Correlations in Bottom Quark Pair Production at the Fermilab TeVatron

Jason Galyardt

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
at
Carnegie Mellon University
Department of Physics
Pittsburgh, Pennsylvania

January 2009

Advisor: Prof. James Russ

Correlations in Bottom Quark Pair Production at the Fermilab TeVatron

Abstract

I present an analysis of $b\bar{b}$ pair production correlations, using dimuon-triggered data collected with the Collider Detector at Fermilab (CDF) in $p\bar{p}$ collisions at $\sqrt{s}$=1.96 TeV during Run II of the TeVatron. The leading order (LO) and next-to-leading order (NLO) b quark production processes are discriminated by the angular and momentum correlations between the $b\bar{b}$ pair. Track-level jets containing a muon are classified by b quark content and used to estimate the momentum vector of the progenitor b quark. The theoretical distributions given by the MC@NLO event generator are tested against the data.

Acknowledgements

Contents

1 Introduction

1.1 The Standard Model

The Standard Model [1] describes the fundamental constituents of matter and how they interact. It is a highly successful gauge field theory in that it describes the vast majority of observed phenomena quite well. The fundamental particles in this model are fermions, possessing half-integral spin (in units of $\hbar$), and bosons, which have integral spin. There are two types of fermions: quarks, and leptons. Quarks and leptons each come in 6 "flavors", organized into 3 generations according to a mass hierarchy:

$$\begin{pmatrix} u \\ d \end{pmatrix}, \quad \begin{pmatrix} c \\ s \end{pmatrix}, \quad \begin{pmatrix} t \\ b \end{pmatrix}$$
$$\begin{pmatrix} e \\ \nu_e \end{pmatrix}, \quad \begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix}, \quad \begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix}$$

The source of the mass hierarchy is as yet unknown. The properties of these fermions are summarized in Table 1.

Quarks			Leptons		
Flavor	Mass [GeV/c^2]	Charge [e]	Flavor	Mass [GeV/c^2]	Charge [e]
up (u)	$\sim$0.002	2/3	electron (e)	0.000511	-1
down (d)	$\sim$0.005	-1/3	ν_e	$\sim$0	0
charm (c)	1.27	2/3	muon (μ)	0.106	-1
strange (s)	0.104	-1/3	ν_μ	$\sim$0	0
top (t)	171	2/3	tau (τ)	1.78	-1
bottom (b)	4.20	-1/3	ν_τ	$\sim$0	0

Table 1: The fundamental fermions and their properties [3]

There are four forces: the electromagnetic (EM) force, the weak force, the strong force, and gravity. The gauge bosons mediate these forces: the photon (γ) mediates the EM force; the massive $W^\pm$ and Z^0 bosons mediate the weak force; the gluon g mediates the strong force; the (undiscovered) graviton mediates gravity. At high energy scale, the EM and weak forces merge into one single force called the 'electroweak' force. The symmetry involved in this unification is broken by the Higgs mechanism via the scalar Higgs boson H^0 (within the Standard Model; other models involve multiple Higgs bosons of varying spin). The properties of these bosons are summarized in Table 2.

Boson	Mass [GeV/c^2]	Charge [e]	Spin [$\hbar$]	Force Mediated
γ	0	0	1	Electromagnetic
$W^\pm$	80.4	± 1	1	Weak
Z^0	91.2	0	1	Weak
g	0	0	1	Strong
Graviton	0	0	2	Gravitation
H^0	> 114	0	0	–

Table 2: The fundamental bosons and their properties [3].

While a quantized theory of gravity still eludes coherent understanding, the gravitational force operates on all massive particles. The EM force operates on all particles with electric charge. The weak force acts on quarks and leptons. The strong force acts on those particles with 'color' charge: quarks and gluons. Quantum Chromodynamics (QCD) is the theory which describes interactions between objects with color charge within the Standard Model.

1.2 The Strong Force and QCD

QCD is a non-abelian gauge field theory with SU(3) symmetry which describes interactions of the strong force upon colored partons (quarks and gluons). A given flavor of quark (*e.g.* strange) can have one of six color charges: red, blue, green, and the corresponding anti-colors. Gluons, on the other hand, are bi-color objects (color/anti-color). This gives a total of eight color charge permutations. Due to the bi-color nature of gluons, they can self-couple. The strong force increases with distance. This leads to two the phenomena of *color confinement* and *asymptotic freedom*.

Color confinement refers to the phenomenon which leads to the absence of observable objects with net color charge. Quarks bind together into colorless states of two (mesons) or three (baryons), called hadrons. Free quarks have never been observed in nature because they are always confined to these color singlet states. The binding force increases with distance, making it difficult to separate a bound quark from its partner(s).

The idea of asymptotic freedom describes the reduction in strength of the strong force between two bound quarks as the distance between them decreases. Because gluons couple to themselves, they can split into a pair of gluons, each carrying half of the color charge of the original gluon. This leads to the situation where a cloud of (virtual) gluons surrounds a (bound) quark; since the gluons will preferentially have the same color charge as the quark, the effective charge of the quark increases with distance. Conversely, at small distances

(large momentum transfer), the quark's effective color charge dwindles; the effective force is reduced until only the color charge of the bare quark is left. The quarks can then be approximated as free. This property is essential in making QCD a calculable theory.

The coupling 'constant' of QCD α_s actually varies with energy scale Q. The scale at which α_s becomes large as Q decreases is defined as Λ_{QCD} ($\sim$200 GeV). At large Q, α_s is small, and the techniques of perturbation theory can be applied. Quantities of interest can be expanded as a power series in α_s and truncated at a designated point. Perturbative QCD (pQCD) calculations with *leading order* (LO) accuracy include terms through $\mathcal{O}(\alpha_s^2)$, while *next-to-leading order* (NLO) accuracy is obtained by including terms through $\mathcal{O}(\alpha_s^3)$.

Processes which occur at scales smaller than Λ_{QCD} are called *non-perturbative* processes. Such processes can only be described by phenomenological models, requiring some sort of input from experiment.

The methods of perturbative QCD pertain to interactions between partons, such as $ab \rightarrow cd$ for partons a and b interacting to produce partons c and d. However the study of partonic interactions would seem to be stymied by the color confinement property mentioned above: there is no way to produce beams of free partons for use in an experiment. One way around this problem is to produce and collide beams of hadrons (*e.g.* protons and anti-protons). The partons a and b within the beam hadrons A and B will have some relative momentum, though. The confinement property also implies that the final state partons c and d will somehow form hadrons C and D. These are both non-perturbative effects. So now the non-perturbative effects are mixed with the perturbative effects. Fortunately, there is a factorization theorem [2] which allows the perturbative component of a cross section to be separated from the non-perturbative component. The cross section $\sigma(AB \rightarrow CD)$ is schematically represented as a convolution of non-perturbative and perturbative terms:

$$\sigma(AB \rightarrow CD) = \sum_{a,b} f_a^A(x_a) f_b^B(x_b) \otimes \hat{\sigma}(ab \rightarrow cd) \otimes F_c^C(z_c) F_d^D(z_d); \qquad (1)$$

the sum is over initial state parton species. The perturbative component is captured in the partonic cross section $\hat{\sigma}(ab \rightarrow cd)$, and the non-perturbative components are capture in the f and F functions. The f functions are called Parton Distribution Functions (PDFs) and describe the probability density of finding a specific parton species inside the corresponding hadron with a particular fraction x of the hadron's momentum. The F functions are called fragmentation functions and parameterize the probability for the hadron to take on a fraction z of the final state quark's momentum.

The parton summation in Eq. 1 is not limited to the primary quark constituents of the beam hadrons A and B. Surrounding these valence quarks is a sea of virtual partons, gener-

ated by the self-interacting gluons as they mediate the binding force. These virtual partons do not have quite the right energy, so they can only exist for a time scale commensurate with the Uncertainty Principle. Due to the complexity of the beam hadrons and the fact that the energy scale in question falls into the non-perturbative regime, the parton distribution functions cannot be calculated from pQCD. Instead, experimental data from many environments are fit into a parameterized framework, such as that of the CTEQ collaboration. The specific series of parton distribution functions published by the CTEQ collaboration which uses a next-to-leading order QCD framework, the CTEQ 6.1M series, can be found in Ref. [17]. This series of PDF fits are used in this analysis. The LHAPDF package [16] provides a standardized interface to various PDF fits, including the CTEQ series.

Another modification to the schematic view of hadronic interactions illustrated in Eq. 1 is the possibility of emission from the outgoing partons. The calculation of the emission probability can be performed within pQCD, but it leads to divergences which must be regularized in order to make predictions on physical observables. This regularization leads to the prediction that emission will be contained within a restricted angular region about the outgoing parton [4]. This collimated spray of particles is called a *jet*.

1.3   Underlying Event

The term 'underlying event' has varying definitions, but in this thesis it will denote physical effects which are ancillary to the physical process of interest. This includes treatment of the hadronic beam remnants, initial state radiation (ISR), and multiple parton interactions (MPI). In a hadronic collider environment, the beam remnants will fragment[1] into a spray of particles which preferentially travel in the initial beam directions. ISR refers to emission from one of the incoming hadrons prior to the 'hard' (high momentum transfer) interaction. Multiple parton interactions are secondary, 'semi-hard' interactions which occur in association with the 'hard' interaction (within the same hadron, in fact).

All of these processes contribute particles to the final state which are effectively uncorrelated with those created through the hard interaction and resulting fragmentation. Thus the underlying event contributes diffuse noise to the system, complicating the reconstruction of the jet structure. On the other hand, much of of these noise particles will have low transverse momentum, and so their degree of influence is small.

---

[1]This is a non-perturbative process, so the term 'jet' is not quite appropriate.

1.4 $b\bar{b}$ Production

Bottom quark pair production at hadron colliders proceeds primarily through strong interactions. The relatively large mass of the bottom quark compared to Λ_{QCD} allows for the application of perturbative QCD to the cross section calculation. In order to truncate a power series after a given term, the remaining terms should give progressively smaller corrections to the quantity under approximation. Somewhat surprisingly, this doesn't happen with the cross section of bottom quark pairs produced in hadron collisions; the next-to-leading order term is numerically comparable in size to the leading order term. This can be explained by considering that the cross section for the leading order process $gg \rightarrow gg$ is approximately two orders of magnitude larger than that of $gg \rightarrow b\bar{b}$. Any of the gluons in the interaction $gg \rightarrow gg$ may split into a bottom quark pair, yielding a higher order bottom quark pair production process which has a numerically significant cross section.

There are three common categories used to describe bottom quark pair production at hadron colliders: flavor creation, flavor excitation, and gluon splitting (also called the 'shower' or 'fragmentation' component). Examples of Feynman diagrams for these production modes are shown in Fig. 1. It should be emphasized that these categories are only illustrative; there is a significant amount of ambiguity in their definitions in an NLO calculation, such as that of Ref. [15], due to interference between the different production diagrams. Flavor creation occurs at both $\mathcal{O}(\alpha_s^2)$ and $\mathcal{O}(\alpha_s^3)$ in the perturbative expansion. Flavor excitation and gluon splitting occur at first at $\mathcal{O}(\alpha_s^3)$ in the perturbative expansion.

Flavor creation processes includes quark annihilation and gluon fusion at $\mathcal{O}(\alpha_s^2)$. These interactions have two body states: $q\bar{q} \rightarrow b\bar{b}$, $gg \rightarrow b\bar{b}$. These flavor creation processes are promoted to $\mathcal{O}(\alpha_s^3)$ via final state radiation (gluon emission from one of the b quark lines: $q\bar{q} \rightarrow b\bar{b}g$ and $gg \rightarrow b\bar{b}g$).

Flavor excitation refers to processes in which one bottom quark of a pair in the initial state attains a large transverse momentum through an interaction with a light parton (u, d, s, or g) yielding three-body final states: $q\bar{q} \rightarrow b\bar{b}g$, $gg \rightarrow b\bar{b}g$, $gq \rightarrow b\bar{b}q$, and $g\bar{q} \rightarrow b\bar{b}\bar{q}$. In the flavor excitation diagram in Fig. 1, the b quark gains a large transverse momentum due to the scattering, while the $\bar{b}$ quark preferentially travels collinearly with its parent gluon. The parent gluon was part of the beam hadron and so had low p_T. Therefore there is a high probability in flavor excitation interactions for one of the bottom quarks to fall outside the detector acceptance due to insufficient transverse momentum.

The source of the $b\bar{b}$ pair in the initial state is due to either gluons splitting ($g \rightarrow b\bar{b}$) or to the intrinsic bottom quark content of the beam hadron. In standard NLO calculations, the initial state partons are considered to be massless; this forbids the consideration of bottom

Flavor Creation $\mathcal{O}(\alpha_s^2)$

Flavor Creation $\mathcal{O}(\alpha_s^2)$

Gluon Radiation $\mathcal{O}(\alpha_s^3)$

Flavor Excitation $\mathcal{O}(\alpha_s^3)$
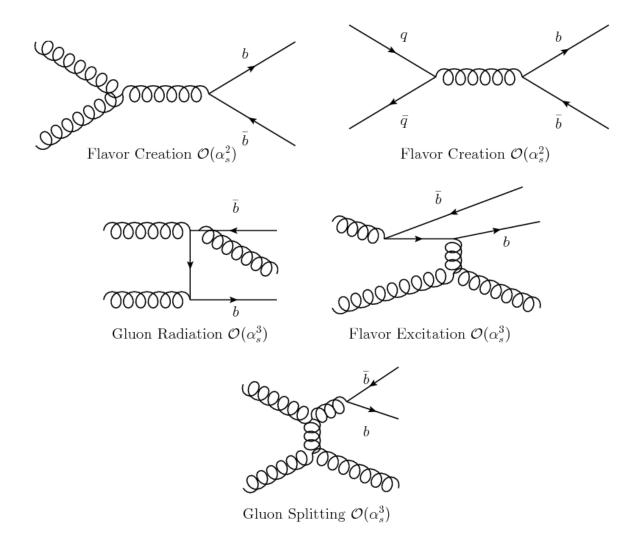
Gluon Splitting $\mathcal{O}(\alpha_s^3)$

Figure 1: Example diagrams for Standard Model $b\bar{b}$ production at the TeVatron

quarks intrinsic to the beam hadron. Thus initial state gluon splitting is the only source of bottom quark pairs considered in NLO calculations. Flavor excitation processes therefore appear for the first time at $\mathcal{O}(\alpha_s^3)$ in the perturbative cross section calculation.

The gluon splitting production mechanism also results in a three-body final state at $\mathcal{O}(\alpha_s^3)$, but the hard scatter involves only light flavor partons. An outgoing gluon splits into a $b\bar{b}$ pair, resulting in interactions with the same final state as for flavor excitation. However, the angular and momentum correlations will be different, as discussed in Section 1.5.

1.5 Bottom Quark Pair Correlations

A phenomenological study of several models of bottom quark production mechanisms and their correlations was performed in Ref. [5] for $p\bar{p}$ interactions at $\sqrt{s} = 1.8$ TeV (*i.e.* Run I of the TeVatron). This study suggests that the different $b\bar{b}$ production mechanisms should have different correlation distributions. Both bottom quarks were required to have a minimum transverse[2] momentum p_T of 5 GeV/c and central rapidity[3] ($|y| < 1$). Figure 2 shows the azimuthal angular difference between two bottom quarks produced by the PYTHIA [7] event generator; the different curves correspond to the different production mechanisms. Figure 3 shows the transverse momentum asymmetry, defined as

$$\mathcal{A}_{p_T} \equiv \frac{p_{T1} - p_{T2}}{p_{T1} + p_{T2}}, \tag{2}$$

where p_{T1} corresponds to the transverse momentum of the b quark, and p_{T2} corresponds to that of the $\bar{b}$ quark. The minimum p_T for the b quark was reduced to zero in the construction of this distribution. Raising the minimum p_{T1} to 5 GeV/c would remove a large portion of the flavor excitation component and some of the gluon splitting component. The bump at negative values of $\mathcal{A}_{p_T}$ in Fig. 3 would be much reduced.

The LO flavor creation production processes yield $b\bar{b}$ pairs from $q\bar{q}$ annihilation ($q\bar{q} \to b\bar{b}$) or gluon fusion ($gg \to b\bar{b}$). With only two partons in the final state, momentum conservation requires that the $b\bar{b}$ pair be balanced in transverse momentum ($\mathcal{A}_{p_T} = 0$) and back-to-back in azimuthal angle ($\Delta\phi(b,\bar{b}) = \pi$). As mentioned in Section 1.4, higher order flavor creation processes involve gluon emission. As the emitted gluon carries off some of the momentum of the $b\bar{b}$ system, the azimuthal angular difference $\Delta\phi(b,\bar{b})$ acquires a non-zero width, and the b and the $\bar{b}$ are no longer balanced in p_T.

---

[2]The beam direction is assumed to be the longitudinal direction; azimuthal angle is measured in the transverse plane.

[3]Rapidity y is defined as $\frac{1}{2}\ln((E - p_z)/(E + p_z))$, with the beam axis defined to be the z-axis. Differences of rapidity are Lorentz invariant. Rapidity is approximated by pseudorapidity $\eta \equiv -\ln\tan(\theta/2)$ when $p \gg m$ and $\theta \gg 1/\gamma$, where $\gamma \equiv (1 - \beta^2)^{-1/2}$, and $\beta \equiv \boldsymbol{p}/E$. Note that η is calculable even when y is not.
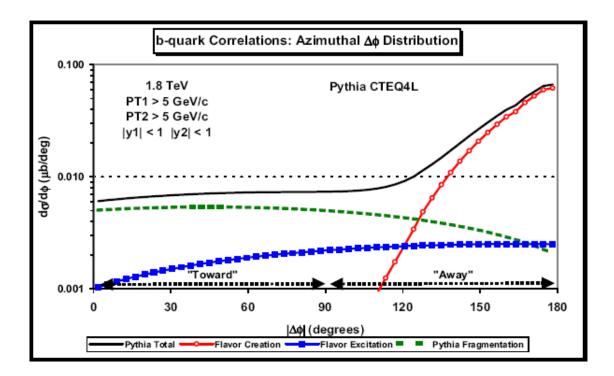
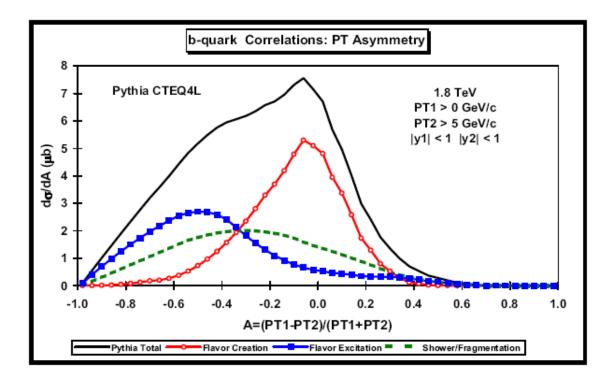Figure 2: Bottom quark pair azimuthal angular difference $\Delta\phi(b,\bar{b})$, from Ref. [5]

Figure 3: Bottom quark pair transverse momentum asymmetry $\mathcal{A}_{p_T}(b,\bar{b})$, from Ref. [5]

The gluon splitting process yields bottom quark pairs with somewhat asymmetric p_T ($\mathcal{A}_{p_T} \neq 0$). There is a slight preference for bottom quarks to be produced at small $\Delta\phi(b,\bar{b})$. Given that this is the only process which contributes in this region, it should be clearly identified in data, given appropriate experimental conditions.

Due to the scattering of one initial state bottom quark off of a light parton in flavor excitation processes, the p_T of the $b\bar{b}$ system is also asymmetric. The azimuthal angular difference is decorrelated for the same reason, resulting in a broad $\Delta\phi(b,\bar{b})$ distribution which is depleted at small angles.

1.6 Motivation of This Analysis

The discussion of bottom quark pair production so far has revolved around the prevailing qQCD theory at NLO accuracy, exemplified by the calculation of Ref. [15]. The experimental picture of the $b\bar{b}$ pair production cross section shows significant discrepancy with the exact NLO theory, as demonstrated in Ref. [6]. This study compared five measurements (references [35] – [38]) of the bottom quark pair cross section $\sigma_{b\bar{b}}$, extracted from various experimental signatures, to the exact NLO prediction of Ref. [15]. The measurements were all made for central rapidities ($|y| < 1$), but at various minimum transverse momenta. The comparison was made via the ratio R_{2b} of the experimental measurement to the NLO prediction, the results of which are summarized in Table 3. As noted in Ref. [6], the experimental results seem to be inconsistent among themselves; in particular the dimuon channels show significant discrepancies with respect to the NLO result, while the dijet results are more consistent.

channel	(experiment)	R_{2b} for $p_T^{\min}$ (GeV/c) $\geq$			
		$6 - 7$	10	15	$\simeq 20$
$b + \bar{b}$ jets	(CDF [35])			0.012 ± 0.25	
$b + \bar{b}$ jets	(CDF [36])				0.010 ± 0.32
$\mu + b$ jet	(CDF [37])		0.015 ± 0.10		
$\mu^+ + \mu^-$	(CDF [39])	0.030 ± 0.20			
$\mu^+ + \mu^-$	(DØ [38])	0.023 ± 0.33			

Table 3: A summary across several studies performed at the TeVatron of R_{2b}, the ratio of the measured bottom quark pair cross section $\sigma_{b\bar{b}}$ to the exact NLO theory, as presented in Ref. [6]. The bottom quarks are all produced centrally ($|y_b| < 1$).

As already mentioned, quarks are not directly observable. In order to quote a quark-level cross section, the measurements reviewed in Ref. [6] all performed some sort of 'unfolding'

procedure which corrected the cross section of the observed experimental signature (*e.g.* dimuons), to the quark level. Such a procedure inevitably involves some model of the non-perturbative fragmentation function (either an actual fragmentation function F as in Eq. 1, or a parton shower MC approximation, as described in Section 3.1). The studies performed in references [42] and [43] showed that such non-perturbative effects can have a significant effect on the shape and integral of the single-b differential cross section $d\sigma_b/dp_T(b)$. Non-perturbative effects should have less influence at higher energy; this is borne out in the agreement with the NLO predictions found in the dijet analyses ([35] and [36]).

Therefore, it is not yet clear whether the discrepancy in the lower p_T range illustrated in Table 3 is due to an intrinsic failure of the pQCD approach or due to contributing experimental and/or theoretical effects. As discussed in Section 1.5, the various production mechanisms correlate the b and the $\bar{b}$ in different ways. A comparison of the shapes of several $b\bar{b}$ correlation distributions between CDF data and NLO QCD predictions in the low to moderate $p_T(b)$ regime will help to clarify the situation.

1.7   Previous Measurements

The pioneering study of $b\bar{b}$ pair production and their angular correlations was performed [33] by the UA1 collaboration at the CERN $p\bar{p}$ collider with a center of mass energy of $\sqrt{s} = 630$ GeV. In 4.7 pb^{-1} of data, a sample of dimuon events was collected in the mass range $4 < m(\mu, \mu) < 35$ GeV/c^2. The minimum dimuon mass threshold was necessary to reject light vector meson resonances, J/Ψ and Ψ' mesons, and sequential charm decays ($b \to \mu\nu c$ followed by $c \to \mu\nu X$). The minimum dimuon mass requirement severely restricts the acceptance for dimuons with small opening angle. The maximum mass threshold was intended to reject Z^0 bosons. The transverse momentum for muons in this sample was required to be greater than 3 GeV/c, and the muon pseudorapidity was required to be in the domain $|\eta(\mu)| < 2.3$.

The dimuon data were corrected back to the quark level using a parton shower Monte Carlo to describe the fragmentation process. The $\Delta\phi(b, \bar{b})$ distribution and the $\Delta R(b, \bar{b})$ distribution[4] for $\Delta\phi(b, \bar{b}) < 2\pi/3$ were studied. This latter requirement was imposed to enrich the higher order production mechanisms; the flavor creation mechanism is entirely removed by this requirement. Reasonable agreement in shape was found between data and the MNR NLO calculation of Ref. [15] for $p_{Tb}^{\max} > 6$ GeV/c and $p_{Tb}^{\max} > 11$ GeV/c, as shown in Fig. 4. These data were binned in $p_{Tb}^{\max}$ in order to observe potential p_T dependence of the admixture of bottom quark production processes. Agreement between data and the MNR calculation for different p_T bins supports the validity of the MNR calculation in this respect.

---

[4]$\Delta R \equiv \sqrt{\Delta\phi^2 + \Delta\eta^2}$

Figure 4: Measured $b\bar{b}$ angular correlations, compared with MNR NLO theory and a parton shower MC (labeled 'eff. QCD' in the legends), from Ref. [33]. The plots on the left were made with a minimum p_{Tb} of 6 GeV/c for the highest p_T quark in each event, while the minimum p_{Tb} threshold for the plots on the right was set at 11 GeV/c. The top row shows the $\Delta\phi(b,\bar{b})$, while the bottom row shows $\Delta R(b,\bar{b})$ for $\Delta\phi(b,\bar{b}) < 120°$. This latter requirement removes all of the flavor creation component.

Two similar measurements of angular correlations in the dimuon channel have been made at the TeVatron with $\sqrt{s} = 1.8$ TeV during the 1992-1993 run by the CDF [39] and DØ [38] collaborations. The DØ measurement examined a nearly identical mass domain of $6 \text{GeV}/c^2 < m(\mu, \mu) < 35 \text{ GeV}/c^2$, while the CDF measurement chose to use a single bounded domain: $m(\mu, \mu) > 5 \text{ GeV}/c^2$. The CDF analysis required $p_T(\mu) > 3 \text{ GeV}/c$ and $|\eta(\mu)| < 0.6$, while the DØ analysis required $p_T(\mu) > 4 \text{ GeV}/c$ and $|\eta(\mu)| < 0.8$. In a departure from the UA1 analysis, the DØ analysis required each muon to be contained within a jet of size $R = 0.8$ and transverse energy[5] $E_T > 12$ GeV. Both analyses corrected the NLO theory for the bias introduced by the dimuon mass requirements; they found the corrected theory and the data to be consistent in shape. However, both of these analyses suffer from the same low acceptance at small opening angles as did UA1.

Figure 5: Dimuon correlations at the TeVatron; the DØ [38] result is on the left, and the CDF [39] is on the right.

Two CDF analyses [40] utilizing data taken during the 1994-1995 run of the TeVatron employed bottom hadron identification techniques which avoided restrictions on the opening angle distribution. Both analyses measured the fraction f_{toward} of bottom hadron pairs produced in the same hemisphere. The first analysis used a secondary vertex identification tool to select a 'double-tagged' bottom hadron sample from single-lepton-triggered data. The trigger leptons were required to have $p_T > 8$ GeV/c and $|\eta| < 1$ for electrons and $|\eta| < 0.6$ for muons. According to simulation, these selection criteria yielded a trigger bottom hadron with $p_T(H_b) > 14$ GeV/c, and an associated bottom hadron with $p_T(H_b) > 7.5$ GeV/c.

---

[5]$E_T \equiv E \sin\theta$, where θ is the polar angle measured from the z-axis.

The second analysis used a dimuon J/Ψ trigger to collect a sample of bottom hadron pairs where an associated lepton identified the other bottom hadron. Muons were required to have $p_T > 3$ GeV/c, while electrons were required to have $p_T > 2$ GeV/c. Ninety percent of the bottom quarks associated with the J/Ψ had $p_T(b) \gtrsim 6$ GeV/c, while about 90% of the b quarks associated with the third lepton satisfied $p_T > 4$ GeV/c. Both of these analyses found f_{toward} to be approximately 25%, consistent with both the parton shower Monte Carlos PYTHIA and HERWIG, as well as the NLO theory.

1.8 Analysis Overview

The strategy employed in this analysis was as follows. The base dataset was collected with the CDF detector in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV with a set of dimuon triggers which cover the pseudorapidity region $|\eta(\mu)| < 1$. This dataset is already highly enric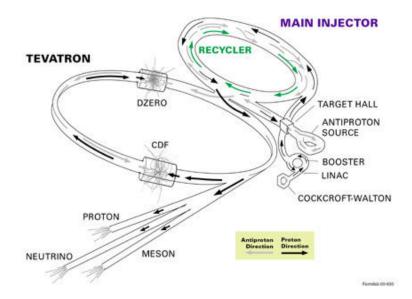hed in heavy flavor content (both bottom and charm) due to the relatively large semi-leptonic branching ratio of heavy flavor hadrons ($\sim$10% for $H_{\mathrm{HF}} \to \mu\nu X$, where H_{HF} is a generic heavy flavor hadron).

A jet-clustering algorithm was run on the tracks reconstructed within the CDF detector. Bottom quark jet (b-jet) candidates were identified through the inclusion of a muon within a given jet's constituents. A multivariate machine learning technique was developed in order to identify true b-jets within the sample of μ-jets; two positively-classified b-jets were required in each event. The bottom quark momenta were estimated using the b-jet momenta. The following correlation distributions were formed from the two b-jets in each event: $p_T(b+\bar{b})$, $\eta(b+\bar{b})$, $\Delta\phi(b,\bar{b})$, $|\Delta\eta(b,\bar{b})|$, $|\Delta\eta(b,\bar{b})|$ given $\Delta\phi(b,\bar{b}) < 2\pi/3$, and $\mathcal{A}_{p_T}$. The distributions $p_T(b+\bar{b})$ and $|\eta(b+\bar{b})|$ serve to characterize the b-jet system in a basic fashion. The requirement $\Delta\phi(b,\bar{b}) < 2\pi/3$ serves to isolate higher-order contributions (demonstrated in Fig. 2). In the experimental definition of $\mathcal{A}_{p_T}$, μ-jets were ordered by decreasing p_T, since momentum requirements of the trigger were symmetric; this makes the p_T asymmetry positive definite (compare to Fig. 3).

The CDF data obtained in this way were compared with the MC@NLO [12] event generator, according to the shape of the $b\bar{b}$ correlation distributions mentioned above. This generator includes the "MNR" calculation of Ref. [15], matched to the HERWIG [10] parton shower MC in such a way as to avoid double counting of $b\bar{b}$ production mechanisms in the cross section (see Section 3.2). MC@NLO provides a fully exclusive description of $b\bar{b}$ events, as given by qQCD at NLO accuracy. Thus the data and qQCD prediction may be analyzed using the same techniques, without the need for an 'unfolding' procedure to correct the data down to the quark level.

2 Experimental Apparatus

2.1 Accelerator Complex

Figure 6: The Fermilab Accelerator Chain

2.1.1 Proton Production

A chain of seven accelerators, depicted in Fig. 6, is used to produce the proton and anti-proton beams which are collided inside the CDF detector. The process of proton acceleration begins with H^- ions. These ions are accelerated by the Cockroft-Walton DC accelerator to an energy of 750 keV. The ions are then transfered to the linear accelerator (linac), which uses RF cavities to accelerate them to 400 MeV. Due to the alternating current nature of the RF cavities, the ions are divided into bunches naturally. The H^- ions go through a carbon foil, which strips them of their electrons. The resulting protons then go to the Booster, a circular synchrotron with 18 RF cavities which provides acceleration up to 8 GeV. The Main Injector is the next accelerator in the chain. It is a circular synchrotron which accelerates the protons to 150 GeV. The TeVatron is the final link in the accelerator chain. It accelerates both protons and anti-protons from 150 GeV to 980 GeV. The proton and anti-proton beams collide at two points on the TeVatron ring, corresponding to the CDF and D0 experiments. The center of mass energy of the $p\bar{p}$ collisions is 1.96 TeV.

20

2.1.2 Anti-Proton Production

Anti-proton production starts with the extraction of 120 GeV protons from the Main Injector. This proton beam hits a nickel target, producing anti-protons, among other things. It takes approximately 10^5 protons on target to produce a single anti-proton. A cylindrical lithium lens focuses the particle spray, and a pulsed dipole magnet spectrometer filters out the undesirable species, resulting in an 8 GeV anti-proton beam. This $\bar{p}$ beam then goes to the triangular shaped Debuncher where the spread of the $\bar{p}$ momentum distribution is reduced.

The Accumulator, which resides inside the Debuncher tunnel, collects the beam from the Debuncher and stores it. The Accumulator is most efficient for smaller "stashes" of anti-protons, so once the stash reaches a certain size, it is transferred to the Recycler. The Recycler is a storage ring located within the Main Injector tunnel. Several stages of cooling are applied to the 8 GeV anti-proton beam in Recycler which reduces the physical spread (both longitudinal and transverse) and the momentum spread of the $\bar{p}$-beam. Once a sufficient collection of anti-protons has accumulated in the Recycler, the beam is transferred to the Main Injector, where it counter-circulates with the proton beam and is accelerated from 8 GeV to 150 GeV. As with the proton beam, the next step is injection into the TeVatron and acceleration to 980 GeV.

2.1.3 Collisions

The beams in the TeVatron are divided into 36 bunches each, spaced such that bunch crossings happen at CDF every 396 ns. At high instantaneous luminosity, several $p\bar{p}$ collisions can occur with each bunch crossing. Such a collection of p and $\bar{p}$ bunches circulating in the TeVatron is called a 'store'. The spread in the physical size and momentum distributions of the beams continually increase while they circulate, reducing the instantaneous luminosity. The decision of how long to keep the store depends in part on the rate of degradation of the beams, but it also depends on the rate of anti-proton production. One day seems about optimal for a given store.

2.2 CDF Run II Detector

The CDF detector ([23]-[31]) is a multi-purpose particle detector possessing cylindrical symmetry. The detector is depicted in Fig. 7. The axis of the cylinder is aligned with the beampipe, which coincides with the z-axis in the detector coordinate system. The positive z-axis points in the direction of the proton beam, the positive x-axis points radially outward from the center of the TeVatron ring, and the positive y-axis points upward.

Figure 7: The CDF Run II Detector

Due to the cylindrical symmetry, it is convenient to define the azimuthal angle ϕ, measured from the positive x-axis in the plane perpendicular to the z-axis. The polar angle θ is measured from the positive z-axis.

There are 3 basic sub-detector types employed by CDF: tracking detectors, particle identification detectors, and calorimetry. The tracking detectors are closest to the beam pipe and are bathed in a uniform 1.4 T magnetic field aligned with the negative z-axis. A superconducting solenoid with a radius of 1.5 m and length of 4.8 m generates this magnetic field. This allows the measurement of particle momenta via the curvature of their helical trajectory through the detector.

The calorimeters are segmented into η-ϕ towers which point back to the origin of the CDF coordinate system. The electromagnetic calorimeter systems measure energy from electrons and photons, while the hadronic calorimeters measure the energy of composite particles.

The muon systems comprise the outer layer of the CDF detector. Muons are minimum ionizing particles, so they do not interact much in the calorimeters. The calorimeter systems act as absorbers for the muon detectors, filtering out almost everything but the muons.

2.2.1 Silicon Tracking System

The CDF silicon tracking system is composed of three sub-systems, shown in Fig. 8: the Silicon Vertex II detector (SVXII), the Intermediate Silicon Layer (ISL), and Layer 00 (L00). The silicon tracking system improves the measurements of track impact parameters (both longitudinal and transverse) given by the COT, reduces the fake track rate, provides stand-

22

Figure 8: An end view of the CDF Run II silicon detectors.

alone tracking for the 'forward' region $(1 < |\eta| < 2)$, and participates in the trigger decision. All three sub-systems are made up of microstrip silicon sensors arranged in a barrel shape. There are three barrels fitted end-to-end in z, totaling $\sim$90 cm. There are 7 layers of sensors in the central barrel and 8 sensor layers on the two 'forward' barrels (see Fig. 9).

The sensors come in three basic flavors, corresponding to the orientation of the microstrips: r-ϕ (axial), small angle stereo, and $90°z$. The small angle stereo (SAS) provides measurements in r, ϕ, and z. The $90°z$ sensors provide measurements along z, with the r and ϕ positions taken from the location of the sensor.

L00 is the innermost system. It is mounted on the beampipe, giving the sensors a nominal radius of 1.1 cm. The single-sided (r-ϕ) radiation hard sensors dramatically improve d_0 resolution at low p_T due to the proximity of the sensors to the interaction point and the comparatively small amount of material between the interaction point and the sensor.

The SVXII surrounds L00, extending from a radius of 2.5 cm to 10.6 cm. It provides 5 double-sided sensor layers; one side of each layer being an axial sensor, the other side being either a SAS sensor or $90°z$ sensor.

The ISL fits between the SVXII and the COT at a radii of 22 cm (central barrel), 20 cm and 28 cm (forward barrels). Each ISL layer is comprised of double-sided sensors. The ISL aids in linking SVXII hits to COT tracks and reduces the fake track rate.

23

Figure 9: An r-z view of the CDF Run II tracking detectors.

2.2.2 Central Outer Tracker

The Central Outer Tracker (COT) is cylindrical drift chamber with of more than 32,000 channels. The gas mixture employed is nominally 50/50 argon/ethane, with a small amount of alcohol and oxygen. The active volume extends radially from 40 cm to 140 cm and to ± 155 cm in z. This gives fully-instrumented coverage for $|\eta| < 1$, as can be seen in Fig. 9.

The instrumentation of the COT consists of 96 layers of sense wires, arranged into 8 concentric "superlayers". The superlayers alternate between stereo ($\pm 2°$) and axial (parallel) orientation with respect to the z-axis, with the innermost superlayer being stereo. The stereo layers allow measurement of the z-component of charged particle trajectories, in addition to $r - \phi$ information.

The superlayers are segmented into azimuthal cells consisting of one sense plane and two grounded field sheets. The sense planes (anodes) are composed of gold-plated tungsten wires 40 μm in diameter. The sense wires alternate with the field shaping wires. The cathode plane is comprised of a 6.4 μm-thick mylar plate, coated on each side with a layer of gold ~ 350 $\mathring{A}$ thick.

The entirety of the tracking system is in a 1.4 T B-field. This causes charged particles created at the interaction point to follow a helical trajectory through the tracking volume.

24

Such a trajectory can be parameterized by 5 quantities: d_0, z_0, ϕ, C, and $cot\theta$. The transverse impact parameter, d_0, measures the distance between the helix and the origin at the point of closest approach in the plane perpendicular to the z-axis. The corresponding longitudinal quantity, z_0, measures the distance of closest approach to the origin along the z-axis. The azimuthal angle of the trajectory at the point of closest approach is denoted ϕ and is measured from the x-axis. The curvature of the trajectory in the transverse plane is labeled C. The sign of the charge of the particle is identically the sign of C, while the transverse momentum of the particle is derived from the magnitude of C. The pseudorapidity of the particle is derived from $cot\theta$.

2.2.3   Central EM Calorimeter

The Central Electromagnetic calorimeter (CEM), is a sampling calorimeter located outside the solenoid and concentric with the COT. The alternating absorber and sampling layers are composed of lead and plastic scintilator, respectively. The coverage in $|\eta|$ extends to 1. The CEM is segmented into wedges measuring 15° in ϕ by 0.11 η units. Each wedge has a projective tower geometry, in which it points back to the origin of the coordinate system. The CEM has a depth of $\sim$18 radiation lengths.

2.2.4   Central and End Wall Hadronic Calorimeters

The Central Hadronic Calorimeter (CHA) abuts the CEM and shares identical geometry. It is also a sampling calorimeter, using steel for the absorber layers, and scintilator for the sampling layers. Its depth is approximately 4.7 pion interaction lengths.

The End Wall Hadronic Calorimeter (WHA) has a similar construction to the CHA and is located on the end face of the central calorimeter arches. This configuration covers θ between 30° and 45° and extends the hadronic calorimeter instrumentation out to $|\eta|$ of 1.

2.2.5   Muon Systems

The Central Muon system is located outside the central calorimeter, at a radius of 347 cm. It has cylindrical geometry, and is made up of 4 layers of rectangular drift cells. There are seven wires in each cell, all parallel to the z-axis. Each cell is 226 cm long, giving an $|\eta|$ coverage up to 0.63. The central calorimeter systems provide 5.5 pion interaction lengths of shielding.

The Central Muon Upgrade (CMP) is located outside the solenoid magnetic field return yoke. This provides an extra 60 cm of steel absorber. The central calorimeter and return yoke together provide 7.8 hadronic absorption lengths of shielding. The CMP is comprised of 4

layers of single-wire drift tubes, arranged in a box configuration around the central detector. Each drift tube is 640 cm long. The average $|\eta|$ coverage extends to 0.6, though due to the box shape, it varies with ϕ. A collection of hits in the drift tubes fit to a trajectory is called a "stub". A muon with stubs reconstructed in both the CMU and the CMP detectors is called a CMUP muon.

The Central Muon eXtension (CMX) extends the muon coverage in $|\eta|$ from 0.6 to 1.0. The 4 layers of drift tubes are arranged in conic sections which are attached to each end of the central detector. Each drift tube is 180 cm long. The calorimeter systems, steel support structure, and solenoid return yoke together provide approximately 6.2 pion interaction lengths of shielding.

Muon objects are created by matching COT tracks to stubs measured in the muon detectors. Due to multiple scattering in the absorber material, the muon stubs will be displaced from the point that the extrapolated COT track enters the muon system. Three variables are used to quantify this displacement: ΔX, $\Delta \Phi$, and ΔZ. The distance between the extrapolated track and the muon stub in the $r\phi$-plane is denoted ΔX. The quantity $\Delta \Phi$ measures the azimuthal angle between the extrapolated track and the muon stub. The longitudinal distance between the extrapolated track and the stub is quantified by ΔZ. Only ΔX is used in the track-to-stub matching algorithm, but the remaining two quantities can be used to distinguish real muons from 'fake' muons.

Fake muons come in two categories: decay-in-flight (DIF), or 'punch-through'. A DIF muon is produced when a charged pion or kaon decays somewhere between the beam pipe and the muon detector. A punch-through muon is generated when a hadron interacts late in the calorimeter and is thus not fully absorbed. This hadron then goes on to interact with the muon detectors.

2.2.6   Trigger System

At the TeVatron, an interaction happens every 396 ns, corresponding to a beam collision rate of 1.7 MHz. This rate is much higher than CDF can record; on the other hand, most of the interactions are uninteresting. In order to preferentially save interesting events and reduce the data flow rate to a manageable level, CDF employs a three-tier trigger system. Each level filters more finely than the previous level. Thus, the processing time available for each subsequent level increases, allowing more complex algorithms and more precise information to be used in deciding whether a given event is worth saving.

The first level (L1) is a synchronous hardware trigger. It renders a decision at $\sim$5 μs after each beam collision. The L1 output rate is $\sim$20 kHz. The rough objects which participate in

the L1 decision are tracks, energy clusters in the calorimeters, muon system stubs, and the total energy balance. Since the bunch crossing rate does not allow enough time to perform tracking using the offline algorithms, a rough pattern matching algorithm is used to group COT hits into pre-defined trajectories. The hardware system which performs this function is the eXtremely Fast Tracker (XFT) [32].

The second tier (L2) is an asynchronous trigger which incorporates both hardware and software algorithms. The average event processing time for L2 is $\sim$30 μs. The L2 output rate is $\sim$350 Hz. Some of the L1 trigger primitives are further refined at this level.

The final trigger level, L3, is implemented entirely in software, through a speed-optimized version of the CDF reconstruction framework. This software runs on a dedicated batch computing farm comprised of several hundred commodity PCs. The L3 output rate is $\sim$75 Hz, which is close to the limit of the mass storage system.

2.2.7   The Dimuon Triggers

A trigger path is defined as a collection of requirements involving each of the three levels of the CDF trigger system. Trigger paths are grouped together according to the physical process they are designed to select. The data comprised of a single grouping of trigger paths is called a dataset.

The primary dataset for this thesis is the Υ dataset, and the primary trigger paths are `BBBAR_CMUP3_CMU1.5` and `BBBAR_CMUP3_CMX2`. As the names suggest, these are dimuon triggers designed to select events in which a pair of bottom hadrons decay semi-leptonicaly to a pair of muons. The mean branching fraction for this family of decay modes is roughly 10%, which is large enough to create a sizeable data sample. These two paths are very similar to the Υ trigger paths, but with different $m(\mu, \mu)$ requirements. Both `BBBAR` paths require a well-measured central muon (`CMUP3`); the second muon is either central (`CMU1.5`, $|\eta(\mu)| < 0.6$), or at intermediate pseudorapidity (`CMX2`, $0.6 < |\eta(\mu)| < 1.0$).

For the central dimuon trigger path, `BBBAR_CMUP3_CMU1.5`, muon primitives are formed at Level 1 by extrapolating an XFT track to the CMU and then looking for a CMU stub in the same spatial region. Both the XFT track and the CMU stub are required to have $p_T > 1.5$ GeV/c. At Level 2, one of the CMU muon primitives is required to extrapolate to a stub in the CMP system. Since the CMP system is behind an additional absorber, the minimum XFT p_T of this muon primitive is increased to 3.0 GeV/c. This both ensures that the muon survived the passage through the absorber and minimizes multiple scattering effects. The track-to-stub matching criteria are tightened at Level 3 by applying cuts on ΔX: for both CMU and CMUP muons, the cut is $\Delta X_{CMU} < 30$ cm; additionally for the CMUP muon,

the criterion is $\Delta X_{CMP} < 40$ cm. The transverse momentum cuts are confirmed using fully reconstructed tracks (including hits in the silicon tracking detectors). Dimuons from light vector mesons and charmonia ($c\bar{c}$ bound states, such as J/Ψ) are rejected by requiring that the dimuon mass, $m(\mu, \mu)$, be greater than 5 GeV/c^2.

The criteria applied to the central muon in the central-intermediate dimuon trigger path, BBBAR_CMUP3_CMX2, are identical to the CMUP criteria applied to the central dimuon trigger path. For the CMX muon primitive, the minimum p_T is set at 2.0 GeV/c, while the maximum ΔX_{CMX} is set at 50 cm. The dimuon mass cut is identical to the central dimuon trigger path.

A set of ancillary trigger paths are also used for classifier performance studies. The dimuon trigger paths JPSI_CMU1.5_CMX2 and JPSI_CMUCMU1.5 seek to record events from the decay $J/\Psi \to \mu^+\mu^-$. The dimuon trigger requirements are presented in Tables 4 and 5, respectively.

Quantity	Criterion
$p_T(\mathrm{CMU})$	> 1.5 GeV/c
$p_T(\mathrm{CMX})$	> 2.0 GeV/c
$q(\mu_1) \cdot q(\mu_2)$	< 0
$|\Delta\phi(\mu_1, \mu_2)|$	$< 120°$
$m_T(\mu_1, \mu_2)$	< 20GeV/c^2
$|\Delta X(\mathrm{CMU})|$	< 30 cm
$|\Delta X(\mathrm{CMX})|$	< 50 cm
$|\Delta z_0(\mu_1, \mu_2)|$	< 5 cm
$m(\mu_1, \mu_2$	$2.7 < m(\mu_1, \mu_2 < 4.0$GeV/$c^2$

Table 4: Trigger muon requirements for the J/Ψ trigger path JPSI_CMU1.5_CMX2

The final trigger path, called JET_20, selects jets with transverse energy $E_T > 20$ GeV. A cone size of $R_{cone} = 0.7$ is used to reconstruct these jets with CDF's primary jet clustering algorithm, JetClu. This trigger path selects primarily light flavor QCD processes, though it also contains QCD heavy flavor and electroweak processes.

Quantity	Criterion
$p_T(\text{CMU})$	$> 1.5 \text{ GeV}/c$
$q(\mu_1) \cdot q(\mu_2)$	< 0
$|\Delta\phi(\mu_1, \mu_2)|$	$< 120°$
$m_T(\mu_1, \mu_2)$	$< 20\text{GeV}/c^2$
$|\Delta X(\text{CMU})|$	$< 30 \text{ cm}$
$|\Delta z_0(\mu_1, \mu_2)|$	$< 5 \text{ cm}$
$m(\mu_1, \mu_2$	$2.7 < m(\mu_1, \mu_2 < 4.0\text{GeV}/c^2$

Table 5: Trigger muon requirements for the J/Ψ trigger path `JPSI_CMUCMU1.5`

3 Simulation

3.1 Anatomy of an Event Generator

The basic methods of perturbative QCD allow only the computation of inclusive cross sections. In order to compare theory with experimental observations, the data need to be corrected back to the parton level. This is an inherently difficult proposition, leading to numerically ill-posed problems and large errors due to the many non-perturbative effects which occur between the high energy ('hard') interaction and the particles observed in the detector.

A better-defined process is implemented in an 'event generator' package. Such a package uses Monte Carlo techniques to generate fully exclusive interactions ('events'). Such a description of an interaction includes probabilistic models of the non-perturbative physics which give rise to the stable particles which interact with a detector. By passing a sample of events generated in this fashion through a simulation of a detector, one can apply the same reconstruction algorithms to simulated data as to real data. The theory involved in the event generator can then be tested by comparing the observed data with the simulation.

The starting point for event generation is the partonic cross section calculation $\hat{\sigma}$ for the process of interest to the user. This partonic cross section, in concert with the parton distribution function, is used to generate specific parton configurations. This configuration includes parton flavors, their four momenta, and their color connections.

The next step is the perturbative parton shower, which probabalistically models the emission from the initial and final state partons involved in the hard interaction. The parton shower implementations available have been shown to reproduce the jet structure of high transverse momentum data in both ep [22] and $p\bar{p}$ [34] collider environments.

Following the shower, a non-perturbative model of the hadronization process is applied to the event. This 'dresses' the bare quarks, grouping them together into colorless hadrons and creating new quark-antiquark and diquark-antidiquark pairs as necessary. The unstable hadrons are decayed using models of varying degrees of precision.

In leading order event generators, only the leading order terms of the cross section are used. For heavy quark (b, c) generation, this means that only flavor creation terms are simulated directly. The higher order terms only arrise through the parton shower. In order to get a sample of events which include higher order processes, one simulates the generic (*i.e.* g, u, d, s, c, b) QCD processes involving 2 partons in both the initial state and the final state; the resulting events are filtered for heavy quark content. This process is quite resource intensive.

Additionally, the sample composition of the various heavy quark production mechanisms will not be correct. This problem has been solved in the past by fitting the sample mixing proportions to data using the $\Delta\phi_{Q\bar{Q}}$ distribution (where Q represents a heavy quark). However, this seems to be an undesireable feature of an analysis which seeks to test the validity of the underlying theory.

3.2   MC@NLO

Due to the numerically significant contribution of the next-to-leading order (NLO) bottom quark production processes to the correlated cross section, it is desireable to employ the NLO corrections in an event generator. However, care must be taken in this enterprise, as the parton shower generates some NLO production effects. A subtraction scheme must be implemented in order to avoid double counting when the NLO cross section is mated to the parton shower. This implies that the NLO production processes generated by the shower must be calculated exactly. The particular parton shower implementation considered will determine the course of this calculation.

The MC@NLO generator [12] implements[6] such a subtraction scheme. Specific partonic configurations are generated according to a differential cross section, which describes the phase space distribution and potentially correlations between the final state partons. Each parton configuration generated is given a weight by the NLO cross section calculation; some weights are negative. These negative-weight events act as 'counter-events', cancelling extra NLO processes generated by the shower. An unweighting proceedure is performed using the BASES/SPRING package [13], leaving a sample of partonic events with positive and negative unit weights. Both positive and negative weights must be used when filling histograms. The partonic events can then be processed by the parton shower package without fear of double counting. The specific parton shower package used in MC@NLO is HERWIG [10].

Heavy quark generation (t and b) described in Ref. [14]. The NLO bottom quark pair cross section $\sigma_{b\bar{b}}$ is provided by the calculation of Mangano, Nason, and Ridolfi [15]. This calculation includes a discription of the transverse momentum and rapidity of the individual quarks via the differential cross section $d^2\sigma_b/dp_T dy$. It also includes a model of the angular correlation between the b quark pair, $d\sigma_{b\bar{b}}/d\Delta\phi_{b\bar{b}}$. The renormalization and factorization scales used in this analysis are set to a common scale μ_0, which is equated with the bottom quark transverse mass, $\mu_0 = \sqrt{p_T^2(b) + m_b^2}$.

While MC@NLO simulates perturbative processes quite carefully and includes models for several important non-perturbative effects, the heavy hadron decay models implemented

---

[6]Version 3.3 of the MC@NLO package is used.

leave something to be desired. The `EvtGen` [20] package is used to correct this deficiency; bottom and charm hadrons are decayed according to a specified decay table.

The advantages of an event generator such as MC@NLO over a leading order generator for the simulation of interactions involving bottom quarks are several. Firstly the sample generation is much less resource-intensive, due to the fact that the NLO processes are included in the cross section calculation. The cross section calculation includes the azimuthal correlations explicitly; it also includes a prediction for the relative proportions of the production mechanisms.

3.3   Cluster Model of Hadronization

HERWIG implements a cluster model for hadronization (also referred to as 'fragmentation'). After the perturbative parton shower, outgoing gluons are split non-perturbatively into quark-antiquark or diquark-antidiquark pairs. Color connected quark-antiquark pairs are grouped into colorless clusters (similarly for diquark-antidiquark pairs). The cluster mass and spatial size distributions peak at low values and fall quite rapidly.

Clusters then decay isotropically into hadrons by introducing a parton pair $f\bar{f}$ to break the color connection between the two cluster constituents, where $f\bar{f}$ can be a quark/anti-quark pair or a diquark/anti-diquark pair. HERWIG has a parameter array `PWT(I)` which sets the *a priori* weight for choosing the flavor of the $f\bar{f}$ pair, where `I` = 1-5, 7, corresponding to $f = u$, d, s, c, b, and $[qq']$ (*i.e.* a diquark). The hadrons are chosen randomly from tables of hadrons containing the chosen quark constituents. This picture is modified if the cluster is too light or too heavy. If the cluster mass falls below that required to generate two hadrons, the cluster becomes lightest hadron of the appropriate flavor.

If the cluster mass is above a fission threshold M_f, the cluster is split iteratively into smaller clusters. The fission threshold is determined by

$$M_f^{\texttt{CLPOW}} = \texttt{CLMAX}^{\texttt{CLPOW}} + (m_1 + m_2)^{\texttt{CLPOW}}, \tag{3}$$

where m_1 and m_2 are the quark masses for the cluster constituents. The probability density funtion used to generate the daughter cluster masses is taken to be $M^{\texttt{PSPLT}}$.

For clusters containing partons from the hard interaction, the daughter hadron optionally 'remembers' the direction of the perturbative quark (set via the `CLDIR` parameter). Gaussian smearing can be added to this hadron's direction through the parameter `CLSMR`.

Clusters containing bottom quarks near the $B\pi$ threshold mass M_{th} are allowed to decay to a single bottom hadron if $M < M_{\lim}$, where

$$M_{\text{lim}} = (1 + \texttt{B1LIM})M_{th}. \tag{4}$$

While this model is relatively simple, it has been tested thoroughly at the Z^0 pole by the LEP experiments (see Ref. [62], for example). Inclusive charge multiplicities as well as rates for an array of common particles were found to be in reasonable agreement with data, after tuning. The LEP-tuned parameters became the defaults for version 5.9 onward.

3.4 Underlying Event Models

After the hard interaction, the beam remnants must be fragmented. HERWIG models this fragmentation as a soft minimum bias interaction between the two beam clusters, according to the UA5 simulation [11]. This model uses a negative binomial distribution to describe the distribution of $p\bar{p}$ inelastic charged multiplicity n:

$$P(n) = \frac{\Gamma(n + k)(\bar{n}/k)^n}{n!\Gamma(k)(1 + \bar{n}/k)^{n+k}}. \tag{5}$$

The mean charged multiplicity $\bar{n}$ is parameterized as $\bar{n} = as^b + c$, where s is the square of the center of mass energy, and a, b, and c were fit to UA5 data. The negative binomial parameter k also depends upon $\sqrt{s}$: $1/k = a \ln s + b$.

In HERWIG, what is actually modeled is the multiplicity of soft clusters, which then hadronize according to the model described in Section 3.3. The mass M of these 'soft underlying event' clusters is modeled as $(M - m_1 - m_2 - a)e^{-bM}$, where m_1 and m_2 are the masses of the consituent quarks. The cluster transverse momentum distribution is flavor specific and modeled as

$$P(p_T) \propto p_T \exp(-b_i\sqrt{p_T^2 + M^2}), \tag{6}$$

where the flavor is specified by the index i and grouped according to $\{\{d, u\}, \{s, c\}, \{qq\}\}$. Since the particles generated by this model are primarily soft, they do not affect the definition of the jets from the hard interaction. The default parameters for this model are given in Table 6.

It was found in Ref. [8] that a complete description of the underlying event must include a model of multiparton interactions in order to match the CDF Run I data. Multiparton interactions are secondary 'semi-hard' interactions between partons in the same hadrons which collided to produce the hard interaction. The JIMMY [9] package was used to model these interactions.

Name	Description	Default
PMBN1	a in $\bar{n} = as^b + c$	9.110
PMBN2	b in $\bar{n} = as^b + c$	0.115
PMBN3	c in $\bar{n} = as^b + c$	-9.5
PMBK1	a in $1/k = a \ln s + b$	0.029
PMBK2	b in $1/k = a \ln s + b$	-0.104
PMBM1	a in $(M - m_1 - m_2 - a)e^{-bM}$	0.4
PMBM2	b in $(M - m_1 - m_2 - a)e^{-bM}$	2.0
PMBP1	p_T slope b_i for d, u	5.2
PMBP2	p_T slope b_i for s, c	3.0
PMBP3	p_T slope b_i for qq, qq	5.2

Table 6: Soft Underlying Event parameters for HERWIG and their default values.

JIMMY views the beam hadrons as relativistically flattened discs of sea partons, the momentum fractions of which are described by the parton distribution functions. The spatial distribution of the sea partons are parameterized as a function $A(b)$ of the impact parameter b between the two hadron discs. The functional form of $A(b)$ is assumed to be the convolution of the form-factor distributions for the beam hadrons:

$$A(b) = \int d^2\mathbf{b}' G_p(\mathbf{b}') G_{\bar{p}}(\mathbf{b} - \mathbf{b}')$$

where the $G(\mathbf{b})$ functions are taken to be the electric form-factor of the proton ($G_p = G_{\bar{p}}$):

$$G_p(\mathbf{b}) = \int \frac{d^2\mathbf{k}}{(2\pi)^2} \frac{\exp(\mathbf{k} \cdot \mathbf{b})}{(1 + \mathbf{k}^2/\mu^2)^2}.$$

The parameter μ^2 is the inverse square of the proton radius, defaulting to 0.71 GeV^2.

The complete view of the interaction between the beam hadrons is then viewed as collection of n individual interactions; one of type b with cross section σ_b (the hard interaction), and $n - 1$ of type a with cross section σ_a. Type a interactions are implemented as strong interactions between light partons. The (leading order) cross section for these interactions is the same as used in standard HERWIG (*i.e.* not MC@NLO). Consequently a transverse momentum threshold $\widehat{p}_T^{\min}$ is required to regulate the infrared divergences as $p_T \to 0$. Type b interactions are assumed to be rare enough that at most one occurs per event. The probability for having exactly n scatters, when at least one has occured, is approximated as

$$P(n) \approx \frac{\int d^2b \, n \frac{(A(b)\sigma_a)^n}{n!} e^{A(b)\sigma_a}}{\sigma_a}, \quad n \geq 1. \tag{7}$$

After generating the number of interactions according to the probability mass function[7] defined in Eq. 7, the standard HERWIG parton shower and cluster hadronization are applied.

3.5   Realistic Simulation

As mentioned in Section 3.1, the best path to compare data and theory is by using the theory to generate 'events' which should then be passed through a simulation of the detector. The detector simulation used is of the realistic type: the responses of the various subdetectors to the stable particles input by the event generator are simulated in some detail. These responses are coded into the same data structures as hardware uses. This allows the reconstruction package to treat data and MC (nearly) identically. The calibrations and

---

[7]A probability mass function (PMF) is analogous to a probability density function (PDF). However a PMF is defined only for discrete random variates, while a PDF is defined only for continuous random variates.

alignments of all the subdetectors are set on a run-dependent basis, increasing the realism of the simulation by tracking the changes to the detector over time.

GEANT3 [19] is used to simulate particle transport through a model of the detector material. Material interactions, multiple scattering, and decay of long-lived particles (*e.g.* K_s) are also included.

A list of 687 representative runs are simulated for this analysis. Additional interactions are simulated by adding some number of minimum bias interactions to each event. The number of minimum bias interactions added is generated according to a Poisson distribution with mean N_{MB} given by

$$N_{MB} = L_1 \times \sigma_{p\bar{p}}^{inelastic} \times 132\text{ns} \times (159/36), \tag{8}$$

with L_1 the average instanteous luminosity for a given run, and $\sigma_{p\bar{p}}^{inelastic}$ the inelastic proton - anti-proton cross section, measured at CDF as 61.7 mb [41]. The remaining terms are appropriate for the beam structure at the TeVatron during RunII.

A simulation of the Level 1 trigger systems is also employed.

4 Evolutionary Optimization

4.1 Definition of Optimization

An optimization algorithm attempts to find the solution vector $\boldsymbol{x}^*$ which optimizes an *objective function* $f(\boldsymbol{x})$:

$$f(\boldsymbol{x}^*) < f(\boldsymbol{x}) \quad \forall \ \boldsymbol{x} \ \in \ \boldsymbol{\mathcal{X}}, \tag{9}$$

where $\boldsymbol{\mathcal{X}}$ is the space of possible values of the vector $\boldsymbol{x}$. This optimization can be either the minization or the maximization sense; for specificity, the minimization sense is assumed here[8]. The objective function will in general be a non-linear, multimodal function of the D dimensions of the problem. While the objective function of an optimization problem is frequently a function in the algebraic sense, it need not be. Any numerical system which generates a response to a set of input conditions can be thought of as an objective function.

In general, there are two types of optimization algorithms: single-objective and multi-objective. *Single-objective optimization algorithms* (SOA) specify the objective function as a scalar function. This works well for a large class of problems. However many problems have multiple, competing objective functions. Such problems are called *Multiobjective Optimization Problems* (MOPs), and specify a vector of objective functions, $\boldsymbol{f}$. A MOP is then specified as

$$\boldsymbol{f}(\boldsymbol{x}^*) < \boldsymbol{f}(\boldsymbol{x}) \quad \forall \ \boldsymbol{x} \ \in \boldsymbol{\mathcal{X}}. \tag{10}$$

A *multi-objective optimization algorithm* (MOA) optimizes each objective funtion f_i independently and simultaneously. Note that problems for which the multiple objectives have been aggregated into a single objective function through *e.g.* a weighted sum, are classified as single-objective optimization problems.

4.2 Motivation for Direct Function Optimization

It has been demonstrated that the default parameters for the HERWIG parton shower Monte Carlo package are not well tuned with respect to the fractional energy propagated from the bottom quark to the bottom hadron through the fragmentation process (see Ref. [63]). Due to the absence of a pre-packaged optimization algorithm which could handle such a

---

[8]a maximization problem can be turned into a minimization problem simply by multiplying the objective function by negative one.

stochastic, resource-intensive objective function as the HERWIG event generator, we decided to develop a SOA to address this problem. The desirable features sought were to:

- execute in reasonable time despite resource-intensive objective function;

- handle non-linear objective functions;

- handle a mixture of discrete and continuous tuning parameters;

- be numerically stable for stochastic objective functions;

- provide interface to external processes for objective function evaluation;

- be parallelizable, and

- submit jobs to a batch computing farm automatically.

The default methods for optimizing parameterized simulation in High Energy Physics are adapted directly from gradient-based algorithms, such as Newton's Method. Typically, a grid search is used to limit the number of points in parameter space and thus the total running time. However, the grid configuration is somewhat subjective and leaves open the possibility of the global optimum lying between grid points without affecting the objective function value at the grid points. Another critique concerns the numerical estimation of the objective function gradient. Such estimation algorithms require several function evaluations per trial solution. Furthermore, applying a numerical gradient algorithm to a stochastic function, such as the HERWIG parton shower simulation package, can lead to an ill-conditioned problem.

"Direct" function optimization algorithms, on the other hand, do not estimate the objective function gradient in the generation of new trial solutions. Instead, information from previously explored regions of parameter space is used to produce subsequent trial solutions. One such algorithm is the classic Nelder-Mead Simplex algorithm [61].

By avoiding numerical estimation of the objective function gradient, the requirement of the existence of the gradient is circumvented. Thus, the method can handle non-continuous objective function topology, and it is possible to formulate problems with a mixture of discrete and continuous parameter space dimensions. The possibility of ill-conditioned problems vanishes due to the absence of a need to calculate a numerical derivative.

4.3 Evolutionary Algorithms for Direct Function Optimization

Evolutionary algorithms comprise a sub-class of direct function optimization algorithms. Borrowing from the evolutionary biology lexicon, we define each iteration of the algorithm as a "generation" and the current collection of N_p trial solutions as the "population," denoted P^G. Each trial solution, or "individual," is a D-dimensional vector of parameters, denoted $\boldsymbol{x}_i^G$. The population structure promotes exploration of the solution space through diversity of individuals and lends itself naturally to parallelization.

The population evolves with successive generations through a procreation process. Stochastic elements of the procreation algorithm enable escape from local minima. A selection process preserves traits which further the optimization goal.

4.4 Differential Evolution

A popular evolutionary algorithm for single-objective optimization is Differential Evolution [58]. The basic algorithm is illustrated with pseudocode in Fig. 10.

Procreation is limited to one child per parent, and the child and parent compete for the parent's spot in the next generation. Thus the population size is fixed. Procreation occurs through two successive steps termed mutation and crossover. Mutation promotes exploration of the parameter space and escape from local minima, while the crossover process promotes diversity in the population in a controlled fashion.

In the first step, a mutant $\boldsymbol{\omega}$ is formed via a linear combination of a "base" vector β and a "difference" vector δ derived from the current population:

$$\boldsymbol{\omega} = \boldsymbol{\beta} + F \cdot \boldsymbol{\delta}$$

The base vector is either the best individual in the current generation $\boldsymbol{x}_{best}^G$ (algorithm variant *DE/best*), or a random individual $\boldsymbol{x}_{r_0}^G$ (variant *DE/rand*, see Fig. 11). The vector $\boldsymbol{\delta}$ is the difference of a pair of unique individuals (the base vector is excluded). The scale factor F ranges from 0 to 1 and controls the amplification of the difference vector $\boldsymbol{\delta}$. Depending on the value of F and the objective function topology, it can aid the algorithm in escaping from a local minimum. Alternatively, a poor choice of F will cause new trial solutions to be located far from their parents; this results in a lengthening of the number of generations required for convergence.

The child $\boldsymbol{x}_c$ is formed in the crossover operation from the mutant $\boldsymbol{\omega}$ and the parent $\boldsymbol{x}_i^G$ (see Fig. 12). At least one randomly chosen element from the mutant is copied to the child. The probability for any of the remaining elements of $\boldsymbol{x}_c$ to come from the mutant is Cr. The probability mass function is then the binomial distribution.

```
Init(P^{G=1})
while not converged():
    // Loop over Population members
    for i ∈ [1, N_p]:
        // Mutation, ω = β + F·δ
        ω = GetMutant(P^G)
        // Crossover
        j_r = randInt(1, D)
        // Loop over vector elements
        for j ∈ [1, D]:
            if rand(0, 1) < Cr or j == j_r:
                x_{c,j} =  ω_j
            else:
                x_{c,j} = x^G_{i,j}
        // Selection
        if f(x_c) <= f(x^G_i):
            x^{G+1}_i = x_c
        else:
            x^{G+1}_i = x^G_i
    G += 1
```

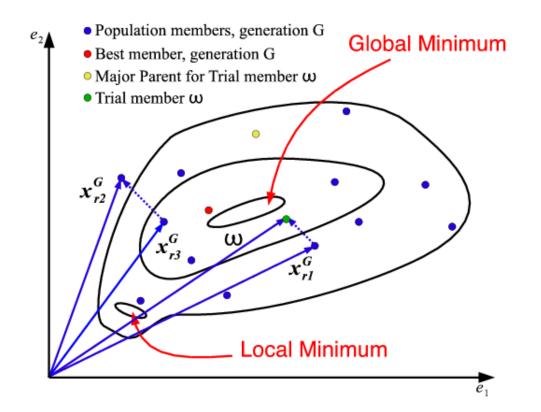Figure 10: The Differential Evolution Algorithm.

Figure 11: An example of the mutant $\boldsymbol{\omega}$ construction using *DE/rand* for a two-dimensional objective function.
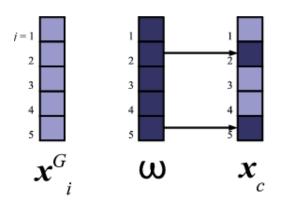

Figure 12: An example of the crossover process used to form the child $\boldsymbol{x}_c$ from the mutant $\boldsymbol{\omega}$ and the parent $\boldsymbol{x}_i^G$.

Once the child has been generated, it's "fitness" is determined by evaluating the objective function. The parent and the child compete according to the greedy criterion: the individual with the more optimal fitness value survives to the next generation.

The convergence criterion is similar to the fractional tolerance used in the Simulated Annealing implementation of Ref. [59]. The best n members of the population are found, and the difference between the objective function values of the most optimal and the least optimal of this set is compared to their average:

$$f_{tol} = \frac{2|f_{hi} - f_{lo}|}{(f_{hi} + f_{lo})} \tag{11}$$

When f_{tol} drops below a user-specified level α for at least m generations, the algorithm declares convergence.

Differential evolution is well-suited to optimize non-linear, non-differentiable objective functions. It has been shown to perform well on a variety of standard test functions [58]. For resource-intensive objective functions, the parallel nature of the algorithm makes it ideal for use with a batch computing farm.

The main drawback of this algorithm is that the performance of the algorithm is dependent, to a moderate degree, on the control parameters F, Cr, and N_p. The choice of these parameters is dependent upon the problem at hand, though a self-adaptive algorithm can help ameliorate this issue. A self-adaptive algorithm is one which adapts its parameters as it runs to suit the changing nature of the problem. In the case of an evolutionary optimization algorithm, the dynamic aspect of the problem is the local geometry of the objective function space being probed at any given time.

4.5 Two-Phase Adaptive Differential Evolution

4.5.1 Introduction

For resource-intensive objective functions, minimizing the overall number of function evaluations was found to be not the best resource meter. The availability of a batch computing farm environment made it clear that what was necessary to minimize was the overall number of generations, with only secondary regard to the number of functions being evaluated in each batch job. One also continues to desire a convergence on the global optimum. These goals can be restated as [64]

1. exploration of parameter space and

2. exploitation of "fit" individuals.

A novel approach to quantifiably satisfying these two goals was developed. Once these two goals have been quantified (as described in Section 4.5.3), it becomes possible to embed a multi-objective optimization problem within the larger single-objective optimization problem. The parameter vector for this embedded optimizer is simply the Differential Evolution control parameters F and Cr used in the problem of interest to the user.

This main problem can then be divided into two phases. In the first phase, the DE parameters are chosen to maximize exploration of the parameter space. In the second phase, the DE parameters are chosen so as to fully exploit any potential optimal points. The population size N_p remains fixed in this algorithm.

The MOA is a simple extension of Differential Evolution[9] to multiple objectives. Therefore it has a population structure. This population is maintained separately from the population of the primary SOA. The comparison operation between two individuals for the self-adaptive algorithm is described in Section 4.5.2; the population update function is described in Section 4.5.4; the objective functions for the self-adaptive algorithm is described in Section 4.5.3; and the choice of Differential Evolution control parameters for a given reproduction operation in the primary SOA is described in Section 4.5.5.

4.5.2 The Pareto Front

MOPs frequently have competing objectives such that it is not possible to find an $\boldsymbol{x}^*$ which minimizes all functions in $\boldsymbol{f}$ simultaneously. Therefore instead of requiring that the solution vector be optimal in all m dimensions of $\boldsymbol{f}$ simultaneously, the concept of domination [65] is employed to rank trial solutions. The solution $\boldsymbol{x}_a$ emphdominates the solution $\boldsymbol{x}_b$, written $\boldsymbol{x}_a \prec \boldsymbol{x}_b$, if and only if

$$
\begin{aligned}
f_i(\boldsymbol{x}_a) \quad &\leq f_i(\boldsymbol{x}_b) \ \forall \ i \ \in \ \{1, 2, \ldots, m\} \\
\text{and} \quad &\exists \ j \ \in \ \{1, 2, \ldots, m\} \text{ where } f_j(\boldsymbol{x}_a) < f_j(\boldsymbol{x}_b).
\end{aligned}
\tag{12}
$$

In other words, $\boldsymbol{f}(\boldsymbol{x}_a)$ is at least as good as all other points in $m-1$ dimensions, and strictly better than all other points in at least one dimension. Solution vectors which are not dominated by any other members of the population are called non-dominated individuals. In general a MOP will have a set of such points, called the Pareto set, which form a surface in the m-dimensional space of objective functions $\mathcal{F}$ of the MOP. This surface is called the Pareto front, and it evolves with each successive generation of the optimization algorithm. The Pareto front represents the objective function compromises currently available to the user.

---

[9]The strategy *DE/rand* was used for reproduction of the individuals in the MOA.

Only one operating point will be chosen among all points in the Pareto set, corresponding to one set of trade-offs between the objective functions.

4.5.3 Objective Functions for Self-Adaptation

The two goals of this self-adaptive algorithm stated in Section 4.5.1 must be quantified before they can be optimized. The *exploration* goal was quantified analogously to the 'crowding distance' of the NSGA-II algorithm [66]. For each member of the primary SOA population, the maximal hypersphere which does not include any other points in solution vector space ($\boldsymbol{\mathcal{X}}$ is found. The larger the hypersphere, the lower the local density of trial solutions. Selecting DE control parameters which maximize the average hypersphere promotes exploration of the solution vector space.

The *exploitation* goal was quantified as the normalized gain achieved during the last generation of the primary SOA in which a particular DE control parameter vector was used:

$$g(\boldsymbol{x}) = \frac{f_{\text{primary}}^{\text{child}} - f_{\text{primary}}^{\text{parent}}}{f_{\text{primary}}^{\text{max}} - f_{\text{primary}}^{\text{min}}}. \tag{13}$$

The normalization factor is the inverse of the range of function values in the primary SOA's current population.

4.5.4 The MOA Update Function

The DE control parameter population is updated via a simple elitist algorithm. After each generation, the dominated individuals in the population are removed, leaving only the non-dominated individuals. These individuals then participate in reproduction during the next generation of the self-adaptive algorithm.

4.5.5 Choice of DE Control Parameters

The motivation for developing this algorithm was to accomplish self-adaptation of the DE control parameters for the primary SOA. However, the MOA presents a set of possible solutions. The particular DE control parameters used in a given generation of the primary SOA are chosen as follows. During the exploration phase, the current MOA population is sorted according to increasing exploration function value. During the exploitation phase, the MOA population is sorted according to increasing exploitation value. In each phase, the MOA individual (containing the vector of DE control parameters) used in each primary SOA reproduction operation is chosen randomly from the best ζ percent of the MOA population, where the population has been sorted according to the current phase. The DE control

parameters for the primary SOA chosen in this way were then used for every reproduction operation in the current SOA generation.

The generation at which the MOA phase shifts from exploration to exploitation is controlled by the user. It is set as a fraction of the maximum number of allowable SOA generations.

5 Simulation Tuning

The default parameter settings for HERWIG were derived through fits to LEP data acquired at the Z^0 pole (*e.g.* Ref. [62]). However Ref. [63] showed that the default fragmentation parameters did not reproduce the bottom hadron energy fraction[10] as measured in e^+e^- collider data at LEP and SLAC. Furthermore, it is not at all obvious that fragmentation in 'clean' e^+e^- collisions should be portable to the hadron collider environment, which is color-rich due to the composite nature of the beam particles [50].

The other major area in which HERWIG has been shown to be lacking is the underlying event model [44]. It produces too many soft particles. The situation is still unsatisfactory when the JIMMY package is used to simulate multiparton interactions. The default JIMMY parameters[11] were set according to preliminary tunes to CDF data; the authors indicated in Ref. [9] that more work was needed in this area.

PYTHIA has been shown to agree with the data to a fair degree in both bottom quark fragmentation and underlying event simulation. The PYTHIA underlying event parameters[12] used were tuned to CDF Run I data [8]. The resulting configuration is termed 'PYTHIA TuneA'. In the course of the CDF B_s mixing analysis, it was shown [47] that PYTHIA TuneA models the data well with respect to the multiplicity and p_T spectra of several particle species found in fixed cones about bottom mesons.

In order to streamline the tuning process and reduce the resource requirements, it was determined that the PYTHIA generator would be used to represent data. This allows the generator data to be compared directly, without the need for the CPU-intensive detector simulation, reconstruction software, or trigger simulation. The difficult and error-prone process of 'unfolding' the data is also avoided by this choice.

An implementation of the Differential Evolution algorithm (see Section 4) is used to perform the optimization.

5.1 Underlying Event Tuning

In order to tune the underlying event model, the observables must first be isolated from the 'interesting' physics, insofar as that is possible. Following the methods of Ref. [8], the starting point is an event topology selection: $b\bar{b}$ events which are 'back-to-back' in azimuth are required. This selection applies to the two ground state bottom hadrons (generically denoted

---

[10]$x_b \equiv E_{H_b}/E_b$

[11]JIMMY was developed for *ep* collisions at HERA. Version 4.1 is the first version available for use with hadron-hadron collisions.

[12]This includes the multiparton model parameters

H_b) with the largest transverse momentum and with $|\eta(H_b)| < 1$ and $p_T(H_b) > 6$ GeV/c. No jet clustering was implemented. 'Back-to-back' is defined as $\Delta\phi(H_{b,1}, H_{b,2}) > 150°$. Two 'transverse' regions can be defined in the r-ϕ plane as $60° < \phi' < 120°$ and $240° < \phi' < 300°$, with ϕ' being the azimuthal angle measured from the flight direction of the highest p_T H_b (the 'lead' H_b). The 'transMIN' region is then the 'transverse' region with the minimum charged particle scalar sum p_T. This region minimizes contributions from initial- and final-state radiation, and thus it is most sensitive to the beam-beam remnant and multiparton interaction components in each event. Figures 13 and 14 compare the default MC@NLO plus JIMMY with PYTHIA TuneA for back-to-back $b\bar{b}$ events.

The charged pion and charged kaon multiplicities in the 'transMIN' region versus p_T(Lead H_b) are shown in Fig. 13. These histograms were constructed by filling the appropriate p_T(Lead H_b) bin in each event with the number of particles found in the 'transMIN' region multiplied by the appropriate weight ($+1$ or -1). The charged particles are required to have $p_T > 0.5$GeV/c and $|\eta| < 1$. After all events were processed, the histograms were scaled by $1/N_{eff}$, with effective number of events $N_{eff} \equiv N_+ - N_-$, N_+ (N_-) being the number of positive (negative) weight events. The contents of each bin were then divided by the bin width to make it a density histogram. The multiplicities defined in this way give the probability that an interaction will produce a given hadron (π,K,p) at a given p_T. Broadly, the pion and kaon multiplicities are too high across the p_T(Lead H_b) spectrum.

The pion and kaon multiplicities were reduced somewhat by allowing gluon splitting to diquarks. The maximum scale for $g \to [qq'][\bar{q}\bar{q}']$ was set at QDIQK = 5.0 GeV, and the rate parameter is set to the default value: PDIQK = 5.0.

The 'transMIN' charged particle p_T spectra are shown in Fig. 14 for pions and kaons. The normalization for these spectra was calculated at the end of event processing in a similar way to the particle multiplicity *vs* p_T(Lead H_b) histograms. First the particle p_T spectra were scaled to the number of effective MC@NLO events, then the contents of each bin were divided by its bin width. The p_T spectrum is too soft for both species. Thus the default HERWIG and JIMMY settings produce too many particles, primarily in the low-p_T regime.

As demonstrated in References [8] and [44], the default HERWIG underlying event model produces primarily soft charged particles. The only sensible way to increase the mean charged particle p_T is by adding a multiparton interaction (MPI) model, such as that of PYTHIA. JIMMY is the only MPI model fully integrated with HERWIG. It has two main parameters: PTJIM, and JMRAD(73). The former sets the minimum transverse momentum threshold for secondary (MPI) interactions, while the latter sets the inverse square proton radius (the inverse square anti-proton radius is set identically to that of the proton). As noted in Section 3.4, the inverse square proton radius is involved in the calculation of the

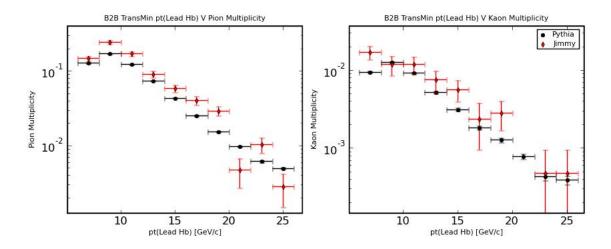Figure 13: UE Defaults: charged pion (left) and charged kaon (right) multiplicity *vs* $p_T(\text{Lead } H_b)$.
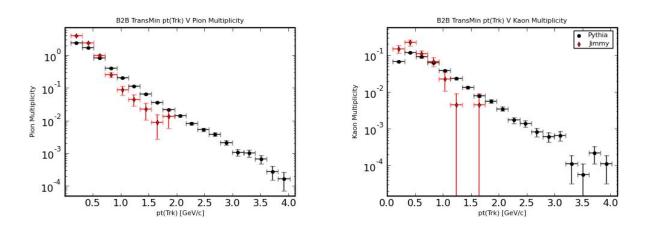

Figure 14: UE Defaults: charged pion multiplicity *vs* $p_T(\pi^{\pm})$ (left) and charged kaon multiplicity *vs* $p_T(K^{\pm})$ (right).

mean number of jet pairs produced by secondary interactions.

Ref. [44] tuned JIMMY to the 'transverse' $\sum E_T$ per unit η-ϕ area versus p_T(Lead jet) as measured in CDF Run II data. `PTJIM` was the only parameter tuned (the result being `PTJIM` = 3.25 GeV/c). The effect was that the 'transverse' charged particle multiplicity increased without increasing the mean charged particle p_T.

Therefore it is necessary to tune the JIMMY parameters (`PTJIM` and `JMRAD(73)`) to both the charged particle multiplicity and p_T simultaneously. Charged pions and charged kaons were considered separately in order to provide a baseline for bottom quark fragmentation tuning, which was expected to be sensitive to the charged particle composition. The objective function under optimization was then formed as follows. For a given trial solution, the charged particle multiplicities were plotted against

- p_T(Lead H_b), and

- $p_T(particle)$,

for $\pi^{\pm}$ and $K^{\pm}$ separately. For the i^{th}plot, a χ^2 measure of similarity between JIMMY and PYTHIA is formed via

$$\chi_i^2 = \boldsymbol{\delta}_i \boldsymbol{V}_i^{-1} \boldsymbol{\delta}_i, \tag{14}$$

where $\boldsymbol{\delta}_i$ is the bin-by-bin difference vector between the PYTHIA distribution and the JIMMY distribution, and $\boldsymbol{V}_i$ is the covariance matrix. The covariance matrix was constructed as the sum of the covariance matrices for PYTHIA and JIMMY, each being diagonal and derived from the statistical errors only. No bin-by-bin correlation was estimated. The objective function was then constructed as a sum of the similarity metrics:

$$f(\boldsymbol{x}) \equiv \sum \frac{\chi_i^2(\boldsymbol{x})}{NDF_i}, \tag{15}$$

with NDF_i corresponding to the number of degrees of freedom for plot i.

For each evaluation of $f(\boldsymbol{x})$, 500,000 $b\bar{b}$ events were generated with MC@NLO. The ground state bottom hadron p_T spectrum was reweighted to the spectrum measured at CDF [45] in both PYTHIA and MC@NLO. The 'accept-reject' method was applied to a randomly selected fiducial H_b. After reweighting, event selection, and canceling of the MC@NLO negative weight events, each function evaluation was left with between 1,000 and 1,500 events.

The Differential Evolution population size N_p was fixed at 20 members. The control parameters were determined using the two-phase adaptive algorithm described in Section 4.5,

using $\zeta = 0.2$; the exploitation phase was initiated after 5 generations. The tuning converged after 42 generations. Fig. 15 shows the evolution of the cost function for the 'best' member, the mean cost (taken over the whole population), and the fractional tolerance (see Eq. 11).
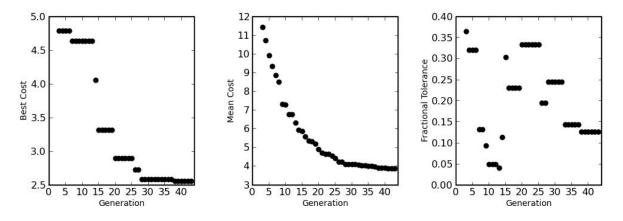

Figure 15: UE tune convergence history

The results of the underlying event tuning are presented in Table 7. The tuned value of the inverse square proton radius is quite close to the JIMMY default of 0.71 GeV^{-2}, which was derived from the proton's electromagnetic form-factor. This indicates that the mean number of jet pairs created in secondary interactions as calculated with the default parameters is about right. The tuned value of PTJIM is larger than that of [44], due primarily to the requirement that the charged particle p_T spectra and multiplicities were tuned simultaneously.

The performance of this tune with respect to the charged particle multiplicity vs p_T(Lead H_b) can be seen in Figures 16. Comparing this figure to Fig. 13, the agreement in the pion multiplicity improves in the medium p_T(Lead H_b) region (12 to 16 GeV/c), while the improvement in the kaon multiplicity is more uniform in p_T(Lead H_b). There are still too many pions in the low p_T(Lead H_b) regime, however.

Figure 17 shows the effect of this tune on the charge particle multiplicity vs p_T. The difference with respect to Fig. 14 cannot be categorized as 'improvement'. Clearly, the JIMMY parameters don't affect the charged particle p_T spectra significantly. It will be demonstrated in the next section that tuning the HERWIG fragmentation and parton shower models can improve the situation.

5.2 'b-Jet' Particle Content Tuning

As this thesis is based upon bottom jets and uses MC for training, the bottom quark fragmentation model implemented in MC must be as near to the data as can be accomplished.

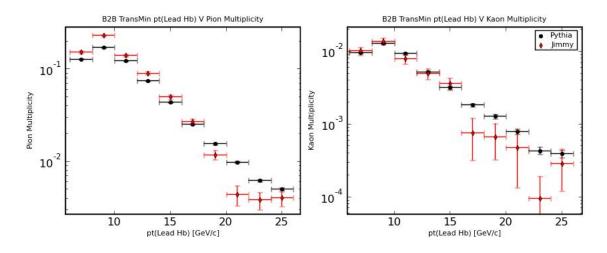| Parameter	Default Value	Tuned Value
PTJIM	3.0 GeV/c	5.09 GeV/c
JMRAD(73)	0.71 GeV^{-2}	0.698 GeV^{-2}

Table 7: Tuned UE parameters

Figure 16: UE tune results: charged pion (left) and charged kaon (right) multiplicity *vs* p_T(Lead jet).
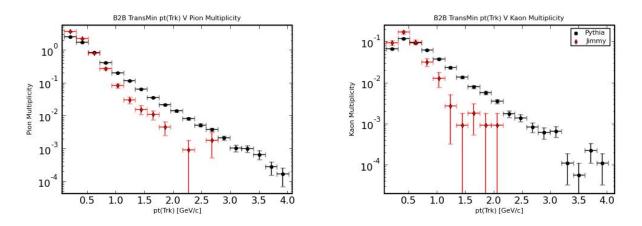

Figure 17: UE tune results: charged pion multiplicity *vs* $p_T(\pi^{\pm})$ (left) and charged kaon multiplicity *vs* $p_T(K^{\pm})$ (right).

As mentioned in Section 5, Ref. [47] found that PYTHIA TuneA models the particle content around bottom hadrons well. The methods used to tune the HERWIG bottom quark fragmentation here incorporate the strategy of Ref. [47], in which no jet clustering is performed. Rather, the charged particle content within a fixed η-ϕ cone of $\Delta R = 0.7$ around a bottom hadron is examined (H_b daughters having been first discarded). The event selection is identical to that used for underlying event tuning in Section 5.1. The $p_T(H_b)$ spectrum was reweighted for both PYTHIA and MC@NLO in the same way.

The default parameter set results are similar to the default underlying event settings: HERWIG generates too many particles with a transverse momentum spectrum which is too soft. The default charged pion, charged kaon and proton multiplicities around light bottom mesons (B_d and B_u) and B_s mesons are compared with PYTHIA TuneA in Figures 18 and 19, respectively. Here 'multiplicity' refers to the mean number of particles for bottom hadrons of a particular species. The default p_T spectra for these species are similarly compared in Figures 20 and 21.
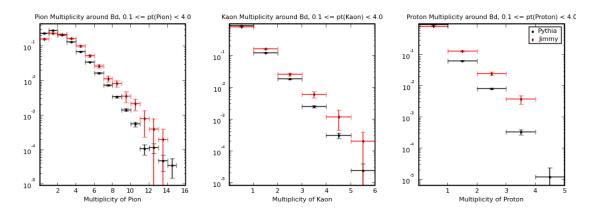

Figure 18: HERWIG Defaults: track multiplicity around B_d mesons.

The starting parameter set for 'b-jet' particle content tuning is the underlying event tune described in Section 5.1. Additional parameter settings and modifications to HERWIG were found to improve the tuning result. In order to reduce the proton multiplicity, gluon splitting to diquark pairs was vetoed in the soft underlying event (*i.e.* the UA5 model, not JIMMY). The HERWIG parameter array PWT(I) controls the *a priori* weight for choosing the parton pair $f\bar{f}$ during cluster hadronization (described in more detail in Section 3.3). The diquark/anti-diquark weight PWT(7) was reduced from the default of 1.0 down to 0.5. This latter setting had the effect of increasing the pion and kaon multiplicities in the low-p_T regime, while decreasing the proton multiplicity significantly.

The valence partons of the colliding hadrons have no intrinsic transverse momentum by default, though the PTRMS parameter allows control of the width of the gaussian distribution
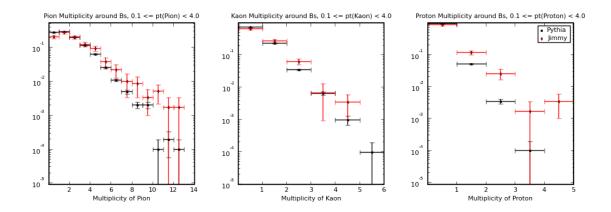
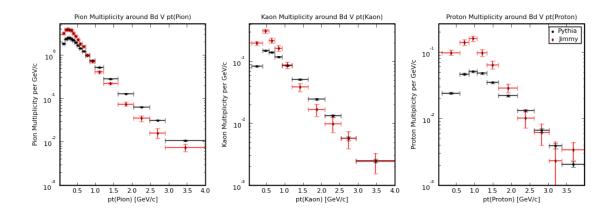Figure 19: HERWIG Defaults: track multiplicity around B_s mesons.

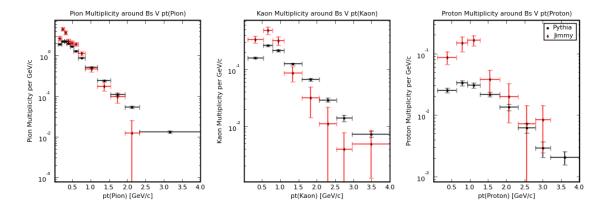Figure 20: HERWIG Defaults: track multiplicity around B_d mesons *vs* p_T(track)

Figure 21: HERWIG Defaults: track multiplicity around B_s mesons *vs* p_T(track)

used to generate such an intrinsic p_T. The value of `PTRMS` was increased from zero to 1 GeV/$c$in order to harden the p_T spectrum of particles produced by initial state radiation; the primary interaction is not affected by this parameter.

In an attempt to reduce the multiplicity of soft pions, the meson weight structure `REPWT(L,J,N)` was set to zero for all values of `L`, `J`, and `N`. Since this structure holds the *a priori* weights for generating $N^{(2S+1)}L_J$ mesons relative to the pionic $N = 1$, $0^{-(+)}$ states, the effect is that only *s*-wave mesons are allowed.

The parameters selected for tuning were the light quark mass (with $m_d = m_u$), the strange quark mass m_s, the bottom quark mass m_b, the gluon effective mass m_g, and the cluster model parameters `CLMAX(I)`, `CLPOW(I)`, `PSPLT(I)`, `CLSMR(I)`, and `PWT(3)`. For the array parameters, the first component (`I=1`) affects non-bottom flavor clusters, while the second component affects only bottom flavor clusters. In the process of tuning the hadronization parameters, it was found that converting the parameters `CLMAX` and `CLPOW` into arrays of length two achieved better performance. The parameter `PWT(3)` sets the *a priori* weight for choosing a strange quark/anti-quark pair when a cluster decays into two hadrons. The cluster model parameters are described in more detail in Section 3.3.

The objective function being tuned was a likelihood. The distributions which enter the likelihood are the charge particle multiplicities and transverse momentum spectra. Separate distributions were used for charged pions, charged kaons, and protons around light bottom mesons (B_d and B_u, hereafter denoted simply as B_d) and B_s mesons, for a total of twelve distributions. The negative log-likelihood was then formed as

$$l = -\frac{1}{\sum_{i,j,k} w_k} \cdot \sum_{i,j,k} w_k \ln f_{\text{mult}}^{(jk)}(n_{ijk}) - \frac{1}{\sum_{i,j,k,m} w_k} \cdot \sum_{i,j,k,m} w_k \ln f_{p_T}^{(jk)}(p_T^{(ijkm)}), \qquad (16)$$

with $i = [1, N_{\text{jet}}]$, $j \in \{B_d, B_s\}$, $k \in \{\pi, K, p\}$, $m = [1, n_{ijk}]$; n_{ijk} is the number of particles of type k in the ith 'b-jet' of type j. Each term of l is multiplied by a weight w_k, which only depends upon the particle type. The probability mass function $f_{\text{mult}}^{(jk)}(n_{ijk})$ is the normalized reference multiplicity histogram. The probability density function (PDF) $f_{p_T}^{(jk)}(p_T^{(ijkm)})$ is the normalized reference transverse momentum spectrum. These p_T reference distributions were fit with log-normal PDFs,

$$f(x; l, s, \sigma) = \frac{1}{\sigma(x - l)\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln((x - l)/s)}{\sigma}\right)^2\right]. \qquad (17)$$

The fits can be seen in Figures 22 and 23. The fit parameters can be found in Tables 8 and 9.

The differential evolution parameters used for this tuning were similar to the underlying

54

Figure 22: Log-normal fits to the PYTHIA TuneA multiplicity *vs* $p_T(particle)$ for pions (left), kaons (middle), and protons(right) around light bottom mesons (B_d and B_u).

Figure 23: Log-normal fits to the PYTHIA TuneA multiplicity *vs* $p_T(particle)$ for pions (left), kaons (middle), and protons(right) around B_s mesons.

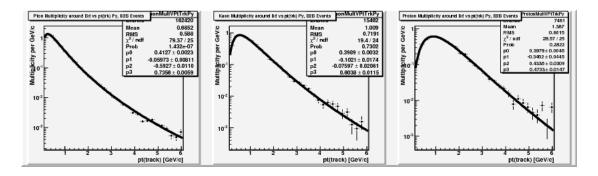| Particle	χ^2/ndf	Parameter	Value
$\pi^\pm$	79.37 / 25	l	-0.05973 ± 0.00811
		$\log(s)$	-0.5927 ± 0.0110
		σ	0.7356 ± 0.0059
$K^\pm$	19.4 / 24	l	-0.1021 ± 0.0174
		$\log(s)$	-0.07597 ± 0.02081
		σ	0.6038 ± 0.0115
$p/\bar{p}$	28.57 / 25	l	-0.3452 ± 0.0445
		$\log(s)$	0.4335 ± 0.0309
		σ	0.4733 ± 0.0147

Table 8: Parameters for the log-normal fits to the PYTHIA TuneA multiplicity *vs* p_T(particle) for pions (left), kaons (middle), and protons(right) around light bottom mesons.

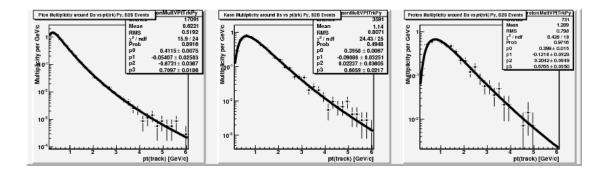| Particle	χ^2/ndf	Parameter	Value
$\pi^{\pm}$	15.9 / 24	l	-0.05407 ± 0.0075
		$\log(s)$	-0.6731 ± 0.0367
		σ	0.7097 ± 0.0186
$K^{\pm}$	24.43 / 25	l	-0.09086 ± 0.03251
		$\log(s)$	0.02237 ± 0.03605
		σ	0.6059 ± 0.0217
$p/\bar{p}$	8.428 / 18	l	-0.1218 ± 0.0923
		$\log(s)$	0.2042 ± 0.0849
		σ	0.5765 ± 0.0550

Table 9: Parameters for the log-normal fits to the PYTHIA TuneA multiplicity *vs* p_T(particle) for pions (left), kaons (middle), and protons(right) around B_s mesons.

event tuning. The population size was set to $N_p = 52$; the control parameters were set using the two-phase adaptive algorithm of Section 4.5 with $\zeta = 0.2$. The exploitation phase was initiated after 5 generations. The weights were set to $w_k = \{0.2, 1.0, 0.2\}$ for $k = \pi, K, p$; this set of weights made it possible to better match the kaon distributions. The proton multiplicity around bottom mesons is small compared with the pion multiplicity, so the proton component of the objective function was nearly negligible. Convergence was obtained after 41 generations; the convergence history is shown in Fig. 24
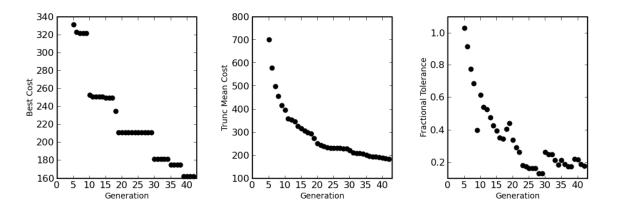

Figure 24: Convergence history for the 'b-jet' particle content tuning.

The tuned parameters are shown in Table 10 and are the average values for the ten best individuals in the population of the 41$^{\text{st}}$ generation. The average parameter values were used rather than the parameters from the best individual because they were found to give better performance in higher statistics evaluations. Converting `CLMAX` and `CLPOW` into arrays of length two can now be seen to be justified, since the tuned values of these parameters differ for bottom flavored clusters and non-bottom flavored clusters.

The comparisons of the tuned HERWIG and PYTHIA TuneA multiplicities are shown in Figures 25, 26. The pion and kaon multiplicity distributions are improved over the defaults (Figures 18 and 19), though the proton multiplicity is still too high. The comparisons of the p_T spectra can be seen in Figures 27, and 28. Clearly the excess at low-p_T across all particle species has been much reduced.

The underlying event performance plots after 'b-jet' particle content tuning are presented in Figures 29 and 30. The multiplicity *vs* p_T(Lead jet) distributions in Fig. 29 are now closer to the reference distributions, though with higher variance compared to Fig. 16. The shape of the multiplicity *vs* p_T(particle) distributions exhibit improved agreement in shape compared with Fig. 17, though there are still too few charged particles with p_T above 1.0 GeV/c.

Parameter	Default Value	Tuned Value
m_d	0.32 GeV/c^2	0.3873 GeV/c^2
m_s	0.5 GeV/c^2	0.4651 GeV/c^2
m_b	4.95 GeV/c^2	5.168 GeV/c^2
m_g	0.75 GeV/c^2	0.9010 GeV/c^2
CLMAX(1)	3.35 GeV/c^2	3.562 GeV/c^2
CLMAX(2)	3.35 GeV/c^2	4.856 GeV/c^2
CLPOW(1)	2.0	1.389
CLPOW(2)	2.0	0.9977
PSPLT(1)	1.0	0.2497
PSPLT(2)	1.0	0.8427
CLSMR(1)	0.0	0.1795
CLSMR(2)	0.0	0.09591
PWT(3)	1.0	1.969

Table 10: Tuned fragmentation parameters

Figure 25: BJet Tune results: track multiplicity around B_d mesons.

Figure 26: BJet Tune results: track multiplicity around B_s mesons.

Figure 27: BJet Tune results: track multiplicity around B_d mesons *vs* p_T(track)
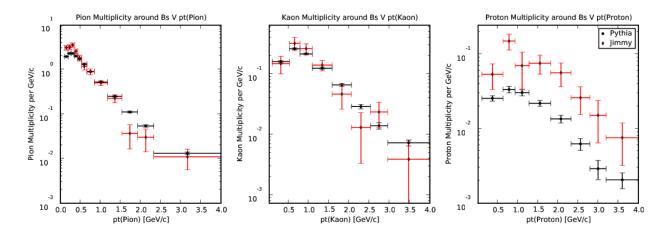

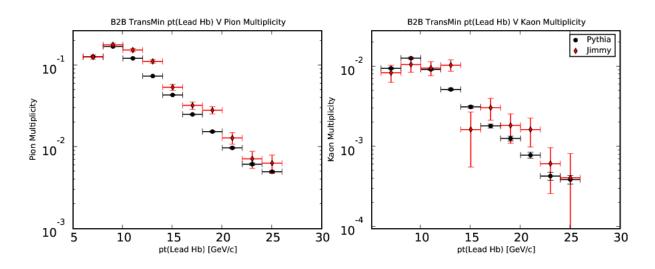Figure 28: BJet Tune results: track multiplicity around B_s mesons *vs* p_T(track)

Figure 29: UE tune results (after BJetTune): charged pion (left) and charged kaon (right) multiplicity *vs* p_T(Lead jet).
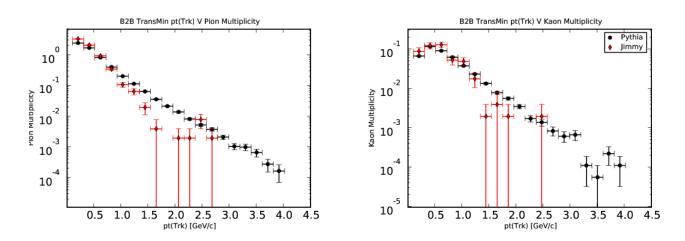

Figure 30: UE tune results (after BJetTune): charged pion multiplicity *vs* $p_T(\pi^\pm)$ (left) and charged kaon multiplicity *vs* $p_T(K^\pm)$ (right).

6   Soft Muon Tagger

It is desireable to reduce the fake muon rate as much as possible. Toward this end, an algorithm developed for the CDF B_s mixing analysis is employed: the Likelihood Soft Muon Tagger [48]. This algorithm uses the muon track-to-stub matching variables, ΔX, $\Delta \Phi$, and ΔZ, as well as the energy deposited in the calorimeters to construct a likelihood function which estimates the probability that the muon is real. The likelihood is constructed as

$$\mathcal{L} = \frac{S}{S + B}, \tag{18}$$

with S as the estimated probability density function (PDF) that the muon is real, and B as the PDF that the muon is fake. The S and B functions are constructed from the PDFs of the individual variables:

$$S = \prod_i S_i \tag{19}$$

$$B = \prod_i B_i, \tag{20}$$

with $i \in \{\Delta x, \Delta \phi, \Delta z, E_{EM}, E_{HAD}\}$. A separate likelihood is constructed for each muon type (CMU, CMUP, CMP, and CMX). The lowercase analogues of the track-to-stub matching variables correspond to p_T-scaled variables and are constructed as $\Delta x = \Delta X/\sigma_{\Delta X}$, $\Delta \Phi = \Delta \Phi/\sigma_{\Delta \Phi}$, and $\Delta z = \Delta Z/\sigma_{\Delta Z}$. This p_T scaling is necessary to account for the p_T dependent nature of the scattering that muons undergo as they traverse the absorber material. Higher-momentum muons have more narrow track-to-stub matching distributions, due to less multiple scattering compared with lower-momentum muons. The p_T scaling functions are:

$$\sigma_{\Delta X}^{CMU}(p_T) = A^{CMU}(p_T) + d \tag{21}$$

$$\sigma_{\Delta \Phi, \Delta Z}^{CMU}(p_T) = C_{\Delta \Phi, \Delta Z}^{CMU}(p_T) \tag{22}$$

$$\sigma_{\Delta X}^{CMP}(p_T, \phi) = A^{CMP}(p_T) + B_{\Delta X}^{CMP}(p_T) \cdot (\sin(4\phi + \frac{\pi}{2}) - 1) \tag{23}$$

$$\sigma_{\Delta \Phi, \Delta Z}^{CMP}(p_T) = C_{\Delta \Phi, \Delta Z}^{CMP}(p_T) \tag{24}$$

$$\sigma_l^{CMX}(p_T) = B_l^{CMX} \cdot (|\eta| - 0.675) + C_l^{CMX}, \tag{25}$$

where

$$A^j(p_T) \equiv \frac{a_j + e^{b_j + c_j p_T}}{p_T} \tag{26}$$

$$B_l^j(p_T) \equiv \min(e_j^l + f_j^l p_T, 0) \tag{27}$$

$$C_l^j(p_T) \equiv g_j^l + \frac{h_j^l}{p_T} + \frac{k_j^l}{p_T^2}, \tag{28}$$

for $j \in \{CMU, CMP, CMX\}$, and $l \in \{\Delta X, \Delta \Phi, \Delta Z\}$. The CMP ΔX scaling function depends on the azimuthal angle ϕ due to the box architecture of the CMP detector.

The PDFs for the EM calorimeter were found to have negligible p_T dependence. However, they do depend slightly on the momentum isolation variable I, defined as,

$$I = \frac{p_T}{\sum_i p_{Ti}}. \tag{29}$$

The sum is taken over all tracks in an η-ϕ cone of size $R = 0.4$ around the muon. An isolated muon is defined to be one for which $I > 0.5$, while non-isolated muons fall into the complementary requirement. Separate E_{EM} PDFs are used for these two isolation regions.

The hadronic calorimeter response was found to be weakly dependent upon the p_T of the muon. Separate PDFs are used for the following domains: $p_T < 2$ GeV/c, $2 < p_T < 3$ GeV/c, and $p_T > 3$ GeV/c for CMU muons, $p_T < 3$ GeV/c, $3 < p_T < 4$ GeV/c, and $p_T > 4$ GeV/c for CMU muons, and $p_T < 3$ GeV/c, $3 < p_T < 5$ GeV/c, and $p_T > 5$ GeV/c for CMU muons. The dependence of the E_{HAD} PDFs on the muon momentum isolation is negligible.

Muons from J/ψ decays are used to construct the real muon template distributions; protons from Λ decay whose tracks are matched to stubs in the muon detectors are used to construct the fake muon templates. The parameters of the scaling functions were determined by fitting the PDF functional forms to the data distributions.

Since J/ψ decays tend to produce isolated muons, the templates used for real muons favors isolated muons. Requiring a minimum-ionizing particle signature in the calorimeters contributes to this bias. This is not a problem when reconstructing exclusive decay modes, but it presents a challenge in a jet analysis. This preference for isolated muons reduces the event selection efficiency somewhat, but the requirements on $\mathcal{L}$ were set to minimize this effect.

7 Jet Clustering

QCD predicts that high-p_T partons will radiate soft and collinear partons before they hadronize. The particles which comprise a jet will interact with some components of a particle detector. Nearby detector elements should be associated with the same jet. However, there is some ambiguity in the definition of 'nearby'; a jet clustering algorithm such as the k_T algorithm fully specifies this definition.

7.1 k_T Algorithm

The k_T jet clustering algorithm is used to group tracks into jets. Jet clustering in a hadron collider environment is somewhat more complex than in a lepton collider environment owing to the fact that the beam particles are composite. The beam remnants are assumed to travel in nearly the same direction as the original beam (*i.e.* either parallel or anti-parallel to the z-axis). The distance metric d_{iB} used to compare a given particle to the beam is simply the squared transverse momentum k_{Ti}^2 of the particle. Since the beam remnants have low transverse momentum (and large $|\eta|$), this choice of d_{iB} separates them from the remainder of the particles in the event (the interesting bits).

The metric of similarity between two particles i and j takes into account their transverse momenta as well as their angular separation in η-ϕ space, compared to a user-defined jet size parameter R_{jet}. This similarity metric $d_{i,j}$ is defined by:

$$d_{i,j} \equiv \min(k_{Ti}^2, k_{Tj}^2)\Delta R_{i,j}^2/R_{\text{jet}}^2, \tag{30}$$

where $\Delta R_{i,j} \equiv \sqrt{(\phi_i - \phi_j)^2 + (y_i - y_j)^2}$, y_i is the the rapidity for particle i, ϕ_i is it's azimuth, and k_{Ti} is it's transverse momentum. Similar particles are recombined into jets in an agglomerative, heirarchical fashion (see Section 7.2), starting with the two particles possessing the minimal $d_{i,j}$. If the beam is the most similar particle to particle i (that is, d_{iB} is smaller than all the $d_{i,j}$), this particle is declared a 'final jet' and no longer participates in clustering. The algorithm then stops when it runs out of particles to recombine.

This description is based on the FastJet[18] implementation, whose algorithm proceeds as follows:

1. For each pair of particles i and j, determine their similarity $d_{i,j}$, given in Eq. 30.

2. Compute each particle's similarity to the beam, $d_{iB} = k_{Ti}^2$.

3. Determine the minimum similarity d_{min} over all the $d_{i,j}$ and d_{iB}. If d_{min} corresponds to a $d_{i,j}$, merge particles i and j according to the chosen recombination scheme. If

d_{min} corresponds to a d_{iB}, particle i is declared a 'final jet' and removed from further consideration.

4. Go back to step 1 and repeat until all particles have been labeled 'final jets'.

7.2 Recombination Scheme

The recombination scheme chosen is the p_T scheme[13]. In this scheme, the four-momenta of particles i and j are recombined into the four-momentum p_r according to

$$
\begin{aligned}
p_{Tr} &= p_{Ti} + p_{Tj}, & (31)\\
\phi_r &= (p_{Ti}\phi_i + p_{Tj}\phi_j)/(p_{Ti} + p_{Tj}), & (32)\\
y_r &= (p_{Ti}y_i + p_{Tj}y_j)/(p_{Ti} + p_{Tj}). & (33)
\end{aligned}
$$

The FastJet implementation of this recombination scheme sets the energy component of the recombined jet E_r to zero. This has the effect of making the recombined particle's mass equal to the magnitude of the jet 3-momentum.

In this thesis, an alternative jet mass estimate is used: the mass of the charged pion is assumed for every track in the jet, yielding an energy estimate for each track; the jet energy is then estimated as the sum of the track energies. The estimated jet mass is found in the usual way from $m_{\mathrm{jet}}^2 = E_{\mathrm{jet}}^2 - |\boldsymbol{p}_{\mathrm{jet}}^2|$.

---

[13]In the FastJet package, this is the 'boost-invariant' p_T scheme.

8 Data Sample

8.1 Dataset

From the Υ dataset, we select the dimuon trigger paths `BBBAR_CMUP3_CMU1.5` and `BBBAR_CMUP3_CMX2`. These trigger paths select both Υ mesons as well as $b\bar{b}$ events in which both H_b's decay semi-muonically. There is an implicit $\Delta\phi(\mu,\mu)$ cut in these trigger paths due to the $m(\mu,\mu) > 5$ GeV/c^2requirement, as shown in Fig. 31. The sensitivity to gluon splitting and flavor excitation mechanisms is thus severely limited in this region.

Figure 31: $\Delta\phi(\mu_1,\mu_2)$ *vs* $m(\mu_1,\mu_2)$ for MC@NLO semi-muonic $b\bar{b}$ events which pass the trigger and good muon selection of Section 8.3 (no jet criteria are applied). The majority of events with $\Delta\phi(\mu_1,\mu_2) < 0.7$ fail the mass requirement imposed by the trigger ($m(\mu_1,\mu_2) > 5$ GeV/c^2).

8.2 Event Selection Overview

Event selection proceeds as follows. A good run list is applied which ensures that the relevant detector components were functioning properly at the time the data were recorded. For each event, the starting point is the 'good' muon selection, in which a subset of muons is extracted from the full collection of muons in the current event, based upon kinematic and reconstruction quality variables (Section 8.3). The primary vertex selection is next (Section 8.4. The three-level trigger requirements for the `BBBAR_CMUP3_CMU1.5` and `BBBAR_CMUP3_CMX2` (Section 2.2.7) paths are emulated in the next step. At this point, the raw track impact parameters can be corrected for the primary vertex position; requirements are made on the

muon transverse impact parameter in order to reject decay-in-flight muons. The jet clustering algorithm (Section 8.5) is applied to the subset of tracks which are associated with muon primary vertex; selected events have at least two jets which contain a muon or one jet which contains multiple muons (Section 8.6). The final step in the event selection puts requirements on the μ-jet flavor classifier responses (Section 10). The resulting events are quite highly enriched in bottom quark pairs.

8.3   Good Muon Selection

The criteria for the 'good' muons involve hits in the tracking detectors, p_T, impact parameter, and the response of the Likelihood Soft Muon Tagger, as outlined in Table 11. The silicon hit requirements (one L00 hit, one L0 hit, and one additional r-ϕ hit) are made in order to increase the transverse impact parameter resolution. These cuts also help to reject muon candidates which are daughters of long-lived hadrons, which have a high probability of decaying outside the beam pipe. These are the same hit requirements made in the CDF correlated bottom quark cross section measurement [46]. The COT hit requirements ensure the quality of the overall track fit.

Quantity	Criterion
L00 Hits	1
SVX-II L0 Hits	1
Add'l SVX-II r-ϕ Hits	1
COT Axial Hits	10
COT Stereo Hits	10
$p_T(\mu)$	$(3.0 \leq p_T < 30.0)$ GeV/c
$d_0(\mu)$	≤ 0.2 cm
$z_0(\mu)$	≤ 5.0 cm
$\mathcal{L}(CMU)$	> 0.75
$\mathcal{L}(CMP)$	> 0.50
$\mathcal{L}(CMUP)$	> 0.75
$\mathcal{L}(CMX)$	> 0.70
ΔZ	< 80.0 cm

Table 11: 'Good' Muon Criteria

The minimum $p_T(\mu)$ criterion removes many fakes (tracks spuriously matched to stubs in the muon chambers). The maximum $p_T(\mu)$ requirement is set to reject muons from $W^{\pm}$

decay.

The Likelihood Soft Muon Tagger (LSMT [48]) is used to select real muons and suppress fake muons. Decay-in-flight muons which pass the track quality criteria will appear as real muons. The LSMT operating points were chosen to balance the efficiency for real muons (true positive rate) against the efficiency for accepting fake muons (false positive rate). The characteristics of the chosen operating points were taken from Ref. [48] and are displayed in Table 12. The CMX cut was reduced from 0.75 to 0.70 in order to raise the true positive rate to be more similar to that of the CMUP. Only the 0.75 CMX likelihood threshold is shown in Table 12 because this operating point's true positive rate (TPR) and false positive rate (FPR) were measured in Ref. [48]. The definitions of TPR and FPR are given in Section 9.

Muon Type	Threshold	TPR [%]	FPR [%]
CMU	0.75	84.4	8.4
CMP	0.50	88.2	27.1
CMUP	0.75	85.9	27.5
CMX	0.75	82.5	11.6

Table 12: Likelihood Soft Muon Tagger true positive rate (TPR) and false positive rate (FPR) for the chosen operating points, subdivided by muon type (from [48]).

The ΔZ requirement removes some small, unphysical humps at large values of the longitudinal track-to-stub matching variable.

After the primary vertex selection has been made, the raw impact parameters of the good muons are corrected for the position of the primary vertex. The requirements made on $d_0(\mu)$ and $z_0(\mu)$ listed in Table 11 were designed to reject muons from other $p\bar{p}$ interactions ('pile-up'), as well as muons from long-lived strange particles, such as $K^{\pm}$.

The good muon collection is then examined for pairs of muons whose invariant mass $m(\mu, \mu)$ is consistent with the dimuon resonances listed in table 13. Any pair of muons found to have an invariant mass within the mass range of one of these resonances is rejected.

Note that no requirement was made on the relative signs of the muon pairs. This approach was taken in order to allow muons from charm hadron decay ($b \rightarrow Xc$ followed by $c \rightarrow \mu\nu X$) to be considered as signal when the charm hadron is itself the decay product of a bottom hadron. Making no requirement on the relative dimuon signs also allows neutral bottom meson mixing to occur naturally in the sample.

State(s)	Mass Minimum [GeV/c^2]	Mass Maximum [GeV/c^2]
ρ, ω	0.70	0.85
ϕ	0.99	1.05
$J\psi$	2.95	3.2
$\psi(2s)$	3.5	3.8
$\Upsilon(1s)$	9.20	9.65
$\Upsilon(2s)$, $\Upsilon(3s)$	9.85	10.55
Z^0	88.0	94.0

Table 13: Dimuon Resonance States

8.4   Primary Vertex Selection

The primary vertex is fit individually for each event using the standard CDF algorithm.

There can be as many as a dozen primary vertices reconstructed in a given data event, spread over an interaction region about 120 cm long in z. The mean number of primary vertices is about 2.5. In order to choose the primary vertex, a mapping is created between the tracks and the primary vertices. For the first two vertices in each event, the tracks which participated in the fit are identified. For other vertices or tracks which did not participate in the fits of the first two vertices, another algorithm must be used to create the track-to-vertex map.

In this alternative algorithm, each track is associated with the closest vertex in the z coordinate (i.e. the minimum $(z_0^{raw}(track) - z_{PV})$). The chosen vertex is required to have two associated muons whose separation in z $(z_0^{raw}(\mu_1) - z_0^{raw}(\mu_2))$ is less than 5.0 cm.

This method of selecting the primary vertex allows 'pile-up' rejection naturally. Pile-up is the term given to the phenomenon of multiple, separate $p\bar{p}$ collisions in a single bunch crossing. By mapping each track to a vertex, the collection of tracks may be restricted to those associated with the multi-muon primary vertex.

It is possible for tracks to be assigned to the wrong primary vertex. However in this analysis, the only effects of concern are those which may skew the b-jet reconstruction. Additional interactions will almost always be minimum bias and distributed quasi-normally along the z-axis with a width of 30 cm. This situation makes it rare for the b-jet reconstruction to be affected, and so the effect is neglected.

8.5 Track-Jet Clustering

The track collection input to the jet clustering algorithm is a subset of the full collection available in each event. The first selection criterion is the pile-up rejection: only tracks associated with the chosen primary vertex are allowed to participate in jet clustering. This association is determined either by the flag indicating that the track participated in the chosen primary vertex, or by minimum distance in the z coordinate.

Similar to the muon selection, tracks are required to be within 5 cm in z of the primary vertex. The transverse momentum of each track is required to be greater than 0.4 GeV/c. At least three r-ϕ silicon hits per track are required, in order to minimize the impact parameter uncertainty. No requirement is made on the number of COT hits per track in order to increase the acceptance for tracks with $1.0 < |\eta| < 2.0$. This ensures that muon jets whose muon is near 1.0 in $|\eta|$ are not artificially starved of tracks.

The proportion of μ-jet tracks with zero COT hits, called 'silicon stand-alone tracks', is actually quite small, as can be seen from Fig. 32 for CDF dimuon data. While some of these tracks may be fake tracks, they do not affect the jet clustering adversely. Their characteristics are shown in Fig. 33. The mean number of silicon hits for these tracks is an appreciable fraction of the maximum, indicating that the majority of tracks in this category simply failed to link to a COT-only track due to reduced coverage for $1.0 < |\eta| < 2.0$. The p_T of these tracks is relatively low, and accounts for only a small fraction of their μ-jet transverse momenta. Therefore, silicon stand-alone tracks were not a significant source of noise in jet-clustering.

For tracks, there is no individual energy measurement from the calorimeter. However, the energy of each track can be determined from it's mass. In order to avoid systematic errors associated with particle identification, the $\pi^{\pm}$ mass is assumed for all tracks uniformly. Depending on the actual particle mass, this assumption will have a small effect (a few percent for protons) on the rapidity calculation.

The k_T jet clustering algorithm parameter which determines the maximum size of jets R_{jet} is set at 0.7. The inclusive jet reconstruction mode is used, with the minimum jet transverse momentum is set at 1.0 GeV/c. Because the four-momenta of the input particles are estimated by their tracks, the resulting jets are called 'track-jets'.

8.6 Muon-Jets

A muon-jet (or μ-jet) is defined as a track-jet which contains at least one good muon, based on track ID code matching. For this analysis, each μ-jet is required to have at least one additional track, besides the muon.
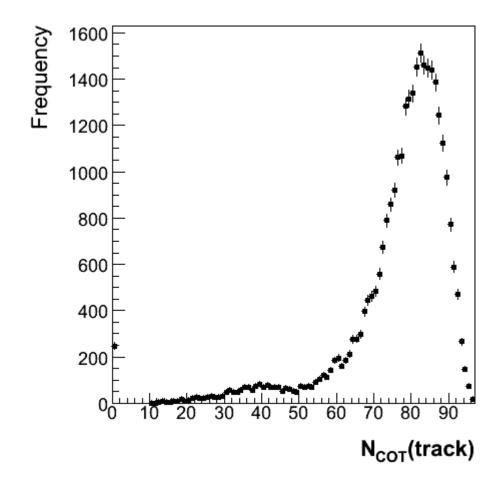
Figure 32: Distribution of the number of COT hits for tracks within μ-jets from CDF data.
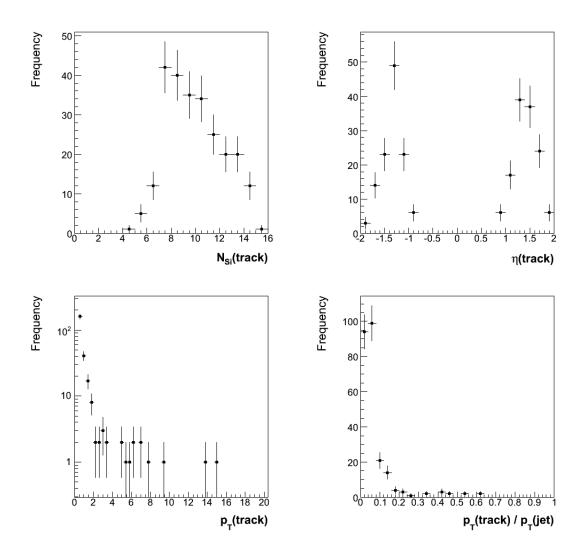
Figure 33: Silicon stand-alone track characteristics in CDF data.

9 Classifiers

9.1 Binary Classifier Performance Metrics

Binary classifiers attempt to separate signal events ('positives') from background events ('negatives'). For many algorithms, the output (response) is continuous, rather than discrete. For these algorithms, all events whose response exceeds a given threshold are labeled positives, all others being declared negatives. Outside the training data, there will be some leakage across the threshold between the two classes. The events correctly predicted to be positives are called 'true positives', while those events correctly predicted to be negatives are called 'true negatives'. Those events which are misclassified as positive are called 'false positives'. The term 'false negatives' is given to those events which are misclassified as negatives. We can thus define a 'confusion matrix' for the number of events in each of these four categories, illustrated in Table 14. The 'True Positive Rate' (TPR) and the 'False Positive Rate' (FPR) can then be defined:

$$TPR \equiv \frac{TP}{P}, \tag{34}$$

$$FPR \equiv \frac{FP}{N}, \tag{35}$$

where $P \equiv TP + FN$ and $N \equiv TN + FP$ are the actual number of positive and negative events, respectively. The predicted number of positives is given by $\widehat{P} = TP + FP$, while the predicted number of negatives is given by $\widehat{N} = TN + FN$. The sum of P and N is equivalent to the total number of events M, as is the sum of $\widehat{P}$ and $\widehat{N}$.

A perfect classifier would have $TPR = 1$ and $FPR = 0$. By varying the threshold separating positives from negatives, a curve can be generated which illustrates the trade-offs between TPR and FPR. A common representation of this trade-off curve is called the *Receiver Operating Characteristic* (ROC) curve [56], in which TPR is plotted against FPR for a variety of threshold values.

9.2 False Discovery Rate

For a given data sample of unknown composition, it is desireable to quantify the background contamination the predicted signal region. The False Discovery Rate [57] (FDR) is the fraction of predicted positive events which were misclassified:

$$FDR \equiv \frac{FP}{FP + TP}. \tag{36}$$

		Total
TP	FP	$\widehat{P}$
FN	TN	$\widehat{N}$
Total P	N	M

Table 14: A confusion matrix for a binary classifier shows the relationship between the number of True Positives (TP), the number of False Positives (FP), the number of False Negatives (FN), and the number of True Negatives (TN). The total number of actual (P, N) and predicted ($\widehat{P}$, $\widehat{N}$) events of each type are also indicated. The total number of events is M.

It is possible to estimate FDR given estimates of TPR, and FPR, $\widehat{P}$, and M. First, rewrite TP in terms of TPR, $\widehat{N}$, and TN:

$$TP = \frac{TPR(\widehat{N} - TN)}{1 - TPR}. \tag{37}$$

Next, rewite FPR as a function of TP:

$$1 - FPR = \frac{TN}{TN + \widehat{P} - TP}. \tag{38}$$

Substituting Eq. 37 and solving for TN gives

$$TN = \left(\frac{1 - FPR}{TPR - FPR}\right)(TPR \cdot M - \widehat{P}) \tag{39}$$

Substituting Eq. 39 back into Eq. 37 and using the fact that $M = \widehat{P} + \widehat{N}$ gives

$$TP = \frac{TPR}{TPR - FPR}\left(\widehat{P} - FPR \cdot M\right). \tag{40}$$

Using the definition of $\widehat{P}$ to rewriting Eq. 36 as

$$FDR = 1 - \frac{TP}{\widehat{P}}, \tag{41}$$

FDR is completely defined by TPR, FPR, $\widehat{P}$, and M. It is thus not necessary to know N or P *a priori* to determine the false discovery rate.

9.3   Decision Trees

A decision tree recursively splits the data into rectangular regions (*nodes*) defined by a collection of simple binary splits (*i.e.* 'cuts'). The exact algorithm used to grow the tree

depends upon the software package used; the description following is appropriate for the StatPatternRecognition package (SPR [51]). At each training iteration and for each node, all possible binary splits in each dimension are examined in order to maximize a figure of merit (FOM) which quantifies the node impurity. Given a parent node of weight W, the first child node is given a weight W_1, and the second is given weight $W_2 = W - W_1$. Each daughter node is also given a temporary class assignment ("signal" for one, "background" for the other); this class assignment can be swapped if the FOM is better. While there are several figures of merit Q available, this analysis uses the negative cross entropy, defined as

$$Q(p, q) = p \log p + q \log q, \tag{42}$$

where p is the fraction of correctly classified events, and $q = 1 - p$ is fraction of misclassified events in the node. Denoting the figures of merit for the daughter nodes as Q_1 and Q_2, the FOM for a split is then given by

$$Q_{\text{split}} = \frac{W_1 Q_1 + W_2 Q_2}{W}. \tag{43}$$

There is only one stopping criterion implemented in SPR: minimal node size. If no split can be found which reduces the node's initial FOM or the node size is equivalent to the user-supplied minimum, the node is declare a 'terminal' or 'leaf' node. Each terminal node is labeled 'signal' (1) or 'background' (0) based upon the weighted majority vote of the events in the node. Training stops when every branch of the tree ends in a leaf node.

According to the SPR package documentation, the minimal leaf size used during training affects the classifier performance to a moderate degree. A model selection tool such as cross validation (Section 9.6) can be used to guide the choice of minimal leaf size.

The training of decision trees is inherently unstable, due to the fact that they are comprised of binary splits. The early splits can affect the remainder quite dramatically. In fact, the error rate for decision trees only a little better than 50%. Such classifiers are termed 'weak'. It turns out that these disadvantages can be overcome spectacularly by employing an ensemble-building algorithm, such as 'bagging' or 'boosting'.

9.4   Bagging

Bagging [53], short for 'boostrap aggregating', is one method of constructing an ensemble of weak classifiers. The bootstrap method [52] estimates the uncertainty of a parameter or model using the training data directly[14]. This is accomplished via resampling the training

---

[14]The method described here is the nonparametric bootstrap method.

data *with replacement*. Let $\mathbf{Z} = \{z_1, z_2, \ldots, z_N\}$ represent the training data. A set of B bootstrap datasets $\{\mathbf{Z}_1^*, \mathbf{Z}_1^*, \ldots, \mathbf{Z}_B^*\}$ are created by resampling with replacement from $\mathbf{Z}$. The bootstrap datasets can then be used to estimate the variability of the original dataset.

Bagging then creates an ensemble of classifiers, each trained on a bootstrap dataset. A prediction for a new data point is then taken to be the mean response across the B classifiers. The response for a bagged classifier is then continuous on the domain $[0, 1]$.

The author of the SPR package found that, in the context of bagged decision trees, the performance of an ensemble classifier had only a weak dependence on the number B of ensembles for B between 100 and 300 (more training cycles than this became computationally prohibitive).

9.5 AdaBoost

The other classifier ensemble creation algorithm mentioned is boosting. The basic idea of boosting is to enhance or 'boost' the weight of misclassified events in the training data. Subsequent iterations of the algorithm then work harder to correctly classify the events with higher weight. One such boosting algorithm is AdaBoost [54]. The algorithm[15] proceeds as follows:

1. For i in 1 to N, set the weight $w_i^{(0)}$ of the ith training data point to $1/N$.

2. For k in 1 to M:

    (a) Train a classifier $G_m(x)$ to the training data, weighted by the $w_i^{(k)}$.

    (b) Calculate the weighted misclassification error as

    $$\epsilon_m = \frac{\sum_{i=1}^{N} w_i^{(k)} I(y_i \neq G_m(x_i))}{\sum_{i=1}^{N} w_i^{(k)}}.$$

    (c) Compute the classifier weight as

    $$\alpha_m = \log((1 - \epsilon_m)/\epsilon_m).$$

    (d) For i in 1 to N, update the training data weights according to

    $$w_i^{(k+1)} = w_i^{(k)} \exp\left[\alpha_m I(y_i \neq G_m(x_i))\right].$$

---

[15]This algorithm is called 'AdaBoost.M1' by the original authors.

3. Compute the overall response as

$$G(x) = \text{sign}\left[\sum_{m=1}^{M} \alpha_m G_m(x)\right].$$

For consistency with the literature, the AdaBoost algorithm described here codes background events as -1, rather than 0. SPR performs a linear transformation in order to satisfy its internal convention, which uses 0 for background events.

As with bagging in SPR, ensemble classifiers created with AdaBoost only have a weak dependence upon the number of boosting cycles used.

9.6   Cross Validation

Cross validation [55] is a method of model selection; the models under consideration can be distinct, or simply the same parametric model with different parameter values. It is particularly useful for small training datasets. The algorithm proceeds by dividing the data sample into k equally sized, mutually exclusive pieces. The model parameter(s) are fit using an agglomeration of $k-1$ blocks of data, and a figure of merit is evaluated on the kth. This procedure is iterated through all k pieces, leaving each out in turn. The resulting k figures of merit are then averaged. Having done this for each model, the practitioner chooses the model with the optimal mean figure of merit.

For a classifier whose training depends upon a set of parameters (*e.g.* the minimal leaf size of a bagged decision tree), the cross-validated set of parameters are determined by training k separate classifiers on the agglomeration of $k-1$ blocks of data. The figures of merit are determined by applying the trained classifier to the reserved block of data. The practitioner, having chosen a set of trial parameter sets before beginning, calculates the mean classifier figure of merit for each parameter set. As mentioned above, the optimal parameter set is chosen as the one yielding the optimal figure of merit.

10 Jet Classification

10.1 Strategy

The main observables which set bottom quark jets apart from light flavor jets are the large inclusive semi-muonic branching ratio of bottom hadrons, the mass of the initiating parton, and the long lifetime of the bottom hadrons.

The latter phenomenon is manifested by tracks with impact parameters significantly larger than the detector resolution. The large b-quark mass gives muons from bottom hadron decay a large transverse momentum with respect to the jet axis.

As muon objects can also be faked by hadrons from light flavor jets, it is desireable to employ an algorithm to separate bottom jets from light flavor jets. Charm quark jets are also a source of muons; charm hadrons have an inclusive semi-muonic branching ratio similar to that of bottom hadrons.

In this analysis, bottom jets were identified through a heirarchy of jet flavor classifiers. First, bottom jets were separated from light flavor jets. Because charm jets can have similar characteristics to bottom jets, there is a non-zero efficiency for charm jets to be classified as bottom jets by the bottom/light flavor classifier. Thus it is necessary to separate bottom jets from charm jets, given that both types are classified as bottom jets by the bottom/light classifier.

10.2 Common Feature Variables

10.2.1 'Generalized Jet Probability'

In order to extract from the jet the most information about its flavor, it is necessary to encorporate the characteristics of every track in the jet. This presents a problem however, as each jet contains an *a priori* unknown number of tracks. This problem is overcome by estimating the probability that the jet is a bottom jet, given the attributes of its constituent tracks. However, bottom jets are not solely comprised of bottom hadron daughters; they contain fragmentation tracks and tracks from the underlying event, both of which will appear as light flavor tracks. This difficulty is met by employing a classifier to separate bottom daughter tracks from light flavor tracks. The response $G_{\mathrm{trk},i}(\boldsymbol{x}_i)$ of this track classifier will be continuous on $[0, 1]$, with background (*e.g.* light flavor) tracks clustered near zero, and signal (bottom daughter) tracks clustered near one. The response for the ith track can then be interpreted as a probability for this track to be of the signal class,

$$P_i(y_i = \text{signal}|\boldsymbol{X}_i = \boldsymbol{x}_i) = G_{\text{trk},i}(\boldsymbol{x}_i), \tag{44}$$

where y_i is the true class of the track, and $\boldsymbol{x}_i$ is its observed feature vector.

The probability of the jet to be signal can then be formed from track probabilities. First we compute the jet log-likelihood ratio r_{jet} for the two classes, keeping in mind that $P_i(y_i = \text{background}|\boldsymbol{X}_i = \boldsymbol{x}_i) = 1 - P_i(y_i = \text{signal}|\boldsymbol{X}_i = \boldsymbol{x}_i)$:

$$r_{\text{jet}} \equiv \sum_{i \in \text{jet}} \ln \left(\frac{P_i(y_{\text{trk}} = \text{signal}|\boldsymbol{X}_i = \boldsymbol{x}_i)}{1 - P_i(y_{\text{trk}} = \text{signal}|\boldsymbol{X}_i = \boldsymbol{x}_i)} \right), \tag{45}$$

The logit transform is applied to r_{jet} in order to map the domain onto $[0, 1]$:

$$P(y_{\text{jet}} = \text{signal}|\boldsymbol{X} = \{\boldsymbol{x}_i\}) \equiv \frac{1}{1 + \exp[-r_{\text{jet}}]} \tag{46}$$

The 'Generalized Jet Probability' is then defined by Eq. 46. This is not a true probability; rather it is a discriminating variable constructed such that signal events (*i.e.* bottom jets) peak at one, and background events (*e.g.* light flavor jets) peak at zero.

10.2.2 Signed Impact Parameter

As in the CDF Jet Probability b-tagging algorithm (distinct from the 'Generalized Jet Probability' defined here; see [49] for the description of the latest version), this analysis adds discriminating information to the impact parameter via its sign. The track impact parameters are signed according to the position of closest approach with respect to the primary vertex, as shown in Fig. 34 for the transverse impact parameter. If the angle between the track's point of closest approach vector and the jet axis is acute (obtuse), the impact parameter is taken to be positive (negative). A similar assignment is made for the longitudinal impact parameter. Prompt tracks will populate the signed impact parameter distribution symmetrically about zero, while tracks from a secondary vertex will preferentially populate the positive region. Tracks from long-lived hadron decay (*e.g.* bottom) fall into the negative region solely through tracking resolution and jet clustering effects. The jet-signed transverse impact parameter d_0 is shown in Fig. 35 for muons from bottom hadron decay, charm hadron decay, and light flavor sources (all from the simulation samples used for jet flavor classifier training, described in Section 10.3).

10.2.3 Relative Transverse Momentum p_T^{rel}

One way to access the mass of the parent hadron for a displaced track is through the relative transverse momentum p_T^{rel}, taken with respect to the jet axis. The distribution of p_T^{rel} for
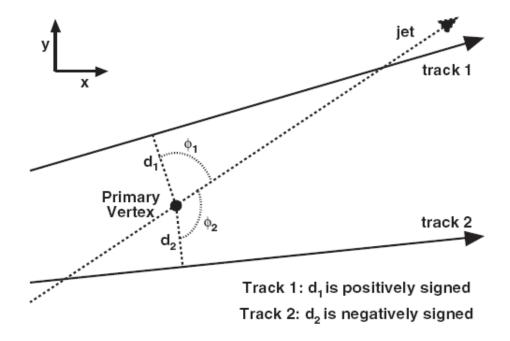
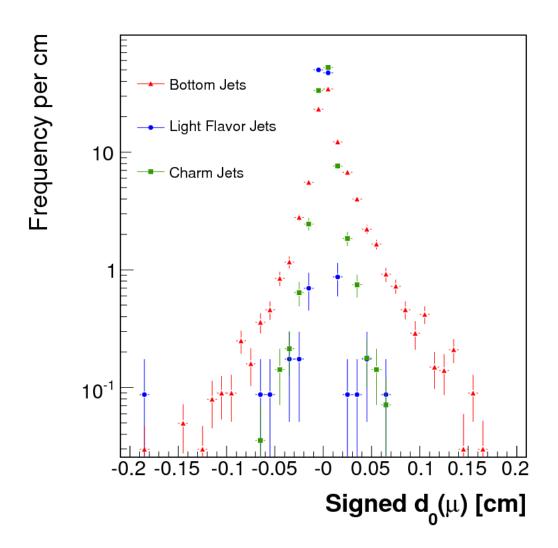Figure 34: Definition of jet-signed d_0, taken from Ref. [49].

Figure 35: Muon jet-signed d_0 from bottom hadron decay (red triangles), charm hadron decay (green squares), and light flavor sources (blue circles; includes decay-in-flight and punch through muons). Histograms are normalized to unit area.

tracks from bottom hadrons will be broader and have a higher mean than those from light flavor sources, as shown in Fig. 36.
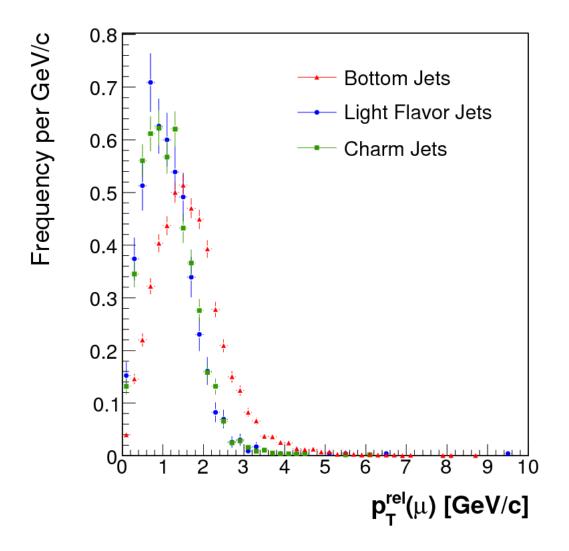

Figure 36: The relative transverse momentum p_T^{rel} for muons from bottom hadrons (red triangles), muons from charm hadrons (green squares), and muons from light flavor sources (blue circles; includes decay-in-flight and punch through muons). Histograms are normalized to unit area.

10.3 Training Samples

In order to train and validate a classifier, the actual class of each point in the training sample must be known. For this reason, simulation data is used in this analysis as the source of the training and validation data.

PYTHIA TuneA [8] is used to generate a sample of light flavor jets which contain either a fake muon or a real muon from a source other than heavy flavor decay. A portion of this sample of μ-jets is used to provide the tracks for training the bottom/light flavor track classifier. PYTHIA TuneA is also used to generate the sample of charm quark jets; only the leading order matrix elements were used, however (*i.e.* charm quarks were generated exclusively, rather than generating the full QCD spectrum and filtering for charm quarks). The bias due to the neglect of the NLO charm jet production mechanisms is assumed to be negligible. The bottom jet training sample was generated with MC@NLO. For both heavy flavor MC samples, the primary heavy hadron was forced to decay semimuonically (*e.g.* $H_b \to \mu\nu X$ with H_b a generic bottom hadron).

All of the training subsamples were processed with the CDF detector simulation, trigger simulation, and reconstruction software. Single-muon event selection requirements were imposed (Table 11), but not the muon pair requirements. This was necessary because the light flavor training sample contains no events with two good muon jets. The selection efficiency for the light flavor MC is roughly 1 in 10^4; this severely limited the light flavor sample size due to computing resource limitations.

For classifier training, the appropriate subsamples were merged into a randomly distributed training samples. All composite samples were comprised of 50% signal and 50% background. Each composite sample was split into a training sample and a validation sample.

10.4   Bottom Daughters *vs* Light Flavor Tracks

For the bottom jet track *vs* light flavor track classifier training, the bottom jet tracks were matched to bottom hadron daughters found in the MC event record. The light flavor tracks were taken from good muon jets found in the light flavor training subsample. The composite training sample consisted of 4622 tracks.

Thirteen feature variables were chosen for this classifier, including information on the number of tracking detector hits, kinematic quantities, and signed impact parameter quantities. The full list can be found in Table 15.

The classifier employed for separation of bottom daughters from light flavor tracks is a bagged, boosted decision tree. For each bagging cycle, an AdaBoost tree was trained using cross entropy as the figure of merit. The AdaBoost tree was set up to train on a bootstrap sample of the training data (appart from the bootstrapping inherent in the bagging algorithm). Then in the bagging phase, the response for each AdaBoost tree was determined for a unique bootstrap sample of the training data. The resulting responses were then averaged to give the overall response. This procedure has the effect of clustering

Track Feature Variable
$N(\text{COT, Axial})$
$N(\text{COT, Stereo})$
$N(\text{Si, } r - \phi)$
$N(\text{Si, } z)$
p_T
p_T^{rel}
$|\eta|$
d_0^{sign}
z_0^{sign}
σ_{d_0}
σ_{z_0}
$d_0^{\text{sign}}/\sigma_{d_0}$
$z_0^{\text{sign}}/\sigma_{z_0}$

Table 15: Bottom daughter *vs* light flavor track classifier feature variables.

the class responses closer to their respective ideals (0 for background, 1 for signal), making it more useful as part of the Generalized Jet Probability than broader response variable distributions.

Due to the limitations of the classification package used, no variable importance estimates are available. Fig. 37 shows the responses for the two classes in validation MC sample of 1156 tracks. The classifier used a minimum leaf size of 50, 50 AdaBoost cycles, and 50 bagging cycles. These parameters were determined by trial and error while watching the file size of the resulting configuration file and subjectively optimizing the shape of the response distributions in validation data. Tighter clustering toward the appropriate ideal value for each class should give better separating power in the Generalized Jet Probability.

While a more rigorous method of model selection might be more satisfying, implementing an algorithm, such as cross validation (Section 9.6), was deemed to be impractical due to the complexity and computational expense of each training pass. The figure of merit guiding model selection for the track classifier must be a measure of the separating power of the Generalized Jet Probability, which was implemented in an analysis package external to SPR. Given that there was no way to preserve the jet structure within the track training sample, a significant amount of logistical support scripting would have been necessary.

The Generalized Jet Probability derived from bottom hadron daughters and light flavor tracks found in the validation MC samples are shown in Fig. 38 (normalized to unit area).

10.5   Bottom Jets *vs* Light Flavor Jets

The bottom jet *vs* light flavor jet classifier was trained on 17 feature variables, including both muon variables and jet variables. The full list of variables given in Table 16. As suggested in Ref. [67], the maximum track transverse and longitudinal impact parameter significances are included. In addition to the Generalized Jet Probability $P_{b,\mathrm{LF}}(y_{\mathrm{jet}} = b | \boldsymbol{X} = \{\boldsymbol{x}_i\})$, the mean $\langle r_{\mathrm{track}}^{b,\mathrm{LF}} \rangle$ and variance $Var(r_{\mathrm{track}}^{b,\mathrm{LF}})$ of the track likelihood ratios for each jet are included.

The jet classifier training subsamples were statistically independent of the track classifier training subsamples. The composite training sample consisted of 1840 muon jets. The validation sample was comprised of 460 muon jets.

A bagged decision tree with 100 cycles comprises the bottom jet *vs* light flavor jet classifier. Cross entropy was used to grow the trees, with the minimal node size set to 10. The relative importance of each variable was estimated according to the change in the split figure of merit it effected. A bar graph showing the variable importance estimated in training data is shown in Fig. 39. As expected, the signed muon transverse impact parameter $d_0^{\mathrm{sign}}(\mu)$ is quite important, along with the muon quality variables $\mathcal{L}(\mu)$, $\Delta\Phi(\mu)$, and $\Delta Z(\mu)$. The
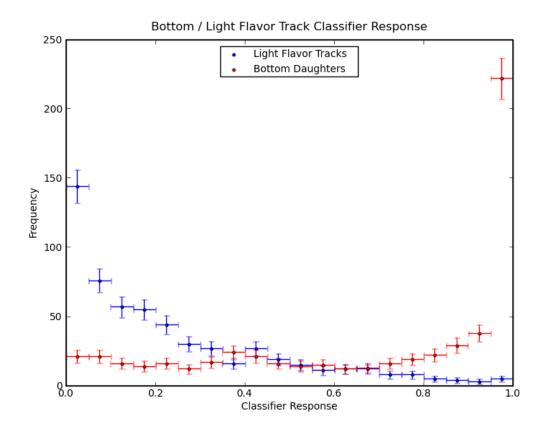
84

Figure 37: The performance of the bottom *vs* light flavor track classifier on validation MC.
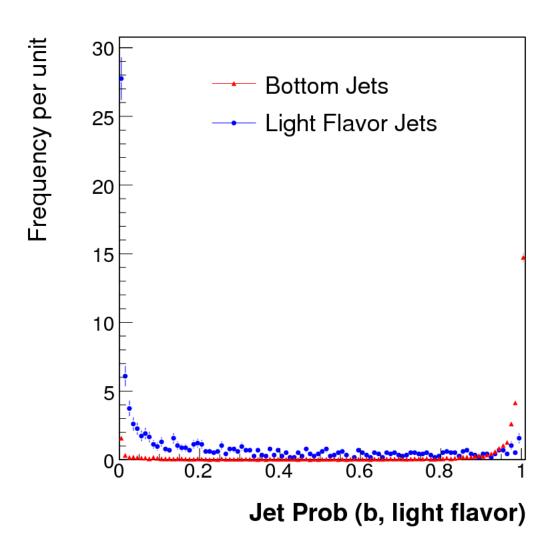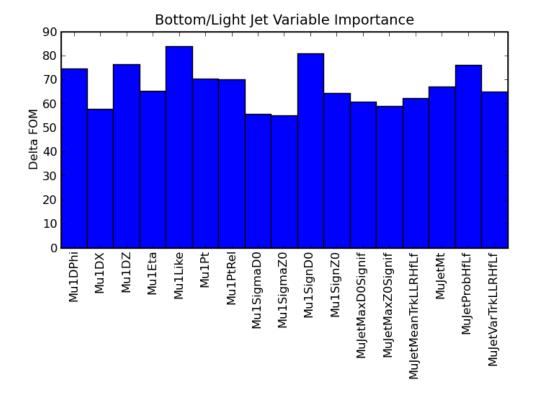
Figure 38: Generalized Jet Probability for bottom jets (red triangles) and light flavor jets (blue circles). Histograms are normalized to unit area.

Jet Feature Variables
$\Delta\Phi(\mu)$
$\Delta X(\mu)$
$\Delta Z(\mu)$
$|\eta(\mu)|$
$\mathcal{L}(\mu)$
$p_T(\mu)$
$p_T^{\text{rel}}(\mu)$
$d_0^{\text{sign}}(\mu)$
$z_0^{\text{sign}}(\mu)$
$\sigma_{d_0}(\mu)$
$\sigma_{z_0}(\mu)$
$\max |d_0(\text{track})|/\sigma_{d_0}$
$\max |z_0(\text{track})|/\sigma_{z_0}$
$m_T(\text{jet})$
$P_{b,\text{LF}}(y_{\text{jet}} = b | \boldsymbol{X} = \{\boldsymbol{x}_i\})$
$\langle r_{\text{track}}^{b,\text{LF}} \rangle$
$Var(r_{\text{track}}^{b,\text{LF}})$

Table 16: Bottom jet *vs* light flavor jet classifier feature variables.

Generalized Jet Probability is also quite important, on par with $\Delta Z(\mu)$.

Figure 39: Variable importance for the bottom *vs* light flavor jet classifier.

The Receiver Operating Characteristic (ROC) curve and classifier responses for validation MC can be found in Fig. 40. Two operating points, labeled A and B, are indicated in the performance plots. Table 17 summarizes the information on these operating points. The FPR values were estimated for fixed TPR values. The errors on the FPR values are statistical only.

The selection of these operating points was guided by a desire to reduce the false positive rate as much as possible, given that the inelastic $p\bar{p}$ cross section exceeds the inclusive b cross section by 3 orders of magnitude.

Label	Cut	TPR [%]	FPR [%]
A	0.91	40	0.0 ± 0.0
B	0.85	45	1.3 ± 0.8

Table 17: Operating points for the bottom jet *vs* light flavor jet classifier.
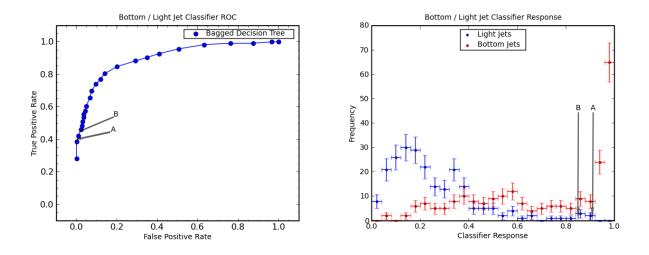
Figure 40: The performance of the bottom *vs* light flavor jet classifier on validation MC.

10.6 Bottom Daughters *vs* Charm Daughters

The bottom daughter *vs* charm daughter training samples were taken from tracks matched to heavy flavor hadrons in MC@NLO and PYTHIA TuneA, respectively. Muon-jets were required to pass the bottom *vs* light flavor jet classifier operating point (either A or B in Table 17) before being allowed to contribute heavy flavor daughters to the training samples for this classifier. The composite training and validation sample sizes were constrained by the charm μ-jet selection efficiency, the total size of the charm sample available, and the necessity of leaving a portion of the sample for jet training. The selection efficiency for charm μ-jets which pass the bottom *vs* light jet response operating point A (B) was about 14% (17%). Working within these constraints for b *vs* light operating point A, a composite sample (b and c) was split into 6186 tracks for training and 1092 tracks for validation. For b *vs* light operating point B, the composite training sample held 8094 tracks, while the validation sample held 1430 tracks.

The 14 feature variables for this classifier are listed in Table 18. Application of this classifier on generic μ-jets highlights the problem that there are actually three classes to separate: light flavor tracks, bottom daughters, and charm daughters. In order to circumvent this problem, the tracks were required to be tagged as bottom daughters (though 'heavy flavor daughters' is perhaps more accurate) by the bottom daughter *vs* light flavor track classifier. The operating point for the bottom daughter *vs* light flavor track classifier was taken to be the naive threshold: $G_{\text{track}}^{b,\text{LF}} > 0.5$. This corresponds to a log likelihood ratio threshold of $r_{\text{track}}^{b,\text{LF}} > 0.0$. The bottom daughter tag was required both for training sample generation and response evaluation. The log likelihood ratio $r_{\text{track}}^{b,\text{LF}}$ was also included in the list of bottom daughter *vs* charm daughter features.

The bottom daughter *vs* charm daughter classifier responses in validation MC can be found in Figs. 41 and 42. The first is for bottom *vs* light flavor jet classifier operating point A, while the second corresponds to operating point B.

10.7 Bottom Jets *vs* Charm Jets

The μ-jet selection for the bottom *vs* charm jet classifier training subsamples proceeded identically to the bottom *vs* charm daughter training subsamples. As with the bottom *vs* light flavor training subsamples, the jet classifier training subsamples were statistically independent of the track classifier training samples. For the b *vs* light flavor operating point A, the composite training sample consisted of 8568 μ-jets, and the validation sample contained 1514 μ-jets. For the b *vs* light flavor operating point A, the composite training sample consisted of 11396 μ-jets, and the validation sample contained 2012 μ-jets.

Track Feature Variable
$r_{\text{track}}^{b,\text{LF}}$
$N(\text{COT, Axial})$
$N(\text{COT, Stereo})$
$N(\text{Si, } r - \phi)$
$N(\text{Si, } z)$
p_T
p_T^{rel}
$|\eta|$
d_0^{sign}
z_0^{sign}
σ_{d_0}
σ_{z_0}
$d_0^{\text{sign}}/\sigma_{d_0}$
$z_0^{\text{sign}}/\sigma_{z_0}$

Table 18: Feature variables employed in the bottom daughter *vs* charm daughter classifier.
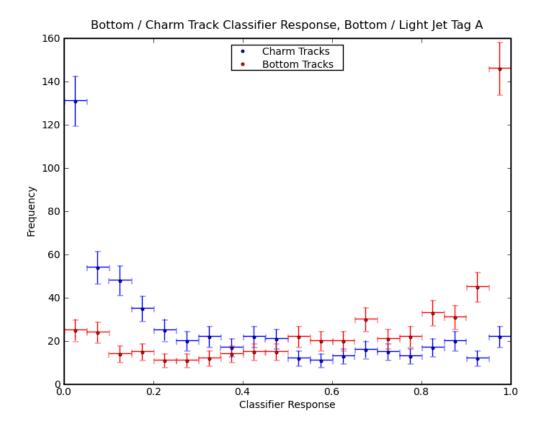
Figure 41: The performance of the bottom *vs* charm track classifier for bottom *vs* light jet operating point A on validation MC.
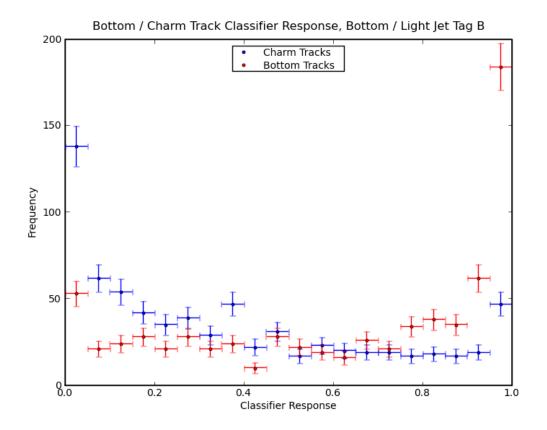
Figure 42: The performance of the bottom *vs* charm track classifier for bottom *vs* light jet operating point *B* on validation MC.
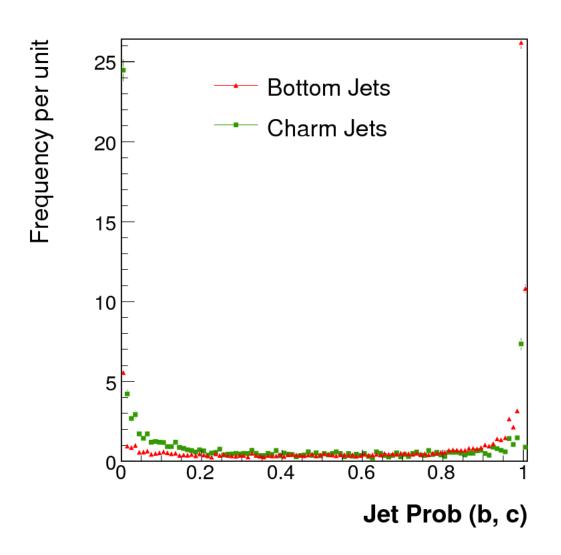
Figure 43: Generalized Jet Probability for bottom jets (red triangles) and charm jets (green squares) for bottom *vs* light jet operating point *A* on validation MC. Histograms are normalized to unit area.

Fifteen feature variables were used in the classifier training, listed in Table 19. The main differences over the bottom *vs* light flavor jet features are the absence of the muon quality variables and the addition of the Generalized Jet Probability related variables formed using the *b vs c* daughter classifier ($P_{b,c}(y_{\text{jet}} = b | \boldsymbol{X} = \{\boldsymbol{x}_i\})$, $\langle r^{b,c}_{\text{track}} \rangle$, and $Var(r^{b,c}_{\text{track}})$).

Jet Feature Variables
$p_T(\mu)$
$p_T^{\text{rel}}(\mu)$
$d_0^{\text{sign}}(\mu)$
$z_0^{\text{sign}}(\mu)$
$\sigma_{d_0}(\mu)$
$\sigma_{z_0}(\mu)$
$\max |d_0(\text{track})|/\sigma_{d_0}$
$\max |z_0(\text{track})|/\sigma_{z_0}$
$m_T(\text{jet})$
$P_{b,\text{LF}}(y_{\text{jet}} = b | \boldsymbol{X} = \{\boldsymbol{x}_i\})$
$\langle r^{b,\text{LF}}_{\text{track}} \rangle$
$Var(r^{b,\text{LF}}_{\text{track}})$
$P_{b,c}(y_{\text{jet}} = b | \boldsymbol{X} = \{\boldsymbol{x}_i\})$
$\langle r^{b,c}_{\text{track}} \rangle$
$Var(r^{b,c}_{\text{track}})$

Table 19: Feature variables employed in the bottom jet *vs* charm jet classifier.

Two bagged decision trees were trained using 100 cycles with the cross entropy as the figure of merit. For *b vs* light flavor operating point A, the minimum node size was set to 10, but it was increased to 15 for operating point B in order to reduce the total number of leaves and to keep the configuration file to a manageable size (the training sample size was larger than for operating point A).

The variable importance bar graphs are shown in Fig. 44 and Fig. 45 for *b vs* light operating points A and B, respectively. For both *b vs* light operating points, the most important variables appear to be $p_T(\mu)$, $p_T^{\text{rel}}(\mu)$, $d_0^{\text{sign}}(\mu)$, and $m_T(\text{jet})$. The Generalized Jet Probabilities and related quantities play a less important role than in the *b vs* light flavor jet classifier. For the variables based on the response of the *b vs c* daughter classifier, the reduced importance could be due to efficiency effects associated with the *b* daughter selection requirement.

The performance plots for the two *b vs c* jet classifiers are in Figs. 46 and 47. Two

operating points are indicated in each figure, giving a set of four possible jet classifier chains: AA, AB, BA, and BB. The true positive rates and false positive rates for AA and AB (BA and BB) are listed in Table 20 (Table 21).

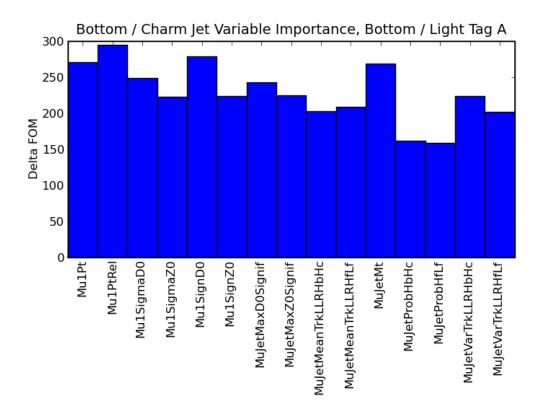Figure 44: Variable importance for the bottom *vs* charm jet classifier for bottom *vs* light flavor operating point A.

Label	Cut	TPR [%]	FPR [%]
AA	0.75	40	5.0 ± 0.8
AB	0.69	50	6.9 ± 1.0

Table 20: Operating points for the bottom jet *vs* charm jet classifier, for bottom *vs* light jet operating point A.
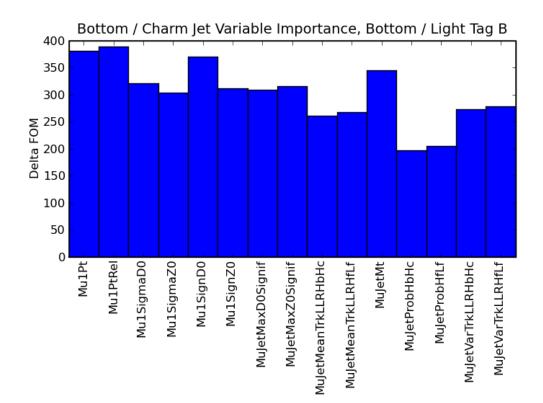
Figure 45: Variable importance for the bottom *vs* charm jet classifier for bottom *vs* light flavor operating point B.

Figure 46: The performance of the bottom *vs* charm jet classifier for bottom *vs* light jet operating point A on validation MC.

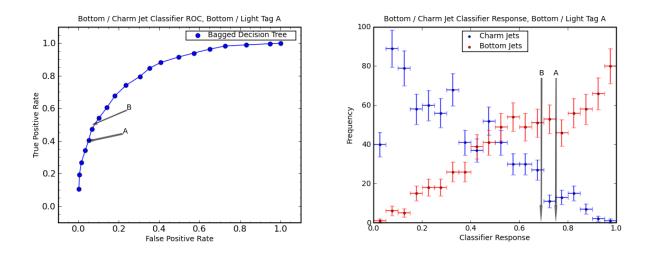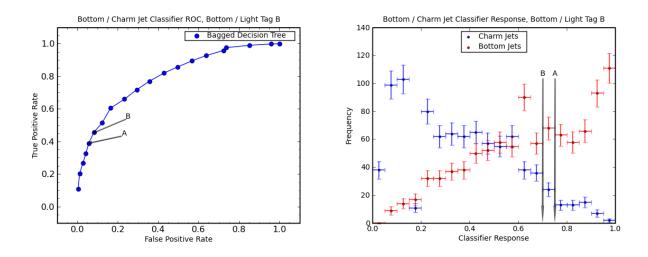Figure 47: The performance of the bottom *vs* charm jet classifier for bottom *vs* light jet operating point B on validation MC.

Label	Cut	TPR [%]	FPR [%]
BA	0.75	39	5.0 ± 0.7
BB	0.70	45	6.9 ± 0.8

Table 21: Operating points for the bottom jet *vs* charm jet classifier, for bottom *vs* light jet operating point B.

10.8   Data Validation

The TPR and FPR values for the classifiers in the previous sections have all been based on simulation. It is desireable to measure these numbers from data in order to gain confidence in the classifier performance metrics.

The true positive rate for both the b vs light flavor jet classifier and the b vs c jet classifier can be measured by leveraging the fact that b quarks are produced in $b\bar{b}$ pairs. Starting with events which contain one identified bottom jet, a sample of unbiased, high purity bottom jets is built by requiring an associated μ-jet (the 'away jet'). The $J/\Psi \to \mu\mu$ dataset is used for this study, so the starting point is three muon events. While this sample of μ-jets has high purity, some background component remains.

In each event, a J/Ψ candidate was reconstructed by requiring that two of the muons fit a vertex and have an invariant mass consistent with the J/Ψ meson. Since J/Ψ mesons can be either created promptly or form part of the decay chain of bottom hadron, a selection must be made to isolate the displaced component (*i.e.* the component from bottom hadron decay). For this purpose the transverse decay length L_{xy} is useful. It is defined as the position of the secondary vertex (SV) with respect to the primary vertex in the transverse plane, projected onto the transverse momentum vector of the J/Ψ:

$$L_{xy} \equiv \frac{X_{xy}^{\vec{SV}} \cdot \vec{p_T}(J/\Psi)}{|\vec{p_T}(J/\Psi)|}. \tag{47}$$

The muons forming the J/Ψ candidate are only required to satisfy the `JPSI_CMUCMU1.5` or `JPSI_CMU1.5_CMX2` trigger requirements, summarized in Tables 5 and 4. In order to select well-measured secondary vertices and ensure a very high purity with respect to the bottom hadron decay component of the sample, the criteria in Table 22 were applied to the J/Ψ meson. The quantity $P(\chi^2_{SV})$ corresponds to the χ^2 probability for the secondary vertex fit. The requirement on $\Delta R(J/\Psi, \mu - \text{jet})$ was made in order to avoid contaminating the away jet with other decay products of the bottom hadron which produced the J/Ψ.

The third muon is required to satisfy the single-muon criteria for this analysis, as summarized in Table 11. The muon jet definition of Section 8.6 is applied.

A complimentary MC sample was generated using MC@NLO. In half of the sample, the decay chains $b \to J/\Psi X$ and $\bar{b} \to \mu X$ were forced. The charge conjugate decays were forced for the remainder of the sample.

The measured and predicted TPR values for the b vs light jet classifier can be found in Table 23. The error estimates are statistical only and derived via the bootstrap method with 5000 bootstrap samples.

Quantity	Criterion
$m(J/\Psi)$	$3.05 < m(J/\Psi) < 3.15 \text{GeV}/c^2$
$P(\chi^2_{SV})$	> 0.001
$\sigma(L_{xy})$	< 0.025 cm
L_{xy}	> 0.1 cm
$\Delta R(J/\Psi, \mu - \text{jet})$	> 0.875

Table 22: J/Ψ meson requirements.

The measured TPR, TPR_{DATA}, disagrees with the complimentary MC estimate, TPR_{MC} by about 16% for operating point A and 13% for operating point B. The MC sample used for this comparison was forced to mimic the selection requirements for the data; specifically, one H_b was forced to decay to $J/\Psi X$, while the other bottom hadron was forced to decay semileptonically. However, the TPR_{DATA} values agree fairly well with those listed in Table 17, the validation MC in which both bottom hadrons were decayed semileptonically.

Label	TPR_{DATA} [%]	TPR_{MC} [%]
A	39.08 ± 0.01	45.15 ± 0.01
B	45.21 ± 0.01	50.90 ± 0.01

Table 23: True Positive Rates (TPR) for the bottom jet vs light flavor jet classifier, as found in $J/\Psi + \mu$-jet data and MC.

The true positive rates for the b vs c jet classifiers are presented in Tables 24 and 25. The disparity between the $J/\Psi + \mu$-jet data and MC is compounded for these classifiers. The source of this discrepancy is most likely a significant violation of the assumption that the μ-jet sample is 100% pure bottom jets. There has been recent theoretical interest in heavy flavor pair production in association with quarkonia [21] of the same flavor (*i.e.* $J/\Psi + c\bar{c}$ or $\Upsilon + b\bar{b}$). Standard QCD processes can also produce heavy flavor pairs through gluon splitting, yielding the final state[16] $\{b\bar{b} + c\bar{c}\}$. Charm jets have a much lower efficiency ($\sim$10% in simulation) for passing the bottom vs light flavor classifier. The presence of charm jets in the $J/\Psi + \mu$-jet sample would thus increase the denomenator in the TPR calculation for data. Multiple heavy flavor pair final states are beyond the scope of this analysis.

Label	TPR_{DATA} [%]	TPR_{MC} [%]
AA	39.68 ± 0.02	53.89 ± 0.01
AB	49.89 ± 0.02	61.87 ± 0.01

Table 24: True Positive Rates (TPR) for the bottom jet vs charm jet classifier, as found in $J/\Psi + \mu$-jet data and MC with bottom vs light flavor operating point A.

The false positive rate for the bottom vs light flavor jet classifier could be measured from data, given a high-purity sample of light flavor muon jets. The JET_20 data were considered

---

[16]Gluon splitting can also yield the final state $\{b\bar{b} + b\bar{b}\}$, but this state is not expected to be a source of the discrepancy illustrated in Tables 24 and 25

Label	TPR_{DATA} [%]	TPR_{MC} [%]
BA	38.57 ± 0.02	50.82 ± 0.01
BB	47.03 ± 0.02	58.69 ± 0.01

Table 25: True Positive Rates (TPR) for the bottom jet *vs* charm jet classifier, as found in $J/\Psi + \mu$-jet data and MC with bottom *vs* light flavor operating point B.

to be close enough to these requirements, despite the small heavy flavor component known to be part of this dataset. A single good muon jet was required for each event.

The FPR values found for operating points A and B are shown in Table 26. As can be seen by comparing Table 26 to Table 17, there is evidence for a heavy flavor component in the sample, on the order of 10%. On the other hand, this can be seen as an upper bound on the FPR for the bottom vs light flavor jet classifier.

Label	FPR_{DATA} [%]
A	9.764 ± 0.004
B	12.584 ± 0.004

Table 26: False Positive Rates (FPR) for the bottom jet vs light flavor jet classifier, as found in JET20 data.

11 Results

11.1 Double-Tag Sample Composition

In order to estimate the false discovery rate FDR for the double-tag sample, estimates of TPR and FPR must be made for each classifier. The operating points A and AA were the focus. Given that the JET_20 sample is known to contain a non-negligible heavy flavor component, the bottom vs light flavor classifier false positive rate estimated from this sample is not ideal for use in an FDR estimate. Therefore the FPR estimate for this classifier is taken to be the one from validation MC. For operating point A, Table 17 indicates an FPR of 0. Eq. 41 then gives an FDR of 0 for this classifier and operating point. From a sample of 122196 multi-muon events comprising 244473 μ-jets, 70143 μ-jets pass the bottom vs light flavor classifier, only 28296 of which occur in events with at least two μ-jets that pass the classifier.

The FDR estimate for the bottom vs charm jet classifier for operating point AA can then be made. The classifier performance metrics FPR and TPR are strictly only valid for single jets. However the only source of an estimate of the double-tag TPR and FPR is the same simulation which is under examination, MC@NLO. Therefore the assumption is made that the two μ-jets are produced independently. For operating point AA, the total number of μ-jets in events with at least two μ-jets passing the b vs light flavor classifier is $M = 28296$. The estimated number of positive b-jets is $\widehat{P} = 12376$. From these values for M and $\widehat{P}$, and the value of $FPR_{AA}(b/c) = 5.0 \pm 0.8\%$ from validation MC, given in Table 20, and an estimate of $TPR_{AA}(b/c)$, the false discovery rate may be estimated. FDR estimates for the b vs c classifier are made using the TPR estimates of Table 24; the results are given in Table 27. These results indicate that the number of charm jets in the final sample of b-jets (where both μ-jets are required to pass the AA bottom vs charm classifier) is small.

	TPR_{AA}^{MC} [%]	TPR_{AA}^{DATA} [%]
FDR_{AA}^{DATA}	2 ± 2	-1 ± 2

Table 27: False Discovery Rate estimates of the bottom vs charm jet classifier at operating point AA for the data sample after two μ-jets are required to pass the bottom vs light flavor classifier for operating point A.

Though the FDR_{AA}^{DATA} value quoted for TPR_{AA}^{DATA} in Table 27 is negative, this should not be cause for concern. The calculation of FDR from Eq. 41 necessarily includes an estimate of the false positive rate taken from validation MC. Such an estimate is most likely different

from the true FPR in data. However, given that it was not possible to measure the bottom vs charm jet classifier FPR from data reliably, the MC estimate must be used. If FDR_{AA}^{DATA} is assumed to be normally distributed, FDR_{AA}^{DATA} is less than 2.6% at 95% confidence.

In approximately 2.4 pb^{-1} of CDF data, there are 122,196 dimuon-triggered events which pass the event selection criteria prior to the jet flavor classifier. The number of events with two μ-jets which satisfy the jet flavor classifier thresholds for operating point AA is 2,889.

11.2   Systematic Uncertainties

Since this thesis is concerned solely with the comparison of the shapes of $b\bar{b}$ correlation distributions, sources of systematic uncertainty which contribute to a uniform shift in the normalization are not considered.

The overwhelmingly dominant source of systematic uncertainty in this analysis is the jet flavor classifier. The following procedure was used to estimate the uncertainty due to the classifier training and choice of operating points (described in Section 10). The nominal operating point used to produce the histograms in Section 11.3 was point AA. For each bin in a given histogram, the maximal difference δ with respect to the nominal distribution was found, among the corresponding histograms of the remaining operating points. The systematic uncertainty for a given bin was taken to be $\delta/\sqrt{3}$.

11.3 $b\bar{b}$ Correlations

All data and MC histograms in this section are normalized to unit area within the domain shown. For this section, the events are required to have two μ-jets which both pass the b vs light flavor and b vs c jet classifier thresholds. Thus μ-jets are now denoted as b-jets. For comparison to MC@NLO, PYTHIA was used to generate a $b\bar{b}$ sample using only the LO flavor creation processes; in particular, it should be noted that the standard PYTHIA parton shower was used. The same processing as used for the MC@NLO sample was applied to the PYTHIA sample, including the event selection.

The transverse momentum of the b-jet system $p_T(b-\text{jet}_1, b-\text{jet}_2)$ is shown in Fig. 48 for the AA classifier operating point. The error bars include both statistical and systematic sources of uncertainty (added in quadrature). The MC@NLO histogram is consistent with the data histogram, within the error bars. On the other hand, the LO PYTHIA histogram is not consistent with the others. This can be understood by recalling that the LO process produces a back-to-back angular topology and balanced in p_T; the vector sum of the two b-jet momenta must therefore have small magnitude and large pseudorapidity. The final state radiation implemented in the parton shower modifies this picture somewhat, but the

overall effect is a depletion at larger $p_T(b - \text{jet}_1, b - \text{jet}_2)$. This depletion is construed as evidence that the higher order processes are necessary to describe the data.
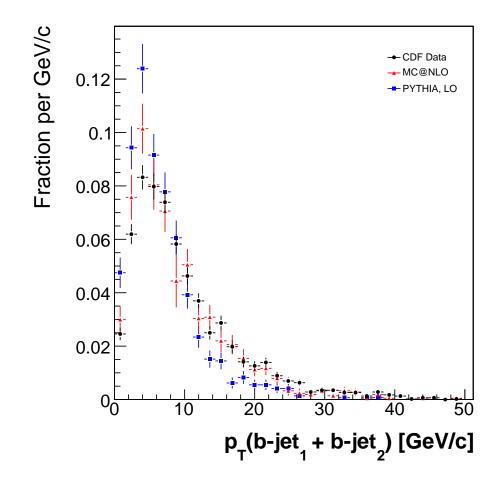

Figure 48: p_T of the b-jet system for CDF data (black points), MC@NLO (red triangles), and LO PYTHIA (blue squares).

The pseudorapidity of the b-jet system is shown in Fig. 49. The data and MC@NLO are in fair agreement for this distribution. The PYTHIA sample is not in agreement, however. The depletion at low $|\eta(b\text{jet}_1 + b\text{jet}_2)|$ in the LO PYTHIA sample indicates a lack of three-body final states where the light parton is non-collinear. The parton shower technique generates primarily collinear emission.

Figure 50 presents the distribution of azimuthal angular difference between the two b-jets. The minimum dimuon mass threshold in the trigger and the jet clustering parameter R_{jet} severely limit the acceptance for $\Delta\phi < 0.7$. The MC@NLO distribution is slightly narrower than the data in the peak region ($\Delta\phi > 2\pi/3$). This indicates a smaller proportion of higher order processes in MC@NLO than in data. The LO PYTHIA distribution is significantly
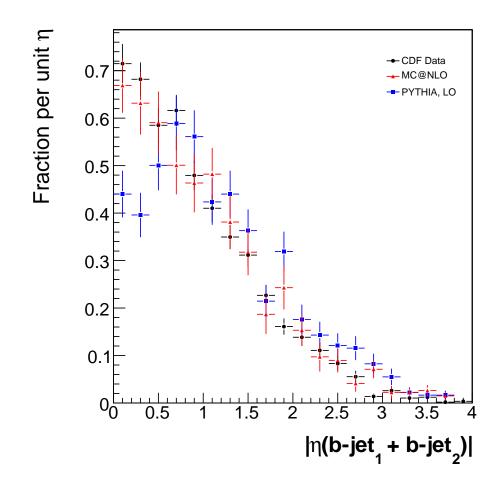
Figure 49: $|\eta|$ of the b-jet system for CDF data (black points), MC@NLO (red triangles), and LO PYTHIA (blue squares).

narrower than the data distribution, and it has negligible occupancy for the region enhanced in higher order b quark pair production processes, $\Delta\phi(b, \bar{b}) < 2\pi/3$.

There is a suggestive enhancement in the data for $0.8 \leq \Delta\phi < 1.0$; it is possible that this is the beginning of an NLO enhancement which was attenuated due to the trigger mass requirement and R_{jet} value. While the NLO differential cross section $d\sigma_{b\bar{b}}/d\Delta\phi_{b\bar{b}}$ predicts an enhancement for $\Delta\phi < 1$ due to gluon splitting, this enhancement is not present in the default MC@NLO when realistic p_T and rapidity requirements are placed on the b quarks, as shown in Fig. 51.
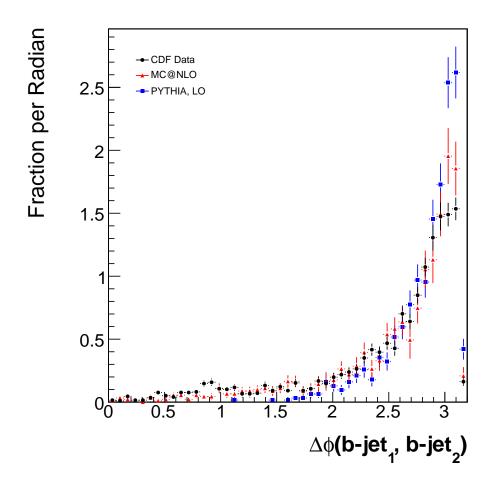

Figure 50: $\Delta\phi$ between the two b-jets for data (black points), MC@NLO (red triangles), , and LO PYTHIA (blue squares).

The distributions of p_T asymmetry between the two b-jets can be seen in Fig. 52. Within the error bars, the data and both simulations are in agreement. It is not surprising that the LO PYTHIA sample is consistent with the data in this distribution, as it includes final state emission through the parton shower model. The primary contribution from higher order production mechanisms is removed by the effective requirement $p_T(b\text{jet}) > 3.4$ GeV/c
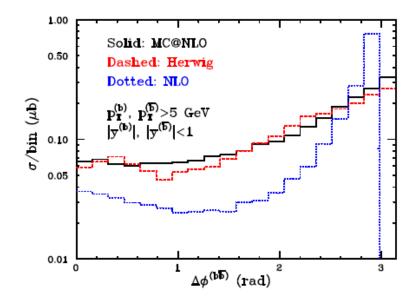
Figure 51: Azimuthal angular difference between bottom quarks with the requirements $p_T(b) > 5$ GeV/c and $|y(b)| < 1$ for MC@NLO (solid), HERWIG (dashed), and partonic NLO theory (dotted).

(derived from the combinations of muon requirements and muon jet definitions). According to the study of $b\bar{b}$ production mechanisms of Ref. [5], flavor excitation is the only production process which would produce any significant deviation in $\mathcal{A}_{p_T}$ from the flavor creation process (see Fig. 3). However the majority of events which might have a large asymmetry in flavor excitation processes produce one b quark which falls outside the CDF detector acceptance.
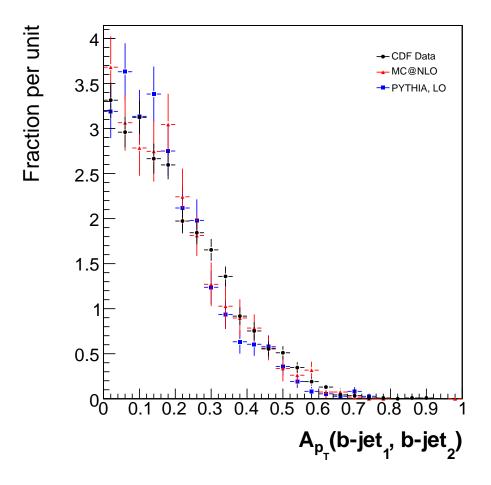

Figure 52: $\mathcal{A}_{p_T}$ for the two b-jets for data (black points), MC@NLO (red triangles), and LO PYTHIA (blue squares).

Figure 53 shows the psuedorapidity difference $\Delta\eta$ between the two b-jets for events which satisfy the azimuthal angular difference criterion, $\Delta\phi < 2\pi/3$. This criterion enhances the higher order component. While the data and MC@NLO are largely consistent within the uncertainties, there is a remarkable shape difference between the two distributions. As mentioned previously, the LO PYTHIA sample has negligible occupancy for this $\Delta\phi$ region. The $\Delta\eta$ distributions in the absence of this $\Delta\phi$ requirement, shown in Fig. 54, show the opposite behavior for $|\Delta\eta| < 0.8$.
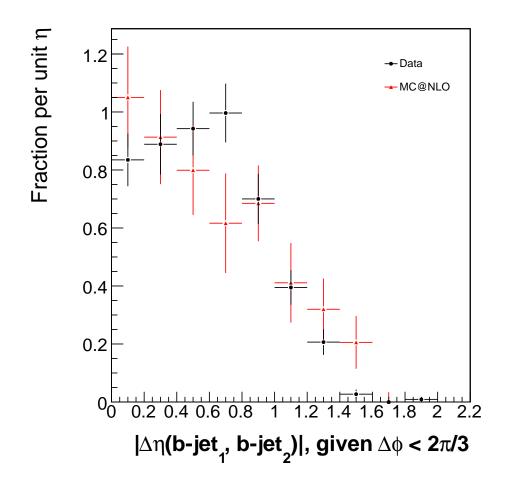
Figure 53: $|\Delta\eta|$ between the two b-jets, given $\Delta\phi < 2\pi/3$ for data (black points) and MC@NLO (red triangles).
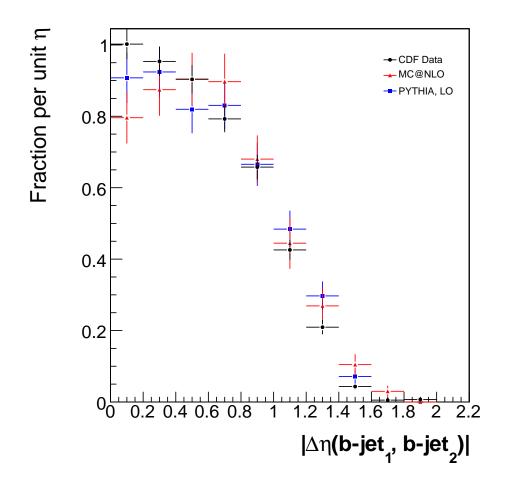
Figure 54: $|\Delta\eta|$ between the two b-jets for data (black points), MC@NLO (red triangles), and LO PYTHIA (blue squares).

12 Conclusion

Angular and momentum correlations between b-jets have been studied in dimuon-triggered data collected with the CDF detector. By examining correlations between b-jet pairs rather than muon pairs, the dependence upon non-perturbative physics in the estimation of the b quark flight direction is minimized. This strategy allowed a clearer test of the NLO theory of bottom quark pair hadroproduction than previous dimuon measurements in the low- and mid-$p_T(b)$ regime. A sub-sample of μ-jet pairs was derived from the dimuon trigger sample. A multivariate jet flavor classification algorithm was developed to identify a high-purity sample of bottom quark jet pairs.

The data are well-represented by the NLO theory predictions of the MC@NLO generator, on the whole. The non-perturbative physics models (underlying event and bottom quark fragmentation) embedded in MC@NLO were tuned to those of the PYTHIA TuneA event generator, which was previously shown to conform to CDF data in these respects. The LO flavor creation processes simulated by the PYTHIA generator are not representative of the data in the distributions of $\Delta\phi(b,\bar{b})$, $p_T(b+\bar{b})$, $\eta(b+\bar{b})$. These distributions are the most sensitive to three-body final states produced by $\mathcal{O}(\alpha_s^3)$ processes with the experimental apparatus and methods available. This analysis has shown that the NLO theory produces a consistent description of nature in the low- to mid-$p_T(b)$ regime, insofar as the non-perturbative effects are modeled well.

This analysis is limited in the small azimuthal angular difference region $\Delta\phi(b,\bar{b}) < 0.7$ by the spatial size of the jets and the mass requirement of the dimuon trigger. Thus this analysis is not sensitive to the portion of the $\Delta\phi(b,\bar{b})$ spectrum predicted to be enhanced in final state gluon splitting. All of the CDF triggers designed to record bottom quark pairs, including the lepton plus secondary vertex trigger (l + SVT), have an implicit or explicit angular requirement. Thus these triggers would suffer from a similarly limited access to the final state gluon splitting process. While these angular requirements have been necessary at CDF to limit the trigger rates at the high instantaneous luminosities achieved during Run II of the TeVatron, a trigger which has no such angular requirement is desirable for the purpose of studying heavy quark pair production. A trigger designed to capture heavy flavor pairs might best avoid the necessity of angular requirements by selecting a J/Ψ meson plus an additional lepton. As pointed out in Ref. [40], the only single particle which can produce this signature is $B_c \to J/\Psi l$, and techniques for estimating fake lepton contributions are well-known. Such a trigger would also allow the study of multiple heavy flavor pairs and the production of charmonia in association with open heavy flavor quarks.

References

[1] S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967); A. Salam, p. 367 of *Elementary Particle Theory*, ed. N. Svartholm(Almquist and Wiksells, Stokholm, 1969); S. L. Glashow, J. Iliopoulos and L. Maiani, Phys. Rev **D2**, 1285 (1970).

[2] J.C. Collins, D.E. Soper, and G. Sterman, Nucl. Phys. **B263**, 37 (1986).

[3] C. Amsler, *et al.*, Physics Letters **B 667**, 1 (2008).

[4] G. Sterman, arXiv:hep-ph/0412013.

[5] R.D. Field, *The sources of b-quarks at the Tevatron and Their Correlations*, Phys. Rev. **D65** 094006 (2002).

[6] F. Happacher, P. Giromini, and F. Ptohos, Phys. Rev. **D73**, 014026 (2006).

[7] T. Sjöstrand, L. Lönnblad, and S. Mrenna, PYTHIA *6.2 Physics and Manual*, arXiv:hep-ph/0108264.

[8] R. Field, *Min-Bias and the Underlying Event at the Tevatron and the LHC*, talk presented by at the Fermilab ME/MC Tuning Workshop, Fermilab, October 4, 2002.

[9] J. M. Butterworth, J. R. Forshaw and M. H. Seymour, Z. Phys. C **72**, 637 (1996) [arXiv:hep-ph/9601371];
J. M. Butterworth, M. H. Seymour, JIMMY4*: Multiparton interactions in* HERWIG *for the LHC*, draft obtained from http://projects.hepforge.org/jimmy/ (2005).

[10] HERWIG 6.5, G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M.H. Seymour and B.R. Webber, JHEP 0101 (2001) 010 [arXiv:hep-ph/0011363; hep-ph/0210213]

[11] G. J. Alner *et al.* [UA5 Collaboration], Nucl. Phys. B **291**, 445 (1987).

[12] S. Frixione and B.R. Webber, *Matching NLO QCD computations and parton shower simulations*, JHEP 0206 (2002) 029 [arXiv:hep-ph/0204244]

[13] S. Kawabata, Comput. Phys. Commun. **88** (1995) 309.

[14] S. Frixione, P. Nason and B.R. Webber, *Matching NLO QCD and parton showers in heavy flavour production*, JHEP 0308 (2003) 007 [arXiv:hep-ph/0305252]

[15] M.L. Mangano, P. Nason, and G. Ridolfi, Nucl. Phys. **B373** (1992) 295.

[16] M. R. Whalley, D. Bourilkov and R. C. Group, arXiv:hep-ph/0508110.

[17] D. Stump, J. Huston, J. Pumplin, W. K. Tung, H. L. Lai, S. Kuhlmann and J. F. Owens, JHEP **0310**, 046 (2003) [arXiv:hep-ph/0303013].

[18] M. Cacciari, G.P. Salam, and G. Soyez, Phys. Lett. **B641**, 57 (2006) [arXiv:hep-ph/0512210]

[19] GEANT3 Detector Design and Simulation Tool, *CERN Program Library Long Writeup W5013*, CERN, 1993.

[20] D.J. Lange, Nucl. Instrum. Methods Phys. Res., Sect. A **462**, 152 (2001).

[21] P. Artoisenet, J.P. Lansberg, F. Maltoni, Phys Lett. B, **653**, 60 (2007).

[22] J. Breitweg, *et al.* [ZEUS Collaboration] Eur. Phys. J. C 2 (1) 61-75 (1998).

[23] F. Abe *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **271**, 387 (1988).

[24] R. Blair *et al.*, Fermilab Report No. FERMILAB-Pub-96/390-E(1996).

[25] C. S. Hill *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **530**, 1 (2004).

[26] A. Sill *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **447**, 1 (2000).

[27] T. Affolder *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **453**, 84 (2000).

[28] T. Affolder *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **526**, 249 (2004).

[29] G. Ascoli *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **268**, 33 (1988).

[30] J. Elias *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **441**, 366 (2000).

[31] D. Acosta *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **461**, 540 (2001).

[32] XFT Reference

[33] C. Albajar *et al.* [UA1 Collaboration], Z. Phys. C **61**, 41 (1994).

[34] D. Acosta *et al.* [CDF Collaboration], Phys. Rev. **D71**, 112002 (2005).

[35] D. Acosta *et al.* [CDF Collaboration], Phys. Rev. **D69**, 072004 (2004).

[36] T. Shears, "Charm and Beauty Production at the Tevatron", talk presented at the Int. Euro-phys. Conf. on High Energy Phys., Lisboa, Portugal (2005)

[37] F. Abe *et al.* [CDF Collaboration], Phys. Rev. **D53**, 1051 (1996).

[38] B. Abbott *et al.* [DØCollaboration], Phys. Lett. **B487**, 264 (2000).

[39] F. Abe *et al.* [CDF Collaboration], Phys. Rev. D **55**, 2546 (1997).

[40] D. E. Acosta *et al.* [CDF Collaboration], Phys. Rev. D **71**, 092001 (2005) [arXiv:hep-ex/0412006].

[41] S. Klimenko, J. Konigsberg, and T. Liss, *Averaging of the inelastic cross-section measured by the CDF and E811 experiments*, CDF-6314.

[42] M. Cacciari and P. Nason, Phys. Rev. Lett. **89** (2002) 122003 [arXiv:hep-ph/0204025].

[43] M. Cacciari, S. Frixione, M. L. Mangano, P. Nason and G. Ridolfi, JHEP **0407**, 033 (2004) [arXiv:hep-ph/0312132].

[44] R.D. Field, *Pythia Tune A, Herwig, and Jimmy in Run 2 at CDF*, CDF-ANAL-CDF-PUBLIC-7822 [arXiv:hep-ph/0510198] (2005).

[45] D. Acosta, *et al.* [CDF Collaboration], Phys. Rev. **D71**, 032001 (2005).

[46] T. Aaltonen *et al.* [CDF Collaboration], Phys. Rev. **D77**, 072004 (2008).

[47] D. Usynin, Ph.D. thesis, University of Pennsylvania, FERMILAB-THESIS-2005-68 (2005).

[48] CDF Likelihood Soft Muon Tagger (CDF 7043)

[49] A. Abulencia *et al.* [CDF Collaboration], Phys. Rev. **D74**, 072006 (2006).

[50] K.R. Gibson, Ph.D. thesis, Carnegie Mellon University, FERMILAB-THESIS-2006-09 (2006).

[51] I. Narsky, *StatPatternRecognition: A C++ Package for Statistical Analysis of High Energy Physics Data*, arXiv:physics/0507143v1 (2005); software available from http://sourceforge.net/projects/statpatrec/

[52] B. Efron and R.J. Tibshirani, Chapman & Hall/CRC, 1993.

[53] L. Breiman, Machine Learning **26**, 123-140 (1996); 553-568 (1997).

[54] Y. Freund and R. Schapire, Journal of Computer and System Sciences, **55**, 325-332 (1997).

[55] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics (2001).

[56] T. Fawcett, Pattern Recognition Letters **27**, 861-874 (2006).

[57] J.D. Storey, Journal of the Royal Statistical Society: Series B, **64**, issue 3, 479-498 (2002).

[58] R. Storn, K. Price, "Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces", *Journal of Global Optimization*, Volume **11**, Issue 4, Pages 341 - 359, 1997.

[59] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes in C: the Art of Scientific Computing*, 2nd ed., Cambridge University Press (1992).

[60] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes: the Art of Scientific Computing*, 3rd ed., Cambridge University Press (2007).

[61] J.A. Nelder and R. Mead, "A Simplex Method for Function Minimization", *The Computer Journal*, Volume **7**, Issue 4, Pages 308 - 313, 1965.

[62] R. Hemingway, "An OPAL tune of HERWIG 6.1 using Z^0 data", OPAL Technical Note TN652 (2000).

[63] G. Corcella and V. Drollinger, "Bottom-quark fragmentation: comparing results from tuned event generators and resummed calculations", arXiv:hep-ph/0508013 (2005).

[64] D.E. Goldberg, "Genetic algorithms in search, optimization, and machine learning", Reading: Addison-Wesley (1989).

[65] C.M. Fonseca, and P.J. Fleming, prodeedings of *The International Conference on Genetic Algorithms in Engineering Systems: Inovations and Applications*, United Kingdom, pp. 12-14 (1993).

[66] K. Deb, *et al.*, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, **6**, 182 (2000).

[67] J. Bastos, Y. Liu, "A multivariate approach to heavy flavor tagging with cascade training", JINST, **2**, 11007 (2007).