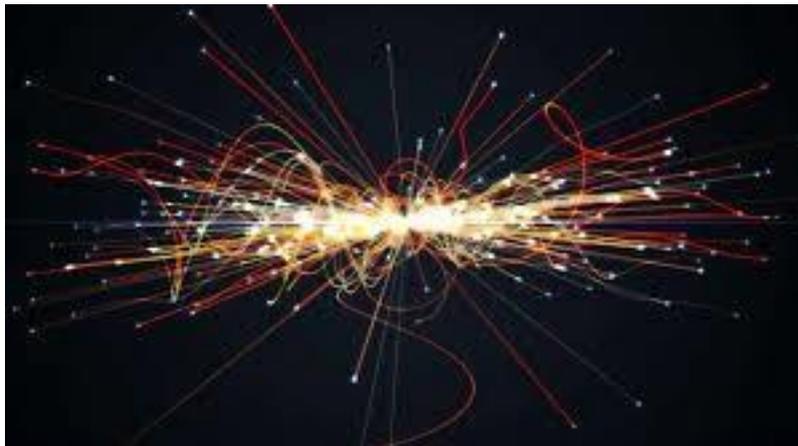


Rapport de stage en laboratoire au Laboratoire de Physique Nucléaire et
de Hautes Énergies (LPNHE), Paris :

**Reconstructions du boson Z dans l'expérience ATLAS auprès du
LHC**



Tuteur de stage : Frédéric Derue

Stage Licence de Physique (L3), Sorbonne Université

2024-2025



Remerciements :

Nous tenons à exprimer notre profonde gratitude à toute l'équipe du LPNHE de nous avoir accueilli au sein de leur laboratoire et de nous avoir fait découvrir avec enthousiasme et passion la vie du laboratoire.

Un merci particulier à notre tuteur de stage Frédéric Derue pour son accueil chaleureux et son encadrement précieux tout au long de notre stage. Cette expérience nous a permis d'enrichir nos connaissances et de développer de nouvelles compétences dans un cadre de travail stimulant et bienveillant. Merci pour le temps consacré à répondre à nos questions et pour les conseils avisés qui nous ont aidé à progresser. Nous avons beaucoup appris grâce aux échanges enrichissants. Ce stage a renforcé notre intérêt pour la recherche et nous a donné envie de poursuivre dans cette voie.

Un grand merci pour cette belle opportunité et pour votre bienveillance tout au long de notre stage.



Introduction

Le Laboratoire de Physique Nucléaire et des Hautes Énergies (LPNHE) de Paris est une unité mixte de recherche du CNRS, de Sorbonne Université et de l'Université Paris Cité situées sur le campus Pierre et Marie Curie (Jussieu). Cette unité participe à de nombreux programmes expérimentaux dans les domaines de la physique des particules, des astroparticules et de la cosmologie. Il participe aux expériences sur le LHC (Large Hadron Collider)¹, au CERN², notamment le détecteur ATLAS³, ou encore en cosmologie pour comprendre la matière et l'énergie noire avec les expériences Euclid⁴ et le télescope LSST⁵ installé au Chili. Nous avons été supervisés par Frédéric Derue appartenant au groupe ATLAS qui étudie actuellement la physique du quark-top et a beaucoup participé à la création de la grille de calcul du LPNHE.

La physique des particules est l'étude de l'infiniment petit, invisible à l'échelle humaine mais qui régit les lois de notre Univers. Suivant les découvertes de l'électron, du proton et du neutron il y a un siècle, cette discipline a développé dans les années 60 un cadre théorique permettant d'expliquer ses découvertes, le Modèle Standard⁶. Celui-ci arrive à décrire avec précision l'ensemble des particules élémentaires et leur interactions fondamentales comme la force électromagnétique, la force faible et la force forte. Les particules de matière, comme les quarks et les leptons, interagissent grâce à des particules élémentaires, appelées bosons.

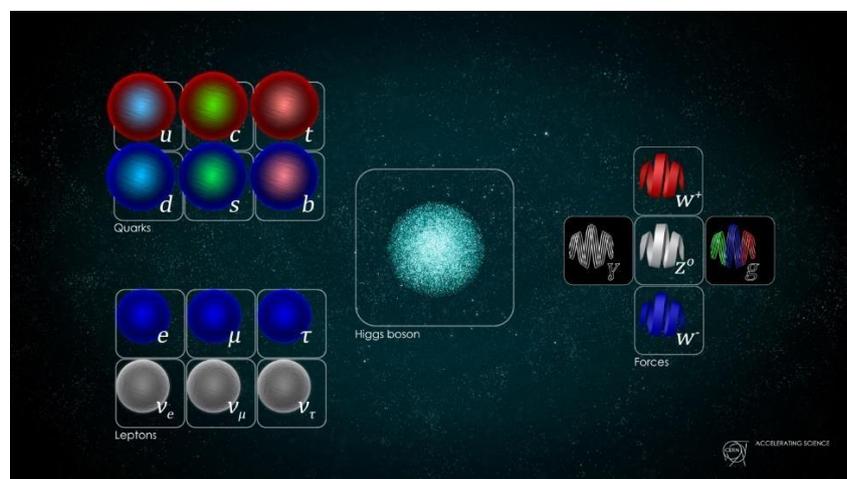


Figure 1 : Particules du Modèle Standard

En 2012, les collaborations ATLAS et CMS font une découverte attendue mais majeure sur le LHC, le boson de Higgs, à l'origine de la masse des particules élémentaires. Parmi ces particules élémentaires, le boson Z occupe une place importante, car il est avec le boson W associé à l'interaction faible ou force faible ; l'origine même du phénomène de désintégration de certains noyaux radioactifs.



Découvert en 1983 au CERN, le boson Z est une particule non-chargée qui a une durée d'existence extrêmement courte (de l'ordre de 10^{-25} secondes). Aucun capteur n'est encore assez rapide pour détecter cette particule mais on peut en déduire son existence grâce à l'étude de sa désintégration.

Les accélérateurs de particules comme le (LHC), situé au CERN, sont les outils principaux pour produire et étudier des particules comme le boson Z . Long de 27 km, il est le plus puissant au monde et fait entrer en collision des protons à des énergies allant jusqu'à 13.6 TeV. Lors de ces collisions à très hautes énergies (semblables à celles de l'Univers jeune), des particules élémentaires sont produites, comme les bosons Z . Le détecteur ATLAS est ainsi utilisé pour observer les désintégrations des différentes particules, ainsi que leurs éventuelles désintégrations.

Durant notre stage, nous avons essayé de reconstituer le boson Z à l'aide de données de collisions venant du détecteur ATLAS et de données simulées. Nous avons mené ce travail en suivant trois approches, chacune se rapprochant davantage de celles utilisées par les chercheurs actuellement. Tout d'abord, nous avons utilisé un programme, appelé Hypatia et utilisé lors de MasterClasses⁷, pour visualiser des collisions protons-protons dans le détecteur ATLAS et sélectionner "à l'oeil" des bosons Z . Puis, en récupérant des lots de données simulant des collisions produisant des bosons Z se désintégrant en paires électrons-positrons, disponibles en Sciences Ouvertes (OpenData⁸), nous avons reconstitué la masse invariante du boson Z à l'aide d'un programme d'analyse écrit en Python dans un Jupyter notebook. Enfin, nous avons utilisé des lots de données simulées, dans un format plus proche de celui utilisé réellement en recherche.



Partie 1 : Étude et identification des électrons dans les données du détecteur ATLAS

Le détecteur ATLAS est l'un des instruments clés du Grand Collisionneur de Hadrons (LHC) au CERN. Conçu pour explorer les fondements du Modèle Standard et rechercher une éventuelle nouvelle physique, il permet d'analyser les produits des collisions de protons en identifiant les particules secondaires générées.

La reconstruction et l'identification des électrons constituent un enjeu crucial pour l'étude des désintégrations de bosons lourds, notamment du boson de Z , qui se désintègre dans environ 6% des cas en une paire électron-positron ou en une paire de muon-antimuon.

Méthodologie et données utilisées

Les données exploitées dans cette étude proviennent de collisions enregistrées en 2012 par la collaboration ATLAS. Elles ont été traitées par un logiciel appelé Hypatia qui traduit les signaux électriques des sous-détecteurs en objets exploitables, tels que des traces dans le trajectographe ou des dépôts d'énergie dans les calorimètres.

Description du détecteur ATLAS

Le détecteur ATLAS est constitué de plusieurs sous-détecteurs :

1. **Le trajectographe interne** : plongé dans un champ magnétique, il reconstruit les trajectoires des particules chargées et permet d'en déduire leur charge et leur impulsion.
2. **Le calorimètre électromagnétique** : il mesure l'énergie déposée par les électrons et photons, générant des gerbes électromagnétiques.
3. **Le calorimètre hadronique** : il mesure la position et l'énergie des hadrons.
4. **Le spectromètre à muons** : situé à l'extérieur, il est dédié à l'identification des muons, qui traversent les autres sous-détecteurs avec une faible interaction.

Sélection et identification des différentes particules

L'identification des particules repose sur plusieurs critères discriminants :

- Un électron est caractérisé par une trace bien définie dans le trajectographe, un dépôt d'énergie compact dans le calorimètre électromagnétique et un rapport énergie/impulsion proche de 1, dû à sa faible masse. À l'inverse, les hadrons produisent des gerbes plus diffuses et ne déposent qu'une fraction de leur énergie dans le calorimètre électromagnétique (voir Figure 2).

- Les muons sont les seules particules qui atteignent le spectromètre à muons. Ainsi il est facile de les distinguer (voir Figure 3).
- Les photons sont assez faciles à reconnaître car ils ne laissent aucune trace dans le trajectographe et déposent de l'énergie dans le calorimètre électromagnétique (voir Figure 4).

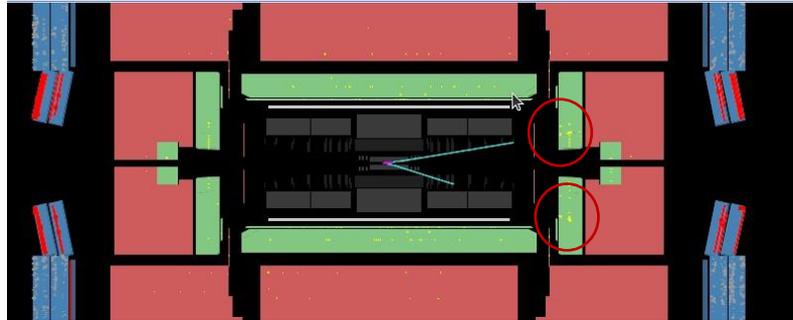


Figure 2 : Une paire d'électrons dans le détecteur ATLAS en coupe transversale. On reconnaît les deux dépôts énergétiques alignés avec leur trace.

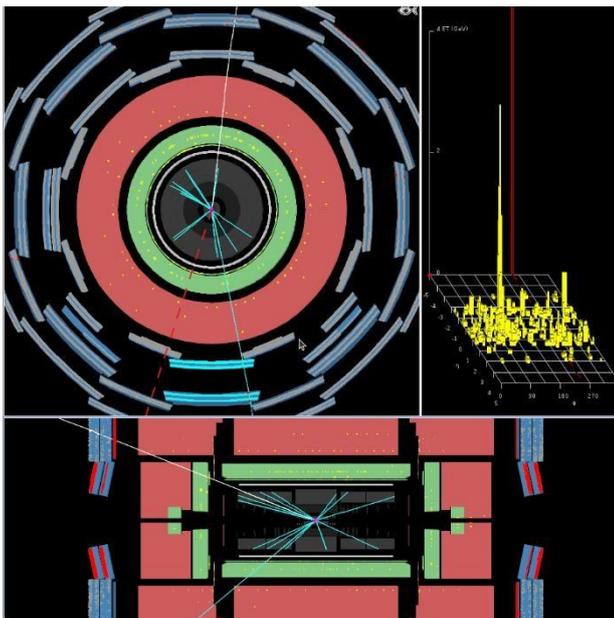


Figure 3 : Une paire de muons. On reconnaît les traces des muons qui atteignent le spectromètre à muons.

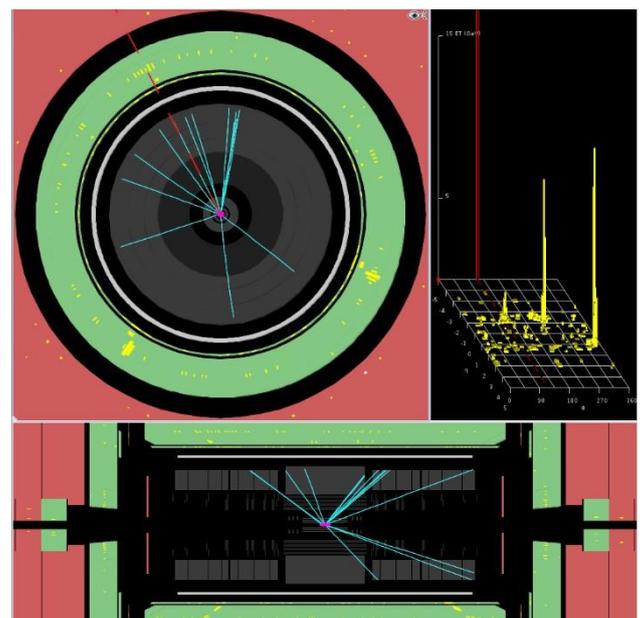


Figure 4 : Une paire de photons. On reconnaît les deux dépôts énergétiques sans trace qui les précède.

Nous avons sélectionné "à l'oeil" quelques 300 bosons Z à partir d'un lot de près d'un millier d'événements/photos avec Hypathia. Ces identifications purement qualitatives nous ont permis de mieux comprendre les comportements des différentes particules dans le détecteur ATLAS. Même si des résultats ont été obtenus durant les tous premiers jours du stage, de bien meilleurs l'ont été ensuite et vont être maintenant décrits.



Partie 2 : Reconstruction du boson Z avec des données OpenData

Suite à nos sélections des différentes particules dans notre détecteur ATLAS, il vient maintenant de reconstruire la masse invariante du boson Z. Nous aurions pu utiliser le fichier construit à partir des traces sélectionnées sur Hypatia mais nous avons un fichier d'un millier de traces environ ce qui est bien trop faible pour une étude du boson Z. Nous avons utilisé un lot de données simulant des collisions produisant des bosons Z, environ 650000, se désintégrant en paires électrons-positrons, disponibles en Sciences Ouvertes (OpenData⁸). Nous avons exploité ces données à l'aide d'un programme d'analyse écrit en Python dans un Jupyter notebook. Le lot de traces se présente sous la forme d'un fichier, dans un format usuel de physique des particules appelé ROOT, qu'on transforme sous la forme d'une dataframe (une sorte de tableau

2×2 utilisé en programmation sous python). La particularité du lot est qu'il ne comporte que des particules ayant passé les critères d'identification de la collaboration et ne comportent presque pas de bruit de fond. Nous pouvons également accéder à son énergie, son impulsion transversale, les angles par rapport à l'axe du faisceau (transversale et longitudinale) notés ϕ et η .

Calcul du boson Z pour la désintégration $Z \rightarrow e^-e^+$

On peut désormais calculer la masse invariante de chaque paire d'électrons à l'aide de la formule suivante :

$$M = \sqrt{\left(\frac{E_{e^-} + E_{e^+}}{c^2}\right)^2 - \left(\frac{p_{e^-} + p_{e^+}}{c}\right)^2}$$

Une fois la masse invariante calculée pour chaque paire, on construit un histogramme avec la masse invariante en abscisse et le nombre de paires d'électrons correspondant à cette masse.

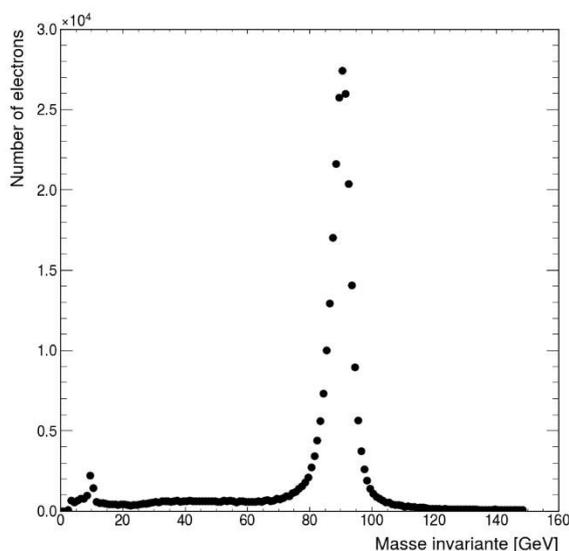


Figure 5 : Reconstruction de la masse invariante

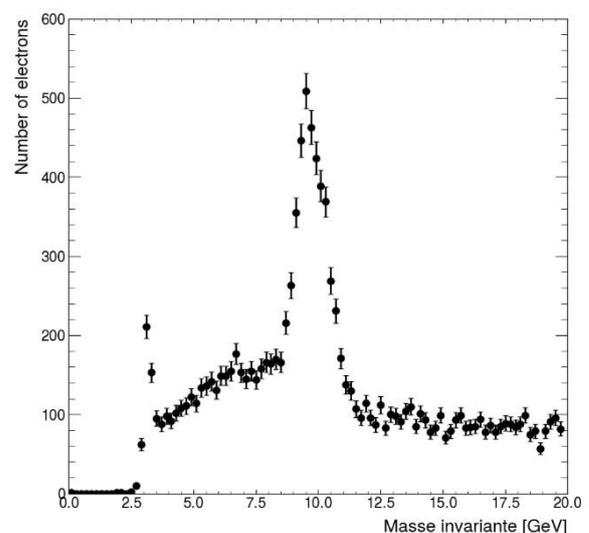


Figure 6 : Zoom entre 0 et 20 GeV de la figure 5 du boson Z



La figure 5 montre cet histogramme en 0 et 160 GeV, et la figure 6 montre un zoom dans la zone de basses masses, en dessous de 20 GeV. Sur la figure 5, nous observons un pic prononcé aux alentours de 91 GeV. Autour de ce pic, se trouve un continuum de valeurs plus basses ou plus hautes, correspondant à des associations fortuites de paires de traces, ne provenant pas de la désintégration d'une même particule.

Sur la figure 6, les deux pics prononcés autour de 3 GeV et 9 GeV correspondent respectivement au J/ψ un méson neutre constitué d'un quark charmé et d'un quark anti-charmé et Υ un méson neutre constitué d'un quark bottom et d'un quark anti-bottom. Ces deux mésons sont aussi issus d'une désintégration en paire d'électrons.



Partie 3 : Reconstruction de la masse invariante du boson Z avec bruit de fond

Dans la section précédente, les données de collisions analysées étaient préalablement nettoyées et exemptes de tout bruit de fond. Dans cette partie, nous utilisons des données simulées. La nature de chaque particule est gardée en mémoire et authentifiée. De plus, ces particules n'ont pas encore passées les critères d'identification de la collaboration. On a ainsi affaire à un format bien plus proche de celui réellement utilisé par les chercheurs dans leurs analyses.

Comme nous l'avons observé dans la première partie, l'identification des particules n'est pas une tâche facile et certaine.

L'objectif de cette dernière partie est donc de reconstruire la masse invariante du boson Z à partir de données, en connaissant le type de particules dans un premier temps et de définir des critères d'identification qui vont nous permettre de nous affranchir de cette connaissance qui n'existe que pour les données simulées. Le fichier nous donne accès à plusieurs informations sur les différentes traces et nous savons comment chaque particule doit se comporter dans le calorimètre ATLAS. Pour reconnaître les paires d'électrons, nous devons donc établir un discriminant pour plusieurs paramètres à l'aide d'une analyse statistique.

Le fichier traité restera issu d'une simulation ce qui nous aidera au début pour avoir accès au type de la particule. Néanmoins, nous nous en détacherons pour construire la masse invariante du boson Z.

1) Détermination des paramètres discriminants

A) Figures

On choisit de construire des histogrammes pour différents paramètres, sur chaque figure nous aurons en abscisse la valeur du paramètre et en ordonnée le nombre de toutes les particules et d'électrons correspondant.

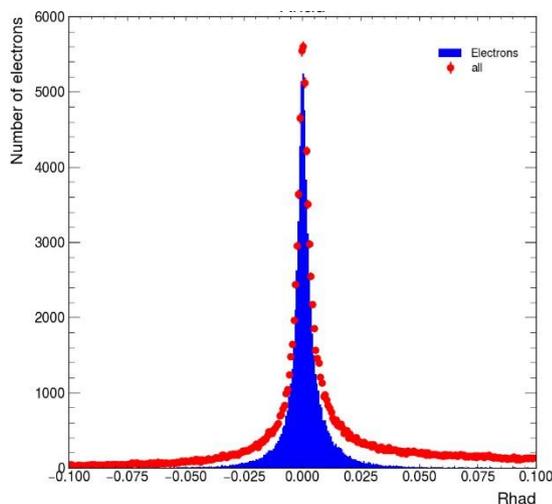


Figure 7: Nombre de pions et d'électrons en fonctions du paramètre $Rhad$

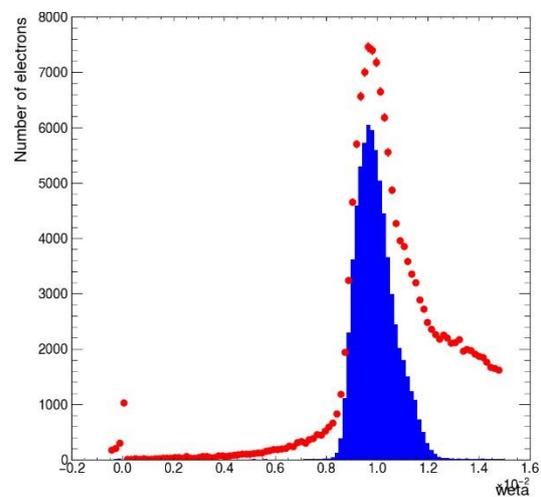


Figure 8: Nombre de pions et d'électrons en fonctions du paramètre $weta$

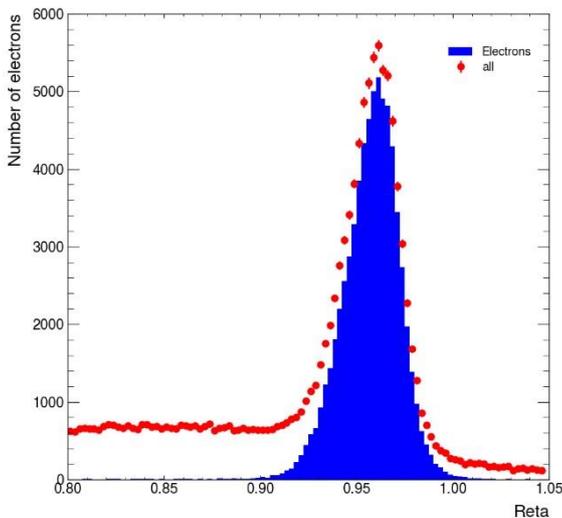


Figure 9 : Nombre de pions et d'électrons en fonctions du paramètre **Reta**

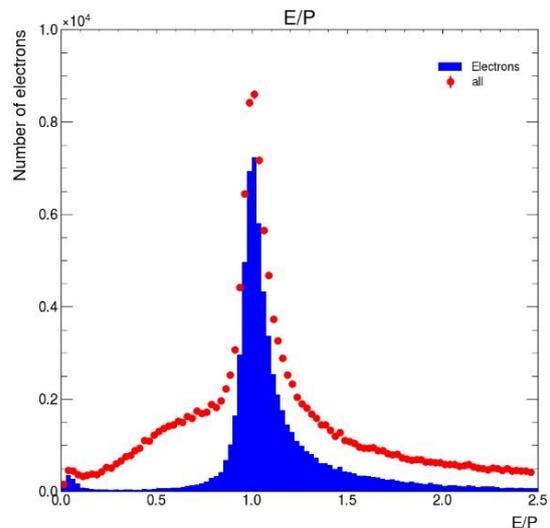


Figure 7 : Nombre de pions et d'électrons en fonctions du paramètre **E/P**

B) Définition des paramètres⁹

La variable **Rhad** correspond au dépôt énergétique de l'électron dans la partie hadronique du calorimètre. Or, nous l'avons vu dans la première partie l'électron dépose seulement son énergie dans le calorimètre électromagnétique. Ainsi **Rhad** pour un électron doit être égal à environ 0. La variable **Reta** est associé à la concentration longitudinale du dépôt d'énergie. L'électron a son dépôt d'énergie qui ne se diffuse presque pas dans le calorimètre électromagnétique au contraire par exemple d'un jet d'hadron. Soit **Reta** doit être aux alentours d'1.

La variable **weta** correspond à la largeur de la gerbe transversale. Comme on l'a dit la trace de l'électron est très concentré ainsi, on s'attend à avoir **weta** très proche de 0 pour un électron. La variable **E/P** est le rapport entre l'énergie et l'impulsion du pions. Or, on sait que la masse de l'électrons est très petite devant son énergie et son impulsion. On en déduit facilement que **E/P** doit se situer aux alentours de 1.

C) Exploitation des graphiques

Les distributions issues des électrons apparaissent sur chaque figure comme un pic et nous voyons bien le bruit de fond apparaître en faisant la comparaison entre l'histogramme des électrons et des pions. Nous définissons pour chaque paramètre un intervalle de sélection dans lequel les traces seront gardées. On peut ainsi calculer, pour ces intervalles, une efficacité de sélection pour les électrons et un facteur de rejet pour les pions.



	Weta	Rhad	Reta	E/P
Intervalle discriminant	[0.008 ; 0.0125]	[-0.02 ; 0.02]	[0.9 ; 1]	[0.6 ; 2]

$$\text{Efficacité de sélection : } \epsilon = \frac{\text{electrons}_{\text{sélectionnés}}}{\text{Particules}_{\text{total}}}$$

$$\text{Facteur de rejet : } R = \frac{\text{Pions}_{\text{tot}}}{\text{Pions}_{\text{sélectionnés}}}$$

En reliant nos intervalles discriminants, nous arrivons à une efficacité de sélection de **90%** et un facteur à rejet d'environ **900** (soit seulement 0,11% des pions sont passés à travers nos filtres).

NB : Nous avons simplement multiplier les efficacités et les facteurs de rejet pour chaque paramètre entre eux afin d'obtenir l'efficacité de sélection et le facteur de rejet total.

D) Ratio Efficacité- Facteur de rejet : Importance du Machine Learning

On peut aussi s'amuser avec les intervalles car on se pose la question : est-ce que contraindre notre efficacité serait intéressant afin d'obtenir un facteur de rejet important ?

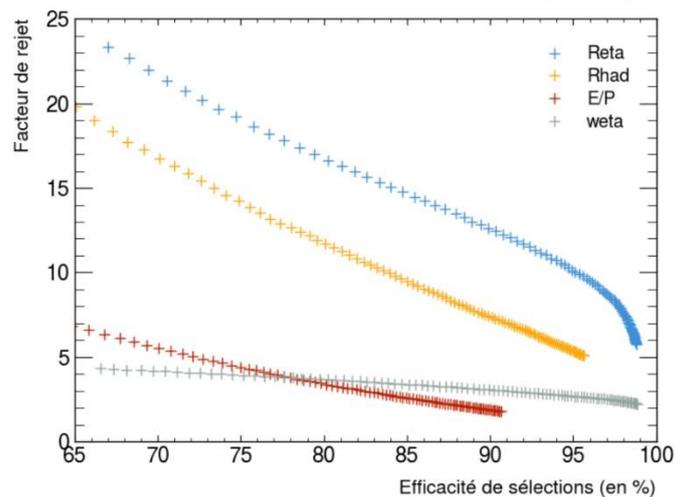


Figure 11 : Efficacité de sélection en fonction du facteur de rejet pour chaque paramètre



Il apparaît sur ce graphique que les paramètres discriminants **E/P** et **weta** sont bien moins utiles que les variables **Reta** et **Rhad**. En effet, ces dernières peuvent rejeter jusqu'à 99% des pions en gardant une efficacité à 90%. De plus, on remarque que réduire notre efficacité de sélection pour certaines variables n'augmente pas le facteur de rejet.

Dans cette partie, nous avons calculé nos efficacités et vérifié notre modèle à la main. Nous voyons bien qu'il serait bien plus compliqué si l'on voulait calculer des intervalles variables pour chaque paramètre (cela nous donnerait des millions de possibilités !). Le machine Learning est une alternative. Il utilise un processus d'intelligence artificielle ; il va relier de manière automatisée tous les paramètres entre eux afin de vérifier la validité du modèle ou non. Lors de ce stage, nous n'avons que très brièvement travaillé avec cette méthode, à travers les bibliothèques python scikit disponibles facilement dans l'environnement de travail python/Jupyter à disposition, mais il est tout de même intéressant de comparer les quelques résultats.

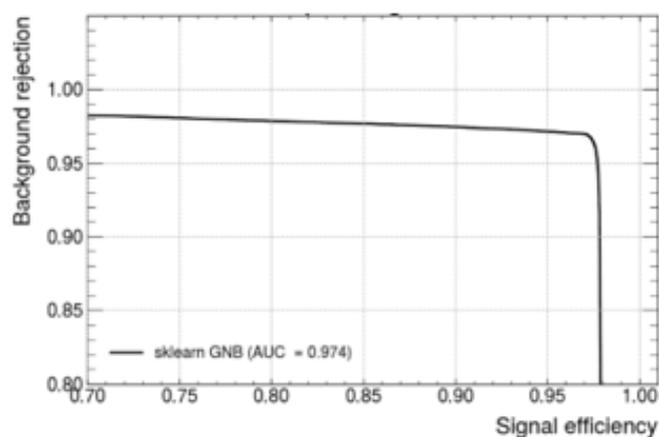


Figure 7 : Efficacité de sélection en fonction de $(1 - \frac{1}{Rejet})$ produit par le machine Learning

La figure 12 montre le facteur de rejet en fonction de l'efficacité de sélection, en utilisant l'une des méthodes disponibles dans scikit, avec un maximum de vraisemblance combinant les informations des histogrammes obtenus précédemment. La figure est obtenue facilement en quelques secondes, bien plus facilement qu'un travail d'optimisation des intervalles que nous aurions dû faire à la main.

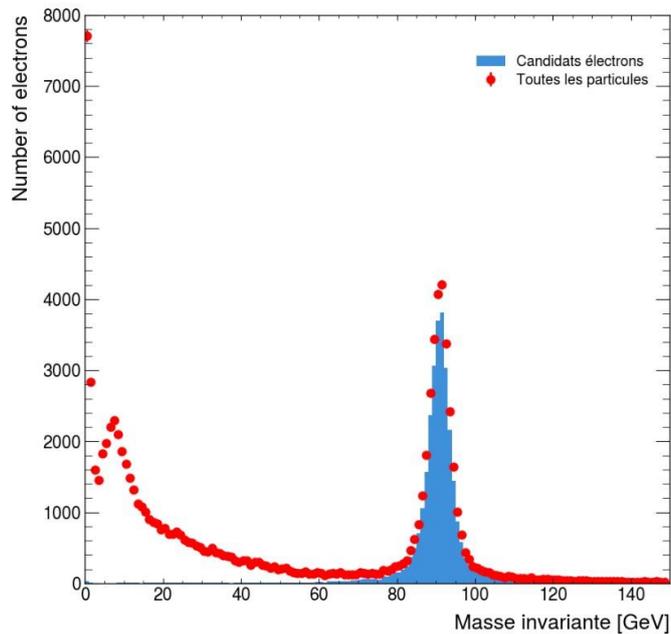
2) Reconstruction de la masse invariante du boson Z

Figure 13 : Reconstruction de la masse invariante du boson Z avec bruit de fond

La figure 13 montre la distribution de masse invariante obtenue pour toutes les paires de traces de charges opposées (en rouge), et pour celles qui sont sélectionnées par les critères définis dans la table ci-dessus (en bleu). Il apparaît clairement que notre méthode de sélection grâce aux paramètres discriminants a fonctionné. En effet, nous voyons bien que nous avons éliminé le bruit de fond sur la Figure 8 (différence entre les points rouges et l'histogramme bleu) et que nous avons le pic prononcé à 91 GeV comme attendu qui résulte de la désintégration d'une paire d'électrons.



Conclusion

Ce stage au LPNHE nous a offert une première immersion dans le domaine de la physique des particules et une opportunité d'acquérir une première expérience en analyse de données issues du détecteur ATLAS. En reconstituant la masse invariante du boson Z , nous avons appris, mis en pratique des concepts clés du modèle standard et compris l'importance des techniques de sélection des particules dans un environnement expérimental complexe.

Grâce à différentes approches, allant de l'utilisation du logiciel Hypatia à l'exploitation de grands ensembles de données sous Python, nous avons observé la signature caractéristique du boson Z autour de 91 GeV. L'introduction du bruit de fond a également mis en évidence les défis liés à l'identification précise des électrons et à l'importance des paramètres discriminants. Cette analyse, bien que simplifiée, illustre les méthodes employées dans la recherche actuellement.

Finalement, l'analyse exposée met en lumière l'efficacité des techniques de reconstruction de la masse invariante pour extraire la signature du boson Z dans les données du LHC, confirmant ainsi de manière objective les prédictions du modèle standard.



Références bibliographiques :

- ¹ Large Hadron Collider, site web de LHC France, <https://www.lhc-france.fr/>
- ² CERN, Organisation Européenne pour la Recherche Nucléaire, <https://home.cern/fr>
- ³ The ATLAS experiment, Monica Lynn Dunford and Peter Jenni (2014), Scholarpedia, 9(10) :32147, http://www.scholarpedia.org/article/The_ATLAS_experiment
- ⁴ Euclid,
- ⁵ Observatoire Vera Rubin/LSST, <https://www.lsst.org/>
- ⁶ « *Introduction à l'aventure du grand collisionneur LHC : du big bang au boson de Higgs* », Daniel Denegri, Claude Guyot, Andreas Hoecker et Lydia Roos, EDP Sciences (2014)
- ⁷ ATLAS Master Classes, <https://atlas.physicsmasterclasses.org/fr/zpath.htm>
- ⁸ ATLAS OpenData, <https://atlas.cern/Resources/Opendata>
- ⁹ « *Reconstruction et identification des électrons dans l'expérience Atlas* », Habilitation à diriger des recherches, Frédéric Derue (2008)