

UNIVERSITÉ DE STRASBOURG

THÈSE

présentée pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE STRASBOURG

Spécialité : Physique des Particules

présentée par

Jérémy Andrea

**Mesure de la section efficace de production de paires de quarks
top dans l'expérience CMS au LHC**

M. D. BLOCH	(Directeur de thèse)
M. J. D'HONDT	(Rapporteur externe)
M. D. FROIDEVAUX	(Rapporteur externe)
M. D. HUSS	(Rapporteur interne)
M. D. CONTARDO	(Examineur)
Mme M-H. SCHUNE	(Examineur)

A Carine,

A mes parents.

*L'intuition est à la connaissance
ce que la canne blanche est à l'aveugle.*

Jérôme Touzalin.

Remerciements

Cette partie de la thèse est réputée pour être la plus difficile à rédiger. C'est vrai... Je vais cependant essayer de n'oublier personne, si je n'y arrive pas, j'offre une bière à chaque personne omise, pour me faire pardonner!

J'ai eu l'occasion de travailler avec beaucoup de membres de CMS, que ce soient des français, des allemands, des russes, des américains ou des italiens... Vous me permettrez donc sans doute d'écrire une partie de ces remerciements en anglais, puisque je ne parle ni allemand, ni russe, ni italien (et à peu près français)...

Je commencerai par remercier très chaleureusement mes rapporteurs Jorgen D'Hondt, Daniel Froidevaux et Daniel Huss. Merci d'avoir pris le temps de lire en détail et de corriger mon manuscrit. Merci également aux examinateurs Marie-Hélène Schune et Didier Contardo, qui ont également suivi en profondeur le travail que j'ai effectué (merci à Marie-Hélène pour les corrections!).

J'aimerais également remercier mes deux encadrants de thèse : Daniel Bloch, mon directeur de thèse, et Denis Gelé. J'aimerais leur exprimer toute ma gratitude et mon amitié. Je considère comme étant une grande chance, très sincèrement, d'avoir pu apprendre de physiciens de leur qualité. Merci pour votre patience, merci d'avoir supporté mes coups de blues et mes coups de sang, merci de votre soutien dans les moments difficiles mais surtout merci pour les connaissances que vous m'avez transmises. Avec le petit recul que j'ai, je pense avoir vécu la thèse comme une sorte de voyage initiatique dont vous avez été des guides éclairés et je vous en remercie. Merci également à Pierre Juillot pour sa gentillesse, les journées et week-ends passés à produire des données et pour ses anecdotes croustillantes. Je n'oublierai pas la théorie de la chemise bleue, variante de la loi de Murphy...

Switching to English, I would like to thank a lot all the people from CMS I have worked with during my PhD. I benefit from their experiences and from their advices in many of my works. I had the opportunity to work with many people and here is a not exhaustive list. First, I would like to thank again Jorgen D'Hondt as a convener of the Top group. I really appreciate (and benefit also many times) the support he is giving to young researchers (I have really enjoyed being at Lake Louise!). I'm also really thankful to Claudio Campagnari and Tim Christiansen, the actual Top conveners, for their support and the very helpful advices. I would like also to thank Kevin Burket, Boris

Mangano (Tracking conveners) and Didier Contardo (Tracker DPG convener) for their guidance through my work on cosmic muons. I would like to thank also Riccardo Bellan who was the first one to propose the method I developed for tracking efficiency estimate. Thank you very much to Ian Tomalin, Thomas Speer, Tommaso Boccali and Wolfgang Adam (b-tagging conveners during my PhD) for their help, guidance and sympathy.

Many thanks to all of the other people that I have worked with or just for their friendships (sorry for any omission!!) : Roberto Cherici, Stephanie Beauceron (thx for kindness and advices for my travel to Australia!), Frank-Peter Schilling, Christophe Saout, Francisco Yumiceva, Victor Bazterra, Meenakshi Narain, Andrea Giammonco, Steven Lowette, Gregory Hammad (congratulations guy!), Andrea Rizzi, Jan Heninck, Petra Van Mulders, Ilaria Vilella, Gen Kukartzev, Slava Krutleyov, Jesus Vizan, Julien Caudron, Roger Wolf, Jake Ribnik, Javier Cuevas Maestro, Pedrame Bargassa.

Je voudrais remercier les membres de l'IPHC que j'ai côtoyé durant ma thèse. Tout d'abord les membres du groupe CMS : Anne-Catherine Lebihan (félicitations encore pour la petite Anouk, à quand le deuxième?), Auguste Besson (trop sexy tes lunettes!), Benjamin Fuks (muf muf et re-muf), Eric Comte (paix et félicité!), Anne-Fleur Barfuss (qui a un joli crochet du droit), Pierre Van Hove, Jean Laurent Agram, Ulrich Georlach, Yoshi Mikami et David Bodin. Un merci tout spécial à Joaquim Speck, la "beaugossitude" luxembourgeoise même, qui m'a supporté pendant deux ans dans le même bureau que lui! Il volera bientôt vers de nouveaux cieus, enjoy my friend! Je n'oublierai pas Yannick Patois, à qui j'ai cassé les pieds un nombre incalculable de fois avec mes problèmes de grille...

Merci également à Isa (merci pour ta gentillesse), Jérôme, Seb (on se revoit dans CMS bientôt?;) et Antonin (l'amateur de Picon Bière). Merci à Leila, Jean Pierre, Frobi and Jean pour leur aide et leur sympathie. Spéciale dédicace à mon chaton, Vincent, je compte sur toi en cas de démangeaisons oculaires... Merci aux membres de l'université de Strasbourg que j'ai côtoyé au cours de mon monitorat : Patrick, Boris, Cécile, Vincent et Amine.

J'ai aussi reçu beaucoup de soutien moral de la part de mes amis proches, sans qui j'aurais certainement sombré dans une profonde dépression... Merci donc à ma Toutounne pour son amitié infaillible, à Hélène (je vous aime gros!) et au p'tit Louis (mon filleul, c'est le plus beau!!!), à Emilyne et JF expatriés dans les îles (veinards!), à Nico, Vanessa et au p'tit Ethan (prenez soin de vous), à Aline (pas toujours maline;)), Scheumel (vivement les vacances!), Bibi (t'es beau mon Bibi en costume!) et Fonfec, à Delphine et Sam (you knowwwww...), à Antho et Isa (quelle manie vous avez tous à vouloir vous marier?), à Doro, à Juju, Cédric, Julien K., Fabucodonosaure, Clément, Greg, Mathieu et les autres... Merci donc à tous, j'espère qu'à l'avenir nous pourrons nous voir plus souvent! A tous ceux que j'ai oubliés, la tireuse à bière est prête!!!

J'aimerais profiter de cette occasion pour remercier mes parents et leur dire que je les aime. J'ai conscience des sacrifices qu'ils ont fait pour financer mes études. Désolé aussi d'avoir été si peu présent ces trois dernières années. À mes frères j'aimerais dire : bon courage pour la fin de ta thèse Jonathan et j'espère que tu feras la grande carrière scientifique que tu mérites (mais pour ça, faut passer chez le coiffeur ! Imagine que tu aies autant de cheveux dans que sur la tête... Ca ne facilite pas la connexion des neurones !). Luc, tout d'abord ce n'est pas parce que tu fais 1m90 que je vais arrêter de t'appeler "le p'tit". Bon courage pour la suite de tes études, en bossant un peu tu arriveras loin, j'en suis sûr !

Finalement, j'aimerais remercier la personne à qui je dois le plus, Carine. Elle s'est occupée de moi, a sacrifié ses soirées, ses week-ends, m'a supporté dans les moments difficiles. Elle m'apporte l'équilibre dont j'ai besoin pour pouvoir avancer. Tous les efforts qu'elle a faits pour me permettre de réaliser mes rêves sont les plus belles preuves d'amour que je puisse espérer. J'aimerais lui dire à quel point je l'aime et à quel point je lui suis reconnaissant. T'es une dondon atomique !

Table des matières

Introduction	1
1 Le Modèle Standard de la Physique des Particules au LHC	3
1.1 Introduction au Modèle Standard	3
1.2 Introduction aux théories de jauge	5
1.2.1 Champs et symétries	5
1.2.2 Théorie de jauge	6
1.2.3 Diagrammes de Feynman et développement perturbatif	8
1.2.4 Renormalisation	10
1.3 Modèle Standard de la physique des particules	10
1.3.1 L'interaction forte	10
1.3.1.1 Symétrie $SU(3)$ et charges colorées	10
1.3.1.2 Chromodynamique Quantique	11
1.3.1.3 La constante de couplage forte	11
1.3.1.4 Production d'événements QCD au LHC	12
1.3.2 Le modèle électrofaible	13
1.3.3 Production de bosons W et Z dans les collisions proton-proton . .	15
1.3.4 Brisure de symétrie et modèle de Higgs	16
1.4 Au-delà du Modèle Standard	18
1.5 Physique du quark top	21
1.5.1 Propriétés du quark top	21
1.5.1.1 Masse du top et contraintes sur la masse du boson de Higgs	22
1.5.1.2 Désintégration du quark top	22
1.5.2 Modes de production	24
1.5.2.1 Production par interaction forte	24
1.5.2.2 Production par interaction faible	24
1.5.3 Canaux de désintégration	25
1.5.4 Motivation pour l'étude du top au LHC	26
1.6 Génération des événements	28

2	Le Dispositif Expérimental	31
2.1	Le LHC : un collisionneur proton-proton	31
2.2	Le détecteur CMS	34
2.3	La trajectographie	36
2.3.1	Le trajectographe	36
2.3.2	Reconstruction de la trajectoire des particules chargées	38
2.3.2.1	Reconstruction des traces	38
2.3.2.2	Reconstruction des vertex	43
2.4	Le spectromètre à muons	43
2.4.1	Le dispositif de détection	44
2.4.2	Reconstruction des muons	45
2.4.2.1	Les muons locaux	45
2.4.2.2	Les muons globaux	45
2.5	La calorimétrie	48
2.5.1	Le calorimètre électromagnétique ECAL	49
2.5.2	Le calorimètre hadronique HCAL	50
2.5.3	Reconstruction des électrons	52
2.5.4	Reconstruction des jets de particules et de l'énergie transverse manquante	54
2.5.4.1	Les algorithmes de reconstruction des jets	54
2.5.4.2	Reconstruction de l'énergie transverse manquante	60
2.6	Acquisition, reconstruction et simulation des données	63
2.6.1	Le système de déclenchement	63
2.6.2	Reconstruction des événements et l'environnement d'analyse	64
2.6.3	La simulation des événements	65
3	Estimation des efficacités de reconstruction des traces à l'aide de muons cosmiques.	67
3.1	Algorithmes de reconstruction des traces pour les données cosmiques	68
3.1.1	Adaptation de l'algorithme <i>Combinatorial Track Finder</i> (CTF) à la prise de données cosmiques	69
3.1.2	Présentation de l'algorithme <i>CosmicTrackFinder</i>	69
3.1.3	Présentation de l'algorithme <i>RoadSearch</i>	70
3.2	Sélection des muons cosmiques	71
3.3	Données cosmiques et simulation Monte-Carlo	73
3.4	Etude comparative des muons locaux et des traces reconstruites dans le trajectographe	80
3.5	Estimation de l'efficacité de reconstruction des traces dans les données <i>Craft</i>	87
3.6	Conclusion et perspectives	94

4	Identification des jets de quark beau.	97
4.1	Algorithmes d'identification des jets de quark beau	98
4.1.1	Algorithmes basés sur le paramètre d'impact des traces	99
4.1.1.1	Le comptage de traces	102
4.1.1.2	La méthode probabiliste	106
4.1.2	Description des autres algorithmes d'identification	116
4.2	Evénements utilisés et définition de la taggabilité	122
4.3	Mesure de l'efficacité d'identification des jets de quark beau	125
4.4	Estimation du taux de mauvaise identification	133
4.4.1	Description de la méthode	134
4.4.2	Etudes des traces provenant de processus déplacés	140
4.4.3	Erreurs systématiques.	146
4.4.4	Paramétrisation de la taggabilité et de l'efficacité d'identification	152
4.5	Conclusion et perspective	154
5	Mesure de la section efficace de production d'événements $t\bar{t}$	157
5.1	Caractéristiques des événements	158
5.1.1	Evénements $t\bar{t}$ di-leptons	158
5.1.2	Bruits de fond	161
5.1.3	Stratégie de sélection	162
5.2	Mesure de la section efficace	163
5.2.1	Les données utilisées	163
5.2.1.1	La production CSA07	163
5.2.1.2	Simulation des événements	164
5.2.2	Chemins de déclenchement	167
5.2.3	L'isolation des leptons	170
5.2.4	Identification des électrons	175
5.2.5	Choix du canal	176
5.2.6	Sélection principale	176
5.3	Estimation du bruit de fond QCD dans la simulation	183
5.3.1	Factorisation	184
5.3.2	Sensibilité au taux de faux électrons	185
5.4	Estimation du bruit de fond à partir des données : la méthode de la matrice di-électrons	187
5.4.1	Description de la méthode	188
5.4.2	Hypothèse de factorisation	189
5.4.3	Estimation de ε_s et ε_{fake} à partir des données	190
5.4.4	Méthode de résolution du système d'équations	191
5.4.5	Test de la méthode	191
5.4.5.1	Efficacités	191
5.4.5.2	Erreurs systématiques	195
5.4.5.3	Résultats	200

5.5	Estimation du bruit de fond Z +jets	207
5.6	Stratégie de mesure de la section efficace	211
5.7	Erreurs systématiques	215
5.7.1	Incertitudes liées au détecteur	216
5.7.1.1	Echelle en énergie et résolutions des jets	216
5.7.1.2	Algorithme d'identification des jets b	217
5.7.2	Incertitude liée à la génération des événements	218
5.7.2.1	Radiation QCD	218
5.7.2.2	Fragmentation	219
5.7.2.3	Evénements sous-jacents	220
5.7.2.4	PDF	221
5.7.2.5	La masse du top	221
5.7.3	Erreurs Systématiques globales	222
5.8	Conclusion	222
6	Conclusion	225
A	Taux de Mauvaise identification	227
A.1	Erreurs systématiques de l'estimation du taux de mauvaise identification	227
A.2	Résultats pour les point de fonctionnement <i>loose</i> et <i>tight</i>	228
A.3	Les différents algorithmes d'étiquetage négatif	228
B	system8	245
	Bibliographie	263

Introduction

Le grand collisionneur de hadrons, LHC, du CERN (laboratoire Européen pour la recherche en physique des particules, situé près de Genève) verra voir ses premières collisions proton-proton à l'automne 2009. Avec une énergie disponible inégalée, 7 TeV par faisceau de protons, le LHC est principalement dédié à la découverte du boson de Higgs (la clef de voûte du Modèle Standard des particules élémentaires, responsable de leurs masses) et de tout signal de nouvelle physique. Le détecteur de l'expérience CMS a été conçu pour réaliser cet ambitieux programme de recherche mais aussi pour réaliser des mesures fines du Modèle Standard.

Le quark top (t) est de loin le fermion le plus lourd et c'est le seul quark à se désintégrer directement sans s'hadroniser. Il sera copieusement produit aux énergies du LHC, permettant d'effectuer des tests de précision du Modèle Standard et pouvant être utilisé comme source de calibration pour les différents sous-détecteurs de *CMS*. Dans cette thèse, je me suis consacré à la mesure de la section efficace de production de paires de quarks $t\bar{t}$. Dans le Modèle Standard, chaque quark top se désintègre essentiellement en un boson W et un quark beau (b), le bosons W se désintégrant soit en paire lepton-neutrino, soit en paire quark-antiquark. Dans 5% des cas, l'état final $t\bar{t}$ est constitué de deux jets de quarks b , de deux leptons chargés (électron ou muon) et d'énergie transverse manquante provenant des neutrinos non détectés. Je me suis intéressé à cet état final, car le bruit de fond correspondant est relativement faible, ce qui permet une détection relativement simple des événements $t\bar{t}$, même à faible luminosité. Les échantillons sélectionnés peuvent être enrichis en signal à l'aide d'algorithmes d'identification des jets de quark beau. Une partie de ma thèse a été consacrée à la calibration de tels algorithmes ainsi qu'à la mesure de leur taux de mauvaise identification. Ces algorithmes utilisent en particulier les traces des particules chargées reconstruites avec le trajectographe de *CMS*. J'ai également mesuré l'efficacité de reconstruction de ces traces à l'aide de muons cosmiques acquis à partir de juillet 2008.

Le premier chapitre introduit le cadre théorique de ce travail : le Modèle Standard de la physique des particules. Dans le second chapitre, l'accélérateur LHC et le détecteur CMS seront présentés. Les différents algorithmes de reconstruction des objets physiques seront brièvement décrits. Le troisième chapitre est dédié à la mesure de l'efficacité de reconstruction des traces, dans les données réelles, à l'aide de muons cosmiques. Le chapitre 4 présente les travaux effectués sur les algorithmes d'identification des jets b .

Plus particulièrement, il s'agit de la calibration de l'algorithme de probabilité par jet et de la mise en place de méthodes de mesure des efficacités d'identification des jets b et du taux de mauvaise identification des jets de quarks légers et de gluons. Finalement, le chapitre 5 présente une analyse dédiée à la mesure de la section efficace de production de paires de quark top dans les canaux di-leptons. La sélection des événements, l'estimation des bruits de fond à partir des données et l'estimation des erreurs systématiques seront présentées.

1

Le Modèle Standard de la Physique des Particules au LHC

1.1 Introduction au Modèle Standard.

Le Modèle Standard de la physique des particules [1, 2, 3, 4, 5, 6, 7] est la théorie qui décrit la physique des particules élémentaires et leurs interactions (cf. figure 1.1). Il a été élaboré après plus de 50 ans de travaux théoriques et se base sur les concepts fondamentaux de théorie quantique des champs et sur des observations expérimentales.

La relativité restreinte et la mécanique quantique servent d'incubateur à la théorie quantique des champs et au Modèle Standard. Les particules y sont représentées par des champs relativistes qui obéissent à des équations de propagation¹. La quantification d'une théorie des champs se fait par l'introduction d'opérateurs de création et d'annihilation, analogues à ceux présents en mécanique quantique, pour rendre compte de la création et de l'annihilation possible des particules.

Utilisant le formalisme de la théorie quantique des champs et les concepts de symétries et d'invariance de jauge, l'Electro-Dynamique Quantique (QED) voit le jour en 1946. En 1954 Yang et Mills proposent une théorie de jauge "non-abélienne" pour décrire les interactions faibles. Dans cette théorie, les fermions interagissent en échangeant des particules appelées bosons d'interaction, de spin entier. Combinant cette théorie avec le modèle de Higgs, Brout et Englert, qui permet d'attribuer des masse aux bosons vecteurs, Glashow [8], Salam [9] et Weinberg [10] proposent une théorie d'unification des

¹En 1928 Dirac propose une équation d'onde relativiste pour l'électron.

interactions électromagnétique et faible (1967) : la théorie électrofaible. Les bosons vecteurs de l'interaction faible, les bosons Z et W^\pm , ont été découverts en 1983 au CERN² par les expériences UA1 [11, 12] et UA2 [13, 14].

En 1964, l'existence de trois quarks (u , d et s) est proposée. La non observation de courant neutre avec changement de saveur laisse supposer l'existence d'un quatrième quark (mécanisme de GIM), le quark charmé découvert en 1974. L'introduction du concept de liberté asymptotique (1973) permet de décrire l'interaction forte en utilisant une théorie de jauge non-abélienne. L'introduction d'une charge colorée portée par les quarks (la couleur) donne naissance à la Chromo-Dynamique Quantique (ou QCD). De plus, l'oscillation des mésons K_0 et \bar{K}_0 est observée, suggérant l'existence d'au moins deux autres quarks, le quark beau (découvert en 1974) et le quark top découvert en 1995 par les expériences CDF et DØ au Tevatron [15].

$$\begin{array}{l}
 \text{première famille} \\
 \text{deuxième famille} \\
 \text{troisième famille}
 \end{array}
 \begin{array}{l}
 \left(\begin{array}{c} \nu_e \\ e^- \end{array} \right)_L \\
 \left(\begin{array}{c} \nu_\mu \\ \mu^- \end{array} \right)_L \\
 \left(\begin{array}{c} \nu_\tau \\ \tau^- \end{array} \right)_L
 \end{array}
 \begin{array}{l}
 e_R^- \\
 \mu_R^- \\
 \tau_R^-
 \end{array}
 \begin{array}{l}
 \left(\begin{array}{c} u \\ d \end{array} \right)_L \\
 \left(\begin{array}{c} c \\ s \end{array} \right)_L \\
 \left(\begin{array}{c} t \\ b \end{array} \right)_L
 \end{array}
 \begin{array}{l}
 u_R \quad d_R \\
 c_R \quad s_R \\
 t_R \quad b_R
 \end{array}$$

TAB. 1.1: Fermions élémentaires du Modèle Standard.

Le Modèle Standard comporte 3 générations de particules (figure 1.1), chacune contenant une paire de leptons et une paire de quarks. Les interactions fondamentales qu'il décrit sont au nombre de trois : l'interaction électromagnétique véhiculée par le photon, l'interaction faible véhiculée par les bosons Z et W^\pm et l'interaction forte véhiculée par les gluons. Le Modèle Standard reste cependant une théorie incomplète. Il n'inclut pas l'une des interactions fondamentales : la gravitation. L'amplitude de l'interaction gravitationnelle est cependant négligeable à l'échelle des particules et aux énergies des collisionneurs (il existe environ 40 ordres de grandeur de différence entre les amplitudes de l'interaction gravitationnelle et électromagnétique).

Dans ce chapitre, la notion de champ et le principe des théories de jauge vont être tout d'abord présentés dans la section 1.2 à l'aide d'un exemple simple, l'électrodynamique quantique. Le Modèle Standard de la physique des particules sera alors introduit dans la section 1.3 et certaines de ses extensions possibles seront exposées brièvement dans la section 1.4. Finalement, la physique du quark top sera introduite dans la section 1.5.

²Organisation Européenne pour la recherche en physique des particules et en physique nucléaire

1.2 Introduction aux théories de jauge

En théorie quantique des champs, les particules sont décrites par des champs relativistes et quantifiés, c'est-à-dire auxquels on a ajouté des propriétés quantiques. D'une manière générale, une théorie des champs est tout d'abord élaborée au niveau classique à l'aide de la mécanique lagrangienne. Ensuite, les principes de symétrie sont utilisés pour décrire les différentes interactions ainsi que les quantités conservées à l'aide de l'invariance de jauge et du théorème de Noether. La théorie est quantifiée en utilisant la quantification canonique par exemple ou les intégrales de chemin de Feynman. Il faut alors déterminer si la théorie est renormalisable, c'est à dire si elle ne conduit pas à des divergences dans le calcul des différentes quantités physiques. Dans ce chapitre, une courte introduction de ces différentes notions sera présentée.

1.2.1 Champs et symétries

La description des particules en théorie des champs est très fortement liée aux concepts de symétries et d'invariance de jauge en mécanique lagrangienne. Une symétrie correspond à une transformation appliquée aux champs qui laisse les équations du mouvement invariantes. La notion de symétrie est tout particulièrement intéressante en physique car elle permet de mettre en évidence les quantités physiques conservées.

Le théorème de Noether indique que pour toute symétrie il existe une quantité physique conservée qui lui correspond. On peut distinguer deux types de symétries distinctes :

- les symétries externes liées aux propriétés d'espace-temps, décrites par des transformations correspondent au groupe de Poincaré : transformation de Lorentz, rotation (quantité conservée : moment angulaire) et translation dans l'espace-temps (quantités conservées : impulsion et énergie),
- les symétries internes, propres aux champs considérés, correspondent aux symétries de jauge.

L'invariance sous le groupe de Poincaré assure que la théorie est relativiste et isotrope. Les symétries de jauge sont utilisées afin de décrire les interactions entre particules.

Ces dernières sont décrites par des champs quantifiés dont la nature va dépendre de leur spin. Les constituants de la matière sont des particules de spin $1/2$ et peuvent être représentés par des champs de type spineur à deux composantes χ_L et χ_R liés l'un à l'autre par une transformation de parité. Les champs χ_L et χ_R correspondent à des chiralités droite et gauche, ils sont regroupés en *quadri-spineur* (χ_L, χ_R) ou spineur de Dirac. Les bosons d'interaction ont un spin égal à 1 et sont représentés par des champs vectoriels. A chaque type de particule correspond un lagrangien libre dont les équations du mouvement décrivent leur propagation (tableau 1.2).

Champs	Spin	Particule	Lagrangien
scalaire φ	0	boson de Higgs	Lagrangien de Klein-Gordon $\mathcal{L}_{KG} = \frac{1}{2}(\partial_\mu\phi\partial^\mu - m^2\phi^2)$
spineur $\psi(x) = \begin{pmatrix} \chi_L(x) \\ \chi_R(x) \end{pmatrix}$	1/2	quarks et leptons	Lagrangien de Dirac $\mathcal{L}_D = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi$
vecteur A^μ	1	bosons d'interaction	Lagrangien de Maxwell pour la jauge U(1) $\mathcal{L}_M = -\frac{1}{4}(F_{\mu\nu}F^{\mu\nu})$ avec $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$

TAB. 1.2: Champs de particules du Modèle Standard et les lagrangiens correspondants.

1.2.2 Théorie de jauge

En mécanique quantique, la phase d'une fonction d'onde peut être traitée comme une rotation arbitraire indépendante en chaque point de l'espace-temps. Comme cette phase n'est pas mesurable, la mécanique quantique possède une *invariance de jauge* ou *symétrie de jauge* pour une transformation de changement de phase. Dans le cas où cette phase est une constante, la symétrie de jauge est qualifiée de globale. Lorsque cette phase dépend de l'espace et du temps, la symétrie de jauge est qualifiée de locale. Dans le cas de jauge globale, les théories sont en général trivialement invariantes. Le cas de l'invariance de jauge locale est cependant plus délicat, il est régi par le *principe de jauge* qui dit que pour garder les lois de la physique invariantes sous une transformation de jauge locale, il doit exister un champ appelé champ de jauge : *“Imposer l'invariance du Lagrangien sous une transformation de jauge locale permet donc de “générer” des interactions et d'introduire des champs vectoriels qui seront les médiateurs des forces dans le langage de la théorie des champs”* [16].

En physique des particules, ces champs de jauge correspondent aux bosons d'interaction (ou bosons de jauge). Les transformations de jauge locales dans le Modèle Standard sont décrites par des groupes de transformation constitués des matrices unitaires de déterminant 1 et de dimensions $n \times n$ notées $SU(n)$. Pour des matrices unitaires $U(\theta^a)$ de paramètres θ^a (pour $a = 1, \dots, d$) et pour une théorie impliquant un multiplet de champs ϕ_i (avec $i = 1, \dots, n$), la théorie est symétrique par rapport à la transformation :

$$\phi_i(x) \rightarrow \phi'_i(x) = U(\theta^a)_{ij}\phi_j(x), \quad (1.1)$$

si les équations du mouvement restent inchangées sous cette transformation. Cela se traduit par l'invariance du Lagrangien de la théorie :

$$\mathcal{L}(\phi, \partial_\mu\phi) = \mathcal{L}(\phi', \partial_\mu\phi') = \mathcal{L}(U\phi, \partial_\mu U\phi). \quad (1.2)$$

Une matrice unitaire U peut s'écrire en fonction d'une matrice hermitienne H : $U = e^{iH}$. Dans le cas d'une transformation infinitésimale on obtient :

$$U(\theta) = 1 + i\theta^a H^a + \mathcal{O}((\theta^a)^2), \quad (1.3)$$

et la transformation du champ $\phi_i(x)$ s'écrit :

$$\phi'_i(x) = \phi_i(x) + i\theta^a H_{ij}^a \phi_j(x) + \mathcal{O}((\theta^a)^2) \quad (1.4)$$

où les paramètres θ^a sont petits. Les matrices hermitiennes H^a sont les représentations matricielles des générateurs du groupe de transformation U et forment une algèbre de Lie :

$$[H^a, H^b] = iC^{abc} H_c. \quad (1.5)$$

Les constantes C^{abc} sont appelées constantes de structure du groupe de transformation. Lorsqu'elles sont nulles, la théorie de jauge est qualifiée d'abélienne (les matrices T^a commutent).

A titre d'exemple, on peut considérer l'équation de Dirac pour un électron libre dans le cas de l'invariance sous la symétrie de jauge locale abélienne $U(1)$ qui correspond simplement à une phase complexe $e^{i\theta(x)}$. Le Lagrangien de Dirac pour un électron libre s'écrit :

$$\mathcal{L}_D = \bar{\psi}(x)(i\gamma^\mu \partial_\mu - m)\psi(x), \quad (1.6)$$

où $\psi(x)$ et $\bar{\psi}(x)$ représentent l'électron et le positron respectivement et m est la masse de l'électron. La densité lagrangienne 1.6 est trivialement invariante sous transformation de la symétrie globale $U(1)$:

$$\psi(x) \rightarrow \psi' = e^{-i\theta} \psi(x), \quad (1.7)$$

$$\bar{\psi}(x) \rightarrow \bar{\psi}' = e^{i\theta} \bar{\psi}(x). \quad (1.8)$$

La symétrie locale correspondante est :

$$\psi(x) \rightarrow \psi' = e^{-i\theta(x)} \psi(x), \quad (1.9)$$

$$\bar{\psi}(x) \rightarrow \bar{\psi}' = e^{i\theta(x)} \bar{\psi}(x). \quad (1.10)$$

Le terme de masse du Lagrangien 1.6 reste invariant pour la transformation locale et le terme de dérivée devient :

$$\bar{\psi}(x)i\gamma^\mu \partial_\mu \psi(x) \rightarrow \bar{\psi}(x)i\gamma^\mu \partial_\mu \psi(x) + \bar{\psi}(x)\gamma^\mu (\partial_\mu \theta(x))\psi(x). \quad (1.11)$$

Pour compenser ce terme supplémentaire, on peut introduire un *champ de jauge* $A_\mu(x)$ se transformant sous la jauge $U(1)$ comme :

$$A_\mu(x) \rightarrow A'_\mu(x) = A_\mu(x) + \frac{1}{q} \partial_\mu \theta(x), \quad (1.12)$$

où q est la charge électrique de l'électron. A partir de ce champ vectoriel, on peut introduire une dérivée covariante ayant les propriétés de transformation suivantes :

$$D_\mu \psi(x) \rightarrow e^{-i\theta(x)} D_\mu \psi(x). \quad (1.13)$$

La dérivée covariante est définie comme :

$$D_\mu \psi(x) = (\partial_\mu + iqA_\mu(x))\psi(x). \quad (1.14)$$

En introduisant cette dérivée covariante et le champ vectoriel $A_\mu(x)$, le Lagrangien 1.6 devient invariant sous une transformation de jauge locale $U(1)$. Pour que le champ de jauge ait une réalité physique (qu'il ne soit pas éliminé lors du calcul des équations du mouvement), il faut ajouter un terme cinétique également invariant sous la transformation de jauge. Ce terme cinétique est de la forme $F^{\mu\nu}F_{\mu\nu}$ où $F^{\mu\nu}$ est le tenseur de Maxwell défini par :

$$F_{\mu\nu} = \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x). \quad (1.15)$$

Finalement, la densité lagrangienne 1.6 devient :

$$\mathcal{L}_D^{QED} = \bar{\psi}(x)(i\gamma^\mu D_\mu - m)\psi(x) - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \quad (1.16)$$

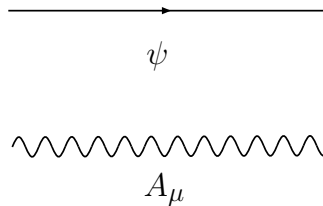
$$= \bar{\psi}(x)(i\gamma^\mu \partial_\mu - m)\psi(x) - q\bar{\psi}(x)\gamma^\mu A_\mu(x)\psi(x) - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}. \quad (1.17)$$

Le Lagrangien \mathcal{L}_D^{QED} décrit la théorie appelée électrodynamique quantique (QED). Le premier terme de ce Lagrangien correspond au terme de propagation et de masse d'un fermion et le second terme correspond au couplage entre un fermion et un boson d'interaction : le photon. Le dernier terme décrit la propagation d'un photon libre. Il n'existe pas de terme de masse pour le photon car la présence d'un tel terme, qui serait de la forme $mA^\mu A_\mu$, briserait l'invariance de jauge. Nous venons de construire une théorie de l'électromagnétisme : la QED .

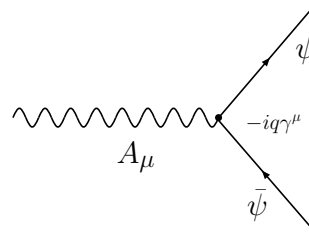
Les choses se compliquent cependant pour des symétries plus complexes lorsque des bosons massifs doivent être présents dans la théorie. Il faut alors introduire un champ scalaire et le mécanisme de brisure spontanée de symétrie (cf section 1.3.4).

1.2.3 Diagrammes de Feynman et développement perturbatif

Une telle théorie doit être quantifiée. Le schéma de quantification par *intégrale de chemin*, combiné aux équations de mouvement, permet de calculer un processus sous la forme d'une somme d'intégrales en effectuant un développement perturbatif en puissances de la constante de couplage. Pour chaque terme de cette somme, un ensemble de *diagrammes de Feynman* peut être introduit dans lequel les fonctions décrivant les particules libres (les propagateurs) sont représentées par des lignes :



et chaque vertex d'interaction par :



où ψ représente les champs de matière et A_μ les champs de jauge.

Prenons l'exemple d'un développement perturbatif au premier ordre du processus $e^+e^- \rightarrow \mu^+\mu^-$ dans le cadre de la QED . Traité de manière inclusive, la seule information utilisable est que l'état initial doit contenir un électron et un positron et que l'état final doit contenir un muon et un antimuon. Les calculs liés à ce processus peuvent alors se représenter sous la forme de la somme des diagrammes de Feynman, comme indiqué sur la figure 1.1.

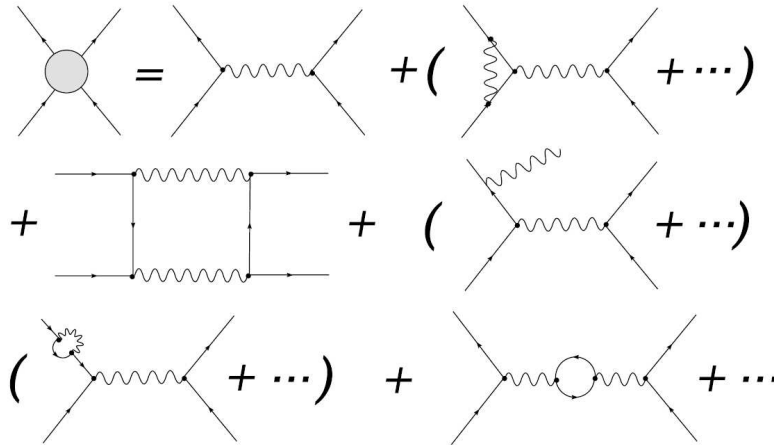


FIG. 1.1: Illustration du développement perturbatif en diagrammes de Feynman du processus $e^+e^- \rightarrow \mu^+\mu^-$ au *Next-to-Leading Order*.

Le premier diagramme est appelé *diagramme au niveau de l'arbre* et correspond au premier ordre du développement perturbatif et à une puissance 2 de la constante de couplage e^2 , on parle de *Leading Order* (ou LO). Les diagrammes suivants sont appelés *Next-to-Leading Order* (ou NLO) et correspondent à l'ordre 2 du développement perturbatif et à une puissance 4 de la constante de couplage e^4 . Entre parenthèses, seul un exemple de diagrammes de même type est représenté. Les ordres supérieurs ne sont pas représentés.

1.2.4 Renormalisation

De tels développements en puissance de la constante de couplage interviennent également dans les propagateurs des fermions (*self interaction*) et des bosons (*polarisation du vide*). Les diagrammes correspondants dans le cas du processus $e^+e^- \rightarrow \mu^+\mu^-$ au LO sont représentés par les deux derniers diagrammes de la figure 1.1.

Ces termes introduisent des divergences ultra-violettes dans le calcul des quantités physiques. Ces divergences sont traitées par la méthode de *renormalisation*. La renormalisation est une procédure qui permet de compenser les contributions infinies en renormalisant les champs, les masses et les constantes. L'idée est de définir dans un premier temps une théorie avec des quantités (champs et constantes) "nues". Après renormalisation, les divergences provenant de la *self interaction* et de la *polarisation du vide* sont absorbées dans les quantités renormalisées. La présence des diagrammes d'émission réelles permet de compenser les divergences infra rouge restantes. Les quantités deviennent ainsi finies et physiques. A cause du processus de renormalisation, les constantes de couplage et les masses des particules évoluent avec l'énergie (cf section 1.3.1.3).

1.3 Modèle Standard de la physique des particules

1.3.1 L'interaction forte

Dès les années soixante, l'existence de trois quarks (u , d et s) a été suggérée par Gell-Mann et Zweig [17] afin d'expliquer le large spectre des particules hadroniques : les mésons, constitués d'un quark et d'un antiquark et les baryons, constitués de 3 quarks. La découverte des baryons Δ^{++} , Δ^- et Ω^- , respectivement constitués de 3 quarks u , d et s a suggéré l'existence d'un nouveau nombre quantique, la couleur. En effet, les quarks des particules Δ^{++} , Δ^- et Ω^- , de spin $3/2$, se trouvent être dans le même état quantique. Les quarks étant des fermions, il y a une contradiction avec le principe d'exclusion de Pauli. De tels états liés ne peuvent exister que s'il existe un nombre quantique différent pour chaque quark qui les constituent : les charges de couleur rouge, verte et bleue (par convention) et les anti-couleurs correspondantes. En chromodynamique quantique (ou *QCD*), les états liés de quarks ne sont possibles que si la couleur globale du hadron est "blanche", c'est-à-dire nulle. Par exemple, un méson ne peut contenir un quark de couleur bleue que s'il contient également un antiquark anti-bleu.

1.3.1.1 Symétrie $SU(3)$ et charges colorées

La chromodynamique quantique est une théorie de jauge basée sur le groupe $SU(3)$, constitué de $n^2 - 1 = 8$ matrices et associée à la conservation de la charge de couleur. L'algèbre de Lie du groupe $SU(3)$ est défini par $[T_a, T_b] = if_{abc}T_c$ où f_{abc} sont les

constantes de structure de l'algèbre de Lie et les générateurs du groupe sont les matrices hermitiennes T , sont proportionnelles aux matrices de Gell-Mann. Le groupe $SU(3)$ est un groupe non-abélien car ses constantes de structure ne sont pas nulles, ce qui permettra par la suite aux gluons d'interagir entre eux. Les bosons de jauge correspondant à l'invariance de jauge du Lagrangien de Dirac sous une transformation de groupe $SU(3)$ sont appelés gluons. Ils sont au nombre de 8 (car il y a 8 générateurs pour le groupe $SU(3)$) et portent eux même deux charges colorées. Ils vont donc pouvoir interagir avec eux-même ce qui n'était pas le cas pour les photons et la symétrie $U(1)$, les photons ne portant pas de charge électrique.

1.3.1.2 Chromodynamique Quantique

Afin de rappeler que la symétrie $SU(3)$ correspond aux charges colorées, elle est usuellement notée $SU(3)_c$. Les quarks dans la représentation de $SU(3)_c$ sont donc décrits par des triplets de couleur. De la même manière que pour la QED , le Lagrangien après introduction de la symétrie de jauge $SU(3)_c$ et de la dérivée covariante correspondante s'écrit :

$$\mathcal{L}_D^{QCD} = \bar{\psi}(x)(i\gamma^\mu D_\mu - m)\psi(x) - \frac{1}{4}G_a^{\mu\nu}G_{\mu\nu}^a \quad (1.18)$$

$$= \bar{\psi}(x)(i\gamma^\mu \partial_\mu - m)\psi(x) - g_s \bar{\psi}(x)\gamma^\mu T_a \psi(x)G_\mu^a - \frac{1}{4}G_a^{\mu\nu}G_{\mu\nu}^a, \quad (1.19)$$

où les $G_{\mu\nu}^a$ sont les champs de gluons et g_s la constante de couplage forte. La dérivée covariante est ainsi définie par :

$$D_\mu = \partial_\mu + ig_s \lambda_a G_\mu^a \quad (1.20)$$

et les tenseurs $G_{\mu\nu}^a$ sont définis par :

$$G_{\mu\nu}^a = D_\mu G_\nu^a - D_\nu G_\mu^a. \quad (1.21)$$

De manière analogue à la QED , le premier terme de la densité lagrangienne 1.19 correspond au terme de propagation d'un fermion et le second terme correspond au couplage entre un fermion et les bosons d'interaction : les gluons. Le dernier terme est lié aux gluons libres et n'a pas d'équivalent QED . En effet il inclut un terme de propagation de gluons (par analogie à la QED), mais aussi un couplage entre les gluons dû au caractère non-abélien de la symétrie de jauge $SU(3)_c$.

1.3.1.3 La constante de couplage forte

A cause du processus de renormalisation, les constantes de couplages dépendent de l'énergie mise en jeu. L'évolution de la constante de couplage forte en fonction de l'énergie transférée Q et de l'échelle d'énergie μ est donnée par la formule :

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu^2)}{1 + \frac{\alpha_s(\mu^2)}{4\pi} b_0 \ln \frac{Q^2}{\mu^2}}, \quad (1.22)$$

avec

$$\alpha_s = \frac{g_s}{4\pi} \quad (1.23)$$

$$b_0 = -\frac{2}{3}N_f + \frac{11}{3}N_c, \quad (1.24)$$

où N_f est le nombre de saveurs et N_c le nombre de couleurs qui valent 6 et 3 respectivement dans le Modèle Standard. Dû à la grande masse du quark top, il n'intervient cependant pas dans le calcul de α_s à l'échelle du Modèle Standard. Le nombre de saveur "active" est donc de 5. On peut déduire de l'équation 1.22 les comportements à courte et à longue distance de l'interaction forte.

- A courte distance (pour des valeurs de Q^2 élevées), α_s tend vers 0. L'intensité de la force forte diminue donc lorsque la distance entre deux quarks diminue. Deux quarks infiniment proches sont alors libres, on parle de **liberté asymptotique**.
- Lorsque la distance entre deux quarks est élevée (pour des valeurs faibles de Q^2), la constante de couplage α_s peut devenir très élevée et la théorie perd son caractère perturbatif. L'amplitude de l'interaction forte augmente alors de manière importante. Ce phénomène s'appelle **le confinement**. Il faut introduire une grande énergie pour "éloigner" deux quarks. Lorsque la quantité d'énergie injectée est suffisante, une nouvelle paire de quark anti-quark et deux nouveaux hadrons sont produits. Ce processus, qui peut engendrer des hadrons en cascade, est à l'origine des gerbes hadroniques et des jets de particules qui sont observés dans les détecteurs.

1.3.1.4 Production d'événements QCD au LHC

Les diagrammes de production de paires de quark anti-quark au niveau de l'arbre sont représentés sur la figure 1.2 dans le cas de collisions proton-proton.

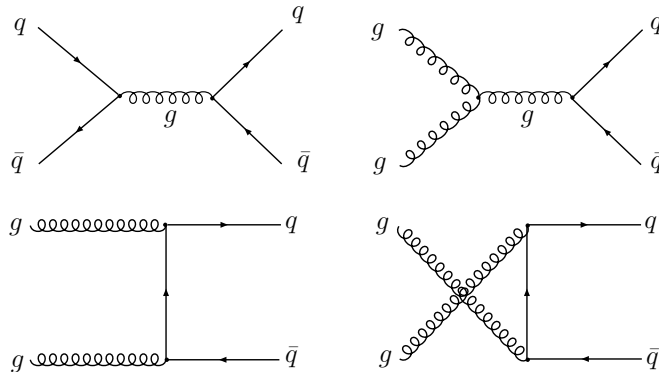


FIG. 1.2: Exemples de diagrammes de production d'un paire quark anti-quark au niveau de l'arbre dans des collisions proton-proton par annihilation quark anti-quark (en haut à gauche) et par fusion de gluons.

Du fait des grandes énergies mises en jeu dans les collisions du LHC, les sections efficaces de production des processus QCD sont très élevées. Ces événements constituent un bruit de fond hadronique important qu'il faudra prendre en compte dans les analyses de physique. Les sections efficaces de production d'événements QCD au LO (calculé avec le générateur PYTHIA) pour différents intervalles en \hat{p}_T sont présentées dans le tableau 5.6, chapitre 5. Le \hat{p}_T est défini comme l'impulsion transverse des partons produits dans le processus dur.

1.3.2 Le modèle électrofaible

Dans la section 1.2, il a été montré comment une théorie de jauge permet de décrire l'interaction électromagnétique. Dans le Modèle Standard, cette dernière interaction est unifiée avec l'interaction nucléaire faible. Ce sont T.D. Lee et C.N. Yang qui les premiers proposèrent la violation de la parité dans l'interaction faibles [18]. Cette hypothèse fut confirmée expérimentalement par C.S. Wu et ses collaborateurs en 1957 [19]. L'interaction faible se comporte donc différemment pour les composantes de chiralité gauche et droite d'un champ fermionique ψ . Les projections de chiralité gauche et droite sont définies par :

$$\psi(x) = \psi_L(x) + \psi_R(x), \text{ avec } \psi_L(x) = \frac{1 - \gamma^5}{2}\psi(x) \text{ et } \psi_R(x) = \frac{1 + \gamma^5}{2}\psi(x) \quad (1.25)$$

L'interaction faible chargée dans le modèle standard viole de manière maximale la parité, les courants faibles chargés ne se couplent en effet qu'avec les fermions de chiralité gauche (ou les anti-fermions de chiralité droite). D'après le projecteur de chiralité gauche $1 - \gamma^5$, on dit que l'interaction faible a une nature V (vecteur) – A (vecteur axial).

On peut alors introduire une nouvelle symétrie de jauge locale associée à une quantité conservée : l'isospin faible, basée sur la symétrie $SU(2)$. Les particules peuvent ainsi être classées par chiralité : les particules de chiralité gauche sont regroupées en doublets et les particules de chiralité droite en singulets, comme il est présenté dans le tableau 1.1.

Chaque doublet possède ainsi un isospin $T = 1/2$ et chaque champ correspond à un projecteur d'isospin $T_3 = \pm 1/2$. Les champs droits étant des singulets de cette symétrie, ils possèdent un isospin nul.

La théorie électrofaible est construite en introduisant une quantité conservée supplémentaire qui lie la charge électrique et l'isospin : l'*hypercharge* définie par

$$Y = 2(Q - T_3). \quad (1.26)$$

La théorie de jauge des interactions faibles est ainsi construite sur l'invariance du Lagrangien sous une transformation locale du groupe de jauge $SU(2)_L \times U(1)_Y$ où L se réfère au caractère gaucher de l'interaction faible et Y à l'hypercharge (Glashow [8], Salam [9] et Weinberg [10]). Le groupe $SU(2)_L$ est associé à trois générateurs T_a (représentés par les matrices de Pauli) correspondant à trois bosons de jauge supplémentaires

Particules	Y	T	T_3	q
$(e, \mu\tau)_L$	-1	1/2	+1/2	-1
$(\nu_{e,\mu,\tau})_L$	-1	1/2	+1/2	0
$(e, \mu\tau)_R$	-2	0	0	-1
$(\nu_{e,\mu,\tau})_R$	0	0	0	0
$(u, c, t)_L$	+1/3	1/2	+1/2	+2/3
$(d, s, b)_L$	+1/3	1/2	-1/2	-1/3
$(u, c, t)_R$	+4/3	0	0	+2/3
$(d, s, b)_R$	-2/3	0	0	-1/3

TAB. 1.3: Table des différents nombres quantiques associés aux particules de matière : l'hypercharge (Y), l'isospin (T) et sa composante (T_3) et la charge électrique q .

W_μ^a (pour a allant de 1 à 3). Le boson associé à la symétrie $U(1)_Y$ est noté B_μ . Le Lagrangien électrofaible est ainsi composé de différents termes correspondant aux doublets et singulets de quarks et de leptons. Prenons l'exemple de la troisième famille de quarks, qui est décrite par un doublet gauche et deux singulets droits :

$$Q_L = \begin{pmatrix} t \\ b \end{pmatrix}_L, \quad b_R, \quad t_R. \quad (1.27)$$

Les termes correspondants du Lagrangien électrofaible s'écrivent :

$$\mathcal{L}_D^{EW} = \bar{Q}_L(i\gamma^\mu D_\mu)Q_L + \bar{t}_R(i\gamma^\mu D_\mu)t_R + \bar{b}_R(i\gamma^\mu D_\mu)b_R \quad (1.28)$$

$$- \frac{1}{4}W_{\mu\nu}^a W_a^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}. \quad (1.29)$$

avec

$$D_\mu = \partial_\mu + igT_a W_\mu^a + i\frac{g'}{2}Y B_\mu, \quad (1.30)$$

où g et g' sont les constantes de couplage de $SU(2)_L$ et $U(1)_Y$. Les tenseurs $W_a^{\mu\nu}$ et $B_{\mu\nu}$ sont définis par analogie avec la QED (équation 1.15) et la QCD (équation 1.21).

Après brisure de symétrie (cf. section 1.3.4), les champs W_μ^a et B_μ ne constituent pas des champs physiques et ne décrivent pas de bosons de jauge. Les champs physiques A_μ (correspondant au photon), Z_μ , W_μ^+ et W_μ^- , correspondant aux bosons Z et W^\pm , sont décrits par des combinaisons linéaires des champs W_μ^a et B_μ :

$$A_\mu = \cos \theta_W B_\mu + \sin \theta_W W_\mu^3, \quad (1.31)$$

$$Z_\mu = -\sin \theta_W B_\mu + \cos \theta_W W_\mu^3, \quad (1.32)$$

$$W_\mu^+ = \frac{W_\mu^1 - iW_\mu^2}{\sqrt{2}}, \quad (1.33)$$

$$W_\mu^- = \frac{W_\mu^1 + iW_\mu^2}{\sqrt{2}}, \quad (1.34)$$

où θ_W est l'angle de mélange électrofaible ou angle de Weinberg. Les constantes de couplage g et g' sont reliées à θ_W et à la constante de couplage électromagnétique e par les relations :

$$\tan \theta_W = \frac{g'}{g} \text{ et } g' \cos \theta_W = e. \quad (1.35)$$

Comme les termes de masse du Lagrangien électrofaible pour les bosons d'interaction violeraient l'invariance de jauge, les bosons doivent nécessairement être non massifs. Or, l'interaction faible est une interaction de courte portée et les mesures expérimentales montrent que les bosons Z et W^\pm ont effectivement des masses qui ont été mesurées avec une très grande précision au *LEP* et au *Tevatron* [20] :

$$m_Z = 91.1876 \pm 0.0021 \text{ GeV}/c^2, \quad (1.36)$$

$$m_W = 80.425 \pm 0.038 \text{ GeV}/c^2. \quad (1.37)$$

Le mécanisme mathématique permettant d'introduire des masses aux bosons de jauge électrofaibles est basé sur la *brisure spontanée de la symétrie électrofaible* en présence d'un champ scalaire : le champ de Higgs. Ce mécanisme sera présenté dans la section 1.3.4.

1.3.3 Production de bosons W et Z dans les collisions proton-proton

Les diagrammes de production des bosons Z et W dans des collisions proton-proton sont présentés au niveau de l'arbre sur la figure 1.3.

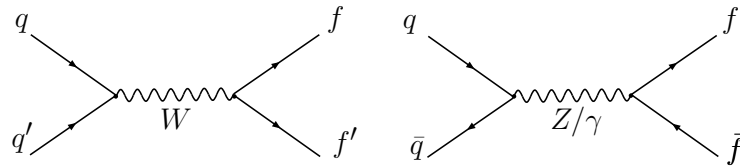


FIG. 1.3: Diagrammes de production de boson W à gauche et de boson Z à droite, au niveau de l'arbre dans des collisions proton-proton, où f se réfère aux fermions.

Le boson W possède deux modes de désintégration : le mode leptonique, lorsqu'il se désintègre en lepton (électron, muon ou tau) et en anti-neutrino associé; le mode hadronique, lorsqu'il se désintègre en paire de quarks constituant un doublet d'isospin. Le boson Z se désintègre quant à lui en paire fermion anti-fermion de chiralité gauche et droite respectivement.

Ces bosons peuvent également être produits en association avec des gluons ou des quarks. On parlera alors d'événements W +jets et Z +jets. Il existe également des diagrammes conduisant à la production de deux bosons faibles au sein du même événement. Ces événements sont appelés événements di-bosons et les diagrammes de productions correspondants sont présentés sur la figure 1.4.

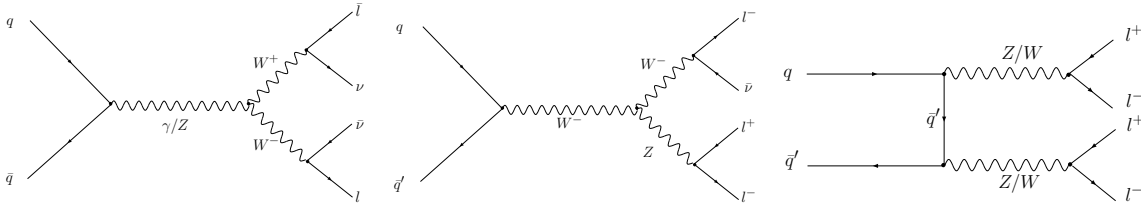


FIG. 1.4: Diagrammes de production d'événements di-bosons WW (en haut à gauche), WZ (en haut à droite) et ZZ (en bas), au niveau de l'arbre dans des collisions proton-proton.

1.3.4 Brisure de symétrie et modèle de Higgs

Comme nous l'avons vu dans la section 1.3.2, la théorie électrofaible n'est valide que si les masses des particules sont nulles si l'on veut préserver l'invariance de jauge $SU(2)_L \times U(1)_Y$. Un mécanisme de brisure spontanée de la symétrie électrofaible a été proposé par Higgs [21, 22, 23] et par Brout et Englert [24] pour générer la masse des bosons de jauge W et Z . Ce mécanisme est appelé communément mécanisme de Higgs. Il s'agit d'introduire dans le Lagrangien électrofaible un champ scalaire complexe iso-spineur de charge électrique nulle :

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \varphi_1 + i\varphi_2 \\ \varphi_3 + i\varphi_4 \end{pmatrix}$$

En plus du terme dynamique du champ de Higgs, un potentiel supplémentaire est introduit : le potentiel de Higgs. Le Lagrangien de Higgs correspondant s'écrit :

$$\mathcal{L}_{Higgs} = D_\mu \phi D^\mu \phi^\dagger - V(\phi) \quad (1.38)$$

$$D_\mu \phi D^\mu \phi^\dagger + \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2. \quad (1.39)$$

Dans le cas où $\mu^2 > 0$ et $\lambda < 0$, le Lagrangien 1.39 décrit un champ scalaire massif et l'état fondamental correspond au cas $\phi = 0$. Dans le cas où $\mu^2 > 0$ et $\lambda > 0$, il existe une

infinité de minima dégénérés pour le potentiel. Ces minima sont atteints, à une phase près, pour :

$$\phi^\dagger\phi = \frac{-\mu^2}{2\lambda} = \frac{v^2}{2}, \quad (1.40)$$

où $v = \sqrt{\mu^2/2}$. Cette forme du potentiel de Higgs est illustrée sur la figure 1.5.

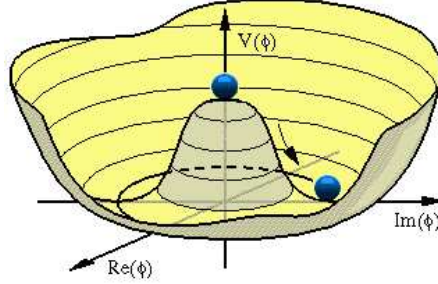


FIG. 1.5: Schéma illustrant le potentiel de Higgs et le mécanisme de brisure de symétrie.

On peut alors briser la symétrie en choisissant un état fondamental tel que :

$$\phi(0) = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix}.$$

A partir de cet état fondamental et pour de petites perturbations, il est possible de développer le champ ϕ en série de Taylor autour du minimum $\phi(0)$:

$$\phi(x) = \frac{\eta_a(x)T^a}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}.$$

Le théorème de Goldstone indique que si un groupe de symétrie interne est brisé spontanément, des bosons non physiques, appelés *bosons de Goldstone* sont créés. Dans notre cas il s'agit des champs $\eta_a(x)$. A l'aide d'une transformation de jauge, il est possible de faire disparaître ces termes. Le seul champ restant est le champ de Higgs $H(x)$. Les termes de couplages entre le boson de Higgs et les bosons de jauge proviennent des dérivées covariantes présentes dans le Lagrangien de Higgs (1.39) et les termes proportionnels à v correspondant aux termes de masses usuels $m_W^2 W_\mu^+ W^{-\mu}$ et $m_Z^2 Z_\mu Z^\mu$. Les masses des particules sont alors des masses effectives provenant de la présence du potentiel de Higgs. On peut également remarquer que le boson de Higgs n'a pas de charge électrique. Il ne se couple donc pas avec le photon, ce dernier est donc de masse nulle.

Les masses des bosons W^\pm , Z et du bosons de Higgs sont finalement :

$$m_W = \frac{gv}{2} \quad (1.41)$$

$$m_Z = \frac{v}{2} \sqrt{g^2 + g'^2} \quad (1.42)$$

$$m_H = \sqrt{2\lambda}v. \quad (1.43)$$

La valeur dans le vide du champ de Higgs peut être mesurée expérimentalement à travers la constante de Fermi par la relation :

$$\frac{G_F}{\sqrt{2}} = \frac{1}{2v^2}, \quad (1.44)$$

ce qui donne une valeur dans le vide du champ de Higgs de $v \approx 246 \text{ GeV}/c^2$.

Il reste à introduire les termes de masse pour les fermions, il faut donc ajouter des couplages entre fermions et le boson de Higgs dans le Lagrangien (1.39) Ces termes de couplages sont appelés couplages de Yukawa et sont de la forme :

$$\mathcal{L} = -g_Y(\bar{f}_L\phi f_R + \bar{f}_R\phi^\dagger f_L), \quad (1.45)$$

où f_L est un doublet de fermions gauches et f_R est un singulet droit. Les g_Y sont les couplages de Yukawa des quarks et des leptons. Au même titre que les constantes de couplage et que la valeur dans le vide du champs de Higgs, ce sont des paramètres libres de la théorie. Il est à noter que les neutrinos dans le Modèle Standard sont sans masse et n'ont pas de singulet droit, il n'existe pas de constante de Yukawa pour ces fermions. De plus, les états propres de masse des quarks ne sont pas des états propres de $SU(2)_L$. Les états propres d'interaction faible sont en fait des combinaisons linéaires des états propres de masse. Il faut donc introduire une matrice (unitaire) de mélange correspondant au changement de la base des états propre d'interaction vers la base des états propres de masse. Cette matrice est appelé matrice de Cabbibo-Kobayashi-Maskawa ou matrice CKM [25],[26] :

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}$$

où le vecteur de quarks *down* à gauche de l'égalité correspond aux états propres d'interaction et le vecteur à droite de la matrice CKM aux états propre de masse. Il est à noter que les éléments diagonaux de cette matrice ont une norme proche de 1, ce qui favorise les changements de saveur au sein d'une même famille.

1.4 Au-delà du Modèle Standard

Le Modèle Standard de la physique des particules, qui regroupe les interaction électrofaible et forte, a été vérifié notamment au LEP et au Tevatron avec une remarquable

précision. Pour différentes raisons qui seront discutées ici, le Modèle Standard reste cependant incomplet. Il n'est qu'une théorie effective, valide aux énergies accessibles actuellement. Il correspondrait à l'approximation à basse énergie d'une théorie plus générale dont les effets seraient perceptibles à plus haute énergie. Les éléments manquant du Modèle Standard sont :

- **La gravitation** : L'interaction gravitationnelle n'est pas prise en compte dans le Modèle Standard (MS), elle est supposée très faible (40 ordres de grandeur de différence en amplitude avec l'interaction électromagnétique) et est négligée. Cependant, pour des échelles d'énergie proches de l'échelle de Planck ($M_P = 2.4 \times 10^{18} \text{GeV}/c^2$), cette interaction n'est plus négligeable. Les difficultés surviennent lorsque l'on tente de construire une théorie de jauge de l'interaction gravitationnelle. Une telle théorie n'est pas renormalisable, il est donc impossible de faire disparaître les divergences dans les calculs de diagrammes de Feynman et la théorie n'a aucun pouvoir prédictif.
- **La masse des neutrinos** : L'observation de l'oscillation des neutrinos prouve que ces dernières particules sont massives. Or, dans le MS, les masses de ces particules sont nulles mais peuvent cependant être inclus en élargissant le MS.
- **La matière noire** : L'étude de la dynamique des galaxies montre que la majeure partie de la matière gravitationnelle n'est pas constituée de matière visible. L'hypothèse selon laquelle cette matière "noire" serait constituée de matière hadronique ordinaire n'est pas en bon accord avec le Modèle Standard cosmologique et les mesures expérimentales. Elle serait alors constituée d'un nouveau type de particules stables, de grande masse, interagissant faiblement et principalement par interaction gravitationnelle (WIMP).
- **L'énergie noire** : L'étude des supernovæ de type Ia a montré que l'univers est en expansion accélérée. La source d'énergie de cette expansion doit encore être déterminée. Elle constituerait environ 70% du contenu en énergie de l'univers.

La matière hadronique connue, décrite par le MS, ne représente ainsi que 5% seulement de l'énergie globale de l'univers. De plus, le MS possède une part d'arbitraire et est, sous certains aspects, intellectuellement insatisfaisant :

- **Convergence des constantes de couplage** : La section 1.3.1.3 introduit comment la constante de couplage forte évolue avec l'énergie. De manière analogue, les constantes de couplage faible et électromagnétique évoluent avec l'énergie. Ces évolutions sont illustrées sur la figure 1.6. On remarque qu'à très haute énergie, les constantes de couplage semblent converger vers une même valeur mais l'unification est imparfaite.
- **Corrections radiatives à la masse du Higgs, problème de hiérarchie** : A très haute énergie (masse de Planck), les corrections quantiques à la masse du boson de Higgs peuvent être très importantes. Il est nécessaire de faire un choix très précis (*fine-tuning* en anglais) du paramètre Λ , échelle d'une possible nouvelle physique, afin de garder la masse du Higgs finie. Mais ce choix n'est pas "naturel" et peu satisfaisant.

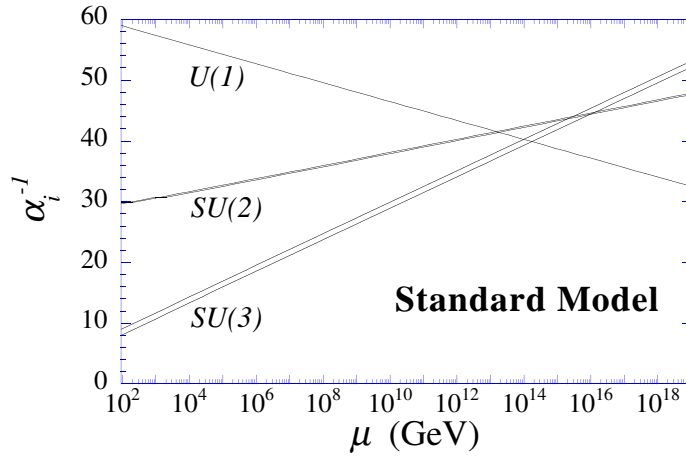


FIG. 1.6: Evolution des inverses de constante de couplage électromagnétique ($U(1)$), faible ($SU(2)$) et forte ($SU(3)$) en fonction de l'énergie.

- **Nombre de paramètres libres** : Le MS comporte au total 19 paramètres libres : les 3 constantes de couplage des interactions forte, faible et électromagnétique, les 9 masses des fermions chargés, les 4 paramètres de la matrice CKM, θ_{QCD} et les deux paramètres du modèle de Higgs. Ce nombre est étendu à 25 s'il on tient compte de la masse de neutrinos. Ces paramètres sont déduits d'observations expérimentales et ne sont pas prédits par la théorie.
- **Trois familles de particules** : Le fait que 3 générations de particules seulement existent n'est pas expliqué par le MS.
- **La hiérarchie des masses** : Les différences de masses entre les particules et entre les générations de particules ne sont pas non plus expliquées. En particulier, la masse du quark top est significativement plus grande que la masse des autres fermions et des bosons W et Z .

Au vu de ces différents arguments, le Modèle Standard est sans doute une théorie effective à basse énergie d'une théorie plus globale. De nombreuses théories possibles pour expliquer la physique au-delà du Modèle Standard sont actuellement à l'étude dans les expériences au Tevatron et seront activement explorées au LHC. L'une des théories reconnues comme prometteuse est la *Supersymétrie* (ou Susy).

L'idée de base est de régler le problème de hiérarchie en introduisant de nouveaux bosons et fermions dont la présence dans le calcul des corrections radiatives à la masse du boson de Higgs compenserait les termes divergents. Il faut alors introduire une symétrie supplémentaire (la *Supersymétrie*) qui permet de transformer un fermion en boson et inversement [27]. Ainsi pour chaque fermion du Modèle Standard il existerait un boson. Comme de telles particules n'ont pas été observées jusqu'à présent, les masses des particules Supersymétriques doivent être supérieures à celles des particules du Modèle Standard et de l'ordre du TeV .

De plus, si R est conservé, la Susy prédit l'existence d'une particule de charge électrique nulle, interagissant faiblement et possédant une masse élevée, ce qui en fait un bon candidat pour expliquer la matière noire. Finalement, dans les théories Supersymétriques les constantes de couplage convergent vers une même valeur à haute énergie.

D'autres modèles (comme les modèles à dimensions supplémentaires, les modèles de Technicouleur, les modèles de grande unification) permettent également de résoudre un certain nombre des problèmes du Modèle Standard. Comme en Susy, des nouvelles particules sont prédites à plus haute énergie. Au LHC, une nouvelle gamme d'énergie devient accessible ce qui pourrait conduire à la découverte d'au moins une de ces nouvelles particules.

1.5 Physique du quark top

1.5.1 Propriétés du quark top

Le quark top a été découvert au Tevatron en 1995 par les expériences CDF et DØ [15]. Il possède une masse très élevée de $172.4 \pm 1.2 \text{ GeV}/c^2$, supérieure aux masses des bosons W et Z , ce qui en fait la particule la plus lourde du Modèle Standard. La masse du top est mesurée par les expériences CDF et DØ [20] pour différents canaux de désintégration (cf figure 1.7).

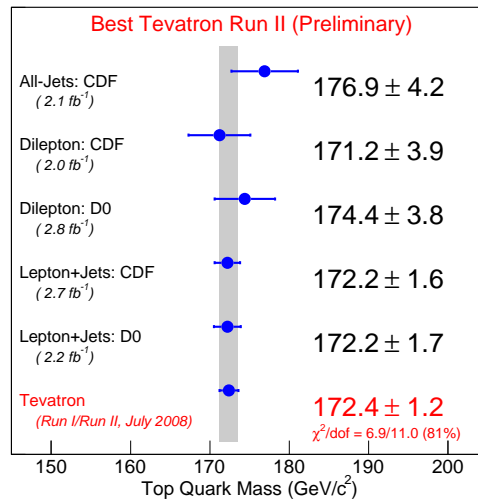


FIG. 1.7: Mesure de la masse du quark top dans les expériences CDF et DØ du Tevatron à partir de différents canaux de désintégration. La valeur combinée est de $172.4 \pm 1.2 \text{ GeV}/c^2$.

1.5.1.1 Masse du top et contraintes sur la masse du boson de Higgs

Du fait de cette masse très grande, le couplage entre le boson de Higgs et le quark top est importante. Le couplage de Yukawa du quark top est ainsi proche de l'unité. De ce fait, le quark top intervient de manière significative (quadratiquement) dans les corrections radiatives des propagateurs des bosons W et Z , comme l'illustre le diagramme de gauche de la figure 1.8. Le boson de Higgs intervient également (logarithmiquement) dans le calcul de ces mêmes corrections (diagrammes de droite de la figure 1.8). Il est donc possible d'imposer des limites sur la masse du boson de Higgs en connaissant avec précision la masse du quark top et des bosons W ou Z .

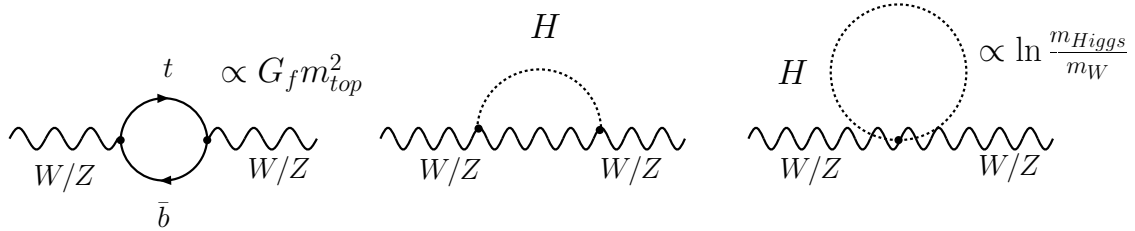


FIG. 1.8: A gauche, contribution du quark top aux corrections radiatives des propagateurs des bosons W et Z . Au milieu et à droite, contributions du boson de Higgs aux corrections radiatives des propagateurs des bosons W et Z .

La figure 1.9 résume les limites sur la masse du Higgs ainsi obtenue à partir des mesures de précision des masses du top et du boson W au LEP et au Tevatron notamment [28, 29]. On remarque très clairement que la masse du boson de Higgs serait plutôt légère. La recherche directe du boson de Higgs à LEP2 à permis d'imposer une limite inférieure à sa masse : $m_{Higgs} > 114 \text{ GeV}/c^2$. De récents résultats du Tevatron ont exclu une masse de $170 \text{ GeV}/c^2$ avec 95% de niveau de confiance [29].

1.5.1.2 Désintégration du quark top

Une autre propriété importante du quark top est que, à la différence des autres quarks, son temps de vie (de l'ordre de 10^{-24} ps) est 10 fois plus faible que le temps typique que met un quark pour s'hadroniser. De plus, le module de l'élément de matrice CKM V_{tb} est très proche de 1 ($0.9990 < |V_{tb}| < 0.9992$ [30]). Les produits de désintégration du quark top sont donc presque exclusivement constitués d'un boson W réel (car $m_t > m_W$) et d'un quark b comme l'illustre le diagramme de désintégration suivant :

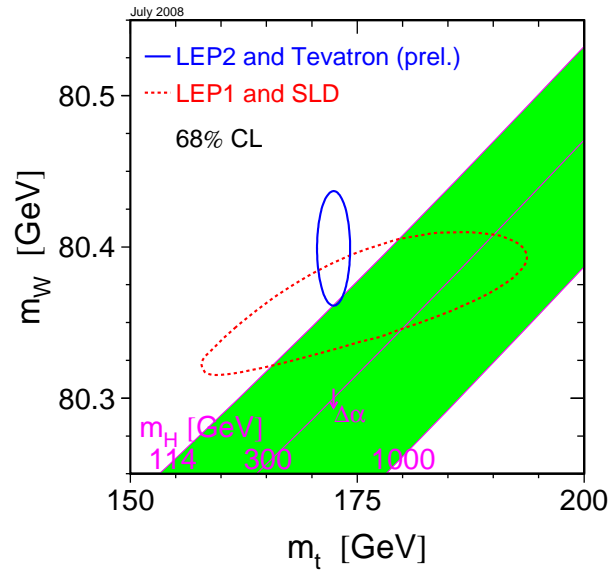
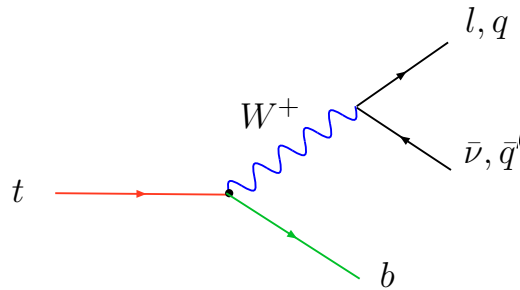


FIG. 1.9: Contrainte de la masse du boson W (axe des ordonnées) et de la masse du quark top (axe des abscisses) sur la masse du boson de Higgs. Le contour bleu correspond à la région de masse du Higgs permise avec un niveau de confiance de 68% et calculée à partir des mesures directes de la masse du W et du top de LEP2 et du Tevatron. Le contour rouge correspond au résultat combiné de LEP1 et de SLD. La bande verte correspond à la région de masse du Higgs permise par le Modèle Standard ($< 1 \text{ TeV}/c^2$) et non exclue par les mesures directes au LEP ($> 114 \text{ GeV}/c^2$) [20].



Comme le top se désintègre avant de s'hadroniser, il est possible d'étudier directement le spin du top en étudiant ses produits de désintégration. Par exemple, l'étude de la polarisation du boson W dans la désintégration du top permet d'étudier directement le vertex tWb et la structure $V - A$ du Modèle Standard. Cet aspect est tout particulièrement intéressant pour la recherche de nouvelle physique, telle que des contributions $V + A$ ou la présence d'un boson de Higgs chargé.

1.5.2 Modes de production

Plusieurs modes de production de quark top sont présents aux collisionneurs hadroniques : par interaction forte (production par paire $t\bar{t}$) et par interaction électrofaible (produisant un seul quark top appelé *single top*).

1.5.2.1 Production par interaction forte

La production de paire de quarks top $t\bar{t}$ est le mode de production principal de quarks top au LHC. Au niveau de l'arbre, quatre diagrammes principaux interviennent : contrairement au Tevatron, au LHC la majorité des paires de quark top, 85% environ, sont produits par fusion de gluons (diagrammes en haut à droite et en bas de la figure 1.10) alors que 15% sont produits par annihilation de quark-antiquark (diagramme en haut à gauche de la figure 1.10). En effet, la fraction de gluon contenue dans un proton augmente avec l'énergie et la fraction d'antiquark est plus faible et provient uniquement de la mer.

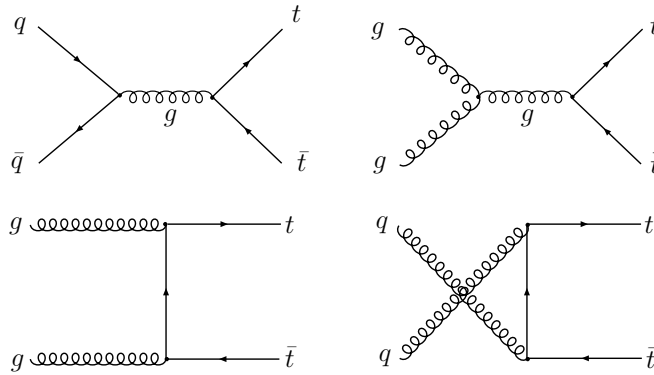


FIG. 1.10: Diagrammes de production de paire de quarks top au niveau de l'arbre dans les collisions pp au LHC.

La section efficace totale de production de paire de quark top au Tevatron est de **6.97 pb** pour des collisions de 1.96 TeV (en prenant en compte la resommation de termes logarithmes NLL [31] correspondant au calcul de radiations molles et colinéaire) au NLO. Cette section efficace est bien supérieure au LHC, où elle est de **833 pb** au NLO+NLL pour des collisions de 14 TeV [31]. Cette valeur sera utilisée dans la suite de cette thèse.

1.5.2.2 Production par interaction faible

Le quark top peut également être produit par voie électrofaible, qui conduit à la génération d'un seul quark top (*single top*). Les trois canaux principaux sont présentés sur la figure 1.11) : la voie s conduisant à un quark top et un anti-quark b dans l'état

final , la voie tW conduisant à un quark top et un boson W et la voie t conduisant à un quark top, un anti-quark b et un quark additionnel.

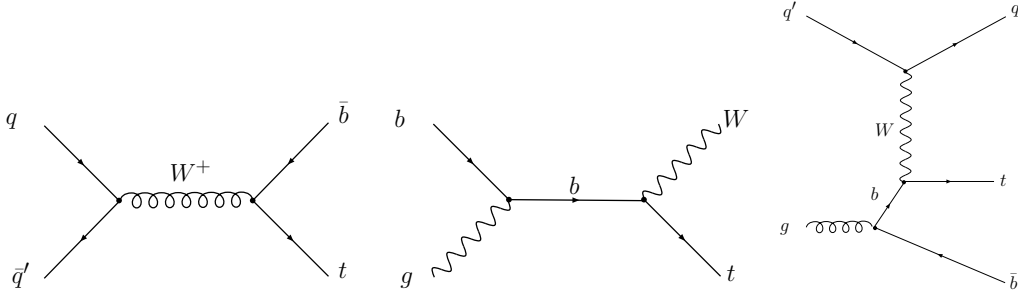


FIG. 1.11: Canaux de production de single top pour la voie s (à gauche), la voie tW (au milieu) et la voie t (à droite)

Les sections efficaces de production de single top au NLO sont résumées dans le tableau 1.4. On remarque que pour les voies s et t les sections efficaces top et anti-top sont différentes dans les collisions pp pour des raisons de conservation de charge.

section efficace (pb)	$p\bar{p}$ à 2 TeV	pp à 14 TeV
voie s (t)	0.4	6.1
voie s (\bar{t})	0.4	3.8
voie t (t)	0.9	150.0
voie t (\bar{t})	0.9	88.5
voie tW (t)	0.1	32.1
voie tW (\bar{t})	0.1	32.1

TAB. 1.4: Sections efficaces de production de single top au Tevatron (collisions $p\bar{p}$ à 2 TeV) et au LHC (collisions pp à 14 TeV) [32, 33].

1.5.3 Canaux de désintégration

Dans les événements $t\bar{t}$, différents canaux de désintégration peuvent être étudiés. Ils correspondent au mode de désintégration du boson W (hadronique ou leptonique) issus de la désintégration des quarks top et anti-top.

- **Le canal di-lepton** : Il correspond au cas où chacun des deux bosons W issus de la désintégration du t ou \bar{t} se désintègre en un lepton (e, μ ou τ) et son neutrino associé (ν_e, ν_μ et ν_τ). L'état final contient donc deux leptons isolés de grande impulsion transverse p_T , deux neutrinos générant une grande énergie transverse manquante et deux jets de particules issus de l'hadronisation de deux quarks b . Ce canal sera étudié en détail dans le chapitre 5. Dû aux rapports d'embranchement $Br(W \rightarrow l\bar{\nu}_l)$ relativement faibles, la fraction d'événements di-leptons est de l'ordre de 10%. Les

bruits de fond associés à ce canal sont très faibles, mais la statistique des événements reste limitée.

- **Le canal lepton+jets** : Dans ce canal, un boson W se désintègre leptoniquement et l'autre se désintègre en une paire de quarks. L'état final contient donc un lepton isolé à grand p_T , de l'énergie transverse manquante provenant de la présence d'un neutrino et quatre jets dont deux issus de l'hadronisation de quarks b . Les fractions d'événements sont de 15% environ pour chaque canal leptonique e +jets, μ +jets et τ +jets. Les canaux électron+jets et muon+jets sont facilement identifiables et les bruits de fond instrumentaux principaux peuvent être rejetés en sélectionnant un lepton de grande impulsion transverse.
- **Le canal hadronique** : Il correspond au cas où les deux bosons W se désintègrent hadroniquement. La fraction d'événements correspondante est de 45%. Bien qu'il s'agisse du canal principal, il est assez difficile à étudier car le bruit de fond hadronique correspondant est difficile à rejeter.

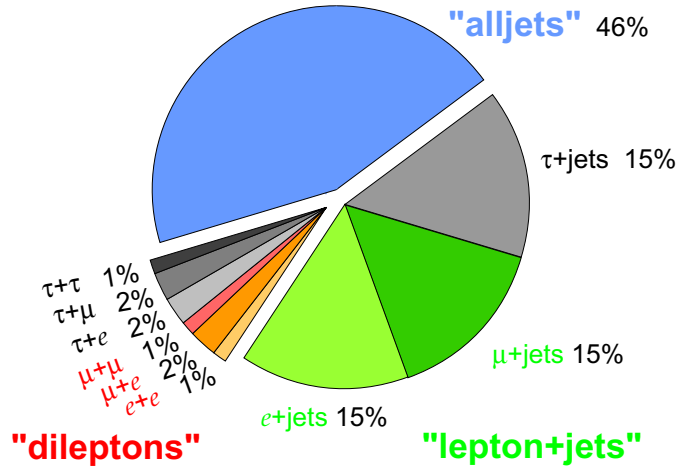


FIG. 1.12: Canaux de désintégration des paires de quarks top $t\bar{t}$ et leurs rapports d'embranchement.

Les différentes fractions relatives des canaux de désintégration de paires de quark top sont représentées sur la figure 1.12.

1.5.4 Motivation pour l'étude du top au LHC

La physique du top au LHC sera essentielle pour comprendre le détecteur mais aussi pour préparer la recherche de nouvelle physique.

Les événements $t\bar{t}$ seront produits en très grand nombre au LHC. Ils permettront ainsi de calibrer en énergie les jets de quarks issus de bosons W [34] et les jets b , mais

également de mesurer l'efficacité des algorithmes d'identification des jets b [35, 36, 37]. De ce fait, ils constitueront un nouvel outil qui permettra de comprendre et de maîtriser le détecteur.

Dans le cadre du Modèle Standard, nous avons vu que les caractéristiques particulières du quark top (grande masse, très faible temps de désintégration) permettent d'explorer le couplage $V - A$ et le caractère "gaucher" de l'interaction faible (cf. section 1.3.2). Il a également été mentionné que la mesure de précision de la masse du quark top permet de contraindre la masse du boson de Higgs. La très grande quantité d'événements produite par le LHC devrait nous permettre de minimiser les erreurs statistiques de ces mesures dans un premier temps, puis de mieux contrôler les erreurs systématiques. Les nombres d'événements $t\bar{t}$ produits par jour au LHC est résumé dans le tableau 1.5. Après quelques années de fonctionnement, il devrait ainsi être possible d'effectuer des mesures de précision des caractéristiques du quark top au LHC. De plus, les événements $t\bar{t}$ présentent les ingrédients propres à de nombreux canaux de nouvelle physique : la présence d'un ou plusieurs leptons chargés émis à grand p_T , une énergie transverse manquante élevée, des jets de grand p_T et éventuellement des jets b . Il sera donc essentiel de bien comprendre ces événements afin de préparer au mieux la recherche directe de nouvelle physique.

Energie dans le centre de masse	$\mathcal{L} = 10^{32} \text{cm}^{-2} \text{s}^{-1}$
10 TeV ($\sigma_{t\bar{t}} = 375 \text{ pb}$)	≈ 300
14 TeV ($\sigma_{t\bar{t}} = 833 \text{ pb}$)	≈ 720

TAB. 1.5: Nombre d'événements $t\bar{t}$ produits par jour au LHC pour différentes énergies de collision et pour différentes luminosités instantanées.

Les quarks top peuvent également être utilisés pour la recherche de nouvelle physique en vérifiant que les mesures de précision de ses propriétés sont en bon accord avec les prédictions du Modèle Standard. Par exemple, l'existence d'un boson de Higgs chargé *Supersymétrique* avec $m_{H^+} < m_t$ conduirait à un nouveau canal de désintégration du top $t \rightarrow H^+ b$ où le Higgs chargé se désintègre préférentiellement en lepton τ et neutrino associé [38]. Cela se traduirait ainsi par un déficit d'événements top contenant un lepton e ou μ ce qui conduit à une valeur de la section efficace de production de paire de quark top plus faible que les prédictions du Modèle Standard dans les canaux contenant des électrons et des muons. En contrepartie, les canaux contenant des leptons τ présenteraient un excès d'événements. Le canal di-lepton est tout particulièrement sensible à ce type de processus.

De plus, des quarks top peuvent être présents dans les produits de désintégration de particules non standard de haute masse, telle que une résonance $Z' \rightarrow t\bar{t}$ [39] ou un boson de Higgs Supersymétrique chargé $H^+ \rightarrow t\bar{b}$ lorsque $m_{H^+} > m_t$ [40].

1.6 Génération des événements

Dans les analyses de physique, il est fondamental de pouvoir comparer une mesure avec sa prédiction théorique. En physique des particules, les prédictions théoriques peuvent être déterminées à l'aide de la génération "artificielle" d'événements à l'aide des méthodes *Monte-Carlo* [41], en utilisant des procédés aléatoires (tirages aléatoires) et probabilistes pour calculer une valeur numérique.

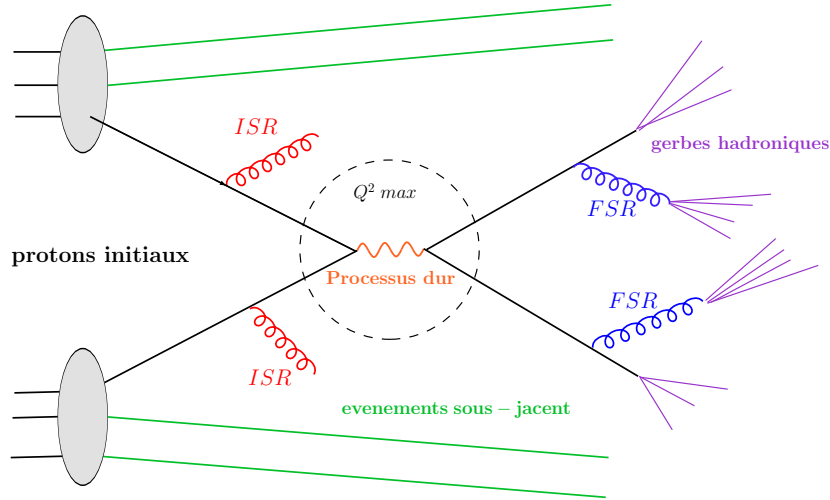


FIG. 1.13: Schéma représentant la modélisation d'une collision proton-proton par les générateurs d'événements.

Dans notre cas, les techniques de génération sont utilisées pour reproduire des événements créés lors de collisions proton-proton. La production des particules finales lors de la génération d'événements se décompose en plusieurs étapes présentées sur la figure 1.13 et qui vont être discutées.

- **Description des protons et des processus durs :** le collisionneur LHC (cf. section 2.1) est un collisionneur accélérant des faisceaux de protons jusqu'à une énergie de 7 TeV, correspondant à des collisions de 14 TeV dans le centre de masse. Grâce à ces hautes énergies, les constituants même des protons interagissent dans les collisions : les quarks et les gluons, également appelés partons. Deux types de quarks sont à considérer : les quarks *de valence* qui constituent le proton et les quarks *de la mer* issus de la polarisation du vide. Ainsi, pour deux partons impliqués dans une interaction dure et qui portent chacun une fraction x_1 et x_2 de l'impulsion totale de leur proton, la section efficace totale d'interaction entre deux protons est donnée par :

$$\sigma(pp \rightarrow X) = \sum_{i,j} \int_0^1 \int_0^1 f_i^p(x_1, Q_F^2) f_j^p(x_2, Q_F^2) \hat{\sigma}_{ij \rightarrow X}(Q_F^2) dx_1 dx_2, \quad (1.46)$$

où $\hat{\sigma}(ij \rightarrow X)$ est la section efficace d'interaction des partons i et j et les $f^p(x_i, Q_F^2)$ correspondent aux distributions des fonctions de probabilité (appelées PDF) pour qu'un parton i possède une fraction x_i de l'impulsion totale d'un proton. Q_F^2 est l'impulsion qui marque la séparation entre les interactions partoniques dures et les effets non-perturbatifs dans le proton. Les PDF sont par exemple déterminées par les groupes CTEQ [42] et MRST [43], notamment à partir des mesures des expériences auprès du collisionneur électron-proton HERA.

- **Les événements sous-jacents** : ils sont principalement constitués des résidus de la collision proton-proton. Considérons l'exemple des processus $2 \rightarrow 2$ (deux partons interviennent dans le processus dur et deux nouveaux partons sont produits) dans le cas de collisions proton-proton. Lorsqu'un quark u d'un proton intervient dans le processus dur, les résidus de ce proton (les quarks ud de valence, les paires $q\bar{q}$ de la mer et les gluons) peuvent interagir avec les résidus de l'autre proton et créer un événement sous-jacent. Lorsque des gluons interviennent dans le processus dur, l'événement sous-jacent contient les interactions dues aux quarks et aux gluons résiduels. Les événements sous-jacents sont relativement difficiles à modéliser car ils font intervenir un grand nombre de processus d'une part et d'autres part car des corrélations provenant de la conservation de la couleur existent entre les partons du processus dur et les résidus de la collision. Les modèles décrivant les événements sous-jacents doivent ainsi être ajustés à l'aide des données réelles. Les événements sous-jacents sont illustrés sur la figure 1.13 par les lignes vertes.
- **Le rayonnements des états initiaux (*ISR*) et finaux (*FSR*)** : les partons mis en jeu dans le processus dur ou les éventuels quarks et gluons produits par l'interaction peuvent également émettre des gluons. On parle alors de radiations de l'état initial (avant le processus dur, *ISR*³) et de l'état final (après le processus dur *FSR*⁴). L'évolution de l'émission des gerbes partoniques est paramétrisée par la virtualité Q^2 du parton (où Q^2 est l'impulsion transférée aux partons). Lorsque la virtualité atteint un seuil Q_{max}^2 , la gerbe s'arrête et le processus d'interaction dure commence. Les radiations *ISR* et *FSR* sont représentées sur le schéma de la figure 1.13 en rouge et en bleu respectivement. La coupure Q_{max}^2 est schématisée par le cercle pontillé.
- **Hadronisation** : il s'agit de la dernière étape de la génération d'événements. Les partons colorés finals se transforment pour produire des hadrons sans couleur. Ce processus, appelé hadronisation, est traité par un modèle de fragmentation de corde [44] dans le générateur PYTHIA [45]. Dans ce modèle, l'interaction entre un quark et un anti-quark est représentée par une corde de couleur qui les relie. Au fur et à

³Initial State Radiation

⁴Final State Radiation

mesure que les quarks se séparent, la corde se tend et cède, conduisant à la création d'une nouvelle paire de quark anti-quark. Cette étape permet de former des jets de particules qui vont par la suite être observés dans les détecteurs.

Différents générateurs d'événements existent pour décrire les processus durs. Ceux qui ont été utilisés dans cette thèse sont PYTHIA [45] et ALPGEN [46]. Le premier a l'avantage de décrire un très grand nombre de processus tandis que le second donne une meilleure description des multiplicités de quarks et de gluons dans les états finaux de certains processus. Dans le cas où le générateur ALPGEN est utilisé, il est interfacé avec PYTHIA qui génère les événements sous-jacents, les radiations *ISR* et *FSR* ainsi que l'hadronisation. Il est important de noter que les générateurs possèdent un grand nombre de paramètres libres qu'il faut déterminer à partir des données. Les erreurs correspondantes doivent être prises en compte dans toutes les analyses de physique.

Finalement, pour prendre en compte les effets d'acceptance, d'interaction dans la matière et de reconstruction du détecteur, une simulation numérique est utilisée. La simulation du détecteur et les algorithmes standard de reconstruction seront discutés dans le chapitre 2. Ces événements générés et simulés sont appelés événements *Monte-Carlo* et peuvent en principe être comparés aux vraies données.

Cependant, au démarrage du LHC, un bon accord entre données Monte-Carlo et données réelles n'est pas attendu, du fait de la non optimisation des paramètres de la génération et de la simulation à l'aide des données. De telles différences entre données et Monte-Carlo sont déjà observées dans les données cosmiques. Cette aspect sera discuté dans le chapitre 3.

2

Le Dispositif Expérimental

Le collisionneur LHC (*Large Hadron Collider*) est situé au laboratoire européen de recherche en physique des particules, le CERN, situé près de Genève. Il a été conçu pour produire des collisions *proton – proton* de 14 TeV à haute luminosité afin de rechercher des processus rares tels que la production de boson de Higgs Standard ou de nouvelles particules ayant des masses de l'ordre du TeV. Quatre grandes expériences ont été conçues et installées auprès du LHC : *Atlas* [47] et *CMS* [48] sont des détecteurs généralistes dédiés en particulier à la recherche du boson de Higgs et de nouvelle physique. L'expérience *LHCb* [49] étudiera la violation de *CP* dans la désintégration des hadrons beaux afin d'essayer d'expliquer la prédominance de la matière sur l'anti-matière dans l'univers. Finalement l'expérience *Alice* [50] analysera les données issues de la collision d'ions lourds relativistes afin d'étudier l'état de la matière hadronique dans des conditions proches de celles du Big-Bang (à très forte densité et à très haute température).

Dans la section 2.1, l'accélérateur de protons LHC sera présenté. Les sections 2.2-2.5 sont dédiées à la description du détecteur CMS et de ses différents composants. Les algorithmes de reconstruction des différents objets physiques seront également discutés.

2.1 Le LHC : un collisionneur proton-proton

Le LHC a été installé dans le tunnel initialement creusé pour le collisionneur e^+e^- LEP (*Large Electron Positron*). Les accélérateurs d'électrons-positrons souffrent d'importantes pertes d'énergie dues au rayonnement synchrotron, proportionnel à $(E^4)/(R m^4)$, où E est l'énergie de l'électron, m sa masse et R le rayon de l'accélérateur. Pour atteindre

des énergies de l'ordre du TeV, le rayon de l'accélérateur étant fixe, des particules chargées plus massives doivent être utilisées : les protons. Cela permet de réduire la perte d'énergie par rayonnement synchrotron d'un facteur 2000⁴ environ pour une énergie E fixée.

Un autre aspect important du LHC est son taux de collisions. Afin de produire des événements rares en quantités suffisantes pour être détectés, le taux de collision doit être élevé. Le faisceau étant divisé en paquets de particules, le taux de production de particules dans les collisions est proportionnel à la luminosité instantanée de l'accélérateur. Celle-ci est donnée par :

$$\mathcal{L} = \frac{f k n_p^2}{4\pi\sigma_x\sigma_y}, \quad (2.1)$$

où f est la fréquence de révolution d'un paquet, k le nombre de paquet, n_p est le nombre de protons par paquet et σ_x et σ_y leurs dispersions transverses en x et en y . Dans les conditions nominales de fonctionnement du LHC à 14 TeV ($\mathcal{L} = 10^{34}\text{cm}^{-2}\text{s}^{-1}$), les valeurs des ces quantités sont $k = 2808$, $n_p = 1.5 \times 10^{11}$ et $\sigma_x = \sigma_y = 16.6 \mu\text{m}$ (et $\sigma_z = 7.6$ cm le long du faisceau). La luminosité intégrée est définie par $L = \int \mathcal{L} dt$.

L'accélérateur Tevatron du Fermilab produit depuis 1992 des collisions proton-antiproton. L'énergie dans le centre de masse a été de 1.8 TeV jusqu'en 1998 et de 1.96 TeV depuis 2001. La principale limitation des collisionneurs proton-antiproton vient de la difficulté de produire des antiprotons en grande quantité. C'est pourquoi le LHC, un collisionneur proton-proton, pourrait atteindre une luminosité instantanée de plus de deux ordres de grandeur supérieure à celle du Tevatron.

Le LHC produira des collisions à une énergie de 14 TeV dans le centre de masse. Il est principalement constitué de 1232 aimants dipolaires supra-conducteurs de 15 m de long, délivrant un champ magnétique de 8.3 T afin de garder le faisceau sur sa trajectoire de 27 km de circonférence. Deux tubes à vide différents sont utilisés pour faire circuler chaque faisceau de protons en sens opposés. Un schéma représentant la section d'un aimant dipolaire est représenté sur la figure 2.1. Plus de 8000 autres aimants sont utilisés pour l'injection des faisceaux, leur collimation, la correction de leur trajectoire ou encore pour leur croisement. L'ensemble des aimants sont refroidis par de l'hélium liquide superfluide jusqu'à une température de 1.9 K

Les faisceaux sont accélérés de l'énergie d'injection dans le LHC (450 GeV) à une énergie de 7 TeV par 16 cavités Radio Fréquence (8 par faisceau) qui augmente l'énergie des faisceaux de 16 MeV par tour à l'aide d'un champ électrique de 5 MV/m oscillant à une fréquence de 400 MHz.

Avant d'être injectés dans le LHC, les faisceaux sont produits et accélérés par les différents composants du complexe d'accélération du CERN. Après avoir été produits à partir d'hydrogène ionisé, ils sont accélérés par l'accélérateur linéaire (LINAC), le

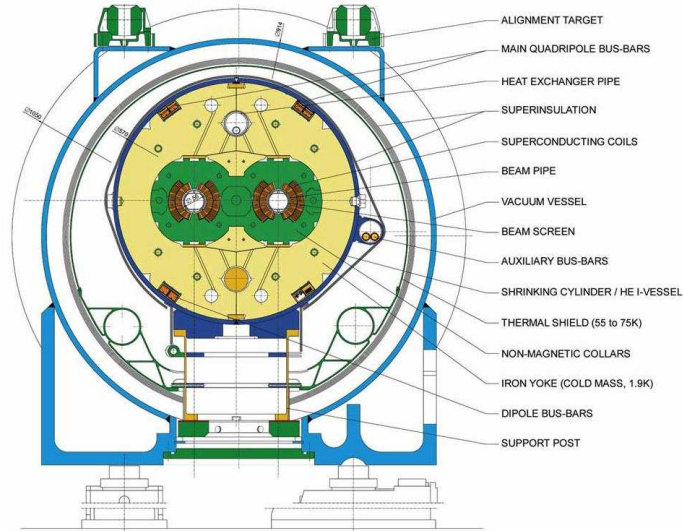


FIG. 2.1: Schéma de la section d'un aimant dipolaire du LHC

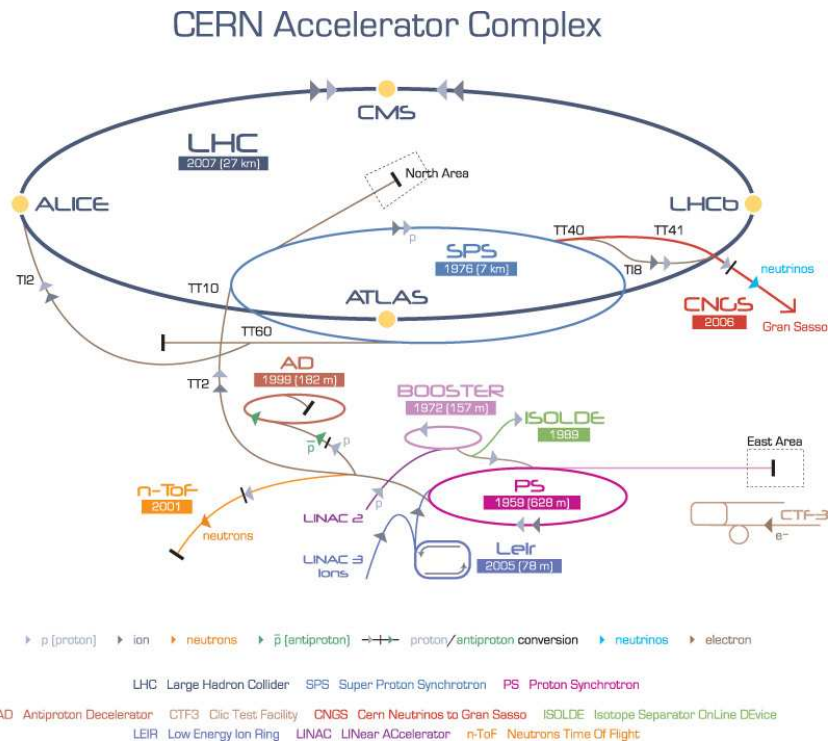


FIG. 2.2: Schéma représentant le complexe d'accélération du CERN

Booster et le Proton Synchrotron (PS) jusqu'à une énergie de 26 GeV, l'espacement entre les paquets étant de 25 ns. Les faisceaux sont alors injectés dans le Super Proton Synchrotron (SPS) où ils vont être accélérés jusqu'à une énergie de 450 GeV. Ils sont alors transférés au LHC et vont être accélérés jusqu'à une énergie de 7 TeV chacun. Le complexe d'accélération du CERN est présenté sur la figure **2.2**.

Le LHC a été officiellement inauguré le 10 septembre 2008. Ce jour, deux faisceaux de protons ont été injectés l'un après l'autre en sens opposés et ont parcouru chacun l'ensemble de l'accélérateur. Les premières collisions auront avoir lieu à l'automne 2009. Après une courte période de tests avec des faisceaux de 450 GeV, l'énergie devrait être de 5 TeV pour les premières analyses de physique. Le nombre de paquets dans chaque faisceau ainsi que le nombre de protons par paquet va augmenter graduellement jusqu'à ce que des luminosités de $10^{31}\text{cm}^{-2}\text{s}^{-1}$ à $10^{32}\text{cm}^{-2}\text{s}^{-1}$ soient atteintes. Des luminosités intégrées de 10pb^{-1} à 100pb^{-1} devraient être enregistrées. L'espacement entre les paquets sera de 75 ns et diminuera progressivement jusqu'à 25 ns, l'énergie des faisceaux devraient passer à 5 TeV en 2009 puis à 7 TeV. Au bout de 1 an de prise de données, le régime de "basse luminosité" $\mathcal{L} \simeq 10^{33}\text{cm}^{-2}\text{s}^{-1}$ devrait être atteint. Cette phase de basse luminosité devrait durer quelques années, permettant l'enregistrement de quelques dizaines de fb^{-1} . Après cette période, le LHC devrait être placé en mode de fonctionnement nominal ($\mathcal{L} \simeq 10^{34}\text{cm}^{-2}\text{s}^{-1}$) et enregistrer quelques centaines de fb^{-1} .

2.2 Le détecteur CMS

Le détecteur CMS¹ [51, 52] est dans sa conception un détecteur standard de physique des particules. De forme cylindrique et centré autour du tube dans lequel les faisceaux circulent, il est constitué d'une partie centrale et de deux bouchons. Trois types de détecteurs distincts permettent de mesurer la trajectoire des particules chargées (trajectographe), l'énergie des électrons, des photon et des jets (calorimètres électromagnétique et hadronique) ou encore de détecter les muons (chambres à muons).

Cependant, le détecteur CMS possède ses caractéristiques propres : le trajectographe ainsi que toute la calorimétrie sont placés dans un champ magnétique homogène de 4 T parallèle aux faisceaux et délivré par un aimant solénoïdal supra-conducteur. Au-delà de l'aimant, une structure en fer soutient les chambres à muons et est utilisé pour refermer les boucles de champ magnétique (*return yoke*). Le but est d'obtenir une résolution en impulsion d'environ 10% sur des muons de 1 TeV. Les détails du détecteur CMS sont représentés sur la figure **2.3**.

Le système de coordonnées adopté par CMS a pour origine le point de collision au centre du détecteur. L'axe y pointe verticalement vers le haut et l'axe horizontal x pointe vers le centre du LHC. L'axe z suit donc la direction des faisceaux et pointe vers le Jura.

¹Compact Muon Solenoid

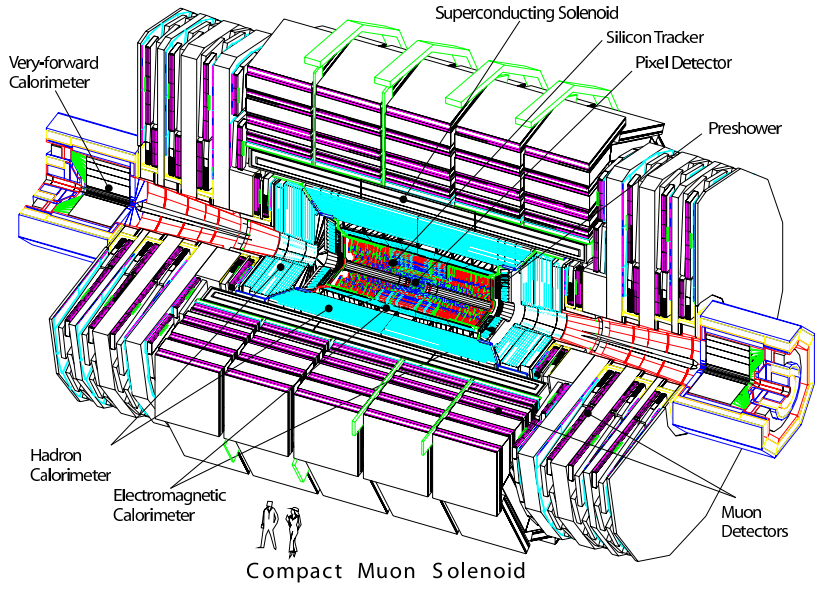


FIG. 2.3: Schéma du détecteur CMS.

L'angle azimutal ϕ est défini à partir de l'axe x dans le plan $x - y$. L'angle polaire θ est mesuré à partir de l'axe z . Les angles ϕ et θ sont présentés sur la figure 2.4.

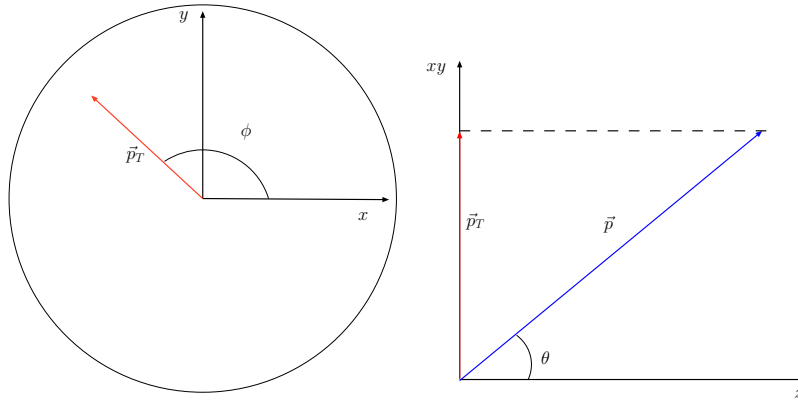


FIG. 2.4: Schéma du système de coordonnées utilisé dans CMS.

En pratique, la pseudo-rapidité η est utilisée à la place de l'angle polaire θ . Elle correspond à une approximation relativiste ($E \gg mc^2$) de la rapidité y définie par :

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right), \quad (2.2)$$

qui devient pour $E \gg mc^2$:

$$y \approx \eta = -\ln \left(\tan \frac{\theta}{2} \right), \quad (2.3)$$

où E est l'énergie, p_z est l'impulsion longitudinale et m la masse de la particule produite. La distance angulaire entre deux objets est alors définie par $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, qui est un invariant de Lorentz le long de l'axe z .

Dans la suite de cette section, il sera présenté une vue d'ensemble des différentes parties du détecteur ainsi que des algorithmes de reconstruction utilisés. La trajectographie, le spectromètre à muons et finalement le système de calorimétrie seront présentés.

2.3 La trajectographie

L'une des contraintes majeures de la reconstruction des traces de particules chargées provient de la grande multiplicité de particules chargées dans les collisions du LHC, qui complique la reconstruction et augmente le taux de fausses traces. De très bonnes efficacités de reconstruction de traces et des vertex primaires étant nécessaires, le trajectographe doit posséder une granularité très fine au plus près de l'axe des faisceaux.

2.3.1 Le trajectographe

La technologie choisie pour le trajectographe de CMS est basée sur des détecteurs semi-conducteurs en silicium : des détecteurs à pixels et des micro-pistes de silicium. Avec un rayon de 110 cm et une longueur de 540 cm, le trajectographe possède une couverture angulaire allant jusque $|\eta| < 2.5$. Le schéma du trajectographe de CMS est représenté sur la figure 2.5.

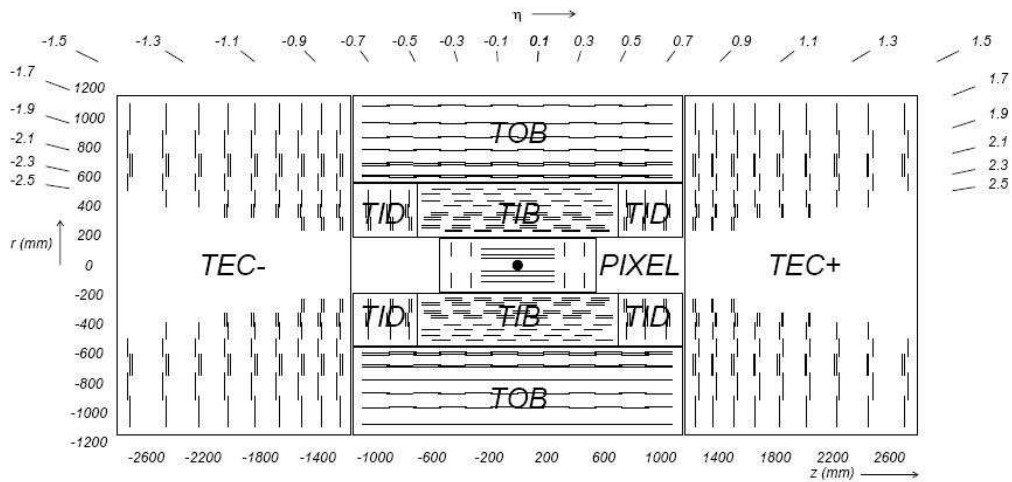


FIG. 2.5: Schéma représentant une coupe du trajectographe de CMS.

Le détecteur à pixels est situé au plus près du point d'interaction. Il est constitué de trois couches cylindriques placées autour de l'axe du faisceau dans la partie centrale.

Deux disques de pixels supplémentaires sont disposés perpendiculairement à l'axe des faisceaux sur les parties avant et arrière du détecteur (les bouchons). La taille d'un pixel est de $100 \mu m \times 150 \mu m$, la résolution en position dans le plan transverse à l'axe du faisceau (r, φ) est de $10 \mu m$ et la résolution suivant l'axe du faisceau z est de $20 \mu m$.

Les détecteurs à micro-pistes de silicium de la partie centrale sont disposés en régions distinctes : le TIB² (4 couches stéréos³) qui entoure le détecteur à pixels et le TOB⁴ (6 couches dont 2 stéréos) situés sur les parties les plus externes du trajectographe. Pour que les angles entre les trajectoires et le plan des modules de silicium ne soient pas trop grands dans les parties avant et arrière, le TIB est plus court que le TOB et l'espace ainsi libéré est comblé par le TID⁵ (2×3 disques dont 2 stéréos) placé sur les bouchons disposés perpendiculairement au faisceau. Le TEC⁶ (2×9 disques dont 1 stéréo) complète la couverture angulaire du trajectographe dans les bouchons jusqu'à $|\eta| = 2.5$. Les modules du TID et du TEC sont disposés en pétales perpendiculairement à l'axe du faisceau. La résolution des détecteurs à micro-pistes de silicium sur la position d'une particule chargée varie de 23 à 52 μm dans le plan (r, φ) et de 230 à 530 μm suivant l'axe z .

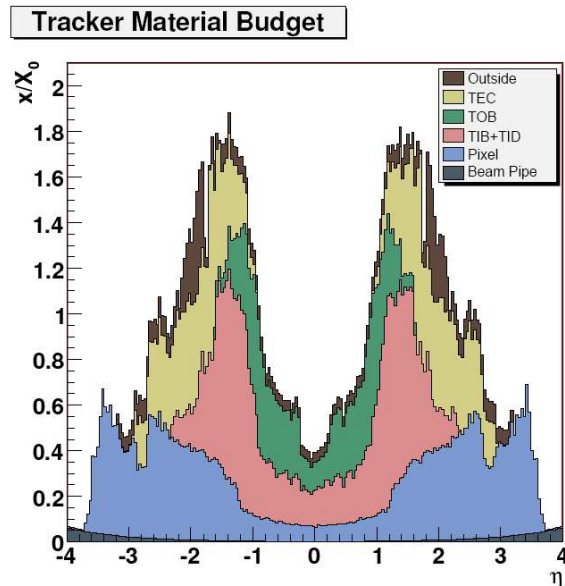


FIG. 2.6: La quantité de matière du trajectographe de CMS en unité de longueur de radiation en fonction de la pseudo-rapacité.

²Tracker Inner Barrel

³Couches constituées de deux faces distinctes de détection dont les pistes sont inclinées de 10° l'une par rapport à l'autre.

⁴Tracker Outer Barrel

⁵Tracker Inner Disk

⁶Tracker End Cap

La figure 2.6 représente la quantité de matière du trajectographe de CMS en unité de longueur de radiation en fonction de la pseudo-rapacité [51]. Elle augmente de $0.4 X_0$ pour $\eta \approx 0$ jusqu'à $1.8 X_0$ pour $\eta \approx 1.4$, qui correspond à la région de transition entre la partie centrale et les bouchons. Dans cette région, les efficacités de reconstruction des traces seront plus faibles. Au delà de $\eta \approx 1.4$, la quantité de matière diminue de nouveau pour atteindre $1 X_0$ pour $\eta \approx 2.5$.

2.3.2 Reconstruction de la trajectoire des particules chargées

2.3.2.1 Reconstruction des traces

Le principal algorithme de reconstruction des trajectoires de particules chargées (ou traces) va être brièvement décrit dans cette section. Il peut être décomposé en quatre étapes :

1. la reconstruction des coups ("points" d'interaction entre les particules chargées et le trajectographe) et l'estimation de leurs positions et de leurs incertitudes,
2. la génération des graines (proto-traces),
3. la reconstruction de la trajectoire,
4. l'ajustement final de la trace.

Le signal électrique induit par la traversée d'une particule chargée dans le trajectographe permet la reconstruction de coups. A partir de ces coups, les graines des traces peuvent être reconstruites. Celles-ci servent de trajectoires initiales pour la reconstruction complète. Les graines sont déterminées à l'aide de triplets de coups reconstruits dans les détecteurs à pixels ou de doublets de coups combinés avec des contraintes liées au vertex d'interaction. Ces coups sont utilisés pour reconstruire les traces à l'aide de l'équation du mouvement de particules chargées dans un champ magnétique. La trajectoire correspondante est décrite par une hélicoïde. Une fois que les graines sont reconstruites, leurs trajectoires sont une première fois extrapolées grossièrement de manière à définir une région dans le trajectographe contenant les coups reconstruits compatibles avec la trace.

Une méthode de filtre de Kalman [53, 54] est alors utilisée et va être brièvement décrite. Cette méthode permet de prendre en compte les effets de diffusions multiples durant le processus d'estimation de la trace et d'obtenir les paramètres de la trajectoire au niveau de chaque couche compatible du trajectographe.

Pour une trajectoire donnée et N couches du trajectographe compatibles, le filtre de Kalman procède de la manière suivante :

- A partir de l'état (pour un jeu de paramètres) de la trace sur la couche k , son état sur la couche $k + 1$ est prédit à l'aide des équations de propagation d'une particule chargée dans un champ magnétique.
- L'information de la couche $k + 1$ est étendue à toutes les couches précédentes en utilisant un *Filtre* qui estime l'état de la trajectoire en considérant les couches

comprises entre la première couche et la couche k incluses, on parle alors de filtre “avant”. Les filtres “arrières” estiment l’état de la trajectoire à l’aide des couches comprises entre la couche N et la couche k .

- Une dernière étape de *lissage* est utilisée. Elle permet d’estimer les paramètres de la trace au niveau de chaque couche compatible. L’estimation *lisse* est obtenue en appliquant 2 filtres, l’un avant (des couches 1 à k), et le second arrière (des couches N à k).

Dans la pratique, le filtre de Kalman démarre de l’estimation grossière des paramètres de la trace à partir des graines. L’ensemble des couches compatibles du trajectographe est déterminé et la trajectoire est extrapolée vers la couche du trajectographe la plus proche ($k + 1$), en partant des couches intérieures ($1 \dots k$) vers les couches extérieures. Les équations du mouvement d’une particule chargée dans un champ magnétique sont utilisées en prenant en compte les effets de diffusion multiple et de perte d’énergie dans le détecteur. L’algorithme recherche alors les coups compatibles de cette couche. Si plusieurs coups compatibles sont trouvés, une trace par coup compatible est créée. Si il n’y a pas de coup, un “coup invalide” virtuel est créé de manière à permettre à l’algorithme de chercher des coups dans les couches suivantes. Les paramètres des trajectoires des traces sont alors recalculés en prenant en compte les nouveaux coups sélectionnés. Cette procédure est répétée jusqu’à atteindre la dernière couche compatible du trajectographe.

Des ambiguïtés peuvent survenir dans la trajectoire des traces. Cela se produit quand, pour une même particule, plusieurs trajectoires sont reconstruites, ce qui peut provoquer un double comptage de traces. Les ambiguïtés sont résolues en comptant le nombre de coups communs à plusieurs traces. Pour une paire de traces donnée, la fraction de coups communs est définie par :

$$f_{commun} = \frac{N_{commun}^{coups}}{\min(N_1^{coups}, N_2^{coups})}, \quad (2.4)$$

où N_1^{coups} (N_2^{coups}) est le nombre de coups de la première (seconde) trace candidate. Si cette fraction est supérieure à 0.5, seule la trace qui possède le plus grand nombre de coups est retenue. Si les deux traces possèdent le même nombre de coups, seule la trace de plus faible χ^2 normalisé⁷ est conservée.

Finalement, les trajectoires des traces sont réajustées une première fois en prenant en compte les pertes d’énergie successives d’un coup à l’autre, en partant de l’intérieur et en allant vers l’extérieur du trajectographe (*lissage* pour tous les coups allant de 1 à N). Ce réajustement est répété une seconde fois en partant des coups les plus extérieurs et en allant vers les coups les plus intérieurs (*lissage* pour tous les coups allant de N à 1). Les deux jeux de paramètres ainsi obtenus sont combinés pour calculer la trajectoire finale de la trace. Cette procédure permet d’éviter les biais possibles provenant de la

⁷ χ^2 divisé par le nombre de degrés de liberté

reconstruction des graines. L'algorithme standard de reconstruction des traces décrit ici s'appelle *Combinatorial Track Finder* ou CTF. D'autres algorithmes de reconstruction seront discutés dans le chapitre 3.

Pour les versions les plus récentes des algorithmes de reconstruction de traces (versions CMSSW_2_1_12 et supérieures du code de CMS) une stratégie de reconstruction itérative des traces a été mise en place. L'idée est d'améliorer l'efficacité de reconstruction des traces en appliquant l'algorithme CTF successivement à des sous-ensembles de coups de plus en plus restreints.

La première étape consiste en la sélection d'une première collection de traces passant des coupures de qualité très sévères, de manière à minimiser le taux de fausses traces. Les coups associés à ces traces sont retirés de la liste des coups et afin de compenser la perte d'efficacité liée à la sélection sévère des traces, l'algorithme de reconstruction CTF est utilisé une nouvelle fois à partir des coups restants. Une seconde collection de traces est alors reconstruite en appliquant des coupures de qualité plus relâchées. Cette procédure peut être répétée plusieurs fois. Cette méthode conduit à une forte diminution du taux de fausses traces dans les jets et à une meilleure efficacité des traces de bas p_T .

Les efficacités de reconstruction des traces sont présentées pour des événements QCD sur la figure 2.7 en fonction de l'impulsion transverse p_T et de la pseudo-rapacité $|\eta|$ des traces pour différentes versions du code de reconstruction de CMS : la version CMSSW_1_6_12 (sans *reconstruction itérative*) et la versions CMSSW_2_1_12 (avec *reconstruction itérative*). Dans ces deux cas, l'efficacité de reconstruction des traces est de l'ordre de 90%. L'utilisation de la méthode de *reconstruction itérative* permet d'augmenter l'efficacité de reconstruction des traces de quelques pourcents pour des événements QCD , principalement pour les traces de faible impulsion transverse. Pour des événements $t\bar{t}$, cette augmentation peut atteindre une dizaine de pourcents.

Une baisse d'efficacité due à l'augmentation de la quantité de matière est observée dans la zone de transition entre la partie centrale et les bouchons ($0.9 < |\eta| < 1.2$). L'efficacité de reconstruction est de 80% à 85% dans les bouchons.

La figure 2.8 représente l'efficacité de reconstruction des traces provenant de muons dans des événements Z +jets générés par ALPGEN et reconstruits avec la version CMSSW_1_6_12 n'incluant pas la *reconstruction itérative*. L'efficacité de reconstruction des traces isolés issues de la désintégration de bosons Z est supérieure à 95% dans la partie centrale et de 90% environ dans les bouchons. L'efficacité de reconstruction des traces est donc meilleure pour les traces isolées.

La figure 2.9 présente le taux de fausses traces reconstruites dans des événements QCD pour les versions CMSSW_1_6_12 et CMSSW_2_1_12. On remarque une très nette diminution du taux de fausses traces dans le cas de la *reconstruction itérative*, allant jusqu'à un facteur 4 dans la zone de transition entre la partie centrale et les bouchons ($|\eta| \approx 1.5$). Nous verrons dans le chapitre 4 que cela aura un effet significatif sur les efficacités d'identification de jets de quark b et sur le taux de mauvaise identification.

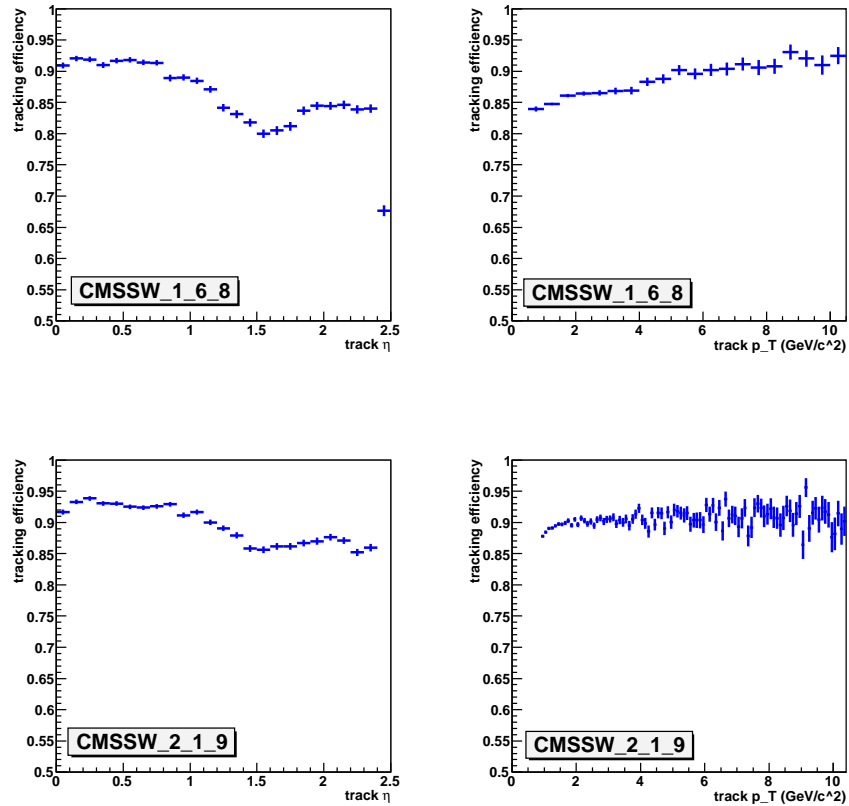


FIG. 2.7: Efficacité de reconstruction des traces dans des événements QCD $80 < \hat{p}_T < 120$ GeV/c pour les versions CMSSW_1_6_12 (en haut) et CMSSW_2_1_12 (en bas) du code de CMS en fonction de l'impulsion transverse (à gauche) et de la pseudo-rapidité des traces (à droite).

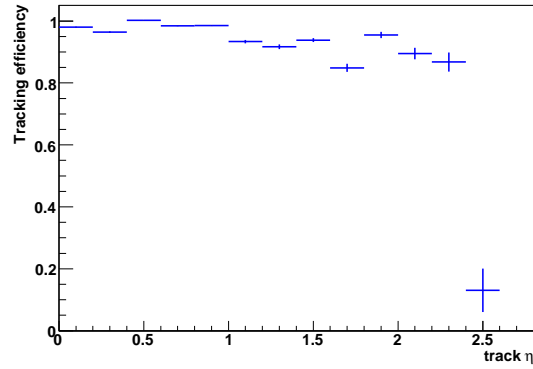


FIG. 2.8: Efficacité de reconstruction des traces dans les événements Z +jets générés par ALPGEN et reconstruits avec la version CMSSW_1_6_12 avec l'algorithme de reconstruction CTF sans *reconstruction itérative*.

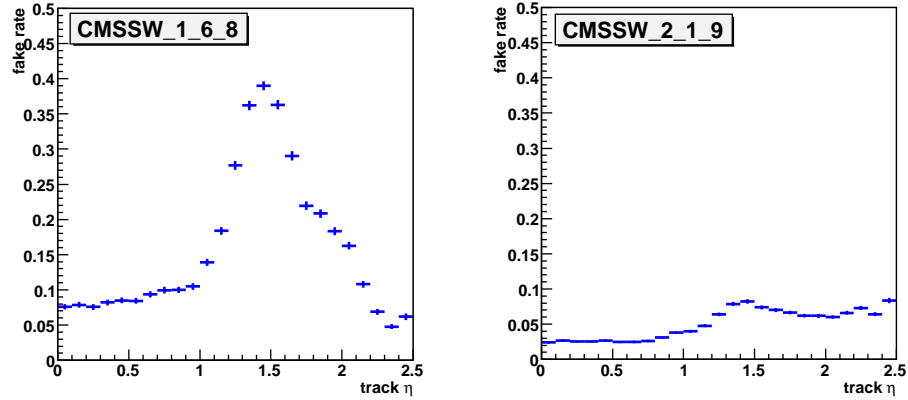


FIG. 2.9: Taux de fausses traces reconstruites dans des événements QCD $80 < \hat{p}_T < 120$ GeV/c pour les versions CMSSW_1_6_12 (à gauche) et CMSSW_2_1_12 (à droite) en fonction de la pseudo-rapacité des traces.

La mesure des efficacités de reconstruction des traces de muons cosmiques dans les données réelles sera présentée dans le chapitre 3.

2.3.2.2 Reconstruction des vertex

La reconstruction des vertex d'interaction permet de rejeter les traces provenant d'éventuels événements d'empilement. Nous verrons dans le chapitre 4 que ces algorithmes de reconstruction de vertex sont également utilisés dans la plupart des algorithmes d'identification de jets de quark b pour calculer par exemple les paramètres d'impact (cf section 4.1.1) ou pour reconstruire les vertex secondaires issus de la désintégration de hadrons beaux.

La reconstruction des vertex primaires se fait en deux étapes. Premièrement les traces reconstruites sont regroupées en candidats vertex en comparant les coordonnées z des points de leurs trajectoires qui sont les plus proches du tube du faisceau. Seules les traces ayant un paramètre d'impact transverse inférieure à 3 cm sont utilisées. Les vertex sont ensuite reconstruits à l'aide d'une méthode séquentielle d'estimation des paramètres du vertex primaire en considérant successivement les traces associées. Un filtre de Kalman est utilisé [55] dans cette procédure.

Pour un événement donné, le vertex primaire est choisi comme celui possédant la somme des impulsions transverses des traces associées $\sum p_T$ la plus élevée. L'efficacité de reconstruction des vertex primaires est proche de 100% et la résolution en position dans des événements $t\bar{t}$ est d'environ 13 μm suivant l'axe des x et 18 μm suivant l'axe des z [52].

Il est également possible de reconstruire les vertex secondaires issus, par exemple, de la désintégration de hadrons beaux (cf. section 4.1.2). La reconstruction des vertex secondaires utilise les traces associées aux jets en appliquant des coupures de sélection supplémentaires : le paramètre d'impact transverse des traces doit être supérieur à 100 μm (pour éviter de prendre en compte les traces provenant du vertex primaire) et inférieur à 2 cm (pour éviter de prendre en compte les traces produites par d'éventuelles interactions avec le tube à vide). Dans des événements QCD avec $80 < \hat{p}_T < 120$ GeV/c, l'efficacité de reconstruction des vertex secondaires dans les jets b est de 65% environ. Pour des jets de hadrons b ayant un p_T compris entre 20 et 70 GeV/c, la résolution de la distance de vol (2 déviations standards) est de 765 μm dans la partie centrale et 2275 μm dans les bouchons [56].

2.4 Le spectromètre à muons

Les détecteurs à muons sont positionnés dans la structure en fer utilisée pour refermer les boucles de champs magnétique. Le retour du champ magnétique est mis à profit pour augmenter la précision sur l'impulsion transverse reconstruite des muons et améliorer l'identification de leur charge électrique. De la même manière que pour le trajectographe,

le spectromètre à muon est divisé en 3 parties : la partie centrale ($|\eta| < 1.2$ du détecteur et les deux bouchons ($0.9 < |\eta| < 2.4$).

Dans une première partie, le dispositif de détection des muons sera présenté. Les algorithmes de reconstruction des muons seront ensuite discutés.

2.4.1 Le dispositif de détection

Le spectromètre à muons permet la reconstruction de ces particules jusqu'à une pseudo-rapacité $|\eta| < 2.4$. Trois types de détecteur à gaz sont utilisés : des chambres de tubes à dérive (*Drift Tube Chambers, DT*), des chambres à pistes cathodiques (*Cathode Strip Chambers CSC*) et des chambres à plaques résistives (*Resistive-plate Chambers RPC*) :

- **DT** : basées sur le principe de fonctionnement classique des tubes à dérive, ces chambres sont constituées d'ensembles de tubes d'aluminium, alignés avec l'axe z , contenant du gaz et traversées par un fil servant d'anode. Le gaz est ionisé au passage d'une particule chargée et les charges électriques ainsi produites sont collectées par l'anode. Il est ainsi possible de déterminer la position en R et en ϕ du muon dans la chambre et de mesurer sa position en z à l'aide des différences de charges collectées aux deux extrémités de l'anode. L'ensemble des 250 chambres *DT* est réparti en quatre "stations" : MB 1, 2, 3 et 4. Ces stations sont disposées les unes par rapport aux autres de manière à ce qu'un muon de grande impulsion transverse traverse au moins trois des quatre stations. Chaque station permet de mesurer la position des muons avec une précision de $100 \mu\text{m}$ environ et leur direction avec une précision de 1 mrad.
- **CSC** : ces chambres sont constituées de six volumes remplis de gaz délimités par sept plaques. Des pistes de cathodes sont disposées sur l'une des faces de chaque plaque. Une grille de fils d'anode est placée dans chaque volume de gaz. Lorsqu'une particule chargée traverse la chambre, l'ionisation du gaz produit une avalanche d'électrons qui se déposent rapidement sur les fils d'anodes. La résolution des *CSC* est d'environ $200 \mu\text{m}$ en position et 10 mrad en direction.
- **RPC** : elles sont constituées de deux volumes de gaz espacés par des plaques de bakélite. Elles possèdent une moins bonne résolution que les chambres *DT* et *CSC* mais sont beaucoup plus rapides. Elles sont particulièrement utilisées par le système de déclenchement. Elles complètent également les autres détecteurs à muons, notamment dans les régions de transition entre la partie centrale et les bouchons.

Dans la partie centrale, les quatre stations constituées de chambres *DT* et *RPC* sont disposées de manière concentrique autour du faisceau à des rayons compris entre 4 et 7 m. Dans les bouchons, quatre disques composés de chambres *CSC* et *RPC* sont placés perpendiculairement au faisceau. La disposition des chambres à muons est présentée sur la figure 2.10.

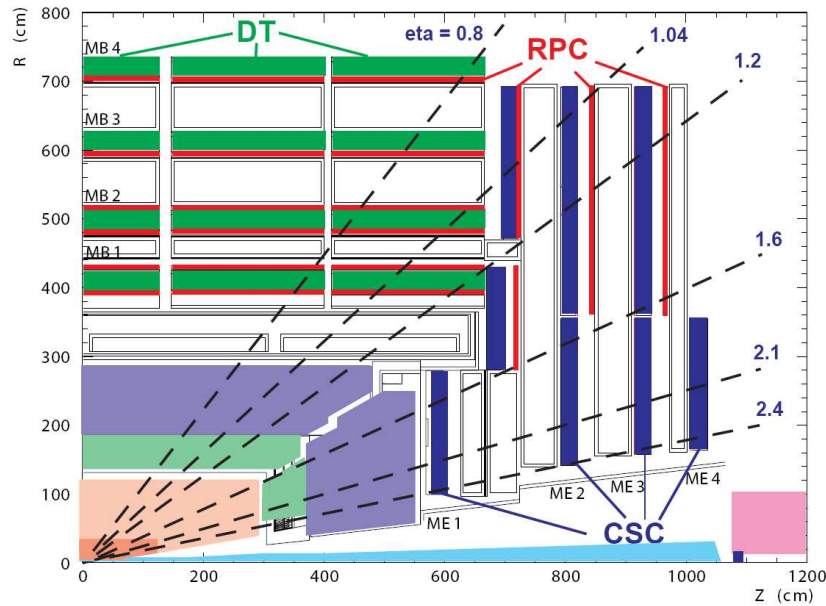


FIG. 2.10: Schéma représentant une coupe d'un quart du détecteur CMS avec ses chambres à muons.

2.4.2 Reconstruction des muons

2.4.2.1 Les muons locaux

À partir des coups enregistrés dans les chambres à muons, des segments de traces sont reconstruits dans chaque chambre. La reconstruction des trajectoires commence à partir du segment de trace reconstruit le plus proche du faisceau. Une méthode de filtre de Kalman (cf. section 2.3.2.1) est utilisée pour propager les traces vers les couches les plus externes du détecteur à muons, en prenant en compte les effets de perte d'énergie, les effets de diffusion multiple et la non-uniformité du champ magnétique. Une fois les chambres les plus externes atteintes, un filtre de Kalman est appliqué en sens inverse, vers le faisceau. La trace est ainsi réajustée et extrapolée au vertex primaire.

2.4.2.2 Les muons globaux

Les muons globaux sont reconstruits à partir des muons locaux et des traces reconstruites dans le trajectographe. Tout d'abord, la trajectoire de muons locaux est extrapolée vers le trajectographe afin de définir la région compatible et les couches correspondantes. Une graine est créée à partir des deux coups les plus proches du faisceau. La trace du muon dans le trajectographe est alors reconstruite en utilisant la méthode de filtre de Kalman (cf section 2.3.2.1) présentée précédemment. Finalement, la trace du muon reconstruite dans le trajectographe est réajustée en prenant en compte l'infor-

mation des muons locaux.

Pour des événements $t\bar{t}$, les efficacités de reconstruction des muons globaux issus de bosons W sont présentées sur la figure 2.11 en fonction de l'impulsion transverse p_T et de la pseudo-rapidité des muons générés associés. Un muon reconstruit est associé au muon généré le plus proche si la distance angulaire ΔR qui les sépare est inférieure à 0.3. Une sélection sur l'impulsion transverse des muons générés ($p_T > 10$ GeV/c) est appliquée. L'efficacité de reconstruction des muons globaux est d'environ 97% et semble peu dépendre du p_T . Dans les régions situées entre les deux roues supportant des chambres DT ($|\eta| \approx 0.25$ et $|\eta| \approx 0.8$) une baisse d'efficacité est observée. Une faible diminution de l'efficacité apparaît également dans la région de transition entre les chambres DT et CSC ($|\eta| \approx 1.2$).

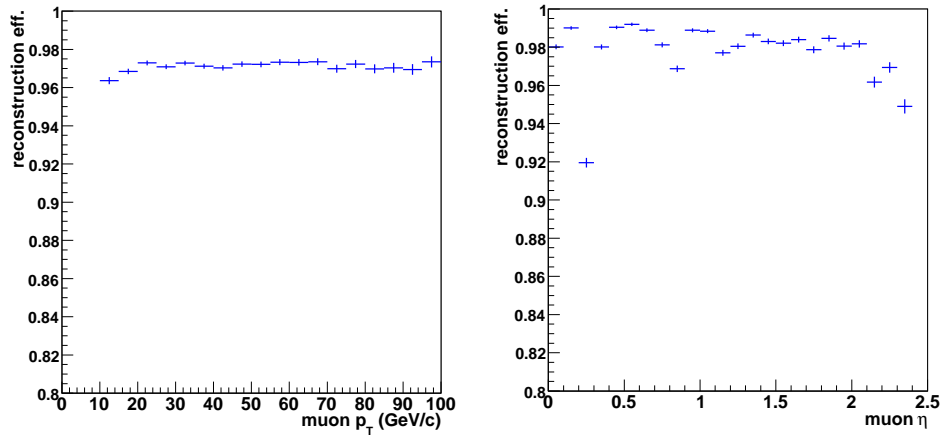


FIG. 2.11: Efficacité de reconstruction des muons issus de la désintégration de bosons W dans des événements $t\bar{t}$ en fonction du p_T et du $|\eta|$ des muons générés associés. Une coupure à 10 GeV/c sur l'impulsion transverse des muons générés est appliquée.

Les résolutions en impulsion transverse et en angle des muons globaux ont été estimées en fonction du p_T du muon généré associé. La résolution relative sur une variable q est calculée en ajustant une fonction gaussienne sur la quantité :

$$R_q = \frac{q_{reco} - q_{MC}}{q_{MC}}, \quad (2.5)$$

où $reco$ se réfère aux muons reconstruits et MC aux muons générés associés. La résolution relative sur la quantité q est alors définie comme l'écart type de la fonction gaussienne noté $\sigma(q)$. Les résolutions relatives sont estimées pour les quantités $1/p_T$, $|\eta|$ et ϕ sur différents intervalles de p_T . Elles sont présentées sur la figure 2.12.

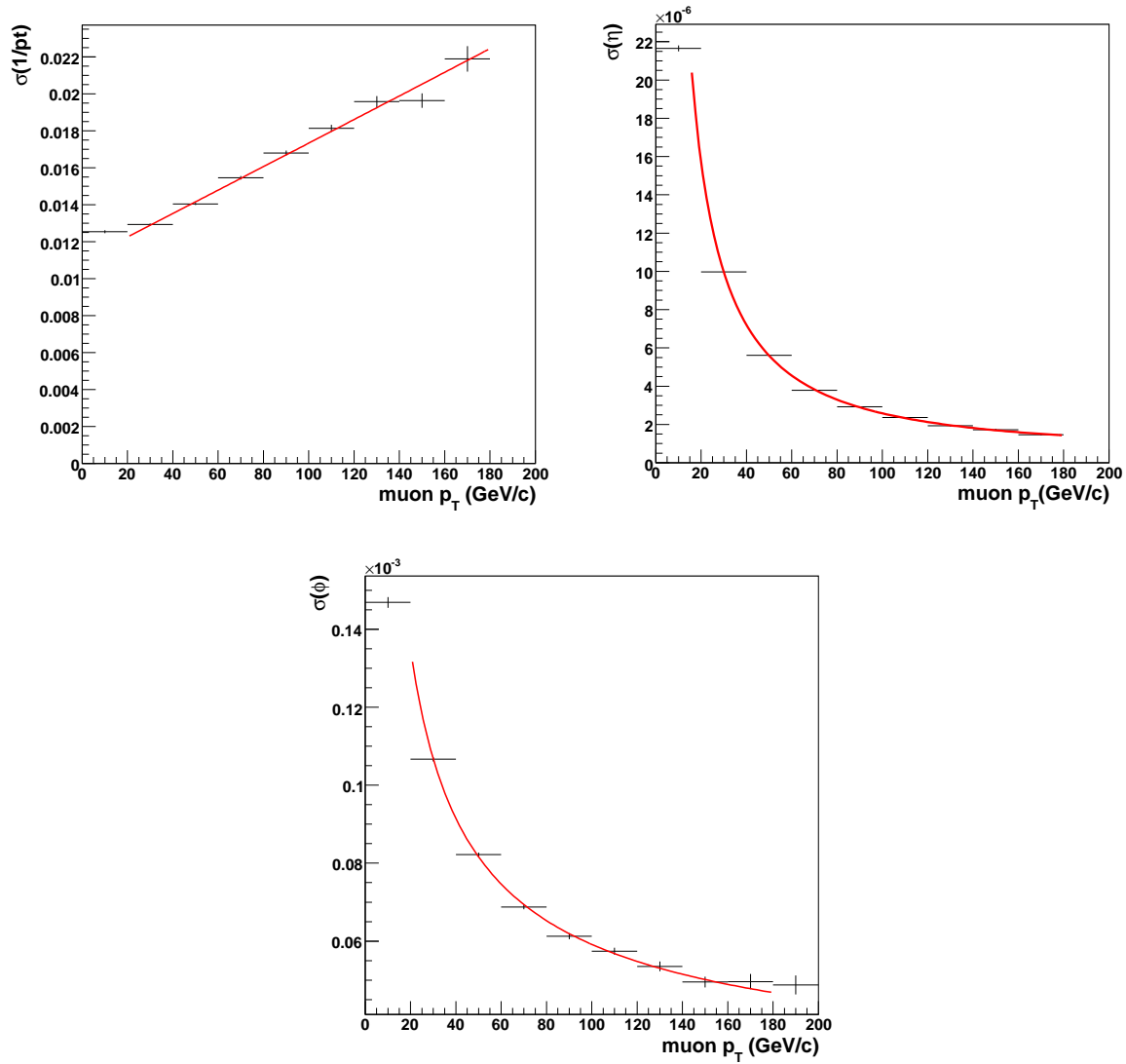


FIG. 2.12: Résolutions relative sur l'impulsion transverse p_T (en haut à gauche), sur la pseudo-rapacité $|\eta|$ (en haut à droite) et sur l'angle azimutal ϕ (en bas) des muons en fonction du p_T des muons générés associés.

La résolution relative sur l'impulsion transverse est de l'ordre 10^{-2} pour des muons de $p_T = 30$ GeV/c et augmente pour des muons de plus haute énergie. Deux différents effets sont à prendre en compte :

1. D'une part, la diffusion multiple dans le détecteur qui tend à diminuer avec p_T .
2. D'autre part, la trajectoire des muons de grande énergie tend vers une ligne droite. La détermination de l'impulsion (mesurée à partir de la double courbure de la trajectoire) est donc plus difficile ce qui tend à dégrader la résolution en p_T pour de grandes impulsions transverses.

L'effet 1 se trouve être moins important, dans le cas des muons, que l'effet 2. Il en résulte une dégradation de la résolution en p_T pour les muons en fonction du p_T .

On remarque que les résolutions relative sur $|\eta|$ et ϕ sont de l'ordre de 10^{-5} et 10^{-4} respectivement pour des muons de $p_T = 30$ GeV/c et diminuent à grand p_T .

Les efficacités de reconstruction des muons seront estimées dans les vraies données par une méthode dite de *tag & probe* [57]. Cette méthode utilise les événements contenant des Z qui se désintègrent en $\mu^+\mu^-$. L'efficacité de reconstruction des muons locaux est estimée à partir des événements contenant un muon bien reconstruit qui passe des coupures d'identification (appelés *tag*) et une trace (appelé *probe*). Le muon *tag* et la trace *probe* doivent avoir une masse invariante compatible avec la masse du Z afin de sélectionner des événements contenant des Z . L'efficacité des muons locaux est alors estimée par le rapport entre le nombre d'événements avec un muon *tag* et une trace *probe* associée à un muon local et le nombre d'événements avec un muon *tag* et une trace *probe*.

L'efficacité de reconstruction des traces de muons peut être estimée de la même manière en définissant le muon *probe* par un muon local. Les efficacités de reconstruction des traces et des muons locaux sont ensuite combinées pour estimer l'efficacité de reconstruction des muons globaux [58].

2.5 La calorimétrie

Le système de calorimétrie de CMS se divise en deux parties. Tout d'abord le calorimètre électromagnétique ECAL, placé au plus près du trajectographe, est dédié à la détection des électrons et des photons. Entre ECAL et la bobine du solénoïde se trouve le calorimètre hadronique HCAL, dédié à la détection des jets de hadrons.

Dans la suite de ce chapitre, le système de calorimétrie de CMS sera présenté. Les algorithmes de reconstruction des différents objets physiques liés aux calorimètres (les électrons et les photons, les jets et l'énergie transverse manquante \cancel{E}_T) seront ensuite discutés.

2.5.1 Le calorimètre électromagnétique ECAL

Le calorimètre électromagnétique est constitué d'un ensemble de cristaux scintillants de tungstate de plomb $PbWO_4$. Ce matériau a été choisi pour sa courte longueur de radiation, sa grande vitesse de scintillation et sa bonne résistance aux radiations.

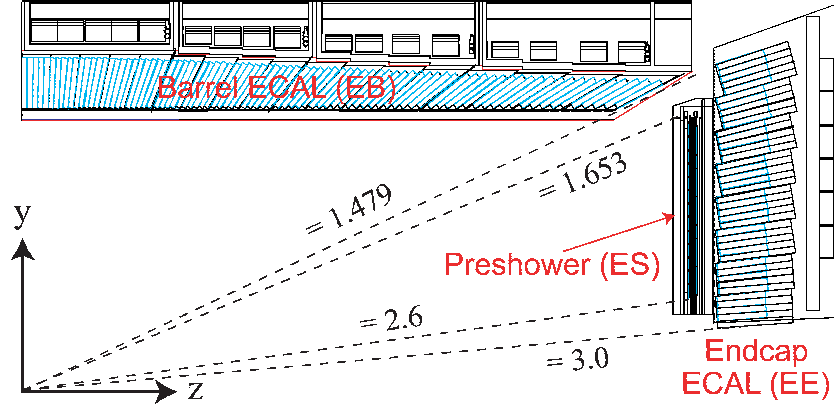


FIG. 2.13: Schéma illustrant la disposition des cristaux du ECAL dans une coupe représentant un quark du détecteur.

Dans la partie centrale du détecteur ($0 < |\eta| < 1.48$) les cristaux pointent vers le point d'interaction. Ils ont une section de $22 \times 22 \text{ mm}^2$ ce qui se traduit par une très grande granularité du détecteur en $\eta - \phi$ le long du faisceau. Les cristaux ont une longueur de 230 mm qui correspond à 25 longueurs de radiation. Ils sont regroupés en 36 super-modules disposés le long du faisceau.

Dans les bouchons ($1.48 < |\eta| < 3$), 4 demi-disques (2 par bouchons) supportent des ensemble de 5×5 cristaux, de section $28.6 \times 28.6 \text{ mm}^2$ et de longueur 220 mm. Les supports de ces cristaux sont perpendiculaires à l'axe du faisceau et les faces des cristaux sont orientées vers le point d'interaction. La disposition des cristaux est représentée sur la figure 2.13.

Un détecteur supplémentaire à pied de gerbe est utilisé pour améliorer l'identification et la détermination de la position des électrons et des photons dans les bouchons ($1.6 < |\eta| < 2.6$). Il est positionné en avant des cristaux et est constitué d'une plaque de plomb, servant à initialiser la gerbe électromagnétique et d'un détecteur à pistes de silicium.

Les performances du calorimètre électromagnétique ont été mesurées en faisceaux test. La résolution en énergie des super-modules $\sigma(E)/E$ est paramétrisée en fonction de l'énergie E des électrons incidents :

$$\frac{\sigma(E)}{E} = \left(\frac{S}{\sqrt{E}} \right) \oplus \left(\frac{N}{E} \right) \oplus C, \quad (2.6)$$

où S est le terme de réponse stochastique, N est le terme incluant les bruits provenant d'événements d'empilement, d'événements sous-jacent ou d'origine électronique et C est

un terme constant. La résolution en énergie en fonction de l'énergie des électrons du faisceau test est illustrée par la figure 2.14. Elle est globalement inférieure à 1% pour des électrons d'énergie transverse de plus de 20 GeV et pourrait atteindre une valeur de 0.4% à haute énergie.

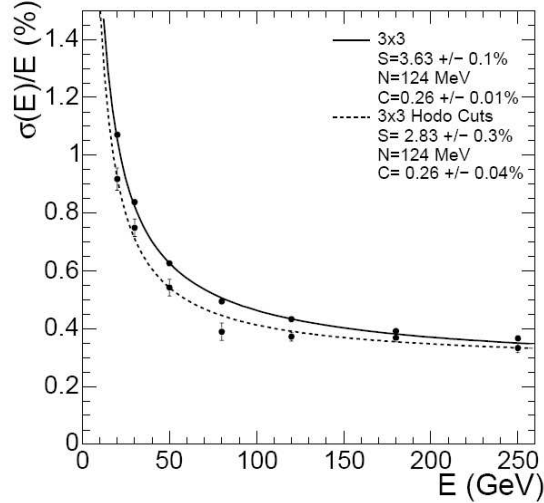


FIG. 2.14: Résolution en énergie du ECAL mesurée en faisceau test.

2.5.2 Le calorimètre hadronique HCAL

Le calorimètre hadronique entoure le calorimètre électromagnétique à l'intérieur du solénoïde. Il est constitué d'une superposition de couches de cuivre, servant d'absorbeur, et de couches de scintillateurs plastiques. Il est complété par des scintillateurs supplémentaires disposés au-delà de l'aimant. On peut distinguer quatre parties : la partie centrale (*Hadron Barrel*), la partie centrale hors de la bobine du solénoïde (*Hadron Outer*), les bouchons (*Hadron Endcaps*) et la partie avant (*Hadron Forward*).

- *Hadron Barrel* ($0 < |\eta| < 1.4$) est composé de 15 plaques de cuivres intercalées par des scintillateurs plastiques de granularité $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. Des fibres scintillantes à décalage de longueur d'onde (WLS⁸) permettent de collecter la lumière émise par scintillation. Deux couches supplémentaires d'acier sont utilisées comme absorbeur ce qui permet d'augmenter la résistance mécanique de la structure.
- *Hadron Outer* couvre une région allant jusqu'à $|\eta| = 1.26$ et est situé hors de la bobine du solénoïde suivant la segmentation des chambres *DT*. Ce détecteur est utilisé pour mesurer l'énergie provenant des gerbes de hadrons qui s'étendent au-delà du détecteur *Hadron Barrel*.

⁸Wavelength Light Shifter

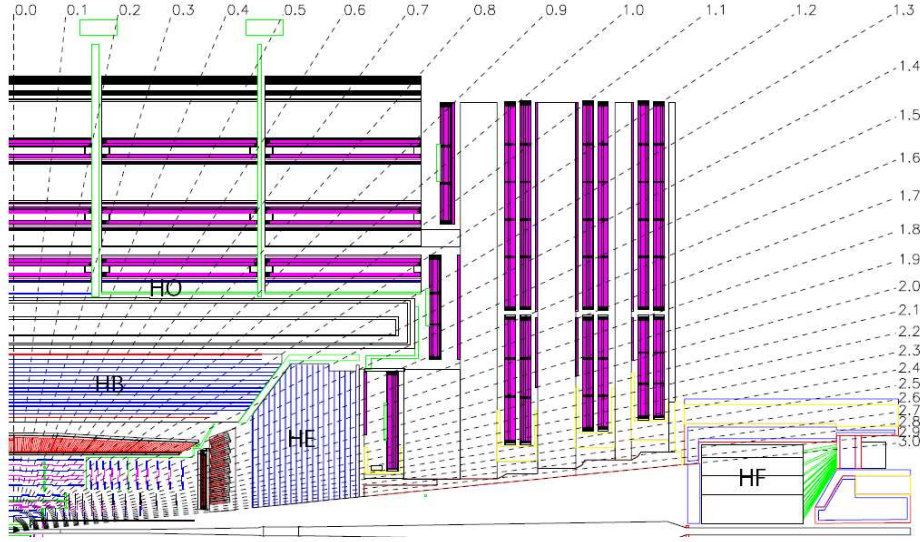


FIG. 2.15: Schéma de l'arrangement des tours calorimétriques de HCAL dans le plan (r, z) pour un quart du détecteur.

- *Hadron Endcaps* couvre la région de pseudo-rapacité $1.3 < |\eta| < 3.0$. Les tours calorimétriques les plus éloignées du point d'interaction ont une segmentation de 5° en ϕ et $0,087$ en η tandis que les tours les plus proches ont une segmentation de 10° en ϕ et comprise entre $0,09$ et $0,35$ en η .
- *Hadron Forward* ($3 < |\eta| < 5$) est composé de fibres d'acier et de quartz. Les fibres de quartz, disposées le long de l'axe du faisceau, agissent comme un détecteur Cerenkov. Ce détecteur est segmenté en 13 tours de section $\Delta\eta = 0.175$ et $\Delta\phi = 10^\circ$. Les longueurs d'interactions du calorimètre hadronique sont comprises entre 5.82 et 10.6 .

Le calorimètre *HF* peut en particulier être utilisé pour mesurer la luminosité [52]. Le nombre d'interactions par croisement de faisceau n est distribué suivant une distribution de Poisson :

$$p(n; \mu) = \mu^n \frac{e^{-\mu}}{n!}, \quad (2.7)$$

où $\mu = \frac{\sigma \mathcal{L}}{f_{BX}}$ est le nombre moyen d'interactions avec \mathcal{L} la luminosité, σ la section efficace totale pp et f_{BX} la fréquence des croisements de faisceau. A basse luminosité, le nombre moyen d'événements μ est faible. Pour une luminosité de $10^{32} \text{ cm}^{-2} \text{ s}^{-1}$, μ est inférieur à 1. Pour une luminosité de $10^{33} \text{ cm}^{-2} \text{ s}^{-1}$ et $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, μ est d'environ 2.5 et 25 respectivement. En distinguant les croisements de faisceaux ne contenant aucun événement et les croisements avec au moins un événement, il est possible de déterminer

μ à l'aide de la formule (dérivé de 2.7 pour $n=0$) :

$$\mu = -\ln p(0), \quad (2.8)$$

la probabilité $p(0)$ est déterminée en comptant le nombre de croisements de faisceaux pour lesquels la somme des énergies transverses mesurées par HF est inférieure à un certain seuil. D'autre dispositif sont utilisés dans CMS pour mesurer la luminosité : le Télescope de Luminosité à Pixels (PLT) ou le détecteur Totem [52]. Les événements W et Z peuvent également être utilisés comme chandelle standard pour estimer la luminosité.

2.5.3 Reconstruction des électrons

La reconstruction des électrons combine l'utilisation du trajectographe et du calorimètre ECAL. Les électrons produits au point d'interaction traversent une grande quantité de matière avant d'atteindre ce calorimètre (en moyenne 26 longueurs de radiation) et peuvent donc perdre une importante quantité d'énergie en émettant des photons par rayonnement de freinage (Brehmstrahlung). En présence du champ magnétique, cela se traduit par un étalement en ϕ des dépôts d'énergie. A l'aide de la simulation, il a été estimé qu'environ 50% des électrons de 10 GeV perdent en rayonnant plus de la moitié de leur énergie avant d'atteindre ECAL. Des photons émis par un électron peuvent également se convertir en paire e^+e^- de plus basse énergie, éventuellement piégée par le champ magnétique et ainsi augmenter les pertes d'énergie dans le détecteur.

Un amas calorimétrique correspond à l'ensemble des cellules où une même particule a déposé de l'énergie. Les algorithmes de reconstruction d'amas calorimétriques [59] doivent prendre en compte autant que possible les effets d'étalement en ϕ et de perte d'énergie dans le détecteur. Pour cela, des algorithmes de reconstruction de super-amas ont été développés. Ils permettent de regrouper les amas calorimétriques provenant des produits de radiation d'un électron.

Les amas et les super-amas de ECAL sont ensuite utilisés pour trouver l'ensemble de paires de coups compatibles dans le détecteur à pixels. Ces derniers vont servir de graine pour la reconstruction des traces d'électrons. Pour cela, la trajectoire et l'énergie des électrons sont estimées à partir des super-amas. La trajectoire est alors propagée vers le détecteur à pixels. Le premier puis le second coup compatible sont recherchés successivement dans une fenêtre en ϕ et z . Cela permet de réduire le nombre de candidats graines et d'augmenter la pureté des échantillons de candidats traces. Cette méthode d'identification des graines a été développée pour le système de déclenchement [60].

La reconstruction de la trajectoire de la trace de l'électron suit le schéma décrit dans la section 2.3.2.1, en utilisant un filtre non-linéaire appelé Filtre à Somme de gaussiennes (GSF) [61]. Dans un filtre de Kalman classique, les erreurs sur la mesure et les différents bruits sont supposés être décrits par une fonction gaussienne dans l'équation de propagation. Afin d'obtenir une meilleure prise en compte des pertes d'énergie dans

le trajectographe et du Brehmstrahlung, le filtre GSF utilise une somme de fonctions gaussiennes pour décrire les erreurs et les bruits. Les efficacités de reconstruction des traces d'électrons sont meilleures à bas p_T .

Une fois la trace de l'électron reconstruite, une association avec les super-amas est réalisée ainsi qu'une première pré-sélection basée sur des critères lâches de sélection géométrique, sur l'énergie et sur l'impulsion. La définition d'un candidat électron est :

- une trace d'électron reconstruite,
- le rapport entre l'énergie du super-amas E_{SC} et l'impulsion de la trace calculée au point d'interaction p_{in} doit être $E_{SC}/p_{in} < 3$,
- la distance $|\Delta\eta_{in}| = |\eta_{SC} - \eta_{trace}^{extrap.}|$ doit être < 0.1 , où η_{SC} est la pseudo-rapidité du super-amas et $\eta_{trace}^{extrap.}$ celle de la trace extrapolée à la surface du super-amas,
- la distance $|\Delta\phi_{in}| = |\phi_{SC} - \phi_{trace}^{extrap.}|$ doit être < 0.1 , où ϕ_{SC} est l'angle azimutal du super-amas et $\phi_{trace}^{extrap.}$ celui de la trace extrapolée à la surface du super-amas,
- le rapport entre l'énergie déposée dans le calorimètre HCAL et l'énergie déposée dans le calorimètre ECAL doit être inférieur à 0.2.

Pour chaque candidat électron ainsi défini, les informations provenant de la trace du super-amas sont combinées de manière à améliorer la résolution sur l'impulsion initiale de l'électron au vertex primaire. Durant ce processus de reconstruction, plus d'une trace par super-amas peut être reconstruite. Dans ce cas, le candidat électron sélectionné est celui possédant le rapport E/p le plus proche de un.

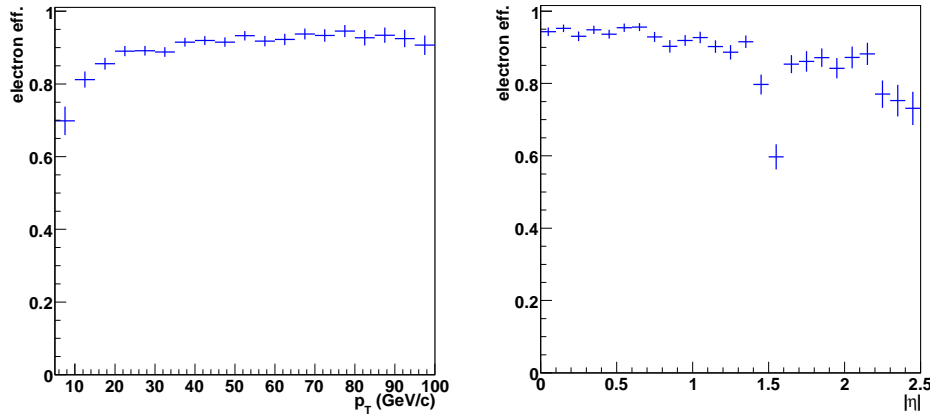


FIG. 2.16: Efficacité de reconstruction des électrons issus de la désintégration de bosons W dans des événements $t\bar{t}$, en fonction de l'impulsion transverse p_T et de la pseudo-rapidité des électrons générés associés.

L'efficacité de reconstruction des électrons est présentée sur la figure 2.16 en fonction de l'impulsion transverse p_T et de la pseudo-rapidité $|\eta|$ pour des électrons provenant de

la désintégration de bosons W dans des événements $t\bar{t}$. L'efficacité augmente avec le p_T jusqu'à atteindre un plateau. Les efficacités de reconstruction des électrons d'impulsion transverse $p_T > 20$ GeV/c sont supérieures à 90%. Autour de $|\eta| = 1.5$, la quantité de matière augmentant, la perte d'énergie des électrons s'en trouve accrue, il en résulte une nette diminution de l'efficacité de reconstruction ($\approx 50\%$).

Les résolutions sur l'énergie transverse et sur la direction des électrons sont calculées de la même manière que pour les muons (cf section 2.4.2). L'évolution de la résolution en énergie d'un électron est décrite par la fonction 2.6, elle diminue avec l'énergie jusqu'à atteindre un seuil. Les résolutions sont représentées par les graphiques de la figure 2.17 et sont de l'ordre de 0.02, 1.5×10^{-5} et 1.7×10^{-3} respectivement pour l'énergie transverse E_T , $|\eta|$ et ϕ pour des électrons provenant de la désintégration de bosons W dans des événements $t\bar{t}$.

Le taux de faux électrons reconstruits peut se révéler très élevé. Des critères d'identification permettant de le réduire seront présentés dans la section 5.2.5.

2.5.4 Reconstruction des jets de particules et de l'énergie transverse manquante

Lorsqu'un quark ou un gluon de haute énergie s'hadronise (cf. 1.6), les hadrons produits ont tendance à être émis dans la même direction que le parton initial. L'ensemble de ces hadrons est appelé jet de particules. La manière la plus simple de reconstruire un jet est de sommer l'énergie des amas calorimétriques présents dans un cône autour de la direction du parton initial. Le cône est décrit dans l'espace (η, ϕ) par sa direction et son ouverture ΔR .

Les trois algorithmes de jets les plus communément utilisés dans *CMS* sont : les algorithmes de "cône itératif", de "point central" (*midpoint*) et de " k_T " [62].

2.5.4.1 Les algorithmes de reconstruction des jets

Les algorithmes de reconstruction des jets utilisent les informations des deux sous-détecteurs ECAL et HCAL en les combinant pour former des tours calorimétriques. Comme la granularité du calorimètre ECAL est supérieur, les tours calorimétriques sont reconstruites en sommant les signaux contenus dans des bins (η, ϕ) correspondant à la taille et à la position des cellules de HCAL. L'énergie associée aux tours calorimétriques est calculée en sommant les énergies mesurées par toutes les cellules qui les composent. Du point de vue de la reconstruction des jets, les tours calorimétriques sont traitées comme des particules de masse nulle ayant une énergie correspondant à l'énergie des tours et une direction définie par le point d'interaction (vertex) et le centre de la tour.

Pour la reconstruction des jets, un schéma de recombinaison doit être choisi parmi deux schémas différents. Dans le schéma "d'énergie", les 4-vecteurs des constituants sont sommés et des jets massifs sont produits. Dans le schéma " E_T ", l'impulsion transverse des

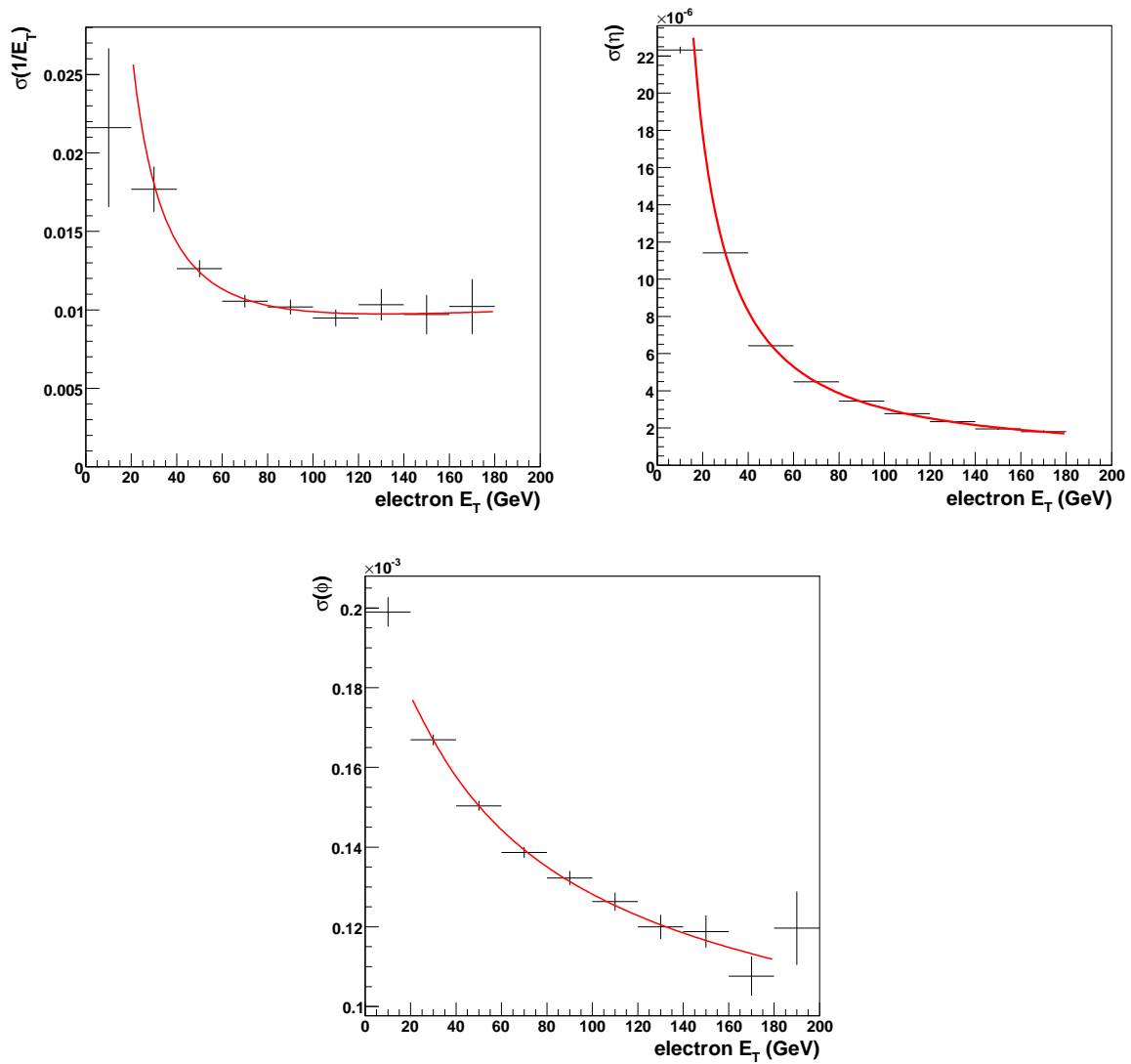


FIG. 2.17: Résolutions sur l'énergie transverse E_T (en haut à gauche), sur la pseudo-rapacité $|\eta|$ (en haut à droite) et sur l'angle azimutal ϕ (en bas) des électrons en fonction de l'énergie transverse E_T des électrons générés associés (de $E_T > 10$ GeV)

jets est définie comme la somme des énergies transverses des constituant $\sum E_T$, la direction des jets étant elle-même calculée de deux manières différentes : (1) $\sin\theta = \sum \vec{E}_T / E$ où E est l'énergie du jet ou (2) $\eta = E_{T,\eta_i} / \sum E_T$ et $\phi = E_{T,\phi_i} / \sum E_T$. La méthode (1) est préférentiellement utilisée pour les algorithmes de cône tandis que la méthode (2) est préférentiellement utilisé dans l'algorithme de k_T .

- **Algorithme du cône itératif** : une liste d'objets (particules ou tours calorimétriques) est créée et ordonnée par E_T décroissante. Les objets ayant une énergie transverse E_T supérieure à un certain seuil sont sélectionnés et un cône d'ouverture fixe R est créé autour de chacun d'eux. En utilisant les objets contenus dans chaque cône, des proto-jets sont créés (schéma E_T). La direction de ces proto-jets va être utilisée pour créer de nouveaux cônes et de nouveaux proto-jets. L'opération est répétée jusqu'à ce qu'entre chaque itération l'énergie change de moins de 1% et que le changement de direction soit inférieur à 0.01. Quand des proto-jets stables sont trouvés, les objets les constituant sont enlevés de la liste des objets et le proto-jet est ajouté à la liste des jets.
- **Algorithme "midpoint"** : l'algorithme midpoint est utilisé pour obtenir une meilleur prise en compte du recouvrement possible entre plusieurs jets. Une procédure similaire au cône itératif est appliquée, mais cette fois sans enlever aucun objet de la liste d'objets d'entrée. Il est alors possible d'obtenir plusieurs proto-jets contenant des objets communs. Pour toutes les paires de proto-jets plus proches que le diamètre d'un cône (typiquement 0.3), un proto-jet supplémentaire est créé en combinant les impulsions des proto-jets initiaux. Une fois cette étape terminée, si un proto-jet ne partage pas d'énergie avec un autre proto-jet, il est alors ajouté à la liste des jets. S'il partage plus de 50% de son énergie avec un autre proto-jet, les deux jets sont combinés pour former un seul jet. Sinon, il est associé au jet qui lui est le plus proches dans l'espace (η, ϕ) .
- **Algorithme k_T** : pour chaque objet ou paire d'objets de la liste, des distances d_i et $d_{i,j}$ sont calculées. Elles sont définies par :

$$d_i = (E_{T,i})^2 R^2, \quad (2.9)$$

$$d_{i,j} = \min\{E_{T,i}^2, E_{T,j}^2\} \Delta R_{i,j}^2, \quad (2.10)$$

où $E_{T,i}$ correspond à l'énergie de l'objet i , $\Delta R_{i,j} = \sqrt{(\varphi_i - \varphi_j)^2 + (\eta_i - \eta_j)^2}$ et R est un paramètre sans dimension généralement égal à l'unité. Les objets correspondant à la distance d la plus petite sont sélectionnés. S'il s'agit d'une distance de type d_{ij} , les objets sont regroupés en un nouvel objet. S'il s'agit d'une distance de type d_i , l'objet i est enlevé de la liste d'objets et ajouté à la liste de jets. La procédure est répétée jusqu'à ce que tous les objets se trouvent dans la liste de jets.

Dans cette thèse, l'algorithme utilisé est l'algorithme du cône itératif avec $\Delta R = 0.5$.

Différents effets peuvent dégrader la mesure de l'énergie des jets. Les pertes d'énergie et la présence de zones mortes dans le détecteur ou l'emprisonnement des particules de

bas p_T par le champ magnétique tendent à diminuer l'énergie mesurées des jets. La présence de particules provenant d'événements sous-jacents, d'empilement ou la présence de bruit électronique dans les calorimètres tendent à augmenter l'énergie des jets mesurés.

Pour corriger ces effets, différentes méthodes de calibration en énergie des jets ont été développées dans CMS et vont être brièvement discutées ici.

- **Corrections Monte Carlo** [62] : Les jets peuvent être reconstruits dans la simulation à partir des particules générées, excluant les neutrinos et les muons : nous les appellons “jet générés”. Un jet reconstruit sera associé à un jet généré si la distance angulaire ΔR qui les sépare est inférieure à 0.2. A partir des distributions du rapport entre l'énergie du jet reconstruit et l'énergie du jet généré $R_{jet} = E_T^{rec}/E_T^{MC}$, un facteur de calibration K peut être calculé. Ce facteur de correction est présenté sur le graphique de gauche de la figure 2.18 en fonction de η_{MC} pour différentes valeurs de E_T^{MC} et pour l'algorithme de cône itératif avec $\Delta R = 0.5$ dans des événements QCD . Il est particulièrement grand dans la partie centrale pour les jets de “basse” impulsion transverse. L'énergie corrigée du jet est donnée par $E_T^{corr} = E_T^{rec}/K$.
- **Equilibre di-jets** [63] : La méthode d'équilibre d'impulsion transverse di-jet est basée uniquement sur les données réelles. Elle utilise le fait que dans un événement QCD di-jets (tel que le processus $gg \rightarrow q\bar{q}$), les jets produits doivent avoir des énergies transverses voisines. Il s'agit d'un traitement statistique en faisant l'hypothèse que, pour l'énergie d'un parton donné, la valeur moyenne des énergies des jets reconstruits est égale à l'énergie du parton initial. Les événements ayant 2 jets de haut p_T dont le jet de plus haut p_T se trouve dans la région $|\eta| < 1$ sont sélectionnés. L'impulsion transverse de ce premier jet (p_T^{barrel}) est comparée à l'impulsion transverse du deuxième jet, p_T^{probe} . L'équilibre di-jets e est alors défini par :

$$e = 2 \frac{p_T^{probe} - p_T^{barrel}}{p_T^{probe} + p_T^{barrel}}. \quad (2.11)$$

La réponse relative di-jets est alors définie par le rapport $\langle e \rangle / (2 - \langle e \rangle)$ avec $\langle e \rangle$ représentant la valeur moyenne de e . Elle peut être utilisée pour déduire des fonctions de calibration paramétrisées en η . La figure de droite 2.18 représente la réponse relative di-jets dans des événements QCD ayant deux jets de hauts p_T dont l'impulsion transverse moyenne est comprise dans l'intervalle [120,250] GeV/c. On observe une différence significative de cette réponse en fonction de $|\eta|$. Cela est dû en partie à la différence de géométrie des tours calorimétriques entre la partie centrale et les bouchons.

- **Equilibre γ -jet** [64] : Au premier ordre, les photons émis dans les processus $qg \rightarrow q\gamma$ et $q\bar{q} \rightarrow g\gamma$ ont une impulsion transverse identique au parton associé. Grâce à la bonne précision de la mesure de l'énergie des photons dans le calorimètre ECAL (calibré en faisceau test et à l'aide des données cosmique [65]), il est possible de calibrer l'énergie des jets à l'aide de ces processus. Les constantes de calibration paramétrisées en fonction de l'impulsion transverse des photons sont simplement le

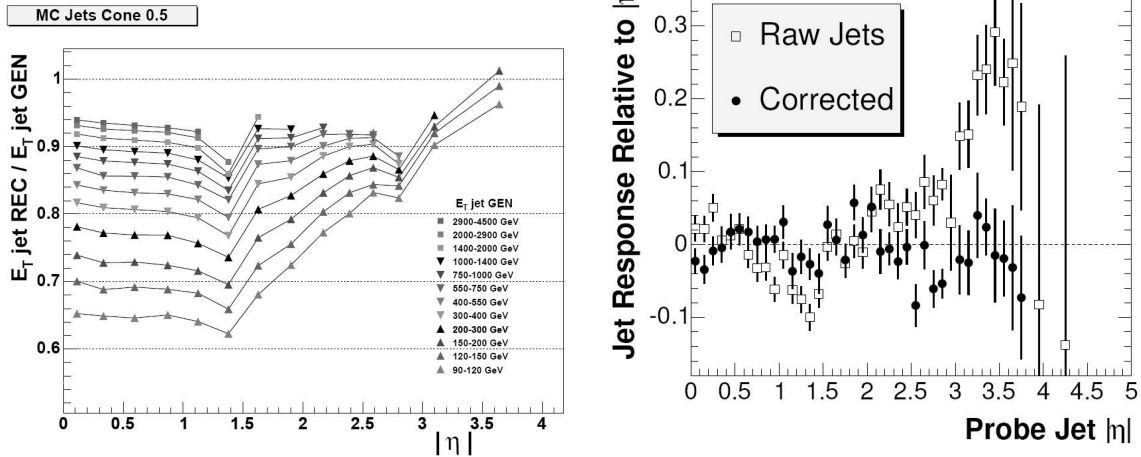


FIG. 2.18: A gauche, rapport des énergies transverses des jets reconstruits et des jets générées en fonction de la pseudo-rapidité des jets générés et pour différentes énergies transverses générées. L’algorithme du cône itératif avec $\Delta R = 0.5$ est utilisé dans des événements QCD . A droite, réponse relative di-jets dans des événements QCD ayant deux jets de haut p_T dont l’impulsion transverse moyenne est comprise dans l’intervalle $[120, 250]$ GeV/c.

rapport entre l’impulsion transverse des jets et des photons, p_T^{jet}/p_T^γ . La principale limitation de cette méthode est la contamination d’événements de bruits de fond QCD lorsqu’un faux photon est reconstruit.

- **Contrainte de la masse du W** [34] : Connaissant la masse du boson W avec une très bonne précision [28], les événements $t\bar{t}$ semi-leptoniques peuvent être utilisés pour extraire un facteur d’échelle sur l’énergie des jets en reconstruisant la masse invariante des jets issus de la désintégration hadronique de bosons W . Les bruits de fond peuvent être rejetés en sélectionnant les événements avec un muon isolé de haut p_T et 4 jets dont 2 jets issus de l’hadronisation de hadrons beaux. L’utilisation d’algorithme d’identification de jet de quark b (cf chapitre 4) permet de distinguer les jets provenant de bosons W et les jets b . Le graphique de gauche de la figure 2.19 montre la distribution en masse invariante des candidats jets provenant de bosons W lorsque ces jets sont ou ne sont pas correctement sélectionnés comme issus de W (la luminosité intégrée correspondante est de 5 fb^{-1}).

L’énergie des candidats jets issus de W est alors multipliée par un facteur $1+\Delta C$. La direction des jets n’est pas changée mais la valeur de l’impulsion p est modifiée de manière à ce que le rapport E/p reste inchangé. De nouveaux quadri-vecteurs sont déterminés et les masses invariantes des candidats di-jets issus de W sont recalculées. Une fonction gaussienne est ajustée sur cette dernière distribution afin de mesurer la masse du boson W . La procédure est répétée sur un large ensemble

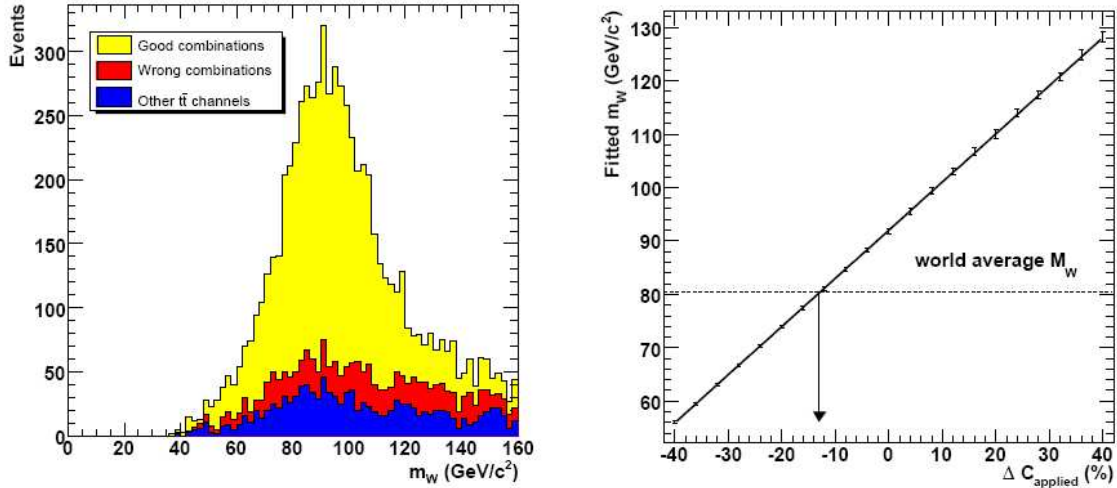


FIG. 2.19: A gauche, distribution en masse invariante des candidats di-jets provenant de bosons W lorsque que ces jets sont ou ne sont pas correctement sélectionnés comme issus de W , pour les événements $t\bar{t}$ et pour une luminosité intégrée de 5 fb^{-1} . A droite, évolution du résultat de l'ajustement $t\bar{t}$ gaussien des distributions an masse invariante des candidats di-jets issus de W pour différentes valeurs du facteur correctif ΔC .

de valeurs ΔC . Le graphique de droite sur la figure 2.19, représente l'évolution de la masse du boson W en fonction de ΔC . La valeur de ΔC correspondant à la masse du W réelle ($80.4 \text{ GeV}/c^2$ [28]) est utilisée comme facteur de calibration. L'incertitude sur le facteur de correction a été estimée à $\pm 0.6\%$ environ pour une luminosité de 1 fb^{-1} .

Dans les chapitres 4 et 5, les jets reconstruits avec l'algorithme du cône itératif $\Delta R = 0.5$ sont utilisés avec la correction en énergie Monte-Carlo.

Le graphique de gauche de la figure 2.20 représente les efficacités de reconstruction des jets corrigés à l'aide de la correction Monte-Carlo en fonction de l'impulsion transverse des jets générés associés pour la partie centrale ($|\eta| < 1.5$) pour des événements générés avec différents \hat{p}_T [66] et pour les algorithmes de cône itératif et midpoint. L'efficacité de reconstruction des jets est supérieure à 95% pour les jets ayant un p_T de plus de 30 GeV/c. Le graphique de droite de la figure 2.20 présente l'efficacité de reconstruction des jets en fonction de η pour des jets ayant un p_T compris entre 30 et 40 GeV/c. Pour l'algorithme du cône avec $\Delta R = 0.5$, elle est supérieure à 95% et légèrement moins efficace dans la partie centrale du détecteur.

La résolution des jets après corrections Monte-Carlo sur leur énergie est déterminée en ajustant une fonction gaussienne sur les distributions du rapport E_T^{reco}/E_T^{MC} pour différents bins en E_T^{MC} . La figure 2.21 montre l'évolution de la résolution en énergie des jets $\sigma(E_T^{reco}/E_T^{MC}) / \langle E_T^{reco}/E_T^{MC} \rangle$ en fonction de E_T^{MC} pour différentes régions

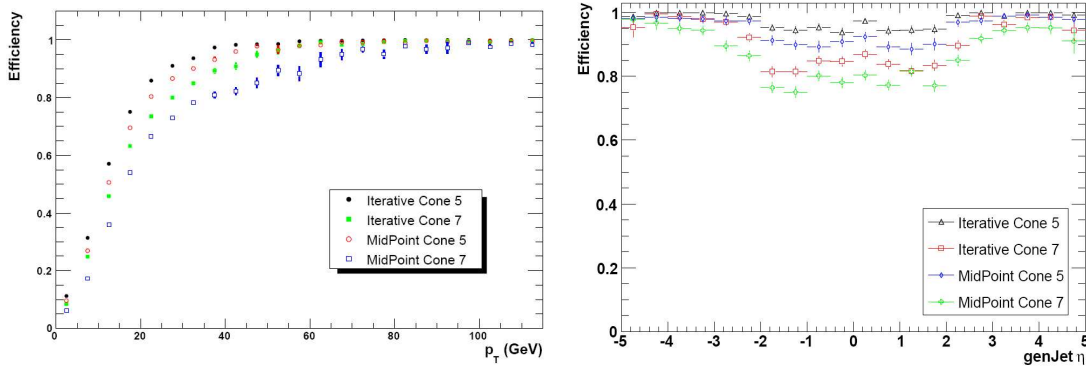


FIG. 2.20: A gauche, efficacité de reconstruction des jets en fonction de l'impulsion transverse des jets générés associés pour la partie centrale ($|\eta| < 1.5$). A droite, efficacité de reconstruction des jets ($30 < p_T < 40$ GeV/c) en fonction de la pseudo-rapidité des jets générés. Des événements QCD avec différents intervalles en \hat{p}_T ont été utilisés.

du détecteur. Des événements QCD ont été utilisés. Le graphique est ajusté par la fonctionnelle suivante (similaire à la formule 2.6) :

$$\frac{\sigma(E_T^{reco}/E_T^{MC})}{\langle E_T^{reco}/E_T^{MC} \rangle} = \frac{a}{E_T^{MC}} \oplus \frac{b}{\sqrt{E_T^{MC}}} \oplus c. \quad (2.12)$$

Dans la région centrale, le résultat de l'ajustement des données de la figure 2.6 par la fonction 2.12 sont $a = 5.6$ GeV, $b = 1.25$ GeV^{1/2} et $c = 0.033$. Les résolutions angulaires $\sigma(|\eta|)$ et $\sigma(\phi)$ sur la direction des jets sont déterminées de la même manière que pour les résolutions des électrons et des muons. Elles sont représentées sur la figure 2.22 pour des événements QCD . Les résolutions en $|\eta|$ et en ϕ sont inférieures à 2.5×10^{-2} et 7×10^{-2} respectivement pour des jets de $p_T > 30$ GeV/c. Les résolutions en énergie et en angle sont significativement moins bonnes que pour les électrons ou les muons reconstruits.

2.5.4.2 Reconstruction de l'énergie transverse manquante

La grande couverture angulaire de CMS permet de reconstruire indirectement l'énergie transverse de l'ensemble des particules qui ne déposent pas (ou très peu) d'énergie dans les calorimètres, tels que les neutrinos par exemple. Dans le plan transverse, la somme vectorielle des impulsions des particules doit être conservée. Ainsi, la non conservation de la somme des énergies transverses se traduit par une énergie transverse "manquante", révélatrice de la présence de particules interagissant très faiblement dans les calorimètres. Les événements sous-jacents ne contribuent en principe que très peu à l'énergie transverse manquante car les particules correspondantes sont produites à très grand $|\eta|$ et ne déposent que peu de leur énergie dans les calorimètres. Dans cette thèse l'énergie transverse manquante, notée \cancel{E}_T , est utilisée pour mettre en évidence la présence de neutrinos.

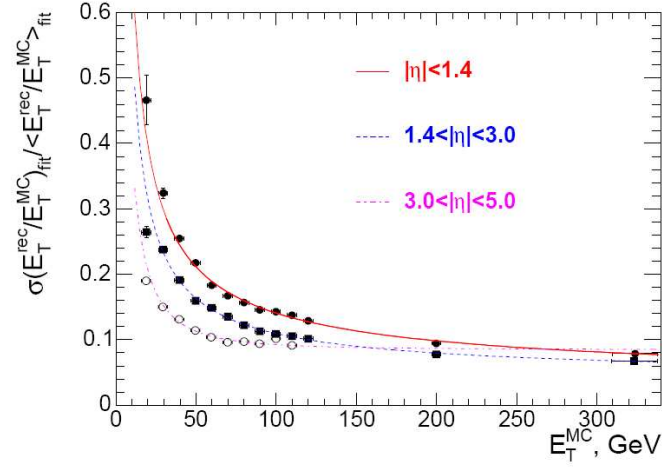


FIG. 2.21: Résolution en énergie transverse des jets reconstruits par l'algorithme de cône itératif $\Delta R = 0.5$ en fonction de E_T^{MC} dans la partie centrale $|\eta| < 1.4$, dans les bouchons $1.4 < |\eta| < 3$ et dans la région très avant $3 < |\eta| < 5$.

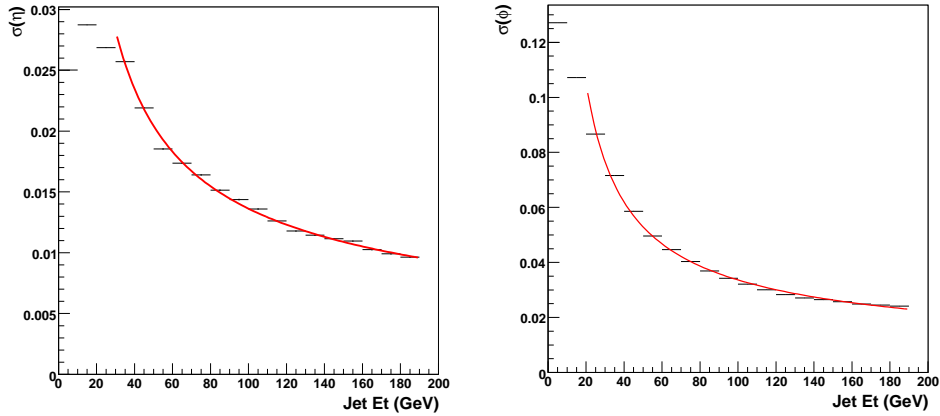


FIG. 2.22: Résolution en pseudo-rapidité (à gauche) et en angle azimutal (à droite) des jets reconstruits par l'algorithme du cône itératif $\Delta R = 0.5$ en fonction de E_T^{MC} pour des événements QCD .

L'énergie transverse manquante vectorielle est calculée en sommant l'énergie des tours calorimétriques E_n en tenant compte leurs angles θ_n et ϕ_n :

$$\vec{E}_T^{miss} = \sum (E_n \sin \theta_n \cos \phi_n \vec{i} + E_n \sin \theta_n \sin \phi_n \vec{j}). \quad (2.13)$$

$$= E_x^{miss} \vec{i} + E_y^{miss} \vec{j}. \quad (2.14)$$

Au LHC, la mesure du vecteur énergie transverse manquante \vec{E}_T^{miss} va être compliquée par la présence d'événements d'empilement. Dans CMS, la mesure sera encore plus difficile car il faut tenir compte des réponses différentes entre les calorimètres ECAL et HCAL et de la présence d'un champ magnétique de 4 T qui peut capturer des particules de bas p_T .

La résolution en énergie transverse manquante E_T [67] en fonction de la somme des énergies transverses des jets reconstruits $\sum E_T$ est représentée sur la figure 2.23 (à gauche) pour des événements QCD .

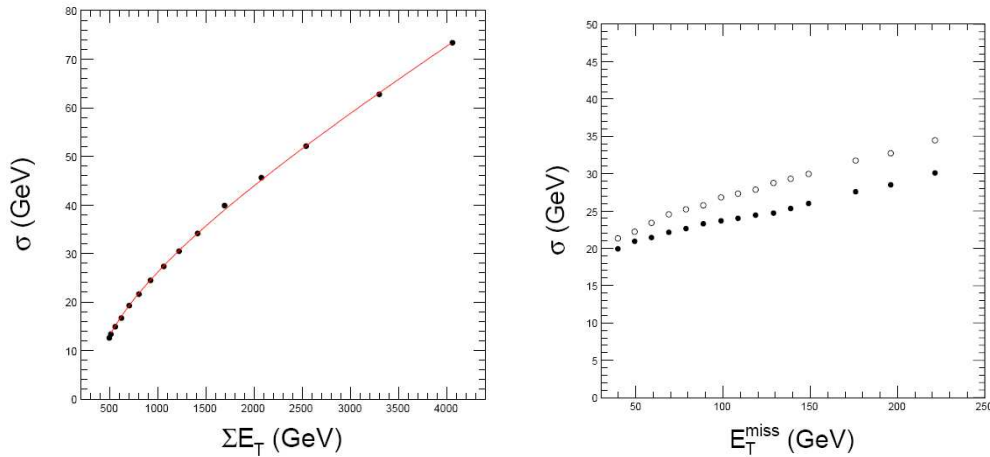


FIG. 2.23: A gauche, résolution en énergie transverse manquante E_T en fonction de la somme des énergies transverses des jets reconstruits $\sum E_T$ pour des événements QCD . A droite, résolution de l'énergie transverse manquante en fonction de E_T^{miss} avec et sans la correction en énergie des jets pour des événements $t\bar{t}$.

Afin d'améliorer la résolution en E_T , celle-ci sera corrigée à l'aide des corrections en énergie des jets. Il a été montré que cette procédure n'apporte une amélioration significative que pour les événements contenant de la vraie E_T (comme les événements $t\bar{t}$ di-leptons). Sur la figure 2.23, le graphique de droite présente la résolution sur l'énergie transverse manquante en fonction de E_T avec et sans la correction en énergie des jets pour des événements $t\bar{t}$.

Dans le chapitre 5, l'énergie transverse manquante sera utilisée avec la correction Monte-Carlo appliquée sur l'énergie des jets.

2.6 Acquisition, reconstruction et simulation des données

2.6.1 Le système de déclenchement

Le système de déclenchement (ou trigger) est le point de départ de la sélection des événements pour les analyses de physique. La décision de retenir ou non un événement doit se faire à une fréquence de 400 MHz (toute les 25 ns) correspondant à la fréquence de croisement des faisceaux.

Ce taux d'interaction très élevé doit être réduit à 100 Hz pour permettre l'acquisition des données. La réduction s'opère en deux étapes [68] :

- **Trigger de niveau 1 (L1)** : il correspond à la sélection d'objets reconstruits (muons, électrons, photons, jets) par des cartes électroniques ("hardware"), à partir d'une quantité d'information réduite afin de limiter le temps de sélection. En effet, le temps alloué pour la transmission des données aux calculateurs de L1 et pour la prise de décision est de $3.2 \mu s$. Après la sélection du trigger L1, le taux d'événements devrait être réduit à 100 kHz environ.

L1 utilise les informations provenant des chambres à muons (DT, RPC, CSC) et des calorimètres. Le spectromètre à muons est utilisé en combinant les informations des différentes chambres et les quatre muons les mieux reconstruits sont transférés au trigger de haut niveau. Du point de vue du calorimètre, les informations des tours calorimétriques (HCAL) et d'ensembles de 5×5 cristaux (ECAL) sont utilisés. La décision est alors basée sur une sélection des tours calorimétriques, des muons ou des candidats jets qui passent des seuils en p_T et E_T . Pour des raisons de temps de calcul, aucune information provenant du trajectographe n'est utilisée.

- **Trigger de Haut Niveau (HLT)** : il analyse les événements acceptés par le trigger L1. Il reconstruit des objets de complexités croissantes et applique différentes sélections d'événements (ou chemins de déclenchement) de manière à répondre aux besoins des analyses de physique (trigger "software"). Le taux d'événements doit ainsi être réduit à 100 Hz. Des exemples de chemins de déclenchement (incluant des sélections sur les muons et sur les électrons) sont présentées dans la section 5.2.2.

Les données brutes enregistrées par l'expérience seront transférées à un premier centre de calcul au CERN appelé *Tier 0*. Un premier traitement informatique y est effectué puis les données sont transférées à différents centres de calcul nationaux (appelés *Tier 1*, au nombre de 7) où la reconstruction complète des événements est effectuée. Les données reconstruites sont alors distribuées à une centaine de centres de calcul (appelés *Tier 2*, au nombre de 40) pour y être analysées.

Pour la gestion de cette structure, un système de partage des ressources informatiques a été mis en place : la grille de calcul LCG⁹. Ce système permet à un utilisateur de profiter des ressources informatiques de toute la collaboration. Courant 2007, un Tier2 a été mis en place et développé à l'IPHC jusqu'à atteindre des capacités de calcul importantes : plus de 680 processeurs sont disponibles et la capacité de stockage de données dépasse les 200 To. Au cours de cette thèse, le centre de calcul de l'IPHC a très largement été utilisé, permettant la validation du système pour les analyses de physique.

2.6.2 Reconstruction des événements et l'environnement d'analyse

Une fois que les données issues des collisions sont enregistrées, les différents programmes de reconstruction des objets physiques sont utilisés. L'ensemble de ces programmes définit un environnement de travail informatique basé sur le langage de programmation C++ orienté objets [69] et sur l'outil d'analyse Root [70]. Le modèle de données de CMS est l'*événement*. Un *événement* donne accès aux données enregistrées pour un seul croisement de faisceau. Cela inclut les données brutes digitisées, les données reconstruites ou les objets complexes de haut niveau utilisé pour les analyses.

Les différents objets physiques sont reconstruits à partir de la réponse du détecteur par l'ensemble d'algorithmes de reconstruction qui ont été présentés dans ce chapitre.

Depuis 2006, un nouvel environnement de travail a été mis en place : *CMSSW*. Différentes versions de cet environnement ont été utilisées au cours de cette thèse pour développer les analyses qui seront présentées.

Une partie du travail de thèse a consisté au développement de ce nouvel environnement. Plus particulièrement, un environnement d'analyse dédié à l'étude des événements *top* a été mis en place [71] au sein du groupe *top* de CMS. Cet environnement d'analyse (appelé TQAF¹⁰) consiste en un ensemble d'interfaces entre les différents objets de haut niveau reconstruits, le but étant de fournir aux utilisateurs des outils simples, flexibles et faciles d'utilisation. Dans ce cadre, l'interface entre les jets et les algorithmes d'identification des jets *b* a été développée et maintenue au cours de cette thèse. L'environnement TQAF a été utilisé dans l'analyse qui sera présentée dans le chapitre 5.

Dans les versions les plus récentes de *CMSSW*, TQAF a été étendu à l'ensemble de la collaboration et est devenu l'environnement d'analyse standard de CMS nommé PAT¹¹.

⁹LHC Computing Grid

¹⁰Top Quark Analysis Framework

¹¹Physic Analysis Toolkit

2.6.3 La simulation des événements

Dans la section 1.6, il a été montré comment des événements pouvaient être générés par des méthodes Monte-Carlo. Ces événements peuvent être combinés avec une simulation du détecteur (Geant4 [72]) de manière à estimer sa réponse et à développer les algorithmes de reconstruction. De tels événements peuvent également être utilisés comme référence dans les analyses de physique.

La simulation du détecteur [73] commence par la description détaillée de la géométrie de tous les sous-détecteurs. Les particules générées sont alors propagées du point d'interaction à travers le champ magnétique jusque dans la géométrie simulée du détecteur. A ce niveau, les particules peuvent se désintégrer et interagir avec la matière (diffusion multiple, interactions électromagnétique, interactions nucléaires). Cette étape conduit à la création de particules et de coups simulés.

La dernière étape de la simulation décrit la réponse du détecteur. Dans le trajectographe, les pertes d'énergie dans les détecteurs au silicium sont converties en dépôts de charges, (en introduisant des bruits). Le signal correspondant est alors digitisé et utilisé par les programmes de reconstruction. Dans les calorimètres, les énergies simulées déposées dans les scintillateurs sont converties en photo-électrons. Dans les chambres à muons, la simulation du passage d'un muon se traduit par la simulation des signaux correspondant aux points d'impact, à l'ionisation du gaz et au temps de dérive des charges déposées. Les positions des coups et des dépôts d'énergie simulés peuvent être dégradés artificiellement de manière à reproduire les problèmes d'alignement et de mauvaise calibrations du détecteur.

Les événements sont ensuite reconstruits avec les mêmes algorithmes que ceux utilisés sur les vraies données. Dans cette thèse, des événements Monte-Carlo (générés et simulés) ont été utilisés pour étudier les algorithmes d'identification des jets b et pour estimer la précision sur la mesure de la section efficace de production de paires de quarks top. Un environnement d'analyse dédié à la physique du top a également été utilisé [74].

Les différents sous-détecteurs ainsi que les différents algorithmes de reconstruction ont été présentés dans ce chapitre. Avant de pouvoir les utiliser dans les analyses de physique, un travail de compréhension de la réponse du détecteur est nécessaire. Un exemple d'analyse visant à étudier les efficacités de reconstruction des traces dans les données est présenté dans le chapitre 3.

3

Estimation des efficacités de reconstruction des traces à l'aide de muons cosmiques.

Les premières collisions proton-proton délivrées par le LHC auront lieu en septembre 2009. Les données enregistrées seront utilisées pour comprendre la réponse des différents sous-détecteurs mais aussi pour les aligner et les calibrer. Cette étape essentielle de compréhension et de préparation du détecteur est appelée *commissioning*.

Les travaux de *commissioning* ont cependant déjà débutés avec les séquences de prise de données “cosmiques” : les muons issus de la désintégration de rayons cosmiques dans l’atmosphère peuvent être détectés et enregistrés. Ces muons cosmiques sont utilisés pour étudier le système de déclenchement, la détection des muons dans le trajectographe, dans les chambres à muons et dans les calorimètres. Les muons cosmiques peuvent également être utilisés pour aligner [75] et calibrer [76, 77] les sous-détecteurs et pour étudier les différents algorithmes de reconstruction.

Dans ce cadre, il sera présenté une analyse originale des algorithmes de reconstruction des muons cosmiques dans les chambres à muons d’une part et dans le trajectographe d’autre part. La méthode d’estimation des efficacités de reconstruction des traces et des muons cosmiques sera présentée dans la section 3.2. Les efficacités de reconstruction des muons locaux et leur résolution en impulsion transverse seront discutées dans la section 2.4.2. Dans cette section, une étude des associations entre traces et muons locaux sera effectuée. Finalement, les efficacités de reconstruction des traces seront présentées dans la section 3.5. Avant cela, les algorithmes de reconstruction et les lots de données utilisés seront exposés dans la section 3.1.

Dans cette analyse, les muons locaux et les traces reconstruits dans la partie centrale du détecteur ont été étudiées.

3.1 Algorithmes de reconstruction des traces pour les données cosmiques

Trois différents algorithmes de reconstruction des traces sont utilisés pour les études en mode de prise de données cosmiques [78]. Les algorithmes *CombinatorialTrackFinder* (*CTF*, cf. section 2.3.2.1) et *Road Search* (*RS*) ont été initialement développés pour la reconstruction des traces issues de collisions et ont dû être adaptés au cas de données cosmiques, où la topologie des événements est très différente :

- les muons proviennent de la désintégration de rayons cosmiques dans l'atmosphère. Ils sont donc produits verticalement et se dirigent vers le bas du détecteur alors que les traces issues de collisions proton-proton sont produites au centre du détecteur suivant un large éventail de directions et se dirigent vers l'extérieur du détecteur. Les algorithmes *CTF* et *RS* doivent donc être adaptés pour permettre la reconstruction de traces verticales qui ne passent pas par le centre du détecteur.
- Dans les collisions, les traces sont reconstruites du point de collision jusqu'à la couche la plus externe du trajectographe. Dans les données cosmiques, les muons traversent le trajectographe de part en part. Les traces doivent donc être reconstruites en prenant en compte à la fois l'hémisphère supérieur et l'hémisphère inférieur du trajectographe.
- Dans les collisions, la reconstruction des graines se fait dans les régions les plus proches du tube du faisceau. Dans les données cosmiques, la reconstruction des graines doit permettre la reconstruction des traces dont la trajectoire ne traverse pas le tube du faisceau.

Même si les événements cosmiques ne sont constitués en majorité que d'une seule trace, leur reconstruction se révèle difficile. Le trajectographe a été conçu de manière à optimiser la couverture spatiale autour du tube du faisceau, afin de maximiser l'efficacité de reconstruction des traces issues du vertex primaire. A cause de cette géométrie, les traces de muons cosmiques peuvent traverser les espaces situés entre deux couches successives. Elles peuvent également traverser des zones avec un grand nombre de modules superposés.

Un algorithme dédié à la prise de données cosmiques, l'algorithme *CosmicTrackFinder* (noté *CosmicTF* ou *Cosmic*) a également été développé [78]. Ces trois algorithmes suivent le schéma de reconstruction des traces présentées dans la section 2.3.2.1 et seront étudiés et comparés.

3.1.1 Adaptation de l'algorithme *Combinatorial Track Finder* (CTF) à la prise de données cosmiques

Cet algorithme est présenté dans la section 2.3.2.1. Dans le mode de reconstruction cosmique, la recherche des graines doit permettre la reconstruction des traces incompatibles avec le point d'interaction. Les spécificités de la recherche de graines sont les suivantes :

- aucune contrainte sur la présence d'un vertex ne doit être demandée,
- les graines doivent pouvoir être reconstruites dans les couches du trajectographe les plus externes car ces couches ont de plus grandes acceptances pour les muons cosmiques.

Une nouvelle procédure de recherche de graines a été définie : des triplets de coups (ensemble de trois coups) situés dans différentes couches de la partie centrale du trajectographe sont utilisés (les couches correspondantes sont TOB 3, 4, 5 et 6, cf figure 2.5), ainsi que des triplets de coups espacés par moins de deux disques dans les bouchons .

A partir de ces graines, la reconstruction des traces suit le schéma présenté dans la section 2.3.2.1. Comme les muons cosmiques peuvent traverser l'ensemble du trajectographe du haut vers le bas, la reconstruction permet la propagation des traces entre les hémisphères supérieurs et inférieurs du trajectographe.

3.1.2 Présentation de l'algorithme *CosmicTrackFinder*

L'algorithme *CosmicTrackFinder* (ou *CosmicTF*) est dédié à la reconstruction de traces uniques sans imposer de région d'origine mais en supposant une direction privilégiée : l'axe vertical.

Comme le nombre de coups reconstruits dans les événements cosmiques est bien moins important que dans les données de collisions, la recherche des graines se fait à partir de tous les triplets de coups géométriquement compatibles dans le trajectographe. Un triplet de coups compatibles est défini comme :

- des triplets de coups provenant de différentes couches (partie centrale) ou roues (bouchons). Dans la partie centrale, toutes les combinaisons de triplet de coups dans les trois couches *TOB* les plus externes ou les trois couches *TIB* les plus internes sont considérées. Dans les bouchons, tous les triplets de coups sont utilisés si les coups ne sont pas séparés par plus d'une roue.
- la distance entre les coups suivant l'axe des z ne doit pas excéder 30 cm. Cette coupure n'est pas appliquée dans le *TEC*.
- la distance entre les coups suivant l'axe global des x doit être inférieure à deux fois la distance entre les coups suivant l'axe global des y . Cette coupure est motivée par le fait que les muons cosmiques doivent avoir un angle faible avec l'axe vertical. Elle n'est pas appliquée dans le *TEC*.

Une fois les graines reconstruites, un filtre de Kalman (cf section 2.3.2.1) simplifié est utilisé : si une graine est reconstruite dans la partie supérieure (inférieure) du trajectographe, tous les coups reconstruits avec une coordonnée globale y inférieure (supérieure) aux coups de la graine sont ordonnés en y décroissant (croissant). La sélection des coups compatibles suit alors le schéma suivant :

1. La trajectoire est propagée à la surface du module correspondant au coup testé. L'incertitude provenant de la diffusion multiple est prise en compte lors de la propagation.
2. La compatibilité des coups avec la trajectoire propagée est évaluée.
3. Si le coup est compatible, la trajectoire est reconstruite en le prenant en compte.

Après l'étape d'ajustement des trajectoires, plusieurs traces peuvent être reconstruites. Comme une seule trace par événement est attendue, une seule trace sera retenue. Les critères de sélection des traces sont (par ordre d'importance) :

- le plus grand nombre de couches ayant un coup reconstruit associé à la trace,
- le plus grand nombre de coups associés à la trace,
- la plus petite valeur du χ^2 de l'ajustement.

Pour être sélectionnée, une trace doit avoir au moins trois coups associés.

3.1.3 Présentation de l'algorithme *RoadSearch*

L'algorithme *Road Search* [78] est initialement dédié à la reconstruction des traces issues de collisions proton-proton. Comme le champ magnétique est orienté suivant l'axe du faisceau, les traces produites dans les collisions ont des trajectoires rectilignes dans l'espace $r - z$ (où r est la coordonnées radiale de la particule dans le plan transverse $r = \sqrt{x^2 + y^2}$). L'algorithme *Road Search* utilise alors un ensemble d'anneaux prédéfinis, appelé *Roads* ou routes, compatibles avec une ligne dans l'espace $r - z$. Le trajectographe est ainsi découpé en anneaux (contenant un ensemble de modules) avec une position $r - z$ donnée, ϕ allant de 0 à 360°. Ces routes d'anneaux vont être utilisées pour reconstruire les traces.

La reconstruction des graines utilise les paires de coups reconstruits dans un ensemble d'anneaux spécifiques appelés anneaux graines. Ces derniers sont constitués des deux couches les plus internes du trajectographe, des deux couches les plus externes ainsi que des couches TOB 1 et 2. Les anneaux graines sont présentés sur la figure 3.1. Les graines sont alors des paires de coups dans les anneaux graines : un coup doit être reconstruit dans un anneau interne et un autre dans un anneau externe. L'ensemble des anneaux utilisés pour définir les routes est déterminé par l'extrapolation linéaire des deux anneaux graines dans l'espace $r - z$. Une coupure sur la différence d'angles $\Delta\phi$ maximum entre les coups constituant la graine est appliquée. Elle correspond à une coupure en énergie transverse sur les traces.

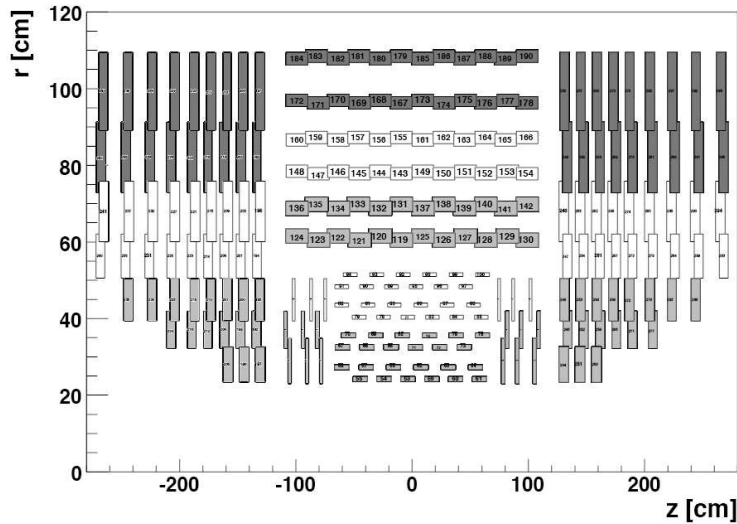


FIG. 3.1: Vue d'ensemble de la structure en anneaux du trajectographe. Les différents rectangles représentent les anneaux définis par l'algorithme *Road Search*. Les anneaux utilisés pour la recherche des graines sont colorés en gris clair pour les couches internes et en gris foncé pour les couches externes.

Une route est déterminée à partir de la graine et de contraintes liées au centre du détecteur. Dans les données cosmiques, cette contrainte a été relâchée. L'ensemble des coups compatibles avec le chemin $r - \phi$ est déterminé par une méthode itérative. La trace est finalement ajustée avec l'ensemble des coups compatibles.

Cet algorithme est, de par sa conception, spécialement dédié au cas des traces produites dans les collisions. Il n'a pas été complètement optimisé pour la reconstruction des muons cosmiques en champ magnétique (2 coups seulement sont utilisés pour définir les graines) ce qui conduira à des efficacités de reconstruction plus faibles que pour les autres algorithmes.

3.2 Sélection des muons cosmiques

La méthode présentée ici utilise de manière indépendante l'information des chambres à muons et du trajectographe. Lorsqu'un muon est reconstruit dans les chambres à muons, l'algorithme recherche s'il y a une trace reconstruite dans le trajectographe qui peut lui être associée. Pour chacune de ces traces, sa trajectoire est extrapolée au module compatible le plus supérieur (inférieur) du trajectographe si le muon considéré a été reconstruit dans l'hémisphère supérieur (inférieur) du détecteur à muons. La trajectoire du muon local correspondant est extrapolée à la même surface. Les extrapolations sont faites suivant l'équation du mouvement d'une particule chargée dans un champ magné-

tique. Des critères d'association peuvent alors être utilisés pour associer les traces et les muons.

On peut ainsi mesurer l'efficacité de reconstruction des traces dans le trajectographe indépendamment de toute information provenant du trajectographe lui-même. L'estimation des efficacités de reconstruction des traces reste néanmoins biaisée : les chambres à muons ont une plus grande couverture spatiale que le trajectographe. L'efficacité mesurée va donc dépendre de la définition de l'acceptance du trajectographe. En effet, pour assurer une bonne efficacité de reconstruction des graines et la restructibilité des traces, la simple géométrie du trajectographe ne peut être utilisée.

De manière à prendre en compte ces effets d'acceptance ainsi que la qualité des muons locaux, différentes coupures sont appliquées sur les muons.

- **Qualité des muons locaux :** De manière à rejeter les muons mal reconstruits, les muons locaux sélectionnés doivent posséder un χ^2 normalisé inférieur à 10 et doivent être associés à au moins 28 coups dans les chambres à muons. Cette dernière coupure correspond à la traversée d'au moins deux chambres à muons.

Certains muons reconstruits de bas p_T peuvent être capturés par le champ magnétique ($p_T \lesssim 5 \text{ GeV}/c$) ou perdre une grande quantité de leur énergie dans les calorimètres. Ces muons n'atteignent pas le trajectographe et induisent des biais dans les mesures d'efficacité de reconstruction des traces. C'est pourquoi les événements sélectionnés doivent contenir un muon reconstruit dans l'hémisphère supérieur du détecteur et un autre dans l'hémisphère inférieur. Ces deux muons locaux correspondent au passage d'un seul muon cosmique. Des critères supplémentaires d'association angulaires entre ces deux muons sont appliqués :

- la différence $|\Delta\phi|$ entre les muons locaux doit être $< 0.045 \text{ rad}$,
- la différence $|\Delta\eta|$ entre les muons locaux doit être < 0.06 ,
- la différence $|\Delta d0|$ entre les muons locaux doit être $< 15 \text{ cm}$,
- la différence $|\Delta dz|$ entre les muons locaux doit être $< 20 \text{ cm}$.

où ϕ est l'angle azimutal, η la pseudo-rapacité, $d0$ et dz sont les paramètres d'impact transverse et longitudinaux par rapport au centre du détecteur. Les coupures ont été déterminées à partir d'événements Monte-Carlo. Le paramètre d'impact est défini comme la distance entre le centre du détecteur et le point de la trajectoire de la trace qui lui est le plus proche. Afin d'éviter toute ambiguïté, seuls les événements contenant exactement 2 muons locaux passant les coupures précédemment énoncées sont retenues.

- **Compatibilité avec la trajectographe :** Comme la couverture des chambres à muons est plus importante que celle du trajectographe, il faut s'assurer que la trajectoire d'un muon local traverse effectivement le trajectographe. La trajectoire des muons locaux reconstruits dans l'hémisphère supérieur des chambres à muons est propagée jusqu'à un cylindre dont l'axe est aligné à l'axe des faisceaux. La propagation suit les équations du mouvement des particules chargées dans un champ

magnétique en prenant en compte les effets de perte d'énergie dans le détecteur. Le rayon du cylindre est de 85 cm et sa longueur de 220 cm. Son centre correspond au point d'interaction de manière à couvrir la région $-110 < z < 110$ cm.

La surface du cylindre correspond ainsi à la couche TOB 4 du trajectographe (cf. figure 2.5). Afin de s'assurer que la reconstruction des graines est efficace, les muons sélectionnés sont ceux traversant ce cylindre de part en part.

Les muons peuvent perdre une fraction significative de leur énergie dans le trajectographe et dans les calorimètres. Cela se traduit par une dégradation des résolutions angulaires des muons inférieurs comparés aux muons supérieurs. Afin de minimiser cet effet, les critères d'association entre les muons locaux et les traces sont appliqués uniquement entre une trace et un muon de l'hémisphère supérieur des chambres à muons (rappelons que deux muons locaux ont été sélectionné, l'un dans l'hémisphère supérieur et l'autre dans l'hémisphère inférieur). Dans la suite de ce chapitre, la sélection d'événements qui vient d'être présentée est appelée sélection **Pour la Recherche des Traces** ou *PRT*.

Une méthode similaire est utilisée pour estimer l'efficacité de reconstruction des muons locaux. L'algorithme sélectionne les événements contenant exactement une trace reconstruite dans le trajectographe et recherche les éventuels muons locaux reconstruits dans l'hémisphère supérieur des chambres à muon. N'étant pas soumise à d'éventuels biais liés à l'acceptance du trajectographe, une telle méthode sera utilisée pour étudier les critères d'association entre muons locaux et traces reconstruites dans le trajectographe. Des coupures de qualité sont appliquées de manière à rejeter les fausses traces ou les traces mal reconstruites : au moins 8 coups doivent être associés aux traces et le χ^2 normalisé de l'ajustement doit être inférieur à 5. Dans la suite de ce chapitre, cette sélection d'événements sera appelée sélection **Pour la Recherche des Muons** ou *PRM*, elle sera utilisée dans la section 3.4.

A partir de ces deux sélections, l'efficacité de reconstruction des traces ou des muons locaux est définie par :

$$\varepsilon_{tracking} = \frac{\text{Nbr. d'événements sélectionnés avec (un muon local + une trace associée)}}{\text{Nombre d'événements sélectionnés}}. \quad (3.1)$$

3.3 Données cosmiques et simulation Monte-Carlo

Les données utilisées pour cette étude ont été enregistrées pendant la prise de données *Craft*¹. Lors de cet exercice, les rayons cosmiques détectés par les chambres à muons ont

¹Cosmic Run At Four Tesla

été utilisés pour déclencher l'enregistrement des données. Le système de déclenchement est principalement basé sur le trigger de niveau 1 des chambres *DT* et *RPC* (cf. section 2.4.1). De septembre à décembre 2008, près de 400 millions d'événements ont été collectés en présence d'un champ magnétique de 3.8 Tesla. Ils sont utilisés pour aligner [75] et calibrer [76, 76] les différents sous-détecteurs.

Dans cette analyse, les données vont nous permettre d'étudier l'efficacité de reconstruction des traces dans le trajectographe ainsi que les différences en impulsion transverse entre muons locaux et traces.

Une analyse similaire a été effectuée dans le cas de données sans champ magnétique (données *Cruzet*, Cosmic Run At Zero Telsa) mais ne sera pas présentée ici.

Les périodes d'acquisition de données sont appelées *runs*. Durant ces périodes d'acquisition, des dysfonctionnements des différents sous détecteur (par exemples des problèmes d'alimentation électrique, de refroidissement ou de synchronisation) peuvent dégrader la qualité des données. Avant toute analyse, il est donc nécessaire de sélectionner les runs pour lesquels le détecteur était dans un bon état de fonctionnement.

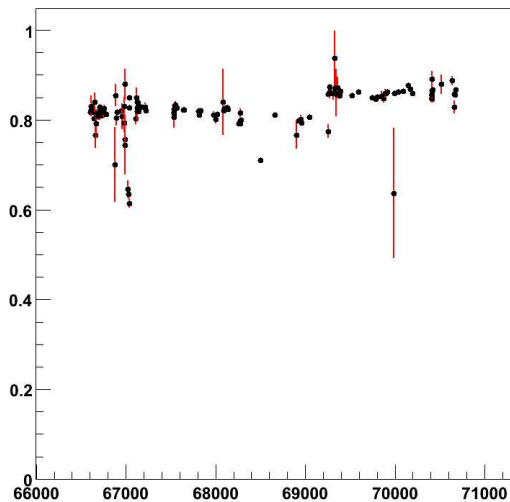


FIG. 3.2: Efficacité de reconstruction des traces en fonction du numéro de *run*, sans coupure d'association traces-muons locaux et dans le cas de la sélection *PRT*.

Cette sélection est faite en évaluant l'efficacité (3.1) avec la sélection *PRT* pour chaque run, lorsqu'aucun critère d'association entre les traces et les muons locaux n'est utilisé. La figure 3.2 présente les efficacités en fonction du numéro de *run*. Ces efficacités peuvent varier de plus de 20% même si elles restent relativement stables d'un run à l'autre ($\geq 80\%$).

On remarque également une augmentation de l'efficacité moyenne de quelques pourcent pour les runs > 69245 . Dans les runs < 69245 , un problème de synchronisation entre le système de déclenchement et la lecture des données dans le trajectographe a été découvert. Il conduisait à des inefficacités de reconstruction des coups dans l'hémisphère supérieur du trajectographe. C'est pourquoi, les runs utilisés dans cette analyse sont ceux dont les numéros sont supérieurs à 69245 et qui ont une efficacité $> 80\%$.

La figure **3.3** illustre les distributions de différentes variables liées aux traces reconstruites pour les algorithmes *CTF*, *RS* et *CosmicTF* lorsque la sélection *PRT* est utilisée. Le graphique en haut à gauche représente la distribution en η . Pour une trace donnée, η est estimé au point de la trajectoire le plus proche du centre du détecteur. Les muons cosmiques ont une direction proche de l'axe vertical, la majorité des traces doivent donc posséder un η proche de 0. On remarque cependant un excès de traces pour $\eta \approx 0.2$ qui correspond à la coordonnée η du puits de la caverne de CMS. Les muons cosmiques traversant le puits ont plus de chance de ne pas être arrêtés par la roche et d'atteindre CMS.

On remarque également que le nombre de traces reconstruites diffère d'un algorithme à l'autre. Les nombres de traces reconstruites avec les algorithmes *CTF* et *CosmicTF* sont comparables alors que le nombre de traces reconstruites avec l'algorithme *RS* est significativement plus faible.

Le graphique en haut à droite sur la figure **3.3** représente la distribution en ϕ des traces reconstruites. Pour une trace donnée, l'angle ϕ est estimé au point de la trajectoire de la trace qui est le plus proche du centre du détecteur. Les angles ϕ sont négatifs car les muons cosmiques traversent le détecteur du haut vers le bas. Le sens de propagation des traces est alors dirigé suivant les ϕ négatifs. A cause de défauts de reconstruction qu'il reste à comprendre, une faible fraction de traces (environ 1.5% pour l'algorithme *CTF*) peuvent posséder une direction ϕ positive. De plus, l'allure globale des distributions en ϕ reste mal comprise pour l'instant. On s'attend à observer des maxima se situant autour de $\phi = \pi/2$, qui correspond à l'axe vertical. Les distributions en ϕ montrent cependant une structure en double pic qui serait liés à des problèmes d'alignement : pour cet échantillon, le détecteur a été aligné avec des données prises sans champ magnétique. Les effets du champ magnétique (comme les angles de Lorentz) ne sont pas pris en compte ici.

Sur la figure **3.3**, le graphique en bas à gauche représente les distributions en paramètre d'impact d_0 des traces. Comme le nombre de couches du trajectographe compatible avec un muon local est plus grand dans la région centrale du détecteur ($|d_0| < 20$ cm), l'efficacité de reconstruction y est plus grande et le nombre de traces reconstruites plus important. Le graphique en bas à droite représente les distributions en impulsion transverse des traces. A cause des muons stoppés dans le détecteur ou retenus par le champ magnétique, les efficacités de reconstruction des traces chutent pour $p_T < 5$

GeV/c. Un maximum de traces est reconstruit pour un $p_T \approx 10$ GeV/c, le nombre de traces produites diminue ensuite avec le p_T .

Sur la figure 3.4, le graphique de gauche présente les distributions en χ^2 normalisé de l'ajustement des traces. On remarque que l'ajustement est significativement meilleur pour l'algorithme *CTF* que pour les algorithmes *RS* et *CosmicTF*. Le graphique de droite sur cette même figure présente les distributions en nombre de coups associés aux traces. Ils sont significativement plus faibles pour l'algorithme *RS* que pour les autres algorithmes. Pour cet algorithme, la recherche des coups compatibles se fait à partir d'une trajectoire définie par une paire de coups (graine). Cela correspond à une approximation linéaire des trajectoires des traces qui n'est valable qu'à grand p_T ou quand le champ magnétique est nul.

Dans les données cosmiques, le nombre de coups associés aux traces peut être significativement plus grand que dans les données issues des collisions *proton – proton*. En effet, le nombre de couches du trajectographe traversées est plus grand dans les données cosmiques. Le nombre de coups peut également être plus grand que le nombre de couches compatibles car deux coups par couche *stereo* peuvent être reconstruits (cf. section 2.3.2.1). Le recouvrement entre les modules de détection d'une même couche permet également de reconstruire plusieurs coups par couche.

Dans cette analyse, les résultats obtenus avec les données *Craft* seront comparés à ceux obtenus par une simulation Monte-Carlo. Le générateur Monte-Carlo *CMSCGEN* [79, 80] est utilisé pour produire des muons cosmiques. Dans ce générateur, les muons sont produits par la désintégration de protons dans la haute atmosphère. A cause de la conservation de la charge électrique, le nombre d'antimuons produits sera plus grand que le nombre de muons ($N^{\mu^+}/N^{\mu^-} \approx 1.33$). Le flux de muons correspondant est calculé au niveau du sol et des muons cosmiques sont générés en surface au dessus de CMS avec des impulsions comprises entre 3 GeV/c et 3 TeV/c. Les angles entre l'axe zénithal et les muons générés sont compris entre 0° et 84° . Les muons sont ensuite propagés jusqu'à la surface de CMS en prenant en compte leur perte d'énergie dans la roche. La réponse du détecteur est alors simulée et les algorithmes de reconstruction standards sont utilisés.

Les graphiques en haut sur la figure 3.5 représentent les distributions en η et ϕ des muons cosmiques reconstruits dans les données et dans la simulation Monte-Carlo pour l'algorithme *CTF*. On observe un bon accord entre données réelles et le Monte-Carlo excepté pour la partie centrale des distributions en ϕ (structure en double pics). Le graphique en bas à gauche de la même figure représente les distributions en paramètre d'impact transverse d_0 . On observe également un bon accord entre données et Monte-Carlo. D'une manière générale, les variables angulaires sont bien décrites par le Monte-Carlo. Le graphique en bas à gauche sur la figure 3.5 représente les distributions en impulsion transverse. Le spectre en p_T des muons cosmiques n'est pas bien décrit dans la simulation Monte-Carlo. Cela peut provenir de l'imperfection du modèle théorique

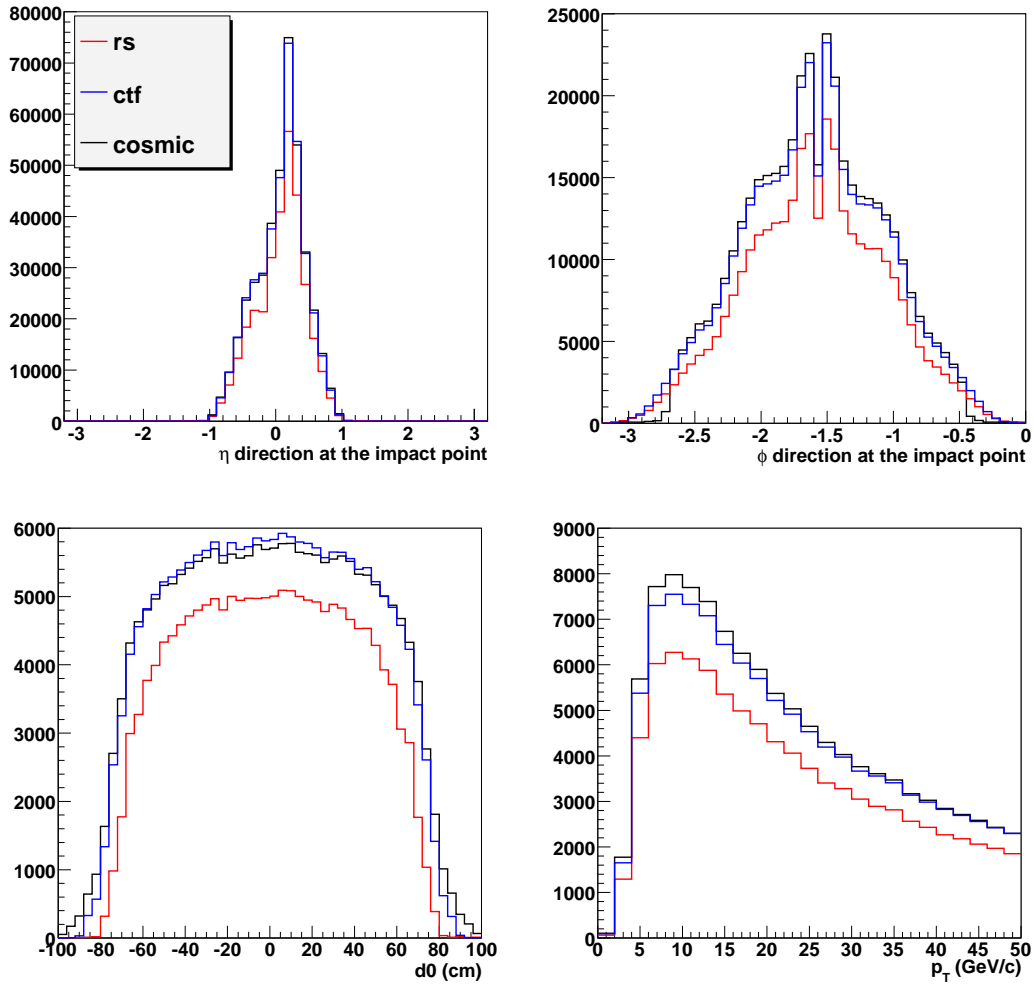


FIG. 3.3: Dans les données *Craft*, distribution de la pseudo-rapidité η (en haut à gauche), de l'angle azimutal ϕ (en haut à droite), du paramètre d'impact transverse d_0 (en bas à gauche) et du p_T (en bas à droite) des traces de muons reconstruits dans le trajectographe. Pour une trace donnée, η et ϕ sont estimés au point de la trajectoire de la trace qui est le plus proche du centre du détecteur. Les histogrammes rouges correspondent à l'algorithme *RS*, les bleus à l'algorithme *CTF* et les noirs à l'algorithme *CosmicTF*.

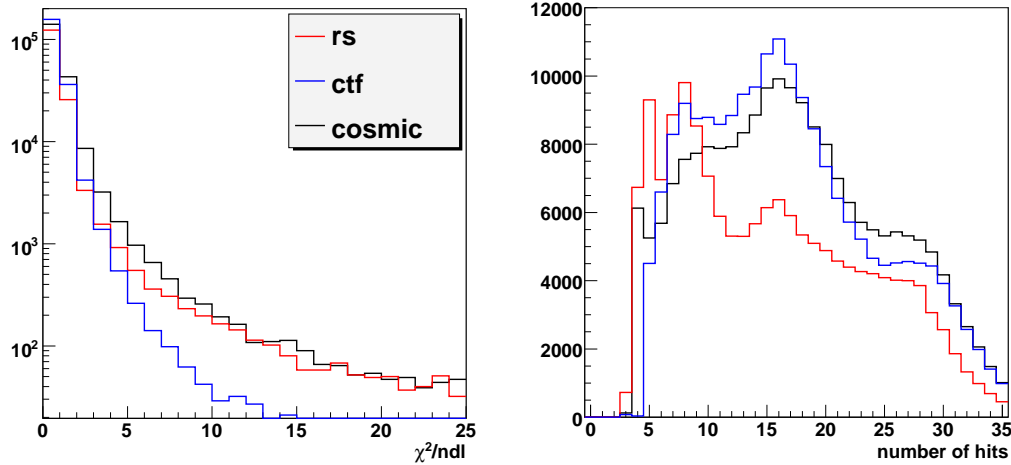


FIG. 3.4: Dans les données *Craft*, distribution du χ^2 normalisé (à gauche) et nombre de coups associés (à droite) aux traces de muons reconstruits dans le trajectographe.

de la génération ou d'un mauvais ajustement des paramètres. Les muons générés ont globalement des p_T plus faibles dans le Monte-Carlo ce qui tend à diminuer les efficacités globales de reconstruction des traces dans le Monte-Carlo par rapport aux données réelles. Nous verrons dans la suite que cet effet n'est pas prépondérant.

Sur la figure 3.6, le graphique de gauche représente le χ^2 normalisé de l'ajustement des traces dans les données et dans le Monte-Carlo. On observe un bon accord entre ces deux distributions. Le graphique de droite présente une comparaison du nombre de coups associés aux traces. On remarque également qu'il y a un bon accord entre données et Monte-Carlo. A première vue, il semblerait que la simulation du détecteur reproduise correctement les données dans le trajectographe.

Les figures 3.7 et 3.8 présentent des comparaisons entre données et Monte-Carlo pour les variables liées aux muons locaux. La figure 3.7 montre les distributions en η , ϕ , $d0$ et p_T . Comme pour les traces, la simulation Monte-Carlo semble bien décrire les distributions angulaires des données, excepté pour la partie centrale de la distribution en ϕ . Le p_T des muons reconstruits est également significativement plus faible dans le Monte-Carlo que dans la simulation.

Sur la figure 3.8, le graphique de gauche représente les distributions du χ^2 normalisé de l'ajustement des muons locaux et celui de droite le nombre de coups (dans les chambres à muons) associés aux traces. On remarque que la qualité des traces dans le Monte-Carlo est largement surestimée. L'efficacité de la recherche de coups compatibles à une traces graine est également surestimée dans la simulation. Les raisons de l'ineffica-

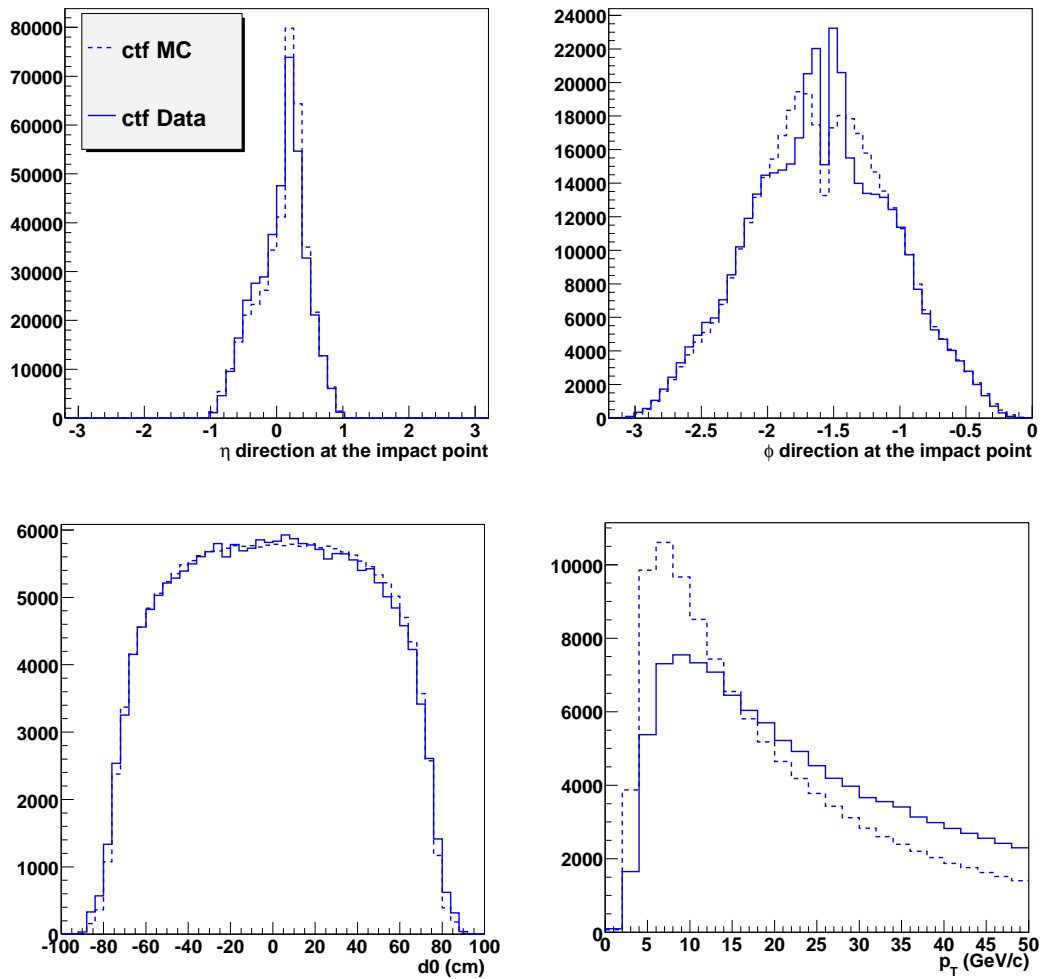


FIG. 3.5: Distribution en η (en haut à gauche), en ϕ (en haut à droite), en d_0 (en bas à gauche) et en p_T (en bas à droite) des traces de muons reconstruites par l'algorithme *CTF* dans les données (trait plein) et dans le Monte-Carlo (trait pointillé). Pour une trace donnée, η et ϕ sont estimés au point de la trajectoire de la trace qui est le plus proche du centre du détecteur. Les distributions sont normalisées aux mêmes nombres d'événements.

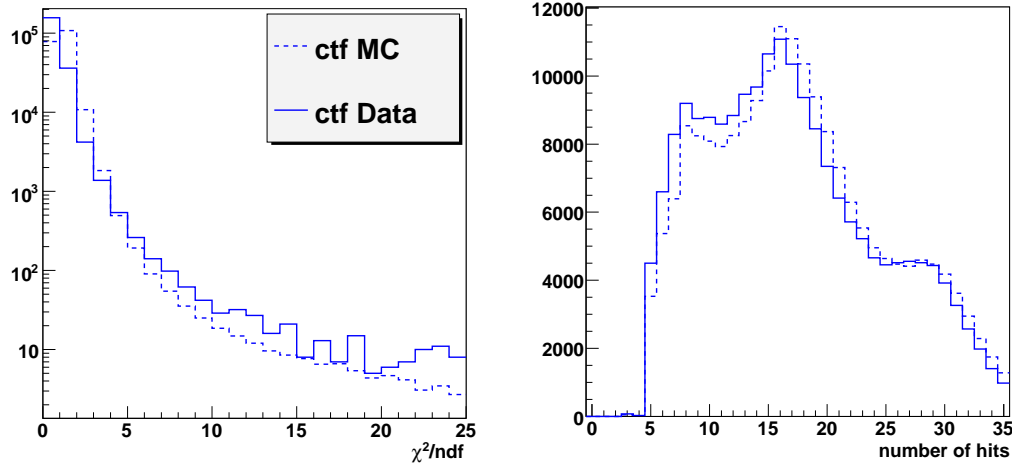


FIG. 3.6: Distribution du χ^2 normalisé (à gauche) et nombre de coups associés aux traces de muons reconstruits (à droite) par l'algorithme *CTF* dans les données (trait plein) et dans le Monte-Carlo (trait pointillé). Les distributions sont normalisées aux mêmes nombres d'événements.

citée de la recherche de coups peuvent être : des inefficacités de reconstruction de coups dans les chambres à muons (de telles inefficacités n'ont pour l'instant pas été observées), une mauvaise description du champ magnétique dans cette région du détecteur qui peut conduire à une mauvaise propagation des trajectoires ou des effets d'alignement mal reproduits (le détecteur est parfait dans la simulation).

Dans le lot d'événements Monte-Carlo utilisé, l'information générée n'est pas présente. Une étude des efficacités de reconstruction des traces à l'aide de l'information du générateur n'a donc pas encore pu être effectuée.

3.4 Etude comparative des muons locaux et des traces reconstruites dans le trajectographe

Dans cette section, les premiers résultats d'étude comparative des muons locaux et des traces reconstruites seront présentés. Ensuite, l'association entre muons locaux et traces reconstruites dans le trajectographe sera étudiée en faisant l'hypothèse que le taux de fausses traces et le taux de faux muons sont négligeables. Les différences en impulsion transverse entre les traces et les muons locaux seront également étudiées.

Dans cette étude, la sélection *PRM* sera utilisée (cf. section 3.2). A partir des événements contenant une trace, l'algorithme recherche des muons locaux reconstruits dans l'hémisphère supérieur des chambres à muon. Cette sélection permet d'éviter d'éventuels

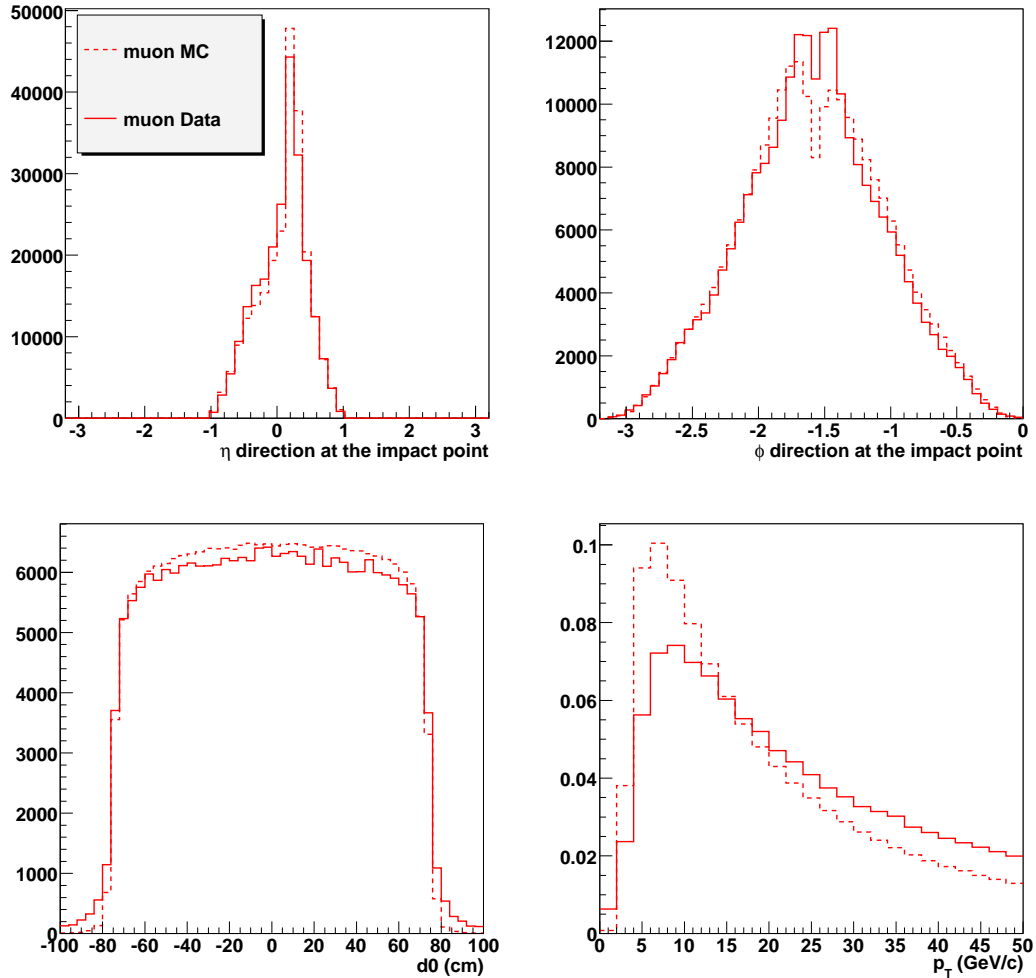


FIG. 3.7: Distribution en η (en haut à gauche), en ϕ (en haut à droite), en d_0 (en bas à gauche) et en p_T (en bas à droite) des muons locaux reconstruits. Pour une trace donnée, η et ϕ sont estimés au point de la trajectoire de la trace qui est le plus proche du centre du détecteur.

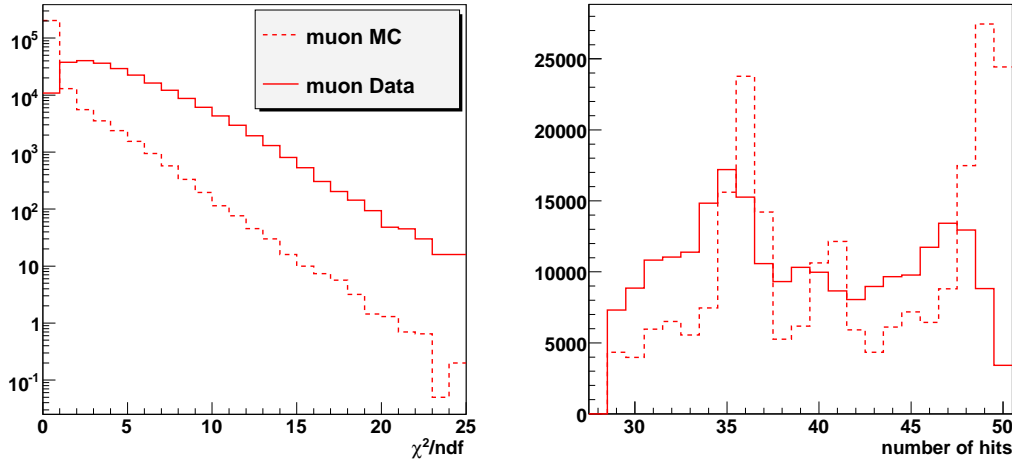


FIG. 3.8: Dans les données *Craft*, distribution du χ^2 normalisé (à gauche) et nombre de coups (à droite) associés aux muons locaux reconstruits.

biais provenant de l'acceptance du trajectographe par rapport aux chambres à muon.

Dans la section 2.3.2.1, nous avons vu que les paramètres d'une trace peuvent être estimés au niveau de chaque coup qui lui est associé. Ces coups correspondent à une certaine localisation (x, y, z) dans le détecteur. Il est important de comparer les paramètres d'un muon local et d'une trace dans une même région de détecteur et pour ce faire d'extrapoler les trajectoires de la trace et du muon local dans cette région. Les extrapolations des trajectoires sont réalisées à l'aide des équations du mouvement des particules chargées dans un champ magnétique en prenant en compte les pertes d'énergie dans les différents sous-détecteurs. Cette méthode est schématisée sur la figure 3.9.

Ainsi, pour chaque événement sélectionné contenant une trace reconstruite, l'état de la trace reconstruite est estimé au niveau du module du trajectographe qui contenant le coup reconstruit qui a le plus grand y . Ce module sera appelé *module compatible le plus externe* dans la suite. Il définit une surface plane tangente à la couche du trajectographe. Le système de coordonnées locales du module est alors utilisé : l'axe Y est l'axe parallèle aux pistes de silicium (parallèle à l'axe des faisceaux z) et l'axe X est l'axe perpendiculaire à Y de manière à former un repère direct positionné au centre du module. La trajectoire du muon local est propagée jusqu'à la surface de ce module en prenant en compte les pertes d'énergie et la diffusion multiple. Les paramètres des trajectoires seront comparés sur cette surface.

La figure 3.10 représente les positions locales X et Y des traces au niveau de leur module compatible le plus externe. La grande majorité des positions se trouve dans la région $|X| < 10$ cm et $|Y| < 10$ cm car elle correspond à la géométrie d'un module du

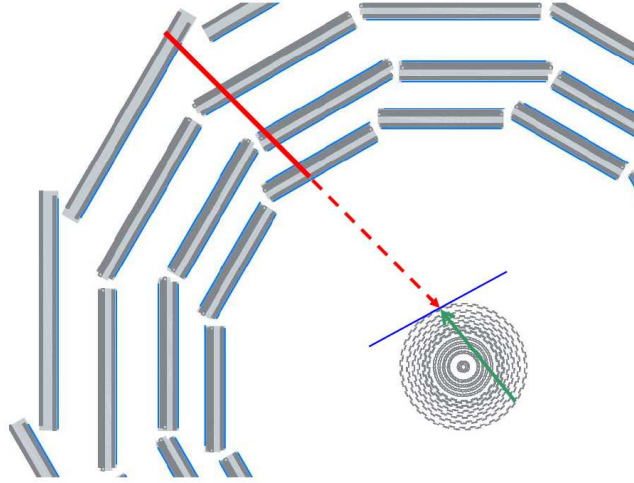


FIG. 3.9: Schéma représentant la propagation des muons locaux et des traces vers la couche compatible la plus externe.

trajectographe. Une petite fraction des traces ont leurs positions situées hors de cette région. En effet, bien que le coup de la trajectoire soit reconstruit sur le module, après l'ajustement de la trace, la position de la trajectoire peut se trouver hors de la surface du module.

Dans le cas de l'extrapolation des muons sur cette même surface (graphique de droite sur cette même figure) on remarque que la fraction de trajectoires de muons qui ne sont pas compatibles avec la région $|X| < 10$ cm et $|Y| < 10$ cm est plus importante. Cet effet est dû à l'erreur sur la propagation des muons.

Cet effet est également illustré sur la figure 3.11 : le graphique de gauche représente la position en X de la trace en fonction de la position en Y du muon, le graphique de droite représente la position en Y de la trace en fonction de la position en Y du muon. Bien que qu'il y a de fortes corrélations entre les traces et les muons, on remarque que la position des muons peut se situer à plusieurs dizaines de centimètres de la position des traces.

La figures 3.12 présente les distances entre traces et muons locaux au niveau du module compatible le plus externe en fonction du p_T de la trace : $\Delta X = X_{trace} - X_{muon}$, $\Delta Y = Y_{trace} - Y_{muon}$ et $\Delta R = \sqrt{\Delta X^2 + \Delta Y^2}$. Les distances diminuent avec le p_T des traces car à grand p_T les courbures des trajectoires sont moins fortes et les résolutions angulaires sont meilleures (moins de diffusion multiple). L'erreur sur la propagation des muons locaux est donc réduite. Il est clair que les distances ΔX peuvent être très supérieures aux distances ΔY . Le champ magnétique, qui agit dans le plan transverse, dégrade significativement la résolution sur la position X et augmente l'erreur sur la propagation des muons locaux dans le plan transverse. De ce fait, il s'avère difficile de

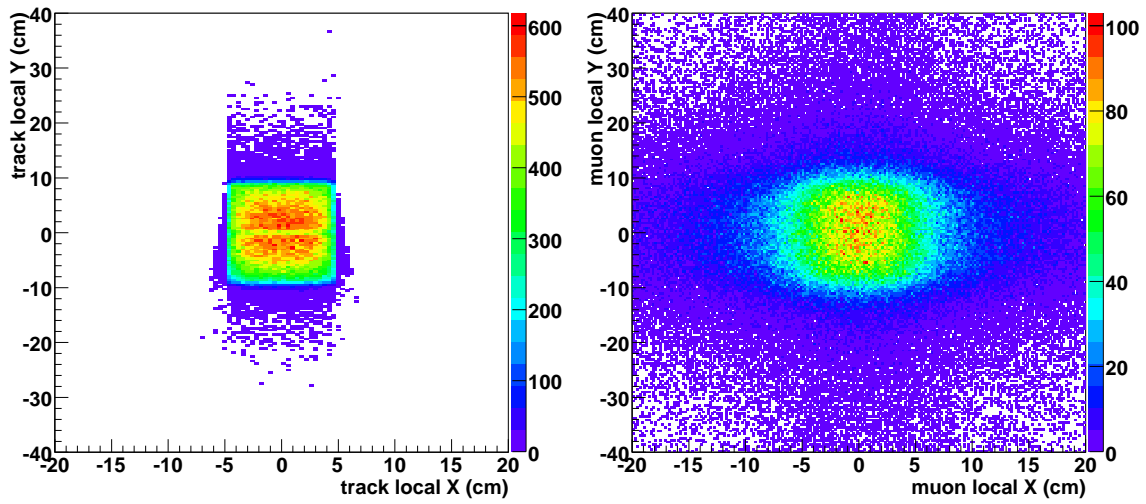


FIG. 3.10: Positions locales Y en fonction de la position locale Y des traces (à gauche) et des muons locaux (à droite) au niveau de leur module compatible le plus externe.

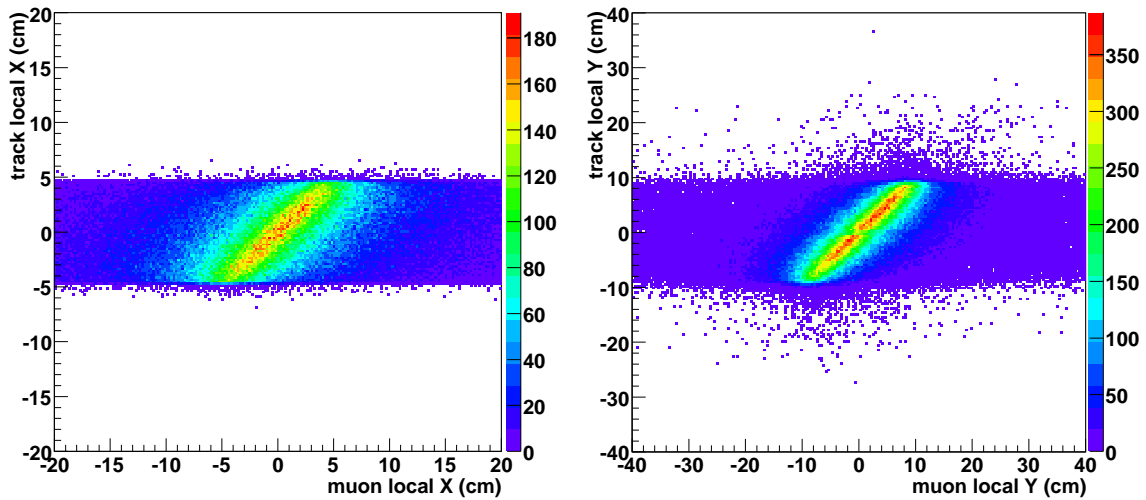


FIG. 3.11: Positions locales X des traces en fonction de la position locale X des muons locaux (à gauche) et positions locales Y des traces en fonction de la position locale Y des muons locaux (à droite) au niveau du module compatible le plus externe.

définir des critères géométriques d'association entre muons locaux et traces.

Les effets introduits par le champ magnétique peuvent être mis en évidence en comparant les impulsions transverses entre traces et muons locaux. Pour la sélection *PRT*, le ΔR est calculé sur le module compatible le plus externe. Si l'on considère que la résolution en p_T des traces est bien meilleure que celle des muons locaux (car le trajectographe a une meilleure précision et le champ magnétique est uniforme dans le solénoïde et donc plus facile à estimer), on peut mesurer la résolution en impulsion transverse de muons locaux par rapport au p_T des traces.

Afin de minimiser les ambiguïtés liées à la présence possible de plusieurs traces par événement, un muon local sera comparé à la trace qui lui est la plus proche sur le module compatible le plus externe (ΔR le plus petit).

Pour la sélection *PRT*, les impulsions transverses des traces et des muons locaux sont comparés en étudiant la distribution de leur différence relativement au p_T des traces. Pour différents intervalles en p_T , les distributions suivantes sont calculées :

$$R_1 = \frac{p_T^{muon} - p_T^{trace}}{p_T^{trace}}, \quad (3.2)$$

$$R_2 = \frac{1/p_T^{muon} - 1/p_T^{trace}}{1/p_T^{trace}}. \quad (3.3)$$

Les distributions en R_1 et R_2 sont illustrés sur la figure **3.13**.

Ces distributions sont ensuite ajustées par une fonction gaussienne. Le graphique en haut à gauche sur la figure **3.14** représente la moyenne de l'ajustement de R_1 pour différents intervalles de p_T des traces et pour les différents algorithmes de reconstruction. Les erreurs correspondent aux erreurs de l'ajustement de la gaussienne. Comme les distributions R_1 ne sont pas parfaitement gaussiennes (cf. figure **3.13**), les barres d'erreurs sont relativement importantes. On peut cependant remarquer que les impulsions transverses des muons locaux sont systématiquement plus grandes que les impulsions transverses des traces. Cet effet peut provenir de la perte d'énergie des muons à la traversée du détecteur (effet majoritaire à bas p_T) ou à une mauvaise description du champ magnétique dans les régions extérieures au solénoïde.

Ce comportement est confirmé par le graphique en haut à droite sur la figure **3.14**. Il représente les moyennes de l'ajustement gaussien de R_2 pour l'algorithme *CTF* et pour différents intervalles en p_T des traces. Les données réelles et Monte-Carlo sont comparées. Les erreurs sont moins importantes ici car les distributions R_2 sont très bien décrites par des fonctions Gaussiennes. Les valeurs moyennes des distributions R_2 sont significativement inférieures à 0 pour les données, même pour de grands p_T . Dans le Monte-Carlo, les valeurs moyennes sont également inférieures à 0 pour de faibles p_T mais se rapprochent de 0 pour les grands p_T .

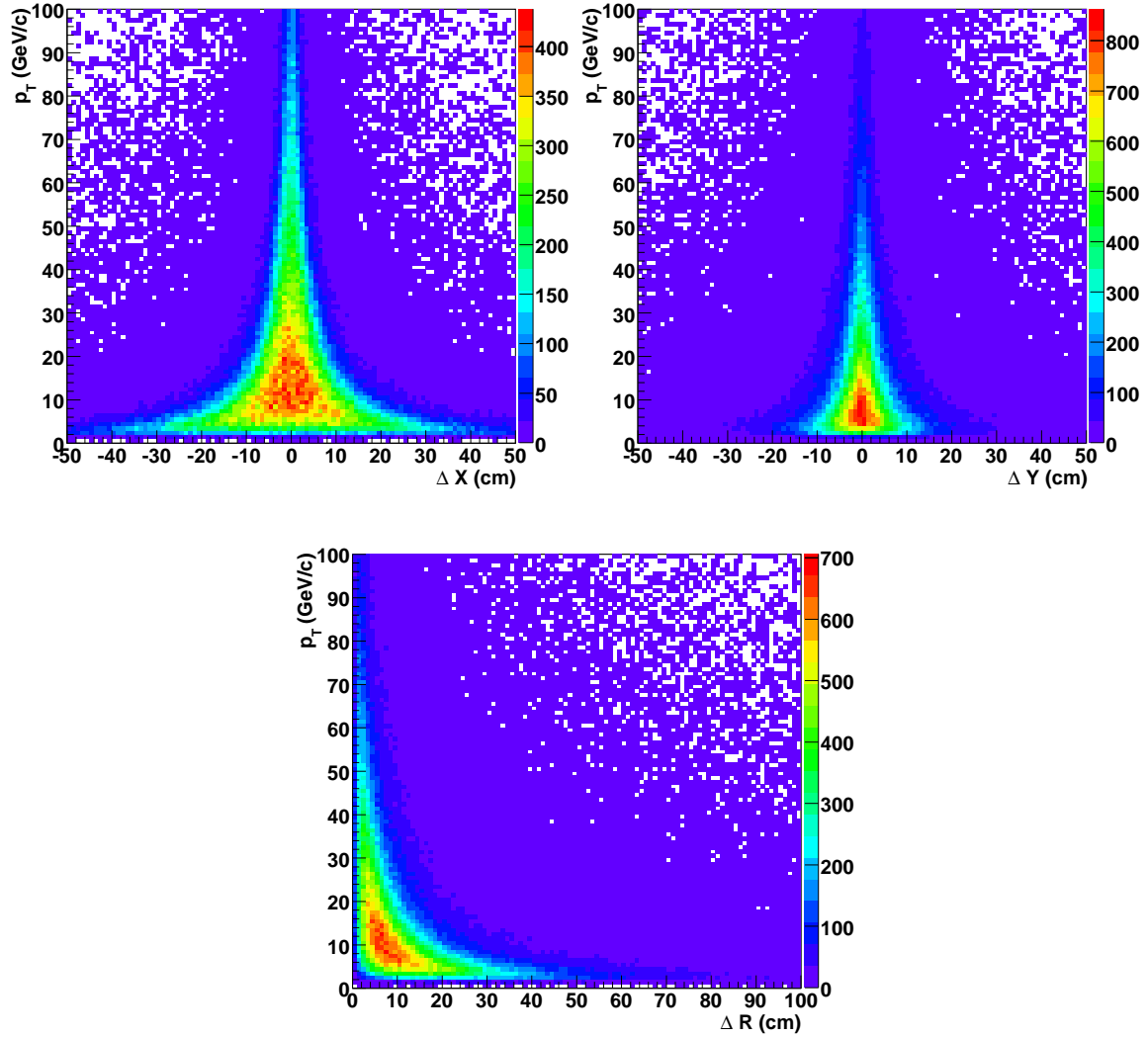


FIG. 3.12: Différences entre les positions locales X (en haut à gauche), Y (en haut à droite) et distances entre ces positions ΔR pour les trajectoires extrapolées au niveau du module compatible le plus externe.

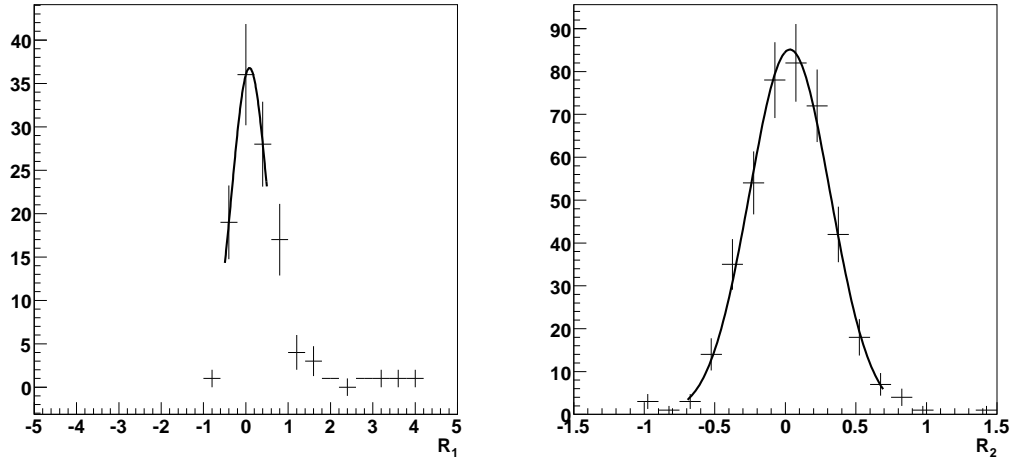


FIG. 3.13: Distributions en R_1 et R_2 pour des traces de $p_T \approx 10$ GeV/c. Les distributions sont ajustés par un fonction gaussienne.

Si l'on considère que les valeurs moyennes négatives dans le Monte-Carlo proviennent principalement de pertes d'énergie à la traversée du détecteur, les différences entre données et Monte-Carlo suggèrent que le champ magnétique n'est pas bien estimé dans les données réelles.

Cet effet proviendrait de la mauvaise estimation du champ magnétique dans les chambres à muons. Le champ magnétique dans le solénoïde est relativement bien connu et bien calculé. Par contre, hors du volume du solénoïde, le champ magnétique est semblait-il surestimé. Cela se traduit par une surestimation des impulsions transverses des muons locaux. Des effets de pertes d'énergie de d'alignement peuvent intervenir mais ne suffisent apparament pas à expliquer les différences observées.

Au moment de la rédaction de cette thèse, les effets induits par la méconnaissance du champ magnétique hors du solénoïde sont activement étudiés. L'analyse qui a été présentée précédemment a contribué à la mise en évidence de ces problèmes.

3.5 Estimation de l'efficacité de reconstruction des traces dans les données *Craft*

A partir de la sélection PRT , les efficacités de reconstruction des traces ont été estimées. Comme nous l'avons vu, il est délicat d'appliquer des critères d'association basés sur les distances entre les position locales ou sur les impulsions transverses. De tels critères biaiserait les résultats. D'autres critères d'association sont actuellement étudiés,

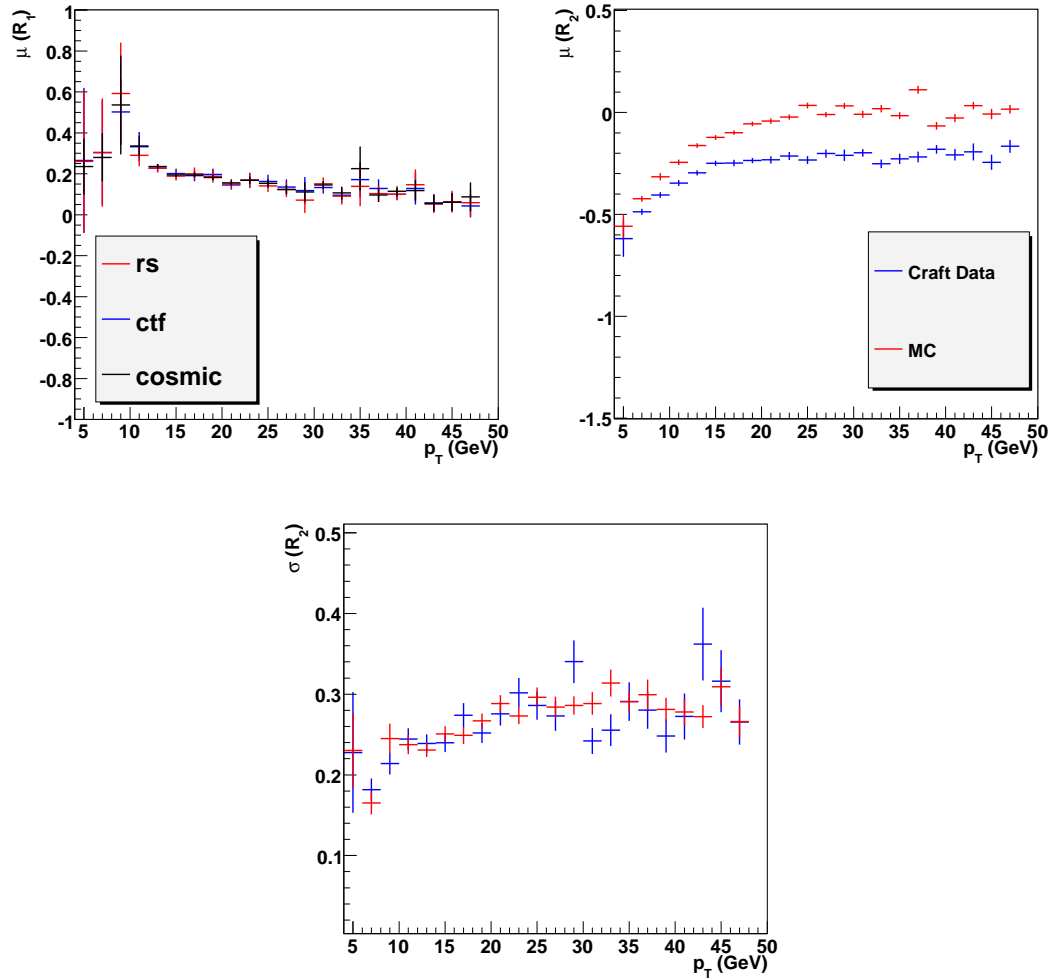


FIG. 3.14: Moyennes des ajustements de R_1 (en haut à gauche), de R_2 (en haut à droite) et déviation standard de R_2 (en bas) en fonction de l'impulsion transverse des traces.

comme la comparaison des positions locales à la surface des chambres à muons. L'incertitude sur la position des traces serait plus faible que l'erreur due à la propagation devraient être réduites. L'erreur introduite par la présence du champ magnétique serait cependant encore présente.

Dans l'estimation des efficacités de reconstruction, nous avons considéré que le taux de fausses traces est faible. Cette approximation est justifiée par la figure 3.15 montrant les multiplicités de traces par événement après la sélection *PRT* et pour les différents algorithmes de reconstruction.

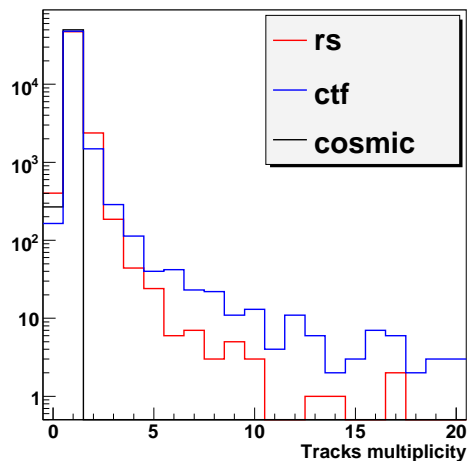


FIG. 3.15: Multiplicité de traces après la sélection *PRT* pour les algorithmes *RS*, *CTF* et *CosmicTF*.

Pour les algorithmes *CTF* et *RS*, on observe que le nombre d'événements contenant plus d'une trace est de deux ordres de grandeur plus faible que le nombre d'événements contenant une seule trace. On s'attend donc à ce que les contributions de fausses traces incluses dans le calcul des efficacités soient de l'ordre de quelques pourcent. De plus, une analyse indépendante tendrait à démontrer qu'à cause de problèmes de reconstruction, un seul muon cosmique pourrait engendrer plusieurs traces le long de sa trajectoire. Ceci est conforté par le fait que l'algorithme *CosmicTF*, qui ne reconstruit qu'une seule trace par événement (cf. section 3.1.2), ait une efficacité de reconstruction comparable à l'algorithme *CTF*, comme nous le verrons.

Les efficacités de reconstruction des traces sont alors évaluées par le rapport entre le nombre d'événements passant les coupures *PRT* contenant au moins une trace reconstruite et le nombre total d'événements passant les coupures *PRT*. Cette estimation est faite pour les algorithmes de reconstruction *CTF*, *CosmicTF* et *RS*.

Cependant, les erreurs sur la propagation des muons locaux peuvent introduire un biais important. Lors de la sélection *PRT* (cf section 3.2), les candidats muons cosmiques sont déterminés en vérifiant la compatibilité des muons locaux avec un cylindre équivalent à la couche TOB4. Pour ce faire, la trajectoire des muons locaux est propagée jusqu'à ce cylindre. La figure 3.16 correspond à la distribution des erreurs sur les positions locales X_{TOB4} et Y_{TOB4} des trajectoires extrapolées au cylindre TOB4. Les erreurs peuvent être plus grandes que les dimensions du trajectographe. Elles sont également plus grandes pour les positions locales en X (situé dans le plan transverse) que pour les positions locales en Y (suivant l'axe du faisceau) car le champ magnétique dégrade la précision de la propagation dans le plan transverse.

Ainsi, la trajectoire d'un muon local propagé peut être compatible avec le cylindre TOB4 alors que la vraie trajectoire du muon cosmique n'atteint pas le trajectographe. Une coupure supplémentaire sur l'erreur sur la position locale X_{TOB4} propagée a été ajoutée à la sélection *PRT*. Comme cette quantité est corrélée à l'impulsion transverse des muons, une coupure très relâchée est utilisée : l'erreur sur X_{TOB4} ($\sigma(X_{TOB4})$) doit être inférieure à 220 cm (ce qui correspond à la longueur de la partie centrale du trajectographe).

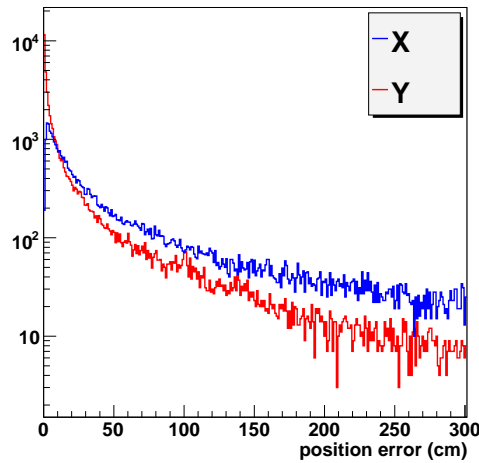


FIG. 3.16: Distribution de $\sigma(X_{TOB4})$ et $\sigma(Y_{TOB4})$ pour des événements après l'application de la sélection *PRT*.

Les efficacités ainsi estimées sont présentées dans le tableau 3.1. Dans les données, les efficacités de reconstruction sont de l'ordre de 90% pour les algorithmes *CTF* et *CosmicTF* et d'environ 70% pour l'algorithme *RS*. Les efficacités correspondantes sont significativement plus élevées dans le Monte-Carlo : de 3 à 4% pour les algorithmes *CTF* et *CosmicTF* et de 16% pour l'algorithme *RS*. Les raisons de ces différences peuvent

être : un mauvais ajustement des paramètres de la simulation et du générateur, une mauvaise description du champ magnétique dans les données, les différences d'alignement entre données et Monte-Carlo ou encore une mauvaise description de la quantité de matière dans le détecteur.

Algorithmes	<i>CTF</i>	<i>CosmicTF</i>	<i>RS</i>
Efficacités dans les données (%)	89.7 ± 0.1	90.2 ± 0.1	71.3 ± 0.1
Efficacités dans le Monte-Carlo (%)	93.7 ± 0.1	93.3 ± 0.1	87.6 ± 0.1
différences absolues données-Monte-Carlo (%)	-4.0 ± 0.2	-3.1 ± 0.2	-16.3 ± 0.2

TAB. 3.1: Efficacités de reconstruction des traces estimées dans les données et dans le Monte-Carlo pour les différents algorithmes après la sélection *PRT*. La dernière ligne correspond à la différence d'efficacité entre données et Monte-Carlo. Les erreurs statistiques sont mentionnées.

Les efficacités de reconstruction des traces en fonction de η , ϕ , $d0$ et du p_T des muons locaux sont présentées sur la figure **3.17** pour les données (traits pleins) et le Monte-Carlo (pointillés). Sur le graphique en haut à droite, on peut remarquer que les efficacités de l'algorithme *CosmicTF* diminuent très fortement pour $\phi \approx -2.7$ et $\phi \approx -0.5$. Cette chute provient de la sélection des graines présentée dans la section 3.1.2.

On observe également que les efficacités dans le Monte-Carlo sont systématiquement supérieures aux efficacités dans les données. Ces différences sont accentuées pour les grandes valeurs de $|\eta|$ et pour les faibles impulsions transverses. Ce dernier aspect suggère que le champ magnétique ou la perte d'énergie ne sont pas correctement décrits dans la simulation Monte-Carlo.

Les graphiques sur la figure **3.17** montrent que les efficacités sont maximales pour les muons verticaux ($\eta \approx 0$ et $\phi \approx -\pi/2$) dans la partie la plus centrale du détecteur ($d0 \approx 0$). La figure **3.18** illustre la dépendance en $d0$ des efficacités de reconstruction des traces. Elle représente les efficacités de reconstruction des traces *CTF* en fonction de ϕ et pour différentes coupures sur $d0$. Les traces situées au centre du détecteur ont une efficacité de reconstruction proche de 100% pour $|d0| < 10$ cm.

Les traces verticales proches du centre du détecteur ont de meilleures efficacités de reconstruction car dans ce cas :

1. le nombre de couches compatibles est maximal,
2. la probabilité qu'un muon passe entre deux couches du trajectographe est plus faible,
3. les muons traversent perpendiculairement les modules du trajectographe : la précision sur la position des coups doit être meilleure,
4. dans les régions les plus centrales, le biais introduit par les erreurs sur les positions X_{TOB4} et Y_{TOB4} discutées précédemment est minimal (l'acceptance du trajectographe y est maximal).

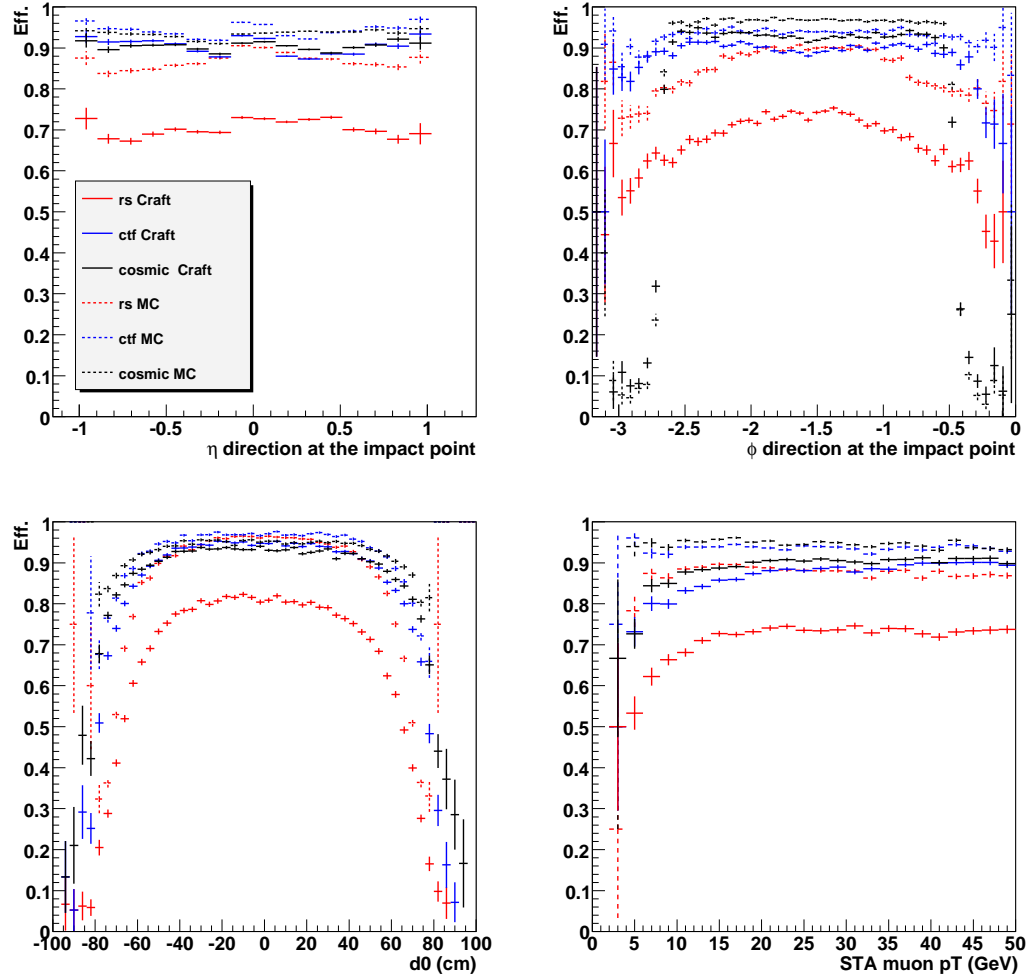


FIG. 3.17: Efficacité de reconstruction des traces après la sélection PRT dans les données Craft (en traits plein) et dans le Monte-Carlo (en pointillés), en fonction du η (en haut à gauche), du ϕ (en haut à droite), de $d0$ (en bas à gauche) et du p_T des muons locaux associés.

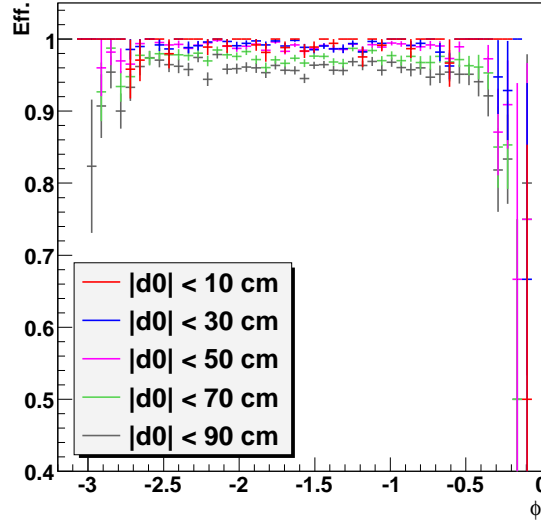


FIG. 3.18: Efficacité de reconstruction des traces (*CTF*) dans les données *Craft* en fonction du ϕ des muons locaux associés et pour différentes coupures sur $|d0|$.

Les points 1, 2 et 3 correspondent aux caractéristiques des traces produites dans les collisions. Afin d'estimer les efficacités de reconstruction des traces qui peuvent être attendues dans les collisions, un ensemble de coupures de sélection est appliqué sur les muons locaux : en plus de la sélection *PRT* et sur X_{TOB4} , les muons locaux sélectionnés doivent vérifier $|dz| < 30$ cm, $|d0| < 30$ cm, $|\eta| < 1$ et $0.5 < |\phi| < 2.5$. Cette sélection est appelé sélection *LHC*.

Les efficacités de reconstruction des traces après la sélection *LHC* sont présentées sur la figure 3.19 en fonction de l'impulsion transverse des muons locaux. Dans les données, les efficacités de reconstruction des algorithmes *CTF* et *CosmicTF* sont comprises entre 95 et 100% et sont compatibles avec les efficacités dans le Monte-Carlo. Les efficacités de l'algorithme *RS* dans les données sont comprises entre 80 et 90%. Elles restent incompatibles avec les efficacités dans le Monte-Carlo pour cet algorithme.

Les efficacités estimées sur tout le spectre en p_T sont résumées dans le tableau 3.2 pour les trois algorithmes de reconstruction des traces, pour les données et le Monte-Carlo. Les efficacités des algorithmes *CTF* et *CosmicTF* sont proches de 100% dans les données et il y a un bon accord entre données et Monte-Carlo.

Les efficacités de reconstruction des traces après la sélection *LHC* sont supérieures à 99% pour l'algorithme *CTF* et supérieures à 98% pour l'algorithme *CosmicTF*. De telles efficacités (pour des traces isolées) pourraient donc être observées dans les événements produits par les collisions. L'algorithme *RS* a cependant une efficacité significativement plus faible ($\approx 85\%$) dans les données alors que l'efficacité calculée à l'aide du Monte-Carlo

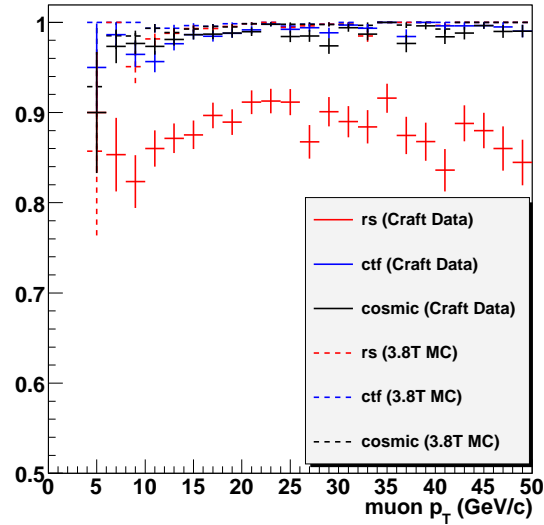


FIG. 3.19: Efficacité de reconstruction des traces *CTF*, *CosmicTF* et *RS* en fonction du p_T des muons locaux associés, dans les données (trait plein) et dans le Monte-Carlo (pointillés) dans le cas de la sélection *LHC*.

est comparable à celles des autres algorithmes ($>99\%$). Ces différences entre données et Monte-Carlo doivent être étudiées dans de futures analyses.

3.6 Conclusion et perspectives

Dans l'analyse qui a été présentée, la compatibilité entre les traces et les muons locaux et les efficacités de reconstruction des traces ont été étudiées dans les vraies données. Une comparaison a été faite avec un échantillon d'événements Monte-Carlo. Cette étude a permis de mettre en évidence un certain nombre de problèmes : inefficacité de l'algorithme *RS*, mauvaise estimation du champ magnétique au niveau des chambres à muons, accord imparfait entre données et Monte-Carlo. Ces problèmes sont encore en cours d'étude et à ce titre les résultats présentés dans ce chapitre sont préliminaires. En particulier, les paramètres de la simulation et la description du champ magnétique doivent être améliorés, les critères d'association des muons locaux aux traces doivent être précisés. Dans la simulation, les efficacités mesurées doivent être comparées aux vraies efficacités calculées à partir de l'information du générateur et de la simulation Geant4, de manière à estimer les biais liés à la sélection des muons. Les erreurs systématiques doivent également être estimées.

Il a cependant été possible de développer un outil permettant d'estimer la qualité des données cosmiques en estimant les efficacités par run. Un module de validation des

Algorithmes	<i>CTF</i>	<i>CosmicTF</i>	<i>RS</i>
Efficacités dans les données (%)	99.1 ± 0.1	98.8 ± 0.1	85.7 ± 0.3
Efficacités dans le Monte-Carlo (%)	$99.8 \pm <0.1$	99.5 ± 0.1	99.5 ± 0.1
différences absolues données-Monte-Carlo (%)	-0.7 ± 0.1	-0.7 ± 0.2	-13.8 ± 0.3

TAB. 3.2: Efficacité de reconstruction des traces estimées dans les données et dans le Monte-Carlo pour les différents algorithmes après la sélection *LHC* décrite dans le texte. La dernière ligne correspond à la différence d'efficacité entre données et Monte-Carlo. Les erreurs statistiques sont mentionnées.

données a été développé et a été introduit dans le code de validation des données *DQM*² de CMS.

²Data Quality Monitoring

4

Identification des jets de quark beau.

Dans de cette thèse, plusieurs algorithmes d'étiquetage des jets b issus de la désintégration de hadrons beaux ont été étudiés. Nous nous sommes plus particulièrement concentrés sur l'algorithme de probabilité par jet. Une partie des travaux mis en oeuvre a consisté à mettre en place des outils de calibration de cet algorithme. Une fois la prise de données commencée, de tels outils seront nécessaires pour calibrer l'algorithme mais également pour éventuellement redéfinir les catégories de traces sélectionnées.

Par souci de simplicité, les jets de particules issus de la désintégration de hadrons beaux (charmés) seront appelés jets b (c). De la même manière, les jets de particules issus de la désintégration de hadrons légers¹ (gluons) seront nommés jets de quarks légers (gluons). Les jets *légers* se référeront aux saveurs u , d , s et aux gluons g .

Pour les analyses de physique, il est nécessaire de connaître l'efficacité d'identification des jets b mais également le taux de mauvaise identification. Dans le cas où les données seraient mal reproduites par la simulation Monte-Carlo, ces efficacités doivent être évaluées à partir des données. La faisabilité des méthodes destinées à mesurer les efficacités d'identification des jets b a été étudiée à l'aide d'échantillons enrichis en jets contenant un muon. Les biais éventuellement introduits devront être étudiés avec les données. Nous discuterons également la mesure du taux de mauvaise identification à l'aide de la méthode dite *d'étiquetage négatif*, qui utilise en particulier la symétrie de la distribution en paramètre d'impact des traces pour les jets de quark léger et de gluon.

¹c'est à dire de hadrons ne contenant que des quarks u , d ou s

Les lots d'événements utilisés pour les études présentées dans ce chapitre ont été produits avec différentes versions du code de *CMS* ainsi que différents scénarios de calibration et d'alignement du détecteur. Cela conduit à une différence significative entre les performances d'un même algorithme pour différentes versions de codes. Les jets sont reconstruits à l'aide de la méthode du cône itératif $\Delta R = 0.5$ (cf. section 2.5.4) et les traces par la méthode *CTF* présentée au chapitre 2.3.2.1.

Les distributions des différents discriminants ainsi que les courbes de performance correspondantes, pour chaque algorithme d'étiquetage, seront calculées à l'aide d'échantillons produits par la version *CMSSW_2_1_12* du code de *CMS*. La première partie de l'analyse du taux de mauvaise identification utilise des échantillons plus anciens produits avec la version *CMSSW_1_3_1*. Une mise à jour avec des données plus récentes (produites avec la version *CMSSW_2_1_12*) sera présentée pour la méthode d'estimation du taux de mauvaise identification dans l'annexe A.1.

Les échantillons produits avec une simulation réaliste (prenant en compte les effets de désalignement et de décalibration du détecteur), suivant les scénarios "10 pb⁻¹" ou "100 pb⁻¹"² [81] permettent d'estimer les performances des algorithmes d'étiquetage de jets *b* dans des conditions proches du démarrage du LHC. Les effets de désalignement sur l'algorithme de probabilité par jet seront plus particulièrement étudiés. La version *CMSSW_1_6_7* a été utilisée pour produire ces échantillons.

Dans une première partie, les algorithmes d'identification utilisés dans l'expérience *CMS* seront présentés en mettant l'accent sur ceux utilisant la signification du paramètre d'impact des traces et plus particulièrement sur l'algorithme de probabilité par jet (cf. section 4.1.1.2). Les autres algorithmes seront présentés brièvement (cf. section 4.1.2). Les méthodes de mesures des performances de ces algorithmes seront présentées : la méthode "*System8*" (section 4.3) permet de mesurer l'efficacité d'étiquetage des jets de quark *b* et la méthode "*d'étiquetage négatif*" (section 4.4) permet de mesurer le taux de mauvaise identification des jets issus de l'hadronisation de quarks légers.

4.1 Algorithmes d'identification des jets de quark beau

La plupart des algorithmes d'identification des jets *b* utilisent le temps de vie significativement élevé des hadrons beaux qui peuvent parcourir, aux énergies du LHC, quelques millimètres avant de se désintégrer. Cela se traduit par la présence d'un vertex secondaire déplacé par rapport au vertex primaire d'interaction et par une distance de vol significativement plus grande que la précision du trajectographe. La masse des hadrons beaux est également plus élevée que la masse des autres hadrons et ils emportent en

²le détecteur est supposé être calibré et aligné avec un nombre d'événements correspondant à de telles luminosités.

moyenne 70 % de l'énergie du quark b initial. Finalement, dans 40 % des cas environ, les jets b contiennent un lepton (électron ou muon) issu de la désintégration faible d'un hadron beau ou charmé.

A partir de ces caractéristiques, il est possible de définir un ensemble de variables topologiques et cinématiques permettant de discriminer les jets b et les jets de saveurs légères : l'impulsion transverse des leptons contenus dans les jets b relativement à l'axe du jet, la présence de vertex secondaires reconstruits, la multiplicité de traces, la masse, la distance de vol ou encore le paramètre d'impact des traces de particules chargées par rapport au vertex primaire.

4.1.1 Algorithmes basés sur le paramètre d'impact des traces

Le paramètre d'impact d'une trace est défini comme la distance minimale entre la trajectoire reconstruite d'une trace de particule chargée et le vertex primaire. Le paramètre d'impact en trois dimensions ne peut être calculé analytiquement. On peut cependant l'estimer en linéarisant la trajectoire³ de la particule à partir du point d'approche le plus proche du vertex primaire. La figure 4.1 schématise la reconstruction du paramètre d'impact des traces. Le segment QC est la distance entre le point de la trajectoire de la trace le plus proche de l'axe du jet C et l'axe du jet lui-même. Cette distance est appelée "distance à l'axe du jet". La trajectoire est linéarisée au point C par la tangente à la trajectoire. Le paramètre d'impact IP est alors défini comme la distance minimum entre la trajectoire linéarisée et le vertex primaire V . Le segment VQ est appelé "longueur de désintégration". Le paramètre d'impact peut être calculé en 3 dimensions, on parlera alors de paramètre d'impact 3D. Si il est calculé dans le plan transverse, on parlera de paramètre d'impact transverse ou 2D.

Il est à noter que le calcul du paramètre d'impact nécessite de reconstruire le vertex primaire et l'axe des jets mais aussi d'associer les traces aux jets. Ainsi, les traces reconstruites sont associées à un jet si elles sont contenues dans un cône d'angle $\Delta R < 0.5$ autour de l'axe de ce jet. Pour prendre en compte les erreurs dues aux résolutions sur la position du vertex primaire, sur la trajectoire des traces et sur l'axe des jets, on utilise généralement la signification du paramètre d'impact, c'est-à-dire le paramètre d'impact divisé par son erreur $S_{IP} = IP/\sigma$.

Les paramètres d'impact sont signés comme le produit scalaire entre la direction du jet et la direction du paramètre d'impact. Les traces issues de la désintégration de particules à grande durée de vie ont préférentiellement de grands paramètres d'impact signés positivement.

³la trajectoire d'une particule chargée dans un champ magnétique est décrite par une hélicoïde

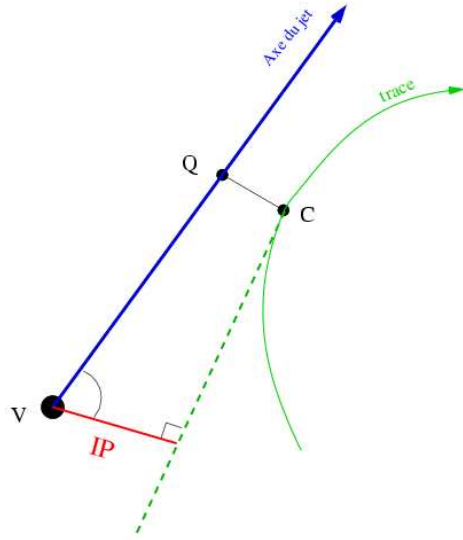


FIG. 4.1: Représentation du paramètre d'impact signé.

Pour des jets légers, on s'attend donc à observer une distribution en signification du paramètre d'impact symétrique par rapport à 0 (effets de résolution) et pour les jets b , une distribution asymétrique avec une plus grande fraction de traces ayant de grands paramètres d'impact positif. Les hadrons charmés pouvant posséder un grand temps de vie. Une asymétrie est également attendue pour les traces issues de jet c , mais dans une proportion moindre.

La figure 4.2 représente les distributions en paramètre d'impact 3D pour différentes saveurs de jets. La saveur des jets est déterminée à l'aide de l'information du générateur suivant la méthode dite "algorithmique" [52]. Les jets reconstruits sont associés aux partons finaux après la phase de radiation de parton (cf. section 1.6). Seuls les partons générés dans un cône $\Delta R = \sqrt{(\eta_{jet} - \eta_{parton})^2 + (\phi_{jet} - \phi_{parton})^2} < 0.3$ autour de l'axe du jet sont considérés. Ainsi, si un gluon produit une paire de quark-antiquark, ces derniers sont utilisés lors de l'association et non le gluon initial. Le jet est considéré comme un jet de b si un quark b généré est trouvé dans le cône, sinon l'algorithme recherche s'il existe un quark c dans ce même cône. Si aucun quark c ou b n'est trouvé, la saveur du jet correspond à la saveur du parton ayant l'énergie la plus élevée.

La fraction de traces avec de grands paramètres d'impact positif, issus de jets b est bien plus importante que dans le cas des traces issues de jets légers. On observe cependant une faible asymétrie dans la distribution des jets légers. Ceci est dû essentiellement à la présence de particules légères à grand temps de vie (les particules K_s^0 et Λ) et de conversions de photon. Ces particules sont appelées particules V_0 . Par la suite, nous

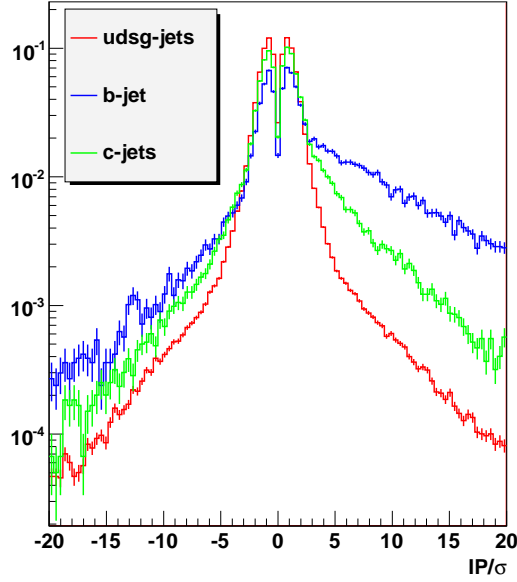


FIG. 4.2: Distribution en significane du paramètre d'impact 3D pour différents saveurs de jets pour des événements QCD $80 < \hat{p}_T < 120 \text{ GeV}/c$.

utiliserons uniquement la significane du paramètre d'impact 3D, celle-ci induisant de meilleures performances aux algorithmes d'identification des jets b .

Afin de minimiser l'influence des fausses traces et des traces mal reconstruites sur les performances des algorithmes d'étiquetage, des coupures de qualités sont également appliquées [82]. Pour chaque trace on demande :

1. un nombre de coups reconstruits dans le détecteur à pixels du trajectographe $e \geq 2$,
2. un nombre total de coups reconstruits dans le trajectographe (pistes de silicium+pixels) ≥ 8 ,
3. une impulsion transverse $p_T > 1 \text{ GeV}/c$,
4. un χ^2 normalisé de l'ajustement de la trace < 5 ,
5. un paramètre d'impact transverse $< 0.2 \text{ cm}$,
6. une longueur de désintégration $< 5 \text{ cm}$,
7. une distance à l'axe du jet $< 0.07 \text{ cm}$.

Ces coupures ont été appliquées pour la figure 4.2 .

Les coupures 1, 2, 3 et 4 servent principalement à rejeter les fausses traces et les traces mal reconstruites. Le taux de fausses traces est ainsi réduit d'un facteur 10 environ. Les traces issues de la désintégration de particules K_s^0 , Λ et de conversion de photon ont d'assez faibles impulsions transverses. La coupure 3 permet donc de rejeter une grande partie de ces traces, comme l'illustre la figure 4.3 qui présente leur distribution en p_T

[83]. Les variables utilisées pour les coupures 5, 6 et 7 permettent de rejeter les traces ne provenant pas de la désintégration de hadrons beaux mais de tout autre processus produisant des traces avec de grande signification telle que les fausses traces, les traces mal reconstruites, les particules V_0 ou les interactions nucléaires dans le détecteur. Ces processus seront discutés dans la section 4.4.2.

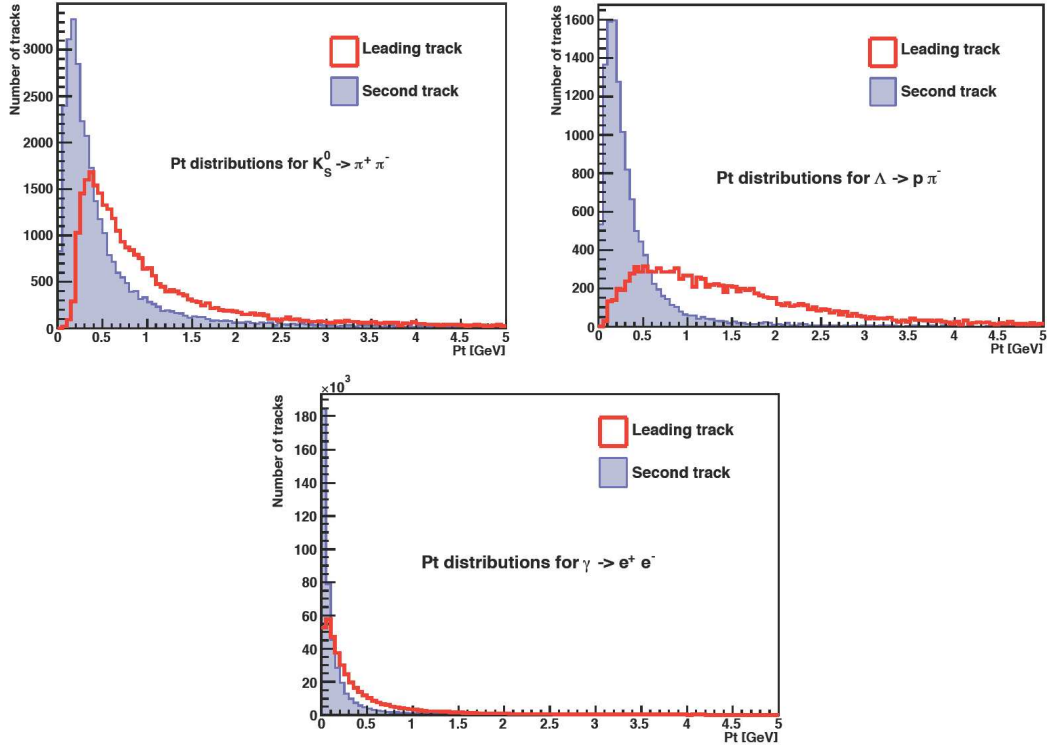


FIG. 4.3: Distributions des impulsions transverses des traces reconstruites issues de la désintégration de particules K_s^0 (en haut à gauche), Λ (en haut à droite) et de conversion de photon (en bas) pour la première trace de plus haut p_T (en rouge) et la seconde (en gris).

4.1.1.1 Le comptage de traces

L'algorithme de comptage de traces est simple et robuste. Il est tout particulièrement intéressant car il reste assez performant lorsque le détecteur est mal aligné [84]. Dans un premier temps, l'algorithme ordonne les traces en significances du paramètre d'impact de manière décroissante. Un jet va alors être identifié comme jet b s'il contient au moins N traces avec une signification du paramètre d'impact plus grande qu'une coupure seuil.

La figure 4.4 représente les distributions des 2ième et 3ième traces ayant les plus grandes significances du paramètre d'impact, ce qui correspond aux distributions des discriminants des algorithmes de comptage de trace $N = 2$ et $N = 3$ respectivement.

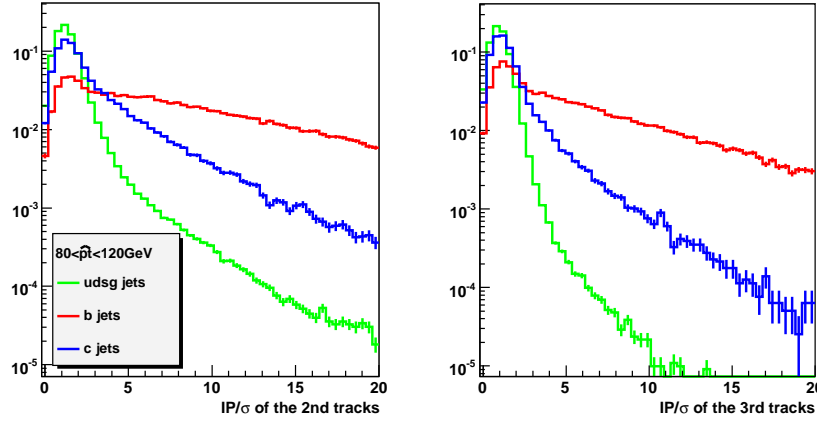


FIG. 4.4: Distribution normalisé à l'unité de la signification du paramètre d'impact 3D pour les 2ièmes (à gauche) et 3ième (à droite) traces.

L'efficacité ε_q d'étiqueter un jet de saveur q comme étant un jet b est définie par :

$$\varepsilon_q = \frac{\text{nombre de jets de saveur } q \text{ étiquetés comme jets } b}{\text{nombre total de jets de saveur } q}. \quad (4.1)$$

L'efficacité d'identification des jets b est notée ε_b et les efficacités d'identification des jets uds , c ou de gluons sont ε_{uds} , ε_c et ε_g respectivement. L'efficacité des jets légers ($uds + g$) sera notée ε_l .

Les courbes de performance (ε_l et ε_c en fonction de ε_b) sont représentées sur la figure 4.5 pour les algorithmes de comptage de traces $N = 2$ et $N = 3$.

En comparant ces algorithmes, on peut remarquer que le cas $N = 2$ permet d'obtenir de meilleures performances pour les grandes efficacités d'étiquetage de jet ε_b , il sera appelé algorithme *TCHE*⁴ dans la suite. Le cas $N = 3$ permet d'obtenir de meilleures performances pour les faibles efficacités d'identification des jets légers, il sera appelé algorithme *TCHP*⁵. Le cas $N = 1$ n'est pas utilisé car il ne permet pas une bonne rejection des jets non b .

4.1.1.1.1 Effets du désalignement Les effets de désalignement du trajectographe et de décalibration sur les algorithmes de comptage de traces sont discutés ici. L'effet de la décalibration joue principalement sur la détermination de l'axe du jet, ce qui peut avoir des conséquences sur le calcul du paramètre d'impact. Cet effet reste cependant négligeable comparé au désalignement du trajectographe.

⁴Track Counting High Efficiency

⁵Track Counting High Purity

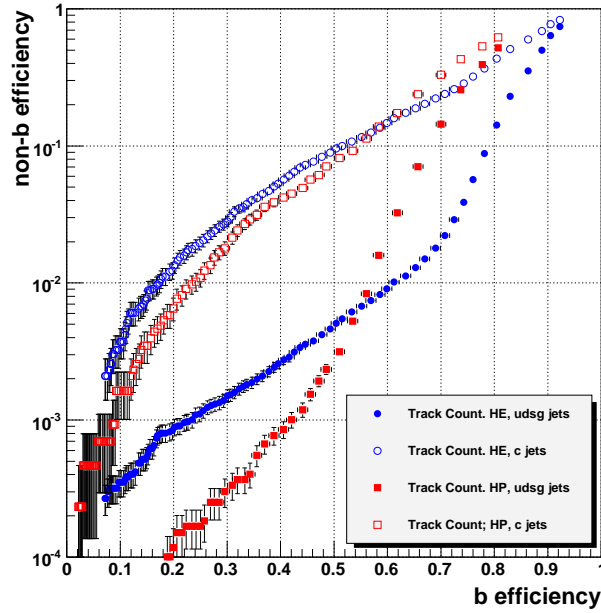


FIG. 4.5: Courbes de performance des algorithmes de comptage de traces pour des données QCD PYTHIA (*CMSSW_2_1_12*) avec $80 < \hat{p}_T < 120 \text{ GeV}$

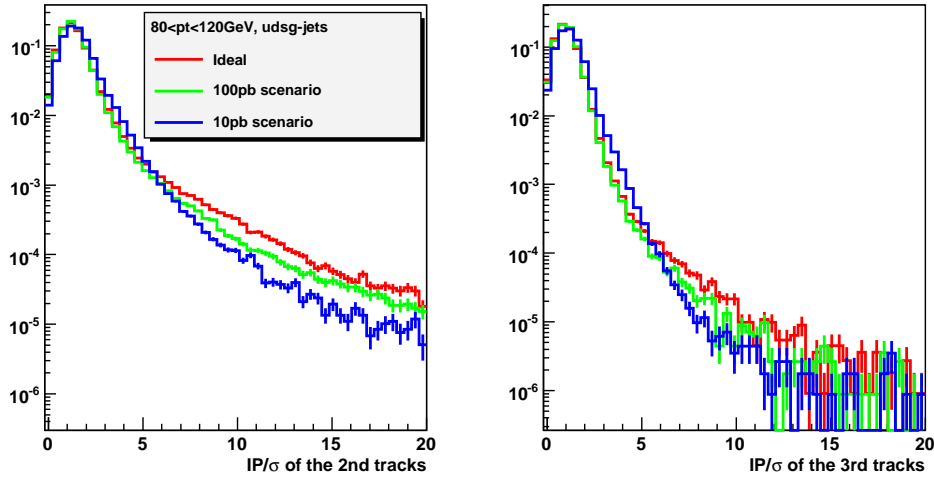


FIG. 4.6: Distributions normalisées à l'unité de la signficance du paramètre d'impact des jets légers pour les 2ième (à gauche) et 3ième (à droite) traces et pour différents scénarios de désalignement et de décalibration (*QCD* $80 < \hat{p}_T < 120 \text{ GeV}/c$).

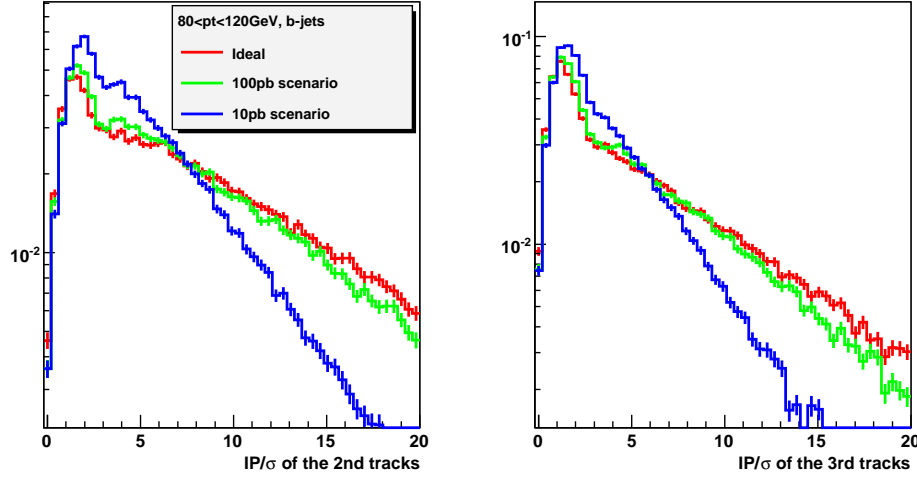


FIG. 4.7: Distributions normalisées à 1 de la signification du paramètre d'impact des jets b , pour les 2ième (à gauche) et 3ième (à droite) traces et pour différents scénarios de désalignement et de décalibration (QCD $80 < \hat{p}_T < 120$ GeV/c).

Des événements avec un désalignement et une décalibration artificiels du détecteur ont été générés de manière à estimer les performances du détecteur au démarrage du LHC. Dans le scénario 10 pb^{-1} , le trajectographe est supposé aligné à l'aide de muons cosmiques et de données provenant des premières collisions. Il s'agit principalement de hadrons isolés, d'événements "biais minimum"⁶ et de muons issus de la désintégration de résonances de basses masses J/Ψ et Υ . Dans le scénario 100 pb^{-1} , des muons de haut p_T provenant de la désintégration de bosons Z ou W seront utilisés. Les scénarios de désalignement et de décalibration sont détaillés dans la référence [81].

Les effets de désalignement sur les algorithmes d'identification des jets b ont été étudiés [84] avec des données produites avec $CMSSW_1_3_1$. Il sera présenté dans cette section une brève comparaison entre les scénarios 10 pb^{-1} , 100 pb^{-1} et le cas d'un détecteur parfait avec des données $CMSSW_1_6_12$ et pour l'algorithme de comptage de traces.

Les figures 4.6 et 4.7 représentent les distributions des deuxième et troisième traces pour les jets de saveurs $udsg$, c et b dans des événements QCD avec $80 < \hat{p}_T < 120$ GeV/c dans le cas des scénarios 10 pb^{-1} et 100 pb^{-1} et d'un détecteur parfait. Pour le scénario 10 pb^{-1} , la fraction des traces ayant de grandes significances est très largement réduite (queues de distribution plus courtes) principalement pour les jets b et c , les performances sont donc grandement diminuées dans ce cas.

⁶correspondant à des événements QCD inclusifs

La figure 4.8 représente les courbes de performance pour les algorithmes TCHE et TCHP. Les performances observées pour le scénario 100 pb^{-1} sont similaires aux performances d'un détecteur parfait. Pour le scénario 10 pb^{-1} ; les performances attendues sont très largement réduites car les erreurs sur le paramètre d'impact sont plus grandes. Pour un taux de mauvaise identification des jets légers de 1%, l'efficacité d'identification des jets b diminue de plus de 15% en valeur absolue.

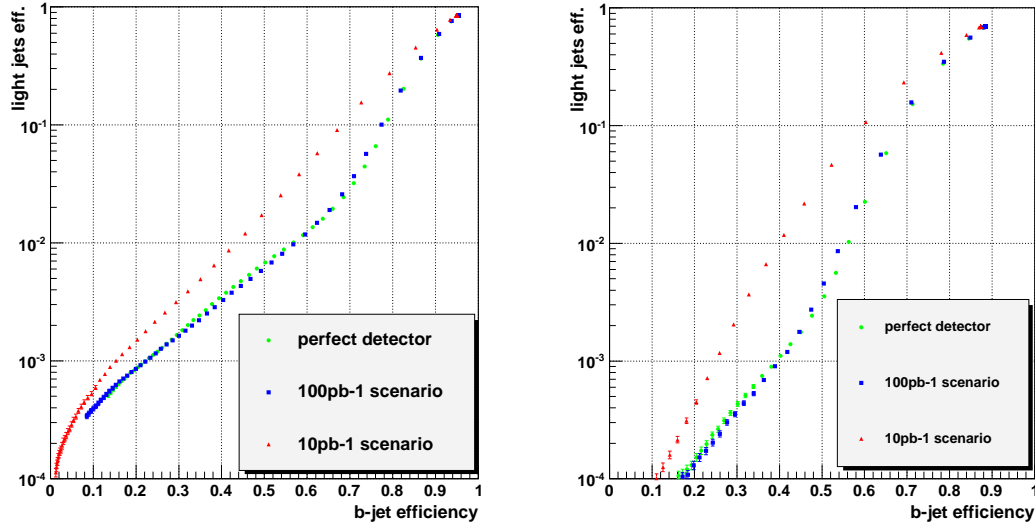


FIG. 4.8: Courbes de performance des algorithmes de comptage de traces TCHE (en haut) et TCHP (en bas) pour les jets légers pour différents scénarios de désalignement de décalibration. Les différences pour les jets c sont négligeables et non représentées ($QCD 80 < \hat{p}_T < 120 \text{ GeV}/c$).

Les algorithmes de comptage de traces n'utilisent que l'information des 2 ou 3 traces avec les plus grandes significances en paramètre d'impact contenues dans un jet. Il est donc en principe possible d'obtenir de meilleures performances en combinant les informations de l'ensemble des traces associées au jet.

4.1.1.2 La méthode probabiliste

La méthode probabiliste [85] calcule la probabilité des traces de provenir du vertex primaire. Les probabilités de toutes les traces associées à un jet sont combinées de manière à calculer la probabilité par jet de provenir du vertex primaire. Si cette probabilité est faible, le jet considéré a plus de chance d'être un jet de quark b .

La probabilité des traces est calculée à l'aide des distributions en $S_{IP} < 0$. Les paramètres d'impact négatifs proviennent majoritairement du vertex primaire et sont utilisés

pour déterminer des *fonctions de densité de probabilité*. L'avantage principal de cette méthode est d'utiliser toutes les traces contenues dans un jet.

Les significances des paramètres d'impact négatif sont donc utilisées pour déterminer des fonctions de résolution R . Elles permettent de calculer la probabilité signée $P_{tr}(S_{IP})$ d'une trace de provenir du vertex primaire, sachant qu'elle possède une signification du paramètre d'impact S_{IP} :

$$P_{tr}(S_{IP}) = \text{sign}(S_{IP}) \int_{|S_{IP}|}^{\infty} R(x) dx. \quad (4.2)$$

où $\text{sign}(S_{IP})$ est le signe du paramètre d'impact.

La partie centrale des fonctions de résolution est bien décrite par une gaussienne de moyenne nulle. Les queues non-gaussiennes sont relativement difficiles à modéliser, principalement à cause de l'erreur sur le paramètre d'impact qui peut être mal calculée. C'est pourquoi le calcul de $P(S_{IP})$ est réalisé à l'aide d'intégrations numériques des histogrammes de significances négatives. Ces histogrammes sont déterminés à l'aide de données réelles. La statistique très importante nécessaire à cette méthode devrait être obtenue assez rapidement après le début de la prise de données. Cette approche permet de ne pas utiliser de fonction d'ajustement pour déterminer l'allure des distributions en signification négative du paramètre d'impact et d'obtenir de meilleures descriptions des fonctions de résolution R .

La distribution en signification négative du paramètre d'impact peut également varier de manière importante en fonction de l'impulsion p , de la pseudo-rapacité $|\eta|$, du χ^2 normalisé de l'ajustement de la trace ou encore du nombre de coups associés dans l'ensemble du trajectographe ou dans le détecteur à pixels uniquement. Afin de prendre en compte ces effets, différentes catégories de traces sont définies de manière à augmenter le pouvoir discriminant de l'algorithme. Pour chaque catégorie de traces, un histogramme en S_{IP} est produit à partir des données pour des valeurs de $S_{IP} < 0$. Par construction la distribution en probabilité des traces de paramètre d'impact négatif est uniformément distribuée entre -1 et 0.

La probabilité d'un jet contenant N traces de provenir du vertex primaire est calculée par la combinaison des probabilités des N traces suivant une distribution de Poisson. Elle est définie par [86, 87, 88] :

$$P_{jet} = \Pi \sum_{j=0}^{N-1} \frac{(-\ln \Pi)^j}{j!},$$

avec

$$\Pi = \prod_{i=1}^N \tilde{P}_{tr}(i). \quad (4.3)$$

où \tilde{P}_{tr} est définie comme la probabilité d'une trace positive (ayant une signification $S_{IP}>0$) ou comme la probabilité redéfinie par $\tilde{P}_{tr} = P_{tr}/2$ pour $S_{IP} > 0$ et $\tilde{P}_{tr} = 1 + P_{tr}/2$ pour $S_{IP} < 0$. Cette dernière transformation permet de prendre en compte toutes les traces (positives et négatives). Les traces de $S_{IP}<0$ ont ainsi tendance à augmenter la probabilité par jet tandis que les traces de $S_{IP}>0$ ont tendance à la diminuer. Par la suite, la probabilité redéfinie sera utilisée.

Les différentes catégories de traces utilisées pour définir les histogrammes de calibration sont les suivantes :

- une catégorie pour les traces ayant un χ^2 normalisé > 2.5 (catégorie 1).
- Pour les traces ayant un χ^2 normalisé < 2.5 :
- 3 catégories pour les traces ayant une pseudo-rapacité $|\eta|$ comprise dans les intervalles $[0;0.8]$, $[0.8;1.6]$ et $[1.6; 2.5]$, au moins 3 coups dans le détecteur à pixels et une impulsion $p < 8$ GeV/c (catégories 2, 3, et 4 respectivement),
 - 1 catégorie pour les traces ayant exactement 2 coups dans le détecteur à pixels et une impulsion $p < 8$ GeV/c (catégorie 5)
 - 3 catégories pour les traces ayant un $|\eta|$ compris dans les intervalles $[0;0.8]$, $[0.8;1.6]$ et $[1.6; 2.5]$, au moins 3 coups dans le détecteur à pixels et une impulsion $p > 8$ GeV/c (catégories 6, 7 et 8 respectivement),
 - 1 catégorie pour les traces ayant exactement 2 coups dans le détecteur à pixels et une impulsion $p > 8$ GeV/c (catégorie 9).

Le découpage en $|\eta|$ correspond au cas où la trajectoire d'une particule chargée de grande impulsion traverse la partie centrale du trajectographe $[0;0.8]$, la région de transition entre la partie centrale et les bouchons $[0.8;1.6]$ ou uniquement les bouchons $[1.6;2.5]$ (cf. figure **2.5**). Ces intervalles en $|\eta|$ sont basés sur la géométrie du détecteur. Le découpage en χ^2 normalisé est quand à lui plus délicat. Il correspond à une coupure relativement arbitraire (χ^2 normalisé < 2.5) qui sélectionne environ 78% de traces reconstruites et moins de 0.1% de fausses traces (cf. figure **4.9**).

Le découpage en nombre de coups reconstruits dans le détecteur à pixels correspond au cas où les traces sont produites majoritairement dans la partie centrale du détecteur ($N_{pixel} = 3$) ou dans les bouchons ($N_{pixel} \geq 3$). Les catégories avec $N_{pixel} = 2$ permettent de compenser les inefficacités éventuelles du détecteur à pixels. Finalement, le découpage en impulsion, bien que raisonnable, reste arbitraire. La définition des différentes catégories reste non optimisée, en particulier pour les découpages en impulsion et en χ^2 normalisé. Elle se fera pour les données réelles à l'aide des outils présentés dans la section 4.1.1.2.1.

Pour tester la méthode, les calibrations ont été produites à l'aide de 1,2 million d'événements *QCD* PYTHIA ($80 < \hat{p}_T < 120$ GeV/c). La figure **4.10** de droite représente la distribution de probabilité pour les traces négatives et pour toutes les saveurs de jets.

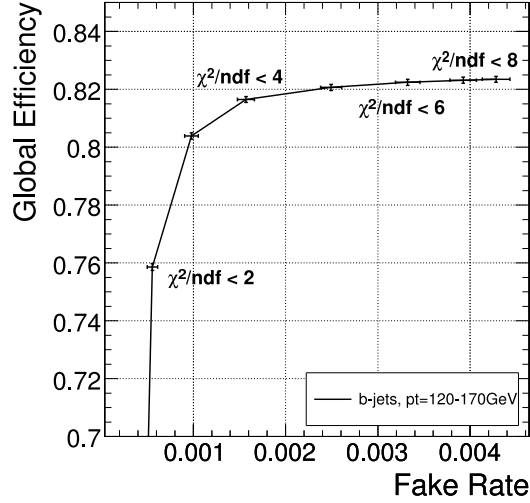


FIG. 4.9: Efficacité de sélection des traces en fonction du taux de fausses traces sélectionnées pour différentes coupures en χ^2 normalisé. Des jets d'impulsions transverses comprises entre 120 et 170 GeV/c provenant d'événements QCD sont utilisés [52].

Comme ces traces sont utilisées pour mesurer les fonctions de résolution R la probabilité des traces ayant un paramètre d'impact négatif est uniformément distribuée entre -1 et 0, par construction. Le graphique de droite de la figure 4.10 représente les distributions normalisées de probabilité des traces pour les jets $udsg$, c et b . On observe que la distribution pour les jets b est très fortement piquée à 0. C'est également le cas de la distribution pour les jets c mais dans une moindre mesure. Dans le cas des jets $udsg$, on observe bien une distribution uniformément distribuée entre -1 et 1, excepté autour de 0. La distribution des probabilités des traces de signification négative pour les jets $udsg$ diminue pour les probabilités proches de 0^- . Ce comportement est dû à l'utilisation de toutes les saveurs de jets pour le calcul des fonctions de résolution (qui contiennent donc une contribution des jets b et c). Le pic pour les probabilités proches de 0^+ est dû à la présence de particules à grande durée de vie dans les jets de saveurs légères telles que les K_s^0 , les Λ ainsi que les conversions de photon. Des algorithmes de réjection de particules V^0 et des conversion de photons existent dans CMS mais ne sont pas encore utilisés par les algorithmes d'identification des jets b .

La figure 4.11 représente la probabilité par jet de provenir du vertex primaire pour les différentes saveurs $udsg$, c et b . Pour les jets b et c , la distribution est piquée à 0. La distribution de probabilité pour les jets légers est quant à elle relativement uniforme entre 0 et 1.

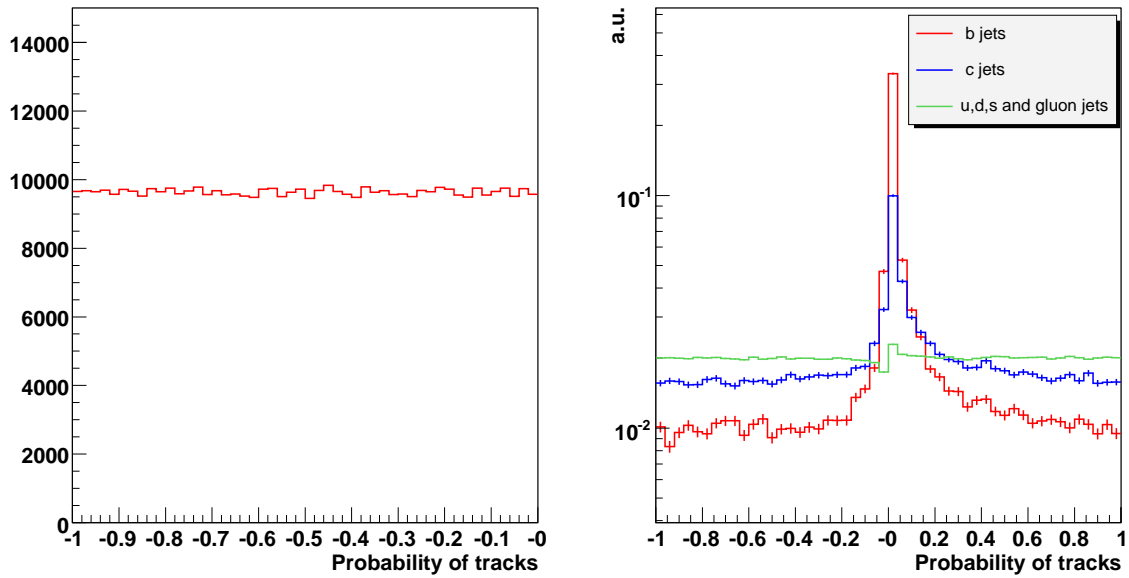


FIG. 4.10: Probabilité des traces de signficance négative (à gauche) pour toutes les saveurs de jets et probabilité des traces de significances positive et négative (à droite) pour les saveurs de jets $udsg$, c et b (données QCD PYTHIA $CMSSW_1_5_2$ avec $80 < \hat{p}_T < 120$ GeV/c).

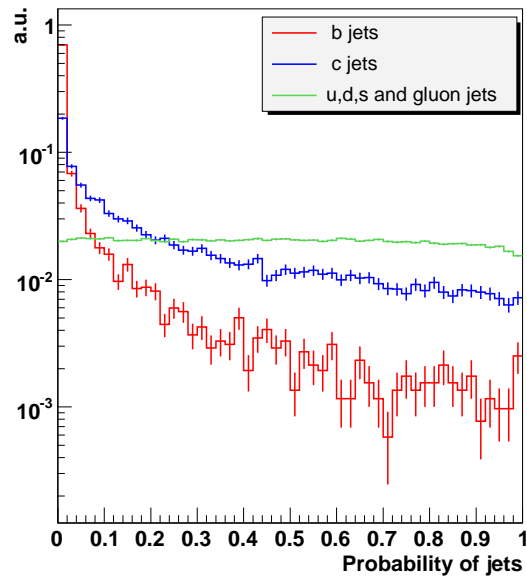


FIG. 4.11: Probabilité par jet pour les saveurs $udsg$, c et b (données QCD PYTHIA $CMSSW_1_5_2$ avec $80 < \hat{p}_T < 120$ GeV/c).

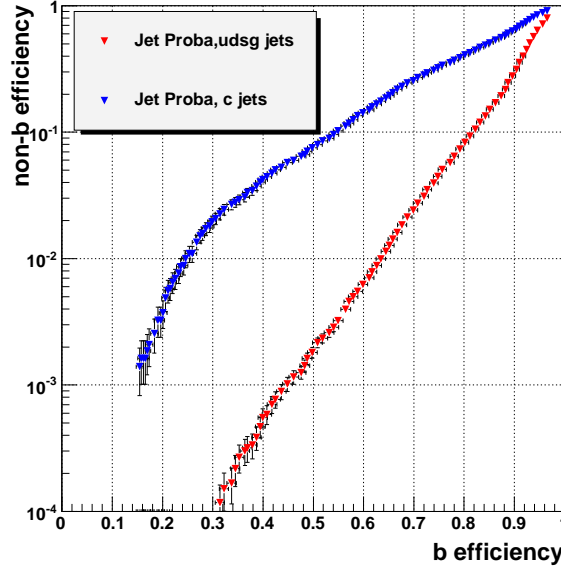


FIG. 4.12: Courbes de performance de l'algorithme de probabilité par jet pour des données QCD PYTHIA (*CMSSW_2_1_12*) avec $80 < \hat{p}_T < 120$ GeV/c.

Les courbes de performance de l'algorithme de probabilité par jet sont présentées sur la figure 4.12 pour des événements QCD ($80 < \hat{p}_T < 120$ GeV/c) produits avec la version *CMSSW_2_1_12* du code de *CMS*.

4.1.1.2.1 Outils de validation des calibrations Dans le cas d'un détecteur parfait, il s'avère que les calibrations sont proches de l'optimum. Cela est certainement dû à la surestimation de la qualité du trajectographe dans la simulation. Pour les données réelles, les effets de reconstruction (comme le désalignement, la présence de modules inactifs dans le trajectographe, une mauvaise estimation des erreurs sur le paramètre d'impact, etc...) devront être pris en compte dans la définition des catégories de traces. Il est donc nécessaire de déterminer un ensemble d'histogrammes de contrôle qui pourront dans un premier temps valider les calibrations et si nécessaire être utilisés pour les redéfinir.

Les distributions en probabilité des traces et des jets sont les histogrammes de contrôle les plus naturels (cf. figure 4.10 et 4.11). Elles permettent de vérifier que les probabilités de traces négatives sont bien uniformément réparties entre -1 et 0 et que la probabilité par jet, lorsqu'elle est calculée uniquement avec les traces de $S_{IP} < 0$, est uniformément répartie entre 0 et 1.

Des histogrammes de contrôle peuvent également être définis à l'aide des variables utilisées pour la paramétrisation des catégories. Nous avons vu que la qualité des traces dépend principalement de variables continues (p , η et χ^2) et de variables discrètes (nombre de coups dans le détecteur à pixels). Un autre paramètre important est le nombre de traces par catégorie. En effet, une catégorie doit contenir suffisamment de traces pour décrire correctement les distributions en signification et pour avoir un impact sur les performances.

Il est donc important de caractériser l'évolution de la signification du paramètre d'impact **néгатif** $S_{IP < 0}$ en fonction des différents paramètres utilisés pour la définition des catégories. Comme il est délicat d'utiliser des fonctions d'ajustement pour étudier les distributions en signification négative, on peut se restreindre dans un premier temps à la partie centrale gaussienne de cette distribution. Elle sera étudiée à l'aide de l'écart type σ_G issu de l'ajustement d'une distribution gaussienne dans l'intervalle de signification $[-2; 0]$. Les queues sont ensuite caractérisées par la fraction de traces $f_{neg}(X)$ ayant une signification du paramètre d'impact inférieure à une valeur **néгатive** X .

On peut alors représenter les variables σ_G et $f_{neg}(X)$ en fonction de p , de η ou du χ^2 normalisé des traces par des graphiques à une dimension, pour les 9 catégories déjà prédéfinies.

Les figures 4.13 et 4.14 représentent l'évolution du paramètre σ_G de l'ajustement gaussien de la partie centrale des distributions en S_{IP} négative pour différents intervalles en χ^2 normalisé et en impulsion p des traces, pour les catégories 2 et 5. Les largeurs des distributions ne semblent pas dépendre significativement de l'impulsion ni du χ^2 normalisé des traces. Les mêmes comportements sont observés pour les autres catégories.

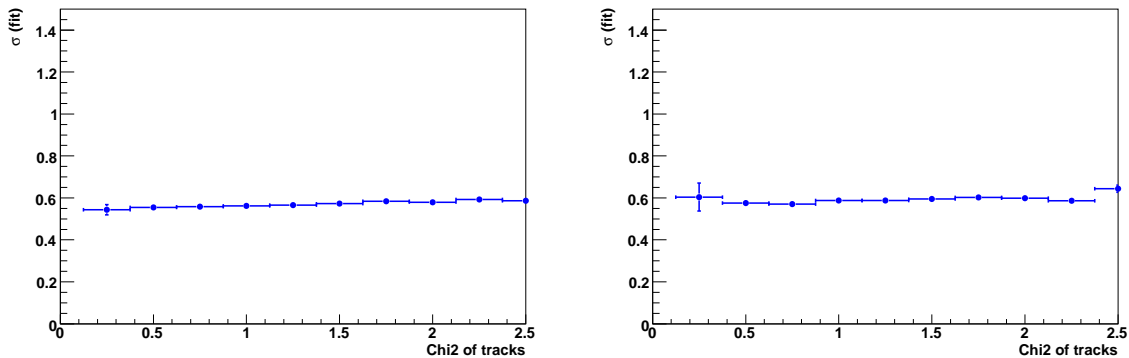


FIG. 4.13: Ecart type de l'ajustement gaussien des distributions en signification sur l'intervalle de signification $[-2; 0]$ en fonction du χ^2 normalisé des traces, pour les catégories 2 (à gauche) et 5 (à droite).

La figure 4.15 représente les fractions $f_{neg}(-2)$, $f_{neg}(-3)$ et $f_{neg}(-4)$ en fonction du χ^2 normalisé des traces pour les catégories 2 et 5. On observe une claire évolution de la queue de distribution en fonction du χ^2 . La figure 4.16 représente $f_{neg}(-2)$, $f_{neg}(-3)$

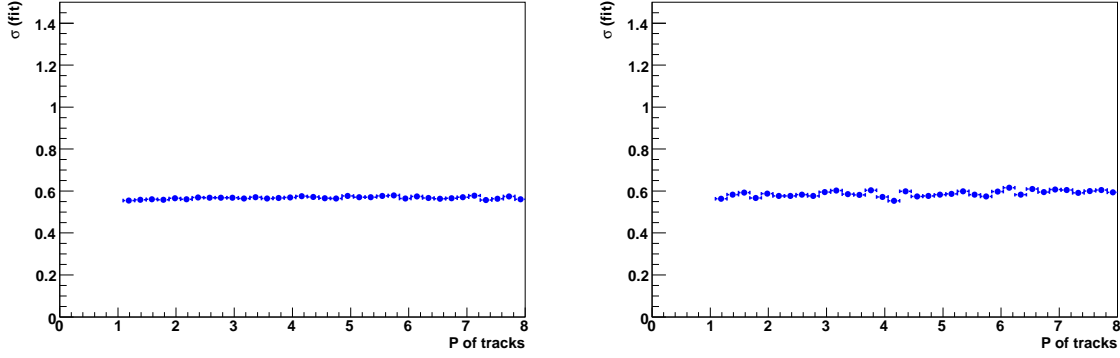


FIG. 4.14: Ecart type de l'ajustement gaussien des distributions en signifiacance sur l'intervalle de signifiacance $[-2;0]$ en fonction de l'impulsion des traces, pour les catégories 2 (à gauche) et 5 (à droite).

et $f_{neg}(-4)$ en fonction de l'impulsion des traces. Les fractions $f_{neg}(X)$ semblent moins sensibles à p .

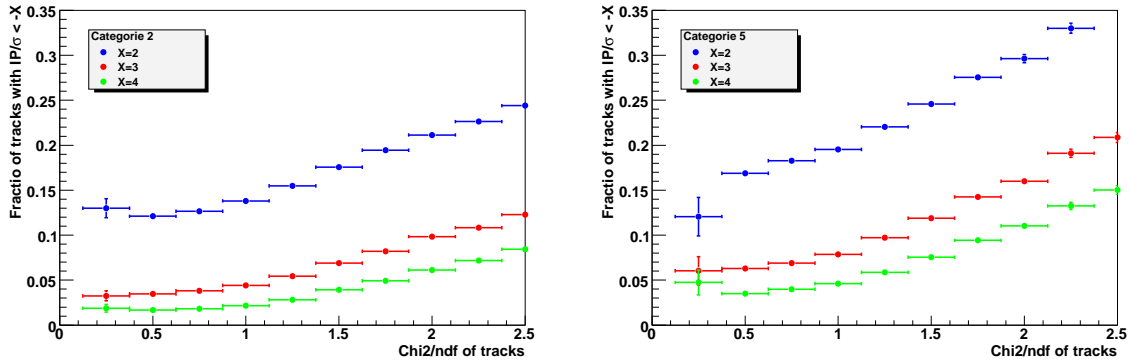


FIG. 4.15: Fraction de traces ayant une signifiacance du paramètre d'impact inférieure à X pour $X = -2, -3,$ et -4 en fonction du χ^2 normalisé des traces, pour les catégories 2 (à gauche) et 5 (à droite).

Cette méthode est simple et permet d'observer l'évolution des distributions en signifiacance, elle ne rend cependant pas compte des corrélations entre les variables. Il est donc nécessaire d'introduire des graphiques à 2 dimensions tels que $\sigma_G(\chi^2;p)$ et $f_{neg}(X)(\chi^2;p)$ pour différents intervalles de $|\eta|$. La figure 4.17 représente le nombre de traces et la fraction $f_{neg}(2)$ en fonction de l'impulsion des traces et de leur χ^2 normalisé pour 860 000 événements QCD ($80 < \hat{p}_t < 120$ GeV/c). De tels histogrammes sont utilisés pour étudier les dépendances en p et χ^2 , mais peuvent également être utilisés pour étudier les

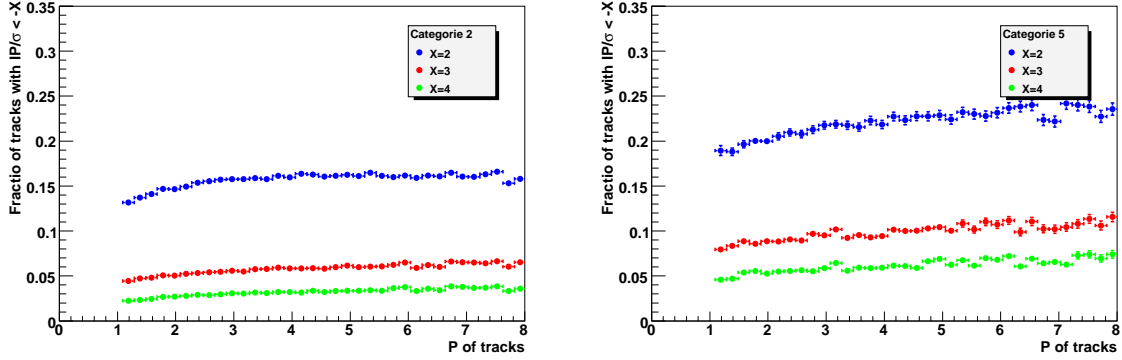


FIG. 4.16: Fraction de traces ayant une signification du paramètre d'impact inférieure à X pour $X = -2, -3,$ et -4 en fonction de l'impulsion des traces, pour les catégories 2 (à gauche) et 5 (à droite).

différentes corrélations telles que p et $|\eta|$ à χ^2 fixé ou χ^2 et $|\eta|$ à p fixé.

4.1.1.2.2 Effets du désalignement L'algorithme et ses calibrations dépendent également des effets de désalignement du détecteur. Le figure 4.18 représente les courbes de performance pour les scénarios 10 pb^{-1} , 100 pb^{-1} et dans le cas d'un détecteur idéal pour des événements produits avec la version *CMSSW_1_6_12* du code de *CMS*. Dans chaque cas, l'algorithme a été recalibré avec des données correspondant au scénario considéré et en utilisant le même nombre d'événements *QCD*. Les performances de l'algorithme pour le scénario 100 pb^{-1} et pour un détecteur parfait sont comparables. Les performances pour 10 pb^{-1} sont fortement diminuées. Pour un taux de mauvaise identification des jets *udsg* d'environ 1%, l'efficacité d'identification des jets *b* passe de 63% pour un détecteur parfait à 45% pour le scénario 10 pb^{-1} .

Comme la recalibration de l'algorithme pour le scénario 10 pb^{-1} ne change pas ses performances de manière significative, les baisses des performances observées ne sont pas dues à des problèmes de définition de catégories mais plutôt aux effets de désalignement.

Pour conclure, un ensemble d'histogrammes de validation et de vérification des calibrations a été défini. Seule une partie des histogrammes de validation a été présentée, par souci de clarté. Il a également été fait une comparaison entre les différents scénarios d'alignement. Il n'y a cependant pas un grand intérêt à optimiser les calibrations pour ces échantillons Monte-Carlo. Des tests ont cependant été faits et conduisent, pour ε_{udsg} d'environ 1%, à une amélioration des efficacités d'identification ε_b d'environ 3% en redéfinissant le découpage en χ^2 .

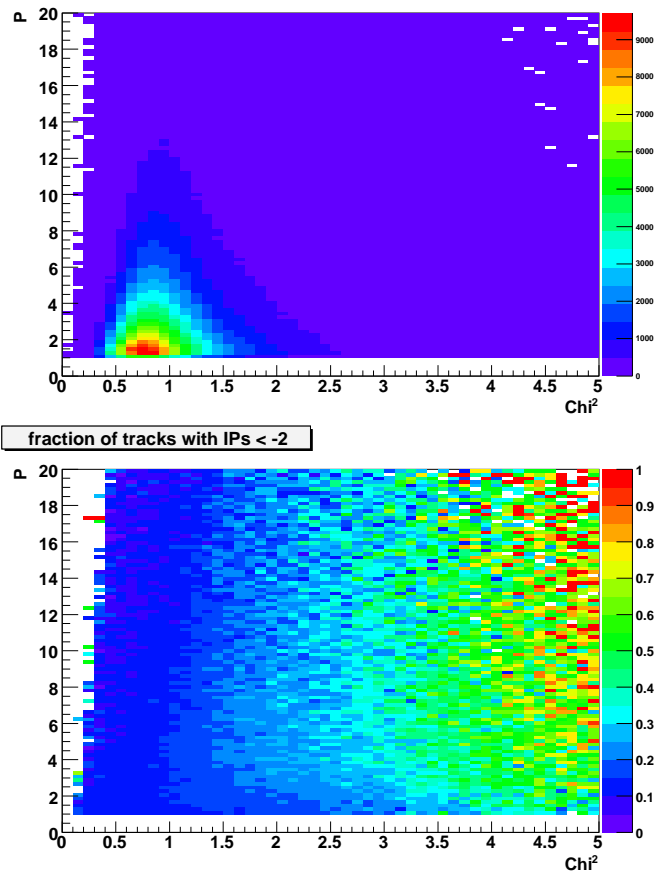


FIG. 4.17: Pour les traces avec $0 < |\eta| < 0.8$ et au moins 3 coups dans le détecteur à pixels, le nombre de traces (en haut) et la fraction $f_{neg}(-2)$ (en bas) en fonction de l'impulsion des traces et de leur χ^2 normalisé pour des événements QCD ($80 < \hat{p}_t < 120$ GeV/c).

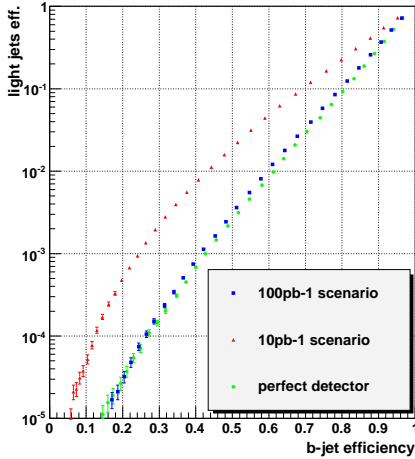


FIG. 4.18: Courbes de performance (ε_{udsq} en fonction de ε_b) pour l’algorithme de probabilité par jet pour différents scénarios de désalignement du trajectographe : détecteur parfait, 100pb^{-1} et 10pb^{-1} pour des événements QCD avec $80 < \hat{p}_t < 120\text{ GeV}/c$.

4.1.2 Description des autres algorithmes d’identification

Les autres algorithmes d’identification des jets b seront brièvement présentés dans cette section : les algorithmes utilisant la reconstruction des vertex secondaires et les algorithmes utilisant les leptons dans les jets.

Les algorithmes de Vertex Secondaires et de Vertex Secondaires Combinés [89] utilisent les traces associées à un jet pour reconstruire des candidats vertex secondaires. Tout d’abord, **l’algorithme du Vertex Secondaire** (cf. section 2.3.2.2) utilise la distance entre le vertex primaire et le vertex secondaire comme discriminant. Cet algorithme est simple et robuste lorsque le détecteur est désaligné [84]. L’efficacité de reconstruction des vertex secondaires n’étant que d’environ 65% dans les événements QCD ($80 < \hat{p}_T < 120\text{ GeV}/c$), cet algorithme ne peut être utilisé pour les grandes valeurs de ε_b . Les performances de l’algorithme du Vertex Secondaire sont présentées sur la figure 4.19.

L’algorithme du Vertex Secondaire Combiné utilise quant à lui les vertex secondaires reconstruits, s’il y en a, en association avec d’autres variables discriminantes à l’aide d’une technique de rapport de vraisemblance. Si aucun vertex secondaire n’est reconstruit, la variable discriminante est définie par d’autres quantités. Dans cette méthode, 3 différentes catégories de vertex secondaires sont définies :

1. *Reco-Vertex* : Lorsqu’au moins un vertex secondaire est reconstruit dans le jet.
2. *Pseudo-Vertex* : Lorsqu’il n’y a pas de vertex secondaire reconstruit, un pseudo-vertex est reconstruit à partir des traces non compatibles avec le vertex primaire

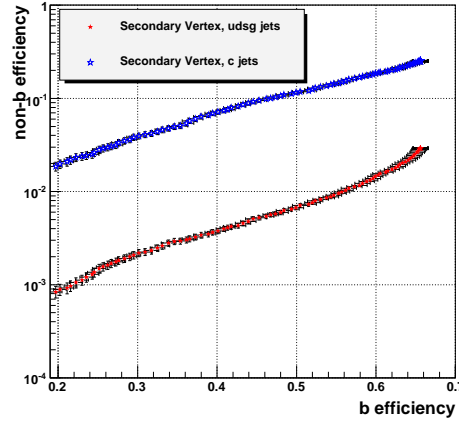


FIG. 4.19: Courbes de performance de l'algorithme du Vertex Secondaire pour des données QCD PYTHIA (*CMSSW_2_1_X*) avec $80 < \hat{p}_T < 120$ GeV/ c .

et qui ont une signification du paramètre d'impact supérieure à 2. Au moins 2 de ces traces sont nécessaires.

3. *No-Vertex* : Lorsque les conditions 1 et 2 ne sont pas remplies.

Ces trois catégories utilisent les significances du paramètre d'impact des traces contenues dans un jet. Différentes coupures sur les traces sont appliquées, en particulier les paires de traces ayant des masses invariantes compatibles avec la masse du K_S^0 sont rejetées.

Pour la catégorie *Reco-Vertex*, des variables telles que la multiplicité des traces associées au vertex secondaire, leur masse invariante, leur énergie, leur pseudo-rapacité par rapport à l'axe du jet, la distance entre le vertex primaire et le vertex secondaire ou encore la plus grande signification du paramètre d'impact dans le jet sont combinées pour calculer le discriminant. Si plusieurs vertex secondaires sont reconstruits, celui avec la plus grande distance de vol est choisi.

Pour la catégorie *Pseudo-Vertex*, les mêmes quantités sont utilisées excepté la distance entre le vertex primaire et le secondaire. La catégorie *No-Vertex* n'utilise que la signification du paramètre d'impact des traces. On peut donc s'attendre à ce que la performance de l'algorithme du Vertex Secondaire Combinée se rapproche de l'algorithme de probabilité par jet pour $\varepsilon_b > 65\%$ à cause de l'efficacité de reconstruction des vertex secondaires dans les jets.

Le dernier type d'algorithme d'identification de jets b utilise les leptons (électrons et muons) issus de la désintégration faible des hadrons beaux ($b \rightarrow W^* X$ avec $W^* \rightarrow l\bar{l}$ ou $b \rightarrow W^* c$ avec $c \rightarrow W^* X$ et $W^* \rightarrow l\bar{l}$). Un jet b va donc contenir un lepton dans

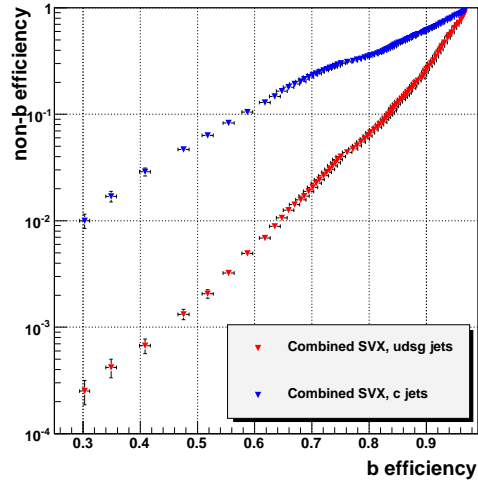


FIG. 4.20: Courbes de performance de l’algorithme du Vertex Secondaire combiné pour des données QCD PYTHIA (*CMSSW_2_1_12*) avec $80 < p_T < 120$ GeV/c.

environ 40% des cas. Par la suite, pour des raisons de simplicité, il ne sera discuté que des algorithmes utilisant un muon dans les jets [90].

Premièrement, la reconstruction standard de muons globaux (cf. section 2.4.2.2) est utilisée. La principale limitation de la méthode est due à la présence du champ magnétique qui provoque des inefficacités de reconstruction des leptons à bas p_T . La présence de tels leptons reconstruits à l’intérieur des jets suffit à rejeter une grande partie des jets légers. Cependant, ces derniers peuvent être identifiés comme jets de b s’ils contiennent de vrais muons provenant de la désintégration de π^\pm ou de K^\pm ou si des hadrons interagissent tardivement dans le calorimètre et qu’ainsi une partie des gerbes hadroniques atteint les chambres à muons.

Différentes variables peuvent être utilisées pour calculer le discriminant correspondant : l’impulsion transverse relative p_T^{Rel} du muon par rapport à la direction de l’axe du système jet+muon est calculé par la formule :

$$p_T^{rel} = \frac{\vec{p}^\mu \times \vec{p}^{\mu+jet}}{|\vec{p}^{\mu+jet}|}, \quad (4.4)$$

la signification du paramètre d’impact de la trace du muon ou encore la combinaison de plusieurs variables à l’aide d’un réseau de neurones. La liste des variables utilisées par ce réseau de neurones est :

- l’impulsion transverse relative p_T^{Rel} ,
- la signification du paramètre d’impact du muon,
- la distance R dans l’espace (η, ϕ) entre le muon et l’axe du jet,

- le rapport entre l'impulsion de la trace du muon mesurée dans le trajectographe et l'énergie calorimétrique du jet.

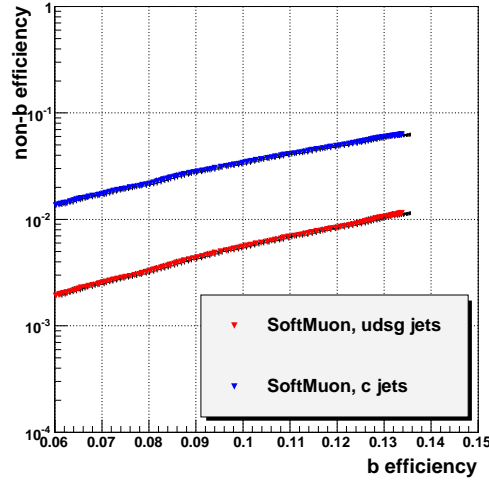


FIG. 4.21: Courbes de performance de l'algorithme *SoftMuon* pour des données QCD PYTHIA (*CMSSW_2_1_12*) avec $80 < \hat{p}_T < 120$ GeV/c.

Les courbes de performance de l'algorithme *SoftMuon* (lorsque le discriminant est le p_T^{rel} du muon) sont présentées sur la figure 4.21. On remarque que l'efficacité maximale qui peut être atteinte avec cet algorithme est limitée. Des critères d'identification plus simple peuvent être utilisés : une coupure sur le p_T^{rel} du muon ou sur la signification du paramètre d'impact.

Pour conclure sur les performances des différents algorithmes, l'algorithme du Vertex Secondaire Combiné donne les meilleurs résultats. Il utilise des données issues de la simulation Monte-Carlo comme calibration et de ce fait manque de robustesse au démarrage du LHC. L'algorithme de probabilité par jet utilise également des calibrations, mais qui sont déterminées à partir des données réelles, même si la définition des catégories reste sensible à la qualité des traces et à la résolution du paramètre d'impact. Les algorithmes de comptage de traces restent robustes et ont des performances suffisamment bonnes pour être utilisables avec les premières données. L'algorithme utilisant les muons dans les jets est le plus robuste dans sa configuration la plus simple : lorsque le discriminant est l'impulsion transverse relative p_T^{Rel} ou la signification du paramètre d'impact du muon. Il ne donne cependant accès qu'à un ensemble limité de points de fonctionnement correspondant à des faibles efficacités ε_b . Les différentes courbes de performance pour les différents algorithmes d'identification de jets *b* sont comparées sur de la figure 4.22.

Pour pouvoir utiliser l'identification des jets *b* dans les analyses de physique, il est nécessaire de déterminer les efficacités d'identification des jets *b* ainsi que le taux de

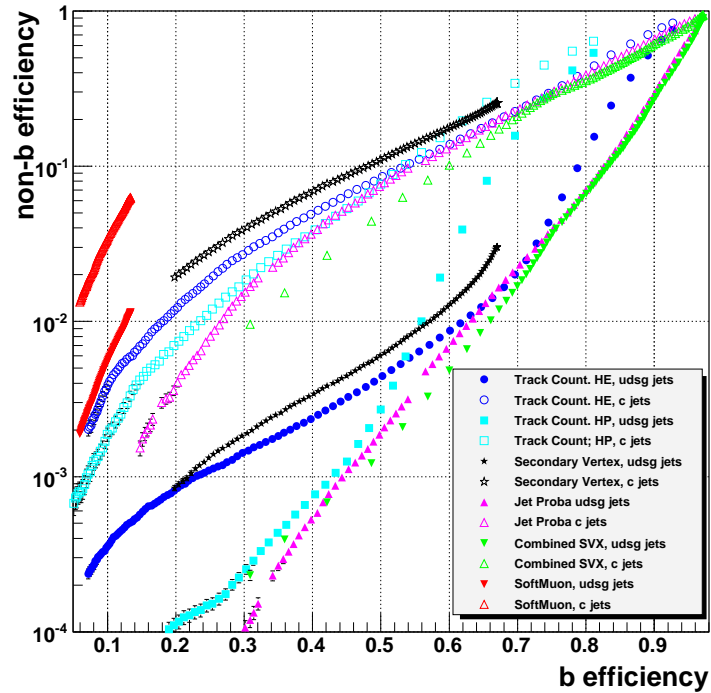


FIG. 4.22: Courbes de performance des tous les algorithmes d'identification des jets b pour des données QCD PYTHIA ($CMSSW_2_1_12$) avec $80 < \hat{p}_T < 120$ GeV/c. Pour un même algorithme (pour une même couleur) les courbes supérieures correspondent à la mauvaise identification des jets c et les courbes inférieures à la mauvaise identification des jets légers $udsg$.

mauvaise identification. Il sera délicat de déterminer ces efficacités à partir de la simulation Monte-Carlo. Des méthodes de mesure d'efficacité des algorithmes d'identification à partir des données ont été développées au cours de cette thèse, elles seront présentées dans les sections qui vont suivre.

4.2 Événements utilisés et définition de la taggabilité

Il va être décrit ici brièvement les données utilisées pour estimer les performances des algorithmes d'identification des jets b . Nous définirons ensuite la notion de "taggabilité".

Différents types d'échantillons Monte-Carlo ont été utilisés (produits avec la version `CMSSW_1_3_1`). Le taux de mauvaise identification est évalué à partir d'environ 3,4 millions d'événements QCD PYTHIA pour des valeurs de \hat{p}_T comprises dans les différents intervalles [20;30], [30;50], [50;80], [80;120], [120;170], [170;230], [230;300], [300;380], [380;470], [470;600] GeV/c. Environ 400 000 événements W +jets (avec $W \rightarrow e\nu_e$ ou $W \rightarrow \mu\nu_\mu$) et 1,2 millions d'événements $t\bar{t}$ ont été utilisés pour faire des comparaisons. Dans la suite, uniquement les jets ayant des impulsions transverses p_T supérieures à 20 GeV/c seront pris en compte dans cette analyse. Des échantillons d'événements enrichis en muons dans les jets sont également utilisés pour déterminer l'efficacité d'identification des jets b . La liste des échantillons utilisés est détaillée dans l'appendice B.

Les propriétés cinématiques des événements QCD sont illustrées sur la figure **4.23** : l'impulsion transverse p_T et la pseudo-rapacité $|\eta|$ des jets, le nombre de traces associées par jet et la distance minimale $\Delta R_{min} = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ entre les jets sélectionnés. En appliquant une sélection sur $\Delta R_{min} (> 2)$, (symboles ouverts de la figure **4.23**), les fractions de quarks issus de processus $g \rightarrow q\bar{q}$ et de gluons issus de radiation peuvent être réduites et l'impulsion transverse ainsi que la multiplicité de traces peuvent être augmentées pour chaque intervalle en \hat{p}_T . Pour l'instant, une telle coupure n'est pas appliquée car elle ne change pas significativement les résultats.

L'efficacité de reconstruction des traces et des jets peut rendre la comparaison entre données et Monte-Carlo délicate dans le cas où ces efficacités ne sont pas bien décrites par la simulation. Dans les analyses de physique, il est nécessaire de découpler les effets des reconstructions des mesures d'efficacités d'identification de jets b . A ces fins, un critère supplémentaire peut être introduit : la *taggabilité*. Un jet peut être défini comme étiquetable s'il a au moins n traces associées (cf. section 4.1.1). Typiquement, $n = 1$ ou 2 ($n = 0$ signifierait qu'il y a un accord parfait entre données et simulation). La *taggabilité* est alors simplement définie comme le rapport entre le nombre de jets étiquetables et le nombre de jets reconstruits. La "taggabilité" a été testée en combinant différents échantillons d'événements QCD PYTHIA (`CMSSW_1_3_1`).

La "taggabilité" pour les événements QCD Monte-Carlo est présentée sur la figure **4.24** pour $n = 1$ et $n = 2$ en fonction du p_T et $|\eta|$ des jets (pour différents intervalles en p_T [20;30], [30;40] et > 40 GeV/c). La taggabilité augmente avec le p_T des jets et atteint 98% pour des impulsions transverses supérieures à 40 (100) GeV/c pour $n = 1$ ($n = 2$). Cela est dû à l'augmentation du nombre de traces reconstruites par jet (pour des jets de grands p_T , les traces associées ont également de plus grands p_T et donc de

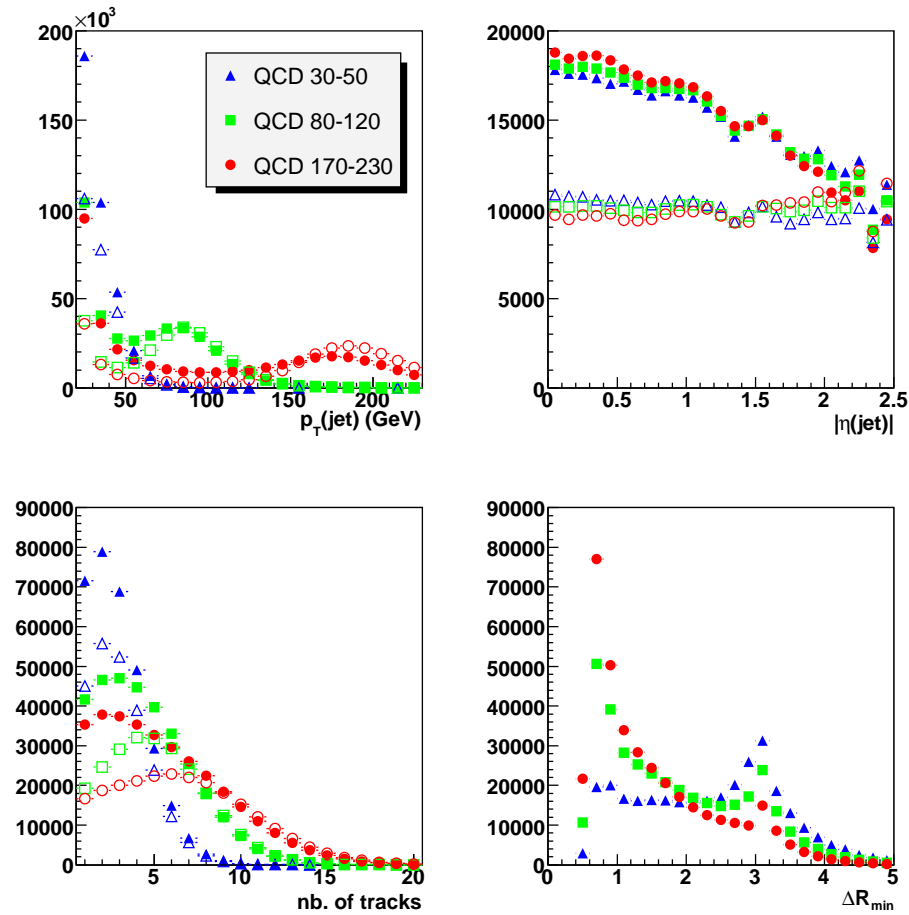


FIG. 4.23: Propriétés cinématiques des jets dans les échantillons *QCD* normalisés au même nombre d'entrée : l'impulsion transverse (en haut à gauche), la pseudo rapidité (en haut à droite) et la multiplicité de traces (en bas à gauche) des jets ainsi que la distance ΔR la plus courte entre deux jets. Symboles pleins : tous les événements, symboles creux : une coupure $\Delta R_{\min} > 2$ est appliquée.

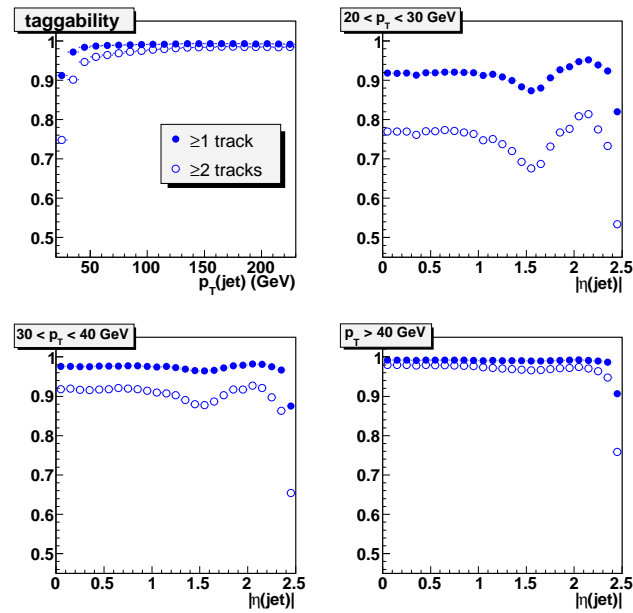


FIG. 4.24: Taggabilité pour $n \geq 1$ or $n \geq 2$ traces associées aux jets pour différents intervalles en p_T dans des événements QCD .

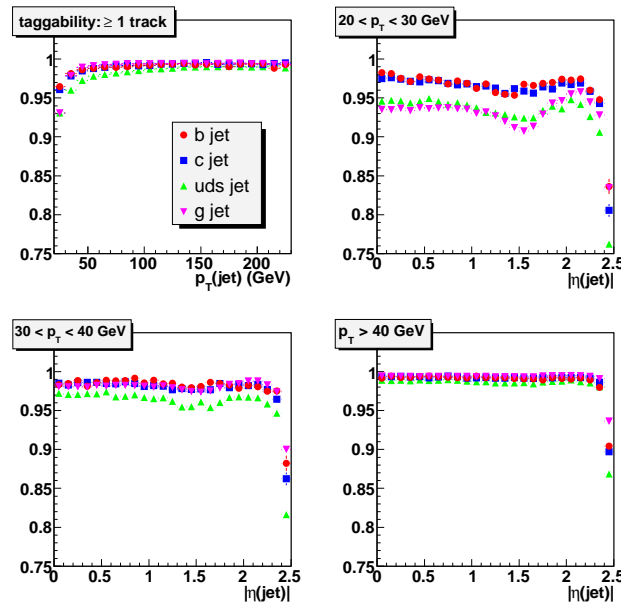


FIG. 4.25: Taggabilité pour $n \geq 1$ pour différentes saveurs de jets et pour différents intervalles en p_T dans des événements QCD .

meilleurs efficacités de reconstruction). Une perte de *taggabilité* est visible dans le cas de l'intervalle $20 < p_T < 30$ GeV/c dans la région $1.2 < |\eta| < 1.8$, qui correspond à la transition entre la partie centrale du détecteur et les bouchons.

Pour $n = 1$, la taggabilité est présentée sur la figure 4.25 pour différents saveurs de jets et pour différents intervalles de p_T . Des différences de l'ordre du pourcent sont observées pour $p_T > 40$ GeV/c. A plus bas p_T , la taggabilité est plus faible dans le cas des jets de gluons et de quarks u , d et s que dans le cas des jets de quarks c et b , le p_T des traces associées au jets légers étant plus faible que celui des traces associées aux jets lourds.

Les algorithmes de comptage de traces et de Vertex Secondaire requièrent au moins deux traces associées par jet tandis que l'algorithme de probabilité par jet peut être utilisé si une seule trace est associée au jet. Ainsi, de manière à ne biaiser aucun algorithme, les jets étiquetables sont définis par $n = 1$. Par la suite, l'efficacité d'identification des jets b sera redéfinie comme :

$$\varepsilon_q = \frac{\text{nombre de jets de saveur } q \text{ identifiés comme jet } b}{\text{nombre de jets de saveur } q \text{ étiquetables}}. \quad (4.5)$$

4.3 Mesure de l'efficacité d'identification des jets de quark beau

La méthode *System8* a été développée pour mesurer l'efficacité d'identification des jets b à partir des données [91] et a été utilisée par la collaboration DØ pendant le run II du Tevatron [92, 93]. Elle utilise plusieurs critères d'identification des jets b faiblement corrélés qui sont appliqués à un même lot d'événements. En combinant ces critères, un système d'équations peut être défini dont la résolution permet d'estimer l'efficacité d'identification des jets b . Une adaptation de la méthode *System8* à l'expérience CMS sera présentée dans ce chapitre [94]. Dans cette méthode, la simulation Monte-Carlo n'est utilisée que pour déterminer les facteurs de corrélation entre les différents critères d'identification. L'application de la méthode nécessite une grande quantité d'événements contenant des muons dans les jets. Ces muons sont principalement produits par la désintégration faible de hadrons beaux. La liste des échantillons utilisés est présentée dans l'annexe B, table B.1. Il s'agit principalement d'événements *QCD*, $b\bar{b}$ et $c\bar{c}$ produits par la version *CMSSW_1_3_1* du code de CMS.

Un mélange des ces événements Monte-Carlo est utilisé pour tester la méthode avec les algorithmes de comptage de traces et de probabilité par jet. Les performances de ces algorithmes dans la version *CMSSW_1_3_1* sont représentées sur la figure 4.26 pour des événements *QCD* ($80 < \hat{p}_T < 120$ GeV/c). Elles sont significativement moins bonnes que pour les versions plus récentes de *CMSSW* (cf. section 4.5) qui présentent de meilleures efficacités de reconstruction des traces (cf. section 2.3.2.1).

Les trois points de fonctionnement (*loose*, *medium* et *tight*) décrits par le tableau 4.1 seront étudiés par la suite. Ils correspondent respectivement à des taux de mauvaise identification des jets légers (ε_{udsq}) de 0.1, 1 et 10 % environ.

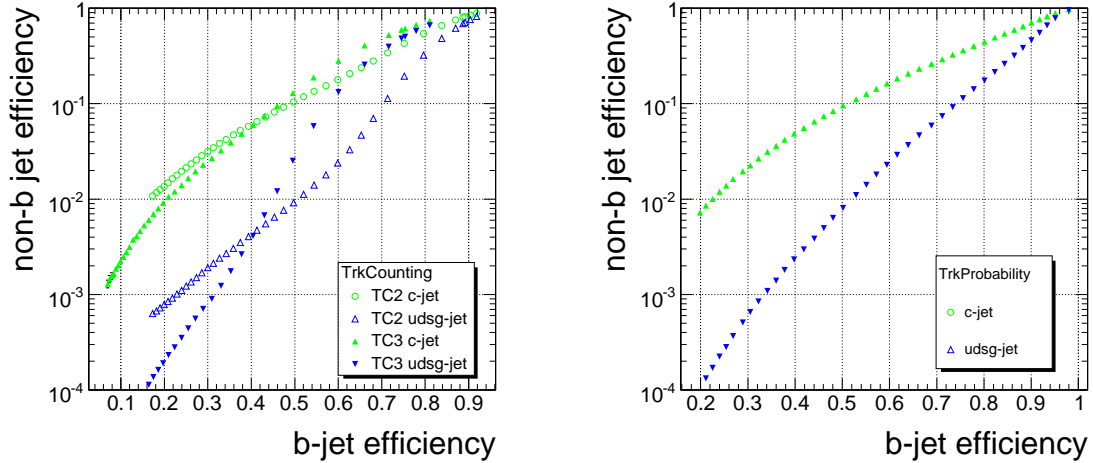


FIG. 4.26: Performances des algorithmes TCHE et TCHP (à gauche) et de l'algorithme de probabilité par jet (à droite) pour des événements *QCD* PYTHIA $80 < \hat{p}_T < 120$ GeV/c générés par la version *CMSSW_1_3_1*.

La reconstruction des muons est détaillée dans la section 2.4 et la reconstruction des jets dans la section 2.5.4. Les jets corrigés sélectionnés vérifient un $p_T > 20$ GeV/c et $|\eta| < 2.5$. Des coupures de qualité sont également appliquées sur les muons :

- $p_T^\mu > 6$ GeV/c,
- $|\eta^\mu| < 2.5$,
- χ^2 normalisé de l'ajustement global < 5 ,
- nombre de coups reconstruits dans le trajectographe > 7 .

Un muon est associé à un jet s'il se trouve dans un cône $\Delta R < 0.4$ autour de l'axe du jet. Si plus d'un muon est trouvé dans le jet, le muon avec le plus grand p_T est sélectionné. L'objet issu de cette association sera appelé par la suite **muon-jet**.

La méthode *System8* utilise donc des événements enrichis en muon-jet. Trois critères d'identification différents sont utilisés,

1. **le premier** correspond à l'algorithme d'étiquetage étudié pour un point de fonctionnement donné.
2. **Le second** est une coupure sur l'impulsion relative p_T^{rel} du muon. Une coupure de $p_T^{rel} > 0.8$ GeV/c est choisie de manière à obtenir des efficacités similaires pour

Algorithme	Discriminant	eff. b	eff. c	eff. uds	eff. g	eff. $udsg$
Comp. de traces						
Loose	$TCHE > 2.30$	70.49 ± 0.20	32.33 ± 0.16	8.64 ± 0.04	10.43 ± 0.03	9.98 ± 0.02
Medium	$TCHE > 5.30$	50.30 ± 0.21	10.77 ± 0.10	0.92 ± 0.02	0.98 ± 0.01	0.96 ± 0.01
Tight	$TCHP > 4.80$	31.94 ± 0.20	2.93 ± 0.06	0.10 ± 0.01	0.11 ± 0.01	0.10 ± 0.01
Proba. par jet						
Loose	> 0.29	74.23 ± 0.19	34.09 ± 0.16	10.00 ± 0.05	10.06 ± 0.03	10.04 ± 0.02
Medium	> 0.54	51.60 ± 0.21	10.32 ± 0.10	0.92 ± 0.02	0.95 ± 0.01	0.95 ± 0.01
Tight	> 0.75	34.12 ± 0.20	3.13 ± 0.06	0.12 ± 0.01	0.11 ± 0.01	0.11 ± 0.01

TAB. 4.1: Points de fonctionnement et efficacités d'identification correspondantes pour les algorithmes de comptage de trace et de probabilité par jet. Seules les erreurs statistiques sont présentées pour des événements QCD $80 < \hat{p}_T < 120$ GeV/ c .

les jets légers et les jets c . Une coupure supplémentaire de $p_T^{rel} < 3.0$ GeV/ c est utilisée pour limiter la sensibilité sur la queue de distribution du p_T^{rel} et sur les muons mal reconstruits.

3. Comme les quarks b sont produits principalement par paire, le **troisième** critère de sélection consiste à demander la présence d'un jet b identifié supplémentaire de manière à augmenter la fraction de jets b . Dans ce sous échantillon, la grande majorité des muon-jets qui sont étiquetés par l'algorithme étudié proviennent effectivement d'un quark b .

Pour l'échantillon sélectionné, on peut alors définir n comme le nombre de muon-jets et p le nombre de muons-jets lorsqu'un jet additionnel est étiqueté par l'algorithme d'identification de jet b étudié. Le critère de sélection lié à l'algorithme dont on veut estimer l'efficacité (labellé " tag ") et la sélection en impulsion relative transverse $p_T^{rel} > 0.8$ GeV/ c du muon (labellé " mu ") sont appliqués séparément. Les jets sont également divisés en deux catégories, les jets b et les jets c +légers $udsg$ (jets cl), il faut donc que les efficacités sur ces types de jet soient voisines, ce qui justifie la coupure en p_T^{rel} .

Les corrélations entre les différents critères de sélection sont estimées à l'aide de la simulation. Tout d'abord, la trace du muon dans le trajectographe peut être prise en compte dans le calcul des discriminants des algorithmes d'identification de jet b et donc introduire une corrélation entre les critères 1 et 2. Les différences d'efficacité d'identification de jets b entre l'échantillon "muon-jet" et "muon-jet + jet additionnel identifié" sont également prises en compte.

Il est alors possible de construire un système de huit équations à huit inconnues :

$$n = n_b + n_{cl} \quad (4.6)$$

$$p = p_b + p_{cl} \quad (4.7)$$

$$n^{\text{tag}} = \varepsilon_b^{\text{tag}} n_b + \varepsilon_{cl}^{\text{tag}} n_{cl} \quad (4.8)$$

$$p^{\text{tag}} = \beta \varepsilon_b^{\text{tag}} p_b + \alpha \varepsilon_{cl}^{\text{tag}} p_{cl} \quad (4.9)$$

$$n^{\text{mu}} = \varepsilon_b^{\text{mu}} n_b + \varepsilon_{cl}^{\text{mu}} n_{cl} \quad (4.10)$$

$$p^{\text{mu}} = \varepsilon_b^{\text{mu}} p_b + \varepsilon_{cl}^{\text{mu}} p_{cl} \quad (4.11)$$

$$n^{\text{tag,mu}} = \kappa_b \varepsilon_b^{\text{tag}} \varepsilon_b^{\text{mu}} n_b + \kappa_{cl} \varepsilon_{cl}^{\text{tag}} \varepsilon_{cl}^{\text{mu}} n_{cl} \quad (4.12)$$

$$p^{\text{tag,mu}} = \kappa_b \beta \varepsilon_b^{\text{tag}} \varepsilon_b^{\text{mu}} p_b + \kappa_{cl} \alpha \varepsilon_{cl}^{\text{tag}} \varepsilon_{cl}^{\text{mu}} p_{cl} . \quad (4.13)$$

Les membres de gauche des équations (4.6)-(4.13) représentent les nombres totaux de muon-jets avant (n , p) et après l'application des critères de sélections 1 (n^{tag} , p^{tag}), 2 (n^{mu} , p^{mu}) séparément ou 1 et 2 simultanément ($n^{\text{tag,mu}}$, $p^{\text{tag,mu}}$). Dans les membres de droite, les huit inconnues sont les nombres de jets b et c +légers dans les deux échantillons (n_b, n_{cl} , p_b , p_{cl}) et les efficacités d'étiquetage des jets b et c +légers par l'algorithme d'identification étudié et par la coupure en p_T^{rel} ($\varepsilon_b^{\text{tag}}$, $\varepsilon_b^{\text{mu}}$, $\varepsilon_{cl}^{\text{tag}}$, $\varepsilon_{cl}^{\text{mu}}$). La méthode fait l'hypothèse que les efficacités d'identification des jets et les efficacités de la coupure en p_T^{rel} sont factorisables. Les corrélations entre l'algorithme d'identification des jets b et la coupure en p_T^{rel} des muons sont prises en compte par l'introduction des paramètres κ_b pour les jets b et κ_{cl} pour les jets c +légers définis par :

$$\kappa_b = \frac{\varepsilon_b^{\text{tag,mu}}}{\varepsilon_b^{\text{tag}} \varepsilon_b^{\text{mu}}} \quad \text{et} \quad \kappa_{cl} = \frac{\varepsilon_{cl}^{\text{tag,mu}}}{\varepsilon_{cl}^{\text{tag}} \varepsilon_{cl}^{\text{mu}}} . \quad (4.14)$$

Les efficacités du critère d'identification sur les jets additionnels sont prises en compte par l'introduction des paramètres β et α définis comme le rapport entre les efficacités d'identification des jets b et c +légers dans les échantillons "muon-jet" et "muon-jet + jet additionnel identifié" :

$$\beta = \frac{\varepsilon_b^{\text{tag}}(\text{muon-jet} + \text{jet additionnel identifié})}{\varepsilon_b^{\text{tag}}(\text{muon-jet})} , \quad (4.15)$$

$$\alpha = \frac{\varepsilon_{cl}^{\text{tag}}(\text{muon-jets} + \text{jet additionnel identifié})}{\varepsilon_{cl}^{\text{tag}}(\text{muon-jet})} . \quad (4.16)$$

Les facteurs de corrélation sont estimés séparément à partir d'événements QCD , $b\bar{b}$ et $c\bar{c}$ semi-leptoniques, $b\bar{b}$ et $c\bar{c}$ inclusifs et sont représentés sur la figure 4.27 pour le point de fonctionnement *medium* des algorithmes de comptage de traces et de probabilité

par jet en fonction du p_T des jets. Pour la résolution du système d'équations (4.6)-(4.13), comme les facteurs β et κ_b dépendent du p_T et du $|\eta|$, leurs valeurs moyennes sont calculées pour les intervalles en p_T des jets : [20;40], [40;60], [60;80] et [80; ∞ [GeV/ c et en $|\eta|$: [0;1.5] et [1.5;2.5]. Pour chaque intervalle, l'erreur de la moyenne des paramètres β et κ_b est propagée au calcul de l'incertitude statistique des efficacités mesurées. Les facteurs α et κ_{cl} semblent peu dépendants du p_T et du $|\eta|$, ils sont ajustés par une constante sur tout le spectre en p_T et $|\eta|$ des jets.

Les erreurs systématiques sont obtenues en faisant varier les facteurs de corrélation β , α , κ_b et κ_{cl} pour différents échantillons Monte-Carlo et en observant les variations de l'estimation de l'efficacité d'identification des jets b , comme nous le verrons plus loin

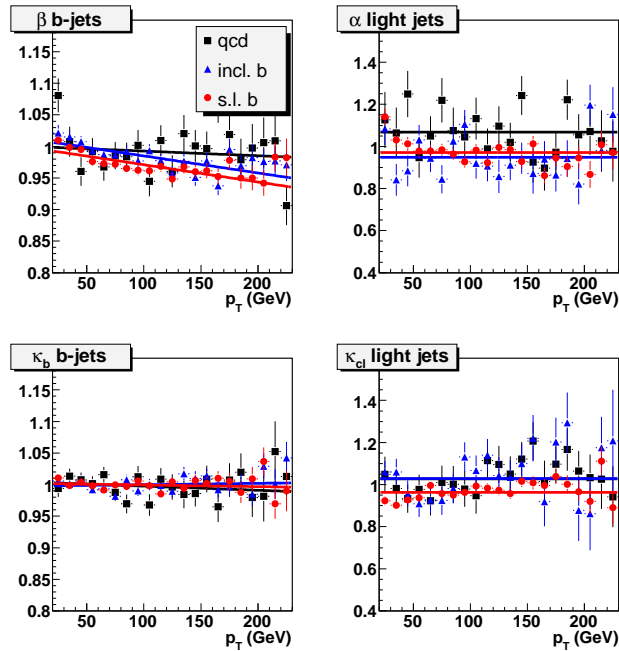


FIG. 4.27: Facteurs de corrélation pour le point de fonctionnement *medium* de l'algorithme de comptage de traces.

Les efficacités d'identification des jets b sont alors estimées en résolvant le système d'équations (4.6)-(4.13) pour chaque intervalle en p_T et $|\eta|$ de jets séparément. Les figures 4.29 et 4.30 présentent les efficacités d'étiquetage prédites par *System8* pour les algorithmes de comptage de traces et de probabilité par jet respectivement, pour le point de fonctionnement *medium* en fonction de l'impulsion transverse p_T et de la pseudo-rapacité $|\eta|$ des jets. Le mélange d'événements décrit dans le tableau B.1 de l'annexe

B est utilisé. La prédiction est comparée avec l'efficacité déterminée avec l'information au niveau du générateur. Les graphiques sont représentés avec des barres d'erreurs provenant uniquement des erreurs statistiques. Un assez bon accord entre mesure et information générateur est obtenu, excepté pour le plus faible intervalle en p_T [20;30] GeV/c, où l'efficacité mesurée est plus faible que l'efficacité attendue.

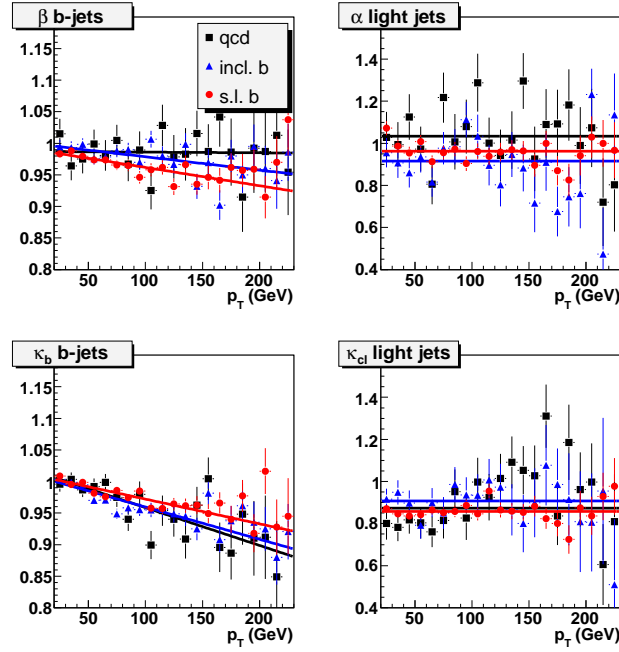


FIG. 4.28: Facteur de corrélation pour le point de fonctionnement *medium* de l'algorithme de probabilité par jet.

L'intervalle [20;30] GeV/c est dominé par des événements *QCD* inclusifs, sa composition en jets *b* est donc différente des autres intervalles en p_T . De plus, l'efficacité de reconstruction et la résolution en énergie des jets de gluons, majoritaires à bas p_T , sont moins bonnes. Cet intervalle est exclu des graphiques présentant les mesures d'efficacité en fonction de $|\eta|$.

Le système d'équations (4.6)-(4.13) ne converge pas pour des impulsions transverses $p_T > 80$ GeV/c. Plusieurs explications sont en court d'investigation : un nombre d'événements trop faible, un manque de discrimination entre jets *b* et jets *c*+légers de la coupure en p_T^{rel} pour les grands p_T ou encore une corrélation entre les critères de sélection 2 et 3. Dans ce dernier cas, de récents résultats [95] montrent que l'introduction de facteurs

de corrélations supplémentaires permet de résoudre ces problèmes de convergence.

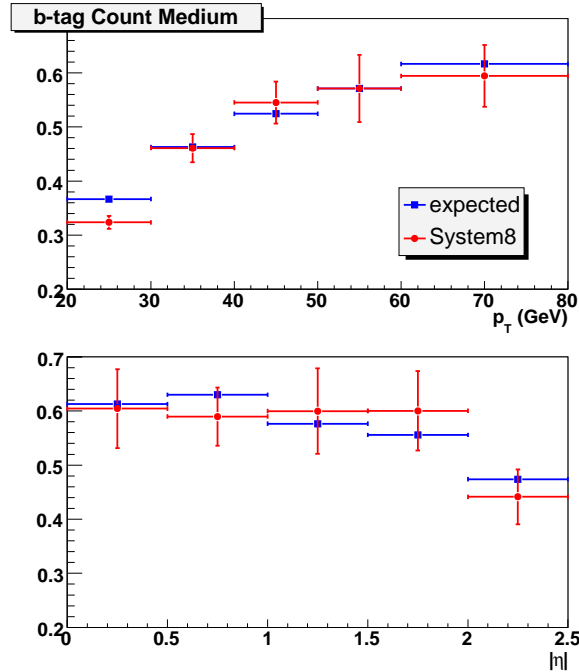


FIG. 4.29: Efficacités d'identification des jets b attendues (information de la génération) et mesurée par la méthode *System8* en fonction du p_T et du $|\eta|$ des jets (avec $p_T > 30$ GeV/c) pour le point de fonctionnement *medium* de l'algorithme de comptage de traces. Seules les erreurs statistiques sont représentées.

Les erreurs systématiques des efficacités mesurées sont évaluées en variant les facteurs de corrélations β , α , κ_b and κ_{cl} par la plus forte différence observée d'un échantillon Monte-Carlo à l'autre. La plus grande différence est observée en comparant les événements QCD et les événements $b\bar{b}$ et $c\bar{c}$. Cela correspond à une incertitude de ± 0.02 pour β et ± 0.01 pour κ_b . Pour α et κ_{cl} les erreurs systématiques sont significativement plus grandes (± 0.2), mais l'incidence sur l'incertitude finale des l'efficacité ε_b est plus faible.

Une source supplémentaire d'incertitude provient des différences de la forme de la distribution p_T^{rel} pour les jets c et $udsg$. Afin d'évaluer cet effet, la coupure en p_T^{rel} est variée entre 0.6 GeV/c et 1 GeV/c et le paramètre κ_b correspondant est estimé. Les variations de l'efficacité d'étiquetage de jets b observées sont considérées comme une incertitude systématique supplémentaire.

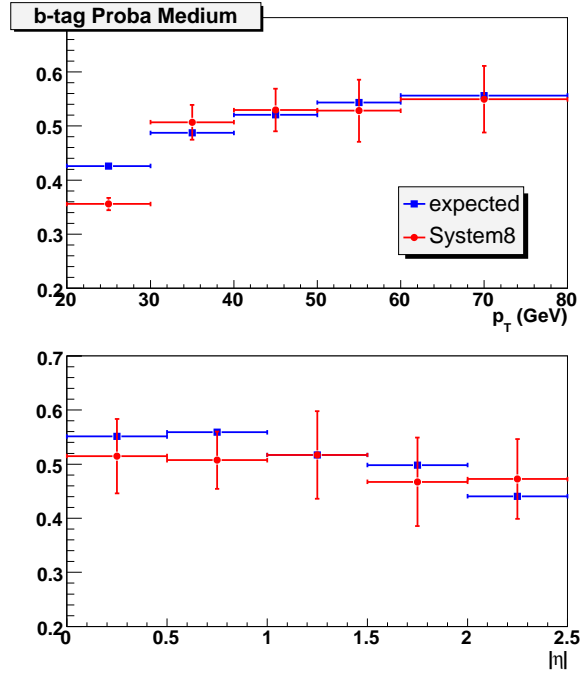


FIG. 4.30: Efficacités d'identification des jets b attendue (information de la génération) et mesurée par la méthode *System8* en fonction du p_T et du $|\eta|$ des jets (avec $p_T > 30$ GeV/ c) pour le point de fonctionnement *medium* de l'algorithme de probabilité par jet . Seules les erreurs statistiques sont représentées.

L'erreur systématique globale sur la mesure de l'efficacité d'identification des jets b a été estimée pour différentes luminosités intégrées. Le taux d'acquisition attendu pour le trigger $b \rightarrow \mu$ HLT [96] est de 4 Hz pour une luminosité instantanée de $L = 10^{32} \text{cm}^{-2} \text{s}^{-1}$. Les nombres d'événements enregistrés attendus sont donc de 400 000, 4 millions et 40 millions pour des luminosités intégrées de 10, 100 et 1000 pb^{-1} , respectivement.

Les erreurs statistiques sont évaluées pour ces trois luminosités à partir des jets contenus dans l'intervalle de p_T [40;50] GeV/ c . Le nombre total d'événements dans le Monte Carlo est d'environ 1.4 fois le nombre d'événements attendus pour une luminosité intégrée de 10 pb^{-1} . L'erreur statistique pour $\mathcal{L} = 10 \text{pb}^{-1}$ est donc simplement obtenue en renormalisant l'erreur statistique du Monte-Carlo par $\sqrt{1.4} = 1.17$. Les erreurs statistiques pour $\mathcal{L} = 100$ et 1000 pb^{-1} sont obtenues en divisant les erreurs pour $\mathcal{L} = 10 \text{pb}^{-1}$ par les facteurs $\sqrt{10}$ et $\sqrt{100}$ respectivement.

Les incertitudes attendues sont résumées dans les tables 4.2 et 4.3 pour des luminosités intégrées de 10, 100, 1000 pb^{-1} et pour les algorithmes de comptage de traces et de probabilité par jet, pour les jets ayant des p_T compris entre 40 et 50 GeV/ c .

Points de fonctionnement Luminosité (pb^{-1})	<i>loose</i>			<i>medium</i>			<i>tight</i>		
	10	100	1000	10	100	1000	10	100	1000
Systématiques (%)									
β	5.8	5.8	2.9	6.3	6.3	3.2	5.7	5.7	2.9
α	0.4	0.4	0.2	0.4	0.4	0.2	0.4	0.4	0.2
κ_b	3.4	3.4	1.7	3.6	3.6	1.8	3.3	3.3	1.7
κ_{cl}	0.2	0.2	0.1	0.2	0.2	0.1	0.2	0.2	0.1
p_T^{rel}	2.8	2.8	2.8	2.9	2.9	2.9	3.0	3.0	3.0
statistique MC (%)	2.3	2.3	2.3	2.6	2.6	2.6	2.7	2.7	2.7
statistique data (%)	7.2	2.3	0.7	8.4	2.6	0.8	8.7	2.7	0.9
Erreur totale(%)	10.5	8.0	6.4	11.8	8.6	5.4	11.6	8.2	5.3

TAB. 4.2: Incertitudes attendues sur les efficacités d'identification des jets b des algorithmes de comptage de traces, mesurées par la méthode *System8* pour différentes luminosités intégrées.

4.4 Estimation du taux de mauvaise identification

Du point de vue des analyses de physique, en plus de l'estimation des efficacités d'identification des jets b , il est important de connaître les efficacités de sélection des jets uds et de gluon, également appelé taux de mauvaise identification des jets légers. La méthode d'estimation du taux de mauvaise identification utilise des algorithmes d'étiquetage négatif [97]. La méthode sera présentée dans la section 4.4.1. Les sources d'erreur systématique seront discutées dans la section 4.4.3, et la précision de la méthode sera estimée pour différentes luminosités. Les différents algorithmes d'étiquetage négatifs

Points de fonctionnement	<i>loose</i>			<i>medium</i>			<i>tight</i>		
Luminosité (pb ⁻¹)	10	100	1000	10	100	1000	10	100	1000
Systématiques (%)									
β	6.0	6.0	3.0	5.8	5.8	2.9	5.4	5.4	2.7
α	1.1	1.1	0.6	1.1	1.1	0.6	1.1	1.1	0.6
κ_b	3.6	3.6	1.8	3.6	3.6	1.8	3.4	3.4	1.7
κ_{cl}	0.2	0.2	0.1	0.2	0.2	0.1	0.2	0.2	0.1
p_T^{rel}	3.2	3.2	3.2	4.0	4.0	4.0	3.8	3.8	3.8
statistique MC (%)	2.4	2.4	2.4	2.6	2.6	2.6	2.9	2.9	2.9
statistique data (%)	7.6	2.4	0.8	8.6	2.7	0.9	9.2	2.9	0.9
Erreur totale (%)	11.1	8.5	5.4	12.0	8.9	6.0	12.2	8.6	5.8

TAB. 4.3: Incertitudes attendues sur les efficacités d'identification des jets b de l'algorithme de probabilité par jet, mesurées par la méthode *System8*, pour différentes luminosités intégrées.

sont présentées dans l'appendice A.3. Pour cette analyse, un ensemble des lots *QCD* a été utilisé, il est décrit dans la section 4.2.

4.4.1 Description de la méthode

Nous nous intéresserons à l'algorithme de comptage de traces qui a été utilisé initialement pour faire l'étude de faisabilité de la méthode. Comme présenté dans la figure 4.2, la distribution en signifiante du paramètre d'impact S_{IP} est symétrique autour de 0 au premier ordre pour les jets légers. Il est alors possible de définir un algorithme d'étiquetage négatif, défini comme le nombre de traces N associées aux jets ayant une signifiante du paramètre d'impact **négatif** plus petit qu'une certaine valeur. Le discriminant de l'algorithme est donc la signifiante $S_{IP < 0}$ de la $N^{\text{ième}}$ trace.

L'utilisation de l'algorithme négatif du comptage de traces permet ainsi d'estimer directement le taux de mauvaise identification de l'algorithme du comptage de traces positives.

Il y a cependant deux corrections à apporter :

- Dû à l'incertitude sur la reconstruction du signe des paramètres d'impact ou à cause de la présence de traces issues du vertex primaire, les jets b et c , ces derniers peuvent également dans un certaine mesure être étiquetés négativement. Cela a pour effet d'augmenter l'efficacité d'étiquetage négatif.
- La présence de particules V_0 et de traces issues de conversions de photons dans les jets légers a pour effet d'augmenter la queue de distribution en signifiante du paramètre d'impact positif et donc d'introduire une asymétrie entre la partie positive et la partie négative des distributions en S_{IP} .

Nous verrons que d'autres effets peuvent également intervenir. Ils peuvent être estimés par l'introduction de facteurs correctifs déterminés à partir de la simulation Monte-Carlo :

- afin de prendre en compte la présence de jets b et c étiquetés négativement, le facteur d'échelle $SF_{hf} = \varepsilon_{neg}^{MC}(udsg)/\varepsilon_{neg}^{MC}(all)$ est introduit, où $\varepsilon_{neg}^{MC}(udsg)$ est l'efficacité d'étiquetage négatif pour les jets légers et $\varepsilon_{neg}^{MC}(all)$ est l'efficacité d'étiquetage négatif pour tous les jets dans la simulation Monte-Carlo.
- le facteur d'échelle $SF_{ll} = \varepsilon_{pos}^{MC}(udsg)/\varepsilon_{neg}^{MC}(udsg)$ est introduit pour compenser la présence de particules V_0 dans la partie positive du discriminant.

L'efficacité ε_{udsg} due aux jets de quarks légers et de gluons peut être alors mesurée par la relation :

$$\varepsilon^{data}(udsg) = SF_{hf} \cdot SF_{ll} \cdot \varepsilon_{neg}^{data}, \quad (4.17)$$

$$= R_{light} \cdot \varepsilon_{neg}^{data}, \quad (4.18)$$

où

- ε_{neg}^{data} est le taux d'étiquetage négatif dans les données. Comme dans le cas de l'efficacité d'identification usuelle, il s'agit du nombre de jets étiquetés négativement (c'est-à-dire étiquetés par l'application de l'algorithme négatif) divisé par le nombre de jets étiquetables.
- $R_{light} = \varepsilon^{MC}(udsg)/\varepsilon_{neg}^{MC}$ est le rapport entre le taux de mauvaise identification ($udsg$) et l'efficacité d'étiqueter les jets négativement (pour toutes les saveurs $udsg+c+b$) dans la simulation.

Il est possible de définir des algorithmes d'identification négatifs pour la plupart des algorithmes existant (cf. annexe A.3). Les premiers tests de faisabilité ont été faits et les résultats correspondants sont présentés.

Pour l'algorithme de comptage de traces décrit dans le chapitre 4.1.1.1, deux discriminants peuvent être choisis en demandant au moins 2 traces ($TCH E$ utilisé pour obtenir une grande efficacité d'identification) ou 3 traces ($TCH P$ utilisé pour obtenir une grande pureté) ayant des significances de paramètre d'impact plus grandes qu'une certaine valeur (S_{IP}). L'algorithme d'identification négatif est quand à lui défini par la condition d'avoir au moins 2 ou 3 traces avec des significances de paramètre d'impact négatives plus petites que l'opposée de cette même valeur ($-S_{IP}$).

Les performances de l'algorithme de comptage de traces sont représentées par la figure 4.26 pour des événements Monte-Carlo PYTHIA QCD avec $80 < \hat{p}_T < 120$ GeV/c produits avec la version $CMSSW_1_3_1$ du code de CMS . Les points de fonctionnement et les efficacités sont présentées dans le tableau 4.1.

L'évaluation du taux de mauvaise identification est sensible à la fraction de jets b et c dans la partie négative des distributions de discriminant (ce qui tend à diminuer R_{light}), et à la fraction de traces venant d'autres processus produisant des traces à grands pa-

ramètres d'impact tels que les particules V_0 (issues de la désintégration de K_S^0 , Λ et de conversions de photon), les interactions nucléaires dans le détecteur et des fausses traces ou traces mal reconstruites (ce qui tend à augmenter R_{light}). Les différences entre les jets de quarks légers et de gluons peuvent également affecter le rapport R_{light} . L'incertitude sur tous ces effets doit être soigneusement prise en compte dans le calcul des erreurs systématiques (cf. 4.4.3).

En utilisant le point de fonctionnement *medium*, les efficacités $\varepsilon^{MC}(udsg)$ et $\varepsilon^{MC}(uds)$, $\varepsilon^{MC}(g)$ et ε_{neg}^{MC} sont présentées sur les figures 4.31 et 4.32 (graphiques en haut à gauche) en fonction de l'impulsion transverse p_T et de la pseudo-rapacité $|\eta|$ des jets pour le point de fonctionnement *medium*. On remarque que ces efficacités augmentent avec le p_T . Cette augmentation est corrélée à l'augmentation de la multiplicité de traces dans les jets de haut p_T . La diminution de ces efficacités à grande pseudo-rapacité $|\eta| \approx 2.4$ est reliée à la réduction de l'acceptance du trajectographe et à une diminution de l'efficacité de reconstruction des traces dans les parties avant du détecteur. Le taux de mauvaise identification est plus important dans les bouchons que dans la partie centrale du détecteur, cela est dû au taux de fausses traces qui est plus important dans les bouchons.

Le rapport R_{light} (figure 4.31, en haut à droite) est d'environ 1,4 pour les jets ayant une impulsion transverse $p_T > 40$ GeV/c. Au delà de cette impulsion transverse, il ne semble pas dépendre de p_T . On remarque également (figure 4.32) que le facteur R_{light} est plus grand dans les bouchons car le taux de fausses traces y est plus élevé.

Sur les mêmes figures, (en bas à gauche, symboles creux) on peut remarquer que les différentes fractions de saveurs de jets évoluent avec p_T et $|\eta|$, en particulier pour les faibles impulsions transverses. Pour $p_T > 40$ GeV/c, les fractions moyennes de jets uds , c , b et g sont de 29%, 6%, 4% et 61% respectivement dans la simulation. Ces mêmes fractions sont représentées sur les mêmes graphiques après étiquetage négatif (symboles pleins) : on observe que les contributions relatives des jets b et c augmentent. L'incertitude sur les fractions de jets g , c et b va avoir un effet sur le taux d'étiquetage et la valeur du facteur R_{light} . Cela doit être pris en compte dans le calcul des erreurs systématiques liées aux fractions de jets b , c et de gluon (graphiques en bas à droite). Elles seront estimées en faisant varier les fractions de jets b , c et de gluons de $\pm 50\%$, $\pm 50\%$ et $\pm 20\%$ respectivement. Il apparaît clairement que les contributions provenant de jets de quark b après étiquetage négatif (ε_{neg}^{MC}) dominent ces erreurs systématiques sur R_{light} . Les différentes erreurs systématiques seront présentées dans la section 4.4.3.

Afin de contrôler ces systématiques, les fractions de jets b et c peuvent être réduites de manière significative par l'application d'un veto sur les jets. Dans l'estimation de l'efficacité d'étiquetage négatif, les jets possédant des traces à grande signification positive du paramètre d'impact peuvent être rejetés. Cela est mis en évidence pour le point de fonctionnement *medium* sur les figures 4.33 et 4.34 où les jets étiquetés négativement

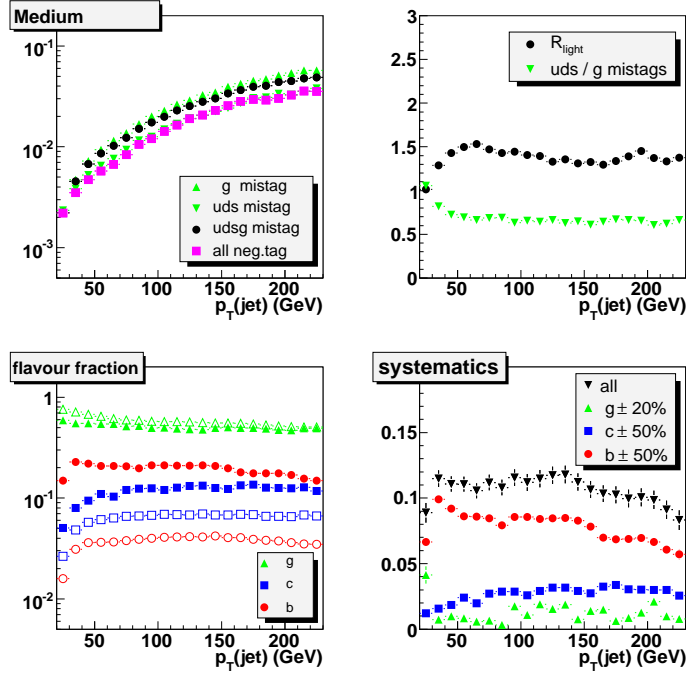


FIG. 4.31: Dans les échantillons QCD et avec le point de fonctionnement *medium*, en fonction du p_T des jets : (en haut à gauche) les taux de mauvaise identification $\varepsilon^{MC}(udsg)$ (cercles), $\varepsilon^{MC}(uds)$ et $\varepsilon^{MC}(g)$ (triangles) et l'efficacité d'étiquetage négatif $\varepsilon_{neg}^{MC}(udsg + c + b)$ pour toutes les saveurs (carrés), (en haut à droite) le facteur R_{light} (cercles) et le rapport $\varepsilon^{MC}(uds)/\varepsilon^{MC}(g)$ (triangles), (en bas à gauche) les fractions de jets b , c et g avant (symbole creux) et après étiquetage négatif (symboles pleins) et (en bas à droite) les erreurs systématiques provenant de l'incertitude sur les fractions de jets g , c et b (cf. section 4.4.3) qui sont mal connues aux collisionneurs hadroniques.

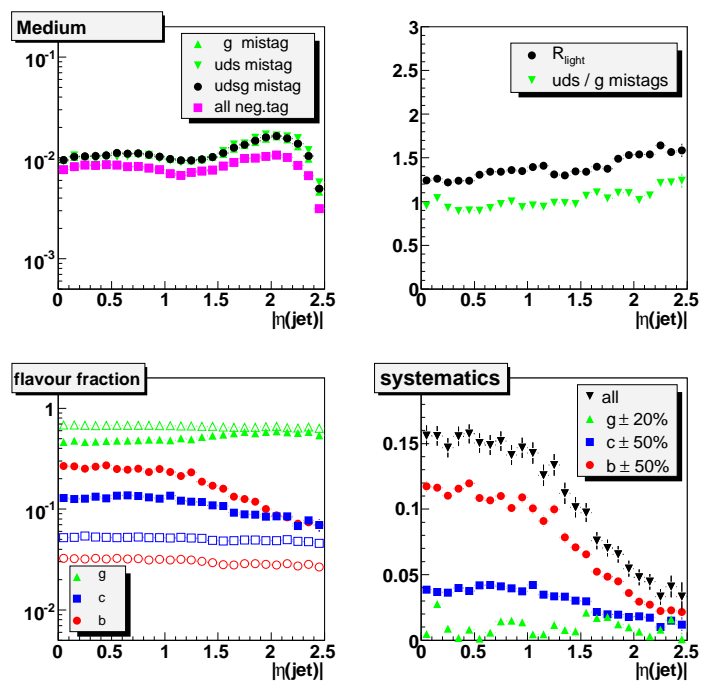


FIG. 4.32: Identique à la figure 4.31 en fonction du $|\eta|$ des jets.

sont rejetés s'ils possèdent au moins une trace avec $S_{IP} > 4$. Une étude plus détaillée des erreurs systématiques sera présentée dans la section 4.4.3. On peut cependant remarquer que le nombre total de jets sélectionnés par l'étiquetage négatif est réduit par un facteur 0.7. Les facteurs de réduction sur les jets légers, c et b étiquetés négativement sont de 0.6, 0.3 et 0.1 respectivement. De ce fait, l'efficacité d'étiquetage négatif est réduite et le facteur R_{light} augmente d'un facteur 1.6 en moyenne lorsque le veto des jets est appliqué.

Les erreurs systématiques totales sont alors réduites par un facteur 5 et dominées par l'incertitude sur la fraction des jets de gluon. Les résultats correspondant pour les points de fonctionnement *loose* et *tight* sont représentés en annexe A.2.

Il est à noter que pour les jets de $p_T > 40$ GeV/c (figure 4.31 en haut à droite) le rapport des efficacités d'identification entre jets uds et jets de gluons est d'environ 0.6 ± 0.1 pour tous les points de fonctionnement. Ceci est principalement dû aux différences, entre jets de quark léger et jets de gluon, de multiplicité de traces et de p_T des traces associées aux jets.

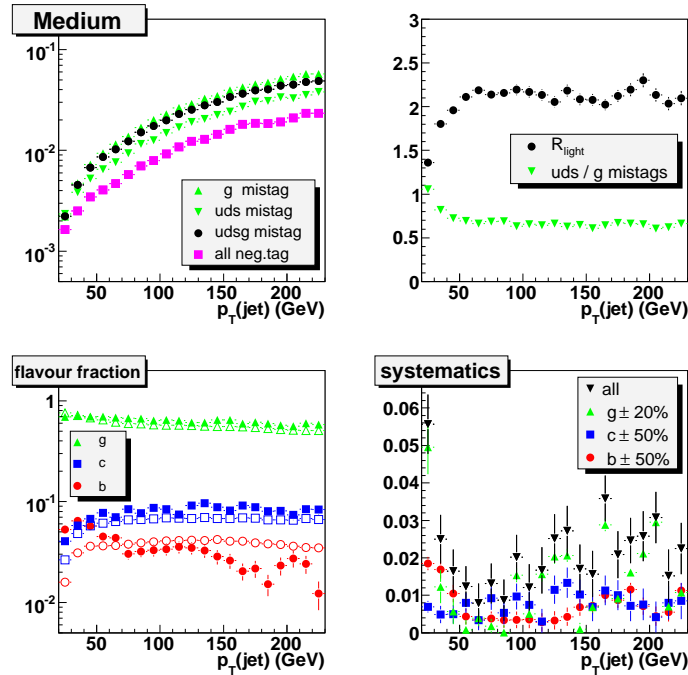


FIG. 4.33: Identique à la figure 4.31 après application du veto sur les jets : les jets étiquetés négativement sont rejetés s'ils possèdent au moins une trace avec $IP/\sigma_{IP} > 4$.

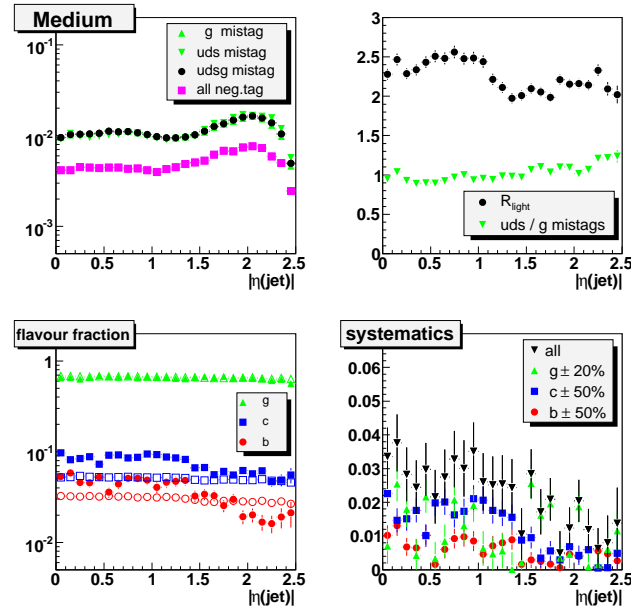


FIG. 4.34: Identique à la figure 4.32 après application du veto sur les jets : les jets étiquetés négativement sont rejetés s'ils possèdent au moins une trace avec $IP/\sigma_{IP} > 4$.

4.4.2 Etudes des traces provenant de processus déplacés

Génériquement, on appellera processus déplacé tout processus physique ou processus provenant de l'interaction avec le détecteur qui produit des traces à grand paramètre d'impact. Il s'agit de processus tels que la désintégration de particules à grande durée de vie comme les K_S^0 et Λ , de conversions de photons en paires e^+e^- , d'interactions nucléaires dans le détecteur, de traces mal reconstruites et de fausses traces (effets de mauvaise reconstruction) [82]. Tous ces processus affectent les algorithmes d'étiquetage basés sur la grande durée de vie des hadrons beaux (comptage de traces, probabilité par jet et algorithmes utilisant les vertex secondaires) et conduisent à des asymétries entre algorithmes d'étiquetage positif et négatif pour les jets légers. Ils doivent donc être pris en compte dans le calcul des erreurs systématiques. Nous appellerons de manière générique traces V^0 les traces provenant de la désintégration des K_S^0 et Λ et de conversions de photons.

L'identification de l'origine des traces est possible grâce l'outil *TrackHistory* [83] qui permet de déterminer la généalogie complète des traces et des vertex associés à l'aide de l'information générée et simulée. Cet outil associe les traces reconstruites aux traces simulées par *Geant4* si elles existent. Ensuite, les différents parents de ces traces simulées ainsi que leurs vertex associés sont identifiés. Les traces reconstruites sont

alors caractérisées par leur généalogie complète. A titre d'illustration, la figure 4.35 représente les distributions en paramètre d'impact longitudinal (LIP), transverse (TIP), la distance entre l'axe du jet et la trace (DJA) et la longueur de désintégration signée (SDL), présentés dans la section 4.1.1, pour les 5 catégories de traces suivantes :

1. *Fausses traces (Fake)* : traces reconstruites qui ne sont associées à aucune trace simulée.
2. *Mauvaises traces (Bad)* : traces ne provenant pas de la désintégration de hadrons b mais dont la différence entre les paramètres d'impact transverse reconstruits et simulés divisé par l'erreur est plus grande que 3. De telles traces correspondent à des traces mal reconstruites.
3. *Traces déplacées (Displaced)* : traces ne provenant pas de hadrons b mais associées à au moins un vertex simulé autre que le vertex primaire.
4. *Traces provenant de hadrons beaux (B)* : traces provenant de la désintégration de hadrons beaux.
5. *Traces ne provenant pas de hadrons beaux (Non B)* : traces ne provenant pas de la désintégration de hadrons b et n'appartenant à aucune autre catégorie.

La figure 4.35 illustre donc une partie des sources de vertex déplacés.

Un jet est défini comme jet V_0 si l'une des deux (trois) traces ayant les significances de paramètre d'impact les plus grandes est issue de la désintégration de particules K_S^0 , Λ ou de conversions de photons pour l'algorithme de comptage de traces *TCHE* (*TCHP* respectivement).

Les fractions de jets V_0 pour les points de fonctionnement *loose*, *medium* et *tight* sont représentées sur les figures 4.36 et 4.37 en fonction du p_T et du $|\eta|$ des jets. On remarque qu'après application des algorithmes d'identification des jets b et d'étiquetage négatif, les fractions des jets contenant des particules V_0 augmentent de manière significative : pour des jets de $p_T > 50$ GeV/C, cette fraction est de 18% pour le point de fonctionnement *loose*, 58% pour le *medium* et 70% pour le *tight*.

Cette fraction est cependant moins importante dans le cas des efficacités d'étiquetage négatifs que dans l'efficacité d'étiquetage positif. Cela illustre le fait que la contribution des particules V_0 au taux de mauvaise identification n'est pas suffisamment prise en compte par la mesure du taux d'étiquetage négatif. Cela introduit des erreurs systématiques supplémentaires.

Les distributions en signification du paramètre d'impact S_{IP} sont représentées sur la figure 4.38 (en haut) pour les algorithmes *TCHE* et *TCHP*, pour les jets b , c et $usdg$. Pour les jets $usdg$, les significances du paramètre d'impact sont présentées avec et sans la contribution provenant des jets V_0 . Les traces provenant de jets V^0 uniquement sont représentées dans les graphiques du bas sur la même figure. La contribution des V^0 à la queue de distribution en $S_{IP} > 0$ est d'environ 50%, la contribution principale provenant de conversions de photons. Les autres contributions, principalement les interactions

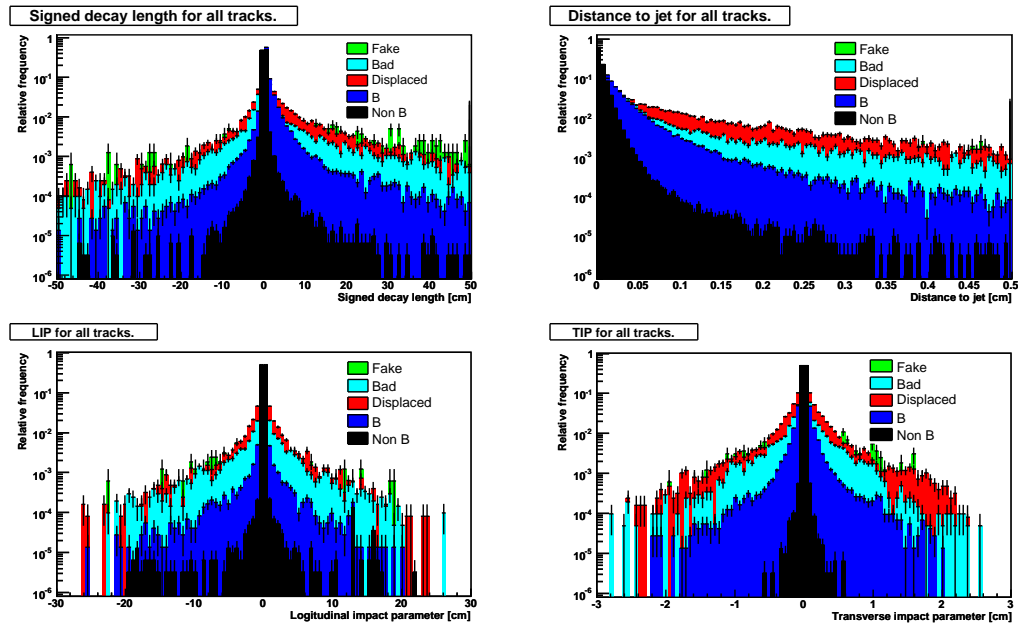


FIG. 4.35: Distributions des différentes propriétés géométriques des traces : la longueur de désintégration (en haut à gauche), la distance à l'axe du jet (en haut à droite), le paramètre d'impact longitudinal (en bas à gauche) et le paramètre d'impact transverse (en bas à droite). Les distributions sont représentées pour les 5 catégories de traces discutées dans le texte.

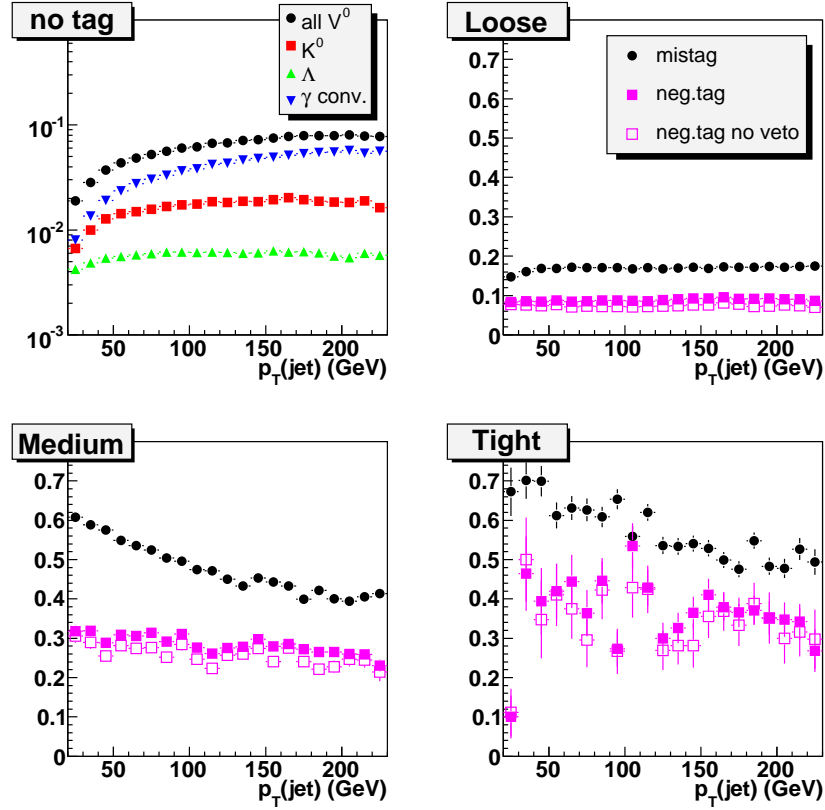
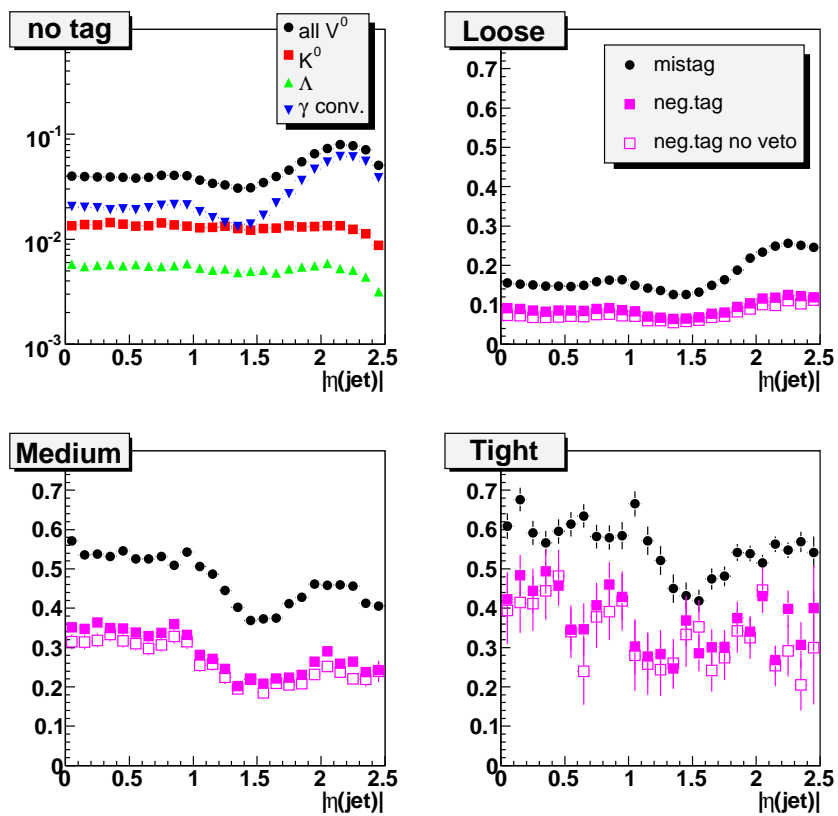


FIG. 4.36: Fractions de jets contenant au moins une trace provenant d'un V^0 pour les jets $udsg$ en fonction du p_T des jets lorsque l'identification des jets b n'est pas appliquée (en haut à gauche). Pour les points de fonctionnement *loose* (en haut à droite), *medium* (en bas à gauche) et *tight* (en bas à gauche) du comptage de traces, fraction de jets contenant au moins une trace provenant d'un V^0 après l'application de l'algorithme d'identification des jets b (point noir) et de l'algorithme d'étiquetage négatif avec (carrés creux) et sans (carrés pleins) l'utilisation du veto sur les jets. Les deux traces ayant les deux (trois) plus grands paramètres d'impact sont utilisées pour déterminer les contributions de particules V_0 pour les points de fonctionnement *loose*, *medium* et *tight*.

FIG. 4.37: Identique à la figure 4.36 en fonction de $|\eta(\text{jet})|$.

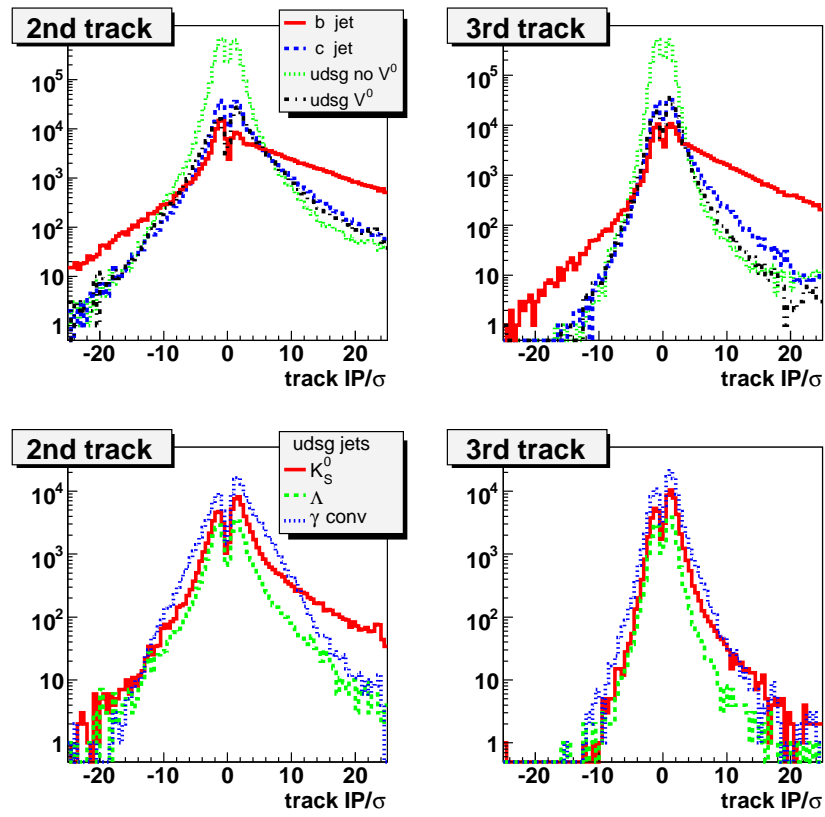


FIG. 4.38: Deuxième (graphique de gauche) et troisième (graphique de droite) plus grande signification du paramètre d'impact par jet pour différentes saveurs (en haut) et différentes contributions de V^0 dans les jets $udsg$ (en bas).

secondaires dans le détecteur ou encore les traces mal mesurées, doivent être prises en compte dans le calcul des erreurs systématiques.

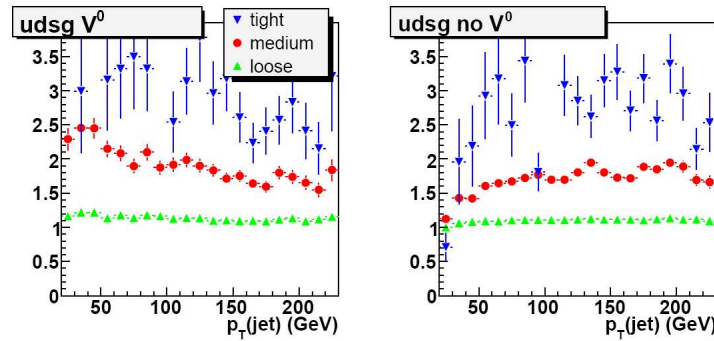


FIG. 4.39: Pour chaque point de fonctionnement, le facteur R_{light} est représenté en fonction du p_T des jets pour les jets $udsg$ contenant les candidats V_0 (à gauche) et les jets $udsg$ sans candidats V_0 (à droite).

Les effets sur le facteur R_{light} sont présentés sur la figure 4.39 avec (à gauche) ou sans (à droite) la contribution des jets V_0 pour les jets $udsg$. Il existe une contribution à R_{light} qui n'est pas due aux V_0 mais dont l'effet est similaire.

Pour conclure cette partie, à l'aide de l'outil *TrackHistory* nous avons mis en évidence le fait que l'excès de traces ayant une grande signification du paramètre d'impact dans les jets $udsg$ n'est pas uniquement dû à la présence de particules V_0 . De ce fait, le calcul des erreurs systématiques devrait également prendre en compte les effets supplémentaires provenant d'interactions nucléaires, de traces mal reconstruites ou de fausses traces.

Il n'est cependant pas possible d'estimer ces dernières contributions dans la version de CMS utilisée pour cette analyse. Nous ferons l'hypothèse que la contribution venant d'interactions nucléaires, de fausses traces ou de traces mal reconstruite est du même ordre de grandeur que la contribution des particules V_0 .

4.4.3 Erreurs systématiques.

Il sera discuté ici des erreurs systématiques liées à la méthode. Tout d'abord, sa précision va dépendre de la luminosité intégrée. Le calcul du taux de mauvaise identification doit donc être évalué pour différentes luminosités intégrées. Il a été estimé que les chemins de déclenchement HLT multi-jets [96] ont un taux d'acquisition de l'ordre de 20 Hz, ce qui correspond à 2, 20 et 200 millions d'événements enregistrés pour des luminosités intégrées de 10, 100, 1000 pb^{-1} respectivement. Il est attendu que les événements issus de chemins de déclenchement multi-jets soient principalement constitués de jets de bas p_T également produits en grande quantité par la simulation Monte-Carlo disponible.

Ceux-ci seront utilisés pour estimer les erreurs systématiques.

Ces dernières proviennent principalement de la confiance que l'on a dans l'estimation du facteur R_{light} à partir de la simulation Monte-Carlo. Les incertitudes sont estimées en faisant varier certains paramètres et en calculant leurs effets sur le facteur R_{light} . Les sources d'erreur prises en compte dans ce calcul sont les suivantes :

- **Fractions de quarks b et c :** Le rapport R_{light} est sensible à l'incertitude sur les fractions de jets b et c produits dans les événements QCD Monte-Carlo. Ces fractions sont généralement mal connues aux collisionneurs hadroniques. Il s'agit de la source principale d'erreur systématique et l'effet sur l'incertitude de R_{light} est lié au nombre de jets b et c étiquetés négativement. Comme nous l'avons vu il est possible de réduire cette incertitude en introduisant un veto sur les jets lors du calcul de $\varepsilon_{neg}^{MC}(all)$ (aucun jet n'ayant une trace avec une signification du paramètre d'impact $S_{IP} > 4$).

L'accord sur la fraction de jets b entre données et simulation peut en principe être évalué à partir des données en comptant les muon-jets dans un échantillon enrichi en jet b (cf. section 4.3). Ces muons proviennent en grande majorité de la désintégration faible de hadrons beaux et peuvent être utilisés pour étiqueter les jets b , ce qui conduit à un échantillon très pur de jets b .

L'efficacité d'étiqueter les jets b peut ensuite être estimée à l'aide d'un lot d'événements indépendants (événements $t\bar{t}$ par exemple [98]), un simple comptage de jets permet d'avoir une estimation de la fraction de jets de saveurs lourdes étiquetés négativement dans les données. Il est ainsi possible de comparer cette fraction avec l'estimation faite dans la simulation Monte-Carlo. Une telle méthode nécessite un nombre d'événements $t\bar{t}$ important. Pour une luminosité intégrée de 1 fb^{-1} , elle permettrait d'estimer la fraction de jets b à 10% près. Dans cette analyse, les fractions de jets b et c sont variées de $\pm 50\%$ (ou $\pm 20\%$) dans la simulation.

- **Fraction de gluons :** Les fractions moyennes de jets de gluons dans les échantillons QCD considérés se situent entre 75% et 50% pour des jets ayant un p_T compris entre 20 et 230 GeV/c. L'incertitude sur ces fractions dépend de l'incertitude sur les densités partoniques utilisées à la génération et des paramètres d'hadronisation. Dans cette analyse, nous ferons l'hypothèse que l'incertitude sur cette fraction de gluon est de l'ordre de $\pm 20\%$. Cette incertitude devrait être réduite pour des luminosités intégrées supérieures à 1 fb^{-1} lorsque les paramètres des générateurs seront optimisés à l'aide des données.
- **Présence de V_0 (K_S^0 , Λ et conversions de photons) :** Le nombre de conversions de photons est sensible à la quantité de matière présente dans le trajectographe. L'incertitude correspondante sur R_{light} sera donc liée à la qualité de la description de cette matière dans la simulation. La production de particules K_S^0 et Λ dans les jets légers est quant à elle reliée au taux de production de quarks étranges, ca-

ractérisé par le paramètre PYTHIA γ_s (facteur de suppression [45]) où les quarks légers sont produits avec une fraction relative $u:d:s = 1:1:\gamma_s$ dans le processus de fragmentation⁷. γ_s a été mesuré par les expériences au *LEP* et à *HERA* ($\simeq 0.2 - 0.3$) [99]. Au *LHC*, le grand nombre d'événements produits devrait permettre d'étudier la production de V^0 de manière précise. Dans cette analyse, les fractions de particules K_S^0 , Λ et de conversion de photons sont variées de $\pm 20\%$ chacune dans la simulation. L'erreur systématique globale est obtenue en sommant quadratiquement les variations correspondantes de R_{light} . Pour une luminosité de 1 fb^{-1} , ces fractions sont supposées être connues et mesurées avec une incertitude de $\pm 10\%$. La dépendance sur R_{light} des erreurs systématiques dues aux V^0 en fonction de p_T et $|\eta|$ des jets est présentée sur la figure 4.41 pour une luminosité de 100 pb^{-1} .

- **Autres processus déplacés** : il a été montré dans la section 4.4.2 que les particules V^0 ne contribuaient que pour la moitié environ de la valeur de $\varepsilon_{pos}^{MC}(udsg)$ observée dans les jets légers. D'autres contributions proviennent d'interactions secondaires dans le détecteur et de fausses traces ou de traces mal reconstruites. Une incertitude du même ordre de grandeur que l'incertitude sur les V_0 est introduite.
- **Taux d'Inversion du signe du paramètre d'impact** : Le calcul du signe du discriminant de l'algorithme d'étiquetage est sensible à la résolution angulaire des jets calorimétriques et des traces. Ceci peut conduire à une mauvaise reconstruction du signe du discriminant (*inversion de signe*) et modifier le nombre de jets b étiquetés négativement. Les différences des taux d'inversion de signe entre données et Monte-Carlo doivent donc également être prises en compte dans le calcul des erreurs systématiques sur R_{light} .

Une méthode pour estimer la fraction de jets b étiquetés négativement (à cause de l'inversion de signe) est proposée dans les données et dans le Monte-Carlo. Les différences observées peuvent être utilisées pour ajuster la simulation et ainsi diminuer les erreurs systématiques liées au taux d'inversion de signe.

L'idée est de sélectionner un lot d'événements enrichis en jets b et d'appliquer les algorithmes d'étiquetage négatif. Il est alors possible de comparer les fractions de jets étiquetés négativement dans les données et dans le Monte-Carlo. L'erreur systématique correspond alors à l'erreur statistique sur la fraction de jets b étiquetés négativement.

Tout d'abord, les événements contenant des muons provenant de la désintégration de hadrons beaux peuvent être sélectionnés suivant la procédure décrite dans la section 4.3. Les distributions des discriminants des algorithmes *TCHE* et *TCHP*

⁷Dans le processus de fragmentation expliqué dans la section 1.6, en moyenne pour un quark u et un quark d produit, γ_s quarks s sont produits.

sont présentés sur la figure 4.40). La fraction de jets b sélectionnés pour le point de fonctionnement *medium* est de $78 \pm 1\%$ ($37 \pm 1\%$) pour l'algorithme d'identification positif (négatif). Pour le point de fonctionnement *tight*, cette fraction devient $91 \pm 1\%$ ($78 \pm 2\%$), la fraction globale de jets étiquetés négativement étant de 0.055 ± 0.002 . Afin d'augmenter la pureté dans le cas du point de fonctionnement *medium*, il est possible d'appliquer une coupure sur le $p_T^{rel} > 1.5$ GeV/ c du muon (bas de la figure 4.40). Pour le point de fonctionnement *medium*, la fraction de jets b après application de l'algorithme d'étiquetage positif (négatif) devient $87 \pm 1\%$ ($70 \pm 1\%$), la fraction globale de jets étiquetés négativement étant de 0.180 ± 0.005 .

En faisant l'hypothèse que les jets b étiquetés négativement correspondent majoritairement à une mauvaise reconstruction du signe des paramètres d'impact⁸, l'erreur relative sur la fraction de jets étiquetés négativement est donc de $\pm(0.005/0.180) \approx \pm 2.8\%$ et $\pm(0.002/0.055) \approx \pm 3.6\%$ pour les points de fonctionnement *medium* et *tight* respectivement. Ces valeurs ont été utilisées par la suite pour estimer l'incertitude due aux effets d'inversion du signe des paramètres d'impact.

Dans la section 4.3, il a été montré qu'environ 400 000 événements contenant un muon-jet devraient être collectés pour une luminosité intégrée de 10 pb^{-1} . Comme précédemment, l'incertitude statistique sur les événements générés peut être extrapolée afin de déterminer l'incertitude relative sur le taux d'inversion du signe du paramètre d'impact pour 10 pb^{-1} . Pour les points de fonctionnement *loose* et *medium*, l'incertitude correspondante sur l'inversion de signe est estimée à $\pm 1.1\%$ et pour le point de fonctionnement *tight* à $\pm 1.4\%$.

En supposant que la quantité d'événements contenant des muon-jets produits par la simulation Monte-Carlo est d'environ 4 millions, l'incertitude sur l'inversion du signe dans le Monte-Carlo est d'environ $\pm 0.4\%$. L'incertitude totale sur le taux d'inversion du signe du paramètre d'impact est alors prise comme la somme quadratique des incertitudes sur les données et sur la simulation Monte-Carlo. L'incertitude correspondante, dominée par l'incertitude sur les données, est propagée à l'incertitude sur le facteur R_{light} en faisant varier la fraction de jets étiquetés positivement et négativement par l'erreur relative.

La figure 4.41 représente les erreurs systématiques dues à la présence des V_0 et à l'inversion du signe pour les trois points de fonctionnement. Pour les points de fonctionnement *loose* et *medium* les erreurs systématiques sont dominées par les incertitudes provenant de la fraction de jets V_0 (principalement de la fraction de conversion de photon. C'est plus particulièrement le cas dans les bouchons où la quantité de matière est plus importante et le taux de conversion plus élevé. Pour le

⁸La contribution provenant des particules émises à proximité du vertex primaire pendant le processus d'hadronisation du quark b est assimilée au taux d'inversion du signe du paramètre d'impact.

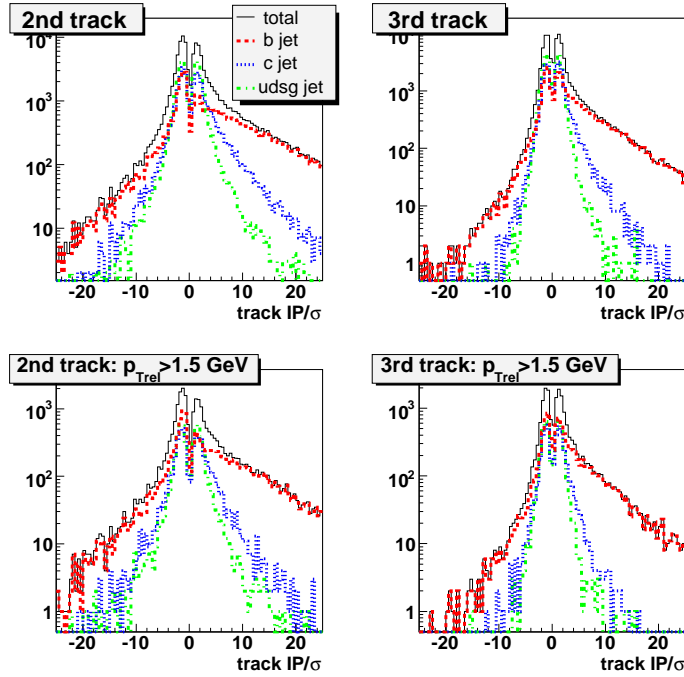


FIG. 4.40: Pour les muon-jets avec un muon de $p_T^\mu > 6 \text{ GeV}/c$, distribution des deuxième (à gauche) et troisième (à droite) plus grandes significances du paramètre d'impact pour tous les jets (en haut) et en appliquant une sélection sur les muons ($p_{Trel} > 1.5 \text{ GeV}/c$, en bas). Aucune normalisation n'est appliquée sur les distributions.

point de fonctionnement *tight*, l'incertitude liée à l'inversion de signe domine car les erreurs statistiques de son estimation deviennent importantes.

- **Nombre d'événements dans la simulation Monte-Carlo et dans les données :** Le nombre d'événements sélectionnés peut être assez important, même pour de faibles luminosités. Cependant, l'efficacité de ε_{udsq} doit être paramétrisée en différents intervalles en p_T , $|\eta|$ et éventuellement en multiplicité de traces. L'erreur statistique est estimée à partir du nombre d'événements attendus, pour une luminosité de 100 pb^{-1} , dans l'intervalle de p_T [80 ; 120] GeV.
- **Dépendance de l'échantillon utilisé :** Pour calculer le rapport R_{light} et l'efficacité d'étiquetage négatif, des échantillons similaires doivent être utilisés dans les données et dans la simulation. De plus, l'utilisation de différents chemins de déclenchement [96] peut changer significativement le contenu d'un échantillon. Une fois les données réelles disponibles, l'erreur due à la composition en événements d'un échantillon donné sera estimée en utilisant différents chemins de déclenchement. De manière à simuler leurs effets (et puisque l'information HLT n'est pas disponible

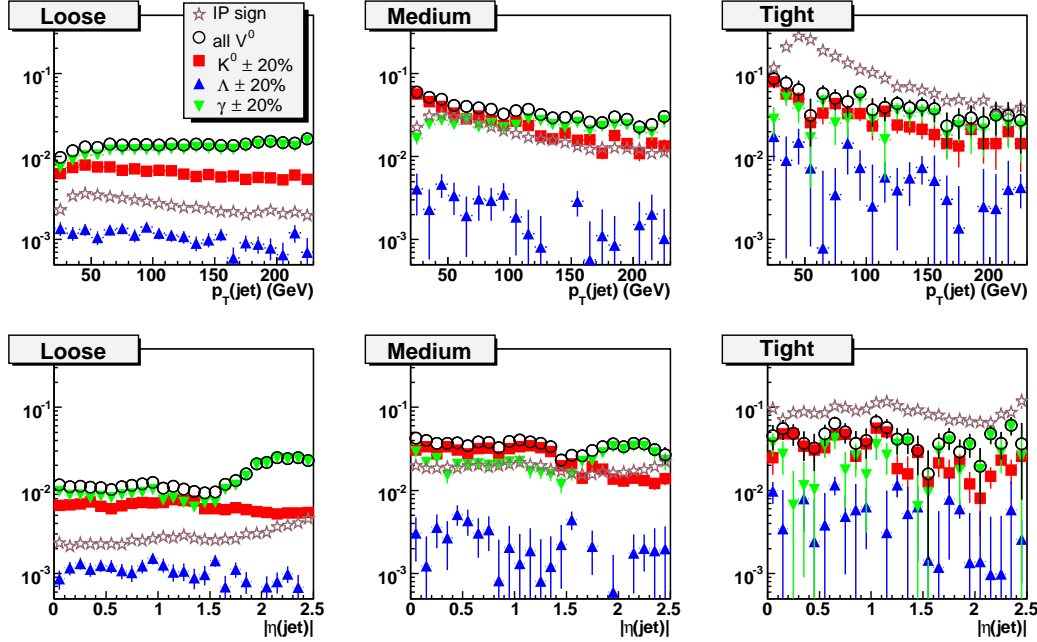


FIG. 4.41: Pour chaque point de fonctionnement des algorithmes de comptage de traces, les erreurs relatives provenant des V_0 et de l'inversion du signe en fonction du p_T (en haut) et de $|\eta|$ (en bas) des jets. Les erreurs correspondent à une luminosité de 100 pb^{-1}

dans les lots d'événements $CMSSW_1_3_1$, des coupures de sélection de type HLT ont été appliquées aux échantillons Monte-Carlo QCD .

Quatre sélections d'événements équivalentes à quatre chemins de déclenchement différents ont été considérées ici :

1. ≥ 1 jet de $p_T > 200 \text{ GeV}$,
2. ≥ 2 jets de $p_T > 150 \text{ GeV}$,
3. ≥ 3 jets de $p_T > 85 \text{ GeV}$,
4. et ≥ 4 jets de $p_T > 60 \text{ GeV}$.

Les échantillons sont comparés en étudiant les rapports entre le taux de mauvaise identification, l'efficacité d'étiquetage négatif et le facteur R_{light} des échantillons 2, 3 et 4 par rapport à l'échantillon 1 (présenté sur la figure 4.42 pour le point de fonctionnement *medium*). Ces rapports ne semblent pas dépendre de l'impulsion transverse p_T , leur moyenne est donc calculée sur tout le spectre en impulsion transverse. Les plus grandes déviations de $\varepsilon^{MC}(udsg)$ par rapport à 1 sont $(2.2 \pm 0.3)\%$, $(3 \pm 1)\%$ et $(4 \pm 2)\%$ pour les points de fonctionnement *loose*, *medium*, *tight* respectivement. Pour le taux d'étiquetage négatif, les plus grandes déviations sont de $(1.9 \pm 0.4)\%$, $(5 \pm 2)\%$ et $(13 \pm 4)\%$. Pour le facteur R_{light} , il n'y a pas de déviation observée (avec la précision relative de $\pm 1\%$, $\pm 2\%$, $\pm 5\%$). Les incertitudes systématiques liées au choix de l'échantillon QCD sont ainsi estimées à $\pm 2\%$, $\pm 5\%$,

$\pm 13\%$ environ pour les points de fonctionnement *loose*, *medium* et *tight* respectivement.

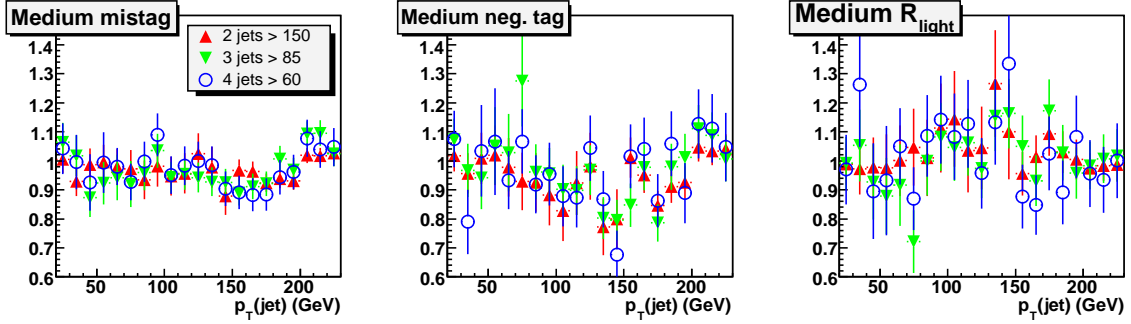


FIG. 4.42: Pour le point de fonctionnement *medium*, rapport des efficacités $\varepsilon(udsq)$ (à gauche), $\varepsilon_{neg}(udsq)$ (au milieu) et facteur R_{light} (à droite) en fonction du p_T des jets, en comparant différents échantillons correspondant à l'émulation de différents chemins de déclenchement : 1. ≥ 1 jet de $p_T > 200$ GeV, 2. ≥ 2 jets de $p_T > 150$ GeV, 3. ≥ 3 jets de $p_T > 85$ GeV et 4. ≥ 4 jets de $p_T > 60$ GeV. Les rapports sont obtenus en divisant $\varepsilon(udsq)$, $\varepsilon_{neg}(udsq)$ et R_{light} des échantillons 2, 3 et 4 par celles de l'échantillon 1.

Les formules utilisés pour le calcul de l'erreur totale sont présentées dans l'appendice A.1. Les erreurs systématiques sont résumées dans le tableau 4.4 pour les jets de $p_T \approx 100$ GeV/c en intégrant sur tout le spectre en η . Pour les points de fonctionnement *loose*, *medium* et *tight*, la source d'erreur systématique principale provient du système de déclenchement. Pour le point de fonctionnement *tight*, dans le cas des faibles luminosités ($< 10 \text{ pb}^{-1}$), la source d'erreur principale provient de l'effet de l'inversion de signe.

Plus généralement, on remarque que l'erreur systématique globale augmente pour les points de fonctionnement de plus en plus serrés. Cela est principalement dû à une diminution du nombre de jets étiquetés et donc à une moins bonne connaissance des différentes sources d'erreur.

4.4.4 Paramétrisation de la taggabilité et de l'efficacité d'identification

Il sera discuté ici de la manière dont l'efficacité d'étiquetage négatif et la taggabilité vont être paramétrisées pour être utilisées dans les analyses de physique, plus particulièrement comment les événements Monte-Carlo vont être pondérés pour être en meilleur accord avec les données.

En raison de la spécificité cinématique des événements étudiés et de leurs sélections, la taggabilité doit être calculée pour chaque analyse de physique en utilisant des événements dédiés. La fonction de taggabilité doit être paramétrisée en fonction du p_T et du

Point de fonctionnement Luminosité (pb ⁻¹)	<i>Loose</i>			<i>Medium</i>			<i>Tight</i>		
	10	100	1000	10	100	1000	10	100	1000
Systematiques (%)									
fraction de <i>b</i>	1.4	1.4	0.6	0.8	0.8	0.3	1.2	1.2	0.5
fraction de <i>c</i>	0.8	0.8	0.3	0.7	0.7	0.3	1.3	1.3	0.5
fraction de <i>g</i>	0.8	0.8	0.4	1.4	1.4	0.7	2.3	2.3	1.2
fraction de <i>V</i> ⁰	1.4	1.4	0.7	3.6	3.6	1.8	4.6	4.6	2.3
autres processus déplacé	1.4	1.4	0.7	3.6	3.6	1.8	4.6	4.6	2.3
Inversion de signe	0.7	0.3	0.2	4.5	1.9	1.4	24.0	10.2	7.6
statistique MC	0.1	0.1	0.1	0.4	0.4	0.4	1.2	1.2	1.2
statistique data	0.4	0.1	—	1.6	0.5	0.2	5.5	1.7	0.6
trigger	2.0	2.0	2.0	5.0	5.0	5.0	13.0	13.0	13.0
Total syst.	3.4	3.4	2.4	8.8	7.6	5.9	28.7	18.1	15.5

TAB. 4.4: Erreurs systématiques relatives sur l'estimation du taux de mauvaise identification de l'algorithme de comptage de traces pour des jets avec $p_T \approx 100$ GeV/c, pour les trois points de fonctionnement et pour différentes luminosités.

η des jets, mais aussi éventuellement en fonction de la multiplicité de jets, de la distance minimale ΔR_{min} entre les jets, de l'intersection en z des jets le long de l'axe du faisceau etc... Un facteur d'échelle $SF_{taggability}$ peut également être utilisé. Il est défini comme le rapport entre la taggabilité dans les données et dans la simulation Monte-Carlo. La taggabilité, ou le facteur d'échelle correspondant, peut être factorisée en fonction du p_T et du η des jets pour une saveur i donnée :

$$Taggabilit(p_T, \eta, i) = Taggabilit(p_T, i) \cdot Taggabilit(\eta, i). \quad (4.19)$$

Une méthode similaire peut être appliquée à la mesure d'efficacité d'identification des jets b ainsi que le taux de mauvaise identification. Pour chaque saveur de jets i , une fonction de taux d'étiquetage TRF^9 peut être introduite et paramétrisée en fonction de l'impulsion transverse p_T et de la pseudo-rapidité η des jets :

$$TR(p_T, \eta, i) = TRF(p_T) \cdot TRF(\eta, i). \quad (4.20)$$

L'hypothèse de factorisation peut être testée avec des données réelles ou simulées et des systématiques supplémentaires peuvent être introduites. Dans le cas où le nombre de jets est suffisant, il est également possible de définir des TRF à deux dimensions (p_T , η). L'erreur systématique liée à l'hypothèse de factorisation sera négligée ici.

Dans le cas de l'estimation du taux de mauvaise identification, la TRF pour les jets légers est calculée à partir de l'équation (4.18) à l'aide d'événements QCD uniquement

⁹Tag Rate Function

produits par simulation Monte-Carlo et un chemin de déclenchement multi-jets :

$$TRF_{light}(p_T, \eta) = \varepsilon_{data}^{mistag}(p_T, \eta) \quad (4.21)$$

$$= \varepsilon_{data}^-(p_T, \eta) \cdot R_{light}(p_T, \eta) \quad (4.22)$$

$$= \frac{\varepsilon_{data}^-(p_T, \eta)}{\varepsilon_{MC}^-(p_T, \eta)} \varepsilon_{MC}^{mistag}(p_T, \eta). \quad (4.23)$$

Dans la pratique, au lieu d'appliquer les algorithmes d'identification dans la simulation Monte-Carlo, chaque jet reconstruit dans le Monte-Carlo peut être pondéré par la taggabilité dans un premier temps, puis par la TRF afin de simuler l'application des coupures d'identification de jets b . Dans notre cas, la TRF calculée utilise l'efficacité d'étiquetage négatif mesurée dans les données en utilisant un trigger multi-jets. Elle est donc destinée à être utilisée dans les données QCD multi-jets. Pour appliquer ces TRF à d'autres échantillons, comme les événements $t\bar{t}$ ou W , les différences avec les événements QCD conduisent à des erreurs systématiques supplémentaires. Pour illustrer cette dépendance dans les échantillons utilisés, la fraction de jets de quarks légers uds en fonction du p_T est illustrée sur la figure 4.43 (en haut à gauche). Pour des jets de $p_T > 50$ GeV/c, la fraction de jets de quarks légers est environ deux fois plus importante dans les événements $t\bar{t}$ que dans les événements QCD et 50% plus importante que dans les événements W . Les autres graphiques de la figure 4.43 présentent le rapport entre l'efficacité $\varepsilon(udsg)$ dans les événements $t\bar{t}$ (ou W +jets) et l'efficacité $\varepsilon(udsg)$ dans les événements QCD . Il faudra donc considérer les fonction TRF_{uds} et TRF_g séparément.

Au lieu d'utiliser ces TRF , il est possible d'utiliser un facteur d'échelle SF entre les données multi-jets et le Monte-Carlo QCD . Le facteur d'échelle SF est défini par :

$$SF_{light}(p_T, \eta) = \varepsilon_{udsg}^{data}(p_T, \eta) / \varepsilon_{udsg}^{MC}(p_T, \eta) \quad (4.24)$$

$$= \varepsilon_{neg}^{data}(p_T, \eta) / \varepsilon_{neg}^{MC}(p_T, \eta), \quad (4.25)$$

qui peut être également factorisé en p_T et η des jets. Dans ce cas, l'efficacité d'étiquetage négatif des jets légers est déterminée en appliquant l'algorithme d'étiquetage négatif directement sur le lot d'événements Monte-Carlo analysé.

4.5 Conclusion et perspective

Au cours de cette thèse, les méthodes d'identification des jets b ont été étudiées et plus particulièrement les calibrations de l'algorithme de probabilité par jet. Un programme de calibration pseudo-automatisé a été implémenté dans le code de CMS et des outils d'étude de ces calibrations ont été mis en place. Le but étant de préparer la prise de données au démarrage du LHC.

Nous nous sommes également intéressés aux méthodes d'estimation des efficacités d'identification et du taux de mauvaise identification à l'aide des données. La méthode

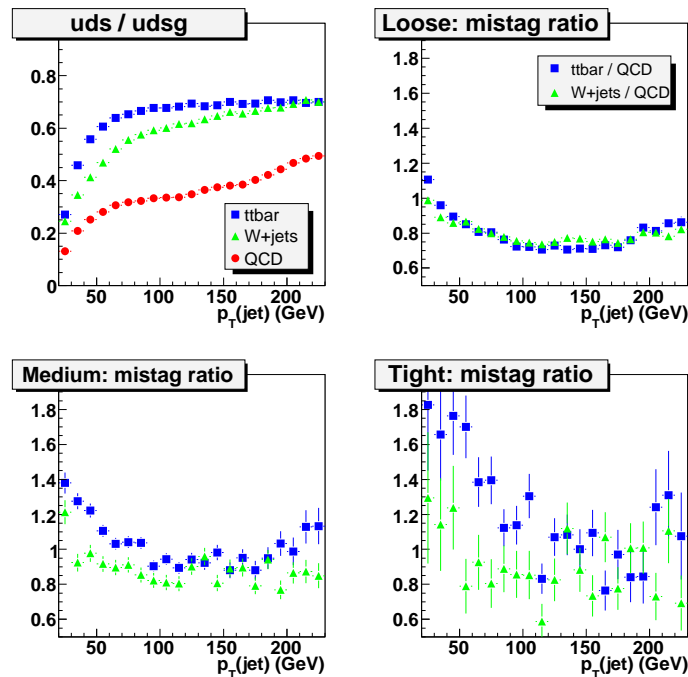


FIG. 4.43: En haut à gauche : fraction des jets uds par rapport à tous les jets $udsg$ pour des événements QCD , W +jets et $t\bar{t}$ semi-leptonique. Les trois autres graphiques représentent le rapport entre les efficacités $\varepsilon(uds)$ dans les échantillons W +jets et $t\bar{t}$ semi-leptonique et l'efficacité $\varepsilon(uds)$ dans l'échantillon QCD pour les trois points de fonctionnement : *loose* en haut à droite, *medium* en bas à gauche et *tight* en bas à droite.

System8 a été introduite dans CMS et des premiers résultats encourageant ont été obtenus. Des problèmes de convergence à haut p_T sont cependant survenus et sont en cours d'étude.

Un effort tout particulier a été porté sur la méthode d'estimation du taux de mauvaise identification. Cette méthode a été testée sur des données plus récentes et les résultats correspondant sont présentés dans l'annexe A.3. En particulier, des algorithmes d'étiquetage positif et négatif ont été développés pour la majorité des algorithmes d'identification. Cela nous permettra d'estimer le taux de mauvaise identification pour les algorithmes de comptage de traces, de probabilité par jet, de vertex secondaire, de vertex secondaire combiné et de soft muon. Il reste cependant un certain nombre de points à explorer lorsque les données réelles seront disponibles :

- il faudra mettre en place des méthode de comparaison entre les échantillons de données et les échantillons de Monte-Carlo. Il sera également important d'étudier différents générateurs.
- La dépendance de l'estimation sur le chemin de déclenchement utilisé devra être étudiée.
- Le taux de mauvaise identification sera estimé à partir de données multi-jets et du Monte-Carlo QCD , il faudra donc extrapoler ces résultats à d'autres types d'événements comme les événements W +jets ou $t\bar{t}$,
- Finalement, la paramétrisation du taux de mauvaise identification et son utilisation dans les analyses de physique doivent être précisées.

Pour conclure, la définition des algorithmes d'identification mais également les méthodes de mesure de performances permettront, à moyen et long terme, d'utiliser l'identification de jets b dans les analyses de physique telles que : les mesures de section efficace de production de quark top et de sa masse, l'étude du vertex V_{tb} , la recherche du boson de Higgs ou encore la recherche de nouvelle physique.

5

Mesure de la section efficace de production d'événements $t\bar{t}$

Au cours de cette thèse, une stratégie de sélection des événements $t\bar{t}$ a été mise en place, afin de mesurer la section efficace de production des canaux **di-leptons** (sans inclure les canaux contenant des τ se désintégrant hadroniquement [100, 101]) au LHC pour une luminosité intégrée de 100 pb^{-1} .

La sensibilité de *CMS* pour la mesure de la section efficace de production d'événements $t\bar{t}$ a déjà été étudiée pour des luminosités de 1 fb^{-1} et au-delà [102, 103, 104, 105, 106, 107, 108]. De nombreuses analyses de *CMS* sont dédiées à la physique du top comme les mesures de corrélation de spin [102], la mesure de la masse du top [109] ou de la masse invariante du système $t\bar{t}$ [110]. Comme les top seront produits en grande quantité au LHC, les événements top pourront également être utilisés comme un outil de calibrations en énergie des jets [34] ou de mesure d'efficacité des algorithmes d'identification des jets b [111].

Le but de l'analyse qui sera présentée dans ce chapitre est cependant différent : les performances du détecteur au démarrage du LHC doivent être considérées avec une grande précaution. L'analyse inclut donc des effets de désalignement et de mauvaise calibration à l'aide d'événements simulés correspondant à un détecteur imparfait [81]. L'un des objectifs majeurs de cette analyse est ainsi de définir une sélection simple et robuste. Le système de déclenchement (ou trigger) a également été introduit dans l'analyse, les différents chemins de déclenchement sont présentés dans la référence [96] pour une luminosité instantanée de $10^{32} \text{ cm}^{-2}\text{s}^{-1}$.

Un effort tout particulier a été porté sur le développement de méthodes d'estimation de bruits de fond à l'aide des données réelles afin de s'affranchir au maximum des imperfections du Monte-Carlo ou de la mauvaise description du détecteur dans la simulation.

Dans un premier temps, la spécificité de l'analyse ainsi que la stratégie de sélection seront détaillées. Nous étudierons plusieurs scénarios de sélection, correspondant à différents critères de sélection des jets b . Différentes méthodes d'estimation des bruits de fond à l'aide des données seront ensuite présentées. La stratégie mise en place pour la mesure de la section efficace et l'estimation des erreurs systématiques correspondantes seront finalement discutées.

5.1 Caractéristiques des événements

5.1.1 Événements $t\bar{t}$ di-leptons

Comme nous l'avons vu dans le chapitre 1.5, les canaux di-leptons correspondent au cas où chaque boson W issu de la désintégration de quark top se désintègre leptoniquement. Nous nous sommes intéressés uniquement aux canaux dont les produits de cette désintégration contiennent un électron ou un muon. Ceux-ci qui peuvent également être produits par les cascades de désintégration $W \rightarrow \tau\bar{\nu}_\tau \rightarrow \bar{\nu}_\tau\nu_\tau e\bar{\nu}_e$ et $\bar{\nu}_\tau\nu_\tau\mu\bar{\nu}_\mu$. Les canaux constituant le signal sont donc les canaux di-électrons ($e - e$), di-muons ($\mu - \mu$) et électron-muon ($e - \mu$).

Les événements de signal sont tout d'abord caractérisés par la présence de deux leptons (électron ou muon) de charges opposées, de grande impulsion transverse produits plutôt dans la partie centrale du détecteur avec des p_T de 55 GeV/c en moyenne. L'impulsion transverse des leptons générés et leur pseudo-rapacité sont présentées sur la figure 5.1.

Deux jets de grande impulsion transverse (80 GeV/c en moyenne), sont également produits préférentiellement dans la partie centrale. Ils proviennent de l'hadronisation de quarks b issus de la désintégration de quarks top. La figure 5.2 (en haut) présente les distributions en impulsion transverse et en pseudo-rapacité des jets générés issus de l'hadronisation de hadrons beaux. Les jets générés sont reconstruits à l'aide de l'algorithme du cône itératif à partir des particules générées. Des jets additionnels de bas p_T peuvent également être produits par la radiation des partons initiaux (avant l'interaction dure) et finaux (après le processus dur). Leurs impulsions transverses et leurs pseudo-rapidités sont présentées sur la figure 5.2.

Une grande énergie transverse manquante (\cancel{E}_T) provient des neutrinos produits par la désintégration leptonique de bosons W . La distributions en \cancel{E}_T générée (reconstruite à l'aide des particules générées sans prendre en compte les neutrinos) est présentée sur la figure 5.3.

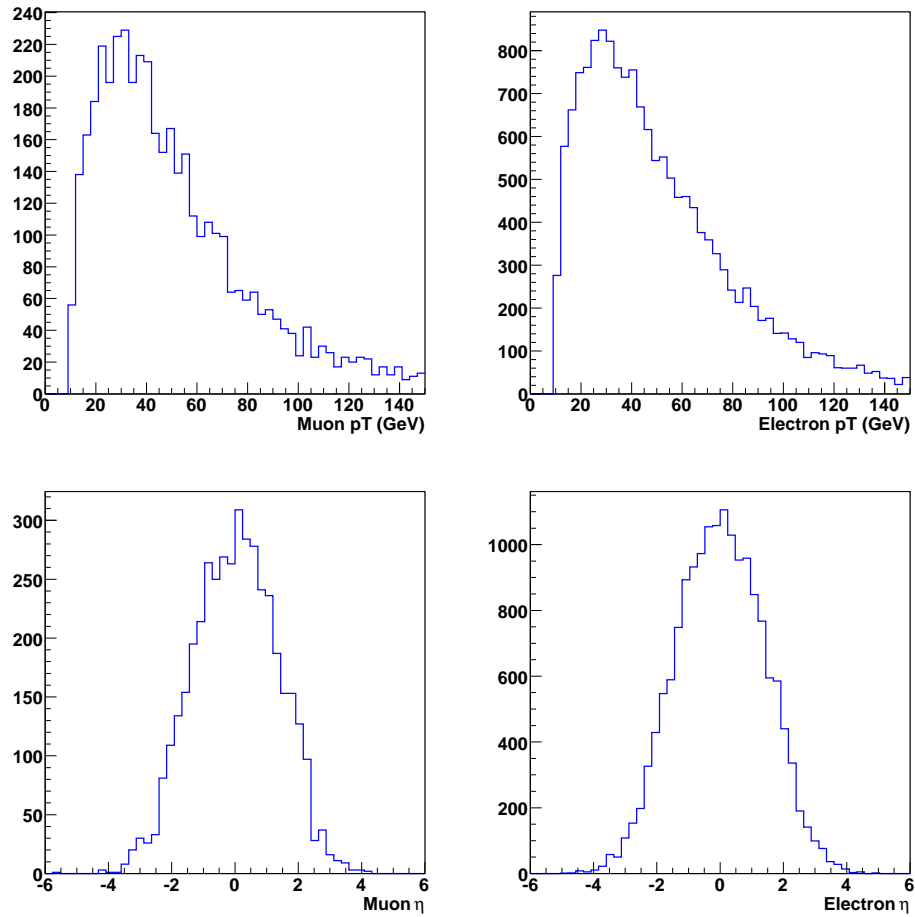


FIG. 5.1: Impulsion transverse (en haut) et pseudo-rapidity (en bas) des leptons générés issus de la désintégration des bosons W dans des événements ALPGEN $t\bar{t}+0$ jet di-leptons pour les muons (à gauche) et les électrons (à droite).

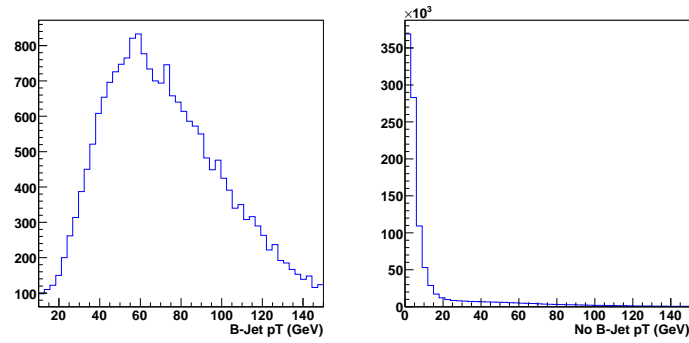


FIG. 5.2: Impulsion transverse (à gauche) des jets générés issus de la désintégration des hadrons beaux (à gauche) et les autres jets générés (droite) dans des événements ALPGEN $t\bar{t}+0$ jet di-leptons.

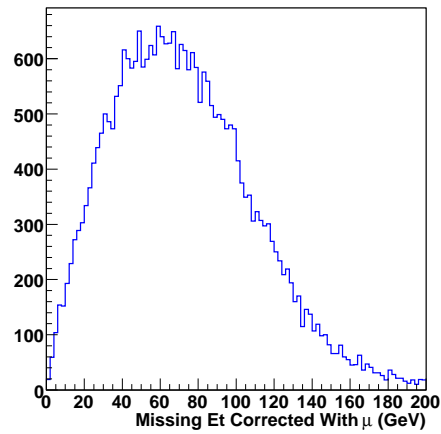


FIG. 5.3: Distribution de la \cancel{E}_T générée pour les événements ALPGEN $t\bar{t}+0$ jet di-leptons.

5.1.2 Bruits de fond

Il y a principalement trois types de bruits de fond liés à la nature des leptons sélectionnés et à leur isolation. Ces trois types de bruits de fond sont classés par multiplicité de leptons issus de la désintégration de bosons Z ou W :

1. **Les bruits de fond physique.** Les événements contenant au moins deux vrais leptons isolés dans l'état final. Il s'agit, par ordre d'importance après sélection, des événements Drell-Yan γ^*/Z , des événements single top et des événements di-bosons WW , ZZ et WZ (cf. section 1.4), avec $W \rightarrow l\nu_l$ et $Z \rightarrow l^+l^-$.

Les événements Drell-Yan peuvent être produits en association avec des partons. De part leurs grandes sections efficaces de production, de l'ordre de 5000 pb au NLO, ils constituent le principal bruit de fond. Le diagramme de Feynman principal de production correspondant est présenté par les diagrammes de droite de la figure 1.3. Ils peuvent être rejetés à l'aide d'une coupure sur la masse invariante du système di-leptons, sur l'énergie transverse manquante ou en demandant des jets identifiés comme jets b .

Les diagrammes de Feynman principaux de production de single top sont illustrés sur la figure 1.11. Avec une section efficace de production de 62 pb, le canal tW contribue le plus au bruit de fond total s'il on applique des coupures d'identification des jets b . Ces événements ne peuvent être rejetés efficacement qu'en demandant la présence de deux jets identifiés comme jets b .

Les événements di-bosons constituent également un bruit de fond important. Plus particulièrement, les événements WW , de par leur plus grande section efficace que les événements WZ et ZZ , peuvent survivre aux coupures de sélection lorsqu'ils contiennent aux moins deux jets issus de la radiation des partons initiaux et finaux ou de la désintégration de W ou de Z . L'identification de jets b permet de les rejeter car ces événements ne contiennent que peu de jets b .

2. **Les bruits de fond semi-instrumentaux.** Une autre catégorie d'événements de bruit de fond est constituée des événements contenant un vrai lepton isolé provenant de la désintégration d'un boson W et d'un faux lepton reconstruit ou d'un lepton non-isolé qui passe les critères d'isolation. Ils peuvent provenir de jets partoniques produits en association avec le boson W ou de jets issus de la radiation des partons initiaux ou finaux. Il s'agit principalement des événements $W + jets$ et $t\bar{t}$ semi-leptoniques. Pour ce dernier type d'événements, les jets de quarks légers provenant d'un boson W hadronique peuvent contribuer. Le mode de production principal de bosons W est présenté par le diagramme de gauche de la figure 1.3.
3. **Les bruits de fond instrumentaux.** La dernière catégorie d'événements de bruit de fond est constituée des événements contenant au moins deux faux leptons ou leptons non-isolés qui passent les critères d'isolation. Il s'agit des événements $t\bar{t}$ hadronique et QCD multi-jets.

Les faux électrons isolés proviennent principalement de jets de quark léger associés à une seule trace dont les particules qui les constituent (comme les π^0) déposent une grande fraction de leur énergie dans le calorimètre électromagnétique. De faux électrons peuvent également être produits par des τ se désintégrant hadroniquement ou des photons associés à une particule chargée ou à une fausse trace. Dans les événements contenant un Z , un muon peut être reconstruit comme électron lorsqu'il émet un photon par conversion interne. Une plus faible fraction de ces faux muons peut être produite par la désintégration de pions ou des gerbes hadroniques qui atteignent les chambres à muons.

Les vrais leptons non-isolés qui apparaissent isolés dans le détecteur proviennent majoritairement de hadrons lourds (hadrons b et c) se désintégrant leptoniquement.

Des problèmes de reconstruction de \cancel{E}_T peuvent également augmenter le nombre d'événements de bruit de fond sélectionnés. La \cancel{E}_T peut être élevée, même dans le cas des événements QCD , Z +jets et ZZ . Ceci est dû à des effets de perte d'énergie dans le détecteur, de mauvaise calibration, de résolution en énergie des jets et des leptons, ou encore de mauvaise reconstruction.

A cause de leur grande section efficace de production, les événements QCD multi-jets peuvent constituer un bruit de fond important, même si l'efficacité de sélection des leptons est très faible dans ces événements. Une étude dédiée au bruit de fond QCD sera présentée dans la section 5.3.

Nous appellerons bruit de fond physique les événements contenant au moins deux leptons de charges opposées (Drell Yan, single top tW , WW , ZZ , WZ) et bruits de fond instrumentaux les événements contenant au moins un faux lepton ou un lepton non-isolé passant les coupures d'isolation (par la suite nous appellerons ces deux catégories de lepton : faux lepton isolé). Le bruit de fond single top produit dans la voie s est négligé, compte tenu de sa faible section efficace (12 pb au NLO). Les quarks single top produits dans la voie t ont une pseudo-rapacité élevée et devraient donc être reconstruits avec une faible efficacité, ils seront également négligés. Faute d'échantillon adéquat, les voies s et t n'ont pu être étudiées. Des résultats récents, effectués dans le cadre de collision à 10 TeV, ont cependant démontré que ces événements sont effectivement négligeables comparés à la voie tW .

Dans les sections 5.4 et 5.5, différentes méthodes d'estimation de bruits de fond à partir des données seront présentées. Les bruits de fond single top et di-bosons seront estimés à partir du Monte-Carlo uniquement.

5.1.3 Stratégie de sélection

Les canaux de désintégration di-leptons ont l'avantage d'avoir une signature claire permettant l'utilisation de simples coupures séquentielles. Compte tenu des caractéristiques du signal et des bruits de fond, la sélection des événements commence par la reconstruction, l'identification et l'isolation des paires de candidats leptons. Une présélection

appliquée sur les événements possédant deux candidats leptons de charges électriques opposées.

Afin de limiter la contamination d'événements de bruits de fond contenant les leptons de bas p_T issus de la désintégration faible de hadrons beaux ou charmés, ou issus de faux leptons, une coupure additionnelle sur le p_T des leptons est réalisée. Une première sélection sur les jets est appliquée afin de rejeter les événements W , Z et di-bosons en demandant au moins deux jets avec des impulsions transverses supérieures à un certain seuil.

Les événements Z +jets sont principalement rejetés par l'application d'une sélection sur la masse invariante des deux leptons.

Un grande partie des événements QCD multi-jets et des événements Z +jets restants est rejetée par l'application d'une coupure en énergie transverse manquante. La sélection est finalisée par l'utilisation d'un algorithme d'identification de jets b : un ou deux jets doivent être identifiés par l'algorithme de comptage de traces. Le point fonctionnement *loose* (cf. section 4.3) afin de conserver une plus grande partie des événements de signal (les efficacités d'identification pour les différents points de fonctionnement sont mentionnées dans le tableau 4.1) tout en gardant un bon contrôle des erreurs systématiques associées à l'identification des jets b . L'identification d'un deuxième jet b permet notamment la rejection des événements single top tW .

Différents scénarios de sélection ont été testés, chacun d'eux correspondant à l'identification d'au moins 0, 1 ou 2 jets b . L'utilisation d'un algorithme d'identification de jet b permet de réduire les incertitudes dues au nombre d'événements de bruits de fond sélectionnés, qui deviennent très faibles, mais augmente les incertitudes systématiques et statistiques. Un compromis doit être trouvé entre identification des jets b et contamination du bruit de fond, afin de minimiser l'erreur totale.

5.2 Mesure de la section efficace

5.2.1 Les données utilisées

Les données Monte-Carlo utilisées dans cette analyse ont été produites dans le cadre de l'exercice CSA07 [112, 113, 114] : plus de 100 millions d'événements ont été générés, simulés et reconstruits pour des collisions de 14 TeV dans le centre de masse suivant le schéma présenté dans la section 1.6. Ces événements ont ensuite été assemblés en échantillons appelés "soupe", qui contiennent un mélange de différents types d'événements pondérés et simulent un lot de vraies données.

5.2.1.1 La production CSA07

Le signal et la plupart des bruits de fond utilisés proviennent de la soupe appelée "Chowder". Cet échantillon est constitué d'événements ($t\bar{t}$ présentés dans le tableau 5.1), d'événements Z +jets (tableau 5.2) et d'événements W +jets (tableau 5.3). La soupe

“Gumbo” est quant à elle constituée principalement d'événements QCD multi-jets générés avec PYTHIA. Le générateur TOPREX est utilisé pour générer des événements single top dans le canal tW et des événements di-bosons (WW , WZ et ZZ).

De manière à n'utiliser qu'un nombre modéré d'événements et afin de tenir compte des efficacités du système de déclenchement, une première présélection est réalisée : un événement doit passer les sélections du système de déclenchement et posséder au moins deux leptons reconstruits avec des impulsions transverses supérieures à 20 GeV/c. Cette présélection est appelée “skimming”.

Afin de décrire de manière réaliste les conditions de démarrage de l'expérience, des scénarios de désalignement et de décalibration ont été définis de manière à reproduire les performances du détecteur attendues pour différentes luminosités intégrées, comme il a été discuté dans la section 4.1.1.1. Pour une analyse utilisant une luminosité intégrée de 100 pb^{-1} , le détecteur est supposé être aligné et calibré avec 10 pb^{-1} de données. C'est donc ce scénario (10 pb^{-1}) qui sera utilisé dans l'analyse qui va être présentée.

5.2.1.2 Simulation des événements

Les événements $t\bar{t}+N$ jets ont été produits avec le générateur ALPGEN [46], où N correspond au nombre de partons produits en association avec les paires de quarks $t\bar{t}$, exclusivement pour N allant de 0 à 3 et inclusivement pour $N > 3$. La liste des échantillons, leurs sections efficaces ainsi que les nombres d'événements générés sont présentés dans la table 5.1. Bien que le générateur ALPGEN permette de calculer les sections efficaces jusqu'au LO, les événements sont pondérés par la section efficace calculée analytiquement au NLO (+ correction NLL). La section efficace de production de paires de quark top est de 447 pb au LO et de 833 pb au NLO [31]. En première approximation, nous supposons que les différences entre variables cinématiques et topologiques au LO et au NLO sont faibles et que les fractions d'événements produits par multiplicité de partons associés restent sensiblement les mêmes.

De la même manière, les sections efficaces au NLO des événements de bruits de fond sont renormalisées au NLO. Les échantillons de données simulées utilisés dans l'analyse sont maintenant présentés.

- Le bruit de fond dominant est la production d'événements Z/γ^* (Drell-Yan, appelé Z +jets dans la suite du texte). Il est simulé avec ALPGEN, qui inclut la production de boson vecteur en association avec N quarks, exclusivement pour N allant de 0 à 4 et inclusivement pour $N > 4$. Les détails des échantillons utilisés sont présentés dans la table 5.2.
- La production de bosons W associés à N quarks, exclusivement pour N allant de 0 à 4 et inclusivement pour $N > 4$, est réalisée à l'aide du générateur ALPGEN. Les détails de ces échantillons sont présentés dans la table 5.3. Les taux de productions d'événements W et \bar{W} seront mesurés indépendamment dans les données [115].

Processus	échantillons	σ LO [pb]	σ NLO [pb]	Nbr d'év.
$t\bar{t} + 0$ jets	/tt0j_mT_70-alpgen/CMSSW_1_5_2-CSA07-2231/GEN-SIM-DIGI-RECO	334.51	622.19	1.2M
$t\bar{t} + 1$ jets	/tt1j_mT_70-alpgen/CMSSW_1_5_2-CSA07-2211/GEN-SIM-DIGI-RECO	90.23	167.8	350k
$t\bar{t} + 2$ jets	/tt2j_mT_70-alpgen/CMSSW_1_5_2-CSA07-2232/GEN-SIM-DIGI-RECO	18.75	34.85	70k
$t\bar{t} + 3$ jets	/tt3j_mT_70-alpgen/CMSSW_1_5_2-CSA07-2212/GEN-SIM-DIGI-RECO	3.16	5.88	12k
$t\bar{t} + 4$ jets	/tt4j_mT_70-alpgen/CMSSW_1_5_2-CSA07-2233/GEN-SIM-DIGI-RECO	0.82	1.52	3k

TAB. 5.1: Echantillons Monte-CarloALPGEN utilisés pour les événements $t\bar{t}+N$ jets. La première colonne correspond au processus (multiplicité de partons associées), la deuxième au nom de l'échantillon, les troisième et quatrième aux sections efficaces LO et NLO et la dernière au nombre d'événements générés.

Processus	échantillons	σ LO [pb]	σ NLO [pb]	Nbr d'év.
$Z + 0$ jets	/Z0jet-alpgen/CMSSW_1_5_2-CSA07-1193756147/GEN-SIM-DIGI-RECO	4400	5280	4.5M
$Z + 1$ jets	/Z1jet_Optw100-alpgen/CMSSW_1_5_2-CSA07-2207/GEN-SIM-DIGI-RECO	935	1122	935k
$Z + 1$ jets	/Z1jet_100ptw300-alpgen/CMSSW_1_5_2-CSA07-2227/GEN-SIM-DIGI-RECO	30	36	30k
$Z + 2$ jets	/Z2jet_Optw100-alpgen/CMSSW_1_5_2-CSA07-2208/GEN-SIM-DIGI-RECO	271	325	275k
$Z + 2$ jets	/Z2jet_100ptw300-alpgen/CMSSW_1_5_2-CSA07-2190/GEN-SIM-DIGI-RECO	28	34	30k
$Z + 3$ jets	/Z3jet_Optw100-alpgen/CMSSW_1_5_2-CSA07-2209/GEN-SIM-DIGI-RECO	68	82	70k
$Z + 3$ jets	/Z3jet_100ptw300-alpgen/CMSSW_1_5_2-CSA07-2210/GEN-SIM-DIGI-RECO	13	16	15k
$Z + 4$ jets	/Z4jet_Optw100-alpgen/CMSSW_1_5_2-CSA07-2234/GEN-SIM-DIGI-RECO	14	17	15k
$Z + 4$ jets	/Z4jet_100ptw300-alpgen/CMSSW_1_5_2-CSA07-2228/GEN-SIM-DIGI-RECO	4	5	5k
$Z + 5$ jets	/Z5jet_Optw100-alpgen/CMSSW_1_5_2-CSA07-2229/GEN-SIM-DIGI-RECO	9	11	10k
$Z + 5$ jets	/Z5jet_100ptw300-alpgen/CMSSW_1_5_2-CSA07-2230/GEN-SIM-DIGI-RECO	5	6	5k

TAB. 5.2: Similaire au tableau 5.1 pour des événements Z +jets ALPGEN.

Processus	échantillons	σ LO [pb]	σ NLO [pb]	Nbr d'év.
$W + 0$ jets	/W0jet-alpgen/CMSSW_1_5_2-CSA07-2203/GEN-SIM-DIGI-RECO	45000	54000	9M
$W + 1$ jets	/W1jet_Optw100-alpgen/CMSSW_1_5_2-CSA07-2219/GEN-SIM-DIGI-RECO	9200	11040	9.25M
$W + 1$ jets	/W1jet_100ptw300-alpgen/CMSSW_1_5_2-CSA07-2220/GEN-SIM-DIGI-RECO	250	300	250k
$W + 2$ jets	/W2jet_Optw100-alpgen/CMSSW_1_5_2-CSA07-2221/GEN-SIM-DIGI-RECO	2500	3000	2.5M
$W + 2$ jets	/W2jet_100ptw300-alpgen/CMSSW_1_5_2-CSA07-2222/GEN-SIM-DIGI-RECO	225	270	225k
$W + 3$ jets	/W3jet_Optw100-alpgen/CMSSW_1_5_2-CSA07-2223/GEN-SIM-DIGI-RECO	590	708	590k
$W + 3$ jets	/W3jet_100ptw300-alpgen/CMSSW_1_5_2-CSA07-2188/GEN-SIM-DIGI-RECO	100	120	100k
$W + 4$ jets	/W4jet_Optw100-alpgen/CMSSW_1_5_2-CSA07-2206/GEN-SIM-DIGI-RECO	125	150	125k
$W + 4$ jets	/W4jet_100ptw300-alpgen/CMSSW_1_5_2-CSA07-2189/GEN-SIM-DIGI-RECO	40	48	40k
$W + 5$ jets	/W5jet_Optw100-alpgen/CMSSW_1_5_2-CSA07-2224/GEN-SIM-DIGI-RECO	85	102	85k
$W + 5$ jets	/W5jet_100ptw300-alpgen/CMSSW_1_5_2-CSA07-2225/GEN-SIM-DIGI-RECO	40	48	40k

TAB. 5.3: Similaire au tableau 5.1 pour des événements W +jets ALPGEN.

Processus	échantillon	σ LO [pb]	σ NLO [pb]	Nbr d'ev.
WZ leptonique	/tW_inclusive/CMSSW_1_6_7-CSA07-1198096703/AODSIM	58.8	64.2	272791

TAB. 5.4: Similaire au tableau 5.1 pour les événements single top tW TOPREX.

Processus	échantillons	σ LO [pb]	σ NLO [pb]	Nbr d'ev.
WZ leptonique	/WZ_incl/CMSSW_1_6_7-CSA07-1195629996/RECO	30.0	49.9	363291
WW inclusive	/WW_incl/CMSSW_1_6_7-CSA07-1196178448/RECO	81.3	114.3	845261
ZZ inclusive	/ZZ_incl/CMSSW_1_6_7-CSA07-1194964234/RECO	11.8	16.1	143113

TAB. 5.5: Similaire au tableau 5.1 pour des événements di-bosons PYTHIA.

- Les événements single top tW sont produits également à l'aide du générateur TOPREX [116] dans le cas d'un détecteur parfait (tableau 5.4). Comme les bruits de fond di-bosons et single top sont des bruits de fond physique, on peut s'attendre à ce que des effets de désalignement et de décalibration réduisent les efficacités de sélection de ces événements et donc diminuent leur contamination. Le détecteur étant parfait, le bruit de fond tW sera surestimé ¹.
- Les événements di-bosons (WW , WZ et ZZ) sont produits à l'aide du générateur PYTHIA (cf. tableau 5.5). Ces événements ont été simulés dans le cas d'un détecteur parfait.
- Pour le bruit de fond QCD , le mélange d'événements Monte-Carlo "Gumbo" PYTHIA est utilisé pour tester la méthode d'estimation de bruit de fond instrumental (cf. section 5.4). Compte tenu des sections efficaces QCD très élevées et des efficacités de sélection très faibles, il est difficile de générer suffisamment d'événements pour effectuer une étude satisfaisante. Il faut ajouter à cela le fait que dans la simulation, d'une manière générale, le taux de faux leptons reconstruits est probablement sous-estimé. Afin d'obtenir une estimation plus correcte de la contamination d'événements QCD , une approche de factorisation, décrite dans la section 5.3, sera présentée à partir des lots d'événements du tableau 5.6 .

¹Des événements avec un détecteur parfait ont été utilisés pour les bruits de fond di-boson et single top tW car il s'agissait des seuls échantillons disponibles au moment où cette analyse a été faite.

Processus	Echantillons	σ [pb]	Nbr d'ev.
$15 < \hat{P}_t < 20 GeV$	/QCD_Pt_15_20/CMSSW_1_5_2-CSA07-2026/GEN-SIM-DIGI-RECO	1.4×10^9	1.24M
$20 < \hat{P}_t < 30 GeV$	/QCD_Pt_20_30/CMSSW_1_5_2-CSA07-2162/GEN-SIM-DIGI-RECO	6.32×10^8	1.38M
$30 < \hat{P}_t < 50 GeV$	/QCD_Pt_30_50/CMSSW_1_5_2-CSA07-2048/GEN-SIM-DIGI-RECO	1.63×10^8	745k
$50 < \hat{P}_t < 80 GeV$	/QCD_Pt_50_80/CMSSW_1_5_2-CSA07-2049/GEN-SIM-DIGI-RECO	2.16×10^7	580k
$80 < \hat{P}_t < 120 GeV$	/QCD_Pt_80_120/CMSSW_1_5_2-CSA07-2027/GEN-SIM-DIGI-RECO	3.08×10^6	720k
$120 < \hat{P}_t < 170 GeV$	/QCD_Pt_120_170/CMSSW_1_5_2-CSA07-2171/GEN-SIM-DIGI-RECO	4.94×10^5	940k
$170 < \hat{P}_t < 230 GeV$	/QCD_Pt_170_230/CMSSW_1_5_2-CSA07-2069/GEN-SIM-DIGI-RECO	1.01×10^5	930k
$230 < \hat{P}_t < 300 GeV$	/QCD_Pt_230_300/CMSSW_1_5_2-CSA07-2050/GEN-SIM-DIGI-RECO	2.45×10^4	725k
$300 < \hat{P}_t < 380 GeV$	/QCD_Pt_300_380/CMSSW_1_5_2-CSA07-2061/GEN-SIM-DIGI-RECO	6.24×10^3	1.09M

TAB. 5.6: Echantillons Monte-Carlo PYTHIA utilisés pour les événements QCD .

La simulation complète de CMS, basée sur Géant 4, est utilisée pour décrire la réponse du détecteur. Les différents algorithmes de reconstruction décrits dans le chapitre 2 sont appliqués à ces événements générés.

5.2.2 Chemins de déclenchement

Le déclenchement du système d'acquisition s'effectue en deux étapes : le déclenchement de niveau 1 ("Level-1 trigger") utilise l'information produite au niveau de l'électronique et le déclenchement de haut niveau "High Level Trigger" (HLT) correspond à des séquences de reconstruction d'objets de complexité croissante². Comme il est décrit dans la section 5.2.1.1, les données présélectionnées correspondent à la combinaison de triggers et de coupures lâches de présélection.

Différents chemins de déclenchement sont considérés, suivant les lots d'événement utilisés :

- pour le canal di-électrons, le chemin de déclenchement (ou *trigger*) "électron simple" correspond à la reconstruction d'un dépôt d'énergie électromagnétique reconstruit au niveau 1 suivi de l'application d'un critère d'isolation lâche sur les électrons reconstruits par les algorithmes HLT à partir de l'information provenant du calorimètre ECAL et du trajectographe. Des coupures sur le rapport entre l'énergie calorimétrique associée à l'électron et l'impulsion de la trace correspondante (E/p) sont appliquées, telles que E/p est inférieur à 1.5 dans la partie centrale du détecteur et à 2.45 dans les bouchons. Le trigger "simple électron" correspond à la reconstruction et à la sélection d'un électron HLT ayant une énergie transverse $E_T > 17$ GeV. Le trigger "double électron" est allumé quand deux électrons HLT sont reconstruits avec des énergies transverses $E_T > 12$ GeV. Cette sélection sera appelée chemin de déclenchement di-électrons.
- Pour les canaux di-muons et électron-muon, les triggers "simple muon" (un muon reconstruit au niveau 1 reconstruit par les algorithmes HLT et ayant un $p_T > 16$ GeV/c), "double muons" (2 muons HLT ayant un $p_T > 3$ GeV/c) et un trigger mixte "électron-muon" correspondant à "simple électron" ET "simple muon" (avec E_t ou p_T supérieur à 10 GeV(/c)). Cette sélection sera appelée chemin de déclenchement muon+X où X peut être un électron ou un muon.

La définition des skims comprend également des coupures additionnelles sur l'énergie transverse E_T et l'impulsion transverse p_T des leptons > 20 GeV(/c). Les efficacités du système de déclenchement ont été estimées à l'aide de la simulation Monte-Carlo. Elles sont déterminées à partir d'événements $t\bar{t}$ di-électrons et di-muons. Les efficacités sont définies par rapport aux événements dont les deux leptons issus de W sont générés avec $p_T > 10$ GeV/c et $|\eta| < 2.4$. Des coupures sur les objets reconstruits sont également appliquées : au moins 2 jets reconstruits avec $E_T > 20$ GeV et $|\eta| < 2.4$ et au moins deux leptons de charge opposées vérifiant $p_T > 15$ GeV/c et $|\eta| < 2.1$.

²HLT correspond à des triggers de niveau 2 et 3

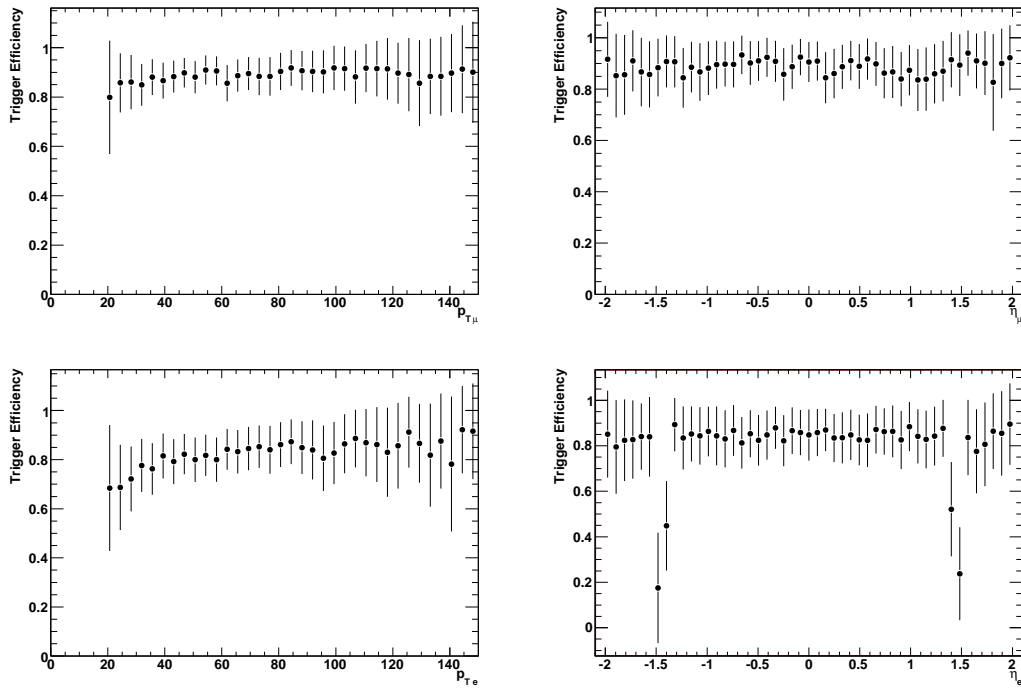


FIG. 5.4: Efficacité des chemins de déclenchements muon-X (en haut) et di-électrons (en bas) en fonction de p_T et η des leptons pour les événements reconstruits passant les sélections au niveau de la génération et de la reconstruction décrites dans le texte, pour les canaux $\mu - \mu$ en haut et $e - e$ en bas. Les erreurs présentées correspondent aux erreurs statistiques pour une luminosité intégrée de 100 pb^{-1} .

Les graphiques de la figure 5.4 [104] représentent l'évolution des efficacités de trigger en fonction de l'impulsion transverse p_T et de la pseudo-rapidité η des leptons reconstruits pour les canaux $\mu - \mu$ et $e - e$. Les erreurs statistiques correspondent à une luminosité de 100 pb^{-1} . Les baisses d'efficacité autour de $|\eta| = 1.4$ correspondent à la zone de transition entre la partie centrale et les bouchons du trajectographe et des calorimètres.

Pour le canal $\mu - \mu$, l'efficacité du chemin de déclenchement "simple muon" est de $99.3 \pm 1.0\%$. Le chemin de déclenchement "double muons" a, quant à lui, une efficacité de $89 \pm 1.5\%$. Pour le canal $e - e$, l'efficacité du chemin de déclenchement "simple électron" est de $91.4 \pm 1.5\%$ tandis que l'efficacité du chemin de déclenchement "double électrons" a une efficacité de $82.5 \pm 1.5\%$. Pour le canal $e - \mu$, les efficacités sont $89.1 \pm 1.5\%$, $71.2 \pm 1.5\%$ et $69.3 \pm 1.5\%$ pour les chemins de déclenchement "simple muon", "simple électron" et "électron-muon" respectivement.

L'efficacité de sélection des triggers sera mesurée dans les vraies données à l'aide d'un chemin de déclenchement indépendant ne contenant aucun critère de sélection sur les leptons, par exemple un trigger multi-jets. L'ensemble de la sélection d'événements peut être appliquée sur les données provenant de ce trigger indépendant. L'efficacité d'un trigger di-leptons est alors définie comme la fraction d'événements provenant du trigger multi-jets qui sont également présents dans l'échantillon de données provenant du trigger di-leptons.

Dans un premier temps, une présélection est appliquée afin de reconstruire des candidats événements $t\bar{t}$ di-leptons, la sélection des électrons et des muons va être présentée dans cette section. Par la suite des coupures supplémentaires seront appliquées de manière à réduire les bruits de fond.

Tout d'abord la collection d'électrons reconstruits dans un événement peut contenir des électrons qui possèdent les mêmes super-amas (cf. section 2.5.1) mais qui ont des traces reconstruites différentes. Pour éviter ce double comptage, seul l'électron ayant le rapport E/p le plus proche de 1 est conservé dans la collection d'électrons.

La première étape de la présélection des événements consiste à ne retenir que ceux contenant au moins deux leptons ayant une impulsion transverse $p_T > 10 \text{ GeV}/c$ (ou une énergie transverse $E_T > 10 \text{ GeV}$ pour les électrons) et une pseudo-rapidité $|\eta| < 2.4$ ³. Les graphiques de la figure 5.5 présentent les distributions en impulsion transverse des muons (en haut) et en énergie transverse des électrons (en bas) pour les leptons issus de bosons W (en bleu) ou de la désintégration faible de hadrons lourds (en rouge) dans des événements $t\bar{t} + 0$ jets. Les graphiques de droite de la même figure présentent les efficacités de sélection de ces leptons pour différentes coupures sur l'impulsion transverse des muons (en haut) et l'énergie transverse des électrons (en bas).

³Les lepton ne vérifiant pas $p_T > 10 \text{ GeV}/c$ ($E_T > 10 \text{ GeV}$ pour les électrons) et $|\eta| < 2.4$ sont rejetés et ne sont pas pris en compte dans l'analyse.

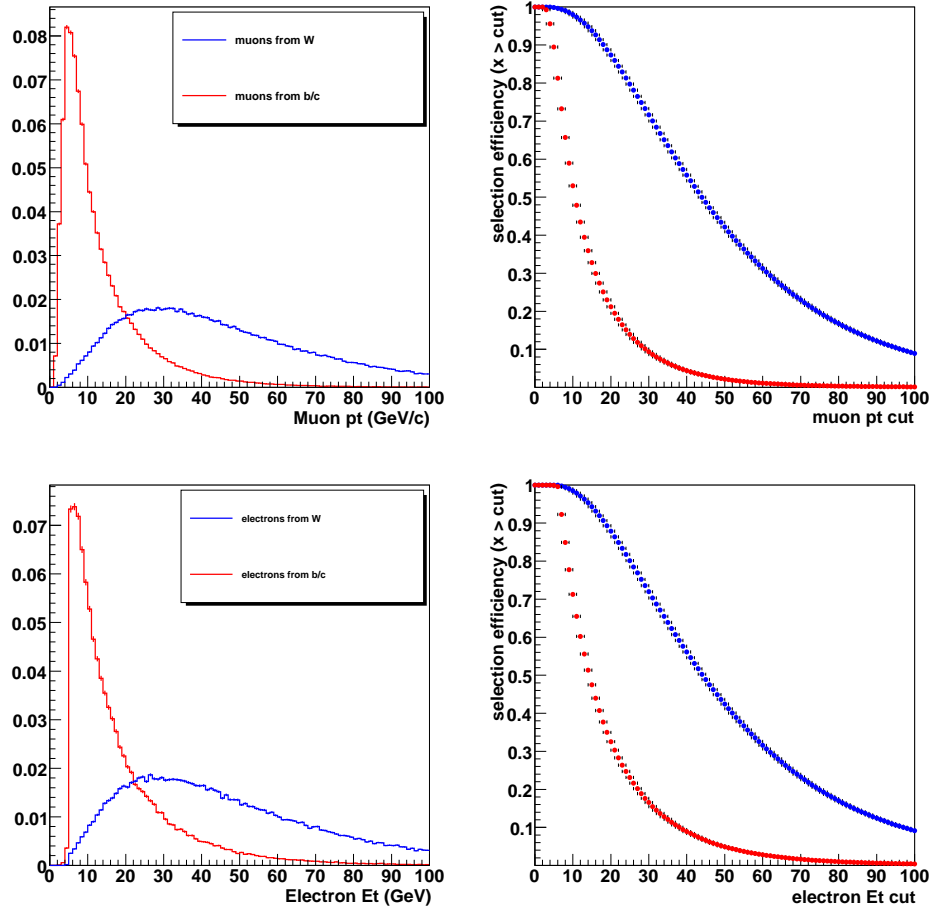


FIG. 5.5: Distribution de l'énergie transverse (à gauche) et efficacité de sélection après l'application de coupures en énergie/impulsion transverse pour les électrons (en haut) et les muons (en bas) provenant de la désintégration de bosons W (en bleu) ou de hadrons lourds (en rouge) pour des événements $t\bar{t}+0$ jet (détecteur parfait).

5.2.3 L'isolation des leptons

Des coupures d'isolation et d'identification sont alors appliquées. Les coupures d'isolation permettent de rejeter les leptons provenant de la désintégration faible de hadrons lourds (b , c) tandis que les coupures d'identification permettent de rejeter une grande partie de faux leptons.

Les critères d'isolation appliqués sur les leptons au niveau de la présélection sont :

- **l'isolation dans le trajectographe** : elle est définie par la somme de l'impulsion transverse de toutes les traces contenues dans un cône de rayon $\Delta R = 0.3$ autour du lepton, l'impulsion de la trace la plus proche du lepton étant soustraite. Un lepton isolé aura ainsi une variable d'isolation proche de 0,

- **l'isolation calorimétrique** : elle est définie comme la somme des énergies des tours calorimétriques (cf. section 2.5.1) contenues dans un cône de rayon $\Delta R = 0.3$ autour du lepton, l'énergie calorimétrique associée au lepton étant soustraite. Dans le cas des électrons, cette énergie est l'énergie *corrigée* du super-amas associé à l'électron. La valeur de l'isolation peut alors être négative.

Les distributions de l'isolation dans le trajectographe, normalisées à 1, sont présentées sur la figure 5.6 (à gauche) pour les électrons et les muons provenant de la désintégration de bosons W ou de hadrons lourds. Le comportement autour de 0 (grand nombre d'entrées à 0 suivi d'une chute) est dû à un effet de seuil provenant de la faible efficacité de reconstruction des traces de bas p_T .

En effet, pour dans la version CMSSW_1_6_12 (utilisée pour produire les échantillons Monte-Carlo de cette analyse) les traces de $p_T < 0.5$ GeV/c ne sont pas reconstruites. L'isolation dans le trajectographe est donc nulle si aucune trace de $p_T > 0.5$ GeV/c n'est reconstruite autour du lepton. Les efficacités de sélection correspondant à ces coupures d'isolation sont présentées dans les graphiques de droite de la figure 5.6.

Les distributions d'isolation calorimétrique normalisées à l'unité sont présentées dans les graphiques de gauche de la figure 5.7 pour les électrons et les muons, respectivement, provenant de la désintégration de bosons W ou de hadrons lourds. Dans le cas des électrons, on observe que l'isolation peut être négative. Cela est dû à l'utilisation dans le calcul de l'isolation de l'énergie corrigée des super-amas.

Les graphiques de la figure 5.8 présentent les efficacités de sélection des leptons issus de bosons W et de hadrons beaux ou charmés pour différentes coupures sur l'isolation calorimétrique et dans le trajectographe. Ils mettent en évidence les corrélations entre les variables d'isolation.

Les valeurs des coupures appliquées dans l'analyse sont :

- électrons : $E_{iso}^{calo} < 6$ GeV, $p_{T,iso}^{traj.} < 3$ GeV/c,
- muons : $E_{iso}^{calo} < 5$ GeV, $p_{T,iso}^{traj.} < 3$ GeV/c.

où E_{iso}^{calo} correspond à l'isolation calorimétrique et $p_{T,iso}^{traj.}$ à l'isolation dans le trajectographe. Aucune optimisation de sélection n'a été faite, les valeurs des coupures ont été choisies de manière à rejeter plus de 80% des leptons issus de hadrons lourds et à sélectionner au moins 80% de leptons issus de bosons W . Les efficacités de sélection des leptons isolés provenant de lepton W et provenant de hadrons beaux sont présentées dans le tableau 5.7 pour des événements $t\bar{t}$.

L'optimisation des valeurs des coupures d'isolation ne sera pas réalisée à partir d'événements Monte-Carlo mais sera étudiée à l'aide des vraies données. Une méthode *tag&probe* sera utilisée [57, 117] : dans les événements contenant 2 leptons ayant une masse inva-

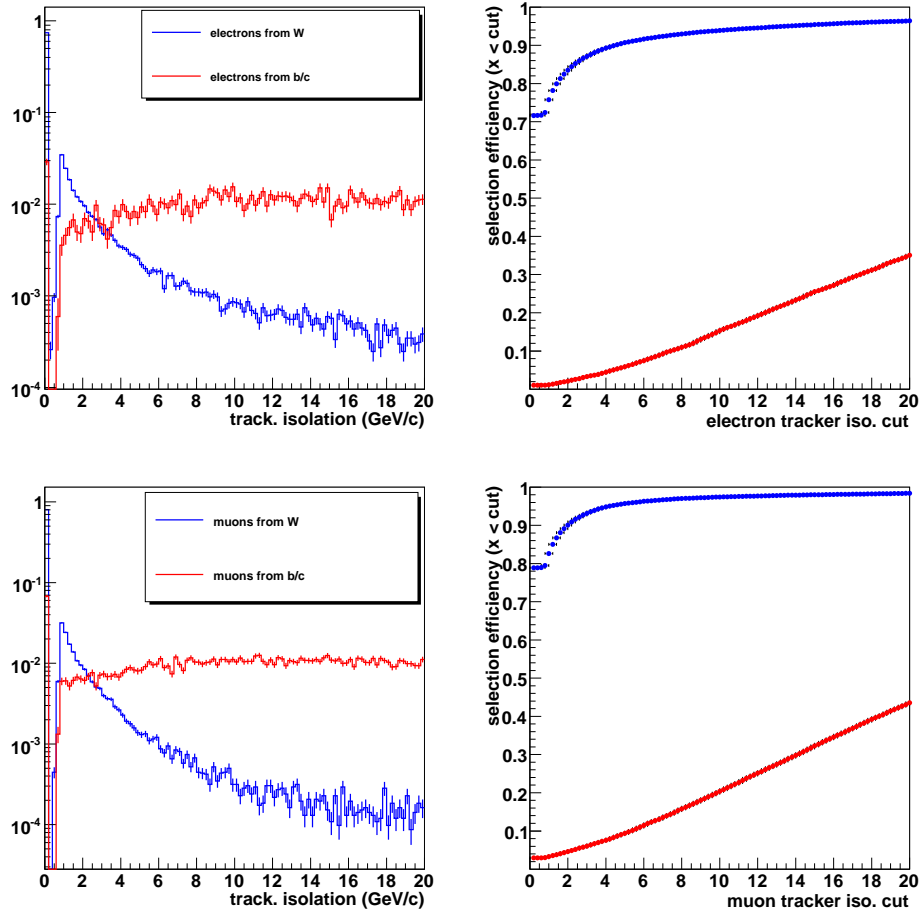


FIG. 5.6: Dans les événements $t\bar{t} + 0$ jet (soupe “Chowder”), isolation au niveau du trajectographe (à gauche) et efficacité de sélection (à droite) pour des électrons (en haut) et des muons (en bas) ayant une énergie (impulsion) transverse > 20 GeV(/c), une pseudo-rapacité $|\eta| < 2.4$ et provenant de bosons W (en bleu) ou de hadrons lourds (en rouge).

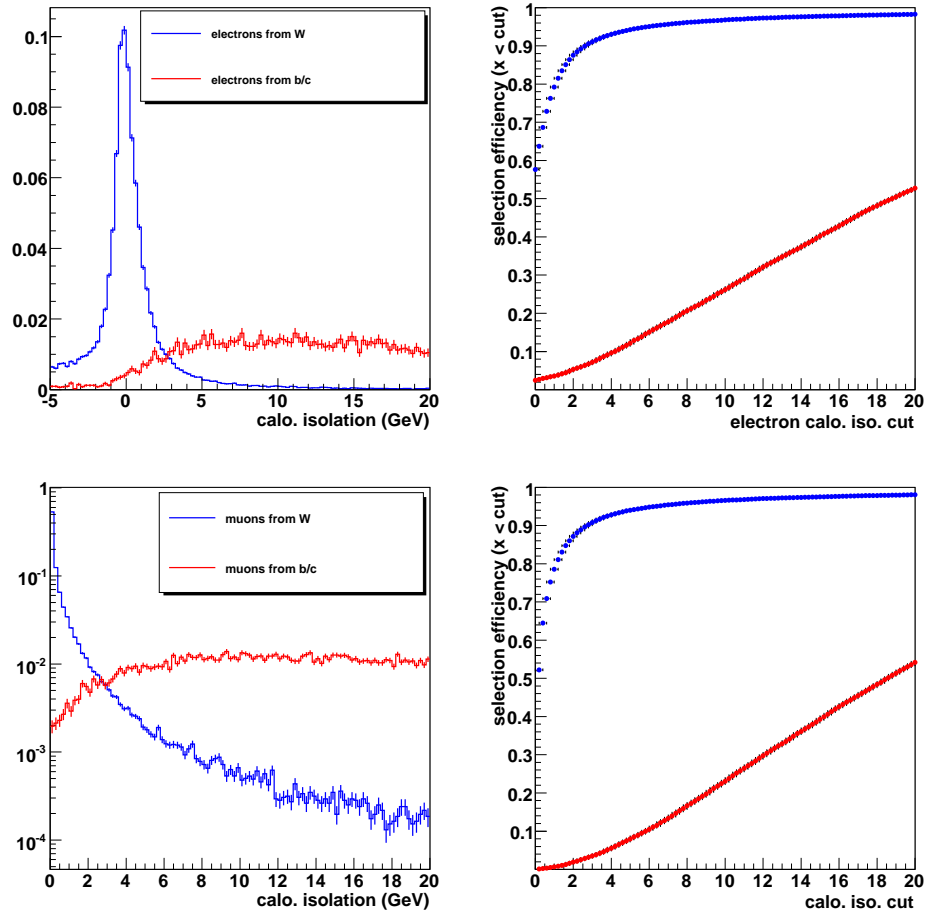


FIG. 5.7: Dans les événements $t\bar{t} + 0$ jet (soupe “Chowder”), isolation calorimétrique (à gauche) et efficacité de sélection correspondante (à droite) pour les électrons (en haut) et les muons (en bas) ayant une énergie (impulsion) transverse > 20 GeV(/c) et provenant de bosons W (en bleu) ou de hadrons lourds (en rouge).

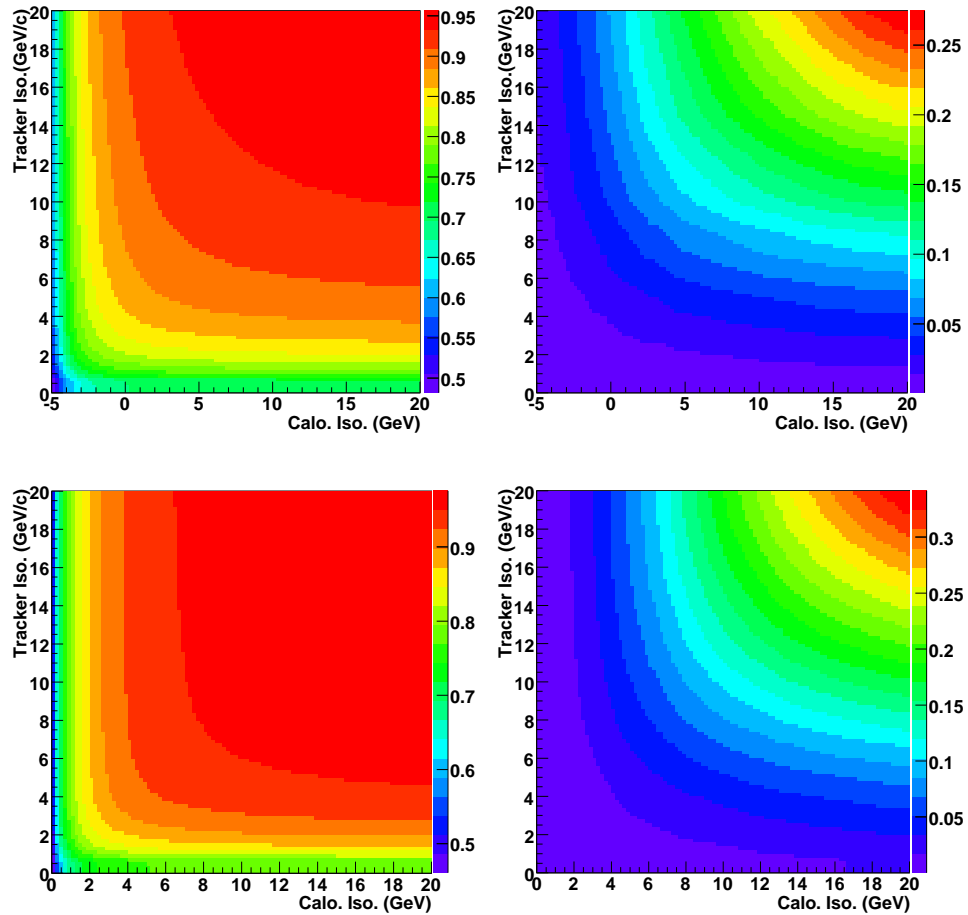


FIG. 5.8: Dans les événements $t\bar{t} + 0$ jet (soupe “Chowder”), efficacité de sélection (axe coloré) des leptons ayant une énergie (impulsion) transverse > 20 GeV(/c) issus de bosons W (à gauche) et de hadrons lourds (à droite) pour les électrons (en haut) et les muons (en bas) pour différentes coupures sur l’isolation calorimétrique (axe des abscisses) et au niveau du trajectographe (axe des ordonnées).

électrons	ε_W	$\varepsilon_{b/c}$
traj. isolation	$87.1 \pm 0.1\%$	$3.3 \pm 0.1\%$
calo. isolation	$89.5 \pm 0.1\%$	$6.1 \pm 0.1\%$
muons	ε_W	$\varepsilon_{b/c}$
traj. isolation	$93.2 \pm 0.1\%$	$6.0 \pm 0.1\%$
calo. isolation	$94.0 \pm 0.1\%$	$7.9 \pm 0.1\%$

TAB. 5.7: Efficacité de sélection des coupures d’isolation pour les leptons ayant une énergie (impulsion) transverse $> 20 \text{ GeV}(/c)$ et issus de bosons W et de hadrons lourds dans les événements $t\bar{t}+0$ jets (soupe “Chowder”).

riante compatible avec la masse du Z , un lepton *tag* est défini comme un lepton bien reconstruit passant des coupures de sélection serrées. Un lepton *probe* correspond à un lepton passant des coupures d’isolation relâchées. La coupure en masse invariante dileptons assure une bonne pureté d’événements contenant des Z . A partir du nombre d’événements ainsi sélectionnés (noté N_1) le nombre d’événements contenant un lepton *probe* qui passe les coupures d’isolation serrées (celles de la sélection) est compté (noté N_2). Les efficacités des coupures d’isolation sont alors estimées par le rapport N_1/N_2 .

5.2.4 Identification des électrons

En plus des critères d’isolation, une variable d’identification des électrons est utilisée afin de rejeter les faux électrons. Ces faux électrons sont en grande majorité des jets hadroniques ayant une faible multiplicité de traces qui déposent une quantité importante de leur énergie dans le calorimètre électromagnétique. Le critère d’identification des électrons est défini par une succession de coupures sur différentes variables comme : le rapport entre l’énergie déposée dans HCAL et l’énergie déposée dans ECAL, les différences d’angles $\Delta\eta$ et $\Delta\phi$ entre le super-amas et la trace associée et l’allure du super-amas suivant la direction η définie par la variable $\sigma_{\eta\eta}$:

$$\sigma_{\eta\eta} = \sum_i (\eta_i - \eta_s)^2 \frac{E_i}{E_{graine}}, \quad (5.1)$$

où la somme s’effectue sur l’ensemble des cristaux d’un amas, η_s est la pseudo-rapacité du cristal de plus grande énergie. Cette variable rend compte de l’étalement en η du super-amas associé à l’électron par rapport au cristal principal.

Pour être identifiés, les électrons reconstruits ne doivent perdre qu’une faible fraction de leur énergie sous forme de photons rayonnés. La perte d’énergie est estimée en comparant les impulsions de la trace de l’électron au vertex d’interaction et au niveau du calorimètre ECAL après la prise en compte des pertes d’énergie dans le trajectographe.

L'efficacité d'identification pour des électrons de $p_T > 20$ GeV/c provenant de la désintégration de bosons W dans les événements $t\bar{t}$ est de 98.1% tandis que le taux de faux électrons de $p_T > 20$ GeV/c qui passent les coupures d'identification dans des événements QCD ($50 < \hat{p}_T < 80$ GeV/c) est d'environ 6.9 %. Le taux de faux muons étant environ dix fois plus faible, aucun critère supplémentaire d'identification des muons n'est appliqué.

5.2.5 Choix du canal

Le choix du canal est fait à partir des leptons passant les coupures d'isolation et d'identification. Afin d'éviter les problèmes de double comptage, les canaux di-leptons sont choisis à partir des deux leptons de plus haut p_t (ou E_T pour les électrons) de manière exclusive au sein d'un même échantillon :

- dans le cas des événements passant le chemin de déclenchement “di-électrons”, si les deux leptons sélectionnés de plus haut E_T sont des électrons de charges opposées, les événements sont comptés comme candidats $t\bar{t}$ di-électrons, les autres événements sont rejetés.
- Dans le cas des événements passant le chemin de déclenchement “muon+X”, si les deux leptons sélectionnés de plus haut p_T et de charges opposées sont des muons, les événements sont comptés comme candidats $t\bar{t}$ di-muons. Si les deux leptons sélectionnés de plus haut p_T (E_T) et de charges opposées sont un électron et un muon, les événements sont comptés comme candidats $t\bar{t}$ électron-muon. Les autres événements sont rejetés.

Un double comptage d'événements peut cependant provenir d'événements identiques contenus dans chacun des chemins de déclenchement. Dans la simulation, ce double comptage a été estimé au niveau de la présélection et se trouve être négligeable ($\approx 1\%$).

Les derniers critères de présélection s'appliquent aux jets. Dans un premier temps, il est nécessaire d'enlever de la collection de jets les électrons reconstruits comme des jets. Ainsi, tous les jets contenus dans un cône $\Delta R < 0.3$ autour d'un électron reconstruit qui passe les coupures d'isolation et d'identification sont rejetés. La présélection est finalisée en sélectionnant les événements possédant au moins 2 jets reconstruits de $p_T > 15$ GeV/c et $|\eta| < 2.4$.

5.2.6 Sélection principale

La présélection des événements conduit à une première sélection des candidats di-leptons. Des sélections supplémentaires sont appliquées afin de rejeter un maximum d'événements de bruit de fond. Les coupures utilisées sont :

- Les muons et les électrons présélectionnés doivent avoir une impulsion transverse ou une énergie transverse > 20 GeV(/c). Comme le montrent les graphiques de la figure 5.5, une simple coupure à 20 GeV/c sur le p_T (ou E_T pour les électrons)

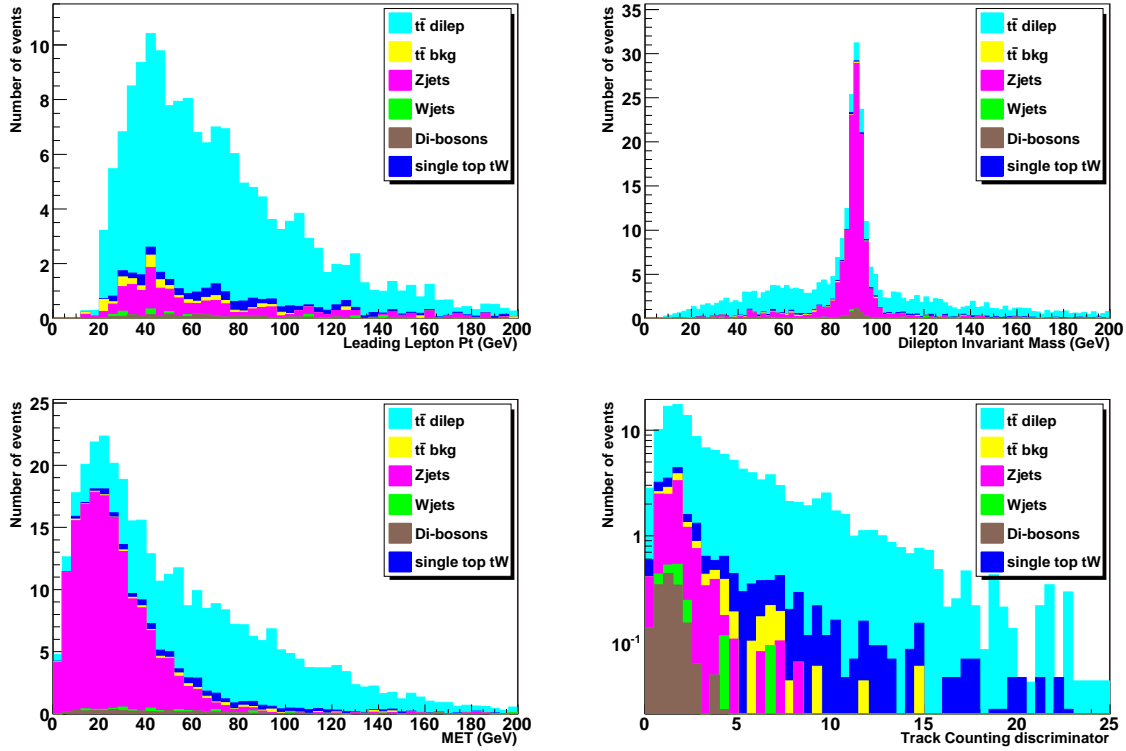


FIG. 5.9: Dans le canal $e - e$, distribution en E_T de l'électron de plus haut E_T (en haut à gauche), de la masse invariante des candidats di-électrons (en haut à droite), de l'énergie transverse manquante (en bas à gauche) et du discriminant de l'algorithme d'identification (TCHE) de jets b de plus haut p_T (en bas à droite), pour les événements de signal et de bruits de fond. Toutes les coupures de sélections sont appliquées, excepté la coupure d'identification des jets b et la coupure sur la variable représentée.

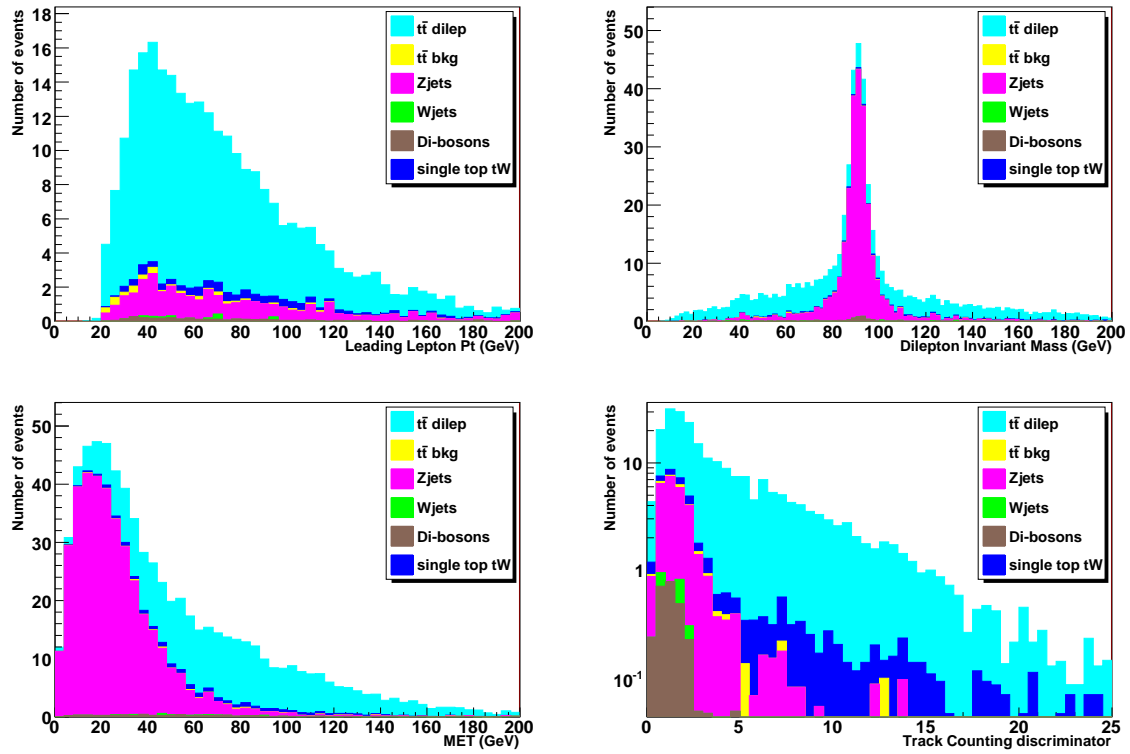
permet de rejeter 80% des muons et 70% des électrons issus de la désintégration de hadrons lourds, pour une efficacité de sélection des leptons issus de W de 90% environ.

- La présence d'au moins deux jets avec une impulsion transverse de plus de 30 GeV/c est requise.
- Afin de réduire le nombre d'événements Z +jets sélectionnés, la masse invariante des candidats di-leptons ne doit pas être voisine de la masse du Z : les événements $e - e$ et $\mu - \mu$ dont la masse invariante du système di-leptons est contenue dans l'intervalle [75, 105] GeV/c² sont rejetés. Cette coupure n'est pas appliquée pour le canal $e - \mu$.
- Les événements de signal ont une grande énergie transverse manquante \cancel{E}_T dû à la présence de deux neutrinos provenant de la désintégration des deux bosons W , tandis que les événements Drell-Yan et QCD doivent avoir une énergie transverse manquante \cancel{E}_T plus faible provenant uniquement d'effet de résolution. Une coupure sur l'énergie transverse manquante doit donc augmenter le rapport signal sur bruit. Dans le cas des canaux di-électrons et di-muons, les plus sensibles au bruit de fond Drell-Yan, la coupure en $\cancel{E}_T > 50$ GeV est appliquée tandis que dans le cas du canal électron-muon, une coupure plus relâchée (>20 GeV) est utilisée. Les sélections en \cancel{E}_T permettent de contrôler le bruit de fond QCD .
- La sélection est finalisée par l'utilisation de critères d'identification des jets b . Différents scénarios de sélection ont été étudiés. Ils correspondent à des sélections avec l'identification de 0, 1 ou 2 jets b . Au démarrage du LHC, les effets de désalignement du trajectographe dégradent significativement les performances des algorithmes d'identification des jets b . La sélection correspondante se doit d'être robuste et aussi simple que possible. C'est pourquoi l'algorithme utilisé est le comptage de trace TCHE (cf. section 4.1.1.1) avec le point de fonctionnement *loose* qui correspond à une coupure de 2.3 sur le discriminant.

Les graphiques présentés sur les figures 5.9, 5.10 et 5.11 illustrent les sélections. Ils correspondent aux distributions des différentes variables utilisées, lorsque toutes les coupures de sélection sont appliquées, excepté la sélection sur la variable représentée (sans appliquer de sélection sur l'identification des jets b).

Au démarrage du LHC, l'application de la coupure d'identification des jets b va dépendre des performances du détecteur. Des sélections avec 0, 1 ou 2 jets b identifiés ont donc été étudiées. Les erreurs systématiques seront estimées dans ces trois cas.

Le tableau 5.8 présente le nombre d'événements sélectionnés après l'application consécutive des coupures de sélection dans le canal di-électrons, pour une luminosité intégrée de 100 pb⁻¹. La ligne "preSel" correspond à la présélection présentée dans la section 5.2.5. La colonne $TTbkg$ correspond au bruit de fond $t\bar{t}$ constitué principalement d'événements lepton+jets. Avant coupure sur l'identification des jets b , les bruits de fond principaux proviennent des événements Z +jets et single top tW et dans une moindre mesure des événements $t\bar{t}$ semi-leptoniques.

FIG. 5.10: Similaire à la figure 5.9 pour le canal $\mu - \mu$

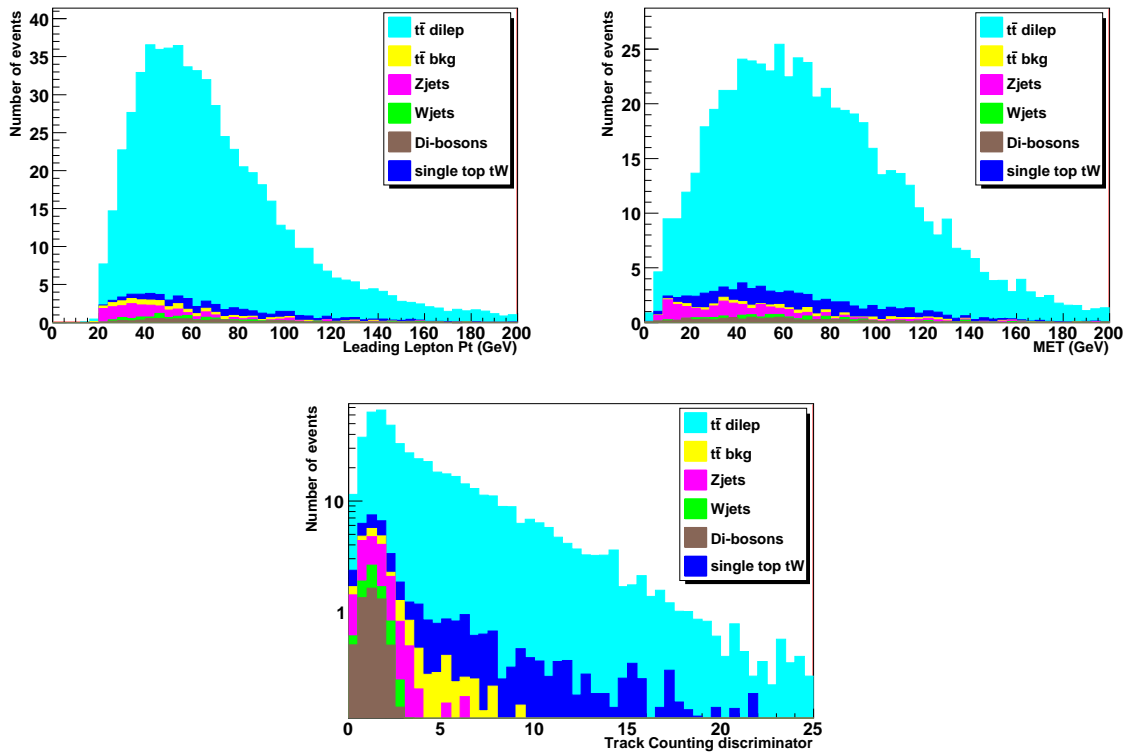


FIG. 5.11: Dans le canal $e - \mu$, distribution en p_T du lepton de plus haut p_T (en haut à gauche), de l'énergie transverse manquante (en haut à droite) et du discriminant de l'algorithme d'identification (TCHE) de jets b de plus haut p_T (en bas), pour les événements de signal et de bruits de fond. Toutes les coupures de sélection sont appliquées, excepté l'identification des jets b et la coupure sur la variable représentée.

sélection ($e - e$)	Signal	TTbkg	Z+jets	W+jets	WW	WZ	ZZ	tW	S/B
sans sélection	1357	81943	577700	8053000	11450	4990	1610	6200	10^{-4}
presel.	271.87 ± 3.08	26.35 ± 1.07	2065.19 ± 13.41	23.25 ± 1.61	4.12 ± 0.26	16.60 ± 0.48	12.30 ± 0.37	17.57 ± 0.63	0.13
2 leptons $p_t > 20$ GeV/c	242.90 ± 2.95	5.19 ± 0.48	1744.00 ± 12.22	2.44 ± 0.51	3.10 ± 0.22	15.58 ± 0.46	11.79 ± 0.36	15.05 ± 0.58	0.13
p_T Jet	241.53 ± 2.94	5.19 ± 0.48	1699.14 ± 12.05	2.44 ± 0.51	2.94 ± 0.22	15.36 ± 0.46	11.61 ± 0.36	14.86 ± 0.58	0.14
Masse inv.	189.14 ± 2.66	3.82 ± 0.41	152.43 ± 3.63	2.08 ± 0.47	2.43 ± 0.20	1.23 ± 0.13	1.21 ± 0.12	11.93 ± 0.52	1.08
$\cancel{E}_T > 50$ GeV	134.59 ± 2.30	2.69 ± 0.34	12.58 ± 0.98	0.85 ± 0.28	1.64 ± 0.16	0.25 ± 0.06	0.28 ± 0.06	8.05 ± 0.43	5.11
1 jet b identifié	112.15 ± 2.12	2.21 ± 0.31	3.33 ± 0.50	0.46 ± 0.20	0.37 ± 0.08	0.05 ± 0.03	0.09 ± 0.03	5.91 ± 0.37	9.03
2 jets b identifiés	49.59 ± 1.44	0.74 ± 0.18	0.83 ± 0.25	-	0.03 ± 0.02	-	0.01 ± 0.01	1.32 ± 0.17	16.92

TAB. 5.8: Nombre d'événements de signal et de bruit de fond sélectionnés dans le canal ($e - e$) après l'application cumulée des différentes coupures de sélection pour une luminosité intégrée de 100 pb^{-1} . La dernière colonne présente les rapports signaux sur bruit. Les erreurs statistiques sont le nombre d'événements Monte-Carlo sont présentés.

sélection ($e - \mu$)	Signal	TTbkg	Z+jets	W+jets	WW	WZ	ZZ	tW	S/B
sans sélection	2661	80638	577700	8053000	11450	4990	1610	6200	10^{-4}
presel.	630.93 ± 4.58	36.02 ± 1.25	86.07 ± 2.84	27.31 ± 1.72	8.35 ± 0.36	1.96 ± 0.16	0.57 ± 0.08	37.32 ± 0.92	3.19
2 leptons $p_t > 20$ GeV/c	577.22 ± 4.44	7.12 ± 0.56	19.59 ± 1.26	3.65 ± 0.63	7.29 ± 0.34	1.50 ± 0.14	0.38 ± 0.07	34.57 ± 0.88	7.79
p_T Jet	574.51 ± 4.43	6.95 ± 0.53	18.77 ± 1.23	3.65 ± 0.63	7.21 ± 0.34	1.48 ± 0.14	0.37 ± 0.06	34.25 ± 0.88	7.90
$\cancel{E}_T > 20$ GeV	544.57 ± 4.34	6.39 ± 0.53	13.89 ± 1.03	3.57 ± 0.63	6.54 ± 0.32	1.26 ± 0.13	0.31 ± 0.06	22.89 ± 0.72	9.93
1 jet b identifié	453.47 ± 4.05	4.86 ± 0.46	3.07 ± 0.46	0.58 ± 0.26	1.24 ± 0.14	0.19 ± 0.05	0.11 ± 0.04	17.07 ± 0.62	16.72
2 jets b identifiés	201.02 ± 2.84	1.59 ± 0.26	0.36 ± 0.16	-	0.14 ± 0.05	0.01 ± 0.01	0.04 ± 0.02	3.57 ± 0.28	35.20

TAB. 5.9: Identique au tableau 5.8 pour le canal ($e - \mu$).

Par rapport au nombre d'événements sélectionnés après la coupure sur l'énergie transverse manquante \cancel{E}_T , l'identification d'au moins un jet b réduit les bruits de fond Z+jets de 75%, tW de 26% et $t\bar{t}$ semi-leptoniques de 18%. L'efficacité de sélection du signal est de 83%. Par rapport au nombre d'événements sélectionnés après l'identification d'au moins un jet b , l'identification d'au moins un deuxième jet b permet de rejeter 86% des événements Z+jets et 78% des événements tW restants tout en conservant 44% des événements de signal.

Le tableau 5.9 présente le nombre d'événements sélectionnés après l'application séquentielles des coupures de sélection dans le canal électron-muon. Dans ce cas, le bruit

sélection ($\mu - \mu$)	Signal	TTbkg	Z+jets	W+jets	WW	WZ	ZZ	tW	S/B
sans sélection	1322	81978	577700	8053000	11450	4990	1610	6200	10^{-4}
presel.	431.41 ± 3.55	7.53 ± 0.57	3375.79 ± 17.09	2.97 ± 0.59	6.02 ± 0.31	23.43 ± 0.57	18.67 ± 0.46	22.84 ± 0.72	0.12
2 leptons $p_t > 20$ GeV/c	404.89 ± 3.49	4.46 ± 0.44	3322.11 ± 16.97	2.34 ± 0.52	6.02 ± 0.31	23.43 ± 0.57	18.67 ± 0.46	22.84 ± 0.72	0.12
p_T Jet	402.38 ± 3.48	4.46 ± 0.44	3234.68 ± 16.72	2.23 ± 0.51	5.90 ± 0.31	23.10 ± 0.56	18.47 ± 0.45	22.68 ± 0.72	0.12
Masse Inv.	315.09 ± 3.23	3.08 ± 0.37	369.73 ± 5.74	1.62 ± 0.43	4.84 ± 0.28	1.96 ± 0.16	2.08 ± 0.15	17.98 ± 0.64	0.78
$\cancel{E}_T > 50$ GeV	223.02 ± 2.84	1.92 ± 0.29	36.19 ± 1.69	0.94 ± 0.33	3.09 ± 0.22	0.49 ± 0.08	0.52 ± 0.08	12.23 ± 0.53	4.02
1 jet b identifié	188.09 ± 2.65	1.31 ± 0.24	8.96 ± 0.82	0.08 ± 0.08	0.49 ± 0.09	0.12 ± 0.04	0.12 ± 0.038	8.75 ± 0.45	9.48
2 jets b identifiés	87.74 ± 1.89	0.36 ± 0.13	1.57 ± 0.34	-	0.06 ± 0.03	0.01 ± 0.01	0.03 ± 0.02	2.09 ± 0.22	21.30

TAB. 5.10: Identique au tableau 5.8 pour le canal ($\mu - \mu$).

de fond Z +jets est bien moins important. Cela justifie la coupure plus relâchée sur l'énergie transverse manquante.

De la même manière, par rapport au nombre d'événements sélectionnés après la coupure en \cancel{E}_T , l'identification d'au moins un jet b réduit les bruits de fond Z +jets de 78%, tW de 24% et $t\bar{t}$ semi-leptonique de 26%. L'efficacité de sélection du signal est de 83%. Par rapport au nombre d'événements sélectionnés après l'identification d'au moins un jet b , l'identification d'au moins un deuxième jet b permet de rejeter 88% des événements Z +jets et 80% des événements tW restants tout en conservant 44% des événements de signal.

Le tableau 5.10 présente le nombre d'événements sélectionnés après l'application consécutive des coupures de sélection dans le canal di-muons.

Par rapport au nombre d'événements sélectionnés après la coupure en \cancel{E}_T , l'identification d'au moins un jet b réduit les bruits de fond Z +jets de 75%, tW de 29% et $t\bar{t}$ de 31%. L'efficacité de sélection du signal est de 84%. Par rapport au nombre d'événements sélectionnés après l'identification d'au moins un jet b , l'identification d'au moins un deuxième jet b permet de rejeter 82% des événements Z +jets et 24% des événements tW restants tout en conservant 47% des événements de signal. Après cette dernière sélection, le bruit de fond résiduel est pratiquement nul ce qui conduit à un échantillon très pur d'événements signal.

Il est à noter que le bruit de fond QCD est complètement négligeable dans les canaux $e - \mu$ et $\mu - \mu$, en raison du faible taux de faux muons reconstruits. Le canal $e - e$ peu dépendre beaucoup plus du taux de faux électrons reconstruits qui peut être important, même si le nombre d'événements de la soupe "Gombo" (QCD PYTHIA) qui passent les coupures des présélection est nul.

Le choix définitif de la sélection va dépendre de l'erreur globale sur la mesure de la section efficace. Dans le cas où l'identification des jets b est utilisée, le nombre d'événements de bruits de fond sélectionnés devient plus faible. Les erreurs systématiques liées à l'estimation des bruits de fond sont alors fortement diminuées. L'erreur systématique liée à l'identification de jets b est cependant plus importante. Le nombre d'événements sélectionnés étant plus faible, les erreurs statistiques augmentent également.

Lorsqu'aucun critère d'identification des jets b n'est utilisé, l'erreur systématique correspondante est évidemment nulle et l'erreur statistique est également réduite. L'erreur liée à l'estimation des bruits de fond devient cependant plus importante. Au démarrage du LHC, l'incertitude sur le bruit de fond principal, les bruits de fond single top tW , sera importante car celui-ci est mal connue.

Une étude des erreurs systématiques principales sera présentée dans la section 5.7. Elle permettra de choisir la sélection finale qui minimise l'erreur globale sur la mesure.

D'une manière générale, les événements de signal peuvent être, dans les trois canaux, sélectionnés avec une très bonne pureté. Le canal $e - \mu$ est tout particulièrement intéressant car il permet d'obtenir un grand rapport signal sur bruit avec un grand nombre d'événements de signal sélectionnés. Ce canal est moins sensible au bruit de fond Z +jets, ce qui permet d'utiliser une coupure relâchée en \cancel{E}_T avec de plus faibles erreurs systématiques dûes à la correction en énergie des jets.

Le canal $\mu - \mu$, bien que plus sensible au bruit de fond Z +jets, est moins sensible aux bruits de fond instrumentaux car le taux de faux muons reconstruits est faible.

Le canal $e - e$ devrait conduire aux erreurs les plus importantes car le taux de faux électrons peut être important. Dans les événements QCD ($50 < \hat{p}_T < 80$ GeV/c), le taux d'électrons reconstruits avec $E_T > 20$ GeV, non associés à un électron généré et passant les coupures de présélection, est d'environ 0.15%. Un traitement spécial de ce canal sera présenté dans la section 5.3 suivante. Une méthode d'estimation des bruits de fond instrumentaux dans ce canal sera également présentée dans la section 5.4.

5.3 Estimation du bruit de fond QCD dans la simulation

Ce chapitre est dédié à l'estimation du bruit de fond QCD à partir du Monte-Carlo. Le but étant d'estimer les performances du détecteur au démarrage et de la contamination de bruit de fond même grossière. Après la sélection, aucun événement de l'échantillon "Gumbo" (QCD PYTHIA, cf. tableau 5.6) n'est sélectionné. Il est cependant important de vérifier ce résultat car l'estimation du bruit de fond QCD dans des événements Monte-Carlo s'avère particulièrement difficile pour principalement deux raisons :

- le bruit de fond QCD est très fortement lié aux bruits de fond instrumentaux et à l'estimation des taux de faux leptons isolés dans la simulation. La sous-estimation de ce taux conduit à la sous-estimation du bruit de fond QCD .

- De très grandes sections efficaces combinées avec de très faibles efficacités de sélection posent le problème du manque d'événements Monte-Carlo. En effet, il est difficile de générer suffisamment d'événements QCD pour obtenir une bonne estimation de ce bruit de fond.

Ces deux aspects sont particulièrement importants dans le cas du canal $e-e$ pour lequel le taux de faux électrons dans les événements QCD est non négligeable.

5.3.1 Factorisation

Les effets de reconstruction liés à l'imperfection du détecteur ne pourront être estimés qu'avec les données réelles, même s'ils sont partiellement pris en compte par la décalibration et le désalignement du détecteur dans la simulation. Pour s'accommoder du manque d'événements, une méthode de factorisation est utilisée. L'idée est d'estimer le bruit de fond QCD en factorisant les efficacités de sélection pour chaque coupure ou pour différents ensembles de coupures. La méthode de factorisation n'est pas le meilleur moyen d'estimer les bruits de fond instrumentaux et ceux liés aux effets de reconstruction car la plupart des variables utilisées sont corrélées et les effets de mauvaise reconstruction sont mal décrits dans la simulation. Le but est cependant d'estimer un ordre de grandeur de la contamination des événements QCD .

Les événements QCD sont rejetés en majorité par les coupures d'isolation et d'identification des électrons. Demander au moins deux électrons isolés de $E_T > 20$ GeV suffit à rejeter la majorité des événements QCD Monte-Carlo. Soit $\varepsilon_{1e, 2jet}^{10, 15}$ l'efficacité de sélection d'un électron isolé et identifié de $E_T > 15$ GeV et de deux jets de $p_T > 10$ GeV/c ; soit ε_{1e}^{10} l'efficacité de sélection d'un électron isolé et identifié de $E_T > 10$ GeV ; l'efficacité de présélection (cf. section 5.2.5) devient alors :

$$\varepsilon_{presel.} = \frac{\varepsilon_{1e, 2jet}^{10, 10} \times \varepsilon_{1e}^{10}}{2}. \quad (5.2)$$

La division par deux rend compte de la coupure sur la charge des électrons. L'efficacité $\varepsilon_{1e, 2jet}^{10, 10}$ permet de réduire les corrélations entre les coupures sur les jets et les coupures sur les électrons. Les nombres d'événements attendus sont alors calculés en multipliant $\varepsilon_{presel.}$ par les sections efficaces⁴.

De la même manière, l'efficacité de la sélection globale peut être décomposée en un produit des différentes efficacités des coupures de sélection. Les nombres d'événements QCD sélectionnés prédits par ces factorisations sont présentés dans le tableau 5.11 pour une luminosité intégrée de 100 pb^{-1} . La colonne (a) représente la présélection.

Considérons l'efficacité de sélection $\varepsilon_{1e, 2jet}^{20, 15}$ des événements contenant 1 électron isolé et identifié de $E_T > 20$ GeV et de deux jets de $p_T > 15$ GeV/c et l'efficacité de sélection

⁴Les sections efficaces de productions des événements QCD sont présentés dans le tableau 5.6 et la luminosité intégrée

ε_{1e}^{20} d'un l'électron isolé et identifié de $E_T > 20$ GeV. Les valeurs de la colonne (b) sont alors obtenues par la formule :

$$\varepsilon_{(b)} = \frac{\varepsilon_{1e, 2jet}^{20, 15} \times \varepsilon_{1e}^{20}}{2}. \quad (5.3)$$

Cette efficacité est ensuite successivement multipliée par l'efficacité de sélectionner deux jets de $p_T > 30$ GeV/c parmi les événements contenant 2 jets de $p_T > 10$ GeV/c (ε_{jets} , colonne (c)), par l'efficacité de sélection en \cancel{E}_T (ε_{met} , colonne (d)) et par l'efficacité d'identification d'au moins un jet b (ε_b , colonne (e)). L'efficacité de la sélection globale est alors donnée par la formule :

$$\varepsilon_{totale} = \frac{\varepsilon_{1e, 2jet}^{20, 15} \times \varepsilon_{1e}^{20}}{2} \times \varepsilon_{jets} \times \varepsilon_{met} \times \varepsilon_b. \quad (5.4)$$

Les nombres d'événements prédits par la méthode de factorisation sont résumés dans le tableau 5.11 pour les différentes coupures de sélection. La factorisation des efficacités ε_{jets} , ε_{met} et ε_b a été vérifiée, les différences relatives induites par la factorisation sont comprise entre 5 et 40% environ et dépendent principalement du \hat{p}_T . Les grandes variations sont principalement dues à des fluctuations statistiques et donc contribuent peu au résultat final. La factorisation des efficacités $\varepsilon_{1e, 2jet}^{20, 10}$ et ε_{1e}^{20} n'a cependant pas pu être vérifiée car la sélection de 2 leptons isolés suffit à rejeter la majorité des événements QCD . Rappelons que le but est ici d'estimer un *ordre de grandeur* du nombre d'événements QCD sélectionnés.

Comme très peu d'événements QCD contiennent deux électrons isolés et identifiés de $E_T > 20$ GeV, les effets de la coupure sur la masse invariante n'ont pas été pris en compte. On s'attend cependant à ce qu'ils restent faibles. Par ailleurs, cette estimation a été faite pour des événements produits dans le cas d'un détecteur parfait, de ce fait la contamination du bruit de fond QCD est certainement sous évaluée.

On remarque qu'après l'application de la coupure en \cancel{E}_T , le nombre d'événements QCD attendu est inférieur à 1. Après l'identification d'un jet b , le bruit de fond QCD devrait être complètement négligeable. Lorsque ni la coupure en \cancel{E}_T ni l'identification des jets b ne sont utilisés, il reste une quinzaine d'événements QCD . Les intervalles $30 < \hat{p}_T < 50$ GeV/c et $50 < \hat{p}_T < 80$ GeV/c contribuent le plus au bruit de fond QCD car l'énergie mise en jeu est suffisante pour créer plusieurs jets de $p_T > 30$ GeV/c et parce que leur section efficace est grande.

5.3.2 Sensibilité au taux de faux électrons

La sensibilité de la sélection en fonction de la valeur du taux de faux électrons ε_{fake} peut être estimée en introduisant un facteur α tel que $\varepsilon_{fake} \rightarrow \alpha \times \varepsilon_{fake}$. Cette variation du taux de faux électrons va avoir un impact sur le bruit de fond QCD . Comme les événements QCD contiennent deux faux électrons, le nombre d'événements QCD

Lots QCD \hat{p}_T	(a) présel.	(b) 2 e sel. ($p_T > 20$ GeV/c)	(c) 2 jets ($p_T > 30$ GeV/c)	(d) \cancel{E}_T (> 50 GeV)	(e) 1 jet b TCHE loose
15 $< \hat{p}_T < 20$	1909 ± 117	1.51×10^{-1} $\pm 0.01 \times 10^{-1}$	1.1×10^{-3} $\pm 0.6 \times 10^{-3}$	5.2×10^{-7} $\pm 2.8 \times 10^{-7}$	2.7×10^{-9} $\pm 1.7 \times 10^{-9}$
20 $< \hat{p}_T < 30$	949 ± 38	7.41×10^{-1} $\pm 0.03 \times 10^{-1}$	3.1×10^{-2} $\pm 0.6 \times 10^{-2}$	6.4×10^{-5} $\pm 1.3 \times 10^{-5}$	1.5×10^{-6} $\pm 0.3 \times 10^{-6}$
30 $< \hat{p}_T < 50$	1991 ± 91	27.3 ± 0.3	6.5 ± 0.7	8.6×10^{-2} $\pm 1.0 \times 10^{-2}$	8.7×10^{-3} $\pm 0.1 \times 10^{-3}$
50 $< \hat{p}_T < 80$	333 ± 18	10.9 ± 0.2	6.3 ± 0.7	3.0×10^{-1} $\pm 0.32 \times 10^{-1}$	7.9×10^{-2} $\pm 0.9 \times 10^{-2}$
80 $< \hat{p}_T < 120$	44 ± 2	1.12 ± 0.02	8.4×10^{-1} $\pm 0.9 \times 10^{-1}$	6.2×10^{-2} $\pm 0.7 \times 10^{-2}$	2.8×10^{-2} $\pm 0.3 \times 10^{-2}$
120 $< \hat{p}_T < 170$	6.0 ± 0.3	1.44×10^{-1} $\pm 0.02 \times 10^{-1}$	1.2×10^{-1} $\pm 0.1 \times 10^{-1}$	1.3×10^{-2} $\pm 0.1 \times 10^{-2}$	8.6×10^{-3} $\pm 0.1 \times 10^{-3}$
170 $< \hat{p}_T < 230$	1.2 ± 0.1	2.59×10^{-2} $\pm 0.03 \times 10^{-2}$	2.3×10^{-2} $\pm 0.2 \times 10^{-2}$	3.6×10^{-3} $\pm 0.4 \times 10^{-3}$	3.1×10^{-3} $\pm 0.3 \times 10^{-3}$
230 $< \hat{p}_T < 300$	2.9×10^{-1} $\pm 0.2 \times 10^{-1}$	7.5×10^{-3} $\pm 0.1 \times 10^{-3}$	7.1×10^{-3} $\pm 0.8 \times 10^{-3}$	1.5×10^{-3} $\pm 0.4 \times 10^{-3}$	1.5×10^{-3} $\pm 0.2 \times 10^{-3}$
300 $< \hat{p}_T < 380$	7.1×10^{-2} $\pm 0.3 \times 10^{-1}$	1.86×10^{-3} $\pm 0.02 \times 10^{-3}$	1.8×10^{-3} $\pm 0.2 \times 10^{-3}$	4.8×10^{-4} $\pm 4.3e-05 \times 10^{-4}$	5.8×10^{-4} $\pm 0.5 \times 10^{-4}$

TAB. 5.11: Nombres d'événements QCD normalisés à une luminosité de 100 pb^{-1} passant les coupures de sélection estimées par factorisation, pour différents intervalles en \hat{p}_T .

sélectionnés évolue comme α^2 . Les bruits de fond W +jets et $t\bar{t}$ semi-leptoniques sont également affectés. Comme ils contiennent un seul faux électron, les nombres d'événements W +jets et $t\bar{t}$ semi-leptoniques sélectionnés évoluent comme α .

La figure 5.12 représente, en fonction de α , le rapport signal sur bruit S/B et les fractions d'événements QCD (N_{QCD}), W +jets (N_W) et $t\bar{t}$ semi-leptoniques ($N_{other\bar{t}t}$) contenus dans l'ensemble des événements bruit de fond⁵, calculés à l'aide de la sélection présentée dans la section 5.2.6 et de la méthode de factorisation.

Pour des faibles valeurs de α (< 5) le bruit de fond $t\bar{t}$ semi-leptonique est le plus sensible au taux de faux électrons. Cela est dû au nombre plus important d'événements $t\bar{t}$ semi-leptoniques sélectionnés. La contamination des événements QCD augmente cependant rapidement avec α . Le bruit de fond QCD devient plus important que le bruit de fond W +jets pour des valeurs de $\alpha > 2$. Pour $\alpha > 6$, le bruit de fond QCD devient alors le bruit de fond principal et la fraction d'événements $t\bar{t}$ semi-leptoniques dans le bruit de fond commence à diminuer.

Pour conclure cette section, on peut dire que si le taux de faux électrons isolés est modéré (de l'ordre de 5%), le bruit de fond QCD devrait être négligeable dans le canal $e - e$. Cependant, si le taux de faux électrons isolés est supérieur à deux fois le taux présent dans la simulation actuelle (environ 10 %), la contribution du bruit de fond QCD n'est plus négligeable. Il est donc important de mettre au point une méthode d'estimation des bruits de fond instrumentaux à partir des données réelles. Une telle méthode est présentée dans la section 5.4.

⁵ $N_{QCD} + N_{other\bar{t}t} + N_W + N_Z + N_{tW}$ où N_Z est le nombre d'événements Z +jets sélectionnés et N_{tW} le nombre d'événements single top

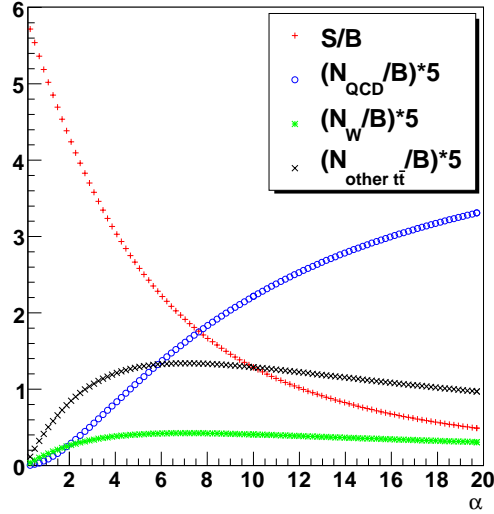


FIG. 5.12: Après la sélection globale des événements et la factorisation des efficacités de sélection, en fonction de α , rapport signal sur bruit (en rouge) et fractions d'événements QCD (en bleu), W +jets(en vert) et $t\bar{t}$ semi-leptoniques (en noir) contenu dans l'ensemble des événements bruits de fond ($\times 5$).

5.4 Estimation du bruit de fond à partir des données : la méthode de la matrice di-électrons

Le signal et les bruits de fond physiques décrits dans la section 5.1 peuvent être estimés avec une bonne précision en utilisant des événements simulés s'il y a un bon accord entre données et simulation Monte-Carlo. Les bruits de fonds instrumentaux sont fortement liés aux performances du détecteur et doivent être estimés à partir des données. Une adaptation originale de la méthode de la matrice, initialement développée dans l'expérience $D\bar{O}$ au Tevatron [91], va être présentée.

A l'origine, elle permettait d'estimer le bruit de fond QCD dans le canal lepton+jets. L'idée principale est de définir deux ensembles de coupures, l'une *tight* et l'autre *loose*. La sélection *tight* correspond aux coupures standards utilisées pour la mesure et la sélection *loose* correspond à une sélection où les critères d'isolation appliqués aux leptons ont été relâchés. A partir des nombres d'événements sélectionnés dans les lots *loose* et *tight*, et connaissant les efficacités des coupures d'isolation, un système de deux équations à deux inconnues peut être défini. Sa résolution permet d'estimer le nombre d'événements de bruit de fond QCD sélectionné.

Dans le cas de notre analyse (canal $t\bar{t}$ di-leptons), les coupures d'isolation des leptons sélectionnés peuvent être relâchées sur **un** ou **deux** leptons. Connaissant les efficacités

des coupures d'isolation, nous pouvons définir un système de trois équations à trois inconnues. La résolution de ce système nous permettra d'estimer la contamination des événements de type W +jets⁶ et des événements de type QCD ⁷ d'une part, et d'estimer le nombre d'événements de type signal⁸ sélectionnés d'autre part.

Comme une grande quantité d'événements est nécessaire, la méthode de la matrice a été testée sans critère d'identification des jets b . De plus, comme le taux de faux électrons est bien supérieur au taux de faux muons, seul le canal di-électrons est étudié. Cela permet d'avoir un nombre raisonnable d'événements pour tester la méthode. Il est également à noter que les bruits de fond physiques di-bosons et tW ont été négligés dans cette étude, mais peuvent facilement être inclus.

5.4.1 Description de la méthode

On définit donc trois ensembles de coupures : *loose*, *medium* et *tight*. La sélection *tight* correspond à la sélection standard définie dans la section 5.2.6 (sélection d'au moins 2 vrais leptons isolés), quand aucun algorithme d'identification des jets b n'est utilisé. La sélection *medium* correspond au cas où les coupures d'isolation sont relâchées sur un des deux leptons sélectionnés (sélection d'au moins un vrai lepton isolé). La sélection *loose* correspond au cas où aucune coupure d'isolation n'est appliquée sur les leptons. Chaque sélection correspond donc à une certaine multiplicité de vrais leptons provenant de la désintégrations de bosons W ou Z . Les échantillons sont ainsi regroupés en trois lots différents :

- **Echantillon de type signal** (N_S), il contient les événements possédant 2 vrais leptons isolés provenant de bosons W et Z . Il s'agit des événements $t\bar{t}$ di-leptons, di-bosons et single top tW .
- **Echantillon de type W + jets** (N_W), il contient les événements possédant 1 vrai et 1 faux lepton isolé. Il s'agit des événements W +jets et $t\bar{t}$ lepton+jets.
- **Echantillon de type QCD** (N_{QCD}), il contient les événements possédant 2 faux leptons isolés. Il s'agit des événements QCD et $t\bar{t}$ hadronique.

On peut ainsi définir un ensemble de trois équations :

$$N^t = N_S^t + N_W^t + N_{QCD}^t, \quad (5.5)$$

$$N^m = N_S^m + N_W^m + N_{QCD}^m, \quad (5.6)$$

$$N^l = N_S^l + N_W^l + N_{QCD}^l, \quad (5.7)$$

où N_S est le nombre d'événements de type signal, N_W le nombre d'événements de type W +jets et N_{QCD} le nombre d'événements de type QCD ; N^t , N^m et N^l sont les nombres

⁶contenant également les événements $t\bar{t}$ lepton+jets

⁷contenant également les événements $t\bar{t}$ hadroniques

⁸ $t\bar{t}$ di-leptons et bruits de fond physique Z +jets, single top tW et di-bosons

d'événements passant les coupures *tight*, *medium* et *loose* respectivement (avec $N^t \subset N^m \subset N^l$) et sont directement données par les mesures expérimentales.

On peut alors introduire les efficacités de passer de la sélection *loose* à *medium* ($\varepsilon_S^{l \rightarrow m}$, $\varepsilon_W^{l \rightarrow m}$, $\varepsilon_{QCD}^{l \rightarrow m}$) et les efficacités de passer d'une sélection *loose* à *tight* ($\varepsilon_S^{l \rightarrow t}$, $\varepsilon_W^{l \rightarrow t}$, $\varepsilon_{QCD}^{l \rightarrow t}$). Dans ce cas, les équations (5.5) à (5.7) deviennent :

$$N^t = \varepsilon_S^{l \rightarrow t} N_S^l + \varepsilon_W^{l \rightarrow t} N_W^l + \varepsilon_{QCD}^{l \rightarrow t} N_{QCD}^l, \quad (5.8)$$

$$N^m = \varepsilon_S^{l \rightarrow m} N_S^l + \varepsilon_W^{l \rightarrow m} N_W^l + \varepsilon_{QCD}^{l \rightarrow m} N_{QCD}^l, \quad (5.9)$$

$$N^l = N_S^l + N_W^l + N_{QCD}^l. \quad (5.10)$$

Si ces efficacités peuvent être évaluées indépendamment, les équations (5.8) à (5.10) forment un système de trois équations à trois inconnues qui correspondent aux nombres d'événements passant les critères de sélection *loose* N_S^l , N_W^l et N_{QCD}^l . Les nombres d'événements de type signal et de type bruit de fond passant les critères de sélection *tight* (sélection standard) sont alors donnés par :

$$N_S^t = \varepsilon_S^{l \rightarrow t} N_S^l, \quad (5.11)$$

$$N_W^t = \varepsilon_W^{l \rightarrow t} N_W^l, \quad (5.12)$$

$$N_{QCD}^t = \varepsilon_{QCD}^{l \rightarrow t} N_{QCD}^l, \quad (5.13)$$

La difficulté principale étant l'estimation des efficacités $\varepsilon_S^{l \rightarrow m}$, $\varepsilon_W^{l \rightarrow m}$, $\varepsilon_{QCD}^{l \rightarrow m}$, $\varepsilon_S^{l \rightarrow t}$, $\varepsilon_W^{l \rightarrow t}$ et $\varepsilon_{QCD}^{l \rightarrow t}$ à partir des données.

5.4.2 Hypothèse de factorisation

Des hypothèses de factorisation peuvent être introduites pour permettre la résolution du système d'équations (5.8)-(5.10). On considérera que l'efficacité des coupures d'isolation pour un lepton provenant de la désintégration de bosons W peut être factorisée avec l'efficacité de sélection globale. Une hypothèse similaire est faite pour l'efficacité des coupures d'isolation d'un faux lepton ou d'un lepton provenant de hadrons lourds.

- Premièrement, nous faisons l'hypothèse que les efficacités de passer des coupures *loose* aux coupures *medium* peuvent s'écrire :

$$\varepsilon_S^{l \rightarrow m} = 2\varepsilon_s - \varepsilon_s^2 \quad (5.14)$$

$$\varepsilon_W^{l \rightarrow m} = \varepsilon_s + \varepsilon_{fake} - \varepsilon_s \cdot \varepsilon_{fake} \quad (5.15)$$

$$\varepsilon_{QCD}^{l \rightarrow m} = 2\varepsilon_{fake} - \varepsilon_{fake}^2 \quad (5.16)$$

où ε_s (ε_{fake}) est la probabilité pour un vrai (faux) électron isolé de passer les coupures d'isolation. Les équations (5.14) et (5.16) représentent la probabilité de sélectionner au moins 1 vrai (faux) lepton isolé à partir de deux candidats et l'équation (5.15) la probabilité de sélectionner un vrai et un faux leptons isolés.

- Deuxièmement, nous faisons l'hypothèse que les efficacités de passer des coupures *loose* aux coupures *tight* peuvent s'écrire :

$$\varepsilon_S^{l \rightarrow t} = \varepsilon_s^2, \quad (5.17)$$

$$\varepsilon_W^{l \rightarrow t} = \varepsilon_s \cdot \varepsilon_{fake}, \quad (5.18)$$

$$\varepsilon_{QCD}^{l \rightarrow t} = \varepsilon_{fake}^2. \quad (5.19)$$

A partir de ces hypothèses, le système d'équation (5.8)-(5.10) devient :

$$N^t = \varepsilon_s^2 N_S^l + \varepsilon_s \cdot \varepsilon_{fake} N_W^l + \varepsilon_{fake}^2 N_{QCD}^l, \quad (5.20)$$

$$N^m = (2\varepsilon_s - \varepsilon_s^2) N_S^l + (\varepsilon_s + \varepsilon_{fake} - \varepsilon_s \varepsilon_{fake}) N_W^l + (2\varepsilon_{fake} - \varepsilon_{fake}^2) N_{QCD}^l, \quad (5.21)$$

$$N^l = N_S^l + N_W^l + N_{QCD}^l \quad (5.22)$$

les solutions analytiques étant :

$$N_S^l = \frac{N^t - \varepsilon_{fake}(N^m + N^t - \varepsilon_{fake}N^l)}{(\varepsilon_s - \varepsilon_{fake})^2}, \quad (5.23)$$

$$N_W^l = \frac{(\varepsilon_s + \varepsilon_{fake})(N^m + N^t) - 2(N^t + \varepsilon_s \varepsilon_{fake} N^l)}{(\varepsilon_s - \varepsilon_{fake})^2}, \quad (5.24)$$

$$N_{QCD}^l = \frac{N^t - \varepsilon_s(N^m + N^t - \varepsilon_s N^l)}{(\varepsilon_s - \varepsilon_{fake})^2}. \quad (5.25)$$

5.4.3 Estimation de ε_s et ε_{fake} à partir des données

Les hypothèses faites précédemment peuvent être testées dans un premier temps à partir d'événements simulés mais doivent être vérifiées sur les données. L'efficacité ε_s peut être estimée à l'aide d'événements $Z \rightarrow l^+l^-$ en utilisant la méthode *tag & probe* [118] décrite dans la section 5.2.5.

L'efficacité ε_{fake} peut être évaluée à partir des données à l'aide d'événements multi-jets en demandant une faible énergie transverse manquante et/ou en utilisant les événements à faible multiplicité de jets (< 2 jets). Ainsi, nous considérons une région de l'espace de phase dominée par le bruit de fond *QCD* (faible contamination d'événements de type signal et de type $W + jets$).

Ces méthodes d'estimation des efficacités ε_s et ε_{fake} sont actuellement en cours de développement. Dans notre cas, les efficacités ε_s et ε_{fake} seront évaluées à partir de l'information générée uniquement. La sensibilité de la méthode à ces efficacités sera cependant estimée dans la section 5.4.5.2 et traitée comme une source d'incertitude systématique.

5.4.4 Méthode de résolution du système d'équations

Pour résoudre le système d'équations, une méthode matricielle est utilisée. Les équations (5.20)-(5.22) peuvent être réécrites comme $y = \mathbb{A}x$, où y et x sont deux vecteurs de dimensions 3 définis comme :

$$y = \begin{pmatrix} N^t \\ N^m \\ N^l \end{pmatrix}, \quad x = \begin{pmatrix} N_S^l \\ N_W^l \\ N_{QCD}^l \end{pmatrix}$$

et \mathbb{A} est une matrice 3×3 définie par :

$$\mathbb{A} = \begin{pmatrix} \varepsilon_S^{l \rightarrow t} & \varepsilon_W^{l \rightarrow t} & \varepsilon_{QCD}^{l \rightarrow t} \\ \varepsilon_S^{l \rightarrow m} & \varepsilon_W^{l \rightarrow m} & \varepsilon_{QCD}^{l \rightarrow m} \\ 1 & 1 & 1 \end{pmatrix}.$$

Les solutions de ce système d'équations sont alors simplement $x = \mathbb{A}^{-1}y$. Une méthode numérique implémentée dans Root [70] est utilisée pour inverser la matrice \mathbb{A} et calculer les solutions.

5.4.5 Test de la méthode

La méthode a été testée à l'aide des simulations Monte-Carlo décrites dans la section 5.2.1.2, uniquement dans le cas du canal $e - e$ pour les raisons décrites précédemment. Il est aussi à noter que les événements di-bosons et single top tW n'ont pas été pris en compte. Ces bruits de fond physiques n'affectent en principe que les efficacités $\varepsilon_S^{l \rightarrow t}$ et $\varepsilon_S^{l \rightarrow m}$ et peuvent conduire à de plus grandes incertitudes systématiques, mais ne devraient pas affecter la stabilité de la méthode elle-même. A cause du faible nombre d'événements QCD sélectionnés, les résultats souffrent de fluctuations statistiques. Un premier test peut cependant être réalisé pour évaluer la faisabilité de la méthode, à l'aide de quelques approximations supplémentaires. Notons que la méthode de la matrice a été paramétrisée en fonction de la multiplicité de jets ayant un $p_T > 30$ GeV/c.

5.4.5.1 Efficacités

La première étape de la méthode consiste à définir 3 échantillons correspondant aux sélections *loose*, *medium* et *tight*.

- **loose** : sélection sans critère d'isolation,
- **medium** : sélection avec au moins un lepton sélectionné passant les coupures d'isolation,
- **tight** : sélection avec au moins deux leptons sélectionnés passant les coupures d'isolation.

Pour un événement donné, le choix des candidats électrons (provenant de W) se fait toujours en favorisant les électrons passant les coupures d'isolation, s'il y en a. Les événements *tight* constituent ainsi un sous-ensemble des échantillons *medium* et *loose* avec des choix de candidats électrons strictement identiques. Il en est de même pour les événements *medium* constituant un sous-ensemble de l'échantillon *loose*. De cette manière, nous évitons les biais qui peuvent provenir d'un mauvais choix de candidats électrons.

Les sélections *loose*, *medium* et *tight* sont illustrées par les figures 5.13, 5.14 et 5.15 en fonction de la multiplicité de jets ayant une énergie transverse $p_T > 30$ GeV/c. Dans le cas de la sélection *tight*, aucun événement de type *QCD* n'est sélectionné.

Les efficacités de passer de la sélection *loose* à *medium* ($\varepsilon^{l \rightarrow m}$) et de la sélection *loose* à *tight* ($\varepsilon^{l \rightarrow t}$) sont représentées par les figures 5.16 et 5.17 respectivement. Comme aucun événement *QCD* ne passe les coupures de sélection *tight*, $\varepsilon_{QCD}^{l \rightarrow t}$ n'est pas représentée.

Ces efficacités devront être estimées à partir des données réelles. Dans la suite de cette analyse, l'information au niveau générateur est utilisée pour les déterminer. L'efficacité de sélection des vrais leptons isolés ainsi que les taux de faux leptons peuvent être estimés de différentes manières dans la simulation, soit en utilisant les efficacités représentées dans la figure 5.16 avec les équations (5.14) à (5.16) ou en extrayant les efficacités illustrées dans la figure 5.17 et utilisant les équations (5.17) à (5.19).

Sur le graphique de gauche de la figure 5.17, on observe que les efficacités ε_S peuvent être différentes entre les échantillons $t\bar{t}$ et Z +jets (à cause d'éventuels partons produits en association). Les efficacités ε_s seront estimées en utilisant l'échantillon de type signal présenté dans la figure 5.17 en prenant en compte les différences d'efficacités de sélection de leptons entre les événements $t\bar{t}$ signal et Z + jets et de la fraction relative des événements Z + jets.

$$\varepsilon_s^2 = \varepsilon_{t\bar{t}}^2 \frac{N_{t\bar{t}}}{N_{t\bar{t}} + N_{Z+jets}} + \varepsilon_{Z+jets}^2 \frac{N_{Z+jets}}{N_{t\bar{t}} + N_{Z+jets}}, \quad (5.26)$$

où $N_{t\bar{t}}$ et N_{Z+jets} sont les nombres d'événements $t\bar{t}$ signal et Z + jets sélectionnés, et $\varepsilon_{t\bar{t}}$ et ε_{Z+jets} sont les efficacités de sélection des vrais électrons isolés dans les événements $t\bar{t}$ signal et Z + jets.

Le taux de faux électrons isolés peut quant à lui être estimé de deux manières différentes :

- à partir de l'efficacité présentée dans la figure 5.16 et utilisant l'échantillon de type *QCD* et l'équation (5.16),
- à partir de l'efficacité présentée dans la figure 5.17 et utilisant l'échantillon de type W +jets et l'équation (5.18).

Puisque dans le premier cas il n'y a pas suffisamment d'événements après la sélection *medium*, le second cas va être utilisé par la suite pour déterminer ε_{fake} :

$$\varepsilon_{fake} = \frac{\varepsilon_W^{l \rightarrow t}}{\varepsilon_s}. \quad (5.27)$$

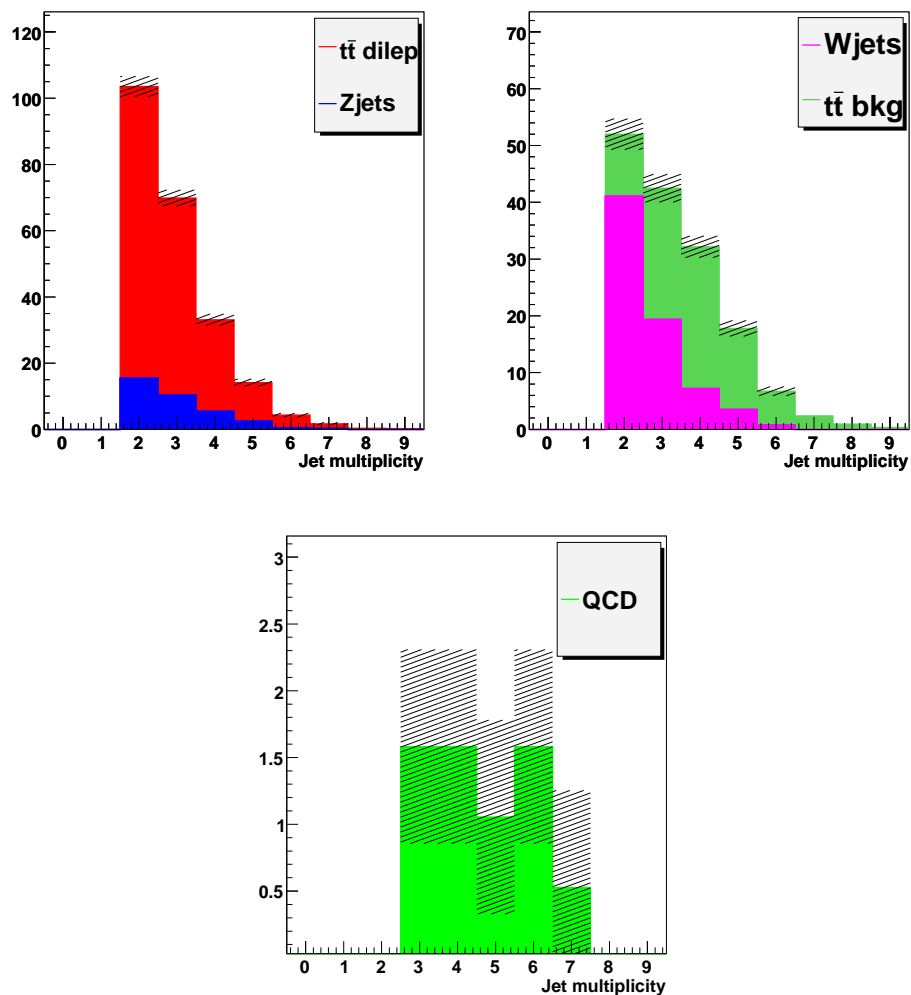


FIG. 5.13: Pour la sélection *loose*, distributions de multiplicité de jets pour les événements de type signal (en haut à gauche), de type W +jets (en haut à droite) et de type QCD (en bas). Les hachures correspondent aux erreurs statistiques du Monte-Carlo.

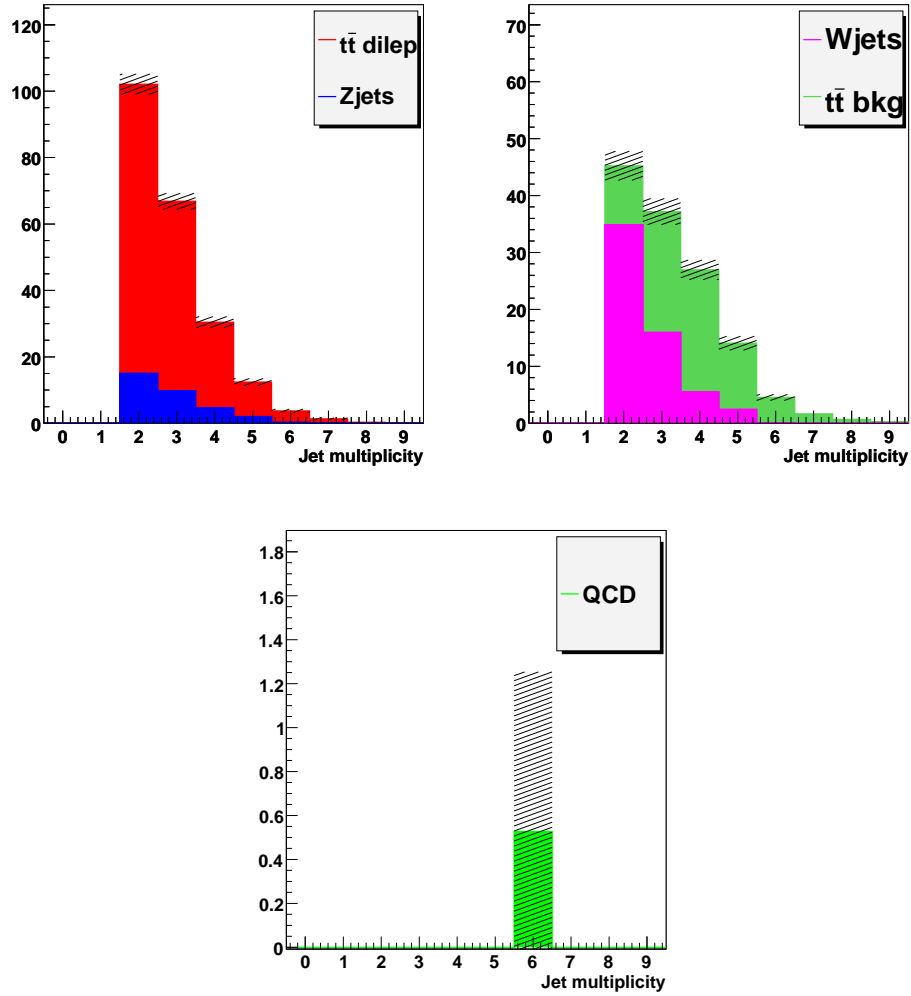


FIG. 5.14: Pour la sélection *medium*, distributions de multiplicité de jets pour les événements de type signal (en haut à gauche), de type W +jets (en haut à droite) et de type QCD (en bas). Les hachures correspondent aux erreurs statistiques du Monte-Carlo.

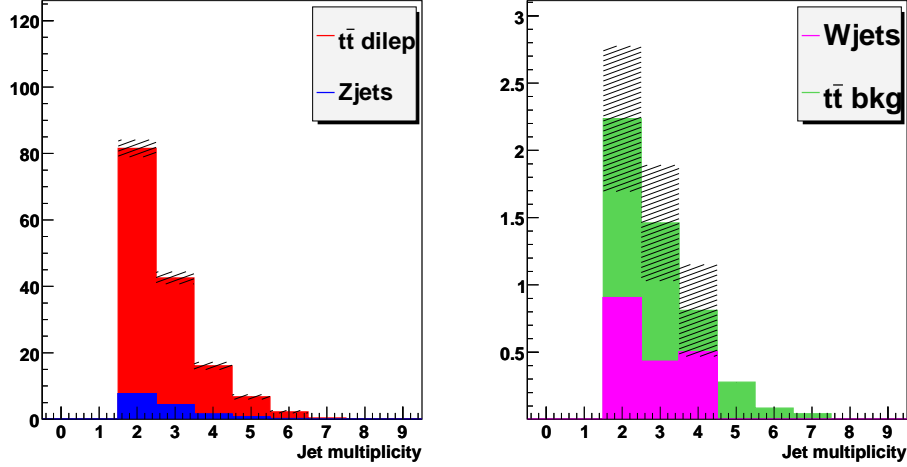


FIG. 5.15: Pour la sélection *tight*, distributions de multiplicité de jets pour les événements de type signal (à gauche), de type W +jets (à droite). Les hachures correspondent aux erreurs statistiques du Monte-Carlo. Aucun événements de type QCD ne passe les coupures.

De plus, dû au manque d'événements QCD Monte-Carlo et afin d'éviter des fluctuations statistiques trop importantes⁹, un même poids est appliqué sur chaque événement QCD qui passe la sélection. Ce poids est calculé comme la moyenne des poids de tous les événements. Cette modification affecte sensiblement l'allure des distributions de type QCD mais, comme la contamination QCD est faible, les résultats de la méthode ne changent pas de manière significative.

5.4.5.2 Erreurs systématiques

Les erreurs systématiques liées à la méthode de la matrice proviennent des erreurs statistiques sur le nombre d'événements sélectionnés dans les lots *loose*, *medium* et *tight* et de l'incertitude sur ε_s et ε_{fake} .

- **(1) Les erreurs statistiques.** Premièrement, le calcul des erreurs systématiques liées aux nombres d'événements sélectionnés est fait en définissant trois échantillons statistiquement indépendants :

$$N_1 = N^l - N_2 - N_3. \quad (5.28)$$

$$N_2 = N^m - N_3 \quad (5.29)$$

$$N_3 = N^t. \quad (5.30)$$

L'erreur est estimée en faisant varier les nombres d'événements N_1 , N_2 et N_3 suivant une distribution de Poisson.

⁹des poids très élevés sont appliqués à certains événements QCD produits avec un faible \hat{p}_T

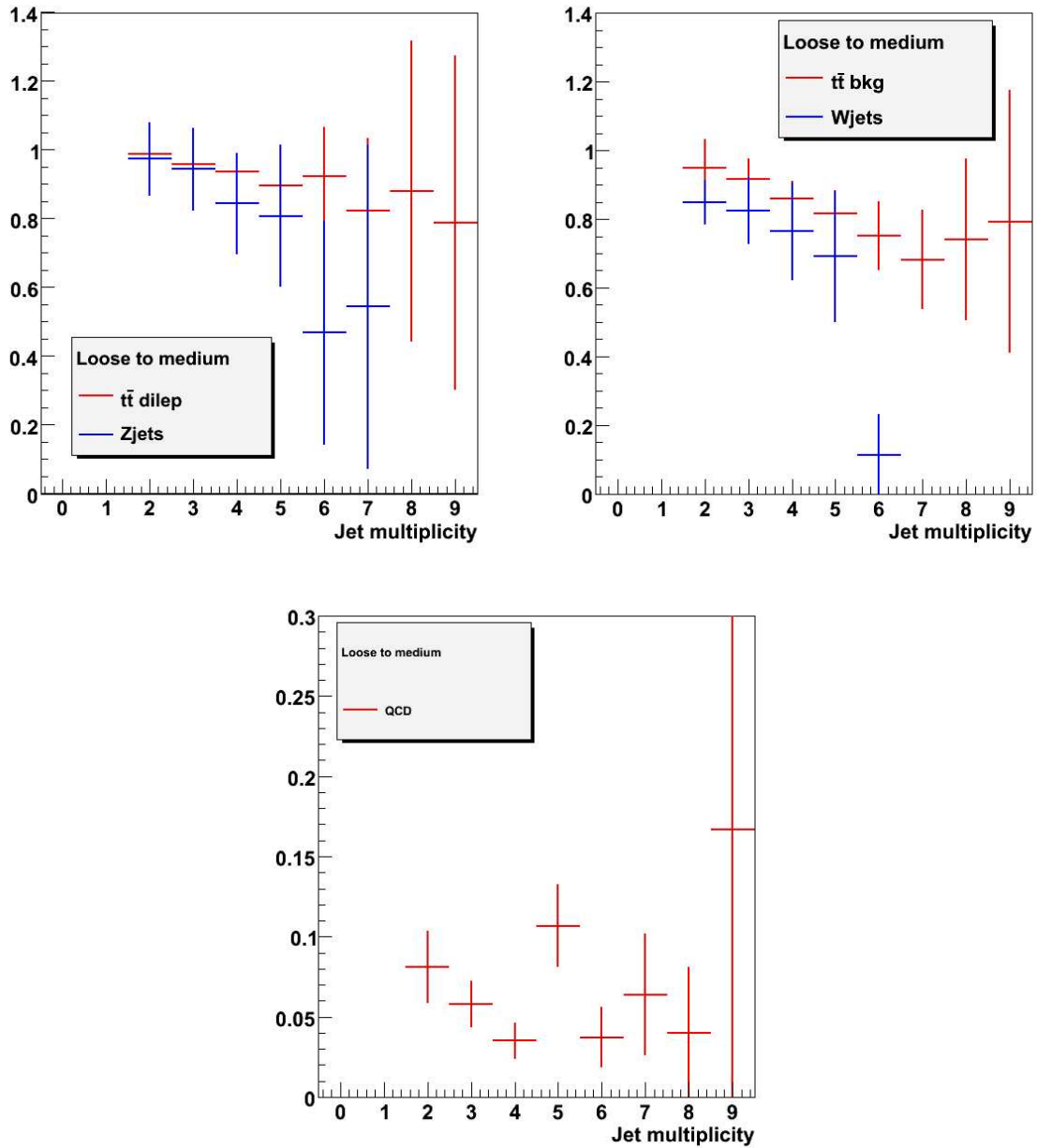


FIG. 5.16: Efficacité de passer de la sélection *loose* à la sélection *medium* $\varepsilon^{l \rightarrow m}$ pour les événements de type signal (en haut à gauche), de type W +jets (en haut à droite) et de type QCD (en bas).

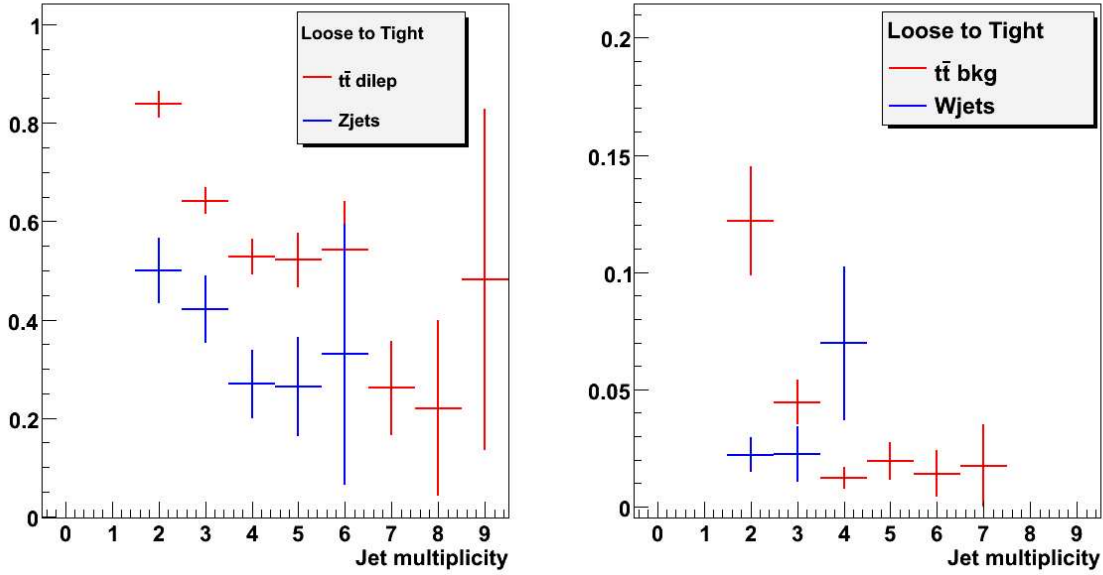


FIG. 5.17: Efficacité de passer de la sélection *loose* à la sélection *tight* $\varepsilon^{l \rightarrow t}$ pour les événements de type signal (à gauche) et de type W +jets (à droite).

- **(2) Les erreurs sur ε_s et ε_{fake} .** L'autre source principale d'erreur systématique est liée aux incertitudes sur les valeurs de ε_s et ε_{fake} . Nous faisons l'hypothèse que ces incertitudes sont connues et que les erreurs correspondantes suivent des distributions gaussiennes.

Les erreurs systématiques globales sont alors estimées à l'aide d'une méthode de pseudo-expériences. Le système d'équations (5.20)-(5.22) est résolu 3000 fois en faisant varier les nombres d'événements sélectionnés suivant une distribution de Poisson à l'aide des transformations (5.28)-(5.30) d'une part (erreur **(1)**) et en variant les efficacités ε_s et ε_{fake} suivant deux distributions gaussiennes de paramètres (μ_s, σ_s) et $(\mu_{fake}, \sigma_{fake})$ d'autre part (erreur **(2)**). μ_s (μ_{fake}) sont les efficacités de sélection de vrais leptons isolés (taux de faux leptons isolés) et σ_s (σ_{fake}) sont les incertitudes attendues. En pratique, les incertitudes sur les efficacités ε_s et ε_{fake} doivent être estimées à partir des données. Dans notre cas, nous faisons l'hypothèse que ε_s et ε_{fake} sont connues avec une précision de $\pm 5\%$ et $\pm 10\%$ respectivement. Cette procédure est appliquée pour chaque multiplicité de jets ayant un p_T supérieur à 30 GeV/c.

Les distributions N_s^l , N_W^l et N_{QCD}^l des solutions des 3000 pseudo-expériences sont présentées sur les figures 5.18 et 5.19 pour les événements de type *signal*, W +jets et QCD et pour les échantillons *loose* et *tight*. Les distributions en noir correspondent aux erreurs statistiques **(1)** et en rouge aux erreurs sur ε_s et ε_{fake} **(2)**. Les écarts types de ces distributions sont pris comme erreurs systématiques tandis que les moyennes

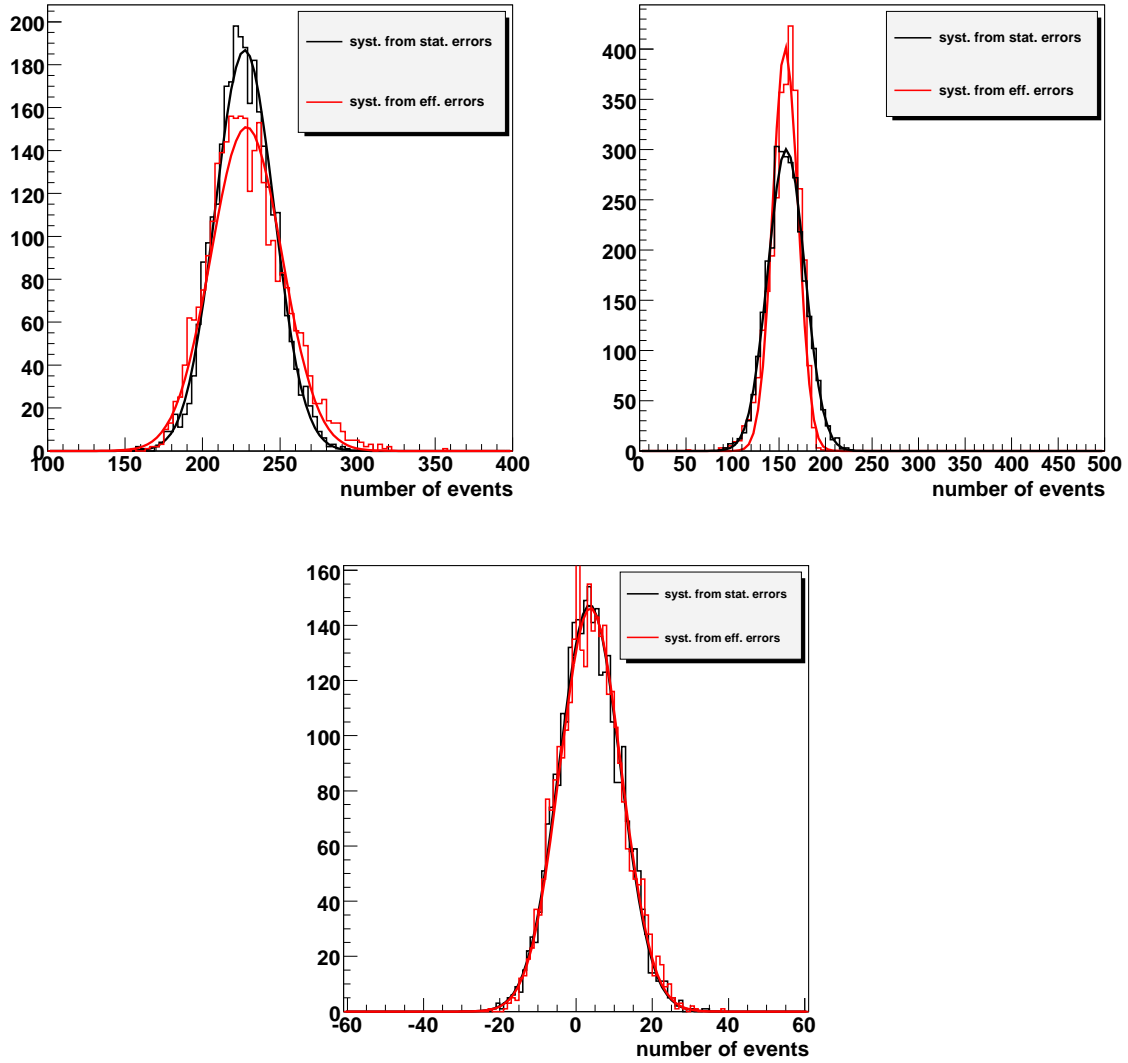


FIG. 5.18: Distributions des 3000 solutions de la méthode de la matrice pour 3000 pseudo-expériences dans le cas de l'échantillon *loose*, pour les événements de type signal (en haut à gauche), W +jets (en haut à droite) et QCD (en bas).

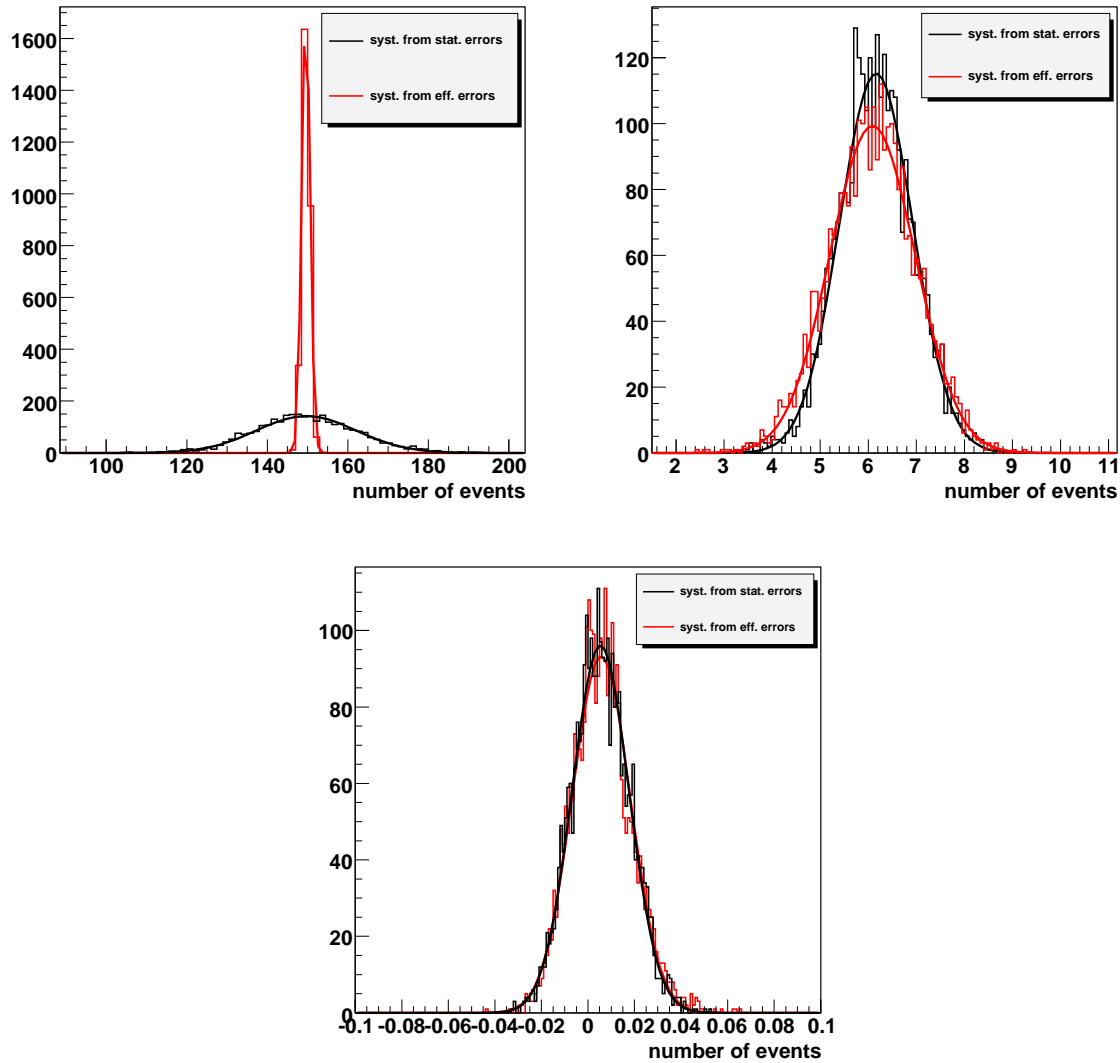


FIG. 5.19: Distributions des 3000 solutions de la méthode de la matrice pour 3000 pseudo-expériences dans le cas de l'échantillon *tight*, pour les événements de type signal (en haut à gauche), W +jets (en haut à droite) et QCD (en bas).

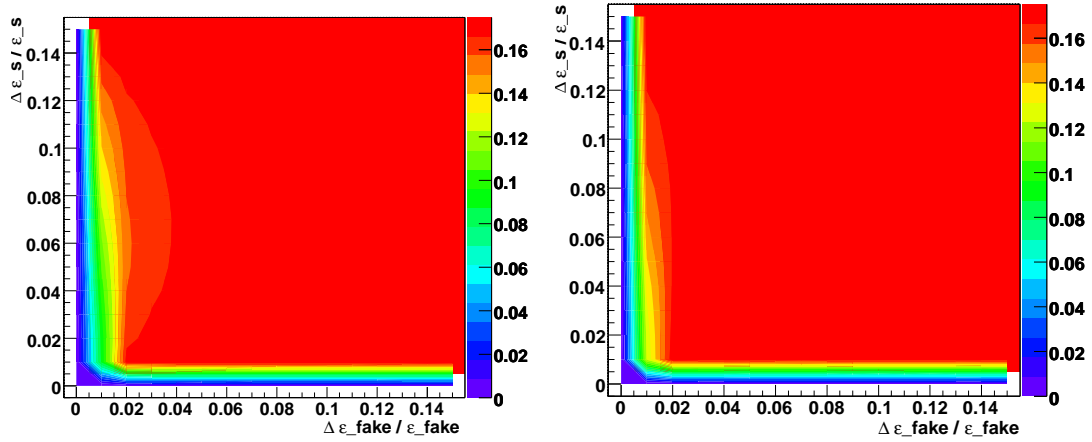


FIG. 5.20: Le graphique de gauche représente la distribution en trois dimensions des incertitudes relatives du taux de faux électrons isolés (axe des x), de l'efficacité de sélection de vrais électrons isolés (axe de y) et du nombre d'événements *loose* prédit par la méthode de la matrice (axe z des couleurs) pour les événements de type signal. Le graphique de droite représente les mêmes quantités pour les événements de type W +jets.

correspondent à la prédiction de la Méthode de la Matrice. La variation du nombre attendu d'événements de type *signal* et W +jets, dans le cas de la sélection *loose* en fonction de la variation relative de ε_s et ε_{fake} est illustrée par la figure 5.20.

Il est à noter que dans le graphique en haut à gauche sur la figure 5.19, la distribution de N_S^l (en rouge) est très étroite. Cela est dû à des effets de corrélation entre N_S^l et ε_s qui apparaissent quand les transformations (5.17) sont appliquées. Les corrélations entre le nombre d'événements attendus dans l'échantillon *loose* et les efficacités ε_s et ε_{fake} sont illustrées sur les figures 5.21 à 5.23. Le facteur de corrélation entre N_S^l et ε_s (cf. figure 5.21) est négatif et proche de -1. Cela conduit à une contribution négative dans la formule de propagation des erreurs de l'équation (5.17) et par conséquent réduit significativement l'incertitude sur le nombre d'événements de type *signal*. A l'opposé, le facteur de corrélation entre N_W^l et ε_s est proche de 1, élargissant les incertitudes systématiques sur le nombre d'événements de type W +jets.

5.4.5.3 Résultats

La figure 5.24 représente les distributions en multiplicité de jets ayant une énergie transverse $E_T > 30$ GeV dans le cas de la sélection *loose*. Les nombres d'événements prédits par la méthode de la matrice et ceux réellement présents dans l'échantillon sont comparés. Ces derniers sont déterminés à l'aide de l'information générée, que nous appellerons *vérité Monte-Carlo*. On observe un très bon accord entre les solutions de la méthode de la matrice (point avec barres d'erreur) et la vérité Monte-Carlo dans

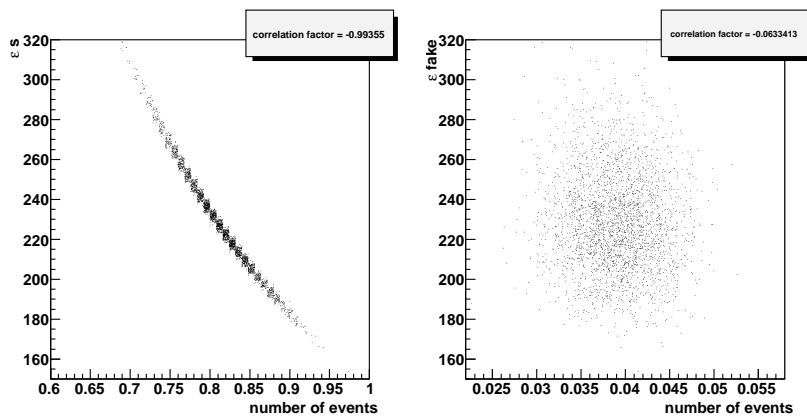


FIG. 5.21: Corrélations entre N_S^l et ε_s (à gauche) et ε_{fake} (à droite). Les facteurs de corrélation sont de -0.99 (à gauche) et -0.06 (à droite).

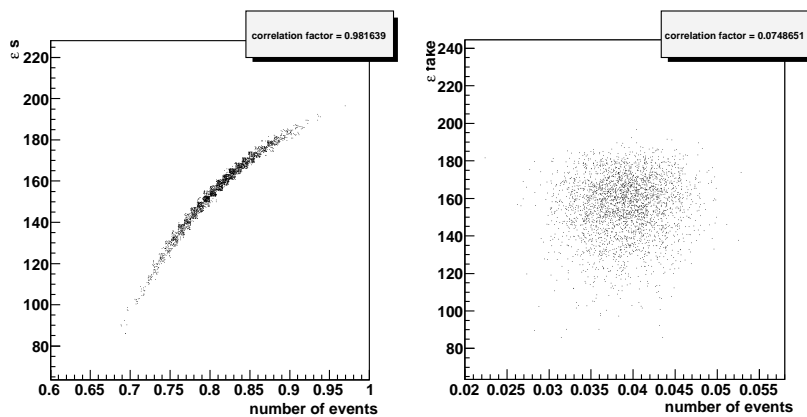


FIG. 5.22: Corrélations entre N_W^l et ε_s (à gauche) et ε_{fake} (à droite). Les facteurs de corrélation sont de 0.98 (à gauche) et 0.07 (à droite).

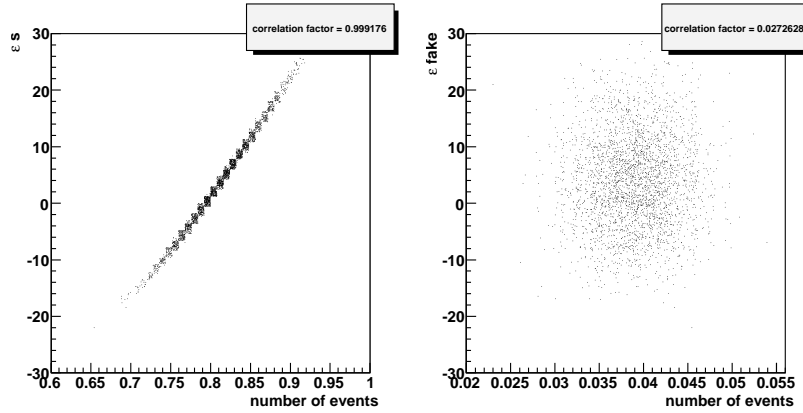


FIG. 5.23: Corrélations entre N_{QCD}^l et ε_s (à gauche) et ε_{fake} (à droite). Les facteurs de corrélation sont de 0.99 (à gauche) et 0.03 (à droite).

le cas des échantillons de type *signal* et W +jets. L'échantillon de type *QCD* souffre d'un manque de statistique, même si les prédictions restent compatibles dans les barres d'erreurs avec la vérité Monte-Carlo.

La figure 5.25 présente une comparaison entre la prédiction de la méthode de la matrice et la vérité Monte-Carlo dans le cas de la sélection *medium*. Les solutions de la méthode de la matrice (points avec barres d'erreur) sont déduites des distributions présentées dans la figure 5.24 en utilisant les équations (5.14) à (5.16). Dans ce cas, il y a également un assez bon accord entre les prédictions de la méthode de la matrice et la vérité Monte-Carlo pour les événements de type *signal* et de type W + jets. Il y a encore une fois de larges fluctuations statistiques dans le cas des événements de type *QCD*.

La figure 5.26 présente les prédictions de la méthode de la matrice, correspondant aux résultats de la figure 5.24 multipliés par les efficacités (5.17), (5.18) et (5.19) pour les événements de type *signal*, W + jets et *QCD*. Une fois encore, il y a un très bon accord entre les prédictions de la méthode de la matrice et la vérité Monte-Carlo dans le cas des événements de type *signal*. Dans le cas des événements de types W +jets, comme les erreurs statistiques sont relativement élevées, il est assez difficile de conclure même si les résultats sont compatibles dans les barres d'erreurs. Dans le cas des événements de type *QCD*, le nombre d'événements prédit est compatible avec 0.

Le nombre d'événements attendus sommé pour toutes les multiplicités de jets ainsi que les erreurs correspondantes sont présentés dans le tableau 5.12. Dans le cas de l'échantillon de type *signal*, la source principale d'erreur est l'incertitude sur le nombre d'événements sélectionnés dans les échantillons *loose*, *medium* et *tight*. Tandis que pour l'échantillon W +jets la source principale d'erreur provient des incertitudes sur ε_s et ε_{fake} .

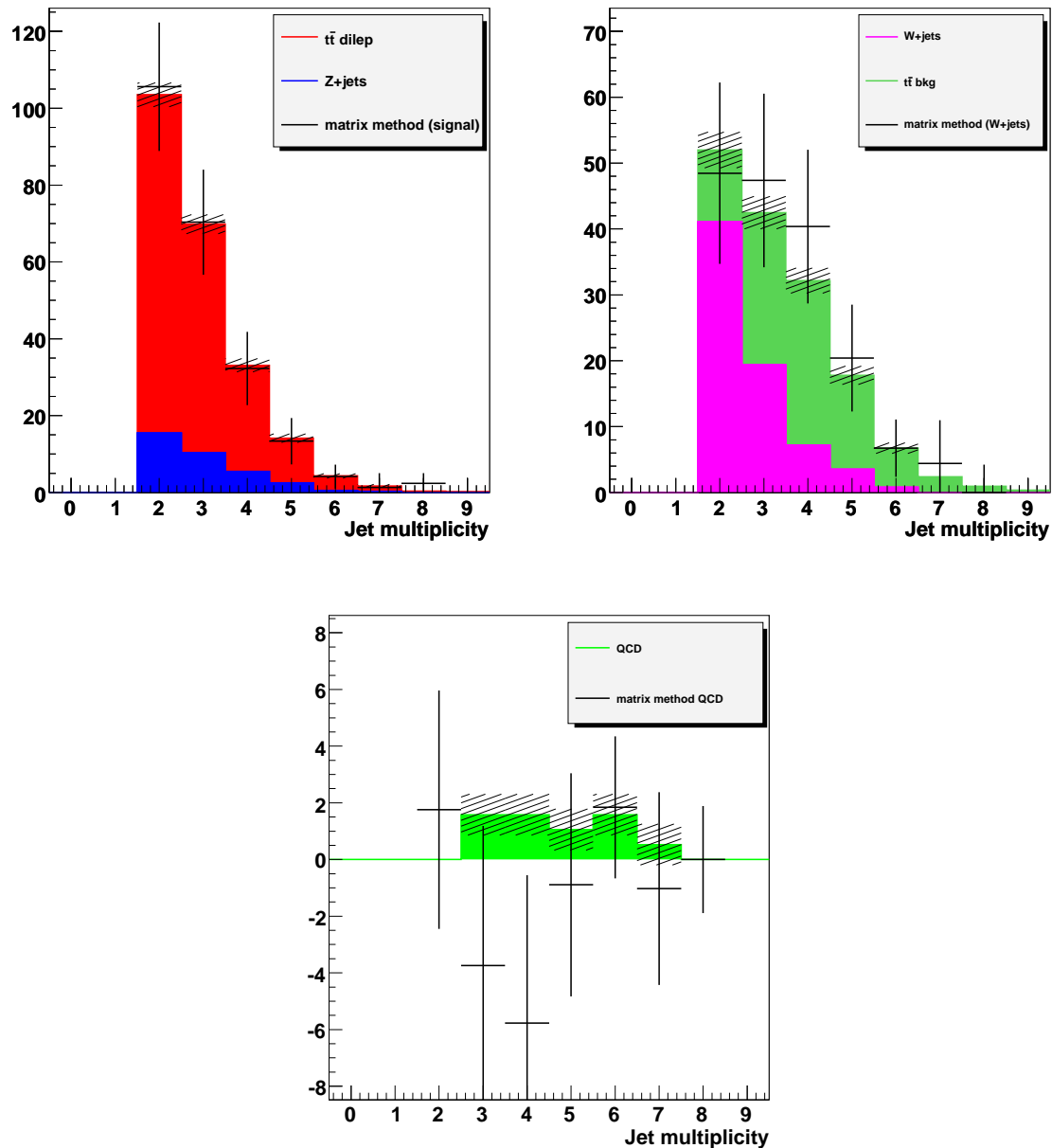


FIG. 5.24: Pour la sélection *loose*, multiplicité de jets pour les événements de type signal (en haut à gauche), W +jets (en haut à droite) et QCD (en bas). Les hachures correspondent aux erreurs statistiques du Monte-Carlo. Les points correspondent aux prédictions de la méthode de la matrice, les erreurs statistiques et systématiques sont représentées par les barres d'erreur.

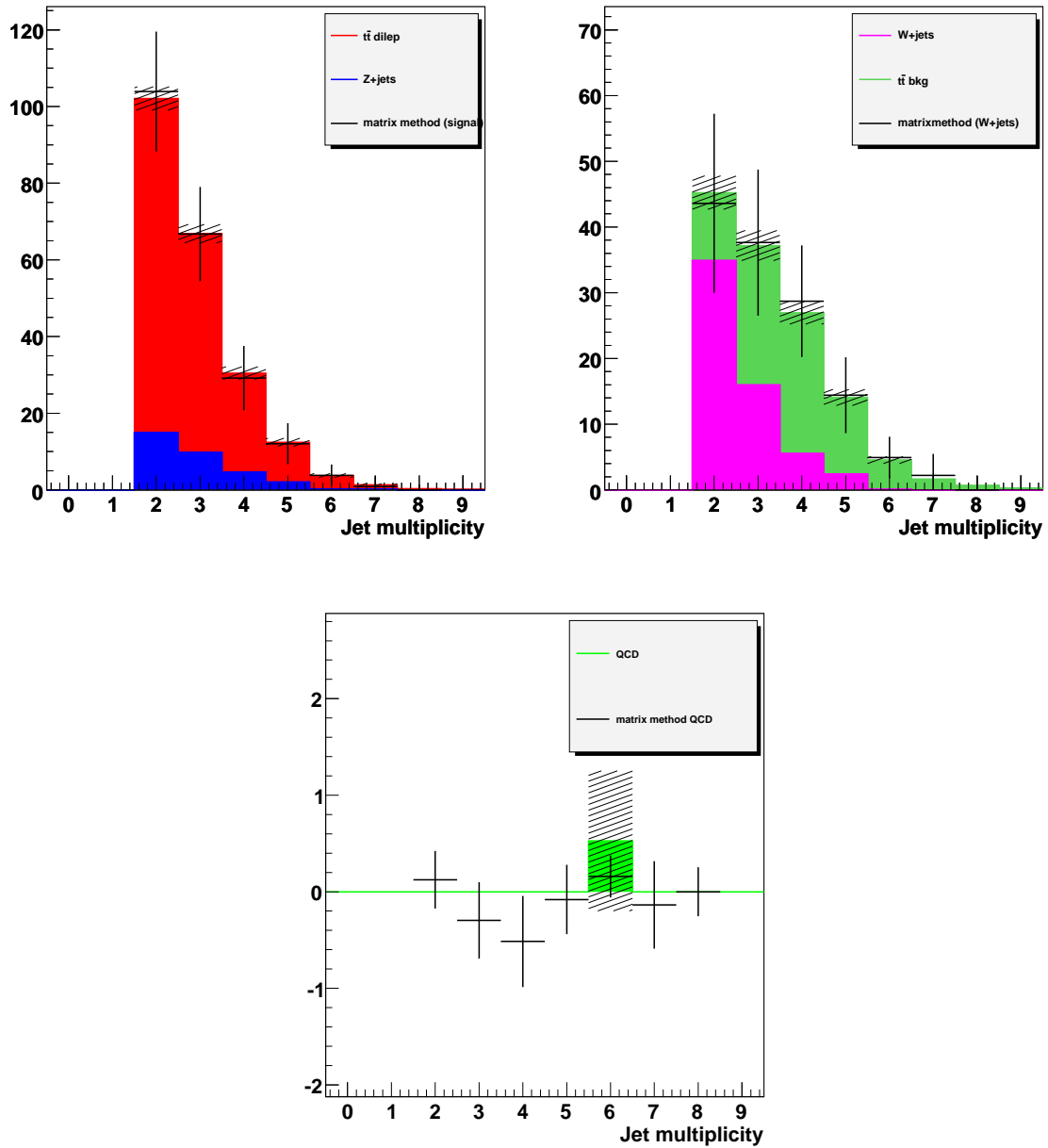


FIG. 5.25: Pour la sélection *medium*, multiplicité de jets pour les événements de type signal (en haut à gauche), W +jets (en haut à droite) et QCD (en bas). Les hachures correspondent aux erreurs statistiques du Monte-Carlo. Les points correspondent aux prédictions de la méthode de la matrice, les erreurs statistiques et systématiques sont représentées par les barres d'erreur.

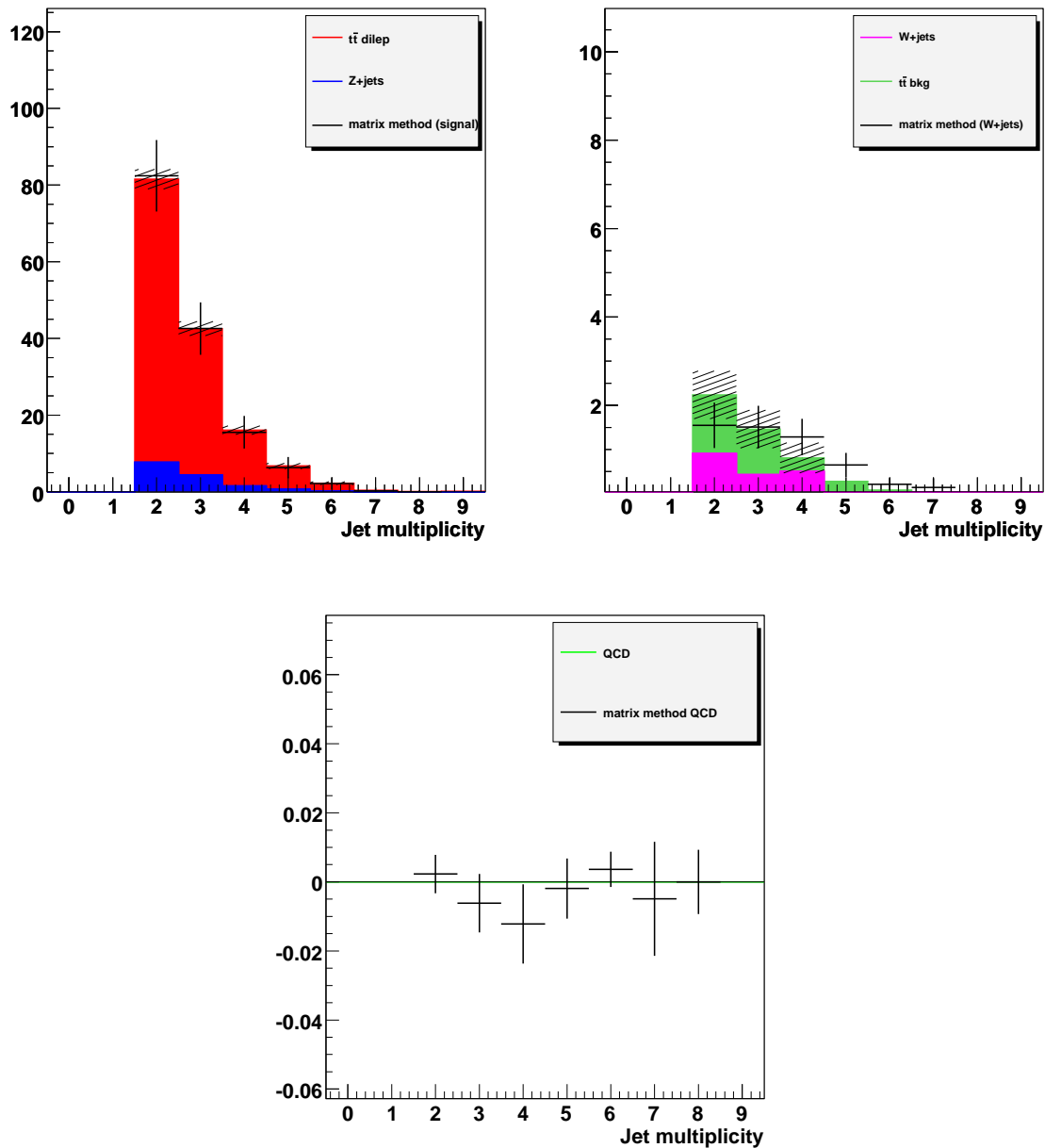


FIG. 5.26: Pour la sélection *tight*, multiplicité de jets pour les événements de type signal (en haut à gauche), W +jets (en haut à droite) et QCD (en bas). Les hachures correspondent aux erreurs statistiques du Monte-Carlo. Les points correspondent aux prédictions de la méthode de la matrice, les erreurs statistiques et systématiques sont représentées par les barres d'erreur.

La méthode a également été testée pour différentes sélections. Les tableaux 5.13 à 5.15 montrent les nombres d'événements et leurs erreurs pour chaque type d'échantillons de type *signal*, *W+jets* et *QCD* calculés comme décrit dans la section 5.4.5.2 inclusivement (sans paramétrisation en multiplicité de jets) pour différentes sélections :

- sélection standard sans coupure d'identification des jets *b* ni sur la masse invariante di-électrons (tableau 5.13),
- sans coupure d'identification des jets *b* ni sur la \cancel{E}_T (tableau 5.14),
- sans coupure d'identification des jets *b*, ni sur la masse invariante di-électrons ni sur l'énergie transverse manquante (tableau 5.15)).

Même en relâchant les coupures de sélection, les prédictions de la méthode de la matrice restent compatibles avec la vérité Monte-Carlo avec des incertitudes raisonnables. Les allures des distributions en multiplicité de jets ne sont cependant pas bien décrites car le bruit de fond *QCD* souffre de fluctuations statistiques.

	Type Signal	Type <i>W+jets</i>	Type <i>QCD</i>
Nbr. d'événements	149.1	5.3	-0.019
Erreur stat.	12.7	0.7	0.012
Erreur $\varepsilon_s, \varepsilon_{fake}$	0.9	0.9	0.013
Erreur totale	12.7	1.1	0.018
Nbr. d'événements (MC)	149.6 ± 3.5	4.9 ± 0.8	0

TAB. 5.12: Nombre d'événements prédits par la méthode de la matrice et erreurs systématiques liées aux erreurs statistiques des lots *loose*, *medium* et *tight* et aux erreurs sur ε_s et ε_{fake} . Le nombre d'événements "vrais" déterminé par l'information du générateur est également indiqué.

	Type Signal	Type <i>W+jets</i>	Type <i>QCD</i>
Nbr. d'événements	305	7.1	-0.029
Erreur stat.	17.9	0.9	0.015
Erreur $\varepsilon_s, \varepsilon_{fake}$	1.8	1.9	0.014
Erreur totale	18.0	2.1	0.020
Nbr. d'événements MC	305.8 ± 6.0	6.5 ± 0.9	0

TAB. 5.13: Identique au tableau 5.12 lorsque la coupure en masse invariante n'est pas appliquée.

Pour conclure, une méthode d'estimation de bruit de fond instrumental a été mise au point et testée avec les échantillons disponibles. Les résultats semblent plus qu'encourageant. Une telle méthode devrait pouvoir être mise en place au démarrage du LHC et être utilisée pour une luminosité de quelques centaines de pb^{-1} . Pour des luminosités

	Type Signal	Type W +jets	Type QCD
Nbr. d'événements	397.1	12.5	4.0
Erreur stat.	20.9	1.5	0.116
Erreur $\varepsilon_s, \varepsilon_{fake}$	21.7	3.7	0.708
Erreur totale	30.1	4.0	0.717
Nbr. d'événements MC	404.3 ± 7.2	8.9 ± 1.1	0

TAB. 5.14: Identique au tableau 5.12 lorsque la coupure en \cancel{E}_T n'est pas appliquée.

	Type Signal	Type W +jets	Type QCD
Nbr. d'événements	2508.0	16.0	4.783
Erreur stat.	51.4	2.9	0.139
Erreur $\varepsilon_s, \varepsilon_{fake}$	16.7	8.2	0.848
Erreur totale	54.1	8.8	0.860
Nbr. d'événements MC	2517.0 ± 17.5	12.2 ± 1.3	0

TAB. 5.15: Identique au tableau 5.12 lorsque ni la coupure en masse invariante ni la coupure en \cancel{E}_T ne sont appliquées.

plus élevées, la précision de la méthode de la matrice devrait s'améliorer de manière significative, ceci grâce à l'augmentation de la précision des mesures d'efficacités de sélection des leptons ainsi qu'à l'augmentation du nombre d'événements disponibles.

Une source d'erreur reste cependant à étudier : un biais peut provenir des critères d'isolation appliqués au niveau du système de déclenchement. Un tel biais est supposé négligeable dans notre cas mais devra être estimé précisément dans de futures études.

5.5 Estimation du bruit de fond Z +jets

Dans le cas d'une sélection n'utilisant pas de critères d'identification des jets b , il est possible d'estimer le nombre d'événements de bruit de fond Z +jets à partir des données à l'aide d'un ajustement de fonctions sur la distribution de masse invariante des candidats di-leptons. En effet, cette distribution (graphique en haut à droite de la figure 5.9) correspond à la superposition d'une distribution polynomiale (contenant principalement des événements de signal) et d'une distribution piquée autour de la masse du boson Z décrite par une distribution de *Breit – Wigner*. La figure 5.27 représente l'ajustement d'un polynôme d'ordre 3 et d'une fonction de *Breit – Wigner* sur les distributions en masse invariante, pour le canal $e - e$ à gauche et le canal $\mu - \mu$ à droite.

L'idée est de déterminer l'allure de la distribution en masse invariante di-leptons dans un échantillon très pur en événements Z +jets, puis d'extrapoler ce résultat à la sélection utilisée. Dans ce lot, on ne peut cependant pas ajuster la distribution en masse

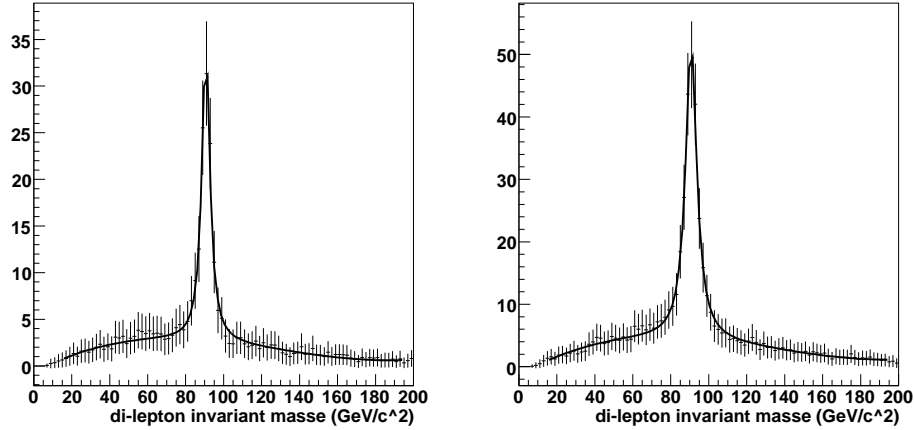


FIG. 5.27: Distributions en masse invariante de candidats di-leptons après la sélection présentée dans la section 5.2.6 sans identification de jets b . La fonction d'ajustement utilisée correspond à l'addition d'un polynôme d'ordre 3 et d'une fonction de *Breit – Wigner* pour le canal $e - e$ (à gauche) et le canal $\mu - \mu$ (à droite).

invariante du Z par une fonction de *Breit – Wigner*. En effet, une telle fonction ne prend pas en compte les effets de résolution en angle et en impulsion des leptons. La distribution en masse invariante du pic du Z est également asymétrique. Ceci est dû à la présence d'événements Drell-Yan (γ) pour les basses masses invariantes. Pour ces raisons, les queues de distribution du Z ne sont par bien décrites par une fonction de *Breit – Wigner* seule. L'ajustement d'une fonction de *Breit – Wigner* et une fonction de Cruiff [119] a été testé. La fonction de Cruiff est définie comme :

$$f(x) = N \times \exp\left(-\frac{(x - m)^2}{2\sigma_{d/g}^2 + \alpha_{d/g}(x - m)^2}\right), \quad (5.31)$$

où σ_d^2 et α_d (σ_g^2 et α_g) correspondent au cas $x - m < 0$ ($x - m > 0$) où m est la valeur moyenne. Une telle fonction ne peut cependant pas être ajustée directement sur l'ensemble des événements signal et bruits de fond de la figure 5.27 pour déterminer les queues de distribution. Il est en effet difficile de dissocier le pic du Z et le signal (+autres bruit de fond). Les queues contenant le signal et les autres bruits de fond sont ajustés par la fonction de Cruiff et non plus par le polynôme. C'est pourquoi un lot d'événements enrichis en événements Z +jets est utilisé.

La méthode utilisée pour estimer le bruit de fond Z +jets consiste alors en :

1. l'ajustement de la fonction "*Cruiff + Breit – Wigner*" sur un lot d'événements enrichi en Z . La sélection des événements présentée dans la section 5.2.6 est utilisée avec $\cancel{E}_T < 20$ GeV,

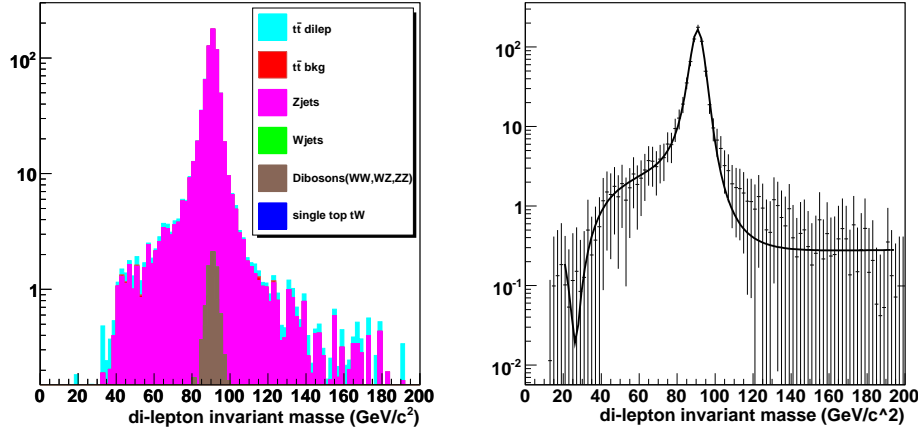


FIG. 5.28: Pour le canal $(e - e)$, distributions de la masse invariante pour tous les événements (signal et bruit de fond) sommés avant (à gauche) et après (à droite) l’ajustement d’une fonction de type “*Cruijff + Breit - Wigner*”.

2. la renormalisation de cette fonction à la quantité d’événements attendus pour la sélection avec $\cancel{E}_T > 50$ GeV,
3. l’intégration de la fonction hors de la fenêtre de masse invariante $[75, 105]$ GeV/ c^2 .

Nous faisons l’hypothèse qu’à basse \cancel{E}_T , l’échantillon est dominé par les événements Z +jets (faible contamination d’événements de signal et des autres événements de bruit de fond, en particulier le bruit de fond QCD), et que la résolution des leptons est la même qu’à grande \cancel{E}_T . Les graphiques de gauche des figures 5.28 et 5.29 représentent les distributions en masse invariante pour la sélection standard mais en demandant $\cancel{E}_T < 20$ GeV, sans identification des jets b , pour les canaux $e - e$ et $\mu - \mu$ respectivement. On vérifie ici que les événements sélectionnés sont en grande majorité des événements Z +jets. Il existe cependant une faible contamination d’événements WZ et ZZ mais qui ne modifie a priori que peu l’allure de la distribution en masse invariante du Z .

Les graphiques de droite des figures 5.28 et 5.29 représentent l’ajustement de la fonction “*Cruijff + Breit - Wigner*” sur les distributions de masse invariante de l’échantillon “ $\cancel{E}_T < 20$ GeV” pour les canaux $(e - e)$ et $(\mu - \mu)$ respectivement. Cet ajustement décrit correctement les distributions de masse invariante excepté pour la queue de distribution à grande masse et à très basse masse, bien qu’il reste compatible avec les barres d’erreurs. Cette différence entre la distribution et l’ajustement conduira toutefois à des erreurs relativement importantes. Aucune masse invariante n’est reconstruite en dessous de 40 GeV/ c^2 à cause d’une coupure au niveau du générateur.

On peut alors mettre en place une procédure de renormalisation de cette fonction au nombre d’événements sélectionnés dans l’échantillon $\cancel{E}_T > 50$ GeV. Dans un premier temps, pour chaque canal, un histogramme est rempli avec 10 000 tirages aléatoires sui-

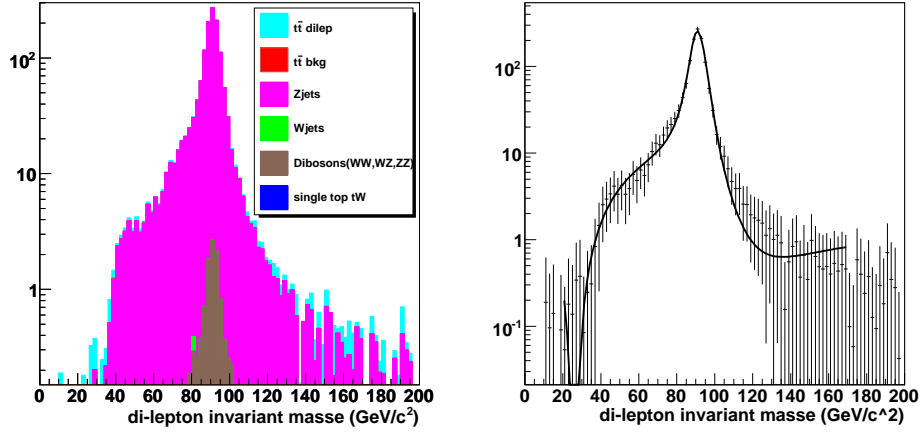


FIG. 5.29: Pour le canal $(\mu - \mu)$, distribution de la masse invariante pour tous les événements (signal et bruit de fond) sommés avant (à gauche) et après (à droite) l'ajustement d'une fonction de type "Cruijff + Breit - Wigner".

avant la distribution "Cruijff + Breit - Wigner" ajustée sur l'échantillon correspondant à la sélection $\cancel{E}_T < 20$ GeV. Nous nommerons cet histogramme h_{rand} . Le maximum de h_{rand} est déterminé, il est noté $Max_{rand}^{hist.}$.

Dans un second temps, la distribution de l'échantillon $\cancel{E}_T > 50$ GeV est ajustée par des fonctions de type (polynôme + Breit - Wigner). Le maximum de la fonction de Breit - Wigner est noté $Max_{fit}^{func.}$.

h_{rand} est alors renormalisé en utilisant le rapport entre $Max_{fit}^{func.}$ et $Max_{rand}^{hist.}$. Il est à noter que dans cette procédure, les histogrammes doivent avoir le même échantillonnage.

Le nombre d'événements bruit de fond Drell-Yan est alors estimé en intégrant les queues de distribution de h_{rand} dans les intervalles masses invariantes $[20;75]$ et $[105;180]$ GeV/ c^2 . Afin d'éviter les erreurs possibles dues à la largeur des bins de h_{rand} lors de l'intégration des queues et pour faciliter le calcul des erreurs, h_{rand} est utilisé pour l'ajustement d'une nouvelle fonction "Cruijff + Breit - Wigner". Le nombre d'événements de bruit de fond Drell-Yan correspond alors à l'intégrale de cette fonction hors de la fenêtre de masse invariante $[75,105]$ GeV. Les sources principales d'incertitudes sont :

- l'incertitude sur les paramètres d'ajustement,
- l'incertitude sur le facteur de renormalisation,
- les différences de résolution des leptons entre les échantillons $\cancel{E}_T < 20$ GeV et $\cancel{E}_T > 50$ GeV.

En effet, les leptons mal reconstruits sont susceptibles de générer de la fausse \cancel{E}_T et de dégrader la reconstruction en masse invariante. Il peut donc y avoir une corrélation relativement importante entre \cancel{E}_T et masse invariante. Une source d'erreur supplémentaire qu'il faudra prendre en compte dans de futures analyses provient des bornes d'intégration. Cette erreur est négligée.

Les incertitudes sur les paramètres d’ajustement sont propagées au calcul des intégrales par le programme d’ajustement (Root [70]) utilisé. Les incertitudes restantes sont estimées à partir de la simulation Monte-Carlo. L’incertitude sur le facteur de renormalisation $Max_{fit}^{func.}/Max_{rand}^{hist.}$ est estimée en comparant $Max_{fit}^{func.}$ et le maximum réel de la distribution en masse invariante des événements Z à l’aide de la vérité Monte-Carlo. Les incertitudes liées aux différences de résolution de l’impulsion des leptons entre les échantillons $\cancel{E}_T < 20$ GeV et $\cancel{E}_T > 50$ GeV sont estimées en comparant les nombres d’événements prédits par la méthode et le nombre d’événements obtenus en intégrant une fonction “*Cruiff + Breit – Wigner*” ajustée sur la distribution de masse invariante des événements Z +jets ($\cancel{E}_T > 50$ GeV) uniquement à l’aide de la vérité Monte-Carlo. Ces différentes erreurs seront estimées dans le Monte-Carlo et propagées aux données réelles.

canal	prédiction	vérité MC	paramètre	facteur d’échelle	résolution	total
$(e - e)$	12.3	12.6	27.3 %	0.9 %	12.8 %	30 %
$(\mu - \mu)$	33.0	36.2	17.6 %	2.6 %	5.8 %	19%

TAB. 5.16: Prédiction de la méthode, vérité Monte-Carlo et incertitudes systématiques relatives sur les paramètres d’ajustement, le facteur d’échelle et la résolution sur l’impulsion des leptons pour les canaux $e - e$ et $\mu - \mu$.

Les prédictions de la méthode et les erreurs systématiques associées sont présentées dans le tableau 5.16. Les bruits de fond Z +jets dans les canaux $e - e$ et $\mu - \mu$ peuvent être prédits avec une incertitude d’environ 30% et 19% respectivement. On observe que l’erreur systématique est plus grande pour le canal $e - e$. Cela est dû à de plus grandes incertitudes sur l’estimation des paramètres d’ajustement, mais aussi à une plus grande sensibilité aux effets de résolution.

Cette méthode, bien que prometteuse, souffre de problèmes liés à la convergence de l’ajustement de la fonction “*Cruiff + Breit – Wigner*” qui se révèle sensible aux paramètres initiaux. De plus, les incertitudes liées à la résolution sur l’impulsion des leptons peut être sous-évaluées car estimées à partir du Monte-Carlo. Cette méthode reste néanmoins un moyen d’estimer le bruit de fond Drell-Yann et de comparer les données et le Monte-Carlo.

5.6 Stratégie de mesure de la section efficace

Plusieurs scénarios de sélection sont étudiés et peuvent conduire à différentes puretés d’échantillons et différentes précisions sur la mesure de la section efficace :

1. lorsqu’aucun critère d’identification de jets b n’est appliqué,
2. lorsqu’au moins un jet sélectionné passe le critère d’identification des jets b ,
3. lorsqu’au moins deux jets sélectionnés passent le critère d’identification des jets b .

Comme la sélection conduit à un échantillon très pur en événements de signal, même lorsqu'aucun critère d'identification des jets b n'est appliqué, il n'y a pas lieu d'utiliser de méthodes statistiques élaborées pour mesurer la section efficace dans les différents canaux. Une simple méthode de comptage d'événements est utilisée.

La section efficace d'un canal est alors calculée comme :

$$\sigma_{t\bar{t}} \times BR = \frac{N_{sel} - N_{bdf}}{\varepsilon_{t\bar{t}} \times \mathcal{L}} \quad (5.32)$$

où N_{sel} est le nombre total d'événements sélectionnés, N_{bdf} est le nombre d'événements de bruit de fond sélectionnés, $\varepsilon_{t\bar{t}}$ est l'efficacité de sélection des événements de signal, \mathcal{L} est la luminosité intégrée et BR est le rapport d'embranchement du canal étudié.

L'efficacité de sélection peut s'écrire : $\varepsilon_{t\bar{t}} = \varepsilon_{t\bar{t}}^{HLT} \times \varepsilon_{t\bar{t}}^{MC}$, où le terme HLT se réfère à l'efficacité de sélection du trigger et MC à l'efficacité du reste de la sélection.

Sélection $e - e$	S/B	efficacité du signal $t\bar{t}$	erreur statistique
1) 0 jet b	5.1	9.9%	$\pm 9.4 \%$
2) 1 jet b	9.0	8.3%	$\pm 9.9 \%$
3) 2 jets b	16.9	3.7%	$\pm 14.6 \%$

TAB. 5.17: Pour le canal $e - e$, rapport signal sur bruit, efficacité de sélection des événements de signal et incertitude statistique correspondante pour une luminosité intégrée de 100 pb^{-1} .

Sélection $e - \mu$	S/B	efficacité du signal $t\bar{t}$	erreur statistique
1) 0 jet b	9.9	20.4%	$\pm 4.5\%$
2) 1 jet b	16.7	17.0%	$\pm 4.8\%$
3) 2 jets b	35.6	7.5%	$\pm 7.2\%$

TAB. 5.18: Identique au tableau 5.17 pour le canal $e - \mu$.

Sélection $\mu - \mu$	S/B	efficacité du signal $t\bar{t}$	erreur statistique
1) 0 jet b	4.1	16.8%	$\pm 7.5 \%$
2) 1 jet b	9.5	14.2%	$\pm 7.7 \%$
3) 2 jets b	21.4	6.6%	$\pm 10.9\%$

TAB. 5.19: Identique au tableau 5.17 pour le canal $\mu - \mu$.

Une estimation de la précision statistique (pour une luminosité de 100 pb^{-1}) sur la mesure de la section efficace est obtenue pour les trois différents scénarios d'identification

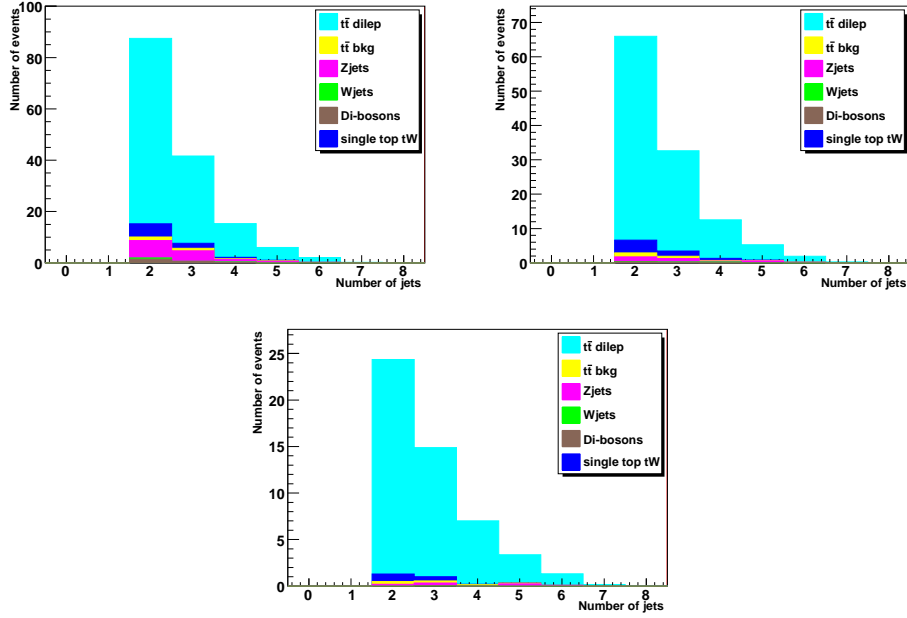
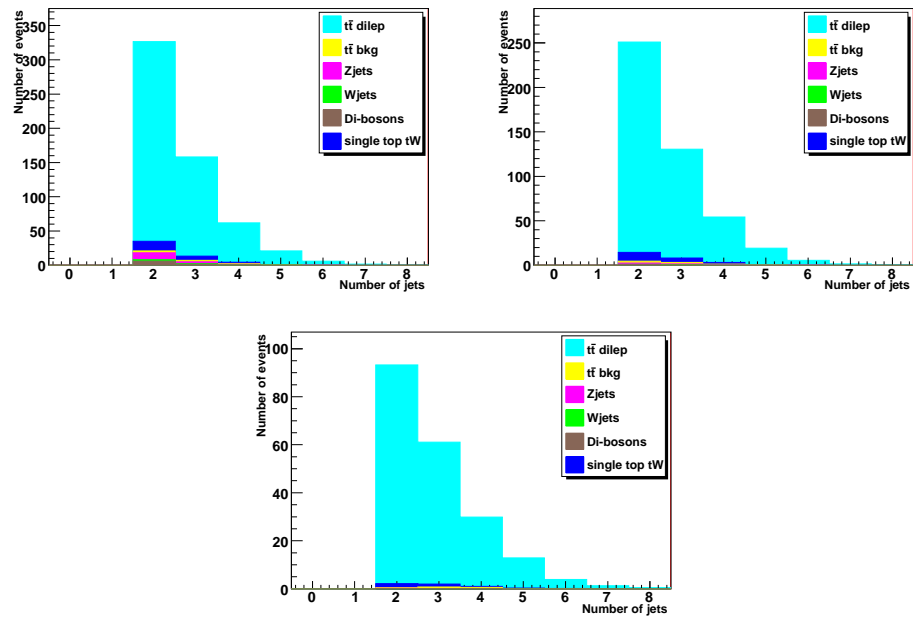
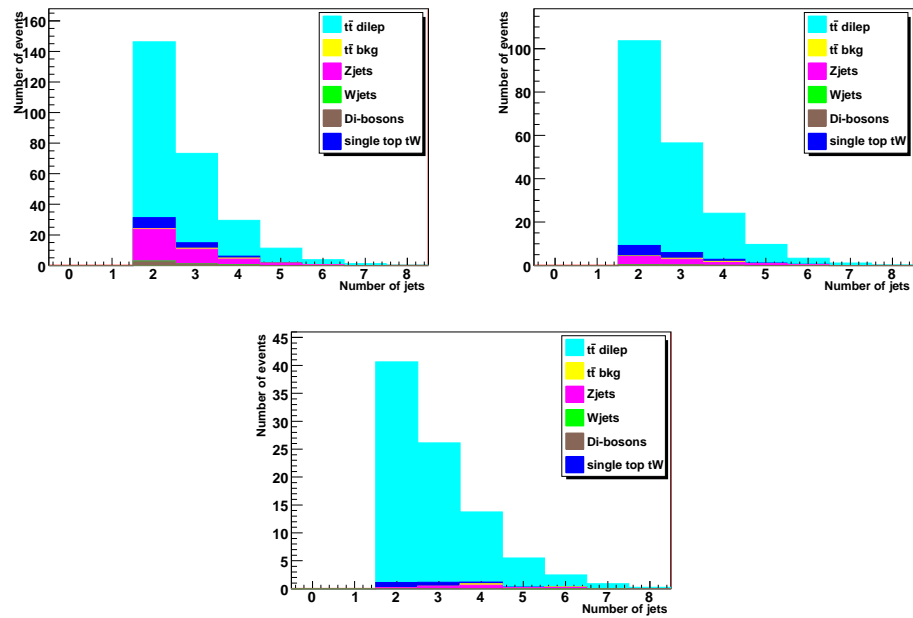


FIG. 5.30: Pour le canal $e - e$, distributions de multiplicité de jets par événement après la sélection sans identification des jets b (en haut à gauche) ou avec au moins 1 jet b (en haut à droite) ou 2 jets b (au milieu) identifiés, pour une luminosité intégrée de 100 pb^{-1} .

de jets b . Les tableaux 5.17 à 5.19 représentent le rapport signal sur bruit, l'efficacité de sélection et l'erreur statistique correspondante dans le cas de sélections contenant au moins 0, 1 ou 2 jets étiquetés comme jets b , pour les canaux $e - e$, $e - \mu$ et $\mu - \mu$ respectivement.

On observe que l'utilisation de l'identification des jets b permet d'augmenter significativement le rapport signal sur bruit. Dans le cas où un seul jet identifié comme jet b est demandé, les rapports signal sur bruit augmentent d'environ un facteur 2 sans changer significativement l'incertitude statistique. L'incertitude systématique liée à l'estimation du bruit de fond tW par la simulation Monte-Carlo peut s'avérer importante. Lorsqu'au moins deux jets identifiés b sont demandés, les rapports signal sur bruit augmentent d'environ un facteur 3 pour le canal $e - e$, d'un facteur 4 pour le canal $e - \mu$ et d'un facteur 5 pour le canal $\mu - \mu$ tandis que les erreurs statistiques relatives augmentent de 5.2%, 2.7% et 3.5% pour les canaux $e - e$, $e - \mu$ et $\mu - \mu$ respectivement. Le bruit de fond tW est alors plus faible, conduisant à de plus faibles erreurs systématiques liées à son estimation.

Le choix de la sélection à utiliser sur les données réelles ne peut donc se faire qu'après une étude approfondie des erreurs systématiques.

FIG. 5.31: Identique à la figure 5.30 pour le canal $e - \mu$.FIG. 5.32: Identique à la figure 5.30 pour le canal $\mu - \mu$.

5.7 Erreurs systématiques

Les erreurs systématiques sur la mesure de la section efficace sont principalement liées aux incertitudes sur l'efficacité de sélection des événements $t\bar{t}$, estimée à partir du Monte-Carlo, et sur l'estimation du nombre d'événements de bruit de fond.

Lorsque aucune sélection sur l'identification des jets b n'est appliquée, l'incertitude sur le bruit de fond instrumental provient de l'erreur systématique sur la méthode de la matrice, évalué à 21% dans le canal $e - e$. Le bruit de fond Z +jets, comme nous l'avons vu dans la section 5.5, peut être estimé avec une incertitude d'environ 30% pour le canal $e - e$ et 21% pour le canal $\mu - \mu$.

Lorsque l'identification d'au moins un jet b est demandée, le bruit de fond est réduit mais les erreurs statistiques Monte-Carlo deviennent plus importantes ($\geq 10\%$) pour les événements Z +jets, W +jets, $t\bar{t}$ leptons+jets et di-bosons, ce qui complique l'estimation des incertitudes systématiques. Dans ce cas, une approche conservatrice est suivie. Les erreurs systématiques dues à ces événements sont supposées être de 25% pour la sélection avec au moins 1 jet b et de 50% pour la sélection avec au moins 2 jet b . Comme le bruit de fond tW sera mal connu au démarrage, nous considérons que l'incertitude sur le nombre d'événements tW sélectionnés est de 50%.

Dans l'estimation des efficacités de sélection du signal à partir du Monte-Carlo, les sources principales d'incertitudes sont liées d'une part à des effets de détecteur et d'autre part au choix des paramètres de la génération :

- **détecteur** : les sources principales d'incertitude proviennent de l'échelle en énergie des jets et de l'efficacité d'identification des jets b . Les erreurs correspondantes sont estimées en faisant varier le p_T des jets et la coupure sur les discriminants des algorithmes d'identification des jets b . Le lot d'événements "Chowder" est utilisé. Les variations relatives des nombres d'événements sélectionnés (par rapport à la valeur nominale pour ce lot) sont utilisées pour calculer les erreurs ;
- **génération** : les incertitudes liées aux différents paramètres des générateurs ALPGEN et PYTHIA ont été étudiées. Pour ALPGEN (production des processus durs), des lots d'environ 70 000 événements $t\bar{t}$ ont été générés¹⁰ en utilisant différentes valeurs pour la masse du quark top et les PDFs. Le générateur PYTHIA traite l'hadronisation des quarks, les radiations ISR et FSR ainsi que les événements sous-jacents. Les incertitudes correspondantes sont estimées en générant des lots de 70 000 événements avec différentes configurations des paramètres de PYTHIA à partir des mêmes événements ALPGEN. La simulation rapide de CMS a été utilisée [120]. Un lot de référence de 70 000 événements a également été généré avec les paramètres standard de la génération dans CMS. Les déviations relatives des nombre d'événements sélectionnés entre ce lot de référence et les lots étudiés sont utilisés pour calculer

¹⁰produits par la simulation rapide FastSim

les erreurs systématiques.

L'incertitude sur la luminosité intégrée doit également être pris en compte. Pour 100 pb^{-1} , l'incertitude sur la luminosité est estimée à 10% environ [103].

5.7.1 Incertitudes liées au détecteur

5.7.1.1 Echelle en énergie et résolutions des jets

A cause des coupures sur l'impulsion transverse des jets, les systématiques sont liées aux incertitudes sur l'échelle en énergie et la résolution en impulsion transverse des jets. Ces deux contributions sont traitées en même temps afin de prendre en compte leurs corrélations. Les erreurs correspondantes sont également propagées à la \cancel{E}_T .

- **Echelle en énergie des jets** : la quadri-impulsion des jets est corrigée par un facteur α tel que :

$$p_{scaled}^\mu = (1 \pm \alpha) p^\mu. \quad (5.33)$$

où p^μ est la quadri-vecteur impulsion des jets. Au démarrage, l'incertitude sur l'échelle en énergie des jets devrait être de $\alpha \approx 10\%$ environ.

- **Résolution en impulsion transverse des jets** : la résolution en énergie des jets peut également être mal décrite par le Monte-Carlo. L'incertitude correspondante est estimée en faisant varier l'énergie transverse des jets :

$$p'_{x(y)} = p_{x(y)} + \text{Gauss}(0, f \times \sigma(p_T, \eta)), \quad (5.34)$$

où $p_{x(y)}$ est la composante en $x(y)$ de l'impulsion des jets et $\text{Gauss}(0, f \times \sigma(p_T, \eta))$ une fonction gaussienne de moyenne 0 et de largeur $f \times \sigma(p_T)$ avec $\sigma(p_T, \eta)$ la résolution en p_T des jets. Le paramètre f est choisi de manière à obtenir une variation de la résolution de 10%.

Les valeurs $\sigma(p_T, \eta)$ sont déterminées à l'aide de la formule (2.12) à partir d'un lot d'événements QCD $80 < p_T < 120 \text{ GeV}/c$, pour $|\eta| < 1.4$ et $|\eta| > 1.4$ séparément. Les paramètres correspondants (cf. section 2.5.4.1) sont $a = 1.78$, $b = 1.3$ et $c = 0.053$ pour $|\eta| < 1.4$ et $a = 1.84$, $b = 1.14$ et $c = 0.027$ pour $|\eta| > 1.4$.

- **Incertitude liée à la \cancel{E}_T** : les variations des impulsions transverses de jets doivent être prises en compte dans le calcul de l'énergie transverse manquante. Pour ce faire, les contributions des jets sont d'abord soustraites à l'énergie transverse manquante :

$$E_{x(y)}^{miss'} = E_{x(y)}^{miss} + \sum_{jet} p_{x(y)}^{jet}, \quad (5.35)$$

où $E_{x(y)}^{miss}$ est la composante $x(y)$ de \cancel{E}_T . La somme est appliquée sur tous les jets de l'événement. On varie alors le p_T des jets ; suivant les formules (5.33) et (5.34) de manière à reproduire les incertitudes sur leur échelle en énergie et leur résolution en $p_{x(y)}$. La \cancel{E}_T correspondante est alors calculée :

$$E_{x(y) var}^{miss} = E_{x(y)}^{miss'} - \sum_{jet} P_{x(y) var}^{jet}, \quad (5.36)$$

où p_T^{jet} est l'impulsion transverse des jets après variation.

Les erreurs systématiques sont déterminées en appliquant la sélection d'événements sur les p_T et la \cancel{E}_T après variation. Les différences relatives entre le nombre d'événements sélectionnés avec et sans variation sont prises comme erreurs systématiques. Elles sont présentées dans le tableau 5.20 pour $\geq 0, 1$ ou 2 jets b identifiés.

$\Delta N/N$ (%)	$(e - e)$	$(e - \mu)$	$(\mu - \mu)$
0 jet b	+5.2	+0.5	+ 4.9
	-10.4	-7.4	- 9.9
1 jet b	+5.5	+0.5	+4.8
	-10.7	-7.1	-10.3
2 jet b	+5.1	+0.6	+5.1
	-10.6	-7.3	-10.2

TAB. 5.20: Variation relative (en %) du nombre d'événements de signal sélectionnés après variation de l'échelle en énergie et de la résolution en p_T des jets pour les différents canaux et pour différentes sélections sur l'identification des jets b .

Pour les événements de signal, les erreurs systématiques sont de +5% à -10% environ pour les canaux di-électrons et di-muons. Pour le canal électron-muon, les erreurs systématiques sont plus faibles, entre +0.5 et -7% environ et proviennent principalement car la coupure en \cancel{E}_T est relâchée pour ce canal. Pour de plus grandes luminosités et lorsque la calibration des jets et de la \cancel{E}_T seront mieux maîtrisées, ces erreurs devraient diminuer significativement.

5.7.1.2 Algorithme d'identification des jets b

Les incertitudes sur l'efficacité d'identification des jets b et sur le taux de mauvaise identification doivent être estimées. Elles vont dépendre du nombre de jets b demandés au cours de la sélection. Dans la section 4.3 (tableau 4.2), l'erreur systématique sur l'efficacité d'identification des jets b a été estimée à $\pm 8\%$ pour le point de fonctionnement *loose* et pour 100 pb^{-1} . Dans la section 4.4.1 (tableau 4.4), l'incertitude sur le taux de mauvaise identification a été estimée à $\pm 3.4\%$.

L'incertitude liée à l'identification des jets b est estimée en variant la coupure sur le discriminant pour les jets b et pour les jets légers $udsg$ indépendamment, de manière à faire varier l'efficacité d'identification des jets b de $\pm 8\%$ d'une part et le taux de mauvaise identification de $\pm 4\%$ d'autre part. Les différences relatives des nombres d'événements sélectionnés sont présentées dans le tableau 5.21 lorsqu'au moins un ou deux jets b sont demandés lors de la sélection.

$\Delta N/N$ (%)	$(e - e)$	$(e - \mu)$	$(\mu - \mu)$
1 jets b	+4.2	+4.4	+4.6
	-5.3	-5.2	-5.0
2 jets b	+16.3	+15.2	+17.0
	-14.6	-14.3	-15.2

TAB. 5.21: Variations relatives (en %) du nombre d'événements de signal sélectionnés dûs à l'erreur systématique sur l'identification des jets b .

L'incertitude liée à la sélection de 2 jets b s'avère importante. Rappelons que le lot de données utilisé a été généré avec un détecteur désaligné. Pour un détecteur aligné et une luminosité plus grande, ces erreurs devraient être significativement réduites.

5.7.2 Incertitude liée à la génération des événements

Les incertitudes liées aux paramètres de la génération sont déterminées en suivant les recommandations présentées dans le document [121].

5.7.2.1 Radiation QCD

Les rayonnements ISR et FSR (présentés dans la section 1.6) sont reproduits à l'aide des équations d'Altarelli-Parisi [122]. Les paramètres principaux sont le paramètre d'échelle de la QCD (Λ_{QCD}) et la virtualité Q_{max}^2 correspondant au seuil où les rayonnements ISR (FSR) s'arrêtent (débutent). La fraction de gluons émis par radiation va dépendre de ces deux paramètres.

Les paramètres de PYTHIA mis en jeu sont :

- Λ_{QCD} : PARJ(61), PARJ(72) et PARJ(81) dont la valeur centrale (0.25 GeV déterminé par l'expérience CDF [121]) est variée de ± 0.1 GeV, comme suggéré dans dans [121];
- Q_{max}^2 : les paramètres PARJ(67) et PARJ(72) contrôlent l'émission de gluons. La valeur centrale de PARJ(67) est modifiée de 0.25 à 4 GeV (avec PARJ(72) fixé à sa valeur par défaut) et celle de PARJ(72) est modifié de 1 à 16 GeV (avec PARJ(67) fixé à sa valeur par défaut) comme suggéré dans [121].

70 000 événements di-leptons ont été générés en faisant varier ces paramètres indépendamment. Une fois ces paramètres modifiés, les variations relatives les plus importantes du nombre d'événements sélectionnés sont utilisées pour le calcul des erreurs. Celles-ci sont présentés dans le tableau 5.22 pour Λ_{QCD} et dans le tableau 5.23 pour Q_{max}^2 .

On remarque que les erreurs provenant de Q_{max}^2 peuvent être importantes et augmentent avec l'identification des jets b . En effet, l'augmentation du nombre de jets de quarks légers et de gluons peut avoir un impact important sur le taux de mauvaise identification des jets b .

Λ_{QCD} $\Delta N/N$ (%)	($e - e$)	($e - \mu$)	($\mu - \mu$)
0 jet b	3.1	7.4	3.6
1 jet b	5.9	9.5	3.2
2 jets b	9.5	7.5	3.5

TAB. 5.22: Variations relatives (en %) du nombre d'événements de signal sélectionnés pour l'erreur systématique sur Λ_{QCD} .

Q_{max}^2 $\Delta N/N$ (%)	($e - e$)	($e - \mu$)	($\mu - \mu$)
0 jet b	11.6	15.4	14.3
1 jet b	14.9	16.3	16.2
2 jets b	20.4	17.1	16.2

TAB. 5.23: Variations relatives (en %) du nombre d'événements de signal sélectionnés pour l'erreur systématique sur Q_{max}^2 .

5.7.2.2 Fragmentation

Le processus de fragmentation (cf. section 1.6) pour la fragmentation des quarks légers est décrit dans PYTHIA par le modèle de Lund [44]. Les principaux paramètres sont $Lund b$, qui intervient dans le calcul de la fraction d'énergie emportée par les hadrons, et le paramètre σ_b qui correspond la dispersion en moment transverse, par rapport à la direction du parton initial, des hadrons produits par rapport à la direction du parton initial.

Les paramètres de PYTHIA correspondants sont PARJ(42)=0.52 ($Lund b$) et PARJ(21)=0.40 (σ_b). Ils sont variés, dans leurs erreurs expérimentales, de ± 0.04 et ± 0.03 respectivement. Les valeurs centrales et ces erreurs sont issues des mesures de l'expérience OPAL au LEP.

Fragmentation jets légers $\Delta N/N$ (%)	($e - e$)	($e - \mu$)	($\mu - \mu$)
0 jet b	3.2	3.7	2.6
1 jet b	3.2	3.1	2.7
2 jets b	14.3	2.7	3.9

TAB. 5.24: Variations relatives (en %) du nombre d'événements de signal sélectionnés pour l'erreur systématique sur la fragmentation des quarks légers.

La description de la fragmentation des quarks lourds est susceptible de modifier les efficacités d'identification des jets b . La fragmentation des quarks b et c est décrite par la fonction de Peterson. Les paramètres principaux de ces fonction sont ϵ_b et ϵ_c . Les paramètres PYTHIA correspondants PARJ(54)=-0.031 et PARJ(55)=-0.0041 sont variés de ± 0.011 et ± 0.0004 respectivement. Ces variations correspondent à leur incertitude

estimée par l'expérience OPAL. Une fois ces paramètres modifiés, les variations relatives les plus importantes du nombre d'événements sélectionnés sont utilisées pour le calcul des erreurs.

Fragmentation jets lourds $\Delta N/N$ (%)	$(e - e)$	$(e - \mu)$	$(\mu - \mu)$
0 jet b	4.3	2.6	3.2
1 jet b	3.2	2.3	2.9
2 jet b	7.3	2.7	4.9

TAB. 5.25: Identique au tableau 5.24 pour la fragmentation des quarks lourds.

Le tableau 5.24 représente les variations relatives des nombres d'événements pour les erreurs sur la fragmentation des quarks légers. Les variations relatives sont de l'ordre de 3%, excepté pour le canal $e - e$ avec ≥ 2 jets b où la variation est de 14%. Le tableau 5.25 correspond à la fragmentation des quarks lourds. Les variations relatives sont là encore de l'ordre de 3%, excepté pour le canal $e - e$ avec ≥ 2 jets b (7%). Les variations relatives importantes dans le cas de ce canal pourraient être révélatrices de corrélations entre les électrons et les jets b , ce qu'il reste à vérifier.

5.7.2.3 Événements sous-jacents

Dans le modèle de description des événements sous-jacents de PYTHIA, le paramètre principal est PARP(82) ("color screen cut-off"), qui correspond à l'échelle en p_T en dessous de laquelle le modèle d'interaction multiple de PYTHIA est valide. Ce paramètre est varié de $\pm 3\sigma$ dans l'intervalle [2.4, 3.4]. Sa valeur et son erreur sont déterminées par l'extrapolation des données de CDF et UA5 aux énergies du LHC [121]. Des lots d'événements (70 000 $t\bar{t}$) sont donc produits pour des valeurs de PARP(82) égale à 2.4 ou 3.4.

Une fois ces paramètres modifiés, les variations relatives les plus importantes du nombre d'événements sélectionnés sont utilisées pour le calcul des erreurs. Elles sont présentées dans le tableau 5.26 pour différentes sélections sur les jets b . Les variations relatives du nombre d'événements signal sélectionné sont comprises entre 2 et 7%.

Événements sous-jacents $\Delta N/N$ (%)	$(e - e)$	$(e - \mu)$	$(\mu - \mu)$
0 jet b	5.5	1.8	2.8
1 jet b	3.2	1.9	4.2
2 jet b	4.8	4.4	6.9

TAB. 5.26: Variations relatives (en %) du nombre d'événements de signal sélectionnés pour l'erreur systématique sur les événements sous-jacents.

5.7.2.4 PDF

Les événements $t\bar{t}$ ALPGEN ont été produits à l'aide des PDFs (cf. section 1.6) CTEQ6 [42]. Les erreurs sur l'utilisation de ces PDFs sont estimées en générant 70 000 événements $t\bar{t}$ avec les PDFs MRST1 [43]. La différence relative des nombres d'événements sélectionnés entre les lots MRST1 et CTEQ6, présentée dans le tableau 5.27, est utilisée comme incertitude systématique.

PDF (CTEQ6 vs MRST1) $\Delta N/N$ (%)	$(e - e)$	$(e - \mu)$	$(\mu - \mu)$
0 jet b	2.9	1.9	3.8
1 jet b	4.0	2.1	3.8
2 jet b	13.6	6.0	3.9

TAB. 5.27: Variations relatives (en %) du nombre d'événements de signal sélectionnés pour l'erreur systématique sur les PDFs.

5.7.2.5 La masse du top

Dans le Monte-Carlo, les événements $t\bar{t}$ ont été générés pour une masse du quark top de $175 \text{ GeV}/c^2$. L'incertitude liée à la valeur de la masse du top est estimée en générant 70 000 événements pour différentes masses : 170, 172, 175 et 178 GeV/c^2 . A titre illustratif, les nombres d'événements sélectionnés (sans identification des jets b) en fonction de la masse du top sont présentés sur la figure 5.33 pour le canal $e - \mu$.

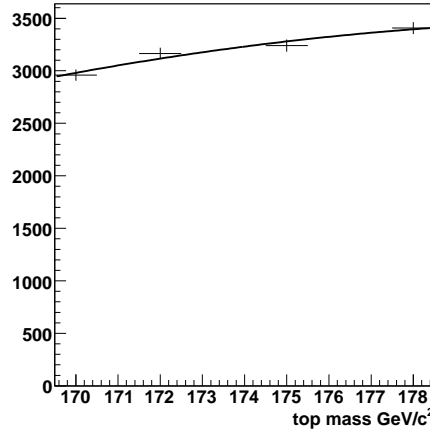


FIG. 5.33: Dans le canal $e - \mu$, nombre d'événements sélectionnés après la génération de 70 000 événements $t\bar{t}$ di-leptons pour différentes masses du top. Les points sont ajustés par un polynôme d'ordre 2.

L'incertitude expérimentale sur la masse du top [29] étant de ± 1.2 GeV/ c^2 , la sensibilité de la sélection sur m_t est estimée en ajustant les valeurs de la figure 5.33 par un polynôme d'ordre 2. A partir de la fonction ajustée, les nombres d'événements sélectionnés pour des masses de $175+1.2$ et $175-1.2$ GeV/ c^2 sont calculés. Les différences par rapport à cas $m_t = 175$ GeV/ c^2 sont de $+2.8$ et -0.6% respectivement. La différence la plus élevée (2.8%) est retenue comme erreur systématique.

5.7.3 Erreurs Systématiques globales

Les erreurs systématiques relatives globales sur la mesure de la section efficace de production $t\bar{t}$ sont calculées en propageant les différentes erreurs dans l'équation (5.32). Elles sont résumées dans le tableau 5.28.

$\Delta\sigma/\sigma(\%)$	$(e - e)$	$(e - \mu)$	$(\mu - \mu)$
0 jet b	18.8	19.6	20.5
1 jet b	21.7	21.7	21.6
2 jets b	36.9	26.7	26.9

TAB. 5.28: Erreur systématique globale sur la mesure de la section efficace de production $t\bar{t}$.

En considérant que les trois canaux sont 100% corrélés, les erreurs systématiques relatives totales sur la mesure de la section efficace $t\bar{t}$ sont de 19.6%, 21.7% et 30.2% lorsqu'au moins 0, 1 ou 2 jets b sont sélectionnés.

Ces résultats montrent que l'identification des jets b n'augmente que légèrement l'erreur systématique globale lorsqu'au moins un jet b est demandé. Lorsqu'au moins deux jets b sont demandés, l'erreur systématique globale augmente plus significativement. Dans le cas où la contamination d'événements QCD serait importante dans les données réelles, l'utilisation d'algorithmes d'identification de jets b pourrait cependant permettre de rejeter ce bruit de fond et donc de limiter l'erreur systématique liée à son estimation.

Dans l'estimation des erreurs systématiques qui vient d'être présentée, certaines sources de systématiques, comme l'incertitude sur l'efficacité de sélection des leptons ou sur les efficacités du système de déclenchement, n'ont pas encore été étudiées.

5.8 Conclusion

Lors de cette analyse, nous avons mis en place une sélection des événements $t\bar{t}$ dileptons pour une luminosité intégrée de 100 pb $^{-1}$ et avec un détecteur désaligné et mal calibré. Cette sélection peut être finalisée par l'application de l'identification des jets b . Des méthodes d'estimation des bruits de fonds $t\bar{t}$, W +jets et QCD d'une part (méthode de la matrice) et du bruit de fond Z +jets d'autre part ont été élaborées et leur précision estimée. Une stratégie de mesure de la section efficace a également été mise en place

et les erreurs statistiques et systématiques principales ont été évaluées. Finalement, les précisions sur la mesure de la section efficace de production $t\bar{t}$ dans le canal di-leptons sont :

$$\frac{\Delta\sigma}{\sigma}(0 \text{ jet } b) = 3.8 \% (\text{Stat.}) \oplus 19.6 \% (\text{Syst.}) \oplus 10.0 \% (\text{Lumi.}) \quad (5.37)$$

$$\frac{\Delta\sigma}{\sigma}(1 \text{ jet } b) = 4.0 \% (\text{Stat.}) \oplus 21.7 \% (\text{Syst.}) \oplus 10.0 \% (\text{Lumi.}) \quad (5.38)$$

$$\frac{\Delta\sigma}{\sigma}(2 \text{ jets } b) = 5.8 \% (\text{Stat.}) \oplus 30.2 \% (\text{Syst.}) \oplus 10.0 \% (\text{Lumi.}) , \quad (5.39)$$

pour des sélections avec ≥ 0 , ≥ 1 ou ≥ 2 jets b . Les erreurs globales correspondantes sont de 22%, 24% et 32%. On remarque que pour une luminosité de 100 pb^{-1} , l'erreur sur la mesure est déjà dominée par les erreurs systématiques. Les contributions principales proviennent des incertitudes sur l'échelle en énergie des jets, sur l'identification des jets b et sur la luminosité.

Notons également que l'erreur liée au paramètre Q_{max}^2 peut être surestimée car on varie ce paramètre d'un grand nombre de déviations standards par rapport à sa valeur nominale. La division de cette erreur par un facteur 2 réduit les erreurs globales à 18%, 19% et 27% pour des sélections avec ≥ 0 , ≥ 1 ou ≥ 2 jets b . Avec un détecteur mieux compris et une luminosité suffisante, un ajustement des paramètres de la génération devrait réduire significativement les incertitudes liées au Monte-Carlo.

Dans le cadre du modèle Supersymétrique MSSM et pour des masses du boson de Higgs chargé H^\pm inférieure à la masse du top, un quark top peut se désintégrer en H^\pm , qui se désintègre lui-même en lepton τ ($H^\pm \rightarrow \tau\nu_\tau$) dans plus de 90% des cas. Ce nouveau canal de désintégration du top conduirait à une diminution du nombre d'événements $t\bar{t}$ avec des électrons et des muons dans l'état final.

A titre d'illustration, pour une valeur de $\tan\beta$ de 20 et pour une masse du Higgs chargé de $140 \text{ GeV}/c^2$, la section efficace du processus $t\bar{t} \rightarrow H^\pm W^\pm b\bar{b} \rightarrow \tau\nu_\tau l\nu_l b\bar{b}$, $\tau \rightarrow \text{hadrons}$, est de l'ordre de 10 pb [38]. Cela se traduirait par une diminution d'environ 20% de la section efficace de production des canaux $t\bar{t}$ di-leptons $e - e$, $e - \mu$ et $\mu - \mu$.

Comme nous avons obtenu une précision sur la mesure de la section efficace di-leptons de l'ordre de 20%, ces canaux pourraient être utilisés pour la mise en évidence indirecte de boson du Higgs chargé ou pour contraindre les paramètres du modèle MSSM.

6

Conclusion

Le travail de thèse qui a été présenté dans ce manuscrit suit trois axes principaux : l'étude des données issues de l'enregistrement de muons cosmiques, la préparation des algorithmes d'identification de jets b et de la mesure du taux de mauvaise identification à partir des données et finalement l'estimation de la précision attendue sur la mesure de la section efficace de production des paires de quarks top dans le canal di-leptons, pour une luminosité de 100 pb^{-1} et une énergie dans le centre de masse de $\sqrt{s} = 14 \text{ TeV}$.

Les événements di-leptons sont sélectionnés en demandant la présence de deux leptons (électrons et/ou muons) isolés de grande impulsion transverse, d'une grande énergie transverse manquante et de deux jets b. La présence des jets b est déterminée par l'utilisation d'algorithmes d'identification de jets b. Les algorithmes de reconstructions des traces, utilisés par les algorithmes d'identification de jets, sont étudiés à l'aide de données cosmiques.

Les efficacités de reconstruction des traces de muons issus de la désintégration de particules cosmiques ont été estimées à partir de données réelles acquises entre Juillet et Novembre 2008. Les muons sont tout d'abord reconstruits à partir des informations des chambres à muons. Les trajectoires des muons reconstruits sont ensuite propagées jusqu'au trajectographe et les traces reconstruites compatibles sont recherchées. Les efficacités de reconstruction des traces ont ainsi été estimées à plus de 99% dans la région la plus centrale du trajectographe et pour des muons verticaux. Ces résultats sont reportés dans une note en cours de publication. Les impulsions transverses entre muons reconstruits dans le trajectographe et dans les chambres à muons ont également été étudiées et une mauvaise description du champ magnétique dans les détecteurs à muons a été mise en évidence.

L'algorithme d'identification des jets b "Probabilité par Jet" a été étudié. Il utilise les distributions en paramètre d'impact signé négativement comme calibration pour déterminer la probabilité pour les traces d'un jet de provenir du vertex primaire. Une procédure de calibration et des histogrammes de contrôle de ces calibrations ont été définis. Des méthodes d'estimation des performances des algorithmes ont également été testées sur des données simulées. Notamment, la méthode "d'étiquetage négatif" permet d'estimer le taux de mauvaise identification à partir de données QCD multi-jets. Une étude détaillée des erreurs systématiques liées à cette méthode a été effectuée. Les erreurs systématiques sont comprises entre 3% et 20%, suivant le point de fonctionnement utilisé.

La précision de la mesure de la section efficace de production pour des collisions à $\sqrt{14}$ TeV dans le centre de masse et pour une luminosité de 100 pb^{-1} a été étudiée. Une sélection d'événements a été définie et a été rendue publique par la collaboration. Elle est finalisée par l'utilisation d'algorithmes d'identification de jets b. Une méthode d'estimation des bruits de fond QCD et W +jets a été définie et testée avec des événements simulés. Les incertitudes systématiques principales ont été étudiées. Les erreurs systématiques attendues sont de l'ordre de 20% pour une sélection sans jet b ou avec au moins un jet b identifié et de 30% pour une sélection demandant au moins deux jets b identifiés. Les erreurs statistiques attendues sont comprises entre 4% et 6%. Les premières collisions pour les analyses de physiques devraient avoir lieu à des énergies dans le centre de masse de $\sqrt{s} = 10$ TeV. Une analyse similaire, utilisant des événements simulés avec $\sqrt{s} = 10$ TeV, est en cours d'approbation par la collaboration. Elle conduit à des erreurs systématiques similaires mais des erreurs statistiques légèrement supérieures ($> 5\%$).

Le travail qui a été effectué au cours de cette thèse visait à préparer les premières analyses du LHC. Malheureusement, compte tenu du retard du LHC, aucune donnée issue de collisions n'a pu être étudiée. Mais les différents outils nécessaires à la mise en évidence des premiers événements $t\bar{t}$ et à la première mesure de section efficace de production de paires $t\bar{t}$ ont été développés. Ces mesures pourront être effectuées dans les deux années à venir avec les premières collisions du LHC. La compréhension des événements issus de la production de quarks top devrait ouvrir la voie à des recherches de physique nouvelle.



Taux de Mauvaise identification

A.1 Erreurs systématiques de l'estimation du taux de mauvaise identification

Les détails du calcul des erreurs systématiques de la méthode d'estimation du taux de mauvaise identification vont être présentés dans cet annexe. Ces erreurs correspondent aux erreurs sur le facteur R_{light} .

A partir de la formule (4.18) de la section 4.4.1 le facteur R_{light} peut se ré-écrire comme :

$$R_{light} = \frac{\varepsilon_{MC}^{mistag}}{\varepsilon_{MC}^-} \quad (A.1)$$

$$= \frac{L^+}{L} \frac{A}{A^-} \quad (A.2)$$

$$= \frac{\ell^+ + g^+}{\ell + g} \frac{\ell + g + c + b}{\ell^- + g^- + c^- + b^-} \quad (A.3)$$

avec

- L le nombre de jets légers étiquetables, L^+ le nombre de jets légers étiquetés comme jets b , L^- le nombre de jets légers étiquetés négativement (cf. 4.4.1) ;
- A le nombre total de jets étiquetables, A^+ le nombre total de jets étiquetés comme jet b , A^- le nombre de jets étiquetés négativement ;
- ℓ (g , c , b) le nombre de jets étiquetables uds (g , c , b), avec les indices $+$ et $-$ correspondant aux nombres étiquetés comme jet b et étiquetés négativement.

A partir de cette définition, les systématiques relatives deviennent :

– sur les fractions de jets b et c :

$$\left(\frac{\sigma(R_{light})}{R_{light}}\right)_c = \frac{\sigma(c)}{c} \left[\frac{c^-}{A^-} - \frac{c}{A} \right], \quad (\text{A.4})$$

$$\left(\frac{\sigma(R_{light})}{R_{light}}\right)_b = \frac{\sigma(b)}{b} \left[\frac{b^-}{A^-} - \frac{b}{A} \right]; \quad (\text{A.5})$$

– sur la fraction de gluon :

$$\left(\frac{\sigma(R_{light})}{R_{light}}\right)_g = \frac{\sigma(g)}{g} \left[\frac{g^+}{L^+} - \frac{g^-}{A^-} - \frac{g}{\ell} \left(\frac{\ell^+}{L^+} - \frac{\ell^-}{A^-} \right) \right]; \quad (\text{A.6})$$

– sur la fraction de V^0 :

$$\left(\frac{\sigma(R_{light})}{R_{light}}\right)_{V^0} = \frac{\sigma(V^0)}{V^0} \left[\frac{L^{0+}}{L^+} - \frac{A^{0-}}{A^-} - \frac{L^0}{L} + \frac{A^0}{A} \right], \quad (\text{A.7})$$

où A^0 (L^0) est le nombre de jets légers étiquetables contenant un candidat V^0 .

– inversion du signe du paramètre d'impact :

$$\left(\frac{\sigma(R_{light})}{R_{light}}\right)_{IP\ sign\ flip} = x_{flip} \left[\frac{A^+}{A_{no\ veto}^-} - \frac{L_{no\ veto}^-}{L^+} \right], \quad (\text{A.8})$$

où x_{flip} est la fraction de jets étiquetés pour lesquels le signe est inversé.

$A_{no\ veto}^-$ et $L_{no\ veto}^-$ sont les nombres de jets totaux et de jets légers étiquetés négativement sans appliquer de veto sur les jets.

A.2 Résultats pour les point de fonctionnement *loose* et *tight*.

Les résultats de la méthode d'estimation du taux de mauvaise identification pour le point de fonctionnement *loose* sont présentés sur les figures **A.1** et **A.2** et pour le point de fonctionnement *tight* sur les figures **A.3** et **A.4**

A.3 Les différents algorithmes d'étiquetage négatif

Une mise à jour de de la méthode d'estimation du taux de mauvaise identification des jets b a été faite à l'aide de données plus récentes produites avec la version *CMSSW_2_1_12* du code de *CMS*. Dans cette version, des algorithmes d'étiquetages négatifs ont été définis pour la majorité des algorithmes d'identification de jets b de *CMS*.

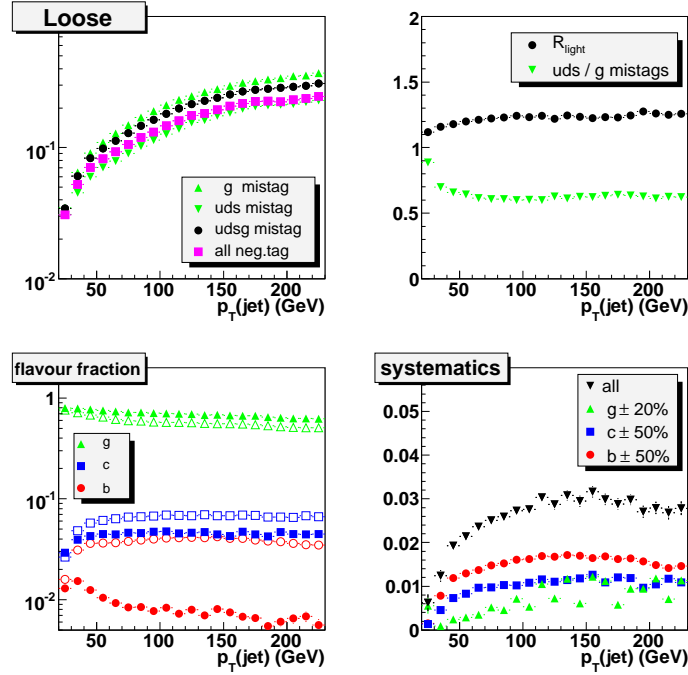


FIG. A.1: Dans les échantillons QCD et avec le point de fonctionnement *loose* avec l'application du veto sur les jets : les jets étiquetés négativement sont rejetés s'ils possèdent au moins une trace avec $IP/\sigma_{IP} > 4$, en fonction du p_T des jets : (en haut à gauche) les taux de mauvaise identification $\varepsilon^{MC}(udsg)$ (cercles), $\varepsilon^{MC}(uds)$ et $\varepsilon^{MC}(g)$ (triangles) et l'efficacité d'étiquetage négatif $\varepsilon_{neg}^{MC}(udsg + c + b)$ pour toutes les saveurs (carrés), (en haut à droite) le facteur R_{light} (cercles) et le rapport $\varepsilon^{MC}(uds)/\varepsilon^{MC}(g)$ (triangles), (en bas à gauche) les fractions de jets b , c et g avant (symbole creux) et après étiquetage négatif (symboles pleins) et (en bas à droite) les erreurs systématiques provenant de l'incertitude sur les fractions de jets g , c et b .

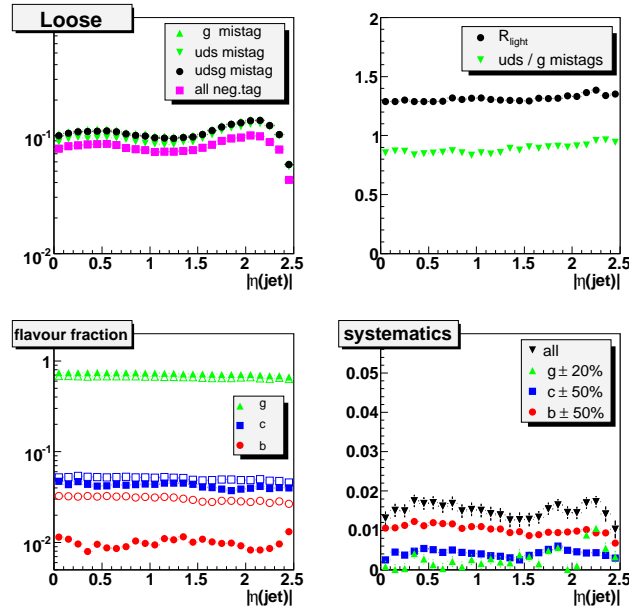


FIG. A.2: Identique à la figure A.2 en fonction de η .

Il sera présenté ici les tests de la méthode pour ces différents algorithmes. Des événements QCD ($p_T > 80$ GeV) simulés à l'aide de PYTHIA ont été utilisés. La génération a été faite pour des collisions à 10 TeV et la simulation correspond au cas d'un détecteur parfait (pas de désalignement ni de mauvaise calibration).

Certains algorithmes d'étiquetage dans leur configuration standard utilisent à la fois des traces ayant un paramètre d'impact positif et négatif. Il est alors nécessaire de définir des algorithmes d'étiquetage positif qui seront utilisés pour l'estimation du taux de mauvaise identification. Les différences entre algorithmes d'étiquetage positif et algorithme d'étiquetage standard doivent être également prises en compte dans l'estimation des erreurs systématiques de la méthode d'estimation du taux de mauvaise identification.

Les performances globales des algorithmes ont également changé dans cette dernière version du code de *CMS*, ce qui est principalement dû à l'amélioration des algorithmes de reconstruction des traces (cf section 2.3.2.1). Il est donc nécessaire de redéfinir des points de fonctionnement pour les différents algorithmes d'identification de jets b . Les résultats seront présentés pour chaque algorithme d'étiquetage.

– **Définition d'un algorithme d'étiquetage négatif pour l'algorithme de comptage de trace**

Les algorithmes de comptage de traces ont déjà été étudiés dans la section précédente. Nous ne verrons ici qu'une mise à jour de ces résultats avec les données *CMSSW_2_1_X*.

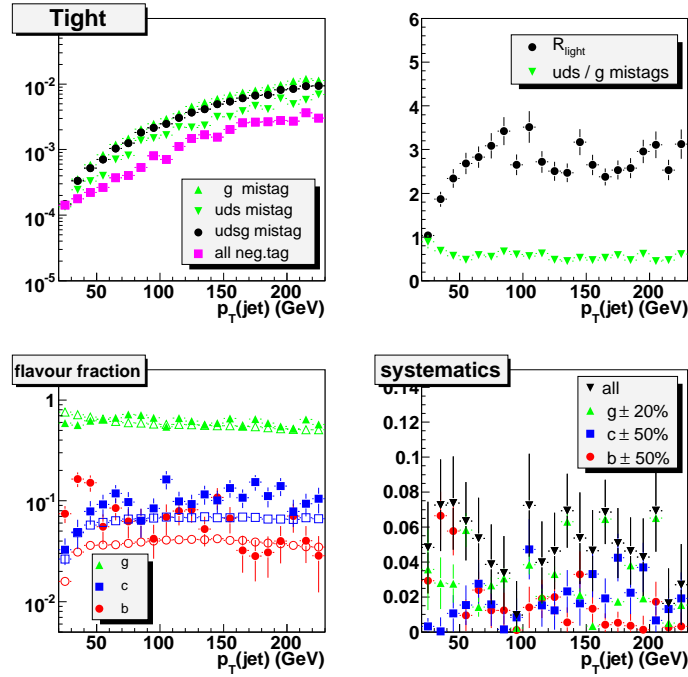


FIG. A.3: Dans les échantillons QCD et avec le point de fonctionnement *loose* avec l'application du veto sur les jets : les jets étiquetés négativement sont rejetés s'ils possèdent au moins une trace avec $IP/\sigma_{IP} > 4$, en fonction du p_T des jets : (en haut à gauche) les taux de mauvaise identification $\varepsilon^{MC}(udsg)$ (cercles), $\varepsilon^{MC}(uds)$ et $\varepsilon^{MC}(g)$ (triangles) et l'efficacité d'étiquetage négatif $\varepsilon_{neg}^{MC}(udsg + c + b)$ pour toutes les saveurs (carrés), (en haut à droite) le facteur R_{light} (cercles) et le rapport $\varepsilon^{MC}(uds)/\varepsilon^{MC}(g)$ (triangles), (en bas à gauche) les fractions de jets b , c et g avant (symbole creux) et après étiquetage négatif (symboles pleins) et (en bas à droite) les erreurs systématiques provenant de l'incertitude sur les fractions de jets g , c et b .

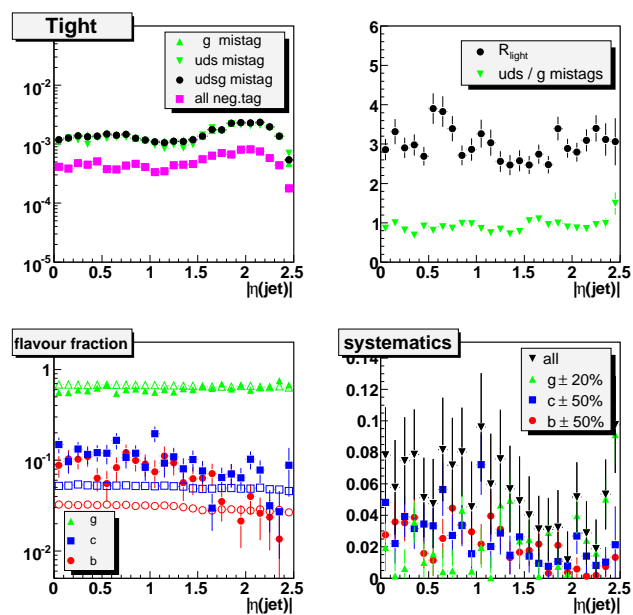


FIG. A.4: Identique à la figure A.4 en fonction de η .

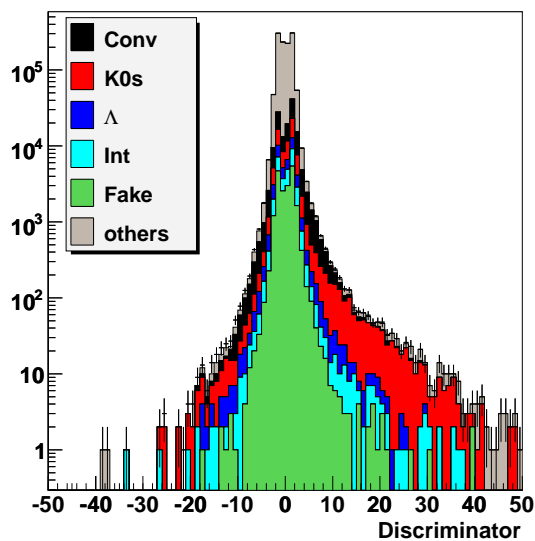


FIG. A.5: Distribution du discriminant de l'algorithme *TCHE* pour différentes origines des jets.

La figure A.5 montre les distributions du discriminant de l'algorithme *TCHE* pour les jets de saveur légère *udsg*. L'outil *TrackHistory* est utilisé ici pour déterminer la provenance des traces : conversions de photons, K_s^0 , Λ , interactions nucléaires (noté "Int") ou fausses traces (noté "fake"). Plusieurs de ces processus peuvent être mis en jeu lors de la production des particules chargés, une hiérarchie est donc instaurée dans la détermination de l'origine des traces. Cette hiérarchie peut être résumée par : $V^0 >$ interaction nucléaire $>$ fausses traces. Par exemple, une trace produite par interaction nucléaire à partir d'une trace provenant de particule V^0 sera prise en compte comme une trace provenant de particules V^0 . Lorsqu'aucun des ces processus n'est trouvé dans la généalogie d'une trace, l'origine de la trace est noté "Others". L'origine d'un jet est déterminé par l'origine des deux (trois) traces ayant les plus grande significances du paramètre d'impact S_{IP} pour l'algorithme *THHE* (*TCHP*). A partir de la figure A.5, on remarque que les contributions principales aux queues de distribution positive proviennent principalement des K_s^0 et de conversion de photons.

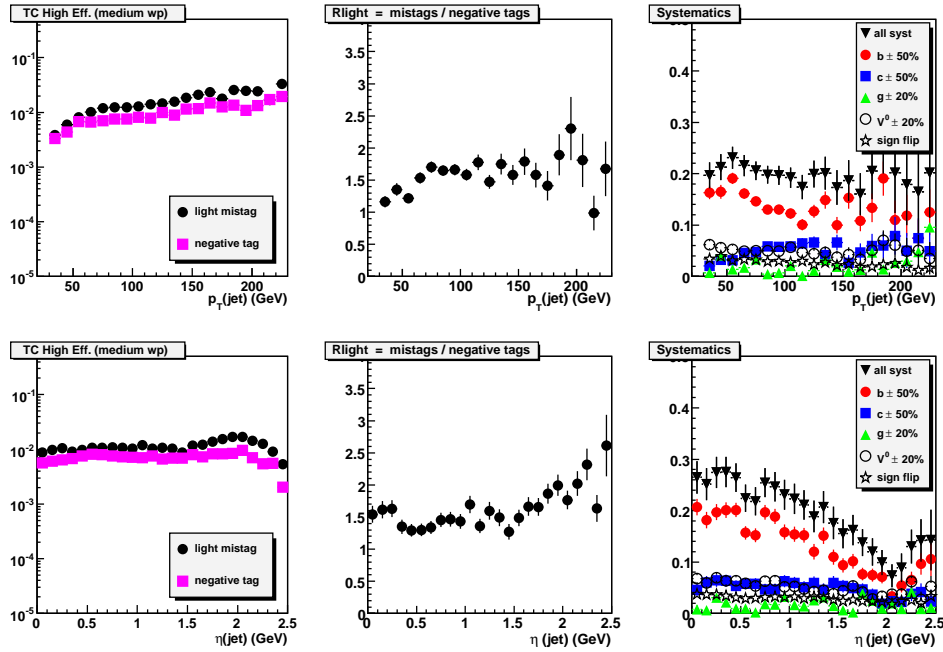


FIG. A.6: Pour le point de fonctionnement *medium* et l'algorithme *TCHE*, efficacités d'identification des jets légers et d'étiquetage négatif (à gauche), rapport R_{light} au milieu et erreurs systématiques principales (à droite) en fonction du p_T (en haut) et du $|\eta|$ (en bas) des jets. Aucun veto sur les jets n'est appliqué.

les figures A.6 et A.7 représentent l'efficacité ε_{udsg} et l'efficacité d'étiquetage négatif (graphique de gauche), le rapport R_{light} (graphique du milieu) et les erreurs systématiques (graphique de droite) en fonction du p_T (graphiques en haut) et de

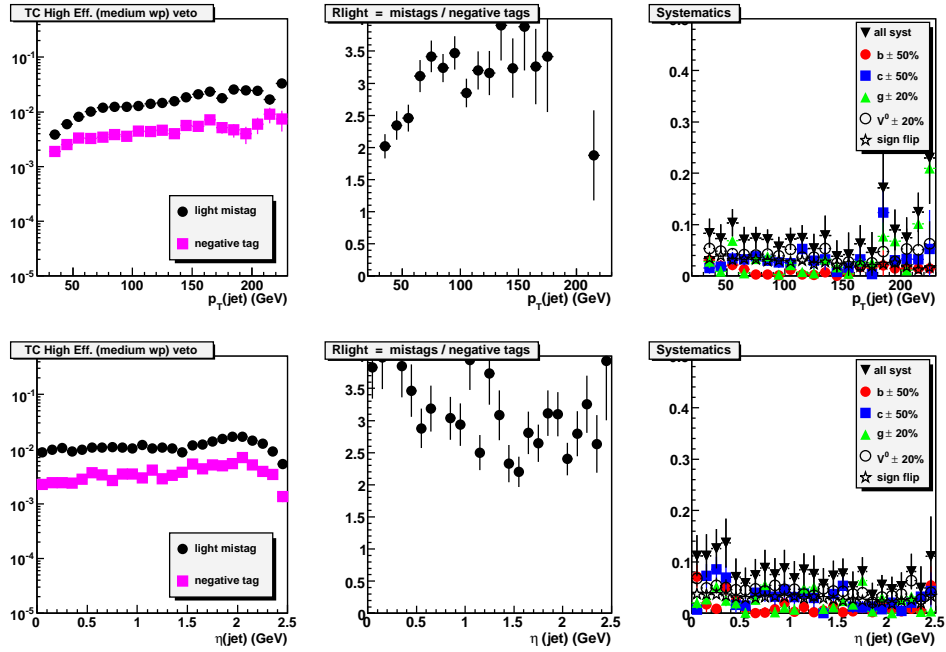


FIG. A.7: Pour le point de fonctionnement *medium* et l'algorithme *TCHE*, efficacités d'identification des jets légers et d'étiquetage négatif (à gauche), rapport R_{light} au milieu et erreurs systématiques principales (à droite) en fonction du p_T (en haut) et du $|\eta|$ (en bas) des jets. Un veto sur les jets est appliqué : les jets étiquetés négativement sont rejetés s'ils possèdent au moins une trace avec $S_{IP} > 4$.

$|\eta|$ (graphique en bas) des jets, pour le point de fonctionnement *medium* avec (A.7 ou sans A.6) veto sur les jets. Lorsque le veto sur les jets est appliqué par les jets étiquetés négativement sont rejetés s'ils possèdent au moins une trace avec $S_{IP} > 4$. Les résultats obtenus pour avec les échantillons *CMSSW_2_1_12* sont globalement comparable à ceux obtenu avec les échantillons *CMSSW_1_3_1* (cf section 4.4.1 figures 4.31 à 4.34).

– **Définition d'un algorithme d'étiquetage négatif pour l'algorithme de probabilité par jet**

Comme discuté dans le chapitre 4.1.1.2, la probabilité par jet peut être calculée en utilisant les traces de S_{IP} positive uniquement ou toutes les traces associées à un jet. La probabilité négative, correspondant à l'algorithme d'étiquetage négatif est alors simplement calculée à partir des traces de S_{IP} négatifs uniquement.

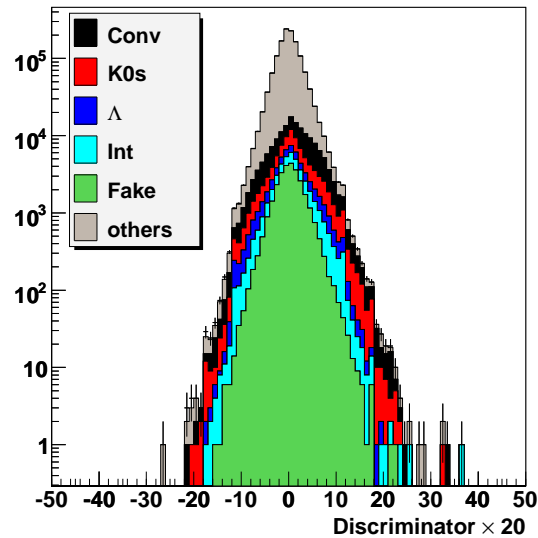


FIG. A.8: Distribution du discriminant de l'algorithme de probabilité par jet pour différentes origines des jets.

La figure A.8 représente la distribution des discriminants ($\times 20$) des algorithmes de probabilité de jets positifs et négatifs pour les jets *udsg*. L'origine de chaque jet est déterminé à partir de l'origine des traces qui lui sont associées à l'aide de l'outil *TrackHistory*. On remarque qu'il n'existe pas de queue de distribution pour les discriminant positifs.

Les figures A.9 à A.10 représentent le taux de mauvaise identification (jets *udsg*) et l'efficacité d'étiquetage négatif (à gauche), le rapport R_{light} (au milieu) et les erreurs systématiques (à droite) en fonction du p_T (graphique en haut) et du $|\eta|$

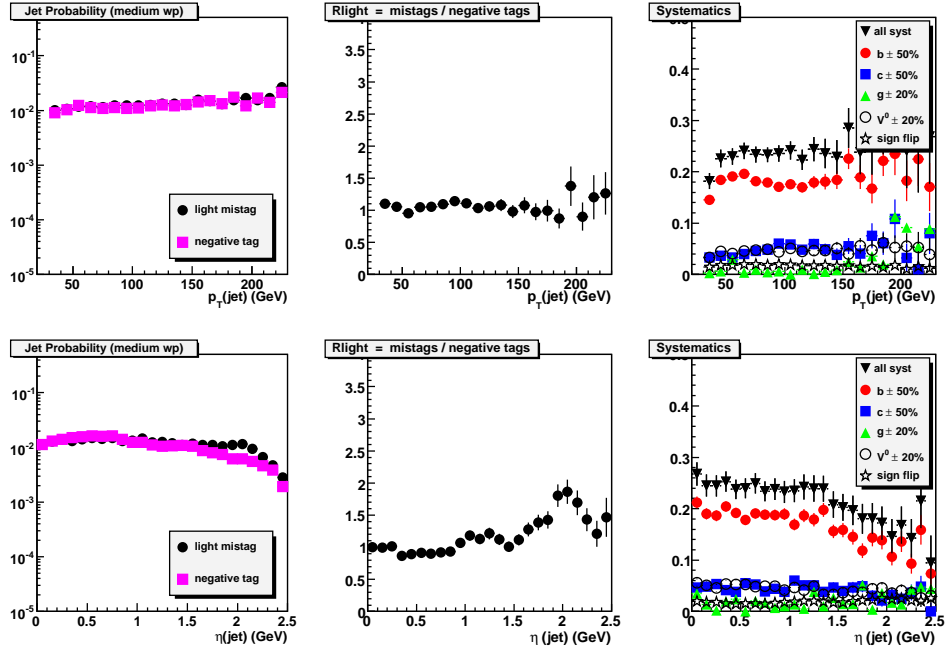


FIG. A.9: Pour le point de fonctionnement *medium* et l'algorithme de probabilité par jet, efficacités d'identification des jets légers et d'étiquetage négatif (à gauche), rapport R_{light} au milieu et erreurs systématiques principales (à droite) en fonction du p_T (en haut) et du $|\eta|$ (en bas) des jets. Aucun veto sur les jets n'est appliqué.

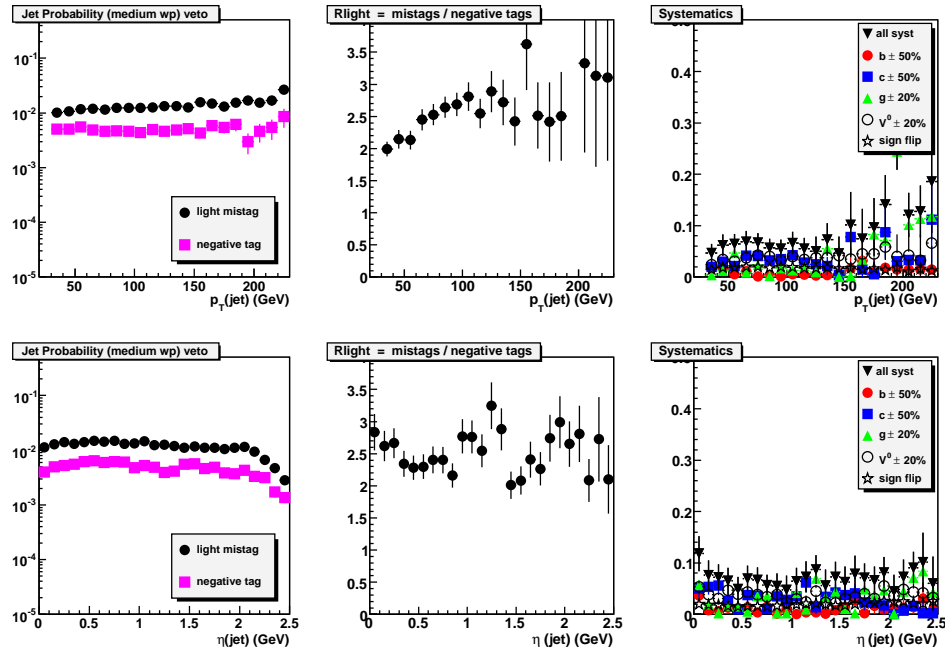


FIG. A.10: Pour le point de fonctionnement *medium* et l'algorithme de probabilité par jet, efficacités d'identification des jets légers et d'étiquetage négatif (à gauche), rapport R_{light} au milieu et erreurs systématiques principales (à droite) en fonction du p_T (en haut) et du $|\eta|$ (en bas) des jets. Un veto sur les jets est appliqué : les jets étiquetés négativement sont rejetés s'ils possèdent au moins une trace avec $IP/\sigma > 4$.

(graphique en bas) des jets, pour le point de fonctionnement *medium*, avec (A.10) ou sans (A.9) veto sur les jets. Comme les calibrations utilisent les traces ayant un paramètre d'impact négatif, le facteur R_{light} très proche de 1. Les erreurs systématiques restent comparables avec celle de l'algorithme *TCHE*.

– **Définition d'un algorithme d'étiquetage négatif pour l'algorithme de Vertex Secondaire**

Il est possible de signer la distance de vol des vertex secondaires reconstruits. Le signe d'un vertex secondaire est alors défini comme le signe du cosinus de l'angle entre la direction de la distance de vol et l'axe du jet. La figure A.11 représente la distribution du discriminant de l'algorithme du Vertex Secondaire transformé par la formule ($discr = discr \times 10 - 10$) pour les jets de saveurs légères. L'origine du vertex secondaire est déterminée à partir de l'origine des traces qui lui sont associées.

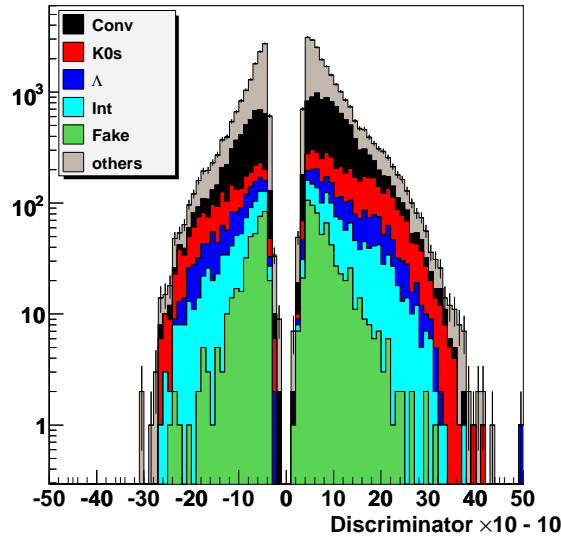


FIG. A.11: Distribution du discriminant de l'algorithme de vertex secondaire pour différentes origines de vertex.

La figure A.12 représente le taux de mauvaise identification des jets légers et l'efficacité d'étiquetage négatif (graphique de gauche), le rapport R_{light} (graphique du milieu) et les erreurs systématiques (graphiques de droites) en fonction du p_T (graphiques en haut) et de $|\eta|$ (graphique en bas) des jets, pour le point de fonctionnement *medium*. On remarque que les erreurs systématiques sont très faibles, même sans l'utilisation du veto sur les jets. Cela est dû à la faible fraction des jets b et c étiqueté par l'algorithme négatif.

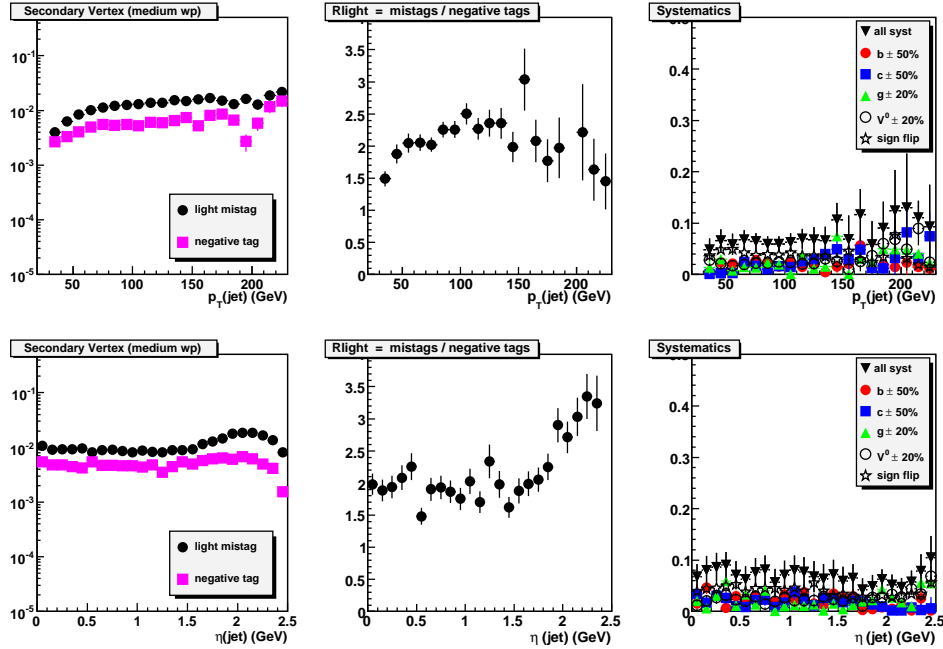


FIG. A.12: Pour le point de fonctionnement *medium* et l'algorithme du vertex secondaire, efficacités d'identification des jets légers et d'étiquetage négatif (à gauche), rapport R_{light} au milieu et erreurs systématiques principales (à droite) en fonction du p_T (en haut) et du $|\eta|$ (en bas) des jets. Aucun veto sur les jets n'est appliqué.

– **Définition d'un algorithme d'étiquetage négatif pour l'algorithme de Vertex Secondaire Combiné**

Le cas de l'algorithme de Vertex Secondaire combiné est un peu plus délicat. En effet, cet algorithme utilise différentes information supplémentaire comme il l'a été expliqué dans la section 4.1.2. Par défaut, l'algorithme du Vertex Secondaire Combiné utilise toutes les traces (positives ou négatives) et tous les vertex secondaires reconstruits (positifs ou négatifs). Il est cependant possible de définir un algorithme positif qui n'utilise que les traces positives et les vertex secondaires positifs. De la même manière, l'algorithme négatif n'utilise que les traces négatives et les vertex secondaires négatifs. La figure A.13 représente la distribution du discriminant de l'algorithme du Vertex Secondaire Combiné pour les jets de saveurs légères. L'origine du discriminant dépend du type de vertex secondaire reconstruit. Dans le cas des vertex ou pseudo-vertex secondaire, l'origine des discriminants est déterminée à partir de toutes les traces associées aux vertex ou pseudo-vertex secondaires. Dans le cas où il n'y a pas de vertex ni de vertex secondaire reconstruits, l'origine du discriminant est déterminé à partir de toutes les traces associées au jet, comme expliqué dans la section 4.1.1.2.

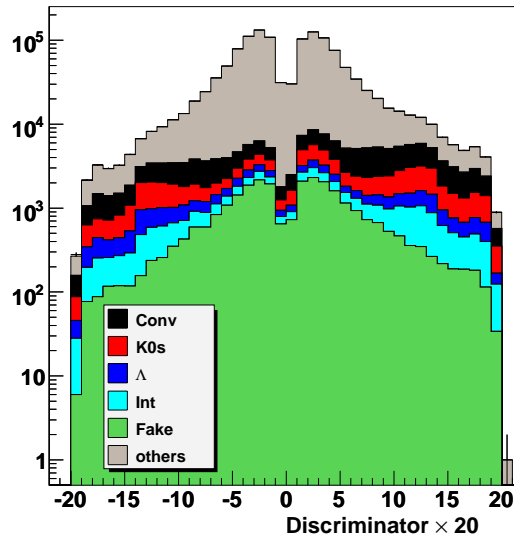


FIG. A.13: Distribution du discriminant de l'algorithme de vertex secondaire combiné pour différentes origines de vertex.

Les figure A.14 et A.15 présentent le taux de mauvaise identification (jets $udsg$) et l'efficacité d'étiquetage négatif (à gauche), le rapport R_{light} (au milieu) et les erreurs systématiques (à droite) en fonction du p_T (en haut) et du $|\eta|$ (en bas) des

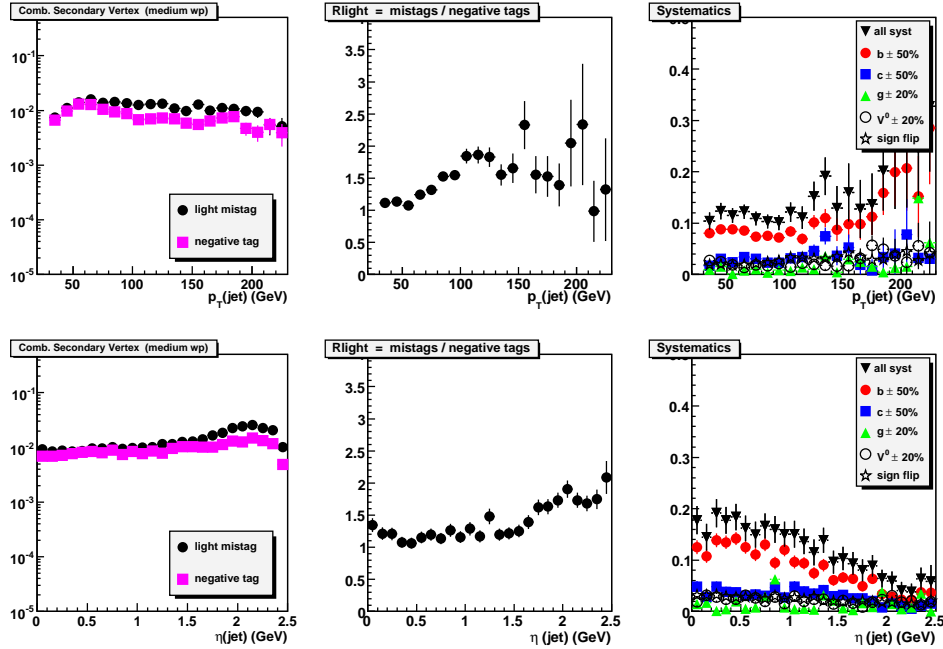


FIG. A.14: Pour le point de fonctionnement *medium* et l'algorithme du vertex secondaire combiné, efficacités d'identification des jets légers et d'étiquetage négatif (à gauche), rapport R_{light} au milieu et erreurs systématiques principales (à droite) en fonction du p_T (en haut) et du $|\eta|$ (en bas) des jets. Aucun veto sur les jets n'est appliqué.

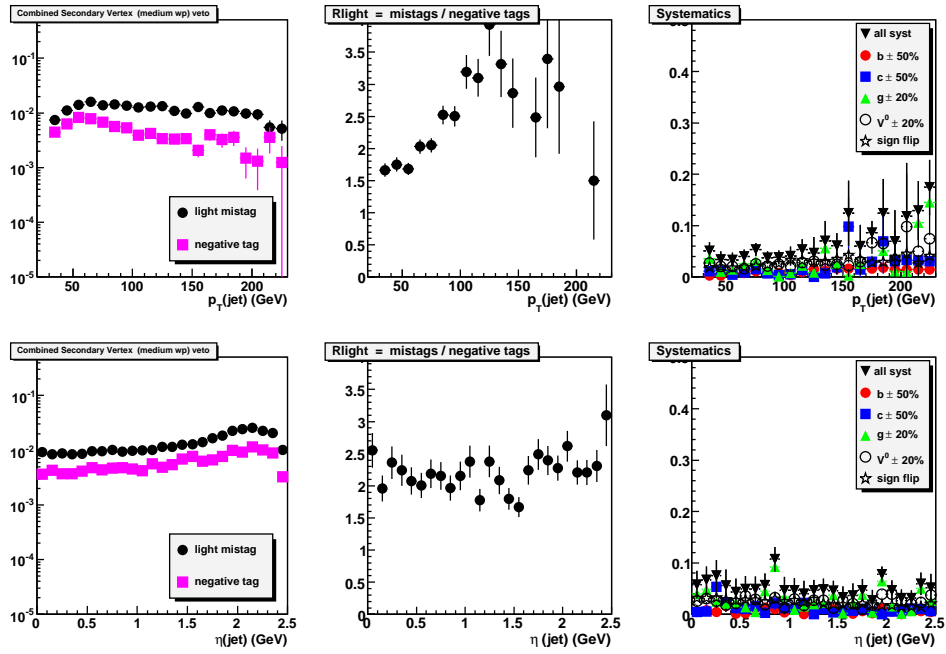


FIG. A.15: Pour le point de fonctionnement *medium* et l'algorithme de probabilité par jet, efficacités d'identification des jets légers et d'étiquetage négatif (à gauche), rapport R_{light} au milieu et erreurs systématiques principales (à droite) en fonction du p_T (en haut) et du $|\eta|$ (en bas) des jets. Un veto sur les jets est appliqué : les jets étiquetés négativement sont rejetés s'ils possèdent au moins une trace avec $IP/\sigma > 4$.

jets, pour le point de fonctionnement *medium* avec (A.15) et sans (A.14) veto sur les jets.

L'algorithme SoftMuon est en cours d'étude.

B

system8

table des échantillons utilisé dans la méthode system8 :

Echantillon Monte Carlo	Nombre d'événement
QCD \hat{p}_T 20-30	159 488
QCD \hat{p}_T 30-50	193 961
QCD \hat{p}_T 50-80	1 375 153
QCD \hat{p}_T 80-120	579 134
QCD \hat{p}_T 120-170	390 110
QCD \hat{p}_T 170-230	401 244
QCD \hat{p}_T 230-300	199 287
semileptonic $b\bar{b}$ 30-50	59 849
semileptonic $b\bar{b}$ 50-80	59 604
semileptonic $b\bar{b}$ 80-120	60 508
semileptonic $b\bar{b}$ 120-170	45 274
semileptonic $b\bar{b}$ 170-230	42 584
semileptonic $c\bar{c}$ 30-50	55 332
semileptonic $c\bar{c}$ 50-80	49 373
semileptonic $c\bar{c}$ 80-120	48 086
semileptonic $c\bar{c}$ 120-170	45 132
semileptonic $c\bar{c}$ 170-230	47 008
<i>Soupe</i>	3 811 127
inclusive $b\bar{b}$ 50-80	44 999
inclusive $b\bar{b}$ 80-120	40 517
inclusive $c\bar{c}$ 30-50	94 831
inclusive $c\bar{c}$ 50-80	154 919
inclusive $c\bar{c}$ 80-120	87 767
inclusive $c\bar{c}$ 120-170	84 297
inclusive $c\bar{c}$ 170-230	42 277
Incl $b\bar{b}$ 30-50	603 306
Incl $b\bar{b}$ 50-80	213 737
Incl $b\bar{b}$ 80-120	164 853
Incl $b\bar{b}$ 120-170	76 390
Incl $b\bar{b} > 170$	220 571
inclusive $t\bar{t}$	1 246 365

TAB. B.1: Nombre d'événements disponibles dans chaque échantillon utilisé dans l'analyse. Le terme *Soupe* se réfèrent à la somme des événements QCD et $b\bar{b}$ et $c\bar{c}$ avec des poids identiques pour chaque événement.

Table des figures

1.1	Illustration du développement perturbatif en diagrammes de Feynman du processus $e^+e^- \rightarrow \mu^+\mu^-$ au <i>Next-to-Leading Order</i>	9
1.2	Exemples de diagrammes de production d'un paire quark anti-quark au niveau de l'arbre dans des collisions proton-proton par annihilation quark anti-quark (en haut à gauche) et par fusion de gluons.	12
1.3	Diagrammes de production de boson W à gauche et de boson Z à droite, au niveau de l'arbre dans des collisions proton-proton, où f se réfère aux fermions.	15
1.4	Diagrammes de production d'événements di-bosons WW (en haut à gauche), WZ (en haut à droite) et ZZ (en bas), au niveau de l'arbre dans des collisions proton-proton.	16
1.5	Schéma illustrant le potentiel de Higgs et le mécanisme de brisure de symétrie.	17
1.6	Evolution des inverses de constante de couplage électromagnétique ($U(1)$), faible ($SU(2)$) et forte ($SU(3)$) en fonction de l'énergie.	20
1.7	Mesure de la masse du quark top dans les expériences CDF et DØ du Tevatron à partir de différents canaux de désintégration. La valeur combinée est de $172.4 \pm 1.2 \text{ GeV}/c^2$	21
1.8	À gauche, contribution du quark top aux corrections radiatives des propagateurs des bosons W et Z . Au milieu et à droite, contributions du boson de Higgs aux corrections radiatives des propagateurs des bosons W et Z	22
1.9	Contrainte de la masse du boson W (axe des ordonnées) et de la masse du quark top (axe des abscisses) sur la masse du boson de Higgs. Le contour bleu correspond à la région de masse du Higgs permise avec un niveau de confiance de 68% et calculée à partir des mesures directes de la masse du W et du top de LEP2 et du Tevatron. Le contour rouge correspond au résultat combiné de LEP1 et de SLD. La bande verte correspond à la région de masse du Higgs permise par le Modèle Standard ($< 1 \text{ TeV}/c^2$) et non exclue par les mesures directes au LEP ($> 114 \text{ GeV}/c^2$) [20].	23
1.10	Diagrammes de production de paire de quarks top au niveau de l'arbre dans les collisions pp au LHC.	24

1.11	Canaux de production de single top pour la voie s (à gauche), la voie tW (au milieu) et la voie t (à droite)	25
1.12	Canaux de désintégration des paires de quarks top $t\bar{t}$ et leurs rapports d'embranchement.	26
1.13	Schéma représentant la modélisation d'une collision proton-proton par les générateurs d'événements.	28
2.1	Schéma de la section d'un aimant dipolaire du LHC	33
2.2	Schéma représentant le complexe d'accélération du CERN	33
2.3	Schéma du détecteur CMS.	35
2.4	Schéma du système de coordonnées utilisé dans CMS.	35
2.5	Schéma représentant une coupe du trajectographe de CMS.	36
2.6	La quantité de matière du trajectographe de CMS en unité de longueur de radiation en fonction de la pseudo-rapidité.	37
2.7	Efficacité de reconstruction des traces dans des événements QCD $80 < \hat{p}_T < 120$ GeV/c pour les versions CMSSW_1_6_12 (en haut) et CMSSW_2_1_12 (en bas) du code de CMS en fonction de l'impulsion transverse (à gauche) et de la pseudo-rapidité des traces(à droite).	41
2.8	Efficacité de reconstruction des traces dans les événements Z +jets générés par ALPGEN et reconstruits avec la version CMSSW_1_6_12 avec l'algorithme de reconstruction CTF sans <i>reconstruction itérative</i>	42
2.9	Taux de fausses traces reconstruites dans des événements QCD $80 < \hat{p}_T < 120$ GeV/c pour les versions CMSSW_1_6_12 (à gauche) et CMSSW_2_1_12 (à droite) en fonction de la pseudo-rapidité des traces.	42
2.10	Schéma représentant une coupe d'un quart du détecteur CMS avec ses chambres à muons.	45
2.11	Efficacité de reconstruction des muons issus de la désintégration de bosons W dans des événements $t\bar{t}$ en fonction du p_T et du $ \eta $ des muons générés associés. Une coupure à 10 GeV/c sur l'impulsion transverse des muons générés est appliquée.	46
2.12	Résolutions relative sur l'impulsion transverse p_T (en haut à gauche), sur la pseudo-rapidité $ \eta $ (en haut à droite) et sur l'angle azimutal ϕ (en bas) des muons en fonction du p_T des muons générés associés.	47
2.13	Schéma illustrant la disposition des cristaux du ECAL dans une coupe représentant un quark du détecteur.	49
2.14	Résolution en énergie du ECAL mesurée en faisceau test.	50
2.15	Schéma de l'arrangement des tours calorimétriques de HCAL dans le plan (r,z) pour un quart du détecteur.	51
2.16	Efficacité de reconstruction des électrons issus de la désintégration de bosons W dans des événements $t\bar{t}$, en fonction de l'impulsion transverse p_T et de la pseudo-rapidité des électrons générés associés.	53

2.17	Résolutions sur l'énergie transverse E_T (en haut à gauche), sur la pseudo-rapacité $ \eta $ (en haut à droite) et sur l'angle azimutal ϕ (en bas) des électrons en fonction de l'énergie transverse E_T des électrons générés associés (de $E_T > 10$ GeV)	55
2.18	A gauche, rapport des énergies transverses des jets reconstruits et des jets générées en fonction de la pseudo-rapacité des jets générés et pour différentes énergies transverses générées. L'algorithme du cône itératif avec $\Delta R = 0.5$ est utilisé dans des événements QCD . A droite, réponse relative di-jets dans des événements QCD ayant deux jets de haut p_T dont l'impulsion transverse moyenne est comprise dans l'intervalle $[120,250]$ GeV/c.	58
2.19	A gauche, distribution en masse invariante des candidats di-jets provenant de bosons W lorsque que ces jets sont ou ne sont pas correctement sélectionnés comme issus de W , pour les événements $t\bar{t}$ et pour une luminosité intégrée de 5 fb^{-1} . A droite, évolution du résultat de l'ajustement gaussien des distributions an masse invariante des candidats di-jets issus de W pour différentes valeurs du facteur correctif ΔC	59
2.20	A gauche, efficacité de reconstruction des jets en fonction de l'impulsion transverse des jets générés associés pour la partie centrale ($ \eta < 1.5$). A droite, efficacité de reconstruction des jets ($30 < p_T < 40$ GeV/c) en fonction de la pseudo-rapacité des jets générés. Des événements QCD avec différents intervalles en \hat{p}_T ont été utilisés.	60
2.21	Résolution en énergie transverse des jets reconstruits par l'algorithme de cône itératif $\Delta R = 0.5$ en fonction de E_T^{MC} dans la partie centrale $ \eta < 1.4$, dans les bouchons $1.4 < \eta < 3$ et dans la région très avant $3 < \eta < 5$	61
2.22	Résolution en pseudo-rapacité (à gauche) et en angle azimutal (à droite) des jets reconstruits par l'algorithme du cône itératif $\Delta R = 0.5$ en fonction de E_T^{MC} pour des événements QCD	61
2.23	A gauche, résolution en énergie transverse manquante \cancel{E}_T en fonction de la somme des énergies transverses des jets reconstruits $\sum E_T$ pour des événements QCD . A droite, résolution de l'énergie transverse manquante en fonction de \cancel{E}_T avec et sans la correction en énergie des jets pour des événements $t\bar{t}$	62
3.1	Vue d'ensemble de la structure en anneaux du trajectographe. Les différents rectangles représentent les anneaux définis par l'algorithme <i>Road Search</i> . Les anneaux utilisés pour la recherche des graines sont colorés en gris clair pour les couches internes et en gris foncé pour les couches externes. . . .	71
3.2	Efficacité de reconstruction des traces en fonction du numéro de <i>run</i> , sans coupure d'association traces-muons locaux et dans le cas de la sélection <i>PRT</i>	74

3.3	Dans les données <i>Craft</i> , distribution de la pseudo-rapacité η (en haut à gauche), de l'angle azimutal ϕ (en haut à droite), du paramètre d'impact transverse $d0$ (en bas à gauche) et du p_T (en bas à droite) des traces de muons reconstruits dans le trajectographe. Pour une trace donnée, η et ϕ sont estimés au point de la trajectoire de la trace qui est le plus proche du centre du détecteur. Les histogrammes rouges correspondent à l'algorithme <i>RS</i> , les bleus à l'algorithme <i>CTF</i> et les noirs à l'algorithme <i>CosmicTF</i>	77
3.4	Dans les données <i>Craft</i> , distribution du χ^2 normalisé (à gauche) et nombre de coups associés (à droite) aux traces de muons reconstruits dans le trajectographe.	78
3.5	Distribution en η (en haut à gauche), en ϕ (en haut à droite), en $d0$ (en bas à gauche) et en p_T (en bas à droite) des traces de muons reconstruites par l'algorithme <i>CTF</i> dans les données (trait plein) et dans le Monte-Carlo (trait pointillé). Pour une trace donnée, η et ϕ sont estimés au point de la trajectoire de la trace qui est le plus proche du centre du détecteur. Les distributions sont normalisées aux mêmes nombres d'événements.	79
3.6	Distribution du χ^2 normalisé (à gauche) et nombre de coups associés aux traces de muons reconstruits (à droite) par l'algorithme <i>CTF</i> dans les données (trait plein) et dans le Monte-Carlo (trait pointillé). Les distributions sont normalisées aux mêmes nombres d'événements.	80
3.7	Distribution en η (en haut à gauche), en ϕ (en haut à droite), en $d0$ (en bas à gauche) et en p_T (en bas à droite) des muons locaux reconstruits. Pour une trace donnée, η et ϕ sont estimés au point de la trajectoire de la trace qui est le plus proche du centre du détecteur.	81
3.8	Dans les données <i>Craft</i> , distribution du χ^2 normalisé (à gauche) et nombre de coups (à droite) associés aux muons locaux reconstruits.	82
3.9	Schéma représentant la propagation des muons locaux et des traces vers la couche compatible la plus externe.	83
3.10	Positions locales Y en fonction de la position locale Y des traces (à gauche) et des muons locaux (à droite) au niveau de leur module compatible le plus externe.	84
3.11	Positions locales X des traces en fonction de la position locale X des muons locaux (à gauche) et positions locales Y des traces en fonction de la position locale Y des muons locaux (à droite) au niveau du module compatible le plus externe.	84
3.12	Différences entre les positions locales X (en haut à gauche), Y (en haut à droite) et distances entre ces positions ΔR pour les trajectoires extrapolées au niveau du module compatible le plus externe.	86
3.13	Distributions en R_1 et R_2 pour des traces de $p_T \approx 10$ GeV/c. Les distributions sont ajustés par un fonction gaussienne.	87

3.14	Moyennes des ajustements de R_1 (en haut à gauche), de R_2 (en haut à droite) et déviation standard de R_2 (en bas) en fonction de l'impulsion transverse des traces.	88
3.15	Multiplicité de traces après la sélection PRT pour les algorithmes RS , CTF et $CosmicTT$	89
3.16	Distribution de $\sigma(X_{TOB4})$ et $\sigma(Y_{TOB4})$ pour des événements après l'application de la sélection PRT	90
3.17	Efficacité de reconstruction des traces après la sélection PRT dans les données Craft (en traits plein) et dans le Monte-Carlo (en pointillés), en fonction du η (en haut à gauche), du ϕ (en haut à droite), de $d0$ (en bas à gauche) et du p_T des muons locaux associés.	92
3.18	Efficacité de reconstruction des traces (CTF) dans les données Craft en fonction du ϕ des muons locaux associés et pour différentes coupures sur $ d0 $	93
3.19	Efficacité de reconstruction des traces CTF , $CosmicTF$ et RS en fonction du p_T des muons locaux associés, dans les données (trait plein) et dans le Monte-Carlo (pointillés) dans le cas de la sélection LHC	94
4.1	Représentation du paramètre d'impact signé.	100
4.2	Distribution en signficance du paramètre d'impact 3D pour différentes saveurs de jets pour des événements QCD $80 < \hat{p}_T < 120 \text{ GeV}/c$	101
4.3	Distributions des impulsions transverses des traces reconstruites issues de la désintégration de particules K_s^0 (en haut à gauche), Λ (en haut à droite) et de conversion de photon (en bas) pour la première trace de plus haut p_T (en rouge) et la seconde (en gris).	102
4.4	Distribution normalisé à l'unité de la signficance du paramètre d'impact 3D pour les 2ièmes (à gauche) et 3ième (à droite) traces.	103
4.5	Courbes de performance des algorithmes de comptage de traces pour des données QCD PYTHIA ($CMSSW_2_1_12$) avec $80 < \hat{p}_T < 120 \text{ GeV}$	104
4.6	Distributions normalisées à l'unité de la signficance du paramètre d'impact des jets légers pour les 2ième (à gauche) et 3ième (à droite) traces et pour différents scénarios de désalignement et de décalibration (QCD $80 < \hat{p}_T < 120 \text{ GeV}/c$).	104
4.7	Distributions normalisées à 1 de la signficance du paramètre d'impact des jets b , pour les 2ième (à gauche) et 3ième (à droite) traces et pour différents scénarios de désalignement et de décalibration (QCD $80 < \hat{p}_T < 120 \text{ GeV}/c$).	105
4.8	Courbes de performance des algorithmes de comptage de traces TCHE (en haut) et TCHP (en bas) pour les jets légers pour différents scénarios de désalignement de décalibration. Les différences pour les jets c sont négligeables et non représentées (QCD $80 < \hat{p}_T < 120 \text{ GeV}/c$).	106

4.9	Efficacité de sélection des traces en fonction du taux de fausses traces sélectionnées pour différentes coupure en χ^2 normalisé. Des jets d'impulsions transverses comprises entre 120 et 170 GeV/c provenant d'événements <i>QCD</i> sont utilisés [52].	109
4.10	Probabilité des traces de signification négative (à gauche) pour toutes les saveurs de jets et probabilité des traces de significances positive et négative (à droite) pour les saveurs de jets <i>udsg</i> , <i>c</i> et <i>b</i> (données QCD PYTHIA <i>CMSSW_1_5_2</i> avec $80 < \hat{p}_T < 120$ GeV/c).	110
4.11	Probabilité par jet pour les saveurs <i>udsg</i> , <i>c</i> et <i>b</i> (données QCD PYTHIA <i>CMSSW_1_5_2</i> avec $80 < \hat{p}_T < 120$ GeV/c).	110
4.12	Courbes de performance de l'algorithme de probabilité par jet pour des données QCD PYTHIA (<i>CMSSW_2_1_12</i>) avec $80 < \hat{p}_T < 120$ GeV/c.	111
4.13	Ecart type de l'ajustement gaussien des distributions en signification sur l'intervalle de signification [-2;0] en fonction du χ^2 normalisé des traces, pour les catégories 2 (à gauche) et 5 (à droite).	112
4.14	Ecart type de l'ajustement gaussien des distributions en signification sur l'intervalle de signification [-2;0] en fonction de l'impulsion des traces, pour les catégories 2 (à gauche) et 5 (à droite).	113
4.15	Fraction de traces ayant une signification du paramètre d'impact inférieure à <i>X</i> pour <i>X</i> = -2, -3, et -4 en fonction du χ^2 normalisé des traces, pour les catégories 2 (à gauche) et 5 (à droite).	113
4.16	Fraction de traces ayant une signification du paramètre d'impact inférieure à <i>X</i> pour <i>X</i> = -2, -3, et -4 en fonction de l'impulsion des traces, pour les catégories 2 (à gauche) et 5 (à droite).	114
4.17	Pour les traces avec $0 < \eta < 0.8$ et au moins 3 coups dans le détecteur à pixels, le nombre de traces (en haut) et la fraction $f_{neg}(-2)$ (en bas) en fonction de l'impulsion des traces et de leur χ^2 normalisé pour des événements <i>QCD</i> ($80 < \hat{p}_t < 120$ GeV/c).	115
4.18	Courbes de performance (ε_{udsg} en fonction de ε_b) pour l'algorithme de probabilité par jet pour différents scénarios de désalignement du trajectographe : détecteur parfait, 100pb^{-1} et 10pb^{-1} pour des événements <i>QCD</i> avec $80 < \hat{p}_t < 120$ GeV/c.	116
4.19	Courbes de performance de l'algorithme du Vertex Secondaire pour des données QCD PYTHIA (<i>CMSSW_2_1_X</i>) avec $80 < \hat{p}_T < 120$ GeV/c.	117
4.20	Courbes de performance de l'algorithme du Vertex Secondaire combiné pour des données QCD PYTHIA (<i>CMSSW_2_1_12</i>) avec $80 < \hat{p}_T < 120$ GeV/c.	118
4.21	Courbes de performance de l'algorithme <i>SoftMuon</i> pour des données QCD PYTHIA (<i>CMSSW_2_1_12</i>) avec $80 < \hat{p}_T < 120$ GeV/c.	119

4.22	Courbes de performance des tous les algorithmes d'identification des jets b pour des données QCD PYTHIA ($CMSSW_2_1_12$) avec $80 < \hat{p}_T < 120$ GeV/c. Pour un même algorithme (pour une même couleur) les courbes supérieures correspondent à la mauvaise identification des jets c et les courbes inférieures à la mauvaise identification des jets légers $udsg$	120
4.23	Propriétés cinématiques des jets dans les échantillons QCD normalisés au même nombre d'entrée : l'impulsion transverse (en haut à gauche), la pseudo rapidité (en haut à droite) et la multiplicité de traces(en bas à gauche) des jets ainsi que la distance ΔR la plus courte entre deux jets. Symboles pleins : tous les événements, symboles creux : une coupure $\Delta R_{min} > 2$ est appliquée.	123
4.24	Taggabilité pour $n \geq 1$ or $n \geq 2$ traces associées aux jets pour différents intervalles en p_T dans des événements QCD	124
4.25	Taggabilité pour $n \geq 1$ pour différentes saveurs de jets et pour différents intervalles en p_T dans des événements QCD	124
4.26	Performances des algorithmes TCHE et TCHP (à gauche) et de l'algorithme de probabilité par jet (à droite) pour des événements QCD PYTHIA $80 < \hat{p}_T < 120$ GeV/c générés par la version $CMSSW_1_3_1$	126
4.27	Facteurs de corrélation pour le point de fonctionnement <i>medium</i> de l'algorithme de comptage de traces.	129
4.28	Facteur de corrélation pour le point de fonctionnement <i>medium</i> de l'algorithme de probabilité par jet.	130
4.29	Efficacités d'identification des jets b attendues (information de la génération) et mesurée par la méthode <i>System8</i> en fonction du p_T et du $ \eta $ des jets (avec $p_T > 30$ GeV/c) pour le point de fonctionnement <i>medium</i> de l'algorithme de comptage de traces. Seules les erreurs statistiques sont représentées.	131
4.30	Efficacités d'identification des jets b attendue (information de la génération) et mesurée par la méthode <i>System8</i> en fonction du p_T et du $ \eta $ des jets (avec $p_T > 30$ GeV/c) pour le point de fonctionnement <i>medium</i> de l'algorithme de probabilité par jet . Seules les erreurs statistiques sont représentées.	132
4.31	Dans les échantillons QCD et avec le point de fonctionnement <i>medium</i> , en fonction du p_T des jets : (en haut à gauche) les taux de mauvaise identification $\varepsilon^{MC}(udsg)$ (cercles), $\varepsilon^{MC}(uds)$ et $\varepsilon^{MC}(g)$ (triangles) et l'efficacité d'étiquetage négatif $\varepsilon_{neg}^{MC}(udsg+c+b)$ pour toutes les saveurs (carrés), (en haut à droite) le facteur R_{light} (cercles) et le rapport $\varepsilon^{MC}(uds)/\varepsilon^{MC}(g)$ (triangles), (en bas à gauche) les fractions de jets b , c et g avant (symbole creux) et après étiquetage négatif (symboles pleins) et (en bas à droite) les erreurs systématiques provenant de l'incertitude sur les fractions de jets g , c et b (cf. section 4.4.3) qui sont mal connues aux collisionneurs hadroniques.	137

4.32	Identique à la figure 4.31 en fonction du $ \eta $ des jets.	138
4.33	Identique à la figure 4.31 après application du veto sur les jets : les jets étiquetés négativement sont rejetés s'ils possèdent au moins une trace avec $IP/\sigma_{IP} > 4$	139
4.34	Identique à la figure 4.32 après application du veto sur les jets : les jets étiquetés négativement sont rejetés s'ils possèdent au moins une trace avec $IP/\sigma_{IP} > 4$	140
4.35	Distributions des différentes propriétés géométriques des traces : la longueur de désintégration (en haut à gauche), la distance à l'axe du jet (en haut à droite), le paramètre d'impact longitudinal (en bas à gauche) et le paramètre d'impact transverse (en bas à droite). Les distributions sont représentées pour les 5 catégories de traces discutées dans le texte.	142
4.36	Fractions de jets contenant au moins une trace provenant d'un V^0 pour les jets <i>udsg</i> en fonction du p_T des jets lorsque l'identification des jets <i>b</i> n'est pas appliquée (en haut à gauche). Pour les points de fonctionnement <i>loose</i> (en haut à droite), <i>medium</i> (en bas à gauche) et <i>tight</i> (en bas à gauche) du comptage de traces, fraction de jets contenant au moins une trace provenant d'un V^0 après l'application de l'algorithme d'identification des jets <i>b</i> (point noir) et de l'algorithme d'étiquetage négatif avec (carrés creux) et sans (carrés pleins) l'utilisation du veto sur les jets. Les deux traces ayant les deux (trois) plus grands paramètres d'impact sont utilisées pour déterminer les contributions de particules V_0 pour les points de fonctionnement <i>loose</i> , <i>medium</i> et <i>tight</i>	143
4.37	Identique à la figure 4.36 en fonction de $ \eta(\text{jet}) $	144
4.38	Deuxième (graphique de gauche) et troisième (graphique de droite) plus grande signification du paramètre d'impact par jet pour différentes saveurs (en haut) et différentes contributions de V^0 dans les jets <i>udsg</i> (en bas).	145
4.39	Pour chaque point de fonctionnement, le facteur R_{light} est représenté en fonction du p_T des jets pour les jets <i>udsg</i> contenant les candidats V_0 (à gauche) et les jets <i>udsg</i> sans candidats V_0 (à droite).	146
4.40	Pour les muon-jets avec un muon de $p_T^\mu > 6$ GeV/c, distribution des deuxième (à gauche) et troisième (à droite) plus grandes significances du paramètre d'impact pour tous les jets (en haut) et en appliquant une sélection sur les muons ($p_{Trel} > 1.5$ GeV/c, en bas). Aucune normalisation n'est appliquée sur les distributions.	150
4.41	Pour chaque point de fonctionnement des algorithmes de comptage de traces, les erreurs relatives provenant des V_0 et de l'inversion du signe en fonction du p_T (en haut) et de $ \eta $ (en bas) des jets. Les erreurs correspondent à une luminosité de 100 pb^{-1}	151

4.42	Pour le point de fonctionnement <i>medium</i> , rapport des efficacités $\varepsilon(udsg)$ (à gauche), $\varepsilon_{neg}(udsg)$ (au milieu) et facteur R_{light} (à droite) en fonction du p_T des jets, en comparant différents échantillons correspondant à l'émulation de différents chemins de déclenchement : 1. ≥ 1 jet de $p_T > 200$ GeV, 2. ≥ 2 jets de $p_T > 150$ GeV, 3. ≥ 3 jets de $p_T > 85$ GeV et 4. ≥ 4 jets de $p_T > 60$ GeV. Les rapports sont obtenus en divisant $\varepsilon(udsg)$, $\varepsilon_{neg}(udsg)$ et R_{light} des échantillons 2, 3 et 4 par celles de l'échantillon 1.	152
4.43	En haut à gauche : fraction des jets <i>uds</i> par rapport à tous les jets <i>udsg</i> pour des événements <i>QCD</i> , <i>W</i> +jets et <i>t\bar{t}</i> semi-leptonique. Les trois autres graphiques représentent le rapport entre les efficacités $\varepsilon(udsg)$ dans les échantillons <i>W</i> +jets et <i>t\bar{t}</i> semi-leptonique et l'efficacité $\varepsilon(udsg)$ dans l'échantillon <i>QCD</i> pour les trois points de fonctionnement : <i>loose</i> en haut à droite, <i>medium</i> en bas à gauche et <i>tight</i> en bas à droite.	155
5.1	Impulsion transverse (en haut) et pseudo-rapacité (en bas) des leptons générés issus de la désintégration des bosons <i>W</i> dans des événements ALPGEN <i>t\bar{t}</i> +0jet di-leptons pour les muons (à gauche) et les électrons (à droite).	159
5.2	Impulsion transverse (à gauche) des jets générés issus de la désintégration des hadrons beaux (à gauche) et les autres jets générés (droite) dans des événements ALPGEN <i>t\bar{t}</i> +0jet di-leptons.	160
5.3	Distribution de la \cancel{E}_T générée pour les événements ALPGEN <i>t\bar{t}</i> +0jet di-leptons.	160
5.4	Efficacité des chemins de déclenchements muon-X (en haut) et di-électrons (en bas) en fonction de p_T et η des leptons pour les événements reconstruits passant les sélections au niveau de la génération et de la reconstruction décrites dans le texte, pour les canaux $\mu - \mu$ en haut et $e - e$ en bas. Les erreurs présentées correspondent aux erreurs statistiques pour une luminosité intégrée de 100 pb^{-1} .	168
5.5	Distribution de l'énergie transverse (à gauche) et efficacité de sélection après l'application de coupures en énergie/impulsion transverse pour les électrons (en haut) et les muons (en bas) provenant de la désintégration de bosons <i>W</i> (en bleu) ou de hadrons lourds (en rouge) pour des événements <i>t\bar{t}</i> +0 jet (détecteur parfait).	170
5.6	Dans les événements <i>t\bar{t}</i> + 0 jet (soupe "Chowder"), isolation au niveau du trajectographe (à gauche) et efficacité de sélection (à droite) pour des électrons (en haut) et des muons (en bas) ayant une énergie (impulsion) transverse $> 20 \text{ GeV}(/c)$, une pseudo-rapacité $ \eta < 2.4$ et provenant de bosons <i>W</i> (en bleu) ou de hadrons lourds (en rouge).	172

5.7	Dans les événements $t\bar{t} + 0$ jet (soupe “Chowder”), isolation calorimétrique (à gauche) et efficacité de sélection correspondante (à droite) pour les électrons (en haut) et les muons (en bas) ayant une énergie (impulsion) transverse > 20 GeV(/c) et provenant de bosons W (en bleu) ou de hadrons lourds (en rouge).	173
5.8	Dans les événements $t\bar{t} + 0$ jet (soupe “Chowder”), efficacité de sélection (axe coloré) des leptons ayant une énergie (impulsion) transverse > 20 GeV(/c) issus de bosons W (à gauche) et de hadrons lourds (à droite) pour les électrons (en haut) et les muons (en bas) pour différentes coupures sur l’isolation calorimétrique (axe des abscisses) et au niveau du trajectographe (axe des ordonnées).	174
5.9	Dans le canal $e - e$, distribution en E_T de l’électron de plus haut E_T (en haut à gauche), de la masse invariante des candidats di-électrons (en haut à droite), de l’énergie transverse manquante (en bas à gauche) et du discriminant de l’algorithme d’identification (TCHE) de jets b de plus haut p_T (en bas à droite), pour les événements de signal et de bruits de fond. Toutes les coupures de sélections sont appliquées, excepté la coupure d’identification des jets b et la coupure sur la variable représentée.	177
5.10	Similaire à la figure 5.9 pour le canal $\mu - \mu$	179
5.11	Dans le canal $e - \mu$, distribution en p_T du lepton de plus haut p_T (en haut à gauche), de l’énergie transverse manquante (en haut à droite) et du discriminant de l’algorithme d’identification (TCHE) de jets b de plus haut p_T (en bas), pour les événements de signal et de bruits de fond. Toutes les coupures de sélection sont appliquées, excepté l’identification des jets b et la coupure sur la variable représentée.	180
5.12	Après la sélection globale des événements et la factorisation des efficacités de sélection, en fonction de α , rapport signal sur bruit (en rouge) et fractions d’événements QCD (en bleu), W +jets(en vert) et $t\bar{t}$ semi-leptoniques (en noir) contenu dans l’ensemble des événements bruits de fond ($\times 5$).	187
5.13	Pour la sélection <i>loose</i> , distributions de multiplicité de jets pour les événements de type signal (en haut à gauche), de type W +jets (en haut à droite) et de type QCD (en bas). Les hachures correspondent aux erreurs statistiques du Monte-Carlo.	193
5.14	Pour la sélection <i>medium</i> , distributions de multiplicité de jets pour les événements de type signal (en haut à gauche), de type W +jets (en haut à droite) et de type QCD (en bas). Les hachures correspondent aux erreurs statistiques du Monte-Carlo.	194
5.15	Pour la sélection <i>tight</i> , distributions de multiplicité de jets pour les événements de type signal (à gauche), de type W +jets (à droite). Les hachures correspondent aux erreurs statistiques du Monte-Carlo. Aucun événements de type QCD ne passe les coupures.	195

5.16	Efficacité de passer de la sélection <i>loose</i> à la sélection <i>medium</i> $\varepsilon^{l \rightarrow m}$ pour les événements de type signal (en haut à gauche), de type W +jets (en haut à droite) et de type QCD (en bas).	196
5.17	Efficacité de passer de la sélection <i>loose</i> à la sélection <i>tight</i> $\varepsilon^{l \rightarrow t}$ pour les événements de type signal (à gauche) et de type W +jets (à droite).	197
5.18	Distributions des 3000 solutions de la méthode de la matrice pour 3000 pseudo-expériences dans le cas de l'échantillon <i>loose</i> , pour les événements de type signal (en haut à gauche), W +jets (en haut à droite) et QCD (en bas).	198
5.19	Distributions des 3000 solutions de la méthode de la matrice pour 3000 pseudo-expériences dans le cas de l'échantillon <i>tight</i> , pour les événements de type signal (en haut à gauche), W +jets (en haut à droite) et QCD (en bas).	199
5.20	Le graphique de gauche représente la distribution en trois dimensions des incertitudes relatives du taux de faux électrons isolés (axe des x), de l'efficacité de sélection de vrais électrons isolés (axe de y) et du nombre d'événements <i>loose</i> prédit par la méthode de la matrice (axe z des couleurs) pour les événements de type signal. Le graphique de droite représente les mêmes quantités pour les événements de type W +jets.	200
5.21	Corrélations entre N_S^l et ε_s (à gauche) et ε_{fake} (à droite). Les facteurs de corrélation sont de -0.99 (à gauche) et -0.06 (à droite).	201
5.22	Corrélations entre N_W^l et ε_s (à gauche) et ε_{fake} (à droite). Les facteurs de corrélation sont de 0.98 (à gauche) et 0.07 (à droite).	201
5.23	Corrélations entre N_{QCD}^l et ε_s (à gauche) et ε_{fake} (à droite). Les facteurs de corrélation sont de 0.99 (à gauche) et 0.03 (à droite).	202
5.24	Pour la sélection <i>loose</i> , multiplicité de jets pour les événements de type signal (en haut à gauche), W +jets (en haut à droite) et QCD (en bas). Les hachures correspondent aux erreurs statistiques du Monte-Carlo. Les points correspondent aux prédictions de la méthode de la matrice, les erreurs statistiques et systématiques sont représentées par les barres d'erreur.	203
5.25	Pour la sélection <i>medium</i> , multiplicité de jets pour les événements de type signal (en haut à gauche), W +jets (en haut à droite) et QCD (en bas). Les hachures correspondent aux erreurs statistiques du Monte-Carlo. Les points correspondent aux prédictions de la méthode de la matrice, les erreurs statistiques et systématiques sont représentées par les barres d'erreur.	204
5.26	Pour la sélection <i>tight</i> , multiplicité de jets pour les événements de type signal (en haut à gauche), W +jets (en haut à droite) et QCD (en bas). Les hachures correspondent aux erreurs statistiques du Monte-Carlo. Les points correspondent aux prédictions de la méthode de la matrice, les erreurs statistiques et systématiques sont représentées par les barres d'erreur.	205

5.27	Distributions en masse invariante de candidats di-leptons après la sélection présentée dans la section 5.2.6 sans identification de jets b . La fonction d'ajustement utilisée correspond à l'addition d'un polynôme d'ordre 3 et d'une fonction de <i>Breit – Wigner</i> pour le canal $e - e$ (à gauche) et le canal $\mu - \mu$ (à droite).	208
5.28	Pour le canal $(e - e)$, distributions de la masse invariante pour tous les événements (signal et bruit de fond) sommés avant (à gauche) et après (à droite) l'ajustement d'une fonction de type " <i>Cruijff + Breit – Wigner</i> ".	209
5.29	Pour le canal $(\mu - \mu)$, distribution de la masse invariante pour tous les événements (signal et bruit de fond) sommés avant (à gauche) et après (à droite) l'ajustement d'une fonction de type " <i>Cruijff + Breit – Wigner</i> ".	210
5.30	Pour le canal $e - e$, distributions de multiplicité de jets par événement après la sélection sans identification des jets b (en haut à gauche) ou avec au moins 1 jet b (en haut à droite) ou 2 jets b (au milieu) identifiés, pour une luminosité intégrée de 100 pb^{-1}	213
5.31	Identique à la figure 5.30 pour le canal $e - \mu$	214
5.32	Identique à la figure 5.30 pour le canal $\mu - \mu$	214
5.33	Dans le canal $e - \mu$, nombre d'événements sélectionnés après la génération de 70 000 événements $t\bar{t}$ di-leptons pour différentes masses du top. Les points sont ajustés par un polynôme d'ordre 2.	221
A.1	Dans les échantillons <i>QCD</i> et avec le point de fonctionnement <i>loose</i> avec l'application du veto sur les jets : les jets étiquetés négativement sont rejetés s'ils possèdent au moins une trace avec $IP/\sigma_{IP} > 4$, en fonction du p_T des jets : (en haut à gauche) les taux de mauvaise identification $\varepsilon^{MC}(udsg)$ (cercles), $\varepsilon^{MC}(uds)$ et $\varepsilon^{MC}(g)$ (triangles) et l'efficacité d'étiquetage négatif $\varepsilon_{neg}^{MC}(udsg + c + b)$ pour toutes les saveurs (carrés), (en haut à droite) le facteur R_{light} (cercles) et le rapport $\varepsilon^{MC}(uds)/\varepsilon^{MC}(g)$ (triangles), (en bas à gauche) les fractions de jets b , c et g avant (symbole creux) et après étiquetage négatif (symboles pleins) et (en bas à droite) les erreurs systématiques provenant de l'incertitude sur les fractions de jets g , c et b	229
A.2	Identique à la figure A.2 en fonction de η	230

A.3	Dans les échantillons <i>QCD</i> et avec le point de fonctionnement <i>loose</i> avec l'application du veto sur les jets : les jets étiquetés négativement sont rejetés s'ils possèdent au moins une trace avec $IP/\sigma_{IP} > 4$, en fonction du p_T des jets : (en haut à gauche) les taux de mauvaise identification $\varepsilon^{MC}(udsg)$ (cercles), $\varepsilon^{MC}(uds)$ et $\varepsilon^{MC}(g)$ (triangles) et l'efficacité d'étiquetage négatif $\varepsilon_{neg}^{MC}(udsg + c + b)$ pour toutes les saveurs (carrés), (en haut à droite) le facteur R_{light} (cercles) et le rapport $\varepsilon^{MC}(uds)/\varepsilon^{MC}(g)$ (triangles), (en bas à gauche) les fractions de jets b , c et g avant (symbole creux) et après étiquetage négatif (symboles pleins) et (en bas à droite) les erreurs systématiques provenant de l'incertitude sur les fractions de jets g , c et b	231
A.4	Identique à la figure A.4 en fonction de η	232
A.5	Distribution du discriminant de l'algorithme <i>TCHE</i> pour différentes origines des jets.	232
A.6	Pour le point de fonctionnement <i>medium</i> et l'algorithme <i>TCHE</i> , efficacités d'identification des jets légers et d'étiquetage négatif (à gauche), rapport R_{light} au milieu et erreurs systématiques principales (à droite) en fonction du p_T (en haut) et du $ \eta $ (en bas) des jets. Aucun veto sur les jets n'est appliqué.	233
A.7	Pour le point de fonctionnement <i>medium</i> et l'algorithme <i>TCHE</i> , efficacités d'identification des jets légers et d'étiquetage négatif (à gauche), rapport R_{light} au milieu et erreurs systématiques principales (à droite) en fonction du p_T (en haut) et du $ \eta $ (en bas) des jets. Un veto sur les jets est appliqué : les jets étiquetés négativement sont rejetés s'ils possèdent au moins une trace avec $S_{IP} > 4$	234
A.8	Distribution du discriminant de l'algorithme de probabilité par jet pour différentes origines des jets.	235
A.9	Pour le point de fonctionnement <i>medium</i> et l'algorithme de probabilité par jet, efficacités d'identification des jets légers et d'étiquetage négatif (à gauche), rapport R_{light} au milieu et erreurs systématiques principales (à droite) en fonction du p_T (en haut) et du $ \eta $ (en bas) des jets. Aucun veto sur les jets n'est appliqué.	236
A.10	Pour le point de fonctionnement <i>medium</i> et l'algorithme de probabilité par jet, efficacités d'identification des jets légers et d'étiquetage négatif (à gauche), rapport R_{light} au milieu et erreurs systématiques principales (à droite) en fonction du p_T (en haut) et du $ \eta $ (en bas) des jets. Un veto sur les jets est appliqué : les jets étiquetés négativement sont rejetés s'ils possèdent au moins une trace avec $IP/\sigma > 4$	237
A.11	Distribution du discriminant de l'algorithme de vertex secondaire pour différentes origines de vertex.	238

A.12	Pour le point de fonctionnement <i>medium</i> et l'algorithme du vertex secondaire, efficacités d'identification des jets légers et d'étiquetage négatif (à gauche), rapport R_{light} au milieu et erreurs systématiques principales (à droite) en fonction du p_T (en haut) et du $ \eta $ (en bas) des jets. Aucun veto sur les jets n'est appliqué.	239
A.13	Distribution du discriminant de l'algorithme de vertex secondaire combiné pour différentes origines de vertex.	240
A.14	Pour le point de fonctionnement <i>medium</i> et l'algorithme du vertex secondaire combiné, efficacités d'identification des jets légers et d'étiquetage négatif (à gauche), rapport R_{light} au milieu et erreurs systématiques principales (à droite) en fonction du p_T (en haut) et du $ \eta $ (en bas) des jets. Aucun veto sur les jets n'est appliqué.	241
A.15	Pour le point de fonctionnement <i>medium</i> et l'algorithme de probabilité par jet, efficacités d'identification des jets légers et d'étiquetage négatif (à gauche), rapport R_{light} au milieu et erreurs systématiques principales (à droite) en fonction du p_T (en haut) et du $ \eta $ (en bas) des jets. Un veto sur les jets est appliqué : les jets étiquetés négativement sont rejetés s'ils possèdent au moins une trace avec $IP/\sigma > 4$	242

Liste des tableaux

1.1	Fermions élémentaires du Modèle Standard.	4
1.2	Champs de particules du Modèle Standard et les lagrangiens correspondants.	6
1.3	Table des différents nombres quantiques associés aux particules de matière : l'hypercharge (Y), l'isospin (T) et sa composante (T_3) et la charge électrique q	14
1.4	Sections efficaces de production de single top au Tevatron (collisions $p\bar{p}$ à 2 TeV) et au LHC (collisions pp à 14 TeV) [32, 33].	25
1.5	Nombre d'événements $t\bar{t}$ produits par jour au LHC pour différentes énergies de collision et pour différentes luminosités instantanées.	27
3.1	Efficacités de reconstruction des traces estimées dans les données et dans le Monte-Carlo pour les différents algorithmes après la sélection PRT . La dernière ligne correspond à la différence d'efficacité entre données et Monte-Carlo. Les erreurs statistiques sont mentionnées.	91
3.2	Efficacité de reconstruction des traces estimées dans les données et dans le Monte-Carlo pour les différents algorithmes après la sélection LHC décrite dans le texte. La dernière ligne correspond à la différence d'efficacité entre données et Monte-Carlo. Les erreurs statistiques sont mentionnées.	95
4.1	Points de fonctionnement et efficacités d'identification correspondantes pour les algorithmes de comptage de trace et de probabilité par jet. Seules les erreurs statistiques sont présentées pour des événements QCD $80 < \hat{p}_T < 120$ GeV/c.	127
4.2	133
4.3	134
4.4	153
5.1	Echantillons Monte-CarloALPGEN utilisés pour les événements $t\bar{t}+N$ jets. La première colonne correspond au processus (multiplicité de partons associées), la deuxième au nom de l'échantillon, les troisième et quatrième aux sections efficaces LO et NLO et la dernière au nombre d'événements générés.	165

5.2	Similaire au tableau 5.1 pour des événements Z +jets ALPGEN.	165
5.3	Similaire au tableau 5.1 pour des événements W +jets ALPGEN.	165
5.4	Similaire au tableau 5.1 pour les événements single top tW TOPREX.	166
5.5	Similaire au tableau 5.1 pour des événements di-bosons PYTHIA.	166
5.6	Echantillons Monte-Carlo PYTHIA utilisés pour les événements QCD	166
5.7	Efficacité de sélection des coupures d'isolation pour les leptons ayant une énergie (impulsion) transverse > 20 GeV($/c$) et issus de bosons W et de hadrons lourds dans les événements $t\bar{t}+0$ jets (soupe "Chowder").	175
5.8	Nombre d'événements de signal et de bruit de fond sélectionnés dans le canal $(e - e)$ après l'application cumulée des différentes coupures de sélection pour une luminosité intégrée de 100 pb^{-1} . La dernière colonne présente les rapports signaux sur bruit. Les erreurs statistiques sont le nombre d'événements Monte-Carlo sont présentés.	181
5.9	Identique au tableau 5.8 pour le canal $(e - \mu)$	181
5.10	Identique au tableau 5.8 pour le canal $(\mu - \mu)$	182
5.11	Nombres d'événements QCD normalisés à une luminosité de 100 pb^{-1} passant les coupures de sélection estimées par factorisation, pour différents intervalles en \hat{p}_T	186
5.12	Nombre d'événements prédits par la méthode de la matrice et erreurs systématiques liées aux erreurs statistiques des lots <i>loose</i> , <i>medium</i> et <i>tight</i> et aux erreurs sur ε_s et ε_{fake} . Le nombre d'événements "vrais" déterminé par l'information du générateur est également indiqué.	206
5.13	Identique au tableau 5.12 lorsque la coupure en masse invariante n'est pas appliquée.	206
5.14	Identique au tableau 5.12 lorsque la coupure en \cancel{E}_T n'est pas appliquée.	207
5.15	Identique au tableau 5.12 lorsque ni la coupure en masse invariante ni la coupure en \cancel{E}_T ne sont appliquées.	207
5.16	Prédiction de la méthode, vérité Monte-Carlo et incertitudes systématiques relatives sur les paramètres d'ajustement, le facteur d'échelle et la résolution sur l'impulsion des leptons pour les canaux $e - e$ et $\mu - \mu$	211
5.17	Pour le canal $e - e$, rapport signal sur bruit, efficacité de sélection des événements de signal et incertitude statistique correspondante pour une luminosité intégrée de 100 pb^{-1}	212
5.18	Identique au tableau 5.17 pour le canal $e - \mu$	212
5.19	Identique au tableau 5.17 pour le canal $\mu - \mu$	212
5.20	Variation relative (en %) du nombre d'événements de signal sélectionnés après variation de l'échelle en énergie et de la résolution en p_T des jets pour les différents canaux et pour différentes sélections sur l'identification des jets b	217
5.21	Variations relatives (en %) du nombre d'événements de signal sélectionnés dues à l'erreur systématique sur l'identification des jets b	218

5.22	Variations relatives (en %) du nombre d'événements de signal sélectionnés pour l'erreur systématique sur Λ_{QCD}	219
5.23	Variations relatives (en %) du nombre d'événements de signal sélectionnés pour l'erreur systématique sur Q_{max}^2	219
5.24	Variations relatives (en %) du nombre d'événements de signal sélectionnés pour l'erreur systématique sur la fragmentation des quarks légers.	219
5.25	Identique au tableau 5.24 pour la fragmentation des quarks lourds.	220
5.26	Variations relatives (en %) du nombre d'événements de signal sélectionnés pour l'erreur systématique sur les événements sous-jacents.	220
5.27	Variations relatives (en %) du nombre d'événements de signal sélectionnés pour l'erreur systématique sur les PDFs.	221
5.28	Erreur systématique globale sur la mesure de la section efficace de production $t\bar{t}$	222
B.1	Nombre d'événements disponibles dans chaque échantillon utilisé dans l'analyse. Le terme <i>Soupe</i> se réfère à la somme des événements QCD et $b\bar{b}$ et $c\bar{c}$ avec des poids identiques pour chaque événement.	246

Bibliographie

- [1] S.L. Glashow. *Nucl. Phys.*, **22** 161 (579).
- [2] S. Weinberg. *Phys.Rev.Lett.*, **19** (1967) 1264.
- [3] S.L. Glashow, J. Iliopoulos, and L. Miami. *Phys.Rev.*, **D2** (1970) 1285.
- [4] S. Weinberg. *Phys.Rev.*, **D5** (1972) 1962.
- [5] D.J. Gross and F.Wilczek. *Phys.Rev.*, **D8** (1973) 3633.
- [6] D.J. Gross and F.Wilczek. *Phys.Rev.*, **D9** (1974) 980.
- [7] *Phys.Rept.*, **14** (1975) 129.
- [8] S.L. Glashow. *Nucl.Phys.*, **22** ,579 (1961).
- [9] A. Salam and J.C. Ward. *Phys. Rev.Lett.*, **13** , 168 (1964).
- [10] S. Weinberg. *Phys. Rev.Lett.*, **19** , 1264 (1967).
- [11] G Arnison et al. *Phys.Lett.B*, **122** ,103 (1983).
- [12] G Arnison et al. *Phys.Lett.B*, **126** ,398 (1983).
- [13] M. Banner et al. *Phys.Lett.B*, **122** , 476 (1983).
- [14] M. Banner et al. *Phys.Lett.B*, **129** , 130 (1983).
- [15] F Abe et al. *Phys.Rev.Lett.*, **74** , 2626 (1995).
- [16] A. DeAndrea. Interactions électrofaibles et introduction à la superSymétrie.
Cours.
- [17] Gell-Mann. *Phys.Rev.Lett.*, **8**, 214 (1964).
- [18] T.D. Lee and C.N. Yang. Question Of Parity Conservation In Weak Interactions.
Phys.Rev., **104**, 254 (1956).
- [19] C.S. WU, E Ambler, R.W. Hayward, D.D. Hoppes, and Hudson R.P. Experimental Test Of Parity Conservation in Beta Decay. *Phys. Rev.*, **109** , 1015 (1958).
- [20] [http ://lepewwg.cern.ch/LEPEWWG](http://lepewwg.cern.ch/LEPEWWG).
- [21] P.W. Higgs. *Phys.Lett.*, **12** , 132 (1964).
- [22] P.W. Higgs. *Phys.Rev.*, **13**, 508 (1964).
- [23] P.W. Higgs. *Phys.Rev.*, **145** , 1156 (1966).
- [24] F. Englert and R. Brout. *Phys.Rev.Lett.*, **13** , 321 (1964).

- [25] N. Cabibbo. *Phys.Rev.Lett.*, **10** , 531 (1963).
- [26] M. Kobayashi and T. Maskawa. *Prog.Theor.Phys.*, **49**, 652 (1973).
- [27] A. Djouadi. The Anatomy of Electr-Weak Summetry Breaking. *arXiv*, Tome II, hep-ph/0503173v2, 2005.
- [28] <http://lepewwg.web.cern.ch/LEPEWWG/>.
- [29] <http://tevwwg.fnal.gov>.
- [30] S. Eidelman et al. *Phys.Lett.B*, **592** , 1 (2004).
- [31] Roberto Bonciani, Stefano Catani, Michelangelo L. Mangano, and Paolo Nason. NLL resummation of the heavy-quark hadroproduction cross-section. *Nucl. Phys.*, B529 :424–450, 1998.
- [32] J. Campbel et al. Single top production and decay at next-to-leading order. arXiv :hep-ph/0408158, 2004.
- [33] J. Campbel et al. arXiv :hep-ph/0506289, 2005.
- [34] J. D’Hondt et al. Energy-Scale Calibration Using W -Mass Constraint in Single-Leptonic $t\bar{t}$ Events. *CMS Note*, 2006/025, 2006.
- [35] S. Lowette. B-Tagging as a Tool for Charged Higgs Boson Identification in CMS. *Thesis*, 2007/003, 2006.
- [36] S. Lowette, J. D’Hondt, and J. Heyninck. Offline calibration of b-jet identification efficiencies. CMSAN 2006/013, 2006.
- [37] J. D’Hondt et al. Measurement of flavour tagging efficiencies using top quark events in CMS. CMSAN 2007/030, 2007.
- [38] M. Baarmand, M. Hashemi, and A. Nikitenko. Light Charged Higgs Discovery Potentiel of CMS in the $H^\pm \rightarrow \tau\nu_\tau$ Decay with Single Lepton Trigger. *CMS Note*, 2006/056, 2006.
- [39] M. Erdmann et al. Measurement of the top quark pair invariant mass distribution. *CMS Analysis Note*, AN-2007/027, 2007.
- [40] S. Lowette, J. D’Hondt, and P. Vanlaer. Charged MSSM Higgs Boson Observability in the $H^\pm \rightarrow tb$ Decay. *CMS Note*, 2006/017, 2006.
- [41] N. Metropolis. The beginning of the Monte Carlo Method. *Los Alamos Science*, nř **15** ,p.125-130 (1987).
- [42] R. Brock et al. *Rev.Mod.Phys.*, **67** ,157-248 (1995).
- [43] A.D. Martin, W.J. Roberts, R.G.and Strling, and R.S. Thorne. *Eur.Phys.J.*, **C23**,73-87 (2002).
- [44] B. Andersson et al. *Z.Phys.*, **C20** (1983) 317.
- [45] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. PYTHIA 6.4 physics and manual. *JHEP*, 05 :026, 2006.

- [46] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, Roberto Pittau, and Antonio D. Polosa. ALPGEN, a generator for hard multiparton processes in hadronic collisions. *JHEP*, 07 :001, 2003.
- [47] <http://atlas.ch/>.
- [48] <http://cms.cern.ch/iCMS/jsp/page.jsp?mode=cms>.
- [49] <http://lhcb.web.cern.ch/lhcb/>.
- [50] <http://aliceinfo.cern.ch/>.
- [51] The CMS Collaboration, S. Chatrchyan, et al. The CMS experiment at CERN LHC. *JINST*, 3 S08004 : <http://www.iop.org/EJ/abstract/1748-0221/3/08/S08004>, 2008.
- [52] Physics Technical Design Report, Vol. 1 : Detector Performance and Software. *CERN/LHCC*, 2006-001, 2006.
- [53] W. Adam, B. Mangano, T. Speer, and T. Todorov. Track Reconstruction in CMS Tracker. *CMS Note*, 2006/041, 2006.
- [54] P. Astier et al. Kalman Filter Track Fits and Track Breakpoint Analysis. *arXiv*, physics/9912034v1 (1999).
- [55] T. Speer, K. Prokofiev, R. Fruhwirth, W. Waltenberger, and P. Vanlear. Vertex Fitting in the CMS Tracker. *CMS Note*, 2006/032, 2006.
- [56] T. Muller, C. Piasecki, G. Quast, and C. Weiser. Inclusive Secondary Vertex Reconstruction in jet. *CMS Note*, 2006/027, 2006.
- [57] D. Acosta et al. Measuring Muon Reconstruction Efficiency from Data. *CMS Note*, 2006/060, 2006.
- [58] D. Acosta et al. Measuring Muon Reconstruction Efficiencies from Data. *CMS Note*, 2006/060, 2006.
- [59] E. Meschi et al. Electron Reconstruction in the CMS Electronmagnetic Calorimeter. *CMS Note*, 2001/034, 2001.
- [60] CMS collaboration. Trigger and Data Acquisition Project. *CERN/LHCC*, volume 1.
- [61] S. Bafioni et al. Electron reconstruction in CMS. *CMS Note*, 2006/040, 2006.
- [62] A. Heister et al. Measurement of Jets with the CMS Detector at the LHC. *CMS Note*, CMSNOTE 2006/036, 2006.
- [63] R. Harris. Jet Calibration from Di-Jet Balancing. *CMS Analysis Note*, 2005/034, 2005.
- [64] V. Konoplyanikov et al. Jet Calibration Using γ +jets Events in the CMS detector. *CMS Analysis Note*, 2006/042, 2006.
- [65] P. Adzic et al. Intercalibration of the barrel electromagnetic calorimeter of the CMS experiment at start-up. *CMS Note*, 2008/018, 2008.

- [66] M. Vázquez Acosta et al. Jet MET performance in CMSSW_1_2_0. *CMS Internal Note*, 2007/053, 2007.
- [67] E.A. Albayrak et al. E_T Performance in CMS. *AN Note*, 2007/041, 2007.
- [68] W.H. Smith et al. The TriDAS Project : Technical Design Report Volume 1, The Trigger Systems. *CERN/LHCC*, 2000-38, 2000.
- [69] B. Stroupstrup. Programming : Principles and Practice Using C++. *Addison-Wesley*.
- [70] <http://root.cern.ch/>.
- [71] S Lowette et al. Top Quark Analysis Framework. *Internal Note*, 2007/068, 2007.
- [72] *Nuclear Instruments and Methods in Physics Research*, **1506** , 250-303 (2003).
- [73] <https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideFrameWork>.
- [74] <https://twiki.cern.ch/twiki/bin/view/CMS/TWikiTopQuarkFramework>.
- [75] W. Adam et al. CMS Tracker Alignment at the Integration Facility. *CMS Note*, 2009/02, 2009.
- [76] P. se Barbaro et al. Validation of HCAL energy calibration using high statistic cosmic ray data. *CMS Internal Note*, 2008/025, 2008.
- [77] C. Adzic et al. Intercalibration of the barrel electromagnetic calorimeter of the CMS experiment at start-up. *CMS Internal Note*, 2008/018, 2008.
- [78] W. Adam et al. Track Reconstruction with Cosmic Ray Data at Tracker Integration Facility. *CMS Note*, 2009/03, 2009.
- [79] P. Biallas, T. Hebbeker, and K. Hoepfer. Simulation of cosmic Muons and Comparison with Data from the Cosmic challenge using Drift Tube Chambers. *CMS Note*, 2007/024, 2007.
- [80] P. Biallas and Hebbeker. Improved Parametrization of the Cosmic Muon Flux for the Generator CMSCGEN. *CMS Internal Note*, 2007/033, 2007.
- [81] C. Saout, A. Scheurer, F.P. Schilling, and A. Schmidt. Update of Misalignment Scenarios for the CMS Tracker. *CMS Analysis Note*, 2007/047, 2007.
- [82] V.H. Bazterra et al. Track Selection for Life-time base b-jet tagging algorithms. *CMS Analysis Note*, 2008/088, 2008.
- [83] V.E. Bazterra, C. Gerber, J. Andrea, and D. Bloch. Track Selection for Lifetime-based b-jet tagging algorithms. *CMS AN2008/088*, 2008.
- [84] N. de Filippis, T. Lampen, F.-P. Schilling, A. Schmidt, and M. Weber. Impact of Tracker Misalignment on the CMS b-Tagging Performance. *CMS Internal Note*, 2007/036, 2007.
- [85] A. Rizzi, F. Palla, and G. Segneri. Track impact parameter based b-tagging with CMS. *CMS Note*, 2006/019, 2006.
- [86] ALEPH Collab. *Phys. Lett.*, **B313**, 535, 1993.

- [87] G. Borisov and C. Mariotti. *Nucl. Instr. & Meth.*, **A372**, 181, 1996.
- [88] DELPHI Collab. *E. Phys.*, **C10**, 415, 1999.
- [89] C. Weiser. A Combined Secondary Vertex Based B-Tagging Algorithm in CMS. *CMS Note*, 2006/014, 2006.
- [90] A. Bocci, P. Demin, and S. Ranieri, R. de Visscher. Tagging b jets with electrons and muons CMS. *CMS Analysis Note*, 2006/043, 2007.
- [91] Benoit Clément. *PhD. thesis, University Louis Pasteur*, http://www-d0.fnal.gov/results/publications_talks/thesis/clement/thesis.pdf, 2006.
- [92] V.M.Abazov et al. *Phys. Lett. B*, **626**, 35, 2005.
- [93] V.M.Abazov et al. *Phys. Rev. D*, **74**, 112004, 2006.
- [94] Meenakshi Narain, Ping Tan, Francisco Yumiceva, Jeremy Andrea, Daniel Bloch, Denis Gelé, Pierre Juillot, Victor Bazterra, and Cecilia Gerber. Performance Measurement of b tagging Algorithms Using Data containing Muons within Jets. *CMS Analysis Note*, 2007/046, 2007.
- [95] F. Yumiceva et al. Measurement of the btagging Performance using a Muon in Jet Sample : Update of the System8 and pTrel Methods using CSA07 simulation data. *CMS Analysis Note*, 2008/081, 2008.
- [96] CMS High Level Trigger. *CERN/LHCC*, 2007-021, 2007.
- [97] Jeremy Andrea, Daniel Bloch, Denis Gelé, Pierre Juillot, Victor Bazterra, Cecilia Gerber, and Francisco Yumiceva. Evaluation of udsg Mistag Rate of b-tag Jets using Negative Tags. *CMS Analysis Note*, 2007/048, 2007.
- [98] V.P. Andreev, D.B. Cline, and S. Otwinowski. Measurement of open beauty production at LHC with CMS. *CMS Note* 2006/120.
- [99] W.-M. Yao et al. *Journal of Physics*, **G 33** p.199, 2006.
- [100] A. Abulencia et al. A search for $t \rightarrow \tau\nu q$ in $t\bar{t}$ production. *Phys. Lett.*, **B639** :172, 2006.
- [101] F. Abe et al. The mu tau and e tau decays of top quark pairs produced in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV. *Phys. Rev. Lett.*, **79** :3585–3590, 1997.
- [102] Physics Technical Design Report, Vol. 2 : Physics Performance. *J. Phys G : Nucl. Part. Phys.*, **34** :995–1579, 2007.
- [103] M. Davids et al. Measurement of top-pair cross section and top-quark mass in the di-lepton and full-hadronic channels with CMS. *CMS Note*, 2006/077, 2006.
- [104] J. Andrea et al. Measurement of the Top Quark Pair Production Cross Section with $\mathcal{L} = 100\text{pb}^{-1}$ Dileptonic Final States. *CMS Analysis Note*, 2008/016, 2008.
- [105] J. D’Hondt, J. Heyninck, and S. Lowette. Measurement of the cross section of single leptonic $t\bar{t}$ events. *CMS Note*, 2006/064, 2006.

- [106] J. D'Hondt et al. Electron and muon reconstruction in single leptonic $t\bar{t}$ events. *CMS Note*, 2006/024, 2006.
- [107] J. D'Hondt, S. Lowette, O. Buchmuller, S. Cucciarelli, F.-P. Schilling, M. Spiropulu, S.P. Mehdibadi, D. Benedetti, and L. Pape. Fitting of Event Topologies with External Kinematic Constraints in CMS. *CMS Note*, 2006/023, 2006.
- [108] O. Buchmuller, L. Edera, M. Mulders, R. Ranieri, D. Tsirigkas, J. Caudron, A. Giammanco, D. Kcira, V. Lemaitre, and F.-P. Schilling. Early measurements of top quark pair events with the first data of CMS. *CMS Analysis Note*, 2007/022, 2007.
- [109] J. D'Hondt, J. Heyninck, and S. Lowette. Top Quark mass measurement in single-leptonic $t\bar{t}$ events. *CMS Note*, 2006/066, 2006.
- [110] M. Erdmann et al. Measurement of the top quark pair invariant mass distribution. *CMS Analysis Note*, 2007/027, 2007.
- [111] Jorgen D'Hondt et al. Measurement of Flavour Tagging Efficiencies using Top Quark Events in CMS. *CMS Analysis Note*, 2007/030, 2007.
- [112] <https://twiki.cern.ch/twiki/bin/view/Main/Prod06>.
- [113] <https://twiki.cern.ch/twiki/bin/view/Main/AlpgenSummer07>.
- [114] <https://twiki.cern.ch/twiki/bin/view/CMS/TWikiTopQuarkCSA07>.
- [115] N. Adam et al. Towards a measurement of the inclusive $W \rightarrow \mu\nu$ and $Z \rightarrow \mu^+\mu^-$ cross sections in pp collisions at $\sqrt{s} = 14$ TeV. *CMS Analysis Note*, 2007/031, 2007.
- [116] S.R. Slabospitsky. The toprex generator manual. arXiv :hep-ph/020129201, 2002.
- [117] G. Daskalakis et al. Measuring Electron Efficiencies at CMS with Early Data. *CMS Analysis Note*, 2007/019, 2007.
- [118] N. Adam et al. Measurement of Inclusive $W \rightarrow e\nu$, $Z \rightarrow e^+e^-$ Cross Section in pp Collisions at $\sqrt{s} = 14$ TeV. *CMS Analysis Note*, 2007/026, 2007.
- [119] M. Bona et al. The $Z \rightarrow \mu\mu$ +jets data candle. *CMS Analysis Note*, 2008/095, 2008.
- [120] <https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideFastSimulation>.
- [121] P. Bartalini, R. Chierici, and A. De Roeck. Guidelines for the estimation of theoretical uncertainties at the LHC. *CMS Note*, 2005/013, 2005.
- [122] G. Altarelli and G. Parisi. *Nucl.Phys.*, **B126** (1977) 298.
- [123] J. D'Hondt, S. Lowette, G. Hammad, J. Heyninck, and P. Van Mulders. Observability of same-charge lepton topology in di-leptonic $t\bar{t}$ events. *CMS Note*, CMSNOTE 2006/065, 2006.
- [124] S. Lowette, J. D'Hondt, J. Heyninck, and P. Vanlaer. Offline Calibration of b-Jet Identification Efficiencies. *CMS Note*, 2006/013, 2006.

-
- [125] Jorgen D'Hondt et al. Measurement of Jet Energy Scale corrections using Top Quark Events. *CMS Analysis Note*, 2007/029, 2007.
- [126] S. R. Slabospitsky and L. Sonnenschein. TopReX generator (version 3.25) : Short manual. *Comput. Phys. Commun.*, 148 :87–102, 2002.



Institut Pluridisciplinaire Hubert Curien

Service de Reprographie
Michel Schlée

23, rue du Loess
67037 Strasbourg Cedex 2