

# Measurement of the $b\bar{b}$ di-jet cross section at CDF

## THÈSE

présentée à la Faculté des sciences de l'Université de Genève  
pour obtenir le grade de Docteur en sciences, mention physique

par

**Sofia Vallecorsa**

*d'Italie*

Thèse N° 3916

La Faculté des sciences, sur le préavis de Messieurs A. G. CLARK, professeur ordinaire et directeur de thèse (Département de physique nucléaire et corpusculaire), M. POHL, professeur ordinaire (Département de physique nucléaire et corpusculaire), X. WU, docteur (Département de physique nucléaire et corpusculaire), et B. HEINEMANN, professeur (University of California, Berkeley – Lawrence Berkeley National Laboratory – Berkeley (CA), U.S.A.), autorise l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 30 novembre 2007

**Thèse - 3916 -**

  
**Le Doyen, Jean-Marc TRISCONE**

# Abstract

The dominant  $b$  production mechanism at the Tevatron is pair production through strong interactions. The lowest order QCD diagrams contain only  $b$  and  $\bar{b}$  quarks in the final state, for which momentum conservation requires the quarks to be produced back-to-back in azimuthal opening angle. When higher order QCD processes are considered, the presence of additional light quarks and gluons in the final state allows the azimuthal angle difference,  $\Delta\phi$ , to spread. The next to leading order QCD calculation includes diagrams up to  $O(\alpha_s^3)$  some of which, commonly known as flavor excitation and gluon splitting, provide a contribution of approximately the same magnitude as the lowest order diagrams.

The study of  $b\bar{b}$  angular correlation gives predictions on the effective  $b$  quark production mechanisms and on the different contributions of the leading order and next-to-leading order terms.

The first experimental results on inclusive bottom production at the Tevatron were strongly underestimated by the exact NLO QCD prediction. Later on this disagreement had been explained and reduced by theoretical and experimental improvements: new QCD calculations that implement the Fixed Order with Next-to-Leading-Logarithms calculation (FONLL); updated parton distribution functions and fragmentation functions; and more precise measurements. Previous measurements of  $b\bar{b}$  azimuthal angle correlation have, instead, reached various level of agreement with parton shower Monte Carlo and NLO predictions.

Here we present a measurement of the  $b\bar{b}$  jet cross section and azimuthal angle correlation performed on about  $260\text{ pb}^{-1}$  of data collected by the CDF II detector at Fermilab from March 2002 to September 2004. This study extends the energy range investigated by previous analyses, measuring jet transverse energies ( $E_T$ ) up to values of about 220 GeV. It relies on the good tracking capabilities of the CDF detector both at the trigger level and offline. Events with heavy quarks are selected online using the Secondary Vertex Trigger (SVT), which can measure in real time the impact parameter of the tracks, in particular those originated from the decay of long-lived particles. The SVT represents the key element for all the heavy flavor measurement performed by CDF, and this analysis describes one of the first cases in which the SVT trigger is used to study high  $p_T$  physics.

The total cross section is measured together with the di-jet differential cross sections as a function of the highest energy jet  $E_T$  and the di-jet invariant mass. The azimuthal angular correlation ( $\Delta\phi$ ) between the two jets is also measured. As expected this distribution proves that the largest contribution to  $b\bar{b}$  production is due to lowest order QCD diagrams, corresponding to a back to back configuration of the two  $b$ -jets (large  $\Delta\phi$  values).

The most interesting fact is, however, that the low  $\Delta\phi$  region also results highly populated, suggesting an important role played by higher order production terms.

To verify this conclusion, results are compared to Monte Carlo predictions at leading order and next to leading order QCD. When technical details are correctly taken into account, as the contribution of the underlying event for example, it is possible to conclude that the data are in agreement with a next to leading order model. Nevertheless the agreement is not perfect and the data present some excess with respect to theoretical predictions.

This thesis describes the analysis steps in details as support to the PRL paper forseen to be published soon.

# Résumé

Ce travail de thèse décrit les mesures de section efficace de production d'un paire de quarks  $b\bar{b}$  dans des collisions  $p\bar{p}$  de  $\sqrt{s} = 1.96 \text{ TeV}$  dans le centre de masse, réalisées par l'expérience CDF. Après une brève introduction théorique, le détecteur CDF et ses divers composants sont présentés, puis la stratégie d'analyse utilisée est expliquée en détail. L'analyse utilise des jets dans la région de pseudo-rapacité centrale ( $|\eta| < 1.2$ ) reconstruits avec l'algorithme 'JetClu' pour des énergies transverses  $E_T^{jet}$  allant jusqu'à  $220 \text{ GeV}$ . Les bonnes performances du système de déclenchement SVT sont exploitées ainsi que la bonne capacité de reconstruction de traces du détecteur CDF pour identifier les jets  $b$  en reconstruisant le vertex secondaire avec l'algorithme 'SecVtx'. Les résultats obtenus sont comparés avec les prédictions théoriques.

## Introduction théorique

Le Modèle Standard (SM) des interactions électrofaibles et fortes décrit les interactions entre les particules élémentaires en utilisant la théorie quantique des champs. Jusqu'à présent, le Modèle Standard a connu un succès remarquable par ces prédictions, qui ont été vérifiées expérimentalement.

Les interactions sont basées sur le groupe de symétrie de jauge  $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ , qui décrit les interactions électromagnétiques, faibles et fortes, par l'échange de champs de jauge. En particulier, la Chromodynamique Quantique est la théorie de jauge des interactions fortes: le groupe de couleur  $SU(3)_C$  sur lequel cette théorie se fonde, implique l'existence de 8 gluons, médiateurs de l'interaction. La portée de l'interaction forte est caractérisée par la constante de couplage forte  $\alpha_s$ , qui décroît lorsque le module carré du quadri-moment d'impulsion  $|Q^2|$  transférée lors de la collision augmente. Les sections efficaces de la QCD peuvent être approximées par des développements en série de  $\alpha_s$ : si  $\alpha_s$  est suffisamment petit, la théorie des perturbations est justifiée. L'énergie dans le centre de masse au Tevatron est déjà suffisamment élevée pour permettre des mesures de précision des prédictions de la QCD, et spécialement pour la production de quarks bottom. Les mesures de la section efficace de production des jets  $b$  fournissent ainsi un important test quantitatif de la QCD.

Au Tevatron, les quarks  $b$  sont produits principalement en paires par l'interaction forte. À l'ordre le plus bas, les diagrammes de production ne comportent que des paires de quarks  $b$  et  $\bar{b}$  dans l'état final. Dans ce cas, la loi de conservation de l'impulsion requiert que les deux quarks soient produits dos à dos avec une grande distance dans l'angle azimutal  $\Delta\phi$ . Lorsque les ordres supérieurs sont considérés, la présence supplémentaire de quarks légers et de gluons dans l'état final permet au  $\Delta\phi$  de s'étendre. En particulier, les calculs au second ordre (NLO) contiennent des diagrammes jusqu'à  $O(\alpha_s^3)$  et certains de ceux-ci, communément connus sous le nom de 'flavour excitation' et 'gluons splitting', apportent une contribution d'environ la même ampleur que ceux d'ordre  $O(\alpha_s^2)$ .

Une mesure de la corrélation angulaire  $\Delta\phi$  entre les quarks  $b$  et  $\bar{b}$  donne donc un aperçu des mécanismes de production du quark  $b$  et sur les différentes contributions des termes au premier et second ordre de la théorie perturbative des interactions fortes. Les prédictions théoriques au second ordre sous-estimaient les premiers résultats expérimentaux sur la production inclusive des quarks  $b$  au Tevatron. Seulement plus tard ce désaccord a été expliqué et réduit par l'amélioration des calculs perturbatifs, grâce à l'introduction du calcul 'Fixed-Order at Next-to-Leading-Logarithms' (FONLL), grâce à la mise à jour des fonctions de distribution partonique et des fonctions de fragmentation, et aussi par des mesures plus précises. Les mesures précédentes de corrélation entre  $b\bar{b}$  ont atteint divers niveaux d'accord avec les simulations des gerbes partoniques (Parton Shower Monte Carlo) et les prédictions au NLO.

## Le détecteur CDF

Le détecteur CDF II a une symétrie cylindrique autour de la ligne de faisceau, rendant avantageuse l'utilisation d'un système de coordonnées cylindriques avec l'axe Z le long de la direction du faisceau de protons. On définit  $r$  comme étant la distance par rapport à la ligne de faisceau et  $\phi$  est l'angle azimutal mesuré à partir de la direction pointant radialement vers l'extérieur dans le plan du Tevatron. Les quantités 'transversales', telles que l'énergie transverse d'un jet,  $E_T$ , ou l'impulsion transverse d'une particule,  $P_T$ , sont définies dans le plan perpendiculaire à la ligne de faisceau.

Une description plus détaillée du détecteur CDF sera donnée dans la thèse. Ici, seulement les composants utilisés dans la mesure sont décrits brièvement: le calorimètre central et le détecteur de traces des particules chargées. La partie centrale du calorimètre, couvrant la région  $|\eta| < 1.1$  est divisée en deux parties: le calorimètre électromagnétique et le calorimètre hadronique, basés tous les deux sur la technologie des scintillateurs. Ils mesurent l'énergie des particules avec des résolutions, exprimée en GeV, de  $13.5\%/\sqrt{E_T} \oplus 1.5\%$  pour les électrons et de  $50\%/\sqrt{E_T} \oplus 3\%$  pour les pions. À l'intérieur du calorimètre se trouve un solénoïde qui produit un champ

magnétique de 1,4 Tesla. Dans le solénoïde une chambre à fils (le COT) est composée de 96 couches alternées de fils de mesure axiaux et stéréo. Le COT est capable de reconstruire les traces des particules dans la région de  $|\eta| < 1.0$  avec une impulsion transverse supérieure à 400  $MeV$ . La résolution sur l'impulsion transverse est de  $\Delta p_T/p_T = 0.15\% p_T/GeV/c$ .

À l'intérieur du COT, à proximité de la ligne de faisceau, il y a le détecteur de vertex au silicium. Il est composé de trois sous-détecteurs: l'ISL, Intermediate Silicon Layer, le SVX, Silicon Vertex Detector et L00. Ces systèmes mesurent la position d'un vertex de désintégration avec une grande précision: La résolution du paramètre d'impact est d'environ 15  $\mu m$ , pour les traces isolées, comparé aux 600  $\mu m$  obtenues à l'aide du COT.

Le système de déclenchement est composé de trois niveaux successifs, chacun diminuant le nombre d'événements d'un ou deux ordres de grandeur. L'information du détecteur au silicium sur les traces est utilisée au deuxième niveau (L2) du système de déclenchement par le SVT, Silicon Vertex Trigger, qui combine cette information avec traces reconstruites au niveau 1 (L1). Le SVT mesure le paramètre d'impact des traces avec une résolution d'environ 37  $\mu m$ , dont 30  $\mu m$  représentent l'incertitude sur la position du faisceau.

## Mesure de la section efficace

L'utilisation de jets pour la production des quarks  $b$  permet de dépasser la limite atteinte par les mesures précédentes, qui utilisaient les désintégrations exclusives de hadrons  $B$ , et permet ainsi d'explorer une large gamme d'énergies transverses. De plus, on s'attend que les corrections dues aux effets de fragmentation auxquels les calculs théoriques sont sensibles, soient petits pour les jets.

L'échantillon est sélectionné au moyen d'un système de déclenchement spécialement conçu pour choisir des événements riches en quarks lourds, en utilisant le SVT au L2. Les événements sont sélectionnés en demandant deux jets avec une énergie transverse de  $E_T > 20 GeV$ , chacun associé à une trace, reconstruite par le SVT, avec un grand paramètre d'impact.

L'énergie transverse mesurée des jets est par définition biaisée par les effets du détecteur (pertes en calorimétrie, efficacité finie du détecteur, effets de lissage dus à la résolution finie du détecteur). Ces sont communément appelés jet calorimétriques, et ils dépendent de façon évidente, de l'expérience. Lorsqu'on compare les sections efficaces ou d'autres mesures avec les prédictions théoriques, les jets impliqués sont reconstruits en utilisant les particules stables (hadrons). Les jets calorimétriques doivent alors être corrigés pour éliminer les effets du détecteur. Pour ce faire il existe des corrections génériques mesurées par CDF. Une correction supplémentaire (de l'ordre de 5 %) est nécessaire pour tenir en compte les caractéristiques spécifiques des

jets  $b$  comme la présence de hadrons  $B$  qui se désintègrent à l'intérieur du jet et qui affectent la mesure de l'énergie du jet. La procédure pour obtenir la correction utilise la simulation. Dans les échantillons Monte Carlo (on a utilisé principalement les jets et le jets  $b$  de PYTHIA), pour former des jets au niveau hadronique, on utilise les informations provenant des particules stables et on compare ces jets à ceux reconstruits avec des informations fournies par le calorimètre.

La stratégie que nous utilisons dans cette mesure consiste à appliquer une sélection plus serrée que celle du système de déclenchement. L'avantage de cette approche est qu'il n'est pas nécessaire de mesurer l'efficacité de sélection du système de déclenchement. Seulement l'efficacité de la sélection finale est mesurée en utilisant le Monte Carlo. Les événements sont sélectionnés s'ils ont deux jets avec des énergies transverses supérieures à  $35 \text{ GeV}$  et  $32 \text{ GeV}$  respectivement, dans la région centrale du détecteur, ce qui correspond à une pseudo-rapacité  $|\eta| < 1.2$ , chacun identifié comme un  $b$ -jet (un jet de saveur  $b$ ).

Une des techniques développées pour l'identification de  $b$ -jets, 'b-tagging', tire partie de la longue durée de vie des hadrons  $B$ . Le temps de vie moyen est de l'ordre de  $1.5 \text{ ps}$ , et la distance propre  $c\tau$  est donc d'environ  $450 \mu\text{m}$ : en considérant l'important changement de référentiel (boost) relativiste des événements pris en compte, le hadron  $B$  parcourt une distance de l'ordre de quelques millimètres avant de se désintégrer. Les produits de désintégration proviennent alors d'un vertex secondaire, déplacé par rapport au point d'interaction.

L'algorithme de 'b-tagging' SecVtx exploite les performances du détecteur de vertex au silicium pour distinguer entre les traces qui dérivent d'un vertex primaire et celle qui proviennent de vertex déplacés. Cet algorithme applique une sélection sur le paramètre d'impact des traces et reconstruit un vertex secondaire à partir des traces sélectionnées. L'efficacité est alors définie comme la fraction des événements avec deux  $b$ -jets identifiés (et incluant chacun une trace SVT) par rapport aux événements avec deux  $b$ -jets. Cette efficacité est très petite, autour de 2%.

Un aspect négatif d'une telle procédure réside dans le fait que les hadrons charmés sont aussi caractérisés par un parcours propre intermédiaire ( $c\tau \approx 20 \mu\text{m}$ ), et il est donc difficile de distinguer les jets  $b$  des jets  $c$ . En plus, les jets légers peuvent aussi être incorrectement identifiés. Il est donc nécessaire d'extraire le contenu en saveur  $b$  des événements sélectionnés. La technique adoptée dans l'analyse présentée ici considère, comme quantité discriminatrice, la forme de la masse invariante des traces utilisées pour trouver le vertex secondaire. En fait, la masse du vertex secondaire est plus grande pour les jets  $b$  que pour les jets  $c$ , ou les jets légers. Comme on sélectionne deux jets dans chaque événement, la somme des masses invariantes pour chaque jet est calculée avec les données et les Monte Carlos. Les distributions fournies par les données sont ajustées à la courbe formée par une combinaison linéaire des deux composantes: le cas  $b\bar{b}$  et le cas où au moins un jet n'est pas un  $b$ -jet. Cette méthode

donne une pureté très élevée, autour de 85%.

La section efficace différentielle des deux jets  $b\bar{b}$  est mesurée au niveau des particules en fonction de l'énergie transverse du jet le plus énergétique et en fonction de l'angle  $\Delta\phi$  entre les deux jets:  $\sigma_{b\bar{b}} = \frac{N_{2svt} \cdot f_{b\bar{b}}}{\epsilon \cdot \Delta\eta \cdot \mathcal{L}} \cdot C$  où  $N_{2svt}$  est le nombre d'événements avec deux 'SVT-tagged' jets;  $\mathcal{L}$  est la luminosité intégrée  $f_{b\bar{b}}$  est la pureté  $b\bar{b}$ ;  $\epsilon$  est l'efficacité de l'identification des événements  $b\bar{b}$ .  $C$  est le facteur de correction spécifique aux jets  $b$  qui corrige la section efficace au niveau des particules.

Les sources principales d'erreurs systématiques sont l'échelle d'énergie des jets et la fraction des événements  $b\bar{b}$  parmi ceux contenant deux 'SVT-tagged' jets. Les sections efficaces sont montrées dans la figure 1 en fonction de l'énergie transverse du jet et de l'angle  $\Delta\phi$ . Elles sont aussi comparées avec les prédictions au premier (LO) et au second (NLO) ordre.

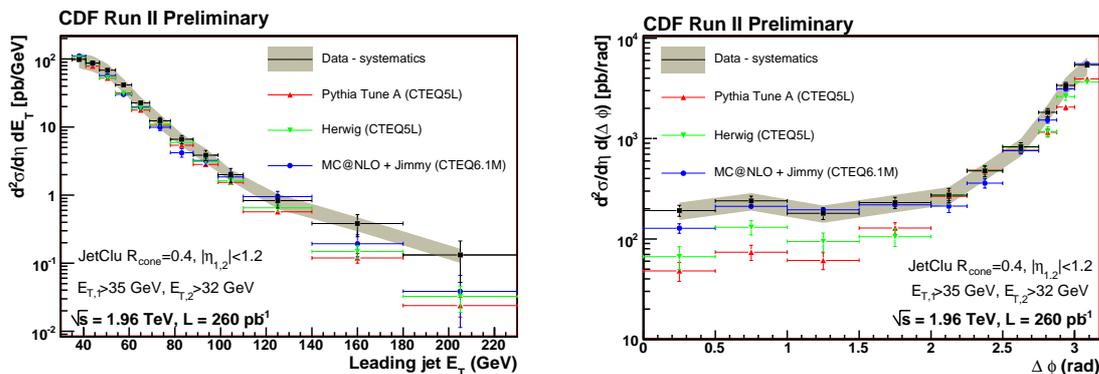


Figure 1: Section efficace différentielle mesurée en fonction de l'énergie transverse du jet (gauche) et en fonction de l'angle  $\Delta\phi$  (droite). Les données sont comparées aux trois Monte Carlos: PYTHIA, HERWIG + JIMMY et MC@NLO + JIMMY.

La distribution  $\Delta\phi$  montre que la plupart des événements  $b\bar{b}$  sont essentiellement produits par les processus au premier ordre de 'flavour creation'. Cependant, l'excès aux petites angles ( $\Delta\phi < 1,5$ ) suggère que la contribution des diagrammes d'ordre supérieur et des interactions multiples ne peut pas être négligée. Dans cette perspective, la prédiction du MC@NLO + JIMMY fournit une bien meilleure description de  $\Delta\phi$  que les Monte Carlos LO (comme PYTHIA et HERWIG). Aussi, un traitement correct des interactions multiples (comme dans le cas de HERWIG + JIMMY) donne une meilleure prédiction du résultat. Finalement l'inclusion de diagrammes d'ordre supérieur (avec MC@NLO + JIMMY) est compatible avec les données à l'intérieur des erreurs systématiques.



# Contents

<b>1</b>	<b>Theoretical introduction</b>	<b>1</b>
1.1	Standard Model . . . . .	1
1.2	The Strong Interaction (QCD) . . . . .	3
1.2.1	Factorization Theorem . . . . .	5
1.2.2	Parton Distribution Functions . . . . .	6
1.3	Monte Carlo methods . . . . .	7
1.3.1	Parton shower generators . . . . .	8
1.3.2	Hadronization . . . . .	9
1.3.3	Matrix Element generators and the problem of combination with showering . . . . .	12
1.4	Bottom production at Next-to-Leding Order . . . . .	13
1.4.1	Production of $b$ quarks at the Tevatron . . . . .	14
1.5	Past $b$ cross section measurements at $p\bar{p}$ colliders . . . . .	16
1.6	$b\bar{b}$ correlated production . . . . .	18
1.6.1	Past $b\bar{b}$ correlation measurements . . . . .	19
<b>2</b>	<b>The Tevatron and the CDF detector</b>	<b>23</b>
2.1	FNAL - the Fermi National Accelerator Laboratory . . . . .	23
2.2	The Fermilab accelerator complex . . . . .	24
2.2.1	Proton source . . . . .	24
2.2.2	Antiproton source . . . . .	25
2.2.3	TEVATRON . . . . .	26
2.2.4	Performance and future perspectives . . . . .	27
2.3	The CDF II detector . . . . .	29
2.3.1	Tracking system . . . . .	31
2.3.2	The calorimetry . . . . .	36
2.3.3	The muon system . . . . .	38
2.4	The CDF Data Acquisition and Trigger . . . . .	38
2.5	The Silicon Vertex Tracker . . . . .	41
2.6	On-line Monitoring . . . . .	44

2.6.1	CDF Luminosity . . . . .	45
<b>3</b>	<b>Event Reconstruction</b>	<b>47</b>
3.1	Track reconstruction . . . . .	47
3.1.1	Tracking in the COT . . . . .	48
3.1.2	Silicon tracking . . . . .	49
3.2	Jet reconstruction . . . . .	51
3.2.1	Cone Algorithms . . . . .	53
3.2.2	The ‘correct’ jet algorithm . . . . .	57
3.3	Evaluation of the Primary vertex position . . . . .	59
3.3.1	Z-vertex reconstruction . . . . .	60
3.3.2	Primary Vertex Finding . . . . .	61
3.4	Data samples . . . . .	62
3.4.1	The HIGH_PT_BJET trigger . . . . .	62
3.4.2	Monte Carlo samples . . . . .	63
3.5	Event selection . . . . .	66
3.6	Jet energy corrections . . . . .	71
3.6.1	Additional corrections on tagged jets . . . . .	73
3.6.2	Systematic uncertainties . . . . .	76
<b>4</b>	<b>Tagging <math>b</math>-jets</b>	<b>81</b>
4.1	Different techniques to identify $b$ -jets . . . . .	81
4.2	The Secondary Vertex (SECVTX) algorithm . . . . .	82
4.3	The $b$ -tagging efficiency . . . . .	86
4.3.1	The effect of SVT selection . . . . .	87
4.3.2	The SVT-tagging scale factor . . . . .	91
4.4	The $b\bar{b}$ content of the tagged jet sample . . . . .	91
4.5	Systematic Uncertainties . . . . .	102
4.5.1	Heavy quark multiplicity in the jets . . . . .	102
4.5.2	Track reconstruction inefficiency . . . . .	115
4.5.3	Other sources of systematic uncertainty on the $b\bar{b}$ fraction . . . . .	115
4.5.4	Total Uncertainty related to $b\bar{b}$ -jet fraction . . . . .	119
4.5.5	The data/Monte Carlo scale factor . . . . .	119
<b>5</b>	<b>Results and comparison to Monte Carlo</b>	<b>123</b>
5.1	The raw cross sections . . . . .	123
5.2	Unfolding procedure . . . . .	124
5.2.1	Systematic uncertainty on unfolding factors . . . . .	125
5.3	Total systematic uncertainties . . . . .	128
5.4	Results . . . . .	134
5.4.1	The total cross section . . . . .	134

5.5	Comparison to Monte Carlo . . . . .	140
5.5.1	The importance of the underlying event . . . . .	144
5.6	Conclusions . . . . .	145
<b>A</b>	<b>Simulation of the SVT beamline</b>	<b>149</b>
<b>B</b>	<b>The SVT-tagging scale factor</b>	<b>151</b>
<b>C</b>	<b>Unfolding factors and jet energy corrections</b>	<b>161</b>
	<b>Bibliography</b>	<b>165</b>



# Chapter 1

## Theoretical introduction

Bottom production cross section measurements represent an important test of perturbative QCD and also, a valuable source of information to study and test its state of the art modeling using Monte Carlo methods. This is even more so for the  $b\bar{b}$  jet cross section measurement where some observables can be found that are sensitive to production mechanisms and to the relative contribution of the Leading Order (LO) and Next to Leading Order (NLO) terms. This chapter represents a brief introduction to different theoretical and phenomenological aspects of bottom production at a hadron collider. After a few words on perturbative QCD in section 1.2, a short description of the Monte Carlo approach to QCD calculations is given, paying special attention to the different methods of Parton Showers and Matrix Elements and to the problem of a correct combination between the two (section 1.3). The last part of the chapter is dedicated to the theoretical prediction and phenomenology of  $b$  production at the Tevatron. Previous bottom production measurements are described in section 1.5. Section 1.6 focuses on the importance of a  $b\bar{b}$  angular correlation study to improve the present understanding on the production mechanisms: in fact, a NLO prediction of the full kinematics of bottom pair production has not yet been calculated in a closed form.

### 1.1 Standard Model

The Standard Model of particle physics is the current theory of the interactions of fundamental particles. The model has been extremely successful and almost all experimental measurements are consistent with the model, within measurement uncertainties. Within the Standard Model, matter consists of fermions with spin  $\hbar/2$  and the strong, weak and electro-magnetic forces are mediated by spin  $\hbar$  gauge bosons. Gravity remains to be incorporated in the Standard Model as a quantized theory. It is assumed to be mediated by a massless particle of spin  $2\hbar$ , called the graviton. For

each fundamental particle, an antiparticle exists with opposite electrical charge and parity.

Two types of fundamental fermions exist in the Standard Model, leptons and quarks. There are three generations of leptons, each of which consists of charged particle (electron, muon or tau) and a neutral partner (electron, muon and tau neutrinos). The generations are arranged in a mass hierarchy, whose source is so far unknown. Leptons are only sensitive to the electroweak and gravitational forces. In an analogous manner, three generations of quarks exist. Each generation consists of a charge  $+2/3$  quark (up, charm and top) and a charge  $-1/3$  quark (down, strange and bottom). Quarks carry strong interaction charge, ‘color’, as well as the electromagnetic charge and therefore can interact via the strong force as well as the electroweak and gravitational forces. Tables 1.1 and 1.2 summarize the characteristics of the fundamental fermions in the Standard Model.

Leptons		
Flavor	Mass (GeV)	Charge
Electron ( $e$ )	0.000511 (est.)	-1
Neutrino $\nu_e$	$< 225 \text{ eV}$ @95% C.L.	0
Muon ( $\mu$ )	0.106	-1
Neutrino $\nu_\mu$	$< 0.19 \text{ MeV}$ @90% C.L.	0
Tau ( $\tau$ )	1.777	-1
Neutrino $\nu_\tau$	$< 18.2 \text{ MeV}$ @95% C.L.	0

Table 1.1: Fundamental fermions in the Standard Model: leptons

Quarks		
Flavor	Mass (GeV)	Charge
Up ( $u$ )	0.003	$2/3$
Down ( $d$ )	0.006	$-1/3$
Charm ( $c$ )	1.3	$2/3$
Strange ( $s$ )	0.1	$-1/3$
Top ( $t$ )	175	$2/3$
Bottom ( $b$ )	4.3	$-1/3$

Table 1.2: Fundamental fermions in the Standard Model: quarks

The photon is the spin  $1\hbar$  gauge boson which mediates the electro-magnetic force. The weak force is carried by 3 massive spin  $1\hbar$  bosons,  $W^+, W^-$  and  $Z^0$ . The strong

Bosons				
	Mass (GeV)	Charge	Force	Affected particles
Graviton	0	0	Gravitation	All
Photon ( $\gamma$ )	0	0	Electromagnetic	All charged
$W^+$	80.4	+1	Weak	Quarks, Leptons
$W^-$	80.4	-1	Weak	Quarks, Leptons
$Z^0$	91.2	0	Weak	Quarks, Leptons
Gluon ( $g$ )	0	0	Strong	Quarks, Gluons
Higgs ( $H^0$ )	> 114	0		Massive particles

Table 1.3: Fundamental bosons in the Standard Model

interaction is transmitted by 8 ‘colored’ spin  $1\hbar$  gluons. Finally, the Higgs Boson ( $H^0$ ) is believed to be the source of electroweak symmetry breaking and mass. It has not yet been observed. Table 1.3 summarizes the fundamental bosons in the Standard Model.

## 1.2 The Strong Interaction (QCD)

Quantum Chromodynamics (QCD), the theory of the strong interaction within the Standard Model, was developed [1] as a non-Abelian SU(3) gauge theory, following the way opened by QED and Yang-Mills theories. The charge within QCD is the ‘color’; each quark has 3 possible colors and gluons, the force carrier in QCD, have the 8 different possible combinations of the 3 fundamental colors. As gluons carry color, gluons can self-couple, unlike photons in QED. The Lagrangian is given by:

$$L_{QCD} = \sum_{flavors} \bar{q}_a (i\gamma^\mu D_\mu - m_q)_{ab} q_b - \frac{1}{4} F_{\alpha\beta}^A F_A^{\alpha\beta}$$

where the sum runs over the six different flavors of the quarks.  $F_{\alpha\beta}^A$  is the field strength tensor derived for the gluon field  $A_\alpha^A$  as

$$F_{\alpha\beta}^A = [\partial_\alpha A_\beta^A - \partial_\beta A_\alpha^A - gf^{ABC} A_\alpha^B A_\beta^C]$$

and the indexes  $A, B, C$  run over the eight color degrees of freedom of gluon field.  $g$  is the coupling constant, which determines the strength of the interaction between quarks and gluons. The third term shows the non-abelian nature of QCD. This term describes the property of the interaction among gluons. Interaction between two, three, or more gluons can occur, resulting to a very different behavior compared to

the electromagnetic interaction. Self-coupling of gluons is the reason why the strong coupling constant,  $\alpha_s = \frac{g^2}{4\pi}$ , is large at small momentum transfer (large distances) and decreases at high momentum transfer (small distance) as shown in figure 1.1.

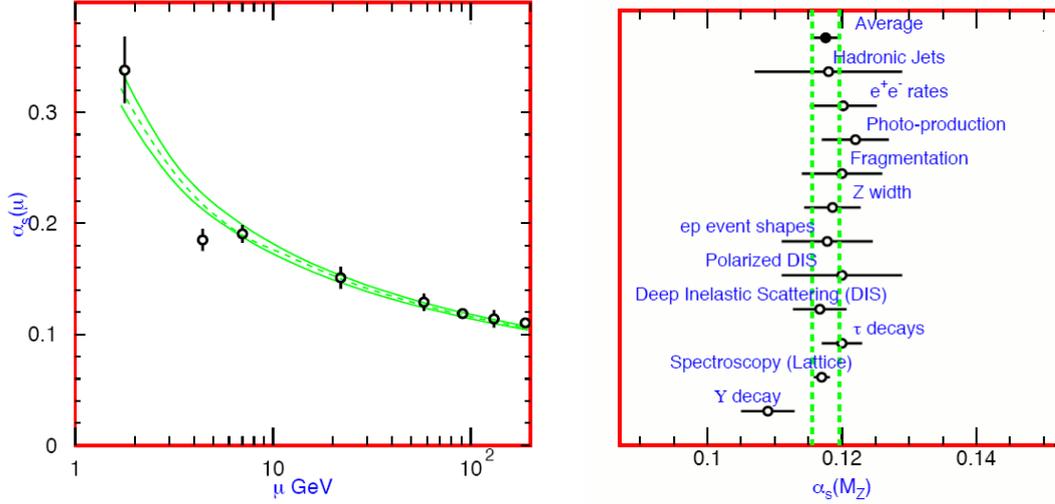


Figure 1.1: Left: Summary of the values of  $\alpha_s$ . The lines show the central value and the  $\pm 1\sigma$  of the PDG average. The data are, in increasing order of  $\mu$ :  $\tau$  width,  $\Upsilon$  decays, deep inelastic scattering,  $e^+e^-$  event shapes from the JADE data at 22 GeV and TRISTAN at 58 GeV,  $Z$  width, and  $e^+e^-$  event shapes at 135 GeV and 189 GeV. Right: Summary of the values of  $\alpha_s(M_Z)$  from various processes extrapolated at  $\mu = M_Z$ . It is shown the total error including theoretical uncertainties. The PDG average  $\alpha_s = 0.1176 \pm 0.002$  is also shown. [2].

This behaviour, known as *running coupling constant* explains two observed phenomena in QCD: asymptotic freedom and color confinement. Asymptotic freedom refers to the weakening of strong coupling at small distances (high momentum transfer). Quarks are surrounded by a ‘cloud’ with virtual gluons and quarks. Because gluons can split into gluon pairs, the color charge of the cloud is preferentially the color of the quark. Thus as the quark is probed at smaller distances, less of the color charge of the virtual particles is seen, eventually leaving only the bare color charge of the quark. Therefore the theory has a small coupling at small distance scales.

The coupling of the strong force becomes large at a scale  $\Lambda_{QCD} \approx 300$  MeV.  $\Lambda_{QCD}$  is approximately the scale where QCD is non-perturbative, because the strong coupling constant  $\alpha_s \rightarrow 1$ . As the force between the colored objects increases with distance, eventually enough potential energy is present to create a  $q\bar{q}$  pair out of the vacuum. This process continues until the quark hadronizes into a color singlet object.

The simplest color singlet is the meson, the pairing of a quark and anti-quark of the same color. The next simplest color singlet is a baryon, which is the combination of three quarks or three anti-quarks, each with a different color. Color confinement is a non-perturbative process and it explains the lack of free quarks in nature; only color singlet objects have been seen.

### 1.2.1 Factorization Theorem

The description of bottom production at proton-antiproton colliders using QCD includes two processes which involve the transfer of soft (low) momentum gluons. Within QCD theory, the proton is a complex multi-body object which consists of three valence quarks and a sea of virtual gluons and quark-antiquark pairs. The exchange of soft gluons within the proton prevents a first principles calculation of the internal structure of the proton in perturbative QCD. Thus, the partons involved in the hard scatter (high momentum) process are not well defined in perturbative QCD. Additionally, color confinement requires the produced bottom quarks be hadronized, which also involves soft interactions that are not calculable in perturbative QCD. Thanks to the Factorization Theorem [3], the short distance parton scattering which produces the bottom quarks is separable from the long distance parton evolution within the proton and the long distance interactions of the partons within the bottom hadrons. Thus, the hard scatter is calculated by perturbative QCD, while the non-perturbative aspects of bottom hadroproduction are determined from empirical model based on experimental measurements.

Another consequence of the factorization theorem is that the distribution of the partons' momentum fraction within the proton and hadronization process are universal for a given quark species and only dependent on the momentum transfer ( $Q^2$ ) involved in the collision. Thus, these non-perturbative effects can be determined in measurements with less complicated experimental environments and/or theoretically precise predictions (such as  $e^-e^+$  colliders, fixed target experiments using hadronic targets and leptonic beams, electroweak boson and high  $p_T$  jet production in hadronic colliders) and then convoluted to the prediction of bottom production.

Schematically, the cross section to produce bottom hadrons C and D from the fragmentation of bottom antiquark  $\gamma$  and bottom quark  $\delta$  from the hard scatter of partons  $\alpha$  and  $\beta$  inside the proton A and antiproton B respectively is

$$\begin{aligned} \sigma(p_A \bar{p}_B \rightarrow B_C \bar{b}_D) = \\ = \Sigma_{\alpha, \beta} f_{\alpha}^A(x_{\alpha}, \mu_F^2) f_{\beta}^B(x_{\beta}, \mu_F^2) \oplus \sigma(\alpha, \beta \rightarrow \bar{b}_{\gamma} b_{\delta}, \mu_R) \oplus F_{\gamma}(z_{\gamma}, \mu_H^2) F_{\delta}(z_{\delta}, \mu_H^2) \end{aligned}$$

where:  $\oplus$  denotes a convolution integral,  $x$  and  $z$  represent the fraction of momentum carried by the partons,  $\hat{\sigma}$  is the cross section of the parton scattering process,  $f$  is

the parton distribution function (PDF) for the proton, and  $F$  is the fragmentation function.

All three components of the QCD prediction depend on the experimentally determined value of  $\Lambda_{QCD}$  used in the calculation, as the value of  $\Lambda_{QCD}$  sets the value of the QCD coupling strength  $\alpha_s$  at energy scale  $Q$ . The quantity  $\mu_F$ , the factorization scale, is an arbitrary parameter. It can be thought of as the scale which separates the soft and the hard processes. Thus a parton emitted with small transverse momentum less than the scale  $\mu_F$  is considered part of the hadron structure and is absorbed into the parton distribution. Processes with transverse momentum larger than  $\mu_F$  are partons of the parton-parton cross section. The scale  $\mu_F$  is typically chosen to be of the order of the hard scale  $Q$ , which characterizes the parton-parton interaction.

In principle, any observable should be invariant under variations of this scale. This is formally expressed as:

$$\mu^2 \frac{d}{d\mu^2} \Gamma = 0$$

where  $\Gamma$  is the observable we are interested in. In the perturbative approach, this equation has to be applied to the perturbative expression of the observable,

$$\Gamma = \Gamma_0 + \alpha_s \Gamma_1 + \alpha_s^2 \Gamma_2 + \dots$$

therefore the equation transforms into

$$\mu^2 \frac{d}{d\mu^2} \sum_{j=0}^N \alpha_s^j \Gamma_j = O(\alpha_s^{N+1})$$

showing that the variation of the observable with the scale is given by terms which are not included in the perturbation expansion. The more terms included in the perturbative expansion, the weaker the dependence of  $\mu$  will be.

Similar to QED, different types of divergences appear in the pQCD calculations. The renormalization is the standard regularization procedure used to solve these divergences [4]. The procedure is not unique, and different renormalization schemes have been used in the literature, depending on the properties of the parameters needed in the calculation. As in the factorization procedure, the observable should be invariant under variations of this scale. However, since the perturbative expansion is performed only to a given order, a theoretical dependence is obtained in pQCD.

## 1.2.2 Parton Distribution Functions

Parton distribution functions describe the longitudinal momentum carried by the various partons in a hadron.  $f_\alpha^A(x_\alpha, \mu_F)$  is the probability distribution for a parton of flavor  $\alpha$  to have a momentum  $x_\alpha P_{proton}$  at energy scale  $\mu_F$ . The PDFs change or

‘evolve’ as a function of the energy scale of the interaction because shorter distances within the proton are probed. As the energy of the probe increases, the effects of the emission of softer gluons from a quark and the splitting of gluons in  $q\bar{q}$  pairs are resolved. Therefore, the PDFs populate lower and lower regions of  $x$  as the factorization scale,  $\mu_F$ , increases. The evolution of the PDFs are determined by a set of evolution equations, first described by Altarelli and Parisi [5], which are solved in perturbative QCD to the same fixed order as the parton cross section. As the PDFs are non-perturbative, the functional form of the PDFs are empirical and must be fit from experimental measurements. As no experiment is sensitive to all partons over the entire  $x$  region, the PDFs have to be determined by a global fit to wide range of experimental data. Two groups which perform such global analyses are the CTEQ [6] and MRST [7] collaborations. Both groups fit a set of PDFs to the following categories of experimental data:

- Deep Inelastic Scattering (DIS) of muons on nucleonic targets ( $\mu N \rightarrow \mu X$ ) at SLAC, FNAL, and CERN.
- DIS of muon neutrinos on nucleonic targets ( $\nu_\mu N \rightarrow \mu X$ ) at FNAL.
- DIS in electron-proton collisions ( $F_2^p$  and  $F_L$  proton form factors) at HERA.
- DIS of proton on nucleonic targets ( $pN \rightarrow \gamma X$  and  $pN \rightarrow \mu^+ \mu^- X$ ) at FNAL and CERN.
- DIS in proton-antiproton collisions (W asymmetry and  $p\bar{p} \rightarrow jet X$ ) at FNAL.

In the analysis performed in this thesis, the CTEQ PDF is used.

### 1.3 Monte Carlo methods

Theoretical predictions form an integral part of any particle physics experiment. They help to design the detectors and to define experimental strategies. They must reproduce as closely as possible the experimental data related to a given analysis. A largely successful way of achieving this goal is through event generator codes, which are used to produce events with the distribution predicted by a given theory <sup>1</sup>.

Any event generator is based on the description of elementary processes (or ‘hard scattering’), for example  $q\bar{q} \rightarrow b\bar{b}$ , which are unfortunately non-physical: quark beams cannot be prepared, and isolated quarks cannot be detected. Initial and final state radiations are expected to play a major role, especially in QCD, and must be added

---

<sup>1</sup>For an unambiguous interpretation of experimental data, other type of codes (cross section integrators) are better suited than event generators. A general description can be found in [8].

to the hard scattering process. This extra radiation taking place on top of the hard scattering subprocess corresponds to considering higher-order corrections in perturbation theory or it can be viewed as a way of describing the ‘dressing’ of a bare quark (‘hadronization’).

Complete perturbative calculations in QCD have been performed only to NLO in most cases. However, higher-order terms cannot be neglected in the case of soft-gluon radiation and collinear configurations. The leading contributions of these soft and collinear topologies, and the corresponding enhanced virtual corrections, can be identified and summed to all orders, improving the convergence of the perturbative series. It is necessary that any event generator includes a way to compute exactly higher order corrections, up to a fixed order as *tree level matrix element* generators do, or to estimate the high order corrections final effects ( this is the strategy implemented in *parton shower* generators).

Parton shower Monte Carlo programs not only include parton showers, allowing partons to split into pairs of other partons, but also a phenomenological model to produce colorless hadrons from the resulting partons. The implementation and modelling of the parton shower and hadronization processes in two of the most used MC programs, PYTHIA [9] and HERWIG [10], are presented below.

### 1.3.1 Parton shower generators

As explained above, the parton shower in the MC serves two main purposes. The first purpose is to provide an estimation of higher-order corrections that are enhanced by large kinematic logarithms. The second purpose is to generate high-multiplicity partonic states which can then be converted into the observed hadrons by a hadronization mechanism.

The Monte Carlo technique models parton shower as a random process: during showering, successive values of a scale  $t$ , a momentum fraction  $z$  and an azimuthal angle  $\phi$  are generated, together with the flavors of the partons emitted. The evolution is based on the Sudakov form factors, which express the probability that a parton does not branch between some initial maximum and minimum scale  $t$ . The branching processes:  $q \rightarrow gq$ ,  $g \rightarrow gg$  and  $q \rightarrow q\bar{q}$  are described by the DGLAP equations [5]. Once the branching has occurred, say  $i \rightarrow jk$  at scale  $t$ , the evolution of the daughter partons  $j$  and  $k$  has to be generated. At the simplest level, their evolution starts at scale  $t_i$  and the next values  $t_k$  and  $t_j$  are obtained using the appropriate Sudakov form factors. Usually,  $t$  is proportional to the virtuality  $Q^2$ , thus the virtualities of daughter particles are constrained by the kinematic relation  $\sqrt{t_j} + \sqrt{t_k} < \sqrt{t_i}$ , where  $t_i = E_i^2(1 - \cos\theta_i)$  being  $\theta_i$  the opening angle in the branching  $i \rightarrow jk$ , and where an angular ordering is imposed. This means that the opening angle  $\theta_j$  of any subsequent branching of parton  $j$  is less than  $\theta$ , a property referred to as ‘angular ordering’.

The final outcome of successive branchings is a parton shower in which each initial parton from the hard process evolves into a jet of partons moving roughly in the same direction. The typical scale of relative transverse momenta between partons at the end of the shower is set by the cutoff  $t_0$ ; beyond that pQCD cannot be applied. Quarks, anti-quarks, and gluons produced by the showering process are not allowed to exist in isolation, as dictated by color confinement. Thus, the next step in the MC programs is to group these colored partons into the observed colorless hadrons using a phenomenological model referred to as hadronization.

### 1.3.2 Hadronization

The principle of color confinement states that colored objects must bind with each other to produce a colorless object, (hadronize). Such hadronization processes involve soft gluons which have a typical scale of the size of hadrons,  $Q \sim 1/R_{had} \sim 300 MeV$ . Due to the low  $Q^2$  of the process and the large value of  $\alpha_s$ , hadronization is not described well by perturbation theory and has to be described by an empirical ansatz based on kinematical arguments that is tuned to experimental data.

The function describing hadronization (for example of the  $b$  quark) is called fragmentation function, and it is parameterized as a function of the fraction of the bottom quark's momentum carried by the bottom hadron  $z = \frac{E(B)+p_{\parallel}(B)}{E_b+p_b}$  where:  $E(B)$  and  $p_{\parallel}(B)$  are the bottom hadron's energy and momentum parallel to the bottom quark direction and  $E_b$  and  $p_b$  are the bottom quark's energy and momentum.

One general approach to hadronization, based on the observation that perturbation theory seems to work well down to rather low scales, is the hypothesis of local parton hadron duality, where one supposes only that the flow of momentum and quantum numbers at the hadron level tends to follow the flow established at the parton level. Hence, the flavor of the quark initiating a jet should be found in a hadron near the jet axis. Two of the most commonly used models are the cluster model [11] and the string fragmentation (Lund) model [12].

#### The string model

The assumption of linear confinement provides the starting point for the string model [13]. As the  $q$  and  $\bar{q}$  partons move apart from their common production vertex, the physical picture is that of a color flux tube being stretched between the  $q$  and the  $\bar{q}$ . If the tube is assumed to be uniform along its length, this automatically leads to a confinement picture with a linearly rising potential  $\sim kl$  where  $k \approx 1 \text{ GeV}/fm$  and  $l$  is the separation between color charges. As the  $q$  and  $\bar{q}$  move apart, the potential energy stored in the string increases, and the string may break by the production of a new  $q\bar{q}$  pair, so that the system splits into two color-singlet systems  $q\bar{q}$  and  $q\bar{q}$ . The

probability of breaking the string is given by the quantum mechanical probability of tunneling through a potential barrier.

$$\frac{d^2 P}{dxdt} = \exp\left(-\frac{\pi m^2}{k}\right) \cdot \exp\left(-\frac{\pi p_T^2}{k}\right)$$

Due to the mass term in the tunneling potential, fragmentation to heavy quark pairs are suppressed  $u : d : s : c \approx 1 : 1 : 0.3 : 10^{-11}$ . In addition, the cutting of the string with a  $q\bar{q}$  pair guarantees momentum and energy conservation. The  $p_T$  of the quark formed in the tunneling is compensated by the antiquark in the pair. The fragmentation continues until all strings end with quarks and the quark-string-antiquark systems all have a mass below a given cut-off.

If the invariant mass of either of these string pieces is large enough, further breaks may occur. In the string model, the string break-up process is assumed to proceed until only on-mass-shell hadrons remain, each hadron corresponding to a small piece of string with a quark on one end and an antiquark in the other. Charm and heavier quarks hence are not expected to be produced in the soft fragmentation, but only in perturbative parton-shower branchings  $gq\bar{q}$ . Baryon production is still a poorly understood area. In the simplest possible approach, a di-quark in a color anti-triplet state is just treated like an ordinary antiquark, such that a string can break either by quark-antiquark or antidiquark-diquark pair production. If several partons are moving apart from a common origin, the details of the string drawing become more complicated. A schematic picture of the production of a multi-hadronic final state according to the string model is shown in figure 1.2 (left). Notice that whenever a gluon splits perturbatively into a quark-antiquark pair during the evolution of the parton shower, an additional string segmentation is produced. On the other hand, gluons which remain at the end of the shower lead to kinks in the string segment which connect them. The Monte Carlo code PYTHIA uses the string fragmentation model.

## Cluster Model

In the independent fragmentation model, the hadronization is calculated as an incoherent sum of independent fragmentation processes for each of the partons. It is based on the color pre-confinement property of pQCD [1]. At the end of the parton shower evolution, the remaining gluons are split non-perturbatively into  $q\bar{q}$  pairs. Neighboring  $q$  and  $\bar{q}$  can be then combined into a singlet cluster with a typical mass of 1–2 GeV. These clusters decay directly into two hadrons unless they are too heavy, (then they decay to two clusters), or too light in which case the cluster decays to a single hadron. Experimentally this model describes quite well the hadronic energy and transverse momentum distribution for final states. Figure 1.2 (right) shows a

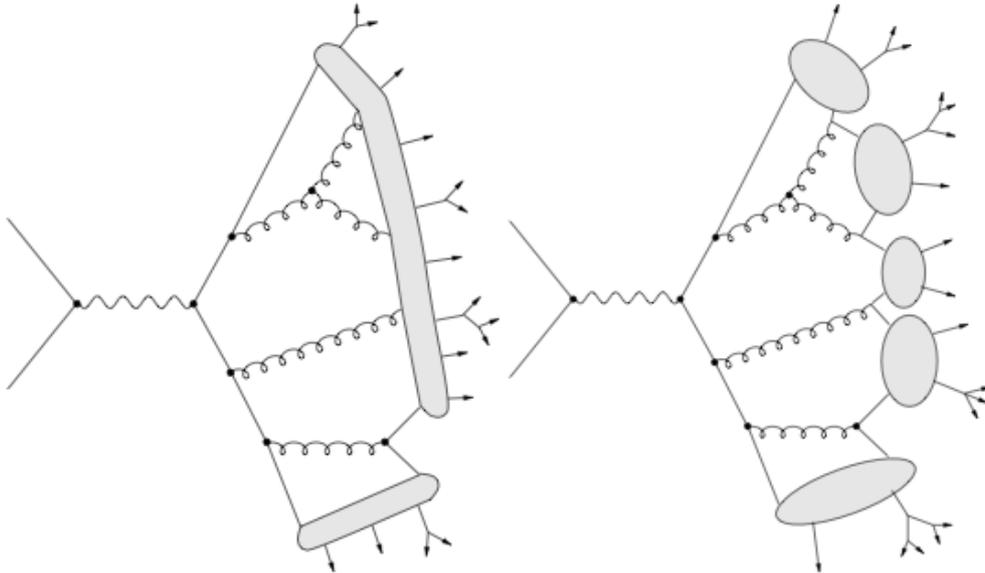


Figure 1.2: Parton shower with string (left) and cluster (right) hadronization models.

schematic picture of the cluster hadronization model. This model has the advantage of being easy to implement, but has some weaknesses. Flavor, energy, and momentum conservation are not guaranteed, as each parton is fragmented independently. In addition, the fragmentation procedure is not Lorentz invariant; hadronization depends on the frame in which the fragmentation is calculated. Momentum and energy conservation can be re-established by various schemes of rescaling the momentum and energy of the generated particles. The Lorentz invariance problem can be circumvented by choosing a convention such that the fragmentation process must always be made in the center-of-mass frame.

### Fragmentation functions

Many fragmentation functions exist which differ according to kinematic arguments used to derive them. Two commonly used fragmentation functions are by Peterson et al. [14] and by Bowler [15]. The fragmentation functions are assumed to be universal, that is to have no dependence on the incoming particle. Thus, fragmentation functions are tuned using measurements of  $e^+e^-$  collisions at the  $Z$  pole [16] where the measurement of the fraction  $x$  is best determined. The fragmentation functions evolve with the scale of the fragmentation ( $\mu_H$ ) in a manner similar to the PDFs, with a set of equations, similar to the Altarelli-Parisi ones. These functions should be calculated to the same fixed order as the PDF and parton cross section. The

quantity  $\mu_H$  is typically set to the same value as the renormalization/factorization scale. The Peterson fragmentation function assumes that the energy lost by the heavy quark due to the light antiquark is small. The transition amplitude is determined by the energy difference between the incoming partons and the outgoing hadron. The function has one free parameter  $\epsilon_B$  which is determined experimentally. The Peterson fragmentation function is

$$F(z) \propto \frac{1}{z(1 - \frac{1}{z} - \frac{\epsilon_B}{1-z})^2}$$

All of the previous bottom cross section measurements at the Tevatron and  $Spp\bar{p}S$  colliders have used the Peterson fragmentation functions with independent fragmentation.

### 1.3.3 Matrix Element generators and the problem of combination with showering

Matrix element generators allow the computation of tree-level matrix elements with a fixed number of legs. These programs generally do not include any form of hadronization, thus the final states consist of bare quarks and gluons. The kinematics of all hard objects in the event are explicitly represented and it is simply assumed that there is a one to one correspondence between hard partons and jets.

However this assumption may cause problems when interfacing these codes to showering and hadronization programs such as PYTHIA or HERWIG. An example of the ‘combination problem’ is a process like  $q\bar{q} \rightarrow Z^0$ , with its higher order correction  $q\bar{q} \rightarrow Z^0 g$ . Events from these two processes should not be blindly combined, since a fraction of the latter events are already included in the former process via gluon radiation in the parton shower. Combining the two processes without special procedures results in double counting some portion of the phase space. Different approaches have been developed to solve this problem [8].

The MC@NLO event generator [17] includes the full NLO order QCD corrections in the computation of hard subprocesses and it uses the subtraction method in order to avoid double counting when combining to a parton shower. The basic idea [18] is to modify the subtraction to take into account the terms that are generated by the parton shower. This results in a set of weighted LO and NLO parton configurations that can be fed into the parton showering generator without the problem of double counting. Each weight distribution is well-behaved in the sense that it has no divergences or tails that would lead to Monte Carlo inefficiency. However, in order to reproduce the NLO corrections fully, some of the configurations have negative weights. Event unweighting can still be achieved successfully by generating a small fraction of ‘counter-events’ that contribute with equal but opposite weight to the events in all distributions.

## 1.4 Bottom production cross section at Next-to-Leading Order

Unlike the light quark cross sections, the bottom quark cross section can be calculated reliably at fixed order in perturbative QCD as  $p_T^b \rightarrow 0$ . The bottom mass acts as a effective low momentum cut-off in the calculation. As  $m_B \gg \Lambda_{QCD}$ , the strong coupling  $\alpha_s$  is small ( $\alpha_s(m_B) \approx 0.24$ ) and therefore perturbative QCD should work well. Predictions to order  $\alpha_s^3$  have been calculated [19]. Such calculations include the following subprocesses:

$$\begin{aligned}
 g + g &\rightarrow b + \bar{b} \\
 q + \bar{q} &\rightarrow b + \bar{b} \\
 g + g &\rightarrow b + \bar{b} + g \\
 g + q &\rightarrow b + \bar{b} + q \\
 g + \bar{q} &\rightarrow b + \bar{b} + \bar{q} \\
 q + \bar{q} &\rightarrow b + \bar{b} + g
 \end{aligned}$$

Figure 1.3 gives an example of the some the Feynman diagrams used in the NLO calculation. Virtual diagrams of  $O(\alpha_s^4)$  interfere with the  $O(\alpha_s^2)$  terms. The NLO order contributions to the cross section can be sizable relative to the LO predictions. The cross section also depends on the renormalization scale ( $\mu_R$ ) used to evaluate the value of  $\alpha_s$ . The scale used is typically of order  $\sqrt{m_b^2 + (p_{T_b}^2 + p_{T_{\bar{b}}}^2)}/2$  which minimizes large logarithmic uncertainties at high  $p_T$ . Nason, Dawson, and Ellis [19] first calculated the NLO inclusive single bottom differential cross section  $\frac{d^2\sigma}{dydp_T}$ . In the calculation, the kinematical variables of the  $\bar{b}$  quark are integrated, and therefore correlations between the bottom quarks cannot be calculated.

The NLO prediction of the full kinematics of bottom pair production has not yet been calculated in closed form. Mangano, Nason, and Ridolfi [20] have produced a fully exclusive parton cross section for heavy quark hadroproduction using numerical integration techniques. Soft and collinear divergences in the calculation are handled with careful organization of the integrals and the inclusion of large negative counter-terms. The calculation of the differential spectra is not predictive in the usual areas in phase space for  $O(\alpha_s^3)$  calculations, when the radiated gluon is collinear with either of the bottom quarks or as the radiated gluon's momenta approaches zero. This condition occurs when  $p_T^{b\bar{b}} \rightarrow 0$ ,  $\Delta\phi^{b\bar{b}} \rightarrow \pi$  and  $R > \pi$ . In such regions, negative differential cross section are encountered due to the large negative terms originating from the virtual diagrams and the collinear subtractions. The bins in the differential

cross section should be widened until the predicted cross section is stable, i.e. the shape of the distribution has a fairly smooth second derivative. Positive-only differential cross sections can only be guaranteed with the inclusion of an arbitrary number of soft gluon emissions [21] (summation of the leading Sudakov logarithms [22]).

### 1.4.1 Production of $b$ quarks at the Tevatron

The dominant  $b$  quark production mechanism at the Tevatron is believed to be pair production via strong interaction. A parton from both a proton and a antiproton hard scatter, producing a bottom quark-antiquark pair. In perturbative QCD, initial states with 2 gluons dominate the production cross section at low momenta. Figure 1.3 shows representative Feynman diagrams which contribute to the NLO QCD calculation.

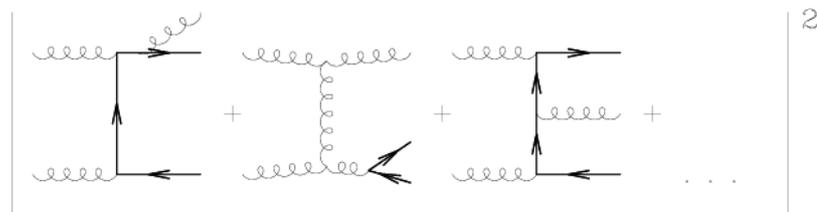
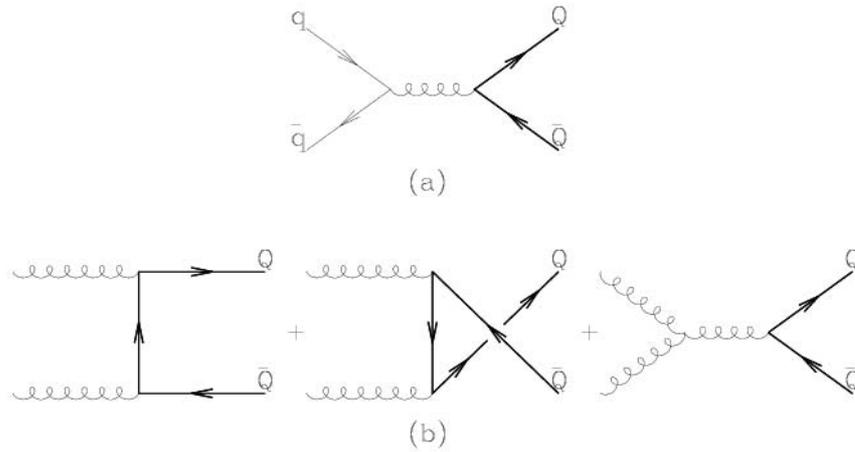
Bottom quark production is an interesting test of QCD because decay topologies, first introduced in NLO order calculations, can have cross sections of the same importance as the LO terms at the Tevatron, as shown by the following simple argument. The  $g + g \rightarrow g + g$  cross section is about a factor of a hundred larger than the  $g + g \rightarrow b + \bar{b}$  cross section. As the rate of gluon splitting to  $b$  quarks ( $g \rightarrow b + \bar{b}$ ) goes as  $\approx \alpha_s$ , a relatively large cross section for such terms is possible. Of course, the cross section is suppressed by the virtuality of the gluon required due to bottom quark's mass. At LHC center-of-mass energies, these terms are predicted to be dominant bottom production terms.

In LO QCD, only  $g + g \rightarrow b + \bar{b}$  and  $q + \bar{q} \rightarrow b + \bar{b}$  processes are included in the calculation and the bottom quarks are always produced back-to-back in the azimuthal angle ( $\Delta\phi^{b\bar{b}} \approx \pi$ ).

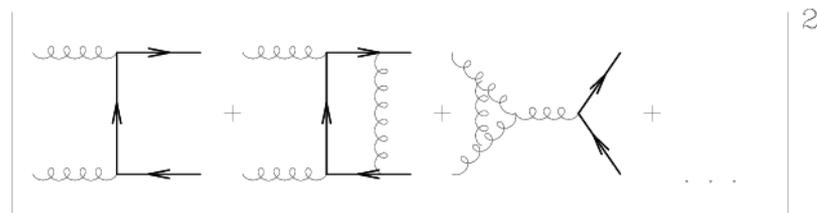
In the NLO calculation, the terms have traditionally grouped into three categories: flavor creation, flavor excitation and gluon splitting.

In perturbation theory, the three categories are not independent, due to interference terms between them. Flavor creation, flavor excitation and gluon splitting are still useful concepts in describing bottom hadroproduction, as they have minimal overlap in phase space. At NLO, flavor creation consists of the  $2 \rightarrow 2$  processes, in addition to diagrams which add a gluon radiation to the  $2 \rightarrow 2$  terms. Flavor excitation includes diagrams in which a initial state gluon splits into a  $b\bar{b}$  pair before interacting with the parton from the other hadron, putting the bottom quarks on-shell. Gluon splitting consists of diagrams where a gluon splits into a  $b\bar{b}$  pair after interacting with the parton from the other hadron (figure 1.3). Due to the new three body final states included in NLO calculations ( $b\bar{b}g, b\bar{b}q$  and  $b\bar{b}\bar{q}$ ), the predicted  $\Delta\phi^{b\bar{b}}$  spectrum is non-zero over the whole range of possible values, but still peaks back-to-back.

In Leading-Log (LL) showering Monte Carlos (such as PYTHIA and HERWIG), the



Real emission diagrams



Virtual emission diagrams

Figure 1.3: Feynman diagrams for the production of a  $Q\bar{Q}$  final state. LO diagrams are shown in figures *a* and *b*. Bottom diagrams represent corrections due to real or virtual emissions.

three categories are generated separately and then added together for the prediction. Since interference is not included, the predictions may include some double counting. At the heart of these generators is a leading-order matrix element calculation. The incoming and outgoing partons are then allowed to radiate using analytical algorithms that are tuned to experimental measurements. The resulting final partons are hadronized using models described in section 1.3.2. The fragmentation and initial state radiation algorithms yields a predicted  $\Delta\phi^{b\bar{b}}$  spectra similar to the next-to-leading order calculation.

## 1.5 Past Experimental $b$ quark cross section measurements at $p\bar{p}$ colliders

The first measurements of the bottom production cross section at a hadron collider were performed by the UA1 collaboration [23]. The analyses used semi-leptonic bottom decays to muons ( $J/\psi \rightarrow \mu\mu$ , di-muons, muon+jet final states) to measure the integrated bottom quark cross section with  $p_T^b > p_T^{min}$  at  $\sqrt{s} = 630$  GeV.

Although the measurements showed a slight excess with respect to the NLO QCD prediction, they were considered consistent with theory within the prediction uncertainties as shown in figure 1.4.

The bottom cross section was studied by the DØ [24] and CDF [25] collaborations at the Tevatron with  $\sqrt{s} = 1800$  GeV. Both measurements used semi-leptonic decays and showed a factor 2-4 excess in the measured cross section with respect to theory, but with a shape consistent with the theoretical predictions. Figure 1.4 shows the integrate bottom quark cross section with  $p_T^b > p_T^{min}$ , measured by the CDF collaboration.

The publication of these results in Run I led to many developments both in theoretical calculation beyond NLO and the experimental approach, resulting in a better agreement between data and theory. A major theoretical improvement was the implementation of the so called Fixed-Order with Next to Leading Log (FONLL) calculation [27], where the re-summation of  $\log(p_T/m_b)$  terms with the next to leading logarithmic accuracy is matched with a fixed order NLO calculation. Also, there have been substantial changes in the bottom fragmentation function as extracted from experimental data and for the PDF functions (now CTEQ6M) used.

A measurement of the B-hadron production cross section [28] has been made by CDF, using Run II data corresponding to a total  $39.7 pb^{-1}$ . The differential cross section of the B-hadrons over the transverse momentum range from 0 to 25 GeV/c is shown in figure 1.5. The superimposed theoretical calculations show a remarkable agreement with data.

More recently CDF has measured the inclusive  $b$ -jet cross section using about 300

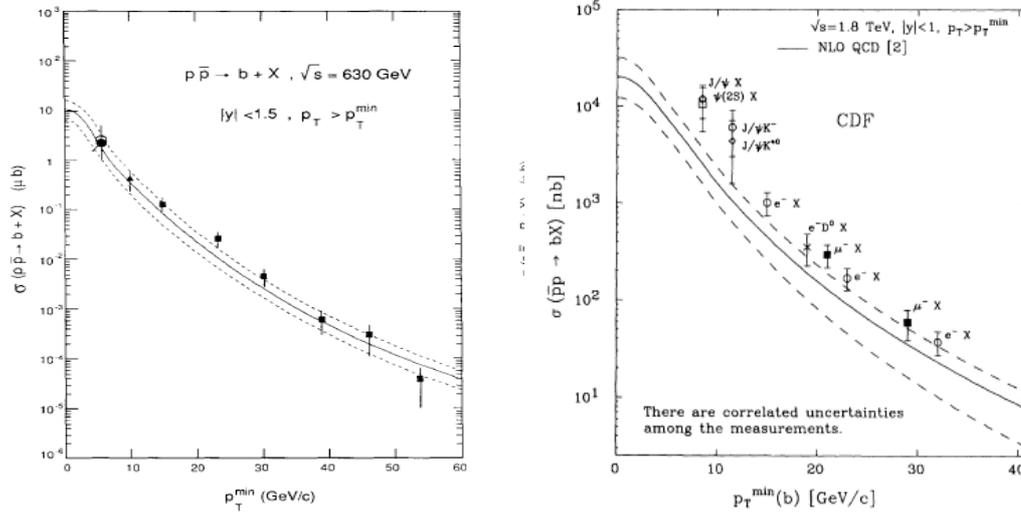


Figure 1.4: UA1 measurement [23] (left) of the  $b$  quark integrated cross section. (right) Integrated  $b$  quark cross section measured by the CDF collaboration [26].

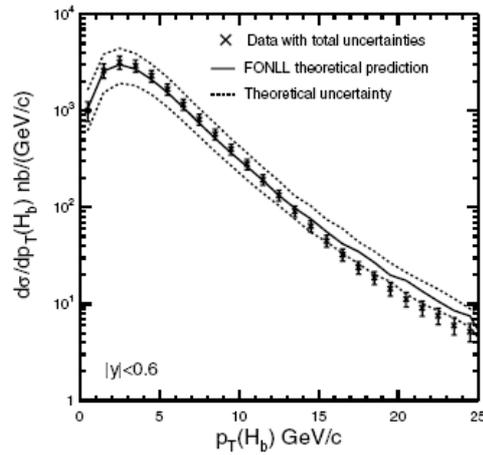


Figure 1.5: Differential cross-section distribution of  $b$ -hadron production. Crosses with error bars are the data with systematic and statistical uncertainties. The solid line is the central theoretical values using FONLL calculation.

$pb^{-1}$  of data [29]. The use of jets not only extends the transverse momentum range available to exclusive decays measurement, it is also less sensitive to fragmentation and hadronization effects. Figure 1.6 shows the  $b$ -jet cross section for jets in the  $p_T$  range from 38 GeV/ $c$  up to 400 GeV/ $c$  and the ratio data to NLO prediction: good consistency is found for jets below 90 GeV/ $c$ ; for jets above 90 GeV/ $c$  agreement is observed within the systematic uncertainties. In particular the big uncertainty related to renormalization and factorization scale suggest that non-included higher order contributions might play a major role in  $b$ -jet production. For instance, in a NLO calculation, gluon splitting only appears at leading order: while at low  $p_T$  effects are expected to be small, logarithmic enhancements of the higher order contribution due to this process could be very important at high  $p_T$ .

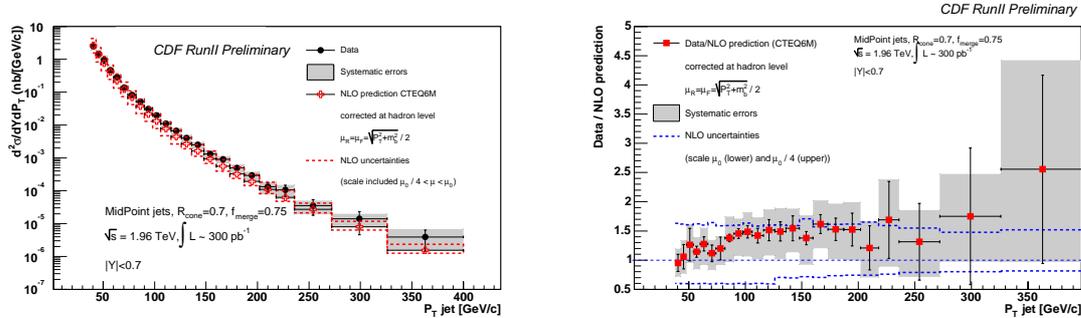


Figure 1.6: Inclusive  $b$ -jet differential cross section measured by CDF II [29], compared to NLO predictions (left). Data/NLO prediction ratio (right). The systematic error on the theoretical curve refers to PDF, jet cone radius and scale uncertainties.

## 1.6 Theoretical motivation for $b\bar{b}$ angular production correlation measurements

As stated in the previous section, the single differential cross section measurements (in  $p_T$  and rapidity) only partially agree with the predicted shape from NLO QCD. The disagreement may indicate the importance of higher order corrections or non-perturbative fragmentation of gluons into bottom quark pairs [30]. The arguments for non-perturbative fragmentation effects are strengthened by the central values of the gluon splitting rates to bottom quark measured by the LEP experiments and SLD [31], which are higher than the NLO predictions [32]. The errors on both theory predictions and experimental measurements are large enough to explain the differences between the predictions and the measurements.

In order to better understand the bottom quark production mechanisms, it was proposed in ref. [33], [34] and [35] to measure correlations between the bottom quarks ( $\Delta p_T$ ,  $\Delta\phi$ ,  $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta y)^2}$ ). Angular correlations are easier to measure than  $p_T$  correlations, because bottom decays are not required to be fully reconstructed. In LO QCD, the bottom quarks are produced back-to-back, while  $2 \rightarrow 3$  terms that first appear in NLO QCD allow the bottom quarks to be produced with any angular relationship. Thus, a low  $\Delta\phi$  or  $\Delta R$  measurement should be able to discern the effects of higher order perturbative or non-perturbative terms in bottom production.

For leading-log showering Monte Carlo, angular correlations are able to distinguish between flavor creation, flavor excitation and gluon splitting (fragmentation) terms. Figure 1.7 shows the PYTHIA predictions of  $\Delta\phi$  and  $\Delta R$  in [36].

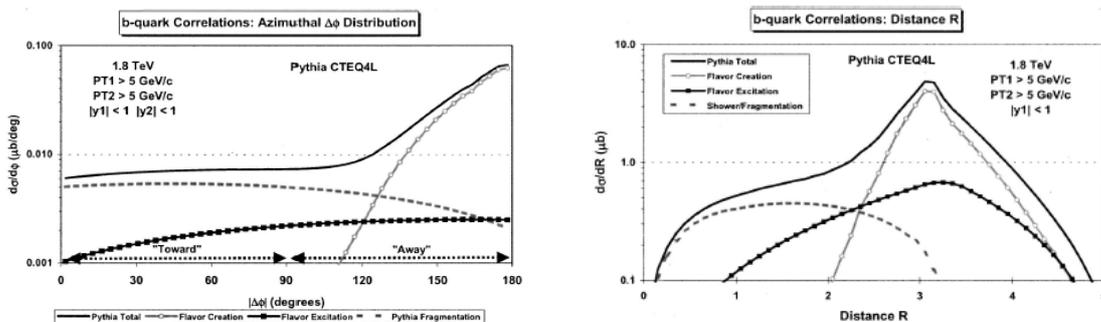


Figure 1.7: PYTHIA 6.158 distribution for the azimuthal angle  $\Delta\phi$  (left) and the  $\Delta R$  difference (right), between a  $b$  quark with  $p_{T1} > 5$  GeV/ $c$  and  $|y_1| < 1$  and  $\bar{b}$  quark with  $p_{T2} > 5$  GeV/ $c$ . The prediction are obtained at the Tevatron Run I center-of-mass energy  $\sqrt{s} = 1.8$  TeV.

A measurement can be used to tune the relative rates of the three mechanisms, which are relatively uncertain in showering Monte Carlos.

### 1.6.1 Past Experimental $b\bar{b}$ angular production correlation measurements at $p\bar{p}$ colliders

The first measurement of  $b$ -quark angular correlations at a hadronic collider was performed by the UA1 collaboration [37] in 1994 at the  $Spp\bar{p}$  collider with  $\sqrt{s} = 630$  GeV. The measurements used the di-muon decay signature, in which both bottom quarks decay semi-leptonically. In order to minimize  $J/\psi$ , double sequential decay muons ( $b \rightarrow c\mu X$  and  $c \rightarrow \mu X'$ ) and Z muons, a di-muon mass cut of  $4 < M^{\mu\mu} < 35$  GeV was applied. This mass cut also minimizes the acceptance of collinear bottom quark pairs, biasing the  $\Delta\phi^{b\bar{b}}$  distribution measured. The analysis corrected

for varying acceptance versus  $\Delta\phi^{b\bar{b}}$ . The  $\Delta\phi$  shape was consistent with theoretical prediction, but the prediction was 30-40% lower than the measurements. The fraction of the total bottom cross section for non-perturbative gluon fragmentation into  $b\bar{b}$  pairs was measured to be  $\sigma_{non-pert}(\Delta R(b\bar{b} < 1.6))/\sigma_{all} < 11\% @ 90\% \text{ c.l.}$

Similar measurements were carried out at the Tevatron with  $\sqrt{s} = 1800 \text{ GeV}$ . Both the DØ [38] and CDF [39] collaborations performed  $\Delta\phi$  measurements in the di-muon channel in the same manner as UA1. CDF's measurement required a di-muon mass of  $M^{\mu\mu} > 5 \text{ GeV}$ , whereas DØ's analysis required a di-muon mass of  $6 < M^{\mu\mu} < 35 \text{ GeV}$ . Unlike the UA1 analysis, both CDF and DØ corrected the theory for the  $\Delta\phi$  bias due to the di-muon mass requirement, instead of correcting the data for the acceptance. The  $\Delta\phi$  shape in both cases is consistent with the NLO QCD predictions, but both the analyses measured a factor 2-3 excess in data relative to the theory predictions.

At CDF,  $\Delta\phi^{\mu b}$  was measured between a muon (presumably from a bottom quark decay) and a bottom quark jet identified using a jet probability algorithm (jetprob) [40]. The jet probability algorithm uses the impact parameters of particles in a jet with a cone size  $R = 0.4$  radius; it calculates the probability of a jet originating from the primary vertex. The jet and the muon were required to be separated by at least 1 unit of  $\eta - \phi$  space, which again lead to a large non-uniformity in the acceptance versus  $\Delta\phi$ . The measurement showed a slight  $\Delta\phi$  shape disagreement and a factor of  $\approx 2$  excess relative to the NLO predictions.

Finally a rapidity correlation measurement was performed by CDF measuring the ratio of a bottom quark being produced with  $2.0 < |y_1| < 2.6$  to  $|y_1| < 0.6$  when the second bottom quark is produced with  $|y_2| < 1.5$  [41]. The first bottom quark was identified with a semi-leptonic decay muon and the second bottom quark was identified by a displaced vertex. The purity of the sample was determined by fitting the transverse momenta of the jet associated with the muon relative to the muon ( $p_T^{rel}$ ) and the pseudo decay length of the displaced vertex. The  $\Delta\phi$  between the muon jet and the displaced vertex was required to be greater than  $\pi/3$  in order to remove the contribution from gluon splitting. The ratios measured were consistent with theory.

The results of the measurements are summarized in table 1.4 and table 1.5. The approximate  $f_{toward}$ , the fraction of  $b\bar{b}$  pairs produced with  $\Delta\phi^{b\bar{b}} < \pi/2$ , for both the measurements and the theory predictions are shown, along with any requirements which yield a non-uniform efficiency versus  $\Delta\phi^{b\bar{b}}$ . The typical theory prediction of  $f_{forward}$  ranges between 16-19% if no  $\Delta\phi$  or  $M^{\mu\mu}$  cuts are applied.

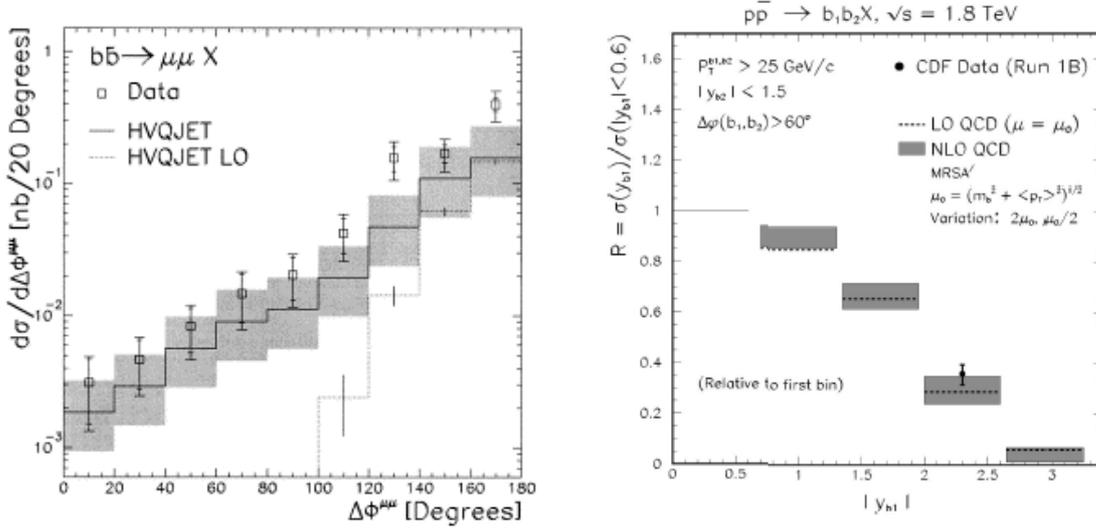


Figure 1.8: Past angular correlation measurements of bottom production at the Tevatron. D0  $\Delta\phi$  measurement using  $\mu, \mu$  (left). CDF  $\Delta R$  measurement using  $\mu, \text{jet}$  using SecVtx (right).

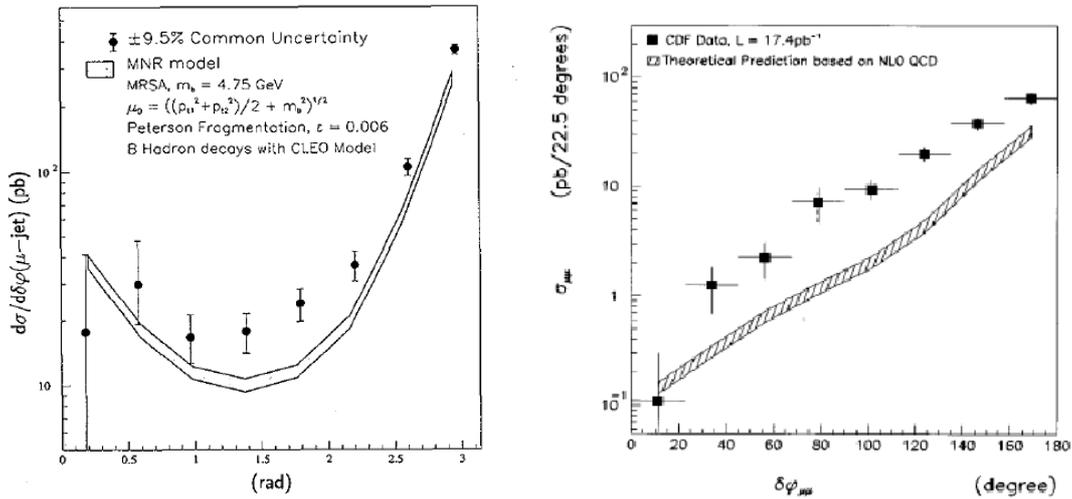


Figure 1.9: Past angular correlation measurements of bottom production at the Tevatron. CDF  $\Delta\phi$  measurement using  $\mu, \mu$  (right). CDF  $\Delta\phi$  measurement using  $\mu, \text{jet}$  using JetProb (left).

Measurement	$b_{pT}^1$ (GeV)	$b_{pT}^2$ (GeV)	$b_y^1$	$b_y^2$
UA1 $\Delta\phi^{\mu,\mu}$	6.0	6.0	2.3	2.3
CDF $\Delta\phi^{\mu,\mu}$	6.5	6.5	0.67	0.67
DØ $\Delta\phi^{\mu,\mu}$	8	8	1.0	1.0
CDF $\Delta\phi^{\mu,jet}$	15	20.7	0.67	1.5
CDF $\Delta R^{\mu,jet}$	25	25	0.6	1.5

Table 1.4: Bottom angular correlation measurement quantities for previous analyses. Top: approximate bottom kinematics ( $p_{Tb}$ ,  $y_b$ ) of the measurements due to selection criteria.

Measurement	$f_{forward}^{exp}$ (%)	$f_{forward}^{theo}$ (%)
UA1 $\Delta\phi^{\mu,\mu}$	18.5	16.6
CDF $\Delta\phi^{\mu,\mu}$	7.7	4.4
DØ $\Delta\phi^{\mu,\mu}$	5.1	7.0
CDF $\Delta\phi^{\mu,jet}$	13.4	18.5
CDF $\Delta R^{\mu,jet}$	N/A	18.5

Table 1.5: Fraction of  $b\bar{b}$  pairs measured ( $f_{forward}^{exp}$ ) and predicted ( $f_{forward}^{theo}$ ) in the same hemisphere in the azimuthal angle,  $\Delta\phi < \pi/2$ .

# Chapter 2

## The Tevatron and the CDF detector

### 2.1 FNAL - the Fermi National Accelerator Laboratory

The Fermi National Accelerator Laboratory ('Fermilab') is located about 35 miles west of Chicago, Illinois. Originally named the National Accelerator Laboratory, it was commissioned by the U.S atomic energy in 1967, and in 1974 it was renamed in honor of the Italian Nobel Prize winner Enrico Fermi, one of the greatest physicists of the atomic age.

Since then several experiments at Fermilab have made important contributions to the understanding of the Standard Model. In 1977, the Fermilab experiment E288 observed a new particle, the Upsilon, composed of a new quark-antiquark couple, giving the first evidence of the existence of the bottom quark. In 1995 the CDF and  $D\bar{0}$  experiments completed the quark sector of the Standard Model with the observation of the top quark.

The laboratory is currently the home of a large diversity of projects mostly related to high energy physics, including CDF,  $D\bar{0}$ , the Dark Energy Survey, the Pierre Auger Observatory, the Sloan Digital Sky Survey, MINERvA, MiniBooNE, MINOS, NOvA and NuTeV.

The Run II of the Tevatron started in March 2001, designed to meet the goals of the new particle physics frontiers: it involved a complete upgrade of the full accelerator complex.

In this chapter we briefly describe the Tevatron and the CDF experimental apparatus.

## 2.2 The Fermilab accelerator complex

The accelerator complex at Fermilab consists of several components, that can be conceptually separated into a series of accelerators preparing the protons, producing and storing anti-protons and finally accelerating both protons and anti-protons to  $\sqrt{s} = 1.96$  TeV to make the collisions. The schematic view of the accelerator chain is shown in Figure 2.1.

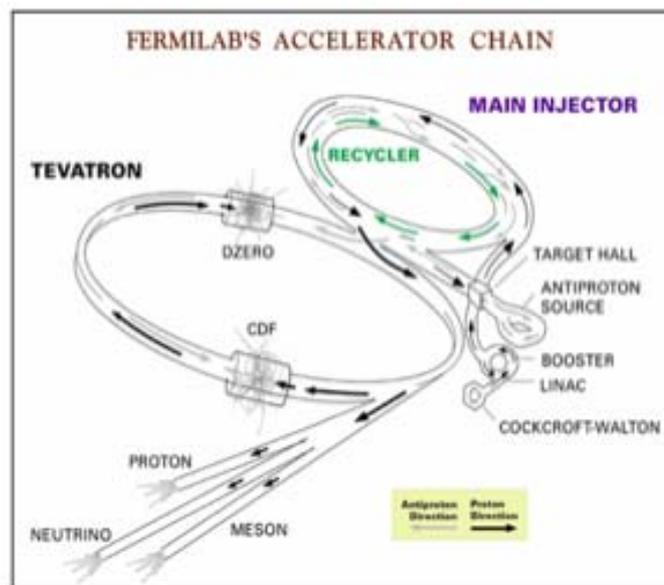


Figure 2.1: A schematic drawing of the Fermilab accelerator complex. The diagram shows the paths taken by protons and anti-protons at the Fermilab's five accelerators. Negative hydrogen ions move from the Cockcroft-Walton accelerator to the Linac. Electrons are stripped off when the ions enter the circular Booster, and remaining protons are injected into the Tevatron. In fixed target mode, the proton beam is extracted and sent down the Fixed Target beam-line to the experimental areas. In colliding beam mode, anti-protons are collected in the Antiproton Storage Rings and injected into the Main Ring traveling in an opposite direction with respect to the protons. The protons and anti-protons collide inside the two detectors - CDF and Dzero (DØ).

### 2.2.1 Proton source

The pre-accelerator is a linear accelerator that produces negative hydrogen ions and accelerates them to 750 keV by applying the electric field to the ionized hydrogen.

The output frequency is 15 Hz and the resulting  $H^-$  ions then enter the Linac or Linear Accelerator [42]. The Linac is the next step in the acceleration process. It accelerates negatively charged hydrogen ions from 750 keV to 400 MeV. The Linac operates with a frequency of 15 Hz. The modes of operation include feeding the beam to the Booster, feeding the beam to the Nuclear Therapy Facility or dumping the beam into a concrete block.

The Booster [43] is the first circular accelerator in the proton accelerator chain. It has a radius of 75 meters and consists of alternating magnets and RF cavities. The Booster strips electrons of the  $H^-$  ions and accelerates the protons from 400 MeV to 8 GeV with a period of 0.033 second. The RF cavities apply the accelerating field while the magnets apply the bending field to keep the protons in the circular orbit, and ramping it up in accordance with the instantaneous energy of the beam. The proton beams travel around the Booster about 20,000 times before they undergo the next acceleration step - acceleration in the main injector.

The Main Injector (MI) [43] is a circular accelerator that serves several purposes. One of its functions is to accelerate the protons transported from the Booster from 8 GeV to 150 GeV. In another mode of operation the MI stacks the anti-protons and accelerates them to 120 GeV. The circumference of this machine is seven times that of the Booster. The ramp up process takes about 1.5 seconds.

## 2.2.2 Antiproton source

The production of anti-protons is a technologically challenging task and thus the anti-protons production and storage capacity is the main limiting factor for the luminosity of the colliding beams. The anti-protons are produced at the Target station when the 120 GeV proton beam from the Main Injector hits a Ni target. In this process different particles are being produced. A system of magnets is used to separate 8 GeV anti-protons and direct them to the next stage accelerator. The purpose of the Debuncher, a triangular synchrotron <sup>1</sup> machine, where anti-protons are directed after being produced, is not to accelerate them but rather to make the momentum of the particles inside the beam more uniform - this process is normally referred to as "cooling" <sup>2</sup>. The mean radius of the Debuncher is 90 meters. The resulting 8 GeV beam of anti-protons is then delivered into the Accumulator. The Accumulator is located in the same tunnel as the Debuncher and is also a triangular synchrotron. It is used for storage and further cooling of the anti-protons.

The Recycler is a relatively recent addition to the Fermilab accelerator chain.

---

<sup>1</sup>A circular accelerator in which both magnetic field and accelerating electric field are varied synchronously as the particles are accelerated, in order to maintain the same particle orbit.

<sup>2</sup>Another way of looking at it is to consider cooling as a process of increasing particle density in momentum phase-space.

It resides in the same tunnel as the Main Injector and is used to decelerate and cool down the anti-protons left at the TEVATRON after the store. Several different cooling techniques are used in the Recycler which serves an important purpose of improving the antiproton utilization efficiency.

### 2.2.3 TEVATRON

The Tevatron [43] is the final stage in the acceleration process. It receives 150 GeV (anti)protons from the Main Injector, and then accelerates them to the final energy of 980 GeV. The Tevatron ring is an underground circular beam tube with a radius of one kilometer buried about 6 meters deep. Its total circumference is about 6.3 km and it holds 816 di-polar superconducting magnets. These magnets are kept at a temperature of 4.3K by what is one of the largest cryogenic systems in the world. At 980 GeV energy the magnetic field of the di-polar magnets is 4.2 Tesla, the current draw of the coils being 4000 A.

In addition to the dipole magnets there are 204 quadrupole pairs that focus the beam to achieve peak luminosity.

The RF system of the Tevatron consist of an array of 8 RF cavities running at a frequency of 53.03 MHz. This frequency does not need to be changed during the ramping since the very small velocity difference of the protons at 150 GeV and 980 GeV. It takes 9 cycles to fill the Tevatron with 36 bunches of protons (or anti-protons), and takes about 40 seconds to ramp the energy from the 150 GeV to 980 GeV. The typical number of protons (anti-protons) in a bunch is about  $27 \cdot 10^{10}$  ( $33 \cdot 10^9$ ).

During Run II the Tevatron is operated in 36x36 mode, which refers to the number of bunches of protons and anti-protons respectively. The bunch crossing occurs every 396 nsec.

The instantaneous luminosity is given by:

$$\mathcal{L} = \frac{N_B N_p N_{\bar{p}} f}{2\pi \sigma_p^2 \sigma_{\bar{p}}^2}, \quad (2.1)$$

where the  $N_B$  is the number of bunches in the machine,  $N_{p,\bar{p}}$  is the number of protons/anti-protons in a bunch,  $f$  is the bunch revolution frequency and  $\sigma_{p,\bar{p}}^2$  is the effective width of the proton/antiproton beams.

Although the instantaneous luminosity is proportional to the number of particles per bunch, its value is limited by the number of interactions per bunch crossing. This limitation is due to the superposition of several  $p\bar{p}$  interactions within the same bunch crossing, resulting in an increase of the event complexity.

The average number of interactions per bunch crossing is shown in figure 2.2 as a function of the instantaneous luminosity for different number of bunches. From the plot it is evident that, at a fixed instantaneous luminosity, to keep the number of

interactions per bunch reasonably small, the number of bunches in each beam has to be increased. More bunches results in a shorter interval between each interaction. This time interval is the clock of the whole apparatus: accelerator and detectors. Presently the Tevatron is running with  $36 \times 36$  bunches resulting in a bunch crossing interval of 396 ns.

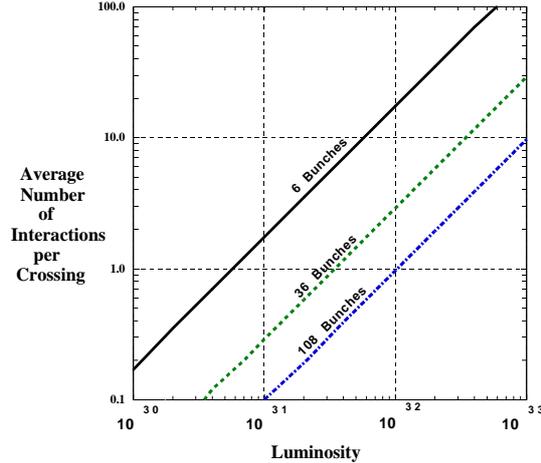


Figure 2.2: Average number of interactions per bunch crossing as a function of the instantaneous luminosity for different beam conditions

There are two detectors along the Tevatron ring located at the B $\emptyset$  and D $\emptyset$  points. The data used for the analysis reported in this thesis were collected by the detector located at B $\emptyset$  - the **C**ollider **D**etector at **F**ermilab (**CDF**).

#### 2.2.4 Performance and future perspectives

Many problems were identified during the start-up period in 2001: the efficiency of the machine was severely limited, for instance, by beam-beam interaction effects. There were major losses of anti-protons during the focusing of the beam (*squeeze*), and smaller but significant ones at 150 GeV and during acceleration. Most of these losses have been overcome by changing the helices to increase the beam separations, with consequently a smaller anti-proton emittance. In addition other improvements have been implemented upgrading the accumulator stochastic cooling and modifying the proton beam loading compensation in the MI.

At present all the accelerator complex is performing very well with the instantaneous luminosity reaching peaks above  $2.8 \times 10^{32} cm^{-2}s^{-1}$  (see figure 2.3) and integrated luminosity of about  $3 fb^{-1}$ , as shown in figure 2.4.

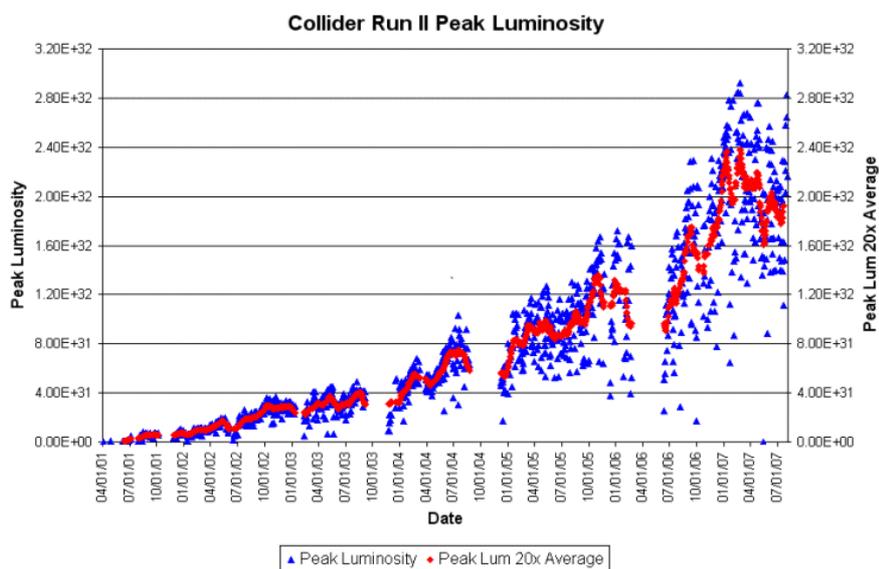


Figure 2.3: Instantaneous luminosities at the beginning of the stores through October 2007.

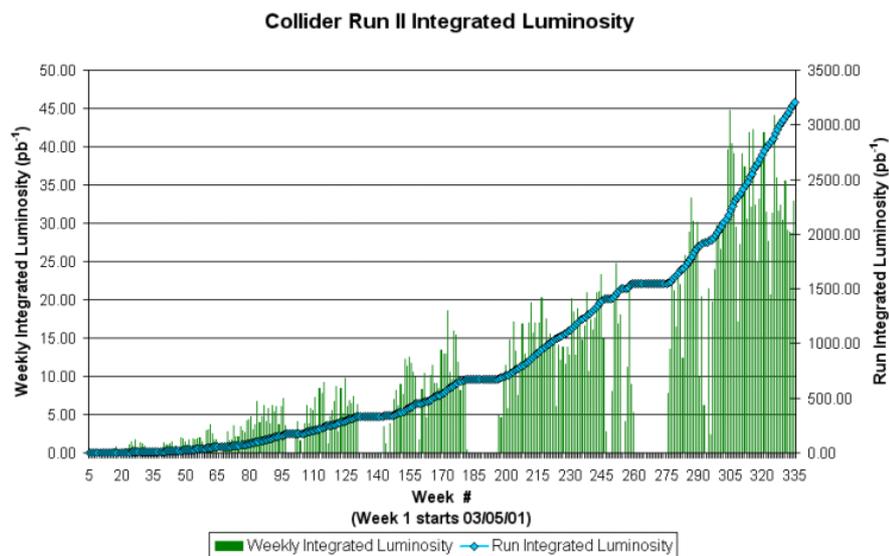


Figure 2.4: Delivered luminosity since the beginning of Run II to October 2007.

Figure 2.5 shows the integrated luminosity estimate for the next 2 years, for different possible configuration of the Tevatron accelerator. The designed luminosity for 2009 was set to  $9 \text{ fb}^{-1}$ . A reasonable estimate, taking into account the good performances achieved by the accelerator complex, is to reach approximately the  $6 - 8 \text{ fb}^{-1}$  by the end of 2009.

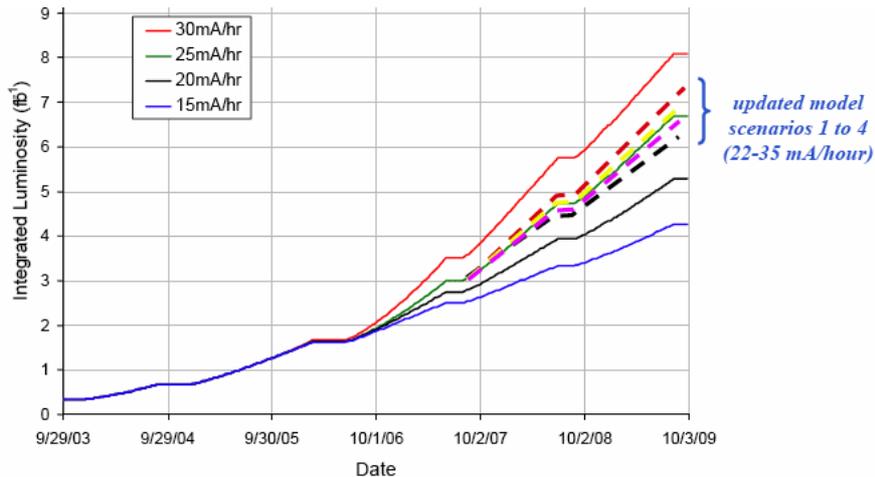


Figure 2.5: Estimates of the integrated luminosity delivered by the Tevatron collider for different accelerator running conditions.

More details about the data taking efficiency and the luminosity acquired by the CDF detector will be given in section 2.6.1, after a description of the CDF detector itself.

## 2.3 The CDF II detector

The Collider Detector at Fermilab (CDF) [44, 45], is a multipurpose particle detector built around the Tevatron  $B\bar{O}$  interaction point. It is approximately 15 meters long and 10 meters high and maintains approximate axial and forward-backward symmetries. The associated Cartesian coordinate system is defined as a right-handed basis with the  $z$ -axis set by the colliding beams and the protons moving in the positive  $z$ -direction.

The  $x$ -axis points radially outwards and the  $y$ -axis vertically upwards. Occasionally it is convenient to work in cylindrical  $(r, z, \phi)$  or polar  $(r, \theta, \phi)$  coordinates, where the azimuthal angle  $\phi$  is the  $(xy)$  plane angle measured from the direction of  $x$ -axis. Another coordinate system is commonly used in collider physics where the polar angle

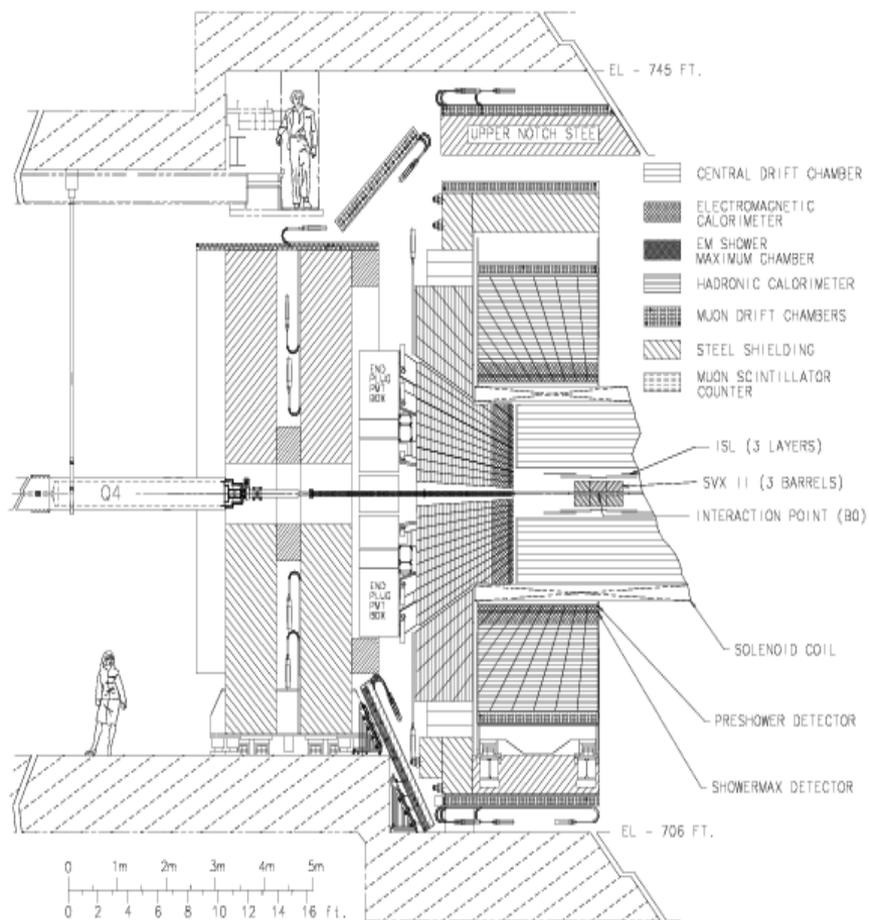


Figure 2.6: Elevation view of one half of the CDF II detector.

$\theta$  is replaced by pseudo-rapidity:

$$\eta = -\ln \tan \frac{\theta}{2}. \quad (2.2)$$

The utility of the  $(r, \theta, \phi)$  coordinate system stems from the axial symmetry of the experiment and from the nature of proton-antiproton collisions. The latter manifests itself in the fact that the interactions of colliding partons often result in considerable longitudinal momentum of the collision products. The rapidity of the system:

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z}, \quad (2.3)$$

transforms under boosts along the  $z$ -axis as  $y' = y + \tanh^{-1} \eta$  and thus  $\Delta y$  are invariant under such transformations. The pseudo-rapidity is the relativistic or massless approximation to the real rapidity of the system:

$$\eta = \frac{1}{2} \ln \frac{p + p_z}{p - p_z} = -\ln \tan \frac{\theta}{2}. \quad (2.4)$$

Many detector components are segmented uniformly in  $\eta$  and  $\phi$ . Typically the pseudo-rapidity  $\eta$  refers to "detector  $\eta$ " is determined with respect to the center of the detector coordinate system. The  $\eta$  can also be evaluated with respect to the interaction point. (in that case it is called "event  $\eta$ ").

The particles produced in each collision transverse through various detector sub-systems. The first system to be encountered is the tracking system, then calorimetry and finally the muon sub-detectors. The schematic overview of the CDF detector is shown in Figure 2.6.

### 2.3.1 Tracking system

The trajectory of particles can give valuable information about the kinematics of physical process, including charge sign and good momentum resolution if magnetic fields are present. The process of reconstructing a particle trajectory is known as *tracking*. In this section we describe the sub-detectors that form the integrated tracking system of the CDF detector as shown in Figure 2.7, followed by a brief overview of the basics in tracking reconstruction.

#### Solenoid:

All the tracking systems are enclosed inside a superconducting solenoid of 1.5m in radius and 4.8m in length. The solenoid provides a very homogeneous magnetic axial field of 1.4 Tesla inside a volume of 2.8m in diameter and 3.5m along the  $z$  direction. In normal operating conditions its current consumption is about 4650 A.

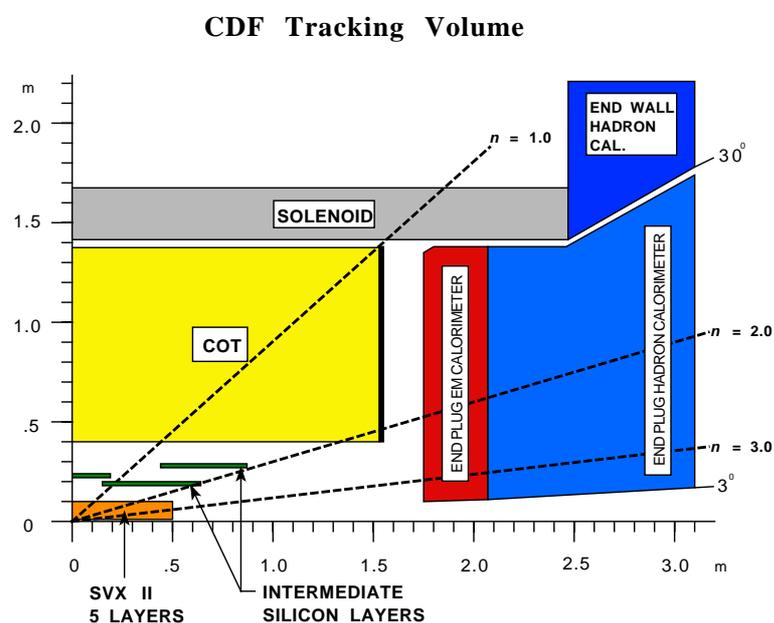


Figure 2.7: Longitudinal  $(r, z)$  view of the CDF tracking volume. its main components are the solenoid, the central outer tracker (COT), which is a gas drift chamber, and the silicon micro-strip detectors - ISL, SVX and L00.

### Central Outer Tracker (COT)

The COT position defines the global CDF reference frame and it is the main component of the tracking system. It is located inside the solenoid in the region of  $|z| < 155$  cm between radii of 44 and 132 cm. The COT is a cylindrical multi-wire open-cell drift chamber. It consists of 8 super-layers, each made of a large number of cells (varies between 168 cells in the inner-most layer and 480 cells in the outer-most layer), Figure 2.8.

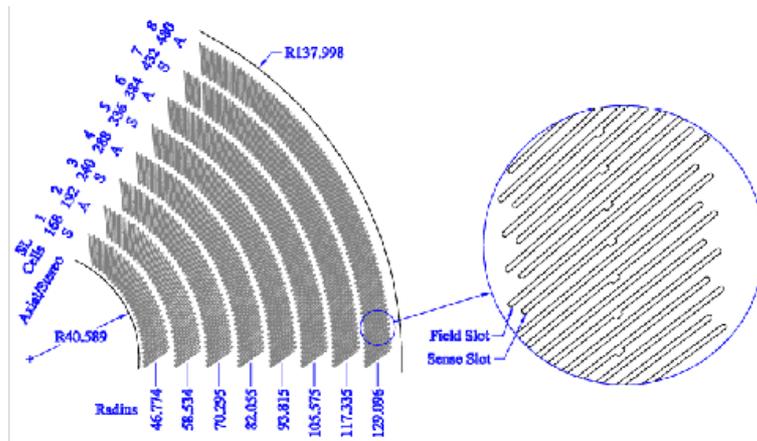


Figure 2.8: 1/6 section of the COT end plate. For each super layer is given the total number of cells, the wire orientation (axial or stereo), and the average radius. The enlargement shows the sense and field slot geometry in detail. Dimensions are in cm.

Each cell is either an 'axial' or 'stereo' cell. Each cell has an approximate size of about 2 cm by 10 cm and with a length of 310 cm spans the whole longitudinal direction of the COT.

When a charged particle passes through the gas it leaves a trail of ionization electrons. These electrons drift towards the sense wire by virtue of the electric field created by the field panels and potential wires. Close to the sense wires the electrons are strongly accelerated because of the local  $1/r$  electric field, producing more ionization in a process known as an avalanche. The signal of this wire is further analyzed and depending on the charge collection (beyond some threshold) the wire is said to be 'hit'. The time it takes from the moment the collision was expected to occur to the time the signal was detected gives information about the distance between the particle trajectory and the sense wire. The high voltage applied to the COT is such that the electric field drift is about  $1.9\text{kv/cm}$ . The drift velocity is about  $54 \mu\text{m/ns}$ .

The single 'hit' position resolution of the COT has been measured to be  $140 \mu\text{m}$ , resulting in a transverse resolution of :

$$\frac{\Delta_{P_T}}{P_T} = 0.15\% \frac{P_T}{\text{GeV}/c} \quad (2.5)$$

### Silicon vertex detector

The silicon tracking and vertex sub-detector of CDF consists of three independent detectors named L00, SVX and ISL. They differ in size, radius and number of active elements, but they all use silicon micro-strip ( $\mu$ -strip) technology. They enable the measurement of the position of secondary vertexes, like those produced in the decay of long lived hadrons such as B mesons, with excellent resolution - a factor 10 better than the resolution obtained using the COT only. Silicon micro-strip detectors were used for the first time in a hadronic accelerator by CDF during Run I of data taking; silicon pad detectors had previously been used by the UA2 experiment at the  $Spp\bar{p}S$  collider.

**SVX** The silicon is supported with a rohacell foam in assemblies called ladders. Each ladder supports double-sided silicon modules, where in each side lies a string of four modules connected to each other by wire bonds, thus quadruplying the length of the strip. Twelve of these ladders are set in a semi-circular configuration to form a layer that surrounds the beam pipe at a given radius. The ladders are supported by two beryllium bulkheads, Figure 2.9.

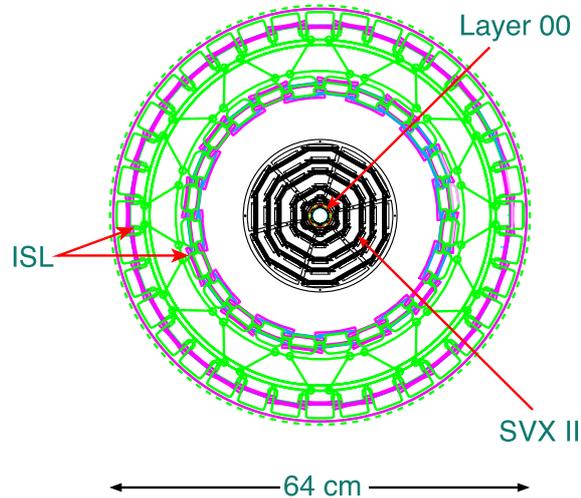


Figure 2.9: An overview of the CDF SVX, ISL and L00 subdetectors.

The bulkhead provides support for 60 ladders in 5 concentric layers, confirming a 29 cm long SVX barrel. The SVX is built by placing three barrels along the beam-pipe. Layers are numbered from 0 (innermost) to 4 (outermost). Layers 0,1 and 3

combine  $(\rho - \phi)$  modules on one side with  $90^\circ$  stereo on the other side. Layers 2 and 4 combine  $(\rho - \phi)$  modules on one side with a small angle stereo on the other. The single hit resolution of the SVX is about  $12\mu m$ . The large number of channels require that the front-end preamplifier electronics is mounted close to the modules. This results in a better signal-to-noise ratio but also in additional multiple scattering due to the extra material in the volume.

The **ISL** consists of three layers. In the central region ( $\eta \leq 1$ ) a single layer is placed at a radius of 22 cm. Two more layers are located in the region  $1 \leq \eta \leq 2$  at a radii of 20 cm and 28 cm. The layers are double sided with  $(\rho - \phi)$  on one side and a small stereo angle on the other side. To reduce the total number of channels to 268,800 only every 2nd channel is read. The ISL single hit resolution is about  $20\mu m$ .

The **L00** consists of a single sided layer of 12 ladders, shown in Figure 2.10. Six of the ladders lie at a radius of 1.35 cm and the other six at 1.62 cm from the beam-line. Each ladder is assembled from six sets of two wire-bonded modules, spanning 95 cm in the  $\hat{z}$  direction. The layer is supported by a carbon fiber structure. In addition, L00, SVX and ISL have dedicated cooling lines running at a nominal temperature of  $-6^\circ C$ . A special effort was made to accurately align the silicon detectors with respect to the beam, as opposed to the COT. This is because a small misalignment of the COT, while not significantly changing the track information obtained with it, can effect the impact parameter of the tracks<sup>3</sup>. which is obtained using silicon hit information and is used for triggering purposes, introducing a dependence on the  $\phi$  coordinate.

**Tracking Parameterization:** In a plane perpendicular to a homogeneous magnetic field, such as the one provided by the CDF solenoid, the trajectory of a charged particle follows a circular pattern to the limit of energy loss due to the traversed material. The longitudinal component of the particle's momentum is not modified by the axial magnetic field, thus in three spacial dimensions the trajectories of charged particles are helices. The helices can be parametrized with five parameters chosen to be:

$C$ : the signed helix curvature defined as  $C = sign(q)/2R$ , where  $R$  is the radius of the circle in the transverse plane and  $q$  is the charge of the particle.

$z_0$ : The position of the point along the  $\hat{z}$  axis at the point of closest approach.

$d_0$ : a signed quantity defined as  $d_0 = sign(q)(r_C - R)$ , where  $r_C$  is the position of the center of the circle in the transverse plane, and its magnitude is the distance from  $z_0$  to the closest point in the track.

$\phi_0$ : the direction of the transverse momentum of the particle at the point of closest

---

<sup>3</sup>Distance to closest approach of the track to the beam, explained later

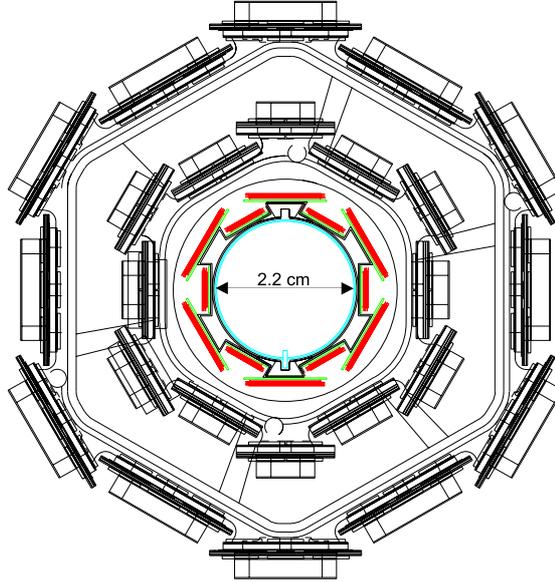


Figure 2.10: End view of the L0 detector

approach.

$\cot(\theta)$ : the ratio of the longitudinal and the transverse momentum ( $P_z/P_T$ ).

### 2.3.2 The calorimetry

Charged particles with energies greater than 350 MeV may leave the tracking volume and propagate beyond the solenoid magnet that surrounds it. Neutral particles, both photons and neutral hadrons, leave the tracking volume undetected. Even if they were seen by the tracking system, no information about their momenta be available since their trajectories do not bend in magnetic field. In many cases the measurement of their momenta is essential for understanding the observed events. The calorimetry subsystem serves the purpose of filling this gap and is based on the fact that as particles that have reasonably high electromagnetic or hadronic interaction cross sections propagate through matter they transfer their energy to the medium, until eventually all of it is absorbed. Some media produce measurable response to such energy depositions by emitting light in amounts dependent on the amount of energy lost by the particles.

The CDF calorimetry system[46, 47] consists of alternating layers of scintillator and absorber material. As particles interact with absorber they produce cascades showers” that penetrate the scintillator. The light from the latter is guided into photo-multipliers. Their response is in turn digitized and with the help of relevant

calibrations converted into the measurement of the deposited energy.

The calorimeters are divided into segments or "towers" in such a way that the division boundaries between them point at the interaction point. The entire calorimetry system consists of two regions, central and forward; the latter is also known as the "plug". The central calorimeters cover the region of  $2\pi$  in  $\phi$  and as far as  $|\eta| < 1.0$  in pseudo-rapidity. They are segmented into the "towers" of  $0.11 \times 15^\circ$  in  $\eta \times \phi$ . The  $\eta$  segmentation of the calorimeter system is shown in figure 2.11. The Central Electro-

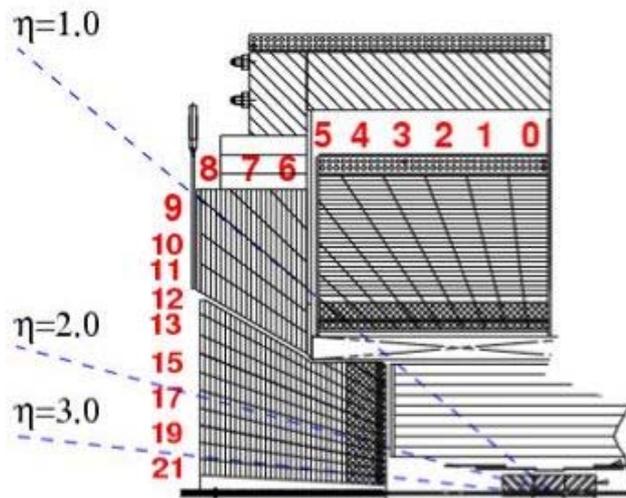


Figure 2.11:  $\eta$  segmentation of the CDF calorimeter system.

Magnetic (CEM) part consists of alternating layers of lead absorber and polystyrene scintillator, while the Central Hadronic part (CHA) uses thicker steel plates as the absorber. The CEM is  $18X_0$  radiation lengths thick, while the CHA is around  $4.5\lambda_0$  attenuation lengths thick. The energy resolution of the CEM is estimated to be:

$$\frac{\sigma_E}{E} = \frac{13.5\%}{\sqrt{E_T}} \oplus 2\%, \quad (2.6)$$

where the notation  $\oplus$  means that the constant part is added in quadrature. The CHA resolution is estimated to be:

$$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E_T}} \oplus 3\%. \quad (2.7)$$

Other parts of the calorimeter system are the the Central Shower-maximum (CES) and Central Pre-Radiator (CPR) detectors. The CES is a gas multi-wire proportional chamber with cathode strips that provide measurements of the  $z$ -position and anode

wires that allow a measurement of the  $\phi$  of the energy deposition. The chamber is embedded into the CEM at approximately  $5.9X_0$  where the maximum of electromagnetic energy deposition occurs. The position resolution in both directions is around 2 mm. The CPR consists of proportional chambers placed between the solenoid and the calorimeter. These two subsystems, CES and CPR, provide both position measurement that helps in matching energy depositions to tracks and in shower profile measurements (information used in particle identification to distinguish between  $e^\pm/\gamma$  and  $\gamma/\pi_0$ ).

It should be noted that the spatial segmentation of the calorimeters is rather coarse and thus it is possible that the energy measured by a particular tower includes contributions by multiple particles. Occasionally this poses a problem as the energies of the individual particles cannot be determined. In other cases, most notably when a "jet" of particles that results from hadronization of a quark or gluon coming directly from hard scattering hits the calorimeter, the total energy of all the particles in the jet is exactly the information that is needed.

### 2.3.3 The muon system

Muons are 207 times heavier than electrons, so they lose substantially less energy due to electromagnetic interactions as they travel through the calorimeter material. This allows the muons to pierce through the calorimetry subsystem after they exit the tracking volume. The CDF muon subsystem consists of several chambers located outside the calorimeters and includes Central Muon detector (CMU), Central Muon upgrade (CMP), Central Muon extension (CMX) [48, 49]. Due to space and design constraints, the muon coverage is incomplete. The CMU is comprised of series of rectangular drift cells four layers deep. The hits registered in at least 3 out of 4 layers form a "stub", which after being properly matched with the corresponding COT track suggests the presence of a muon. The CMP sub-detector consists of both drift chambers and scintillator plates and functions similar to the CMU by providing a "stub" that is used in muon reconstruction. The coverages of CMU and CMP partially overlap. The CMP is placed behind additional 60 cm of steel and thus is less sensitive to remnant hadrons that penetrate through CHA (punch-through hadrons). The CMX subsystem, like the CMP, combines drift cells and scintillator plates arranged in semi-conical arches that cover the pseudo-rapidity region of  $0.6 < |\eta| < 1.0$ .

## 2.4 The CDF Data Acquisition and Trigger

A schematic view of the CDF Data Acquisition (DAQ) and trigger system is given in Figure 2.12. The trigger plays an important role to efficiently extract the most interesting physics events from the large number of minimum bias and background

events and to reduce the amount of data to a reasonable volume. The CDF trigger

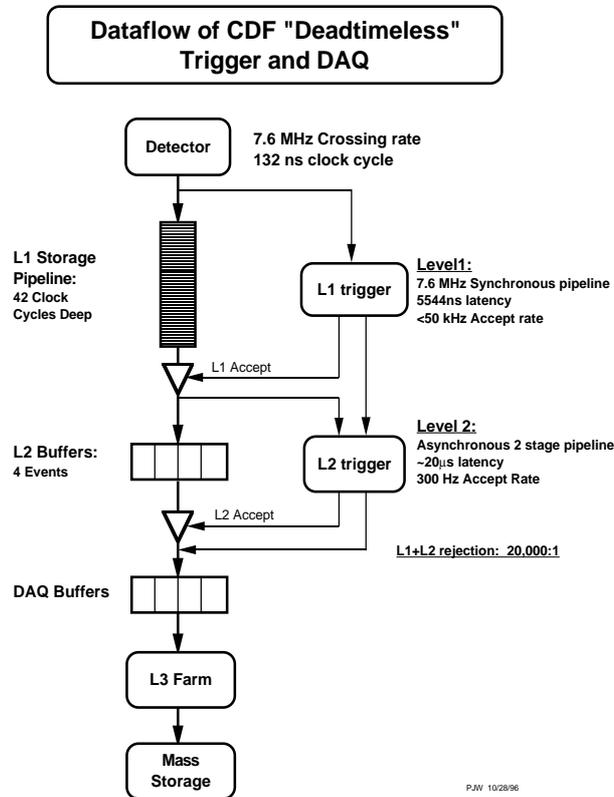


Figure 2.12: Data flow schematic of the three level pipe-lined and buffered trigger system.

is a three level system. The time available for event processing increases in each level of the trigger, which permits the use of an increasing amount of information to either accept or reject an event. While Level-1 (L1) and Level-2 (L2) triggers are based only on parts of the detector information, the Level-3 (L3) triggers makes use of the complete event data. A signal is defined as an event where a variable (for instance the energy in the calorimeter) lies above a certain trigger threshold. A list of quantities that can be cut on at the different trigger levels is given in [50]. L1 and L2 are hardware triggers while L3 is a software trigger. An optimized version of the reconstruction executable is running on a Linux PC farm with about 100 nodes. The design processing rates for L1, L2 and L3 are 50 kHz, 300 Hz and 50 Hz respectively. The typical event size is approximately 250-300 kB. The L1 triggers base their decisions on information of the calorimeters, the muon system, the forward detectors and the drift chamber(see Figure 2.13). The eXtremely Fast Tracker(XFT)

reconstructs  $r/\phi$  tracks in the COT in time to take part in the Level-1 decision making process. The XFT produces tracks with a transverse momentum resolution of  $\delta P_T/P_T^2 = 0.01651 \text{ GeV}^{-1}$  and an angular resolution of 5.1 mrad. An important

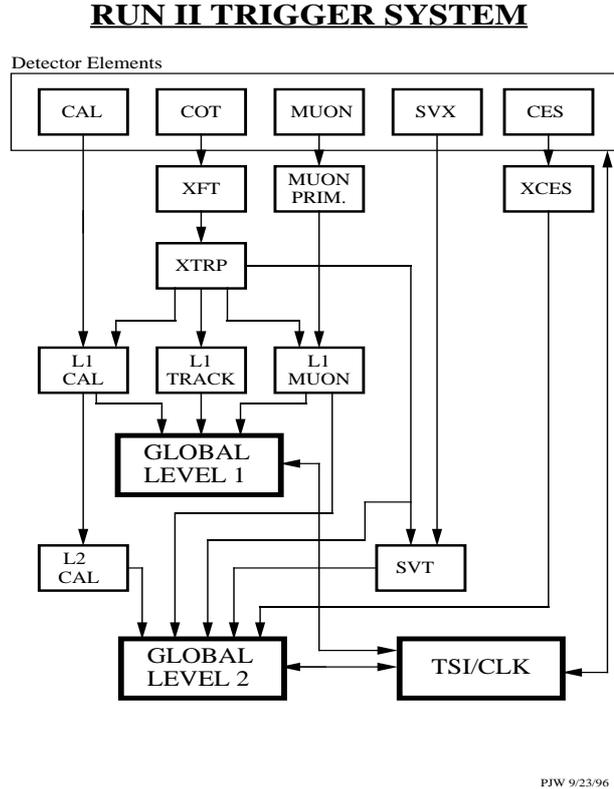


Figure 2.13: Functional block diagram of the CDF L1 and L2 trigger system.

feature of L2 is the Silicon Vertex Tracker(SVT). It adds silicon  $r/\phi$  hits to the L1 XFT tracks. This makes a large number of important processes involving the hadronic decays of bottom hadrons accessible as it calculates the impact parameter with respect to the nominal  $z$  axis and it allows to select events with two tracks having a large impact parameter in order to identify secondary vertexes. Further details on the SVT are given in the section below 2.5.

Full event reconstruction takes place on the L3 trigger farm and hence a wide variety of requirements can be imposed on the events passing L3 [51]. Events passing the final trigger level belong to a certain trigger path. Each "path" is a unique combination of L1, L2 and L3 triggers. The trigger decisions are combined via a logical "AND". Many paths combined by a logical "OR" can be used to feed a single data set. The data are written to approximately 20 streams and stored on tape.

After reprocessing the events they are split up into more specific data sets. During measurements the data quality is monitored on-line [52].

## 2.5 The Silicon Vertex Tracker

The Silicon Vertex Trigger inputs are the list of COT tracks found by the XFT and the data from four axial silicon layers. It associates a set of silicon hits to a XFT track and fit the result to a circle in the transverse plane, determining the track parameters with a precision comparable to the offline. SVT is divided in several subsystems as shown in the figure 2.14. The SVT core is organized as 12 identical

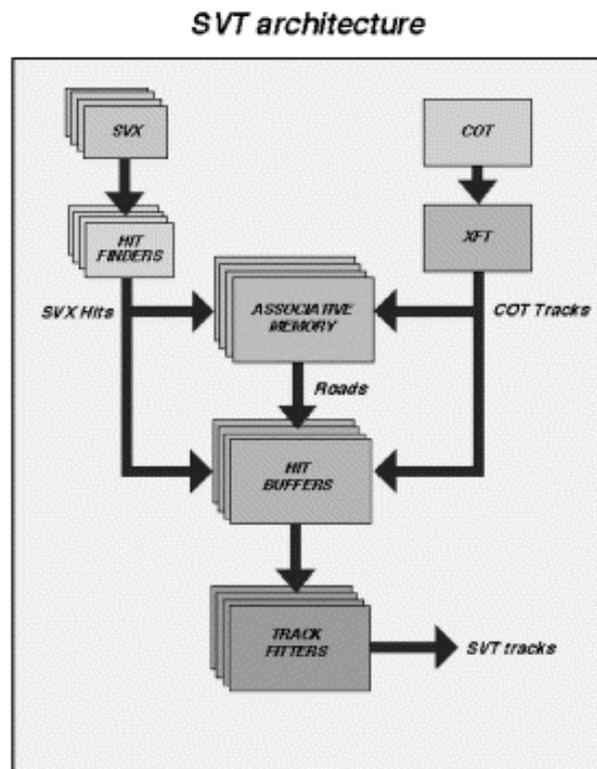


Figure 2.14: SVT component list and data flow

systems which correspond to the 12 silicon wedges in  $\phi$ . The Hit Finders, three per sector, calculate the cluster centroid position of the silicon hits and sends them to the Hit Buffer where they are stored for future reference and to the Associative Memory (AM) units. This system perform the pattern recognition. Upon receiving the list of silicon hits and XFT tracks, each AM chip searches for all the coincidences of 4

silicon hits and XFT tracks which represent a legitimate particle trajectory (roads). This is done by comparing the input data with a stored set of precalculated patterns. In order to limit the number of roads that would be needed to match all the possible tracks the AM system groups clusters into "superstrips" each covering about  $250 \mu m$ . In this way the number of channels is reduced but the coarse resolution increases the number of fake tracks and may cause multiple candidates to fall within the same road. This width is a good compromise between cost, performances and processing time. For each Hit Finder there are 2 boards with 128 chips each and 128 roads per chip which corresponds to 32000 roads approximately, and to a coverage of about 95 %. As soon as the last hit of an event is read the pattern recognition is ended and the list of the 'active' roads is sent to the Hit Buffer. It retrieves the full XFT and SVX clusters information and send them to the Track Fitter. This last subsystem checks all clusters combinations in each road and outputs the tracks parameters. Additional details on the fitting algorithm can be found in [53].

Tracks reconstruction is made only in the transverse plane and it is strongly related to the beam-line position: each detector element has to be correctly aligned to each other and to the beam. The internal detector alignment was performed with high precision during the assembling, but some misalignment and time variations of the beam position, sometimes of the order of  $100 \mu m$ , are possible during the stores.

Each run the SVT fits the beam position every few minutes and it uses it to correct on-line the impact parameter of the tracks, according to the formula:

$$d_0 = y_{beam} \cdot \cos(\phi) - x_{beam} \cdot \sin(\phi)$$

where  $x_{beam}$  and  $y_{beam}$  are the beam coordinate and  $\phi$  is the track azimuthal angle with respect to the beam. A simple schema is drawn in figure 2.15.

The beam position is also sent to the Tevatron for correction when the displacement between two following stores is so large that a new set of patterns would be necessary.

The current impact parameter resolution is  $\sigma_d = 48 \mu m = 33 \mu m \oplus 35 \mu m$ , where the first number is the beam spot and the second one is SVT resolution. The SVT performances have been evaluated using a sample  $J/\psi \rightarrow \mu\mu$  collected with a muon trigger. In figure 2.16 the efficiency is shown as function of the track impact parameter. The average value is around 80% with a decrease over 1 mm. This behavior is due to a partial coverage of AM patterns over 1 mm.

The efficiency is also shown as a function of pseudo-rapidity  $\eta$  (upper plot) and track momentum  $p_T$  (lower plot). The flat  $\eta$  dependence between  $-1$  and  $+1$  is expected while the low efficiency at low transverse momentum is due to the fact that the AM patterns are generated flat in curvature causing a lower coverage at low track  $p_T$ . Both efficiency are around 80 %.

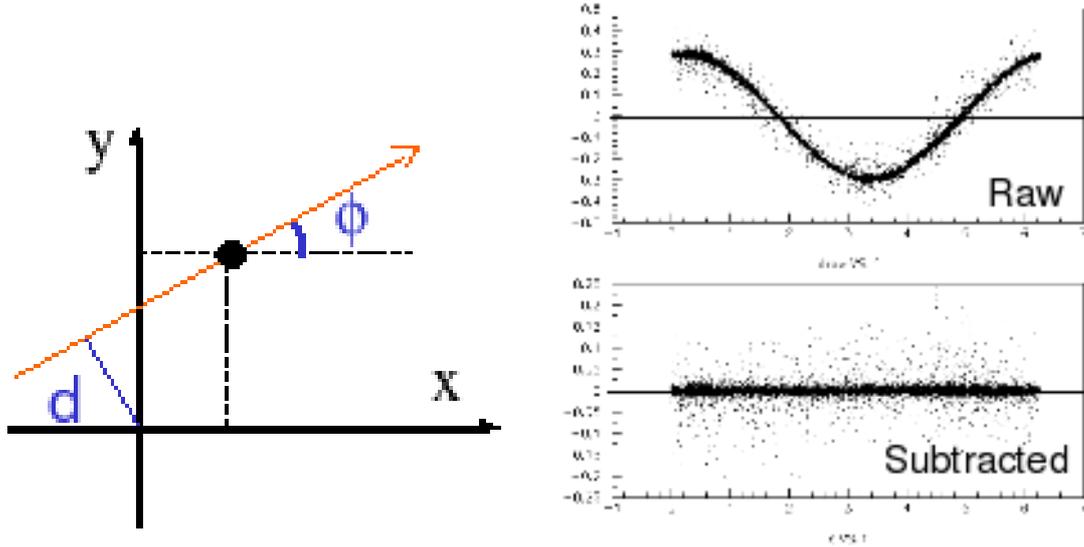


Figure 2.15: A simple representation of the correction applied by the SVT to the impact parameter of the track (left). The impact parameter with respect to the track  $\phi$  before and after the correction is applied (right).

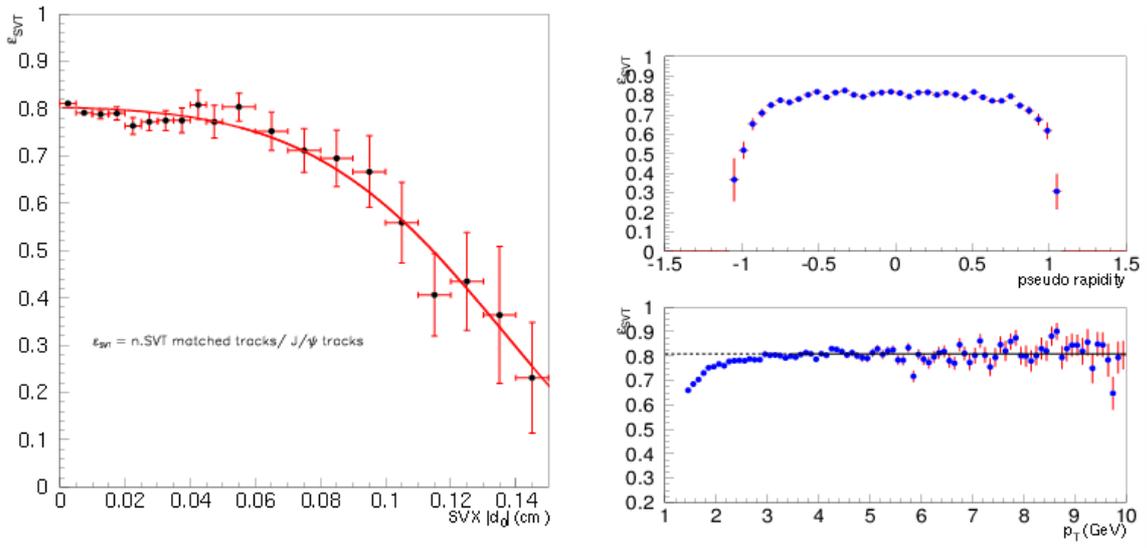


Figure 2.16: SVT efficiency calculated as a function of track impact parameter (left) and track  $p_T$  and  $\eta$  (right)

The SVT is, in fact, so effective in selecting heavy-flavors that a simple on-line selection based on the requirement of two large impact parameter tracks gives the possibility of reconstructing the decay  $D^0 \rightarrow K\pi$  (shown in figure 2.17) or  $B^0 \rightarrow K\pi$ . The picture shows the two tracks invariant mass reconstructed on very early data of Run II, collected in 2002 (about  $1.3 \text{ pb}^{-1}$ )

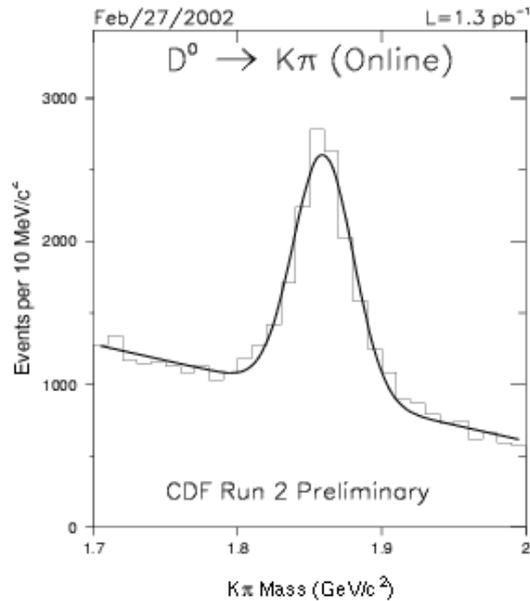


Figure 2.17:  $D^0 \rightarrow K\pi$  decay reconstructed on-line using a two track trigger path

## 2.6 On-line Monitoring

After passing L3 triggers, the data is saved in storage systems. Some of the data stream are picked up by on-line data validation programs, called ‘Consumers’. The data streams are arranged by Consumer Server Logger (CSL) to meet hardware and software requirements from up and down streams. The consumer programs consist of 13 subprograms which are meant to verify the condition of the different sub-detectors and the quality of data when data taking is performed.

Their role is extremely important. In fact it is the information taken from the on-line monitoring system that is used to perform a preliminary selection on the data with respect to the sub-detector that are relevant for the analysis. Details will be given in chapter 3. Information about the consumers is found at [54].

### 2.6.1 CDF Luminosity

The luminosity at CDF is measured both on-line and offline with the Cerenkov Luminosity Counters (**CLC**), using the process of  $p\bar{p}$  inelastic scattering.

The CLC (Cerenkov Luminosity Counter) is placed at the sides of the detector, in the  $3^\circ$  gap between the plug calorimeter and the beam pipe. It spans from 184 cm to 405 cm in the  $\hat{z}$  axis.

It measures the average number of interactions per bunch crossing ( $\mu$ ), which allows the calculation of the instantaneous luminosity  $\mathcal{L}$  from:  $\mu f_{bc} = \sigma_{p\bar{p}} \mathcal{L}$ , where  $f_{bc}$  is the frequency of bunch crossing at the Tevatron and  $\sigma_{p\bar{p}}$  is the total  $p\bar{p}$  cross section of about 100mb at the Tevatron's  $\sqrt{s} = 1.96 TeV$ .

To measure  $\mu$  the CLC takes advantage of Cerenkov radiation - the light emitted when particles travel in a medium with a velocity higher than the speed of light. This light is radiated in a fixed angle ( $\delta$ ) with respect to the particle's momentum. The angle  $\delta$  depends on the refraction index of the medium ( $n$ ) and the particle's velocity -  $\cos(\delta) = 1/n\beta$  where  $\beta = v/c$ .

The CLC is composed of an array of Cerenkov counters in the shape of long cones (cone-modules). More detailed information about the CLC is found in [55].

The on-line luminosity is available in real time and takes into account multiple interactions automatically but does not include further possible corrections, which are done offline. Every CDF event contains both the on-line and offline luminosity information: the average instantaneous and the integrated value up to that event. It is then possible to find the integrated luminosity for a given dataset.

The systematic error on the luminosity measurement is dominated by the uncertainty on the elastic  $b\bar{p}$  cross section, by the one on the CLC acceptance and stability, together with the uncertainties on beam losses, beam position and statistics. The total systematic error on luminosity is  $\delta\mathcal{L}/\mathcal{L} \approx 6\%$ .

Figure 2.18 shows the integrated luminosity collected by CDF until September 2007. However the analysis presented in this thesis used about  $300 \text{ pb}^{-1}$  of data, until September 2004, mainly because of the time that was necessary to process the data.

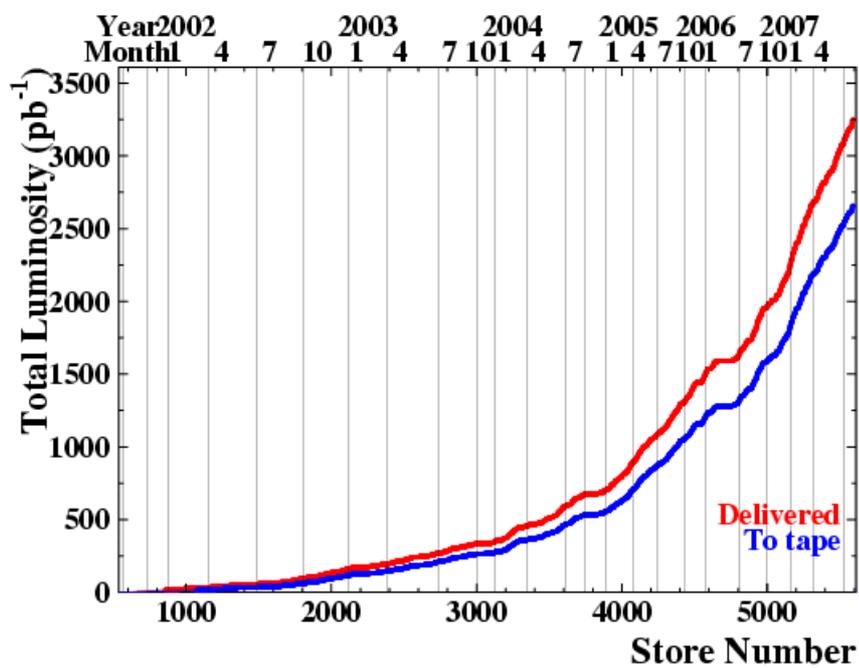


Figure 2.18: Integrated luminosity collected by the CDF detector until October 2007, together with the total integrated luminosity delivered by Tevatron

# Chapter 3

## Event Reconstruction

The previous chapter has given an overview of the CDF detector, paying special attention to those subsystems that were most relevant to this analysis. Here a description of the different steps of the software event reconstruction is given. The first part of the chapter is dedicated to an overview of common tools: the reconstruction of charged particle trajectory in the tracking system (COT and silicon detectors SVX/ISL/L00) in section 3.1, the algorithms that define jets of particles in the calorimeter (section 3.2) and the determination of the position of the primary interaction point in section 3.3.

The second part of the chapter goes into the details of the analysis. Section 3.4 describes the data sample used in this measurement: the three level trigger that selects data events on-line and the basic quality requirements that determine the available integrated luminosity. It also includes a brief description of the Monte Carlo samples that were used to study the effect of the event selection, to determine corrections to the jet energy scale and to estimate systematic uncertainty. The analysis strategy and the event selection are described in section 3.5. Finally, the problem of the determination of the jet energy scale and an estimate of the necessary corrections are described in section 3.6. Details on the  $b$ -jet identification procedure are given in the next chapter.

### 3.1 Track reconstruction

The detection and tracking of charged particles ('tracks') is an essential part of this analysis. On-line, tracks are reconstructed using the SVT trigger. Offline, the reconstruction of primary and secondary vertexes depends on the software capabilities to define the direction and curvature of the particle's path together with its momentum.

The tracking algorithms can either use combined informations from COT and SVX/ISL/L00, or use standalone input from one of the two sub-systems. The main

procedures implemented are summarized in the next sections.

### 3.1.1 Tracking in the COT

The Central Outer Tracker is able to reconstruct tracks and their momenta up to  $|\eta| \approx 1$ , the maximal coverage of the drift chamber. The pattern recognition algorithm is performed according to many successive steps [56], summarized as follows:

- In each of the three COT super-layers, hits in three consecutive wires are grouped and fit to a straight line by the method of the least square. These segments are listed in decreasing  $p_T$  and are used as seeds. Other hits in the super-layer within a distance of  $\approx 1$  mm from the segment (20 ns in road time considering a drift time of  $55 \mu\text{m}/\text{sec}$ ) are added to the straight line fit using an iterative procedure. Seed-segments defined in this step can use axial or stereo super-layer hits.
- Tracks are initially reconstructed using axial ( $r - \phi$ ) super-layer seed-segments only. The segments with good angle and position matches are then linked together to form axial tracks. To increase the reconstruction efficiency, two algorithms are run in parallel for this purpose:
  - The ‘segment linking’ algorithm matches segments from different super-layers and makes a fit on all hits in the segments using a  $\chi^2$  minimization;
  - The ‘histogram linking’ algorithm starts with one segment position and the beam position, makes a circle fit to these points and considers a band of  $\pm 1$  cm around this circle. It then looks for any hit within this band: a  $200 \mu\text{m}$  binned histogram is filled with the radius (the distance to the center of the track-circle) of each hit. If the most populated bin contains more than 10 hits, a track is made from those hits. Finally, the procedure attempts to add other hits within  $750 \mu\text{m}$  of the track and re-fits it: if the fit succeeds and the track has at least 20 axial hits (15 if the seed segment includes hits from the fourth COT super-layer, SL4), the track is added to the track collection.

Track duplicates (reconstructed by both algorithms) are removed.

- In a second step of the pattern recognition procedure, stereo angle super-layer information is added to the axial tracks. Again two algorithms are implemented, this time running sequentially:
  - The stereo segment linking algorithm matches stereo segments to existing axial tracks starting from the outer stereo super-layer; thus it performs a

re-fitting of the track to estimate the  $z$  and  $\cot\theta$  helix parameters and then continues the procedure looking to inner layers.

- Following this re-definition, the hit linking algorithm tries to recover stereo information for axial tracks that failed the stereo-segment matching. To do this, stereo tracks are used to reconstruct the  $z$ -coordinate of vertex seeds in the event and scans the  $\cot\theta$  parameter of the helix for the best hit usage.
- Finally, tracks are re-fitted to take into account any underestimation of the material used in the procedures above, or variations in the value of the applied magnetic field. Furthermore, a refit is necessary to obtain the best energy loss corrections.

The tracking efficiency is a function of the track transverse momentum: it is found to be above 99% for tracks with  $p_T > 1.5$  GeV/ $c$  ( $p_T > 1.35$  GeV/ $c$  in case of isolated muons); while it decreases down to 95% for 500 MeV/ $c$  tracks.

### 3.1.2 Silicon tracking

Tracking algorithms using the silicon information are either stand-alone or combine silicon and COT inputs. The main features are summarized below.

- The silicon clustering algorithm (‘Si-standalone’) uses strip data from the silicon detectors and produces a set of strip clusters that represent charge deposited by a single particle as it traverses the silicon, considering that particles often deposit charge among several adjacent strips. The profile of the resulting ‘cluster’ depends on many factors, including, for example, the strip pitch, the angle of incidence of the track, the type of charge carrier, the diffusion of charge carriers in the silicon and magnetic fields. The purpose of the clustering algorithms is to identify groups of contiguous strips that represent the charge from a single particle, gather them into a strip cluster (hit) and estimate the exact location of the track impact. Once this set has been identified as a cluster, the position and resolution of the track are calculated.
- The Outside-In algorithm (‘OI’) takes COT tracks and extrapolates them into the silicon detectors, adding hits via a progressive fit. Only tracks with at least three hits are retained. As each layer of silicon is reached (going outside-in), a ‘road’ is established around the ‘seed track’: the road is four standard deviations wide, based on the error matrix of the track. Hits that are within the road are added to the track, and the track parameters and the error matrix are re-fit with this new information. A new track candidate is generated for each hit in

the road, and each of these new candidates are then extrapolated to the next layer in, where the process is repeated. As the extrapolation proceeds, the track error matrix is inflated to reflect the amount of scattering material encountered. At the end of this process, there may be many track candidates associated with the original COT track. The candidate that has hits in the largest number of silicon layers is chosen; if more than one candidate has the same number of hits, the  $\chi^2$  of the fit in the silicon is used as discriminating quantity.

One of the most important track features for the  $b$ -tagging algorithm is the impact parameter resolution. As already mention in section 2.5, this quantity measures the distance of the track from the primary interaction point and it is used to select ‘displaced’ tracks: tracks that do not originate from the primary interaction but from secondary decay vertexes, considered as heavy flavor candidates. The single track impact parameter resolution is approximately  $40 \mu m$  including a  $30 \mu m$  contribution from the beam-line. This is a considerable improvement with respect to the past (when  $\sigma_{d_0} \approx 50 - 60 \mu m$ ), due to the introduction of L00 information in the track reconstruction. Figure 3.1 shows the impact parameter resolution as a function of track  $p_T$  with and without L00 hits included in the reconstruction of isolated tracks.

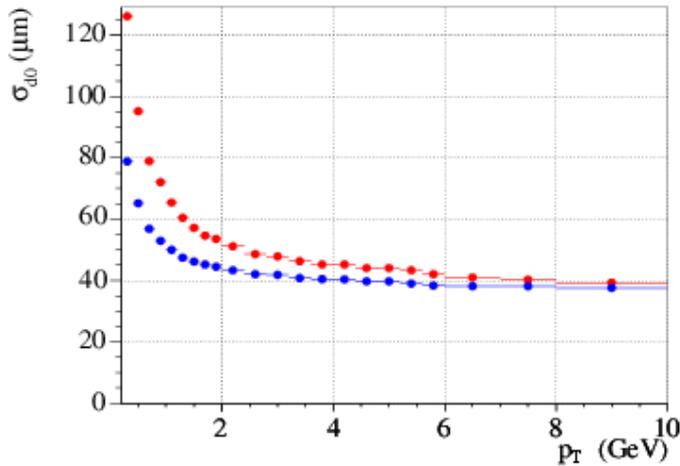


Figure 3.1: Resolution on the track impact parameter ( $\sigma_{d_0}$ ) as a function of the track transverse momentum ( $p_T$ ). Sample with (without) L00 clusters is shown in blue (red).

## 3.2 Jet reconstruction

A very important part of many QCD analyses is the ability to reconstruct jets. Some particles, such as charged leptons, or photons, traverse the detector without any decay or the production of other particles. This results in a very clear signature in the detector: a single track in the tracker for the case of charged particles and no track in the case of neutral particles, together with a well-localized energy deposit in the calorimeters.

Hadronic particles, on the other hand, are more complicated and complex to study. Because of the nature of the strong interaction, a single hard interaction involving a quark or a gluon will result in a large number of neutral or charged hadrons seen in the detector. This evolution from a single parton to a large number of final state particles is governed by QCD processes, in particular by parton shower processes as well as hadronization as seen in chapter 1.

The signature of a hadronic interaction in our detector is, for the purpose of this study, a large track multiplicity collimated in real or phase space, with an energy deposition in the calorimeters spread over a similar region to that of the tracks. This shower of observed particles is experimentally referred to as a jet [57].

At hadron colliders the definition of a jet is not straightforward. Because of the possibility of having gluon-mediated interactions, which by far dominate the interaction cross section, there is a possibility of having several jets in the event. This can lead to the shower of particles from one parton overlapping with those from another. There are also many more processes occurring, when protons and anti-protons collide, than just the hard interaction. These other processes are what is referred to as the underlying event [58]. All of these effects result in a large number of tracks and calorimeter energy deposits which must be disentangled into objects such as jets or isolated particles. This disentangling is made using jet algorithms.

There are two main classes of jet algorithms which have been studied and used so far at CDF. The first class of algorithms, the most widely used, are called cone algorithms. They look at the energy deposited in the towers of the calorimeter and define all particles within a physical cone of fixed size to belong to the same jet. Different cone algorithms differ in the method used to find the final cones, in the size of the cones, or in the way of dealing with cones that overlap. This is the class of algorithms which is discussed below.

The second class of algorithms, called the  $k_T$  algorithms, define jets in terms of towers which are not necessarily spatially collimated but are collimated in  $k_T$  [59] [60].  $k_T$  is proportional to the relative transverse momentum between two particles. This class of algorithms will not be discussed further but more information about this algorithm can be found in the references mentioned above.

The choice of the algorithm is non-trivial, it somehow defines the shape of the

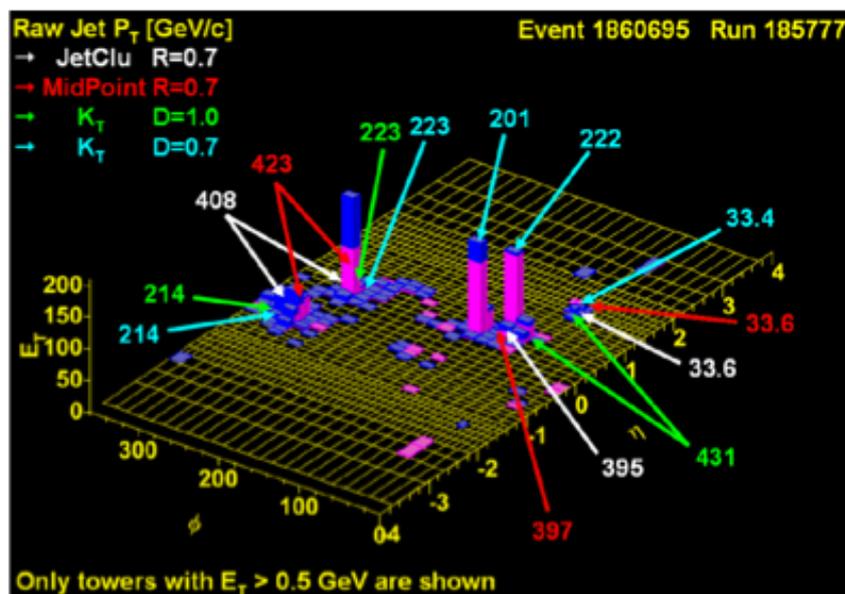


Figure 3.2: Calorimeter towers mapping for a real event as recorded by the CDF Event Display. The picture shows how different jet algorithms give different descriptions of the same event. The JetClu cone algorithm and the Midpoint algorithm reconstruct the same multiplicity of jets (3 jets) but associate to each one of them different energies. The  $k_T$  type algorithm gives the same description only when a larger size ( $D=1.0$ ,  $D$  defines a sort of jet radius) is used.

event under study: figure 3.2 represents a real event calorimeter cluster deposition. Each algorithm gives a completely different description of the event in terms of jets.

### 3.2.1 Cone Algorithms

The cone-based algorithm is still the most widely used at hadron colliders. The jets are characterized in term of variables that are invariant under boosts along the beam axis. The variables are: transverse momentum,  $p_T$ , or the transverse energy,  $E_T$ ; azimuthal angle around the beam direction,  $\phi$ ; and pseudo-rapidity,  $\eta$ . The algorithm forms jets by associating together particles whose trajectories lie within a circle of specific radius  $R$  in space.

#### Initial cone algorithms

In the 1980's, the UA1 and UA2 experiments were the first to use cone-based algorithms in a  $p\bar{p}$  collider. The UA1 algorithm started ordering in decreasing transverse  $E_T$  the cells of the calorimeter with  $E_T > E_{Tseed}$ . The value used for  $E_T$  seed was 2.5 GeV. The cell with the highest transverse energy initiated the first jet. The next cell was added to the first if it was within a distance  $R_0 = 1$ . If the cell was outside this radius then a new jet was initiated. This procedure was repeated until all cells above the  $E_{Tseed}$  threshold had been assigned to a jet. Finally, the cells with  $E_T < E_{Tseed}$  were then added to each jet if  $R < R_0$ . In contrast UA2 used another approach. In the UA2 algorithm the cluster did not have a limited size in  $(\eta, \phi)$  space. Once all the calorimeter cells were ordered in decreasing  $E_T$ , starting for the highest one, all the neighboring cells were joined into the cluster if the  $E_T$  exceeded a given threshold. In that case, the threshold was 0.4 GeV.

#### The CDF cone algorithms

Figure 3.3 (top left) shows a schematic picture of a jet containing tracks and calorimeter energy deposits. Calorimeter deposits are represented as red towers, tracks are red lines and the jet is represented by everything which is inside the yellow cone.

All current cone algorithms are based on the so-called ‘Snowmass Algorithm’ which defines both the stability conditions and the properties of the jets [57]. The stability condition is usually defined as the requirement that the change of the jet properties from small variations in the location of the center of the jet should be minimal. The properties of a cone algorithm jet are:

- The cone size,  $R$ , usually 0.4 or 0.7 but sometimes 1.0, depending on the analysis.
- The energy ( $E$ ) and momentum ( $p$ ).

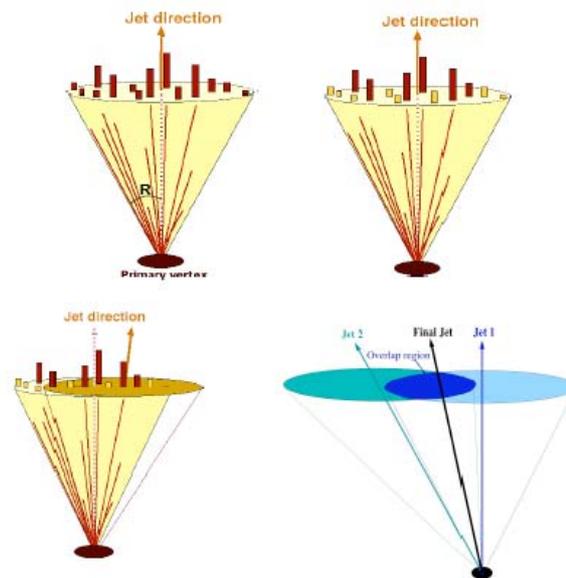


Figure 3.3: Schematic drawing of a cone algorithm jet showing the tracks and towers inside the jet cone (top left) and illustration of the seed mechanism (top right) which selects tracks above a certain threshold (in black) as potential centers for the jets. The ratcheting process (bottom left) constraints all the seed towers initially inside the jet cone to remain inside the jet even if they are outside the final cone. The splitting/merging procedure when the jets are merged according to the amount of overlap region they share (bottom right).

- The direction  $(\eta, \phi)$ .

The cone size defines the spatial extent (in rapidity,  $\eta$ , and  $\phi$ ) of the jet i.e. the region covered by towers considered to be inside the jet. To be more precise the cone size is defined as  $R = \sqrt{\eta^2 + \phi^2}$  which is the angular separation in  $(\eta, \phi)$ -space between the jet direction and the outer edge of the cone.

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

$$P^{jet} = \sum P^{tow}$$

where  $P^{jet}$  is the 4-vector momentum  $(p_x, p_y, p_z, E)$  of the jet and  $P^{tow}$  is that of the towers [61]. The sum is over all towers inside the jet. This defines the energy and direction of the jet.

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

$$E_T^{jet} = \sum E_T^{tow}$$

and the jet direction as

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

and

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

where  $\eta$  is the pseudo-rapidity.

In an ideal world, to find the final jets, each tower of each event should be considered as a potential center of a jet. The algorithm would iterate over each possible jet until the local minimum of the jet potential would be found. The minimum of the potential is the point where the change in the kinematic jet quantities used by the algorithm, when moving the center of the jet in space, is the smallest. These local minima would then be the center of the jets. In the real world there are a number of compromises which must be made in order to cope with limited computing power and high trigger decision rates. These are

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

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

Table 3.1: Comparison between the JetClu and MidPoint cone algorithms

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

Seed towers are defined as towers with a higher threshold (usually set to 1 GeV) than the threshold to be included in a jet (usually set to 0.1 GeV). Only the seed towers are considered as potential centers for the jets as illustrated in figure 3.3 (top right). This greatly reduces the time needed to find stable cones. There are a number of problems with the use of seed towers, some of which are discussed below.

Ratcheting is a compromise used in the JetClu algorithm only. It constrains all seed towers initially inside a jet to remain inside a jet even if they are outside the final jet cone. This leads to jets which can have strange shapes and high amounts of transverse energy outside the cone radius, see figure 3.3 (bottom left). There is no reasonable way to simulate the role of ratcheting in theoretical calculations, since its role depends in detail on the level of soft radiation in the event as well as on the details of the calorimeter.

The final compromise mentioned here is the merging/splitting procedure. It is possible, because of the use of seed towers, that, in the final state, two or more jets can be separated by less than twice their cone size, see figure 3.3 (bottom right). In this case there are a number of towers which are inside both jet cones, something which would have no physical meaning. In both cone algorithms described here, this problem is dealt with by introducing a merging/splitting fraction,  $f_{merge}$ . If the fraction of energy of a jet which overlaps with another jet is larger than this fraction  $f_{merge}$ , the jets are merged together into a single jet. All the towers in the initial jets then belong to the new jet and the kinematic quantities of the jet are re-calculated from this new set of towers. If this overlap fraction is smaller than  $f_{merge}$ , the towers in the overlap region are individually associated with the jet whose center is closest

and again the kinematic quantities are re-calculated. This fraction  $f_{merge}$  can be set to different values, the default for both JetClu and MidPoint is 75%. For JetClu the number of these iterations is limited to one, no more than two jets may be merged together, leading to a maximum distance between the outermost towers and the jet axis of just less than twice the cone size. The JetClu algorithm has so far been the preferred jet reconstruction algorithm at CDF Run II principally because of the desire for backward compatibility with Run I results.

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

### ‘Hadron Level’ jets in Monte Carlos

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

### 3.2.2 The ‘correct’ jet algorithm

To illustrate one of the problems with jet algorithms, we consider the situation where there are two localized clusters of towers (or two tracks) that are separated by just less than twice the cone size [61]. This situation is illustrated in the left hand diagram of figure 3.4. Seeds would be placed at each of these clusters and stable cones would be found centered approximately on each of the clusters, leading to two distinct jets. At parton level, this topology could come from one parton with high momentum which splits into two roughly equal energy partons, with a fairly large angular separation, before hadronization occurs, in which case the ideal jet solution would be one where both these clusters are included in a single jet centered roughly halfway between them. The MidPoint algorithm partially solves this problem by artificially placing a seed

tower exactly half way between all of the seed towers (hence the name MidPoint). This is shown on the right hand diagram of figure 3.4.

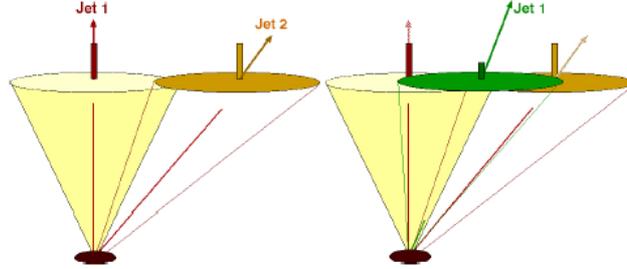


Figure 3.4: Schematic drawing of the jet reconstruction of a two parton event, where the partons are separated by just less than twice the cone radius. In JetClu the event would be reconstructed as two jets (left) whereas in MidPoint the use of the extra seed leads to one single jet (right). JetClu is not Infrared safe because the emission of a soft parton between the two hard partons (right) yields a different jet configuration: the same as the default MidPoint configuration. MidPoint on the other hand is infrared safe

In fact it turns out that in practice it is only necessary to place a midpoint seed between seeds separated by less than twice the cone radius. In MidPoint, the stable configuration would be one in which the center of the jet is roughly half-way between the two partons. Now let's add a very soft gluon emitted somewhere between these two partons and see what happens. The configuration is now one with three seeds centered on each of the partons. JetClu will reconstruct this configuration as one jet, centered half-way between the two high  $p_T$  partons, just as MidPoint did for the case without the soft gluon. MidPoint will reconstruct exactly the same as what it reconstructs for the leading order process. From this, one can see that JetClu is not infrared safe.

In fact there is one small subtlety. Experimentally, the algorithm most likely doesn't see the soft gluon; the  $p_T$  of the soft gluon is most probably much smaller than the threshold used by the algorithm for towers to be included in the jet. So in that case JetClu would still reconstruct two jets as the third calorimetric cluster will be very small.

Of course this still doesn't answer the question of which of the configurations, two jets centered each on one of the hard partons or a single jet centered half-way between the two, is the 'correct answer'. There is no correct answer because there is no way to know, experimentally, if these partons should be considered as a single gluon jet or as two parton jets!

The best solution would probably be the use of a  $k_T$  type of algorithm, which doesn't define a real jet-cone. The performance of such algorithms has been tested for the first time at a hadron collider with the measurement of the inclusive jet cross section [62] and [63]. Figure 3.5 shows a comparison between two independent measurements of the inclusive jet cross section performed at CDF using a  $k_T$  algorithm (left) and a cone algorithm (right). The two measurements are in perfect agreement over a large  $p_T$  and  $\eta$  range.

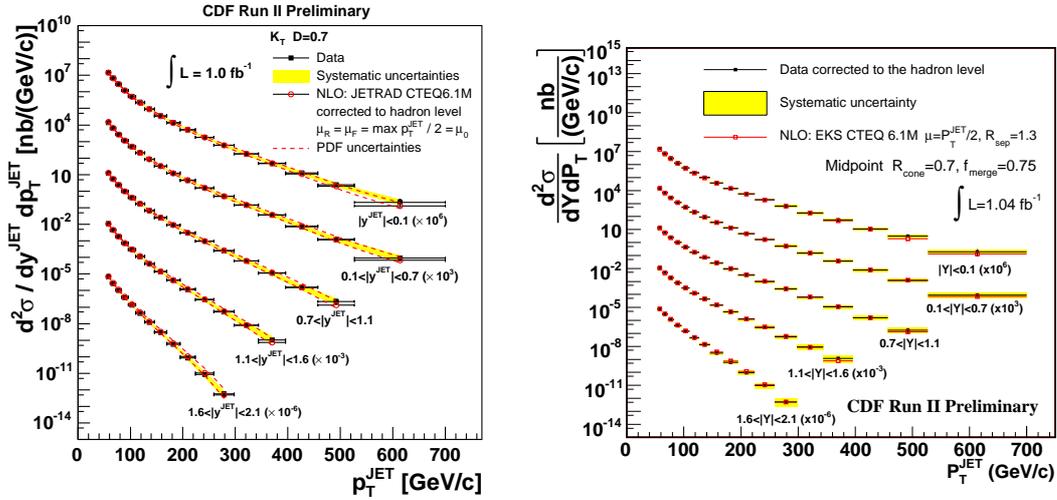


Figure 3.5: Measurement of the inclusive jet cross section at CDF in five rapidity regions:  $k_T$  jets (left), MidPoint jets (right). Data are compared to NLO predictions.

The application of a  $k_T$  type algorithm to heavy flavor jets is nevertheless limited by the problem of  $b$ -tagging: in order to identify a  $b$ -jet it is still necessary to define a cone inside the jet, to look for the  $B$ -hadron. This fact would certainly limit the advantages of a  $k_T$  type algorithm. Discussion is ongoing, to propose different types or reconstruction algorithm suitable for  $b$ -jets, and allowing a transparent comparison to theoretical predictions [64] and [65].

### 3.3 Evaluation of the Primary vertex position

The position of the primary interaction vertex is a necessary information to the reconstruction of the high-level variables in the event, (jets for example) and to ensure that the results from the secondary vertex tagging algorithms are correct. A first estimate on the position of the interaction point is measured in the  $z$  direction. Information on the  $z$  coordinate is sufficient to rebuild offline jets, which at the L3 trigger are

reconstructed assuming the primary vertex coordinates to be zero. In fact, at the stage of jet reconstruction, information related to the transverse plane ( $xy$ ) is not used, since to a first approximation the vertex coincides with the beam axis. Once  $z$  is defined, each calorimetric tower 4-momentum is recalculated and it is used to reconstruct the jets. The  $z$ -coordinate is used as a starting point to measure the 3-dimensional position of the vertex: this information is necessary to correctly identify decay vertexes ('secondary vertexes'), and measure their distance from the primary interaction point to estimate the life-time of the decaying particle.

### 3.3.1 $Z$ -vertex reconstruction

The  $z$ -vertex algorithm is seed-driven and in order to reduce fakes, it uses vertexes created upstream as input and it associates them to tracks reconstructed in the event. Two list of primary vertex candidates are considered, the one created by *PVFinder* - a pre-tracking finder which attempts to find vertexes using the 2D hits from small-angle stereo layers - and a COT based vertex collection *COT standalone vertexes*. The collections are then merged to provide a single list of seeds. To reconstruct primary vertexes the  $Z$ -vertex finder uses a subset if tracks. The tracks are selected and classified as follows:

- COT-only and COT-SVX tracks with at least 3 axial and 3 stereo COT segments (good COT tracks) and  $\chi^2(COT)/DOF < 4$ : track quality <sup>1</sup>;
- SVX-only tracks with at least 5 axial and 3 ( $z$ +small angle stereo) SVX hits and  $\chi^2(SVX)/DOF < 8$ : track quality 4. COT+SVX tracks found by inside out algorithm are required to pass the same requirements as good SVX tracks;
- track with at least 2 axial and 2 stereo COT segments and  $\chi^2(COT)/DOF < 4$  and at least 4 axial and 3 ( $z$ +small angle stereo) SVX hits and  $\chi^2(COT)/DOF < 8$ :track quality 2.

The highest  $p_T$  COT track also makes a vertex, to increase the efficiency of vertex finding. In this case, the quality assigned to the track is 12. All the participating tracks are also required to have impact parameter  $|d_0| < 1$  cm. The list of the reconstructed seed vertexes is pruned: 2 seeds are merged if the distance between them is less than 3 cm. For each vertex, the quality code is defined as  $\Sigma track_{quality}$ : a good vertex has a quality above or equal to 12. A list of  $z$ -vertex candidates is

---

<sup>1</sup>The quality of the track is a flag defined in order to have good vertexes only with specified combinations: *i.e.* since a good vertex should have quality 12, at least 2 quality-6-tracks must be used

made and in the case that two or more are found, it is chosen the vertex with highest total scalar sum of transverse momentum, in order to maximize the efficiency.

The efficiency of the algorithm, tested on Monte Carlo events, is found to be above 99% for  $|Z| < 50$  cm for inclusive jet samples [66].

### 3.3.2 Primary Vertex Finding

The  $z$ -coordinate of the primary vertex and the average beam-line are used as starting point of the algorithms providing the vertex position in the  $xy$  plane. By default, the beam-line is the time-dependent SVX beam-line read from the CDF database.

Tracks within  $\pm 1$  cm window from the  $z$ -seed are considered: the procedure start fitting a vertex using all the tracks in this window, where tracks are also required to have an impact parameter significance, relative to the average beam position,  $|d_0/\sigma_{d_0}| < 3$ :  $d_0$  is the impact parameter and  $\sigma_{d_0}$  is its uncertainty (including the beam position error). If a track added to the fit results giving a  $\chi^2 > 10$  it is removed from the procedure (‘pruning’). The 3-dimensional primary vertex position is determined, with a precision of the order of  $10 - 20$   $\mu m$ , by the fit performed on the tracks passing the initial pruning, iterated until no tracks below the  $\chi^2$  cut are found. In case no tracks survive the selections, the beam position is assumed as the primary vertex. Figure 3.6 shows the distributions of the primary vertex  $x$  and  $y$  for the events in the *HIGH\_PT\_BJET* dataset (described below): the interaction point in the  $xy$  plane changes in time and does not coincide with the origin coordinates  $(0, 0)$ .

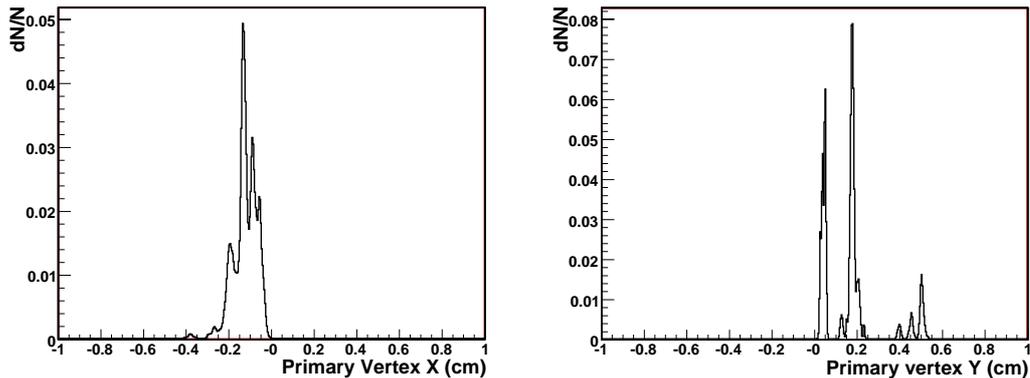


Figure 3.6: Primary vertex  $x$  coordinate (left) and  $y$  coordinate (right) for events selected using the *HIGH\_PT\_BJET* trigger: the interaction point changes with time in the  $x - y$  plane and doesn’t coincide with the origin coordinates  $(0, 0)$ .

## 3.4 Data samples

The data sample used in this analysis is selected using the **HIGHT\_PT\_BJET** trigger. A list of good runs has been compiled using the CDF Data Quality Monitoring system to insure a meaningful reconstruction of calorimeter, tracking and SVT trigger data. Applying this requirements results in a total integrated luminosity of  $260 \pm 16 \text{ pb}^{-1}$ .

Additional control samples including inclusive muon samples have been used to study the performance of the  $b$ -tagging algorithm.

### 3.4.1 The HIGH\_PT\_BJET trigger

High  $p_T$   $b$ -jets have already been used in many CDF Run I analyses to search for new particles such as Higgs or SUSY. In general the data were selected requiring high  $p_T$  leptons or multi-jet triggers. However, leptons trigger datasets have the disadvantage that the searches are limited to particles associated with  $W/Z$ . On the other hand, multi-jet triggers must have relatively high thresholds to reduce the background due to light quark or gluon jets.

In Run II, it is possible to select high impact parameter tracks at L2, using the SVT trigger. It is therefore possible to design a high  $p_T$   $b$ -jet trigger by associating a large impact parameter track to the calorimeter cluster. This requirement reduces the jet energy threshold needed to select  $b\bar{b}$  events. It also reduces the necessity of applying a pre-scale to the trigger.

The HIGH\_PT\_BJET trigger evolved overtime, with increased luminosity and with a better understanding of the detector. The first versions simply required two jets and two SVT displaced tracks in the event:

- **Level 1:**
  - Two central hadronic 5 GeV towers
  - Two XFT tracks with  $p_t > 2 \text{ GeV}$
- **Level 2:**
  - Two central calorimeter clusters with  $E_t > 20 \text{ GeV}$  and  $E_T > 10 \text{ GeV}$
  - Two SVT tracks with  $|d_0| > 80 \mu m$
  - at least one SVT track with  $|d_0| > 100 \mu m$
- **Level 3:**
  - Two  $|\eta| < 1.5$  jets with  $E_t > 20 \text{ GeV}$  and  $10 \text{ GeV}$

- at least one jet tagged as a heavy flavour jet according to the SecVtx algorithm, described in the next chapter

Table 3.2 shows the trigger cross section, and the consequent rate for a luminosity of  $10^{32} \text{ cm}^{-1} \text{ s}^{-1}$ , estimated in [67] for the first trigger version.

Trigger Level	Cross Section	Rate
L1	$27 \mu b$	2.7 kHz
L2	$200 \text{ nb}$	20 Hz
L3	$41 \text{ nb}$	4 Hz

Table 3.2: Summary of cross sections and rates of the first version of the high pt  $b$ -jet trigger [67].

Thresholds for the cluster transverse energy and track impact parameter have changed subsequently and recent versions apply a exclusive matching between the jets and the SVT tracks (two L2 cluster each associated to a SVT track). Details of the trigger implementations corresponding to different run periods are given in table 3.3.

Each change had a direct consequence on the trigger cross sections and rates. A few examples are shown in figures 3.7 and figures 3.8. Figure 3.7 represents the trigger cross sections for two different implementations of the L2 trigger path corresponding to runs up to number 152949 and for the later version corresponding to runs in the range 178407 – 179866. Figure 3.8 shows the early (runs below 152949) and late (runs in 175154 – 178744) versions of the L3 trigger. Both figures show the same reduction in the trigger cross section due to tighter selection. Since run 177486 an additional L2 pre-scaled path has been introduced to cope with the increased luminosity delivered by the Tevatron. However, this latest path has not been considered in this analysis.

Figure 3.9 shows the SVT track  $d_0$  averaged over the run number. As it will be better explained in section 3.5 a  $d_0$  average value as large as the one observed for the two runs 144424 and 178853 is indicative of some problems with the beam-line position. Those runs have been excluded from the good run list.

### 3.4.2 Monte Carlo samples

Monte Carlo event samples are used to study background and understand detector response: to calculate jet energy corrections, to measure selection efficiency ( $b$ -tagging efficiency in particular) and to get the  $b\bar{b}$  jet purity of the final sample.

Run range	Trigger		
	L1	L2	L3
14088 – 152949	2 clus. $E_T > 5$ GeV 2 XFT tracks: $\Sigma p_T = 0$ GeV $p_T > 2.04$ GeV/c $0 < \Delta\phi < 180$ 4 XFT layers	2 L1 clus. $E_T > 5$ GeV 2 SVT tracks: $p_T > 2$ GeV/c $ d_0  > 100 \mu m, \chi^2 < 25$	2 jets $E_T > 20$ GeV 2 COT-SVT tracks: $p_T > 2$ GeV/c $ d_0  > 100 \mu m$
152953 – 164259	no change	L2 cluster: $E_T > 15$ GeV, $ \eta  < 1.5$	no change
164261 – 177485	no change	no change	COT-SVT track confirmed with Si: $ d_0  < 80 \mu m$
168086 – 178744	no change	SVT-cluster match: $\Delta\phi < 0.8$	no change
178757 – 179866	no change	SVT $\chi^2 < 15$	no change

Table 3.3: Summary of the HIGH\_PT\_BJET L1, L2, L3 changes during the different runs periods covered by this analysis.

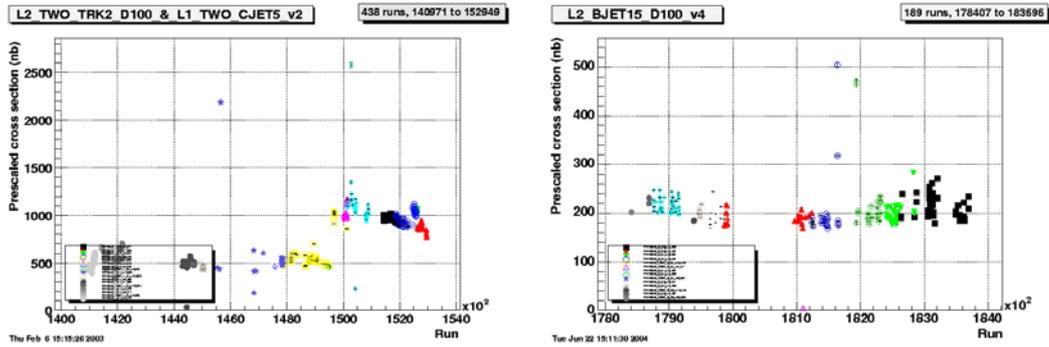


Figure 3.7: Cross section for trigger path L2\_TWO\_TRK2\_D100\_&\_TWO\_CJET\_v2 corresponding to runs up to 152949 (left). Cross section for the trigger path L2\_BJET15\_D100\_v2 corresponding to runs in the range 178407 – 179866.

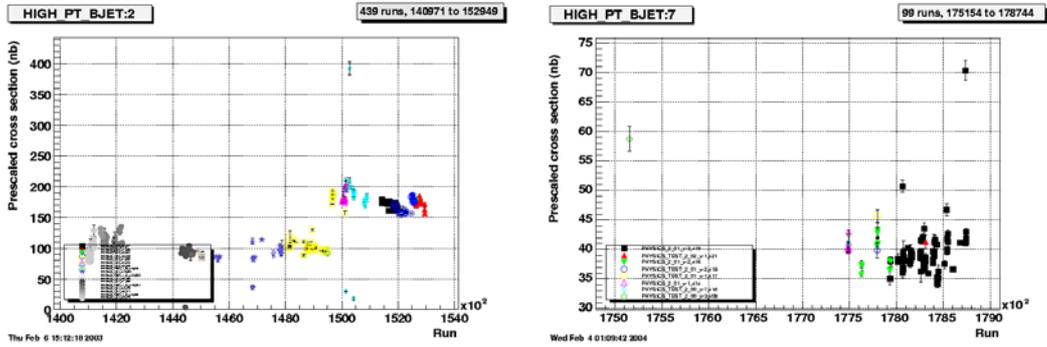


Figure 3.8: Cross section for trigger path L3\_HIGHPT\_B\_JET\_v2 corresponding to runs up to 152949 (left). Cross section for the trigger path L3\_HIGHPT\_B\_JET\_v4 corresponding to runs in the range 175154 – 178744.

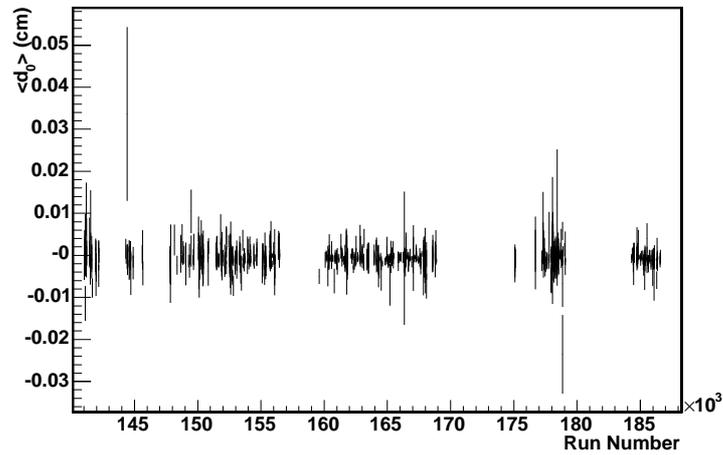


Figure 3.9: The mean  $d_0$  as a function of run number for events selected using the *HIGH\_PT\_BJET* trigger.

The generated events are passed through a full detector simulation, based on GEANT3 [68], where the GFLASH package [69] is used for the energy deposition model in the calorimeters. All the event variables and measurable quantities (such as jets or tracks) are reconstructed using the same software version as for data. Different generic di-jet event samples are generated using PYTHIA 6.203 [9] and HERWIG 6.4 [10]. They are listed in table 3.4 and 3.5; CTE5L PDFs functions are used. The samples differ for the threshold applied to the transverse momentum of the hard scattering  $\hat{p}_T$ . Before any selection, their content in terms of  $b\bar{b}$  ( $c\bar{c}$ ) pair production at the level of the hard scattering is on average about 4% (11%), for both PYTHIA and HERWIG samples.

PYTHIA events are used as default samples and generated according to a specific tuning ('Tune A', [71]) based on Run I studies to better simulate the underlying event. HERWIG events are used as control samples.

In addition, samples filtered in order to have at least one  $b$ -quark in the initial parton list have been produced, using both PYTHIA and HERWIG: those samples are completely equivalent to the inclusive ones but for the  $b$ -filter and they have been used to study  $b$ -tagging efficiency and to measure the  $b\bar{b}$  content of the data.

In order to have Monte Carlo events as realistic as possible, those are simulated using a set of parameters corresponding to real data runs<sup>2</sup>, as any difference, between the data and the simulation, can imply different  $b$ -tagging performance it needs to be taken into account. A particularly important issue concerns the simulation of the L2 trigger, and the SVT in particular. A correct alignment of the beam-line is crucial to insure good track reconstruction by the SVT trigger. In fact the SVT calculates the impact parameter of the track correcting on-line for the real beam position. A special study on the simulation of the beam-line in all the Monte Carlo samples has been performed. Details are given in the appendix A.

## 3.5 Event selection

The event selection is chosen in such a way that the trigger is 100% efficient on all the events passing the offline selection<sup>3</sup>.

The strategy proposed in this analysis is to avoid the measurement of trigger efficiency, to apply offline (on the reconstructed quantities) a much tighter selection and

---

<sup>2</sup>Beam positions are simulated using information on 732 runs, corresponding to an integrated luminosity of about 250 pb<sup>-1</sup>. For details see [70].

<sup>3</sup>A study performed on an inclusive jet sample (jet  $E_T > 20$  GeV) yields to only 0.1% events that satisfy the selection criteria and have the HIGH\_PT\_BJET trigger bit off. A detailed study on the systematic uncertainties will show that this effect is negligible and the trigger efficiency on selected events can be considered 100%.

Di-jet sample	Cross Section (pb)	Luminosity (pb <sup>-1</sup> )
PYTHIA		
$\hat{p}_T > 18 \text{ GeV}/c^2$	$49.8 \cdot 10^6$	0.1
$\hat{p}_T > 40 \text{ GeV}/c^2$	$1.33 \cdot 10^6$	3.76
$\hat{p}_T > 60 \text{ GeV}/c^2$	$0.178 \cdot 10^6$	5.58
$\hat{p}_T > 90 \text{ GeV}/c^2$	$20.5 \cdot 10^3$	75
HERWIG		
$\hat{p}_T > 18 \text{ GeV}/c^2$	$39.29 \cdot 10^6$	0.02
$\hat{p}_T > 40 \text{ GeV}/c^2$	$1.059 \cdot 10^6$	0.94
$\hat{p}_T > 60 \text{ GeV}/c^2$	$0.141 \cdot 10^6$	7.1
$\hat{p}_T > 90 \text{ GeV}/c^2$	$16.3 \cdot 10^3$	61

Table 3.4: Generic di-jet Monte Carlo samples.

$b\bar{b}$ Sample	Cross Section (pb)	Luminosity (pb <sup>-1</sup> )	efficiency
PYTHIA			
$\hat{p}_T > 10 \text{ GeV}/c^2$	$563.6 \cdot 10^6$	0.073	0.0351
$\hat{p}_T > 18 \text{ GeV}/c^2$	$49.8 \cdot 10^6$	0.5	0.0426
$\hat{p}_T > 40 \text{ GeV}/c^2$	$1.33 \cdot 10^6$	15.22	0.0558
$\hat{p}_T > 60 \text{ GeV}/c^2$	$0.178 \cdot 10^6$	48.077	0.0591
$\hat{p}_T > 90 \text{ GeV}/c^2$	$20.5 \cdot 10^3$	1014	0.0623
$\hat{p}_T > 120 \text{ GeV}/c^2$	$3.967 \cdot 10^3$	2115	0.0615
HERWIG			
$\hat{p}_T > 18 \text{ GeV}/c^2$	$39.29 \cdot 10^6$	0.108	0.0492
$\hat{p}_T > 40 \text{ GeV}/c^2$	$1.059 \cdot 10^6$	3.89	0.0571
$\hat{p}_T > 60 \text{ GeV}/c^2$	$0.141 \cdot 10^6$	32.9	0.0583
$\hat{p}_T > 90 \text{ GeV}/c^2$	$16.3 \cdot 10^3$	299.8	0.0631

Table 3.5:  $b\bar{b}$  di-jet Monte Carlo samples: they are obtained from generic di-jet samples to which a generator level requirement ( $b\bar{b}$  event) is applied. The last column represent the generator level selection efficiency.

to calculate directly a total offline efficiency. In fact, an estimate of the trigger efficiency results difficult using Monte Carlo events or the data. In the case of simulated events, any measurement relies on the quality of the trigger simulation. Using the data requires an unbiased and unrescaled reference sample. The same approach has been used in [74]. The advantage, as explained in the next chapter, is that a single trigger-tagging efficiency can be calculated per jet using the Monte Carlo simulation. The efficiency is then corrected using a scale factor, specifically measured to take into account any difference between real and simulated events.

The offline selection is based on the following criteria:

- At least one good primary vertex, reconstructed in the range  $|Z_v| < 50 \text{ cm}$ . Protons and anti-protons are delivered in bunches which extend for about  $50 \text{ cm}$  in the beam-pipe direction. As a consequence, the distribution of the  $z$ -vertex position is approximately Gaussian with  $\sigma \approx 30 \text{ cm}$ , centered near  $z = 0$ . Thus, a selection on  $z$  assures good energy measurement of the jets; besides, events are in this way selected within the acceptance limit in the  $b$ -tagging algorithm. Finally, the selection on primary vertex allows the removal of cosmic rays.
- Two  $b$ -tagged jets with  $E_t > 30 \text{ GeV}$  in the central region of the detector ( $|\eta| < 1.2$ ). The jets are reconstructed using the JetClu algorithm with a cone of radius  $R=0.4$ . Each of the jets is required to be geometrically matched (the difference in the  $\eta\phi$  space is  $\Delta R < 0.4$ ) to trigger level jet quantities:
  - a L3 jet with  $E_t > 20 \text{ GeV}$ ;
  - a L2 cluster with  $E_t > 15 \text{ GeV}$ ,  $|\eta| < 1.5$ .

Figure 3.10 shows the efficiencies for matching the jet to the trigger objects: the L2 cluster and the L3 jet. It is interesting to note that L3 jets are reconstructed on-line assuming the primary vertex is in the  $z = 0$  position, whereas offline jets are reconstructed around the real primary vertex.

- Two COT tracks  $p_t > 2 \text{ GeV}$ ,  $|\eta| < 1.5$ . Each one of them is required to be associated ( $\delta\rho < 0.00015$ ,  $\delta\phi < 0.015$ ) to a SVT trigger track with impact parameter  $|d_0| > 120 \mu\text{m}$ . The cut on the SVT track impact parameter is tighter than the threshold on the same quantity applied at trigger level (from  $80 \mu\text{m}$  to  $100 \mu\text{m}$  depending on the trigger version). The two COT tracks are also required to be associated ( $\delta R < 0.015$ ) to two tracks reconstructed in the silicon detectors and having  $|d_0| > 120 \mu\text{m}$ . This cut is applied on the impact parameter that has to be corrected for the position of the beam-line, following the same procedure applied on-line by the SVT trigger, as explained in section 2.5. Figure 3.11 show the impact parameter of offline reconstructed tracks

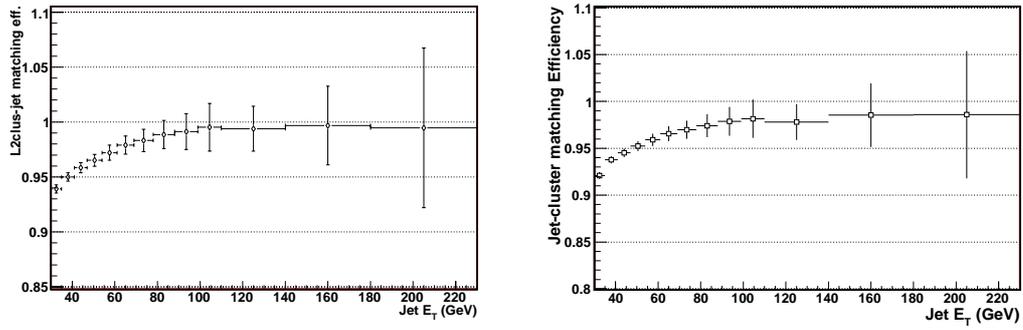


Figure 3.10: L2cluster -jet matching efficiency (left). L3jet- offline jet matching efficiency (right).

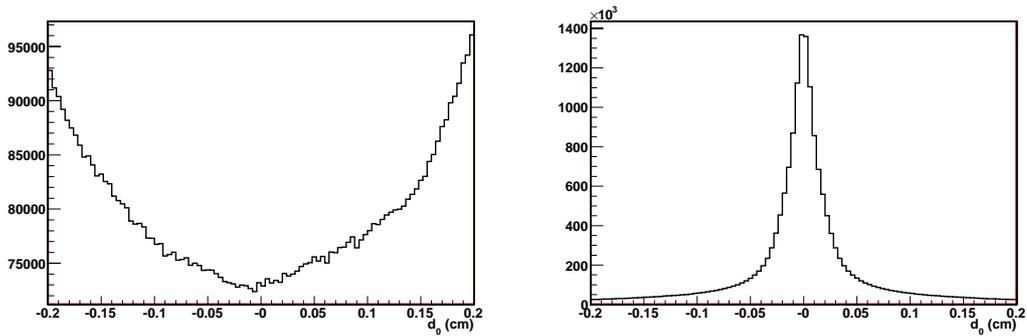


Figure 3.11: Uncorrected (left) and corrected (right) impact parameter after the offline reconstruction.

before and after correction. For SVT tracks the same correction is performed on-line by the SVT trigger. The number of tracks that satisfy this selection per event is shown on the left side of figure 3.12

- Each one of the two tagged jets in the event is geometrically matched to one of the two tracks selected above. The matching is required to be 3-dimensional, (*i.e.* there is a cut on the distance between the track and the jet axis in the  $\eta\phi$  plane,  $\Delta R < 0.3$ ). This cut makes sure each of the two jets contains one of these tracks inside its cone. It is to be noted that at trigger level the association was done in only one dimension,  $\phi$  (*i.e.* requiring  $\Delta\phi < 0.8$ ). Figure 3.12 shows, on the right, that requiring a three dimensional matching between a jet and a SVT track has an efficiency of approximately 25% for generic jet, that rises to  $\approx 35\%$  for tagged jets.

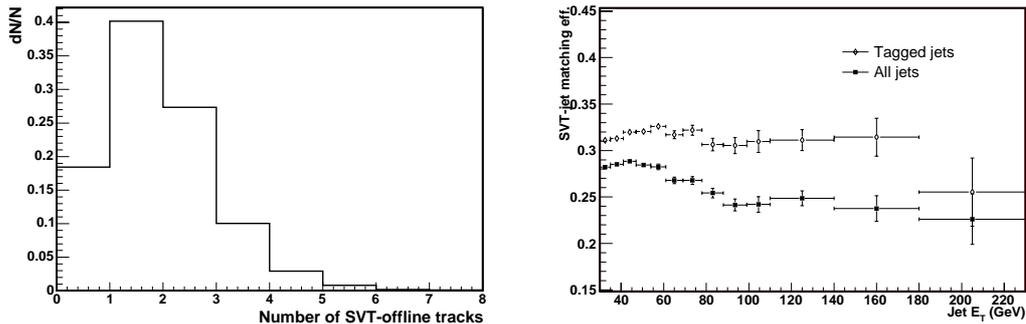


Figure 3.12: Left: Number of tracks per event passing the offline selection described in the text. Right: SVT track -jet matching efficiency. The empty diamonds represent SVT-tagged jets, the full squares are generic tagged jets.

A tagged jet containing a SVT track according to the definition above will be referred to in the following as a *SVT-tagged*. As already stated this definition is chosen so that events with two such objects always pass the trigger. The advantage of this choice is that it allows to reasonably neglect the problem of calculating trigger efficiency: a sort of trigger and tagging efficiency in one single step is measured instead. Details about the tagging algorithm and the SVT-tagging efficiency will be given in the next chapter.

### 3.6 Jet energy corrections

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

Sampling calorimeters are non-compensating. The hadronic calorimeter was calibrated based on charged pions from a test beam with a transverse momentum of 57 GeV/ $c$ . Only pions which did not interact with the EM calorimeter were included in the calibration. In real jets however, a large fraction of hadrons do interact with the EM calorimeter. Because the EM calorimeter is calibrated based on electrons its response to hadrons is lower. This fact reduces the overall response to single hadrons. The effect is larger for low transverse momentum particles because they interact in the EM calorimeter more often, and it can therefore contribute to the non-linearity of the calorimeter response to hadrons. Hadronic showers have a larger fraction of neutral pions when the incident hadron has a higher transverse momentum. Because the calorimeter has a higher response to EM showers (i.e.,  $\pi^0$  decays), this also contributes to the non-linear nature of the calorimeter response to charged hadrons.

In general the calorimeter response goes up as the transverse momentum of the incident hadron increases and it is not linear. This causes a systematic shift down in the energy response to jets because they include multiple hadrons with lower transverse momentum, rather than one hadron with the full jet  $E_T$ .

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

The CDF collaboration has determined a set of corrections [75], divided into different steps, to get results corrected to the hadron or parton level. The sketch in figure 3.13 shows the general idea of it. The correction from the calorimeter level jet to the particle level jet accounts for detector effects and is necessary for a meaningful comparison to theory.

For the second step, from particle level jets to parton level jets, the situation is not so clear. The correction depends on the fragmentation and hadronization processes and it can be model dependent. Here we choose not to correct further to the parton level. The data are in fact compared to Monte Carlo simulation that produces particle level jets according to its own hadronization model (PYTHIA or HERWIG).

So the calorimeter jet  $E_T$  is corrected using the following steps. First, an  $\eta$ -dependent relative correction is applied to the data and MC in order to equalize the

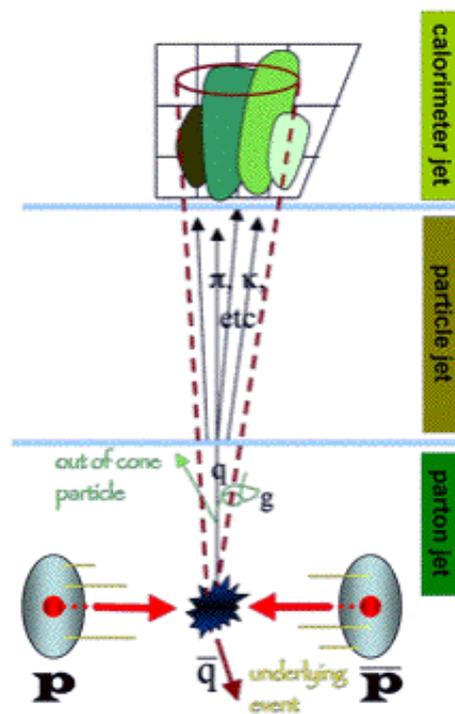


Figure 3.13: A representation of the different levels of corrections that are necessary to compare the jet measurement to a theoretical prediction.

response of the CDF calorimeters to jets in  $\eta$ . The equalized jet  $E_T$  is corrected for the pile-up effect.

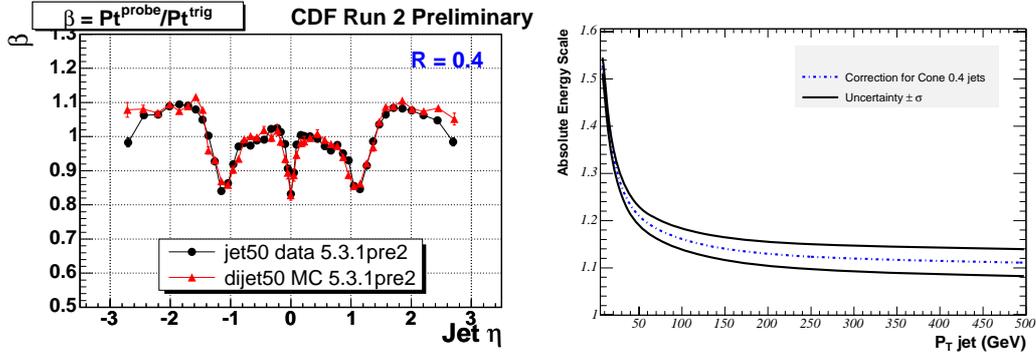


Figure 3.14: Relative correction (left) and absolute energy correction as a function of jet  $E_T$  (right)

Then, the absolute correction is applied to correct on average for the hadron energy that is not measured by the calorimeter. The ‘ $\eta$ -dependence’ correction and the ‘absolute’ correction are shown in figure 3.14. These corrections are applied jet by jet to all the events in this analysis.

After that, the hadron and calorimeter level jet  $E_T$  distributions are compared in Monte Carlo to derive a bin-by-bin correction in order to remove resolution effects. This procedure is called unfolding and it will be better described in section 5. At this point, the data have been corrected to the particle level.

This study doesn’t apply corrections relative to the underlying event, that accounts for the presence of ”spectator” particles that don’t take part in the interaction, or ”out of cone” corrections that calculate the radiation lost out of the clustering cone and correct the jet energy to its parent parton energy.

### 3.6.1 Additional corrections on tagged jets

The generic corrections described above are derived for ordinary jets and do not take into account different factors originating from the presence of a  $b$ -quark inside the jet.  $b$ -flavored jets undergo different fragmentation, and it is necessary to take into account the presence of the  $B$ -hadron decay inside the jet: about 23% of  $B$ -hadrons decay semileptonically, so there is an underestimation of the parton energy due to lost neutrinos. Also, transverse energy corrections for tagged  $b$ -jets might be biased by the selection on charged tracks applied by the tagging algorithms.

Overall a residual difference between calorimeter jets and hadron level jets, in

the case of  $b$ -jets, tagged jets and SVT-tagged jets, is expected even after generic jet corrections were applied.

Figure 3.15 proves that, in fact, this is the case. It shows the ratio of calorimeter jet  $E_T$ , corrected using generic corrections to hadron level jet  $E_T$ , measured as the sum of the energies of the particles within the jet cone. This ratio is averaged as a function of the calorimeter level jet  $E_T$  and hadron level jet  $E_T$ . In both cases the ratio is different from one. The distributions are obtained, using PYTHIA events, for tagged jets and SVT-tagged jets in the central calorimeter region  $|\eta| < 1.2$ . Each calorimetric jet is associated to a hadronic jet requiring that the distance between the two satisfies  $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} < 0.4$ .

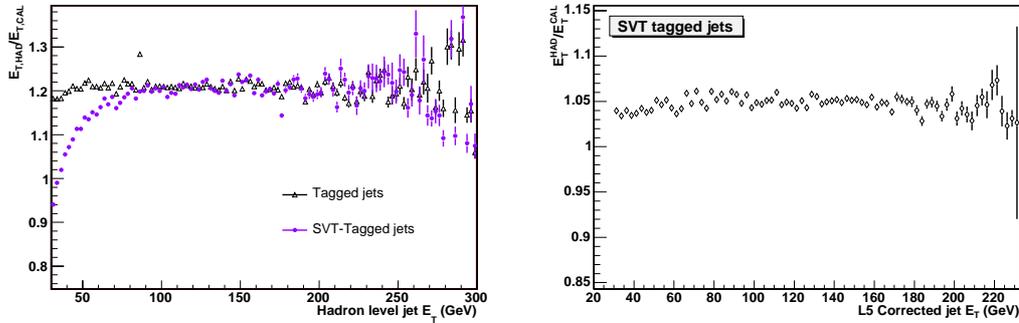


Figure 3.15: Left: Ratio of hadronic jet  $E_t$  to calorimeter jet  $E_t$  as a function of hadronic jet  $E_t$  for tagged jets (black) and SVT-tagged jets (blue). The different behavior at low  $E_t$  points out the bias introduced by the matching at trigger level of the jet to L2 cluster ( $E_t > 15$  GeV) and L3 jets ( $E_t > 20$  GeV). Right: ratio of hadronic jet  $E_t$  to calorimetric jet  $E_t$  as a function of calorimetric jet  $E_t$  for SVT-tagged jet.

The fact that the ratio is larger than one means that the corrected jet  $E_T$  doesn't exactly describe the 'true' particle level energy in the case of tagged jets or SVT-tagged jets. An additional correction specific to tagged jets is necessary. It is derived using the following procedure:

- Hadronic and calorimetric jets are associated if  $\Delta R < 0.4$ ;
- The correlation  $E_{T,CAL}$  versus  $E_{T,HAD}$  is measured, and fit to a polynomial;
- The average correction is applied jet by jet;

The choice of calculating corrections for tagged jets (and not  $b$ -jets) has the advantage of allowing a straightforward application to data jets (where no information

about the flavor is obviously available). However it is important to take into account the fact that not all positively tagged jets contain a  $b$  hadron inside their cone. The residual difference between tagged jets and tagged  $b$ -jets, will be taken into account by the unfolding procedure described in chapter 5.

Figure 3.16 shows the calorimeter vs hadronic jet  $E_t$  correlation ( $E_{T,CAL}$  vs  $E_{T,HAD}$ ). The calorimeter energy range is divided in 5 GeV bins, figure 3.16 shows the  $E_{T,HAD}$

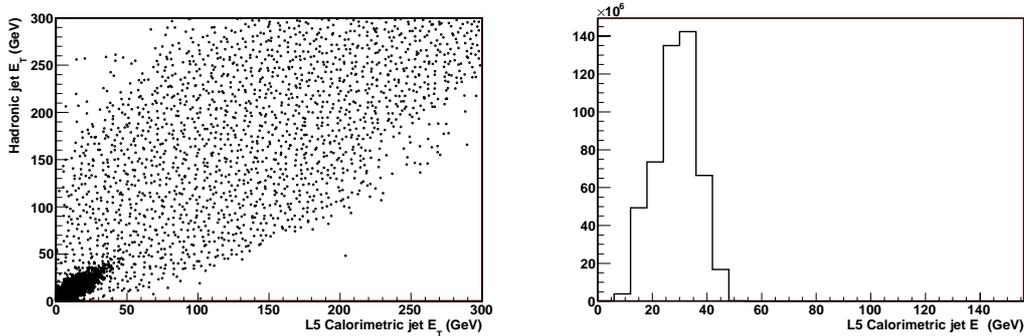


Figure 3.16: (left)  $E_{T,CAL}$  vs  $E_{T,HAD}$  correlation. (right)  $E_{T,HAD}$  distribution for jets in the range  $35 < E_{T,CAL} < 40$  GeV.

distribution for a low energy bin ( $35 < E_T < 40$  GeV). The distribution has non negligible tails and a Gaussian fit doesn't describe it very well. Therefore the statistical mean of the distribution is used rather than the Gaussian mean to estimate the central value and a fit is performed as shown in figure 3.17.

The fit range is chosen according to the range the measurement is performed on data ( $30 < E_T < 250$  GeV).

A fourth order polynomial fit of the form:

$$a + b \cdot x + c \cdot x^2 + d \cdot x^3 + e \cdot x^4$$

is chosen as fitting function. The values are reported in the picture. The correction is rather large (with a offset of about -2 GeV).

The quality of the fits is represented by the fit residuals shown in fig 3.17.

Residuals are defined, here, as

$$r_i = \frac{E_{T,HAD} - E_{T,CAL}^{corr}}{E_{T,HAD}}$$

and are plotted as a function of corrected jet  $E_{T,CAL}^{corr}$ . Their low values (below 1%) show the fit is trustable even though the values raise at the edges (as expected since

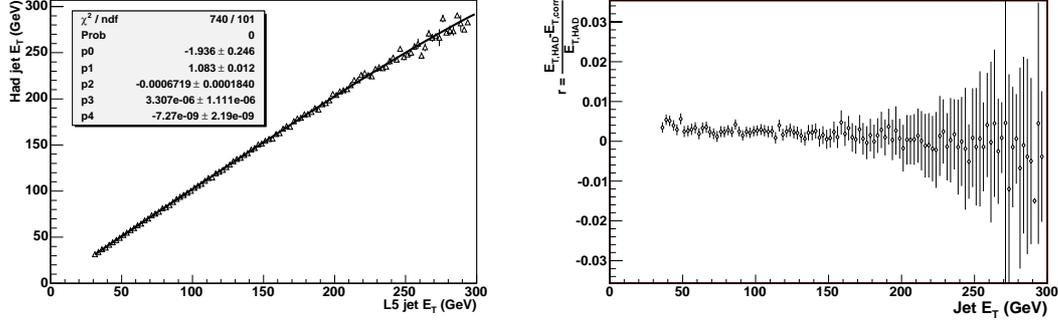


Figure 3.17: (left)  $E_{T,CAL}$  vs  $E_{T,HAD}$  for SVT tagged jets and fit residuals (right) from Pythia.

points in the tails are not used for the fit). A higher threshold on the corrected  $E_T$  is applied to take into account the additional energy correction and insure a meaningful comparison to Monte Carlo hadronic jets. The final threshold on the two SVT-tagged jets  $E_T$  is:

- Leading jet:  $E_T > 35$  GeV,  $|\eta| < 1.2$ ;
- Second jet:  $E_T > 32$  GeV,  $|\eta| < 1.2$ ;

### 3.6.2 Systematic uncertainties

The total systematic uncertainty on the jet energy corrections includes contributions from the generic corrections and the tagged jets specific ones. The generic contribution has been evaluated for common use in CDF by the jet energy scale and resolution group. The total systematic is the quadratic sum of the uncertainty of the single corrections applied. Figure 3.18 shows the uncertainty relative to the corrections for jets in the central calorimeter region  $0.2 < |\eta| < 0.6$ , corresponding to the different correction steps. For jets out of this region the relative correction uncertainty (also shown in figure 3.18) has to be summed. The figure gives an idea of the uncertainty value for a single jet.

To estimate the specific correction contribution to the uncertainty the values of the polynomial coefficients are changed (in turn raised and reduced) by 10%.

The total uncertainty on the differential cross sections functions of the leading jet  $E_T$ , the invariant mass and the di-jet  $\Delta\phi$  are estimated changing each jet  $E_T$  by a quantity that is equal to  $\pm 1\Delta\sigma_{generic} \pm \Delta E(10\%)_{specific}$ . The difference in the cross sections with respect to the central values are taken bin by bin as systematic uncertainty.

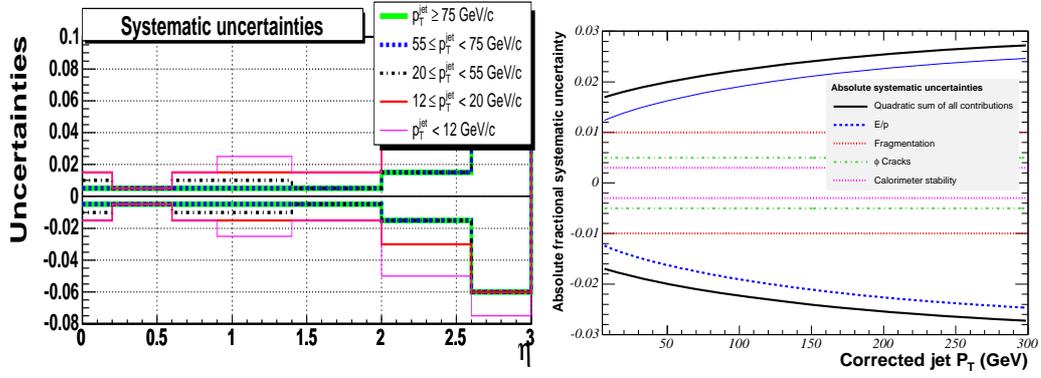


Figure 3.18: Systematic uncertainty on the relative correction (left) as a function of the jet  $\eta$  and total uncertainty as a function of the jet  $E_T$  (right). The components due to different correction are separated.

Fig. 3.19, fig. 3.20 and fig. 3.21 show the total uncertainties as a function of leading jet  $E_T$ , invariant mass and the azimuthal distance  $\Delta\phi$  respectively, the values are summarized in tables 3.7 and 3.8.

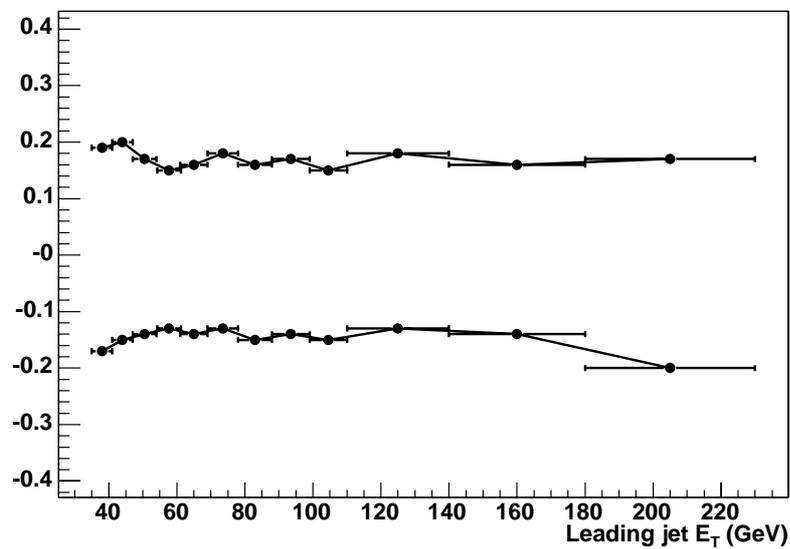


Figure 3.19: Systematics from jet energy corrections as a function of the leading jet  $E_T$ .

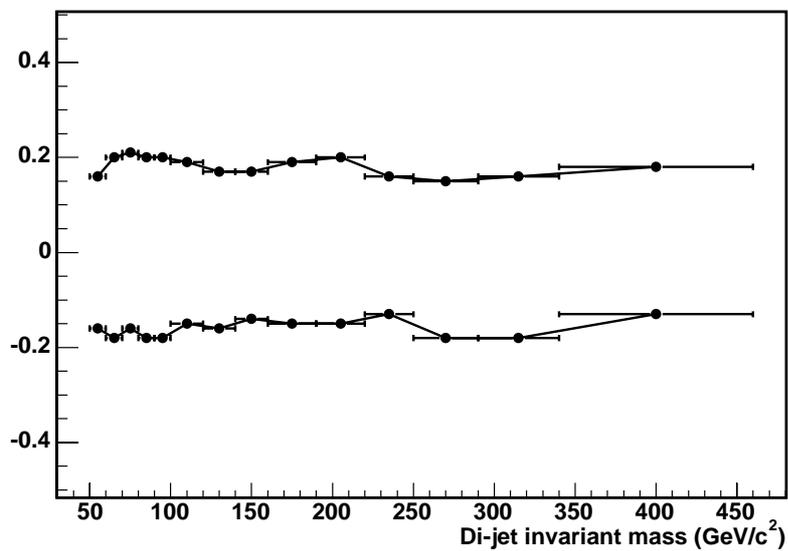
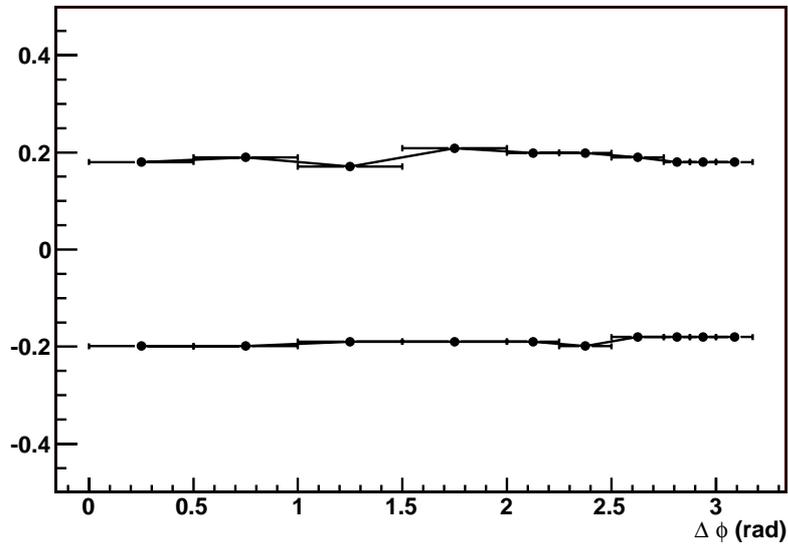


Figure 3.20: Systematics from jet energy corrections as a function of the di-jet invariant mass.

Figure 3.21: Systematics from jet energy corrections as a function of  $\Delta\phi$ 

$E_T$ bin (GeV)	+1 $\sigma$ (%)	-1 $\sigma$ (%)
(35,41)	19	17
(41,47)	20	15
(47,54)	17	14
(54,61)	15	13
(61,69)	13	11
(69,78)	18	13
(78,88)	11	10
(88,99)	18	18
(99,110)	10	15
(110,140)	18	7
(140,180)	14	14
(180,230)	17	24

Table 3.6: Jet energy correction relative systematic uncertainties for each  $E_T$  bin.

Di-jet invariant mass bin ( $\text{GeV}/c^2$ )	$+1\sigma$ (%)	$-1\sigma$ (%)
(50,60)	16	16
(60,70)	20	18
(70,80)	21	15
(80,90)	20	18
(90,100)	20	18
(100,120)	20	13
(120,140)	15	16
(140,160)	15	13
(160,190)	20	15
(190,220)	21	15
(220,250)	15	13
(250,290)	13	18
(290,340)	15	18
(340,460)	18	13

Table 3.7: Jet energy correction relative systematic uncertainties for each invariant mass bin.

$\Delta\phi$ bin (rad)	$+1\sigma$ (%)	$-1\sigma$ (%)
(0.,0.5)	17	18
(0.5,1.0)	18	17
(1.0,1.5)	16	17
(1.5,2.0)	20	18
(2.0,2.25)	19	18
(2.5,2.75)	19	17
(2.75,2.875)	17	17
(2.875,3.0)	16	15
(3.0,3.175)	15	15

Table 3.8: Jet energy correction relative systematic uncertainties for each  $\Delta\phi$  bin.

# Chapter 4

## Tagging $b$ -jets

Identifying  $b$ -flavor jets (*i.e.* tagging  $b$  jets) is fundamental for measuring  $b$ -jet cross sections. Different techniques exist for this purpose. A short overview will be given in section 4.1.

The tagging algorithm used in this analysis exploits the long life-time of the  $B$ -hadrons: the decay products, in fact, originate from the location of the decay, a secondary vertex, and not from the primary vertex like all other particles in the event. The tagging algorithm is based on the reconstruction of secondary vertices inside jets and it is described in section 4.2. Section 4.3 describes the tagging efficiency and the way it is modified by the presence of SVT tracks inside the jet.

The main limitation of the secondary vertex tagging method is the fact that fake secondary vertices can be reconstructed. Moreover, hadrons that contain a  $c$ -quark can have decay lengths, in the detector, similar to those of  $B$ -hadrons: it is therefore necessary to find out which is the real  $b$ -jet content of a tagged jet sample (purity). A procedure to calculate the  $b\bar{b}$  purity of a double tagged jets sample is summarized in section 4.4. Section 4.5 finally describes the systematic uncertainties which are related to the tagging procedure.

### 4.1 Different techniques to identify $b$ -jets

Methods to identify  $b$ -jets can be roughly divided into two main groups: those relying on semi-leptonic decays of the  $B$ -hadron inside the jet and those measuring the decay length of the  $B$ -hadron.

The so-called ‘soft lepton tagger’ identifies low momentum electrons or muons, inside jets. It disentangles leptons originating from semi-leptonic decays of the  $B$ -hadrons from prompt leptons. The limitations of this method are mainly due to the low semi-leptonic branching ratio which is of the order of 10% and to possible wrong identification of leptons (lepton fakes). This tagger is of limited use for this analysis

because of the low branching ratio.

A different technique takes advantage of the long life-time of the *B*-hadrons, which is of the order of 1.5 ps and corresponds to a proper decay length,  $c\tau$ , of 450  $\mu\text{m}$ . The large relativistic boost of the events makes the *B*-hadrons travel a few millimeters before decaying so that daughter particles derive from a secondary vertex well displaced from the primary interaction point. Due to the complexity of the final states, not all the decay products trajectories will be reconstructed. However it is possible to use more inclusive techniques that exploit the silicon vertex tracker to distinguish between tracks deriving from the primary or secondary vertex. For example, the SecVtx algorithm, used for this measurement, applies a selection on track impact parameter and reconstructs a secondary vertex from the selected tracks.

One drawback of this procedure is that charm hadrons are also characterized by intermediate ( $c\tau \approx 200 \mu\text{m}$ ) proper path length, so it is difficult to distinguish *b* from *c* jets. For this reason a further selection is needed to extract the real bottom flavour content in the analyzed sample.

## 4.2 The Secondary Vertex (SECVTX) algorithm

The SecVtx algorithm [76] is based on the reconstruction of secondary vertices inside a jet as illustrated in figure 4.1. It measures the 2-dimensional projection along the jet axis of the distance between the primary and secondary vertex and it applies a cut on its significance <sup>1</sup>.

The secondary vertex is reconstructed using tracks which lie within a *sub-cone* of radius 0.4 around the jet axis:

$$\Delta R = \sqrt{(\eta_{track} - \eta_{jet})^2 + (\phi_{track} - \phi_{jet})^2} < 0.4$$

The *sub-cone* size is kept at 0.4 despite the fact that jets can be reconstructed with cones of size 0.4, 0.7 or 1.0. This choice is justified by the fact that, most of the times, the heavy flavour hadron trajectory is close to the jet axis and the decay products follow the direction of the initial hadron. Making use of low energy tracks further away from the jet axis can also increase the reconstruction of fake secondary vertices.

In order to reduce the number of tracks which are not well measured, a preliminary track selection is applied [77]:

- $r - \phi$  hits in at least 3 SVX layers
- $\chi^2$  per degree of freedom less than 8.0

---

<sup>1</sup>Here the significance of a quantity  $x$  is defined as  $\sigma x/x$ .

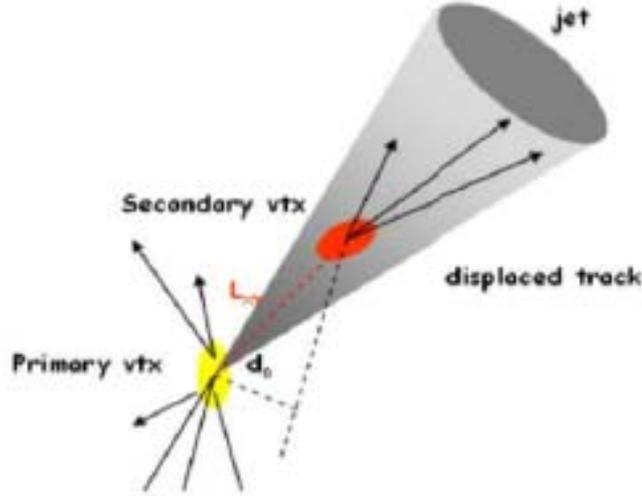


Figure 4.1: Schematic representation of a secondary vertex reconstruction inside a jet and the definition of  $L_{xy}$  and  $d_0$  with respect to the primary vertex.

- $p_T > 0.5$  GeV
- $d_0$  with respect to the primary vertex  $< 0.15$  cm
- $\delta(Z_0)$  with respect to the primary vertex  $< 2$  cm
- $K_s$  and  $\lambda$  removal
- Conversion track removal

$d_0$  is the distance of closest approach of the track with respect to the primary vertex (impact parameter), as illustrated in figure 4.1.  $\delta(Z_0)$  is the distance along the  $z$ -axis between the  $z$ -projection of the point of closest approach and the  $z$ -position of the primary vertex. The  $\chi^2$  of a track is the sum of the squares of the deviations between the location of the silicon hits and the fitted track. The number of degrees of freedom is the number of parameters used when fitting the track to the hits in the silicon.

The cuts on the  $d_0$  and the  $\delta(Z_0)$  are applied to reduce the effects of interactions in the detector material. Tracks which pass these basic selection criteria are labeled as being *good* tracks.

Having selected *good* tracks, an attempt at reconstructing a secondary vertex is made using a two steps procedure:

- *Pass1*: at least three tracks with  $P_T > 500 \text{ MeV}/c$  and  $|d_0/\sigma_{d_0}| > 2.5$  are required: out of them, an attempt to reconstruct a secondary vertex is made, where at least one of the tracks used in the fit must have  $p_T > 1 \text{ GeV}/c$ . The tracks considered must fulfill the requirements listed in table 4.1.
- *Pass2*: If no vertex is reconstructed after Pass1, a second attempt is performed using tighter track requirements ( $p_T > 1 \text{ GeV}/c$  and  $|d_0/\sigma_{d_0}| > 3$ , plus one track  $p_T > 1.5 \text{ GeV}/c$ ), but with looser particle multiplicity (two displaced tracks are enough). Track requirements in terms of hits and  $\chi^2$  are again summarized in table 4.1.

For each tagged jet the multiplicity of *good* tracks inside the jet cone is shown in figure 4.2, together as the number of *Pass1* tracks, *Pass2* tracks, and the number of tracks used to reconstruct the secondary vertex.

SecVtx Track Selection			
N hit	3 hits	4hits	$\geq 5$ hits
Pass 1	gd=3; m=0; $\chi^2 < 12$ gd=3; $m \leq 2$ ; $\chi^2 < 6$ gd=2; $m = 0$ ; $\chi^2 < 6$ gd=3; m=0; $\chi^2 < 12$ $m \leq 1$ ; $\chi^2 < 12$	gd=4; m=0; $\chi^2 < 18$ gd=4; $m \leq 1$ ; $\chi^2 < 12$ gd=4; $m \leq 2$ ; $\chi^2 < 6$	gd=5; $\chi^2 < 18$ gd=4; $\chi^2 < 12$ $m \leq 1$ ; $\chi^2 < 12$
Pass 2	gd=3; $m \leq 1$ ; $\chi^2 < 6$ hits on SVX L0 or L1	gd=4; $m \leq 1$ ; $\chi^2 < 6$ gd=4; m=0; $\chi^2 < 6$	gd=5; $m \leq 1$ ; $\chi^2 < 12$ $\chi^2 < 6$ ; $m \leq 0$

Table 4.1: SVX hits requirement, where ‘gd’ is the number of good hits, and ‘m’ is the number of missing hits. In addition, associated COT tracks must have a total of at least 19 axial hits and 16 stereo hits.

Once a secondary vertex is found inside the jet, additional selection is applied on the distance between the primary vertex and the secondary vertex: for this purpose the quantity  $L_{xy}$  is calculated. It is defined as the distance between primary and secondary vertex in the  $r-\phi$  plane. Figure 4.1 shows the scheme of a secondary vertex within a jet cone and the distance from the primary vertex ( $L_{xy}$ ). The selections of the SecVtx algorithm are actually performed on  $L_{2D}$ , the projection of  $L_{xy}$  onto the jet axis<sup>2</sup>. The sign of  $L_{2D}$  is defined with respect to the jet direction by the angle  $\theta$  between the jet axis and the secondary vertex vector: if  $\theta < \pi/2$ ,  $L_{2D}$  is positive. The sign of  $L_{2D}$  is also used in the selection. In fact, displaced vertices from heavy flavor

<sup>2</sup>Sometimes the same quantity is referred to as  $L_{xy}$ , since the difference is quite small and the two can be treated as equivalent at the first approximation

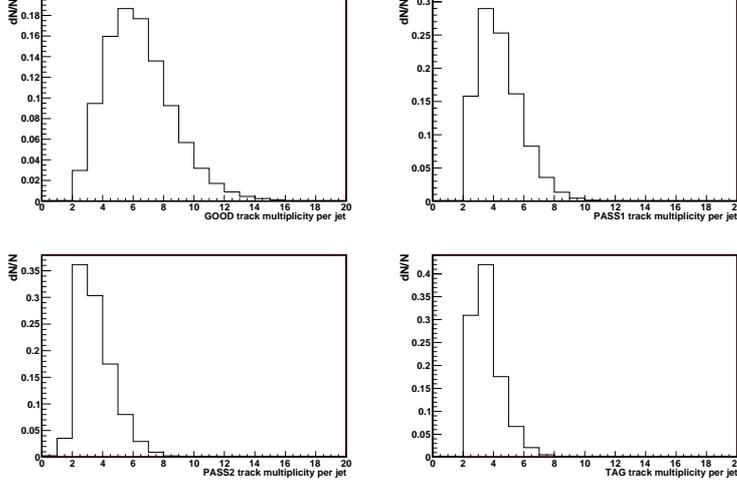


Figure 4.2: Multiplicity of good tracks inside jets (top-left), number of Pass 1 (top right) and Pass2 (bottom left) tracks; the bottom-right plot refers to tracks effectively used to reconstruct secondary vertex in jets positively tagged.

decays are more likely to have a large and positive  $L_{2D}$ , while mis-tagged vertices usually present a smaller displacement with respect to the primary interaction point and  $L_{2d}$  is supposed to be randomly positive or negative, thus symmetric around 0.

A secondary vertex is required to have  $L_{2D}/\sigma_{L_{2D}} > 7.5$  for positive tag jets, and  $L_{2D}/\sigma_{L_{2D}} < -7.5$  for negative tag jets, where  $\sigma_{L_{2D}}$  is the total estimated uncertainty on  $L_{2D}$ .

Figure 4.3 shows a simple sketch of positive and negative SecVtx tags. Additional requirements on  $L_{2D}$  are finally applied to make a further rejection of vertices found because of material interactions:

- $|L_{2D}| < 5 \text{ cm}$  and  $L_{2D} < 2 \text{ cm}$  for two tracks vertices;
- $|pseudo - c\tau| < 1 \text{ cm}$ , defined as  $|L_{2D}|xM_{sec.vtx}/p_T^{sec.vtx}$ , where  $M_{sec.vtx}$  and  $p_T^{sec.vtx}$  are respectively the mass and the momentum of the secondary vertex, reconstructed from the tracks used in the fit;
- $|z_{sec.vtx} - z_{prim.vtx}| < 5 \text{ cm}$ .

Positively tagged jets are the  $b$ -flavour jet candidates considered in this analysis. Negatively tagged jets are typically used as an estimate of mis-tag rate.

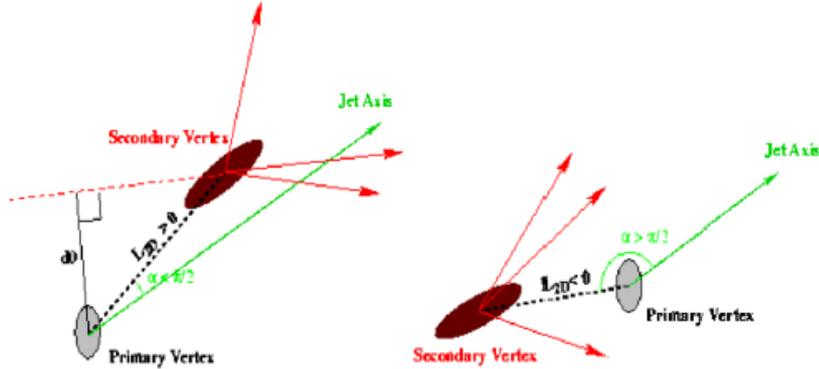


Figure 4.3: Schematic representation of a jet positively and negatively tagged by SecVtx.  $L_{2D}$  is the same as the  $L_{xy}$  in the text.

### 4.3 The $b$ -tagging efficiency

The  $b$ -tagging efficiency is generally defined as the fraction of positively tagged  $b$ -jets with respect to all  $b$ -jets in the event. In order to do a direct measurement of the  $b$ -tagging efficiency, the heavy flavor content of the event should be determined before and after the tagging procedure.

The approach CDF has adopted up to now [78] is to use a calibration data sample of jets, whose heavy flavour fraction can be measured independently of the tagger, and to derive a per-jet tagging efficiency. However, this estimate cannot be directly extrapolated to other samples: in general jets from different samples can have different features. Therefore, the  $b$ -tagging efficiency is also derived from Monte Carlo events, specifically generated to match the data of the calibration sample. The ratio between the efficiencies in data and Monte Carlo is measured.

This ratio is called *Scale Factor* (SF) and it is independent of the dataset. It is used to correct, for any data sample, the  $b$ -tagging efficiency measured in the Monte Carlo, in order to account for any difference between real and simulated events.

In other words the simulation is used to describe the energy dependence and geometrical acceptance of the tagger, while the scale factor takes into account imperfections in the Monte Carlo such as the difference in tracking efficiency and resolution, uncertainties in B-hadron decay models, etc.

Usually inclusive leptons samples (electrons or muons) are chosen as calibration samples. Their enhanced heavy flavour content - thanks to the sensitivity of the trigger to semi-leptonic decays of bottom and charm hadrons - allows a measurement of the  $b$ -jet fraction and the  $b$ -tagging efficiency in the data, using techniques that

are independent of displaced vertices and that are based on the good electrons and muons identification capability of the detector.

Figure 4.4 shows the  $b$ -tagging efficiency measured using a muon data sample, as a function of jet  $E_T$  and run number [79]: the  $b$ -tagging efficiency is of the order of 40% and the scale factor around 90%.

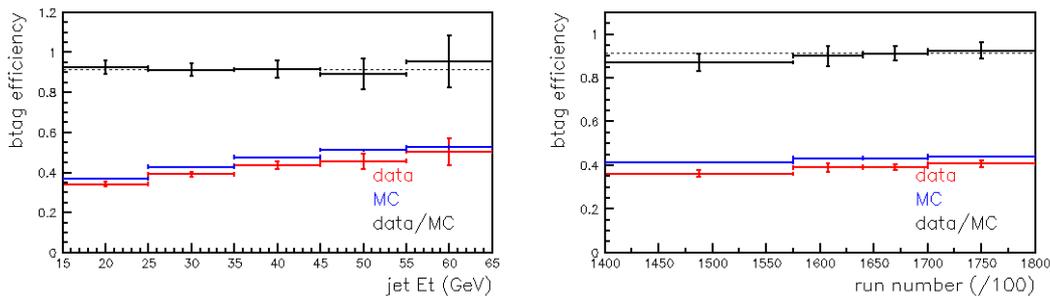


Figure 4.4:  $b$ -tagging efficiency in data and Monte Carlo and scale factor as a function of jet  $E_T$  (left) and run number (right) [78].

This analysis makes use of the approach described above to calculate in the Monte Carlo the efficiency for tagging a jet which includes at least one SVT track in its cone. A specific scale factor, that takes into account the presence of the SVT track inside the jet, is measured and it is used to correct the efficiency to data.

Next section describes how the presence of a SVT track inside the jet increases the tagging efficiency on the jet itself, but biases the performances of the  $b$ -tagging algorithm in a way that requires a careful study.

### 4.3.1 The effect of SVT selection

It is reasonable to assume that a jet containing a SVT track within its cone is more likely to be tagged. These tracks, in fact, must satisfy selection criteria, specifically designed to identify heavy flavour events: for example a large impact parameter.

Figure 4.5 shows the  $b$ -tagging efficiency for these jets, estimated from a PYTHIA  $b\bar{b}$  sample. It is larger than 60% over the all  $E_T$  range. In comparison the  $b$ -tagging efficiency for a generic  $b$ -jet is approximately 40%, as shown in figure 4.4.

In fact, SVT-tagged jets have a number of good tracks, pass1 and pass2 tracks that is larger than generic jets. This quantities are represented in figure 4.6 and can be compared to those in figure 4.2 corresponding to generic jets.

As explained in section 3.5 this analysis applies a tight offline selection so to insure that the final events are free of any residual trigger effect.

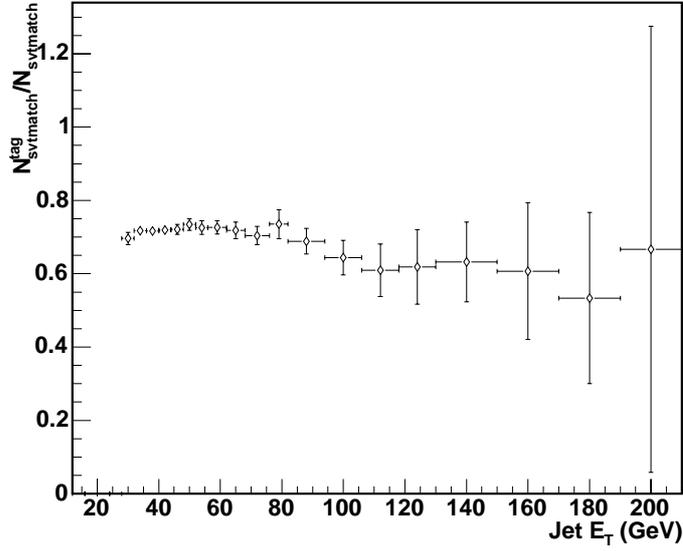


Figure 4.5:  $b$ -tagging efficiency for jets including a SVT track inside their cone, calculated using PYTHIA  $b\bar{b}$  events.

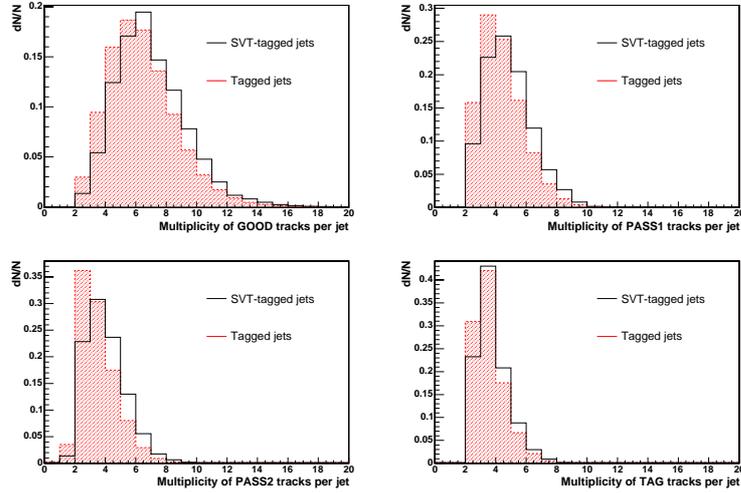


Figure 4.6: Multiplicity of good tracks inside jets (top-left), number of Pass 1 (top right) and Pass2 (bottom left) tracks the bottom-right plot refers to tracks effectively used to reconstruct secondary vertex in jets positively tagged. SVT-tagged jets (empty histograms) and generic tagged jets (shaded histograms) are compared.

Details on these requirements have been given in section 3.5. The most important issue is the request of two SVT-tagged jets in the event: two tagged jets each associated to a track with  $|d_0| < 120 \mu m$  and  $p_T > 2 GeV/c$ , reconstructed by the COT, the silicon detector and triggered by SVT. As already explained the matching is made, requiring that the  $\Delta R$  distance between the jet axis and the track is lower than 0.4. As shown in section 3.5 only about 30% of the tagged jets satisfy this criteria.

As explained below, the overall event selection, which requires two such jets is therefore very low: only a few percents.

There is an additional effect that has to be taken into account. A certain correlation between tagged jet is usually expected to be positive (once a jet is tagged the second one has a higher chance to be tagged as well) but generally negligible: figure 4.7 shows on the left the tagging efficiency for tagging a jet in a pair in which the second jet is not tagged (full squares) or it is tagged (empty triangles). The two curves overlap. It has been proved, instead, that requiring a three dimensional matching between a tagged jet and a good SVT track ( $p_T > 2 GeV/c$  and  $|d_0| > 120 \mu m$ ) results in a negative correlation as shown in figure 4.7.

Events from a  $b\bar{b}$  di-jet sample have been selected requiring at least two ‘good’ SVT tracks and two  $b$ -jets (that are ordered in  $E_T$ ).

The figure shows that the SVT-tagging efficiency for the second jet is much lower when the leading jet has been tagged (empty triangles in the plot) compared to the case no tagging requirement is made on the leading jet (full triangles in the plot). The efficiency for tagging the first jet is approximately 15-17%, a value that can be also estimated taking into account that the tagging efficiency for a  $b$ -jet is about 40% and that the efficiency for finding a SVT track inside a tagged jet cone is  $\approx 30\%$ . For the second jet, instead, the efficiency is always below 13%.

The correlation is also independent on the criteria chosen to order the jets.

The same result has been observed when the two jets are ordered according to other variables (jet  $\phi$  and  $|\eta|$  for example). This behavior seems related instead to the way SVT tracks are distributed in space and to the three dimensional matching between the track and the jet.

To prove it, all the ‘good’ (i.e. satisfying the  $p_T$  and  $d_0$  requirements) SVT tracks are identified in the event, a couple tagged jet-SVT track is found requiring the track to be inside the jet cone ( $\Delta R_{jet-track} < 0.4$ ) and the  $\Delta R$  between this SVT-tagged jet and all the other SVT tracks in the event is calculated.

The result is shown in figure 4.8. The  $\Delta R$  distribution proves that most of the ‘other’ SVT tracks in the event are close to the selected one and so they are more likely to fall inside the same jet ( $\Delta R < 0.4$ , the jet cone) or very close to it. The probability that a second jet is associated to a second ‘good’ SVT track is, therefore, reduced.

To take this effect into account, a per-event efficiency is measured using  $b\bar{b}$  PYTHIA

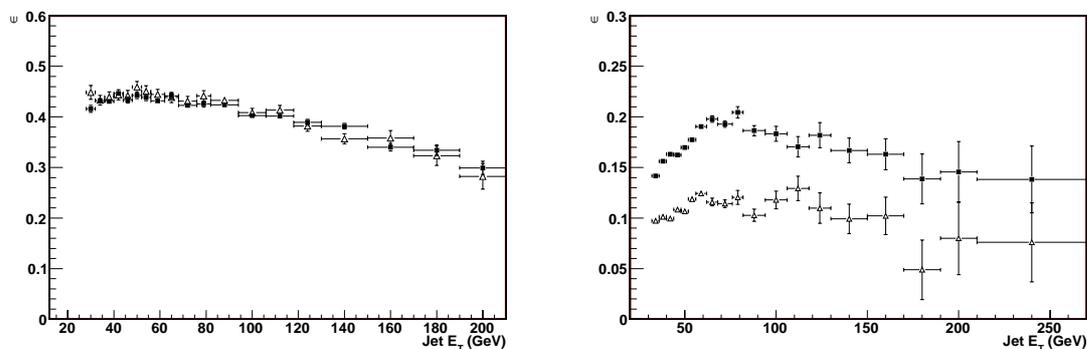


Figure 4.7: Left: Tagging efficiency for the second leading jet associated to "any leading jet" (full squares) or associated to tagged leading jet (empty triangles). Right: SVT-tagging efficiency for the second leading jet associated to "any leading jet" (full squares) or associated to SVT-tagged leading jet (empty triangles).

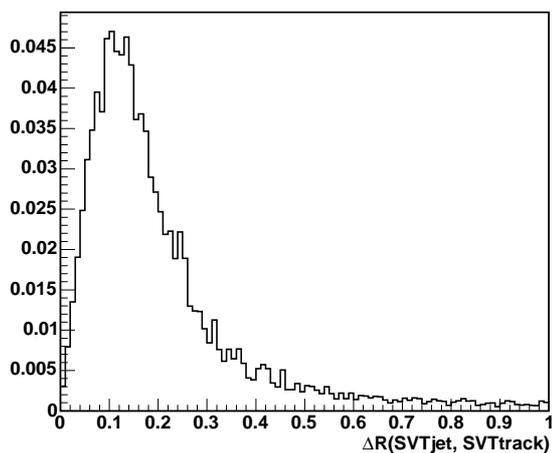


Figure 4.8: **left:**  $\Delta R$  between the SVT-tagged jet and other "good" SVT tracks in the same event.

di-jet events. It is calculated as

$$\epsilon = \frac{N_{2SVT}}{N_{2b}}$$

where  $N_{2SVT}$  is the number of  $b\bar{b}$  events with two SVT tagged  $b$ -jets in the region  $|\eta| < 1.2$  and  $N_{2b}$  is the number of  $b\bar{b}$  events with two  $b$ -jets in the region  $|\eta| < 1.2$ . The result is summarized in figure 4.9 as a function of the leading jet  $E_T$  and the di-jet invariant mass  $M_{jj}$ . The  $\Delta\phi$  between the two jets is shown in figure 4.10.

### 4.3.2 The SVT-tagging scale factor

The efficiencies shown in figures 4.9 and 4.10 are already corrected for the scale factor:

$$SF = 1.029 \pm 0.009(stat) \pm 0.034(syst)$$

It is the value CDF has estimated in [74] for tagged jets including a SVT track inside their cone. The measurement makes use of slightly different cuts with respect to the ones used in this analysis, even though the selection strategy is the same: for example, jets are reconstructed using a cone of 0.7 instead of 0.4.

In order to prove that the official value can be applied to the measurement presented in this thesis, a specific study, detailed in the appendix B has been made.

## 4.4 The $b\bar{b}$ content of the tagged jet sample

As explained already, a tagged jet sample is not a pure  $b$ -jet sample: it includes not only  $b$ -jets but also mis-tagged  $c$ -jets and light or gluon jets. The  $b\bar{b}$  purity of a two SVT-tagged jet sample can be defined as:

$$f_{b\bar{b}} = \frac{N_{++}^{bb}}{N_{++}}$$

where:

- $N_{++}^{bb}$  is the number of two SVT-tagged  $b$ -jets events;
- $N_{++}$  is the number of two SVT-tagged jets events.

Several techniques have been developed to determine the heavy flavor content of a tagged jet in order to extract the fraction of  $b$ -jets. Usually one or more variables, whose shape is different for  $b$ -jets and mis-tagged jets, is chosen to discriminate among the different components of a tagged jet sample. For example, the lepton  $p_T$ , measured

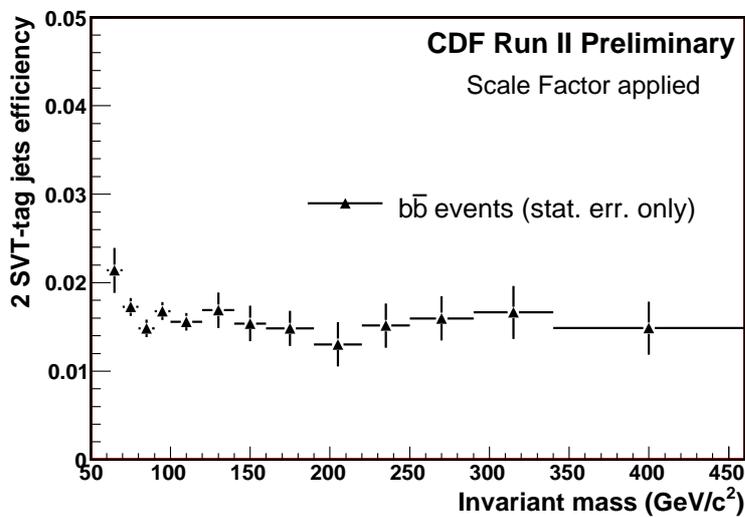
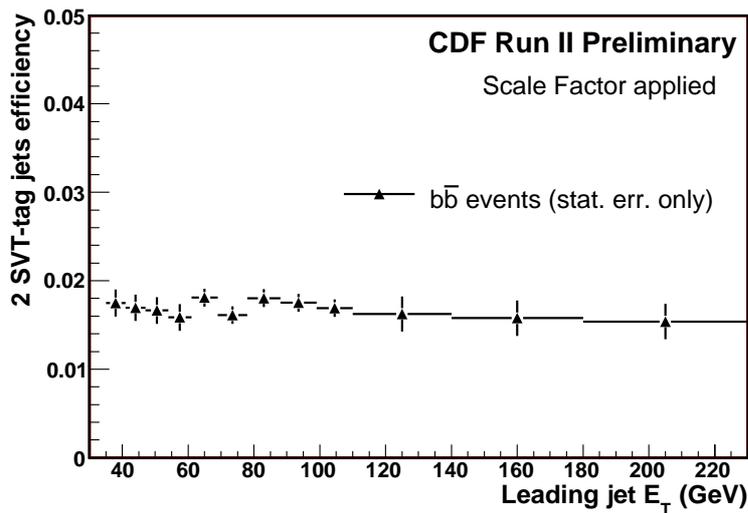


Figure 4.9: **top:** SVT-tagging efficiency as a function of the leading jet  $E_T$ . **bottom:** SVT tagging efficiency as a function of the invariant mass of the two jets .

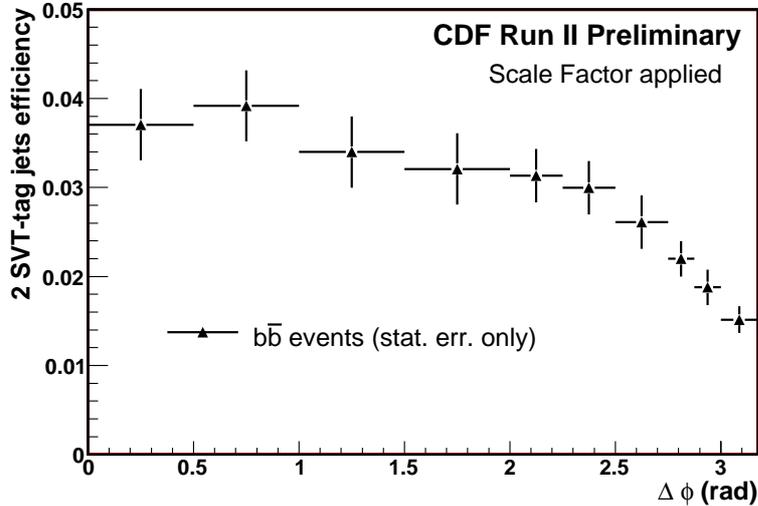


Figure 4.10: SVT tagging efficiency as function of  $\Delta\phi$  between the two jets

with respect to the jet axis, if the semileptonic  $B$  decay is reconstructed; the distance, projected on the transverse plane, between the primary vertex and the secondary vertex; or the secondary vertex invariant mass.

Here, the shape of the secondary vertex invariant mass is used. A secondary vertex, well separated from the primary interaction point, is usually identified for long-lived  $b$  or  $c$  hadrons. Due to the presence of neutral particles and the energy lost because of detector resolution, a full reconstruction of the hadron invariant mass is not possible. However, the invariant mass of tracks used to find the secondary vertex constitutes a good discrimination between jets containing  $b$ ,  $c$  or light quarks as the shapes are very different in each case, as shown in figure 4.11.

Using PYTHIA di-jet Monte Carlo samples it is possible to obtain  $b$ ,  $c$  and light jet mass distributions for positively tagged jets, that can be used as templates to fit the data distribution. Following this procedure the  $b$  purity can be extracted directly from the data.

In this case, the  $b$  content is the only fraction of interest, so the  $c$  and light templates are merged: the secondary vertex mass from the data is fitted to find the coefficient of a linear combination of the 2 components ( $b$  and non- $b$ ) using the ROOT routine TFractionFitter [80]. The method is stable within the whole jet  $E_T$  range.

The downside of this choice is the dependence on the  $c$  and light jet relative contribution to the non- $b$  template. The measurement is based on the assumption that the composition of non- $b$  templates in terms of the  $c$ -tagged and mis-tagged light

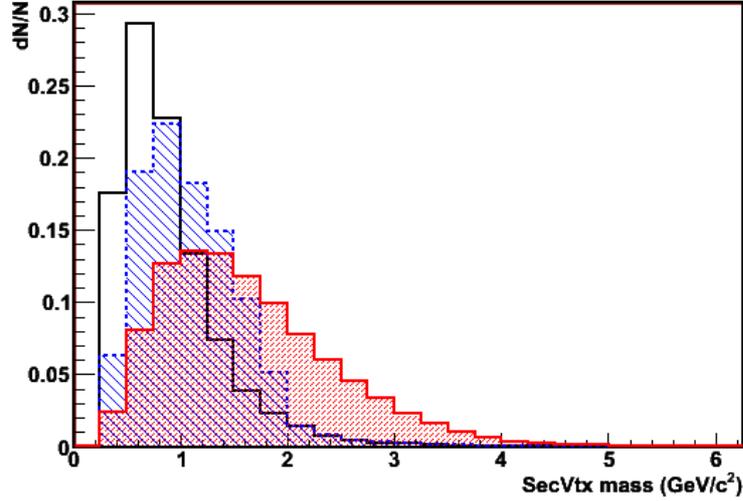


Figure 4.11: Secondary vertex mass templates for  $b$ -jets (full line, shaded area),  $c$ -jets (dashed line, shaded area) or light jets (full line).

jet fraction, given by the Monte Carlo, is correct. In fact it will be shown in the next chapter that a larger contribution of  $c$  jets (light jets) would correspond to a larger (smaller) average value of the non- $b$  mass distribution. A systematic uncertainty is associated to this assumption as shown in section 4.5.

Again, the presence of the SVT track in the jet modifies the structure of the secondary vertex, and it changes the distribution of the associated invariant mass as shown in figure 4.12. The SVT-tagged  $b$ -jets have a slightly larger invariant mass with respect to simple tagged  $b$ -jets. Therefore  $b$ -jet templates are built using SVT-tagged jets, even though this choice reduced the available number of events.

For this measurement, events are selected requiring two SVT-tagged jets and the  $b\bar{b}$  purity has to be estimated. The templates are chosen, accordingly, to be the sum of the two SVT-tagged jet secondary vertex masses. A two components fit is performed, using a ‘signal’ template corresponding to the  $b\bar{b}$  case and a ‘background’ template that is built merging all the possible contributions coming from non- $b$  jets:  $c\bar{c}$ ,  $b\bar{c}$ ,  $b\bar{l}$ , and so on.

The main problem associated to this procedure is the lack of statistics, especially for what concerns non- $b$  jets templates: the number of events having two mis-tagged jets is not sufficient to build proper templates. Therefore the following method is used: samples of tagged jets for different flavours are collected and, from each one of them, jets are picked in couples to build the sum of the secondary vertex masses. The

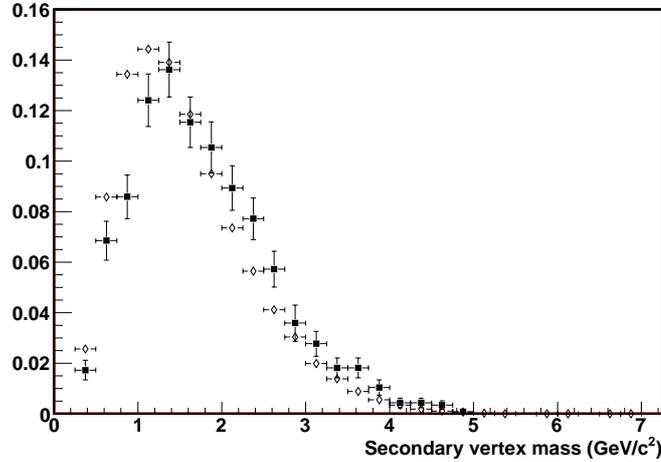


Figure 4.12: Tagged  $b$  jets secondary vertex mass (empty diamonds) and SVT-tagged (black squares)  $b$ -jets secondary vertex mass

relative normalization is chosen according to PYTHIA prediction and a systematic uncertainty will be assigned to it as described in section 4.5. Figure 4.13 shows a comparison between templates built following this procedure and ‘real’ tagged  $b\bar{b}$  events as a function of the two jets mean  $E_T$ .

A Kolmogorov test is also run to compare the shape of the distribution and the resulting probabilities and maximal differences are summarized in tables 4.2. The difference in shape is negligible, if compared to other sources of systematic uncertainty.

$E_T$ bin (GeV)	Test Probability	Max Distance
(32,40)	0.35	0.32
(40,50)	0.82	0.11
(50,60)	0.39	0.29
(60,80)	0.71	0.15
(80,100)	0.79	0.12
$E_T > 100$	0.68	0.19

Table 4.2: Results of the Kolmogorov test on  $b$ -quark mass templates as a function of jet  $E_T$ : probabilities and maximum differences

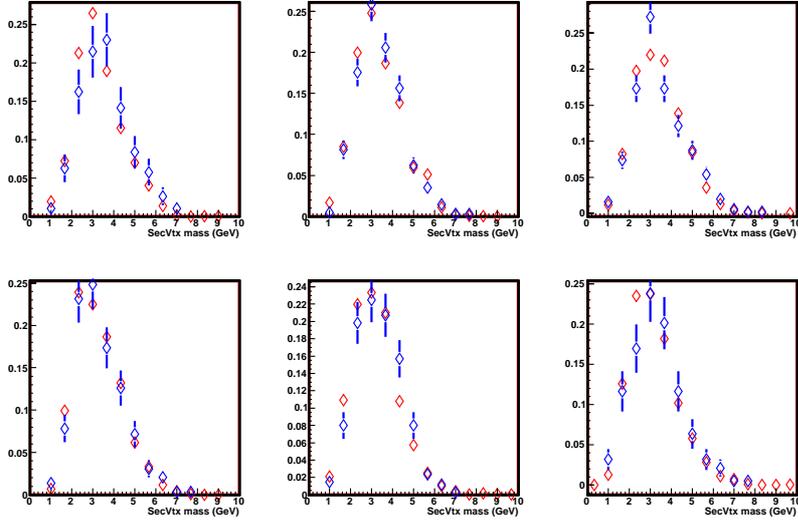


Figure 4.13: Mass templates for couples of SVT-tagged  $b$ -jets coming from independent (red) events or from the same event (blue), for jet  $E_T$  in the ranges (in GeV): (32,40), (40,50), (50,60), (60,80), (80,100),  $E_T > 100$  GeV.

The average di-jet  $E_T$  is divided in six intervals defined (in GeV) by the ranges: (32,40), (40,50), (50,60), (60,80), (80,100) and  $E_T > 100$  GeV, which are chosen to maximize the statistical power of the fit. Fits are performed independently in each interval and, for comparison, for all the events in the whole  $E_T$  range. To do so, different signal and background templates are built, corresponding to each  $E_T$  bin, they are shown in figure 4.14. The ‘inclusive’ signal and background templates are shown in figure 4.15.

The corresponding fit results are shown in figure 4.16 and figure 4.17. The  $b\bar{b}$  fraction values as a function of  $E_T$  are summarized in table 4.3 and figure 4.18. The value measured over the whole  $E_T$  range is  $0.87 \pm 0.04$ . This high purity is one of the advantages associated to the use of the SVT-tagging selection.

Those values are used in the calculation of the cross section to assign a weight to each event according to the mean value of the  $E_T$  of the two jets. Following this procedure the  $b\bar{b}$  purity of the two tagged jets can be displayed as a function of the leading jet  $E_T$ , the di-jet invariant mass and  $\Delta\phi$  as shown in figures 4.19 and 4.20.

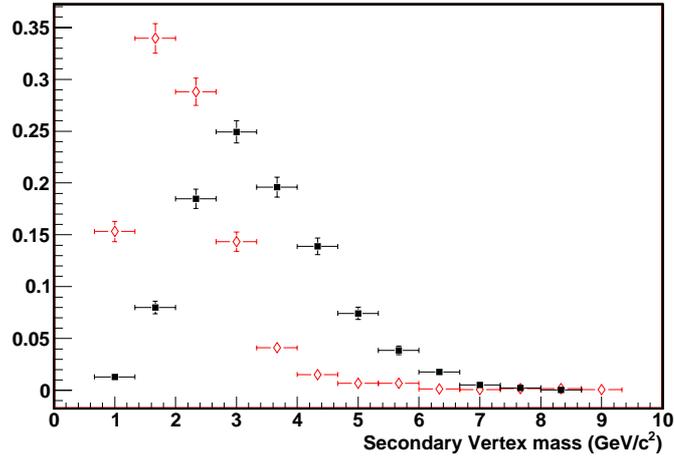


Figure 4.14:  $b\bar{b}$  secondary vertex mass templates (black squares) and background templates (empty diamonds) for  $E_T > 32$  GeV

$E_T$ bin (GeV)	b fraction
(32,40)	$0.87 \pm 0.03$ %
(40,50)	$0.88 \pm 0.03$ %
(50,60)	$0.86 \pm 0.04$ %
(60,80)	$0.78 \pm 0.05$ %
(80,100)	$0.84 \pm 0.06$ %
$E_T > 100$	$0.75 \pm 0.09$ %
$E_T > 32$	$0.87 \pm 0.04$ %

Table 4.3:  $b\bar{b}$  fraction for the different di-jet average  $E_t$  intervals. The last line shows the integrated result

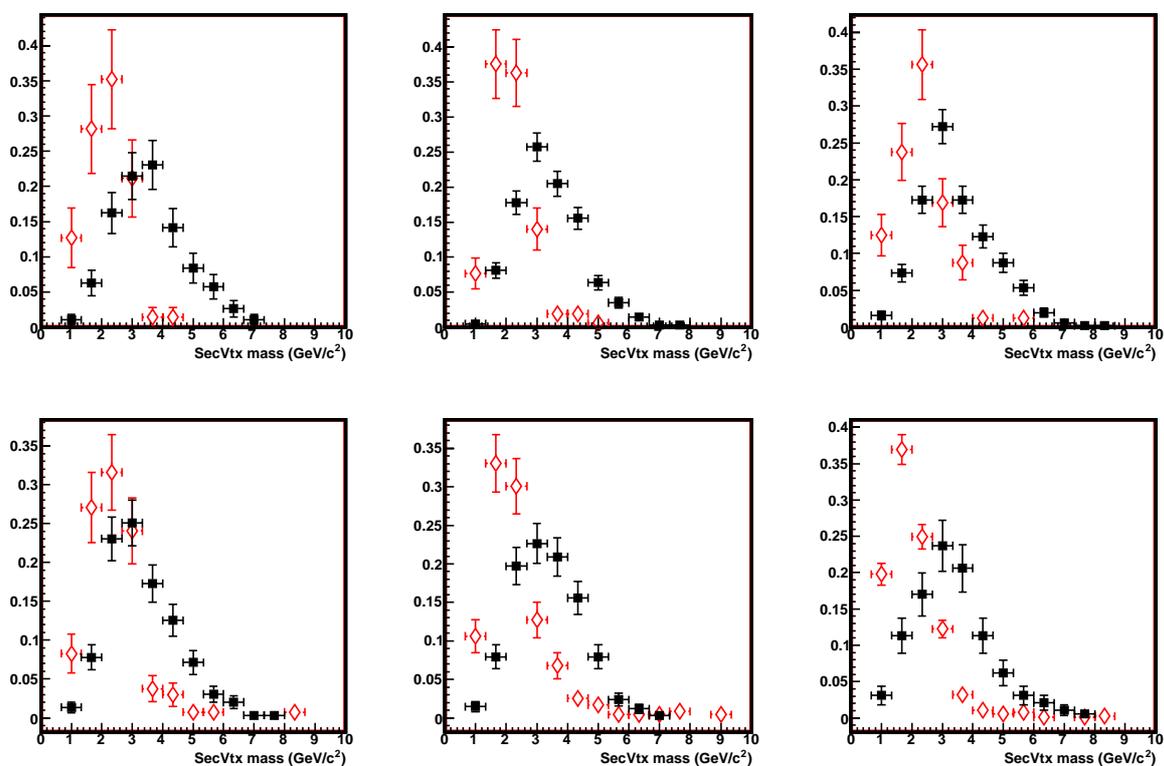


Figure 4.15:  $b\bar{b}$  secondary vertex mass templates (black squares) and background templates (empty diamond)

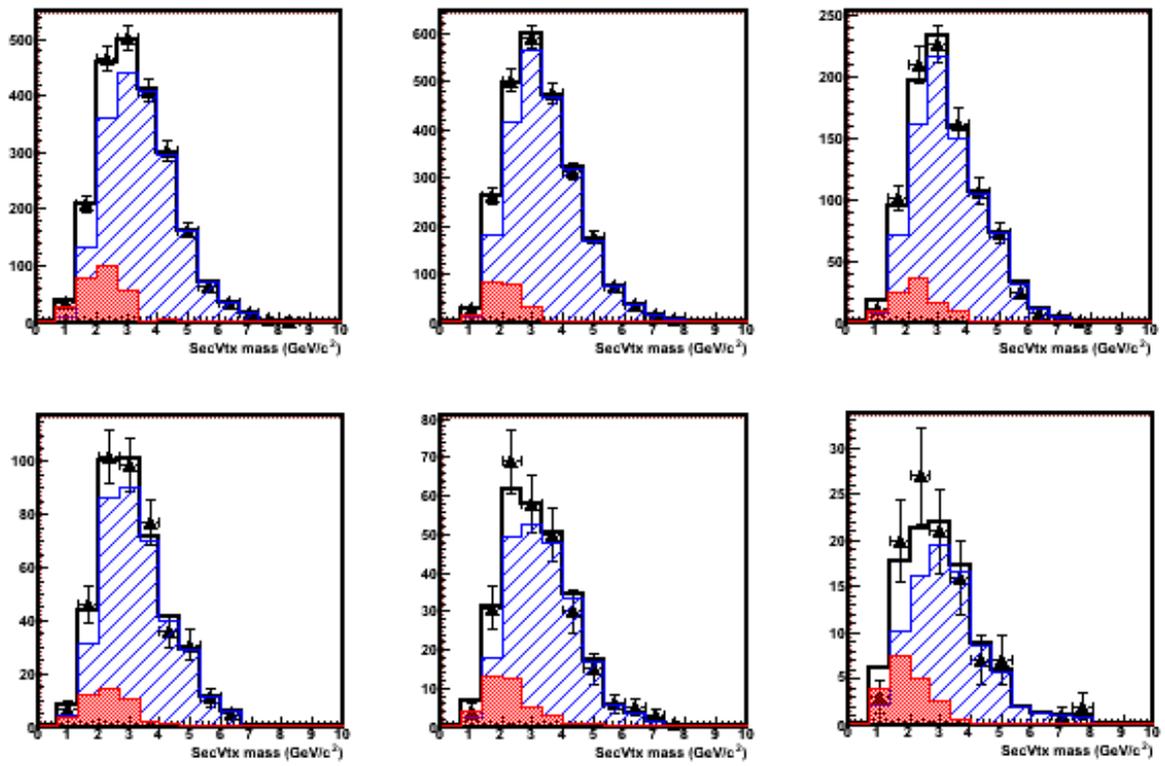


Figure 4.16: Result from the fit in each  $E_T$  bin: data is overlapped to fit prediction and Monte Carlo signal (blue) and background (red) templates

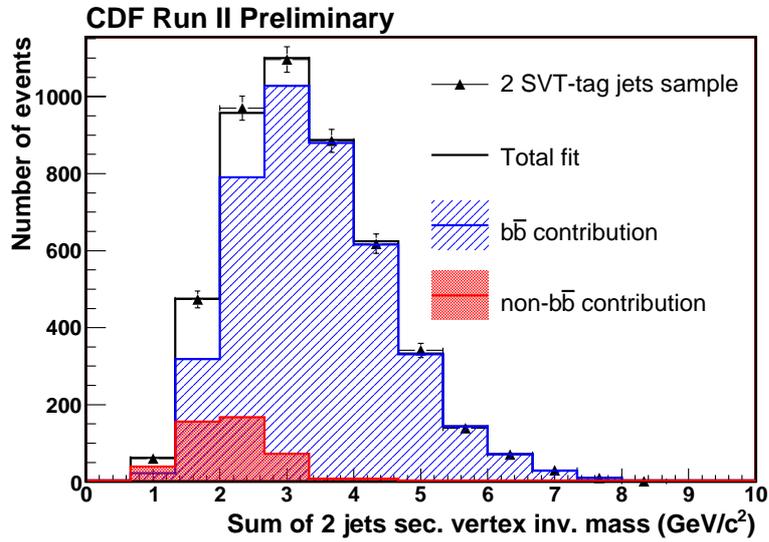


Figure 4.17: Result from the fit: data is overlapped to fit prediction and MonteCarlo templates

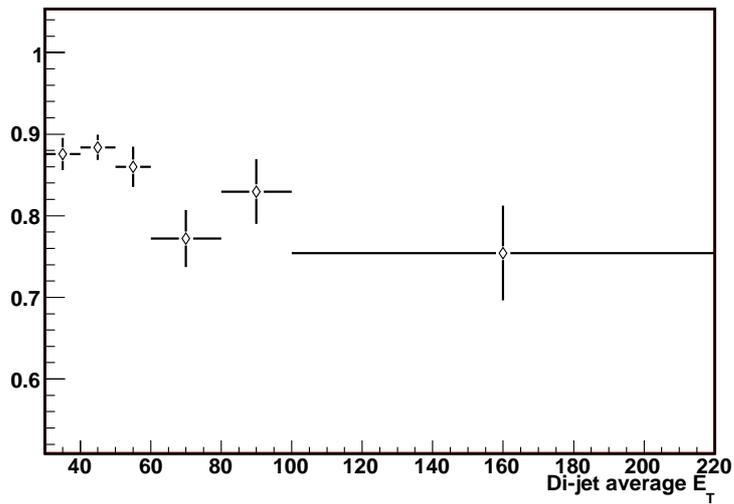


Figure 4.18:  $b\bar{b}$ -jet fraction as a function of the two jets average  $E_T$

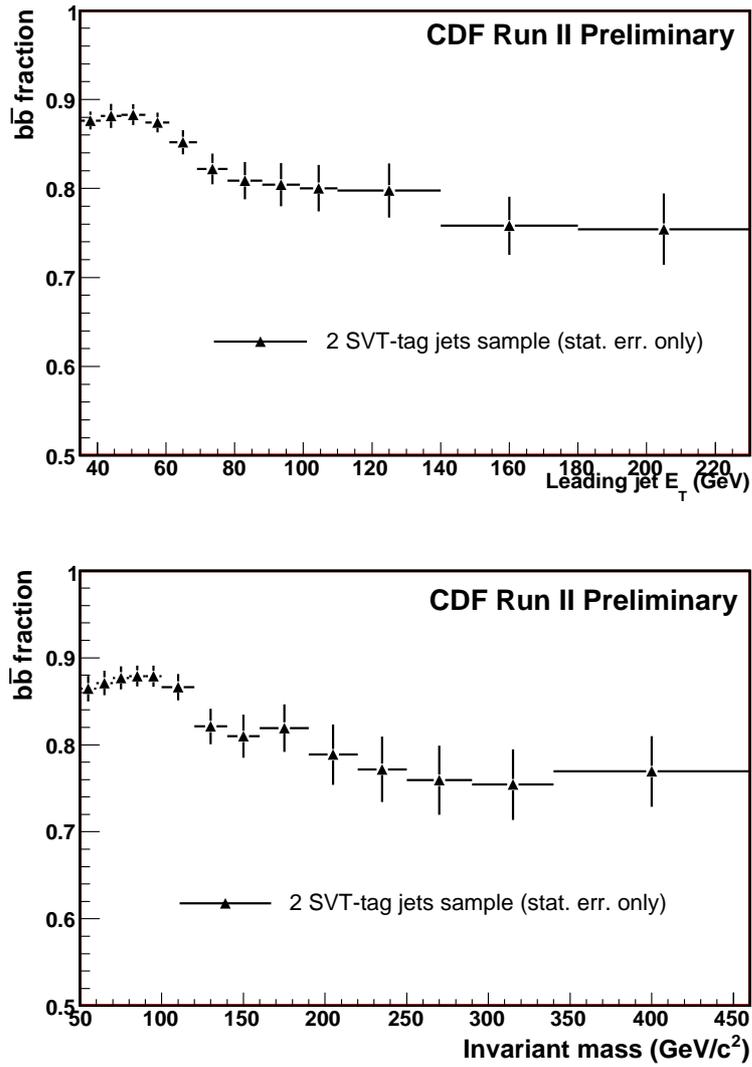


Figure 4.19:  $b\bar{b}$ -jet fraction as a function of the leading jet  $E_T$  (top) and of the di-jet invariant mass (bottom)

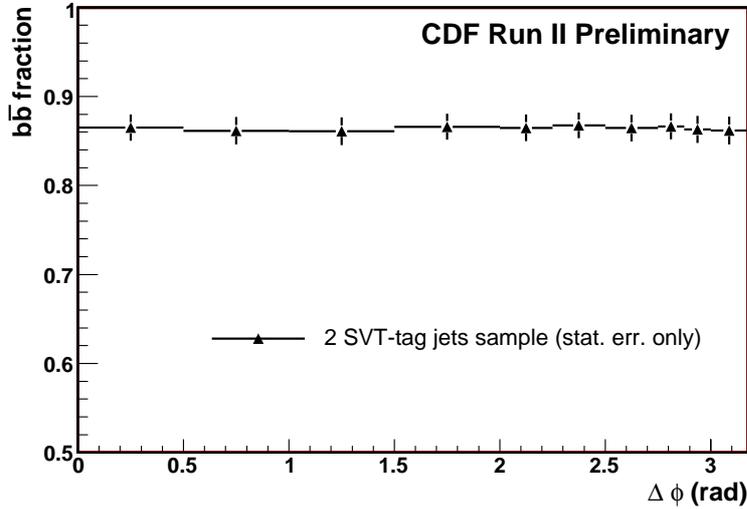


Figure 4.20:  $b\bar{b}$  -jet fraction as a function of the  $\Delta\phi$  between the two jets.

## 4.5 Systematic Uncertainties

We present a study of the systematic uncertainties related to the tagging procedure in this section. The topic concludes the discussion on  $b$ -tagging. A complete overview of all the different sources of systematic uncertainties will be described in the next section before a discussion of the results.

### 4.5.1 Heavy quark multiplicity in the jets

The quark multiplicity inside a jet modifies the tagging algorithm and the secondary vertex mass fit performances. A jet including two  $b$ -quarks, for example, is likely to have a larger number of good tracks and therefore a higher probability of being tagged with respect to a single  $b$ -quark jet. Moreover the case of a jet including two  $c$ -quark can fake the presence of a  $b$ -quark inside the jet.

This problem can affect the measurement both at the level of the SVT-tagging efficiency determination and at the level of the  $b\bar{b}$  purity of the sample. In fact, a higher number of tracks attached to the secondary vertex can modify the shape of the secondary vertex mass distribution. Figure 4.21 shows a comparison of the different categories of tracks considered by the SecVtx algorithm for jets including one  $b$  quark or two  $b\bar{b}$  quarks inside their cone. Figure 4.22 shows the same quantities for jets including one or two  $c$  quarks.

The pictures show that jets containing a couple of  $b\bar{b}$  quarks, have an higher

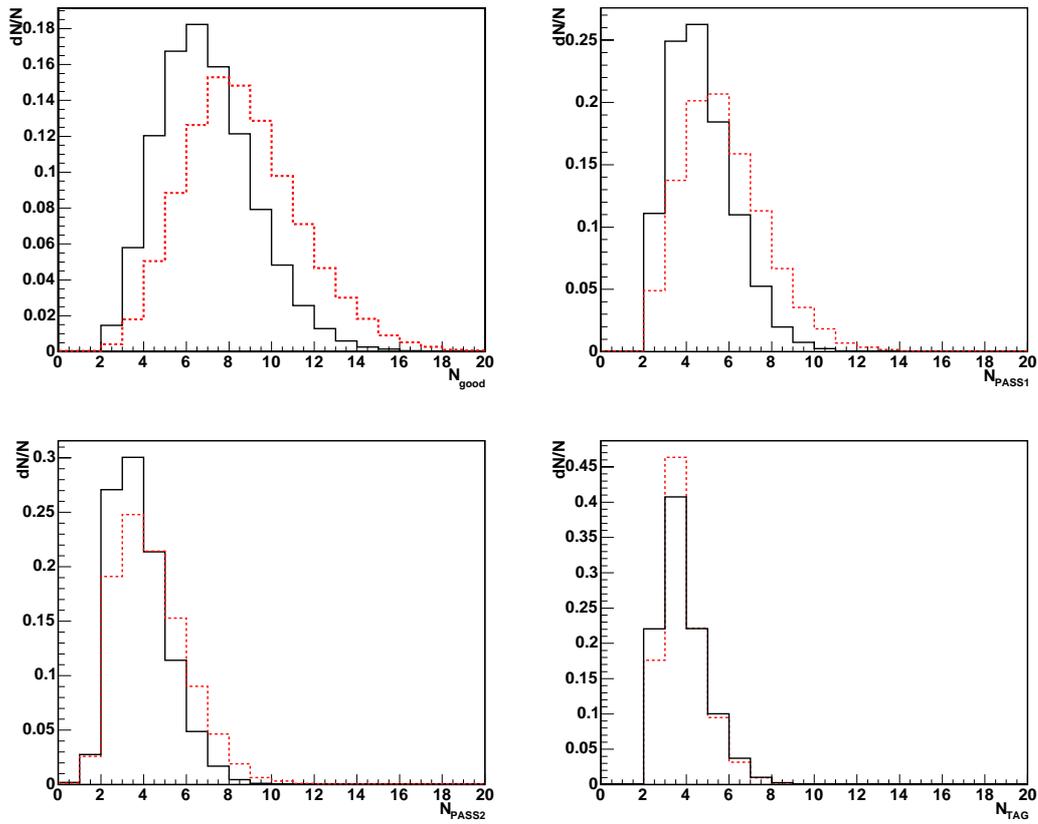


Figure 4.21: Track multiplicities for jets including one  $b$  quark (full line) or two  $b\bar{b}$  quarks, clockwise: number of good tracks, number of pass1 tracks, number of pass2 tracks, number of tag tracks (dashed line).

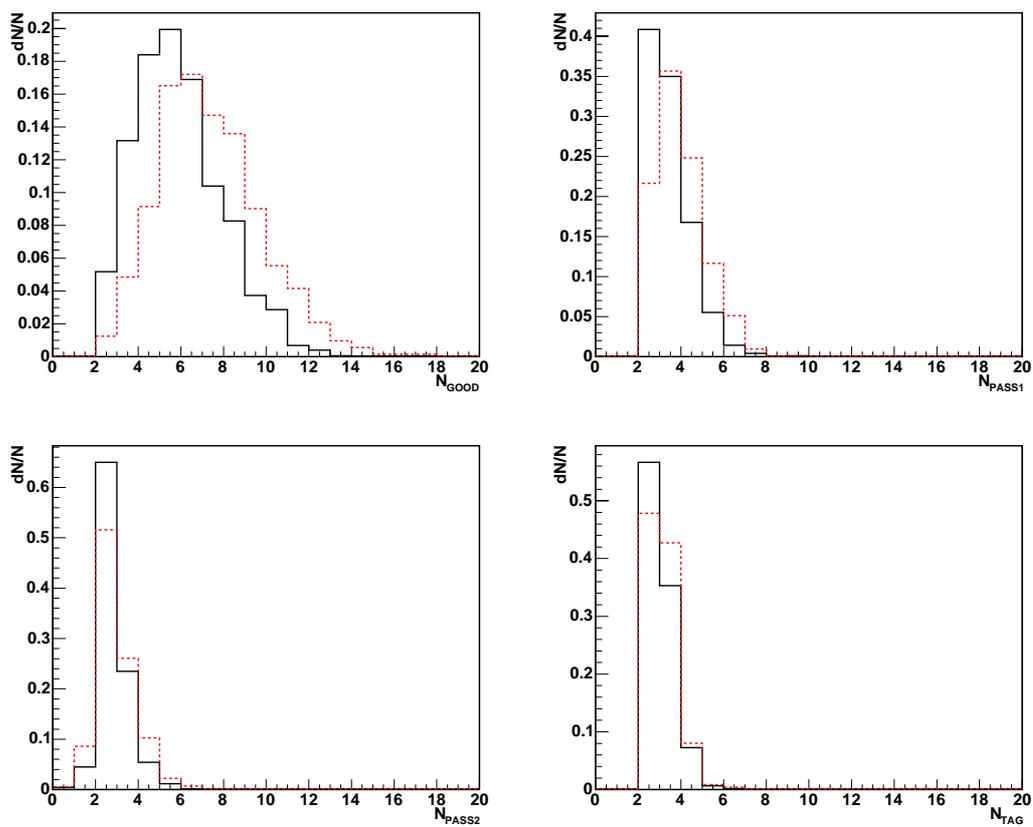


Figure 4.22: Track multiplicities for jets including one  $c$  quark (full line) or two  $c\bar{c}$  quarks, clockwise: number of good tracks, number of pass1 tracks, number of pass2 tracks, number of tag tracks (dashed line).

average number of pass1 and pass2 tracks with respect to single  $b$  quarks jets. This difference turns into a higher tagging efficiency as shown on the left of figure 4.23: the presence of the second  $b$  quark inside the same jet increases the efficiency from 10% to 13%, approximately.

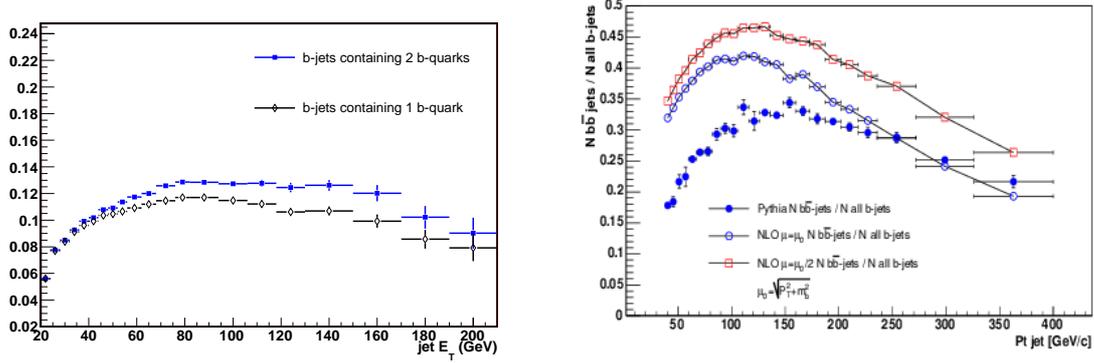


Figure 4.23: Left: SVT-tagging efficiency for jets containing two  $b$  quarks (empty diamonds) or one  $b$  quark (black squares). Right: Ratio of the number of  $b\bar{b}$ -jets with the number of all  $b$ -jets for Pythia (filled marker) and NLO prediction, as calculated for  $\mu_R = \mu_F = \mu_0$  (empty blue circle) and  $\mu_R = \mu_F = \mu_0/2$  (empty red square) [72].

PYTHIA predicts that 20-30% of  $b$ -jets reconstructed in the central region ( $|\eta| < 1.2$ ) include two  $b$  quarks inside their cone, as shown in figure 4.23 (right). A previous study proved that this rate is actually a factor of two below the rate predicted at the NLO [72]. So the problem is generally relevant. However in our case the situation is better: requiring two  $b$ -jets in the same event considerably lowers the fraction of double  $b$  quark jets.

An estimate of the fraction of events with one of the two  $b$ -jets including two  $b$  quarks together with the case in which both jets include two  $b$  quarks is shown in figure 4.24: in the first case the fraction is always lower 1.2 % while in the second case it goes below 0.5 %. It is reasonable to assume that even though the analysis relies on the fraction predicted by PYTHIA the final effect on the tagging efficiency and  $b\bar{b}$  fraction will be small.

In order to prove it, the uncertainties are estimated, as described below, recalculating the tagging efficiency and the  $b\bar{b}$  fraction assuming that the number of jets including two  $b$  quarks is twice the value predicted by PYTHIA (for the upper limits) and assuming it is one half of the PYTHIA value (for the lower limits). This choice gives a reasonable estimate on how the two quantities would be affected by a different double  $b$  jet rate than the one given by PYTHIA.

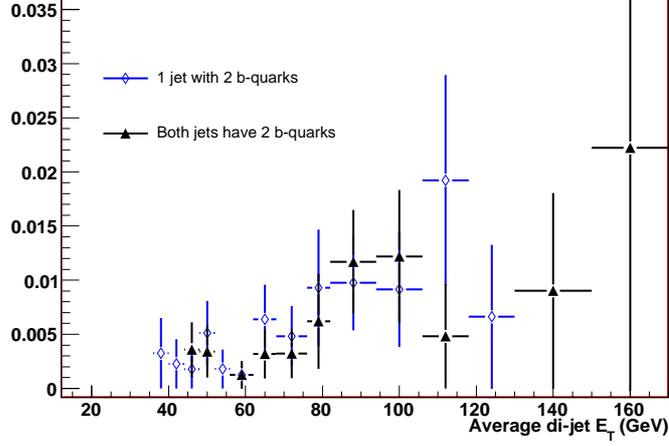


Figure 4.24: (left) Fraction of two  $b$ -jets events in which both  $b$ -jets (black triangles) or only one of the two (empty diamonds) contains two  $b$  quark.

### Effect on the tagging efficiency estimate

To calculate the effect of quark multiplicity inside the jet on the tagging efficiency, the following expression is used:

$$\epsilon = f_{2b2b} \cdot \epsilon_{2b}^2 + f_{2b1b} \cdot \epsilon_{2b} \cdot \epsilon_{1b} + f_{1b1b} \cdot \epsilon_{1b}^2$$

where

- $f_{2b2b}$  is the fraction of events in which each of the two  $b$  jets include two  $b$  quarks;
- $f_{2b1b}$  is the fraction of events in which one of the two  $b$  jets include two  $b$  quarks while the other contains only one  $b$  quark. It includes contribution from both jets, so that no factor 2 is needed in the tagging efficiency formula;
- $f_{1b1b}$  is the fraction of events in which both  $b$  jets include a single  $b$  quark each;
- $\epsilon_{2b}$  is the tagging efficiency for jets including two  $b$  quarks;
- $\epsilon_{1b}$  is the tagging efficiency for jets including one  $b$  quark.

In this estimate the efficiency curves are assumed, for simplicity, to be the same for each jet independently on the presence of other tagged jets in the event.

Figure 4.25 shows, on the left side, the result as a function of the di-jet average  $E_T$ : the efficiency corresponding to the PYTHIA estimate is compared to the two cases corresponding to doubling or reducing to one half the number of jets including two b quarks (namely, changing  $f_{2b2b}$  and  $f_{2b1b}$ ).

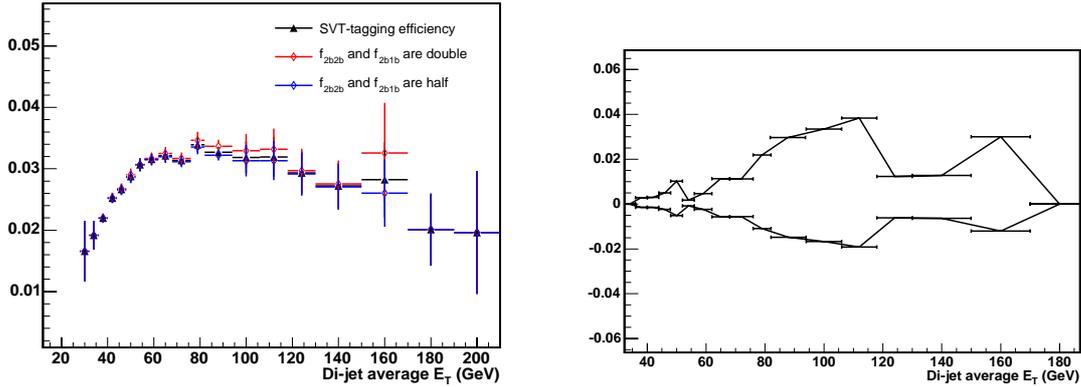


Figure 4.25: **left:** Tagging efficiency for  $b\bar{b}$  jet events (black triangles). Also shown are the curves associated to an increased fraction of jets including 2 b quarks (empty diamonds) and the case in which this fraction is reduced (green empty diamonds). **right** Systematic uncertainties corresponding to a doubled (or reduced to one half)  $f_{2b2b}$  and  $f_{2b1b}$  as a function of the leading jet  $E_T$ .

The effect is very small ( $< 3\%$ ) as a consequence of the fact that the two fractions  $f_{2b2b}$  and  $f_{2b1b}$  are both below 2%. Figure 4.25 also shows on the left the difference calculated with respect to the default values as a function of the di-jet average  $E_T$ : as expected increasing  $f_{2b2b}$  and  $f_{2b1b}$  increases the efficiency, while reducing their values consequently reduces the efficiency. The systematic uncertainty is summarized in table 4.4 and figure 4.26 as function of the leading jet  $E_T$ , in table 4.5 and figure 4.27 as a function of the di-jet invariant mass and, finally, in table 4.6 and figure 4.28 as a function of the di-jet invariant mass.

### Effect on the $b\bar{b}$ fraction estimate

The heavy quark content of each jet modifies the shape of the secondary vertex invariant mass templates and therefore it can be a source of systematic uncertainty on the measurement: the case of  $c$  jets is particularly interesting as the presence of two  $c$  quarks inside the same jet could fake the mass of a  $b$  jet. It turns out, however that this effect is very small on the final fraction of  $b$ -jets. As for the  $b\bar{b}$  case, to

$E_T$ bin (GeV)	syst. err. ( $f_{2b} \times 2$ ) (%)	syst. err. ( $f_{2b} \times 0.5$ ) (%)
(35,41)	1.2	-1.3
(41,47)	1.3	-1.4
(47,54)	1.4	-1.8
(54,61)	1.9	-2
(61,69)	1.8	-2.1
(69,78)	2.0	-2.4
(78,88)	2.1	-2.2
(88,99)	2.3	-1.9
(99,110)	2.5	-1.9
(110,140)	2.1	-2
(140,180)	1.9	-1.8
(180,230)	2	-1.7

Table 4.4: Tagging efficiency systematic uncertainties as a function of the leading jet  $E_T$ .

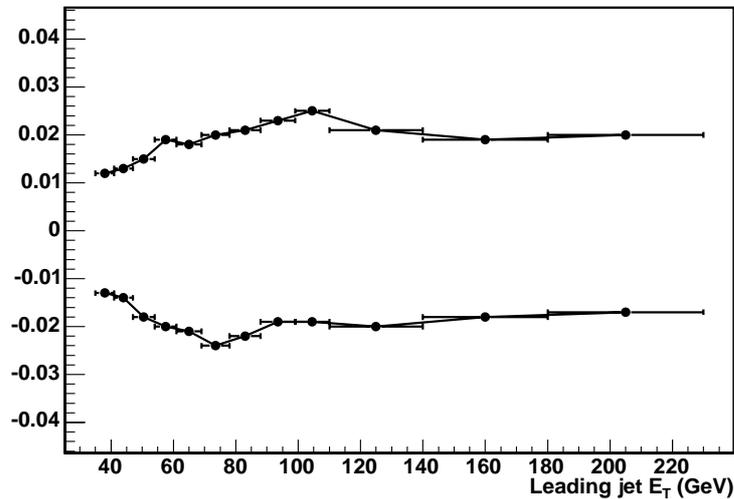


Figure 4.26: Tagging efficiency systematic uncertainties as a function of the leading jet  $E_T$ .

Invariant mass (GeV/ $c^2$ )	syst. err. ( $f_{2b} \times 2$ ) (%)	syst. err. ( $f_{2b} \times 0.5$ ) (%)
(50,60)	1	-1.6
(60,70)	1.5	-1.5
(70,80)	1.8	-1.5
(80,90)	2.1	-1.8
(90,100)	2.2	-2
(100,120)	2.5	-2.2
(120,140)	2.6	-1.9
(140,160)	2.4	-1.8
(160,190)	2.3	-1.7
(190,220)	2.2	-1.5
(220,250)	2.1	-1.6
(250,290)	2	-1.8
(290,340)	2.1	-1.6
(340,460)	1.9	-1.5

Table 4.5: Tagging efficiency systematic uncertainties as a function of the di-jet invariant mass  $M_{jj}$ .

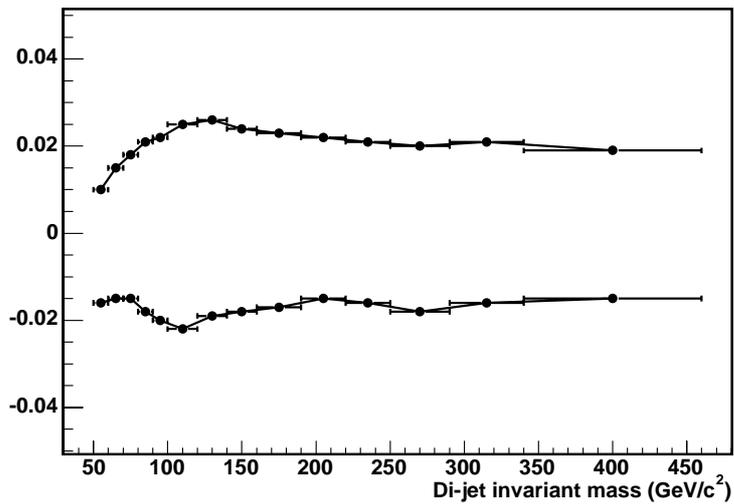
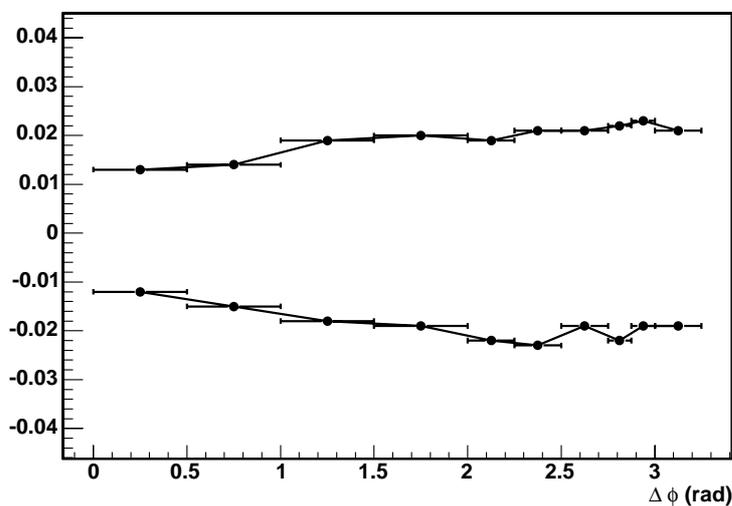


Figure 4.27: Tagging efficiency systematic uncertainties as a function of the di-jet invariant mass  $M_{jj}$ .

$\Delta\phi$ (rad)	syst. err. ( $f_{2b} \times 2$ ). (%)	syst. err. ( $f_{2b} \times 0.5$ ) (%)
(0.,0.5)	1.3	-1.2
(0.5,1.0)	1.4	-1.5
(1.0,1.5)	1.9	-1.8
(1.5,2.0)	2	-1.9
(2.0,2.25)	1.9	-2.2
(2.25,2.5)	2.1	-2.3
(2.5,2.75)	2.1	-1.9
(2.75,2.875)	2.2	-2.2
(2.875,3.0)	2.3	-1.9
(3.0,3.175)	2.1	-1.9

Table 4.6: Tagging efficiency systematic uncertainties as a function of the di-jet  $\Delta\phi$ .Figure 4.28: Tagging efficiency systematic uncertainties as a function of the di-jet  $\Delta\phi$ .

evaluate the effect the fraction of jets which include two  $b$ -quarks (or two  $c$ -quarks) has been doubled (and reduced to one half) with respect to the PYTHIA prediction.

Mass templates have been calculated in the four cases, (separately for  $b$ -jets and  $c$ -jets). Figure 4.29 and 4.30 show the results of the fits when the  $2b$ -quark contribution has been doubled or divided by two in the  $b\bar{b}$  invariant mass template. The fit are performed in 6 average jet  $E_t$  bins:  $32 < E_T < 40$ ,  $40 < E_T < 50$ ,  $50 < E_T < 60$ ,  $60 < E_T < 80$ ,  $80 < E_T < 100$  and  $E_T > 100$ .

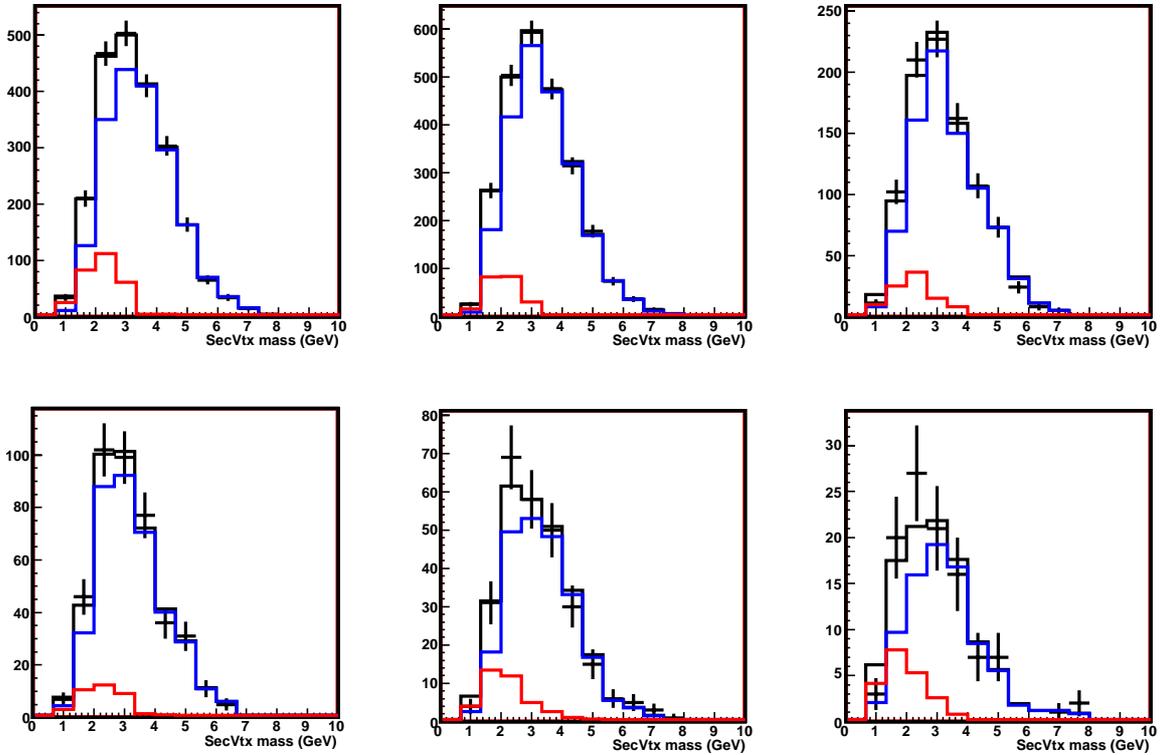


Figure 4.29: Fit to the secondary vertex mass for couple of jet with average  $E_t$  in different ranges (the contribution of jets with  $2b$  quarks inside has been doubled)

The resulting fraction are listed in table 4.7.

The two tables prove that the effect of  $b$ -quark content inside the jet is very small. The same result is found regarding the composition of  $c$ -jets templates: figures 4.31 and 4.32 show the fit in the six energy bins and table 4.8 the resulting  $b\bar{b}$  di-jet fractions.

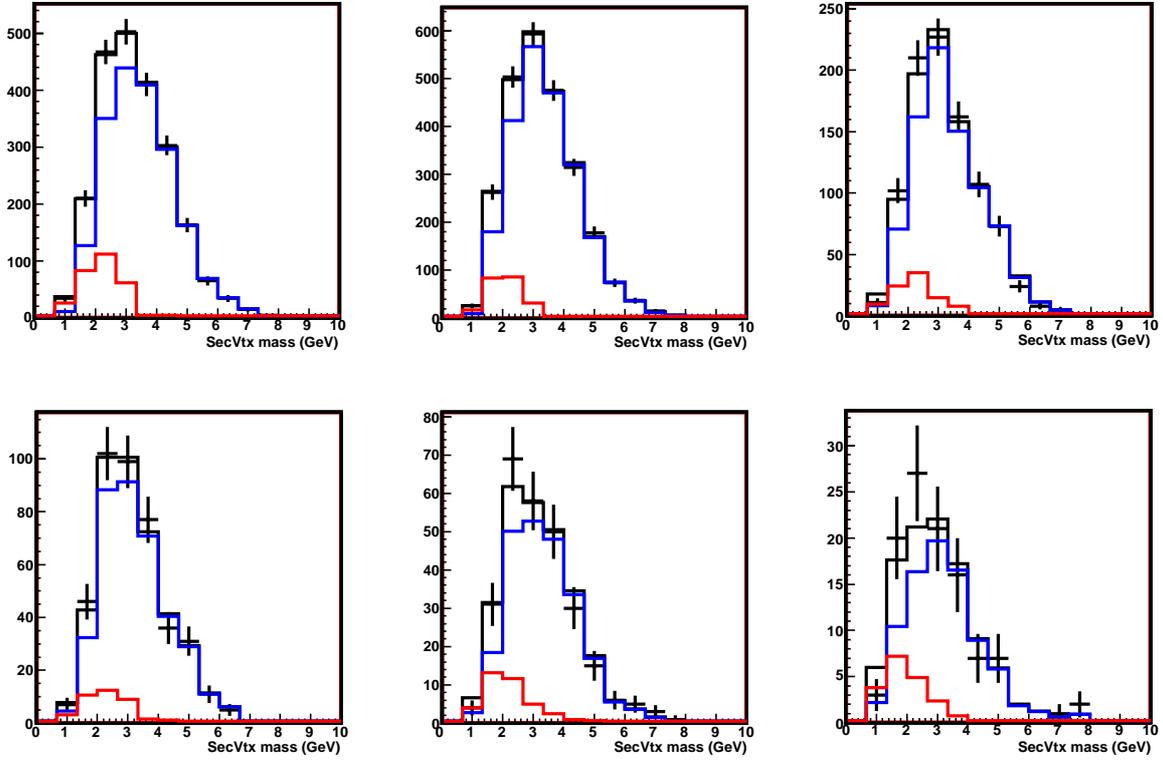


Figure 4.30: Fit to the secondary vertex mass for couple of jet with average  $E_t$  in different ranges (the contribution of jets with 2b quarks inside has been divided by two)

$E_T$ range (GeV)	$bb$ fraction ( $f_{2b} \times 2$ ).	$bb$ fraction ( $f_{2b} \times 0.5$ )
(32,40)	$0.87 \pm 0.04$	$0.88 \pm 0.04$
(40,50)	$0.88 \pm 0.03$	$0.88 \pm 0.03$
(50,60)	$0.86 \pm 0.05$	$0.83 \pm 0.05$
(60,80)	$0.78 \pm 0.07$	$0.79 \pm 0.07$
(80,100)	$0.83 \pm 0.07$	$0.84 \pm 0.07$
$E_T > 100$	$0.79 \pm 0.09$	$0.78 \pm 0.09$

Table 4.7:  $b\bar{b}$  fraction in di-jet average  $E_t$  ranges: the number of jets containing 2  $b$  quarks has been doubled and reduced to one half of the PYTHIA value. The last line shows the integrated result

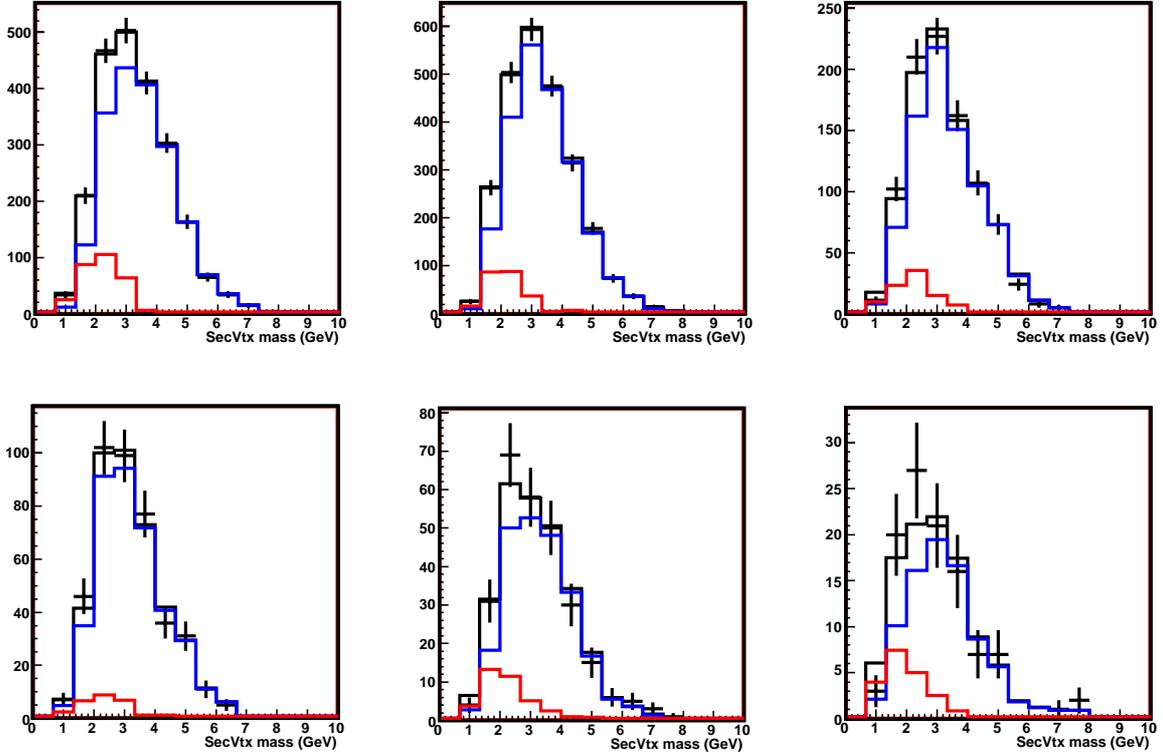


Figure 4.31: Fit to the secondary vertex mass for couple of jet with average  $E_t$  in different ranges (the contribution of jets with  $2c$  quarks inside has been doubled)

$E_T$ range (GeV)	$b\bar{b}$ fraction ( $f_{2c} \times 2$ ).	$b\bar{b}$ fraction ( $f_{2c} \times 0.5$ )
(32,40)	$0.88 \pm 0.04$	$0.88 \pm 0.04$
(40,50)	$0.87 \pm 0.03$	$0.88 \pm 0.03$
(50,60)	$0.86 \pm 0.05$	$0.87 \pm 0.05$
(60,80)	$0.80 \pm 0.07$	$0.79 \pm 0.07$
(80,100)	$0.83 \pm 0.07$	$0.84 \pm 0.07$
$E_T > 100$	$0.76 \pm 0.09$	$0.79 \pm 0.09$

Table 4.8:  $b\bar{b}$  fraction in di-jet average  $E_t$  ranges: the number of jets containing  $2c$  quarks has been doubled and reduced to one half of the PYTHIA value. The last line shows the integrated result

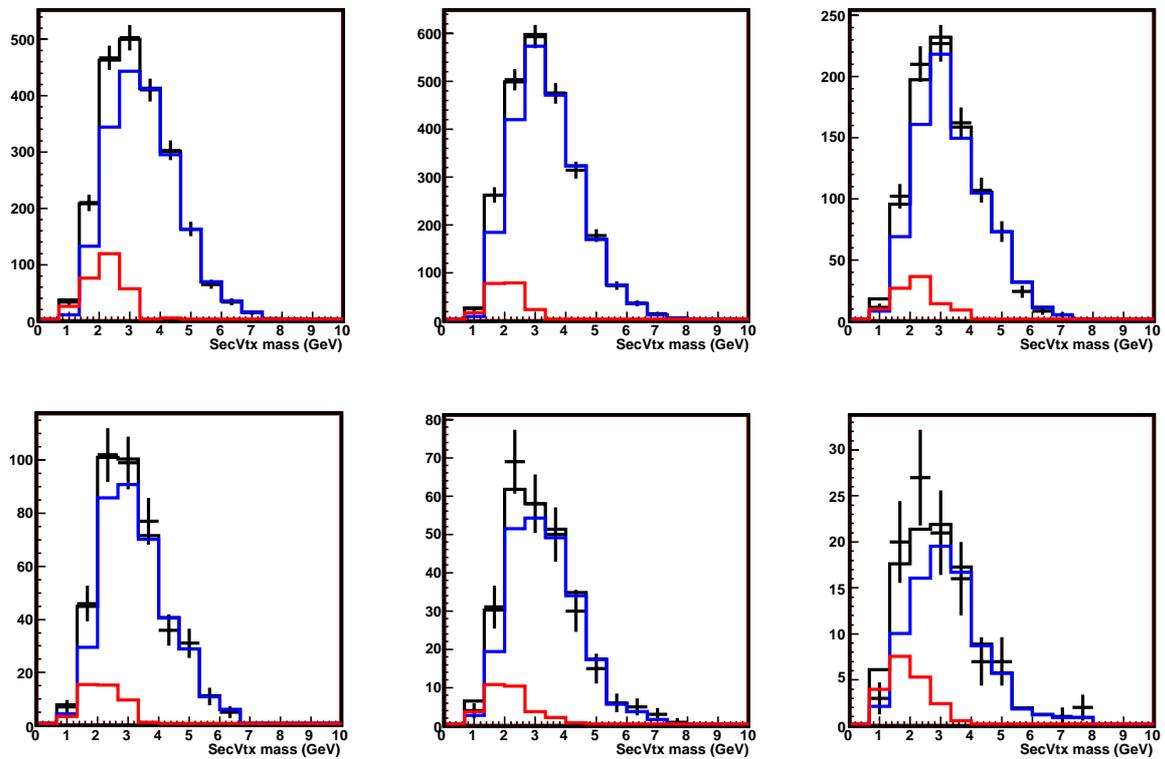


Figure 4.32: Fit to the secondary vertex mass for couple of jet with average  $E_t$  in different ranges (the contribution of jets with 2c quarks inside has been divided by two)

### 4.5.2 Track reconstruction inefficiency

Reconstruction efficiency for tracks inside a jet is, in general, expected to be different than for isolated tracks because of the merging or pattern recognition confusion of near hits. A study has been performed in [81] measuring the tracking efficiency for tracks within a jet as a function of jet  $E_T$ , track momentum and track direction in the jet. The results point out a non-negligible difference between data and Monte Carlo events: figure 4.33 shows the efficiency in Monte Carlo and data events<sup>3</sup> as a function of jet  $E_T$ . A 3% difference has been found in the low jet  $E_T$  region which can have an effect on the reconstruction of the invariant mass. Similar analysis involving secondary vertex tagging [82] have estimated that this problem can change the determination of the  $b$  fraction as consequence of a change in the shape of the secondary vertex invariant mass templates. The change has been estimated applying a 3% downward shift in mass templates as if the full efficiency difference turned into an equal shift on the mass templates: this assumption overestimates the uncertainty.

In the case of two jets simultaneously tagged, the sum of the invariant masses of the two jets is used to fit the fraction. Therefore a 3% shift is applied to each mass. The total templates are built and are used to re-fit the data. Figure 4.34 shows the result as a function of the di-jet average  $E_T$  compared to the  $b\bar{b}$  fraction in the case of regular templates. The full difference is taken as systematic uncertainty and summarized in tables 4.11, 4.12 and 4.13 as a function of the leading jet  $E_T$ , the di-jet invariant mass and the di-jet  $\Delta\phi$ , respectively: it represents the highest source of uncertainty on the determination of  $b\bar{b}$  fraction.

### 4.5.3 Other sources of systematic uncertainty on the $b\bar{b}$ fraction

- **Fragmentation:** Generators such as PYTHIA and HERWIG use different fragmentation functions to fragment heavy quarks. In addition, they use a different modeling of the underlying event. However the difference between HERWIG and PYTHIA have already been found negligible in many analyses ([66], [83]) and will not be considered at this level.
- **Construction of the  $b\bar{b}$  templates:** As already explained mass templates are built adding up together jets from different event in the Monte Carlo sample, therefore assuming the absence of any correlation between the masses of two tagged jets in the same event. To test this assumption the mass template for couples of  $b$  jets in the same event is compared to the mass templates built

---

<sup>3</sup>The ‘data’ events are actually represented by Monte Carlo events in which hits have been added to each track, (“embedded” in the reconstruction) to simulate the data.

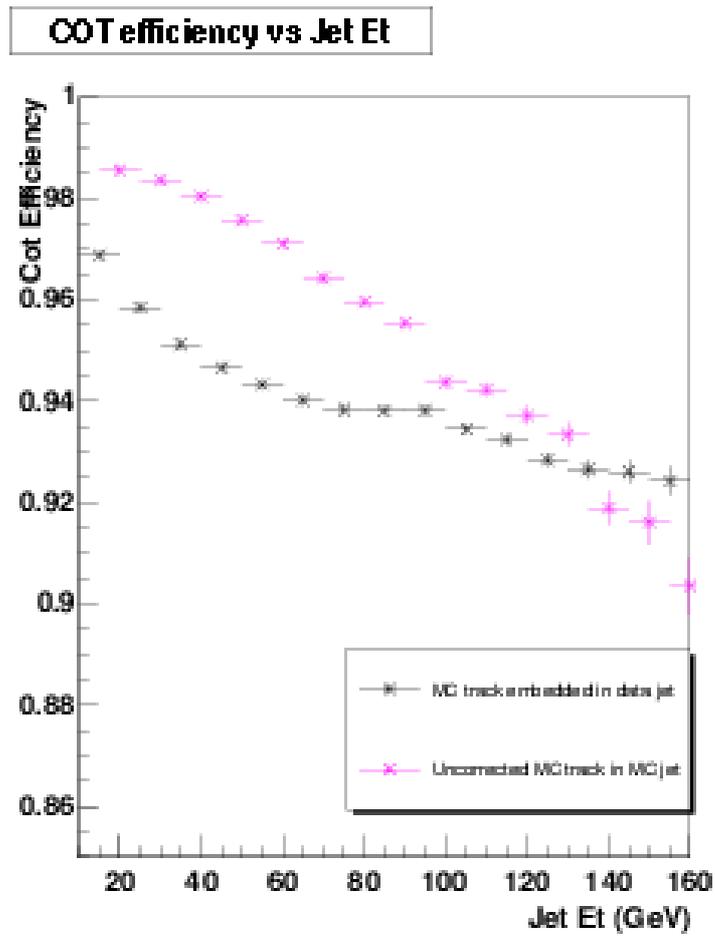


Figure 4.33: Reconstruction efficiency for tracks inside a jet in data and Monte Carlo samples.

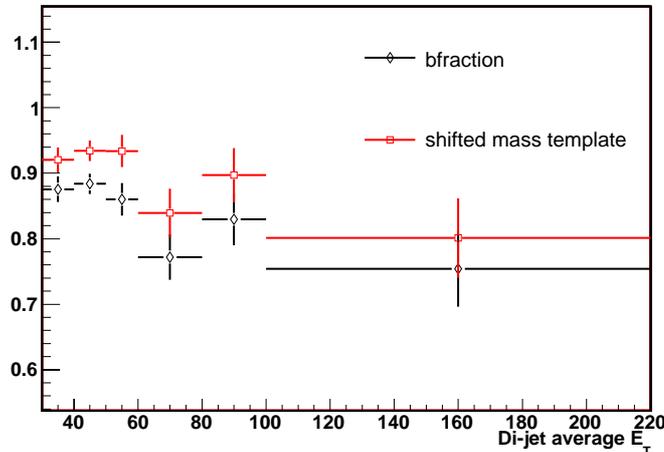


Figure 4.34: b fraction for b templates shifted downward of 3% to take into account the track reconstruction inefficiency effect

using two  $b$  jets from two different events. A Kolmogorov test is performed to verify the compatibility of the two distributions, shown in figure 4.35 and the difference is found to be negligible with respect to the total systematic uncertainty for the  $b\bar{b}$  fraction.

- **c/light jets mixing:** A larger systematic uncertainty is related to the contribution of light jets relative to heavy flavour jets in the non- $b\bar{b}$  template. Each combination has, in fact, a different average invariant mass and, since all the possible combination have been merged together to build the non- $b\bar{b}$  template, the final shape is dependent on the relative weight of each component (for example  $b\bar{l}$  with respect to  $c\bar{c}$  or to  $b\bar{c}$ ).

These fractions rely on the prediction obtained using PYTHIA, and it is necessary to estimate how big is the uncertainty related on this assumption. To do so the weight of light jets has been changed (to one half of its value and to the double of its value) and the corresponding  $b\bar{b}$  fraction has been recalculated.

The  $b\bar{b}$  fraction is compared to the central value in figure 4.36: the difference is taken as a systematic uncertainty on the  $b\bar{b}$  fraction. The values are listed in tables 4.11, 4.12 and 4.13 as a function of the leading jet  $E_T$ , the di-jet invariant mass and the di-jet  $\Delta\phi$ .

- **Monte Carlo statistics:** In order to fit the uncertainty due to statistical errors on the fit templates each bin of each template was treated as an independent

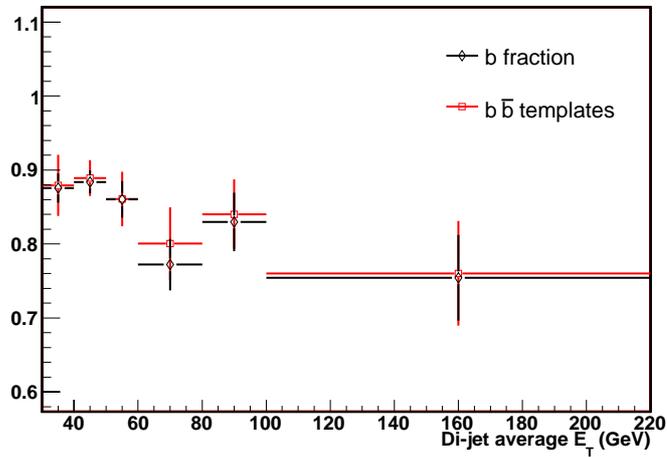


Figure 4.35:  $b$  fraction for  $b$  templates built adding  $b$  jets from independent events (black) or real  $b\bar{b}$  events (red). The systematic effect is neglected. The central values are compatible whereas the error bars are larger in the second case due to lower statistics.

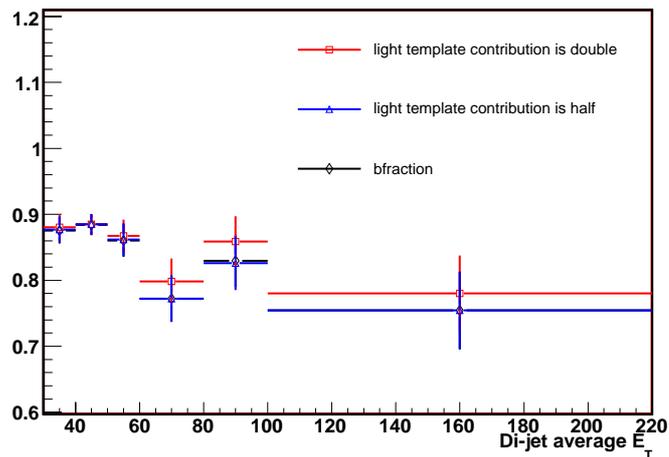


Figure 4.36:  $b$  fraction for templates in which the light jet contribution has been doubled (or reduced) with respect to the  $c$  jets

Bin ( $-1\sigma$ )	$b\bar{b}$ fraction	$\Delta(f_{nominal} - f_{-1\sigma})$ $f_{nominal} = 0.86 \pm 0.05$
1	0.86	0.00
2	0.86	0.00
3	0.86	0.00
4	0.87	-0.01
5	0.87	0.01
6	0.87	-0.01
7	0.87	0.01
8	0.86	0.00
9	0.87	0.01
10	0.86	0.00
11	0.86	0.00
12	0.86	0.00

Table 4.9: Test on the effect of Monte Carlo template statistics on  $b\bar{b}$  fraction:  $b$  template bin content are lowered in turn by  $1\sigma$

variable.

Starting with the first bin of the template, one by one the bins are adjusted downwards (or upwards) by one sigma. After each adjustment the fit is repeated and the difference between the result and the result from the standard fit using unmodified templates is recorded. This process is repeated in turn for the  $b\bar{b}$  template and the non- $b$  template, corresponding to the average  $E_T$  range in (60,80) GeV. The results are summarized in table 4.9 and 4.10. The effect on the total uncertainty is negligible.

#### 4.5.4 Total Uncertainty related to $b\bar{b}$ -jet fraction

Tables 4.11, 4.12 and 4.13 summarize the  $b\bar{b}$  fraction related systematic uncertainty as a function of the leading jet  $E_T$ , the di-jet invariant mass and  $\Delta\phi$ .

#### 4.5.5 The data/Monte Carlo scale factor

The official scale factor [74] is used with its systematic error. As explained in [74] different sources of systematic uncertainties are considered: MonteCarlo statistics, choice of non- $b$  template, the jet direction and  $b$ -decay multiplicity. The total value is  $\Delta = 3.3\%$ .

Bin ( $-1\sigma$ )	$b\bar{b}$ fraction	$\Delta(f_{nominal} - f_{-1\sigma})$ $f_{nominal} = 0.86 \pm 0.05$
1	0.86	0.00
2	0.86	0.00
3	0.85	0.01
4	0.85	-0.01
5	0.85	0.01
6	0.86	0.00
7	0.85	-0.01
8	0.87	0.01
9	0.86	0.00
10	0.86	0.00
11	0.86	0.00
12	0.86	0.00

Table 4.10: Test on the effect of Monte Carlo template statistics on  $b\bar{b}$  fraction: non- $b$  template bin content are lowered in turn by  $1\sigma$

$E_T$ bin (GeV)	syst. err. (%) Track ineff.	syst. err. (%) non- $b$ Templ. comp.	syst. err. (%) Total
(35,41)	5	0.5	5
(41,47)	5	0.5	5
(47,54)	5	0.5	5
(54,61)	6	0.7	6
(61,69)	6	0.5	8
(69,78)	7	0.5	8
(78,88)	8	1	8
(88,99)	8	1	9
(99,110)	8	2	9
(110,140)	8	2	8
(140,180)	7	2	8
(180,230)	7	2	8

Table 4.11:  $b\bar{b}$  fraction systematic uncertainties for each  $E_T$  bin. First column summarizes the uncertainty due to the track reconstruction efficiency. The second column summarizes the uncertainty due to the composition of the non- $b$  template. The third

Di-jet invariant mass (GeV/ $c^2$ )	syst. err. (%) Track ineff.	syst. err. (%) non- $b$ Templ. comp.	syst. err. (%) Total
(50,60)	5	0.5	6
(60,70)	5	0.5	6
(70,80)	5	0.5	5
(80,90)	6	0.5	5
(90,100)	6	0.5	6
(100,120)	7	0.6	6
(120,140)	8	1	7
(140,160)	8	1	7
(160,190)	7	1	8
(190,220)	7	1	8
(220,250)	8	2	8
(250,290)	6	2	8
(290,340)	6	2	8
(340,460)	6	2	8

Table 4.12:  $b\bar{b}$  fraction systematic uncertainties for each  $M_{jj}$  bin. First column summarizes the uncertainty due to the track reconstruction efficiency. The second column summarizes the uncertainty due to the composition of the non- $b$  template.

$\Delta\phi$ bin (rad)	syst. err. (%) Track ineff.	syst. err. (%) non- $b$ Templ. comp.	syst. err. (%) Total
(0.,0.5)	6	0.6	6
(0.5,1.0)	6	0.6	6
(1.0,1.5)	5	0.5	6
(1.5,2.0)	6	0.6	6
(2.0,2.25)	6	0.7	6
(2.25,2.5)	6	0.6	6
(2.5,2.75)	6	0.6	6
(2.75,2.875)	6	0.6	6
(2.875,3.0)	6	0.6	6
(3.0,3.175)	6	0.5	6

Table 4.13:  $b\bar{b}$  fraction systematic uncertainties for each  $\Delta\phi_{jj}$  bin. First column summarizes the uncertainty due to the track reconstruction efficiency. The second column summarizes the uncertainty due to the composition of the non- $b$  template.



# Chapter 5

## Results and comparison to Monte Carlo

The results for the  $b\bar{b}$  di-jet cross section measurement are presented in this chapter. The events are selected requiring two  $b$  tagged jets, and two SVT tracks according to the selection described in chapters 3 and 4. The  $b\bar{b}$  content has been extracted directly from data, following the procedure described in chapter 4. An additional correction is necessary to compare the measurement to hadron level Monte Carlo prediction and it is described in section 5.2. Section 5.4, finally presents the differential di-jet cross sections as a function of the leading jet  $E_T$ , the di-jet invariant mass and the di-jet  $\Delta\phi$  together with the total cross section. These results are compared to predictions obtained using PYTHIA, HERWIG and MC@NLO in section 5.5. An additional discussion on the role played by the underlying event can be found in section 5.5.1. Conclusions, including a brief comparison to a preliminary measurement performed by CDF at the beginning of Run II, are reported in section 5.6.

### 5.1 The raw cross sections

The raw  $b\bar{b}$  di-jet cross section is calculated as function of the leading jet  $E_T$ , the di-jet invariant mass and the di-jet  $\Delta\phi$ , using the formula:

$$\frac{d\sigma_{b\bar{b}}^2}{dXd\eta} = \frac{N_{2SVT} \cdot F_{2SVT}^{2b}}{\epsilon_{2SVT} \cdot SF_{E_{T,1}} \cdot SF_{E_{T,2}} \cdot \mathcal{L} \cdot d\eta dX}$$

where  $X$  stands for each of the three variables ( $E_T$ ,  $M_{jj}$  and  $\Delta\phi$ ) and:

- $N_{2SVT}$  is the number of events in the sample including two SVT tagged jets and satisfying the conditions:  $E_{T,1} > 35$  GeV,  $E_{T,2} > 32$  GeV,  $|\eta_{1,2}| < 1.2$

- $F_{2SVT}^{2b}$  is the fraction of these events that contain two  $b$ -jets.
- $\epsilon_{2SVT}$  is the efficiency for requiring two SVT tagged jets in the event.
- $SF_{E_{T,1}}$  and  $SF_{E_{T,2}}$  are the scale factor values as a function of jet energy;
- $\mathcal{L}$  is the integrated luminosity; the correction factor 1.019 (for combined CDF and  $D\bar{O}$  luminosity measurements) is applied in the final calculation.

The differential cross sections are ‘raw’ in the sense that they are still affected by residual detector effects due to calorimeter energy resolution and need an additional correction in order to be compared to theoretical predictions. This correction is described in the following section.

## 5.2 Unfolding procedure: correction to the hadron level

The specific energy scale correction applied jet by jet, as described in chapter 3, is an average correction: it doesn’t account for smearing effects or jet migration from a bin to the adjacent ones as a consequence of the finite energy resolution of the calorimeter and the jet energy correction itself.

To measure a particle level differential cross section free of detector effect, an additional correction, generally referred to as *unfolding*, is necessary. The impact of this correction, carried out bin-by-bin using Monte Carlo samples, is quite important.

The procedure applied to find the unfolding correction factors is summarized as follows:

- Hadronic and calorimetric jets are reconstructed using generator level final state particles or calorimeter towers, respectively.
- Hadronic and calorimetric jets are identified as  $b$ -jets, requiring that a  $B$  hadron is found inside the jet cone ( $\Delta R < 0.4$ ). No requirement on the  $B$  hadron is applied, nor any matching between the calorimeter and hadron level jets.
- Hadronic and calorimetric jet differential cross sections are calculated as a function of the leading jet  $E_T$ ,  $M_{jj}$  and  $\Delta\phi_{jj}$ . No cuts (other than  $|\eta| < 1.2$  and  $E_T > 35$  GeV for the leading jet and  $E_T > 32$  GeV for the second jet) are applied to hadron or calorimeter level jets.
- The unfolding factors, defined as the bin-by-bin ratio of the calorimeter jet and hadronic jet cross sections, are calculated as

$$C_{unfold}^i = \left[ \frac{d^2\sigma_{HAD}/dXd\eta}{d^2\sigma_{CAL}/dXd\eta} \right]^i = \frac{N_{HAD}^i}{N_{CAL}^i}$$

$X$  represents, in turn, the leading jet  $E_T$ ,  $M_{jj}$  and  $\Delta\phi_{jj}$ .

$N_{HAD}^i$  is the number of hadron level  $b$  jets in bin  $i$  and  $N_{CAL}^i$  the number of calorimeter level  $b$  jets in bin  $i$ .

- The bin-by-bin factors are applied to the raw  $b$  jet cross sections to obtain the cross sections at particle level:

$$\left. \frac{d^2\sigma}{dXd\eta} \right|_i^{unfolded} = \left. \frac{d^2\sigma}{dXd\eta} \right|_i^{raw} \cdot C_{unfold}^i$$

Figures 5.1 and 5.2 show the unfolding factors as a function of leading jet  $E_T$ , of  $M_{jj}$  and  $\Delta\phi_{jj}$  in PYTHIA. Factors calculated using HERWIG are reported in appendix C.

As the figures show, the correction factors are of the order 20-50%, depending on the energy range. Two different effects contribute to these factors: the correction for the acceptance of the basic event selection cuts, namely the  $Z$  vertex cut, and the real smearing convoluted with the specific correction for  $b$  jets. The acceptance contributes only about 10% so that smearing and  $b$  specific jet corrections are dominant.

This is expected, since the average  $E_T$  correction procedure is implemented for general tagged jets and not for real  $b$  jets; together with the bias due to the tagger, we have to consider, for example, that about 23% of  $b$  hadrons decay semileptonically ( $e, \mu, \tau$ ), resulting in an underestimation of the parton energy due to lost neutrinos.

Additional studies on the connection between unfolding factors and average energy corrections can be found in appendix C.

### 5.2.1 Systematic uncertainty on unfolding factors

The main source of systematic uncertainty to take into account in the calculation of the unfolding factors arises from the difference related to the  $E_T$  spectrum in data and Monte Carlo.

Fig. 5.3 shows the ratio of the unfolded cross section in data as a function of leading jet  $E_T$  to the cross section predicted by PYTHIA. If the Monte Carlo  $E_T$  spectrum perfectly matched the one on data, the ratio would be flat in  $E_T$ . The fact that it is not the case points out a difference between the jet  $E_T$  spectrum in the MonteCarlo and data. The effect has already been described in [66]. It is associated to the  $E_T$  shape given by CTEQ5L PDF functions implemented in PYTHIA and it causes a bias when calculating the unfolding factors.

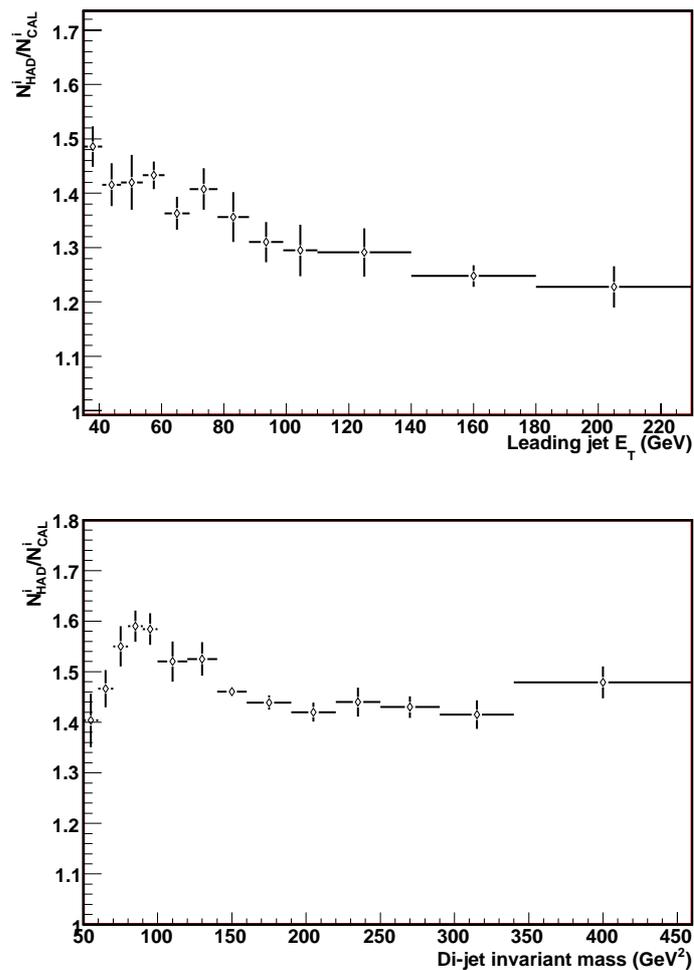


Figure 5.1: Unfolding factors calculated using PYTHIA as a function of the leading jet  $E_T$  (top) and the invariant mass  $M_{jj}$  of the two jets (bottom).

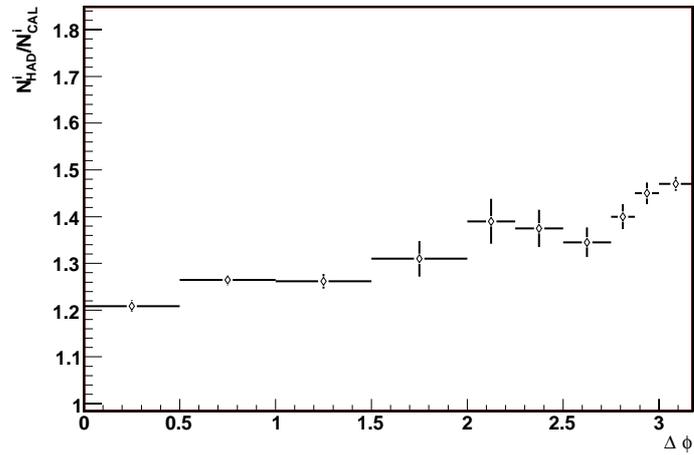


Figure 5.2: Unfolding factors calculated using PYTHIA as a function of the  $\Delta\phi_{jj}$  between the two jets.

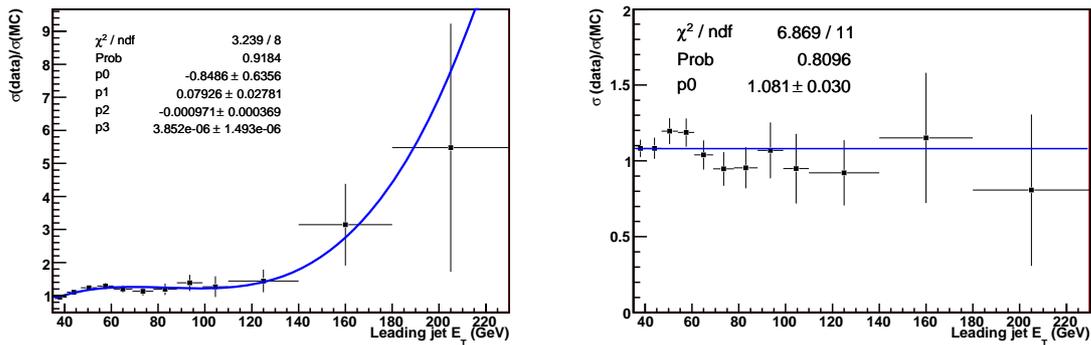


Figure 5.3: (Left) Data/MC ratio of the  $b\bar{b}$  cross section as a function of the leading jet  $E_T$ . (Right) Same ratio after re-weighting Pythia.

$E_T$ bin (GeV)	syst. err. (%)
(35,41)	2.1
(41,47)	0.6
(47,54)	5.2
(54,61)	1.4
(61,69)	0.2
(69,78)	0.7
(78,88)	1.2
(88,99)	1.7
(99,110)	2.4
(110,140)	1.9
(140,180)	4.2
(180,230)	6.7

Table 5.1: Data/MC  $E_T$  spectrum systematic uncertainties for each  $E_T$  bin.

To estimate the associated systematic, this ratio is fit to a third order polynomial and the unfolding factors are recalculated after re-weighting each event in the Monte Carlo for the result of the fit. Fig. 5.3 shows, on the right, the ratio Data/MC after the re-weighting procedure: the distribution is compatible with a flat fit. The full difference on the unfolded cross sections between the two cases is taken as a systematic uncertainty.

The total uncertainties are summarized in table 5.1, table 5.2 and table 5.3.

### 5.3 Total systematic uncertainties

The systematic uncertainty on the  $b$  jet cross sections results from five main sources: calorimeter response, resolution,  $b$ -tagging and tagged jet purity, unfolding factors and luminosity. While the uncertainty on the luminosity has no phase space dependence, the other systematic errors are found as a function of the jet  $E_T$ , the di-jet  $M_{jj}$  and  $\Delta\phi_{jj}$ . Each uncertainty has been described in the section related to the quantity it refers to, in chapters 3 and 4. The statistical errors on the tagging efficiency, the  $b\bar{b}$  fraction and the unfolding factors are not included among the systematic errors. They are accounted for in the total statistical uncertainty as explained in the next section. Here a brief summary of the different systematic uncertainty is given.

- **Jet energy scale:** The systematic uncertainty associated to the jet energy scale is evaluated varying the jet  $E_T$  in data (leaving Monte Carlo unchanged) by  $\pm 1\sigma$ . As already shown in section 3.6. This quantity has been evaluated for

Di-jet invariant mass bin ( $\text{GeV}/c^2$ )	syst. err. (%)
(50,60)	8.1
(60,70)	2.5
(70,80)	3.8
(80,90)	2.3
(90,100)	1.3
(100,120)	0.5
(120,140)	0.8
(140,160)	0.6
(160,190)	5.6
(190,220)	6.2
(220,250)	9.1
(250,290)	8.8
(290,340)	9.7
(340,460)	10.1

Table 5.2: Data/MC  $E_T$  spectrum systematic uncertainties for each invariant mass bin.

$\Delta\phi$ bin (rad)	syst. err. (%)
(0.,0.5)	3.1
(0.5,1.0)	2.6
(1.0,1.5)	2.7
(1.5,2.0)	1.5
(2.0,2.25)	3.6
(2.25,2.5)	0.5
(2.5,2.75)	0.6
(2.75,2.875)	3.6
(2.875,3.0)	4.1
(3.0,3.175)	3.4

Table 5.3: Data/MC  $E_T$  spectrum systematic uncertainties for each  $\Delta\phi$  bin.

common use in CDF [75] and arises from the calorimeter simulation, modeling of the fragmentation by the Monte Carlo, and stability of the calorimeter. It represents the largest source of systematic uncertainty, being of the order of 15% to 22% in different regions of the phase space.

- **SVT tagging efficiency:** The uncertainty on the SVT-tagging efficiency is due to two main parts: the first one, related to the tagging algorithm, is due to the multiplicity of heavy flavour quarks inside the jet cone which, as we have seen, changes by about 10% the probability for a jet to be tagged. The second part is due to the fact that the tagging efficiency is measured together with the SVT efficiency from the Monte Carlo, assuming that the SVT simulation is correct with respect to data. This assumption has not been proved. In fact each event is corrected for a scale factor data/Monte Carlo to correct for possible differences between data and simulation. This part of the uncertainty is accounted for in the scale factor uncertainty which is about 3%, as described in chapter 4. The total uncertainty on  $b$ -tagging efficiency is of the order of 4-5%.
- **$b\bar{b}$  purity:** This quantity is strongly dependent on the shape of the Monte Carlo templates, which can change the result of the fit. The main source of systematic (7-8%) is related to the tracking efficiency inside the jet, which affects the number of tracks attached to the secondary vertex and changes the value of the associated invariant mass. A second contribution of the order of 3-4% is due to the composition of the non  $b$  jet template and, in particular, to the proportion of the light jets component to the heavy flavour component as predicted by PYTHIA. The multiplicity of heavy flavour quarks inside the jet has a small effect on the final  $b\bar{b}$  fraction value.
- **Unfolding:** the main source of systematic uncertainty affecting the unfolding is related to the  $E_T$  spectrum. It has been evaluated in the previous section as few percents: about 1-2% at low  $E_T$  and 6 – 7% at high  $E_T$ .
- **Luminosity:** the integrated luminosity enters in the cross section as a normalization factor. Thus, its uncertainty enters directly in the cross section. The official CDF 6% value is used.

The contributions from different sources are summed in quadrature to obtain the overall uncertainty as a function of the leading jet  $E_T$ ,  $M_{jj}$  and  $\Delta\phi$ . The final values are summarized in tables 5.4, 5.5 and 5.6 and figures 5.4, 5.5 and 5.6.

$E_T$ bin (GeV)	Total syst. err.(%)	Total syst. err. (%)
(35,41)	25	-24
(41,47)	26	-24
(47,54)	23	-23
(54,61)	23	-22
(61,69)	21	-19
(69,78)	24	-19
(78,88)	20	-18
(88,99)	24	-21
(99,110)	19	-16
(110,140)	25	-21
(140,180)	23	-19
(180,230)	25	-21

Table 5.4: Total systematic uncertainties for each  $E_t$  bin. Luminosity 6% uncertainty is not included.

Di-jet invariant mass bin ( $\text{GeV}/c^2$ )	Total syst. err (%) .	Total syst. err. (%)
(50,60)	23	-21
(60,70)	22	-20
(70,80)	25	-22
(80,90)	27	-22
(90,100)	26	-22
(100,120)	25	-21
(120,140)	27	-20
(140,160)	24	-19
(160,190)	26	-21
(190,220)	22	-20
(220,250)	26	-21
(250,290)	27	-22
(290,340)	28	-26
(340,460)	28	-26

Table 5.5: Total systematic uncertainties for each  $E_t$  bin. Luminosity 6% uncertainty is not included.

$\Delta\phi$ (rad)	syst. err. (%)	syst. err. (%)
(0.,0.5)	21	-21
(0.5,1.0)	19	-19
(1.0,1.5)	20	-19
(1.5,2.0)	21	-19
(2.0,2.25)	19	-19
(2.25,2.5)	22	-21
(2.5,2.75)	22	-19
(2.75,2.875)	19	-19
(2.875,3.0)	19	-19
(3.0,3.175)	19	-19

Table 5.6: Total systematic uncertainties for each  $\Delta\phi$  bin. Luminosity 6% uncertainty is not included.

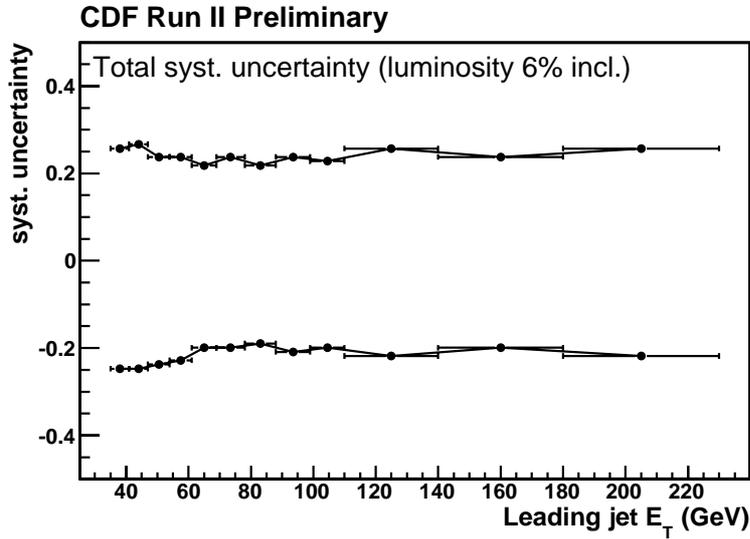


Figure 5.4: Total systematic uncertainties as a function of leading jet  $E_t$ . Luminosity 6% uncertainty is included.

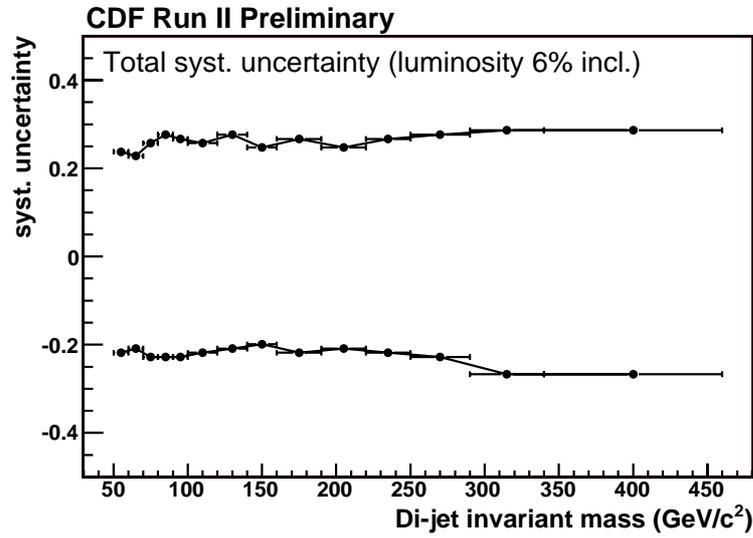


Figure 5.5: Total systematic uncertainties as a function of di-jet invariant mass. Luminosity 6% uncertainty is included.

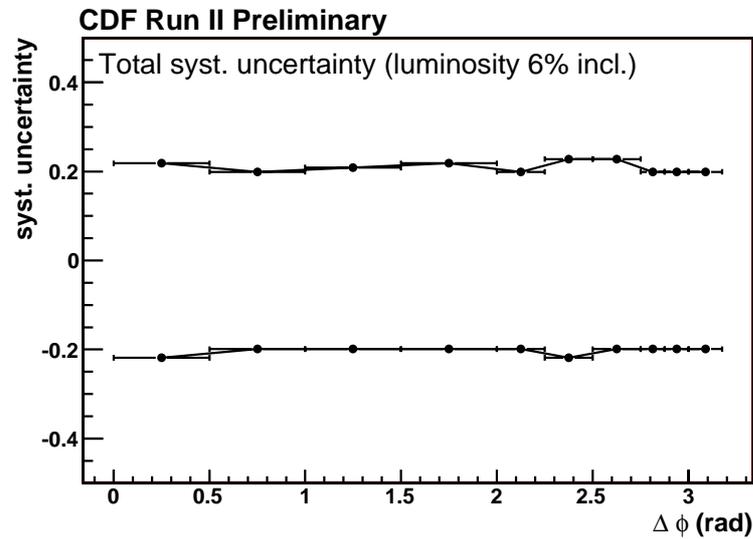


Figure 5.6: Total systematic uncertainties as a function of  $\Delta\phi$ . Luminosity 6% uncertainty is included.

## 5.4 Results

Figure 5.7 shows the differential cross section, corrected to the hadron level, as a function of the leading jet  $E_t$ . The cross section drops three orders of magnitude over a range that goes from 35 GeV up to 220 GeV in energy jet. The same behavior is observed when measuring the differential cross section as a function of the di-jet invariant mass  $M_{jj}$  in figure 5.8. In this case the two jets have an invariant mass that goes as low as 50 GeV/ $c^2$  and up to 450 GeV/ $c^2$ . The rising of the cross section in the low mass region, the first few bins of the distribution, is due to a residual kinematic effect generated by the double asymmetrical cut on the transverse energy of the two jets ( $E_{T,1} > 35$  GeV and  $E_{T,2} > 32$  GeV).

The di-jet  $\Delta\phi$  distribution is shown in figure 5.9. It clearly peaks at large  $\Delta\phi$  values, showing that the back-to-back configuration, mainly due to flavour creation processes, is dominant. However the range corresponding to a small opening angle ( $\Delta\phi < 1.5$ ) is still highly populated, suggesting a non negligible contribution of the higher order processes.

The statistical uncertainty on the unfolded cross sections is calculated as

$$\delta \left( \frac{d^2\sigma}{dXd\eta} \right) = \delta \left( \frac{N \cdot f_{b\bar{b}}}{\epsilon \cdot \epsilon \cdot SF \cdot SF \cdot \Delta X \cdot \Delta\eta \cdot \mathcal{L}} \cdot C_i \right)$$

$$= \frac{d^2\sigma}{dXd\eta} \cdot \sqrt{(\delta N/N)^2 + (\delta f_{b\bar{b}}/f_{b\bar{b}})^2 + (\delta\epsilon/\epsilon)^2 + (\delta\epsilon/\epsilon)^2 + 2 \cdot (\delta SF/SF)^2 + (\delta C_i/C_i)^2}$$

where  $\delta N/N$  comes from the counting of double tagged jet events,  $\delta f_{b\bar{b}}/f_{b\bar{b}}$  is the statistical error on the  $b\bar{b}$  fraction,  $\delta\epsilon/\epsilon$  represents the statistical error on the SVT-tagging efficiency and  $\delta C_i/C_i$  accounts for the statistical uncertainty on the unfolding factors. The term related to the scale factor  $\delta(SF)/SF$  is considered twice because two jets are selected in the event.

The measurement is compared to PYTHIA and HERWIG Monte Carlo and to next to leading order prediction obtained using the MC@NLO generator combined to HERWIG, for the parton shower, and to JIMMY for the underlying event as described in section 5.5.

The Monte Carlo distributions do not include any theoretical uncertainty. The error bars represent statistical errors only. The cross section values as a function of leading jet  $E_T$ , the di-jet  $M_{jj}$  and the di-jet  $\Delta\phi_{jj}$  are summarized in table 5.7, table 5.8 and table 5.9 respectively.

### 5.4.1 The total cross section

The integration of the differential cross section gives the value of the total cross section, shown in table 5.10, corresponding to the phase space  $E_{T,1} > 35$  GeV,

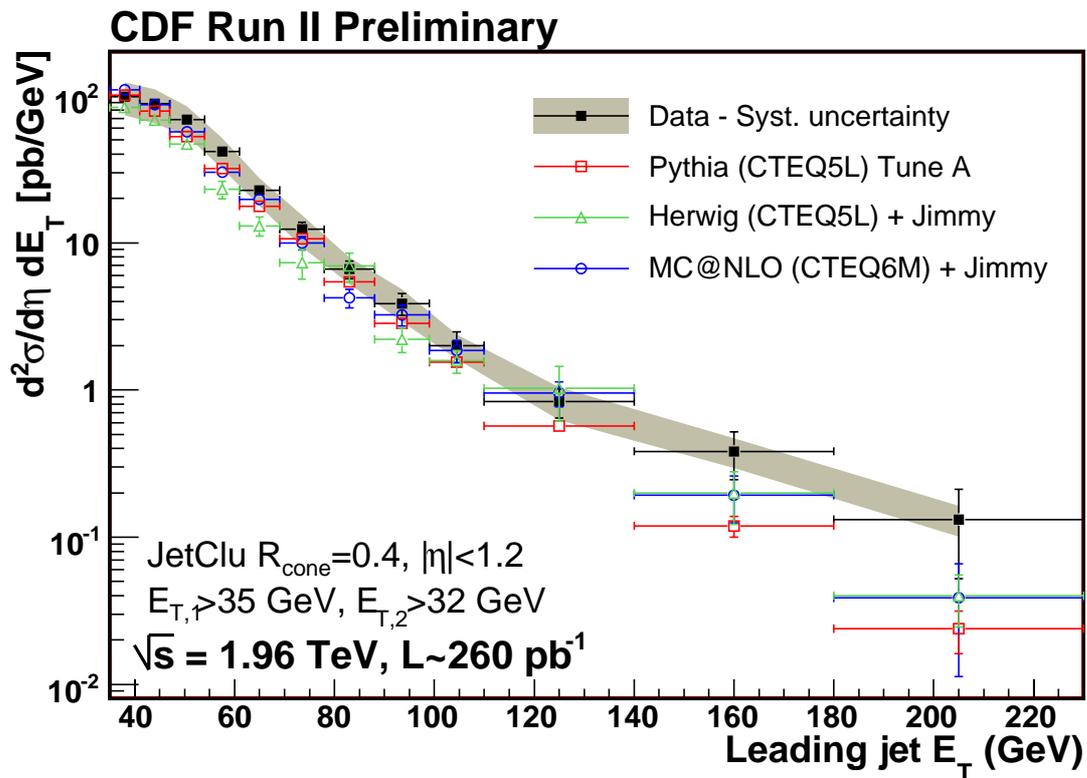


Figure 5.7:  $b\bar{b}$  differential cross section for jets as function of leading jet  $E_t$ . Data is compared to Pythia hadron level cross section and Herwig hadron level cross section. The shaded area represents the systematic total uncertainty on the data.

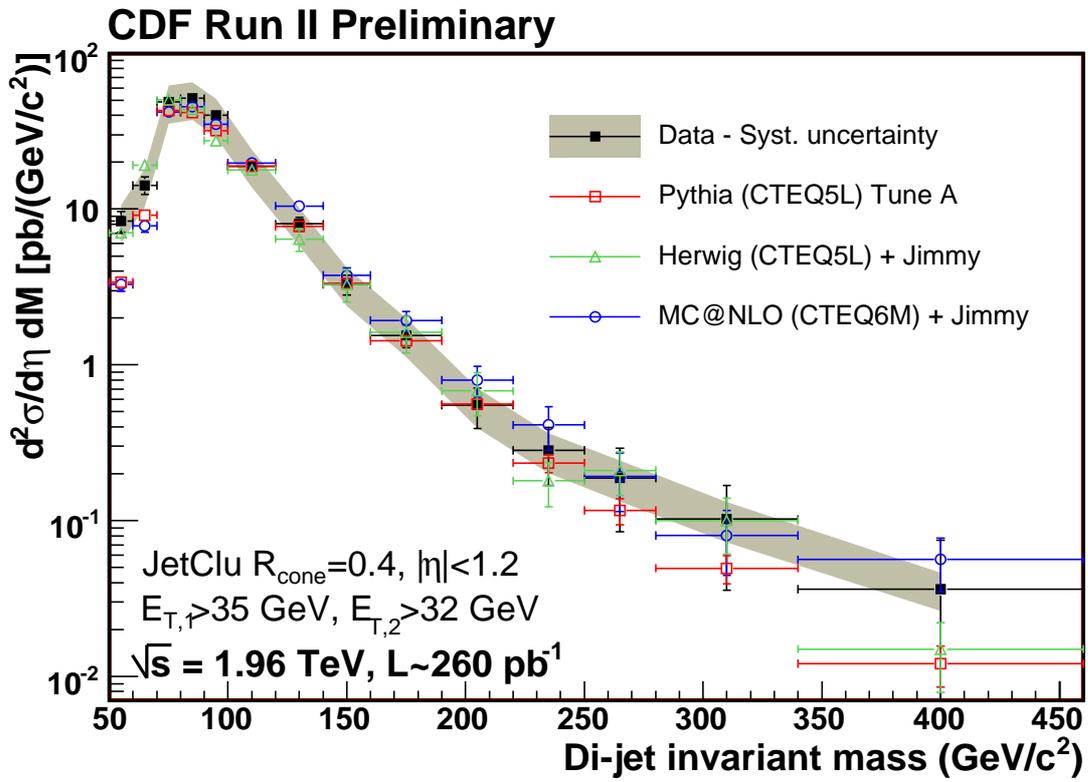


Figure 5.8:  $b\bar{b}$  differential cross section for jets as function of di-jet invariant mass. Data is compared to Pythia hadron level cross section and Herwig hadron level cross section. The shaded area represents the systematic total uncertainty on the data.

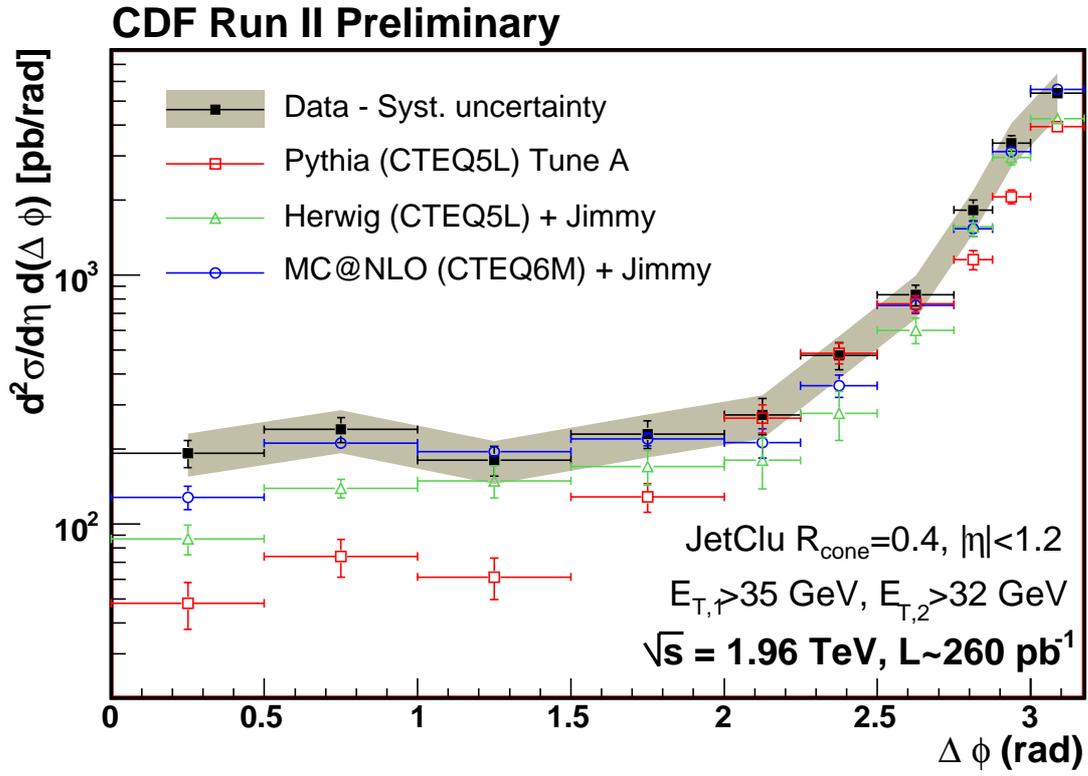


Figure 5.9:  $b\bar{b}$  differential cross section for jets as function of di-jet azimuthal angle. Data is compared to Pythia hadron level cross section and Herwig hadron level cross section. The shaded area represents the systematic total uncertainty on the data.

$E_T$ [GeV]	$d\sigma^2/dE_T d\eta \pm stat. \pm syst.$ [pb/GeV]
(35,41)	$98 \pm 6^{+25}_{-24}$
(41,47)	$88 \pm 6^{+22}_{-21}$
(47,54)	$69 \pm 5^{+15}_{-15}$
(54,61)	$42 \pm 3^{+9}_{-9}$
(61,69)	$23 \pm 2^{+5}_{-4}$
(69,78)	$12 \pm 1^{+2.8}_{-2.2}$
(78,88)	$6.6 \pm 0.9^{+1.3}_{-1.2}$
(88,99)	$3.9 \pm 0.6^{+0.93}_{-0.82}$
(99,110)	$2.0 \pm 0.4^{+0.38}_{-0.32}$
(110,140)	$0.83 \pm 0.18^{+0.21}_{-0.17}$
(140,180)	$0.38 \pm 0.14^{+0.08}_{-0.07}$
(180,230)	$0.13 \pm 0.08^{+0.03}_{-0.03}$

Table 5.7: Cross section values as a function of the leading jet  $E_T$ .

$M_{jj}$ [GeV/c <sup>2</sup> ]	$d\sigma^2/dM_{jj} d\eta \pm stat. \pm syst.$ [pb/GeV/c <sup>2</sup> ]
(50,60)	$8.3 \pm 1.2^{+1.9}_{-1.7}$
(60,70)	$14.2 \pm 1.8^{+3.1}_{-2.8}$
(70,80)	$48.6 \pm 3.7^{+12.5}_{-10.7}$
(80,90)	$51.4 \pm 3.4^{+13.8}_{-11.3}$
(90,100)	$40.0 \pm 3.1^{+10.4}_{-8.8}$
(100,120)	$18.8 \pm 1.4^{+4.7}_{-3.9}$
(120,140)	$8.0 \pm 0.8^{+2.1}_{-1.6}$
(140,160)	$3.3 \pm 0.5^{+0.8}_{-0.6}$
(160,190)	$1.5 \pm 0.2^{+0.4}_{-0.3}$
(190,220)	$0.55 \pm 0.15^{+0.12}_{-0.11}$
(220,250)	$0.28 \pm 0.11^{+0.07}_{-0.06}$
(250,290)	$0.19 \pm 0.11^{+0.05}_{-0.04}$
(290,340)	$0.10 \pm 0.06^{+0.03}_{-0.02}$
(340,460)	$0.036 \pm 0.021^{+0.01}_{-0.01}$

Table 5.8: Cross section values as a function of the di-jet invariant mass  $M_{jj}$ .

$\Delta\phi_{jj}$ [rad]	$d\sigma^2/d\Delta\phi_{jj}d\eta \pm \text{stat. pm syst.}$ [pb/rad]
(0.0,0.5)	$192 \pm 24^{+40}_{-40}$
(0.5,1.0)	$239 \pm 27^{+45}_{-45}$
(1.0,1.5)	$180 \pm 24^{+36}_{-34}$
(1.5,2.0)	$229 \pm 29^{+48}_{-43}$
(2.0,2.25)	$273 \pm 45^{+52}_{-52}$
(2.25,2.5)	$475 \pm 59^{+104}_{-100}$
(2.5,2.75)	$831 \pm 79^{+183}_{-156}$
(2.75,2.875)	$1822 \pm 171^{+346}_{-346}$
(2.875,3.0)	$3386 \pm 242^{+643}_{-643}$
(3.0,3.175)	$5376 \pm 264^{+1021}_{-1021}$

Table 5.9: Cross section values as a function of the di-jet invariant mass  $\Delta_{jj}$ .

CDF Run II Preliminary	$\sigma$ [pb]
	$ \eta_{1,2}  < 1.2, E_{T,1} > 35 \text{ GeV}, E_{T,2} > 32 \text{ GeV}$
Data	$\sigma = 5664 \pm 168 \text{ (stat.)} \pm 1270 \text{ (syst.)}$
PYTHIA	$\sigma = 5136 \pm 52 \text{ (stat.)}$
HERWIG	$\sigma = 5296 \pm 98 \text{ (stat.)}$
MC@NLO + JIMMY	$\sigma = 5421 \pm 105 \text{ (stat.)}$

Table 5.10: Total  $b\bar{b}$  di-jet production cross section.

$E_{T,2} > 32 \text{ GeV}$  and  $|\eta_{12}| < 1.2$ :

$$\sigma = 5664 \pm 168(\text{stat.}) \pm 1270(\text{syst.}) \text{ pb}$$

The systematic error is calculated adding in quadrature the error relative to those quantities that are uncorrelated from bin to bin, as tagging efficiency and  $b\bar{b}$  fraction. The uncertainties on the luminosity and scale factor are added as overall correction factors. The uncertainty on the jet corrections are calculated varying the jet  $E_T$  by  $\pm 1\sigma$ , as already described in section 3.6. The corresponding change in the number of events that are above the  $E_{T,jet1} > 35 \text{ GeV}$  and  $E_{T,jet2} > 32 \text{ GeV}$  thresholds is taken to measure the uncertainty on the total cross section.

## 5.5 Comparison to Monte Carlo

At the leading order the measurement is compared to PYTHIA and HERWIG. As already described, Tune A configuration is used for PYTHIA events.

In order to compare it to data, a HERWIG  $b\bar{b}$  sample has been generated using JIMMY, version 4.3 [85], a generator that produces multiple partons interactions: soft interaction from beam remnants after the hard scattering, in particular, while multiple hard scatterings are not considered.

The NLO prediction is obtained using the MC@NLO generator [86]. MC@NLO  $b\bar{b}$  events are generated with parton  $p_T$  above 10 GeV/ $c$  and pseudo-rapidity  $|\eta| < 1.75$ . CTEQ6M PDF functions are used and the renormalization and factorization scales are chosen so that  $\mu_R = \mu_F = \sqrt{p_T^2 + m^2}$ . The NLO events are combined to HERWIG for the parton shower process and to JIMMY for the underlying event.

Both HERWIG and MC@NLO samples are generated running JIMMY in the default configuration. The minimum transfer momentum of secondary scatters,  $PTMIN$ , is set to 2.5 GeV/ $c^2$ . The inverse proton radius squared,  $JMRAD$  is chosen as 0.71 GeV $^{-2}$ , for both the proton and the anti-proton. The probability of soft underlying event,  $PRSOE$ , is 1.0.

The generated events are passed through the full detector simulation and the same reconstruction code used for data events. The hadronic jet cross section is calculated and it is compared to the unfolded data measurement.

The three Monte Carlo predictions roughly show the same agreement with the data, for what concerns the total cross section or the differential cross section as a function of the leading jet  $E_T$ . Figures 5.10 and 5.11 represent the data to Monte Carlo ratio as function of the leading jet  $E_T$ , for the three predictions. The gray band represents the experimental systematic uncertainty only: no systematic error is considered for the theoretical prediction.

The measurement is  $\approx 10\%$ ,  $\approx 20\%$  higher than the PYTHIA and HERWIG predictions. The agreement to MC@NLO appears to be better. However, due to the large systematic uncertainty on data it is not possible to state a real difference between the predictions at different orders in perturbative QCD.

While there's no big difference in the  $E_T$  distribution between the different Monte Carlo models, both LO Monte Carlo predictions fail to show a good agreement with data in the  $\Delta\phi$  distribution, in figure 5.9, especially in the low  $\Delta\phi$  region. The MC@NLO + JIMMY prediction provides a much better description.

Data to Monte Carlo ratios for the three samples are reported in figure 5.12. It is evident that while PYTHIA is far off the systematic band in the low  $\Delta\phi$  region, HERWIG + JIMMY gives a better prediction and MC@NLO + JIMMY is compatible with data inside the systematic errors.

The difference between the three predictions is remarkable. However it cannot be

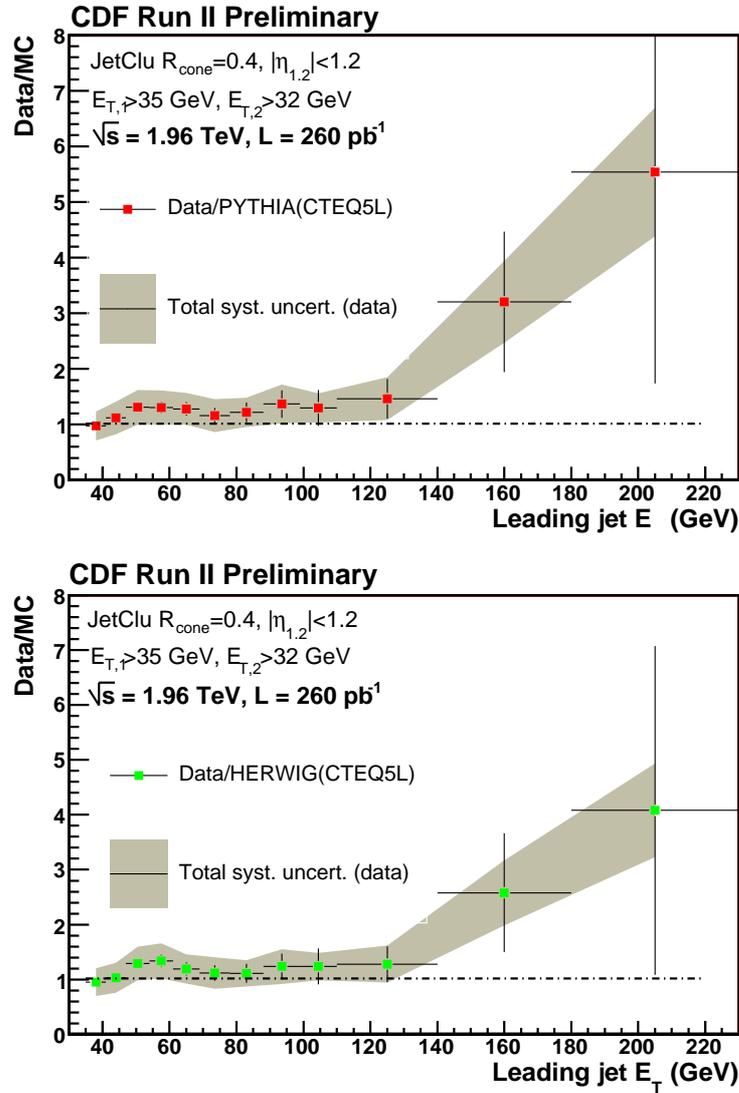


Figure 5.10: Data/Monte Carlo differential cross section ratio as a function of the leading jet  $E_T$ . The gray band represents the experimental systematic uncertainty. PYTHIA is shown in the top plot. HERWIG Monte Carlo is shown in the center bottom plot.

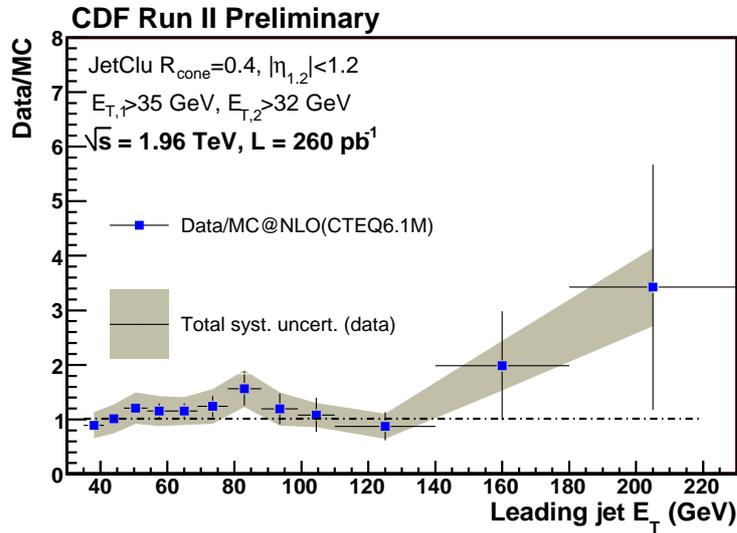


Figure 5.11: Data/MC@NLO differential cross section ratio as a function of the leading jet  $E_T$ . The gray band represents the experimental systematic uncertainty.

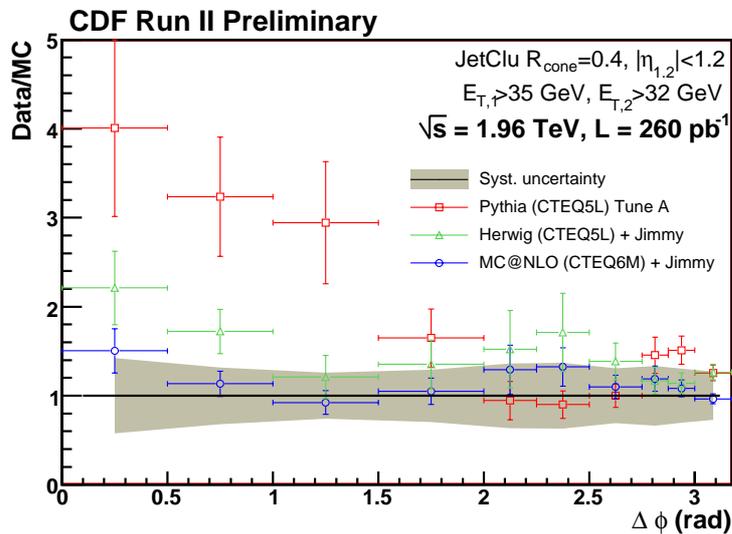


Figure 5.12: Data/Monte Carlo ratio as a function of di-jet azimuthal angle: PYTHIA (red) and HERWIG + JIMMY (green) and MC@NLO + JIMMY (blue).

totally addressed to the contribution of higher order terms in the perturbative QCD calculation: in fact the contribution due to LO and NLO terms is expected to be the same, when two  $b$  quarks in the central region are required [87]. This behavior is shown in figure 5.13. It represents the  $\Delta\phi$  distribution between the two  $b$  quarks, measured at generator level, i.e. before applying the detector simulation and the event data reconstruction. The  $\Delta\phi$  predictions by MC@NLO and HERWIG become compatible when tighter cuts on the pseudo-rapidity of the two quarks are applied: if the two  $b$  quarks are required to have  $|\eta| < 1.5$ , the distributions are very different at LO and NLO. When the cut is tightened to  $|\eta| < 1$  the LO and NLO predictions almost overlap.

The cut applied in the analysis is  $|\eta| < 1.2$ , which corresponds to an intermediate difference between the two distributions, as the picture on the bottom of figure 5.13 shows. For comparison, the PYTHIA prediction is also shown in red.

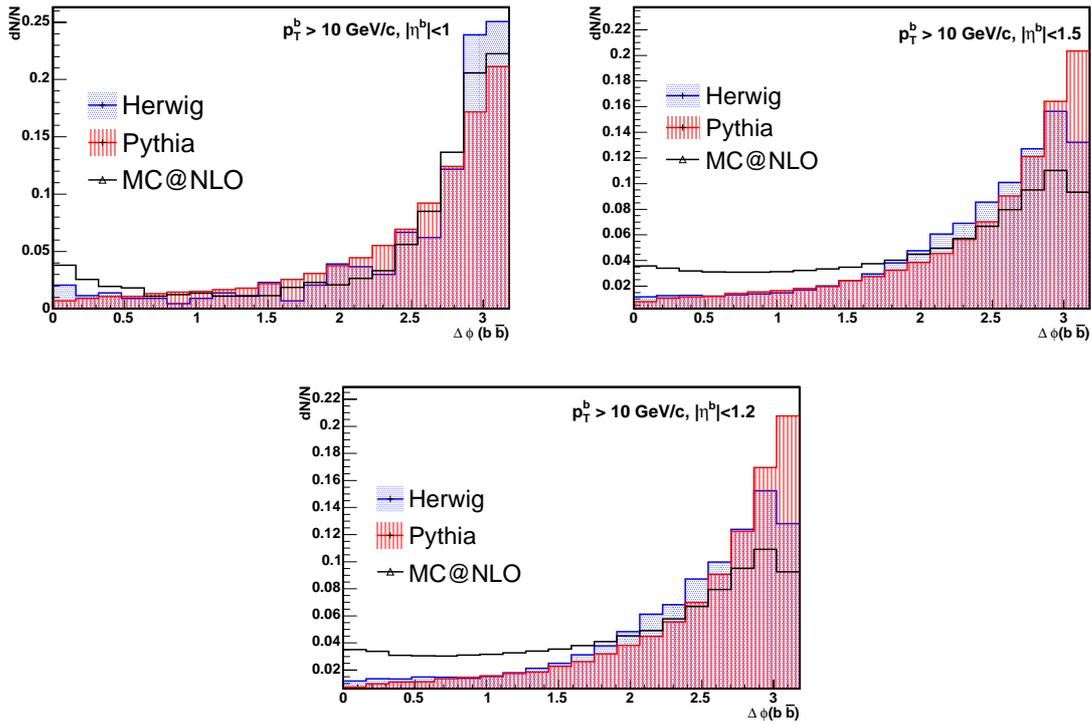


Figure 5.13:  $\Delta\phi$  distribution for  $b\bar{b}$  quarks in the region  $p_T^b > 10$  GeV,  $|\eta| < 1$  (top),  $|\eta| < 1.5$  (center) and  $|\eta| < 1.2$  (bottom).  $|\eta| < 1.2$  is the pseudo-rapidity cut applied in the analysis. All the distributions are normalized to unit area.

The reason of the different agreement of data to LO and NLO prediction has to be

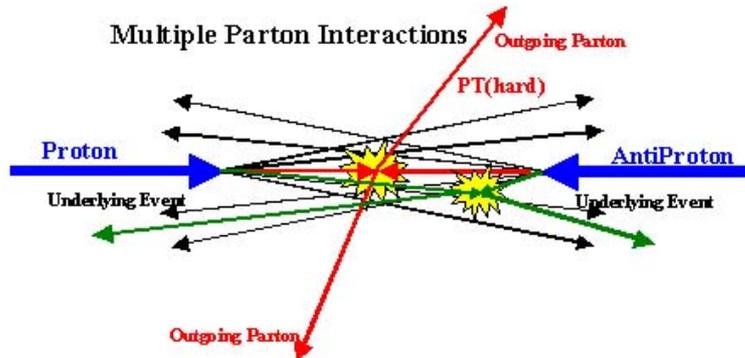


Figure 5.14: A representation of the ‘hard’ 2 to 2 parton scattering and the underlying event in hadronic collisions.

attributed to two different sources: LO versus NLO calculation and non perturbative effects. Next section presents an additional study performed to further investigate the role played by multiple partons interaction. The predictions of the same order in perturbation theory are compared with and without the addition of the soft underlying event to the hard scattering.

### 5.5.1 The importance of the underlying event

The underlying event is an unavoidable background to most collider observables and a good understanding of it will lead to more precise measurements at the Tevatron and the LHC. At the Tevatron both the inclusive jet cross section and the  $b$  jet cross section are sensitive to the underlying event. This measurements show strong sensitivity, in particular for what concerns the di-jet  $\Delta\phi$  distribution.

The ‘hard scattering’ component of the event consists of particles that result from the hadronization of the two outgoing partons plus the particles that arise from initial and final state radiation. The ‘underlying event’ consists of particles that arise from beam-beam remnants and possible multiple parton interactions, as sketched in figure 5.14. Of course, in a given event it is not possible to uniquely determine the origin of the outgoing particles and any observable will have contributions from both the underlying event and the hard scattering.

The underlying event is in a non-perturbative QCD region. As in the description of the hadronization process, Monte Carlo programs include models to simulate this soft component. These models are parametrized to describe experimental results. In the special parametrization of the underlying event, used with PYTHIA, Tune A, the parameters controlling the multiple partons interactions and initial state radiation have been adjusted to reproduce the energy density transverse to the leading jet

observed in the CDF Run I data [71].

To better study the contribution due to multiple interactions, four different measurement of the di-jet  $\Delta\phi$  distribution are performed: two at LO, using HERWIG alone and HERWIG+JIMMY; two at NLO, using MC@NLO and MC@NLO + JIMMY. The distributions are compared to data events in figure 5.15. Figure 5.16 shows the ratios data to Monte Carlo for each of the four predictions. It is clear that HERWIG prediction is strongly enhanced by the addition of the underlying event using JIMMY.

It is interesting that the NLO distribution also shows a rather large enhancement due to JIMMY, the contribution of the underlying event is to be considered necessary to insure a good agreement at low  $\Delta\phi$ , both at LO and NLO order.

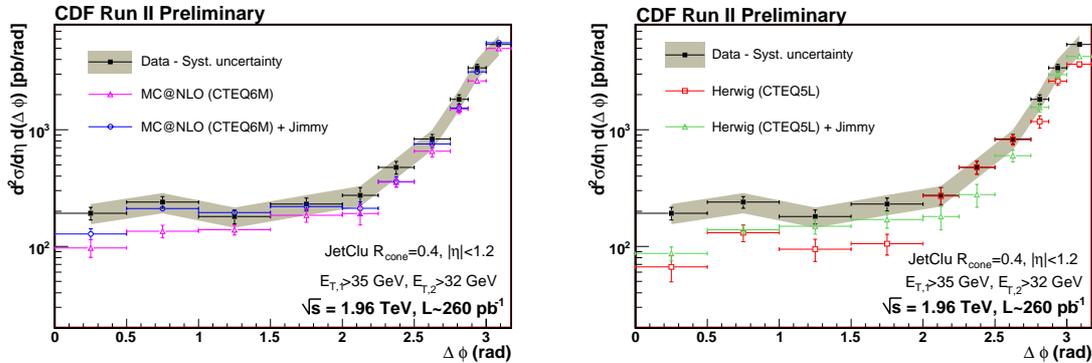


Figure 5.15: Cross section as a function of di-jet azimuthal angle: on the left, MC@NLO and MC@NLO + JIMMY. On the right: HERWIG and HERWIG + JIMMY.

## 5.6 Conclusions

A study on the  $b\bar{b}$  di-jet production in  $p\bar{p}$  collisions at 1.96 TeV has been presented here.

The total cross section has been measured for di-jet events in the phase space region defined by  $|\eta_{jet1,jet2}| < 1.2$ ,  $E_{T,jet1} > 35$  GeV and  $E_{T,jet2} > 32$  GeV. The differential cross sections as a function of the leading jet transverse energy and of the di-jet invariant mass have also been measured together with the  $\Delta\phi$  correlation between the two jets.

The analysis is based on the good tracking capabilities of the CDF detector. Online, events have been selected using the SVT trigger and requiring two large impact parameter tracks.

Offline,  $b$ -jets have been identified using a tagging algorithm based on the reconstruction of secondary vertices originating from  $B$  hadron decays.

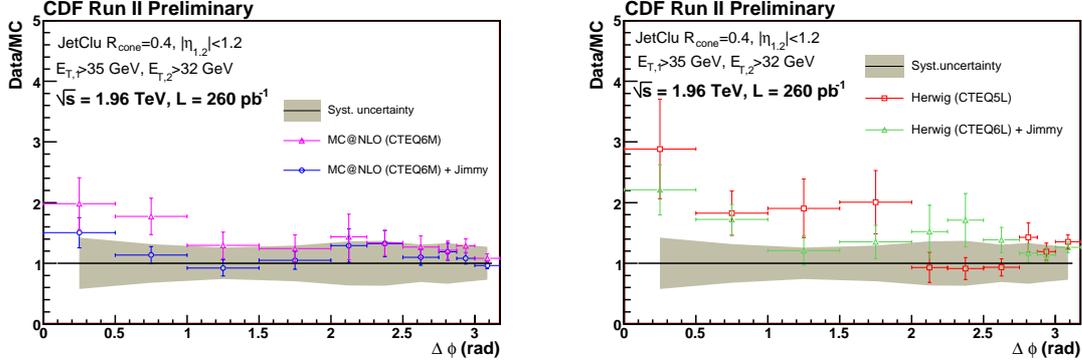


Figure 5.16: Data/MC ratio as a function of di-jet azimuthal angle: on the left, MC@NLO and MC@NLO + JIMMY. On the right: HERWIG + JIMMY (pink) and HERWIG (green).

The measurement covers jet energies going from 35 GeV up to 250 GeV, corresponding to a cross section range of three orders of magnitude. It is the first time a measurement of the  $b\bar{b}$  correlation is performed at high  $p_T$ . Previous Run I measurements and similar Run II analyses had studied exclusive  $B$  decays and were limited to energies below 100 GeV.

This analysis makes use of only  $260 \text{ pb}^{-1}$  of data corresponding to the first two years of Run II data taking, 2002-2004. Nevertheless, the use of the SVT trigger assures that the final  $b\bar{b}$  sample has a large statistical size if compared to inclusive jet samples selected using calorimetric triggers.

The data are corrected to the particle level and compared to Monte Carlo predictions at LO and NLO. An overall agreement is found between the data and the NLO prediction, taking into account the experimental systematic uncertainty.

In particular, the measurement of the  $b\bar{b}$   $\Delta\phi$  correlation confirms that the  $b\bar{b}$  jets are mainly produced via LO mechanisms as the flavour creation process, which corresponds to two body final states ( $b\bar{b}$ ) for which momentum conservation requires a back-to-back configuration. Those events are characterized by a large  $\Delta\phi$  between the two  $b$ -jets.

It is interesting to note, however, that there is a large number of events for which  $\Delta\phi$  is small ( $\Delta\phi < 1.5$ ). When higher order QCD processes are considered, in fact, the presence of additional light quarks and gluons in the final state allows the  $\Delta\phi$  distribution to spread out. The small  $b\bar{b}$  opening angle region is, therefore, of interest because in this region, the higher-order  $b\bar{b}$  production mechanisms become more important. The data presented here cannot be described solely by flavor creation. A significant fraction of the  $b\bar{b}$  pairs are produced with  $\Delta\phi < 1.5$ , in agreement with the

conclusions from previous analyses that flavour excitation and gluon splitting play a significant role in the  $b\bar{b}$  production at the Tevatron.

This conclusion is supported by the fact that the data shows a large excess at small opening angles, with respect to LO prediction (both PYTHIA and HERWIG), while it agrees within the uncertainties with the MC@NLO prediction.

The reason why PYTHIA gives a lower prediction probably suggests that the contribution due to gluon splitting or flavor excitation processes is not correctly described. Other measurements at low  $p_T$ , as a recent chrmed boson angular correlation study, pointed out the same problem [88].

Besides, a detailed comparison between the data and the different Monte Carlo predictions has also shown the importance of a correct description of the underlying event, that represents an important issue for any measurement performed at a hadron collider. The  $\Delta\phi$  distribution predicted by HERWIG, for which additional multiple interactions are simulated using JIMMY, shows a much better agreement to data, compared to the case they are not.



# Appendix A

## Simulation of the SVT beamline

As mentioned in section 3.5 ‘realistic’ Monte Carlo events are simulated. It means that the detector simulation is tuned to reproduce events from real data run using calibration tables from the database. The beam positions, in particular, are simulated using the information from 732 runs, that correspond to a total integrated luminosity of  $250 \text{ pb}^{-1}$  approximately. Unfortunately a large percentage of the simulated samples turned out in having wrong beam-line positions. It has already been mentioned, that the alignment of the SVT to the beam position is crucial to a correct measurement of the track impact parameter. A way to check everything is working correctly is to check the SVT tracks impact parameter distribution as a function of the run number, to verify its stability as in figure A.1: the left plot points out a problem for those run having a value far off the average. Figure A.2 shows the presence of a pattern

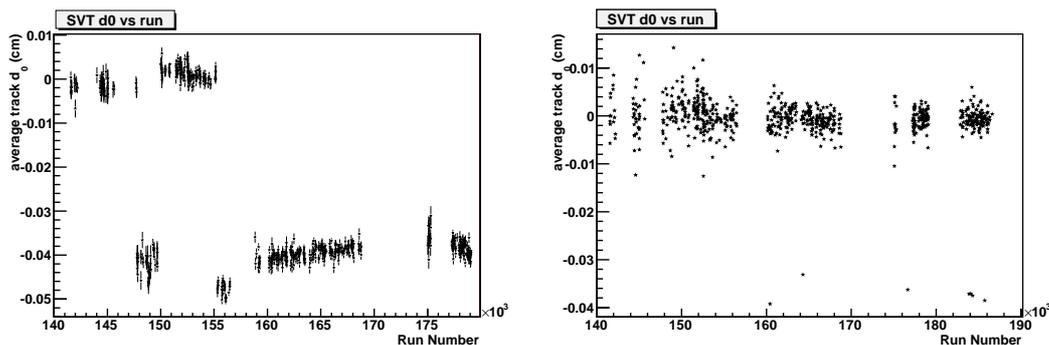


Figure A.1: SVT track impact parameter averaged over run number as originally simulated using the default testSVTbeam file (left) and after correction (right)

indicating an incorrect simulation of the beamline. Figure A.1 shows the average  $d_0$  as a function of run number after the correct beamline simulation is applied.

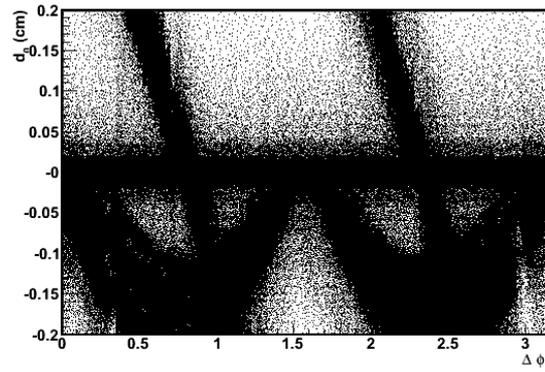


Figure A.2: SVT track impact parameter as a function of the SVT track  $\phi_0$ . The sinusoidal patterns point out that a wrong correction for the SVT beamline is applied and that the beam line position changed with time.

# Appendix B

## The SVT-tagging scale factor

The following study shows that the scale factor measured to correct the SVT-tagging efficiency to data can be applied to this analysis. The procedure is the one described in [74]. The SVT-tagging efficiency is calculated using muon data and Monte Carlo samples, which have an enriched heavy flavor content. Nevertheless the sample is not pure and only about one third of the jets including a muon are expected to be  $b$  flavor jets, mainly because of fake muon identification.

To enhance the fraction of  $b$ -jet, the additional requirement of having the away jet tagged is applied: jets including a muon are identified ('muon jet') and are associated to 'away' jets (applying a cut on the  $\Delta\phi$  between the two jets). It should be noted, however, that even in this case the fraction of heavy-flavor muon jets is not 100% as the away jet could be, for example, a mis-tagged light jet or a  $c$ -jet.

The most straightforward way to calculate the tagging efficiency consists in the *single-side tagging* method, counting the number of muon  $b$ -jets in the case the away jet is tagged and in the case it is not. Under the assumption that the contamination from  $c$ -jets is negligible this method allows the efficiency to be calculated directly. In this case the efficiency is defined as:

$$\epsilon = \frac{N_{a+}^{\mu+} \cdot f_{\mu+}^b}{N_{a+} \cdot f_{\mu}^b}$$

where:

- $N_{a+}^{\mu+}$  is the number of events with a tagged muon jet when the away jet is tagged
- $f_{\mu+}^b$  is the fraction of  $b$  jets among tagged muon jets
- $N_{a+}$  is the number of events with a tagged away jet
- $f_{\mu}^b$  is the fraction of  $b$  jets among all muon jets associated to a tagged away jet

The transverse momentum of the muon relative to the jet axis,  $p_t^{rel}$  is used as a discriminant parameter to measure the  $b$  jet fraction of tagged and untagged muon jets. Being an invariant with respect to a boost along the jet axis, the  $p_T^{rel}$  is a measurement of the transverse muon momentum. In the  $B$ -hadron rest frame, the muon coming from the  $B$  decay has a large momentum due to the mass difference between the  $B$ -hadron and its decay products. This momentum is different for muons originated by a  $D$  hadron decay or a light hadron decay, because the mass difference between the decaying particle and decay products is smaller. The shape of  $p_T^{rel}$  distribution can, then, be used to disentangle different flavour contributions to the muon jet sample.

A muon data samples based on a trigger selecting muons with a  $p_T > 8 GeV$  is used. MonteCarlo samples are HERWIG di-jet samples including a jet with  $E_T > 20 GeV$ , and filtered at generator level requiring at least a muon with  $E_t > 9 GeV$ .

A preliminary selection on muon tracks is performed before associating them to a jet. Good muons are defined as:

- Muon  $P_t > 9 GeV/c$
- CMU stub  $|dx| < 3 cm$  and CMP  $|dx| < 5 cm$
- Track  $|z_0| < 60 cm, |z_0 - z_{vtx}| < 5 cm$
- Isolation (cone 0.4)  $> 0.1$
- Muon track hits all SVX layers

If more than a muon passes this selection, the one with largest  $p_T$  is considered. It good muon is then associated to a jet with  $E_t > 20 GeV$  and  $|\eta| < 1.2$ . The muon jet  $E_t$  and momentum are corrected for the presence of the muon, according to the equation:

$$E_{corr} = E_{raw} \cdot \frac{E_{raw} - 2GeV}{E_{raw}} + p_{t,\mu}$$

$$\vec{P}_{corr} = \vec{P} + \vec{p}_\mu \left( 1 - \frac{2GeV}{p_\mu} \right) E_{raw}$$

The correction factor of  $2 GeV$  has been measured in studies of muon identification [84].

Once the muon jet has been identified, an ‘away’ jet is associated to it if the  $\Delta\phi$  between the two is larger than 2 radian. The ‘away’ jet is also required to have  $E_t > 20 GeV$ ,  $|\eta| < 2$ . On data an additional cut on the secondary vertex invariant mass of the ‘away’ jet (greater than  $1.5 GeV/c^2$ ) is applied to further enhance the heavy flavour content.

Figure B.1 shows the jet  $E_T$  distribution for muon jets and away jets in data and Monte Carlo events.

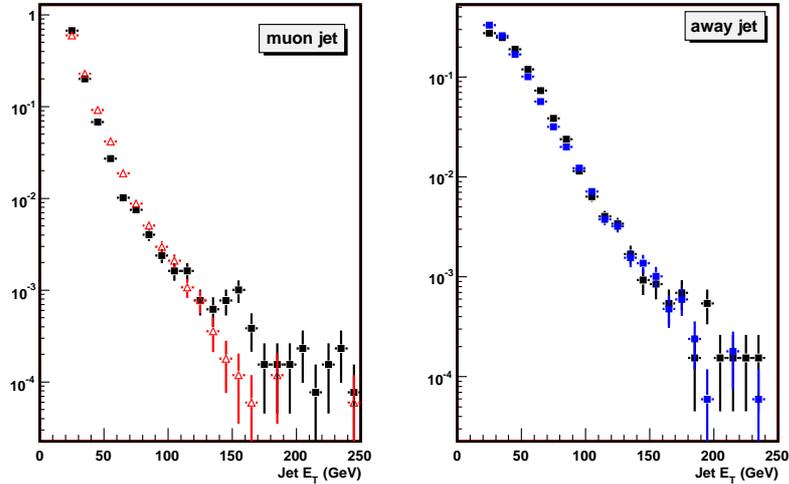


Figure B.1: **left**: Data (black) vs MonteCarlo (red) muon jet  $E_T$  comparison. **right**: Data (black) vs MonteCarlo (blue) away jet  $E_T$  comparison.

Figure B.2, B.3 and B.4 show a comparison between some tagging algorithm related variables in data and Monte Carlo: the number of good tracks, number of pass1 and pass2 tracks, the number of tracks attached to the secondary vertex and their invariant mass.

The same quantities are also shown for SVT-tagged muon jets in figures B.5 and B.6.

The muon  $p_T$  relative to the corrected jet axis is reconstructed, and template distributions for  $b$ ,  $c$  and light jets are generated from Monte Carlo events, in order to fit the fraction of muon  $b$ -jets directly from data.

Special care is necessary in choosing transverse momentum templates: the presence of the SVT track in the jet could in fact bias the shape of the templates. This effect is in fact well known in the case of reconstructing the secondary vertex invariant mass: jets including a SVT track peak at slightly larger mass as shown in chapter 4. Figure B.7 (left) shows a comparison of the relative transverse momentum of tagged muon  $b$ -jet and SVT-tagged muon  $b$ -jets: as the difference is negligible, simple tagged  $b$ -jets templates are used instead of SVT-tagged jet in order to have higher statistic templates, and reduce the uncertainty on the fit.

Light jets and  $c$ -jets  $p_T^{rel}$  distribution are very similar as shown in figure B.7 on the right. To separate the contribution of the two is very difficult and beyond the purpose of this study. Therefore a simplified two templates fit is performed isolating

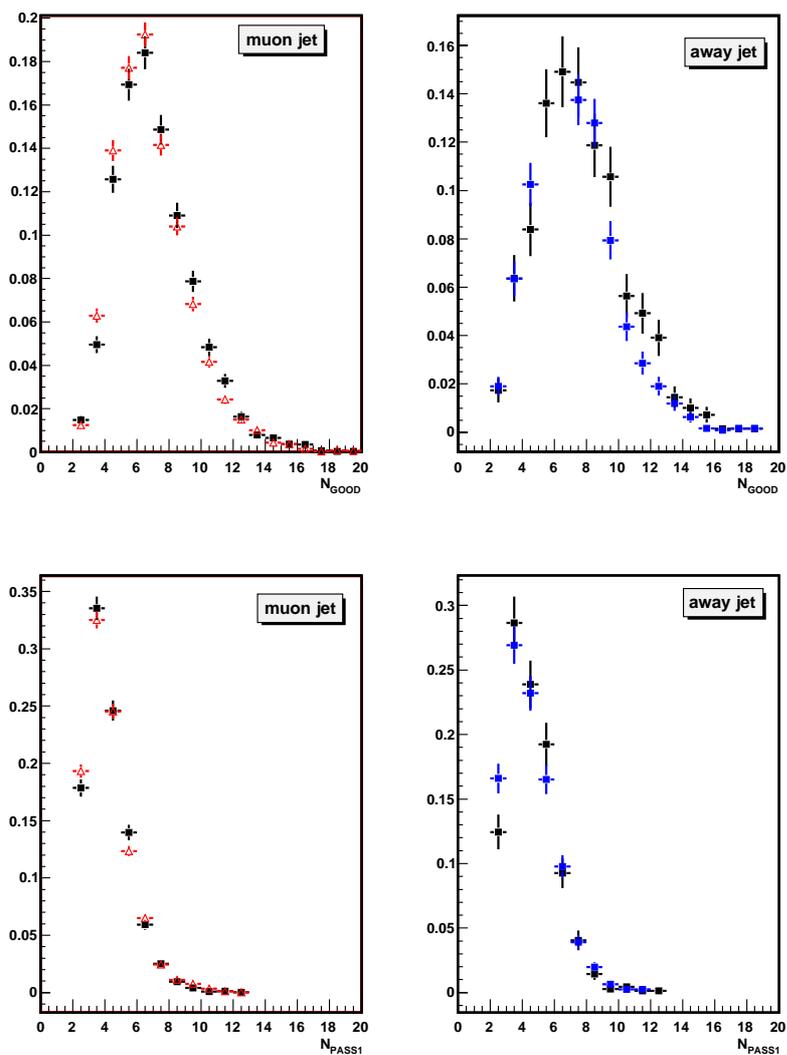


Figure B.2: **left**: Data (black) vs MonteCarlo (red) muon jet  $N_{GOOD}$  (top) and  $N_{PASS1}$  (bottom) comparison. **right**: Data (black) vs MonteCarlo (blue) away jet  $N_{GOOD}$  (top) and  $N_{PASS1}$  (bottom) comparison.

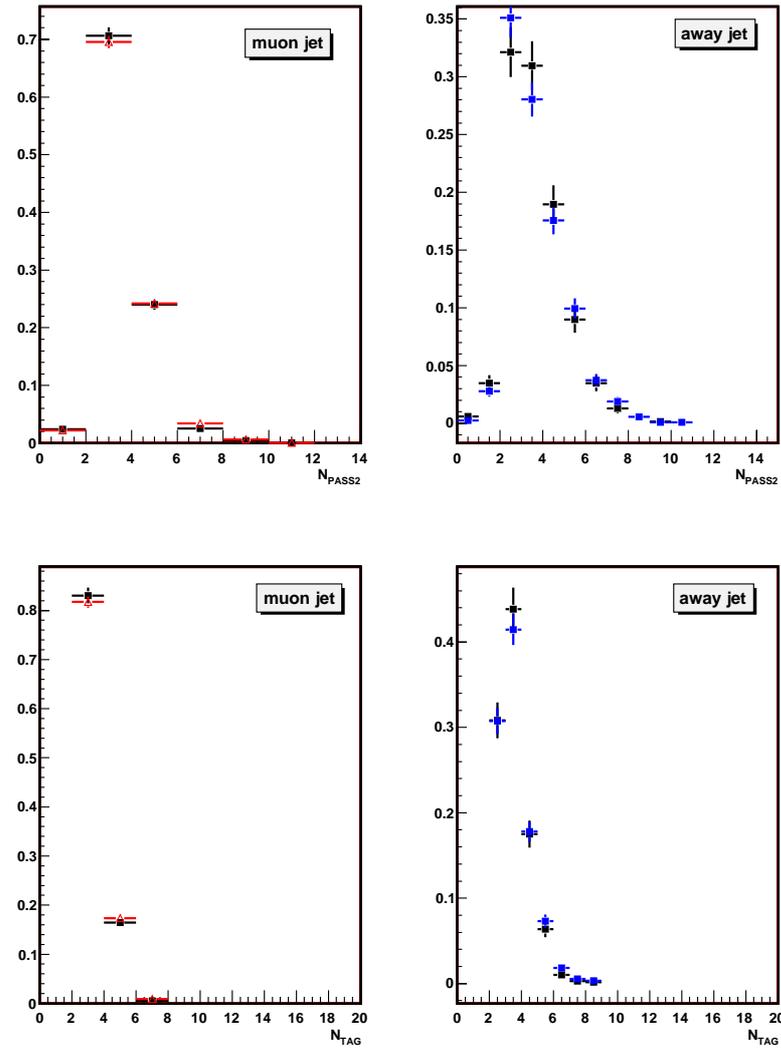


Figure B.3: **left**: Data (black) vs MonteCarlo (red) muon jet  $N_{PASS2}$  (top) and  $N_{TAG}$  (bottom) comparison. **right**: Data (black) vs MonteCarlo (blue) away jet  $N_{PASS2}$  (top) and  $N_{TAG}$  (bottom) comparison.

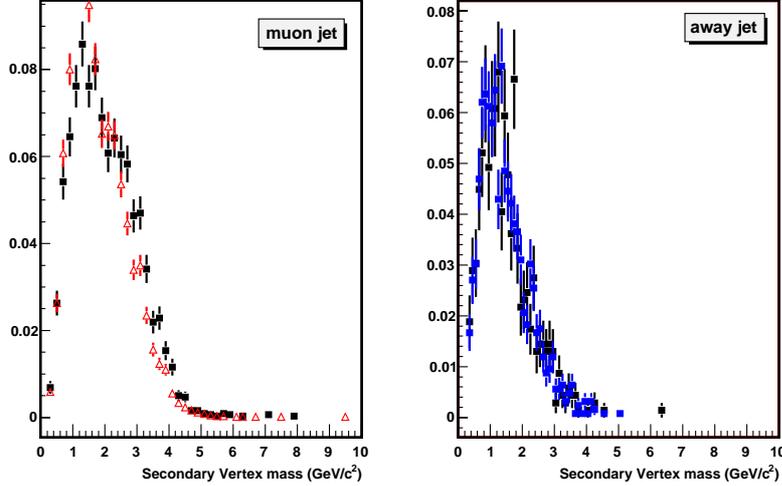


Figure B.4: **left**: Data (black) vs MonteCarlo (red) muon jet secondary vertex invariant mass comparison. **right**: Data (black) vs MonteCarlo (blue) away jet secondary vertex invariant mass comparison.

$b$ -jets against non- $b$  templates, the last being chosen in two different ways:

- $c$  templates obtained from muon  $c$ -jets in muon MonteCarlo samples;
- light muon jets templates collected from generic Pythia di-jet samples.

A comparison between the two cases is shown in figure B.7 on the left side.

Two fits are performed on muon jets in data events, using the two different templates and the  $b$  jet fractions before and after tagging are shown in figure B.8: on the top plot the fit results relative to  $c$ -jet template and on the bottom plot the result for the light jet template.

The results obtained in the two cases very well agree with each other:

$$\begin{aligned}
 f_{tag}^c &= 0.92 \pm 0.03 & f_{tag}^{light} &= 0.91 \pm 0.03 \\
 f_{untag}^c &= 0.65 \pm 0.02 & f_{untag}^{light} &= 0.63 \pm 0.02
 \end{aligned}$$

The  $b$ -jet fraction is close to one for tagged muon jets and above 60% for untagged muon jets. A large purity is expected since muon jets are specifically selected to have a high  $b$  purity, further enhanced by associating them to  $b$ -tagged ‘away’ jets.

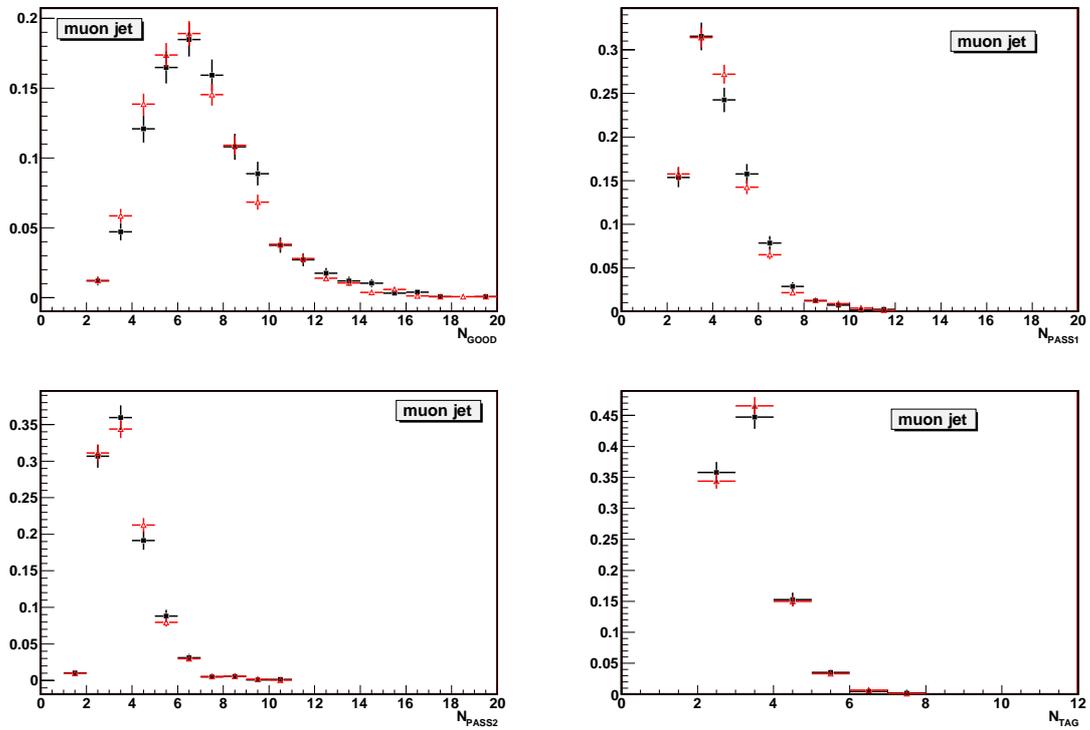


Figure B.5: **left:** Data (black) vs MonteCarlo (red) muon jet  $N_{GOOD}$  (top) and  $N_{PASS1}$  (bottom) comparison. **right:**  $N_{PASS2}$  (top) and  $N_{TAG}$  (bottom).

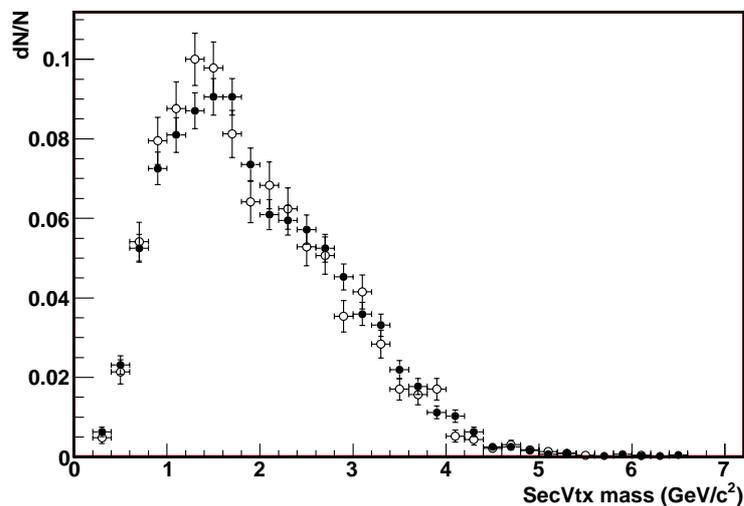


Figure B.6: Data (black) vs MonteCarlo (red) muon jet secondary vertex invariant mass comparison.

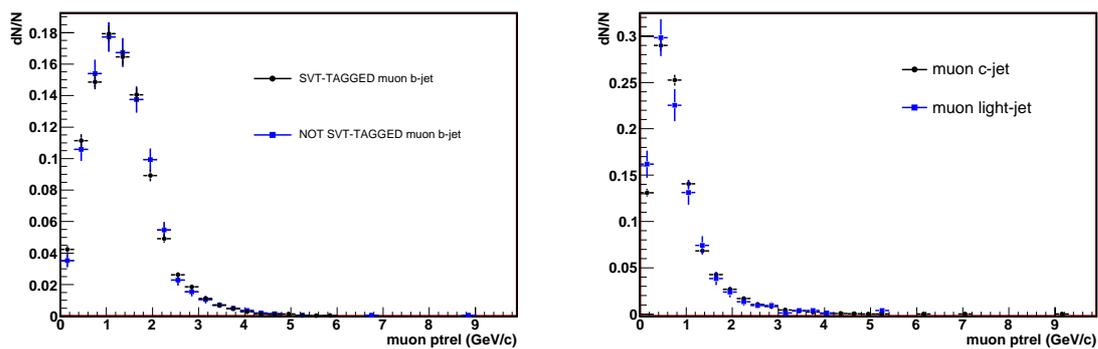


Figure B.7: **left**: Light jet templates compared to c jet templates. **right** Template muon relative  $p_t$  in b jet (black) and in SVT-tagged b jet (blue).

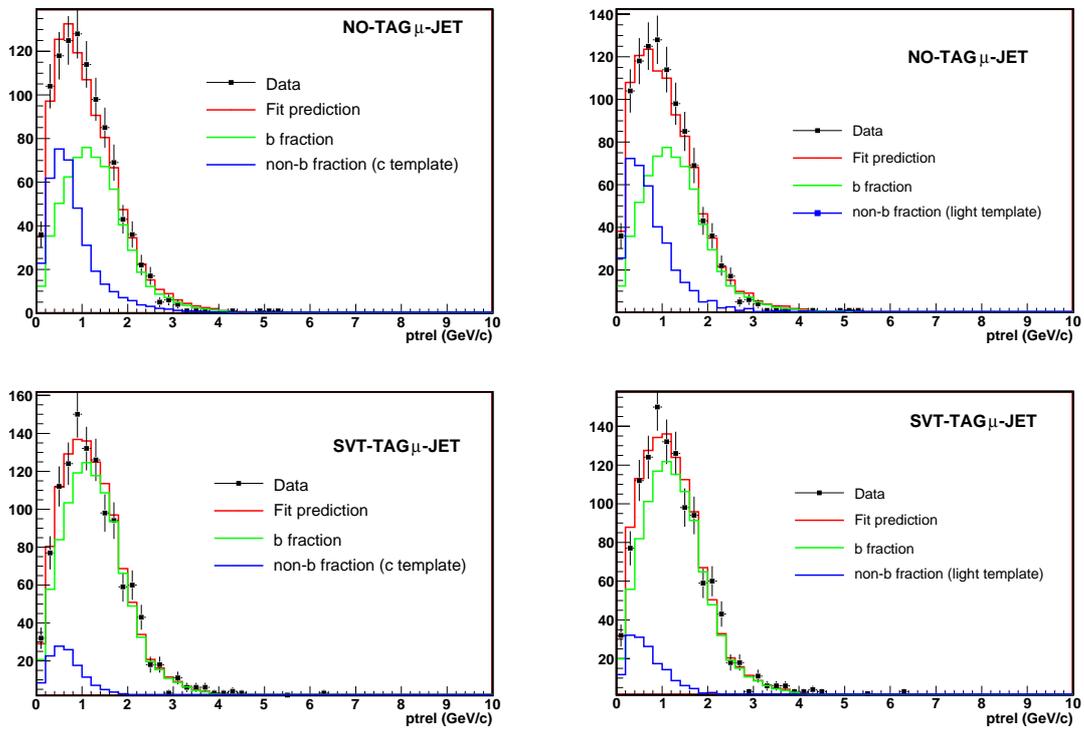


Figure B.8: **top:** relative  $p_t$  fit for not tagged muon jets in data, using c templates (left) and light templates (right). **bottom:** same fit for SVT tagged muon jets in data.

Using the values obtained for the  $b$ -jet fraction (in particular choosing the  $c$ - $jt$  templates), the efficiency is measured in data and MonteCarlo and the result is summarized in figure B.9 and in table B.1.

	$N(\mu^+, away^+)$	$N(away^+)$	$\epsilon$
data	1191	7485	$0.22 \pm 0.01$
MC	1791	7952	$0.225 \pm 0.007$

Table B.1: Results from scale factor measurement.

The scale factor is calculated as

$$SF = 0.99 \pm 0.05$$

This value is compatible with the official one,

$$SF = 1.029 \pm 0.009(stat) \pm 0.034(syst)$$

which is the one that will be used in the cross section measurement.

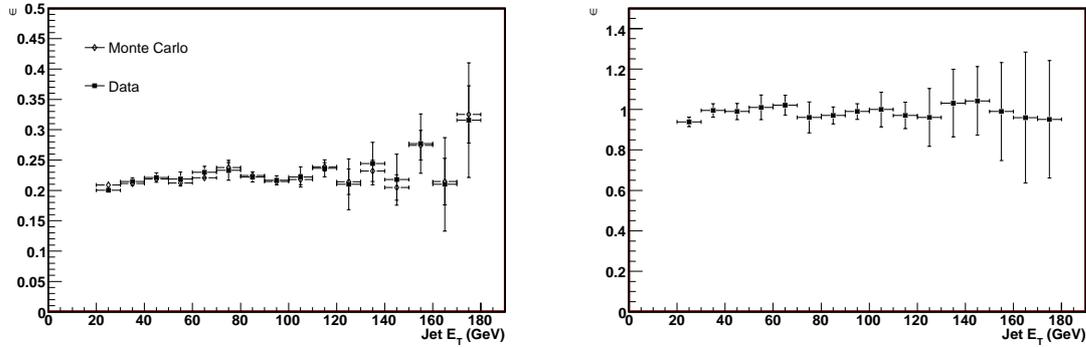


Figure B.9: (left) SVT tagging efficiency in data (full squares) and Monte Carlo (empty diamonds) and scale factor (right) as a function of jet  $E_T$ .

# Appendix C

## Unfolding factors and jet energy corrections

It is interesting to spend some time on the understanding of the relationship between unfolding factor values and the choice of the average jet energy correction. In fact, it might seem surprising that even after correction and energy scale adjustments, there is still the need of a large  $\approx 20\%$  factor to match the measured cross section to the ‘expected’ one, i.e. the particle level cross section. A short study has been performed to prove that the distinction between average jet corrections and unfolding factors is in fact purely practical. The unfolding factors obtained in two different cases are compared:

- Jet energy is corrected applying standard corrections, measured for generic jets.
- Jet energy is corrected for standard generic corrections, plus the additional specific  $b$  jet correction described in section 3.6, which are of the order of 5%.

The two cases will lead to two different sets of factors, as shown in figure C.1 (left) for the leading jet  $E_T$ . It is a general behaviour that applying additional corrections reduces the unfolding factors to be applied to the raw cross section. The lower the correction applied on average, the larger the unfolding factors needed. However the procedure is fully coherent. Figure C.1 compares, on the right, the differential cross sections obtained applying the two different sets of corrections and unfolded using the corresponding factors. The final result is the same in the two cases.

This fact proves that there’s always a residual effect which cannot be taken into account using average energy corrections only. And that different average corrections are compensated by different unfolding factors.

This measurement, however, chooses to apply the additional correction found for tagged jets in order to reduce the size of the unfolding factors.

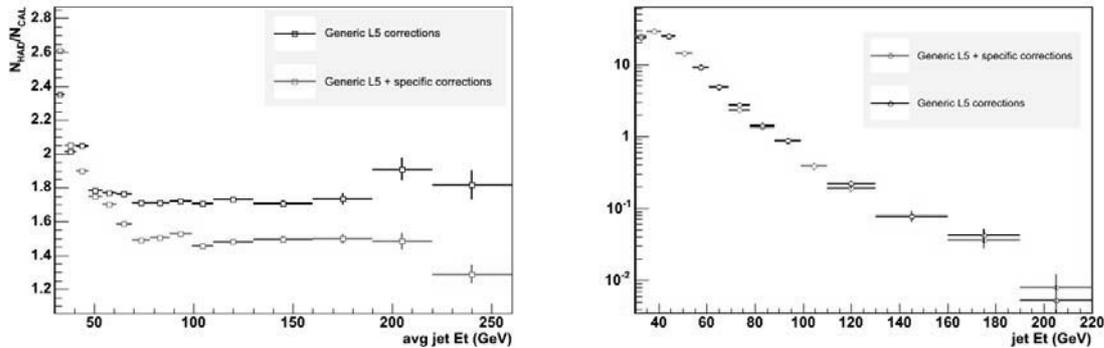


Figure C.1:  $b\bar{b}$  cross section for jets corrected using L5 generic correction (black) and for jets corrected for the additional specific function resulted from the fit (blue) (left). On the right, the corresponding unfolding factors.

There is also a second problem that should be considered when dealing with  $b$ -flavored jets. As already mentioned in section 3.6 about 23% of  $B$ -hadrons decay semileptonically, so there is an underestimation of the parton energy due to lost neutrinos. Of course, generic  $b$ -jet  $E_T$  corrections cannot account for this lost energy and larger unfolding factors are needed to correct the measured cross section to the particle level one. A brief study has been performed to verify this hypothesis: using a PYTHIA  $b\bar{b}$   $\hat{p}_T > 40$  GeV/ $c$  sample,  $b$ -jets in which the  $B$ -hadron decays semileptonically have been identified and their energy has been corrected using generic jet  $E_T$  corrections. The  $b\bar{b}$  di-jet cross section has been measured together with the corresponding unfolding factors. Figure C.2 shows a comparison between the unfolding factors for generic  $b$ -jets and for  $b$ -jets associated to a semileptonic decay of the  $B$ . When the same generic  $E_T$  corrections are applied the unfolding factors, for the semileptonic case, are significantly larger than in the generic case.

### Unfolding factors in HERWIG

The unfolding factors are calculated using the same procedure in HERWIG. Figure C.3 represents the unfolding factors as a function of the leading jet  $E_T$ , the di-jet invariant mass and the di-jet  $\Delta\phi$ .

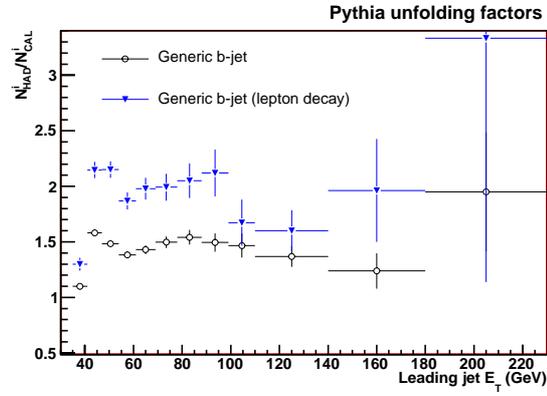


Figure C.2: Unfolding factors for jet energy corrected using L5 generic correction.  $b$ -jets are in black and semileptonic  $b$ -jets are in (blue).

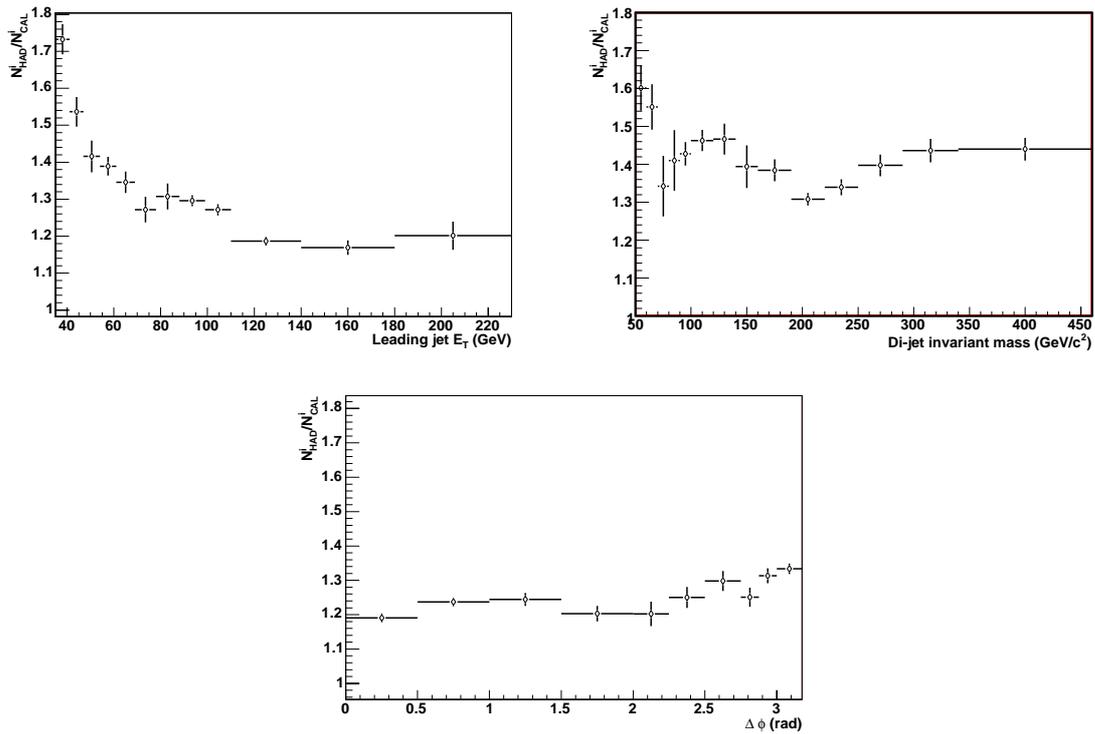


Figure C.3: Unfolding factors calculated using HERWIG as a function of the leading jet  $E_T$  (left),  $M_{jj}$  (right) and  $\Delta\phi_{jj}$  (bottom).



# Bibliography

- [1] H. Fritzsch, M. Gell-Mann and H. Leuthwyler, Phys Lett. **B47**, 365 (1973).
- [2] W.-M. Yao *et al.*, J. Phys. G **33**, 1 (2006).
- [3] J.C. Collins, D.E. Soper and G. Sterman, Nucl. Phys. **B263**, 37 (1986).
- [4] J.C. Collins, arXiv:hep-ph/9510276 (2006).
- [5] G. Altarelli and G. Parisi, Nucl. Phys. **B126** 298 (1977)
- [6] H.L. Lai *et al.*, Phys. Rev **D51**, 4763 (1995).
- [7] A.D. Martin *et al.*, Eur. Phys. J. **C4**, 463 (1998).
- [8] H. Baer *et al.*, hep-ph/0403045v2.
- [9] Pythia webpage, <http://www.thep.lu.se/~torbjorn/Pythia.html>
- [10] Herwig webpage, <http://hepwww.rl.ac.uk/theory/seymour/herwig/>
- [11] R.D. Field and R. P. Feynman, Nucl. Phys. **B136**, 1 (1978).
- [12] T. Sjostrand and M. Bengtson, Comp. Phys. Commun. **46**, 43 (1987).
- [13] T. Sjostrand, CERN-TH.6488/92 (1992).
- [14] C. Peterson *et al.*, Phys. Rev. **D27** 105 (1983).
- [15] B. Anderson *et al.*, Phys. Rep. **97**, 31 (1983).
- [16] G. Baker *et al.*, Contributed paper for ICHEP 2002.
- [17] S. Frixione, B.R. Webber, hep-ph/0204244.
- [18] S. Frixione, P. Nason and B. Webber, hep-ph/030552v2.
- [19] P. Nason *et al.*, Nucl. Phys. **B327**, 49 (1989).

- [20] M. Mangano *et al.*, Nucl. Phys. **B373**, 295 (1992).
- [21] J.C. Collins, D.E. Soper and G. Sterman, Phys. Lett. **B112**, 445 (1982).
- [22] J. Kodaira and L. Trentadue, Phys. Lett. **B194**, 445 (1982).
- [23] C.Abajar *et al.*, Phys. Lett. B **256** (1991) 121.  
C.Abajar *et al.*, Phys. Lett. **B213** (1988) 405.
- [24] S. Abachi *et al.*, Phys. Rev. Lett. **74** (1995) 2632.  
S. Abachi *et al.*, Phys. Lett. **B370** (1996) 239.  
B. Abbott *et al.*, Phys. Rev. Lett. **85** (2000) 5068.
- [25] F. Abe *et al.*, Phys. Rev. Lett. **69** (1992) 3704.  
F. Abe *et al.*, Phys. Rev. Lett. **71** (1993) 500.  
F. Abe *et al.*, Phys. Rev. Lett. **71** (1993) 2396.  
F. Abe *et al.*, Phys. Rev. Lett. **69** (1992) 2537.
- [26] D. Acosta et al. *Measurement of the  $B^+$  total cross section and  $B^+$  differential cross section  $d\sigma/dp_T$  in  $p\bar{p}$  collision at  $\sqrt{s} = 1.8$  TeV.* Phys. Rev. D **65** (2002) 052005.  
F. Abe et al. *Measurement of the  $B$  meson differential cross section  $d\sigma/dp_T$  in  $p\bar{p}$  collision at  $\sqrt{s} = 1.8$  TeV.* Phys. Rev. Lett **75** (1995) 1451.
- [27] M. Cacciari and P. Nason, *Is there a significant excess in bottom hadroproduction at the Tevatron?* Phys. Rev. Lett. **89** (2002) 122003.
- [28] The CDF collaboration, Phys. Rev. **D71**:032001 (2005).
- [29] Monica D'Onofrio, Ph.D. thesis submitted at the University of Geneva, Switzerland, 2005.
- [30] F. Halzen, W.Y. Keung and D.M. Scott, *Production of Heavy quarks: a non-perturbative approach.* Phys. Rev. D **27** (1983) 1631.
- [31] K. Abe et al., *Improved measurement of the probability for gluon splitting into  $b\bar{b}$  in  $Z^0$  decays.* Phys. Lett. B **507** (2001) 61.

- [32] M.H. Seymour, *Heavy quark multiplicity in  $e^+e^-$  events*. Nucl. Phys. B 436 (1995) 163.  
S. Frixione, M. L. Mangano, P. Nason, G. Ridolfi, *Heavy-quark production* Adv. Ser. Direct. High Energy Phys. 15 (1998) 609.
- [33] R.D. Field, *The sources of Q quarks at the Tevatron and their correlations*. Phys. Rev. D 65 (2002) 094006.
- [34] P. Nason et al. *Bottom Production*. hep-ph/0003142.
- [35] E. Norrbin and T. Sjostrand, *Production and hadronization of heavy quarks*. Eur. Phys. J. C17 (2000) 137.
- [36] R. D. Field, *The sources of Q quarks at the Tevatron and their correlations*, Phys. Rev **D65** (2002) 094006.
- [37] C. Albaja et al. *Measurement of  $b\bar{b}$  correlations at the CERN  $p\bar{p}$  collider*. Z. Phys. C 61 (1994) 41.
- [38] B. Abbott et al. *The  $b\bar{b}$  production cross section and angular correlations in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.8$  TeV*. Phys. Lett. B 487 (2000) 264.
- [39] F. Abe et al. *Measurement of the  $b\bar{b}$  production correlations,  $B^0$  anti- $B^0$  mixing and limit on  $\epsilon_B$  in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.8$  TeV*. Phys. Rev D 55 (1997) 2546.
- [40] F. Abe et al. *Measurement of the correlated  $\mu - \bar{b}$  jet cross section in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.8$  TeV*. Phys. Rev D 53 (1996) 1051.
- [41] F. Abe et al. *Measurement of  $b\bar{b}$  rapidity correlations in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.8$  TeV*. Phys.Rev D 61 (2000) 032001.
- [42] M. Popovic *et al.*, "Fermilab Linac Injector, revisited", proceedings for the 16th Biennial Particle Accelerator Conference, Dallas, USA 1995, published by APS.
- [43] Fermilab Beam Division, Run II Handbook, <http://www-bd.fnal.gov/runII/index.html>
- [44] F. Abe *et al.* Nucl. Instr. Meth. **A271**, 357 (1988).
- [45] The CDF collaboration, FERMILAB-Pub-96/390-E (1996).
- [46] L. Balka *et al.*, Nucl. Instrum. Meth **A267**, 272 (1988).
- [47] S. Bertolucci *et al.*, Nucl. Instrum. Meth **A267**, 301 (1988).
- [48] A. Artikov *et al.*, Part. Nucl. Lett 114, 25, (2002).

- [49] A. Artikov *et al.*, Nucl. Instr. Meth. **A538**, 358 (2005).
- [50] Trigger and Dataset working Group, <http://www-cdf.fnal.gov/internal/upgrades/daq-trig/twg/tools/trigopts.ps>
- [51] K. Anikeev *et al.* CDF internal note 5051 (1999).
- [52] K. Maeshima *et al.* [http://www-cdfonline.fnal.gov/internal/mon/consumer/home/consumer\\_home](http://www-cdfonline.fnal.gov/internal/mon/consumer/home/consumer_home)
- [53] SVT TDR, CDF internal note 3108.
- [54] H. Wenzel *et al.*, <http://www-cdfonline.fnal.gov/internal/mon/consumer/framework/index.htm>
- [55] CDF Collaboration, Nucl. Instr. Meth **A411** 366, 2000.
- [56] C. Hays CDF internal note 6992
- [57] J. E. Huth *et al.*, FERMILAB-CONF-90-249-E, 1990.
- [58] R. Field. hep-ph/0108264.
- [59] S.D. Ellis and D. E. Soper, Phys. Rev. **D48**, 3160 (1993).
- [60] Catani *et al.*, Nucl. Phys. **B406**, 187 (1993).
- [61] S. D. Ellis *et al.*, hep-ph/0111434.
- [62] R. Lefevre *et al.*, Phys. Rev. **D75**, 092006 (2007).
- [63] CDF public page, <http://www-cdf.fnal.gov/physics/new/qcd/QCD.html>
- [64] A. Banfi, P. Salam Gavin, G. Zanderighi, Eur. Phys. J. **C47** 113 (2006).
- [65] G. Zanderighi, arXiv:hep-ph/0705.1937.
- [66] CDF internal note 7806.
- [67] CDF internal note 4901.
- [68] R. Brun *et al.*, Tech. Rep. CERN-DD/EE/84-1,1987.
- [69] G. Grindhammer *et al.* Nucl. Instrum. Meth. **A290** 469 (1990).
- [70] CDF internal note 7055.
- [71] R. Field, Proceedings of the Fermilab ME/MC tuning workshop, Fermilab, October 2002.

- [72] CDF internal note 7984.
- [73] CDF internal note 8097.
- [74] CDF internal note 8060.
- [75] A. Bhatti, F. Canelli, B. Heinemann *et al.*, arXiv:hep-ex/0510047
- [76] CDF internal note 2925.
- [77] C. New *et al.*, CDF Note 7578, 2005.
- [78] CDF internal note 7343.
- [79] CDF internal note 7448.
- [80] R. Barlow, C. Beeston, Comput. Phys. Commun. **77**, 219, 1993.
- [81] CDF internal note 6894.
- [82] CDF internal note 8329.
- [83] CDF internal note 6985.
- [84] CDF internal note 7197.
- [85] J.M. Butterworth, J.R. Forshaw, M.H. Seymour, Multi-parton interactions in photoproduction at HERA.
- [86] S. Frixione, B.R. Webber, Matching NLO QCD computations and parton shower simulations, JHEP 0206 (2002) 029 [hep-ph/0204244].
- [87] S. Frixione, P. Nason, B.R. Webber, Matching NLO QCD and parton showers in heavy flavour production, JHEP 0308 (2003) 007 [hep-ph/0305252].
- [88] CDF internal note 8487.



# Acknowledgements

The years of a PhD are probably the most important in the formation of a young physicist. Here, I would like to thank all the persons that helped me going through these four years at the university of Geneva.

First of all, my supervisor, prof. Allan Clark, for offering me the opportunity to join the CDF/ATLAS group, and also for the way he has always been supportive when I've needed it. Thank you to all the other members of the Geneva CDF group, Maria Kientzle, Xin Wu and Mario Campanelli for the nice discussions during our weekly meetings. And in particular I really thank Mario, always supportive: sharing the office with me, he had the hard task of listening to my many negative letting offs, and he always had useful advices to give me. Many thanks to Regis Lefevre: even though we didn't really work closely together, his advices and suggestions, as QCD convener and also as member (for some months) of our group in Geneva, have been crucial. And I want to thank, as well, the whole CDF QCD group as every single meeting I have attended has been an important help for my analysis.

I would like to thank also other many wonderful people that have helped me since my very first day at the University of Geneva, Catherine, in particular, our secretary at the University, for being so helpful and so patient! And Yann for his help whenever everything, in my computer, seemed lost.

The real friends now! None of these years would have been the same without them! So... Thanks to Mercedes, first of all, for welcoming me the very first day I landed in Geneva and I knew no one! Thanks for letting me among your friends and making me meet one of them, in particular!!!. Thanks to Chiara, for sharing many of my cares and concerns, and thanks to Mauro. Above all, thanks to Shulamit! Thanks for sharing with me "the house, the office, the car" ... the life... when we've been together at Fermilab. Thanks for the birthdays, thanks for the support, the long discussions we had.

Thank you so much!!

E... Grazie! ...Alla mia famiglia: a Mamma, a Papà e a Irene! Decidere di partire non è stato facile, ma con il vostro sostegno, i vostri consigli e anche la vostra

allegria sono riuscita a sentirmi come se fossi rimasta sempre a casa! Un bacione grandissimo!

Gracias a ti, consegui!

Com a tua ajuda e atenção deste-me força e apoiaste-me dia a dia. Sem ti perto de mim, tenho a certeza que nunca poderia concluir o PhD neste tempo. Obrigada.

Um beijo

