An Algorithm to Measure Parton Fragmentation at Large Hadron Colliders

Tom Macey

Supervisor: Professor G. Thompson

Submitted August 2013

Accepted January 2014

Submitted in partial fulfillment of the requirements of the Degree of Doctor of Philosophy

Declaration

The work presented in this thesis is my own. Wherever contributions taken from other sources are involved, they are identified as such. The material produced has not previously been presented in identical or similar form, either in whole or in part, to any other examination board.

Acknowledgements

I would like to thank my supervisor, Graham Thompson. I have (almost) always enjoyed our exchange, he has taught and inspired me, and has been a brilliant mentor. He has shown continuous commitment to his role, I will always remember and value this gift.

I would also like to thank Dr Eram Rizvi and Professor Steve Lloyd for their role in my progression throughout the past four years, and Dr John Morris and Dr Dan Traynor, for their time and teachings for which I am grateful.

I would like to thank my mother for her love, encouragement, and of course money. Without her, I may not have reached the final phases of this PhD.

I would like to thank my sister, Emily, for her encouragement, belief, and for bringing into this world my wonderful niece, Evie, who I look forward to getting to know as she grows up in the future.

I would like to thank my father for his belief in me, his advice, and our conversations which inspire me.

I would like to thank all my office companions, especially Jack Goddard, for their company and advice over the years.

I would like to thank the computing team who kindly fix my computer when I cannot.

Finally I would like to thank the STFC for funding my activities for most of the past four years. They have made it possible for me to pursue my interests, and for that I am truly grateful.

Abstract

The Standard Model of particle physics is discussed with emphasis on light quark QCD, and existing data on light quark fragmentation from e^+e^- annihilation and deep inelastic scattering experiments.

A method is developed to measure the directionally correlated pionic scaled momentum distribution, or partonic fragmentation function, in large hadron collider conditions. Jet algorithms are used to provide partonic momentum estimates, which in turn scale the hadronic momenta. The associated resolution is unfolded.

Hadronic profiles about the parton are examined at Monte Carlo "truth" level. There is found to be a uniform uncorrelated background, which may be estimated event-by-event in regions away from jets and then subtracted statistically from the final distributions.

A variable radius cone sampling method is used to count correlated charged hadrons and this also provides a method of coping with any poor directional resolution of jet algorithms. Extrapolation techniques make an estimated measurement possible when the largest safe sampling radius is not large enough to include all correlated hadrons.

A novel method to calculate jet mass using jet collimation information available from the FAPS method is demonstrated.

The algorithm was tested over an order of magnitude in hard scale ($\sim$100GeV $\rightarrow$ 1TeV) with two standard ATLAS reconstructed level Monte Carlos, Pythia and HERWIG, and the calculated fragmentation function is found to be in agreement with the trend of previous data at the hard scale overlap. These models have very different hadronisation models, so may be used to estimate systematic error and test feasibility for a possible full large scale measurement in data. Such work could support the concept of quark universality by establishing propagator invariance.

Contents

1 Introduction **1**

1.1 Motivation 1

1.2 Measurement in the Hadronic Environment 2

2 Partons **3**

2.1 The Standard Model 3

2.2 Quantum Chromodynamics 4

2.3 The Hard Scale 5

2.4 Scaling Violations 9

2.4.1 Higher Orders 9

2.4.2 The "Running" Coupling Constant 9

2.4.3 Softening of the Fragmentation Function 9

3 Monte Carlo Models **11**

3.1 Hadron Collisions 11

3.1.1 Practical Interpretation 12

3.2 Approximating pQCD 13

3.2.1 Parton Showers 13

3.2.2 Lund String Model 14

3.3 Event Samples 15

4 Hadronic Profiles **17**

4.1 Motivation 17

4.2 Detector Coordinate System 17

4.3 p_T Deposited as $f(\eta, \phi)$ 18

4.4 Hadronic Profiles of Partons 19

4.5 Background Variation with the Hard Scale 22
4.6 Hadronic Profile Shape 22
4.7 Conclusions to Hadronic Profiles Study 23

5 Definitions of x_p 24

5.1 Transverse Fragmentation 27

6 Parton Fragmentation 30

6.1 Introduction 30
6.2 Single Parton Fragmentation 31
6.3 Definition of a Parton Fragmentation Function 32
6.4 Previous Measurements 35

7 Parton Fragmentation Algorithm 38

7.1 Method Philosophy 38
7.2 Algorithm Overview 38
7.3 Background Subtraction 40
7.4 Background Proportions in Measuring Cones 42
7.5 Background Solid Angle Size 43
7.6 Statistical Uncertainty on Signal 44
7.7 Systematic Uncertainty on Signal 45
7.7.1 Systematic Uncertainty on Background Subtraction 45
7.7.2 Estimation of Systematic Uncertainty on the Background 45
7.8 Extrapolation and Optimum Solution 48
7.8.1 Extrapolation of D_{signal} vs R Curves 48
7.8.2 Optimal Solutions 51
7.8.3 Extrapolation at Larger Hard Scale 51

8 FAPS Transverse Fragmentation 54

8.1 Jet Shape Parametrisation 54
8.2 Interpretation of Jet Shape 56
8.3 Results 58

9 Parton Estimator Algorithms **60**

9.1 TRAPS . . . 60

9.2 TRAPS Leading Order Event Selection . . . 63

9.3 Parton Estimator Algorithm Comparison . . . 66

9.4 Algorithmic Unfolding . . . 67

9.4.1 The TUnfold Package . . . 68

9.4.2 Unfolding Results . . . 68

9.4.3 TUnfold Instabilities . . . 70

9.5 Extrapolation using Jet Finders . . . 71

9.6 Jet Correlations in η . . . 74

9.7 Optimal Solutions Results . . . 75

9.8 Comparison with the Anti-K_T Algorithm . . . 78

9.8.1 Anti-k_T Selection . . . 79

10 Hadron Measurement **83**

10.1 Calorimetry . . . 83

10.1.1 ATLAS Calorimetry . . . 84

10.2 Tracking . . . 86

10.2.1 ATLAS Tracking . . . 87

10.2.2 Jet Energy Resolution . . . 88

11 Reconstructed Tracks **89**

11.1 Track Selection . . . 89

11.2 Track Reconstruction Efficiency . . . 92

11.3 Unfolding of Track Momenta . . . 95

11.4 Bin-by-bin Correction . . . 100

11.5 Measured Level x_p Bin Efficiencies and Purities . . . 101

11.5.1 Measured vs Truth Efficiencies and Purities . . . 102

11.5.2 Anti-k_T and TRAPS Comparison . . . 102

12 Systematic Error in Resolution Unfolding **104**

12.1 Monte-Carlo Model Systematics . . . 104

12.2 Model Differences . . . 105

12.3 Systematic Uncertainties of Unfolding Techniques 105
12.4 Comparison of Uncertainties 109

13 *Jet* Fragmentation Comparison **110**

14 Preparation For Data Analysis **113**
14.1 Event Cleaning 113
14.1.1 Data Quality 114
14.1.2 Collision Event Selection 114
14.1.3 Jet Selection 115
14.2 Fragmentation Data Analysis Model 118
14.2.1 Trigger 118
14.2.2 Event Selection 119
14.2.3 Monte Carlo Concatenation 119
14.2.4 Comparing Monte Carlo and Data 120

15 Conclusions **122**
15.1 Outlook 125

Bibliography **129**

List of Figures

2.1 4

2.2 Diagram to indicate beyond some separation the colour string between two quarks "splits", producing a new quark-antiquark pair since this is more energetically favourable. 4

2.3 Two body scattering 5

2.4 Feynman diagrams of s (left), t (centre), and u (right)-channel scattering. The direction of time is from left to right, A solid line with and arrow pointing (backward)fowrard in time indicates a (anti)fermion and a dashed line indicates a force carrying boson. 6

2.5 Feynman diagrams of the $2 \to 2$ particle scattering processes which occur at leading order at the LHC. A fermion line with an arrow pointing from left to right indicates a quark, an arrow in the opposite direction indicates an anti-quark and the curly lines represent gluons. In each line the different possible sub-processes with the same initial and final states are shown. . . 7

4.1 The azimuthal and pseudorapidity deposition of p_T. 19

4.2 A ϕ profile is the p_T weighted $\Delta\phi$, in a ± 0.5 slice $\Delta\eta$ away from the parton. 20

4.3 Hadronic p_T deposition profiles measured with three event selections. Each row corresponds to a different requirement on the minimum η separation of the two outgoing partons, $\Delta\eta > 0$ (top), $\Delta\eta > 0.5$ (middle) and $\Delta\eta > 1.0$ (bottom). Pythia J5 Monte Carlo has been used. 21

4.4 In the Monte Carlo, parton hadronic profiles are wider in η than in ϕ. Profiles measured with Pythia J3 Monte Carlo. 23

5.1 The Breit frame. 25

5.2 28

5.3 29

6.1 The physics describing the short (between dashed lines) and long distance is factorised. [32] 31
6.2 The fragmentation function obtained with a naïve unit cone, calculated with J3, J5, and J7 Pythia Monte Carlo samples. 34
6.3 Results from previous fragmentation studies, obtained from e^+e^- (DELPHI [35], TASSO [4], MARKII [36] and AMY [37] and DIS (H1 [38] and ZEUS [39]) experiments. 36

7.1 The background fragmentation function measured using a unit square solid angle in each of the four (eight considering $+\pi$ and -π in azimuth) topology-specified background locations defined in Table 7.3. Measured with Pythia (J5) Monte Carlo at the truth level. 41
7.2 The ratio of background to total hadrons measured using a cone of unit radius. Measured with Pythia J3, J5 and J7 Monte Carlo. 42
7.3 Background per unit LISA as a function of solid angle size used to measure background. Study with on J5 Pythia Monte Carlo data. 43
7.4 A cartoon extrapolation of a D_{signal} vs R curve which has not reached plateau with the largest safe radius. 48
7.5 A cartoon illustrating an exponential extrapolation from $R_1 = 0.2$. 49
7.6 Exponential extrapolations and error bars of equal magnitude as the extrapolations for $0 \leq x_p < 0.02$ J3 Monte Carlo. 50
7.7 Exponential extrapolations and error bars of equal magnitude as the extrapolations for $0.7 \leq x_p < 1.0$ J7 Monte Carlo. 52
7.8 Zoomed version of Figure 7.7. 52

8.1 Mean values of a as a function of x_p, calculated at the truth level with HERWIG Monte Carlo samples at increasing hard scale from ∼100 GeV to ∼1 TeV. The straight line fits omit the first two $< x_p >$ values in each case. 55
8.2 Cartoon diagram of how parton branching axis results in k_t^{brem} between 0 and $M/2$. 56

8.3 The defined jet mass as a function of hard scale, using HERWIG Monte Carlo samples at the truth level. 59

9.1 61
9.2 62
9.3 Processes which may degrade the momentum resolution of TRAPS. . . . 63
9.4 65
9.5 69
9.6 The fragmentation function with TRAPS used to provide the parton momenta estimates, as a percentage of that obtained using the truth supplied partonic momenta values at low momentum (J3), measured on Pythia Monte Carlo at the truth level. Also shown is the result using TRAPS with the truth supplied magnitude of the partonic momenta and the unfolded result. 70
9.7 D_{signal} as a function of R for $0 > x_p \geq 0.02$. Pythia J3 Monte Carlo has been used at the truth level. 71
9.8 A cartoon η, ϕ map of a parton, ISB, the corresponding background solid angle and the TRAPS supplied parton direction. The dashed lines represent successive iterative search cones. 72
9.9 A cartoon diagram of D_{signal} as a function of R using the TRAPS supplied parton information. The first measured value is discarded since it yields an unphysical extrapolated solution. 73
9.10 The curve of $D_{signal}(R)$ for $0.05 \geq x_p \geq 0.1$, produced using truth supplied parton estimator information from the J3 Monte Carlo sample. 73
9.11 The fragmentation cone divided into quadrants. 74
9.12 D_{signal} as a function of R for $\Delta\eta^2 > \Delta\phi^2$ and $\Delta\phi^2 > \Delta\eta^2$. 74
9.13 76
9.14 77
9.15 80
9.16 81
9.17 82

10.1 The trajectory of a charged particle in the presence of a magnetic field. . 86

11.1 The distribution of matching probabilities for truth particles with a single matched track. 90
11.2 91
11.3 93
11.4 94
11.5 The unfolded distribution of hadronic momenta of reconstructed tracks with efficiency loss correction, shown as a percentage of the truth particle distribution. Measured on the J7 Monte Carlo sample using a unit cone. . 95
11.6 The fraction of all charged hadrons which are pions, in a unit cone centered on the truth parton direction, calculated at the truth level using Pythia J3 and J7 Monte Carlo samples. 96
11.7 The noise x_p distribution as a percentage of that of the measured tracks passing the selection. Calculated with the J3 and J7 Monte Carlo samples using a unit cone. 97
11.8 The ratio R_{noise} measured on the J3 and J7 Monte Carlo samples with a unit cone. 98
11.9 99
11.10 The fragmentation function measured with unit cone at the raw measured level using Anti-k_T jets as input as a percentage of the truth pion fragmentation function, measured with Pythia (J3) Monte Carlo. Also included is the bin-by-bin corrected result. 100

12.1 The ratio of tracks (HERWIG/Pythia) at small and large hard scale (J3 and J7 Monte Carlo) in the signal unit cone and background solid angle at the truth level. 105
12.2 The ratio of tracks (HERWIG/Pythia) at small and large hard scale (J3 and J7 Monte Carlo) in the D_{total} unit cone at the measured level. 106
12.3 Raw measured level D_{total} distribution measured in Pythia with a unit cone, corrected by the bin-by-bin and noise removal/unfolding methods with correctional vectors/matrices measured with HERWIG, at small hard scale (J3 Monte Carlo). 107

12.4 Raw measured level D_{total} distribution measured in Pythia with a unit cone, corrected by the bin-by-bin and noise removal/unfolding methods with correctional vectors/matrices measured with HERWIG, at large hard scale (J7 Monte Carlo). 108

13.1 Official ATLAS *jet* fragmentation function (400-500 GeV) data measurement and FAPS *parton* fragmentation function (400-500 GeV reconstructed Monte Carlo). 111

13.2 The (FAPS) measured fragmentation function as a function of measuring angle, R, in the range $0 < x_p \leq 0.02$, at the D_{total}, D_{signal} and $D_{signal,extrapolated}$ levels. The optimum solution is also shown. 112

14.1 The n90 observable in data and Monte Carlo for the inclusive jet distribution after application of jet timing and ECal noise cuts. The Monte Carlo distribution is shown before (dashed line) and after sporadic noise bursts in the HEC are removed by cutting on the f_{HEC} variable. Results from ATLAS note (left) and from this study (right). 116

14.2 The f_{HEC} observable in data and Monte Carlo for the inclusive jet distribution after application of jet timing and ECal noise cuts. The Monte Carlo distribution is shown before (dashed line) and after single cell jets are removed by cutting on the n90 variable. Results from ATLAS note (left) and from this study (right). 117

14.3 The $f_{quality}$ observable in data and Monte Carlo for the inclusive jet distribution after application of single cell and jet timing noise cuts. Results from ATLAS note (left) and from this study (right). 118

14.4 The f_{EM} observable in data for the inclusive jet distribution after application of single cell and jet timing noise cuts. Results from ATLAS note (left) and from this study (right). 119

14.5 The jet time in data for the inclusive jet distribution after application of single cell and ECal noise cuts. Results from ATLAS note (left) and from this study (right). 120

14.6 p_T of the inclusive jet distribution after all jet cleaning cuts. The difference above 100 GeV is due to limited available Monte Carlo events. Results from ATLAS note (left) and from this study (right). 121

14.7 Diagram showing how MC samples will be concatenated based on transverse momentum values of jets, $p_{T,jet}$, reconstructed with the ($R = 0.6$) Anti-k_T algorithm from available J* samples, which are defined in terms of $\hat{p}_{T,parton}$. 121

15.1 The quark to charged hadron fragmentation function calculated with the FAPS method from LHC Monte Carlo presented alongside previous quark to charged hadron fragmentation data. 124

List of Tables

2.1 The Standard Model of Particle Physics, excluding gravity. 3

2.2 Mean $\hat{p_T}$ and $\sqrt{-t}$ values for the Monte Carlo samples. 8

3.1 Number of expected LHC data events in the ATLAS J $\hat{p_T}$ ranges. 15

4.1 Absolute background density in the different Monte Carlo samples, also expressed as the percentage of the transverse momentum of the parton. . 22

7.1 η_{parton} is the η value of the parton of which the background is being measured. $\eta_{background}$ is the η location chosen to sample that background. . . 40

7.2 The systematic uncertainty estimated limits as a function of x_p range from the J2 Monte Carlo sample. 47

7.3 The systematic uncertainty estimated limits, estimated at the lowest x_p interval with Pythia Monte Carlo at the truth level for the samples used in later analysis. 47

7.4 Extrapolated D_{signal} values and corresponding uncertainties shown as a function or R for $0 \leq x_p < 0.02$ calculated using the J3 Monte Carlo sample. 51

7.5 Extrapolated D_{signal} values and corresponding uncertainties shown as a function of R for $0.7 \leq x_p < 1.0$ calculated using the J7 Monte Carlo sample. 53

8.1 The values of M and $M/\sqrt{-t}$, measured on the HERWIG Monte Carlo samples at the truth level. 58

9.1 Event acceptance after leading order event selection. 64

9.2 Efficiencies and purities in the chosen analysis intervals using the respective parton estimators at low transverse momentum (J3 Monte Carlo). Statistical errors are well below the 1% level. 66

9.3 Efficiencies and purities in the chosen analysis intervals using the respective parton estimators at high transverse momentum (J7 Monte Carlo). Statistical errors are well below the 1% level. 67

10.1 The geometry and $\Delta\eta \times \Delta\phi$ granularity of the ATLAS Calorimeters. . . . 85
10.2 The geometry, media, spacial resolution, number of channels and average number of hits per track of the ATLAS Inner Detector tracker. 87
10.3 ATLAS tracking inverse p_T resolution. 88

11.1 Efficiencies and purities in the chosen analysis intervals using the respective parton estimators at low transverse momentum (J3 Monte Carlo). Statistical errors are well below the 1% level. 101
11.2 Efficiencies and purities in the chosen analysis intervals using the respective parton estimators at high transverse momentum (J7 Monte Carlo). Statistical errors are well below the 1% level. 101

12.1 Systematic uncertainties due to the Monte Carlo Model. The systematic uncertainty quoted is the difference of the subtraction of the Pythia corrected result from the HERWIG corrected result. 107
12.2 Systematic uncertainties due to the Monte Carlo Model. The systematic uncertainty quoted is the difference of the subtraction of the Pythia truth result from the HERWIG corrected result. 108
12.3 Systematic uncertainties due to the Monte Carlo Model. The systematic uncertainty quoted is the difference of the subtraction of the Pythia truth result from the weighted HERWIG corrected result. 109

14.1 The number of events in data and Monte Carlo in the present study after successive stages of the jet sample selection. 117

15.1 The quark to charged hadron fragmentation function and percentage errors calculated with the FAPS method. 127
15.2 The quark to charged pion fragmentation function and percentage errors calculated with the FAPS method. 128

Chapter 1

Introduction

It has been shown that the fragmentation of partons into hadrons has been established within the electromagnetic interaction environment as both factorisable and independent of the production process [1]. The same fragments are produced, having the same share of momentum, whether the parton comes from a pair-decay of a time-like virtual photon or from being knocked out of a proton via the space-like virtual photon of deep inelastic scattering (DIS).

Parton fragmentation has been studied in detail in electron-positron annihilation [2, 3, 4, 5] and DIS [1, 6] experiments. However only *jet* fragmentation [7] has been studied, to a lesser extent, in hadron collisions. Such collisions are a prolific source of fragmentation data, however at the high transverse momenta available recently, background has been an insuperable problem.

The main purpose of this thesis is to demonstrate a large acceptance method (FAPS, Fragmentation Algorithm for Parton Scatters) which may be used to to measure the fragmentation of partons in the hadron collision environment of the LHC [8].

1.1 Motivation

Quantum chromodynamics (QCD) [9] is only well understood at large values of the hard scale. Fragmentation studies allow experimental measurement in both the perturbative and non-perturbative regimes of this theory. Measurements of hadronic momenta within jets at high momenta allows a derivation of the hadron splitting function of partons and could also allow a determination of the strong force coupling constant, α_s [10]. Mea-

surements at low momenta allow comparison of models describing the strong interaction where quantum mechanical perturbation theory cannot make firm predictions.

Any BSM (beyond the standard model) physics search requires a precise knowledge of all standard model processes so that the products of new processes may be distinguished. The majority of the background to new physics signatures in a hadron collision environment comes from the fragmentation of hard scattered partons, which result in jets of hadrons.

Partonic fragmentation measurements in the hadron collision environment also provide for a test of "propagator universality", i.e. whether the partons produced in strong interactions fragment in the same way as those produced via electromagnetic processes, i.e. whether virtual gluons can take the place of the virtual photon without altering the subsequent fragmentation properties.

1.2 Measurement in the Hadronic Environment

In order to study partons through the hadrons they produce it is important to understand other sources of background hadrons. This background, considered without "pile-up" (multiple proton-proton collisions in the same data-taking time interval), is comprised of hadrons from initial state bremsstrahlung (ISB) and multiple interactions of other partons within a given proton (MI). In ATLAS [11] terminology, all particles originating from processes, other than that of interest, are referred to as underlying event (UE). However, in this study, all uncorrelated hadrons which are removed from the *total* fragmentation function to give the *signal* fragmentation function are referred to as "background".

ISB is radiation by the parton entering the hard scatter. In hadron-hadron collisions ISB is usually peaked in the "forward" direction in the laboratory reference frame, but not necessarily at the LHC, where the high *virtuality* of the incoming parton even allows "backward" radiation. Multiple interactions are interactions of partons from the two incoming hadrons in addition to the primary hard scatter. It will be shown that both ISB and MI sources result in a "rapidity plateau" (discussed in Chapter 4) deposition of low transverse momentum in the laboratory reference frame which is isotropic in azimuth.

Chapter 2

Partons

2.1 The Standard Model

In the current theory of particle physics, the Standard Model, matter consists of twelve fundamental "building blocks" or fermionic particles, the interactions of which are mediated by four force-carrying bosonic particles, see Table 2.1 [12]. The standard model

matter constituents				*force carriers*	*higgs*
quarks	u	c	t	γ	H
	d	s	b	g (8 colours)	
leptons	ν_e	ν_μ	ν_τ	Z^0	
	e	μ	τ	$W^\pm$	

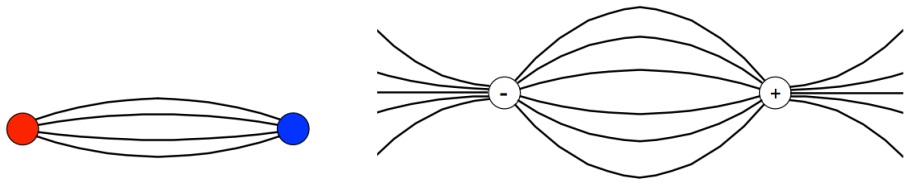
Table 2.1: The Standard Model of Particle Physics, excluding gravity.

matter constituents are the six strongly interacting quarks organised into weak-isospin doublets, the three weakly interacting leptons, and the three corresponding lepton neutrinos, also in weak-isospin doublets. The standard model force mediators are the photon, γ which mediates the electromagnetic force, the gluon, g, which mediates the strong colour force and the $W^\pm$ and Z^0 particles, which mediate the weak force. There are also a set of antimatter fermions, which have equal but oppositely signed quantum numbers. These antifermions are the building blocks of antimattter. The Standard Model also incorporates the higgs mechanism by which the $W^\pm$ and Z^0 are given mass upon breaking of the electroweak symmetry. The fermions may also acquire mass by coupling to the higgs field in a different way to the gauge bosons. This thesis concerns quarks and gluons which are collectively called partons.

2.2 Quantum Chromodynamics

The force between any two coloured quarks is thought to remain constant after a separation approximately equal to the diameter of the proton, being equal to that required to lift a tonne on the surface of the Earth [13].

The coupling of the quarks to the gluons occurs due to the quarks having colour charge. The peculiar quality of the mediators of the strong force, the gluons, is that they themselves also possess colour charge, making QCD (quantum chromodynamics) a non-Abelian theory, in contrast to the Abelian theory of the electromagnetic interaction QED (quantum electrodynamics) in which the gauge bosons are neutral [14]. Consequently gluons are self coupling and therefore field lines of force between two partons are attracted to one another as shown in Figure 2.1(a) which may be compared to the case of electromagnetism shown in Figure 2.1(b). Increasing the separation between two coloured objects causes



(a) Colour field lines between two colour sources

(b) Electromagnetism field lines between electric charges

Figure 2.1: Diagram to indicate gluons are self coupling and are attracted to one another, sub-figure (a), and photons have no such property, sub-figure (b).

the field lines to become tube or *string* like as shown in Figure 2.2. The colour string

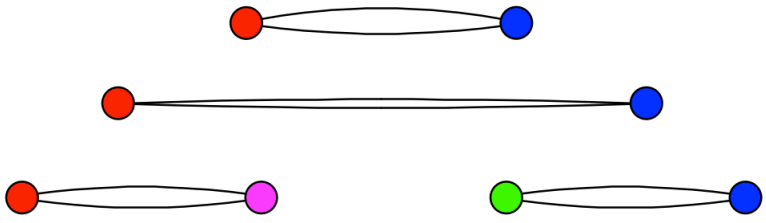


Figure 2.2: Diagram to indicate beyond some separation the colour string between two quarks "splits", producing a new quark-antiquark pair since this is more energetically favourable.

has constant energy density per unit length of ∼1 GeV/fm [15], thus increasing the separation of two quarks increases the potential energy stored in the string. At some point it becomes energetically more favourable for a new quark-antiquark pair to be created

than for the string length to increase, which therefore occurs, as depicted in Figure 2.2. Though no analytic proof exists, the fact that the gluons are colour charged is believed to be the cause for quarks being confined to bound states. These colour neutral states of two or three *valence* (anti-) quarks are known as mesons and baryons respectively, or collectively as hadrons.

Hadrons collided at high energy act as partons sources, but whilst the partons are scattered off one another, they are never detected alone, only the resultant hadrons are.

2.3 The Hard Scale

The centre-of-momentum (CMS) energy, momentum and scattering angles of any $2 \rightarrow 2$ particle scattering process (Figure 2.3) may be expressed as a function of the Lorentz invariant Mandelstam variables s, t, u, each of dimension $(\text{GeV})^2$. These are given in Equations 2.1 to 2.3, where $(\widetilde{p_1}, \widetilde{p_2})$ and $(\widetilde{p_3}, \widetilde{p_4})$ are the particle initial and final state

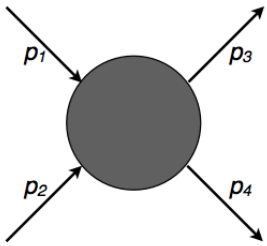


Figure 2.3: Two body scattering

four-momenta respectively. The corresponding Feynman diagrams are given in Figure 2.4. The possible leading order QCD $2 \rightarrow 2$ particle scattering Feynman diagrams are shown in Figure 2.5.

$$s = (\widetilde{p_1} + \widetilde{p_2})^2 = (\widetilde{p_3} + \widetilde{p_4})^2 \tag{2.1}$$

$$t = (\widetilde{p_1} - \widetilde{p_3})^2 = (\widetilde{p_2} - \widetilde{p_4})^2 \tag{2.2}$$

$$u = (\widetilde{p_1} - \widetilde{p_4})^2 = (\widetilde{p_2} - \widetilde{p_3})^2 \tag{2.3}$$

Except the low cross-section Higgs, there are no high mass $qq/q\bar{q}$ or qg resonances expected in the hard scale range analysed and $|t|$ is very much smaller than s. The $s(t)$-channel propagator is roughly proportional to $1/s(1/t)$ for massless propagators in the absence of resonances. The cross section is inversely proportional to the square of the propagator

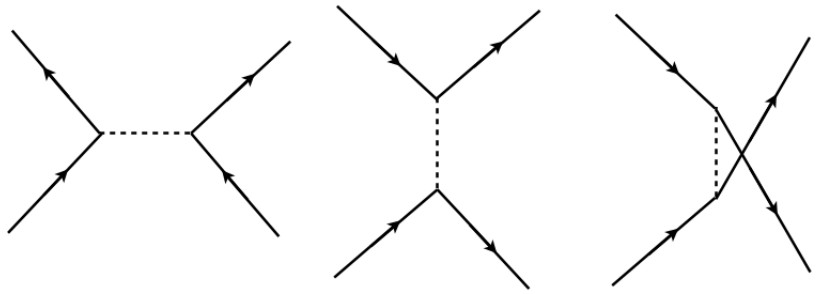


Figure 2.4: Feynman diagrams of s (left), t (centre), and u (right)-channel scattering. The direction of time is from left to right, A solid line with and arrow pointing (backward)fowrard in time indicates a (anti)fermion and a dashed line indicates a force carrying boson.

term, so if the overall E^* is only twice as large as the transverse momentum, the s channel cross section will be approximately sixteen times smaller. At the very large fractional momentum of the parton, x, which is necessary for high p_T events there are fewer anti-quarks than quarks in the collision protons at the LHC, further suppressing the annihilation channel with respect to the scattering channels. Thus the cross section is very much dominated by t-channel scattering. This kinematic variable then provides the appropriate hard scale for the vast majority of parton interactions.

Hard scatters are those involving large values of this hard scale. The hard scale variable may be used in the perturbative expansion of the running strong coupling constant and this choice makes the series converge more quickly.

In previous *DIS* fragmentation measurements the invariant momentum transferred between the scattering objects, $Q = \sqrt{-t}$, was used to represent the hard scale, simply because there are no particles with *eq* quantum numbers. The same observable ($\sqrt{-t}$) is used in the present measurement for relevant comparison.

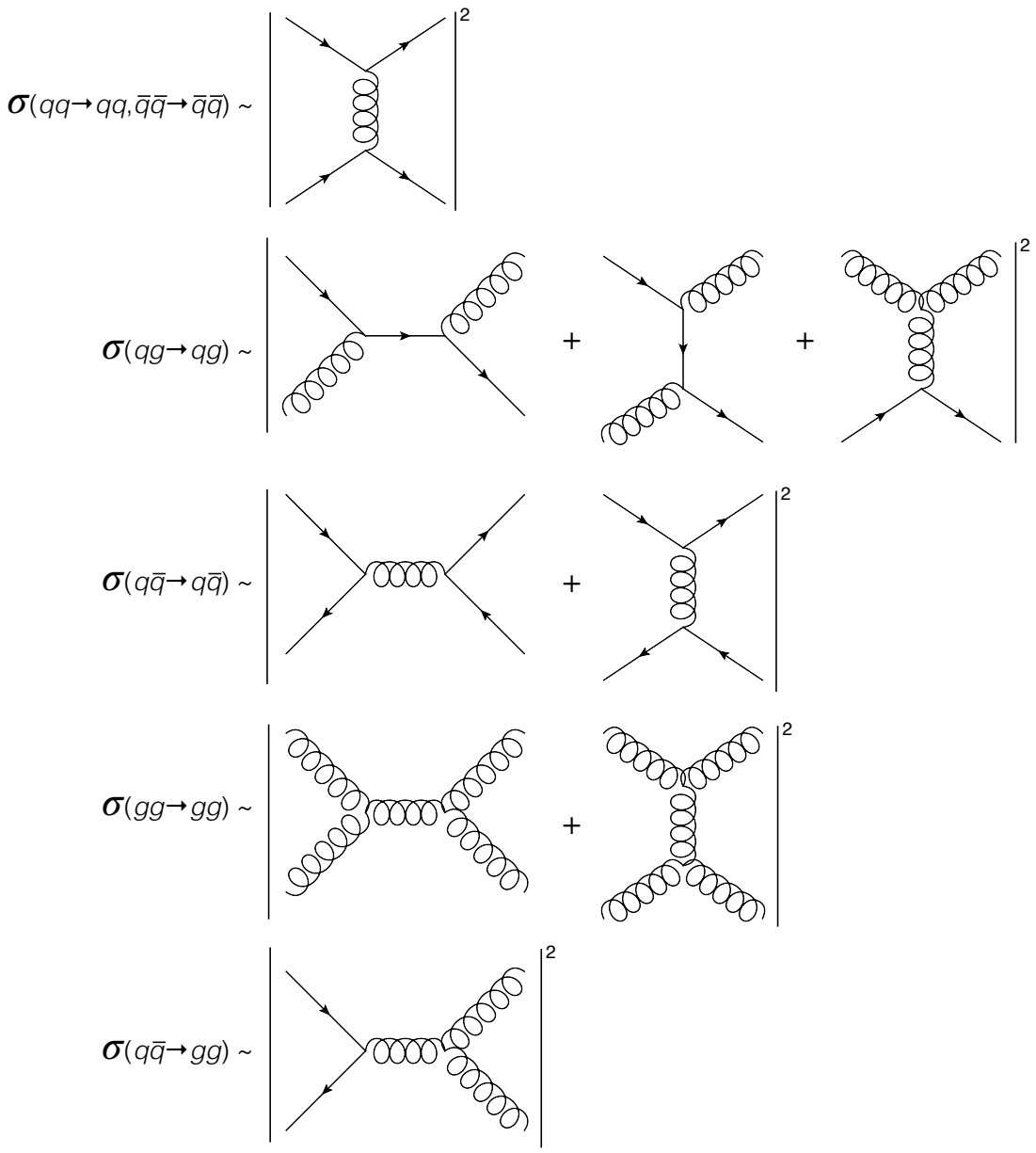


Figure 2.5: Feynman diagrams of the $2 \rightarrow 2$ particle scattering processes which occur at leading order at the LHC. A fermion line with an arrow pointing from left to right indicates a quark, an arrow in the opposite direction indicates an anti-quark and the curly lines represent gluons. In each line the different possible sub-processes with the same initial and final states are shown.

Monte Carlo Sample $\hat{p_T}$-range (GeV)	Mean $\hat{p_T}$ (GeV)	Mean $\sqrt{-t}$ (GeV)
70-140	87	132
140-280	172	245
280-560	335	459
560-1120	669	886
1120-2240	1238	1738

Table 2.2: Mean $\hat{p_T}$ and $\sqrt{-t}$ values for the Monte Carlo samples.

The relationship between $\sqrt{-t}$ and the transverse momentum in the centre of momentum frame, $\hat{p_T}$, may be derived starting from Equation 2.2. In the centre of momentum frame $\widetilde{p_1} = -\widetilde{p_2}$, $\widetilde{p_3} = -\widetilde{p_4}$ and $|\widetilde{p_1}| = |\widetilde{p_3}| = \hat{p}$, and so for massless partons

$$-t = 2|\hat{\vec{p}}|^2(1 - \cos\hat{\theta}), \tag{2.4}$$

where $\hat{\theta}$ is the angle through which the parton under consideration is scattered within that frame. Substituting $\hat{p_T} = \hat{p}\sin\hat{\theta}$, and using trigonometric identities

$$-t = \hat{p_T}^2 \frac{2}{1 + \cos\hat{\theta}} \tag{2.5}$$

and so the relationship between $\sqrt{-t}$, and $\hat{p_T}$ is given by the expression

$$\sqrt{-t} = \hat{p_T}\sqrt{\frac{2}{1 + \cos\hat{\theta}}}. \tag{2.6}$$

Note that in *jet* fragmentation [16] the jets are considered individually and therefore p_T is commonly chosen as the appropriate hard scale observable. This is a good approximation, however $\sqrt{-t}$ would be more suitable.

The next chapter will detail the Monte Carlo production which forms the input to this feasibility study and shows how these comprise of non-overlapping ranges of $\hat{p_T}$. The mean $\sqrt{-t}$ values of these Monte Carlo samples used are compared to those of $< \hat{p_T} >$ in Table 2.2 with both calculated assuming that initially the collinear CMS partons entering the hard scatter have no net momentum transverse to the beam, i.e. that the beam remains on the z-axis after the transform. For a CMS scattering angle of $\pi/4$, $\sqrt{-t}$ is only approximately 14% larger than $\hat{p_T}$, whereas for pure transverse scatter of $\pi/2$, $\sqrt{-t}$ is approximately 40% larger. For a rare backwards scatter, which can not be identified in practice, $\sqrt{-t}$ is much larger, e.g. almost three times larger for $\hat{\theta} = 3\pi/4$.

2.4 Scaling Violations

2.4.1 Higher Orders

In the following, the "matrix-element" contribution to the interaction probability is considered rather than its product with the phase space which gives the total cross-section.

At any energy there are many possible final states to an initial state two-body interaction, i.e. $2 \to 2$, $2 \to 3$, etc. Even allowing for closed loops, each process leading to a higher multiplicity must occur with decreased probability by a factor of an absolute, not energy dependent (i.e. "scaling") coupling constant, the value of which must be below unity to allow perturbative theories such as QED and pQCD.

With decreasing energy, a QCD $2 \to 3$ process will coalesce into a $2 \to 2$ one, due to self coupling of the force carriers. The $2 \to 2$ process therefore has an enhanced probability at lower energy, at the cost of the suppressed $2 \to 3$ process, which itself is enhanced by the $2 \to 4$ diagram, and so on.

In essence, an *extra* loop now contributes to the $2 \to 2$ process and a similar mechanism occurs for the higher orders. The decreasing probability for higher order final states, by the factor of the coupling constant, means that while there are enhancements to the higher orders, they are not as large as the suppressions due to the coalescence, and the largest enhancement is to the $2 \to 2$ process.

2.4.2 The "Running" Coupling Constant

The above may be described in terms of a running coupling constant, $\alpha_S(Q)$, which accounts for all such loops and vertex diagrams and therefore has a larger value at lower energies. The suppression of higher orders by the factor $\alpha_S(Q)$ is then more significant at lower energies, again enhancing the $2 \to 2$ process with respect to the higher orders. Accounting for loops and vertex diagrams contributing to a process of some order in this way is known as renormalisation.

2.4.3 Softening of the Fragmentation Function

Viewed in reverse, with increasing energy, a scaling violation is seen, which results in an increased number of higher multiplicity final states than would be expected with strict scaling. A classically resolved jet will therefore contain more low momentum hadrons

with increasing hard scale. This will be at the expense of higher momentum hadrons, i.e. higher topologies have become discernable at the expense of loops.

Such scaling violations may explain the softening of the fragmentation function with increasing hard scale, as will be shown in Section 6.4.

Chapter 3

Monte Carlo Models

The purpose of a Monte Carlo generator is to produce simulated final states in comparable detail as would appear in real experiments. This "pseudo-data" can then be reconstructed and analysed with exactly the same software as used with real data.

Analysis of generated Monte Carlo collisions has many fewer problems associated than data. For example, they are free from the problems caused by apparatus malfunction, they are in well defined kinematic regions, and sufficiently large samples may be generated such that statistical uncertainties are minimal. In addition, such events have the huge advantage of being accompanied by the physics information at generated or "truth" level. These are often used to compare to data for insight into improving such models and the correlations between these and reconstructed values will be used later in this study in resolution unfolding methods.

For all these reasons the algorithm was tested on such events during development. There are several popular general purpose (leading order) event generators, (e.g. Pythia [17], HERWIG [18] and SHERPA [19]), Pythia is principally used in this chapter to provide real examples of event generation techniques.

3.1 Hadron Collisions

The perturbative theory describing strong interactions, pQCD, predicts collimated "jets" of hadrons as a result of radiation from hard scattered partons. This theory is, however, incomplete in that it breaks down at low momentum. Lattice QCD [20] provides an alternative framework for (non-perturbative) QCD but is totally computationally impractical

for our purposes.

To a first approximation, interactions between fundamental particles are simple. For example much about the structure of e^+e^- annihilation events at LEP [21], may be understood from the skeleton process $e^+e^- \to Z^0 \to q\bar{q}$ [17]. Corrections to this approximation may be divided into three categories, namely those accounting for bremsstrahlung, "higher order" loop diagrams in matrix elements, and confinement. Perturbative processes remain the dominant factor in governing event shape energy flow, and corrections due to hadronisation of partons after a hard scatter then merely "smear" resolution.

Perturbative corrections to the basic 2→2 (LO) process have traditionally been done in Monte Carlo generation with two approaches, namely by the inclusion of higher order diagrams in matrix element calculations, and by parton showering.

Bremsstrahlung type corrections provide a simple way to model higher order effects, for example higher multiplicities. Probabilistic approximations to full perturbative calculations may be used, providing an alternative approach to the otherwise necessary, increasingly complex calculations, which are too computationally demanding for practical generation of appropriately large samples.

Matrix element corrections may be referred to as "true higher order corrections" and include interference effects. The required perturbative calculations are very complicated and results of calculations beyond one trivial order, i.e. beyond one extra branching (NLO), have rarely been presented. Some corrections in this category are however more trivial, well known, and frequently applied, i.e. those solved by the running of the strong coupling constant through loop diagrams. For accurate prediction of, say, the rate of well separated jets, higher order matrix elements must be used.

There is no reliable way to know the error involved in the approximation of *not* including higher order terms, without calculating them, but generally each higher order term provides of the order of a 10% correction.

3.1.1 Practical Interpretation

Events containing jets observed in real data are modelled reasonably well using 2→2 matrix elements [17]. Even events with well separated jets may be simulated with this approach, using parton showering with parameters tuned to data.

It is practically useful, and a good approximation, to consider the transverse mo-

mentum of each of the two highest p_T jets in an event as originating from two outgoing hard scattered partons. The momentum transfered between the scattered partons may then be calculated assuming forward scattering giving access to the highest Q-scale of the interaction to a good approximation.

3.2 Approximating pQCD

A useful and very successful technique in dealing with functions describing the physics of particle interactions is to factorise them into two, or more, functions each depending only on the physics involved at one momentum scale. Factorisation into calculable functions describing high momentum physics (e.g. parton cross sections), and those describing the physics at long distances at low energies which may be measured with global fits to data (e.g. fragmentation function) enables a full prediction of strong reactions. This is the technique commonly used in event generation, where, initial and final state parton showers and the hard interaction matrix element are calculated separately. The full matrix element is calculated only for the hard scattering process, to leading order (2→2), or next-to-leading order (2→3), depending on the generator.

3.2.1 Parton Showers

In initial and final state parton showers successive parton branchings e.g. $q \to qg$, $g \to gg$ and $g \to q\bar{q}$, are performed, with each parton being given a "virtuality" Q^2 $(= E^2 - p^2)$.[1]

The process to construct initial state showers is somewhat different from those of the final state. In this Sudakov picture [22], the parton begins on-mass-shell (massless) and given that a hard scatter will occur, branching is performed such that one branch is given a positive virtuality (bremsstrahlung) and the other a negative virtuality, thus conserving energy and momentum at each vertex. Further space-like ($E^2 - p^2 < 0$) branchings are performed up to the scale Q^2 where the (negatively off-mass-shell, $p^2 > E^2$) parton enters the hard scatter, at which point the mass squared of the parton is between 0 and $-Q^2_{hard}$, with the value given according to the Sudakov form factor. Momentum is then transfered to the parton in the hard scatter, making it positively off-mass-shell and thus enabling a time-like parton shower and ultimately fragmentation.

[1] Note that time ordering of parton showers is only quantum mechanically meaningful if the only reversible measurement is the hard interaction.

After a hard interaction (at some scale Q^2_{hard}, corresponding to an associated $\hat{p_T}$ in the generation range), each of the outgoing two (or three) partons are given positive virtualities Q^2 ($\leq Q^2_{hard}$) according to the Sudakov form factor, and final state showers are evolved with time-like ($E^2 - p^2 \geq 0$), virtuality-decreasing branchings. The Q^2 interaction scale falls, until some defined cut-off value, Q_0, where hadronisation occurs.

In practice, the likelihood of a scatter at scale Q^2_{hard} is calculated first, and the initial state shower is generated backwards (time reverse order) from that scale, "guessing" the partons which branched, given the incoming pair, and matching to an initial parton with a probability derived from known PDFs (parton density functions) from DIS analyses.

Like many of the ideas in this chapter the Sudakov parton shower picture is a heuristic theory whose language, but not verity, has been useful in developing the FAPS algorithm, in this case especially the content of Chapter 8.

3.2.2 Lund String Model

The inclusion of the Lund String model [23] in Pythia (and thus SHERPA, which interfaces with Pythia for hadronisation) in addition to the basic parton shower, gives a closer approximation to QCD at all orders.

In the string model outgoing partons from a hard scatter are connected to the remnant proton by gluon fields. With increasing separation of the colour charged objects, the field lines become tube-like giving rise to a constant energy density ($\sim$1 GeV/fermi), due to gluon self coupling. When the objects are separated by more than a fermi, the potential energy stored in the string *increases* linearly with length. This may be contrasted with QED, in which the potential energy *decreases* with the inverse-square of the charge separation.

As discussed in Chapter 2 with increasing separation, at some point it becomes energetically favourable to "snap" the string producing a new $q\bar{q}$ pair, which then form the new end points of the two shorter strings. The process is repeated until the parton combinations or clusters are at small separations, and may be considered as effectively colourless hadrons, which may, however, still be massive enough to undergo further hadronic decays.

The model produces hadrons roughly isotropically about the parton, but also dependent on the string direction. Such production of hadrons preferentially in the η-direction is consistent with observations in data [24]. There are, however, other models for the

causal process.

The independent and cluster hadronisation models have also been popular. In the former, quarks fragment independently depending on momentum (energy) not virtuality, however this model is not Lorentz invariant.

The cluster model [25] starts by non-perturbatively splitting gluons after the parton shower. Colour singlet $q\bar{q}$ combinations are assumed to form clusters which decay isotropically into pairs of hadrons, taking into account the density of states with appropriate quantum numbers. This model is used in HERWIG. It contains fewer parameters than the Lund model, however it has been less successful at describing data.

More complicated models of fragmentation exist, which have also not been as successful at describing experimental data, for example the colour dipole model [26], which treats coloured objects connected by strings as dipole aerials, which emit partonic quanta.

3.3 Event Samples

Five Pythia [17] and HERWIG [18] Monte Carlo samples which are tuned to ATLAS 7TeV data have been used throughout this study, with each Pythia sample containing approximately 1.4 million events, and each HERWIG sample containing approximately one million events. These are of the same format (Event Summary Data, ESD) as official ATLAS data, and the standard ATLAS Athena computing framework [27] has been used for analysis. The $\hat{p_T}$-ranges of the event samples are repeated in Table 3.1 which now uses the ATLAS J* nomenclature of the appropriate trigger to use to most efficiently collect such events. The calculated LO cross-section has been used to find the expected number of events per fb^{-1} of collected luminosity, each contributing two scattered partons. Development and testing of the algorithm at the reconstructed level, i.e. using generated collisions which have been passed through the GEANT simulation [28], has enabled performance to be studied with realistic measurement resolutions. The algorithm has

Sample	$\hat{p_T}$ range [GeV]	Expected number of events per fb^{-1}
J3	$70 - 140$	1.3×10^8
J4	$140 - 280$	8.7×10^6
J5	$280 - 560$	4.3×10^5
J6	$560 - 1120$	1.2×10^4
J7	$1120 - 2240$	87

Table 3.1: Number of expected LHC data events in the ATLAS J $\hat{p_T}$ ranges.

been tested on sufficiently low $\hat{p_T}$ samples to enable comparison to data of previous quark fragmentation measurements, and up to a maximum $\hat{p_T}$ where significant measurement would be possible given planned integrated luminosity.

Chapter 4

Hadronic Profiles

In this chapter the event-normalised deposition of transverse momentum in typical high jet p_T hadron collider events is studied at generator level in Monte Carlo samples. The preliminary survey represents all final state particle sources and without reference to their charge, species, or momentum fractions. The shapes of hadronic profiles around the scattered parton direction are also shown. The levels of background below these profiles are estimated, and their variations are studied, over an order of magnitude in hard scale.

4.1 Motivation

In order to measure the fragmentation of partons in the background rich hadron collision environment the angular distribution of the expected background must be known and a suitable region to sample the background must be chosen. To be sure to measure all of the charged hadrons which result from a fragmenting parton, it is necessary to consider the expected dispersion of those hadrons with respect to the parton direction.

4.2 Detector Coordinate System

Kinematics in a hadron collider detector such as ATLAS are described using the variables x, y, z and η, ϕ. Using ATLAS as and example, the variables are defined such that x points to the centre of the LHC ring, z along the beam direction and y upwards. At point 1 (where ATLAS is located), looking to point 8 is the direction of positive z. The azimuthal angle, ϕ, is defined in the range $-\pi$ to $+\pi$, such that $\phi = 0$ along the positive x-axis, increasing in the clockwise direction while facing the positive z-direction. The

polar angle, θ, is measured from the positive z-direction and pseudorapidity, η, the high energy limit of rapidity, y, is;

$$\eta = -ln(tan\frac{\theta}{2}). \tag{4.1}$$

Transverse momentum, p_T, is defined to be the momentum perpendicular to the beam axis.

4.3 p_T Deposited as $f(\eta, \phi)$

Since both beams are identical and along the z-axis, the expectation for energy or p_T at a hadron collider is for azimuthal symmetry and longitudinal symmetry about $\eta = 0$. The Feynman prediction is then for a plateau in the rapidity variable for any given small hadronic p_T-range.

Intuitively, the origin of this prediction is related to the fact that for a given small track p_T range it is 'easier' to produce that p_T with a large angle scatter which is rare, however only a small partonic momentum fraction of the parent proton, x_{Bj}, is required which is very common. Conversely, with a very common small angle scatter it is 'more difficult' to produce the same track p_T, i.e. a larger x_{Bj} is required, which is rare.

For a fixed beam energy the largest track p_T may only be achieved around $\eta \sim 0$ with the largest possible momentum share, x, of the beam energy. Therefore the Feynman rapidity plateau would be expected to be narrow for relatively high p_T hadrons and vice versa, the final total expectation would be for a peaked structure being taller for high values of the hard scale and broader for lower values.

These expectations encourage the introduction of a Lorentz Invariant Solid Angle (LISA) in (η, ϕ) space formed by a unit opening angle in η and ϕ, to measure the total amount of deposited p_T in a given direction. This should then be invariant to boosts along the beam axis. As may be seen in Figure 4.1 these expectations are indeed seen in the Monte Carlo samples.

The momentum of a bremsstrahlung transverse to a hard scattered parton direction, $k_t^{brem.}$, may be related to its momentum transverse to the z-axis, $p_T^{brem.}$, in the following way. In the transverse direction

$$k_t^{brem.} = p_T^{brem} \sin \Delta\phi \approx p_T^{brem} \Delta\phi$$

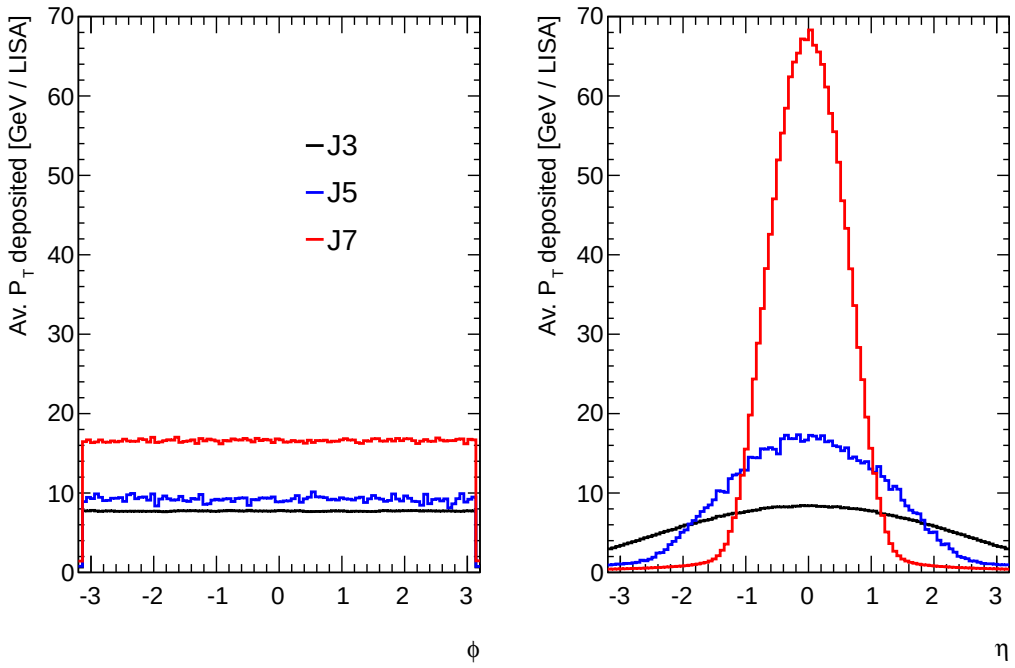


Figure 4.1: The azimuthal and pseudorapidity deposition of p_T.

and in the longitudinal (beam) direction

$$k_t^{brem.} = p^{brem} \sin\Delta\theta \approx p^{brem}\Delta\theta$$

$$= p^{brem} \sin\theta\Delta\eta = p_T^{brem}\Delta\eta,$$

and so there is an ∼isotropic deposition of k_t in (η, ϕ) space with respect to the scattered parton direction, supporting the use of a cone in (η, ϕ) space for sampling hadrons correlated with a hard scattered parton and a LISA to sample p_T deposited in a given direction.

4.4 Hadronic Profiles of Partons

The p_T weighted $\Delta\phi$, in a ± 0.5 slice $\Delta\eta$ away from the parton gives the ϕ profile, see Figure 4.2. An η profile is produced in a similar way. The hadronic profiles allow observation of the angular deposition of hadrons with respect to the parton. The profiles measured at the generator level are shown in Figure 4.3.

The red and magenta sections represent MI + ISB partons respectively, identified using *truth*. Pythia doesn't allow hadrons to be associated directly with any partons for

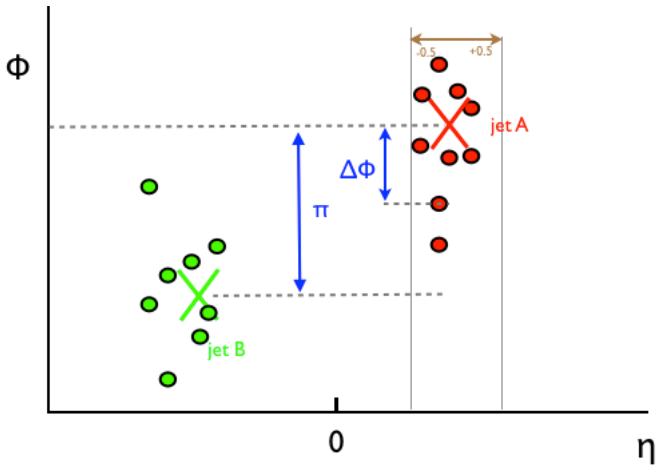


Figure 4.2: A ϕ profile is the p_T weighted $\Delta\phi$, in a ± 0.5 slice $\Delta\eta$ away from the parton.

clear reasons of colour conservation. Therefore the p_T of *partons* are plot to measure the background. On average the hadrons must be found at the same locations as the partons which produce them and so this procedure is justified, the effect of instead measuring *hadrons* from ISB or MI would be a smearing in resolution, which is not important since there is no fine detail in these distributions.

The secondary peaks in the ϕ profiles in Figure 4.3 are shown to be a result of measuring the profile of partons which occur at similar η locations. Measuring the ϕ profile of one parton picks up some hadrons from the other. The secondary peaks are removed by selecting events in which partons are separated by a minimum $\Delta\eta$, and for future analysis the event selection of $\Delta\eta > 1.0$ is applied to the two highest p_T jets such that a measuring cone may be used to which is large enough to sample all hadrons correlated with one scattered parton, while avoiding those from another.

When weighting with their p_T, hadrons coming from the interactions of other partons within the proton, will naturally show a central peak for exactly the same reason as those emerging from the primary collision, namely the symmetry of the beams and the higher contributing p_T for a given x_{Bj} at angles close to 90°. Thus the appearance of an MI hadron peak below the main jet peak reflects the common kinematics of all parton-parton collisions in a hadron collider and should be regarded as correlated but not causal. Both reflect the kinematics of superimposed Feynman plateaux.

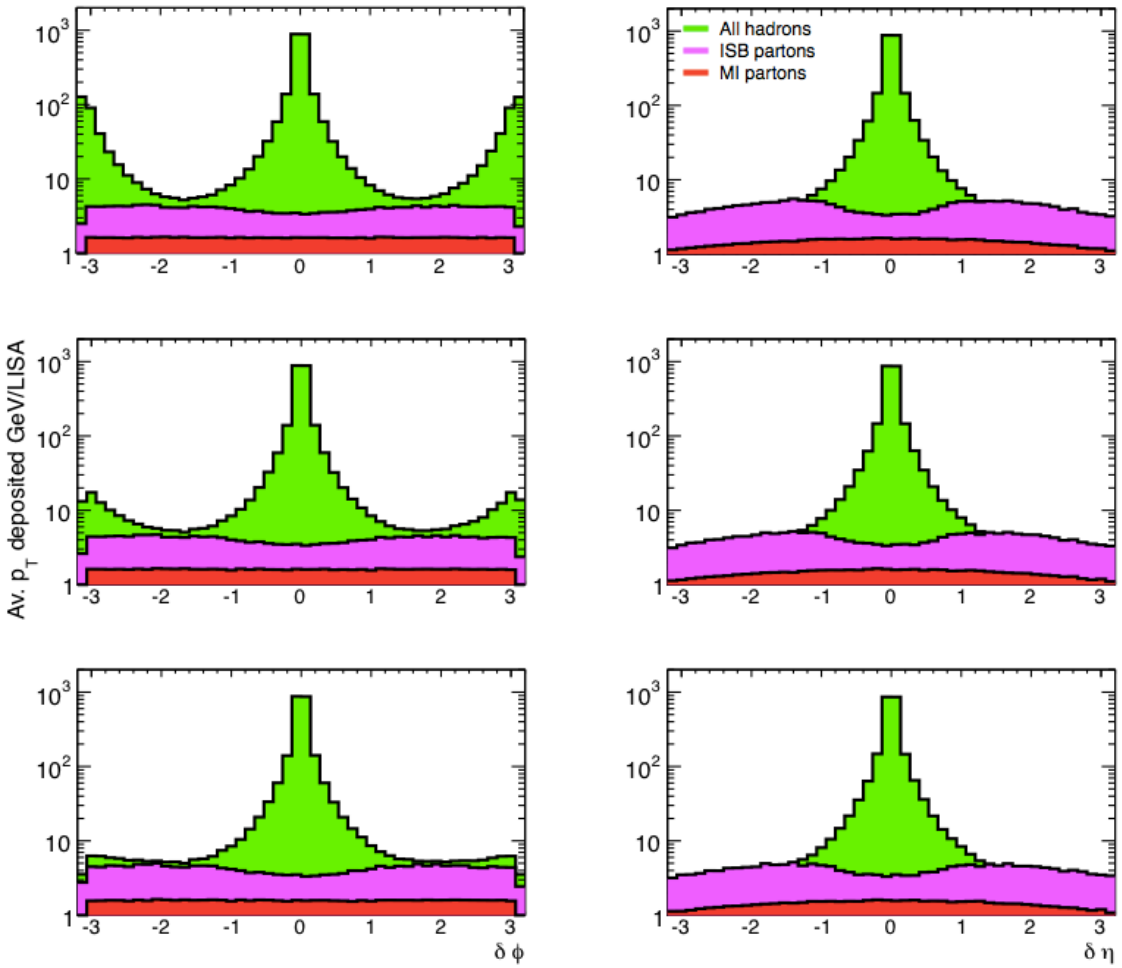


Figure 4.3: Hadronic p_T deposition profiles measured with three event selections. Each row corresponds to a different requirement on the minimum η separation of the two outgoing partons, $\Delta\eta > 0$ (top), $\Delta\eta > 0.5$ (middle) and $\Delta\eta > 1.0$ (bottom). Pythia J5 Monte Carlo has been used.

The variation of ISB may be due to the fact that quanta produced in the Model, near to the outgoing hard scattered parton, are likely to be defined as FSB, rather than ISB. Due to the relative magnitude of the background, note the logarithmic axis, the details of these variations are not considered a point of concern.

4.5 Background Variation with the Hard Scale

Background was superimposed on the profiles for each of the different Monte Carlo samples, giving a rough measure of the quantitative variation with parton the hard scale. The results are summarised in Table 4.1. Background increases with the hard scale, but at such a rate that it becomes a less significant *proportion* of the partonic p_T. This type of effect has been noted elsewhere [29] and it would seem that selecting, or triggering, on very high p_T localised (jet) energy has the reasonable effect of selecting lower p_T in ISB activity.

Monte Carlo Sample	Background (GeV/LISA)	% Background of Parton
J3	$\sim$2.7	1.4%
J5	$\sim$4.0	0.4%
J7	$\sim$4.5	0.1%

Table 4.1: Absolute background density in the different Monte Carlo samples, also expressed as the percentage of the transverse momentum of the parton.

4.6 Hadronic Profile Shape

The Lund string model of hadronisation is used in the Monte Carlo sample studied. In the Lund model all but the highest energy gluons are treated as field lines which are attracted to each other due to gluon self-interaction, causing narrow tubes or *strings* of colour force. If resolution is sufficient, then parton hadronic profiles will be measured as wider longitudinally than azimuthally in the Monte Carlo, due to the colour string joining the jet partons to those in the remnant protons. Comparison of similar measurements on data could enable a statement to be made about that model.

The parton hadronic profiles in η and ϕ are superimposed in Figure 4.4. They are similar but slightly wider in η, in the range $|\Delta\eta, \Delta\phi| < 1.0$. Outside this range, due to the effect of the other scattered partons close in η, the ϕ profile is artificially wider in azimuth. There is sufficient resolution to observe a shape in the correlated hadrons in generator level Monte Carlo, further study is needed to determine whether resolution will be sufficient to observe the effect in data.

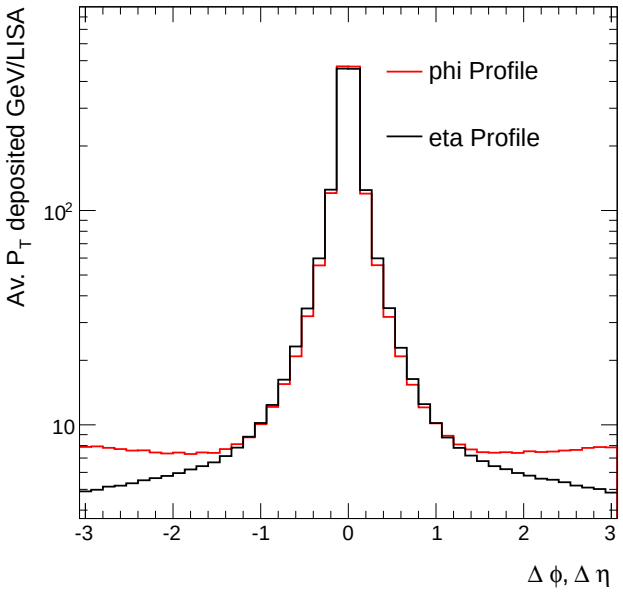


Figure 4.4: In the Monte Carlo, parton hadronic profiles are wider in η than in ϕ. Profiles measured with Pythia J3 Monte Carlo.

4.7 Conclusions to Hadronic Profiles Study

The background to measuring partons at typical LHC energies has been studied in Monte Carlo and found to be ∼5 GeV/LISA. Variation of the background with the hard scale is minimal (2.7 GeV/LISA @ J3, to 4.5 GeV/LISA @ J7). Sampling the background to partons at $\Delta\phi = \pm\frac{\pi}{2}$ has been justified, and sensitivity at the generator level is sufficient to observe the hadronic deposition of scattered partons to be slightly wider in η than in ϕ. This could be due to dynamic effects of the Lund String model, or merely because of the greater phase space in the η-direction for any uncorrelated ISB activity. Similar phenomena have been seen and studied as "long range correlations" in e.g. [24].

Chapter 5

Definitions of x_p

It is the purpose of this thesis to study the momentum distribution of hadrons created from the fragmentation of partons. Clearly higher momentum hadrons will be found in higher momentum jets so to allow comparison it is obvious that the momentum must be scaled in some way by the kinematics of the originating parton.

The Lorentz invariant expression of the scaled hadronic momentum x_p, given in equation 5.1, has the advantage of being invariant to boosts along the parton direction.

Initial fragmentation studies, carried out in e^+e^- annihilation experiments [2, 3, 4, 5], used the definition of x_p given in equation 5.2.

$$x_p = \frac{(E + p_{||})_{hadron}}{(E + p_{||})_{parton}} = \frac{(E + p_{||})_{hadron}}{(E + p)_{parton}} \tag{5.1}$$

In e^+e^-, a quark-antiquark pair are produced out of the vacuum with equal and opposite momenta, each with exactly half of the beam energy, E_{beam} which was known very accurately.

$$x_{p,e^+e^-} = \frac{p_{hadron}}{E_{beam}} \tag{5.2}$$

This definition of x_p may be seen as an approximation to the Lorentz invariant form for hadrons produced at small angles to the parton, $p_{||,hadron} \cong p_{hadron}$, giving the first approximation in equation 5.3. Fragmentation studies in deep inelastic scattering experiments used a similar approximation for comparison purposes, for example to test quark universality.

The approximation also requires that in the laboratory frame the hadrons and partons are of such large momenta that their rest mass energy is small enough to make their energy

approximately equal to their momenta, $E_{parton,hadron} \approxeq p_{parton,hadron}$, as in the second approximation in 5.3. These two approximations are both good and have the additional advantage of changing x_p in opposite ways.

$$x_p = \frac{(E + p_{||})_{hadron}}{(E + p)_{parton}} \approxeq \frac{(E + p)_{hadron}}{(E + p)_{parton}} \approxeq \frac{2p_{hadron}}{2p_{parton}} = \frac{p_{hadron}}{E_{beam}} \tag{5.3}$$

In the e^+e^- experiments all charged particles in all selected hadronic events in the analysis were in principle included and counted as the multiplicity due to two partons. In the *DIS* measurements [1, 6] the Breit frame [30] (see Figure 5.1) was used to measure the hadronic momenta p_{hadron}.

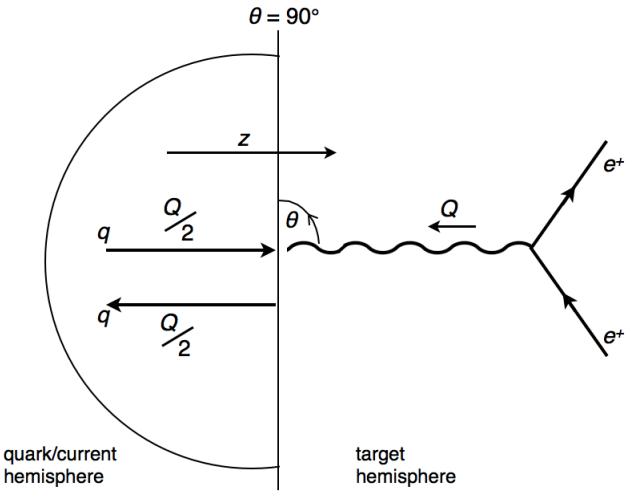


Figure 5.1: The Breit frame.

In this inertial reference frame events are boosted and rotated such that the energy of the exchange photon is zero, i.e. it is completely space-like ($E^2 - p^2 < 0$), and its momentum is the transferred momentum, Q. The momentum of the photon is then chosen to define the z-axis, and the plane of the outgoing and incoming electron/positron defines the $\phi = 0$-direction. The "current hemisphere" $\theta > \pi/2$ then contains the outgoing quark. In the naïve quark parton model (QPM) [23], the incoming and outgoing quarks have momenta $\frac{-Q}{2}$, and $\frac{Q}{2}$ respectively, and thus p_{parton} is $\frac{Q}{2}$. The "target hemisphere" contains the scattered electron/positron and proton remnant. The fragmentation hadrons have boosted momenta and are thus claimed to be easily separated from background [1, 6].

The definition of x_p used is given in Equation 5.4.

$$x_{p,DIS,Breitframe} \approxeq \frac{p_{hadron}}{p_{parton}} = \frac{p_{hadron}}{\frac{Q}{2}} \tag{5.4}$$

Equations 5.2 and 5.4 are thus the same quantity.

The *DIS* experiments allowed a test of the validity of a different variable to be used to represent the hard scale. In e^+e^- annihilation the centre of mass energy is chosen as the appropriate variable, while in *DIS* the momentum transfered to the struck quark, Q, was used. This was known very accurately through accurate measurement of the beam and recoil electron. Note that it is an approximation that the scattering involves a single parton which is struck and ejected. Vector boson fusion, for example, also contributes at low Q.

In hadronic collisions an alternative method must be used to define and measure signal, as will be described in Chapter 7. In such experiments p_{parton} is not measured so easily or accurately as in e^+e^- and *DIS*, and so accurate measurement of the fragmentation function relies on the ability of a parton/jet reconstruction algorithm being able to obtain a good approximation for the parton. Then, in addition the resolution must be well known such that unfolding with very high statistics may compensate for the lack of precision.

In this study the same definition of x_p will be used as in e^+e^- and *DIS* for the same comparison reasons above. The "resolution" associated with using the approximation with respect to the Lorentz invariant form is shown in Figures 5.2(a) and 5.2(b), with the parton momentum resolution superimposed for comparison purposes. The parton momentum resolution is that associated with using the TRAPS algorithm [31] to provide the input instead of the truth supplied value, as will be described in Chapter 9. The resolutions are the difference, dx_p, divided by the sum, $\sum x_p$.

The asymmetry in the "resolution" definition at low x_p is consistent with being due to the approximation being worse for lower momentum hadrons. The resolution on the parton is bad enough that the definition resolution is insignificant. The only exception being possibly at low hadron *and* high parton momentum, where low momentum tracks are emitted at large angles to the parton and the rest mass of hadrons may not be insignificant.

5.1 Transverse Fragmentation

Longitudinal fragmentation may be measured if there is a well defined direction associated with fragmenting object, as is the case in *jet* fragmentation studies. For a hadron of momentum, p_{hadron}, at opening angle χ to a jet of momentum p_{jet}, the longitudinal fractional momentum, x_L, is then given by;

$$x_L = \frac{p_{hadron} \cdot p_{jet}}{|p_{jet}|^2} \tag{5.5}$$

$$x_L = \frac{|p_{hadron}||p_{jet}| \cos\chi}{|p_{jet}|^2} \approx \frac{|p_{hadron}|}{|p_{jet}|}. \tag{5.6}$$

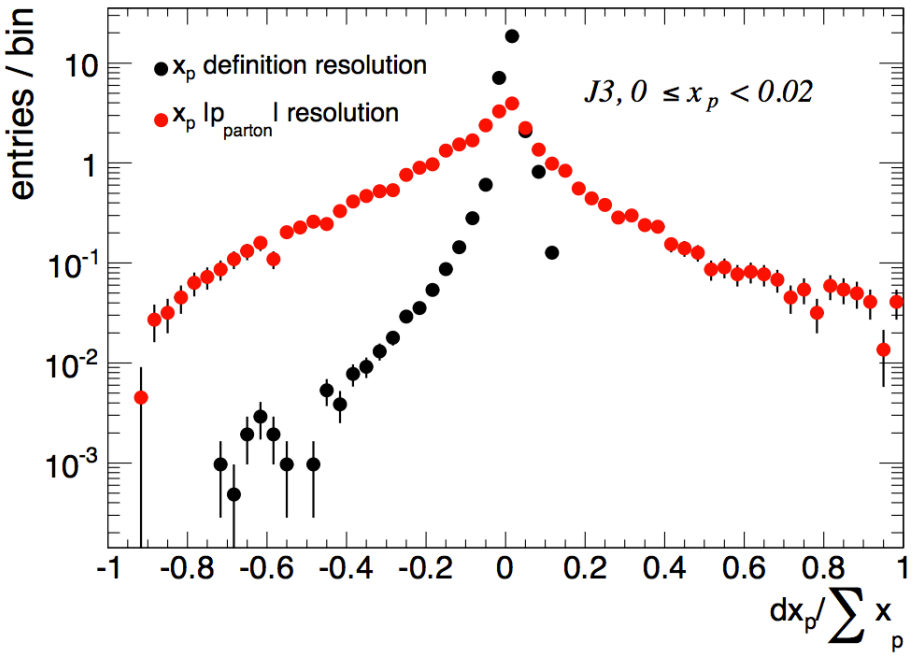
A longitudinal fragmentation measurement is less appropriate using the FAPS method since the direction of the fragmenting object is not *defined* as it is in jet measurements.

The transverse fragmentation function provides an independent measurement of fragmentation in the transverse direction to the jet. The transverse momentum fraction, x_T, of a hadron within a jet is defined as;

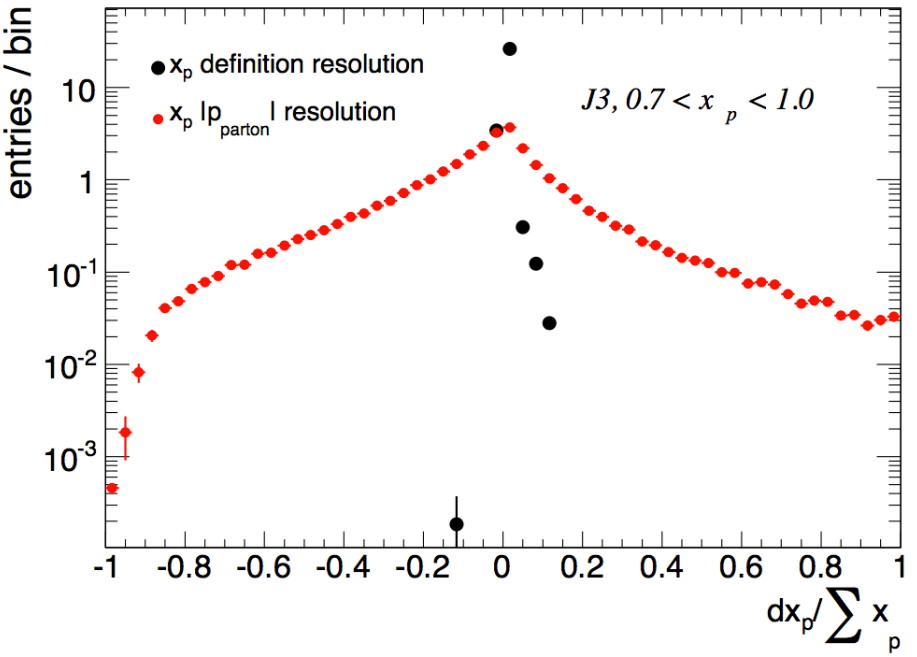
$$x_T = \frac{p_{hadron} \times p_{jet}}{|p_{jet}|^2} \tag{5.7}$$

$$x_T = \frac{|p_{hadron}||p_{jet}| \sin\chi}{|p_{jet}|^2} \approx \frac{|p_{hadron}|\chi}{|p_{jet}|}. \tag{5.8}$$

A transverse fragmentation measurement is in principle possible with the FAPS method, though significant resolution unfolding would be required, generating much larger errors than those to be described in Section 9.4. In any case, it should be noted that there is a correlation of x_T with x_p (e.g. high x_p tracks have smaller opening angles), and so without also making a longitudinal measurement, a transverse fragmentation measurement would be less complementary.

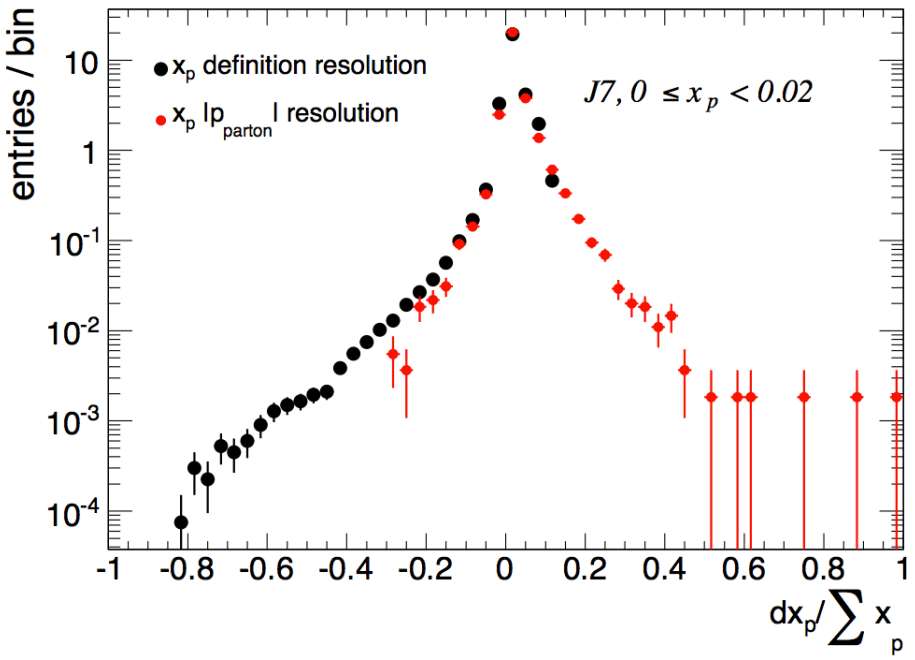


(a)

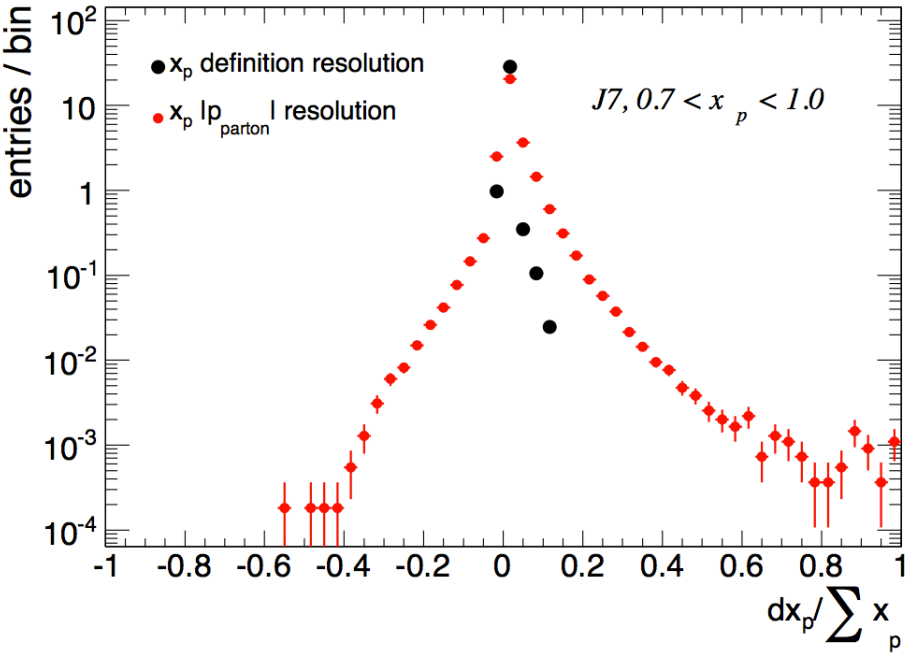


(b)

Figure 5.2: The resolution asymmetry of the x_p approximation used in this study with respect to a Lorentz invariant definition (black) at low x_p (a) and high x_p (b). The resolution associated with using TRAPS to supply the parton momentum with respect to the truth supplied value is superimposed (red). The resolutions have been calculated using Pythia J3 Monte Carlo.



(a)



(b)

Figure 5.3: The "resolution" of the x_p approximation used in this study with respect to a Lorentz invariant definition (black). The resolution associated with using TRAPS to supply the parton momentum with respect to the truth supplied value is superimposed (red). The resolutions have been calculated using Pythia J7 Monte Carlo.

Chapter 6

Parton Fragmentation

6.1 Introduction

Fragmentation is a part of the theory of strong interactions, QCD [32]. To apply QCD to hadron collisions we rely on the factorisation theorem; that we may effectively separate long and short distant interactions between fundamental particles. At short distance or large momentum transfer squared (Q^2) the partons interact in hard scatters, described by perturbative QCD and at long distances, well "before" and "after" the hard scatter, non-perturbative effects dominate [33].

The total inclusive cross section for proton + proton $\rightarrow$ hadrons of type X, may be calculated by factorising in the following way [32];

$$\sigma(Q^2)_{pp\rightarrow X} = \sum_{ijk} \int f_i(x_1,\mu^2) f_j(x_2,\mu^2) \hat{\sigma}_{ij\rightarrow k}(x_1,x_2,z,Q^2,\alpha_s(\mu^2),\mu^2) D_k^X(x_p,\mu^2) dx_1 dx_2 dx_p. \tag{6.1}$$

The structure functions $f(x_i,\mu^2)$ determine the momentum fraction, x, of the parent hadrons the interacting partons have, at renormalisation/factorisation scale μ, and the convolution of the two gives the parton luminosity. The hard parton inclusive scattering cross section, $\hat{\sigma}_{ij\rightarrow k}$, may in principle be calculated to all orders of the strong coupling constant, $\alpha_s(\mu^2)$, using perturbation theory which describes partons interacting at short distances, i.e. of the order $1/Q$ [32]. The final state X, which may denote hadrons or even jets, includes a transition from perturbative hard (coloured) partons to (colourless) hadrons. This may be achieved with the experimentally known fragmentation function, $D_k^X(x_p,\mu^2)$. For such processes Monte Carlo based showering algorithms are crucial tools.

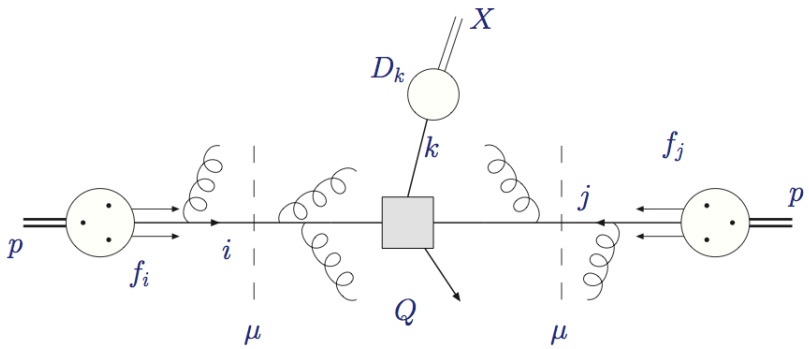


Figure 6.1: The physics describing the short (between dashed lines) and long distance is factorised. [32]

In Figure 6.1 the part between the inner dashed lines (k, short distance parton branching) is, in principle, calculable in perturbation theory as long as the soft infrared and collinear gluon emission divergences are avoided. That outside the lines, including the fragmentation function, is not and so must be extracted from data measurements. Once measured or fit to data at some scale Q, the fragmentation function may then (at least for $x_p \geq 0.1$, since lower than this non perturbative effects are significant) be evolved to larger scales using the DGLAP (Dokshitzer-Gribov-Lipatov-Altarelli-Parisi) QCD evolution equations [34]. There are different models of the non-perturbative hadronisation process, as described in Chapter 3, which all require a specified cut off scale at which point the hadronisation process takes place. This is usually the same scale as is chosen to stop parton branching becoming infinite and is about 1 GeV.

6.2 Single Parton Fragmentation

There is a philosophical problem with the concept of measuring the fragmentation of an individual hard scattered coloured parton to colourless hadrons since conservation of colour means that, in a given event there must be one hadron which is associated with at least *two* scattered partons.

In this study, no claim is made that a single "free" coloured parton results solely in colourless hadrons, however an assumption is made which grants freedom to try to measure parton properties. The assumption is that the properties of an *effectively* free parton may be indirectly measured through the hadrons strongly directionally correlated with it.

If there were a hadron associated with both hard scattered partons, or their emissions, it would be between the two in momentum space and thus have a low x_p value. Given that two partons would be measured per event in the present study at the data level, an ambiguous hadron would likely be included even if associated with the *wrong* parton. Incorrect inclusion or exclusion of this ambiguous *single* low x_p hadron, in a bin with large multiplicity, would not alter the measured values beyond assigned statistical, let alone other, uncertainties.

Earlier e^+e^- experiments measured di-parton fragmentation simultaneously, circumventing this issue. In *DIS* experiments, something purporting to be single parton fragmentation was measured using one hemisphere of the Breit frame.

As already stated, there is no theory of QCD at low momentum, and so the fact that the FAPS measurement may be philosophically uncertain here is no surprise. Even if the detailed theory of the measurement is incomplete, the result of measuring is certainly worthwhile since it may aid theoretical progress, as occurred in *DIS* with the substantiated claim for "quark universality".

6.3 Definition of a Parton Fragmentation Function

The fragmentation function, $D_i^{h^\pm}(x_p, Q, R)$, gives the average hadron of type $h^\pm$ multiplicity, at a given hard scale, Q, due to a fragmenting parton of type i (where i = u, d, s..., g), as a function of x_p, the momentum fraction the hadron takes of the fragmenting parton, within the radius, R, of a sampling cone used to measure it, i.e.

$$D_i^{h^\pm}(x_p, Q, R) = \frac{1}{\sigma_i}\frac{d\sigma^{h^\pm}}{dx_p}. \tag{6.2}$$

Previous parton fragmentation measurements (described in Section 6.4) have not included an R-dependence. In the measurement of D as a function of R in the hadron collision environment, with subtraction of uncorrelated background, there is an expectation of a limiting plateau in the amount of correlated signal measured.

It will be shown (Section 6.4) that dependence on Q is small, and R dependence is removed by techniques of variable cone sampling, extrapolation, and the choice of an "optimum" solution from compatible solutions at large R (Section 7.8).

The inclusive cross section, σ_i, to produce partons of a given type i, is $\sigma_i = 2 \times$

$N_{event} \times \frac{1}{\mathcal{L}}$, where $\mathcal{L}$ is the luminosity, and the inclusive hadron cross section for any charged hadrons is $\sigma^{h^\pm} = n^{h\pm} \times \frac{1}{\mathcal{L}}$, where $n^{h\pm}$ is the charged hadron multiplicity in the same events. This means that the luminosity $\mathcal{L}$ cancels, and $D_i^{h^\pm}(x_p)$ may be written independently of luminosity as

$$D_i^{h^\pm}(x_p) = \frac{1}{N_i} \frac{dn^{h^\pm}}{dx_p}, \tag{6.3}$$

where the superscript $h^\pm$ on $D(x_p)$ indicates the intention to sum over all charged hadrons and N_i is the number of fragmenting partons.

In principle, measurements can be made in which $h^\pm$ is decomposed to its constituents, e.g. K^+, π^-..., in which case the fragmentation function corresponding to parton (of type i) fragmenting to, for example, specifically K^+ may be selected. This type of measurement is in general difficult, since it involves hadron identification to measure any particular mode. This study, like many past fragmentation studies, first takes the summation over all hadron types to measure the fragmentation function of a parton fragmenting to any charged hadron. Later, the hadron fragmentation function is corrected using Monte Carlo to give the pion fragmentation function.

Note from the definition that the fragmentation functions for a given type of parton to specific hadron species, e.g. pions and kaons, may be simply added in the following way

$$D_i^{\pi^\pm + K^\pm} = D_i^{\pi^\pm} + D_i^{K^\pm}, \tag{6.4}$$

however, in order to add fragmentation functions for different partons, e.g. quarks and gluons, the number of each parton type must be accounted for, i.e.

$$(N_{quark} + N_{gluon}) D_{quark+gluon}^{h^\pm} = N_{quark} * D_{quark}^{h^\pm} + N_{gluon} * D_{gluon}^{h^\pm}. \tag{6.5}$$

In this feasibility study quarks are selected, using truth information, as described in Section 9. This enables comparison to previous *quark* fragmentation data, providing a benchmark for the algorithm to test whether the calculated fragmentation function is in agreement, as it should be since that data was used to tune the Monte Carlo models used. Of course, further testing of the algorithm with quark and gluon samples would be performed, in preparation for a data measurement, since at present it is not possible to

select a sample of light quark jets.

In order to begin this study a few arbitrary choices are made which are tested and verified later. A unit radius cone, or "unit cone", (in η, ϕ) around the parton may be used as an initial example. Using this naïve method the resulting fragmentation function is

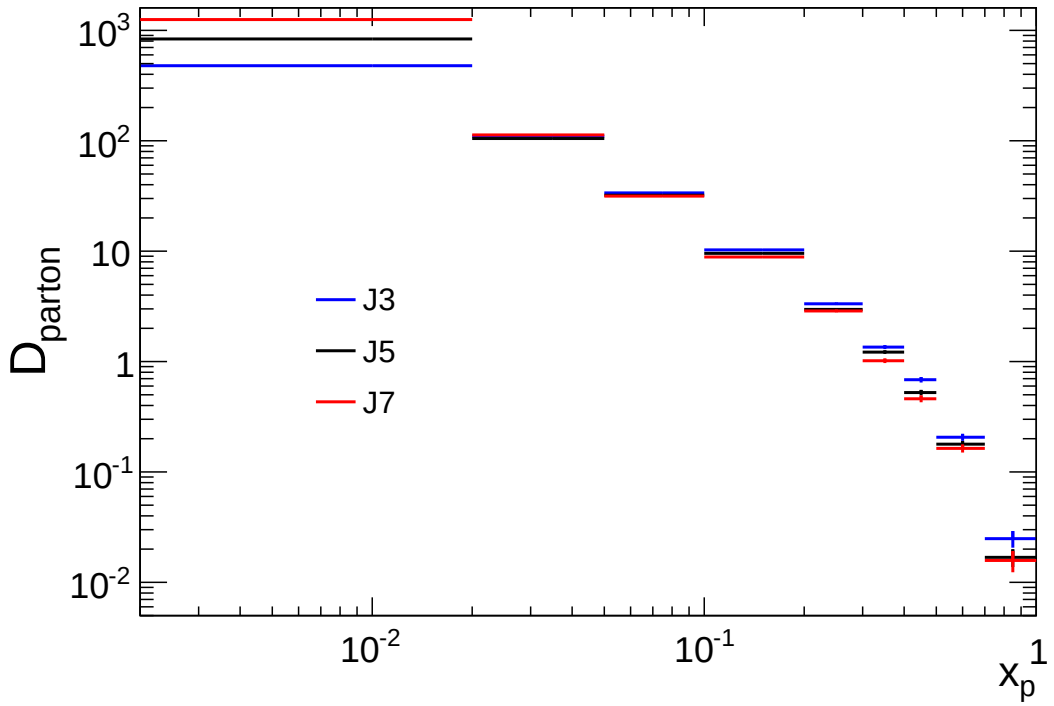


Figure 6.2: The fragmentation function obtained with a naïve unit cone, calculated with J3, J5, and J7 Pythia Monte Carlo samples.

shown in Figure 6.2. It is a "soft" function, i.e. a fragmenting parton will fragment into *many* hadrons, most with a small momentum fraction of the parton. Very rarely will it fragment into a hadron taking a large fraction of its momentum. Variable width binning is used to compensate for lower average multiplicities in the higher x_p bins.

There are four orders of magnitude between the average multiplicity in the first and last x_p ranges. There is approximate scaling of the fragmentation function with the hard scale of the Monte Carlo samples, i.e. the multiplicity is approximately only varying as a function of x_p. A closer look shows scaling violation, i.e. the fragmentation function is not only varying with x_p but also as a function of hard scale, becoming "softer" at higher hard scales as expected, since the hard scale enables more hadrons to be seen in each parton, though at this stage background levels are not accounted for.

6.4 Previous Measurements

Previously, detailed measurements of fragmentation functions have been made using electron-positron (e^+e^-) annihilation [2, 3, 4, 5] and deep inelastic scattering (DIS) ep data [1, 6]. The results of these studies are shown in Figure 6.3. The same binning scheme has been used by previous experiments, and will be used in this study, to allow comparison of results.

In the figure $D(x_p, Q)$ is shown for quarks as a function of the hard scale of the interaction which produced them, each of the nine sub-figures shows the multiplicity per quark for the specified x_p range.

Notice the abscissa has two labels, Q (transferred momentum) and E^* (the centre-of-mass energy). As discussed in Section 2.3 in DIS the momentum transferred from the photon to the quark, Q, is the best measure of the hard scale, while in e^+e^- annihilation, all incoming energy is converted to make the outgoing quark-antiquark ($q\bar{q}$) pair and so this energy is the best measure of the hard scale. The momentum of the scattered electron was accurately measured in the DIS experiments, and the beam energy was well known in e^+e^-, and so the hard scale was accurately known in both of these experiments.

There is only DIS data in the first x_p interval. In e^+e^- experiments, as in any particle collision experiment, there is a difficulty in measuring the lowest momentum tracks and therefore the fragmentation function in the lowest x_p range, due to their low momenta being insufficient to escape the beam pipe or at least be well-measured after doing so. However, the DIS experiments were able to measure the fragmentation function in this low x_p range by using the Breit frame, in which the boost and rotation sometimes gives tracks of high momentum in the laboratory frame, low momenta, and thus low x_p values, in the Breit frame.

Looking at the different ordinates of all nine sub-figures the same four orders of magnitude variation of the data are seen, however, use of a linear range for each bin in x_p shows the hard scale dependence, or scaling violations more clearly. The use of a "scaled variable" (x_p) is noted. Since this quantity is dimensionless D "should" scale.

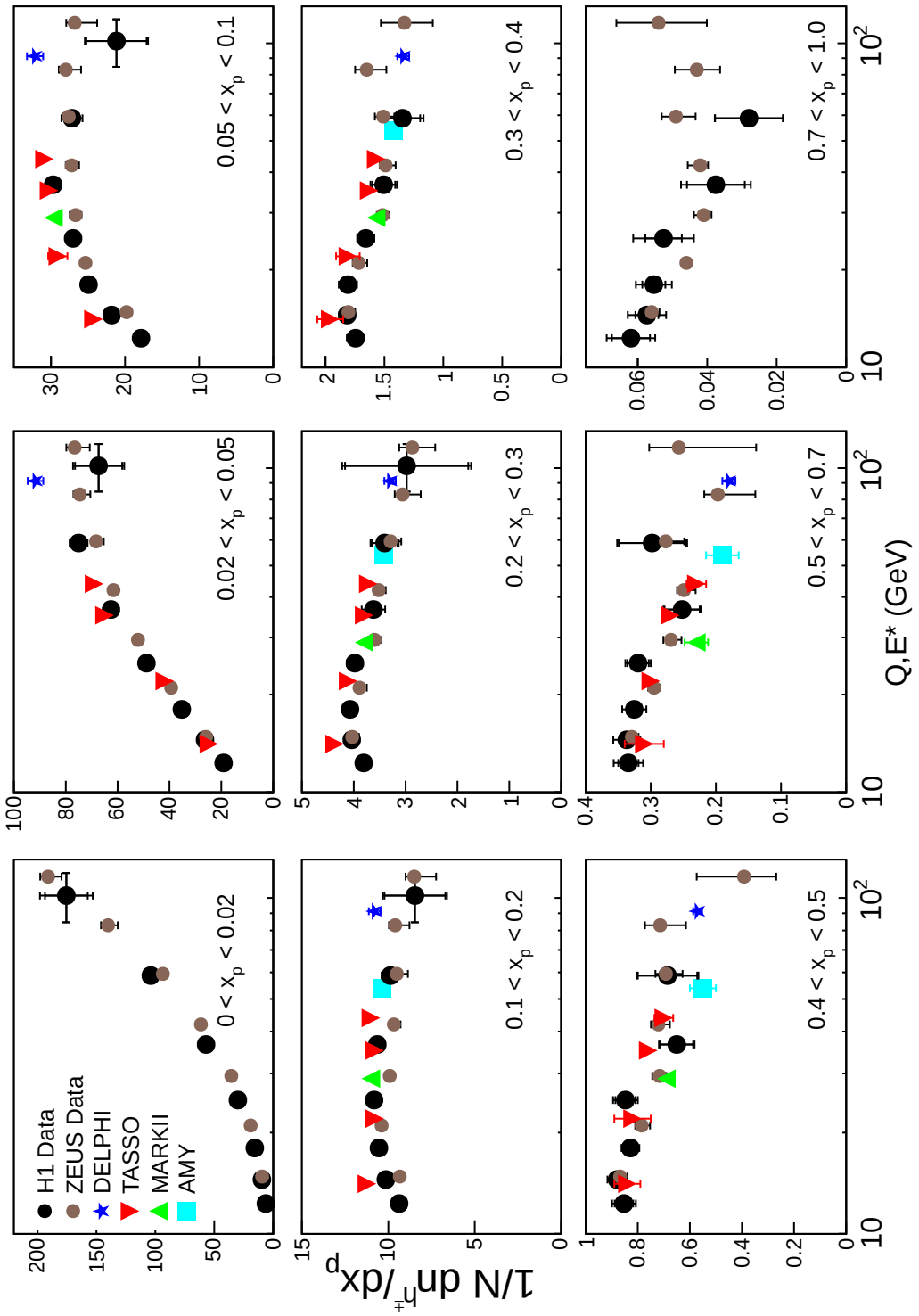


Figure 6.3: Results from previous fragmentation studies, obtained from e^+e^- (DELPHI [35], TASSO [4], MARKII [36] and AMY [37] and DIS (H1 [38] and ZEUS [39]) experiments.

Note the agreement of fragmentation functions of the quarks produced "out of the vacuum" in s channel e^+e^- annihilation and quarks knocked out of protons by photons in t channel DIS scattering experiments. This agreement makes a statement about the quarks produced in these two processes, that the objects produced are in fact the same objects, in that they fragment (on average) in the same way. This is evidence for the "universal nature" of a common concept of a quark, i.e. that quarks have transferable behaviour (properties) and thus can be said to exist as definite objects.

Having commented on the agreement observed, it should be said that Figure 6.3 also displays some disagreement between e^+e^- and DIS data, perhaps best demonstrated in the second, third and fourth x_p range sub-figures. The use of the Breit-frame, which might not include the decay products of a very off-mass-shell quark, is a possible cause of such a difference.

The e^+e^- experiment was limited in its ability to alter the hard scale and in the DIS experiment there was a limited cross section, and the Breit-frame was only a hemisphere. A measurement with proton-proton collisions, which can provide huge statistics due to the strong coupling strength α_s, may enable comment on the disagreement of previous data. However, there is a quark/gluon ambiguity which would rely on (reliable) cross-section calculations to provide gluon fragmentation measurements.

Chapter 7

Parton Fragmentation Algorithm

7.1 Method Philosophy

The fragmentation function of the parton is different in concept from that of a jet. Jets are well defined objects, according to various algorithms with fixed cones and/or recombination methods and arbitrary "seeding" parameters. Nevertheless a measurement pertaining to the actual partons themselves allows fundamental statements about nature.

In order to measure parton fragmentation a radius independent strategy is employed which aims to measure a signal which is itself *independent*. This is defined to be those charged final state hadrons correlated with the parton direction after subtraction of uncorrelated background. The former are referred to as FSB, and the latter as ISB and/or underlying event, both for practical purposes and since they are referred to in that way within the Monte Carlo models being used for correction and test purposes.

7.2 Algorithm Overview

The parton position in this method is provided either by a *truth* scattered parton or by some jet finding algorithm which provides *parton estimators*. For this study *truth* is taken to be the most perfect position a jet finding algorithm could provide. The fragmentation function calculated with the parton direction provided by a jet algorithm and truth may be compared as a quality touchstone of the algorithm. Note that none of the previous quark fragmentation measurements, described in Section 6.4, utilised a jet algorithm to provide partonic momentum estimates.

Resolution on the x_p variable, due to algorithms used to associate tracks and estimate partons, and due to the intrinsic resolution of the detector, is unfolded in a single step, as will be described in Sections 9.4 and 11.3.

To measure the signal, charged hadrons in a given x_p interval are counted in cones of various radii centred on the (estimated) parton position. For a meaningful measurement rather than a qualitative view, it is necessary to consider the background underneath the parton. The background is sampled using a solid angle at some large angle to the parton and to the beam. A square solid angle is used instead of a cone, since there is not expected to be a centre of activity as there is for measuring partons. The measured background is scaled by the ratio of cone to sampling solid angle and then subtracted from the total leaving a measurement of the signal, which still has radial dependence.

Once all events have been accumulated an exponential function is calculated at each radius utilising the two neighbouring points. As the radius increases statistical error decreases whilst background subtraction, and hence its error, increases. Extrapolation error, assessed as the total from each point to the calculated limiting value, also decreases. Thus an array of solutions ($\sim$6 in practice) is available from which that with the minimum error is selected. In this way, directional resolution of the parton estimators is resolved. Such a choice is justified given the compatibility of all extrapolated solutions, which in turn demonstrates the independence of the final measured fragmentation function from a radius parameter.

The method has been developed using the five Monte Carlo samples described in Section 3.3 in order to test the feasibility of measuring the partonic fragmentation function over a realistic order of magnitude of hard scale given integrated luminosity expectations.

The only selection criteria used are the requirements that the outgoing partons are in the range $|\eta_{parton}| < 1.0$, such that a large measuring cone would still be inside the ATLAS tracking fiducial volume, and are sufficiently well separated such that fragmentation sampling cones will not overlap other jets.

7.3 Background Subtraction

In order to find a location that is far from partons, $\pm\frac{\pi}{2}$ from the parton in consideration is chosen for the azimuthal location, since, in general, omitting ISB perturbations, scattered partons are back-to-back in azimuth and so the farthest possible location from any parton, in a two parton event, is here. Both ϕ locations are sampled for each parton. At these locations it is expected that, for all practical purposes, there are only background hadrons.

The η location which is far from scattered partons in a two parton event is not such a simple location to define, since when considering one of the partons, no such correlation exists in the laboratory reference frame, as with the azimuthal angle, of the other parton. For this reason a classification of two parton events is made, such that all events fall into one of four classes.

The four event topology categories are defined based on (a) whether both partons point in the same z direction or not, and (b) how close the partons are to either zero (for same z direction partons), or the mean η position of the two partons (for opposite z direction partons).

For partons pointing in the same z direction, the background location is chosen to be at either $(\eta,\phi)_1 = (-\eta_{parton}, \pm\frac{\pi}{2})$ for partons which are far from zero, or in the opposite z direction to the parton, at $(\eta,\phi)_2 = (\frac{\eta_{max}}{2}, \pm\frac{\pi}{2})$ where η_{max} is the maximum η value for which partons are accepted in this study.

For partons on opposite sides of the detector the background locations are chosen to be at either $(\eta,\phi)_3 = (\eta_{mean}, \pm\frac{\pi}{2})$, where η_{mean} is the mean η position of the two partons, when the parton is far from η_{mean}, or in the opposite z direction to the parton being measured at $(\eta,\phi)_4 = (\frac{\eta_{max}}{2}, \pm\frac{\pi}{2})$ otherwise.

This is summarised in Table 7.3.

Event type	Location 1	Location 2	Location 3	Location 4
η_{parton}	Partons on same side		Partons on opposite sides	
	Far from zero	Near zero	Far from mean	Near mean
$\eta_{background}$	$-\eta_{paron}$	$-sgn(\eta_{parton}) \times \frac{\eta_{max}}{2}$	η_{mean}	$-sgn(\eta_{parton}) \times \frac{\eta_{max}}{2}$

Table 7.1: η_{parton} is the η value of the parton of which the background is being measured. $\eta_{background}$ is the η location chosen to sample that background.

This procedure was designed in order to keep the $|\eta|$ for the background the same as for the parton as far is possible without compromising picking up signal from the other parton.

The background fragmentation function was measured in each of these locations, to test that each location gives roughly the same measurement of ISB/MI background within expected variation with η. Figure 7.1 shows the background measured in each of the topology specified locations. Each colour represents only the background measured in that location for events with topology corresponding to the location.

The multiplicities measured in each solid angle are of the same order of magnitude given the proportion of background to signal.

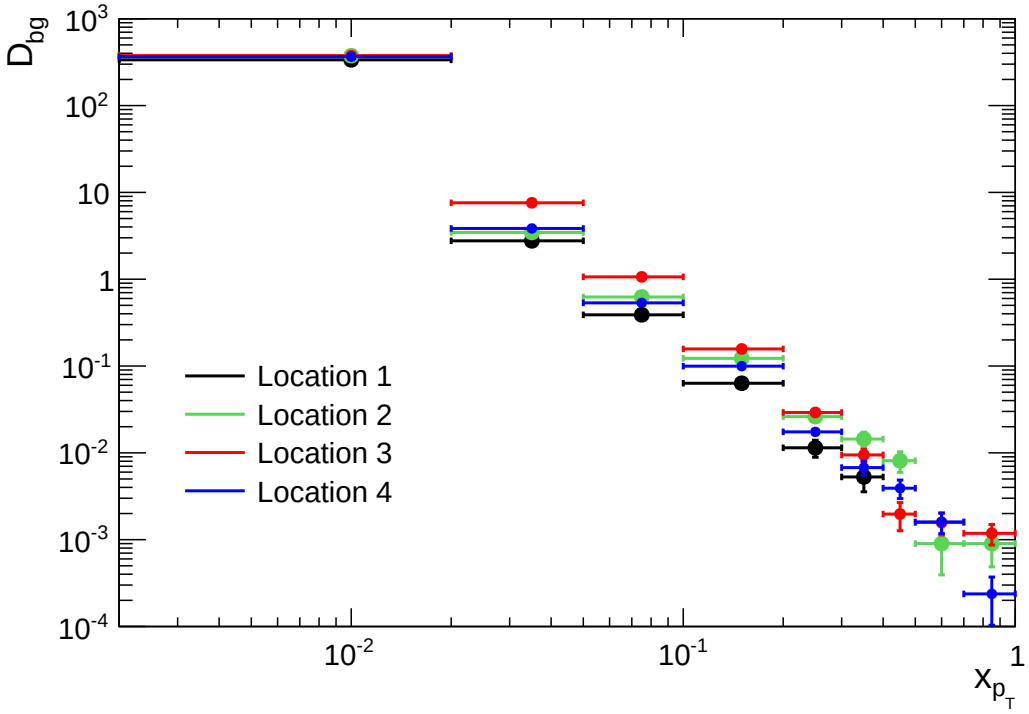


Figure 7.1: The background fragmentation function measured using a unit square solid angle in each of the four (eight considering $+\pi$ and $-\pi$ in azimuth) topology-specified background locations defined in Table 7.3. Measured with Pythia (J5) Monte Carlo at the truth level.

7.4 Background Proportions in Measuring Cones

When a cone is used to measure the fragmentation function, naturally some background will be included in the cone. The ratio of signal to background calculated using a cone of given radius is presented in Figure 7.2. The ratio is consistently higher for the lower p_T MC

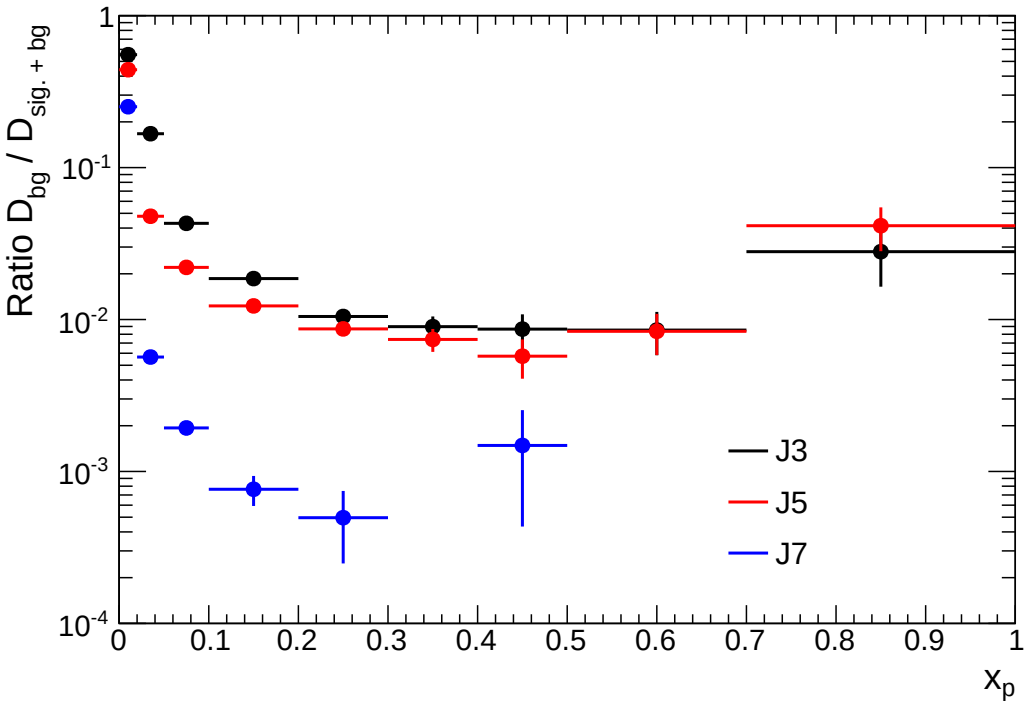


Figure 7.2: The ratio of background to total hadrons measured using a cone of unit radius. Measured with Pythia J3, J5 and J7 Monte Carlo.

samples in all x_p bins. In the range $0< x_p <0.02$, and the ratio $\frac{background}{total}$ is greater than one half at low momentum (J3), therefore any method which does not include background subtraction may make mistakes of more than 100% here.

Smaller sampling cones are sufficient to encompass the higher x_p signal and so a unit cone would be expected to have a higher noise/total ratio for higher x_p, as shown in Figure 7.2.

7.5 Background Solid Angle Size

The background sampling solid angle is chosen with a compromise between the solid angle being large enough that good statistics are measured to provide an accurate background measurement with minimum uncertainty, but not so large such that it begins to include signal hadrons.

To check the safety of an appropriate solid angle size, the background per unit LISA was measured with different solid angles. In Figure 7.3 the measured background per unit

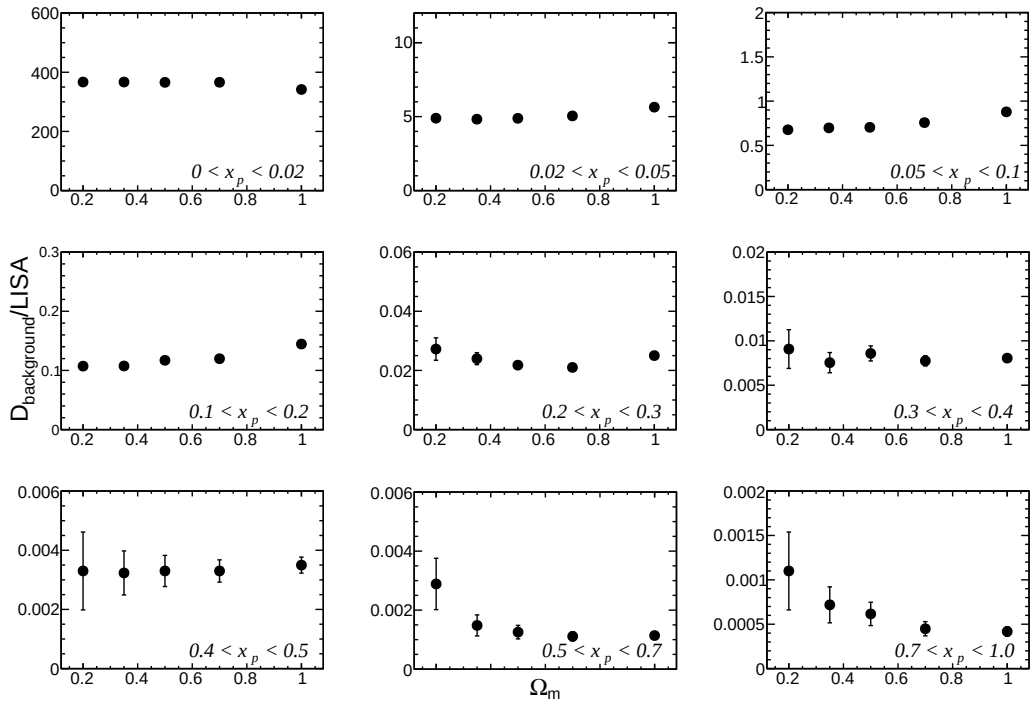


Figure 7.3: Background per unit LISA as a function of solid angle size used to measure background. Study with on J5 Pythia Monte Carlo data.

LISA is shown as a function of the measuring solid angle, again in the nine separate x_p intervals. As will be shown the background is of significant magnitude in the x_p interval $0< x_p <0.1$, and thus the first three sub-figures are of most interest. For this x_p range, the calculated background per unit LISA is independent of the measuring solid angle, Ω_m, for small values, i.e. Ω_m <0.5. No signal is measured in this range, so measuring background with a solid angle in this range is justified.

The last two sub-figures show a fluctuation by a factor of two or three with varying solid angle size. The fluctuation here can be seen to be due to limited statistics. The last sub-figure is consistent with measuring one hadron in the smallest solid angle, and not

measuring another until the largest angle is used. The background in this x_p range is not significantly large ($O(1\%)$) and so this fluctuation is not a point of concern.

7.6 Statistical Uncertainty on Signal

In discussion of the statistical uncertainty the following nomenclature is used. The total charged hadron multiplicity per parton, per unit of x_p , including background, using a measuring cone of radius $R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, is

$$D_{T,R} = \frac{1}{N_p}\frac{n_{T,R}}{\Delta x_p} \tag{7.1}$$

Once background has been subtracted the signal is written

$$D_{S,R} = \frac{1}{N_p}\frac{n_{S,R}}{\Delta x_p}. \tag{7.2}$$

Where N_p is the total number of partons used, $n_{T,R}$ is the total, not background subtracted, charged hadron multiplicity measured and $n_{S,R}$ is the corresponding, background subtracted, signal multiplicity, both within that cone of radius R.

A change in $D_{S,R}$ with radius may be written,

$$D_{S,2} = D_{S,1} + \Delta D_S \tag{7.3}$$

i.e. ΔD_S is the difference between $D_{S,1}$ and $D_{S,2}$, calculated at those two radii.

Writing $D_{S,R}$ in terms of total and background components,

$$D_{T,2} - A_2 D_B = D_{T,1} - A_1 D_B + \Delta D_S. \tag{7.4}$$

Where A_1 and A_2 (LISA) are the areas of the two measuring cones and D_B is the background density per unit LISA. Written in this way, all terms are uncorrelated with each other, this is not true if the equation is expressed in terms of $D_{S,1}$ and $D_{S,2}$ since a similar background has been subtracted from both of them. Rearranging,

$$D_{T,2} = D_{T,1} + (A_2 - A_1)D_B + \Delta D_S. \tag{7.5}$$

The errors on these quantities may now be written,

$$\delta^2 D_{T,2} = \delta^2 D_{T,1} + (A_2 - A_1)^2 \delta^2 D_B + \delta^2 \Delta D_S. \tag{7.6}$$

Where $\delta^2 D_B$ may be written in terms of the component systematic and statistical contributions,

$$\delta^2 D_B = \delta^2 D_{B_{STAT}} + \delta^2 D_{B_{SYST}}. \tag{7.7}$$

7.7 Systematic Uncertainty on Signal

There may be systematic uncertainty introduced in the measured value of D_{signal} due to the sampling and subsequent subtraction of background from D_{total}.

7.7.1 Systematic Uncertainty on Background Subtraction

In Chapter 4 on hadronic profiles, background from ISB and MI was shown to have a small η dependence. This fact is important in considering the quantitative estimation of background to a parton which is being measured. Ideally, the parton background would be sampled at the same $|\eta|$ as the parton as shown before. While an attempt is made to sample at this value, it is not always possible because of contamination from the parton in the other semi-cylinder. Sampling background in a different $|\eta|$ from that of the parton may thus introduce a systematic uncertainty, which must therefore be estimated.

7.7.2 Estimation of Systematic Uncertainty on the Background

The behaviour of D_{signal} as a function of the measuring cone radius, R, is considered in order to estimate an upper limit on the systematic uncertainty on the background. As previously stated, this curve results from measuring the total charged hadron multiplicity around a fragmenting parton, then sampling and subtracting what is believed to be background. The maximum cone radius used is 1.3, in order to avoid proximity with the other parton. With the correct amount of background sampled and subtracted, and a sufficiently large measuring cone, the cumulative D_{signal} vs R curve, should plateau. However, if the incorrect amount of background is subtracted from the measuring cone, or if an insufficiently large measuring cone is used, then a plateau may not be observed.

There are two scenarios, which are treated differently in terms of the method adopted to estimate a limit of the systematic background uncertainty. The D_{signal} vs R curve may be turning over after subtraction of unscaled background. Since the D_{signal} vs R curve is a cumulative function, then if it is turning over, this is unphysical and too much background *must* have been subtracted. If the curve is still increasing at the largest safe measuring radius, then either too little background has been subtracted or the curve may be increasing due to signal. The first method gives an estimate and a lower limit of the systematic error, the second gives only an upper limit.

At this point, it is useful to give a formal definition of the curve turning over. The difference in D_{signal}, ΔD_{signal}, measured at two different radii is considered. When the difference is negative, and the magnitude of the difference is greater than the uncertainty in the difference, $\delta(\Delta D_{signal})$, then the curve is said to turn over. The conditions, which must both be satisfied, for the curve to be turning over are thus

$$\Delta D_{Signal} < 0 \text{ and } |\Delta D_{Signal}| > \delta(\Delta D_{signal}). \tag{7.8}$$

In order to estimate a limit on the uncertainty on the background when the D_{signal} vs R curve is turning over, the amount of background which is subtracted from the measuring cone is scaled downward until the curve no longer turns over. The scaling factor, α, which just makes the curve flat is used to scale all background subtractions, in this hard scale range, shifting all values of D_{signal} after this. The D_{signal} measurement is then obtained from D_{total} and $D_{background}$ from

$$D_{signal} = D_{total} - \alpha D_{background}. \tag{7.9}$$

To be safe, a systematic uncertainty error bar, $\delta_{syst} D_{signal}$, of the same magnitude as the correction which has been made, is applied to the D_{signal} data points. The systematic uncertainty is thus expressed as

$$\delta_{syst}(D_{background}) = |\alpha - 1| D_{background}. \tag{7.10}$$

In the case where the D_{signal} vs R curve is flat or still increasing at the largest safe radius, the amount of background is scaled upward until the curve turns over. The minimum value

of the scaling factor, α, which makes the curve turn over is then used to calculate the maximum systematic uncertainty, as in the above equation. However, no correction is made to the amount of background to be subtracted since the possibility of signal causing the rising curve is still very possible.

The estimated systematic uncertainties in the background subtraction procedure are summarised in Table 7.2 and 7.3. Exceptionally, the estimation was performed at a lower value of hard scale (using J2 Monte Carlo, of $\hat{p_T}$-range $35 - 70$GeV), where excessive background would preclude analysis. It increases with increasing x_p, due to the lack of

x_p range	Scaling Correction Factor	$\lvert 1-\alpha \rvert$	$\delta_{syst}(D_{background})/D_{total}$
0→0.02	0.89	0.11	8%
0.02→0.05	1.0	<0.03	<11%
0.05→0.1	1.0	<0.69	<10%
0.1→0.2	1.0	<2.18	<9%
0.2→0.3	1.0	<2.61	<5%
0.3→0.4	1.0	<2.31	<3%
0.4→0.5	1.0	<2.02	<2%
0.5→0.7	1.0	<2.18	<3%
0.7→1.0	1.0	<0.65	<2%

Table 7.2: The systematic uncertainty estimated limits as a function of x_p range from the J2 Monte Carlo sample.

Monte Carlo Sample	Scaling Correction Factor	$\lvert 1-\alpha \rvert$	$\delta_{syst}(D_{background})/D_{total}$
J3	0.96	<0.04	2%
J4	1.0	<0.08	4%
J5	0.99	<0.007	<1%
J6	1.0	<0.007	<1%
J7	1.0	<0.16	4%

Table 7.3: The systematic uncertainty estimated limits, estimated at the lowest x_p interval with Pythia Monte Carlo at the truth level for the samples used in later analysis.

background available to constrain the uncertainty, which decreases rapidly with increasing x_p. The less-than symbols are included to highlight that the estimates are expected to be generous for this reason, and that the uncertainties at larger x_p are believed to be compatible with that at the lowest x_p. Since the amount of background present at higher x_p is minimal, a relatively large uncertainty makes little contribution to the error on D_{signal}. A similar reasoning is applied to the evolution with p_T of the estimated limit of the background systematic uncertainty.

The *systematic* uncertainty in the background is assumed to be independent of x_p , and since the majority of background is of low x_p and the best limit is expected to be

obtained here, it is estimated, generally, in this interval and applied to all others. While this study is presented at the truth level, to ensure relevant estimation and correction the same procedure is repeated later at the reconstructed level.

7.8 Extrapolation and Optimum Solution

In the case where the largest safe measuring cone is not sufficiently large to encompass all of the signal, extrapolation may be used to estimate the missing signal (Figure 7.4). The

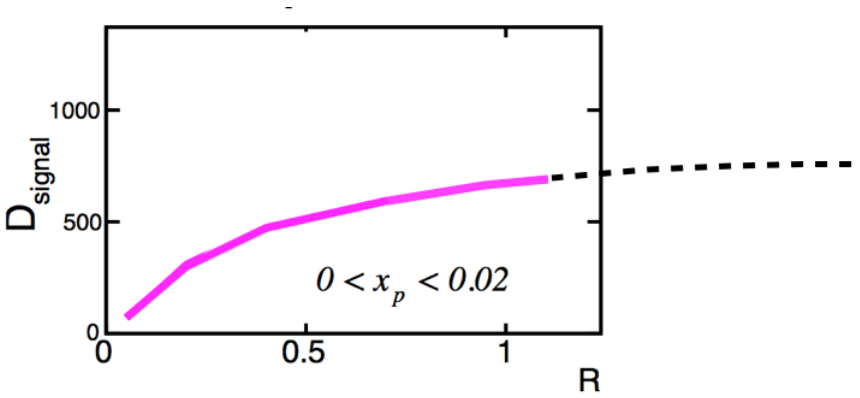


Figure 7.4: A cartoon extrapolation of a D_{signal} vs R curve which has not reached plateau with the largest safe radius.

method utilises multiple D_{signal} extrapolations for each R value within a given (x_p, Q) range. An extrapolation from a small value of R has a large error from the distance to be extrapolated but a small error from the lack of background. The reverse is true for an extrapolation at large R so from this variety of solutions the one with the smallest total error is selected as being "optimum".

7.8.1 Extrapolation of D_{signal} vs R Curves

The probability of a bremsstrahlung with transverse momentum k_t with respect to the original parton direction, has an approximate exponential dependence such that small values are common and large values are rare. This approximate relationship may be expressed as in Equation 7.11

$$P_{brem.}(k_t) = P_{brem.}^{k_t=0} e^{-bk_t} \tag{7.11}$$

where $P_{brem.}(k_t)$ is the probability of bremsstrahlung with some k_t value, $P_{brem.}^{k_t=0}$ is the probability of bremsstrahlung with $k_t = 0$, where b, with units of GeV^{-1}, relates to the rate of decrease in expectation with increasing k_t.

Given the relationship between k_t and R discussed in Chapter 4, the D_{signal} vs R curves are then expected to be reasonably well described by the following cumulative function.

$$y = D(1 - e^{-aR}) \tag{7.12}$$

D is the value of the extrapolation in the limit $R \to \infty$ and a gives the rate at which the function is approaching the limit. The value, y_i, of the function at radius R_i, and the gradient at R_i is approximated as $(y_{i+1} - y_i)/(R_{i+1} - R_i)$. To find the unknown variables the differential of the exponential function is used, equation 7.13, as illustrated in Figure 7.5.

$$\frac{dy}{dR} = Dae^{-aR} \tag{7.13}$$

Equation 7.13 gives a value of the product Da at $R = 0$ as in equation 7.14.

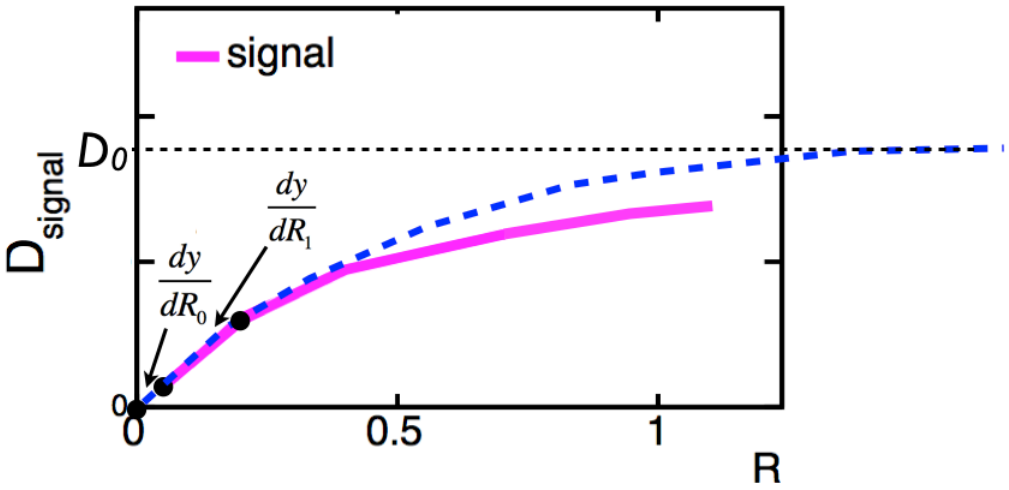


Figure 7.5: A cartoon illustrating an exponential extrapolation from $R_1 = 0.2$.

$$\frac{dy_0}{dR_0} = D_0 a_0 \tag{7.14}$$

The subscripts in equation 7.14 indicate the radius used to obtain the variables. The product $D_0 a_0$ is used to find a better estimate or *updated* value for the variable a, a_1,

using the following substitution

$$\frac{dy_1}{dR_1} = D_0 a_0 e^{-a_1 R_1}. \tag{7.15}$$

The assumption has been made that the product aD does not vary significantly with radius, justifying the use of $D_0 a_0$ in the calculation of a_1. Substitution of a_1 into the original exponential function yields D_1;

$$D_1 = \frac{y_1}{1 - e^{-a_1 R_1}} \tag{7.16}$$

where y_1 is set to be the value of D_{signal} at R_1. For safety an "extrapolation" error of the magnitude of the total extrapolation is placed on D, i.e. $\delta D_i = (D_i - y_i)$.

The process to find a_1 and D_1 is repeated to find further updated values of a and D for increasing radii up to $R = 1.15$. Multiple compatible D solutions result from which an *optimal* solution which has the smallest total error is then quoted as $D_q(x_p, Q)$.

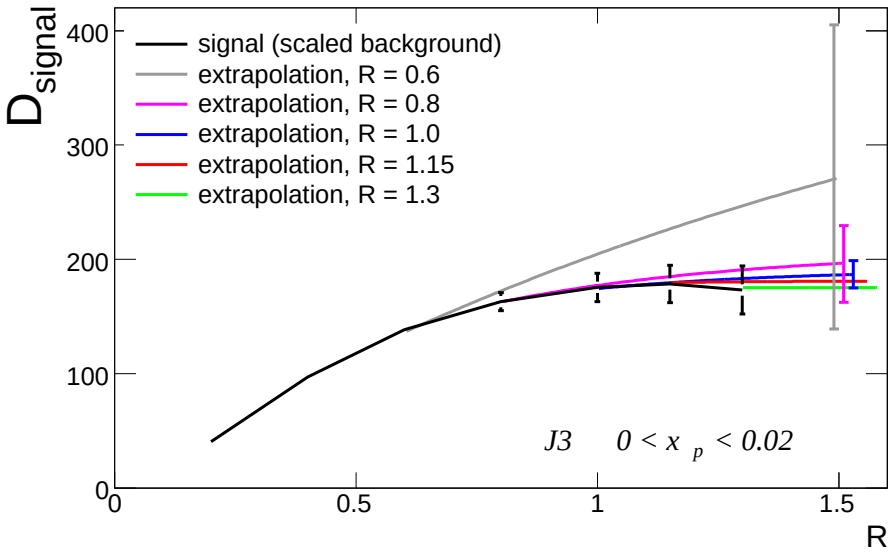


Figure 7.6: Exponential extrapolations and error bars of equal magnitude as the extrapolations for $0 \leq x_p < 0.02$ J3 Monte Carlo.

7.8.2 Optimal Solutions

The total error on an extrapolated value of D_{signal} is a summation in quadrature of the statistical, systematic and extrapolation components, $\delta_{statistical}$, $\delta_{systematic}$, and $\delta_{extrapolation}$ respectively.

$$\delta^2_{total} = \delta^2_{statistical} + \delta^2_{systematic} + \delta^2_{extrapolation} \tag{7.17}$$

The total error as a function of R is shown for one interval of x_p, at one hard scale in Table 7.4. The significant contributions to uncertainty come from systematic background uncertainty which increases with R and that due to extrapolation, which decreases with R. The smallest total error corresponds to the solution at $R = 1.15$, and this is then the optimal solution for the lowest x_p range at J3. The table excludes the columns of $R =$

R	0.6	0.8	1.0	1.15
Extrapolated D_{signal}	477.3	208.7	191.6	181.2
$\delta_{statistical}(D_{signal})$	15.1	6.7	6.8	7.0
$\delta_{systematic}(D_{signal})$	5.3	9.3	14.6	19.3
$\delta_{extrapolation}(D_{signal})$	342.1	46.1	17.1	1.5
$\delta_{total}(D_{signal})$	341.5	47.5	23.5	20.6
% Error	72%	23%	12%	11%

Table 7.4: Extrapolated D_{signal} values and corresponding uncertainties shown as a function or R for $0 \leq x_p < 0.02$ calculated using the J3 Monte Carlo sample.

0.2 and 0.4 since the method for calculating a yields a negative value. In this case the exponential extrapolation is replaced with a linear one which uses the gradient at that point with an extrapolation to $R = \frac{\pi}{2}$.

7.8.3 Extrapolation at Larger Hard Scale

So far the low $\sqrt{-t}$ and x_p scenario has been described in which significant background is present and fragmenting partons produce hadrons at large angles from the parton direction. In contrast, at large $\sqrt{-t}$ and x_p a small radius is adequate to encompass the signal. Figure 7.7 is a demonstration of the extrapolation method applied to the J7 $0.7 \leq x_p \leq 1.0$ D_{signal} vs R curve. In Figure 7.8 zoom has been used to illustrate that the method does not break down at large hard scale and still provides a variety of compatible solutions. The solutions and corresponding errors are summarised in Table 7.5.

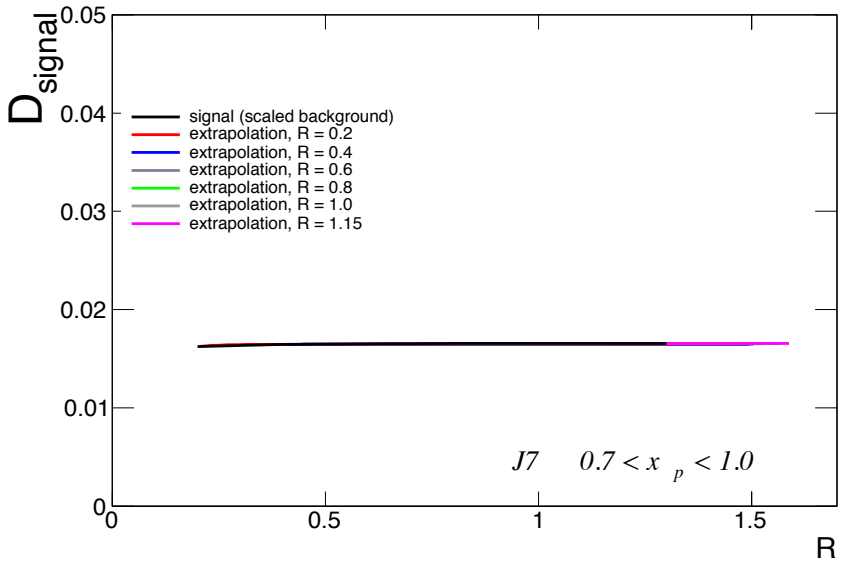


Figure 7.7: Exponential extrapolations and error bars of equal magnitude as the extrapolations for 0.7$\leq x_p <$1.0 J7 Monte Carlo.

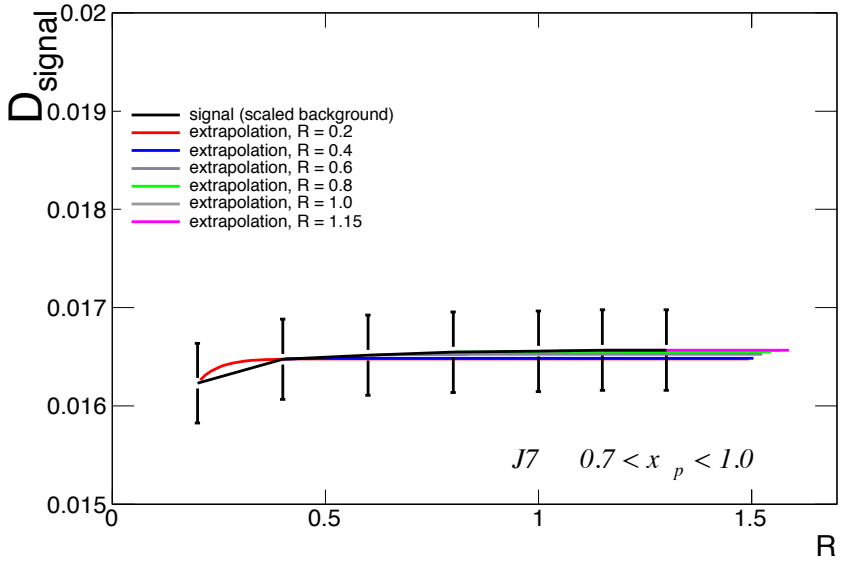


Figure 7.8: Zoomed version of Figure 7.7.

R	0.2	0.4	0.6	0.8	1.0	1.15
Extrapolated D_{signal}	0.0168	0.0165	0.0165	0.0166	0.0166	0.0166
$\delta_{statistical}(D_{signal})$	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
$\delta_{systematic}(D_{signal})$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\delta_{extrapolation}(D_{signal})$	0.0002	0.00001	0.00001	0.00000	0.00000	0.00000
$\delta_{total}(D_{signal})$	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004
% Error	3%	2%	2%	2%	2%	2%

Table 7.5: Extrapolated D_{signal} values and corresponding uncertainties shown as a function of R for $0.7 \leq x_p < 1.0$ calculated using the J7 Monte Carlo sample.

Chapter 8

FAPS Transverse Fragmentation

This chapter gives an alternative to describing the jet shape and measuring the transverse fragmentation function from the method referred to in Section 5.1. FAPS provides for a more physically based measurement, thus also enabling the task of Monte-Carlo tuning by reference to parameters closer to the models employed. This study is conducted at the truth level, and therefore truth supplied values of partonic and hadronic momenta are used as the algorithm inputs.

8.1 Jet Shape Parametrisation

The extrapolation process provides information about the transverse behaviour of the correlated hadrons. The coefficient, a, of the extrapolation function (Equation 7.12) gives a measure of how fast the cumulative $D_{signal}(x_p, Q, R)$ curves are approaching their limit with increasing R, i.e. how collimated the hadrons are in that x_p range. This then constitutes a measure of the jet "shape", as a function of fractional momenta, x_p, rather than a p_T or energy deposition profile as given in Chapter 4. Even though all local extrapolations give compatible final plateau values, the $D_{signal}(x_p, R)$ curves are only *approximately* exponential over the entire R range, and therefore the values of a in a given x_p range vary with R. Generally they are larger at smaller values of R. Since the behaviour is exponential, the inverse of the mean values of a, i.e. $1/ < a >$, may be taken as a measure of the mean value of R at a given x_p, but because of this variation, the standard error on the mean of the a values for a given x_p range is taken as typical error. The mean value of a, for a given x_p range is shown as a function of the mean value of x_p

for hadrons within that range, and for each hard scale of the Monte Carlo data sets in Figure 8.1.

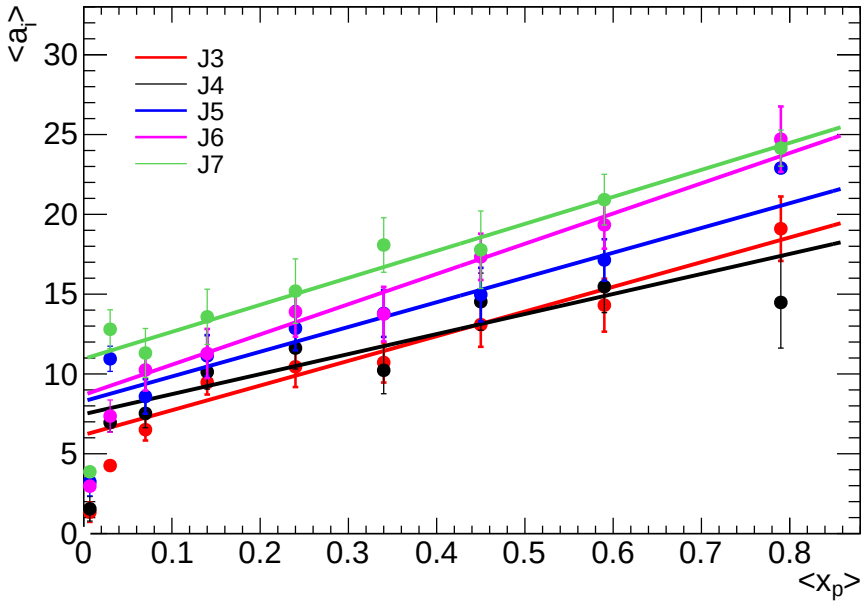


Figure 8.1: Mean values of a as a function of x_p, calculated at the truth level with HERWIG Monte Carlo samples at increasing hard scale from ∼100 GeV to ∼1 TeV. The straight line fits omit the first two $< x_p >$ values in each case.

There is a linear variation of $< a >$ vs $< x_p >$ beyond the first two points where the value of $< a >$ changes quite rapidly with $< x_p >$, however, the gradient of this linear proportion changes minimally with hard scale. Thus jet shape is found to be modelled as momentum fraction dependent; there is increased hadronic collimation at higher fractional momenta, and the shape almost scales with jet momentum.

Increased collimation might be expected at higher x_p given that as $x_p \to 1$, there is, by definition, no hadronic momentum transverse to the jet direction.

As $x_p \to 0$, hadrons are produced almost in no preferred direction, having momenta ∼zero, and randomised hadronisation produces randomised directions, isotropically about the parton direction, with momenta of the order of the hadronisation scale, i.e. a few hundred MeV, and very much less than the bremsstrahlung scale. These factors will result in little collimation of the jet in the low x_p ranges. The observed resultant hadronic angular distributions, shown in Figure 8.1, indeed show an initial steep rise at low $< x_p >$, which is consistent with the expectation of additional low momentum hadrons from hadronic decays and hadronisation processes. There is then an increasing collimation, a,

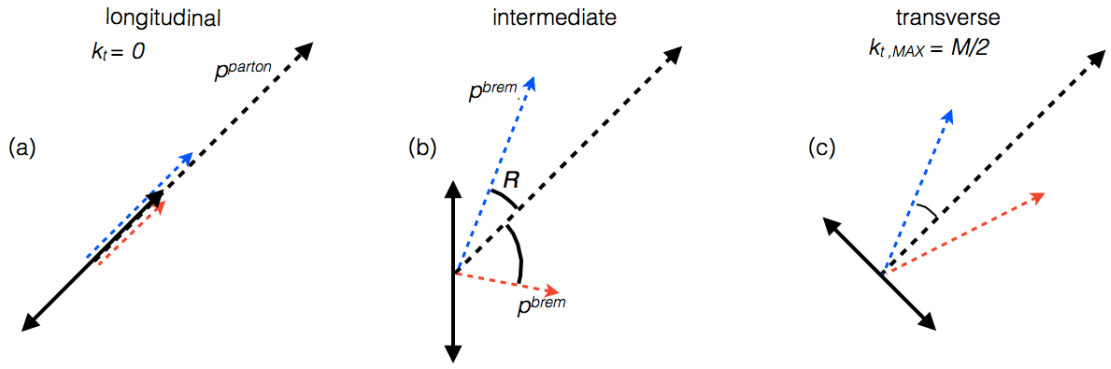


Figure 8.2: Cartoon diagram of how parton branching axis results in k_t^{brem} between 0 and $M/2$.

linear beyond the lowest x_p values.

8.2 Interpretation of Jet Shape

In order to interpret and quantify the above remarks, it is now assumed that the spread of a jet at larger x_p is dominated by the first sub-branching of a massive, (off-mass-shell) parton, into two comparatively massless final state bremsstrahlung partons (also off-mass-shell) which subsequently radiate on a much lower scale.

In the following, a lower case t and an upper case T are used to indicate the directions transverse to jet and z-axis respectively. An approximate relationship is derived between the transverse momentum, k_t^{brem}, of a bremsstrahlung and the hadronic R dependence, as characterised by $1/ < a >$.

As shown in Chapter 4 the bremsstrahlung k_t may be related to its p_T as follows

$$k_t^{brem.} \approx p_T^{brem} R^{brem.}.$$

For a *transverse* decay each *bremsstrahlung* takes an equal momentum fraction, $X = p^{brem}/p^{jet}$, i.e. $X = 0.5$, and for small R, $p_T^{brem} = p_T^{jet}/2$. This configuration, as shown in the right hand side of Figure 8.2, gives the largest possible $k_t^{brem.}$, $k_{t,max}^{brem.}$, for a given off-shell mass of the original parton.

$$k_{t,max}^{brem.} = \frac{p_T^{jet}}{2} R_{X=0.5}^{brem.}$$

Different masses give different values of $k_{t,max}^{brem.}$, which in turn give different values of $R_{X=0.5}^{brem.}$, the average value of $k_{t,max}$ over all possible masses, $< k_{t,max}^{brem.} >_M$, is then

$$< k_{t,max}^{brem.} >_M = \frac{p_T^{jet}}{2} < R_{X=0.5}^{brem.} >_M,$$

where $< R_{X=0.5}^{brem.} >_M$ is unknown. However the average R of a hadron, over all possible masses, $< R_{x_p=0.5}^{hadron} >_M$, *is* known, and using the further assumption that in an event containing a hadron of $x_p = 0.5$, the other hadrons may be considered together (vector addition) as a pseudo particle, which would have a combined x_p value of 0.5. This is then a similar kinematic situation to the initial transverse bremsstrahlung branching to relatively massless partons. It therefore expected that $< R_{x_p=0.5}^{hadron} >_M$ is similar to the unknown quantity giving:

$$< k_{t,max}^{brem.} >_M = \frac{p_T^{jet}}{2} < R_{x_p=0.5}^{hadron} >_M .$$

The mean value of R expected for a given ($x_p = 0.5$) hadron is then $< R >_{x_p=0.5}^{hadron} = 1/a_{x_p=0.5}$, and therefore

$$< k_{t,max}^{brem.} >_M = \frac{p_T^{jet}}{2a_{x_p=0.5}}. \tag{8.1}$$

For a transverse decay of an off-mass-shell parton, of mass M,

$$M = 2k_{t,max}^{brem},$$

considering that different masses give different values k_t, the average mass, $< M >$, is then

$$< M >= 2 < k_{t,max}^{brem} >_M \doteq \frac{< p_T^{jet} >}{a_{x_p=0.5}}. \tag{8.2}$$

The scale independent Sudakov form factor (referred to in Section 3.2) gives the mass of an off-mass-shell scattered parton exiting an interaction at some hard scale, Q. This mass then enables the branching.

Since the defined quantity $< M >$ has dimensions of mass and, like the Sudakov mass, the scale of this quantity, is an order of magnitude less than the hard scale, it is therefore assumed to be a closely related "mass" property, and, by scaling is expected to vary linearly in proportion with the hard scale. Due to the details of the assumptions

made, the value of $< M >$ may differ from the mean mass in the Sudakov picture, and so Equation 8.2 should be regarded as a formula *defining* such a typical mass.

8.3 Results

The values of a at $x_p = 0.5$ are calculated by fitting a straight line to the points in Figure 8.1, excluding the first two points. Corresponding values of $< k_{t,max}^{brem.} >_M$ and $< M >$, are calculated by equations 8.1 and 8.2 respectively, the latter given in the left-hand side of Table 8.1. The variation of $< M >$ with $\sqrt{-t}$ of the Monte Carlo samples (related to J* in Table 2.2), is shown in Figure 8.3. A fit of a straight line, passing through the origin has an associated χ^2/degrees-of-freedom of ∼1.6/(5-1), compatible with the expectation of scaling, as are the compatible values of $M/\sqrt{-t}$, shown in the right hand side of Table 8.1. Scaling violations of the fragmentation function (discussed in Section 2.4) which enhance production of low x_p hadrons, at the expense of high x_p hadrons, may lead to a more rapid increase of $< M >$ with hard scale than would be expected by scaling alone. Effects of scaling violation, especially the faster than scaling predicted jet mass increase with hard scale, have not been observed in the HERWIG Monte Carlo samples used.

Monte Carlo Sample	$< M >$ [GeV]	$\delta < M >$ [GeV]	$< M > /\sqrt{-t}$	$\delta(< M > /\sqrt{-t})$
J3	12.4	±2.2	0.09	±0.02
J4	24.8	±6.1	0.10	±0.02
J5	38.9	±4.8	0.08	±0.01
J6	73.1	±13.2	0.08	±0.01
J7	127.7	±20.8	0.07	±0.01

Table 8.1: The values of M and $M/\sqrt{-t}$, measured on the HERWIG Monte Carlo samples at the truth level.

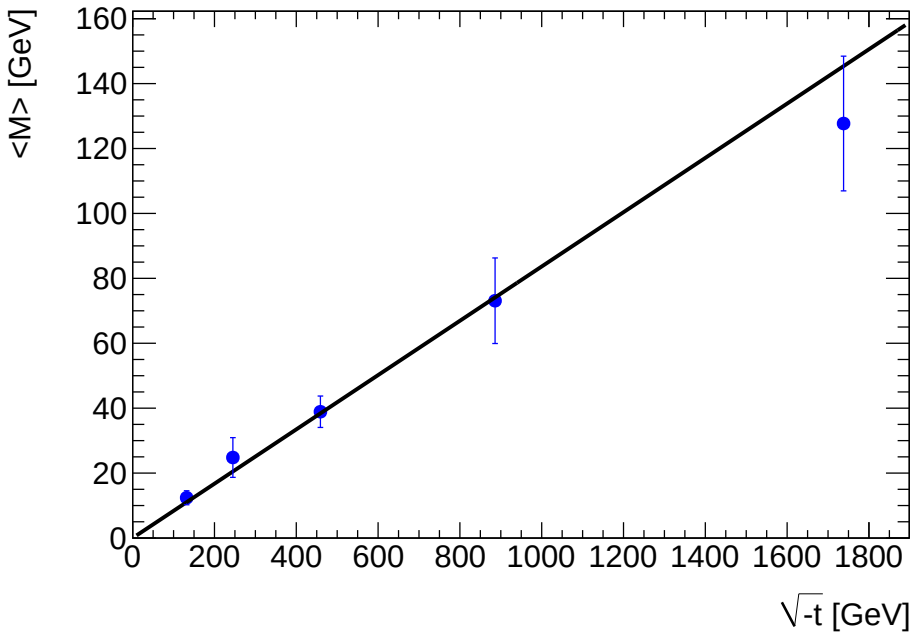


Figure 8.3: The defined jet mass as a function of hard scale, using HERWIG Monte Carlo samples at the truth level.

Chapter 9

Parton Estimator Algorithms

9.1 TRAPS

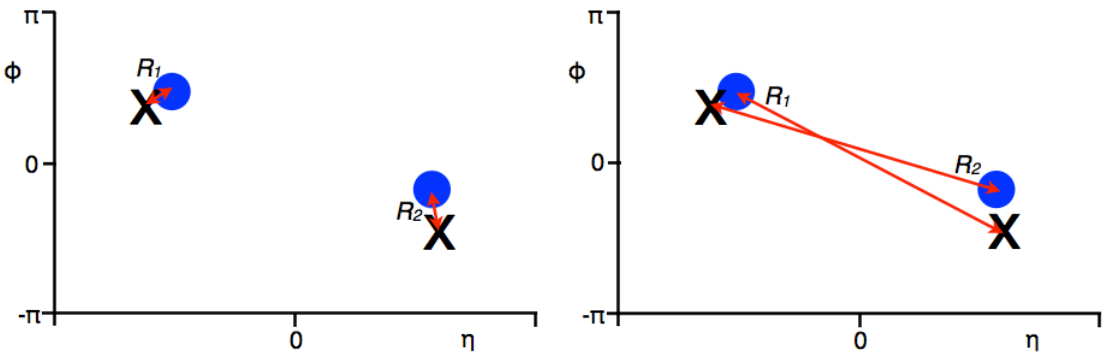
Thus far in the development of the fragmentation algorithm the "true" parton momentum has been used to calculate the true hadronic x_p. The use of a unit cone about that direction gave a fair approximation to the optimal solutions and allowed development of other aspects of the algorithm.

In this section a newly developed algorithm, TRAPS (Topological Reconstruction Algorithm for Parton Scatters) [31], is used to estimate the parton momentum and later compared with the performance of the standard Anti-k_t jet finder.

Quarks have again been selected by matching the reconstructed parton candidate to a quark, by comparing the sum of the angles between the two TRAPS parton candidates and two truth partons, see Figure 9.1. The configuration with the smallest sum of angles is taken to be the appropriate match. The truth information about parton type may then safely be used for the TRAPS parton estimators.

The results shown in Figures 9.2(a) and 9.2(b) with TRAPS are in significant disagreement with the case where the truth parton momentum is used. The disagreement is attributed to poor momentum resolution of TRAPS causing migration across the x_p ranges. Since the hadronic x_p is the fractional hadronic momentum divided by the parton momentum, overestimation of the partonic momentum decreases the x_p value and vice versa.

Due to the soft nature of the function, i.e. that the scales of neighbouring x_p ranges vary by large factors, migrations to the higher x_p ranges are more noticeable than to the

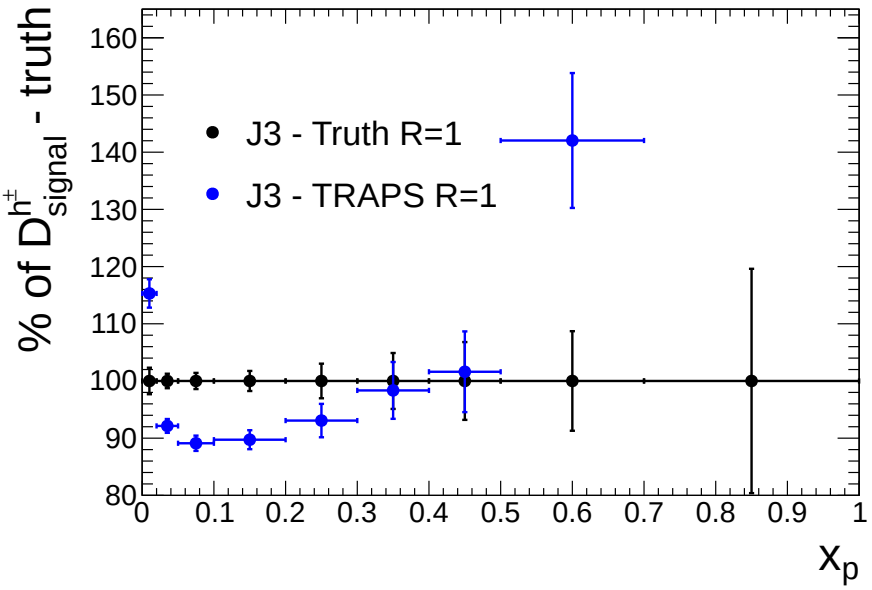


(a) Configuration A (b) Configuration B

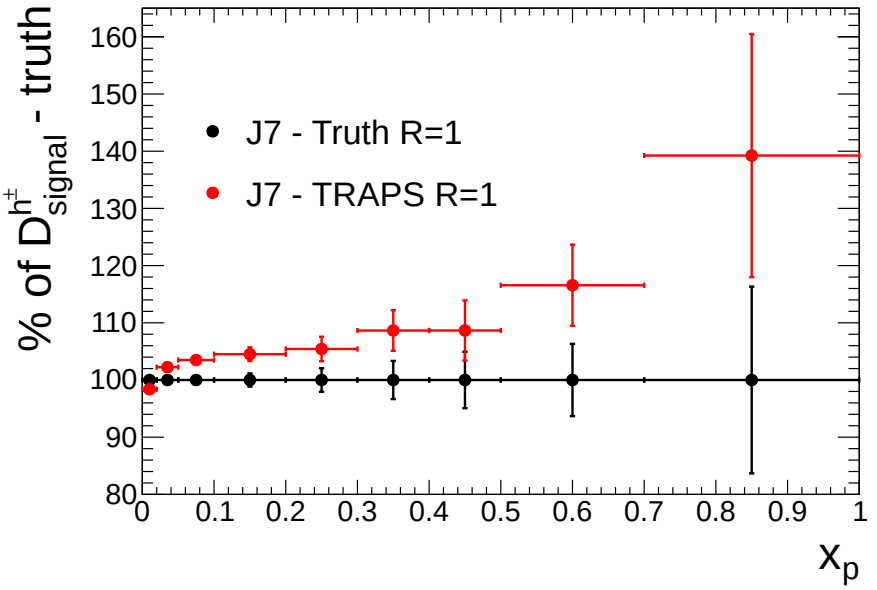
Figure 9.1: The sums of $R_1 + R_2$ in the two possible configurations are compared to match TRAPS parton candidates to truth partons. Blue circles correspond to truth partons, black crosses correspond to TRAPS parton candidates.

lower ranges, and they then cause large discrepancies in the value of D. In the case of underestimation of partonic momenta, where a hadronic x_p value is unphysically larger than unity, this value is then set to be equal to 0.999. By doing this, they do not contribute to acceptance losses, and a better estimate of the true momentum has clearly been assigned.

Although the differences with truth values are significant, they are not large compared to errors on existing data as can be seen in Figure 6.3. The poor resolution that is the ultimate cause is well-understood and there are huge Monte Carlo statistics presently available for correction procedures which will be used later in Sections 9.4 and 11.3 with tolerable systematic error. For the next few sections, however, methods of ameliorating the problem using event selections will be discussed.



(a) J3



(b) J7

Figure 9.2: The fragmentation function with TRAPS used to provide the parton momenta estimates, as a percentage of that obtained using the truth supplied partonic momenta values, at small hard scale (J3), left, and large hard scale (J7), right, calculated with Pythia Monte Carlo at the truth level.

9.2 TRAPS Leading Order Event Selection

Poor assessment of partonic momenta by the TRAPS algorithm causes migration of x_p values. The problem is worse at low momentum (J3) and might be improved using event selection. The fragmentation algorithm and TRAPS assume 2→2 parton scatters. By removing clear next-to-leading order (NLO) type events, where possible, both algorithms become more relevant.

It is the intention of TRAPS to sum over FSB (final state bremsstrahlung), as much as possible excluding ISB and MI which are largely uncorrelated with the parton. Processes which may inhibit the assessment of partonic momenta by TRAPS, or any jet finder which reconstructs only two jets in each event, are indicated in Figure 9.3. The arrow from the

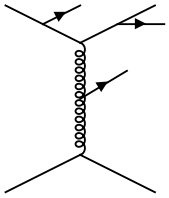


Figure 9.3: Processes which may degrade the momentum resolution of TRAPS.

top left fermion line indicates an ISB which may be incorrectly included. The arrow from the top right fermion line indicates an FSB, which should be included but may be missed. The arrow on the propagator represents a NLO parton which should be ignored from the point of view of efficient background calculation. Pile-up and MI background may also be incorrectly included. Removing events in which these processes occur is expected to improve the estimation of partonic momenta by TRAPS.

Di-parton events must be planar, those with more final state partons need not be. A selection was applied to remove acoplanar events using the TRAPS acoplanarity variable, which is defined as the p_T out of a plane divided by the total p_T, where the plane is that from which the transverse momenta is minimised. NLO type events are not planar, and so a cut on the event planarity may remove them, however such a requirement might also preferentially select events which contain high x_p hadrons, which would also be planar.

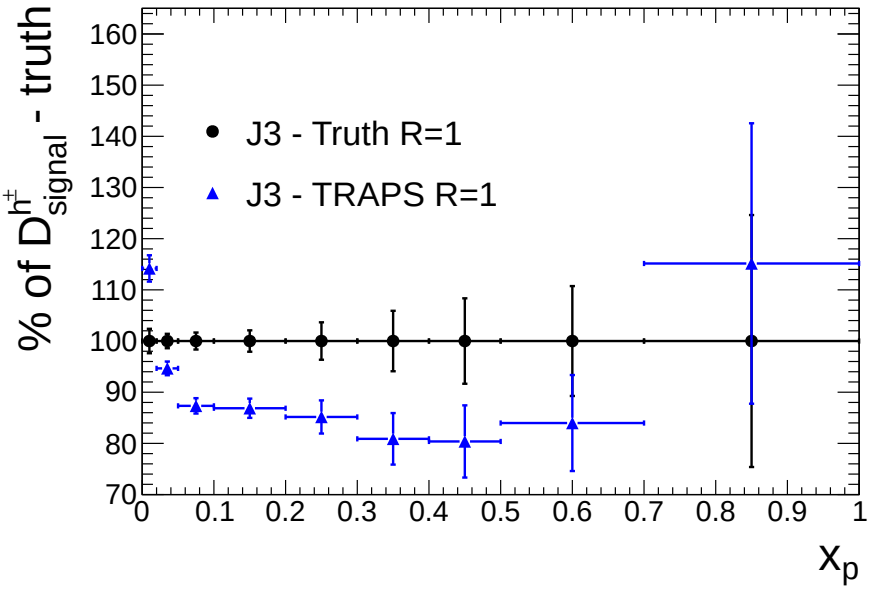
The selection of events with *acoplanarity*< 0.05 removes approximately 90% of events in the lowest $\hat{p_T}$ sample, and only 10% in the highest $\hat{p_T}$ sample. Comparison of the fragmentation function calculated with selected events to that calculated with all events confirms the introduction of a bias.

An alternative requirement is made, to attempt to remove NLO type events, the partons are required to be balanced in p_T within 10% and back to back in azimuth within 10%. Some *good* events will be removed, such as events in which an ISB boosts the hard scatter system. The event acceptance after this requirement is shown in Table 9.1. At low

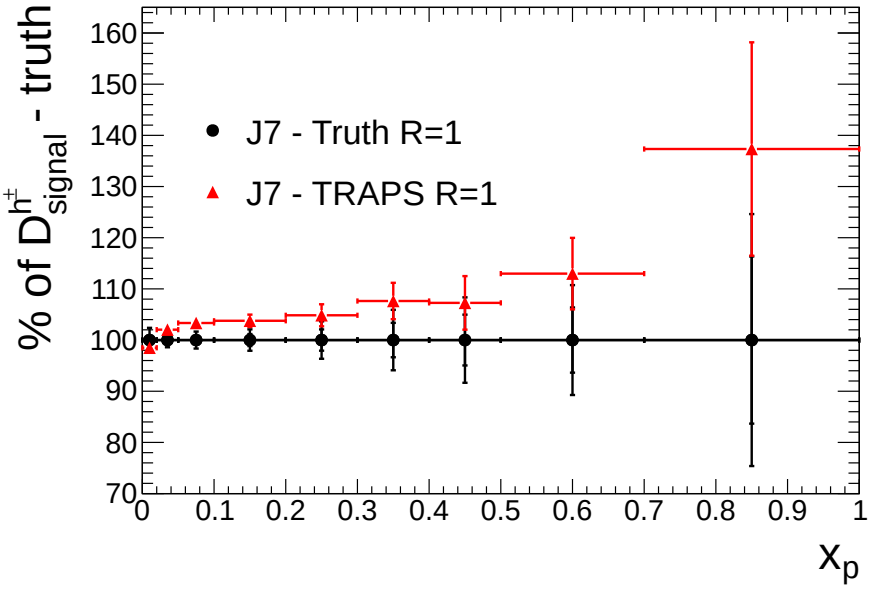
Monte Carlo Sample	Truth Parton	TRAPS Parton Estimator
J3	$\approx 70\%$	$\approx 65\%$
J4	$\approx 80\%$	$\approx 80\%$
J5	$\approx 85\%$	$\approx 90\%$
J6	$\approx 85\%$	$\approx 90\%$
J7	$\approx 98\%$	$\approx 97\%$

Table 9.1: Event acceptance after leading order event selection.

momentum, migration into the highest x_p range is reduced without the introduction of a bias, as shown in Figures 9.4(a) and 9.4(b), which may be compared to Figures 9.2(a) and 9.2(b).



(a) J3



(b) J7

Figure 9.4: The fragmentation function with TRAPS used to provide the parton momenta estimates for events passing the event selection, as a percentage of that obtained using the truth supplied partonic momenta values, at low momentum (J3), left, and high momentum (J7), right, calculated using Pythia Monte Carlo at the truth level.

x_p	Efficiency		Purity	
	Anti-k_T	TRAPS	Anti-k_T	TRAPS
$0 \leq x_p < 0.02$	98%	97%	98%	97%
$0.02 \leq x_p < 0.05$	91%	85%	90%	86%
$0.05 \leq x_p < 0.1$	89%	81%	88%	82%
$0.1 \leq x_p < 0.2$	90%	82%	89%	84%
$0.2 \leq x_p < 0.3$	85%	75%	84%	76%
$0.3 \leq x_p < 0.4$	90%	69%	79%	69%
$0.4 \leq x_p < 0.5$	82%	66%	76%	64%
$0.5 \leq x_p < 0.7$	79%	76%	84%	73%
$0.7 \leq x_p \leq 1.0$	86%	80%	86%	70%

Table 9.2: Efficiencies and purities in the chosen analysis intervals using the respective parton estimators at low transverse momentum (J3 Monte Carlo). Statistical errors are well below the 1% level.

9.3 Parton Estimator Algorithm Comparison

The efficiencies and purities in the chosen analysis intervals, calculated using hadrons within a unit cone about the jet or parton estimator (η, ϕ) position are shown in Tables 9.2 and 9.3. Quarks have been selected as described in Section 9.1. Since only the jet direction is used in this exercise, migration of x_p values may only occur due to mis-estimation of the parton momentum.

The efficiency in a given x_p bin is the total number of hadrons with an x_p value in that range at the truth (partonic) *and* algorithmic level, divided by the number of hadrons with an x_p value in that range at the truth level. The purity is the number of hadrons with an x_p value in that range at the truth and algorithmic level, divided by the number in that range at the algorithmic level. That is, the efficiency is the percentage of hadrons that remain in a given bin after the resolution of a given algorithm is considered, and the purity is the percentage of such hadrons compared to the new bin content.

Higher algorithm efficiencies and purities indicate less dependence on the unfolding, and the lower of the two indicates whether emigration (smaller efficiency) or immigration (smaller purity) is the larger resolution problem to be unfolded in a given bin. As expected from such a "soft" function, immigration into the higher x_p intervals can be seen from Table 9.2 to be the chief problem.

While Tables 9.2 and 9.3 clearly indicated the Anti-k_T algorithm [40] to be the preferred choice, the resolutions of both parton estimators are sufficient for a measurement with the same binning as used in previous measurements.

x_p	Efficiency		Purity	
	Anti-k_T	TRAPS	Anti-k_T	TRAPS
$0 \leq x_p < 0.02$	100%	99%	100%	100%
$0.02 \leq x_p < 0.05$	98%	96%	97%	95%
$0.05 \leq x_p < 0.1$	98%	96%	97%	94%
$0.1 \leq x_p < 0.2$	98%	96%	97%	94%
$0.2 \leq x_p < 0.3$	97%	94%	96%	91%
$0.3 \leq x_p < 0.4$	96%	93%	95%	88%
$0.4 \leq x_p < 0.5$	95%	91%	94%	86%
$0.5 \leq x_p < 0.7$	98%	96%	96%	89%
$0.7 \leq x_p \leq 1.0$	98%	99%	96%	87%

Table 9.3: Efficiencies and purities in the chosen analysis intervals using the respective parton estimators at high transverse momentum (J7 Monte Carlo). Statistical errors are well below the 1% level.

The fact that both algorithms may be used as input to the fragmentation algorithm, which then produces compatible output, i.e. agreement with truth values, demonstrates a lack of dependence of the fragmentation algorithm on the specific algorithm chosen to supply the parton estimates.

9.4 Algorithmic Unfolding

Any physical measurement is subject to a resolution problem which may, or may not be a significant problem. The resolution will be degraded by the detector used or similarly by inefficiencies of any tool or algorithm employed as part of the measurement. As long as the true result and the result affected by resolution may be linked then such effects may be removed on a statistical basis.

The FAPS method is thought to be safe against directional accuracy, via the optimal solutions method, however, it is sensitive to poor resolution in the magnitude of partonic momenta.

The ROOT ([41]) package *TUnfold* [42] is applied to remove the effect of a jet finding algorithm on the measured distribution of fractional hadronic momenta, i.e. to remove the affect of imperfect assessment of partonic momenta magnitude. It is noted that two jet algorithms will be used to estimate the partonic momenta, to test for dependence of FAPS on the exterior algorithm.

9.4.1 The TUnfold Package

TUnfold iteratively solves Equation 9.1 to minimise the χ^2 as a function of x, the true distribution.

$$\chi^2 = (y - Ax)^T V_{yy^{-1}}(y - Ax) + \tau^2(L(x - x_0))^T L(x - x_0) + \Lambda \sum_i (y_i - (Ax)_i) \quad (9.1)$$

There are three separate terms on the right. The first term gives the standard χ^2 of the difference between the "measured" distribution and the expected distribution Ax, taking account of correlation, where A is the migration matrix found by relating the true to measured distributions in Monte Carlo. The covariance matrix, $V_{yy^{-1}}$, is calculated from the known errors on y.

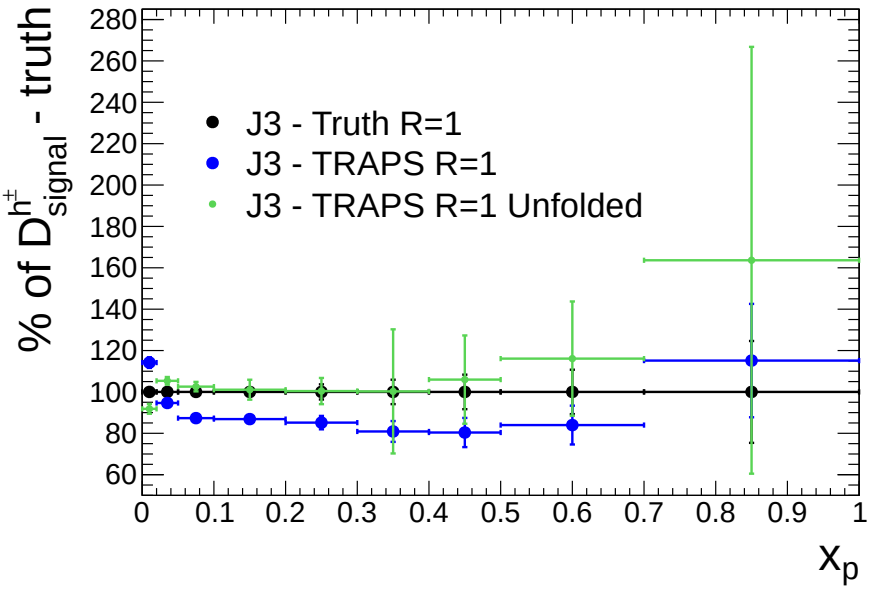
In the present case the measured distribution, y, is the number of hadrons measured in each bin of x_p using the jet algorithm estimated partonic momenta.

The second term applies a vector of offsets of the mean unfolded "true" result from the diagonal, x_0. The matrix of regularisation conditions, $L(x-x_0)$, ensures the minimisation is performed with respect to the values of best fit to the migration distribution, thus accounting for the bias in the measurement. The parameter τ defines the significance of the second term, with relation to the first. The third term ensures the same number of hadrons in the measured distribution as the unfolded one (i.e. unitarity) and the Lagrangian multiplier, Λ, allows minimisation with respect to this constraint.

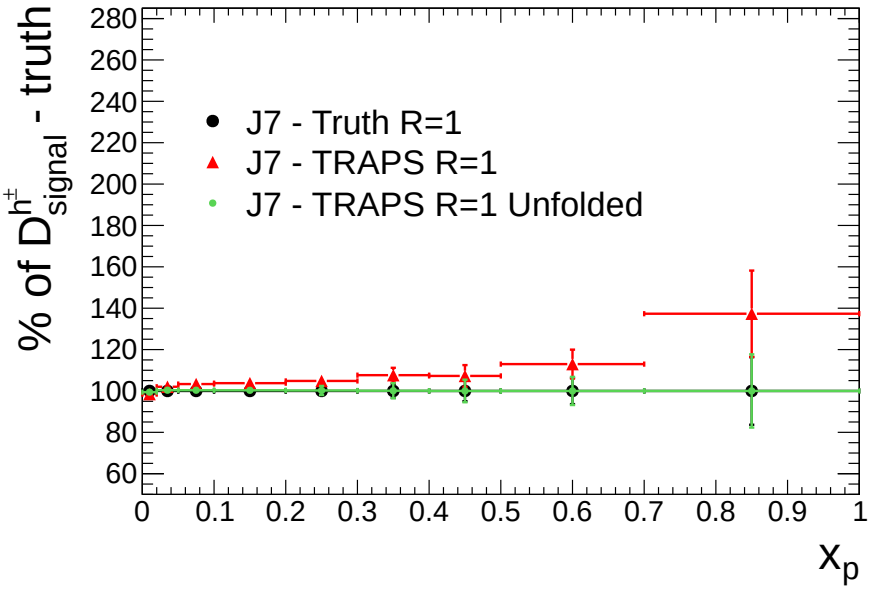
9.4.2 Unfolding Results

To test the unfolding process and its error handling, 90% of the available Monte Carlo events are used to fill the matrices A, V and the vector x_0. The remaining 10% were treated as "data" only and unfolded.

The τ parameter is chosen to give the smallest optimal χ^2 possible for which χ^2 is still sensitive to τ, found by plotting the logarithm of τ against the logarithm of χ^2_{min} and choosing the value of τ on the kink of the resulting curve. The results of unfolding at high p_T are shown in Figure 9.5(b) for the J7 sample.



(a) J3



(b) J7

Figure 9.5: The fragmentation function with TRAPS used to provide the parton momenta estimates, as a percentage of that obtained using the truth supplied partonic momenta values, at low momentum (J3), left, and high momentum (J7), right, measured on Pythia Monte Carlo at the truth level. Also shown is the unfolded result.

The unfolding works well at this high p_T. The results of unfolding at the lowest p_T are shown in Figure 9.5(a). At low p_T the unfolding does not get back to the truth result, within errors, in the lowest two x_p ranges. The discrepancy between the truth and the unfolded result will be shown to be attributed to less than perfect directional resolution of the TRAPS algorithm (Figure 9.6). This discrepancy is not a problem since a larger cone may be used and if necessary extrapolation. The unfolded result is compared to that obtained using the truth supplied magnitude of the parton momentum with the TRAPS supplied parton direction of the parton in Figure 9.6. The unfolding of the magnitude

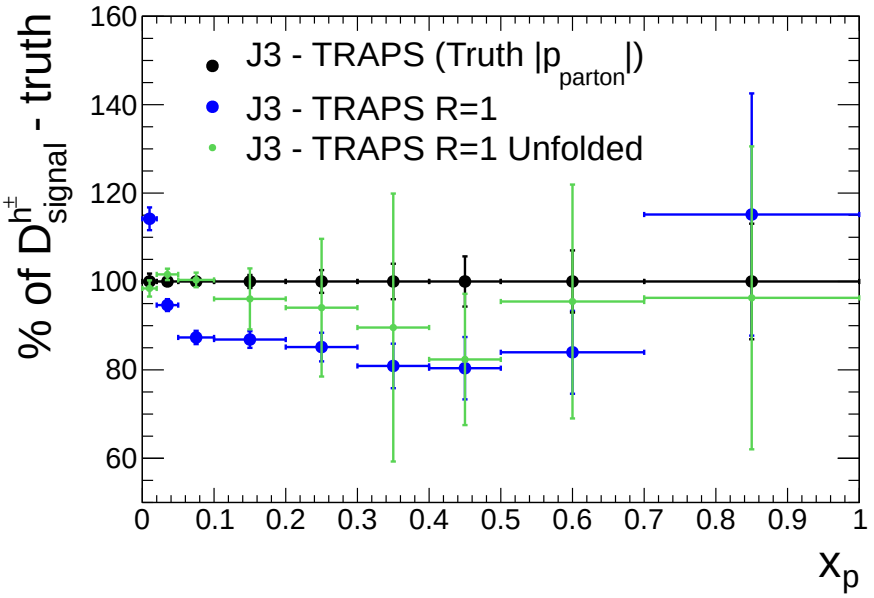


Figure 9.6: The fragmentation function with TRAPS used to provide the parton momenta estimates, as a percentage of that obtained using the truth supplied partonic momenta values at low momentum (J3), measured on Pythia Monte Carlo at the truth level. Also shown is the result using TRAPS with the truth supplied magnitude of the partonic momenta and the unfolded result.

of the parton momenta gets back to the distribution of fractional hadronic momenta obtained using the TRAPS supplied direction for the measurement cone with the truth supplied magnitude, and therefore there is a problem of direction resolution when using the TRAPS algorithm.

9.4.3 TUnfold Instabilities

The method is unstable, small changes in the input lead to large changes in the output and the errors associated are large, errors bars and sometimes even central values can unphysically extend to less than zero. This is especially a problem when correcting the

background, which is mainly at low x_p, meaning that the higher x_p bins of the matrices and inputs vectors to the unfolding are sparsely populated. Many more statistics could be used in an attempt to sufficiently populate the higher x_p bins, however an impractical number would probably be required.

The author of TUnfold has been contacted with regard unphysical outputs. The question was asked as to whether there is a setting available to "impose positivity" by perhaps constraining the chi-squared fit to make negative results impossible. However, there is no such option, this could be a useful option which may make results more accurate/sensible, by removing the possibility of unphysical solutions.

9.5 Extrapolation using Jet Finders

When using a jet finder to provide the parton estimator information poor directional resolution of the algorithm may cause a lower starting value and some delay of the rise of the curve of D_{signal} as a function of R, as shown in the low R region of Figure 9.7. In essence, the estimated direction is displaced from the truth parton direction, as shown

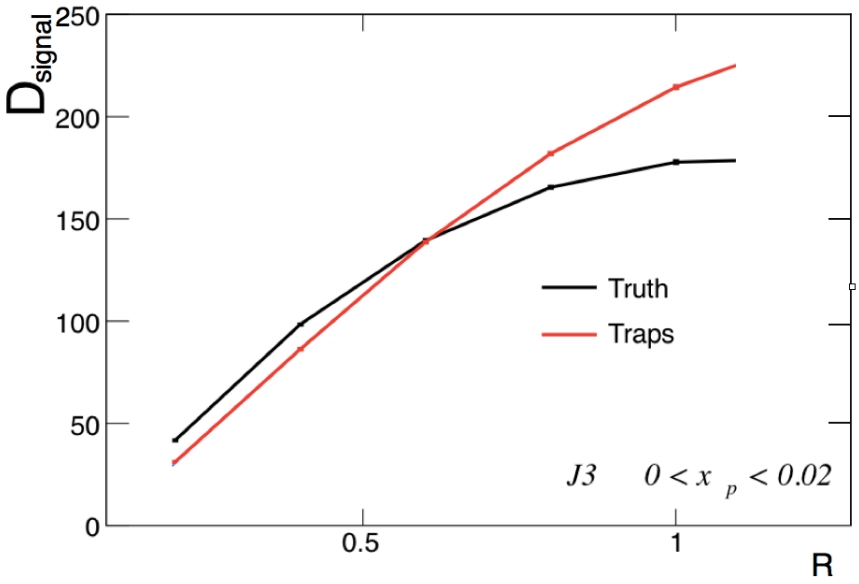


Figure 9.7: D_{signal} as a function of R for $0 > x_p \geq 0.02$. Pythia J3 Monte Carlo has been used at the truth level.

in Figure 9.8, in which a significantly large ISB has caused TRAPS to reconstruct the parton direction slightly away from the truth parton direction and towards the ISB. Fewer hadrons will now be counted in the smallest cone than would have been if it were centered on the true parton direction. As a result, the first gradient of the $D_{signal}(R)$ curve may

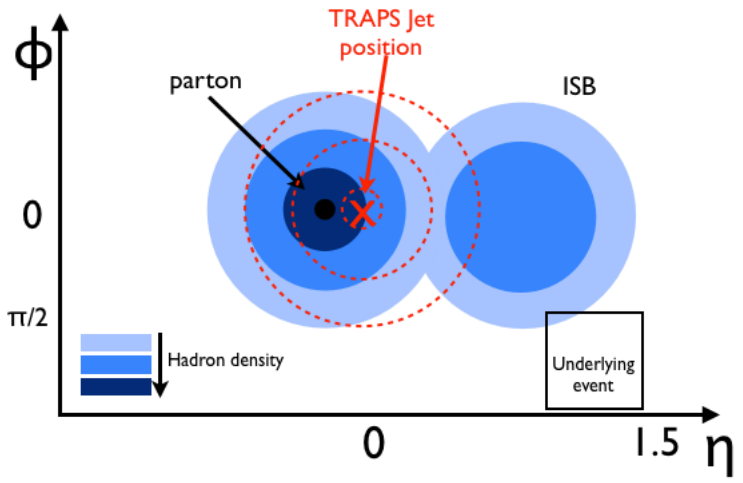


Figure 9.8: A cartoon η, ϕ map of a parton, ISB, the corresponding background solid angle and the TRAPS supplied parton direction. The dashed lines represent successive iterative search cones.

be more shallow than the next, (see Figure 9.9), which would give a negative value for the exponential coefficient, and hence an unphysical solution. Such solutions are avoided by discarding the first measured value when there is a rising gradient, and calculating the first gradient using the next point which invariably then yields a viable solution.

There is more background included in a sampling cone of large R in this configuration, compared to when the ISB is far from the parton, and this may lead to a $D_{signal}(R)$ curve which does not asymptotically plateau, as shown in the large R section of the TRAPS curve in Figure 9.7. Calculating an asymptotically plateauing function to such a curve, and extrapolating from each measured value may lead to a variety of inconsistent solutions, as shown in Figure 9.10. This is a problem since the freedom to choose the solution with the smallest total error is only acceptable if all solutions are compatible. This problem is avoided by iteratively discarding a solution if it is incompatible with any smaller R solutions.

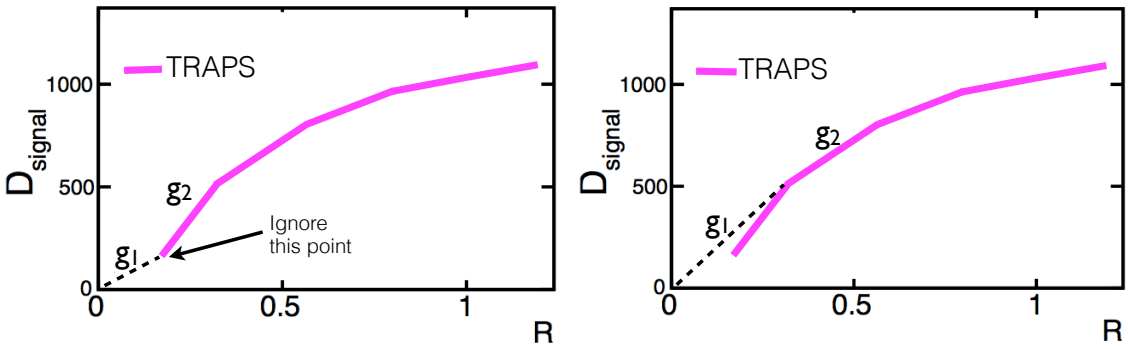


Figure 9.9: A cartoon diagram of D_{signal} as a function of R using the TRAPS supplied parton information. The first measured value is discarded since it yields an unphysical extrapolated solution.

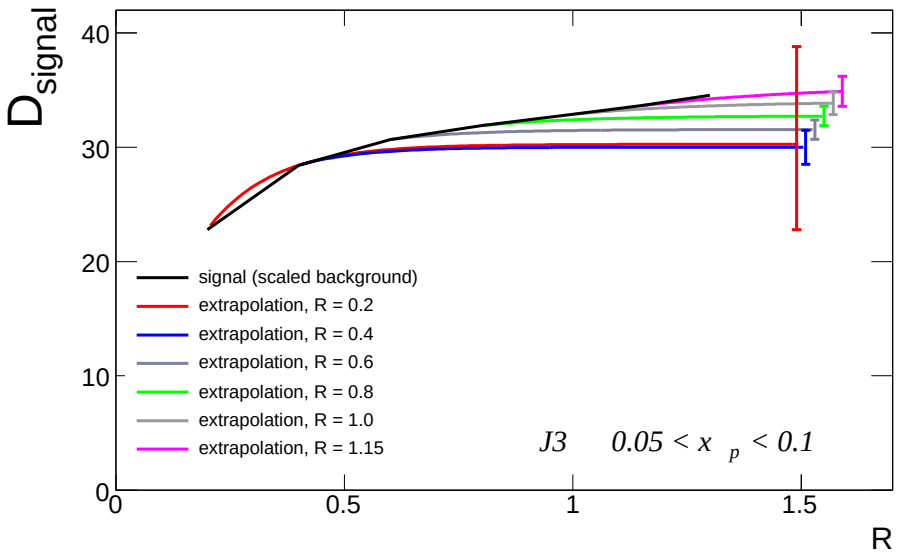


Figure 9.10: The curve of $D_{signal}(R)$ for $0.05 \geq x_p \geq 0.1$, produced using truth supplied parton estimator information from the J3 Monte Carlo sample.

9.6 Jet Correlations in η

Following on from the problem case found in the extrapolation and optimal solutions study a test done to investigate whether the behaviour has a direction dependence.

The fragmentation cone was divided into quadrants for which $\Delta\eta^2 > \Delta\phi^2$ for two of the quadrants and $\Delta\phi^2 > \Delta\eta^2$ for the other two, see Figure 9.11. The D_{signal} vs R curve

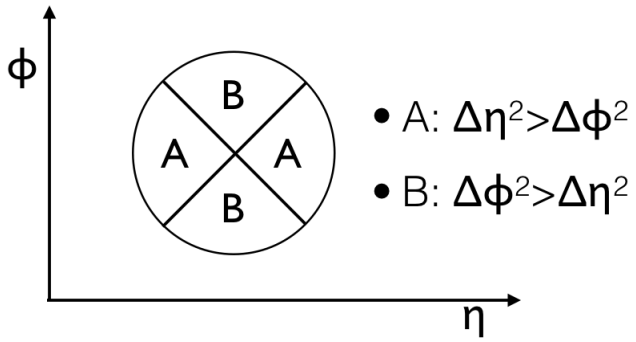


Figure 9.11: The fragmentation cone divided into quadrants.

for $\Delta\eta^2 > \Delta\phi^2$ exhibits the behaviour in question while the other curve does not, Figure 9.12. Two possible causes for this behaviour are colour strings connecting the jet to the

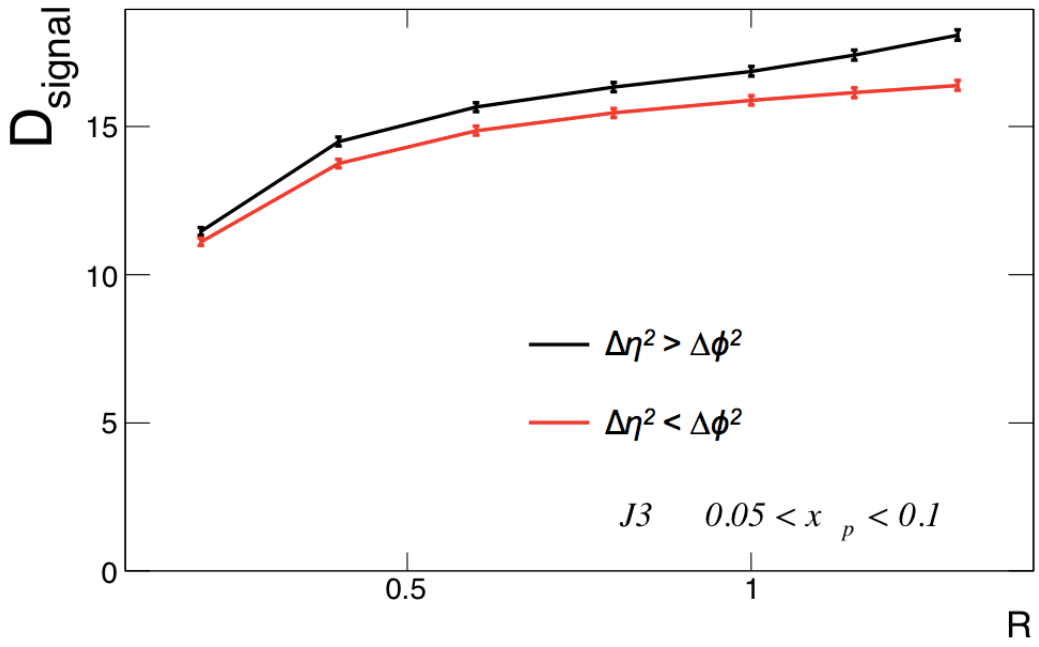
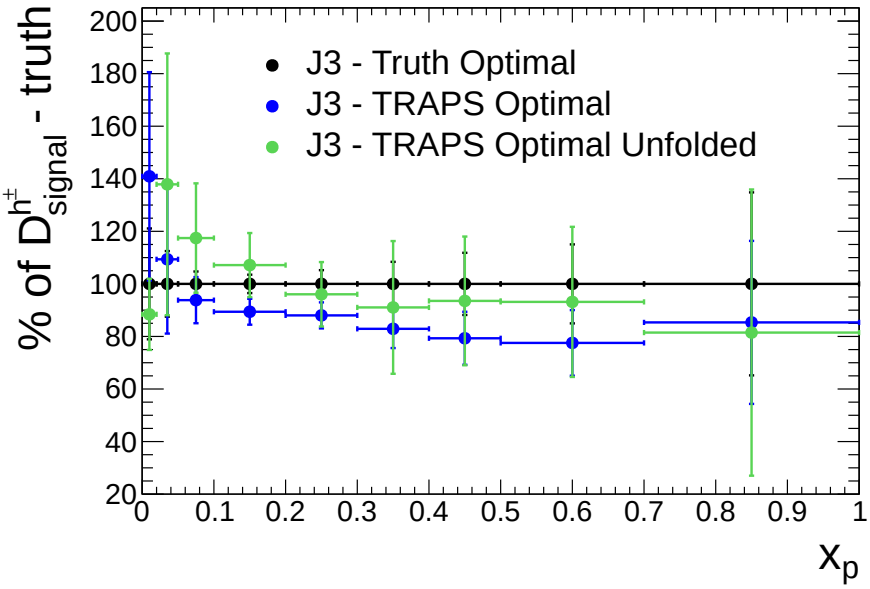


Figure 9.12: D_{signal} as a function of R for $\Delta\eta^2 > \Delta\phi^2$ and $\Delta\phi^2 > \Delta\eta^2$.

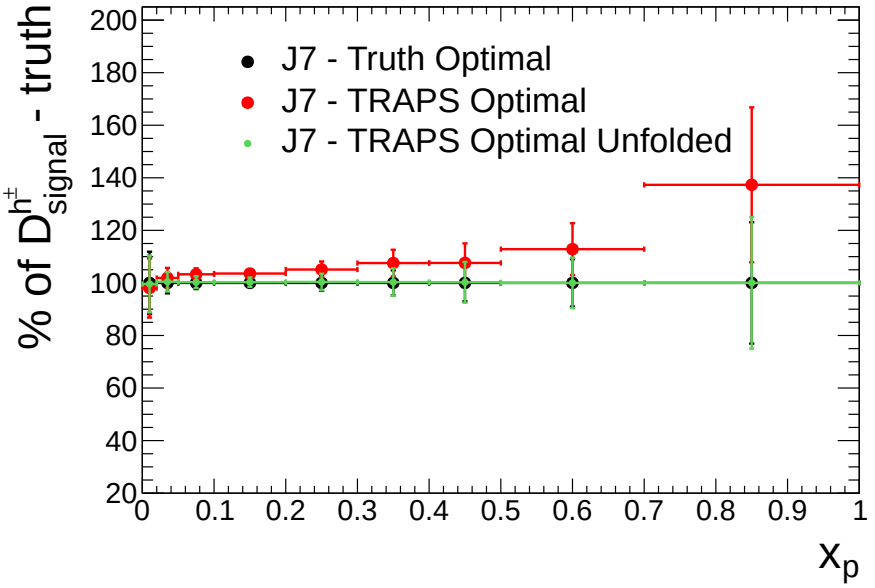
proton remnants and ISB which is preferentially in the η direction.

9.7 Optimal Solutions Results

The results using the optimal solutions method are shown in Figures 9.13(a) and 9.13(b). There is very good agreement between the results of the algorithm on pure truth and the algorithm using TRAPS and unfolding. In Section 9.4 it was shown that the unfolding of the TRAPS momentum works well in all but the lowest p_T and x_p. The excellent agreement in the lowest x_p range between the pure truth and TRAPS unfolded results at J3 should be contrasted to the poorer agreement in the next x_p range. The $D_{signal}(R)$ curves are included to explain the discrepancy. The non-plateau in Figure 9.14(b) results in a large extrapolation, while Figure 9.14(a) has a plateau and hence no large extrapolation.

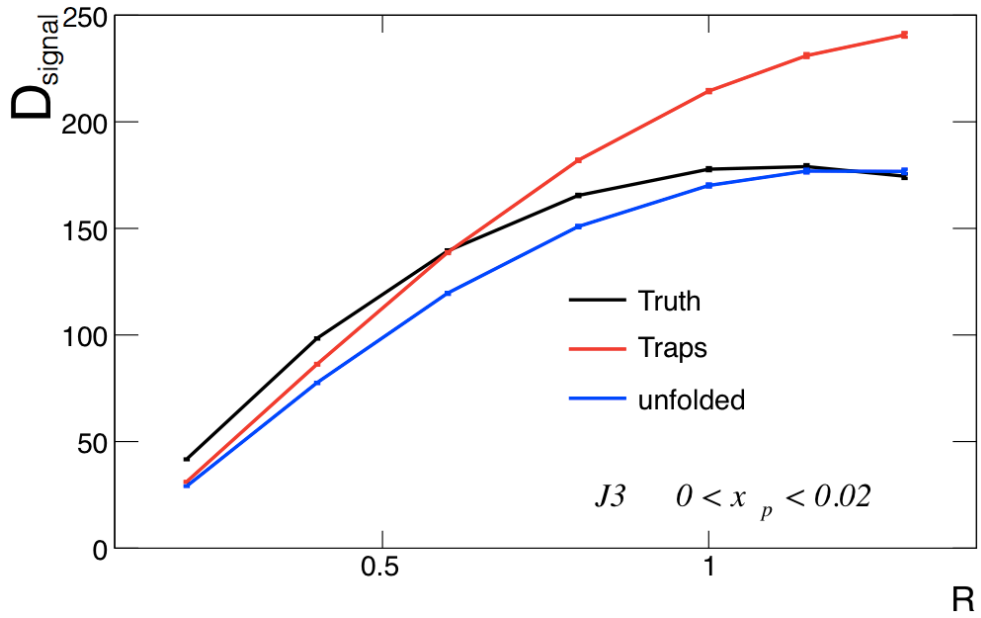


(a) J3

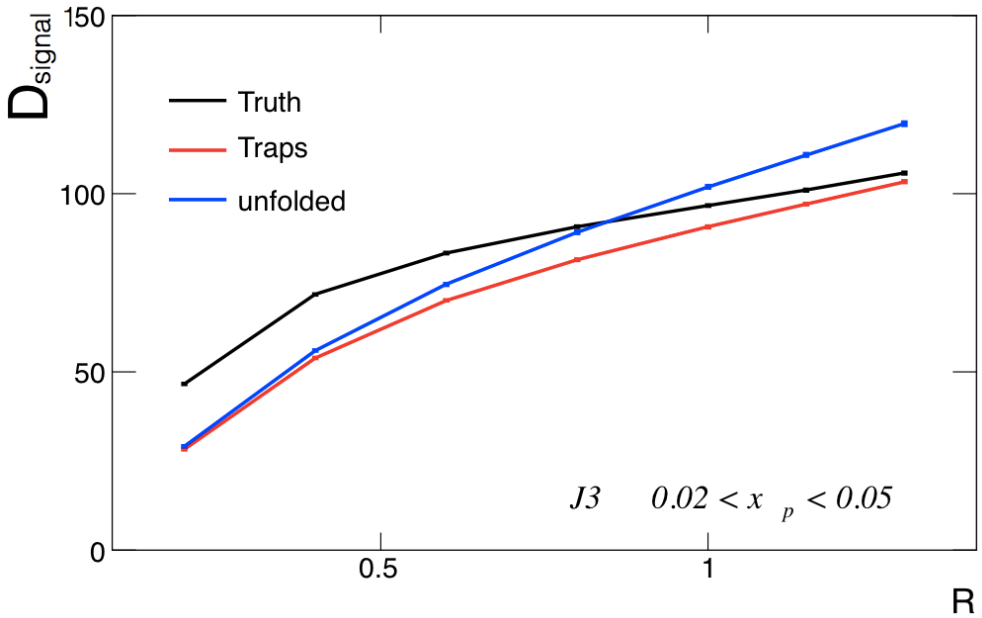


(b) J7

Figure 9.13: The fragmentation function obtainted using the optimal solutions method, with TRAPS used to provide the parton momenta estimates for events passing the event selection, as a percentage of that obtained using the truth supplied partonic momenta values, at low momentum (J3), left, and high momentum (J7), right, measured on Pythia Monte Carlo at the truth level. Also shown is the unfolded result.



(a) $D_{signal}(R)$, $0 > x_p \geq 0.02$ J3.



(b) $D_{signal}(R)$, $0.02 \geq x_p \geq 0.05$ J3

Figure 9.14: D_{signal} as a function of R.

9.8 Comparison with the Anti-K_T Algorithm

The fragmentation algorithm requires as input the vector momentum of a parton estimator, from which the measuring cones are centred and which also serves to normalise the fractional momenta of the hadrons within each cone. The parton estimator information may be supplied by TRAPS or another algorithm which at least provides a good approximation to a fragmenting parton's momentum vector.

While the TRAPS algorithm claims to measure the partons themselves, the algorithm is new and as yet not used widely. It is useful to demonstrate input insensitivity of the fragmentation algorithm being developed. Thus the present algorithm of choice of the ATLAS collaboration, the Anti-k_T algorithm [40], has also been used to supply the parton estimator momenta. While Anti-k_T is well recognised, it claims only to be measuring 'jets' rather than estimating partons specifically, let alone those partons which are the result of a hard scatter.

The Anti-k$_T$ algorithm is a sequential recombination clustering algorithm, based on the following distance measures which are expressed in terms of the particle rapidity, y, azimuth, ϕ, and transverse momentum p_T. The distance, d_{ij}, between any two particles i and j

$$d_{ij} = min(p_{Ti}^{-2}, p_{Tj}^{-2})\frac{\Delta_{ij}}{R}, \Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$$

and the distance, d_{iB}, between any particle, i, and the beam, B

$$d_{iB} = p_{Ti}^{-2}.$$

Clustering is performed by first calculating all distances d_{ij} and d_{iB}. If the smallest is a d_{ij} the two particles are combined (by summing their four-momenta) and the new distances are computed. If the smallest is a d_{iB} then particle i is removed from the clustering sample. This process is repeated until all particles are either discarded or clustered into jets. The parameter R scales the d_{ij} with respect to the d_{iB} such that any pair of final jets are separated by at least that value.

In this study the R parameter is chosen to be 0.6, one of the two ATLAS defaults, the other being 0.4. It has previously been demonstrated within this thesis that a maximum cone radius of 0.6 will not include all of the fragments correlated with the parton direction,

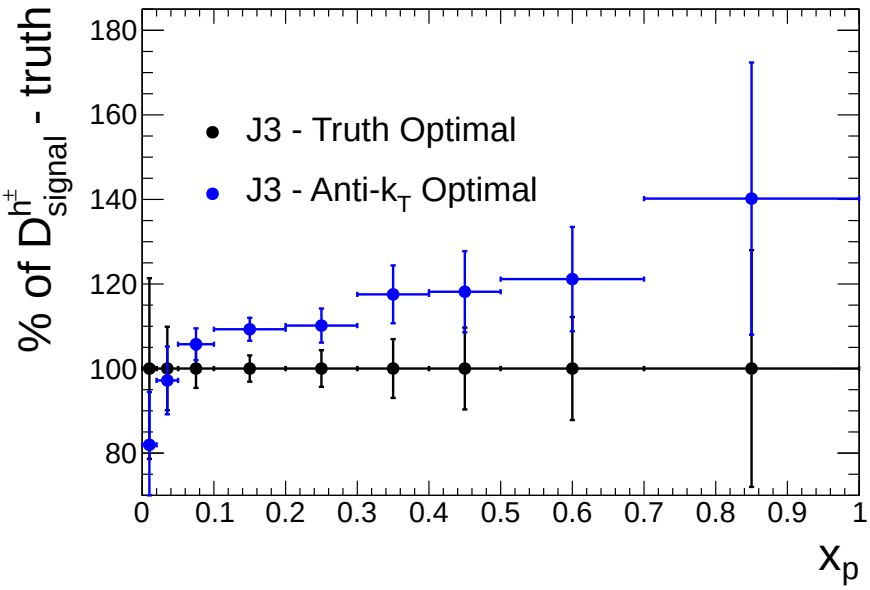
see, for example, Figure 9.14(a). The algorithm makes no intrinsic underlying event (UE) subtraction as TRAPS does, this will cause lower x_p values due to overestimation of parton momentum. Using, say, a 0.6 cone radius will cause opposite migration. In Figures 9.15(a) and 9.15(b) the results using Anti-k_T are shown, with no event selection used and the two highest p_T quark initiated jets considered, the matching of jets to truth partons was done the same way as for TRAPS, Section 9.1. There is uniform migration to higher x_p, indicating the net effect is underestimation of the partonic momenta.

9.8.1 Anti-k_T Selection

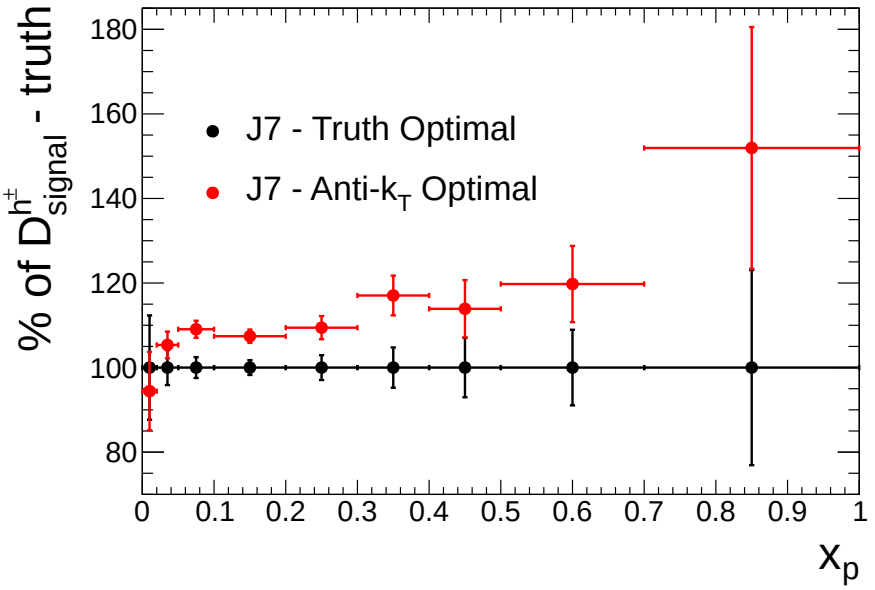
The 2→2 event selection (see Section 9.2) is applied based on the momentum of the parton estimators as measured by the Anti-k_T algorithm, Figures 9.16(a) and 9.16(b). A bias is introduced to the truth level scaled momentum distribution.

The Anti-k_T algorithmic resolution is unfolded, see Figures 9.17(a) and 9.17(b). Agreement with results where the truth supplied value of the partonic momenta had been used are good, indicating that the Anti-k_T algorithmic partonic momentum resolution has been successfully removed. The event selection is not needed. The agreement is excellent at high momentum. At low momentum x_p resolution is adequate, and comparable to that obtained with the TRAPS algorithm (Figure 9.13(a)), which had poorer agreement at low x_p, 40% difference in the range $0.02 < x_p < 0.05$, which is of the order of the difference observed at high x_p with the Anti-k_T algorithm.

From this point onwards, the fragmentation algorithm will be developed and tested using both the above algorithms in parallel either to demonstrate input independence, or to justify the choice of one while measuring the systematic error with respect to the other.

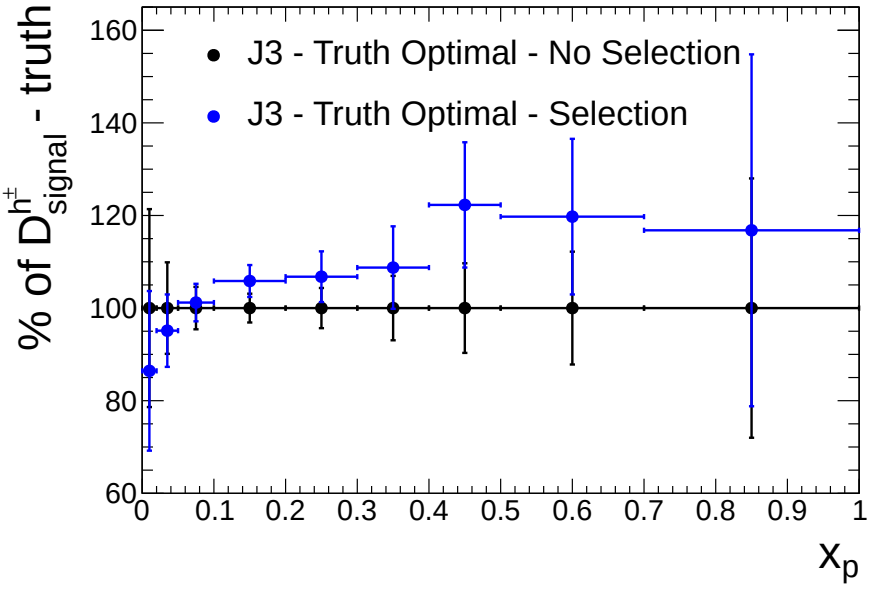


(a) J3

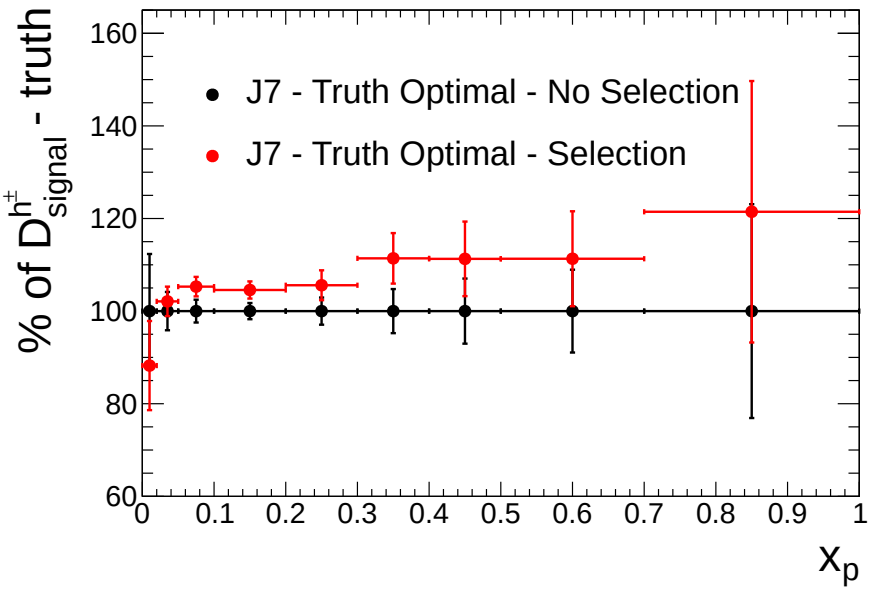


(b) J7

Figure 9.15: The fragmentation function with the Anti-k_T algorithm as input, as a percentage of that obtained using the truth supplied partonic momenta values, at low momentum (J3), left, and high momentum (J7), right, measured on Pythia Monte Carlo at the truth level. The optimal solution is used with no event selection.

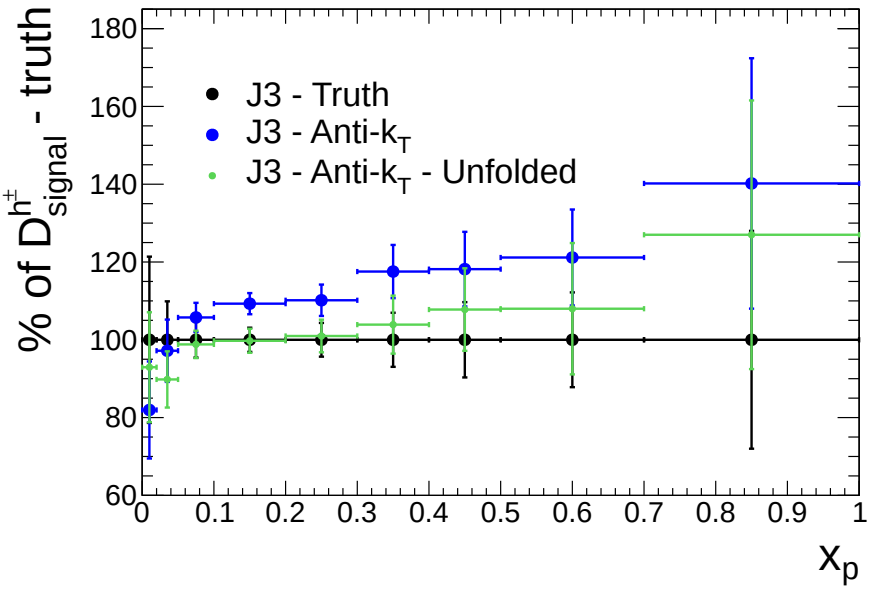


(a) J3

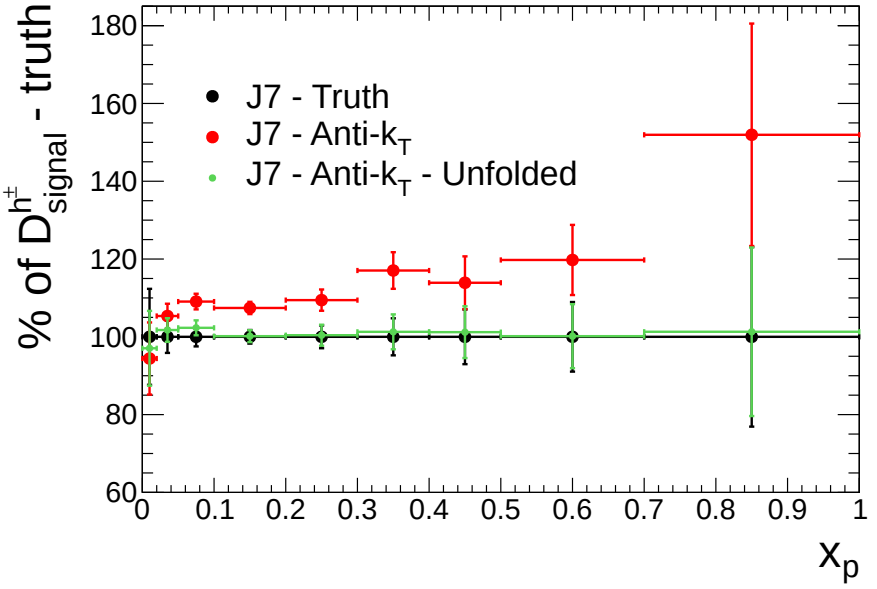


(b) J7

Figure 9.16: The fragmentation function with truth supplied values of partonic momenta as input, with and without application of the event selection at the Anti-k_T level, at low momentum (J3), left, and high momentum (J7), right, measured on Pythia Monte Carlo at the truth level. Both are displayed as a percentage of the result obtained without selection of events. The optimal solution is used in both cases.



(a) J3



(b) J7

Figure 9.17: The fragmentation function using Anti-k_T and unfolding, displayed as a percentage of that using the truth supplied partonic momenta, at low momentum (J3), left, and high momentum (J7), right, measured on Pythia Monte Carlo at the truth level. The optimal solution is used.

Chapter 10

Hadron Measurement

To measure charged particles and jets, and therefore the x_p variable and the directions for fragmentation sampling cones, algorithms are applied to tracking and calorimetry measurements, which also then define these measured objects. These measured objects then have associated resolutions due to the physical apparatus and due to the algorithms used.

Unfolding is used to correct for the momentum resolution of the tracks and parton estimates, to recover what would have been measured, as described by the Monte Carlo 'truth' variables. Directional resolution of the parton estimators is resolved by the variable cone and extrapolation techniques as explained in Section 7.8.

In the present study, no attempt is made to address optimisation beyond using standard selections and calibrations. However, in the interest of algorithmic resolution, two completely different jet algorithms are used as input.

10.1 Calorimetry

The energy of a particle is measured by collecting, or essentially by counting, the secondary charged particle multiplicity produced during showering in an active medium. The fractional error on the multiplicity is of the form $\delta_{N_{particles}}/N_p = 1/\sqrt{N_{particles}}$, i.e. it is smaller for larger energies, as, then, is the fractional energy resolution of an incident particle or group of particles.

Hadrons deposit some of their energy in both the electromagnetic and the hadronic calorimeters. Both the ATLAS electromagnetic and the hadronic calorimeters have dif-

ferences in their response to electromagnetic and hadronic particles (i.e. they are non-compensating) and are calibrated at the EM scale.

10.1.1 ATLAS Calorimetry

The ATLAS electromagnetic calorimeter (ECAL) [43] is of sampling type, utilising liquid Argon (LAr) as the active medium, interspersed with lead absorber plates. The ECAL electrodes have accordion geometry. In the barrel region the sampling layers are 2.1 mm thick, while in the end-caps the LAr layers are of variable thickness due to the increasing amplitude of the accordion waves with increasing radius and the constant thickness of the absorption layers. The ATLAS hadronic calorimeter (HCAL) barrel and extended barrel sections are of sampling type, utilising lead absorber interspersed with scintillating tiles. The hadronic end-caps consist of copper plates filled with 8.5 mm thick layers of LAr. Three electrodes divide each LAr layer, providing four drift layers, each approximately 1.8 mm thick. A summary of the ATLAS calorimetry is given in Table 10.1.1. The reconstructed energy of a hadron is expressed using a set of energy independent corrections in the 'Benchmark method' [11] (equation 10.1)

$$E_{rec} = a.E_{had} + E_{em} + b.E_{em}^2 + c.\sqrt{a.E_{had1} + E_{em3}}. \tag{10.1}$$

The coefficient a accounts for ECAL and HCAL differences in their response to pions. A first order correction for non-compensation is made by the quadratic term, and the last term estimates the energy lost in the cryostat between the two calorimeters, E_{em3} is the energy in the last LAr compartment and E_{had1} is the energy measured in the first Tile Calorimeter compartment. The fractional energy resolution may then be parameterised as

$$\sigma/E = (A/\sqrt{E} + B) \oplus C/E. \tag{10.2}$$

Minimising the energy, E, resolution for 300 GeV pions gives coefficient values $A = (59.5 \pm 0.3)$ % GeV$^{1/2}$, $B = (1.8 \pm 0.2)$ % and $C = (2.0 \pm 0.1)$ GeV.

EM Calorimeter	
Barrel $\|\eta\| < 1.475$	
3 Sampling layers	$\Delta\eta \times \Delta\phi$ granularity
1st sampling	0.003×0.1
2nd sampling	0.025×0.025
3rd sampling	0.05×0.025
End-caps $1.375 < \|\eta\| < 3.2$	
$1.5 < \|\eta\| < 2.5$	3 sampling layers
$1.375 < \|\eta\| < 1.5$	2 sampling layers
$2.5 < \|\eta\| < 3.2$	
Hadronic Calorimeter $\|\eta\| < 4.9$	
Barrel $\|\eta\| < 1.0$	
3 sampling layers	$\Delta\eta \times \Delta\phi$ granularity
1st and 2nd samplings	0.1×0.1
3rd sampling	0.2×0.1
Extended Barrel $0.8 < \|\eta\| < 1.7$	
3 sampling layers	$\Delta\eta \times \Delta\phi$ granularity
1st and 2nd samplings	0.1×0.1
3rd sampling	0.2×0.1
End-caps $1.5 < \|\eta\| < 3.2$	
4 Sampling layers	$\Delta\eta \times \Delta\phi$ granularity
$1.5 < \|\eta\| < 2.5$	0.1×0.1
$2.5 < \|\eta\| < 3.2$	0.2×0.2

Table 10.1: The geometry and $\Delta\eta \times \Delta\phi$ granularity of the ATLAS Calorimeters.

10.2 Tracking

The transverse momentum, p_T, of a particle with charge q, may be measured using the curvature of its trajectory in a magnetic field, B. The relationship between the particle p_T and its radius of curvature, r, is given by; $r = p_T/qB$. In practice, this radius is determined by helical fits to a finite set of measured points.

For large values of r, or small traversed angles, the sagitta, s, is given by; $s \approx L^2/8r$ where L is the chord length of the track, see Figure 10.1.

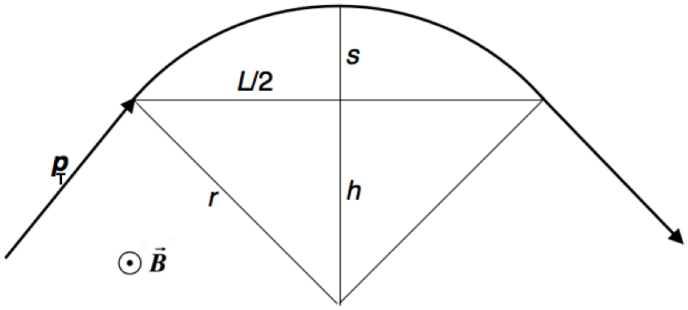


Figure 10.1: The trajectory of a charged particle in the presence of a magnetic field.

The track p_T is related to the sagitta by;

$$p_T = \frac{L^2 qB}{8s}, \left(\frac{1}{s} = \frac{8p_T}{L^2 qB}\right). \tag{10.3}$$

Therefore the fractional p_T resolution, which is given in Equation 10.4, increases linearly with the magnitude of the track p_T.

$$\frac{\delta p_T}{p_T} = \frac{\delta s}{s} = \frac{8\delta s}{L^2 qB} p_T, \tag{10.4}$$

This expression is reasonable for tracks of sufficiently large p_T, for low p_T tracks uncertainty due to multiple scattering becomes significant (Ø $\approx 1.5\%$, [11]). The track p_T resolution is *much* ($\delta p_T \propto p_T^2$) worse at high jet energies for the same x_p. The resolution could be improved by increasing the detection medium traversed (large L) by the particle, or increasing the B-field strength.

Pixel Detector	Radial Extension	5-12 cm
	Detector Medium	50×400 μm Silicon Pads
	Spacial Resolution	10×115 μm
	Number of Channels	80M
	Average number of hits per track	3
SCT	Radial Extension	30-52 cm
	Detector Medium	80 μm×12 cm Stereo Silicon Strips
	Spacial Resolution	17×580 μm
	Number of Channels	6.3M
	Average number of hits per track	8
TRT	Radial Extension	56-107 cm
	Detector Medium	4 mm diameter straw tubes
	Spacial Resolution	130 μm
	Number of Channels	3.5×10^5
	Average number of hits per track	30

Table 10.2: The geometry, media, spacial resolution, number of channels and average number of hits per track of the ATLAS Inner Detector tracker.

10.2.1 ATLAS Tracking

The ATLAS tracker is composed of three sub-detectors, the Pixel Detector (PD), the Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT), known collectively as the Inner Detector (ID) [44]. The inner detector extends 7 m in the z-direction, has an outer radius of 1.15 m and is immersed in a 2 T magnetic field, from the central solenoid. A complete overview of the tracker is given in Table 10.2.

It is the product of charge and *inverse* transverse momentum which may be deduced from the sagitta. The sagitta measurement is itself dependent on the basic positional measurement error, and is thus Gaussian. It is therefore that quantity, q/p_T, and not the transverse momentum, on which the errors are Gaussian [45]. The ATLAS tracking inverse transverse momentum resolution, σ_{1/p_T}, may be parameterised as follows [11],

$$\sigma_{1/p_T} = \sigma_{1/p_T}(\infty)(1 \oplus C_{p_T,equ.}/p_T) \tag{10.5}$$

where $\sigma_{1/p_T}(\infty)$ is the asymptotic resolution expected at infinite momentum and $C_{p_T,equ.}$ is a constant representing the value of p_T for which the intrinsic (first term in bracket) and multiple scattering (second) terms are equal. Specific resolution values are shown in Table 10.3 [11].

Angular coverage	p_T [GeV]	σ_{1/p_T}
$0.25< \|\eta\| <0.5$	44	0.34 TeV^{-1}
$1.5< \|\eta\| <1.75$	80	0.41 TeV^{-1}

Table 10.3: ATLAS tracking inverse p_T resolution.

10.2.2 Jet Energy Resolution

Fundamentally, calorimeter jets are composed of energy deposits from incident hadrons, the fractional errors on which behave as $\approx 1/\sqrt{E_{hadron}}$, hence the jet energy resolution has the same approximate energy dependence. Since the ATLAS calorimeters are calibrated at the EM scale, a correction must be applied to measure energies of hadronic jets. This correction is referred to as the jet energy scale (JES). The JES has a significant systematic uncertainty, which results primarily due to uncertainty in the degree of non-compensation and of the jet electromagnetic fraction, $F(E)$, Ø30%, (due to π_0 production, decaying to photons), [46]. The JES systematic uncertainty is estimated at ATLAS using *in situ* methods and using single pion test beams. Jets are reconstructed with the Anti-k_T algorithm using R parameter values of 0.4 and 0.6. Uncertainties on the amount of material in the ATLAS detector, the description of electromagnetic noise and the Monte Carlo models used are taken into account. The JES systematic uncertainty is less than 2.5% for $60 < p_{Tjet} < 800$ GeV, $|\eta_{jet}| < 0.8$, [46]. Background from additional proton-proton interactions per bunch crossing leads to a further 1.5% per additional interaction for $p_{Tjet} > 50$ GeV. This implies similar systematic measurement errors on $\frac{\delta x_p}{x_p}$ which are small even compared to jet algorithm errors and are ignored for this study.

Chapter 11

Reconstructed Tracks

11.1 Track Selection

The ATLAS tuned Monte Carlo is passed through the Geant4 [28] model of the ATLAS detector. The same standard ATLAS reconstruction algorithms are then run on the simulated detector read out as on real data, resulting in fully reconstructed events, but which also contain a record of the simulation at the truth level.

The fragmentation algorithm was tested on reconstructed Monte Carlo (Pythia 6.4 - ATLAS tune:mc11 [47]) to check its performance with realistic resolutions and reconstruction efficiencies. Migration due to track momentum mismeasurement and track reconstruction inefficiency is corrected along with that due to jet algorithm resolution migration in one step of unfolding. An identical track selection is used as in the ATLAS jet fragmentation paper [16], namely;

- $p_T > 500$ MeV - Tracks of lower momentum suffer a high percentage of energy loss, or large angle scatter and are hence unreliable.
- $|\eta| < 2.4$ - Inner detector (ID) tracking safe angular coverage limit.
- $|d_0| < 1.5$ mm - The impact parameter, the shortest distance from the extrapolated track to the primary vertex. This strongly reduces non-primary tracks.
- $|z_0.sin\theta| < 1.5$ mm - The shortest distance from the extrapolated track to the primary vertex along the z-axis. This offers discrimination not based on extrapolated track bending, but instead on pointing of the track towards the primary vertex.

- At least five hits in the Semiconductor Tracker (SCT) - This ensures tracks are relatively long, since shorter tracks have poorer momentum resolution. It also reduces secondary tracks, which begin further away from the primary vertex.

- At least one pixel hit - Reduces secondaries.

To measure the reconstruction efficiency, track matching links available in the Athena framework [27], have been used to link a given reconstructed track to the corresponding truth particle, to which it is best matched. The distribution of matching probabilities

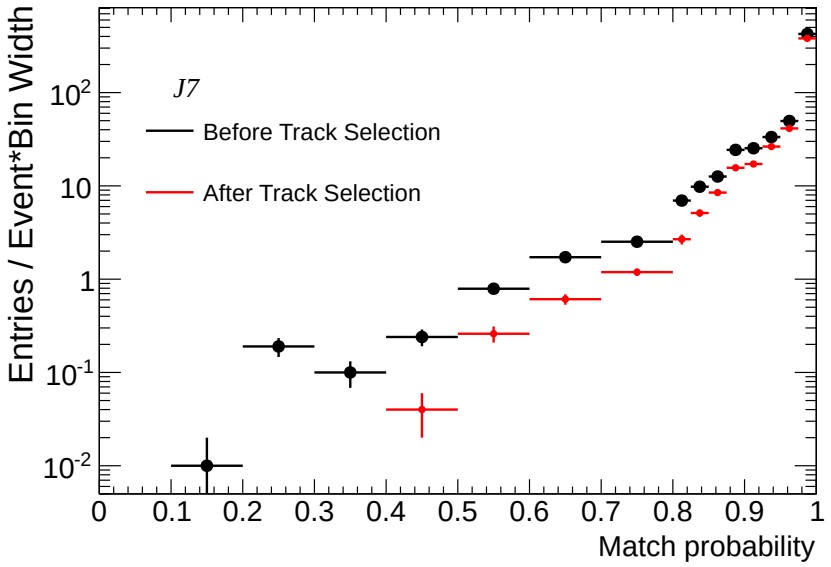
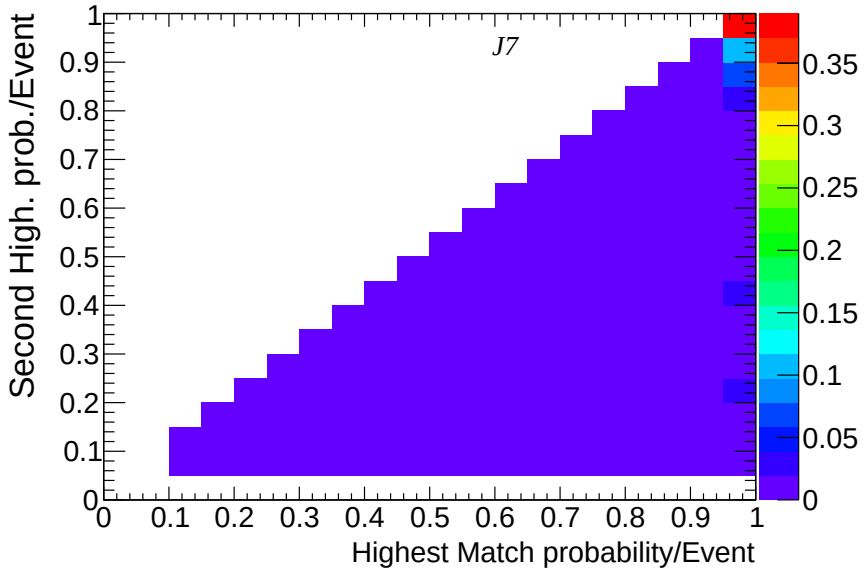


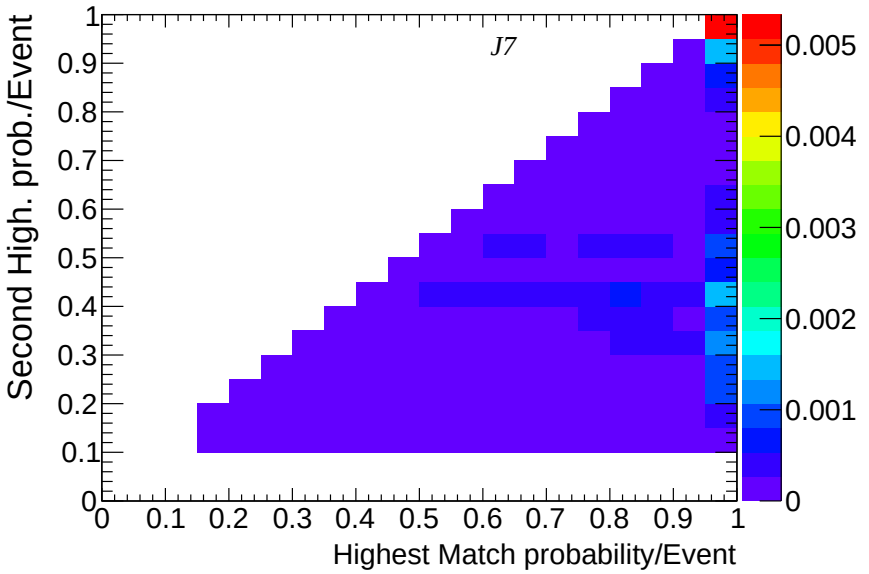
Figure 11.1: The distribution of matching probabilities for truth particles with a single matched track.

between tracks and truth particles is shown in Figure 11.1 at the highest energy scale where tracking is most difficult. This includes the 92% of truth particles with one and only one matched track. The majority of matches have a probability close to one and even more so, as expected, after the track selection is applied.

There may be more than one track matched to a given truth particle. The distribution of the highest and second highest matching probabilities is given for truth particles with more than one matched track in Figures 11.2(a),(b).



(a) Before track selection



(b) After track selection

Figure 11.2: The highest matching probability vs the second highest for truth particles with more than one matched track for the J7 Atlas Pythia sample.

A truth particle which has undergone a small angle secondary scatter may have two shorter reconstructed tracks associated, one before, and one after the secondary scatter, both of which are well matched. The majority of double matches occur with both matches having a probability close to one, indicating such tracks may be a cause.

11.2 Track Reconstruction Efficiency

The efficiency for reconstructing a track matched to a given truth particle is shown in Figures 11.3(a)-(d), for the J3 and J7 samples as a function of the truth particle transverse momentum. Note that (a) resolution implies that the requirement that *tracks* are of at least 500 MeV results in a "soft" cutoff at the truth level, since the requirement is applied at the reconstructed level. In Figures 11.3(b)-11.3(c) it can be seen that the efficiency goes up to approximately 90% after selection, before falling to 75% at around 300 GeV (J7) or 30 MeV (J3), afterwards it rises again to approximately 90% at the highest momentum. This behaviour may result due to higher p_T jets in a given p_T range being composed of only few high momentum tracks, which then are reconstructed well since they are in a relatively sparse region of the detector. Conversely the lower p_T jets in that sample, comprising of many more lower p_T tracks which may be more poorly reconstructed due to now crowded regions.

Significant correction for efficiency loss is needed, predominantly in the lowest x_p bin due to the low efficiency at approximately 500 MeV.

The spectra and efficiencies as a function of η and ϕ are shown in Figures 11.4(b)-(d). The efficiencies are flat as a function of ϕ but reduced at large $|\eta|$, probably due to the larger traversed beam pipe thickness at small angles and thus limited track acceptance.

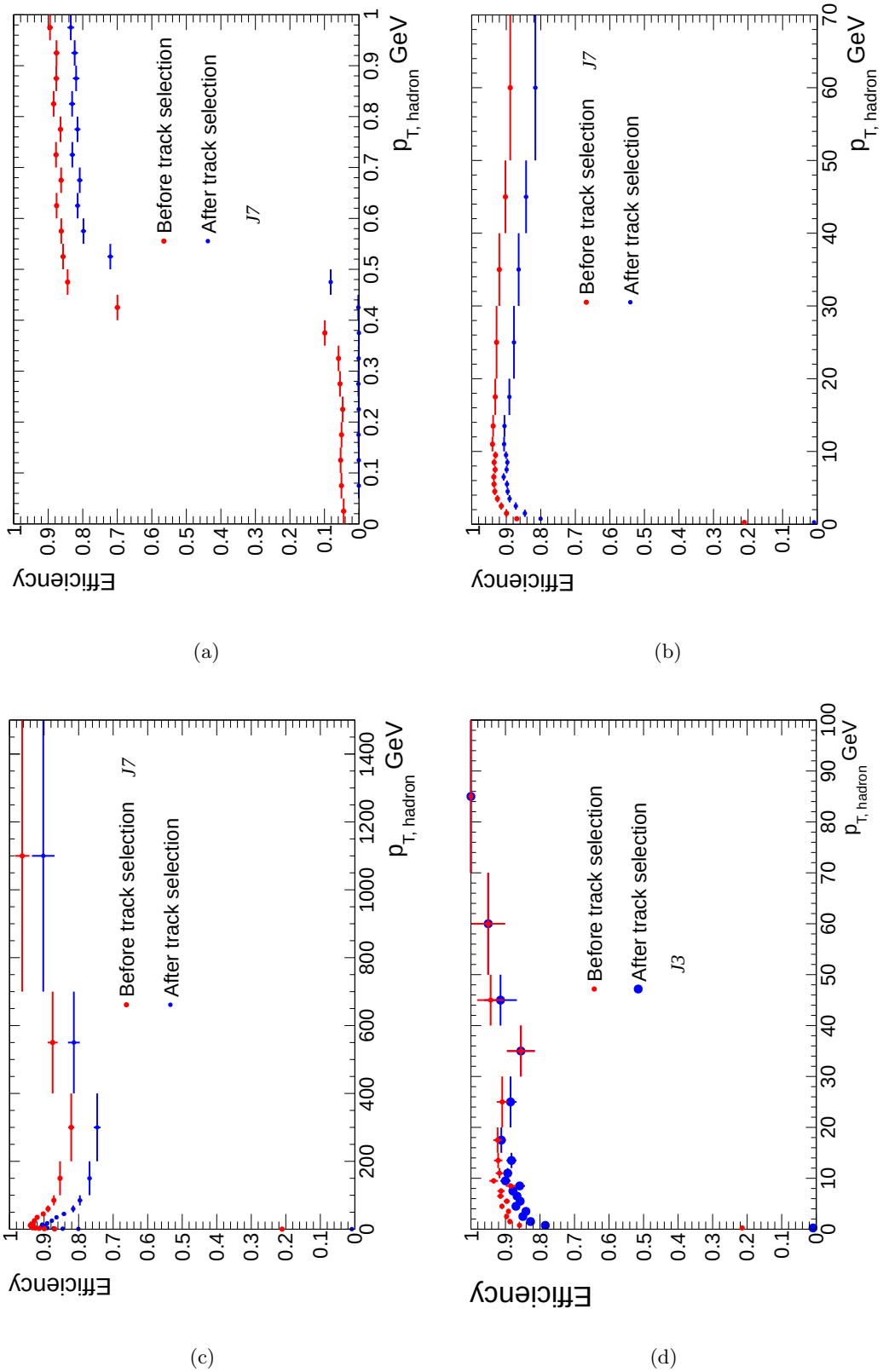
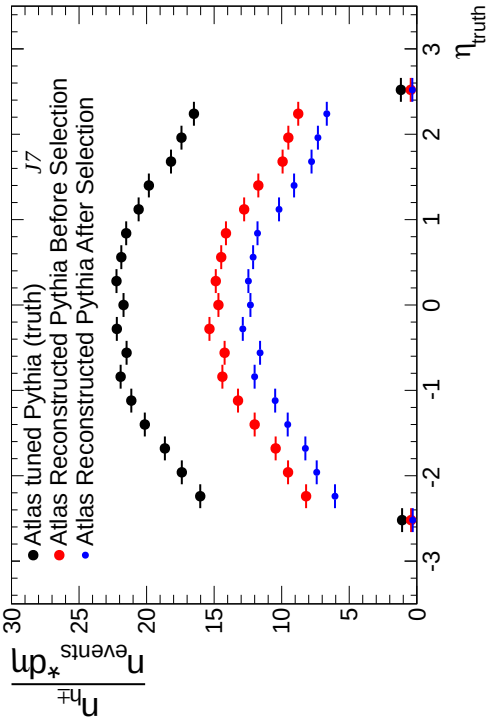
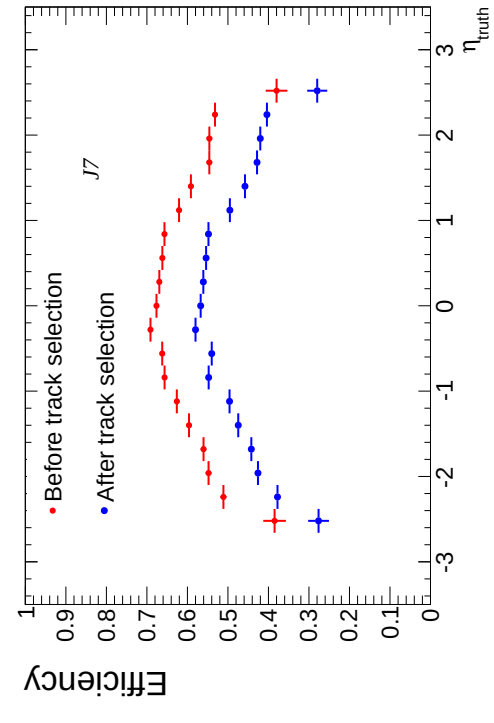


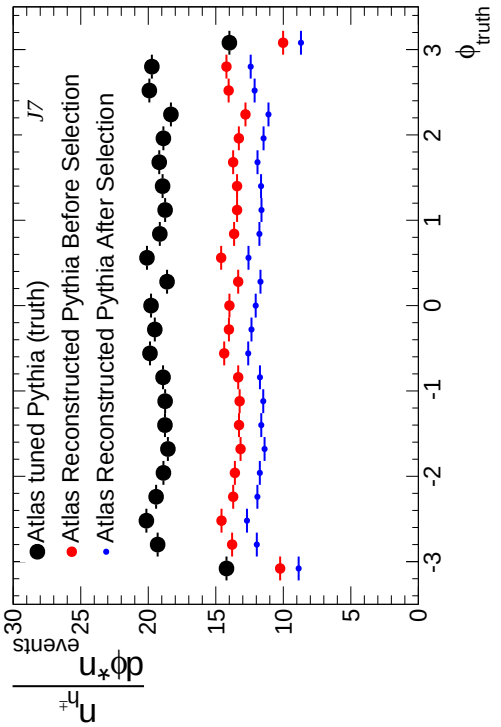
Figure 11.3: Efficiency of reconstructing a track given a truth particle, as a function of the truth particle momentum. Measured before and after the track selection, on the J7 sample (Figures a, b, and c) and the J3 Atlas Pythia sample (d).



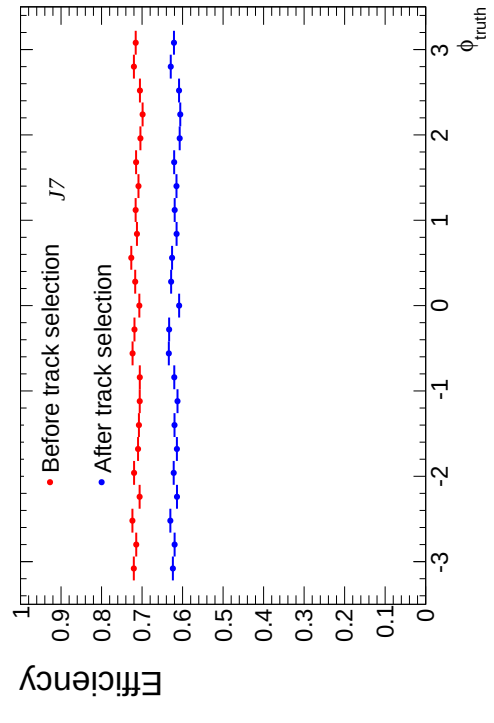
(a) Pseudorapidity spectra of charged hadrons and reconstructed tracks



(b) Efficiency of reconstructing a track given a truth particle, as a function of the truth particle pseudorapidity



(c) Azimuthal spectra of charged hadrons and reconstructed tracks



(d) Efficiency of reconstructing a track given a truth particle, as a function of the truth particle azimuthal angle

Figure 11.4: Spectra and efficiencies, shown before and after the track selection is applied, measured on the J7 Atlas Pythia sample.

11.3 Unfolding of Track Momenta

The resolution involved in measuring tracks causes migration of x_p. The algorithmic migration associated with measurement of the parton estimator has already been corrected for by unfolding, now unfolding is used to correct for track momentum resolution. This is complicated by the limited track reconstruction efficiency and noise. For the purposes of observing and hence correcting for migration and efficiency losses without complication from migration from any other source, the truth supplied values of the fragmenting partons momenta are used and the fragmentation algorithm is run with a unit cone.

The tracks passing the selection are unfolded using TUnfold with a response matrix filled by matching tracks to truth particles in the way described above. The underflow bin of the response matrix is filled with truth particles which do not have a matched track passing the selection, this underflow bin is used in TUnfold to correct for track reconstruction inefficiency. The results when tracks passing the selection are unfolded and efficiency corrected are shown in Figure 11.5.

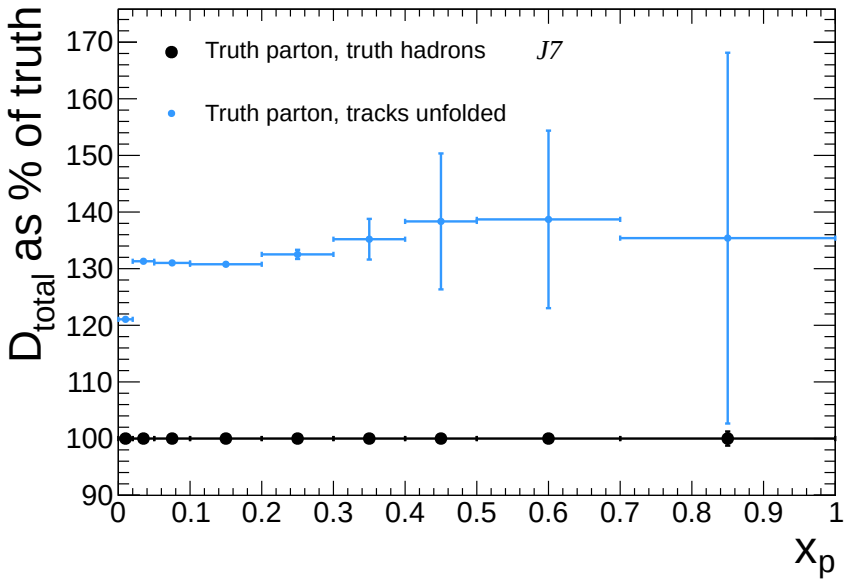


Figure 11.5: The unfolded distribution of hadronic momenta of reconstructed tracks with efficiency loss correction, shown as a percentage of the truth particle distribution. Measured on the J7 Monte Carlo sample using a unit cone.

The unfolded result on is average 30% larger than the truth result, due to noise. Noise is defined as tracks originating from interaction with the detector and decays in flight as well as tracks that migrate into a measuring cone due to directional resolution. The first two types of noise are known as secondaries and are identified using matching to a list of

truth particles labeled as secondary. A fourth type of noise is henceforth defined as all non-pions tracks, i.e. the Monte Carlo will be used to make a correction to the charged hadron fragmentation function, to give the charged pion fragmentation function. The percentage of charged hadrons which are pions, in a unit cone about the truth parton direction, is shown in Figure 11.6. A larger fraction of charged hadrons are pions in the lower x_p intervals at low hard scale (J3). This remains to be true at larger hard scale (J7), however, there are more non-pions at high x_p, and fewer at low x_p at lower hard scale. Decays of unstable non-pions would produce more tracks with lower x_p values. In any

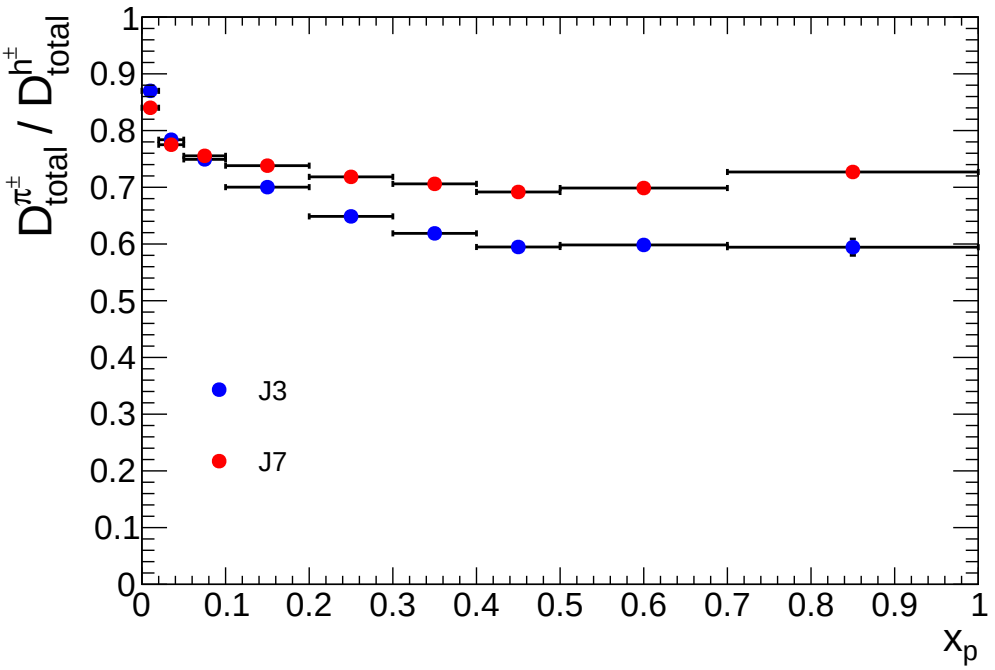


Figure 11.6: The fraction of all charged hadrons which are pions, in a unit cone centered on the truth parton direction, calculated at the truth level using Pythia J3 and J7 Monte Carlo samples.

two, or indeed multi, particle production process, of which $q^* \rightarrow h_1, h_2...$ is an example, kinematically the heaviest particle will take a greater share of the momentum. Clearly this is a scale dependent statement since, when *all* particles are relatively massless, they will take equal shares - thus at J7 there is a flatter curve.

The above curve would produce a flatter fragmentation function $D(x_p)$ for pions rather than for all charged hadrons. Kinematically decays can not affect $x_p = 1$ particles. The effect is cumulative as x_p gets larger. So this is a probable cause for this (Figure 11.6) curve. If this curve is that of a kinematic decay effect rather than a dynamic parton shower

effect, then it could be the cause of spurious scaling violations. Such scaling violations have been claimed in [1].

The lowest x_p bin in Figure 11.5 is 10% smaller than the others. This is due to the track selection cut at 500 MeV and the fact that the acceptance correction corrects for missed pions but not for missed non-pions. Missed non-pions are not corrected for since they would be removed again as noise, and hence the total error involved would be unnecessarily increased.

The percentage of tracks which are noise, including non-pions, is shown in Figure 11.7. Note that the total noise level of around 30% (J7) is in agreement with the excess

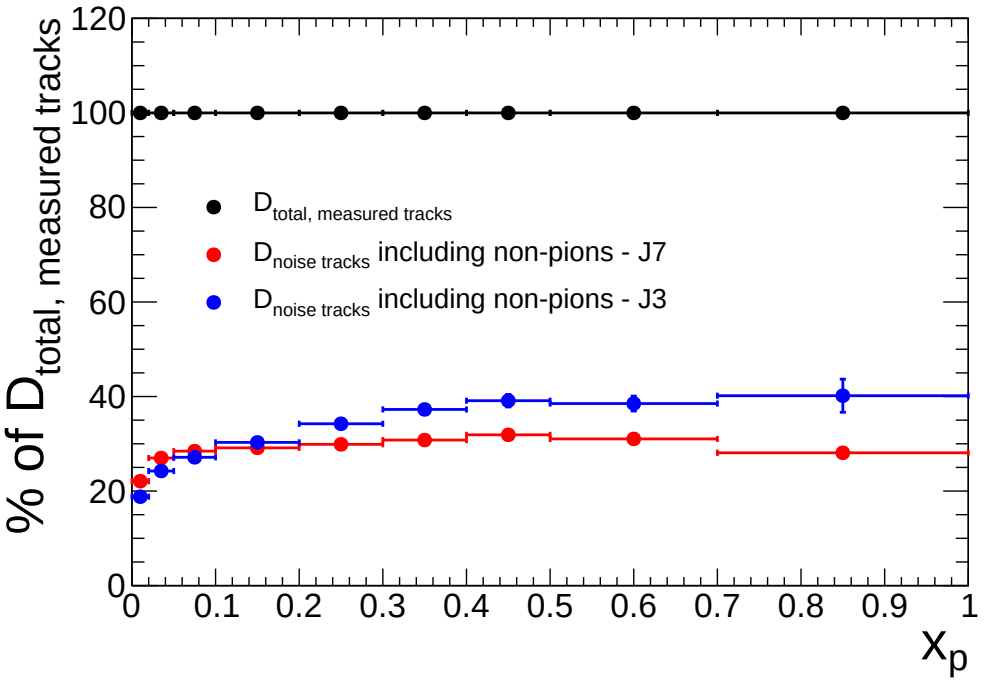


Figure 11.7: The noise x_p distribution as a percentage of that of the measured tracks passing the selection. Calculated with the J3 and J7 Monte Carlo samples using a unit cone.

observed in Figure 11.5. This noise is removed by measuring the ratio of pions, to pions plus noise at the truth level, denoted R_{noise}.

$$R_{noise} = \frac{N_{\pi}}{N_{\pi+noise}} \tag{11.1}$$

In order to correct for migration due to track direction resolution, the particles which are selected in a unit cone about the parton estimator (the truth parton in this instance), are selected based upon the direction of the track to which they are matched. The ratio R_{noise} is shown in Figure 11.8.

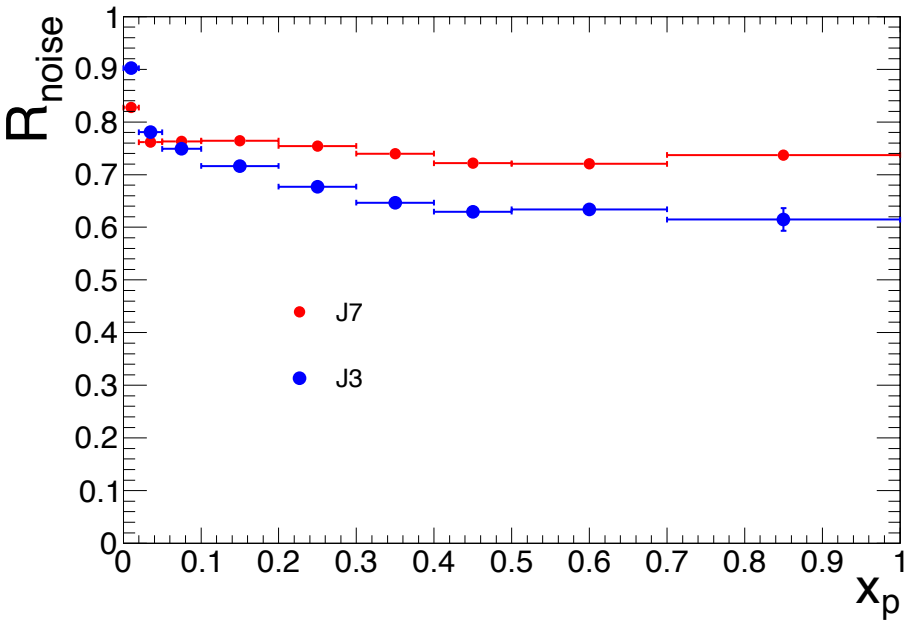
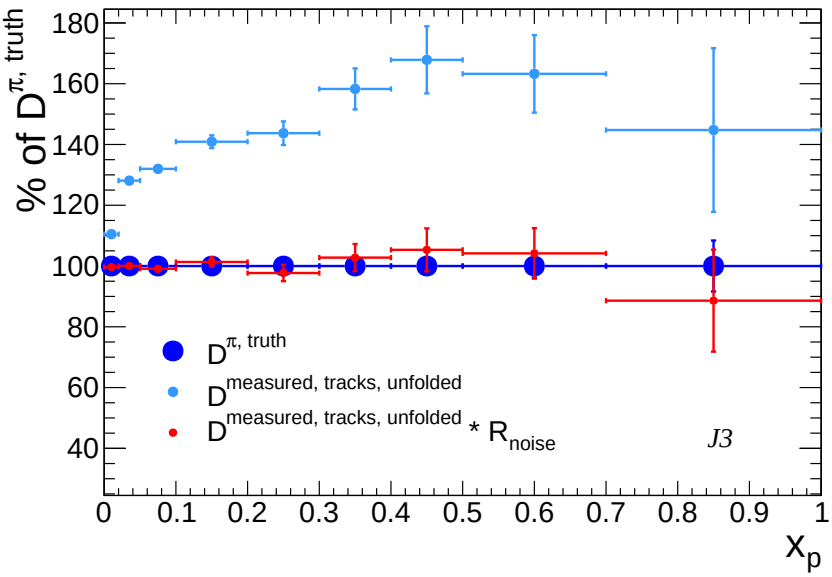
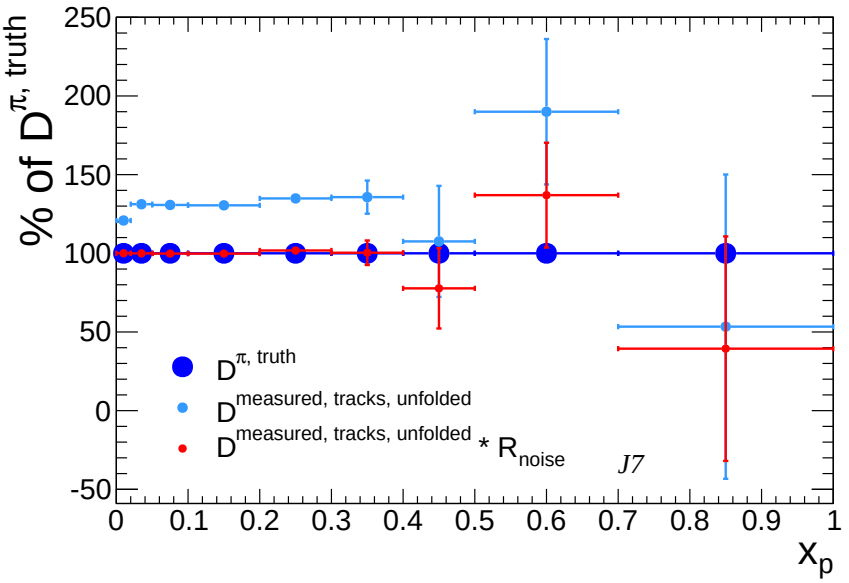


Figure 11.8: The ratio R_{noise} measured on the J3 and J7 Monte Carlo samples with a unit cone.

The results after unfolding and scaling by R_{noise} are shown in Figures 11.9(a) and 11.9(b).



(a) J3



(b) J7

Figure 11.9: The unfolded quark to pion fragmentation function using measured level tracks, before and after removal of noise, as a percentage of the truth level quark to pion fragmentation function. Reconstructed level Pythia Monte Carlo has been used.

11.4 Bin-by-bin Correction

The TUnfold method is unstable as described in Section 9.4.3. A simpler bin-by-bin method of unfolding has also been used to see if it can do better. The method guarantees positivity and does not suffer the same instability as TUnfold.

The ratio ("truth" pions)/(raw measured level) is measured as a correction factor, using reconstructed level Monte Carlo. This is then applied to the raw measured 'data' (in practice reconstructed level Monte Carlo). This way noise (including non-pions), acceptance and migration are corrected for in a single step.

To test the method 90% of the available Monte Carlo events are used to measure the correction vector and the remaining 10%, representative of typical data samples, are corrected as pseudo-data, see Figure 11.10. The corrected "data" results are consistent

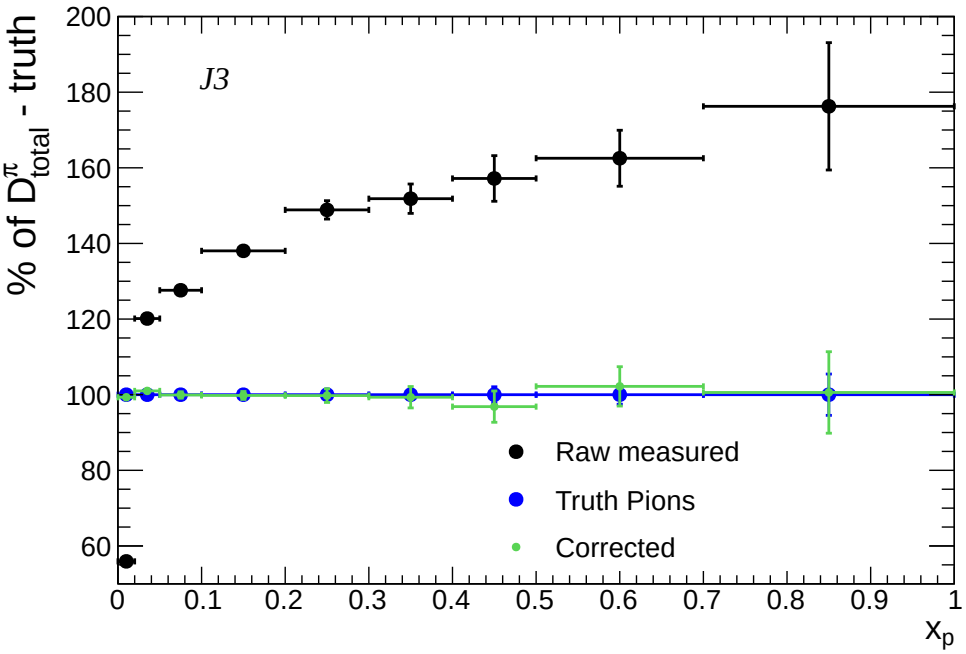


Figure 11.10: The fragmentation function measured with unit cone at the raw measured level using Anti-k_T jets as input as a percentage of the truth pion fragmentation function, measured with Pythia (J3) Monte Carlo. Also included is the bin-by-bin corrected result.

with the truth values within statistical uncertainties. The bin-by-bin unfolding is more stable than the TUnfold method. However the errors are suspiciously small.

x_p	Efficiency		Purity	
	Anti-k_T	TRAPS	Anti-k_T	TRAPS
$0 \leq x_p < 0.02$	94%	92%	96%	96%
$0.02 \leq x_p < 0.05$	86%	83%	85%	81%
$0.05 \leq x_p < 0.1$	83%	78%	81%	76%
$0.1 \leq x_p < 0.2$	85%	80%	81%	76%
$0.2 \leq x_p < 0.3$	75%	69%	72%	63%
$0.3 \leq x_p < 0.4$	65%	60%	64%	54%
$0.4 \leq x_p < 0.5$	57%	53%	55%	44%
$0.5 \leq x_p < 0.7$	67%	66%	68%	54%
$0.7 \leq x_p \leq 1.0$	58%	75%	59%	42%

Table 11.1: Efficiencies and purities in the chosen analysis intervals using the respective parton estimators at low transverse momentum (J3 Monte Carlo). Statistical errors are well below the 1% level.

x_p	Efficiency		Purity	
	Anti-k_T	TRAPS	Anti-k_T	TRAPS
$0 \leq x_p < 0.02$	99%	98%	99%	100%
$0.02 \leq x_p < 0.05$	91%	90%	89%	87%
$0.05 \leq x_p < 0.1$	81%	81%	81%	79%
$0.1 \leq x_p < 0.2$	69%	70%	72%	70%
$0.2 \leq x_p < 0.3$	42%	42%	45%	43%
$0.3 \leq x_p < 0.4$	27%	27%	28%	26%
$0.4 \leq x_p < 0.5$	19%	19%	17%	16%
$0.5 \leq x_p < 0.7$	25%	26%	16%	16%
$0.7 \leq x_p \leq 1.0$	58%	60%	4%	4%

Table 11.2: Efficiencies and purities in the chosen analysis intervals using the respective parton estimators at high transverse momentum (J7 Monte Carlo). Statistical errors are well below the 1% level.

11.5 Measured Level x_p Bin Efficiencies and Purities

The efficiencies and purities in the chosen x_p analysis intervals, measured as defined in Section 9.3 but now using *reconstructed* tracks within a unit cone about the reconstructed jet or parton estimator (η, ϕ) position are shown in Tables 11.1 and 11.2 at extreme values of the hard scale. As before, quarks have been selected for the matched parton. Acceptance losses are now a consideration, for example, a truth particle which was in the signal cone may no longer be after the reconstruction, i.e. directional resolution of the tracks is now an issue, though, as before, directional resolution of the jets is absorbed in the radial dependence.

Migration of x_p values now occurs due to mis-estimation of the parton *and/or* the track momenta. Measurement of track momenta and the inputs to the exterior jet algorithm introduces an additional possible cause of calculated values of x_p which are unphysically

larger than unity. Any such calculated values (which amount to less than 1% of entries in the highest x_p bin) are dealt with in the same fashion, i.e. by setting their value to be 0.999. At worst, this still improves the resolution.

Low x_p values (from low momenta, well resolved tracks in the numerator) are better resolved at larger jet momenta, where calorimeter resolution is also better to improve the denominator also. Migration into the higher x_p intervals remains the dominant problem and is most severe at high p_T, in the top four x_p bins, where track momentum uncertainty is largest.

11.5.1 Measured vs Truth Efficiencies and Purities

Referring to Tables 11.1 and 11.2 and the truth equivalents Tables 9.2 and 9.3, at low p_T the efficiencies and purities are approximately 4% worse at the measured level than at the truth level in the lowest x_p bins, and approximately 20% worse in the mid-higher bins, where tracking resolution is a problem. The high p_T lowest x_p efficiencies and purities are only decreased by approximately 1% at the measured level with respect to the truth level, however there is a major degradation in the highest x_p interval, efficiencies and purities achieved are of the order of 90% at the truth level and approximately 20%(efficiency) and 5%(purity) at the measured level.

11.5.2 Anti-k_T and TRAPS Comparison

Similar efficiencies and purities are achieved with both algorithms at high p_T, however at low p_T the situation is slightly less straightforward. Higher values (approximately 5% better) are achieved with the Anti-k_T algorithm in all but the highest x_p interval, where the efficiency is 15% higher using TRAPS and the purity is significantly (40% vs 60%) better doing so.

For efficiencies throughout, and for purities at the low momentum, the absolute percentage differences between estimators remain about the same as they were at the truth level, but the smaller numbers at the reconstructed level mean bigger *proportionate* difference.

The Anti-k_T algorithm remains to be the preferred choice, but the resolutions of both parton estimators at the measured level are just sufficient for a measurement with the same binning as used in previous measurements in all but the higher x_p bins at high p_T.

Analysis with wider bins at high x_p could be used to reduce uncertainties here whilst new, narrower bins (e.g. as in the jet analysis, Figure 13.1) could be used at low x_p with no significant degradation.

Chapter 12

Systematic Error in Resolution Unfolding

12.1 Monte-Carlo Model Systematics

Unfolding necessarily depends on a model and the best way of validating a model is to use an alternative model. Ideally, n different models (*which describe the data*) would be tested, giving $(n(n-1))$ independent combinations to be entered into a distribution, the associated RMS would be used as the systematic uncertainty estimate. However, reconstructing millions of events with many models would be unfeasible.

Thus to assess the systematic uncertainty associated with using "truth" information to unfold the jet algorithm and tracking resolutions, one model (HERWIG) is used to correct another (Pythia) and the results are compared to that model (Pythia) correcting itself. If both models give a good description of the resolving process, then correcting one model with the other should yield unfolded results compatible with those from the first model correcting itself. If incompatibilities are observed when the above comparison is made, any differences are interpreted to be due to a Monte Carlo systematic uncertainty. The model which best describes the measured data would be chosen, and to be safe, an error of $\pm$ the difference used as the uncertainty, since a further alternate model might have produced a similar difference in the other direction.

12.2 Model Differences

The model differences have been compared at the truth and measured levels (see Figures 12.1 and 12.2) for all parton interactions.

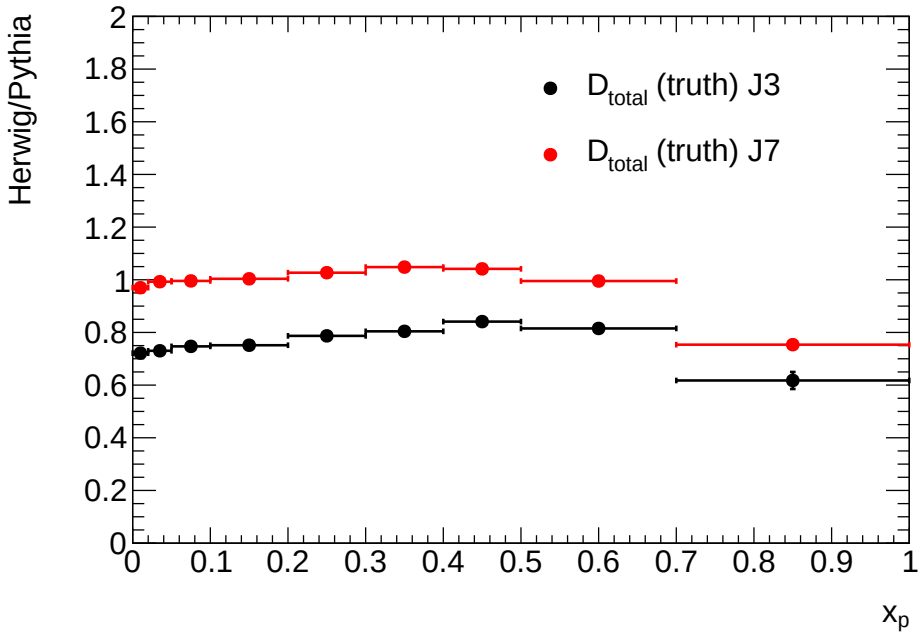


Figure 12.1: The ratio of tracks (HERWIG/Pythia) at small and large hard scale (J3 and J7 Monte Carlo) in the signal unit cone and background solid angle at the truth level.

A unit cone has been used, since this is sufficiently large to measure most of the correlated tracks - the ratio of tracks sampled in Pythia/HERWIG should be the close to that of optimal solution. At the truth level, there are on average more charged hadrons in Pythia than HERWIG in the D_{total} sampling unit cone (see Figure 12.1), and less in the corresponding background solid angles. At large hard scale (J7), there are 5% more low x_p tracks in Pythia than HERWIG in the D_{total} sampling unit cone and almost 25% more in the highest x_p range at the truth level.

At large hard scale (J7), but at the raw measured level (Figure 12.2), there are 7% more low x_p and 6% less mid-x_p range tracks in Pythia. The differences are larger at small hard scale (J3) and are smallest in the mid-x_p range.

12.3 Systematic Uncertainties of Unfolding Techniques

TUnfold and bin-by-bin methods are used to correct for resolution. The systematic differences associated with using the Monte Carlo model to unfold with matrix and bin-by-bin

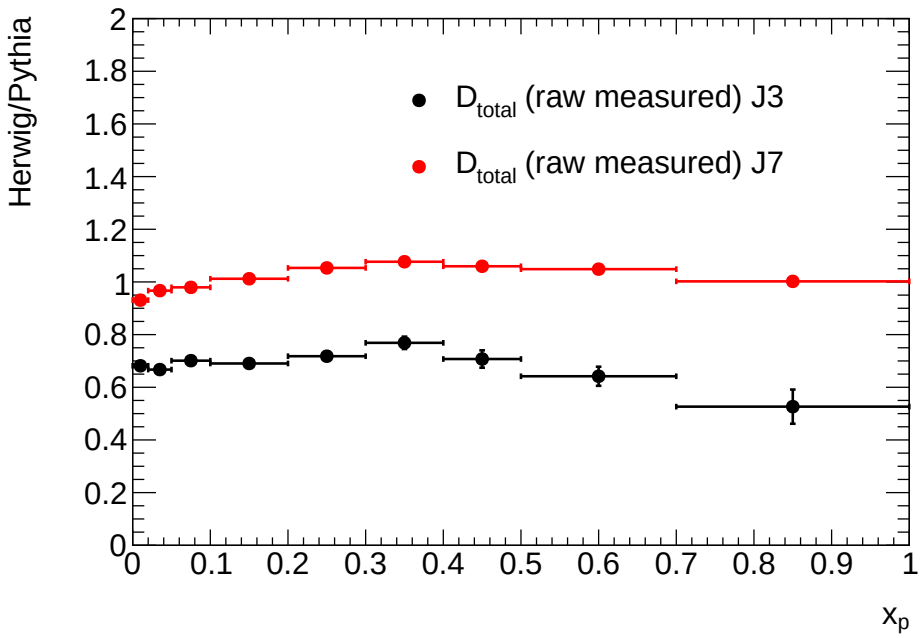


Figure 12.2: The ratio of tracks (HERWIG/Pythia) at small and large hard scale (J3 and J7 Monte Carlo) in the D_{total} unit cone at the measured level.

methods are shown in Table 12.1. The model differences as estimated using the two correctional methods are compatible at low x_p - see Figures 12.3 and 12.4. The bin-by-bin method has a smaller associated systematic uncertainties at the highest x_p compared to the TUnfold method, the latter of course involving separate noise removal.

The TUnfold differences, at low x_p, are generally of the order 12%, 5% and 2%, for the J3, J4 and J7 samples respectively. They are much larger in the higher x_p bins, due to low efficiencies and purities as seen in both correctional methods, though TUnfold is clearly unstable, probably due to low efficiency and purity. Note the severely low purity ($\sim$5%) and low efficiency ($\sim$40%) in the top bin in Table 9.3.

While the systematic error differences are large at the lowest p_T, where the jet resolution is poorest, they are largest at high p_T and x_p, where tracking resolution is poorest and dependence on unfolding is largest.

Clearly the fragmentation algorithm is sensitive to the hadronisation differences between the two models. Some of the higher x_p bins may be combined to achieve resolution sufficient to make a data measurement feasible. Given that the raw measured level ratio of Pythia/HERWIG is not compatible with unity in many of the x_p bins, at least one of these models must not be correctly describing nature.

	Systematic Uncertainty(Error on difference)				
	TUnfold Method			Bin-by-bin Method	
x_p range	J3	J4	J7	J3	J7
$0 \leq x_p < 0.02$	-12(±3)%	+9(±1)%	+2(±1)%	+6(±1)%	+4(±1)%
$0.02 \leq x_p < 0.05$	+16(±1)%	+5(±1)%	+1(±1)%	+10(±1)%	+2(±1)%
$0.05 \leq x_p < 0.1$	+9(±1)%	+5(±1)%	+1(±1)%	+8(±1)%	+1(±1)%
$0.1 \leq x_p < 0.2$	+13(±2)%	+7(±1)%	-1(±1)%	+9(±1)%	-1(±1)%
$0.2 \leq x_p < 0.3$	+13(±4)%	+6(±2)%	-7(±3)%	+11(±3)%	-3(±1)%
$0.3 \leq x_p < 0.4$	+20(±7)%	+2(±5)%	-4(±12)%	+10(±4)%	-3(±1)%
$0.4 \leq x_p < 0.5$	+11(±8000)%	+3(±8)%	-9(±63)%	+15(±6)%	-4(±2)%
$0.5 \leq x_p < 0.7$	+18(±13)%	+3(±9)%	-32(±45)%	+28(±8)%	-6(±2)%
$0.7 \leq x_p \leq 1.0$	+48(±67)%	+28(±19)%	+322(±175)%	+17(±15)%	-25(±2)%

Table 12.1: Systematic uncertainties due to the Monte Carlo Model. The systematic uncertainty quoted is the difference of the subtraction of the Pythia corrected result from the HERWIG corrected result.

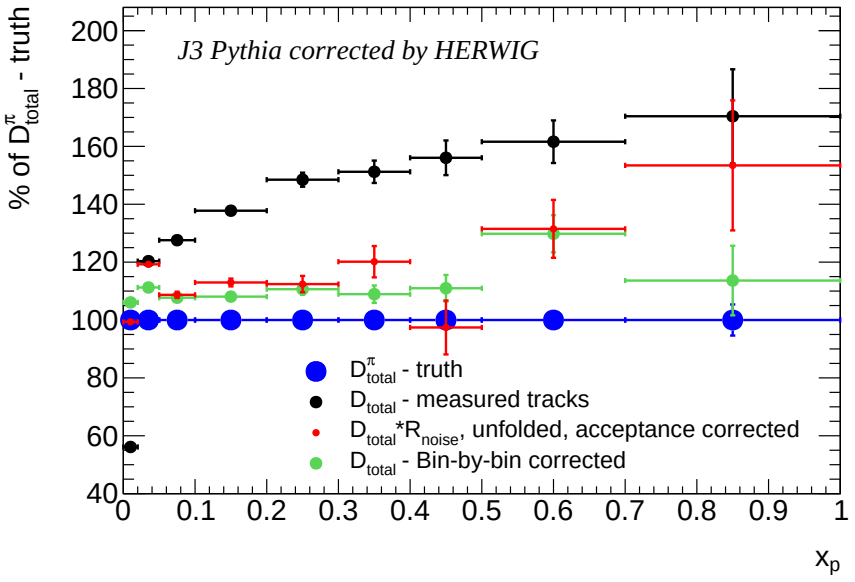


Figure 12.3: Raw measured level D_{total} distribution measured in Pythia with a unit cone, corrected by the bin-by-bin and noise removal/unfolding methods with correctional vectors/matrices measured with HERWIG, at small hard scale (J3 Monte Carlo).

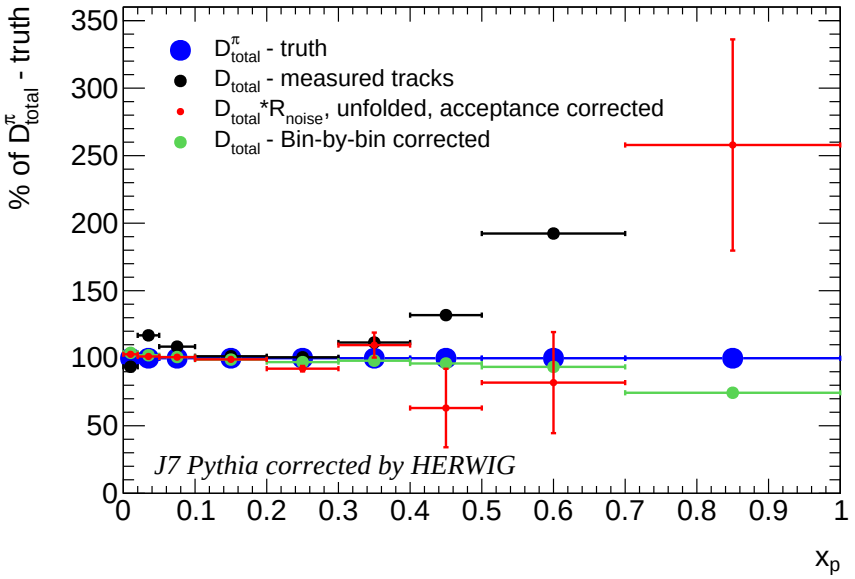


Figure 12.4: Raw measured level D_{total} distribution measured in Pythia with a unit cone, corrected by the bin-by-bin and noise removal/unfolding methods with correctional vectors/matrices measured with HERWIG, at large hard scale (J7 Monte Carlo).

The errors associated with the instabilities of the TUnfold method are being accounted for twice in the present comparison of Pythia unfolded by Pythia, to Pythia unfolded with HERWIG. It is argued that this gives an overestimation of the systematic uncertainty. In an attempt to measure a smaller - more fair estimate, a comparison of Pythia truth to Pythia unfolded with HERWIG is done. The resulting quantified model differences are show in Table 12.2. Some over estimation of this error is removed from the systematic uncertainties estimation by comparing Pythia unfolded with HERWIG to Pythia truth.

	Systematic Uncertainty(Error on difference)			
	TUnfold Method		Bin-by-bin Method	
x_p range	J3	J7	J3	J7
$0 \leq x_p < 0.02$	-12(±3)%	+2(±1)%	+8(±1)%	+3(±1)%
$0.02 \leq x_p < 0.05$	+16(±1)%	+1(±1)%	+8(±1)%	+2(±1)%
$0.05 \leq x_p < 0.1$	+9(±1)%	+1(±1)%	+7(±1)%	+1(±1)%
$0.1 \leq x_p < 0.2$	+13(±2)%	-1(±1)%	+8(±1)%	-1(±1)%
$0.2 \leq x_p < 0.3$	+13(±4)%	-7(±3)%	+11(±1)%	-3(±1)%
$0.3 \leq x_p < 0.4$	+20(±7)%	-4(±12)%	+9(±2)%	-3(±1)%
$0.4 \leq x_p < 0.5$	+11(±8000)%	-9(±63)%	+15(±3)%	-4(±1)%
$0.5 \leq x_p < 0.7$	+18(±13)%	-32(±45)%	+27(±3)%	-6(±1)%
$0.7 \leq x_p \leq 1.0$	+48(±67)%	+322(±175)%	+22(±7)%	-26(±2)%

Table 12.2: Systematic uncertainties due to the Monte Carlo Model. The systematic uncertainty quoted is the difference of the subtraction of the Pythia truth result from the HERWIG corrected result.

	Systematic Uncertainty(Error on difference)			
	TUnfold Method		Bin-by-bin Method	
x_p range	J3	J7	J3	J7
$0 \leq x_p < 0.02$	-18(±1)%	+1(±1)%	+8(±1)%	+3(±1)%
$0.02 \leq x_p < 0.05$	+12(±1)%	+1(±1)%	+8(±1)%	+2(±1)%
$0.05 \leq x_p < 0.1$	+7(±1)%	0(±1)%	+8(±1)%	+1(±1)%
$0.1 \leq x_p < 0.2$	+8(±1)%	+1(±1)%	+8(±1)%	-1(±1)%
$0.2 \leq x_p < 0.3$	+9(±3)%	+5(±2)%	+12(±1)%	-2(±1)%
$0.3 \leq x_p < 0.4$	+23(±5)%	+29(±10)%	+13(±2)%	-1(±1)%
$0.4 \leq x_p < 0.5$	-3(±9)%	+72(±30)%	+12(±3)%	-2(±1)%
$0.5 \leq x_p < 0.7$	+25(±9)%	+46(±38)%	+22(±3)%	-4(±1)%
$0.7 \leq x_p \leq 1.0$	+15(±20)%	+80(±80)%	+7(±7)%	-27(±2)%

Table 12.3: Systematic uncertainties due to the Monte Carlo Model. The systematic uncertainty quoted is the difference of the subtraction of the Pythia truth result from the weighted HERWIG corrected result.

However the differences and their associated errors remain to be large.

The two models don't describe one another, as demonstrated in Figures 12.1 and 12.2. In a further attempt for a smaller, more suitably measured systematic uncertainty, the failure of the HERWIG 'model' to describe the Pythia 'data' is addressed. As an alternative to the impractical method to address such differences, namely re-tuning the Monte Carlo, instead HERWIG is weighted to describe Pythia when filling the unfolding tools. Consequently, the differences and associated errors are slightly reduced, see Figure 12.3. The largest improvement is at high x_p, where the differences were largest.

12.4 Comparison of Uncertainties

It is not completely understood why a more crude method ought to have smaller associated uncertainties. The bin-by-bin method clearly has more consistency within errors from bin to bin. If the TUnfold package had been used with increased statistics, perhaps its behaviour would be more stable and the more sophisticated method might have smaller associated errors. The larger uncertainties are a sign of over-parameterisation or lack of constraint of, for example, positivity.

Chapter 13

Jet Fragmentation Comparison

The pion content correcting factor may be relaxed to provide the inclusive *charged hadron* results of the parton fragmentation algorithm at the measured-unfolded level to compare with those from the ATLAS *jet* fragmentation paper [16], see Figure 13.1. The jet fragmentation measurement is longitudinal, as defined in Section 5.1, this would give a softer function, since $x_L < x_p$. For these comparison purposes, all partons are included and a similar jet selection is used, (400-500 GeV p_T of $R = 0.6$, Anti-k_T algorithm, selected from the J5 Pythia reconstructed Monte Carlo). The same track and event selections are applied in the present study, and "data" corrected (using TUnfold) as previously discussed.

The binning scheme used in the present study is motivated by comparison (at low hard scale) to relevant data, as in the previous section. Taking into account the measured efficiencies and purities (Section 11.5), finer binning would be used in a data measurement at low x_p, where average multiplicity is large and resolution is sufficient. To measure with reasonable uncertainty given the low efficiency and purity at high p_T and x_p, more coarse binning (than here) would be used.

There appears to be a statistically significant slightly sharper fall at low x_p (FAPS), which is contrary to that which would be expected given the differences of x_p and x_L, but the agreement of the *jet* and *parton* functions is remarkable, even in the lowest x_p range, where uncorrelated background is considerable and a large angle sampling cone is required to measure all of the hadrons correlated with a parton's direction, illustrated in Figure 13.2. It may be seen in the figure, that measuring the total number of hadrons confined to a jet cone of R=0.6, without background subtraction, gives a fortuitously

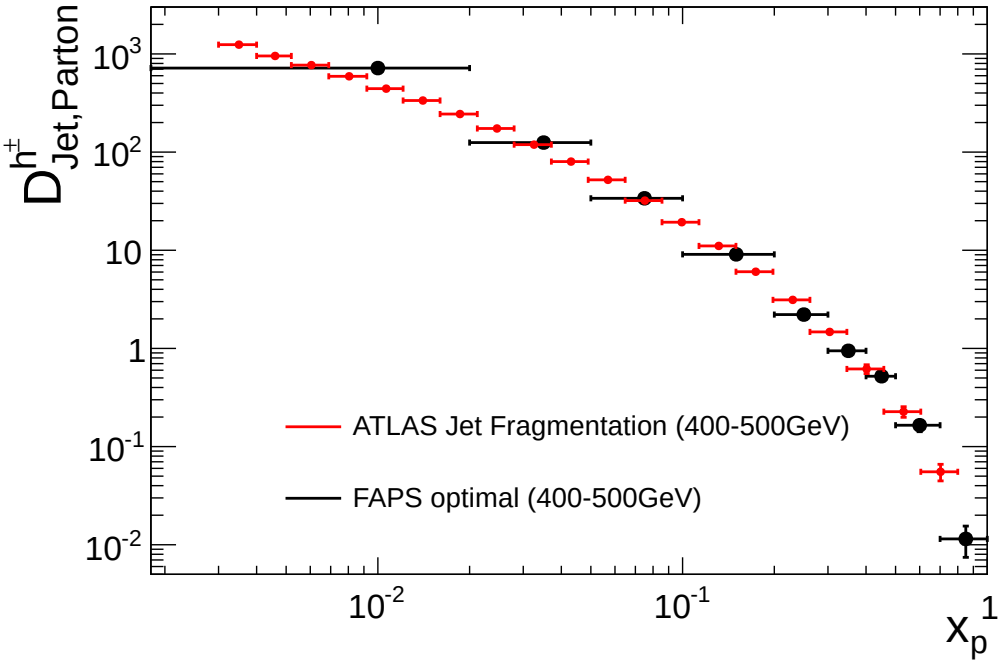


Figure 13.1: Official ATLAS *jet* fragmentation function (400-500 GeV) data measurement and FAPS *parton* fragmentation function (400-500 GeV reconstructed Monte Carlo).

good estimate of the total number of correlated signal hadrons.

This might be the reason for the popular use of the (R=0.6) Anti-k_T algorithm jets. Of course, in using such a basic definition for a fragmentation measurement, no claim of *parton* measurement could be justified.

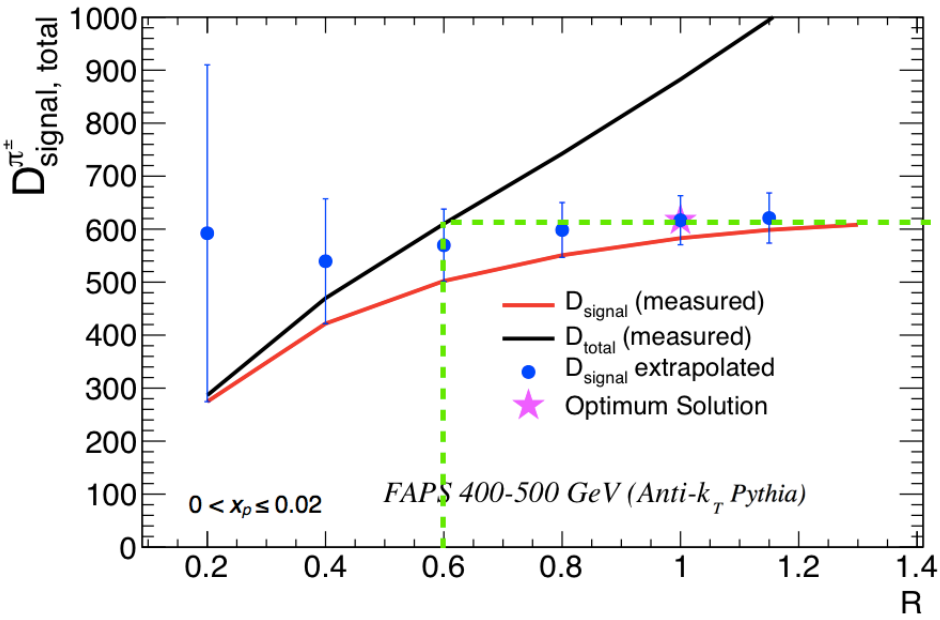


Figure 13.2: The (FAPS) measured fragmentation function as a function of measuring angle, R, in the range $0 < x_p \leq 0.02$, at the D_{total}, D_{signal} and $D_{signal,extrapolated}$ levels. The optimum solution is also shown.

Chapter 14

Preparation For Data Analysis

In the first section of this chapter the selection of well measured jets but unbiased which may later be unfolded for resolution is demonstrated. Then, in the next section, data selection techniques which could be used to measure the fragmentation function in data are described.

Data collected in a low efficiency range of a trigger might be biased or not well measured. Even in the optimally efficient transverse momentum range of a jet trigger, calorimeter noise, for example, may result in fake and poorly resolved jets. Such jets could be simulated in Monte Carlo for unfolding purposes but removal from data results in smaller uncertainties.

14.1 Event Cleaning

Finite measurement precision gives all measured objets associated resolutions and component malfunction is inevitable in such a vast measuring apparatus as ATLAS. Cuts are used to select runs of high quality and then to form a dividing line between those measured objects which are reasonably well resolved from those too hopeless to unfold, and thus must be compensated for with acceptance corrections. In this chapter, selection methods are applied to jets in ATLAS data and Monte Carlo. The experience was gained in summer 2010 when selection cuts that were to be used in an early paper on jet cleaning, [48] were validated. The work shows practical examples of bad measurements and how well measured key variables are used to choose appropriate selection cuts.

14.1.1 Data Quality

The first stage of selecting data which is believed to be well measured involves choosing "good" runs and luminosity blocks. These are data taking periods identified by the data quality team, which are defined according to whether some possible known issues are expected to have degraded the integrity of collected data. The following were taken into account in this early phase of data collection by ATLAS;

- Validity of physics run and luminosity block verified by data quality team.
- Trigger was functional.
- "Stable beams" declared.
- Solenoid and toroid magnets on and stable.
- All LAr, Tile and endcap calorimeters performed without major problems.

14.1.2 Collision Event Selection

In the next stage of filtering, "collision events" in which two beam protons undergo a head-on inelastic collision, are selected. Known backgrounds to collision events are beam gas collisions, beam halo events, and a small contribution from cosmic ray events. Beam gas collisions are collisions between a beam proton and residual nuclei within the beam pipe. Beam halo "events" are caused by pions and muons travelling in the beam halo, originating upstream in either direction from ATLAS, for example due to interaction with beam collimators.

Beam-backgrounds are removed using the assumption that collision events are expected to produce activity in both sides of the LAr endcap and forward calorimeters with similar timings, such that they are consistent with particles originating from the detector centre. Beam-backgrounds originate from outside ATLAS and will therefore produce signal in one side before the other, with significant timing difference. They are removed from the collision events by requiring the following;

- A trigger with signal in the beam pickup timing devices (located either side of ATLAS 175 m along the beam pipe) and at least one minimum bias trigger scintillator hit on both sides of ATLAS.

- At least one vertex per event, with at least 5 tracks of $p_T > 150$ MeV

- The mean arrival times of particles in each side of the end-cap calorimeter inner wheels and the forward calorimeters are considered. The mean side-arrival time is the average time of all particles arriving in that side. The difference in the mean arrival time of the two sides, δ_{time}, is considered. The distribution of δ_{time} has a peak at zero, and secondary peaks at $\pm 30(\pm 23)$ ns. The selection requires $\delta_{time} <$ 5 ns for the LAr ECAL, and $\delta_{time} < 10$ ns for the minimum bias trigger scintillator.

14.1.3 Jet Selection

The Anti-k_T algorithm was then run over the surviving events with a resolution parameter, ΔR, of 0.6 (0.4 in the note [48]) and jets of p_T greater than 20 GeV (10 GeV in the note), measured at the EM scale are accepted in this study.

After the selection of collision events this minimum bias sample still mainly consists of poorly measured and fake jets. Poorly measured jets are largely due to out-of-time energy depositions, and jets from real energy depositions, but where the energy measurement is of low confidence due to not well calibrated sections of the calorimeter. Fake jets are due to poor LHC beam conditions, hardware problems, and coincidental cosmic ray showers.

Jet cleaning cuts are used to address only jets from fake or out-of-time energy depositions. The following three criteria define fake or mis-reconstructed jets and each are accounted for with a dedicated selection on one or two related cleaning observables;

- Sporadic noise bursts in the Hadronic calorimeter EndCap (HEC). Removed when *both* the fraction of the jet energy in the HEC, labeled f_{HEC}, is greater than 80% and a small number of *cells*, labeled n90, in the HEC whose sum contains at least 90%, of the jet energy, i.e. $f_{HEC} > 0.8$ and n90 ≤ 5.

- Noise bursts in the Electromagnetic Calorimeter (ECal). Removed by excluding jets with *both* the fraction of jet energy in the ECal, labeled f_{EM}, greater than 95% and the fraction of jet energy is measured in bad quality calorimeter cells, labeled $f_{quality}$, greater than 80 %, i.e. $f_{EM} > 0.95$ and $f_{quality} > 0.8$.

- "Jets" reconstructed with large out-of-time energy deposits with respect to collision time, due to e.g. cosmic rays. Removed by demanding that the energy-squared-

weighted cell time of a jet, labeled t_{jet}, is within 50 ns of the collision, i.e. $|t_{jet}| <$ 50 ns.

Figures from the ATLAS note [48] are included alongside the jet observable distributions produced for validation in this study. Those from the note are measured with the 7 TeV 0.3 nb^{-1} (April 2010) data-set and compared with the ATLAS minimum bias (Pythia MC09 tune) Monte Carlo, and those from the present validation, are measured with similar data differing by being reprocessed in April 2010 and Monte Carlo, see Figures 14.1-14.6. To observe the effect of a cut on each cleaning observable the other two *independent* selection criteria are applied. The Monte Carlo jet samples are normalised to the data.

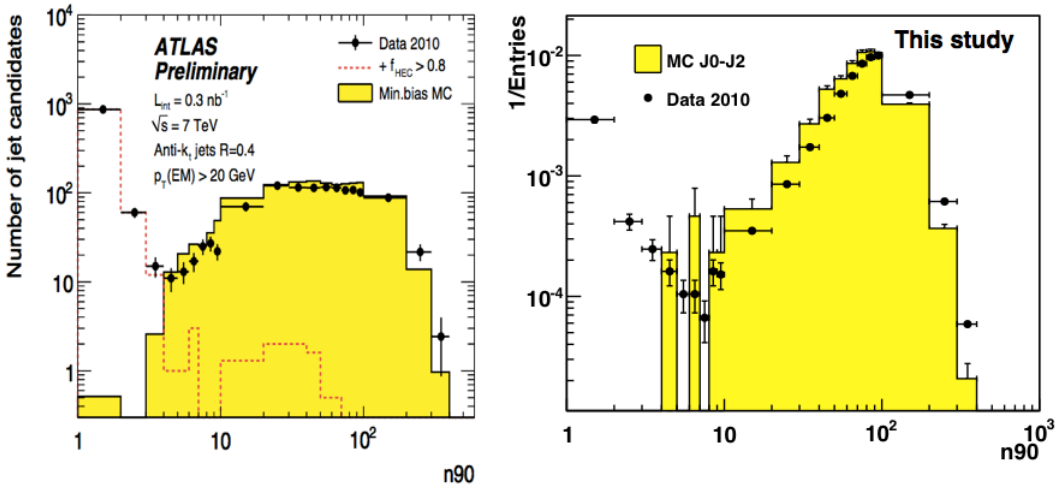


Figure 14.1: The n90 observable in data and Monte Carlo for the inclusive jet distribution after application of jet timing and ECal noise cuts. The Monte Carlo distribution is shown before (dashed line) and after sporadic noise bursts in the HEC are removed by cutting on the f_{HEC} variable. Results from ATLAS note (left) and from this study (right).

For jets with $p_T > 20$ GeV, the percentage of the defined mis-reconstructed jets is 5%. The majority (99%) of mis-reconstructed jets are from sporadic out-of-time HEC noise bursts, which leave energy depositions in only a few cells. These jets produce the peaks in the data at n90=1 and f_{HEC} >0.8, which are not present in the Monte Carlo, see the dashed lines in the n90 and f_{HEC} distributions which indicate jets with $f_{HEC} > 0.8$ and n90 ≤ 5 respectively. Fake jets from noise bursts in the ECAL are rare, at the sub-percent level, illustrated by the good agreement of Monte Carlo and data in the f_{EM} observable (Figure 14.4).

The numbers of events remaining in the sample after applying the selections and cuts are shown in Table 14.1. The transverse momentum spectra for surviving jets after

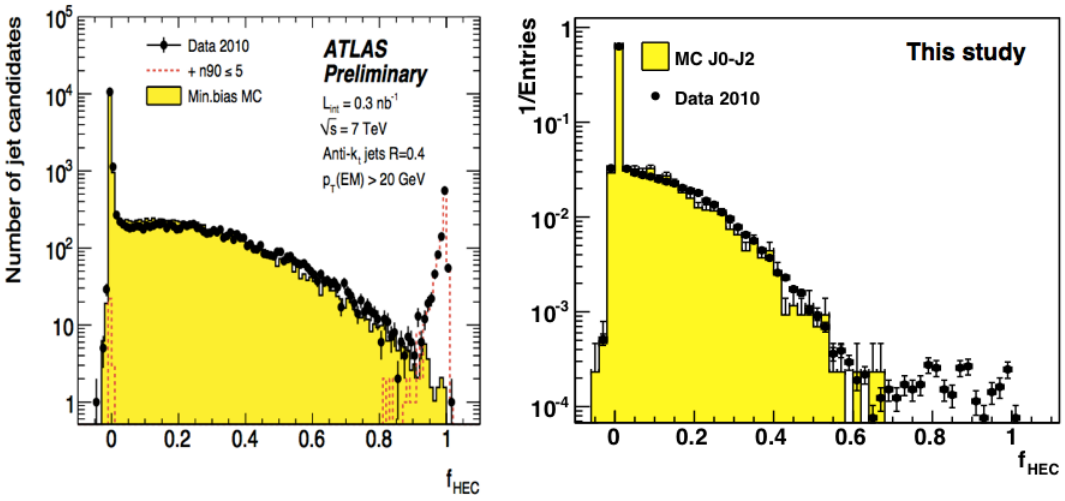


Figure 14.2: The f_{HEC} observable in data and Monte Carlo for the inclusive jet distribution after application of jet timing and ECal noise cuts. The Monte Carlo distribution is shown before (dashed line) and after single cell jets are removed by cutting on the n90 variable. Results from ATLAS note (left) and from this study (right).

successive application of the cleaning cuts are shown in Figure 14.6. Almost all jets in the tail of the distribution are rejected.

	Number of Events	After Collision Cut	After Jet Cleaning
Data - Minimum Bias	30×10^6	6.8×10^6	106×10^3
ATLAS (J0-J2) Monte Carlo	250×10^3	210×10^3	5.6×10^3

Table 14.1: The number of events in data and Monte Carlo in the present study after successive stages of the jet sample selection.

The J0-J2 MC analysed in this study was generated in the parton CMS transverse momentum range $8 < \hat{P}_{T\,parton} < 70$ GeV, rather than the complete minimum bias value of >2.5 GeV. This is believed to account for some differences between the ATLAS plots and those reproduced in this study. Also the dataset in the present study was only a small proportion of that used in the ATLAS note. Further differences are expected since the data analysed in the present study had been reprocessed in April, those in the ATLAS note had not. The significant differences in Figure 14.1 are not expected, and further work would be needed to establish the cause.

Note that jets which would be measured in the proposed fragmentation analysis would be much cleaner, since they are of much higher transverse momenta than those studied in this section. Note also that this study used minimum bias data, meaning that "collision events" must be selected among the triggered events, this is not required when using a

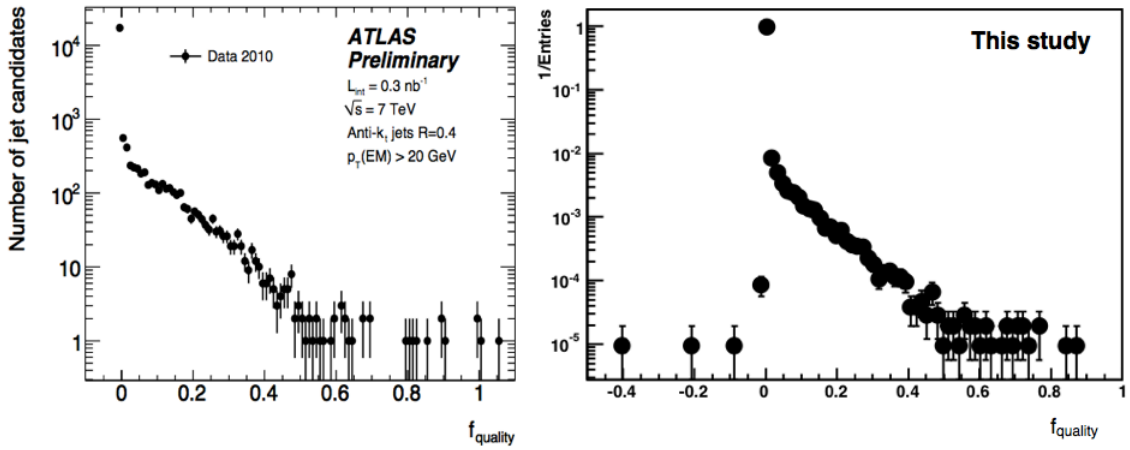


Figure 14.3: The $f_{quality}$ observable in data and Monte Carlo for the inclusive jet distribution after application of single cell and jet timing noise cuts. Results from ATLAS note (left) and from this study (right).

jet trigger, where events are almost certainly from collisions.

14.2 Fragmentation Data Analysis Model

There are many inverse femtobarns of recorded jet data available. This section details how data would be selected for a fragmentation analysis. In addition to the track selections described in Chapter 11, this will involve selecting efficiently triggered events containing central jets, and forming equivalent Monte Carlo samples to unfold the data, and for comparison purposes.

14.2.1 Trigger

The selection of data events should be based on an observable which is well-measured and efficiently related to the trigger variable, such that the data could be included with known high efficiency and thus less bias. For this reason, and to lessen Monte Carlo wastage, events are selected in defined p_T ranges according to the (offline) reconstructed transverse momentum of the highest p_T jet in each event, using triggers of the form "at least one Anti-k_T jet with greater than some p_T". Level 1 calorimeter (missing transverse energy) trigger efficiencies have been studied in data as part of an ATLAS service task, however since not directly relevant they are not included in this thesis. Effects of trigger turn on curves, such as bias, would be avoided by only selecting events collected with a given trigger, where that trigger had reached full efficiency.

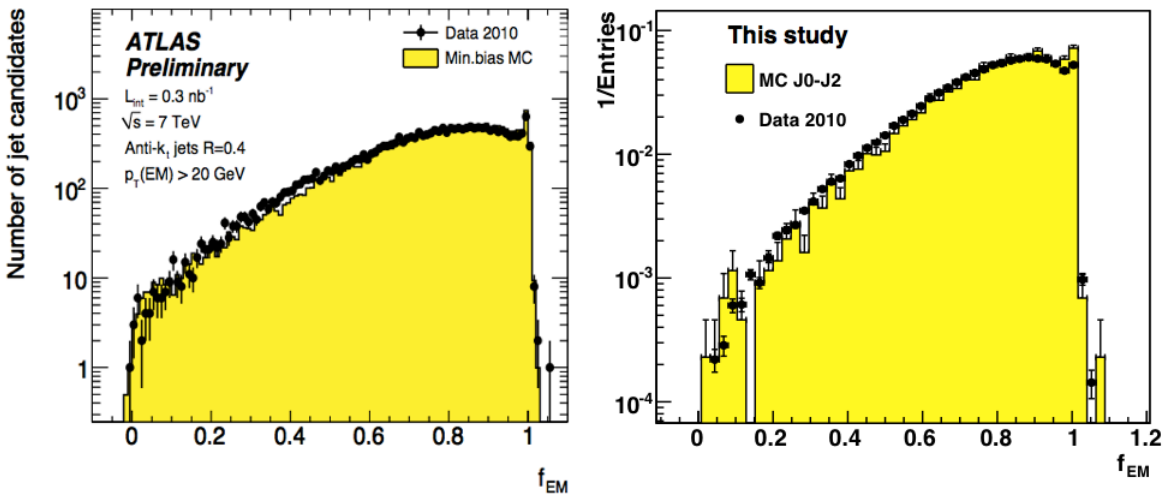


Figure 14.4: The f_{EM} observable in data for the inclusive jet distribution after application of single cell and jet timing noise cuts. Results from ATLAS note (left) and from this study (right).

14.2.2 Event Selection

Exclusive data samples will be selected each from only one trigger. If an event passes more than one trigger, it will only be included in the sample corresponding to the highest p_T (optimally efficient) trigger it passed. Removing such events spoils inclusivity.

To measure with large angle sampling cones without acceptance losses (due to detector geometry), all selected events will be further required to have the two highest p_T jets both with $|\eta_{jet}|$ <1.0.

14.2.3 Monte Carlo Concatenation

The Monte Carlo samples used thus far are defined according to the transverse momentum of the outgoing partons in the centre of momentum system, $\hat{p}_{T,parton}$, at the truth level. As stated, data samples will be defined in terms of the (offline) reconstructed jet transverse momentum, $p_{T,jet}$, and so the available Monte Carlo samples will be concatenated to have similar spectra in the same variable so that unfolding tools trained on them are relevant. Events will be combined from more than one sample, see Diagram 14.7, appropriately weighted by σ/N, where σ is the (Monte Carlo supplied) cross section corresponding to the N generated events being analysed. To avoid a possible bias to the concatenated Monte Carlo samples due to the difference in $p_{T,jet}$ and $\hat{p}_{T,parton}$ values, jets with p_{jet} values in the J3 p_T range should be included from the J2 sample.

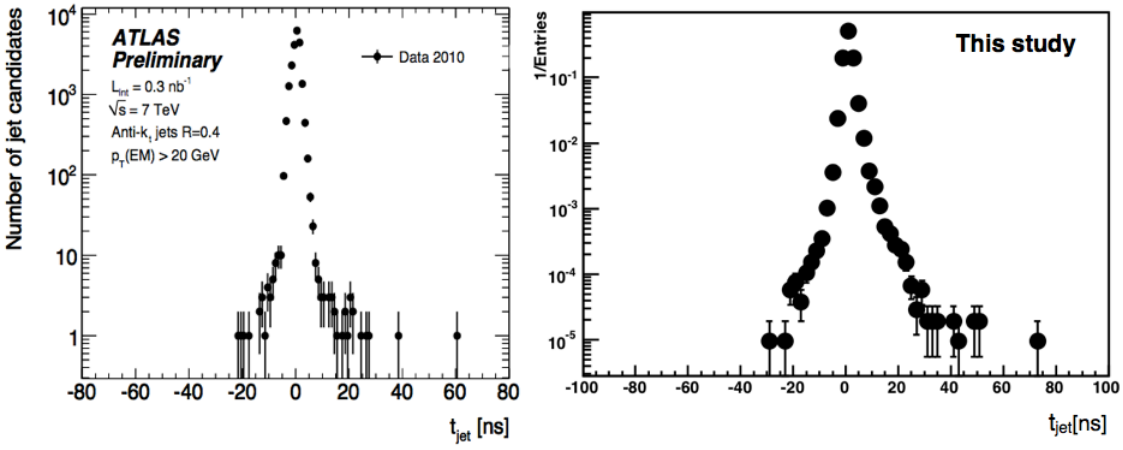


Figure 14.5: The jet time in data for the inclusive jet distribution after application of single cell and ECal noise cuts. Results from ATLAS note (left) and from this study (right).

14.2.4 Comparing Monte Carlo and Data

Final results will be plotted as a function of the hard scale observable ($\sqrt{-t}$) calculated in the appropriate Monte Carlo sample in the interests of both the fundamental nature of this variable and for comparison with existing *DIS* data. The cross section per event will be measured in data and Monte Carlo and used to weight events (data events weighted by S/L, where S and L are the prescale and luminosity, respectively), such that raw spectra (e.g. p_{Tjet}) in addition to normalised spectra may be compared.

It is noted at this stage that, while the Monte Carlo used to unfold a data sample should ideally have an identical p_T selection, small differences will not significantly affect the accuracy of the results, since the evolution of the function being measured varies only with the logarithm of the hard scale.

Since the function is normalised by the number of partons measured, no cross section is measured, and there is only slow variation of the fragmentation function with the logarithm of the hard scale, absolute inclusivity, i.e. that every event ends up in one of the concatenated samples, is not crucial, and may be relaxed within existing errors.

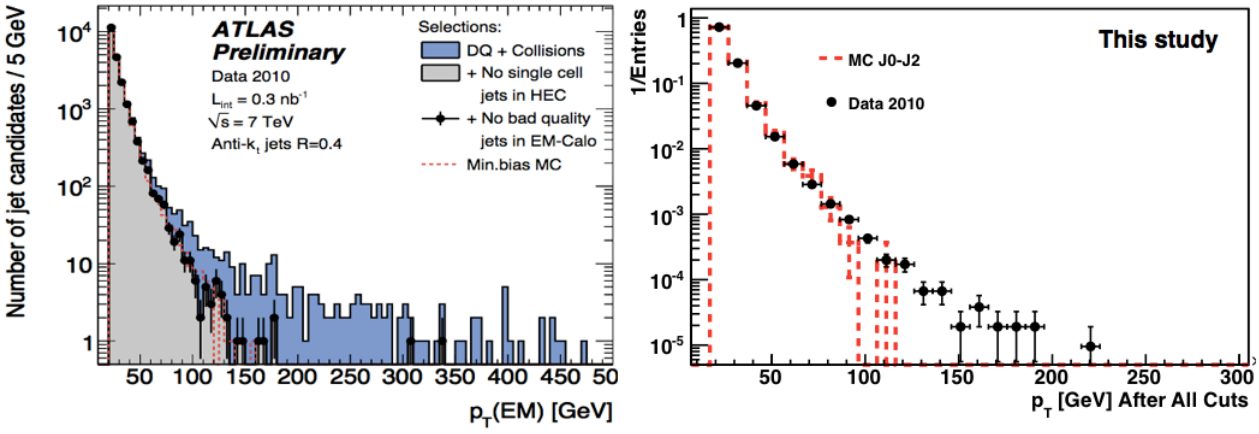


Figure 14.6: p_T of the inclusive jet distribution after all jet cleaning cuts. The difference above 100 GeV is due to limited available Monte Carlo events. Results from ATLAS note (left) and from this study (right).

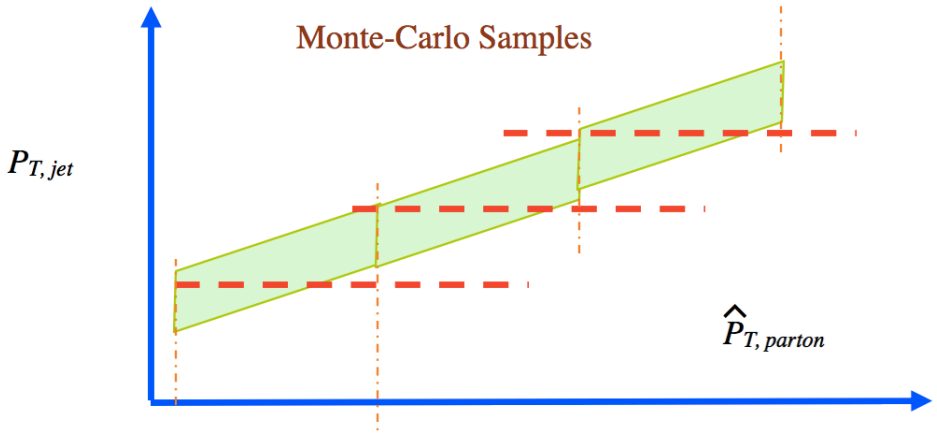


Figure 14.7: Diagram showing how MC samples will be concatenated based on transverse momentum values of jets, $p_{T,jet}$, reconstructed with the ($R = 0.6$) Anti-k_T algorithm from available J* samples, which are defined in terms of $\hat{p}_{T,parton}$.

Chapter 15

Conclusions

A review of the Standard Model of particle physics has been presented with emphasis on light quark QCD, and existing data on light quark fragmentation from e^+e^- annihilation and deep inelastic scattering experiments.

Methods of modelling hadronic collisions, including perturbative parton showering techniques and non-perturbative models for hadronisation have been discussed and a brief description has been given of particle measurement processes using the ATLAS detector as an example.

The feasibility of measuring the partonic fragmentation function in large hadron collider conditions has been tested and a novel method (FAPS) to do so has been developed.

The Anti-k_T and TRAPS algorithms have been used to provide initial partonic momentum estimates, which in turn scale the hadronic momenta. Minimal dependence on the specific exterior jet algorithm chosen has been demonstrated, and it has been shown that parton estimator resolution may be unfolded using bin-by-bin and matrix methods. Tracking and calorimetry measurement resolution has also been unfolded simultaneously within the same processes. The two jet algorithms have been compared for their partonic momentum resolutions: the Anti-k_T algorithm has better resolution at low momentum, however they are competitive at high momentum at the measured level.

Examination of hadronic profiles about the parton at Monte Carlo truth level has shown that there is an approximately uniform uncorrelated background from ISB and underlying event. There is also an additional uncorrelated background from pile-up. Background has been estimated on an event-by-event basis, in regions away from jets. At low x_p and at the lowest hard scales tested, background is almost twice as large as

signal. However after subtraction there is a satisfactory and expected plateau behaviour and good agreement with previous fragmentation data in the hard scale overlap region (Figure 15.1).

A variable radius cone sampling method with background subtraction has been used to count correlated charged pions with Monte Carlo reliance for correction vectors. The varying radius technique provides a way of coping with any bias or poor partonic directional resolution of jet algorithms. Extrapolation techniques have been used to estimate total numbers of correlated pions, where measurement with a large enough sampling cone is not viable, for example, due to proximity to the fiducial region or to other jets.

A novel method to extract information on transverse fragmentation based on a jet collimation parameter has been demonstrated, which is in principle sensitive to scaling or scaling violations.

The fragmentation algorithm has been tested over an order of magnitude in hard scale ($\sim$100 GeV $\rightarrow$ 1 TeV) with two standard ATLAS reconstructed level Monte Carlos and using standard data quality selection cuts. Notably these Monte Carlos have different hadronisation models. The systematic error associated with unfolding data has been estimated by the difference observed when using one of the models to unfold the "data" from the other.

The algorithm is limited at the highest momentum by significant tracking uncertainty and as a result, efficiencies and purities are as low as 5%. Wider binning may be used here to compensate, while much narrower binning than DIS/e^+e^- may be used at low momentum, where statistics are enormous, acceptance is high and tracks are well-measured. There are instabilities associated with the matrix unfolding method. The use of a much larger Monte Carlo sample might reduce errors due to matrices which are sparsely populated at high x_p.

Results of the algorithm at the reconstructed level have been compared to ATLAS *jet* fragmentation data. Compatibility is observed, however the limitations of using a fixed cone are highlighted through the observable differences.

Selections have been applied to ATLAS data to gather events and objects suitably resolved such that unfolding may be done. A description has been given of how events would be selected from available Monte Carlo and data for an unfolded fragmentation data measurement. Such measurement could support the concept of quark universality

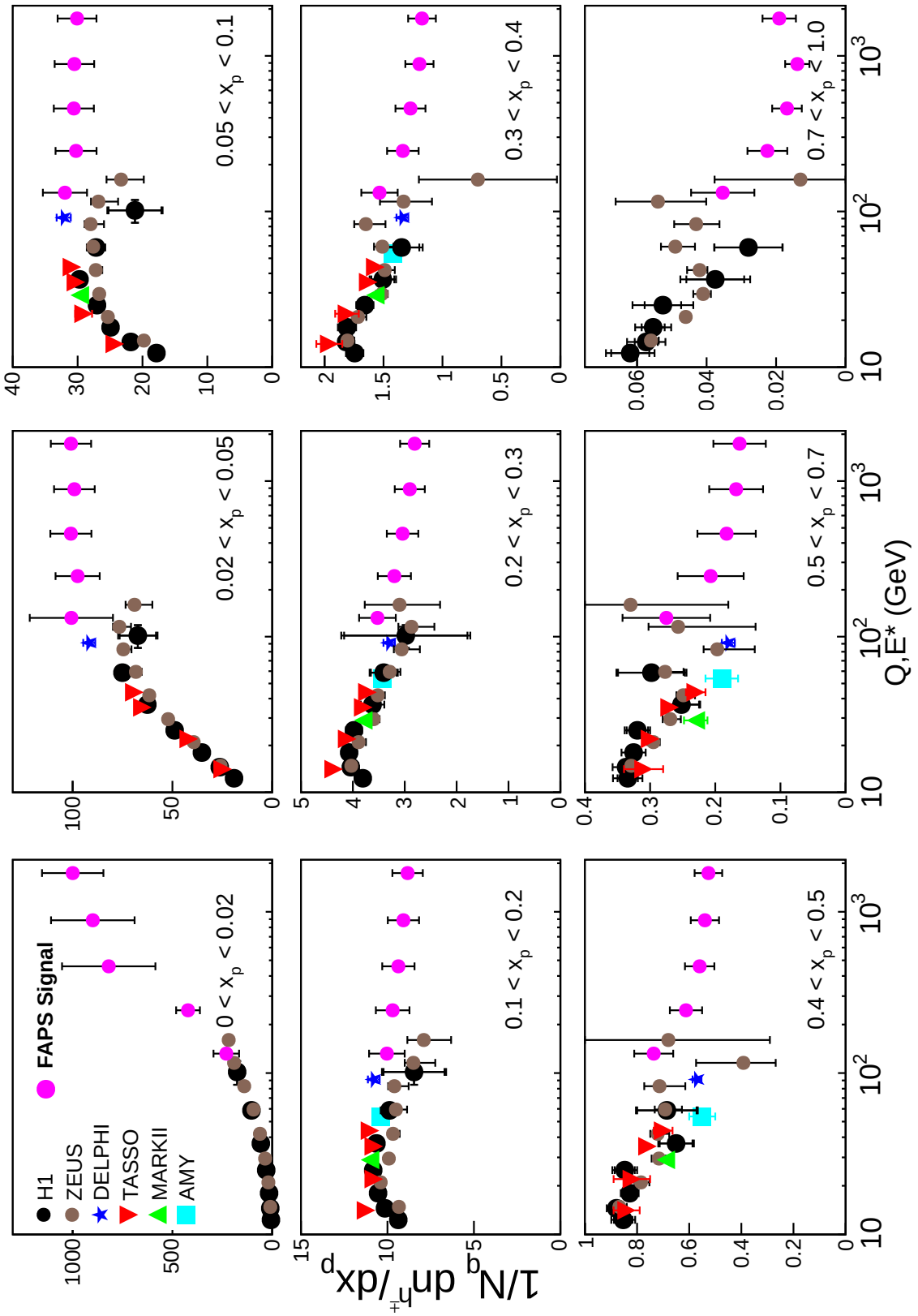


Figure 15.1: The quark to charged hadron fragmentation function calculated with the FAPS method from LHC Monte Carlo presented alongside previous quark to charged hadron fragmentation data.

by establishing propagator invariance, and enable comment on the disagreement of e^+e^- and DIS quark fragmentation data.

For comparison, the correction to pions was relaxed and the quark to charged hadron fragmentation function calculated with the FAPS method, using Pythia reconstructed Monte Carlo as "data" and HERWIG Monte Carlo for the bin-by-bin correction, is shown in Table 15.1 and alongside previous quark to charged hadron fragmentation data (described in Section 6.4) in Figure 15.1.

The pion fragmentation function can be used more suitably (for the reasons given in Section 11.3) than the charged hadron fragmentation function, to observe whether or not scaling has been violated. The quark to charged pion fragmentation function calculated with the FAPS method, using Pythia reconstructed Monte Carlo as "data" and HERWIG Monte Carlo for the bin-by-bin correction is shown in Table 15.2.

There is agreement of the fragmentation function as calculated by this algorithm with previous data, thus propagator invariance is at least modelled in the Monte Carlo, as are scaling violations, or a softening of the fragmentation function with increasing hard scale, but only, presently, within a systematic error.

Realistic statistical uncertainties given the presently available 20 fb^{-1} of ATLAS data are included in Tables 15.1 and 15.2, they are not significant apart from at the very highest momenta, i.e. $x_p \geq 0.2$ for the J7 sample. A systematic Monte Carlo model uncertainty of 10% is included for the first seven x_p bins, and one of 25% is included in the two highest bins. This is the most significant error in all of the x_p intervals excluding the lowest, where there is much background, extrapolation is most necessary, and corresponding errors are considerable, and at the highest momenta where statistical uncertainty is large. The background error could be reduced by suitable Monte Carlo retune, or re-weighting to reproduce data, and with greater confidence the extrapolation error could be reduced by around half from the present estimate of over the entire extrapolation range.

15.1 Outlook

Background is not an insuperable problem preventing parton fragmentation measurement in hadron collisions. The LHC is presently operating at a centre-of-mass energy of 8 TeV and the parton fragmentation algorithm would be tested at this energy in preparation

for data analysis. The cross section for events at the upper limit of the hard scale range considered is larger at 8 TeV. The cross section for events at the lower hard scale limit is also larger, however an increased prescale further reduces the proportion of such events recorded. More background would be produced in a given hard scale range, however this is not thought to be a point of concern due to the ability to cope with large proportionate background demonstrated in this thesis.

Due to the relative stability, simplicity and guarantee of positivity, the bin-by-bin method of unfolding would be the preferred choice over the TUnfold package for a measurement in data. Such a measurement could then be used to predict charged pionic or hadronic multiplicities from hard scattered partons in hadron collisions using well known partonic cross sections. The jet mass definition may be useful by enabling a theoretical prediction which could be compared to experimental results.

Given the similarity of the parton and the jet fragmentation function, comparison where possible over the entire hard scale range considered would be performed in future. This would enable comment on whether the jet function measured at ATLAS is in fact the same as that of the parton.

MC Sample	x_p	D_{signal}	$\frac{\delta_{stat.}}{D_{signal}}$	$\frac{\delta_{background}}{D_{signal}}$	$\frac{\delta_{extrap.}}{D_{signal}}$	$\frac{\delta_{model}}{D_{signal}}$	$\frac{\delta_{total}}{D_{signal}}$
J3	$0> x_p \geq 0.02$	198.3	1%	26%	16%	10%	32%
	$0.02> x_p \geq 0.05$	81.5	1%	2%	24%	10%	26%
	$0.05> x_p \geq 0.1$	30.5	1%	1%	5%	10%	11%
	$0.1> x_p \geq 0.2$	9.69	1%	1%	4%	10%	11%
	$0.2> x_p \geq 0.3$	3.46	1%	1%	2%	10%	10%
	$0.3> x_p \geq 0.4$	1.52	1%	1%	1%	10%	10%
	$0.4> x_p \geq 0.5$	0.73	1%	1%	1%	10%	10%
	$0.5> x_p \geq 0.7$	0.26	1%	2%	4%	25%	25%
	$0.7> x_p \geq 1.0$	0.04	1%	4%	0	25%	25%
J4	$0> x_p \geq 0.02$	390.8	1%	8%	8%	10%	15%
	$0.02> x_p \geq 0.05$	91.3	1%	1%	7%	10%	12%
	$0.05> x_p \geq 0.1$	29.0	1%	1%	4%	10%	11%
	$0.1> x_p \geq 0.2$	9.49	1%	1%	2%	10%	10%
	$0.2> x_p \geq 0.3$	3.16	1%	1%	1%	10%	10%
	$0.3> x_p \geq 0.4$	1.33	1%	1%	1%	10%	10%
	$0.4> x_p \geq 0.5$	0.61	1%	1%	1%	10%	10%
	$0.5> x_p \geq 0.7$	0.20	1%	1%	4%	25%	25%
	$0.7> x_p \geq 1.0$	0.023	1%	3%	0	25%	25%
J5	$0> x_p \geq 0.02$	637.1	1%	20%	29%	10%	37%
	$0.02> x_p \geq 0.05$	96.7	1%	1%	3%	10%	11%
	$0.05> x_p \geq 0.1$	29.8	1%	1%	3%	10%	10%
	$0.1> x_p \geq 0.2$	9.23	1%	1%	2%	10%	10%
	$0.2> x_p \geq 0.3$	3.01	1%	1%	1%	10%	10%
	$0.3> x_p \geq 0.4$	1.27	1%	1%	1%	10%	10%
	$0.4> x_p \geq 0.5$	0.59	1%	1%	1%	10%	10%
	$0.5> x_p \geq 0.7$	0.18	1%	1%	3%	25%	25%
	$0.7> x_p \geq 1.0$	0.017	1%	3%	0	25%	25%
J6	$0> x_p \geq 0.02$	718.5	1%	11%	25%	10%	29%
	$0.02> x_p \geq 0.05$	95.6	1%	1%	2%	10%	11%
	$0.05> x_p \geq 0.1$	29.9	1%	1%	2%	10%	10%
	$0.1> x_p \geq 0.2$	8.98	1%	1%	1%	10%	10%
	$0.2> x_p \geq 0.3$	2.89	1%	1%	1%	10%	10%
	$0.3> x_p \geq 0.4$	1.19	1%	1%	1%	10%	10%
	$0.4> x_p \geq 0.5$	0.54	2%	1%	1%	10%	10%
	$0.5> x_p \geq 0.7$	0.16	2%	1%	2%	25%	25%
	$0.7> x_p \geq 1.0$	0.014	5%	2%	0	25%	26%
J7	$0> x_p \geq 0.02$	891.3	1%	7%	12%	10%	17%
	$0.02> x_p \geq 0.05$	98.7	3%	1%	2%	10%	11%
	$0.05> x_p \geq 0.1$	29.7	4%	1%	1%	10%	11%
	$0.1> x_p \geq 0.2$	8.77	5%	1%	1%	10%	11%
	$0.2> x_p \geq 0.3$	2.80	9%	1%	1%	10%	14%
	$0.3> x_p \geq 0.4$	1.14	14%	1%	3%	10%	18%
	$0.4> x_p \geq 0.5$	0.52	20%	1%	2%	10%	22%
	$0.5> x_p \geq 0.7$	0.16	22%	1%	1%	25%	33%
	$0.7> x_p \geq 1.0$	0.013	47%	2%	0	25%	53%

Table 15.1: The quark to charged hadron fragmentation function and percentage errors calculated with the FAPS method.

MC Sample	x_p	D_{signal}	$\frac{\delta_{stat.}}{D_{signal}}$	$\frac{\delta_{background}}{D_{signal}}$	$\frac{\delta_{extrap.}}{D_{signal}}$	$\frac{\delta_{model}}{D_{signal}}$	$\frac{\delta_{total}}{D_{signal}}$
J3	$0> x_p \geq 0.02$	177.2	1%	45%	15%	10%	49%
	$0.02> x_p \geq 0.05$	67.6	1%	3%	20%	10%	23%
	$0.05> x_p \geq 0.1$	23.7	1%	1%	4%	10%	11%
	$0.1> x_p \geq 0.2$	7.00	1%	1%	3%	10%	11%
	$0.2> x_p \geq 0.3$	2.28	1%	1%	2%	10%	10%
	$0.3> x_p \geq 0.4$	0.91	1%	1%	1%	10%	10%
	$0.4> x_p \geq 0.5$	0.42	1%	1%	1%	10%	10%
	$0.5> x_p \geq 0.7$	0.15	1%	2%	4%	25%	25%
	$0.7> x_p \geq 1.0$	0.022	1%	4%	2%	25%	25%
J4	$0> x_p \geq 0.02$	317.16	1%	10%	0	10%	14%
	$0.02> x_p \geq 0.05$	74.4	1%	1%	6%	10%	12%
	$0.05> x_p \geq 0.1$	22.6	1%	1%	4%	10%	11%
	$0.1> x_p \geq 0.2$	6.94	1%	1%	2%	10%	10%
	$0.2> x_p \geq 0.3$	2.14	1%	1%	1%	10%	10%
	$0.3> x_p \geq 0.4$	0.82	1%	1%	1%	10%	10%
	$0.4> x_p \geq 0.5$	0.36	1%	1%	1%	10%	10%
	$0.5> x_p \geq 0.7$	0.12	1%	2%	4%	25%	25%
	$0.7> x_p \geq 1.0$	0.015	1%	4%	2%	25%	26%
J5	$0> x_p \geq 0.02$	498.9	1%	12%	11%	10%	19%
	$0.02> x_p \geq 0.05$	79.0	1%	1%	3%	10%	10%
	$0.05> x_p \geq 0.1$	23.2	1%	1%	2%	10%	10%
	$0.1> x_p \geq 0.2$	6.81	1%	1%	1%	10%	10%
	$0.2> x_p \geq 0.3$	2.08	1%	1%	1%	10%	10%
	$0.3> x_p \geq 0.4$	0.81	1%	1%	1%	10%	10%
	$0.4> x_p \geq 0.5$	0.34	1%	1%	1%	10%	10%
	$0.5> x_p \geq 0.7$	0.11	1%	2%	1%	25%	25%
	$0.7> x_p \geq 1.0$	0.013	1%	4%	1%	25%	25%
J6	$0> x_p \geq 0.02$	647.0	1%	10%	14%	10%	20%
	$0.02> x_p \geq 0.05$	77.6	1%	1%	3%	10%	11%
	$0.05> x_p \geq 0.1$	23.4	1%	1%	2%	10%	10%
	$0.1> x_p \geq 0.2$	6.70	1%	1%	1%	10%	10%
	$0.2> x_p \geq 0.3$	2.02	1%	1%	1%	10%	10%
	$0.3> x_p \geq 0.4$	0.78	1%	1%	1%	10%	10%
	$0.4> x_p \geq 0.5$	0.34	2%	1%	1%	10%	10%
	$0.5> x_p \geq 0.7$	0.11	2%	1%	2%	25%	25%
	$0.7> x_p \geq 1.0$	0.011	5%	2%	1%	25%	26%
J7	$0> x_p \geq 0.02$	742.8	2%	4%	9%	10%	14%
	$0.02> x_p \geq 0.05$	79.7	3%	1%	2%	10%	11%
	$0.05> x_p \geq 0.1$	23.2	4%	1%	1%	10%	11%
	$0.1> x_p \geq 0.2$	6.58	5%	1%	1%	10%	11%
	$0.2> x_p \geq 0.3$	1.98	10%	1%	1%	10%	14%
	$0.3> x_p \geq 0.4$	0.76	15%	1%	3%	10%	18%
	$0.4> x_p \geq 0.5$	0.33	21%	1%	2%	10%	24%
	$0.5> x_p \geq 0.7$	0.11	24%	1%	2%	25%	35%
	$0.7> x_p \geq 1.0$	0.011	50%	2%	0	25%	56%

Table 15.2: The quark to charged pion fragmentation function and percentage errors calculated with the FAPS method.

Bibliography

[1] F.D. Aaron et al. Charged Particle Production in High Q^2 Deep-Inelastic Scattering at HERA. *Phys.Lett.*, B654:148–159, 2007.

[2] A. Petersen et al. Multihadronic Events at $E_{c.m.}$=29 GeV and Predictions of QCD Models from $E_{c.m.}$=29 GeV to $E_{c.m.}$=93 GeV. *Phys.Rev.*, D37, 1998.

[3] Y.K. Li et al. Multi-hadron Event Properties in e^+e^- Annihilation at $\sqrt{s} = 52$ GeV to 57 GeV. *Phys.Rev.*, D41:2675, 1990.

[4] W. Braunschweig et al. Global Jet Properties at 14 GeV TO 44 GeV Centre-of-mass Energy in e^+e^- Annihilation. *Z.Phys.*, C47:187–198, 1990.

[5] P. Abreu et al. Determination of α_s from the Scaling Violation in the Fragmentation Functions in e^+e^- Annihilation. *Phys.Lett.*, B311:408–424, 1993.

[6] H. Abramowicz et al. Scaled Momentum Spectra in Deep Inelastic Scattering at HERA. *JHEP*, 1006:009, 2010.

[7] B. Webber. Jet Fragmentation. QCDNET99, 1999.

[8] L. Evans and P. Bryant. LHC Machine. *JINST*, 3:S08001, 2008.

[9] B. Martin and G. Shaw. *Particle Physics.* Wiley, 2008.

[10] D. Buskulic et al. Measurement of α_s from Scaling Violations in Fragmentation Functions in e^+e^- Annihilation. *Phys.Lett.*, B357:487–499, 1995.

[11] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[12] W. N. Cottingham and D. A. Greenwood. *An Introduction to the Standard Model of Particle Physics.* Cambridge, 2007.

[13] H. Fritzsch. *Elementary Particles: Building Blocks of Matter*. World Scientific, 2005.

[14] D. Griffiths. *Introduction to Elementary Particles*. Wiley, 2008.

[15] F. Halzen and A. Martin. *Quarks and Leptons*. Wiley, 1984.

[16] G. Aad et al. Measurement of the Jet Fragmentation Function and Transverse Profile in Proton-Proton Collisions at a Center-of-Mass Energy of 7 TeV with the ATLAS Detector. *Eur.Phys.J.*, C71:1795, 2011.

[17] T. Sjostrand, S. Mrenna, and P. Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.

[18] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, et al. HERWIG 6: An Event Generator for Hadron Emission Reactions with Interfering Gluons (including Supersymmetric Processes). *JHEP*, 0101:010, 2001.

[19] SHERPA Collaboration. SHERPA User Manual. *Online, [Cited 2013]*, http://sherpa.hepforge.org/doc/SHERPA-MC-1.4.3.html.

[20] R. Gupta. Introduction to Lattice QCD. *arXiv*, hep-lat/9807028, 1997.

[21] The LEP Collaboration. LEP Design Report. 1984.

[22] V.V. Sudakov. Vertex Parts at Very High-Energies in Quantum Electrodynamics. *Sov.Phys.JETP*, 3:65–71, 1956.

[23] R. K. Ellis, W. J. Stirling, and B. R. Webber. *QCD and Collider Physics*. Cambridge University Press, 2003.

[24] Serguei Chatrchyan et al. Observation of Long-Range Near-Side Angular Correlations in Proton-Lead Collisions at the LHC. *Phys.Lett.*, B718:795–814, 2013.

[25] B. R. Webber. Fragmentation and Hadronization. *Online, [Cited July 2013]*, http://www.slac.stanford.edu/econf/C990809/docs/webber.pdf.

[26] N. N. Nikolaev and B. G. Zakharov. Colour Transparency and Scaling Properties of Nuclear Shadowing in Deep Inelastic Scattering. *Z.Phys.*, C49, 1991.

[27] ATLAS Collaboration. The ATLAS Athena Framework. *Online*, https://twiki.cern.ch/twiki/bin/viewauth/Atlas/AthenaFramework.

[28] S. Agostinelli et al. GEANT4: A Simulation Toolkit. *Nucl.Instrum.Meth.*, A506:250–303, 2003.

[29] C Albajar et al. Analysis of the Highest Transverse Energy Events Seen in the UA1 Detector at the Sp$\overline{p}$S Collider. *Z. Phys. C*, RAL-87-037, Jun 1987.

[30] D. Kant. QCD Studies in the Breit Frame at HERA. *Nucl.Phys.Proc.Suppl.*, 71:31–38, 1999.

[31] K. Ellis. *TRAPS: Topological Reconstruction Algorithm for Parton Scatters.* PhD thesis, Queen Mary University of London, 2012.

[32] S. Moch. Expectations at LHC from hard QCD. *J.Phys.*, G35:073001, 2008.

[33] T. Muta. *Foundations of Quantum Chromodynamics: An Introduction to Perturbative Methods in Gauge Theories.* World Scientific, 2009.

[34] S. Albino, F. Anulli, F. Arleo, D. Z. Besson, W. K. Brooks, et al. Parton Fragmentation in the Vacuum and in the Medium. 2008.

[35] P. S. L. Booth. The DELPHI Experiment. *Philisophical Transactions: Physical Sciences and Engineering*, 336 No. 1642:213–222, 1991.

[36] G. S. Abrams et al. The MARK-II Detector for the SLC. *Nucl.Instrum.Meth.*, A281:55, 1989.

[37] Virginia Polytechnic Inst. and Blacksburg. Dept. of Physics. State Univ. The AMY Experiment: Progress Report. 1989.

[38] I. Abt et al. The H1 detector at HERA. *Nucl. Instrum. Meth.*, A386:310–347, 1997.

[39] ZEUS Collaboration. The ZEUS Detector. Status Report (unpublished). 1993.

[40] M. Cacciari, G. P. Salam, and G. Soyez. The Anti-k_T Jet Clustering Algorithm. *Journal of High Energy Physics*, 04-63, 2008.

[41] CERN. The ROOT User's Guide. *Online, [Cited 2013]*, http://root.cern.ch/drupal/content/users-guide.

[42] S. Schmidt. Tunfold. *Online, [Cited November 2011]*, http://root.cern.ch/root/html/TUnfold.html.

[43] The ATLAS Collaboration. Calorimeter Performance - Technical Design Report. CERN/LHCC-96-40, 1997.

[44] G. Aad et al. Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics. SLAC-R-980, 2009.

[45] R. J. Dankers. *The Physics Performance of the Level 2 Trigger for the Inner Detector of ATLAS*. PhD thesis, NIKHEF, 1998.

[46] The ATLAS Collaboration. Jet Energy Scale and its Systematic Uncertainty in Proton-Proton Collisions at $\sqrt{s}$ =7 TeV with ATLAS 2011 Data. ATLAS-CONF-2013-004, 2013.

[47] The ATLAS Collaboration. ATLAS Tunes of PYTHIA 6 and Pythia 8 for MC11. ATL-PHYS-PUB-2011-009, Jul 2011.

[48] Data-Quality Requirements and Event Cleaning for Jets and Missing Transverse Energy Reconstruction with the ATLAS Detector in Proton-Proton Collisions at a Center-of-Mass Energy of $\sqrt{s} = 7$ TeV. ATLAS-CONF-2010-038, 2010.

[49] P. Dixon, D. Kant, and G. Thompson. Fragmentation Function Scaling violations in the Breit Frame. *J.Phys.*, G25:1453–1456, 1999.

[50] B Andersson, G Gustafson, G Ingelman, and T. Sjostrand. Parton Fragmentation and String Dynamics. *Phys. Rept.*, 97:31–145, 1983.

[51] P. Dixon. *A Study of the Fragmentation of Quarks in* ep *Collisions at HERA*. PhD thesis, Queen Mary University of London, 1998.

[52] B. Brzozowska et al. Scaled Momentum Spectra in the Current Region of the Breit Frame at HERA. 2007.

[53] G. G. Hanson. Jet Production and Quark Fragmentation Results from PEP. SLAC-PUB-3194, 1983.

[54] N. Wainer. A Study of Fragmentation in e^+e^- Annihilation and of Electron Identification in Calorimeters. 1989.

[55] C. Z. Jarlskog. *Luminosity Evaluation and Fragmentation Studies for the DELPHI Experiment at LEP*. PhD thesis, Lund U., 2001.

[56] J. Turney. *A Study of Identified Hadron Fragmentation in* ep *Collisions at HERA Using the H1 Detector*. PhD thesis, Queen Mary University of London, 2002.

[57] D Traynor. *Hadronic Fragmentation Studies in Diffractive Deep Inelastic Scattering at HERA*. PhD thesis, Queen Mary University of London, 2002.

[58] S. Eno. Recent Results from the AMY Experiment. 1989.

[59] The ATLAS Collaboration. Trigger Performance - Technical Design Report. CERN/LHCC-98-15, 1998.

[60] RIVET Collaboration. Rivet User Manual. *Online, [Cited 2010]*, http://projects.hepforge.org/rivet/rivet-manual.pdf.

[61] M. Wing. Standard Model (lecture notes). University of London, 2009.

[62] The ATLAS Collaboration. ATLAS Detectors. *Online, [Cited July 2010]*, https://twiki.cern.ch/twiki/bin/view/Atlas/WebHome.

[63] S. Aid et al. A Study of the Fragmentation of Quarks in *ep* collisions at HERA. *Nucl.Phys.*, B445:3–24, 1995.

[64] J. Bromley. Fragmentation Functions at ZEUS. 1997.

[65] The ATLAS Collaboration. ATLAS: Detector and Physics Performance Technical Design Report. Volume 1. CERN-LHCC-99-14, 1994.

[66] H. Ogren and P. Cwetanski. TRT Wiki. *Online, [Cited July 2010]*, https://twiki.cern.ch/twiki/bin/view/Atlas/TrtWiki.

[67] A. J. Barr. *Studies of Supersymmetric Models for the ATLAS Experiment at the LHC*. PhD thesis, Cambridge, 2007.

[68] R. Ruber. ATLAS Magnet System. *Online, [Cited July 2010]*, http://atlas-bt.web.cern.ch/atlas-bt/info/project/.

[69] J. Pequenao. The ATLAS Detector. *Online, [Cited March 2008]*, http://cdsweb.cern.ch/record/1095924.

[70] F. E. Close. *An Introduction to Quarks and Partons*. Academic Press, 1979.

[71] A. Smilga. *Lectures on Quantum Chromodynamics*. World Scientific, 2001.