

Tagging Boosted Objects at the LHC

Résumé

Pendant ce stage nous nous sommes intéressés à la phénoménologie au LHC. Il s'agissait de s'attarder sur les techniques permettant d'améliorer le traitement des données expérimentales au LHC dans le cas d'évènements comportant un objet boosté, typiquement un Z, un W ou un H. En effet les énergies atteintes au LHC change la donne, au LEP et au Tevatron les bosons électrofaibles ainsi que les tops étaient des particules lourdes peu susceptibles d'être détectées avec une grande impulsion transverse, et les techniques d'analyse des évènements utilisées précédemment se trouve maintenant devant de nouveaux problèmes pour traiter ces cas. Le but de ce stage aura alors été d'étudier ces difficultés, et de proposer des alternatives, aux techniques précédemment utilisées pour l'analyse des évènements. Nous avons pu construire une méthode originale pour discriminer le Higgs boosté du bruit QCD en s'intéressant aux connections de couleur des diagrammes de désintégration d'un Higgs et d'un gluon.

Mots clefs : Higgs boosté, sous-structure jet, connections couleurs, quark bottom

Stage encadré par :

Gregory Soyez

gregory.soyez@cea.fr

IPhT

Institut de Physique Théorique

Orme des mérisiers bâtiment 774

Point courrier 136

CEA/DSM/IPhT, CEA/Saclay

F-91191 Gif-sur-Yvette Cedex

<http://ipht.cea.fr/>

Remerciements

Je voudrais commencer par remercier mon encadrant de stage Gregory Soyez pour m'avoir pris en stage, avoir été très disponible et accepté de discuter parfois de très long moments de mes problèmes conceptuels pourtant sans grands intérêts pratiques, ainsi que pour m'avoir laissé très libre dans mon travail, et permis de découvrir plusieurs aspects de la QCD. Je suis aussi reconnaissant à Edmond Iancu et Gavin Salam que j'ai pu rencontrer à Genève en février et grâce à qui j'ai pu effectuer ce stage. Je les remercie pour le temps qu'ils m'ont consacré le jour de ma venue et par la suite.

Je voudrais aussi remercier les thésards de l'IPhT, pour leur accueil dès mon arrivée, une mention spéciale ira à Romain et Nicolas pour m'avoir supporté dans leur bureau. Une autre mention ira à la voiture de Julien pour avoir porté l'usure supplémentaire due à ma présence. Et évidemment je remercie les autres membres du laboratoire avec qui j'ai pu discuter. Merci pour cette très bonne ambiance de travail mais aussi pour la bonne ambiance générale.

Table des matières

| | | |
|----------|---|-----------|
| 1 | Position du problème. | 3 |
| 1.1 | Phénoménologie QCD | 3 |
| 1.2 | Un Higgs léger et beaucoup de bruit. | 5 |
| 2 | Solution proposée. | 6 |
| 2.1 | Arguments heuristiques. | 6 |
| 2.2 | Calculs en QCD perturbative. | 9 |
| 2.2.1 | Cas du Higgs. | 9 |
| 2.2.2 | Cas du gluon. | 13 |
| 3 | Tests avec PYTHIA | 17 |
| 3.1 | Comparaison PYTHIA/Calculs analytiques. | 17 |
| 3.2 | Cut en ν et ω | 19 |
| 3.3 | Une variable robuste | 21 |
| 4 | Conclusion. | 22 |

Introduction

Le LHC est maintenant en fonctionnement réel depuis environ deux ans. Il ne fonctionne pas encore à pleine puissance, cependant comme disent les expérimentateurs : “La machine fonctionne très bien”. À l’heure actuelle elle permet la collisions de faisceaux de proton et d’anti-protons à une énergie dans le centre de masse $\sqrt{s} = 7$ TeV. Le LHC est donc le premier accélérateur à explorer le régime multi-TeV. Sa luminosité n’a beau pas être celle de son fonctionnement nominal la quantité de données recueillies est gigantesque et égal voir dépasse la quantité de données recueillies pendant le fonctionnement du Tevatron. À la conférence EPS qui s’est tenu à Grenoble fin juillet 2011, les expérimentateurs s’attendaient à pouvoir décider de l’existence du Higgs à la fin de l’année 2012. En effet, ceux-ci s’attendent à une augmentation d’un facteur dix dans la quantité de données accumulées au cours de l’année 2012. Mais ces performances exceptionnelles ne sont pas sans nouvelles difficultés. Le LHC est, comme son nom l’indique, un collisionneur hadronique, le bruit de fond QCD est donc beaucoup plus important qu’au LEP et rend le traitements des résultats expérimentaux plus difficile. Ce bruit existait certes déjà au Tevatron mais la luminosité et l’énergie atteinte au LHC rendent le problème beaucoup plus difficile. À chaque interaction des deux faisceaux de p et de \bar{p} la probabilité de collisions multiples est plus grande qu’au Tevatron, du fait de la densité des paquets de protons.

L’expérience acquise au LEP et surtout au Tevatron, a permis de mieux maîtriser la QCD, théorie dont la compréhension précise est essentielle à l’utilisation de collisionneur hadronique, mais aussi de préparer des techniques d’analyses des données. Des travaux récents ont donné des méthodes de reconstruction des événements fiables [12, 4, 11]. Cependant, une meilleure compréhension de celles-ci reste nécessaire. De plus de nouveaux défis apparaissent. Des objets lourds, en particulier les bosons électrofaibles sont susceptibles d’être produits avec une grande impulsion transverse, ce sont des objets boostés. L’identification de ces objets est rendu difficile car leurs produits de désintégration sont très collimatés dans le référentiel du laboratoire. Cependant leur étude est intéressante car ils permettent de simplifier certains problèmes de bruit de fond¹[1].

Pendant ce stage je me suis donc intéressé à ces objets boostés, en particulier le Higgs boosté, dans le but d’améliorer leur identification. Nous avons pu trouver une piste intéressante pour améliorer le traitement de ce type d’évènements. Je commencerai donc par présenter des aspects généraux de la phénoménologie, ainsi que le problème plus spécifique du canal de désintégration dominant d’un Higgs léger. Ensuite viendra une présentation d’une solution originale qui pourrait être utilisée. Enfin nous aurons l’occasion de tester cette idée à l’aide d’un logiciel Monte Carlo : PYTHIA, ainsi que de présenter quelques résultats qui ne sont que préliminaires, mais qui semble montrer que la voie est intéressante.

1. Principalement des événements $t\bar{t}$ avec une désintégration du W trop proche de l’axe du faisceau.

1 Position du problème.

1.1 Phénoménologie QCD

Lors d'un évènement dans un accélérateur hadronique² on observe des états finaux qu'il est difficile de relier au processus dur qui a lieu au niveau partonique. En effet la QCD est une théorie confinante, et l'on observe pas directement les partons, mais des particules, les hadrons, qui sont des états liés de ces mêmes partons. Du fait des divergences collinéaires, on observe ces hadrons en gerbes collimatées : les jets. Pour reconstituer l'évènement dur partonique il a fallu développer un ensemble de techniques permettant de remonter le processus de formation de jets. Cependant le manque de connaissances techniques sur la QCD (qui présente toutes les difficultés, théorie quantique des champs, avec des interactions fortement non linéaires, un groupe de jauge non abélien, une constante de couplage forte...), rend impossible l'analyse exacte de la formation des jets. Ont alors été développés des algorithmes sur des bases heuristiques et pratiques permettant de reconstituer au mieux l'évènement partonique à partir des états finaux observés. Cependant nombre de ces algorithmes n'étaient pas cohérent vis à vis de la QCD. Ceci est lié à la difficulté qu'il y a à définir un jet. Il est difficile de trouver une définition de jets qui soit à la fois utilisable expérimentalement et théoriquement. La première définition de jet cohérente vis à vis de la QCD a été proposée par Sterman et Weinberg [15], celle-ci est : " Un évènement est classé comme un évènement à deux jets si au moins une fraction $1 - \epsilon$ de l'énergie de l'évènement est contenue dans deux cônes d'ouverture angulaire 2δ ". Seulement cette définition a été donnée dans le cadre des collisions $e^+ e^-$. Les évènements dans ces accélérateurs sont beaucoup mieux maîtrisés, l'énergie de chaque collision est connue. Dans les collisionneurs hadroniques ce n'est pas le cas, si l'on connaît l'énergie portée par un proton on ne connaît pas celle portée par un parton, on ne connaît que la distribution en énergie à travers les PDF³ [6].

Actuellement, les jets sont définis à l'aide d'un ensemble de règle qu'un ordinateur peut utiliser pour, à partir d'une liste des impulsions des particules d'un évènement, donner une liste des jets [11]. C'est une définition algorithmique, et chaque définition donne à priori un ensemble de jets différents. Il existe deux principaux types d'algorithmes : les algorithmes de cônes et les algorithmes de recombinaison. On peut en résumer le fonctionnement ainsi : pour un évènement donné l'algorithme cherche la particule la plus dure (i.e. celle avec la plus grande impulsion) c'est la particule de "seed". On forme alors un cône de rayon $R = \sqrt{\delta\phi^2 + \delta y^2}$ autour du seed⁴, puis l'on détermine la direction de la somme des impulsions de toutes les particules contenues dans ce cône. Si cette direction ne coïncide pas avec la direction de la particule de seed originale, alors on utilise cette nouvelle direction comme seed, l'on trace un nouveau cône et calcule à nouveau le seed associé à ce nouveau cône etc... Jusqu'à ce que l'on obtienne un cône stable. On a alors obtenu un jet au sens des algorithmes de cônes. Cependant les algorithmes de cônes

2. C'est vrai aussi dans les accélérateurs leptonique cependant ces derniers sont plus propres.

3. PDF = parton distribution function, elles donne la probabilité qu'un parton de type i porte une fraction x de l'impulsion longitudinale du proton

4. y , ϕ la rapidité et l'azimut respectivement.

ne sont pas stables au niveau collinéaire, c'est à dire que le résultat de l'algorithme dépend de l'ajout d'un gluon collinéaire. On le peut le voir simplement. Si l'on prend un évènement à deux jets qui au niveau partonique correspond à la création de deux quarks alors si l'on rajoute un gluon collinéaire la particule ayant la plus grande impulsion peut ne plus l'être après le splitting collinéaire. On obtient alors un évènement avec un ensemble de jets différents⁵. Au Tevatron a été utilisé une variante, cependant cette variante qui permettait d'éviter l'écueil collinéaire, était sensible aux gluons infrarouges. Un algorithme de cône sûr au niveau collinéaire et infrarouge a alors été développé, SIScone en évitant d'utiliser une particule de seed [12].

Un autre type d'algorithme est utilisé : les algorithmes de recombinaison. Ils définissent les jets à partir de la distance entre particules. Le choix de la distance est motivé par la physique du splitting de quark et de gluon. On choisit une distance qui permet de voir un gluon issu d'un splitting collinéaire comme très proche de son parent⁶. Pour former les jets, on calcule la distance entre chaque particule et on recombine en une seule pseudo-particule les paires de particules les plus proches. Puis on réitère la procédure avec les pseudo-particules jusqu'à ce que la distance entre toutes les pseudo-particules soit supérieure à une certaine valeur y_{cut} . Les pseudo-particules restantes définissent les jets de l'évènement. On peut alors définir un ensemble d'algorithme de recombinaison en changeant la distance utilisée. On a par exemple l'algorithme k_t , anti- k_t , k_t^{2p} etc... Pour l'exemple on peut expliquer un peu plus précisément l'algorithme de k_t . La distance entre particules est $d_{ij} = \min(p_{ti}^2, p_{tj}^2) \frac{\Delta R_{ij}^2}{R^2}$, avec $R_{ij} = \sqrt{(\phi_i - \phi_j)^2 + (y_i - y_j)^2}$, on introduit aussi une "distance par rapport au faisceau" $d_{if} = p_{ti}^2$. R joue un rôle analogue à celui qu'il joue dans les algorithmes de cône. En effet l'algorithme sélectionne la plus petite distance entre d_{ij} et d_{if} , si il s'agit de d_{ij} alors les particules i et j sont recombinaées, si il s'agit de d_{if} alors la particule i est appelée un jet, la procédure étant répétée jusqu'à ce qu'il n'y ait plus de particules. Si la distance ΔR_{ij} est plus petite que R alors d_{ij} est plus petite que d_{if} et les particules seront recombinaées, alors que si la distance entre la particule i et toutes les autres particules est plus grande que R alors la particule i est appelée à former un autre jet. Cependant la plupart des algorithmes de cônes ne sont pas stables vis à vis de l'émission de particules molles. Un autre écueil qui a pu être résolu est le temps de calcul de ces algorithmes, longtemps le temps de calcul était en $\mathcal{O}(N^3)$, N le nombre de particules, cependant des techniques élégantes de calcul géométrique ont permis de les implémenter en $\mathcal{O}(N \ln N)$ [3].

5. Ce qui pose problème car les divergences sont de trois types : virtuelle correspondant à l'émission d'un gluon virtuel, collinéaire et infrarouge correspondant respectivement à l'émission d'un gluon mou et collinéaire, et s'annulent entre elles, cependant si un évènement à un jet avec un gluon collinéaire est classifié comme évènement à deux jets alors les divergences virtuelles et collinéaires contribuent séparément aux sections efficaces à respectivement 1 et 2 jets et ne s'annulent plus.

6. La distance est simplement, dans l'algorithme de k_t , le dénominateur de l'élément de matrice de splitting collinéaire. Quand ce dénominateur est petit on a une divergence, mais la distance est petite et les particules sont recombinaées.

1.2 Un Higgs léger et beaucoup de bruit.

Les expériences conduites au LEP ont permis d'exclure une masse du Higgs inférieure à 115 GeV. Les résultats des runs du LHC, présentés à la conférence EPS en juillet dernier, ont permis d'exclure les masses du Higgs au-dessus de 139 GeV. On s'attend⁷ donc à observer un Higgs léger. Malheureusement, c'est aussi dans cette gamme de masse qu'il sera le plus difficile à découvrir. Les canaux qui deviennent intéressants dans cet intervalle de masse sont [1] : fusion de gluon suivi de $H \rightarrow \gamma\gamma$, fusion de W suivi de $Hjj \rightarrow \tau^+\tau^-jj$ et $VH \rightarrow Vb\bar{b}$ ⁸, ce dernier étant le canal dominant mais il est noyé dans le bruit de fond QCD. Aux énergies atteintes par le LHC on peut s'attendre à produire des particules autrefois considérées comme lourdes avec de grandes impulsions transverses, supérieure à la masse aux repos de ces particules, ce sont des objets boostés. Le canal $H \rightarrow b\bar{b}$ s'est révélé intéressant après l'étude de Butterworth, Davidson, Rubin et Salam [2], en s'assurant que le Higgs était produit avec une grande impulsion transverse et se désintégrait en une paire très collimatée de $b\bar{b}$. Ce canal devenant intéressant à $\sqrt{s} = 14$ TeV. L'un des problèmes qui survient dans la recherche du Higgs est la faiblesse du signal, celui-ci se retrouve noyé dans le bruit QCD. De plus le fait que les objets lourds puissent maintenant être boostés fait apparaître de nouvelles difficultés, ils sont moins différenciables des objets légers tels que les quarks ou les gluons (à ces énergies tous les bosons électrofaibles sont virtuellement sans masse), et l'étude de la sous-structure du jet est difficile, les jets étant très collimatés dans le référentiel du laboratoire.

Pour illustrer ce qu'il se passe en pratique nous avons placé une figure simplifiée (cf. Figure 1) de ce qui est observable dans le cas d'un Higgs boosté qui se désintègre en une paire $b\bar{b}$. Si l'on ne fait rien de précis dans le traitement des événements on observe la figure à gauche, la figure centrale correspond au traitement des événements proposé par Butterworth, Davidson, Rubin et Salam et la figure de droite correspond à l'utilisation complémentaire de notre résultat. Il est donc évident que bien discriminer le bruit de fond du signal est essentiel pour découvrir le Higgs.

Pour y arriver on procède en trois étapes, premièrement une étude de la sous-structure du jet par une analyse de Mass Drop permet de déterminer deux sous-jets candidats. Donné un jet de masse m , on identifie deux sous-jets en revenant à l'avant dernière étape de recombinaison, on obtient alors deux sous-jets 1 et 2 avec des masses respectives m_1, m_2 , où l'on choisit $m_1 > m_2$ sans perdre de généralité. Intervient le paramètre de mass drop⁹ : μ . Si le jet 1 vérifie la condition $m_1 < \mu m$ et que le splitting est assez symétrique, alors on considérera le jet initial comme issu du Higgs. On introduit alors le b-tagging, c'est-à-dire que l'on veut que les deux sous-jets 1 et 2 contiennent des hadrons B, que l'on détecte par leurs canaux de désintégration. Enfin le filtrage permet de traiter le problème du bruit, on veut éliminer la plus grande partie possible du pile-up et de l'underlying event¹⁰ afin d'obtenir la meilleure précision sur la masse du Higgs. Si

7. Si il existe bien, sait-on jamais...

8. V étant un boson de jauge électrofaible.

9. Si l'on regarde la désintégration d'un Higgs au repos en Mercedes, une valeur de $\mu = 1/\sqrt{3}$ assure qu'il passera la conditions.

10. Le pile-up est le phénomène d'évènements multiples à chaque collision des paquets de proton au

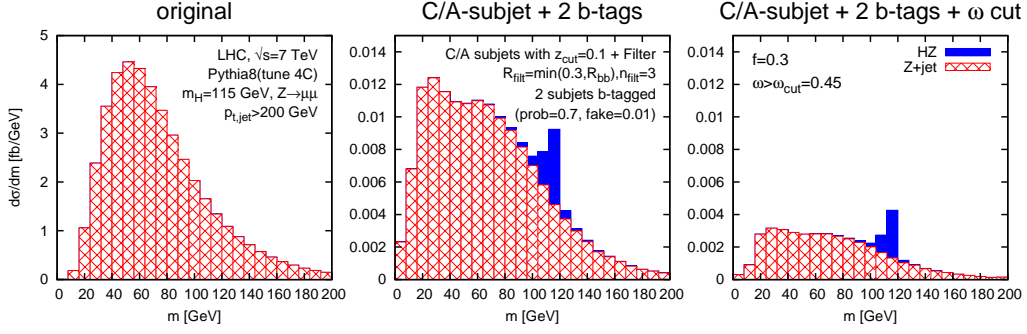


FIGURE 1 – Ce graphe représente de manière très simplifiée le signal que l’on devrait observer expérimentalement lors de la recherche du Higgs. En bleu la contribution du Higgs, en rouge le bruit dû à la QCD pure. De gauche à droite, si l’on ne fait rien, si l’on applique une technique similaire à celle de l’article de BDRS et si l’on utilise la méthode proposée pendant mon stage.

l’on appelle R_{bb} la distance entre les deux sous-jets de b alors on introduit un nouveau paramètre, $R_{filt} = \min(\alpha, \frac{R_{bb}}{2})$ où α est un paramètre à déterminer comme optimal. On reconstruit alors les trois sous-jets les plus durs contenus dans des cônes d’ouverture angulaire $2R_{filt}$. C’est ce qui permet de garder le rayonnement de la désintégration du Higgs à l’ordre dominant en α_s , tout en limitant la contamination de la masse par le pile-up et l’underlying event.

2 Solution proposée.

2.1 Arguments heuristiques.

Pour résoudre le problème j’ai proposé de regarder la distribution d’une variable permettant de discriminer le splitting du Higgs de celui du gluon. Dans un premier temps (et cela m’a servi d’initiation), nous avons regardé la distribution en z ¹¹, en effet il y a une différence car le Higgs possède un spin nul, alors que le gluon est de spin 1. Cette technique est en fait déjà couramment utilisée. En imposant des coupures sur z on peut améliorer la visibilité du signal. J’ai alors voulu trouver une autre variable, qui permettrait de faire la différence entre la QCD et le reste. Or la QCD c’est la physique des objets colorés, dans l’idéal il faudrait pouvoir observer la charge de couleur, ce qui n’est pas possible. On peut par contre s’attaquer aux connections de couleur (voir Figure 2). C’est ce que j’ai essayé de faire.

LHC. Il est devenu particulièrement important au LHC car la densité de proton dans les paquets est beaucoup plus grande que celle du Tevatron. L’underlying event est le phénomène de réinteraction avec les faisceaux de protons rémanents. À cause de ces phénomènes on perd de la précision sur les reconstruction de masse et de p_t des jets car une partie de la masse et du p_t mesuré vient de ces phénomènes et non de l’évènement dur.

11. z est la fraction d’énergie emportée par le quark b lors des splittings $H \rightarrow b\bar{b}$ et $g \rightarrow b\bar{b}$.

Dans ce but j'ai créé une variable qui révèle les connections de couleur. Des analogies avec la QED et l'utilisation de la limite $N_c \rightarrow +\infty$ permettent de construire cette variable. En QED il existe un phénomène appelé effet Sudakov. Si l'on considère le splitting d'un photon (virtuel) en une paire e^+e^- , le rayonnement de cette paire est très supprimé à l'extérieur du cône formé par l'électron et le positron, à cause d'effets d'interférences quantiques. Physiquement cela vient du fait que de loin la paire est vue comme un objet de charge nulle. Il existe un effet analogue en QCD, cependant la symétrie $SU(3)$ complique un peu le constat. Si l'on prend un quark et anti-quark, la représentation $3 \otimes \bar{3}$ se décompose en $3 \otimes \bar{3} = 8 \oplus 1$ ce qui correspond à une représentation singlet et une représentation adjointe¹², or la représentation adjointe est celle des gluons¹³. On a donc pas un seul cas possible comme en QED. On va en fait utiliser ceci pour discriminer le Higgs du gluon. En effet le Higgs est un objet sans couleur, il est donc dans une représentation singlet de $SU(3)$, et ainsi la paire dans laquelle il se désintègre sera dans la représentation singlet. Étant dans la représentation singlet la paire quark anti-quark issue de la désintégration du Higgs sera soumise à l'effet Sudakov, le rayonnement de gluon sera fortement éliminé à l'extérieur de la paire. Il ne doit pas en être de même pour le gluon qui lui est dans la représentation adjointe, la paire issue de la désintégration du gluon n'a donc pas lieu d'être soumise à l'effet Sudakov. Elle doit donc produire plus de gluons à l'extérieur du cône que celle issue de la désintégration du Higgs. On doit prendre un effet supplémentaire en compte, la paire issue du gluon doit moins rayonner à l'intérieur du cône.

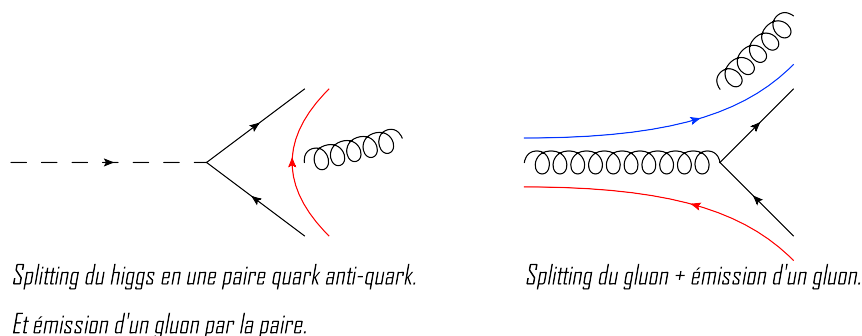


FIGURE 2 – Illustration des différences de connections de couleur dans le splitting d'un Higgs ou d'un gluon.

On le voit facilement à la limite de grand N_c des théories de jauge $SU(N_c)$. Dans cette limite ne subsiste que des diagrammes planaires [16], le gluon peut-être vu comme composé de quark et d'anti-quark. Un argument que l'on peut donner pour cela est que

12. D'une manière générale la décomposition du produit de la représentation fondamentale de $SU(N)$ et de sa contragrédiente est singlet \oplus adjointe i.e. $N \otimes \bar{N} = 1 \oplus N^2 - 1$.

13. Et ces derniers peuvent rayonner d'autres gluons. Dans le cas de la QED on ne peut obtenir, en associant e^+e^- , qu'un objet qui ne porte pas de charge électrique globale et donc ne rayonne pas. En QCD les gluons peuvent rayonner donc si la paire $q\bar{q}$ est un octet elle peut rayonner, c'est en ça que le constat est plus compliqué en QCD.

la dimension de la représentation adjointe de $SU(N_c)$ est $N_c^2 - 1$ et celle du produit de la représentation fondamentale et de sa contragrédiente est N_c^2 , quand N_c est grand $N_c^2 - 1 = N_c^2 + \mathcal{O}(1/N_c^2)$: le gluon est équivalent à deux quarks. Dans cette limite il n'y a plus que des dipôles de couleur et le rayonnement est supprimé hors des cônes formés par ces dipôles. Si l'on observe les graphes de la théorie à grand N_c on voit qu'il n'y a pas d'émission possible entre les deux quarks issus du splitting du gluon. En QCD cependant il reste la possibilité d'émettre à l'intérieur du cône mais elle est supprimée en $1/N_c^2$. On peut voir en Figure 3, une illustration de la limite $N_c \rightarrow \infty$ et des diagrammes que l'on obtient.

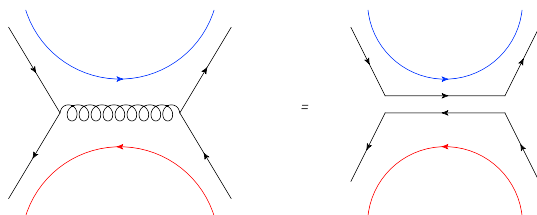


FIGURE 3 – On peut voir à gauche le graphe QCD de l'annihilation d'une paire $q\bar{q}$ en gluon et rematérialisé en paire $q\bar{q}$. Les flux de couleur sont tracés. À droite est tracé le graphe correspondant à la limite $N_c \rightarrow \infty$. Le gluon est vu comme composite, et la réaction se réduit à une diffusion de deux dipôles $q\bar{q}$. Le rayonnement n'est possible que depuis les lignes de couleur (i.e. à l'intérieur des dipôles).

On peut maintenant construire une variable intéressante. Pour cela il nous faut s'intéresser aux processus $H \rightarrow b\bar{b}g$ et $g \rightarrow b\bar{b}g$, dans le premier cas, au vu des arguments précédents le gluon final sera plus souvent situé dans le cône de $b\bar{b}$, tandis que dans le second cas le gluon final sera plutôt émis à l'extérieur de la paire. Dans l'état final le b , le \bar{b} et le gluon auront hadronisé et seront observés sous forme de jets (ou plutôt de sous-jets). Dans notre étude le jet peut-être vu comme étant contenu dans un cône de rayon $R = \sqrt{(\delta\phi)^2 + (\delta y)^2}$ assez grand pour contenir la paire $b\bar{b}$, où ϕ et y sont respectivement l'azimut et la rapidité (voir Figure 4). Si l'on appelle θ la distance dans le plan (y, ϕ) ¹⁴ entre le $b\bar{b}$ (i.e. les deux sous-jets les plus durs), θ_1 l'angle entre le sous-jet le plus dur (que l'on prendra issu du b sans perte de généralité) et le troisième sous-jet le plus dur, θ_2 l'angle entre le deuxième sous-jet le plus dur et le troisième sous-jet, on peut construire une variable qui semble adaptée à la symétrie du problème

$$\nu = \frac{\theta_1^2 + \theta_2^2}{\theta^2}. \quad (1)$$

14. Compte tenu de la disposition des détecteurs, la rapidité est directement relié à la position en z (où l'axe z suit l'axe du faisceau) dans l'espace "réel", et ϕ aux positions en x et y .

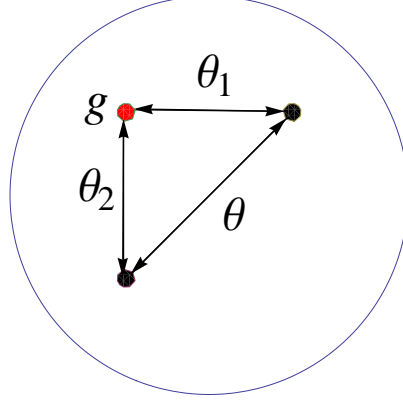


FIGURE 4 – Schéma représentatif du jet contenant les $b\bar{b}$ ainsi que le gluon dans le plan (y, ϕ) . En noir les deux fermions. En rouge le gluon.

La distribution du troisième sous-jet selon cette variable, permet d'accéder à la distribution d'émission du gluon final selon des cercles de rayon $\sqrt{\nu}\theta$ autour du centre du dipôle $b\bar{b}$. D'après les arguments précédents on s'attend à ce que la distribution d'un troisième sous-jet soit plus importante à grand ν pour un splitting de gluon que pour un Higgs, et vice versa à petit ν . Le choix de cette variable fut ma contribution principale pendant ce stage.

2.2 Calculs en QCD perturbative.

2.2.1 Cas du Higgs.

On fera dans un premier temps le calcul pour le Higgs, d'abord parce qu'il est plus aisé, deuxièmement car une remarque simple permet d'obtenir le résultat pour le gluon à partir du résultat pour le Higgs. Dans le cas de l'émission d'un gluon mou par la paire quark anti-quark, la section efficace totalement différentielle se factorise [10] et peut s'écrire

$$d\sigma(H \rightarrow b\bar{b}g) = (d\sigma(H \rightarrow b\bar{b})) \frac{\alpha_s C_F}{\pi^2} \frac{p_1 \cdot p_2}{(p_1 \cdot k)(p_2 \cdot k)} \delta(k^2) d^4k. \quad (2)$$

Où p_1, p_2 sont les impulsions du quark b et de l'anti-quark \bar{b} respectivement, k l'impulsion du gluon final. Cette factorisation nous permet de définir la distribution en ν du gluon final comme

$$\frac{dP_H}{d\nu} = \int d^4k \frac{\alpha_s C_F}{\pi^2} \frac{p_1 \cdot p_2}{(p_1 \cdot k)(p_2 \cdot k)} \delta(k^2) \delta\left(\nu - \frac{\theta_1^2 + \theta_2^2}{\theta^2}\right). \quad (3)$$

On écrira l'élément de matrice à l'aide de θ , θ_1 et θ_2 . Dans ces grandeurs on peut écrire

$$\begin{aligned} \frac{\alpha_s C_F}{\pi^2} \frac{p_1 \cdot p_2}{(p_1 \cdot k)(p_2 \cdot k)} &= \frac{\alpha_s C_F}{\pi^2} \frac{z(1-z)p_t^2(1-\cos(\theta))}{z(1-z)p_t^2 k_t^2 (1-\cos(\theta_1))(1-\cos(\theta_2))} \\ &\simeq \frac{\alpha_s C_F}{\pi^2} \frac{2\theta^2}{k_t^2 \theta_1^2 \theta_2^2}, \end{aligned} \quad (4)$$

où l'on a fait deux hypothèses : les fermions sont pris sans masses et les angles sont petits, ce qui pour un Higgs boosté est le cas¹⁵. On remarque le terme en $\frac{1}{k_t}$ qui signe la divergence infrarouge, de même que la dépendance en $\frac{1}{\theta_1^2 \theta_2^2}$ est le signe de divergence collinéaire, ces divergences sont connues en QCD (les mêmes existent en QED), et viennent dans ce cas de la limite de masse nulle pour les fermions. Cependant le théorème de Kinoshita [14, 10] prouve que ces divergences s'annulent ordre par ordre en théorie des perturbations. La mesure $d^4 k$ s'écrit, en terme de y et ϕ

$$d^4 k = k_t dk_t m_t dm_t d\phi dy, \quad (5)$$

m_t étant la masse transverse (pour la cinématique on se référera à [5]). La contrainte de couche de masse nous permet de simplifier la mesure, $\delta(k^2) = \delta(k_t^2 - m_t^2)$, et donc

$$\delta(k^2) d^4 k = \delta(k_t^2 - m_t^2) k_t dk_t m_t dm_t d\phi dy = \frac{1}{2} k_t dk_t d\phi dy. \quad (6)$$

Il faut maintenant réécrire y et ϕ en terme des θ_i . Pour de petites valeurs des θ_i , on a

$$\theta_1^2 = \phi^2 + \left(y + \frac{\theta}{2}\right)^2, \quad \theta_2^2 = \phi^2 + \left(y - \frac{\theta}{2}\right)^2. \quad (7)$$

Inversant ces relations on trouve

$$y = \frac{\theta_1^2 - \theta_2^2}{2\theta}, \quad (8)$$

$$\phi = \frac{1}{2\theta} \sqrt{2(\theta^2 \theta_1^2 + \theta_1^2 \theta_2^2 + \theta_2^2 \theta^2) - (\theta_1^4 + \theta_2^4 + \theta^4)}. \quad (9)$$

On calcule alors le déterminant $|J|$ du jacobien de changement de variable :

$$|J| = \frac{2\theta_1 \theta_2}{\sqrt{2(\theta^2 \theta_1^2 + \theta_1^2 \theta_2^2 + \theta_2^2 \theta^2) - (\theta_1^4 + \theta_2^4 + \theta^4)}}. \quad (10)$$

On écrira dorénavant $\Lambda[\theta_1, \theta_2, \theta] = 2(\theta^2 \theta_1^2 + \theta_1^2 \theta_2^2 + \theta_2^2 \theta^2) - (\theta_1^4 + \theta_2^4 + \theta^4)$ et l'on remarquera qu'il est invariant sous n'importe quelle permutation de θ_1 , θ_2 , θ . La distribution (3) se réécrit alors

15. Du reste c'est une approximation très utilisée et qui fonctionne très bien.

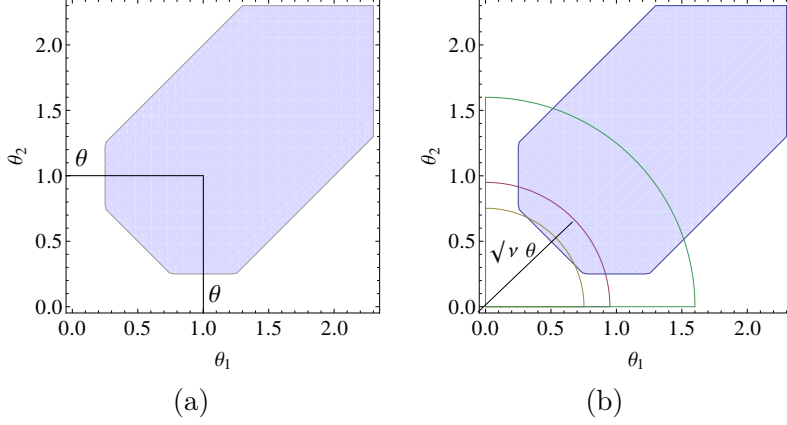


FIGURE 5 – (a) Domaine d'intégration en hachuré, (b) Les 3 différents cas d'intégration sur γ fonction des valeurs de ν . On aura pris $\theta = 1$.

$$\frac{dP_H}{d\nu} \simeq \frac{C_F}{\pi} \int \alpha_s \frac{dk_t}{k_t} d\theta_1 d\theta_2 \frac{2\theta^2}{\theta_1 \theta_2 \sqrt{\Lambda[\theta_1, \theta_2, \theta]}} \delta\left(\nu - \frac{\theta_1^2 + \theta_2^2}{\theta^2}\right). \quad (11)$$

Il reste à évaluer cette intégrale. Pour cela il faut intégrer sur le domaine D (cf. Figure 5) défini à l'aide des inégalités triangulaires en oubliant pas que les divergences collinéaires nous interdisent d'intégrer en $\theta_i = 0$, cependant il existe un cut-off R_0 qui vient de la taille des sous jets

$$D(\theta, R_0) = \{\theta_1, \theta_2 \in \mathbb{R}^2 | \theta_1 \geq R_0, \theta_2 \geq R_0, \theta_1 + \theta_2 \geq \theta, |\theta_1 - \theta_2| \leq \theta\}. \quad (12)$$

On peut évaluer cette intégrale en effectuant une série de changement de variable. Le premier nous fait passer en coordonnées polaires en posant

$$\theta_1 = r \cos(\gamma + \frac{\pi}{4}), \quad \theta_2 = r \sin(\gamma + \frac{\pi}{4}). \quad (13)$$

Pour simplifier le calcul on va faire l'hypothèse supplémentaire que la constante de couplage variable ne dépend pas de θ_1 et θ_2 (la dépendance est en fait logarithmique). On peut donc factoriser l'intégrale sur k_t et écrire

$$\frac{dP_H}{d\nu} = \left(\frac{C_F}{\pi} \int \alpha_s \frac{dk_t}{k_t}\right) F_H(\nu), \quad (14)$$

$$F_H(\nu) = \int d\theta_1 d\theta_2 \frac{2\theta^2}{\theta_1 \theta_2 \sqrt{\Lambda[\theta_1, \theta_2, \theta]}} \delta\left(\nu - \frac{\theta_1^2 + \theta_2^2}{\theta^2}\right). \quad (15)$$

On a

$$\Lambda[\theta_1, \theta_2, \theta] \rightarrow \tilde{\Lambda}[r, \gamma, \theta] = \Lambda[r \cos(\gamma + \frac{\pi}{4}), r \sin(\gamma + \frac{\pi}{4}), \theta] \quad (16)$$

$$= 2\theta^2 r^2 - r^4 \sin^2(2\gamma) - \theta^4. \quad (17)$$

Ainsi $F_H(\nu)$ se réécrit :

$$F_H(\nu) = \int \frac{dr^2}{2} d\gamma \frac{4\theta^2}{r^2 \cos(2\gamma) \sqrt{2\theta^2 r^2 - \theta^4 - r^4 \sin^2(2\gamma)}} \delta\left(\nu - \frac{r^2}{\theta^2}\right). \quad (18)$$

Puis intégrant sur r à l'aide de la distribution de Dirac et en posant $t = \sin(2\gamma)$, on a

$$F_H(\nu) = \frac{1}{4} \int dt \frac{1}{\nu(1-t^2) \sqrt{2\nu-1-\nu^2 t^2}}. \quad (19)$$

On peut effectuer un dernier changement de variable qui nous permettra de calculer :
 $u = \frac{\nu t}{\sqrt{2\nu-1}}$,

$$F_H(\nu) = \frac{1}{4} \int \frac{du}{\nu^2} \frac{1}{\left(1 - \frac{2\nu-1}{\nu^2} u^2\right)} \frac{1}{\sqrt{1-u^2}}. \quad (20)$$

Il nous faut déterminer les bornes de l'intégrale. Premièrement on se doit de remarquer que $\nu > 1/2$. De plus, si l'on appelle $f = \frac{R_0}{\theta}$, un peu de géométrie nous permet de voir que l'on peut diviser l'intégrale en 3 cas (Figure 5.b) qui dépendent de la valeur de ν . On a

$$\begin{aligned} \text{cas 1 : } & f^2 + (1+f)^2 < \nu \\ \text{cas 2 : } & f^2 + (1-f)^2 < \nu \leq f^2 + (1+f)^2 \\ \text{cas 3 : } & 1/2 < \nu \leq f^2 + (1-f)^2. \end{aligned} \quad (21)$$

Les bornes en γ se trouvent alors facilement dans chacun des cas

$$\begin{aligned} \text{cas 1 : } \gamma_{1,\pm} &= \pm \text{Arcsin}\left(\frac{1}{\sqrt{2\nu}}\right) \\ \text{cas 2 : } \gamma_{2,\pm} &= \pm \left(\text{Arccos}\left(\frac{f}{\sqrt{\nu}}\right) - \frac{\pi}{4}\right) \\ \text{cas 3 : } \gamma_{3,\pm} &= \pm \text{Arccos}\left(\frac{1}{\sqrt{2\nu}}\right). \end{aligned} \quad (22)$$

On en déduit les bornes en u

$$\begin{aligned} \text{cas 1 et cas 3 : } & u_{\pm} = \pm 1 \\ \text{cas 2 : } & u_{2,\pm} = \pm \frac{\nu-2f^2}{\sqrt{2\nu-1}}. \end{aligned} \quad (23)$$

En utilisant aussi la parité de l'intégrande en u , on peut écrire au final :

$$F_H(\nu) = \frac{1}{2} \int_0^{u_{i,+}} \frac{du}{\nu^2} \frac{1}{\left(1 - \frac{2\nu-1}{\nu^2} u^2\right)} \frac{1}{\sqrt{1-u^2}} \quad (24)$$

On trouve alors :

$$\frac{dP_H}{d\nu} = \left(\frac{C_F}{\pi} \int \alpha_s \frac{dk_t}{k_t} \right) \frac{1}{\nu|\nu-1|} \quad (25)$$

$$\cdot \begin{cases} \frac{2}{\pi} \text{Arctan} \left(\frac{(\nu-2f^2)|\nu-1|}{\nu \sqrt{4f^2 - (\nu-1-2f^2)^2}} \right) & \text{si } \nu_- \leq \nu \leq \nu_+ \\ 1 & \text{Autrement.} \end{cases} \quad (26)$$

Où l'on a $\nu_{\pm} = f^2 + (1 \pm f^2)$. Cette expression donne la distribution en ν d'un troisième sous-jet dans le cas de la désintégration du Higgs en une paire de $b\bar{b}$ et un gluon dans la limite de gluon mou. Dans l'optique d'étudier le comportement du bruit QCD, il nous faut maintenant faire le calcul dans le cas $g \rightarrow b\bar{b}g$. L'élément de matrice possède une structure plus complexe, cependant il arrive une simplification qui nous donne le résultat en fonction de la distribution pour la désintégration du Higgs.

2.2.2 Cas du gluon.

En considérant la limite $N_c \rightarrow \infty$ (Figure 6), on peut déterminer l'élément de matrice d'émissions de gluon. Il y a deux dipôles caractérisés par une impulsion q et l'impulsion p_1 ou p_2 respectivement pour le b et le \bar{b} , le terme supplémentaire est le terme singlet, supprimé en $1/N_c^2$, qui corrige la limite $N_c \rightarrow \infty$. L'élément de matrice est :

$$\frac{\alpha_s C_F}{\pi^2} \left(\frac{q \cdot p_1}{(q \cdot k)(p_1 \cdot k)} + \frac{q \cdot p_2}{(q \cdot k)(p_2 \cdot k)} - \frac{1}{N_c^2} \frac{p_1 \cdot p_2}{(p_1 \cdot k)(p_2 \cdot k)} \right) \quad (27)$$

On a dans la limite à petit R , $\frac{q \cdot p_1}{q \cdot k} = \frac{z p_t}{k_t} (1 + \mathcal{O}(\theta_1^2))$, pour les mêmes raisons nous avons $\frac{q \cdot p_2}{q \cdot k} = \frac{(1-z) p_t}{k_t} (1 + \mathcal{O}(\theta_2^2))$, l'élément de matrice évalué à petit R est alors

$$\frac{\alpha_s C_F}{\pi^2} \left(\frac{2}{k_t^2 \theta_1^2} + \frac{2}{k_t^2 \theta_2^2} - \frac{1}{N_c^2} \frac{2\theta^2}{k_t^2 \theta_1^2 \theta_2^2} \right) \quad (28)$$

Introduisant l'expression de $\frac{dP_g}{d\nu}$

$$\frac{dP_g}{d\nu} = \frac{C_F}{\pi^2} \int \alpha_s \frac{dk_t}{k_t} \int d\theta_1 d\theta_2 \frac{2\theta_1 \theta_2}{\sqrt{\Lambda[\theta_1, \theta_2, \theta]}} \left(\frac{\theta_1^2 + \theta_2^2}{\theta_1^2 \theta_2^2} - \frac{1}{N_c^2} \frac{\theta^2}{\theta_1^2 \theta_2^2} \right) \delta \left(\nu - \frac{\theta_1^2 + \theta_2^2}{\theta^2} \right) \quad (29)$$

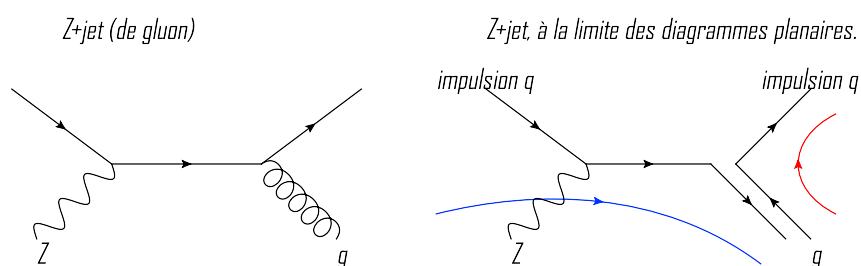


FIGURE 6 – Diagrammes complets correspondant au bruit $g \rightarrow b\bar{b}$. À droite le diagramme planaire (limite $N_c \rightarrow \infty$) correspondant. Les dipôles ont été tracés en couleur.

Utilisant le fait que $f(x)\delta(a-f(x)) = a\delta(a-f(x))$, nous pouvons réécrire la distribution pour le splitting du gluon en fonction de celle pour le splitting du Higgs. On pose :

$$\begin{aligned} F_g(\nu) &= \int d\theta_1 d\theta_2 \frac{2\theta^2}{\theta_1\theta_2\sqrt{\Lambda[\theta_1, \theta_2, \theta]}} \left(\nu - \frac{1}{N_c^2}\right) \delta\left(\nu - \frac{\theta_1^2 + \theta_2^2}{\theta^2}\right) \\ &= \left(\nu - \frac{1}{N_c^2}\right) F_H(\nu) \end{aligned} \quad (30)$$

d'où :

$$\frac{dP_g}{d\nu} = \left(\nu - \frac{1}{N_c^2}\right) \frac{dP_H}{d\nu}. \quad (31)$$

Pour avoir le résultat final, il nous faut nous occuper de l'intégration sur k_t , plusieurs

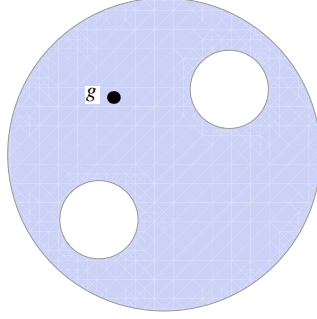


FIGURE 7 – Domaine de resommation des gluons, pour le calcul du terme de Sudakov il faut intégrer l'élément de matrice sur la partie hachurée. Les deux disques blancs sont les deux sous-jets les plus durs, ils correspondent aux sous-jets initiés par les $b\bar{b}$. Le point noir correspond au gluon émis.

solutions s'offrent à nous. La première est de considérer α_s comme constante, on peut aussi utiliser une constante de couplage variable. Il faut aussi prendre en compte le fait que si l'on considère que le gluon est le troisième sous-jet le plus dur alors il faut qu'aucun autre gluon plus dur n'ait été émis dans le disque de rayon R en dehors des sous-jets des b et \bar{b} . On peut voir une illustration de ce domaine sur la Figure 7. Dans l'approximation où les émissions successives sont indépendantes ceci revient à introduire un terme de Sudakov dans l'intégrale, terme qui resomme les gluons mous. Ce terme s'écrit

$$S_H = \exp \left[-\frac{C_F}{\pi^2} \int d^4l \alpha_s \frac{p_1 \cdot p_2}{(p_1 \cdot l)(p_2 \cdot l)} \delta(l^2) \right] \quad (32)$$

$$S_g = \exp \left[-\frac{C_F}{\pi^2} \int d^4l \alpha_s \left(\frac{q \cdot p_1}{(q \cdot l)(p_1 \cdot l)} + \frac{q \cdot p_2}{(q \cdot l)(p_2 \cdot l)} - \frac{1}{N_c} \frac{p_1 \cdot p_2}{(p_1 \cdot l)(p_2 \cdot l)} \right) \delta(l^2) \right] \quad (33)$$

Le domaine d'intégration étant celui de la Figure 7. Le terme de Sudakov s'insère alors dans l'intégrale $\int \frac{dk_t}{k_t} \alpha_s S_i$ où i est H ou g selon que l'on considère le splitting du Higgs

ou du gluon. Il faut se poser la question de l'échelle de la constante de couplage. L'échelle la plus naturelle qui apparait dans le cas du Higgs¹⁶ est $\mu^2 = \frac{(p_1 \cdot k)(p_2 \cdot k)}{(p_1 \cdot p_2)}$, l'intégrale étant dominée par la limite collinéaire on peut facilement éliminer la dépendance géométrique de la constante de couplage, en effet à la limite collinéaire nous avons $\theta_2 \simeq \theta$ et $\theta_1 \simeq R_0$ (et vice et versa). Nous voyons :

$$\mu \simeq \sqrt{\frac{k_t^2 \theta_1^2 \theta_2^2}{2\theta^2}} \simeq k_t R_0 \quad (34)$$

Dans le cas gluon l'intégrale est aussi dominée par la limite collinéaire ainsi on peut choisir la même échelle et faire la même approximation. On prendra donc la constante de couplage à l'échelle $k_t R_0$ dans les deux cas. On imposera aussi des cuts en k_t pour les intégrales sur l_t et k_t , le cut supérieur se justifiant par le fait qu'aucun gluon ne peut être émis par la paire avec une impulsion transverse supérieure à p_t ¹⁷ (impulsion transverse du Higgs). Pour les cuts inférieurs dans l'intégrale sur l_t cela vient du fait que nous voulons qu'aucun gluon ne soit émis à des impulsions comprises entre k_t (impulsion du troisième sous-jet considéré) et p_t , la borne inférieure de cette intégrale est alors k_t . Pour l'intégrale sur k_t on peut imposer un cut inférieur afin d'éviter de confondre le troisième sous-jet qui vient d'un processus dur avec un jet du pile-up ou de l'underlying event. La constante de couplage est donnée pour des valeurs de μ grandes devant Λ_{QCD} par¹⁸

$$\alpha_s(\mu) = \frac{1}{2\beta \ln\left(\frac{\mu}{\Lambda_{QCD}}\right)} \quad (35)$$

On appelle $I_i(h) = \int_{1/2}^{2h^2} d\nu F_i(\nu)$, où $h = \frac{R}{\theta}$ est un cut naturel sur la valeur de ν qui vient du fait que le jet est défini, pour le calcul¹⁹, comme les particules contenues dans un cône de rayon R . On calcule alors facilement le terme de Sudakov.

$$\begin{aligned} S_i &= e^{-\frac{C_F}{2\beta\pi^2} I_i(h) \int_{k_t}^{p_t} \frac{1}{\ln\left(\frac{R_0 k_t}{\Lambda_{QCD}}\right)} \frac{dl_t}{l_t}} \\ &= e^{-\frac{C_F}{2\beta\pi^2} I_i(h) \ln\left(\frac{\alpha_s(R_0 k_t)}{\alpha_s(R_0 p_t)}\right)} \\ &= \left[\frac{\alpha_s(R_0 k_t)}{\alpha_s(R_0 p_t)} \right]^{-\frac{C_F}{2\beta\pi^2} I_i(h)} \end{aligned} \quad (36)$$

16. Cela correspond en fait à l'impulsion du gluon dans le centre de masse.

17. Plus précisément c'est zp_t qui devrait être choisi, cependant la dépendance étant logarithmique on se permet cette approximation afin de limiter le nombre de paramètres à garder sous contrôle.

18. Λ_{QCD} est le paramètre dimensionnel que l'on obtient de la théorie par la procédure de transmutation dimensionnelle, elle est définie comme l'échelle d'énergie à laquelle la constante de couplage est de l'ordre de 1. Sa valeur dépend du schéma de renormalisation mais est de l'ordre de 200 MeV (217 MeV dans le schéma minimal), ce que l'on prendra comme valeur. β est la constante calculée par Wilczek, Gross et Politzer.

19. Dans le monde réel un jet est défini par l'algorithme de clusterisation utilisé, ici Cambridge/Aachen. Cependant pour que le calcul soit faisable il nous faut une définition effective. En pratique la taille du jet varie.

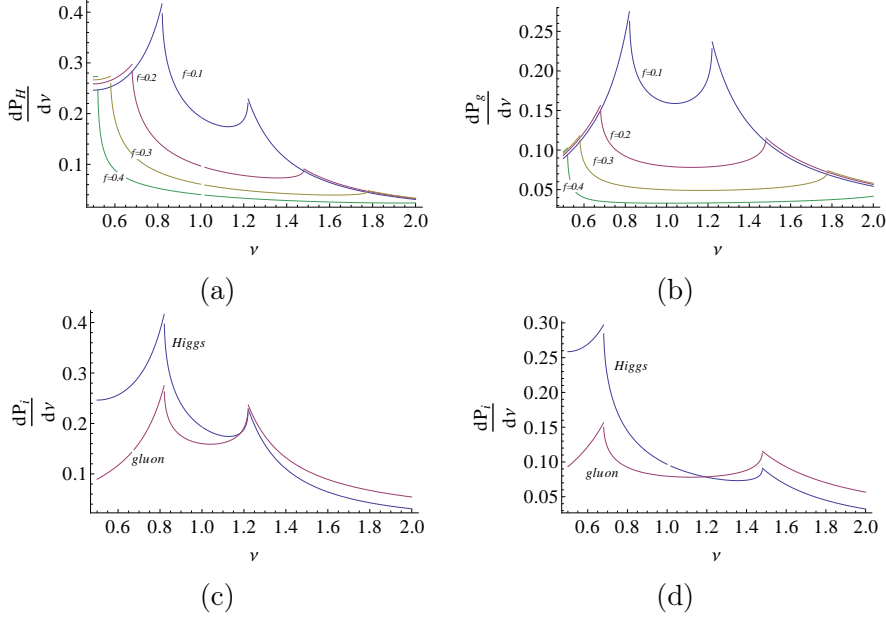


FIGURE 8 – (a) Distribution en ν cas du Higgs pour différentes valeurs de f , (b) Distribution en ν cas du gluon pour différentes valeurs de f , (c) Comparaison des distributions entre le cas du Higgs (bleu) et le cas du gluon (magenta) à $f = 0.1$, (d) Comparaison des distributions entre le cas du Higgs (bleu) et le cas du gluon (magenta) à $f = 0.2$.

Il nous reste à calculer l'intégrale sur k_t avec le terme de Sudakov.

$$\begin{aligned}
\int_{k_c}^{p_t} \frac{dk_t}{k_t} \alpha_s(R_0 k_t) S_i &= \int_{k_c}^{p_t} \frac{dk_t}{k_t} \alpha_s(R_0 k_t) \left[\frac{\alpha_s(R_0 k_t)}{\alpha_s(R_0 p_t)} \right]^{-\frac{C_F}{2\beta\pi^2} I_i(h)} \\
&= \frac{\pi^2}{C_F I_i(h)} \left[1 - \left(\frac{\alpha_s(R_0 p_t)}{\alpha_s(R_0 k_c)} \right)^{\frac{C_F}{2\pi^2\beta} I_i(h)} \right] \quad (37)
\end{aligned}$$

À la limite où l'on prend un cut inférieur nul, le terme de Sudakov introduit simplement une normalisation. On a maintenant le résultat complet. Pour le Higgs, nous avons

$$\begin{aligned}
\frac{dP_H}{d\nu} &= \frac{\pi}{I_i(h)} \left[1 - \left(\frac{\alpha_s(R_0 p_t)}{\alpha_s(R_0 k_c)} \right)^{\frac{C_F}{2\pi^2\beta} I_i(h)} \right] \cdot \frac{1}{\nu|\nu-1|} \\
&\cdot \begin{cases} \frac{2}{\pi} \text{Arctan} \left(\frac{(\nu-2f^2)|\nu-1|}{\nu\sqrt{4f^2-(\nu-1-2f^2)^2}} \right) & \text{si } \nu_- \leq \nu \leq \nu_+ \\ 1 & \text{Autrement.} \end{cases} \quad (38)
\end{aligned}$$

Pour le gluon :

$$\frac{dP_g}{d\nu} = \frac{\pi}{I_i(h)} \left[1 - \left(\frac{\alpha_s(R_0 p_t)}{\alpha(R_0 k_c)} \right)^{\frac{C_F}{2\pi^2\beta} I_i(h)} \right] \frac{\nu - 1/N_c^3}{\nu|\nu - 1|} \cdot \begin{cases} \frac{2}{\pi} \text{Arctan} \left(\frac{(\nu - 2f^2)|\nu - 1|}{\nu\sqrt{4f^2 - (\nu - 1 - 2f^2)^2}} \right) & \text{si } \nu_- \leq \nu \leq \nu_+ \\ 1 & \text{Autrement.} \end{cases} \quad (39)$$

On présente en Figure 8 les résultats (quasi-)analytiques²⁰ pour différentes valeurs de f , et des comparaisons de cas gluons et Higgs, on aura pris $p_t = 350$ GeV et $m_H = 115$ GeV²¹. Nous avons aussi utilisé une autre variable : $\omega = \frac{1}{2\nu}$ qui permet d'obtenir toute la distribution sur un intervalle borné en ω . On obtient facilement la distribution en ω à partir de celle en ν . ω est utilisée dans la suite pour la comparaison avec Pythia.

3 Tests avec PYTHIA

Qu'est-ce que PYTHIA ? PYTHIA est un générateur d'évènement pour la physique des hautes énergies [13]. Il est capable de traiter toute la physique de l'évènement en utilisant des méthodes de Monte-Carlo. Pour ce faire, il divise le problème en sous-problèmes plus simples, en traitant d'abord le processus dur (au niveau partonique dans notre cas) puis en rajoutant les douches de partons, les interactions multiples et réinteractions avec le faisceau, et enfin traite l'hadronisation avec des modèles de fragmentation de cordes. On obtient alors des états finaux qui sont les types d'états que l'on devrait observer en accélérateurs. Tout ces processus ne sont pas analysables analytiquement, ainsi PYTHIA est un outil nous permettant de tester notre idée dans un cadre bien plus complet que celui (très simple) dans lequel nous avons pu faire les calculs analytique. Nous avons utilisé presque exclusivement PYTHIA 8.

3.1 Comparaison PYTHIA/Calculs analytiques.

Nous avons utilisé PYTHIA pour déterminer la distribution en ν , dans le cas HZ et Zj dans l'état final, qui doivent correspondre respectivement aux calculs pour le splitting du Higgs et pour le splitting du gluon. L'analyse des évènements s'effectue avec Cambridge/Aachen et avec une technique analogue à [8] (le Higgs est identifié de la même manière), le tag du Higgs se faisant avec une technique différente du Mass Drop. On choisit un R assez grand pour contenir les $b\bar{b}$. On sélectionne les jets de $p_t > 200$ GeV, puis on identifie les sous-jets avec Cambridge/Aachen²², on applique un b-tagging pour chacun des 2 sous-jets tout en mesurant la distance entre ces deux sous jets, puis on applique un filtre (voir introduction de [9]) pour conserver les 3 sous-jets les plus durs

20. Seul I_i est calculé numériquement.

21. On a pris $R=1.2$ valeur utilisée dans les articles [9, 2], le cut inférieur $k_c = 10$ GeV.

22. Distance JADE $J = p_{t,i} p_{t,j} (\Delta R_{ij})^4$. Similaire à la masse du splitting. Coupure en $z = 0.1$, voir[8]

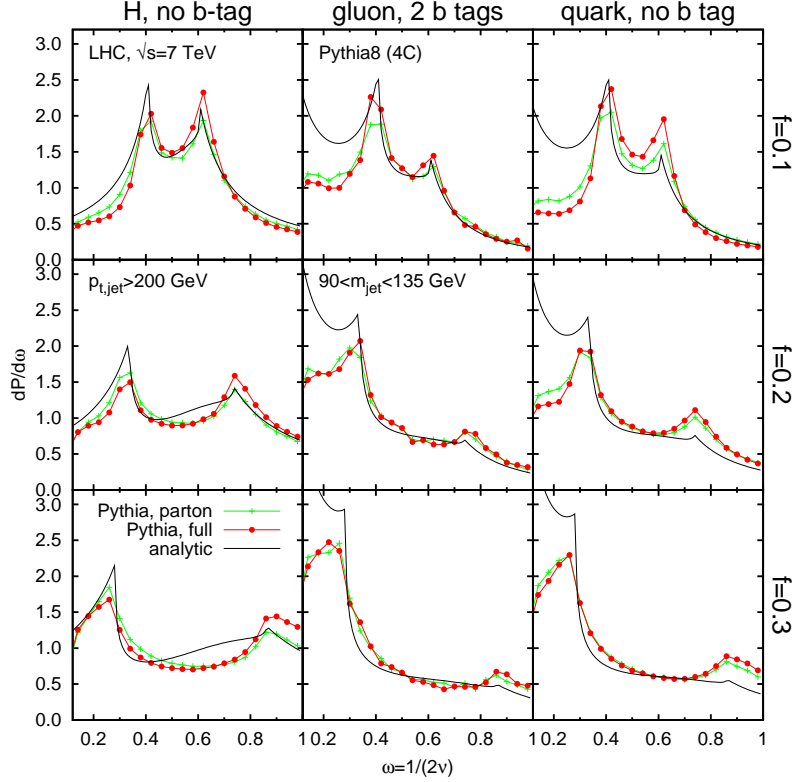


FIGURE 9 – Distributions $\frac{dP}{d\omega}$ où $\omega = 1/2\nu$. Comparaison entre les résultats de PYTHIA et le calcul analytique à $p_t > 200\text{GeV}$ au LHC (faisceaux $p\bar{p}$) à $\sqrt{s} = 7\text{ TeV}$. On a dans chaque colonne de gauche à droite les cas : $H \rightarrow b\bar{b}g$, $g \rightarrow b\bar{b}g$ et $q \rightarrow qgg$. Dans le cas gluon on applique un b-tagging. On remarquera ici que tout est normalisé à 1, en effet pour simplifier l'analyse on a pris $k_c = 0$.

avec un rayon $R_{filt} = \min(R_{bb}/2, 0.3)$, enfin on impose la coupure en p_t . Pour les jets passant cette procédure, l'analyse est la suivante, on reclustérise le jet avec un rayon $R_0 = fR_{bb}$, on garde alors les 3 sous-jets les plus durs. ν est calculé à partir de ces trois sous-jets.

On peut comparer en Figure 9 les résultats de Pythia à plusieurs niveaux (partons, hadronisation + UE) avec le calcul analytique. On commencera par PYTHIA au niveau partons avec le calcul analytique qui est lui uniquement partonique. On notera en premier lieu, que dans le cas du Higgs le calcul analytique semble être meilleur à petit f . Analysant d'abord les résultats pour le Higgs à $f = 0.1$, on peut voir deux pics, le premier étant un peu sur-estimé par rapport aux résultats de PYTHIA, le calcul analytique sur-estime aussi un peu la distribution à grand ν (i.e. à petit ω). Quand on rajoute l'hadronisation, on remarque que le second pic est plus marqué que ne le prévoit le calcul analytique. L'accord reste raisonnable compte tenu des approximations effectuées. À $f = 0.2$ et surtout $f = 0.3$ on peut voir que l'accord devient plutôt mauvais dans la zone centrale

du graphe, hors c'est la zone qui est affectée par les cuts en R_0 autour des sous-jets. Une explication proposée est qu'il s'agit d'effets de géométrie des sous-jets, en effet quand les sous-jets sont reclusterisés avec Cambridge/Aachen, ce ne sont pas des cercles parfaits ainsi les domaines d'intégration "vrais" ne sont pas aussi simples que ceux utilisés²³.

Dans le cas gluon l'accord est très mauvais à grand ν (petit ω) et comme remarqué précédemment on ne retrouve pas le mauvais comportement dans la région centrale. Le désaccord à grand ν pourrait là aussi venir d'effets de géométrie autour non des sous-jet mais du jet dans sa globalité, de la même manière le jet n'est pas un cône de rayon R dans les algorithmes de recombinaison. Ces effets de géométrie seraient plus sensibles à grand ν dans le cas du gluon que dans le cas du Higgs car le gluon a plus tendance à rayonner à l'extérieur du cône formé par la paire $b\bar{b}$ et ainsi la géométrie du jet est plus variable d'un évènement à l'autre. On peut invoquer la même explication dans le cas du bruit QCD $q \rightarrow qg$. On peut remarquer que dans le cas $q \rightarrow qg$ le calcul analytique sous-estime de beaucoup le second pic, de plus la différence entre PYTHIA avec hadronisation et UE, et la solution analytique est assez grande à petit f , cas assez insensible à l'UE ce qui laisse penser que les corrections d'hadronisation deviennent importantes à petit f .

Pour améliorer ces résultats on pourrait dans un premier temps essayer de relâcher l'approximation de gluon soft (en introduisant des fonctions de splitting complètes [7]). Dans les résultats analytiques nous avons pris $\frac{R}{\theta}$ fixe, ce n'est pas le cas avec PYTHIA, il se peut que ça ait un effet à grand ν , même si à priori la dépendance est faible. On pourrait le faire analytiquement en intégrant sur le splitting du Higgs ou du gluon. Les corrections d'hadronisation sont plus difficiles à prendre en compte, il faut se fixer un modèle type modèle de Lund. Cependant, on pourrait essayer d'obtenir un critère permettant de savoir à partir de quelles valeurs de f ces corrections deviennent importantes. Une chose à essayer, d'après moi, est d'améliorer la prise en compte de la géométrie dans l'échelle de la constante de couplage (où l'on avait éliminer toute dépendance en θ_1, θ_2), en effet l'hadronisation est un processus régi par la valeur de α_s . Par exemple, on pourrait faire un développement de cette dépendance et voir si l'on peut calculer la forme des corrections. On peut faire le calcul analytique pour le quark, il a été fait par mon maître de stage. Cependant le b-tagging réduit énormément cette source de bruit et à priori si il y a du b-tagging les effets du bruit $q \rightarrow qg$ sont faibles.

3.2 Cut en ν et ω .

La question que l'on est en mesure de se poser maintenant est : en quoi l'utilisation de cette variable nous permettrait-elle d'améliorer l'analyse des évènements au LHC ? L'idée est d'imposer des cuts sur cette variable discriminant le signal $H \rightarrow b\bar{b}$, du bruit QCD $g \rightarrow b\bar{b}$ et $q \rightarrow qg$. En ce qui concerne le dernier le b-tagging nous permet d'éliminer 98%

23. On pourrait arguer que si ce sont simplement des effets de géométrie alors on devrait observer le même phénomène pour les cas $g \rightarrow b\bar{b}g$ et $q \rightarrow qgg$. La procédure de clusterisation étant la même. Cet argument est vrai uniquement au niveau perturbatif, cependant comme souligné dans la section 1.1 la forme d'un jet est totalement non perturbative et dépend de l'évènement dans sa globalité phénomène de rayonnement mou et d'hadronisation compris, en effet il suffit de rajouter un gluon pour modifier la forme des jets.

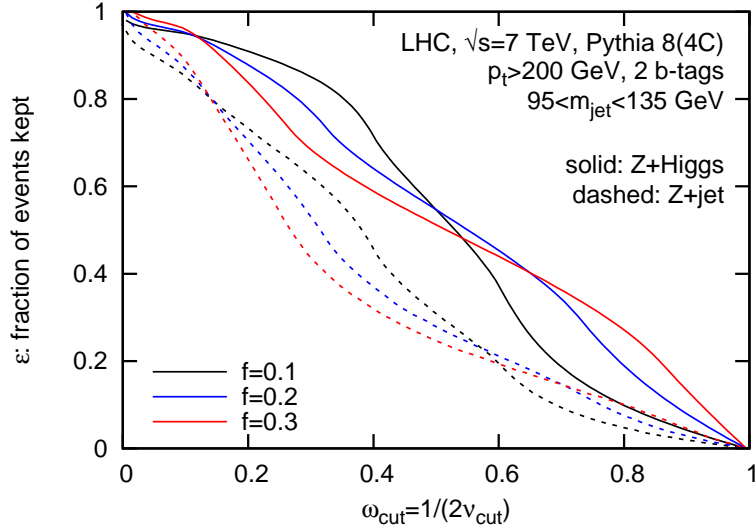


FIGURE 10 – Fraction d'évènements conservés en fonction de la coupure en ω . Résultats de PYTHIA. Pour différentes valeurs de f . En traits pleins le cas du Higgs, en traits pointillés le cas du gluon.

de celui-ci (en réduisant en contrepartie le signal de 30-40 %). Cependant le cas $g \rightarrow b\bar{b}$ est plus coriace. Nous allons donc nous attacher à étudier ce dernier cas.

Il y a alors deux grandeurs qui nous intéressent : le rapport signal/bruit et la signifi-
cance qui est le rapport $\text{signal}/\sqrt{\text{bruit}}$. Le premier nous informe sur la visibilité du pic
que l'on devrait observer dans la section efficace si le Higgs existe, tandis que le second
nous donne une information sur la valeur statistique de ce pic, i.e. si il est dû unique-
ment à des fluctuations ou si il est réellement dû à l'existence du Higgs. Le but étant
d'améliorer le plus possible ces deux grandeurs.

On définit²⁴ la fraction de signal par $\epsilon_S = \int_{1/2}^{\nu_{cut}} \frac{dP_H}{d\nu} d\nu = \int_{\omega_{cut}}^1 \frac{dP_H}{d\omega} d\omega$ de même la
fraction de bruit est définie de la même manière : $\epsilon_B = \int_{1/2}^{\nu_{cut}} \frac{dP_g}{d\nu} d\nu = \int_{\omega_{cut}}^1 \frac{dP_g}{d\omega} d\omega$. Les
résultats de PYTHIA sont en Figure 10.

Les prévisions du calcul analytique pour la signifi-
cance en fonction du rapport si-
gnal/bruit sont montrées en Figure 11. On constate qu'il y a seulement un accord de
principe entre les résultats analytiques et PYTHIA. Les courbes ont des comportements
similaires. On voit que le gain en signifi-
cance est très faible, tout au plus de l'ordre de
10%. Ainsi le choix d'une coupure en ω sera principalement motivée par la volonté d'aug-
menter le rapport signal/bruit sans diminuer la signifi-
cance. Dans cette optique les cas
 $f = 0.2$ et $f = 0.3$ semblent être les plus intéressants. On a affiché sur la Figure 11 le point
 $\omega_{cut} = 0.4$ qui correspond à la valeur de la coupure pour laquelle, dans le cas $f = 0.1$, la
signifi-
cance analytique re-passe en dessous de 1. Pour cette même valeur de la coupure

24. Cette définition n'est valable que dans le cas où l'on prend le cut minimum k_c en impulsion nul, sinon il faut introduire une normalisation.

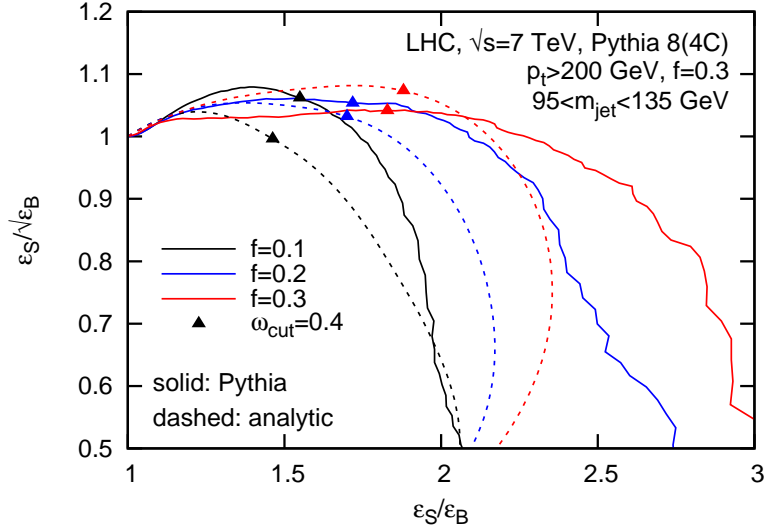


FIGURE 11 – Tracé de la signficance en fonction du rapport ϵ_S/ϵ_B . Les tracés en traits pleins sont les résultats de PYTHIA, en traits pointillés les résultats analytiques. Les triangles désignent les points correspondant à la valeur $\omega_{cut} = 0.4$.

les cas $f = 0.2$ et $f = 0.3$ confirment être plus intéressants. Cependant on ne veut pas choisir une valeur de la coupure trop grande pour éviter de perdre trop d'évènements, la section efficace en deviendrait trop réduite (en effet on est dans un cas où elle est déjà très faible.). Une augmentation du rapport signal/bruit d'un facteur compris entre 1.5 et 2 semble raisonnable. Sur la Figure 10 on peut voir qu'une coupure à $\omega_{cut} = 0.4$ permet de garder environ 60% des évènements de désintégration d'un Higgs dans les cas $f = 0.2$ et $f = 0.3$, et cette coupure nous permet d'augmenter le ratio signal/bruit d'environ 2.

3.3 Une variable robuste

L'étude de cette variable sous PYTHIA nous a permis de remarquer que celle-ci était très robuste aux différents facteurs qui n'étaient pas sous contrôle pour le calcul analytique, que ce soit les variations de p_t ou les phénomènes non perturbatifs tels l'hadronisation et l'underlying event. Elle est sûre dans l'infrarouge, et les cuts en R_0 permettent d'assurer la sureté collinéaire. Nous pouvons constater cette robustesse en Figure 9 où le passage du niveau parton au niveau partons+hadrons+UE se fait sans changement de la forme de la distribution. Il est possible vérifier la robustesse vis à vis des variations de p_t Figure 12. Nous avons pu tester la variable pour des p_t allant de 200 GeV (cas où l'approximation d'un Higgs de masse 115 GeV très boosté n'est pas vraiment réaliste) à 400 GeV. Cette robustesse en fait une variable intéressante.

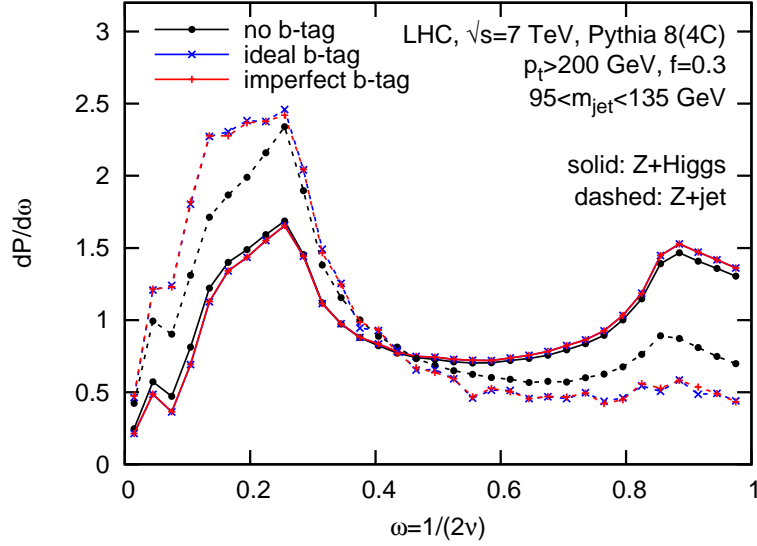


FIGURE 12 – Représentation des effets du p_t sur la distribution en ω , pour deux valeurs de p_t 200 GeV et 400 GeV. Ici $f = 0.3$.

4 Conclusion.

Dans ce rapport nous avons présenté certaines problématiques de la recherche du Higgs en collisionneur hadronique, introduit les bases de la physique des jets ainsi que le principe des méthodes de phénoménologie utilisées pour l'analyse des événements au LHC. Nous avons ensuite pu présenter un problème plus spécifique directement relié à l'expérience : le traitement du bruit QCD dans l'étude du canal de désintégration du Higgs en une paire $b\bar{b}$, canal important si il en est car c'est le canal dominant dans la gamme de masse où est attendu le Higgs, cependant l'importance du bruit de fond l'a longtemps rendu difficilement exploitable. Les travaux de Butterworth, Davidson, Rubin et Salam l'ont rendu plus accessible. Nous avons alors pu proposer une nouvelle méthode de réduction du bruit de fond en étudiant les connections de couleur du bruit QCD, étude rendue possible dans le cas des jets d'objets boostés car ils possèdent une sous-structure. Cette méthode originale pourrait éventuellement être appliquée à la recherche d'autres objets boostés non-colorés qui souffriraient d'un bruit QCD trop important.

Il reste pourtant encore beaucoup à étudier dans le cas présenté dans ce rapport. Premièrement une meilleure compréhension analytique pourrait être obtenue en relâchant certaines approximations, en particulier l'approximation de gluons mous (qui n'est pas toujours très raisonnable dans notre travail). Essayer de calculer, ou au moins d'estimer les effets de pile-up et d'underlying event, ainsi que les corrections d'hadronisation. Il reste encore à vérifier la stabilité de la variable par rapport à certains effets non perturbatifs d'émission de gluon mou. Enfin j'aimerais pouvoir généraliser la variable à plus de sous-jets de gluons (deux me semble un bon chiffre), les différences de connections de couleur devrait être plus marquée en considérant plus de sous-jets.

D'un point de vue plus personnel ce stage aura été l'occasion d'apprendre quelques rudiments de phénoménologie QCD aux cotés d'un des meilleurs du domaine, mais il aura aussi été l'occasion d'élargir mes connaissances en QCD et, de manière plus générale, en théorie des champs. Ainsi je me suis découvert un grand intérêt pour l'étude des états liés en QCD, aussi bien que pour la chromodynamique quantique à température finie.

Références

- [1] A. Abdesselam et al. Boosted objects : a probe of beyond the Standard Model physics. *Eur. Phys. J.*, C71 :1661, 2011.
- [2] Jonathan M. Butterworth, Adam R. Davison, Mathieu Rubin, and Gavin P. Salam. Jet substructure as a new Higgs search channel at the LHC. *Phys. Rev. Lett.*, 100 :242001, 2008.
- [3] Matteo Cacciari and Gavin P. Salam. Dispelling the N³ myth for the k(t) jet-finder. *Phys. Lett.*, B641 :57–61, 2006.
- [4] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti- k_t jet clustering algorithm. *JHEP*, 04 :063, 2008.
- [5] John C. Collins. Light cone variables, rapidity and all that. 1997.
- [6] John C. Collins. What exactly is a parton density? *Acta Phys. Polon.*, B34 :3103, 2003.
- [7] Michael E. Peskin and Daniel V. Schroeder. An Introduction to quantum field theory. Reading, USA : Addison-Wesley (1995) 842 p.
- [8] Tilman Plehn, Gavin P. Salam, and Michael Spannowsky. Fat Jets for a Light Higgs. *Phys. Rev. Lett.*, 104 :111801, 2010.
- [9] Mathieu Rubin. Non-Global Logarithms in Filtered Jet Algorithms. *JHEP*, 05 :005, 2010.
- [10] Gavin P. Salam. Elements of QCD for hadron colliders. 2010.
- [11] Gavin P. Salam. Towards Jetography. *Eur. Phys. J.*, C67 :637–686, 2010.
- [12] Gavin P. Salam and Gregory Soyez. A practical Seedless Infrared-Safe Cone jet algorithm. *JHEP*, 05 :086, 2007.
- [13] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178 :852–867, 2008.
- [14] George Stermán. Kinoshita’s theorem in yang-mills theories. *Phys. Rev. D*, 14(8) :2123–2125, Oct 1976.
- [15] George Stermán and Steven Weinberg. Jets from quantum chromodynamics. *Phys. Rev. Lett.*, 39(23) :1436–1439, Dec 1977.
- [16] Gerard ’t Hooft. A PLANAR DIAGRAM THEORY FOR STRONG INTERACTIONS. *Nucl. Phys.*, B72 :461, 1974.