MEASUREMENT OF THE INCLUSIVE JET CROSS SECTION USING THE MIDPOINT ALGORITHM IN RUN II AT THE COLLIDER DETECTOR AT FERMILAB (CDF)

By

ROBERT CRAIG GROUP

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2006

I dedicate this work to all who devote their careers to the study of science and the selfless pursuit of knowledge.

ACKNOWLEDGMENTS

I would like to thank my wife, Nicole, for her understanding of my time away from home during this work. Her encouragement and confidence in my ability carried over into my professional demeanor; my success is directly correlated with her support.

I would like to thank my advisors for their guidance over the last four years. Professor Field's tireless dedication to answer the toughest questions of QCD kept me working hard to keep up. Professor Matchev's requirement to give many public presentations, and travel support from his Outstanding Junior Investigator award, helped me to conquer my fear of public speaking. Also, partial support for my stipend this semester from Dr. Matchev, made it possible for me to concentrate on this dissertation and obtaining employment. Without professor Field's and professor Matchev's combined efforts on my behalf, I do not know where I would have ended up after UF. I also want to thank Dr. Kenichi Hatakeyama, of the The Rockefeller University, for his patience while teaching me how to do an experimental analysis. Without his constant collaboration, I would have had no chance to complete a CDF analysis without being located at Fermilab. I am proud of the work we did together over these last two years. Finally, I want to thank Professor Field, Professor Matchev, and Dr. Hatakeyama for writing the recommendation letters that helped me obtain more employment options than I could possibly have expected.

Professor Paul Avery and Dr. Jorge Rodriguez allowed me to work with the UF Grid computing project my first summer at UF. This allowed my wife and me to move to Gainesville early and buy a house. I am thankful for this. I appreciate

the patience and time extended to me by Dr. Dimitri Bourilkov. The technical skills he helped me to develop will be invaluable to me in my future.

The High Energy Theory Group supported me for travel to the TeV4LHC workshop, and the TASI summer school. These experiences helped me build exposure and confidence in my place in the high energy physics community. The High Energy Experimental Group gave me partial support this last semester so that I could search for a job, and finish my degree. I appreciate this support and hope to take full advantage of the freedom it allows me. The HEE group also made it possible for me to spend three months at FNAL last Winter to fulfill my Ace shift responsibility. I took advantage of this time to make great progress on my research, and meet many of the people at CDF I had only worked with over email. I am truly thankful for this opportunity. I also thank Yvonne Dixon for her patience with travel paperwork. She always made it easy to travel. The four year full support from the Alumni Fellowship offered to me by the University of Florida allowed me to pursue additional research projects outside of the scope of this dissertation. The generous support from all of the above groups has made my experience at UF an exciting and successful one.

I also have had the benefit of many great friends over my graduate career whom I have been able to share my experiences with. At FSU, Jose Lazoflores, and Jorge O'Farrill helped me find my place in the department and have remained good friends over the years. At UF, Wayne Bomstad, Ethan Siegal, Leanne Duffy, Alberto Cruz, Bobby Scurlock and many others contributed indirectly to my success. The selfless support extended to me by other students in the UF advanced graduate program has been exceptional.

I want to thank my mother and father for always being there. Without their encouragement and support throughout my lifetime, I could never have dreamed

I would obtain a PhD in physics. They always said I was capable of anything I wanted to do, and never showed surprise when I was successful.

TABLE OF CONTENTS

CHAPTER

APPENDIX

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

MEASUREMENT OF THE INCLUSIVE JET CROSS SECTION USING THE MIDPOINT ALGORITHM IN RUN II AT THE COLLIDER DETECTOR AT FERMILAB (CDF)

By

Robert Craig Group

December 2006

Chair: Richard D. Field
Major Department: Physics

A measurement is presented of the inclusive jet cross section using the Midpoint jet clustering algorithm in five different rapidity regions. This is the first analysis which measures the inclusive jet cross section using the Midpoint algorithm in the forward region of the detector. The measurement is based on more than 1 fb^{-1} of integrated luminosity of Run II data taken by the CDF experiment at the Fermi National Accelerator Laboratory. The results are consistent with with the predictions of perturbative quantum chromodynamics.

CHAPTER 1
INTRODUCTION

Since the dawn of intellectual thought, human beings have questioned the origin of the matter that composes the world in which they live. The earliest well documented attempts to explain the physical world were constructed by philosophers. Democritus (ca 460-370 BC) postulated that invisible and indivisible “atoms” made up everything around us. In our modern world, we have the technology to go beyond mere speculation and design experiments to directly probe the nature of matter. Thanks to experimental studies science has advanced immeasurably since the time of Democritus; however, to this day, one of the major goals of the physical sciences is still to identify and understand the most fundamental building blocks of nature. This is the defining quest of particle physics.

Many profound observations in the late 19th and early 20th centuries led to a revolution in the perceptions which guided microscopic physics. Modern particle physics was born with the discovery of the electron in 1897 by J.J. Thompson. This discovery demonstrated that what we call “atoms” in the modern world, are actually not fundamental. Einstein’s explanation for the photoelectric effect through the quantization of light led to the quantization of the electric field. In 1911, the scattering experiments of Rutherford led to the Bohr model of the atom. Compton’s scattering experiments in 1923 demonstrated the particle nature of light. In 1927, Dirac wrote down an equation which predicted that all fermionic particles should also have antiparticles and, in 1931, the positron was discovered by Anderson. A theory of strong interactions mediated by massive particles was introduced by Yukawa in 1934 to describe the binding of nucleons.

These advances and countless others over the last one hundred years have led to the development of the *Standard Model* (SM) of particle physics. This model describes the electromagnetic, weak, and strong interactions between particles. Under the Standard Model, the strong and electroweak interactions are unified under the gauge group $SU(3) \otimes SU(2) \otimes U(1)$.

The observation by Louis de Broglie that moving bodies have a wave nature has profound consequences in particle physics. The resolution of an optical microscope is approximately proportional to the wavelength of the incident light. Assuming the probing beam consists of particles, then the resolution is limited by the de Broglie wavelength of these particles

$$\lambda = \frac{h}{p}, \tag{1–1}$$

where λ is the de Broglie wavelength, h is Plank's constant, and p is the momentum of the particle beam. This prediction, that an object's momentum is inversely proportional to its wavelength, implies that as the momentum transfer, Q, of the probing beam is increased it is possible to resolve smaller distance scales. Low energy particles only probe large spatial regions, while high energy particles can resolve short distance effects. This observation by de Broglie motivates the use of high energy particle accelerators as the laboratory of particle physics in the modern age.

Through the use of particle colliders, the Standard Model has compiled an impressive history of experimental success. For example, the W and Z bosons were predicted before their discovery by the electroweak theory of Glashow, Weinberg and Salam. The running of the electromagnetic and the strong coupling constants have been verified by experiment. The final quark, the top quark, of the Standard Model was discovered at the Tevatron collider by the two collider experiments, CDF and D0. The SM also survived the plethora of precision electroweak measurements of LEP, the CERN e^+e^- collider. In summary, the Standard Model is a successful model of high energy particle physics at all energies accessible to the experimental community so far.

No deviations from the SM have been observed; however, there are imperfections with the Standard Model. So far the Higgs boson has not been observed. This particle must exist in the Standard Model to allow the basic building blocks of the model to obtain mass. Also, there are many theoretical arguments that suggest that the SM breaks down at higher energy scales. One such argument is related to the mass of the Higgs boson itself. This *hierarchy problem* is related to the sensitivity of the Higgs mass to physics at high energy scales and requires a *fine tuning* (*i.e.,* cancellation to a precision of 10^{32}) of the Standard Model which is undesirable to many theorists. Supersymmetric (SUSY) models, which require a symmetry between fermions and bosons, can provide an elegant solution to this problem. However, so far, no SUSY particles have been discovered. Many models of physics beyond the SM (BSM) such as SUSY require additional particles which are heavier than the SM particles. The particles must be heavy or they would have already been discovered in previous measurements. High energy collisions are required to produce heavy particles in the laboratory. This need to search for heavy particles further motivates the need for high energy colliders in particle physics.

Theorists have used compelling arguments, such as hierarchy, for many years to motivate the need for new particles or new forces to be observed at the TeV scale. Without experimental guidance at the TeV scale, theorists who study physics beyond the Standard Model have had freedom to pursue countless possibilities. The Tevatron at Fermilab is only just beginning to threaten the TeV scale with a center of mass energy of 1.96 TeV, but has the potential to constrain many of these theories. At the same time, the particle physics community is waiting eagerly for the first collisions at the Large Hadron Collider (LHC) which will begin colliding protons in the next few years at a center of mass energy of 14 TeV.

Precision measurements at high energy colliders have been an immensely valuable tool both to validate the SM and to constrain its properties. As higher energy colliders become available, it is possible to make discoveries of BSM physics or to validate and constrain our understanding of the SM. In this dissertation, a measurement which probes the smallest distance scales ever probed by studying the collisions of the highest energy particle accelerator in the world will be discussed. This measurement provides validation of quantum chromodynamics (QCD), the theory of the strong force, at the highest energy ever directly probed, and at the same time provides constraints on the quantum nature of the proton which will improve theoretical predictions for the high energy colliders of the future.

CHAPTER 2
PHENOMENOLOGY OF QUANTUM CHROMODYNAMICS

The Standard Model of particle physics is a quantum field theory (QFT) based on the principle of local gauge invariance of the gauge group

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \tag{2–1}$$

where $SU(3)_C$ is the symmetry group of the strong interactions, called quantum chromodynamics, and $SU(2)_L \otimes U(1)_Y$ represents the symmetry group of the electroweak theory. A complete discussion of QFT [1] and the standard model of particle physics [2, 3] is beyond the scope of this experimental dissertation. Instead, the particle content of the standard model will be reviewed, and only aspects of QCD phenomenology which are relevant to the inclusive jet cross section measurement will be addressed in detail. A more complete discussion of QCD can be found in many references [4, 5, 6, 7].

2.1 The Particle Content of the Standard Model

The particle content of the standard model of particle physics includes six quarks, six leptons, and four gauge bosons (an anti-particle also exists for each particle). Quarks and leptons are spin $\frac{1}{2}$ fermions. The quarks and leptons of the SM can each be arranged into three doublets. Each lepton doublet includes a charged lepton partnered with a neutral neutrino. The quark and lepton content of the standard model is listed in table 2–1 and table 2–2 along with some of the fermion measured properties as listed in the Particle Data Book [8].

The fermions of the standard model interact through the exchange of the integer spin gauge bosons. The four gauge bosons are shown in table 2–3 with some of their properties. The mass-less photon is the propagator of the electromagnetic

Table 2–1. Some properties of the quark content of the standard model. Quark properties are taken from the Particle Data Book.

Flavor	Symbol	Electric Charge (e)	Mass (GeV/c^2)
Up	u	$+\frac{2}{3}$	$1.5 - 3 \times 10^{-3}$
Down	d	$-\frac{1}{3}$	$3 - 7 \times 10^{-3}$
Charm	c	$+\frac{2}{3}$	1.25 ± 0.09
Strange	s	$-\frac{1}{3}$	$95 \pm 25 \times 10^{-3}$
Top	t	$+\frac{2}{3}$	174.2 ± 3.3
Bottom	b	$-\frac{1}{3}$	4.2 ± 0.07

force, the ±W and the Z bosons are the force carriers of the weak interactions, and the eight mass-less gluons (g_i where $i = 1..8$ correspond to the $3^2 - 1$ generators of the $SU(3)$ symmetry group) mediate the strong interaction. A major difference between quarks and leptons is that quarks carry an additional internal degree of freedom called *color*. This is the charge of the strong force and is commonly denoted as red, green, or blue (RGB). Of the fermions, only quarks participate in the strong interactions of QCD.

Table 2–2. Some properties of the lepton content of the standard model. Lepton properties are taken from the Particle Data Book.

Flavor	Symbol	Electric Charge (e)	Mass (MeV/c^2)
Electron	e	-1	0.511
Electron Neutrino	ν_e	0	$< 3 \times 10^{-6}$
Muon	μ	-1	105.7
Muon Neutrino	ν_μ	0	< 0.19
Tau	τ	-1	1777
Tau Neutrino	ν_τ	0	< 18.2

Gravity is not mentioned in the above discussion. Although all massive particles couple to gravity, it is the weakest force and is typically only important on macroscopic scales. Gravity is not included in the standard model. It is not important for the research discussed here and will not be discussed further.

It is a remarkable triumph of the SM that all of the interactions of matter in the observed universe (barring gravity) can be described with amazing precision based on the simple particle content discussed here.

Table 2–3. Some properties of the gauge boson content of the standard model. The gauge boson properties are taken from the Particle Data Book.

Boson	Symbol	Electric Charge (e)	Mass (GeV/c^2)
Photon	γ	0	0
W	W	±1	80.403 ± 0.029
Z	Z	0	91.1876 ± 0.0021
Gluon	g	0	0

2.2 Feynman Rules of QCD

The theory of QCD describes the interactions of the spin $\frac{1}{2}$ quarks and the spin 1 gluons. Requiring that QCD be a gauge theory based on the group $SU(3)$ with three color charges fixes the Lagrangian density to be

$$\mathcal{L} = -\frac{1}{4}\sum_{A=1}^{8} F^{A\alpha\beta}F^{A}_{\alpha\beta} + \sum_{j=1}^{n_f} \bar{q}_j(i\not{D} - m_j)q_j, \tag{2–2}$$

where q_j are the quark fields of n_f different flavors and mass m_j. The γ^α are the Dirac matrices and D_α ($\not{D} \equiv D_\alpha\gamma^\alpha$) is the covariant derivative defined by

$$(D_\alpha)_{ab} = \partial_\alpha\delta_{ab} + ig(t^C\mathcal{A}^C_\alpha)_{ab}, \tag{2–3}$$

where g is the gauge coupling of QCD, and t^C are the matrices of the fundamental representation of $SU(3)$. These generators obey the commutation relations

$$[t^A, t^B] = if^{ABC}t^C, \tag{2–4}$$

where f^{ABC} are the complete antisymmetric structure constants of $SU(3)$. The normalization of the structure constants and of g is specified by

$$Tr[t^At^B] = \delta^{AB}/2. \tag{2–5}$$

In analogy with Quantum Electrodynamics (QED) we may also make the definition

$$\alpha_s = \frac{g^2}{4\pi}. \tag{2–6}$$

The quantity $F^A_{\alpha\beta}$ is the field strength tensor derived from the gluon field $\mathcal{A}^A_\alpha$

$$F^A_{\alpha\beta} \;=\; \partial_\alpha g^A_\beta - \partial_\beta g^A_\alpha \;-\; g f^{ABC} \mathcal{A}^B_\alpha \mathcal{A}^C_\beta . \tag{2–7}$$

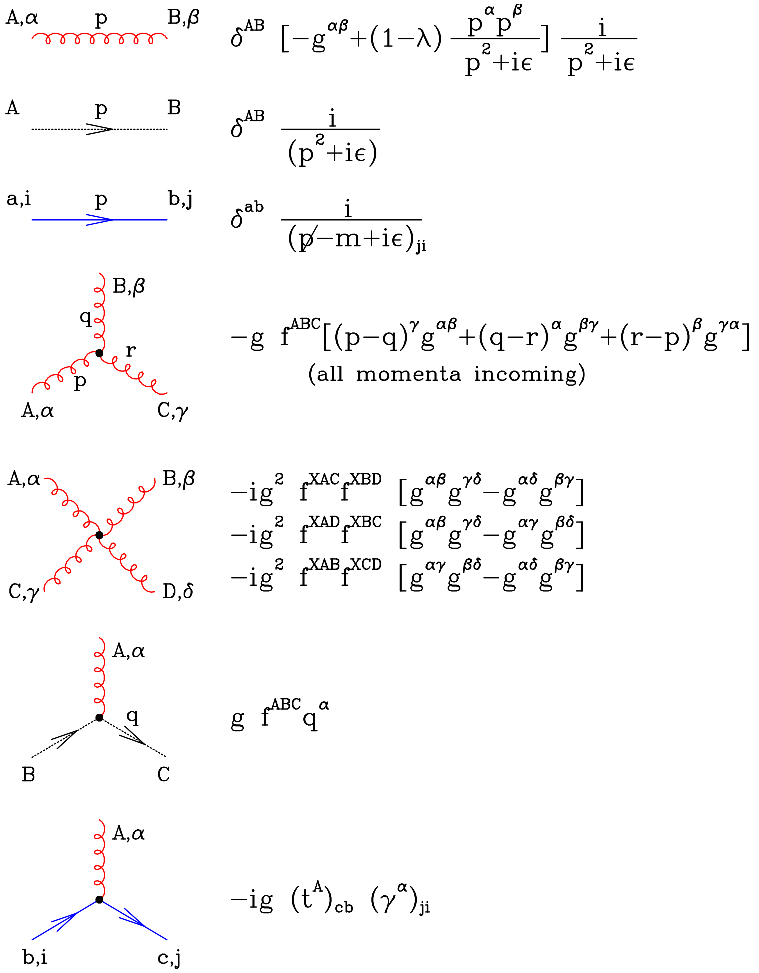


Figure 2–1. Feynman rules for QCD in a covariant gauge. The solid lines (black) represent fermions, curly lines gluons, and dotted lines (black) ghost propagators.

The Feynman rules of QCD can be calculated directly from the QCD Lagrangian after fixing a gauge. The gauge must be fixed in order to define the gluon propagator. The Feynman rules in a covariant gauge are given in Fig 2–1 [9]. The physical vertices in QCD include the gluon-quark-antiquark vertex. This vertex and the physical propagators of the quark and gluon are analogous to the coupling and propagators of the electron and photon of QED. However, there is also the three-gluon and four-gluon vertices, of order g and g^2 respectively. These gluon self couplings have no analogue in QED since the photon does not carry electric charge and therefore does not interact with other photons. They arise from the third term of Eq. 2–7 which is not present in the QED field strength.

The Feynman rules discussed above can be used to make many predictions of QCD. For example, one can compute the probability that a given initial state will interact to yield a final state, $P(A + B \rightarrow C + D)$. In particle physics, these probabilities are called *cross sections*, σ. Cross sections are expressed in units of area called *barns*, b. One barn is equal to $10^{-24}\ cm^2$. In this analysis, cross section results are expressed in nano-barns ($1\ nb = 10^{-9}\ b$). Once a cross section for a process has been calculated, and the total number of collisions which have occurred in some time period (*integrated luminosity*, $\mathcal{L}$) has been measured, it is possible to predict the number of events of that process that have occurred

$$N = \mathcal{L}\sigma.^{1} \tag{2–8}$$

The instantaneous luminosity is a measure at a specific time of the number of collisions per unit time occurring in the collider. The integrated luminosity is a measure of the number of $p\bar{p}$ collisions that have occurred over some period of time. Instantaneous luminosity is usually measured in units of $cm^{-2}s^{-1}$. Recently, the Tevatron has achieved instantaneous luminosities greater than $200 \times 10^{30}\ cm^{-2}s^{-1}$. Integrated luminosity is measured in nb^{-1}. A process with a cross section of 1 nb will be created approximately 2000 times per day at current CDF luminosities

---

[1] Due to the substructure of the proton, the cross section calculated from the Feynman rules must be convoluted with the quark and gluon density functions of the proton before it can be used to predict the number of events expected at $p\bar{p}$ colliders like the Tevatron.

(*i.e.,* the integrated luminosity for a 24 hour period at CDF can be on the order of 2000 nb^{-1}).

2.3 Color Confinement and Asymptotic Freedom

The Feynman rules shown in figure 2–1 are made up of *tree-level* diagrams (*i.e.,* diagrams which do not contain loops). Most calculations made in QFT are an approximation based on an expansion in powers of the coupling constant. This is called a *perturbative* expansion, and is only valid in the limit that the coupling constant is small (*i.e.,* $\alpha << 1$). Perturbative predictions for observables such as scattering amplitudes are affected by higher order loop corrections. The vacuum polarization diagram for QED shown in figure 2–2 is an example of such a correction. This correction to the photon propagator diverges logarithmically at the one loop level as the four-momentum squared of the virtual photon ($q^2 = -Q^2$) increases. In the high energy limit ($Q^2 >> m_e^2$) the contribution is

$$\alpha_0 B(Q^2) = -\frac{\alpha_0}{3\pi}\{log(\lambda^2/Q^2) + \frac{5}{3}\} \tag{2–9}$$

where λ is an ultraviolet cutoff and α_0 is known as the *bare* electric charge ($\alpha_0 = e_0^2/4\pi$). An effective coupling may be defined which sums the vacuum bubbles to all orders

$$\alpha_{eff}(Q^2) = \alpha_0(1 + \alpha_0 B(q^2) + \alpha_0 B(q^2 \alpha_0 B(q^2) + ...) = \frac{\alpha_0}{1 - \alpha_0 B(q^2)}. \tag{2–10}$$

This procedure for defining an effective coupling to absorb the ultraviolet diverges in the theory into the unobservable bare coupling is called renormalization. The effective coupling now varies (*i.e., runs*) with the energy scale of the problem. The long distance behavior of the effective coupling is used to define the experimental coupling

$$\alpha \equiv \alpha_{eff}(Q^2 = 0) \sim \frac{1}{137}. \tag{2–11}$$

In the high energy limit (large Q^2)

$$\alpha_{QED}(Q^2) \equiv \alpha_{eff}(Q^2) = \frac{\alpha}{1 - \frac{\alpha}{3\pi} log(\frac{Q^2}{m_e^2})}. \tag{2–12}$$

From Eq. 2–12, we see that the coupling constant of QED increases with energy. At low energy, α_{QED} is small and a perturbative expansion in the coupling is relevant. At high energy, the coupling gets large and eventually diverges (Landau pole). Perturbation theory is not valid in this regime. Luckily, this occurs at a very high energy scale for QED, and it is expected that physics beyond the standard model should modify the running of the coupling at an energy far below the Landau pole.

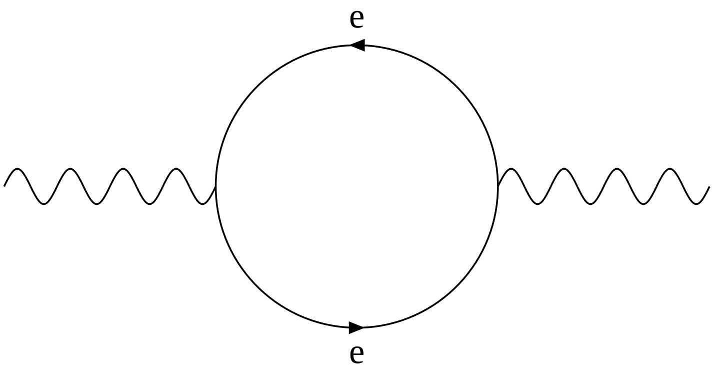


Figure 2–2. One-loop Vacuum polarization diagram of QED. Diagrams like this affect the rate for electron-positron scattering in QED. This diagram diverges logarithmically.

The QCD diagram analogous to the QED diagram shown in figure 2–2 is shown in figure 2–3. In QCD, there is an additional contribution to the propagator of the gluon due to its self coupling. This extra diagram (shown in figure 2–4) leads to profound consequences with respect to the running of the coupling constant. By following similar arguments as those applied to QED and including the extra QCD diagrams, one arrives at the analogous equation to 2–12

$$\alpha_s(Q^2) \equiv \alpha_{eff}(Q^2) = \frac{\alpha(\mu_r^2)}{1 - \frac{\beta_0 \alpha(\mu_r^2)}{4\pi} log(\frac{Q^2}{\mu_r^2})}. \tag{2–13}$$

The quantity β_0 is the one-loop β-function of QCD

$$\beta_0 = 11 - \frac{2}{3} n_f, \tag{2–14}$$

where n_f is the number of quark flavors. In QED, the one-loop β-function was

$$\beta_0^{QED} = -\frac{4}{3}. \tag{2–15}$$

The sign of the β-function is different for QED and QCD if $n_f < 16$. This has the consequence that the effective coupling of QCD runs in the opposite direction of the QED effective coupling. Equation 2–13 represents only the leading order

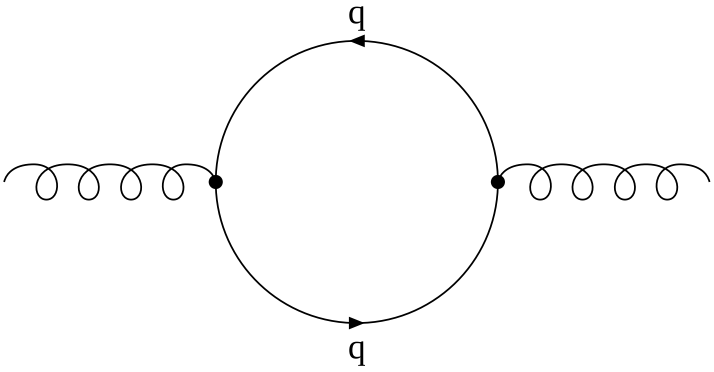


Figure 2–3. One-loop vacuum polarization diagram of QCD. This diagram is analogous to the vacuum polarization diagram of QED.

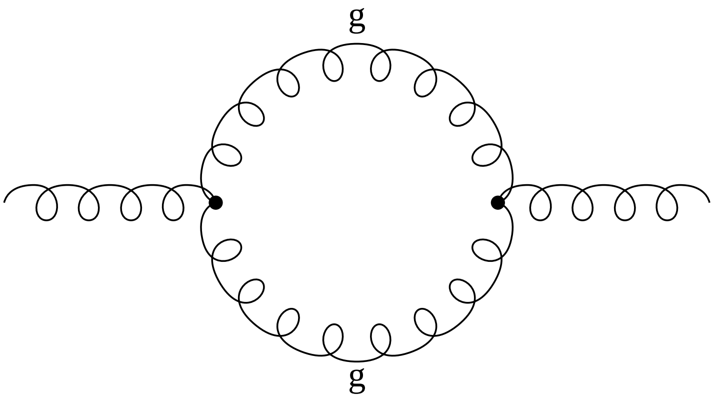


Figure 2–4. One-loop Vacuum polarization diagram of QCD which arises from the gluon self coupling. Contributions from this diagram are responsible in the sign flip for the β-function of QCD with respect to QED.

behavior of the coupling constant in QCD, where the experimental charge $\alpha(\mu_r^2)$ has been defined at an arbitrary renormalization scale ($Q^2 = \mu_r^2$). The low energy limit is not a useful scale in QCD. This is because the coupling α_s diverges in the low energy limit. Typically the arbitrary scale μ_r is taken to be the mass of the Z boson where

$$\alpha_s(M_Z) \sim 0.12. \tag{2–16}$$

The main features of Eq. 2–12 and Eq. 2–13 are sketched in figure 2–5. For QED, we see the dynamics that were mentioned above; for low Q^2 (large distances) the coupling constant is small and the coupling increases with Q^2 until at some very high energy scale ($\sim\ 10^{34}\ GeV$) it diverges. For QCD, the dynamics are very

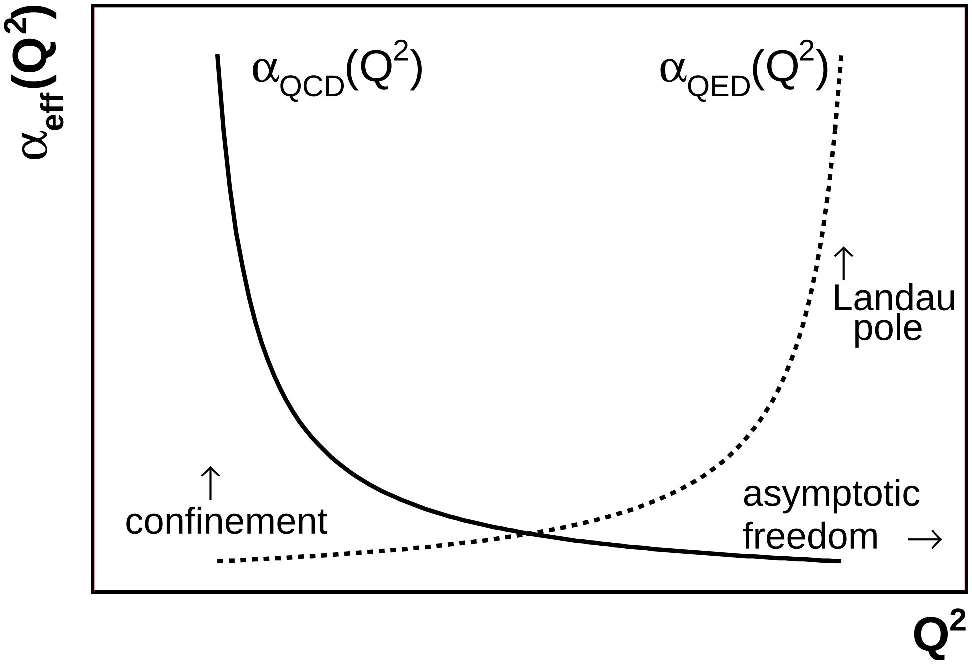


Figure 2–5. Running of the the QED and QCD coupling constants. In QED the effective coupling is small at large distances, but diverges at very high energy (“Landau pole”). In QCD the coupling diverges at large distances (“color confinement”) and goes to zero asymptotically at large energy (“asymptotic freedom”). Color confinement and asymptotic freedom are important qualities of QCD.

different. At low Q^2 the coupling diverges. This is known as color confinement and is the reason why free quarks and free glouns are not observed in nature. As objects connected by color fields are separated the field strength becomes so strong that new quark-antiquark pairs are pulled from the vacuum. These quark-antiquark pairs form color neutral singlets called hadrons. This process of colored *partons* (*i.e.,* quarks and gluons) forming color singlet hadrons is referred to as hadronization. Through this process a colored parton can hadronize into many hadrons which are roughly co-linear with respect to the momentum vector of the original parton. These clusters of roughly co-linear hadrons are called jets.

At high Q^2 (particles resolve small distances) the coupling constant of QCD becomes small and even vanishes asymptotically. This is the phenomena known as asymptotic freedom. A consequence of asymptotic freedom is that perturbative methods are valid at high energies in QCD. Because of this property of QCD, the long distance (low Q^2) and short distance (high Q^2) behavior of QCD may be separated (*i.e., factorized*). This factorization allows the methods of perturbative QCD (pQCD) to be applied to the large Q^2 component and phenomenological models to be applied to low Q^2 component. This factorization property means that the partonic cross section calculated with the methods of pQCD is useful for hadronic collisions in the high energy limit.

Another important result that may be obtained directly from Eq. 2–3 is the energy scale at which the coupling constant in QCD diverges (Λ_{QCD}). Solving for the energy scale (Q^2) where the denominator of Eq. 2–3 vanishes yields

$$log(\Lambda^2_{QCD}) = \frac{4\pi}{\beta_0 \alpha(\mu_r^2)} + log(\mu_r^2). \tag{2–17}$$

This result can be used to obtain the effective couping in terms of Λ_{QCD}

$$\alpha_s(Q^2) = \frac{4\pi}{\beta_0 log(Q^2/\Lambda^2_{QCD})}. \tag{2–18}$$

The quantity Λ_{QCD} has been determined experimentally to be roughly 200 MeV. Therefore, the QCD effective coupling gets large for $Q^2 \sim 1\ GeV^2$. Perhaps, it is more than coincidence that this is roughly the mass of the proton.

2.4 The Factorization Theorem

The observation by J.D. Bjorken in 1969 from deep inelastic scattering experiments, that protons when probed with sufficiently high momentum behaved like free partons, had profound implications [10]. A few years later Drell and Yan extended these parton model ideas to some hadron-hadron processes [11]. The idea of the factorization model can be seen pictorially in figure 2–6 [9]. It means that the hadronic cross section may be written as

$$\sigma(\mathrm{P_1, P_2}) = \sum_{\mathrm{ij}} \int \mathrm{dx_1 dx_2 f_i(x_1, \mu_f^2) f_j(x_2, \mu_f^2) \hat{\sigma}_{ij}(p_1, p_2, \alpha_s(\mu_f^2), Q^2/\mu_f^2)}, \qquad (2\text{–}19)$$

where the momenta of the partons which engaged in the partonic scattering is $p_1 = x_1 P_1$ and $p_2 = x_2 P_2$, x_1 and x_2 are the momentum fraction of the hadron carried by the interacting partons, $f_{i/p(\bar{p})}(x_i)$ are the quark and gluon parton density functions (PDFs) defined at the arbitrary factorization scale μ_f, and $\hat{\sigma}_{\mathrm{ij}}$ is the partonic (short-distance) cross section for the scattering of partons of type i and j. At leading order in QCD the parton cross section is directly calculated from the leading order tree diagrams. However, at higher order, there are long-distance contributions which must be factored out and absorbed into the PDFs of the incoming hadrons. This factorization is possible to all orders in perturbation theory, and is the property of QCD which makes it a useful tool in calculations for hadron collisions.

In principle, the factorization scale (μ_f) and the renormalization scale (μ_r) can differ. However, in practice, it is convenient to set both scales equal to the hard scattering scale ($\mu = \mu_r = \mu_f = Q$). For the inclusive jet cross section, the hard scattering scale (Q) is often taken to be one half of the jet transverse momenta.

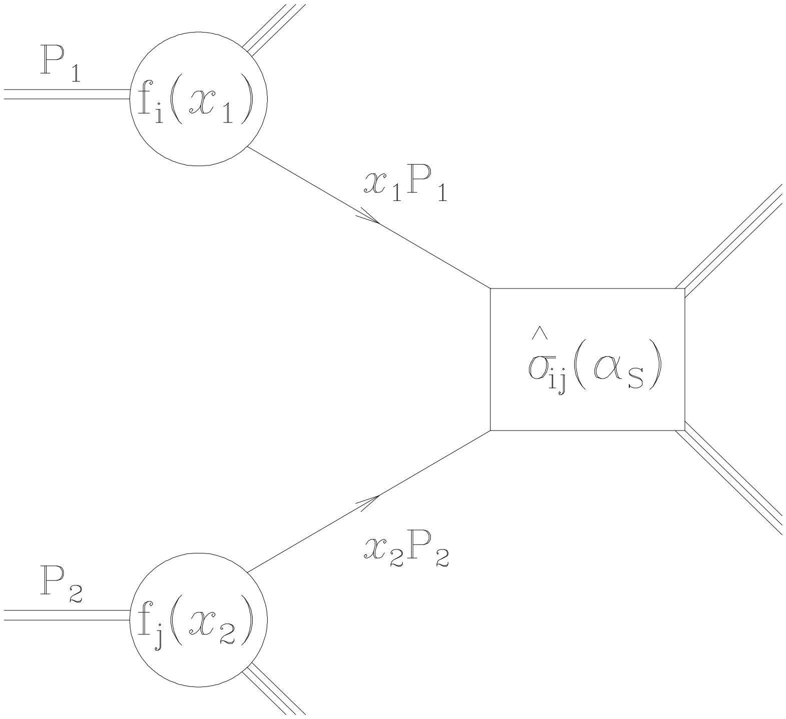


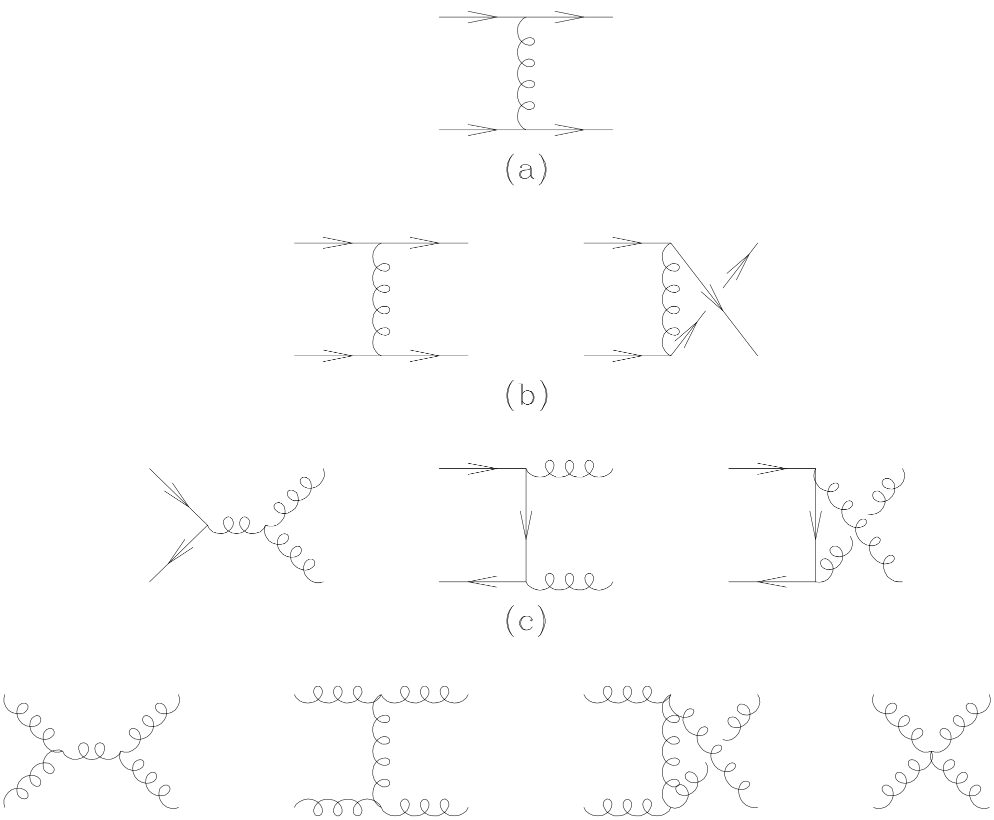
Figure 2–6. Schematic of the QCD factorization theorem. The partonic cross section must be folded in with the parton density functions of the hadron.

2.5 Jet Production Cross Sections

Diagrams contributing to jet production at leading order are shown in figure 2–7 [9]. These diagrams may be read from left to right, or bottom to top. For example, 2–7(c) can be interpreted as $q\bar{q} \rightarrow gg$ when read from left to right, or it may be interpreted as $g\bar{q} \rightarrow g\bar{q}$ when read from bottom to top.

Lowest order (LO) calculations have uncertainties for multiple reasons. The leading order result quite often has a large dependence on renormalization and factorization scales. This dependence is reduced by going to higher order in the perturbative expansion. Another source of uncertainty on LO predictions is that additional processes may become possible only when going beyond leading order.

At next to leading order (NLO), all Feynman diagrams which contribute an additional factor of α_s to the scattering amplitude must be considered when calculating the scattering cross section. Extra factors of the strong coupling



(a)

(b)

(c)

(d)

Figure 2–7. Diagrams which contribute to leading order jet production at a hadron collider.

constant can be added in two ways. Real radiation may be added to the initial or final state, or diagrams may contain one loop. As a simple example, consider the case of electron-positron annihilation to hadrons through a virtual photon exchange($e^+e^- \to q\bar{q}$). The diagrams which contribute to this process at NLO are shown in figure 2–8 [9].

The diagrams in figure 2–8(a) include the tree-level diagram for the process as well as one-loop diagrams with a virtual gluon emission. Although the virtual gluon diagrams have two extra factors of g_s due to the two extra vertices, they still contribute at NLO. Since these diagrams have the same final state as the tree-level diagram, the matrix element for the sum of the four diagrams must be squared. Cross terms from this squared matrix element are at NLO in α_s (*i.e.,*

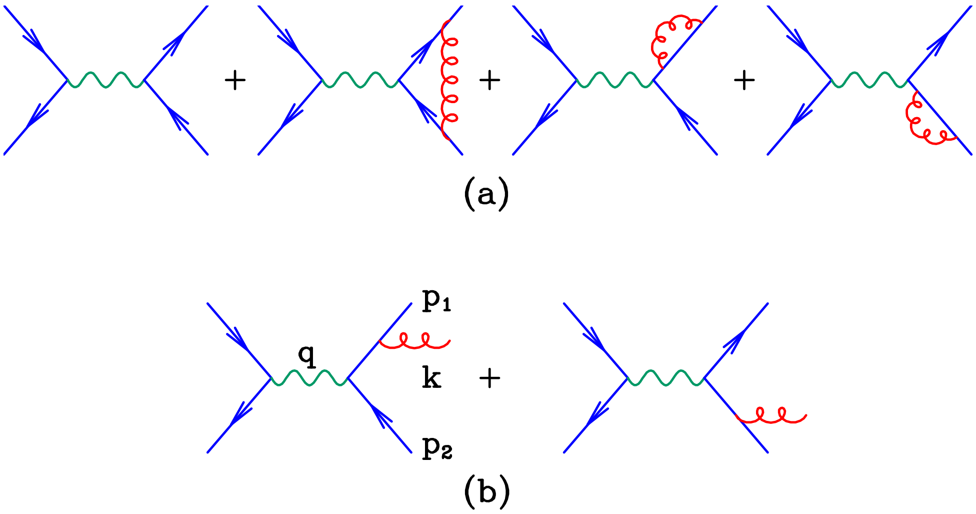


Figure 2–8. Diagrams which contribute to e^+e^- annihilation to hadrons at NLO. The divergences in the real gluon emission diagrams cancel the divergences in the loop diagrams.

$O(\alpha_s^2)$). These cross terms have infrared ($Q^2 \rightarrow 0$) divergences related to the integral over the loop momenta. Figure 2–8(b) shows the NLO diagrams with a real gluon emission. The real gluon emission diagrams diverge in the limit that the gluon is collinear to the quark. A very powerful theorem of QFT states that soft and collinear divergences cancel to all orders in perturbation theory [12, 13, 14]. This cancellation means that inclusive quantities will be free of divergences. The inclusive requirement means that the observable can not include only the diagrams in figure 2–8(a) or only the diagrams in figure 2–8(b) because contributions from all of these diagrams must be included to ensure that the divergences are canceled. It is from arguments of this type that the jet cross section must be inclusive, and jet algorithms must be defined in such a way as to not be sensitive to infrared and co-linear effects.

2.6 Structure of Hadronic Collisions

When protons of equal and opposite momentum collide at high energy there is a large probability that the protons will break up and the resulting hadrons will continue in roughly the same direction with very little transverse momentum relative to the beam direction. Sometimes however, there is a hard scattering where particles with large transverse momentum are generated in the collision. The factorization theorem of QCD discussed above allows one to factor out the hard scattering component of the hadron collider event and calculate it perturbatively. However, there are other components that must be included for a complete model of the hadron collider event.

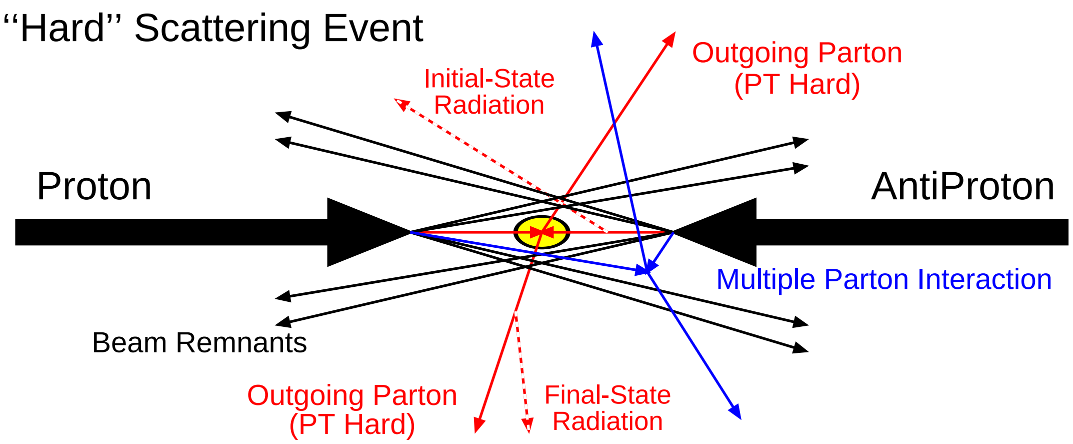


Figure 2–9. A cartoon description of a typical hadron collider event at the parton level (before hadronization of colored partons into color singlet hadrons). The hard scattering, initial state radiation, final state radiation, multiple parton interactions, pileup, and the beam remnants are the components of an event at a hadron collider.

Figure 2–9 shows a simplified description of a hard scattering event at a hadron collider. The schematic shown can be considered to represent what goes on within the radius of a proton around the hard collision. Once the colored partons move outside of the radius of the proton, they must hadronize into color neutral hadrons due to the requirement of color confinement. The state of the event before hadronization is not a physical observable, but is useful when discussing the phenomenology of hadron collider events and will be referred to as the parton level. This simple parton level model of hadron-hadron scattering is used as the basis for the QCD Monte-Carlo (MC) event generator programs.

As partons begin to *feel* the effect of other partons, they radiate quarks and gluons. These quarks and gluons can also radiate more quarks and gluons and this series of partons radiating more partons leads to an *avalanche* or *shower* of partons. This chain of radiation is called a parton shower. In the context of a QCD event generator, a parton shower is an approximate perturbative treatment of QCD parton splitting which is valid above some cut-off value ($Q_0 \sim 1\ GeV$). The parton shower is based on identifying and summing to all orders the logarithmic enhancements due to soft gluon emission and gluon splitting functions. Because the parton shower is based on enhancements due to soft gluon emission (small angles) it is only an approximation of the hard gluon emission component (large angles). Parton showers are used to model initial state radiation (ISR), and final state radiation (FSR) in QCD MC generators. These models for the parton shower can be combined with phenomenological models of hadronization which take over for energy scales below the cut-off scale ($Q < Q_0$).

As shown in figure 2–9, it is also possible that there is a second parton-parton interaction within the same proton-antiproton collision. This is referred to as multiple parton interactions (MPI). Because the protons and antiprotons collide in bunches, it is possible that multiple proton-antiproton collisions occur in the same bunch crossing. This is commonly referred to as *pileup*. The rate of pileup collisions is proportional to the luminosity and can be studied be by looking at the number of secondary vertices in the event.

Beam-beam remnants and multiple parton interactions define the *underlying event* (UE). The underlying event is always present at hadron colliders and increases the difficulty of resolving the properties of the hard scattering process. Separating particles from the UE and particles which come directly from the hard scattering is not possible on an event by event basis. However, the topological structure of hadron-hadron collisions can be used to study the the average properties of the UE contribution. An analysis based on this strategy was published in Run I at CDF[15, 16], and has been continued and improved in Run II at CDF[17, 18, 19]. By studying distributions which are sensitive to the underlying event the MC programs are *tuned* to fit the effects observed in data.

The leading order matrix elements, PDFs, parton showers, hadronization models, the underlying event, and pileup are required components of a QCD event generator for complete generation of hadron collider events.

After all of the partons in the event hadronize, the particle content of the event is referred to as the hadron level. The particles at the hadron level are observable, and they are the states which interact with the detector. After the particles interact with the detector, the resulting description of the event is referred to as the detector level. These three levels of the event (parton, hadron, and detector) will be referred to throughout this draft and are depicted in figure 2–10 [20].

Experimental measurements are only available at the detector level. However, the MC generators, when combined with a detector simulation program, can be used to make predictions at all three levels [2] . MC generators combined with detector simulations are a useful tool for deriving corrections to the data.

---

[2] The hadron level in the Monte Carlo generators is defined as all final-state particles with lifetime above 10^{-11} s.

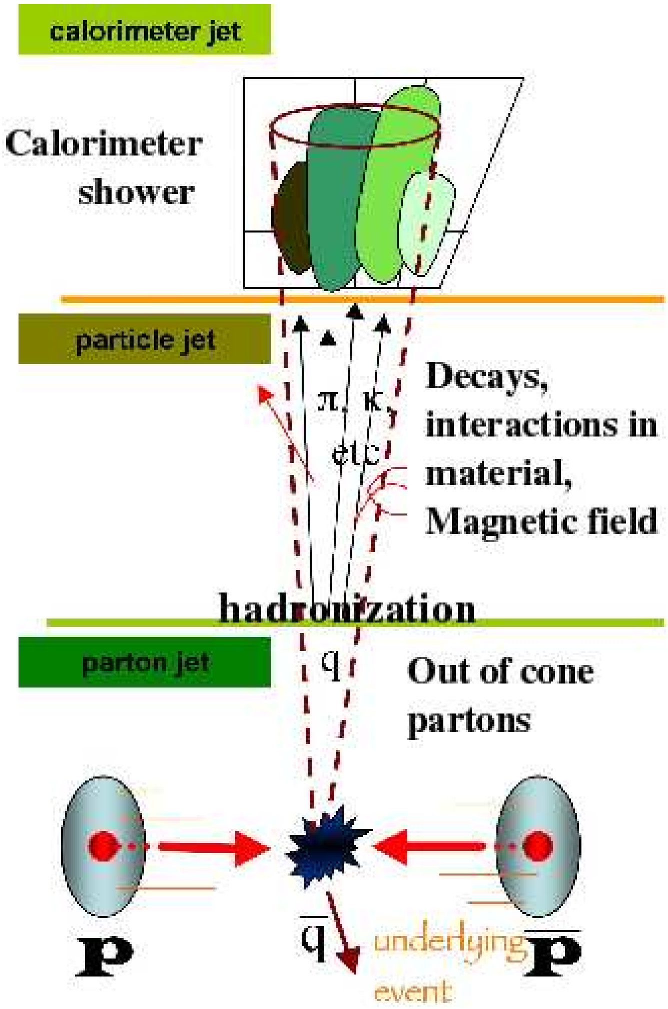


Figure 2–10. A cartoon description of the different levels of a jet event at CDF. The parton level is the state before the partons hadronize and is not physical observable. The hadron or particle level is the state after hadronization but before the particles have interacted with the detector. Finally, the detector level is the result of the event as reported by the detector.

CHAPTER 3
EXPERIMENTAL APPARATUS

3.1 Coordinates and Conventions

The CDF detector is naturally described by a cylindrical coordinate system (ρ, z, ϕ). The z-coordinate is taken along the proton beam direction with the origin at the center of the detector. Defining the x-axis to point away from the center of the accelerator ring fixes the azimuthal angle ϕ. The ρ-coordinate is the perpendicular distance from the z-axis. It is also useful to define the polar angle θ, which is usually expressed through the pseudorapidity

$$\eta \equiv -\ln \tan(\theta/2). \tag{3–1}$$

A value of $\eta = 0$ corresponds to $\theta = 90°$, and a value of $\eta = 1$ corresponds to $\theta \sim 40°$. The rapidity Y is defined as

$$Y \equiv -\frac{1}{2} \ln \left\{ \frac{E + P_z}{E - P_z} \right\}, \tag{3–2}$$

where E denotes the energy and P_z is the momentum component along z-direction. Rapidity is a useful quantity at a hadron collider because it is invariant under boosts in the z-direction up to a constant. In the limit that the momentum of a particle is much larger than its mass, Y and η are equivalent. One difference between Y and η is that Y does not correspond to a definite θ value.

Because the interacting partons need not balance in momentum along the z-direction, the colliding system can have an arbitrary boost in the lab frame. Momentum of the interacting partons is not known in the direction parallel to the beam; however, the initial parton momenta is roughly zero in the direction perpendicular to the beam. The transverse momentum vector is a two dimensional vector perpendicular to the beam direction and is defined by

$$\vec{P}_T \equiv P_x\hat{i} + P_y\hat{j}, \tag{3–3}$$

where P_x and P_y are the x and y components of the momentum. The sum of the transverse momentum vectors of all particles in the event should sum to zero in the absence of any particles escaping detection. The magnitude of the transverse momentum is given by $P_T \equiv P\sin\theta$, where P is the magnitude of the momentum vector. Similarly, the transverse energy vector is defined as

$$\vec{E}_T \equiv E\sin\theta\cos\phi\hat{i} + E\sin\theta\sin\phi\hat{j} \equiv E_T(\cos\phi\hat{i} + \sin\phi\hat{j}). \tag{3–4}$$

The magnitude of the transverse energy is $E_T = E\sin\theta$.

Some other useful quantities are the missing transverse energy vector ($\vec{\not{E}}_T$), and the missing E_T significance. The missing transverse energy vector is defined by

$$\vec{\not{E}}_T \equiv -\sum_i E_T^i\mathbf{\hat{n}_i}, \tag{3–5}$$

where $\mathbf{\hat{n}_i}$ is a unit vector perpendicular to the beam axis and pointing at the ith calorimeter tower. Missing E_T significance is defined by

$$\widetilde{\not{E}_T} \equiv \not{E}_T/\sqrt{\sum E_T}, \tag{3–6}$$

where $\not{E}_T$ is the magnitude of $\vec{\not{E}}_T$. $\widetilde{\not{E}_T}$ is a useful variable to remove backgrounds which do not originate symmetrically from the center of the detector (*e.g.,* cosmic rays).

3.2 The Tevatron

The Tevatron accelerator at the Fermi National Accelerator Laboratory (FNAL or Fermilab) provides the highest energy proton-antiproton collisions available in the world with a center of mass energy of 1.96 TeV. Experimental discoveries would not be possible at CDF without the beam provided to the experiment by Fermilab.

The colliding beams at the CDF and D0 experiments are the result of the complex accelerator chain shown in figure 3–1[21]. A Cockroft-Walton accelerator starts the process by accelerating Hydrogen ions (Hydrogen atoms with one extra electron) to 750 keV. The ions are then injected into the 500 ft long Linac where their energy is boosted to 400 MeV by oscillating electric fields. The electrons are then stripped from the ions by a carbon foil. The remaining protons then enter a fast-cycling synchrotron ring called the Booster. Here the beam is accelerated by radio frequency (RF) cavities at each revolution until they reach an energy of 8 GeV. Bunches of protons, each containing about 5×10^{10} protons, are passed on to the Main Injector.

Proton bunches from the Main Injector are also used to create antiprotons ($\bar{p}$) by collisions with a nickel-copper target. This technique produces antiprotons with a wide range of momentum which must be *cooled* into a mono-energetic beam. The antiprotons are first focused with a lithium collector lens and then passed into the the Debuncher. The Debuncher applies complex computer-controlled RF techniques to cool the antiproton beam as much as possible. Correction signals are applied to individual particles in order to further stocastically cool the antiproton beam. An 8 GeV beam emerges and is passed on to the Accumulator where $\bar{p}$ bunches are *stacked* (*i.e.,* accumulated) at rates as high as 10^{12} antiprotons per hour.

The beam is then passed on to the Recycler ring, which is an 8 GeV magnetic storage ring that utilizes stochastic cooling systems. As its name suggests, the Recycler is also capable of recovering antiprotons left over at the end of a *store* (*i.e.,* period of colliding beam time). Once the accumulated antiproton beam reaches 8 GeV it can be extracted into the Main Injector and accelerated to 150 GeV.

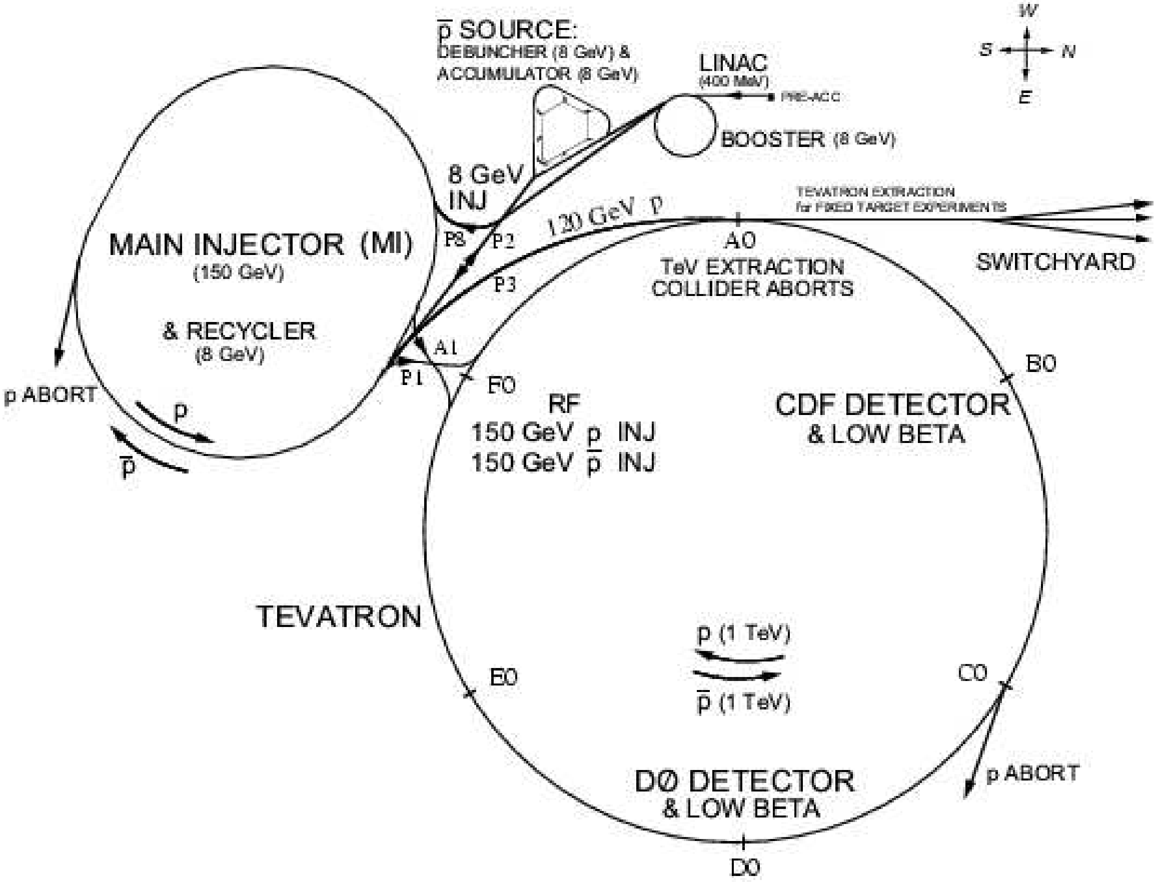


Figure 3–1. A schematic of the accelerator complex used for Run II at Fermilab. The accelerator process may be divided into eight steps. Each step in this process is summarized in the text.

The Main Injector is a synchrotron ring located next to the Tevatron. It was a Run II upgrade to replace the Main Ring. The Main Ring was located in the Tevatron tunnel and was replaced with the Main Injector because it caused beam backgrounds in the collider detectors. Proton bunches exiting the Booster are combined by the Main Injector into a single high intensity bunch of approximately 10^{12} protons. Protons are transferred to the Tevatron after reaching an energy of 150 GeV.

The Tevatron is the largest of the Fermilab accelerators, with a circumference of approximately 4 miles. It is a circular synchrotron with eight accelerating cavities. The Tevatron accepts both protons and antiprotons from the Main

Injector and accelerates them from 150 GeV to 980 GeV in opposite directions. Once the beams energy has reached 980 GeV, they are squeezed to small transverse dimensions by quadrupole magnets at the interaction points (the centers of the CDF and D0 detectors). The beam can be stored in the Tevatron while colliding for many hours. Typically, collisions continue until there is some failure, or the remaining collider luminosity is low and the antiproton stack is large enough to begin a new store.

3.3 The CDF Detector

The CDF detector is described in detail in [22, 23]. Here, those components of the detector which are crucial to this measurement are briefly discussed. A detailed schematic drawing on the CDF detector is shown in figure 3–3 [23].

Although it is not shown in the figure, the Cerenkov Luminosity Counter (CLC) is a critical component of this analysis [24]. When charge particles travel faster than the speed of light in a medium the radiation that they emit becomes coherent. This is a similar phenomena to a sonic boom which occurs when something travels faster than the speed of sound. This effect is used to measure the average number of inelastic $p\bar{p}$ collisions per bunch crossing in order to calculate the instantaneous luminosity delivered by the Tevatron. The instantaneous luminosity provided by the CLC must be integrated with respect to time to calculate the integrated luminosity. The total integrated luminosity included in this measurement is approximately 1000 pb^{-1} or 1 fb^{-1}, and is used for the normalization of the cross section. The Cerenkov counters are located in the region $3.7 < |\eta| < 4.7$.

The central tracking system consists of a silicon vertex detector inside a cylindrical drift chamber. The drift chamber is referred to as the central outer tracker or COT. Charge particles ionize atoms in the gas (argon-ethane $50:50$) of the COT as they pass through the detector. The liberated electrons are accelerated by an electric field towards the nearest positively charged wire to energies high enough to cause secondary ionization. The electrons arising from this chain of ionization are collected on the wire and an electronic pulse is read out. The COT provides charged particle reconstruction up to $|\eta| = 1.0$. Surrounding the tracking detectors is a superconducting solenoid which provides a 1.4 T magnetic field.

Tracks are the reconstructed paths of charged particles in the magnetic field based on the wires that collected electronic signals. These tracks can be traced back to their point of closest approach to the beam line (*i.e.,* impact parameter). A place along the beam line where multiple tracks intersect is called an interaction point or vertex. Vertices are signs that a $p\bar{p}$ interaction occurred at that position. At CDF there may be multiple interaction points in the same event due to pileup. For each event a *primary* vertex is reconstructed. This is defined as the vertex with the highest sum P_T (sum of the P_T of all tracks pointing to the vertex). The position of the primary vertex can vary significantly with respect to the center of the detector at CDF due to the length of the colliding bunches. The length of the bunches is roughly 50 cm in the z-direction. For calculating jet properties, the primary vertex is used as the origin of the coordinate system. Shifting the origin of the coordinate system along the z-direction changes θ (η) and therefore the values of P_T and E_T are modified. The number of extra vertices in the event is a good indicator of the number of multiple interactions within a bunch crossing (pileup). In summary, the tracks reconstructed by the COT are used in the inclusive jet analysis to reconstruct the vertices in each event for two reasons: to determine the primary vertex in the event which is used to define the origin of the event coordinate system, and to determine the number of secondary vertices in the event which is used to count the total number of interactions that occurred in the bunch crossing.

The tracking chambers above only detect charged particles. However, on average, approximately 40% of the energy in an event is carried by uncharged particles. Calorimeters are used to determine the energy and position of both charged and uncharged particles by their total absorption. Electromagnetic particles (photons and electrons) and hadronic particles interact with the detector material differently. The electromagnetic and hadronic calorimeters are all sampling calorimeters (*i.e.,* alternating layers of absorber and detector material); however, each component is composed of different material and has different depth based on the interaction properties of the particles it was designed to measure.

At high energies, when electrons or positrons interact with matter, the dominant way in which they lose energy is through radiation of photons (*i.e., bremsstrahlung*: $e^- \rightarrow e^-\gamma$). For high energy photons the dominant interaction process is pair production (*i.e.,* $\gamma \rightarrow e^+e^-$). An initial electron or photon will interact through these two processes to produce a shower of photons and electrons in the detector. This phenomena is sketched in figure 3–2 and is referred to as an electromagnetic (EM) shower. The shower develops until the energy reaches a critical energy ($E_c \sim 600\ MeV/c$) and ionization losses equal those of bremsstrahlung. The depth is governed by the radiation length (X_0) of the material and only increases logarithmically with the energy of the particle which initiated the shower. The energy resolution of the electromagnetic calorimeter is limited by statistical fluctuations in shower development [25].

Hadrons interact with matter through much different interactions than the ones which lead to the EM shower described above. However, a similar phenomena occurs which is referred to as a hadronic shower. An incident hadron undergoes an inelastic collision with nuclear matter in the detector resulting in secondary hadrons. These hadrons also undergo inelastic collisions. Because many different processes contribute to the development of a hadronic shower, the modeling of the

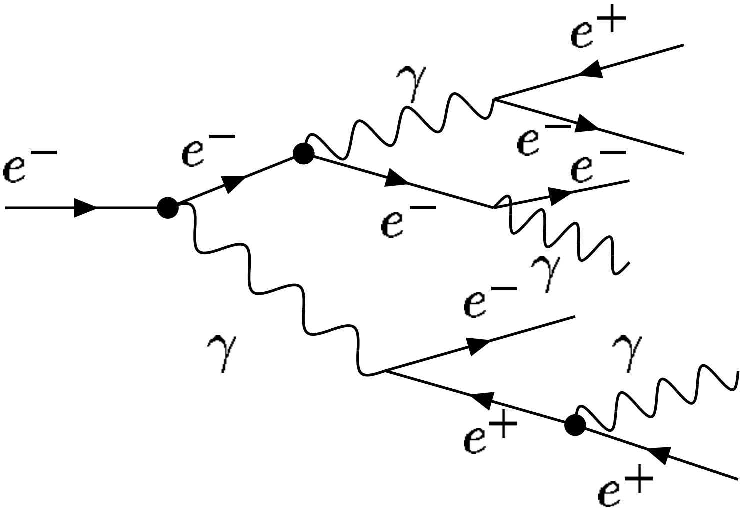


Figure 3–2. Development of an electromagnetic shower. Electrons and photons shower when they interact with the detector through the processes of Bremsstrahlung and pair production.

shower is much more complex than an EM shower. For example, neutral π mesons (π^0's) may be produced. Neutral π mesons primarily decay into two photons which instigate an EM shower. Fluctuations, such as the number of π^0's which are produced early on in the hadronic shower, lead to an energy resolution which in general is much worse than the resolution of EM calorimeters. The depth of a hadronic shower is governed by the nuclear interaction length of the detector material. For most materials the nuclear interaction length is much larger than the radiation length. This means that hadronic showers typically pass through more material before starting to shower, and the showers typically take up more detector volume. This is the reason that hadronic calorimeters are located outside of EM calorimeters and are typically much thicker.

As mentioned above, the CDF calorimeter consist of alternating absorber and detector layers. The absorber consists of a dense material (lead or iron) with a large radiation or interaction length for the purpose of instigating the showers discussed above. This material is *dead* in the sense that it has no detection capabilities. When charged particles reach energies below some critical energy they lose energy to the detector material through excitation and ionization of atoms. In some materials, called *scintillators*, a fraction of this excitation energy emerges as visible light as the excited atoms return to their ground state. Some of this light can be transferred to photomultiplier tubes which converts the light to electronic signals through the photoelectric effect. The detection layers of the CDF calorimeters are composed of various scintillating material. The energy of a particle absorbed by the calorimeter is proportional to the amount of light measured by the scintillating material, and this proportionality constant must be determined through calibration. The process of converting charged particle interactions into electronic signals described above is very fast. Information from the calorimeter is available very quickly at CDF and is used to make the first decisions on whether an event is interesting or not (*i.e.,* it is useful in the level-1 *trigger*).

The size, granularity, location, and resolution of the various CDF calorimeter modules will now be described. The central calorimeter is located outside (*i.e.,* farther away from the interaction point) of the solenoid magnet and is divided into electromagnetic (CEM) and hadronic (CHA) sections. The central calorimeter is segmented in $Y - \phi$ space into 480 *towers* which point back towards the interaction point. The granularity of the towers is $\Delta\eta \times \Delta\phi \approx 0.1 \times 0.26$. The central calorimeter covers a pseudo-rapidity range up to 1.1. The CEM is a lead-scintillator calorimeter with a depth of about 18 radiation lengths; the CHA is an iron-scintillator calorimeter with a depth of approximately 4.7 interaction lengths. The energy resolution of the CEM for electrons is

$$\frac{\sigma(E_T)}{E_T} = \frac{13.5\ \%}{\sqrt{E_T(\mathrm{GeV})}} \oplus 2\ \%, \qquad (3\text{–}7)$$

while the average energy resolution of the CHA for charged pions is

$$\frac{\sigma(E_T)}{E_T} = \frac{50\ \%}{\sqrt{E_T(\mathrm{GeV})}} \oplus 3\ \%. \tag{3–8}$$

In Run II at CDF, a new forward scintillating calorimeter replaced the original gas calorimeter. The forward region, $1.1 < |\eta| < 3.6$, is covered by the *Plug Calorimeters* which consist of lead-scintillator for the electromagnetic section (PEM) and the region, $1.3 < |\eta| < 3.6$ is covered by the iron-scintillator for the hadronic section (PHA). The positions of the various plug calorimeter modules with respect to detector η are shown in figure 3–4 [23]. The PEM and PHA are identically segmented into 480 towers of size which varies with η ($\Delta\eta \times \Delta\phi \approx 0.1 \times 0.13$ at $|\eta| < 1.8$ and increases to $\Delta\eta \times \Delta\phi \approx 0.6 \times 0.26$ at $|\eta| = 3.6$). The energy resolution of the PEM for electrons is

$$\frac{\sigma(E_T)}{E_T} = \frac{16\ \%}{\sqrt{E_T(\mathrm{GeV})}} \oplus 1\ \% \tag{3–9}$$

while the average energy resolution of the PHA for charged pions is

$$\frac{\sigma(E_T)}{E_T} = \frac{80\ \%}{\sqrt{E_T(\mathrm{GeV})}} \oplus 5\ \%. \tag{3–10}$$

The region between the central and forward calorimeters, $0.7 < |\eta| < 1.3$, is covered by an iron-scintillator hadron calorimeter (WHA) with similar segmentation to the central calorimeter. The WHA has a depth of approximately 4.5 interaction lengths, and a resolution for charged pions of

$$\frac{\sigma(E_T)}{E_T} = \frac{75\ \%}{\sqrt{E_T(\mathrm{GeV})}} \oplus 3\ \%. \tag{3–11}$$

EM calorimeters are designed to contain most EM showers. However, some EM shower energy may spill over into the hadronic calorimeter. It is also possible that hadrons may begin to interact in the EM calorimeter. In some analyses the features of the EM and hadronic showers are important for particle identification.

However, for the inclusive jet cross section, no distinction based on particle type is necessary. Only the total energy deposited within a $Y - \phi$ region of the calorimeter is important to this jet measurement. Therefore, the energy deposited in electromagnetic and hadronic sections of each tower are combined into *physics* towers. The position of each section is defined by the vector joining the interaction point to the geometrical center of the section. Each section is assumed to have no mass (*i.e.,* $E \propto P$), and the sum of the momentum four-vector for the EM and hadronic section are taken as the momentum vector of the physics towers. Physics towers are the detector objects which are clustered into jets.

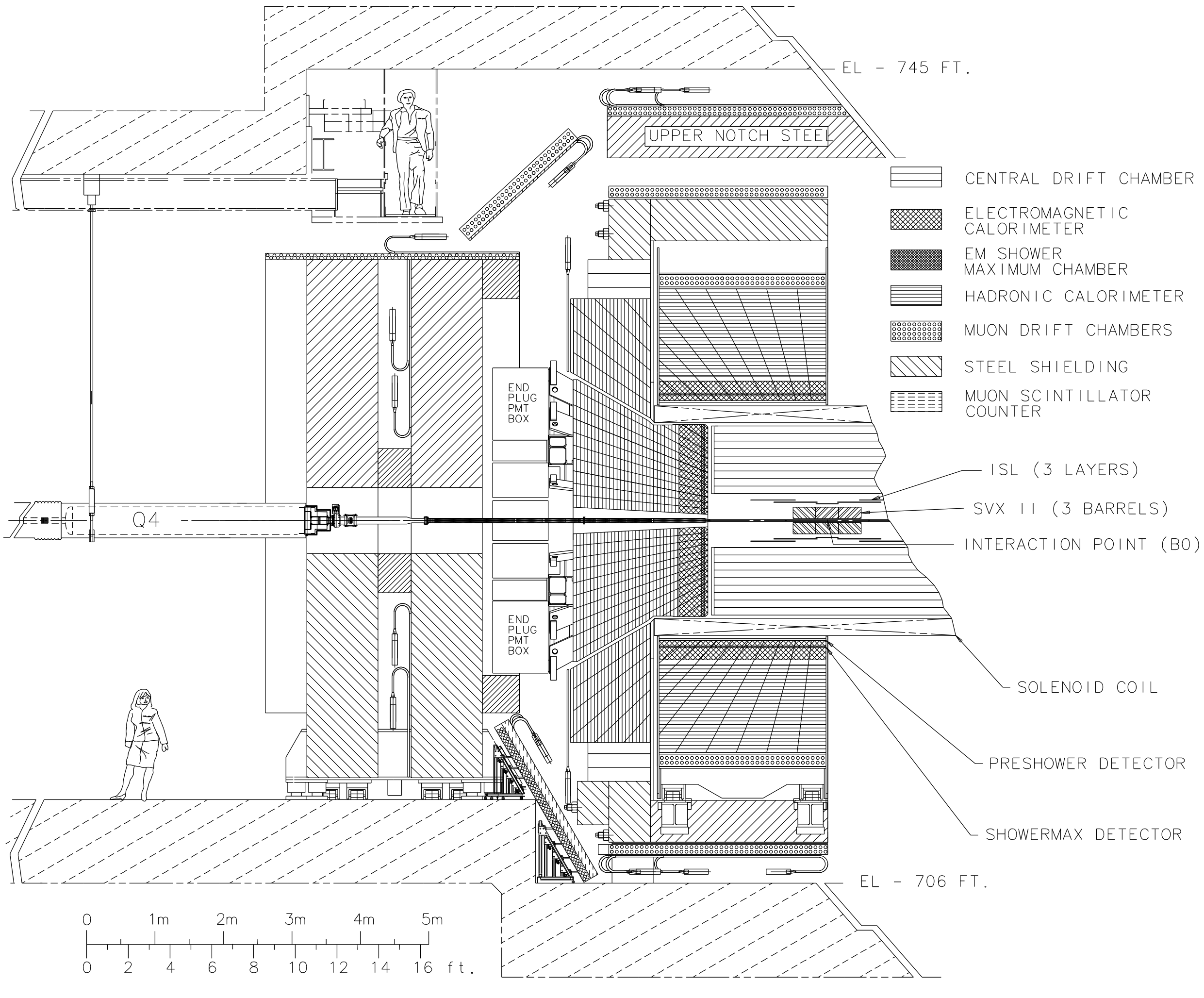


Figure 3–3. CDF detector. The CLC, COT, and the calorimeters are the important components for this analysis. They are described briefly in the text.

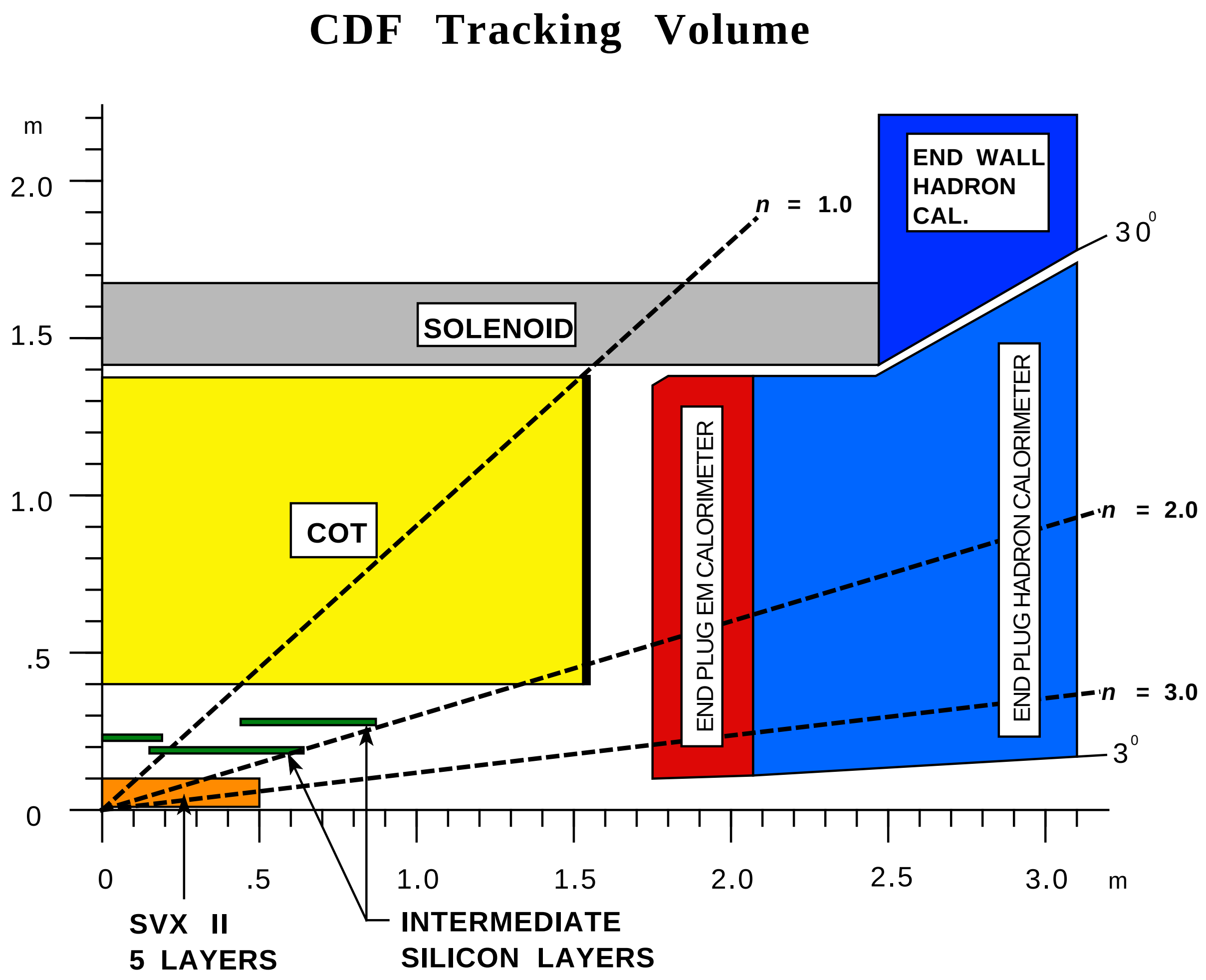


Figure 3–4. Longitudinal view of the CDF II Tracking System and plug calorimeters. The position of the calorimeter modules with respect to detector η is shown in the figure.

CHAPTER 4
JET DEFINITION

Cross sections of the hard partonic scattering processes can be calculated to a fixed order in perturbation theory. However, due to hadronization of the partons, and other aspects of the hadron collider environment (see section 2.6), it is not clear what physical observables (if any) will yield a clear interpretation of the original hard interaction. Jet clustering algorithms are designed to cluster the complex structure of final state objects from each collider event into jets. These jets must be a map to the physical properties of the partons from the hard scattering to be a useful construct for comparison with theoretical predictions. Currently at CDF there are three jet clustering algorithms in use:

- JetClu: JetClu [26] is the cone algorithm used in Run I at CDF. Cone algorithms combine objects based on relative separation in $Y - \phi$ space, $\Delta R \equiv \sqrt{\Delta Y^2 + \Delta\phi^2}$.
- Midpoint: The Midpoint algorithm is a cone algorithm similar to JetClu. It has certain advantages over JetClu and is the cone algorithm of CDF in Run II.
- K_T: The K_T algorithm combines objects based on their relative transverse momentum as well as their relative separation in $Y - \phi$ space [27].

In this chapter the Midpoint jet clustering algorithm will be described in detail. For completeness, the K_T algorithm and JetClu algorithm will also be summarized. After defining the algorithms some technical issues on the topic of jet definition will be discussed.

4.1 The CDF Midpoint Jet Clustering Algorithm

When clustering jets with a cone algorithm a cone size, R_{cone}, must be specified. The cone size determines the maximum amount of angular separation particles can have in $Y - \phi$ space and still be combined into a jet. At CDF, jets are reconstructed with three different cone sizes: 0.4, 0.7, and 1.0. What cone size is useful depends on the details of the role that jets play in an analysis. This analysis uses a cone size of $R_{cone} = 0.7$.

The first step in any jet algorithm is to identify the list of objects to be clustered. In this analysis, jets will be clustered at four different levels (see figure 2–10). The list of objects to be clustered is different in each case:

- Detector level (data or MC with detector simulation): four-vectors of the calorimeter physics towers are used as the basic elements of the clustering. To reduce the effect of electronic noise, only towers with $P_T > 100\ MeV/c$ are included in the list.

- Particle or hadron level (MC): four-vectors of the stable particles (*i.e.,* hadrons) are the basic elements to be clustered.

- Parton level (MC): four-vectors of the partons before hadronization are clustered into jets. For MC such as HERWIG or PYTHIA this will include the many quarks and gluons from the parton shower and multiple parton interactions.

- Parton level (NLO parton level): four-vectors of the partons are clustered into jets. There are at most three partons in the list at NLO.

The next step is to identify a list of *seed* objects. This is a subset of the list of objects to be clustered with the extra requirement that the P_T of the object be above some threshold (1 GeV/c). It would be preferable theoretically to include seeds corresponding to every point in $Y - \phi$ space; however, by searching for jets only at seed locations the CPU-time to run the algorithm is greatly reduced. At each seed location a cone of radius, $R = R_{cone}/2$, in $Y - \phi$ space is constructed. This reduced cone size, or *search cone*, is not a feature of the *standard* Midpoint algorithm. The standard Midpoint algorithm uses $R = R_{cone}$ for all clustering steps. CDF uses a modified version of the Midpoint algorithm which is often called the *Search Cone* algorithm. The momentum four-vectors of all objects located in the search cone are summed. This four-vector sum is called the *centroid* of the cluster. The four-vector of the centroid is then used as a new cone axis. From this axis a new cone is drawn and the process of summing up the four-vectors of all particles in the cone is conducted again. This process is iterated until the cone axis and the centroid coincide, indicating that the configuration is stable. Once the stable configuration is found, the cone axis is expanded to the full cone size ($R = R_{cone}$), and the four-vector of a *protojet* is formed by adding up all of the four-vectors of the objects in the expanded cone. The expanded cone is not iterated for stability. This procedure of finding stable cones is applied to every object in the seed list.

The next step in the algorithm is the one for which it is named. Additional seeds are added at the midpoint between all protojets whose separation in $Y - \phi$ space is less than two times R_{cone} (*i.e.,* if $\Delta R < 2R_{cone}$). A cone of radius $R = R_{cone}$ is then drawn around the midpoint seed and iterated until a stable configuration is found. If this configuration is not already in the list of protojets, it is added to the list. After all midpoint seeds have been iterated to stable cone configurations, the list of protojets is complete. The process of adding seeds at the midpoint between all stable cones reduces the sensitivity of the algorithm to soft radiation.

It is possible that many of the protojets will overlap (*i.e.,* objects may appear in more than one protojet). Overlapping protojets must be split or merged to make sure that the same object is not included in more than one jet. Before splitting and merging begins, protojets are sorted according to their P_T. If the sum-P_T (four-vector sum) of shared objects between two protojets is more than the fraction, $f_{merge} = 0.75$, of the protojet with lower P_T, then the two protojets are merged. If the sum-P_T of shared objects between two protojets is less than f_{merge}, then the shared objects between the two protojets are split and assigned to the closer cone in $Y - \phi$ space. The Midpoint algorithm may be summarized:

1. A list of all objects to be merged into jets is constructed.
2. A seed list which includes only objects with transverse momentum greater than 1 GeV/c is generated.
3. Stable cones are constructed around each seed ($R = R_{cone}/2$).
4. The radius of all stable cones is extended to $R = R_{cone}$.
5. An additional seed is added for midpoints between each pair of stable cones separated by less than twice the cone radius. Each additional seed is searched for stable cones ($R = R_{cone}$) that have not already been discovered.
6. The stable cones are P_T-ordered and splitting and merging is performed for overlapping cones.

At NLO parton level, there are at most three particles in the event. In this case, the idea of a jet algorithm becomes very simple. The algorithm must decide if the two particles which are closest in $Y - \phi$ space should be combined or not. There is no complicated splitting and merging step needed. Here, the default Midpoint algorithm would merge any two particles which are separated by less than $2 \times R_{cone}$ in $Y - \phi$ space. At the detector level it is observed that particles are almost never merged by the Midpoint algorithm if they are separated by $2 \times R_{cone}$. For this reason, R_{sep} has been introduced for NLO calculations with the Midpoint algorithm. At NLO, the algorithm is modified by R_{sep} so that particles are only merged if their separation in in $Y - \phi$ space is less than $R_{sep} \times R_{cone}$. A value of $R_{sep} = 1.3$ is consistent with detector level studies. R_{sep} plays the role of splitting and merging at the NLO parton level [26].

4.2 Other Jet Clustering Algorithms

JetClu is a cone algorithm similar to the Snowmass parton clustering algorithm [28]. The steps for jet clustering at the detector level used by JetClu are described below:

1. An E_T ordered list of *seed* towers with E_T >1.0 GeV is created .

2. Beginning with the highest E_T tower, preclusters are formed from adjacent seed towers, provided that the towers are within a 0.7x0.7 window centered at the seed tower. Any tower outside of this window is used to form a new precluster. This clustering step is dependent on detector geometry, and it cannot be conducted in the same way at the hadron or particle level.

3. The preclusters are ordered in decreasing E_T and the E_T weighted centroid is formed by adding the energy from all towers with more than 100 MeV within R=0.7 of the centroid.

4. A new centroid is calculated from the set of towers within the cone and a new cone drawn about this position. Steps 3 and 4 are iterated until the set of towers contributing to the jet is stable. The property of the JetClu algorithm, that all towers included in the original cluster remain in the cluster even when they no longer lie within the cone radius, is called *Ratcheting.*

5. Clusters are reordered in decreasing E_T and overlapping jets are merged if they share $\geq$75% of the smaller jet's energy. If they share less, the towers in the overlap region are assigned to the nearest jet.

The K_T algorithm handles particle combination much differently than the cone algorithms described so far. The procedure for combining objects into jets is exactly the same at the parton, hadron, and detector levels for the K_T algorithm and they are described below:

1. The quantities Y, ϕ, and P_T are constructed for each object in the list of objects to be combined.

2. For each object in the list $d_i = P_{T_i}^2$ is calculated, and for each pair of partons the quantity $d_{ij} = min((di)^2, (dj)^2) * (R_{ij})^2/D^2$ is defined. The D parameter plays a similar role in the K_T algorithm as R_{cone} does in the cone algorithms.

3. Find the minimum of all d_i and d_{ij}.

4. If the minimum is one of the d_{ij} remove particles i and j from the list and replace them with an object defined by the sum of their momentum four-vectors.

5. If the minimum is one of the d_i then remove it from the list to be checked for merging and add it to the list of jets.

6. If any particles remain in the list, go to second step above.

4.3 Jet Definition Issues

The JetClu algorithm used in Run I at CDF has several flaws. One major problem with JetClu is that the first step of the algorithm, when applied at the detector level, is to "form preclusters from adjacent seed towers". This step is dependent on the detector granularity. This type of dependence makes it impossible to define an equivalent algorithm at the parton and hadron level. The Midpoint algorithm clusters towers based on their separation in $Y - \phi$ space and is therefor less sensitive to the detector granularity.

As mentioned above, it is preferable to place seeds at every point in $Y - \phi$ space. This is because the use of seeds adds an infrared sensitivity. Fluctuations in tower energy due to soft radiation can push the energy just above or below the seed threshold. In this way the jets clustered in a given event may depend on soft radiation. This sensitivity can be minimized if the seed threshold is low, and only high P_T jets are studied. If high energy jets always have at least one physics tower with P_T far above the seed threshold, then recognizing the seed is not sensitive to fluctuations due to soft radiation. The sensitivity introduced to this measurement is not significant, since only jets with P_T greater than 54 GeV/c are included.

The CDF Midpoint jet clustering algorithm is less sensitive to soft radiation than JetClu [26]. The reason this is true is illustrated in figure 4–1 [29]. In the configuration on the left, two seeds may have lead to two stable cone configurations. If a seed had been constructed between these two cones, a different configuration may have been found. A stable cone may have been found which included the other two objects so that only one jet was constructed in the event. Soft radiation can push objects just below or just above seed threshold. In this way, the event topology is sensitive to soft radiation. The Midpoint algorithm reduces this sensitivity by adding an extra seed at the midpoint between all stable cones. The result of the midpoint seed is that the event topology is not as sensitive to fluctuations above or below seed threshold caused by soft radiation. For this reason, and the reduced dependence on detector granularity, the Midpoint algorithm has replaced JetClu as the cone clustering algorithm used at CDF in Run II.

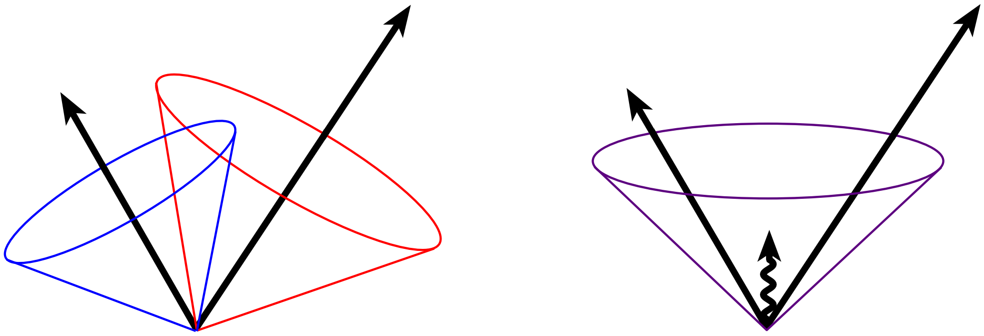


Figure 4–1. The Midpoint jet clustering algorithm checks for a stable cone configuration at the midpoint of all stable cones located from searching at seed locations. In this way the algorithm is less sensitive to fluctuations due to soft radiation.

At CDF, it was observed that in some events, after clustering jets with the standard Midpoint algorithm, there were significant clusters of energy which were not included in any jet. An event display of an event in which this occurs is shown in figure 4–2 [26]. The figure shows the energy deposited in each tower of the detector on the $\eta - \phi$ plane. The clusters of energy shown in black were not clustered into any jet in this event. These clusters of energy which are not included in a jet are referred to as *dark-towers*. Dark towers occur because a cone that started from a seed within the dark-tower cluster moves away from this cluster of energy towards a larger cluster of energy (*i.e.,* the cone migrates to a neighboring cluster of energy).

The effect of dark towers is significant. For example, approximately two percent of events with a 400 GeV/c jet have more than 50 GeV/c of un-clustered transverse momentum. Because this dark-tower effect is not included at NLO, it was decided that the issue of un-clustered energy needed to be addressed. The

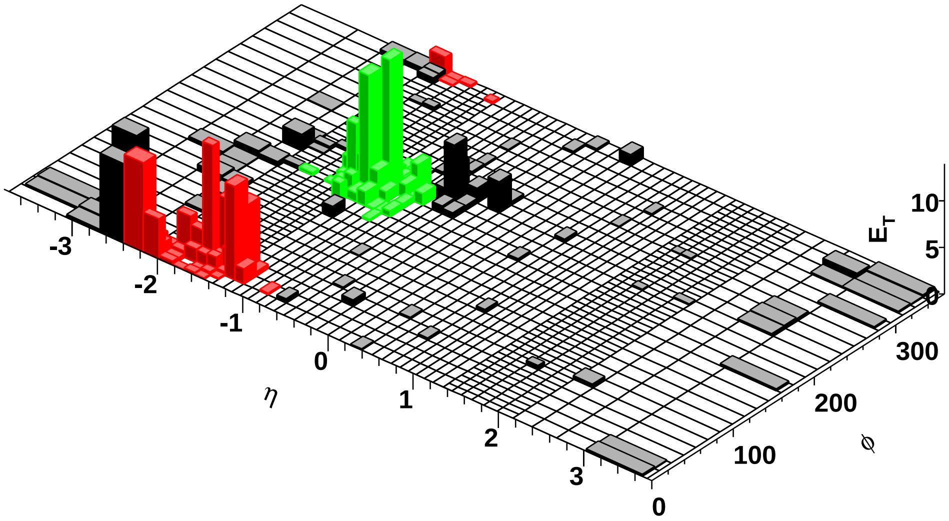


Figure 4–2. Dark towers observed by the original Midpoint algorithm. Not all clusters of energy in the detector were being included in a jet.

solution applied at CDF is to use a smaller initial search cone ($R_{cone}/2$) (*i.e.,* use the Search Cone algorithm). The net result of using the Search Cone algorithm is a five percent increase in the inclusive jet cross section, which is roughly independent of jet P_T [30].

The CDF Search Cone algorithm applied in this analysis is not perfect. It is slightly more sensitive to tower and seed thresholds than the standard Midpoint algorithm. Also, there is a sensitivity to the size of the the search cone. Varying the search cone radius from $R/2$ to $R/\sqrt{2}$ leads to a variation in the cross section by less than 2 percent. When comparing to NLO predictions the imperfections with the Search Cone algorithm are minor when compared with the five percent shift in the cross section caused by the dark towers. However, the Search Cone algorithm has serious theoretical issues if applied to NNLO predictions. Finding a stable configuration with the reduced cone size, and then expanding the cone without further iteration yields an infrared sensitivity at NNLO [30].

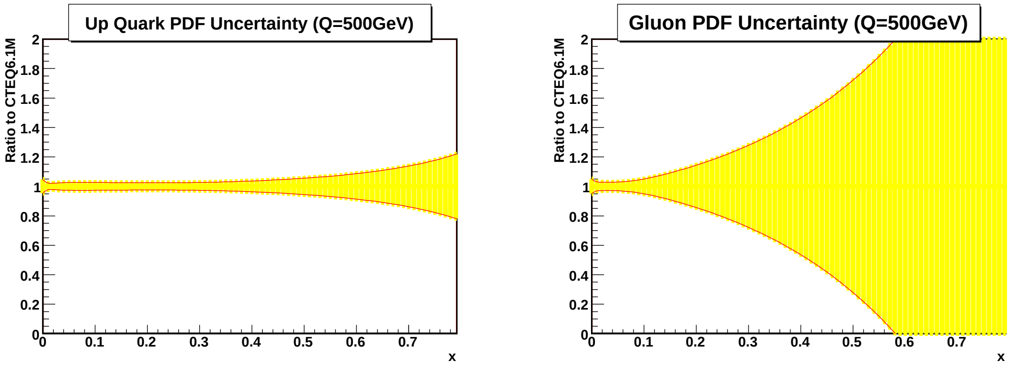
Although the Search Cone is not perfect, and cone algorithms in general have more issues theoretically than the K_T algorithm, cone algorithms have some advantages. With a standard cone algorithm the user has complete control over what is included in the cone through the variable R_{cone}. The fact that the jet is composed of a specified cone size in the detector is a useful property for making corrections. For example, when correcting for multiple $p\bar{p}$ interactions in the same bunch crossing a cone located randomly in the detector can be used to study the extra energy that is included on average in a jet due to the extra interaction. This cannot be done with a K_T algorithm since it does not use a fixed cone size. It is also useful to make measurements with two different types of algorithms. The difference in results from different algorithms should be predictable by MC and can be used to learn about the jet clustering properties of each algorithm.

CHAPTER 5
INCLUSIVE JET MEASUREMENTS

The prediction in the early 1970's, that jets of hadrons whose momentum when summed up would be equal to the momentum of the initially scattered partons, has led to a rich history of theoretical predictions and jet measurements at hadron colliders [31, 32, 33]. The first inclusive jet cross section measurement at a hadron collider, as well as the first direct observation of a clear two-jet event topology, came from the $Sp\bar{p}S$ collider at CERN [34, 35] with a center-of-mass energy of 540 GeV ($\sqrt{s} = 540\ GeV$). A detailed history of the evolution of theoretical predictions and experimental measurements of the inclusive jet cross section can be found in reference [36].

The measurement of the differential inclusive jet cross section at CDF reaches the highest momentum transfers ever studied in collider experiments. Thus, it is potentially sensitive to physics beyond the standard model [37]. Studying the highest energy events at the Tevatron is equivalent to probing distances on the order of $10^{-19}\ m$. This measurement is probing distance scales more than one thousand times smaller than the radius of the proton and is sensitive to whether or not the quark has substructure [38]. The inclusive jet cross section measurement is also a fundamental test of predictions of perturbative QCD [39, 40] over eight orders of magnitude in cross section. Jet events in the central region of the detector with transverse momentum higher than 530 GeV/c have a cross section of approximately 30 fb ($30 \times 10^{-6}\ nb$). These are among the smallest cross sections ever measured at a collider experiment. The current measurement spans 600 GeV/c in jet transverse momentum, and it can therefore be used to observe the running of α_s in a single measurement. It is a common misconception that the QCD force gets small at small distances. This is certainly not the case. The effective coupling does decrease at high momentum transfer, but the effective force is inversely proportional to square of the distance($F \propto \frac{\alpha_s}{r^2}$). In fact, the forces occurring in high energy dijet events are of among the largest ever observed in a laboratory environment.

Perhaps the most useful aspect of these measurements, is that they can be used to constrain the proton PDF, which in turn improves the theoretical predictions in all physics channels for experiments at the Tevatron and the future experiments at the LHC. This is important because the probability of a gluon carrying a large fraction of the momentum of the proton (i.e. large Bjorken-x or high-x) is not well known [41]. In fact, the gluon PDF is the dominant source of theoretical uncertainty in the inclusive jet cross section and many other processes at hadron colliders. The uncertainty on the quark and gluon PDFs for $Q = 500\ GeV$ is shown as a function of Bjorken-x in figure 5–1.



(a) (b)

Figure 5–1. Uncertainty on the up quark , 5–1(a), and gluon , 5–1(b), PDFs for $Q = 500\ GeV$ as a function of Bjorken-x (momentum fraction carried by the parton). The uncertainty is much larger on the gluon PDF because this component of the proton is not probed directly in DIS experiment.

The strongest constraints on the parton density functions come from deep inelastic scattering experiments (DIS). The DIS experiments scatter electrons off of protons. This process is depicted in figure 5–2 [9]. The electron interacts with the quarks in the proton through the exchange of a virtual photon. Because photons do not couple directly to gluons, the DIS experiments can not make strong constraints on the gluon content of the proton. The strongest constraints on the gluon PDF come from jet measurements at hadron colliders, such as the inclusive jet cross section measurement discussed here.

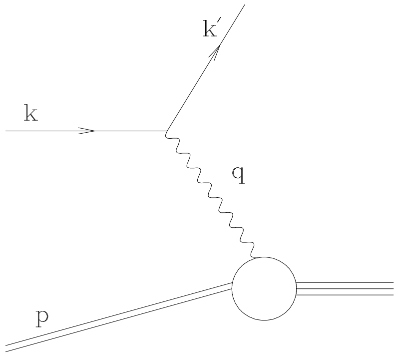


Figure 5–2. Dominant process in deep inelastic scattering experiments. The virtual photon probes the quark content of the proton. Since photons do not couple directly to gluons, this process can not be used to place strong constraints on the gluon PDF.

Because the inclusive jet measurement in the forward region (large Y) probes a kinematic range which is not expected to be sensitive to new physics, it should lead to a powerful constraint on the gluon PDF. Dijet events produce high P_T jets at high rapidity when the momentum fraction of the two incoming partons do not balance. In this topology, the high-x component of the proton is probed at a lower momentum transfer than for an equivalent energy jet in the central region. In other words, the high-x component of the proton can be probed at a lower energy scale with forward jets. Because lower energy scales have been studied extensively in the central region, and agree with the predictions of pQCD, one can be confident that physics beyond the SM is not affecting the result. Since the forward region jets is not as sensitive to new physics as the high P_T jets in the central region, measurements in the forward region are crucial to confirming that observations can be attributed to PDF effects and not physics beyond the standard model.

The CDF experiment has a history of making important inclusive jet measurements. In Run I ($\sqrt{s} = 1.8\ TeV$), CDF made several measurements of the inclusive jet cross section in the central region of the detector ($0.1 < |\eta| < 0.7$) using the JetClu cone jet clustering algorithm. In Run IA, with 19.6 pb^{-1} of data collected during the period 1991-1993, an excess of data over theoretical predictions was observed at high jet transverse momentum [42]. This result alluded at the need for larger gluon content at high-x in the proton. The Run IB CDF central region result, with 87 pb^{-1} of data collected during the period 1994-1995 [36], still showed an excess at high jet P_T. When these results were combined with the D0 result including the higher rapidity region [43], with 95 pb^{-1} of data up to $\eta = 3.0$, it was confirmed that the gluon content at high-x had been underestimated in the proton structure functions. This experience from Run I revealed a need to be able to quantify the effect of PDF uncertainty on collider observables such as cross section [44, 45].

The increased center-of-mass energy in Run II ($\sqrt{s} = 1.96\ TeV$) significantly increases the jet cross section at high jet P_T. Quantitatively, the cross section for dijet production is approximately five times higher for jet P_T of approximately 600 GeV/c. A new inclusive jet cross section measurement by CDF (with 385 pb^{-1} of data collected during the period 2001-2004), in the central region ($0.1 < |Y| < 0.7$), using the Midpoint jet algorithm in Run II, has been accepted for publication recently [46]. CDF also recently published an inclusive jet measurement in the central region for jets clustered by the k_T algorithm [27].

In this document, the Run I cone jet analysis is updated with approximately ten times the integrated luminosity resulting in over 1 fb^{-1} of data. This is the first inclusive jet measurement at CDF to include the forward region ($|Y| < 2.1$) using a cone algorithm for jet clustering. The techniques applied in this measurement will be fully motivated and briefly described. For complete details and all relevant distributions please see reference [47].

CHAPTER 6
DATA SAMPLE AND EVENT SELECTION

This analysis includes data taken from Summer 2001 until November 2005 and corresponds to an integrated luminosity of 1.04 fb^{-1}. The jet data used in this analysis were collected using four different paths of the CDF three-level trigger system [48]. The rate at which collisions occur within the CDF detector is much higher than the rate at which data can be collected and stored. Bunch crossings occur at CDF at a rate of approximately 1.7 MHz, while data can only be written to tape at about 75 Hz. Various triggers are designed to extract events that are useful for physics analysis at CDF. The trigger is split into three levels, and at each level the events which pass the trigger requirements are passed on to the next level. The trigger requirements at each level must produce a rate reduction large enough for processing at the next level to be possible.

The Level 1 jet trigger consists of two trigger streams; requiring a calorimeter trigger tower to have $E_T > 5\ GeV$ for the jet20 and jet50 triggers, and $E_T > 10\ GeV$ for the jet70 and jet100 triggers. At Level 2, the calorimeter towers are clustered using a nearest neighbor algorithm. Four trigger paths with cluster $E_T > 15$, 40, 60, and 90 GeV are used. Events in these paths are required to pass jet $E_T > 20$, 50, 70, and 100 GeV thresholds at Level 3, where the clustering is performed using the JetClu cone algorithm with a cone radius $R_{cone} = 0.7$. The CDF trigger, and all trigger studies required for this analysis are described in more detail in Appendix A.

The selection criteria for the inclusive jet cross section is minimal, since all jet events are included in the analysis. Since cosmic ray backgrounds originate from outside the detector, the transverse energy deposited in the detector is not balanced. Cosmic rays can therefore be efficiently removed by applying a missing E_T significance requirement ($\widetilde{\not{E}_T}$). The $\widetilde{\not{E}_T}$ selection criteria varies according to the jet sample and it is 4, 5, 5, and 6 GeV$^{1/2}$ for jet20, 50, 70, and 100 triggers respectively. With any selection criteria it is possible that some jet events which came from hard scattering events are removed. The percentage of real jet events which survive the selection criteria is called the *efficiency.* The $\widetilde{\not{E}_T}$ requirement is approximately 100% efficient at low P_T and varies to approximately 90% at the highest jet P_T included in the measurement. In order to ensure that particles from the $p\bar{p}$ interactions are in a region of the detector with good tracking coverage, primary vertices are required to be within 60 *cm* of the center of the detector ($|Z| < 60\ cm$). The efficiency of this requirement is measured to be 95.8% at CDF [49].

The jet20, 50, and 70 triggers are prescaled to avoid saturating the bandwidth of the trigger and data acquisition system. Prescaling by a factor of n means that only 1 out of n events satisfying the trigger requirement are stored to tape. The jet70 trigger is prescaled by a constant factor of 8 for all data used in this analysis. The prescales for the jet20 and 50 triggers have changed during the data taking period considered. The *effective* prescales of the jet20 and 50 triggers for all the data were found to be 776.8 and 33.6, respectively, by luminosity-weighting the inverse of prescale factors

$$\frac{1}{P_{\text{effective}}} = \frac{1}{\mathcal{L}_{\text{total}}} \cdot \sum_i \left(\frac{\mathcal{L}_i}{P_i}\right), \tag{6–1}$$

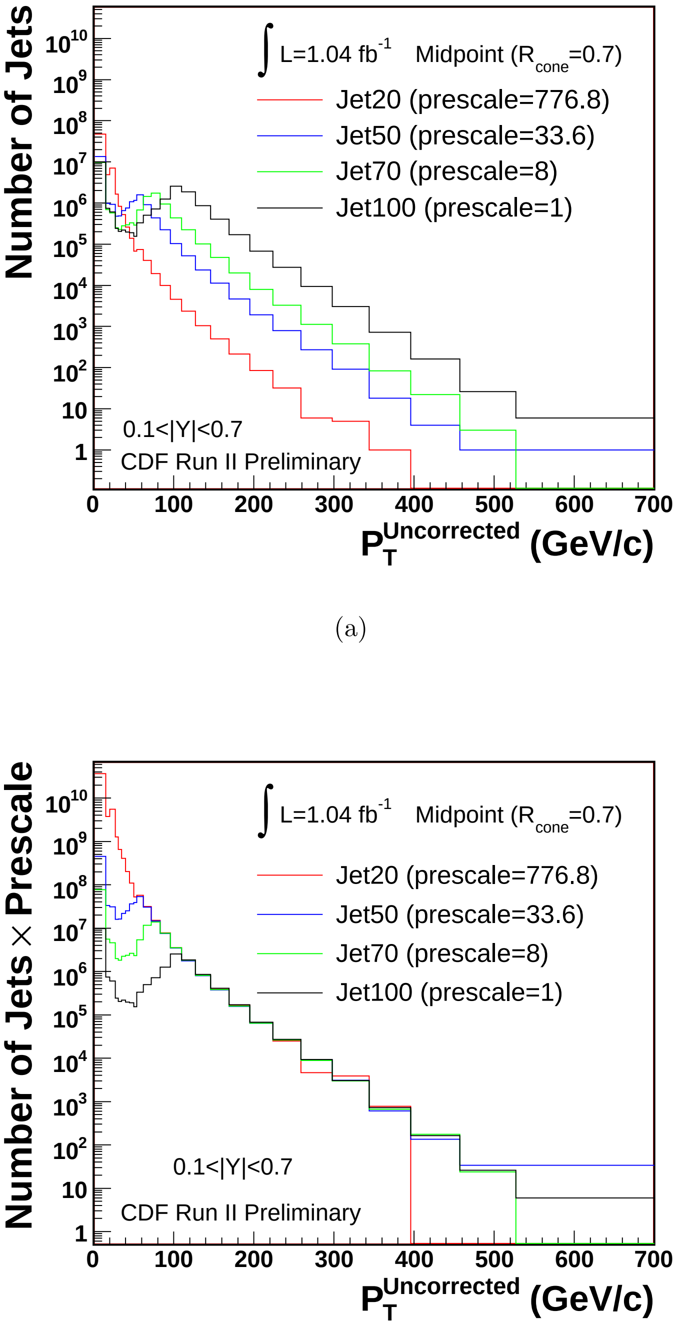
where $P_{\text{effective}}$ is the effective prescale, $\mathcal{L}_{\text{total}}$ is the total integrated luminosity, and $\mathcal{L}_i$ is the integrated luminosity of a period when a prescale factor P_i is used. As a cross-check of the jet20 and jet50 effective prescales, the cross section ratios (before prescale correction) of jet70 to jet20 triggers and of jet70 to jet50 triggers in the jet P_T region where jet70 is efficient were studied. The results were found to be consistent to better than 1% between the two methods. The distributions for the number (*i.e., jet yield*) of jets as function of P_T in the central region $(0.1 < |Y| < 0.7)$ before and after correcting for trigger prescales are shown in figure 6–1.

The inclusive differential jet cross section is determined as follows:

$$\frac{d^2\sigma}{dP_T dY} = \frac{1}{\Delta Y}\frac{1}{\int \mathcal{L}dt}\frac{N_{jet}/\epsilon}{\Delta P_T}, \tag{6–2}$$

where N_{jet} is the number of jets in the P_T range ΔP_T, ϵ is the trigger, $\widetilde{\not{E}_T}$ cut, and vertex cut efficiency, $\int \mathcal{L}dt$ is the effective integrated luminosity which is corrected for trigger prescales, and ΔY is the rapidity interval.

Because the jet triggers do not use the Midpoint algorithm they are not 100% efficient near the trigger threshold. A trigger efficiency greater than 99.5 % is required to include the jets collected by a given jet trigger. A complete study of the trigger efficiency (*i.e., turn-on* curves) in all rapidity regions and the study of the effective prescales is included in Appendix A. Figure 6–2 shows the uncorrected (*i.e., raw*) inclusive differential jet cross section for the five rapidity regions: $|Y| < 0.1$, $0.1 < |Y| < 0.7$, $0.7 < |Y| < 1.1$, $1.1 < |Y| < 1.6$, and $1.6 < |Y| < 2.1$. This rapidity binning is roughly based on detector segmentation. The region $|Y| < 0.1$ corresponds to central crack region where the detector modules meet, $0.1 < |Y| < 0.7$ corresponds to the well understood central region of the detector, $0.7 < |Y| < 1.1$ and $1.1 < |Y| < 1.6$ correspond to the region where central calorimeter and the plug calorimeter are connected by the WHA, and $1.6 < |Y| < 2.1$ corresponds to the plug (forward) region of the calorimeter.



(a)

(b)

Figure 6–1. Jet yield distributions as a function of P_T in the central region before, 6–1(a), and after, 6–1(b), correcting for trigger prescales.

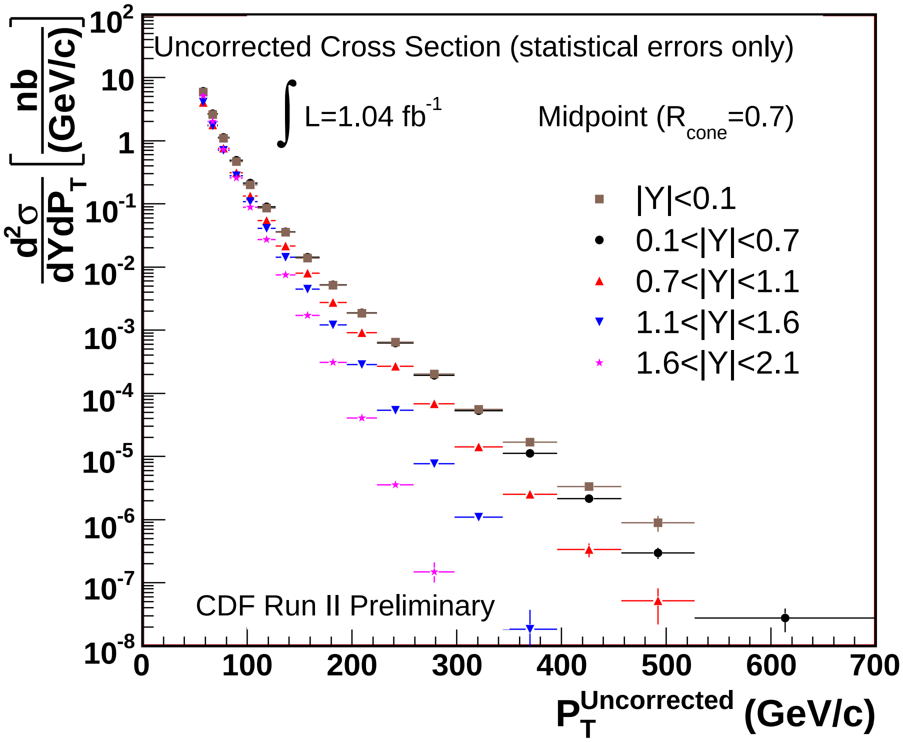


Figure 6–2. Measured raw jet cross section for the five rapidity regions. The raw jet cross section has not been corrected to remove detector effects.

CHAPTER 7
JET CORRECTIONS

The jet energy measured by the calorimeters must be corrected for detector effects, such as calorimeter non-linearity and energy smearing, before comparing experimental measurements with theoretical predictions. In addition to detector effects, corrections must also be made for some physics effects such as pileup and the underlying event before the measurement may be compared with NLO parton level perturbative predictions.

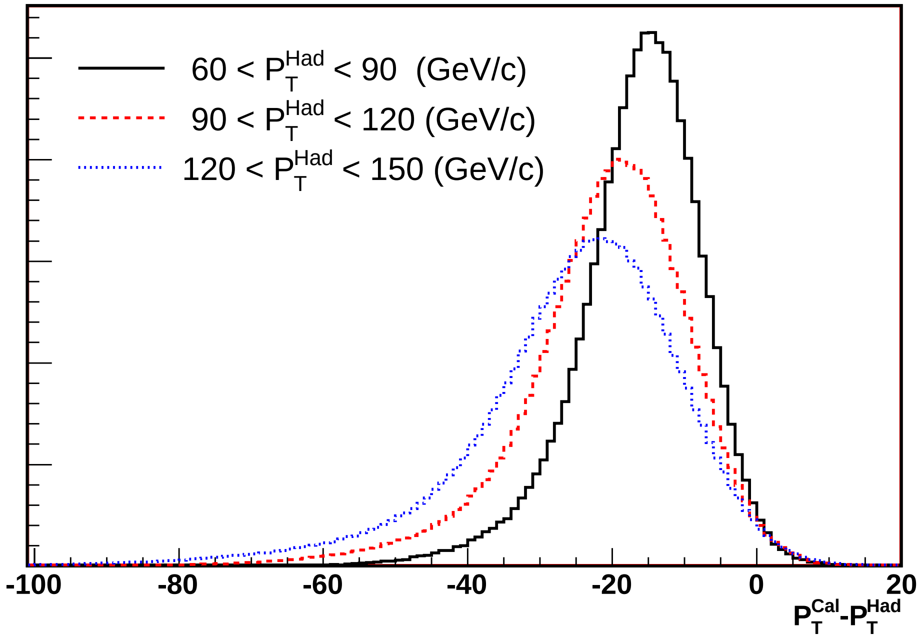


Figure 7–1. Difference between calorimeter jet P_T and hadron level jet P_T for three different jet P_T ranges. The calorimeter jet P_T is systematically lower than the hadron level jet P_T. Also, there is a smearing effect due to the fluctuation in the energy measured in the calorimeter for a given hadron level jet.

Two detector effects, which the jet data must be corrected for, are illustrated in Figure 7–1. The difference between calorimeter jet transverse momentum (P_T^{Cal}) and hadron level jet transverse momentum (P_T^{Had}) for three different jet P_T ranges is shown in the figure. Hadron level jets were matched to calorimeter jets by their separation in $Y - \phi$ space ($\Delta R < 0.7$) using MC generator results which were passed through the CDF detector simulation. The distributions peak below zero because P_T^{Cal} is systematically lower than P_T^{Had}. This reflects the non-compensating nature of the sampling calorimeters. The hadronic calorimeter was calibrated based on charged pions from a test beam with a transverse momentum of 57 GeV/c. Only pions which did not interact with the EM calorimeter were included in the calibration. In real jets however, a large fraction of hadrons do interact with the EM calorimeter. Because the EM calorimeter is calibrated based on electrons its response to hadrons is lower. This lowers the overall response to single hadrons. This effect is larger for low transverse momentum particles because they interact in the EM calorimeter more often, and it can therefore contribute to the non-linearity of the calorimeter response to hadrons. Hadronic showers have a larger fraction of neutral pions when the incident hadron has a higher transverse momentum. Because the calorimeter has a higher response to EM showers (*i.e.,* π^0 decays), this also contributes to the nonlinear nature of the calorimeter response to charged hadrons. In general the calorimeter response goes up as transverse momentum of the incident hadron increases and is not linear. This causes a systematic shift down in the energy response to jets because they include multiple hadrons with lower transverse momentum, rather than one hadron with the full jet P_T.

The jet energy smearing effect is caused by the limited jet energy resolution of the calorimeters, and is reflected in figure 7–1 by the width of the peak. Fluctuations in shower development due to the probabilistic nature of the interactions between the particles in the jet and the detector material cause the detector response to particle jets with a fixed energy to vary.

Figure 7–2 illustrates the flow of the jet correction scheme used to obtain results corrected to the hadron or parton level. First, an η-dependent relative correction is applied to the data and MC in order to equalize the response of the CDF calorimeters to jets in Y. The equalized jet P_T is then corrected for the pileup effect. Then, the absolute correction is applied to correct on average for the hadron energy that is not measured by the calorimeter. After that, the hadron and calorimeter level jet P_T distributions are compared in Monte Carlo to derive a bin-by-bin correction in order to remove resolution effects. This is called *unfolding*. At this point, the data have been corrected to the hadron level. In order to compare directly with pQCD predictions, the effects of the underlying event and hadronization need to be removed from the data. After this final correction, the data have been corrected to the parton level. The Monte Carlo simulation used to derive the corrections, and the details of each correction step will now me described.

7.1 Monte Carlo Simulation

The parton shower MC programs PYTHIA 6.2 [50] and HERWIG 6.4 [51], along with the CDF detector simulation, are used to derive the various corrections which are applied to the data, and to estimate systematic uncertainties. Structure functions (*i.e.,* parton distribution functions) for the proton and anti-proton are taken from CTEQ5L [52]. The CDF detector simulation is based on GEANT3 [53] in which a parametrized shower simulation, GFLASH [54], is used to simulate the energy deposited in the calorimeter [55].

The GFLASH parameters are tuned to test-beam data for electrons and high momentum charged pions and to the *in-situ* collision data for electrons from Z decays and low momentum charged hadrons. However, the CDF simulation does

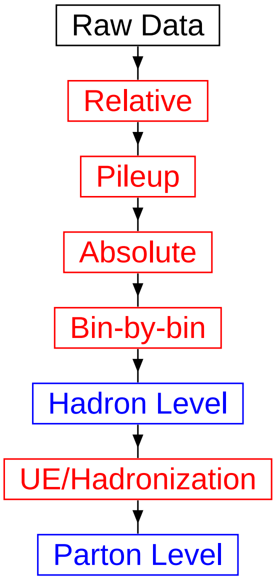


Figure 7–2. Flow diagram for the jet corrections used in the inclusive jet analysis. Correction steps are shown in red, while the hadron level and parton level corrected states are shown in blue. Relative, pileup and average (absolute) corrections are applied directly to the jet P_T before binning.

not describe energy deposition in the calorimeters perfectly, especially in the regions corresponding to the plug calorimeters and cracks between calorimeter modules. Since the MC simulation is used to derive various jet corrections to be made on the data, differences between the real calorimeter response to jets and the calorimeter simulation need to be well understood. Differences in the relative jet energy response and jet energy resolution between the collision data and MC simulation events were investigated using dijet P_T balancing in dijet events [55] and the *bisector* method [56], respectively. The dijet P_T balance and bisector studies are briefly described below. More details can be found in Appendix C.

Comparisons of dijet P_T balance between data and MC reveal that the relative jet energy scale versus η is different between data and MC, and that the difference depends on jet P_T at high rapidity ($|Y| > 1.1$). For example, the jet energy scale in the plug calorimeter region is higher in MC than in data by $\sim$ 2% and the difference increases slightly with jet P_T. This difference is accounted for by the relative corrections which are described in detail in section 7.2.

The bisector method allows one to compare the energy resolution of the CDF detector and the CDF detector simulation. In the central region ($0.1 < |Y| < 0.7$) the detector simulation reproduces the detector jet energy resolution well. In the other rapidity regions, small differences were found between data and MC. To account for these differences, modifications to the unfolding factors were derived and can be found in Appendix C. The corrections are less than 6% in most bins, and less than 10% in the most extreme cases.

7.2 Relative Correction

The calorimeter response to jets is not flat in detector-η (η_d). The non-uniformity in η_d arises from cracks between calorimeter modules and also from the different energy responses of the central and plug calorimeters. The leading two jets in dijet events are expected to be nearly balanced in P_T in absence of QCD radiation. This dijet P_T balance provides a useful tool to study the jet energy response as a function of η_d, and to derive the relative correction. To study the dijet balance, a dijet event topology is required. The CDF calorimeter response to jets is well understood and almost flat in η_d in the central region. For this reason, a jet with $0.2 < |\eta_d| < 0.6$ is required in the event, and this jet is defined as the *trigger* jet. The other jet in the event is defined as a *probe* jet. Figure 7–3 [55] shows a measure of the dijet balance, $\beta \equiv P_T^{probe}/P_T^{trig}$, observed in the CDF calorimeter and in detector simulation, as a function of detector-η of the probe jet. If the response of the calorimeter was the same in all regions of the detector, this distribution would be approximately flat and very close to one. The relative correction is applied to equalize the response in η of the CDF calorimeter to jets.

To determine the η_d-dependent relative jet energy correction, the P_T balance of the probe and trigger jet is studied as a function of the probe jet η [55]. The

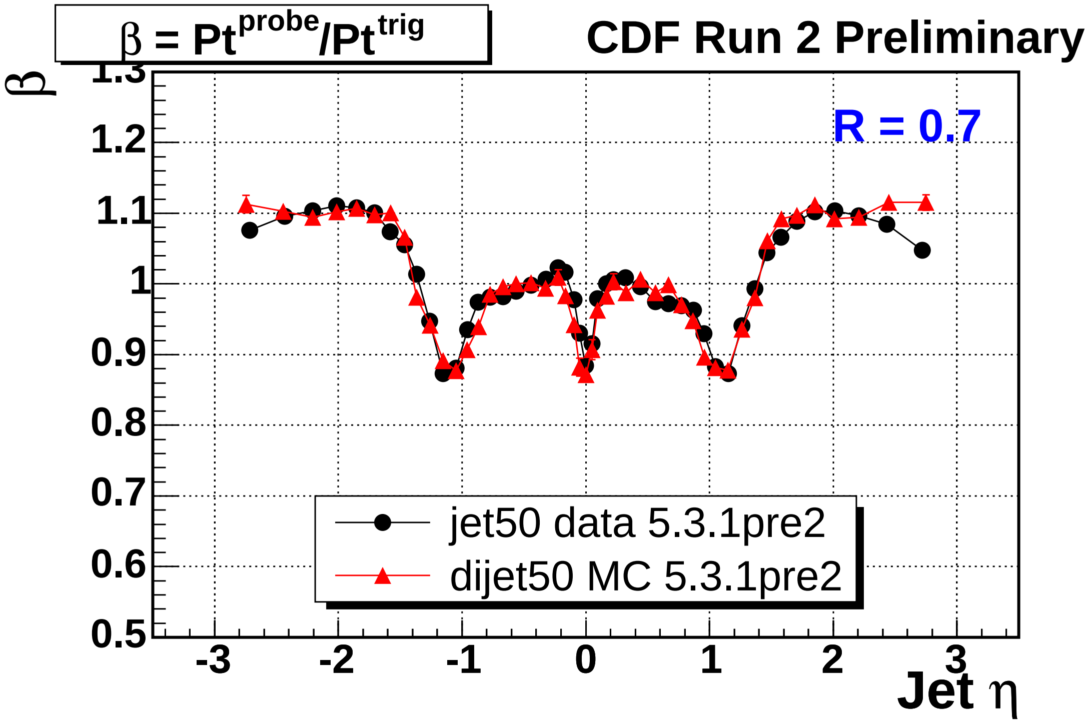


Figure 7–3. Degree of dijet balance (β) observed in the CDF calorimeter and and MC with CDF detector simulation as a function of detector η for the probe jet in the event. The relative correction is applied to equalize the response in η of the CDF calorimeter to jets.

η_d-dependent relative corrections are obtained by making a fit to the β distribution at a fixed jet transverse momentum. Since the relative jet energy response is different between data and MC (see figure 7–3), corrections are derived separately for data and MC. The relative corrections for this analysis were derived at a fixed value ($P_T = 117.5 GeV/C$) of jet P_T so that the difference in the P_T dependence of the response observed in data and MC simulation could be handled more directly. More details regarding the relative jet energy corrections can be found in Appendix B.

As mentioned earlier, the data-MC difference in the relative jet energy scale depends on jet P_T for $|Y| > 1.1$. Therefore, a P_T-dependent correction derived for the two corresponding rapidity regions is also applied to the MC in order to force the β distributions to agree with data for all jet P_T. Requiring a dijet event topology for high P_T jets in the forward region greatly reduces statistics. The lack of statistics in data and simulated MC lead to a significant uncertainty in this P_T-dependent correction. Details of the P_T-dependent correction to the dijet P_T balance in the MC are included in Appendix C.

7.3 Pileup Correction

Hard scattering events with additional $p\bar{p}$ interactions in the same bunch crossing produce additional particles which can contribute to the jet energy in an event. The number of reconstructed vertices is a good estimator of the number of inelastic collisions occurring in a bunch crossing. The correction for the additional $p\bar{p}$ interactions is derived by studying the P_T measured in a randomly chosen cone in $Y - \phi$ space as a function of the number of vertices reconstructed in *minimum bias* events. The minimum bias trigger only requires that there be coincidence in the CLC on both sides of the detector. This trigger criteria is effectively only requiring that there was an inelastic scattering within the bunch crossing. The P_T in the randomly located cone scales linearly with the number of additional vertices in the event. The pileup correction is derived from the slope of this line, and estimates the average amount of transverse momentum to be subtracted from the jet P_T per additional vertex reconstructed in the event. The pileup correction is approximately 1 GeV/c for each additional vertex in the event [55, 57].

7.4 Absolute Correction

As hadrons pass through the CDF calorimeter, all of their energy is not collected. This effect is mostly due to the non-compensating nature of the calorimeters [discussed above]. The absolute correction corrects the jet for the average energy loss and is derived by comparing hadron level and calorimeter level jets using PYTHIA and the CDF detector simulation. Hadron level and calorimeter level jets are matched by their position in the $Y - \phi$ space ($\Delta R \leq 0.7$). The average hadron level jet energy is then studied as a function of the calorimeter jet energy in each rapidity region. This distribution is fit to a fourth order polynomial and the fit is applied as a correction to the P_T of each jet in the data sample. The correction is of the order of 20% for $P_T^{Cal} \sim 50\ GeV/c$ and decreases to the order of a few percent for $P_T^{Cal} \sim 600\ GeV/c$. This correction is slightly different for each rapidity region.

7.5 Unfolding Correction

The next step in correcting the jet P_T distribution to the hadron level is the unfolding correction. It removes smearing effects due to the finite energy resolution of calorimeter and accounts for the efficiency of the event selection cuts. After the absolute correction has been applied, P_T distributions from PYTHIA MC at hadron level and calorimeter level are compared on a bin-by-bin basis to derive the multiplicative correction. The selection criteria, Z vertex position and $\widetilde{\not{E}_T}$ cuts, are applied at the calorimeter level, but not at the hadron level. These unfolding factors are slightly different for each rapidity bin; they vary from roughly 1.4 at low jet P_T to just below 2.5 at high jet P_T.

After this final correction is applied to the data, the measurement has been corrected to the hadron level. It is now possible to compare the data directly with MC at the hadron level. In all rapidity regions, the PYTHIA transverse momentum distributions fall off slightly faster than the data at high P_T. This is due to the fact that the PYTHIA samples were generated with the CTEQ5L PDF, which do not include the enhanced high-x gluon distribution that was required to fit the Run I inclusive jet data [36, 43]. If PYTHIA is used to correct the data back to the hadron level, the shapes of the predicted P_T distributions should be the same as the data in order to avoid introducing any bias with the corrections.

The ratios of data corrected to the hadron level to the PYTHIA predictions are fit to polynomials. These fits are used to re-weight the PYTHIA P_T distributions; thereby, forcing the shapes of the P_T distributions of the MC to agree with the data. The unfolding correction factors obtained from the weighted PYTHIA distributions are applied to the data. The modification due to the re-weighting of PYTHIA is less than 1% in all rapidity regions except for the region $0.7 < |Y| < 1.1$. In this region at high P_T the correction is still less than 6%.

7.6 Hadron to Parton Correction

Before hadron level results can be directly compared to theoretical predictions of pQCD at the parton level, the effects of the UE and fragmentation must be removed. The hadron to parton level correction is obtained from comparing PYTHIA-Tune A [58] MC results at the hadron and the parton level to derive a bin-by-bin correction.

Tune A was tuned to fit the underlying event observables measured at CDF in Run I [1] . It is used for all PYTHIA calculations mentioned in this text, but is especially important for the UE correction [15, 18]. Multiple parton interactions are turned off in PYTHIA to generate the parton level distributions. The hadron to parton level correction for the central region is shown in figure 7–4. The corrections for the other regions are similar in magnitude.

---

[1] PYTHIA-Tune A implies that the following parameters are set in PYTHIA (CTEQ5L): PARP(67)=4, MSTP(82)=4, PARP(82)=2, PARP(84)=0.4, PARP(85)=0.9, PARP(86)=0.95, PARP(89)=1800, PARP(90)=0.25.

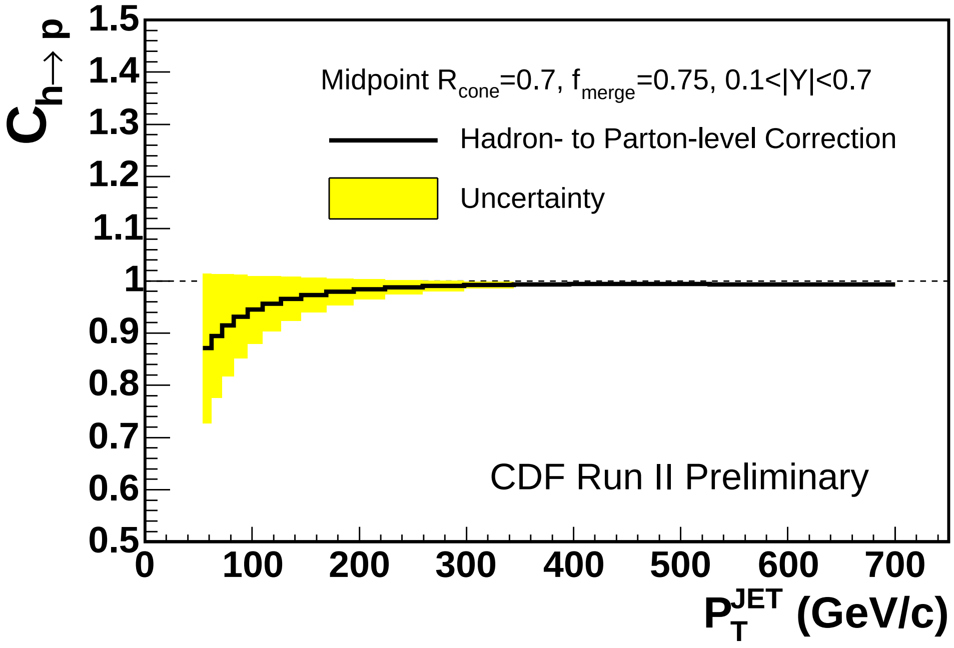


Figure 7–4. Hadron to parton level correction applied in the central region. The difference between HERWIG and PYTHIA predictions for this correction is conservatively taken as the systematic uncertainty (shaded bands).

CHAPTER 8
SYSTEMATIC UNCERTAINTIES

The uncertainty of the jet energy scale (JES) is the dominant source of systematic error on this measurement. Imperfections in the tuning of the detector simulation for the central calorimeter ($0.2 < |\eta_d| < 0.6$) energy response make the largest contribution to this uncertainty. In this region, the jet energy scale is known to better than 3% and has been expressed in a functional form [55]. The corrected jet P_T, in the PYTHIA MC with detector simulation, was varied up and down according to this parametrization. The resulting distributions were compared with the central value in order to derive the systematic uncertainty due to the JES. Even though the jet energy scale is know to better than 3%, when convoluted with the steeply falling jet P_T distributions, the uncertainties on the cross section are large. They vary from approximately 10% at low jet P_T up to as high as 60% at high transverse momentum in some rapidity regions.

There is an additional uncertainty on the jet energy scale in the higher rapidity regions ($|Y| > 1.1$). Statistics are limited when a dijet topology is required at high jet transverse momentum; as a result, the P_T -dependent correction to the MC, based on dijet P_T balance, is not very well constrained at high jet P_T . This uncertainty is approximately 40% in the highest transverse momentum bins, and is discussed in more detail in Appendix C.

The remaining sources of systematic uncertainty in this analysis are summarized below:

- Unfolding: The unfolding correction is sensitive to the momentum distributions of particles within jets. HERWIG and PYTHIA rely on different fragmentation models; therefore, the difference between their unfolding correction can be used as a measure of the sensitivity of this correction to the fragmentation model. The systematic uncertainty on the measured cross section is taken from the ratio of the unfolding factors obtained from PYTHIA and HERWIG. The difference in the unfolding factors, as obtained with weighted and un-weighted PYTHIA, is taken as additional systematic uncertainty.

- Jet energy resolution: Due to the sharply falling spectrum of the inclusive jet cross section, any imperfection in the jet energy smearing of the detector simulation will affect the unfolding correction. The bisector study revealed that the resolution difference between data and MC varied by approximately 10%, with no dependence on the transverse momentum of the jets. This variation was taken as the uncertainty on the resolution of the MC response to jet energy. The calorimeter level jets in the PYTHIA simulation have been smeared by additional Gaussian of width equal to 10% of the nominal simulated resolution. In other words, the calorimeter jet P_T was modified by

  $$P_T^{SYS} = P_T^{cal} + \sigma_{mc} \times Gauss(0, \sqrt{F^2 - 1}), \tag{8–1}$$

  where F was taken as 1.10, σ_{mc} is the nominal resolution for jets in the CDF simulation, and $Gauss(0, \sqrt{F^2 - 1})$ is a number randomly pulled from a Gaussian distribution of width $\sqrt{F^2 - 1}$ and centered at zero. The resulting P_T^{SYS} distribution was then compared with the nominally smeared result, P_T^{cal}, to obtain the systematic uncertainty on the cross section. This uncertainty is small all rapidity regions. It is less than 5% for most jet transverse momentum and still less than 10% in the highest jet P_T bins.

- Pileup correction: The pileup correction applied in this analysis was obtained from minimum bias data. The energy away from jets in dijet events, electrons in ($W \rightarrow e\nu$) events, and photons and jets in photon-jet events, as a function of the number of reconstructed vertices was also studied. The quoted uncertainty of 30% on the pileup correction covers variations from all of these cross-checks [55]. A 30% error in the pileup correction results in an uncertainty of less than 3% on the cross section measurement.

- Luminosity: There is a 6% uncertainty in the normalization if the cross section. This is a direct consequence of the uncertainty in the measurement of the luminosity at CDF [24].

- Hadron to parton level correction: The systematic uncertainty on the hadron to parton level correction is estimated from the difference in the predictions for this correction from HERWIG and PYTHIA. HERWIG does not include multiple parton interactions in its underlying event model, and instead relies on initial state radiation and beam remnants to populate the UE. PYTHIA includes MPI as well as the components included in HERWIG. PYTHIA predicts a larger correction at low P_T due to MPI. The difference between the HERWIG and PYTHIA prediction for the hadron to parton level correction is a conservative estimate of the systematic uncertainty since PYTHIA is known to reproduce the UE observables at CDF better than HERWIG [18]. The size of this uncertainty is similar in all rapidity regions, and is represented by the yellow band in figure 7–4 for the central region. It is on the order of 15% at low jet transverse momentum and is negligible for higher jet P_T .

The uncertainty in the jet energy scale leads to the largest systematic error on the inclusive jet cross section and is limited by the simulation of the calorimeter response. The total systematic uncertainty is the quadratic sum of all uncertainties listed above. When results are corrected to the parton level the hadron to parton level systematic must be included, and this error is significant for low jet transverse momentum.

CHAPTER 9
THEORETICAL PREDICTIONS AND UNCERTAINTIES

In order to use the inclusive jet cross section to extract information on the structure of the proton, the partonic cross section must be known. Any uncertainty on this perturbative calculation will limit the precision with which the parton distribution functions can be extracted. For this reason, it is important to go beyond the leading log predictions of PYTHIA. The NLO predictions have a smaller dependence on the factorization and renormalization scale, and should also be more precise since they are at a higher order in perturbation theory.

Several programs are available which make predictions for the inclusive jet cross section at next-to-leading order parton level. Three similar examples which will be mentioned in this document are: EKS [59], JETRAD [60], and FastNLO [61, 62, 63, 64]. These programs do not include the effects of hadronization, the underlying event, or parton showers. They include all diagrams that contribute to the NLO cross section. At NLO it is possible to have 2 → 2 processes like those shown in figure 2–7, and the one-loop diagrams with equivalent final states. It is also possible at NLO to have 2 → 3 processes, where an additional parton has been radiated from one of the legs or propagators in any of the diagrams of figure 2–7. These programs have at most three particles in the final state for each event. The calculations are done in the mass-less limit with five quark flavors (u, d, s, c, b, and their anti-particles), and do not include any processes besides the ones outlined above. Jets from other processes such as Z boson decays to hadrons, W boson plus jets, and top quark decays all contribute jets which are included in the inclusive jet measurement. However, the QCD cross section is so large in comparison to these processes that it is a good approximation to neglect them. For example, the $t\bar{t}$ production component of the jet sample is approximately 0.01% [65]. Contributions from other processes should even be smaller [36].

The predictions of the programs listed above depend on many input parameters such as: the method for clustering partons, factorization and renormalization scale, and parton distribution functions. There is an uncertainty on the theoretical prediction related to each of these input choices.

At CDF, the Midpoint algorithm is modified for NLO parton level calculations in order to mimic the splitting and merging step of the Midpoint algorithm ($R_{sep} = 1.3$). The value of 1.3 is considered a reasonable choice by CDF; however, other values for R_{sep} close to this value are also reasonable. The effect, on the NLO cross section prediction of FastNLO , due to varying the parameter R_{sep} is shown in figure 9–1. Increasing R_{sep} from 1.3 to 2 increases the cross section, while decreasing R_{sep} to 1 (equivalent to the K_T algorithm at NLO parton level) decreases the cross section. This variation from of 1 to 2 represents the maximal reasonable range which R_{sep} can be varied. The size of this effect is similar in all rapidity regions, and is never greater than 5% away from the prediction of $R_{sep} = 1.3$.

The factorization scale, μ_f, is often taken to be one half of the transverse momentum of the jet. This is convenient because this is the scale used to determine the proton structure functions by the CTEQ group. Using $\mu_f = P_T^{jet}$ or $\mu_f = 2P_T^{jet}$ gives approximately 10% and 20 % smaller predictions for the cross section, respectively [46]. However, if one extracted parton density function fits based on $\mu_f = P_T^{jet}$ or $\mu_f = 2P_T^{jet}$in order to use a consistent value of μ_f for the PDF and the calculation, the dependence would be reduced.

The dominant theoretical uncertainty, on the inclusive jet cross sections at NLO, is due to the error introduced into the calculation by the uncertainty on the proton structure functions. The Run I inclusive jet experience revealed the need for tools which could quantify this effect. Experimental constraints must be incorporated into the uncertainties of parton distribution functions before they can be propagated through to errors on the predictions of observables. In recent years, tools [66] such as the Hessian Method [67], have been developed to make error propagation to final state observables possible. The CTEQ6.1M [37] error sets have 41 PDF members: a central value, and 40 error set members. The error set members correspond to 20 eigenvector directions which have been varied in the positive and negative directions. The eigenstates are obtained from diagonalizing the Hessian error matrix. The matrix is obtained by varying each parameter used in the global fit within the tolerances of the experimental data included. In order to approximate the PDF uncertainty on the prediction of a physical observable with this method, the observable must be calculated with each PDF set member (*i.e.*, the observable must be calculated 41 times). After the prediction has been calculated with each member, the uncertainty on the observable is calculated with the following equations [68, 69, 70]:

$$\Delta X^{+}_{max} = \sqrt{\sum_{i=1}^{N}[max(X_i^{+} - X_0, X_i^{-} - X_0, 0)]^2} \tag{9–1}$$

and

$$\Delta X^{-}_{max} = \sqrt{\sum_{i=1}^{N}[max(X_0 - X_i^{+}, X_0 - X_i^{-}, 0)]^2}, \tag{9–2}$$

where the X_0 is the central value of the observable; X_i^{+} and X_i^{-} are the values obtained from the plus and minus variations along the i^{th} eigenvector direction; and ΔX^{-}_{max} and ΔX^{+}_{max} are the positive and negative uncertainties on the physical observable. The PDF error, calculated in this way, is included on all comparisons of data with the NLO predictions in section 10.

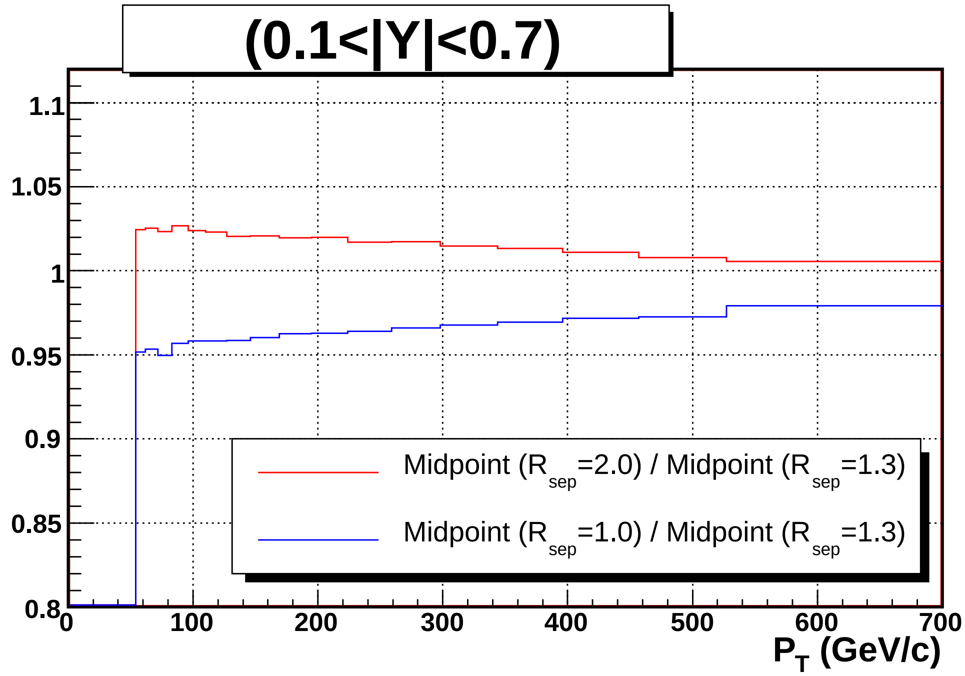


Figure 9–1. Effect of varying the parameter R_{sep} on the NLO cross section prediction of FastNLO for the central rapidity region. Increasing R_{sep} from 1.3 to 2 increases the cross section, while decreasing R_{sep} to 1 (the equivalent of the K_T algorithm) decreases the cross section.

CHAPTER 10
RESULTS

The results of the inclusive jet cross section measurement with over 1 fb^{-1} of Run II integrated luminosity are shown in figures 10–1 - 10–5. In each figure, the inclusive differential jet cross section corrected to the hadron level is shown in (a), and the ratio of data (corrected to the parton level) to the NLO parton level predictions of EKS, with CTEQ6.1M parton distribution functions, is shown in (b). The yellow band shows the experimental systematic uncertainty. All sources of systematic error are considered independent and have been added in quadrature. The blue band also includes the modeling uncertainty associated with the hadronization and underlying event corrections. This systematic associated with the hadron to parton level correction is added in quadrature to the total experimental systematic. There is an additional 6 % normalization uncertainty due to the uncertainty on the integrated luminosity which has not been included in the figures. The uncertainty on the theoretical prediction due to estimated error on the proton structure functions is drawn in red on the ratio plots.

Figure 10–1(a) shows the inclusive jet cross section corrected to the hadron level in the well understood central region of the detector. The vertical axis is plotted on a log scale. The cross section varies by more than eight orders of magnitude as the jet transverse momentum increases from 55 GeV/c to approximately 650 GeV/c. The differences between the measured result and the theoretical predictions of EKS are not resolvable on the log scale. In figure 10–1(b), the ratio of the measured cross section corrected to the parton level to the NLO prediction is shown. Exceptionally good agreement is observed. The systematic uncertainty varies from approximately 20% at low jet P_T up to 80% in the highest transverse momentum bin. The systematic uncertainty is larger than statistical errors on every bin. The systematic errors are slightly smaller than the PDF errors for the central P_T range indicating that this measurement will be useful to constrain the parton densities.

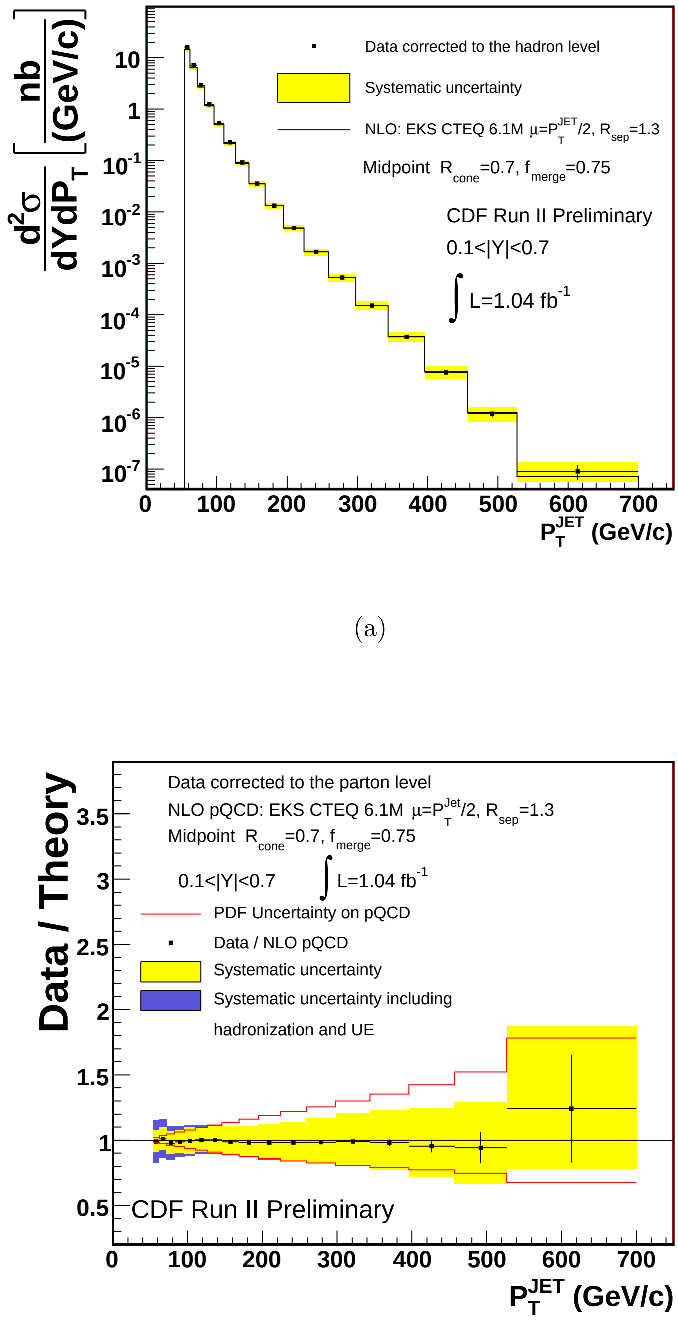
Figure 10–2(a) shows the inclusive jet cross section corrected to the hadron level in the central crack region. The cross section varies by approximately seven orders of magnitude as the jet transverse momentum increases from 55 GeV/c to approximately 650 GeV/c. In figure 10–2(b), the ratio of the measured cross section corrected to the parton level to the NLO prediction is shown. The agreement is good in most bins. There is a slight excess in the highest two transverse momentum bins. It is possible that the effect of the crack needs more attention in this region, or these fluctuations could be statistical. Regardless, the measured result is consistent with the theoretical prediction when the error on the parton distributions functions is also considered. The systematic uncertainty varies from approximately 20% at low jet P_T up to 100% in the highest transverse momentum bin. The systematic uncertainty is larger than statistical errors on every bin except for the highest jet P_T bin. PDF error and systematic error are roughly of the same order in this rapidity region.

The two regions where the central calorimeter and the plug calorimeter modules overlap have similar features and will be discussed in tandem. Figures 10–3(a) and 10–4(a) show the inclusive jet cross section corrected to the hadron level in in the rapidity regions $0.7 < |Y| < 1.1$ and $1.1 < |Y| < 1.6$, respectively. The measurement includes jet P_T up to approximately 500 GeV/c in region $0.7 < |Y| < 1.1$, and only up to 400 GeV/c in region $1.1 < |Y| < 1.6$. In figures 10–3(b) and 10–4(b), the ratio of the measured cross section corrected to the parton level to the NLO prediction is shown. The agreement is good in most bins; however, the NLO prediction seems to be systematically higher than the measured result. This is a small effect and the NLO predictions are consistent with the measured result for all jet transverse momenta. In both regions the systematic uncertainty is slightly smaller than the PDF uncertainty. These results will be useful to constrain proton structure functions.

Figure 10–5(a) shows the inclusive jet cross section corrected to the hadron level for the highest rapidity bin. The measurement only goes up to approximately 300 GeV/c in jet transverse momentum because the jet cross section falls off much more rapidly at high rapidity. In fact, the cross section varies by approximately seven orders of magnitude as the jet transverse momentum varies from 55 GeV/c to approximately 300 GeV/c. In figure 10–5(b), the ratio of the measured cross section corrected to the parton level to the NLO prediction is shown. The the NLO prediction of EKS is systematically higher than the measured cross section over the full range of jet P_T; however, when systematic and PDF uncertainties are considered it is still consistent with the measured result. The systematic errors are largest in this region and approach 170% for the highest bin in jet P_T. The steeper shape of the P_T distribution in this region, combined with the additional systematic on the jet energy scale due to the P_T-dependent correction, are responsible for this increased uncertainty. Even so, the PDF uncertainty is still significantly larger than the systematic errors for most jet P_T. The result in this rapidity region will lead to the strongest constraint on parton density functions out of all the regions included in this measurement.

There is a trend in the data for the last bin to fluctuate higher than the NLO prediction. There is a simple explanation for this systematic effect. There are very few events in the last bins. If this number fluctuated down significantly then there would not be events in the bin, and therefor it would not be shown on the figure. By this reasoning, bins with very few events are more likely to be statistical fluctuations which added to the bin contents rather than fluctuations which subtracted from the contents of the bin.

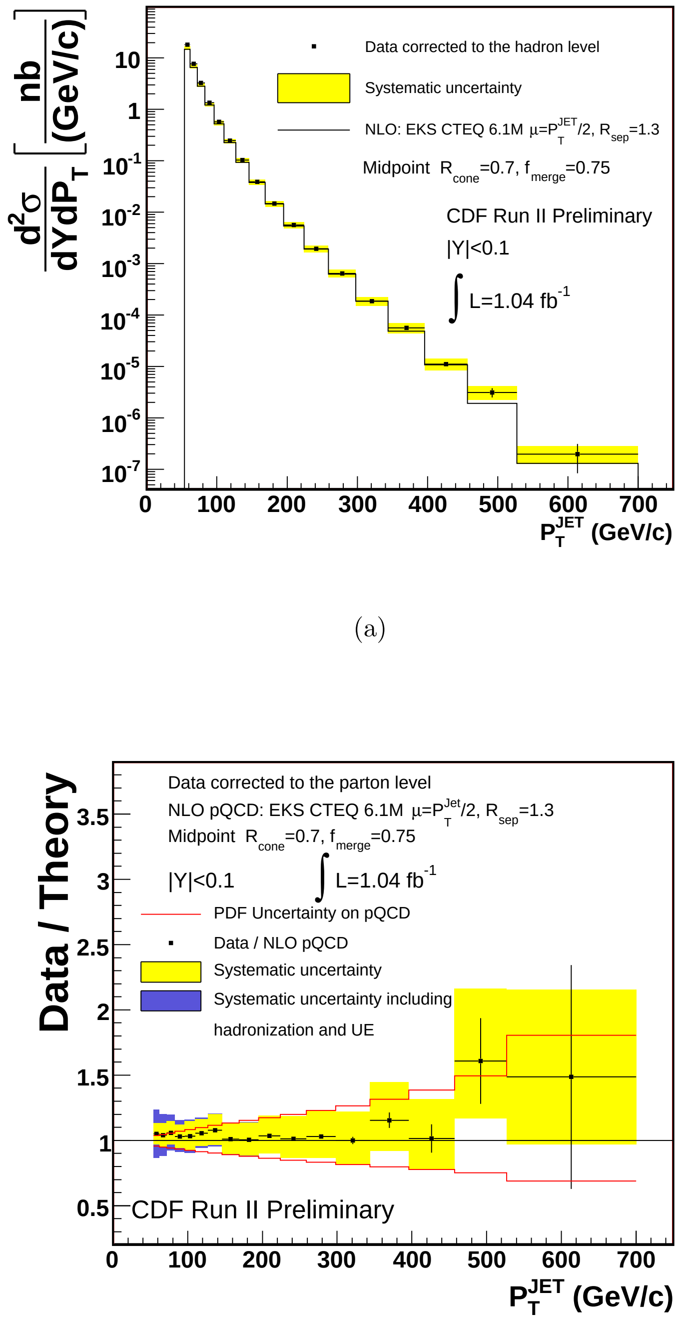
The cross sections for the various rapidity regions are presented on figure 10–6 where they have been scaled by different factors so they would be distinguishable when plotted on the same axis. The region $0.7 < |Y| < 1.1$ has note been scaled. The regions $|Y| < 0.1$, $0.1 < |Y| < 0.7$, $1.1 < |Y| < 1.6$, and $1.6 < |Y| < 2.1$ have been scaled by the factors of 10^6, 10^3, 10^{-3}, and 10^{-6}, respectively. Figure 10–7 shows the ratios to NLO pQCD predictions for the different rapidity regions.



(a)

(b)

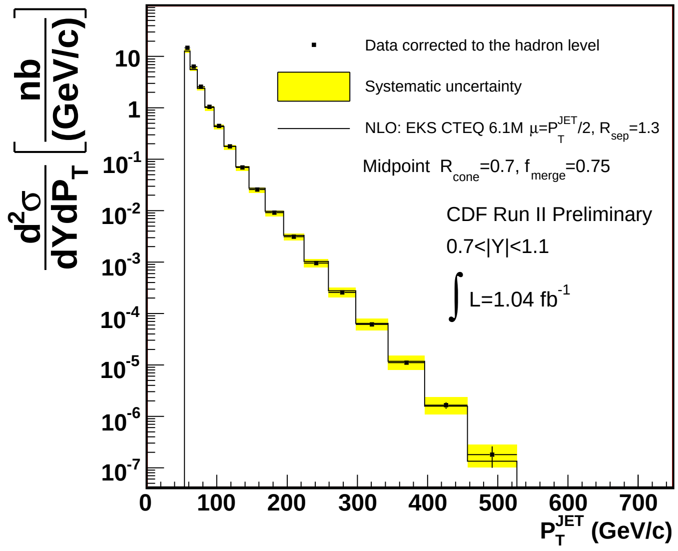
Figure 10–1. Measured inclusive jet cross section with the Midpoint algorithm in the region $0.1 < |Y| < 0.7$. The distribution for the hadron level cross section is shown in figure 10–1(a). The ratio of data corrected to the parton level to the parton level pQCD prediction of EKS is shown in figure 10–1(b).



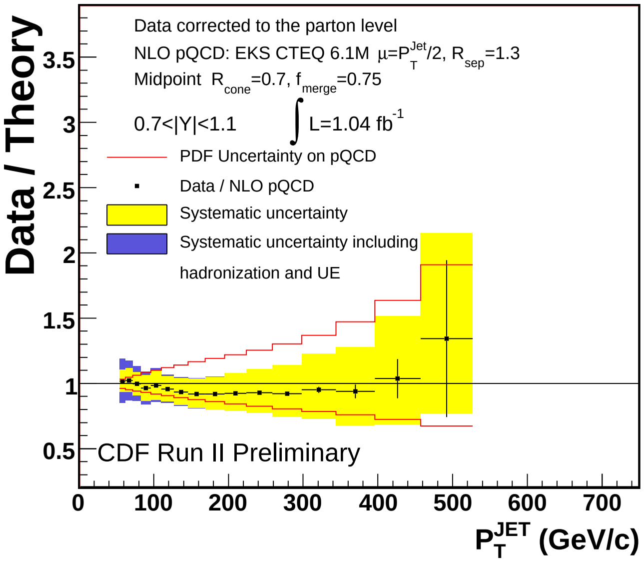
(a)

(b)

Figure 10–2. Measured inclusive jet cross section with the Midpoint algorithm in the region $|Y| < 0.1$. The distribution for the hadron level cross section is shown in figure 10–2(a). The ratio of data corrected to the parton level to the parton level pQCD prediction of EKS is shown in figure 10–2(b).

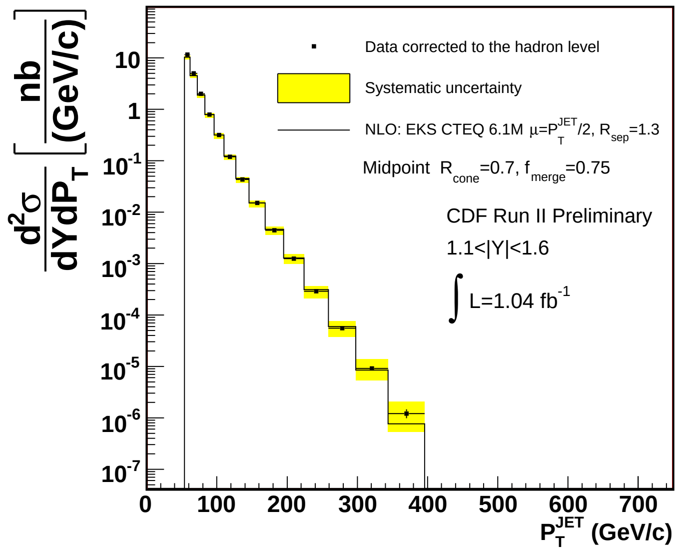


(a)

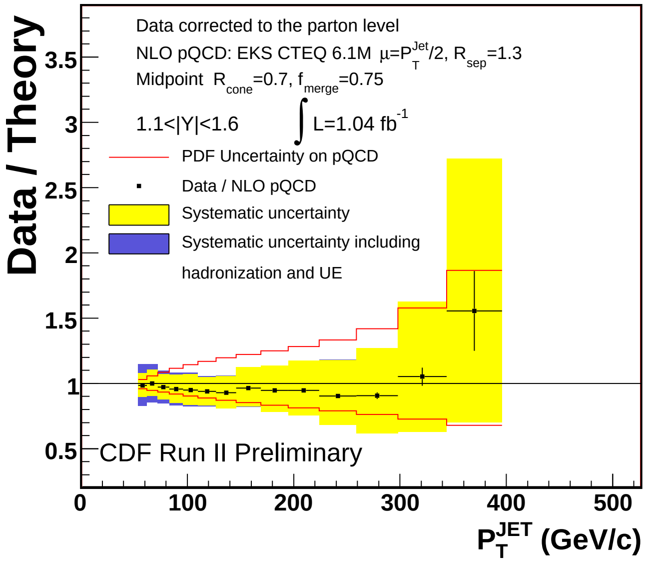


(b)

Figure 10–3. Measured inclusive jet cross section with the Midpoint algorithm in the region $0.7 < |Y| < 1.1$. The distribution for the hadron level cross section is shown in figure 10–3(a). The ratio of data corrected to the parton level to the parton level pQCD prediction of EKS is shown in figure 10–3(b).

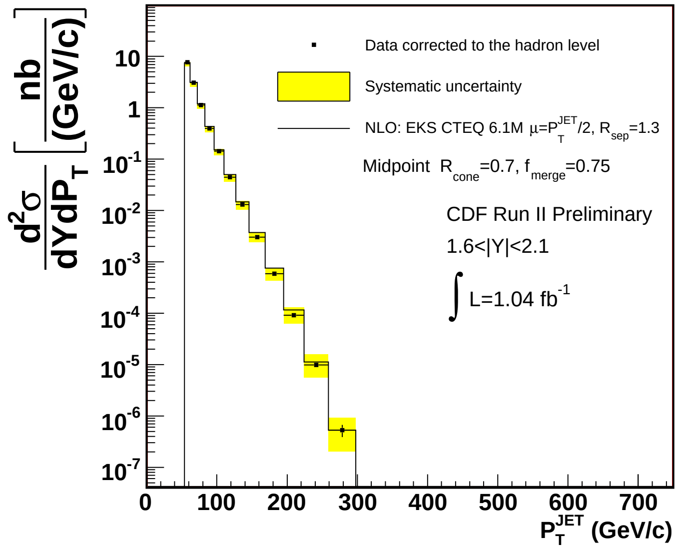


(a)

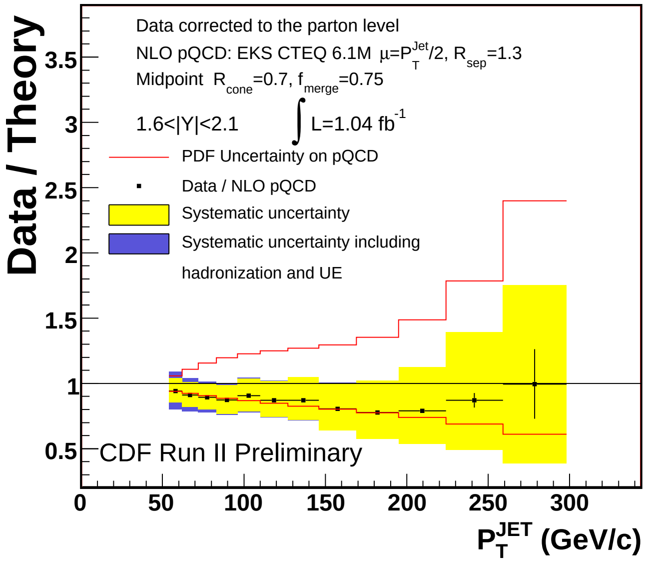


(b)

Figure 10–4. Measured inclusive jet cross section with the Midpoint algorithm in the region $1.1 < |Y| < 1.6$. The distribution for the hadron level cross section is shown in figure 10–4(a). The ratio of data corrected to the parton level to the parton level pQCD prediction of EKS is shown in figure 10–4(b).



(a)



(b)

Figure 10–5. Measured inclusive jet cross section with the Midpoint algorithm in the region $1.6 < |Y| < 2.1$. The distribution for the hadron level cross section is shown in figure 10–5(a). The ratio of data corrected to the parton level to the parton level pQCD prediction of EKS is shown in figure 10–5(b).

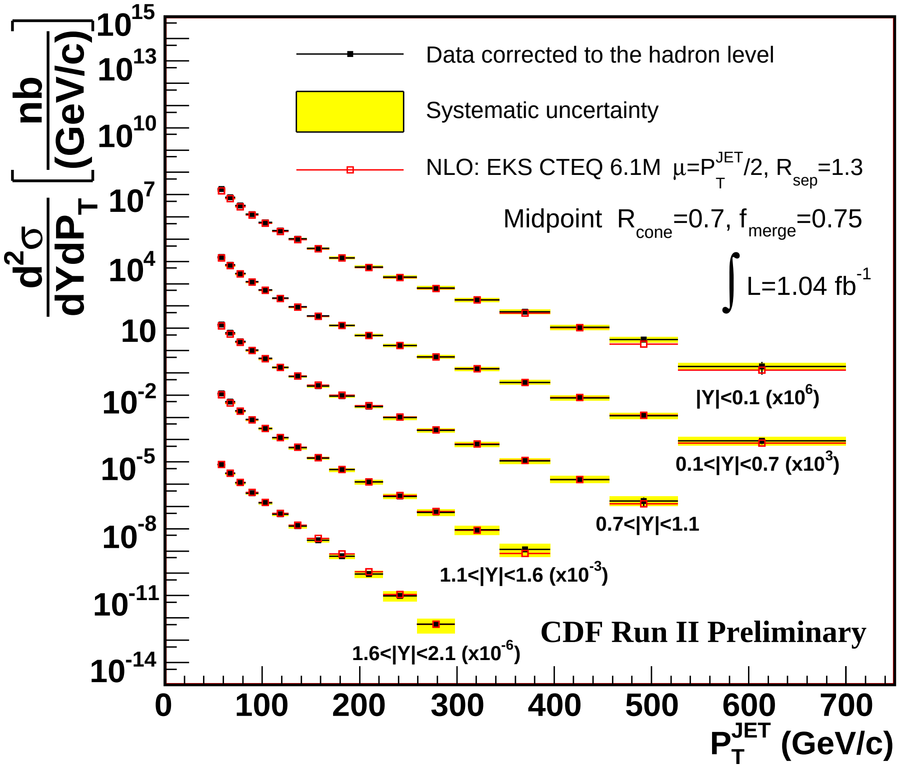


Figure 10–6. Measured inclusive jet cross section at the hadron level with the Midpoint algorithm for all rapidity regions.

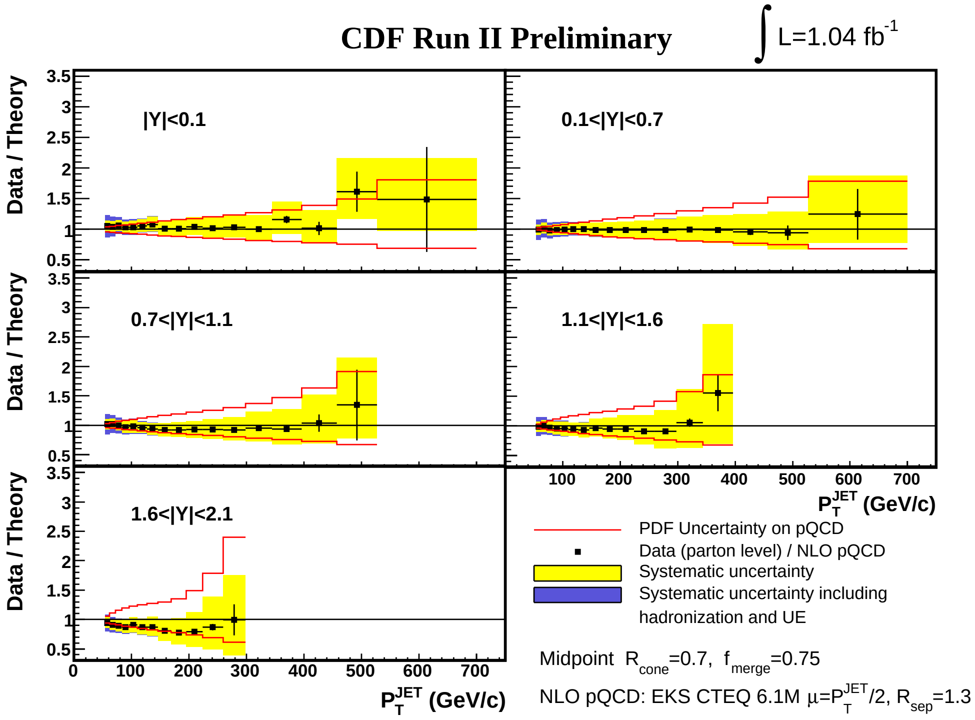


Figure 10–7. Ratio of the measured inclusive jet cross section at the parton level with the Midpoint algorithm to the pQCD prediction of EKS in the different rapidity regions.

CHAPTER 11
COMPARISON WITH THE KT ALGORITHM

CDF has made inclusive jet cross section measurements with the Midpoint and the K_T jet clustering algorithms in Run II. Here, a simple comparison of these measurements in the central region ($0.1 < |Y| < 0.7$) is presented. In order to require that the algorithms use a similar scale for clustering in $y-\phi$ space, the cone size for the Midpoint algorithm and the D-parameter of the K_T algorithm will both be taken as 0.7 ($R_{cone} = D = 0.7$). It is important to note that different algorithms correspond to different observables. The inclusive jet cross section is not expected to be exactly the same for jets clustered with these two different algorithms (even though $R_{cone} = D = 0.7$). The NLO predictions for the cross sections, as well as the hadron to parton level corrections, will also be compared for the two algorithms.

Historically, the K_T jet clustering algorithm has been used successfully at electron-positron colliders, and electron-proton colliders [71, 27]. Only recently has this algorithm been applied to jets in the more challenging hadron-hadron collider environment of the Tevatron: studied first by D0 [72] in Run I, and more recently by CDF in Run II. D0 reported only marginal agreement with NLO prediction; however, the more recent CDF result reports good agreement. The ratio of the CDF result to the NLO prediction of the JETRAD program is shown in figure 11–1 [73]. In the figure, the JETRAD result has been corrected to the hadron-level to include the effects of jet fragmentation and the underlying event. This correction is shown in figure 11–2 [73].

Figure 11–3 shows the ratio of the inclusive jet cross section measurement for jets clustered with the K_T algorithm to the result for jets clustered with the Midpoint algorithm (black). Only statistical errors are shown. The data used in the

K_T measurement is a subset of the data included in the Midpoint measurement, so the statistics are clearly correlated; however, the errors were propagated based on no correlation because all information required for full error propagation was not available. This results in a slight over-estimate of the true statistical error on the ratio. Systematic errors are not included, and could provide an additional large contribution to the uncertainty. The prediction of this ratio from the NLO program FASTNLO is also shown in the figure (blue). At the NLO parton level, where there are at most three objects in the event, the Midpoint algorithm with $R_{sep} = 1$ is equivalent to the K_T algorithm. However, the parameter R_{sep} is taken as 1.3 for the Midpoint algorithm at CDF. The standard value of R_{sep} being equal to 2 would result in a prediction of a larger difference between the two algorithms at NLO (*i.e.,* setting $R_{sep} = 2$ results in a larger jet cross section as seen in figure 9–1). The P_T dependence and magnitude of the measured ratio observed in the data is close to the NLO prediction. This result provides confidence that the algorithm definitions are consistent at calorimeter level and NLO parton level.

Figure 11–3 shows the ratio of the hadron to parton level correction derived with the K_T algorithm (inversion of the distribution shown in 11–2) to the one derived with the Midpoint algorithm (figure 7–4). These corrections were derived from PYTHIA TUNE-A, as described in section 7. The multiplicative corrections are both less than one, so the ratio means that the size of the correction derived with the K_T algorithm is larger (*i.e.,* farther away from one) than the correction derived with the Midpoint algorithm. This result indicates that the K_T algorithm is slightly more sensitive to the underlying event.

The consistency of the K_T inclusive jet cross section measurement with NLO predictions, combined with only a sightly larger underlying event correction, supports the use of K_T -type algorithms at the Tevatron and hadron-hadron colliders of the future, provided that one has a good understanding of the UE. The agreement, between the NLO prediction and the measured result for the ratio of the the two jet clustering algorithms, adds credence that the jet definitions are defined consistently at the parton and detector levels.

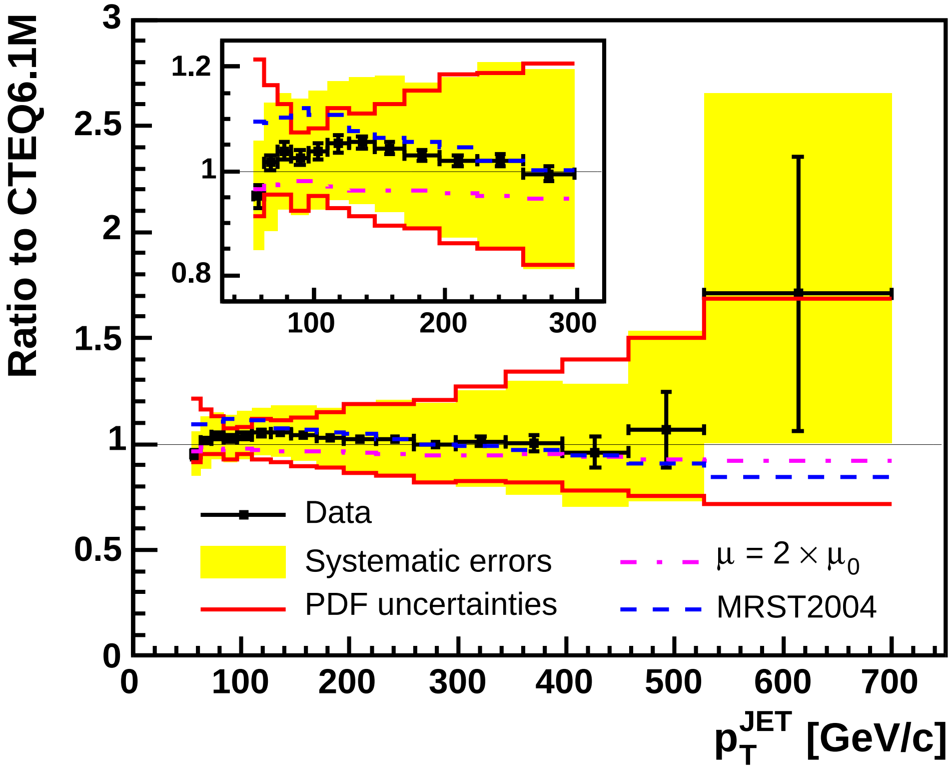


Figure 11–1. Measured inclusive jet cross section with the K_T algorithm in the rapidity region $0.1 < |Y| < 0.7$. The ratio of the hadron level cross section to the NLO prediction of JETRAD is shown. JETRAD has been corrected to the hadron-level to include the effects of fragmentation and the underlying event.

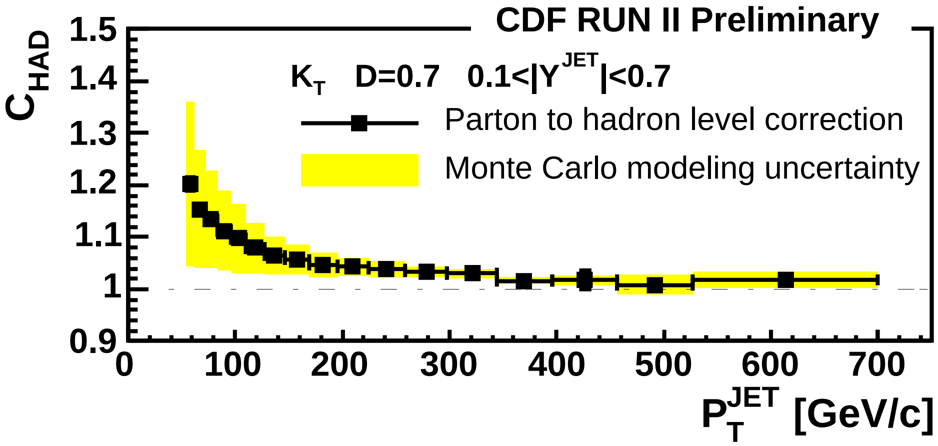


Figure 11–2. Parton-to-hadron correction used by the K_T inclusive jet cross section analysis to correct the NLO prediction to the hadron level.

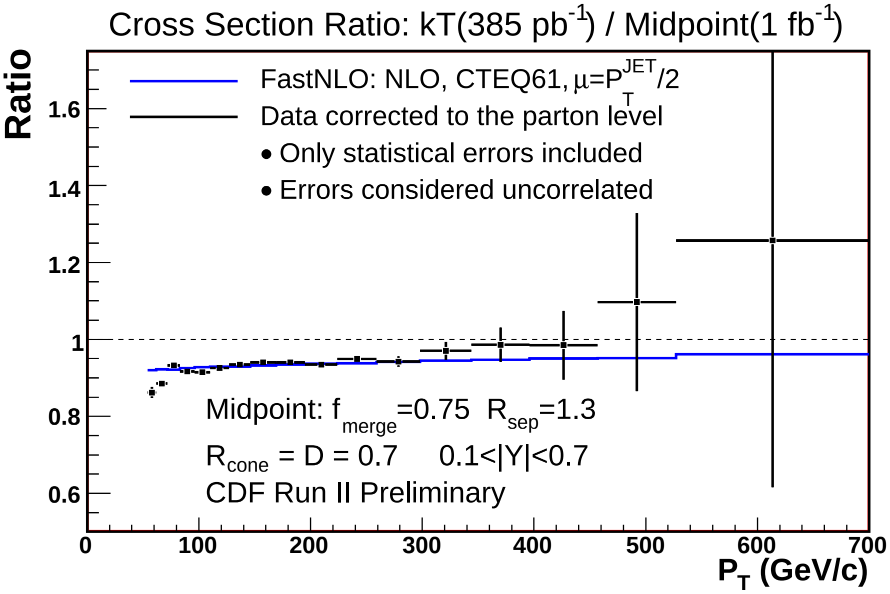


Figure 11–3. Ratio of the inclusive jet cross section measured with the K_T algorithm to that measured by the Midpoint algorithm (black). The prediction of this ratio from the NLO program FastNLO is also shown in the figure (blue).

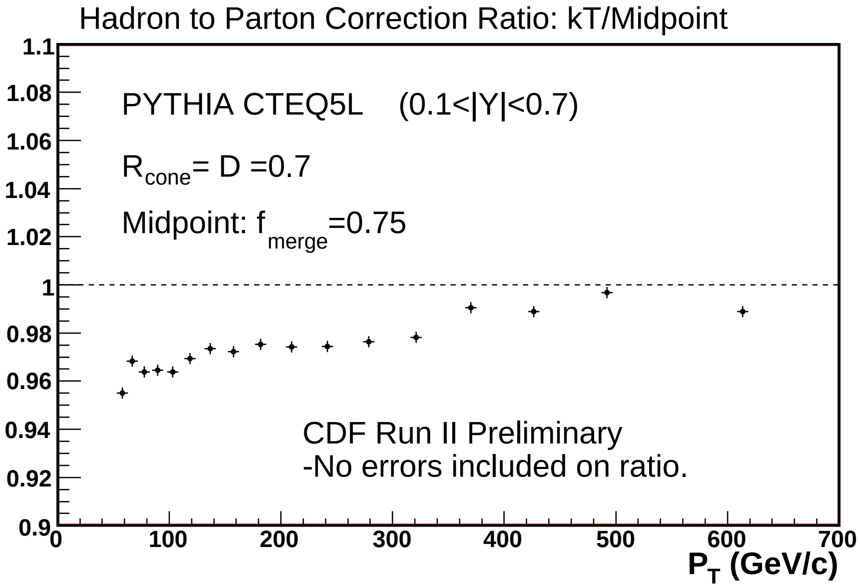


Figure 11–4. Ratio of the hadron to parton level correction derived with the K_T algorithm to that derived with the Midpoint algorithm. The multiplicative corrections are both less than one. The correction derived with the K_T algorithm is larger (farther away from one). The corrections were derived from PYTHIA TUNE-A.

CHAPTER 12
CONCLUSIONS

In this dissertation, a measurement was presented of the inclusive jet cross section for jets clustered by the CDF Midpoint jet finding algorithm using over 1 fb^{-1} of data collected by the CDF experiment. Good agreement is observed in all rapidity regions with the pQCD predictions of the NLO program EKS using CTEQ6.1M parton distributions. This confirms the theory of QCD at a higher energy scale than previous measurements, and implies that quarks are *point-like* down to a scale of approximately 10^{-19} m.

This measurement also provides important constraints on the parton content of the proton, which improves the precision of theoretical predictions for all proton colliders. In the most forward region ($|Y| > 1.6$), the systematic uncertainty on the measurement is significantly smaller than the PDF uncertainty. The measurement in this region will lead to constraints on global PDF fits which will reduce the uncertainty on many important theoretical predictions. For example, reducing the PDF uncertainties is important for measurements regarding the Higgs boson. The uncertainty on the NLO cross section for Higgs production due to the PDF error is on the order of 5-10% at the Tevatron and the LHC [74]. Uncertainties of this order do not determine if the Higgs boson will be discovered or not; however, they do limit the precision with which the coupling of the Higgs to other particles can be measured after its discovery.

In all rapidity regions, and over the full range of jet P_T excluding the highest bins, the systematic uncertainty is larger than the statistical error on this measurement. This means that the precision of the measurement is limited by systematics and not by statistics. Simply obtaining more data will not improve the accuracy of the measurement significantly. The systematic uncertainty due to the jet energy scale can be reduced by improving the simulation of the CDF detector. Currently, CDF is working to improve the detector simulation. This will lead to a reduced uncertainty on the jet energy scale, and to a smaller systematic error on future inclusive jet measurements.

There are aspects of the inclusive jet cross section which have implications for the LHC. It is important for the LHC to choose a consistent set of jet algorithms early on between experiments and stick with these jet definitions throughout the life of the experiment. No jet algorithm is perfect, and some algorithms are better suited for certain types of analyses. Making a strong effort to understand the flaws of an algorithm may be a better approach than to try to correct all of the imperfections of an algorithm after it is in use. Perhaps, this lesson should be learned from Tevatron experiments, where an effort to improve the Midpoint algorithm only introduced a new set of imperfections and led to the two collider experiments using different jet definitions.

The fact that no significant deviations from the predictions of pQCD have been observed in the inclusive jet cross section is a useful result for LHC experiments. In the early running of the LHC it is likely that many deviations from the standard model predictions will be observed. Any deviations observed in jets with transverse momentum below 600 GeV/c are likely detector effects which are not completely understood, rather than some sign of physics beyond the standard model. In other words, confidence in QCD predictions at this energy scale provides a kinematic region where it is safe to calibrate the detector. Some of the techniques applied in this analysis will also be useful at the LHC. For example, the method of dijet P_T balancing and the bisector method can both be used to understand the detector response, and to fine tune the detector simulation for the

LHC experiments. In general, the experience gained at the Tevatron improves techniques and tools that will be applied at the LHC.

Sometimes, it is assumed that the range of discovery at an experiment is only limited by the integrated luminosity and center of mass energy provided by the accelerator. The inclusive jet cross section provides an example in which the precision is not limited by statistics. In many cases the level at which the detector is understood is just as crucial to making important discoveries as the performance of the accelerator.

Over the next few years, the physicists of the LHC experiments will be working to understand their detectors. This effort should lead to exciting discoveries regarding the most fundamental laws of the universe. The interpretation of these results would not be possible without the wealth of previous measurements of standard model properties, such as the inclusive jet cross section. The results from the LHC will define the future of the fascinating field of high energy particle physics, and will bring human beings one step closer to a complete understanding of the universe in which they live.

Hopefully, it will be a giant step!

APPENDIX A
JET TRIGGERS AT CDF

The rate at which collisions happen within the CDF detector is much higher than the rate at which data can be collected and stored. Therefore, the CDF trigger is a very important part of the data acquisition system. Quantitatively, bunch crossings occur at CDF at an approximate rate of 1.7 MHz, while data can only be written to tape a about 75 Hz. Various triggers are designed to extract events that are useful for physics analysis at CDF. The trigger is split into three levels at CDF. At each level the events which pass the trigger requirements are passed to the next level. The trigger requirements at each level must produce a rate reduction large enough for processing at the next level.

Soft collisions (low energy) dominate the spectrum of events if no trigger is in place. In order to statistically populate the full range of jet energy available at CDF Run II, multiple jet triggers with different energy thresholds are used. The flow of the four jet triggers used for physics studies in this analysis is shown in figure A–1.

A detailed description of the CDF trigger is available in reference [48]. Each of the three levels of the CDF jet trigger are briefly described below:

- Level-1: The level-1 trigger has 132 ns to decide if an event will pass on to level-2 or not. This time constraint limits the hardware access to information from the calorimeter and the COT. The level-1 jet triggers rely on applying thresholds to the sum of hadronic and electromagnetic energy in individual calorimeter towers. Because jets are made of more then one calorimeter tower, the thresholds at level-1 must be less than the jet energy requirements at the higher trigger levels. As shown in figure A–1 STT5 (Single Trigger Tower 5), and STT10 (Single Trigger Tower 10) are the 2 triggers relevant for jets at level-1. STT5 and STT10 require 5 GeV and 10 GeV in a single tower. The level-1 trigger must reduce the 1.7 MHz bunch crossing rate to the level-2

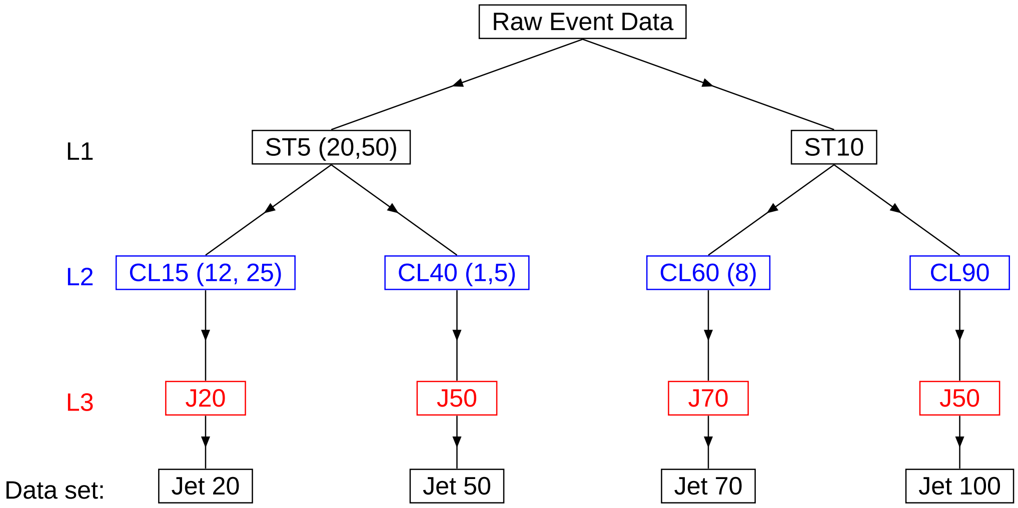


Figure A–1. Trigger flow diagram for the four jet triggers. Prescales are given in parentheses. If multiple prescales are shown then the prescale for that trigger was changed during the data taking period.

design rate of 40 kHz. For the level-1 STT5 trigger a prescale is required in addition to the above mentioned requirements to make this rate reduction. At CDF jet triggers are prescaled by accepting a fixed fraction of events which pass the trigger requirements. The prescale required may need to be changed based on luminosity. Due to increases in the instantaneous luminosity, the jet20 and jet50 L1 and L2 prescales have changed during the period of data taking for this analysis. The details of these prescale changes are given in section A.2.

- Level-2: At level-2 a "nearest neighbor" cluster finding algorithm is used to cluster adjacent calorimeter towers with non-trivial energy. Clusters are formed around all seed towers with energy above a certain threshold. All neighboring towers with energy above a second, slightly lower threshold, are then added to the cluster until no towers adjacent to the cluster have energy above this second threshold. The total EM and HAD energies are recorded for each cluster as well as the number of towers and the (η, ϕ) of the seed tower. The calorimeter towers are summed into trigger towers, whose energies are weighted by $sin\theta$. As shown in figure A–1, the 4 level-2 jet triggers are CL15, CL40, CL60, and CL90 for jet[3]20, 50, 70, and 100 respectively. It is the job of these level-2 triggers to reduce the 40 kHz rate passed by level-1 to approximately 300 Hz which can be handled by the level-3 system. CL15, and CL60 have prescales that have varied over the data included in this analysis and are given in the trigger flow diagram. See section A.2 for details on calculating their effective prescales.

- Level 3: At Level3 CDF uses the JetClu algorithm to form jets. The JetClu algorithm is briefly described Chapter 4 and in more detail in reference [36]. The 300 Hz passed from level-2 must be reduce to approximately 75 Hz before the data is written to tape.

A.1 Jet Trigger Efficiency

It is important that the kinematic region of data used in a physics analysis is not biased due to trigger requirements. A detailed study of the trigger efficiencies was conducted for each trigger to discern in which P_T regions the data is is unbiased.

A trigger is 100 % efficient for a bin of a distribution when the contents of the bin would be unchanged if the trigger requirements were removed. Therefore, in order to study the efficiencies of a trigger, data taken with a looser trigger requirement is used. The trigger requirements for the tighter trigger are then applied to the data and the trigger is considered efficient for a given bin when its efficiency at all levels is greater than 99.5 %. The trigger efficiency of a bin can be found by dividing the number of events in the bin that passed the trigger requirements being studied, N_{trig}, by the number of events in the bin before making the extra trigger requirements, N_{tot},

$$\epsilon = \frac{N_{trig}}{N_{tot}}. \tag{A–1}$$

The distributions obtained from these ratios are called *turn-on* curves.

A complete study of the trigger efficiencies was done including as much of the data sample as possible. Each trigger, data samples used to study the efficiency of each trigger, and the data samples which each trigger contributes to shown in table A–1. Since the L1 trigger becomes efficient at much lower P_T than L2 or L3 (see figures A–2 and A–3), its contribution to the trigger efficiency in the bins of interest is negligible. Therefore, L2 and L3 efficiencies have been combined and the results are shown in figures A–4 - A–7. For completeness, the trigger turn-on curves were fit to the function

$$\epsilon(P_T) = \frac{1}{1 + e^{-P_0(P_T+P_1)}}, \tag{A–2}$$

and the fit parameters (P_0 and P_1) are included on the figures.

Table A–1. Data samples used to study the efficiency (middle) of each trigger. The data samples, which each trigger contributes, to is also shown(right).

Trigger	Sample used	Data set affected
L1:ST05	High P_T μ	Jet20 and Jet50
L2:J15	ST05	Jet20
L3:J20	ST05	Jet20
L2:J40	Jet20	Jet50
L3:J50	Jet20	Jet50
L1:ST10	Jet20	Jet70 and Jet100
L2:J60	Jet50	Jet70
L3:J70	Jet50	Jet70
L2:J90	Jet70	Jet100
L3:J100	Jet70	Jet100

A.2 Jet Trigger Prescales

As mentioned above, some triggers require prescales (in addition to the kinematic requirements placed on the detector information available) to meet the reduction in rate required by the next trigger level. The jet100 trigger is not prescaled. The jet70 trigger is prescaled by a factor of 8. It was cross-checked by taking the ratio of the jet100 to jet70 P_T distributions. The correct value of 8 is obtained by fitting this ratio to a constant value in the region where jet100 and jet70 are both efficient as shown in the top left plot of figure A–8.

The study of prescales for the jet20 and jet50 triggers are complicated by the fact that the samples are prescaled at multiple levels and these prescales have not remained constant during the the data period considered. The jet20 and jet50 L1 prescale changed from 20 to 50 at run number 184444. The jet20 L2 trigger increased from 12 to 25 after run 153068, while the jet50 L2 is only prescaled for runs after run 194917 with initial instantaneous luminosity greater than 150×10^{30} cm^{-2}sec^{-1}. In order to deal with these changes and to treat all events equally, an ‘effective’ prescale was calculated.

The effective prescale was calculated by luminosity-weighting the inverse of prescale factors. Eq. 6–1 from Chapter 6 clarifies this method. This method requires calculating the integrated luminosity separately for each prescale period. For jet20 the effective prescale is

$$P_{eff}^{J20} = \frac{967}{\frac{47.4}{20\times12} + \frac{259.7}{20\times25} + \frac{659.9}{50\times25}} = 776.8. \tag{A–3}$$

Similarly, for jet50 the effective prescale can be calculated as

$$P_{eff}^{J50} = \frac{967}{\frac{307.0}{20\times1} + \frac{639.3}{50\times1} + \frac{20.7}{50\times5}} = 34. \tag{A–4}$$

As a cross check, the effective prescale were also be obtained from the ratio of jet70 to jet50 for the jet50 trigger, and the ratio of jet70 to jet20 for the jet20 trigger. In Figure A–8 these distributions are fit to a constant value in the region where both relevant distributions are efficient. In order to take the jet70 prescale into account these ratios are scaled by a factor of eight. The prescales obtained are also shown in the figures. The effective prescales obtained from the two methods are consistent to better than 1.0 %.

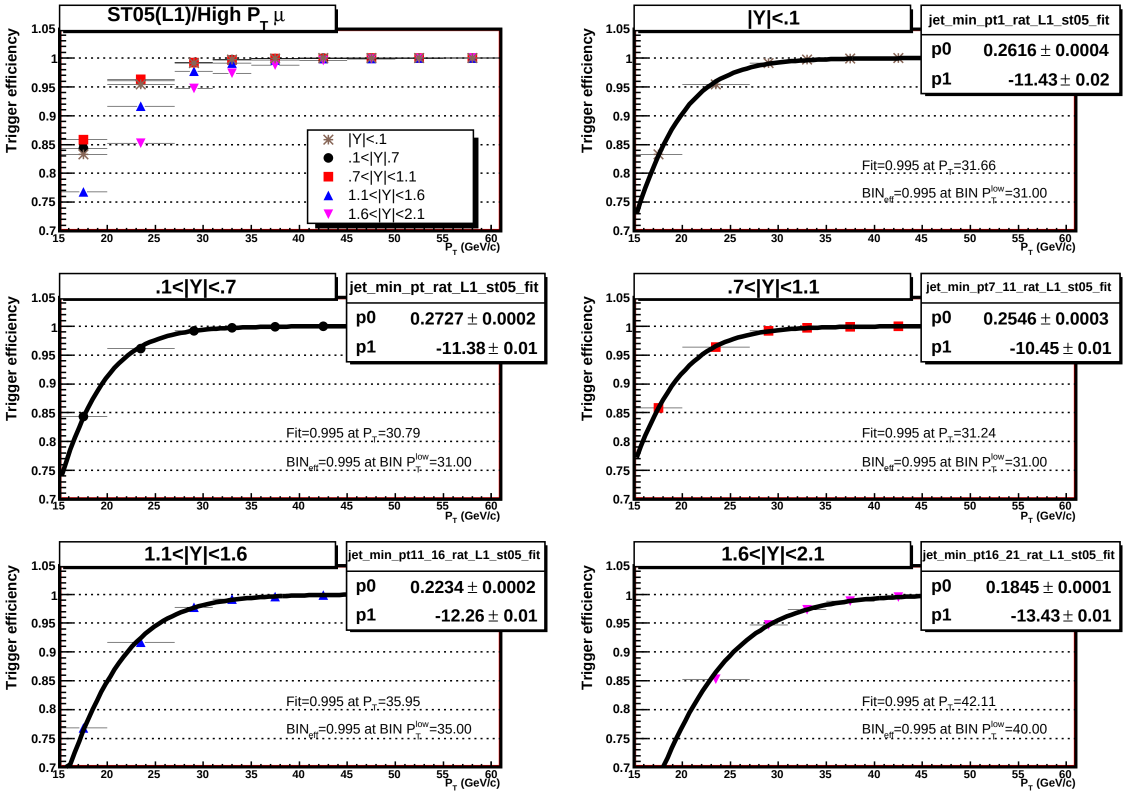


Figure A–2. Trigger efficiencies as function of uncorrected jet P_T for the L1-ST5 trigger. This is the level-1 trigger for the jet20 and jet50 trigger samples.

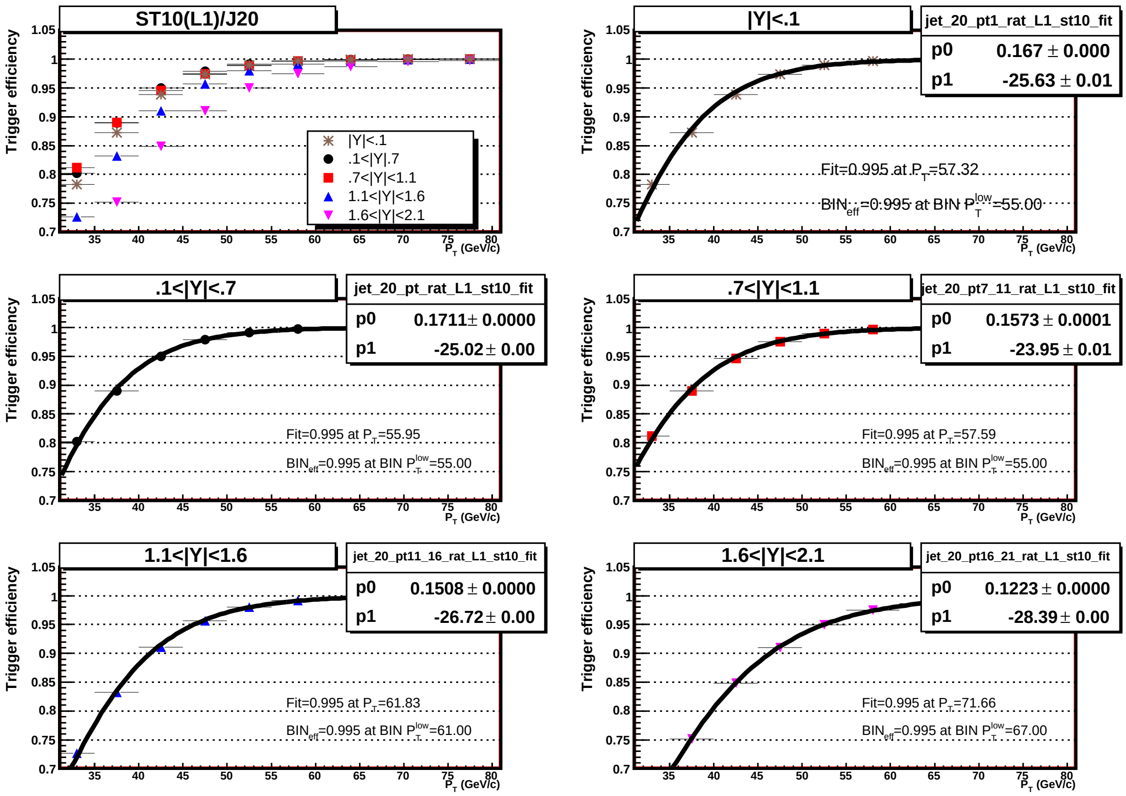


Figure A–3. Trigger efficiencies as function of uncorrected jet P_T for the L1-ST10 trigger. This is the level-1 trigger for the jet70 and jet100 trigger samples.

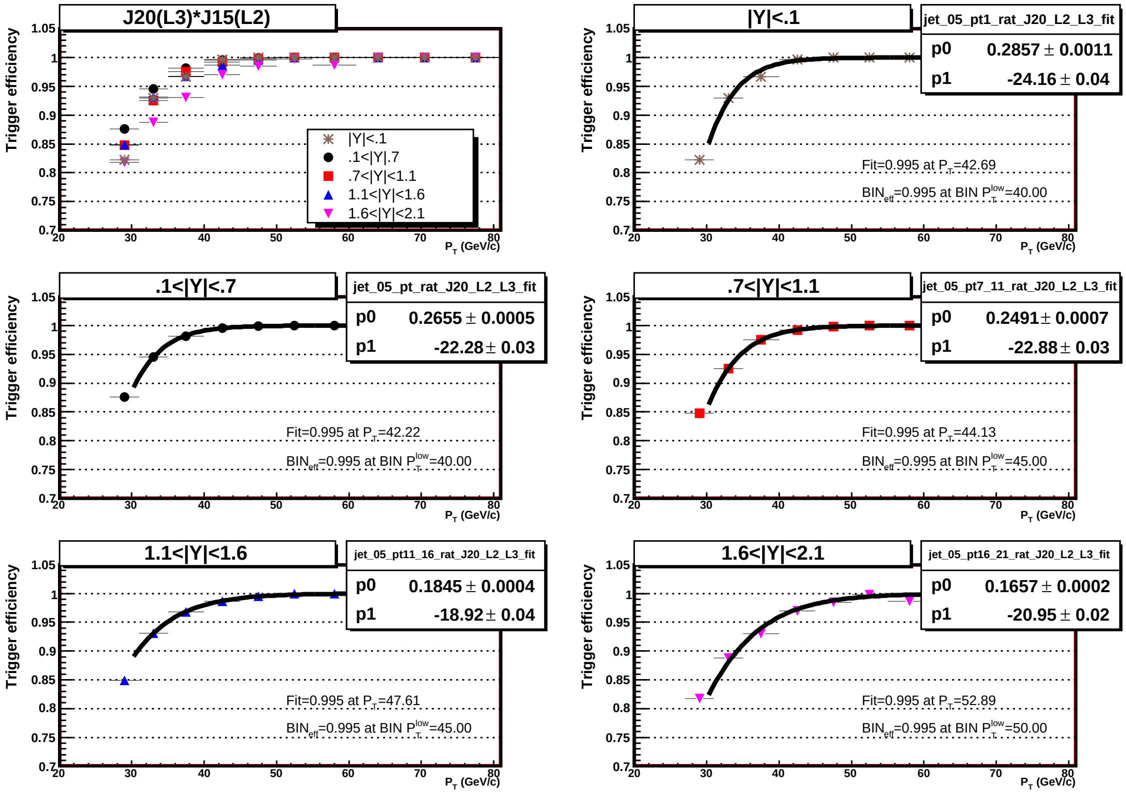


Figure A–4. Trigger efficiencies as a function of jet P_T for the jet20 trigger. L2 and L3 turn-on curves have been combined to calculate the efficiency of the jet20 trigger.

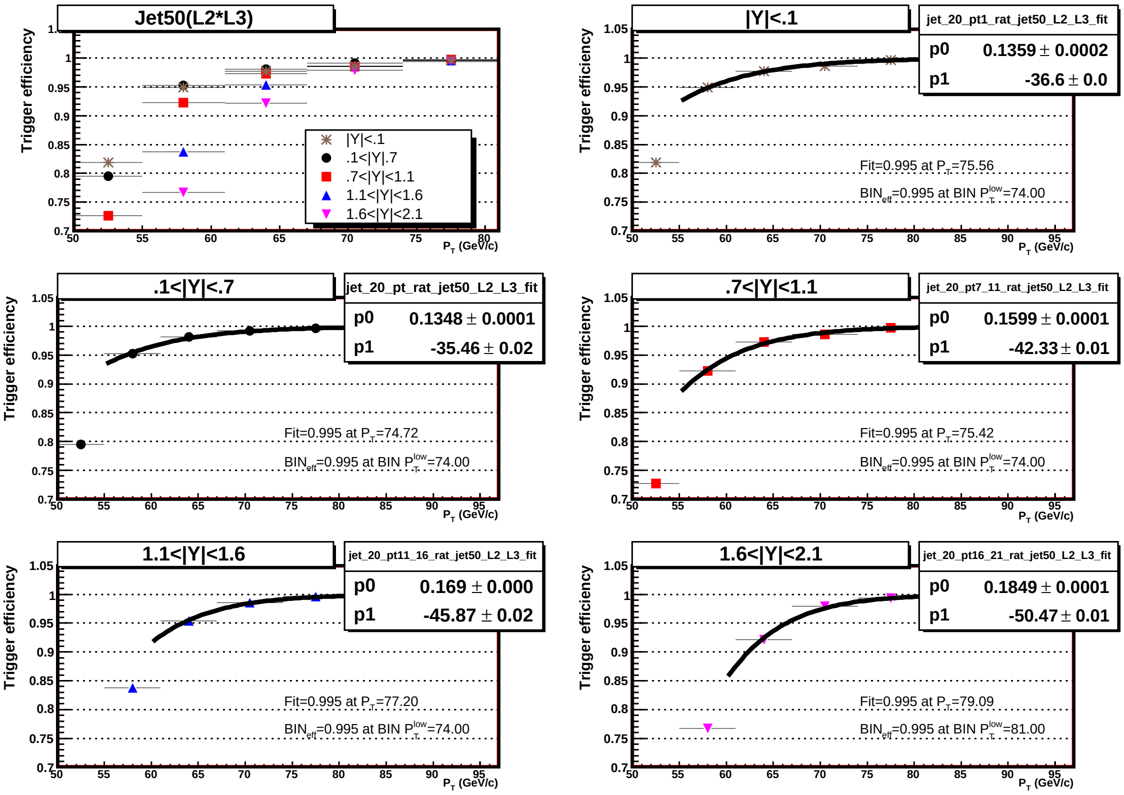


Figure A–5. Trigger efficiencies as a function of jet P_T for the jet50 trigger. L2 and L3 turn-on curves are combined to calculate the efficiency of the jet50 trigger.

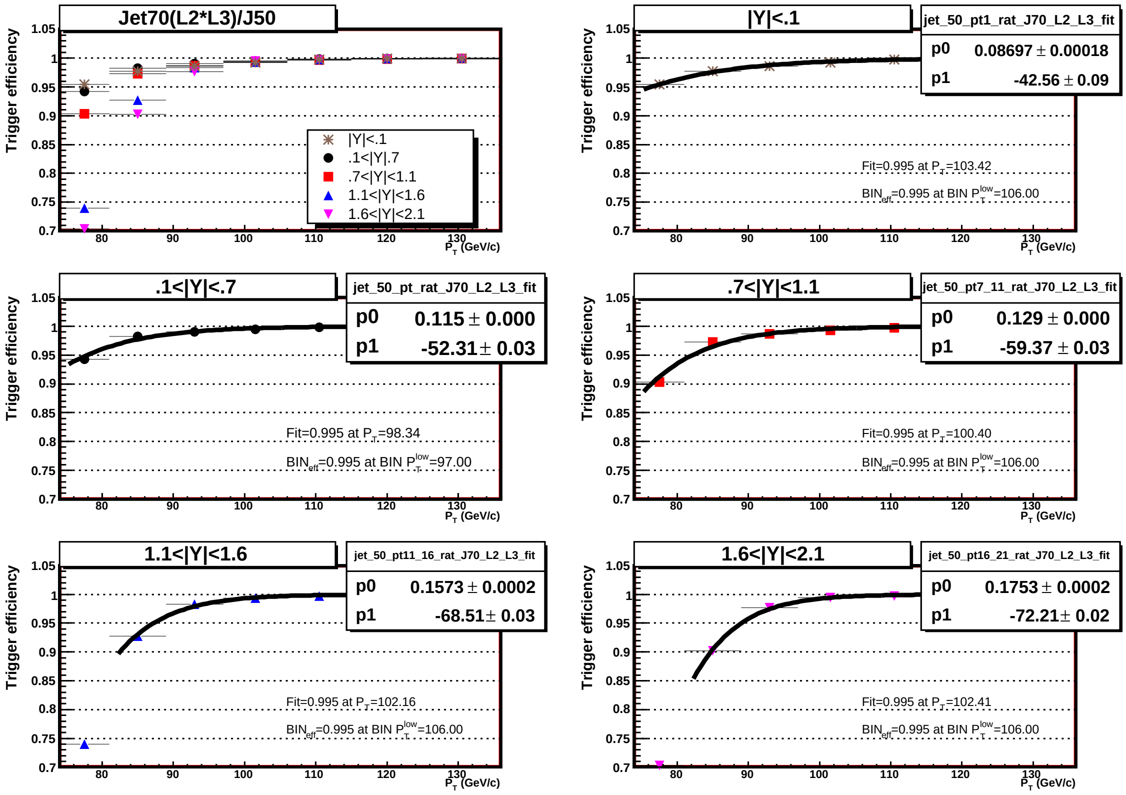


Figure A–6. Trigger efficiencies as a function of jet P_T for the jet70 trigger. L2 and L3 turn-on curves are combined to calculate the efficiency of the jet70 trigger.

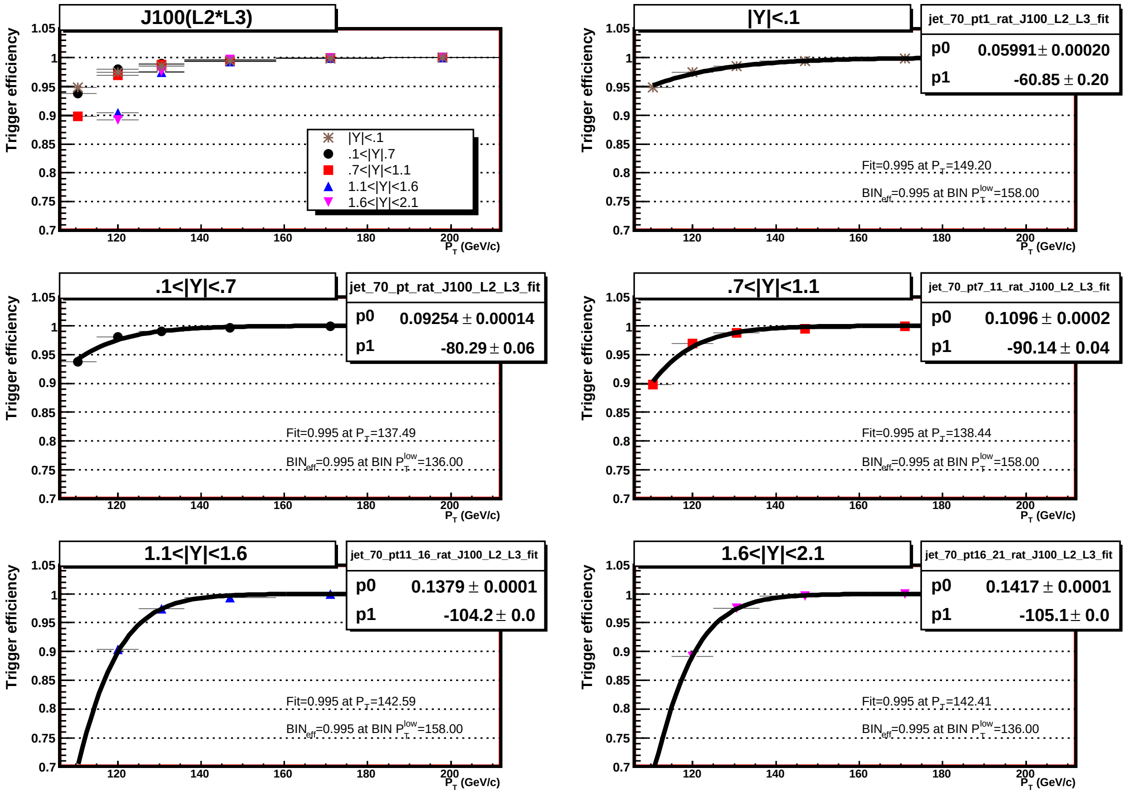
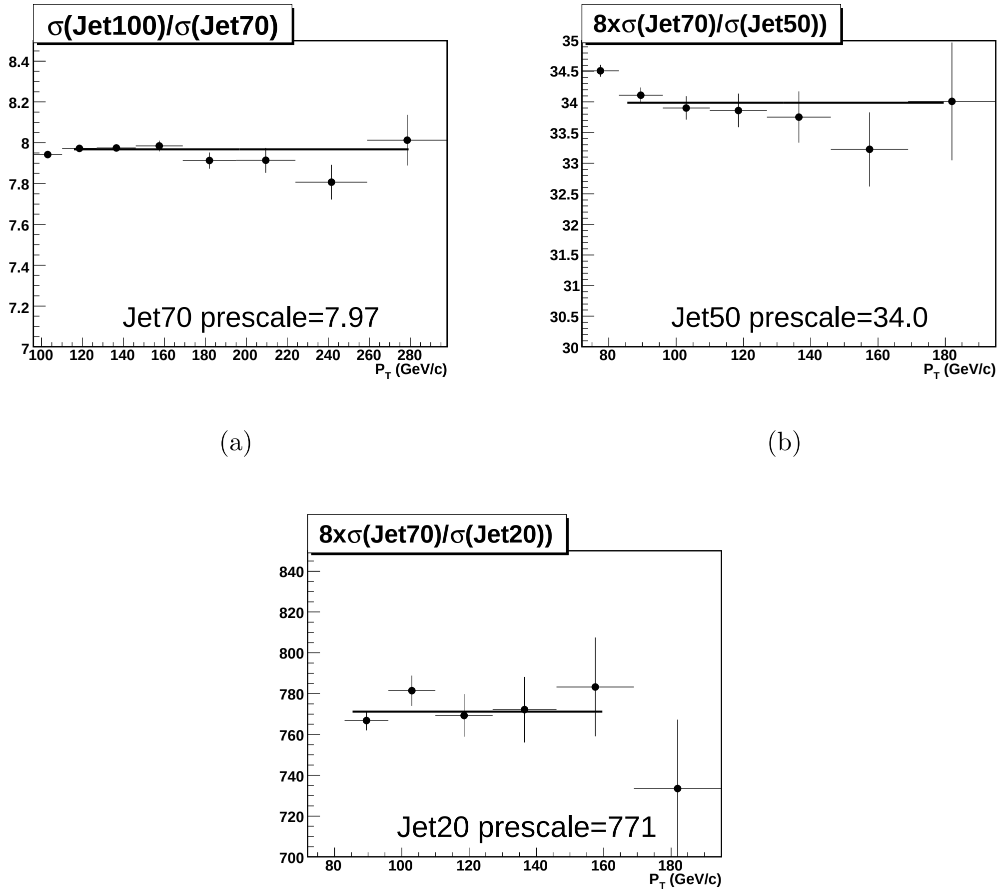


Figure A–7. Trigger efficiencies as a function of jet P_T for the jet100 trigger. L2 and L3 turn-on curves are combined to calculate the efficiency of the jet100 trigger.



(a) (b)

(c)

Figure A–8. Cross section ratio of the jet100 to jet70 data samples, A–8(a); the ratio of the jet70 to jet50 samples multiplied by 8, A–8(b); and the ratio of the jet70 to jet20 samples multiplied by 8, A–8(c). The fits are consistent with the jet70, jet50 and jet20 prescales obtained from the luminosity weighting method to better than 1%.

APPENDIX B
RELATIVE CORRECTIONS

Relative jet energy corrections are applied to correct jets for any variation in the energy response as a function of detector-η (η_d). In Run II at CDF, the relative corrections are derived using P_T balancing of the leading two jets (ΔP_T) [55, 75]. A **trigger** jet is required to be in the well understood central region of the calorimeter ($0.2 < |\eta_d| < 0.6$). The other jet in the event is defined as the **probe** jet. The probe jet is assigned randomly if both jets are found in the central region. The P_T fraction ($\Delta P_T f$) is then defined as:

$$\Delta P_T f \equiv \frac{\Delta P_T}{P_T^{ave}} = \frac{P_T^{probe} - P_T^{trigger}}{(P_T^{probe} + P_T^{trigger})/2} \tag{B–1}$$

The correction factor (β) that will, on average, equate P_T^{probe} to $P_T^{trigger}$ is:

$$\beta \equiv \frac{P_T^{probe}}{P_T^{trigger}} = \frac{2 + \langle \Delta P_T f \rangle}{2 - \langle \Delta P_T f \rangle} \tag{B–2}$$

At CDF, the relative corrections have been derived using the generation 5.3 data (data taken before the 2004 shutdown) with jets clustered by the JetClu algorithm. Here the effectiveness of these corrections applied to generation 5.3 jets clustered using the Midpoint algorithm is studied. Both algorithms form jets by grouping together calorimeter towers which are nearby in angle. The exact prescription for grouping the towers is defined by the algorithm. The algorithms are defined in Chapter 4 of this document.

B.1 Event selection for relative correction studies

The relative corrections were derived from five data samples, ST5 (single-tower 5), jet20, 50, 70, and 100. This data was taken from February 25, 2002 to August

22, 2004. For the data and MC comparisons, dijet MC samples generated with PYTHIA tune A and processed with the 5.3.3 release CDF simulation are used.

The following event selection was applied to the data used in this study unless stated otherwise:

- The good run list for “QCD no silicon” was applied for good run selection[1] .
- N$_{jets} \geq 2$
- Missing E_T significance $\not{E}_T/\sqrt{\sum E_T} < max(2 + 5/400 \times P_T^{j1}, 3)$ where P_T^{j1} is in GeV/c.
- One or more jets in the probe region $(0.2 < |\eta_d| < 0.6)$
- $|Z_{vtx}| < 60cm$
- Azimuthal angle between leading two jets, $\Delta\phi_{12} > 2.7$ radians
- Average P_T of the leading two jets, $P_{T_{min}}^{ave} < P_{T_{j_1,j_2}}^{ave} < P_{T_{max}}^{ave}$
- P_T of the third jet, $P_T^{j_3} < P_{T_{max}}^{j_3}$ (The P_T selection cuts for each jet sample are given in Table B–1)

Table B–1. Selection cuts applied to require a dijet event topology for the five jet data sets used to study the relative corrections.

Data Set	$P_{T_{min}}^{ave}$ (GeV/c)	$P_{T_{max}}^{ave}$ (GeV/c)	$P_T^{j_3}$ (GeV/c)
ST5	15	25	8
Jet20	25	55	8
Jet50	55	75	8
Jet70	75	105	10
Jet100	105	130	10

---

[1] The good run list lists runs where all components of the detector were reported to be working properly. “QCD no silicon” means that the silicon tracking system was not required to be operational

B.2 Relative Corrections with the Midpoint Algorithm

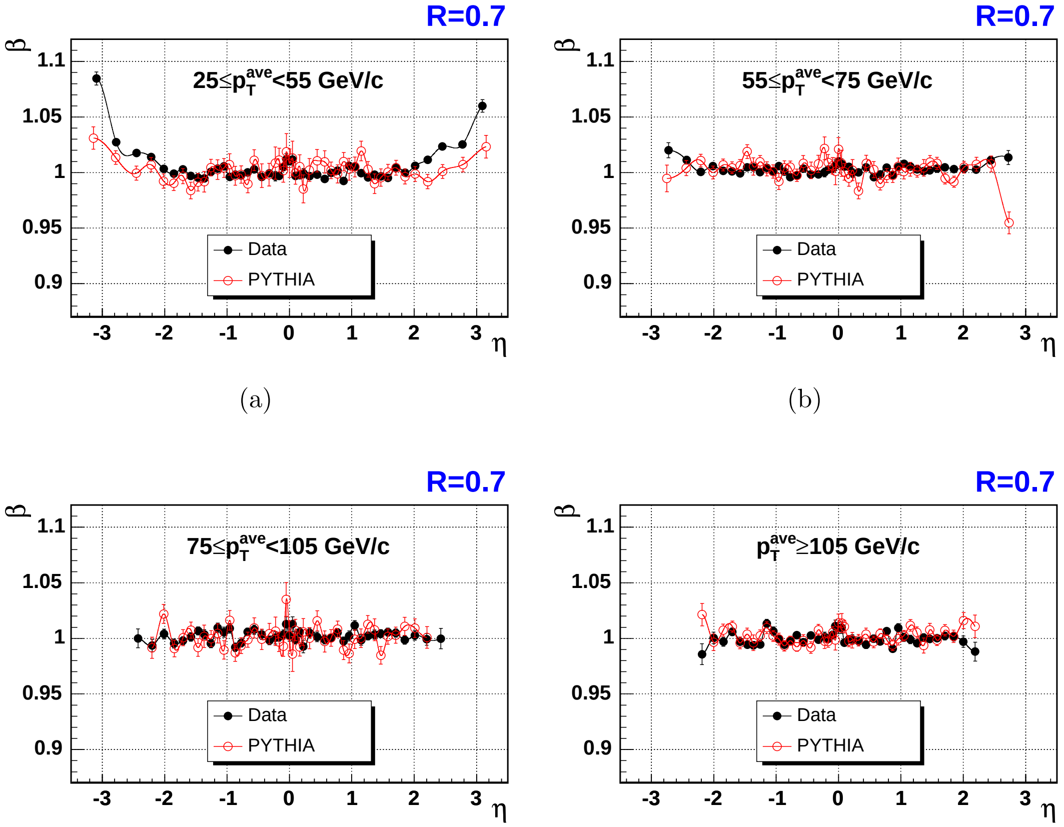
Since JetClu and Midpoint are both cone based algorithms and they use the same merging fraction of 75% by default, it is conceivable that corrections derived from jets clustered with JetClu will be satisfactory for jets clustered with Midpoint with the same cone size without further modifications.

Figure B–1 shows the β versus η_D for jets clustered by Midpoint with $R_{cone} = 0.7$ after the relative corrections derived from jets clustered by JetClu have been applied. The relative corrections are determined to make the β versus η_D flat at 1, therefore deviations of β from one indicates the imperfection of the corrections. It is clear that the corrections derived from jets clustered by JetClu do a reasonable job in jets clustered by Midpoint at $P_T \geq 55$ GeV/c. In the plot for $25 \leq P_T^{ave} < 55$ GeV/c, the β distribution is reasonably flat up to $|\eta_D| = 2$; however, β starts to deviate from one upward with increasing $|\eta_D|$ at $|\eta_D| > 2$. This rising trend at $|\eta_D| > 2$ is more significant in data than in MC.

Figure B–2 shows the β versus η_D plots combining bins to improve statistics after the relative corrections have been applied. Points are shown for the default P_T cut on the third jet (black), and a tighter (red) and looser (blue) cuts. In this figure, the systematic trend of β vs η_D is better presented. In the plots for $25 \leq P_T^{ave} < 55$ GeV/c, while significant variation from one is observed in the high η_D regions, the relative corrections do a reasonable job making dijet balancing distributions flat in the region of $|\eta_D| < 2$. The relative corrections derived from jets clustered by JetClu with $R_{cone} = 0.7$ are valid for jets clustered by Midpoint with $R_{cone} = 0.7$ up to $|\eta_d| = 2$.

The trend observed at $|\eta_D| > 2$ and $P_T < 55$ GeV/c may be attributed to the fact that the physical size of calorimeter towers is smaller in the region of $|\eta_D| > 2$ than in $|\eta_D| < 2$. However, even though the physical size is smaller, individual calorimeter towers occupy larger (Y, ϕ) spaces than towers in $|\eta_D| < 2$.

Jets at $|\eta_D| > 2$ may be sensitive to details of the jet clustering algorithm, especially splitting and merging procedures, and in MC jet clustering in the forward region would be sensitive to the lateral profile tuning of particle showers in the calorimeters. Analyses that use jets much above $|\eta_D| > 2$ will require more studies.



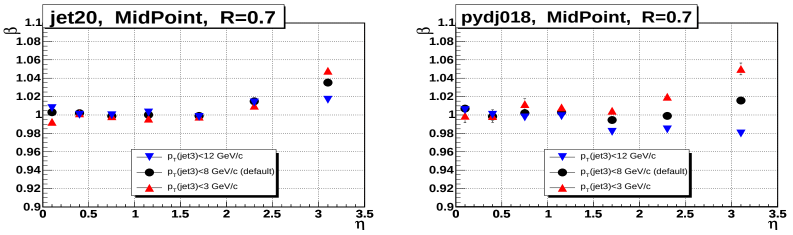
(a)

(b)

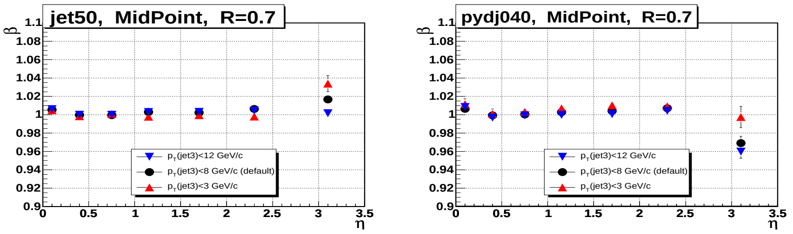
(c)

(d)

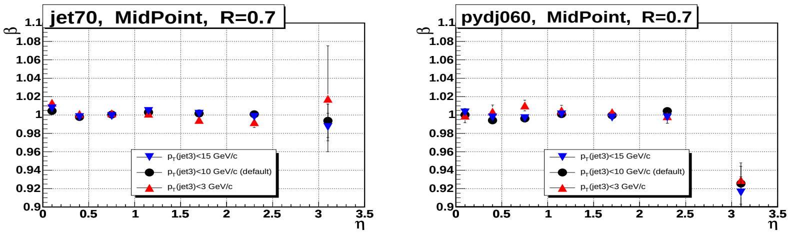
Figure B–1. β versus η_D for jets clustered by the Midpoint algorithm with the cone radius $R_{cone} = 0.7$ after applying the relative corrections derived from jets clustered by JetClu. Shown are for dijets with $25 \leq P_T^{ave} < 55$ GeV/c, $55 \leq P_T^{ave} < 75$ GeV/c, $75 \leq P_T^{ave} < 105$ GeV/c and $P_T^{ave} \geq 105$ GeV/c, respectively.



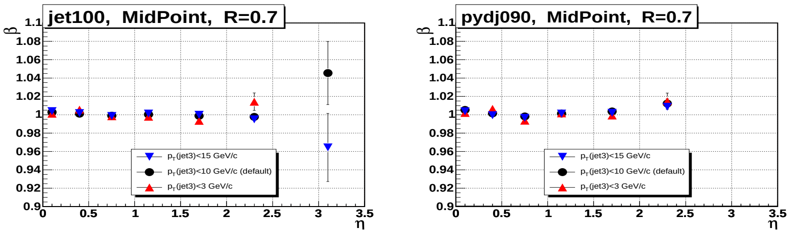
(a) $25 \leq P_T^{ave} < 55$ GeV/c



(b) $55 \leq P_T^{ave} < 75$ GeV/c



(c) $75 \leq P_T^{ave} < 105$ GeV/c



(d) $P_T^{ave} \geq 105$ GeV/c

Figure B–2. β versus η_D for jets clustered by the Midpoint algorithm with $R_{cone} = 0.7$ after applying the relative corrections derived from jets clustered by JetClu. Points are shown for the default P_T cut on the third jet (black), a tighter cut (red), and a looser cut (blue). Shown for data (left) and MC (right) in four different ranges of jet P_T .

APPENDIX C
SIMULATION OF DETECTOR RESPONSE AND RESOLUTION

Before the CDF simulation can be used to derive corrections for the jet data that will unfold detector effects, it must be checked that the simulation correctly describes the CDF detector. Corrections must to be made to the simulation if imperfections are observed. The bisector method and the method of dijet P_T balancing were both used to check that the MC and CDF detector simulation correctly describe the data. Details of these two studies are given below. Cases in which the MC simulation needed to be corrected are emphasized.

C.1 Jet Energy Resolution: Bisector Method

The bisector method is a technique used to quantify the resolution of the jet P_T. Because it may be used on data and MC results after CDF detector simulation it is useful as a check of the reliability of the MC simulation. Events with a two jet topology are used for the study by requiring that two and only two jets have $P_T > 10$ GeV/c. In order to remove the effects of pileup from the data only events with exactly one reconstructed Z vertex are used. The $\widetilde{\not{E}_T}$ cut is also applied to reduce background effects in the data. Also, one jet is required to be in the central region of the detector $(0.1 < |Y| < 0.7)$ and is referred to as the **trigger** jet. The second jet is called the **probe** jet and may be in any other rapidity region $(|Y| < 2.1)$. A coordinate system may be drawn with one axis aligned with the bisector of the two jets. The vector P_T sum of these two jets represents the P_T imbalance in the event. With reference to figure C–1 the following objects are defined for use with the bisector study and are constructed for the jet data and the MC at the detector level:

- $P_T^{MEAN} \equiv (P_{T_1} + P_{T_2})/2$
- $\Delta\phi_{12} \equiv |\phi_1 - \phi_2|$, $\Delta\phi_{12}$ corresponds to the angle between the two jets.
- $\Delta P_{T_\perp} = \pm(P_{T_1} + P_{T_2})\cos(\Delta\phi_{12}/2)$, $\Delta P_{T_\perp}$ is the component of the vector P_T sum of the dijet pair that is oriented along the direction of the bisector. This component of P_T imbalance should be mostly sensitive to physics effects.
- $\Delta P_{T_{||}} = \pm(P_{T_1} - P_{T_2})\sin(\Delta\phi_{12}/2)$ $\Delta P_{T_{||}}$ is the component of the vector P_T sum of the dijet pair that is along the axis perpendicular to the bisector axis. This is the parallel direction in dijet topology because this is the direction closest to the jet direction in the transverse plane. This component of P_T imbalance should be sensitive to physics and detector effects.

After the above variables have been calculated for each event, histograms are filled by bins of P_T mean for the following quantities:

- $\sigma_\perp$ is the rms of the $\Delta P_{T_\perp}$ distribution.
- $\sigma_{||}$ is the rms of the $\Delta P_{T_{||}}$ distribution.
- $\sigma_D \equiv \sqrt{\sigma_{||}^2 - \sigma_\perp^2}$ (σ_D should be most sensitive to detector effects since the physics effects of σ_{perp} are removed. It should also be noted that since σ_D has both jets contributing to it, for a single jet $\sigma = \sigma_D/\sqrt{2}$.)

Ultimately, σ_D is used to compare the jet energy resolution in the real CDF detector and in the CDF detector simulation.

Because the results of the bisector method may be sensitive to the relative corrections, the study has been conducted after the relative corrections have been applied. The detector simulation does a good job reproducing the resolution of jets observed in the central region ($0.1 < |Y| < 0.7$). Figures C–2 - C–5 show the result of the bisector study in all other rapidity regions.

In the two regions where the CDF detector simulation is seen to underestimate the jet energy resolution ($0.7 < |Y| < 1.1$ and $1.6 < |Y| < 2.1$), jets are smeared by an extra Gaussian of width:

$$\Delta\sigma = \sigma_{mc} \times \sqrt{F^2 - 1} \qquad \text{(C–1)}$$

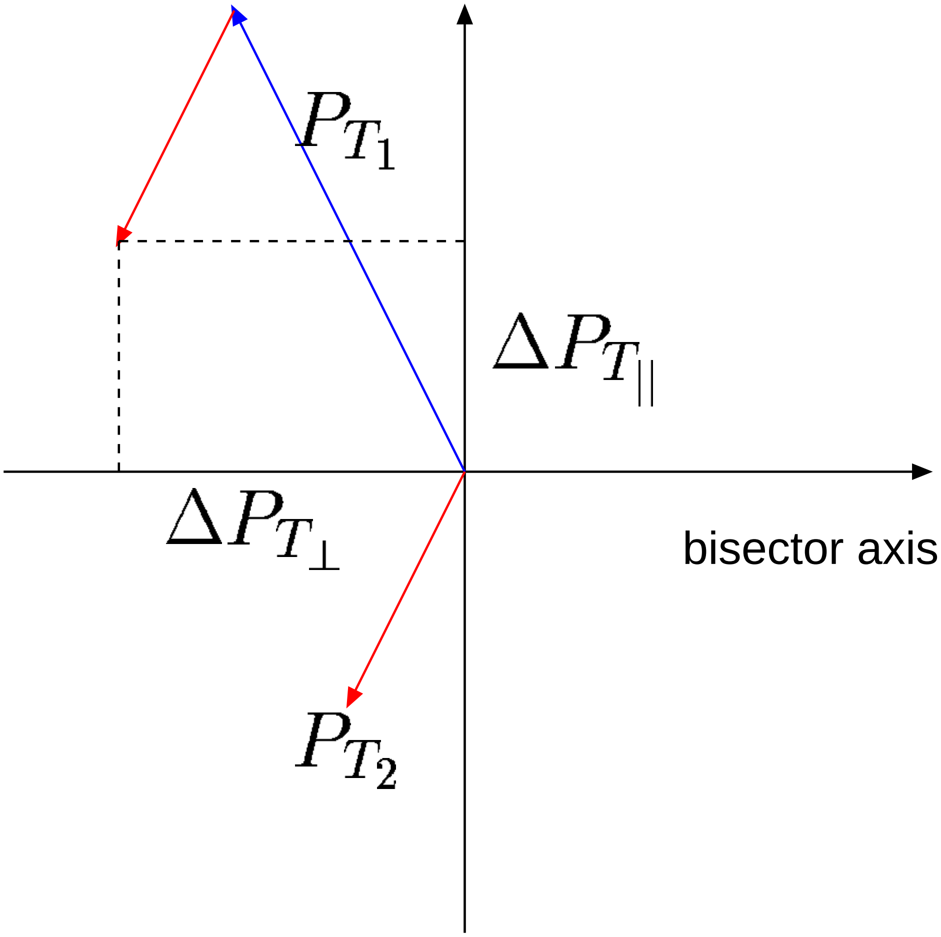


Figure C–1. Bisector variables, described in the text, are labeled in the diagram of the transverse plane. $\Delta P_{T_\perp}$ is defined to be the component of the sum of the jet P_T along the bisector axis.

where σ_{mc} refers to the nominal resolution of the CDF simulation applied to the MC. The resolution of the MC was studied in each rapidity region. For bins of P_T^{had}, the distribution of $P_T^{had} - P_T^{cal}$ (hadron and detector level jets were matched based on separation in $Y - \phi$ space) was fit by a Gaussian and the width of the fit was taken as σ_{mc} for each P_T^{had} bin.

By smearing the MC jets in the regions $0.7 < |Y| < 1.1$ and $1.6 < |Y| < 2.1$ as defined above with $F = 1.10$, the bisector results for the ratio of σ_D between MC and data for both correction methods are consistent with one.

For the region $1.1 < |Y| < 1.6$ and $|Y| < .1$ the simulation over smears the jet P_T. In these regions the jet P_T in the data is smeared in order to quantify the difference between the resolution in the jet data and MC. Smearing the jet data by $F = 1.04$ brings the resolution of the data and MC into agreement in both over smeared regions.

These F values where used to derive bin-by-bin corrections by smearing hadron level jets by a Gaussian of width:

$$\sigma = \sigma_{mc} \times F \tag{C–2}$$

for the case where the MC simulation is not smearing enough, and with a Gaussian of width:

$$\sigma = \sigma_{mc} \times (2 - F) \tag{C–3}$$

for the case where the MC simulation is found to be smearing too much. The difference in the resulting distributions and distributions smeared by only σ_{mc} is used to derive bin-by-bin corrections to the final cross section measurement. The bin-by-bin corrections are small and are shown in figure C–6

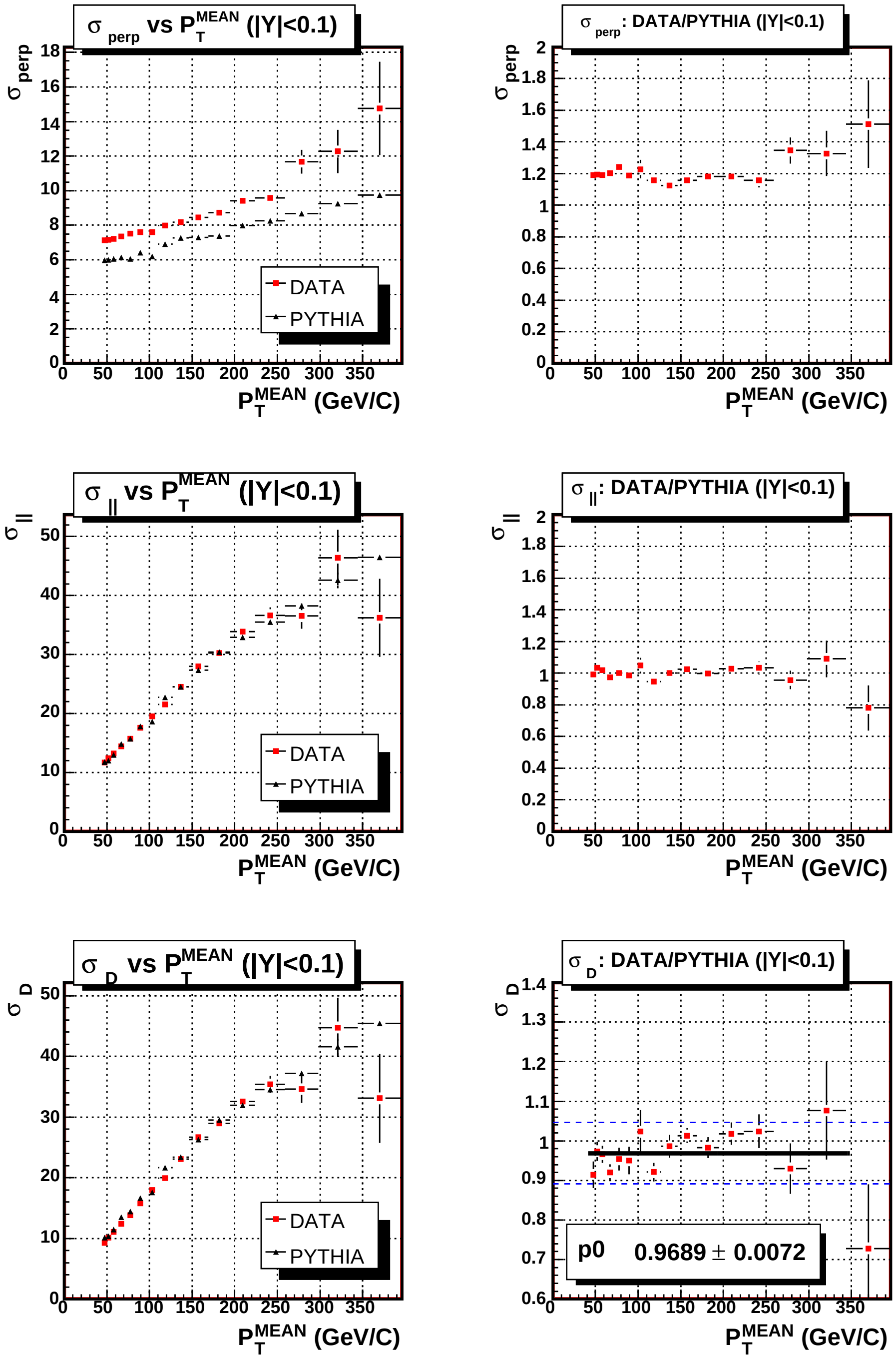


Figure C–2. Results of the bisector study for the rapidity region $|Y| < 0.1$. In this Y region the CDF detector simulation smears the jet energy too much.

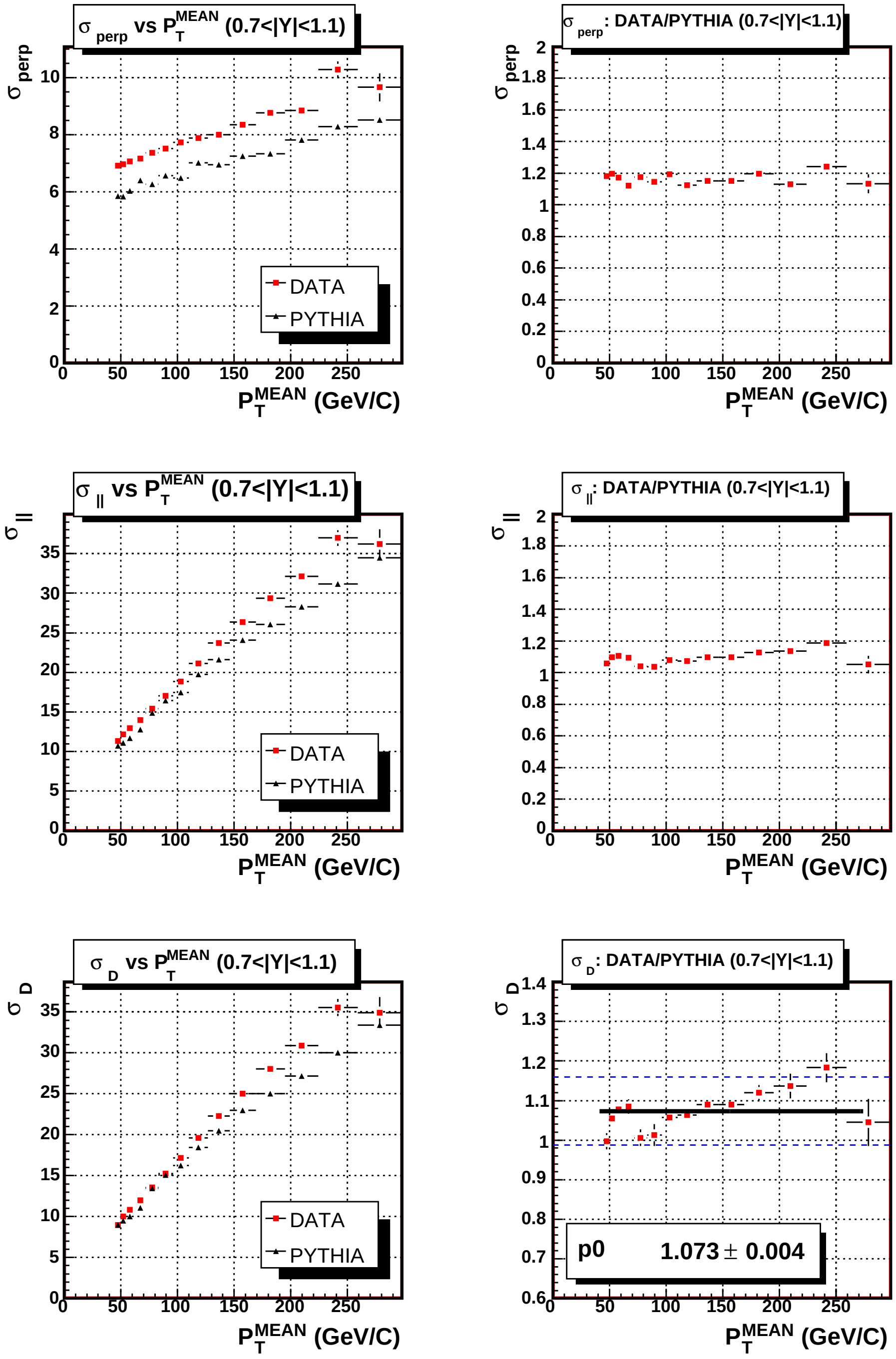


Figure C–3. Results of the bisector study for the rapidity region $0.7 < |Y| < 1.1$. In this Y region the CDF detector simulation does not smear the jet energy enough.

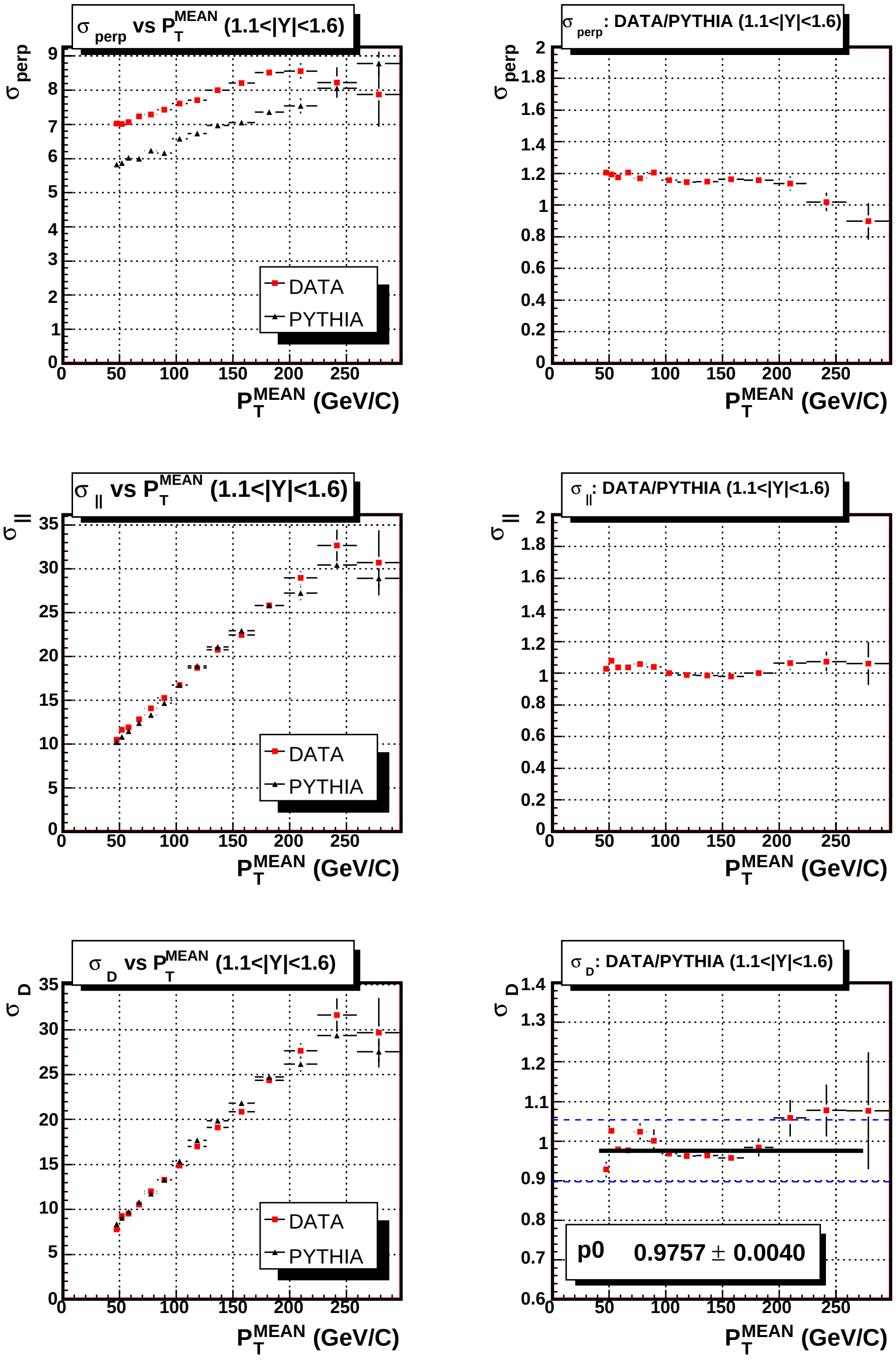


Figure C–4. Results of the bisector study for the rapidity region $1.1 < |Y| < 1.6$. In this Y region the CDF detector simulation smears the jet energy too much.

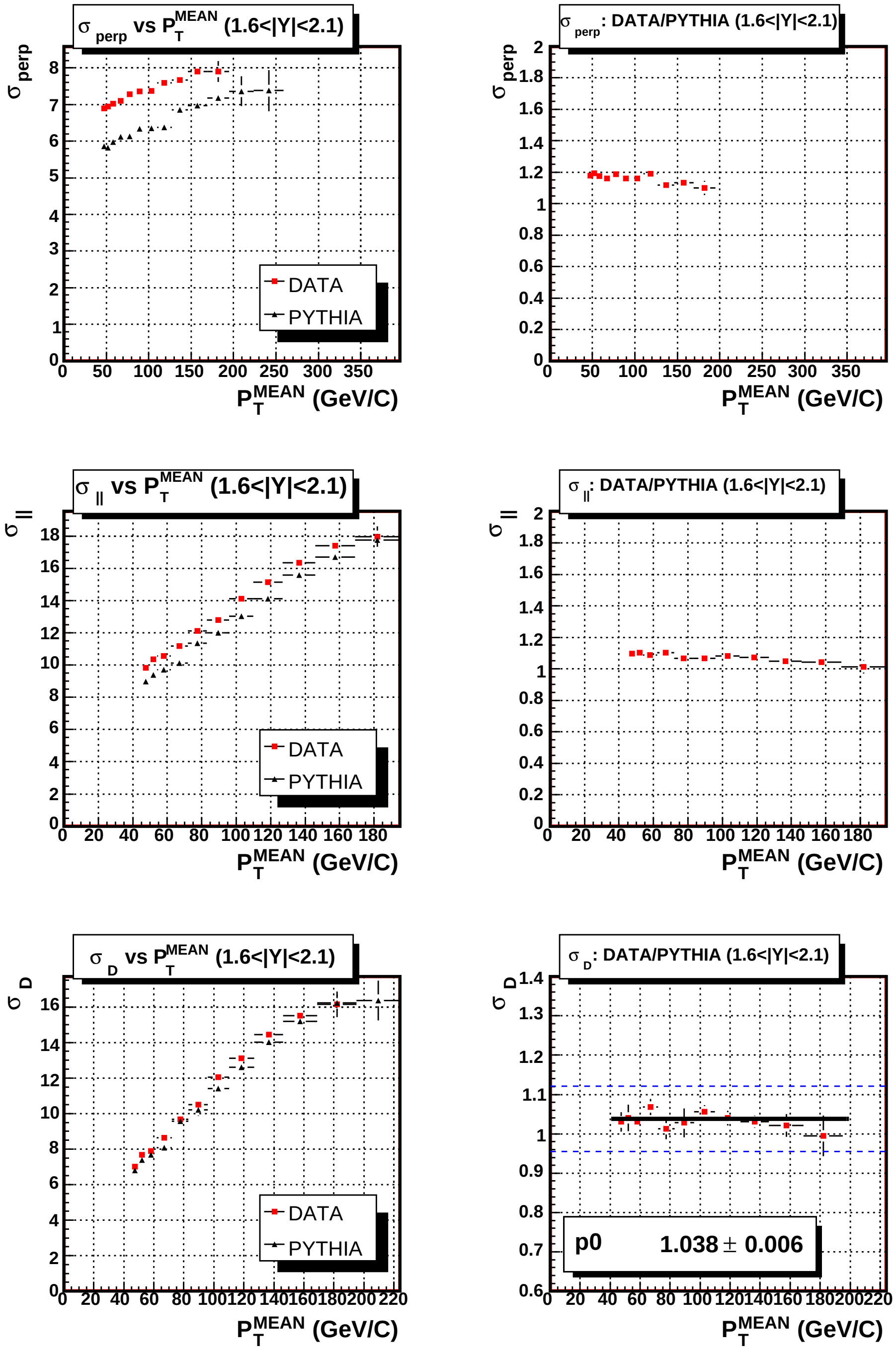


Figure C–5. Results of the bisector study for the rapidity region $1.6 < |Y| < 2.1$. In this Y region the CDF detector simulation does not smear the jet energy enough.

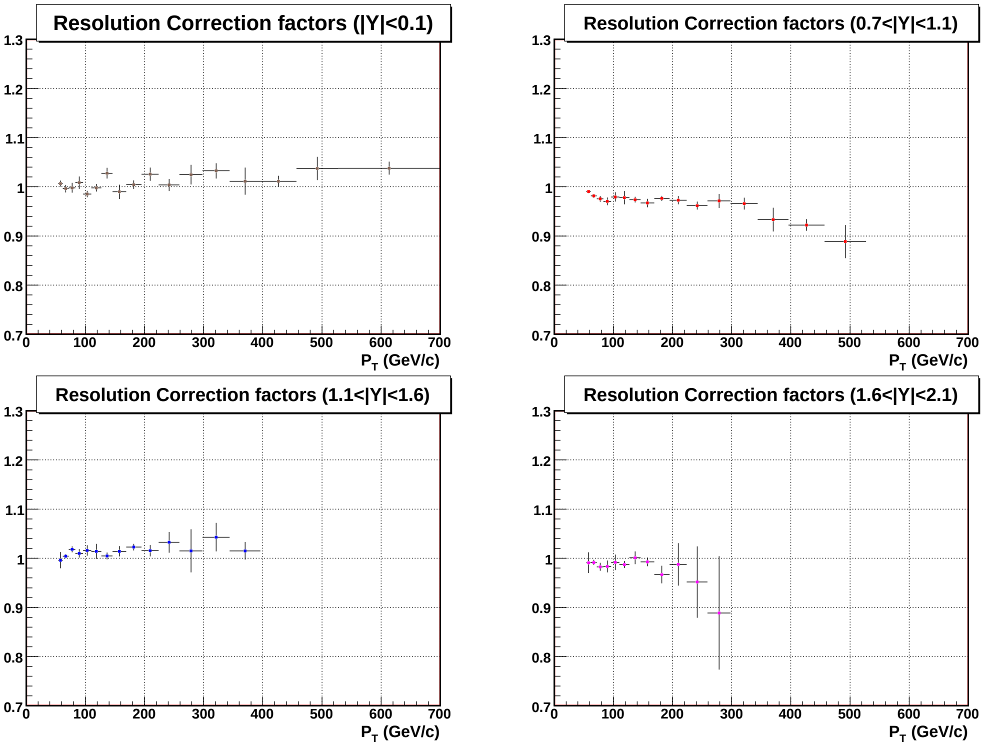


Figure C–6. Modifications to the bin corrections to account for imperfections in simulation of the jet P_T resolution.

C.2 Jet Energy Response: Dijet P_T Balance

subsec:dijet Events with a two jet topology are also used to study dijet P_T balancing. Here, the dijet P_T balance is studied as a function of rapidity for both data and simulated events as another check of the CDF detector simulation. As in the bisector method, events are required to have 2 and only 2 jets with $P_T > 10$ GeV/c. A **trigger** jet is required to be in the well understood central region of the calorimeter ($0.2 < |\eta_d| < 0.6$). The other jet in the event is defined as the **probe** jet. The probe jet is assigned randomly if both jets are found in the central region. The P_T fraction ($\Delta P_T f$) and β are defined in the same way as in the relative corrections of Appendix B, Equations B–1 and B–2.

Because a P_T independent correction is applied for the relative correction, the dijet balance correction to the MC as a function of jet P_T may still be needed in some rapidity regions. Figures C–7 and C–8 show the results of the dijet balance study after the relative corrections are applied. The relative corrections make the data and MC dijet balance agree to 0.5% or better for ($|Y| < 1.1$).

The energy scale in the higher rapidity regions is not as well understood as the central region of the calorimeter, and therefore must be treated with special care. Statistics for the data and PYTHIA samples are limited at high jet P_T in these regions, and this makes it very difficult to check the validity of the simulation at high P_T. In figure C–8 a first order polynomial is used to fit the dijet balance ratio between data and MC at low P_T. The fit is then modified for the higher P_T points. The dotted lines represent what is taken as an additional uncertainty on the jet energy scale in this region due to the limited constraint on the fits for the dijet balance correction. This corresponds to a significant contribution to the systematic uncertainty on the cross section in the high rapidity regions ($\sim 40\%$).

In an effort to improve statistics at high P_T for this study, the third jet cut was relaxed from 10 GeV/c to 10% of the mean P_T of the dijet pair. Unfortunately, the error is not greatly reduced for the relaxed cut. The bin error is proportional to the RMS which increases with a relaxed third jet cut. Even though statistics increased by a factor of two by relaxing the cut, the bin errors are largely unchanged. The same P_T dependence in the ratio of data to MC in the higher rapidity regions is observed.

The corrections to the MC dijet balance for each rapidity region were calculated with the standard 10 GeV/c P_T cut on the third jet, and are summarized in table C–1.

Table C–1. Dijet balance correction applied to the PYTHIA MC simulation for each rapidity region.

Y region	P_T region	Dijet balance correction
$1.1 < \|Y\| < 1.6$	$P_T < 125\mathrm{GeV}/c$	$P_T \rightarrow 1.021 \times P_T - 1.97e^{-4} \times P_T^2$
	$P_T > 125\mathrm{GeV}/c$	$P_T \rightarrow 1.009 \times P_T - 9.87e^{-5} \times P_T^2$
$1.6 < \|Y\| < 2.1$	$P_T < 110\mathrm{GeV}/c$	$P_T \rightarrow 1.018 \times P_T - 2.20e^{-4} \times P_T^2$
	$P_T > 110\mathrm{GeV}/c$	$P_T \rightarrow 1.005 \times P_T - 1.10e^{-4} \times P_T^2$

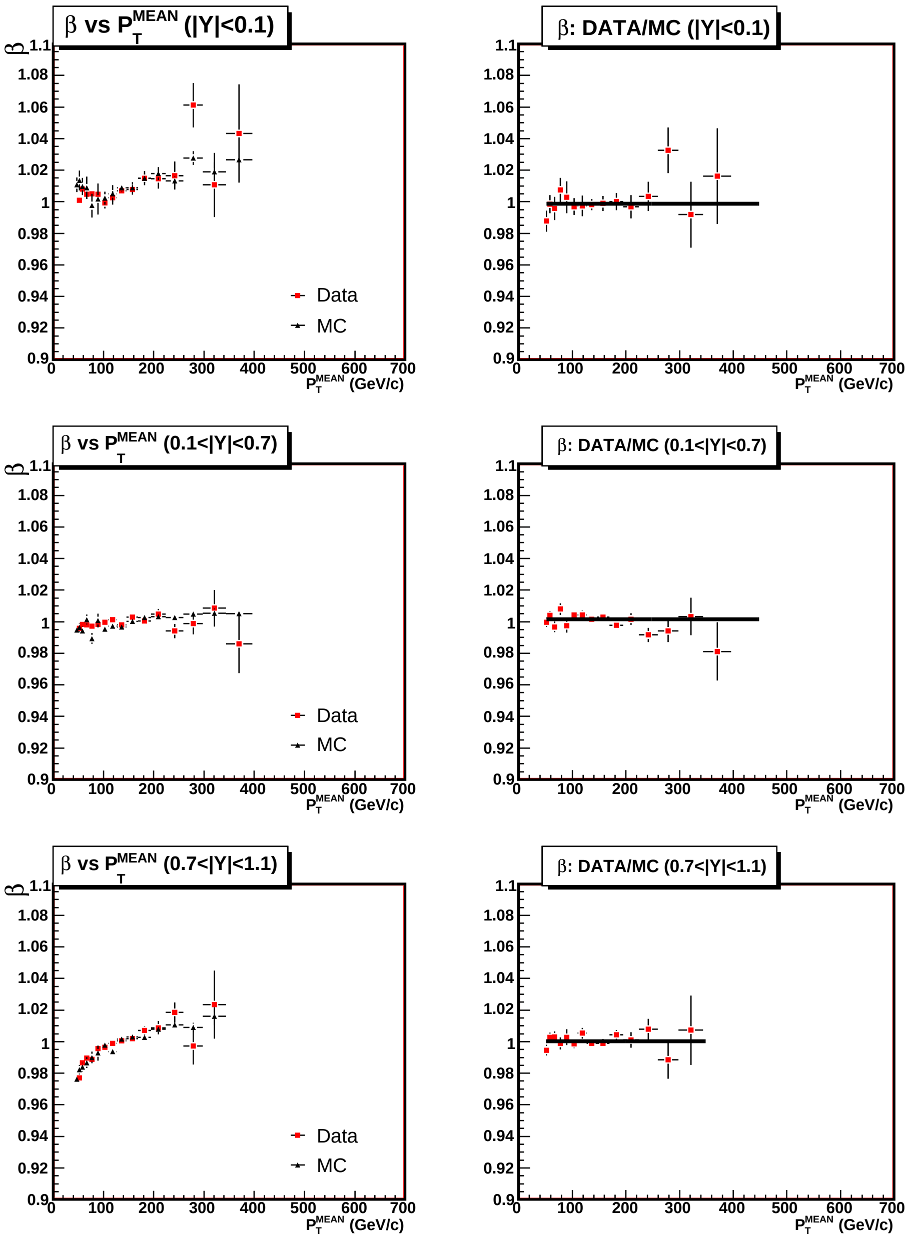


Figure C–7. β as functions of P_T^{MEAN} in data and MC (left) and the data/MC ratio (right) for $|Y| < 0.1$ (top), $0.1 < |Y| < 0.7$ (middle) and $0.7 < |Y| < 1.1$ (bottom) after the the relative corrections (P_T independent) are applied. The data/MC ratios are fit to constant numbers.

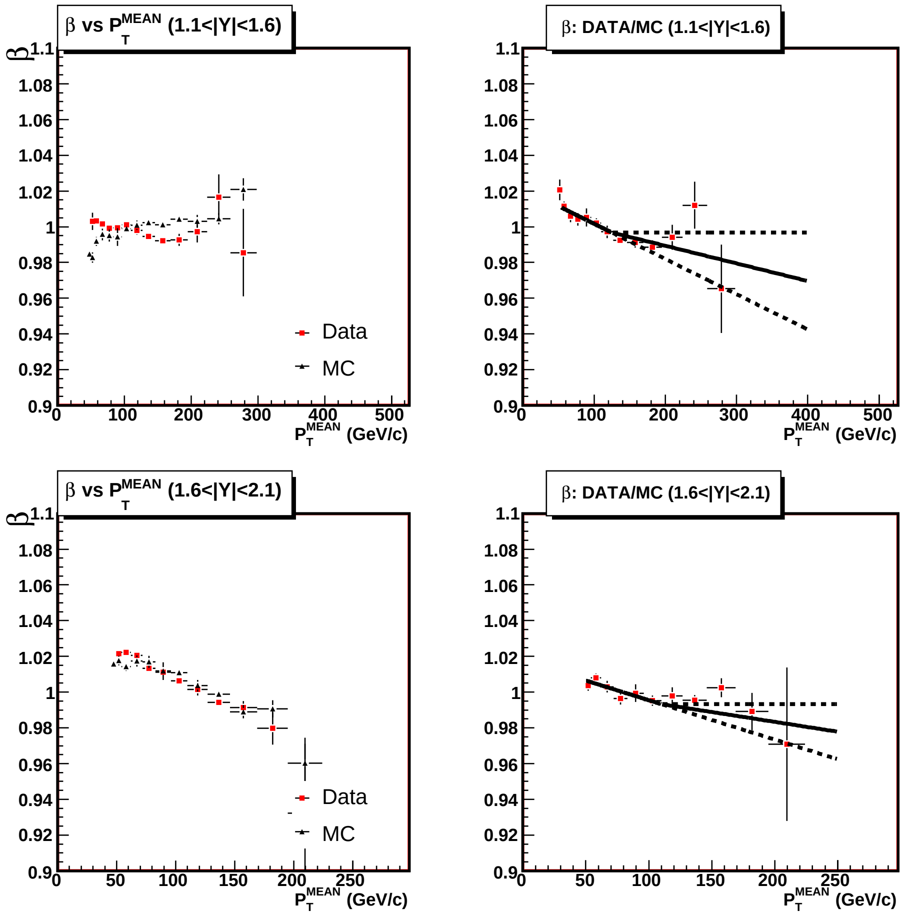


Figure C–8. β as functions of P_T^{MEAN} in data and MC (left) and the data/MC ratio (right) for $1.1 < |Y| < 1.6$ (top) and $1.6 < |Y| < 2.1$ (bottom) after the the relative corrections (P_T independent) are applied. Linear fits are made to the data/MC ratios at low P_T, and the slopes of the fits are modified at high P_T. The dotted lines represent the uncertainty on the fit at high P_T that will be taken as the systematic uncertainty for this correction.

REFERENCES

[1] M. Peskin and D. Schroeder, *An Introduction to Quantum Field Theory* (Reading: Addison-Wesley, 1995).

[2] C. Quigg, *Gauge Theories of the Strong, Weak and Electromagnetic Interactions* (Boulder: Westview Press, 1983).

[3] J. Donoghue, E. Golowich, and B. Holstein, Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. **2**, 1 (1992).

[4] R. Field, *Applications of Perturbative QCD* (Redwood City: Addison-Wesley, 1989).

[5] F. Halzen and A. Martin, *Quarks and Leptons* (New York: Wiley, 1984).

[6] R. Ellis, J. Stirling, and B. Webber, *QCD and Collider Physics* (Cambridge: University Press, 1996), pp. 1–435.

[7] J. Campbell, J. Huston, and J. Stirling, Hard Interactions of Quarks and Gluons: a Primer for LHC Physics, 2006.

[8] W. Yao *et al.*, (Particle Data Group Collaboration), J. Phys. **G33**, 1 (2006).

[9] QCD Figures Available Online, http://theory.fnal.gov/people/ellis/BookFigs/Figs.html (last accessed Oct. 2006).

[10] J. Bjorken and E. Paschos, Phys. Rev. **185**, 1975 (1969).

[11] S. Drell and T. Yan, Ann. Phys. **66**, 578 (1971).

[12] F. Bloch and A. Nordsieck, Phys. Rev. **52**, 54 (1937).

[13] T. Kinoshita, J. Math. Phys. **3**, 650 (1962).

[14] T. Lee and M. Nauenberg, Phys. Rev. **133**, B1549 (1964).

[15] A. Affolder *et al.*, (CDF Collaboration), Phys. Rev. **D65**, 092002 (2002).

[16] J. Pumplin, Phys. Rev. **D57**, 5787 (1998).

[17] R. Field, (CDF Collaboration), Acta Phys. Polon. **B36**, 167 (2005).

[18] R. Field and C. Group, PYTHIA Tune A, HERWIG, and JIMMY in Run 2 at CDF, 2005, contributed to the HERA-LHC workshop in Geneva, CH.

[19] A. Cruz, Ph.D. thesis, University of Florida, 2005.

[20] Jet energy corrections group at CDF public web page, http://www-cdf.fnal.gov/physics/new/top/public/jets/cdfpublic.html (last accessed Oct. 2006).

[21] Fermilab Accelerator Division: Accelerator Concepts Rookie Book, http://www-bdnew.fnal.gov/operations/rookie_books/rbooks.html (last accessed Oct. 2006).

[22] F. Abe *et al.*, (CDF Collaboration), Nucl. Instr. Meth. **A271**, 387 (1988).

[23] P. Lukens *et al.*, (CDF II collaboration), The CDF IIb Detector:Technical Design Report, 2003, fermilab-TM-2198.

[24] D. Acosta, S. Klimenko, J. Konigsberg, A. Korytov, G. Mitselmakher, V. Necula, A. Nomerotsky, A. Pronko, A. Sukhanov, A. Safonov, D. Tsybychev, S. Wang, and M. Wong, Nucl. Instrum. Meth. **A494**, 57 (2002).

[25] B. Martin and G. Shaw, *Particle Physics* (Chichester: Wiley, 1992).

[26] S. Ellis, J. Huston, and M. Tonnesmann, eConf **C010630**, P513 (2001).

[27] S. Ellis and D. Soper, Phys. Rev. **D48**, 3160 (1993).

[28] *Research directions for the decade: 1990 Summer Study on High Energy Physics*, edited by E. L. Berger (Singapore: World Scientific, 1992).

[29] C. Blazey, J. Dittmann, S. Ellis, D. Elvira, K. Frame, S. Grinstein, R. Hirosky, R. Piegaia, H. Schellman, R. Snihur, V. Sorin, and D. Zeppenfeld, in *QCD and weak boson physics in Run II* (Batavia: Fermilab, 2000).

[30] M. Albrow *et al.*, (TeV4LHC QCD Working Group), Tevatron-for-LHC Report of the QCD Working Group, 2006, hep-ph/0610012.

[31] S. Berman and M. Jacob, Phys. Rev. Lett. **25**, 1683 (1970).

[32] S. Berman, J. Bjorken, and J. Kogut, Phys. Rev. **D4**, 3388 (1971).

[33] R. Feynman, *Photon Hadron Interactions* (New York: Benjamin, 1972).

[34] G. Arnison *et al.*, (UA1 Collaboration), Phys. Lett. **B123**, 115 (1983).

[35] M. Banner *et al.*, (UA2 Collaboration), Phys. Lett. **B118**, 203 (1982).

[36] A. Affolder *et al.*, (CDF Collaboration), Phys. Rev. **D64**, 032001 (2001).

[37] D. Stump, J. Huston, J. Pumplin, W. Tung, H. Lai, S. Kuhlmann, and J. Owens, JHEP **10**, 046 (2003).

[38] E. Eichten, K. Lane, and M. Peskin, Phys. Rev. Lett. **50**, 811 (1983).

[39] D. Gross and F. Wilczek, Phys. Rev. **D8**, 3633 (1973).

[40] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, Phys. Lett. **B47**, 365 (1973).

[41] J. Huston, S. Kuhlmann, H. Lai, F. Olness, J. Owens, D. Soper, and W. Tung, Phys. Rev. **D58**, 114034 (1998).

[42] F. Abe *et al.*, (CDF Collaboration), Phys. Rev. Lett. **77**, 438 (1996).

[43] B. Abbott *et al.*, (D0 Collaboration), Phys. Rev. Lett. **86**, 1707 (2001).

[44] J. Huston, E. Kovacs, S. Kuhlmann, H. Lai, J. Owens, D. Soper, and W. Tung, Phys. Rev. Lett. **77**, 444 (1996).

[45] W. Giele and S. Keller, Phys. Rev. **D58**, 094023 (1998).

[46] A. Abulencia *et al.*, (CDF Run II Collaboration), (2005), recently accepted for publication in PRD-RC, hep-ex/0512020.

[47] A. Bhatti, F. Chlebana, G. Flanagan, C. Group, K. Hatakeyama, J. Huston, and G. Latina, internal CDF publication 8192, 2006 (unpublished).

[48] B. Winer, Int. J. Mod. Phys. **A16S1C**, 1169 (2001).

[49] W. Sakumoto, internal CDF publication 8318, 2006 (unpublished).

[50] T. Sjostrand, P. Eden, C. Friberg, L. Lonnblad, G. Miu, S. Mrenna, and E. Norrbin, Comput. Phys. Commun. **135**, 238 (2001).

[51] G. Corcella, I. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, H. Seymour, and B. Webber, JHEP **01**, 010 (2001).

[52] H. Lai *et al.*, (CTEQ Collaboration), Eur. Phys. J. **C12**, 375 (2000).

[53] R. Brun, F. Bruyant, M. Maire, A. McPherson, and P. Zanarini, GEANT3, 1987, cern-DD/EE/84-1.

[54] G. Grindhammer, M. Rudowicz, and S. Peters, Nucl. Instrum. Meth. **A290**, 469 (1990).

[55] A. Bhatti, D. Ambrose, J. Arguin, A. Barbaro-Galtieri, H. Budd, Y. Chung, K. Chung, B. Cooper, C. Currat, M. D'Onofrio, *et al.*, Nucl. Instrum. Meth. **A566**, 375 (2006).

[56] P. Bagnaia *et al.*, (UA2 Collaboration), Phys. Lett. **B144**, 283 (1984).

[57] G. Flanagan, Ph.D. thesis, University of Michigan, 2004.

[58] R. Field, 2002, presentation at Fermilab ME/MC tuning workshop in Batavia, IL.

[59] S. Ellis, Z. Kunszt, and D. Soper, Phys. Rev. Lett. **64**, 2121 (1990).

[60] W. Giele, E. Glover, and D. Kosower, Nucl. Phys. **B403**, 633 (1993).

[61] T. Kluge, K. Rabbertz, and M. Wobisch, fastNLO: Fast pQCD calculations for PDF fits, 2006, contribution to TeV4LHC workshop in Batavia, IL.

[62] FastNLO Web Interface, http://hepforge.cedar.ac.uk/fastnlo/ (last accessed Oct. 2006).

[63] Z. Nagy, Phys. Rev. **D68**, 094002 (2003).

[64] Z. Nagy, Phys. Rev. Lett. **88**, 122003 (2002).

[65] F. Abe *et al.*, (CDF Collaboration), Phys. Rev. Lett. **80**, 2773 (1998).

[66] M. Whalley, D. Bourilkov, and C. Group, The Les Houches accord PDFs (LHAPDF) and LHAGLUE, 2005, contributed to the HERA-LHC workshop in Geneva, CH.

[67] J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, J. Kalk, H. Lai, and W. Tung, Phys. Rev. **D65**, 014013 (2002).

[68] P. Nadolsky and Z. Sullivan, eConf **C010630**, P510 (2001).

[69] Z. Sullivan, Phys. Rev. **D66**, 075011 (2002).

[70] D. Bourilkov, C. Group, and M. Whalley, LHAPDF: PDF use from the Tevatron to the LHC, 2006, contributed to the TeV4LHC workshop in Batavia, IL.

[71] S. Bethke *et al.*, (JADE Collaboration), Phys. Lett. **B213**, 235 (1988).

[72] V. Abazov *et al.*, (D0 Collaboration), Phys. Lett. **B525**, 211 (2002).

[73] A. Abulencia *et al.*, (CDF II Collaboration), Phys. Rev. Lett. **96**, 122001 (2006).

[74] A. Djouadi and S. Ferrag, Phys. Lett. **B586**, 345 (2004).

[75] K. Hatakeyama and C. Group, internal CDF publication 8030, 2006 (unpublished).

BIOGRAPHICAL SKETCH

Robert Craig Group was born in Columbia, SC, on May 28, 1977. He grew up in rural low-country South Carolina in the town of Allendale, and attended high school in Barnwell, SC. He has two loving parents, Cindy and Bob, and a slightly younger brother, Eric. While at Barnwell High School Craig dated Nicole L. Tyler, and lettered in academics and golf. Craig attended Erskine College in Due West, SC. He served as president of the mathematics honor society (Sigma Pi Sigma) and also vice-president and then president of the Philomathean Literary Society. Craig earned a B.A. degree in physics with a minor in mathematics from Erskine. While a Junior at Erskine, Craig engaged in a summer research project for undergraduates at Florida State University studying experimental nuclear physics. This experience contributed to his choice to attend Florida State for graduate school directly after his graduation from Erskine College.

Craig spent two summers at the Fermi National Accelerator Laboratory (FNAL) and studied high energy particle physics while at FSU. He received his MS in physics from FSU and then took time off from school to be reunited with Nicole. They were married in August of 2002. One week later, Craig returned to graduate school, this time at the University of Florida. At UF, Craig took the upper level graduate courses his first two semesters. After completion of the course requirements, he engaged himself completely with research for the next three years. His hard work and determination led to this dissertation. Craig will continue his research at FNAL as a research associate next year. He looks forward to the challenges and rewards that life will surely provide.