measured b-quark jet shapes. The total errors on the measured b-quark jet shapes are shown as yellow bands. This shows that the b-quark jet shapes in Pythia Tune A MC using the fitted f_{1b} fraction agree much better with the data than the default b-quark jet shape predictions.

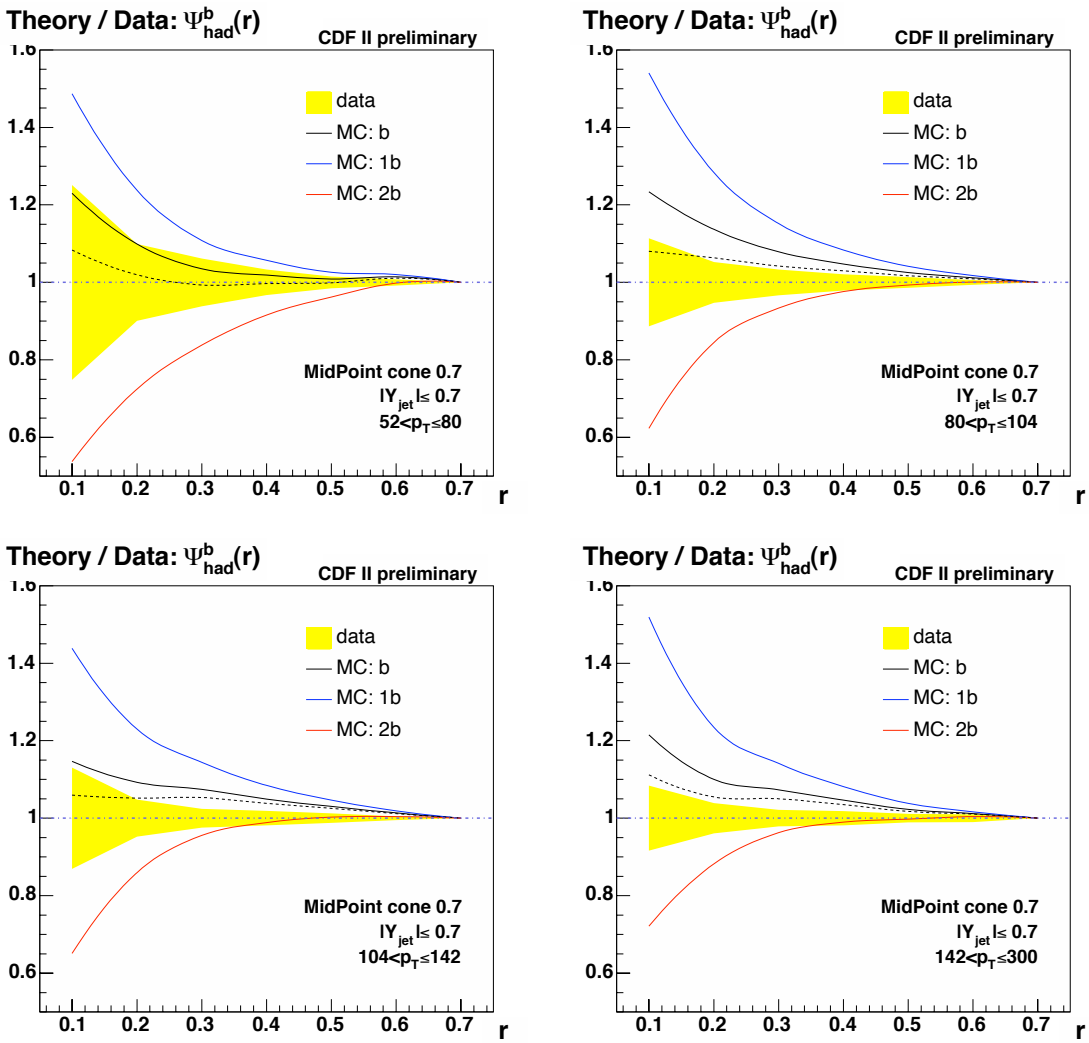

FIGURE 8.18: Ratio of the hadron level integrated jet shapes for b-quark jets for various theoretical predictions over the values obtained in data. The Pythia Tune A MC predictions using the default and the fitted f_{1b} fractions are shown (dotted and full black lines, respectively) along with the predictions for single and double b-quark jet shapes (red and blue lines, respectively). The yellow bands represent the total errors on the measured b-jet shapes.

Figure 8.19 shows the hadron level b-quark jet shapes in data compared to the Pythia Tune A predictions using the default values for the f_{1b} fractions (black lines). The Pythia Tune A predictions are also shown for the cases where the f_{1b} fraction is decreased by 10% and 20% (absolute) with respect to the default values (dotted and dashed black curves, respectively). This emphasises that the fraction of gluon splitting events is underestimated in Pythia Tune A.

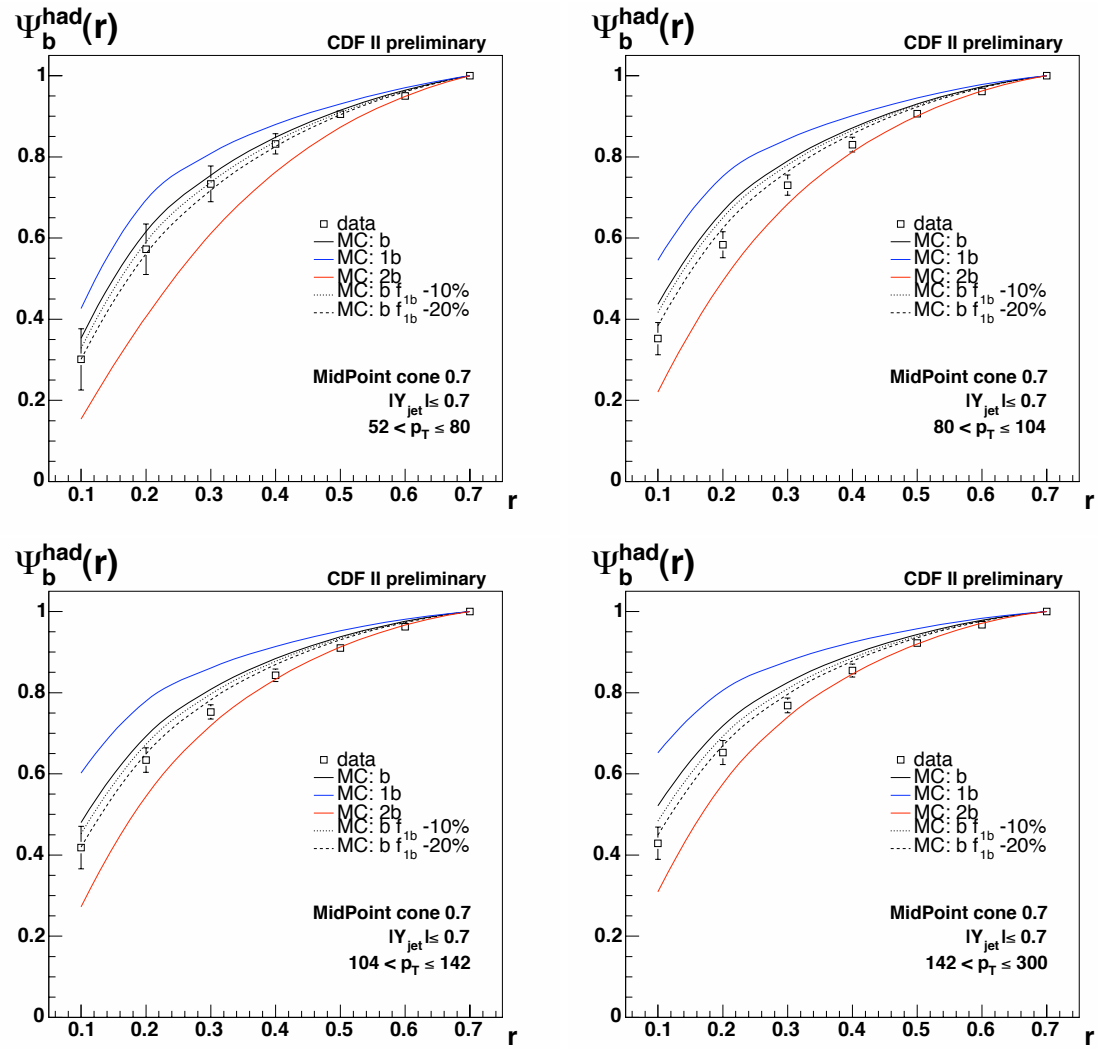

FIGURE 8.19: Hadron level integrated jet shapes for b-quark jets in data (black points). The total errors are shown. The Pythia Tune A MC predictions using the default f_{1b} fractions are shown as black lines. The Pythia Tune A MC predictions using f_{1b} fractions 10% and 20% (absolute) below the MC values are also shown (dotted and dashed black lines, respectively) along with the predictions for single and double b-quark jet shapes (red and blue lines, respectively).

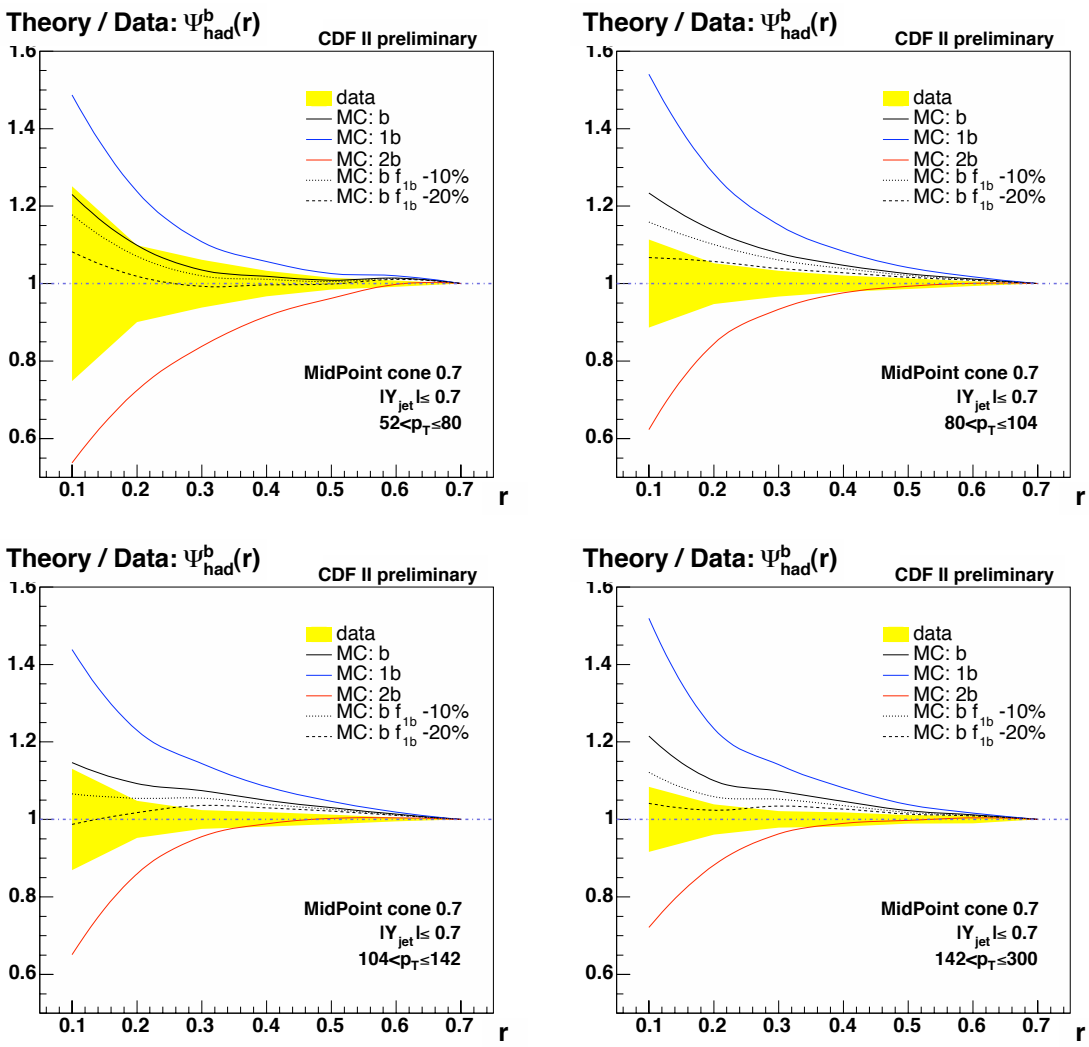

FIGURE 8.20: Ratio of the hadron level integrated jet shapes for b-quark jets for various theoretical predictions over the values obtained in data. The Pythia Tune A MC predictions using the default f_{1b} fractions are shown as black lines. The Pythia Tune A MC predictions using f_{1b} fractions 10% and 20% (absolute) below the MC values are also shown (dotted and dashed black lines, respectively) along with the predictions for single and double b-quark jet shapes (red and blue lines, respectively). The yellow bands represent the total errors on the measured b-jet shapes.

Chapter 9

Summary and Conclusions

For the first time at hadron colliders, the shapes of b-quark jets have been measured. This measurement uses approximately $300\,\mathrm{pb}^{-1}$ of data from the CDF detector. Because it is not possible to obtain a jet sample which contains only b-quark jets, a sophisticated unfolding method was needed in order to extract the b-quark jet shapes from the measured shapes for all jets and those measured for tagged jets. The tagging of the jets was achieved using the SecVtx tight tagger which increases the b-jet content from about 5% before any tagging requirement to $20 - 40\%$ after tagging, depending on the p_T of the jet. It was necessary to verify that this tagging algorithm could be applied not only to the default jet cone algorithm (JetClu with a cone size of 0.4) but also to other cone algorithms, in particular the one used in this analysis (MidPoint with a cone size of 0.7). It was found that the algorithm could be used in exactly the same way as for the default case, as long as the sub-cone inside which tracks used to reconstruct the secondary vertex can be found is kept at 0.4. Changing this sub-cone size to 0.7 significantly increases the tagging efficiency but the number of mistags also increases. A full re-optimisation of the algorithm would be needed before increasing this sub-cone to the full jet cone size of 0.7.

Other than the secondary vertex tagging, the analysis presented in this thesis involves the calculation of a number of different parameters before the b-quark shapes can be extracted statistically. These other parameters are the biases due to the tagging requirement, both on b- and nonb-jets, and also the hadron level corrections which correct the detector level jet shapes back to hadron level where they can more easily be compared to other experimental results or to theoretical models. The b-quark jet shape results show a disagreement with respect to the Pythia Tune A MC predictions. This is thought to be mainly due to the fact that the fraction of b-jets that contain one or two b-quarks is not correctly modelled in Pythia Tune A. The results, despite the relatively large systematic errors, can be used to extract the most probable value of this single b-quark jet fraction for each p_T bin considered. This fraction is found to be between 10% and 20%, in absolute terms, lower in data than in Pythia Tune A. The agreement between the data and the Pythia Tune A MC predictions using this new, fitted, fraction is better.

The systematic errors on the measurement are still relatively high. The statistical errors on the other hand are very small. The dominant sources of systematic errors are the fraction of b-quark jets which contain one or two b-quarks and similarly the fraction of c-quark jets, part of the background, which contain one or two c-quarks inside the jet cone. The difference between the hadron level shapes reconstructed from raw shapes using tracks and the default shapes, reconstructed using calorimeter towers, is also a dominant source of systematic error. Despite the fact that the Pythia Tune A MC statistics was limited, in particular for the calculation of the tagging biases on nonb-jets, generating more MC would not have significantly decreased the total systematic errors for any of the measurement points.

The measurement of b-quark jet shapes has been shown to be a good method for checking the MC modelling of the ratio between flavour creation and gluon splitting in b-quark jet production. The analysis methodology is somewhat sensitive to this ratio through the different parameters of the unfolding but this dependence is not very strong. It does not enter linearly into the unfolding equations and an uncertainty on the single b-quark jet fraction can be accounted for in the systematic error calculation. Despite this weak dependence on the single b-quark jet fraction, it is possible to obtain a relatively precise measurement of the b-quark jet shapes at CDF. This method should also be applicable to LHC studies, where b-jets will be involved in many searches for new particles. A good understanding of b-jets, in particular the relative amount of gluon splitting to flavour creation jets, is very important for such searches.

The b-quark jet shapes could also be used as parameters to apply cuts on in order to enhance or reduce the relative contributions of gluon splitting or flavour creation b-quark jets. For example, by requiring the fractional jet energy outside a fixed cone radius to be below a certain threshold, it is possible to enhance the flavour creation component of the b-quark jets.

A summary of the work presented in this thesis on the measurement of the b-quark jet shapes can be found in the CDF internal note number 8143. This note contains additional sections on comparisons of general jet related quantities as well as an appendix which compares the inclusive jet shape results using my version of the analysis code with the previously published inclusive jet shape results. I expect this analysis to be approved by the CDF collaboration in the very near future. These results would then become official CDF preliminary results.

Appendix A

Differential b-quark Jet Shapes

It is possible to write a similar equation to equation 6.8 but using the differential instead of the integrated shapes

$$\rho^{\mathrm{b}}_{\mathrm{had}}(r) = C_{\mathrm{had}}(r)\frac{\rho_{\mathrm{meas}}(r) - (1-p_{\mathrm{b}})b^{\mathrm{nonb}}(r)\rho^{\mathrm{nonb}}_{\mathrm{det}}(r)}{p_{\mathrm{b}}b^{\mathrm{b}}(r)} \tag{A.1}$$

This would be the simplest equation for the b-quark jet unfolding. The bias terms would be defined, in analogy to the integrated jet shape definition, as

$$b^{\mathrm{b}}(r) = \frac{\rho^{\mathrm{tag}}_{\mathrm{b\,MC}}(r)}{\rho^{\mathrm{incl}}_{\mathrm{b\,MC}}(r)} \tag{A.2}$$

for b-jets and

$$b^{\mathrm{nonb}}(r) = \frac{\rho^{\mathrm{tag}}_{\mathrm{nonb\,MC}}(r)}{\rho^{\mathrm{incl}}_{\mathrm{nonb\,MC}}(r)} \tag{A.3}$$

for nonb-jets. The hadron level corrections would become

$$C_{\mathrm{had}}(r) = \frac{\rho^{\mathrm{had}}_{\mathrm{MC}}(r)}{\rho^{\mathrm{det}}_{\mathrm{MC}}(r)}. \tag{A.4}$$

Combining all these terms into equation A.1 gives

$$\rho^{\mathrm{b}}_{\mathrm{had}}(r) = \frac{\rho^{\mathrm{had}}_{\mathrm{MC}}(r)}{\rho^{\mathrm{det}}_{\mathrm{MC}}(r)}\frac{\rho_{\mathrm{meas}}(r) - (1-p_{\mathrm{b}})\frac{\rho^{\mathrm{tag}}_{\mathrm{nonb\,MC}}(r)}{\rho^{\mathrm{incl}}_{\mathrm{nonb\,MC}}(r)}\rho^{\mathrm{nonb}}_{\mathrm{det}}(r)}{p_{\mathrm{b}}\frac{\rho^{\mathrm{tag}}_{\mathrm{b\,MC}}(r)}{\rho^{\mathrm{incl}}_{\mathrm{b\,MC}}(r)}} \tag{A.5}$$

It is apparent from this equation that replacing the measured quantities by the MC quantities, the equation unfolds correctly. The problem with the use of the differential shapes is that the normalisations are not easy to take into account because of the constraint on the integral of the jet shapes. This constraint must be true both for tagged shapes and the shapes corrected back to detector level, for both the data and the MC,

$$\int_0^R \rho^{\mathrm{b}}_{\mathrm{meas}}(r)dr = \int_0^R \rho^{\mathrm{b}}_{\mathrm{det}}(r)b^{\mathrm{b}}(r)dr = 1 \tag{A.6}$$

and

$$\int_0^R \rho^{\mathrm{b}}_{\mathrm{det}}(r)dr = 1 \tag{A.7}$$

It is not trivial to obtain bias terms which preserve this normalisation. In the numerator it would be possible to redefine the terms involving nonb-jets in order to maintain the normalisation. The $b^{\mathrm{nonb}}(r)$ terms could be replaced by

$$b^{\mathrm{nonb}}_{\mathrm{new}}(r) = \frac{b^{\mathrm{nonb}}(r)}{\int_0^R b^{\mathrm{nonb}}(r)\rho^{\mathrm{nonb}}_{\mathrm{det}}(r)dr}. \tag{A.8}$$

The normalisations would depend on the measured nonb-jet shapes (i.e. on the raw inclusive jet shapes using our approximation) but would not depend on the shapes we want to extract from this equation, the b-quark jet shapes.

The bias terms in the denominator, on the other hand, are where the problems lie. The tagging bias terms on b-quark jets must have additional renormalisations in order to obtain properly normalised b-quark jet shapes. The problem is that these renormalisations will depend on the detector level b-quark jet shapes which we want to measure. The $b^{\mathrm{b}}(r)$ terms should be replaced by something which would have the form

$$b^{\mathrm{b}}_{\mathrm{new}}(r) = \frac{b^{\mathrm{b}}(r)}{\int_0^R b^{\mathrm{b}}(r)\rho^{\mathrm{b}}_{\mathrm{det}}(r)dr}, \tag{A.9}$$

where $\rho^{\mathrm{b}}_{\mathrm{det}}(r)$ are the detector level b-quark jet shapes we want to measure. We would therefore end up with a dependence on the b-quark jet shapes on the left hand side of the unfolding equation but also on the right hand side of the equation. It is therefore impossible to obtain such a simple unfolding equation for the differential shape as the one used for the integrated jet shape, even requiring additional renormalisation of the shapes.

An alternative approach would be to obtain an unfolding equation for the differential b-jet shapes by differentiating the equation for the integrated shapes. This would lead to the correct normalisation of the jet shapes because the normalisation of the integrated shapes is taken care of in the integrated shape unfolding equation and will be maintained after differentiation. This equation turns out to be relatively complicated even for the detector level b-quark jet shapes (i.e. without considering the hadron level jet shape corrections)

$$\begin{aligned}\rho^{\mathrm{b}}_{\mathrm{det}}(r) = \frac{\partial \Psi^{\mathrm{b}}_{\mathrm{det}}(r)}{\partial r} &= \frac{1}{p_{\mathrm{b}} b^{\mathrm{b}}(r)}\Bigg[\frac{-\frac{\partial b^{\mathrm{b}}(r)}{\partial r}}{b_{\mathrm{b}}(r)}\Big(\Psi_{\mathrm{meas}}(r) - (1-p_{\mathrm{b}})b^{\mathrm{nonb}}(r)\Psi^{\mathrm{nonb}}_{\mathrm{det}}(r)\Big) + \rho_{\mathrm{meas}}(r) \\ &\quad -(1-p_{\mathrm{b}})b^{\mathrm{nonb}}(r)\rho^{\mathrm{nonb}}_{\mathrm{det}}(r) - (1-p_{\mathrm{b}})\frac{\partial b^{\mathrm{nonb}}(r)}{\partial r}\Psi^{\mathrm{nonb}}_{\mathrm{det}}(r)\Bigg],\end{aligned} \tag{A.10}$$

where

$$\frac{\partial b^{\mathrm{b}}(r)}{\partial r} = \frac{\rho^{\mathrm{tag}}_{\mathrm{b\,MC}}(r)}{\Psi^{\mathrm{incl}}_{\mathrm{b,MC}}(r)} - b^{\mathrm{b}}(r)\frac{\rho^{\mathrm{incl}}_{\mathrm{b\,MC}}(r)}{\Psi^{\mathrm{incl}}_{\mathrm{b\,MC}}(r)} \tag{A.11}$$

and

$$\frac{\partial b^{\mathrm{nonb}}(r)}{\partial r} = \frac{\rho^{\mathrm{tag}}_{\mathrm{nonb\,MC}}(r)}{\Psi^{\mathrm{incl}}_{\mathrm{nonb,MC}}(r)} - b^{\mathrm{nonb}}(r)\frac{\rho^{\mathrm{incl}}_{\mathrm{nonb\,MC}}(r)}{\Psi^{\mathrm{incl}}_{\mathrm{nonb\,MC}}(r)}. \tag{A.12}$$

This equation involves both differential and integrated terms for all the r dependent parameters of the unfolding equation and for the measured shapes. Moreover, the error calculation would be extremely complex. It was therefore deemed not feasible to carry out this analysis using the differential jet shapes.

For the integrated jet shape, these problems are not present because the bias terms and hadron level corrections tend to 1 for $r \to R$ without any additional requirements. Also, because of the nature of the unfolding equation, the integrated b-jet shapes tend to 1 as $r \to R$.

Appendix B

CMS ECAL Detector Control System

In order to get data of good enough quality for Physics analyses, the particle detection must be very well understood and the detectors must be kept under very stable conditions for extended periods of time. It is therefore necessary to continuously monitor the status of the detectors. This task is usually done by the Detector Control Systems (DCS). During my first year as a PhD student I was involved in the development of such a system for the electromagnetic calorimeter (ECAL) of the Compact Muon Solenoid (CMS) experiment.

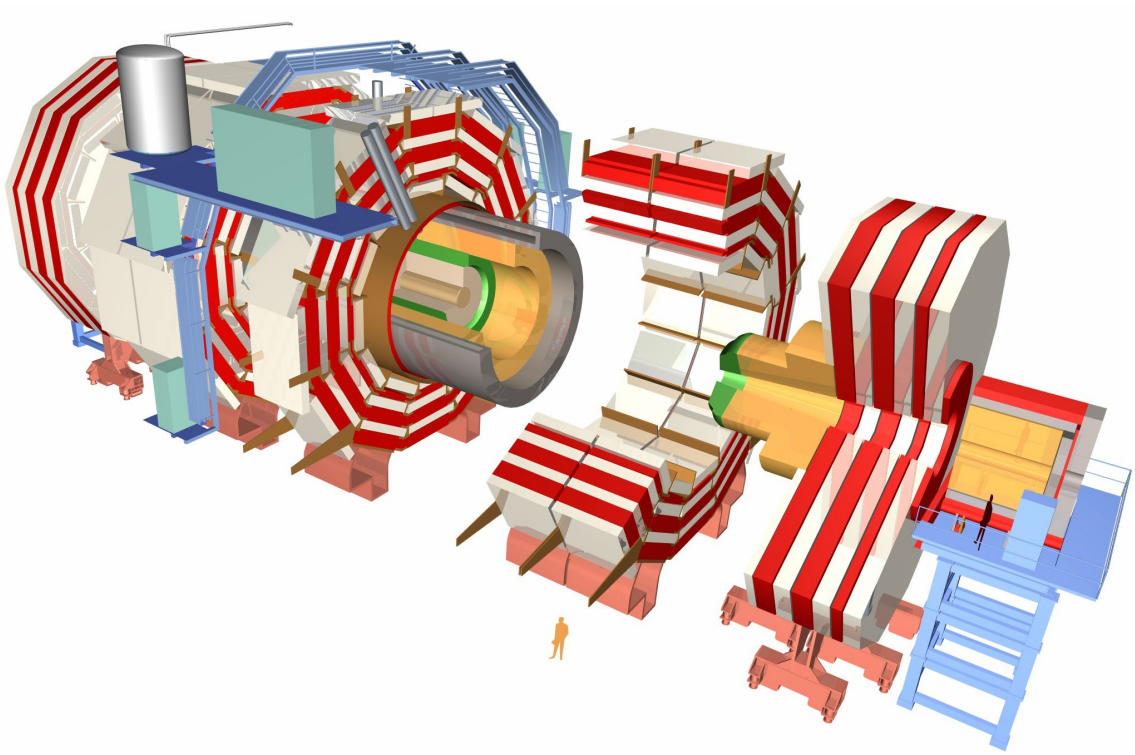

FIGURE B.1: Schematic drawing of the CMS detector. From inside to outside, the different detectors are: the silicon tracking (grey), the ECAL (green), the hadron calorimeter (HCAL) (yellow), the solenoid and it's housing (grey), the muon chambers (white) and the magnet return yoke (red).

At CERN a new machine, the Large Hadron Collider (LHC), is being built along with four experiments placed at the beam collision points. One of these experiments is the Compact

Muon Solenoid - CMS. This is a multi-purpose detector, a schematic picture of which is shown in figure B.1. More details on the CMS experiment can be found in reference [37]. One of the main Physics goals of CMS is the discovery the the Higgs particle. In the low Higgs boson mass range, the decay $H \rightarrow \gamma\gamma$ is a potential discovery channel. This decay has a very clean signature but has a very low cross section and requires an excellent calorimetric measurement of the photon energy as well as a good spatial resolution. This is achieved by the electromagnetic calorimeter - ECAL, one of the key detectors at CMS. The ECAL detector is made up of 76′000 lead tungstate ($PbW0_4$) crystals whose purpose is to measure the energy of most electromagnetic particles (e^+, e^- and γ in particular) by measuring the amount of scintillation photons produced by the particles as they go through the material and converting the number of photons back to the energy deposited in the crystal [38]. A schematic drawing of the ECAL is shown in figure B.2. The detector, which covers a pseudo-rapidity region up to $|\eta| < 3$, is split up into two half-barrels each containing 16 supermodules and two endcaps each split into two D-shaped substructures. Each supermodule has the same number of crystals. All the data acquisition (DAQ) and detector control is done independently for each supermodule.

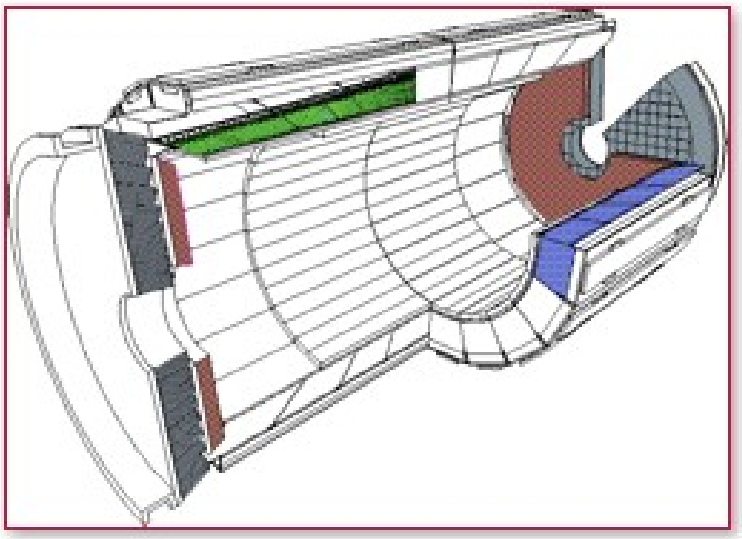

FIGURE B.2: Schematic view of the ECAL detector showing the barrel supermodules and the end-cap D-shapes sections.

The DCS system for the ECAL is discussed below with particular emphasis on the software control of this system [39] and the precision temperature monitoring system. An overview of the whole ECAL DCS system is shown in figure B.3.

The crystal light yield (basically the number of photons produced for each unit of energy deposited in the crystals) is very sensitive to temperature, a variation of 1°C changes the light yield by 2%. For this reason it is necessary to keep the detector at a very constant temperature with variations smaller than 0.1°C at 18°C. This requires not only a very good self-regulating cooling system to keep the temperature as stable as possible, but also a precise absolute measurement of this temperature to check the stability and to determine calibration periods over which it can be assumed that the temperature of the crystals is constant. For this, very precise thermistors (temperature dependent resistors) are used which must be individually calibrated to a precision significantly better than that provided by the manufacturer. The calibration is done using a thermal bath. The temperature of the bath is varied and the variation of resistance in the thermistors is measured. The absolute temperature value of the bath is given by a very well

calibrated thermistor, calibrated by the company. All the information about the calibration of the individual sensors is saved in a database which can later be read back by the control software. It should be noted that all sensors used in the ECAL DCS must be radiation hard to survive for the entire duration of the LHC running as in most cases there will be no access to these sensors once the supermodules are assembled.

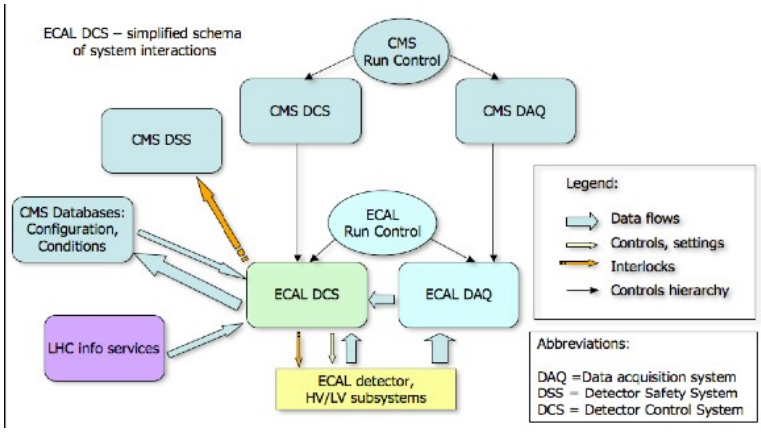

FIGURE B.3: Schematic view of the control chain at CMS.

The ECAL DCS controls not only the precision temperature monitoring (PTM) mentioned above but also a number of other parameters of the detector. These are: the high voltage (HV), low voltage (LV), cooling, humidity monitoring (HM) and safety system (ESS). These applications need to be running basically 24h a day, 7 days a week, for most of the year. The safety system needs to be particularly reliable as it produces hardwired interlocks to the HV and LV if there is a problem in the detector. The ESS is controlled by a set of programmable logic controllers (PLC) and only displays alarms and statuses to the rest of the system. The situation is similar for the cooling application. The DCS has the task of not only monitoring the HV and LV used by the electronics but also must ramp these up and down to predefined values depending on the needs of the experiment.

The software used to control all DCS systems at the LHC is based on a commercial SCADA product called PVSS II [40]. On top of this a framework has been developed by CERN which provides a unified set of tools usable by all the LHC experiments [41].

There are two different ways of looking at the DCS software system. The first is the hardware view which splits the detector up into each of the 36 barrel supermodules and two endcaps, segmented into eight quadrants. Each supermodule view contains all the individual sub-systems which can be controlled individually. This view is particularly important for spotting localised hardware problems and for testing and calibrating individual supermodules.

The more relevant view for controlling the whole system is the logical view. This view is based on the idea of Finite State Machines (FSM) whereby controls are passed from a supervisory system to all the relevant supermodules and sub-systems. An example of this would be a button which would turn on or off all the high voltage channels on all the supermodules, extracting from a database the values at which each channel needs to be set.

An important part of any control system is the logging of data. It is important to log all the data of the different DCS systems in order to improve the different systems or for understanding

problems as they occur. This writing to databases of the DCS data is done both in an internal PVSS log as well as to an Oracle database maintained by CERN. This system was tested successfully during the summer 2003 ECAL test-beam and prompted the PVSS developers to provide the possibility of automatically writing to Oracle databases.

The first full-scale test of any DCS system at CMS was carried out during the summer 2003 ECAL test-beam. One module (about 1/4 of a supermodule) was fully equipped with electronics and the DCS was fully controlled by PVSS II 2.12 applications with the exception of the supervisory system. The following year all the applications were updated to the latest version of PVSS II (3.0) and a supervisory system was able to control all the applications. The applications were all written in such a way as to be easily scalable to integrate all of the ECAL supermodules.

Figure B.4 shows a few examples of the front end view of the applications used in the 2004 test-beam. The top row shows the supervisory level generalised for the entire ECAL detectors. On the lower row are shown the front end view of the applications for the different sub-systems.

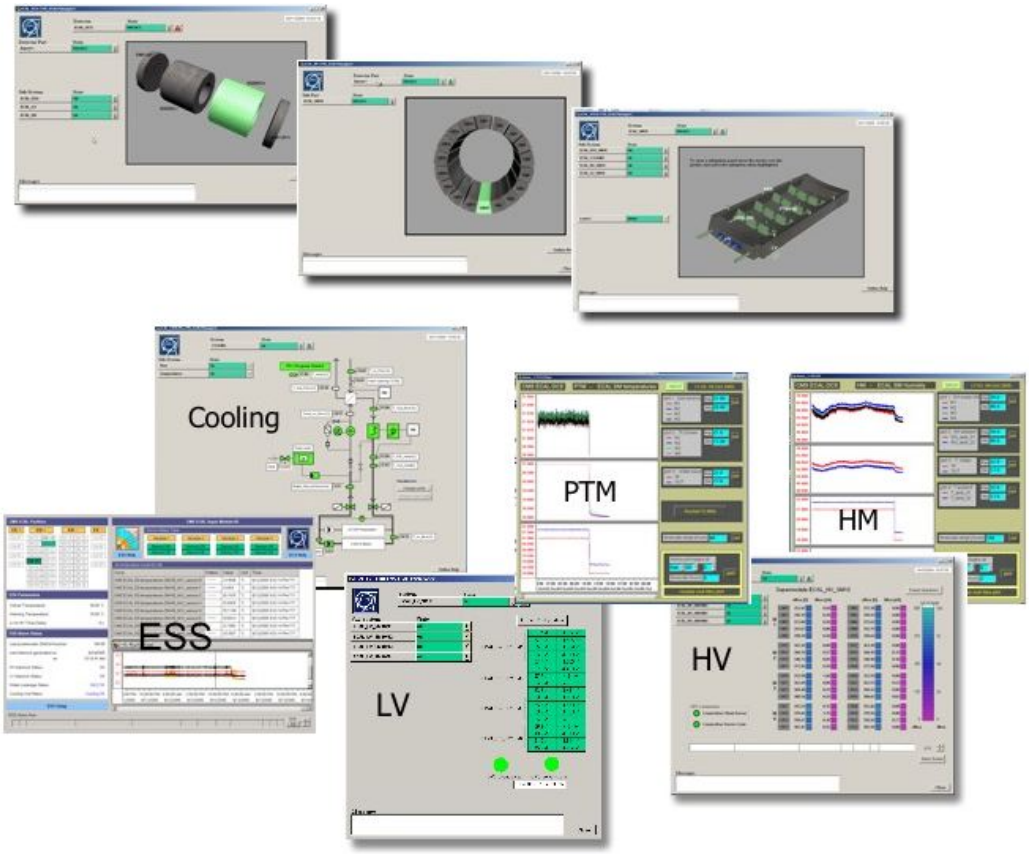

FIGURE B.4: Schematic view of the user interface panels of the PVSS applications for ECAL. The top row shows the physical view of the detector and the bottom shows the different sub-systems for each supermodule

My contribution to the ECAL DCS system was mainly the PVSS II 2.12 software development for the 2003 test-beam as well as the calibration of the precision temperature sensors. I wrote the applications to control the cooling, HV, LV and additional temperature sensors used for the temperature mapping specific to that test-beam. The systems were written in such a way to be easily scalable for later use in the final systems. I was also involved in the general de-bugging of PVSS II 2.12 and the development of the JCOP framework. This led to many

improvements in the software which were then implemented in patches or in PVSS II 3.

Bibliography

[1] A. Lister. Measurement of b-quark Jet Shapes at CDF. *CDF Note* 7973, 2005.

[2] A. Lister. Measurement of b-quark Jet Shapes at CDF: Inclusive Method. *CDF Note* 8143, 2006.

[3] D.H. Perkins. *Introduction to High Energy Physics, 4th edition.* Cambridge University Press, 2000.

[4] The figures are taken from http://particleadventure.org.

[5] F. Abe *et al* [CDF Collaboration]. Evidence for Top Quark Production in anti-p p Collisions at $\sqrt{s} = 1.8$ TeV, hep-ex/9405005. *Phys. Rev. Lett.*, 73(225), 1994.

[6] P. Higgs. Broken Symmetries, Massless Particles and Gauge Fields. *Phys. Lett.*, 12(132), 1964. Broken Symmetries and the Masses of Gauge Bosons *Phys. Rev. Lett.*, 13(508), 1964.

[7] S. Martin. A Supersymmetry Primer, hep-ph/9709356 v3. 1999.

[8] R.K. Ellis, W.J. Stirling and B.R. Webber. *QCD and Collider Physics.* Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology, 1996.

[9] G. Dissertori, I. Knowles and M. Schmelling. *Quantum Chromodynamics: High Energy Experiments and Theory.* Oxford Science Publications, 2003.

[10] G. Altarelli and G. Parisi. Aymptotic Freedom in Parton Language. *Nucl. Phys B*, 126(298), 1977.

[11] T. Sjöstrand *et al.* Pythia 6.2: Physics and Manual. hep-ph/0108264.

[12] B. Anderson. *The Lund Model.* Cambridge University Press, 1988.

[13] D. Acosta *et al* [CDF Collaboration]. Momentum Distribution of Charged Particles in Jets in Dijet Events in p anti-p Collisions at $\sqrt{s} = 1.8$ Tev and Comparisons to Perturbative QCD Predictions. *Phys. Rev. D*, 68(012003), 2003.

[14] R. D. Field. The Underlying Event in Hard Scattering Processes. hep-ph/0201192.

[15] D. Acosta *et al* [CDF Collaboration]. Study of Jet Shapes in Inclusive Jet Production in p anti-p Collisions at $\sqrt{s} = 1.96$ Tev. *Phys. Rev. D*, 71(112002), 2005.

[16] S. Frixione *et al.* Heavy Quark Production. *CERN-TH/97-16*, 1997. hep-ph/9702287.

[17] Tevatron Rookie Book, http://www-bdnew.fnal.gov/operations/rookie_books/Tevatron_PDF/TeV_Rookie_Book.htm.

[18] J. Tseng [CDF Collaboration]. Update of Run 1b Results at CDF. *FERMILAB-CONF-02/169-E*, 2002.

[19] The CDF II collaboration. The CDF II Detector Technical Design Report. *FERMILAB-PUB-96/390-E*, 1996. http://www-cdf.fnal.gov/upgrades/tdr/tdr.html.

[20] A. Sill *et al.* CDF Run II Silicon Tracking Projects. *Nucl. Instrum. Meth. A*, 447(1), 2000.

[21] J.E. Huth *et al.* Toward a Standardization of Jet Definitions. *FERMILAB-CONF-90-249-E*, 1990.

[22] S. D. Ellis and D.E. Soper. Successive Combination Jet Algorithm for Hadron Collisions. *Phys. Rev. D*, 48(3160), 1993.

[23] Catani, Dokshitzer, Seymour and Webber. Longitudinally Invariant k_{T} Clustering Algorithms for Hadron Hadron Colliders. *Nucl. Phys B*, 406(187), 1993.

[24] S.D. Ellis, J. Huston and M. Tönnesmann. On Building Better Cone Jet Algorithms. hep-ph/0111434.

[25] G.C. Blazey *et al.* RunII Jet Physics: Proceedings of the RunII QCD and Weak Boson Physics Workshop. hep-ex/0005012.

[26] C. New *et al.* SecVtx Optimization Studies for 5.3.3. Analyses. *CDF Note* 7578, 2005.

[27] P. Aurenche *et al.* The QCD and Standard Model Working Group: Summary Report from Les Houches. hep-ph/0005114.

[28] M. D'Onofrio, A. Lister. SecVtx Using the MidPoint Cone Algorithm: Comparison to JetClu. *CDF Note* 7045, 2005.

[29] M. D'Onofrio. Preliminary Results on Inclusive b-jet Cross Section. *CDF Note* 6984, 2004.

[30] H.L. Lai *et al* [CTEQ Collaboration]. Global QCD Analysis of Parton Structure of the Nucleon: CTEQ5 Parton Distributions. *Eur. Phys. J*, C. 12(375), 2000. hep-ph/990328.

[31] G. Corcella *et al.* Herwig 6: An Event Generator for Hadron Emission Reactions with Interfering Gluons (Including Supersymmetric Processes). *JHEP*, 0101(010), 2001. HERWIG 6.5. hep-ph/0011363.

[32] H.L. Lai *et al* [CTEQ Collaboration]. CTEQ6 Parton Distributions With Heavy Quark Mass Effects. *Phys.Rev.D*, 69(114005), 2004. http://user.pa.msu.edu/wkt/cteq/cteq6/cteq6pdf.html.

[33] C. Peterson *et al.* Scaling Violation in Inclusive $e^+ e^-$ Annihilation Spectra. *Phys. Rev. D*, 27(105), 1983.

[34] The CDF Collaboration. J/ψ, $\psi' \to \mu^+\mu^-$ and $B \to J/\psi$, ψ' Cross Sections. *FERMILAB-CONF-94-136-E*, 1994. hep-ex/9412013.

[35] M. D'Onofrio. Updated Results and Preliminary Comparison with Next to Leading Order Predictions: ADDENDUM to b-jet Cross Section Measurement. *CDF Note* 7806, 2005.

[36] J. Adelman *et al.* Generic Jet Scale Corrections for Run II. *CDF Note* 7358, 2005.

[37] The CMS collaboration. The Compact Muon Solenoid Technical Proposal. *CERN/LHCC 94-38*, 1994.

[38] The CMS collaboration. ECAL CMS TDR 4. *CERN/LHCC 97-33*, 1997.

[39] P. Adzic *et al.* The Detector Control System for the Electromagnetic Calorimeter of the CMS Experiment at the LHC. *CERN-CMS-CR-2005-028*, 2005.

[40] ETM. http://www.etm.au.

[41] JCOP. http://itco.web.cern.ch/itco/Projects-Services/JCOP.

Acknowledgements

I would like to thank my professor and supervisor Günther Dissertori for his unconditional support during my thesis. I learnt a lot during my thesis and I hope I have acquired the necessary tools to launch me further into the world of particle physics. I would like to thank Thomas Gehrmann for accepting to be my co-exmainer.

I am grateful to Felicitas Pauss for leading such a dynamic group at the IPP and for allowing us to have such good conditions in which to carry out our projects. I would like to thank all the members of the ETHZ who contributed to my experience as a PhD student. In particular I would like to thank our secretary Gabriele Kogler without whom I would have been lost many times with various administrative problems. I would like to thank Michael Dröge, Christian Haller, Dieter Schinzel and the rest of the "lunch crew" for providing endless entertainment and interesting discussions. Let us not forget the people who shared the ups and downs of my daily office life: Radek Ofierzynski (author of the Frustratiometer), Jan Ehlers (go Roma or Panera?), Giovanna Davatz and Fabian Stöckli as well as Anne-Sylvie Giolo.

I would like to thank Allan Clark and the other members of the University of Geneva CDF group for allowing me to carry out my thesis analysis on CDF and for their support during my thesis.

I would like to thank everyone at CDF who contributes or contributed to the smooth running of the experiment, from the technicians to the spokespeople. I would especially like to thank Monica D'Onofrio, Ken Hatakeyama and Daniel Jeans for the numerous discussions and their support and without whom this analysis would not have been possible.

Finally, I would like to thank my family and friends both in Switzerland and in US who contributed to making my life outside of work interesting and fulfilling: from the skiing to the climbing without forgetting some parties and much more.