University of California

Los Angeles

The Cosmic Ray Spectrum as Measured by the Surface Detector of the Pierre Auger Observatory and its Theoretical Implications

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Physics

by

Joong Yeol Lee

2007

The dissertation of Joong Yeol Lee is approved.

Abby Kavner

Vladimir Vassiliev

Katsushi Arisaka, Committee Chair

University of California, Los Angeles

2007

Table of Contents

List of Figures

List of Tables

Acknowledgments

First of all, I would like to thank all the present and former members of the UCLA Auger group for without their help and guidance this would not have been possible. I would like to thank Matt Healy, David Barnhill, and Tohru Ohnuki for their support and help with technical matters and otherwise for the past 4 years. I am deeply indebted to the tutelage and guidance of Arun Tripathi who more than anyone was responsible for pointing me in the right direction when I first started this journey as a clueless graduate student 4 years ago. I am also deeply indebted to Professor Katsushi Arisaka for his help, guidance, and support for the past 4 years. I have learned more than I could have imagined when I first started, in the past 4 under his tutelage. I would also like to thank Professor Bill Slater. And finally, I would like to thank our theorist colleagues in the group, Graciela Gelmini, Alexander Kusenko, Dmitri Semikoz, and Oleg kalashev for their help.

Beyond the core of the UCLA Auger group, I would like to thank all the people I have learned from here at UCLA. Professors Cousins, Hauser, Saltzberg, and Wallny are just some of the names that come to mind. Also, I would like to thank the entire Auger Collaboration for I have benefited both directly and indirectly from all the hard work done by the members of the collaboration.

It should be note that several of the chapters of in this dissertation contain material from previously written Auger internal notes (GAP Notes).

Chapter 6 (Energy Determination and Systematic Uncertainties) is a GAP Note to submitted co-written by J. Lee and M. Healy.

Chapter 7 (Spectrum) is an updated version of GAP-2006-039, co-written by J. Lee, K. Arisaka, D. Barnhill, P. Boghrat, M. Healy, and A. Tripathi, titled

"The Cosmic Ray Spectrum with 2 Years of Data."

Chapter 8 (Theoretical Implications of the Auger Spectrum) is an updated version of GAP-2007-012. co-written by J. Lee, O. Kalashev, K. Arisaka, D. Barnhill, P. Boghrat, G. Gelmini, and M. Healy, titled " The Latest Auger Spectrum and Theoretical Implications."

My research was supported by a grant from the U.S. Department of Energy.

Vita

1976	Born, Paju, Gyeonggi-do, Republic of Korea
1998	B.A. in Physics, University of California at Berkeley
1999-2001	Electronics Engineer, Ernest Orlando Lawrence Berkeley Laboratory
2001–2003	Teaching Assistant, UCLA
2003–2007	Graduate Student Researcher, UCLA.

Publications

J. Lee, An Estimate of the Cosmic Ray Spectrum from the Pierre Auger Observatory in *2006 APS April Meeting* , Dallas, Texas, April 22–25, 2006

Abstract of the Dissertation

The Cosmic Ray Spectrum as Measured by the Surface Detector of the Pierre Auger Observatory and its Theoretical Implications

by

Joong Yeol Lee

Doctor of Philosophy in Physics

University of California, Los Angeles, 2007

Professor Katsushi Arisaka, Chair

Even some 40 years after the discovery of a 10^{20} eV cosmic ray by Linsley *et al.*, the origin of these ultra high energy cosmic rays still mystifies us. The Pierre Auger Observatory, the largest cosmic ray experiment ever consisting of the surface detector and the fluorescence detector, was designed to unlock the mystery.

Two previous experiments, HiRes and AGASA, reported very different spectra, both in the absolute flux and the presence (or absence) of the so-called GZK feature. The GZK feature, theorized by Greisen, Zatsepin, and Kuzmin in 1966, is the sharp reduction and departure of the flux of cosmic rays, if they are protons, above $\sim 5 \cdot 10^{19}$ eV from that expected from the simple continuation of the spectrum from below $\sim 5 \cdot 10^{19}$ eV. This arises as a result of the interaction of protons with the cosmic microwave background radiation. Iron and other nuclei also undergo interactions with the cosmic microwave background and other radiation which result in a reduction of the flux similar to the GZK feature.

I present the cosmic ray energy spectrum above $\sim 3 \cdot 10^{18}$ eV as is measured

by the surface detector using the data from January 1, 2004 to February 28, 2007. I show that by using either a Monte Carlos simulation based energy calibration method or the hybrid energy calibration method, where events seen by both the surface detector and the fluorescence detector are used to calibrate the surface detector with the fluorescence detector, it is possible to reproduce both the HiRes and AGASA spectrum. Regardless of the energy calibration method used, however, the spectrum shows a reduction of the flux above $\sim 5 \cdot 10^{19}$ eV, in disagreement with the AGASA spectrum which engendered excitement in the theory circle with its lack of the GZK feature and the exotic scenarios for the top down models as possible explanation for it.

I compare the spectrum with the theoretical model by Oleg Kalashev with a homogenous source distribution and pure proton and iron simple power law injections at the sources, using the binned maximum liklihood method. The pure proton injection case yield a theoretical spectrum that exhibit the GZK feature with a reduction in the flux beyond $\sim 5 \cdot 10^{19}$ eV, and the iron injection models also exhibit reduction in the flux beyond $\sim 5 \cdot 10^{19}$ eV. I show that the spectrum cannot distinguish between the proton and iron injection cases as the spectrum agrees well with both cases. This also means, however, that the spectrum can be explained with the conventional bottom up acceleration and physics interaction processes without having to invoke new exotic physics.

CHAPTER 1

Introduction

The earth is constantly bombarded by particles commonly known as cosmic rays. Some cosmic rays have been observed with energies as high as 10^{20} eV. The discipline of cosmic ray physics itself has a beginning tracing back to Victor Hess's discovery of the increased rate of ionization with altitude in 1912, leading him to conclude that the radiation was of extraterrestrial origin [1].

Although modern cosmic ray physics is more allied with astrophysics, dealing with the origin and composition of cosmic rays, early cosmic ray physics was more closely allied with elementary particle physics. Before the advent and the subsequent rise of the particle accelerators, most of the early elementary discoveries were made with cosmic rays. The earth is constantly bombarded with cosmic rays with energies many orders of magnitude beyond those attainable on earth. This gives rise to many secondary particles of high energies in the atmosphere creating conditions favorable to creating meta- and unstable particles (See Ch. 2). And, it is these high energy particles that played the role of modern accelerator beams in the discovery of new particles. Starting with the discovery of positrons by Anderson [2] in 1931, many particles, such as the muon (μ) and the pion (π), were discovered with cosmic rays.

The advent of the high energy particle accelerators meant the end of an era of the particle physics discovery through cosmic rays. Modern cosmic ray physics has centered around identifying the source and the composition of cosmic rays.

Most cosmic rays that we encounter on earth are secondary, tertiary, and higher order particles from the "extensive air showers" (see Ch. 2) resulting from interaction of the cosmic ray primaries and the atmosphere. Pierre Auger, the namesake of the Pierre Auger Observatory in Argentina, can be credited with the discovery of the extensive air shower. He saw many multiple coincidences from Geiger counters placed on the ground in 1938. He deduced from the electromagnetic cascade theory that the these were from showers triggered by a particle of 10^{15} eV [3]. Because of the low flux of cosmic rays at high energies, an indirect detection method that detects the extensive air shower, the same kind of method employed by Pierre Auger, must be used.

In 1963, Linsley at Volcano Ranch made the first discovery of a 10^{20} eV cosmic ray [5]. Considering that the highest energy attainable at particle accelerators here on earth is only $\sim 10^{12}$ eV, the source and the acceleration mechanism is obviously of utmost interest for physicists. The source and the acceleration mechanism of cosmic rays with energies that high are, however, still a mystery even to this day. We will study cosmic rays in the $10^{18} - 10^{20}$ eV energy range and contribute to unlocking the mystery.

The flux of cosmic rays as seen on earth is shown in Fig. 1.1. The cosmic ray spectrum falls off steeply roughly following an E^{-3} power law. It covers many orders of magnitude in energy as high as 10^{20} eV. In the energy range that we are interested in, $10^{18} - 10^{20}$ eV, the flux is very small. It goes from 1 particle per km$^2\times$year at 10^{18} eV to 1 particle per km$^2\times$century at 10^{20} eV. To overcome this extremely low flux, it is necessary to cover a large area. The Pierre Auger Observatory (PAO) was designed with this in mind as it has the specific aim of studying Ultra High Energy Cosmic Rays (UHECR) above 10^{18} eV. When completed, the PAO will cover 3000 km^2 which is much larger than any previous

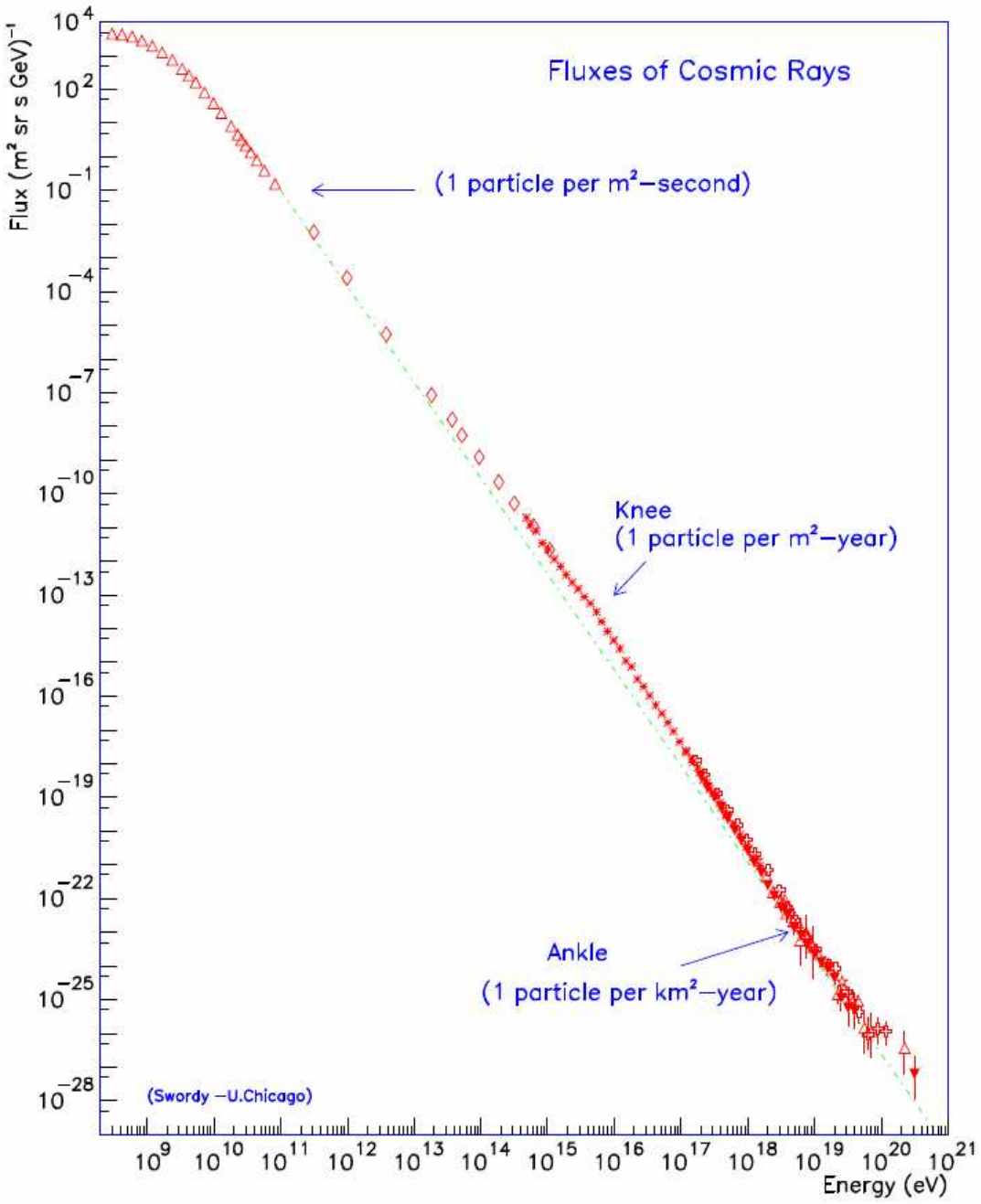


Figure 1.1: Differential energy spectrum of cosmic rays. Dotted line is E^{-3}. Figure from [4].

experiment and will run for 20 years. Hence, it will provide us with unprecedented statistics and help us tackle the mystery that is UHECR.

CHAPTER 2

Cosmic Ray Physics

When studying cosmic rays, one naturally wonders what they are, where they come from, and how they attain their energies as cosmic rays can have energies that are well beyond attainable in accelerators (see Fig. 1.1). Or in cosmic ray physics terms, what are the sources and the chemical composition of cosmic rays? As we will see shortly, these questions are intertwined.

2.1 Chemical Composition

What exactly are these cosmic rays? It turns out that they are just composed of different chemical elements. The comparison of the relative abundances of the chemical elements in cosmic rays and in the solar system are shown in Fig. 2.1. Although they do not exactly match, the overall relative abundance of the cosmic rays are similar to the relative abundance of elements found in the solar system. There are key differences, however. Cosmic rays are richer in Li, Be, and Bo as well as elements just lighter than Fe. On the other hand, cosmic rays have a relative deficiency in proton and He. These differences can be understood by assuming that cosmic rays start out with the same composition as solar matter, and as they propagate they interact with other particles and spallate into lighter nuclei [6].

At TeV-PeV range, cosmic rays consists of 50 % protons, 25 % α particles,

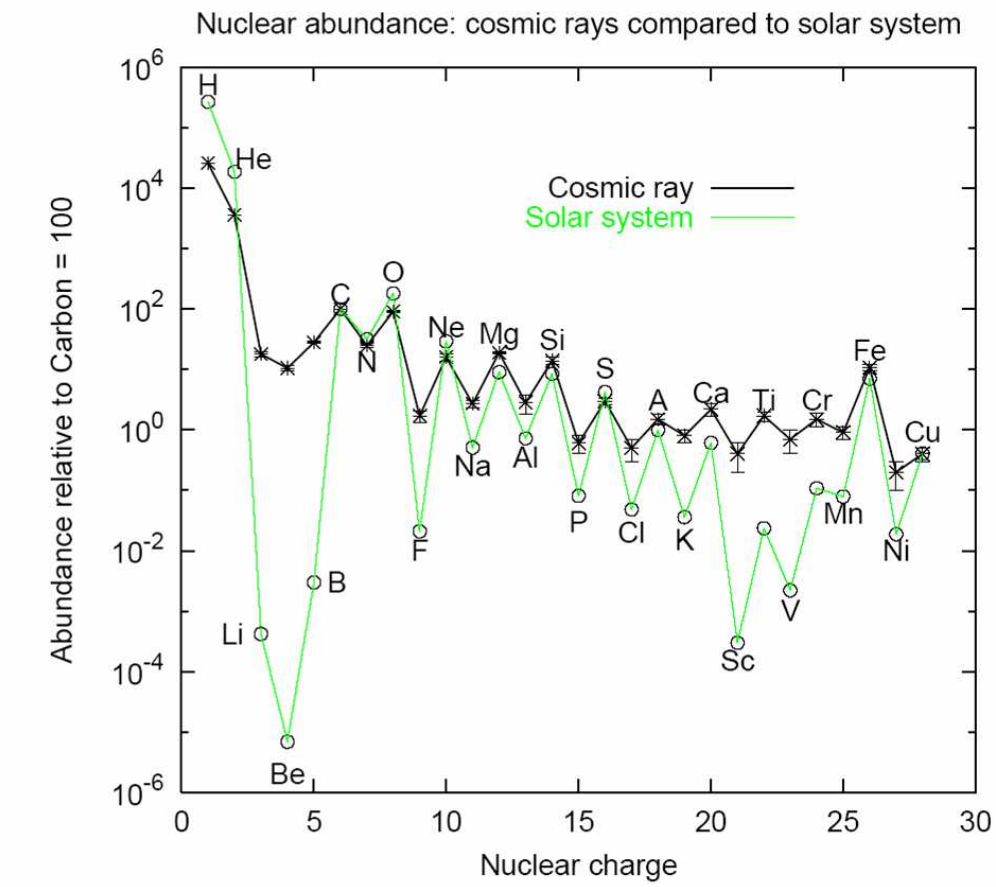


Figure 2.1: Composition of cosmic rays (open circles) and solar system abundances (asterisks). Figure from [7].

13 % CNO, and 13 % Fe [8]. At higher energies, the flux of cosmic rays is too sparse for direct measurement, so it is necessary to rely on indirect measurements. One way to measure the composition of cosmic rays above 10^{18} eV is to measure the X_{max}, or the maximum of the shower depth, of the extensive air shower (see the section on extensive air shower). X_{max} has a dependence on the composition. Fig. 2.2 shows the composition of cosmic rays above 10^{17} eV as given by the X_{max} measurement technique [9]. Although the composition above 10^{19} eV is proton-like in Fig. 2.2, there is still no definitive proof of the composition of cosmic rays above 10^{18} eV. It is widely accepted, however, that cosmic rays above 10^{18} eV ordinary matter between proton and Fe. Due to the lack of statistics and uncertainties in hadronic models, it will be a long time before we have an exact measurement of the composition as in Fig. 2.1 for low energies.

Besides the embarrassment of not knowing what they are, the lack of knowledge of the composition is problematic in many respects. The effect of the galactic and intergalactic magnetic fields depends on the charge of the particle, so the lack of knowledge of the composition means large uncertainties in the bending angle and the path length of the cosmic rays. And as we will see later, the lack of knowledge of composition contributes to the uncertainties in energy determination.

2.2 Acceleration Mechanisms and Sources

The general consensus regarding the source of cosmic rays is that the source of cosmic rays are astronomical objects where they are created and accelerated. These are the so-called bottom-up models. The power law over many decades seen over many decades in energy in the cosmic ray spectrum in Fig. 1.1 implies that the source must generate a power law spectrum and gives us a clue on

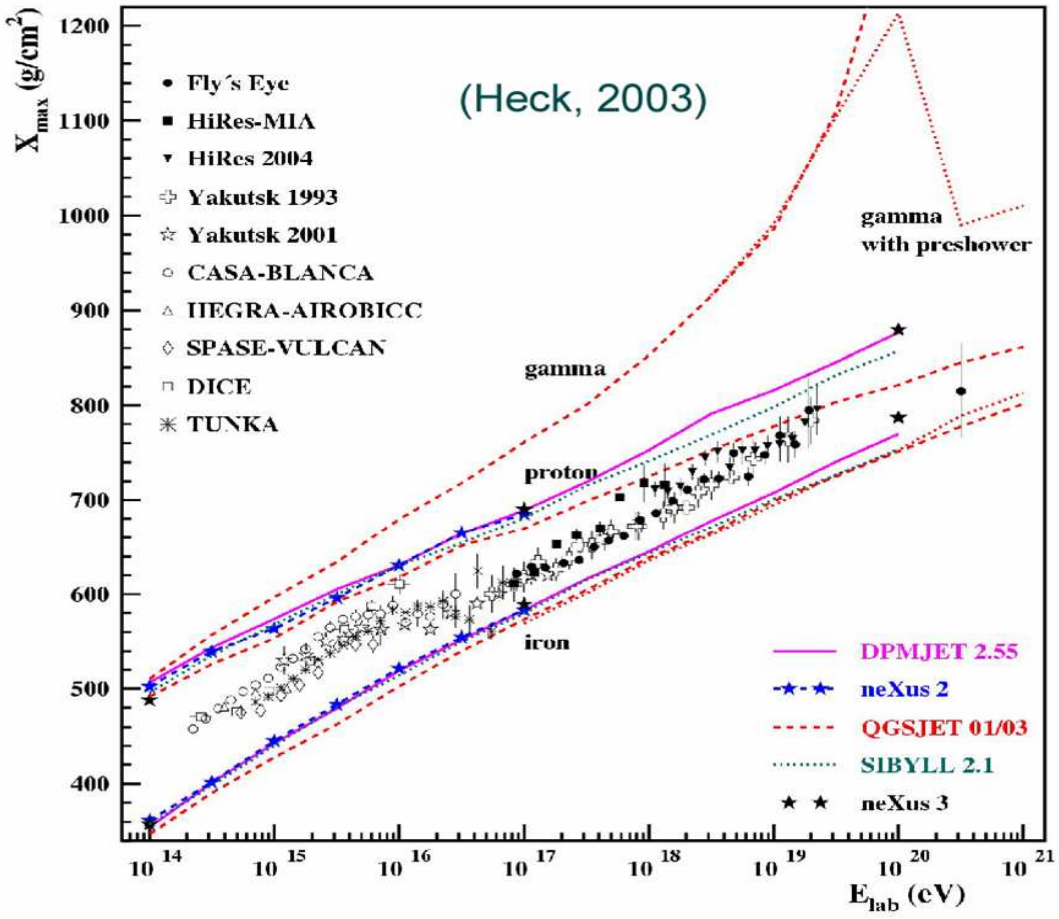


Figure 2.2: Xmax measurements by various experiments. Expectation values for proton and iron primaries from simulation are also shown [9].

possible sources and the acceleration mechanisms.

The cosmic ray energy density makes up a significant portion of the total energy of the universe. The cosmic ray energy density is approximately 1 eV/cm^3 whereas the energy densities of the starlight and the galactic magnetic field are 0.6 and 0.2 eV/cm^3 respectively [6]. Given the present observation level of 10^{20} eV cosmic rays, the corresponding energy density is 10^{-8} eV. Assuming the cosmic rays fill the local super cluster of galaxies with a lifetime of 10^8 years, the local super cluster must pump out approximately $5 \cdot 10^{41}$ eV per second at 10^{20} eV to keep the flux constant [10]. Considering the fact that is comparable to the entire radio band energy output of the galaxies M87 or Cen A, the cosmic ray sources cannot follow a black body radiation spectrum [6]. Whatever the source may be, it must have a nonthermal acceleration mechanism to impart such high energies to cosmic rays that also generates a power law spectrum.

There are two broad categories of acceleration mechanism known as statistical and direct acceleration. In Statistical acceleration schemes, largely owing to Fermi acceleration devised by Fermi [11], cosmic rays gain energy from collisions with magnetic clouds or shockwaves in astronomical objects over a long period of time. In direct acceleration, cosmic rays get accelerated directly by high EMF found in objects such as pulsars, and the duration of acceleration is short in comparison to that of statistical acceleration.

There are two versions of Fermi acceleration. In the first version, charged particles collide with magnetic clouds. There are many magnetic clouds moving in random direction. If we take the frame of an external observer, a particle is just as likely to make a 'head-on' collision as it is to make a 'following' collision. Suppose for a moment that there is an infinitely massive magnetic cloud moving at velocity V and a particle moving at velocity v. Then for a head-on collision,

the change in energy is (See [8] for more details)

$$\triangle E = 2\gamma^2 E \ \frac{V}{c}(\frac{V}{c} + \frac{v}{c}) \tag{2.1}$$

And for a following collision, the change in energy is

$$\triangle E = -2\gamma^2 E \ \frac{V}{c}(\frac{V}{c} - \frac{v}{c}) \tag{2.2}$$

The probability of making a head-on and a following collisions are $1/2((V+v)/v)$ and $1/2((v-V)/v)$,respectively. Then the net energy gained per collision is

$$\triangle E = \frac{1}{2}(\frac{v+V}{v})2\gamma^2 E \ \frac{V}{c}(\frac{V}{c} + \frac{v}{c}) - \frac{1}{2}(\frac{v-V}{v})2\gamma^2 E \ \frac{V}{c}(\frac{V}{c} - \frac{v}{c}) \tag{2.3}$$

Or in a simplified form

$$\frac{\triangle E}{E} \ = \ 4\gamma^2(\frac{V}{c})^2 \tag{2.4}$$

If $V \ll c$, then the rate of gain in energy is

$$\frac{dE}{dt} \ = \ 4M(\frac{V}{c})^2 E \ = \ \alpha E \tag{2.5}$$

where M is the number of collisions per second. Assuming the escape time for the particle from the accelerating region is τ, the diffusion equation for particle acceleration looks like this,

$$\frac{dN}{dt} \ = \ D\nabla^2 N + \frac{\partial}{\partial E}[b(E)N(E)] - \frac{N}{\tau} + Q(E) \tag{2.6}$$

Since we are interested in steady state solution, dN/dt=0. Assuming there are no sources and no diffusion, the $D\nabla^2 N$ and Q(E) terms can be dropped. The energy loss term is b(E)=-dE/dt=-αE. Then the solution to the diffusion equation above is

$$N(E) \ = \ constant \cdot E^{-(1+\alpha^{-1}\tau-1)} \tag{2.7}$$

The resulting spectrum follows a power law. It is actually unclear, however, what the exponent is exactly from the equation above. Other shortcomings of

this model is that since the velocities of clouds are small compared to the speed of light, and the mean free path for collisions is large, it is very difficult get large energy from this mechanism. That we have not considered energy loss is cause for more pessimism. The first version of Fermi acceleration is also known the second order Fermi acceleration due to the V^2 dependence in gains in energy.

Before we take a look at the second version of Fermi acceleration, first, we put the previous treatise in simpler terms. We rewrite $E = E_o\beta$ where β is the energy gain per collision and P as the probability of the particle remaining the accelerating region after one collision. Then after k collisions, $E = E_o\beta^k$ and $N = N_o P^k$. Then it follows that

$$\frac{N}{N_o} = (\frac{E}{E_o})^{\ln P/\ln\beta} \tag{2.8}$$

Then $dN(E) = constant \cdot E^{-1+\ln P/\ln\beta}dE$, and we get a power law spectrum again.

In the second version, the particle gets accelerated by a strong shock wave propagating at a velocity much higher than the speed of sound. This time the particle makes only head-on collisions and the energy increase has a $\triangle E/E \sim 2V/c$ dependence, rather than a V^2 dependence in the first version. A shockwave moves with a velocity $-u_1$. Some of the cosmic rays pass through the shockwave and gain kinetic energy in the process and moves with a velocity u_2 relative to the shockwave (In the lab frame, the shocked particles move at a velocity $-u_1+u_2$ relative to the unshocked particles upstream. In other words, the shocked particles have gained energy). These particles get istropised by the gas behind the shockwave. Some particles recross the shockwave. These particles get isotropised again by the usual scattering processes upstream, but these particles have gained energy in the process. This is the so-called first order Fermi acceleration. When the shockwave catches up with the particles, the process repeats. It turns out for

this mechanism

$$\frac{\triangle E}{E} \;=\; \frac{4}{3}\frac{\triangle u}{c} \tag{2.9}$$

Thus

$$\beta \;=\; \frac{E}{E_o} \;=\; 1+\frac{4}{3}\frac{\triangle u}{c} \tag{2.10}$$

It follows that for $u \ll c$,

$$\ln\beta \;=\; \ln(1+\frac{4}{3}\frac{\triangle u}{c}) \;=\; \frac{4}{3}\frac{\triangle u}{c} \tag{2.11}$$

According to Bell, the number of particles crossing unit surface area is $1/4N_1v$ where N_1 is the particle density [12]. Then for our case of ultrarelativistic particles crossing shockwaves we get $1/4N_1c$. At the same time, u_2N_2 particles get swept away. The particle flow back upstream is then $1/4N_1c - u_2N_2$. Now we make the approximation, $N_1 = N_2$, since cosmic rays hardly feel the shockwave. Then we write

$$P \;=\; 1-\frac{4u_2}{c} \tag{2.12}$$

Then for $u_2 \ll c$, $\ln P = -4u_2/c$. Then,

$$\frac{\ln P}{\ln\beta} \;=\; -\frac{3u_2}{u1-u2} \tag{2.13}$$

To tie it all in, we make use of conservation of mass $\rho_1u_1 = \rho_2u_2$. For strong shockwaves, $\rho_2/\rho_1 = 4$ [13]. Then it follows $\ln P/\ln\beta = -1$. Finally, we have $dN(E) = constant \cdot E^{-2}dE$. We get a power law spectrum which is close to the observed spectrum that is like $E^{-2.5}$ (and for pure iron injection spectrum above 10^{19} eV, E^{-2} is favored (see chapter 8)). Unfortunately, this elegant mechanism probably cannot be responsible for acceleration of cosmic rays beyond 10^{19} eV which is what we are interested in.

The maximum energy attainable in the processes above is believed to be about 10^{15} eV [14]. Acceleration up to 10^{15} eV may be possible for interactions with

multiple supernova remnants [15]. There is no definitive proof, however, of any acceleration mechanism yet. What is clear is that it is very difficult to accelerate cosmic rays to 10^{20} eV, and there are only a handful of source candidates. The larmour radius must be of the size of the acceleration region to confine the particle in the acceleration region, and the magnetic field must be weak enough to prevent synchrotron radiation dissipating energy gained from acceleration.

The maximum energy attainable in a diffusive shock acceleration process was shown by Drury [16] to be $E = kZeBR\beta c$, where Ze is the charge of the particle, B is the magnetic field, R is the size of the acceleration region, k is a efficiency coefficient less than 1. To calculate the absolute maximum energy attainable, we consider the optimal acceleration case by letting $\beta = 1$ (ultrarelativistic shock) and k=1. Then we arrive at $E = 0.9ZBR$, where E is in 10^{18} eV, B in μG, and R in kpc. The size and the magnetic field strength of candidate source astronomical objects are shown in Fig. 2.3. As can be seen, there is only few candidate sources capable of accelerating particles to 10^{20} eV even at $\beta = 1$ (see the solid and the dotted diagonal lines). If $\beta = 1/300$, then the size of the acceleration region required becomes too large, and there is no astronomical object capable of accelerating particles to 10^{20} eV! This illustrates the difficulty in devising schemes to accelerate particles to 10^{20} eV with the current state of knowledge.

Direct acceleration occurs in compact astronomical objects with high magnetic fields such as neutron stars and pulsars [8, 17]. They can rotate at $\sim$ 30 Hz, and the surface magnetic field can be as high as 10^{12} G. The induced EMF is $\sim 10^{18}$ eV. Thus, cosmic rays can be accelerated to high energies and the duration of acceleration can be quite short unlike statistical acceleration. It is unclear, however, where the EMF drop occurs, and that has an effect on the

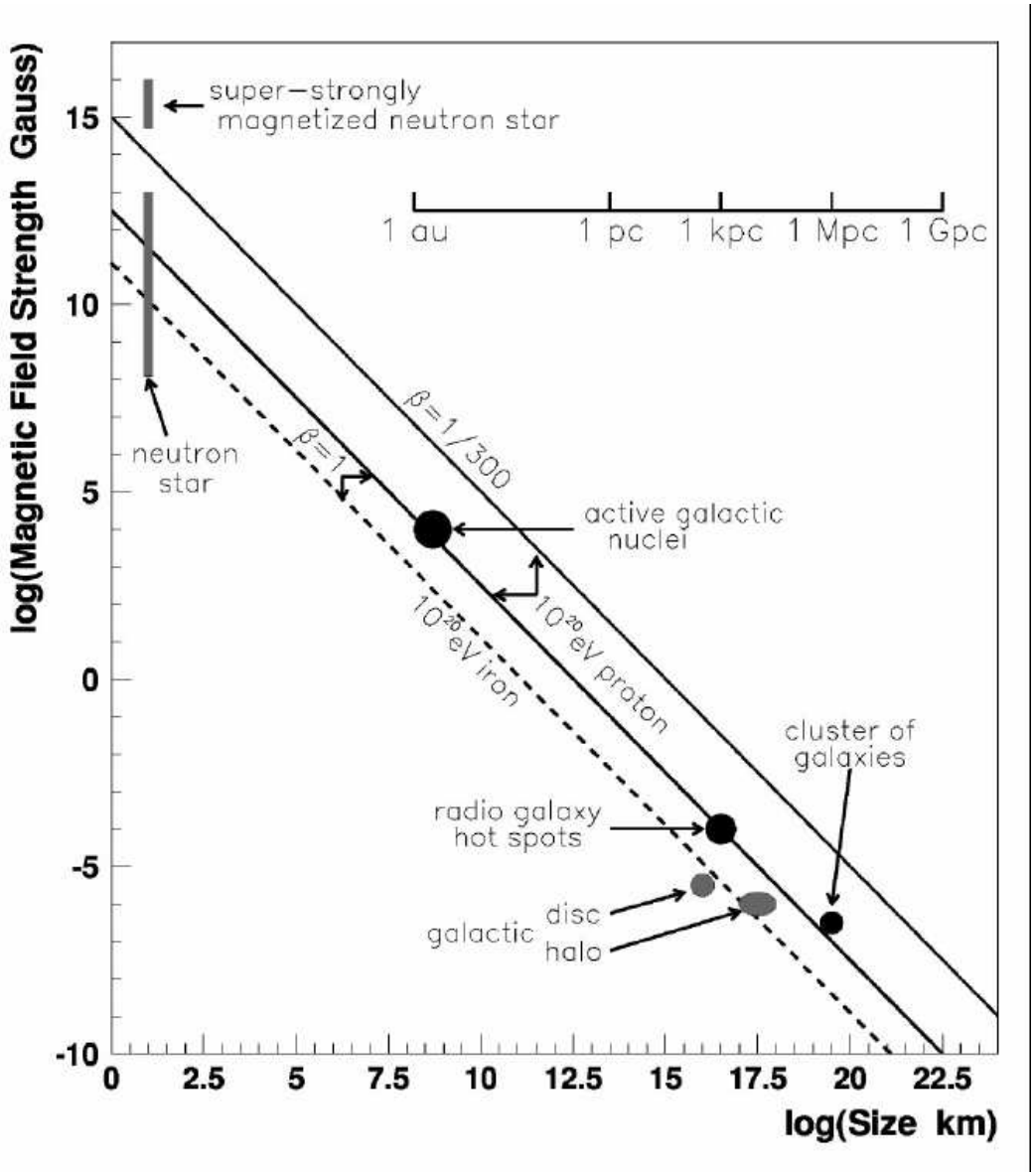


Figure 2.3: The size and the strength of the magnetic field required to accelerate particles to 10^{20} eV and possible candidates. Not so many candidates satisfy the minimum requirements. The Plot is from [14].

energy loss. Longair puts the upper bound of the energy attainable at pulsars at $3 \cdot 10^{19}$ eV, with reasonable assumptions. Though this may not be entirely true, it illustrates the point that a particle can be accelerated to a high energy through direct acceleration. On the downside, however, it is unclear how direct acceleration results in a power law spectrum.

Besides the conventional 'bottom-up' models above, there are more exotic models called 'top-down' models in which cosmic rays are byproducts of the decays of exotic particles. Topological defects such as cosmic strings and monopoles are created in early universe and are sources of super massive particles which then subsequently decay [71]. Super Heavy Dark Matter are weakly interacting massive particles that have a long lifetime [72]. In Z-Bursts, ultrahigh energy neutrinos enter the GZK sphere (see next section) and interact with relic neutrinos created in early universe. The resulting Z boson decays into pions and nucleons. Top-down models all predict a significant fraction of flux at the highest energies to be gamma rays. Measurement of gamma ray flux at the highest energies can test these models directly.

As can be seen, the subject of source and acceleration of cosmic rays is rather messy as there are myriad of source and acceleration models. And as we saw earlier, there are only a few candidate sources that are capable of accelerating a particle to 10^{20} eV even assuming maximum efficiency. And some models even invoke very exotic physics. Clearly, understanding how particles are accelerated to 10^{20} eV is no trivial matter.

2.3 Propagation

So after cosmic rays get accelerated and leave the acceleration site, how do they get to us and what happens in transit? We saw in Fig. 2.1 that the chemical composition of cosmic rays agrees with the solar abundances, at least for cosmic rays below PeV. The mean free path of spallation for heavy nuclei is 10 g/cm^2. Hence, cosmic rays cannot encounter more than 5 g/cm^2 of matter, otherwise the chemical composition of cosmic rays would be drastically different from the solar abundances.

For high energy cosmic rays, interactions with radiation background present throughout the intergalactic space, such as the cosmic microwave background (CMB), infrared, and radio backgrounds, become important. For protons, interactions with the CMB loom large. Soon after the discovery of the CMB by Penzias and Wilson [18], Greisen and Kuzmin and Zatsepin independently concluded that if cosmic rays are protons then there would be a sharp drop-off in the cosmic ray spectrum around $6 \cdot 10^{19}$ eV due to interactions with the CMB [66, 67]. This is the so-called GZK cutoff. The actual interactions primarily takes place as follows,

$$p + \gamma_{CMB} -> \triangle -> p + \pi^0 \tag{2.14}$$

$$p + \gamma_{CMB} -> \triangle -> n + \pi^+ \tag{2.15}$$

The CMB gets blueshifted in the rest frame of the proton, and when the energy of a proton reaches $6 \cdot 10^{19}$ eV, the $\triangle^+$ resonance occurs, which then subsequently decays to protons (or neutrons) plus pions. This process results in a 20% loss in energy in the proton [14]. This interaction reduces the energy loss length of a proton from $\sim$ 1 Gpc at 10^{19} eV to $\sim$ 10 Mpc at 10^{19} eV (see Fig. 2.4. This means the 'GZK sphere' is 1 million times smaller in volume than the 'transparent

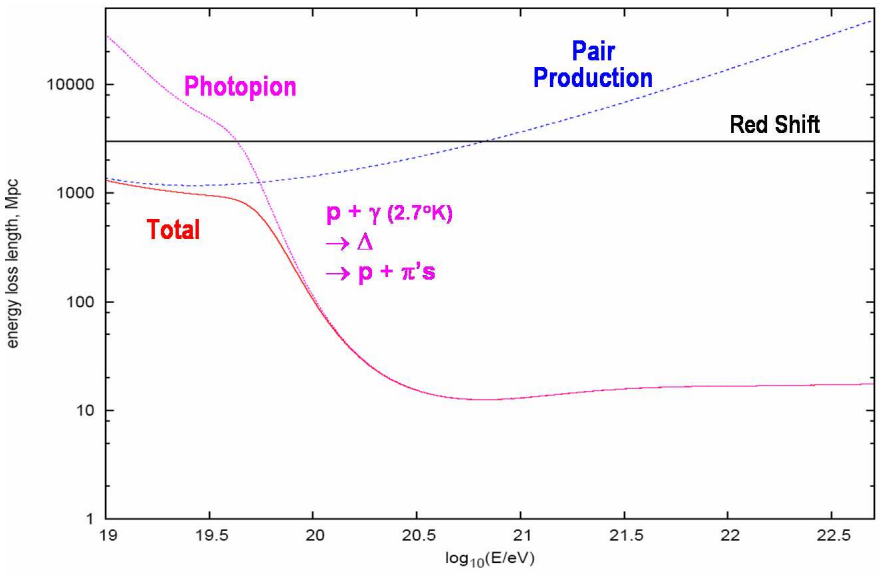


Figure 2.4: Energy loss length of proton. Photopion production becomes dominant above ~ 60 EeV. Plot courtesy of O. Kalashev.

universe' at 10^{19} eV. This is responsible for the drop-off in the spectrum above $6 \cdot 10^{19}$ eV). Therefore, 10^{20} eV protons, if they exist, must come from nearby, in cosmological terms.

Another important interaction channel for the proton-CMB interaction is

$$p + \gamma_{CMB} -> p + e^+ + e^- \tag{2.16}$$

This process has a threshold of 10^{18} eV and a mean free path of 1 Mpc, but the energy loss from this interaction is only 0.1% [14]. As can be seen in Fig. 2.4, this process, while dominant at low energies before the pion production becomes dominant around $6 \cdot 10^{19}$ eV, is still very weak and results in an energy loss length of ~ 1 Gpc.

For heavy nuclei, photopion production process becomes less important, and photodisintegration and pair production processes are the important processes.

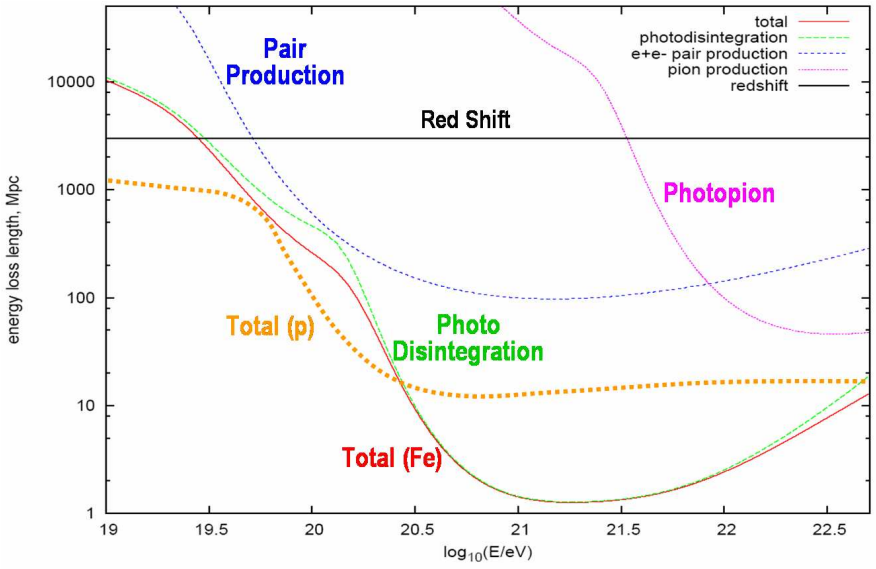


Figure 2.5: Energy loss length of iron. Photodisintegration is the dominant process. Plot courtesy of O. Kalashev.

Photodisintegration process takes place as follows:

$$A + \gamma -> (A-1) + N \tag{2.17}$$

$$A + \gamma -> (A-2) + 2N \tag{2.18}$$

Heavy nuclei of mass A interacts with CMB to produce nucleons, N (below $5 \cdot 10^{19}$ eV, interactions with infrared photons become important). The energy loss rate through the single nucleon channel is about one order of magnitude higher than that of the double nucleon channel. Fig. 2.5 shows the contribution of different processes to the energy loss length of Fe. Clearly, photo disintegration is the dominant process. Unlike with protons, photopion production process is almost inconsequential for Fe. As with protons, heavy nuclei do not disintegrate through pair production, but the energy loss through pair production is more significant than with protons. The energy loss length through pair production is actually comparable to that of photodisintegration around 10^{20} eV. If cosmic rays

at the acceleration sites are Fe, then it should exhibit a drop-off in the spectrum akin to the GZK cutoff as the energy loss length for Fe also drops precipitously with increasing energy. Also, even if cosmic rays are all Fe at the sources, a mixed composition is expected at earth as the Fe nuclei (and the daughter nuclei from photodisintegration) will undergo photodisintegration during propagation.

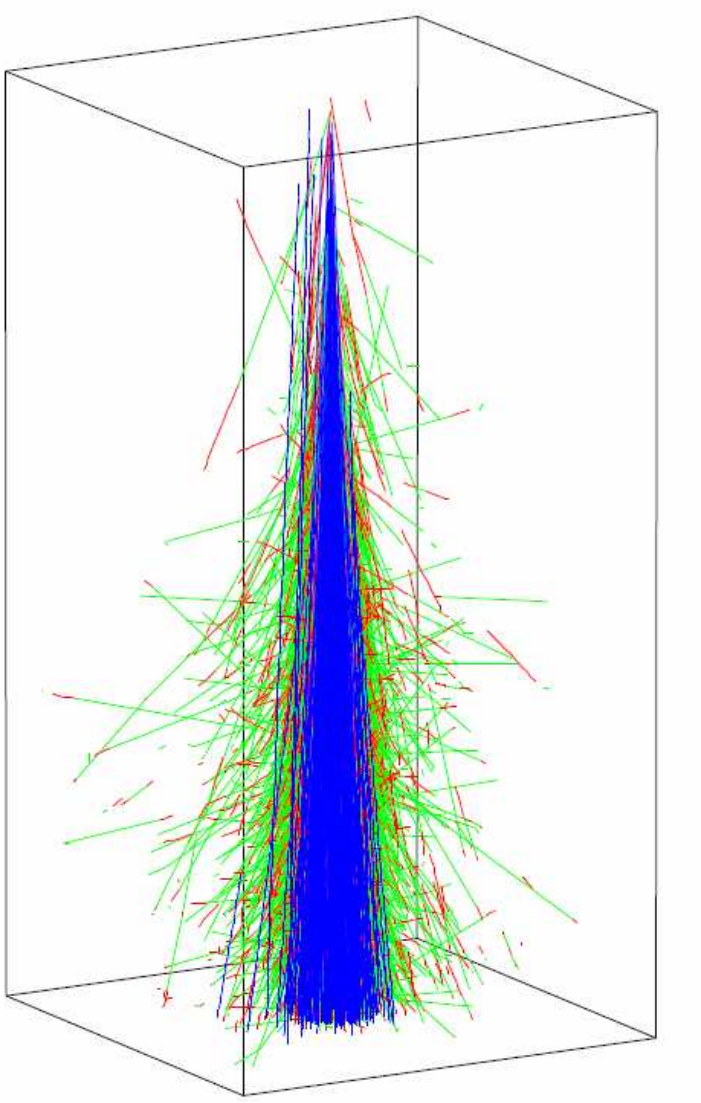
2.4 Extensive Air Shower

Cosmic rays incident on earth interact with particles in the atmosphere and creates secondary, tertiary, and higher order particles. This interaction initiates a cascade of particles called extensive air shower (EAS) (see Fig. 2.6). At energies above 10^{18} eV, the flux of cosmic rays is too small to make it practical to directly detect cosmic rays. Instead, cosmic rays at this energy scale are studied indirectly by a detector covering a large area on the ground detecting the EAS produced by cosmic rays.

Assuming the cosmic ray primary is a nucleon, hadronic interaction between the primary and the air molecules feeds the cascade. In subsequent interactions, the number of hadrons increases. As can be seen in Fig. 2.7, in each generation, π^0's decay and about 30% of the energy is transferred to the electromagnetic cascade [19]. Eventually, 90% of the energy of the primary goes to the electromagnetic cascade, and the remaining 10% is carried by muons and neutrinos.

2.4.1 Electromagnetic Cascade

A simplified picture of the electromagnetic cascade illustrates how the EAS develops. A photon of energy E_0 traverses distance R before undergoing a pair creation. Then the electron-positron pair, carrying on average 1/2 of E_0 each,



Simulation of a 10^{19} eV proton EAS using the MOCCA program. A sample of tracks at > 300 m from the shower axis are shown. Frame box: $6 \times 6 \times 12$ km high. Color code: γ green, e red, μ blue. *Drawn by Clem Pryke — University of Chicago*

Figure 2.6: Simulation of Extensive Air Shower. Note: Although the EM component is obscured by muons (blue) in this drawing, the EM component is by far the dominant component.

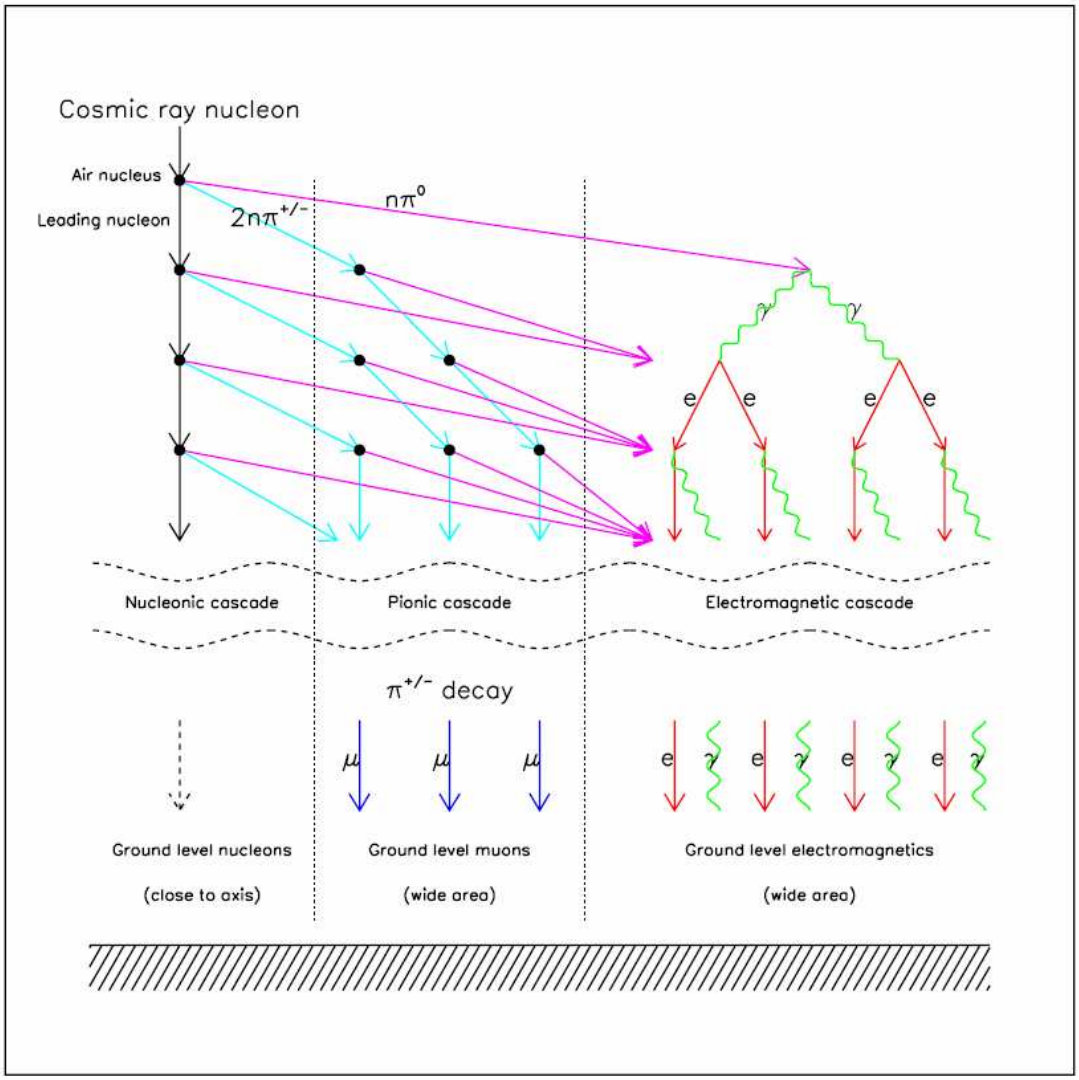


Figure 2.7: Simplified picture of the cascade processes in EAS. Figure from [19].

produces a photon of energy $E_0/4$ each traveling a distance R. Thus we can see that after nR, 2^n particles will be created with an average energy of $E_0/2^n$. The shower is going to keep on growing until the average energy of the particle drops below the critical energy. The critical energy, E_c, is defined as the energy below which the energy loss through ionization outstrips the energy loss through bremsstrahlung. For photons, E_c corresponds to the energy where Compton scattering overtakes pair production. The shower grows to maximum when the average of energy of the shower particles are equal to the critical energy, and the shower size should be proportional to E_0 divided by E_c. After the shower maximum, with the particle multiplication process halted, no particle will be added while existing shower particles dissipate. The shower size is going to shrink as a result.

In the high energy regime of interest to us, the pair production length, ϵ_0, is more or less equal to the radiation length for bremsstrahlung [6]. The distance nR required to reach shower maximum, N_{max}, is given by using the relation $E_0/2^n = E_c$. It follows that $n = \ln(E_0/E_c)/\ln 2$. Thus the depth of the shower maximum, X_{max}, has a logarithmic dependence on the incident energy, while N_{max} is proportional to E_0/E_c, hence linearly dependent on the incident energy. For real showers generated by hadrons that are of interest to us, the linear dependence of shower maximum on the incident energy still holds. Over the range of different models, $N_{max} = (1.1\ to\ 1.6)\ E_0(\mathrm{GeV})$ [5]. For example, a 10^{18} eV particle will have an N_{max} of 10^9. The actual longitudinal development of a shower is well described by the so-called Gaisser-Hillas function,

$$N(X) \ = \ N_{max} \left(\frac{X - X_0}{X_{max} - X_0} \right)^{(\frac{X_{max} - X_0}{\lambda})} \exp\left(\frac{X_{max} - X_0}{\lambda} \right) \qquad (2.19)$$

where X_0 is the point of initial interaction and $\lambda = 70$ g/cm^2.

2.4.2 Hadronic Cascade

The depth of first interaction, X_0, is 70 g/cm^2 for protons and 15 g/cm^2 for irons at 10^{15} eV. X_{max} is strongly influenced by X_0 and the energy loss that takes place in the first interaction. Protons have longer interaction lengths than irons. Thus, X_{max} for protons will have larger fluctuations. Furthermore, as mentioned earlier, hadronic showers can be treated as just a superposition of many electromagnetic showers initiated by pion decays fed by the hadronic core.

Irons with energy E can be considered as a superposition of nucleons with energy E/A, where A is the atomic number. Since X_{max} has a logarithmic dependence on energy, each subshower initiated by the nucleons will be shallower than a shower initiated by a proton with energy E. Thus, X_{max} for irons is shallower than for protons at the same energy. Also, the number of muons produced grows as $E^{0.85}$ [20]. Treating irons with energy E as a superposition of nucleons with energy E/A again, we find that the number of muons produced by irons and protons at the same energy has the relation $N_A = A^{0.15}N_p$. Since A=56 for irons, iron showers will have $\sim 80\%$ more muons than proton showers.

As number of particles in the shower scales with the energy of the primary, measuring the number of particles should give us the energy of the primary. Most experiments measure the lateral distribution of the shower particles rather than the longitudinal profile of the shower. The shower is completely dominated by electromagnetic component and hence the lateral distribution is determined largely by multiple scattering of the electrons. Experimental data show that, if muons are excluded, the lateral distribution agrees well with the expectations from a purely electromagnetic cascade. Fig. 2.8 shows the lateral distribution of electrons and muons expected from a 10^{19} eV shower. Details of energy determination and direction vector of cosmic ray primaries using the lateral distribution

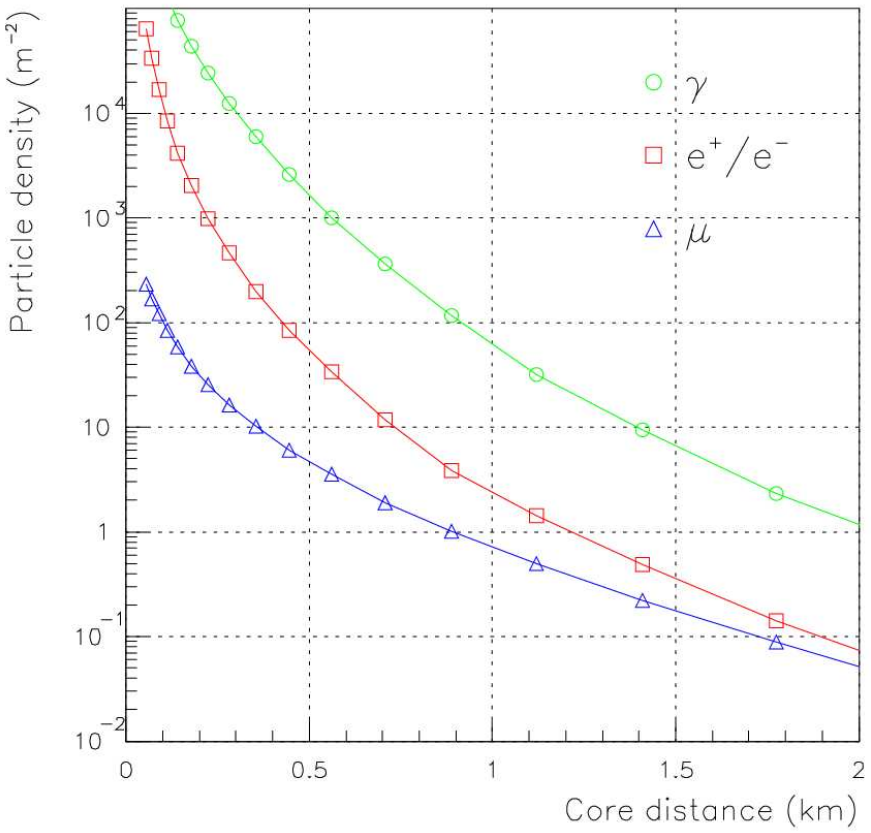


Figure 2.8: Lateral distribution of a simulated 10^{19} eV proton shower. The electromagnetic magnetic component completely dominates near the core. Figure from [19].

and longitudinal development of the shower will be discussed in later chapters.

2.5 Past Experiments and Summary

Now that we have had a brief overview of the theory and background information, we are in a position to discuss past experiments and the current state of things. At present, our understanding of cosmic rays above 10^{18} eV, or UHECR, is murky. We do not know the composition of UHECR or the source of UHECR. At 10^{15} eV

and above, we cannot use direct detection methods to determine the composition. The composition above 10^{15} eV is determined by indirect measurements such as X_{max} measurements. Fig. 2.2 shows the results of X_{max} measurements by several experiments. X_{max} measurements lie between those predicted by pure proton and pure iron compositions. According to HiRes, Fly's Eye, and Yakutsk in Fig. 2.2, X_{max} goes from iron-like at low energies to proton-like at high energies which implies that the composition changes from heavy nuclei to light nuclei — This is somewhat contradictory to the latest composition studies at UCLA by M. Healy [84] as well as other analyses in Auger, elongation rate measurement by M. Unger *et. al.* for example, which indicate that the composition goes from proton-like to iron-like above 10^{19} eV, but there is only scant amount data at this energy scale for other experiments. What we would like is the kind of detailed measurement of composition that is obtained with direct detection methods at TeV to PeV scale.

Cosmic ray energy spectrum also is in a state of confusion. Two recent experiments, HiRes and AGASA, report two very different spectra, shown in Fig. 2.9. The absolute flux reported by the two experiments do not agree at all. Not only that, the HiRes spectrum show a sharp decrease in flux above $\sim 6 \cdot 10^{19}$ eV similar to the GZK cutoff. AGASA spectrum, on the other, has an excess of super-GZK events and do not agree with the GZK cutoff. The two spectra have very different implications. HiRes spectrum is consistent with consistent with the conventional bottom-up models and the propagation models. It also means that the super-GZK events coming from relatively nearby within the GZK sphere. AGASA spectrum, on the other hand, implies that we need more than the bottom-up models to make sense of it. Top-down models have been suggested as a way to do that.

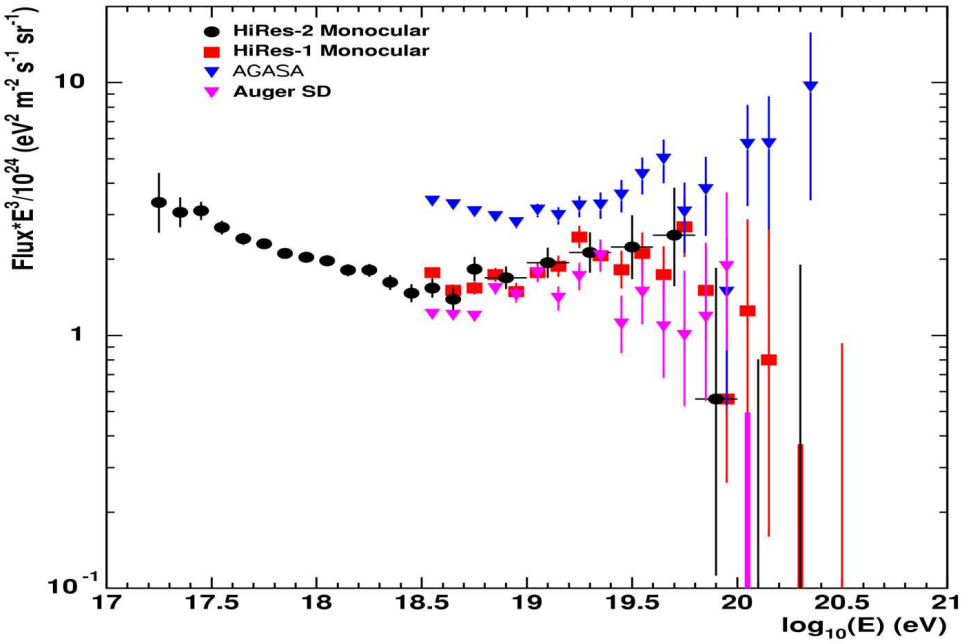


Figure 2.9: HiRes and AGASA spectra [21].

Anisotropy measurements so far has yielded no clear evidence of correlation with sources. Above 10^{19} eV, the bending angle of a proton is expected to be very small, so we expect to see clustering and correlation with sources. If the primaries are irons, then the bending angle would be hopelessly large to see any kind of anisotropy. AGASA found a clustering of events above $4 \cdot 10^{19}$ eV within 2.5° of one another — 5 doublets and 1 triplet out of 57 events in total (see Fig. 2.10) [22]. The reported significance was 10^{-4}. However, with penalty factors taken into account the significance drops to 0.003 [23]. Gorbunov *et. al.* reported a correlation between BL-Lac objects and HiRes events above 10^{19} eV at 0.7° separation with a significance of $\sim 10^{-4}$ (see Fig. 2.11) [24]. The fact that the significances reported by HiRes and AGASA are relatively low notwithstanding (a significance of 10^{-6} is desired to claim correlation), neither result has been confirmed by other experiments.

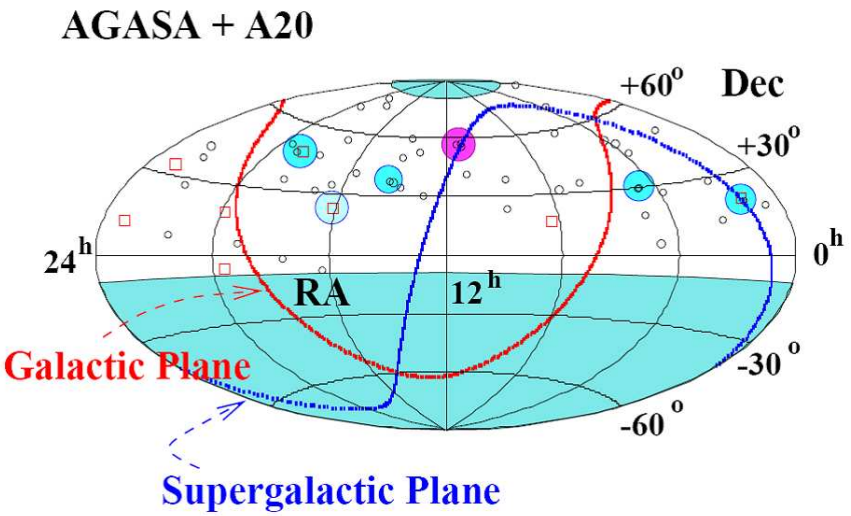


Figure 2.10: Sky map of AGASA events above 40 EeV. 5 doublets within 2.5° and 1 triplet within 1° out of 57 events give a significance of 10^{-4}. Fig. is from [22].

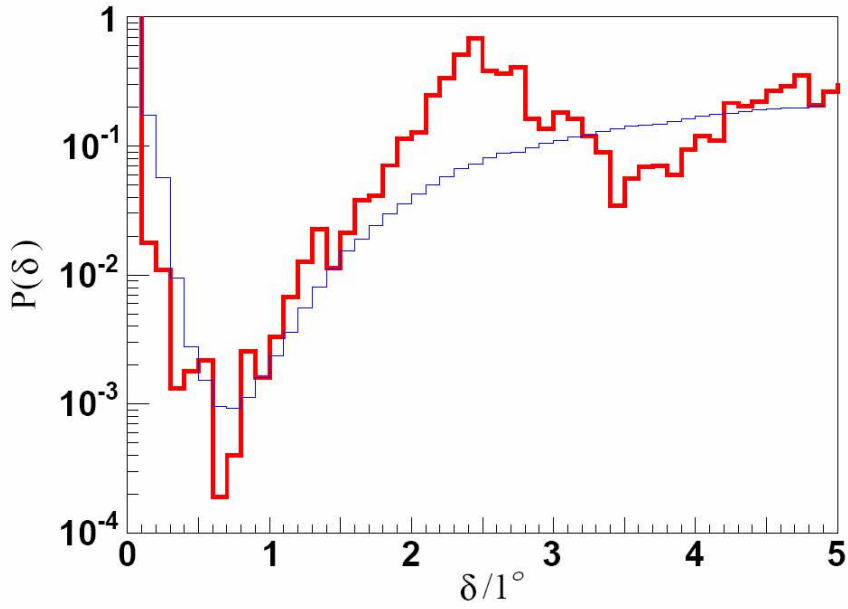


Figure 2.11: HiRes events above 10 EeV correlation with 157 BL Lacs with M > 18. Fig. is from [24].

Our brief overview of the composition, spectrum, and anisotropy studies illustrate the necessities of the multi pronged approach to solving the mystery of the UHECR. The present state allows for several possibilities for the origin of the UHECR. If Auger sees a continuous spectrum like that of AGASA, then top-down models could be a possibility. Top-Down models predict significant gamma ray flux, thus a gamma ray flux analysis can be done to determine whether our spectrum does agree with a top-down model. On the other hand, if Auger spectrum is consistent with the GZK cutoff, then all the super-GZK events have to come from nearby. Thus, if the super-GZK events are protons, we should be able to see some kind of anisotropies. If they are irons or heavy nuclei, on the other hand, then they would be hopelessly jumbled by the intergalactic and galactic magnetic fields, and it would be hopelessly to see any kind of anisotropy — this is a particularly depressing scenario as we would not be able to identify the sources seriously limiting the kinds of questions we can answer. And composition studies should tell us what cosmic rays are exactly. Auger can do all kinds of studies with more statistics than any previous experiment and will help answer the all of these outstanding questions with more precision.

CHAPTER 3

Pierre Auger Observatory

The Pierre Auger Observatory (PAO) was designed with the goal of understanding the ultra high energy cosmic rays (UHECR) above 10^{19} eV. As mentioned earlier, the cosmic ray flux above 10^{19} eV is extremely low — about 1 per km^2·century above 10^{20} eV. In order to overcome the extreme low flux, it is necessary to build a detector covering a large area and collect data over a long period of time. At these energies, the extreme low flux makes direct detection methods impractical, and past experiments, AGASA and HiRes to name a few, employed indirect detection methods measuring either the longitudinal development or the lateral distribution of the shower.

The PAO is located in the pampas in Mendoza province of Argentina. When completed, the PAO will cover ~ 3000 km^2 which is much larger than any previous experiment and will run for 20 years giving us unprecedented large statistics for UHECR's. It is a hybrid detector that consists of the fluorescence detector (FD), which measures both the longitudinal development of the shower, and the surface detector (SD), which measures the lateral distribution of the shower (see Fig. 3.1). The SD consists of an array of 1600 water Cherenkov tanks in a triangular grid with a 1.5 km spacing. The FD consists of 4 fluorescence detector sites, or eyes, on the periphery of the large are covered by the SD (see Fig. 3.2 for the actual layout).

The SD measures the lateral distribution of the shower by sampling the

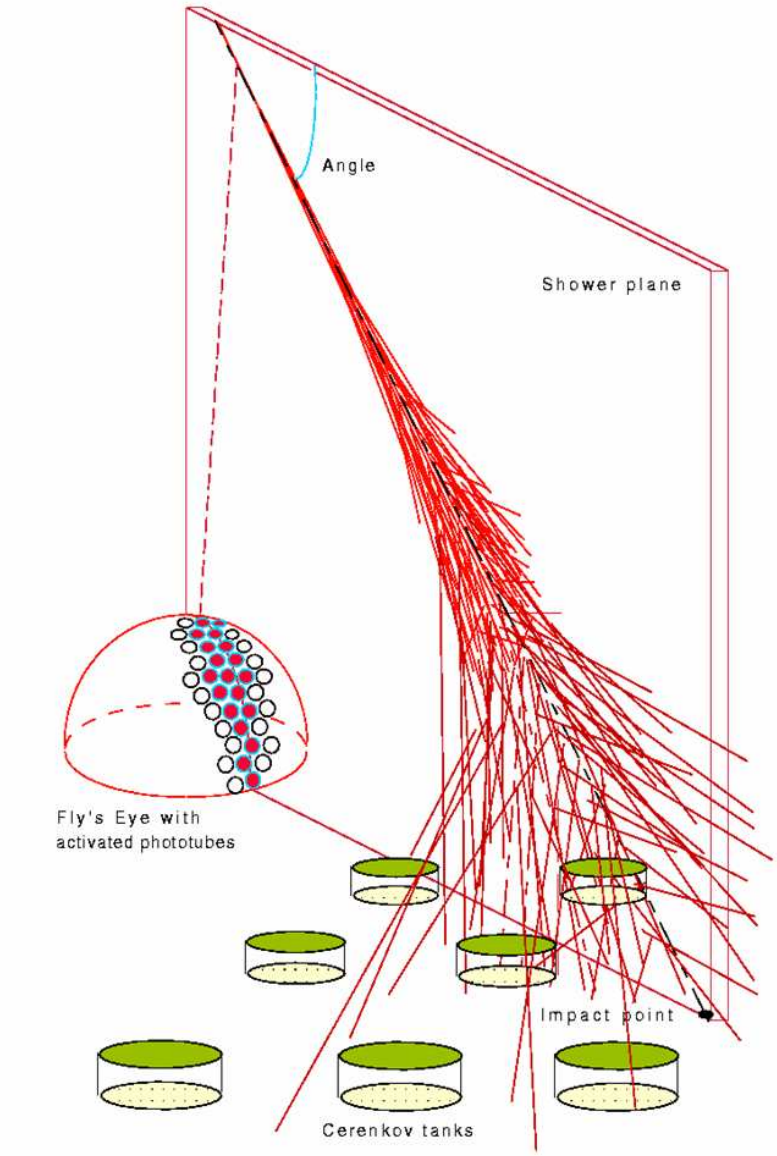


Figure 3.1: Artist's concept of the PAO. The FD detects the fluorescence light, while the water Cherenkov tanks in the SD samples the shower particles.

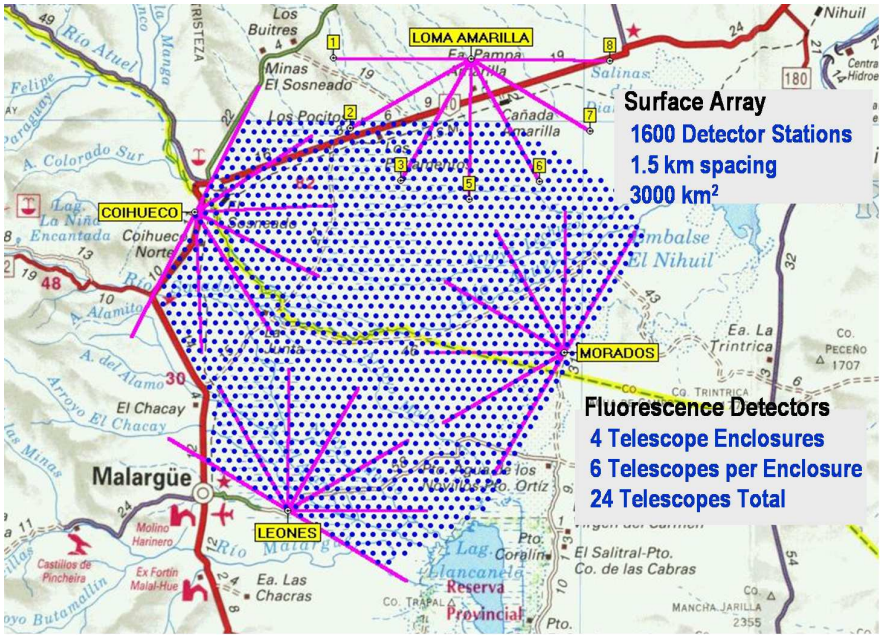


Figure 3.2: Map showing the layout of the PAO. There are 1600 tanks (dots) covering 3000 km^2 with four FD eyes on the periphery.

shower particles at different distances from the core of the shower with the water Cherenkov tanks. Each tank (or station) measures the Cherenkov light emitted by the shower particles in the water (Note: The terms 'tank' and 'station' will be used interchangeably throughout this document). The FD measures the longitudinal development of the shower by detecting the fluorescence light from the excitation of nitrogen molecules in the atmosphere by the shower particles (see Fig. 3.1). The FD and the SD have different systematics. And it has been suggested that the differences in the HiRes and AGASA spectra (see Fig. 2.9) are due to the systematics of the two methods — fluorescence detection was employed by HiRes and the surface array method was used by AGASA. The PAO, thus, offers an opportunity to check the systematics of the two methods and cross-calibrate.

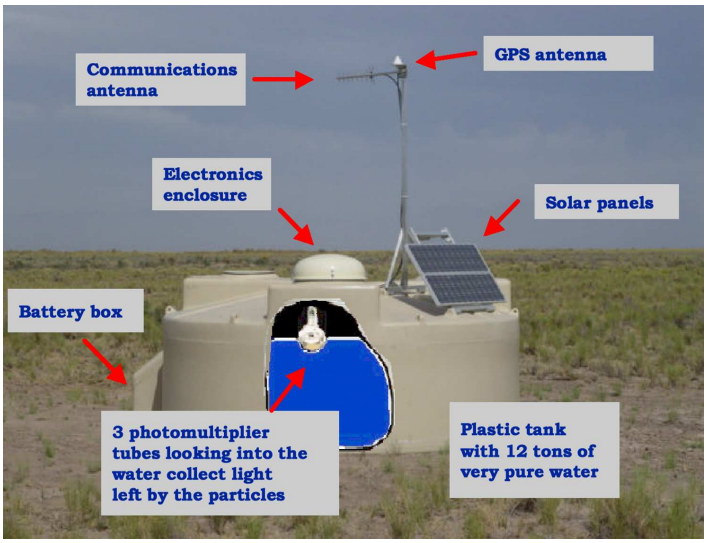


Figure 3.3: Water Cherenkov Detector. Notice the PMT looking down at the water. There are 3 PMT's per tank.

3.1 Surface Detector

3.1.1 Detector Description

Previously, the water Cherenkov detector was employed successfully for 20 years in Haverah Park. The water Cherenkov detector in Auger is a plastic cylindrical tank with a cross sectional area of 10 m^2 and a height of 1.5 m (see Fig. 3.3). Besides the Photomultiplier tube (PMT) and the electronics that measures the signals from showers, each water tank has its own solar panel, batteries, GPS and communications systems. Each water tank has two lead-acid batteries. Two solar panels in series charge the batteries as well as power the electronics onboard. The GPS system onboard allows for accurate determination of the position of the tanks as well as establish a common time base for correlating data taken at different stations. The timing resolution of the GPS systems are calibrated to

within 15 ns [25]. Each station has a wireless communications system that relays information back and forth to the Central Data Acquisition System (CDAS).

The water inside the tank is ultra-pure water with resistivity above 15 Mohm-cm and will remain stable over the 20 years of operation [25]. The water is enclosed in a Tyvek liner bag with a height of 1.2 m for a total of 12 tons of water. The Tyvek liner diffusively reflect Cherenkov photons produced by shower particles traversing through the water, and the Cherenkov photons reach the PMT's. There are three 9" Photonis PMT's inside the tank placed in three windows at the top of the Tyvek liner.

The PMT's are designed to operate at a relatively low gain of $\sim 2 \times 10^5$ at the dynode channel and also maintain linearity of 5% at a relatively large anode current of ~ 50 mA at this gain [25]. This is necessary to achieve a large dynamic range as the PMT needs to be able to see signals from single muons, for calibration purposes, as well as large signals from large showers. The signals from the PMT are split into two channels, dynode and anode. The dynode channel has a gain ~ 32 times that of anode. In case of saturation of signals at the dynode, the signal at the anode is used.

The PMT signals from the six PMT's are filtered through 20 MHz filters and fed to six 40 MHz 10-bit Flash Analog to Digital Converters (FADC's). Since the dynode gain is $\sim$32 times that of the anode, the nominal dynamic range is, then, 15 bits, or corresponding to a few to 20,000 photoelectrons.

3.1.2 Triggering Scheme and Event Selection

As the PAO was originally designed to study cosmic rays above 10^{19} eV, the trigger condition is set up to trigger efficiently on showers of energy above 10^{19} eV. In considering the basic criterion for the trigger, some characteristics of shower

need to be considered. Higher energy showers tend to be more dispersed in time. Since muons on average have larger energy —average muon energy is on the order of 1 GeV whereas the electromagnetic component is on the order of a few MeV)— they give larger Cherenkov signals. Since the muonic component of showers are composition dependent, this exposes trigger to the possibility of compositional dependence.

This gives us two basic conditions for trigger. The triggered signal should be extended in time. Pulse height threshold should be kept low to minimize the dependence of trigger on the muon content.

We have a hierarchical triggering scheme in the SD. T1 is the lowest level trigger and is triggered at the hardware level at the frequency of 100 Hz. When a signal is above 3 VEM (refer to chapter 4 for the discussion on VEM) on the three PMT's, or 13 bins out of a 120 bin window have a signal over 0.2 VEM, that triggers T2. The decision for T3 is made at the observatory campus based on temporal and spatial correlation of T2's. The bare minimum for a T3 trigger is a three-fold triggers (i.e. three neighboring tanks not on the same line that are correlated in time). As we will see later, we need at least a three-fold event to reconstruct the event. Then, there is an offline trigger, T4, that is applied to T3's in the analysis step that decides whether a T3 is actually a real cosmic ray shower event [26].

There are actually two different modes for T1. The first kind is a Time over Threshold (ToT) trigger, in which 13 bins in a 120 bin window are above a threshold of 0.2 I^{est}_{VEM} in 2 PMT's where I^{est}_{VEM} is the estimated current for a vertical equivalent muon (see chapter 4). The trigger rate for this trigger is only about 1.6 Hz. This trigger is very efficient for selecting small, spread out signals that are from distant high energy showers or low energy showers, while sifting

out single muon backgrounds which would leave large pulses in short duration. The second type of T1 is a 3-fold coincidence of 1.75 I_{VEM}^{est} threshold. The trigger rate for this is about 100 Hz. This trigger is for selecting fast signals from muonic components that are from horizontal showers.

Then, the T2 trigger is applied in the station controller to cull T1 signals that are likely to have come from showers. All ToT triggers are automatically promoted to T2 while the T1 threshold triggers are required to pass a 3.2 I_{VEM}^{est} in 3 PMT's in coincidence, which trims the trigger rate down to $\sim$ 20 Hz.

The T3 trigger decision is made at the observatory campus based on the spatial and temporal correlation of stations. The main T3 requires a coincidence of 3 tanks that passes the ToT trigger (3ToT) that form a triangle. 90% of the events selected by this condition are real (vertical) showers, and the trigger rate is about 1.3 events per day per triangle of 3 neighboring stations [26]. The other T3 trigger requires a 4-fold coincidence (4C1) of any T2 with three tanks forming a triangle and one tank being allowed to be as far as 6 km away from the other tanks. This trigger is primarily for the detection of the horizontal showers, and only 2% are real showers [26].

The next level trigger, T4, selects events out of T4 that are real showers offline. The process starts with a seed, the station with highest signal with 2 neighbors in a non-aligned configuration forming a triangle. The tanks in an event must be compatible with the propagation of the shower plane front at the speed of light, otherwise they will be marked accidental and removed from the event. In the event that not enough tanks survive the removal of accidentals, these events are rejected as they cannot be reconstructed — at least 3 non-aligned tanks are required for reconstruction. The 3ToT T4 triggers on $\sim$ 95% of showers below 60° while 4C1 recovers the remaining $\sim$ 5% as well as showers above 60° [26].

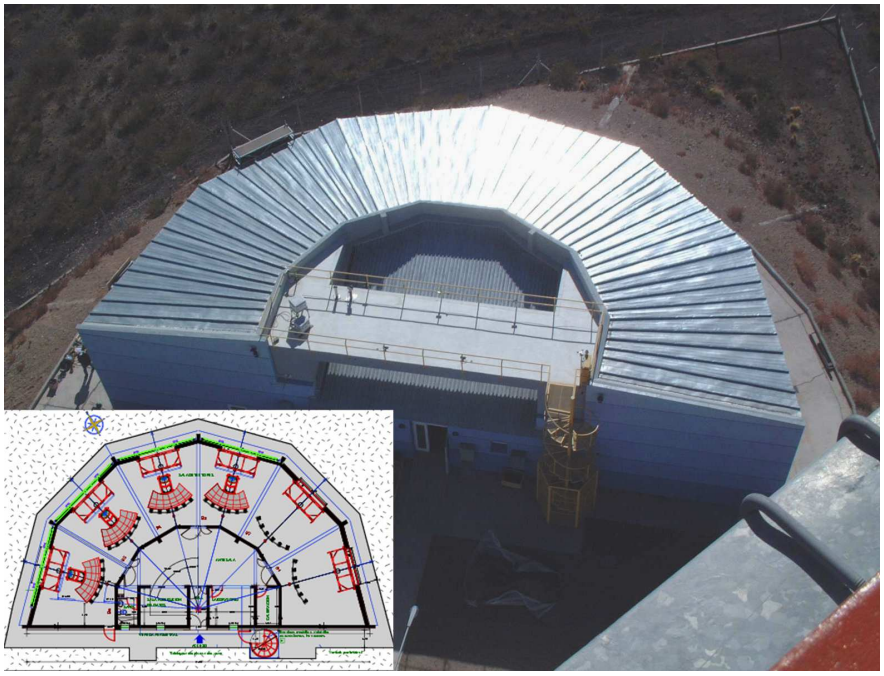


Figure 3.4: Fluorescence Detector. The inset shows the schematic of the FD with six telescopes.

The number of T4 events per day has a temperature dependence of about 1% per degree, which must be taken into account when the acceptance estimation is done for the low energies where the acceptance is not saturation.

T5 is an offline quality trigger that ensures good reconstruction accuracy and ease of computation of the acceptance of the detector. T5 requires the tank with the highest signal be surround by 6 working (but not necessarily all triggered) tanks to ensure that the core of the shower is within the seed — if a shower falls at the edge of the shower, a situation can arise where the seed catches only a part of the shower away from the core of the shower, and the reconstructed core is inside the triangle, while real core is outside the triangle. The exposure calculation just amounts to adding up the number of operational hexagons during the live-time of the detector (see chapter 7).

3.2 Fluorescence Detector

The FD consists of 4 'eyes' (see Fig. 3.4) that measure the longitudinal profiles of showers by detecting the fluorescence photons from excitation of N_2 molecules in the air by shower particles. Each eye contains 6 telescopes, and each telescope covers 30° x 30° (refer to the inset in Fig. 3.4). In the telescope, the photons enter through a Schmidt diaphragm that has a radius of 1.5 m. The fluorescence photon are mostly in the 300–400 nm range, and each telescope has a UV filter that only transmits photons in the said range. The photons are collected using a 3.5 m x 3.5 m spherical mirror (see Fig. 3.5). The Schmidt diaphragm was chosen to eliminate coma aberration, which results in the telescope optics becoming nearly spherically symmetrical [27]. The photons are detected by the camera which has 440 pixels. Each pixel covers 1.5° and has a hexagonal PMT. The signal from the PMT is fed to a 12 bit ADC every 100 ns. As is the case with the SD, when there is an interesting pattern in the FD, the FD sends the data to the CDAS. When there is an FD event trigger, the CDAS looks for a pattern in the SD that may be correlated with the FD trigger. These are the so-called 'hybrid events' as they are seen by both the FD and the SD at the same time, and these events are valuable as they allow for a direct comparison of the two methods.

3.2.1 Calibration

The calibration strategy pursued in the FD is an end-to-end calibration. What this means instead of evaluating the effects of the diaphragm, the filter, the mirror, the PMT response, *etc.* one by one, we fold everything in and just measure the response of a pixel to given flux of photons. Drum Calibration is such a method. This technique uses a portable light source that illuminates all the PMT's in a camera with a uniform pulsed photons (see Fig. 3.6). A pulsed

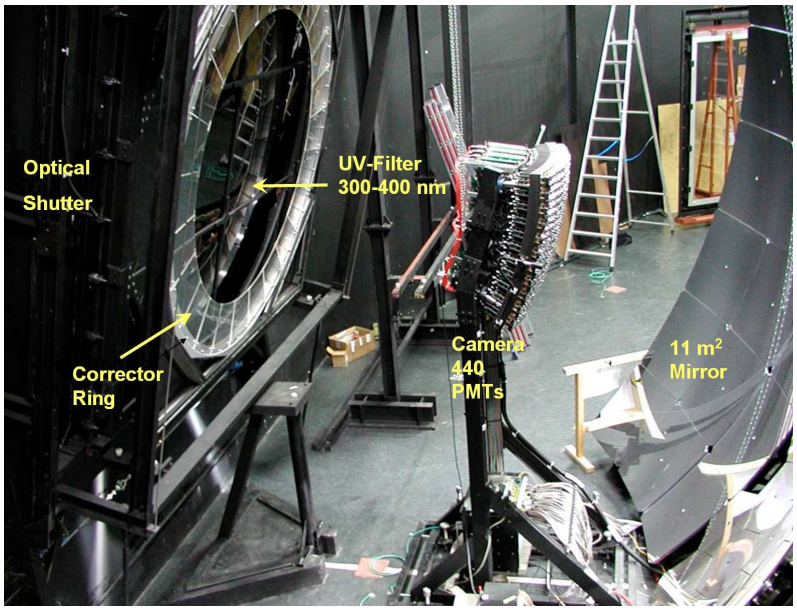


Figure 3.5: Optics inside a telescope. There are 440 PMT's detecting the light collected by the spherical mirror.

UV LED is housed in a small Teflon sphere and illuminates the interior of a drum that is 2.5 m and 1.25 m deep. The drum diffusively reflects the light onto the PMT's. There is a silicon detector mounted near the LED that measures the relative intensity of each flash. The geometry of the source and the drum is such that the intensity of the light is uniform throughout the camera. The uniformity of the intensity is measured with a CCD camera 15 m away in a laboratory setting, and the non-uniformities are less than 5%. The UV LED is of wavelength 375 nm, and the intensity and the duration of the pulse can be varied. The technique as described gives us relative calibration. We still need to know the number of photons from the output of the ADC from the pixels. In other words, we need absolute calibration. That is done by using a UV silicon detector that is absolutely calibrated by the National Institute of Standards and Technology (NIST). A PMT is, then, absolutely calibrated by comparing its output to that of

Figure 3.6: Picture of a drum mounted on the outside of the FD building for calibration.

the NIST silicon detector under identical conditions in a laboratory setting. Then the PMT is used to calibrate the drum, i.e. we now have a hard number for the photons in relation to the intensity of the drum (we cannot directly use the NIST Si detector, due to its low sensitivity, to calibrate the drum). When the drum illuminates the pixels in the camera, we can now relate the output of the pixels with a number of photons. The uncertainty in the overall calibration is dominated by the indirect calibration of the drum described above. The contributions to the uncertainty from the non-uniformities of the diffuse surface of the light source and the stability of the LED monitoring Si detector are minimal. The overall uncertainty in the drum calibration is 12% [28].

The drum calibration is done on a periodic basis but not on a nightly basis. Instead, the tracking of calibration is done on a nightly basis with a relative calibration system with a UV (470 nm) LED and Xenon flash bulbs with variable intensities/wavelengths placed at three different locations in a telescope. The light is fed to an optical fiber, and there is a diffuser at the end of the optical fiber to equalize the light going to the PMT's. There is a Si photodiode that records the intensity of the light. Using the relative calibration run within one hour after the absolute calibration measurement as reference, any nightly changes in the relative calibration is recorded and corrected for. The system stays stable within a few percent over a long term, and the uncertainty is only in the range of 1 to 3% [29].

3.2.2 Atmospheric Monitoring

WE need to be able to relate the observed light intensity, I, with the light intensity at the fluorescence source, I_0. In order to do that, we need to correct for the geometric factors and atmospheric transmission factors. The Cherenkov photons get scattered on the way to the telescopes. There are two relevant types of scattering, Rayleigh and Mie scattering. Rayleigh scattering is the scattering that takes place in a pure or molecular atmosphere. The properties of molecular scattering are well-known effects and can be easily dealt with precisely using the pressure and the temperature at the FD and the adiabatic model for the atmosphere. Mie scattering, on the other hand, is the scattering of light by the aerosol in the atmosphere, in the form of clouds, dust, smoke, and other pollutants. It is necessary to monitor the atmosphere to correct for the Rayleigh and Mie scattering.

At each eye, there is (or will be at completion of the PAO) a weather sta-

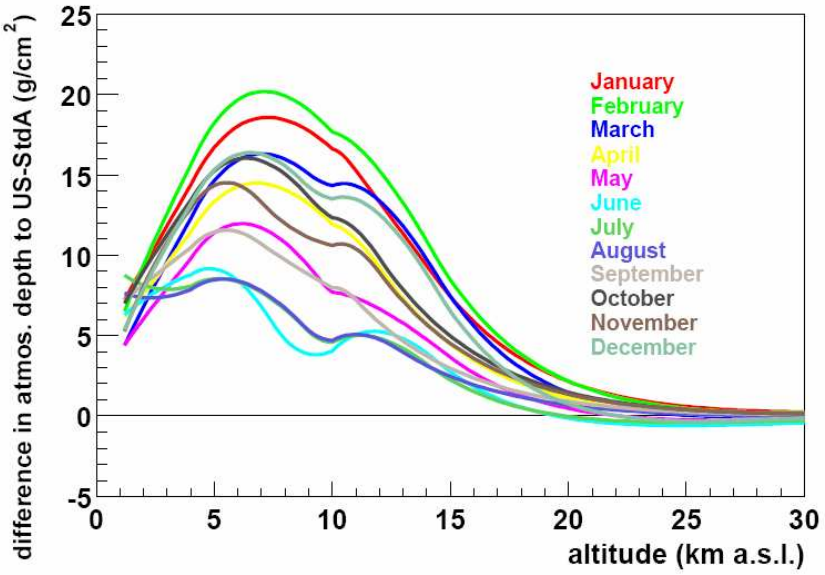


Figure 3.7: Profile of Atmospheric Depth at Malargüe. Figure from [30].

tion that monitors the local pressure and temperature, wind speed and direction and humidity. The temperature and the pressure information is important for Rayleigh scattering. Wind speed and direction are important for safe operation of the FD, i.e. if the wind is too strong, the shutters are closed to protect optical instruments.

It is important to know the atmospheric temperature and pressure profile as the fluorescence photon yield has a pressure and temperature dependence. So in addition to the ground-based pressure and temperature monitoring, the atmospheric pressure and temperature profile is measured by launching radiosondes intermittently. The radiosondes take data about every 20 m during ascent up to 25 km above sea level. More than 100 measurements have produced average monthly atmospheric profiles of the atmosphere at Malargüe (see Fig. 3.7) [30].

Each FD site will be equipped with a backscatter LIDAR to monitor the verti-

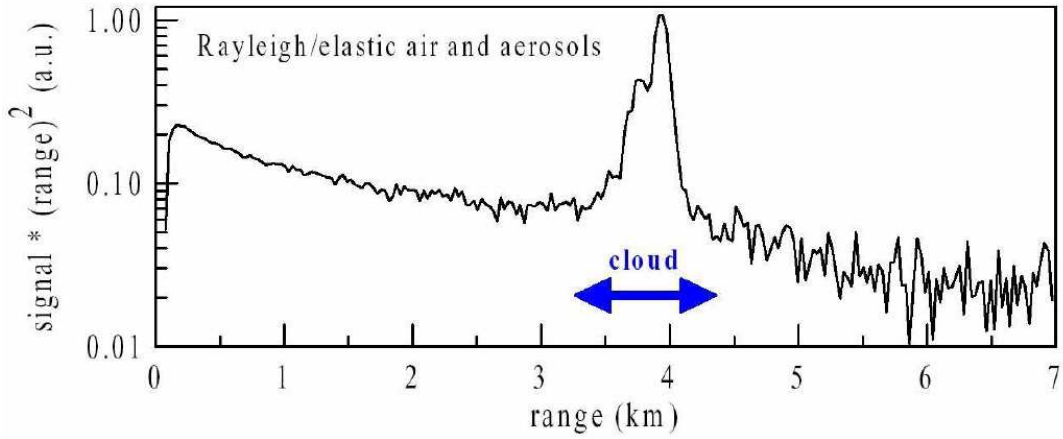


Figure 3.8: Example of LIDAR measurement. The spike at 4 km is the backscattered light in the presence of aerosols. Figure from [31].

cal aerosol profile (see Fig. 3.8). Currently, there are 2 LIDAR's operational. The LIDAR measures the vertical aerosol optical depth by detecting the backscatter light from pulsed UV lasers (351 nm) with PMT's. The Horizontal Attenuation Monitors (HAM's) measure the attenuation length between the eyes near ground level every 1 hour. Each HAM consists of a DC light source located at one eye and a receiver at another eye. The DC light source emit a broad spectrum including 300-400 nm. The light detector is a UV enhanced CCD array and monitors at 365, 404, 436, and 542 nm. These measurements determine the horizontal attenuation length at 365 nm, when combined with the local pressure and temperature, and its wavelength dependence.

The observed light from an EAS will contain both Cherenkov and fluorescence photons. The Cherenkov photons are strongly forward beamed but appears as background both directly and indirectly due to multiple scattering in the atmosphere. The aerosol phase function monitors (APFs) are designed to measure the differential scattering cross section $d\sigma/d\Omega$ of the aerosol [32]. The measurement is

done by firing a collimated beam of light from Xenon flash lamp across the front of an eye. The track generated from the light contains a wide range of scattering angles, thus allowing us to measure $d\sigma/d\Omega$. This measurement has been done hourly at Coihueco since September 2004.

Clouds have large optical depths and can dramatically affect scattering and transmission. To obtain a detailed skymap of cloud distributions, cloud cameras are used. The infrared cameras have a field of view of 46° x 35° and are on steerable mounts to cover the entire sky. They will provide each FD pixel with a cloud/cloud-free decision.

The Central Laser Facility (CLF) is located in the middle of the SD array. Every hour, the CLF shoots several hundred 355 nm laser pulses with a 7 ns width which can be steered to any part of the sky with an accuracy of 0.2°. For a vertical shot, the scattering is dominated by the well-known molecular scattering processes. The predicted intensity of the scattered light at each height is compared with the measured intensity to produce the vertical aerosol optical depth.

Because we need to accurately measure the fluorescence light from the shower particles traversing the atmosphere, we need to closely monitor the atmospheric conditions to correct for changing conditions which affects the transmission and scattering properties of the atmosphere. All these tools make close monitoring of the atmosphere possible.

CHAPTER 4

Surface Detector Calibration and Monitoring

4.1 Calibration

The shower particles that enter the stations emit Cherenkov photons , and these photons are in turn detected by the PMT's and leave a signal in the FADC. Until these signals are calibrated, these signals are just some arbitrary quantities. The calibration of the SD is done via atmospheric muons. More specifically, the calibration is done by measuring the charge deposit by a vertical and central throughgoing muon which should correspond to an energy deposit of 240 MeV in the 120 cm of water inside the tank. This quantity is referred to as the vertical equivalent muon (VEM). The particle density in the tank is measured in units of VEM by converting the signal left by shower particles in units of VEM. The remoteness and the independent standalone operation of the stations necessitate a self-contained calibration method that can be carried out at each station. The abundance of the atmospheric muons makes this possible. And the abundance (relatively high flux) of these muons also means a constant update and monitoring of the calibration is possible.

In reality, the atmospheric muons actually enters through all angles, so measuring Q_{VEM} (charge deposit by a vertical through-going muon) requires some effort. Fig. 4.1 shows a distribution of Q_{VEM} (red line) in units ADC channel that was measured by placing a scintillator above and below the tank each to

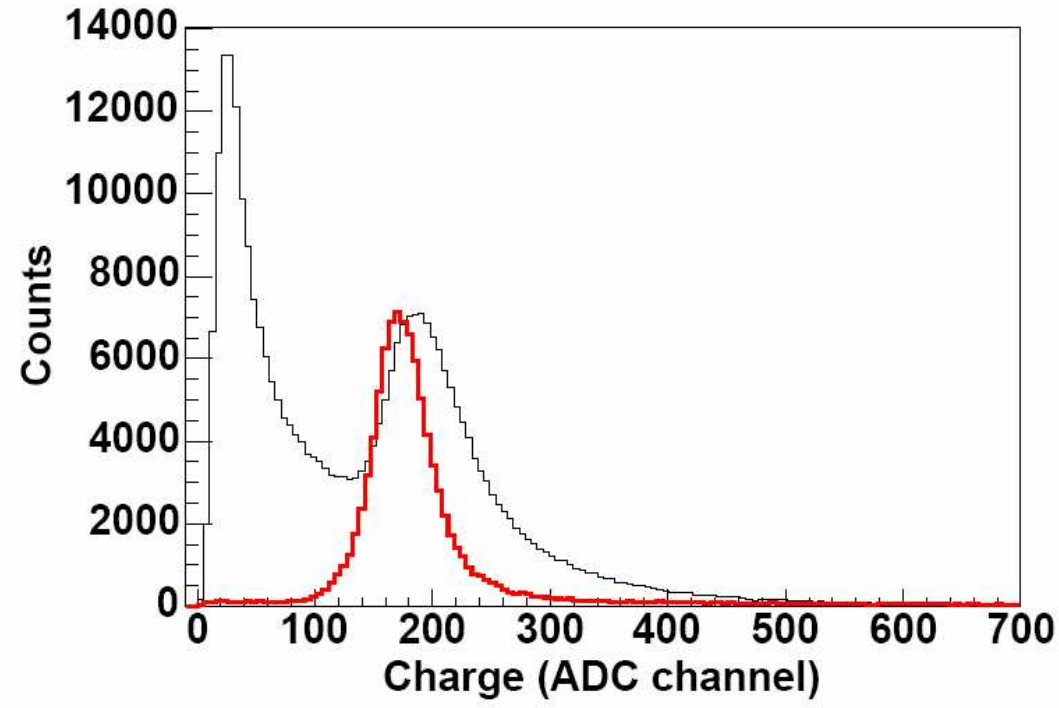


Figure 4.1: Muon charge histograms. The Red line is the charge distribution for the vertical throughgoing muon triggered by scintillators. The peak Q_{VEM}. The black line is charge histograms for all muons that triggered a 3-PMT coincidence. The second peak is Q_{VEM}^{Peak}. Figure from [33].

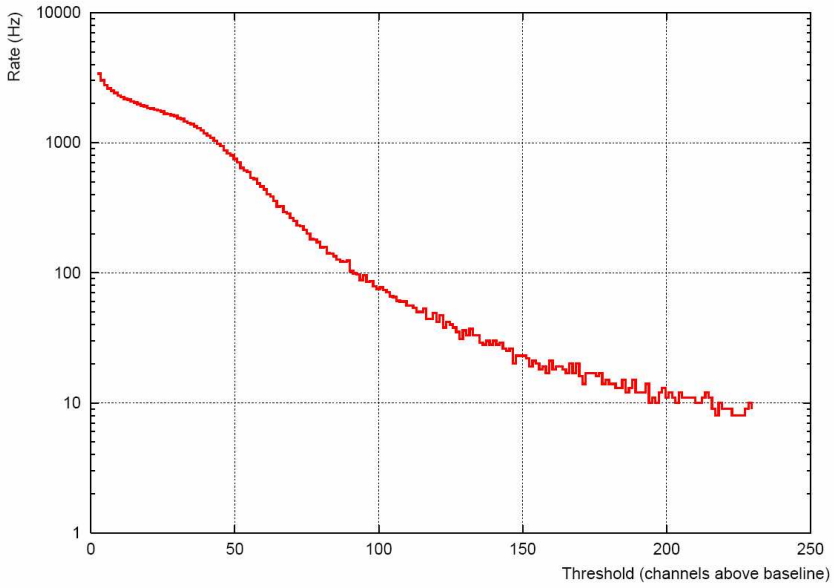


Figure 4.2: The atmospheric muon flux as seen by engineering array station Laura. Figure from [34].

trigger on vertical through going muons only [33]. The charge histogram for 3 PMT coincidences at a trigger level of 5 channels above the baseline is shown on the same plot as well. The first peak is a trigger artifact. The second peak is mostly due to the vertical through-going muons with the peak smeared out by non-vertical through-going muons and is close to Q_{VEM}. The second peak, Q_{VEM}^{Peak}, is at 1.05 VEM [33]. Since we do not have the luxury of having scintillators above and below the stations to trigger on vertical throughgoing muons only, we have to rely on the charge histogram of atmospheric muons to determine Q_{VEM}^{Peak} which in turn gives us Q_{VEM}.

Fig. 4.2 shows the cosmic ray spectrum as measured by a station [34]. The engineering array (prototype surface detector) local station hardware and software originally had a constraint of low trigger rate of 100 Hz. The trigger levels used for the calibration was, thus, chosen to be on the order of 100 Hz. The

calibration method is a trigger rate based one. The reasoning behind this is that the atmospheric muon flux should be more or less the same throughout the array, and the PMT's inside a tank should also see the same rate. To ensure uniformity across the array, a common trigger threshold is desired. The reference unit for the threshold level is I_{VEM}^{peak}, the second peak in a maximum pulse height histogram. I_{VEM}^{peak} is related to the peak photocurrent and has the same conversion factor as Q_{VEM}^{Peak} to Q_{VEM}. Moreover, I_{VEM}^{peak} and Q_{VEM}^{Peak} has a roughly constant ratio of 3.4, which means a trigger level set using I_{VEM}^{peak} is in effect in units Q_{VEM}^{Peak} — the relation between I and Q can be seen in Fig. 4.5 where I is peak of the muon pulse and Q is the area of the pulse. Setting the trigger level in units of I_{VEM}^{peak} frees us from having to integrate the FADC trace to set the trigger level in terms of Q_{VEM}^{Peak} which would require too many CPU clock cycles.

The target value of I_{VEM}^{peak} for the VEM calibration procedure is chosen to be 50 ADC channels above the baseline. A trigger rate of a single PMT at 150 ADC channels above the baseline was chosen as the target calibration point for all tanks. From a study with a reference tank [34], the 100 Hz singles trigger rate roughly corresponds to roughly $3I_{VEM}^{peak}$. This satisfies the target value of 50 ch/I_{VEM}^{peak}.

The initial end-to-end gain setup procedure is as follows. All 3 PMT's in the tank are required to meet the 100 Hz trigger rate at 150 channels above the baseline when the local station electronics are turned on which balances the PMT's to within 10% initially [35]. Since the end-to-end gain will drift inevitably due to temperature responses and other factors, there is little need for accuracy in this step.

Once the gains of the 3 PMT's are set up, the high voltage is unchanged. The drift in the end-to-end gain is compensated for by changing the trigger level

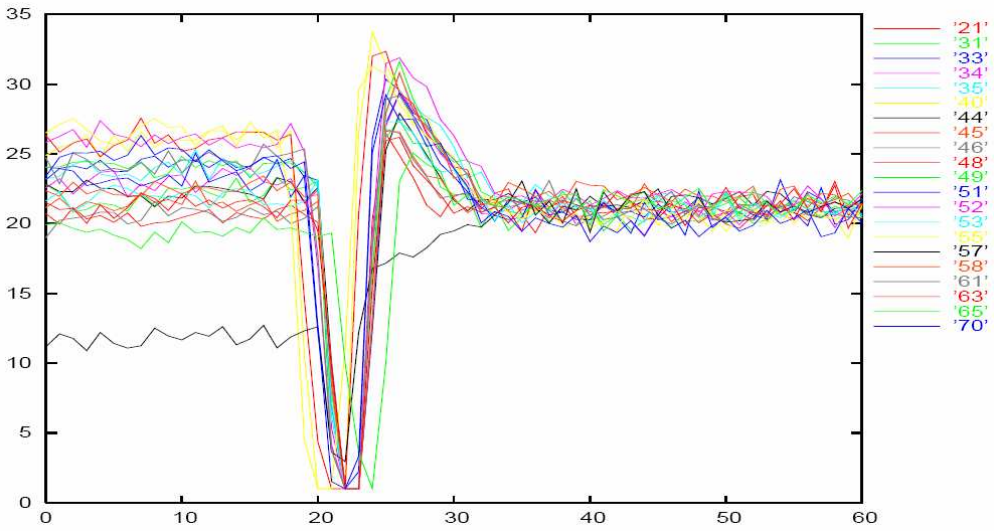


Figure 4.3: Convergence of the trigger rate at 3.2 I_{VEM}^{peak} to 20 Hz after stations are reconfigured with the convergence algorithm based on the singles rate of 70 Hz at 2.5 I_{VEM}^{est}. Figure from [34].

of the PMT's. This means that the dynamic range will not be the same for all stations. But the drift in the end-to-end gain is minimal. The actual average value of I_{VEM}^{peak} is 46 ± 4 ch [35].

The online calibration is updated continuously. For the continuous online calibration, 3-fold coincidence trigger of the 3 PMT's is required with a trigger level of 1.75 I_{VEM}^{peak} — the calibration begins with an estimate of I_{VEM}^{peak} at 50 ch, denoted I_{VEM}^{est}. This 3-fold coincidence trigger should result in a trigger rate of 100 Hz. Within those 100 Hz triggers, events with threshold above 2.5 I_{VEM}^{est} are recorded for each PMT. These events are counted for $t_{cal} = 5\ s$, and the rate should be 70 Hz. The online calibration is adjusted by using a σ–δ convergence algorithm, with $\sigma = 2$ Hz and $\delta = 1$ ch. For every σ greater than 70 Hz for the singles trigger rate, the threshold is increased by one δ, and vice versa. This pro-

cess is repeated and the singles trigger rate converges to 70 Hz. This convergence algorithm works well, as in Fig. 4.3 it is clear that when stations are reconfigured with the convergence algorithm, the T1 trigger rates at 3.2 I^{peak}_{VEM} settle nicely at 21 Hz [34]. Q^{est}_{VEM} from the online calibration is quite close to Q_{VEM} — the ratio is 0.96 ± 0.03 [35].

The online calibration takes place every minute and sent to the CDAS every 6 minutes for monitoring [36]. Quantities from sensors mounted such as temperature, PMT voltage, current, *etc.* and other quantities such as dynode/anode ratio, and baseline values for the 6 FADC in each tank are available for monitoring as well. There is a separate trigger with a low threshold, 0.1 I^{est}_{VEM}, to collect high statistics histograms ($\sim 150,000$ Entries) of charge distribution of responses to muons (see Fig. 4.1). These histograms are sent with every candidate shower event.

4.2 Dynode/Anode Ratio and Linearity

The VEM is determined to an accuracy of 5%. But what about larger signals? For instance, an inclined muon will have a longer track length inside the tank and will leave a larger signal, hence. Is the signal, hence energy deposit, for inclined muons simply scales linearly with the track length? This is important as there should an asymmetry in the Cherenkov light seen by the 3 PMT's for the inclined muons (other particles for that matter) since some PMT's may see direct light while other PMT's will see the diffuse light from bouncing off the liner. As can be seen Fig. 4.4, when the signals in the 3 PMT's are averaged, the signals simply scale linearly with the track length. This is true for all angles out to the zenith angle of 60°, but beyond that angle the asymmetry is too large that the average signal does not scale linearly with the track length. The asymmetry seen

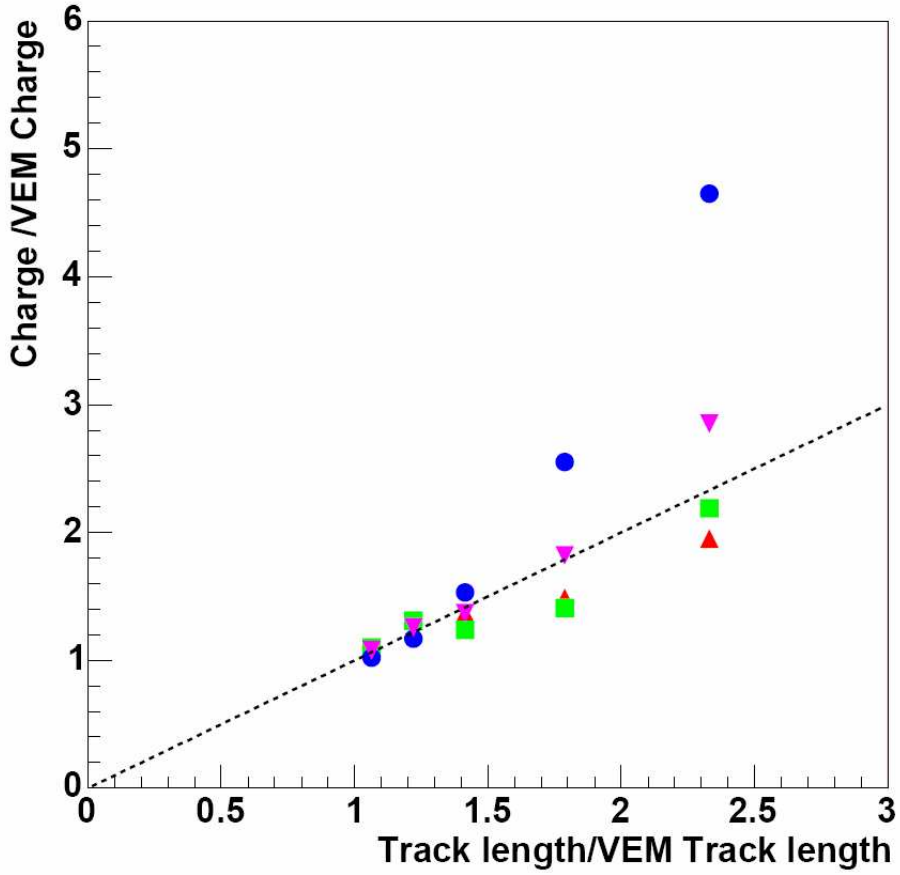


Figure 4.4: Charge deposit as function of angle of incidence (track length). The groups of points are 5 different zenith angles (20°, 35°, 45°, 56°, and 70°). The upside down triangle is the average charge deposit for the PMT's, and other points are the charge deposit seen by each PMT. Even though the asymmetry in the charge measured by the 3 PMT's grows larger with the zenith angle, the average charge deposit scales linearly with the track length out to 56°. Beyond 60°, the average charge is not linear. Figure from [33].

in Fig. 4.4 is actually due to the location of the scintillators for triggering where the scintillators are closer to one PMT than the other. That the average of the 3 PMT signals do not scale linearly for the 70° case means that is most likely due to direct light seen by the PMT with the largest signals.

As mentioned earlier, there are two gain settings, the dynode and anode. The dynode gain is nominally 32 times higher than that of the anode. This setup extends the dynamic range of the SD. Often, We find that the dynode is saturated by a moderate signal size since the FADC only goes to 1024 channels — one VEM is roughly 50 channels. In that case, the dynode signal is not accurate, so we have turn to the anode to measure large signals. The dynode channel is well calibrated with the muon data as we have seen before. Ideally, the anode channel calibration would be simply 1/32 that of the dynode. Unfortunately, the dynode/anode (D/A) gain ratio depends on several factors such as the high voltage, amplifier, filters, individual tube characteristics, *etc.* The D/A ratio not only varies from PMT to PMT, but it also fluctuates. An accurate measurement and tracking of the D/A ratio is crucial as an uncertainty the D/A ratio would result in an uncertainty in energy for events with large signals.

Unfortunately, the D/A ratio is not simply a ratio of peak of the anode and dynode signal as the amplification stages add a small delay to the dynode signal. The relation between the anode and the dynode signals are modeled as:

$$f(x) = \frac{1}{R}((1-\epsilon)d(x) + \epsilon d(x+1)) \tag{4.1}$$

where R is the D/A ratio, $d(x)$ is the dynode signal in bin x, and ϵ is the fractional offset in the dynode [37]. Fig. 4.5 shows a fit of the anode to the dynode. About 100 samples of $f(x)$ and $d(x)$ are taken every 3 minutes and averaged. The fit is performed by the CDAS to determine the D/A ratio, R. The D/A ratio is determined to an accuracy of 5%.

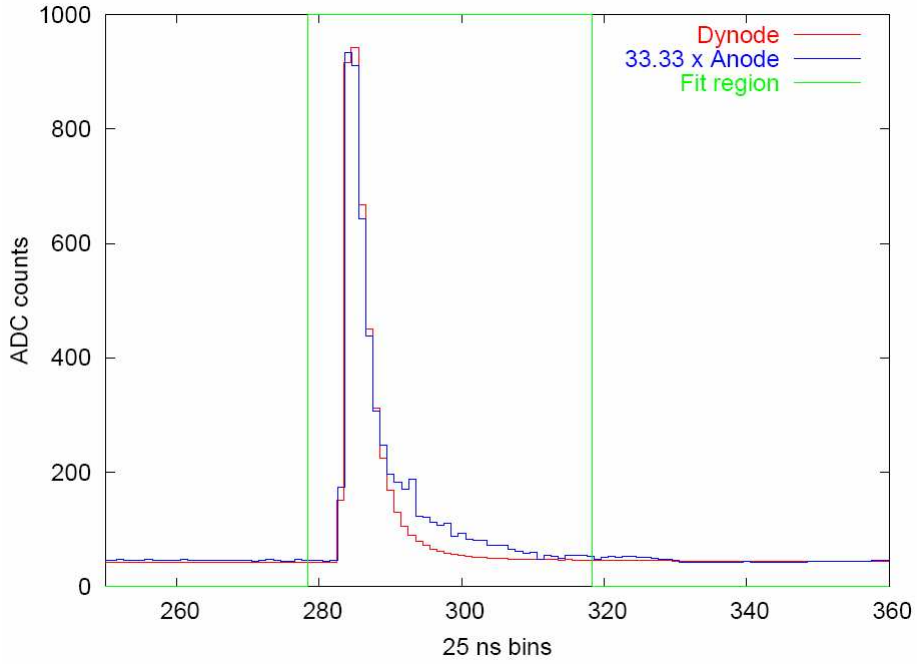


Figure 4.5: The result of a fit of the anode to the dynode. Figure from [37].

Non-linearity of the detector can be checked with 2 LED's (A and B) mounted in the center of the liner bag. The idea is to measure the PMT output from pulses from A and B separately and A and B simultaneously. Letting $f(A)$ be the integrated charge measured by the PMT for pulse A, we see that if the detector was linear, then $f(A+B)=f(A)+f(B)$. Indeed the detector is linear within 5%.

4.3 monitoring

The calibration method for the SD is carried out continually on a station to station basis without human intervention. The SD could in principle run on its own. It is necessary to monitor the detector performance as not only do we need to know how the detector is performing, but invariably there will be problems with the detector, and they need to be identified and rectified.

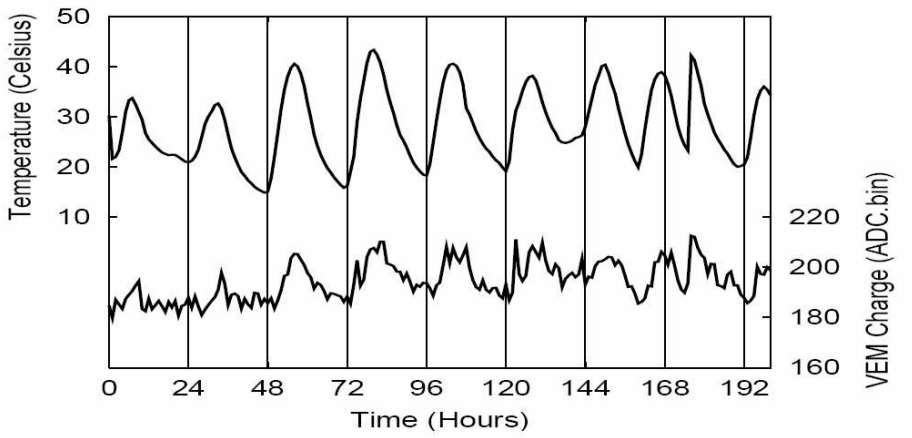


Figure 4.6: The charge deposit for VEM (bottom) and temperature (top) over time. The two are correlated. Figure from [36].

As mentioned earlier, the monitoring data files contains an extensive list of quantities such as HV, temperature, baseline, VEM charge, *etc.* recorded continually and sent to the CDAS every 6 minutes and allows us to closely monitor the performance of the tanks. What we are most interest in is the quantities that directly affect the physics results, namely the VEM charge, D/A ratio, anode and dynode baselines, and VEM area/peak ratio. Because the detector condition changes over time, the detector response will change, and we need to monitor these quantities.

The charge deposit for a VEM changes over time and is actually correlated with temperature (see Fig. 4.6). This figure illustrates the need for a frequent update of calibration. If there were only sparse updates of the VEM charge calibration, say once every 12 hours, then clearly it could lead to a large error when converting the signal in the tank to its equivalent VEM value in between calibration updates. The fluctuation of the VEM is monitored to ensure they are relatively stable over time, and for the period of November and December of

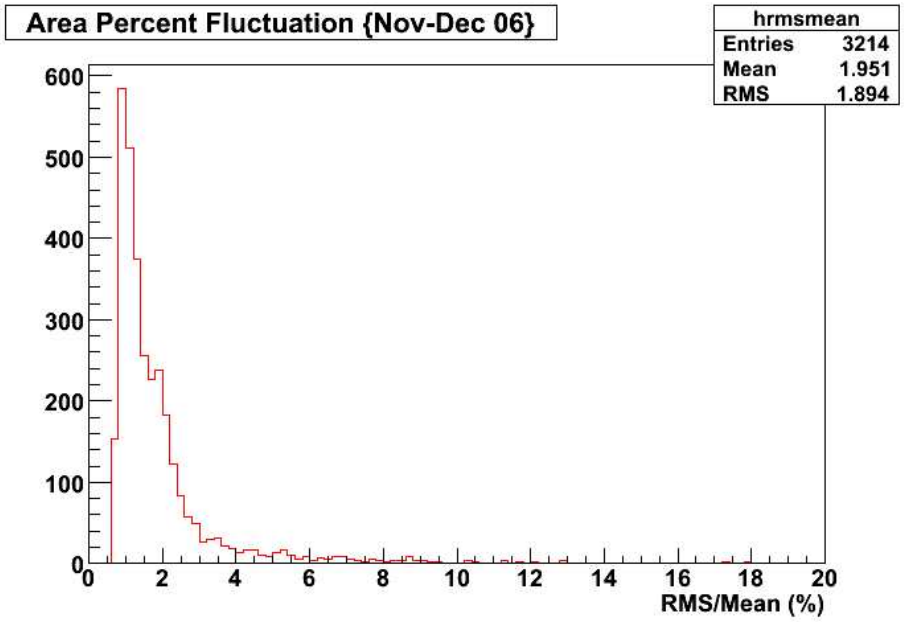


Figure 4.7: The distribution of fluctuations in VEM charge deposit (area) for Nov-Dec 06.

2006, the distribution of the fluctuation for that period for all the PMT's out in the field is shown in Fig. 4.7. Vast majority of the PMT's have a VEM charge fluctuation of less than 5% with the mean at 2%. Even in the cases where the fluctuations are large, the calibration update happens frequently enough that VEM charge determination should be pretty accurate.

The anode and dynode baseline are also monitored as large fluctuations in the baselines (the part of the FADC trace before and after the muon pulse in Fig. 4.5) could potentially affect the accuracy of determination of the charge deposit as we have to subtract the baseline from the signal to get the charge deposit. The baseline fluctuation in the period is miniscule as the fluctuation in both the dynode and anode baselines are all less than 1% (see Fig. 4.8 and Fig. 4.9).

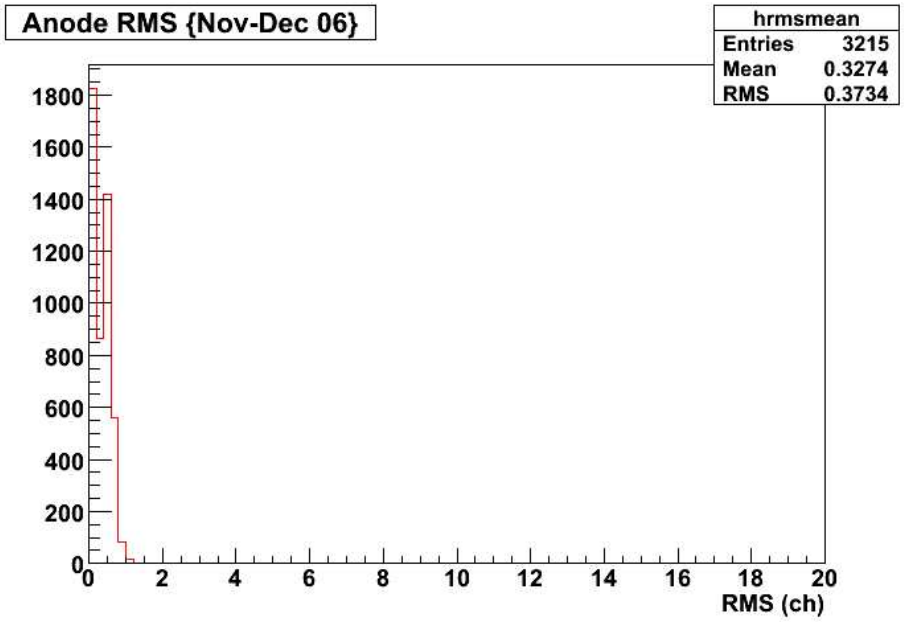


Figure 4.8: The distribution of fluctuations in anode baseline for Nov-Dec 06.

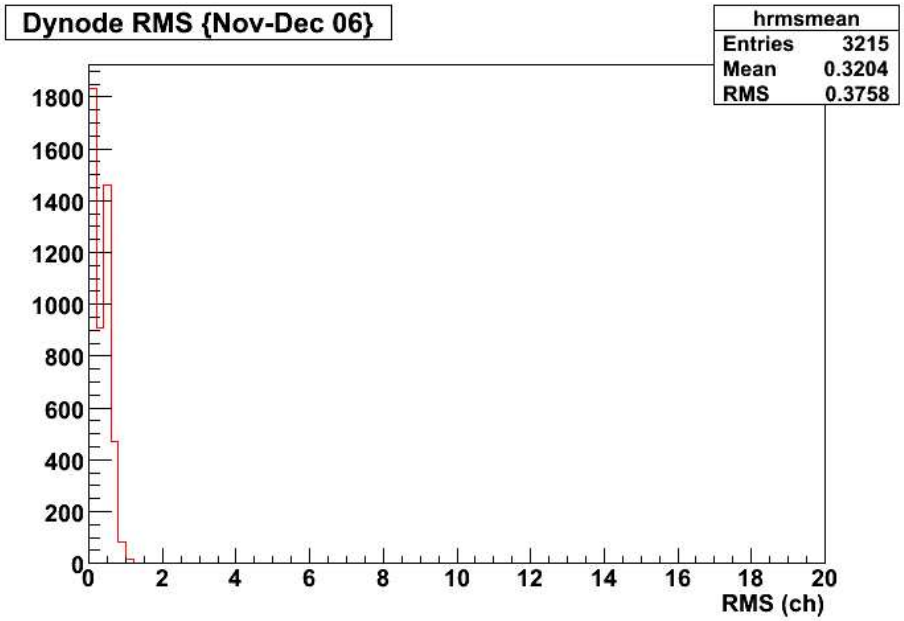


Figure 4.9: The distribution of fluctuations in dynode baseline for Nov-Dec 06.

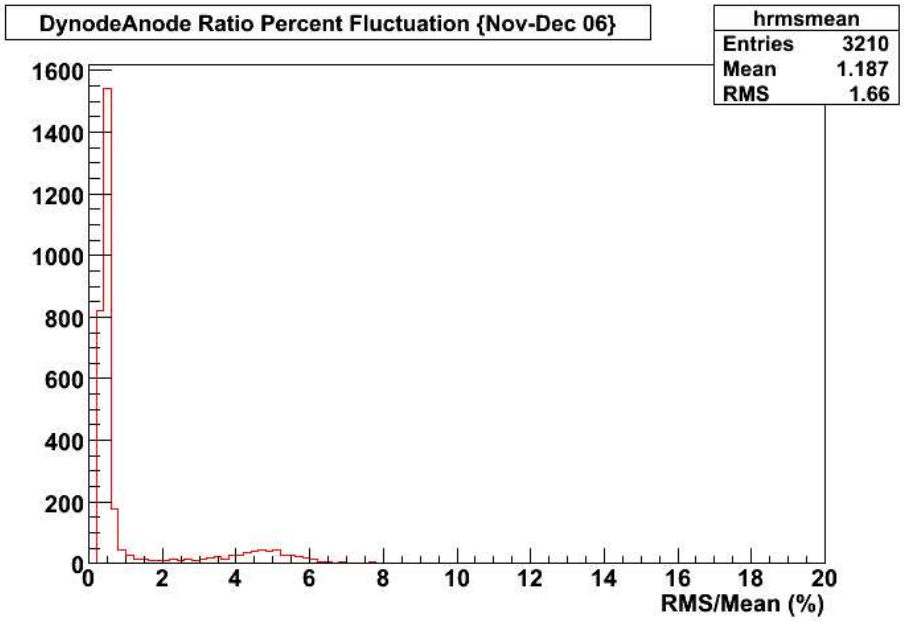


Figure 4.10: The distribution of fluctuations in D/A ratio for Nov-Dec 06.

The D/A ratio affects the accuracy of large signals. The fluctuation of the D/A ratio is also well under control as the mean fluctuation is 1% and almost all are within 5% (see Fig. 4.10).

The quality of water inside the tank must also be monitored as that affects the signal as well. Fig. 4.11 shows a typical muon FADC trace with an exponential fit to the tail to determine the decay constant. The decay constants are typically ~ 65 ns. As the water quality worsens after deployment of stations in the field, more diffuse light gets absorbed in the water leading to a faster decay time in the muon signal. The decay constant is therefore a good indicator of the water quality. The VEM Area to peak (Area/Peak) ratio is the ratio between the total area of the muon pulse to the peak of the pulse (see Fig. 4.11). This is also a good indicator of the water quality as it is correlated with the decay constant as shown in Fig. 4.12. The indication so far is that some stations have an initial

period where the Area/Peak ratio decreases followed by stabilization and then modulation with the seasons, while others are stable from the beginning and only show seasonal modulations (see Fig. 4.13 and Fig. 4.14). All the indications point to water quality worsening to the point where the the decay constant falling to 0 will not happen — a decay constant of 0 means all the diffuse light gets absorbed in the water meaning no signal from shower particles. The water quality will be stable for the 20 years of operation. As far as the fluctuations of the Area/Peak ratios for the 2 month period are concerned, the mean fluctuation is less than 2% and most are well within 5%. As seen from above, the calibration and the stability of the detector is well under control.

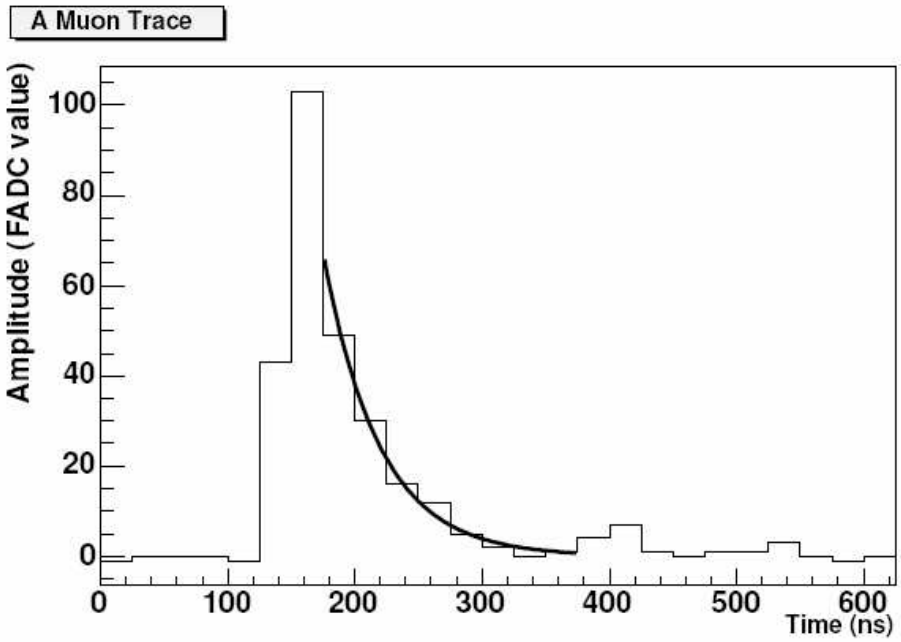


Figure 4.11: Typical muon FADC trace with an exponential decay fit to the tail. A typical decay constant is ∼65 ns. VEM charge is the area under the curve, and VEM peak is the peak of the pulse. Baseline is the channel value of the FADC trace when there is no signal (∼ 0 in this figure, but usually it is around 50). Figure from [53].

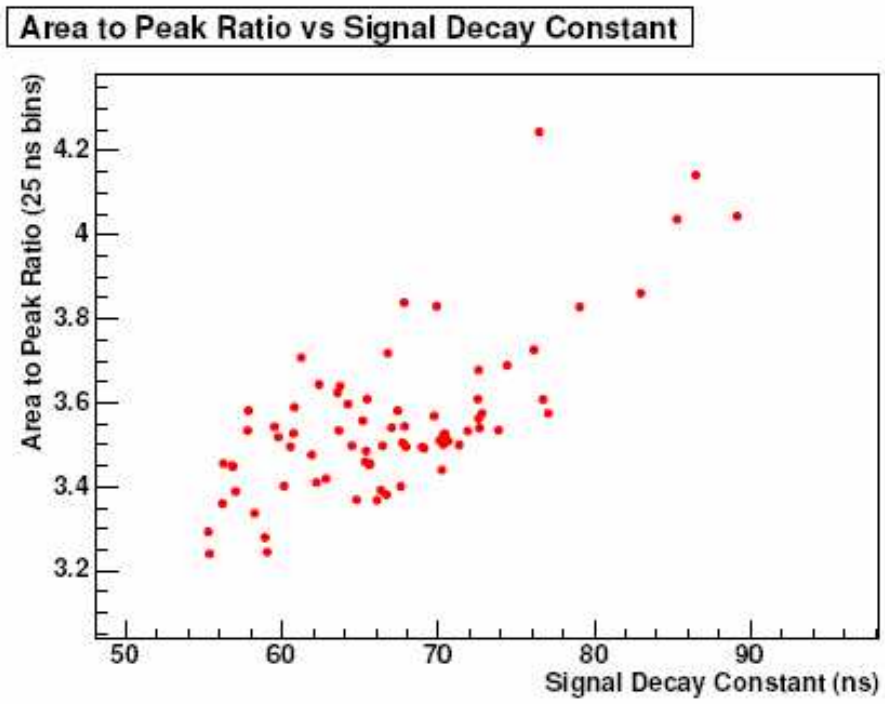


Figure 4.12: The correlation between the decay constant and the Area/Peak ratio. Figure from [38].

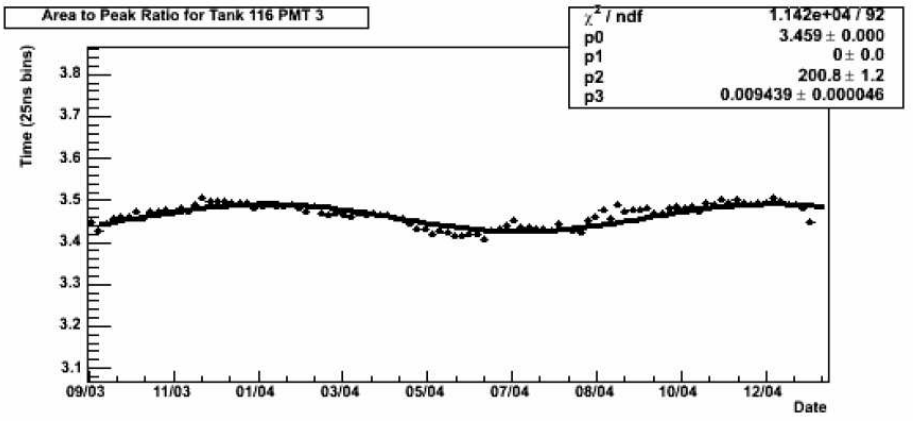


Figure 4.13: The evolution of the Area/Peak ratio over time. There is a seasonal modulation in sinusoidal form. Figure from [39].

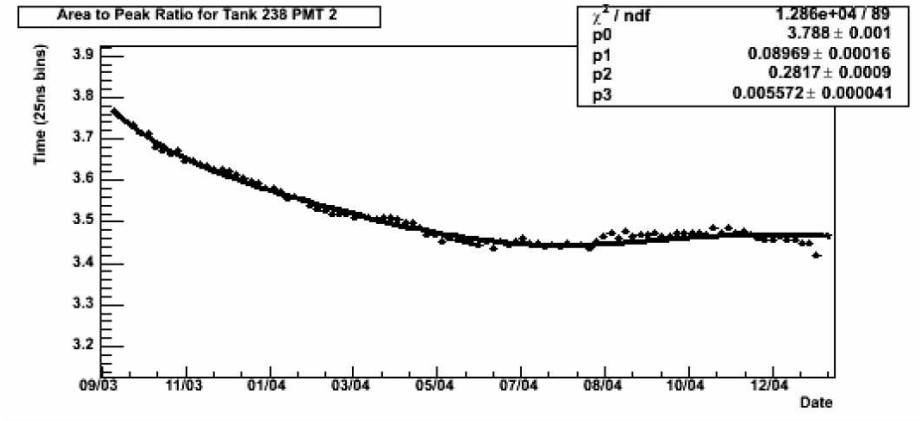


Figure 4.14: The evolution of the Area/Peak ratio over time. There is an initial decay followed by a seasonal modulation in sinusoidal form. Figure from [39].

CHAPTER 5

Reconstruction

When a shower is detected by the PAO, the detector records an event with a geometrical, timing, and signal pattern on the tanks (for the SD) and pixels (for the FD). We use the information to reconstruct events to determine the direction vector and the energy as well as other information.

5.1 FD Reconstruction

Fig. 5.1 is an illustration of Cherenkov and fluorescence light flux at the detector induced by a shower, and Fig. 5.2 shows an event display of a shower as seen by the FD. As can be seen, the EAS is seen by the triggered pixels. We need to reconstruct the event using the information recorded by the detector. The first step is to do the geometric reconstruction. The shower profile reconstruction requires information such as the direction vector and the distance to the shower from the geometric reconstruction the geometric reconstruction, so it is deferred until the completion of the geometric reconstruction.

The first step in the geometric reconstruction is to find the shower-detector-plane (SDP) which contains the triggered pixels and the shower axis (see Fig. 5.3). The SDP is found by minimizing the sum of the dot product of the normal vector to the SPD and the pointing direction of the pixels weighted by the signal at each pixel.

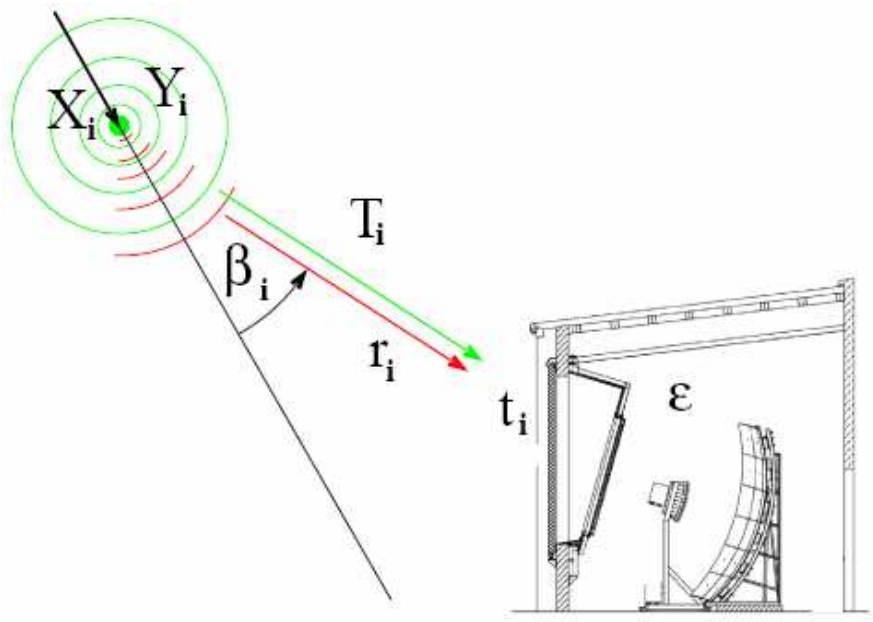


Figure 5.1: Illustration of the Cherenkov (red) and fluorescence light (green) reaching the FD. Figure from [40].

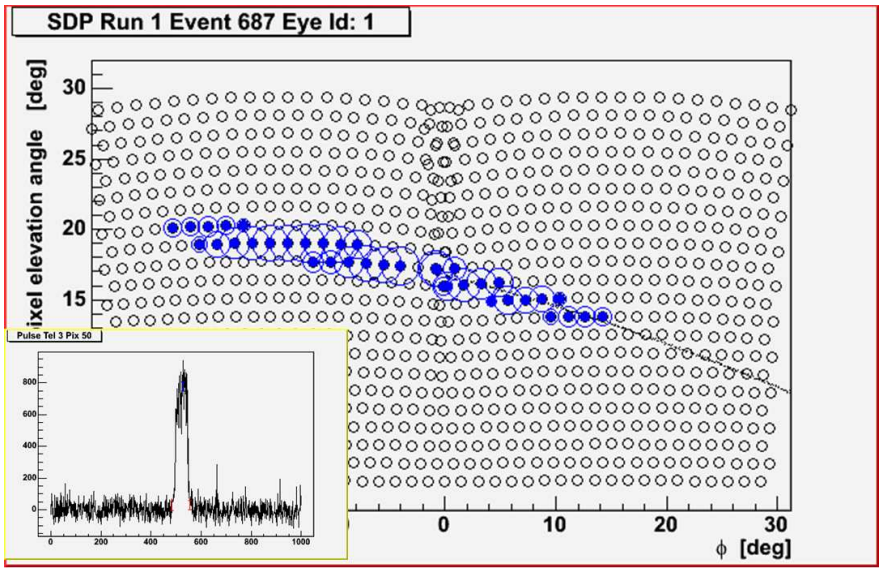


Figure 5.2: An event as seen by the FD. The blue dots are the triggered pixels. The inset is an ADC trace from a triggered pixel.

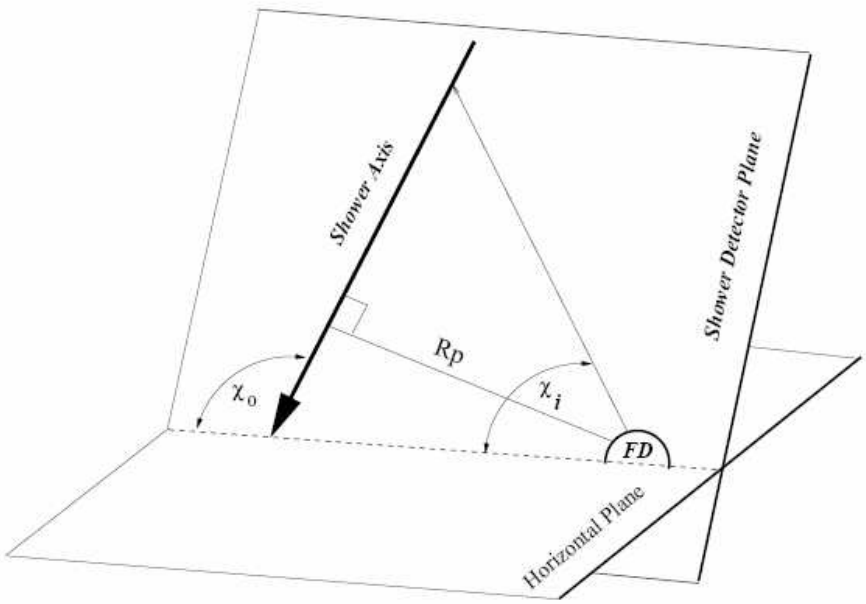


Figure 5.3: Shower-detector-plane containing the shower axis and the triggered pixels, and other relevant geometries. Figure from [41].

Once the SDP is found, the shower axis within the SDP is determined next. The expected flight time of the photons within the SDP to the i-th triggered pixel is given by

$$t_i = T_0 + \frac{R_p}{c} \tan\left(\frac{\chi_0 - \chi_i}{2}\right) \tag{5.1}$$

where T_0 is the time at which the shower front passes closest to the eye, at distance R_p (refer to Fig. 5.3). Then the shower geometrical parameters found by minimizing

$$\chi^2 = \sum_i^N \frac{(t_i - t_i^{obs})}{\sigma^2} \tag{5.2}$$

where t_i^{obs} is the observed trigger time at i-th pixel and σ is uncertainty in the time measurement. In hybrid events, the geometrical reconstruction can be improved by using the timing information at the tank to further constrain the shower axis within the SDP.

Once the shower geometry have been determined, we have the necessary information to proceed with the shower profile reconstruction, i.e. having calculated R_p and the shower axis, we have the necessary geometry information to figure out the slant depth, X_i, along the shower profile. The profile reconstruction relies on the fact that the electromagnetic component dominates far outstrips the number of the muons. Over an interval of slant depth, $\triangle X_i$, the number of photons emitted is

$$N_\gamma^F = Y_i^F dE/dX_i \triangle X_i \tag{5.3}$$

where dE/dX_i is the energy deposited at slant depth X_i. The fluorescence photon yield, Y_i^F, is proportional to dE/dX [42, 43]. The fluorescence photons are emitted isotropically over $4\pi r_i^2$ where r_i is the distance to the detector. And because of the Rayleigh and Mie scattering, only a fraction, T, of the photons reach the detector. Given a detector efficiency, ε, the fluorescence photon flux y_i^F at the FD is then

$$y_i^F = d_i N_\gamma^F \tag{5.4}$$

where $d_i = \frac{\varepsilon T_i}{4\pi r_i^2}$.

The light flux at the detector is contaminated by the Cherenkov photons. The number of the Cherenkov photons is proportional to the number of charged particles above the cutoff energy for Cherenkov photon production. The number of Cherenkov photons, N_γ^C, emitted is

$$N_\gamma^C = Y_i^C N_e(X_i) \triangle X_i \tag{5.5}$$

where Y_i^C is the Cherenkov light yield, and $N_e(X_i)$ is the number of electrons and positrons above the cutoff energy (Remember, electrons and positrons dominate the number of charged particle). Then the Cherenkov light flux at the detector is

$$y_i^C = d_i f_c(\beta_i) N_\gamma^C \tag{5.6}$$

where $f_c(\beta_i)$ is the fraction of Cherenkov light emitted at an angle β_i (see Fig. 5.1). Now, the total energy deposit, dE/dX_i, is

$$dE/dX_i = N_e(X_i) \int_0^\infty f_e(E, X_i) dE/dX_e(E, X_i) dE \tag{5.7}$$

where $f_e(E, X_i)$ is the normalized electron energy distribution. The electron energy distribution $f_e(E, X_i)$ is universal in shower age which allows us to simplify dE/dX_i to $dE/dX_i = N_e(X_i)\alpha_i$, where α_i the average energy deposit per electron given a shower age $s_i = 3/(1 + 2X_{max}/X_i)$ [44, 45]. Fig. 5.4 shows the energy deposit per electron as a function of the shower age. For a range of shower age of interest to us, $0.8 < s < 1.2$, the shower to shower fluctuation is small.

Now, We have

$$dE/dX_i = \frac{y_i}{d_i \triangle X_i} (Y_i^C f_c(\beta_i)/\alpha_i)^{-1} \tag{5.8}$$

and we can determine dE/dX directly from the light flux at the detector, y_i. The situation, in reality, is a bit more complicated by presence of the scattered Cherenkov light which we have ignored. The wavelength dependence of light production and transmission have also been ignored. But they are accounted for in the reconstruction.

Now, the detector does not see the entire shower profile. The so-called Gaisser-Hillas function describes the longitudinal development well, and it is fit to the profile observed by the detector and allows us to extract information about the shower profile outside the range of the detector. It is

$$f_{gh}(X) = \ dE/dX_{max} \left(\frac{X - X_0}{X_{max} - X_0} \right)^{(\frac{X_{max} - X_0}{\lambda})} \exp\left(\frac{X_{max} - X_0}{\lambda} \right) \tag{5.9}$$

where X_0 and λ are interpreted as the initial interaction point and the absorption length, respectively [46]. Fig. 5.5 shows an example of light flux at the detector and the reconstructed profile along with the Gaisser-Hillas fit to the profile. Now

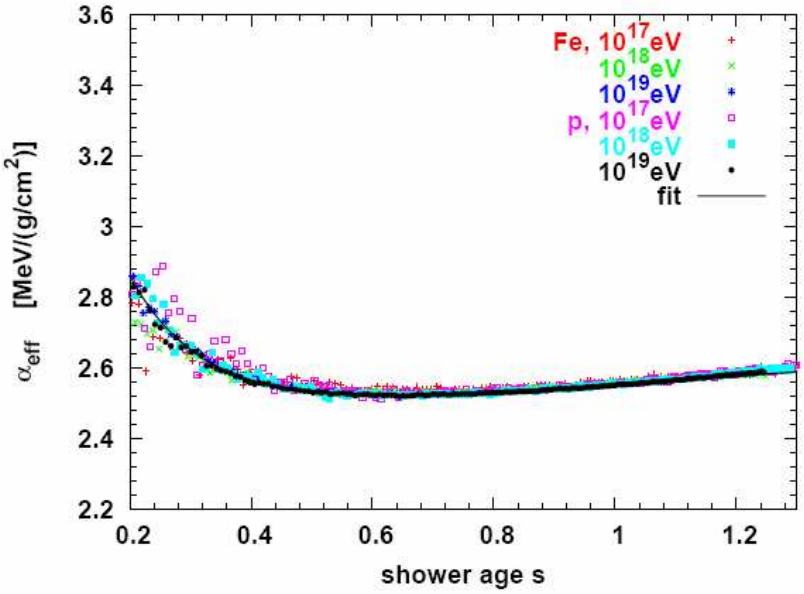


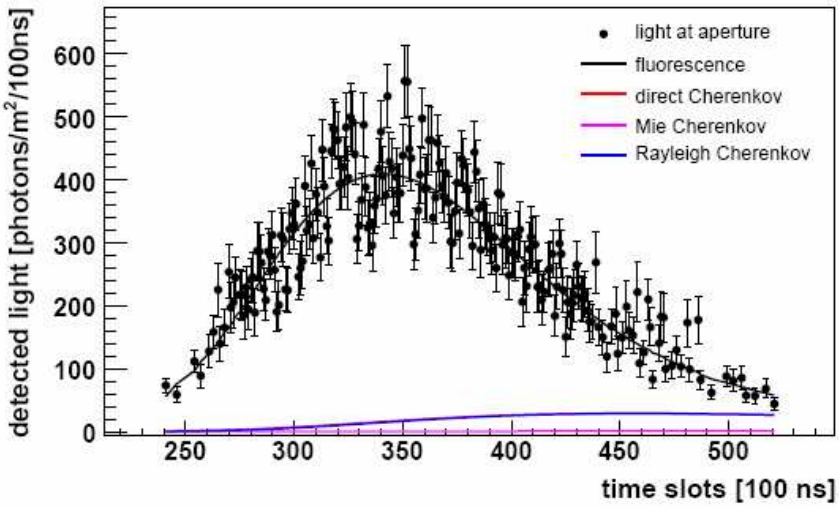
Figure 5.4: Average energy deposit per electron as a function of shower age as given by Corsika. Figure from [44].

the total energy deposit can be determined by integrating the Gaisser-Hillas function,

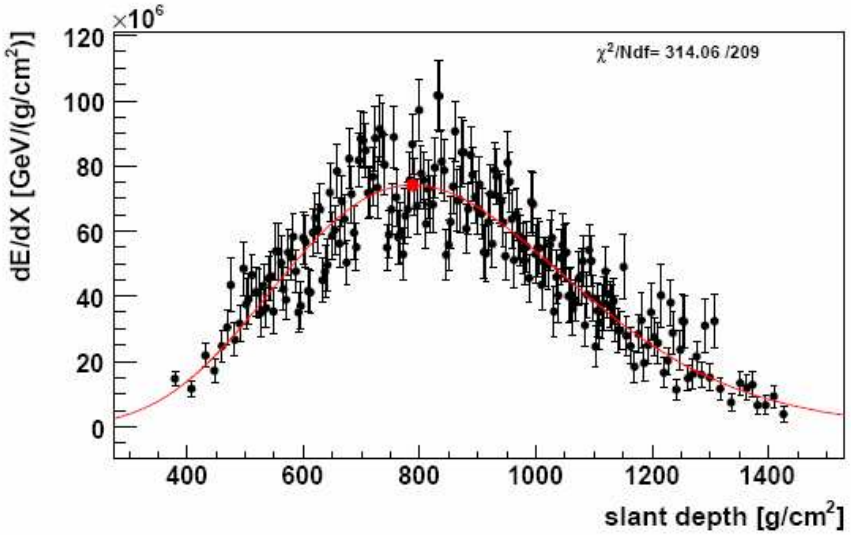
$$E_{dep} = \int f_{gh}(X)dX \tag{5.10}$$

Finally, the invisible energy must be accounted for. This is done by multiplying E_{dep} with a correction factor, f_{inv} — $E_{tot} = f_{inv}E_{dep}$. f_{inv} is parameterized as $f_{inv} = (a + b \cdot E_{dep}^{c})^{-1}$, where a, b, and c are constants depending on the primary and hadronic interaction model [47].

The reconstruction method described above was formulated by M. Unger [40]. There are other algorithms as well, but the general idea is the same. They involve reconstructing the shower profile from the photon flux at the detector and using Gaisser-Hillas function to infer the shower profile at all depths [48, 49].



(b) *Light flux at Los Leones.*



(d) *Energy deposit at Los Leones.*

Figure 5.5: Reconstructed shower profile along with Gaisser-Hillas fit. Top figure shows the fraction of direct (red) and scattered (purple) Cherenkov light for this event as well. Figure from [40].

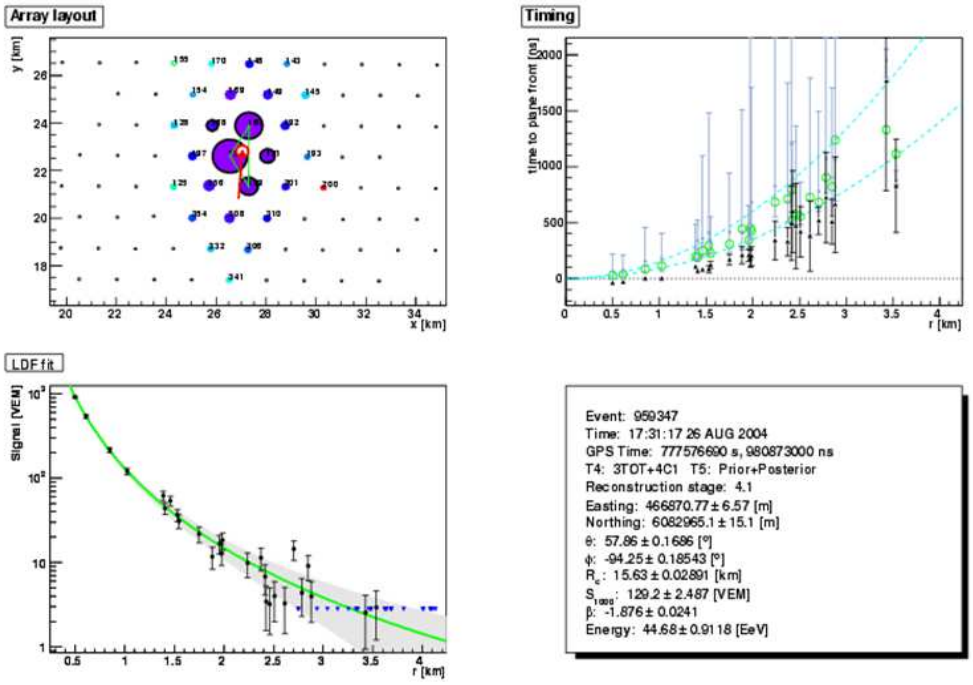


Figure 5.6: SD event display. Top left plot shows SD array with each dot representing tank. Colored circles triggered tanks with size of circles indicating signal size. Bottom left shows an LDF fit for this particular event.

5.2 SD Reconstruction

Fig. 5.6 shows an event display of an SD event. From the timing and signal information from each tank, we need to reconstruct an event to determine the direction vector and the energy. SD reconstruction can be divided into two parts—angular reconstruction and lateral distribution reconstruction. We need to perform the angular reconstruction first as the direction vector is required for determining the lateral distribution.

The shower front is assumed to be a plane, to begin with. If we let $u = \sin\theta\cos\phi$, $v = \sin\theta\sin\phi$, and $w = \cos\theta$ where θ is the zenith angle and ϕ is the

azimuth angle, then the expected arrival time at each tank i is given, for a planar shower front, by

$$t_i^{th} = T_0 - (ux_i + vy_i + zw_i)/c \tag{5.11}$$

where T_0 is the arrival time on the ground at the shower core. T_0, u, and v can be determined by minimizing the following

$$\chi^2 = \sum_i (t_i^{meas} - t_i^{th})/\sigma_i^2 \tag{5.12}$$

where σ_i is the uncertainty in time.

Strictly speaking, the shower front is actually not planar. Simulations indicate that the shower front is actually very close to a parabolic shape and is modeled as either a parabolic or spherical shape [50, 51]. Given a radius of curvature, R, the shift in the longitudinal position of the shower front from the planar front at a distance d from (and perpendicular to) the axis is $d^2/2R$ [50]. Letting $1/2R = \gamma$, the expected shower front arrival time for a curved front is then

$$\begin{aligned} t_i^{th} &= T_0 - (ux_i + vy_i + zw_i)/c + \gamma d_i^2/c \\ &= T_0 - (ux_i + vy_i + zw_i)/c + \gamma(x_i^2 + y_i^2 + z_i^2 - (ux_i + vy_i + zw_i)^2)/c \end{aligned} \tag{5.13}$$

There are four free parameters in the curvature—(u, v, T_0, γ). At least four tanks are required to do a curvature fit.

As mentioned earlier, in the SD we measure the lateral distribution to determine the energy of the shower. The lateral distribution changes as a function of energy—higher the energy the higher the signal at the tanks. The lateral distribution changes as function of the zenith angle as well, as the shower age increases with the zenith angle which means more shower particles dissipate resulting in smaller signals at the tank. The lateral distribution is described well by the lateral distribution function (LDF). There are several forms of LDF, but a commonly used LDF is the NKG function [52]. The NKG function is parameterized

as a function of the distance from the core in the perpendicular plane to the axis as

$$S(r) = S_{1000} \left(\frac{r}{r_{1000}} \right)^{-\beta} \left(\frac{r + r_{700}}{r_{1000} + r_{700}} \right)^{-\beta} \tag{5.14}$$

where S_{1000} is the expected signal at 1000 m from the core, and $r_{700} = 700\ m$ and $r_{1000} = 1000\ m$, respectively. β is the slope of the LDF given as $\beta = 3.3 - 0.9 \sec\theta$ — this is the angular dependence of the LDF. The LDF is fitted to the signal at each tank, with distance to the core projected to the perpendicular plane, using a maximum likelihood function (see Fig. 5.6) [51]. S_{1000} is closely correlated with the energy and thus a good indicator of the energy. S_{1000} is converted to energy using an energy converter (see next Chapter). S_{1000} is actually not an absolutely defined quantity in that for different LDF's the value of S_{1000} is actually different even at the same energy and zenith angle. But what matters is the final conversion to energy, and each LDF has its own energy converter. In the end, the systematic effect on energy is due to the choice of LDF only $\sim$5%.

CHAPTER 6

Energy Determination and Systematic Uncertainties

Given a signal from a shower at the detector we need to be able to convert that signal to the energy of the primary. The atmosphere acts as a calorimeter, and the shower particles deposit their energies in passing through the atmosphere. After a few generations, the shower physics boils down to low energy interactions of the electromagnetic component of the shower. Thus the determination of energy reduces to measuring the energy deposit along the shower track (see Ch. 5), and the influence of the uncertainties due to the unknown primary and hadronic interaction models should be minimal.

Since the FD tracks the shower development process, the FD energy measurement is calorimetric. Energy measurement from the SD is not calorimetric in the sense that the SD catches the shower at the tail end of the development process, thus the energy deposit occurs only for the thin slice out of the entire shower development process that are detected by the SD (see Fig. 6.1). Thus, the SD is a differential energy detector.

To determine the energy, the energy deposit, more specifically signals in units of VEM, in the SD must be somehow be related to the total shower energy without the benefit of having the information on the manner in which the shower develops like with the FD. Unfortunately shower development can be quite complicated

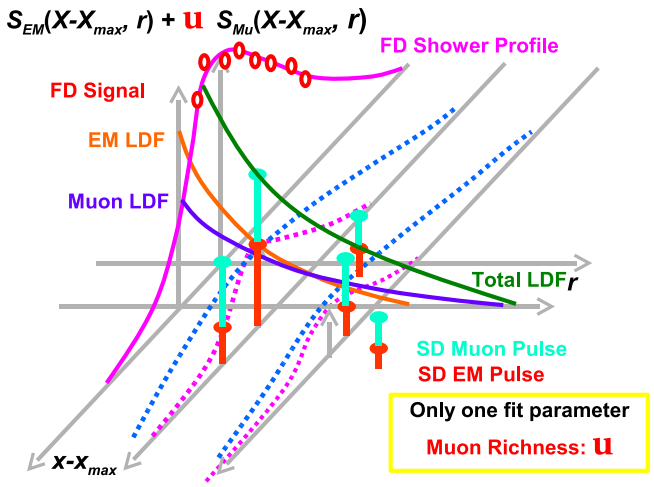


Figure 6.1: 3-Dimensional Shower Development showing the principle of the FD and SD energy measurement. The FD measures the shower development at the core. The SD measures the LDF, hence the energy deposited by particles in a slice perpendicular to the direction of shower development. X is the slant depth of the ground level given by $X = X_{vert} \cdot \sec(\theta)$, where X_{vert} is the shower depth for the vertical shower. X_{max} has no angular dependence. Plot courtesy of K. Arisaka.

and depends on a range of variables, some related to the mass of the primary, and some on the interaction of the cosmic ray with the atmosphere above the detector. The goal of energy conversion is to relate the energy deposited at the ground to the initial input energy by accounting for those effects.

6.1 Theoretical Considerations and the Origin of S(1000)

Although the lateral distribution is correlated with energy, due to shower to shower fluctuations, the lateral distribution also fluctuates from shower to shower. And also the sparseness of the SD has an influence on the LDF fit. Thus we

need to devise a way to minimize these effects. S(1000), the interpolated signal at 1000 m from the core from the LDF fit, turns out to be the quantity that minimizes the aforementioned effects, hence is the best estimator of the energy of the shower.

6.1.1 Intrinsic Statistical Uncertainties

The X_{max} fluctuates on a shower by shower basis even for the same primary at the same energy. From Fig. 6.1, it is clear that a fluctuation in the location of the X_{max} means $X - X_{max}$, where X is the slant depth at the ground, will also fluctuate resulting in a fluctuation of the lateral distribution. Since the SD does not 'see' the X_{max}, there is no easy way to correct for the effects due to the fluctuation of X_{max} and will result in a fluctuation of S(1000) (along with other factors such as the sparseness of the stations and the reconstruction resolution).

6.1.2 Intrinsic Systematic Uncertainties

The X_{max} has a dependence on the mass of the primary — proton showers are deeply penetrating while the X_{max} for iron showers are shallow. The muon content of the showers also has a compositional dependence. As discussed earlier, iron showers contain more muons than proton showers. Fig. 6.2 shows the compositional dependence of the muon content and the X_{max}. The water cerenkov tank greatly enhances the signals from muons. This combined with the systematic difference in the X_{max} of protons and irons means that even proton and iron primaries of the same energy will give different S(1000). This means there is degeneracy in converting S(1000) to energy, i.e. one could covert to different energies with different primary assumptions for the same S(1000). Since we do not know the composition, we are forced to make an assumption on the composition,

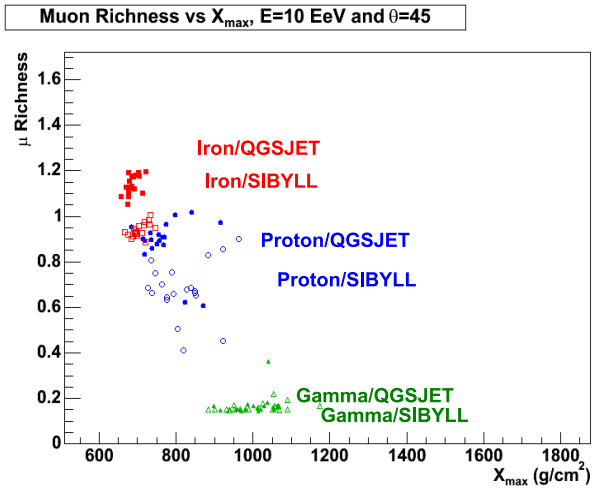


Figure 6.2: Muon Richness versus X_{max}: Proton (blue), Iron (red), and Gamma (green) have a Systematic Difference in X_{max} and Muon Richness ($\frac{N_\mu}{N_{em}}$ at respective X_{max}'s). The Hadronic Interaction Model Assumed also Affects these Quantities. Plot courtesy of D. Barnhill.

thus resulting in a systematic uncertainty in energy determination.

6.2 Parameterized Monte Carlo (MC) Energy Conversion

One way to obtain a conversion to the primary cosmic ray energy is through the use of a library of simulated showers combined with a realistic detector simulation [53]. The ultimate goal of this technique is the creation of a function that models the response of the surface detector based on the observables, S(1000) and zenith angle, to arrive at the *input four-vector* of the primary cosmic ray (E, p).

6.3 Shower Library and Fitting Procedure

A library of simulated showers was created with a thinning level of 10^{-6}. The relative weight factor for electro-magnetic particles to hadrons was eighty-eight for Aires, whereas Corsika used a maximum weight factor of one hundred for hadrons and ten thousand for electro-magnetic particles. The shower library covered the following combinations, each combination containing about 20 showers.

- Aires[54], Corsika[55]
- Proton, Iron
- QGSJet[56], QGSJet II[57], Sibyll[58], Fluka[59], Gheisha[60]
- Zeniths, θ – $0°, 25°, 36°, 45°, 53°, 60°, 66°, 72°, 84°$
- Azimuths, ϕ – $0°, 36°, 72°, 108°, 144°, 180°, 216°, 252°, 288°, 324°$
- Energy 3, 10, 31, 100 EeV

Due to computer limitations, the Aires (proton/iron and QGSJet/Sibyll) combinations were used as the standard set. The core locations of the simulated showers were randomly assigned to lie in a triangle of adjacent stations and the detector response was simulated with the offline software assuming a perfect auger detector (all stations are identical and function flawlessly) [61]. The same reconstruction algorithm, LDFFinder, is used for both the real data and the simulated data [51]. This reconstruction used a fixed beta regardless of the saturation status of the event or the number of stations.

The S(1000) value obtained from the reconstruction of 10EeV showers is averaged in each zenith bin and then plotted as a function of the $\sin^2(\theta)$. That

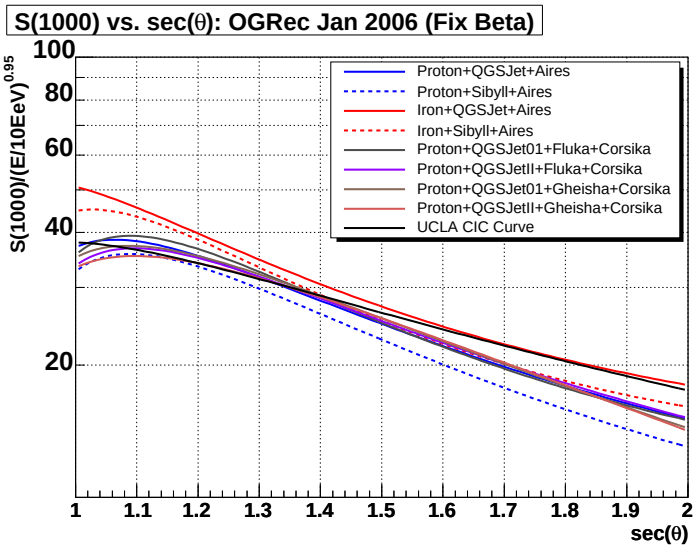


Figure 6.3: S(1000) Angular Dependence: Black is the Constant Intensity Cut Curve (Normalized here to Proton QGSJet); Blue is Proton (Solid - QGSJet, Dashed - Sibyll); Red is Iron (Solid - QGSJet, Dashed - Sibyll); Gray is Proton (QGSJet01, Fluka, and Corsika); Purple is Proton (QGSJetII, Fluka, and Corsika); Brown is Proton (QGSJet01, Gheisha, and Corsika); Gold is Proton (QGSJetII, Gheisha, and Corsika). Plot courtesy of M. Healy.

plot is then fit with a third degree polynomial in the range of zero to sixty-six degrees with equal weighting given to each zenith bin, giving us parameterizations of S(1000) at 10 EeV, denoted as $S(1000)(10\text{ EeV}, \theta)$. This procedure is repeated for the different compositions and models. The resultant parameterizations $S(1000)(10\text{ EeV}, \theta)$ for different composition and model are shown in Fig. 6.3.

As discussed earlier, S(1000) for the iron primary (red lines) is clearly larger than that for the proton primary. The shape of the S(1000) attenuation curves are different as well. The S(1000) attenuation curve for the proton has a bump

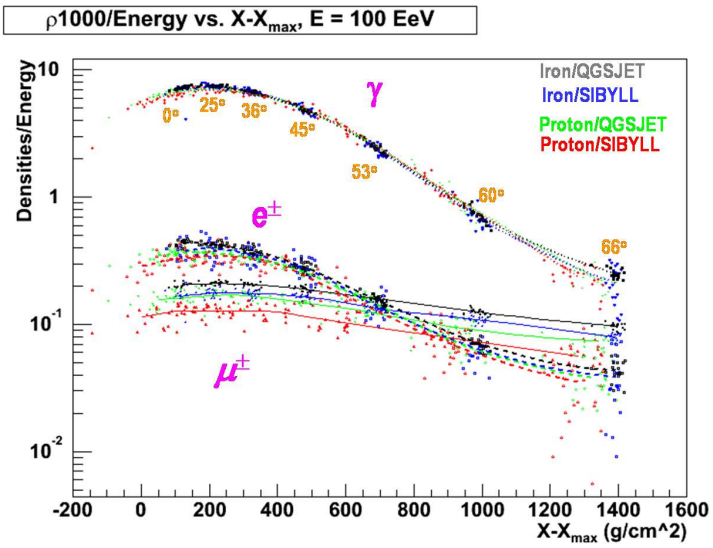


Figure 6.4: Particle density at 1000 m from the core given by simulation of shower development. The EM part of the shower development process is independent of the composition and only depends on $X - X_{max}$. For protons, the X_{max} is actually below the ground at $\theta = 0°$. Plot courtesy of D. Barnhill.

around $\sec(\theta) = 1.1$ (or $\theta \sim 25°$), whereas the S(1000) curve for the iron does not have such a feature. This is due to the fact that at 1000 m from the core, the X_{max} for the proton is actually below the ground for $\theta = 0°$ at Malargüe (see Fig. 6.4). It is only around 25° that the slant depth X becomes deep enough for the X_{max} to be above the ground. The X_{max} for iron, on the other hand, is shallow enough that it is above the ground for all zenith angles — hence, the difference.

Fig. 6.3 tells us the exact expected S(1000) values at all zenith angles at 10 EeV. We also need to figure out the S(1000) values at all energies. The energy dependence is found by using the showers from all energies and averaging the S(1000) values for a single zenith angle and energy and then fitting them as a

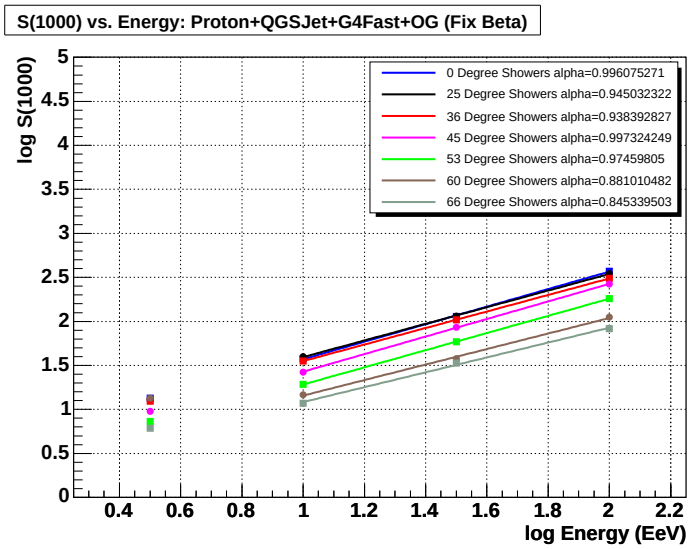


Figure 6.5: S(1000) Energy Dependence for Various Zenith Angles; 0.95 is the Value Taken for the Exponent of the Power Law (3 EeV showers are not used). Plot courtesy of M. Healy.

function of the input energy (see Fig. 6.5). When this is done the relationship is found to be logarithmic especially if the 3 EeV events, which generally have inferior reconstructions, are neglected. The assumption that the energy and angular dependence are separable can be checked by doing this for several angles and observing whether the exponent obtained depends significantly on the zenith angle. We take the energy dependence of S(1000) to be $S(1000) \sim E^{0.95}$.

A plot of S(1000) at all zenith angles and energies normalized to 10 EeV assuming $S(1000) \sim E^{0.95}$ is shown in Fig. 6.6. The S(1000) values agree with one another well within the error bar for all angles less than 60°. S(1000) for $\theta > 60°$ is attenuated significantly and dominated by muons as the EM part of the shower dissipates for the most part and the shower is not described by the LDF quite as well. Thus, reconstruction accuracy is poor $\theta > 60°$, and the

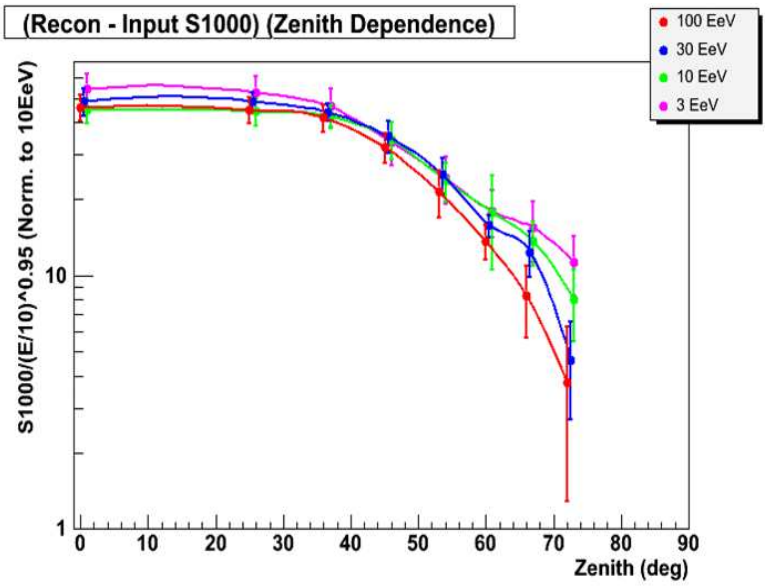


Figure 6.6: S(1000) at different energies normalized to 10 EeV assuming S(1000) grows as $S(1000) \sim E^{0.95}$.

spectrum analysis is carried out only for events with $\theta < 60°$.

In the final step of the parameterization the energy and angular function described above are combined to yield a conversion to energy based on the individual shower observables S(1000) and zenith angle. For each primary and hadronic interaction model combination, an energy converter is parameterized as follows,

$$E = \left(\frac{S(1000)(E,\theta)}{S(1000)(10 \text{ EeV},\theta)} \right)^{1/0.95} \tag{6.1}$$

where $S(1000)(10 \text{ EeV},\theta)$ is S(1000) at θ as given by the S(1000) attenuation curve in Fig. 6.3 for a given primary and hadronic interaction model.

The statistical uncertainties in S(1000), or equivalently energy, as well as S(r) at other distances (as predicted by the proton QGSJet combination) at different energies are shown in Fig. 6.7. The statistical uncertainties is smallest for S(1000) confirming our choice of S(1000) as the energy estimator is well justified. That

S(1000) has the smallest fluctuations is mostly due to the geometry of the SD array [62]. The closest tank contains the most information on energy (it has the highest signals as well as more EM component). The tank spacing 1.5 km. S(600) is susceptible to large fluctuations as there is a good chance there is no tank closer than 600 m which means one has to extrapolate S(600) from the LDF fit. S(1200) and beyond are too far from the core and suffer from too much muon contamination. For S(1000), we are likely to have a nearest tank within 1 km of the core and several tanks further than 1 km from the core, so it is an interpolated quantity and is the most stable quantity. The statistical uncertainties in S(1000) (energy) is around 15% at 3 EeV and improves to $\sim$10% for $E > 30$ EeV. The statistical uncertainties has an angular dependence as well, especially at low energies. The uncertainties are largest for the nearly vertical cases, below 25°, and inclined cases, above 60°. The large uncertainties for the nearly vertical case stems mainly from the fact that the tank multiplicity (number of tanks with signals) in a triggered event is small, so the sampling of the shower is sparse and the reconstruction accuracy is not as good as the moderately inclined case. The inclined case suffers from severe attrition of the EM component of the shower and is susceptible to the fluctuations in the relatively small number of muons hence the large statistical uncertainties.

The systematic uncertainties in energy due to unknown composition is shown Fig. 6.8. Different primaries give systematically different S(1000) even at the same energy (see Fig. 6.3). Conversely, given an S(1000) value, it could be converted to different energies depending on the primary assumption one chooses. Since we do not know the composition of the primary, a systematic uncertainty in energy arises as a result. Thus we take the extreme cases, proton and iron, and assign the difference in the energy between the two primary assumptions as the systematic uncertainties in energy due to unknown composition. Except for the

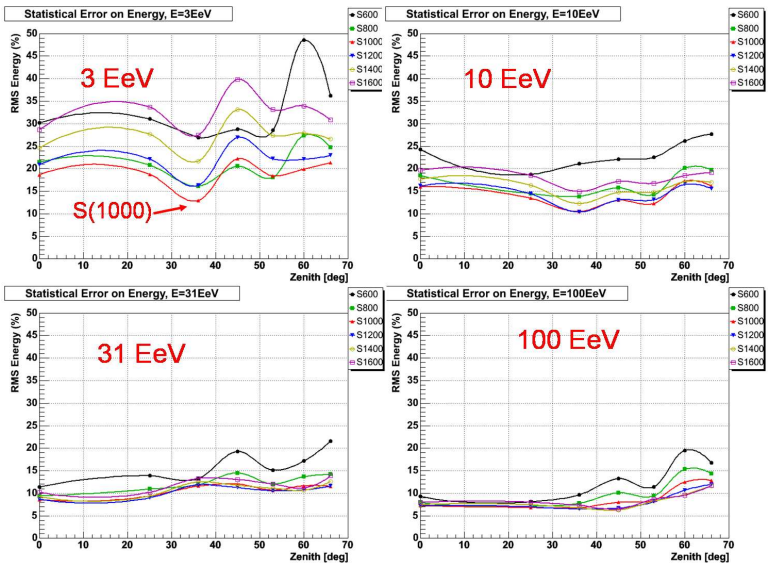


Figure 6.7: Statistical uncertainties in energy (S1000). The uncertainties become smaller with increasing energy. Plot courtesy of D. Barnhill.

3 EeV case where the reconstruction accuracy is marginal and the multiplicity of the events is small, the systematic uncertainties are actually smallest for S(600). This is expected as the closer one goes to the core the more EM component one picks up, and the EM component is nearly universal for all nuclei. To minimize the systematic uncertainties in energy due to unknown composition, one should measure the energy as close to the core as possible to pick up the EM component and minimize the influence of the muonic component. But given the sparse spacing of the SD, S(600) is not as well determined as S(1000). If the tanks were more closely spaced and the detector had a larger dynamic range, S(600) would be a better choice as the energy estimator. Another salient feature of Fig. 6.8 is that the systematic uncertainties have an angular dependence. The systematic uncertainties are larger for the nearly vertical and the highly inclined cases ($\theta > 60°$). For nearly vertical showers, the X_{max} for proton is below the ground whereas the X_{max} for iron is above the ground. This results in a large difference in

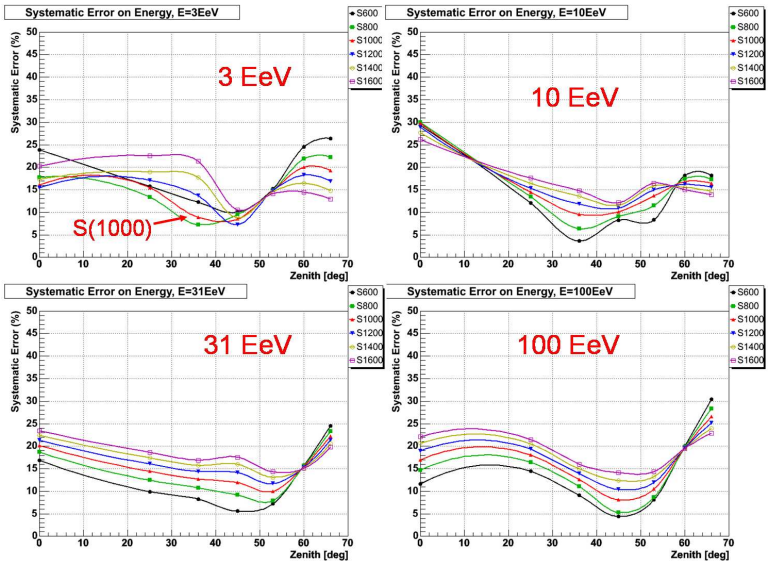


Figure 6.8: Systematic uncertainties in energy (S1000) due to unknown composition. The systematic uncertainty is defined as the difference energy between proton and iron assumptions. Plot courtesy of D. Barnhill.

S(1000). For highly inclined showers, muons start to dominate which have large compositional dependences. The systematic uncertainties are the smallest around 35-45°. For this reason we will often use the S(1000) at 38° at 10 EeV, or S38, of different hadronic interaction models/primary as well as different methods (see below) for comparison and absolute energy calibration. If we had large statistics, we could even restrict the data sample to 30-45° to minimize the systematic uncertainties due to unknown composition.

6.4 Constant Intensity Cut Energy Conversion

There is an alternative to relying on Monte Carlo (MC) simulation approach for energy determination. The constant intensity cut (CIC) method is an empirical method relying primarily on real showers observed by the detector to arrive at

the angular dependence of the atmospheric attenuation, hence zenith dependence, of S(1000). The CIC method relies on the assumption that cosmic rays arrive isotropically from all zenith angles on large scales (large scales being on the same order as the angular binning used to determine the constant intensity cut curve). This assumption is supported by angular analysis to date [53] with the Auger detector.

6.5 Reconstruction and Fitting Procedure for Real Showers

The parameterization of S(1000) in the constant intensity cut method requires a set of real reconstructed showers. The showers are reconstructed in the same fashion as those of Monte Carlo using the offline reconstruction with a fixed beta for all events. For the SD, the acceptance of the detector is uniform in $\sin(\theta)\cos(\theta)$ (see Ch. 7), thus if the events are divided into bins of equal size in $\sin^2(\theta)$, then each bin should receive equal intensity in cosmic ray flux, provided that the acceptance saturation condition is satisfied, (i.e. the energy of the cosmic rays should be high enough ($>\sim$ 3 EeV) to always trigger the array for all zenith angles (see Ch. 7)). The events are divided into zenith bins of 0.1 in $\sin^2(\theta)$. The cosmic rays in each bin are ordered from highest to lowest S(1000). Since each bin sees equal intensity, the N_{th} event in each bin should be the result of a shower with approximately the same primary energy. In other words, equal integral flux in each bin corresponds to the same energy. The N in N_{th} should be sufficiently large so as not be affected by the low statistics of highest S(1000) events — the flux of the highest energy events is so small that the flux of those events across all the bins is definitely not uniform.

A cut is then imposed at the N_{th} event, say the 150th highest event, for each zenith angle bin (Fig. 6.9) and a linear fit using the logarithm of the fifteen S(1000) values above and below the 150th event is applied to smooth out statistical fluctuations. The central value of the fit is taken as the S(1000) of the 150th event and plotted for each $\sin^2(\theta)$ bin. Those points are fit with a third degree polynomial (see Fig. 6.10) using a weighting equal to the S(1000) difference between the N_{th} event and the N_{th} + sqrt[N_{th}] event (the error bar in each bin in Fig. 6.10) for each bin to obtain an S(1000) attenuation curve.

Fig. 6.10 tells us what the relative attenuation of S(1000) as a function of zenith angle at one particular energy, but we do not know what energy Fig. 6.10 corresponds to. By the same token, we can choose other numbers for the N_{th} event and obtain different CIC cuts. The curves obtained from different CIC cuts correspond to S(1000) attenuation curves at different energies. Fig. 6.11 shows different CIC curves at different (approximate) energy cuts. All the different curves are normalized to the same value at 38° to compare attenuation at energies, and the shapes of the curves are different. That the shape of the CIC curves at different energies are different is expected as the X_{max} and the muon richness, which affect the shape of the attenuation, have energy dependences. The difference between different curves are at $\sim 5-10\%$ level. So we choose the CIC curve for the 150th event (which roughly correspond to 10 EeV) for energy determination.

6.6 Absolute Energy Scale

CIC only yields the shape of an S(1000) attenuation curve, i.e. how S(1000) changes as a function of zenith angle for a given energy. One does not know, however, what that given energy is from CIC alone. For example, Fig. 6.10 was

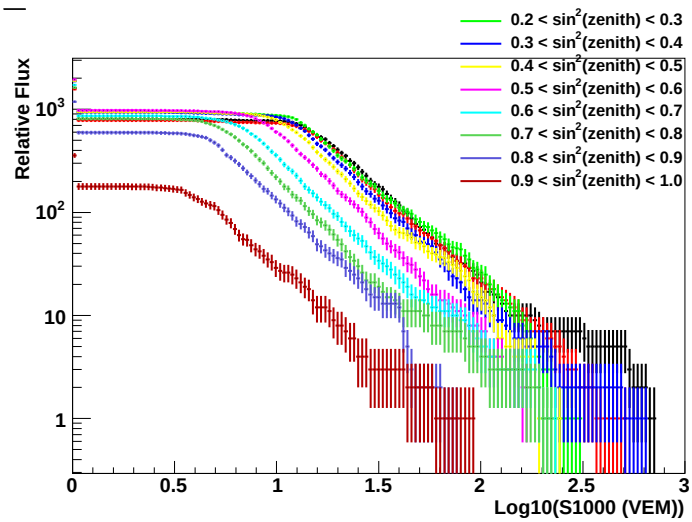


Figure 6.9: Integral Flux versus log S(1000) for Various Zenith Bins. The Constant Intensity Cut Points are Obtained by Drawing a Horizontal Line and Using the S(1000) Values from the Intersection Points. Plot courtesy of M. Healy.

obtained for the 150th highest events in each bin, but we do not know have the energy information for that curve. One natural way to set the absolute energy scale is to use hybrid events to transfer the FD energy calibration to the SD events with FD energy. This section will detail how that is accomplished.

6.6.1 Hybrid Data

The data sample used for this analysis is from Jan 01, 2004 to June 29, 2006. Before a cross calibration between SD and FD is carried out, it is imperative the hybrid events that are used are high quality events. Quality cuts recommended by the Adelaide group who also kindly provided the hybrid data are followed to ensure high quality of the hybrid events.

The exact quality cut criteria for the FD side are as follows.

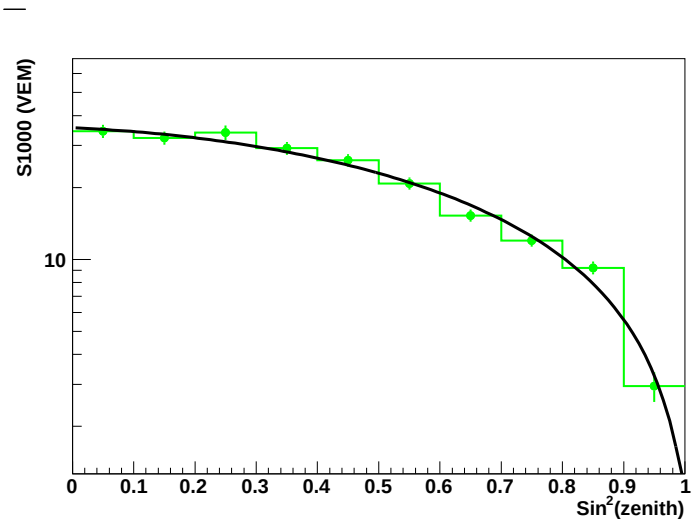


Figure 6.10: Fit to the 150th Highest Event in each Zenith Bin. The Curve is Later Normalized to Obtain a Conversion at 10 EeV (Fig. 6.3; black curve). Plot courtesy of M. Healy.

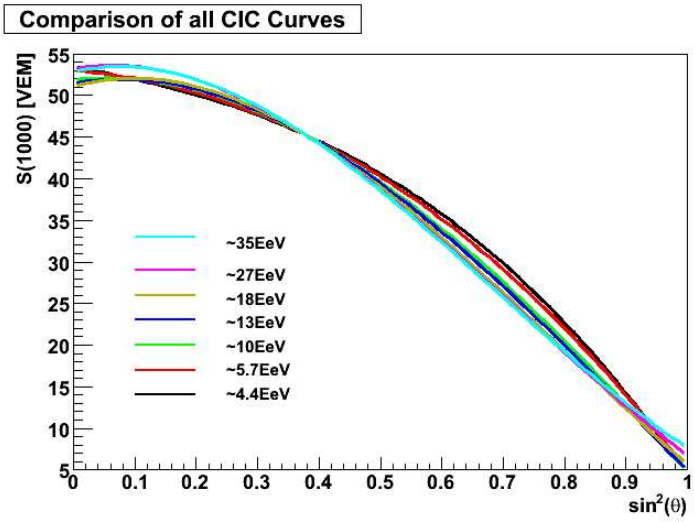


Figure 6.11: CIC curves for different CIC cuts. Energies are approximate as we do not actually know the absolute energy scale. Plot courtesy of M. Healy.

- Data from Los Leones and Coihueco only
- Geometrical reconstruction must be successful.
- X_{max} must be in the field of view and the observed track length is $>$ 200 g/cm^2.
- Minimum viewing angle is $> 20°$.

Requiring a minimum viewing angle eliminates events with large Cerenkov light contamination.

6.6.2 Cross Calibration

Now we set the absolute normalization of the CIC curve thereby setting the absolute energy scale. First, the constant intensity curve (see Fig. 6.10) that was obtained in the $\sin^2(\theta)$ space is now plotted in the $\sec(\theta)$ space. This step is really not necessary, but since the shower depth at the ground has a $\sec(\theta)$ dependence, it seems logical to plot in $\sec(\theta)$. Then the constant intensity curve is moved up and down in the scatter plot of S(1000) and $\sec(\theta)$ of the hybrid events until the best fit is obtained. The normalization is the only free parameter (See Fig. 6.12). S(1000) for each event in the plot is actually normalized to its equivalent value at 10 EeV at that particular angle, where we have used the FD energy of each hybrid event and the relation $S(1000) \sim E^{0.95}$ obtained in the Monte Carlo simulation studies. The fit is performed with equal weight for each point. Fig. 6.12 shows only events with the FD energy above 3 EeV, and the resulting in S(1000) = 59.6 VEM for $\theta = 0°$ (or, equivalently S38 = 50 VEM) at 10 EeV.

The result of this analysis method will vary depending on the size and the quality of the data set. One way to check the robustness of the analysis method

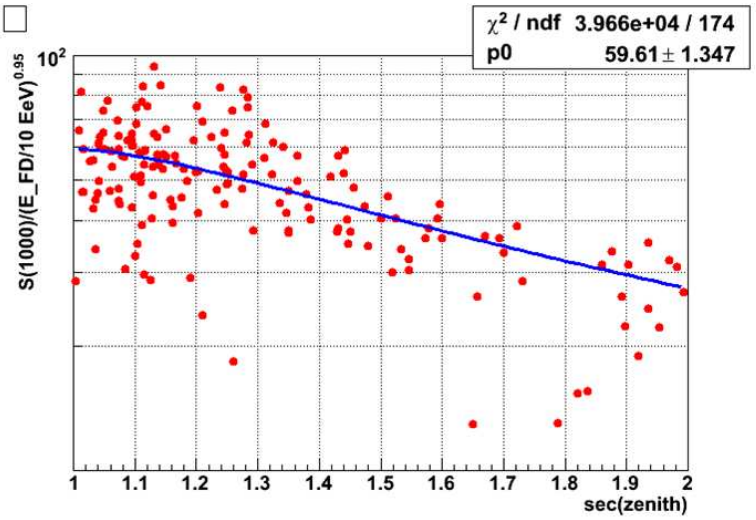


Figure 6.12: Absolute energy calibration of the SD using the hybrid events. Each dot is a hybrid event with S(1000) normalized to equivalent value to 10 EeV, and the line is the constant intensity curve.

is to vary the FD energy cut using only events above the cut in the analysis, and examine its effect on the fit. The higher the energy cut it is likely the quality of the hybrid data will also improve, but that will also result in a smaller data set. If the analysis method is robust and there is no hidden energy dependent systematics, the result should be robust against a change in the energy cut. The result is shown in Fig. 6.13. S38 is stable at around 50 VEM for $\theta = 38°$ for the FD energy cuts above 3 EeV. The disagreement below 3 EeV is not surprising as the SD reconstruction accuracy below 3 EeV is poor, and the small size of the hybrid dataset beyond 10 EeV affects the accuracy of the fit for FD energy cut beyond 10 EeV.

In the analysis above, we made use of a result from simulation, and S(1000) was assumed to have energy dependence, $S(1000) \sim E^{0.95}$. But the validity of

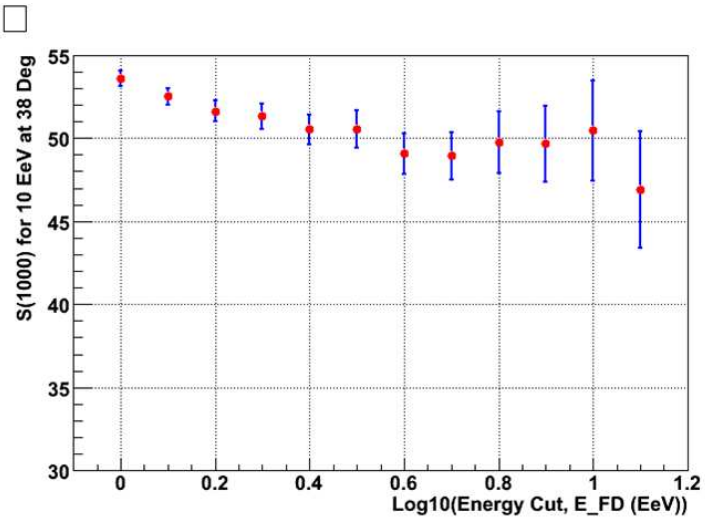


Figure 6.13: S38 vs. Energy Cut. S38 is $\sim$ 50 VEM for energy cuts above 3 EeV, below which the SD reconstruction accuracy is poor.

the $E^{0.95}$ dependence itself should be checked using the real data. Analysis by several groups using the real data only show $S(1000) \sim E^{0.95}$ to be valid (see Fig. 6.14).

Now that we have set the absolute energy scale using the FD energy, we compare the SD energy with the FD energy for the hybrid events. Fig. 6.15 shows the difference in the SD and FD energy for all the hybrid events above 3 EeV. As expected, the mean energy difference is negligible. The spread is reasonable at $\sigma = 0.23$. Although it is not obvious from the CIC method what the fluctuation in S(1000) should be, from MC studies, we expect the statistical fluctuation in S(1000) for events below 10 EeV (most of the hybrid events are below 10 EeV) to be $\sim$ 15% (see Fig. 6.7). So it is likely that half of the σ is due to the fluctuation in S(1000) and half of it is due to statistical uncertainties in the FD energy.

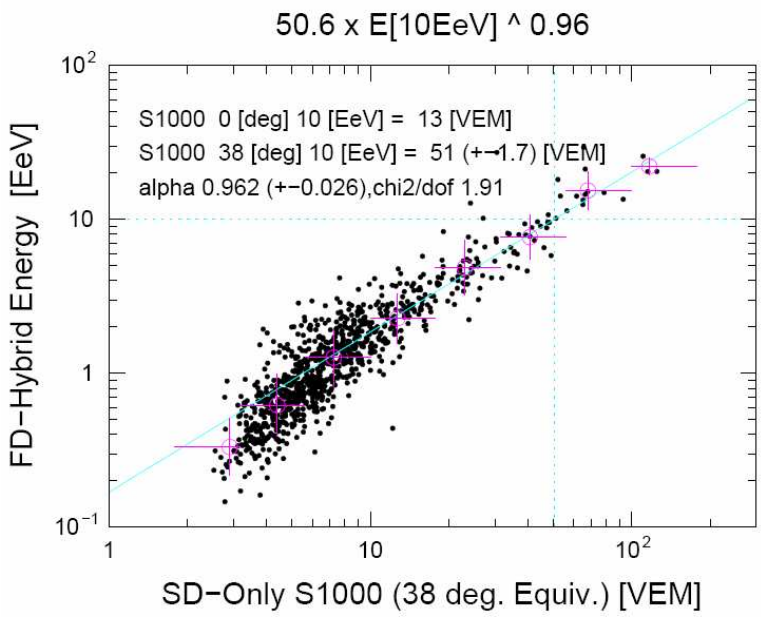


Figure 6.14: Correlation between the FD energy and S(1000). S(1000) is converted to its equivalent value at 38° using the constant intensity curve. $S38 = 51$ VEM and $S(1000) \sim E^{0.96}$ in this analysis which is consistent with our results, $S38 = 50$ VEM and 0.95 power law. Figure from [63].

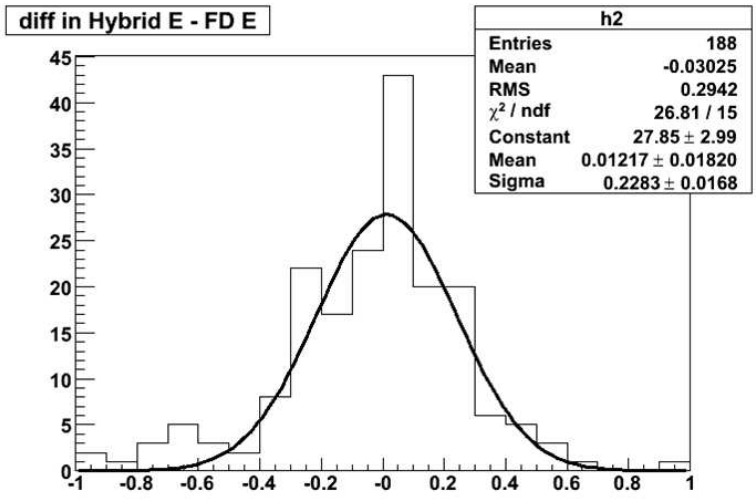


Figure 6.15: Fractional difference between SD and FD energy for hybrid events above 3 EeV. Since the SD energy scale is set with FD scale, we expect the mean difference to be 0. The result is consistent with the expected value.

Finally, in anticipation of the AirFly results which should reduce the fluorescence yield by $\sim 20\%$, the S38 value is reduced from 50 VEM to 40 VEM as the AirFly fluorescence yield measurement will reduce the FD energy by $\sim$20%. This only affects the absolute normalization of the S(1000) curve, and everything else remains the same. In all subsequent analyses, the S38 value for the hybrid energy calibration method will be 40 VEM.

6.6.3 MC+CIC

Just as we have set the absolute energy scale on the SD events using the constant intensity cut method and FD energy, we could set the absolute energy scale by setting the absolute normalization on the constant intensity curve using the results from the simulation studies. As we saw in Fig. 6.3, there are several MC

combinations. The way we choose to set the absolute energy scale for the constant intensity curve with MC simulation is as follows. We choose proton QGSJet and set the absolute normalization of the constant intensity curve at $\theta = 38°$, where the systematic uncertainty due to unknown composition is minimized. We will call this method 'MC+CIC'. S38 = 34.7 VEM for proton QGSJet, hence we set the normalization of the constant intensity curve at 10 EeV at S38 = 34.7 VEM for MC+CIC also.

6.7 Comparison

6.7.1 FD vs. MC

Now we can compare the S(1000) attenuation curves and energy scales of all the different methods. As can be seen in Fig. 6.16, FD calibrated constant intensity curve (FD+CIC) and MC+CIC are exactly the same S(1000) attenuation curve with different normalizations from two different calibration methods. Since S38 = 40 VEM for FD+CIC and S38 = 34.7 VEM for MC+CIC, the two energy converters will give a $\sim 15\%$ difference for the same events. We could use pure MC based energy converters to determine the energy as well. Given an S(1000) value, it could be converted to several different energies depending on the choice of the energy converter. As we will see in the next chapter, we can get both an AGASA-like spectrum and a HIRES-like spectrum depending on the energy converter confirming that the difference in the HIRES and AGASA spectra are due to systematic differences in the energy determination.

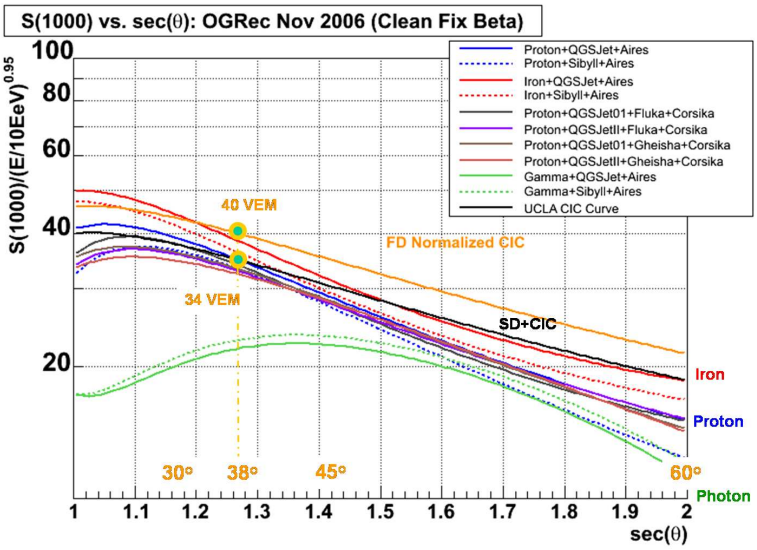


Figure 6.16: S(1000) curves for different MC combinations, MC+CIC, and FD+CIC. Plot courtesy of M. Healy.

6.7.2 MC vs. Constant Intensity

Although it would be preferable to fine tune the MC simulation to exactly reproduce the real data which would give us the confidence we understand the physics behind what we observe, the fact is that there is no accelerator data available yet for the energy scale of the UHECR , and we are forced to rely on the hadronic interaction models in MC simulations. As we see in Fig. 6.16, there is several combinations for the primary and hadronic model assumptions. If we believe constant intensity to be true, then we can see that the MC simulation cannot reproduce the constant intensity curve which is what the nature says how S(1000) attenuates as a function of the zenith angle.

A closer look at the differences between the constant intensity curves and the MC S(1000) attenuation curves reveals further insight into the makeup of the real data as well as improvements required of the MC simulation packages. Fig. 6.17

shows the ratio of the MC curves and the constant intensity curve. Ignoring the normalizations for now, if the MC simulation could reproduce the constant intensity curve perfectly then, the ratio would be a flat horizontal line. Instead, All the MC curves deviate significantly from the constant intensity curve. The proton MC curves, especially proton QGSJet, are close to the horizontal line at low zenith angles whereas they tail away like an exponential decay function at high zenith angles. On the other hand, the iron MC curves, especially iron QGSJet, deviate significantly from the constant intensity curve at low zenith angle while they are nearly flat at high zenith angles.

Low zenith angles are sensitive to X_{max}. The proton curves have maxima around 25° due to the fact that proton is deeply penetrating and the X_{max} is below the ground for the vertical shower, whereas the X_{max} is above the ground for iron for all zenith angles which is why the iron curves look more like exponential decay functions. That the shape of proton QGSJet S(1000) curve is close to that of the constant intensity curve imply that the X_{max} in the real data is proton-like.

High zenith angles are sensitive to the muon richness in the shower — at high zenith angles, the EM component dissipates to a large degree, and muons begin to dominate. At high zenith angles, the constant intensity curve closely resembles the iron QGSJet curve. So the real data is iron-like in terms of muon richness.

This means that we need a simulation package that has proton-like X_{max} and iron-like muon richness in order to reproduce the real observables. As we will see in chapter 7, pure MC based energy converters result in spectra that are non-constant intensity. More specifically, they result in too many highly inclined events.

Constant intensity cut method, on the other hand, has a drawback in that we need to start with an assumption that the cosmic ray flux is isotropic in

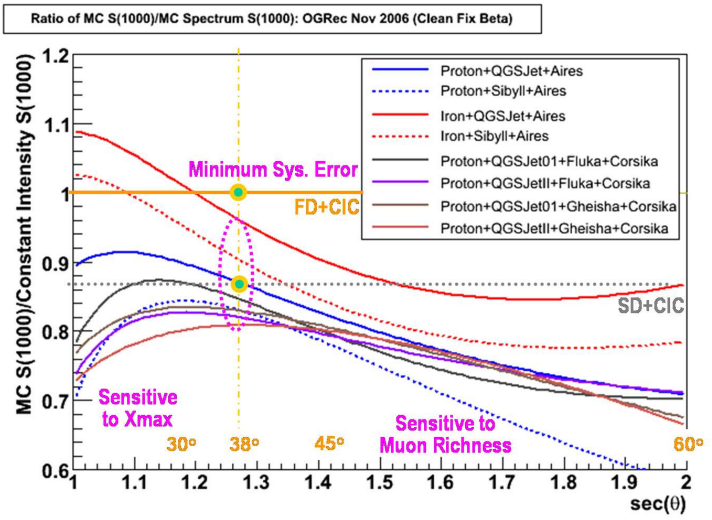


Figure 6.17: Ratios of S(1000) curves for different MC combinations, and MC+CIC and FD+CIC. Plot courtesy of M. Healy.

zenith angle however robust that assumption seems to be. The energy calibration method using the hybrid data set has the drawback of having to use a very small subset of the real data which is also mostly in the low energy range.

6.8 Systematic Uncertainties

In this section we will closely examine the systematic uncertainties in energy. The systematic uncertainties influence the appearance of the spectrum, so it is important to have a good understanding of the systematic uncertainties.

The systematic uncertainties can be broken down into different categories. To determine the energy of a cosmic ray, we need to first calibrate the detector, determine S(1000) and then rely on the MC simulation or the correlation with the FD energy scale to convert the S(1000) to energy. Each one of these steps will contribute to the total systematic uncertainties.

To begin with, the systematic uncertainties from the SD calibration are as follows (see Chapter 4, [53], and [64]). The uncertainties due to the accuracy of the VEM is within 5%. Then the systematic uncertainties due to non-linearity and D/A Ratio are known to $\sim 5\%$ each. The uncertainties in the water depth contribute $\sim 3\%$, and the contribution of the thermal effects are $\sim 3\%$. Response of the detector to the EM component (remember the calibration is carried out using muons) is estimated to contribute $\sim 5\%$. All in all, they all add up in quadrature to about 10% — these are conservative estimates, and some of these quantities are probably smaller, so the systematic uncertainties from calibration may be smaller than 10%.

For determination of S(1000), there are additional systematic uncertainties. First, the systematic uncertainties due to the assumption of an LDF model is $\sim 3\%$. The ignoring of the azimuth asymmetry and the shift of core location contributes $\sim 3\%$. The effect of LDF model assumed and the core shift is minimal as long as the same LDF model is used for the energy calibration process, be it MC-based or FD energy based as the systematic uncertainties due to the LDF model should cancel out to first order. And as the systematic uncertainties are small, even if the energy calibration that we obtained in this chapter, for example, are used for some other reconstruction using another LDF model, the resulting systematic error should be small. Recovery of PMT saturation is estimated to contribute $\sim 3\%$, and the contribution of the reconstruction algorithm is conservatively estimated at $\sim 3\%$. All in all, the S(1000) determination process contributes $\sim 6\%$.

Now there are two paths for the energy calibration process as we saw above. For the MC-based method, first the difference between Corsika and Aires which is at a 5% level should be accounted for. The energy dependence of the compo-

sition and the models as well as the sampling and the resampling of the shower particle has a systematic uncertainty of 5% each. The atmosphere density used in the simulation is estimated to contribute 5%. The S(1000) dependence on the hadronic interaction model and the mass composition was seen above, and they each contribute about 10% — these have zenith dependences, of course, and around 38°, the systematic uncertainties are minimized.

Adding all these up in quadrature yields a total systematic uncertainties of 18-19% for the MC-based energy determination method at 10 EeV. Of course, at 100 EeV additional systematic uncertainties due to energy extrapolation (S(1000) does not follow the 0.95 power law exactly) and possible mass composition change in the real data are both estimated at 5%. Hence, at 100 EeV the systematic uncertainties are 21-22%.

For the hybrid based method, the systematic uncertainties in the FD energies are carried straight over to the SD side [65]. The fluorescence yield has a systematic uncertainty of 15% [42] (Or in the case of AirFly, the systematic uncertainties will be in the range of 1-2%). The atmospheric correction has a 10% uncertainty, and the detector calibration has a 12% uncertainty [28]. Profile reconstruction, cerenkov subtraction, and invisible energy each contribute 5%. All in all, the FD energy has a systematic uncertainty of $\sim 25\%$.

The CIC method as we saw earlier has an energy dependence in that the shape of the constant intensity curve changes with energy. We assume, however, that the S(1000) attenuation curve is the same for all energies, and we choose the constant intensity curve for the 150th event cut to be the S(1000) attenuation curve. The energy dependence is $\sim 5\%$, and the accuracy of the constant intensity curve is estimated at $\sim 5\%$.

Summing up the FD contributions and SD contributions from the calibration,

S(1000) determination and CIC determination, the total systematic uncertainties in the FD-based energy is $\sim 30\%$ at 10 EeV. At 100 EeV, additional systematic uncertainties are present as was the case for the MC-based energy. Once again, the extrapolation of energy to 100 EeV introduces systematic uncertainties as for this method there is no hybrid event at 100 EeV, so the energy correlation obtained for low energies is extrapolated out to 100 EeV. If we incorporate the results from MC simulation (the 0.95 power law) we can reduce the systematic uncertainties to 5%. And the unknown composition must be taken into account. The hybrid calibration method is blind to the composition in that we just determine the correlation between the average of S(1000) and FD energy without any knowledge of the composition. As long as the composition does not change in the extrapolation region, the energy calibration should be valid. But if the composition were to change at 100 EeV, say from proton to iron, then the energy calibration would be wrong at 100 EeV. We know from MC simulation, the effect of the composition on the S(1000) value. We estimate the systematic uncertainty due to the unknown composition at 10%. At 100 EeV, the FD-based method is estimated to have systematic uncertainties of $\sim 33\%$.

The MC-based and FD-based energy determination methods have their own merits and drawbacks. The MC-based method is based on known physics, and we (or author of the MC package) know exactly the inner workings, except for the extrapolation out to the region where there is no accelerator data — Indeed, some contend that S(1000) values may not be bound by Proton Sibyll (lower bound) and Iron QGSJet (upper bound), therefore the systematic uncertainties in the hadronic interaction models are unknowable. The drawback of this method is that no MC package can reproduce the observables exactly, namely all the pure MC-based energies too many inclined events which violate constant intensity (see chapter 7). The FD-based method has the advantage of being independent

of hadronic interaction models and primary assumptions. But the drawback is that first we must assume constant intensity of cosmic rays — though seemingly robust, the less the number of assumptions the better. And second, we must calibrate energy at low energies for a small subset of well reconstructed events then apply it to all events including high energies. This method exposes us to a systematic error if the composition is energy dependent. Presently, the MC simulation packages are imperfect, so FD-based method, which in principle is better because FD measurements are in principle nearly calorimetric, is the preferred method. **Thus, FD-based energy will be our default choice**. Since we know the shortcomings of MC simulation packages, Auger can help the effort to produce a 'perfect' simulation package — if not directly, we can at least validate the results of the MC simulation with real data. Having a 'perfect' simulation package means we can get at the essence of cause and effect rather than just correlations. And it would help us perform more detailed and thorough analyses.

CHAPTER 7

Spectrum

7.1 Introduction

The Pierre Auger Observatory (PAO) is a hybrid detector with a Surface Detector (SD) and a Fluorescence Detector (FD) at its disposal [25]. With the two different types of detectors that AGASA and HiRes employed and much larger statistics, the PAO is particularly well suited to settle an outstanding question of gargantuan proportions in the cosmic ray community — Whose spectrum is right? AGASA and HiRes reported spectra that disagree both in the absolute flux of the cosmic rays and the existence of the GZK cutoff [66, 67, 68, 69]. It has been posited that the discrepancy in the spectra reported by AGASA and HiRes are due to the systematic uncertainties associated with determination of energy in the surface array and fluorescence detectors. A shift in energy of 30% for either AGASA or HiRes would result in consistent spectra. Since these are two independent experiments, the agreement of the two spectra after the 30% shift in energy does not prove that the disagreement is indeed due to the different energy calibration methods. To prove it, we need to apply the two energy calibration methods to the same set of events. The PAO with the hybrid detector should be able to not only settle the difference between AGASA and HiRes but determine the cosmic ray spectrum with unprecedented precision using its unmatched statistics.

An accurate measurement of the cosmic ray spectrum is of particular inter-

est to the cosmic ray community at large due to the fact that an AGASA-like spectrum with the absence of the GZK cutoff could mean exciting possibilities. An absence of the GZK cutoff could signal new physics, the so-called top-down models being the favorite among many [70, 71, 72, 73]. Such models all predict a significant flux of gamma rays at energies beyond the GZK cutoff energy. There have been studies to determine the gamma flux in the PAO to directly test the top-down models [74, 75].

The cosmic ray spectrum also has implications on the locations of the cosmic ray sources and composition. In the conventional bottom-up scenarios, the sources of super-GZK events must be within the GZK sphere (see chapter 8) in which case one should be able to correlate super-GZK events with candidate sources provided that the extragalactic and galactic magnetic fields are small enough, and that the cosmic ray primaries are protons. If super-GZK events do not correlate with sources, that implies heavy element and/or strong intra- and inter-galactic magnetic fields. Hence, an accurate determination of the cosmic ray spectrum has an importance well beyond setting the absolute flux of cosmic rays – it is an integral part of the search for the sources and the composition of cosmic rays, and a test of conventional physics theories.

An exhaustive study of different contributing factors to inaccuracies in the cosmic ray spectrum is examined here. The consistency of Monte Carlo simulation (MC) based and Constant Intensity Cut (CIC) based energy converters on the spectrum are examined. The effect of inaccuracies in reconstruction on the spectrum are closely monitored with a MC cosmic ray spectrum. In the end, it will be shown that it is possible to obtain two spectra — one that is AGASA-like, and one that is HiRes-like.

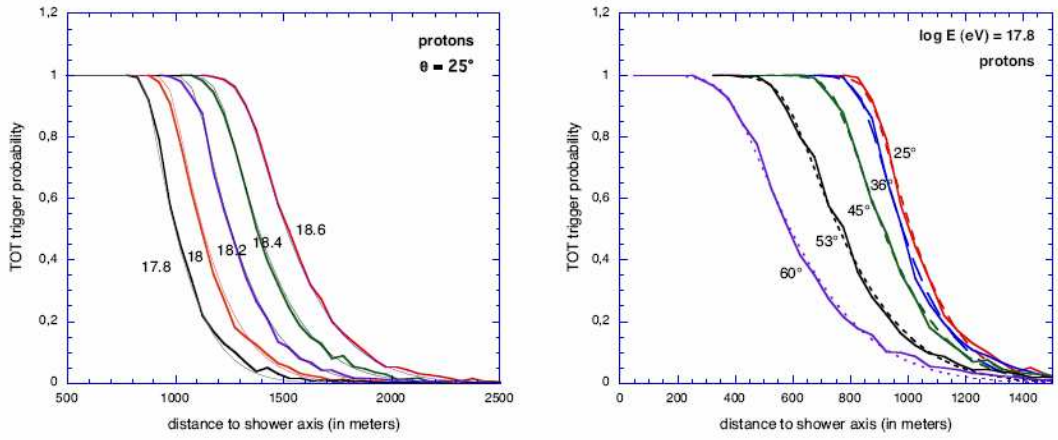


Figure 7.1: Lateral trigger probability function for Corsika proton QGSJet showers. Energy dependence at $\theta = 25°$ (left) and zenith dependence at $E = 10^{17.8}$ eV. Figure from [76].

7.2 Data Set

The Data Set used for this analysis is from January 04 to February 07. Only T5 events with $E > 3$ EeV and $\theta < 60°$ are used as the array is fully efficient above 3 EeV (see below). As mentioned earlier, in light of imminent release of the new fluorescence yield measurement by AirFly which will have much smaller systematic uncertainties ($\sim 1 - 2\%$) than the previous fluorescence yield measurements, **the FD-based energy scale is readjusted from S38 = 50 VEM to S38 = 40 VEM** as the AirFly results reduce the fluorescence yield by 20%. The FD-energy based spectrum will be compared with spectrum obtained with MC-based energies along with some additional checks. If nothing else, this should help identify the shortcomings of the MC packages and suggest ways to improve them.

7.3 Exposure

To calculate the exposure of the SD, we need to first understand the acceptance and the aperture of the detector. The instantaneous aperture of the SD is changing constantly as the configuration of the detector is changing as more tanks are added to the array periodically, and stations go online and offline. As mentioned briefly in Chapter 3, the T4 and T5 triggers make the calculation much easier. T4 selects showers with $\theta < 60^\circ$ very efficiently with the 3ToT trigger. And T5 is a quality trigger that ensures high quality of reconstruction while simplifying the calculation of the aperture.

First, we need to understand the effects of the granularity of the SD and its effects on the aperture. While high energy showers are large enough that they trigger many tanks, depending where they fall, low energy showers may not trigger enough stations to be recorded. So we do not expect the aperture to be uniform for all energies. To figure this out, we need to know first the lateral trigger probability (LTP) function which gives the local TOT trigger probability as function of distance away from the core [77]. This can be obtained by averaging over many simulated showers to calculate the ratio of the number of triggered tanks at a distance from the core to the total number of station at that distance from the core. An example of LTP function is shown in Fig. 7.1 for Corsika Proton QGSJet showers. As expected, at a given distance, LTP is higher for higher energies (larger showers) and smaller zenith angles (younger hence larger showers).

The T5 condition breaks down the array in terms of hexagonal cells with a central station surrounded by the 6 closest neighbors — otherwise, that event is rejected. Now that we know the LTP function we can figure out the probability of a 3ToT trigger. Labeling each tank in the hexagonal cell with index $i = 1$ to

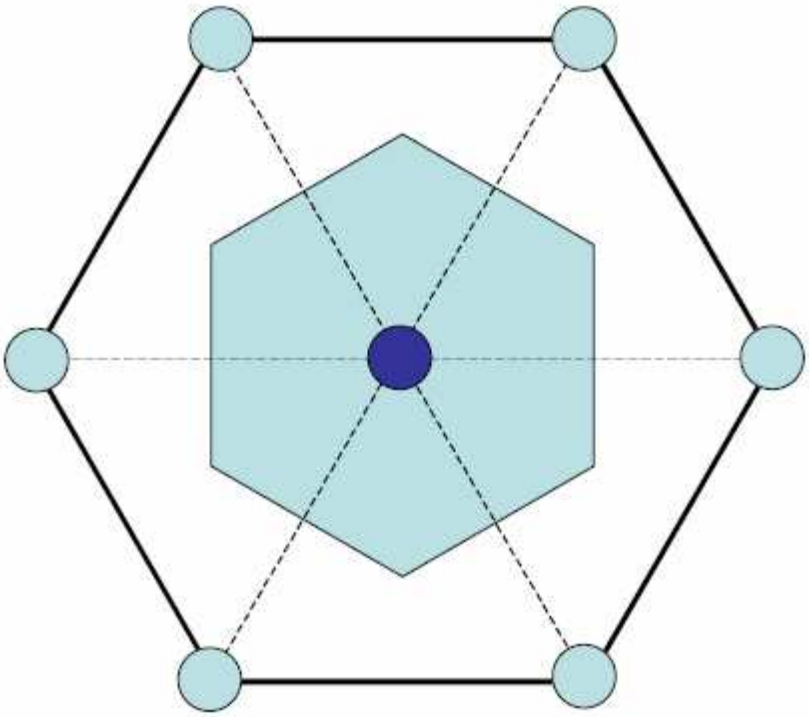


Figure 7.2: Elementary hexagonal cell per requirement of T5. The central station must be surrounded 6 working stations. The shaded area is the effective detection area of the cell. Figure from [76].

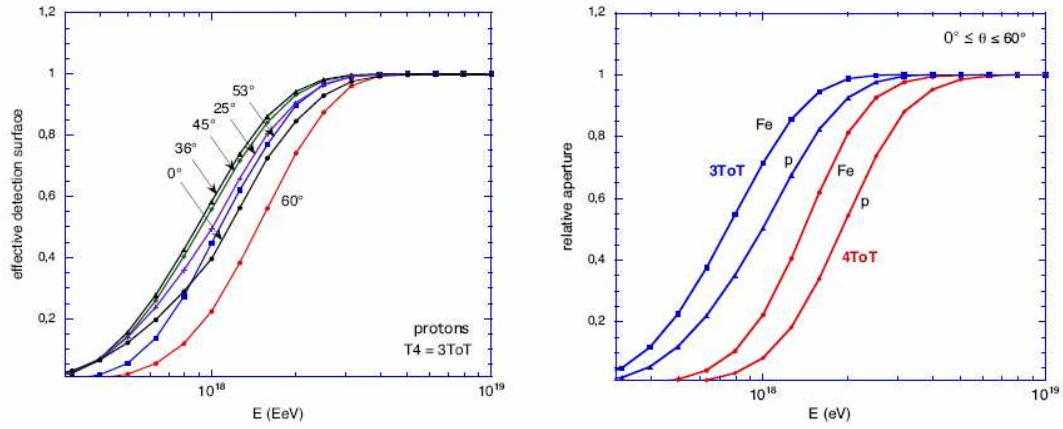


Figure 7.3: Relative effective detection surface (left) and relative aperture (right). Above 3 EeV, the effective detection surface and the aperture saturate, i.e. reach maximum values. Figure from [76].

7, and letting P_i be the probability that station i triggers and $Q_i = 1 - P_i$, for an arbitrary core position (x_c, y_c), the probability of a T4 trigger is

$$P_{T4}(x_c, y_c) = 1 - \left(\prod_{i=1}^{i=7} Q_i\right) \times \left(1 + \sum_i \frac{P_i}{Q_i} + \sum_{j>i} \frac{P_i}{Q_i}\frac{P_j}{Q_j}\right) \tag{7.1}$$

This quantity averaged over all allowed core locations yields the global shower detection efficiency of the SD [76]. The shaded area in Fig. 7.2 indicates where the core position must be inside a hexagonal cell in order to satisfy T5.

The *effective detection surface* of the hexagonal cell is, then, given by

$$S_{eff} = \int_{cell} P_{T4}(x_c, y_c) P_{T5}\, dS \tag{7.2}$$

where P_{T5} is 1 or 0 depending on whether the T5 condition is met or not. Fig. 7.3 shows that S_{eff} is saturated above 3 EeV for moderately inclined showers and 4 EeV at 0° and 60° [76]. The effective surface area of the SD is the normal area

to the angle of the incidence of the showers, i.e. $A(\theta) = A \cdot \cos(\theta)$. Thus, the aperture is obtained by integrating the following over solid angle.

$$a(E) = \int_{\theta \leq 60^\circ} S_{eff}(E, \theta, \phi) \cos\theta \sin\theta d\theta d\phi \tag{7.3}$$

Fig. 7.3 shows that the aperture is saturated above 3 EeV for protons and for irons the aperture is saturated at a lower energy [76].

Above 3 EeV, then, the aperture is uniform and purely geometrical. Calculating the instantaneous aperture boils down simply adding up the number of active hexagonal cells. At full detection efficiency, Each hexagonal cell has a detection area of $(1.5 \text{ km})^2 \cdot \sqrt{3}/2 \simeq 1.95 \text{ km}^2$ (see Fig. 7.2), and the corresponding aperture for $\theta < 60^\circ$ is $a_{cell} \simeq 4.59 \text{ km}^2 sr$.

The integrated exposure of the SD is calculated by counting the number of cell-seconds. The array configuration changes as more stations are added periodically and stations sometimes fail. The changes in the array configuration can monitored second by second by monitoring the T2 rate. To calculate the integrated exposure, for each new configuration, the number of hexagon cells, N_{cell}, is counted and the exposure for that particular configuration, $N_{cell} \cdot a_{cell} \cdot \triangle t$, is incremented, where $\triangle t$ is the duration of the configuration.

The calculation of the exposure is purely geometrical, the accuracy of this calculation depends on the handling of hidden deadtimes. The T2 files which give us information on the single trigger rate at the station level do not give us information on the performance of the central trigger and CDAS. Exposure should only be calculated for periods where the stations, the central trigger and the CDAS are all working at the same time. An additional check is added to accomplish this.

Assuming a constant rate λ for the T5 event rate per hexagon — although

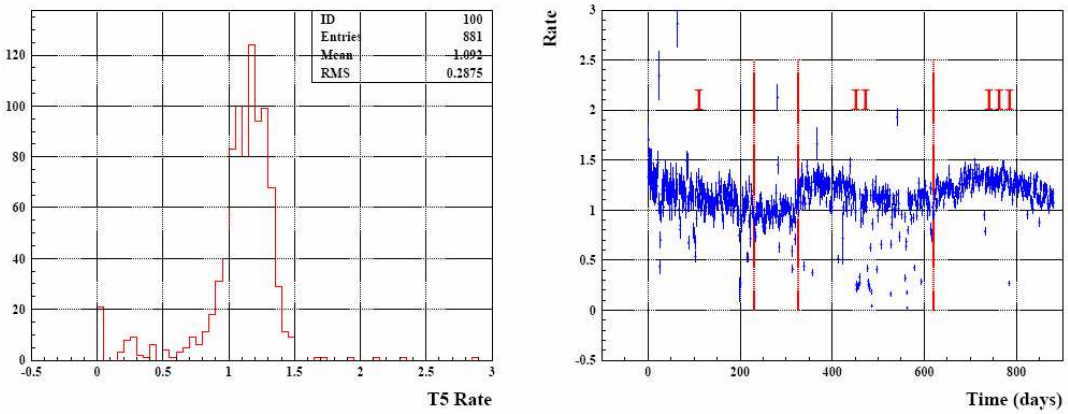


Figure 7.4: Distribution of the T5 event rate (per hexagonal cell per day) for 881 days (left) and T5 rate over time (right). Figure from [78].

there is seasonal variations, to first order this is true (see Fig. 7.4) — the probability P of the time interval between two consecutive events being larger than T is,

$$P(T) = \exp^{-\lambda T} \tag{7.4}$$

If T is much longer than the period expected λ, or $P(T)$ is very small, then that is an indication that the array is not working properly for that period. The procedure to eliminate bad period is, then, to simply reject the time periods where P is below a threshold α.

Due to a bug in the central trigger algorithm and a software change, the data was divided into 3 periods (marked I, II, and III in Fig. 7.4). The distribution of event rate per hexagon (in units of events per hexagon per day) is shown in Fig. 7.5. The event rates from each period follow a Gaussian distribution, except for small non-Gaussian tails. The μ from the Gaussian fits are taken as λ for each period [78].

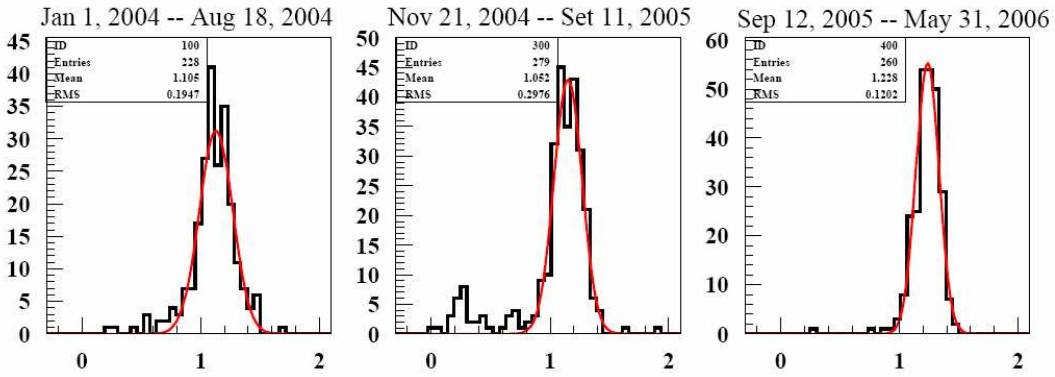


Figure 7.5: Distribution of the T5 event rate (per hexagonal cell per day) for the 3 periods. Figure from [78].

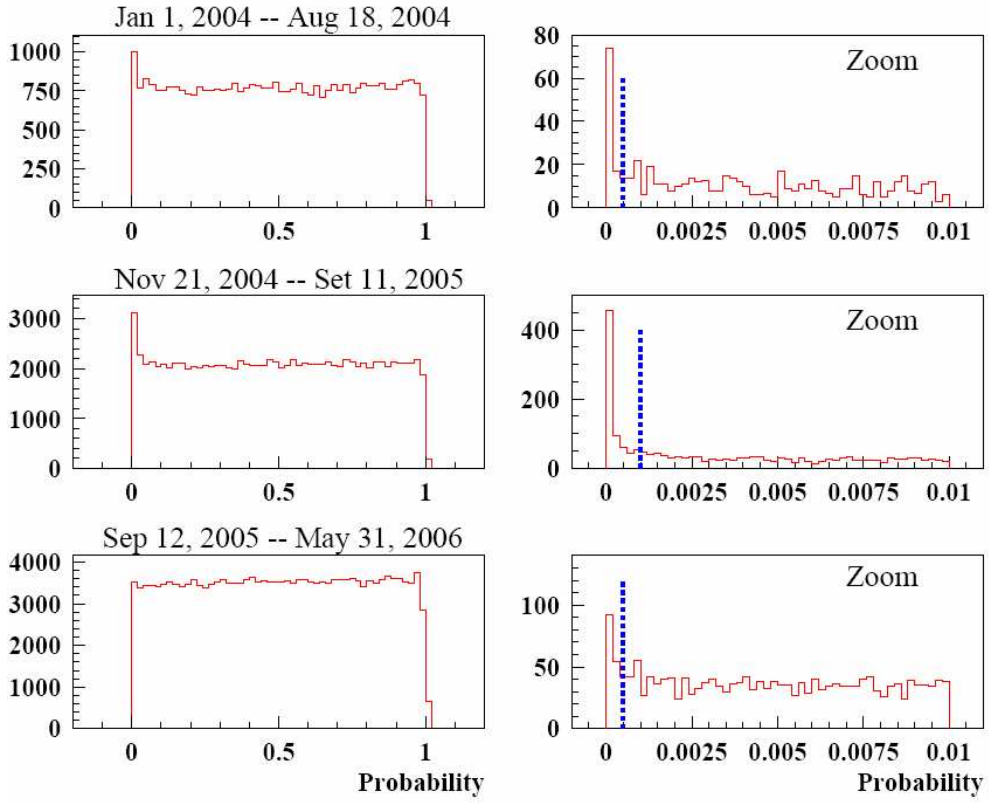


Figure 7.6: Distribution of the probability of the time interval between two consecutive events. Figure from [78].

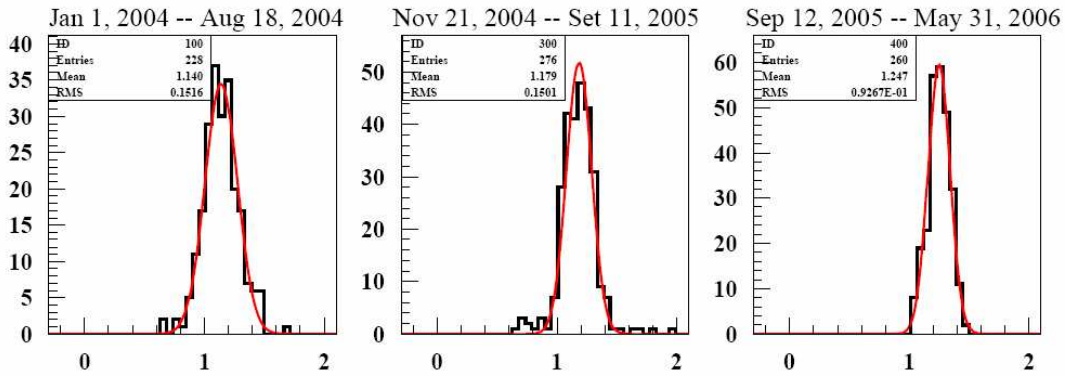


Figure 7.7: Distribution of T5 events after rejecting bad periods. Notice the absence of the non-Gaussian tails that were seen in Fig. 7.5. Figure from [78].

For each period above, the probability of the interval between two consecutive events are calculated using the equation above. The distribution of probability is shown in Fig. 7.6. As expected, the probability distribution is flat except for a small spike for very small probabilities, i.e. T is much longer than expected, for each period. Those spikes indicate an anomalous behavior in the operation of the SD, so α (the dotted lines in Fig. 7.6) is chosen to remove the spiky periods but not too much of the normal periods. The percentage of the bad periods cut out of each period is 3.6%, 15.1%, and 1.9% respectively, but the actual number of events lost are 1% or less in 3 periods [78]. The loss of events are minimal, and the rejected periods are indeed bad periods. Fig. 7.7 shows that after the rejection of the bad periods, and all the non-Gaussian tails seen in Fig. 7.5 are now gone.

The rejection of the bad periods remove hidden deadtimes, and the aperture and the exposure is now truly determined by the T2 files only as discussed above. After removing bad periods, the exposure of the SD from January 1, 2004 to February 28, 2007 is 5100 $\mathrm{km}^2 \cdot \mathrm{sr} \cdot \mathrm{yr}$ for all energies > 3 EeV, according to

P. Ghia *et al.*. The uncertainty in the exposure is estimated at 5% [76]. For Comparison, the integrated exposure reported at ICRC 2005 was 1750 km^2·sr·yr. The exposure has now nearly grown 3 times. And the integrated exposure of AGASA and HiRes were 1619 km^2·sr·yr and $\sim$ 5000 km^2·sr·yr, respectively. The exposure of the SD has eclipsed those of both HiRes and AGASA.

7.4 Determination of Cosmic Ray Flux

The cosmic ray spectrum approximately follows a power law: $dF/dE = kE^{-\alpha}$, where α is between 2 and 3. The spectrum is traditionally plotted in the log(E) space with a bin size of 0.1 to overcome the limited number of events that are available at the energies of interest to us. The differential flux, dF/dE, can be related to the number of events in each bin, which is straight forward to determine, in the following way. For some bin center at 10^c eV, we have for that bin

$$N = A \int_{10^{c-0.05}}^{10^{c+0.05}} \frac{dF}{dE} dE = kA \int_{10^{c-0.05}}^{10^{c+0.05}} E^{-\alpha} dE \tag{7.5}$$

where N is the number of events in the bin, and A is the exposure of the detector. After evaluating the integral, it follows that

$$\frac{dF}{dE}(E = 10^c) = \frac{N}{A} \frac{1}{f(\alpha)} 10^{-c} \tag{7.6}$$

where $f(\alpha) = \frac{1}{\alpha - 1}(10^{0.05(\alpha-1)} - 10^{-0.05(\alpha-1)})$. Although $f(\alpha)$ changes with α, the difference in $f(\alpha)$ is less than 10% for α between 2 and 3. Thus, we take $\alpha = 3$ and use the relation above when plotting the spectrum [79].

7.5 Spectrum

The spectrum is presented in this section using the method outlined above. There are many choices for energy converters to convert S(1000) to energy. For instance,

the oft-used 4 primary/model combinations (Proton QGSJet, Proton Sibyll, Iron QGSJet, and Iron Sibyll) all yield different energy converters (see Fig. 6.16), and each leads to a different spectrum. As mentioned earlier, one weakness of the MC energy converters is that currently MC simulation does not exactly reproduce the real data, i.e. the non-constant intensity problem where MC simulation energy converters give seemingly too many highly inclined showers compared to the vertical showers at the same energy (see below).

The current energy converter of choice Constant Intensity Cut (CIC) method solves this problem by assuming that the cosmic ray intensity has to be isotropic. CIC only gives the shape of the attenuation curve of S(1000) and can be normalized with either a MC simulation combination (MC+CIC) or hybrid events (FD+CIC). The spectrum resulting from MC+CIC and FD+CIC will be presented and will then be compared with those spectra obtained with the pure MC converters.

7.5.1 Spectrum with CIC

As mentioned in chapter 6, there are two CIC energy converters. MC+CIC is normalized with Proton QGSJet and results in S(38) = 34.7 VEM, and FD+CIC is calibrated with hybrid events and results in S(38) = 40 VEM (see Fig. 6.16). Auger is the first hybrid detector with both a fluorescence detector and a surface array, and with these two CIC energy converters we are employing the two different absolute energy calibration methods used by AGASA and HiRes to the same set of events and thus should allow us to identify the source of the discrepancies of the spectra reported by AGASA and HiRes.

The two spectra with MC+CIC and FD+CIC converters are shown in Fig. 7.9 and Fig. 7.11, respectively (see also Fig. 7.8 and Fig. 7.10 for the spectra in

$\frac{dF}{dln(E)}$ for an easy comparison with the ICRC 05 spectrum). The spectra reported previously by HiRes and AGASA are also shown on the same plots. It is apparent that FD+CIC yields a spectrum that is HiRes-like, and that MC+CIC produces a spectrum that is AGASA-like (albeit with a bit lower flux and a GZK-like feature). This means that the discrepancies in the AGASA and HiRes spectra, as conjectured by some, are due to systematic uncertainties in FD-based and MC-based energies. As it stands right now, the systematic uncertainties in FD-based energy is $\sim 30\%$ and MC-based energy is $\sim 20\%$ (see Chapter 6). Until the systematic uncertainties in both are reduced, it will be difficult to reject either the MC-based or FD-based spectrum.

The HiRes' spectra have the spectral shapes that are consistent with the GZK cutoff whereas the AGASA spectrum does not. Interestingly enough, despite the discrepancies in the MC+CIC and FD+CIC spectra, both methods produce super-GZK events beyond 100 EeV. This is not surprising as even with conventional bottom up scenarios we do expect some flux beyond the GZK cutoff energy. It is also quite clear, however, that both methods yield spectra with GZK-like features as we see steep drop-off beyond 50 EeV in both spectra. This seems to indicate that the part of the spectrum beyond the GZK cutoff energy of the AGASA spectrum, which engendered excitement for the possibility of top-down scenarios because of the absence of the GZK-like feature, may be wrong. The question of the existence of the GZK cutoff will be addressed in chapter 8 where we will do comparisons of the Auger spectrum with the theoretical spectra for proton and Iron primaries.

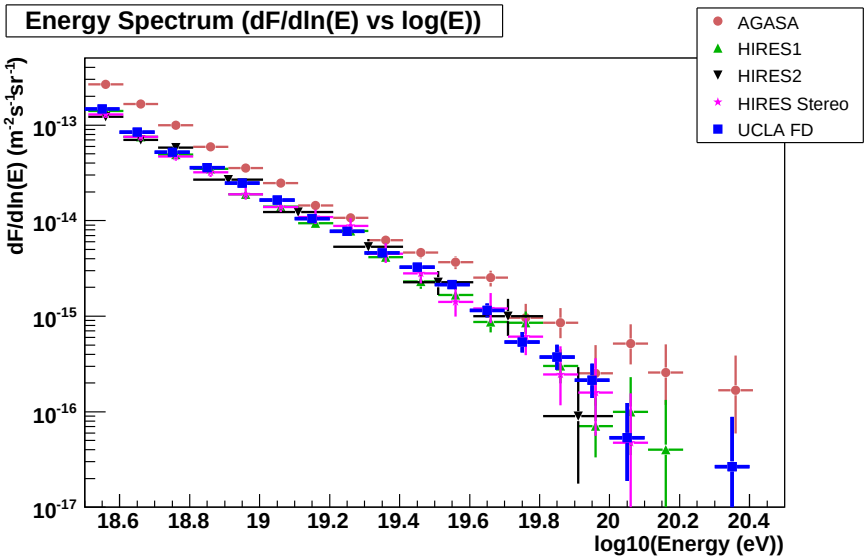


Figure 7.8: FD+CIC spectrum in $\frac{dF}{dln(E)}$ for easy comparison with the ICRC 05 spectrum.

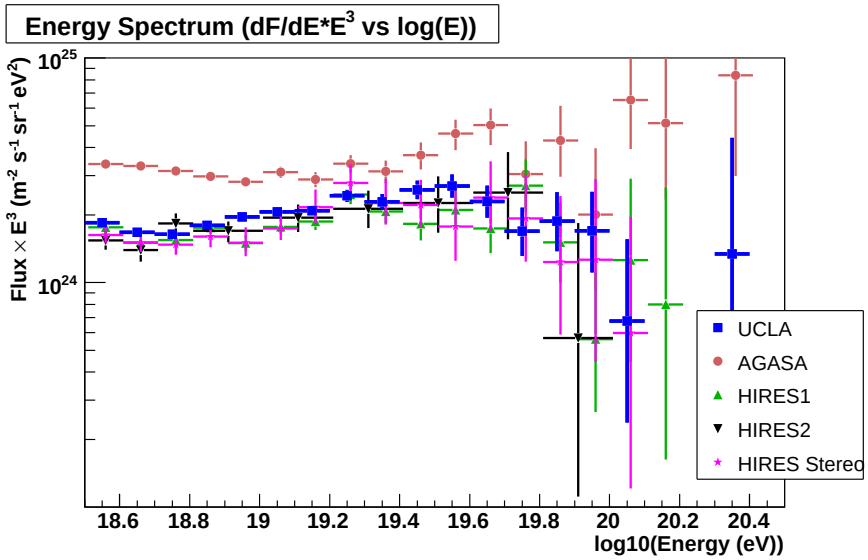


Figure 7.9: FD+CIC spectrum. Energy is calculated using CIC based converter normalized by hybrid events.

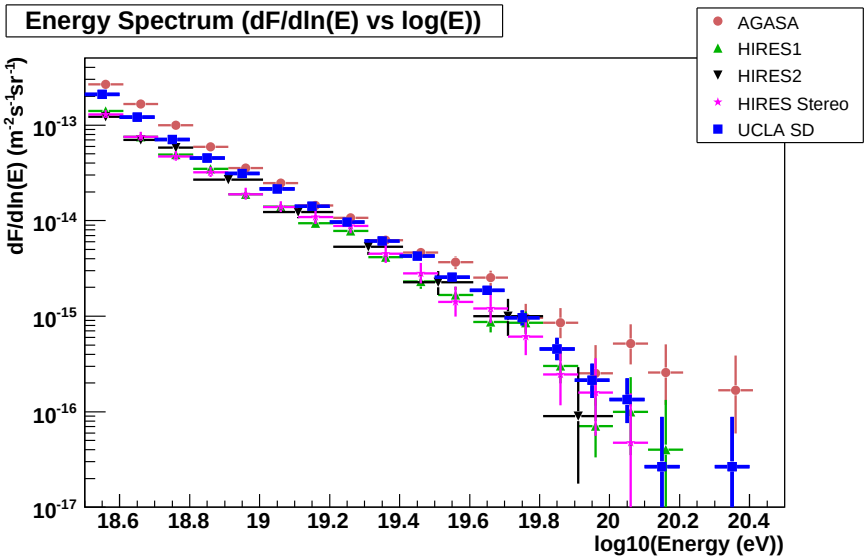


Figure 7.10: MC+CIC spectrum in $\frac{dF}{dln(E)}$ for easy comparison with the ICRC 05 spectrum.

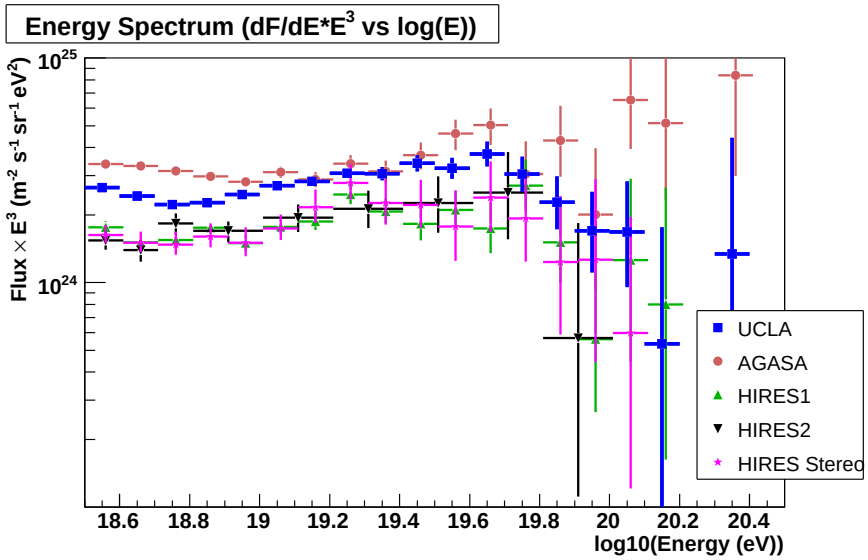


Figure 7.11: MC+CIC spectrum. Energy is calculated using CIC based converter normalized with Proton QGSJet at 38°.

7.5.2 Spectrum with pure MC energy converters

Although CIC is built on the seemingly robust assumption that the flux of the cosmic rays should be isotropic, this reliance on the assumption of the isotropy in the zenith angle distribution of cosmic rays is also its weak point, especially at high energies. At high energies, the cosmic ray flux is definitely not isotropic, due to the lack of statistics, and this necessitates an assumption of constancy of the shape of the S(1000) attenuation curve at all energies in order to assign energy. The effect of this assumption is small as the indication so far is that the constant intensity curves at different energy cuts agree to 5–10% level (see chapter 6).

At present, MC, regardless of primary/model combination, does not exactly reproduce the real observables as it does not produce constant intensity (see below), but it would be nonetheless useful to compare the spectra obtained using CIC converters with those obtained using the MC energy converters. It would at the very least point out what needs to be improved and what is being done right in the simulations. The 4 most commonly used primary/model combinations are proton Sibyll, proton QGSJet, Iron Sibyll, and Iron QGSJet (see Fig. 6.16 for the attenuation curves of the 4 combinations). Proton has a deeper X_{max} than iron and is less muon rich than iron. Sibyll gives a deeper X_{max} and less muons than QGSJet. Thus Proton Sibyll has the deepest X_{max} (below the ground at Malargüe) and least amount of muons which explains why the S(1000) curve has a peak around 25° zenith and the tail drops off sharply. Iron QGSJet has the shallowest X_{max} and most muons which is why the S(1000) curve resembles an exponentially decaying function, and the tail end is flatter. Iron Sibyll and Proton QGSJet lie somewhere in the middle. These differences mean that for the same S(1000), different energy converters will give different energies — from the highest to the lowest Proton Sibyll, Proton QGSJet, Iron Sibyll, and Iron

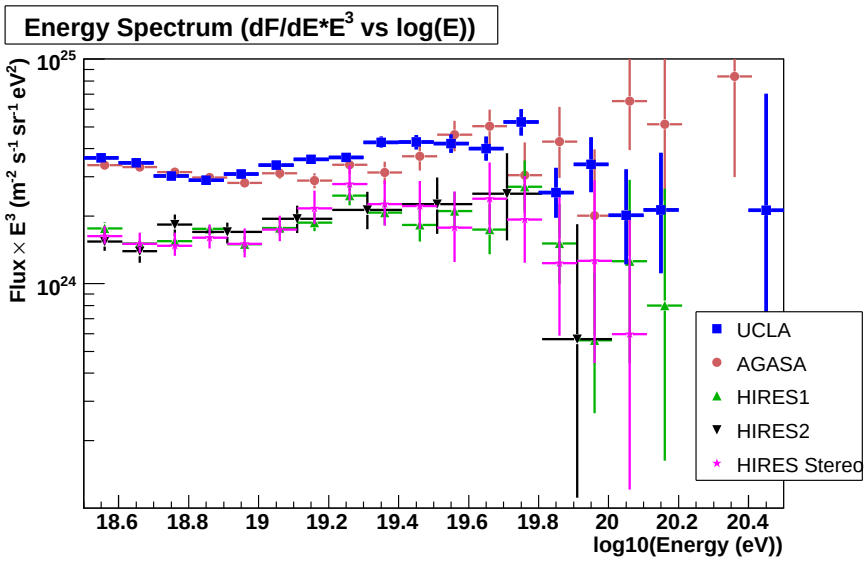


Figure 7.12: Spectrum with Proton Sibyll energy converter. Absolute flux and shape closely resembles those of AGASA spectrum.

QGSJet. The relation S(1000) $\sim$ $E^{0.95}$ is assumed for all 4 MC combinations.

The spectrum with the Proton Sibyll energy converter is shown in Fig. 7.12. This spectrum is very close to the AGASA spectrum in the absolute flux as well as the shape. The difference, however, is in the apparent drop-off beyond the GZK cutoff energy seen in the Proton Sibyll spectrum.

Proton QGSJet yields a spectrum with lower flux than Proton Sibyll, between the AGASA and HiRes spectra (see Fig. 7.13). It looks more consistent with the GZK cutoff than Proton Sibyll as the highest energy bins are tailing off more prominently.

The Iron Sibyll spectrum looks somewhat similar to the Proton QGSJet spectrum (see Fig. 7.14) but lower in flux. It also lies between the AGASA and HiRes spectra, and the highest energy bins tail off. The Iron QGSJet spectrum closely resembles the HiRes spectra except that the flux is higher than that of the HiRes

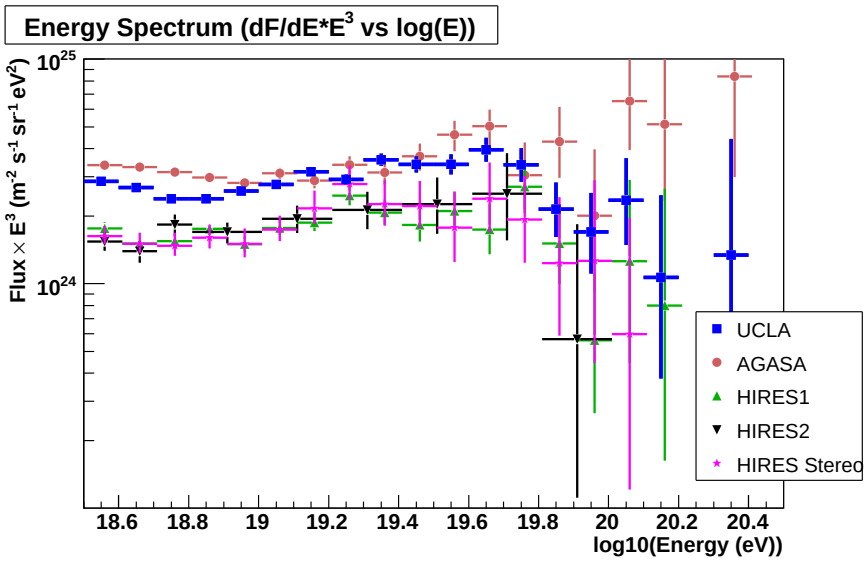


Figure 7.13: Spectrum with Proton QGSJet energy converter. Absolute flux lies between AGASA and HiRes spectra with a hint of GZK cutoff in high energy bins tailing off.

spectra (see Fig. 7.15).

Proton Sibyll produces the highest flux and Iron QGSJet the lowest as expected from the fact that Proton Sibyll gives the largest energy and Iron QGSJet gives the lowest for the same S(1000). It is apparent from this study that the change in the absolute energy scale can reproduce either the AGASA or the HiRes spectra. It is also clear that regardless of the energy converter used, the spectra all exhibit tailing off of the spectrum beyond the GZK cutoff energy. This was also seen in the CIC-based spectra. The GZK-like feature seems to be for real, and the unabated continuation seen in the AGASA spectrum seems to be incorrect. We will quantify the significance of this in chapter 8. What is also interesting is that regardless of the choice of the energy converter, be it a MC based converter or CIC based one, the dip in the spectrum at around $10^{18.8}$ eV

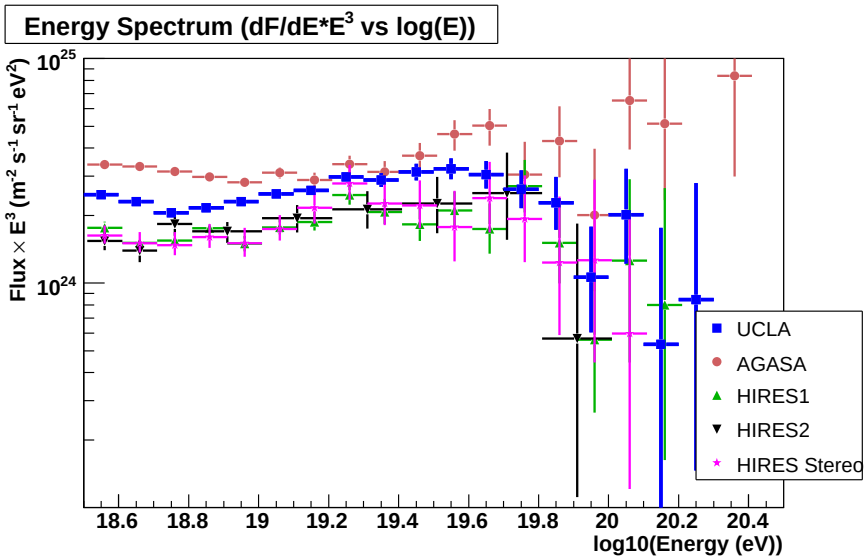


Figure 7.14: Spectrum with Iron Sibyll energy converter. Absolute flux lies between AGASA and HiRes spectra. It somewhat resembles the Proton QGSJet based spectrum.

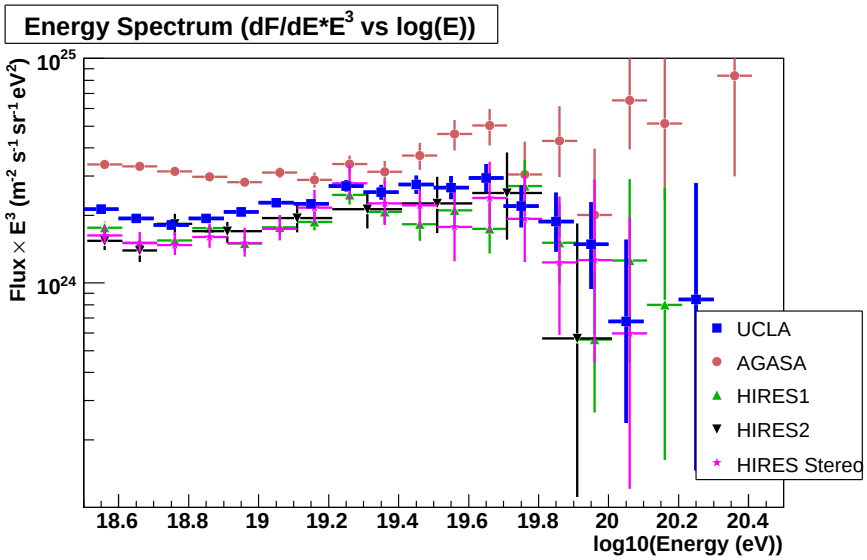


Figure 7.15: Spectrum with Iron QGSJet energy converter. Absolute flux is close to HiRes spectra with a hint of GZK cutoff in high energy bins tailing off.

(or some other energy depending on the choice of the energy converter) is present in all the spectra. For that matter, HiRes and AGASA also exhibit the dip. It is quite clear that the dip, also commonly known as the ankle, is not some artifact but a real feature. This is where the extragalactic component of the cosmic rays take over from the galactic component [8]. A successful theory of the origin and the propagation of the cosmic ray will have to be able to explain the origin of the ankle.

7.6 Consistency Check

The consistency of CIC is checked in this section. To elaborate, we wish to check whether forcing constant intensity at 10 EeV, as we have done, results in constant intensity at other energies as well. The level of agreement of the 4 MC energy converters with constant intensity is also examined.

7.6.1 Consistency check of CIC

CIC, by definition, produces constant intensity at around 10 EeV. But it is unclear whether forcing constant intensity at 10 EeV will also produce constant intensity at other energies. To check the consistency of the CIC method, the spectrum is divided up into three groups in zenith angles, 0-30°, 30-45°, 45-60°. Since these bins are equal in $\sin^2(\theta)$, hence exposure, the 3 resulting spectra should agree with one another.

Such plots are shown in Fig. 7.16 and Fig. 7.17. Both MC+CIC and FD+CIC indicate forcing constant intensity at 10 EeV results in constant intensity at all energies. The 3 spectra for each method agree with one another within statistical fluctuations from 3 EeV to the highest energies. This indicates that choosing the

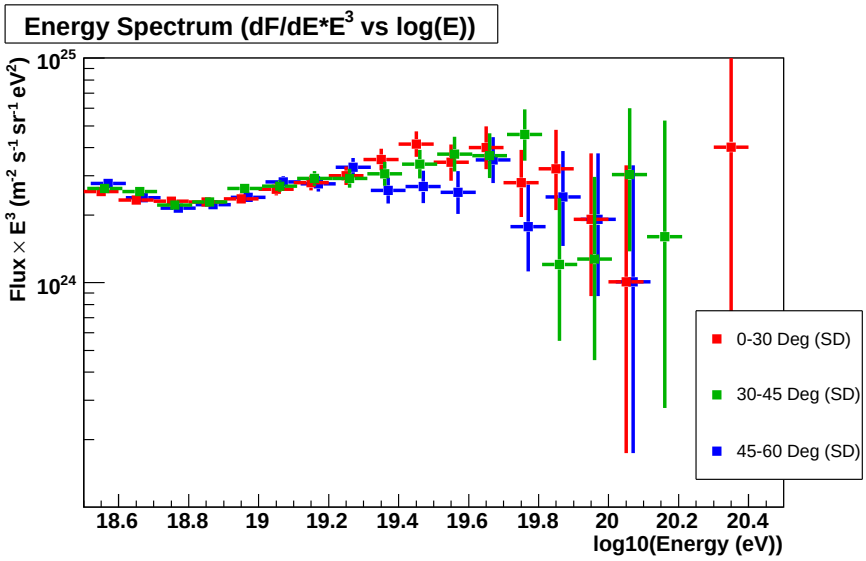


Figure 7.16: MC+CIC based spectrum broken up into three groups, 0-30°, 30-45°, and 45-60°, equal in $\sin^2(\theta)$. The number of events in each differential spectrum are consistent with one another.

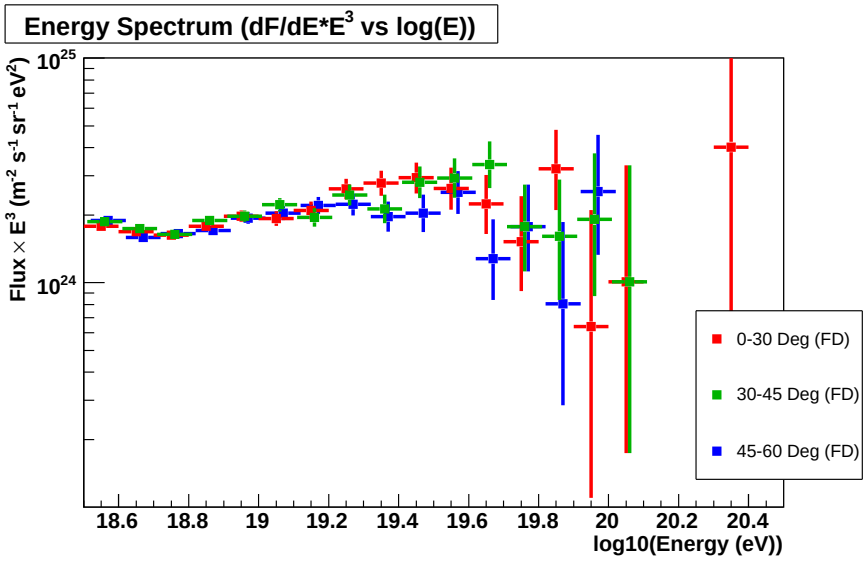


Figure 7.17: Same as Fig. 7.16, except energy is based on FD+CIC. The number of events in each differential spectrum are consistent with one another.

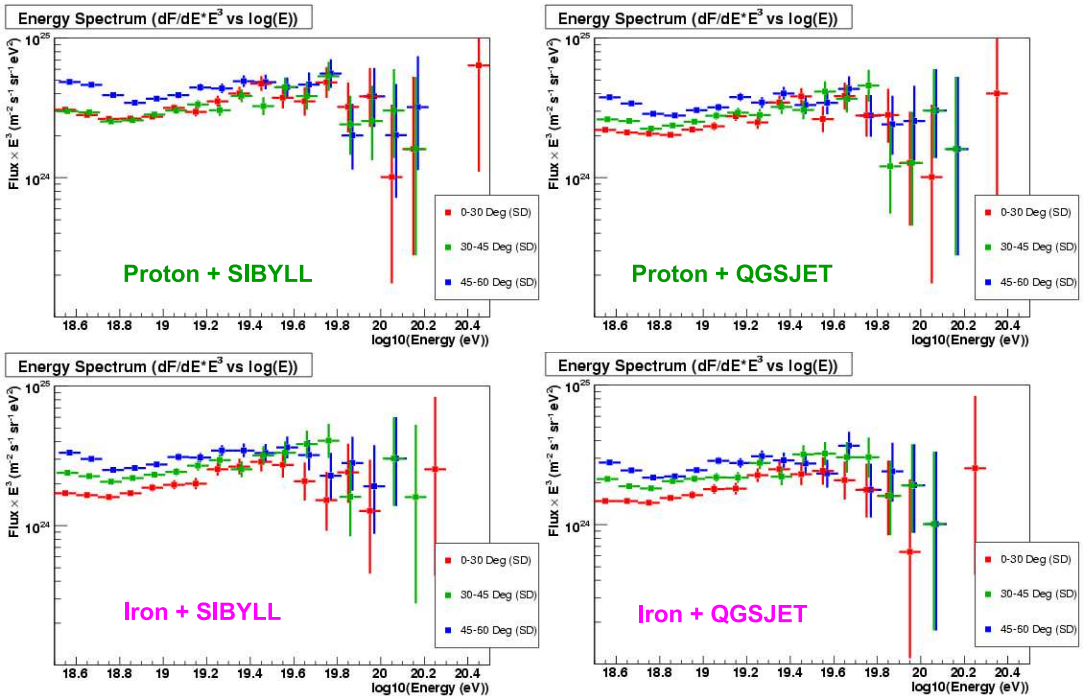


Figure 7.18: 4 spectra calculated with 4 pure MC energy converters divided up into 3 bins, 0-30° (red), 30-45° (green), and 45-60° (blue).

150th event cut constant intensity curve as the reference S(1000) attenuation for all energies results in no discernable systematic errors as the 3 spectra agree with one another. If there were any significant difference between the shape of the S(1000) attenuation curves at different energies, then our method of using the 150th cut constant intensity curve would have resulted in non-constant intensity at other energies. There seems to be, however, a hint of divergence at high energies, as 45-60° has the smallest flux for a series of high energy bins. They are, however, all within statistical fluctuations so until we collect more statistics, they are deemed consistent with one another.

7.6.2 MC vs. Real Data

The strongest argument for using CIC is that none of the MC simulations can reproduce the real data — the number of events is not constant as a function of $\sin^2(\theta)$ for a given energy. The behavior of the 4 primary/model combinations in zenith angle for > 3 EeV are shown in Fig. 7.18. The energy for each real data event is calculated using the MC energy converters. The data is divided up into 0-30°, 30-45°, and 45-60°, equal in $\sin^2(\theta)$. If a MC energy converter produced constant intensity, the three groups should contain more or less the same number of events at all energies.

Systematic uncertainties in energy is smallest at 30-45°, therefore the 30-45° should give us the most accurate spectrum(see chapter 6). Therefore, we take 30-45° as a reference and compare the other two groups to it. The level of agreement between the three groups should shed light on the present state of the MC simulations.

All 4 MC combinations exhibit some common behaviors. There are more events in 45-60° than in 0-30° or 30-45° for all 4 MC combinations, especially at low energies. At 45-60°, the EM part of the shower should dissipate to a large degree, and muons become the dominant component. This excess at 45-60° is due to the fact that the MC packages produce less muons than seen in the real data (MC curves falls off steeper than CIC curves), thus the reconstructed energy is assigned a higher value than it should be for the inclined showers — due to the steeply falling power law nature of the spectrum, a small fraction of low energy events reconstructed to high energy adds significantly high energy event statistics. Iron QGSJet is closest to constant intensity as its 30-45° and 45-60° spectra agree the closest. This implies that the real data is even more muon rich than Iron QGSJet, which produces the largest amount of muons. Fig. 7.18

confirms what was seen in Fig. 6.16 — In order to reproduce constant intensity with MC simulation, more muons are needed in MC simulation, and X_{max} needs to be proton-like.

It is possible that the reason that the MC simulations cannot exactly reproduce the real data is due to the fact comic rays may be of mixed composition. But all 4 primary/model's have a muon deficit relative to the real data, so no matter what combination of primary/model's one chooses to form a mixed composition, MC simulations cannot reproduce the real data. In the event that theorists find it impossible to fine tune MC simulations to produce more muons and more proton-like X_{max}, this may be an indication of some new unknown physics process.

7.7 Monte Carlo Spectrum

Although the uncertainties associated with the spectrum, for example in energy determination and calibration, are known to a reasonable degree, deconvolving the input spectrum from the reconstructed spectrum is a difficult matter. A Monte Carlo (MC) spectrum is an important tool in understanding how the reconstructed spectrum is altered from the input spectrum. With an MC spectrum, one knows exactly all the input parameters, such as energy and the direction vector, unlike with real data. In the past, MC studies in the collaboration have been done with MC events generated at discrete zenith angles and energies, and the behavior within the range covered by MC were interpolated (see chapter 6, for example). Of course, cosmic rays cover a continuous spectrum, so any short comings in previous MC studies due to their discrete nature can be checked and improved upon by comparing a continuous MC spectrum and the resulting spectrum after reconstruction.

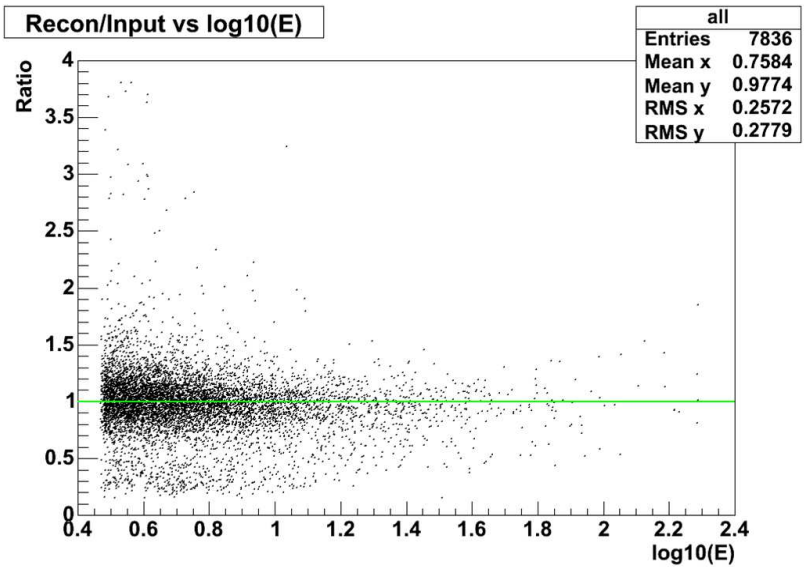


Figure 7.19: Ratio of reconstructed and input Energy. Each dot represents each event. The green line is inserted to guide the eyes.

7.7.1 Input vs. Reconstructed Energy

The MC spectrum was generated using Proton QGSJet showers for this study. All in all, 7200 Events covering > 3 EeV and 0-80° zenith were generated. Even though only events with the input and reconstructed zenith angles between 0° and 60° were used to plot the input and reconstructed spectra, events in 60-80° zenith were generated to allow for the possibility of events in that range being reconstructed to below 60°. The reconstruction chain in Offline that is used for real data is used for the MC spectrum events, and the energy converter is the one that was obtained previously in the aforementioned discrete MC studies (see Fig. 6.16). Any shortcomings in the previous energy determination method deriving from the discrete MC studies will manifest itself in a distortion of the reconstructed MC spectrum.

The difference in input and reconstructed energies for each event is examined

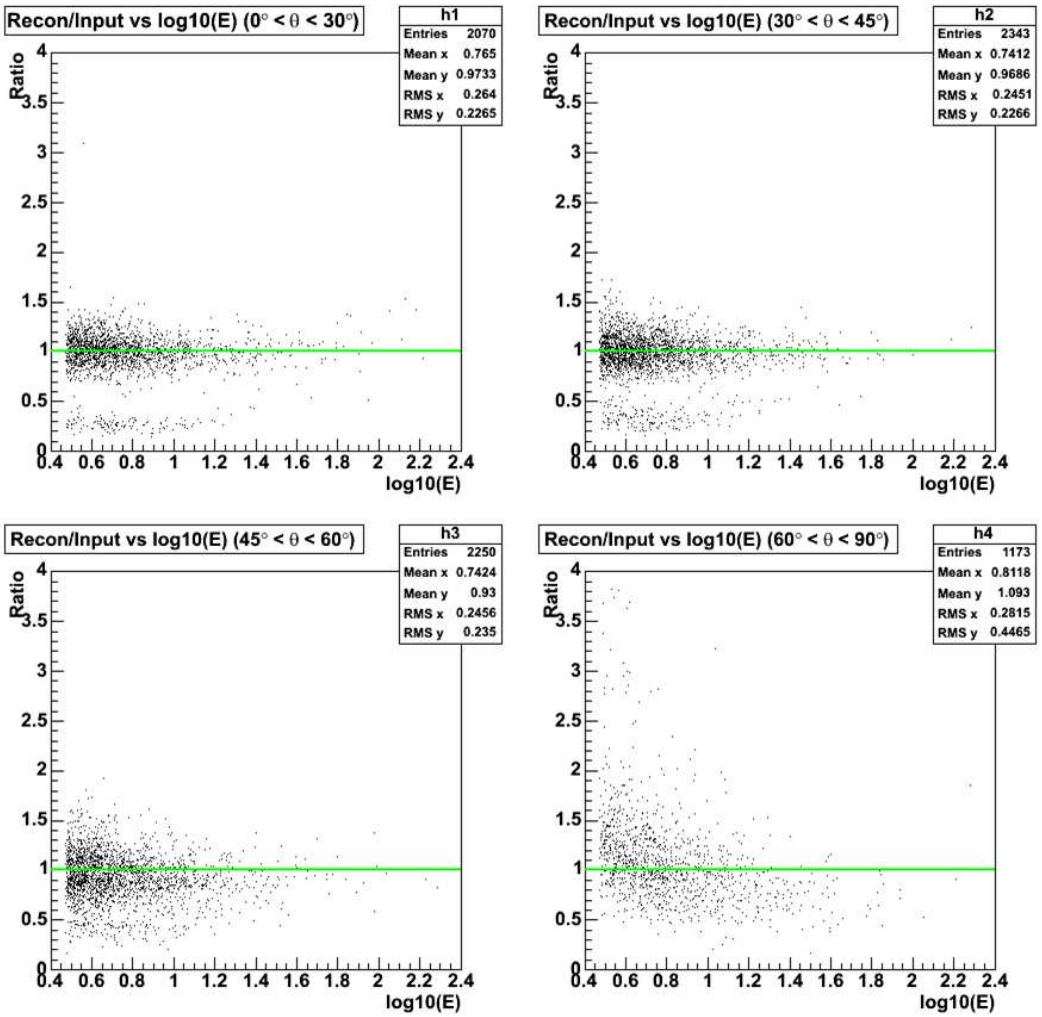


Figure 7.20: Ratio of reconstructed and input energy in 4 zenith bins, 0-30°, 30-45°, 45-60°, and 60-90°.

in Fig. 7.19. As expected, at low energies (around 3 EeV), there is a large scatter. There is a quite a few non-Gaussian tail events, those with underestimated and overestimated energies, at around a ratio (reconstructed energy/input energy) of 0.5 and beyond 2, respectively. The scatter becomes narrower with energy signaling the reconstruction becoming more stable with an increase in energy. Non-Gaussian tails at low and high ends also decrease with energy. However, beyond $\sim$ 40 EeV the trend reverses and the spread widens again. This is probably due to saturation of the PMT's resulting in inaccurate signals, causing relatively large reconstruction error — the effect of saturation should become significant at high energies.

The cosmic ray spectrum is a steeply falling spectrum with $\sim E^{-3}$ dependence. Therefore, underestimation of energies is, though undesirable, not as problematic as overestimation. Those events with underestimated energies will be pushed into lower energy bins where they will be only a small fraction of the events in those bins. Those events with overestimated energies could, however, very well dominate the bins to which they spill over. As the Auger collaboration and the cosmic ray community as a whole are interested in the existence or absence of the GZK cutoff and super-GZK events, we should be wary of those misreconstructed events with overestimated energies creating artificial super-GZK events.

The ratio of reconstructed and input energies are broken up in to 4 different zenith bins in Fig. 7.20. When broken up into different zenith bins, it becomes more apparent that the energy reconstruction improves with energy as for 0-60° the scatter becomes much tighter and converges around 1, meaning the reconstructed energy is very close to the input energy. Fig. 7.20 also shows beyond 60°, due to the fact that the EM part of the shower dissipates to a large degree and only muons survive, energy determination is quite difficult, as expected. The

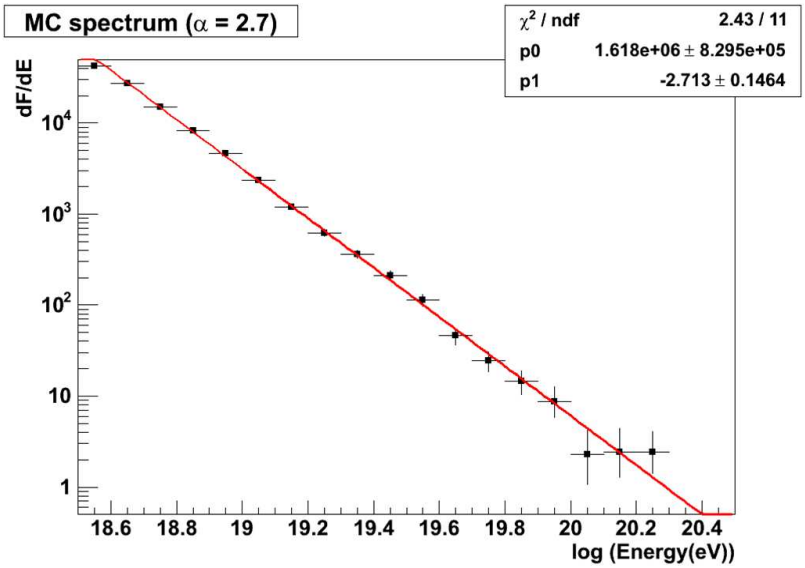


Figure 7.21: MC Input spectrum generated with $\alpha = 2.7$. The only uncertainty is from the statistical fluctuations due to the finite size of the data sample. A power law fit yields $\alpha = 2.71 \pm 0.14$, consistent with 2.7, as expected.

scatter is quite large and the convergence around 1 is not seen for the beyond 60° case. All three bins below 60° also show a fraction of events, with underestimation of energy at low energies. The distance between the non-Gaussian tails with low ratios and the ratio of 1 (green line) is largest at 0-30° and smallest at 45-60°, indicating the difficulty of reconstructing low energy vertical showers due to low multiplicity in those events. The three zenith bins below 60° all exhibit the same behavior without any grossly large difference in one particular bin. This gives us reasonable confidence that the energy reconstruction works more or less equally well in all three bins.

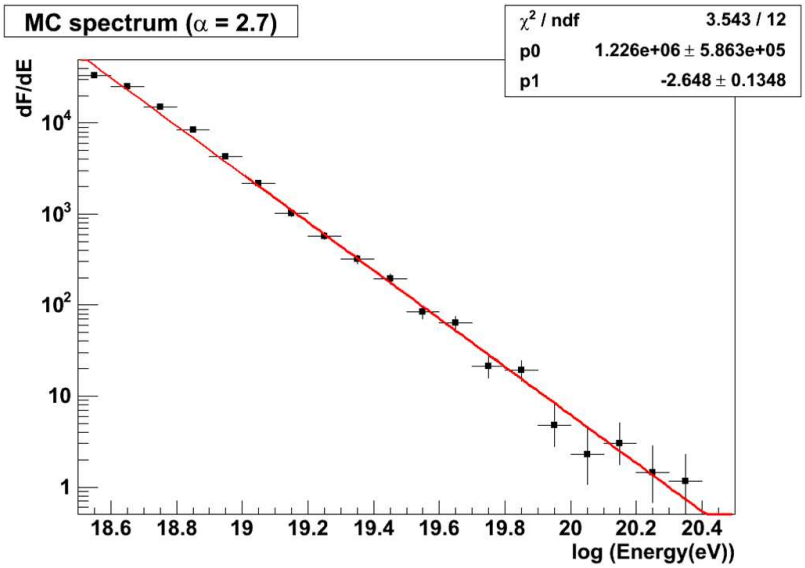


Figure 7.22: Reconstructed spectrum of the MC input spectrum. A power law fit yields $\alpha = 2.65 \pm 0.13$, consistent with 2.7.

7.7.2 MC Spectrum

The MC spectrum used for this study was generated with $\sim E^{-\alpha}$ dependence with $\alpha = 2.7$. The input spectrum is shown in Fig. 7.21. The only uncertainty in the spectrum is the statistical fluctuations in each bin due to the finite size of the MC data sample. A power law fit to the input spectrum above 10 EeV yields $\alpha = 2.71 \pm 0.14$, consistent with 2.7, as expected. The 10 EeV cut is used since when theoretical spectra are fit to the data, the fits usually are done for 10 EeV and above. 10 EeV is used as this is the energy above which the extragalactic component of the cosmic rays is thought to take over, and below which the galactic component is thought to be dominant. Also, 10 EeV is where the multiplicity of the vertical event becomes sufficient for reliable reconstruction of events.

The spectrum plotted with the reconstructed energy is shown in Fig. 7.22. The reconstructed spectrum closely resembles the input spectrum at first glance. A power law fit to the reconstructed spectrum yields $\alpha = 2.65 \pm 0.13$, which is consistent with $\alpha = 2.7$. But a closer inspection reveals a key difference between the two spectra. The reconstructed spectrum has an extra event in the log(E) = 20.3 bin that is not present in the input spectrum. As mentioned earlier this is the spillover from a lower energy bin, and this is dangerous as this kind of behavior can create artificial super-GZK events if some low energy events are assigned significantly higher energies. The energy resolution, however, above 10 EeV is around 20% or better, and even those events in the non-Gaussian tail that are assigned higher energies are within 50% of the input energy, so in all likelihood, a super-GZK event is really a super-GZK event, albeit with some uncertainty in energy. Except for some minor differences in the high end tails, the input and reconstructed spectra agree with each other.

The effect of spillover can be seen more clearly in Fig. 7.23 (and Fig. 7.24). The spectrum was plotted using the reconstructed energy of the events with the input energy less than $10^{19.6}$ eV ($\sim$ 40 EeV). There are some events beyond $10^{19.6}$ eV bin, these are the kind of events that could result in artificial super GZK events, in the worst case scenario — A comparison with Fig. 7.22 suggests, however, that the spillover events are only a small fraction of events. As mentioned earlier, the same kind of spillover to the lower energy bins, however, would have a minimal effect as they would be overwhelmed by the large statistics in those bins. The spillover results in the smearing the spectrum as seen in the slight disagreement between the spectral powers obtained from the input and reconstructed spectra. A good energy resolution as well as a good understanding of the energy resolution is, therefore, imperative.

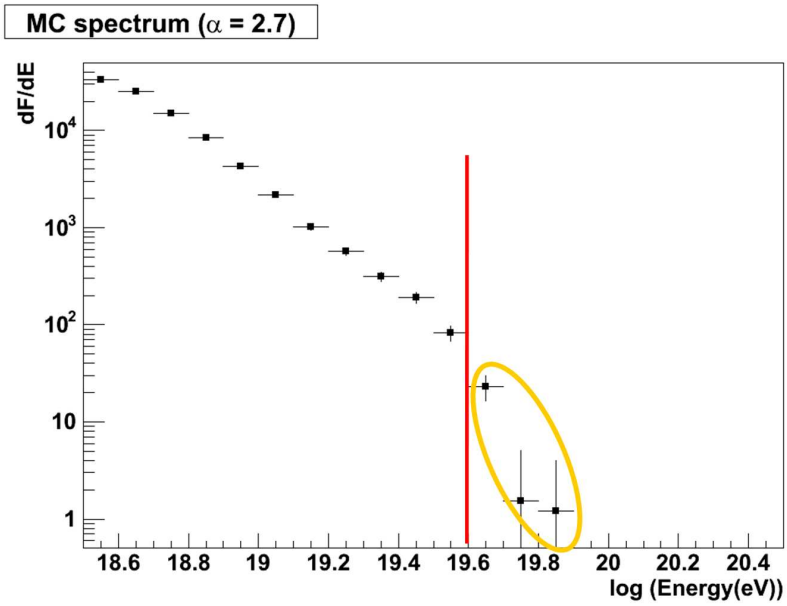


Figure 7.23: Reconstructed spectrum using only those events with input energy less than $10^{19.6}$ eV. There is some spillover into bins larger than $10^{19.6}$ eV.

It is a common practice in the cosmic ray community to plot the spectrum by multiplying by E^3. Since the spectra are multiplied by E^3, for convenience, we define and use $\alpha' = 3 - \alpha$, instead of α (Note: $E^{-\alpha} \cdot E^3 = E^{3-\alpha} = E^{\alpha'}$. For the input spectrum $\alpha = 2.7$, so $\alpha' = 0.3$). Power law fits to the two spectra multiplied by E^3 (Fig. 7.25 and Fig. 7.26), above 10^{19} eV again, yield $\alpha' = 0.28 \pm 0.07$ (input) and $\alpha' = 0.33 \pm 0.07$ (reconstructed), consistent with each other and the expected α' of 0.3. When the input and reconstructed spectra are multiplied by E^3, however, the differences between the two spectra are magnified. And the contrast in the shape of the spectra at the low energies are much starker – there is a small plateau at around $10^{18.9}$ eV and a steeper drop-off for the bins below in the reconstructed spectrum. The important lesson is that the E^3 spectrum, while great for human convenience, can deceive the human eye.

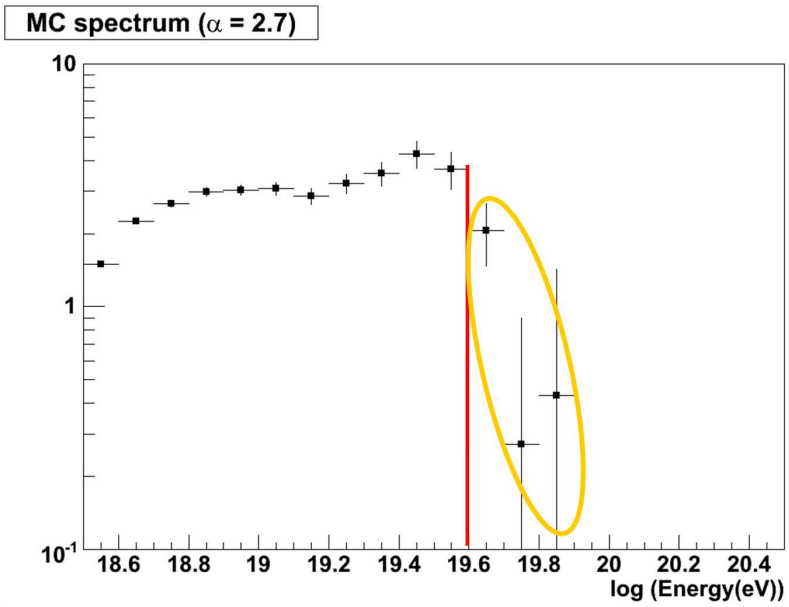


Figure 7.24: Reconstructed spectrum multiplied E^3 using only those events with input energy less than $10^{19.6}$ eV. There is some spillover into bins larger than $10^{19.6}$ eV.

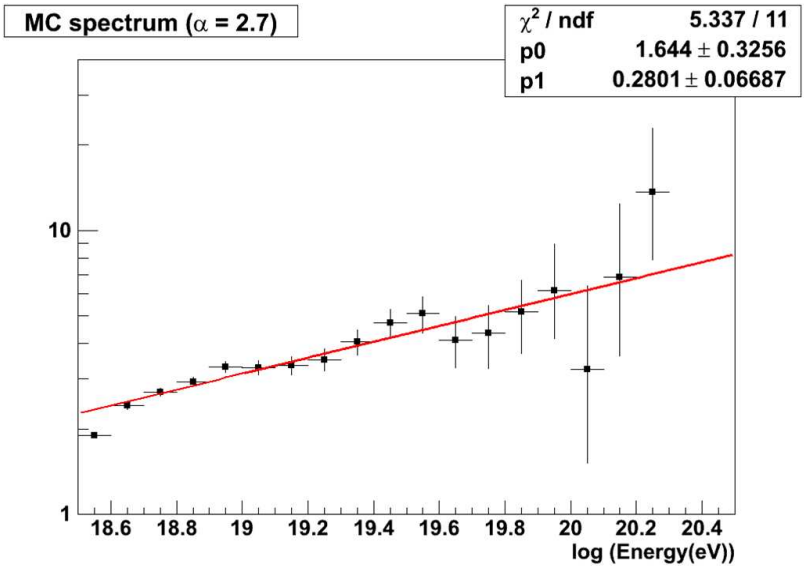


Figure 7.25: MC Input spectrum. Same as Fig. 7.21 except multiplied by E^3. A power law fit yields $\alpha' = 0.28 \pm 0.07$, consistent with 0.3, as expected.

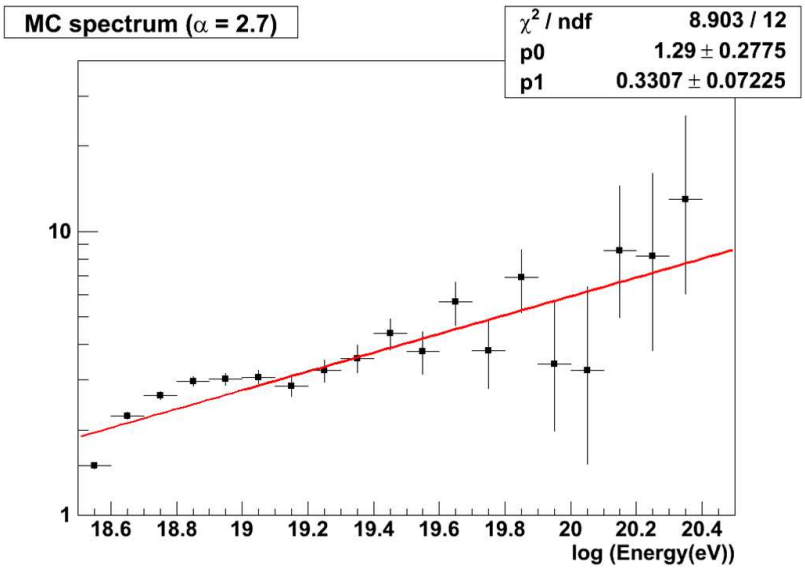


Figure 7.26: Reconstructed spectrum. Same as Fig. 7.22 except multiplied by E^3. A power law fit yields $\alpha' = 0.33 \pm 0.07$, consistent with 0.3, as expected.

7.8 Conclusion

A long standing mystery of the two contradictory spectra reported previously by HiRes and AGASA has now been resolved. The two different levels of absolute fluxes are also seen in the Auger spectra — the spectrum calibrated with MC simulation agrees with that of AGASA, and the spectrum calibrated with hybrid events agree with that of HiRes. The differences in the absolute fluxes of the two previous experiments were due to the systematics in energy, as was conjectured by some. Regardless of which energy converter is used, we see a GZK-like feature in the spectrum. The significance of this will be checked in chapter 8.

The next step forward would be to reduce the systematic uncertainties in both MC simulations and FD-based energies as they are both very large right now, $\sim 20\%$ and $\sim 30\%$ respectively as the large uncertainties mean that both MC-based and FD-bases spectra are in principle correct within error. It would

be desirable also to fine tune the MC packages in the way described earlier — proton-like in X_{max} and iron-like in muon richness. Once we have an MC simulation package that is accurate, it will free us from having to rely on a small subset of hybrid events for energy calibration which is especially dodgy at high energies as there is a lack of high energy hybrid events, and the composition could be energy dependent which would render the hybrid energy calibration at low energies inaccurate at high energies. That the MC packages cannot reproduce the observables is an indication that our current understanding of the physics processes is incomplete, and this presents an opportunity advance our knowledge.

CHAPTER 8

Theoretical Implications of the Auger Spectrum

8.1 Introduction

Soon after the discovery of the Cosmic Microwave Background Radiation (CMBR), Greisen, Zatsepin, and Kuzmin independently arrived at the conclusion that, if cosmic rays are protons, the cosmic ray spectrum should exhibit a suppression of flux beyond $\sim 6 \cdot 10^{19}$ eV [66, 67]. This is the so-called GZK feature. The reason for the GZK suppression is shown in Fig. 2.4. The proton primaries and the CMBR interact and the photopion production becomes significant at $\sim 6 \cdot 10^{19}$ eV. The energy loss length for the proton, consequently, drops precipitously from $\sim$1000 Mpc below $\sim 6 \cdot 10^{19}$ eV to $\sim$10 Mpc beyond 10^{20} eV. Iron (and other nuclei) undergo similar physics processes (photodisintegration) and suffer a reduction of the energy loss length in the same energy range (see Fig. 2.5). Therefore, cosmic rays above $6 \cdot 10^{19}$ eV would have to come from nearby, whether they are protons or irons. Previously, HIRES reported spectra that exhibit a GZK-like feature, whereas the AGASA spectrum did not exhibit such a feature [68, 69]. There are theoretical models that predict an absence of the GZK feature in the cosmic ray spectrum. These are the so-called Top Down models [70, 71, 72, 73]. The common thread among all these top down models are that they predict a significant flux of gamma primaries. Gamma ray flux studies by D. Barnhill *et al.* conclude that these top down models are strongly disfavored [75]. The aim of this

chapter is to do a direct comparison of the observed spectrum and of theoretical spectrum with the GZK feature as predicted by phenomenological source models by O. Kalashev, starting with pure proton and iron injections [80, 81]. We also examine the simple power law case to see if, indeed, we can exclude it.

8.2 Phenomenological Source Model

A brief description of the Phenomenological Source Model(PSM) is in order. PSM starts with a source distribution model and cosmic ray spectrum emitted at the sources following a simple power law. PSM is described by the following equation,

$$F(E,z) = f \cdot E^{-\alpha}(1+z)^{3+m}\Theta(E_{max}-E)\Theta(z-z_{min})\Theta(z_{max}-z)\ , \qquad (8.1)$$

where z is the redshift of the sources, and E denotes the energy, and Θ denotes the step function. The cosmic ray injection at the sources follows a simple power law dependence, $f \cdot E^{-\alpha}$, where f is an arbitrary constant. E_{max} denotes the endpoint of the spectrum, or the maximum energy of the cosmic ray at the sources. The sources are distributed according to $(1+z)^{3+m}$. The sources have a minimum and a maximum distance cutoff in the form of z_{min} and z_{max}. m is the so-called evolution factor, which accounts for the (possible) cosmological evolution of the sources. For instance, if $m > 0$, then the density of the sources increases with an increasing z (or equivalently further back in time or further away in distance). $m = 0$ means no cosmological evolution, and $m < 0$ means source the density decreases the further back in time one goes. The cosmic ray primaries from the PSM are then subjected to physics processes during propagation to earth (see Fig. 2.4 and Fig. 2.5), and finally we arrive at the theoretical predictions for the observed cosmic ray spectrum at earth [80, 81]. We will consider two cases — a pure proton injection case and a pure iron injection at the sources while varying

the 4 parameters (see Table 8.1).

8.2.1 α, E_{max}, z_{min}, and m dependence

A brief discussion on the effect of the 4 free parameters α, E_{max}, z_{min}, and m is now in order. For this discussion, we use the pure proton injection case, but the 4 parameters have the same effect on the pure iron injection case. The Default values for the 4 parameters are as follows — $\alpha = 2.5$, $E_{max} = 2.56 \cdot 10^{21}$ eV, $z_{min} = 0$, and $m = 0$. These values are chosen arbitrarily. Thus, the theoretical spectrum resulting from the 4 values above is not the best fit to the observed spectrum. Nevertheless, the theoretical spectrum closely resembles the observed spectrum. We will change one parameter at a time while holding the other 3 parameters constant to for a cursory look at the effects of the 4 parameters. We will try to get the best fit in later sections.

First, the dependence of the theoretical spectrum on α is shown in Fig. 8.1. Even though the injection spectrum at the sources goes as $E^{-\alpha}$, after cosmic rays undergo physics processes the theoretical spectrum does not resemble strict power law dependence any more (Note: For a comparison, refer to Fig. 8.17 for a look at unadulterated strict power law dependence). Still, it is clearly the case that the $\alpha = 2.9$ case (pink line) falls off faster than the $\alpha = 2.5$ case (red line), in line with the expectations of the $E^{-\alpha}$ dependence of the injection spectrum. The precipitous steepening seen in the theoretical spectra in Fig. 8.1 starting around $10^{19.7}$ eV are the GZK features. The observed spectrum is in very good agreement with the theoretical predictions from the low energies to the energies beyond the GZK cutoff energy.

The endpoint of the spectrum (E_{max}) dependence is shown in Fig. 8.2. E_{max} only affects very high energy components as there is no difference between the

different E_{max}'s below $10^{19.4}$ eV. The $E_{max} = 1.6 \cdot 10^{20}$ eV (red line) has a lower flux than the $E_{max} = 10^{22}$ eV (light blue line) at the high energies above the GZK feature.

As we saw earlier, cosmic rays with energies beyond the GZK cutoff energy interact with the CMBR and lose energy in propagation. In order to have a significant flux above $6 \cdot 10^{19}$ eV, the cosmic rays must, then, necessarily come from nearby sources. As z_{min} decreases, the distance to the nearest source allowed also decreases — hence, an increase in the number of nearby sources. Thus we expect more flux for energies beyond the GZK feature. As such, the $z_{min} = 0$ case has a higher flux beyond the GZK feature than the other cases with $z_{min} > 0$, as shown in Fig. 8.3.

As explained earlier, m has an effect of changing the relative density of the sources as a function of z (Remember, the source distribution follows $(1+z)^{3+m}$). The $m = 3$ case means higher density of sources in the past (further or high z) than in the present (closer or low z). The $m = -3$ case means higher density of sources now (closer or low z) than in the past (further or high z). As explained earlier, more sources at low z means more flux beyond the GZK feature, and more sources at high z means more flux at low energies. That is seen in Fig. 8.4 where the $m = -3$ case has a higher flux beyond the GZK feature than the $m = 3$ case, and the situation is reversed at low energies.

For the rest of this note, we will fix z_{min} and m at 0, where they seem to yield the best fit. We will only vary E_{max} and α for a best possible fit to the observed spectrum and leave a more complete scan of the phase spaces, including z_{min} and m, for the future.

Parameter	Name	Values
Power of the Injection Spectrum ($E^{-\alpha}$)	α	1.0-3.0
End point of the Energy Spectrum	E_{max}	Z·($10^{19} - 10^{22}$) eV
Evolution factor: $(1+z)^{3+m}$	m	-3, 0, 3
Red shift of the nearest source	z_{min}	0–0.1
Red shift of the farthest source	z_{max}	3

Note: Z is the charge of the nuclei.

Table 8.1: Parameters in the Phenomenological Source Model.

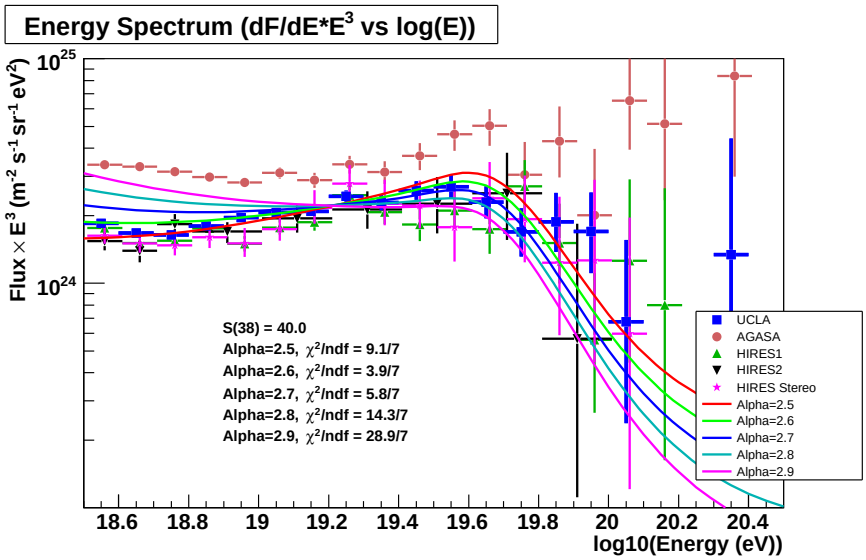


Figure 8.1: α dependence of the theoretical prediction. Red line, $\alpha = 2.5$ (or highest flux at high energies). Purple line, $\alpha = 2.9$ (or lowest flux at high energies)

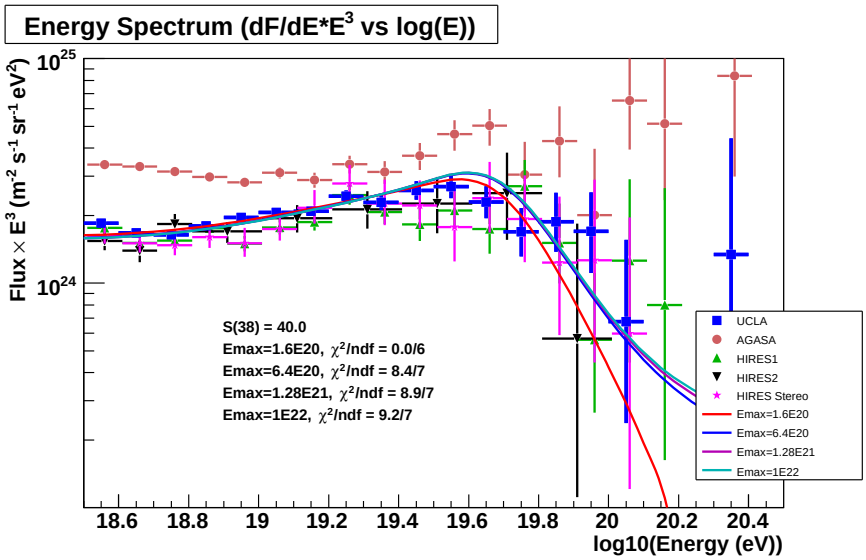


Figure 8.2: E_{max} dependence of the theoretical prediction. Red line, $E_{max} = 1.6 \cdot 10^{20}$ eV (lowest flux at high energies). Light blue, $E_{max} = 1 \cdot 10^{22}$ eV (highest flux at high energies).

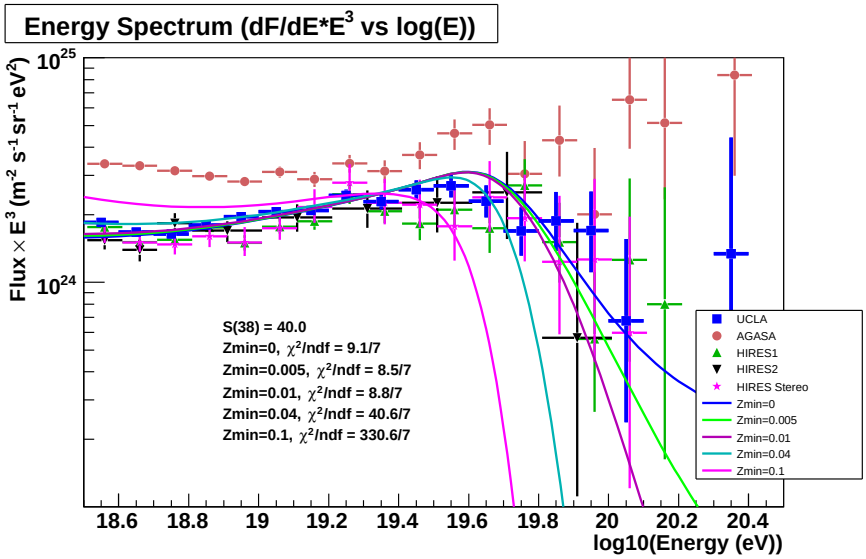


Figure 8.3: z_{min} dependence of the theoretical prediction. Purple line, $z_{min} = 0.1$ (spectrum terminates earliest). Blue line, $z_{min} = 0$ (spectrum extends furthest).

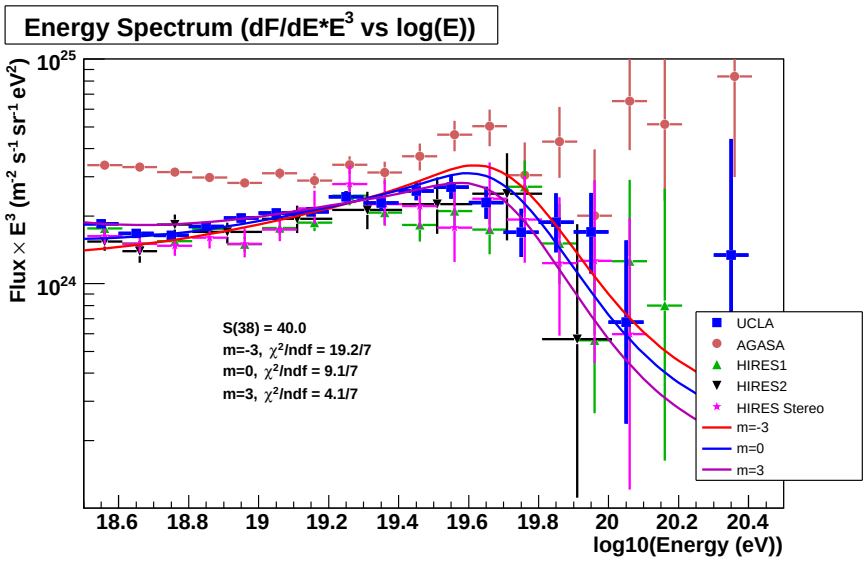


Figure 8.4: m dependence of the theoretical prediction. Red line, $m = -3$ (highest flux at high energies). Purple line, $m = 3$ (lowest flux at high energies).

8.3 Statistical Method

In this section we discuss the statistical method used to quantify the level of agreement between the theoretical predictions and the observed spectrum.

8.3.1 Binned Maximum Likelihood Function

We use the binned maximum likelihood method to fit the theoretical spectrum to the observed spectrum [82]. This is a natural choice since the cosmic ray spectrum has been traditionally plotted with a bin width of 0.1 in log(Energy). For each value of E_{max} and α, we get a theoretical spectrum. The absolute flux of the theoretical spectrum is arbitrary (f from Eq. 8.1), and is the only free parameter in the fit to the observed spectrum. We only use the data points above 10^{19} eV for the fit to minimize the effect of the magnetic field and restrict the

range of the fit to the range where the extragalactic components are expected to dominate [17].

A brief discussion of the binned maximum likelihood method is in order. The number of events, n_i, in each bin is regarded as a Poisson distribution with the number of events expected, ν_i, as predicted by the theoretical prediction. The likelihood function is then,

$$L(\mathbf{n};\, \boldsymbol{\nu}) = \prod_i^N \frac{\nu_i^{n_i}}{n_i} e^{\nu_i} \ , \tag{8.2}$$

where N is the total number of bins.

As is customary, we maximize the log-likelihood function rather than the likelihood function

$$log\ L(\mathbf{n};\, \boldsymbol{\nu}) = \sum_i^N n_i\ log\ \nu_i - \nu_i \ , \tag{8.3}$$

where constants have been dropped.

Although this gives us the best possible fit, there is no way to tell how significant the result is. We need a goodness-of-fit test like that of χ^2 test for the Gaussian case. Luckily, there is a way to make use of the χ^2 test for the Poisson case. First we have to take the ratio $\lambda = \mathrm{L}(\mathbf{n};\, \boldsymbol{\nu})/\mathrm{L}(\mathbf{n};\ \mathbf{n})$. Then

$$\chi^2 = -2\ log\ \lambda = 2\ \sum_i^N\ (n_i\ log\ n_i - n_i\ log\ \nu_i + \nu_i - n_i) \tag{8.4}$$

follows the χ^2 distribution for $N - m$ degrees of freedom, for m parameters, in the ***large n limit***. -2 log λ does not behave like χ^2 when (some) n_i's are small.

One can overcome this problem by generating toy Monte Carlo (MC) simulation data sets with the theoretical predictions and apply the same analysis procedure. The fraction of the fake MC data sets with worse fit results is, then, the significance level (p-value). The inadequacy of the standard goodness-of-fit formula when the large n limit is not satisfied is evident in Fig. 8.5. The p-values

obtained from the standard χ^2 distribution, falsely assuming that the large n limit is satisfied, is compared with the p-values obtained from comparing against the fit results of the 25000 Monte Carlo data sets. The standard formula for p-value underestimates the real significance in the plot. The small number of events in the bins above $10^{19.7}$ eV causes -2 log λ to deviate from the χ^2 distribution.

On the downside, generating thousands of Monte Carlo data sets can be time consuming. The problem of small n can be solved by simply widening the binning hence ensuring each bin contains a large enough number of events. We implemented this method by treating all the energies above $10^{19.7}$ eV as one bin while keeping the other bin widths at 0.1. The result is that -2 log λ gives the expected p-value. It is identical to the p-value obtained from Monte Carlo data sets (see Fig. 8.6). The pathological behavior seen previously in the p-value that is caused by the small number of events above the GZK feature is eliminated by the consolidation of the bins above $10^{19.7}$ eV. Thus instead of generating thousands of Monte Carlo data sets for each E_{max} and α to calculate the p-value, we will treat all events above $10^{19.7}$ eV as one bin while keeping the bin width at 0.1 for the bins below.

Interested readers in the binned likelihood method and the goodness-of-fit associated with it are referred to Cowan [82] and Heinrich [83] for more details.

8.4 Proton Primary

We will examine the pure proton injection case in this section varying E_{max} and α as outlined earlier.

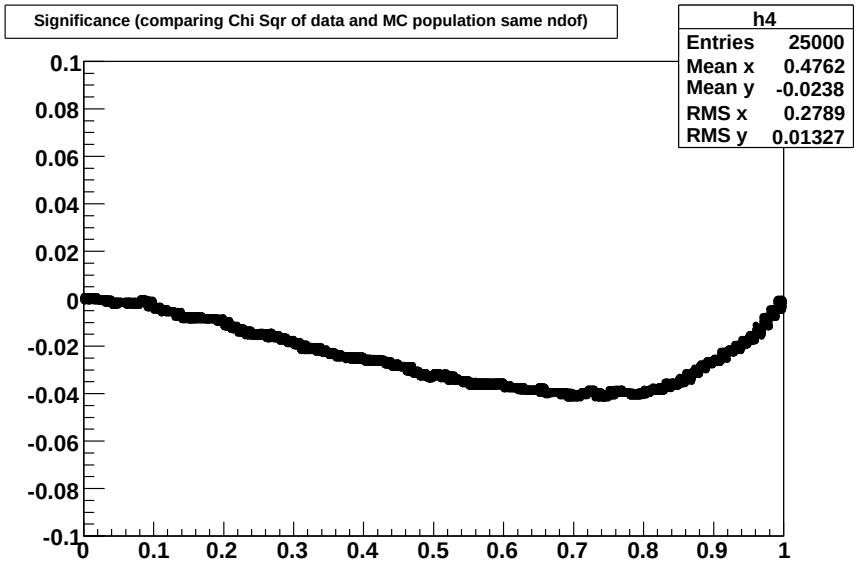


Figure 8.5: Inaccuracies in the goodness-of-fit as given by the standard formula for the χ^2 distribution when the number of events in some bins are too small. For a given χ^2 and degrees of freedom, the goodness-of-fit is calculated (x-axis). It is compared with the goodness-of-fit obtained by comparing the χ^2 with those of those of 25000 fake MC data sets. Y-axis is the difference of the two goodness-of-fit's. According to this plot, the standard goodness-of-fit formula for the χ^2 distribution underestimates the significance. For example, if the standard formula give a significance of 0.8, then 80% of the MC data sets should have a larger χ^2. Instead, 84% of the MC data sets give a larger χ^2. Hence, the real significance is 0.84.

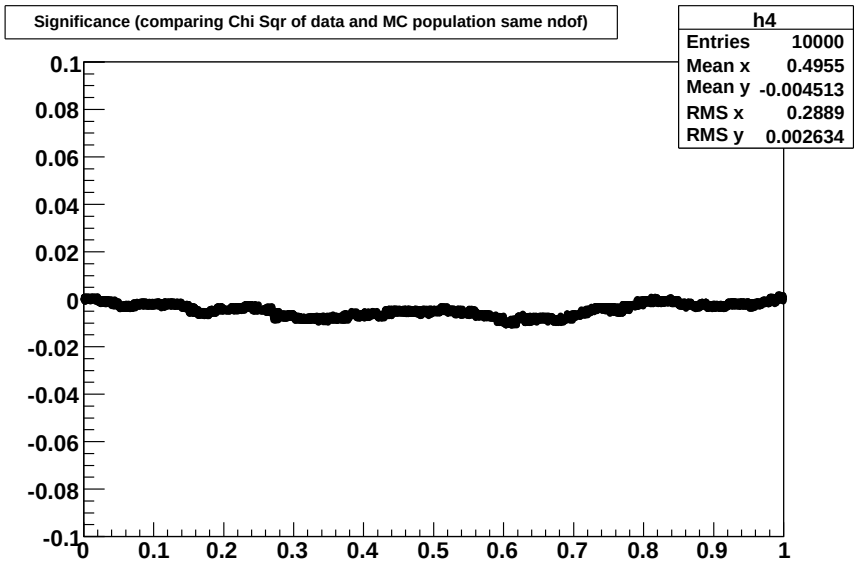


Figure 8.6: The inaccuracies in the previous plot is remedied by ensuring there is enough events in every bin by treating energy above $10^{19.7}$ eV as one bin. Now the two goodness-of-fit, calculated with the standard formula and MC method, agree, i.e. the difference between the two methods is 0.

8.4.1 Further Discussion on α Dependence

The dependence of the theoretical spectrum on α and E_{max} (also z_{min} and m) was discussed earlier. One salient feature in Fig. 8.1 is that the $\alpha = 2.9$ (pink line) falls off faster than the $\alpha = 2.5$ (red line). This is expected from the $E^{-\alpha}$ dependence of the spectrum. Another salient feature of the plot is that the difference between the $\alpha = 2.9$ and the $\alpha = 2.5$ cases is that the $\alpha = 2.9$ case has an excess of events well above the observed spectrum at E $< 10^{19}$ eV, whereas the $\alpha = 2.5$ case has a deficit. The deficit of the events below 10^{19} eV for the $\alpha = 2.5$ case is not as problematic as the $\alpha = 2.9$ case. We consider only the extragalactic component in the PSM and fit only above 10^{19} eV where the extragalactic component is thought to take over [17]. So there is some room to wiggle for the $\alpha = 2.5$ case as we can probably add the galactic component to match the observed spectrum below 10^{19} eV. The excess below 10^{19} eV seen in the $\alpha = 2.9$ case, however, is not as easy to deal with, as one would have to concoct a scenario to reduce the flux of the extragalactic component below 10^{19} eV match the observed flux below 10^{19} eV. To be fair, this excess is largely due to the fitting of spectrum only above 10^{19} eV. But if the fit were adjusted for the $\alpha = 2.9$ case to fit the low energy range better, that would result in a terrible fit with a large deficit at the high energies. On these grounds, $\alpha = 2.9$ (and higher) can be safely ignored.

8.4.2 Results

We varied E_{max} and α and fit the theoretical spectrum to the observed spectrum for each E_{max} and α. For each fit, we obtain the goodness of fit using the method outlined earlier. The result is shown in Fig. 8.7. The hatched region is a forbidden region, forbidden by E_{max} consideration, i.e. if $E_{max} = 10^{19}$ eV, then it is physically impossible to observe events above 10^{19} eV (not taking fluctuations

in energy measurements into account). Red indicates a good agreement, and blue indicates a level of disagreement to $\sim 5\sigma$ level. We can safely rule out a spectrum with $\alpha < 2.2$ regardless of E_{max}. The best agreement is for $\alpha \sim 2.5$–2.7. The E_{max} dependence is weak which is expected from the foregoing discussion (see Fig. 8.2). According to the plot, the observed spectrum agrees with the proton primary assumption with GZK cutoff and $\alpha \sim$ 2.6–2.8 to better than a 1σ level.

Now we allow for 25% systematic uncertainties in the energy determination, and see how that affects the level of agreement of the observed spectrum with the theoretical predictions. To do this, we take $E_{max} = 1.28 \cdot 10^{21}$ eV, where the significance is the largest, and for each α, we vary S38 from 30 to 50 VEM to allow for the 25% systematic uncertainties in energy and obtain the goodness-of-fit value. Once again, $\alpha < 2.2$ can be ruled out (see Fig. 8.8). However, Fig. 8.8 also implies that across the entire S38 range allowed by the systematic uncertainties, the observed spectrum agrees well with the theoretical spectrum to better than 2σ, thus the observed spectrum strongly favors the existence of the GZK feature. On the other hand, the wide range of S38 values over which the observed spectrum agrees well with the theoretical spectrum also means that a reduction in the systematic uncertainties in energy does not help us pinpoint the exact value of α.

8.5 Iron Primary

In this section, we study the pure iron injection case. While the justification for a pure iron injection at the source may be weak (One would have to worry about spallation at the sources. If anything, a pure proton or mixed composition case is probably more likely), there is no reason to rule out the iron injection case since there is no indisputable proof of the sources and the composition of the cosmic

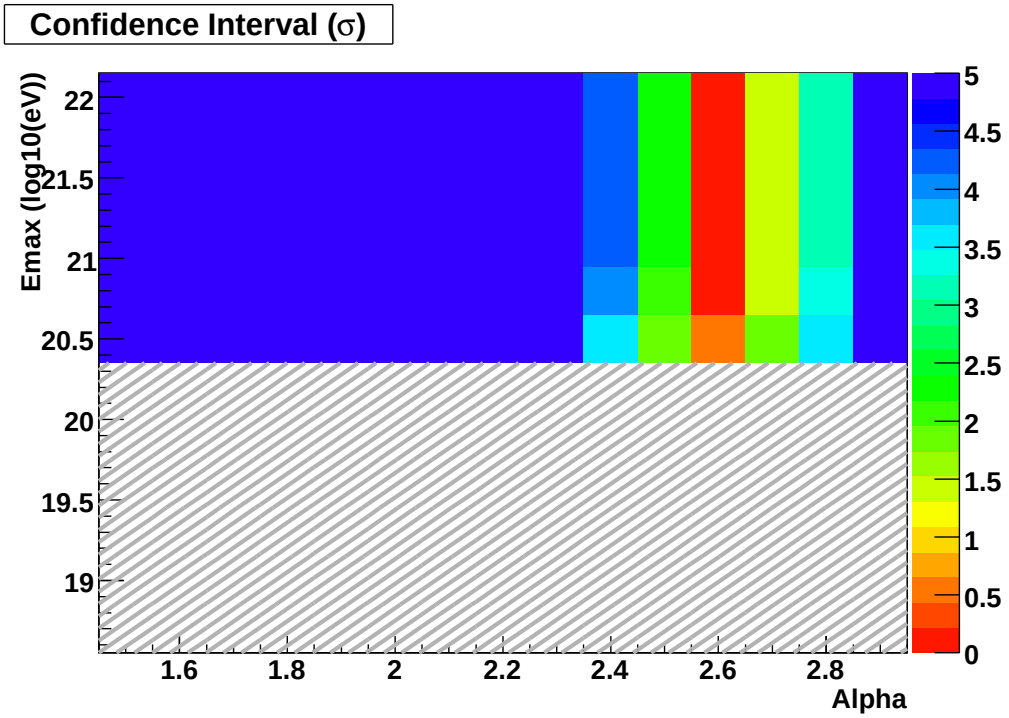


Figure 8.7: Contour plot of the significance of the goodness-of-fit test for each E_{max} and α of the pure proton injection case. A significance of 1 (Red) means the chance probability of getting a smaller χ^2 is 0. The most significant region is where $\alpha = 2.5$–2.7. The hatched region is forbidden by E_{max} considerations.

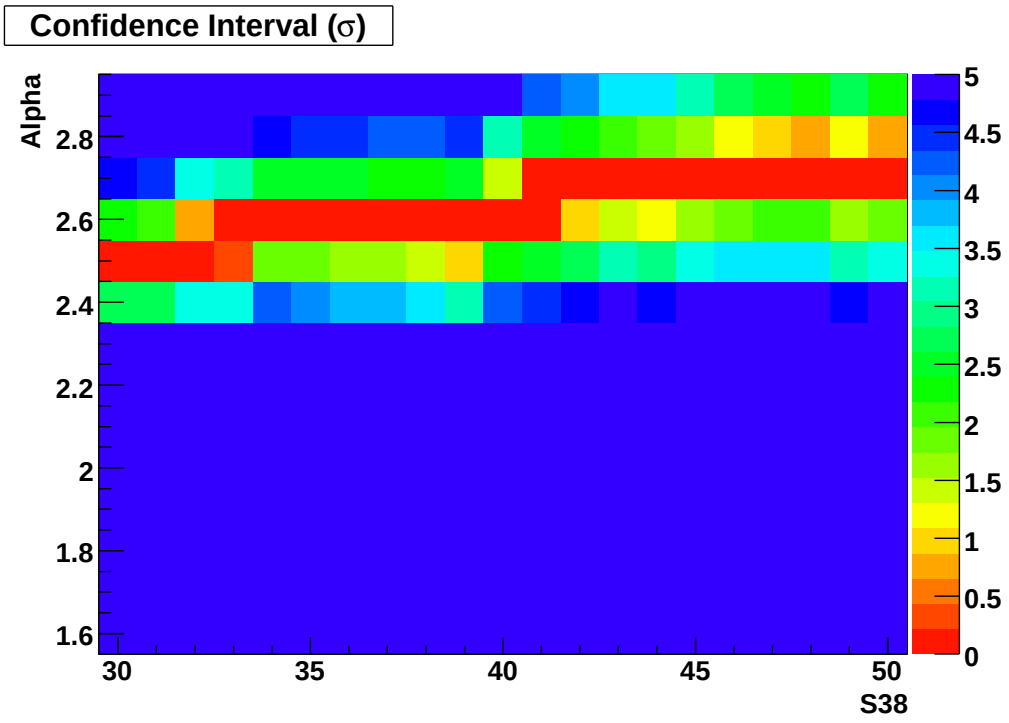


Figure 8.8: Significance plot for the pure proton injection case taking the systematic uncertainties in energy into account. S38 is varied from 30 to 50 VEM to allow for 25% systematic uncertainties in energy. Across the entire allowed S38 values, the observed spectrum is consistent with the theoretical predictions ($\alpha > 2.4$). $E_{max} = 1.28 \cdot 10^{21}$ eV.

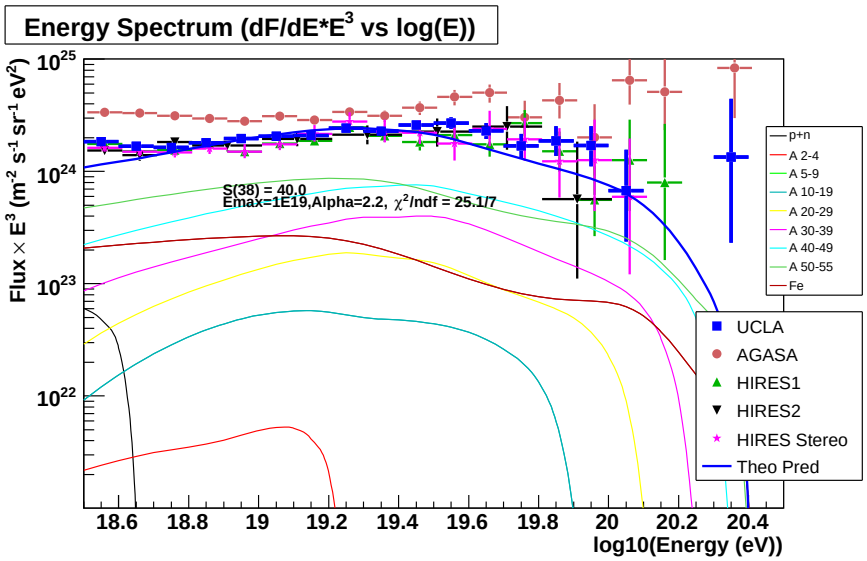


Figure 8.9: Theoretical spectrum for pure iron injection case. The spectrum is broken up, and different nuclei groups resulting from photodisintegration are shown. $E_{max} = 26 \cdot 10^{19}$ eV. Notice the small proton flux ending at 1/56th of E_{max} (Black line).

rays above 10^{19} eV. It should serve well at least as a limiting case. We compare the theoretical predictions from a pure iron injection with the observed spectrum.

8.5.1 Composition from Pure Iron Injection

From Fig. 2.5, it is clear that all the iron primaries will photodisintegrate beyond $10^{20.5}$ eV. But below $10^{20.5}$ eV, the energy loss length of iron is actually less than that of proton. The iron primary undergoes photodisintegration, and the resulting spectrum is composed of different nuclei. Unlike proton, the composition as well as the spectral shape change with a change in the E_{max}. That evolution is shown in Figs. 8.9–8.12. For $\alpha = 2.2$, we start with $E_{max} = 26 \cdot 10^{19}$ eV. For this case, the composition is heavy element dominant. The proton flux, from

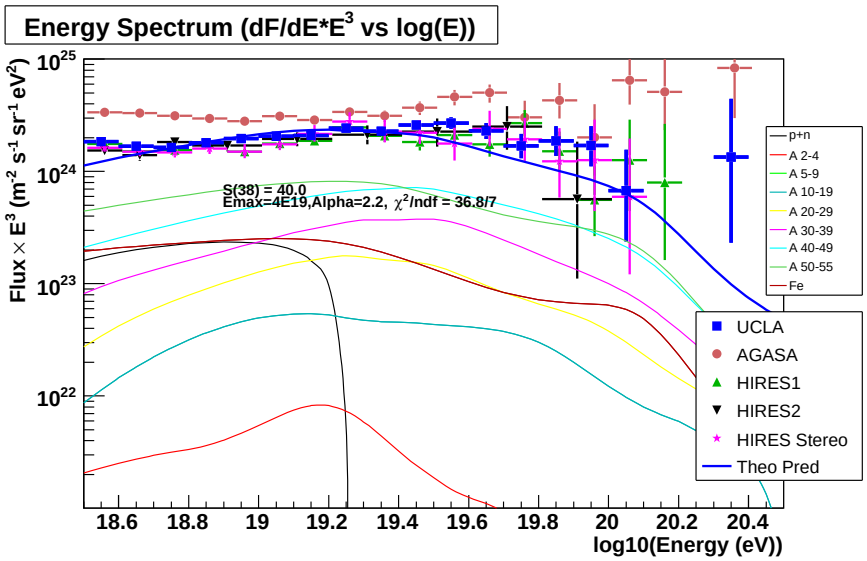


Figure 8.10: $E_{max} = 26 \cdot 4 \cdot 10^{19}$ eV. More photodisintegration proton flux and higher endpoint than the previous plot.

photodisintegration, is quite low and has an endpoint at 1/56th of the E_{max}. In subsequent plots, as the E_{max} increases, so does the proton flux as well as its endpoint. By the time E_{max} is equal to $26 \cdot 10^{20.5}$ eV, the proton flux is dominant. For $E_{max} > 26 \cdot 10^{20.5}$ eV, the proton dominance is only more pronounced.

Armed with the knowledge of the evolution of the resulting composition as function of E_{max}, next, we vary α and E_{max} and find the cases with the best agreement with the observed spectrum as we have done earlier for the pure proton injection case. The results are shown in Fig. 8.13. We can safely rule out cases with $\alpha < 1.6$ and $\alpha > 2.5$ regardless of E_{max}. Unlike the proton case, however, there are two regions of high significance in the iron case. Both the $\alpha = 2.1$–2.2 with $E_{max} > 26 \cdot 10^{20.5}$ eV and $\alpha = 1.9$–2.0 with E_{max} around $26 \cdot 10^{19}$ eV cases agree with the observed spectrum to better than 1σ. The two cases paint two different pictures of composition. The former is a proton dominant case where

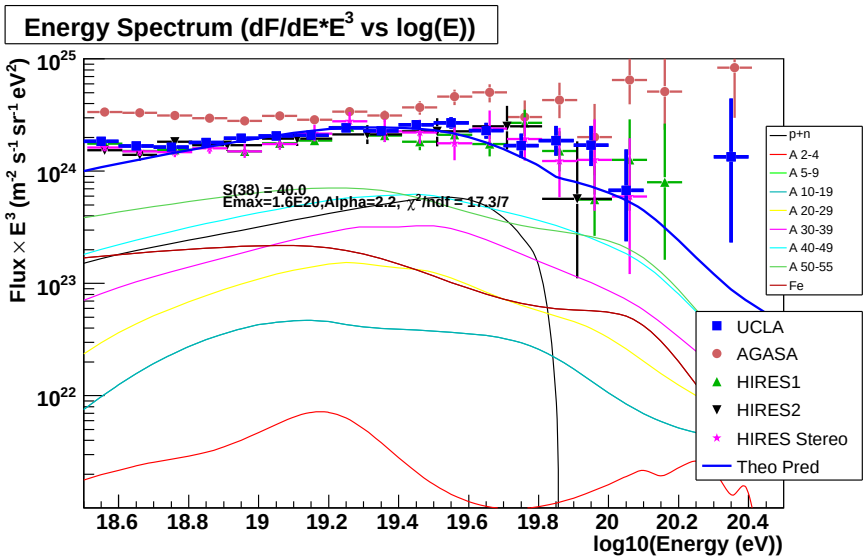


Figure 8.11: $E_{max} = 26 \cdot 1.6 \cdot 10^{20}$ eV. More photodisintegration proton flux and higher endpoint than the previous plot.

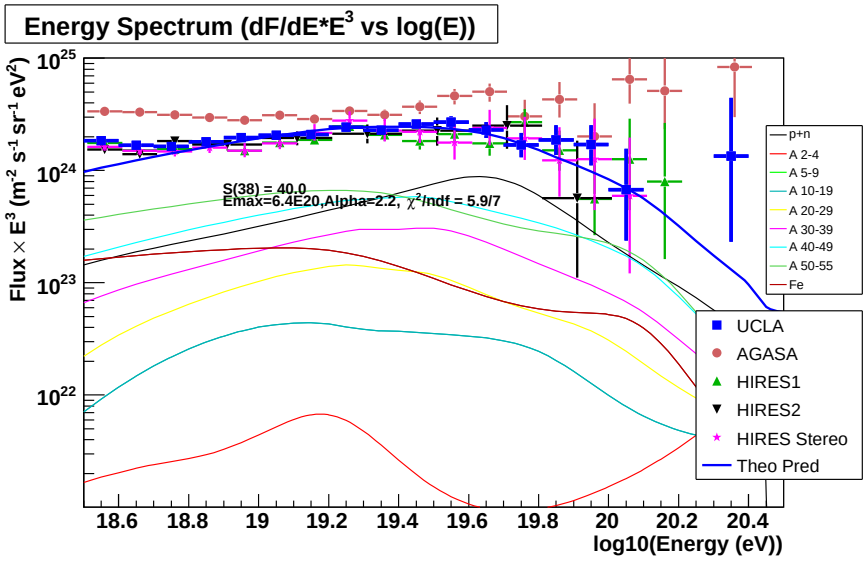


Figure 8.12: $E_{max} = 26 \cdot 6.4 \cdot 10^{20}$ eV. Protons begin to dominate at high energies (above $\sim 6 \cdot 10^{19}$ eV). Proton dominance is more pronounced for an even higher E_{max}.

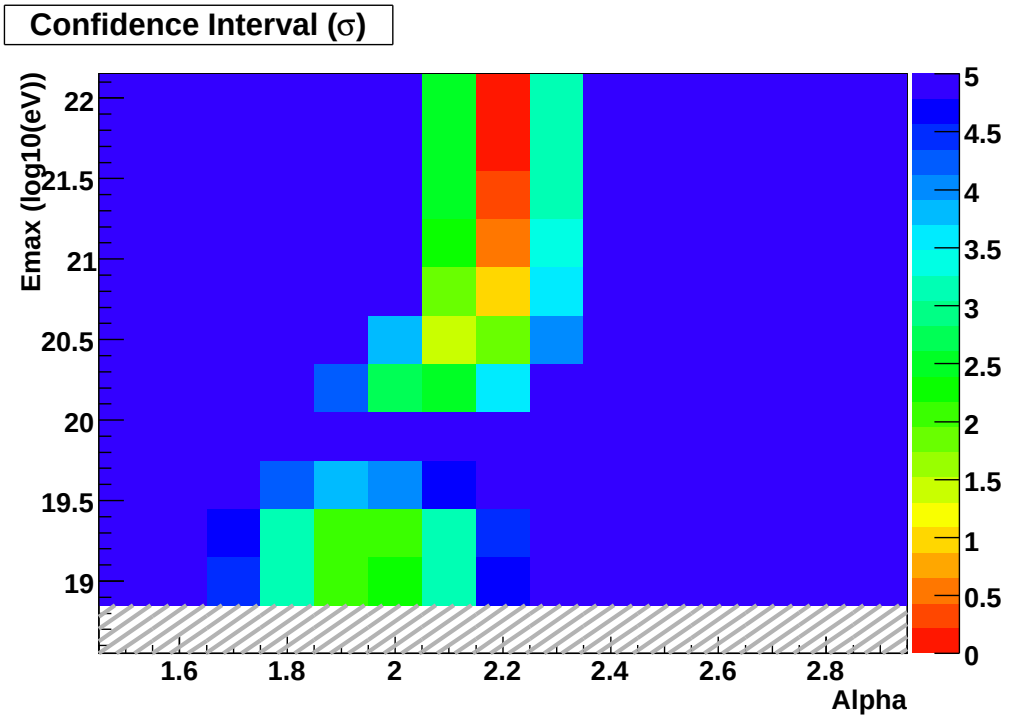


Figure 8.13: Significance plot for the pure iron injection case. There are two regions high significance. $\alpha \sim 2.0$ for $E_{max} \sim 26 \cdot 10^{19}$ eV and $\alpha \sim 2.2$ for $E_{max} > 26 \cdot 3 \cdot 10^{20}$ eV. Note: multiply the values on the y-axis by 26 to get the actual E_{max}.

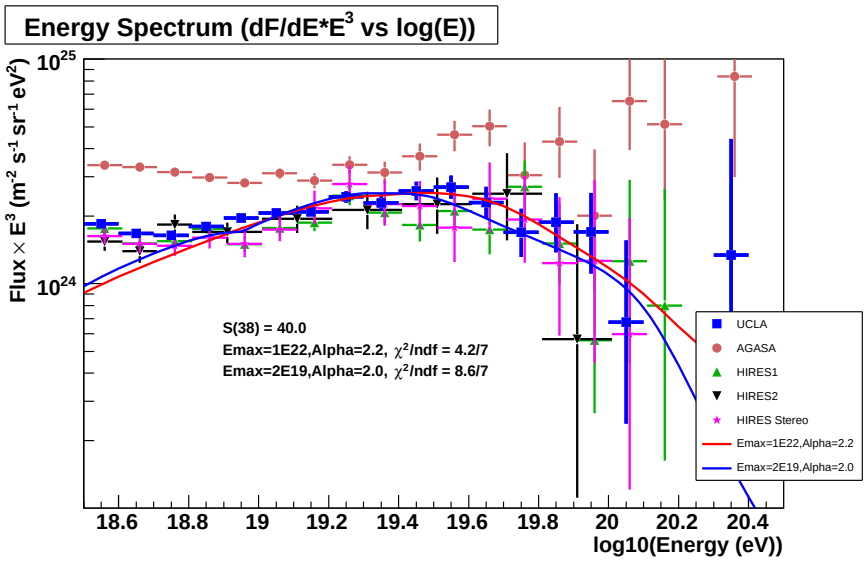


Figure 8.14: The two theoretical spectra of high significance. Blue line, $\alpha = 2.0$ for $E_{max} = 26 \cdot 2 \cdot 10^{19}$ eV. Red line, $\alpha = 2.2$ for $E_{max} = 26 \cdot 1 \cdot 10^{22}$ eV.

the proton flux from the photodisintegration is dominant. The latter is a heavy element dominant case where the proton flux is insignificant due to the low E_{max} resulting in an early termination of the proton flux (Remember, proton endpoint is 1/56th of the iron end point).

The two cases are actually shown in Fig. 8.14. Although the two cases have different final compositions, they are nearly identical in shape. The high E_{max} case has α of 2.2, whereas the low E_{max} case has α of 2.0. This situation can be understood as follows. The low E_{max} case has a deficit of events at the high energies. Thus α must be smaller to make the spectral shape flatter to make up for the deficit at the high energies — Hence the α of 2.2 for the high E_{max} and 2.0 for the low E_{max} cases. Interestingly enough, Fermi acceleration favors an α of 2 which is in godd agreement with that of iron whereas the best fit for proton happens for α of 2.6.

The low E_{max} case with its heavy element dominance presents an intriguing possibility. The cosmic ray spectrum at the sources is most likely to be a mixed composition, rather than all irons. The low E_{max} case naturally lends to a scenario where, for a mixed composition injection, each nucleus has an end point of the spectrum at Z·10^{19} eV, where Z is charge of the primary. In that case, high energy events above $5 \cdot 10^{19}$ eV, would be dominated by heavy elements. This picture is in agreement with the results obtained in the composition studies using rise time and X_{max}, where the composition changes from proton-like at 10^{19} eV and lower to Iron-like at high energies [84]. The high E_{max} case for the mixed composition has already been studied and reported in [85, 86]. While this may be more realistic than the pure iron injection case, the high E_{max} mixed composition injection yield a spectrum on earth with a predominance of protons especially at high energies which is not in line with [84]. The low E_{max} mixed composition injection case has not been studied. The low E_{max} pure iron injection suggests that a mixed composition injection with low E_{max} may yield a more realistic spectrum that is also in line with the preliminary observed compositional trend.

As we have done before for the pure proton injection, we check the effect of the systematic uncertainties in energy on the level of agreement between the theoretical spectrum for iron injection and the observed spectrum. We scan in S38 from 30 to 50 VEM to allow for 25% systematic uncertainties in energy and in alpha for the low and high E_{max} cases (see Figs. 8.15 and 8.16). The observed spectrum agrees with both low and high E_{max} cases with broad ranges of α's (non-blue regions) to 2σ level, at worst, regardless of the S38 value.

The observed spectrum, thus, agrees with both the proton and iron assumptions to better than 1–2σ level. These highly significant agreements are insensitive to the systematic uncertainties in energy. The insensitivity to the systematic

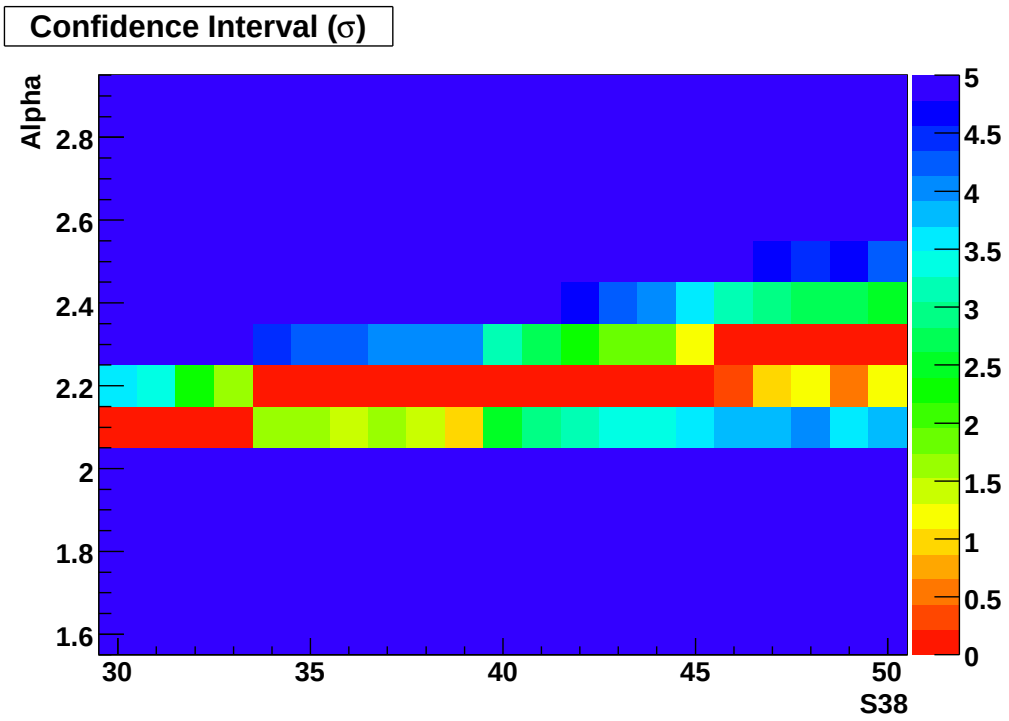


Figure 8.15: Significance plot for the $E_{max} = 26 \cdot 1 \cdot 10^{22}$ eV case allowing for 25% systematic uncertainties in energy as was also done for the pure proton injection case earlier. The theoretical predictions and the observed spectrum agree to no worse than 2σ level in the non-blue region.

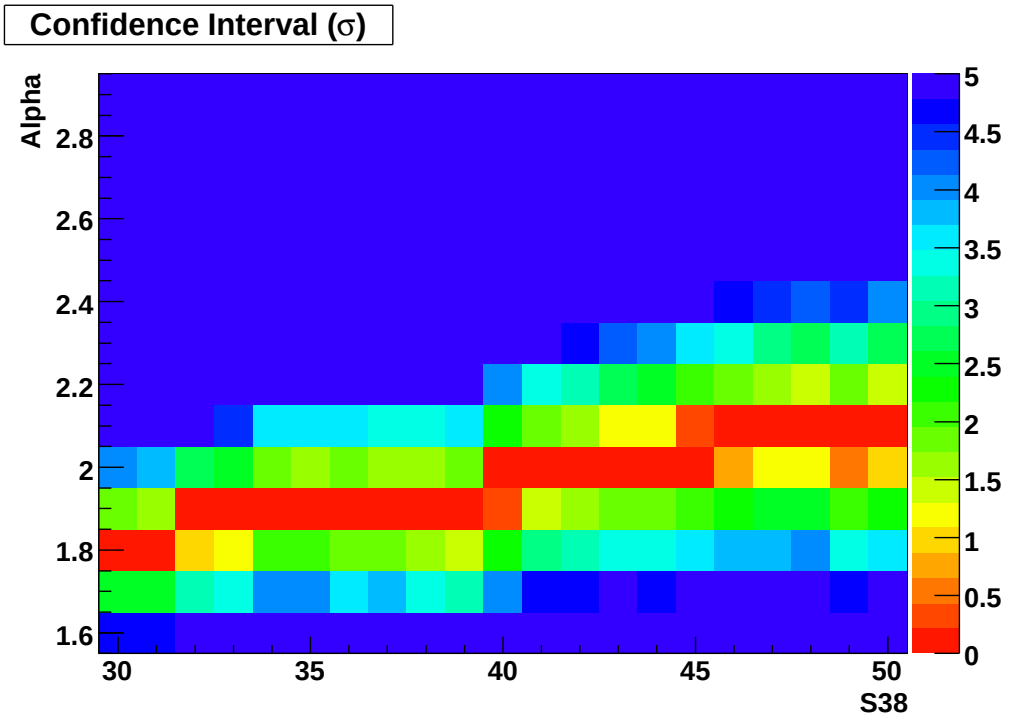


Figure 8.16: Significance plot for the $E_{max} = 26 \cdot 2 \cdot 10^{19}$ eV case. The theoretical predictions and the observed spectrum agree to no worse than 2σ level. α's that give the best agreement with the observed spectrum are smaller than the $E_{max} = 26 \cdot 1 \cdot 10^{22}$ eV case.

uncertainties also suggest that a reduction in the systematic uncertainties will not help us distinguish between iron and proton primary assumptions, let alone different α's. In the future, the spectral shape in the observed spectrum will become more well defined with more statistics and thus enable us to distinguish subtle differences in the spectral shapes of the proton and iron injection spectra. At present, all we can conclude is that the observed spectrum agrees with both proton and iron injection spectra with the GZK-like feature. It is worth mentioning, however, that there is one significant difference between the iron and proton injection cases. From what we have seen earlier, iron has a significant deficit of events below 10^{19} eV whereas proton does not. Since we have a wiggle room at the low energies in cases where the observed spectrum has a deficit of events, we will not dwell on this point.

8.6 Simple Power Law

In this section, we check how well the simple power law assumption agrees with the observed spectrum.

8.6.1 Merits of Simple Power Law

The simple power law assumption used to fit the observed spectrum is just that — a simple spectrum with $E^{-\alpha}$ dependence with no physics processes considered, thus without any deviation from the $E^{-\alpha}$ dependence for all energies. That is in contrast with the aforementioned theoretical spectra that start with a $E^{-\alpha}$ dependence but deviates from a strict $E^{-\alpha}$ dependence by the time they arrive at the earth after the cosmic rays undergo physics processes during propagation to the earth. Questions have been raised as to whether it is necessary to con-

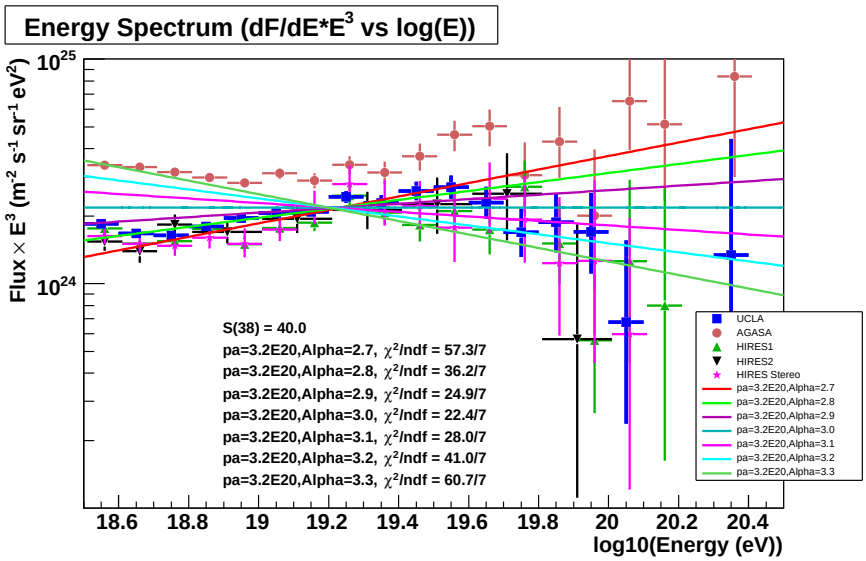


Figure 8.17: Simple Power Law Case.

sider the simple power law assumption since it is not well motivated by physics with its lack of any built-in physics processes during propagation. But since all theoretical spectra start with a $E^{-\alpha}$ dependence at the source, and there have been previous attempts to compare the simple power law assumption with the cosmic ray spectrum (see, for example, [87]), we will also compare the observed spectrum against the simple power law assumption. Excluding the simple power law assumption will eliminate the most naive approach.

8.6.2 Results

Examples of the simple power law fits to the observed spectrum are shown in Fig. 8.17. The E_{max} has a much more pronounced effect for the simple power law case. Due to the lack of a GZK feature in the simple power law assumption, the flux at the high energy can be quite significant for a high E_{max} case. It is, thus, necessary to place an endpoint of the theoretical spectrum close to the endpoint

of the observed spectrum to obtain the best fit in our maximum likelihood fit. A spectrum without the GZK feature that ends abruptly like the ones seen in Fig. 8.17 may seem a bit ad hoc, but as there must be a ceiling on the energy attainable from a bottom-up acceleration, this may not be so unnatural after all.

Varying E_{max} and α, once again we obtain a contour plot as before, shown in Fig. 8.18. The observed spectrum has a best agreement with the simple power law assumption at $\alpha\sim3$ to a 3σ level. Otherwise, the simple power law assumption is rejected emphatically. For the E_{max} range considered, the E_{max} dependence is very weak besides the hatched region. The simple power law cannot be ruled out completely given the 3σ level agreement at $\alpha\sim3$. This is one case where an increase in the statistics will definitely help. $\alpha = 3.0$ fits $1\text{–}5 \cdot 10^{19}$ eV range well (see Fig. 8.17). But there is an obvious disagreement above the GZK cutoff energy. An increase in the statistics in that region in the future will mean smaller error bars and more significant disagreement.

Now, we check, again, the effect of the systematic uncertainties in energy on the level of agreement between the theoretical and the observed spectrum. Varying the S38 values from 30 to 50 VEM, we discover that the significance can slightly improve as the simple power law assumption agrees with the observed spectrum to almost 2σ level for S38 from 30 to 35 VEM and 45 to 50 VEM (see Fig. 8.19). This significance, while quite small, is not small enough to rule out the simple power law case. While it may very well disappear, this 2σ significance may also very well stay. In that case, it would be quite troublesome as if an increase in statistics does not make it go away, that means we may have to reduce the systematic uncertainties in energy so as to eliminate 30–35 VEM and 45–50 VEMfrom allowed values for S38. That is, if we are really serious about rejecting the simple power law case emphatically. Otherwise, we may have to live

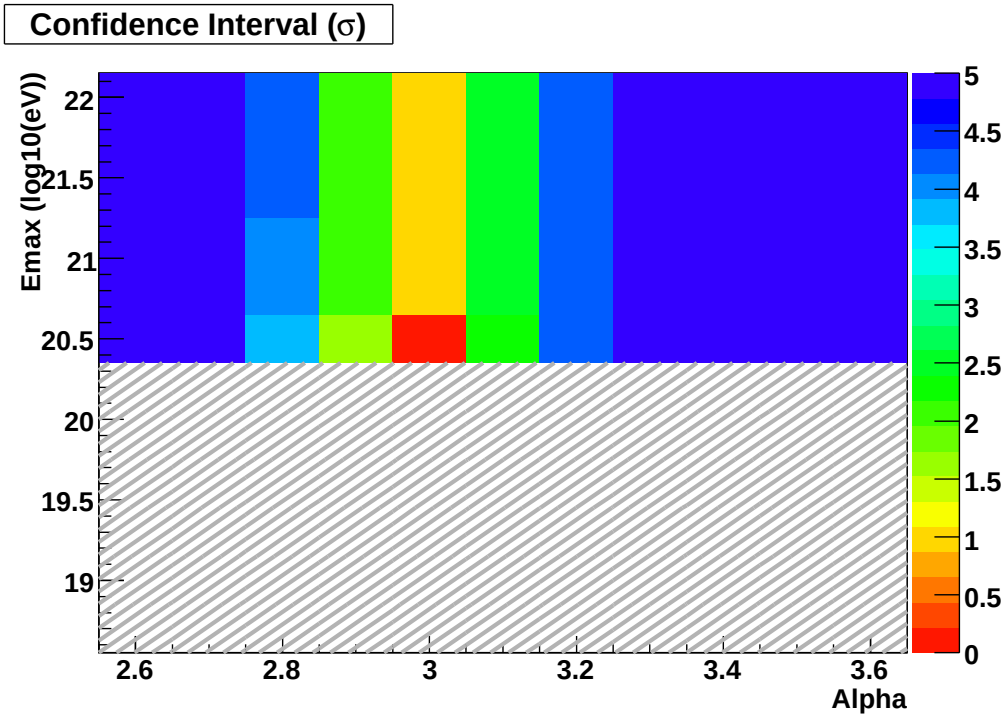


Figure 8.18: Significance plot for the simple power law assumption varying α and E_{max}. The agreement between the simple power law case and the observed spectrum is very weak. The hatched region is the forbidden region.

with the nagging possibility of the simple power law assumption being a viable candidate within the range allowed by the systematic uncertainties in energy no matter how unphysical we may think the simple power law assumption is.

8.7 Conclusions

We have presented our results of the comparisons between the observed spectrum and various theoretical predictions. The situation is summed up in Fig. 8.20, where we plot the maximum p-values for each S38 value for different theoretical

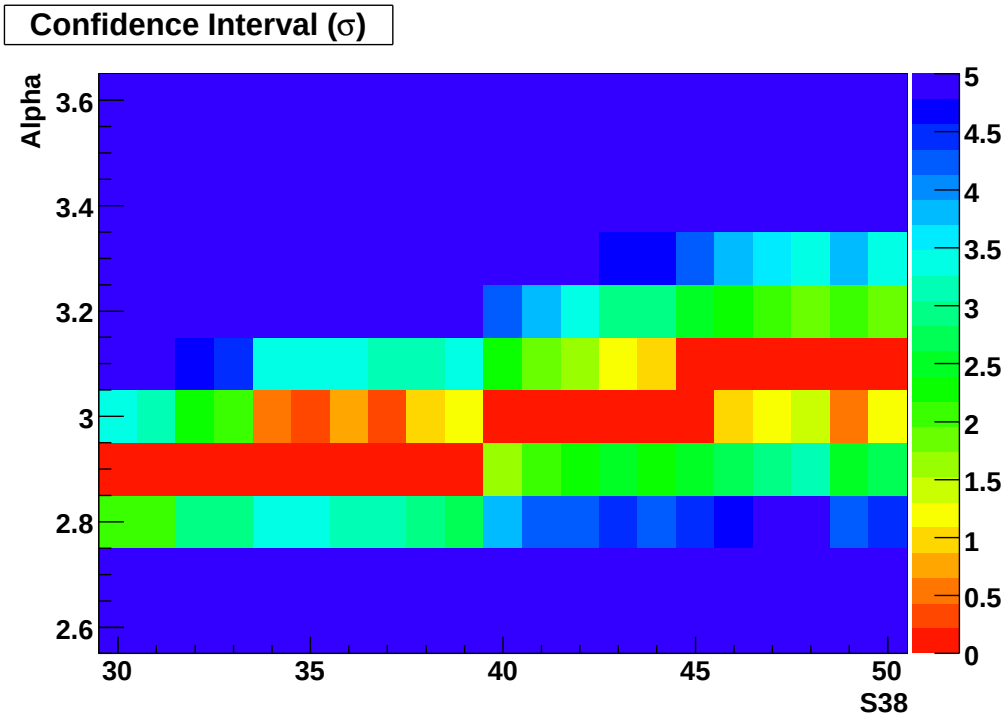


Figure 8.19: Significance plot allowing for the systematic uncertainties in energy, for $E_{max} = 3.2 \cdot 10^{20}$. The observed spectrum agrees with to $\sim 2\sigma$ level for S38 between 30 and 35 VEM and between 45 and 50 VEM.

cases. The observed spectrum agrees very well with both the proton, and the low and high E_{max} iron injection cases to better than 1σ level. Even when we scan S38 from 30 to 50 VEM to allow for the 25% systematic uncertainties in energy, the level of agreement is almost always better than 2σ. The good news is that this result is insensitive to energies. On the other hand, a reduction of the systematic uncertainties will not help us distinguish between the different cases whether it be different α's or different primaries. An increase in statistics will shrink the error bars in the observed spectrum, and the spectral shape will become more definite. Since there are subtle differences in the spectral shape of different theoretical assumptions, more statistics may be our only hope of sorting out different theoretical assumptions. The simple power law case can be ruled out to 2–3σ level. While this may not be significant, this significance will decrease with increasing statistics. The simple power law case is one case where the systematic uncertainties in energy matters. The 2σ level agreement between the observed spectrum and the simple power law at S38 of 30–35 VEM and 45–50 VEM (see Fig. 8.20) may persist even with increasing statistics in the future. Then, if we are serious about ruling out the simple power law case completely, we may have to reduce the systematic uncertainties to eliminate S38 of 30–35 VEM and 45–50 VEM as allowed values.

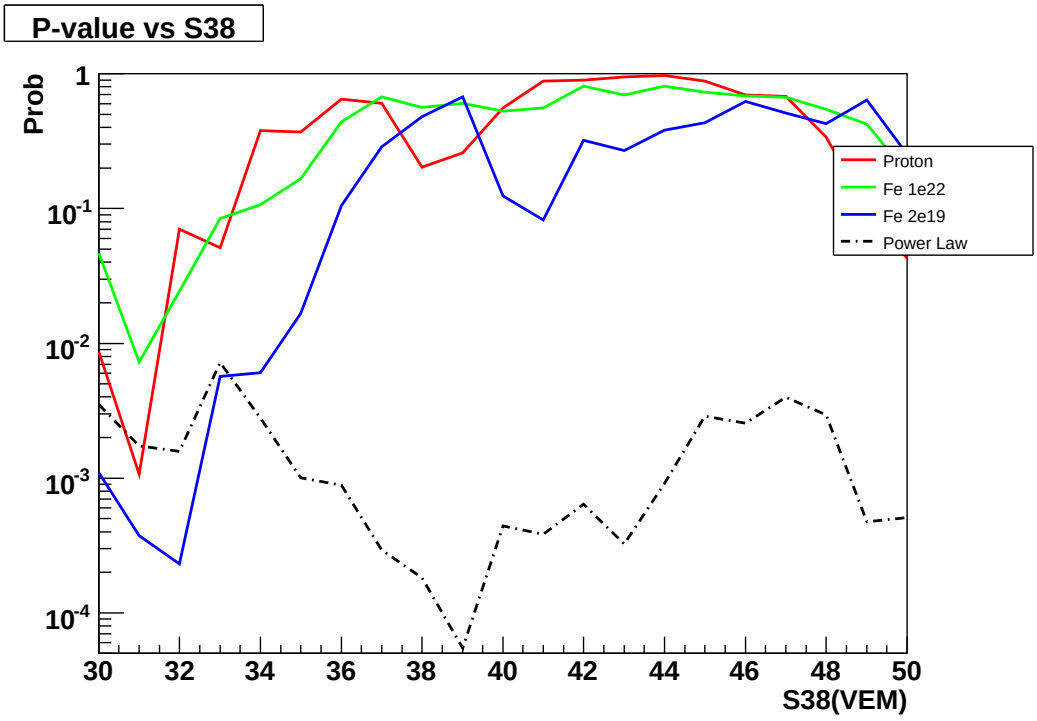


Figure 8.20: Summary plot of all cases. For each S38 value, maximum p-values are plotted for each case. (Dotted black line: simple power law. Red line: Proton. Green Line: Iron $26 \cdot 1 \cdot 10^{22}$ eV. Blue Line: Iron $26 \cdot 2 \cdot 10^{19}$ eV.)

CHAPTER 9

Conclusion

The Pierre Auger Observatory is the largest cosmic ray experiment ever and has already surpassed the two prominent forerunners, HiRes and AGASA, in exposure and will be the place to study cosmic rays in years to come. With the hybrid detector, we were able to cross calibrate the SD and the FD and compare with the Monte Carlo simulation based energy determination methods. By doing so, we were able to confirm that the difference in the HiRes and AGASA spectra were due to the two different energy determination techniques by reproducing the HiRes and AGASA spectra with the same set of events. Though this had been long suspected, this could not be proved as the two different energy determination methods could not be used on the same set of events as HiRes and AGASA were two independent experiments, and each had only one detection technique.

Even more important is the agreement of the observed spectrum with the GZK feature. Though we are not yet able to distinguish between proton and iron injection cases, what seems clear is that the conventional bottom up scenarios are favored as the agreement is better than 1σ. We cannot rule out the simple power law assumption yet, but the increase in statistics in the future should help to eliminate the simple power law case as well as distinguish between the proton and iron assumptions. The agreement with the GZK feature also means we should be able to identify the sources of these high energy events eventually as they must come from nearby unless cosmic rays are heavy elements or the extra-

and galactic magnetic fields are too strong. As the anistropy and the composition situations unfold with more statistics, we should be able to dig deeper into the question of the origin of the ultra high energy cosmic rays, when combined with the spectrum analysis.

Presently, the MC-based energy calibration method has ∼20% uncertainties (or more for those in the camp that believe that the nature may not be bound by Proton Sibyll and Iron QGSJet), and the hybrid-based energy calibration method has ∼30% uncertainties. While we are well aware of the shortcomings of the MC simulation, and that the FD is nearly calorimetric and is, in principle, better than the MC simulation, the large systematic uncertainties need to be reduced — the hybrid calibration of the SD means the systematic uncertainties of the FD gets carried over to the SD, so the SD is saddled with its own systematic uncertainties and those of the FD. The AirFly results with much smaller systematic uncertainties at 1–2% in the fluorescence yield will certainly go a long way as will an increase in hybrid events which will invariably bring high energy hybrid events which will help us cross calibrate at the high energies as opposed to having to just extrapolate as we do now. Better MC simulation, more hybrid data, and a reduction in the systematic uncertainties of the FD along with more statistics will allow us to determine the cosmic ray spectrum, the composition, and the anisotropies more accurately. As for the theory side, the low E_{max} mixed composition case mentioned in chapter 8 is an intriguing case and could take care of the cosmic ray spectrum and the composition in one shot and therefore should be investigated. So far, all the indications point to the conventional bottom up scenario with the GZK feature, and the PAO is in a good position to answer this question more accurately in the near future.

References

[1] V. F. Hess, Observations in Low Level Radiation During Seven Free Balloon Flights, Phys. Z. **13**, 1084 (1912).

[2] C. D. Anderson, Early Work on the Positron and Muon, Am. J. Phys. **29**, 825 (1961).

[3] P. Auger, Extensive Cosmic-Ray Showers, Reviews of Modern Physics **11**, 288 (1939).

[4] J. Cronin, T. Gaisser, and S. Swordy, Cosmic Rays at the Energy Frontier, Sci. Amer. **276**, 44 (1997).

[5] J. Linsley, Evidence for a Primary Cosmic-Ray Particle with Energy 10^{20} eV, Phys. Rev. Lett. **10**, 146 (1963)

[6] P. Sokolsky, *Introduction to Ultrahigh Energy Cosmic Ray Physics*, Addison-Wesley, 1984.

[7] T. Gaisser and T. Stanev, High-energy Cosmic Rays, arXiv:astro-ph/0510321.

[8] M. S. Longair, *High Energy Astrophysics*, Cambridge Univ. Press, 1981.

[9] D. Heck, M. Risse, and J. Knapp, Comparisons of Hadronic Interaction Models at Auger Energies, Nucl. Phys. B (Proc. Suppl.) **122**, 364 (2003).

[10] G. V. Kulikov and Yu. A. Fomin, *Ultrahigh Energy Cosmic Rays* (in Russian), Moscow, Atomizdat, 1975.

[11] E. Fermi, On the Origin of the Cosmic Radiation, Physical Review **75**, 1169 (1949).

[12] A. R. Bell, The Acceleration of Cosmic Rays in Shock Fronts, Mon. Not. Roy. Astron. Soc. **182**, 147 (1978).

[13] L. D. Landau and E. M. Lifshitz *Fluid Mechanics*, Oxford: Pergamon Press, 1978.

[14] M. Nagano and A. A. Watson, Observations and Implications of the Ultrahigh-Energy Cosmic Rays, Reviews of Modern Physics **72**, 689 (2000).

[15] W. I. Axford, The Origins of High-Energy Cosmic Rays, Astrophys. J., Suppl. Ser. **90**, 937 (1994).

[16] L. O. Drury, Acceleration of Cosmic-Rays, Contemp. Phys. **35**, 232 (1994).

[17] A. M. Hillas, The Origin of Ultra-High-Energy Cosmic Rays, Annu. Rev. Astron. Astrophys. **22**, 425 (1984).

[18] A. A. Penzias and R. W. Wilson, A Measurement of Excess Antenna Temperature at 4080 Mc/s, Astrophys. J. **142**, 419 (1965).

[19] The Auger Collaboration, *The Pierre Auger Observatory Design Report*, second edition, November 1996, Revised March 14, 1997.

[20] T. K. Gaisser, *Cosmic Rays and Particle Physics*, Cambridge Univ. Press, 1990.

[21] S. Yoshida, in *Proceedings of the 29th ICRC*, volume 10, page 297, 2005.

[22] N. Hayashida, Updated AGASA Event List above 4×10^{19} eV, arXiv:astro-ph/0008102.

[23] C. Finley and S. Westerhoff, On the Evidence for Clustering in the Arrival Directions of AGASA's Ultrahigh Energy Cosmic Rays, Astropart. Phys., **21**, 359-367 (2004).

[24] D. S. Gorbunov *et al.*, Testing the Correlations between Ultra-High-Energy Cosmic Rays and BL Lac Type Objects with HiRes Stereoscopic Data, JETP Lett., **80**, 145-148 (2004).

[25] The Auger Collaboration, *The Pierre Auger Project Technical Design Report*, March 2004.

[26] D. Allard *et al.* for the Auger Collaboration, The Trigger System of the PAO Surface Detector: Operation, Efficiency and Stability, in *Proceedings of the 29th ICRC*, 2005.

[27] A. Cordero *et al.*, Proposal for the Optical System of The Fluorescence Detectors of the Auger Project, Internal GAP Note 1996-039, 1996.

[28] P. Bauleo *et al.* for the Auger Collaboration, Absolute Calibration of the Auger Fluorescence Detectors, in *Proceedings of the 29th ICRC*, 2005.

[29] C. Aramo *et al.* for the Auger Collaboration, Optical Relative Calibration and Stability Monitoring for the Auger Fluorescence Detector, in *Proceedings of the 29th ICRC*, 2005.

[30] J. Bluemer *et al.* for the Auger Collaboration, Atmospheric Profiles at the Southern Pierre Auger Observatory and Their Relevance for the Air Shower Measurement, in *Proceedings of the 29th ICRC*, 2005.

[31] V. Rizi, M. Iarlori, S. Petrera, R. Caruso, and A. Grillo, The Use of Raman Lidar for the Atmospheric Monitoring at Pierre Auger Observatory, Internal GAP Note 2002-004, 2002.

[32] R. Cester *et al.* for the Auger Collaboration, Atmospheric Aerosol Monitoring at the Pierre Auger Observatory, in *Proceedings of the 29th ICRC*, 2005.

[33] M. Aglietta *et al.* for the Auger Collaboration, Response of the Pierre Auger Observatory Water Cherenkov Tanks to Shower Particles, in *Proceedings of the 29th ICRC*, 2005.

[34] P.S. Allison, P. Bauleo, X. Bertou, and C.B. Bonifazi, Surface Detector Calibration in the Engineering Array, Internal GAP Note 2002-004, 2002.

[35] M. Aglietta *et al.* for the Auger Collaboration, Calibration of the Surface Array of the Pierre Auger Observatory, in *Proceedings of the 29th ICRC*, 2005.

[36] The Auger Collaboration, Performance of the Pierre Auger Observatory Surface Array, in *Proceedings of the 29th ICRC*, 2005.

[37] P.S. Allison, X. Bertou, and C. Grunfeld, Pulse Shape Agnostic Methods of Measuring the Dynode-anode Ratio, Internal GAP Note 2004-033, 2004.

[38] I. Allekotte *et al.* for the Auger Collaboration, Observation of Long Term Stability of Water Tanks in the Pierre Auger Surface Detector, in *Proceedings of the 29th ICRC*, 2005.

[39] M. Healy *et al.*, Observation of the Long Term Stability of Water Tanks in the Pierre Auger Surface Detector, Internal GAP Note 2005-028, 2005.

[40] M. Unger, Shower Profile Reconstruction from Fluorescence and Cherenkov Light, Internal GAP Note 2006-010, 2006.

[41] P. Sommers and B. Fick, A Numerical Recipe for Fluorescence Detector Geometric Reconstruction, Internal GAP Note 2003-045, 2003.

[42] M. Nagano, K. Kobayakawa, N. Sakaki, and K. Ando, New Measurement on Photon Yields from Air and The Application to the Energy Estimation of Primary Cosmic Rays, arXiv:astro-ph/0510321.

[43] F. Kakimoto, E. C. Loh, M. Nagano H. Okuno, M. Teshima, and S. Ueno, A Measurement of the Air Fluorescence Yield, Nucl. Instrum. Methods, **A372** (1996).

[44] F. Nerling J. Blümer, b, R. Engel and M. Risse, Universality of Electron Distributions in High-energy Air Showers–Description of Cherenkov Light Production, 24:6 421 (2006).

[45] M. Giller *et al.*, Energy Spectra of Electrons in the Extensive Air Showers of Ultra-high Energy, J. Phys. G **30**, 97 (2004).

[46] T. K. Gaisser and A. M. Hillas, in *Proceedings of 15th ICRC*, volume 8, page 353, 1977.

[47] H. M. J. Barbosa *et al.*, Determination of the Calorimetric Energy in Extensive Air Showers, Astropart. Phys. **22**, 159 (2004).

[48] B. Dawson, M. Debes, and P. Sommers, Shower Profile Reconstruction with Engineering Array FD Data, Internal GAP Note 2001-016, 2001.

[49] D.Allard *et al.*, A New Method for the Longitudinal Profile Reconstruction of the AUGER Fluorescence Detector Events, Internal GAP Note 2006-026, 2006.

[50] P. Billoir, Reconstruction of First Year EA events from the Surface Detector, Internal GAP Note 2002-044, 2002.

[51] D. Veberic and M. Roth, SD Reconstruction; Offline Reference Manual, Internal GAP Note 2005-035, 2005.

[52] K. Kamata and J. Nishimura, The Lateral and Angular Structure Functions of Electron Showers, Progr. Theor. Phys. Suppl. **6**, 93 (1958).

[53] T. Ohnuki, Small Scale Angular Clustering of the Highest Energy Cosmic Rays in the Surface Detector Data of the Pierre Auger Cosmic Ray Observatory, PhD thesis, UCLA, 2005.

[54] S. J. Sciutto, *AIRES: A System for Air Shower Simulations*,, Departmento de Fisica Universidad Nacional de la Plata, Available from http://www.fisica.unlp.edu.ar/auger/aires.

[55] D. Heck, J. Knapp, J. N. Capdeville, G. Schatz, and T. Thouw, 1998, Forschungszentrum Karlsruhe, available from http://www-ik.fzk.de/ heck/corsika/physics_description/corsika_phys.html.

[56] N. N. Kalmykov, S. S. Ostapchenko, and A. I. Pavlov, Quark-gluon-string model and EAS simulation problems at ultra-high energies, Nucl. Phys. B (Proc. Suppl.) **52B**, 17 (1997).

[57] S. Ostapchenko, QGSJET-II: towards Reliable Description of very High Energy Hadronic Interactions, arXiv:hep-ph/0412332

[58] R. Engel, T. K. Gaisser, P. Lipari, and T. Stanev, in *Proceedings of the 26th ICRC,* volume 1, page 415, 1999.

[59] A. Fassò, A. Ferrari, J. Ranft, and P. R. Sala, FLUKA: Status and Prospective of Hadronic Applications, in *Proc. Monte Carlo 2000 Conf.*, edited by A. Kling, F. Barao, M. Nakagawa, and P. Vaz, page 955, Spring (Berlin), 2001, http://www.fluka.org/heart/rh.html.

[60] H. Fesefeldt, GHEISHA–the simulation of hadronic shower–physics and applications *Report* PITHA 85/02 RWTH Aachen, Germany, 1985.

[61] S. Argiro *et al.* for the Auger Collaboration, The Offline Software Framework of the Pierre Auger Observatory, in *Proceedings of the 29th ICRC*, 2005.

[62] D. Newton, J. Knapp and A.A. Watson, The Optimum Distance at which to Determine the Size of a Giant Air Shower, Astropart. Phys. **26**, 414 (2007).

[63] D. Allard *et al.*, A Guide-line to the Auger-Surface-Detector Analysis, Internal GAP Note 2006-024, 2006.

[64] K. Arisaka, Presentation at Auger Analysis Meeting in Chicago, September 2006.

[65] The Auger Collaboration, Performance of the Fluorescence Detectors of the Pierre Auger Observatory, in *Proceedings of the 29th ICRC*, 2005.

[66] K. Greisen, End to the Cosmic-Ray Spectrum?, Phys. Rev. Lett., **16**, 748 (1966).

[67] G. Zatsepin and V. Kuzmin, Upper Limit of Spectrum of Cosmic Rays, JETP Lett., **4**, 78 (1966).

[68] M. Takeda *et al.*, Extension of the Cosmic-Ray Energy Spectrum Beyond the Predicted Greisen-Zatsepin-Kuzmin Cutoff, Phys. Rev. Lett., **81**, 1163 (1998).

[69] R.U. Abbasi *et al.* for the High Resolution Fly's Eye Collaboration, Observation of the Ankle and Evidence for a High-Energy Break in the Cosmic Ray Spectrum, Phys. Lett. B, **619**, 271 (2005)

[70] T. J. Weiler, Cosmic-ray Neutrino Annihilation on Relic Neutrinos Revisited: a Mechanism for Generating Air Showers above the Greisen-Zatsepin-Kuzmin Cutoff. Astropart. Phys., 11:303, 1999.

[71] P. Bhattacharjee and G. Sigl, Origin and Propagation of Extremely High-energy Cosmic Rays. Physics Reports, 327:109, 2000.

[72] V. Berezinsky, M. Kachelrieß, and A. Vilenkin, Ultra-high Energy Cosmic Rays without GZK Cutoff. Phys. Rev. Lett., **79**, 4302 (1997).

[73] P. Blasi, R. Dick, and E. W. Kolb, Ultra-high Energy Cosmic Rays from Annihilation of Superheavy Dark Matter. Astropart. Phys., 18:57, 2002.

[74] M. Risse, Upper Limit on the Primary Photon Fraction from the Pierre Auger Observatory, in *Proceedings of the 29th ICRC*, 2005.

[75] D. Barnhill, Composition Analysis of Ultrahigh Energy Cosmic Rays Using the Pierre Auger Observatory Surface Detector, PhD thesis, UCLA, 2005.

[76] D. Allard *et al.* for the Auger Collaboration, Determination of the Aperture of the PAO Surface Detector, in *Proceedings of the 29th ICRC*, 2005.

[77] I. Lhenry, E. Parizot, D. Allard, P. Ghia, and J. Aublin, The Lateral Trigger Probability Function (LTP) for UHE Showers Detected by the PAO Surface Detector, Internal GAP Note 2003-088, 2003.

[78] C. Bonifazi and A. Letessier-Selvon, Event Selection Using the T5 Time Distribution, Internal GAP Note 2006-042, 2006.

[79] J. W. Cronin, A Simple Determination of the Cosmic Ray Energy Spectrum at the Auger Observatory, Internal Note.

[80] O. E. Kalashev, V. A. Kuzmin, D. V. Semikoz, G. Sigl, Ultrahigh Energy Cosmic Rays from Neutrino Emitting Acceleration Sources?, Phys. Rev. D **65**, 103003 (2002).

[81] O. Kalashev *et al.*, In preparation.

[82] G. Cowan, *Statistical Data Analysis*, Oxford Science Publications, 1998.

[83] J. Heinrich, The Log Likelihood Ratio of the Poisson Distribution for Small μ, CDF Internal Note 5718, 2001.

[84] M. Healy, D. Barnhill, K. Arisaka, J. Lee, P. Boghrat, A Study of Composition Trends Using Rise Time and Curvature Data, Internal GAP Note 2006-092, 2006.

[85] D. Allard, E. Parizot, A. V. Olinto, On the Transition from Galactic to Extragalactic Cosmic-rays: Spectral and Composition Features from Two Opposite Scenarios, Astropart. Phys. **27** 61 (2007), also arXiv:astro-ph/0512345.

[86] D. Hooper, S. Sarkar, and A. M. Taylor, The Intergalactic Propagation of Ultra-high Energy Cosmic Ray Nuclei, Astropart. Phys. **27** 199-212 (2007), also arXiv:astro-ph/0608085.

[87] D. R. Bergman for the HiRes Collaboration, Fitting the HiRes Spectra, arXiv:astro-ph/0507484.