## Rapport de Stage

# Séparation de gerbes hadroniques proches dans le Calorimètre Hadronique Semi-Digital SDHCAL pour ILC

Mars - Juillet 2013

Rémi Été

Master II Physique Fondamentale et Astrophysique

Parcours Physique Subatomique

Institut de Physique Nucléaire de Lyon Université Claude Bernard Lyon 1

Tuteur de stage: Imad Laktineh







## Remerciements

Je tiens à remercier Imad Laktineh, mon maître de stage, pour le sujet très intéressant qu'il m'a proposé et pour m'avoir pris en stage de Master II. Je le remercie également d'avance pour le trois ans de thèse qui suivront ce stage de recherche.

Je remercie bien évidement Arnaud Steen, thésard sous la direction d'Imad, pour avoir partager des idées de physique et du code au quotidien ainsi que son expérience dans la collaboration ILC et CALICE.

Un grand merci à Stéphane Poss et Giampiero Mancinelli pour m'avoir pousser en avant dans mes études et fait entrer dans le monde la physique des particules et au CERN.

Finalement, merci à Alexey Petrukhin et Gerald Grenier (et non pas Gérard Grenier!) pour l'aide qu'ils m'ont apporté durant mon stage.

# Table des matières

1	Cor	ntexte expérimental et reconstruction dans l'ILD	5
	1.1	ILC : le collisionneur linéaire international	5
	1.2	Le détecteur ILD : International Large Detector	6
	1.3	Le Particle Flow Algorithm	7
	1.4	Environnement de travail et logiciels	8
2	Le	calorimètre hadronique semi-digital SDHCAL	10
	2.1	Description du calorimètre	10
	2.2	Les chambres résistives GRPC	11
	2.3	L'électronique d'acquisition	11
	2.4	Résolution en énergie	12
3	$\mathbf{Pro}$	océdure de superposition des gerbes hadroniques	14
	3.1	Procédure	14
	3.2	Hits superposés : quel seuil?	16
4	Le	s algorithmes pour la séparation	17
	4.1	Étiquetage des hits isolés	17
	4.2	Reconstruction de traces : la transformée de Hough	18
	4.3	Analyse en composante principale	19
	4.4	L'algorithme de séparation dans sa totalité	19
	4.5	Les autres algorithmes développés	20
5	An	alyse des résultats	21
	5.1	Deux gerbes de 10 GeV	21
	5.2	Gerbes de 10 GeV et 30 GeV	23
	5.3	Conclusion de l'analyse et perspectives	25
C	onclu	ısion	26
Bi	iblios	graphie	28

## Introduction

Ce document retranscrit mon travail de stage de Master II recherche réalisé pendant 4 mois au sein de l'IPNL à Lyon. Ce travail vise à identifier les différentes contributions des gerbes hadroniques proches dans un calorimètre hautement granulaire (1 cm<sup>2</sup>) et trouver une façon efficace de séparer ces contributions pour deux gerbes hadroniques proches.

Un premier chapitre **situera le cadre** de ce stage vis à vis du groupe ILC et de la reconstruction d'événements qui sera appliquée dans ces détecteurs.

Un chapitre sera dédié au calorimètre hadronique semi-digital développé à Lyon.

Le chapitre suivant donnera les détails de la **procédure de superposition** des gerbes hadroniques, les données de simulation et des tests en faisceau ne contenant que des événements à une seule particule.

Enfin, après avoir dressé la liste des **algorithmes de reconstruction** développés au cours de ce stage, une estimation de la séparation de deux gerbes hadroniques proches sera présentée ainsi que **l'analyse de ces résultats**.

Une **conclusion** résumera ce travail et proposera des nouvelles idées d'algorithmes à appliquer afin d'**améliorer** ces résultats.

## Chapitre 1

# Contexte expérimental et reconstruction dans l'ILD

Le présent chapitre expose le contexte expérimental de ce stage. Une première section sera dédié au projet de **collisionneur ILC**. Une autre section détaillera un des deux détecteurs placés au point de collision de l'ILC : le détecteur ILD. L'application des techniques dites de PFA (Particle Flow Algorithm) dans l'ILD sera exposé dans la section suivante. Enfin, quelques lignes seront dédiés à mon environnement de travail et aux logiciel utilisés quotidiennement lors de mon stage.

#### Contenu

1.1	ILC: le collisionneur linéaire international	5
1.2	Le détecteur ILD : International Large Detector	6
1.3	Le Particle Flow Algorithm	7
1.4	Environnement de travail et logiciels	8

#### 1.1 ILC: le collisionneur linéaire international

L'ILC est un projet d'accélérateur linéaire e<sup>+</sup> e<sup>-</sup> proposé pour succéder au Large Hadron Collider (LHC) actuellement en fonctionnement au CERN à Genève. Suite à la publication du "*Technical Design Report*" en 2012, le début de sa construction serait estimé pour 2016 - 2018.

Il est conçu pour atteindre une énergie de collision de  $500~{\rm GeV}$  soit  $28~{\rm fois}$  moins d'énergie au LHC après le "Long~Shutdown" (LS) prévu en 2015, période d'arrêt du collisionneur durant laquel des parties du détecteur améliorées seront installées (upgrade) et la luminosité augmentée.

Cependant, étant donnée la nature leptonique des faisceaux à l'ILC, l'énergie de la collision sera connue avec une bonne précision ( $\leq 1\%$ ) et les événements seront nettement plus facile à analyser. Dans un tel environnement, il est possible de suivre les particules chargées avec une grande précision et de reconstruire les énergies dans les calorimètres avec une grande résolution.

L'ILC ne permettra pas d'atteindre les énergies du LHC ( 7 TeV - 14 TeV ) mais sera un outil de précision. Autrement dit les résolutions qui y seront atteintes sont sans précédent. Le

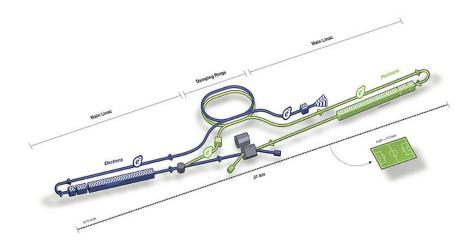


Figure 1.1 – Schéma du collisionneur linéaire prévu dans le TDR

pic de luminosité attendu sera de  $2.10^{34}$  cm<sup>-2</sup> s<sup>-2</sup>. En doublant la longueur des accélérateurs linéaires principaux, il sera possible d'étendre le domaine d'énergie a 1 TeV, le domaine d'énergie de départ se situant autour de 250 GeV. Son but principal sera d'étudier le boson de Higgs potentiellement découvert au LHC en mesurant sa masse et ses propriétés avec plus de précision. Sa masse étant estimée autour de 125 GeV par les expérience LHC, l'ILC permettra de produire le boson de Higgs et de l'étudier. Les autres études seront d'une part des mesures de précision pour la QCD ( constante de couplage  $\alpha_s$  ) et la masse du quark top et d'autre part la recherche de nouvelle physique ( matière noire et particules supersymétrique légères ).

Au point d'interaction, deux différents détecteurs, l'ILD (International Large Detector) et SiD (Silicon Detector) seront installés sur plate-forme mobile afin de pouvoir les inter-changer.

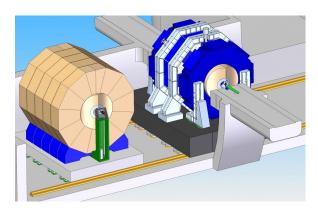


Figure 1.2 – Les détecteurs **ILD** et **SiD** 

Au niveau de la calorimétrie, les deux détecteurs serait équipés de calorimètres électromagnétiques et hadroniques différents. Le calorimètre hadronique SD-HCAL développé ici à Lyon serait installé dans le détecteur ILD. Dans le détecteur SiD serait installé un calorimètre digital développé par le LAPP à Paris. Ici le seul détecteur qui attirera notre attention sera le détecteur ILD. C'est l'objet de la prochaine section.

#### 1.2 Le détecteur ILD : International Large Detector

L'ILD est un détecteur générique  $4\pi$  avec une structure standard en physique des particules dite en couche d'oignons. La figure 1.3 montre une vue en coupe du détecteur ILD.

Au plus proche de la ligne de faisceau se trouve un détecteur à pixel pour tracer les particules chargées finement. Il permet de reconstruire les vertex de désintégrations des particules à courte durée de vie comme les mésons  $B^0$ . La couche suivante du détecteur, le trajectomètre, reconstruit lui aussi les trajectoire des particules chargées mais sur une grande distance. Un aimant supra-conducteur situé après les calorimètres permet de courber les trajectoire des

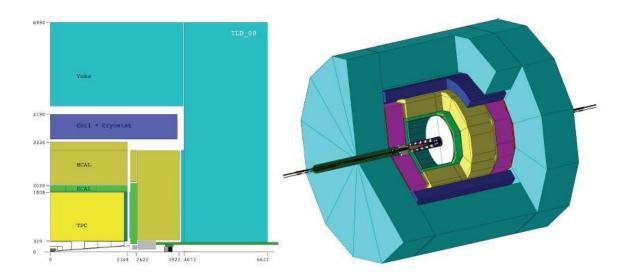


Figure 1.3 – Vue en coupe du détecteur ILD

particules chargées. La courbure des particules chargées est directement reliée à l'impulsion via la relation :

$$p = qBr (1.1)$$

Ainsi, le trajectomètre apporte l'information de l'impulsion. Après le trajectomètre se trouve un calorimètre électromagnétique (ECAL) suivi d'un calorimètre hadronique (HCAL). Le premier permet d'arrêter et de mesurer l'énergie des particule de type électromagnétique, c'est à dire les photons, les électrons et les positrons; le deuxième, les hadrons composés de quarks. Les deux sont des calorimètres à échantillonnage c'est à dire une alternance de couches d'absorbeur, qui forcent l'interaction des particules et de couches de milieu actif qui détectent une partie des particules produites. Les calorimètres permettent de mesurer l'énergie des particules. Après les calorimètres se trouve l'aimant supra-conducteur qui permet de courber la trajectoire des particules chargées. La dernière couche est constitué de chambres à muons afin de reconstruire les traces de ces derniers.

En combinant la mesure de l'impulsion dans le trajectomètre et la mesure de l'énergie dans les calorimètres pour une particule, on peut alors reconstruire son quadri-vecteur impulsion  $p^{\mu}$  pour remonter à la cinématique de l'événement.

#### 1.3 Le Particle Flow Algorithm

Afin d'atteindre des précisions jamais égalées, les détecteurs de l'ILC ont été optimisés pour appliquer un algorithme de reconstruction nommé PFA (Particle Flow Algorithm). Ce dernier utilise différentes informations dans les sous détecteurs afin de suivre les particules lors de la reconstruction. Les sous détecteurs seront donc optimisés afin de d'atteindre ce but et en particulier, les calorimètres électromagnétiques et hadroniques devront être finement segmentés afin de pouvoir distinguer les différentes contributions des particules venant s'y déposer et le trajectomètre devra avoir le plus rayon externe possible afin de mieux reconstruire les traces des particules chargées et de mieux identifier les points d'entrées de ces traces dans le calorimètre électromagnétique.

Les premiers codes de reconstruction utilisant le PFA pour ILC étaient initialement un unique fichier contenant la géométrie et les paramètres des algorithmes. Après une complète réécriture il y a 3 ans, le code actuel consiste en une librairie d'algorithmes modulaire

nommé PandoraPFA, indépendant de tout cadre d'application et peux être appliqué à n'importe quel détecteur de physique des particules si la technologie des sous détecteurs le permet (e.g grande granularité dans les calorimètres et espace faible entre les sous détecteurs).

Cependant, le code actuel de reconstruction ne possède aucune option concernant la technologie du calorimètre hadronique semi-digital (cf chapitre 2). De ce fait, les algorithmes développés dans PandoraPFA utilisent des options par défaut si un calorimètre hadronique semi-digital est utilisé. Un développement d'algorithmes optimisés pour un calorimètre semi-digital devient alors une nécessité afin d'améliorer la résolution en énergie des particules seules et ainsi la séparabilité des gerbes hadroniques proches si un calorimètre semi-digital est utilisé pour l'ILD.

La figure 1.4 ci-contre montre les produits d'une collision  $e^+e^-$  dans le détecteur ILD. Le premier volume circulaire sur la gauche correspond au trajectomètre. On y voit les traces des particules chargées reconstruites. La zone centrale où les amas sont le plus présents est le calorimètre électromagnétique (ECAL). C'est ici que les particules de type électromagnétique, neutres ou chargées, viennent ensuite se déposer. Les particules de type hadronique, neutres ou chargées, se déposent ensuite dans le calorimètre hadronique (HCAL). L'épaisseur du calorimètre électromagnétique représentant environ une longueur de radiation  $\lambda_I$ , il est courant que certains hadrons commencent leurs interaction dans le ECAL.

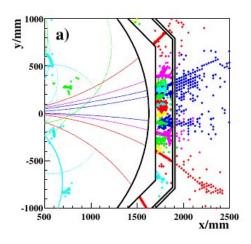


FIGURE 1.4 – Événement suite à une collision  $e^+e^-$  dans le détecteur ILD

Le Particle Flow Algorithm commence par identifier les différents amas dans les calorimètres pointant au bout des traces reconstruites dans le trajectomètre. Un algorithme agglomère ensuite des hits contenu dans cone d'ouverture paramétrable. Cette procédure reconstruit les particules chargées, électromagnétiques ou hadroniques. Les amas des particules chargées reconstruites dans les calorimètres sont ensuite retirés afin de reconstruire les particules neutres.

Les particules reconstruites pouvant avoir des énergies très différentes que celles suggérées par les impulsions reconstruites dans le trajectomètre. Une stratégie dite de re-clustering est ensuite adoptée afin de reconsidérer les différents amas précédemment reconstruits pour qu'ils s'ajustent mieux avec les impulsions.

A ce stade de la reconstruction, des objets appelés *Particle Flow Objects* (PFOs) sont construits. Ils contiennent toutes les informations d'une particule : 4-vecteur impulsion, PID, charge, etc ...

### 1.4 Environnement de travail et logiciels

Afin de traiter les données de simulation GEANT4 et de tests en faisceau, des formats de fichier ont été choisis. Les interfaces de lecture, écriture et traitement de fichiers sont contenus dans un regroupement de logiciels nommé ILCSOFT [4] (ici v01-16-01). Parmi tous ces logiciels, ceux utilisés pour cette étude sont :

- LCIO [8] (Linear Collider Input Output) [v02-03-03]: Système de fichier pour la lecture et écriture de fichier lcio (\*.slcio). Un fichier comprends une collection d'évènements contenant différentes collections correspondant bien souvent au différentes informations stockées dans les différents sous détecteurs. A titre d'exemple, l'unique type de collection utilisé ici est la collection "CalorimeterHit".
- Marlin [6] (Modular Analysis & Reconstruction for the LINear collider) [v01-04]: Interface de traitement d'événements LCIO. A chaque traitement de série d'événements correspond un *processeur* exécutant une série de fonctions sur un événement donné.
- ROOT [9] [v5.34.01-64]: Environnement d'analyse optimisé pour traiter des grandes quantités de données, connu pour le traitement des données au LHC. ROOT fournit également un module de géométrie utilisé ici pour visualiser les données du calorimètre.

Tout le code développé lors de ce stage a été écrit en C++ et été mis sous forme de librairies formant à elle même un environnement de travail pour le prototype m³ de Lyon. Un effort important été mis dans le développement d'un code propre et réutilisable pour tout traitement des données concernant le prototype et ses futures études comme l'ajout d'un calorimètre électromagnétique devant le prototype. Une documentation du code de la version de développement est disponible dans [5].

## Chapitre 2

# Le calorimètre hadronique semi-digital SDHCAL

Après une description du calorimètre hadronique construit à Lyon, le détail de fonctionnement d'un élément de zone active de détection (GRPC) sera présenté. Une section sera consacrée à l'électronique d'acquisition et permettra de caractériser la réponse du calorimètre pour ainsi extraire sa résolution en énergie.

#### Contenu

2.1	Description du calorimètre	10
2.2	Les chambres résistives GRPC	11
<b>2.3</b>	L'électronique d'acquisition	11
2.4	Résolution en énergie	12

#### 2.1 Description du calorimètre

Un prototype de calorimètre hadronique à haute granularité a été développé par l'équipe lyonnaise de ILC. C'est un calorimètre à échantillonnage constitué de 48 chambres à plaques de verres résistif (GRPC) espacées d'absorbeurs d'acier ( $\lambda_I=16.76~cm$  et  $X_0=1.77~cm$ ) de 15 mm d'épaisseur. Les chambres GRPC sont insérées dans des cassettes en acier de 5 mm d'épaisseur formant ainsi une barrière totale d'acier de 20 mm entre chaque zone active.

Les 48 couches d'acier représentent au total environ 6 longueurs d'interaction soit une épaisseur théoriquement suffisante pour arrêter la plupart des gerbes hadroniques qui atteindront le HCAL dans l'ILD.



FIGURE 2.1 – Le SDHCAL lors d'un test en faisceau au SPS (CERN)

#### 2.2 Les chambres résistives GRPC

Les chambres à plaque de verre résistif (GRPC) sont des détecteurs gazeux (cf. figure 2.2). Lorsqu'une particule chargée traverse la couche de gaz de la chambre, celle ci va ioniser le gaz pour produire des électrons et des ions. En appliquant une tension de part et d'autre de la couche de gaz, les électrons vont dériver vers l'anode et la charge va être récolter sur les carreaux de cuivre.

Si le champs électrique fourni est suffisamment fort dans la GRPC, comme dans notre cas, les électrons qui dérivent vers l'anode produisent à leur tour des d'autres électrons, créant ainsi un phénomène d'avalanche électronique. C'est le principe de base d'un détecteur gazeux.

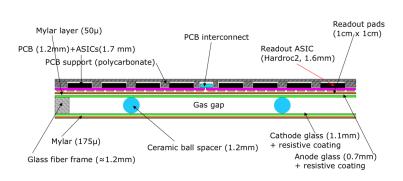


Figure 2.2 – Chambre GRPC utilisée comme zone active de détection dans le SDHCAL

segmentation du calorimètre est directement relié à la dimension des cellules en cuivre utilisées pour récolter la charge dérivée dans la couche de gaz. Celle ci est fixée à 1 cm<sup>2</sup> et fait partie intégrante de l'électronique d'acquisition (cf section 2.3). Ainsi la résolution spatiale transverse du calorimètre lorsque une seule cellule est touché est donnée par le calcul (2.1).

$$\sigma_t^2 = \frac{\int_{-\frac{\delta l}{2}}^{\frac{\delta l}{2}} l^2 dl}{\delta l} = \frac{\delta l^2}{12} \qquad \Longrightarrow \qquad \sigma_t = \frac{\delta l}{\sqrt{12}} = \frac{1}{\sqrt{12}} \simeq 0.288cm \tag{2.1}$$

Différentes informations sur les GRPC telles que le gaz utilisé ou la haute tension appliquée sont référencés dans la table 5.2 en annexe de ce document.

#### 2.3 L'électronique d'acquisition

Les cellules en cuivre qui par induction détectent la charge produite par une avalanche, sont directement reliés à l'électronique de lecture du calorimètre. Des puces de lecture couramment nommées ASIC (Application Specific Integrated Circuit) servent à faire suivre le signal aux DIFs (Digital InterFace board) qui elles même conduisent le signal vers les ordinateurs.

Des puces de lecture ont été développé spécialement pour le calorimètre hadronique semidigital : les HARDROC (HAdronic RPC Detector Read Out Chip).

Les différents seuils de charges actuels appliqués ainsi que le codage binaire est résumé dans la table 2.1.

Seuil		1	2	3
Intervalle de charge	$c < 0.114 \ pC$	$0.114 < { m c} < 5 { m \ pC}$	$5 < \mathrm{c} < 15~\mathrm{pC}$	m c > 15~pC
Codage binaire	00	01	10	11

Table 2.1 – Intervalle de charge et codage des différents seuils

La première colonne de cette table correspond au cas où la charge n'est pas suffisante pour activer une cellule. Le *hit* n'est donc pas enregistré. Plus le nombre de particules traversant une cellule est grand plus le seuil a de chance d'être grand aussi. Les seuils permettent dans ce cas de réduire le phénomène de saturation pouvant avoir lieu à haute densité de particules.

A cause du grand nombre de canaux électroniques (cf. table 5.3 en annexe), une grande consommation d'électricité est attendue lors du fonctionnement du calorimètre ce qui entraîne une forte dissipation thermique. Un système d'alimentation pulsée à été mis au point afin d'éviter ce phénomène de surchauffe de l'électronique. Pour le SDHCAL, le composant utilisant ce système est la puce HARDROC qui enregistre l'information des seuils. L'idée est d'activer l'électronique de lecture uniquement au moment où le faisceau utilisé pour produire des gerbes hadronique est actif. Pour l'ILC ceci représentera 0.5% du temps.

#### 2.4 Résolution en énergie

Le seuil apporte une information sur la concentration en particules qui touchent une cellule et permet de limiter les effets de saturation lorsque cette concentration est trop importante. Il apporte d'autre part une amélioration à la résolution en énergie en assignant un poids différent pour chaque seuil lors du calcul de l'énergie.

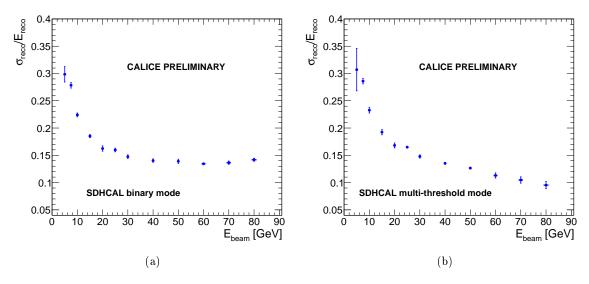


FIGURE 2.3 – Résolution en énergie pour les modes binaire (a) et multi-seuil (b) du calorimètre hadronique

La figure 2.3 montre la résolution en fonction de l'énergie suivant les deux modes de fonctionnement du calorimètre [2] : multi-seuils ou binaire suivant si l'information du seuil est utilisée ou non. L'énergie reconstruite pour les deux modes est donnée respectivement par les formules (2.2) et (2.3) pour le mode binaire et multi-seuils :

$$E_{reco} = (a + b \cdot N_{hit}) \cdot N_{hit} \tag{2.2}$$

$$E_{reco} = \alpha \cdot N_1 + \beta \cdot N_2 + \gamma \cdot N_3 \tag{2.3}$$

où  $N_{hit}$  est le nombre de hits dans la gerbe hadronique,  $N_1$ ,  $N_2$ ,  $N_3$  respectivement le nombre de hits de seuil 1, 2 et 3 dans la gerbe et a et b des constantes. Dû à la complexité de la structure des gerbes hadroniques, les paramètres  $\alpha$ ,  $\beta$  et  $\gamma$  sont fonctions du nombre total de hits  $N_{hit}$ :

$$\alpha = p_1 + p_2 \cdot N_{hit} + p_3 \cdot N_{hit}^2 \tag{2.4}$$

$$\beta = p_4 + p_5 \cdot N_{hit} + p_6 \cdot N_{hit}^2$$

$$\gamma = p_7 + p_8 \cdot N_{hit} + p_9 \cdot N_{hit}^2$$
(2.5)
(2.6)

$$\gamma = p_7 + p_8 \cdot N_{hit} + p_9 \cdot N_{hit}^2 \tag{2.6}$$

avec  $p_{\{1,2,3,4,5,6,7,8,9\}}$  des constantes. Toutes les constantes utilisées pour les calculs d'énergie sont obtenues par minimisation de  $\chi^2$ . Les valeurs utilisées lors de mes analyses sont résumées dans la table 5.4 en annexe de ce document.

## Chapitre 3

# Procédure de superposition des gerbes hadroniques

Afin d'appliquer des algorithmes de séparation de gerbes hadroniques dans le SDHCAL, des événements contenant deux gerbes hadroniques de pi- ont été construits à partir de deux événements contenant une seule gerbe hadronique de pi-. Un code a été développé afin d'automatiser cette superposition négociant un simple jeu de paramètres. Après une description de la **procédure de superposition**, la question de l'affectation du nouveau seuil des **hits superposés** sera discuté.

#### Contenu

3.1	Procédure	14
3.2	Hits superposés : quel seuil?	16

#### 3.1 Procédure

La figure 3.1 montre le schéma de la procédure de superposition des gerbes hadroniques à partir de deux fichiers lcio [8] contenant des événements à une seule particule. Un nouveau fichier est produit, contenant les deux collections de hits avec un décalage global d'une valeur donnée entre le barycentre des deux gerbes.

Les paramètres d'entrée pour les deux fichiers sont de même nature : le nom de la collection lcio utilisé pour enregistrer les hits du calorimètre et la chaîne de caractère utilisée pour décoder la position du hit par rapport à son codage binaire. Pour les paramètres concernant le fichier de sortie, la distance entre les deux gerbes et le nombre d'événements viennent s'ajouter.

La procédure se déroule alors comme suit.

Pour chaque gerbe, le barycentre est calculé en pondérant par les seuils et translaté dans la direction x et y de sorte qu'il soit au milieu du calorimètre. Il est ensuite translaté encore une fois dans la direction x d'une distance  $\pm$  d/2 avec d la distance entre les deux gerbes. La collection totale est ensuite réécrite dans un nouveau fichier lcio en assignant des étiquettes a, b et c respectivement pour les *hits* de la collection 1, de la collection 2 et appartenant aux deux collections.

Si certains *hits* se retrouvent en dehors du calorimètre lors de ce décalage, alors ils sont supprimés de la collection et ne seront donc pas écrits dans la collection finale. Ce pourcentage de *hits* moyen supprimés lors du décalage a été estimé à moins de 1% pour deux gerbes de 50 GeV à 30 cm de distance (distance de séparation la plus grande lors de mon étude avec deux gerbes de grande étendue latérale). De ce fait, cet effet sera négligé par la suite lors de la séparation et des différents calculs d'énergie.

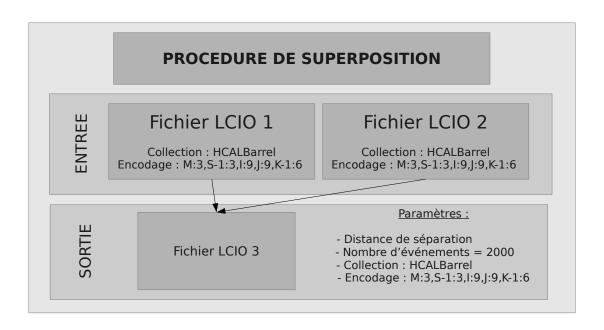


Figure 3.1 – Schéma de la procédure de superposition

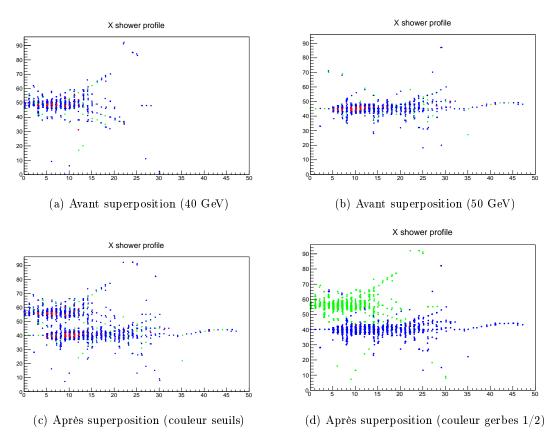


Figure 3.2 – Avant et après la procédure de superposition des gerbes hadroniques à  $15~\mathrm{cm}$  de distance

D'autre part, si deux *hits* viennent à se superposer lors de la translation, la question se pose de savoir quel seuil affecter à ce *hit* commun. C'est l'objet de la section suivante.

#### 3.2 *Hits* superposés : quel seuil?

Suite à une étude plus approfondie [7], les valeurs des seuils de charge ont été fixées aux valeurs présentées dans la table 5.3. Une fois le signal digitalisé, l'information de la valeur de la charge devient inaccessible. Il est donc impossible lorsque deux hits se superposent d'affecter la valeur correcte de charge et par conséquent la bonne valeur du nouveau seuil. Une approximation est alors indispensable lors de la superposition.

La table 3.1 donne la valeur du nouveau seuil en fonction du seuil du hit de la gerbe 1 et du hit de la gerbe 2 se superposant. Cette table peut se résumer en prenant le seuil le plus haut des deux hits. Ce choix de ce nouveau seuil ne résulte que d'une étude intuitive et ne constitue peut etre pas le meilleur choix de valeur du nouveau seuil. Une étude plus approfondie est nécessaire en utilisant les informations avant et après digitalisation afin de savoir l'erreur commise sur le nouveau seuil et ainsi sur le calcul de l'énergie (cf section 2.4 du chaptire 2)

2	seuil 1	seuil 2	seuil 3
seuil 1	1	2	3
seuil 2	2	2	3
seuil 3	3	3	3

Table 3.1 – Nouveau seuil appliqué lors de la superposition en cas de hits superposés

## Chapitre 4

# Les algorithmes pour la séparation

Les algorithmes décrits dans les différentes sections suivantes ont été pour une grande partie développés durant mon stage. D'autres ont été re-codés afin d'être adaptés à mon environnement de travail comme la transformée de Hough. Beaucoup de ces algorithmes sont inspirés de techniques de traitement d'image. Les **différents algorithmes utilisés** pour séparer de gerbes proches seront d'abord décrits dans différentes sections. La section qui suivra expliquera comment ces algorithmes sont agencés pour la séparation de gerbes hadroniques proches. Enfin une partie sera dédiée aux algorithmes non utilisés pour cette séparation mais tout de même développés lors de ce stage. Ces algorithmes seront utiles lors des éventuelles futures améliorations de la séparation.

#### Contenu

4.1	Étiquetage des $hits$ isolés	17
4.2	Reconstruction de traces : la transformée de Hough	18
4.3	Analyse en composante principale	19
4.4	L'algorithme de séparation dans sa totalité	<b>19</b>
4.5	Les autres algorithmes développés	20

### 4.1 Étiquetage des hits isolés

L'algorithme d'isolation est basé sur un algorithme de traitement d'image appelé "blob detection". L'algorithme itère sur les pixels d'une image et regarde différentes informations sur les pixels environnants. Les images utilisés comportent très souvent uniquement des pixels noirs ou blancs. Par exemple, l'algorithme peut regrouper des amas en fonction de s'il trouve un pixel activé (noir) à coté de lui.

Ici le meme principe va s'appliquer. L'algorithme d'isolation itère sur les différents *hits* et compte ses plus proches voisins en pondérant par les seuils. Le tout est ensuite renormalisé par 27, le nombre de total de *hits* voisins y compris le hit central. Une coupure inférieure est ensuite appliquée sur cette "concentration". La démarche se résume par l'équation suivante :

$$\sum_{i=1}^{27} \frac{seuil_i}{27} < c \tag{4.1}$$

Cet algorithme peut être utilisé dans différents buts. Dans cette étude, son but sera de réduire les gerbes hadroniques en retirant ces hits isolés pour ne garder que le bulbe de la

gerbe hadronique. Des études sont aussi en cours pour améliorer la résolution en énergie du calorimètre en pondérant ces *hits* avec un poids différent des autres (cf section 2.4).

#### 4.2 Reconstruction de traces : la transformée de Hough

La transformée de Hough est un algorithme utilisé principalement dans le traitement d'images numériques [10]. L'algorithme a été inventé par Paul Hough en 1962 et permettait initialement de reconnaître des lignes sur une image 2D. Richard Duda et Peter Hart ont ensuite amélioré l'algorithme en 1972 pour ainsi reconnaître n'importe quelle forme : c'est la transformée généralisée de Hough.

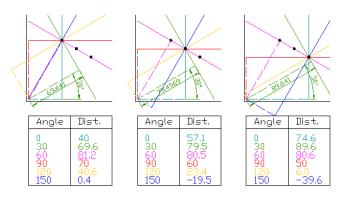


Figure 4.1 – Différentes étapes de la transformée de Hough

Pour une image (en 2 dimensions), il existe un nombre infini de lignes passant par chaque pixel dont la seule différence est l'angle de la droite. L'algorithme consiste en la construction d'un espace appelé espace de Hough en fonction de deux paramètres  $\theta$ l'angle de la droite et  $\rho$  la longueur du segment perpendiculaire à la droite d'angle  $\theta$  et passant par l'origine. On construit alors une sinusoïde en calculant toutes les valeurs de  $\rho$  pour différents  $\theta$ . Si des points sont alignés, alors ces sinusoïdes se couperont aux même endroit (cf. la droite violette sur la figure 4.1).

Le calorimètre hadronique possédant un grande granularité, peut être considéré comme une image en 3 dimensions. L'information des seuils fourni aussi un dégradé sur la concentration en particule. L'algorithme a été généralisé à 3 dimensions pour reconnaître les traces issus des gerbes hadroniques. La même procédure décrite précédemment est appliquée dans les plans X-Z et Y-Z.

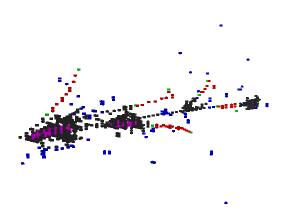


Figure 4.2 – Reconstruction de traces dans le calorimètre

Cependant, un traitement préliminaire doit être fait. Tous les *hits* ne sont pas éligible à la construction de trace tel que les *hits* appartenant au coeur de la gerbe. Ici on construira des amas ou *clusters* en agglomérant des *hits* dans chaque plan. Les amas isolés, c'est à dire des amas de petite taille, seront les "pixels" de la transformée de Hough.

La figure 4.2 présente un événement d'une gerbe de 40 GeV après une transformée de Hough. Les *hits* en rouge représentent les traces reconstruites, en bleu les *hits* isolés éligibles à la reconstruction de traces, en magenta les *hits* considérés comme le cœur de la gerbe et en gris le reste de la gerbe.

#### 4.3 Analyse en composante principale

L'analyse en composante principale (ACP) est une technique d'analyse statistique de la famille des analyses multivariées. Elle consiste à transformer des variables corrélées les unes des autres en un nouveau système de variable dé-corrélées. Cela permet souvent de réduire le nombre de variables de l'étude. On veut compresser un ensemble de N variables en estimant la variance et en la minimisant. On dit alors que ces variables sont réduites aux n premiers axes.

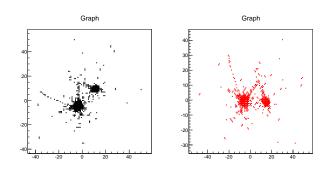


Figure 4.3 – Avant et après rotation du système grâce aux vecteurs de la ACP

Dans notre cas, l'ACP permet de faire des rotations dans l'espace des positions des hits. La rotation sera faite dans le plan XoY. Trouver l'axe principale se résume alors à trouver la matrice de covariance noté  $C_x^y$  (matrice 2x2) et de la diagonaliser. Les deux vecteurs propres trouvés donnent les deux axes principaux, celui ayant la valeur propre la plus grande étant l'axe principal. Celui ci permettra de faire tourner les gerbes dans le plan XoY afin de les aligner sur l'axe x comme sur la figure 4.3.

#### 4.4 L'algorithme de séparation dans sa totalité

L'algorithme appliqué pour séparer deux gerbes hadroniques proches consiste en une série d'algorithme. Le principe repose sur une séparation droite/gauche des gerbes. Toutefois, des suppositions importantes seront faites :

- Les gerbes se développent suivant la direction z sans angle d'incidence  $\theta$ .
- Seulement deux gerbes sont présentes dans l'événement.
- On applique la séparation tout en sachant que deux gerbes doivent être séparé.

C'est cette dernière supposition qui est à mon sens la plus importante. En effet, aucun indice n'est donné quant à la présence d'une seule gerbe hadronique ou plusieurs dans l'événement. Ces suppositions étant faites, on applique l'algorithme suivant :

- Un algorithme d'isolation est lancé sur tous les *hits* dans le calorimètre. La coupure inférieure effectuée est de 0,05.
- Par dessus l'algorithme d'isolation, est lancée la transformée de Hough pour identifier les traces. Les *hits* appartenant aux traces et qui étaient des isolés sont remplacés.
- Une nouvelle collection de *hits* est ensuite construite en retirant les *hits* isolés et les traces. Cette opération est faite afin de ne conserver que les bulbes des gerbes.
- Une analyse en composante principale sur la position des *hits* dans le plan XoY est ensuite effectuer afin de trouver la composante principale du système et d'effectuer une rotation des gerbes dans ce plan pour les aligner sur l'axe oX (cf figure 4.3).
- On se place dans le plan XoZ et on projette les *hits* sur l'axe oX. On obtient une accumulation de *hits* provenant des deux bulbes formant deux pics dans un histogramme. Au bout de 10 fois, si aucun pic secondaire n'est trouvé, l'algorithme s'arrête et un drapeau signifiant "je n'ai pas trouver deux pics pour la séparation" est déclaré. Ce drapeau permettra d'estimer l'efficacité de l'algorithme.
- Une fois les deux pics trouvés, il faut séparer les gerbes en coupant entre les deux pics. Mais un des deux pics pourrait être plus large que l'autre. Le centre est alors défini

• Les deux sommets des pics sont ensuite trouvé grâce à un algorithme codé dans ROOT [9] et encapsulé dans la classe TSpectrum. La paramétrisation de cet algorithme est important. Ici, il est demandé à la classe TSpectrum de trouver exactement deux pics. En fixant ce paramètre il ne peut en trouver plus de deux, mais il est possible de n'en trouver qu'un. Dans ce dernier cas, on diminue le rapport entre la valeur en ordonnée du plus haut pic et la valeur d'un pic cherché par 2; on répète cette étape 10 fois si aucun pic n'est trouvé.

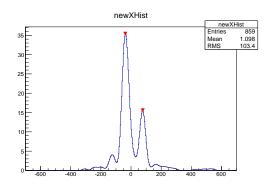


Figure 4.4 – Projection des hits sur l'axe oX

comme le barycentre des deux pics pondérés par leurs largeurs respectives :

$$centre = \frac{\frac{pic_1}{\sigma_1} + \frac{pic_2}{\sigma_2}}{\frac{1}{\sigma_1} + \frac{1}{\sigma_2}}$$

$$\tag{4.2}$$

• Les gerbes sont ensuite séparées en utilisant tous les hits de l'événement (isolés et traces compris).

Une fois les deux gerbes séparés l'énergie est estimée grâce à la formule 2.3 de la section 2.4 du chapitre 2.

#### 4.5 Les autres algorithmes développés

L'algorithme des coeurs est un algorithme basé sur la même principe que le l'algorithme d'isolation. Il itère sur les *hits* et compte ses plus proche voisins en pondérant par les seuils. La coupure appliquée est cette fois ci une coupure supérieure afin d'étiqueter des zones à grande densité. Ceci se résume par l'équation :

$$\sum_{i=1}^{27} \frac{seuil_i}{27} > c \tag{4.3}$$

Plus la valeur c sera grande et donc la densité minimum demandée élevée, plus le coeur sera petit.

L'algorithme de clustering est un algorithme permettant de rassembler des *hits* en amas (ou *clusters*) de proches en proches par une distance donnée entre deux *hits*.

La spécificité de cet algorithme vient de sa paramétrisation souple et avancée. En effet, il est possible de donner une distance d'agglomération entre deux hits (par défaut la plus proche cellule), de mettre une limite inférieure ou supérieure sur la taille d'un cluster, de demander à agglomérer des hits avec une ou plusieurs étiquettes particulières ou bien de rejeter une ou plusieurs étiquettes particulières. Par exemple on peut demander d'agglomérer tous les hits avec une distance entre chaque cellule de 1, une taille minimale de cluster de 5 et de n'utiliser que les étiquettes de "coeur".

## Chapitre 5

## Analyse des résultats

Ce chapitre présente l'analyse des résultats de la séparation des gerbes hadroniques. La séparation a été étudié pour différentes distances (5, 10, 15, 20, 25, et 30 cm) entre les gerbes et pour différentes énergies (10, 20, 30, 40, 50, 60, 70 et 80 GeV). Seulement, à la vue de toutes les combinaisons possibles pour une distance donnée et deux énergies, un choix a dû être fait. Dans le but de comparer ces résultats à une note de la collaboration CALICE [1], le choix de deux couples d'énergies a été fait : 10 - 10 GeV et 10 - 30 GeV. L'analyse de la séparation pour les deux couples d'énergie pour différentes distances sera faite dans deux sections séparées et la dernière section conclura cette analyse.

#### Contenu

5.1	Deux gerbes de 10 GeV	21
5.2	Gerbes de 10 GeV et 30 GeV	<b>23</b>
5.3	Conclusion de l'analyse et perspectives	25

Une fois les gerbes séparées, les énergies sont calculés avant séparation et après séparation. Les résultats étudient principalement la différence d'énergie de la première gerbe et donc dans les deux cas la gerbe de 10 GeV. L'application de l'algorithme a été faite sur les données de simulation (courbes en noires) et des données de test en faisceaux (courbes en rouges) réalisés en novembre 2012 au SPS au CERN. Le label "FTFP\_BERT" dans les légendes pour la simulation corresponds au modèle phénoménologique de physique hadronique utilisé dans Geant4 [3].

#### 5.1 Deux gerbes de 10 GeV

La figure 5.1 montre la différence d'énergie mesurée avant l'application de l'algorithme de séparation et l'énergie reconstruite pour une distance entre les gerbes de 5 cm (gauche) et 30 cm (droite). La distance de 5 cm représente dans cette étude la plus petite distance entre les gerbes. Similairement, la distance de 30 cm entre les deux gerbes est la plus grande des distances considérées. On s'attend dans ce cas à une grande efficacité des algorithmes et peu d'imperfections et de confusion.

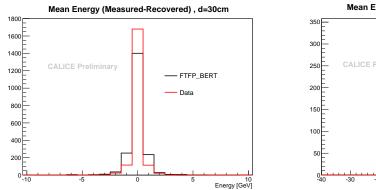


Figure 5.1 – Différence entre l'énergie mesurée et reconstruite à  $30~\rm cm$  (gauche) et  $5~\rm cm$  (droite) de distance entre deux gerbes de  $10~\rm GeV$ 

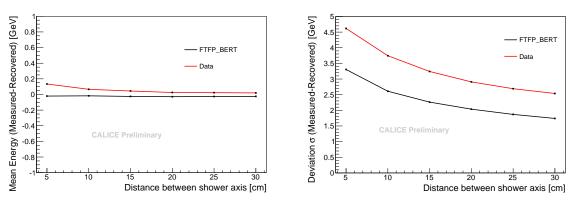


Figure 5.2 – Moyenne et dispersion de la différence en énergie pour différentes distances

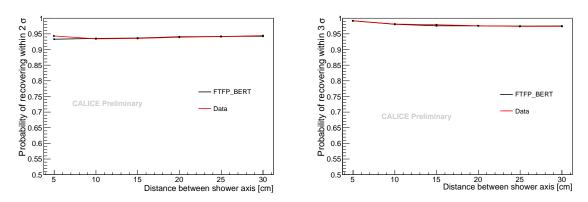


Figure 5.3 – Probabilité de retrouver une gerbe dans  $2\sigma$  (gauche) et  $3\sigma$  (droite) de deviation

On remarque que dans les deux cas la différence est centrée autour de 0. Cependant la dispersion de cette différence est plus importante pour une distance de 5 cm. En effet, la confusion devient plus importante au fur-et-à-mesure que la distance entre les gerbes diminue. La part d'énergie gagnée ou perdue reste cependant la même en moyenne dans le cas de deux gerbes de même énergie.

Les figures 5.2 montrent la moyenne et la dispersion de la différence d'énergie mesurée et reconstruite pour différentes distances entre les gerbes hadroniques. La moyenne reste ici centrée autour de 0 de façon constante avec un léger décalage pour les données. La dispersion quant à elle diminue quand la distance entre les gerbes augmente : la séparation devient de plus en plus efficace quand la distance augmente.

Les figures 5.3 montrent les probabilités de retrouver la gerbe de 10 GeV dans  $2\sigma$  et  $3\sigma$  autour de la moyenne. Si la distribution est gaussienne, on s'attends à retrouver 95% de chance de retrouver la gerbe dans  $2\sigma$  autour de la moyenne et 99.7% pour  $3\sigma$  autour de la moyenne. Ces graphiques permettent de vérifier une éventuelle déviation par rapport à une gaussienne en terme de probabilité. Ici, on retrouve les deux valeurs prédites avec un plateau à  $\simeq 94\%$  dans  $2\sigma$  de déviation et un plateau à  $\simeq 98\%$  dans  $3\sigma$  de déviation.

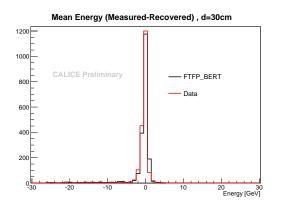
#### 5.2 Gerbes de 10 GeV et 30 GeV

Les figures 5.4 montrent la différence d'énergie mesurée et reconstruite pour deux distances entre les gerbes hadroniques : 5 cm et 30 cm. Ici les différences ne sont plus centrées même dans le cas à 30 cm. Le décalage se fait vers une différence négative comme le montre la figure 5.5 (gauche), c'est à dire que la gerbe de 10 GeV à tendance en moyenne à gagner de l'énergie. Le comportement est totalement différent comme on peut le voir sur la différence d'énergie à 5 cm (cf. figure 5.4 à droite).

L'explication est simple : la gerbe de 30 GeV est suffisamment énergétique pour se fragmenter en plusieurs morceaux. Lors de la projection sur l'axe oX et de la recherche des deux pics avec TSpectrum, la gerbe de 10 GeV est complètement mêlé dans la gerbe de 30 GeV et ne forme plus qu'un seul pic. Mais la gerbe de 30 GeV a produit un fragment dans la direction opposée à la gerbe de 10 GeV. La gerbe de 10 GeV et le premier fragment de 30 GeV sont alors identifiés comme étant la gerbe de 10 GeV et le deuxième fragment produit par la gerbe de 30 GeV est identifié comme la gerbe de 30 GeV. Ainsi, la gerbe de 10 GeV gagne de l'énergie et la gerbe de 30 GeV en perds. La même distribution pour la gerbe de 30 GeV aurait montré le même pic mais du coté positif.

En conséquence la valeur moyenne de la différence en énergie chute sévèrement jusqu'à -6,5 GeV à 5 cm de distance et la dispersion augmente jusqu'à une valeur critique de 12 GeV à 5 cm de distance. Cette déviation importante autour de la moyenne montre de manière significative la limitation de la méthode employée pour la séparation.

Les figures 5.6 montrent les probabilités de retrouver la gerbe de 10 GeV proche d'une gerbe de 30 GeV dans  $2\sigma$  et  $3\sigma$  autour de la moyenne. A faibles distances (de 5 à 15 cm), les probabilités montent anormalement vers 1. Ceci s'explique par la forme non gaussienne de la distribution avec un creux au centre de la distribution et deux pics de chaque cotés contrairement à la gaussienne. De ce fait un écart de  $2\sigma$  prendra plus de valeurs en compte que pour une distribution gaussienne. On voit cependant qu'à partir de 20 cm on retrouve les valeurs standards des probabilités pour une distribution gaussienne.



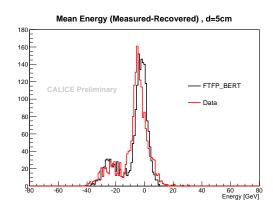
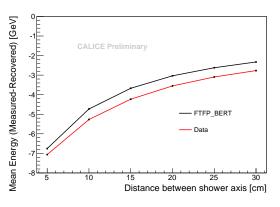


Figure 5.4 — Différence entre l'énergie mesurée et reconstruite à  $30~{\rm cm}$  (gauche) et  $5~{\rm cm}$  (droite) de distance entre deux gerbes de  $10~{\rm GeV}$  et  $30~{\rm GeV}$ 



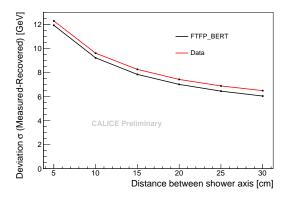
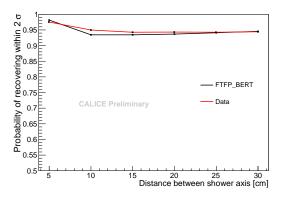


Figure 5.5 – Moyenne et déviation standard de la différence en énergie pour différentes distances



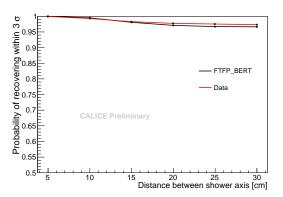


Figure 5.6 – Probabilité de retrouver une gerbe dans  $2\sigma$  (gauche) et  $3\sigma$  (droite) de deviation

#### 5.3 Conclusion de l'analyse et perspectives

L'analyse de ces résultats a montré que pour deux gerbes avec la même énergie, une séparation droite/gauche donne des résultats convenables. La dispersion de la différence d'énergie est faible et la valeur moyenne centrée en 0. Cependant, pour deux gerbes avec des énergies différente, la déviation augmente fortement et la moyenne de la différence est déplacée. L'apparition d'un nouveau pic dans le cas 10 et 30 GeV montre les limites de l'algorithme. La confusion devient statistiquement importante (apparition du pic secondaire) et ne permet pas de séparer convenablement deux gerbes hadroniques trop proches.

D'autres algorithmes décrits à la fin du chapitre 4 comme l'algorithme de cœur de gerbes pourrait palier au défaut de l'algorithme actuel. En effet pour deux gerbes proches, le cœur des gerbes, moins large que la gerbe dans sa totalité car contenu, permettrait de mieux les séparer. D'autre algorithmes basés sur des algorithmes de traitement d'image comme la reconnaissance de forme ou de type topologique sont actuellement en développement. Par exemple, des cœurs de gerbes pourrait être connectés entre eux par proximité ou bien par connexion par trace.

## Conclusion

Durant ce stage, j'ai travaillé dans le groupe ILC de Lyon sur un algorithme de reconstruction pour séparer des gerbes hadronique proches. J'ai pour cela utilisé des données de test en faisceau réalisés au SPS au CERN et des données de simulation Geant4 réalisé avec un prototype de calorimètre hadronique semi-digital ultra-granulaire (cf chapitre 2).

J'ai pû développé des algorithmes (cf chapitre 4) d'une part pour séparer des gerbes hadroniques proches comme l'algorithme d'isolation ou la transformée de Hough et d'autre part pour étudier les gerbes hadroniques.

Un algorithme pour séparer des gerbes hadroniques proches a été développé et exposé dans le chapitre 4. L'analyse des résultats dans le chapitre 5 a montré les limitations de l'algorithme et de nouveaux algorithmes sont en cours de développement afin de palier à ses défauts.

D'autres part, j'ai pu développer des outils pour le groupe travaillant sur le développement du SDHCAL ici à Lyon. Un environnement de travail numérique commence à se former pour la reconstruction et pour étudier les gerbes hadronique dans le prototype de calorimètre hadronique semi-digital. Ces outils sont développés de tel sorte à pouvoir inclure facilement des nouveaux éléments comme par exemple un détecteur additionnel lors de futurs tests en faisceau à la reprise du SPS en juillet 2014.

Je tiens à remercier Imad Laktineh pour la place qu'il m'a accordé au sein de l'équipe ILC de l'IPNL et pour le sujet de stage qu'il m'a proposé. Ce sujet m'a permis de trouver aussi mon sujet de thèse au sein de ce même groupe qui permettra d'améliorer les résultats de séparation dans le prototype et d'inclure ce travail dans des études de PFA dans l'ILD.

## Annexes

Longueur de radiation $X_0$	Longueur d'interaction $\lambda_I$	Segmentation transverse
1.77 cm	$16.76~\mathrm{cm}$	$1x1 \text{ cm}^2$
Nombres de chambres GRPC	Nombre de cellules	Volume total
48	96 x 96 x 48	$\sim 1.3~\mathrm{m}^3$

Table 5.1 – Quelques caractéristiques du SDHCAL

Épaisseur de la couche de gaz	Mélange gazeux	Potentiel d'ionisation du gaz
1.2 mm	$TFE(93\%)-SF_6(2\%)-CO_2(5\%)$	$10.12 \pm 0.02 \text{ V}$
	75 4 1 4 1 4 4 A	<del></del>
Haute tension usuelle	Résistivité du verre	Épaisseur totale

Table 5.2 – Quelques caractéristiques des GRPC

#  cellules/ASIC	# ASICs/DIF	$\# \ \mathrm{DIFs/Chambre}$
64	48	3
# Canaux/Chambre	# total de canaux	Valeurs des seuils de charges <b>usuels</b>
9316	442368	$0.114/5/15  \mathrm{pC}$

Table 5.3 – Quelques caractéristiques de l'électronique d'acquisition (DAQ)

$p_1$	$p_2$	$p_3$
0.0227148	$4.77839 e^{-5}$	$-4.05221 e^{-8}$
$p_4$	$p_5$	$p_6$
0.112287	$6.24094 e^{-6}$	$2.9642 e^{-8}$
$p_7$	$p_8$	$p_9$
0.156503	0.00010774	$4.97735 e^{-8}$

Table 5.4 — Meilleurs paramètres issus de la minimisation pour le calcul d'énergie

# Bibliographie

- [1] CALICE Collaboration. Pandorapfa tests using overlaid charged pion test beam data. CALICE Analysis Note CAN-024, 2010.
- [2] CALICE Collaboration. First results of the calice sdhcal technological prototype. CAL-ICE Analysis Note CAN-037, 2012.
- [3] Geant4 Collaboration. Geant4. http://geant4.cern.ch/.
- [4] ILC Collaboration. ILCSOFT. http://ilcsoft.desy.de/portal.
- [5] R. Eté. Baboon documentation. ftp://lyoftp.in2p3.fr/ilc/rete/Baboon.
- [6] F. Gaede. MARLIN. http://ilcsoft.desy.de/portal.
- [7] R. Kieffer. Caractérisation d'un calorimètre hadronique semi-digital pour le futur collisionneur ILC. PhD thesis, Université de Lyon, 2012.
- [8] The LCIO Team. LCIO. http://lcio.desy.de.
- [9] The ROOT Team. ROOT. http://root.cern.ch/root/html/.
- [10] Wikipédia. Hough Transform. http://fr.wikipedia.org/wiki/Transform%C3%A9e\_de\_Hough.

BIBLIOGRAPHIE 29