



מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE

Thesis for the degree
Doctor of Philosophy

עבודת גמר (תזה) לתואר
דוקטור לפילוסופיה

Submitted to the Scientific Council of the
Weizmann Institute of Science
Rehovot, Israel

מוגשת למועצה המדעית של
מכון ויצמן למדע
רחובות, ישראל

By
Raz Alon

מאת
רז אלון

מבנה של סילונים אנרגטיים
בעלי מסה גבוהה
Substructure of
Highly Boosted Massive Jets

Advisor:
Prof. Ehud Duchovni

מנחה:
פרופ. אהוד דוכובני

October 2012

חשון התשע"ג

Contents

1	Introduction	6
1.1	The Standard Model	7
1.2	Quantum Chromodynamics	9
1.3	The Tevatron	14
1.4	The CDF Detector	15
1.5	The Large Hadron Collider	17
1.6	The ATLAS Detector	18
2	Methods	20
2.1	Jet Algorithms	20
2.1.1	Midpoint	21
2.1.2	Midpoint with Search Cones	23
2.1.3	<i>anti-k_t</i>	24
2.2	Event Selection	25
2.2.1	Data and Monte Carlo Samples	25
2.2.2	Selection Criteria and Jet Calibration	26
2.2.3	Expected Sources of Events	29
2.2.4	Characteristics of $t\bar{t}$ Events and Top Rejection	31
2.3	Pileup Correction	34
2.4	Unfolding	36
2.5	Jet Substructure Observables	36
2.5.1	Jet Mass	36
2.5.2	Angularity	38
2.5.3	Planar Flow	39
3	Results	40

3.1	Pileup Correction	40
3.2	Jet Substructure Observables	43
3.2.1	Jet Mass	43
3.2.2	Angularity	46
3.2.3	Planar Flow	48
3.3	Systematic Uncertainties	48
3.3.1	Internal Jet Energy Scale	48
3.3.2	Energy Flow from Multiple Interactions	54
3.3.3	Uncertainties on the PYTHIA Prediction	55
3.3.4	Unfolding Related Uncertainties	55
4	Search for Boosted Top Quarks	56
4.1	Methods	56
4.2	Results	59
4.3	Uncertainties	60
5	Discussion	62
6	References	65

List of Abbreviations and Symbols

	LAr Liquid Argon.
E_T Transverse Energy.	LHC Large Hadron Collider.
S_{MET} Missing Transverse Energy Significance.	MC Monte Carlo.
\cancel{E}_T Missing Transverse Energy.	MI Multiple Interactions.
p_T Transverse Momentum.	NLO Next-to-Leading Order.
ATLAS A Toroidal Lhc ApparatuS.	NNLO Next-to-Next-to-Leading Order.
BSM Beyond the Standard Model.	PDF Parton Distribution Function.
CDF Collider Detector at Fermilab.	pQCD perturbative QCD.
CERN European Center for Nuclear Research.	QCD Quantum Chromodynamics.
COT Central Outer Tracker.	RGE Renormalization Group Equation.
EM Electro Magnetic.	SM Standard Model.
GRL Good Runs List.	SSB Spontaneous Symmetry Breaking.
HAD Hadronic.	UE Underlying Event.
IR Infra Red.	
JES Jet Energy Scale.	VEV Vacuum Expectation Value.

Abstract

Modern particle accelerators enable researchers to study new high energy frontiers which have never been explored before. This realm opens possibilities to further examine known fields such as Quantum Chromodynamics. In addition, it allows searching for new physics and setting new limits on the existence of such.

This study examined the substructure of highly boosted massive jets measured by the CDF II detector. Events from 1.96 TeV proton-antiproton collisions at the Fermilab Tevatron Collider were collected out of a total integrated luminosity of 5.95 fb^{-1} . They were selected to have at least one jet with transverse momentum above $400\text{ GeV}/c$. The jet mass, angularity, and planar flow were measured and compared with predictions of perturbative Quantum Chromodynamics, and were found to be consistent with the theory. A search for boosted top quarks was conducted and resulted in an upper limit on the production cross section of such top quarks.

תקציר

מאיצי חלקיקים מודרניים מאפשרים לחקור גבולות אנרגיה חדשים אשר טרם נמדדו בעבר. אופק אנרגיה חדש זה מאפשר להמשיך וללמוד תחומי פיזיקה ידועים, כגון דינמיקת צבע קוונטית (Quantum Chromodynamics). בנוסף, מתאפשרים החיפוש אחר פיזיקה חדשה והצבת חסמים על קיומה.

מחקר זה בדק את המבנה הפנימי של סילונים (Jets) כבדים ומואצים מאד, אשר נמדדו בגלאי CDF II. מאורעות נאספו מהתנגשויות פרוטונים באנטי פרוטונים באנרגיה של 1.96 TeV במאיץ החלקיקים Tevatron אשר ב-Fermilab מתוך עוצמת הארה כוללת (integrated luminosity) של 5.95 fb^{-1} . מתוכם נבחרו מאורעות המכילים לפחות סילון אחד בעל תנע רוחבי מעל 400 GeV/c. המסה, הזוויתיות (angularity) והזרימה המישורית (planar flow) של סילונים אלו נמדדו. התוצאות הושאו לתחזיות של תאורית דינמיקת צבע קוונטית ונמצאו תואמות. בוצע חיפוש אחר קווארקים מואצים מסוג top והושם חסם עליון לחתך הפעולה ליצירת קווארקים כאלה.

1 Introduction

The study of substructure of jets with high Transverse Momentum (p_T) and a high mass, produced by Quantum Chromodynamics (QCD) processes (see *e.g.* [1, 2, 3] for recent reviews) sheds light on the QCD showering mechanism and provides excellent environment for perturbative QCD (pQCD) studies. Moreover, events containing such jets constitute a dominant background in searches for boosted top quarks, for some of the Higgs boson search channels (*e.g.* Wh , Zh) [4], as well as searches for possible new physics signatures that may give rise to massive boosted jets [5, 6, 7, 8]. At the Large Hadron Collider (LHC), and even at the Tevatron, the decay product of a highly boosted heavy object are strongly collimated and, at high enough p_T , are clustered into a single jet and cannot be resolved by standard jet algorithms. One then has to study the jet's substructure, *i.e.* the energy flow within the jet, in order to identify such jets and separate them from genuine boosted massive QCD jets. This collimation effect is schematically illustrated in Fig. 1 for a semileptonic $t\bar{t}$ event.

Prior to the study presented here, Tevatron results were available for jets with p_T lower than 400 GeV/c [9, 10]. The present Collider Detector at Fermilab (CDF) study [11, 12] is the first to investigate the substructure of massive jets with p_T in excess of 400 GeV/c . It is already followed by recent publications on the substructure of highly boosted massive jets produced at the LHC [13, 14, 15, 16]. The present study makes use of several jet algorithms: Midpoint [17], Midpoint with Search Cones [18], and *anti- k_t* [19], with typical jet sizes of $R = 0.4$, 0.7, and 1.0. A data-driven technique was developed and used to correct the various measured observables for the effect of pileup. This term is used for non-coherent energy deposits caused by multiple hadron collisions, which are a major impediment in jet's substructure studies. Three substructure observables are presently studied and measured; the measured distribution of jet mass is compared with

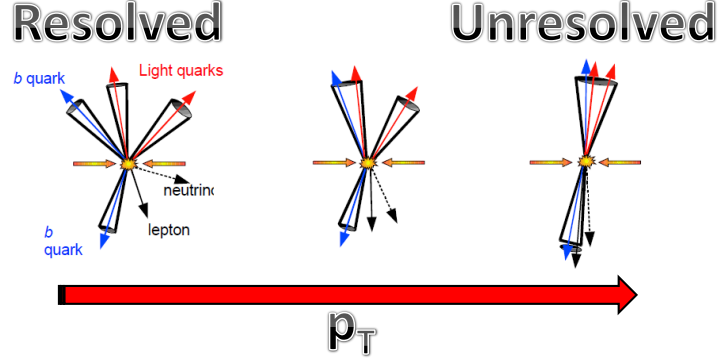


Figure 1: Schematic illustration of the reconstruction of a semileptonic $t\bar{t}$ event as a function of the characteristic p_T . At low p_T the top's decay product are well separated, while at high p_T they are contained in a single boosted heavy jet

theoretical Next-to-Leading Order (NLO) prediction of the jet function [20]. The measured distributions of angularity and planar flow [21] corroborate the hypothesis of two-prong nature of QCD jets. Finally, a dedicated search for boosted top quarks [22] based mostly on the jet mass is conducted.

The following sections give a brief theoretical background for this study.

1.1 The Standard Model

Decades of work of thousands of theoreticians and experimentalists uncovered the fundamental structure of matter; the twelve elementary matter particles, leptons and quarks, and the three fundamental forces that determine their mutual interactions through three types of gauge bosons. The gamut of these efforts is neatly wrapped up in a theory called the Standard Model (SM) [23, 24, 25, 26]. It successfully explains a variety of experimental results without (up to date) a failure.

The Lagrangian of the SM is required to be invariant under transformation of the gauge group $SU(3)_c \times SU(2)_L \times U(1)_Y$. The half-integer-spin fermion particles transform under representations of this gauge group and carry corresponding charges. They interact via vector bosons which transform under the adjoint representation. The renormalizable invariant terms in the Lagrangian result in massless matter particles and gauge bosons. The Higgs mechanism [27, 28, 29, 30, 31, 32] was proposed as the way to break the symmetry and endow the elementary particles with mass. It introduces an $SU(2)$ doublet of complex scalar fields with a “Mexican Hat” shaped potential of the form

$$V(\phi) = -\mu^2 |\phi|^2 + \lambda |\phi|^4 \quad (1)$$

which for positive μ^2 and λ has a minimum at $|\phi|^2 = \frac{\mu^2}{2\lambda}$. Spontaneous Symmetry Breaking (SSB) occurs when the field acquires a Vacuum Expectation Value (VEV) by choosing an arbitrary vacuum state. This introduces mass terms into the Lagrangian through Yukawa and gauge interactions. The charged leptons and all the quarks acquire mass. The W^\pm and Z^0 bosons become massive as the Goldstone modes give them their longitudinal component. The final degree of freedom left from the original doublet is the Higgs boson. The long hunt for the Higgs boson has recently reached an historic stage as experiments at the LHC reported the finding of a new Higgs-like particle [33, 34]. The matter content of the SM is summarized in Table 1.

An important part the SM is the theory that sets the rules in the realm of quarks and gluons, namely Quantum Chromodynamics. The essence of QCD is presented in the next section.

	$SU(3)_c \times SU(2)_L \times U(1)_Y$	Particles	Spin
Left Handed Leptons	$L_L(1, 2)_{-\frac{1}{2}}$	$(\nu_e)_L, (\nu_\mu)_L, (\nu_\tau)_L$	$\frac{1}{2}$
Right Handed Leptons	$e_R(1, 1)_{-1}$	e_R, μ_R, τ_R	$\frac{1}{2}$
Left Handed Quarks	$Q_L(3, 2)_{\frac{1}{6}}$	$(u)_L, (c)_L, (t)_L$	$\frac{1}{2}$
Right Handed Quarks	$u_R(3, 1)_{\frac{2}{3}}, d_R(3, 1)_{-\frac{1}{3}}$	$u_R, c_R, t_R, d_R, s_R, b_R$	$\frac{1}{2}$
Gauge Bosons	$G(8, 1)_0, W(1, 3)_0, B(1, 1)_0$	$8 \times g, W^\pm, Z^0, \gamma$	1
Higgs	$H(1, 2)_{\frac{1}{2}}$	H^0	0

Table 1: The Standard Model particle content

1.2 Quantum Chromodynamics

The introduction of quarks as an explanation to the observed particle spectrum, led to the emergence of a paradox related to the Δ^{++} baryon, which was discovered in 1951 [35]. The flavor and spin content of this particle are given by the symmetric configuration;

$$|\Delta^{++}\rangle = |u_\uparrow u_\uparrow u_\uparrow\rangle \quad (2)$$

Since the Δ^{++} is a fermion it must however be described by an asymmetric wave function. The introduction of color as a new quantum number [36, 37] solved this puzzle by enabling the asymmetric arrangement of the three quarks using the new “color” degree of freedom:

$$|\Delta^{++}\rangle = \epsilon^{ijk} |u_{i\uparrow} u_{j\uparrow} u_{k\uparrow}\rangle \quad (3)$$

QCD is based on the gauge group $SU(3)$. Gluons in color space are represented by the adjoint representation. Since there are 8 independent directions in this representation, gluons are octets.

The QCD Lagrangian is:

$$\mathcal{L} = \bar{\Psi}_q^i (i\gamma^\mu)(D_\mu)_{ij} \Psi_q^j - m_q \bar{\Psi}_q^i \Psi_{qi} - \frac{1}{4} F_{\mu\nu}^a F^{a\mu\nu} \quad (4)$$

where ψ_q^i are the quark fields with color index i , $\psi_q = (\Psi_{qR}, \Psi_{qG}, \Psi_{qB})^T$, γ^μ are Dirac matrices expressing the vector nature of the interaction. m_q represents the possible non-zero mass, and $F_{\mu\nu}^a$ is the gluon field strength tensor for a gluon with color index a in the adjoint representation. D_μ is the covariant QCD derivative:

$$(D_\mu)_{ij} = \delta_{ij}\partial_\mu - ig_s t_{ij}^a A_\mu^a \quad (5)$$

here g_s is the strong coupling ($g_s^2 = 4\pi\alpha_s$), A_μ^a is the gluon field with color index a , and t_{ij}^a is proportional to a Gell-Mann matrix, a generator of $SU(3)$.

The QCD coupling α_s runs logarithmically with energy:

$$Q^2 \frac{\partial \alpha_s}{\partial Q^2} = \beta(\alpha_s) \quad (6)$$

where $\beta(\alpha_s)$ can be expanded in a power series of α_s whose coefficients contain contributions from gluon loops, quark loops, and mixed loops. At higher order they depend explicitly on the renormalization scheme used. As a side remark, additional colored Beyond the Standard Model (BSM) particles will contribute to this function and modify the running of the coupling at high scales. The value of α_s is usually derived from its value at $Q^2 = M_Z^2$:

$$\alpha_s(Q^2) = \alpha_s(M_Z^2) \frac{1}{1 + b_0 \alpha_s(M_Z^2) \ln \frac{Q^2}{M_Z^2} + O(\alpha_s^2)} \quad (7)$$

where b_0 is the first coefficient in the expansion of $\beta(\alpha_s)$ mentioned above. The QCD coupling effectively decreases with energy, as shown in Fig. 2. This phenomenon is called asymptotic freedom and its discoverers D. Gross, H Politzer, and F. Wilczek received the Noble prize in physics in 2004 [38]. As a consequence of this asymptotic freedom, pQCD calculations converge faster at higher energies. On the low end of the energy scale, the perturbative coupling becomes infinite as

$$\alpha_s(Q^2) = \frac{1}{b_0 \ln \frac{Q^2}{\Lambda^2}} \quad (8)$$

where $\Lambda \sim 200 \text{ MeV}$ is called the Landau pole. In this energy region purely perturbative results are not reliable.

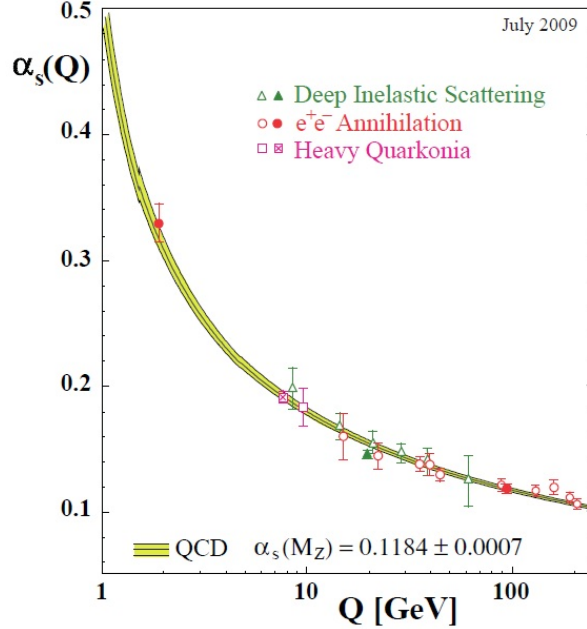


Figure 2: Summary of measurements of α_s as a function of the energy scale Q . The curves are QCD predictions for the combined world average value of $\alpha_s(M_Z)$, in 4-loop approximation and using 3-loop threshold matching at the heavy quark pole masses $M_c = 1.5 \text{ GeV}/c^2$ and $M_b = 4.7 \text{ GeV}/c^2$. Full symbols are results based on N3LO QCD, open circles are based on NNLO, open triangles and squares on NLO QCD. The crossfilled square is based on lattice QCD. The filled triangle at $Q = 20 \text{ GeV}$ (from DIS structure functions) is calculated from the original result which includes data in the energy range from $Q = 2$ to 170 GeV [39].

In order to calculate any related perturbative amplitude, the partonic structure of the colliding hadrons must be addressed.

The factorization theorem [40] shows how the soft processes within the hadron are independent of the hard processes. Thus the cross section for hadron-hadron scattering can be factorized as a convolution of a non-perturbative but process

independent Parton Distribution Function (PDF) and a perturbatively calculable partonic scattering cross section:

$$d\sigma_{h_1, h_2} = \sum_{i, j} \int_0^1 dx_i \int_0^1 dx_j \sum_f \int d\Phi_f f_{i/h_1}(x_i, \mu_F^2) f_{j/h_2}(x_j, \mu_F^2) \frac{d\hat{\sigma}_{ij \rightarrow f}}{dx_i dx_j d\Phi_f} \quad (9)$$

where x_i is the fraction of hadron momentum carried by parton i and f enumerating all possible (partonic) final states, with Lorentz-invariant phase space, Φ . The partonic cross section $d\hat{\sigma}$ is calculable in perturbation theory. The PDF $f_{i/h}$ parametrize the distribution of partons inside the hadron while the factorization scale μ_F is the arbitrary boundary between the perturbative and non-perturbative regions. The PDF are non-perturbative functions and therefore must be measured. In practice they are the outcome of physics-driven fit to data. They can however evolve using a Renormalization Group Equation (RGE) with μ_F . In the example given in Fig. 3 one can clearly see that at low $Q = \mu_F = 2 \text{ GeV}$ the proton is dominated by a few hard (valence) quarks that produce the bump around $x \sim 0.2$, while at higher scales $Q = \mu_F = 100 \text{ GeV}$ it is dominated by a large number of gluons and sea-quarks, all with distributions with rather small x values, while the number of valence quarks play a smaller role.

The observation of jets is an evidence for the phenomenon called color confinement, which relates to the fact that the potential of the strong force grows linearly with the distance between colored objects. Thus free colored objects are bound to form color neutral objects. Jets of particles are formed in roughly three steps that are depicted in Fig. 4. First, partons are scattered away from the mother hadron by hard collisions. Next the prolific parton showers are formed. Lastly, color charge is neutralized by the formation of new hadrons.

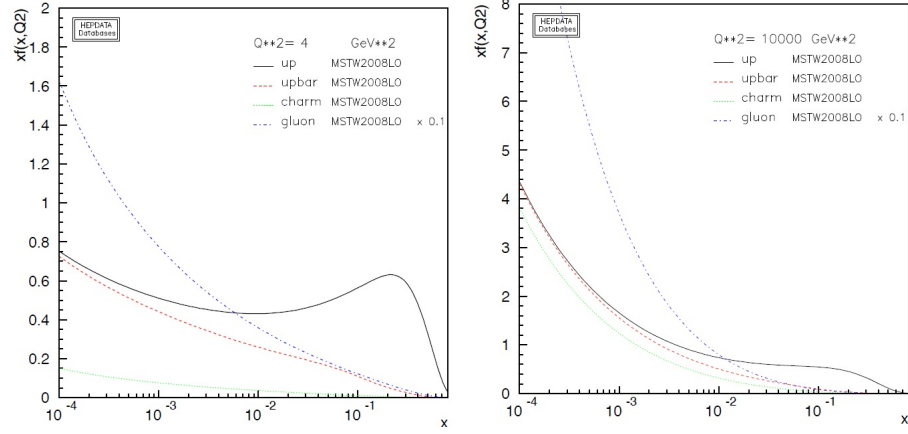


Figure 3: The change of the u (black), \bar{u} (red, dashed), c (green, dotted), and g (blue, dot-dashed) distributions, from $Q = \mu_F = 2 \text{ GeV}$ (left) to $Q = \mu_F = 100 \text{ GeV}$ (right) [38, 41].

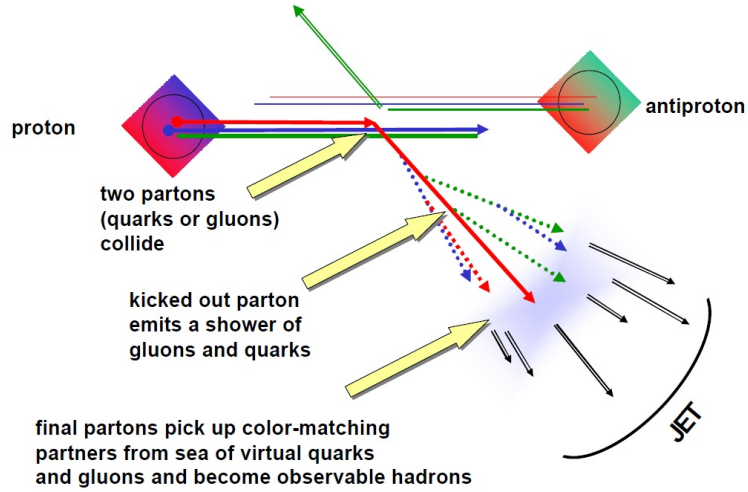


Figure 4: Illustration of the three conceptual steps of jet formation.

1.3 The Tevatron

The Tevatron is a 4-mile-long proton-antiproton collider located at the Fermi National Accelerator Laboratory in Batavia, Illinois, USA. First collisions were observed by CDF in 1985. The Run I physics program ran from 1992 to 1996. Run II with collisions at the center of mass energy of 1.96 TeV started in March 2001, and collected close to 12 fb^{-1} of data until its shutdown in 2011. The collected integrated luminosity by fiscal year is shown in Fig. 5.

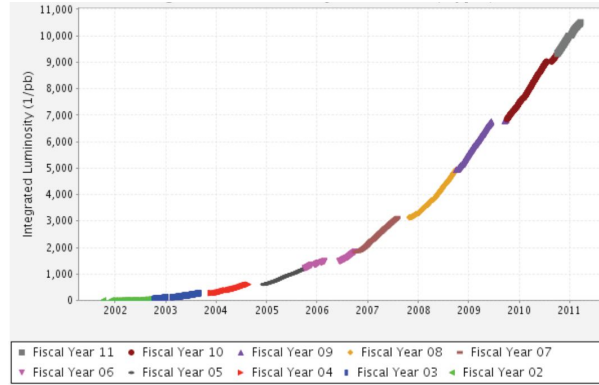


Figure 5: Run II integrated luminosity by fiscal year [42].

The acceleration chain of the Tevatron begins with negatively charged hydrogen ions H^- which are accelerated in a Cockcroft-Walton accelerator up to 750 keV and then further accelerated up to 400 MeV in a linear accelerator. They are injected into the Booster through a carbon foil which strips the electrons and leaves the bare protons. The Booster, a 470 meter synchrotron, brings up the protons energy to 8 GeV and they are passed on to the Main Injector, a synchrotron 7 times larger than the Booster. The Main Injector serves several purposes; it accelerates the protons up to 150 GeV and passes them to the Tevatron. Some of the protons are accelerated to 120 GeV and are sent upon a fixed nickel alloy target for anti-proton production.

An 8 GeV beam of anti-protons is collected in a spectrometric method and passed

to the Debuncher. This is a triangular synchrotron with a mean radius of 90 *m*. Its purpose is to stochastically cool down the anti-protons and lower their transverse momentum spread. They are then stored in the Accumulator which is 75 *m* mean radius synchrotron housed in the same tunnel as the Debuncher. There the anti-protons are further cooled and stored. The rate of anti-proton production decreases as the Accumulator becomes saturated, therefore when enough anti-protons fill the Accumulator, they are passed on the the Recycler, another 8 *GeV* synchrotron housed along the ceiling of the Main Injector tunnel. There the anti-protons are further focused. Before collisions begin, they are accelerated up to 150 *GeV* in the Main Injector.

Collisions start by injecting bunches of protons and anti-protons into the Tevatron. Due to their properties, the same electro-magnetic fields are used to accelerate and turn both beams. A 4.2 *T* magnetic field keeps the 36 bunches of protons and anti-protons in each beam circling while they are accelerated up to 980 *GeV* and collide with $\sqrt{s} = 1.96 \text{ TeV}$ in a rate of 2.5 *MHz* [43].

1.4 The CDF Detector

The CDF detector consists of a calorimeter extending up to $|\eta| < 3.6$. The central calorimeter, covering the range of $|\eta| < 1.1$, is segmented into towers of size $\Delta\eta \times \Delta\phi = 0.11 \times 0.26$ ¹. This calorimeter consists of lead and steel absorbers interleaved with scintillator tiles that measure the deposited energy. The inner calorimeter compartment consists of lead absorbers providing an Electro Magnetic (EM) measurement, while the outer compartment consists of steel absorbers to measure Hadronic (HAD) energy. Transverse energy deposited in the EM calorimeter is measured with a resolution of approximately $0.15\sqrt{E_T} \text{ GeV}$

¹In the coordinate system being used, ϕ and θ are the azimuthal and polar angles around the proton beam axis. The pseudorapidity is $\eta = -\ln \tan(\theta/2)$ and $R = \sqrt{(\delta\eta)^2 + (\delta\phi)^2}$

where E_T is expressed in units of GeV , while the resolution of the HAD calorimeter is approximately $1.1\sqrt{E_T}$ GeV . Two plug calorimeters in the forward and background regions provide energy measurement in the interval $1.1 < |\eta| < 3.6$ using lead and steel absorbers interleaved with scintillator tiles that measure the deposited energy. The calorimeter system is used to measure the jet's energy and the event's Missing Transverse Energy (\cancel{E}_T). Measurement of \cancel{E}_T is made by vectorially summing, for towers with $|\eta| < 3.5$, the energy deposition in each calorimeter tower. This has a resolution of approximately $1.1\sqrt{\sum E_T}$ GeV , where the sum is over the transverse energy observed in all calorimeter towers. This resolution has been determined by detector studies of events with and without significant missing transverse energy [44]. A measure of how large the observed \cancel{E}_T in an event is relative to its uncertainty is, therefore, provided by the Missing Transverse Energy Significance (S_{MET}), defined as:

$$S_{MET} \equiv \frac{\cancel{E}_T}{\sqrt{\sum E_T}} \quad (10)$$

where the numerator is the missing energy, and the summation in the denominator is over the transverse energy of all calorimeter towers.

A superconducting 1.416 Tesla solenoidal charged particle spectrometer resides inside the calorimeter volume and provides charge particle momentum measurement over $|\eta| < 1.5$. Tracks are reconstructed with the help of a set of silicon microstrip detectors arranged in a barrel geometry around the collision point. This is followed by a cylindrical drift chamber, known as the Central Outer Tracker (COT) that provides charged particle tracking from a radius of 40 to 137 cm . The detector also includes systems for electron and muon identification, but these are not used in the present study.

Further details on the CDF detector can be found in [45].

1.5 The Large Hadron Collider

In 2010, at the European Center for Nuclear Research (CERN) in Geneva, Switzerland, the Large Hadron Collider (LHC) provided its first proton-proton collisions at a beam energy of 3.5 TeV . The acceleration process begins with the linear accelerator LINAC2 which delivers 50 MeV protons. Next, they go through a series of circular synchrotrons; the Proton Synchrotron Booster brings their energy up to 1.4 GeV , the Proton Synchrotron accelerates them further up to 26 GeV , and after the Super Proton Synchrotron they reach an energy of 450 GeV [46]. Finally they go into the LHC tunnel inside which is a 27 km long synchrotron. Currently the protons collide at 8 TeV while the goal energy of the LHC is 14 TeV with a luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ [47]. The proton bunches are spaced 25 ns apart. Superconducting bending magnets operating at 1.9 K produce a magnetic field of 8.3 T to achieve such a high beam energy with reasonable power consumption. Figure 6 shows the integrated luminosity at the LHC delivered in 2011

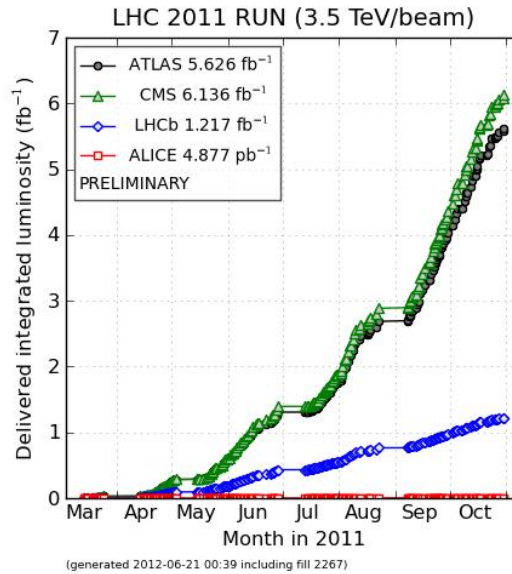


Figure 6: Integrated luminosity at the LHC delivered in 2011 [48].

1.6 The ATLAS Detector

A Toroidal Lhc ApparatuS (ATLAS) is one of the two general purpose detectors at the LHC. It consists of an inner detector immersed in a 2 T solenoidal field. A combination of high resolution silicon made pixel and strip detectors, together with straw tube tracking detectors, achieve the pattern recognition, momentum and vertex measurements, as well as electron identification up to $|\eta| < 2.5$ [49].

The ATLAS calorimeter in the barrel is composed of high granularity Liquid Argon (LAr) electromagnetic sampling calorimeters that cover the pseudorapidity range up to $|\eta| < 3.2$. The hadronic calorimeter in the barrel is made of scintillating tiles and covers the range $|\eta| < 1.7$. In the range $|\eta| > 1.5$, namely the end caps, the hadronic calorimeter uses LAr technology. At higher $|\eta|$, up to $|\eta| < 4.9$, LAr is used for both EM and HAD energy measurements.

The ATLAS muon system is based on an air core toroid system which gives ATLAS its name and typical look, and also minimizes multiple scattering, thus achieving excellent muon momentum resolution. The muon system includes four different technologies; the Monitored Drift Tubes (MDT) cover the range $|\eta| < 2.7$, the Cathode Strip Chambers (CSC) cover the range $2.0 < |\eta| < 2.7$, the Resistive Plate Chambers (RPC) cover the range $|\eta| < 1.05$, and the Thin Gap Chambers (TGC) cover the range $1.05 < |\eta| < 2.7$ (2.4 for triggering). The muon system trigger capabilities have a timing resolution of the order of $1.5 - 4\text{ ns}$. Table 2 summarizes the performance of the ATLAS detector.

Component	Required Resolution	η Coverage	
		Measurement	Trigger
Tracking	$\sigma_{p_T}/p_T = 0.05\% p_T \oplus 1\%$	± 2.5	
EM Calorimetry	$\sigma_E = 10\%/\sqrt{E} \oplus 0.7\%$	± 3.2	± 2.5
Hadronic Calorimetry (jets)			
barrel and end cap	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$	± 3.2	± 3.2
forward	$\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$	$3.1 < \eta < 4.9$	$3.1 < \eta < 4.9$
Muon Spectrometer	$\sigma_{p_T}/p_T = 10\%$ at $p_T = 1 \text{ TeV}$	± 2.7	± 2.4

Table 2: General performance of the ATLAS detector. The units for E and p_T are in GeV [49].

2 Methods

2.1 Jet Algorithms

The observation of hadronic jets in e^+e^- annihilation provided one of the most striking confirmations of the parton picture. In the 70' jets from QCD started to have more rigorous theoretical descriptions (see *e.g.* [50]) which led to the emergence of jet algorithms. Jet algorithms are aimed at associating clusters of particles² of a common origin into jets. This facilitates making the correspondence between the energetic partons produced in the hard scattering process, and the reconstructed jets. Thus jets are the observed traces of those partons. Once the particles that constitute a jet are chosen by the algorithm, a scheme is used to calculate the four-momentum of the jet as a combination of its constituents which are usually taken to be massless. This is called the recombination scheme. In this study, a simple four-vector summation of the massless constituents is used to deduce the jet's four-momentum.

Modern jet algorithms are required to retain the following two important properties:

1. *Infrared safety*: the algorithm should not have Infra Red (IR) singularities, and should therefore be insensitive to soft radiation of particles in the event. Seeded algorithms, namely those which depend on the presence of a particle with energy above a certain threshold (known as seed, or initiator), are an example of IR-unsafe algorithms. This is illustrated in Fig. 7.
2. *Collinear safety*: the algorithm should not have collinear singularities, and should therefore be insensitive to collinear radiation of particles in the event.

Example of such a sensitivity is illustrated in Fig. 8.

²Particles in this context can be partons, hadrons, detector objects (*e.g.* calorimeter towers or clusters), or any list of four-momenta.

The following subsections describe the jet algorithm variants used in this study.

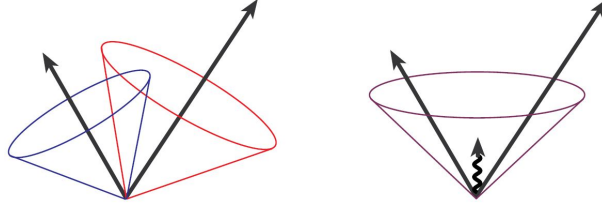


Figure 7: An illustration of infrared sensitivity in cone jet clustering. In this example, jet clustering begins around seed particles, shown here as arrows with length proportional to their energy. The jet algorithm ends up with two distinctive jets (on the left). The presence of an additional soft radiation between these two jets (right) may lead to a merger between the two jets [17].

2.1.1 Midpoint

The Midpoint jet algorithm [17] is a cone algorithm. Due to computational limitations, cone algorithms usually start by looking for seeds, namely particles that pass some energetic threshold. Once found, a seed serves as a temporary jet axis, and all particles lying within a circle of a given radius R in the $\eta \times \phi$ plane around that seed are clustered into a jet candidate. This circle defines a cone-like shape in 3-D space, hence the name cone algorithms. The sum of momenta of all particles in this cone is computed and serves as the jet axis in the next iteration. This process is repeated until the jet axis in a given iteration is identical to that in the previous iteration. The cone is then named stable.

Often two stable cones share some constituents. Particles should not be fully assigned to more than one jet. A split and merge stage usually attempts to rectify such a situation. In this stage, overlapping cones are either merged when there is a large overlap, or the shared particles are split between the cones. If the least

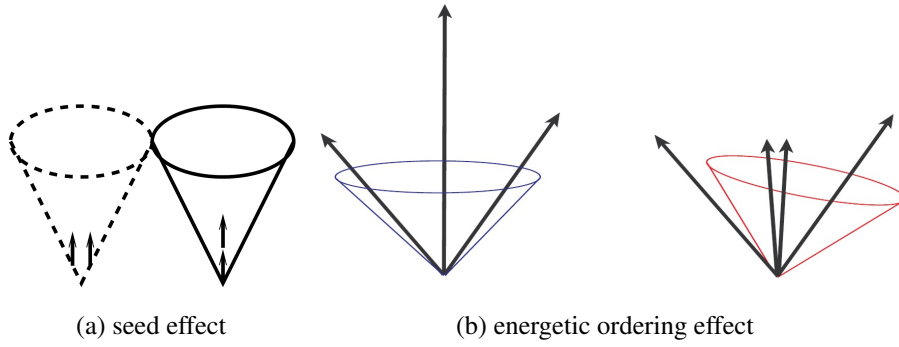


Figure 8: Illustrations of collinear sensitivity in jet reconstruction: (a) the configuration on the left fails to produce a seed because its energy is split among several particles. The configuration on the right produces a seed because its energy is more narrowly distributed; (b) The clustering proceeds from the most energetic to the low energy initiator. On the left one cluster that is centered on the highest energy constituent encompass all and one jet is formed. Once this particle undergoes collinear radiation the first jet is formed around a different initiator and the outcome of the jet algorithm is different [17].

energetic of a pair of cones contains a fraction of shared energy which is above a predefined cutoff value - the pair of cones is merged into a large cone. If the fraction of shared energy is below that threshold, particles belonging to more than one cone are assigned to the cone that is closest in the $\eta \times \phi$ plane.

The process described so far creates jets whose IR sensitivity is ranked as IR_{2+1} ³. The Midpoint algorithm got its name since it is introducing an additional stage. In this new stage, a virtual seed is added between every two seeds whose relative distance is less than twice the required cone size. This reduces the sensitivity to soft radiation that is illustrated in Fig. 7 and promotes the ranking of the algorithm to IR_{3+1} .

2.1.2 Midpoint with Search Cones

During the stable cone search, while the cone slightly wobbles, some energy can drop out of the cone, sometimes leaving a substantial part of the energy in the event unassigned to any of the jets, as illustrated by the literally dark towers in Fig. 9. This poses the risk that the energy of the dark towers will be lost and the interpretation becomes erroneous. Search cone is a technique which reduces this effect by first using a smaller radius, half the size of the final cone, to find stable cones and then expanding the cones to their full radius, without further adjustment. Midpoint seeds are then defined between pairs of stable cones as usual. Finally the jet candidates are passed through the split and merge procedure [18]. The use of smaller search cone reduces the phase space for configurations that lead to dark towers.

A disturbing feature of the search cone variant of the Midpoint jet algorithm arises when its outcome is compared to higher orders in perturbation theory. The

³ IR_{n+1} means that given n hard particles in a common neighborhood, the addition of 1 extra soft particle can modify the number of final hard jets [2]

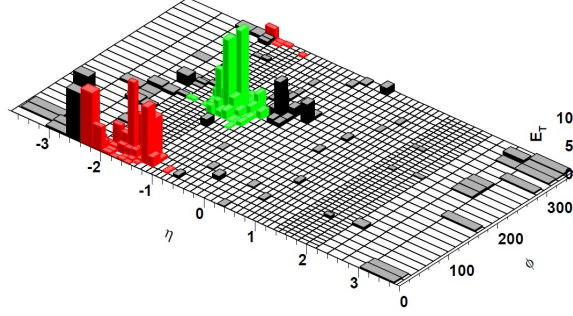


Figure 9: An example of an event where the Midpoint algorithm has left substantial energy unclustered. The colored columns represent energy deposits which belong to a jet. The dark columns are unclustered energy deposits [51].

search cone may identify a small, stable, and soft cone right between two energetic cones, exactly as shown in the example illustrated in Fig. 7 thus reintroducing IR-unsafety issues which demote the ranking of this variant back to IR_{2+1} .

As will be exhibited later, the issue of IR safety is not a theoretical one. It manifests itself very strongly when one studies the substructure of highly boosted massive jets.

2.1.3 *anti- k_t*

The *anti- k_t* jet algorithm [19] is a member of a group of jet reconstruction algorithms known as sequential recombination algorithms [2]. This class of algorithms is seedless and combines pairs of particles together repeatedly, according to some matching criteria, until some stopping criterion is reached. Such algorithms are usually fully IR-safe and fully collinear safe algorithms. The *anti- k_t* algorithm is in wide use by a large variety of analysis. It works as follows:

For each pair of particles, the variable $d_{ij} = \min(p_{T,i}^{-2}, p_{T,j}^{-2}) \frac{\Delta R_{ij}^2}{R^2}$ is calculated, where ΔR_{ij} is the distance between the two particles. In addition, for each particle the value of $d_i = p_{T,i}^{-2}$ is calculated. These d_{ij} s and d_i s are put into a common list

which is then sorted. The minimal item in the list is selected. If this item is of a d_{ij} type, then the i^{th} particle and the j^{th} particle are removed from the list and replaced by a new quasi-particle which is their combination. The procedure is then repeated. If the minimal item is of a d_i type, the i^{th} particle is taken out of the list and is promoted to a jet. The procedure is repeated as long as the list is populated. In addition to the IR and collinear safety properties of this algorithm, hard particles are clustered before soft ones, due to the selection of the minimum of the p_T with a negative power. This reconstructs jets which are not sensitive to soft particles and have a well defined circular edges.

2.2 Event Selection

2.2.1 Data and Monte Carlo Samples

The data used for this study comprise of an integrated luminosity of 5.95 fb^{-1} collected by CDF in Run II. The detector was triggered by requiring the presence of a jet with Transverse Energy (E_T) in excess of 100 GeV . Such a trigger had practically 100% efficiency for events which are the subject of this study. A total of 76 million events pass this trigger.

The PYTHIA 6.216 [52] Monte Carlo (MC) generator was used to produce a sample of QCD events. Partons were described, in this run, by the CTEQ5L PDF [53] parameterization and was limited to partons with $\hat{p}_T > 300 \text{ GeV}/c$. The simulated sample represented an equivalent integrated luminosity of about 800 fb^{-1} . The partonic \hat{p}_T cut on $300 \text{ GeV}/c$ was shown to have negligible effect once events with at least one jet with $p_T > 350 \text{ GeV}/c$ were selected. An average of 0.4 additional collisions per beam crossing were included in the simulation in order to model the effect of Multiple Interactions (MI) in this MC sample. In practice, the number of MI (as measured by the number of primary vertices) in the

data is significantly higher. This issue will be addressed below. The MC events were passed through a full CDF detector simulation.

A similar PYTHIA MC sample was used to study the $t\bar{t}$ event kinematics. This sample had an equivalent integrated luminosity of about 600 fb^{-1} .

Both data and MC events were reconstructed using the standard CDF event reconstruction software, and were further subjected to the selected procedure that is described in the next section.

2.2.2 Selection Criteria and Jet Calibration

The selected events had to be included in a standard Good Runs List (GRL) which ensure that all the relevant parts of the detector were properly functioning. Furthermore, selected events had to contain a high quality reconstructed interaction vertex, which is within 60 cm from the nominal interaction point.

1. $|Z_{\text{vx}}| < 60\text{ cm}.$

The Midpoint jet [17] algorithm with cone sizes $R = 0.4, 0.7$ and 1.0 was applied and the presence of at least one jet having

2. $p_T > 400\text{ GeV}/c$

3. $|\eta| < 0.7$

was required. The tight $|\eta|$ requirement ensured that the leading jet is in the well-understood CDF central calorimeter.

The jets were reconstructed using Fastjet [54], from calorimeter towers. The towers, of size $\Delta\eta \times \Delta\phi = 0.11 \times 0.26$, were assumed to be massless, and their four-momenta vector sum constituted the jet four-momentum. The jet level selection was done after a standard CDF η -dependent p_T correction was performed. These correction accounts for inhomogeneities in the detector response, calorimeter response non-linearities, and jet energy corrections due to MI. These corrections are

described in [55].

Since the *anti- k_t* jet algorithm [19] is not a native algorithm in the CDF software, a private calibration method for it was used. The correction factors used for Mid-point jets were binned in detector η , and the mean correction for each bin was used to correct the relevant *anti- k_t* jets. This method was validated to reproduce the jet energy of corresponding jets without calorimeter effects up to a few per-cent.

The S_{MET} (Missing Energy Significance) distribution for data events with a good vertex and a boosted jet is shown in Fig. 10. Cosmic ray backgrounds and poorly measured events were rejected by requiring:

$$4. S_{MET} < 10 \text{ GeV}^{1/2}$$

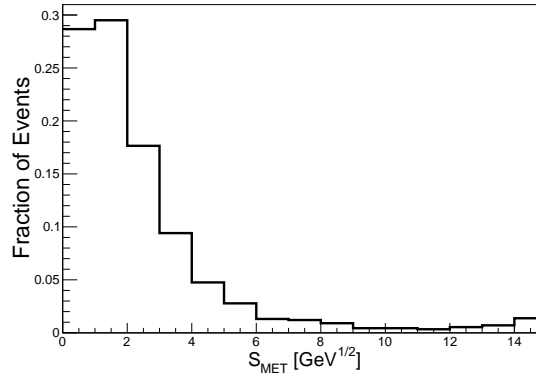


Figure 10: The S_{MET} distribution for data events with a good vertex and a boosted jet

As jets transverse through the detector they leave energy deposits along their track. Comparing energy deposits from various parts of the detector allows additional quality assurance for jets. This is exploited by the two variables:

$$f_{tr} \equiv \frac{\sum_i^{N_{ch}} p_T^i}{p_T^{jet}} \quad (11)$$

where N_{ch} is the number of charged tracks associated with the jet, and p_T^i is the p_T of the i^{th} track, and

$$f_{EM} \equiv \frac{E_{EM}}{(E_{EM} + E_{HAD})} \quad (12)$$

where f_{EM} is the EM energy fraction of the jet candidate. The distributions of these quantities for events satisfying the first three requirements is shown in Fig. 11. Each jet candidate had to satisfy:

5. either $f_{tr} > 0.05$ or $f_{EM} > 0.05$

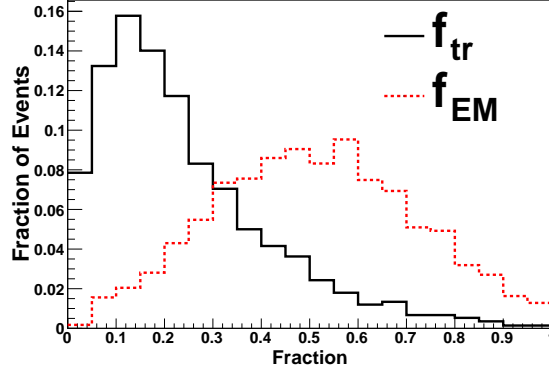


Figure 11: The f_{tr} and f_{EM} distributions for events with a good vertex, a boosted jet, and good S_{MET}

This cut rejects 1.4% of the data events, and a negligible fraction in the MC sample.

2699 events passed these selection criteria for a cone size of $R = 0.7$. In 22% of these events (591) the sub-leading jet also satisfied these conditions, resulting in a total of 3290 good jets with $p_T > 400 \text{ GeV}/c$. The p_T distributions of these jets are shown in Fig. 12 for data, as well as for the QCD and $t\bar{t}$ MC samples.

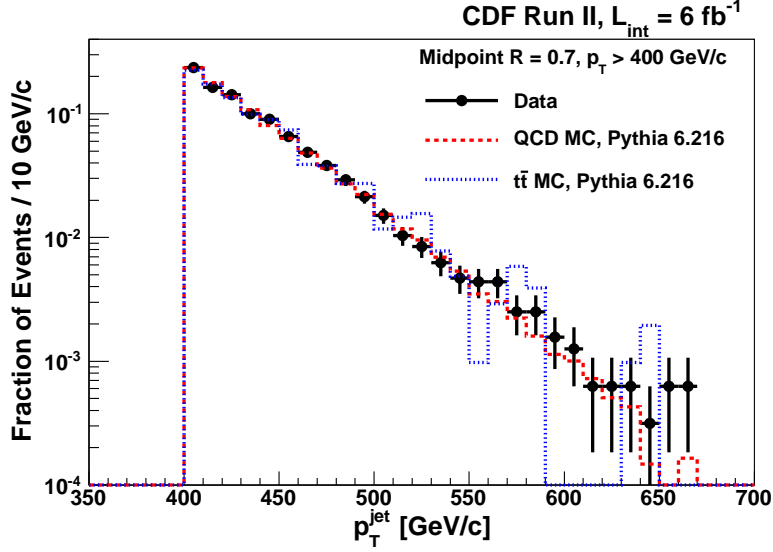


Figure 12: The p_T distribution for all the jets with $p_T > 400 \text{ GeV}/c$ in events passing the requirements described in the text. Overlaid are the distributions from the PYTHIA MC for QCD jets and $t\bar{t}$ production. All samples are area normalized.

2.2.3 Expected Sources of Events

Fig. 13 depicts the MC-based estimation of the relative importance of various possible processes in $p\bar{p}$ collisions as a function of the p_T^{min} cut imposed on the leading jet (cuts 2 and 3). The dominant QCD source is suppressed by a factor of 250 for presentational purposes.

Tevatron studies have shown, as expected, that the production of light quarks and gluons, followed by their fragmentation and hadronization, is the main source of high p_T jets [44, 56]. The nice agreement in shape shown in Fig. 12 supports this assertion. Results from an NLO QCD calculation using POWHEG [57, 58, 59, 60] and CTEQ6m PDF [61] show that about 80% of the jets with $p_T > 400 \text{ GeV}/c$ originate from the production of high p_T quarks, and the rest from the production of gluons. This is in agreement with previous measurements at lower jet energies [10].

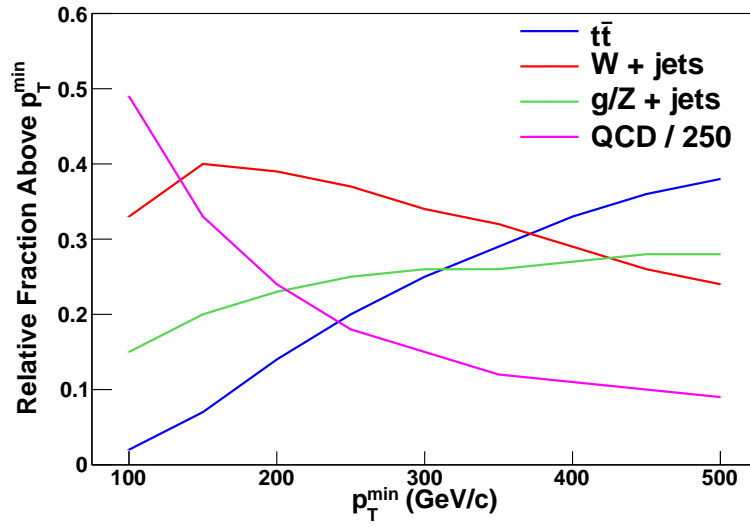


Figure 13: The fractional contribution of the various SM sources as a function of the minimum p_T of the leading jet for $p\bar{p}$ collisions at 2 TeV . Note that the rate of jets from light quarks and bottom quarks (QCD) is suppressed by 250. The Z+jet cross section is separated from the Drell-Yan process by applying a mass cut on the out-coming daughters.

The computed production cross-section for $W + \text{Jets}$ and $Z + \text{Jets}$ with p_T above $400 \text{ GeV}/c$ is 4.5 fb and 3 fb respectively. Hence, after inclusion of the hadronic branching fractions and jet reconstruction acceptance and efficiencies, these processes are expected to contribute about 20 jets, with jet masses in the range of 50 to $100 \text{ GeV}/c^2$ to the sample. In this mass range, 296 events were observed. Due to the small cross section and since the present study is focused on higher jet masses, no attempts to remove the W/Z events has been done.

Fig. 13 shows the p_T behavior of $t\bar{t}$ jets which constitute an irreducible background for the present study. One deduced from this figure that about 1% of the events is indeed due to the SM boosted $t\bar{t}$ production process. However, QCD jets are very unlikely to be massive while fully collimated t-quark jets are heavy by definition, as shown in Fig. 14.

Since search for boosted $t\bar{t}$ is part of this thesis, a better level of understanding has to be sought. Hence, two calculations for the boosted top cross section were considered. Using the MSTW 2008 PDF with a top quark mass of $m_{top} = 173 \text{ GeV}/c^2$, an approximate Next-to-Next-to-Leading Order (NNLO) calculation predicts the cross section of $t\bar{t}$ with $p_T > 400 \text{ GeV}/c$ to be $4.55^{+0.50}_{-0.41} \text{ fb}$ [62]. A PYTHIA MC result is $5.67 \pm 0.37 \text{ fb}$, which is based the measured total $t\bar{t}$ cross section of $7.5 \pm 0.48 \text{ pb}$ [63]. This is in reasonable agreement with the theoretical NNLO prediction.

The former calculation was deemed more reliable and was used in relevant parts of this study.

2.2.4 Characteristics of $t\bar{t}$ Events and Top Rejection

Since $t\bar{t}$ is a major background for the QCD study, in this section, the characteristics of $t\bar{t}$ events and their associated jets are compared to that of QCD events.

The mass distribution of leading jets with $p_T > 400 \text{ GeV}/c$ is shown in Fig. 14.

It demonstrates a large peak around a mass of $40 \text{ GeV}/c^2$ with a long tail towards high jet mass, very similar to the shape measured in data as shown in Fig. 18. The $t\bar{t}$ distribution, on the other hand, has a broad peak around the mass of the top quark, coming from hadronically decaying boosted tops fully collected into a single jet. It also has a secondary peak around the mass of the W boson, coming from events in which the light quarks from the W boson are contained in a single jet, while b flavored quark is reconstructed into a jet of its own. It is also worth mentioning that there are very few $t\bar{t}$ events with jet masses below $70 \text{ GeV}/c^2$ and above $200 \text{ GeV}/c^2$.

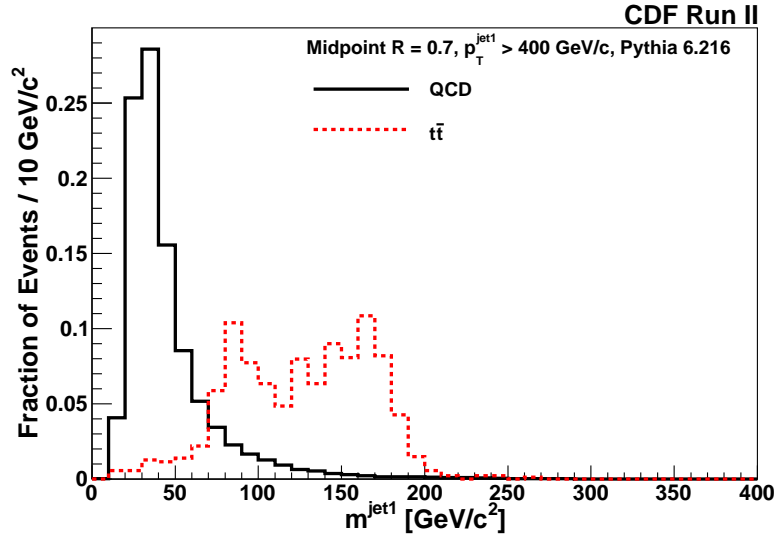


Figure 14: The jet mass distribution for leading Midpoint jets with $R = 0.7$, satisfying $p_T > 400 \text{ GeV}/c$ and $|\eta| \in (0.1, 0.7)$ in QCD (solid black) and $t\bar{t}$ (dashed red) MC events.

Approximately 40% of the $t\bar{t}$ events are expected to have a semileptonic decay in which the leading jet is a highly boosted massive jet, while the second jet is softer and less massive. These events are also characterized by having large missing energy due to the escaping neutrino. The S_{MET} distribution, namely the

significance of the missing energy, is shown in Fig. 15 in which the $t\bar{t}$ tail at large S_{MET} is higher than that of QCD.

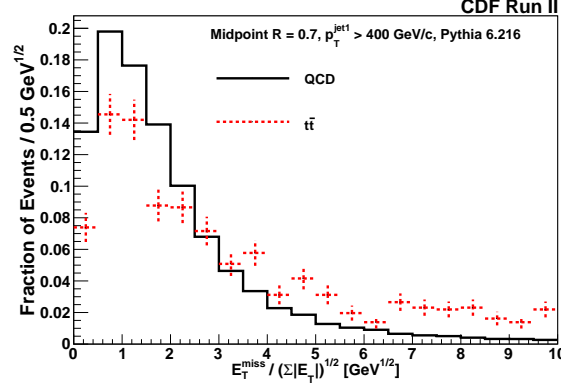


Figure 15: The S_{MET} distributions for QCD (solid black) and $t\bar{t}$ (dashed red) MC events, requiring that the leading jet satisfy $p_T > 400 \text{ GeV}/c$ and $|\eta| \in (0.1, 0.7)$.

Since the main goal of the first part of this study is the systematic investigation of the properties of highly boosted, massive QCD jets, it is important to remove the $t\bar{t}$ contribution which falls exactly in this region of phase space. The above observations were quantized and utilized for rejecting the two $t\bar{t}$ configurations. Rejecting events in which the second jet has a mass above $100 \text{ GeV}/c^2$ strongly suppresses fully hadronic events. Requiring additionally that $S_{MET} < 4 \text{ GeV}^{1/2}$ strongly suppresses semileptonic events. A requirement on the second jet $p_T > 100 \text{ GeV}/c$ was also imposed in order to have a sufficiently energetic recoil jet, and further improve the selection of typical QCD dijet events. These top rejection cuts have an efficiency of 78% for the simulated QCD sample, while in the simulated $t\bar{t}$ sample only 26% of the events survive these cuts, leaving 4 expected events in the data sample. The number of remaining events in the data after these top rejection cuts were applied for each cone size are summarized in Table 3.

p_T Interval (GeV/c)	Cone Size		
	$R = 0.4$	$R = 0.7$	$R = 1.0$
$400 \leq p_T < 500$	1729	1988	2737
$p_T > 500$	107	120	175

Table 3: The number of observed data events with at least one jet in the high p_T intervals studied and for three different cone sizes. All events were required to have at least one Midpoint jet of the given cone size with $p_T > 400 \text{ GeV}/c$ and $|\eta| \in (0.1, 0.7)$. The selection used to reject top quark candidates has been applied.

2.3 Pileup Correction

More than one interaction may occur during a single bunch crossing in a pp or $p\bar{p}$ collider. The number of reconstructed primary vertices, N_{vx} , is a good estimate for the number of actual multiple interactions. In the present CDF data sample an average of ~ 3 interactions per bunch crossing was measured, while at the LHC this number is much higher. The additional collisions add a cloud of particles that, upon impinging on the calorimeter, give rise to incoherent additional recorded energy. Such incoherent energy deposits partially mask the jet substructure. Moreover, the pileup in the MC samples used is not necessarily modeled perfectly. Therefore, a data-driven technique was developed and used to measure the shift in various observables due to the inclusion of this additional incoherent energy in the jet and correct for it [11, 64].

In order to estimate the amount of the incoherent energy deposits, a complementary cone, one with the same size parameter R as used for the jet, is set at 90° in azimuth away from the leading jet axis. The energy deposits contained by this cone are rotated back and added to the leading jet. Then the shifts in the relevant

observables are measured as a function of the raw values of these observables. It is assumed that the shifts are small and within the linear regime. Indeed this is verified and reported in Sec. 3.1. These positive shifts can be subtracted from any measured observable to remove the pileup contribution. The MI contribution exhibits large event-by-event fluctuations and therefore the mean shifts are used as the corrections. Furthermore, some additional coherent energy deposits are coming from the Underlying Event (UE) and should also be taken into account. This is done by the following method: events with a single reconstructed vertex are selected. These are assumed to have underlying activity with no multiple interactions. The average shifts to the relevant quantities in these events are computed using the complementary cone technique and later subtracted from the shifts incurred to these quantities in events with more than one vertex. The latter events are assumed to contain both underlying event and multiple interaction contributions. The resulting shifts are due to multiple interactions only. This is demonstrated schematically in Eq. 13.

$$N_{vtx} > 1 \text{ (pileup and underlying event)} - N_{vtx} = 1 \text{ (underlying event)} = \text{pileup only} \quad (13)$$

The mass shift due to MI is expected to scale as R^4 [2]. This scaling was used to derive the expected shifts for the $R = 1.0$ case since in this case the complementary cone overlaps with the primary jet and therefore the technique cannot be applied without modifications. In the case of $R = 0.4$ jets, there were too few events with sufficiently massive jets to reliably derive the shifts, and scaling down from $R = 0.7$ was performed.

The functional form of the observable shift due to multiple interactions is estimated for several observables [64]. For example, denote the mass after a small

pileup contribution is added to it by $m' \equiv m + \delta m$. Thus

$$\underbrace{m'^2 - m^2}_{\delta m^2} = \underbrace{(m' - m)}_{\delta m} \underbrace{(m' + m)}_{\sim 2m} \quad (14)$$

and therefore $\delta m \sim \frac{1}{m}$, namely, the shift in jet mass due to pileup is expected to behave as one over the jet mass, namely the higher the jet mass is the smaller is the correction.

2.4 Unfolding

The data presented in this study was unfolded back to particle level, thus reducing detector effects and allowing a better comparison with the theoretical predictions. The PYTHIA QCD event generator was used to simulate and reconstruct jets with various cone sizes at both particle level and after full CDF detector simulations. All interesting distributions are then calculated from both kinds of jets. The bin-to-bin ratio of these normalized distributions is calculated, namely the ratio of the particle level distribution over that of the fully simulated reconstructed jets. This ratio is fit with a polynomial, which is later used to correct the measured data. An example for the jet mass unfolding factors for $R = 0.7$ is shown in Fig. 16. The unfolding corrections for planar flow and angularity were found to be negligible, except for the case of planar flow for jets with $R = 1.0$, in which case the distributions were corrected by about 10%.

2.5 Jet Substructure Observables

2.5.1 Jet Mass

The typical QCD jet mass distribution has a peak at a value comparable to a small fraction of the jet momentum and a long tail reaching masses which are a significant fraction of the jet's p_T . This is demonstrated in Fig. 12. Based on QCD

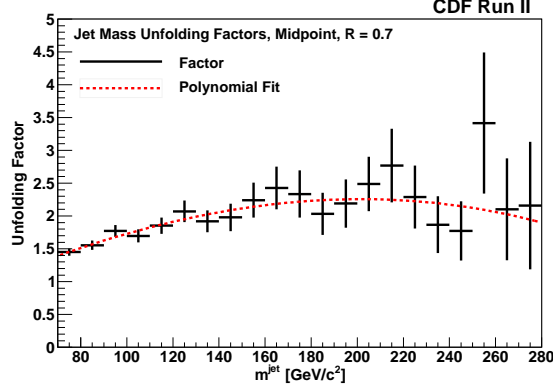


Figure 16: The bin to bin ratio of the jet mass distributions for particle-level jets and detector-level jets with $R = 0.7$ in PYTHIA MC events. The dashed red curve is the result of a polynomial fit to the MC points that is used to make the bin-by-bin correction. The uncertainties shown are from the statistics of the MC sample

factorization (see *e.g.* [65]), a semi-analytic calculation of the QCD jet mass distribution was derived for this high-mass tail where the jet mass is assumed to be dominated by a single gluon emission [20], giving the jet, at leading order, a distinctive two-prong geometry. The probability for such gluon emission is given by the “jet functions,” J^q and J^g for quarks and gluons, respectively. These are defined via the total differential rate

$$\frac{d\sigma(R)}{dp_T dm^{jet}} = \sum_{q,g} J^{q,g}(m^{jet}, p_T, R) \frac{d\hat{\sigma}^{q,g}(R)}{dp_T}, \quad (15)$$

where R is the radius of the jet cone used to define the jets and $\hat{\sigma}^{q,g}$ is the factorized Born cross section. Corrections of $O(R^2)$ are neglected and the analysis is applied to the high mass tail, $m^{\text{peak}} \ll m^{jet} \ll p_T R$. The jet functions are approximated by

$$J(m^{jet}, p_T, R) \simeq \alpha_s(p_T) \frac{4C_{q,g}}{\pi m^{jet}} \log\left(\frac{R p_T}{m^{jet}}\right), \quad (16)$$

where $\alpha_s(p_T)$ is the strong coupling at the appropriate scale and $C_{q,g} = 4/3$ and 3 for quark and gluon jets, respectively.

The jet function predicts both the shape of the jet mass distribution, as well as its absolute normalization. The above result is the leading-log approximation to the full expression where the NLO corrections are yet unknown [2, 66, 67]. These corrections are of order of $1/\log(R^2 p_T^2/(m^{jet})^2)$, *i.e.* at the level of $\sim 30\%$ of the expression in Eq. 16. Thus, Eq. 16 provides a simple yet powerful description for the qualitative behavior of the high m^{jet} tail.

2.5.2 Angularity

Angularity is a class of jet observables defined as [21, 68]:

$$\begin{aligned}\tau_a(R, p_T) &= \frac{1}{m^{jet}} \sum_{i \in jet} E_i \sin^a \theta_i [1 - \cos \theta_i]^{1-a} \\ &\sim \frac{2^{a-1}}{m^{jet}} \sum_{i \in jet} E_i \theta_i^{2-a},\end{aligned}\tag{17}$$

where the sum runs over the jet constituents, E_i is the energy of a constituent, and θ_i is its angle with respect to the jet axis. The second line in Eq. 17 is an approximation for small angles. This expression remains infra-red and collinear safe for $a < 2$. In this study $a = -2$ was used. Angularity was shown to qualitatively distinguish between QCD jets and jets that originate from the two-body decays of heavy boosted objects like the W and Z bosons [21]. The two-prong hypothesis leads to the existence of a lower and upper bound on the allowed values of angularity. The lower bound, $\tau_{-2}^{min} \sim (m^{jet}/2p_T)^3$, is obtained from decay configurations in which both daughter particles are emitted at the same angle with respect to the direction of the mother particle and equally share the available energy. The upper bound, $\tau_{-2}^{max} \sim 2^{-3} R^2 m^{jet}/p_T$, is obtained when one of the decay daughters is hard and almost collinear with the mother particle, whereas the second decay daughter is soft and emitted at the largest allowed angle. The angle between these daughters and the direction of flight is limited by the size parameter R .

2.5.3 Planar Flow

Planar flow [20, 21] describes the way energy is deposited on the plane perpendicular to the jet axis. It is derived from the momentum tensor

$$I_e^{kl} = \frac{1}{m_{jet}} \sum_i E_i \frac{p_{i,k}}{E_i} \frac{p_{i,l}}{E_i}, \quad (18)$$

where the sum runs over the jet constituents, E_i is the constituent energy, and $p_{i,k}$ is the k^{th} component of the constituent transverse momentum relative to the jet axis. Planar flow is defined by

$$P_f = 4 \frac{\det(I_e)}{\text{tr}(I_e)^2} = \frac{4\lambda_1\lambda_2}{(\lambda_1 + \lambda_2)^2}, \quad (19)$$

where $\lambda_{1,2}$ are the eigenvalues of I_e . The planar flow of events with a two-prong nature, including massive and boosted ones, should be very small since the two out-coming jets are, by definition, contained in a plane and their energy deposits in the calorimeter will always be on a straight line. This ideal configuration, therefore, has a planar flow value of zero. The soft contribution inside the jet shifts the expected value of the planar flow to higher than zero value.

On the other hand, consider a hadronic decay of a boosted top quark, in which the b quark and the $q\bar{q}$ pair are clustered together in a single jet. The three-pronged energy deposits are unlikely to be on a straight line, therefore, giving rise to relatively high planar flow values, with a maximum value of 1 when $\lambda_1 = \lambda_2$ in Eq. 19. This makes planar flow a good handle for separating two-prong decays from three-prong decays, *e.g.* in boosted top quark searches.

3 Results

3.1 Pileup Correction

The mean jet mass shift due to pileup as calculated by the complementary cone technique, as a function of the raw jet mass in both data and MC is shown in Fig. 17. The data is shown to roughly follow an expected $1/m^{jet}$ behavior, as predicted in Eq. 14. Energy flow from MI creates an average shift in the jet mass of $3 - 4 \text{ GeV}/c^2$. This first order approximation assumes that the mass shift is small compared to the mass itself, and therefore not valid at low jet mass.

A comparison between the jet mass distribution for events with a single primary vertex, and for events with many vertices, both before and after the application of the mass correction, is shown in Fig. 18. After applying this correction, the average jet mass difference between the jets in single and multi-vertex events is reduced to $\sim 2 \text{ GeV}/c^2$, and the low mass peaks almost coincide. This residual difference may be due to the fact that our correction procedure does not take into account the relatively rare cases where the underlying event and/or multiple interactions produces a large shift in jet mass.

The small number of events left after imposing the requirement of having a high p_T jet within a certain high mass range did not allow the separation of the underlying event contribution in the cases of angularity and planar flow. The corrections for these observables were also found to be relatively small. No MI corrections were therefore applied to these observables. The results, however, do fit the predicted approximations reasonably well, as shown in Figs. 19 and 20.

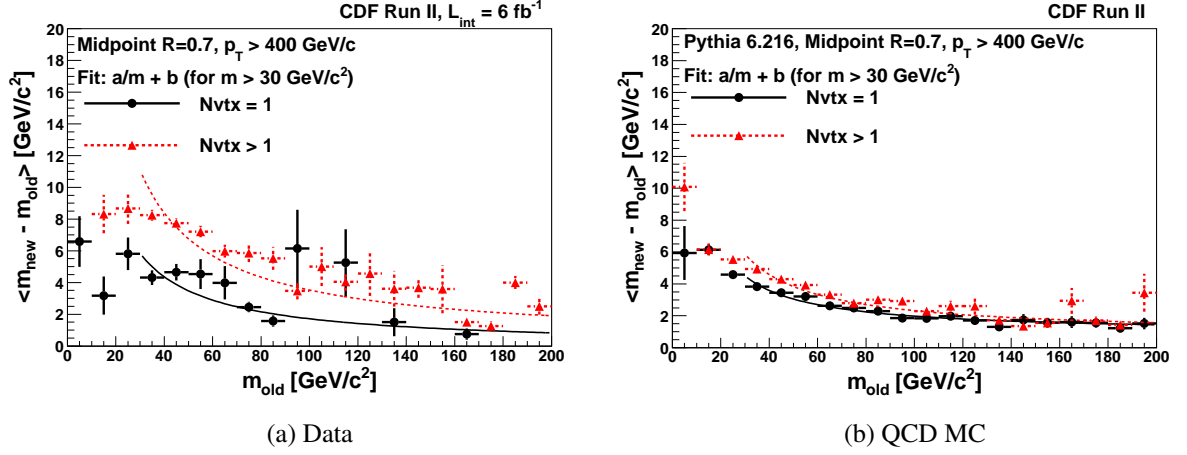


Figure 17: The average shift in the reconstructed jet mass due to underlying event (black points) and due to both underlying event and multiple interactions (dashed red points) for selected jets as a function of the raw jet mass m_{old} . Also shown are the parametrizations of these corrections (solid line for underlying event and dotted line for underlying event and multiple interactions) used for the correction. [64]

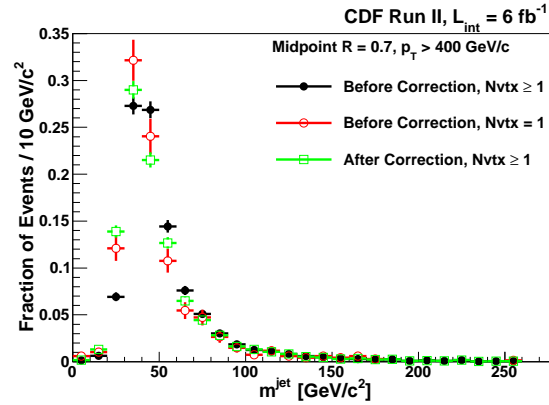


Figure 18: The jet mass distribution for high p_T jets for events with $N_{\text{vtx}} = 1$ (open red circles) and for events with $N_{\text{vtx}} \geq 1$ before (black points) and after (green open squares) the multiple interactions correction [12]

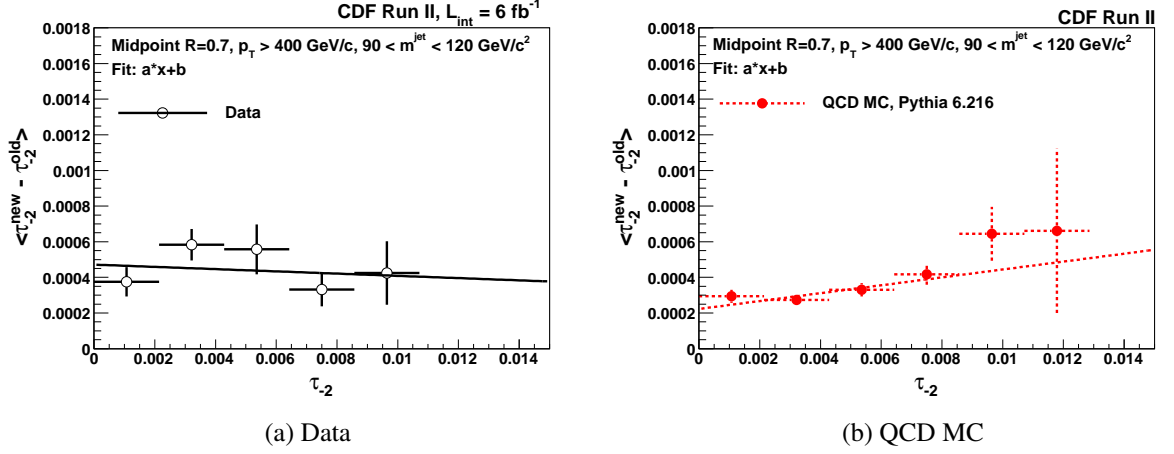


Figure 19: The average shift in the reconstructed jet angularity in data (a) and MC (b) due to underlying event and multiple interactions for selected jets as a function of the raw angularity. Also shown are the parametrizations of these corrections (straight lines) [64]

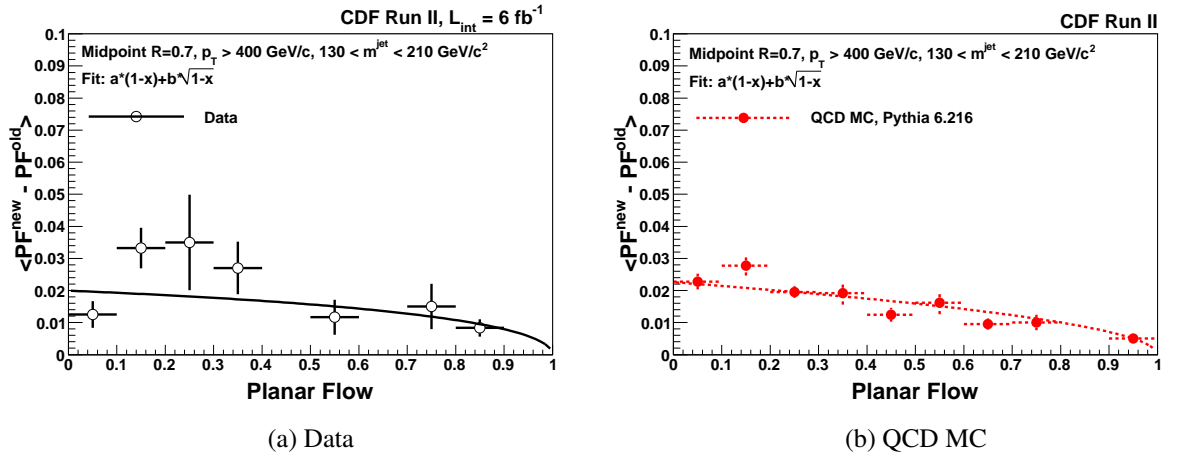


Figure 20: The average shift in the reconstructed jet planar flow in data (a) and MC (b) due to underlying event and multiple interactions for selected jets as a function of the raw planar flow. Also shown are the parametrizations of these corrections (lines of the form $a(1-x) + b\sqrt{1-x}$) [64]

3.2 Jet Substructure Observables

3.2.1 Jet Mass

A comparison of the jet mass distribution for a cone size $R = 0.7$ with the analytic predictions of the jet function (Sec. 2.5.1) is shown in Fig. 21. This is done in the region in which the approximation is valid, namely, for jet masses above $70 \text{ GeV}/c^2$ and up to $280 \text{ GeV}/c^2$. The analytical prediction for quark jets describes approximately the shape of the distribution of jets but tends to over-estimate the rate for jet masses from 130 to $200 \text{ GeV}/c^2$. According to a pQCD prediction, indeed $\sim 80\%$ of these highly boosted Tevatron jets arise from quarks [69]. PYTHIA describes the data with reasonable accuracy. Furthermore, as shown in the inset plot, a good agreement is observed between the Midpoint and the *anti- k_t* algorithms results. Since the jet mass is a higher order in pQCD effect (the bare quark or gluon carry negligible or zero mass) this is a very interesting observation considering the fact that Midpoint is shown to be IR_{3+1} [2] whereas *anti- k_t* is IR-safe. Moreover, one can observe that the Midpoint with Search Cones algorithm, shown to be “only” IR_{2+1} , tends to produce more massive jets. This is another indication to the level of sensitivity the jet mass has to IR effects in highly boosted massive jets.

For completeness, Fig. 22 shows the jet mass distribution at the high mass region for $R = 0.4$ and $R = 1.0$. In general, the same properties were observed for all cone sizes. As expected, the larger the cone size, the more energy is included into the jet, particularly at larger distances from the jet axis. Thus average jet mass increases with cone size. This effect is also shown in Fig. 23.

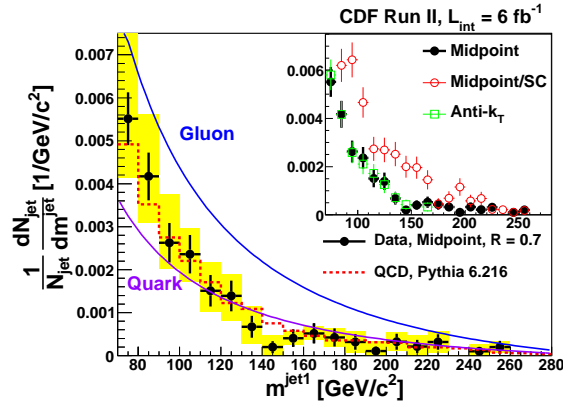


Figure 21: The normalized jet mass distribution for Midpoint jets with $p_T > 400 \text{ GeV}/c$ and $|\eta| \in (0.1, 0.7)$. The uncertainties shown are statistical (black lines) and systematic (yellow bars). The theory predictions for the jet function for quarks and gluons are shown as solid curves and have an estimated uncertainty of $\sim 30\%$. Also shown is the PYTHIA MC prediction (red dashed line). The inset compares Midpoint (full black circles), Midpoint with Search Cones (open red circles), and *anti-k_t* (open green squares) jets [12]

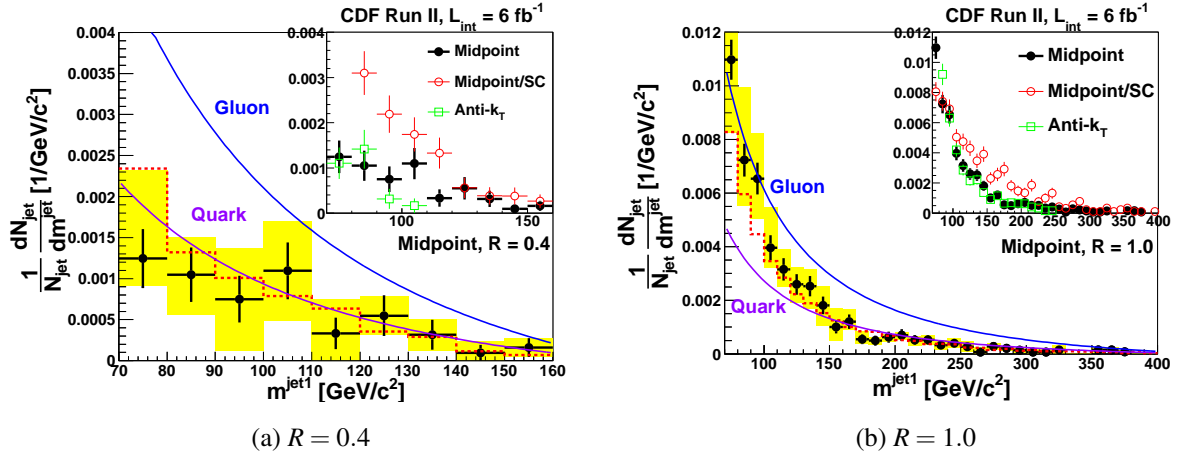


Figure 22: The normalized jet mass distribution for Midpoint jets with $p_T > 400 \text{ GeV}/c$ and $R = 0.4$, $|\eta| \in (0.1, 0.7)$ (a) and $R = 1.0$, $|\eta| < 0.7$ (b). The uncertainties shown are statistical (black lines) and systematic (yellow bars). The theory predictions for the jet function for quarks and gluons are shown as solid curves and have an estimated uncertainty of $\sim 30\%$. Also shown is the PYTHIA MC prediction (red dashed line). The inset compares Midpoint (full black circles), Midpoint with Search Cones (open red circles), and $\text{anti-}k_T$ (open green squares) jets [12]

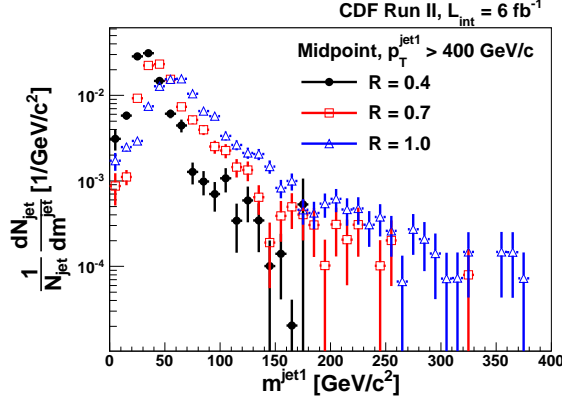


Figure 23: The jet mass distributions with cone sizes $R = 0.4$, 0.7 and 1.0 (black points, open red squares and open blue triangles, respectively) in the data sample for all jets with $p_T > 400 \text{ GeV}/c$

3.2.2 Angularity

The angularity distribution is shown in Fig. 24 for $R = 0.4$ and $R = 0.7$. The jets were required in this case to have a mass between 90 to $120 \text{ GeV}/c^2$. The upper edge of this window is limited by the number of highly boosted massive jets available. This mass window was a compromise between having enough statistics and avoiding contamination from W and Z bosons. This distribution indeed lies between the two expected limits (Sec. 2.5.2) and hence corroborate the two-prong hypothesis. PYTHIA prediction agrees nicely with the data. Furthermore, as in the jet mass case, an agreement is observed between the Midpoint results and those of $anti-k_t$. Finally, the Search Cone variant tends to create jets wider than expected by pQCD as seen by the tail of the relevant distribution extending beyond τ_{-2}^{max} .

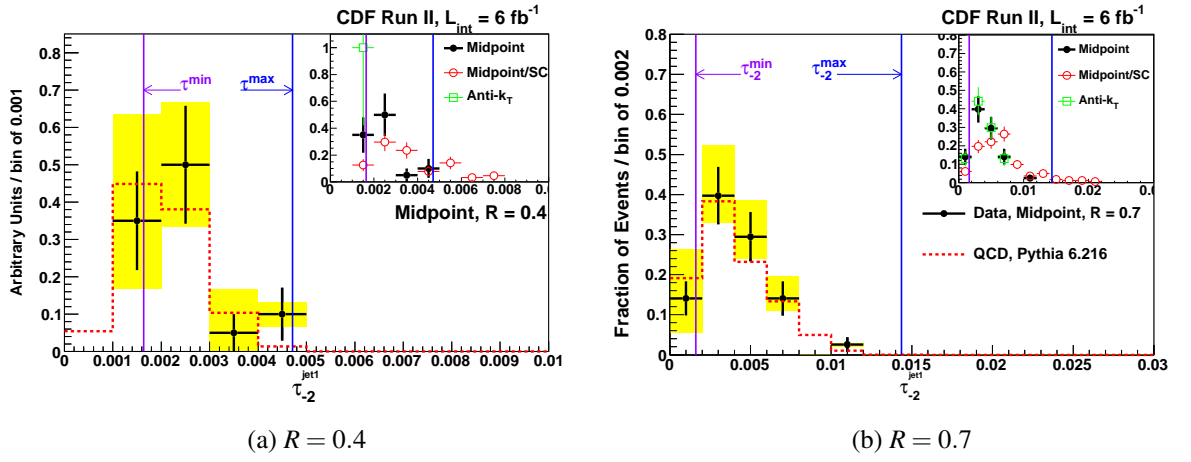


Figure 24: The angularity distribution for Midpoint jets of $R = 0.4$ (a) and $R = 0.7$ (b) with $p_T > 400 \text{ GeV}/c$, $|\eta| \in (0.1, 0.7)$, and $m^{jet1} \in (90, 120) \text{ GeV}/c^2$. Also shown are the PYTHIA calculation (red dashed line) and the pQCD kinematic endpoints. The inset compares the distributions for Midpoint (full black circles), Midpoint with Search Cones (open red circles), and $anti-k_t$ (open green squares) jets [12]

3.2.3 Planar Flow

The planar flow distributions for jets of all masses is shown in Fig. 25. This distribution is dominated by low mass jets, in which soft gluon radiation is spread across the face of the jet, increasing the value of its planar flow. The MC does not seem to nicely reproduce the data in this sample. Fig. 26 shows the planar flow distributions for highly boosted jets in a mass window of 130 to 210 GeV/c^2 , a range relevant for pQCD studies as well as search of boosted tops. The data exhibits the expected QCD like behavior and peaks at a low planar flow value. This should be compared to the $t\bar{t}$ MC sample that exhibits a flatter distribution. Once again Midpoint and $anti-k_t$ are in good agreement. This is particularly interesting in the case of planar flow which deals with three-body configurations and considering the IR_{3+1} ranking of Midpoint. The $R = 0.7$ distribution suffers from low number of events due to the high mass cut. The picture is perhaps more clear in the $R = 1.0$ distribution in which one can also observe that the Search Cone variant tends to produce more planar jets.

3.3 Systematic Uncertainties

3.3.1 Internal Jet Energy Scale

The precision of the Jet Energy Scale (JES) at CDF is well established [55]. The uncertainties on the JES were measured to be less than 3% for jets with $p_T > 400 GeV/c$ as shown in Fig. 27. This result significantly constraints any possible variation in energy response across the face of the jet, one that can have an effect on jet substructure observables. This section describes the procedure that was used to assess the uncertainties of the jet mass scale due to possible such variations.

The ratio of charged p_T to calorimeter transverse energy was measured in three concentric rectangular regions in $\eta - \phi$ space, centered around the jet axis.

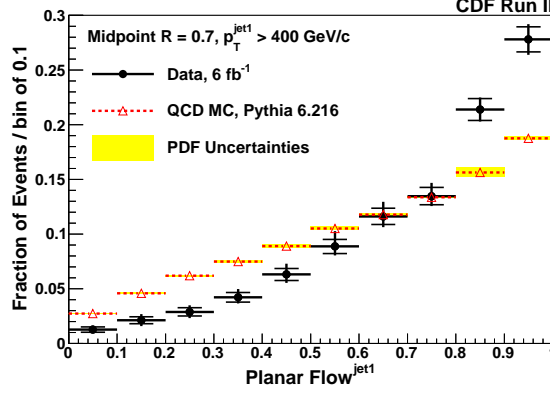


Figure 25: The planar flow distributions for Midpoint jets with $p_T > 400 \text{ GeV}/c$, $|\eta| \in (0.1, 0.7)$, and no explicit mass requirement, in data (black circles) and pythia QCD (red triangles). All distributions have been separately normalized to unity [12]

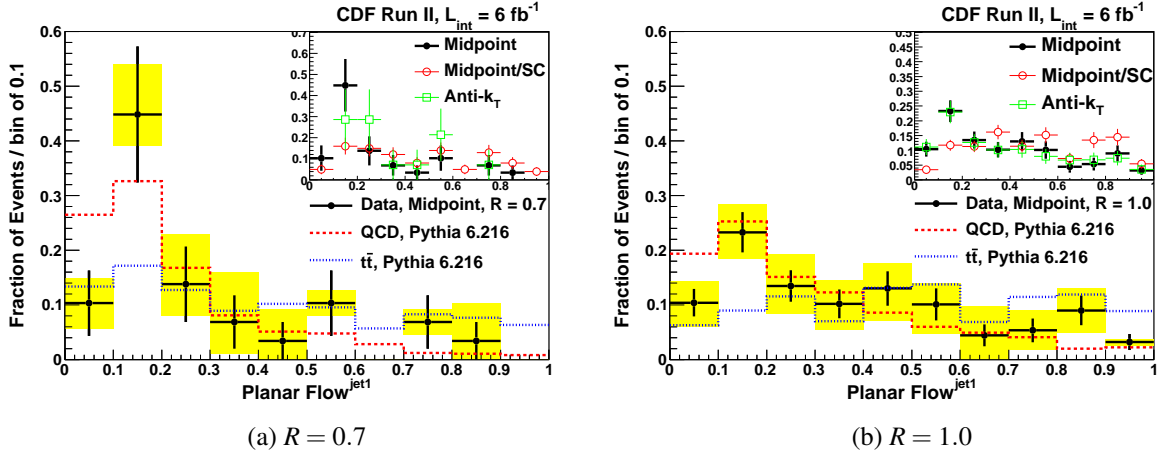


Figure 26: The planar flow distributions for Midpoint jets with $p_T > 400 \text{ GeV}/c$, $R = 0.7$, $|\eta| \in (0.1, 0.7)$ (a) and $R = 1.0$, $|\eta| < 0.7$ (b), and $m^{jet1} \in (130, 210) \text{ GeV}/c^2$. Also shown are the pythia QCD (red dashed line) and $t\bar{t}$ (blue dotted line) jets, as well as the results from the three jet algorithms (inset). All distributions have been separately normalized to unity [12]

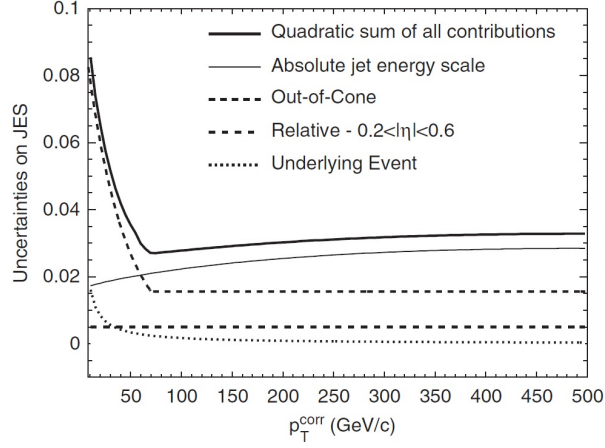


Figure 27: Systematic uncertainties as a function of the corrected jet p_T [55]

The definition of these regions is as follows:

- Region 1 is formed of 4 towers in η and 2 towers in ϕ with the one of the four innermost towers closest to the jet centroid.
- Region 2 is formed of 8 towers in η and 4 towers in ϕ centered on Region 1 and excluding it.
- Region 3 is formed of 12 towers in η and 6 towers in ϕ centered on Region 1 and excluding the interior two regions.

These regions are shown schematically in Fig. 28 overlaid by a jet cone of radius 0.7 for illustration purposes.

For each region i , the following ratio was measured:

$$f_{track/cal,i} \equiv (p_T/E_T)_i = \frac{\sum_{\substack{tracks \\ in\ region}} p_T}{\sum_{\substack{towers \\ in\ region}} E_T} \quad (20)$$

in which the numerator is the sum of the p_T of all charged tracks reconstructed in the COT and intercept the given region when projected to the plane of the

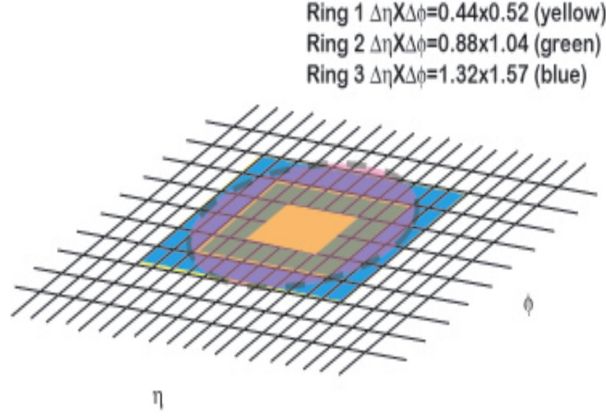


Figure 28: A schematic of the three calorimeter regions used in the verification of the internal energy calibration within the jet. The dashed circle represents a cone of radius $R = 0.7$

calorimeter. The reconstruction efficiency of very soft tracks is unknown and may vary in different regions of the jet. Therefore, the charged tracks were required to have $p_T > 1 \text{ GeV}/c$. The denominator is the sum of the transverse energy deposited in each calorimeter tower in the region. In this study, the number of reconstructed vertices, N_{vtx} was required to be equal to one in order to minimize the effect of multiple interactions. The distributions of this ratio for data and MC in the three regions are shown in Figs. 29, 30, and 31.

It was already shown in another CDF study [70] that the tracking efficiency drops as one goes nearer the core of the jet. This effect is also seen here as the mean ratio becomes smaller when going inwards from region 3 to region 1.

In order to assess the systematic effect of local energy scale variations one can let the JES become a free parameter (JES_i) in the three regions defined above. Using the 3% uncertainty of the global JES one can then write Eq. 21:

$$0.97 E_T^{average} < JES_1 \rho_1 A_1 + JES_2 \rho_2 A_2 + JES_3 \rho_3 A_3 < 1.03 E_T^{average} \quad (21)$$

where $E_T^{average}$ is the average jet transverse energy, ρ_i is the average energy density

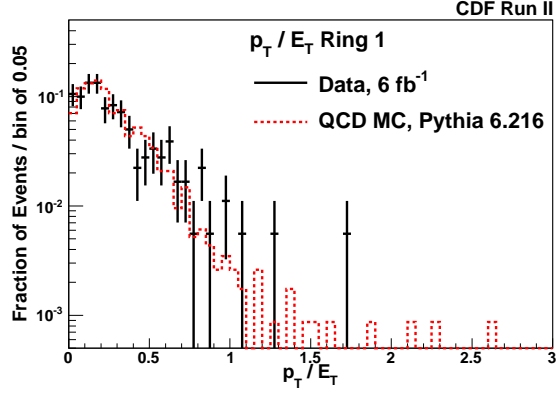


Figure 29: The distribution for region 1 of the ratio of charged track p_T extrapolated to the region and the calorimeter transverse energy in the region, for selected jets with $p_T \in (400, 500)$ and $|\eta| \in (0.1, 0.7)$. The MC prediction for this distribution is given by the red dashed line. $N_{vtx} = 1$ is required

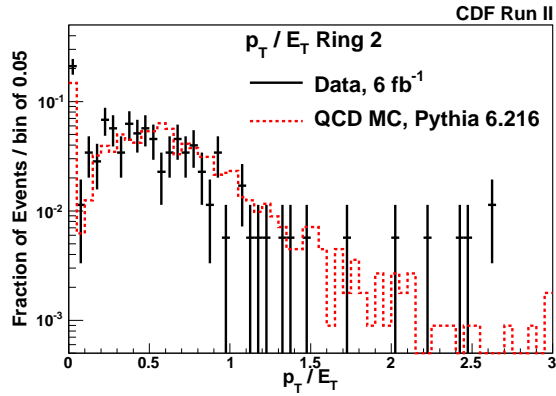


Figure 30: The distribution for region 2 of the ratio of charged track p_T extrapolated to the region and the calorimeter transverse energy in the region, for selected jets with $p_T \in (400, 500)$ and $|\eta| \in (0.1, 0.7)$. The MC prediction for this distribution is given by the red dashed line. $N_{vtx} = 1$ is required

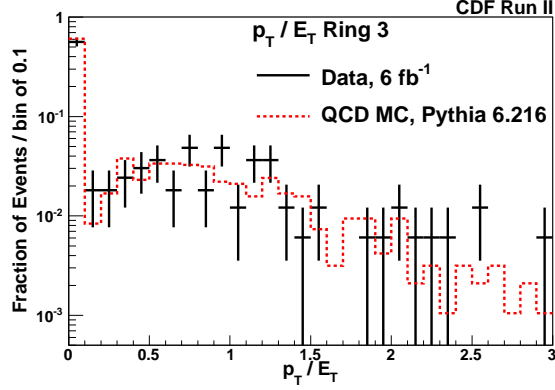


Figure 31: The distribution for region 3 of the ratio of charged track p_T extrapolated to the region and the calorimeter transverse energy in the region, for selected jets with $p_T \in (400, 500)$ and $|\eta| \in (0.1, 0.7)$. The MC prediction for this distribution is given by the red dashed line. $N_{vtx} = 1$ is required

for each region, and A_i is the area in the $\eta - \phi$ space of each region relative to the total area. The measured values of these parameters are shown in Table 4. Most

	Region 1	Region 2	Region 3
Relative Area (A_i)	0.111	0.333	0.555
Mean E_T density ($\rho_i[GeV/\Delta\eta\Delta\phi]$)	1744	33.7	1.5
Mean $f_{track/cal,i}$ (data)	0.176 ± 0.008	0.436 ± 0.012	0.815 ± 0.020
Mean $f_{track/cal,i}$ (MC)	0.150 ± 0.005	0.538 ± 0.006	0.790 ± 0.012
Fractional E_T in region i	0.941	0.055	0.004

Table 4: The relative areas of each calorimeter region, the average E_T densities in the three regions for jets with $p_T \in (400, 500)$ GeV/c and $|\eta| \in (0.1, 0.7)$, and the mean ratio of tracking and calorimeter response for the data jets and the MC jets. The last line shows the average E_T deposited in each region for an average jet in this sample

of the jet's energy is deposited in the Region 1, for which the MC and data are in reasonable agreement. The overall energy scale uncertainty of 3% therefor places the strongest single constraint on JES_1 . On average, Region 1 captures 94% of the total energy of the leading jet in the sample. Thus the uncertainty of JES_1 from the jet energy systematic uncertainty is at most $0.03/0.94 = 0.032$. Next, the two following double ratios are defined:

$$R_{1i} \equiv \frac{f_{track/cal,1}^{MC}/f_{track/cal,1}^{data}}{f_{track/cal,i}^{MC}/f_{track/cal,i}^{data}} \quad (22)$$

where $i = 2, 3$. From Table 4 one can calculate $R_{12} = 0.69$ and $R_{13} = 0.88$ with an estimated uncertainty of ± 0.1 .

The uncertainty on the jet mass was estimated using a model of a jet, namely, using a uniformly distributed energy over a circular area in $\eta - \phi$ taking into account the calorimeter segmentation, such that the average E_T densities are as shown in the table, and the total mass is $115 \text{ GeV}/c^2$. An iterative search was done to find the shift in the energy scale of region 1, consistent with a drop in the energy scale of region 2 and region 3 of as much as one standard deviation. During this search it was required that the jet energy scale remained within the 3% uncertainty range. The maximal mass shift found while keeping these constraints was found to be $10 \text{ GeV}/c^2$.

3.3.2 Energy Flow from Multiple Interactions

As was shown in Sec. 3.1, energy flow from MI creates an average shift in the jet mass of $3 - 4 \text{ GeV}/c^2$, for jets with masses above $70 \text{ GeV}/c^2$ and a cone size of $R = 0.7$. The uncertainty on the mass shift was conservatively set to be $2 \text{ GeV}/c^2$, which is half the value of the average correction.

3.3.3 Uncertainties on the PYTHIA Prediction

Uncertainties arising from the choice of PDF and renormalization scale are estimated using the eigenvector approach [71]. The MC events are re-weighted after shifting each of the eigenvectors and choices of scale describing the PDF parameterization by one standard deviation up and down. One then plots the normalized distributions of the relevant observables using these weights. There is one distribution for each PDF parameter shift. In each of these distributions, each bin is slightly shifted due to the re-weighting of the events. The shifts that arise in each bin is summed in quadrature to give the PDF uncertainty in that bin. These uncertainties were approximately 10% for the jet mass distributions and 5% for angularity and planar flow.

3.3.4 Unfolding Related Uncertainties

The unfolding factors were determined using a fit on a MC sample with finite statistics. This is manifested by the error bars in Fig. 16. The statistical uncertainty introduces uncertainties on the resulting factors, which can be estimated by the uncertainties of the fit. Uncertainties introduced by the choice of model for jet fragmentation and hadronization can be studied by the change in the unfolding factors when these are calculated before and after fragmentation and hadronization, *i.e.* at the partonic level and the hadronic level. These changes were found to be on the order of 10%. An uncertainty of an additional 10% is arising from the uncertainty in the jet energy scale and was estimated by calculating the unfolding factors after re-selection of jets according to a shifted energy scale.

The above results are all consistent with pQCD prediction. According to pQCD the very energetic parton that was produced in the hard scattering process gives off a gluon that carries roughly half of the parton's energy. The mass and angularity distributions are clearly in good agreement with this hypothesis. In the next section one inverts the anti-top-quark cuts and looks for the production of boosted $t\bar{t}$ events.

4 Search for Boosted Top Quarks

4.1 Methods

The production of $t\bar{t}$ pairs was a background for the QCD studies and an effort was made to reject it in Sec. 2.2.4 by a few dedicated cuts. By inverting the top-quark rejection cuts and utilizing the acquired understanding of jet substructure, particularly the jet mass, one can get a $t\bar{t}$ enriched sample and carry out a boosted $t\bar{t}$ study. Such a study is very interesting since new physics may be linked to the top quark and since an anomaly in the forward-backward asymmetry of $t\bar{t}$ events, in particular boosted ones, has been reported by CDF [76]. The main background for $t\bar{t}$ events is, as expected, QCD processes. In order to cope with this hard-to-estimate background, a data-driven background estimation technique was developed.

The typical cone size of a fully hadronic decay of a boosted top is $\sim \frac{m}{p_T/2} \sim 1$. In this part of the study, due to the relatively low available beam energy at the Tevatron, Midpoint jets with $R = 1.0$ were used to increase the efficiency for collecting the full hadronic decay of the boosted top quark inside a single jet. The pseudorapidity range was changed to $|\eta| < 0.7$ to increase the acceptance. There were 4230 events in the data having such leading jets that also satisfy $p_T > 400 \text{ GeV}/c$, as well as a reasonable recoil jet with $p_T > 100 \text{ GeV}/c$.

In principle, for QCD events, one does not expect to find a positive correlation between the masses of the two leading jets in a typical dijet event. On the other hand, in hadronic $t\bar{t}$ decays, one expects that the presence of a massive leading jet is highly correlated with the presence of a secondary massive jet. These correlations are shown for the data and $t\bar{t}$ MC sample in Fig. 32. An upper cut on the S_{MET} was applied at this stage to reject semileptonic $t\bar{t}$ events. The S_{MET} was defined in Eq. 10.

Hadronic $t\bar{t}$ events are thus expected to populate a region in phase space in which both leading jets have high mass, the so called signal region. Control regions in which one or more of the two leading jets have low mass are also defined. Using these control regions, one can predict the number of expected QCD events in the signal region. The definitions of the signal region, D , and the control samples, A , B , and C , are given in Table 5. The predicted number of QCD events in the signal region shown in the bottom line of the table is given by:

$$N_D^{pred} = \frac{N_B N_C}{N_A R_{mass}}, \quad (23)$$

where N_X is the number of events in region X and $R_{mass} \equiv (N_B N_C)/(N_A N_D)$ is a correction factor that was introduced in [72] and represents the residual correlations that might exist between the m^{jet2} and m^{jet1} in QCD events. This ratio was calculated with several different NLO QCD event generators giving rise to R_{mass} values in the range of 0.76 to 0.89 [73]. A value calculated using POWHEG [57, 58, 59, 60] of $R_{mass} = 0.89$ was used in this study.

One may attempt applying a similar technique focusing now on semileptonic $t\bar{t}$ events. In QCD events, no correlation is expected between the missing energy, which originates mostly from instrumental effects, and the leading jet mass. In semileptonic $t\bar{t}$ events, on the other hand, one expects such a correlation. These correlations are shown in Fig. 33 for data and simulated $t\bar{t}$ events.

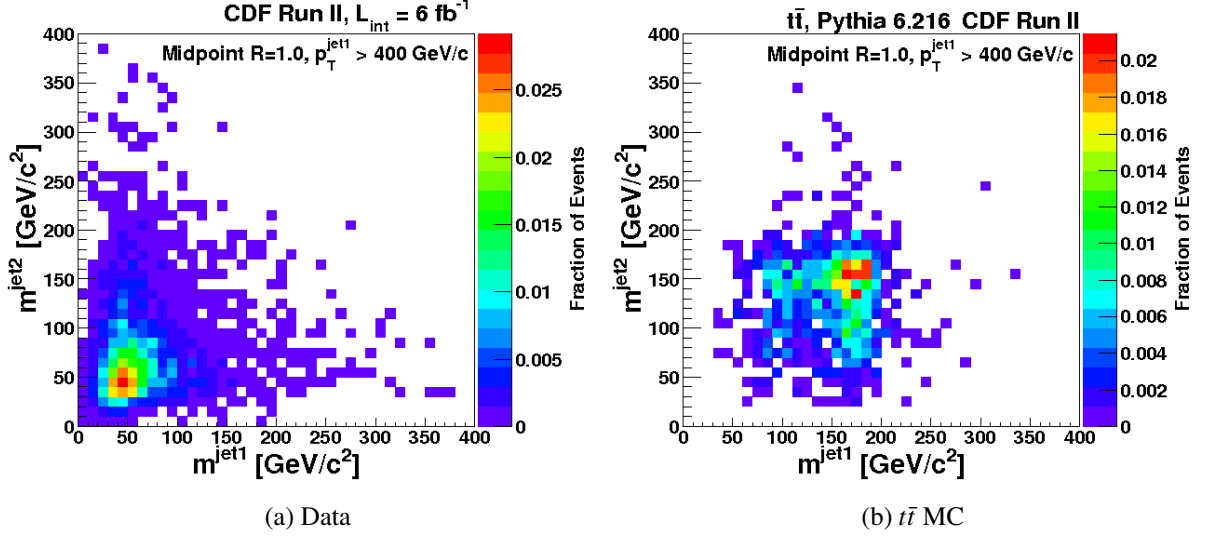


Figure 32: The distribution of m^{jet2} versus m^{jet1} for selected events in data (a) and in a $t\bar{t}$ MC (b) with at least one jet with $p_T > 400 \text{ GeV}/c$ and $|\eta| < 0.7$ using $R = 1.0$ Midpoint cones. All events are required to have $S_{MET} < 4 \text{ GeV}^{1/2}$ and a second jet satisfying $p_T > 100 \text{ GeV}/c$ [22]

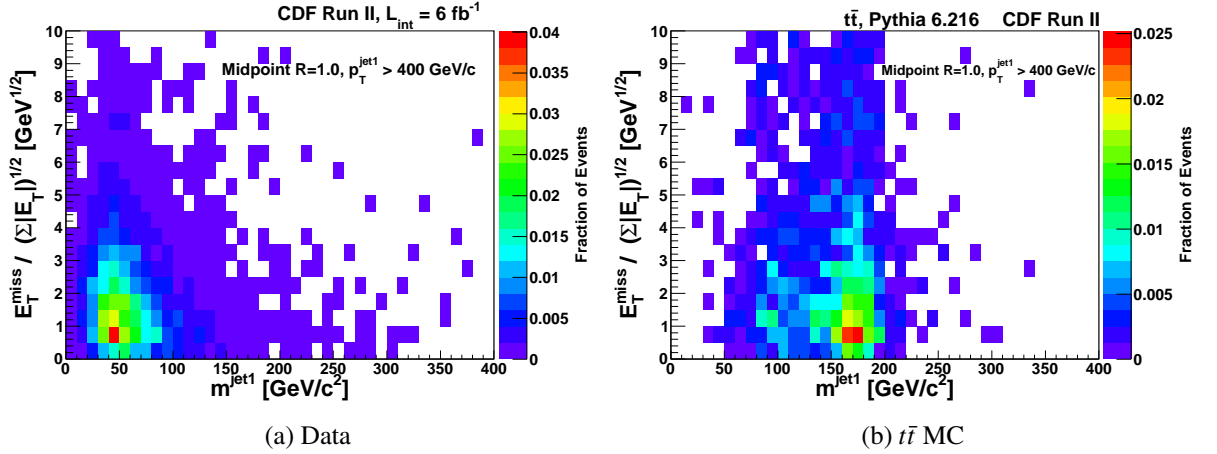


Figure 33: The distribution of S_{MET} versus m^{jet1} for selected events in data (a) and in a $t\bar{t}$ MC (b) with at least one jet with $p_T > 400 \text{ GeV}/c$ and $|\eta| < 0.7$ using $R = 1.0$ Midpoint cones [22]

4.2 Results

The definitions of the control regions and the number of counted events in each region are given in Table 5 for the fully hadronic case (including R_{mass} as described in Sec. 4), and in Table 6 for the semileptonic case. The tables also show the expected number of $t\bar{t}$ events in each region.

Region	m^{jet1} (GeV/c^2)	m^{jet2} (GeV/c^2)	Data (Events)	$t\bar{t}$ MC (Events)
A	(30,50)	(30,50)	370	0.00
B	(130,210)	(30,50)	47	0.08
C	(30,50)	(130,210)	102	0.01
D (signal)	(130,210)	(130,210)	31	3.03
Predicted QCD in D			14.6 ± 2.76	

Table 5: The observed number of events in the three control regions (A, B, and C) used to predict the background rate in the signal region (region D) in the fully hadronic channel. The $t\bar{t}$ MC event rates in each region are also shown [22]

The acceptances for $t\bar{t}$ events in the all-hadronic and in the semileptonic cases are expected a-priori to be comparable. Thus the two channels were combined and used to set an upper limit on the rate of $t\bar{t}$ production for top quarks with $p_T > 400 \text{ GeV}/c$. A 95% C.L. limit on the $t\bar{t}$ cross section was calculated, using the CL_s approach that performs a frequentist upper limit calculation, using pseudo experiments to combine statistical and systematic uncertainties [74, 75]. The resulting 95% C.L. upper limit on the cross section is 38 fb at 95% C.L.. This is approximately an order of magnitude higher than the estimated SM rate, and is limited by the size of the expected QCD background rates and the systematic uncertainty on this number. It is, however, the most stringent limit on boosted top

Region	m^{jet1} (GeV/c^2)	S_{MET} ($\sqrt{GeV/c^2}$)	Data (Events)	MC (Events)
A	(30, 50)	(2, 3)	256	0.01
B	(130, 210)	(2, 3)	42	1.07
C	(30, 50)	(4, 10)	191	0.03
D (signal)	(130, 210)	(4, 10)	26	1.90
Predicted QCD in D			31.3 ± 8.1	

Table 6: The observed number of events in the three control regions (A, B, and C) used to predict the background rate in the signal region (region D) in the semileptonic channel. The $t\bar{t}$ MC event rates in each region are also shown [22]

quark production to date. The expected limit was calculated by using the background estimated from the data-driven technique and assuming an observation of $t\bar{t}$ events at the expected level of 4.9 events using both channels. The expected upper limit is $33fb$ at 95% C.L., which is somewhat lower than the observed limit since a small excess of events above the expected signal plus background is observed. It is interesting to set a limit on the fully hadronic channel, as this creates a selection that is sensitive to pair production of two massive objects near the mass of the top quark that decay primarily hadronically. As the main interest in this case is in beyond-SM contributions to this final state, the background estimate for the expected $t\bar{t}$ contribution of 3 ± 0.8 events is included. Taking out the top quark hadronic branching fraction of $4/9$, the upper limit is $20fb$ at 95% C.L.

4.3 Uncertainties

The largest source of uncertainty affecting the search for boosted top quarks is the uncertainty on the jet mass which is $\pm 10.2 GeV/c^2$ for massive jets, combining

both internal JES uncertainties and MI corrections uncertainties. In the ABCD method described in Sec. 4, the high mass window was shifted up and down by this mass uncertainty. This yielded an uncertainty of 30% on the combined background estimation.

The top quark acceptance is affected by the jet energy scale uncertainty. Changing the jet p_T by 3% [55] results in a $\pm 24\%$ change in the acceptance.

An additional $\pm 6\%$ was added due to integrated luminosity uncertainty, and $\pm 0.3\%$ for variations in the top quark mass.

All these sources of uncertainties are assumed to be independent of each other and are added in quadrature, giving a total of $\pm 44\%$ systematic uncertainty in the predicted $t\bar{t}$ cross section.

5 Discussion

The present research project is focused on the first measurements of the substructure of highly boosted massive jets. It was done using the CDF detector at the Tevatron. Valuable knowledge was acquired on the substructure of such jets and its possible sources.

Pileup is a major impediment for many jet studies and in particular jet substructure studies. In order to cope with it, a data-driven pileup correction technique was developed and used. It was shown to reduce the average mass difference between jets in single and multi-vertex events to about $2 \text{ GeV}/c^2$ and therefore remove systematic mass shifts due to MI. The analytical prediction for the shifts in angularity and planar flow seem to match the observed ones, although the relevant data samples were very small and, therefore, the correction procedure was not applied to them.

The jet mass, angularity, and planar flow were measured for the first time for jets with $p_T > 400 \text{ GeV}/c$, using Midpoint, Midpoint with Search Cones, and *anti- k_t* jet algorithms. A good agreement between the data and the predictions of the PYTHIA event generator was observed for the jet mass distribution. At large mass, the data was found to be consistent with the two-prong hypothesis of pQCD and is well represented by the NLO QCD jet function. An agreement between the Midpoint jet algorithm and the IR-safe *anti- k_t* algorithm was observed, in contrast to the IR irregularity of the Search Cone variant of the Midpoint jet algorithm which was manifested in the form of a longer mass distribution tail.

The values of the jet's angularity fall within the expected kinematic limits τ_{-2}^{\min} and τ_{-2}^{\max} , corroborating the two-prong hypothesis. This was observed for a cone size of $R = 0.7$ in a high mass window, as well as for a lower $R = 0.4$ cone size, albeit the smaller number of events available for this cone size at larger jet masses. This observable also exhibits the importance of the IR-safety issue of the Search

Cone variant, having a distribution tail that exceeds the predicted τ_{-2}^{max} .

A planar flow distribution with no jet mass requirements applied, showed some disagreement between data and PYTHIA event generator, although the two have the same general trend of a monotonously increasing planar flow distribution. Perhaps this hints that the MC fails to model correctly the soft radiation which is ubiquitous in the sample dominated by low mass jets. At larger masses however, a better agreement was observed between data and MC. This is better discernible in the $R = 1.0$ sample which contains a larger number of events at high jet mass.

All these features are consistent with pQCD expectations emphasizing that at the LHC high p_T jets will be an excellent tool for pQCD studies.

A dedicated search for boosted top quark based mainly on jet mass yielded an upper limit of 38 fb at 95% C.L. on the cross section of $t\bar{t}$ production at high p_T . This was done combining both the fully hadronic and the semileptonic channels. The somewhat lower expected limit of 33 fb at 95% C.L. represents the fact that a small excess of events over the predicted number was observed in the fully hadronic channel. The latter channel is interesting on its own, being sensitive to BSM scenarios in which a pair of massive objects are produced and decay primarily hadronically. Assuming a given SM $t\bar{t}$ content and taking out the top quark hadronic branching fraction, an upper limit of 20 fb at 95% C.L. was set based on the fully hadronic channel alone. These results come hand in hand with recent evidence measured by CDF for mass dependent forward-backward asymmetry in top quark pair production [76]. They have already inspired interpretations of the two observations [73].

The regime of highly boosted objects becomes exceedingly relevant at the energies obtainable at the LHC [15, 16]. Understanding high p_T massive jets is a

prerequisite for current and future analyses, and in some cases these complex jets replace the resolved objects used in classic analyses. This study sets up the stage for further jet substructure investigations, SM studies and beyond-SM searches that can benefit from the knowledge obtained here and put the inspected observables into practical use.

6 References

- [1] S. Ellis, J. Huston, K. Hatakeyama, P. Loch, and M. Tonnesmann, “Jets in hadron-hadron collisions,” *Prog.Part.Nucl.Phys.* **60** (2008) 484–551, 0712.2447.
- [2] G. P. Salam, “Towards Jetography,” *Eur.Phys.J.* **C67** (2010) 637–686, 0906.1833.
- [3] P. Nath, B. D. Nelson, H. Davoudiasl, B. Dutta, D. Feldman, *et al.*, “The Hunt for New Physics at the Large Hadron Collider,” *Nucl.Phys.Proc.Suppl.* **200-202** (2010) 185–417, 1001.2693.
- [4] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, “Jet substructure as a new Higgs search channel at the LHC,” *Phys.Rev.Lett.* **100** (2008) 242001, 0802.2470.
- [5] J. Butterworth, B. Cox, and J. R. Forshaw, “ WW scattering at the CERN LHC,” *Phys.Rev.* **D65** (2002) 096014, hep-ph/0201098.
- [6] K. Agashe, A. Belyaev, T. Krupovnickas, G. Perez, and J. Virzi, “LHC Signals from Warped Extra Dimensions,” *Phys.Rev.* **D77** (2008) 015003, hep-ph/0612015.
- [7] B. Lillie, L. Randall, and L.-T. Wang, “The Bulk RS KK-gluon at the LHC,” *JHEP* **0709** (2007) 074, hep-ph/0701166.
- [8] J. M. Butterworth, J. R. Ellis, A. R. Raklev, and G. P. Salam, “Discovering baryon-number violating neutralino decays at the LHC,” *Phys.Rev.Lett.* **103** (2009) 241803, 0906.0728.

- [9] **D0** Collaboration, V. Abazov *et al.*, “Subjet multiplicity of gluon and quark jets reconstructed with the k_T algorithm in $p\bar{p}$ collisions,” *Phys.Rev.* **D65** (2002) 052008, hep-ex/0108054.
- [10] **CDF** Collaboration, D. Acosta *et al.*, “Study of jet shapes in inclusive jet production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV,” *Phys.Rev.* **D71** (2005) 112002, hep-ex/0505013.
- [11] **CDF** Collaboration, T. Aaltonen *et al.*, “Study of Substructure of High Transverse Momentum Jets Produced in Proton-Antiproton Collisions at $\sqrt{s} = 1.96$ TeV,” *Phys.Rev.* **D85** (2012) 091101, 1106.5952.
- [12] **CDF** Collaboration, T. Aaltonen *et al.*, “The Substructure of High Transverse Momentum Jets Observed by CDF II,” *CDF/PUB/JET/PUBLIC/10199* (June, 2011).
- [13] **ATLAS** Collaboration, G. Aad *et al.*, “Study of Jet Shapes in Inclusive Jet Production in pp Collisions at $\sqrt{s} = 7$ TeV using the ATLAS Detector,” *Phys.Rev.* **D83** (2011) 052003, 1101.0070.
- [14] **ATLAS** Collaboration, “Measurement of Jet Mass and Substructure for Inclusive Jets in $\sqrt{s} = 7$ TeV pp Collisions with the ATLAS Experiment,” *ATLAS-CONF-2011-073* (May, 2011).
- [15] **ATLAS** Collaboration, G. Aad *et al.*, “ATLAS measurements of the properties of jets for boosted particle searches,” 1206.5369.
- [16] **ATLAS** Collaboration, G. Aad *et al.*, “Jet mass and substructure of inclusive jets in $\sqrt{s} = 7$ TeV pp collisions with the ATLAS experiment,” *JHEP* **1205** (2012) 128, 1203.4606.

- [17] G. C. Blazey, J. R. Dittmann, S. D. Ellis, V. Elvira, K. Frame, *et al.*, “Run II jet physics,” hep-ex/0005012.
- [18] S. Ellis, J. Huston, and M. Tonnesmann, “On building better cone jet algorithms,” *eConf* **C010630** (2001) P513, hep-ph/0111434.
- [19] M. Cacciari, G. P. Salam, and G. Soyez, “The Anti-k(t) jet clustering algorithm,” *JHEP* **0804** (2008) 063, 0802.1189.
- [20] L. G. Almeida, S. J. Lee, G. Perez, I. Sung, and J. Virzi, “Top Jets at the LHC,” *Phys.Rev.* **D79** (2009) 074012, 0810.0934.
- [21] L. G. Almeida, S. J. Lee, G. Perez, G. F. Sterman, I. Sung, *et al.*, “Substructure of high- p_T Jets at the LHC,” *Phys.Rev.* **D79** (2009) 074017, 0807.0234.
- [22] **CDF** Collaboration, T. Aaltonen *et al.*, “A Search for Boosted Top Quarks by CDF II,” *CDF/ANAL/TOP/PUBLIC/10234* (June, 2011).
- [23] S. L. Glashow, “Partial-symmetries of weak interactions,” *Nuclear Physics* **22** (1961), no. 4, 579 – 588.
- [24] S. Weinberg, “A Model of Leptons,” *Phys. Rev. Lett.* **19** (Nov, 1967) 1264–1266.
- [25] A. Salam, “Weak and electromagnetic interactions, in Elementary particle physics: relativistic groups and analyticity,” *Proceedings of the eighth Nobel symposium N. Svartholm, ed., p. 367. Almqvist & Wiskell* (1968) 367.
- [26] G. ’t Hooft and M. Veltman, “Regularization and Renormalization of Gauge Fields,” *Nucl.Phys.* **B44** (1972) 189–213.

- [27] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons,” *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323.
- [28] P. W. Higgs, “Broken symmetries, massless particles and gauge fields,” *Phys.Lett.* **12** (1964) 132–133.
- [29] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons,” *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509.
- [30] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, “Global Conservation Laws and Massless Particles,” *Phys. Rev. Lett.* **13** (Nov, 1964) 585–587.
- [31] P. W. Higgs, “Spontaneous Symmetry Breakdown without Massless Bosons,” *Phys. Rev.* **145** (May, 1966) 1156–1163.
- [32] T. W. B. Kibble, “Symmetry Breaking in Non-Abelian Gauge Theories,” *Phys. Rev.* **155** (Mar, 1967) 1554–1561.
- [33] **ATLAS Collaboration** Collaboration, G. Aad *et al.*, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” 1207.7214.
- [34] **CMS Collaboration** Collaboration, S. Chatrchyan *et al.*, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Phys.Lett.B* (2012) 1207.7235.
- [35] K. A. Brueckner, “Meson-Nucleon Scattering and Nucleon Isobars,” *Phys. Rev.* **86** (Apr, 1952) 106–109.
- [36] O. W. Greenberg, “Spin and Unitary-Spin Independence in a Paraquark Model of Baryons and Mesons,” *Phys. Rev. Lett.* **13** (Nov, 1964) 598–602.

- [37] M. Y. Han and Y. Nambu, “Three-Triplet Model with Double SU(3) Symmetry,” *Phys. Rev.* **139** (Aug, 1965) B1006–B1010.
- [38] P. Skands, “Introduction to QCD,” 1207.2389.
- [39] S. Bethke, “The 2009 World Average of $\alpha(s)$,” *Eur.Phys.J.* **C64** (2009) 689–703, 0908.1135.
- [40] J. C. Collins and D. E. Soper, “Parton distribution and decay functions,” *Nuclear Physics B* **194** (1982), no. 3, 445 – 492.
- [41] A. Buckley and M. Whalley, “HepData reloaded: Reinventing the HEP data archive,” *PoS ACAT2010* (2010) 067, 1006.0517.
- [42] A. Valishev, “Tevatron accelerator physics and operation highlights,” 1202.5525.
- [43] Fermilab, “Accelerator Concepts,”
http://www-bdnew.fnal.gov/operations/rookie_books/Concepts_v3.6.pdf.
- [44] **CDF** Collaboration, T. Aaltonen *et al.*, “Measurement of the Inclusive Jet Cross Section at the Fermilab Tevatron p anti-p Collider Using a Cone-Based Jet Algorithm,” *Phys.Rev.* **D78** (2008) 052006, 0807.2204.
- [45] **CDF** Collaboration, D. Acosta *et al.*, “Measurement of the J/ψ meson and b –hadron production cross sections in $p\bar{p}$ collisions at $\sqrt{s} = 1960$ GeV,” *Phys.Rev.* **D71** (2005) 032001, hep-ex/0412071.
- [46] O. S. Bruning *et al.*, “LHC Design Report,” (2004).
- [47] e. Evans, Lyndon and e. Bryant, Philip, “LHC Machine,” *JINST* **3** (2008) S08001.

- [48] CERN, “Integrated Luminosity,” (2011)
<http://lpc.web.cern.ch/lpc/lumiplots.htm>.
- [49] **ATLAS** Collaboration, G. Aad *et al.*, “The ATLAS Experiment at the CERN Large Hadron Collider,” *JINST* **3** (2008) S08003.
- [50] G. F. Sterman and S. Weinberg, “Jets from Quantum Chromodynamics,” *Phys.Rev.Lett.* **39** (1977) 1436.
- [51] **TeV4LHC QCD Working Group** Collaboration, M. G. Albrow *et al.*, “Tevatron-for-LHC Report of the QCD Working Group,” hep-ph/0610012.
- [52] T. Sjostrand, L. Lonnblad, and S. Mrenna, “PYTHIA 6.2: Physics and manual,” hep-ph/0108264.
- [53] **CTEQ** Collaboration, H. Lai *et al.*, “Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distributions,” *Eur.Phys.J.* **C12** (2000) 375–392, hep-ph/9903282.
- [54] M. Cacciari and G. P. Salam, “Dispelling the N^3 myth for the k_t jet-finder,” *Phys.Lett.* **B641** (2006) 57–61, hep-ph/0512210.
- [55] **CDF** Collaboration, A. Bhatti *et al.*, “Determination of the jet energy scale at the collider detector at Fermilab,” *Nucl.Instrum.Meth.* **A566** (2006) 375–412, hep-ex/0510047.
- [56] **D0** Collaboration, V. Abazov *et al.*, “Measurement of the inclusive jet cross-section in $p\bar{p}$ collisions at $s^{91/2}=1.96$ -TeV,” *Phys.Rev.Lett.* **101** (2008) 062001, 0802.2400.
- [57] P. Nason, “A New method for combining NLO QCD with shower Monte Carlo algorithms,” *JHEP* **0411** (2004) 040, hep-ph/0409146.

- [58] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method,” *JHEP* **0711** (2007) 070.
- [59] S. Alioli, P. Nason, C. Oleari, and E. Re, “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX,” *JHEP* **1006** (2010) 043, 1002.2581.
- [60] S. Alioli, K. Hamilton, P. Nason, C. Oleari, and E. Re, “Jet pair production in POWHEG,” *JHEP* **1104** (2011) 081.
- [61] J. Pumplin, D. Stump, J. Huston, H. Lai, P. M. Nadolsky, *et al.*, “New generation of parton distributions with uncertainties from global QCD analysis,” *JHEP* **0207** (2002) 012, hep-ph/0201195.
- [62] N. Kidonakis and R. Vogt, “Next-to-next-to-leading order soft gluon corrections in top quark hadroproduction,” *Phys.Rev.* **D68** (2003) 114014, hep-ph/0308222.
- [63] **CDF** Collaboration, T. Aaltonen *et al.*, “Combination of CDF top quark pair production cross section measurements with up to 4.6 fb^{-1} ,” *CDF/PUB/TOP/PUBLIC/9913* (October, 2009).
- [64] R. Alon, E. Duchovni, G. Perez, A. P. Pranko, and P. K. Sinervo, “A Data-driven method of pile-up correction for the substructure of massive jets,” *Phys.Rev.* **D84** (2011) 114025, 1101.3002.
- [65] J. C. Collins, D. E. Soper, and G. F. Sterman, “Factorization of Hard Processes in QCD,” *Adv.Ser.Direct.High Energy Phys.* **5** (1988) 1–91, hep-ph/0409313. To be publ. in ‘Perturbative QCD’ (A.H. Mueller, ed.) (World Scientific Publ., 1989).

- [66] H. Contopanagos, E. Laenen, and G. F. Sterman, “Sudakov factorization and resummation,” *Nucl.Phys.* **B484** (1997) 303–330, hep-ph/9604313.
- [67] M. Dasgupta and G. P. Salam, “Event shapes in e^+e^- annihilation and deep inelastic scattering,” *J.Phys.G* **G30** (2004) R143, hep-ph/0312283.
- [68] C. F. Berger, T. Kucs, and G. F. Sterman, “Event shape / energy flow correlations,” *Phys.Rev.* **D68** (2003) 014012, hep-ph/0303051.
- [69] R. Ellis, W. Stirling, and B. Webber, “QCD and collider physics,” *Camb.Monogr.Part.Phys.Nucl.Phys.Cosmol.* **8** (1996) 1–435.
- [70] **CDF** Collaboration, T. Aaltonen *et al.*, “Measurement of the k_T Distribution of Particles in Jets Produced in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96\text{-TeV}$,” *Phys.Rev.Lett.* **102** (2009) 232002, 0811.2820.
- [71] J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, *et al.*, “Uncertainties of predictions from parton distribution functions. 2. The Hessian method,” *Phys.Rev.* **D65** (2001) 014013, hep-ph/0101032.
- [72] Y. Eshel, O. Gedalia, G. Perez, and Y. Soreq, “Implications of the Measurement of Ultra-Massive Boosted Jets at CDF,” *Phys.Rev.* **D84** (2011) 011505, 1101.2898.
- [73] K. Blum, C. Delaunay, O. Gedalia, Y. Hochberg, S. J. Lee, *et al.*, “Implications of the CDF $t\bar{t}$ Forward-Backward Asymmetry for Boosted Top Physics,” *Phys.Lett.* **B702** (2011) 364–369, 1102.3133.
- [74] T. Junk, “Confidence level computation for combining searches with small statistics,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **434** (1999), no. 23, 435 – 443.

- [75] **CDF** Collaboration, T. Aaltonen *et al.*, “A Search for the Associated Production of the Standard-Model Higgs Boson in the All-Hadronic Channel,” *Phys.Rev.Lett.* **103** (2009) 221801, 0907.0810.
- [76] **CDF** Collaboration, T. Aaltonen *et al.*, “Evidence for a Mass Dependent Forward-Backward Asymmetry in Top Quark Pair Production,” *Phys.Rev.* **D83** (2011) 112003, 1101.0034.

Declaration

This study was conducted by Raz Alon, with supervision from Prof. Ehud Duchovni (Weizmann Institute of Science), Prof. Pekka K. Sinervo (University of Toronto), and Dr. Gilad Perez (Weizmann Institute of Science).

Acknowledgments

I would like to express my deep gratitude to all of those who advised, supported and helped me to carry out this work.

In particular, I would like to thank my advisor, Prof. Ehud Duchovni, without whose help, encouragement and patience, not to mention his unsurpassed knowledge in the field of particle physics, this thesis would not have been possible. His good advice, support and friendship have been invaluable on both an academic and a personal level, for which I am extremely grateful.

I am grateful to Prof. Pekka K. Sinervo and Dr. Gilad Perez, for their guidance and for conversations that clarified my thinking on this and other matters. Their friendship and professional collaboration meant a great deal to me.

My utmost gratitude goes to Prof. Giora Mikenberg and Prof. Eilam Gross for their endless support and invaluable discussions along the way.

I have learned a lot from those who guided me at Fermilab and gratefully acknowledge my debt to them, especially Dr. Sasha Pronko and Dr. Christina Mesropian.

A number of colleagues and students cooperated and shared interest along the way, making the group a pleasant and congenial environment. In this regard, I am indebted to Dr. Amit Klier, Gal Dadiani, Ohad Silbert, Ofer Vitells, Liron Barak, Ofir Gabizon, Dmitry Milstein, Itamar Roth, and Tal Frank.

In conclusion, I would like to express my deepest gratitude to Meir Shoa and Dr. Vladimir Smakhtin for letting me take a small part in the Super LHC research and development that took place in the lab.