

UNIVERSITE PARIS VI – PIERRE ET MARIE CURIE

ECOLE DOCTORALE DE PHYSIQUE

La Physique de la particule à la matière condensée (ED389)

Doctorat de Physique

PHAM Thi Tuyet Nhung

Contribution à l'étude des grandes gerbes à l'aide du
détecteur de surface de l'Observatoire Pierre Auger

Thèse dirigée par

BILLOIR Pierre, LPNHE, Université Paris VI-UPMC

et

DARRIULAT Pierre, Ha Noi University of Sciences, Hanoi

Soutenue le 18 décembre 2009

Jury:

De KERRET Hervé
NGUYEN Mau Chung
TRAN Minh Tam
URBAN Marcel

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

PHAM Thi Tuyet Nhung

Nghiên cứu mưa rào khí quyển năng lượng siêu cao sử dụng
hệ đo bề mặt của Đài quan sát Pierre Auger

Cán bộ hướng dẫn:

DARRIULAT Pierre, Đại học Khoa học Tự nhiên Hà Nội
và
BILLOIR Pierre, LPNHE, Université Paris VI-UPMC

Ngày bảo vệ luận án: 18/12/2009

Hội đồng chấm luận án:

De KERRET Hervé
NGUYEN Mau Chung
TRAN Minh Tam
URBAN Marcel

Résumé

Ce travail porte sur des observations réalisées à l'aide du détecteur de surface (SD) de l'Observatoire Pierre Auger qui étudie les rayons cosmiques d'énergies supérieures à 10 EeV. Il détecte les grandes gerbes produites dans leur interaction avec l'atmosphère au moyen d'un réseau de 1600 compteurs Cherenkov (CC) qui couvre 3000 km². Les données ont la forme d'un enregistrement digital des temps d'arrivée et des amplitudes des signaux enregistrés par les trois photomultiplicateurs (PMT) de chaque CC. La thèse comporte des études de leurs propriétés, d'une asymétrie observée entre les trois PMT d'un même CC et de la désintégration de muons stoppant dans les CC.

En ce qui concerne la première, les incertitudes qui affectent la mesure ont été évaluées et les différences observées entre les trois PMT d'un même CC ont été identifiées et attribuées à deux causes bien maîtrisées : impulsions retardées et asymétrie de première lumière. Un algorithme de recherche de pics, basé sur la déconvolution de la décroissance exponentielle de la lumière détectée, a été affiné, sa performance évaluée et ses limites identifiées, ouvrant ainsi la voie à son utilisation systématique dans des études ultérieures.

Une corrélation entre l'azimut de la gerbe et l'asymétrie entre les trois PMT d'un même CC, observée avant que la lumière n'ait le temps d'être suffisamment diffusée par les parois, a été mise en évidence et exploitée pour mesurer la divergence de la gerbe et illustrer la puissance de la méthode et sa sensibilité.

Enfin, on a mis en évidence l'existence de muons stoppant à l'intérieur du volume des CC, identifiés par le signal produit par l'électron de désintégration. La difficulté de cette étude réside dans la petitesse des signaux recherchés et permet de mettre à l'épreuve la connaissance qu'on a du détecteur et des outils utilisés pour son analyse. Un bruit de fond de très faible amplitude a été décelé, suggérant la présence vraisemblable de neutrons, une possibilité qui reste à explorer.

Tóm tắt

Luận án trình bày nghiên cứu sử dụng số liệu của hệ đo bề mặt (SD), Đài quan sát Pierre Auger. Đài quan sát này ghi nhận mưa rào khí quyển diện rộng sinh ra do tia vũ trụ siêu năng lượng cao (trên 10 EeV) tương tác với bầu khí quyển. Hệ SD gồm 1600 bình đo Cherenkov nước trải rộng trên diện tích 3000 km². Với mỗi bình đo, thông tin về thời gian và độ lớn tín hiệu được ghi nhận bởi ba ống nhân quang điện (PMT) và được lưu dưới dạng số. Luận án tập trung nghiên cứu đặc điểm của bình đo Cherenkov, tính bất đối xứng tín hiệu giữa các PMT vào thời điểm xuất hiện tín hiệu và phân rã của muon ở trong bình đo.

Nghiên cứu đầu tiên đã đánh giá những yếu tố bất định ảnh hưởng tới phép đo và đưa ra bằng chứng cho thấy sự không đồng nhất xảy ra ở một số thời điểm giữa các PMT của một bình đo là do hiện tượng sau xung và sự bất đối xứng tín hiệu lúc bắt đầu được ghi nhận. Hai hiệu tượng này đều có thể kiểm soát được. Nghiên cứu đã phát triển thuật toán xác định đỉnh tín hiệu dựa trên việc loại bỏ phần suy giảm theo hàm mũ của ánh sáng ghi nhận bởi các PMT đồng thời đánh giá về khả năng cũng như hạn chế của nó, tạo tiền đề cho việc áp dụng phương pháp một cách hệ thống trong các nghiên cứu sâu hơn.

Bất đối xứng tín hiệu xảy ra trước khi ánh sáng phân tán đều do khuếch tán nhiều lần trên thành bình. Nghiên cứu cho thấy hiện tượng này có tương quan với góc tới của trục mưa rào khí quyển và có thể sử dụng để xác định độ phân kỳ của mưa rào, chứng tỏ khả năng và minh họa cho độ nhạy của phương pháp.

Nghiên cứu phân rã muon trong bình đo dựa vào việc xác định tín hiệu của sản phẩm phân rã là electron. Nghiên cứu này đã giải quyết một số khó khăn gây ra do biên độ tín hiệu của electron rất nhỏ, cung cấp thêm một phép đánh giá khả năng hoạt động của bình đo cũng như phương pháp phân tích tín hiệu. Nghiên cứu cũng cho thấy tồn tại một phong nền thấp có thể gây ra bởi các neutron, điều này cần được làm rõ bằng các nghiên cứu sâu hơn.

Contribution to the study of ultra high energy showers using the surface detector of the Pierre Auger Observatory

Summary

The present thesis deals with observations made using the surface detector (SD) of the Pierre Auger Observatory that studies cosmic rays having energies in excess of 10 EeV. It detects the extensive air showers produced by such cosmic rays in their interactions with the atmosphere in an array of 1600 water Cherenkov counters (CC) that covers 3000 km². The information available from the SD is in the form of digitized records of the time of arrival and amplitude of the signals recorded in each CC by three photomultiplier tubes (PMT). The thesis includes studies of their properties, of the early time PMT asymmetry and of the decay of muons stopping in the counters.

Concerning the former, the uncertainties affecting the measurement have been evaluated and evidence has been given that the occasional apparent inconsistencies between the three PMTs of a same CC reduce to only two types, after pulses and early time asymmetries, both of which are under control. A peak finding algorithm consisting in unfolding the exponential decay of the collected light has been refined, its performance has been assessed and its limitations have been identified, opening the road toward its systematic use in further studies.

A PMT asymmetry, occurring before the light has a chance of being randomized by multiple diffusions on the CC walls, has been shown to be correlated with the azimuth of the shower axis, which has been exploited to evaluate the shower divergence, to show the power of the method and illustrate its sensitivity.

Finally, a search for muons stopping in the water volume of the CCs, identified by the signal produced by the decay electron, has overcome the difficulties resulting from their small amplitude and has given an opportunity to assess the detector performance, providing a test of both the detector and the tools available for its analysis. Evidence has been found for a very low charge background that might be associated with neutrons, a possibility that remains to be explored.

Mot-clé: rayons cosmiques d' énergies extrêmes

Từ khóa: tia vũ trụ năng lượng cao

Keyword: ultra high energy cosmic rays

Résumé substantiel en français

Les travaux présentés ici ont été réalisés dans le cadre de la Collaboration Pierre Auger qui exploite l'Observatoire Pierre Auger (PAO) dans la pampa argentine et cherche à répondre à un certain nombre de questions qui n'ont pas encore reçu de réponses satisfaisantes concernant la nature et les propriétés des rayons cosmiques d'énergie supérieure à 1 EeV, dits "d'ultra haute énergie" (UHECR). La construction de l'observatoire a été menée à terme en juin 2008 mais la prise de données a commencé dès janvier 2004 de telle sorte qu'une fois terminée la construction l'observatoire avait déjà accumulé le plus grand ensemble au monde d'observations portant sur de tels rayons cosmiques.

À l'époque où l'observatoire a été conçu, il était possible d'identifier un petit nombre de questions fondamentales auxquelles il convenait de répondre en priorité ; elles ont guidé les choix qui ont présidé à la conception de l'observatoire. Parmi elles on comptait la dépendance en énergie du flux, et en particulier l'étude de l'interaction des UHECRs avec le rayonnement fossile du fond cosmique (CMB) qui devait causer une décroissance rapide du spectre aux plus hautes énergies (coupure GZK); la nature des sources et le mécanisme d'accélération; la nature des UHECRs et, dans l'hypothèse généralement admise qu'il s'agit de noyaux d'atomes ionisés, leur distribution massique.

Afin de répondre à ces questions, le PAO a été conçu comme un détecteur hybride des grandes gerbes atmosphériques induites par les UHECR lorsqu'ils pénètrent dans l'atmosphère terrestre. Il combine deux méthodes de détection très différentes qui se complètent de bien des points de vue : un détecteur de fluorescence (FD) qui mesure le profil longitudinal de la gerbe et un réseau au sol de compteurs Cherenkov à eau (SD) qui mesure le profil transversal à partir de l'empreinte laissée par la gerbe au sol.

Aujourd'hui, le PAO a déjà répondu, pour l'essentiel, à la première de ces questions et a mis en évidence l'effet de la coupure GZK. Au cours de la décennie précédente, les astronomies en rayons X et gamma ont fait d'importants progrès dans l'identification de restes de supernovae comme étant les sources des rayons cosmiques galactiques (au dessous du domaine d'énergie des UHECRs) et dans la mise au clair du mécanisme d'accélération. Toutefois, malgré les percées réalisées par le PAO dans le domaine des autres questions, d'importantes incertitudes demeurent quant à la distribution en masse des UHECRs et l'identification de leurs sources.

La thèse ne s'adresse pas directement à ces questions mais contribue à améliorer notre connaissance du détecteur de surface (SD) de sorte à faciliter et optimiser son utilisation dans les recherches esquissées ci-dessus et, en particulier, dans les études portant sur la distribution massique. Elle est structurée en quatre parties.

Une première partie sert d'introduction et survole l'état actuel de nos connaissances. Un premier chapitre passe en revue les progrès récents de la physique des rayons cosmiques en s'intéressant plus particulièrement aux questions pertinentes à l'étude des UHECRs telles que les avancées récentes en astronomies X et gamma qui ont permis d'identifier comme sources galactiques certains restes de supernovae (SNR) et comme mécanisme d'accélération les passages répétés des particules d'aval en amont et d'amont en aval du front de choc.

Un second chapitre sert d'introduction générale à l'Observatoire Pierre Auger en insistant sur les caractères essentiels du détecteur de surface (SD) qui sont d'une importance particulière pour les travaux présentés dans la thèse. On y trouve également un résumé très bref des mesures de la dépendance du flux sur l'énergie et des progrès accomplis dans l'identification de certaines sources à des galaxies de l'univers proche.

Un troisième chapitre est consacré aux problèmes posés par les mesures de distribution massique qui sont en rapport étroit avec le sujet de la thèse. On y passe rapidement en revue les méthodes utilisées, en particulier celles qui sont basées sur la mesure de l'abondance relative des muons au sol (comparée à celle des électrons, positons et photons), quantité censée permettre de distinguer entre primaires légers et primaires lourds. La quatrième partie de la thèse présente une mesure de l'abondance des muons de basse énergie (à l'arrêt dans le volume d'eau des compteurs Cherenkov où ils se désintègrent). Ces muons de basse énergie sont minoritaires et ne sont pas censés dépendre de la masse des primaires: leur étude permet de vérifier le bien-fondé des modèles hadroniques et des simulations du détecteur dont on dispose indépendamment de la distribution massique.

Une seconde partie étudie les propriétés générales de ce qui constitue la source essentielle des données sur la quelle la thèse se base, les enregistrements appelés "traces FADC". Le réseau de détecteurs au sol du PAO est constitué de cuves d'eau dans lesquelles les particules chargées relativistes de la gerbe produisent de la lumière Cherenkov. Cette lumière est détectée par trois tubes photomultiplicateurs (PMT) de neuf pouces dont les signaux sont enregistrés par des convertisseurs analogue digital flash (FADC) dont la

résolution en temps est de 25 ns et celle en charge de 0.5% du signal produit par un muon relativiste d'incidence verticale. L'information que contiennent ces traces est à la fois très riche et très difficile à décoder. L'apport essentiel de la thèse est de contribuer à en faciliter l'interprétation et l'usage.

Un premier chapitre porte sur la fiabilité de l'information disponible, en particulier sur la compatibilité des signaux enregistrés par les trois PMT d'un même compteur. L'étude est abordée sans préjugé ni a priori quant aux causes de possibles incohérences et conclut que les seuls désaccords existants, à l'exclusion de compteurs proches du cœur de la gerbe où les PMT sont surchargés, ont deux causes bien connues: l'occurrence éventuelle de post impulsions individuelles et une asymétrie de réponse entre les trois PMT, dépendant de l'incidence des particules qui traversent le détecteur. Cette dernière cause est en rapport étroit avec la matière de la troisième partie de la thèse. Les compteurs du SD constituent d'excellentes cavités optiques et la lumière Cherenkov subit de nombreuses diffusions sur leurs parois avant de s'échapper vers la photocathode d'un des trois PMT. Ces diffusions ont pour effet de rapidement égaliser les réponses des trois PMT. Cependant, dans les premiers quelque 30 ns, soit moins que deux bins, cette égalisation n'est pas complète (le temps séparant deux diffusions successives est en moyenne de 10 ns) et le PMT qui se trouve être le mieux exposé à la lumière Cherenkov enregistre un signal supérieur à celui des deux autres.

Un second chapitre s'emploie à réduire les traces FADC en une somme de pics associés à des particules individuelles. On y exploite l'excellente qualité optique des compteurs qui permet de soustraire l'effet de la décroissance exponentielle de la lumière détectée (produite presque instantanément mais diminuant avec un temps de relaxation de l'ordre de 75 ns).

Une troisième partie concerne l'asymétrie dont on vient de parler, entre les trois PMT d'un même compteur, et l'exploite pour étudier la divergence des gerbes. Elle débute par une introduction à la méthode et aux calculs qu'elle implique et se poursuit en mettant en évidence l'existence d'une forte corrélation entre l'asymétrie et la direction d'incidence des particules sur le compteur. En combinant les informations associées à tous les compteurs d'une même gerbe, on peut évaluer la divergence moyenne de la gerbe ou, ce qui revient au même, l'altitude moyenne de la source le long de l'axe de la gerbe. La grande quantité de données disponibles réduit considérablement les incertitudes statistiques et la

sensibilité de la méthode est illustrée par une étude de la dépendance de la divergence sur divers paramètres caractéristiques des propriétés de la gerbe.

Une quatrième et dernière partie est consacrée à l'étude des muons de basse énergie qui appartiennent aux grandes gerbes et sont ralentis et stoppés à l'intérieur des compteurs où ils se désintègrent. Une telle étude implique la définition d'un ensemble de critères permettant d'identifier et sélectionner des signaux associés aux électrons et positons produit lors de la désintégration des muons. Ce sont des signaux de faible amplitude, difficiles à bien mesurer. Une fois sélectionné un échantillon d'électrons candidats, un certain nombre de sources possibles de bruit de fond sont identifiées et soustraites. Une troisième étape consiste à mesurer le temps de vie des muons, ce qui implique de faire des hypothèses sur la distribution des temps auxquels les muons mères se sont arrêtés. Finalement, les résultats sont comparé aux prédictions de simulations. Bien qu'un accord général soit le résultat dominant, un signal de bruit de fond de très faible amplitude, absent des simulations, est mis en évidence, suggérant la présence d'une composante neutronique. L'intérêt de cette étude réside dans le fait qu'elle exploite les qualités du détecteur jusqu'à ses limites et permet de ce fait d'acquérir une confiance accrue en sa fiabilité.

La thèse se termine par un bref résumé et quelques considérations portant sur les voies nouvelles qu'elle a permis d'ouvrir dans les directions qu'elle a explorées.

Contents

Preamble

1. Ultra high energy cosmic rays and the Pierre Auger Observatory	17
1.1 Generalities on cosmic rays	17
1.1.1 A brief history	17
1.1.2 The main features	19
1.1.3 Galactic sources	20
1.1.4 Diffusive shock acceleration	22
1.1.5 Extra galactic sources	24
1.2 The Pierre Auger Observatory	27
1.2.1 General description	27
1.2.2 The surface detector	29
1.2.3 Cherenkov tanks	32
1.2.4 Simulations	34
1.2.5 Energy spectrum and the GZK cut-off	35
1.2.6 Correlations with astronomical sources	36
1.3 Identification of the primaries	39
1.3.1 General considerations	39
1.3.2 Longitudinal profiles	40
1.3.3 Risetime	41
1.3.4 Muon abundance	44
1.3.5 Summary	45
2. FADC traces	47
2.1 General features	47
2.1.1 Noise	47
2.1.2 Base-line	49
2.1.3 Photoelectron statistics	53
2.1.4 After-pulsing	57
2.1.5 Spikes	62
2.1.6 Summary	67
2.2 Pattern recognition	69
2.2.1 Introduction	69
2.2.2 Preliminary data reduction and selection	70
2.2.3 Subtraction algorithm	71
2.2.4 Early time asymmetries	76
2.2.5 Muons	79
2.2.6 Adjacent signals	81
2.2.7 Summary	83
3. PMT asymmetries and shower divergences	84
3.1 Introduction	84

3.1.1 Motivation	84
3.1.2 Arithmetics	84
3.2 Overview of the method	87
3.2.1 Azimuth-asymmetry correlation.....	87
3.2.2 Shower divergence	90
3.3 Intrinsic asymmetry	93
3.3.1 Introduction	93
3.3.2 Results	94
3.4 Single vertex approximation.....	96
3.4.1 Dependence on energy and zenith angle	96
3.4.2 Dependence on other parameters	99
3.5 Summary	103
4. On the decay of muons stopping in the SD tanks	104
4.1 General considerations.....	104
4.1.1 Time range.....	104
4.1.2 Muon energies	105
4.1.3 Muon lifetime	105
4.1.4 Electron signals	106
4.1.5 Scope of the study and data sample.....	106
4.1.6 Simulated data	108
4.1.7 General outline	109
4.2 Selection criteria	110
4.2.1 Baseline correction	110
4.2.2 After-pulses	111
4.2.3 Charge cuts	112
4.2.4 Trace-by-trace definition of the late region.....	115
4.3 Background sources	120
4.3.1 Simulated data	120
4.3.2. Real data	126
4.3.3. Systematic uncertainties	132
4.3.4. Nucleon background.....	133
4.4 Counting stopping muons	136
4.4.1 Method.....	136
4.4.2 Validation of the method.....	139
4.5 Results.....	140
4.6 Summary and conclusions	141
5. Summary and perspectives.....	143
References	145

Acknowledgement

Preamble

The work presented here has been performed within the framework of the Pierre Auger Collaboration that operates the Pierre Auger Observatory (PAO) in the Argentinean pampas and aims at answering a number of open questions concerning the nature and properties of cosmic rays having energies in excess of 1EeV , which are referred to as ultra high energy cosmic rays (UHECR). The construction of the Observatory was completed in June 2008 but it started taking data as soon as January 2004 and, by the time of completion, had already accumulated the world's largest data set of cosmic ray observations

At the time of conception, a few major questions could be singled out as having to be addressed in priority and as governing the main options chosen for the design. These included the energy dependence of the flux, and in particular the study of the interaction of UHECRs with the cosmic microwave background (CMB), expected to cause a steep decrease in the energy spectrum (referred to as GZK cut-off); the nature of the sources and of the mechanism of acceleration; the nature of UHECRs and, under the generally accepted hypothesis that they are ionized nuclei, their mass composition.

In order to answer such questions, the PAO was conceived as a hybrid detector of the atmospheric showers induced by UHECRs penetrating in the Earth atmosphere. It combines two very different methods of detection which complete each other in many respects: a fluorescence detector (FD) measuring the shower longitudinal profile and a ground array of water Cherenkov counters (SD) measuring the shower transverse profile on ground.

Today, the PAO has essentially answered the first of the above questions and given evidence in favour of the GZK cut-off. During the past decade, X-ray and gamma-ray astronomy have made important progress at identifying sources of galactic cosmic rays (below the UHECR energy range) and at elucidating the acceleration mechanism. However, in spite of major progress achieved by the PAO toward answering the latter questions, major uncertainties still remain today on the mass composition of UHECRs and on the identification of their sources.

The present work does not address directly these questions but contributes to the understanding of the performance of the surface detector (SD) in a way that should help

making optimal use of it, in particular in mass composition studies. It is organized in four parts.

A first part introduces the subject by giving an overview of current knowledge. A first chapter reviews in very broad terms the general status of present days cosmic ray research with particular emphasis on topics of relevance to the study of UHECRs such as recent progress in X-ray and gamma-ray astronomy, leading to the identification of Super Nova Remnants (SNRs) as galactic sources and of diffusive shock acceleration as the associated acceleration mechanism.

A second chapter gives a general introduction to the Pierre Auger Observatory, with emphasis on the essential features of the SD, which are of particular relevance to the present work. It also reviews very briefly the measurement of the energy spectrum, the evidence for the GZK cut-off and the progress made toward identifying sources with nearby galaxies.

A third chapter focuses on the problem of measuring the mass composition, which is of more direct relevance to the present work. It reviews very briefly some of the methods being used, in particular those which rely on the relative muon population on ground (compared to electrons, positrons and photons), a quantity expected to be a discriminant between light and heavy primaries. Part 4 of the present work studies low energy muons (stopping and decaying within the volume of the SD water Cherenkov counters). Such low energy muons are a minority and are not supposed to depend on the mass of the primary: their study provides a check of both hadronic models and detector simulation independent from the mass composition.

A second part studies the main features of the essential source of information on which the present work is based, the so called FADC traces. The ground PAO array is made of water Cherenkov counters detecting the Cherenkov light emitted by fast shower charged particles crossing the water volume. The light is detected by three 9" photomultiplier tubes (PMT), the signals of which are recorded by flash analog to digital converters (FADC) with a bin size of 25 ns and a charge resolution of 0.5% of the signal produced by a vertical relativistic muon. The information contained in such traces is both extremely rich and extremely difficult to disentangle. The main contribution of the present work is to progress toward being able of making better use of it.

A first chapter addresses the question of the reliability of the available information, in particular by evaluating the consistency between the FADC traces of individual PMTs. The problem is tackled without any a priori presumptions on the nature and the cause of possible inconsistencies and concludes that, far enough from the shower core where PMTs are overloaded, the only significant inconsistencies are due to two well known sources: the occurrence of after pulses in individual PMTs and an asymmetry between the three PMT responses depending upon the angle of incidence and impact of the detected particles. The latter is of direct relevance to Part 3 of the present work. The Cherenkov SD tanks happen to have excellent optical properties and the Cherenkov light makes many diffusions on the tank walls before being absorbed or escaping in the photocathode of one of the PMTs. Such diffusions randomize the light and equalize the PMT responses. However, in the first 30 or so ns, i.e. less than two FADC time bins, randomization is not yet complete (a typical light path from wall to wall takes 10 ns) and the PMT that happens to be most efficiently illuminated records a larger signal than the others.

A second chapter attempts at resolving the FADC traces as sums of individual peaks associated with individual particles. It makes use of the already mentioned excellent optical properties of the Cherenkov tanks which make it possible to unfold the exponential decay of the detected light (produced almost instantaneously but decreasing with a characteristic decay time of some 75 ns).

A third part studies the previously mentioned asymmetry between the three PMT of a same tank and makes use of it to evaluate the shower divergence. It starts with an introduction to the method and to the relevant arithmetics and goes on by providing evidence in favour of a strong correlation between the direction of incidence of the particles on a tank and the PMT asymmetry. Combining the information associated with all tanks of a same shower makes it possible to evaluate its average divergence or, equivalently, the average altitude of the source. The large size of the available data sample allows for very small statistical errors and the sensitivity of the method is illustrated by a study of the dependence of the location of the average altitude on parameters that characterize the shower.

A fourth and last part studies low energy muons (part of a UHECR shower) stopping and decaying in the SD Cherenkov tanks. Such a study implies a set of criteria allowing for the selection of FADC signals associated with electron produced by the decay

of stopping muons. These are small signals, difficult to measure. Once such signals have been selected as decay electron candidates, various sources of background are identified and subtracted. In a third step, the decay time distribution is measured, implying an educated guess of the time at which the parent muon had stopped. Finally the results are compared with expectation. The interest of the study is in its pushing quite far the limits of the detector performance and therefore, inasmuch as it succeeds, in its giving increased confidence in its reliability.

The report closes with a brief summary and some considerations on possible continuations along the roads open by the work reported here.

1. Ultra high energy cosmic rays and the Pierre Auger Observatory

1.1 Generalities on cosmic rays

1.1.1 A brief history

At the end of the XIXth century, scientists were puzzled by the spontaneous discharge of their electroscopes, suggesting that some kind of an ionizing radiation was present on Earth. In 1909, Wulf took his electroscope on top of the Eiffel Tower, suspecting Earth radioactivity – that had been recently discovered – to be the cause. However, he noted that the discharge rate was not decreasing with altitude as fast as he had expected, suggesting the presence of a downward component [1]. Between 1911 and 1913 the Austrian physicist Viktor Hess (Figure 1.1) performed balloon measurements reaching up to five kilometres in altitude and established the existence of an “unknown penetrating radiation coming from above and most probably of extraterrestrial origin” [2]. He shared the 1936 Nobel Prize with Carl Anderson.

In the following years cosmic rays became the subject of intense research, in particular with Millikan (who coined the name in 1925) and Anderson at Pikes Peak. In 1927 the measurement of the east-west asymmetry and of the dependence of the rate on latitude established unambiguously that cosmic rays were charged particles, not photons [3]. In 1938, Pierre Auger (Figure 1.1), using counters in coincidence, discovered extensive air showers (EAS) and understood that they were produced by very high energy (up to at least 10^{15} eV) primaries interacting with the Earth atmosphere [4].

In the thirties and forties, when accelerators were not yet dominating the scene, cosmic rays became the laboratory for the study of particle physics. Anderson (Figure 1.1) discovered the positron [5] in 1932 and the muon [6] in 1936. Powell and Occhialini discovered the pion [7] in 1947. Then came strange particles: kaons, hyperons and many others. In the fifties, accelerators took over and cosmic rays got studied for their own sake.

For many years following, major effort was devoted to the study of cosmic rays, trying to understand their origin [8]. Ground detectors, large arrays and fluorescence

telescopes, reached very high energies (John Linsley at Volcano Ranch saw the first 10^{20} eV shower [9] in 1962). Space astronomy has been a break through for the study of low energy cosmic rays, in particular solar energetic particles. A recent example of space measurements in solar astronomy is the NASA's Advanced Composition Explorer which was launched in 1997 to the Lagrange point between Sun and Earth.

In the past 20 years, spectacular progress in astrophysics and long time scales implied in the construction of very high energy accelerators have caused a renaissance of interest in cosmic rays physics under the name of astroparticle physics. In particular TeV gamma ray detectors have been constructed and operated. Their main asset is that they can point to the sources without suffering deflections from magnetic fields.

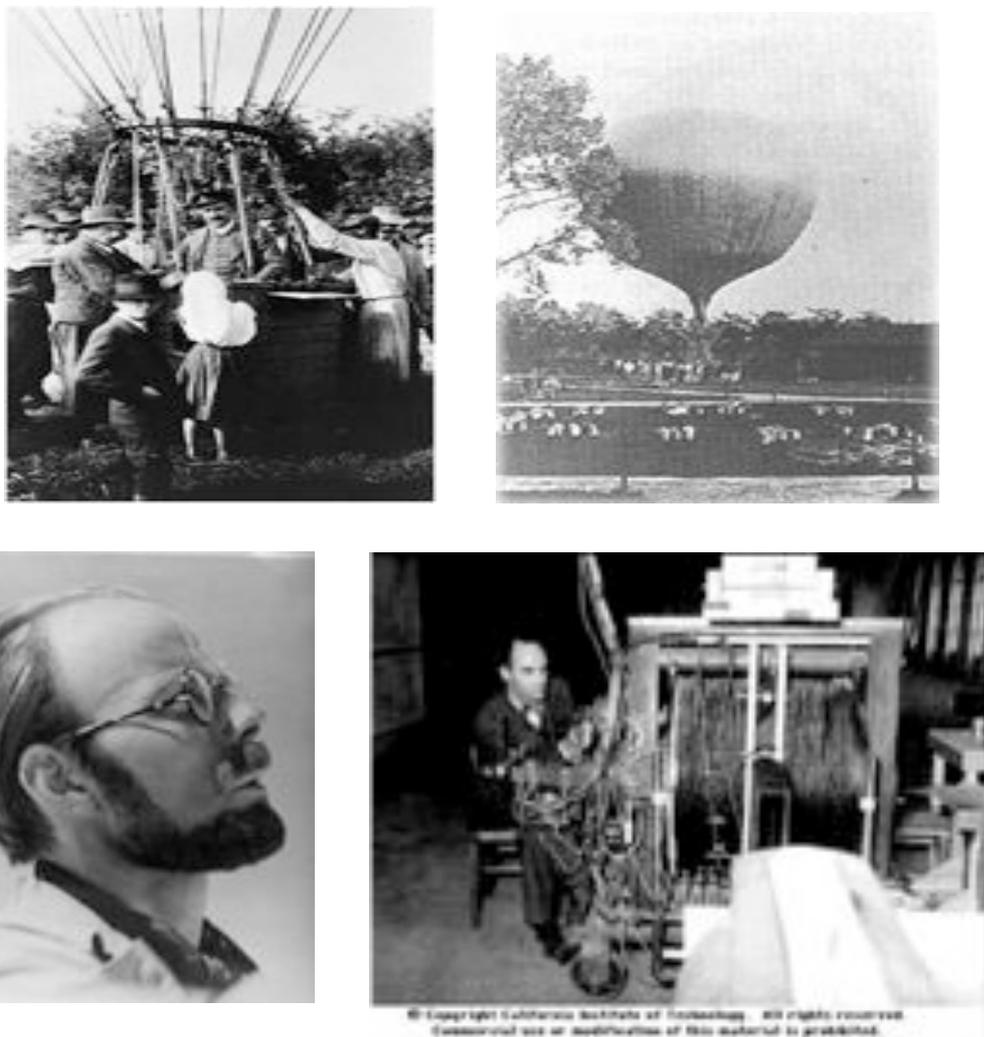


Figure 1.1 The pioneers: Viktor Hess and his balloon (upper panels), Pierre Auger at the Jungfrauoch (lower left), and Anderson with his cloud chamber (lower right).

To study cosmic rays, a new generation of ground detectors was born. Plans to use the whole Earth atmosphere as a radiator observed from space are being implemented and neutrino astronomy is currently being pioneered.

1.1.2 The main features

Cosmic rays are ionized nuclei that travel in space up to extremely high energies of the order of 10^{20} eV=16 Joules! There are very few of them but their contribution to the energy density of the Universe is similar to that of the CMB or of the visible light or of the magnetic fields, namely ~ 1 eV/cm³. Their power law energy spectrum (Figure 1.2), spanning 32 decades (12 decades in energy), is of the approximate form $E^{-2.7}$ [10].

Whenever they have been measured, cosmic rays abundances are similar to elemental abundances observed in their environment, suggesting that they have been accelerated from interstellar matter. As in any galactic environment, hydrogen and helium dominate, even-even nuclei are naturally favoured and the iron region, which corresponds to the strongest nuclear binding, is enhanced. The main difference is that the valleys are now filled by spallation reactions on the matter encountered by the cosmic ray during its journey in the interstellar medium, ~ 7 gcm⁻² on average.

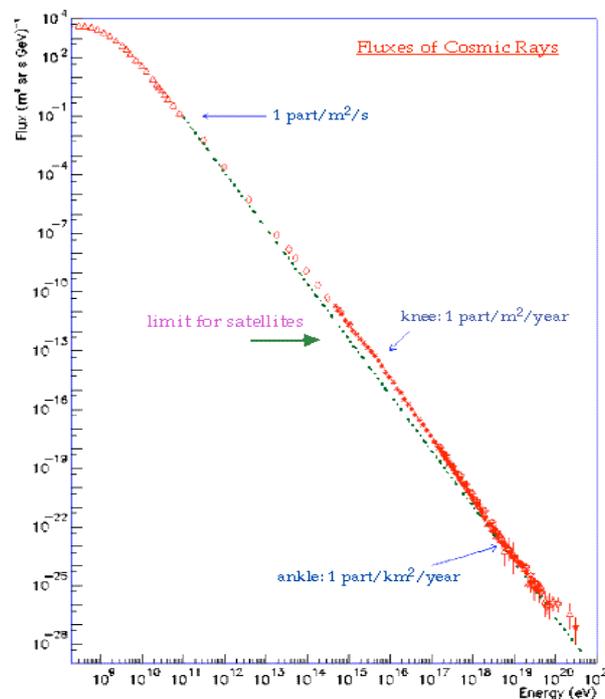


Figure 1.2. The cosmic ray energy spectrum displaying its main features.

While the very low energy part of the cosmic rays spectrum is of solar origin, most of it does not reach the Earth, which is shielded by its magnetic field. The bulk of the energy spectrum on Earth corresponds to an energy density of $\sim 10^{-12}$ erg/cm³. Most of it must have a galactic origin because of the magnetic trapping in the Milky Way disk with a galactic escape time of $\sim 3 \cdot 10^6$ y. The cosmic rays power amounts therefore to some $\sim 10^{-26}$ erg/cm³s which can be compared with the power delivered by SN explosions, $\sim 10^{-25}$ erg/cm³ ($\sim 10^{51}$ erg/SN and ~ 3 SN explosions per century in the disk). Namely cosmic rays carry some 10% of the power delivered by SN explosions [8].

It is only in the higher energy part of the spectrum that an extra galactic component can be found. Its energy density is estimated to some $2 \cdot 10^{-19}$ erg/cm³ implying a power of $\sim 10^{37}$ erg/Mpc³/s [8]. Both active galactic nuclei (AGN) and gamma ray bursts (GRB) stand, from the point of view of energy, as possible sources.

1.1.3 Galactic sources

Particles coming from the Sun reach up to a few MeV and are mostly associated with solar activity and flares. Coronal mass ejections and resulting interplanetary shocks are similarly correlated. On the contrary, galactic cosmic rays are anticorrelated as solar activity increases the Earth magnetic field, which acts as a shield.

Contrary to cosmic rays, gamma rays travel straight in the universe and point back to their sources. They are good at detecting the high energy decay photons coming from neutral pions produced in the interaction of very high energy cosmic rays with interstellar matter. Gamma ray astronomy (Figure 1.3) has shown that several sources have an X ray counterpart identified as an SNR (Figure 1.4) and has established this way that most galactic cosmic rays are likely to originate from SNRs.

There exist two main types of SNRs: Ia and II. Type Ia occurs when a white dwarf, member of a binary, accretes matter from its companion until it reaches Chandrasekhar mass limit of 1.4 solar masses. The core is fully burned; the SNR shell is nearly empty. Type II occurs when a massive star collapses into a neutron star that remains in the centre, possibly detected as a pulsar, the wind of which gives energy to the remnant (one speaks of a plerion).



Figure 1.3: The High Energy Stereoscopic System (HESS, Namibia) [11] includes four telescopes at the corners of a $120 \times 120 \text{ m}^2$ square, operating above 100 GeV. Its field of view is 5° and its resolution a few arc minutes. To take a picture of the Crab takes only 30 seconds.

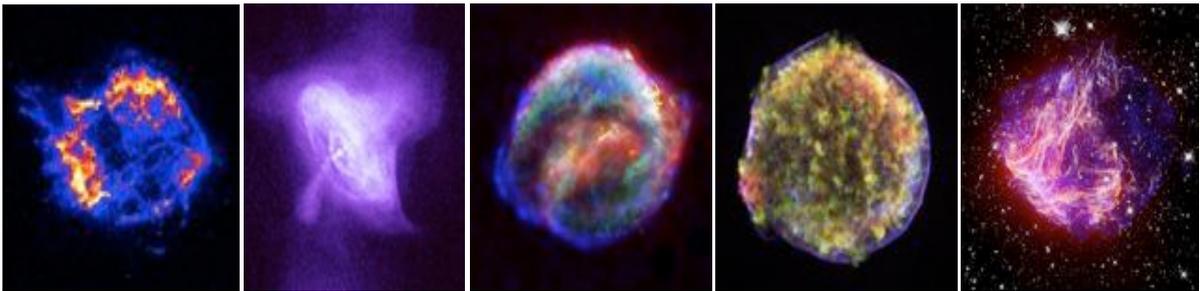


Figure 1.4. Very high resolution X ray images of SNRs (Chandra) [12].
From left to right: Cassiopeia A, the Crab, Kepler (SN 1604), Tycho (SN 1572) and N49.

Figure 1.5 illustrates the correlation observed between high energy γ rays and X-rays emitted by an SNR source [13], which establishes that they come from the shell. The main features of SNR shell structures are reasonably well understood: the explosion blast wave sweeps up the inter-stellar matter (ISM) in the forward shock. As mass is swept up, the forward shock decelerates and ejecta catch up. Then, the reverse shock heats the ejecta and nuclear reactions produce new heavy elements. Once enough mass has been swept up the SNR enters the so-called Sedov phase and slowly dilutes in the ISM. While thermal particles and magnetic field are concentrated in the shell, relativistic particles extend to much larger distances and synchrotron emission is confined to magnetic field regions. The shock structure depends on the SNR age.

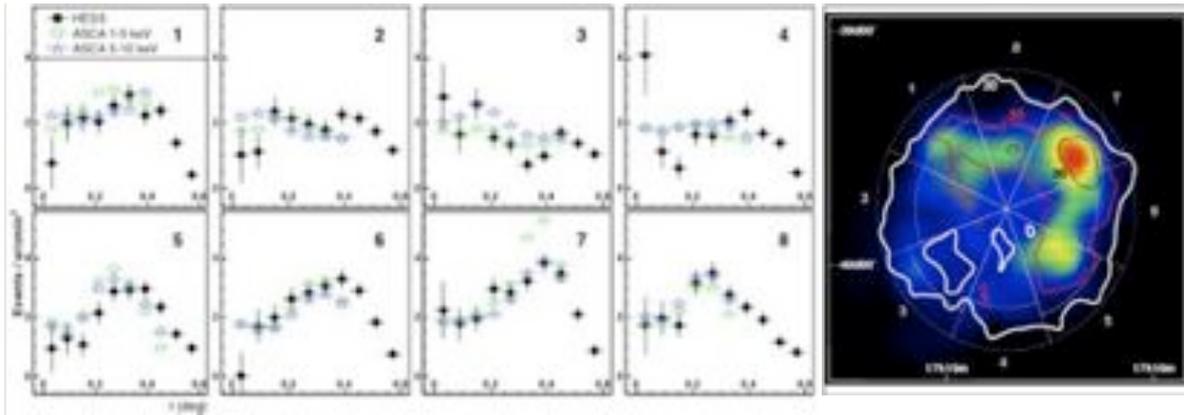


Figure 1.5. Comparison of radial intensity profiles measured in X-rays (ASCA) and γ rays (HESS) in separate octants of SNR RX J1713. The overall correlation coefficient between the two radial distributions is 80%.

1.1.4 Diffusive shock acceleration

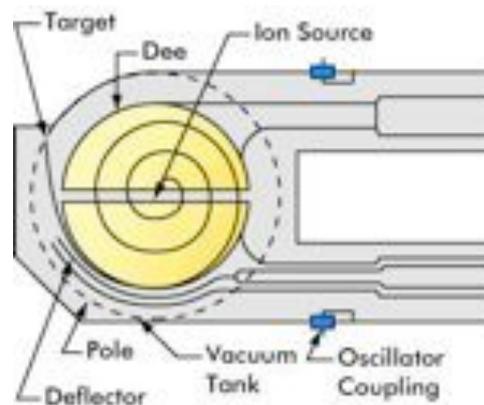


Figure 1.6. Principle diagram of a cyclotron.

The identification of SNRs as sources of galactic cosmic rays has suggested an acceleration mechanism, called diffusive shock acceleration [14], which is now accepted as the most likely candidate for accelerating cosmic rays [15]. The situation is reminiscent of what happens in a cyclotron (Figure 1.6) where the particle is accelerated locally in the gap between the cyclotron dees and is guided by magnetic fields on either side. However both the acceleration and guiding processes are very different from the cyclotron case. Both upstream and downstream of the shock front, are microGauss to milliGauss magnetic fields that are essentially frozen in very diluted plasmas aiming at each other at very high relative velocities. These are not uniform – as in a cyclotron – but are stochastic and significantly enhanced by turbulences and plasma waves that are themselves partly generated by the

cosmic ray stream. Yet, a fraction of the cosmic rays ultimately return to the front without having lost energy. The lower the cosmic ray energy, the larger the probability that they return to the front. At some high energy they will not return and their escape probability is an essential parameter in fixing the shape of the finally energy spectrum.

In a reference frame at rest in the medium where the particle happens to be, whether upstream or downstream the shock, there is no acceleration. This is sometime expressed by talking of "acceleration by change of reference frame". It is somewhat misleading, however: in a reference frame at rest in the medium where the particle happens not to be, there is continuous acceleration from the electric field generated by the moving magnetic field.

Let V_{shock} be the relative velocity of the upstream and downstream media. Each time a cosmic ray particle returns to the shock, it has the same energy as when it last left the shock (bending in a magnetic field does not cause any energy loss and the ISM density is so low that collisions can be neglected). Hence the cosmic ray particle acquires an energy ΔE at each traversal of the shock front with $\Delta E/E = V_{shock}/c$ where c is the light velocity. The time between successive encounters is $\Delta t = kE$, with k a constant, and the escape probability, marking the end of the acceleration process, is equal to V_{shock}/c (as is $\Delta E/E$). Calling r the shock compression ratio (the ratio between upstream and downstream densities), the energy spectrum takes the form $dN/dE \approx E^{-\alpha}$ with $\alpha = (r+2)/(r-1)$. For monatomic gases, $r=4$ and $dN/dE \approx E^{-2}$. The prediction of a power spectrum, with an index not too different from that observed, is a major success of the model.

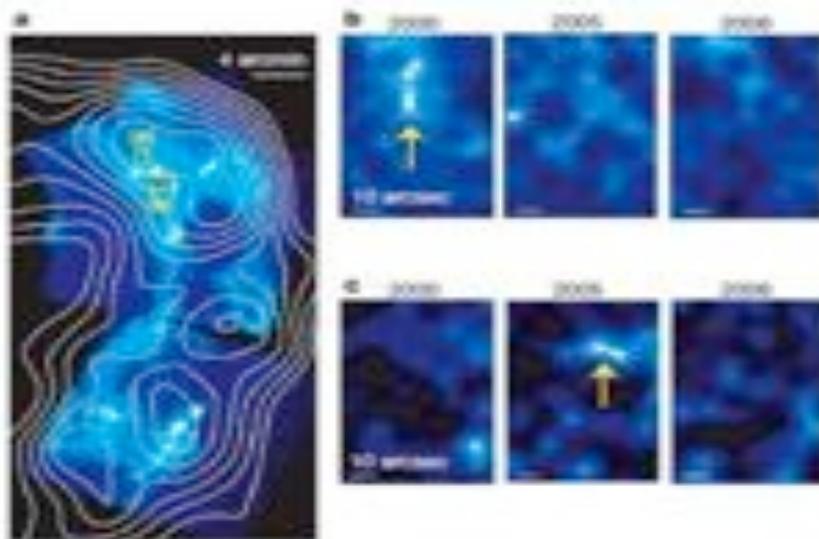


Figure 1.7 Evidence for time varying turbulences in the shell of RX J1713.

Quantitatively, good results have been obtained after it had been realized that the magnetic fields in the shock region are much stronger than was originally thought. There exists indeed copious evidence in favour of strong magnetic turbulences and magnetic field amplification in the shock region of young SNRs. For example, RX J1713 (Figure 1.7) shows a variable shock structure implying strong turbulences and magnetic field amplification [16]. Important variations are detected as a function of time, zones of turbulence becoming quiet and conversely on a few years time scale. Evidence for magnetic field amplification is obtained from the ratio of radio to TeV emission as a same distribution of electrons produces synchrotron (radio, X-ray) and TeV Inverse Compton (IC) but synchrotron depends directly on field while IC and pion decays do not. Shock front compression is a revelator of field amplification. Magnetic fields are enhanced by factors of up to hundred, much larger than the factor of 4 associated with the compression factor of an ideal hydrodynamic shock. For example, in Cass A, one observes a strong front compression implying a magnetic field level of 500 μG instead of the 10 μG expected otherwise [17].

Cosmic rays and the magnetized plasma carry similar energy densities: they do interact on each other. Accelerated particles tend to stream ahead upstream, which causes the generation of streaming instabilities and makes the evolution non linear, resulting in a strong amplification of the mean field: the structure of the shock is modified by cosmic ray retroaction. The higher field, in turn, depresses IC with respect to synchrotron emission, implying faster scattering and increased maximum momentum.

Sharply peaked X-rays at forward shock are evidence that the field is large and increases sharply at the shock, implying that diffusive shock acceleration is efficient and nonlinear at SNR outer blast wave shocks. Older remnants do not show such field amplification: The excitation of turbulences decreases with shock velocity, while damping (by non-linear wave interactions and ion-neutral collisions) does not.

1.1.5 Extra galactic sources

Only the higher energy UHECRs are expected to point to their sources within a few degrees but this expectation rests on the assumption that the magnetic fields that they meet on their journey to the Earth are small enough. Little is known of these fields except that galactic fields are at μG scale while extragalactic fields are expected to be smaller, by an

order of magnitude or so. Obvious geometric arguments imply that nearby fields, i.e. galactic fields, cause much more deviation of the apparent direction of the source than extragalactic fields of the same amplitude. If sources could be reliably identified, their uniform distribution in the sky would then be evidence for their extragalactic origin (otherwise they would cluster around the disk of the Milky Way). Of relevance is the mass composition of cosmic rays: in a same magnetic field, the trajectory of a fully ionized iron nucleus is 26 times more bent than that of a proton.

Very general arguments limit the possible UHECR acceleration sites to a very few. The argument is that, whatever the acceleration mechanism (it is obviously true in the case of diffusive shock acceleration), the product of the size of the site by its mean magnetic field must exceed some value to contain the orbits [18,19,15]. This is illustrated in Figure 1.8, the so-called Hillas plot, that shows as possible acceleration sites AGNs, their jets and radio lobes, GRBs, magnetars (neutron stars having extremely high magnetic fields) and colliding galaxies. For the mechanism of diffusive shock acceleration to be valuably extended to extragalactic cosmic rays, one needs much larger shock sites than provided by SNRs, such as present in colliding galaxies (Figure 1.9).

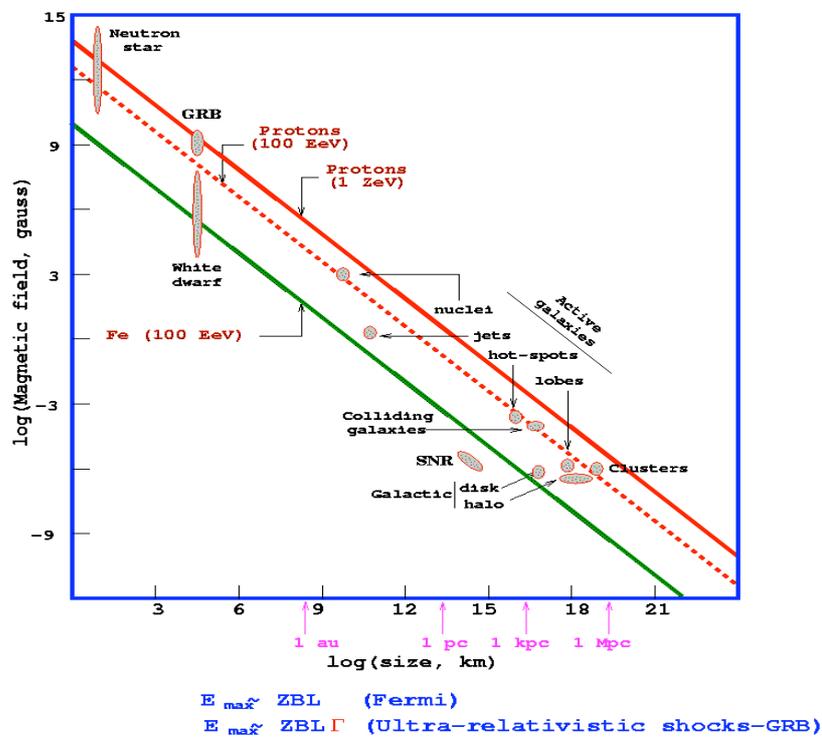


Figure 1.8. Hillas plot for protons (red lines) at 10^{20} eV (dashed) and 10^{21} eV (full). The green line is for 10^{20} eV iron. The decimal logarithm of the size (km) is in abscissa and that of the field (Gauss) in ordinate.

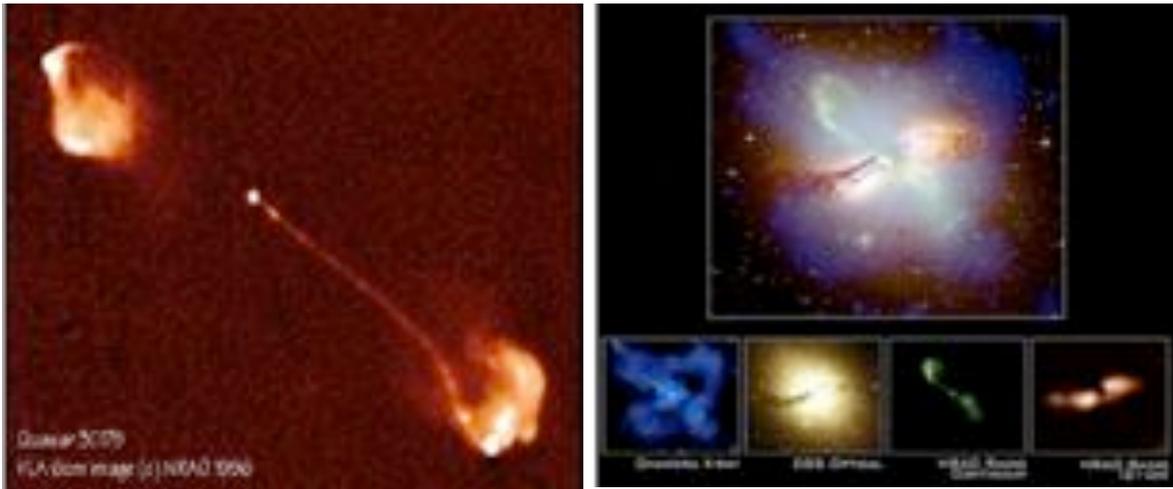


Figure 1.9. Left: Radio image of a quasar. Right: Centaurus A, merging of an elliptical galaxy with a smaller spiral, has an AGN in its centre (the AGN closest to us) and is the site of large shocks.

Recent observations and studies of colliding galaxies and merging galaxy clusters [20] suggest that these were common phenomena in the early denser Universe. Such collisions are now believed to have played an important role in the process of galaxy formation. Galaxy collisions usually do not imply direct star collisions but the strongly increased gravity field enhances the collapse of hydrogen clouds and the formation of new stars, many of which being very massive and therefore having a short lifetime. Galaxy collisions are sites of very violent events on large scales and are therefore most probably sites of large shocks. AGNs also, in particular their jets, are possible sites for UHECR acceleration.

Until recently, it had not been possible to do cosmic rays astronomy because the images of the sources were blurred by magnetic fields. The coming into operation of the Pierre Auger Observatory (PAO) with the collection of a large sample of UHECR showers, is a major step toward making it possible [21,22].

1.2 The Pierre Auger Observatory

1.2.1 General description

The Pierre Auger Observatory (PAO) is a hybrid detector covering 3000 km² where showers are detected from the fluorescence they produce in atmosphere and by their impact on a ground detector array (Figure 1.10) . Its aim is to measure the properties of ultra-high energy cosmic rays (UHECR), i.e. cosmic rays having energy in excess of 1 EeV (10¹⁸eV), in particular the angular and energy dependence of their flux and their mass composition, and to elucidate the question of their origin and of the mechanism of acceleration [21,22].



Figure 1.10 Plan view of the PAO

Construction of the baseline design was completed in June 2008. With stable data taking starting in January 2004, the world's largest data set of cosmic ray observations had been collected already during the construction phase of the Observatory.

When a primary cosmic ray enters the Earth atmosphere, it interacts with it and produces a large number of mesons which, in turn, interact with the atmosphere, and so on until the primary energy is exhausted in ionization losses. The result is a cascade of interactions (Figure 1.11) producing an extensive air shower (EAS). Their longitudinal profile evolves slowly with energy, in proportion to its logarithm, while its energy content, in the form of ionization losses, is proportional to energy.

A major fraction of the mesons produced are pions, either neutral or charged. The former decay promptly into two photons and are therefore lost for the development of the hadronic cascade. They generate instead electromagnetic showers consisting mostly of electrons, positrons and photons and developing longitudinally at the scale of a radiation length, twice as short as the interaction length which governs the development of the hadronic cascade. The charged pions will have a chance to decay into a muon-neutrino pair if their decay length, 56 m/GeV , is short enough in comparison with the interaction length. As a result, the muon to electron/photon ratio increases with depth.

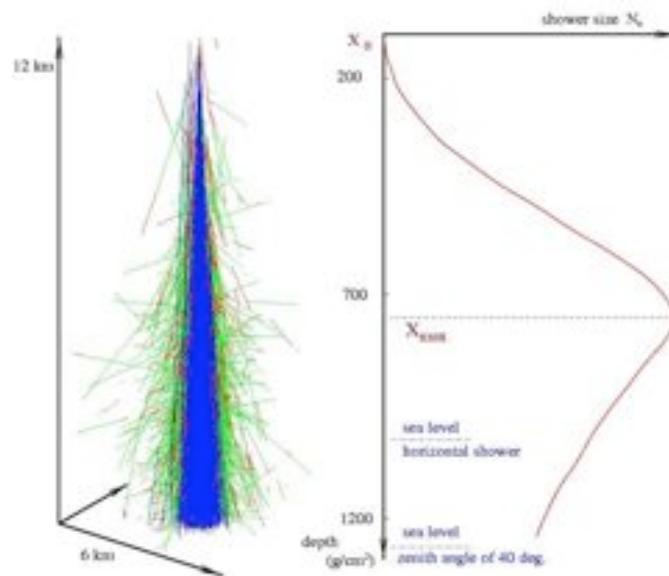


Figure 1.11 Longitudinal development of an extensive air shower [23].

Around 30 EeV , the UHECR flux is about $0.2 \text{ km}^{-2}\text{century}^{-1}\text{sr}^{-1}\text{EeV}^{-1}$ and drops rapidly at higher energies, implying a very large coverage, but the showers contain billions of particles when reaching ground and cover several square kilometers, allowing for a thin sampling [24]. The PAO covers 3000 km^2 in the Argentinean pampas, of which only 5 ppm are covered by detectors. These include 1600 Cherenkov detectors making up the surface detector (SD), and 24 fluorescence telescopes making up the fluorescence detector (FD). Data are transferred by radio to an acquisition centre which filters them and sends them out for subsequent dispatching to the laboratories associated with this research, including VATLY in Ha Noi.

The SD is described in detail in the next section.

The FD is organized in four stations of six telescopes each, which overlook the PAO area (Figure 1.12). They measure the fluorescence light (near UV) produced in the

interaction between the shower charged particles and the nitrogen molecules of the atmosphere. They can only operate during clear moonless nights, which implies a duty cycle of 13%. Each telescope covers a field of view of 30° in azimuth and 28.6° in elevation. After having been filtered, the light is reflected by a concave mirror onto an array of 440 hexagonal PMT pixels. In principle, a single telescope is sufficient to measure the direction of the shower axis from the measurement of the times at which each pixel is hit. But, in practice, a precise measurement requires either binocular detection or, less demanding, the simultaneous detection of the time at which at least one of the ground Cherenkov detectors has been hit by the shower [25,26]. The energy is measured from the longitudinal profile [27] which, when accurately and fully measured, provides a direct calorimetric evaluation of the shower energy (the energy carried away by neutrinos and muons penetrating in ground is of the order of 10% and does not much fluctuate from shower to shower). However, in practice, this measurement is difficult: it implies a good knowledge of the air transparency and of the atmospheric Cherenkov light contamination and, most of the time, the shower is only partly contained in the field of view.

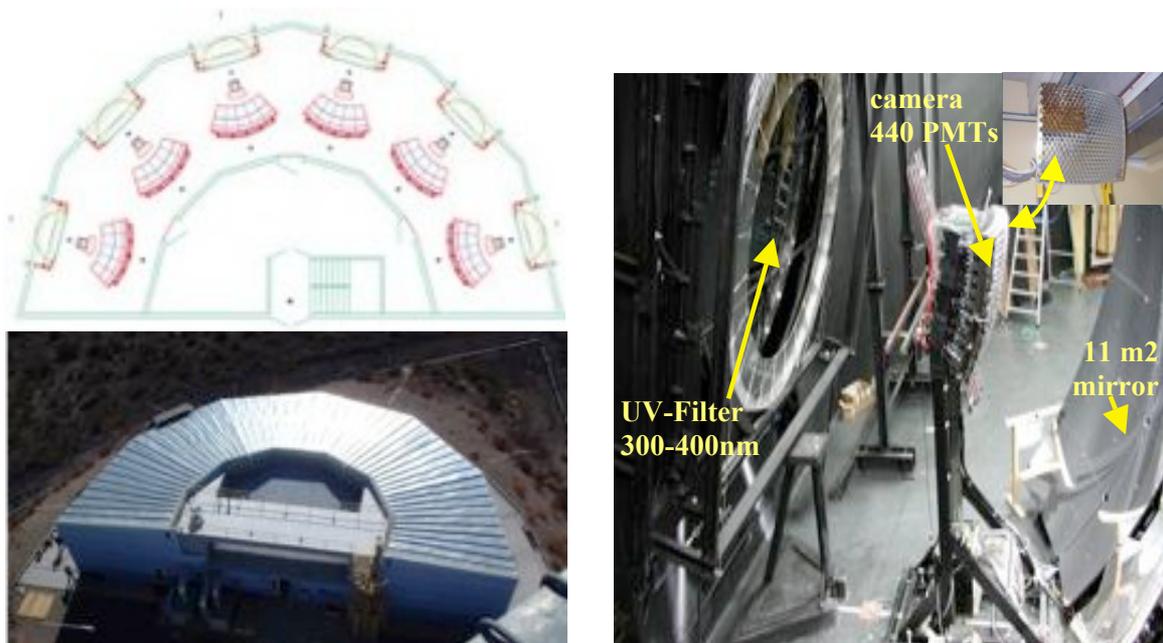


Figure 1.12 Left: A fluorescence station: schematic view (on top) and its photograph. Right: Photograph of an eye.

1.2.2 The surface detector

The SD samples the footprint of the showers on ground. It is made of a triangular array of water Cherenkov counters having a mesh size of 1.5 km deployed on flat ground

at an altitude of 1400 meters above sea level, near the maximum of shower development for the highest energy vertical UHECRs. When reaching ground, showers consist essentially of low energy electrons, positrons and photons as well as of muons having a kinetic energy of a few GeV. In both water Cherenkov counters and scintillator plates, the muon signal is proportional to track length; on average, when averaging over the detector section normal to the direction of incidence, the signal is therefore proportional to the detector volume independently from the angle of incidence. On the contrary, electrons and photons produce small showers at radiation length scale that are fully contained in a water Cherenkov counter but only partially in a scintillator plate. The net result is that they provide a sky coverage twice as large as would be obtained with an array of scintillator plates.

When shower particles are detected in at least three counters, the measurement of the time at which they are hit allows for a precise measurement of the azimuth and zenith angle of the shower axis accounting for the slight curvature of the shower front [28].

The energy measurement is indirect but much easier than in the FD case. It implies the construction of a standard function [28], called lateral distribution function (LDF), which gives the average signal measured in a Cherenkov tank as a function of shower energy, distance to the shower axis and zenith angle. The zenith angle dependence is evaluated under the hypothesis of an isotropic cosmic ray flux. The energy is essentially measured by the normalization of the measured signals to the standard LDF at a distance of 1000 meters from the shower axis (one calls it $S(1000)$). The choice of such a reference is dictated by two scales: the tank spacing, 1.5 km, and the size of the shower detectable footprint on ground, which increases only slowly, logarithmically, with energy. In practice the influence of the former is dominant. The final energy scale is calibrated [29] using FD data in hybrid events as illustrated in Figure 1.13.

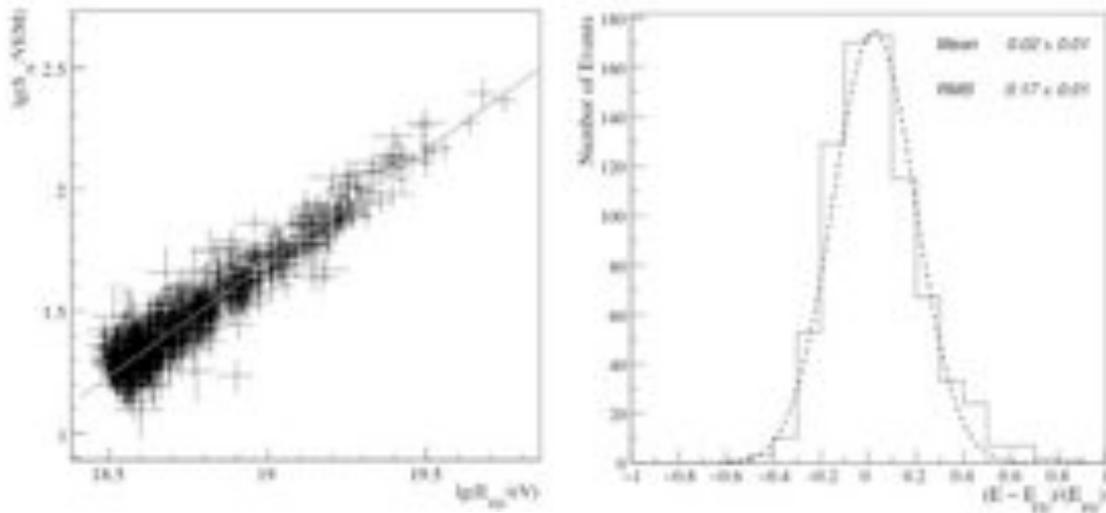


Figure 1.13 Left: Correlation between the decimal logarithms of the energy measured in the FD (abscissa) and of the normalization (ordinate) of the measured SD signals to the value of $S(1000)$ (referred to 38° zenith angle for technical reasons) for the 795 hybrid events used in the fit. The line represents the best fit. Right: Fractional difference between the calorimetric energy, E_{FD} , and the energy estimate of the surface detector, E , obtained by the calibration curve, for the 795 selected events.

Figure 1.14 summarizes the information gathered by the SD [22], showing both the footprint of the shower on ground and the fit to the LDF. Figure 1.15 shows the first four-fold hybrid event recorded in May 2007 with all FD stations active.

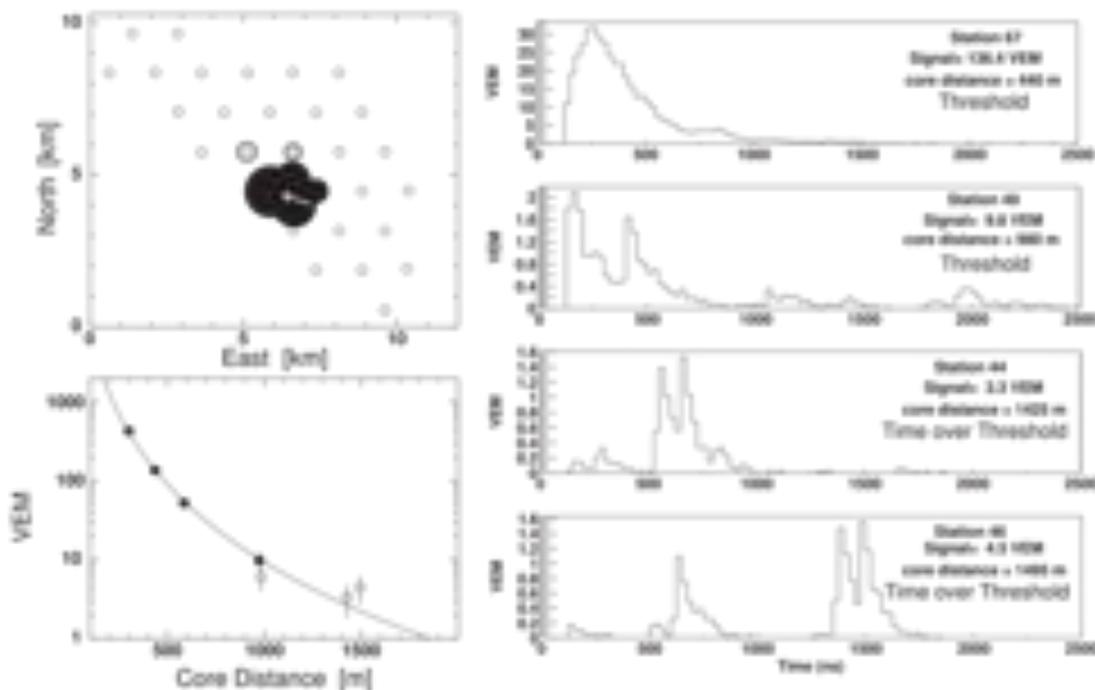


Figure 1.14: Event 211377: a typical event of about 5×10^{18} eV: Top left: The top view of triggered tanks. Lower left: The fit to the LDF. Right: FADC traces from four detectors. The signal sizes are in units of VEM.

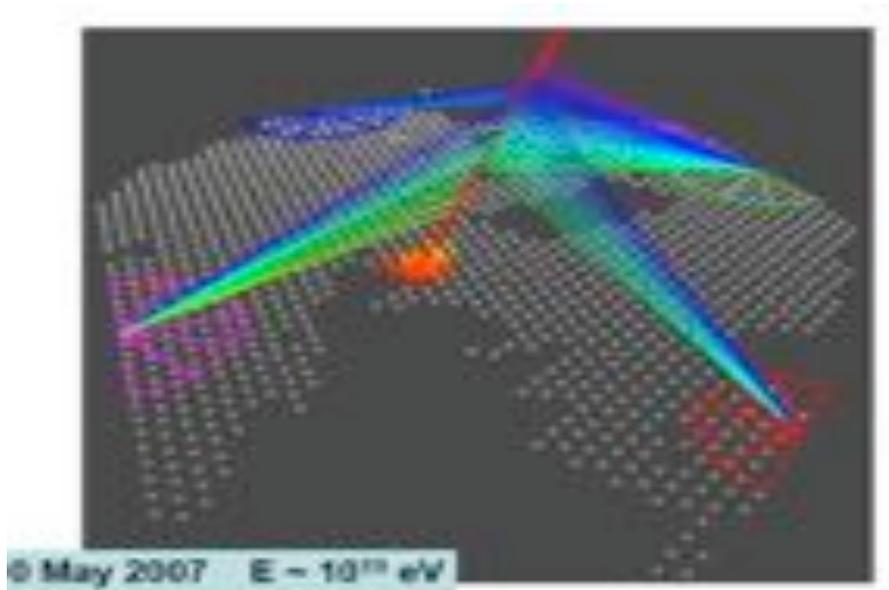


Figure 1.15 The first four-fold hybrid event

1.2.3 Cherenkov tanks

Each Cherenkov counter is made of a resin tank shaped to host a cylindrical volume of ultra pure water, 1.2 m in height and 3.6 m in diameter (Figure 1.16). The water is contained in a highly diffusive plastic bag fitting closely in the resin tank and the Cherenkov light produced in the water volume is seen by three 9" spherical photocathode photomultiplier tubes (PMT) through high transparency windows. The PMTs are not shielded from the Earth magnetic field but are all oriented in a same way meant to maximize their response [30]. The amplification chain of each PMT is made in two parts: a central foil dynode and a standard linear focus chain of seven dynodes. The charge collected from the last dynode is amplified in such a way as to exceed the anode charge by a factor 32. Both are read under 50Ω in 10 bits 40 MHz flash analog to digital converters (FADC). The very high dynamical range implied by the steep slope of the LDF near the shower core results in occasional saturation of the dynode signal.

Energy calibration is constantly monitored by recording locally low energy atmospheric muons inbetween triggers. As such muons are mostly relativistic and feed through the tank, their charge spectrum is essentially a replica of the distribution of track lengths across the water associated with the proper (typically cosine square) zenith angle distribution. As small zenith angle muons are an important fraction of the total, they

produce a peak in the charge distribution which is used to monitor the energy scale (Figure 1.17).

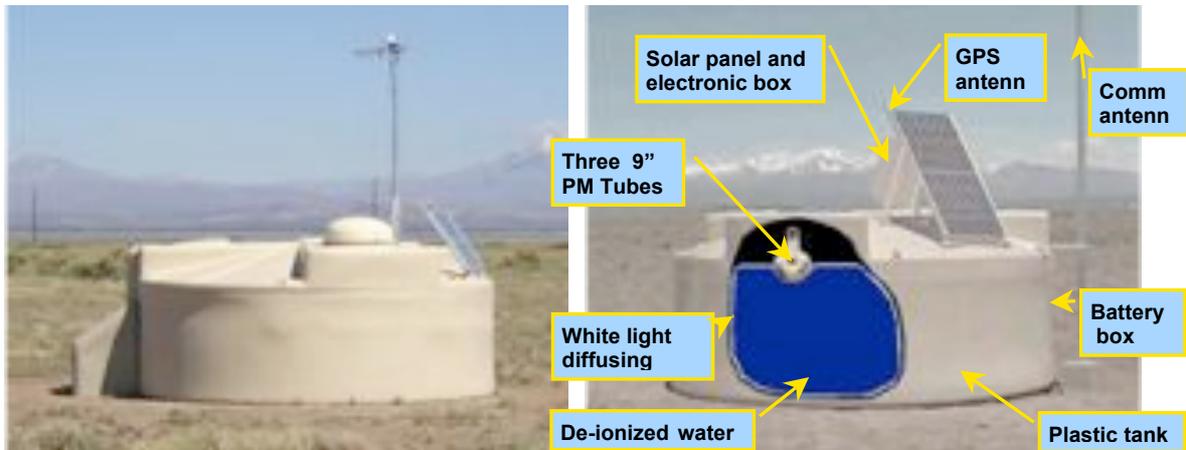


Figure 1.16. A photograph (left) and an exploded view of an Auger water tank.

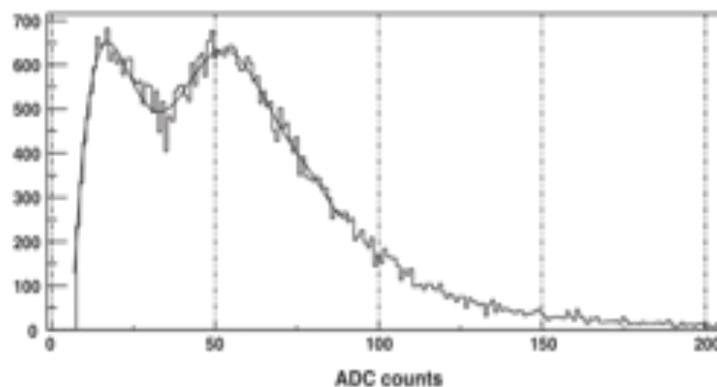


Figure 1.17 Histogram of signals from one PMT in one of the stations of the SD. The peak due to single muons is clearly visible at around 50 ADC channels. The peak at about 20 channels is artificial and is due to the cut made in plotting the data.

The unit used is called VEM for Vertical Equivalent Muon and corresponds to the charge associated with a vertical relativistic muon impinging in the centre of the tank. The calibration of the muon peak displayed in Figure 1.17 in terms of VEM units was done once for all using a scintillator hodoscope bracketing a Cherenkov tank from above and below.

Low level triggers are produced locally by each station whenever some conditions are satisfied, such as a three-fold coincidence of signals exceeding 1.75 VEM or a two-fold coincidence of signals exceeding 0.2 VEM per bin in at least twelve FADC time bins within a 3 μ s window. The main trigger is built centrally from the first level triggers

received from the stations by requiring coincidences in time and in space, the latter being done using a hierarchy of concentric hexagon. Higher level triggers have been designed this way to suppress random coincidences and to provide a trigger efficiency close to unity for showers having energy in excess of $10^{18.5}$ eV.

The electronics in each tank is powered using solar panels feeding a 12 V battery and the data are transferred to the central data acquisition system in the 7 GHz band.

1.2.4 Simulations

Analyzing SD data often requires the help of simulations reproducing the shower development and/or the detector response.

The latter is in principle straightforward but, in practice, quite complex. It requires a good knowledge of the water transparency and liner diffusivity (Lambertian and specular) as a function of wave length, of the quantum efficiency of the photocathode, again as a function of wave length, of the collection efficiencies at the first and second dynodes as a function of photon impact, of the PMT gains, of the electronic and thermal noises, of the after-pulsing characteristics, etc [31]. Much effort has been dedicated in the PAO collaboration to produce adequate codes [32].

The former, however, addresses an energy range in which the characteristics of the hadronic interactions of nuclei, baryons and mesons with air are unknown. One needs to rely on hypotheses, some of which are highly conjectural. To quote a few: the adequacy of the Glauber model to mimic nucleus-nucleus interactions, the extrapolation to higher energies of total cross-sections, rapidity and transverse momentum distributions, inelasticities, multiplicities, particle compositions (including resonances) of baryon-air and meson-air interactions, etc. In addition to these fundamental problems, a technical difficulty results from the very large number of shower particles, which precludes following each of them individually in a Monte Carlo code [33,34]. In order to cope with the need to keep computer time within reasonable limits, various techniques have been developed, such as the “thinning” method, which consists in following, in a well controlled way, only part of the shower particles. Moreover, a same shower may be used many times by simply changing the location of its impact on ground with respect to the detector array, thereby providing a large sample of simulated events. However, in such a case, attention

must be paid to the fact that such simulated events are not at all statistically independent. For example, if the shower starts at significantly lower altitude than average, it has important consequences on several of its properties of relevance to studies of the mass composition of the primary: ignoring it would strongly bias the results of the analysis.

General programs are available to simulate extensive air showers. In particular, CORSIKA [35] and AIRES [36] offer general frames that can accommodate a number of hadronic interaction models. It has been used to generate a library of proton and iron showers covering the energy range from 10^{17} to 10^{20} eV and a range of zenith angles between 0° and 70° .

1.2.5 Energy spectrum and the GZK cut-off

The PAO has already given two particularly important contributions to the physics of UHECRs. One is the evidence for the so-called GZK cut-off, the other is the observation of a correlation between the direction of arrival of the highest energy UHECR and nearby galaxies.

For some time, the differential spectral index of the energy spectrum has been known to change at $\sim 3 \cdot 10^{15}$ eV from 2.7 to 3.0, this is referred to as the knee, and again back to 2.7 near the upper end of the spectrum, this is referred to as the ankle. The latter is often attributed to the transition from galactic to extra galactic sources, although some models accommodate extra galactic origins below the ankle. Sensible scenarios can be produced which reproduce the data.

Of particular relevance to such scenarios are the interactions of cosmic rays with the cosmic microwave background (CMB), producing either electron-positron pairs or new mesons. Of these, the pion photoproduction threshold is of particular importance and causes the so-called Greisen-Zatsepin-Kuzmin (GZK) cut-off at the end of the spectrum, from the name of the physicists who first predicted the effect [37]. Until recently, the existence of such a cut-off was controversial but the Pierre Auger Observatory has settled the issue and given evidence for it. With a typical interaction length in the few 10 Mpc scale, cosmic rays coming from larger distances cannot make it to the Earth without interacting, and therefore lose energy: their flux is significantly damped and only nearby (<100 Mpc) sources can contribute to the UHECR spectrum.

The most recent PAO data [38], combining both SD and FD data, are illustrated in Figure 1.18 showing the fractional difference of the spectrum with respect to an assumed flux of spectral index 2.6. Two spectral features are evident: an abrupt change in the spectral index near 4 EeV (the “ankle”) and a more gradual suppression of the flux beyond about 30 EeV corresponding to the GZK cut-off.

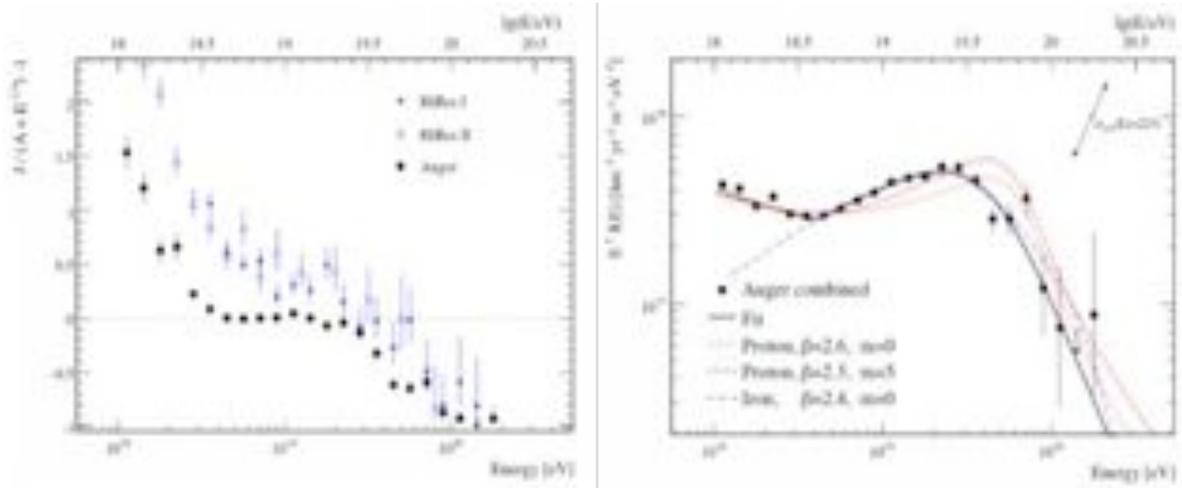


Figure 1.18 Left: Fractional difference between the combined energy spectrum of the Pierre Auger Observatory and a spectrum with an index of 2.6. Data from the HiRes instrument [39] are shown for comparison. Right: Combined energy spectrum compared with several astrophysical models including a pure composition of protons (red lines) or iron (blue line).

1.2.6 Correlations with astronomical sources

The large UHECR statistics accessible to the PAO has revealed a correlation with extragalactic counterparts [40]. Of relevance to this study is the fact that the nearby universe (100 Mpc radius), in which detected UHECRs are confined by the GZK cut-off, is highly inhomogeneous (Figure 1.19). Selecting UHECR having an energy in excess of $6 \cdot 10^{19}$ eV and comparing the direction in the sky where they come from with a catalogue of nearby (<75 Mpc) galaxies, revealed a clear correlation (Figure 1.20). Both numbers corresponded to values giving the best statistical significance to the observed correlation and were in agreement with reasonable expectations based on favoured estimates of the galactic and extra galactic magnetic fields for the former, and on the size of the GZK horizon for the latter. There was an even better correlation with nearby AGNs (of which, however, there exists no complete catalogue). The correlation disappeared when including lower energy cosmic rays (pointing accuracy) or farther away galaxies (GZK cut-off).

An update of these data has been presented recently [41], including data collected through 31st March, 2009 and corresponding to an exposure of $17040 \text{ km}^2 \text{ sr yr}$ ($\pm 3\%$), nearly twice the former value. There are now 31 additional events above the energy threshold of 55 EeV. The systematic uncertainty on energy is $\sim 22\%$ with a resolution of $\sim 17\%$ while the angular resolution of the arrival directions is better than 0.9° . During the period reported earlier, 18 out of 27 events arrive within 3.1° of an AGN in the VCV catalogue [42] with redshift less than 0.018 while of the 31 additional events, 8 have arrival directions within the prescribed area of the sky, not significantly more than the 6.5 events that are expected to arrive on average if the flux were isotropic: the degree of correlation with objects in the VCV catalogue has decreased with the accumulation of new data, as clearly illustrated in Figure 1.21. Yet, possible biases have been carefully explored and discarded. In particular, the parameters used to select the data sample (angular separation, maximum red shift and energy threshold) still apply: of the subset of 44 events which had not been used to define these parameters, 17 correlate when using them, a correlation that has less than 1% probability to occur by chance for an isotropic distributions of arrival directions.

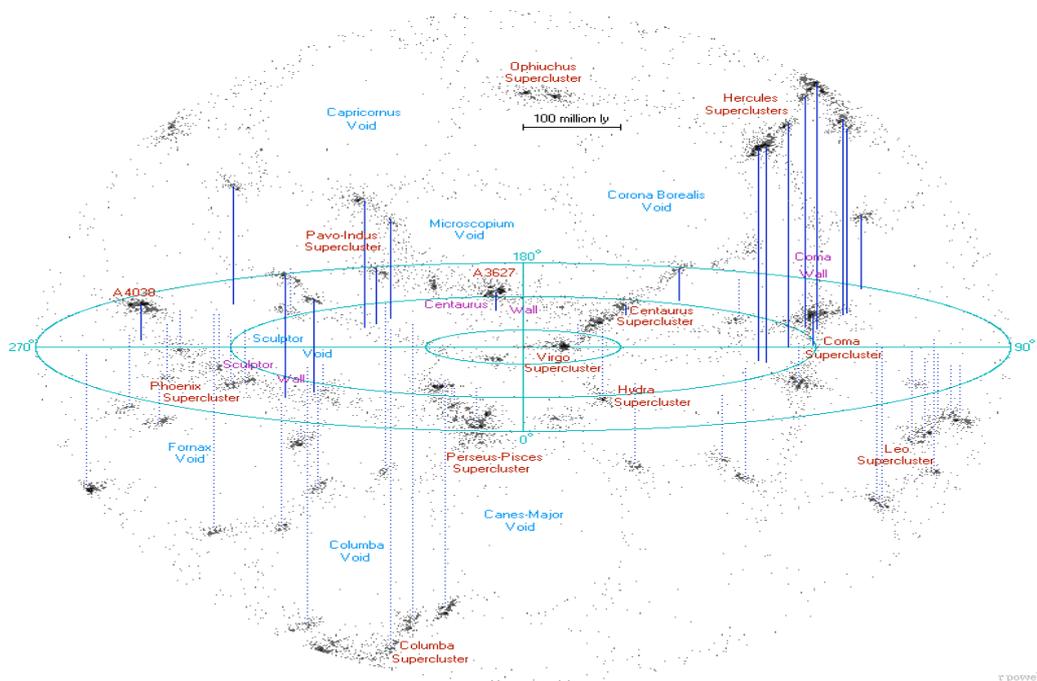


Figure 1.19 The nearby Universe

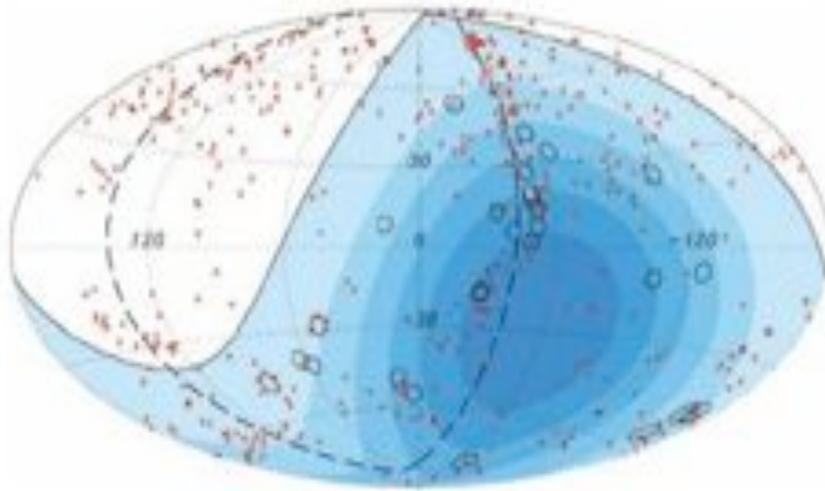


Figure 1.20 Circles of 3.1° are drawn around 27 UHECR detected by the PAO up to year 2007 [38]; red crosses are 472 AGN (318 in field of view) having $z < 0.018$ ($D < 75$ Mpc). The solid line shows the field of view (zenith angle $< 60^\circ$) and the colour tells the exposure. The dashed line is the super galactic plane.

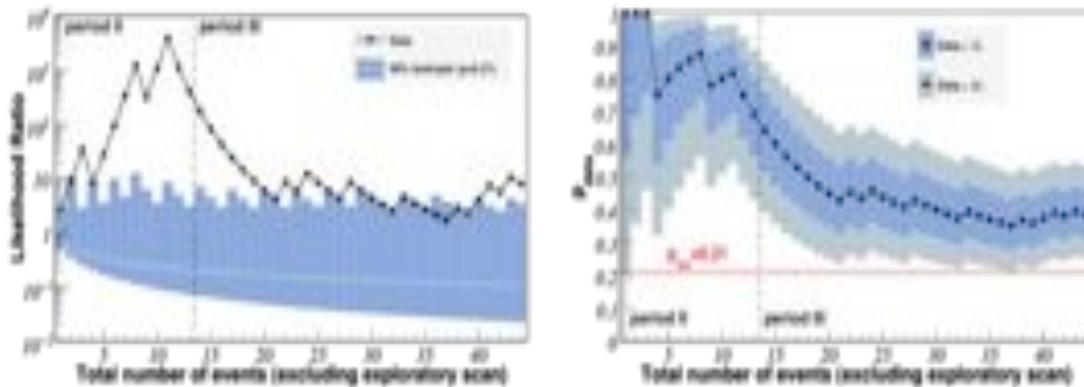


Figure 1.21 Monitoring the correlation signal. Left: The sequential analysis of cosmic rays with energy greater than 55 EeV arriving after 27 May, 2006. The likelihood ratio for the data is plotted in black circles. Events that arrive within $\psi_{max} = 3.1^\circ$ of an AGN with maximum redshift $z_{max} = 0.018$ result in an up-tick of this line. Values above the area shaded in blue have less than 1% chance probability to arise from an isotropic distribution. Right: The most likely value of the degree of correlation with objects in VCV catalogue is plotted (black circles) as a function of time. The 1σ and 2σ uncertainties in the observed value are shaded. The horizontal dashed line shows the isotropic value. The current estimate of the signal is 0.38 ± 0.07 [40].

An accumulation of events is observed in the Centaurus A region and has been the subject of studies taking into account different models of the galactic magnetic field. The data collected so far establish the ability of the PAO to point to sources in the sky, typically

within 3° , which was not a priori obvious because of uncertainties in magnetic fields met by UHECR during their journey to the Earth.

While these results open new perspectives for cosmic ray astronomy, many questions remain unanswered, in particular understanding why such and such a galaxy, AGN or else, is a source while such and such another is not. Several events point to regions in the sky where there seems to be essentially no matter within the GZK horizon. Is it because of unexpectedly large magnetic fields or because of high charge ions, such as iron, or because of a real but invisible source?

The observations made at the PAO in the years to come will help with the identification of preferred acceleration sites. Yet, the need for a significantly larger exposure is obvious: the construction of a larger array is currently being planned in the Northern hemisphere. However, the increased coverage and limited resources imply significant losses in the present outstanding quality of the detector: a smaller detector coverage (by one order of magnitude) and a single PMT per tank.

1.3 Identification of the primaries

1.3.1 General considerations

Low energy cosmic rays are known to have abundances similar to those found in interstellar matter with a predominance of protons. At UHECR energies, however, the mass composition of primaries is uncertain [43-45]. There are even conjectures suggesting particles other than atomic nuclei to populate the higher energy range. Disregarding such exotic scenarios, the question remains of measuring the mass distribution of the primaries in a range spanning essentially from protons to iron nuclei, higher mass nuclei being much less likely.

The main difference between showers induced by protons and iron nuclei results from the very different natures of their first interaction in the upper atmosphere. The proton shower starts to develop on average after having crossed one interaction length and the depth of its starting point fluctuates with a variance also equal to one interaction length. The iron shower, in an oversimplified picture, may be seen as the superposition of 56 proton showers (protons and neutrons are equivalent at such energies), each carrying $1/56$

of the nucleus energy. As a result it starts much earlier, and the location of its starting point fluctuates much less, than in the proton case. From then on proton and iron showers develop in the same way. While such a description is useful to provide a simple qualitative explanation of what is going on, the reality is far more complex and its details are not well understood. Not all nucleons of the colliding nuclei interact the same way. In a simplified picture, some nucleons – one refers to them as wounded nucleons – interact as if they were independent nucleons while the other nucleons – one refers to them as spectator nucleons – are unaffected. This, again, is an oversimplified view of reality. Glauber model [46] provides a recipe to evaluate the number of wounded nucleons.

Nevertheless, as a general rule, in order to distinguish between light and heavy incident nuclei one will aim at measuring quantities that are sensitive to the early shower development.

The interest of such measurements is obvious and their importance is increased by the results of the preceding section, some of which may invoke a significant iron population to explain why some of the highest energy UHECR do not seem to point to any known counterpart.

1.3.2 Longitudinal profiles

The fluorescence detector (FD) of the Pierre Auger Observatory can be used to measure with good resolution the shower longitudinal profile and the depth at which the shower reaches its maximum (X_{max}). At a given energy, the mean and the width of the X_{max} distribution are both correlated with the cosmic ray mass composition [47]. Proton showers penetrate deeper into the atmosphere (larger values of X_{max}) and have wider X_{max} distributions than heavier nuclei.

In practice, however, such a measurement is difficult and a strict selection of useful events is mandatory. A good geometry (average angular resolution of 0.6°) [48] is obtained by requiring the simultaneous detection of shower particles in at least one Cherenkov tank of the SD and by rejecting showers pointing toward the telescope (the time over which pixel hits are recorded must exceed $5\mu\text{s}$). Moreover, the reconstructed X_{max} should be clearly identified and, obviously, lie within the field of view [49]. This is achieved by requiring that the observed profile spans at least 320 g cm^{-2} and that the reduced χ^2 of a fit to a reference profile (showing a maximum) does not exceed 2.5 and is smaller than that of

a straight line fit by at least 4 units. Finally, the estimated uncertainties of the shower maximum and total energy must be smaller than 40 g cm^{-2} and 20%, respectively. The uncertainty on the X_{max} measurement is evaluated from stereo events and found to be $21 \pm 1.5 \text{ g cm}^{-2}$. The most recent PAO results [50] are shown in Figure 1.22 together with predictions of popular hadronic models for both protons and iron nuclei.

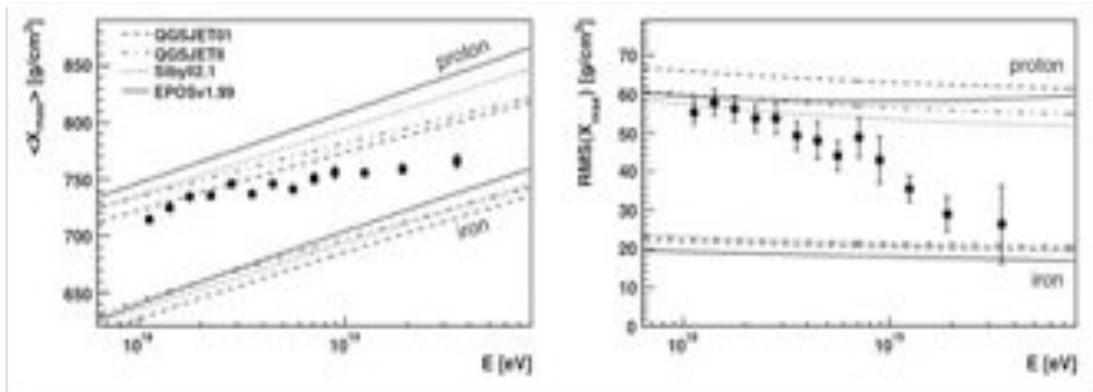


Figure 1.22 $\langle X_{max} \rangle$ and $RMS(X_{max})$ compared with air shower simulations [51] using different hadronic interaction models [34,35].

1.3.3 Risetime

The time profile of particles reaching ground is sensitive to the shower development as the first portion of the signal is supposed to be dominated by muons which arrive earlier and over a period of time shorter than electrons and photons [52]. A risetime ($t_{1/2}$) is defined for each tank FADC trace as the time to go from 10% to 50% of the total integrated signal. To the extent that both risetime and X_{max} are expected to be sensitive to the primary mass composition, they should display a clear correlation. Evidence for it is obtained by unfolding the dependence of the risetime on zenith angle and distance to the shower axis. This is done by defining a standard function, in the same spirit as was done for the LDF, and using a particular energy (10^{19} eV) as reference – as one uses the particular distance of 1000 meters for $S(1000)$ [53]. The resulting quantity, called Δ_i , increases on average with energy as expected for showers developing deeper into atmosphere (Figure 1.23) and is indeed clearly correlated with X_{max} [54] as shown in Figure 1.24.

One might of course exploit this correlation to calibrate the risetime scale in terms

of mass composition but this would not bring additional information – except, to some extent, for what concerns the relative energy dependences of X_{max} and Δ_i .

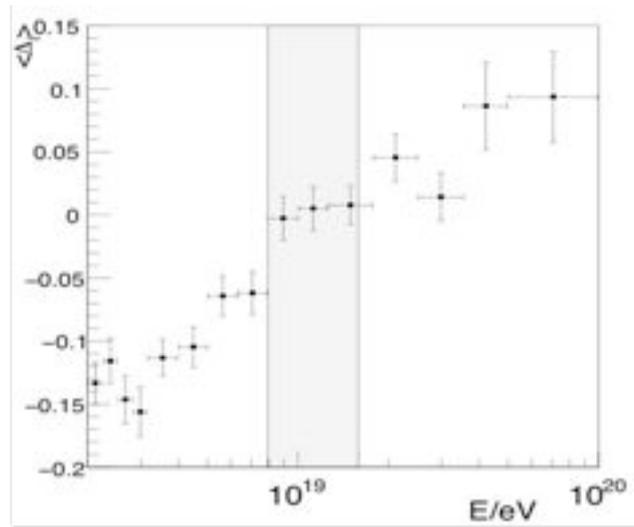


Figure 1.23. SD events: Dependence of the mean value of Δ_i on energy.

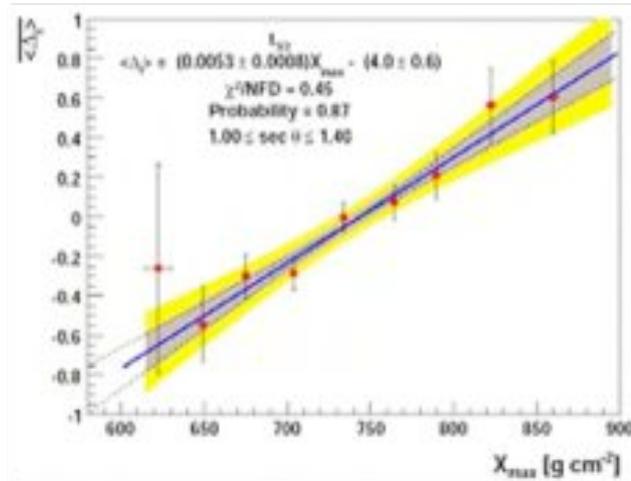


Figure 1.24. Hybrid events: Dependence of the mean value of Δ_i on X_{max} . A correlation is found which is parameterised with a linear fit. The shaded areas show the estimated uncertainty (one and two σ), obtained by fluctuating each point randomly within the measured error bar and repeating the fitting procedure.

Another property displayed by the risetime is its dependence on tank azimuth ξ measured around the shower axis, the more so the more inclined is the shower. When an inclined shower reaches ground, the upstream tanks are hit first and the downstream tanks are hit last. The former probe the shower at an earlier stage of development than the latter do. But there is also a pure geometric effect [55] that differentiates between upstream and

downstream tanks. The path length for particles to reach an upstream tank from the shower axis is much shorter than that to reach a downstream tank with the result that the former are seen under a larger solid angle than the latter and therefore detect a larger signal. Moreover, as noted earlier, the response to muons – most muons having sufficient energy to feed through the tanks – is independent, on average, from the angle of incidence. On the contrary, that of electrons and photons – generating small showers in water – depends on the angle of incidence in the same way as does the tank section normal to the incoming particle momenta. The net effect is an azimuthal asymmetry of the tank responses around the shower axis, trivially increasing with the distance r of the tank to this axis. This asymmetry is in particular visible on the azimuthal dependence of the risetime and is observed to reach a maximum for a value θ_{max} of the zenith angle θ which is sensitive to the depth at which the shower density starts declining. In practice a fit of the form $t_{1/2}=(a+b\cos\xi)r$ allows to measure the dependence on $\sec\theta$ of the asymmetry b/a , which is found to be maximal around $\theta=50^\circ$ independently from energy [54]. This is illustrated in Figure 1.25 where it can be seen that popular hadronic models predict instead an increase with energy of the zenith angle at which the azimuthal asymmetry is maximal. If one were to interpret this result in terms of mass composition, one would conclude that the mean primary masses increase with energy. Indeed, this result is consistent with the FD measurements of the longitudinal profile, suggesting a transition from proton dominance – light nuclei – to iron dominance – heavier nuclei – when the energy increases from 1 to 30 EeV.

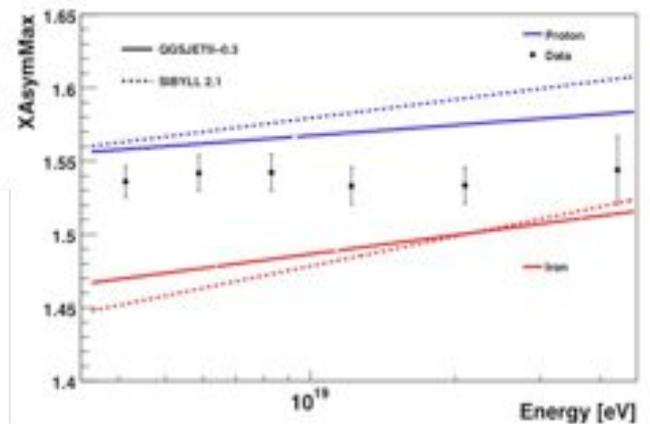


Figure 1.25 Measured dependence of the position of maximum asymmetry on primary energy. Lines correspond to fitted distributions of MC samples for proton (blue) and iron (red) primaries.

1.3.4 Muon abundance

An indicator of the shower age is the relative muon abundance, which increases with age: at a same depth, iron showers are therefore expected to be more muon-rich than proton showers are. While no direct measurement of the muon abundance has yet been made, numerous attempts at measuring quantities closely related to the muon abundance have been explored. Such is the risetime, which was presented in the preceding section.

Other approaches include attempts at identifying muons from sudden jumps [56] in the FADC traces (the “jump method”) and a direct evaluation of the muon signal by subtraction of the electron-photon contribution from the FADC trace [57].

This latter method implies that the electron-photon signal (i.e. the contribution given by electrons and photons to the FADC traces) is a function of energy, zenith angle and depth (measured with respect to X_{max}) having a zenith angle dependence obtained from the hypothesis that the bulk of detected showers are isotropic and an energy dependence known from hadron models. Under such assumptions, the muon abundance is the only unknown. When measured relative to that predicted for proton primaries, it is 1.53 ± 0.08 (stat.) ± 0.21 (syst.). Pure iron composition would predict a lower factor, of the order of 1.3.

Additional evidence is obtained by the analysis of hybrid events [58] where the longitudinal profile is used to choose between a proton and an iron hypothesis, whatever is best, and to then predict the amplitude of the signal on ground (Figure 1.26).

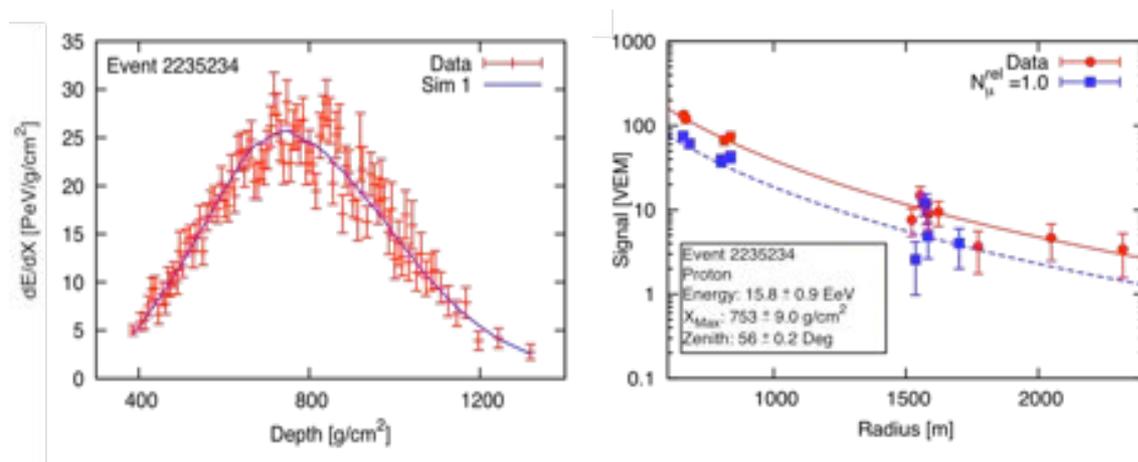


Figure 1.26 Measured longitudinal (left panel) and lateral (right panel) profiles for one of the hybrid events. The best-matching simulation is shown as squares and dashed line in the lateral distribution while the measured SD signal (circles, full lines) is more than twice as large.

A similar conclusion is also reached from analyses of the FADC traces such as

done with the jump method: the muon abundance inferred from such analyses is significantly larger than that predicted for iron by popular hadronic models. Figure 1.27 below summarizes the results. A possible interpretation is to dispose of the problem by blaming it on a 30% underestimate of the FD energy scale, as illustrated in Figure 1.27. Even so, measurements exceed the expected muon abundance for iron. Moreover, a recent energy independent analysis [59] would rather suggest that the hadronic models used in the simulation predict too steep a muon lateral distribution function. This is indeed another way to increase the amplitude of the muon component in the D range explored by the SD. Moreover it would explain why the azimuthal asymmetry of the risetime gives results in agreement with the FD X_{max} measurement, as both probe the longitudinal profile independently from the lateral distribution function.

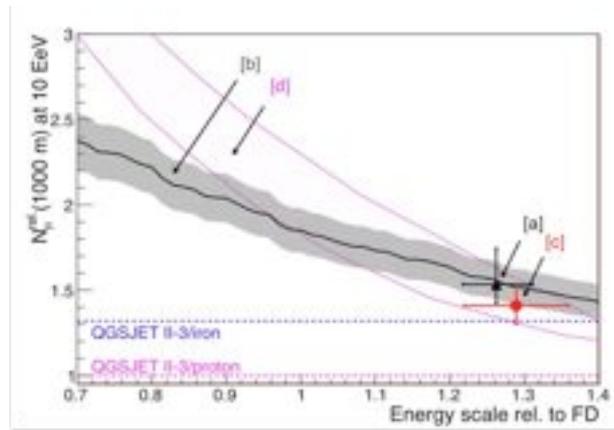


Figure 1.27. Number of muons at 1000 m relative to QGSJET-II/proton vs. the energy scale from different SD analyses (see text). The events have been selected for $\log_{10}(E/\text{eV}) = 19.0 \pm 0.02$ and $\theta \leq 50^\circ$. According to the tested model, iron primaries give a number of muons 1.32 times bigger than that from protons (horizontal lines in the figure).

1.3.5 Summary

In summary, the mass composition of UHECR primaries remains an open question. Major progress has been achieved in the analysis of FD data where a rigorous treatment of possible biases and systematic uncertainties is now available. The results are consistent with the predictions of hadronic models and, in such a picture, provide evidence for a transition from proton-like to iron-like primaries over the energy range covered by the PAO, say 1EeV to 30 EeV where the GZK threshold becomes effective. This conclusion is also reached, with lesser accuracy, by the analysis of the azimuthal risetime asymmetry in the SD, an analysis sensitive to the depth at which the longitudinal shower profile starts

declining. Yet, SD analyses that are sensitive to the amplitude of the muon signal can only be made consistent (barely) with the predictions of hadronic models at the price of a 30% increase of the energy scale. A possible cause might be the inadequacy of hadronic models to reproduce the lateral distribution function of muons. Another possible cause might be the inadequacy of the detector simulation to describe the response to muons. The study presented in Part 4 of the present report is a contribution to the understanding of that latter point: it analyses the response of the detector to low energy muons, the characteristics of which are expected from simulations to be essentially independent from the nature – proton-like or iron-like – of the primary.

2. FADC traces

The information contained in the SD data is in the form of FADC traces recording the times of arrival and amplitudes of individual PMT signals in each tank hit by the shower that gave the trigger. A detailed understanding of such traces is therefore essential. In particular, a preliminary to any attempt at resolving the FADC traces in signals that may be interpreted in physical terms implies a good understanding of the uncertainties attached to FADC data. These include at least three components: a constant noise level, an error proportional to the square root of the signal due to photoelectron statistics and an error proportional to the signal due to systematic differences between the three PMTs of a same tank (geometry, timing, etc...). It also implies a good understanding of the anomalies, dysfunctions or any feature of relevance to the response of the system.

2.1 General features

The approach followed in the present section consists in comparing the three FADC traces of a same tank in order to spot, and hopefully explain, possible inconsistencies. It will turn out ultimately that the only inconsistencies revealed by the analysis fall in two well-known families: after pulses and early time PMT asymmetries. The analysis is nevertheless conducted without any a priori assumption on what can be expected.

2.1.1 Noise

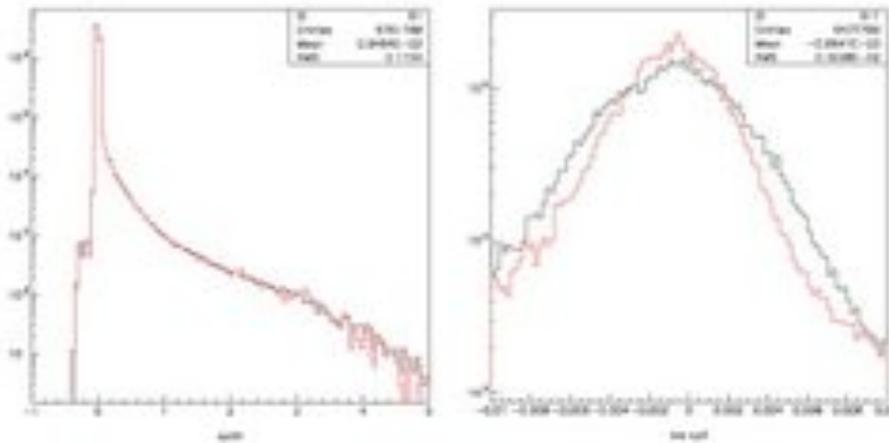
Unless otherwise specified, the study¹ is restricted to dynode spectra, to channels that do not saturate and of course to tanks having no PMT tagged as not working properly. Data are base-line subtracted and scaled to VEM units using the standard procedures² (the baseline amounts typically to 51 ± 7 ADC counts and the ADC count to VEM conversion factor to 184 ± 23 , corresponding to 0.0054 ± 0.0007 VEM per channel).

Figure 2.1 shows the distributions of the channel contents for the whole spectrum and for its lower part. In each case the individual PMT spectra (full line) are shown

¹ Using data from *sdt4_v0r4_2006_05_01[2345]_12h00.root*.

² The subtraction of the baseline is made according to the following algorithm: the rms of the signals measured in a same PMT in channel intervals 0 to 50, 50 to 100, 100 to 150, etc... are calculated until one of them is found to be less than 2 ADC counts. The mean signal in this interval is retained as baseline

together with the mean spectrum (scaled up by a factor of 3 for convenience). The noise component appears clearly and is seen to be Gaussian as shown in Figure 2.2 for both the individual PMT spectra (with $\sigma=0.0032$ VEM) and for their mean ($\sigma=0.0024$ VEM). If one ADC count were much narrower than σ we should expect these two numbers to be in the ratio $1/\sqrt{3}$, namely the mean spectrum to have $\sigma=0.0018$. However, one ADC count is in fact wider than σ (0.0054 VEM compared to 0.0032 VEM) and this effect must be taken in due account. This was done so using a toy Monte Carlo generating a Gaussian noise having $\sigma=0.0028$ VEM such that the individual spectra in bins of 0.0054 VEM have $\sigma=0.0032$ VEM. The resulting σ of the mean spectrum is 0.0018 VEM instead of 0.0024 VEM as observed in the data, suggesting a correlation between the 3 PMT noises (if they were fully correlated, their mean would have an rms of 0.0032 VEM). The toy Monte Carlo was therefore modified to give each PMT noise two contributions, both Gaussian, one specific to each PMT and the other common to the three PMTs. For equal contributions of these two components, namely 0.0020 VEM, the rms of the PMT distributions remains at 0.0032 VEM as expected but that of the mean becomes 0.0024 VEM as found in the data. The mean of the rms to zero (namely the square root of one third of the sum of the squares of the three PMT signals) is 0.0027 VEM, compared to 0.0028 VEM observed in the data.



Figures 2.1 Distributions of the channel contents for the whole spectrum (left) and for its lower part (right). Individual PMT spectra (full lines) are shown together with the mean spectra (scaled up by a factor of 3 for convenience).

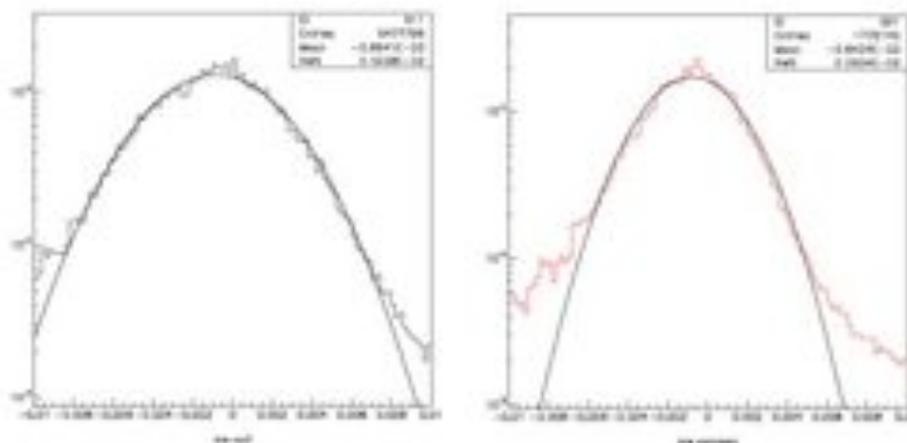


Figure 2.2. Distribution of the noise component for both the individual PMT spectra (left) and for their mean (right).

In conclusion evidence is obtained for a Gaussian noise with an rms of 0.0028 VEM consistent with two similar Gaussian components having an rms of 0.0020 VEM each, one specific to each PMT and the other common to the three PMTs. Due to the large width of an ADC count compared to the noise, the noise spectrum appears in fact wider with an rms of 0.0032 VEM. In practice a cut at 0.015 VEM on the individual PMT spectra provides a very good rejection against noise.

2.1.2 Base-line

Scanning event displays has revealed the existence of traces becoming negative at large times. A preliminary investigation of such cases gave indeed evidence that, in the late part of the FADC trace, the baseline is sometime getting negative, the more so the higher the charge measured in the early part of the FADC trace. The phenomenon is common to anode and dynode traces.

For its study, raw dynode data (before baseline subtraction and ADC count to VEM conversion) are used. The base-line is defined using an improved algorithm. In a first step the maximum and minimum FADC bin contents in each of the sixteen 48-bin intervals of the traces of each of the three PMTs are searched for. The distribution of the difference between the maximum and the minimum is shown in Figure 2.3 for some typical intervals. Whenever the interval is not completely occupied by signals, the noise fits within 4 counts, namely minimum, minimum+1, minimum+2 and minimum+3 (there are nearly no cases

where the maximum and the minimum are equal). In a second step, for each interval of each PMT trace, the mean and rms values of the FADC content (measured in ADC counts) are calculated for all bins where this content does not exceed minimum+3. Figure 2.4 compares this evaluation of the baseline to the standard evaluation used in the official Auger software for some typical intervals.

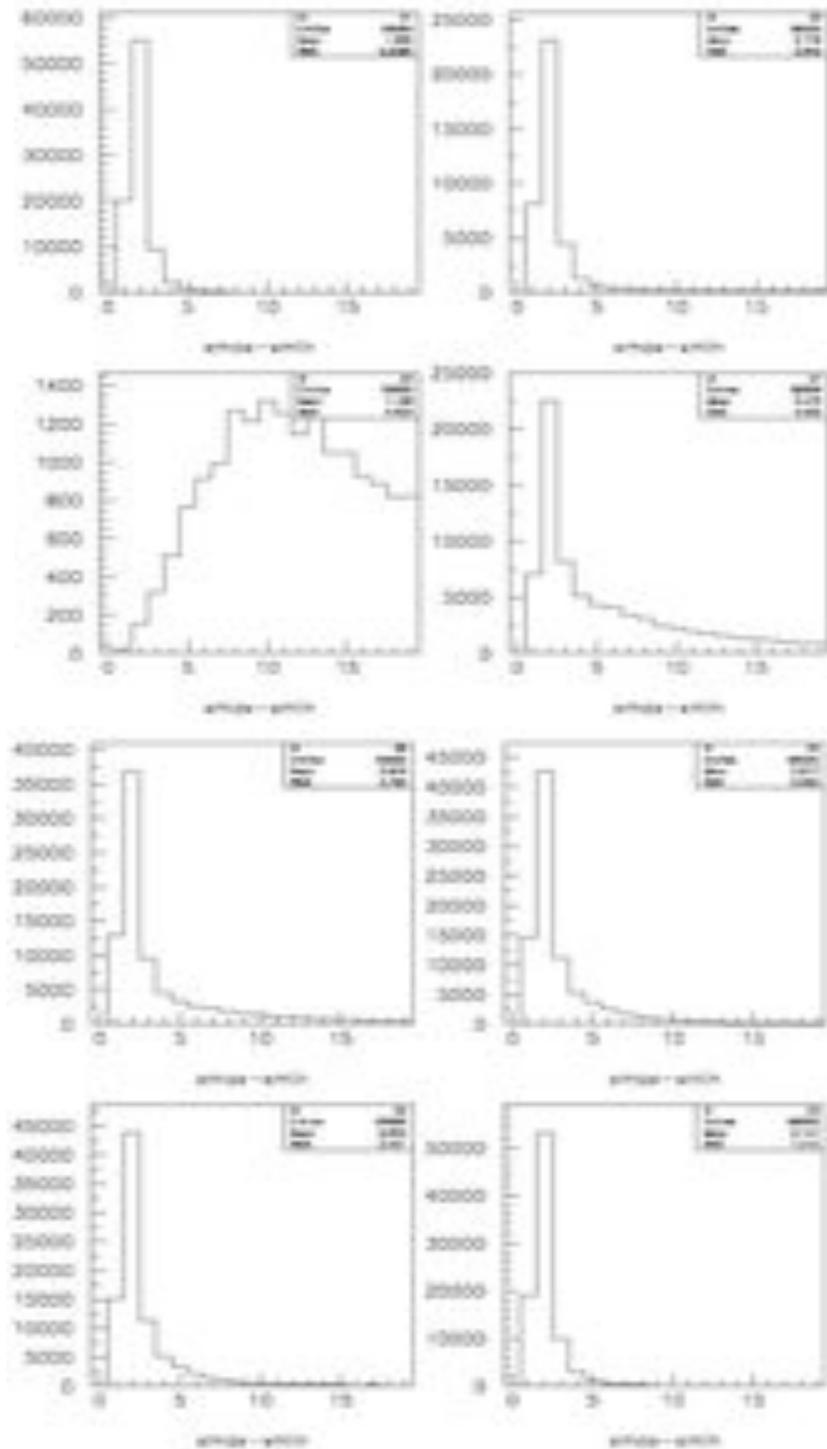


Figure 2.3. Distributions of the difference between maximum and minimum ADC counts in some typical intervals (nr 1, 5, 6, 7, 8, 10, 12 and 16 from left to right and top to bottom).

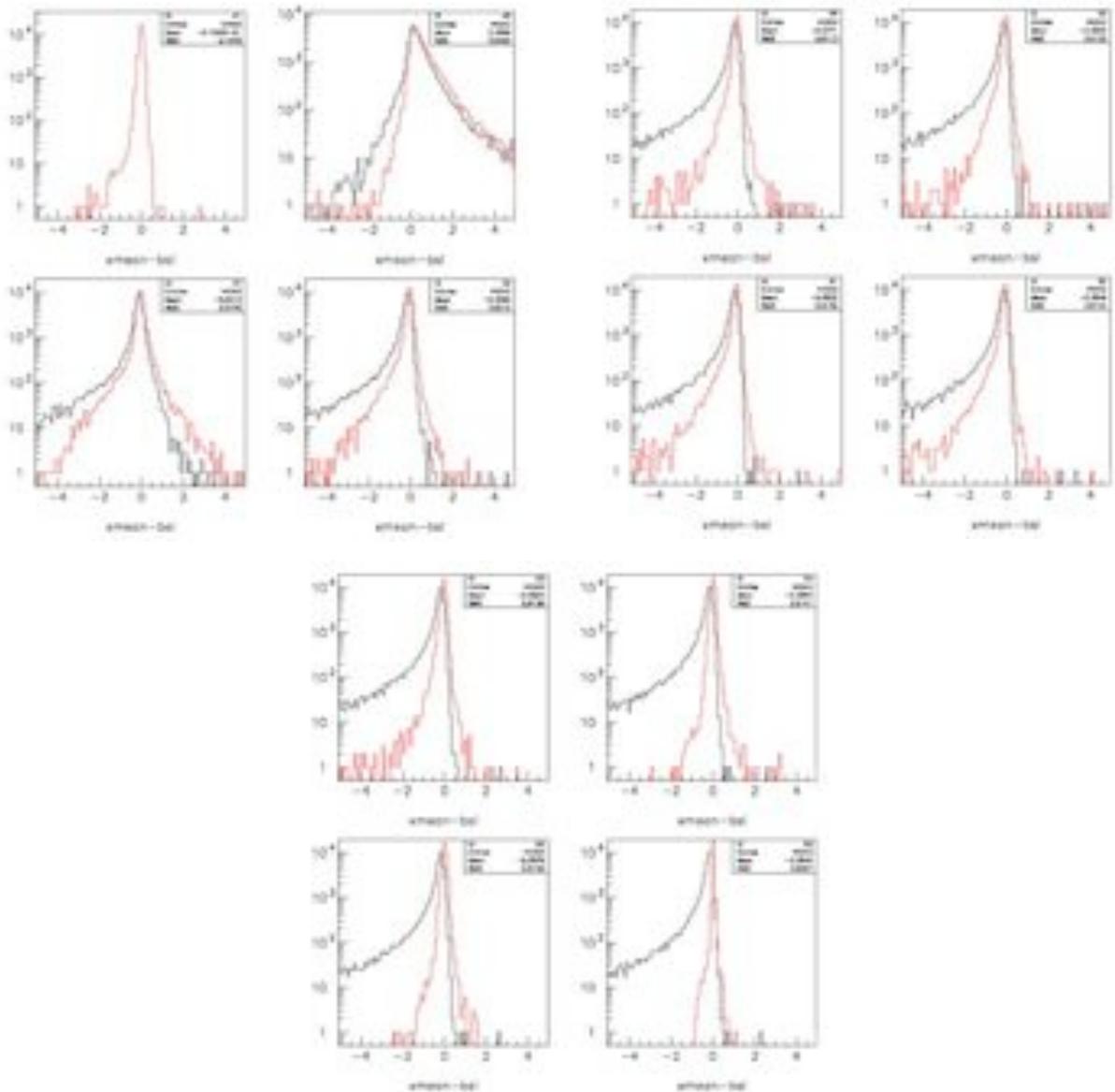


Figure 2.4. Distribution of the difference between the baseline in intervals 1, 6, 7, 8 (upper left), 9, 10, 11, 12 (upper right) and 13, 14, 15, 16 (lower) and the Auger official baseline (corresponding essentially to interval 1). The red curves show the effect of the present correction (see text).

In early intervals, the present and official baseline evaluations are in perfect agreement: indeed the official baseline is *de facto* essentially calculated from the first interval. In intervals 6 and 7, the signal density is such that the present baseline definition often fails, as could be expected. However, from the eighth interval on, several traces show a lower baseline than evaluated in the early intervals. Moreover, one observes that when the present baseline agrees with the official baseline in the last interval, it agrees with it everywhere, and when it disagrees with the official baseline in the last interval, it disagrees with it nearly everywhere.

A base-line correction is therefore defined accordingly with a standard shape displayed in Figure 2.5.

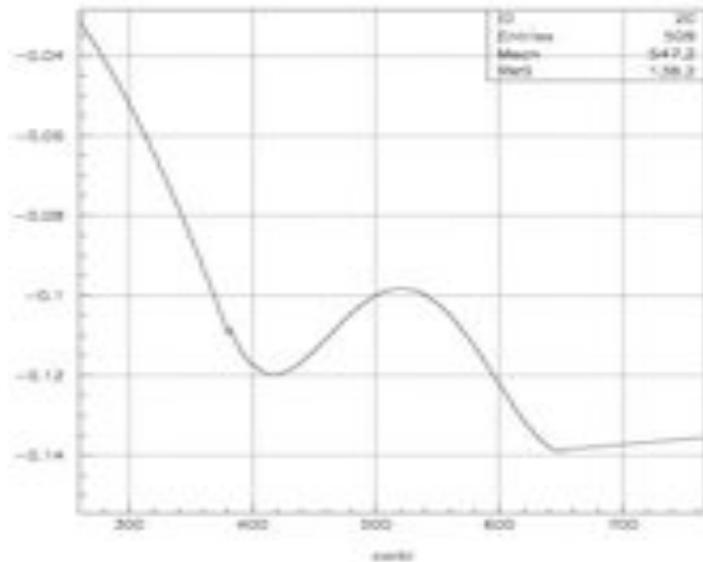


Figure 2.5. Shape of the baseline correction (below bin 260 there is no correction)

For a given trace, the correction uses the last interval, where the baseline is well measured due to the very low signal density, as a reference to scale the standard shape as required. Then, by construction, the baseline is now correct in the last interval as well as in the first intervals. Figure 2.4 illustrates the progress accomplished by this baseline correction. The new baseline is now good, on average, to better than 0.1 ADC count in all intervals except for traces having a very strong early signal, typically 200 VEM or more, which have to be discarded in studies requiring the identification of small signals (of course, in studies where data near the shower core are of relevance, such as the evaluation of the lateral distribution function, they must be retained).

The effect described here has in fact been described previously [60] and is understood as being due to the recovery time of the divider chain capacitors. As it is seen in both dynode and anode traces, the dynode amplifier does not play an important role. However, the presence of a bump around bin 500 cannot be explained by such an effect. Indeed, it has been shown to be associated with after pulses, which increase in proportion with the amplitude of the main (earlier) signal. As is shown later on (Section 2.1.4) such pulses have low amplitude, 0.12 VEM on average, and it is their density, rather than their individual amplitude, that increases with that of the main signal. Depending on the threshold set on defining significant signals, some of such after pulses will be identified as

such and some will be blamed on the base line. Where to set such a threshold depends on what one is after. For the time being, it is sufficient to be satisfied with the improvement illustrated in Figure 2.4 and to keep in mind that the bump displayed in Figure 2.5 is of a different nature than the underlying undershoot. The point will be taken up again in Section 4.2.1 when discussing stopping muon decays.

2.1.3 Photoelectron statistics

A preliminary estimate of the contribution to the uncertainty of photoelectron statistics, namely of the term proportional to the square root of the charge, gives 50 photoelectrons per VEM. As it is known from other sources that this number is rather of the order of 100, it would be interesting to attempt a direct evaluation from the data. However, such an evaluation is made particularly difficult because the mode of operation of the SD PMTs is far from being optimized for such a task and there is essentially no hope to resolve a single photoelectron peak from multiple photoelectrons.

The analysis presented here uses the baseline correction described in the preceding section and raw ADC counts (i.e. before conversion to VEM). In each trace, the segment that carries the larger charge, called the “strong segment”, will be found to determine the level of disturbance caused by after pulses and other related effects. In general, traces having the strong signal charge in excess of 20 VEM or the total charge in excess of 40 VEM have been excluded from the analysis. Two-bin isolated segments are selected with the idea that they should include some low intensity signals having a photoelectron (either emitted by photoelectric effect or by thermal noise) origin. A single photoelectron should easily fit within a few 10 ns (48 ns is amply sufficient for a < 20ns LED pulse in Reference 61): there is no need to consider segments wider than 2 bins. Isolation is obtained by requiring that for each PMT, in each of the two bins preceding and following the segment there be less than 3 ADC counts (this is therefore a requirement on 12 numbers being less than 3). The condition on the presence of a signal requires that the content of each of the other two PMT segments exceed some threshold S_{thr} , a requirement that should not introduce any bias. The segment content of PMT i , S_i , is defined as the sum of the contents of the two bins that make up the segment. Namely S_i is considered if, and only if, $S_m > S_{thr}$ and $S_n > S_{thr}$ with i , m and n being a permutation of 1, 2 and 3. The threshold S_{thr} is chosen large enough to suppress the noise but not too large not to suppress the 1 photoelectron

component. Moreover early segments (below bin 200) where there is only noise are distinguished from late segments (above bin 260) where there may be a signal.

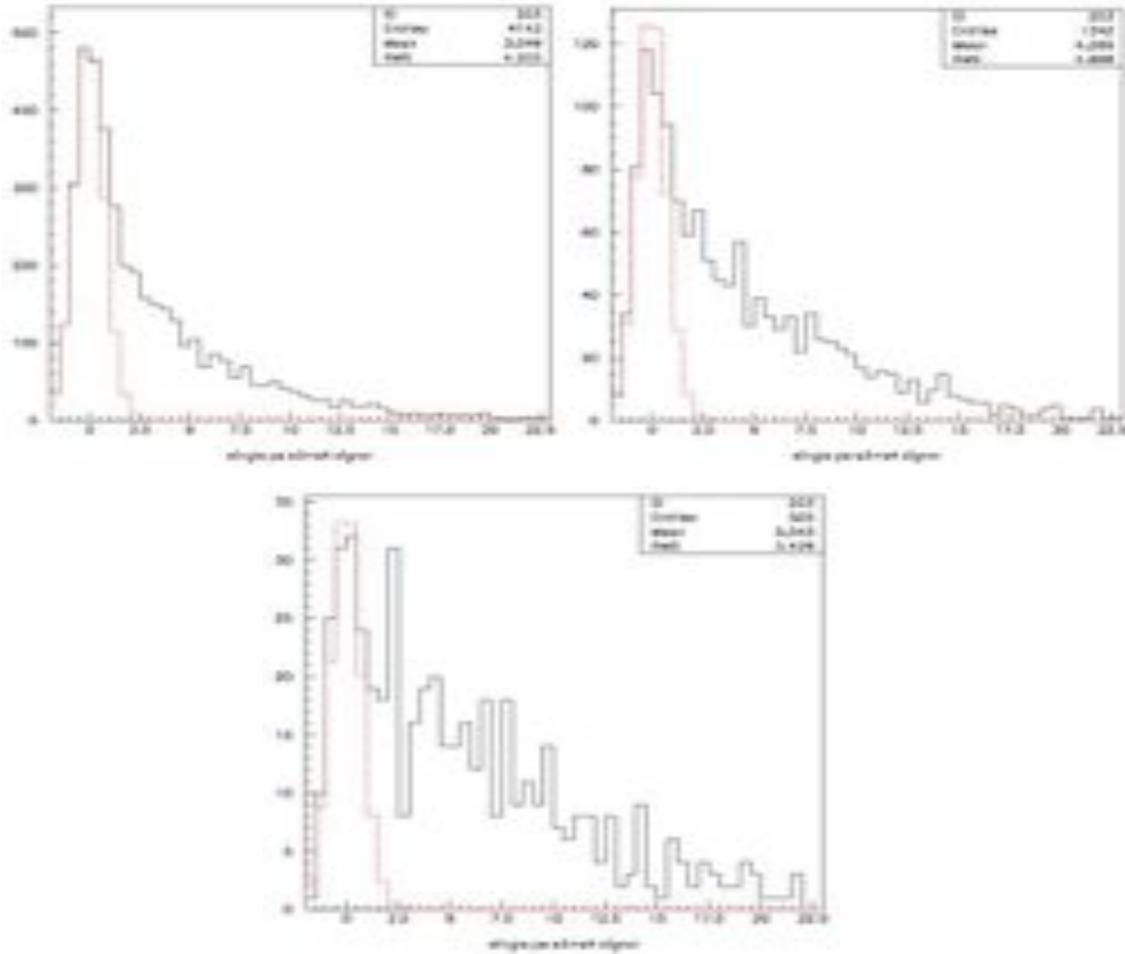


Figure 2.6. S_i spectrum for isolated segments above bin 260 having $S_{m,n} > 4$ (upper left), 6 (upper right) and 8 (lower panel). The 0 photoelectron contribution (normalized in the negative region) is superimposed.

Figure 2.6 shows the distribution of S_i for late segments and $S_{thr} = 4, 6$ and 8 ADC counts respectively. In these figures, S_i has been corrected for the PMT gains – that fluctuate significantly as shown in Figure 2.7 – by multiplying it by the VEM/count value for that PMT and dividing by the mean value, 0.0054. There are 21688 isolated segments in the noise (early) region and 53032, namely a factor 2.445 more, in the signal (late) region. Using this as a normalization the noise contribution corresponding to traces where there is no light in the tank but where PMTs m and n fluctuate in noise enough to pass the cut is evaluated. This evaluation is an overestimate to the extent that there is relatively less room available for noise segments in the signal region; but in fact it accounts for only a

small fraction of the 0-photoelectron peak in Figure 2.6. Indeed, when there is light in the tank reaching PMTs m and n , it may very well be that no photoelectron be emitted in PMT i if none of the light reaches PMT i or if no photoelectron is produced (the quantum efficiency is of the order of 20%).

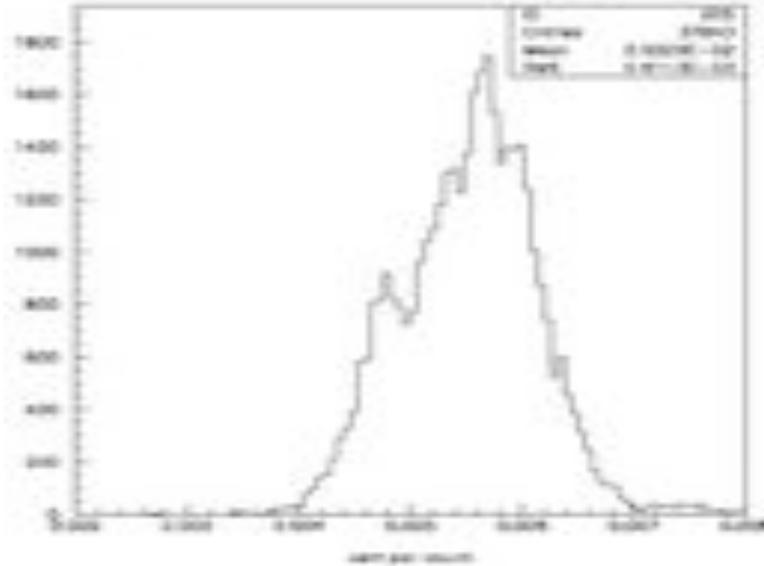


Figure 2.7. Distribution of the (VEM/ADC count) values showing the dispersion of the PMT gains.

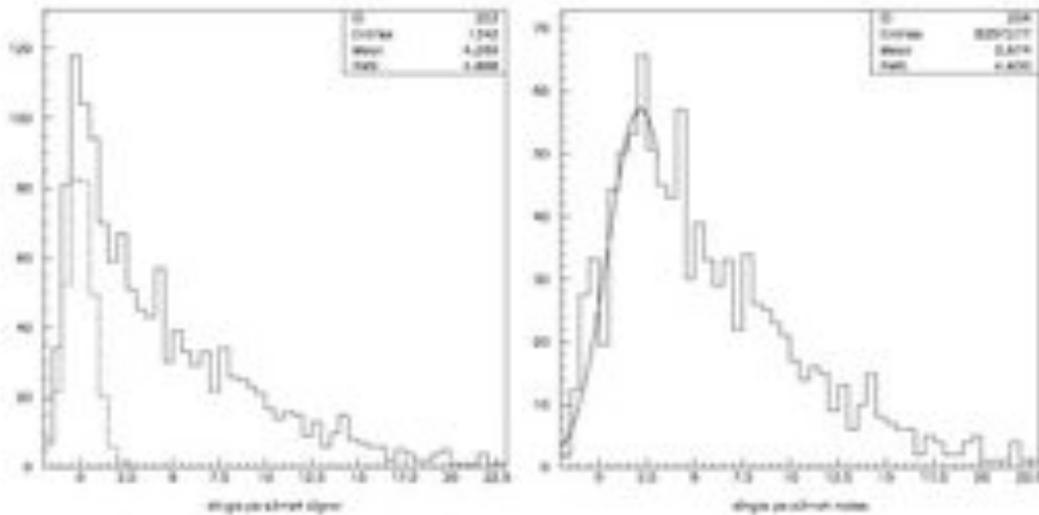


Figure 2.8. Left: S_i spectrum for isolated segments above bin 260 having $S_{m,n} > 6$. The normalized 0 photoelectron contribution is superimposed. Right: noise-subtracted (normalized in the negative region) S_i spectrum for isolated segments above bin 260 having $S_{m,n} > 6$. Also shown is the Gaussian fit of the rising edge.

The shape of the 0-photoelectron contribution is the same whatever it is caused by (no light in the tank but PMTs m and n fluctuating in noise or light in the tank but no photoelectron reaching the dynode chain of PMT i). It is only due to the noise that is causing baseline fluctuations: its shape is known and can be subtracted as long as its amount is also known. It is observed to be Gaussian, centred at 0 with an rms of 0.75 ADC counts, namely a factor $\sqrt{2}$ larger than for a single bin as could have been expected. After subtraction by scaling the noise spectrum in the negative region, one obtains spectra such as that shown in Figure 2.8.

The single photoelectron spectrum is commonly expected [61,62,63] to be approximately Gaussian (with rms/mean of the order of 2/3 in Reference 62) with the idea that once a photoelectron reaches the first dynode, the fluctuations of the PMT gain should be dominated by statistics in the multiplication process. Here, the Gaussian should be slightly broader than in Reference 62 because of small errors on the baseline evaluation and digitization in broader bins: its rms should reach at least 1.5 ADC counts. Using a threshold of 6 ADC counts (the result is insensitive to this choice) the 0-photoelectron contribution, accounting for a 20% tail of the single photoelectron contribution below it, is best fit by a Gaussian having a mean of 2.1 and an rms of 1.7 ADC count (Figure 2.8). This results in a single photoelectron response peaking at 2.1 ± 0.3 ADC counts, corresponding to 88 ± 12 pe/VEM. The uncertainty of ± 0.3 ADC counts is estimated from the dependence of the result on various parameters (amount of noise subtraction, interval over which one fits the leading edge, edge effect in the 50 ns window, etc...).

However, recent measurements [64] of the dependence of the PMT response on the photon impact on the photocathode have shown that most of it is due to the first dynode (which is of the foil type). The single photoelectron spectrum is therefore the sum of many Gaussians having different mean values depending on the location of the photon impact on the photocathode. This has the effect of significantly increasing both the tail of the single photoelectron spectrum below the 0-photoelectron peak and the width of the single photoelectron response. Both imply that the result quoted above is an underestimate, by a significant amount, of the number of photoelectrons per VEM. The result is in agreement with that of more direct, and probably more reliable, measurements of the same quantity.

As previously mentioned, from the dependence on their mean S_{mean} of the rms of the three PMT signals (Figure 2.9) one can infer a S -dependent contribution to the uncertainty. A quadratic fit gives $\Delta^2 = 0.66 E^{-5} + 0.0225 S_{mean} + 0.00104 S_{mean}^2$ implying

that the photoelectron statistic term corresponds to 46 photoelectrons per VEM. This apparent discrepancy is again probably the result of the non Gaussian response of the first dynode, its dependence on the photon impact causing a very significant broadening.

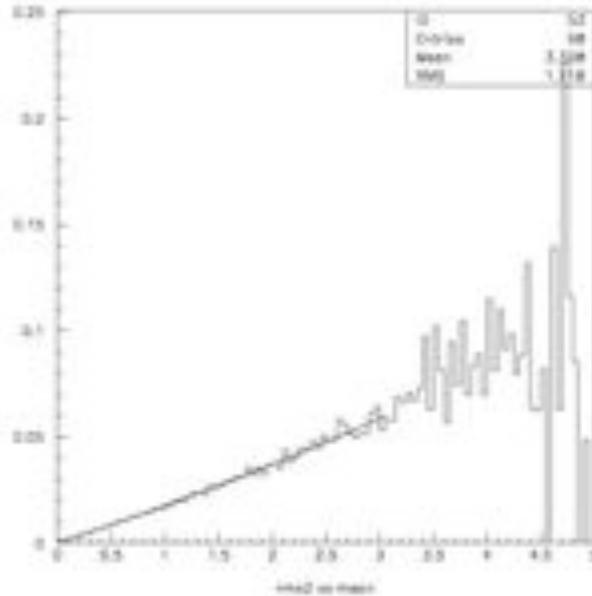


Figure 2.9. Dependence of Δ^2 on S_{mean} (see text)

Having evaluated the uncertainty attached to the charge measurement in each time bin of the FADC traces, a quantitative assessment of the relative consistency of the three FADC traces of a same tank has become possible.

2.1.4 After-pulsing

The present section deals with one-bin segments, where segments are defined as sets of successive bins having all three PMT signals in excess of 0.015 VEM. Calling S_{max} the largest of the three PMT signals in that bin, its distribution is shown in Figure 2.10 (left). It is strongly peaked at low values where the tail of the noise is visible. Moreover, it does not exceed half a VEM. Figure 2.10 (right) shows the distribution of the signals in the two other PMTs as a function of S_{max} . While S_{max} extends up to nearly 0.5 VEM, the other two PMT signals remain very low, giving evidence that one-bin segments are dominated by a single PMT.

As a measure of the asymmetry between the three PMT charges, a quantity $\chi^2_{spike} = (x_{max} - x_{else})^2 / (\Delta_{max}^2 + \Delta_{else}^2)$ is introduced, x_{max} being the largest of the three charges and x_{else} the average of the two others, Δ being the associated uncertainties. "Spikes" are defined as having χ^2_{spike} in excess of 3.5 and correspond to bins where one PMT charge exceeds clearly the other two.

Cutting the tail of the noise by requiring $S_{max} > 0.1 \text{ VEM}$, Figure 2.11 provides evidence for one-bin segments to be related with very strong signals, suggesting that they are due to after-pulsing.

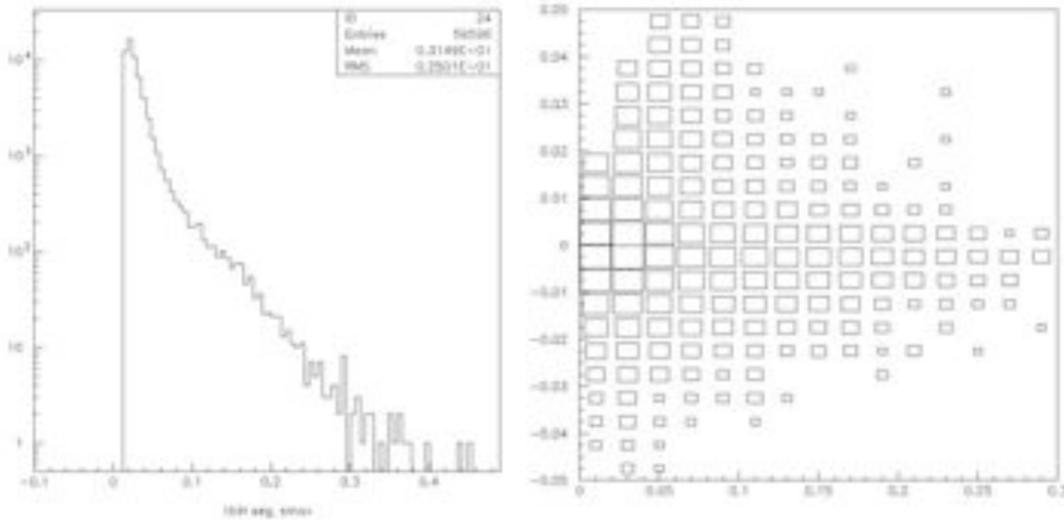


Figure 2.10. One bin segments. The distribution of S_{max} , the largest charge of the three PMT signals, is shown on the left panel. The dependence on it of the charges of the two other PMT signals is illustrated on the right panel.

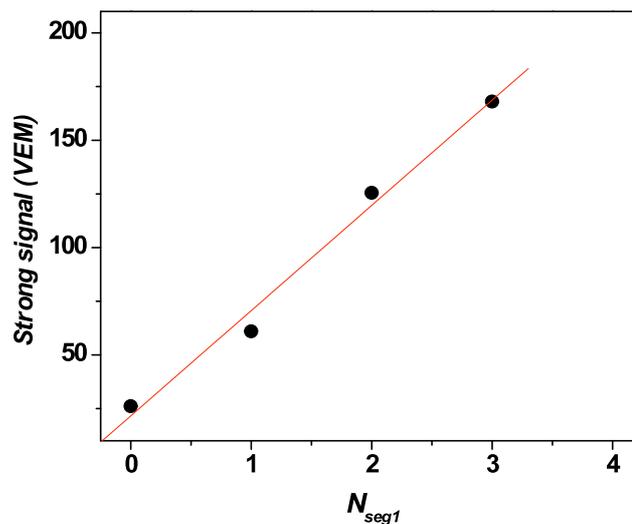


Figure 2.11. Dependence of the strength of the total signal on the number of one-bin segments.

While after pulses are indeed expected to occur [65] the question is now whether all observed spikes are compatible with being due to after pulses.

In order to pursue the study in cleaner conditions, segments are redefined as made of successive bins where each PMT has a signal exceeding 0.025VEM, (i.e. some 5 ADC counts and 2 to 3 photoelectrons), avoiding artificial segments caused by noise fluctuations. To cope with cases where one of the three PMTs starts later recording light than the two others, causing the segment to start after the two other PMTs are on, the bin preceding the segment is included in the calculation of the segment charge or other similar quantities. Traces having one or more segments having a charge in excess of 20 VEM are ignored (this implies the rejection of 13% of all traces). In addition, the one-bin segment is required to be isolated (each PMT in the two bins preceding and the two bins following the segment must contain less than 0.015 VEM) and the spike charge is required to exceed 0.07 VEM. Moreover, no segment in the trace should have a charge in excess of 20 VEM.

The spike charge distribution is shown in Figure 2.12 (left). It rarely exceeds 0.2 VEM and has a mean of only 0.12 VEM. The location, with respect to the last bin of the strong signal segment, of the bins explored for a possible spike (namely satisfying the isolation criterion but not subjected to the cut on χ^2_{spike}) is displayed in Figure 2.12 (right). It is essentially flat. The location, defined in the same way, of the bins that satisfy the χ^2_{spike} selection is shown on the same figure. It has a distribution that shows instead significant structure around 100 and 300 bins after the end of the strong signal segment (namely 2.5 and 7.5 μs respectively). The second of these bumps corresponds to the accumulation of after pulses mentioned earlier (Section 2.1.2 and Figure 2.5). The probability to have a spike in a given bin is only $1.8 \cdot 10^{-4}$ and the probability to have a spike in a given trace is 11%. Moreover the present spikes are found to be correlated with the strong signal not only in number and position but also in charge. This is illustrated in Figure 2.13 that displays the distribution of the charge of the strong signal for all isolated bins and for spiky bins separately: the latter are associated with strong signals having a high charge.

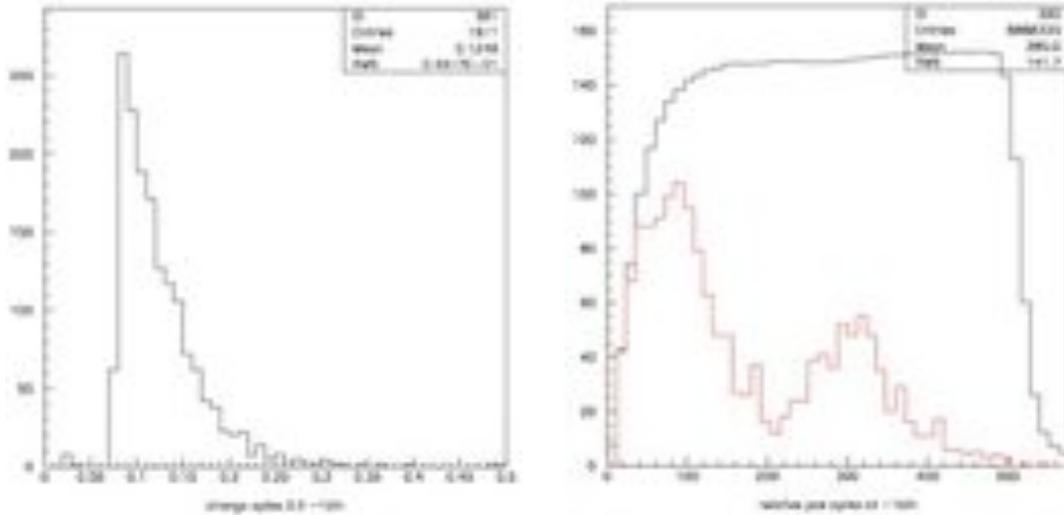


Figure 2.12. Spiky one-bin segments. Left: distribution of their charges. Right: distribution of their time distribution with respect to the strong signal (red). The time distribution of all isolated one-bin segments (including spiky) is shown as comparison.

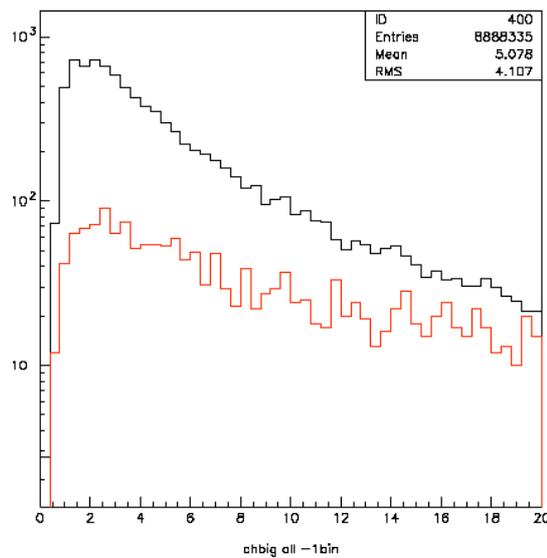


Figure 2.13. Distributions of the charge of the strong segment associated with spiky one-bin segments (red) and that associated with any isolated one-bin segments.

In conclusion, all above features confirm that the spiky one-bin segments are due to after-pulsing. Their spikes occur with a low probability, their charges are small, they show correlations with the strong signal in number, in charge and in timing. They are probably due to photoelectrons ionizing the residual PMT gas and producing ions returning to the photocathode or elsewhere in the dynode chain but are of a different nature than the after-pulses observed with low illuminations and occurring at the microsecond scale (Figure

2.14), which are mostly due to secondary electrons emitted toward the photocathode and turning back to the first dynode.

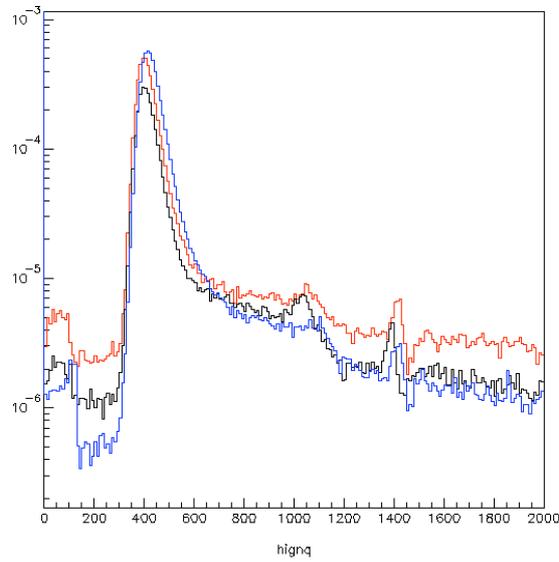


Figure 2.14. Time distributions measured for single photoelectrons impacting the photocathode on different locations [64]. The horizontal scale is TDC bins of 0.1 ns.

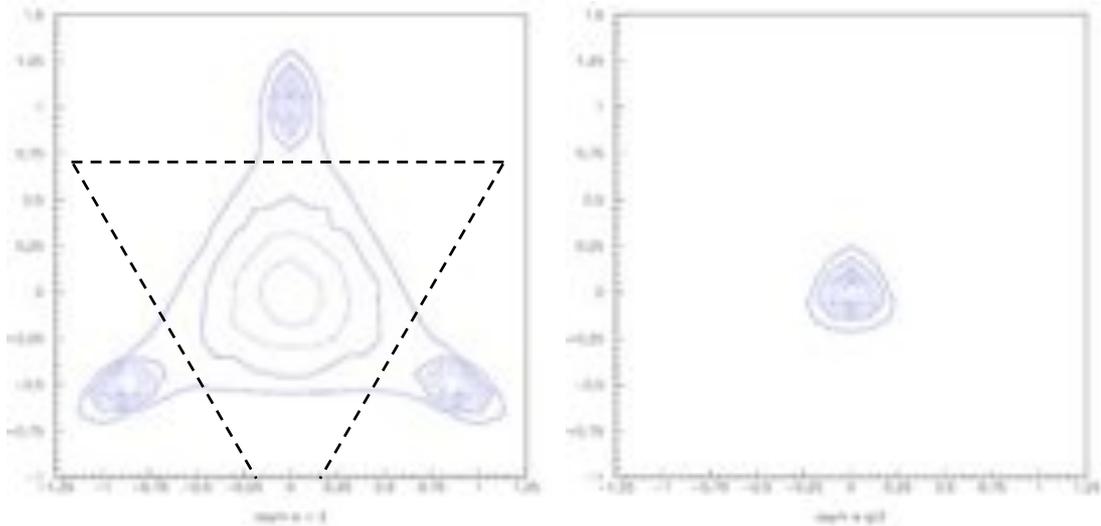


Figure 2.15. a) Left: Trilinear plot of the PMT asymmetries as defined in the text for one-bin segments. The dashed triangle corresponds to the cut applied. b) Right: Same as (a) for wider segments.

After-pulses being short and occurring in only one of the three PMTs at a given time are easy to identify. Figure 2.15 shows the ratios of the PMT charges to the average charge for one-bin segments and for wider segments respectively. Precisely, what is plotted are the three ratios $r_i = q_i / (q_1 + q_2 + q_3)$ where q_i are the segment charges defined separately

for each PMT. As $r_1+r_2+r_3=1$, it is convenient to plot these in trilinear coordinates. The difference between one-bin segments and wider segments is striking. Any one-bin segment having at least one of its r_i in excess of 0.7 is from now on rejected as being caused by after-pulsing.

2.1.5 Spikes

If after-pulsing occurs, there is no reason for it to take place in between real signals exclusively: it may very well fall on top of a real signal. The search for bins where one of the PMT charges exceeds that of the other two is now extended to segments more than one bin wide. Searching for such spikes, one finds oneself swamped by events where nearly all FADC bins satisfy the spike condition. These are events where a very “strong signal” occurs at the beginning. At later times on the FADC trace the signals of the three phototubes are essentially uncorrelated and even seem to have different baselines. Typically, the early “strong signal” exceeds 100 VEM at maximum and the later traces differ not only in their patterns but also in their baselines, often by more than 0.05 VEM. It looks as if the PMTs had been sufficiently disturbed by the high current at the beginning of the trace to have become unreliable to the level of a fraction of a VEM, say one half or so, at later times.

Accordingly, in what follows, as was done earlier, traces having one or more segment having a charge in excess of 20 VEM are ignored.

For each segment bin (including the bin preceding the segment) a quantity χ^2_{bin} is defined as the χ^2 of the three PMT charges to their mean, from which 2, the number of degrees of freedom, is subtracted. In doing so the error formula³ given previously, $\Delta^2 = 0.66 E - 5 + 0.0225 S_{mean} + 0.00104 S_{mean}^2$, is being used.

³ In fact, a single error formula is used to describe two different situations. In principle, one should distinguish between a χ^2 to the hypothesis that there was no light in the tank, hence all three PMTs should measure zero within the baseline fluctuation of 2.5 mVEM, and the χ^2 to the hypothesis that there was light in the tank, hence the PMTs see a number of photoelectrons having a Poisson distribution of a given mean. However, many complications come in addition to this oversimplified picture (eg timing and geometry differences between the three PMTs). It is therefore reasonable, if not strictly correct, to use an approximate χ^2 method but one should be careful to estimate the mean value of the signal correctly, using a *same* error for each PMT signal, namely writing $Mean(x) = (x_1 + x_2 + x_3)/3$ and not $(x_1/d_1^2 + x_2/d_2^2 + x_3/d_3^2)/(1/d_1^2 + 1/d_2^2 + 1/d_3^2)$ where the d s would have been calculated using the standard formula applied to the values of x separately. When calculating χ^2 one then must use as error the value given by the standard formula applied to $Mean(x)$ as defined above. Not doing so might introduce important errors.

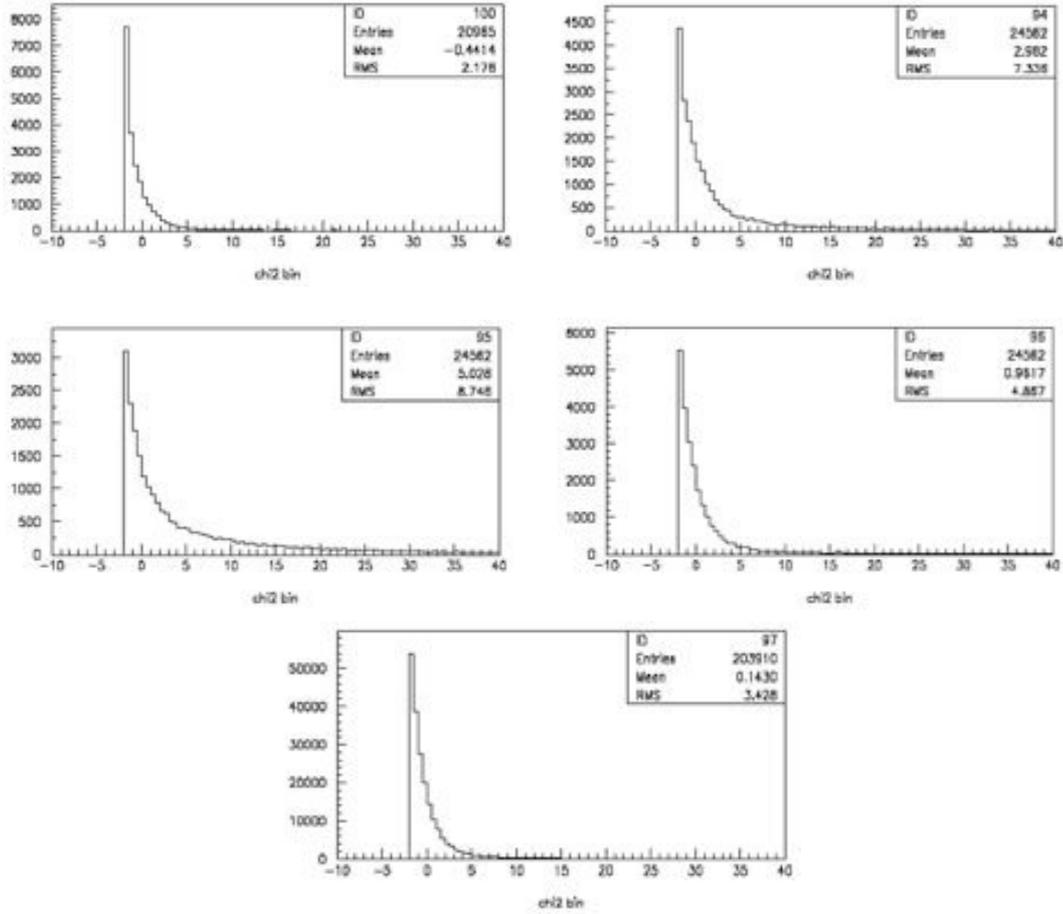


Figure 2.16. χ^2_{bin} distributions (shifted by 2 units, see text). From left to right and from top to bottom: second and first bins before the segment⁴, first and second bins in the segment, all other bins in the segment. Their respective means are: -0.4 , 3.0 , 5.0 , 1.0 and 0.1 ; their respective rms values are: 2.2 , 7.3 , 8.7 , 4.9 and 3.4 .

The χ^2_{bin} distribution is displayed in Figure 2.16 for five cases separately: the two bins preceding the segment⁴, the first and second bins of the segment, and the other bins in the segment. On average, χ^2_{bin} is significantly larger in the bin preceding the first bin and the first bin itself than elsewhere. Comparing the values of χ^2_{bin} in these two bins shows that when one is large, the other tends to be small. Simple attempts at explaining the origin of such spikes as a result of small timing differences between the PMTs of a same tank are failing. In particular, if such shifts were causing the spike, one would expect the charge of the spiky PMT segment to be equal to that of the corresponding segments in the two other PMTs. But such is not the case.

⁴ The earliest of these two bins may belong to a preceding segment, in which case it is excluded. There is a significant number of such cases, where two segments are separated by only one bin.

On the contrary, for spike bins, the spike charge, $x_{spike} = x_{max} - x_{else}$, comes in addition to the segment charges of each individual PMTs. This shows that the effect is not caused by small timing differences between the three PMTs, which would leave the segment charges of each individual PMT invariant. This important result is illustrated in Figure 2.17 where the segment charges of the three PMTs are shown separately after subtraction of the spike charge from the PMT labelled *max*. Their averages (2.9, 3.0 and 3.0) are equal while the average charge distribution of the spike, amounts to 0.5 VEM.

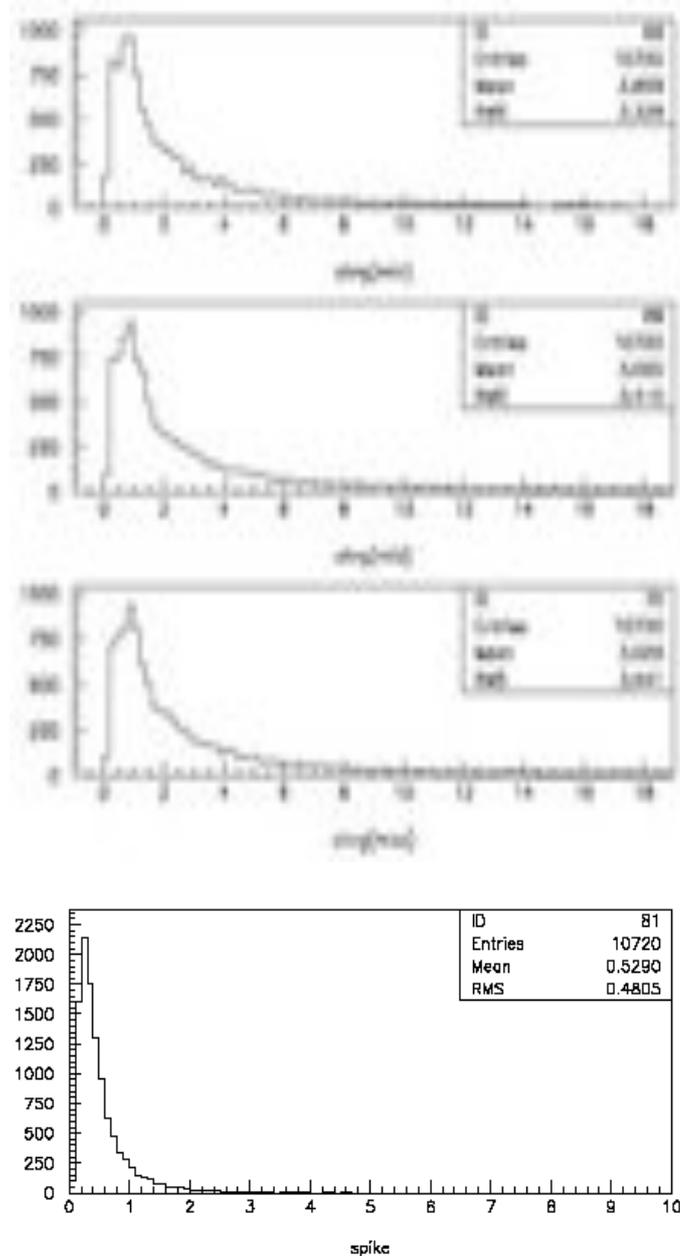


Figure 2.17. Upper panels: segment charges (in units of VEM); the *min*, *mid* and *max* distributions are shown separately (top to bottom); the spike charge has been subtracted from the *max* charge. Lower panel: spike charge distribution for the same segments.

This has been confirmed by a number of additional checks which confirm that in a significant fraction of the segments (when cutting $\chi^2_{spike} > 3.5$, 44% of all segments or 68% of the segments having the largest value of χ^2_{bin} in the bin preceding the segment or in the first bin of the segment) the bin showing the largest disagreement between the three PMTs (i.e, having the largest value of χ^2_{bin}) occurs in the bin preceding the segment or in the first bin of the segment. It is due to the presence of a spike in one of the three PMTs.

All features displayed by the spikes considered in the present section are consistent with what can be expected from the early collection of light by one of the PMTs. This phenomenon has been studied [66] using a simulation of the tank optics in cases where the light source is the Cherenkov radiation of a feed-through relativistic muon. They show that light reaching the PMTs in the first 50 ns (called direct if it reaches the PMTs without being diffused by the tank walls and semi-direct otherwise) is asymmetrically distributed among the three PMTs. What happens is that it takes some time for the light to be randomized by multiple diffusion on the tank walls and a typical path length between successive diffusions corresponds to some 10 ns. Once this is achieved, all three PMTs receive equal amount of light. However, at the beginning, when this randomization is not yet completed, light aims preferentially at one (or possibly two) of the PMTs generating the observed spike.

Several detailed checks have been performed, which validate this interpretation. It is sufficient, here, to report on one of these, the existence of a strong correlation between the shower azimuth and the PMT asymmetry.

The azimuth ψ_i of the three PMTs of a same tank [66] are given in the table below:

i	1	2	3
ψ_i ($^\circ$)	-150	90	-30
$\cos\psi_i$	$-\sqrt{3}/2$	0	$\sqrt{3}/2$
$\sin\psi_i$	-1/2	1	-1/2

The study compares signals measured in any bin having χ^2_{bin} in excess of 3.5 and $x_{max} - x_{min}$ in excess of 0.2 VEM to a form $\sigma_i = a + b \sin\theta \cos(\psi_i - \varphi)$ where θ and φ are the zenith angle and azimuth of the shower axis and the σ_i 's are the ratios between the

PMT signals ($i=1, 2, 3$) in the bin considered and the total charge S_0 of the signal, evaluated by summing the charges collected in the three time bins following the large χ^2_{bin} bin⁵.

Using the above parameterization, a , $b\sin\theta$ and φ are calculated from the data and compared with the values obtained from the tank timings. Explicitly,

$$a = (\sigma_1 + \sigma_2 + \sigma_3)/3 \quad u = (2\sigma_2 - \sigma_1 - \sigma_3)/3 \quad v = (\sigma_3 - \sigma_1)/\sqrt{3}$$

$$b\sin\theta = \sqrt{u^2 + v^2} \quad \varphi = \text{Atan}(u/v)$$

Figure 2.18 shows the distribution of the difference between φ calculated in this way and φ measured from the tank timings. The correlation is strong but in many cases the two values do not match (the mean value is less than 1° but the rms value is nearing 90°). The two values of φ are not expected to be exactly the same: one is the azimuth of the shower axis, the other is that of the particles incident on the tank, and the divergence of the shower is significant. Yet, this effect can be ignored at the present level (it will be extensively exploited in Part 3).

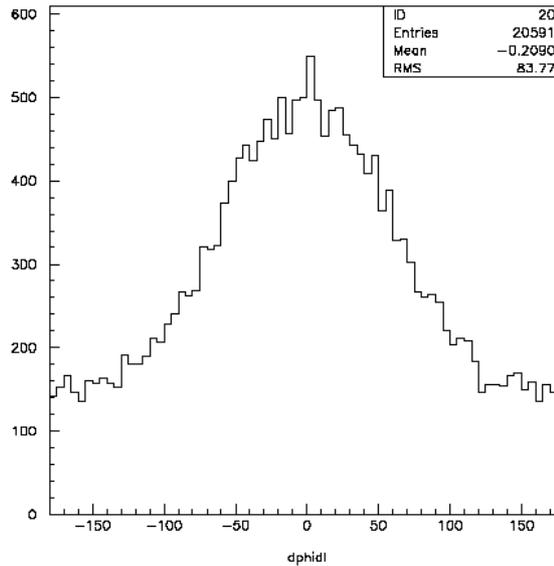


Figure 2.18. Distribution of the difference between the shower axis azimuths measured from the PMT asymmetry and from the tank timings.

In conclusion, it has been established that most, and probably all, occurrences of a bad χ^2_{bin} are associated with an asymmetric sharing of the early (<50ns) light of any signal.

⁵ We ignore cases where less than three bins follow the large χ^2_{bin} bin in the segment. We also ignore cases where any of these three bins has a χ^2_{bin} in excess of 3.5. Both correspond to very few cases.

The asymmetry is strongly correlated with the azimuth of the shower axis, but not strongly enough to allow for a one-to-one correspondence between these two quantities: other factors – among which the location of the impact point on the tank – smear very significantly the azimuth-asymmetry correlation.

The revealed asymmetry may be studied for its own sake or be made use of in practical applications. One of these is the study of the shower divergence, which is the subject of Part 3. Another exploits the idea that spikes might be used to tag the beginning of segments and help the resolution of overlapping signals into their individual components. It has been the subject of an extensive study, which, however, concluded that such a hope was too ambitious: it will be briefly explored in Section 2.2.4. Here, it is sufficient to mention that all spikes occurring within segments are consistent with being due to overlapping signals or, in rare occasions, to after-pulsing.

2.1.6 Summary

The preliminary studies reported above have established a number of important facts which will be made use of in subsequent studies.

The uncertainty attached to each FADC bin receives contributions from three sources:

- a Gaussian noise of 0.0028 VEM (approximately half uncorrelated and half fully correlated between the three PMTs) appearing as an effective Gaussian noise of 0.0032 VEM because of the quantization in ADC counts (0.0054 VEM).
- a photoelectron statistics contribution, well described by a Gaussian with $\sigma^2=0.023 S$ (VEM), S being the bin content in VEM, and corresponding to 46 ± 7 photoelectron. The discrepancy between this number and direct evaluations (of the order of 100 pe/VEM) implies another source of broadening than photoelectron statistics and was interpreted as being the result of the dependence of the response of the first PMT dynode on the photon impact on the photocathode. Indeed, an attempt to reveal the single photoelectron peak has been presented, establishing a lower limit of 88 ± 12 pe/VEM.

The expression retained for the uncertainty on a bin content S is therefore:

$$\Delta = \{ 0.66 E^{-5} + 0.0225 S_{mean} + 0.00104 S_{mean}^2 \}^{1/2}$$

Evidence has been found that, in the late part of the FADC trace, the baseline is sometime getting negative, the more the higher the charge measured in the early part of the FADC trace, the presence of a very strong signal at the beginning of the FADC trace affecting severely the remaining of the trace (not only by shifting the baseline but also by producing erratic peaks at the level of a fraction of a VEM).

A new baseline correction has been evaluated that gives good results over the whole trace except for those that have a very strong early signal and should anyhow be excluded from searches relying on a proper understanding of low charge signals.

A systematic search, free of a priori presumptions, of inconsistencies between the traces of individual PMTs of a same tank has shown that they were restricted to two types, each of which was previously known:

- A first family includes small (0.12 VEM on average) isolated spikes occurring preferentially 2.5 and 7.5 μ s after a strong signal and with a low probability (11% per trace). They occur in a single PMT at a given time and are strongly correlated in number and time with the charge of the strong signal. They are interpreted as after-pulses and are essentially harmless as being very easily identified.

- A second family includes larger spikes (0.5 VEM on average) occurring at the beginning of signals induced by inclined showers in the PMT that receives the first light. They come in addition to the signals recorded in the other PMTs. Indeed, most, and probably all, occurrences of a bad χ^2_{bin} are associated with an asymmetric sharing of the early (< 50ns) light of any signal. The asymmetry is strongly correlated with the azimuth of the shower axis, but not strongly enough to allow for a one-to-one correspondence between these two quantities: other factors – among which the impact point on the tank – smear very significantly the azimuth-asymmetry correlation.

2.2 Pattern recognition

2.2.1 Introduction

Many SD studies, in particular those concerned with the primary mass composition, imply an evaluation of the muon fraction on ground.

Resolving the FADC traces of the PAO surface detector into individual signals is one of the prerequisites to the separation of the muon and electron-photon components of a shower. It is therefore of major interest to the analysis of the PAO data as a step toward the identification of the primaries. Many approaches have been investigated so far, including subtraction of the exponential decay tails [67], standard pattern recognition techniques [68,69] or searches for upward jumps [70,71]. Other approaches make use of additional information, such as using the fact that muons are expected to be associated with higher statistical fluctuations than electron-photons [72] or to arrive earlier than electron-photons [73]. Such information may of course be used in addition but the present section addresses exclusively the problem of pattern recognition. The approach presented here borrows much from such earlier studies, in particular from Reference 67, however refining it and adding information from the early time asymmetry of the responses of the three PMT's of a same tank to a common light source. It makes maximal use of the information available from the FADC data alone. It is the outstanding optical qualities of the Cherenkov tanks, with a typical decay time of 75 ns, implying some 8 successive wall diffusions on average, and the low noise performance of the associated electronics, that make this possible [74].

The subtraction algorithm presented in [67] relies on the fact that any light produced in a Cherenkov tank decays slowly as the result of three possible main causes: absorption in the walls, absorption in water and escape into one of the PMT photocathodes. The resulting decay time is a property of the tank: it varies only little from event to event as light is rapidly randomized. Therefore any amount of light collected in a given FADC bin (25 ns wide) must be followed in the successive bin by light having the same origin, scaled down by a constant factor f depending exclusively on the optical properties of the tank. Any light observed in addition must have a different origin. Subtracting from each bin the content of the preceding bin scaled down by f should therefore resolve the trace in a set of individual one-bin peaks⁶, each associated with a different source. The resulting

⁶ However, as the signal does not start exactly at the beginning of a bin, one expects in fact the trace to be resolved in a set of peaks made of a pair of adjacent bins. This matter will be addressed later on in some detail.

algorithm should not be seen as a mere mathematical trick but as accounting for basic physical properties of the tank optics.

Any new source of light occurring inside a tank should therefore be associated with an upward jump in the FADC trace and therefore with a peak after subtraction using the algorithm presented in [67]. Another feature characteristic of the appearance of a new light source is a strong early time asymmetry between the responses of the three PMT's [66,75]. This asymmetry was shown in the preceding section to be strongly correlated with the azimuth of the trajectory of the particle(s) which are causing the appearance of new light. Its occurrence is a very general feature and may therefore be used to tag new signals and possibly help the algorithm of Reference 67.

Limitations related to the non-zero width of the FADC bins are inherent properties of the detector and cannot be overcome. If several signals appear within a same bin they cannot be resolved in any way. Moreover, when two peaks appear in successive FADC bins after application of the subtraction algorithm of Reference 67, it is in general not possible to decide whether they should be associated with a single signal or with two or more. The effects of such limitations can only be taken into account on a statistical basis, but not on an event by event basis.

2.2.2 Preliminary data reduction and selection

The data used in the present study⁷ to illustrate the method presented have been selected as obeying the offline trigger. The comparison with the results of Reference 67 uses a total of 31 files selected in the same way and covering the whole month of May 2006. In general, dynode data are used; only when they show saturation are anode data used instead. Tanks having one or more PMTs tagged as not working properly are rejected.

The base-line correction presented in the preceding section is applied in order to help the separation of significant signals from noise fluctuations. At the present stage traces having a very large charge collected in the early bins (in tanks close to the shower core) are excluded, the cut used for this selection excluding 13% of the tanks.

As discussed in the preceding section, each bin charge q_{ij} (measured in FADC bin i of PMT j) is associated an uncertainty $\Delta q_{ij} = \sqrt{\{0.66E-5 + 0.0225q_i + 0.00104q_i^2\}}$ where q_i

⁷ Using data from *sdt4_2006_05_01[and 2, 3, 4 and 5]_12h00.root*

is the arithmetic mean (not weighted) of the three PMT charges collected in bin i (in VEM units, remember that one FADC count corresponds to 0.0054 VEM on average).

Segments are defined in each tank trace (meaning the arithmetic mean, bin by bin, of the individual PMT traces) as sets of consecutive bins having charges in excess of 0.025 VEM in each of the three PMTs. There are on average 1.7, 0.9 and 2.4 segments per event having a width of 1, 2 or ≥ 3 bins respectively. The former are due either to noise fluctuation or to after pulsing [65,76] and can be safely ignored by rejecting segments being less than 3 bins wide. In the remaining of this section “segment” stands for “segment being at least three bins wide”. The width distribution of such segments has mean and rms values of 10.8 and 7.6 bins respectively. The zenith angle and energy distributions of the events used in the present study are shown on Figure 2.19.

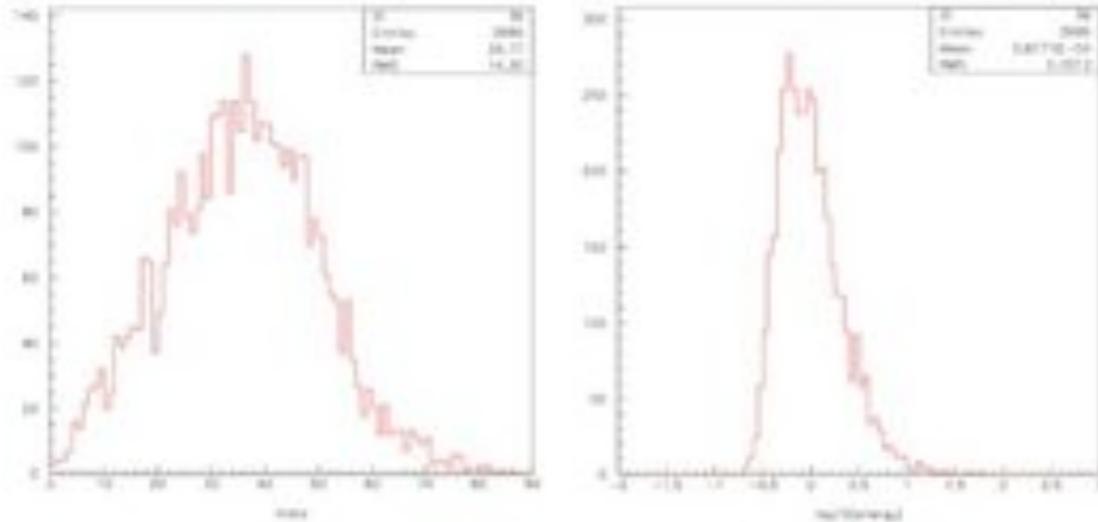


Figure 2.19. Left: zenith angle distribution (in degrees) of the event sample used in the present section. Right: distribution of the decimal logarithm of the energy (EeV) of the event sample used in the present note.

2.2.3 Subtraction algorithm

The subtraction algorithm of Reference 67 implies the knowledge of the light decay time, or equivalently of the factor f introduced in the preceding paragraph. Such a factor is the same for the three PMT’s of a same tank and for the whole time span of the FADC trace. Too large a value of f results in excessive subtraction and generates FADC bins having negative charges. On the contrary, too small a value of f results in excess charges in most FADC bins. One may therefore evaluate f in each tank by maximizing the number of

bins having zero charge. This is done by maximizing the sum Σ (over all bins of all segments of a same tank trace) of the significance levels associated with the χ^2 of each bin to the hypothesis that it is empty after subtraction. As all bins have the same width, the maximization procedure finds the value of f giving the largest time span consistent with zero charge. It is not obvious a priori that the data are of sufficient quality to allow for such a procedure to work, but they turn out to be.

Precisely, for bin i , $\chi^2_{i|empty} = (q_i - fq_{i-1})^2 / (\Delta q_i^2 + f^2 \Delta q_{i-1}^2)$. The significance level to be maximized is $\Sigma = \sum_i [1 - \int_0^u \exp(-x/2) dx / \sqrt{2\pi x}]$ where the integral spans from 0 to $u = \chi^2_{i|empty}$.

Then, $f = \sum \{ \exp(-u/2) / \sqrt{u} s_i s_{i-1} / (\Delta q_i^2 + f^2 \Delta q_{i-1}^2) \} / \sum \{ \exp(-u/2) / \sqrt{u} s_i^2 / (\Delta q_i^2 + f^2 \Delta q_{i-1}^2) \}$, where s is the bin content after subtraction.

The maximization is made in four iterations, which is sufficient to achieve convergence. Figure 2.20 shows the distribution of Σ for the average value $f = 0.74$ and, superimposed on it, for the particular value of f obtained for each tank.

On average, the mean value of Σ has increased by only 3% after the optimization procedure. Figure 2.21 shows the distribution of the values of f obtained after optimization⁸. It is remarkable that they are confined within a narrow range, 0.74 ± 0.16 with an rms value of 0.04 . Having fine-tuned the value of f for each tank in each event separately, the effect of fluctuations on the subtraction algorithm can now be studied.

⁸ The values of f obtained in this manner are larger than those obtained from the inclusive muon signal collected in between events. The latter distribution for the period under consideration shows two distinct peaks, one at 0.66 and another one, three times smaller, at 0.70. The distribution is quite narrower, with an rms of only 0.02 compared to 0.04 here. There is no correlation between the values of f evaluated in these two different ways. Using the values of f from the inclusive muon signal gives a mean value of Σ which is 9% smaller than that shown in Figure 2.21 (6% smaller than that using $f=0.74$ for all tanks). It must be noted that in 18% of the cases the inclusive value of f is not available: in the comparison the corresponding events have been ignored. It must be noted that slightly smaller f values are found far from the shower core than close to it, reflecting the fact that when many signals overlap there is a trivial bias toward large f values. This effect is of little relevance to the present work. Whenever useful, a cut was applied on events having a f value more than three rms away from the mean.

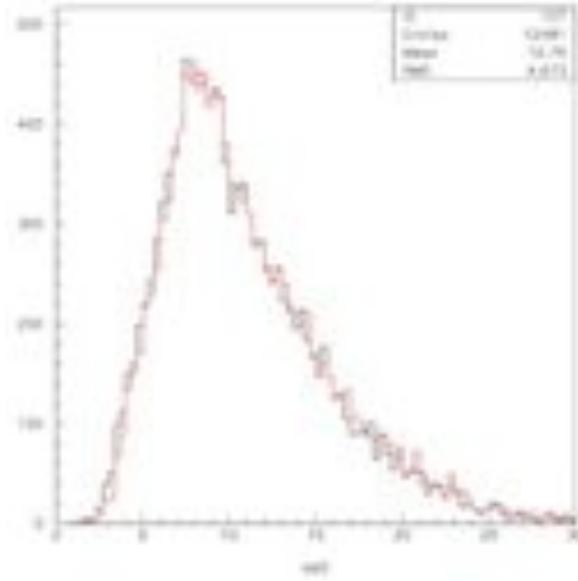


Figure 2.20. Distribution of Σ (see text) before (black) and after (red) optimization of the light decay time in each tank.

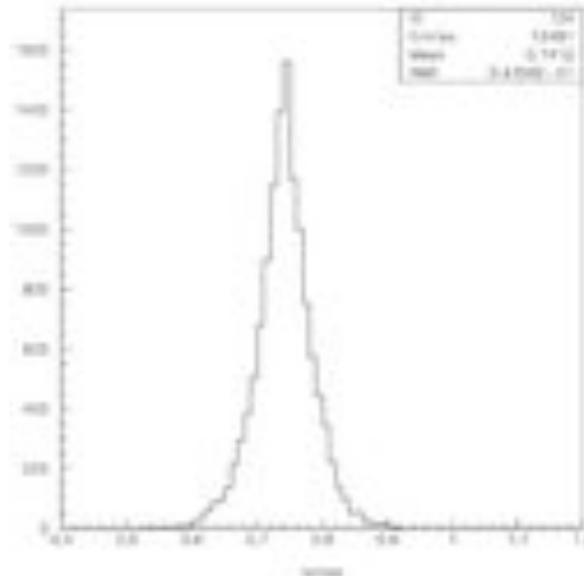


Figure 2.21. Distribution of the decay factors f adjusted in each tank to maximize Σ (see text).

Consider a sequence of three successive bins having charges q_{i-1} , q_i and q_{i+1} . If q_i fluctuates upward, $q_i - f q_{i-1}$ is overestimated and $q_{i+1} - f q_i$ is underestimated. Here again, q_i may be fine-tuned by maximizing Σ . For each bin i , calling m_i the result of this fine tuning, all that needs to be done is to adjust m in order to minimize the value of

$$\chi^2_{i|fluct} = (m - q_i)^2 / \Delta q_i^2 + (m - f q_{i-1})^2 / (f^2 \Delta q_{i-1}^2) + (q_{i+1} - f m)^2 / \Delta q_{i+1}^2.$$

Figure 2.22 shows the distribution of the minima of $\chi^2_{i|fluct}$ obtained using this procedure. It strongly peaks at zero, corresponding to bins for which the fine-tuning is meaningful. Applying it to bins having $Min(\chi^2_{i|fluct}) < 2$ gives the values of $m_i - q_i$ displayed in Figure 2.23 with a mean of -0.006 and an rms of 0.029 .

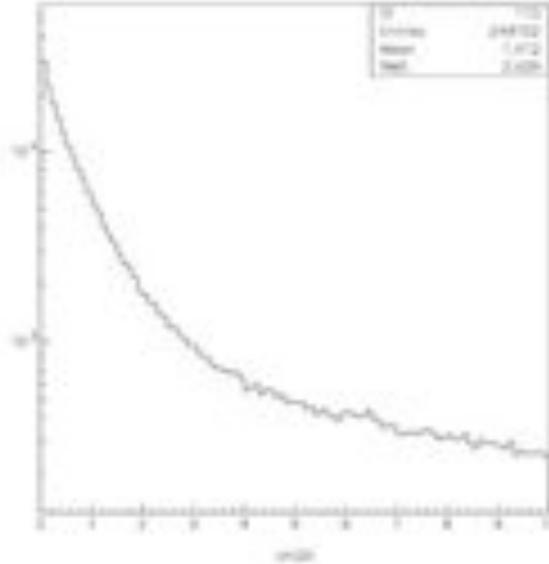


Figure 2.22. Distribution of $\chi^2_{i|fluct}$ after minimization (see text).

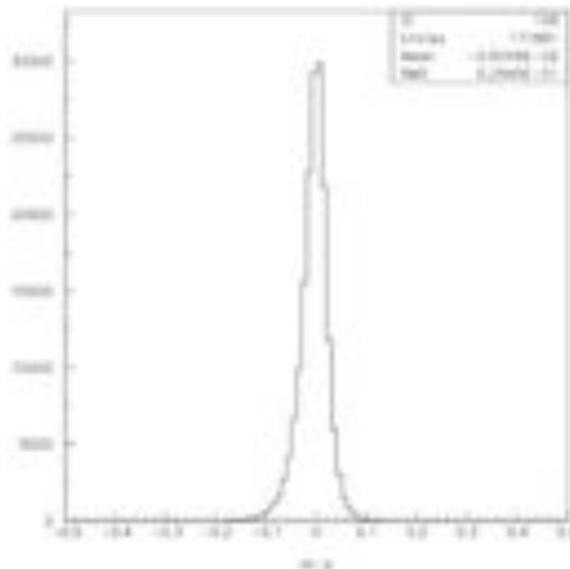


Figure 2.23. Distribution of the charge adjustments, $m_i - q_i$, resulting from the minimization of $\chi^2_{i|fluct}$ and restricted to bins having $Min(\chi^2_{i|fluct}) < 2$ (see text).

In a last step, all q_i having $Min(\chi^2_{i|fluct}) < 2$ are replaced by the corresponding m_i value and $\chi^2_{i|empty} = (q_i - f q_{i-1})^2 / (\Delta q_i^2 + f^2 \Delta q_{i-1}^2)$ is recalculated. Its distribution is shown in Figure 2.24. Figure 2.25 displays the associated evolution of the Σ distribution. The fine-tuning of m has gained 1.8 units per event on average.

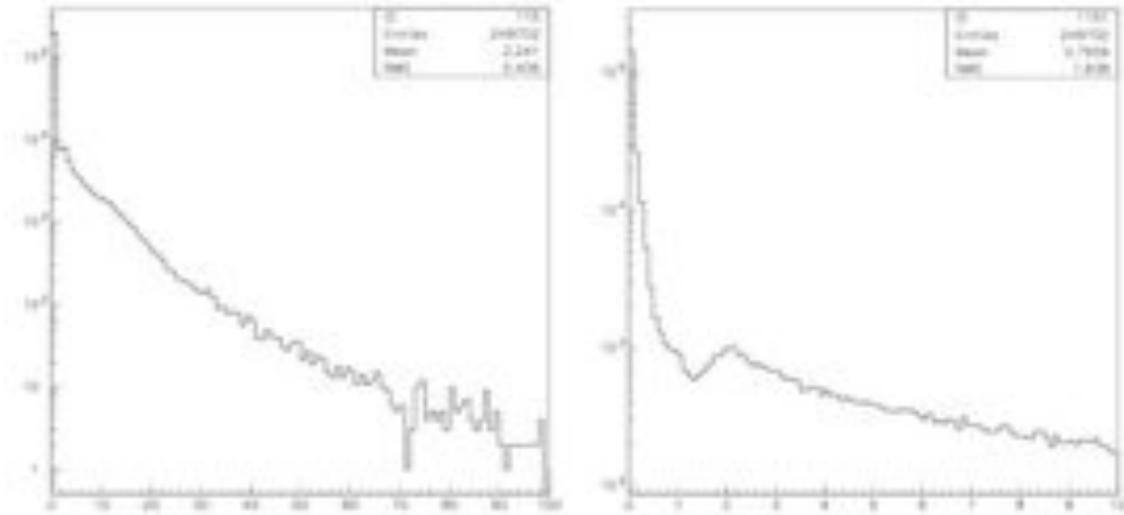


Figure 2.24. Distribution of $\chi^2_{i|empty}$ after correction for statistical fluctuations (see text). A zoom of the lower part is shown in the right panel.

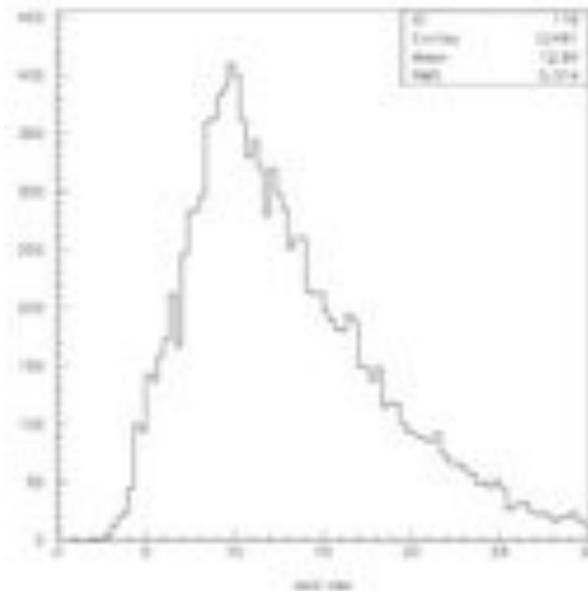


Figure 2.25. Distribution of Σ (see text) after correction for statistical fluctuations.

The distribution of $\chi^2_{i|empty}$ displays a sharp peak at zero, suggesting to define as peak candidates bins having $\chi^2_{i|empty} > 1$. In principle they might contain positive or negative charges (after subtraction). It is remarkable (Figure 2.26) that they nearly always contain positive subtracted charges, corresponding to new signals.

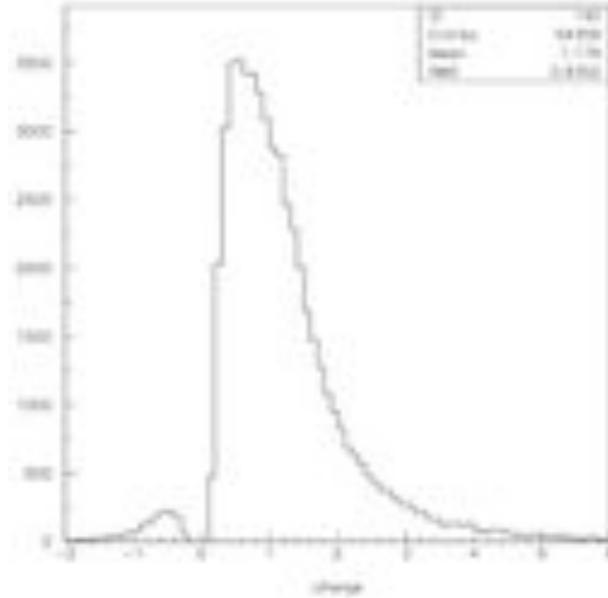


Figure 2.26. Distribution of subtracted charges s_i for bins having $\chi^2_{i|empty} > 1$ after application of all corrections.

2.2.4 Early time asymmetries

When light is produced within a tank, the part of it which escapes first to the PMT photocathodes is not evenly distributed among the three PMT's. This has been studied both on data [77] and on simulations [66,75] and is relatively well understood. The data, both muons and electron-photon showers, display a strong correlation between the shower azimuth and the early time asymmetry (Figure 2.18) and suggest that the asymmetry is not strongly influenced by the nature of the particles detected. Using a simple Monte Carlo simulation of the tank optics [75], we have studied the asymmetry induced by muons and verified that it displays the same correlation. The simulation shows that a correlation with impact point, not accessible to real data, is also important in the case of muons (and probably for electrons-photons as well). The point of relevance is that, whatever its cause, some early time asymmetry is to be expected as long as the signal is large enough. At least the simulation shows that muons associated with no significant asymmetry all have short track lengths (Figure 2.27) and it is reasonable to assume that this result applies as well to electron-photon showers.

Defining a quantity $\chi^2_{i|asym} = \{(q_{i1} - q_i)^2 + (q_{i2} - q_i)^2 + (q_{i3} - q_i)^2\} / \Delta q_i^2$ in each bin of the tank FADC trace, asymmetric bins have a large value of $\chi^2_{i|asym}$. The distribution of the decimal logarithms of $\chi^2_{i|asym}$ for all bins of all segments is shown in Figure 2.28 where it is

compared to that of the subsample having $\chi^2_{i|empty} > 1$. As there are 2 degrees of freedom, a mean value of 2 is expected for symmetric events. The mean value is in fact 3.1 when all events are included, measuring the effect of the long tail of asymmetric events. To get a mean value of 2 requires retaining only events having $\chi^2_{i|asym} < 20$. The subsample of peak candidates ($\chi^2_{i|empty} > 1$) gives a mean value of 8.4 while the complementary sample ($\chi^2_{i|empty} < 1$) gives a mean value of 1.6. A cut at 2 in $\chi^2_{i|asym}$ is efficient at retaining asymmetric events but also rejects a large number of peak candidates. Where to cut in $\chi^2_{i|asym}$ must be decided by finding the optimum compromise. The reason is that some peak candidates displaying small asymmetry (small track lengths) may very well be honest peaks. To understand this better a two-dimensional analysis is necessary. Figure 2.29 displays the correlation between the decimal logarithms of $\chi^2_{i|asym}$ and $\chi^2_{i|empty}$. It shows a clear separation in two families, one of empty bins having no significant asymmetry, and one of peaks having significant asymmetry. Moreover, in the latter family, a clear correlation is observed between $\chi^2_{i|asym}$ and $\chi^2_{i|empty}$.

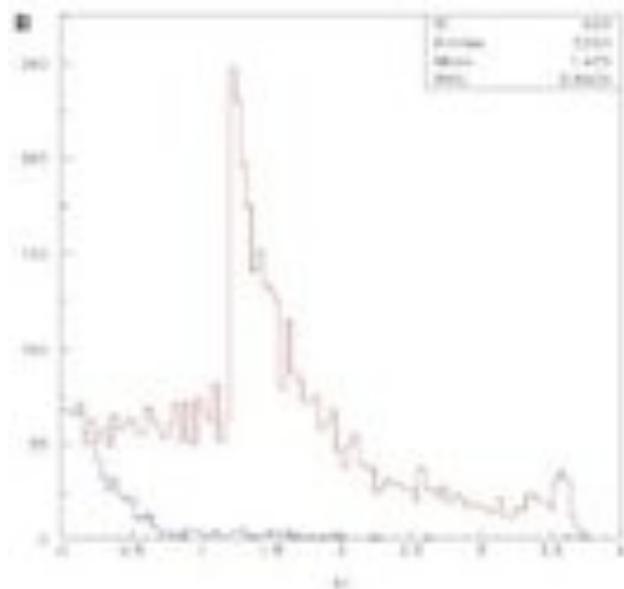


Figure 2.27. Track length distribution of simulated events displaying an insignificant early time asymmetry (blue) is compared to that of all simulated events (red).

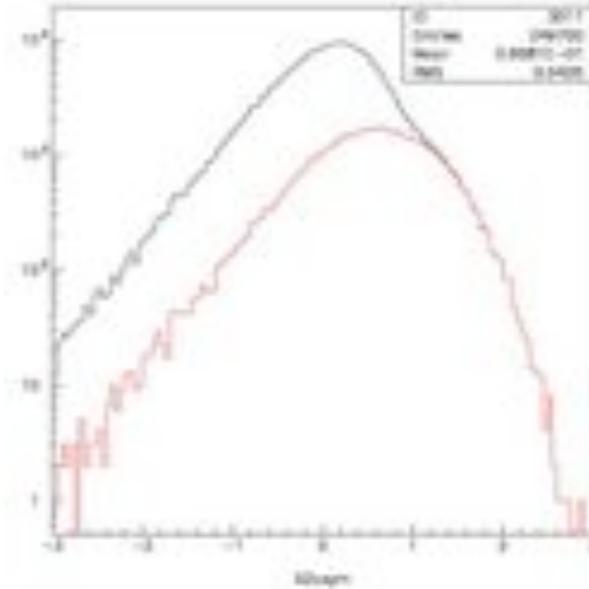


Figure 2.28. Distribution of the decimal logarithms of $\chi^2_{i|asym}$ for all bins of all segments (black) and that of the subsample having $\chi^2_{i|empty} > I$ (red).

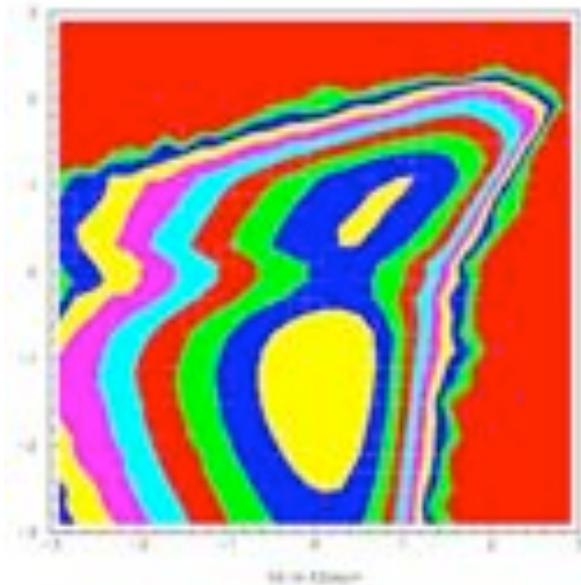


Figure 2.29. Correlation between the decimal logarithms of $\chi^2_{i|asym}$ (abscissa) and of $\chi^2_{i|empty}$ (ordinate).

A moment analysis of the event density in the domain $\log_{10} \chi^2_{i|asym} > -0.5$ and $\log_{10} \chi^2_{i|empty} > 0$ gives the centre of the distribution at $\log_{10} \chi^2_{i|asym} = 0.65$ and $\log_{10} \chi^2_{i|empty} = 0.77$. The angle of the major axis with the horizontal axis is 27° . In principle, selecting peaks within such an ellipse might help resolving adjacent peaks. In practice, however, it can only be done at the expense of efficiency.

2.2.5 Muons

The present section illustrates the method by comparing its results to those of Reference 67. It is recalled that the refinements which have been added to it include: a more sophisticated base line subtraction, an improved estimate of the uncertainty attached to each FADC bin, a tank-by-tank evaluation of the light decay factor f , a correction for statistical fluctuations and, possibly, the additional use of the asymmetry between the charges measured in each PMT separately. Subtracted charges, $s_i = (q_i - f q_{i-1}) / (1 - f)$, are systematically and exclusively used for this study. Figure 2.30 shows a set of distributions of subtracted charges corresponding to different intervals of zenith angle and distance to the shower core, which can be directly compared to those of Figures 3 of Reference 67.

One distribution (D1) is meant to exactly reproduce the results of [67], applying the same cuts aimed at selecting isolated peaks. A second distribution (D2) is using the same cuts but makes use, in addition, of the baseline subtraction, of the tank-by-tank f fit and of the correction for statistical fluctuations. A third distribution (D3) requires in addition $\chi^2_{i|empty}$ to exceed unity. A fourth distribution (D4) requires also $\chi^2_{i|asym}$ to exceed unity (one unit smaller than the number of degrees of freedom). Each of these distributions has been fitted with the sum of an exponential (meant to describe the electron-photon background) and of a Gaussian (meant to describe the muon component).

The results of these fits are summarized in the table below and show that the application of the χ^2 cuts do not loose a significant fraction of the muon component while reducing significantly the lower part of the spectrum.

Fits of the subtracted charge distributions D1 to D4 to a form $a \exp(-0.5(s_i - s_0)^2 / \Delta s^2) + b \exp(-s_i / \sigma)$. The values listed are of $a \cdot \Delta s$, proportional to the area of the Gaussian.

$\cos\theta$	1.0-0.9	0.9-0.8	0.8-0.7
D1	179 ± 6	158 ± 5	112 ± 6
D2	205 ± 7	177 ± 7	130 ± 6
D3	169 ± 6	154 ± 6	117 ± 6
D4	167 ± 6	154 ± 6	120 ± 6

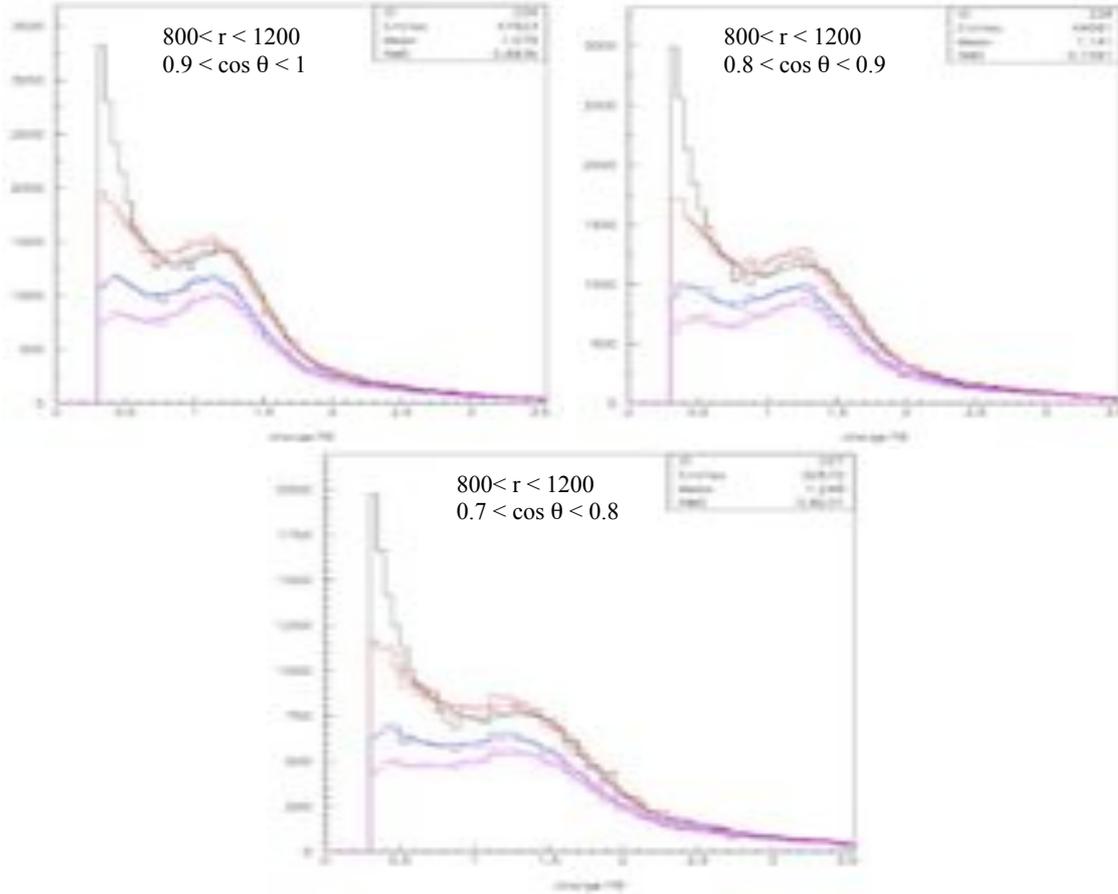


Figure 2.30. Four different distributions of $s_i + s_{i+1}$ (see text) are shown for three different bins of zenith angle and distances to the shower core between 800 and 1200 meters. They are defined as follows: before any of the corrections described in the present note (black), after such corrections (red), requiring in addition $\chi^2_{empty} > 1$ (blue) and requiring in addition $\chi^2_{asym} > 1$ (purple). The curves are the results of fits of the form Gaussian + exponential (see text).

The distribution of the subtracted charges of all bins having $\log_{10} \chi^2_i |_{empty} > 0$ is shown on Figure 2.31 for data having a zenith angle between 0° and 26° and a distance to the shower core between 800 m and 1200 m. There is no visible evidence for a muon peak, at strong variance with Figure 2.30. There is still no strong evidence for a muon peak when using the selection criteria of Reference 67, but the muon peak becomes apparent when the charge of the following bin is added. This shows that it is not the selection criterion used to produce Figure 2.30 that makes the muon peak appear but rather the addition of the charge of the following bin. It points to the importance of addressing the question of adjacent signals.

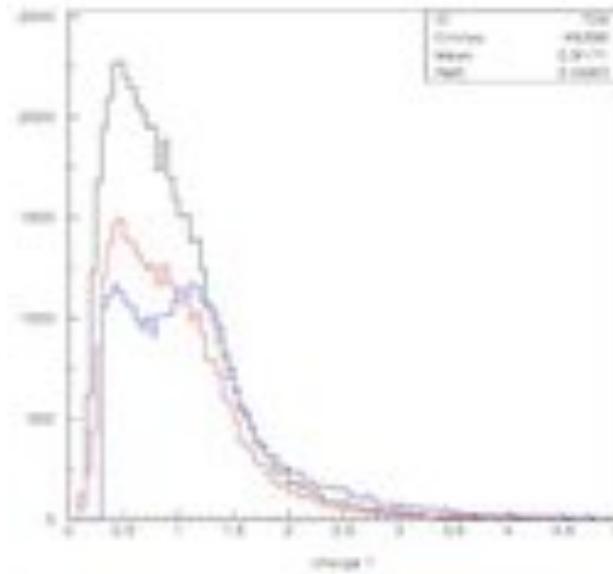


Figure 2.31. Distributions of the subtracted charges of bins having $\log_{10}\chi^2|_{empty} > 0$ for data having a zenith angle between 0° and 26° and a distance to the shower core between 800m and 1200 m: All bins (black), bins obeying the selection criteria of Reference 67 (red) and pairs of adjacent bins (blue).

2.2.6 Adjacent signals

Groups of adjacent bins having $\log_{10}\chi^2|_{empty} > 0$ are distributed as follows: 76% are one bin wide, 20% two bins wide and 4% three bins wide or more. In the one bin case, calling i the bin index, the distributions of $\chi^2|_{empty}$ and $\chi^2|_{empty}$ are shown in Figure 2.32. Calling i_1 and i_2 the indices $i \pm 1$ having the larger and respectively smaller value of $\chi^2|_{empty}$ the distributions of s_i , $s_i + s_{i_1}$ and $s_i + s_{i_2}$ are compared in Figure 2.33. While adding the bin having the largest value of $\chi^2|_{empty}$ makes the muon peak more apparent, the addition of the bin having the lowest value of $\chi^2|_{empty}$ keeps the distribution unchanged.

Attempts at dealing with adjacent bins by making use of the asymmetry information [77] have illustrated the complexity of the problem but have been inconclusive.

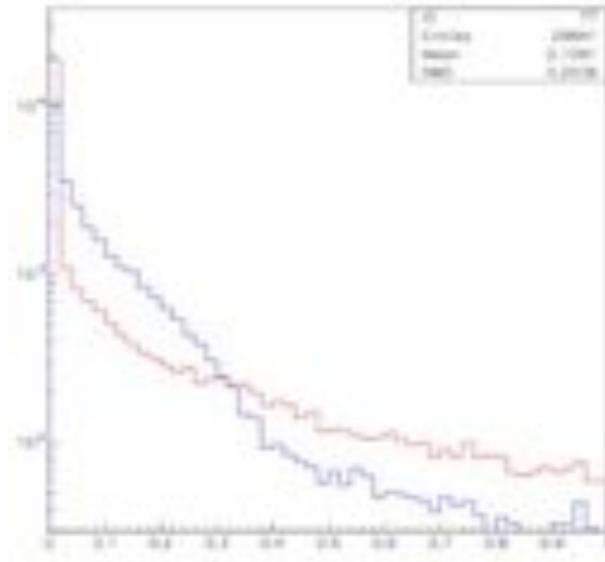


Figure 2.32. Distributions of $\chi^2_{i-1|empty}$ (red) and $\chi^2_{i+1|empty}$ (blue) for isolated bins i having $\chi^2_{i|empty} > 1$.

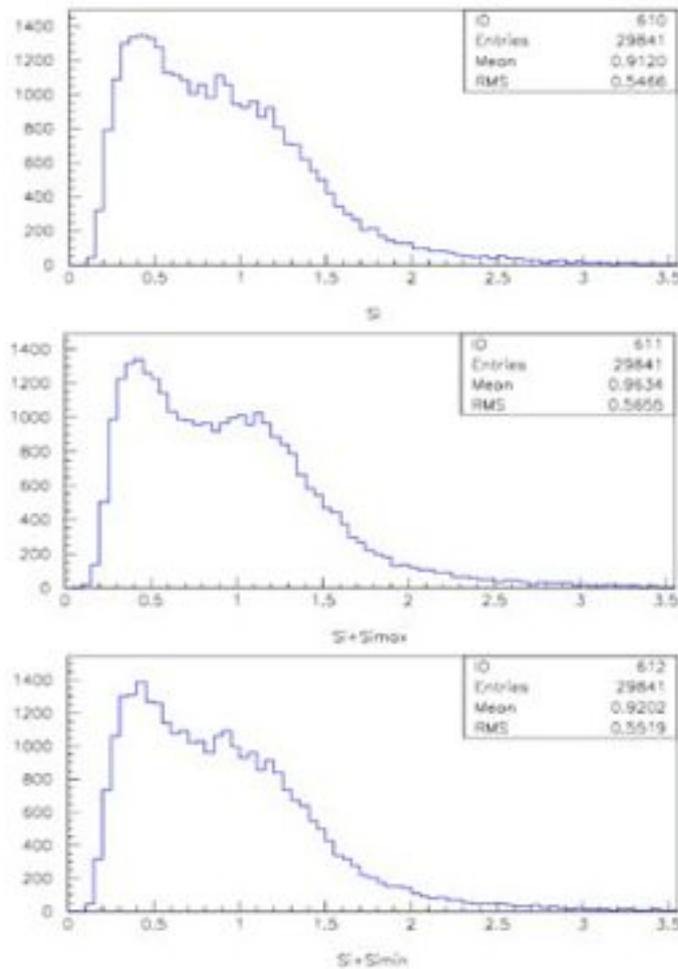


Figure 2.33. Distributions of s_i (upper panel), s_i+s_{i1} (middle panel) and s_i+s_{i2} (lower panel) for isolated bins i having $\chi^2_{i|empty} > 1$ (see text).

2.2.7 Summary

An attempt at resolving FADC traces into individual signals has been presented; it follows an earlier proposed method that consists in unfolding from each trace the exponential decay of the light collected on the PMT photocathodes. The method has been refined in several respects: more sophisticated base line subtraction, improved evaluation of the uncertainty attached to each FADC bin, correction for statistical fluctuations and tank by tank evaluation of the light decay factor. A small, but significant, improvement has been obtained and illustrated by a slightly easier identification of the muon peak in small zenith angle showers. An attempt at using the early time PMT asymmetry as a tag for new signals has given interesting results but has proven of limited practical use in separating overlapping, nearly simultaneous, signals.

3. PMT asymmetries and shower divergences

3.1 Introduction

3.1.1 Motivation

The preceding chapter (Section 2.1) discussed the relation between the shower azimuth and the PMT asymmetries of the SD tanks which display a strong correlation (Figure 2.18), the point being that particles incident on a tank away from the vertical illuminate differently the three PMTs of the tank. It was remarked, on this occasion, that the azimuths that are compared in Figure 2.18 are not exactly the same. The shower azimuth directly measured by the tank timings is that of the shower axis while the azimuth to which the PMT asymmetries are sensitive is that of the particles incident on the particular tank. Hence the possibility to measure the shower divergence by combining the asymmetries measured in all tanks of a same shower. The high statistics presently available in SD data makes this possible. Moreover the algorithm presented in Section 2.2 can now be used to the advantage of the study. The measurement of the mean altitude from which the shower particles are coming is of obvious potential interest to the extent that it is likely to be correlated with the location of the shower maximum and to be dependent on the nature of the particles detected (muons or electron/photons).

3.1.2 Arithmetics

For the purpose of this study, a shower is modelled as a set of secondaries originating from a same point M on the shower axis, at distance D from its impact on ground, O (Figure 3.1 left). For a given tank T , calling \mathbf{u} the unit vector along the shower axis, $\mathbf{TM} = \mathbf{OM} - \mathbf{OT} = D\mathbf{u} - \mathbf{OT}$ has unit vector \mathbf{v} such that $\mathbf{TM} = V\mathbf{v}$. Calling θ and φ the zenith angle and azimuth of \mathbf{u} , α and β those of \mathbf{v} and (X, Y, Z) the coordinates of \mathbf{OT} , one has

$$\begin{aligned}
 D\sin\theta\cos\varphi - X &= V\sin\alpha\cos\beta \\
 D\sin\theta\sin\varphi - Y &= V\sin\alpha\sin\beta \\
 D\cos\theta - Z &= V\cos\alpha \\
 V &= \sqrt{\{D^2 + X^2 + Y^2 + Z^2 - 2D(X\sin\theta\cos\varphi + Y\sin\theta\sin\varphi + Z\cos\theta)\}}
 \end{aligned} \tag{1}$$

Here one must take care that D , V , $\sin\theta$ and $\sin\alpha$ are all positives.

Calling \mathbf{w}_i the unit vector along the tank radius pointing to PMT i with $\mathbf{w}_i = \{\cos\psi_i, \sin\psi_i, 0\}$ and Q_i the charge measured in PMT i the simplest expression one may use for Q_i is $Q_i = Q_0 (1 - \lambda \mathbf{v} \cdot \mathbf{w}_i) = A - B \sin\alpha \cos(\psi_i - \beta)$.

Here $\mathbf{w}_1 = \{\sqrt{3}/2, 1/2, 0\}$, $\mathbf{w}_2 = \{0, -1, 0\}$ and $\mathbf{w}_3 = \{-\sqrt{3}/2, 1/2, 0\}$.

Note that with this expression one should expect B to be positive as \mathbf{v} is directed against the particle direction rather than along with it. Therefore:

$$\begin{aligned} Q_1 &= A + \frac{1}{2}\sqrt{3}B\sin\alpha\cos\beta + \frac{1}{2}B\sin\alpha\sin\beta \\ Q_2 &= A - B\sin\alpha\sin\beta \\ Q_3 &= A - \frac{1}{2}\sqrt{3}B\sin\alpha\cos\beta + \frac{1}{2}B\sin\alpha\sin\beta \end{aligned} \quad (2)$$

From which one gets

$$\begin{aligned} A &= \frac{1}{3}(Q_1 + Q_2 + Q_3) \\ f &= (Q_2 - A)/A = -(B/A) \sin\alpha\sin\beta \\ &= -(B/A)(D\sin\theta\sin\varphi - Y)/V \\ g &= (Q_3 - Q_1)/\sqrt{3} = -(B/A) \sin\alpha\cos\beta \\ &= -(B/A)(D\sin\theta\cos\varphi - X)/V \end{aligned}$$

One is left with two unknowns, D and $\lambda = -B/A$, and two measured quantities, f and g .

In principle, one can solve

$$D\sin\theta = (gY - fX)/(g\sin\varphi - f\cos\varphi), \quad B = -V(g\cos\varphi - f\sin\varphi)/(Y\cos\varphi - X\sin\varphi)$$

In practice, however, it does not work, as the following example will illustrate.

Consider the case of a tank located on the ground projection of the shower axis and take $Z=0$ for simplicity (Figure 3.1 right).

$$\text{Then, } X = R\cos\varphi, \quad Y = R\sin\varphi, \quad R = \sqrt{X^2 + Y^2}$$

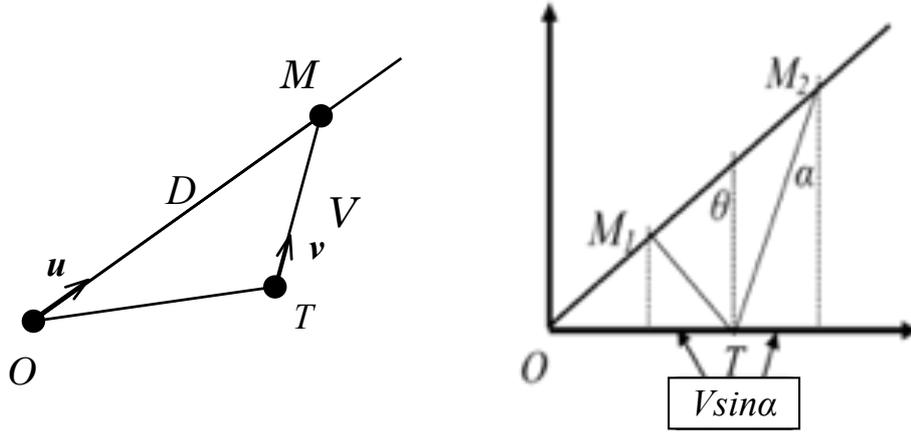


Figure 3.1 Left: geometry. Right: case of a tank located on the projection of the shower axis.

The system of equations becomes

$$(D \sin \theta - R) \cos \varphi = V \sin \alpha \cos \beta$$

$$(D \sin \theta - R) \sin \varphi = V \sin \alpha \sin \beta$$

$$D \cos \theta = V \cos \alpha$$

$$\text{Hence } D \sin \theta = R \pm V \sin \alpha = R \pm D \cos \theta \tan \alpha,$$

where the \pm sign corresponds to the fact that $\beta = \varphi$ or $\beta = \varphi + \pi$ in order for V and $\sin \alpha$ to be positive.

The measurement of f and g , such that $\tan \beta = f/g$, is therefore irrelevant to the evaluation of β . Moreover, we can only get the value of the product $\lambda \sin \alpha$ but there is no way to obtain the two factors independently.

The lesson from this simple example is 1) one should not attempt an analytical evaluation of λ and D but rather find the values of these variables which minimize the χ^2 to the measurements of f and g and 2) one must, obviously, fit all the tanks of a same shower together.

The arithmetic is summarized below:

$$\chi^2 = \sum ((f - \lambda \sin \alpha \sin \beta) / \Delta f)^2 + ((g - \lambda \sin \alpha \cos \beta) / \Delta g)^2 \text{ where the sum runs over all tanks.}$$

Calling $\sigma_0 = \sum (f_i^2 / \Delta^2 f_i + g_i^2 / \Delta^2 g_i)$ and $C = 1/D$, the divergence of the shower

$$\begin{aligned} \sigma_1(C) &= \sum (f \sin \alpha \sin \beta / \Delta^2 f + g \sin \alpha \cos \beta / \Delta^2 g) \\ &= \sum (f [\sin \theta \sin \varphi - C Y] / [C V] / \Delta^2 f + g [\sin \theta \cos \varphi - C X] / [C V] / \Delta^2 g) \\ \sigma_2(C) &= \sum (\sin^2 \alpha \sin^2 \beta / \Delta^2 f + \sin^2 \alpha \cos^2 \beta / \Delta^2 g) \end{aligned} \quad (3)$$

$$= \sum ([\sin\theta\sin\varphi - CY]^2 / [CV]^2 / \Delta^2 f + [\sin\theta\cos\varphi - CX]^2 / [CV]^2 / \Delta^2 g)$$

$$\chi^2(C) = \sigma_0 - 2\lambda\sigma_1(C) + \lambda^2\sigma_2(C),$$

which is minimum for $\lambda(C) = \sigma_1(C)/\sigma_2(C)$ giving $\chi^2(C) = \sigma_0 - \sigma_1^2(C)/\sigma_2(C)$.

All what needs to be done is then to find the value of C which minimizes $\chi^2(C)$.

3.2 Overview of the method

3.2.1 Azimuth-asymmetry correlation

Having defined the arithmetic of relevance, it is now time to look at the data⁹. Usual corrections and selections are performed as described earlier, namely base line correction, rejection of tanks having a total charge in excess of 40 VEM, exclusion of after pulsing candidates [77]. In the wake of the study presented earlier of early time asymmetries of segment charges, one associates with each tank three charges, q_1 , q_2 and q_3 , one for each of the three FADC traces, corresponding to the early time bins of resolved signals. Here *resolved signals* means signals which remain after deconvolution of the time dependence of the light decay, defined as having¹⁰ $\chi^2_{empty} > 1$, namely what was called “peaks” in the preceding chapter. As the time bin preceding a peak is known to contribute to the early time asymmetry, peak segments are defined as sets of adjacent peaks. The peak segment charges are then defined as the sum of the charges of the FADC bins in each peak segment (this time before the deconvolution procedure!), including that of the bin preceding the peak segment, and the q_i 's are defined as the sum of the charges measured in the peak segment bins of a same FADC trace (one per PMT).

Calling $\langle q \rangle$ the tank charges, $\langle q \rangle = (q_1 + q_2 + q_3)/3$, its distribution is shown in Figure 3.2. In order to avoid dealing with too small signals, for which an asymmetry would hardly be significant, tanks must have $\langle q \rangle$ in excess of 0.13 VEM in order to be retained. It is important to remember that the q 's are calculated as sums of charges measured before deconvolution over bins selected as having a significant content after deconvolution. It is recalled that in the ideal case of an isolated exponential signal starting exactly at the

⁹ The data sample used here contains showers having $S_{1000} > 5$ selected from $T4$ data collected between January 2004 and December 2008. Tanks with a total charge exceeding 40 VEM are systematically excluded.

¹⁰ A very few cases of bins having $\chi^2_{empty} > 1$ but a negative charge are also excluded.

beginning of a time bin, the charge contained in this first time bin is typically $1/(1-0.74)=3.85$ times smaller than the total signal charge. The cut on $\langle q \rangle$ corresponds therefore to a cut on total charge of some 0.4 to 0.5 VEM (a precise evaluation requires taking proper account of the fact that the bin preceding the segment has also been included).

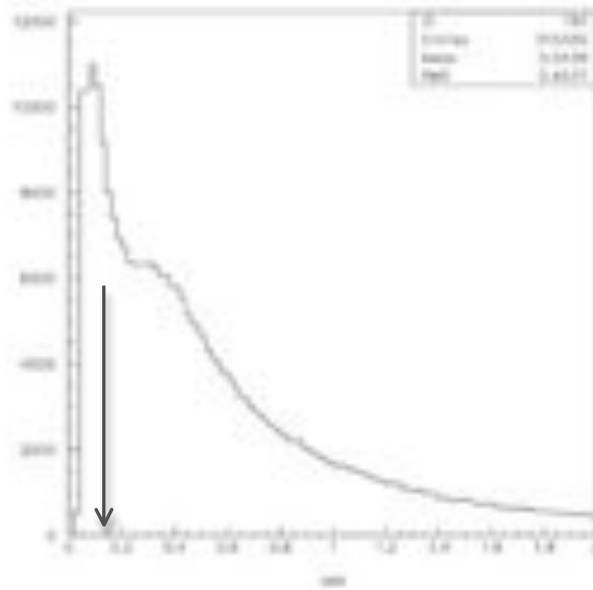


Figure 3.2. Distribution of the tank charges $\langle q \rangle$. The arrow indicates the cut applied in the tank selection ($\langle q \rangle$ in excess of 0.13).

Using Relations (2) and ignoring a possible divergence of the shower, namely assuming that all shower particles come from infinity and point in the (θ, φ) direction of the shower axis, the q_i 's defined above can be used to evaluate the correlation between the asymmetry and the shower direction as was done previously in Figure 2.18 using a different approach.

Figure 3.3 displays the distribution of the difference between the value of φ obtained from the early time asymmetry and that obtained from the shower axis reconstruction using the tank timings. It shows a strong correlation – even stronger than that found in Figure 2.18 – but it is only a correlation, not a one-to-one correspondence: it is not possible to evaluate the direction of the shower axis from the early time asymmetry on an event-to-event basis.

Relations (2), under the assumption of zero divergence, allow also for the calculation of $\lambda \sin \theta$ for each trace. However, the sign of this quantity is not defined:

changing its sign and increasing at the same time φ by 180° leaves Relations (2) invariant. Making the choice of sign which corresponds to the smaller azimuth difference (the quantity displayed in Figure 3.3), and using the known zenith angle θ of the shower axis to divide by $\sin\theta$, one obtains the λ distribution displayed in Figure 3.4. It is clearly positive on average, as expected, and its mean value, 0.30, is a measure of the correlation.

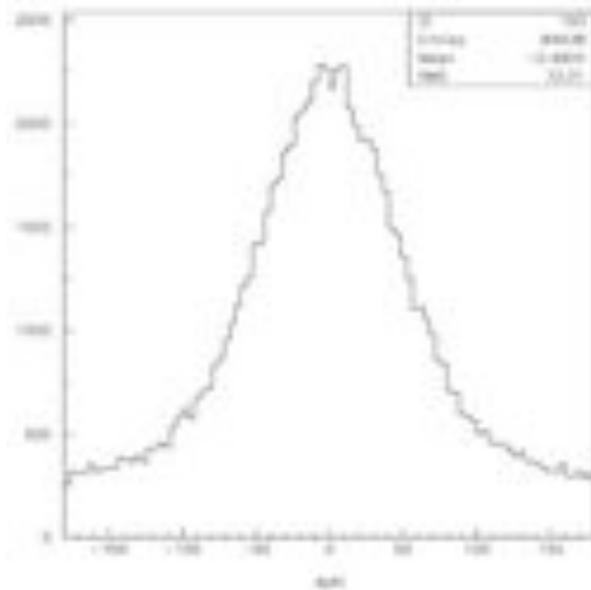


Figure 3.3. Distribution of the difference between the shower axis azimuths measured from the PMT asymmetry and from the tank timings under the assumption of no divergence.

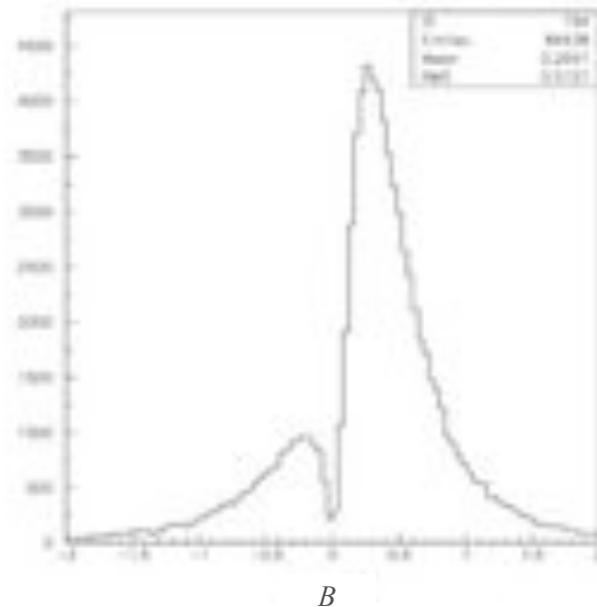


Figure 3.4. Distribution of the correlation strength λ in the assumption of no divergence.

3.2.2 Shower divergence

Having re-established the existence of the correlation, Relations (3) are now used to evaluate the divergence C and the correlation coefficient λ for each shower having at least two tanks retained. The resulting χ^2/N_{tank} distribution, where N_{tank} is the number of tanks contributing to the fit, is shown on Figure 3.5. While this quantity cannot be given the same statistical significance as a normal χ^2 , it gives a useful measure of the quality of the fit.

In the following, showers having χ^2/N_{tank} in excess of 20 are excluded (they amount to only 4% of the total number of showers). The distributions of C and λ are shown in Figures 3.6 and 3.7 respectively. While very broad, they give evidence for a positive correlation¹¹, $\lambda \sim 0.25$, and for a divergence deviating significantly from zero, $C = 0.094 \pm 0.006$, corresponding to a vertex located ~ 10 km along the shower axis from its impact on ground. The similarity between the present evaluation of λ and that made in the preceding paragraph gives confidence in the procedure.

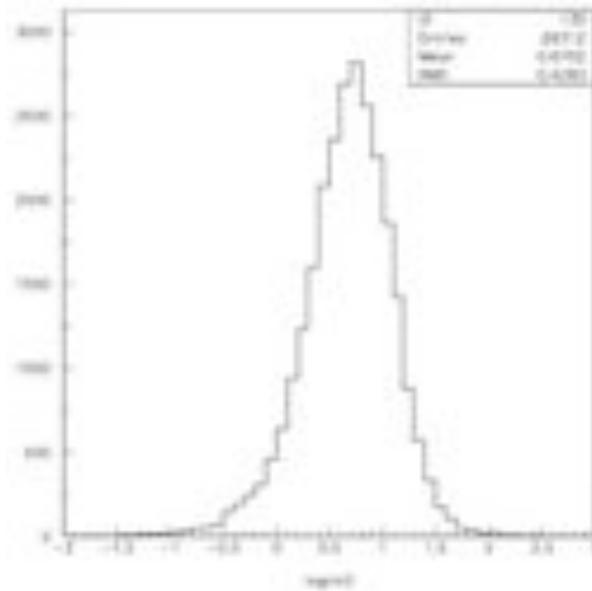


Figure 3.5. Distribution of the decimal logarithm of the quantity χ^2/N_{tank} . The cut applied is at 1.3.

¹¹ The value of λ found here, 0.25, is slightly lower than the value found when assuming no divergence, 0.30, but in that later case a favourable choice of sign had been made, which probably explains the difference.

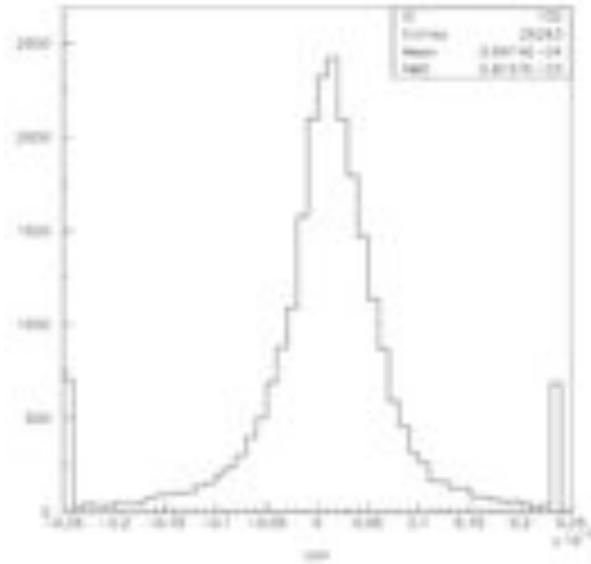


Figure 3.6. Distribution of the divergence of retained showers (in units of m^{-1}).

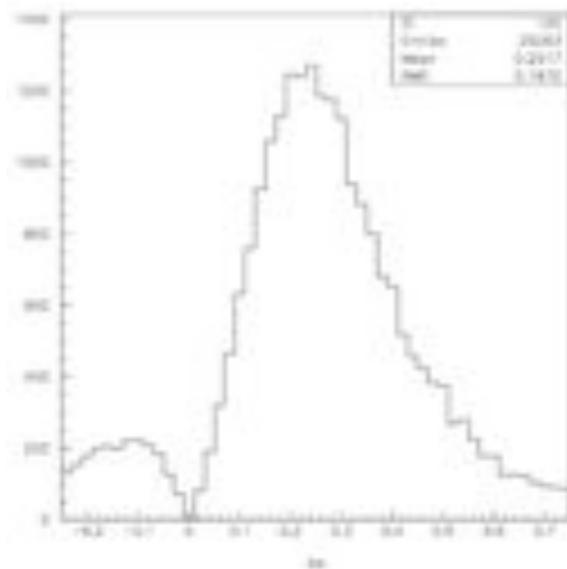


Figure 3.7 Distribution of the correlation coefficient λ for retained showers.

To understand this result, one must remember that the large widths of the distributions in Figures 3.6 and 3.7 are not the results of measurement errors but the results of the imperfection of the correlation which is being exploited. For the same reason, the events which are observed to have a divergence differing strongly from average are found to display no other particularity: their distributions in other variables, such as the number of tanks retained, or the total charge (i.e. the sum of the $\langle q \rangle$'s over the tanks retained in the shower), are the same as for events having C near its average value.

To close this section, Figure 3.8 displays the dependence of the mean values of C , λ and $\log_{10}(\chi^2)$ on zenith angle for the sample of retained showers. Above 30° , C is seen to decrease with zenith angle, meaning that D increases. This evolution is compared with the decrease of C which would result from M being at constant altitude $H=D\cos\theta$. A curve having $H=7km$ fits the data between 30° and 50° or so; then an even stronger decrease of C is required. Further investigations are presented below.

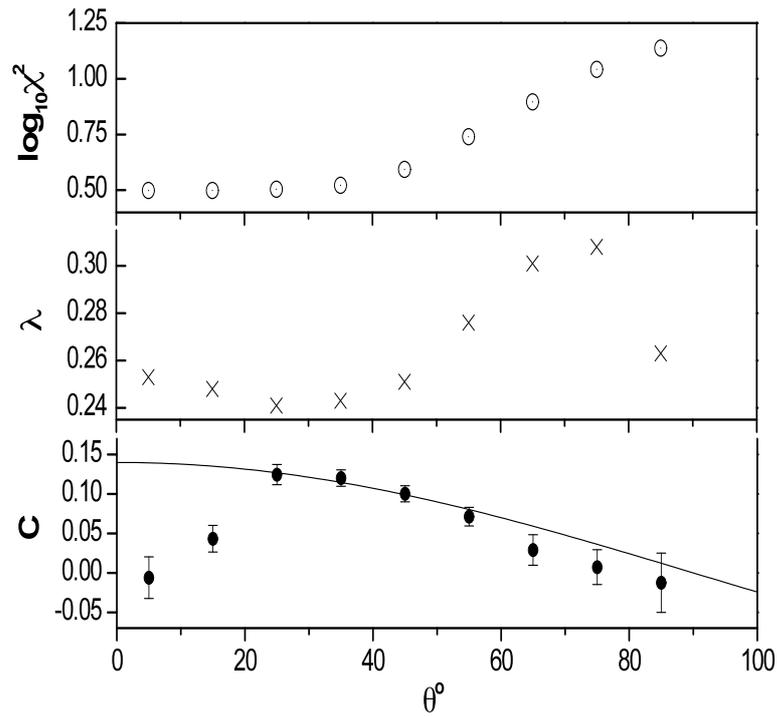


Figure 3.8. Distributions of the mean values of $\log_{10}(\chi^2/N_{tank})$, λ and C [$10^{-3}m^{-1}$] as a function of zenith angle. The line in the lower panel is for $C=\cos\theta/7km$ (see text).

3.3 Intrinsic asymmetry

3.3.1 Introduction

All PMTs of the SD detector are oriented in a same way with respect to the north (precisely with pin 20, called the short pin, westward). The reason for that is that there being no magnetic shield around the photocathode, it has been observed that PMT responses are modulated as a function of the angle between the magnetic north and the PMT orientation in a universal way [64,78,79]. The adopted orientation maximizes the response. Note that this implies that there exists an intrinsic asymmetry within each tube which is essentially the same from tube to tube. On average, each PMT photocathode receives more light from its face looking toward the centre of the tank than from its face looking toward the tank wall (Figure 3.9).

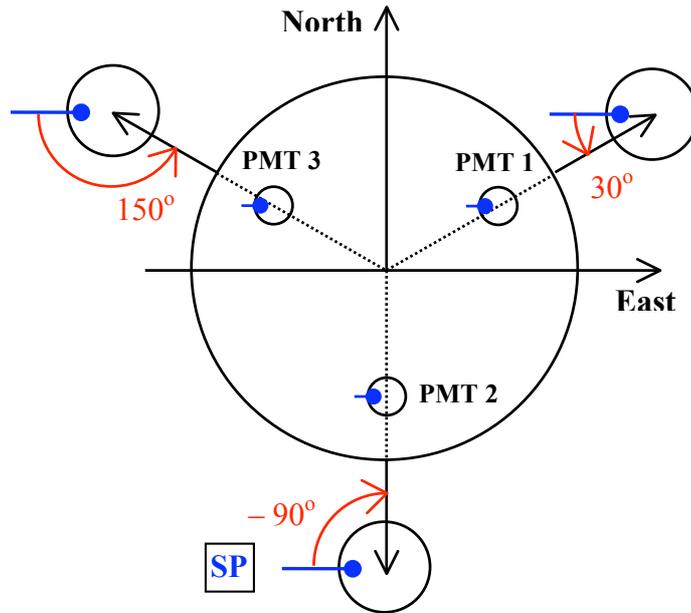


Figure 3.9. Top view of a tank. PMTs are oriented with their short pins westward.

This implies an additional factor of the form $1 - \mu \cos(\psi_i - \psi_0)$ on the expression of Q_i . Including this factor in the expression giving the χ^2 results in the following relations:

$$\chi^2 = \Sigma(Q_2/Q_0 - 1 - f)^2/\Delta^2 + \Sigma((Q_3 - Q_1)/(\sqrt{3}Q_0) - g)^2/\Delta^2$$

$$f = (1 + f_0)(1 + \mu \sin \psi_0) - 1 \text{ and } g = g_0 + \mu([1 - 1/2 f_0] \cos \psi_0 - 1/2 g_0 \sin \psi_0)$$

which reduce, to first order in λ and μ , to:

$$f = f_0 + \mu \sin \psi_0 \text{ and } g = g_0 + \mu \cos \psi_0.$$

Setting $\mu=0$ recovers the relations given in the preceding section as it should,

$$f=f_0=\lambda(D\sin\theta\sin\varphi-R_y)/V \text{ and } g=g_0=\lambda(D\sin\theta\cos\varphi-R_x)/V.$$

Here, \mathbf{R} is the vector joining the impact point of the shower axis on ground to the particular tank and V is the distance from the tank to the vertex. It is recalled that changing λ in $-\lambda$ and φ in $\varphi+\pi$, or similarly μ in $-\mu$ and ψ_0 in $\psi_0+\pi$, leaves the above relations invariant. For this reason, λ and μ are taken positive.

3.3.2 Results

The results concerning the intrinsic asymmetry are presented in Figure 3.10 showing the dependence of μ and ψ_0 on the zenith angle of the shower axis for two distinct cases, defined according to the value of the charge measured on the FADC trace averaged over the three PMTs. More precisely, the average trace having been reduced to a set of peaks as described earlier and a charge having been defined for each peak, the peaks are accordingly separated in two families: those having a charge smaller than 0.4 VEM and those having a charge larger than 0.4 VEM. In the low charge case, the Q_i are calculated as the sum of the charges of the low charge peaks. In the high charge case, the Q_i are calculated as the sum of the charges of the high charge peaks.

The low charge data give no strong evidence for any universal intrinsic asymmetry, the values of μ being generally less than 1.5 standard deviations away from 0. However, the high charge data indicate a preferred orientation $\psi_0 \sim -57^\circ \pm 13^\circ$ with, however, a very small value for the asymmetry: $\mu \sim 1.63 \pm 0.13 \cdot 10^{-3}$. For showers having a zenith angle smaller than 45° , for which the sensitivity to a universal intrinsic asymmetry should be best, $\psi_0 \sim -34^\circ \pm 5^\circ$. This can be compared with the measurements of Reference 64 which establish the existence of a preferred direction of incidence of the photoelectrons entering the multiplication chain. Here (Figure 3.9), it is the value of $\psi_0+\pi$ that gives the azimuth (with respect to the centre of the tank and to the east direction) of the centre of a PMT giving the largest response (because of the factor $-\mu$ rather than $+\mu$). This means that the point on the photocathode of such a PMT that faces the centre of the tank is at azimuth ψ_0 (with respect to the centre of the PMT and to the east direction). With respect to the north, as in Reference 64, the preferred direction of impact of the photons on the photocathode is therefore at $-124^\circ \pm 5^\circ$. This happens to check precisely with Reference 64 which predicts a preferred direction of about -120° . However, this coincidence may well be accidental.

Indeed, the very low value of μ found here, $\mu \sim 1.63 \pm 0.13 \cdot 10^{-3}$, is not commensurable with the expected anisotropy which is at least of the order of 0.2 at maximum [64]. Its smearing over the photocathode area can not reduce this number by more than one order of magnitude.

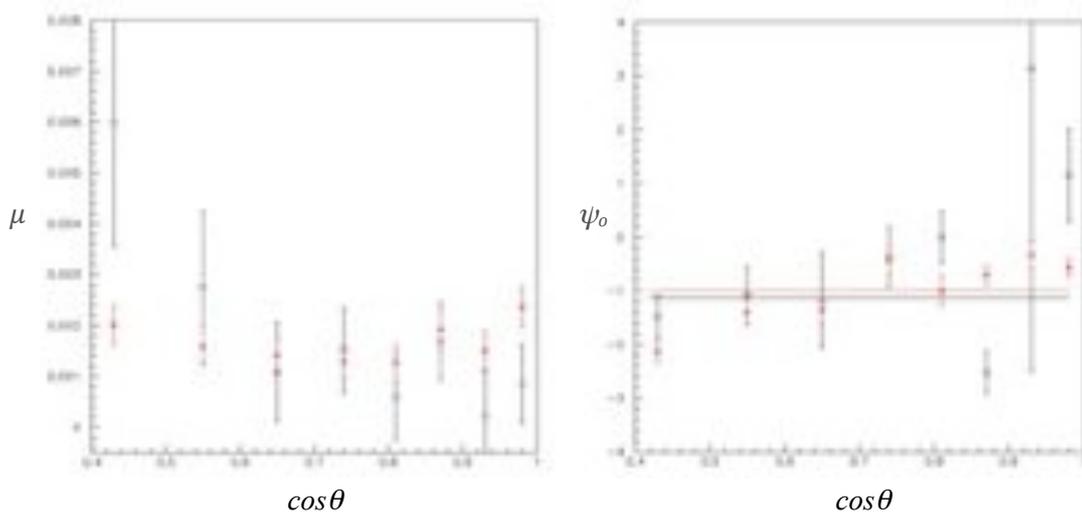


Figure 3.10. Dependence of μ (left) and ψ_0 (right, in radians) on $\cos\theta$ (high charges in red, low charges in black).

The universal intrinsic asymmetry measured *in situ* is therefore probably one order of magnitude smaller than that expected from the laboratory measurements. However, this could be explained by the fact that in Malargüe, the PMTs are calibrated with real events (atmospheric muons): the calibration should account for the effect and cancel the asymmetry. Differences in illumination between calibration and normal operation would then be responsible for the small remaining anisotropy. This interpretation, however, is highly conjectural and is only mentioned here as a suggestion.

The universal intrinsic asymmetry measured here is small enough to have a negligible effect and is ignored in the studies presented below.

3.4 Single vertex approximation

3.4.1 Dependence on energy and zenith angle

In the single vertex approximation, the asymmetry is fully defined by only two parameters, C and λ . It is better to use the convergence C than its reciprocal D as a parameter to avoid the singularity at $D=\infty$. The relation between C , the altitude h of the vertex and the zenith angle θ of the shower axis is simply $hC=\cos\theta$. The parameter λ describes the strength of the correlation between the azimuth of incidence and the asymmetry. The results are displayed in Figure 3.11 for the two data samples considered earlier. Data have been split in 8 bins of $\cos\theta$ between ~ 0.3 and 1. The correlation becomes much stronger at large zenith angles (in addition to the trivial $\sin\theta$ factor coming from the $\mathbf{v}\cdot\mathbf{w}$ scalar products). The convergence decreases with zenith angle but less than would result from a constant altitude. The fits shown in the figure are

$$C_{high}=(-0.133\pm 0.002)+(0.433\pm 0.004)\cos\theta \text{ and } C_{low}=(-0.223\pm 0.007)+(0.671\pm 0.010)\cos\theta$$

for high and low charges respectively. Constant altitude would imply no constant terms. The low charge data originate on average from lower altitudes than the high charge data, which seems reasonable. It is remarkable that both expressions cancel for $\cos\theta\sim 0.32$, corresponding to vertices located on an hyperboloid having an asymptotic cone of half aperture $\theta\sim 72^\circ$ and reaching its minimum value on the vertical at a distance of 3.3 km for high charges and 2.2 km for low charges. This encourages rewriting C as $C=C_0(\cos\theta-0.32)/0.68$ and using C_0 as a parameter applying to all zenith angles.

The above results suggest to concentrate on high charge data and to parameterize them in the form $C=A(\cos\theta)+B(\cos\theta)\log_{10}E$, A and B taking constant values in each of the eight intervals of $\cos\theta$. Their dependence on $\cos\theta$ is illustrated in Figures 3.12 and 3.13 and found linear, the best linear fit being

$$C^*(\cos\theta, \log_{10}E)=(-0.133+0.448\cos\theta)+(0.032-0.088\cos\theta)\log_{10}E.$$

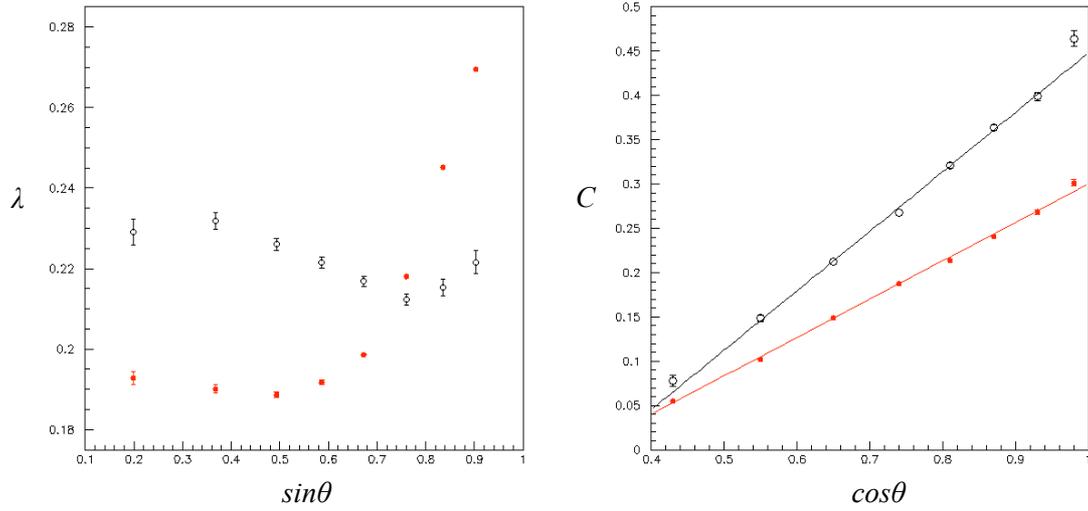


Figure 3.11. Dependence of λ on $\sin\theta$ (left) and of C on $\cos\theta$ (right). High charges are shown in red and low charges in black).

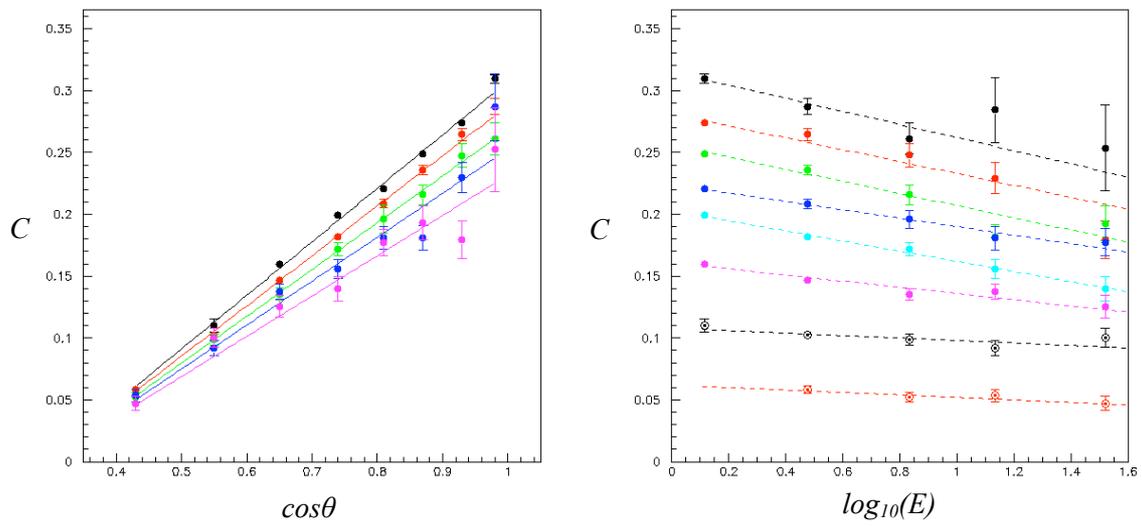


Figure 3.12. Left: Dependence of C on $\cos\theta$ for the five energy intervals considered in the analysis. Right: Dependence of C on the decimal logarithm of the energy (EeV) for the eight $\cos\theta$ intervals considered in the analysis.

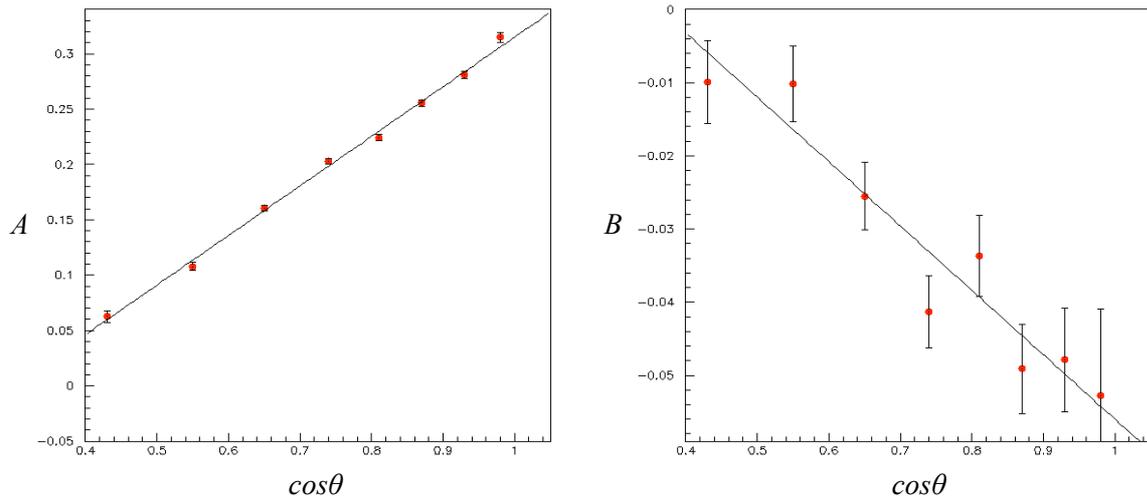


Figure 3.13. Dependence of A (left) and B (right) on $\cos\theta$.

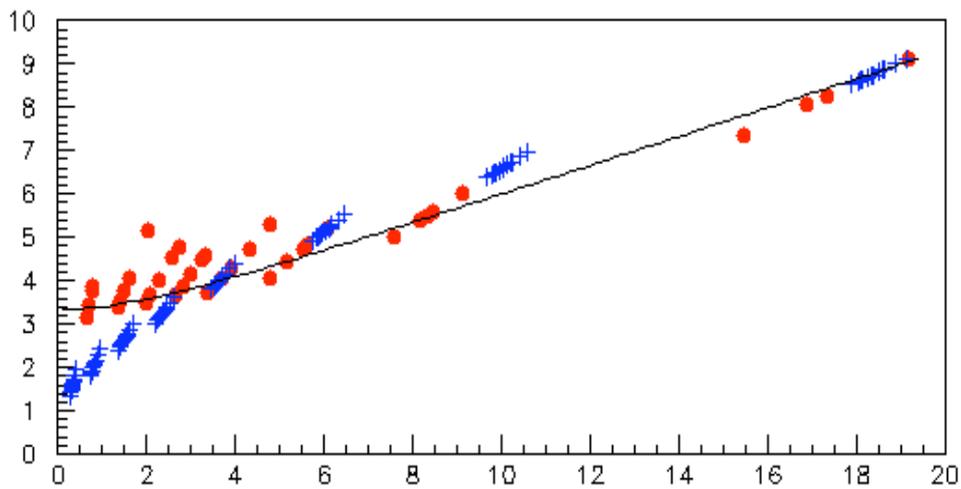


Figure 3.14. Dependence on radius of the altitude of the vertices measured from the PMT asymmetries in the single vertex approximation. The vertical and radial distances are measured in kilometers. Data are shown as red dots, one for each energy and for each zenith angle interval. The black line is the result of a fit to a hyperbola. Blue crosses are for X_{max} as taken from Reference 80 in the same energy intervals as for the data. The shower axis hits ground at the origin of coordinates.

The dependence of C on energy and angle is also illustrated in Figure 3.14 where the altitude $z_v = \cos\theta/C$ of the vertex (measured from ground) is plotted against its distance $r_v = \sin\theta/C$ from the impact on ground for each of the five energy intervals and compared with the location of the maximum of the shower development [80] used as a reference, X_{max} . There is no reason to expect single vertex and shower maximum to match, but for any physical process, the vertex location can be expected to be associated with a well defined

depth with respect to shower maximum. Such an association is however partly obscured by the absorption in the atmosphere between vertex and ground, but some correlation is surely to be expected. The X_{max} coordinates are evaluated under the assumption of an atmospheric depth (gcm^{-2}) having an exponential dependence on altitude z (km) (this time measured above sea level) of the form $1030exp(-z/8.7)$ and in the approximation of a constant PAO altitude of $1.4 km$: $z_{max}=8.7\ln\{1030/(X_{max}\cos\theta)\}-1.4$ and $r_{max}=z_{max}\tan\theta$.

The single vertex is observed to be at smaller depth than shower maximum for nearly vertical showers and at slightly larger depths for inclined shower, the cross over occurring at zenith angles of about 45° . The interpretation of this result requires a comparison with simulated events which has not been done yet.

3.4.2 Dependence on other parameters

Finally, in order to illustrate the sensitivity of the shower divergence to various quantities of relevance to the physics analysis of the data, Relation (2) is now used to study the dependence on such quantities. Precisely, C is written as $\Gamma C^*(\cos\theta, \log_{10}E)$ the best fit values taken by the parameters λ and Γ are evaluated as a function of the quantities under study. On average, Γ will be close to unity (not exactly because of small changes in the data sample with respect to the fit in Relation 2) and $\Gamma < 1$ (> 1) will correspond to larger (respectively smaller) vertex altitudes than average.

In order to illustrate the relevance of such measurements to the underlying physics, it is interesting to compare the asymmetries expected to be produced by muons and electrons respectively, of direct interest to mass composition studies.

A muon feeding through the tank from an incident direction (θ, φ) illuminates unevenly the walls of the Cherenkov tank, sending more light, on average, on the wall opposite to the incident direction (i.e. around $\varphi + \pi$). After the first diffusions, the light is randomized but the unbalanced “early” light causes the asymmetry observed between the PMTs, increasing with θ .

For incident photons, the situation is different: most of them are of low energy, and the Cherenkov light is mainly emitted by low energy conversion and Compton electrons: their momenta are at significant angles to the incident photon direction and suffer important multiple scattering. Moreover, the Cherenkov light is emitted over a short

trajectory, at the scale of a radiation length, generally less than a few 10 cm. The Cherenkov light is therefore emitted in nearly all directions from a compact source near the entrance wall of the incident photon. As a result, the asymmetry is smaller than for feed-through muons and may even be reversed for large values of θ .

The effect has been evaluated quantitatively [81] using a Monte Carlo simulation of the tank optics with uniform fluxes of muons or photons incident on the tank. Fig. 3.15 illustrates the effect. It shows the average relative light excess in a given PMT as a function of its azimuth ψ measured with respect to $\varphi + \pi/2$, at $\theta = 50^\circ$, for a muon of 500 MeV kinetic energy (upper panels) or for a 10 MeV photon (lower panels). Fits of the forms $I + A \sin \psi$ (left panels) and $I + A' \sin \psi + B \sin 2\psi + C \cos 2\psi$ (right panels) have been performed. For values of θ not exceeding 60° , the asymmetry coefficient takes the approximate form $\lambda \sin \theta$ with larger values of λ for muons than for photons, the difference becoming smaller at higher energies. Moreover, low energy muons that emit Cherenkov light in only the early part of their trajectory in water are found to behave as low energy photons, as expected; on the contrary, higher energy electrons or photons, producing a longer and narrower shower in water, are found to behave more and more as feed through muons, again as expected.

Figure 3.16 shows the dependence of I on the time of arrival of the shower particles. Precisely, in each tank, the FADC trace is resolved into peaks and peaks having a charge smaller than 0.3 VEM are ignored. The FADC trace is then divided in five equally populated time intervals¹² and the divergence is calculated in each interval using exclusively the peaks of this interval. A similar procedure is applied for the other quantities considered here (Figures 3.17 and 3.18). In all cases the statistical uncertainties are negligible, the quality of the fit is excellent and the sensitivity is large (putting all fits together, the values taken by λ are Gaussian distributed around a mean value of 23 % with an rms of 3 %).

¹² This very crude way to operate is obviously meaningless in cases of very short traces. A more serious study would need to use a more sophisticated definition of the time intervals.

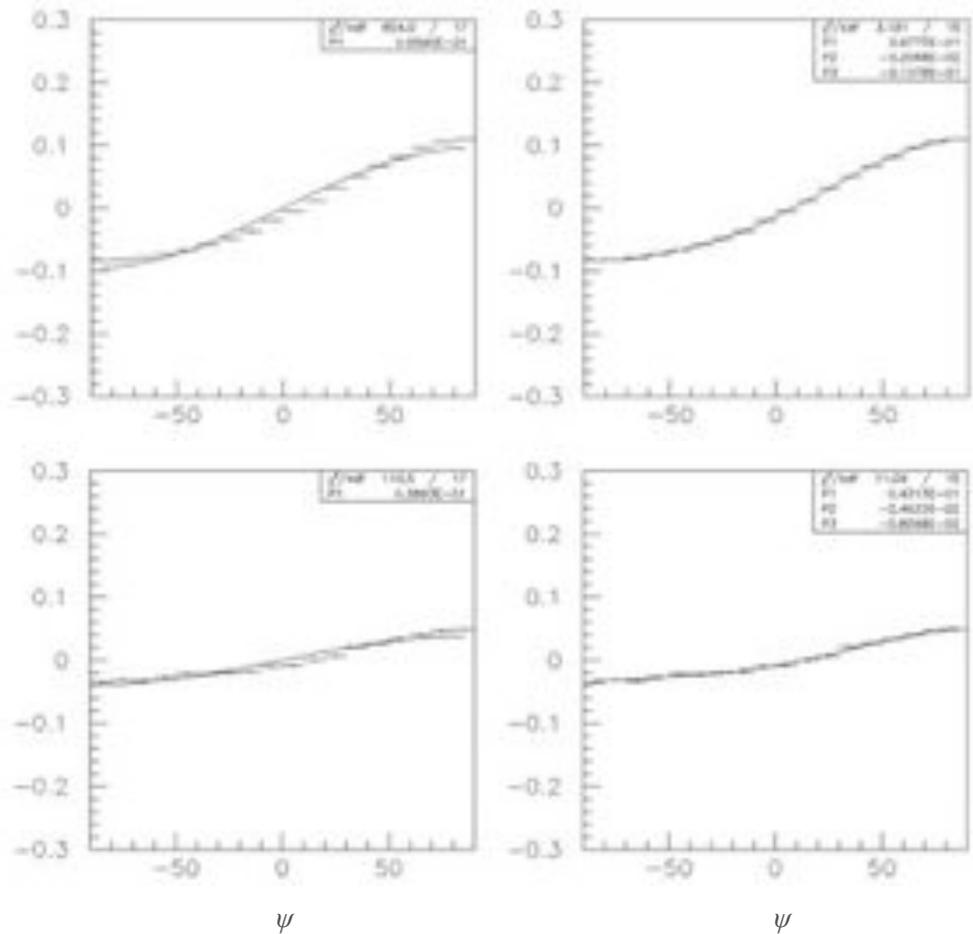


Figure 3.15. Average relative light excess in a given PMT as a function of its azimuth ψ (see text).

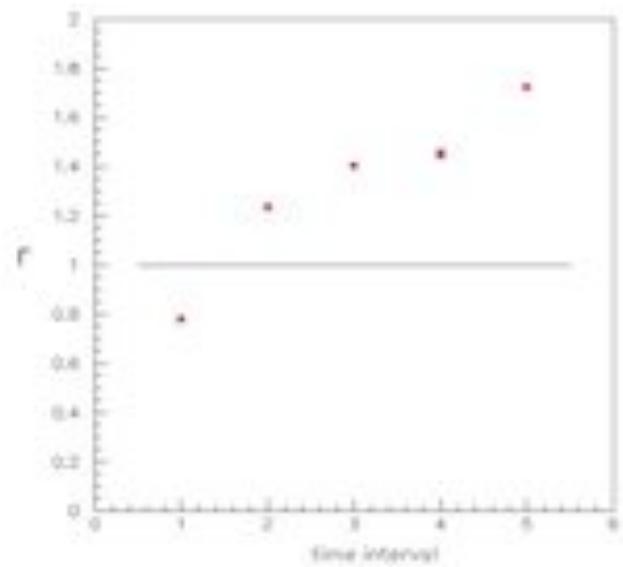


Figure 3.16. Dependence of Γ on the relative time of arrival (see text).

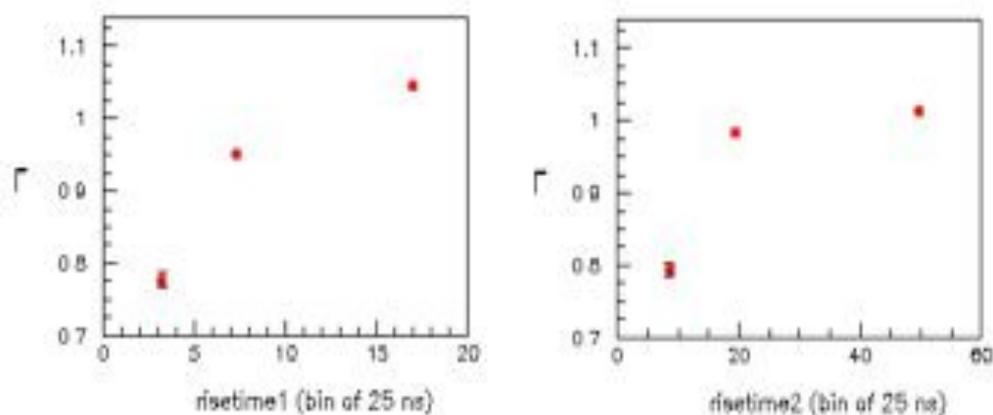


Figure 3.17. Dependence of Γ on rise time defined from 10% to 50% (left) or 80% (right).

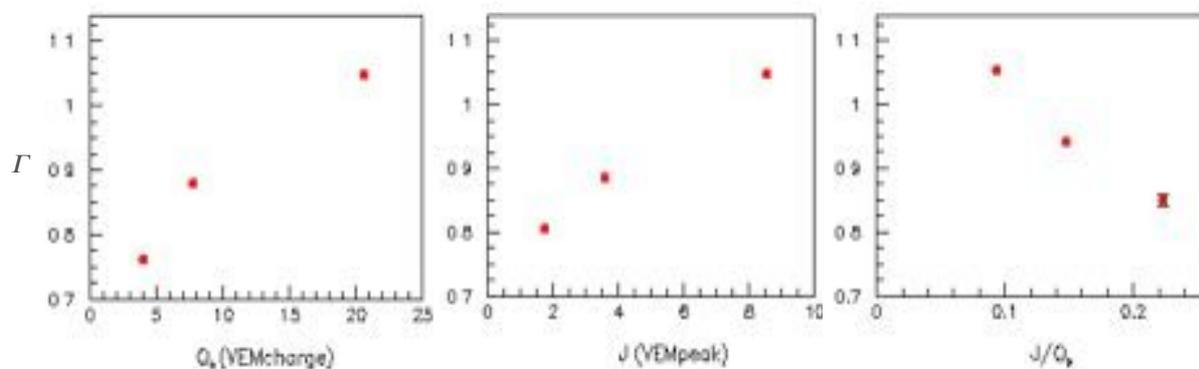


Figure 3.18. Dependence of Γ on (from left to right) Q , J and J/Q (see text).

The dependence of the divergence on the time of arrival (Figure 3.16) shows a strong correlation, with over a factor 2 between late and early times. As muons are expected to preferentially populate early times, this would suggest that they tend to come from higher altitudes than average. Figure 3.17 shows the dependence of divergence on rise time. Here two commonly used definitions of rise time have been considered: between 10% and 50% and between 10% and 80%. Both display a similar dependence with a clear preference of short rise times for higher altitudes (by typically 20%), again in agreement with the higher muon population associated with short rise-time traces. Figure 3.18 shows the dependence of the divergence on parameters commonly used in jump analyses [71,82]: the total charge Q , the jump value J and their ratio¹³. A serious analysis should select tanks in specific intervals of their distance to the shower axis but the aim here is only to illustrate the sensitivity of the divergence to such quantities. However, the ratio J/Q is expected not

¹³ Note that Q is traditionally measured in VEM charge units and J in VEM peak units but when calculating their ratio J/Q , VEM peak units have been used for both.

to be very sensitive to the cuts applied and to be larger for muon rich events, suggesting again that, on average, muons come from a higher altitude than average.

Figures 3.16 to 3.18 illustrate the potential of the shower divergence as a tool to study shower properties. They demonstrate the sensitivity of the divergence (found to vary by large factors) to many parameters of relevance to the physics of the shower formation. They illustrate the quality of the fits with important and well defined asymmetries (23 ± 3 % on average). As statistical uncertainties are essentially negligible, they strongly encourage the pursuit of such analyses in a more systematic way.

3.5 Summary

The previously revealed correlation between the shower azimuth and the PMT early time asymmetries has been studied in detail using the peak finding algorithm presented in the preceding chapter. Evidence for a minute (1.6 permil) correlation of the early time asymmetry with a universal direction has been presented and commented upon. The strong correlation – 20 to 30% typically – between azimuth and asymmetry has been exploited to measure the apparent divergence of the shower under the assumption that all particles hitting the tanks come from a same vertex on the shower axis. The use of the peak finding algorithm has made it possible to study the apparent shower divergence as a function of various selection criteria applied to the peaks: small and large charges, early or late arrival times. Similarly it has been possible to select traces obeying other selection criteria: short and long rise times, low and high jump value, etc... These studies have illustrated the very good power and sensitivity of this tool to study global shower properties. They have suggested that muons usually aim at a higher altitude than average but more work, and in particular the availability of a reliable simulation, would be necessary to pursue and complete such studies.

4. On the decay of muons stopping in the SD tanks

4.1 General considerations

Low energy muons associated with Ultra High Energy Cosmic Ray (UHECR) showers may stop in the Cherenkov water tanks of the Surface Detector (SD) and decay after some time. At variance with high energy muons that feed through the tanks, their flux is expected from simulations to be essentially independent from the nature of the primary – light or heavy nucleus. The detection of the decay electrons provides therefore useful information of relevance to the detailed understanding of the detector and to the quality of the simulation of shower development and detector response. Their study is the subject of the present chapter. Some preliminary general considerations will illustrate the main features of such a study.

4.1.1 Time range

The Flash Analog to Digital Converters (FADC) that equip the Photo Multiplier Tubes (PMT) of the SD tanks cover a time interval of nearly 20 μs , sampled in bins of 25 ns, which can be split in three regions:

- a first region, called the “early region”, precedes the arrival of the shower particles in the tank and extends typically over 2.5 μs . It may contain accidental signals and can therefore be used to monitor their rate;

- a second region, called the “main signal region”, covers typically 5 μs or so and contains the bulk of the shower particles, electrons, photons and muons. Its extension, density and particle composition depend strongly on the zenith angle of the shower and on the distance D of the tank to its axis;

- a third region, called the “late region”, covers typically 12 μs and is available for the study of late signals such as late shower particles, PMT after-pulses and, of specific interest to the present study, electrons from muon decays.

4.1.2 Muon energies

Depending on the muon impact point and zenith angle θ , the maximum path length in water varies between 1.2 m at $\theta=0^\circ$ and 2.4 m at $\theta=60^\circ$. These are the ranges of muons having momenta of 340 MeV/c and 550 MeV/c respectively. Lower energy muons have shorter path lengths. Stopping muons correspond therefore to the low energy fraction of the muon spectrum. The identification of the parent muon is in general not possible. In most cases it is not sufficiently delayed with respect to the shower front to emerge from the main signal region. Indeed, a slow muon having just enough energy to penetrate 10 cm into the tank had a velocity of 90% that of light only 300 m before reaching the tank: ionization losses in the atmosphere are such that the delay caused by the longer time of flight of low energy muons applies only to the last hundred or so metres of their path. The Cherenkov momentum threshold is 120 MeV/c and muons having a momentum of 200 MeV/c radiate only half the number of photons that relativistic muons radiate: even if the parent muon were isolated, its identification would often be uneasy.

4.1.3 Muon lifetime

Showers develop through a sequence of hadronic interactions on target nuclei in the atmosphere that contain approximately equal numbers of protons and neutrons, namely on nucleons having an average electric charge of $+0.5$ units. However, such interactions produce a large multiplicity of secondaries and the excess of positively over negatively charged mesons is diluted and strongly reduced. *In fine* one expects a slight excess of positive over negative muons. These decay with different lifetimes as negative muons may be captured in water¹⁴ while positive muons cannot: the μ^+ lifetime [83] is $2.19703 \pm 0.00004 \mu\text{s}$ while the μ^- lifetime [84] is $1.7954 \pm 0.0020 \mu\text{s}$ (if captured by oxygen). In an extensive air shower, the positive and negative muons are produced in practically equal numbers so the decay distribution after the arrival of stopping muons is proportional to $(\exp(-t/\tau_+)/\tau_+ + \exp(-t/\tau_-)/\tau_-)$. When shifting the origin by ΔT (typically $8\mu\text{s}$ for the present study) the decay distribution becomes $K_+ \exp(-t/\tau_+) + K_- \exp(-t/\tau_-)$, with $K_+ = \exp(-\Delta T/\tau_+)/\tau_+$ and $K_- = \exp(-\Delta T/\tau_-)/\tau_-$. To first order in t , it is proportional to $\exp(-t/T_{\text{eff}})$ where $1/T_{\text{eff}}$ is the weighted average of $1/\tau_+$ and $1/\tau_-$ with respective weights

¹⁴ In practice, negative muons are captured by oxygen rather than hydrogen atoms.

K_+ and K_- : $1/T_{eff} = (K_+/\tau_+ + K_-/\tau_-)/(K_+ + K_-)$. The later the observation, the larger the μ^+/μ^- ratio and the closer T_{eff} to τ_+ . For $\Delta T = 8\mu s$, $T_{eff} = 2.04\mu s$.

4.1.4 Electron signals

Decay electrons have a universal momentum spectrum starting at zero with a mean value of 37 MeV/c and an end-point at 53 MeV/c. The range of such electrons in water is less than 20cm, so most of them are contained inside the tank: the light collected by the PMTs has a universal distribution not exceeding 0.25 VEM with a mean value of 0.14 VEM. One FADC count corresponds to 0.0054 VEM on average. The detection and identification of such decay electrons requires therefore a good control over the measurement of small signals, in particular over possible base-line fluctuations.

The Cherenkov light produced by an electron shower (in one or so nanosecond) is usually diffused several times on the tank walls before reaching the PMTs. The excellent quality of the optical cavity (very small absorption in water and in the walls) results in a signal [77] decaying with a characteristic decay time of the order of 70 ns. The first four bins of the signal contain typically 78% of the total charge.

4.1.5 Scope of the study and data sample

The aim of the present study is the identification and counting of muon decay electrons (MDE) in SD data and a comparison with the predictions of simulations, according to which high energy muons are a good discriminator between proton and iron primaries, while low energy stopping muons are only little dependent, if at all, on the mass of the primary nucleus. This is illustrated in Figure 4.1 where simulated events associated with primary protons and primary iron nuclei respectively are compared. It may be explained by the fact that showers from heavy nuclei develop earlier, so the muons have a longer path to ground which disfavours the muons of lower energy. Because the energy of real events is obtained from $S(1000)$ [85], one should compare simulated events having a same value of $S(1000)$, i.e. a same value of the *reconstructed* energy (not the same *simulated* energy). To a good approximation, $S(1000)$ and the number of muons are nearly proportional to energy: a small relative variation of energy gives the same relative variations of $S(1000)$ and of the number of muons. As a consequence, in order to obtain

simulated showers giving the same averaged value of $S(1000)$ as a given sample of data, one simply scales showers simulated at E_{sim} and giving $S_{sim}(1000)$ by a factor $S(1000)/S_{sim}(1000)$: the number of muons is then scaled by the same factor. In practice, it is sufficient to compare the ratios of the numbers of muons to $S(1000)$.

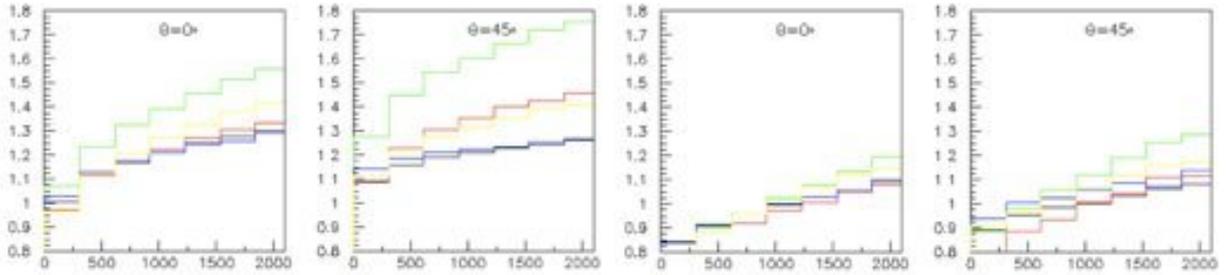


Figure 4.1. Ratio of the number of muons (normalized to S_{1000}) produced in iron showers to that in proton showers as a function of D for different models and shower libraries (AIRES QGSJET1, AIRES QGSJET2, AIRES SYBILL1, AIRES SYBILL2 and CORSIKA QGSJET2+ FLUKA displayed in black, red, green, blue and yellow respectively). The leftmost are for high energy muons ($E > 550$ MeV); the rightmost are for low energy muons (stopping muons, $100 \text{ MeV} < E < 300$ MeV); in each pair, the left and right panels correspond to zenith angles of 0° and 45° respectively.

In the following, simulated showers having fixed energy E_{sim} and zenith angle θ_{sim} are used. Their distribution of $S(1000)$ is observed to cover an interval having a mean value S_{mean} . Showers having a zenith angle in a given interval around θ_{sim} are then selected from the real data. An interval of $S(1000)$ is then chosen such that its mean value is S_{mean} . The widths of the energy and angle intervals are chosen not too small for the statistics to be sufficient and not too large for second order corrections to be negligible.

Muon decays provide therefore a check of the validity of the simulation independently from the mass of the primary nucleus and validate, at least in part, conclusions that may be drawn on the primary composition from the study of high energy muons. Moreover, independently from the nature of the primary, the comparison with different shower models predicting different abundances of stopping muons may provide some discrimination between these models. Figure 4.2 compares the numbers of muons predicted by different shower models, using AIRES+QGSJET01 as a reference. The numbers of stopping muons predicted by different models show significant differences, sometime exceeding 20%.

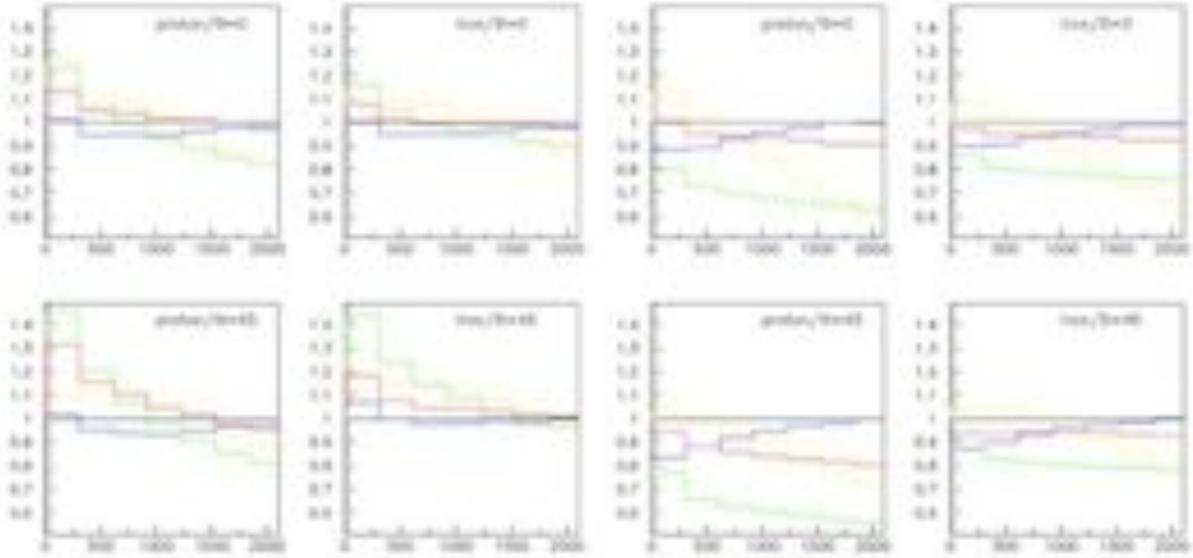


Figure 4.2. Ratio of the number of muons (normalized to S_{1000}) predicted by different models to that predicted by AIRES QGSJET01 (AIRES QGSJET1, AIRES QGSJET2, AIRES SYBILL1, AIRES SYBILL2 and CORSIKA QGSJET2+FLUKA displayed in black, red, green, blue and yellow respectively) as a function of D . The four panels on the left are for high energy muons ($E > 550$ MeV), the four panels on the right are for low energy muons (stopping muons, $100\text{MeV} < E < 300$ MeV).

In order to be free of trigger biases and still take advantage of the large statistics of SD data, the study concentrates on a shower sample having an average energy of $\sim 10^{18.5}$ eV selected according to the corresponding value of S_{1000} . A preliminary selection of showers having $S_{1000} > 5$ and $\theta < 60^\circ$ was made from $T4$ data [86] collected between January 2004 and April 2008. Tanks with a total charge exceeding 40 VEM are systematically excluded as having too noisy a decay region (after-pulses and late shower particles); this cut corresponds to an effective lower bound on the distance of the tank to the shower axis $D_{min} \sim 750$ m.

4.1.6 Simulated data

Comparison is made with simulated data at $10^{18.5}$ eV. Showers are taken from the AIRES [36] shower library with primary protons or iron nuclei at zenith angles $\theta = 0, 25, 36, 45, 53$ and 60° (corresponding to $\cos\theta = 1, 0.9, \dots, 0.5$). There are typically hundred different showers for each type of primary and each angle. Each shower is set at random in 500 different positions over the tank array, producing 500 showers with the same primary,

energy and zenith angle. Muon decays are generated with a common lifetime set at $2.2 \mu\text{s}$ independently from the muon charge.

The interaction of the shower particles with water is simulated in each tank using an earlier code [87] written with no particular emphasis on low energies. Photon interactions other than pair creation and Compton scattering are negligible and neglected. Charged particles are followed stepwise, accounting for energy loss and multiple scattering. Radiation losses by bremsstrahlung are taken into account but the radiated photons, which are mostly low energy, are ignored. Hadrons are also ignored. At each step, Cherenkov photons are emitted on the Cherenkov cone in proportion to the square of the sine of its opening angle and are followed until hitting a PMT or being absorbed in water or in the tank walls. The light reaching a PMT decreases exponentially with time with a decay constant of $\sim 70 \text{ ns}$ dominated by the absorption in the walls and in water. The number of photons per centimetre of water is taken to obey the Cherenkov law and the conversion to VEM units is evaluated from simulated vertical muons for self consistency. FADC traces are evaluated in units of VEM and 25ns time bins. Baseline fluctuations mimicking real data are added but the under-shoot following large signals is ignored, as are after-pulses. Accidental signals extracted from the early part of real traces are used to add an accidental component to the simulated traces having both the right rate and the right charge distribution.

4.1.7 General outline

In Section 4.2 following, selection criteria are devised in order to select decay electron candidates with a good efficiency and a small background contamination. These are often the result of successive iterations which are described in detail elsewhere [88]. The results of Section 4.2 are used in Section 4.3 to identify backgrounds and to define how to reject them. Section 4.4 describes the method used to count stopping muons in the SD data. Attention is paid to identify as explicitly as possible the steps which make use of the predictions of the simulation. The last section, 4.5, presents the results, compares them with predictions and comments on their implications.

4.2 Selection criteria

The selection of MDE candidates includes four steps: a preliminary baseline correction, rejection of after-pulses, charge cuts and a trace-by-trace definition of the late region. Throughout the present chapter FADC trace stands for the bin-by-bin average of the three individual PMT traces of the tank being considered.

4.2.1 Baseline correction

Following the study presented in Chapter 2, an estimator for the base line is taken in 64 intervals of 12 bins each as the minimal value taken by the charge content of a bin in that interval. For times exceeding 260 FADC bins (i.e. $6.5\mu\text{s}$) and total charges Q_{tot} not exceeding 200 VEM (the present study is restricted to $Q_{tot} < 40 \text{ VEM}$) it is observed to have a universal shape and variable amplitude. The shape is well described by a parabola between bins 260 and 400 and a constant value, q_{ref} , above bin 400 (Figure 4.3). The amplitude increases with Q_{tot} as shown in Table 2.1 below.

Average value of q_{ref} as a function of Q_{tot} .

Q_{tot} range	q_{ref} (VEM)
$Q_{tot} < 40 \text{ VEM}$	-0.05
$40 \text{ VEM} < Q_{tot} < 100 \text{ VEM}$	-0.10
$100 \text{ VEM} < Q_{tot} < 200 \text{ VEM}$	-0.17
$200 \text{ VEM} < Q_{tot} < 500 \text{ VEM}$	-0.26
$500 \text{ VEM} < Q_{tot} < 1000 \text{ VEM}$	-0.38
$Q_{tot} > 1000 \text{ VEM}$	-0.97

As was mentioned and discussed earlier (Section 2.1.2) the bump appearing around bin 500 (Figures 2.5 and 4.3) is associated with after pulses of low amplitude (0.12 VEM on average) and density proportional to the charge of the earlier signals. In the present study where the total charge of the trace is restricted to 40 VEM, the effect is anyhow nearly negligible and the bump has been excluded from the base line subtraction procedure.

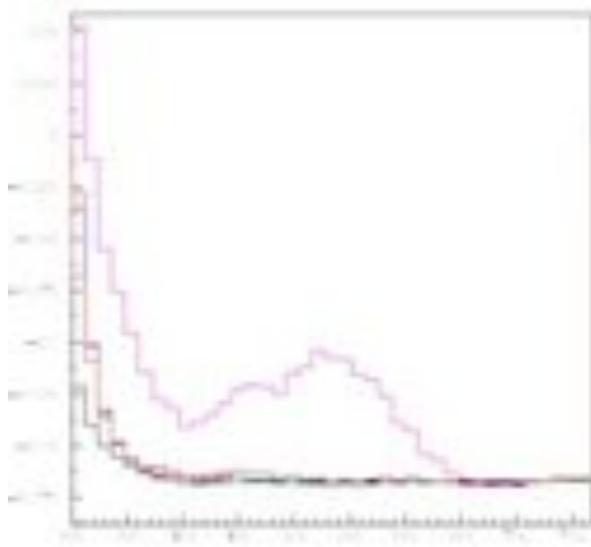


Figure 4.3. Baseline shapes normalized to late times. The outstanding curve is for $Q_{tot} > 1000$ VEM, all other curves are for different intervals of $Q_{tot} < 1000$ VEM (see table).

It is recalled that the algorithm used to subtract the base line on a trace-by-trace basis consists in 1) finding the minimum value, q_{min} , of the bin charge over the 48 last bins of the trace; 2) finding the average value in this interval, q_{av} , of the bin charges not exceeding $q_{min} + 3$ FADC counts; 3) subtracting a baseline defined as the reference curve scaled by q_{av}/q_{ref} .

4.2.2 After-pulses

Again following the study presented in Chapter 2, standard segments are defined as a string of FADC bins having charges above a threshold of 0.02 VEM and separated by at most two bins having charges below this threshold. After-pulses are expected to be short (typically one bin) and to occur in only one of the three PMTs at a given time. These two features make their identification easy. Figure 4.4 shows the ratios of the PMT charges to the average charge for one-bin segments and for wider segments respectively.

We want to plot the three ratios $r_i = q_i / (q_1 + q_2 + q_3)$ where q_i are the segment charges defined separately for each PMT by summing over the segment bins and the preceding and following bins. As $r_1 + r_2 + r_3 = 1$, it is convenient to plot these in trilinear coordinates. In practice Figure 4.4 is a plot of $(q_1 - \frac{1}{2}(q_2 + q_3)) / (q_1 + q_2 + q_3)$ vs $\sqrt{3}/2(q_2 - q_3) / (q_1 + q_2 + q_3)$. The difference between one-bin segments and wider segments is striking making it easy to reject as after-pulse any one-bin segment having at least one of its r_i in excess of 0.7 . This

procedure allows for an efficient rejection of any after pulse segment having a charge in excess of 0.02 VEM.

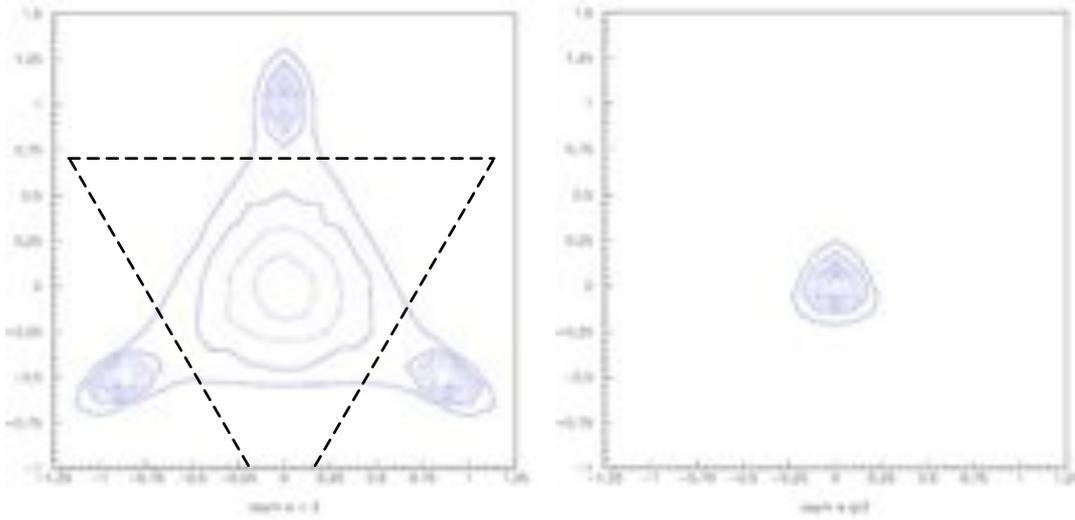


Figure 4.4 Left: Trilinear plot of the PMT asymmetries as defined in the text for one-bin segments. The dashed triangle corresponds to the cut applied. Right: Same as left for wider segments.

4.2.3 Charge cuts

Standard segments (defined above) are used to build the main signal (the string of standard segments separated by at most 10 FADC bins having the maximal total charge, see next subsection) and to define MDE candidates. The latter should not significantly exceed four time bins in length: they are ideally exponential with an integral not exceeding 0.25 VEM and a decay time of 70 ns, nearly three time bins. Keeping in mind that bin charges smaller than 0.02 VEM (*i.e.* at most 3 FADC counts) cannot reliably be distinguished from noise, it should be sufficient to look for signals contained within four time bins. A strong asymmetry may occur in the first slot before the randomization of the light within the tank, which takes about 20ns; after which the PMT signals are similar within fluctuations.

In view of the above comments, rather than simply using standard segments to define MDE candidates, it is preferable to use fixed length (four bins) segments. Such a four-bin segment, starting at bin i , is associated with each standard segment starting at bin i_{start} , i being that of i_{start} , $i_{start}+1$ and $i_{start}+2$ which gives the largest charge, q_{max} , to the four-bin segment. At the same time, two control regions are defined, one between bins $i-4$ and $i-1$ and the other between bins $i+4$ and $i+7$. One calls q_- and q_+ the charges of the

preceding and following control regions respectively. When the electron candidate signal is isolated, one expects q_- to cancel (within baseline fluctuations) and q_+ not to exceed $\sim 1/3$ of q_{max} (corresponding to the exponential tail of the signal).

Simulated data are now used to define charge cuts aimed at selecting MDE candidates. The distributions of q_{max} , q_- and q_+ for signals in the early and late regions (to be defined precisely in the next subsection) are shown in Figures 4.5 and 4.6 respectively.

In the early region the presence of muons (off scale in Figure 4.5 left) causes the bump in the q_+ distribution (Figure 4.5 right, blue) which is absent in the late region (Figure 4.6 right, blue). The q_- distribution is confined to low values in both the early and late regions (Figures 4.5 and 4.6, right, red). This shows that in both the early and late regions the isolation is already quite good. The q_- distribution has a slightly smaller tail in the late region than in the early region, because in the case of the larger signals occurring in the early region it may happen (as found by scanning traces associated with the tail of the q_- distribution) that bin $i-1$ contains some significant charge. This effect is essentially negligible for the smaller charges expected for MDEs. The q_{max} distributions differ at high charges, as expected from the higher relative muon abundance in the early than late region, but none of them suffers significant background at low charges.

Isolation cuts are applied to both q_- and q_+ . A cut at 0.025 VEM on q_- is safely above baseline fluctuations (Figures 4.5 and 4.6 right, red). The cut on q_+ should depend on q_{max} , as a constant fraction (typically 1/3) of the main signal is expected to spill out in this control region. In the present range of charges, below 0.5 VEM, one expects a cut on q_+ of the form $q_+ < 0.015 + 0.29 q_{max}$ to be adequate. Figure 4.7 shows the 2-dimensional distributions of q_+ versus q_{max} together with the areas selected by the cut for the early and late regions separately.

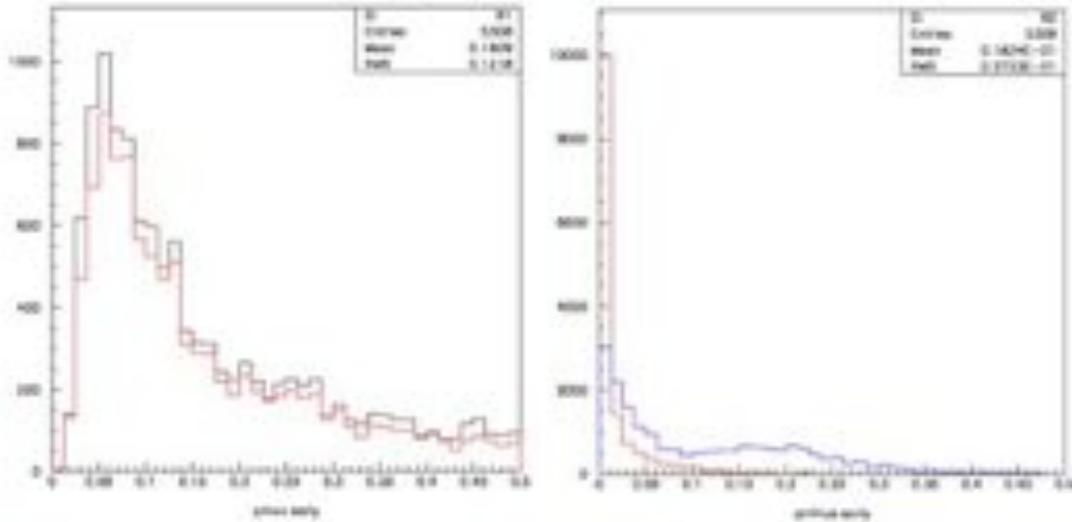


Figure 4.5 Charge distributions q_{max} (left panel), q_- (right panel, red) and q_+ (right panel, blue) for signals occurring in the early region. The q_{max} distribution (left panel) is shown before (black) and after (red) application of the isolation cut.

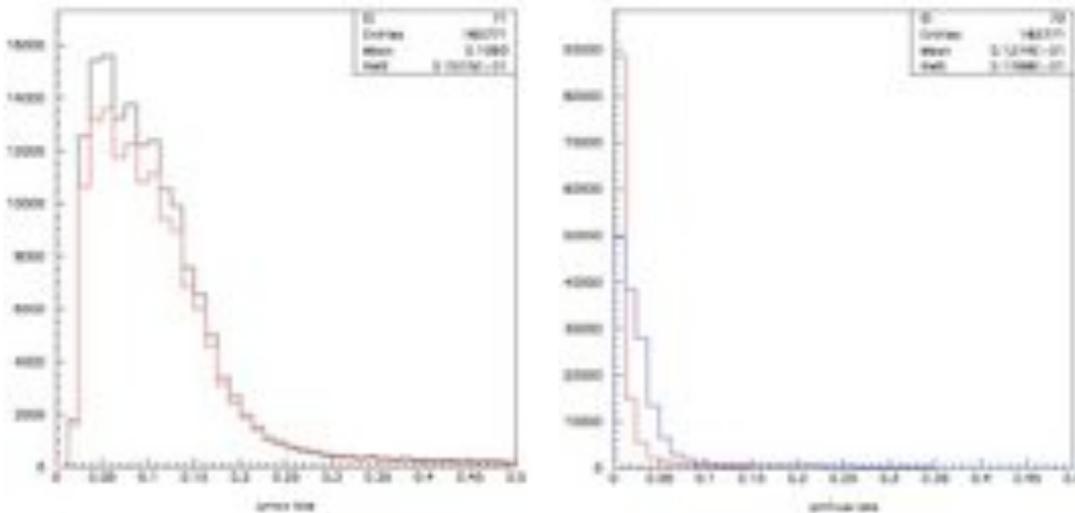


Figure 4.6 Charge distributions q_{max} (left panel), q_- (right panel, red) and q_+ (right panel, blue) for signals occurring in the late region. The q_{max} distribution (left panel) is shown before (black) and after (red) application of the isolation cut.

The application of the isolation cuts results in charge distributions displayed in the left panels of Figures 4.5 and 4.6 in red colour. As seen from Figures 4.7 and 4.6 left, the loss of events is very small (12%). One might argue that a less strict cut or a stricter cut should be used depending on the relative importance given to efficiency versus background rejection.

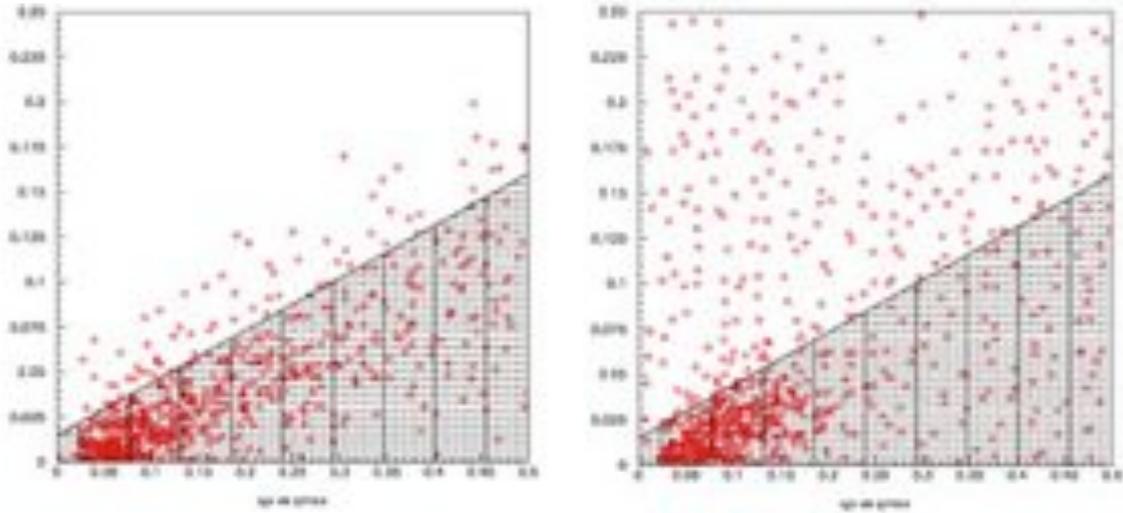


Figure 4.7 Two-dimensional distributions of q_+ (VEM, ordinate) versus q_{max} (VEM, abscissa) for the early region (left panel) and the late region (right panel). The areas selected by the cut have been shaded.

In summary, the following charge cuts are retained for the definition of MDE candidates:

- $q_{max} < 0.25 \text{ VEM}$;
- $q_- < 0.025 \text{ VEM}$;
- $q_+ < 0.015 \text{ VEM} + 0.29 q_{max}$.

As will be shown in Section 4.4 an additional cut, $q_{max} > 0.08 \text{ VEM}$, will be needed to reject a low charge background.

4.2.4 Trace-by-trace definition of the late region

The late region (starting at bin i^*) should be chosen as large as possible but it has to satisfy two conditions:

- i^* should be after the tail of the main signal, hence $i^* > i_l + 100$ where i_l is the end of the main segment. The origin of time is taken as the start of the main segment which is the best available approximation for the edge of the shower front. The front itself cannot be evaluated exactly; it is not even precisely defined. Figure 4.8 shows distributions of i_l for FADC traces taken from $10^{18.5}$ eV simulated proton showers having $Q_{tot} < 40 \text{ VEM}$ and $D < 1800 \text{ m}$. It extends down to 120 (140) time bins for $\cos\theta = 0.6$ (1) respectively, *i.e.* $i_l + 100$ extends down to 220 (240) time bins.

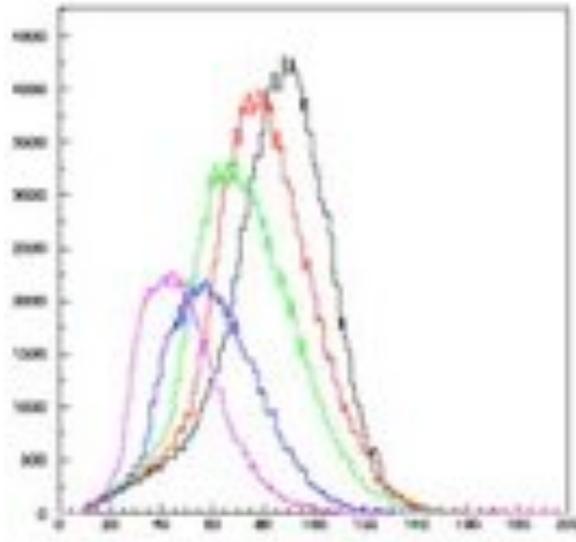


Figure 4.8. Distribution of i_l for different values of $\cos\theta$ (from 1, black, to 0.6, purple, in steps of 0.1). FADC traces are from $10^{18.5}$ eV simulated proton showers and have $Q_{tot} < 40$ VEM and $D < 1800$ m.

– Moreover, in order to make the analysis as transparent as possible, it is preferable to avoid any significant overlap with the time interval during which the parent muons stop in the tank. The probability for a muon stopping in water to arrive between time t and time $t+dt$ is called $F(t)$ with $\int F(t)dt=1$, integrated from 0 to ∞ . $F(t)$ depends on shower properties: zenith angle, energy, etc. as well as on the distance D of the tank to the shower axis. Distributions of $F(t)$ are shown in Figure 4.9 for different zenith angles for the simulated traces used above. Figure 4.10 shows the dependence of their mean and *rms* values over the cosine of the zenith angle. At smaller zenith angles the distribution of $F(t)$ extends to larger values. It is seen from Figure 4.9 that $F(t)=0$ beyond 200 time bins, meaning that in the late region all decay electrons have parent muons stopping before 200 time bins.

Combining these results, the value $i^* = \text{Max}(i_l+100, 200)$ is retained as the definition of the first bin of the late region.

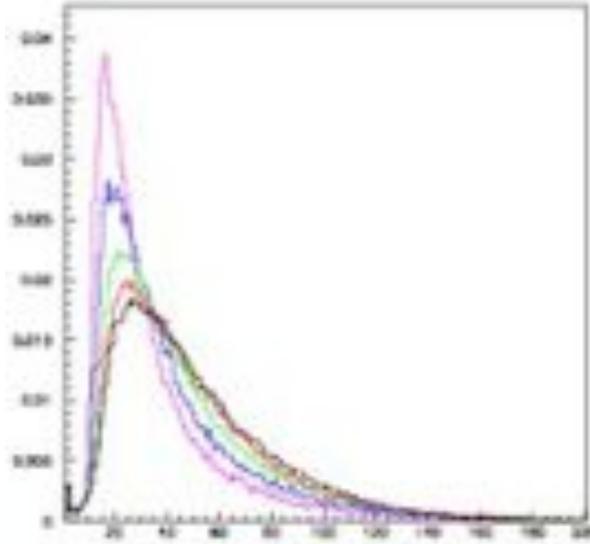


Figure 4.9. Distributions of $F(t)$ for different values of $\cos\theta$. FADC traces are from $10^{18.5}$ eV simulated proton showers and have $Q_{tot} < 40$ VEM and $D < 1800$ m. Bins of 25 ns are used as a unit and the normalization is such that the sum of F over all bins is equal to 1. Hence F is measured in units of $(\text{bin of } 25 \text{ ns})^{-1}$.

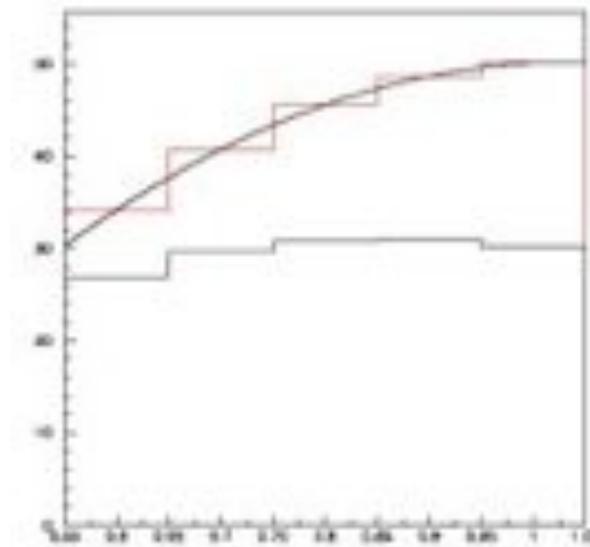


Figure 4.10. Dependence of the mean (red) and *rms* (black) values of $F(t)$ on $\cos\theta$. The traces considered here are the same as displayed in Figure 4.9. The curve is a quadratic fit to the mean values.

Figure 4.11 compares the distributions of the times of stopping muons measured with respect to i^* for two intervals of the distance to the shower axis: $D < 1200$ m and $1200 \text{ m} < D < 1800$ m. The fraction of stopping muons extending beyond zero, namely into the late region, is always small but increases with the distance to the shower axis. This is seen more directly on Figure 4.12 which displays the ratio, in percent, between the

distributions of the distances to the shower axis of stopping muons having respectively positive and negative times measured with respect to i^* . Restricting the distance to the shower core to the region below 1800 m keeps this fraction below a percent.

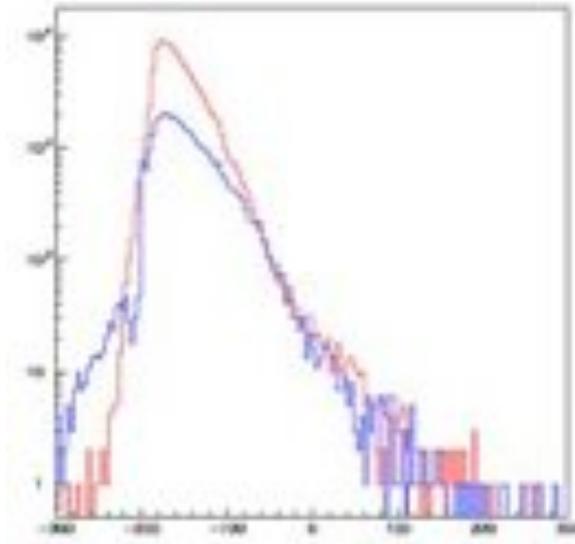


Figure 4.11 Distributions of the times of stopping muons measured with respect to i^* for two intervals of the distance of the tank to the shower axis: $D < 1200 \text{ m}$ (red) and $1200 \text{ m} < D < 1800 \text{ m}$ (blue). Bins of 25 ns are used as unit.

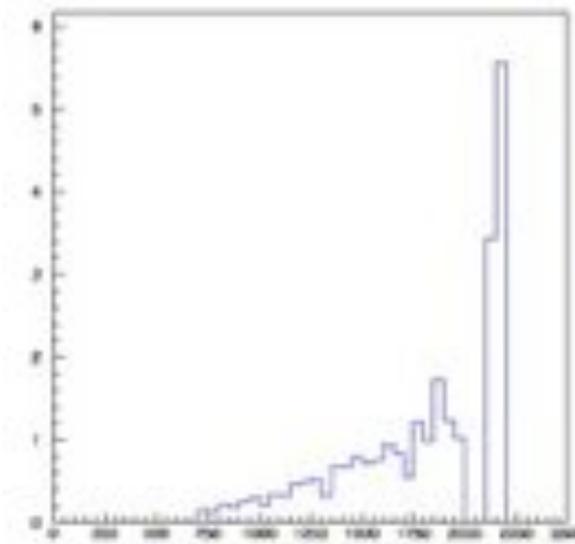


Figure 4.12 Ratio, in percent, between the distributions of the distances to the shower axis (measured in meters) of stopping muons having respectively positive and negative times measured with respect to i^* .

Figure 4.13 displays the distribution of i^* for both intervals of the distance to the shower axis $D < 1200 \text{ m}$ and $1200 \text{ m} < D < 1800 \text{ m}$. Close to the shower axis, $i_l + 100$ is larger

than i_0+200 while far from the shower axis the latter is larger. But, on average, (i_0+200) is always larger than i_1+100 . This is best seen in Figure 4.14 which displays the dependence of the mean value of $(i_0+200) - (i_1+100)$ on the tank distance to shower axis.

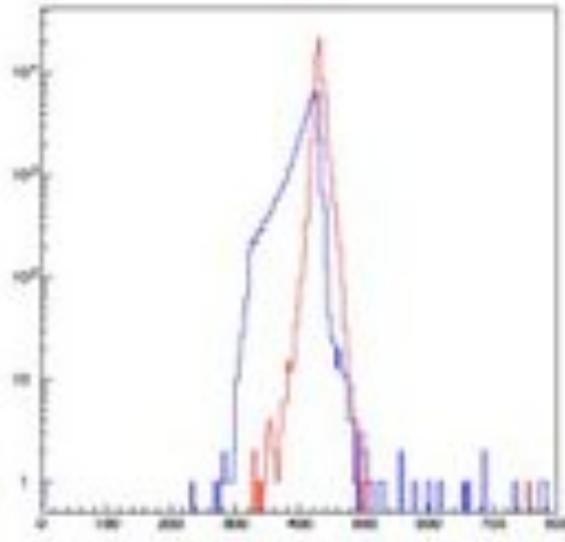


Figure 4.13 Distribution of i^* for both intervals of the tank distance to the shower axis, $D < 1200\text{m}$ (red) and $1200\text{ m} < D < 1800\text{ m}$ (blue).

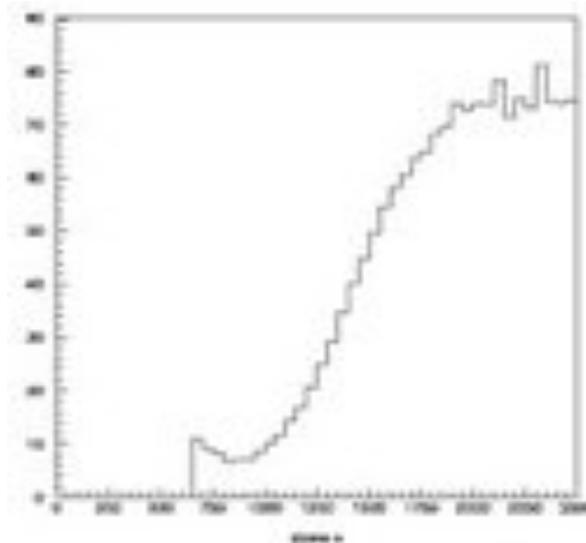


Figure 4.14 Dependence of the mean value of $(i_0+200) - (i_1+100)$ (measured in 25 ns time bins) on the tank distance to the shower axis (measured in meters).

Putting the above results together, the tank distance to the shower axis is required not to exceed 1800 m. As this result has been established at $10^{18.5}$ eV, it is generalized to include slightly different energies by defining an energy dependent cut [71] on the distance

to the shower axis (expressed in meters) as $D < 700 \log_{10} E(eV) - 11150$. As all above results were obtained from simulated showers, the validity of their extension to real data needs to be carefully examined.

4.3 Background sources

4.3.1 Simulated data

In a first step, a particular sample of simulated proton shower data having $\theta = 0^\circ$ and $D < 1200$ m is used as an illustration. Extension of the results to larger values of θ and D is kept for later. Generated electrons giving a charge q_{gen} in the tank and arriving at time t_{gen} corresponding to bin i_{gen} in the FADC trace are considered. In parallel, MDE candidates having charge q_{max} and starting at bin i are identified. Most generated electrons and most MDE candidates are associated with each other as evidenced in Figure 4.15 which displays the difference $i_{gen} - i$ for all selected (i_{gen}, i) pairs. Accordingly, the condition $-2 \leq i_{gen} - i \leq I$ is retained as association criterion.

The data fall in three families: generated electrons associated with an MDE candidate, generated electrons not associated with an MDE candidate and MDE candidates not associated with a generated electron. Their relative populations are 19079¹⁵, 6188 and 6662.

Considering the first family, the distributions of q_{gen} and q_{max} are compared in Figure 4.16. Their mean values are 0.151 and 0.117 respectively, in a ratio of 1 to 0.78, consistent with the earlier estimate (assuming that all electrons start at the beginning of the segment the fraction should be 0.76 for a light decay time of 70 ns but the real situation is more complicated). Also shown in Figure 4.16 is the distribution of $q_{max}/0.78$ which is in very good agreement with that of q_{gen} . By definition (association criterion) the time distributions of the generated electrons and of the associated segments are the same. They are displayed in Figure 4.17 taking i^* as origin of time; shifting the origin of time does not change the slope, but simply multiplies the distribution by a factor, as muon decays are described by a single exponential in the simulation. Exponential fits starting at least 20

¹⁵The number of electrons associated with segments is 19079, the number of segments associated with electrons is 18989. The difference (0.5%) occurs because it happens that more than one electron is associated with a segment.

time bins above bin i^* give a decay time of $2.20 \mu\text{s}$ (as used in the simulation) to within less than 2% with excellent values of χ^2 .

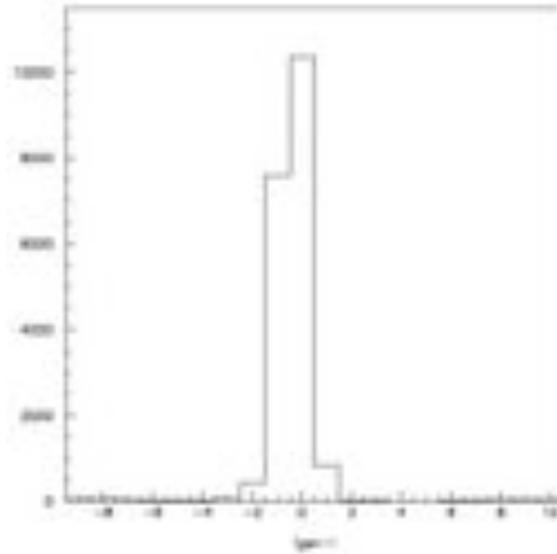


Figure 4.15. Distribution of the difference between the first bins of all (generated electron, MDE candidate) pairs which satisfy the selection criteria.

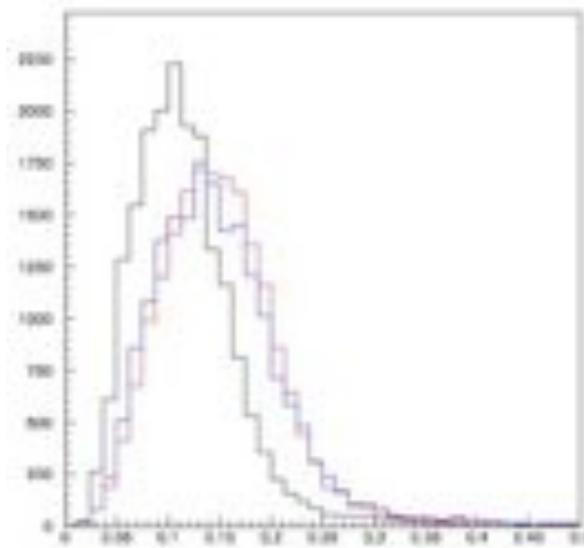


Figure 4.16 Charge distributions of electron candidates of the first family (see text). Black: q_{max} of the MDE candidates; blue: $q_{max}/0.78$ of the MDE candidates; red: q_{gen} of the generated electrons.

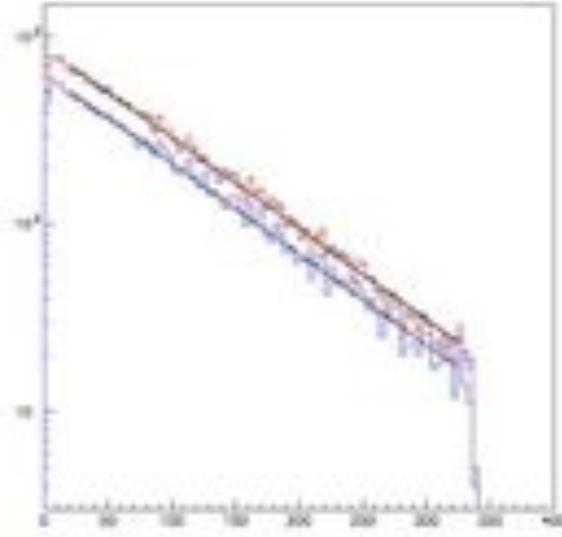


Figure 4.17 Time distributions of electrons of the first family (see text). Black: all MDE candidates; red: all generated electrons; blue: MDE candidates having $q_{max} > 0.08$ VEM. The lines are the result of exponential fits.

The second family (6188 electrons), generated electrons not associated with an MDE candidate, is made of two components. One corresponds to electrons (2544) associated with non-isolated four-bin-segments. It has the same charge distribution as those of the first family and amounts to 14% of it. The other contains electrons (3644) having generally low values of $0.78q_{gen}$, the expected value of q_{max} if there had been an associated segment, as illustrated in Figure 4.18. Of this second component, only 7% have $0.78q_{gen}$ in excess of 0.08 VEM: the reason for them not to be associated with a segment is simply that this segment would have too low a charge to be retained. Globally, when requiring the additional condition $q_{max} > 0.08$ VEM used to reject the low charge background (see below) the global efficiency of the applied selection criteria (on families 1 and 2 together) is $13969/25267 = 55.3$ %. The time distribution of the selected electrons is displayed in Figure 4.17. An exponential fit gives again a decay time of $2.20 \mu s$ to better than 2%.

The third family (6662 segments, of which 2419 have $q_{max} > 0.08$ VEM), MDE candidates not associated with a generated electron, contains an important contribution of accidentals as illustrated in Figure 4.19 where their time distribution is displayed (again taking i^* as origin of time). Also shown in the figure is the time distribution of MDE candidates in the early region.

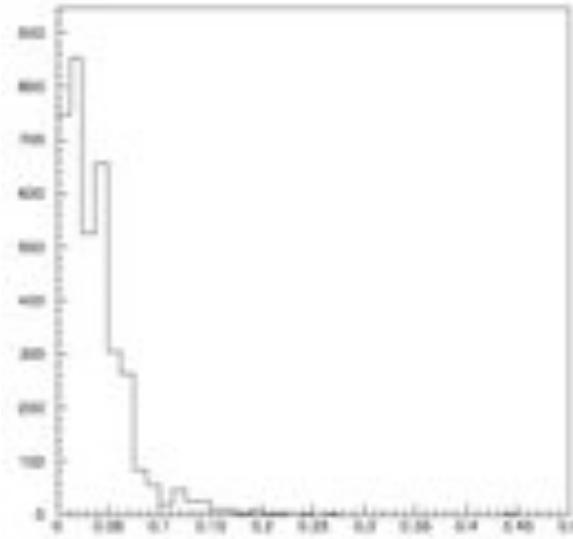


Figure 4.18 Generated electrons not associated with a four-bin segment (whether isolated or not). The charge q_{gen} has been multiplied by 0.78 to simulate what would have been the value of q_{max} if there had been a segment associated.

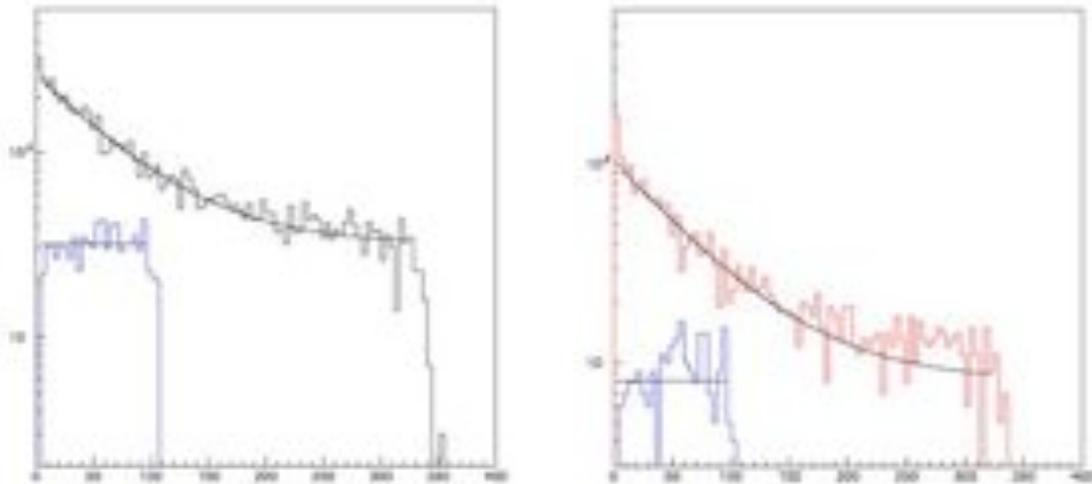


Figure 4.19 Time distributions of non associated MDE candidates (left, black) and of non associated MDE candidates having $q_{max} > 0.08$ VEM (right, red). 25ns wide bins are used as a unit and the origin of time is i^* . The blue histograms are for segments in the early region, the origin of time being now bin 1 (see text). The fits are the sum of an accidental component equal to that in the early region and an exponential describing the tail contamination.

The accidental level can be obtained by fitting the time distribution in bins 5 to 95 to a constant $A_{early}=33.3\pm 1.2$ per 100 ns bin. In addition to this accidental component a population of non associated segments is seen to concentrate toward low time values and is found to be associated with the tail of the main signal. A fit to a form $A_{early}+A_{tail}exp(-t/T_{tail})$ measures this additional component and gives $A_{tail}=237\pm 7$ per 100 ns bin and $T_{tail}=1.57\pm 0.05 \mu s$. Restricting the fits to charges $q_{max}>0.08 VEM$ gives $A_{early}=9.5\pm 0.7$ per 100 ns bin, $A_{tail}=98.7\pm 4.9$ per 100 ns bin and $T_{tail}=1.46\pm 0.07 \mu s$. The time dependence of the tail contamination is nearly unaffected by the $q_{max}>0.08 VEM$ cut and is steeper, but not much, than the muon decay distribution. Integrating the tail contamination gives a total of $N_{tail}=1451\pm 77$ segments out of 2419, namely 60% of the third family population.

The table below extends this analysis to simulated data having larger values of D and θ . The S_{1000} window is now made angle dependent in order to keep the mean energy near the value of $10^{18.5} eV$: $12.3-46.0 (1-\cos\theta)^2 < S_{1000} < 18.3-55.2 (1-\cos\theta)^2$.

Parameters of the tail contamination background obtained from simulated data.

D interval (m)	$\theta(^{\circ})$	$f_{tail}(\%)$	$T_{tail}(\mu s)$
<1050 $\langle D \rangle = 887$	0	8.6 ± 0.5	1.39 ± 0.05
	25	10.4 ± 0.5	
	45	16.8 ± 0.9	
$[1050, 1300]$ $\langle D \rangle = 1173$	0	12.8 ± 0.8	1.18 ± 0.03
	25	20.9 ± 0.9	
	45	33.3 ± 1.5	
$[1300, 1800]$ $\langle D \rangle = 1470$	0	18.7 ± 1.2	1.11 ± 0.10
	25	34.5 ± 1.2	
	45	42.0 ± 2.1	

The global cut efficiencies (fraction of generated decay electrons obeying the charge cuts including $q_{max}>0.08 VEM$) are consistent with a single value independent from zenith angle and from distance to the shower axis, $55.4 \pm 0.3\%$.

At a same distance the values of T_{tail} are observed to be nearly independent from zenith angle. The tail contamination is therefore evaluated by using a single value of T_{tail} for each distance D . From these results, parameterizations of the tail background have been

obtained, f_{tail} being the ratio N_{tail}/N_{sel} where N_{sel} is the number of MDE candidates having $q_{max} > 0.08$ VEM after subtraction of the accidental background, A_{early} :

$$T_{tail} = 2.93 - 2.50 \cdot 10^{-3} D + 0.855 \cdot 10^{-6} D^2$$

$$f_{tail}(\%) = (D - 595)(0.025 + 0.105(1 - \cos\theta)).$$

At large values of θ and D , where the main signal region is not densely populated, the tail contamination evaluated here is seen to become important and a less reliable counting of stopping muons is to be expected.

The charge distribution of the decay electrons has been checked against real data obtained from a local trigger [89] requiring, at the single tank level, that at least one bin of the FADC trace of each PMT exceed 3 ADC counts (~ 0.05 VEM) over a time window of 500 ns. Such data are free of any contamination by after-pulses or trailing shower particles. They are mostly muons and are essentially equivalent to the accidental signals of high energy shower events.

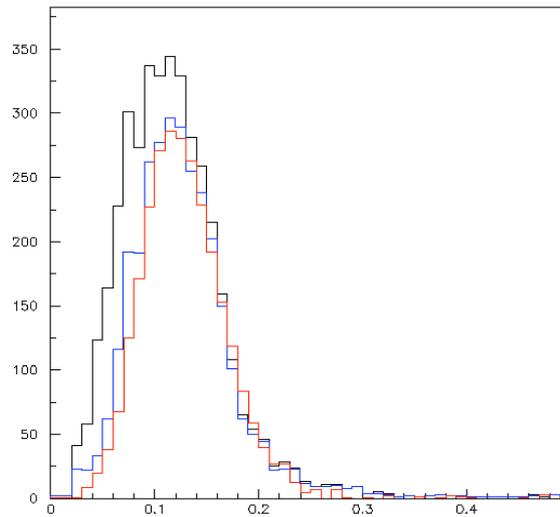


Figure 4.20 Charge distributions (q_{max} measured in VEM units) of decay electrons obtained from the local trigger data (red). It is compared with that obtained from the simulated data (black) and from the simulated data accounted for trigger condition (blue). The red and blue distributions are normalized in the charge interval [0.10 VEM, 0.25 VEM].

Some of these muons stop within the water volume and their decay electrons may generate a subsequent trigger. Their charge distribution can then be obtained by a subtraction method [89]. As shown in Fig. 4.20, this distribution is in good agreement with that of simulated decay signals when accounting properly for the trigger condition. This agreement validates the evaluation of the selection efficiency described above. However,

the trigger condition prevents making a reliable comparison for charges significantly smaller than 0.1 VEM.

4.3.2. Real data

In a second step real data are used, the aim no longer being to evaluate the efficiency of the selection criteria but to understand how well the data are described by the simulation and to identify possible additional backgrounds.

Figure 4.21 compares the charge (q_{max}) distributions of MDE candidates in real and simulated data. Accidentals have been subtracted using the early region as a reference.

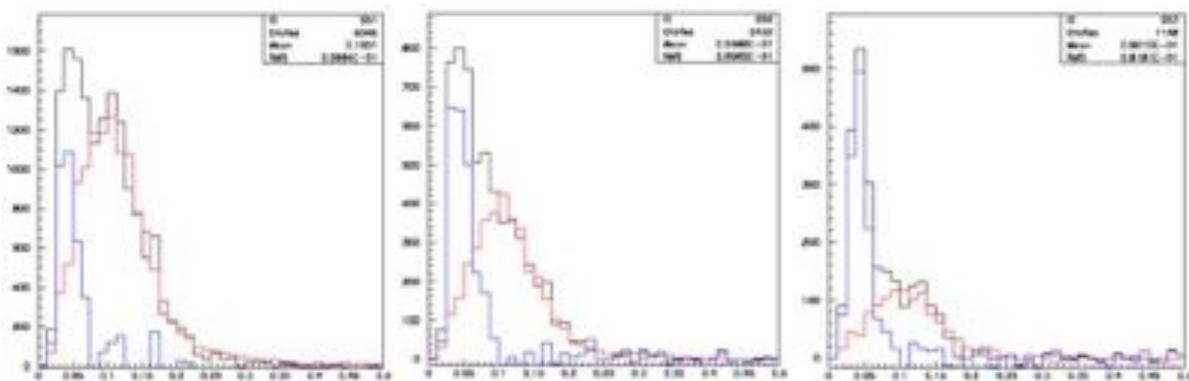


Figure 4.21 Charge (q_{max}) distributions in three equal time intervals measured from i^* as origin; from left to right: early, middle and late respectively. The real data (black) are compared with the simulated data (red) normalized in the charge interval [0.10 VEM, 0.25 VEM]. The blue histogram is the difference between the real and normalized simulated data, showing an excess at very low charges.

Here, as in the preceding subsection, a particular sample having low values of θ and D is used as an illustration. Real data have $\theta < 30^\circ$, $D < 1200$ m and ¹⁶ $12 < S_{1000} < 18$ while simulated data have an energy of $10^{18.5}$ eV and $\theta = 0^\circ$. The comparison is made in three equal intervals of $i - i^*$ with the idea that the background contamination might depend on the proximity of the main signal (at least the background associated with the tail of the main signal does, as it was found in the simulation to decrease more steeply than muon decays). The normalization between real and simulated data is such as to have a same number of events for charges q_{max} between 0.1 VEM and 0.25 VEM.

¹⁶ This S_{1000} interval gives a mean energy of $10^{18.4}$ eV.

An excess of events having charges smaller than 0.1 VEM is present in the real data in each of the three time intervals. Two examples of traces with such low charges are shown in Figure 4.22.

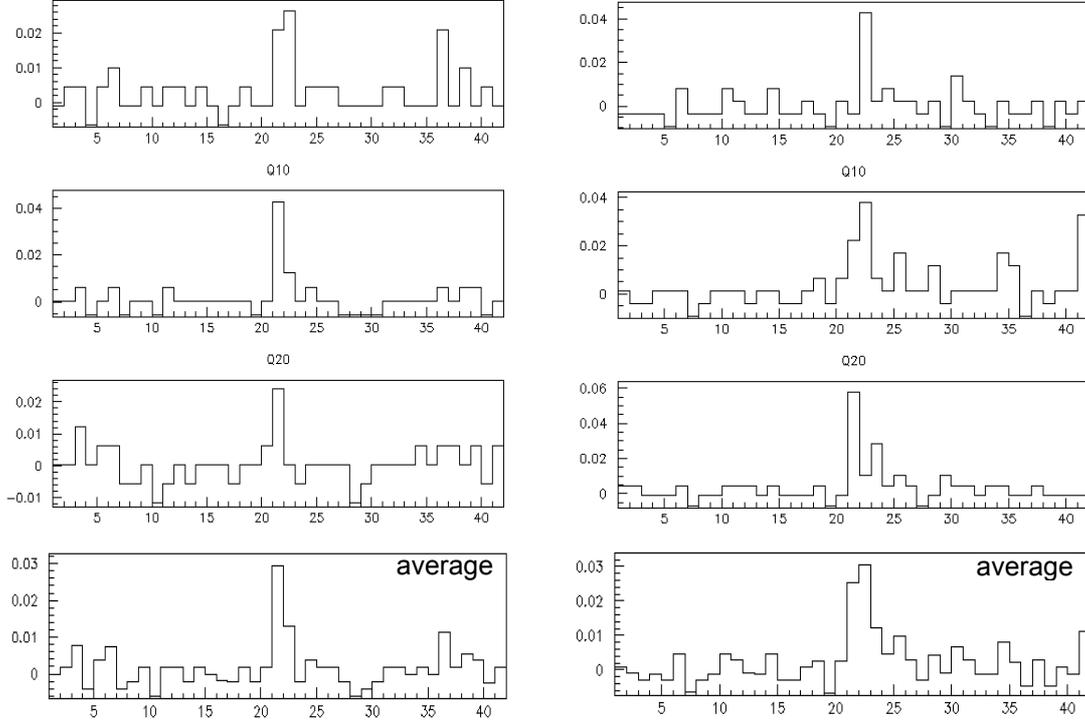


Figure 4.22 Two examples of FADC traces displaying low charge signal (< 0.08 VEM). The three upper panels correspond to the three PMT traces and the lower panel to the average trace.

The correlation between the three PMT traces makes it unlikely that they be due to baseline fluctuations or to uncorrelated noise such as after-pulses. A possible interpretation is presented in Subsection 4.3.4 in terms of late shower nucleons of low energy, producing signals just above the proton Cherenkov threshold (~ 450 MeV). The analysis has been repeated for tanks having D between 1200 m and 1800 m; the results are given in the three lower lines of the table below which lists, for each of the time and D intervals, the following quantities (after normalization of the simulated data as described above):

- N_1 is the number of real data MDE candidates having $q_{max} > 0.08$ VEM, after subtraction of the level of accidentals measured in the early region;
- N_2 is the number of simulated data MDE candidates having $q_{max} > 0.08$ VEM, after subtraction of the level of accidentals measured in the early region;
- N_3 is the number of N_2 signals associated with a real electron (in that case, one must be careful not to subtract the accidental contamination beforehand). As was remarked earlier a

number of simulated MDE candidates are not associated with real electrons. Some are associated with accidentals, but those have already been subtracted. The others (*i.e.* $N_2 - N_3$), are associated with the tail of the main electron/photon signal and are concentrated at early times;

– N_4 is the difference between the numbers of real and simulated MDE candidates having $q_{max} < 0.08$ VEM. It decreases with time and with D , but much less fast than N_1 does, as qualitatively expected for low charge after-pulses. The possible contamination of such a background for $q_{max} > 0.08$ VEM is estimated not to exceed 1% globally of the retained MDE candidate samples and is ignored in what follows.

Lines 1 to 3: $D < 1200$ m, lines 4 to 6: 1200 m $< D < 1800$ m.

<i>Time</i>	N_1	N_2	N_3	$(N_2 - N_3)/N_1$	N_4	N_4/N_1
<i>early</i>	1531	1528	1368	10.5±0.8 %	592	0.4
<i>middle</i>	471	443	411	6.8±1.2 %	365	0.8
<i>late</i>	136	131	121	7.4±2.3 %	205	1.5
<i>early</i>	752	749	614	18.0±1.5 %	215	0.3
<i>middle</i>	210	222	201	10.0±2.2 %	174	0.8
<i>late</i>	48	44	44	0.0±3.1 %	96	2.0

A number of additional checks have been made to validate the comparison between real and simulated data. In particular a comparison between the time distributions of real and simulated data (using i^* as origin of time) is displayed in Figure 4.23 for three q_{max} intervals separately (0 to 0.08 VEM, 0.08 to 0.16 VEM and 0.16 to 0.24 VEM) and a comparison of the distributions of the distances to the shower axis is shown in Figure 4.24 for MDE candidates having $q_{max} > 0.08$ VEM.

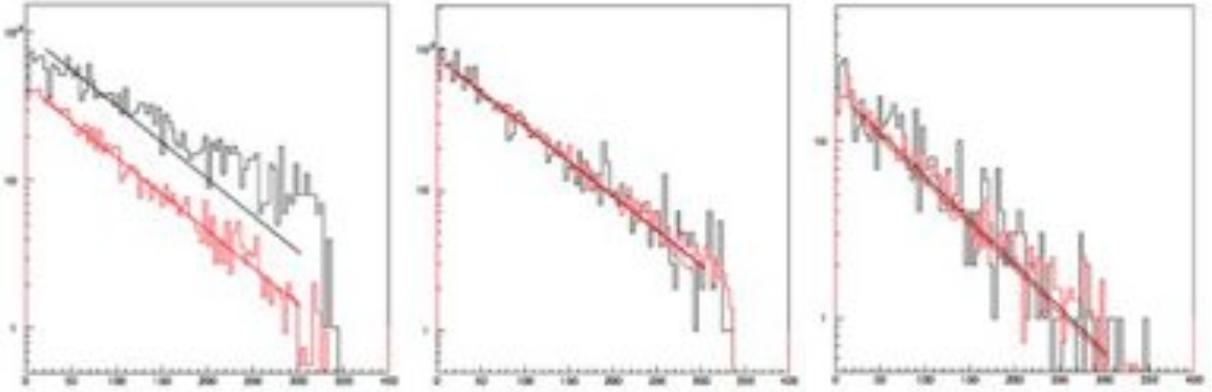


Figure 4.23 Comparison of the time distributions (units are 25ns bins) of real (black) and simulated (red) MDE candidates measured from i^* as origin in different charge (q_{max}) intervals (from left to right, 0 to 0.08 VEM, 0.08 to 0.16 VEM and 0.16 to 0.24 VEM). The normalization factor is common to the three panels and is simply the ratio of real to simulated MDE candidates having $q_{max} > 0.08$ VEM.

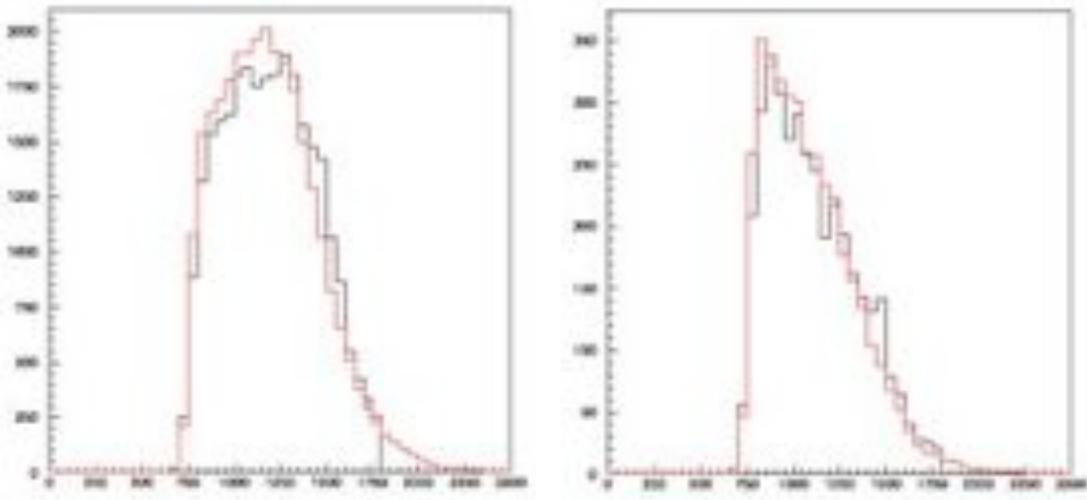


Figure 4.24 Distributions of the distance D (m) to the shower axis for real (black) and simulated (red) data; left panel: all traces; right panel: MDE candidates having $q_{max} > 0.08$ VEM. Real and simulated distributions are normalized.

Figure 4.25 compares the time distributions measured from i^* for the samples of real and normalized simulated MDE candidates having $q_{max} > 0.08$ VEM and $D < 1200$ m. In both cases accidentals have been subtracted using the early region. The results of an exponential fit give life times of 1.94 and 2.16 μ s respectively. The distribution of simulated signals displayed in Figure 4.26 but not associated with an electron is shown in Figure 4.27; an exponential fit gives a decay time of 1.39 μ s. Figure 4.28 is the same as Figure 4.26, after subtraction of the non-electron-associated background displayed in Figure 4.27. Exponential fits give lifetimes of 2.04 ± 0.05 and 2.28 ± 0.05 μ s respectively. As

the simulation uses $2.2 \mu\text{s}$ as muon lifetime while a shorter effective lifetime, of the order of $2.0 \mu\text{s}$, is expected when muon capture is accounted for, these results are consistent with expectation and imply the absence of significant additional background contaminations.

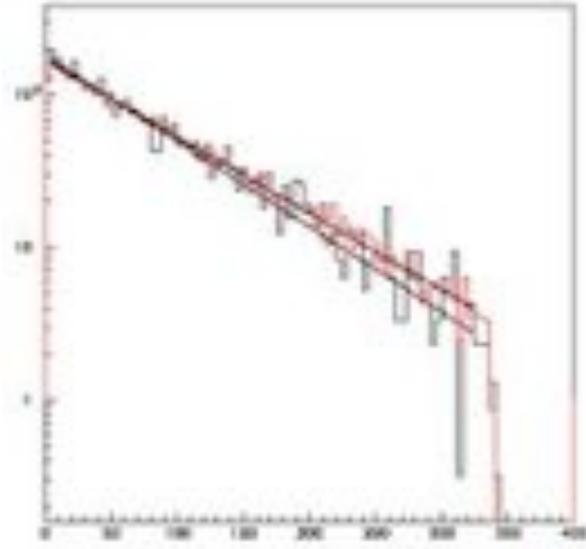


Figure 4.25 Time distributions measured from i^* for the samples of real and normalized simulated MDE candidates having $q_{max} > 0.08 \text{ VEM}$. In both cases accidentals have been subtracted using the early region. 25ns time bins are used as a unit.

In summary, after application of the charge cut $q_{max} > 0.08 \text{ VEM}$ one is left with two main background contaminations. One, due to accidentals, dominates at large times and the other, associated with the tail of the main signal, dominates at short times. The first one is subtracted by using the early region as reference and the uncertainty on this subtraction is essentially statistical. The second one is subtracted by using the simulation and the uncertainty on this subtraction is difficult to evaluate as we do not understand in sufficient detail the mechanism which is causing the contamination.

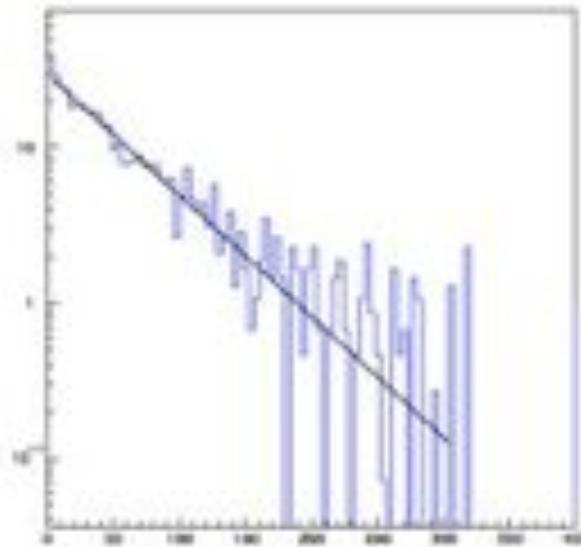


Figure 4.26 Time distribution of simulated MDE candidates displayed in Figure 4.25 but not associated with an electron. An exponential fit gives a decay time of 1.39 μs . 25ns time bins are used as a unit.

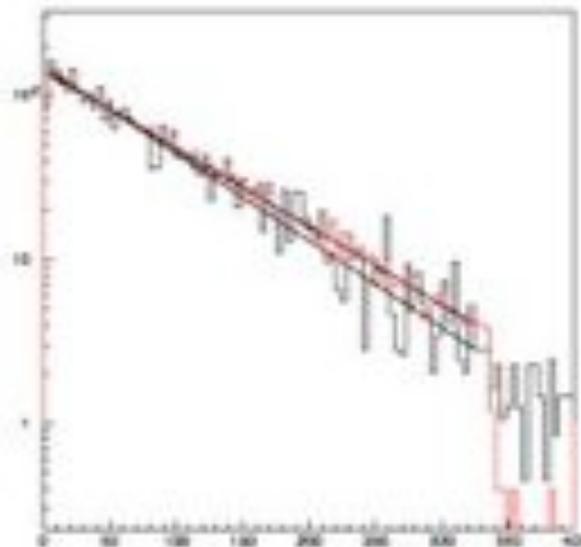


Figure 4.27 Time distributions (in units of 25ns bins) measured from i^* for the samples of real and normalized simulated MDE candidates displayed in Figure 4.25 after subtraction of the non-electron-associated background displayed in Figure 4.26. In both cases accidentals have been subtracted using the early region. Exponential fits give lifetimes of 2.03 and 2.27 μs respectively.

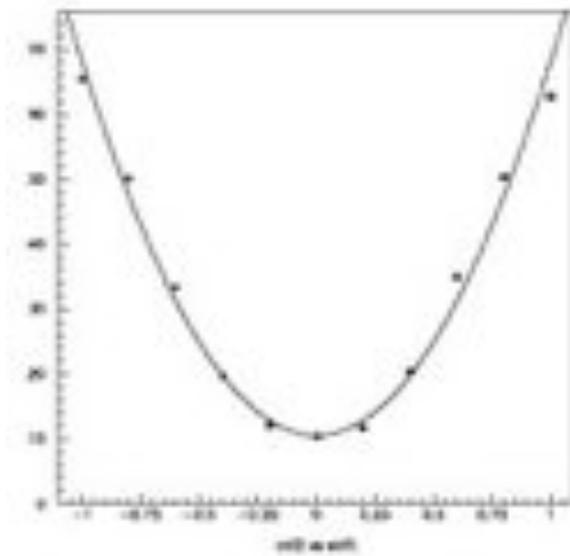


Figure 4.28. Dependence on the charge shift (measured in units of $0.01 VEM$) of the χ^2 measuring the quality of the match between the charge distribution of real data and that, shifted, of normalized simulation data. The χ^2 is calculated for charges (q_{max}) between 0.08 and $0.25 VEM$.

4.3.3. Systematic uncertainties

A potentially important uncertainty is not associated with background subtraction but with the global cut efficiency (fraction of generated decay electrons obeying the charge cuts including $q_{max} > 0.08 VEM$) which, as was seen earlier, amounts to $55.4 \pm 0.3\%$ independently from the values taken by θ and D .

In order to evaluate the systematic uncertainty attached to this number, the reliability of the description of the charge distribution given by the simulation needs to be assessed. This is done by evaluating by which quantity the simulated charge distribution can be shifted with respect to the measured charge distribution while still retaining a good match. Figure 4.28 shows the dependence of the associated χ^2 over the charge shift. The minimum χ^2 is equal to 10.5 for 16 degrees of freedom. The acceptable shift is evaluated as the shift that causes an increase of the χ^2 by one unit. It amounts to $0.0013 VEM$. Such a shift would change the fraction of retained events by 1.0% implying a cut efficiency of $55.4 \pm 1.0\%$.

This, however, addresses only one aspect of the cut efficiency. The 45% of electrons rejected by the cuts are in fact distributed in two families: 24% are simply not identified with MDE candidate segments and 21% are identified but have a value of q_{max} smaller than $0.08 VEM$. The 1.0% uncertainty which has just been evaluated applies to this

second family exclusively. As far as the first family is concerned, the reasons causing an electron not to be identified with an MDE candidate segment are either that the electron is not associated with any segment, usually because the charge is too small, (14%) or that it is associated with a non isolated segment (10%). We estimate that the global uncertainty on these two numbers should not exceed 10% of the total 24%, namely, combining the relative uncertainties on both families, we retain a final value of 55 ± 3 % for the global cut efficiency.

4.3.4. Nucleon background

As protons loose energy continuously in the atmosphere and as neutron decays are negligible during the shower development, the nucleonic component on ground contains mainly neutrons. Their delay increases with decreasing energy, so after a few microseconds their energies do not exceed 1 GeV. This is illustrated in Figure 4.29 for 3 EeV vertical proton showers (CORSIKA [35]).

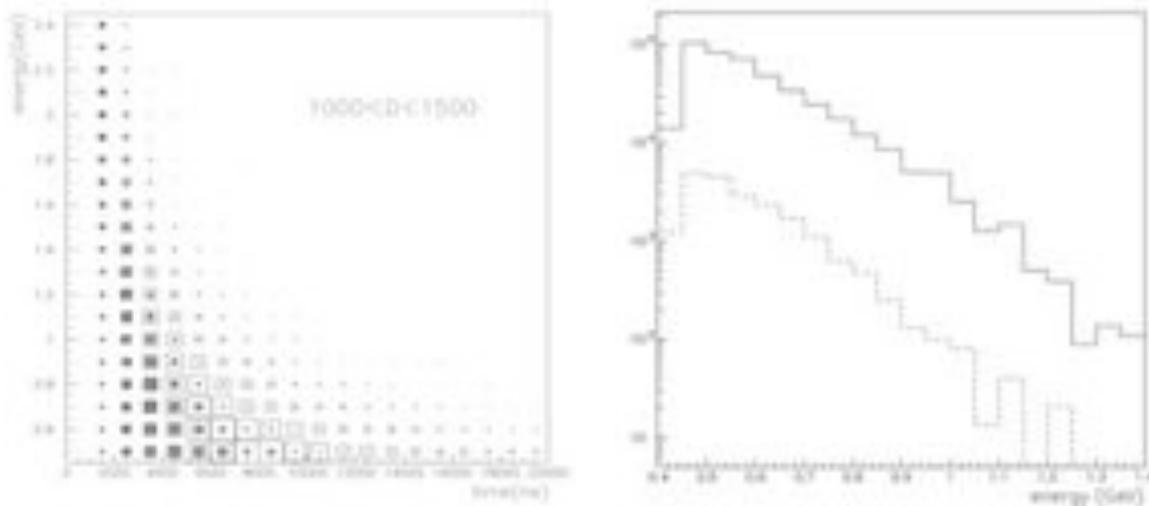


Figure 4.29. Left: Distribution in time and energy of the nucleons from simulated showers with $1000 < D < 1500$ m; neutrons are represented by white boxes and protons by grey ones. Right: Energy spectrum of neutrons (full line) and protons (dashed line) after $5 \mu\text{s}$ for $800 < D < 1800$ m.

The neutrons that reach a tank interact in water and transfer part of their energy to protons or nuclei which may produce Cherenkov light, together with the remaining component of shower protons. The signal from protons is given in Figure 4.30 as a function of the kinetic energy: with a steep energy spectrum, most protons are below 650

MeV, giving signals below 0.1 VEM. At medium or high energy, nucleon interactions can produce mesons which give additional light, but the fraction of such nucleons is very small.

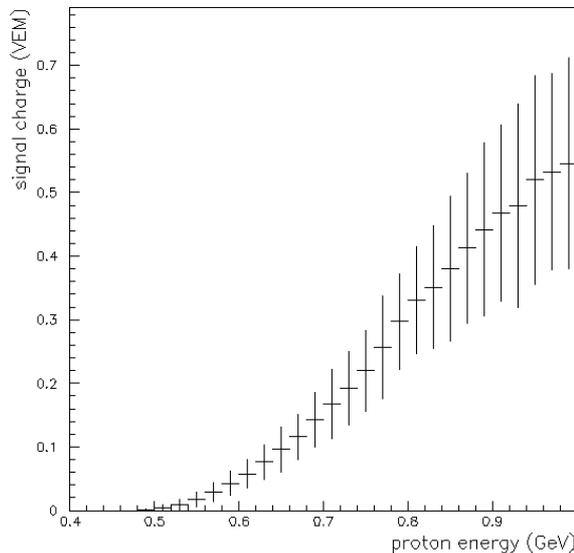


Figure 4.30. Mean signal (in VEM) associated with protons as a function of the kinetic energy.

A realistic estimate of the nucleon signal requires a reliable simulation of the hadronic cascade down to low energies; from this point of view CORSIKA with FLUKA is probably a better combination than AIRES, which predicts less nucleons at ground level. It also requires an accurate description of the elastic and inelastic interactions of nucleons in water. Here we use a very simple model to obtain an order of magnitude estimate of the effect:

- Protons produce Cherenkov light until they interact; after the interaction, light is no longer emitted.
- Neutrons scatter elastically on protons that may then produce Cherenkov light; when interacting with oxygen nuclei (with a relative probability of 7:3 with respect to elastic proton scattering), they keep their energies and only change direction (no Cherenkov light is emitted but the neutron gets a new chance to scatter on a proton).
- For both neutrons and protons, the interaction length is taken to be 60 cm.

The model probably overestimates the signal. However if pions are produced, they may give light at energies much lower than 500 MeV.

The charge distribution of nucleon signals obtained from this simulation is shown Figure 4.31. It shows an accumulation at low charges (below 0.1 VEM), as seen in the real data, and a steep decrease at higher charges.

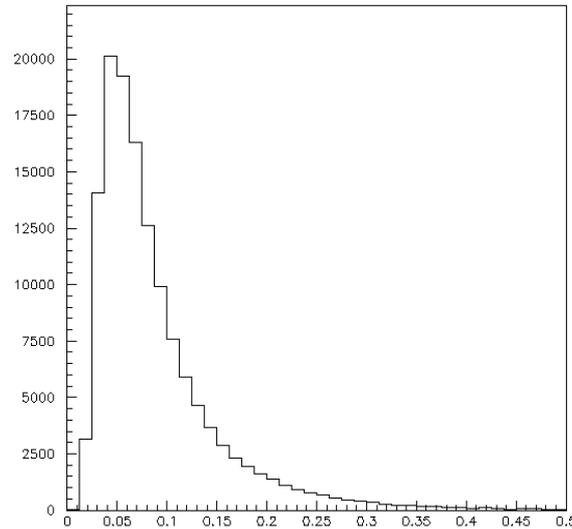


Figure 4.31. Charge distribution (q_{max} measured in VEM units) associated with nucleons as obtained from the simple model described in the text.

Such a simple model predicts that 27% to 41% of the charge distribution associated with nucleons extends above 0.08 VEM, depending on the delay with respect to the shower front. The fraction above 0.3 VEM is too small to allow a measurement beyond the electron signal.

The data displayed in Figure 4.21 have been analyzed in terms of the sum of two properly scaled components: a nucleon component having a charge distribution of the form displayed in Figure 4.31 and an electron component having a charge distribution of the form displayed in Figure 4.20. The result implies nucleon contaminations (i.e. ratio of nucleon component to electron + nucleon component) of 28, 39 and 58% in the early, middle and late time intervals respectively (Figure 4.32). This provides an upper limit to the nucleon contamination as it blames the totality of the low charge signal on nucleons. However, such a contamination increasing rapidly with time would imply a number of muon decays decreasing faster than expected. We note that the low charge spike observed in the subtracted real data background (Figure 4.21) is systematically sharper than in the simulated nucleon associated charge distribution (Figure 4.31). Under the assumption that the latter is properly simulated, this would imply that the actual contribution of the nucleon

associated background is significantly smaller, by a factor 2 or so, than the upper limits evaluated above; this would imply, in turn, the presence of an additional source of very low charge background. Putting together the above arguments we may estimate a maximal nucleon contamination increasing with time from some 10% to some 30% when moving from early to late times. However, lacking a reliable quantitative estimate, we prefer not to subtract any nucleon contamination and quote a final result including a possible nucleon contamination within the above limits.

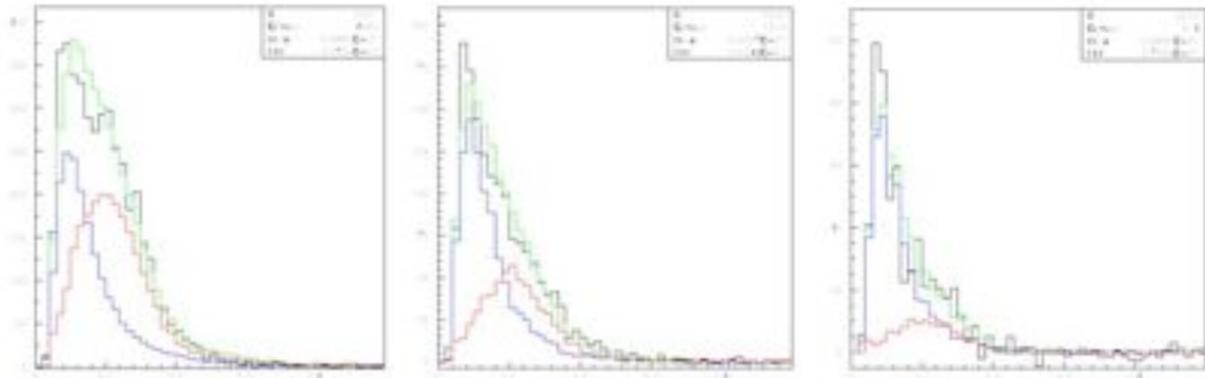


Figure 4.32. Charge (q_{max}) distributions in three equal time intervals (from left to right: early, middle and late respectively) as in Figure 4.21. Black: real data; red: simulated decay electrons; blue: simulated nucleons; green: sum of simulated signals. The blue and red were scaled in such a way that the simulated total have the same number of events as the real data in the two charge intervals [0 VEM, 0.08 VEM] and [0.08 VEM, 0.25 VEM].

4.4 Counting stopping muons

4.4.1 Method

The principle of the method can be summarized as follows:

MDE candidates are now defined as described in the preceding sections and as having in addition $q_{max} > 0.08 \text{ VEM}$. It has been shown that they can be detected with good efficiency ($55 \pm 3 \%$) and, at least for not too large values of θ and D , with low background contamination (accidentals evaluated from the early region and main signal tail contamination evaluated using the simulation).

Assuming that these MDE candidates are all coming from stopped muon decays, one still does not know at which time t the parent muon stopped in water. An evaluation of

the time elapsed between t (when the muon stops in water) and the muon decay time (when the electron candidate is detected) is however necessary to evaluate how many muons have stopped. This was done in Section 4.2 (Figure 4.9) using a probability distribution F evaluated from simulated data.

Taking the muon lifetimes (of positive and negative muons) as known quantities and assuming that the positive and negative muons are produced in equal quantity with the same time distribution, the probability for a given electron candidate to be observed somewhere in the late region can then be calculated. Weighting each candidate by the reciprocal of this probability should give the number of stopping muons.

The probability for a stopping muon (μ_+ or μ_-) to decay between time t and time $t+dt$ is $dP=(dt/\tau)\int F(t')\exp(-(t-t')/\tau)dt'$, integrated from $t'=0$ to $t'=t$. Here, τ is the relevant muon life time.

Writing $W(t) = \int F(t')\exp(t'/\tau)dt'$, where the integral runs from $t'=0$ to $t'=t$

$$dP=(dt/\tau)\exp(-t/\tau)W(t)$$

As $F(t)=0$ beyond 200 time bins and $i^*= \text{Max}(t_i+100, 200)>200$, $W(t)$ is a constant: $W(t)=W(200) = \int F(t')\exp(t'/\tau)dt'$, where t' runs from 0 to 200, for t over the late region. Figure 4.33 shows the dependence on $\cos\theta$ of $W(200)$ for τ having the value of the effective lifetime, 2.04 μs (see Section 4.1). A good fit is obtained using a quadratic form in D and in $1-\cos\theta$. The precise forms for different values of τ 's read:

$$W_+(200)=(0.43+0.22 \cdot 10^{-2}D-0.73 \cdot 10^{-6}D^2)(1-1.04(1-\cos\theta)^2) \text{ for } \tau_+ = 2.20 \mu\text{s},$$

$$W_-(200)=(0.12+0.31 \cdot 10^{-2}D-0.99 \cdot 10^{-6}D^2)(1-1.21(1-\cos\theta)^2) \text{ for } \tau_- = 1.80 \mu\text{s}.$$

The uncertainty associated with the evaluation of $W(200)$ is evaluated by remarking that a time shift Δt , small in comparison with τ , implies a multiplication by $\exp(\Delta t/\tau) \sim 1+\Delta t/\tau$. The scale of Δt is given by the width of the $F(t)$ distribution. Taking $\Delta t=Rms(F(t))/3$ as a reasonable estimate implies a relative uncertainty on W of $\Delta W/W=Rms(F(t))/(3\tau)$. The average relative uncertainties on W_+ and W_- are 11% and 14% respectively. A common value of 13 % is retained.

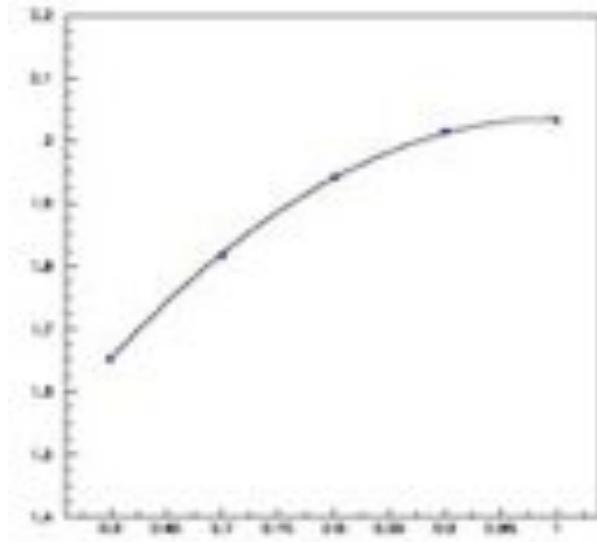


Figure 4.33 Dependence on $\cos\theta$ of $W(200) = \int F(t')\exp(t'/\tau)dt'$, where t' runs from 0 to 200 and $\tau=2.04 \mu\text{s}$. The line is the result of a quadratic fit.

If all muons decay with the same lifetime τ (as in the simulation where $\tau = \tau_+ = 2.20 \mu\text{s}$), the probability for a given electron candidate to be observed in the late region $[t_{start}, t_{end}]$ is:

$$P_v(t_{arr}) = \frac{1}{\tau} \int_{t_{start}}^{t_{end}} e^{-(t-t_{arr})/\tau} dt = \frac{e^{t_{arr}/\tau}}{\tau} \int_{t_{start}}^{t_{end}} e^{-t/\tau} dt = e^{t_{arr}/\tau} \left[e^{-t_{start}/\tau} - e^{-t_{end}/\tau} \right]$$

$$P_v(t_{arr}) = e^{t_{arr}/\tau} K(t_{start}, t_{end}, \tau)$$

Then the number of electrons observed in the late region is:

$$N_v = N_p \int P_v(t_{arr}) F(t_{arr}) dt_{arr},$$

where N_p is the number of stopping muons that give an electron within the charge cuts.

It can be written as:

$$N_v = N_p \int e^{t_{arr}/\tau} K(t_{start}, t_{end}, \tau) F(t_{arr}) dt_{arr}$$

$$N_v = N_p K(t_{start}, t_{end}, \tau) \int e^{t_{arr}/\tau} F(t_{arr}) dt_{arr}$$

$$N_v = N_p K(t_{start}, t_{end}, \tau) W(t_{arr})$$

Note that K is calculated for each trace and $F(t_{arr})$, therefore $W(t)$, depends on θ and D .

The number of stopping muons per tank, N_s , is obtained by weighing each electron candidate with a weight $1/(GK(t_{start}, t_{end}, \tau)W(200))$ where G is the value of the cut efficiency, 55 ± 3 %.

The above evaluation is now extended to the case where μ^+ and μ^- have different lifetimes. With the assumption that μ^+ and μ^- are produced in equal numbers, the probability for a given electron candidate to be observed in the late region becomes $P_v = 0.5(K_+W_+ + K_-W_-)$, where $K_{+/-}$, $W_{+/-}$ are calculated using the forms defined above with τ taking value of $\tau_{+/-}$. The number of stopping muons per tank is then $N_d = \Sigma 2/G(K_+W_+ + K_-W_-)$ summing over the number of observable electrons.

4.4.2 Validation of the method

In order to validate the method outlined above, we apply it to simulated data (single lifetime for muons, no nucleons). The data sample used here is, as before, traces having $Q_{tot} < 40$ VEM and $D < 700(m) \log_{10} E(eV) - 11150(m)$ from vertical $10^{18.5}$ eV proton showers. The number of stopping muons per trace is obtained after subtracting the contributions of accidental and tail backgrounds. The contribution of accidentals is evaluated from the early region as explained earlier. The contribution of the tail background is evaluated from simulated MDE candidates not associated with an electron as displayed in Figure 4.26. The error on this number is calculated by combining three components in quadrature: the statistical uncertainty on the raw data (before subtraction), the statistical uncertainty on the accidental background and one third of the tail background. The result is given in the table below for three D intervals. For D in excess of 1 km, a good agreement with the direct evaluation (N_{s0}) of the simulation is obtained. At shorter distances, however, the number of muons is overestimated by 4%.

Number of stopping muons per tank in the simulation

	N_{s0}	N_s
$D < 1050$ m	2.17	2.26 ± 0.02
$1050m < D < 1300m$	1.07	1.05 ± 0.02
$1300m < D < 1800m$	0.64	0.66 ± 0.02

The global relative systematic uncertainties, estimated earlier as $3/55=5.5\%$ associated with the cut efficiency and as 13 % associated with the knowledge of $F(t)$, should not apply here where simulated data are being used. Only after having compared real data with the predictions of the simulation will one be able to evaluate more precisely which global uncertainties should be retained.

4.5 Results

The method described in the previous section (accounting for different decay lifetimes of μ^+ and μ^-) is now applied to real data. The energy range of the data sample is restricted to 2 to $4 \cdot 10^{18}$ eV and the S_{1000} window is $12.3-46.0(1-\cos\theta)^2 < S_{1000} < 18.3-55.2(1-\cos\theta)^2$. The number of stopping muons are evaluated in different intervals of zenith angle θ and distance D to the shower axis over the whole range ($\theta < 60^\circ$ and $D < 700 \log_{10}E(\text{eV}) - 11150$). The data are split in nine different samples corresponding to three different θ intervals and three different D intervals.

The parameterizations of $W(200)$ as a function of θ and D read:

$$W_+(200) = (0.43 + 0.22 \cdot 10^{-2}D - 0.73 \cdot 10^{-6}D^2)(1 - 1.04(1 - \cos\theta)^2) \text{ for } \tau_+ = 2.20 \mu\text{s}$$

$$W_-(200) = (0.12 + 0.31 \cdot 10^{-2}D - 0.99 \cdot 10^{-6}D^2)(1 - 1.21(1 - \cos\theta)^2) \text{ for } \tau_- = 1.80 \mu\text{s}$$

The contributions of accidental and shower tail backgrounds are subtracted but the nucleon contamination is not. The main signal tail background is subtracted using the parameterization:

$$f_{tail}(\%) = (D - 595)(0.025 + 0.105(1 - \cos\theta)).$$

For each of the nine samples the table below lists N_d , the measured number of stopping muons per trace from real data, and N_s , N_{s0} , the measured value and direct evaluation from simulated data ($10^{18.5}$ eV proton showers at $\theta = 0^\circ, 25^\circ$ and 45°). The first error on N_d is statistical, the second (5%) is systematic and includes a 3% contribution accounting for errors inherent to the method and estimated from a comparison of N_s and N_{s0} , and a 4% contribution estimated as the point to point uncertainty on $F(t)$. This leaves a global uncertainty of 14% ($\sqrt{(13^2 - 4^2)} = 12.4\%$ for $F(t)$ and of 5.5% for the cut efficiency) that applies to all data together.

Number of stopping muons per tank for different intervals of θ and D . The quoted uncertainties do not include a global 14% uncertainty that applies to all points.

D interval (m)	$\theta(^{\circ})$	$\theta(^{\circ})$ sim	N_d	N_s	N_{s0}
<1050	[0, 20]	0	$1.90 \pm 0.01 \pm 0.10$	2.26 ± 0.02	2.17
	[20, 31]	25	$1.86 \pm 0.01 \pm 0.09$	1.85 ± 0.02	1.77
	[41, 49]	45	$1.05 \pm 0.01 \pm 0.05$	0.87 ± 0.02	0.86
[1050, 1300]	[0, 20]	0	$1.07 \pm 0.01 \pm 0.05$	1.05 ± 0.02	1.07
	[20, 31]	25	$0.83 \pm 0.01 \pm 0.04$	0.91 ± 0.02	0.89
	[41, 49]	45	$0.49 \pm 0.01 \pm 0.03$	0.44 ± 0.02	0.48
[1300, 1800]	[0, 20]	0	$0.50 \pm 0.01 \pm 0.03$	0.66 ± 0.02	0.64
	[20, 31]	25	$0.51 \pm 0.02 \pm 0.03$	0.57 ± 0.03	0.56
	[41, 49]	45	$0.25 \pm 0.02 \pm 0.01$	0.36 ± 0.03	0.36

As expected, the measurement method gives satisfactory results when applied to simulated data. The estimations from real data are compatible with the simulation to within 20%. The ratio between real and simulated data has no obvious dependence of the distance to the shower axis nor on zenith angle. Accounting for a possible nucleon associated contamination as discussed above, high values of this ratio (1.3 or more) are excluded for the whole D and θ range explored here.

4.6 Summary and conclusions

A search for electrons having their origin in the decay of shower muons stopping in the Cherenkov tanks of the PAO surface detector has been presented. In order to deal with small isolated signals in clean conditions while using large enough statistical samples, the analysis has been restricted to showers having energies around $10^{18.5}$ eV and zenith angles not exceeding 60° and to tanks located at moderate distances from the shower axis (typically 750 to 1800 m). While muons are generally considered as good discriminant between light and heavy primaries, stopping muons are far less efficient than high energy muons: the latter are generally produced at higher altitudes, the former being more likely to decay before reaching ground. The interest of such a search is to provide a sensitive test of detector and simulation performance in a regime of low intensity signals (the average

charge deposited by an MDE is 0.14 VEM). In addition, it excludes models that predict an abundance of stopping muons that differs too much from what has been observed.

Attention has been given to identify as clearly as possible those aspects of the search that are dependent from the shower simulation and those which are not. The search is performed over a broad interval of time, around 8 μ s, over which the acceptance is uniform. The electron signal has been found to be well described by the detector simulation down to charges of the order of 0.08 VEM. The main backgrounds of relevance are from the tail of the main signal and from accidentals. The former is found to decrease with time faster than the muon decay rate but its contribution needs to rely on the simulation. The latter is of course uniform in time and can be accurately evaluated from early times preceding the arrival of the shower front. Data display in addition a background of very low charges (below 0.1 VEM) that are not properly described by the simulation and cannot all be interpreted as random coincidences between fluctuations in different PMTs. A possible source is from nucleons produced in the showers, but no reliable prediction exists.

Selection criteria have been defined, including charge and isolation cuts, in order to suppress background. Their global efficiency is $55\pm 3\%$. When evaluating the fraction of stopping muons per ADC trace, however, an important source of uncertainty is the lack of knowledge of the time distribution of the arrival of stopping muons in the tank. Its contribution must fully rely on the validity of the shower development simulation. An estimate of 13% for the associated uncertainty has been considered reasonable and arguments have been presented in favour of its reliability.

The final evaluations are given with typical uncertainties of 5% point to point and 14% global. Neglecting a possible contamination by nucleons that might reach up to 20% to 30% in the selected range of charges, they are compatible within 20% with the predictions of a simulation based on AIRES with QGSJET01. There are other simulations quoted in Section 4.1, whether of proton or iron initiated showers, that give also acceptable predictions. If the nucleon contamination were significantly larger than quoted above the results would imply a deficit of low energy muons with respect to prediction.

The results underline the need for a reliable simulation of the contribution of ground nucleons to the Cherenkov signals in the PAO surface detector.

5. Summary and perspectives

The studies that have been presented here are a significant step toward a better and deeper understanding of the surface detector of the Pierre Auger Observatory, which is today the world leading observatory for the study of ultra high energy cosmic rays. At the same time, they have opened new roads on which to pursue the exploration of the properties of the extensive air showers induced by the interaction of UHECRs with the Earth atmosphere.

Their direct contributions are related with the properties of the FADC traces, with the early time PMT asymmetry and with the decay of stopping muons.

Concerning the FADC traces, the uncertainties affecting the charge content of FADC time bins have been evaluated and evidence has been given that the occasional apparent inconsistencies between the three traces of a same tank reduced to only two types, after pulses and early time asymmetries, both of which were previously known to occur and under control. The peak finding algorithm consisting in unfolding the exponential decay of the collected light from the FADC traces has been refined and its performance has been assessed; at the same time, its limitations, resulting essentially from the simultaneous occurrence of overlapping signals, have been identified. This algorithm is the most powerful pattern recognition tool available and the present study opens the road toward its systematic use in studies concerned with the muon fraction and the primary mass composition.

The early time PMT asymmetry has been shown to be strongly correlated with the azimuth of the shower axis, a correlation that has been exploited to evaluate the shower divergence under the single vertex approximation. The study has shown the power of the method and illustrated its sensitivity. When used together with the peak finding algorithm previously mentioned, it allows for measurements of the shower divergence associated with selected peaks, opening the road toward new systematic studies that could be again of relevance to the primary mass composition.

Finally, a search for muons stopping in the water volume of the Cherenkov tanks, identified by the signal produced by the decay electron or positron, has been presented. The complexity of the problem, resulting in particular from the small amplitude of the expected signals, has given an opportunity to probe the detector performance in particularly difficult conditions and to test the current understanding of its response and of

the tools available for such analyses, in particular the simulation codes used to describe the shower development. While the bulk of the results is consistent with expectation and provides a good check, evidence has been found for a very low charge background that is not properly described by the simulations. It has been suggested that it might be associated with neutrons, a possibility that remains to be explored.

References

- [1] Th. Wulf, *Phys. Zeitschr.* 10 (1909) 152.
- [2] V.F. Hess, *Z. Phys.* 13 (1912) 1084.
- [3] J. Clay, *Proc. Roy. Acad, Amsterdam* 30 (1927) 1115.
- [4] P. Auger, P. Ehrenfest, R. Maze, J. Daudin and A.F. Robley, *Rev. Mod. Phys.* 11 (1939) 288;
P. Auger, R. Maze and T. Grivet-Mayer, *Académie des sciences, Paris*, 206 (1938) 1721
- [5] C.D. Anderson, *Science* 76 238 (1932); *Phys. Rev.* 43 (1933) 491.
- [6] S.H. Neddermeyer and C.D. Anderson, *Phys. Rev.* 51 (1937) 884;
J.C. Street and E.C. Stevenson, *Phys. Rev.* 52 (1937) 1003.
- [7] C.M.G. Lattes, H. Muirhead, G.P.S. Occhialini, and C.F. Powell, *Nature* 159 (1947) 694;
C.M.G. Lattes, G.P.S. Occhialini, and C.F. Powell, *Nature* 160 (1947) 486; *Nature* 160 (1947) 453.
- [8] P. Bhattacharjee and G. Sigl, *Phys. Rept.* 327 (2000) 109–247.
- [9] J. Linsley, L. Scarsi, and B. Rossi, *Phys. Rev. Lett.* 6 (1961) 485;
J. Linsley, *Phys. Rev.* 97 (1955) 1292.
- [10] J.W. Cronin, *Rev. Mod. Phys.* 71 (1999) S165;
M. Nagano and A.A. Watson, *Rev. Mod. Phys.* 72 (2000) 689.
- [11] R. Cornils et al., *APh* 20 (2003) 129;
K. Bernlohr et al., *APh* 20 (2003) 111.
- [12] M. C. Weisskopf, H. D. Tananbaum, L. P. Van Speybroeck and S. L. O'Dell, *Proc. SPIE* 4012, 2000, arXiv:astro-ph/0004127v1.
- [13] F.A. Aharonian et al., *Nature* 432 (2004) 75;
F.A. Aharonian et al. (H.E.S.S. Col laboration) *A&A* 449 (2006) 223.
- [14] E. Fermi, *Phys. Rev.* 75, 1169 (1949)
- [15] L.O'C. Drury, *Rep. Prog. Phys.* 46 (1983) 973;
R.D. Blandford and D. Eichler, *Phys. Rept.* 154 (1987) 1;
F.C. Jones and D.C. Ellison, *Sp. Sc. Rev.* 58 (1991) 259;
L.O'C. Drury, *Contemp. Phys.*, 35 (1994) 232.
- [16] Y. Uchiyama et al., *Nature* 449 (2007) 576.
- [17] E.G. Berezhko and H.J. Völk, *A&A* 419 (2004) L27.
- [18] K. Greisen, *Proc. 9th ICRC*, 2 (1965) 609.
- [19] A.M. Hillas, *Ann. Rev. Astron. Astrophys.*, 22 (1984) 425.
- [20] R.A. Dupke, N. Mirabal, J.N. Bregman and A.E. Evrard, *ApJ* 668 (2007) 781;

- N.T. Huong, *Galaxy collisions as possible sources of ultrahigh energy cosmic rays*, dissertation at the Hanoi University of Education, 2008.
- [21] The Pierre Auger Project Design Report, The Auger Collaboration, 2nd Edition, Fermi Laboratory, November 1996, Revised March 1997.
<http://www.auger.org/admin/DesignReport/index.html>
- [22] J. Abraham et al., Nucl. Instrum. Meth. A523 (2004) 50.
- [23] R. Engel, presented at 40th Rencontres de Moriond, Italy (2005).
- [24] A.M. Hillas, Proc. 17th ICRC, 8 (1981) 193.
- [25] B. Dawson [Pierre Auger Collaboration], Proc. 30th ICRC, Merida, (2007) #0976, arXiv:astro-ph/0706.1105.
- [26] D. Bird et al., Nucl. Inst. Meth., A349 (1994) 592;
D. Bird et al., Astrophys. J. 424 (1994) 491.
- [27] T.K. Gaisser and A.M. Hillas, Proc. 15th ICRC, 8 (1977) 353.
- [28] J. Linsley and L. Scarsi, Phys. Rev. 128 (1962) 2384;
J. Linsley, L. Scarsi and B. Rossi, Phys. Rev. Lett. 6 (1961) 458.
- [29] M. Ave [Pierre Auger Collaboration], Proc. 30th ICRC, 4 (2007) 307.
- [30] A.K. Tripathi, Nucl. Instr. and Meth. A 504 (2003) 1;
P. Ranin et al., UCLA-Cosmic/2001-02;
A. Tripathi et al., UCLA-Cosmic/2001-02 and GAP 2002-013.
- [31] S. Agostinelli et al., Nucl. Instrum. Meth. A506 (2003) 250;
J. Allison et al., IEEE Trans. Nucl. Sci. 53 (2006) 270.
- [32] C. Pryke, GAP 1996-008;
C. Pryke, Auger internal GAP 1997-004;
T. McCauley and T. Paul, GAP 2000-055;
W.E. Slater, A. Tripathi and K. Arisaka, GAP 2002-063;
T. Ohnuki, G. Rodriguez-Fernandez, D. Barnhill, A. Tripathi, T. McCauley and T.K. Arisaka, GAP-2004-043;
M. Zha and J. Knapp, GAP 2005-031;
A. Creusot and D. Veberic, GAP 2007-073.
- [33] N. Kalmykov, S. Ostapchenko and A. Pavlov, Nucl. Phys. Proc. Suppl. B52 (1997)17;
S. Ostapchenko, Phys. Rev. D74 (2006) 014026;
S. Ostapchenko, Nucl. Phys. Proc. Suppl. 151 (2006) 143;
S. Ostapchenko, Phys. Lett. B636 (2006) 40;
G. Battistoni et al., AIP Conf. Proc. 896 (2007) 31.
- [34] R. Engel, T. Gaisser, T. Stanev, and P. Lipari, Proc. 26th ICRC, 1 (1999) 415;
R. Fletcher, T. Gaisser, P. Lipari, and T. Stanev, Phys. Rev. D50 (1994) 5710;

- K. Werner and T. Pierog, AIP Conf. Proc. 928, 111 (2007), arXiv:0707.3330.
- [35] D. Heck, J. Knapp, J. Capdevielle, G. Schatz, and T. Thouw, FZKA-6019 (1998).
- [36] S. Sciutto, AIRES. <http://www.fisica.unlp.edu.ar/auger/aires>
- [37] K. Greisen, Phys. Rev. Lett. 16 (1966) 748;
G. Zatsepin and V. Kuzmin, JETP Lett. 4 (1966) 78.
- [38] F. Shussler [Pierre Auger Collaboration], Proc. 31th ICRC, Lodz, Poland, 2009.
- [39] R. U. Abbasi et al. Phys. Rev. Lett., 100 (2008) 101101;
R. U. Abbasi et al. Phys. Lett., B619 (2005) 271.
- [40] J. Abraham et al. [Pierre Auger Collaboration], Astropart. Phys. 29 (2008) 188;
J. Abraham et al. [Pierre Auger Collaboration], Science 318. no. 5852 (2007) 938.
- [41] J.D. Hague, [Pierre Auger Collaboration], Proc. 31th ICRC, Lodz, Poland, 2009.
- [42] M.P. Veron-Cetty and P. Veron, Astron. Astrophys. 455 (2006) 773.
- [43] D. De Marco and T. Stanev, Phys. Rev. D72 (2005) 081301;
T. Wibig and A. W. Wolfendale, J. Phys. G31 (2005) 255.
- [44] D. Allard, E. Parizot, and A. Olinto, Astropart. Phys. 27 (2007) 61, astro-ph/0512345.
- [45] V. Berezhinsky, S. Grigoreva, and B. Hnatyk, Nucl. Phys. Proc. Suppl. 151 (2006) 497.
- [46] R.J. Glauber and G. Matthiae, Nucl. Phys. B21 (1970) 135;
K.G. Boreskov and A.B. Kaidalov, Sov. J. Nucl. Phys. 48 (1988) 367.
- [47] T.K. Gaisser et al., Proc. 16th ICRC 9 (1979) 258;
T.K. Gaisser et al., Phys. Rev. D47 (1993) 1919;
J. Linsley and A.A. Watson, Phys. Rev. Lett. 46 (1981) 459;
J. Matthews, Astropart. Phys. 22 (2005) 387;
W. Heitler, "The Quantum Theory of Radiation", Oxford University Press, 1954.
- [48] C. Bonifazi et al. [Pierre Auger Collaboration], Nucl. Phys. Proc. Suppl. 190 (2009) 20, arXiv:0901.3138.
- [49] M. Unger et al., Nucl. Instrum. Meth. A 588 (2008) 433.
- [50] M. Unger et al. [Pierre Auger Collaboration], in preparation (2009).
- [51] T. Bergmann et al., Astropart. Phys. 26 (2007) 420.
- [52] R. Walker & A.A. Watson, J. Phys. G7 (1981) 1279.
- [53] D. Newton et al., Astropart. Phys. 26 (2007) 414.
- [54] H. Wahlberg [Pierre Auger Collaboration], Proc. 31th ICRC, Lodz, Poland, 2009.
- [55] X. Bertou and P. Billoir, GAP 2000-017;
X. Bertou, P.D. Silva and P. Billoir, GAP 2002-074;
- [56] M. Healy [Pierre Auger Collaboration], Proc. 30th ICRC, 4 (2007) 377.
- [57] F. Schmidt et al., Astropart. Phys. 29 (2008) 355.

- [58] A. Castellina [Pierre Auger Collaboration], Proc. 31th ICRC, Lodz, Poland, 2009.
- [59] P.N. Diep, VATLY internal note nr 33, "*More on the jump method*", August, 2009.
- [60] B. Genolini et al., GAP 2001-021 and 2003-051;
G. Navarra, GAP 2005-006.
- [61] P.S. Allison and D. Barnhill, GAP 2004-046.
- [62] A. Creusot *et al.*, GAP 2002-042.
- [63] M. Aglietta, GAP 2005-010.
- [64] P.N. Dong and P.T. Nhung, VATLY internal note nr 31, *The PMTs of the PAO surface detector: New measurements*, August 2009 and GAP 2009-115.
- [65] Burle Photomultiplier Handbook (TP-136), Lancaster, PA (1989);
S. Atulugama and S.Coutu, GAP 2005-100 and references therein.
- [66] A. Lopez-Agüera, V. M. Olmos-Gilbaja and G. Rodriguez Fernandez, GAP 2003-112,
GAP 2004-026 and GAP 2004-061.
- [67] P. Billoir, GAP 2002-076 and GAP 2005-074.
- [68] P.T. Nhung, *Reducing FADC traces to a sum of peaks*, VATLY internal note 15,
August 2005.
- [69] B. Smith, C. Wileman and A. Watson, GAP 2007-092 and references therein.
- [70] M. Urban et al, GAP 2007-020;
- [71] P.N.Diep, GAP 2008-136;
P.N. Diep, *More on the jump method*, VATLY internal note nr 27, September 2008.
- [72] Aaron Chou, GAP 2004-057.
- [73] J. Cronin, GAP 2003-076.
- [74] J. Abraham et al., Nuc. Inst. Meth. A523 (2004) 50.
- [75] D. T. The, graduation thesis, 2007, Hanoi University of Education.
- [76] P.T. Nhung and P. Darriulat, *Spikes in the raw FADC traces*, VATLY internal notes
17, November 2006 and 17add., December 2006.
- [77] P.T. Nhung and P. Darriulat, *Using the asymmetry information to help the reduction
of FADC traces into peaks: a first look*, VATLY internal note 19, February 2007;
P.T. Nhung and P.Billoir, GAP 2007-131.
- [78] P.N. Dong and N.T. Thao, VATLY internal note nr 24, *A VATLY test bench for the
study of XP1805 photomultiplier tubes*, and note nr 24 add, 2008.
- [79] A. Tripathi *et al.*, GAP 2002-013.
- [80] M. Unger *et al.*, Proc. 30th ICRC, Merida, Mexico (2007).
- [81] P. Billoir, private communication.

- [82] X. Garrido *et al.*, GAP 2007-060 and 2009-023.
- [83] D.E. Groom *et al.*, Eur. Phys. J. C15 (2000) 1.
- [84] T. Suzuki, Phys Rev C35 (1987) 2212.
- [85] D. Newton et al., Astropart. Phys. 26 (2007) 414.
- [86] The event selection group, GAP 2005-023.
- [87] P. Billoir, GAP 2000-025.
- [88] P.T. Nhung, VATLY internal note nr 16 add. (2006), nr 22 (2008) and nr 26 (2009).
- [89] F.Arneodo et al., in preparation (2009);
F. Arneodo [Pierre Auger Collaboration], Proc. 29th ICRC 8 (2005) 299.

Acknowledgement

This thesis was made under joint supervision by Pr Pierre Billoir and Pr Pierre Darriulat, both of whom I express my deepest gratitude for their constant support and invaluable guidance. In particular, I am very grateful to Pr Pierre Billoir for having made my stays in Paris both efficient and enjoyable and for his kindness and patience in giving me suggestions, explanations and advice. I also express my deepest gratitude to Pr Pierre Darriulat, for his invaluable guidance and his enthusiasm that makes his students like science and motivates them to pursue research.

I thank my colleagues in the Pierre Auger Collaboration for their understanding and constant support, in particular the members of the Auger groups in LPNHE, IPN/Orsay and LAL.

I acknowledge the help and support of my professors in Hanoi University of Science, in particular Pr Nguyen Mau Chung, Pr Pham Quoc Hung, Pr Dao Tien Khoa and Pr Bui Duy Cam. Warm thanks are also expressed to my colleagues in the Institute for Nuclear Science and Technology for their help and encouragement.

I warmly thank the members of the VATLY group for their friendly help, discussion (fruitful or not) and kind friendship that makes the life in the lab so pleasant.

I also express my deepest gratitude to my family for their patience and moral support.

Finally, I acknowledge financial support from World Laboratory, French Ministère des Affaires Étrangères (bourse Évariste Galois), Rencontres du Vietnam (bourse Odon Vallet), French CNRS, Vietnam Atomic Energy Commission and Vietnam Ministry of Science and Technology.