

---

# Rapport de Stage de L3

---

Vanessa ARNAL

Magistère de Physique

Université Denis Diderot Paris 7

Identification des électrons provenant de la désintégration du  
 $J/\psi$  dans l'expérience Atlas au LHC

*LPNHE-Paris, Universités Paris 6 et Paris 7, et IN2P3-CNRS*

Directeur de stage : Frédéric Derue



<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Présentation du laboratoire . . . . .	2
1.2	Le groupe Atlas au LPNHE . . . . .	2
1.3	Le Modèle Standard de la Physique des Particules . . . . .	2
<b>2</b>	<b>L'expérience Atlas</b>	<b>3</b>
2.1	Le LHC . . . . .	3
2.2	Présentation du détecteur . . . . .	4
2.2.1	Le détecteur interne . . . . .	4
2.2.2	Le calorimètre électromagnétique . . . . .	4
2.2.3	Le calorimètre hadronique . . . . .	6
2.2.4	Le spectromètre à muons . . . . .	6
<b>3</b>	<b>Reconstruction du <math>J/\psi</math></b>	<b>6</b>
3.1	Les données utilisées . . . . .	6
3.2	Reconstruction des électrons . . . . .	7
3.3	Identification des électrons . . . . .	7
3.3.1	Variables discriminantes . . . . .	7
3.3.2	Efficacité d'identification des électrons . . . . .	11
3.4	Reconstruction du $J/\psi$ . . . . .	11
<b>4</b>	<b>Etudes sur l'amélioration de l'identification des électrons</b>	<b>12</b>
4.1	Méthodes utilisées . . . . .	12
4.2	Performances des différentes méthodes . . . . .	14
<b>5</b>	<b>Conclusion</b>	<b>17</b>

## 1.1 Présentation du laboratoire

Le Laboratoire de Physique Nucléaire et de Hautes Energies (LPNHE), situé sur le campus universitaire de Jussieu, est une unité mixte de recherche de l'Institut National de Physique Nucléaire et de Physique des Particules (IN2P3) qui constitue le département de physique nucléaire et corpusculaire du CNRS. Cet institut regroupe plus de 2000 scientifiques impliqués dans une cinquantaine d'expériences nationales et internationales, notamment au travers de partenariats avec le CERN (Laboratoire Européen de Physique des Particules, Genève), le Fermilab (Chicago) et le Slac (Stanford).

Le laboratoire regroupe environ 110 personnes dont 51 chercheurs ou enseignants chercheurs, principalement impliqués dans les domaines expérimentaux de la physique des hautes énergies, la cosmologie et la physique des astroparticules, au travers de nombreuses expériences telles que LHCb, BABAR, CDF, D0, AUGER, HESS et Atlas.

## 1.2 Le groupe Atlas au LPNHE

Le groupe Atlas, dirigé par Philippe Schwemling, est constitué de 17 physiciens, dont 4 thésards, et d'une équipe de 3 techniciens et ingénieurs en informatique, électronique et mécanique. Il est impliqué dans la réalisation du calorimètre électromagnétique du détecteur Atlas ainsi que dans la préparation des analyses de physique.

## 1.3 Le Modèle Standard de la Physique des Particules

Les nombreuses avancées expérimentales et théoriques dans le domaine de la physique des particules ont conduit dans les années 60 et 70 au développement du Modèle Standard. C'est la théorie actuelle qui permet de décrire les constituants élémentaires de la matière ainsi que trois des quatre interactions fondamentales (électromagnétique, faible et forte). Il distingue deux types de particules : les particules constituant la matière qui sont les fermions et les particules véhiculant les interactions qui sont les bosons de jauge.

Les fermions sont divisés à leur tour en deux catégories de particules : les leptons et les quarks (cf Tab. 1).

	1 <sup>ère</sup> famille	2 <sup>ème</sup> famille	3 <sup>ème</sup> famille
Quarks	$\begin{pmatrix} \textit{haut } u \\ \textit{bas } d \end{pmatrix}$	$\begin{pmatrix} \textit{charmé } c \\ \textit{étrange } s \end{pmatrix}$	$\begin{pmatrix} \textit{top } t \\ \textit{beau } b \end{pmatrix}$
Leptons	$\begin{pmatrix} \textit{neutrino } \nu_e \\ \textit{électron } e^- \end{pmatrix}$	$\begin{pmatrix} \textit{neutrino } \nu_\mu \\ \textit{muon } \mu \end{pmatrix}$	$\begin{pmatrix} \textit{neutrino } \nu_\tau \\ \textit{tau } \tau \end{pmatrix}$

TAB. 1 – *Les particules de type fermion du Modèle Standard*

Les leptons, comme les quarks, sont regroupés en trois familles, chacune étant constituée de deux leptons (un lepton chargé et de son neutrino associé) ou de deux quarks. Parmi ces familles, seule la première est stable et suffit à décrire la matière que nous connaissons. A ces trois familles, il faut ajouter les trois familles composées des antiparticules.

Les bosons de jauge sont les messagers des interactions entre les particules. Les photons  $\gamma$  véhiculent l'interaction électromagnétique qui agit sur les particules chargées. Les bosons



Le détecteur Atlas (A Toroidal LHC ApparatuS) (cf Fig. 2) est le plus grand des détecteurs jamais construit pour une expérience de physique des particules. De forme cylindrique, il mesure 25 m de haut et 44 m de longueur pour un poids total d'environ 7000 tonnes.

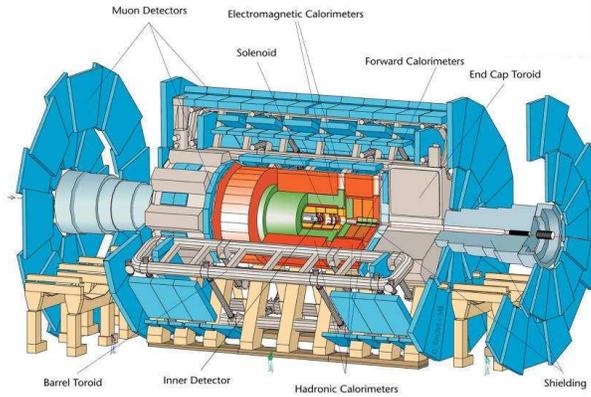


FIG. 2 – *Vue schématique du détecteur Atlas*

En partant du point d'interaction des deux faisceaux de protons au centre du détecteur, on distingue quatre sous-détecteurs : le détecteur interne, le calorimètre électromagnétique, le calorimètre hadronique et le spectromètre à muons.

Le système de coordonnées utilisé pour repérer les événements dans Atlas est tel que l'axe des  $z$  corresponde à celui des deux faisceaux et les axes  $x$  et  $y$  définissent le plan transverse. Du fait de la symétrie cylindrique du détecteur, on définit l'angle azimuthal  $\phi$  correspondant aux rotations autour de l'axe  $z$  et l'angle polaire  $\theta$  par rapport à l'axe  $z$ . Cependant on préférera employer la pseudo-rapacité  $\eta$  donnée par :  $\eta = -\ln \tan \left( \frac{\theta}{2} \right)$ .

### 2.2.1 Le détecteur interne

Le détecteur interne ou trajectographe, subdivisé en trois sous-détecteurs, a pour but de déterminer la trajectoire des particules chargées et de reconstruire les vertex primaires (point d'interaction des deux faisceaux) et secondaires (désintégration de particules). Les deux premiers sont composés de trois couches de pixels de silicium (détecteur à pixels) et de détecteur à micropistes (SCT, "semi-conductor tracking"), ayant une technologie similaire mais avec une granularité ( $\Delta\eta \times \Delta\phi$ ) moins grande. Le troisième est un sous-détecteur à tubes à pailles (TRT, Transition Radiation Tracker). Chaque trace chargée est échantillonnée sur 36 points par ce détecteur.

L'ensemble de ce détecteur est entouré d'un aimant solénoïdal qui crée un champ magnétique permettant de courber la trajectoire des particules. Il est ainsi possible de mesurer la charge et l'impulsion des particules.

### 2.2.2 Le calorimètre électromagnétique

Le rôle du calorimètre électromagnétique est de mesurer avec précision la position et l'énergie des électrons et des photons. Il est fait d'un assemblage de plaques de plomb pliées en "accordéon" (milieu absorbeur) et séparées par des électrodes et de l'argon liquide (milieu actif). Une telle structure élimine les espaces non instrumentalisés et permet

incidentes est transformée en gerbe (cf Fig. 4).

Ce détecteur est constitué de deux parties distinctes : la partie centrale en forme de tonneau (barrel) couvre le domaine de pseudo rapidité  $|\eta| < 1.37$  et les deux “bouchons” (end cap) couvrent le domaine  $1.52 < |\eta| < 3.2$ . La région  $1.37 < |\eta| < 1.52$ , qui correspond à la zone de transition entre le tonneau et les bouchons, est peu instrumentalisée ; par la suite elle sera exclue de l’algorithme d’identification des électrons (cf paragraphe 3.2).

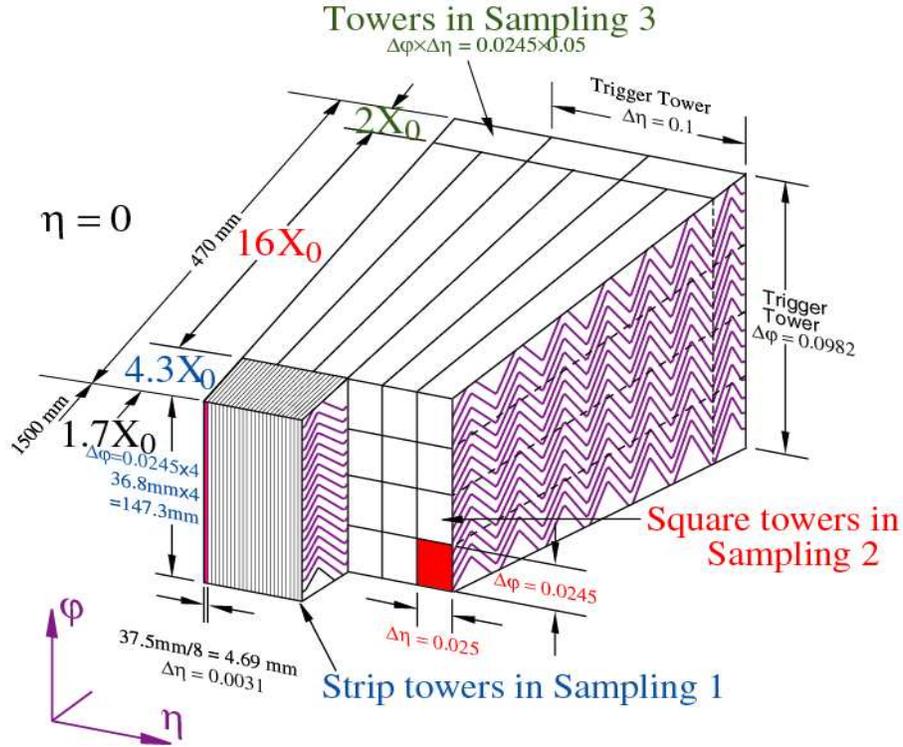


FIG. 3 – Structure en compartiments et granularité des cellules du calorimètre électromagnétique

La calorimètre électromagnétique est divisé en trois compartiments (cf Fig. 3) qui seront utilisés pour identifier les électrons :

- le premier compartiment possède une granularité très fine afin de mesurer avec une grande précision l’angle et la position du début du développement de la gerbe.
- le second compartiment contient l’essentiel de la gerbe et permet une bonne séparation entre les gerbes électromagnétiques et celles issues des pions.
- le troisième compartiment permet la mesure des gerbes de très grandes énergies qui ne seraient pas entièrement contenues dans les deux premiers compartiments.

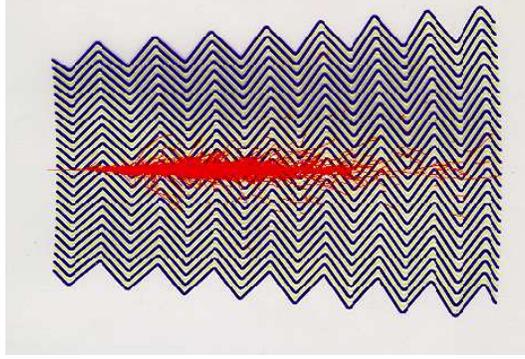


FIG. 4 – *Simulation du passage d'un électron dans le calorimètre électromagnétique (développement d'une gerbe).*

### 2.2.3 Le calorimètre hadronique

Le calorimètre hadronique ou HCAL est composé d'absorbeurs en fer séparés par des tuilles scintillantes. Il sert à mesurer l'énergie déposée et la direction des hadrons (protons, pions, kaons, ...) qui, lors de leur passage dans les absorbeurs développent des gerbes hadroniques.

### 2.2.4 Le spectromètre à muons

Le spectromètre à muons constitue avec les bobines supraconductrices qui l'entourent la couche externe du détecteur. Son but est de reconstruire la trajectoire des muons, seules particules à atteindre cette partie du détecteur et n'interagissant que peu avec la matière. Comme dans le trajectographe, les bobines induisent un champ électromagnétique qui permet de mesurer l'impulsion des muons.

## 3 Reconstruction du $J/\psi$

Le  $J/\psi$  est un méson composé d'un quark  $c$  et d'un anti-quark  $c$  ayant une masse de  $3.09687 \text{ GeV}/c^2$ . Dans cette étude, on s'est intéressé aux électrons de basse impulsion transverse, typiquement  $5 - 7 \text{ GeV}/c^2$ , provenant de la désintégration  $J/\psi \rightarrow e^+e^-$ .

### 3.1 Les données utilisées

On a utilisé les données qui ont été simulées à l'aide du générateur Monte Carlo Pythia dans le cadre de la production de données dite CSC (Computing System Commissioning). Il permet de simuler la production de particules issues d'une collision proton-proton ainsi que les différentes désintégrations qui peuvent avoir lieu et fournit en sortie un fichier avec l'ensemble des particules produites.

A l'issue de cette simulation, on utilise le logiciel Geant qui simule les éléments du détecteur et permet ainsi d'obtenir la réponse des différentes parties du détecteur lors du passage des particules.

Les données obtenues précédemment ne sont pas directement exploitables dans des analyses, elles doivent être préalablement reconstruites à l'aide du logiciel Athena.

Au niveau du détecteur interne, les traces des particules chargées sont reconstruites à partir des différents coups dans les détecteurs de précision (pixels + SCT) et dans le TRT. Cependant, on ne veut reconstruire que celles pouvant être par la suite identifiées comme étant des électrons. Pour cela on applique une série de coupures de “bonne qualité” des traces :

- une impulsion transverse  $p_T$  supérieure à 2 GeV ;
- au moins deux points de mesure dans le détecteur à pixels ;
- au moins un point de mesure dans la couche “b” du détecteur à pixels ;
- au moins sept points de mesure dans les détecteurs de précision (pixels + SCT).

L'ensemble des traces est alors extrapolé jusqu'au calorimètre électromagnétique, où plusieurs étapes sont nécessaires à la reconstruction :

- Autour du point extrapolé les cellules sont regroupées en un amas de taille  $\Delta\eta \times \Delta\phi = 0,025 \times 0,025$  ( $5 \times 5$ ). L'énergie et la position sont alors mesurées.
- Les différents paramètres caractérisant les gerbes sont mesurés dans ces amas.

A la fin de la reconstruction on obtient un fichier contenant les données reconstruites dans chaque partie du détecteur sous forme d'un tableau (Ntuple).

L'algorithme développé en C++ au cours de ce stage utilise ces données afin d'identifier les électrons et les pions parmi l'ensemble des particules produites lors d'une collision proton-proton.

## 3.3 Identification des électrons

### 3.3.1 Variables discriminantes

Afin de reconstruire le  $J/\psi$ , il faut séparer les électrons provenant de sa désintégration et constituant le signal des autres particules, dans notre cas des pions, qui composent le bruit de fond. La méthode la plus simple consiste à appliquer une série de coupures sur un certain nombre de variables discriminantes provenant des calorimètres et du détecteur interne. On suit la procédure standard d'identification des électrons dans Atlas [F.D08].

#### *Calorimètre hadronique*

Les électrons ne déposent qu'une faible fraction de leur énergie dans le calorimètre hadronique. Ainsi on définit les fuites hadroniques, notées hadleak, comme le rapport entre l'énergie transverse déposée dans le premier compartiment du calorimètre hadronique dans une fenêtre  $\Delta\eta \times \Delta\phi = 0,2 \times 0,2$  et l'énergie transverse déposée dans le calorimètre électromagnétique. La figure 5 montre la distribution obtenue pour des électrons et des pions.

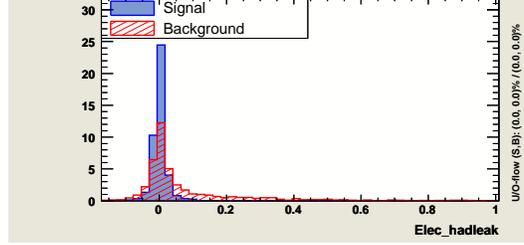


FIG. 5 – *Fuites hadroniques, définies comme le rapport entre l'énergie transverse reconstruite dans le premier compartiment du calorimètre hadronique, dans une fenêtre  $\Delta\eta \times \Delta\phi = 0,2 \times 0,2$  et l'énergie transverse reconstruite dans le calorimètre électromagnétique. Les distributions sont montrées pour des électrons (histogramme en trait plein) et des pions (histogramme hachuré).*

### *Second compartiment du calorimètre électromagnétique*

Dans le calorimètre électromagnétique, les électrons sont caractérisés par des gerbes plus étroites que celles provenant des pions. Les variables caractérisant ces gerbes vont être utilisées pour rejeter une partie du bruit de fond (cf Fig. 6) :

- le développement latéral  $R_\eta(37)$  de la gerbe est défini comme le rapport entre l'énergie déposée dans une fenêtre  $\Delta\eta \times \Delta\phi = 3 \times 7$ , exprimée en nombre de cellules, et l'énergie déposée dans une fenêtre  $7 \times 7$ .
- la largeur latérale  $\omega_{\eta^2}$  est calculée dans une fenêtre  $3 \times 5$ .

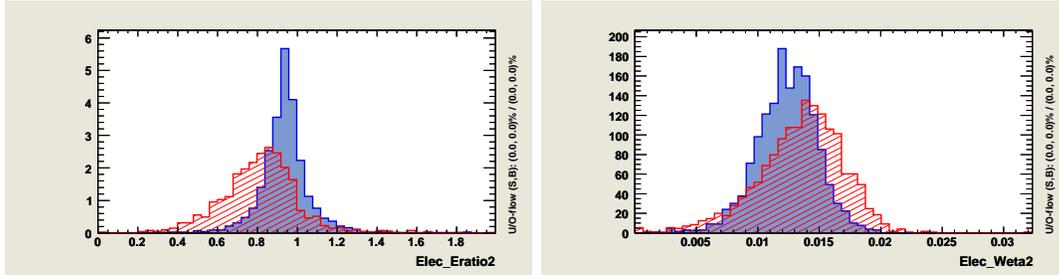


FIG. 6 – *Développement latéral de la gerbe  $R_\eta(37)$  (à gauche) et largeur latérale  $\omega_{\eta^2}$  (à droite). Les distributions sont montrées pour des électrons (histogrammes en trait plein) et pour des pions (histogrammes hachurés).*

### *Premier compartiment du calorimètre électromagnétique*

Les variables utilisées précédemment permettent d'exclure les pions de hautes énergies et formant de larges gerbes. D'autres critères, utilisant le premier compartiment du ECAL, sont employés afin de rejeter les pions de plus basses énergies. Parmi elles on utilise la fraction d'énergie F1 reconstruite dans le premier compartiment du calorimètre électromagnétique. Sa distribution est montrée sur la figure 7 pour des électrons et des pions.

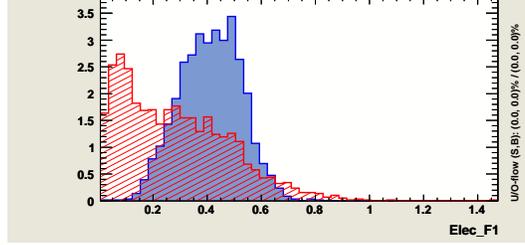


FIG. 7 – Fraction d'énergie  $F1$  reconstruite dans le premier compartiment du calorimètre électromagnétique. Les distributions sont montrées pour des électrons (histogramme en trait plein) et des pions (histogramme hachuré).

Deux autres variables discriminantes caractérisant les gerbes sont également employées (cf Fig. 8) :

- $\omega_{3strips}$  donne la largeur de la gerbe dans 3 cellules autour du maximum.
- La largeur totale de la gerbe  $\omega_{tot1}$  est calculée dans une fenêtre  $\Delta\eta \times \Delta\phi = 0,0625 \times 0,2$  correspondant environ à 40 cellules.

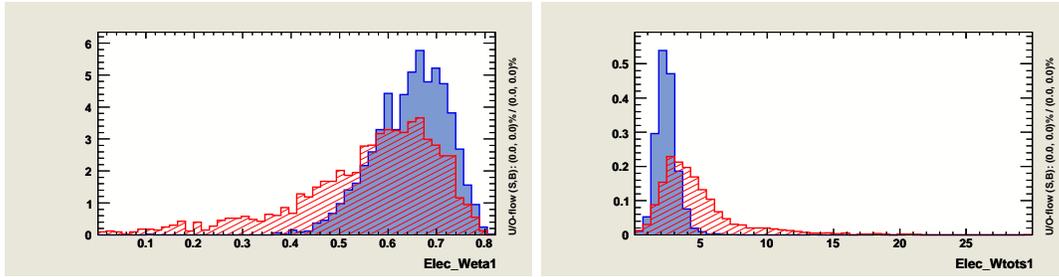


FIG. 8 – Largeur totale de la gerbe  $\omega_{tot1}$  (à gauche) et dans trois cellules  $\omega_{3strips}$  (à droite) dans le premier compartiment du calorimètre électromagnétique. Les distributions sont montrées pour des électrons (histogrammes en trait plein) et des pions (histogrammes hachurés).

### Détecteur interne

Afin de rejeter un plus grand nombre de pions, on peut utiliser les informations combinées du calorimètre électromagnétique et du détecteur interne (cf Fig. 9) :

- $E/p$  est défini comme le rapport entre l'énergie déposée dans le calorimètre électromagnétique et l'impulsion reconstruite dans le détecteur interne.
- $|\Delta\eta|$  donne la différence de position en  $\eta$  calculée sur  $\pm$  sept cellules autour de la trace extrapolée dans le calorimètre électromagnétique.

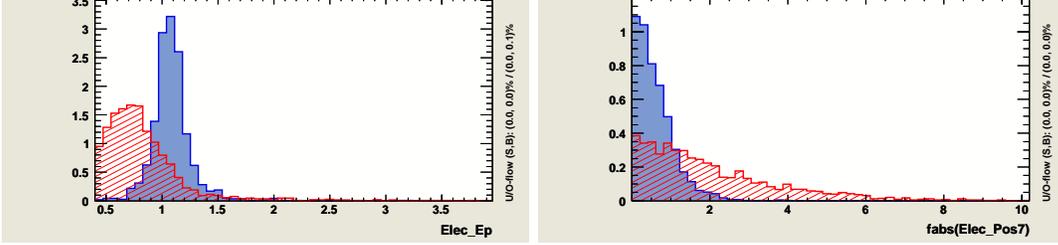


FIG. 9 – Rapport  $E/p$  entre l'énergie reconstruite dans le calorimètre électromagnétique et l'impulsion reconstruite dans le trajectographe (à gauche) et différence angulaire en pseudorapidité  $|\Delta\eta|$  entre la position de la trace extrapolée au calorimètre électromagnétique et celle de l'amas électromagnétique (à droite). Les distributions sont montrées pour des électrons (histogrammes en trait plein) et des pions (histogrammes hachurés).

Enfin une dernière coupure est appliquée sur une des variables provenant du TRT. Il s'agit du rapport  $N_{\text{TRT}}/N_{\text{straw}}$  entre le nombre de points de mesure de haut seuil  $N_{\text{TRT}}$  et le nombre total de points de mesure dans le TRT  $N_{\text{straw}}$  (cf paragraphe 2.2.1). Sa distribution est montrée sur la figure 10.

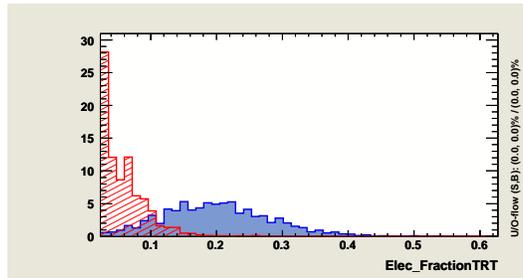


FIG. 10 – Rapport  $N_{\text{HTR}}/N_{\text{straw}}$  entre le nombre de points de mesure de haut seuil dans et le nombre total de points de mesure dans le TRT. Les distributions sont montrées pour des électrons (histogramme en trait plein) et des pions (histogramme hachuré).

L'ensemble des coupures qui ont été appliquées sont présentées dans le tableau 2.

hadleak	$R_\eta(37)$	$\omega_{\eta^2}$	F1	$\omega_{3\text{strips}}$	$\omega_{\text{tot1}}$	$E/p$	$ \Delta\eta $	$N_{\text{HTR}}/N_{\text{straw}}$
0,1	0,65	0,02	0,15	0,45	6	0,7	3	0,07

TAB. 2 – Valeurs des coupures “personnelles” appliquées sur les variables discriminantes dans l'algorithme d'identification des électrons.

Parallèlement à cette série de coupures, on applique une seconde série de coupures dite “officielle” (ISem) définie par la collaboration Atlas qui utilise les variables décrites précédemment ainsi que d'autres variables discriminantes que nous ne détaillerons pas dans ce rapport.

Pour estimer les performances de l'algorithme, on utilise l'efficacité d'identification des électrons  $\varepsilon_e$  définie par  $\varepsilon_e = \frac{N_e^t}{N_e}$ , où  $N_e^t$  est le nombre d'électrons qui ont été identifiés après l'application de la série de coupures, et  $N_e$  est le nombre initial d'électrons. A cette efficacité, on associe l'incertitude statistique :  $\sigma(\varepsilon) = \sqrt{\frac{\varepsilon_e(1 - \varepsilon_e)}{N_e}}$ .

On défini également un facteur de rejet des pions  $R_\pi = \frac{N_\pi}{N_\pi^t}$ , où  $N_\pi$  est le nombre initial de pions et  $N_\pi^t$  est le nombre de pions ayant été identifiés comme étant des électrons. L'incertitude statistique associée est donnée par :  $\sigma(R) = \frac{R_\pi}{\sqrt{N_\pi^t}}$ .

Enfin, on défini la pureté  $P$  par  $P = \frac{N_e^t}{N^t}$ , où  $N^t$  correspond au nombre d'électrons "candidats" restant après l'application des coupures. On associe à la pureté l'incertitude statistique suivante :  $\sigma(P) = \sqrt{\frac{P(1 - P)}{N^t}}$ .

Les résultats obtenus sont regroupés dans le tableau 3.

	$\varepsilon_e$ (en %)	$R_\pi$	$P$ (en %)
sélection (1)	$77 \pm 0,13$	$3 \pm 0,03$	$81 \pm 0,13$
sélection (2)	$55 \pm 0,16$	$186 \pm 20$	$90 \pm 0,12$

TAB. 3 – Efficacité d'identification des électrons  $\varepsilon_e$ , facteur de rejet des pions  $R_\pi$  et pureté  $P$  obtenus avec les coupures "personnelles" (sélection (1)) et les coupures "officielles" (sélection (2)) appliquées dans l'algorithme d'identification.

La sélection (2) applique plus de critères que ceux de la sélection (1), près d'une quinzaine. De plus les seuils des critères ont été optimisés par intervalle de  $\eta$  et de  $E_T$  (énergie transverse), soit près de 500 seuils définis! La sélection (1), très simple, nous a permis facilement d'identifier les électrons avec une bonne efficacité. Mais le facteur de rejet des pions est modeste, et la pureté seulement de 80 %. La sélection (2) a été optimisée pour obtenir un compromis entre une bonne efficacité, un facteur de rejet important et une bonne pureté.

Ces résultats seront comparés à ceux obtenus avec des méthodes d'analyse statistique dans le paragraphe 4.2.

### 3.4 Reconstruction du $J/\psi$

On apparie les électrons qui ont été identifiés avec l'algorithme précédent afin de former des paires  $e^+e^-$  susceptibles de provenir de la désintégration d'un  $J/\psi$ . A l'aide des données reconstruites dans les diverses parties du détecteur (cf paragraphe 3.2), on peut obtenir les différentes composantes de l'impulsion ( $p_x$ ,  $p_y$  et  $p_z$ ), l'impulsion transverse  $p_T$  ainsi que l'énergie totale pour chacun des électrons de la paire, qui vont permettre de reconstruire la masse invariante du  $J/\psi$ .

Les distributions de masse invariante des paires d'électrons en utilisant d'une part les informations du détecteur interne et d'autre part les données du calorimètre électromagnétique sont montrées sur la figure 11.

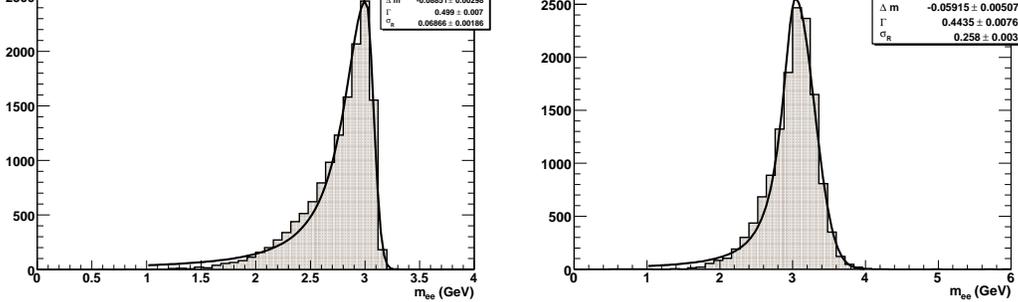


FIG. 11 – *Distribution de masse invariante des paires d'électrons en utilisant uniquement les informations du détecteur interne (à gauche) et avec l'énergie reconstruite dans le calorimètre électromagnétique (à droite). Le signal est ajusté, dans le cas de l'utilisation du détecteur interne, par une distribution de Breit-Wigner à gauche du pic et par une gaussienne à droite. Lors de l'utilisation du calorimètre électromagnétique le signal est ajusté par deux gaussiennes à gauche et à droite du pic*

Dans le cas du détecteur interne, la courbe est ajustée par une distribution de Breit-Wigner à gauche du pic et par une gaussienne à droite du pic. La masse du  $J/\psi$  ainsi reconstruite présente un décalage de  $88 \text{ MeV}/c^2$ . Elle présente à gauche du pic une queue de distribution qui est due aux pertes d'énergie dans le trajectographe. Les particules chargées, sous l'effet du champ magnétique, rayonnent et perdent de l'énergie par rayonnement de freinage : il y a émission de photons. Afin d'obtenir une meilleure reconstruction de la masse invariante du  $J/\psi$  il faudrait prendre en compte l'énergie manquante et donc identifier les photons émis par les particules chargées.

Dans le cas du calorimètre électromagnétique, la courbe est ajustée par deux gaussiennes à gauche et à droite du pic. La masse du  $J/\psi$  reconstruite est décalée d'environ  $59 \text{ MeV}/c^2$ .

## 4 Etudes sur l'amélioration de l'identification des électrons

La méthode d'identification, définie dans le paragraphe 3.3, se contente d'appliquer des coupures sur une série de variables discriminantes, sans prendre en compte la forme des distributions ainsi que les possibles corrélations existant entre ces variables. Le logiciel TMVA [ea] (Toolkit for Multivariate Data Analysis) regroupe un grand nombre de méthodes d'analyse statistique (réseau de neurones, discriminant de Fisher, ...) qui exploitent ces informations dans le but d'améliorer les performances d'identification.

### 4.1 Méthodes utilisées

#### *Optimisation des coupures*

Cette méthode ("cuts") est semblable à celle que l'on a développé précédemment. Elle consiste à faire des coupures sur l'ensemble des variables tout en cherchant à optimiser leurs valeurs afin d'obtenir le meilleur rapport signal sur bruit. L'avantage principal par

et ne dépend pas de la personne qui l'effectue.

### Méthode de maximum de vraisemblance

La plus commune des méthodes d'analyse statistique repose sur un rapport de vraisemblances (likelihood) construit à partir des distributions des variables discriminantes pour le signal et le bruit de fond qui doivent être séparées :

$$X_{\text{RL}} = \frac{\prod_i g_i^s(x_i)}{\prod_i g_i^b(x_i) + \prod_i g_i^s(x_i)}, \quad (1)$$

où  $g_i^s(x_i)$  (resp.  $g_i^b(x_i)$ ) est la distribution de densité de probabilité à une dimension de la variable  $x_i$  pour les événements de signal (resp. de bruit de fond). Ces distributions, appelées fonctions de densité de probabilité (PDF - Probability Density Function) sont construites à partir d'une fraction (environ 20%) des données, qui ne sont alors plus utilisées pour estimer les performances. Cette méthode ne prend pas en compte les corrélations entre les variables.

### Le discriminant de Fisher

Dans la méthode développée par Fisher dans les années 30, les  $N$  variables d'entrée ne sont plus traitées séparément mais combinées linéairement. L'objectif est de déterminer un axe dans l'espace  $R^N$  des variables discriminantes, de telle sorte qu'une fois projetés sur cet axe les événements correspondant à chacune des deux hypothèses (ici électron ou pion) soient les plus éloignés possibles. Cette méthode est d'autant plus performante que les distributions des variables discriminantes pour le signal et le bruit de fond sont des gaussiennes et qu'il existe des corrélations entre ces mêmes variables.

### Le réseau de neurones

Contrairement au discriminant de Fisher, le réseau de neurones (MLP) est une méthode non-linéaire, qui présente une structure en couches (cf Fig. 12). La couche d'entrée contient autant de neurones qu'il y a de variables discriminantes, alors que la couche de sortie n'est composée que d'un seul neurone qui traduit l'appartenance à l'une des deux hypothèses (signal ou bruit de fond). Entre ces deux couches, dont les compositions sont déterminées par les données du problème à étudier, on trouve  $k$  couches internes constituées d'un nombre aléatoire de neurones.

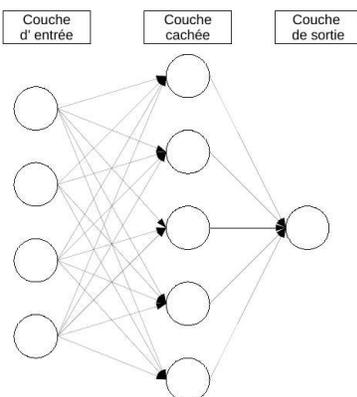


FIG. 12 – Architecture d'un réseau de neurones à une couche cachée

La méthode BDT se caractérise par une succession de noeuds définissant une architecture en forme d'arbre (cf Fig. 13). Au niveau de chaque noeud, une coupure sur une des variable discriminante est appliquée permettant ainsi de séparer les évènements considérés comme du signal (dirigés vers la droite du noeud) de ceux considérés comme du bruit de fond (dirigés vers la gauche du noeud). Ce processus de séparation s'arrête lorsqu'un noeud atteint soit un nombre minimum d'évènements soit un minimum ou un maximum de pureté pour le signal.

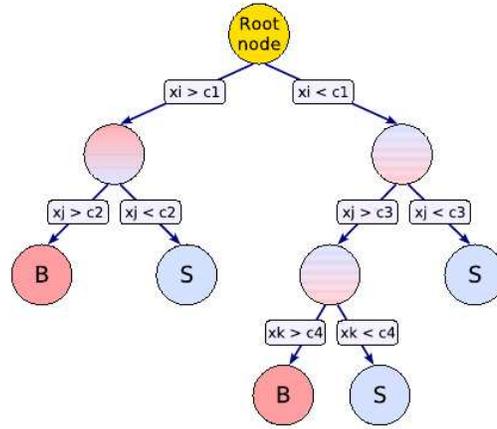


FIG. 13 – Schéma d'un BDT

Comme pour la méthode de maximum de vraisemblance, on peut définir pour les méthodes du discriminant de Fisher, du réseau de neurones et du BDT des fonctions discriminantes  $X_{FI}$ ,  $X_{MLP}$  et  $X_{BDT}$ .

## 4.2 Performances des différentes méthodes

Afin de pouvoir comparer les performances de ces différentes méthodes avec celles obtenues avec l'algorithme d'identification des électrons, les mêmes coupures de présélection des traces ont été appliquées.

A l'exception de la méthode d'optimisation des coupures, les quatre autres méthodes d'analyse statistique utilisent les fonctions discriminantes définies précédemment afin de séparer au maximum les évènements appartenant au signal de ceux appartenant au bruit de fond. La figure 14 montre les distributions des fonctions discriminantes  $X_{RL}$ ,  $X_{FI}$ ,  $X_{BDT}$  et  $X_{MLP}$  pour les électrons et les pions.

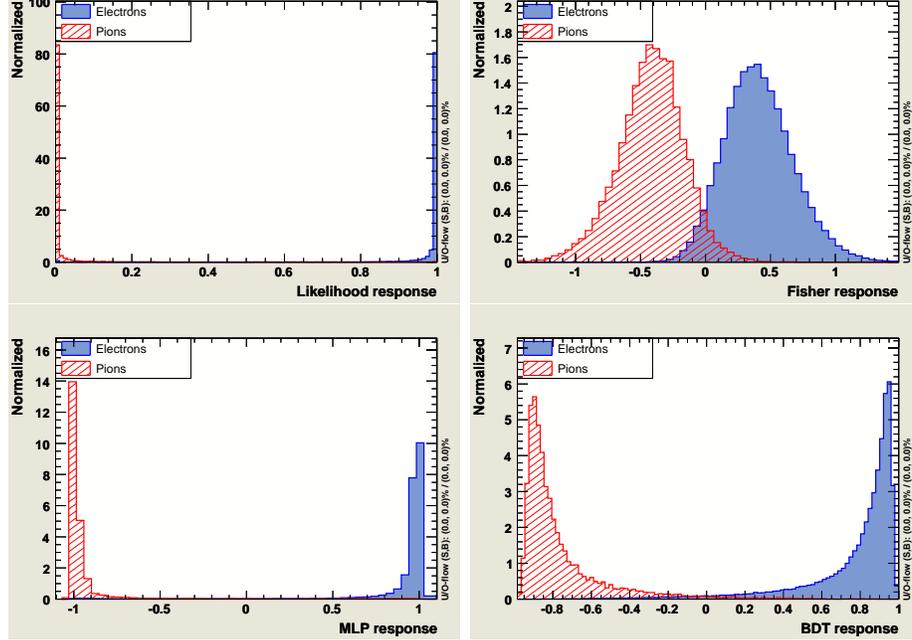


FIG. 14 – Fonctions discriminantes  $X_{\text{RL}}$  (en haut à gauche),  $X_{\text{FI}}$  (en haut à droite),  $X_{\text{MLP}}$  (en bas à gauche) et  $X_{\text{BDT}}$  (en bas à droite) pour des électrons (histogrammes en trait plein) et pour des pions (histogrammes hachurés).

Pour analyser les performances de ces différentes méthodes d’analyse on utilise l’efficacité d’identification des électrons et le facteur de rejet des pions qui ont été définis dans le paragraphe 3.3.2. En faisant varier la coupure appliquée sur la fonction discriminante  $X$  on peut déterminer ces facteurs de rejet pour différentes valeurs de l’efficacité. On peut alors tracer le facteur de rejet  $R_{\pi}$  en fonction de l’efficacité d’identification  $\varepsilon_e$  comme montré sur la figure 15.

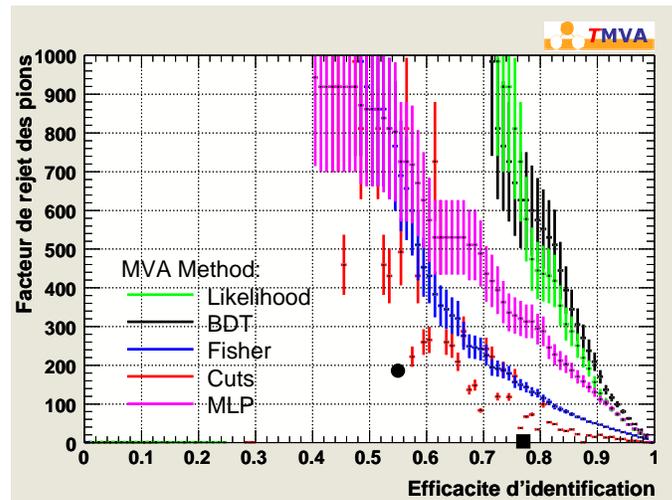


FIG. 15 – Facteur de rejet des pions  $R_{\pi}$  en fonction de l’efficacité d’identification des électrons  $\varepsilon_e$ . Les courbes sont montrées pour les différentes méthodes utilisées. Le point et le carré correspondent respectivement aux valeurs obtenues en appliquant la série de coupures “officielle” et la série de coupures “personnelle” dans l’algorithme d’identification.

l’algorithme d’identification. Le point est associé aux coupures “officielles” et le carré correspond quant à lui aux coupures “personnelles”.

Les performances de notre algorithme d’identification sont plus faibles que celles des méthodes statistiques, ce qui était prévisible étant donné qu’on a traité individuellement les variables discriminantes.

Comparé à la méthode d’optimisation des coupures, les écarts sont moins marqués mais persistent. Cette méthode semble plus performante que notre algorithme, cependant les résultats obtenus sont discutables en particulier du point de vue des coupures appliquées (cf Tab. 4).

$\varepsilon_e$ (en %)	$R_\pi$	hadleak	$R_\eta(37)$	$\omega_{\eta^2}$	F1	$\omega_{3strips}$	$\omega_{tot1}$	$E/p$	$ \Delta\eta $	$N_{\text{HTR}}/N_{\text{straw}}$
$77 \pm 0,13$	$53 \pm 3$	0,49	0,79	0,03	0,12	0,012	10,8	0,52	7,96	0,11
$78 \pm 0,13$	$71 \pm 5$	0,65	0,04	0,02	0,04	0,42	22	0,87	2,9	0,11

TAB. 4 – Valeurs des coupures pour des efficacités de 77% et 78%.

Pour deux efficacités proches (77% et 78%), les valeurs des coupures présentent d’importantes différences notamment dans le cas des variables  $R_\eta(37)$ ,  $\omega_{tot1}$  et  $|\Delta\eta|$  et constituent un des problèmes d’utilisation du logiciel TMVA. La durée de ce stage (6 semaines) n’a pas été suffisante pour résoudre ce problème.

Etant donné les statistiques que l’on a utilisé pour réaliser cette étude et les incertitudes associées aux différentes méthodes, les seuls résultats significatifs sont ceux compris dans l’intervalle d’efficacité [0,7-1] comme montré sur la figure 16.

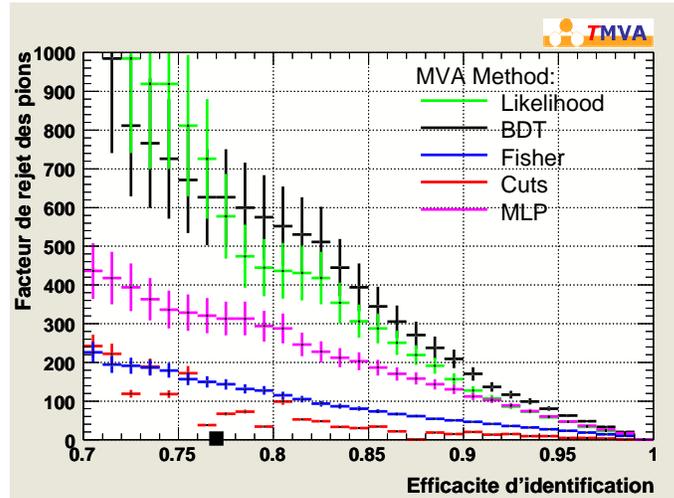


FIG. 16 – Facteur de rejet des pions  $R_\pi$  en fonction de l’efficacité d’identification des électrons  $\varepsilon_e$  dans l’intervalle [0.7-1].

Pour une efficacité de 77 % le facteur de rejet des pions obtenu est de l’ordre de 53 avec la méthode d’optimisation des coupures, de  $580 \pm 120$  avec la méthode de maximum de vraisemblance, de 150 pour le discriminant de Fisher, de 450 pour le réseau de neurones, de 630 pour la méthode BDT alors que l’algorithme d’identification des électrons ne permet d’obtenir qu’un facteur de rejet environ égal à 3.

semblance, du fait de sa relative simplicité d'utilisation, pourra être exploitée dans de prochaines analyses afin d'améliorer de manière significative les performances d'identification des électrons et de rejet des pions.

## 5 Conclusion

Le développement d'un algorithme d'identification des électrons de basses impulsions transverses combinant les données issues du détecteur interne (impulsion, nombre de points de mesure dans le TRT) et du calorimètre électromagnétique (caractéristiques des gerbes dans les différents compartiments) est essentiel dans le processus de reconstruction du  $J/\psi$ . Des coupures ont été appliquées sur un ensemble de neuf variables discriminantes afin d'identifier les électrons provenant de la désintégration du  $J/\psi$  et de rejeter un maximum de pions constituant le bruit de fond. A l'issue de cette sélection on a pu reconstruire la masse du  $J/\psi$  dans le détecteur interne et dans le calorimètre électromagnétique. Les performances obtenues avec cet algorithme sont plutôt modestes en particulier en ce qui concerne le facteur de rejet des pions. Une analyse plus précise par intervalle de pseudo-rapacité  $\eta$  et d'énergie transverse  $E_T$  peut être réalisée pour améliorer l'identification des électrons.

Différentes méthodes d'analyse statistique (maximum de vraisemblance, discriminant de Fisher, ...) prenant en compte la totalité des informations disponibles pour chaque variable (forme des distributions, corrélations) ont par la suite été utilisées dans le but d'améliorer le rapport signal sur bruit. Les performances obtenues sont certes plus élevées que celles provenant de l'algorithme d'identification mais un certain nombre de difficultés persistent.

## Références

- [ea] A. Hocker et al. Tmva - toolkit for multivariate data analysis. <http://arxiv.org/abs/physics/0703039>.
- [F.D08] F.Derue. Reconstruction et identification des électrons dans l'expérience atlas. participation à la mise en place d'un tier 2 de la grille de calcul., 2008. [http://tel.archives-ouvertes.fr/docs/00/26/57/29/PS/HDR\\_Derue.ps](http://tel.archives-ouvertes.fr/docs/00/26/57/29/PS/HDR_Derue.ps).