

MARCILLAUD DE GOURSAC Axel

## RAPPORT DE STAGE

**La Physique du quark top :  
Mesures de sections efficaces dans l'expérience D0**



Directeur de stage : Ursula BASSLER

Dates du stage : 09/01/06 - 24/02/06

Organisme :

Laboratoire de Physique Nucléaire et Hautes Energies

Université de Paris VI

**Résumé :**

Ce stage, effectué au sein de l'équipe  $D\bar{O}$  du LPNHE, m'a permis de découvrir les principaux aspects d'une grande expérience de physique des particules moderne : celle du Tevatron au Fermilab. Avec les données expérimentales, j'ai étudié la physique du quark top, spécialement sa désintégration dans le canal "électron+jets". J'ai pu constater l'importance de l'outil informatique dans une expérience de ce type-là, et c'est surtout avec celui-ci que j'ai travaillé à la détermination de la section efficace de la désintégration semi-leptonique du quark top.

**Abstract :**

This work placement has been made in the team  $D\bar{O}$  of the LPNHE and has allowed me to discover a big modern experience of particles physics : Tevatron in Fermilab. I have studied the quark top's physics with experimental data, especially the disintegration in the way "electron+jets". I have also noticed the importance of computer science in this experience, and I have worked with it on the cross-section of the semileptonic disintegration of the quark top.

**Remerciements :**

Je tiens tout d'abord à remercier M. Debu de m'avoir donné la possibilité de travailler dans son laboratoire.

Je remercie tout particulièrement Ursula Bassler pour la conduite de ce stage, pour son aide, son attention, son accueil. Un grand merci à Jean-Roch Vlimant, qui a conçu ce code que je n'ai fait qu'adapter, merci pour toute son aide et ses conseils.

Merci encore à ceux qui ne sont pas cités ici, mais qui m'ont aidé pendant ce stage.

## Table des matières

<b>1</b>	<b>Description de l'expérience</b>	<b>5</b>
1.1	Théorie . . . . .	5
1.2	L'accélérateur Tevatron et le détecteur $DØ$ . . . . .	8
1.3	Reconstruction des événements et environnement informatique	11
<b>2</b>	<b>Travail effectué au cours du stage</b>	<b>12</b>
2.1	Mon rôle . . . . .	12
2.2	Déroulement du stage et difficultés rencontrées . . . . .	13
2.3	Résultats et interprétation . . . . .	14
2.4	Découvertes imprévues . . . . .	19
2.5	Ce qu'il reste à compléter . . . . .	21



FIG. 1 – Le Tevatron

**Introduction :**

Le Modèle Standard, construit à partir des années 1970, permet de décrire avec une grande précision trois des quatre interactions fondamentales : l'électromagnétisme, l'interaction nucléaire forte et l'interaction nucléaire faible. Depuis, les physiciens des particules le testent de manière de plus en plus fine, surtout dans les accélérateurs actuels, pour vérifier son pouvoir prédictif ou découvrir une physique "au-delà du Modèle Standard". Dans cette optique, le Tevatron, situé dans la banlieue de Chicago, permet aujourd'hui de produire des collisions entre des protons et des antiprotons d'énergie de l'ordre du TeV. C'est dans son anneau principal de deux kilomètres de diamètre que se trouve entre autre le détecteur de l'expérience  $D\bar{O}$  qui détecte les particules émises lors de la collision proton-antiproton, la plupart des particules du Modèle Standard pouvant être produites à cette échelle d'énergie. Et en effet, en 1995, la collaboration  $D\bar{O}$ , de concert avec l'expérience CDF, découvrait le dernier quark du Modèle Standard, le quark top. Celui-ci est principalement créé sous la forme de quarks top-antitop ( $t\bar{t}$ ). Parmi les différents modes de désintégration possibles de cette paire de quarks, il y a notamment le canal "électron+jets", qui nous intéressera tout au long de ce texte.

C'est au sein de l'équipe  $D\bar{O}$  du LPNHE, s'occupant de la mesure de la section efficace de désintégration des quarks top dans le canal "électron+jets", que j'ai effectué mon stage. Cette équipe dispose d'un accès aux données numériques de l'expérience au Fermilab, sous un certain format informatique, format de données qui évolue avec les années. Et cette équipe a développé des outils graphiques et de calcul permettant d'interpréter ces quantités phénoménales de données. Mon travail durant ce stage a consisté tout d'abord à interfacier ces outils de calcul avec le dernier format informatique des données, puis à tester la qualité de ces données numériques et la cohérence des outils graphiques.

Nous pourrions donc étudier dans un premier temps ce qui se rapporte à l'expérience en elle-même, c'est-à-dire d'une part donner les réactions et les principes théoriques mis en jeu dans l'expérience et d'autre part, décrire brièvement le détecteur  $D\bar{O}$ . Pour une étude plus détaillée, on pourra se reporter utilement à la thèse de Jean-Roch Vlimant. Dans une seconde partie, nous aurons l'occasion de voir en détail quelle était ma fonction au sein de l'équipe, les objectifs de mon stage, les difficultés que j'ai rencontrées et les résultats auxquels j'ai abouti.

# 1 Description de l'expérience

## 1.1 Théorie

Dans un premier temps, nous pouvons rappeler les principes du Modèle Standard qui sont requis par les réactions que nous observerons par la suite. Tout d'abord l'interaction électrofaible. On rappelle qu'après brisure de la symétrie  $SU(2)_L \times U(1)_Y$  électrofaible, nous obtenons des doublets gauches de quarks et de leptons, des singulets droits de ces mêmes quarks et leptons, trois bosons intermédiaires massifs  $W^\pm$ ,  $Z^0$ , un boson de masse nulle : le photon, et un champ scalaire réel massif : le champ de Higgs. Les quarks et les leptons sont regroupés en trois générations et a priori, il n'y a pas d'échange possible entre ces générations. Cependant, les états propres de masse et d'interaction électrofaible étant différents pour les quarks, cela se voit dans le lagrangien de Yukawa, il se produit un mélange entre les saveurs résumé par une relation où intervient la matrice CKM. Le mélange entre les quarks des deux premières générations est relativement important alors que les quarks de la troisième génération ne se mélangent que très peu. Nous utiliserons ce résultat par la suite, que le quark top a une désintégration électrofaible quasiment à 100% en quark b de la façon suivante :  $t \longrightarrow qW^+$ .

Les jets de hadrons font intervenir la partie Chromodynamique quantique du Modèle Standard. Les particules "élémentaires" soumises à l'interaction forte sont les quarks. Ils portent une des trois charges de "couleur", car le groupe de symétrie est ici  $SU(3)$ . Il y a huit bosons vecteurs de jauge appelés les gluons, de masse et de charge électrique nulles. Cependant, ni charge de couleur ni les charges fractionnaires des quarks ne sont observées expérimentalement. Les gluons et les quarks ne sont donc pas présents à l'état libre. Comme la constante de couplage devient importante à grande distance, le calcul perturbatif n'est plus valable pour étudier la fragmentation et le confinement des quarks. On utilise alors des modèles phénoménologiques, selon lesquels les quarks et les gluons émis par les réactions, ne pouvant exister que sous forme de hadrons, se transforment en jets de hadrons par émissions successives de partons.

Voyons à présent les suites d'une collision proton-antiproton et le calcul de la section efficace de désintégration, le but ultime de tout ce qui est entrepris dans ce rapport. A chaque processus physique est associée une section efficace ( $\sigma$ ), qui a la dimension d'une surface ( $1\text{pb} = 10^{-40}\text{m}^2$ ). Plus cette quantité est grande, plus le processus a de chances de se produire. Pour un faisceau donné, on appelle luminosité ( $\mathcal{L}$ ) le taux de collisions possibles par unité de surface. La luminosité instantanée est donnée en  $\text{pb}^{-1}\text{s}^{-1}$  et la luminosité intégrée en  $\text{pb}^{-1}$ . On retrouve ces notions dans l'expression du nombre d'événements observés pour un processus X en l'absence de bruit

de fond :

$$N_X^{obs} = \sigma_X \cdot \mathcal{L} \cdot \epsilon$$

où  $\epsilon$  représente le produit de l'acceptance du détecteur (fraction du signal détectée) par l'efficacité de sélection (fraction des événements détectés passant les coupures de présélections).

Nous passerons sous silence le processus de production de  $t\bar{t}$  à partir de la collision proton-antiproton, et le calcul de sa section efficace, pour nous intéresser plus particulièrement à la désintégration de  $t\bar{t}$ . La durée de vie du quark top étant très inférieure au temps typique d'hadronisation, nous pouvons considérer que  $t\bar{t}$  se désintègre de manière prépondérante par ce processus :

$$t\bar{t} \longrightarrow W^+ + b + W^- + \bar{b}$$

comme notifié dans les rappels sur l'interaction électrofaible. Le boson W se désintègre alors hadroniquement ou leptoniquement avec les rapports d'embranchement suivants :

$$\begin{aligned} BR(W \rightarrow hadrons) &= 67.96 \pm 0.35 \% \\ BR(W \rightarrow leptons) &= 3 \cdot (10.68 \pm 0.12) \% \end{aligned}$$

Cela donne trois modes de désintégration possibles. Le canal dominant statistiquement est le canal hadronique, c'est-à-dire que les deux W se désintègrent en paires de quarks. Le rapport d'embranchement est de 45%, ce qui est a priori très intéressant, mais ce canal souffre d'un important bruit de fond QCD. Les canaux les moins importants statistiquement sont les canaux dilepton, les deux W se désintégrant en leptons. Le rapport d'embranchement de chaque canal est de 1.2%. Pour finir, le canal auquel nous allons le plus nous intéresser est le canal intermédiaire "électron+jets", l'un des W se désintégrant en paire de quarks, l'autre en paire de leptons. C'est un bon compromis entre statistique et bruit de fond QCD. Il possède un rapport d'embranchement de l'ordre de 15%. Dans le canal "électron+jets", nous devons cependant prendre en compte la correction due au processus  $W \rightarrow \tau + \dots \rightarrow W + \dots \rightarrow e + \dots$ . Donc :

$$BR(W \rightarrow e+X) = BR(W \rightarrow e) + BR(W \rightarrow \tau) \cdot BR(\tau \rightarrow e) = 12.58 \pm 0.12\%$$

Puis le rapport d'embranchement du canal e+jets s'écrit :

$$BR_{e+jets} = 2 \cdot BR(W \rightarrow e+X) \cdot BR(W \rightarrow hadrons) = 17.106\%$$

Et le processus se résume en :

$$p + \bar{p} \longrightarrow e + 4 \text{ jets} + \text{énergie transverse manquante (MET)}.$$

Étudions maintenant les principales sortes de bruit de fond pouvant se mêler au signal. Le bruit de fond irréductible est issu d'un processus physique qui se caractérise par une signature expérimentale identique et indiscernable du signal. Le bruit de fond irréductible que l'on prendra en compte ici est celui du "W+jets", qui correspond à la production d'un W et de trois ou quatre partons dans l'état initial. Au contraire, si le bruit de fond réductible possède une signature identique au signal, cela est dû à l'imperfection du détecteur. Le canal dilepton avec production de jets supplémentaires en est un exemple, mais que nous négligerons dans cette étude. En revanche, le bruit de fond QCD, correspondant au canal hadronique suivi d'une désintégration électromagnétique d'un des hadrons en "faux électron", est très important et sera pris en compte dans cette analyse. Nous retenons donc deux sources principales de bruit de fond : W+jets et QCD.

Afin d'éliminer ces deux sources de bruit de fond lors de l'analyse des données expérimentales, nous procédons de la façon suivante.

D'une part, les événements sont soumis à deux coupures différentes pour l'électron, une sélection lâche (loose) et l'autre plus stricte (tight). Cela conduit à deux lots de données : le lot Loose (L) des événements passant la coupure loose, le lot Tight (T) de ceux passant les deux coupures. Un troisième peut être déduit : le lot QCD des événements passant la coupure loose mais pas la coupure tight. Le lot L contient alors des événements de bruit de fond QCD, de bruit de fond W+jets et du signal recherché : la désintégration semileptonique de  $t\bar{t}$ . Le lot T comporte les mêmes composantes mais en des proportions différentes : la composante QCD est très réduite du fait de la sélection sur l'électron. Le lot QCD a une grande composante QCD et une faible proportion de signal et de W+jets. Nous pouvons alors définir les efficacités signal et QCD ( $\varepsilon_{signal}$  et  $\varepsilon_{QCD}$ ) pour que respectivement un vrai électron et un faux électron issu d'un événement QCD passent la coupure stricte tight. Une fois ceci défini, on peut simuler le signal et le bruit de fond W+jets grâce à un simulateur Monte Carlo, mais, la simulation du bruit de fond QCD étant difficile, on estime celui-ci directement à partir des données. Puis la probabilité de vraisemblance entre cette estimation et l'observation permet de discerner le bruit de fond QCD et les événements avec un vrai électron (signal et W+jets).

D'autre part, on construit un discriminant topologique à partir de variables topologiques et ceci afin de discerner le bruit de fond W+jets et le signal. Une probabilité de vraisemblance entre le modèle (les simulations Monte Carlo) et les données est calculée, ce qui permet d'obtenir le résultat souhaité.

Nous sommes donc en mesure de déterminer le nombre d'événements

des trois principales composantes considérées (ce qui sera le travail du programme TopFit) et notamment le nombre de signal  $N_T^{t\bar{t}}$  tight. La section efficace s'obtient alors :

$$\sigma_{t\bar{t}} = \frac{N_T^{t\bar{t}}}{\mathcal{L} \cdot \epsilon \cdot BR_{e+jets}}$$

où  $BR_{e+jets}$  est connu, la luminosité  $\mathcal{L}$  et l'efficacité de présélection  $\epsilon$  doivent être déterminées à partir de l'expérience.

## 1.2 L'accélérateur Tevatron et le détecteur $D\bar{O}$

Voyons dans un premier temps la structure de l'accélérateur Tevatron (cf Figure (1)).

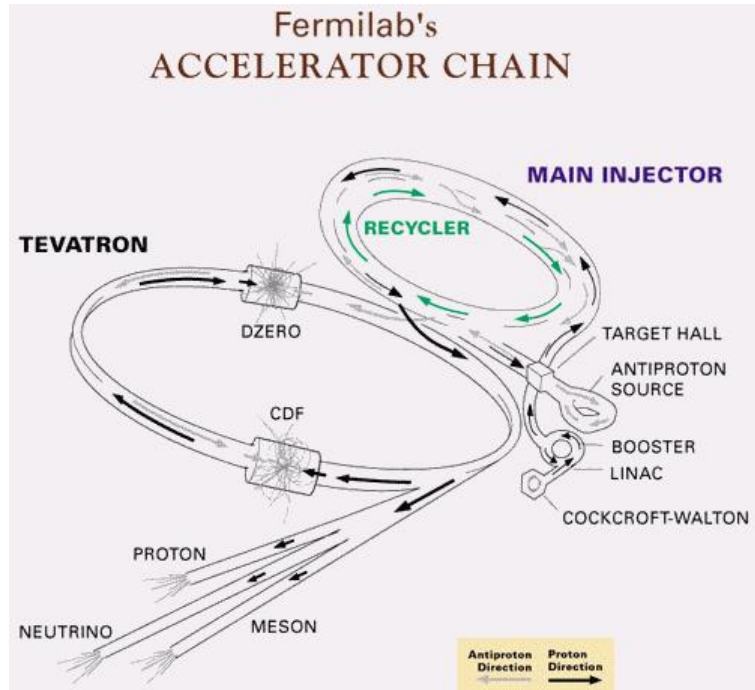


FIG. 2 – Chaîne d'accélération du Fermilab

La source primaire des faisceaux est une bonbonne d'hydrogène liquide alimentant un canon à ions  $H^-$  au Césium. Les ions hydrogène sont injectés à 750 keV dans l'accélérateur linéaire, succession de cavités radio-fréquences classiques. En bout de l'accélérateur linéaire, avec une énergie de 400 MeV et regroupés par paquets, les ions sont alors injectés dans l'anneau du synchrotron appelé Booster, succession de cavités radio-fréquences et d'aimants courbant la trajectoire des particules, et projetés sur une feuille de carbone afin de produire un faisceau de protons. Celui-ci est porté à une énergie de 8



GeV et segmenté en paquets, puis il passe dans le Main Injector, synchrotron à aimants conventionnels. L'énergie des protons est augmentée jusqu'à 150 GeV et des antiprotons sont produits par collision d'une partie des protons sur une cible de Nickel, et accélérés dans le sens opposé des protons jusqu'à 150 GeV. Mais le faible taux de production d'antiprotons limite la durée de vie du faisceau à quelques heures, cela définit un store.

Le tunnel principal de deux kilomètres de diamètre abrite un synchrotron à aimant supra-conducteur appelé Tevatron. Les faisceaux de protons et d'antiprotons tournent alors en sens inverse et sont accélérés jusqu'à 980 GeV, puis déviés aux points de croisement pour assurer une collision face à face avec une énergie de 1.96 TeV dans le centre de masse. Les protons et antiprotons sont répartis en 3 super-paquets de 12 sous-paquets, et les collisions de deux sous-paquets se font toutes les 396 ns. Une prise de données est partitionnée en runs de quelques centaines de milliers d'événements, et ceci pour pouvoir adapter le protocole de déclenchement à la variation de luminosité qui diminue de façon exponentielle au cours du store.

Intéressons nous à présent au détecteur  $D\bar{O}$  (cf Figure (2)).

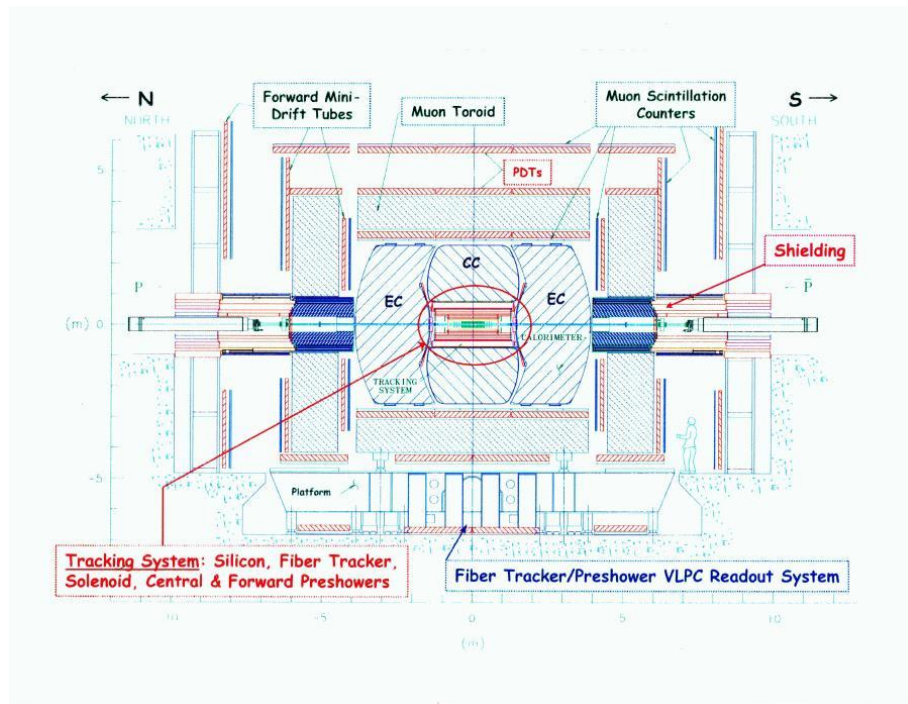


FIG. 3 – Schéma du détecteur  $D\bar{O}$

Assemblé au début des années 1990, le détecteur  $D\bar{O}$  possède une symétrie de révolution axiale. Les coordonnées utilisées lors des expériences ne sont pourtant pas les coordonnées cylindriques. En effet, en physique hadronique,

on utilise la rapidité  $Y$ , et dans l'approximation hyper-relativiste qui sera vérifiée ici, la pseudo-rapidité  $\eta \sim Y$ .

$$Y = \frac{1}{2} \ln \left( \frac{E + p_z}{E - p_z} \right)$$

$$\eta = -\ln \left( \tan \frac{\theta}{2} \right)$$

Pour le stockage des données expérimentales, on utilisera donc les coordonnées pseudo-projectives  $(\varphi, \eta, z)$ .

Comme il est impossible et inutile d'enregistrer toutes les collisions (taux de 2.5 MHz) car la plupart d'entre elles sont des événements QCD à basse énergie ou issus du bruit de fond machine, il existe un système électronique complexe de déclenchement à trois niveaux qui filtre les événements sur la base de leur signature physique. Le niveau 1 effectue sa décision à un taux de 2 kHz à partir d'informations non-croisées des différents sous-détecteurs, excepté le détecteur de vertex. Le déclenchement de niveau 2 combine les informations des sous-détecteurs et du détecteur de vertex et fonctionne à 800 Hz. Le niveau 3 de déclenchement collecte des informations encore grossières mais à partir des événements reconstruits, ce qui augmente la précision, et son taux de décision est de 50 Hz. Nous ne donnerons pas plus de détails sur l'électronique complexe permettant de collecter les données expérimentales mais nous reviendrons prochainement sur les formats informatiques sous lesquels elles sont stockées et mises à disposition des physiciens.

Mais avant cela, donnons quelques détails à propos des sous-détecteurs. La luminosité instantanée est mesurée par le luminomètre, composé de disques de 24 secteurs de scintillateurs. Deux trajectographes, l'un à micro-piste de silicium, l'autre à fibres scintillantes, servent à la reconstruction des traces chargées et par conséquent à celle des vertex d'interactions. La trajectoire des particules chargées est courbée par un solénoïde supra-conducteur afin de permettre une mesure de leur impulsion transverse. Le détecteur de pied de gerbes sert à détecter les particules qui auraient développé leur gerbe électromagnétique avant d'avoir atteint le calorimètre. Le calorimètre est constitué d'une succession de couches d'absorbeurs où les particules interagissent avec la matière pour développer une gerbe de particules dont l'énergie est mesurée. Il y a deux parties dans ce calorimètre : une partie électromagnétique qui repère les objets électromagnétiques comme l'électron et le photon, qui sont freinés en premier, et une partie hadronique qui détecte les jets de hadrons. Les muons n'étant pas arrêtés par le calorimètre, un spectromètre à muons a été installé pour les détecter. Ceci conclue notre brève description du détecteur  $DØ$ .

### 1.3 Reconstruction des événements et environnement informatique

Afin d'expliquer le travail que j'ai effectué au sein de la collaboration  $DØ$ , je dois d'abord décrire la reconstruction des événements et l'environnement "soft" de cette expérience, c'est-à-dire le format informatique des données après traitement.

En effet, les signaux enregistrés par les différents détecteurs sont filtrés au niveau du système de déclenchement à trois niveaux que l'on a déjà décrit. Puis, ces signaux sont traités pour en améliorer la qualité et l'on peut reconstruire les événements, c'est-à-dire les collisions proton-antiproton, en recoupant les informations des différents détecteurs que nous venons de voir. Tout d'abord les traces chargées sont reconstruites à partir du détecteur à micropistes de silicium et du trajectographe à fibres scintillantes.

Ensuite, comme il peut se produire plus d'une interaction par croisement de faisceaux proton-antiproton, il peut y avoir plusieurs vertex d'interactions. On utilise les traces déterminées précédemment, et plus précisément leurs croisements pour déterminer le vertex primaire d'interaction, et lui associer les traces qui conviennent.

Puis les objets électromagnétiques, électrons et photons, sont reconstruits dans la section électromagnétique du calorimètre à l'aide des trajectographes, et de même, les jets de hadrons typiques de l'hadronisation d'un parton ou de la désintégration hadronique d'un lepton tau sont reconstruits dans la partie hadronique de ce même calorimètre. Les muons sont mesurés par le détecteur le plus externe. A chaque particule, on associe donc quantités d'informations comme la trace associée, l'énergie, l'impulsion, la charge,...

Enfin, les particules n'interagissant que faiblement, comme les neutrinos, ne sont pas détectés, mais il est possible de reconstruire partiellement leur cinématique transverse.

A un événement sont associées toutes ces informations, en plus d'informations générales comme le numéro du run, le numéro de l'événement, la luminosité instantanée, ..., et cet événement est alors codé numériquement, muni de toutes ses informations, dans un langage informatique : le C++ par l'intermédiaire du logiciel root, développé par le CERN. Ce langage est orienté objet, ce qui facilite donc la hiérarchie des informations et l'utilisation d'arbres, structure informatique optimisée pour ce type de traitement, pour coder des événements.

Cependant, ce format, sous lequel sont disponibles les données de l'expérience  $DØ$  à toute la collaboration de physiciens qui travaillent dessus, a évolué depuis le début de cette expérience. En simplifiant un petit peu les choses, nous pouvons dire que les événements munis de leur informations se présentaient sous la forme d'arbres appelés TMBTrees, et c'est sous ce format que l'équipe du LPNHE travaillant sur le top dans  $DØ$  utilisait ces données. Mais, la collaboration  $DØ$ , dirigée depuis le Fermilab pour ces aspects, a développé

depuis un ensemble d'outils regroupés sous la dénomination CAF, outils qui facilitent le travail des physiciens et qui deviennent standards pour toute la collaboration. Les événements sont maintenant codés par des arbres de type CAFTrees. Afin de bénéficier de ces nouveaux outils et pour rester au format standard de la collaboration  $DØ$ , ce qui est important notamment pour comparer les résultats, l'équipe du LPNHE dans laquelle j'ai effectué ce stage a dû adapter son ancien travail à ce nouveau format. Et c'est dans cette optique que se place mon stage.

## 2 Travail effectué au cours du stage

### 2.1 Mon rôle

L'équipe du LPNHE travaillant sur la désintégration du top dans  $DØ$  dans la voie "électron+jets", et notamment l'ancien thésard Jean-Roch Vliant de ma directrice de stage Ursula Bassler, a mis au point un format informatique en root et C++ : TopDump, utilisant aussi la structure d'arbre, mais ne gardant que les informations potentiellement utiles pour cette équipe, ce qui réduit grandement l'espace mémoire utilisé et donc accélère tous les calculs informatiques. Une interface avait donc été développée, qui implémentait des TopDumps à partir de TMBTrees, l'ancien format de  $DØ$ . Et sur ces TopDumps, on pouvait faire tourner des logiciels, deux notamment : TopFit et Util, qui donnaient des résultats en un temps considérablement moins long que s'il avait fallu le faire sur des TMBTrees. L'interface se chargeait aussi de faire les présélections qui convenaient sur les événements en imposant à ceux-ci de satisfaire à certaines coupures. Par exemple, il était notifié si l'électron choisi par l'interface comme électron candidat, passait la coupure loose, et s'il passait la coupure tight, coupures dont nous avons déjà parlé en première partie.

Détaillons maintenant l'utilité de ces deux logiciels. Le premier, TopFit prend comme paramètre quatre TopDumps, l'un décrivant les données, les trois autres, les simulations Monte Carlo du signal qu'on cherche à observer, du bruit de fond QCD et du bruit de fond W+jets. Dans ces quatre TopDumps, tous les événements ont passé la coupure loose, ils appartiennent donc tous au lot L; certains appartiennent aussi au lot T, ils ont passé la coupure tight, les autres au lot QCD, comme indiqué en première partie. Ce logiciel TopFit prend aussi comme argument des courbes de référence de certaines variables topologiques, et les efficacités signal ( $t\bar{t}$ ) et QCD. A partir de cela, l'algorithme utilise d'une part la méthode de la matrice, puis d'autre part, avec les courbes de référence, il calcule le discriminant topologique, ce qui permet de comparer le modèle (les simulations Monte Carlo) aux données expérimentales, et donc de donner une estimation statistique du nombre d'événements des données (loose et tight) provenant vraiment du bruit de fond QCD, du bruit de fond W+jets, et du signal recherché. Ce

programme reprend ainsi exactement la stratégie que nous nous étions fixés dans la première partie pour déterminer les trois composantes des données expérimentales.

Une fois obtenus ces précieux nombres, le logiciel Util produit des graphes de toutes les variables jugées utiles sur les données en précisant la part de signal recherché, la part de bruit de fond QCD et la part de bruit de fond W+jets.

Mais le nouveau format des CAFTrees est arrivé, et c'est ici que mon stage intervient. En effet, les CAFTrees, s'il permettent d'utiliser les outils de CAF, prennent aussi un gigantesque espace mémoire. Et il a donc fallu faire une sélection des informations des événements pour ne garder que celles utiles. Ce fut mon travail dans un premier temps de réaliser une interface entre le format des CAFTrees et celui des TopDumps. Celle-ci devait alors se servir des nouveaux outils de CAF pour effectuer les présélections, puis implémenter par l'intermédiaire d'un Processor, autre outil développé dans CAF, les TopDumps à partir des informations des CAFTrees. Dans une deuxième partie du stage, j'ai dû vérifier l'adaptation des deux logiciels TopFit et Util aux TopDumps provenant des nouveaux formats CAF et déterminer les nouveaux paramètres de TopFit pour la version du programme de reconstruction p17 : courbes de références, efficacités.

## 2.2 Déroulement du stage et difficultés rencontrées

Pour être capable de coder cette interface, j'ai dû apprendre un certain nombre de choses. Tout d'abord, je me suis initié au langage de programmation C++ et aux développements apportés par le logiciel Root. Cela m'a pris une bonne semaine d'apprentissage intensif car je ne connaissais rien sur le sujet. Et dans le même temps, j'ai lu la thèse de Jean-Roch Vlimant pour comprendre les notions de physique expérimentale utilisées ici, par exemple ce que nous avons vu dans la première partie : la théorie, le principe du détecteur et l'environnement informatique de l'expérience. J'ai ensuite pu apprendre à me servir des sélecteurs développés dans CAF, qui effectuent les coupures sur les jets détectés, les électrons,... Il a fallu les récupérer un à un en vérifiant que c'était la bonne version, que celle-ci était compatible avec les autres sélecteurs et le reste de mon code, et comprendre comment bien les utiliser. Si je mentionne ici ce qui peut paraître des évidences, c'est pour insister sur le fait que ces étapes ne se sont pas déroulées sans une certaine difficulté : je suis quelquefois resté une journée entière bloqué sur des problèmes de compatibilité de version entre les différents "packages" des sélecteurs, n'étant moi-même qu'un débutant en C++ et Root. En parallèle, j'ai effectué un travail de compréhension du code de l'ancienne interface TMBTrees-TopDumps de Jean-Roch Vlimant afin de repérer les informations pertinentes à récupérer dans les événements. Puis vint le temps

du codage effectif de l'interface CAFTrees-TopDumps, ce qui, là aussi, n'a pas été sans poser des problèmes à quelqu'un qui ne maîtrisait que les rudiments de Linux. Cela a donné lieu à des erreurs de compilation, à des erreurs d'exécution, erreurs qui heureusement se sont faites de moins en moins nombreuses avec le temps grâce à l'aide précieuse de ma directrice de stage et de Jean-Roch Vlimant. Une fois les éléments fondamentaux de l'interface codés et avant d'envisager toute amélioration, il a fallu tester cette interface sur des données pour savoir ce qu'il y avait à corriger. Ma directrice de stage et moi avons donc choisi trois fichiers dans les bases de données du Fermilab, un contenant des données expérimentales, les deux autres étant des simulations Monte-Carlo du signal recherché (désintégration de  $t\bar{t}$  en électron+jets) et du bruit de fond W+jets. J'ai lancé la nouvelle interface produisant des TopDumps avec ces trois fichiers contenant des CAFTrees et créé un petit programme qui permet de visualiser rapidement les graphes des variables importantes de ces TopDumps produits. Cela a permis de repérer certaines imperfections du code, et en refaisant plusieurs fois cette procédure : codage puis vérification graphique, de parvenir à une interface de plus en plus opérationnelle.

L'autre volet de ce stage a consisté en l'adaptation des deux programmes TopFit et Util. Il y eut dans un premier temps une phase de compréhension des fonctionnalités de ces deux programmes, puis une de vérification du bon fonctionnement de ceux-ci avec les nouveaux TopDumps produits. Pour cela, j'ai dû m'enquérir des paramètres nécessaires à TopFit : les courbes de références des variables topologiques, les efficacités. N'ayant pas les moyens de déterminer ceux-ci pour p17 au début, je dus prendre ceux de p14, en considérant que ce n'était qu'une approximation, avec pour résultat les premiers graphiques détaillant la part de signal et de bruit de fond.

Ce stage s'est achevé par deux exposés oraux, traitant du travail de l'équipe  $D\bar{O}$  du LPNHE sur cette interface, en visioconférence avec les collaborateurs étrangers s'intéressant aussi à ces aspects de l'expérience  $D\bar{O}$ . Nous pouvons maintenant regarder l'essentiel de ces deux présentations que j'ai effectuées, qui forment une sorte de bilan de ce qui a été accompli pendant ce stage.

### 2.3 Résultats et interprétation

Cet examen de la situation peut se diviser en trois parties : les présélections, l'implémentation des TopDumps par l'interface, et l'utilisation de TopFit et d'Util.

Commençons par les sélecteurs intervenant sur les données et les simulations Monte Carlo du signal et du W+jets pendant la phase de présélection des événements. Tout d'abord vient le sélecteur sur la qualité des données.

Il supprime tous les événements produits par un bruit de fond repéré. Par exemple, tous les jours à une certaine heure au Fermilab, on constatait un bruit de fond très important dans les données. Finalement on en trouva la source : c'était un atelier de soudure proche du détecteur qui fonctionnait à heures fixes. Les deux sélecteurs suivants n'interviennent que sur les données : le sélecteur de Trigger, qui agit pour que la sélection soit cohérente avec le tri au niveau du système de déclenchement du détecteur, et le sélecteur qui supprime les événements dupliqués, qui, pour l'instant, semble ne pas fonctionner.

Les sélecteurs suivants agissent sur tous les types de fichiers, données comme simulations. Celui de l'électron impose toute une série de coupures (loose), notamment sur l'impulsion transverse :  $pT > 20$  GeV). Il impose qu'il y ait au moins un loose dans le calorimètre central, choisit le bon électron candidat s'il y a plusieurs objets électromagnétiques dans le calorimètre central, et regarde si ce bon électron passe les coupures tight. Celui du vertex impose qu'il y ait moins de 60 vertex reconstruits par événement, et au moins 3 traces associées par vertex. Celui des jets choisit les bons jets reconstruits, passant les coupures  $pT > 20$  GeV et  $\eta > 2.5$ , et en impose au moins un par événement. Celui de l'énergie transverse manquante (MET) exige à l'événement  $MET > 20$  GeV. Le sélecteur d'électron véto demande qu'il y ait au plus un électron tight. On ne parlera pas pour l'instant du Processeur, qui lui aussi peut effectuer certaines présélections.

TAB. 1 – Efficacités des présélections

Sélecteur	données	données	signal MC	W+jets MC
	$N_{evts}$	eff. intégrée	eff. intégrée	eff. intégrée
Initialement	29528499	100%	100%	100%
Qualité des données	24445493	82.79%	96.13%	95.89%
Trigger	3596394	12.18%		
Événements dupliqués	3596394	12.18%		
Electron	1020069	3.45%	19.00%	36.03%
Vertex	996440	3.37%	18.82%	35.20%
Jets	842909	2.85%	18.80%	3.96%
MET	112970	0.38%	17.72%	3.19%
Electron véto	112600	0.38%	16.80%	3.19%
Processeur	112174	0.38%	16.80%	3.19%

Nous trouvons dans le tableau (TAB 1) les efficacités de présélection intégrées au fur et à mesure des sélecteurs. On se souvient que l'efficacité de présélection, après tous les sélecteurs, intervient dans le calcul de la section efficace. J'ai donc déterminé celle-ci (0.38%) pour les présélections que j'ai

décidé d'imposer pendant mon stage.

L'interface entre les CAFTrees et les TopDumps que j'ai développée, a été codée dans le Processeur (MyProcessor). Après les présélections que nous avons vues, le Processeur extrait les informations suivantes des événements pour remplir le TopDump : des informations générales (numéro du run, numéro de l'événement, luminosité instantanée,...), les bons jets (ayant passés les coupures des présélections) avec toutes leurs informations (énergie, impulsion, charge, trace associée,...), le bon électron candidat et toutes ses informations, des informations sur l'énergie transverse manquante (masse transverse, énergie transverse, énergie du calorimètre), et des variables topologiques (Energie scalaire transverse  $H_t$ , centralité, sphéricité, aplanarité,  $\Delta\phi$ , discriminant topologique).

Nous pouvons à présent regarder quelques graphiques issus de mon programme de contrôle, qui illustreront les affirmations que nous avons faites dans la première partie.

Nous avons vu dans la première partie que l'on soumettait les électrons à deux types de coupures : loose et tight, dans le but de recueillir trois lots de données : Loose, Tight et QCD. Nous avons ensuite fait l'hypothèse qu'une grande proportion des vrais électrons des événements signal et  $W$ +jets passaient la coupure tight alors que les faux électrons des événements QCD étaient en majorité rejetés par cette coupure. Les trois graphes (Figure (3)) nous donnent l'occasion de le vérifier : en effet, sur le graphe des données, les événements QCD étant majoritaires, on constate que 62% des événements ont un électron candidat ne passant pas la coupure tight. En revanche, pour le signal et  $W$ +jets, la grande majorité des événements, respectivement 85% et 92%, ont un électron tight.

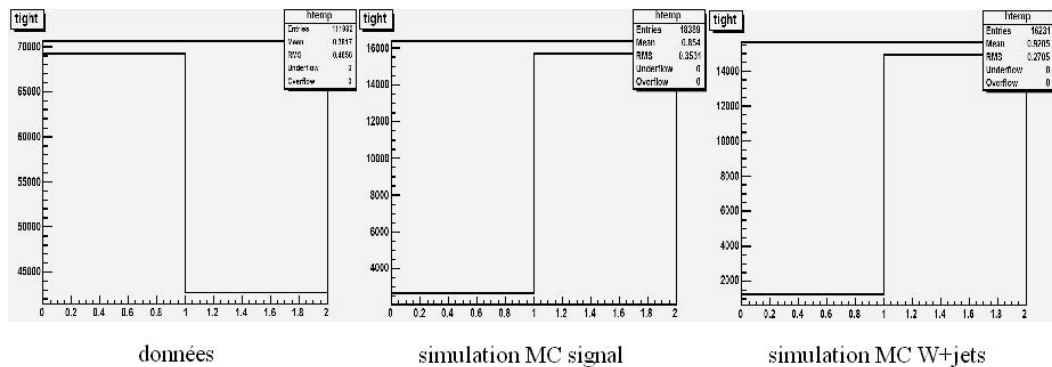


FIG. 4 – Nombre d'électrons tight et non tight



Vérifions à présent que le discriminant topologique qui a été construit permet de discerner le bruit de fond irréductible W+jets du signal recherché. On constate effectivement (cf Figure (4)) que ce discriminant est statistiquement proche de 0 pour le bruit de fond W+jets, et proche de 1 pour le signal.

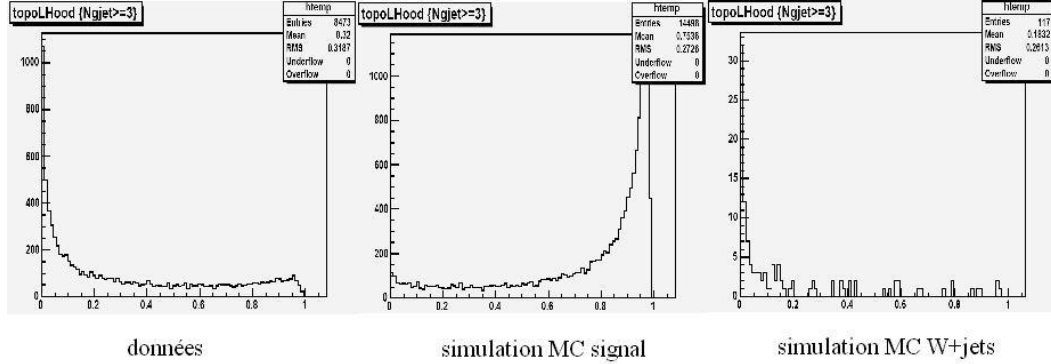


FIG. 5 – Discriminant topologique

Nous venons donc de vérifier sur un exemple avec ces graphiques de contrôle la justesse des hypothèses qui sont les axiomes du programme TopFit. Dans la partie suivante, nous pourrions encore nous apercevoir de l'utilité de ces graphes. Mais avant cela, donnons quelques résultats des programmes TopFit et Util.

TopFit prend en paramètre les courbes de référence de certaines variables topologiques et les efficacités. Pendant le stage, j'ai dû prendre, pour commencer, ces paramètres dans l'ancienne analyse de p14, mais par la suite, j'ai déterminé les nouvelles courbes de référence pour p17. C'est donc avec ces courbes de p17, mais les efficacités de p14 :  $\varepsilon_{QCD} = 15.9\%$  et  $\varepsilon_{signal} = 82.9\%$ , et les trois fichiers de données, simulations signal et W+jets déjà cités, que le programme TopFit a fonctionné et nous pouvons consulter les résultats produits dans le tableau (TAB 2).

TAB. 2 – Résultats du programme TopFit

	nombre de loose	nombre de tight	erreur sur les tight
signal	383.1	317.4	30.4
W+jets	532.1	441.1	43.8
QCD	7577.7	1204.8	15.9

Ici aussi, on vérifie que presque tous les électrons du signal et de W+jets

sont tight, alors que la plupart des (faux) électrons des événements QCD ne sont pas tight. On remarquera également que le nombre d'événement tight pour le signal ( $N_T^{t\bar{t}} = 317.4$ ) est donné à 10% près (erreur de 30.4), ce qui constituerait un progrès dans la précision par rapport à l'analyse p14.

Nous arrivons maintenant au programme Util et aux graphes produits par celui-ci avec les paramètres de TopFit. Tout d'abord, le discriminant topologique (Figure 5), où sont détaillées les trois composantes des données : le signal (16.2%), W+jets (22.6%) et QCD (61.2%), les données étant indiquées par les points noirs et les événements étant tight, ce qui explique leur faible nombre par rapport aux graphiques ci-dessus. La composante QCD est ici beaucoup plus importante que celle du W+jets, et ceci n'est pas normal en comparaison des autres analyses. Cela s'explique peut-être par le fait que les efficacités proviennent de p14 et qu'il faudrait les changer sensiblement, et il faudrait ajouter les poids du trigger, ce dont nous parlerons dans la dernière partie.

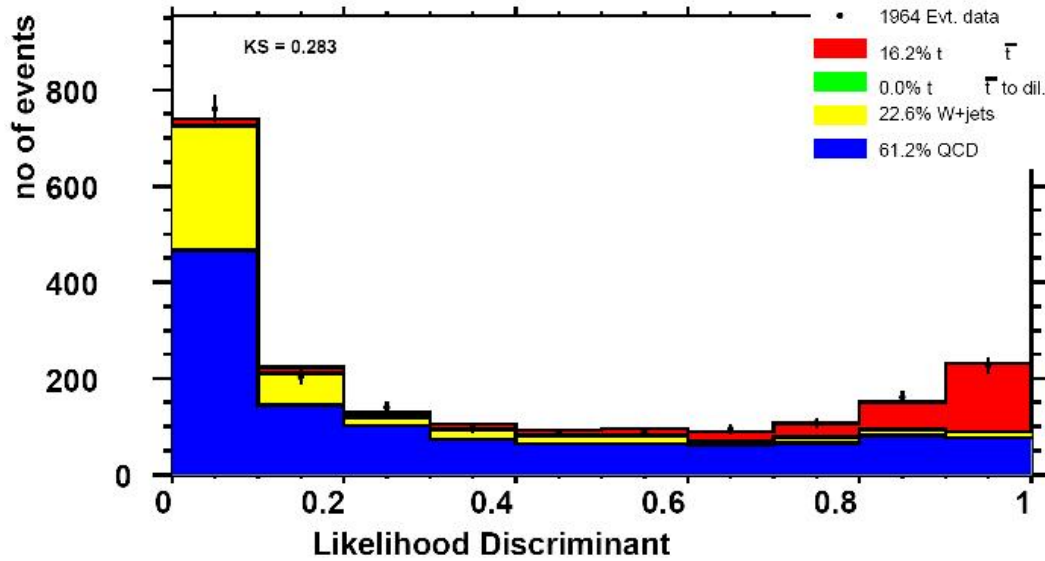


FIG. 6 – Discriminant topologique

D'autres exemples de graphiques d'Util (Figure 6) : la pseudo-rapidité et l'impulsion transverse du premier jet et de l'électron. On remarque sur ces figures les coupures sur l'impulsion transverse imposées pendant les présélections ( $p_T > 20\text{GeV}$ ) pour le jet et l'électron ; on observe aussi que l'électron candidat retenu par les présélections est dans le calorimètre central ( $|\eta| < 1.1$ ), alors que le premier jet n'a pas cette contrainte.

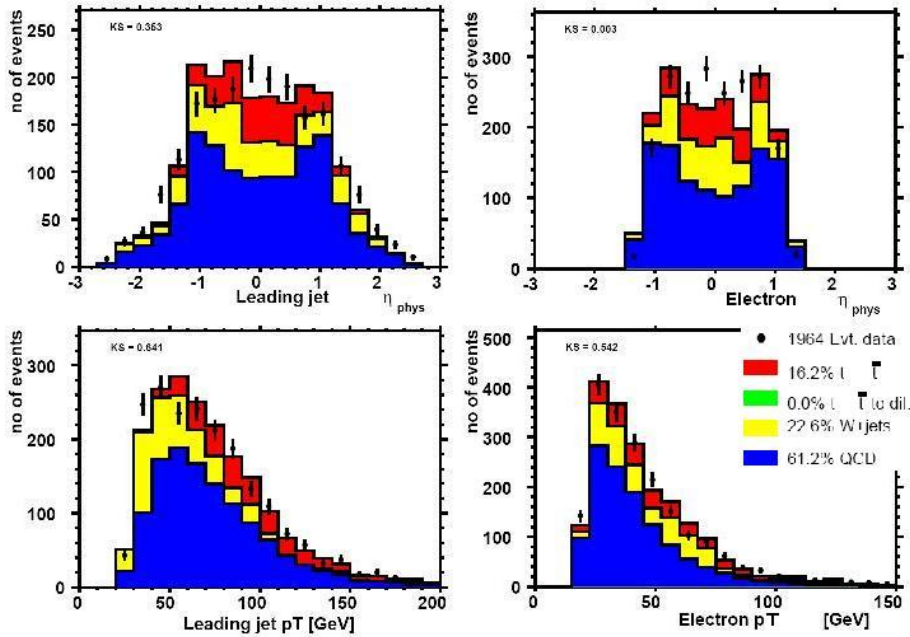


FIG. 7 – Premier jet et électron

## 2.4 Découvertes imprévues

Je vais vous présenter dans cette partie deux découvertes imprévues pendant mon stage : un problème dans le sélecteur de qualité des données, et une incohérence dans le choix de l'électron candidat.

Les graphes de contrôle que j'ai produits nous ont permis de constater qu'il y avait un problème dans le sélecteur de qualité des données. En effet, sur le graphe (Figure (7)), nous observons qu'il y a une asymétrie pour la pseudo-rapidité ( $\eta$ ) et le phi de l'électron. De même, on aperçoit pour les jets une bosse dans l'énergie et un grand pic dans le phi. Ces dissymétries ne devraient pas être présentes. A la demande de ma directrice de stage, j'ai produit d'autres graphiques (Figure (8)) qui nous ont permis de cerner ce bruit de fond comme étant dans le "run range" v12.

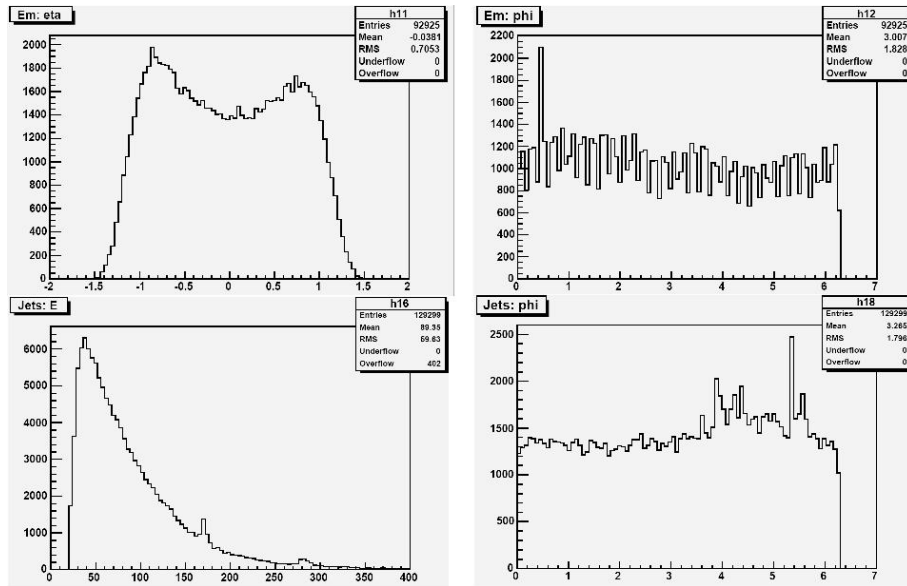


FIG. 8 – Problème de sélecteur de qualité des données

Ce résultat fut communiqué à d'autres membres de la collaboration  $D\phi$ , et bientôt on trouva que le bruit de fond résidait dans trois "runs" : 177976, 177977, 177978, comme l'attestent les graphes (Figure (9)). Ces trois runs, contaminés par un bruit de fond dans le calorimètre, avaient en effet échappé au sélecteur de qualité des données.

Manuellement, je fis donc une sélection en coupant ces trois runs dans le Processeur, et je pus obtenir des graphes "propres" (Figure (10)).

En observant plus attentivement les sélections effectuées sur l'électron candidat par des membres de la collaboration, j'ai pu constater qu'ils le choisissaient de la manière suivante : on exige au moins un électron loose dans le calorimètre central par événement, et on impose un et un seul tight dans le calorimètre central, ce tight étant choisi comme candidat. Ce qui fait que tous les événements du lot QCD étaient rejetés. Cela n'affectait pas l'interface de l'équipe du LPNHE car la sélection était effectuée "à la main", et le programme TopFit, qui a besoin des événements du lot QCD, fonctionnait bien grâce à cela. Cependant, ayant fait remarqué ce souci à la collaboration, j'ai programmé avec l'aide de Jean-Roch Vlimant un sélecteur de type CAF qui choisit l'électron candidat comme il convient : on impose à l'événement d'avoir au moins un électron loose dans le calorimètre central et au plus un tight. S'il y a un tight dans le calorimètre central, il est choisi comme candidat, sinon on prend le loose de plus haute impulsion transverse. Nous voyons ici qu'il peut y avoir des événements passant cette coupure ne possédant pas d'électron tight, et donc appartenant au lot QCD. Ce sélecteur

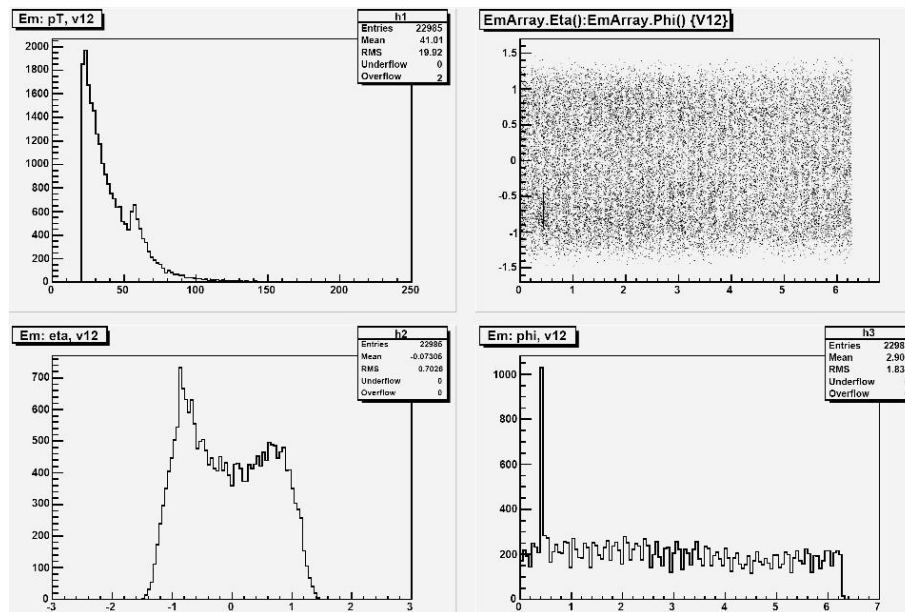


FIG. 9 – Problème dans le run range v12

de type CAF ne modifie pas sensiblement nos résultats, mais, sous réserve de vérification et d'approbation, ayant le format CAF, il servira peut-être à toute personne de la collaboration  $D\emptyset$  qui le souhaite.

## 2.5 Ce qu'il reste à compléter

Lors des présélections intervient un sélecteur sur le trigger. En effet, il s'agit d'enlever les événements des données, incohérents avec la version du système de déclenchement (ou trigger) choisie. Nous avons décrit le système de déclenchement comme rejetant successivement une grande masse d'événements sur trois niveaux de critères. Mais il existe plusieurs versions de ces critères, qui sont plus ou moins sélectives, et le fichier des données expérimentales contient donc des événements passant certaines versions du trigger et pas d'autres. Ainsi il convient de rappeler dans les présélections la version choisie pour ce système de déclenchement et de retirer les événements superflus. Cependant ce tri risque d'induire un biais dans les données. Pour que ce biais n'intervienne pas dans la comparaison entre modèle et expérience, on ajoute des poids statistiques dans les simulations Monte Carlo, appelés poids de trigger, qui simulent le même tri que celui du système de déclenchement sur les données. Il y a un programme qui a été développé dans CAF vers la fin de mon stage, donnant ces poids de trigger. J'ai donc essayé de l'appliquer dans l'interface CAFTrees-TopDumps, mais il reste toujours des problèmes de compatibilité que je n'ai pas eu le temps de résoudre. Ce programme fournirait aussi les luminosités intégrées, qui

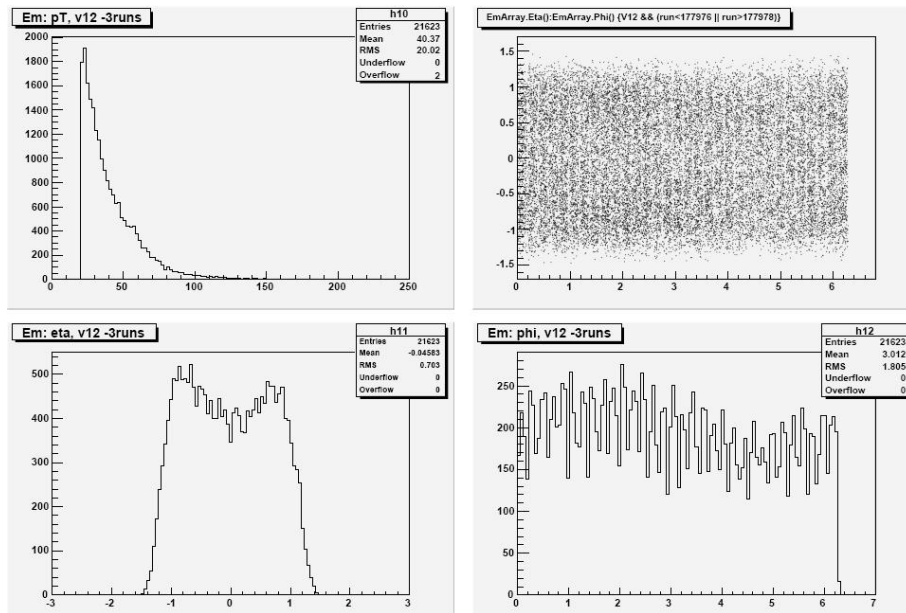


FIG. 10 – Trois runs seulement sont contaminés

interviennent dans le calcul de la section efficace. Ainsi, en mettant à jour l'interface et ce programme, on devrait récupérer les poids de trigger et les luminosités.

Il reste encore à déterminer les efficacités  $\varepsilon_{QCD}$  et  $\varepsilon_{signal}$  pour p17, dont on a déjà dit l'importance pour les résultats de TopFit. Jean-Roch Vlimant a développé au cours de sa thèse des programmes donnant ces efficacités à partir des données et des simulations, mais par manque de temps, je n'ai pas pu les adapter et les utiliser. L'absence des poids de trigger et des efficacités devrait normalement expliquer le problème rencontré (Figure (5)), à savoir que la proportion de bruit de fond réductible QCD est beaucoup plus importante que celle de W+jets. Il sera intéressant de voir, une fois ces choses codées, si la proportion de W+jets devient supérieure à celle de QCD, comme cela devrait se produire.

Il y a aussi d'autres petits problèmes annexes comme celui du sélecteur de vertex, par exemple. Une option de celui-ci, assurant une sélection conjointe des vertex reconstruits et des traces associées, semble ne pas être compatible avec la version de l'interface, elle a donc été supprimée. Il faudrait en toute logique trouver le problème et la rajouter, ce qui modifierait peut-être un petit peu l'efficacité de présélections trouvées.

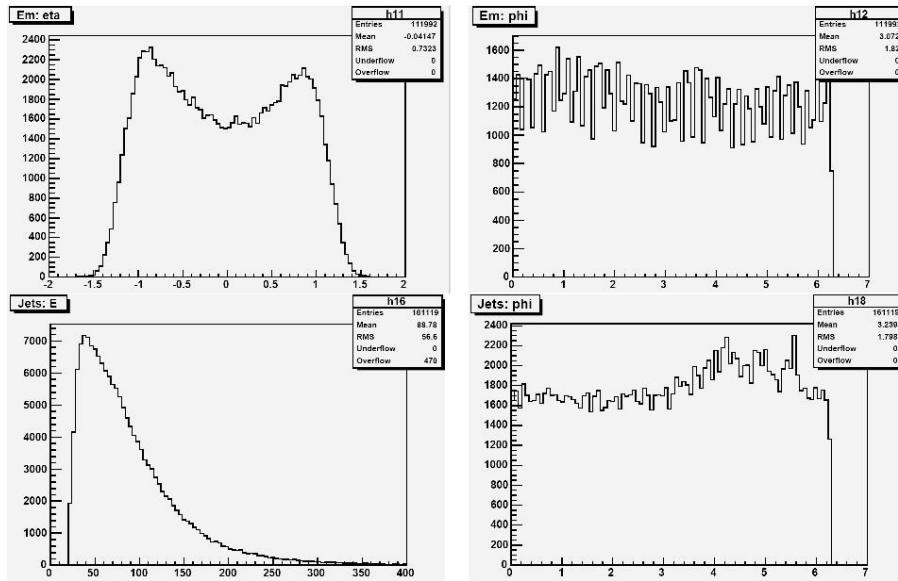


FIG. 11 – Graphes sans contamination

## Conclusion

Ainsi nous avons pu constater, tout le long de ce rapport de stage, dans quelle mesure l'équipe  $D\bar{O}$  du LPNHE se rapproche de l'un de ses objectifs principaux, à savoir la mesure de la section efficace de la désintégration semi-leptonique du quark top, pour p17. En effet, nous avons rappelé quels sont les principes théoriques mis en jeu dans le calcul de cette section efficace. Nous avons décrit comment sont collectées les données expérimentales utilisées par ce calcul, dans le détecteur et dans l'environnement informatique de CAF. Nous avons ensuite commenté l'utilité des différents programmes développés, l'interface sélectionnant les informations nécessaires au calcul de la section efficace, le programme TopFit qui détermine le nombre  $N_T^{t\bar{t}}$  intervenant dans la formule de la section efficace, le programme Util qui permet de contrôler la validité de ces calculs à l'aide d'outils graphiques. Pour utiliser ce programme TopFit, des nouvelles courbes de référence des variables topologiques ont dû être produites pour p17; il reste en revanche à déterminer les efficacités  $\varepsilon_{QCD}$  et  $\varepsilon_{signal}$ , et à inclure les poids de trigger dans le code. Nous avons à notre disposition  $\epsilon$  connu (acceptance du détecteur et efficacité de présélections), même s'il pourrait subir des variations suite à de légères modifications de présélections; nous avons aussi  $BR_{e+jets}$ , et nous avons dit de quelle façon il conviendrait de récupérer la luminosité intégrée. Une fois les modifications énoncées ci-dessus effectuées, l'équipe sera donc en mesure d'exhiber la section efficace recherchée pour p17.

Ce stage m'aura permis de contribuer à la détermination de certains éléments du calcul de cette section efficace, il aura surtout servi à la mise

à jour des moyens de calcul au format informatique CAF, et nous avons déjà souligné à quel point il pouvait être important dans une collaboration d'utiliser les programmes officiels pour la discussion et la comparaison des résultats. Avec ce stage, j'ai eu ainsi un aperçu assez précis de ce qu'est la physique des particules dans des expériences mettant en jeu des accélérateurs géants.