CHEMICAL COMPOSITION OF PRIMARY COSMIC RAYS WITH ICECUBE

by

Chen Xu

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Physics

Winter 2012

CHEMICAL COMPOSITION OF PRIMARY COSMIC RAYS WITH ICECUBE

by

Chen Xu

Approved: ____
George Hadipanayis, Ph.D.
Chair of the Department of Physics and Astronomy

Approved: ____
George Watson, Ph.D.
Dean of the College of Arts and Sciences

Approved: ____
Charles G. Riordan, Ph.D.
Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Thomas Gaisser, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Todor Stanev, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

John Gizis, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

David Seckel, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Louis Rossi, Ph.D.
Member of dissertation committee

ACKNOWLEDGEMENTS

I wish to thank all the people from the IceCube collaboration who have designed and built the detector, developed the simulation and data processing softwares and keep everything running. Without them, this thesis would not exist.

Especially, I'd like to thank people from University of Delaware, Bartol Research Institute. They are Tom Gaisser, my advisor, who read my thesis carefully and gave me invaluable advice; Todor Stanev, together with Dr. Gaisser, taught me cosmic ray physics; David Seckel, who always had sharp questions; Serap Tilav, who monitored and checked everything; Daniel De Marco, who kept the Bartol computers running; Patrick Berghaus, Shahid Hussain, Takao Kuwabara and Paul Evenson, who gave suggestions from time to time.

I'd like to thank some people who used to work at Bartol. They are Peter Niessen, who taught me how to use the Bartol cluster system; Tilo Waldenmaier, who helped me solved the bug in the simulation before leaving IceCube; Xinhua Bai, my mentor at IceCube.

I'd also like to thank people from other institutions in cosmic ray group of IceCube. Jonathan Eisch (Madison), with whom I fought a lot, but who managed to provide the majority of simulation we needed; Karen Andeen (Madison) and Katherine Rawlins (Alaska), who worked on IC40 composition and gave me a lot of suggestions; Alex Olivas (Maryland) who helped me understand more in GCD and python programming; Hermann Kolanoski, who questioned me a lot to prepare me for the defense.

TABLE OF CONTENTS

Chapter

Appendix

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Ground detector arrays have been used to measure high energy cosmic rays for decades to overcome their very low rate. IceCube is a special case with its 3D deployment and unique location–the South Pole. Although all 86 strings and 81 stations of IceCube were completed in 2011, IceCube began to take data in 2006, after the completion of the first 9 strings. In this thesis, experimental data taken in 2009 with 59 strings are used for composition analysis albeit some techniques are illustrated with the 40-string data.

Simulation is essential in the composition work. Simulated data must be compared against the experimental data to find the right mix of cosmic ray components. However, because of limited computing resources and complexities of cosmic rays, the simulation in IceCube is well behind the experiment. The lower and upper bounds of primary energy in simulation for events that go through IceTop and the deep arrays of IceCube are $10^{14}eV$ and $10^{17}eV$. However, since IceCube has a threshold energy about several hundred TeV, and an upper limit of $10^{18}eV$, the full energy range cannot be explored in this thesis.

The approach taken to the composition problem in this thesis is a 2D Bayesian unfolding. It takes account of the measured IceTop and InIce energy spectrum and outputs the expected primary energy spectrum of different mass components. Studies of the uncertainties in the results are not complete because of limited simulation and understanding of the new detector and South Pole environment.

Chapter 1

OVERVIEW

The primary composition of cosmic rays at high energy has been pursued for many decades by using large scale ground based detector arrays. The general idea to solve the problem is to compare composition-sensitive parameters in real experimental data to that in simulation with a known composition.

Some experiments use muon multiplicity or depth of shower maximum as the composition-sensitive parameters. IceTop–the surface part of IceCube is mostly sensitive to the electro-magnetic component in air showers. The high energy muons at the shower core that have energies above 500 GeV can reach or penetrate the InIce part of IceCube (from 1450m to 2450m below the ice). Thus, IceTop-InIce coincident events in IceCube might tell us both the shower size and high energy muon bundle size. The ratio of two sizes is a unique composition indicator in IceCube and complementary to using the ratio of low energy muons at the surface to the electro-magnetic component in the shower front as is done in some other experiments.

The high energy muon bundles measured by IceCube can be reconstructed with several independent energy proxies. One such energy proxy called MuE is used to represent the total muon bundle energy at the center of gravity of InIce hits in this thesis. Since the problem we are facing is to solve the composition at certain primary energies, we need to get the distributions of MuE at precise primary energies. The mean error in the reconstructed energy in $\log_{10}$(E/GeV)

(1PeV $\leq \log_{10}$(E/GeV) $<$ 100PeV) is between 0.1 to 0.2, about the bin size chosen for energy. This is due to the physical fluctuations between different showers of the same primary energy. The inaccuracy in primary energy reconstruction indicates that any 1-dimension unfolding method is not adequate in solving the composition problem. The problem needs to be treated with at least two dimensions in unfolding.

Since the problem has two dimensions, one of which is mass, the other is primary cosmic-ray energy, we need at least two variables from IceCube measurement. We already have the composition sensitive parameter MuE, and we can simply choose reconstructed energy in IceTop as the other one. The problem then becomes how to transform the measured distribution in $\log_{10}(\mathrm{E_{reco}}/\mathrm{GeV})$ and $\log_{10}$ MuE to the physics in $\log_{10}(\mathrm{E_{prim}}/\mathrm{GeV})$ and average mass $\ln A$. A 2D Bayesian unfolding method is used to tackle the problem. However, due to the steep spectrum in energy and lack of statistics, the result is subject to a large error. A smoothing method is taken to reduce the statistical error. Systematic errors also contribute a lot to the uncertainties. The snow height, pressure/temperature, hadronic interaction and reconstructions effects are estimated.

Chapter 2

INTRODUCTION TO COSMIC RAYS

Cosmic rays are ionized nuclei that hit the earth from the outer space. About 89% of them are protons, nearly 9% are alpha particles, with the rest made up of other elements. The word "ray" here is a little misleading since cosmic rays are not real rays and they come to visit the earth individually. Cosmic rays can have energies as high as 10^{20}eV, much higher than the energy that the accelerators can produce (10^{12} to 10^{13}eV). The differential flux of cosmic rays can is well described by an inverse power law function

$$\frac{dN}{dE} \propto E^{-(\gamma+1)}. \tag{2.1}$$

The spectrum has a γ around 1.7 from above 1GeV to 10^6GeV and then steepens with a γ about 2 (see Fig. 2.1). For particles with energy below 100TeV, direct measurement can be conducted by using spectrometers or calorimeters with high resolution in both energy and mass. At higher energies, ground-based detectors covering large areas have to be used due to the low flux. This is an indirect measurement, which usually gives limited information about the size of air showers produced by cosmic rays in the atmosphere. Because we are interested in high energy cosmic rays above 1PeV, the effect of solar wind on low energy cosmic rays in heliosphere should not be a concern here. And this thesis will not address the effect of the earth's magnetic field.

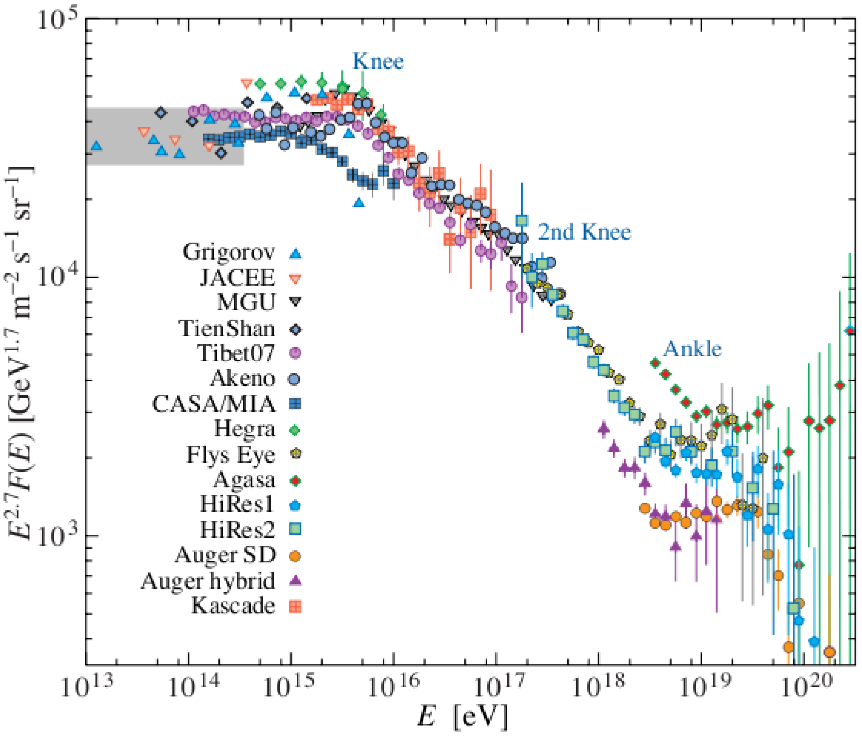


Figure 2.1: The all-particle spectrum from air shower measurements, graph is from [1]. The shaded area shows the range of the direct cosmic ray spectrum measurements.

2.1 Sources of Cosmic Rays

Since the flight paths of cosmic rays are deflected by the magnetic fields of the galaxy, the solar system and the earth, they are uniformly distributed at all directions in the sky and not pointing back to their origins. Assuming a typical cosmic ray density $\rho_E = 1\ eV/cm^3$, the power required to supply all the cosmic rays is [2]

$$L_{CR} = \frac{V_D \rho_E}{t_{GD}} \approx 5 \times 10^{40}\ erg/s, \tag{2.2}$$

where V_D is the volume of the galactic disk and t_{GD} is the residence time of cosmic rays in the volume. The suggested candidate of the source of cosmic rays is supernova remnants, which are large, long-lived and have higher magnetic fields than the average interstellar medium [4]. Supernova with remnants of 10 solar masses and velocity of 5×10^8 cm/s can produce 3×10^{51} $ergs$ every 30 year. Thus, with a efficiency of about a few percent, the supernovae can provide all the power that is needed by the cosmic rays.

The acceleration mechanism is believed to be stochastic acceleration at supernova blast shocks, The matter ejected by the supernova explosion has a much higher velocity than the sound velocity of the interstellar medium, which drives the shock that accelerates cosmic rays. The spectral index γ of the accelerated cosmic rays can be predicted by shock acceleration. For a large plane shock, the escape probability is the ratio of the rate of convection downstream away from the shock to the rate of shock encounters

$$P_{esc} = \frac{\rho_{CR} u_2}{c\rho_{CR}/4} = \frac{4u_2}{c}. \tag{2.3}$$

Note u_2 is the fleeing away velocity of downstream shocked matter, ρ_{CR} is the number density of particles undergoing acceleration. Thus, for acceleration at a strong shock,

$$\gamma = \frac{P_{esc}}{\xi} = \frac{4u_2}{c} \times \frac{3c}{4(u_1 - u_2)} = \frac{3}{u_1/u_2 - 1} \sim 1, \tag{2.4}$$

where u_1 is the velocity of shock front, ξ is the efficiency of energy gain of each particle per encounter. The energy that a particle can achieve in shock acceleration has a cap due to the finite lifetime of the supernova blast wave.

The supernova remnant shock model has several problems: 1) it only explains cosmic rays with energies lower than 100 TeV, though cosmic rays with energies as high as 10^{20} eV are observed; 2) the observed spectral index γ is about 1.7, but diffuse shock acceleration gives 1.1; 3) the expected TeV gamma rays from certain supernova remnants have not been confirmed.

The supernova blast wave mechanism can be modified to support the creation of cosmic rays with energies higher than 100 TeV. In the diffuse source method, the characteristics of the interstellar medium, such as the magnetic configuration, are changed to accommodate cosmic rays with one or two orders magnitude higher than 100 TeV; in the point source method, local compact objects with very high magnetic fields, such as neutron stars, are suggested to accelerate cosmic rays to higher energies. Other solutions include introducing the precursor in the upstream region, treating upstream scattering efficiency a time variable in later phases of supernova expansion, recognizing magneto-hydrodynamic turbulence generated by cosmic-rays upstream of the shock, and so on [5].

2.2 Propagation of Cosmic Rays

Almost all of the cosmic rays come from outside the solar system but within the galaxy. The galaxy has a radius of about 20 kpc. The thickness of the galaxy is about 8.5 kpc at the galactic center and $200 - 300$ pc at the radius of the earth. The main components of interstellar diffuse matter are atomic neutral hydrogen, molecular hydrogen and ionized hydrogen. Atomic hydrogen has an average density about 1 cm^{-3} in the galactic arms with a scale height of $100 - 150$ pc. Molecular hydrogen is concentrated in the solar circle with an average density about 1 cm^{-3}. The density of the giant molecular clouds can be thousands of times above the average. The ionized hydrogen contributes to a small fraction of the interstellar medium with a density of 0.03 cm^{-3} and a height of 700 pc [6]. The ionized gas and galactic magnetic field form a magnetohydrodynamic (MHD) fluid that supports waves that travel with the Alfven velocity $v_A = B/2\sqrt{\rho\pi}$, on which cosmic rays scatter in their propagation.

Let's consider some simplified propagation models of cosmic rays. In the leaky box model, the cosmic rays propagate freely in the galaxy that contains their sources with a certain escape probability. Thus, for the observed $E^{-(\alpha)}$, the source must have a harder spectrum of $E^{-\alpha+\delta}$, where $\delta = 0.6$ denotes the mean amount of matter traversed by the particle. For iron, it predicts a flatter spectrum from 18 GeV/nucleon till about 1 TeV/nucleon. The second model is called nested leaky box model. It assumes the regions close to the sources have high density that make particles diffuse for a short but energy dependent time. In this model, an observer inside a source region would measure a differential spectrum proportionate to $E^{-(\alpha+\delta)}$ due to energy dependent leakage out of the source; an observer outside the source region (the Earth) would see spectrum of the source of $E^{-\alpha}$. A variation of leaky box model is the closed galaxy model, which treats the Earth as inside the source. This model gives the observed decline in the ratio of secondary to primary nuclei with energy. Diffusion models are more realistic models since they include the full transport equation. Compared to various leaky box models, the cosmic rays inside the containment volume is not uniformly distributed, even in steady state.

2.3 Chemical Composition of Cosmic Rays

The aforementioned propagation models interpret the composition and spectra of cosmic nuclei. About 79% of the primary nucleons are free protons and about 70% of the rest are bounded in helium nuclei. The major components of the primary nuclei are shown in Fig. 2.2 with energies from above 2 GeV/nucleus to less than 1 PeV/nucleus. The fractions of primary nuclei are nearly constant over this energy range. The intensity of the primary cosmic nucleons as a power law function of the energy-per-nucleon could be expressed as follows from several GeV to about 100 TeV [1]

$$I_N(E) \approx 1.8 \times 10^4 (E/GeV)^{-2.7} \frac{\text{nucleons}}{m^2\ s\ sr\ GeV}. \tag{2.5}$$

Comparing the relative abundance of elements in cosmic rays with that in solar material is an interesting way to infer the origin and propagation of cosmic rays. Fig. 2.3 shows the difference is two fold:

- Heavy nuclei ($Z > 1$) take a larger portion in cosmic rays than in solar materials. This indicates that there fewer protons go into the acceleration process

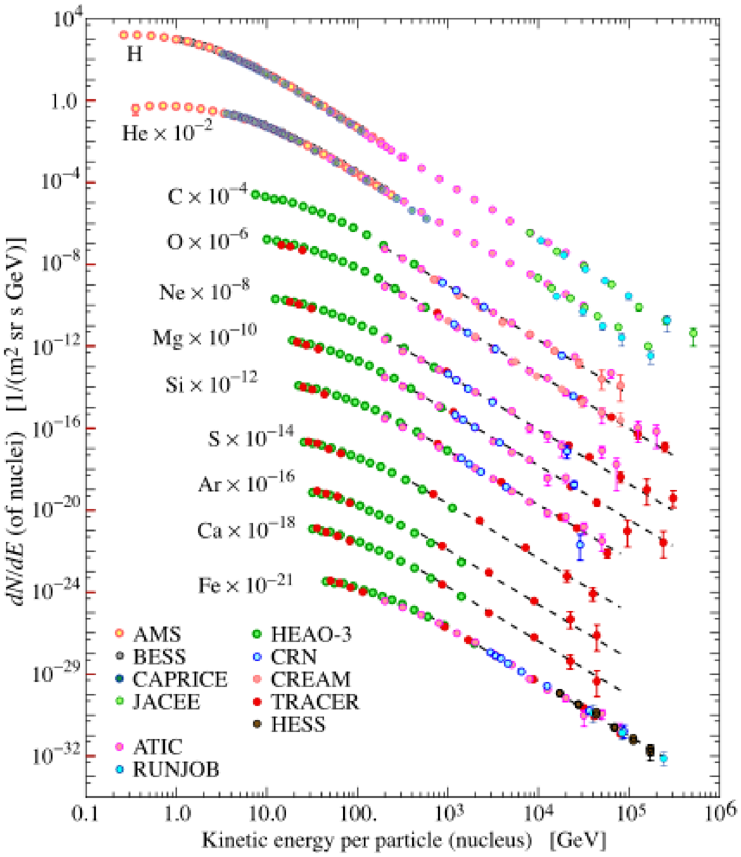


Figure 2.2: Major components of the primary cosmic radiation [1]. The figure was created by P. Boyle and D. Muller.

either because it's relatively harder to ionize hydrogen or the composition at cosmic ray source is really different from the solar system.

- Two groups of elements, Li, Be, B and Sc, Ti, V, Cr, Mn are more abundant in cosmic rays by many orders of magnitude. These elements are the spallation products created by the collisions of the abundant carbon and oxygen with the interstellar medium [7]. With the known spallation cross section, the mean amount of matter that the bulk of cosmic rays traversed is about $5-10\ g/cm^2$. By assuming a $1\ cm^{-3}$ density of the galaxy, the lower limit of the length traveled by cosmic rays of about 1000 kpc is obtained, which is far longer than the thickness of the galactic disk. This supports the idea that the propagation of cosmic rays is a diffuse process.

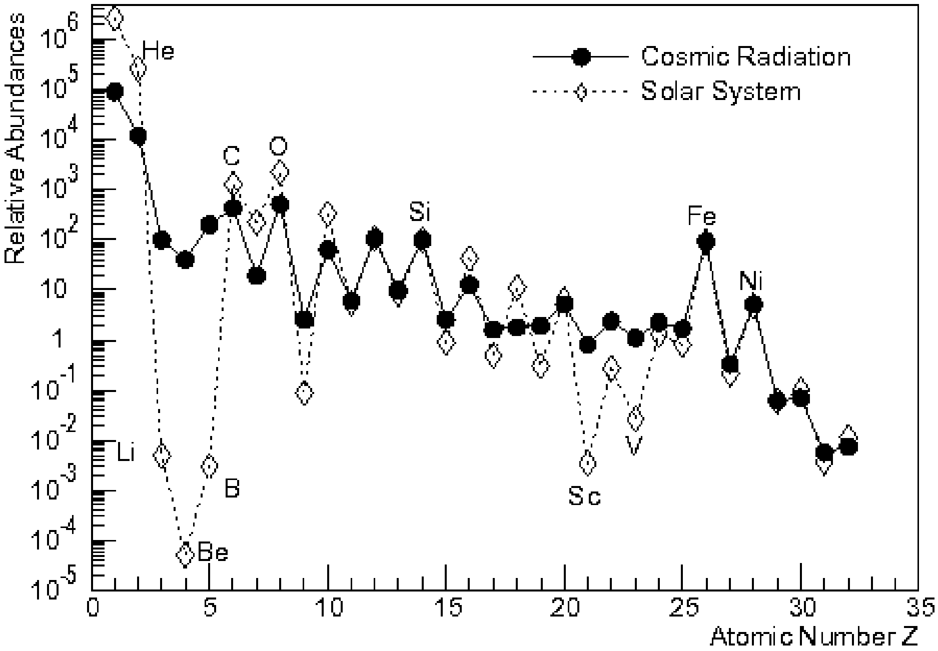


Figure 2.3: The cosmic ray elemental abundances (He-Ni) measured at the earth compared to the solar system abundances, all relative to silicon [8], reproduced by Tilo Waldenmaier.

2.4 The Origin of Ultra-High-Energy Cosmic Rays

The observed $\sim 10^{20}$ eV cosmic rays pose a question of their sources in the leaky box model. In which those particles can not be retained in the disk of our galaxy by its magnetic fields. The gyroradius r_g of a particle with Z charges in a magnetic field B is

$$r_g/m = 3.3 \times \frac{p_\perp/(GeV/c)}{|Z|(B/T)}, \tag{2.6}$$

where $p_\perp$ is the relativistic momentum perpendicular to the direction of the magnetic field of the charged particle. For a 10^{20} eV proton traveling in a 3 μG magnetic field, the gyroradius is about 1.1×10^{21} m, equivalent to 35.6 kpc. This value becomes

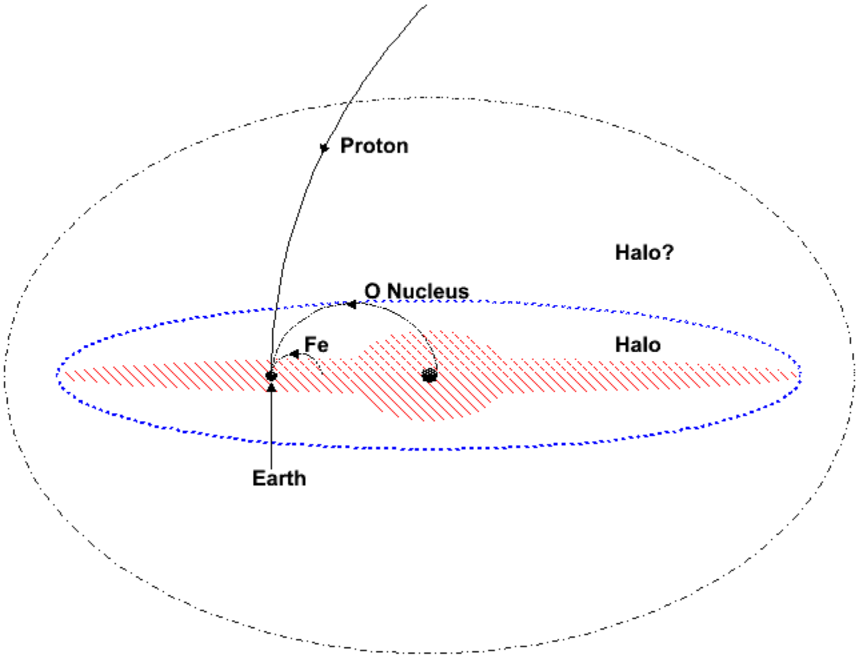


Figure 2.4: Size of the trajectories of 10^{20} eV cosmic-ray nuclei in relation to the size of the galaxy in a uniform 3 μG magnetic field. Reproduced from [9]

4.45 kpc for oxygen and 1.37 kpc for iron, as is shown in Fig. 2.4. Since the essential part of the accelerating region containing the magnetic field must be much larger than twice of the gyroradius [9], protons at 10^{20} eV level would originate outside our galaxy. If the galactic halo extends to several kpc to support the necessary magnetic field, some heavier nuclei of 10^{20} eV level could have sources within our galaxy. The flattened tail in Fig. 2.1 suggests cosmic rays with energies beyond 10^{19} eV falls less rapidly. Hypothesis is that cosmic rays originating from our own galaxy ends near 10^{19} eV, beyond which a different source is active. However, according to the work of Allard and et al [10] the transition from extra-galactic to galactic nuclei could occur from $10^{17.5}$ to 10^{19} eV. IceCube has a capacity to detect cosmic rays as high as 10^{18} eV, which stands a chance to see this transition.

2.5 Air Showers Development and Detection

When a cosmic ray particle, such as a proton, strikes the atmosphere of the earth, the collisions between this particle and the nuclei of the air produce many secondary particles, which hit the air further and produce more ionized particles and electromagnetic radiation. The burst of charged particles in this process is called an extensive air shower or cascade that usually extends over tens of kilometers in altitude, from the very top of the atmosphere to the ground of the earth.

The air shower starts with a single primary particle and then proliferates as it goes deeper and deeper in the atmosphere and the size of the shower keeps getting bigger before reaching a maximum. Here we define the depth of shower maximum as X_{max}, which is the depth in the atmosphere that the number of particles in the shower is greatest. Due to the low flux of showers initiated by high energies (greater than 100 TeV), large ground-based detector arrays have been built to measure the light emitted during the shower development or the electromagnetic and muonic components of shower when it arrives at the detector.

Cosmic ray induced air showers consist of electromagnetic, muonic and hadronic components. In a hadronic shower, a large fraction of the nucleon is transferred to the secondary mesons through charged and neutral channels in the first interaction. If there is any energy left in the nucleon, it would interact again in the same way after traversing about one more interaction length. The mesons interact themselves at about the same time and generate a second generation of hadrons. As is shown in Fig. 2.5, secondary neutral pions have short lives and decay into γ pairs quickly and start electromagnetic cascades. The charged pions can either interact or decay. High energy pions that have long decay length interact while low energy pions decay into muons and muon neutrinos. The competition between interaction and decay is a complicated process and depends on energy and depth in the atmosphere. The electromagnetic component of air showers can only be measured by a detector on

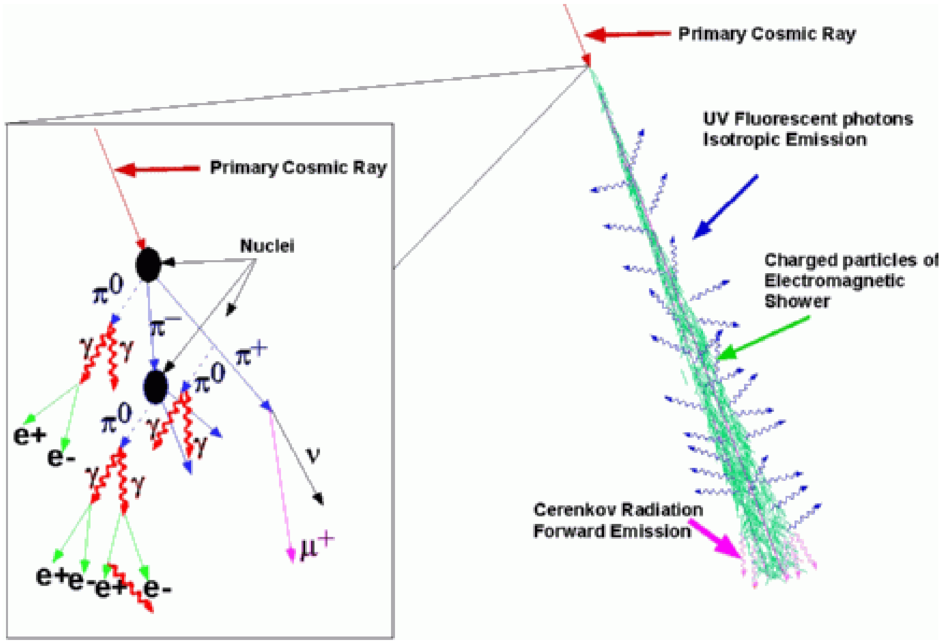


Figure 2.5: A schematic of the hadronic shower interaction from [11]. The right part shows the general profile of a air shower as it develops toward the ground; the left part shows the interaction channels at the injection point. The fluorescent photons should emit isotropically rather than backward as shown in the figure.

the surface while high energy muons might travel several kilometers underground.

Note that IceTop (the surface part of IceCube) can see neither the isotropic fluorescence light or the forward Cerenkov light shown in the right panel of Fig. 2.5. In an experiment that implements the air fluorescence technique such as Pierre Auger detector array, a grid of focusing mirrors with PMTs at the focal planes collect the emitted ultraviolet light in air showers that are up to 15 kilometers away [12]. The shower maximum can thus be measured by using this technique. An air Cerenkov detector like HEGRA uses mirrors to reflect Cerenkov light to the PMTs at the focal planes, which convert the optical signal into an electronic signal. This technique is typically used in gamma ray detectors [13].

Generally, heavy nuclei interact earlier (shorter first interaction length) and develop more rapidly than their lighter counterparts in the atmosphere. Measuring

X_{max} (measured in g/cm^2) of an air shower enables us to identify the primary particle type of this shower once the primary energy is known. Experiments that can determine the shower maximum measure the light emitted from air showers. They are usually telescopes located at high mountains because they only work on clear, dry and moonless nights. Those experiments detect either the Cerenkov light emitted by the shower electrons or fluorescence light excited by the ionized atmospheric nitrogen atoms. The air Cerenkov experiments are able to measure small showers with energies as low as 1 TeV because of the small refraction index of the atmosphere. The fluorescence detectors usually have a threshold of 10^{17} eV because the isotropically emitted fluorescence is dominated by the Cerenkov light if the trajectories are near the detectors. Representatives of these two detection techniques are VERITAS [1] and Fly's Eye experiments that are deployed in Argentina and Utah, USA respectively.

KASCADE [2], on the other hand, is a classic ground-based air shower array that determines air showers by fast timing the shower front. It is built about 100 meters above the sea level. The 252 scintillator counters of the array cover an area of $4 \times 10^4\ m^2$. The ability to distinguish the muonic component from the electronic one not only gives KASCADE a better measure of the primary energy, but also offers an alternative way to determine the chemical composition of the primary cosmic rays. Large air shower arrays are exceptionally useful to observe high energy events, since their broad area and ability to operate round-the-clock compensate for the low flux of ultra high energy cosmic rays.

KASCADE-Grande is a further extension of the previous project by reassembling 37 stations that are at a mutual distance of about 130 m covering an area of 0.5

---

[1] The Very Energetic Radiation Imaging Telescope Array System

[2] Karlsruhe Shower core and Array Detector Array

km^2 next to the KASCADE site. The primary energy range covered by KASCADE-Grande is from 10^{16} to 10^{18} eV [14].

The Auger Observatory is the largest observatory used to study cosmic rays with energies larger 10^{19} eV with unprecedented precision and statistics [15]. It employs two independent methods to detect high-energy cosmic rays. One technique detects Cerenkov light emitted by the high energy particles through their interaction with water placed in the 1600 surface tanks that cover an enormous section of the Pampa Amarilla in Argentina. The fluorescence detector tracks the development of air showers by observing ultraviolet light emitted in the interaction between charged particles and atmospheric nitrogen [16].

IceCube is another unique detector arrays located at the South Pole. Unlike the conventional arrays that only sample showers at one depth, the pure South Pole ice separates the muons from the bulk shower particles naturally from the surface level. By measuring the electrons and muons simultaneously with its thousands of optical modules, IceCube has a potential to measure cosmic rays with energy up to 10^{18} eV.

Chapter 3

ICECUBE DETECTOR ARRAY

IceCube Detector Array is a multi-functional detector [1] built at South Pole, spanning one kilometer in length, width and 2.5 kilometers in height as shown in 3.1. The two components of IceCube are IceTop and In-Ice detector arrays. The IceTop detector array is on the surface at South Pole while the In-Ice array is buried between 1450m and 2450m beneath the surface. IceTop and In-Ice array share the same hexagonal cross section and use the same basic optical units – digital optical modules (DOMs). The DOM is a photo-multiplier tube (PMT) combined with other electronic devices such as a mainboard, enclosed by transparent spherical glass. It is the individual DOM that records the photons generated by all kinds of events that happen in the deep ice and then send waveform signals to the data acquisition system (DAQ). The DAQ collects and filters the information for different types of events. The simulation in IceCube starts with the simulation of source particles, with tools such as CORSIKA or neutrino generators. After propagating the muons by Muon Propagation Code (MMC) [17], the detector simulation comes into play. PMT, DOM and the triggering system completes the rest of the simulation chain. The raw data, either simulated or observed, will be calibrated to feed the waveform processor and feature extractor, which summarize the time and charge information for reconstruction use.

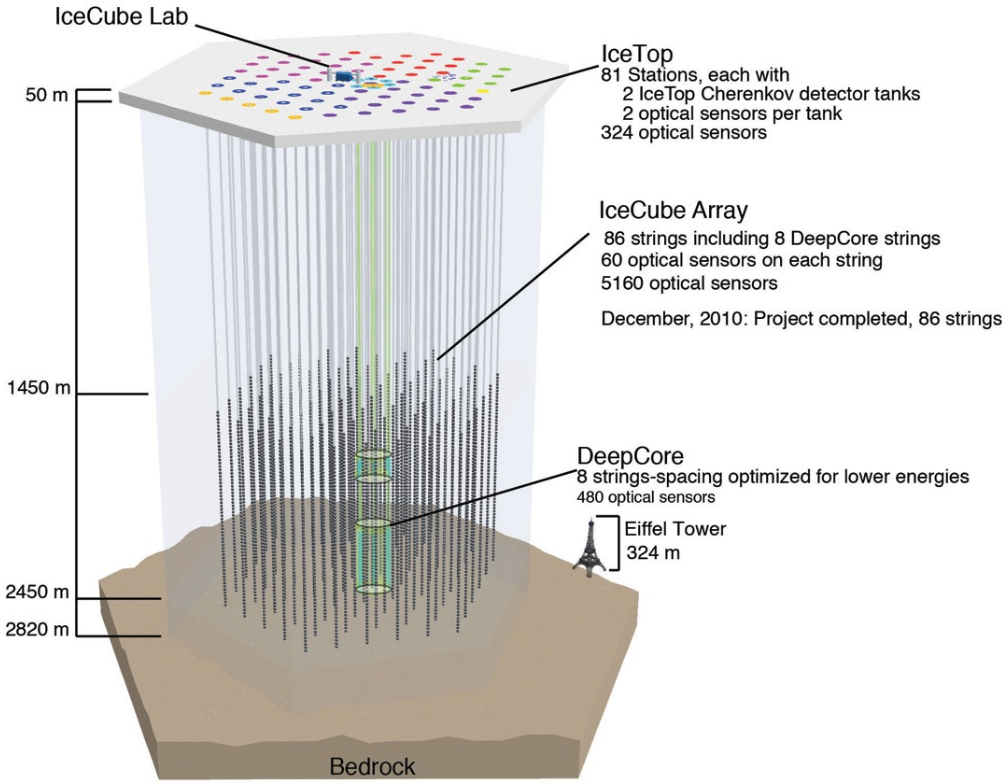


Figure 3.1: The landscape of IceCube, taken from IceCube gallery. The IceTop and IceCube Lab are located at the surface at South Pole, different colors represent different planed vintage years of the stations. The In-Ice detector array is deployed from 1450m to 2450m below the surface with each string (60 DOMs) associated with a station generally. (The DeepCore strings do not have their IceTop counterpart.) The IceCube is so large that it dwarfs the Eiffel Tower.

3.1 A General View of IceCube

The size of IceCube is about one kilometer three-dimensionally, and it was constructed over seven years. The completed IceCube has 86 strings (including 8 DeepCore strings) and 81 IceTop stations on the surface.

The geometrical arrangement of IceTop is depicted in Fig. 3.2. Each IceTop

---

[1] The main purpose of IceCube is to search for high energy cosmic neutrinos.

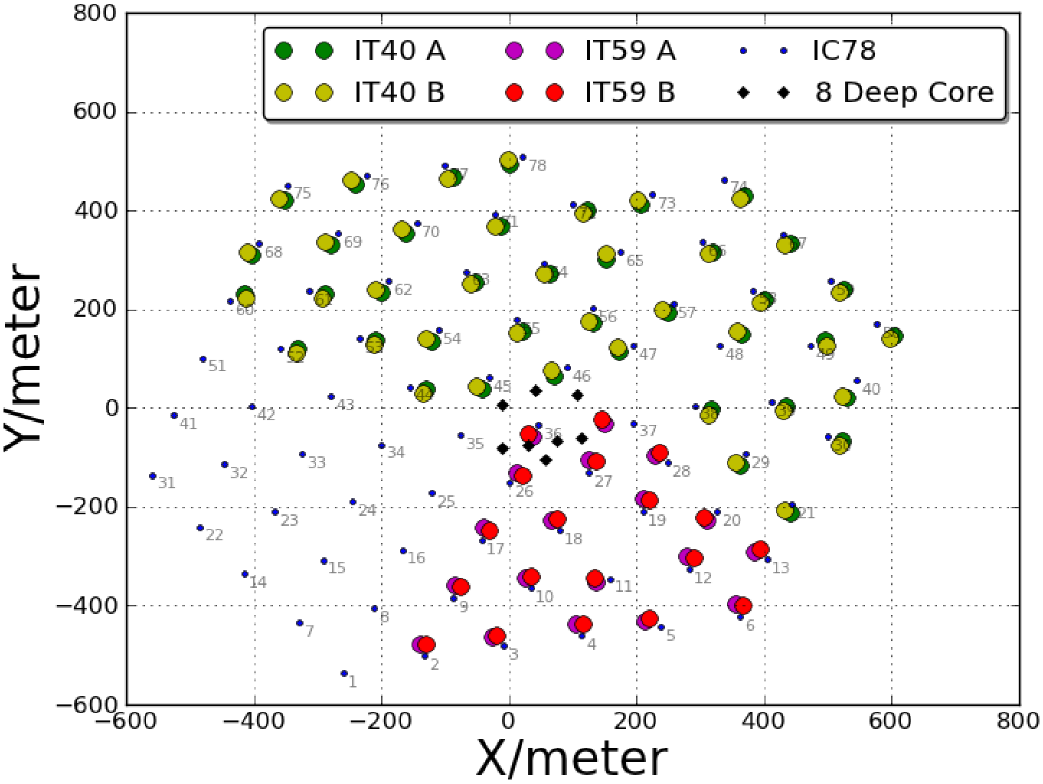


Figure 3.2: IceCube cross-section. The two tanks in IceTop stations and the first DOM under the ice are plotted. Stations constitutes so called IC40 and IC59 are shown in different colors. Note the overlapping of dots of tank A and B doesn't mean the real tanks are overlapping. The distance between two neighboring stations varies from about 120m to 160m.

station consists of two IceTop tanks, 10m apart from each other. A tank is a cylinder with two DOMs (one high gain and one low gain DOM) in it that is filled with ice. As is shown in Fig. 3.3, the cross section radius of a tank is about 0.9m and the height is about 1.2m. The height of the ice in the tank is 0.9m. Cables are used to connect the DOMs in the tanks to the data acquisition system.

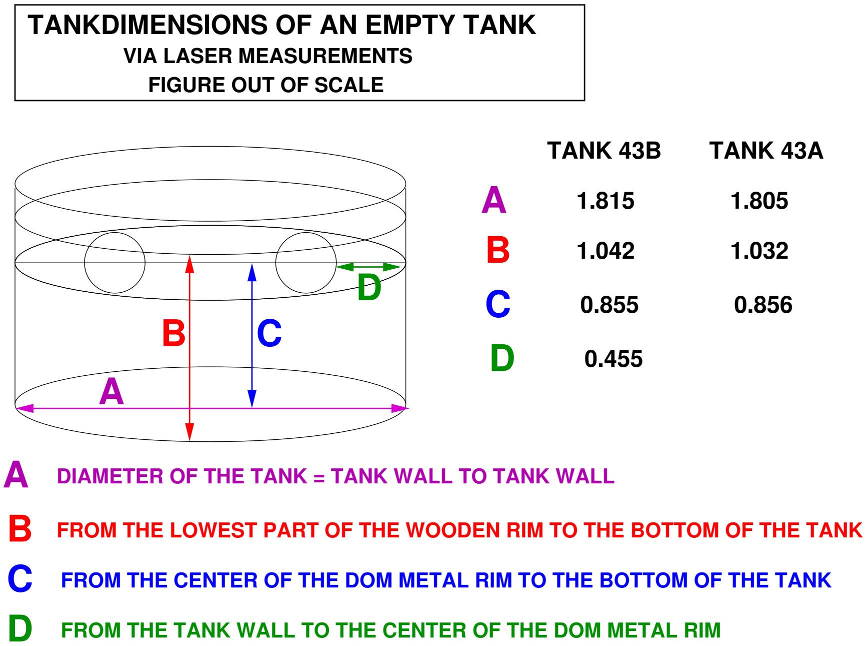


Figure 3.3: Tank dimensions (in meters) made by Arne Van Overloop for station 43.

3.2 Science Motivation

The primary goal of the IceCube project is to search for very high energy neutrinos from extraterrestrial sources. The largest possible effective area of IceCube is needed to compensate for the low flux of high energy neutrinos and the small neutrino cross section with matter. The standard neutrino searching technique is to look for up-going muon neutrinos that have penetrated the earth. The signal is a convolution of muon range R_μ, neutrino interaction cross section in the earth σ_ν and neutrino flux ϕ_μ

$$\text{Signal} \sim \text{Area} \otimes R_\mu N_A \otimes \sigma_\nu \otimes \phi_\nu, \tag{3.1}$$

where N_A is the Avogadro's number. At high energy nutrino attenuation in the earth also must be accounted for. The optimum detection range for muon neutrino is from 1 TeV to 1 PeV, in which the muon energy loss is greater than minimum ionizing. The predicted neutrino event rate is $f \times 30$ events/year·km^2, assuming both extra-galactic cosmic-rays and neutrinos have a power law spectrum of $E^{-2.0}$, where f is the efficiency for production of neutrinos relative to cosmic rays [18]. Different models predict different rates. For $f = 0.3$ this estimate gives a diffuse flux at the level of $E_\nu^2 dN/dE_\nu \sim 10^{-8}$ GeVcm^{-2}s^{-1}sr^{-1}, comparable to the Waxman & Bahcall upper bound [19]. IceCube will reach a sensitivity for diffuse fluxes of a few $10^{-9}E^{-2}$GeV^{-1}cm^{-2}s^{-1}, which is more than one order of magnitude below that conservative ”upper bound”. Point source searches in IceCube will achieve a sensitivity of at least 10^{-12}cm^{-2}s^{-1} for energies greater than 10 TeV. IceCube is expected to detect 10–1000 gamma ray bursts per year if certain models are correct. IceCube can improve the flux limits of relativistic monopoles by two orders of magnitude, complement future WIMP experiments, answer questions in neutrino oscillations, etc [20]. IceCube also detects a large number of atmospheric neutrinos from cosmic ray interactions in the atmosphere.

3.3 Digital Optical Module (DOM)

As the fundamental detecting unit of the IceCube Detector Array, the DOM (Fig. 3.4) takes samples of the photon density around the light sources, such as muon tracks with well-defined Cerenkov cones, or point-like cascades. The DOM is a glass sphere with a 25 cm diameter that encompasses a PMT, a main board, a high voltage generator, a LED flasher board with six pairs of LEDs. The thickness of the glass sphere is 13 mm. The DOM also has a μ-metal grid to shield the effect of magnetic field. Because the DOM is buried in the ice deeply, it must withstand the ice pressure. The required protection is provided by the pressure housing. All the DOMs are connected by network cables to the DOM hub at surface for online data processing. Some important DOM elements or characteristics are described as follows:

PMT [28] stands for photo-multiplier tube, which lies at the heart of the DOM.

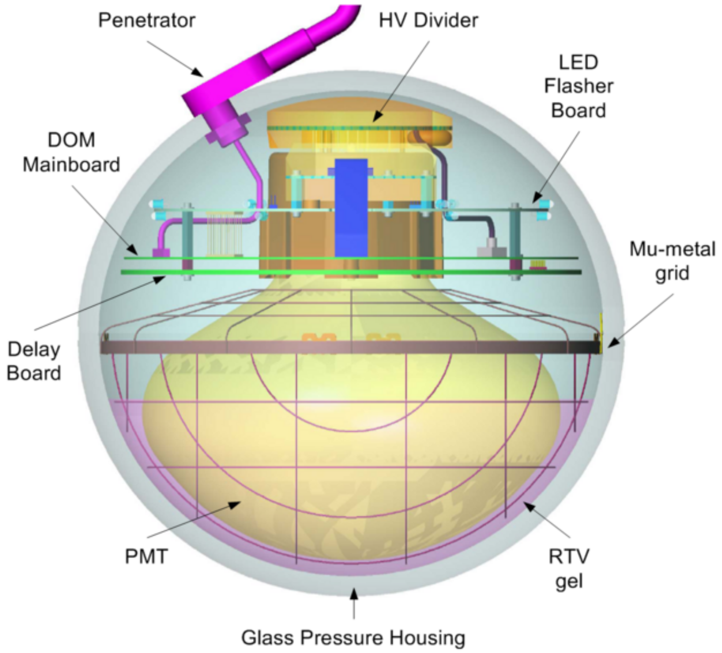


Figure 3.4: IceCube DOM. Included in the glass shell is the μ-metal grid, PMT, DB, LED flasher board, HV generator and divider, etc.

IceCube uses 10-stage PMTs made by Hamamatsu Photonics [2]. It has a wavelength range from 300 to 550nm and a cathode area of 470 to 530 cm^2. The peak quantum efficiency is 24%-28% and transit time is around 60 ns.

When a photon hits the photo-cathode, which is usually made of bi-alkali material, it knocks out a photo-electron. The electron is focused by a focusing electrode and then pumped to a series of dynodes. The photo-electron is accelerated and many secondary electrons are generated during this process. Finally the anode collects the all the signals, which usually multiplies the original electron by 1×10^5 to 1×10^7. A typical InIce DOM is running with a gain of 1×10^7 with dynode voltage of 1600 V, which gives a single photon pulse of 8 mV (the electronic noise level is 0.1 mV). IceTop DOMs have lower gains around 5×10^6 for high gain DOMs due to large amount of photons from air showers.

LED is the acronym of light emitting dynode. The LED flasher board (flashers)

---

[2] Hamamatsu Photonics K.K. is a Janpanese corporation that specializes in developing and manufacturing optical sensors, electric light sources and other optical devices.

contains 12 LEDs (6 horizontal and 6 tilted ones at 40 degree above the horizon) that can emit laser beams to the neighboring DOMs in IceCube, which is very useful in the calibration, verification of IceCube data and developing reconstruction algorithm of cascade events [24]. The LED light pulse width varies between 5 ns to 65 ns. The LED's light output is a function of brightness and pulse width. At full brightness it is able to emit 5.7×10^9 photons for a 50 ns pulse, comparable to the a muon track that sheds 1GeV energy per meter. The flash rate is usually set between 1.2Hz and 610Hz [25]. Currently all LEDs are 405 nm and plans are to install UV flashers (340 mn and 370 nm) and blue/green flasher (450nm, 505nm) [26].

DOM Mainboard integrates all the electronics that are used to amplify and digitize the signals of the PMT. The mainboard also performs communication between the DOM and data acquisition system on surface through a single twisted pair cable. The main components of the mainboard are

- 2kV power supply,
- analog input amplifiers,
- PMT trigger discriminator,
- local coincidence trigger,
- analog transient waveform digitizer (ATWD),
- fast analog to digital converter (FADC),
- field-programmable gate array (FPGA) to implement most of the low-level functions of the DOM,
- DOM oscillator clock,
- communication circuit,
- and on-board LED pulser.

SPE stands for single photo-electron. The DOM generates an SPE pulse when a single photon hit the photo-cathode of the PMT [3]. The waveform digitizers are triggered when the pulse reaches a quarter of the SPE peak amplitude [4]. The time resolution [5] of SPE hits is about 2 ns [28]. If a single pulse has contribution from multiple photons, the pulse is called multi-photoelectron

---

[3] Due to the less than 1 quantum efficiency, not every single photo can produce a pulse in the PMT

[4] For PMTs at gain of 1×10^7, the SPE peak is around 8 mV.

[5] The time resolution of the SPE hit is represented by the σ of a Gaussian fit.

(MPE). The most common pulses we encounter in cosmic ray reconstruction are MPE pulses.

The MPE and SPE discriminators triggers the signal captured with ATWD and FADC respectively [29]. The threshold for MPE discriminator in IceTop is 20 PE for high-gain and 200 PE for low-gain DOMs. The SPE discriminator is used only for solar cosmic ray studies and has a threshold below 1 VEM.

3.4 Data Acquisition System [30]

The digital output of a DOM upon capturing a photon is called a hit, which contains at least a time-stamp, the origin of the hit, a coarse measure of charge or waveform information. The waveform information (if it exists) has a time interval of 6.4 μs, which is larger than the duration of most IceCube events that can light IceCube. Once a pulse is generated in the PMT, a variety of electronic instruments transform the information to timestamped, time-calibrated and digitized data, which is sent to the DOM hub on the surface through cable network. The following are some main elements and techniques in the data acquisition system of IceCube:

ATWD stands for Analog Transient Waveform Digitizer, which is a waveform digitizer that takes four analog inputs. The four analog input channels include three PMT signals amplified at different levels of x16, x2 and x0.25, corresponding to ATWD0, ATWD1 and ATWD2 channels respectively. The fourth channel of the ATWD chip is two four-channel analog multiplexer that can be individually selected externally. The PMT waveform is sampled 128 times within 422 ns by the PMT channels of the ATWD chip. The time interval between two samples is about 3.3 ns. After the sampling is complete, the FPGA's ATWD readout engine initiates the digitization if certain trigger conditions are met.

During the digitization, the 128 Wilkinson 10-bit analog to digital converters (ADCs) digitize the analog singals stored in analog memory (capacitors). The samples are stored in a 128-word deep internal shift register and transferred into the FPGA after digital conversion. The ATWD0 (driven by the x16 amplifier) is converted first. If any sample in this channel exceeds 768 counts, ATWD channel 1 (x2 amplified input) is digitized. If channel 1 overflows, channel 2 (x0.25 amplified input) is digitized. This logic is shown in Fig. 3.5. The ATWD has a dead time of 29 μs, during which the captured waveform is being digitized and no other waveforms can be digitized. To minimize this

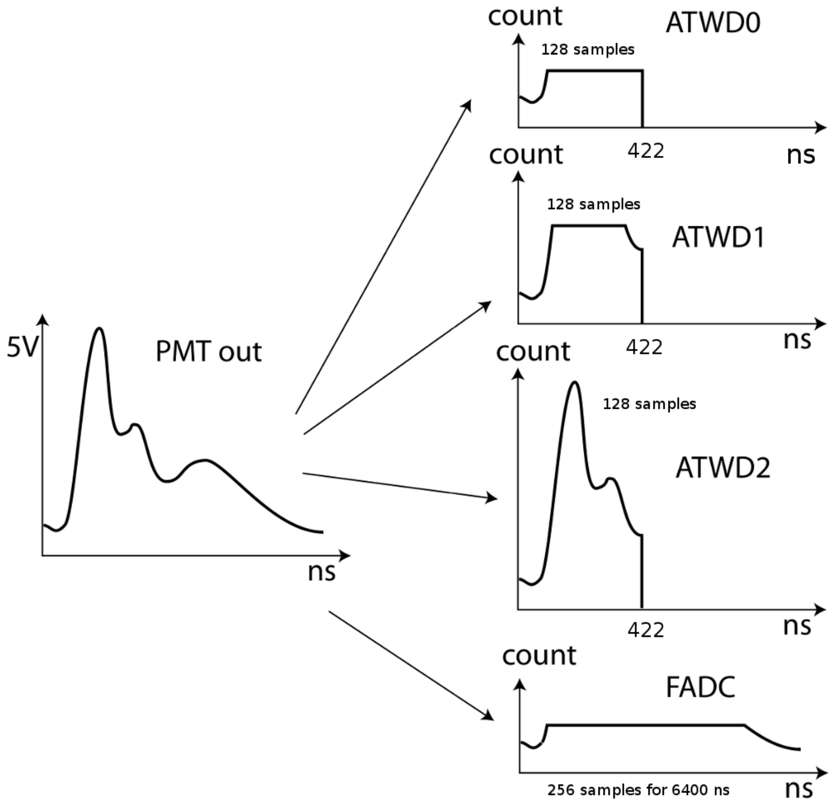


Figure 3.5: Schematic digitization logic taken from [32]. The PMT output signal is fed into three ATWD channels and the FADC channel. Each ATWD channel has a different amplification. Channel with the highest gain channel 0 saturates first, then the medium gain channel 1, the lowest gain channel 2. The ATWD channels take 128 samples and cover 422 ns, while the FADC channel takes 256 samples and covers a total of 6400 ns.

effect, the DOM is equipped with two identical ATWDs to separate the signal processing and capturing. If the trigger conditions to digitization are not met, the FPGA rests ATWD sampling circuitry and the signals are cleared [30].

FADC is the abbreviation of Fast Analog to Digital Converter. It is used together with ATWD for physics signals that last longer than can be covered by ATWD (about 422 ns). The 10-bit wide, parallel output and pipelined FADC digitizes the PMT waveform after passing a three-stage waveform-shaping amplifier with a 180 ns shaping time. The FADC continuously samples 256 PMT output signals for 6.4 μs, each with a bin size of 25 ns as depicted in Fig. 3.5. The FADC can be used to create a charge stamp for SLC hits, since an SPE signal from the PMT produces approximately 13 counts above the FADC baseline, sufficient to detect its presence [30].

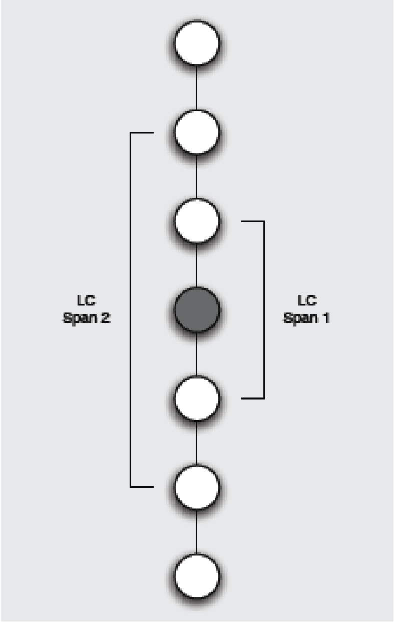


Figure 3.6: Schematic LC logic taken from [33]. Two modes of local coincidence: LC span 1 only considers coincidences between the direct neighbor DOMs; LC span 2 count the next next DOMs as well.

LC represents local coincidence. It basically means one DOM's neighbor or neighbors are also triggered within a certain time window. The LC capability is realized by connecting adjacent DOMs with bidirectional copper wires so that a triggered DOM can transmit this information to its neighbors. When the PMT signal exceeds the discriminator threshold (typically 0.25 for InIce DOMs), a hit occurs and the DOM opens a receptive time window of 1 μs. If the DOM receives a tag signal from its neighboring DOM within this time window, the local coincidence requirement will be satisfied. The two modes of LC are shown in Fig. 3.6, which are different in the definition of neighboring DOMs.

For a silent DOM, the LC tag it receives from its neighbors also initiates a receptive time window in case it receives a hit at a later time. Thus, the time symmetry feature of a LC signal is preserved. A PMT hit that is LC-tagged indicates there is a triggered neighboring DOM present. The LC tagged hits account for only a few percent of all hits that select signals from real particle rather than from PMT noise. By excluding those untagged hits, the data size is reduced dramatically and the data selection is more efficient.

The LC in IceTop has similar meaning but is implemented differently. Unlike in IceCube where neighboring DOMs are considered, the tank (which includes a high gain and a low DOM) is the basic unit in IceTop local coincidence. If both tanks in a station contain hits, the LC condition is met.

SLC hits are hits that do not satisfy the LC condition. They are recorded for monitoring use.

The cable network connects the IceCube and IceTop DOMs to the DOMhub. It supplies powers for the DOMs, transfers signals from one DOM to the DOMhub, and sends LC siginals between DOMs. The cable at surface connects the DOMhub in IceCube Laboratory (ICL), the surface junction box located near the top of each string hole, and the IceTop tanks. The DOMhub is a computer in the ICL that communicates with the main boards of all the DOMs in a string.

The master clock is one of the central components of the IceCube timing system that provides each DOMhub with a high precision internal clock synchronized to UTC [31] by using the Global Positioning System (GPS). The GPS receiver in ICL is synchronized to the UTC master clock at the US Naval Observatory. The system has an accuracy of about ±10 ns over 24 hours. The fan-out subsystem distributes the 10 MHz, the 1 Hz, and encoded time-of-day data from the GPS receiver to DOMhubs through an active fan-out [30].

The performance of IceCube data acquisition system has been tested with flasher data in timing, PMT linearity, temperature variation and reliability. The time resolution of the DAQ system has been determined from LED flasher data and cosmic ray muons. One way to look at the time resolution is to measure the photon travel time from one DOM to an adjacent one on the same string [6]. The 1.5 ns standard deviation in the result indicates that the ATWD time resolution is less than 1 ns assuming the dominant contribution to time resolution is the error in time calibration. Another measurement is to let a DOM emit light to its two closest neighbors with ±17 m from it, which gives a very close result to the first one. The time resolution for FADC is about 4.7 ns. The time resolution can also be obtained from the reconstructed muon tracks in the InIce part of the IceCube.

---

[6] The two DOMs are 17 m apart.

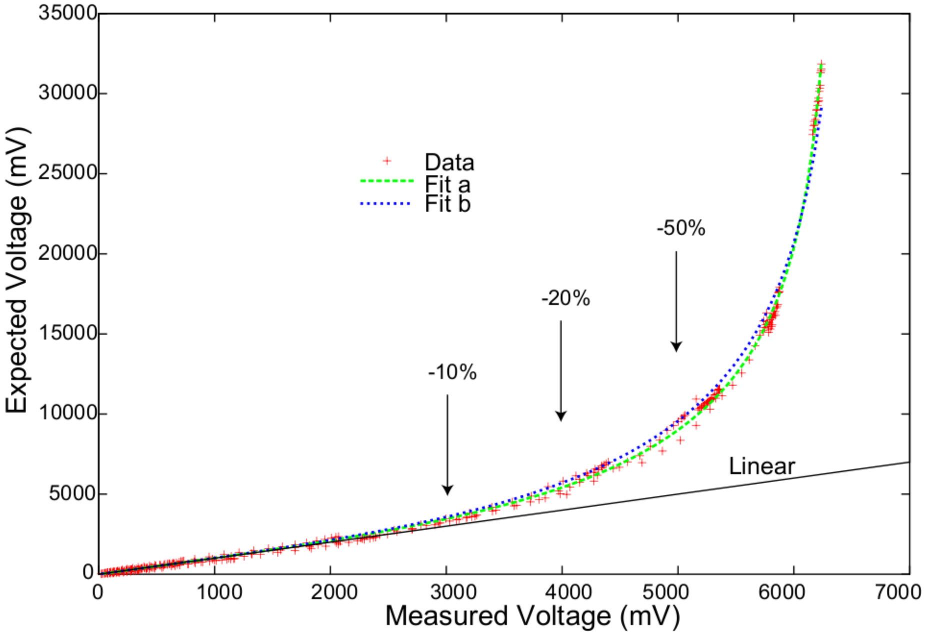


Figure 3.7: Measured PMT linearity for one phototube from [30].

The PMT charge is obtained by calculating the area of the waveform, which is affected by the linearity of the electronic signals at different gains. Fig. 3.7 shows the deviation reaches 10% at ∼400 photoelectrons/15 ns (3000 mV). Calibrating the PMT response in the non-linear region can extend the dynamic range beyond 1000 photoelectrons/15 ns. Fig. 3.8 shows the charge distribution for one day for IceTop-InIce coincident events that will be used in composition analysis in this thesis. The fact that the majority of pulses only have charges of a few PEs indicates the PMT works in its linear range most of the time.

The timing variation of InIce DOMs is constant throughout the year, while IceTop DOMs are subject to local and structural changes in temperature. The overall change in IceTop DOM launch rates [7] is about 20%, 15% of which is due to the day-to-day changes of barometric pressure and the remaining variation could

---

[7] DOM launch rate is the rate that a DOM participates in an event.

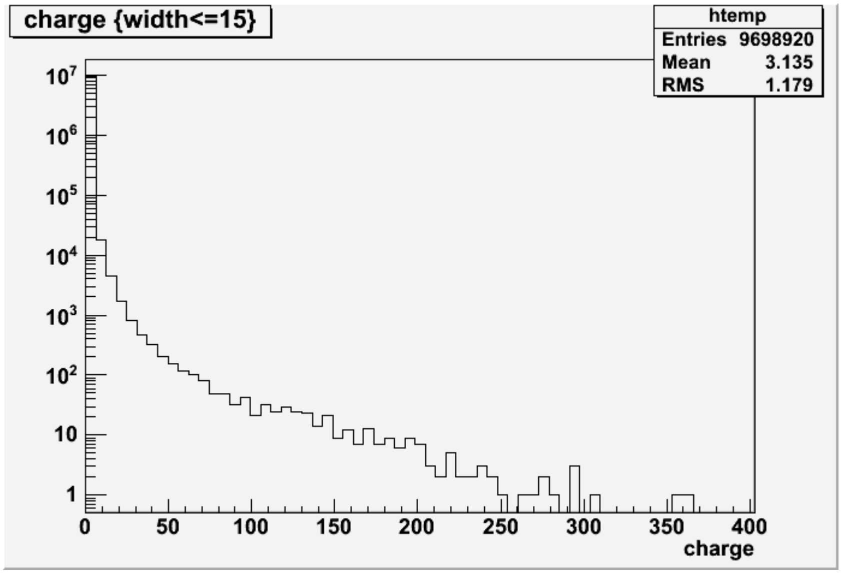


Figure 3.8: Charge distribution of IceTop SMT 3 and InIce Coincidence events on Jan. 01, 2010.

come from the seasonal changes of the atmosphere. Though IceCube DOMs are highly reliable, some of them failed after deployment due to various reasons, such as the loss of PMT vacuum, electronics failure and so on. The failure rate of IceCube DOMs is about 0.5% at the end of the fouth operational year, which suggests 97% of DOMs may survive in 25 years. At any given time during operation of the IceCube and IceTop array for data taking, 97%–98% of the deployed DOMs are operating properly [30].

3.5 IceTop DAQ

The IceTop DAQ is completely integrated into the IceCube DAQ with different configurations from their InIce counterpart. Each IceTop DOM is connected to one wire pair and 32 DOMs share one DOM hub. The IceTop trigger is based on individual DOMs rather than tanks or stations. The HLC (Hard Local Coincidence) in IceTop requires that both HG DOMs in a station detect a signal within $1\mu s$. LG DOMs listen to HG DOMs in the other tank. Only LC-hits participate in the trigger. SLC (Soft Local Coincidence) hits in IceTop are described by time and charge stamps. There is no waveform in a SLC hit. SLC hits are read out for every triggered event. Currently there are three IceTop triggers, they are simple majority trigger, minimum bias trigger and calibration trigger. The simple majority trigger requires 6 LC-hits within $5\mu s$ and has a readout window of $10\mu s$ before and after the trigger time. The minimum bias trigger picks up one event in 10000 and shares the same readout window with simple majority trigger. The calibration trigger is for muon calibration. IceTop is always read out with any InIce trigger and vice versa.

3.6 IceTop Calibration

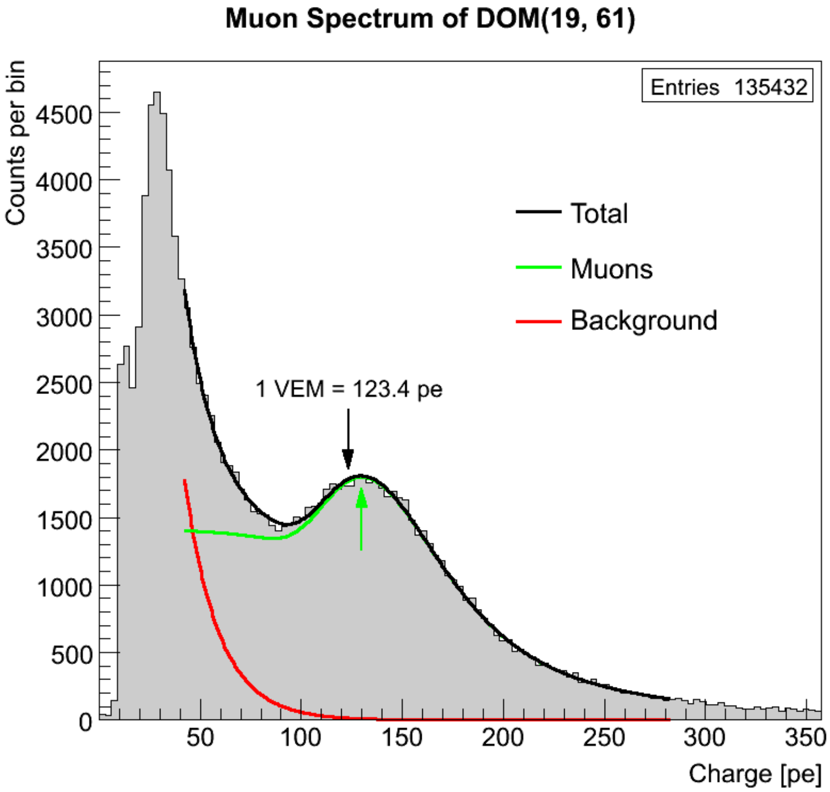


Figure 3.9: Muon spectrum of DOM(57,61) on Dec.25,2010. The muon peak is fitted by function $p(x) = \frac{p_0}{p_2\sqrt{2\pi}}\exp\left(-0.5\left(\frac{x-p_1}{p_2} + \exp\left(-\frac{x-p_1}{p_2}\right)\right)\right) + \exp(p_3 + p_4 x)$ [34]. The plot is taken from [35].

The IceTop part of the IceCube consists of stations associated with the InIce strings. Each of the two tanks of a station contains two DOMs that are configured differently from their InIce counterparts. The high gain and low gain DOMs are operating at 5×10^6 and 5×10^5 gain respectively. The threshold is set to 1023 DAQ setting, which is about 10 pe at 5×10^6 gain, compared to the 0.25 pe for InIce DOMs. The DOM signal is calibrated to the number of photoelectrons (PEs) first and then converted to vertical equivalent muons (VEMs). An example of the charge spectrum of an IceTop DOM is shown in Fig. 3.9. The peak of the vertical muon spectrum is at 123 PEs (in black), which is about 95% of the peak of the full spectrum (in grey). Applying the 95% factor to the measured peak in the experimental data gives the VEM calibration for each DOM. The SLC is calibrated to agree with corresponding HLC charge with about 3% error. The snow height can be obtained by taking the ratio of muon to the background signals as shown in Fig. 3.10. This is because as the snow accumulates, electromagnetic signals (background) are reduced but muon signals remain much less affected. The relationship between the signal to background ratio and snow depth is

$$\frac{S_\mu}{B_{EM}} \approx \exp\left(\frac{h_{snow}}{1.3655} + 1.7521\right), \tag{3.2}$$

where S_μ and B_{EM} are computed as the muon signal spectrum and electro-magnetic background integrated from 0.3 to 2.0 VEM.

3.7 Event Builder

As a multi-purpose detector array, IceCube serves various scientific goals for different working groups that have very different requirements. Since each working group in IceCube needs different kinds of events, assuring working groups an easy way to access the category of data they require is critical in IceCube online data processing as well as keeping the total data size at a reasonable level. In IceCube, the Triggering, Filtering and Transmission (TFT) board takes proposals from each

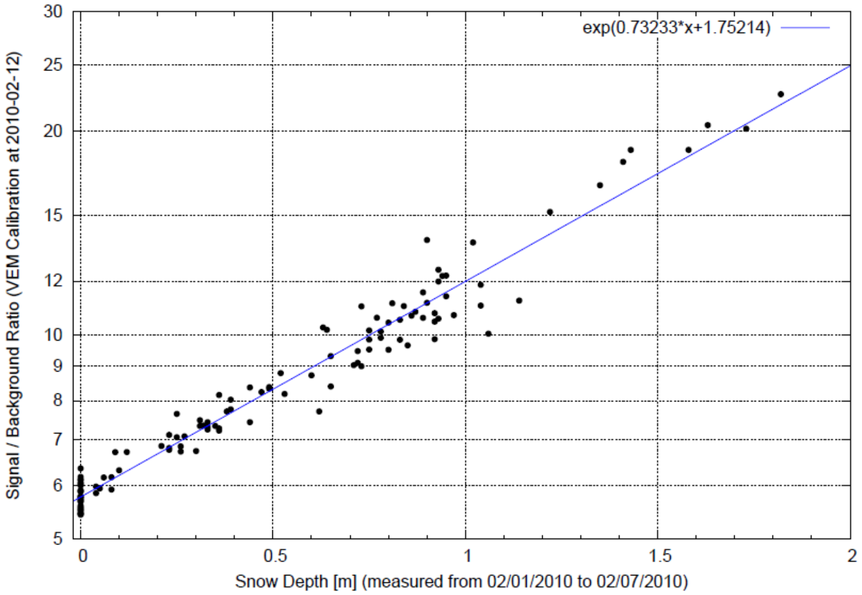


Figure 3.10: IC59 data. Integral from 0.2 to 2.0 VEM.

physics working group and decides the DAQ software and trigger settings, online filter settings and satellite transmissions resources. Selecting triggered events is the first process of the South Pole online data processing. A typical detector trigger is simple majority trigger (SMT), which requires at least 8 InIce hits [8] or 6 IceTop hits that are generated within a time window of 5 μs to trigger the detector [36]. The readout window for InIce SMT trigger is $(-4000\ ns, +6000\ ns)$ and the one for IceTop is $(-10000\ ns, +10000\ ns)$ [37]. The global trigger looks for overlap and merges the triggers, potentially lengthening the events even more. After that events will be sifted through filters for different sources based on the results of the online fits. The filtered events will be sent to the North by satellite. The IC59 IceCube has a trigger rate of 1500 Hz and a filter passing rate of 112 Hz. Examples of filters in IC59 IceCube are [38]

- muon filter tries to select muon-like events over the whole sky. It has an

---

[8] They could be HLC or SLC hits.

estimated rate of 35.3 Hz and requests 13.1 GB bandwidth per day;

- cascade filter looks for point-like events in the detector. The event rate is about 24 Hz and requested bandwidth is 8.6 GB per day.
- EHE filter selects big events with a rate of 2.4 Hz and requests 3.0 GB bandwidth per day.
- IceTopSTA3 filter filters out events with less than 3 stations with a prescale [9] of 8. The rate is about 2.4 Hz and the bandwidth is 2.2 GB per day.
- IceTopSTA3_InIceSMT requires both passing IceTopSTA3 and InIce SMT trigger with a prescale of 3. The rate is about 1.5 Hz and the bandwidth is 1.8 GB per day.
- InIceSMT_IceTopCoincidence selects events with InIce SMT trigger and any IceTop station hit with a prescale 80. The event rate is about 0.4 Hz and the requested bandwidth is 1.0 GB per day. . . .

The total data size is less than the sum of the data size required by the individual filters since some of the filters are overlapping each other. The raw ATWD and FADC event data are then calibrated online/offline to before doing any physics analysis.

3.8 Detector Simulation and Offline Processing

Detector simulation and offline data processing in IceCube are conducted throught ICETRAY framework [10]. The simulation chain, as depicted in Fig. 3.11, starts from a generator, which simulate the physics the detector observes. The

---

[9] Prescale is the fraction of events transferred from South Pole by satellite.

[10] Detailed information about ICETRAY can be found in [40].

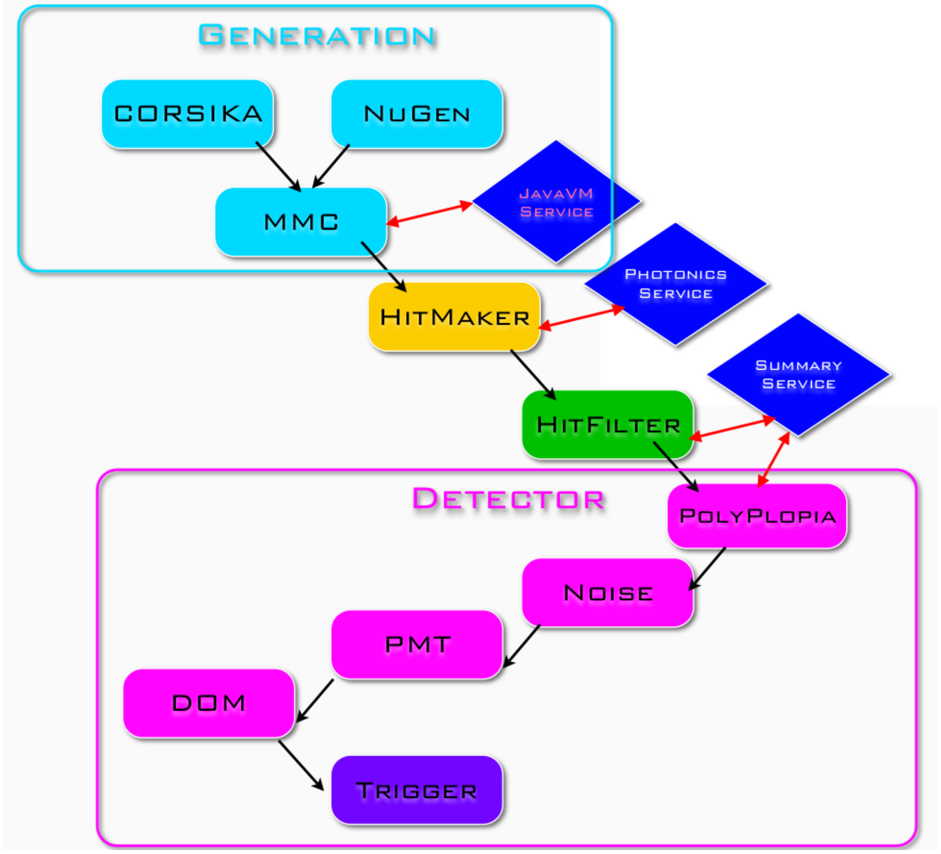


Figure 3.11: Schematic of simulation chain taken from [39]. Some parts are subject to changes according to the physics goals.

generator could be CORSIKA [11], dCORSIKA, or neutrino generator, which performs the simulation of cosmic ray showers or single neutrinos and feed the muons to MMC that propagates the muons through the underground part of the detector. The Hit-Maker makes the expected number of photoelectrons in the PMT out of the muons by using photonics tables [12]. The detector simulation includes PMT, DOM and triggering simulator, each of which simulates an essential part of the IceCube detector arrays [13].

---

[11] Cosmic Ray Simulation for KASCADE

[12] An alternative is to use Photon Propagation Code.

[13] IceTop requires an additional simulator.

The IceTop-InIce coincident simulation starts with the generation of CORSIKA air showers. The showers are produced with an energy spectrum of E^{-1} and evenly distributed with respect to $\cos\theta$ for $0 \leq \theta_{max}$. The energy range is from 100 TeV to 100 PeV. Each shower is then thrown randomly to the South Pole area 100 times. The area is a disk of a 1200 meter radius with IceTop detector array at the center. This area is very sufficient since the analysis only uses IceTop and InIce contained events. This means each 100 generated events share the same CORSIKA shower so that they have the same primary and secondary particles but different surface core locations. Because of the cuts, especially the containment cuts, only about 1% of the events will be selected. Thus, each CORSIKA shower is used only once on average. The *TopSimulator* simulates the IceTop response to the showers and passes the muons to the InIce part of the simulation. The MMC propagates those muons and feeds them to hitmaker, which makes hits out of the muons followed by a full InIce detector simulation. The triggers and filters test if the simulated event satisfies a variety of trigger and filter conditions [14].

The offline data processing contains the following modules: a calibrator calibrates the DOM signals, waveform processor/extractor extracts charge and timing information from the waveforms, and a variety of track and energy reconstructions. The reconstruction results are the inputs of physics analysis in IceCube events. The flexibility of ICETRAY framework allows users to process the simulation or experimental data in their own way if the standard data processing doesn't satisfy their needs. For example, the users can switch among different calibration files, adding their own reconstruction algorithm, and changing the configuration of the standard

---

[14] One thing that all IceTray programmers should pay attention to is that creating an I3Particle from copying an existing one would not give the new I3Particle a new minor ID. This would become a serious problem when using I3MCTree utilities to find I3Particles. The ignorance of this problem in the original developing of TopSimulator caused a disaster in the InIce simulation until I found the bug in December, 2010.

reconstruction modules.

Located at the South Pole, IceCube detector array distinguished itself from other large ground-based detectors by using the ice as the detecting medium for moving charged particles. Though great experience of in-ice detector response has been obtained from the AMANDA [15], new studies on ice property and photon propagation must be done with IceCube because of its unprecedented volume and depth in the ground. These will be discussed in the following chapter.

---

[15] Antarctic Muon and Neutrino Detector Array

Chapter 4

GEOMETRY, CALIBRATION AND DETECTOR STATUS (GCD)

The geometry, calibration data and detector status information are key factors in IceCube data analysis. Separate binary GCD data files are used along with physics data in data processing and detector simulation. The GCD file stores all the physical settings and detector configuration of IceCube at a certain time. Unlike the data, it's relatively static so that the same GCD file is used in detector simulation for a whole year. However, to make sure the GCD reflects an accurate detector status at any point of time, each run in experimental data has its own GCD file. In this chapter, we'll discuss the sources of GCD, how the values are stored and where they go in data processing. Separating the GCD files from physics file reduces the size of data significantly since a typical IC59 GCD is about 66 MB.

4.1 Geometry

The geometry class of the GCD file, namely, carries the information on where and how IceCube detectors are located. The major elements in the geometry class are start time, end time, DOM position, DOM type, DOM orientation, DOM effective area and the relative rotation angle of DOM in azimuth. The start and end time are the time period of the run associated with the GCD file. The DOM type tells whether this is a IceCube or an IceTop DOM [1]. The orientation of the IceCube

---

[1] The type of the DOM could also be an AMANDA DOM. Since we consider only IceCube here, we ignore this type. AMANADA was the predecessor of IceCube.

DOMs are down which means all the DOMs are facing down. This option is to optimize the detection efficiency for upward-going muons from the other side of the earth. The effective area of the DOM is 0.0444 m^2 and the azimuth rotation angle is 0. The physical positions of the IceCube DOMs are given in the IceCube coordinate system.

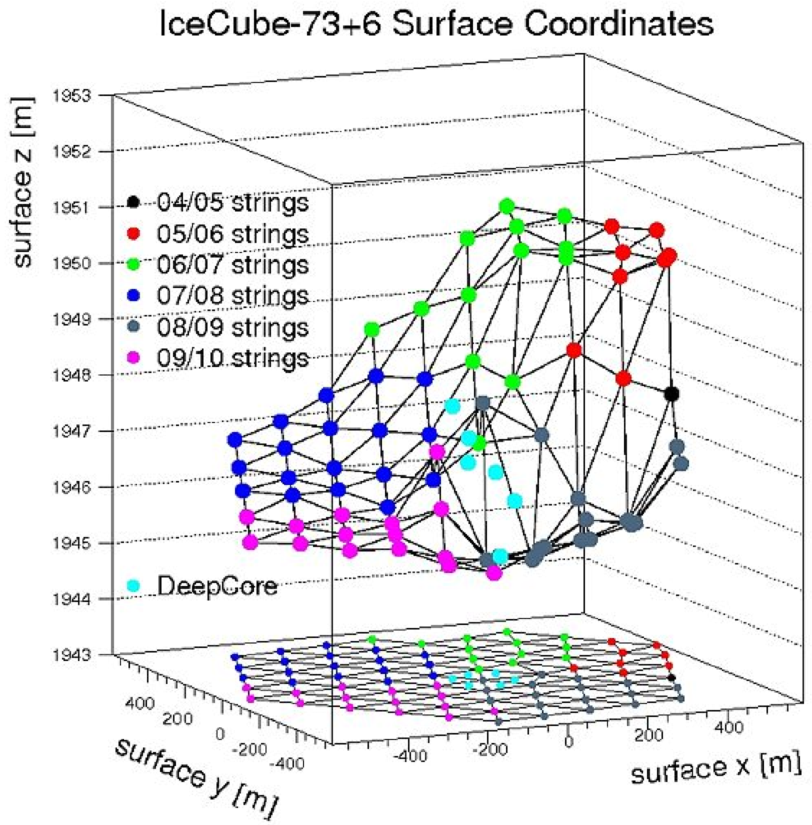


Figure 4.1: IC79 surface coordinates from [41]. The snow surface is sloping toward the southwest and the 07/08 and 09/10 strings have systematically lower surface coordinates than the earlier strings.

The origin of the IceCube coordinate system is 46500’E, 52200’N, in easting and northing measurement, at an elevation of 2900 ft [42]. The y axis is Grid North that is aligned with Prime meridian, pointing towards Greenwich, UK. The x axis is Grid East that points 90 degrees clock-wise from Grid North. The z axis is normal to the Earth's surface, pointing up, completing a right-handed coordinate system. The origin of the coordinate system is chosen to be the center of the In-Ice array. The surface coordinates for the IceCube holes are measured with surveys of the drill tower shortly before deployment. The offical depth in IceCube is the vertical distance from the floor of the tower used for deployment of the first string in IceCube–string 21. Since this distance is 1948.07 m, the relationship between depth d and z coordinate is

$$d = 1948.07 - z.$$

Fig. 4.1 shows the surface coordinates for the IC79 stations that are calculated from the surveyed corners of the drill tower. The lower levels of the newly completed stations are due to the inclined snow surface. The (x,y,z) coordinates for all DOMs in IceCube are obtained from geometry calibration, which is done in three stages. The first stage uses non-optical data collected during string deployment and a pre-deployment survey of the drill tower. The IceTop shares the same coordinate system with IceCube. The locations of IceTop DOMs are determined by direct survey. The locations of the DOMs in the tanks are calculated from the known tank geometry from the locations of the center of the tank and the survey pipe as surveyed by the professionsl surveyors.

The second stage of geometry calibration uses inter-string flasher data to determine the relative depth offsets between the strings. This is a correction to the first stage.

The third stage uses down-going muon data to track the deformations of the array over time due to the ice shear. In this stage the strings are not assumed to be straight any more so DOMs can be tracked individually. The abundance of down-going muons reduces the systematics due to track reconstruction. The accuracy of DOM location is 1 meter.

4.2 Calibration and Detector Status

The calibration files store the data required in the DOM calibration of waveform, charge and time. The users are allowed to develop their own algorithms to convert the waveforms to physical units. Details of DOM calibration files can be found in Appendix D.

The detector status is the state of the aspects of the detector that people have direct control over. It contains the DOM, trigger and DOM hub configurations of the IceCube. Details of status files can be found in Appendix E.

4.3 Calibration Usage

This section discusses the principle of ATWD and FADC waveform calibration.

4.3.1 ATWD Waveform Calibration

The first step in ATWD waveform calibration is to convert raw ATWD waveforms into voltage by using a linear fit for each waveform bin. A front-end bias voltage (pedestal) is subtracted from this value. Any remnant baseline will be subtracted further. This value can be expressed as

$$\mathrm{V(id,ch,bin)} = m\mathrm{(id,ch,bin)}\cdot\mathrm{waveform(id,ch,bin)}+b\mathrm{(id,ch,bin)}-\mathrm{V_{bias}}-\mathrm{baseline(id,ch,bin)},$$

where m and b are the slope and intercept of the linear fit. The final voltage has to be divided by the channel gain for the appropriate channel. If data are pedestal subtracted the calibration is different. After pedestal suppression, a constant offset remains to allow the waveform to be positive. This offset is currently the average of the pedestal waveform for that ATWD channel. Thus, the baseline is reconstructed for each ATWD bin and subtracted from the waveform before the waveform is converted to voltage. The amplified charge of ATWD waveform is obtained by integrating the voltage over time:

$$\text{charge (pC)} = 10^{12} \cdot \frac{1}{Z/\Omega} \cdot \frac{1}{F/Hz} \cdot \sum_i V_i,$$

where Z is the front-end impedance in ohm and F is the ATWD sampling frequency in Hz. Divide the amplified charge by the gain of the PMT for the operating high voltage gives the integrated charge of ATWD. The relationship between the gain and voltage is

$$\log_{10} Gain = \text{slope} \cdot \log_{10} V + \text{intercept}.$$

Details about ATWD time calibration can be found in Appendix F.

4.3.2 FADCWaveform Calibration

The FADC waveform is converted to voltage by multiplying the baseline-subtracted waveform by a constant G. The time calibration of FADC is described as follows [44]:

1. Determine the FADC sample position of the feature.
2. Determine the offset T_{offset} from the FADC launch, using the FADC sampling frequency of 40 MHz.
3. Add the offset to the ATWD waveform start time and the FADC inherent time offset Δ_{FADC} from the calibration data file:

$$\begin{aligned} T_{hit} &= T_{offset} + T_{ATWD} + \Delta_{FADC} \\ &= \frac{bin \cdot 10^3}{F_{FADC}/MHz} + T_{launch} - (T_{transit} + \Delta_{ATWD}) + \Delta_{FADC}, \end{aligned}$$

where F_{FADC} the FADC sampling frequency is 40 MHz. The Δ_{FADC} is negative in the calibration results indicating that the FADC digitization window starts before the ATWD.

Chapter 5

LIGHT IN THE ICE

The InIce part of the IceCube detector array consists of 5160 of digitial optical modules(DOMs), each of which has a transparent pressure sphere that encompasses a photomultiplier tube (PMT). The PMT records the photons that arrive at it in the ice[1]. The number and timing the photoelectrons recognized by the DOMs are used in various angular and energy reconstruction of IceCube events. Thus, understanding the sources and propagation of light is the key to an accurate event reconstruction and composition analysis.

5.1 Light Sources

The IceCube detector is designed to detect the Cerenkov light from charged particles between 1450 m to 2450 m underground in the ice. Most of the signals in IceCube are atmospheric muons produced in the air showers. Many of the tools in IceCube are designed for neutrino science but they work well for the muon bundles. There are two types of light sources, the point-like cascades and muon tracks. The cascades could be generated in either a charged or neutral current process. In the charge-current process, a lepton is produced that carries 50% to 80% of the energy of ν_e and ν_τ with the rest energy transferred to the nuclear target and released in the form of a hadronic shower. In the neutral current interaction, no lepton is produced and the cascade is the hadronic shower. Cascades from all the above interactions

---

[1] The acceptance of the PMT varies with the wavelength of photons. The most efficiency region is between 300 nm and 600 nm.

are indistinguishable in IceCube except that τ lepton with high enough energy can produce two cascades about several hundred meters away from each other [45]. This effect is called "double bang". The muon tracks are muons either from the collision of ν_μ with a nucleus in the ice inside the detector volume, or the remaining muon bundles in the extensive air showers. Besides the continuous ionization energy loss, muons lose energy stochastically by bremsstrahlung, electromagnetic interactions and e^+e^- pair production along the track.

5.1.1 Cerenkov Radiation

Cerenkov radiation occurs when a charged particle moves with a velocity greater than the local phase velocity of the light in a medium. The angle θ_C of the Cerenkov radiation respect to the particle direction is $\cos\theta_C = 1/n\beta$, where β is the ratio of the speed of particle and the speed of light and n is the index of refraction in that medium. The threshold velocity β_t is $1/n$ (see Fig. 5.1). The number of photons produced per unit path length of a particle with charge ze and per unit energy interval of the photons can be expressed as [1]

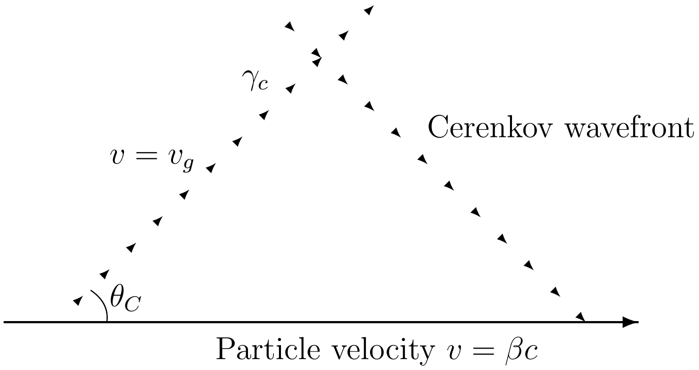


Figure 5.1: Cerenkov light emission and wavefront angles.

$$\frac{\mathrm{d}^2N}{\mathrm{d}E_\gamma \mathrm{d}x} = \frac{\alpha z^2}{\hbar c}\left(1 - \frac{1}{\beta^2 n^2(E_\gamma)}\right) \simeq \frac{\alpha z^2}{\hbar c}\sin^2\theta_C(E_\gamma). \tag{5.1}$$

Taking into account the geometry effect described in Fig. 5.2, the number of photons detected by the optical module is approximately [1]

$$\frac{2\alpha}{\hbar c}\frac{r^2}{\pi d}\exp(-\frac{d}{\lambda_a \sin\theta_C})\sin\theta_C\int Q(E_\gamma)\mathrm{d}E_\gamma,$$

without considering the photon scattering and angular efficiency and assuming a single absorption length λ_a for all photon energies. $Q(E_\gamma)$ is the quantum efficiency of the photomultiplier.

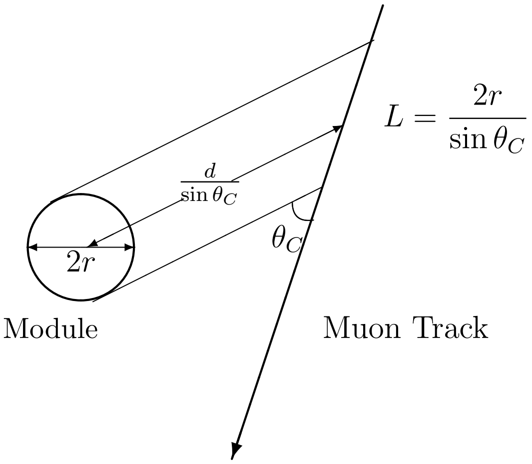


Figure 5.2: Detection of Cerenkov light emitted from a muon track with an optical module (DOM). The DOM has a radius r. θ_C is the Cerenkov light, L is the radiation length and d is the impact parameter.

5.1.2 Muon Energy Loss in Ice

The dominant muon energy loss in ice is through the processes of the ionization of the medium through which the muon passes, bremsstrahlung, electromagnetic interaction with nuclei and e^+e^- pair production [46]. While ionization energy loss is continuous, the other processes include stochastic bursts along the muon trajectory, and are only important at high energy [47]. The average energy loss rate for muon could be written as [47]

$$\frac{\mathrm{d}E}{\mathrm{d}X} = -\alpha - E/\xi, \tag{5.2}$$

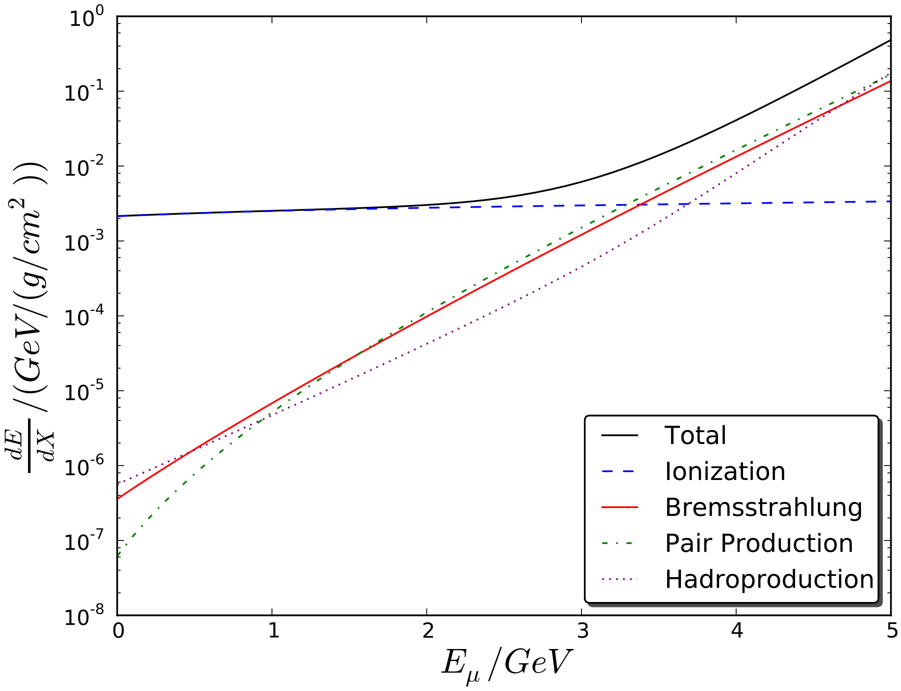


Figure 5.3: Muon energy loss rate in ice for ionization, bremsstrahlung, pair and hadronic production respective for muon energy up to $10^5 GeV$. The data is obtained from Todor Stanev's simulation results present in course PHYS638 in University of Delaware in March, 2009.

where the first term α is the continuous energy loss, which depends somewhat on the medium and nearly constant for relativistic particles. For south pole ice, the magnitude is about $2MeV \cdot cm^2/g$. The second term in equation 5.2 describes the stochastic energy loss, which is the sum of bremsstrahlung, pair and hadronic production $\xi^{-1} = \xi_B^{-1} + \xi_{pair}^{-1} + \xi_{hadronic}^{-1}$. The denominator of the coefficient ξ is also energy dependent and levels at high energy. For the south pole ice, ξ is about $2.5k$ m.w.e. Above the critical energy $\alpha\xi \approx 500$ GeV, the discrete radiative processes are more important than continuous energy loss [48]. The partial energy loss rate due to different processes are shown in Fig. 5.3.

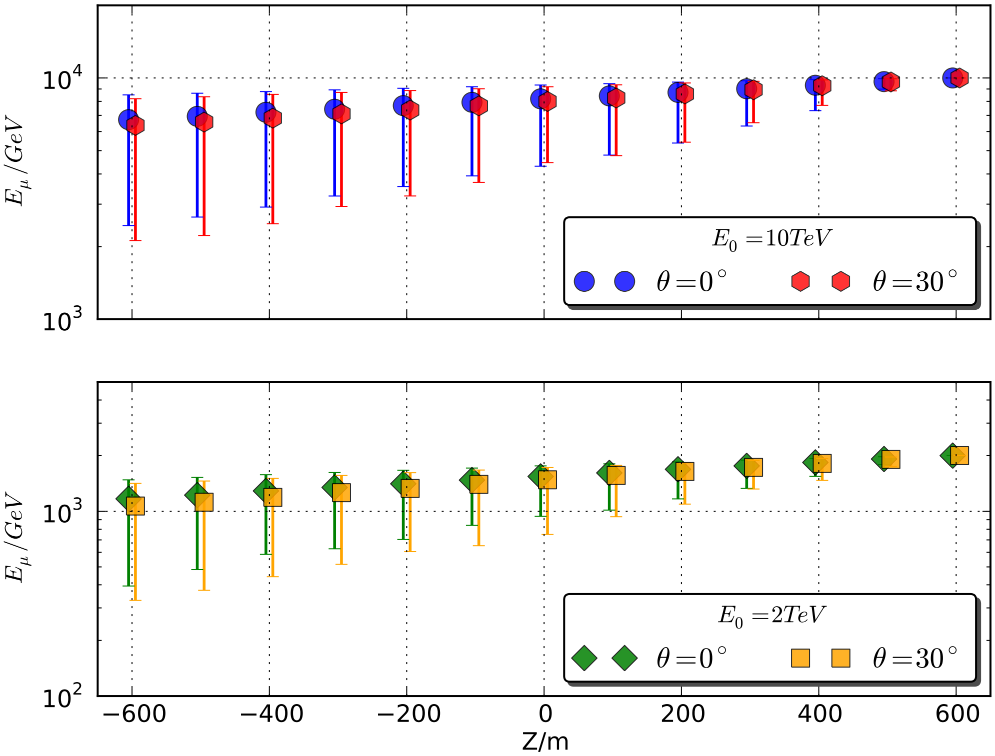


Figure 5.4: Muon energy in IceCube Coordinates (the zero is 1948 meters from the ice surface) for vertical and 30° muons from simulation. 1000 single muons are thrown at (100m, 250m, 600m) with respect to IC40 geometrical center at 0° and 30° with 2TeV and 10TeV starting energy respectively by using SIMPLE-INJECTOR. The muons are then propagated by MMC, which determines the ionization and stochastic energy losses of muons. Average muon energy at depths with 100m interval is taken and error bars represent 90% of total events. The inclined muons have lower energy of vertical ones because of the longer track length traveled by them for the same depth interval. Note the X axis is from the bottom to the top of IceCube.

However, the real muon energy loss rate is far less smooth than what equation 5.2 describes. Fig. 5.4 and Fig. 5.5 show vertical and inclined single muon energies decreasing as they go through the IceCube InIce detector at rates approximately consistent with equation 5.2 but with high fluctuations. Moreover, the distribution of muon energy loss is not symmetric, which makes it difficult to determine the muon energy on an event-by-event basis by using the analytical formulas.

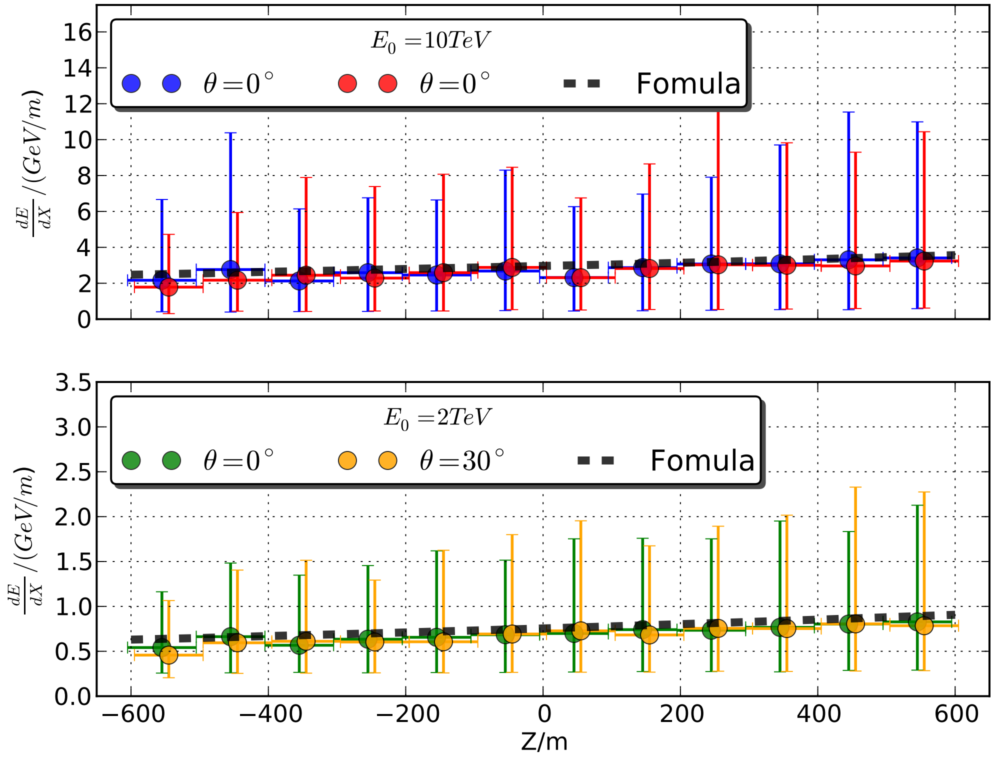


Figure 5.5: The same sets of muons are used as in Fig. 5.4. The energy loss rate is computed as the difference between the muon energy at two neighboring points divided by the distance between these two points ($100m/\cos(\theta)$), where θ is the zenith angle of the muon. which in this case is 0° or 30°. Also the average muon energy loss rate is taken from 1000 sample events and error bars represent 90% of total events. The intermittent line is taken from the vertical muon energy data by using the formula $\frac{dE}{dX} = 0.26 + 3.57 \times 10^{-4} \cdot (E_\mu/GeV)$.

5.2 Light Propagation in Ice

The optical properties of the South Pole ice are of great importance to the analysis of data recorded in the IceCube detector array. To determine the trajectories and energies of muons and electromagnetic cascades by detecting the Cerenkov photons with arrays of deeply buried DOMs, the effects of scattering and absorption of light at wavelengths in the visible and near ultraviolet must be taken into account. Dust, sea salt, minerals, acid, soot, etc., all affect the scattering and absorption properties of the ice in their own way. A whole picture of the ice profile is indispensable in IceCube event simulation and reconstruction.

5.2.1 Scattering

Light is scattered in deep ice by microscopic centers, such as sub-millimeter-sized air bubbles and micron-sized dust grains [49]. An important assumption in photon multiple scattering is described by $\langle\cos\theta\rangle_n = \langle\cos\theta\rangle^n$ [50], where θ is the angle for a single scattering and n is the number of scatterings of the light field. A detailed numerical calculation conducted by IceCube collaboration based on Mie theory shows that $\langle\cos\theta\rangle$ is approximately 0.94 [51]. However, the value used in standard AHA ice model discussed in Section 5.3 is 0.80.

Let λ_s denote the average distance between scatters — the scattering mean free path. For strongly forward peaked anisotropic scatters($\cos\theta > 0$), the effective scattering length λ_e could be written as $\lambda_s \sum_{i=0}^{n} \langle\cos\theta\rangle^i$. For sufficiently large n, λ_e becomes $\lambda_s/(1 - \langle\cos\theta\rangle)$. It is intuitive to think of λ_e as the length traveled by the center of the photon cloud of the incident light propagating through a turbid medium. The benefit to use the scattering length is a reduction from two independent variables to one. The effective scattering coefficient b_e is defined as the reciprocal of λ_e. For light between 300 and 600 nm, scattering off dust has a wavelength dependence that can be described as

$$b_e \propto \lambda_e^{-\alpha}. \tag{5.3}$$

The exponent α in equation 5.3 depends strongly on the dust composition and is close to 1 [51]. The scattering probability function in IceCube detector simulation is the Henyey-Greenstein approximation [52],

$$p(\cos\theta) = \frac{1}{2}\frac{1 - \langle\cos\theta\rangle^2}{(1 + \langle\cos\theta\rangle^2 - 2\langle\cos\theta\rangle\cos\theta)^{3/2}}. \tag{5.4}$$

5.2.2 Absorption

The absorption length λ_a is the distance at which the survival probability drops to $1/e$, the reciprocal of which is the absorption coefficient

$$a = \frac{1}{\lambda_a}. \tag{5.5}$$

The absorption coefficient is related to the imaginary part of the index of refraction n by

$$a = \frac{4\pi Im(n)}{\lambda}, \tag{5.6}$$

where λ is the wavelength of the light source. The absorptivity of ice in the far-ultraviolet through infrared can be parameterized by a three component model [53]

$$a(\lambda) = A_U e^{-B_U\lambda} + C_{dust}\lambda^{-\kappa} + A_{IR}e^{-\lambda_0/\lambda}, \tag{5.7}$$

A fit to measurement of ice absorption in the far-ultraviolet (180-186 nm) yielded $A_U = 8.7 \times 10^{-39} m^{-1}$ and $B_U = 0.48 nm^{-1}$ [55]. Component 2 is the contribution of insoluble dust and dominates in the near UV and blue. Fitting measurements in 1974 [56] and 1994 [57] in the region of $300 \leq \lambda \leq 700$ nm led to a wavelength dependence of λ^{-2}, which is independent of depth. Component 3 is the intrinsic absorption of pure ice. The fitting measurement gives $A_{IR} = 81$ and

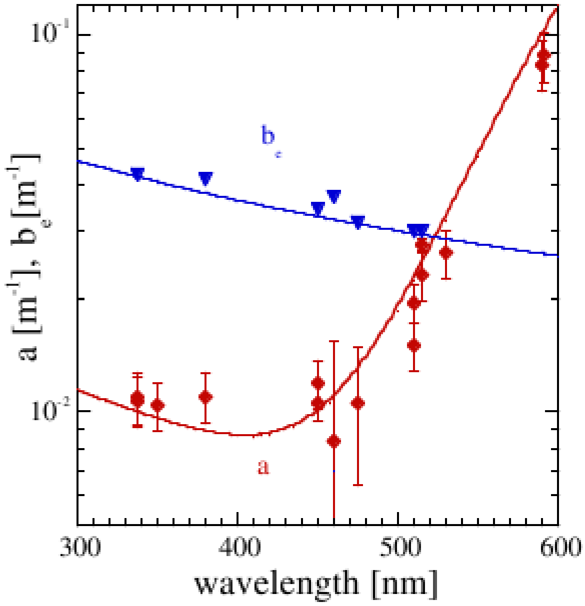


Figure 5.6: Scattering and absorption at average depth of 1.7 km, compared with the theory of He and Price (1998).

$\lambda_0 = 6.7\mu m$ at 600-800 nm [53]. Fig. 5.6 depicts the wavelength dependence of scattering and absorption coefficient [58]. The depth dependence of the two variables will be addressed in section 5.3. The combined effect of scattering and absorption over large distances can be described by the propagation length and coefficient [53]

$$c_p = \frac{1}{\lambda_p} = \sqrt{3ab_e}, \tag{5.8}$$

which is a parameter for a diffusive photon that transports from a point source [2].

5.3 Photonics and Ice Model

Light propagating through ice (even the clearest one) is affected by scattering and absorption, which can not be analytically calculated for distances between light sources and receivers comparable to the photon mean free path. Detailed simulation

---

[2] The commonly used attenuation length $\lambda_{att} = 1/(a + b)$, where $b = 1/\lambda_s$, describes the probability that a photon is removed from a beam by either scattering or absorption.

of photon propagation is required to provide information necessary for event [3] simulation and reconstruction. The problem is further complicated by the anisotropy of the light emitted in particle interactions and the heterogeneity of detector media. The PHOTONICS software package is a freely available software package developed to determine photon flux and time distribution throughout volume containing a light source through Monte Carlo simulation [59].

The PHOTONICS tables are usually binned in six dimensions: space (three dimensions) and time relative to the photon emission angle at the source, and photon incidence angle when it is recorded. Once a PHOTONICS table set has been generated for a class of light sources and detectors, the light yield and time distributions can be obtained from the table by specifying the geometrical relations between light sources and receiver units, the energy scaler, the type of sources, etc, without doing any real-time simulation on photon propagation. Roughly speaking, the PHOTONICS tables store the results of simulating photon propagation with various light sources at different angles and depths, while the PHOTONICS software provides the users or programs an interface to access those simulation results.

Knowing the ice property is crucial in determining the photon flux and time distribution. A simplified ice model [4] is used to generate each PHOTONICS table set. The ice model has been modified from time to time. The model that is used at the time of the draft of this thesis in IceCube event simulation is called additionally heterogeneous absorption (AHA) ice model as shown in Fig. 5.7. The AHA model was corrected for a systematic smearing of the dust layer structure introduced by the analysis methods used in AMANDA. The extrapolation of the optical properties to larger depths was redone with new ice core data on dust concentration to produce

---

[3] The most common IceCube events are muon bundles from cosmic rays, single muons from neutrino interaction in the ice and cascade events.

[4] Ice models in IceCube are parameterized scattering and absorption properties of ice dependent on the depth underground.

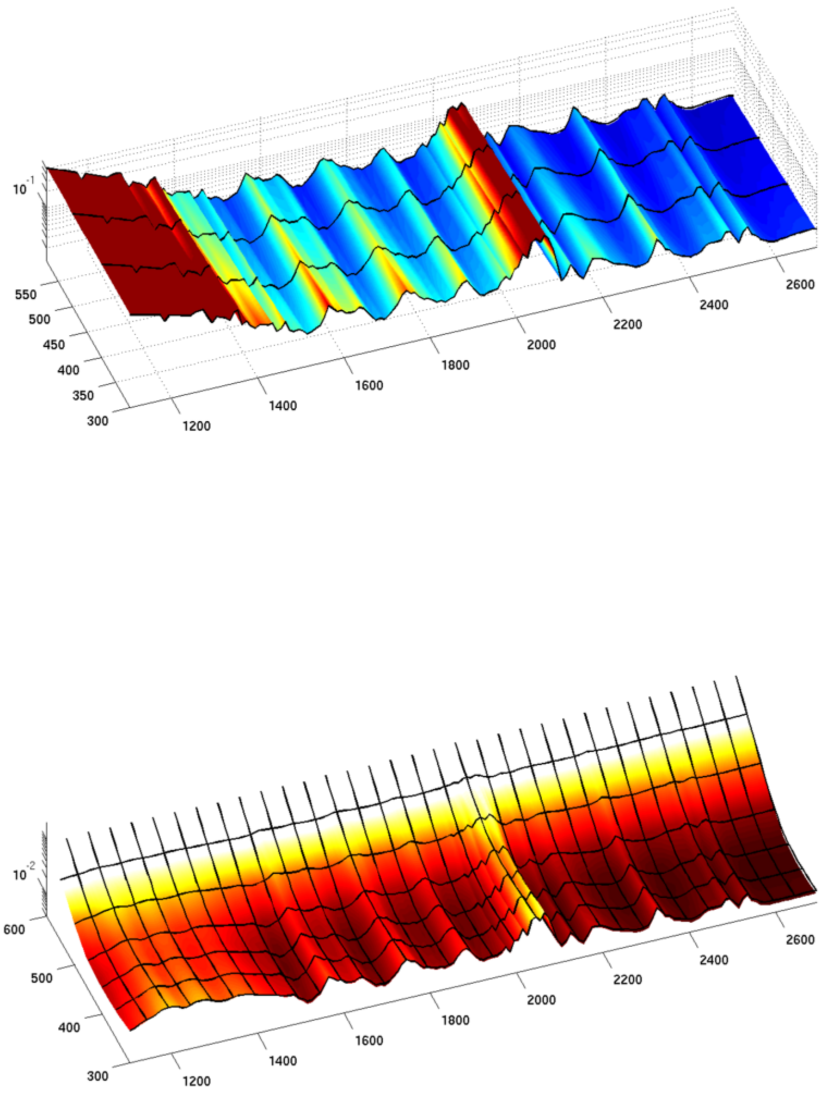


Figure 5.7: Depth and wavelength dependence of optical properties in AHA model taken from [60]. The upper panel is the scattering coefficient and the lower panel is the absorptivity, both are in unit m^{-1}. The horizontal axes are the depth in the ice in meters and the axes penetrating the paper is the wavelength of the light in nanometers.

cleaner ice below the big dust peak (peak D in Fig. 5.8).

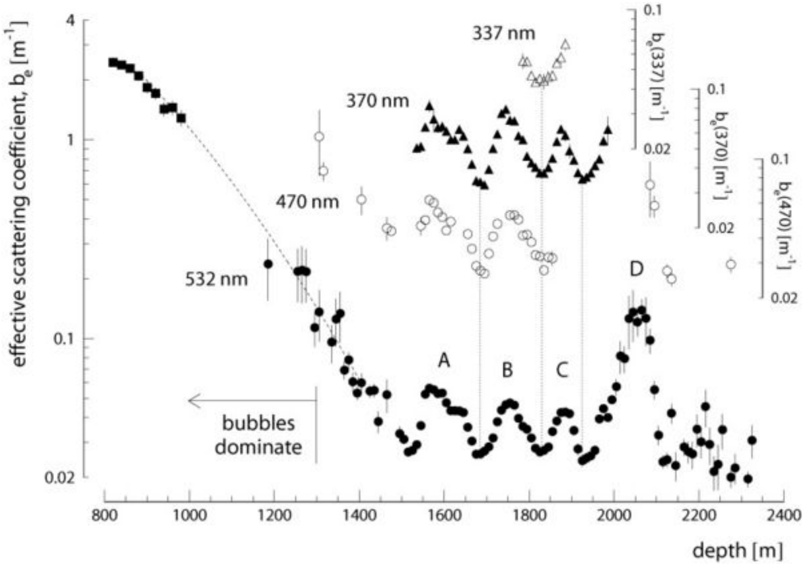


Figure 5.8: A dust peak is an increase in dust concentration over a certain time period which manifests itself as an increase in optical scattering and absorption over the corresponding depth range in glacial ice. In South Pole ice we have identified four dust peaks by comparing AMANDA scattering data with ice core data. They are commonly labeled A, B, C, and D. These peaks correspond to stadials during the last glacial period in the late Pleistocene. Figure is from [51].

5.4 DOM Response

There is a variety of simulated events produced in IceCube, such as single muons, electrons, neutrinos and muon bundles from CORSIKA showers for physics reseach and flasher and standard candle simulation for calibration. All these simulations are generated in the ICETRAY framework by applying associated source generators, such as CORSIKA and neutrino generator.

In the simulation of muon or muon bundle events, the energy and angular spectrum of the input muons (usually at surface, but could be underground) are taken care of by the CORSIKA program. The muons from their parent particles are then passed on to MMC program [17]. MMC propagates the muons fed to it and determines the interactions that happen between the muon and the medium within the specified volume on route, such as bremsstrahlung, electron pair production and so on. Once all the sources of light in effect are known the ICETRAY module HITMAKER is used to make hits out of these sources (muon tracks and cascades) by calling PHOTONICS-SERVICE to look up the dictionary of PHOTONICS tables.

Because of the way the muon interaction in the ice is modeled, there are only two fundamental table types of simulation in PHOTONICS, the shower and the muon tables [5]. While the muon tables give the information about the photon yields due to the Cerenkov light shed by the primary muon track [6] in a well-defined cone, the cascade tables describes the bright bursts ignited by the point-source-like stochastic energy loss along the muon track. A real high energy muon could be considered a superposition of many cascades (electro-magnetic showers and hadrons) on a bare muon track and a muon bundle consists of many such muons with a specific muon energy distribution.

5.4.1 Cascades

A cascade is a burst of light from a point source, which represents the light from stochastic energy loss of muons that happens in the detection medium in simulation of muons with PHOTONICS. The causes of a cascade could be either an electro-magnetic or hadronic shower in PHOTONICS. The photon density could be extremely high close to the source (several meters usually) but drops dramatically outwards. To examine the photon density distribution around a cascade, PHOTONICS is used to call the AHA tables with the input of the location and energy of the cascade. The cascade could be put in any place within the detector volume with any energy. In Fig. 5.9, the source location is selected to be at the IC40 InIce detector

---

[5] We ignore the flasher and standard candle here.

[6] The track could be either infinite or semi-infinite. The finite track could be derived by taking the difference between two semi-infinite tracks.

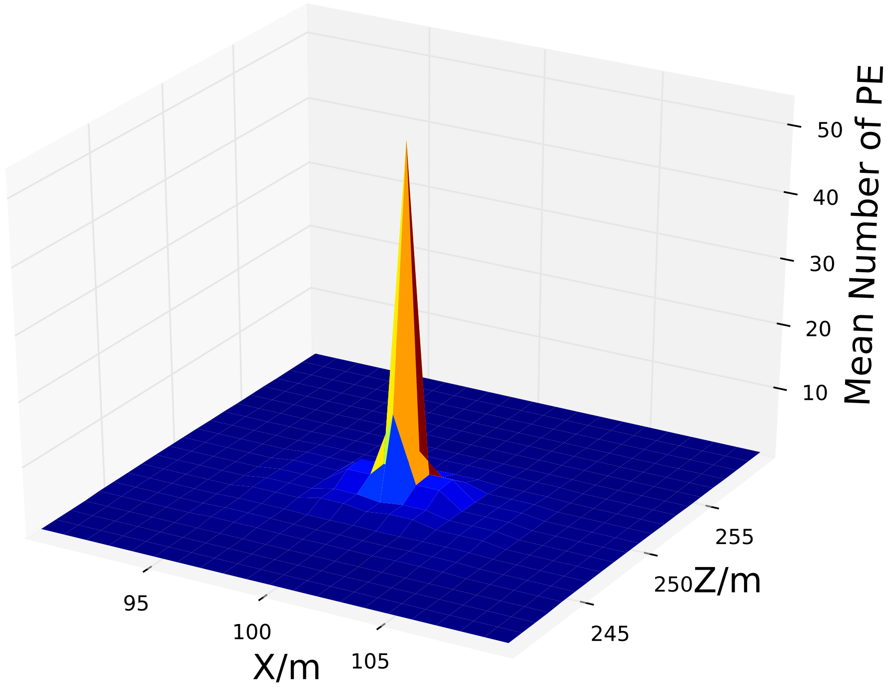


Figure 5.9: A 1 GeV cascade is thrown at IC40 InIce detector center (100, 250, 0) in IceCube coordinates (x, y, z). Mean photoelectron numbers are collected for virtual OMs around the source. Because of the isotropic assumption for any ice layer plane, only variation in x direction is taken for any depth selected. The OMs are $1m$ apart from each other and within $10m$ from the source in x and z direction respectively.

center. To check the OM responses from this cascade, OMs 10 meters away from the source are selected with $1m$ interval in Fig. 5.9. These OMs need not really exist, PHOTONICS treats all the OMs in the same way and incorporates the OM information. So once it is fed the OM location, it gives the photon number recorded due to the source assuming there is really an OM put at that point.

The cascade energy could be arbitrarily high since the effect of it is just like a scaler as seen in Fig. 5.10. The relationship between the mean photoelectron number and cascade energy thus could be written as $N_{pe} = b(\vec{x}_{SRC \to OM})E_{SRC}$, where the coefficient b depends on the position of the OM relative to the source.

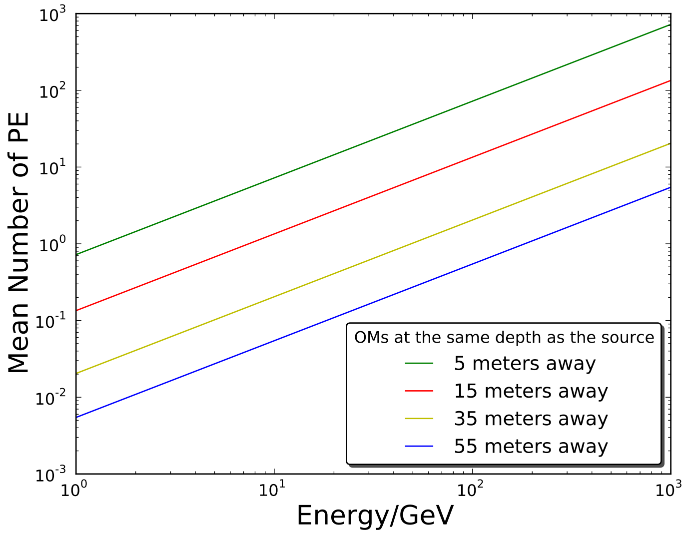


Figure 5.10: The mean number of photoelectron numbers is proportional to the energy of the cascade source. The scaler decreases for more distant OMs. Same DOMs are taken as in Fig. 5.11. The source is still at the detector center except its energy varies from $1GeV$ to $1TeV$.

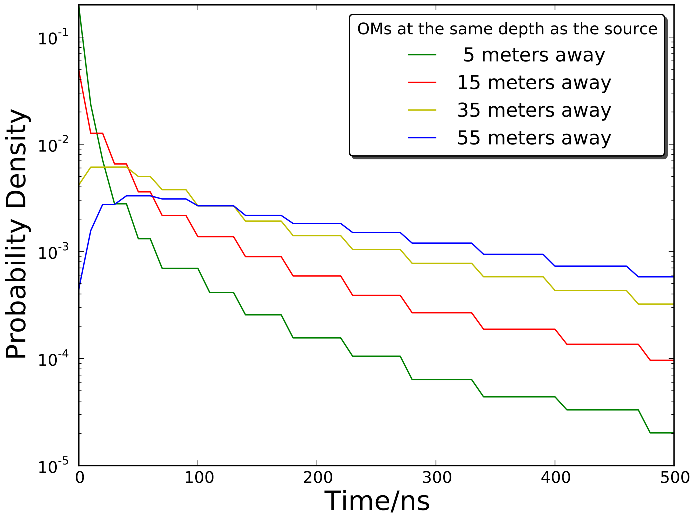


Figure 5.11: The time distribution of arriving photons, referenced to the first direct photon time. The same cascade is thrown as in Fig. 5.9. Four virtual OM timing samples are taken at the same depth as the source along x direction $(5, 15, 35, 55m)$.

The photons measured by the OM do not come all at once. Instead, they have a time distribution, which could be describe by Pandel probability design functions [73] as show in Fig. 5.11. It is easy to notice that the further the OM is from the source, the flatter the time distribution it has.

5.4.2 Bare Muons

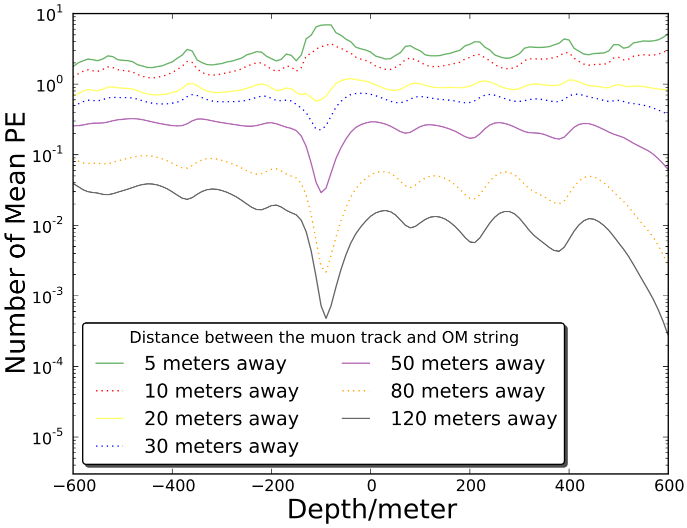


Figure 5.12: Mean photoelectron numbers seen by a string of virtual OMs parallel to the Cerenkov muon track with different distances. The muon starts from $600m$ in the detector coordinates with $1TeV$ energy. The vertical interval of OMs is $10m$.

In contrast to cascade tables, muon tables are for muons with only Cerenkov radiation due to ionization. The energy of the source (the muon energy) plays a much less important role in this case than in the cascade. As long as the OM sits somewhere far from the end of muon track relative to the closest approach to the track, the muon energy has little effect on its signal. The OM signals due to bare muons should reflect the ice property with contamination of smearing effect 5.12. [7]

---

[7] The smearing effect here is two fold. First, the photons are emitted toward

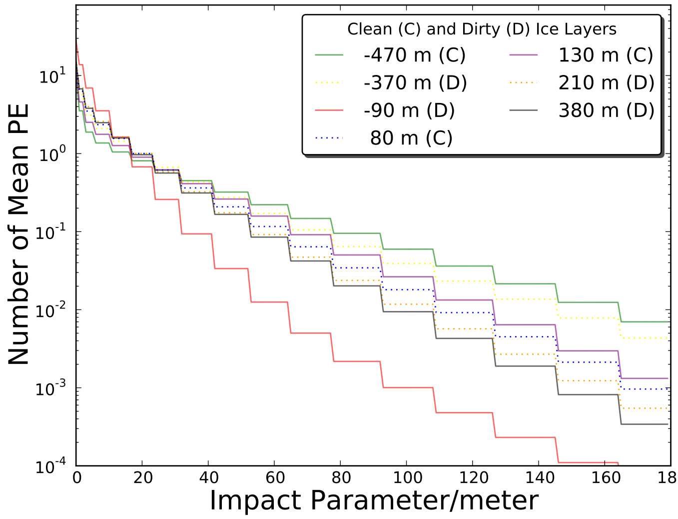


Figure 5.13: Mean PE lateral distribution for the same muon as in Fig. 5.12.

The lateral and time distribution of photon density are displayed in Fig. 5.13 and Fig. 5.14 respectively. The lateral distribution is stepper at a dirty layer than at a clean one.

5.4.3 Muon Bundles

A muon bundle is the group muons left by an extensive air shower under the surface of ice at South Pole. The number of muons in a bundle varies from several to tens of thousand depending on the energy and type of primary particle, the depth in the ice we look at, etc. For a certain type of primary particles, the muon multiplicity at any depth in the ice or at surface has a strong correlation with the energy of primary particle. However, the muon multiplicity is not measured directly by IceCube. The digital optical modules take samples of the photon density

---

42.° thus may penetrate two ice layers. Second, even if the direct photon track is within a single ice layer, scattered photons could go through more than one layers.

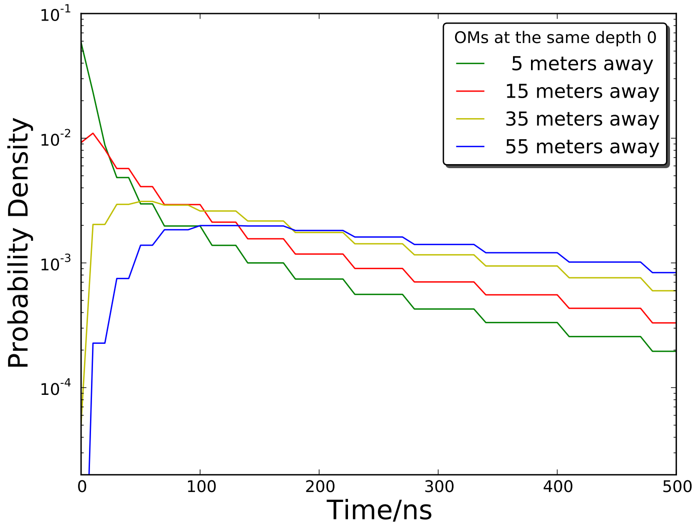


Figure 5.14: Time distribution for DOMs at depth 0 using the same muon as in Fig. 5.12.

surrounding the muons, which is related to the energy deposition of muons. Fig. 5.15 shows the profile of a typical muon bundle and the waveform registered in one of the DOMs of IceCube detector. Though both the constant and stochastic energy loss from the muons contribute to the DOM signals, a nearby cascade is most likely to give a strong boost to the DOM.

The total charge recorded by the IceCube detector is strongly correlated to the energy deposition of the muon bundles as shown in Fig. 5.16. The first 1000 files from simulation dataset 2546 [8] are used to probe the relationship between the muon bundle energy loss and detector response. Big vertical showers ($E_{prim} > 1PeV$ and $\theta < 30^\circ$) that land well inside the detector volume (InIceSiz < 0.4) are selected from those events, which amounts to 15601 in total. In Fig. 5.16, the total charge per

---

[8] Filtered Level2b simulated IC40+TWR CORSIKA-in-ice single muon, Polygonato model with weighted (dslope=-1) spectrum of Hoerandel, using AHA07v2 photon tables. Angular range of $0^\circ < \theta < 89.99^\circ$ and energy range of $600GeV < E_{prim} < 10^{11}GeV$.

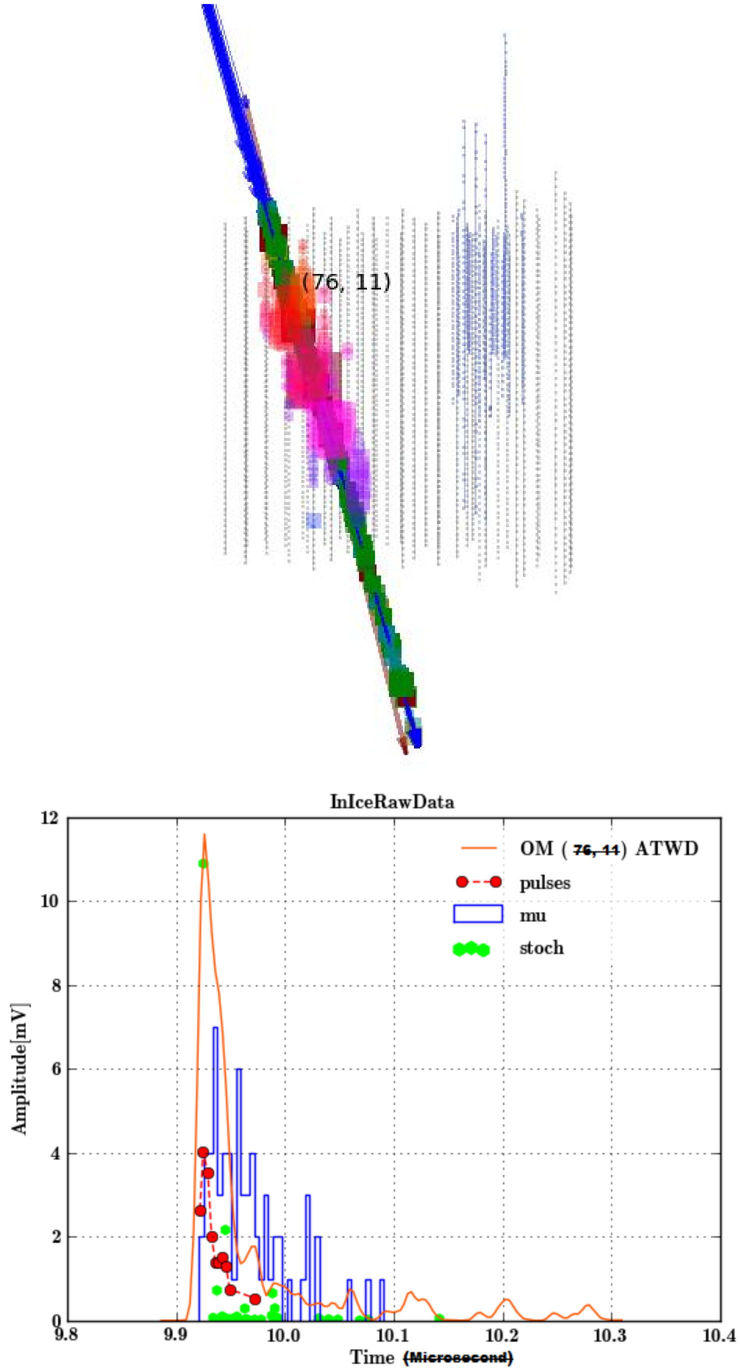


Figure 5.15: The profile of a muon bundle and the waveform of one triggered DOM. The event is from CORSIKA-Inice dataset 2546 with a primary energy about 3PeV. The green arrows in the profile represent the muon tracks while the red clouds are the charge amplitude. The lower graph shows the waveform from the DOM marked in the upper graph. The green dots represent the amplitude of the stochastic energy loss close to this DOM. The blue histogram registers entries of muons that pass the DOM in each time window. Red dots are feature extracted charges.

event in photoelectrons is plotted against total energy deposited by muon bundles, both in logarithmic scale. A linear regression illustrates the relationship between the charge and energy loss very well. With the R^2 equals 0.814, above 90% of the total charge in IceCube detector can be explained by muon energy deposition in the detector volume for these selected events. Believing the simulation is reliable, muon bundle energy loss, (and possibly the muon multiplicity) could be inferred from experimental data in the range where the linearity between detector response and energy deposition holds.

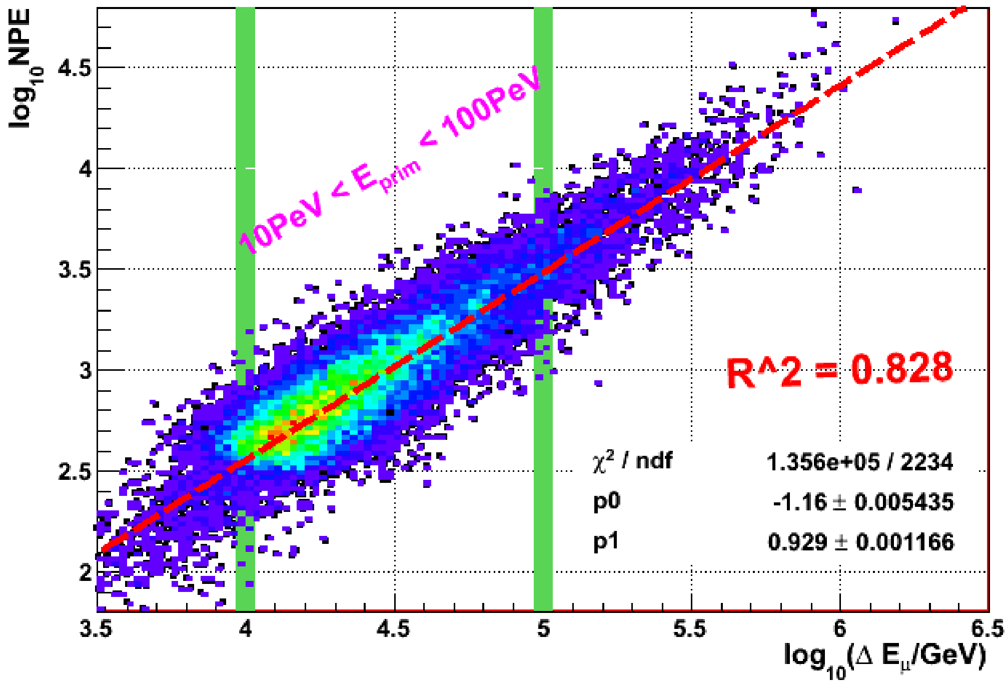


Figure 5.16: $\log_{10}$(Total Charge in PE per event) versus $\log_{10}$(Total energy loss by all muons in each event). The bundle energy loss is calculated by taking the difference between bundle energy at $\pm 800m$ with respect to the vertical center of InIce detector. The energy spectrum of the primary particle is flattened by 1 in $\log_{10}$ scale, which is equivalent to give each event a weight of E_{prim}.

Fig. 5.17 breaks down the muon bundle energy loss and detector response into different depths. The fluctuation of bundle energy loss is significant compared to the magnitude of energy loss at any selected depth, which causes great difficulty in energy estimation, let alone the asymmetry in fluctuation. The detector response is distorted by the ice property at varying depths, shown clearly in Fig. 5.17, thus correction for ice property is the key in the event reconstruction.

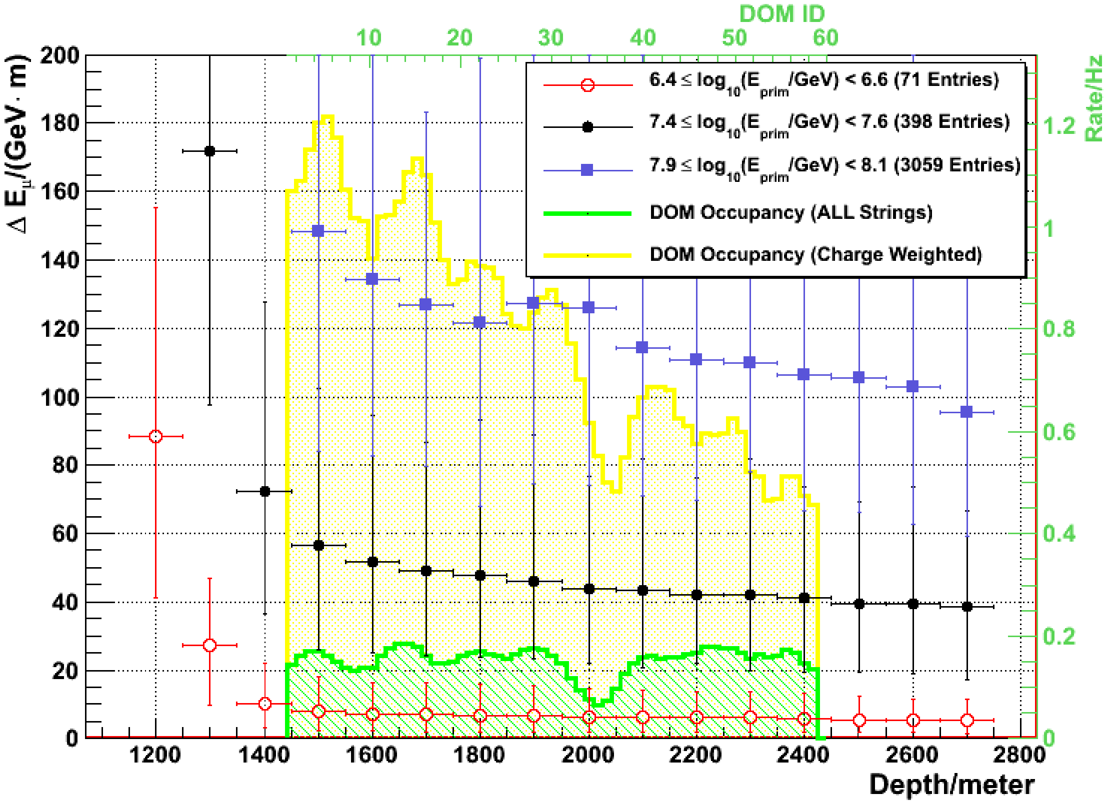


Figure 5.17: DOM Occupancy integrated over all strings; Bundle Energy Loss Rate. The bundle energy is calculated every 100m from +800 to -800 with respect to the vertical detector center. Then the difference between each pair of neighboring energy is divided by the distance traveled by the muons between the two depth. The error bars represent the edges of 90% of total events in each depth bin. Note the left Y labels are associated with the bundle energy loss (profiles) and the right labels are rates for occupancy (histograms).

5.5 PPC versus Photonics

The PPC (photon propagation code [61]) is a project developed by Dmitry Chirkin for, but not limited to, flasher simulation. The PPC propagates photons through heterogeneous ice described by the six-parameter model [62]. The development of PPC is supposed to test to what extent PHOTONICS could be corrected with the remaining errors due to our understanding in ice. The advantage of PPC over Photonics is avoiding binning, parameterization and interpolation problems in Photonics. In report [62], Chirkin used IceCube runs 111738-111744 data [9] to fit the simulated PPC flasher events and obtained a new ice property that is shown in Fig. 5.18.

Paulo Desiati compared the single muon bundle simulation generated by using PHOTONICS and PPC (CORSIKA dataset 1540 for PHOTONICS and CORSIKA dataset 2284 for PPC). Some of those compared parameters are shown in Fig. 5.19.

5.6 OM Acceptance for Photonics tables

We have discussed the OM response to cascade and muon sources in IceCube simulation. It should be noted that the OM is not a hundred percent efficient. Not all the photons that reach the DOM are recorded. Various factors effect the OM efficiency, such as PMT sensitivity to wavelength, inclination of sources, gel and glass transmissivity, hole ice effect, etc. Understanding the efficiency and acceptance of IceCube DOMs is of great importance for event construction and data analysis. In this section we will address the wavelength and angle dependence of the OM acceptance. We will also discuss the effect of hole ice on DOM response.

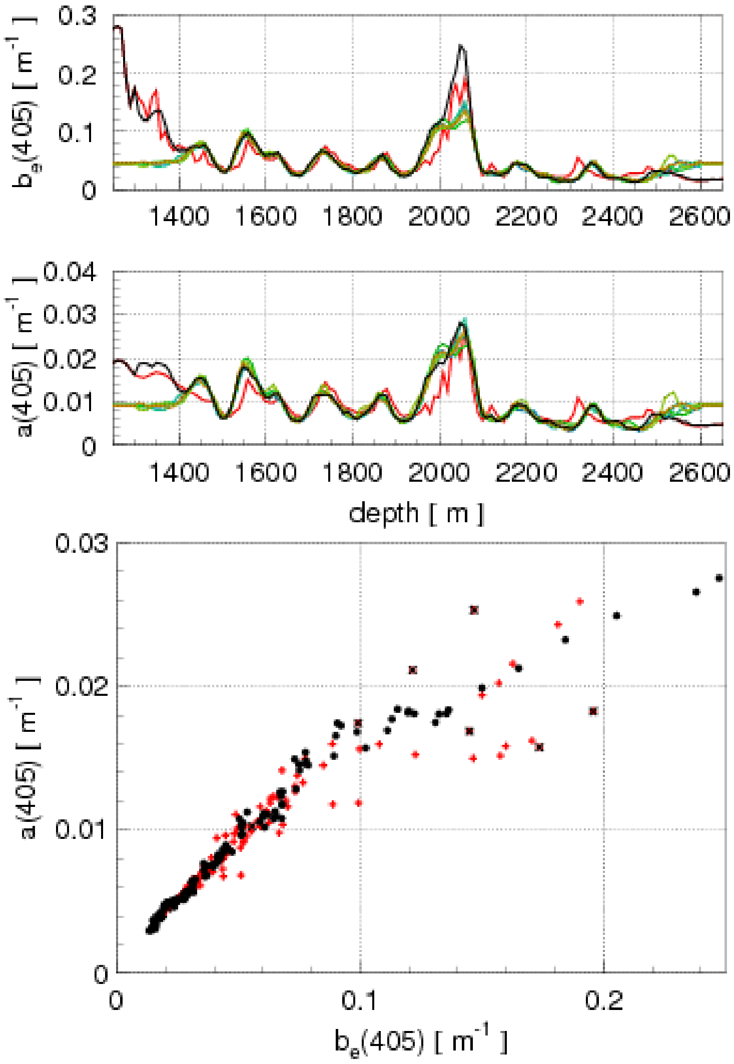


Figure 5.18: From top to bottom are scattering coefficient versus depth, absorption coefficient versus depth and a versus b_e taken from [63]. The red color represents the original AHA ice model; the black represents the fitted data for $p_y = 1.9$, starting from AHA model; five shades of green represent fitted data for $p_y = 1.7, 1.8, 1.9, 2.0, 2.1$, all starting from bulk ice. Where p_y is the photon yield factor, which is the number of unit bunches that correspond to a given number of photons. E.g., 4.5×10^{10} photons emitted by a flasher board correspond to a photon yield factor of $p_y = 2.034$ [62].

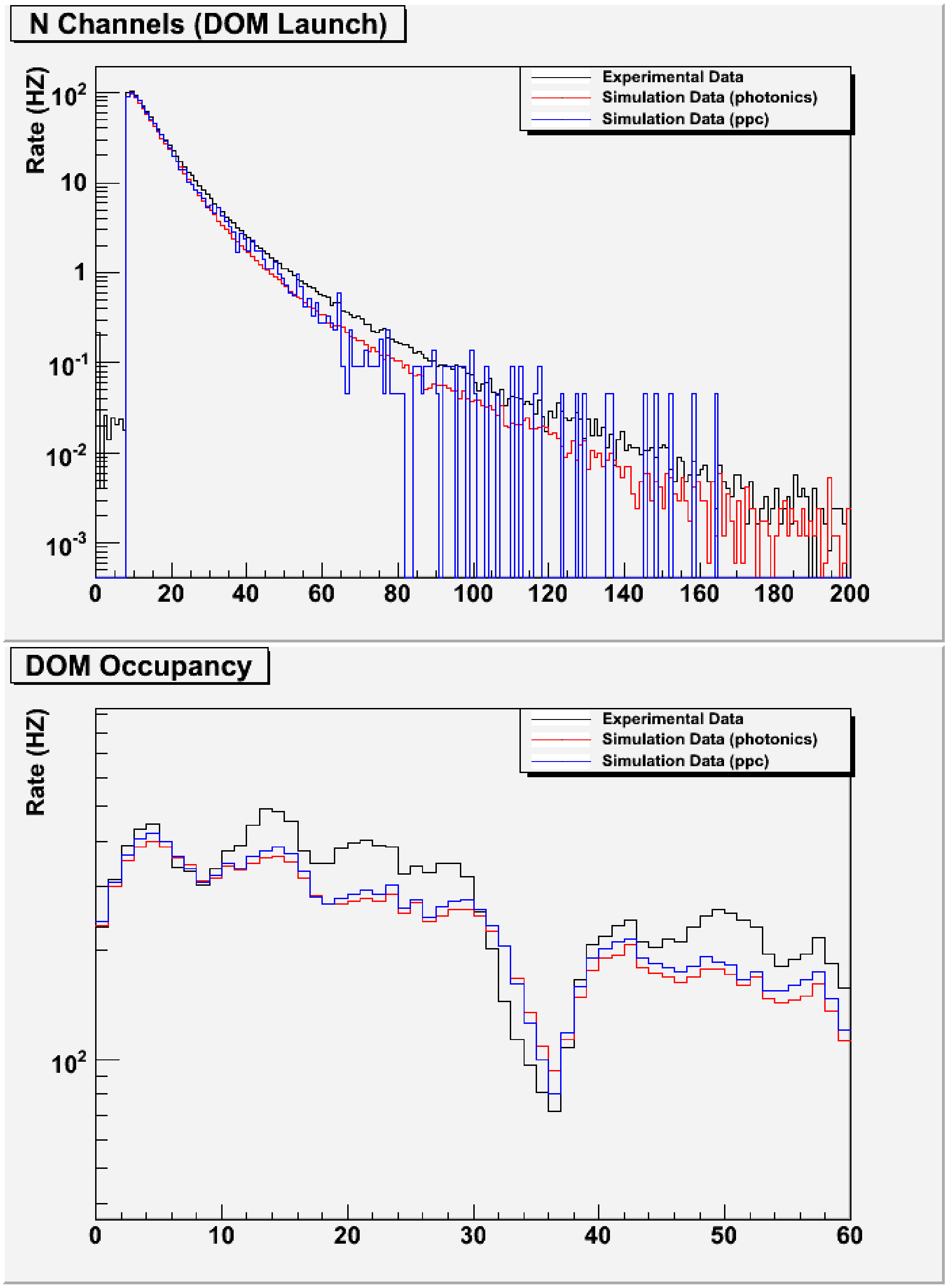


Figure 5.19: Comparison of an IC40 experiment run (black) with PHOTONICS single muon bundle (red) and PPC single muon bundle (blue) taken from [64]. The comparison is done at SMT trigger level. The top figure is the distribution of number of triggered OMs; the bottom is the OM occupancy. The overall event rate in simulation is less than 1/2 of the experimental rate, which causes the discrepancy between the experimental and simulated DOM occupancies.

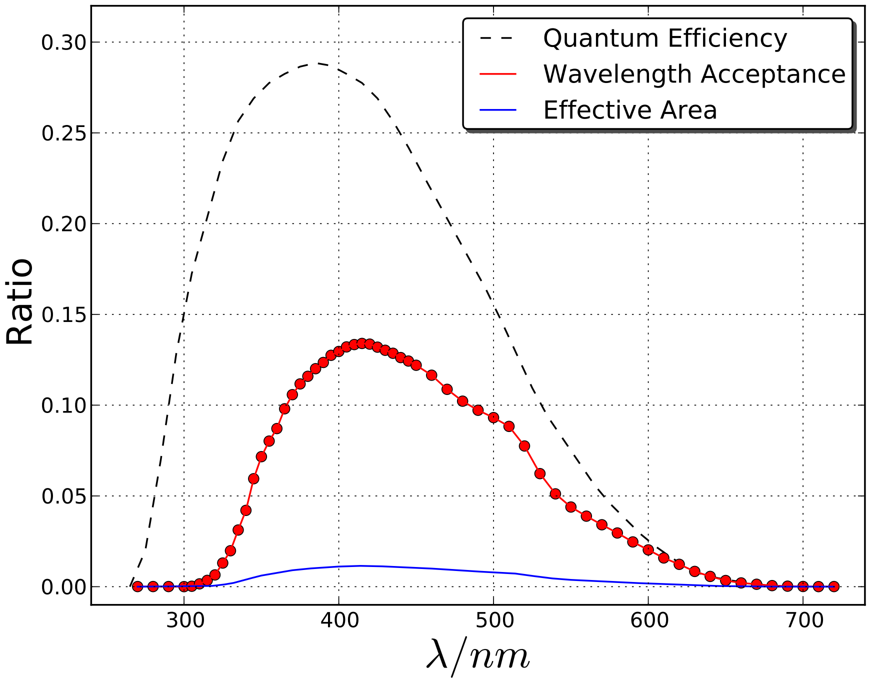


Figure 5.20: The dependence of OM efficiency, acceptance and effective area on wavelength, extracted from file *efficiency.h* in PHOTONICS. The acceptance and effective area are the photo-electron acceptances of the IceCube PMT after through the glass+gel+PMT photo-cathode as a function of wavelength which corresponds to 0 PE threshold and 0 degree injection angle.

5.6.1 Wavelength Acceptance

Quantum efficiency and wavelength acceptance are similar except for the difference in units and PE threshold. The confusion is clarified in IceCube by the

---

[9] IC40 string 63 OM 1-60, 250 flasher events for each of the flashing OMs, all 6 horizontal LEDs are at their maximum brightness and width

following definitions:

$$\begin{aligned}\text{efficiency} &= \frac{\text{number of generated PEs}}{\text{number of photons injected to photo cathode}},\\ \text{acceptance} &= \frac{\text{number of generated PEs}}{\text{number of photons injected to OM section (13-inch circle)}},\\ \text{effective area} &= \frac{\text{number of generated PEs}}{\text{number of photons injected to } 1m^2},\end{aligned}$$

where the “number of generated PEs” is the number of seed electrons generated by photo-cathode.

5.6.2 Angle Acceptance

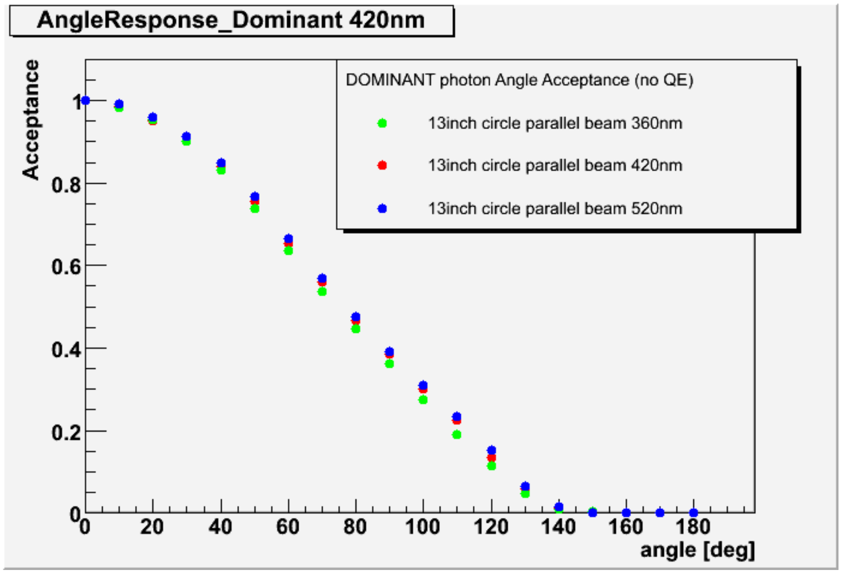


Figure 5.21: The angle acceptance of IceCube DOMs, taken from [65] which shows photon acceptance of OM for three different wavelength, calculated by DOMINANT program (based on GEANT4). Acceptance curve of lowest wavelength drops slightly at higher angle, since the photon absorption inside the gel and glass is greater for photons with shorter wavelength so that their path length increases more as angle increases in these medium. The 420nm curve is the standard PHOTONICS because it is the wavelength with the highest quantum efficiency.

Although the OM shell is a sphere, the PMT in it is not. The angular sensitivity is thus inevitably involved in the OM acceptance curve. If the highest PMT response is achieved at the 0 °, in which case the photon beam comes in perpendicularly to the PMT, at larger the angle, fewer photoelectrons are generated till 180 ° when the photo beam points to the back of PMT. In Fig. 5.21, the angle acceptance is normalized to that at 0 ° declining all the way to 180 °. Also evident from from Fig. 5.21 is that the angular acceptance depends little on the wavelength, which minimizes the error of PHOTONICS. The DOMINANT angle acceptance could be fitted by a six order polynomial function as following

$$\begin{aligned} \text{acceptance} &= 1.0 - 3.59 \times 10^{-3}\theta + 5.11 \times 10^{-5}\theta^2 - 4.25 \times 10^{-6}\theta^3 \\ &\quad +5.56 \times 10^{-8}\theta^4 - 2.73 \times 10^{-10}\theta^5 + 4.76 \times 10^{-13}\theta^6, \end{aligned}$$

where θ is the angle in degree. The fitting results are shown in Fig. 5.22.

The photo-cathode effect should also play role in the acceptance shape, as shown in Fig. 5.22, since photons that reach the edge of the PMT would produce fewer photoelectrons than those hit the center [10]. The overall acceptance becomes lower and not so smooth after introducing the photo-cathode convolution.

5.6.3 The Effect of Hole Ice

The hole ice, against the bulk ice (the pristine glacial ice), is the refrozen column of ice in which the DOM strings are embedded. The glacial ice at South Pole is drilled with hot water and OM strings are then deployed into the holes. The water in the holes refreezes quickly but is different from the bulk ice, which results in a stronger scattering effect near the OMs. The characteristic of this effect is an increasing probability of down-going light [11] to be scattered into the PMT area and recorded. This is light that would otherwise have passed by the OM. Similarly, the

---

[10] This is described by collection efficiency.

[11] This is the light coming from the other side of the PMT.

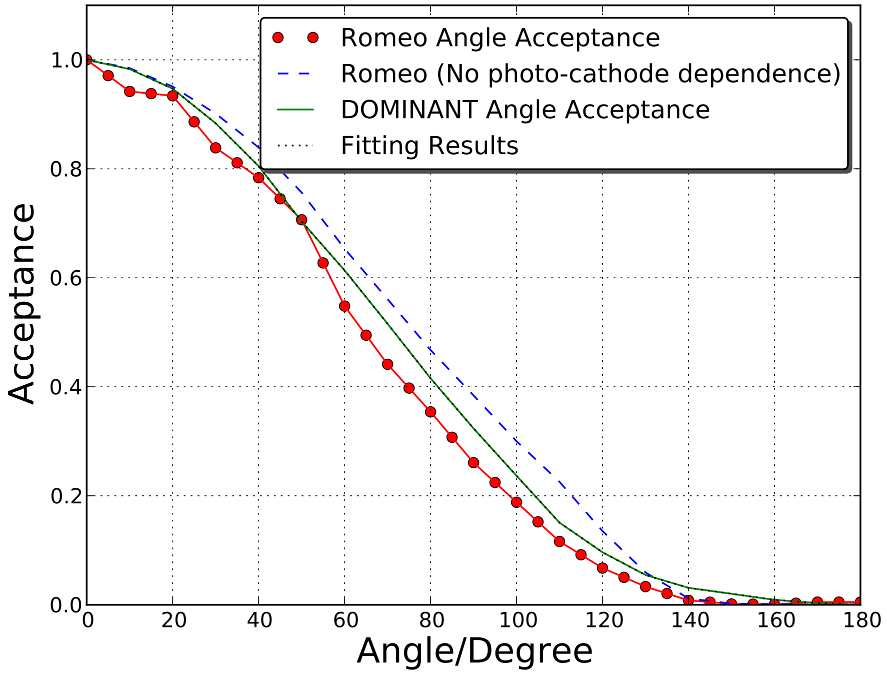


Figure 5.22: The angle acceptance of IceCube OMs from DOMINANT, ROMEO and ROMEA without photo-cathode sensitivity dependence on location, extracted from *efficiency.* in PHOTONICS. The photo-electron acceptances of the IceCube PMT after through the glass+gel+PMT photocathode corresponding to 0 PE threshold are defined by: acceptance = $\frac{\text{number of PEs accepted by PMT}}{\text{number of photons injected to OM}}$ /degree.

probability of up going light escaping the OM acceptance is also increased. The hole ice effect in PHOTONICS simulation is quantified by applying a correction to the DOM's angle acceptance measured in the laboratory [67], which is described in Fig. 5.23.

The conclusion of the hole ice effect from several experimental tests is mixed. In work of Michelangelo DAgostino [68], the flasher experimental data is consistent with a simulation featuring a high scattering hole ice effect [12]. However, Lisa Gerhardt in her recent work compared the time residual of DOMs surrounded by

---

[12] Michelangelo found that the OM observed fewer photons from tilted LED beams

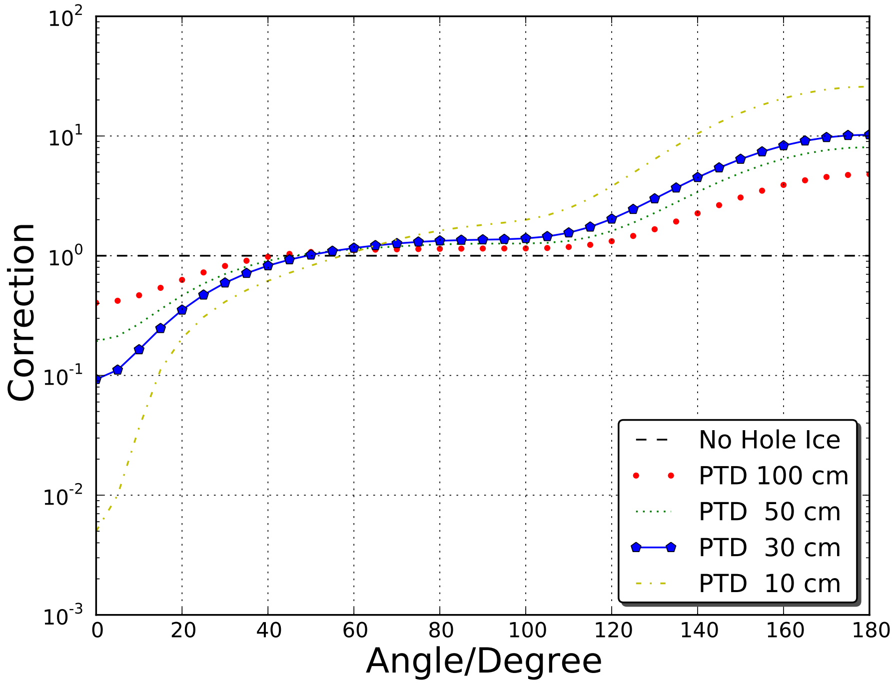


Figure 5.23: Correction to angle acceptance of the OM due to the effect of hole ice for different assumption of the scattering length of the bubbles in hole ice. In AHAv2 model which is currently used for muon simulation, the hole ice model 2 is used.

various amount of hole ice, which led to a conclusion that the hole ice had negligible effect on the arrival time or number of hits observed in the OM [69].

The combined effects of angle acceptance and hole ice are shown in Fig. 5.24. The hole effect dominates at low angle and becomes negligible at large angle.

---

while observing more from the horizontal ones. Also he discovered an azimuthally asymmetry in the hole ice scattering.

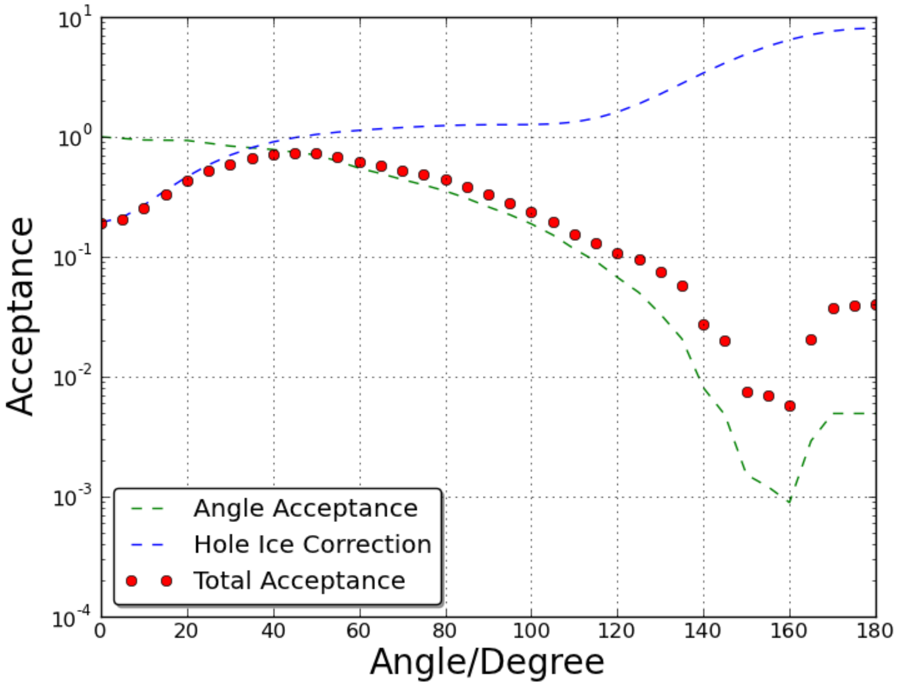


Figure 5.24: Combined effects of angle and hole ice used in simulation. Data extracted from *efficiency.h* in PHOTONICS with 2007 DOMs and hole ice model 2 (50 cm).

Chapter 6

AIR SHOWER RECONSTRUCTION IN ICECUBE

This chapter discusses how various reconstructions of air showers are done with IceCube. Reconstruction is a process of inferring the source (air shower) information based on the detector response. The reconstructed shower parameters in IceCube include but not are limited to the shower/muon bundle vertex, the track direction and the size or energy indicators. The reconstruction algorithms are implemented within the IceTray framework, and the results are pushed into physics frames as particles. The reconstructions are generally categorized into two groups: IceTop and InIce reconstructions, based on the data used in the process. IceTop reconstruction uses charge and timing information from the IceTop stations while InIce reconstruction uses data from InIce DOMs. Both reconstructions are used in this primary composition study.

6.1 Reconstruction Using IceTop Data

The analysis in this thesis uses IC59 IceTop detector array that consists of 59 IceTop stations. Each station contains two tanks which are about 10 meters from each other. Each tank has one high gain and one low gain DOM in it. The DOMs register the waveforms upon being hit by the photons in an air shower event. The calibrated waveforms are the input to the IceTop wave processor, which converts the waveform series to pulse series. The event builder selects the pulses that belong to a single event and converts them from photoelectrons to VEMs. The charges and times are fit with a lateral distribution function to get the direction, position and size of the air shower.

6.1.1 Wave Processor

The IceTop wave processor is an IceTray module that translates the IceTop data from waveforms to pulses. Simply put, the waveform series are ATWD voltages at different times while the pulse series are the number of photoelectrons at different times. This is the first step to convert the electronic response to physics values. For a HLC waveform, the processor integrates all bins to get the total charge and find the maximum amplitude of the waveform. Dividing the total charge by the SPE peak charge, we obtain the number of photoelectrons. The width of the pulse is the time between the leading and trailing edge of the wave form. The definition of leading edge is the time bin which the 10% to 90% slope of the first peak intersects with the baseline. The trailing edge is the last bin that is greater than or equal to 2% of the wave maximum. The start time of the pulse is the leading edge less the PMT transit time. For an SLC hit, there is no waveform so the charge is obtained directly from the charge stamp.

6.1.2 IceTop Event Builder

The IceTop event builder takes the pulse series generated by the wave processor and combines hits from individual high gain and low gain DOMs in each tank to form a tank signal. Furthermore the charge is converted from photoelectrons to vertical equivalent muons (VEM). Afterwards a station signal is created from the individual tank signals and finally events are created from the station signals by checking their casualty. The tank signal builder loops over the wave processor pulse series and figures out the tanks that have hits. It keeps the high gain DOM signal as the tank signal if the high gain is not saturated. If the high gain pulse is saturated, it is replaced by the best matched low gain pulse. The station signal builder tests if

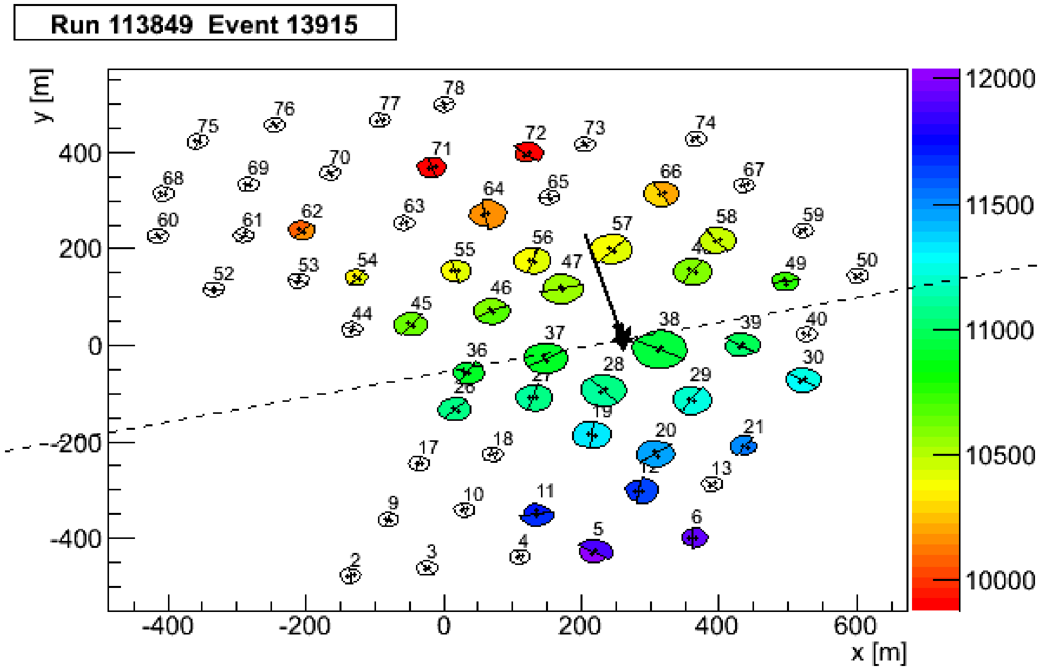


Figure 6.1: IceTop event builder tank signals made by using *topeventbrowser* for event 13915 from run 113849. The color represents the times of pulses and the size represents the magnitude of charges. Reconstruction results are shown as the arrow.

a pulse has a counterpart in the neighboring tank in the same station that is close enough in time. Any isolated pulse is discarded. The station signals is calculated as the average of the two tank signals. The event builder sorts the stations by ascending time. It figures out whether the time difference between two successive pulses from two stations is consistent with geometry of a shower front. Thus, the IceTop data is grouped into different events. Generally, only the first event is the physics event which is followed by some after pulses. A typical IceTop event is shown in Fig. 6.1.

6.1.3 IceTop Tank Variation

Five special experimental runs (run 111787-111791) in 2009 are used to study the IceTop variations. In the five runs, both DOMs in the tank were set to high gain

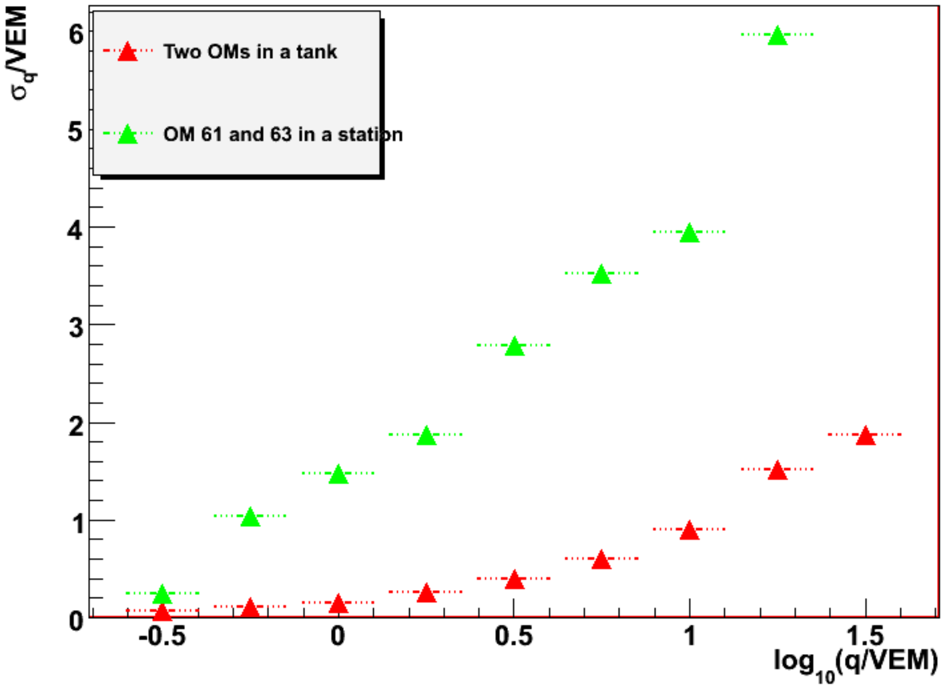


Figure 6.2: Standard deviation of charge versus $\log_{10} q$, where q is the average of the charges of the two DOMs.

that enables us to explore the intrinsic tank fluctuation. Let q_1 and q_2 denote the charges in VEM of two high gain DOMs (61 and 63). We find $\log_{10} q_1$ and $\log_{10} q_2$ are normally distributed with a standard deviation of σ. So $\log_{10} q_1 - \log_{10} q_2$ is normally distributed with a mean of 0 and a standard deviation of $\sqrt{2}\sigma$. The standard deviation of q_1 and q_2 is thus

$$\sigma_q = \sqrt{10^{2\log_{10} q + \sigma^2 \log(10)}(10^{\sigma^2 \log(10)} - 1)}. \tag{6.1}$$

Applying this method to the high gain pairs in 1)two DOMs in the same tank and 2)DOM 61 and 63 in the same station, we get the fluctuation of the individual tanks and stations (Fig. 6.2). Similar studies are done to obtain the time fluctuations as shown in Fig. 6.3 and Fig. 6.4. The study shows that the charge and time fluctuations within a tank are small compared to the intrinsic fluctuations in the shower front as measured by differences between two tanks at the same station.

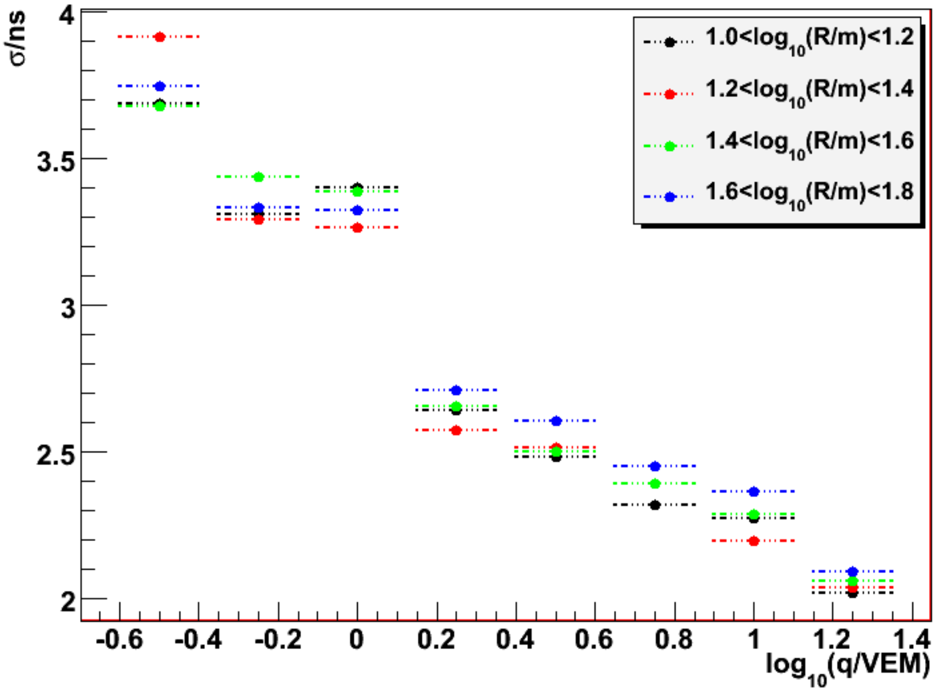


Figure 6.3: Standard deviation of time versus $\log_{10} q$ for the two DOMs in a tank, where q is the average of the charges of the two DOMs.

The charge variation model used in *toprec* is close to the measurement, as shown in Fig. 6.5.

6.1.4 IceTop Reconstruction

The IceTop reconstruction is a process of fitting the event builder pulses described above to extract the basic quantities of the detected air showers, such as shower direction (θ, ϕ), shower core at the surface (x, y), time T_0, snow correction parameter, shower size and age related parameters. Since the likelihood fit requires

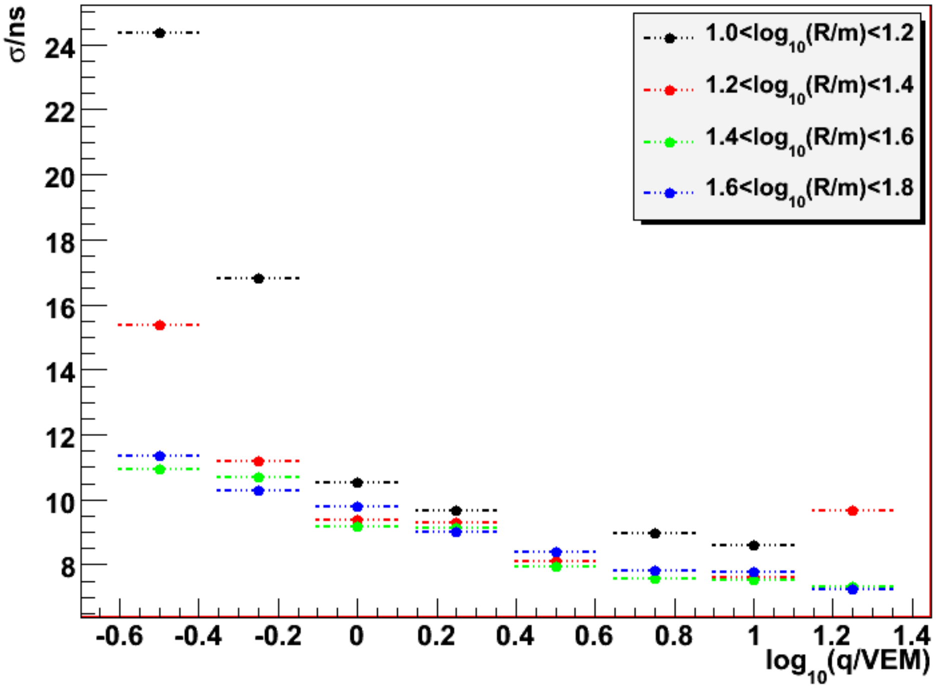


Figure 6.4: Standard deviation of time versus $\log_{10} q$ for the two HG DOMs in a station, where q is the average of the charges of the two DOMs.

at least one data point more than the degrees of freedom to work, the IceTop reconstruction is applied to IceTop events that can trigger at least five IceTop stations, or ten tanks equivalently.

6.1.5 Initial Guess Reconstructions

The initial guess reconstruction serves as the seed to the likelihood fit. The guess of the shower core location and shower direction are performed by two IceTray modules I3TopRecoCore and I3TopRecoPlane respectively. The core guess calculates the center of gravity (COG) of the charges:

$$\vec{x}_{COG} = \frac{\sum_i \sqrt{q_i}\vec{x}_i}{\sum_i \sqrt{q_i}}, \qquad (6.2)$$

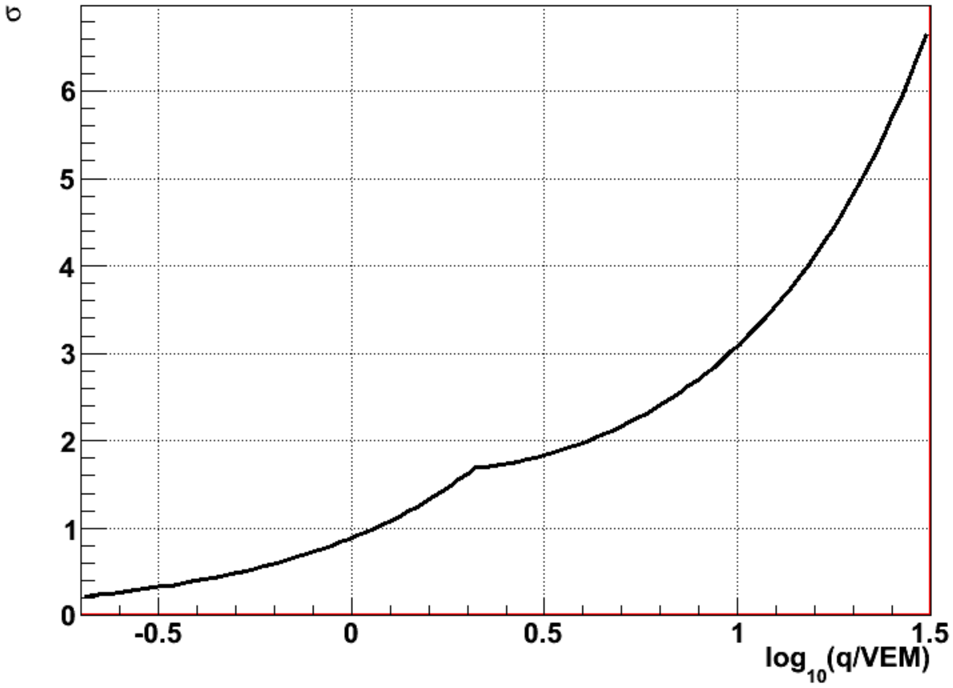


Figure 6.5: Charge fluctuation model used in *toprec*, compared with Fig. 6.2.

where q_i is the charge of the ith pulse in this event. Note that one tank or DOM could have more than one pulse, so the sum in equation 6.2 is for all the pulses from all the tanks. The square root of the signal is used instead of the signal because the signals decrease rapidly at larger distances from the shower axis.

The direction guess assumes a plane shower front and uses the core location obtained from the core guess. The direction of the shower is $\vec{n} = (n_x, n_y, -\sqrt{1-n_x^2-n_y^2})$ and $|\vec{n}| = 1$. The χ^2 function of the arrival times of the signals is

$$\begin{aligned} \chi^2(n_x, n_y) &= \sum_i w_i (t_i^m - t_i^{plane})^2 && (6.3) \\ &= \sum_i \frac{1}{\sigma_i^2} (t_i^m - \frac{n_x x_i + n_y y_i}{c})^2, && (6.4) \end{aligned}$$

where t_i^m are the measured signal times, x_i and y_i are the signal positions. t_i^m

and $(x_i,\ y_i)$ are relative to the center of gravity time and location. The standard deviation of arriving time σ_i is 5 ns. Solving equations $\frac{\partial\chi^2}{\partial n_x} = 0$, $\frac{\partial\chi^2}{\partial n_y} = 0$ gives us

$$\begin{aligned} n_x &= c \cdot \frac{\sum_i t_i^m x_i \sum_i y_i^2 - \sum_i t_i^m y_i \sum_i x_i y_i}{\sum_i x_i^2 \sum_i y_i^2 - (\sum_i x_i y_i)^2}, && (6.5) \\ n_y &= c \cdot \frac{\sum_i t_i^m y_i \sum_i x_i^2 - \sum_i t_i^m x_i \sum_i x_i y_i}{\sum_i x_i^2 \sum_i y_i^2 - (\sum_i x_i y_i)^2}. && (6.6) \\ & && (6.7) \end{aligned}$$

The direction $(\theta,\ \phi)$ is

$$\begin{aligned} \theta &= \arccos(-\sqrt{1 - n_x^2 - n_y^2}), && \theta \in [0, \pi), && (6.8) \\ \phi &= \arctan(\frac{n_y}{n_x}), && \phi \in [-\pi, \pi). && (6.9) \end{aligned}$$

The wrong assumption that all tanks are at the same height is corrected by a second iteration in which the times of all pulses are incremented by $\Delta z/c \cdot \cos\theta$, where Δz is the relative height of the tank of the pulse to the mean altitude of the triggered detector.

6.1.6 Likelihood Fit

The fit of the IceTop data is performed by I3TopLateralFit module that combines the fits for hit, no-hit and time signals. The general form of the log likelihood function is

$$\log \mathcal{L} = \log \mathcal{L}_{hit} + \log \mathcal{L}_{nohit} + \log \mathcal{L}_{time}. \tag{6.10}$$

The charge lateral distribution function is modeled by a double logarithmic parabola (DLP) function and defined as [70]

$$S(r) = S_{125} \left(\frac{r}{125m}\right)^{-\beta - \kappa \log(r/125m)}, \tag{6.11}$$

where r is the distance to shower core, perpendicular to shower axis, S_{125} is referred to as the shower size, β is the slope of DLP function at 125 meters perpendicular to the shower axis from the shower core. The curvature parameter κ is assumed constant at 0.303 in the fitting process. Assuming the measured charges are log-normally distributed around the expectation, the hit likelihood function can be written as

$$\mathcal{L}_{hit} = \prod_{i}^{NHits} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{\left[(\log(S_i^m) - \log(S_i^{fit}(r_i)) \right]^2}{2\sigma^2(s_i^{fit}(r_i))} \right), \tag{6.12}$$

where r_i and S_i^m are the measured distance and charge of the ith pulse while S_i^{fit} is the expected value for S_i^m from the DLP function. The charge fluctuation is taken from Fabian's diploma thesis [71]:

$$\log\sigma(S) = \begin{cases} -0.5519 - 0.078\log(S) & \log(S) < 0.340 \\ -0.373 - 0.685\log(S) + 0.158\log^2(S) & 0.340 \leq (S) < 2.077 \\ -1.114\text{const.} & 2.077 \leq \log(S). \end{cases}$$

The above function is plotted in Fig.6.5.

Omitting a constant term, the logarithmic likelihood function is

$$\log\mathcal{L}_{hit} = -\sum_i \frac{\left[(\log(S_i^m) - \log(S_i^{fit}(r_i)) \right]^2}{2\log\sigma(S_i^{fit}(r))} - \sum_i \log\sigma(S_i^{fit}(r)). \tag{6.13}$$

The silent tanks also contribute to shower reconstruction. Since any tank is totally independent with of other tanks and the station is considered silent if any one of the two tanks is silent, the probability of getting a silent stations is $1 - P^2_{tankhit}(r)$ and the likelihood that a tank can be hit is

$$\begin{aligned} P_{tankhit} &= 1 - \frac{1}{\sqrt{2\pi}\sigma_0} \int_{-\infty}^{S_{thr}} \exp\left(-\frac{\left(\log S^m - \log S^{fit}\right)^2}{2(\sigma_0)^2} \right) \mathbf{d}\log S^m & (6.14) \\ &= \frac{1}{2}\left[1 - erf\left(\frac{\log S^{thr} - \log S^{fit}}{\sqrt{2}\sigma_0} \right) \right], & (6.15) \end{aligned}$$

where the charge threshold 0.3 VEM and σ_0 is the charge fluctuation extrapolated to zero.

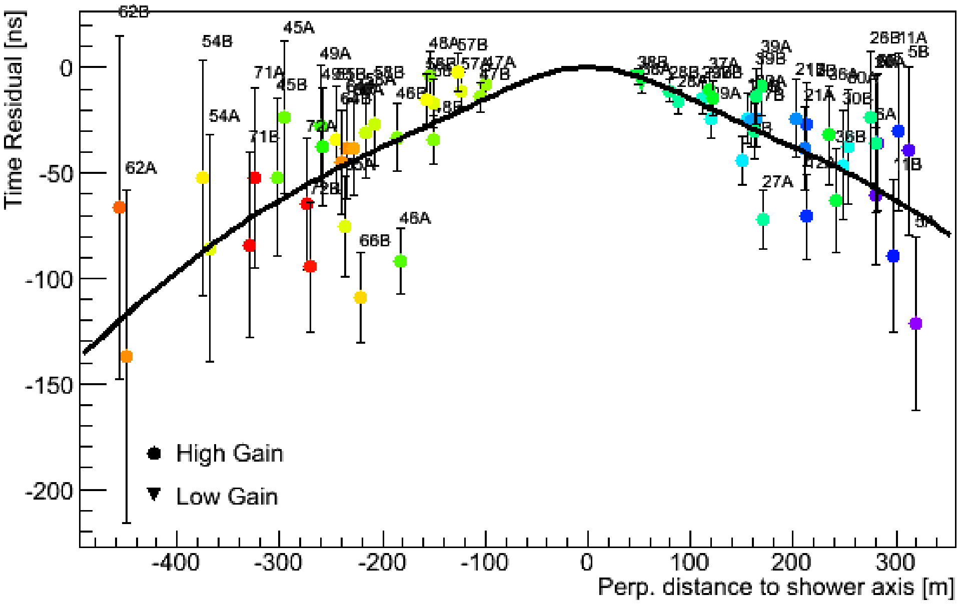


Figure 6.6: Time residual of tank signals for the same event shown in Fig. 6.1 which is fitted by equation 6.1.6.

For the direction reconstruction, a constant curvature of the shower front is assumed. The time residuals relative to a hypothetical shower plane are given by

$$\Delta t(r) = 4.823 \times 10^{-4} ns - 19.4 ns \left(\exp \left(-\frac{r^2}{118.1^2} \right) - 1 \right), \tag{6.16}$$

with a fluctuation of

$$\sigma_t(r) = 2.92 ns + 3.77 \times 10^{-4} r^2. \tag{6.17}$$

Equation 6.1.6 is umbrella-like and displayed in Fig. 6.6. The likelihood of observing a charge at time t^m is thus given by

$$P_{time}(t^m) = \frac{1}{\sqrt{2\pi}\sigma_t(r)} \exp \left(\frac{(t^m - (t^{planefit} - \Delta t(r)))^2}{2\sigma_t^2(r)} \right). \tag{6.18}$$

The final logarithmic likelihood is arrived by adding up the three components and doing the iterative fit. Fig. 6.7 displays the fit result for the same high energy particle in Fig. 6.1.

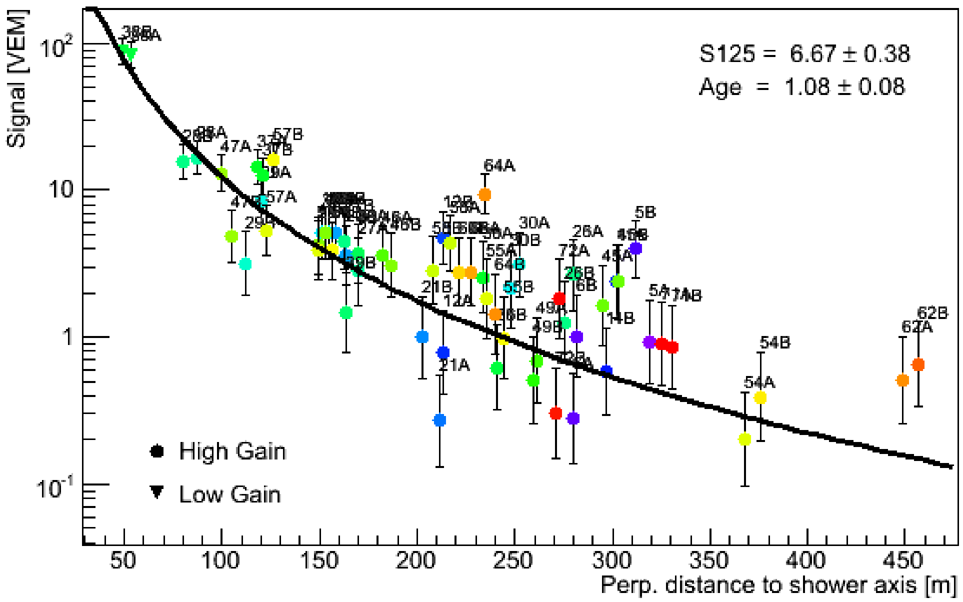


Figure 6.7: Reconstruction results for run 113849 event 13915. The fitted core position ($x = 260.85$, $y = 13.65$)m, shower direction ($\theta = 48.56$, $\phi = 104.53$)degree, $S_{125} = 6.67$ and $\beta = 2.77$ are put into equation 6.1.6 to get the fitted signal curve. The primary energy of this shower is about 19.6 PeV. The plot is made by using topEventBrower project.

6.1.7 Quality Cuts

Setting quality cuts on event data is a balance between minimizing reconstruction errors and preserving the detection efficiency. Though the cuts are loosened or tightened for different experimental tasks, the general settings are

- IceTopSTA3_InIceCoincidence
- Successful IceTop reconstruction [1]
- $2 \leq \beta < 4.5$, the slope of the lateral distribution at 125 meters is greater than 2 and less than 4.5.
- The reconstructed zenith angle is less than 40 °.

---

[1] This implicitly requires that the number of triggered stations is at least 5.

- The station with the largest charge signal is not on the border of the IceTop detector array .
- $\sqrt{\sigma_{X_c}^2 + \sigma_{Y_c}^2} < 20m$, where σ_{X_c} and σ_{Y_c} are the errors of reconstructed shower core (x, y), which is obtained as the final step sizes of the shower core in the iterative process.
- IT size [2] is less than 0.9

6.1.8 Reconstruction Efficiency

A special run of 1×10^5 proton and iron CORSIKA showers with an E^{-1} spectrum from 1 to 10 PeV are generated for the study of the efficiency study. The generation radius is 800 meters centered on $(0, 0)$ and the zenith angle is from $1°$ to $40°$. Fig. 6.9 shows the number of reconstructed events that survive the quality cuts. The reconstruction efficiency is about 80% in this energy range for contained events and remains constant at the high energy end. Table 6.1 boils down the reconstruction efficiency to different sources.

6.1.9 Energy Estimation

The primary energy of air showers is computed as a combination of a first guess and secondary correction [3]:

$$E_0 = E^{fg} \cdot 10^{-C(\theta)}. \tag{6.19}$$

---

[2] The reconstructed shower core at the surface has a high accuracy (about tens of meters). The IceTop containment size is a measure of how close the reconstructed core is to the detector boundary. Size 1.0 means the shower lands right on the border and 0.9 is found to be an optimal maximum value for this parameter. The effect of the containment size parameter is shown in Fig. 6.8.

[3] The reason for a secondary correction is because IceTop detector array is not as symmetric as a ring detector array, which was studied in [70]. Fabian Kislat worked out the new parameters for IC59.

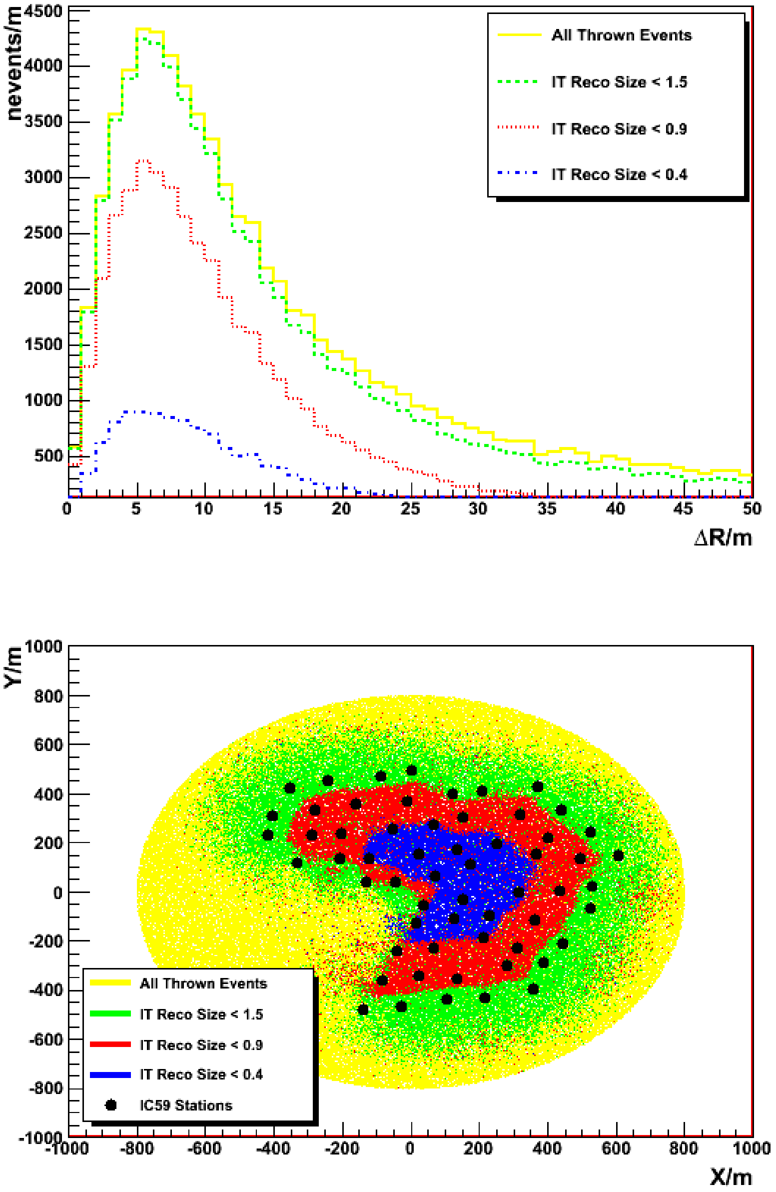


Figure 6.8: The upper panel is the core error assuming three different containment size values and $\Delta R = \sqrt{(x_{reco} - x_{true})^2 + (y_{reco} - y_{true})^2}$. The lower panel shows the true core positions for different containment size cuts for IC59 events. The smaller the containment size cut, the more contained are the events, and thus the better control in shower core reconstruction error.

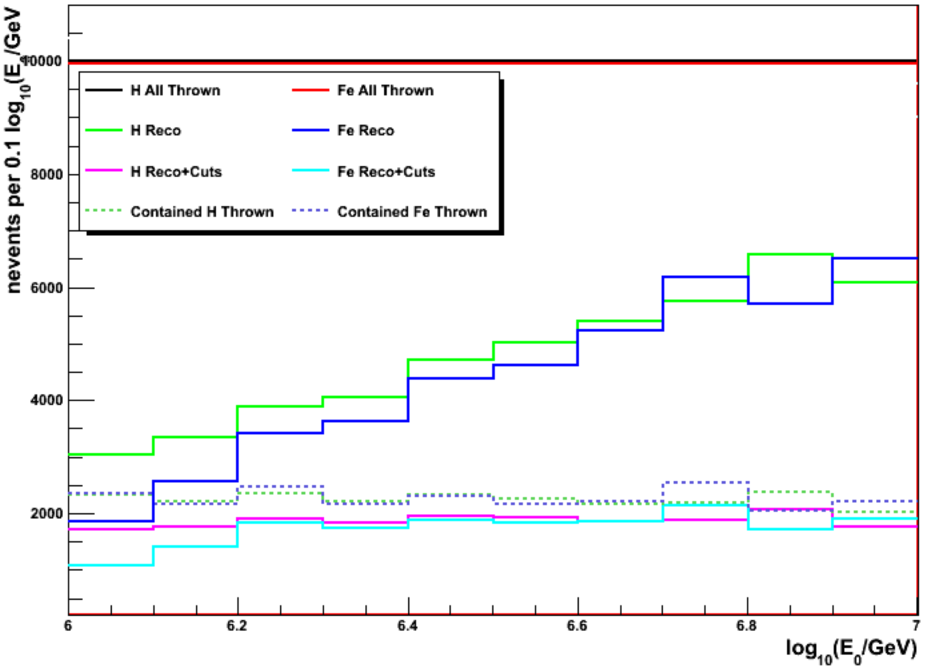


Figure 6.9: Number of events passing the cuts. Bin size of the histogram is 0.1. The number of proton and iron showers in each bin is 10^4. The number of events that can be reconstructed successfully increases as the primary energy increases. $\log_{10}(E/GeV)$.

Table 6.1: Cumulative Reconstruction Efficiency (H/Iron, percent of thrown events)

$\log_{10}(E_0/GeV)$	6.0-6.2	6.2-6.4	6.4-6.6	6.6-6.8	6.8-7.0
IT SMT Trigger	48.7/43.1	54.7/53.4	63.7/60.8	71.6/71.2	79.7/81.1
+nStation > 5	32.3/22.5	40.3/35.3	49.1/46.0	56.9/55.6	65.5/65.9
+NotOnBorder	18.9/14.6	20.5/19.9	21.7/21.1	22.4/22.5	24.6/24.4
Coverged	17.9/13.6	19.6/18.7	20.8/19.8	21.5/21.4	23.8/23.2
$+\Delta R < 20m$	17.9/13.5	19.6/18.7	20.8/19.8	21.5/21.4	23.8/23.2
+Size < 0.9	17.5/13.2	18.9/18.1	19.5/18.9	18.9/19.3	19.6/19.5
$+2.4 < \beta < 3.8$	15.4/9.5	17.7/15.3	19.0/17.7	18.7/18.7	19.6/19.1

The method that converts S_{125} to energy is described in Fabian Kislat's thesis [71]. The first guess energy is determined by

$$\log_{10} E^{fg}(\sec\theta, \log_{10} S_{125}) = p_0 + p_1 \sec\theta - \sqrt{p_2 + p_3 \sec\theta - p_4 \log_{10} S_{125}}, \quad (6.20)$$

where

$$\begin{aligned} p_0 &= 25.49, \\ p_1 &= 11.66, \\ p_2 &= 499.92, \\ p_3 &= 877.28, \\ p_4 &= 74.24. \end{aligned}$$

The secondary correction is

$$C(\theta) = 0.9797 - 1.635 \sec\theta + 0.5483 \sec^2\theta. \quad (6.21)$$

The reconstruction error is less than 2% for showers from 1PeV to 50 PeV as shown in Fig. 6.10 and has a minimum around 10 PeV.

6.2 InIce Reconstruction

IC59 IceCube InIce reconstruction uses the data from 3540 DOMs deployed along 59 strings in the South Pole Ice buried 1450 meters below the surface. The feature extractor converts the calibrated DOM waveforms to pulses. The MPE fit uses the pulses to reconstruct the direction and vertex of the muon track while MuE calculates the muon energies in the ice. For details in feature extractor please see Appendix C.

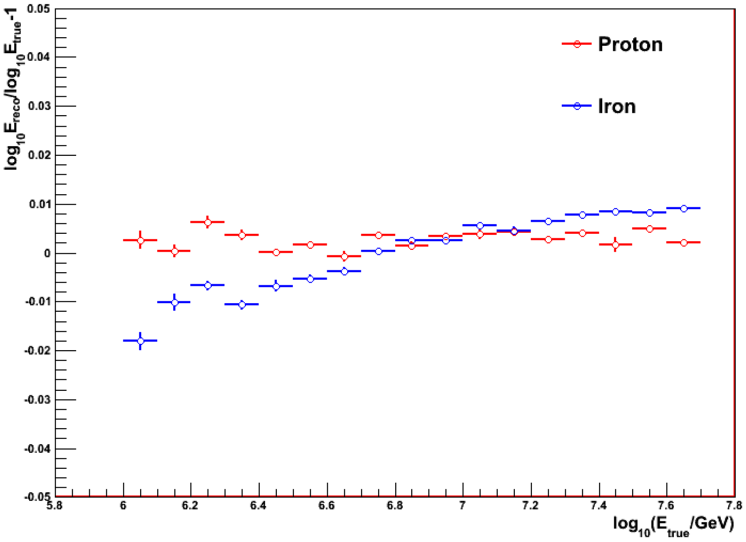


Figure 6.10: The energy reconstruction error based on simulation of 1×10^5 proton and iron showers from 1 to 50 PeV. The setting of this simulation is identical to that used in calculating the IceTop reconstruction efficiency.

6.2.1 InIce Track Direction

The In-Ice multi-photoelectrons (MPE) track reconstruction is accomplished by minimizing the likelihood function described by

$$\mathcal{L} = \prod_{OMs} N_{pe} \cdot \mathcal{F}_\sigma(\rho, \xi, t_1) \left[\int_{t_1}^{\infty} \mathcal{F}_\sigma(\rho, \xi, t) \right]^{N_{pe}-1}, \tag{6.22}$$

where N_{pe} is the integrated charge of the DOM over the whole event and t_1 is the arrival time of the first photon. The probability p is a Gaussian convoluted Pandel function as follows [73]

$$\begin{aligned} \mathcal{F}_\sigma &= \int_0^\infty \frac{\mathbf{d}x}{\sqrt{2\pi\sigma^2}} p(\rho, \xi, t) e^{-(t-x)^2/2\sigma^2}, &(6.23) \\ &= \int_0^\infty \frac{\mathbf{d}x}{\sqrt{2\pi\sigma^2}} \left[\frac{\rho^\xi t^{\xi-1}}{\Gamma(\xi)} e^{-\rho t} \right] e^{-(t-x)^2/2\sigma^2}. &(6.24) \end{aligned}$$

where ρ is related to the absorptivity of the ice and around 0.004 ns^{-1}. ξ represents the distance between the emission and detection location of a Cerenkov photon in units of the mean photon scattering length λ and $\xi = d/(\lambda \sin\theta_c)$. The standard deviation σ represents the time resolution of the detector, which is set to 4 ns in MPE track reconstruction. The time residual t is

$$t \equiv t_{hit} - t_{geo} \quad , \tag{6.25}$$

where t_{hit} is the leading edge of the first pulse of the DOM, t_{geo} is the photon arrival time for the case of no scattering and no absorption [73]. Fig. 6.11 depicts how t_{geo} is calculated. Given $\cos(\theta_c) = 1/n_{ph}$, the expected photon arrival time is

$$t_{geo} = t_0 + \frac{1}{c}\left[\hat{v}\cdot\vec{r} + d\frac{n_{gr}n_{ph} - 1}{\sqrt{n_{ph}^2 - 1}}\right] \quad , \tag{6.26}$$

where n_{ph} and n_{gr} are the phase and group refractive indices respectively.

Fig. 6.12 shows the angular resolution of InIce direction reconstruction. The resolution is between 1 ° to 2 ° over the primary particle angles of interest.

6.2.2 IceTop and InIce Direction

So far the most advanced IceTop direction reconstruction is ShowerCombined fit [4]. Which InIce direction reconstruction performs the best for muon bundles is uncertain, but this analysis uses MPEFit. A non-rigorous comparison [5] of the reconstructed zenith angles in Fig. 6.13 and Fig. 6.14 indicates that ShowerCombined outperforms MPEFit in angular reconstruction for contained events but underperforms for uncontained events. For comparisons of the azimuthal angle and core locations of the two reconstructions, please check wiki [74].

---

[4] A combined fit of shower direction, core location, curvature and S_{125} implemented in toprec/I3TopLateralFit.

[5] This comparison was done with IC40 simulation dataset 2615.

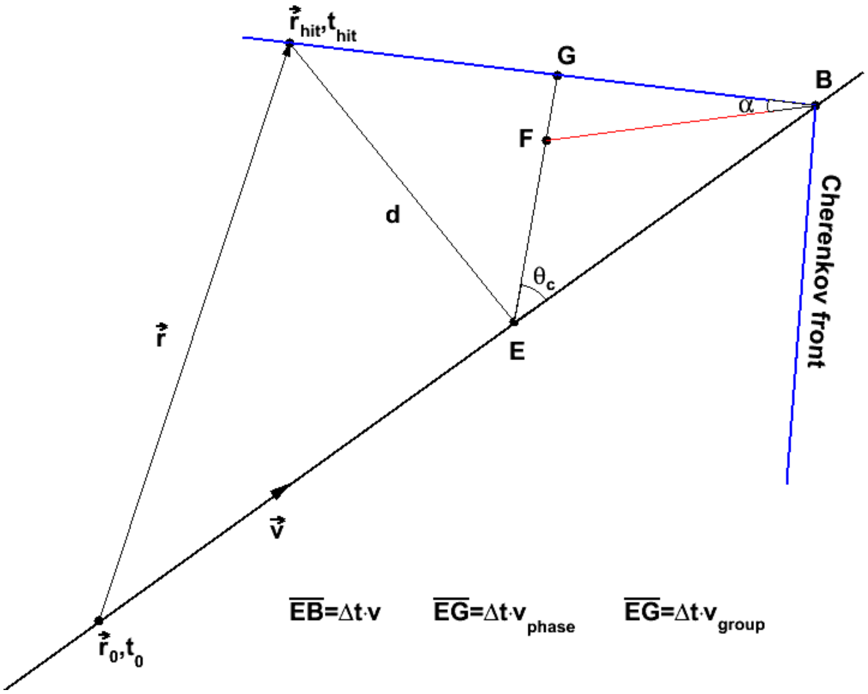


Figure 6.11: Geometry of the signal generation process. E is the closest approach from the DOM to the particle track and G is the closest approach from E to the Cerenkov front. B is particle position when the emitted photon hits the DOM. Suppose Δt is the time it takes for the particle to travel from E to B, the Cerenkov front moves with a phase velocity $v_{phase} \equiv c/n_{ph}$ from E to G in Δt, where n_{ph} is the phase refractive index in the ice. However, the detector is sensitive to the real photon that travel with a group velocity $v_{group} \equiv c/n_{gr}$, where n_{gr} is the group refractive index. Since $v_{group} < v_{phase}$, the real photon lags behind the Cerenkov front. The real photon wavefront is considered to be line BF [73].

6.2.3 Muon Energy Reconstruction

Several independent InIce energy reconstruction tools are available for the study of cosmic ray composition. They are mue, muon-bundle-reco, photorec-llh, muonbundle-llh, which are described as follows:

1. The *MuE* energy reconstruction is performed by using the track reconstruction

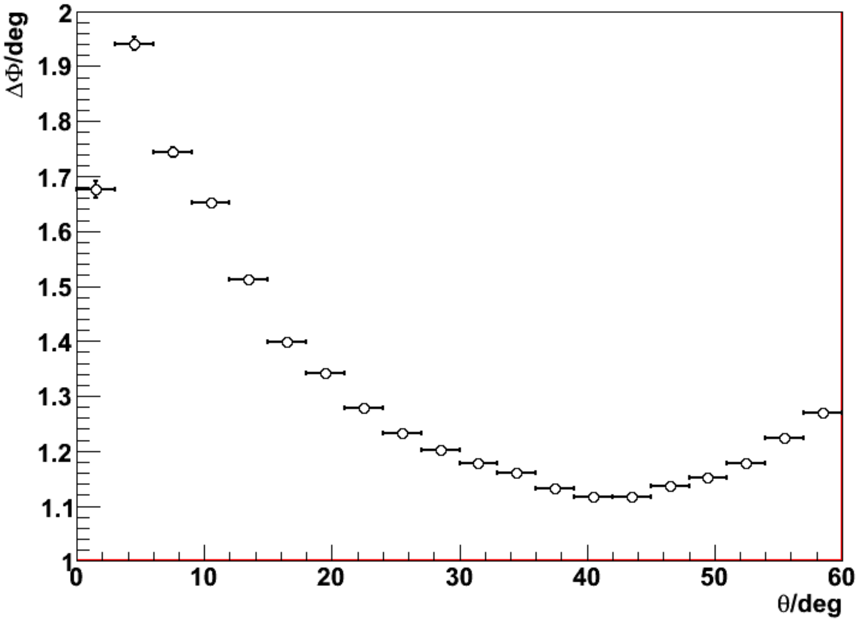


Figure 6.12: Angular resolution of InIce track reconstruction from dataset 4930 (Polygonato model with unweighted spectrum) that is calculated as $\Delta\Phi = \arccos(\cos\theta_{\rm true}\cos\theta_{\rm reco} + \sin\theta_{\rm true}\sin\theta_{\rm reco}\cos(\phi_{\rm reco} - \phi_{\rm true}))$, where $(\theta_{\rm true}, \phi_{\rm true})$ is true direction of the primary particle and $(\theta_{\rm reco}, \phi_{\rm reco})$ is the reconstructed direction.

result as an input. The log likelihood function is

$$\begin{aligned}
-\ln\left(\prod_{OMs} P(n_i|\mu_i)\right) &= -\ln\left(\prod_{OMs}\prod_{i=1}^{k}\frac{\mu_i^{n_i}}{n_i!}e^{-\mu_i}\right) \\
&= \sum_{OMs}\left(-\sum_{i=1}^{k} n_i \ln\frac{\mu_i}{\mu} + \sum_{i=1}^{k}\ln(n_i!) - \ln(N!)\right) \\
&\quad + \sum_{OMs}\left(-N\ln\mu + \mu + \ln(N!)\right) \\
&= \mathcal{L}_{track} + \sum_{OMs}\left(-N\ln(N_l\mu_0) + N_l\mu_0 + \ln(N!)\right) \quad ,
\end{aligned}$$

where n_i and μ_i are the observed and expected photons in each DOM, $\mu = \sum_{i=1}^{k}\mu_i = N_l \cdot \mu_0(d)$ is the expected total number of photons observed by the

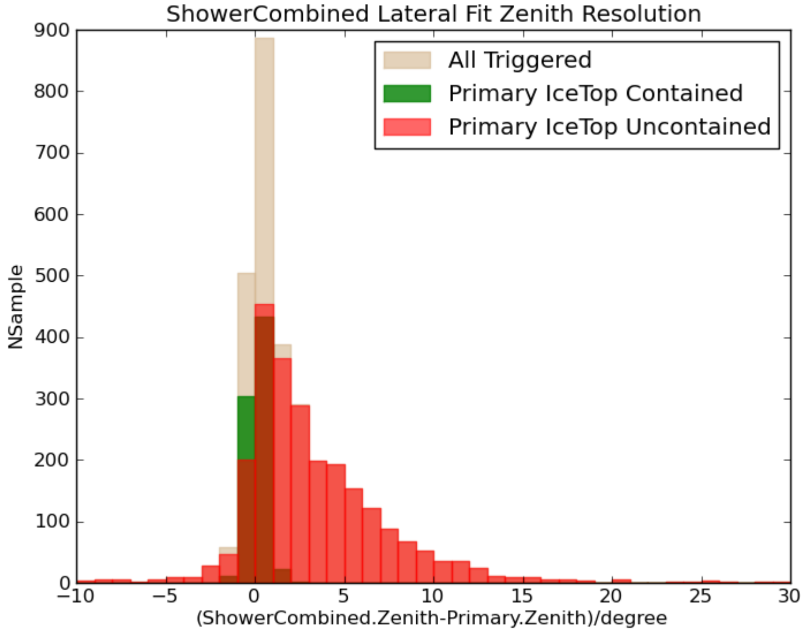


Figure 6.13: ShowerCombined zenith difference from the true zenith in muon bundle events.

DOM, N_l is the average number of photons emitted per unit length by a muon track. Thus, N_l is our energy proxy here. Since μ_i/μ in the first term of the above equation reduces to the Pandel function, N_l only appears in the second term. Assume the input track is optimal, the first term becomes a constant. Differentiating the log likelihood function about N_l and making it equal zero leads to the following equation

$$N_l = \frac{\sum_{OMs} N}{\sum_{OMs} \mu_0(d)} \quad , \tag{6.27}$$

where N is the total observed photons in a DOM and $\mu_0(d) = \mu/N_l$ is the total expected number of photons from unit length of the track. This flux can be calculated as [75]

$$\mu_0(d) = \left[A_{eff} \cdot \frac{f(\eta) + 4g(d)\bar{(\eta)}}{1 + g(d)}\right] \frac{\exp\left(-\frac{d}{\lambda_a \sin\theta_c}\right)}{2\pi\sqrt{\lambda_\mu d}\,\tanh(d/\lambda_\mu)} \quad , \tag{6.28}$$

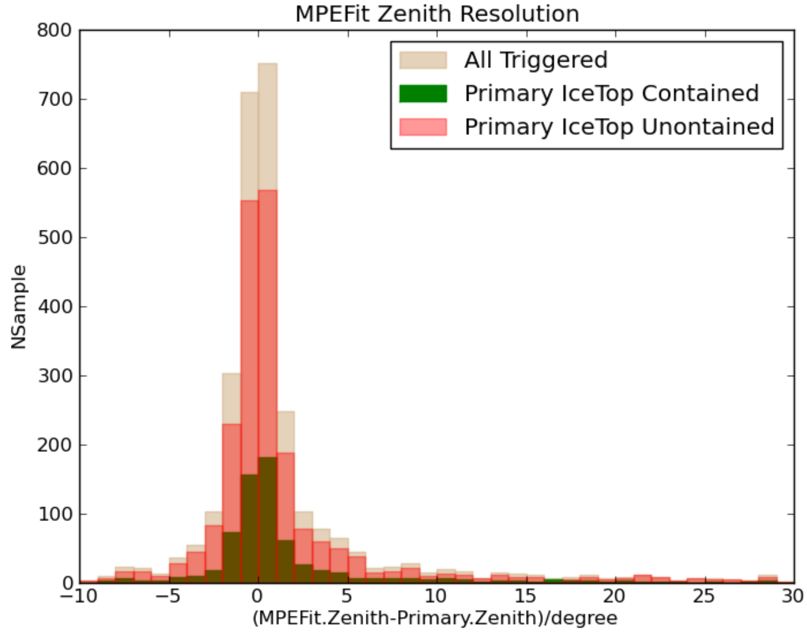


Figure 6.14: MPEFit zenith difference from the true zenith in muon bundle events.

where A_{eff} is the effective DOM PMT area, $f(\eta)$ is the relative sensitivity of the DOM for the photons arriving from an angle η relative to the PMT axis, λ_a is the absorption length, $g(d)$ is a correction to the DOM. PMT area visible to the Cerenkov photons in the near and far regions.

The *MuE* reconstruction can be calibrated to the muon bundle energy at the COG of in-ice IceCube by

$$\log_{10}(E_\mu/GeV) = 1.0911 \cdot \log_{10} \mathrm{MuE} - 1.0382 \tag{6.29}$$

2. photorec-llh is reconstructs the muon energy by using the waveform information obtained from feature extractor, which is binned into K bins. It requires the interfaces of photonics tables and gulliver framework. The probability of

observing n_i photons in the ith waveform bin given an expectation of μ_i photons in the same bin assumes Poissonian statistics. The likelihood function is thus [76]

$$\mathcal{L}(f(t)|\vec{x}, E) = \prod_{i=1}^{K} \frac{e^{-\mu_i}}{n_i!} \mu_i^{n_i}, \tag{6.30}$$

where $f(t)$ represents the waveform, $\vec{x}$ is the geometry and E the energy. The photorec-llh models the average muon energy loss per meter as a chain of equally space and monoenergetic cascades. This is called "lightsaber" model. Special photonics tables were created to calculate the expected charge values at any specific time. The expected charge is proportional to the expected average muon energy loss rate if it is greater than 1GeV/meter. The average muon energy loss rate is a single free parameter in the minimization, thus it is constant at all depths in an event. When applied to muon bundles, this value represents the size of the bundles in the ice. The drawbacks of using photorec-llh for muon bundles are obvious. First, it doesn't account for the muon range-out effect since it averages the muon energy loss through all depths. Second, it assumes a smooth muon energy loss while the true muons lose energy stochastically. Third, the PDF doesn't describe the waveforms of occasional individual photoelectrons well [76].

3. muon-bundle-reco (also known as K70) is originally developed for AMANDA then extended to reconstruct the in-ice muon bundles in IceCube. Compared to photorec-llh, muon-bundle-reco uses only the total charge observed in each DOM and accounts for the muon range-out effect. The expected charge signal in each DOM is

$$Q_{expect} = N N_{\mu}(X) \frac{1}{\sqrt{\lambda_{eff} d}} \exp^{-d/\lambda_{eff}}, \tag{6.31}$$

where N is a normalization factor, $N_{\mu}(X) = K[(\frac{a}{b})(e^{b_{eff}X} - 1)]^{-\gamma_{\mu}}$ is the assumed muon multiplicity at slant depth X. a and b are the coefficients in the single muon energy loss equation $\frac{dE}{dX} = a + bE(X)$. λ_{eff} is the effective attenuation length. The likelihood of an event is computed as

$$\mathcal{L} = \prod_{OM} P(Q_{measured}|Q_{expected}), \tag{6.32}$$

where P is a Poisson probability. The energy related fit result K70 is the average IceCube charge amplitude measured at 70 meters from the reconstructed track axis. In Chapter 7 and Appendix H a comparison between MuE and K70 is shown.

4. muonbundle-llh coded by Tom Feusels is a project that aims at fixing the problems of photorec-llh in reconstructing muon bundles. The single energy scale is substituted for the muon energy loss rate that describes the muon-range out effect as a function the slant depth,

$$\frac{dE}{dX} = e^{-bX} K \gamma^{\mu} \left(\frac{E_0}{A}\right)^{\gamma - 1} \left[\left(\frac{E_0}{A}\right)^{-\gamma^{\mu}} \left(\frac{a}{\gamma^{\mu}} - \frac{bE_0/A}{1-\gamma^{\mu}}\right) + E_{min}^{-\gamma} \left(\frac{a}{\gamma^{\mu}} - \frac{b}{E_{min}(1-\gamma^{\mu})}\right)\right], \tag{6.33}$$

where $a = 0.24$, $b = 3.3 \times 10^{-4}$, E_0 is the primary energy, A is the primary mass, $K = 14.5$, $\gamma^{\mu} = 1.75$ and $E_{min} = (a/b)(e^{bX} - 1)$ is the minimum muon energy required at surface for this muon to reach this slant depth. E_0/A is treated as a free parameter in the minimization. Plugging the minimized E_0/A and a certain slant depth X to Eq. 4 would give the reconstructed muon bundle energy at X.

Chapter 7

COMPOSITION OF COSMIC RAYS

The primary composition of cosmic rays varies with their primary energy. The ultimate goal of this analysis is to find the composition within the energy range that can be covered by IC59 detector. The detector has a trigger threshold around 300 TeV, which is the lower bound of the cosmic rays detected. The largest event could have an energy greater than 10^{18} eV. However, the limitation of simulation brings the upper bound to 100 PeV. Looking for cosmic ray composition within this energy range requires simulating events of different primary particle types. The heaviest primary particles that can be generated are iron nuclei so that iron are assumed to be the heaviest cosmic rays we observe. In the 2-component model, only proton and iron showers are simulated and all observed events are assumed to be either protons or irons. In the 5-component model, three more elements are simulated and they are helium, oxygen and silicon. So the problem becomes what is the number of those five kinds of particles that reach the earth within a certain area, solid angle, time period and primary energy.

7.1 Energy Proxies

The composition of the cosmic rays with certain primary energy has a correlation with the electron and muon ratios of the air showers on the ground. The electron and muon numbers of vertical showers at some discrete energies are shown in Fig. 7.1.

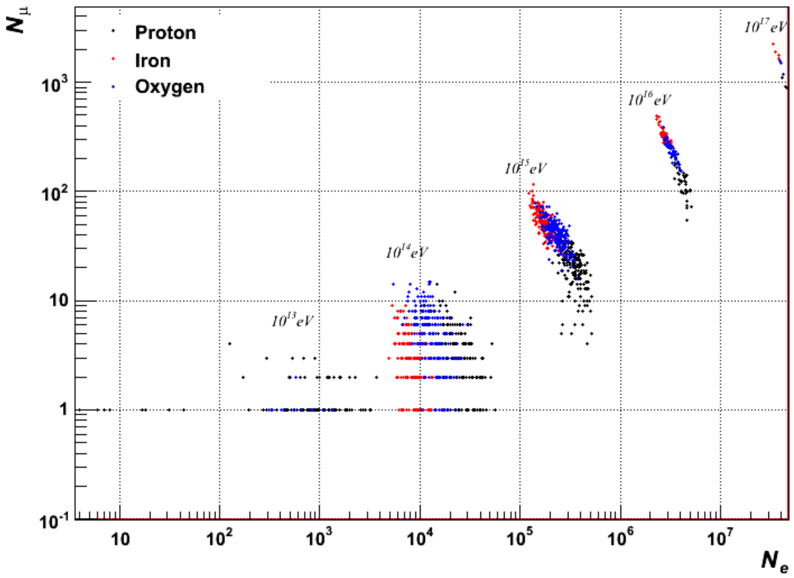


Figure 7.1: Number of muons and electrons for vertical showers from $10^{11} - 10^{17}$eV with SYBILL CORSIKA production. The energy cutoffs are 300GeV and 0.01GeV for muons and electrons respectively.

In order to figure out the primary energy and particle type, we need to find some reconstruction parameters that bear the features of the electrons and muons in the shower. The candidates parameters are S_{125} and E_{reco} from the IceTop reconstruction, E_μ [1], photorec-llh and K70 from the InIce reconstruction. Those parameters were described in Chapter 6. Since the composition is also our goal as well as the energy spectrum, the selected energy proxies should be sensitive to the composition. The IceTop reconstruction S_{125} and E_{reco} has strong correlations to the primary energy but almost no sensitivity to composition. The InIce energy proxies are sensitive to both the primary energy and particle type. So how many energy proxies do we need? The answer is at least two. A single InIce parameter cannot fulfill the job because an increase in the energy proxy could be either from an increase in the primary energy or increase in the heaviness of the primary particle.

---

[1] in-ice muon-bundle (COG) energy, Linearly related with MuE, see Eq. 6.29.

Thus, E_{reco} and E_μ are chosen as the energy proxies in this analysis to uncover the primary energy spectrum and particle type of cosmic rays. In fact, the reconstructed surface energy E_{reco} is quite close to the true primary energy (see Fig. 7.2 and Appendix A). The sensitivity to both the primary energy and composition of the InIce energy proxies can be seen from the Elbert's formula [77]:

$$< N_\mu > \approx A \times \frac{14.5\mathrm{GeV}}{E_\mu \cos\theta} \left(\frac{E_0}{AE_\mu}\right)^{0.757} \left(1 - \frac{AE_\mu}{E_0}\right)^{5.25}, \tag{7.1}$$

where N_μ is the number of muons with an energy greater than E_μ at the ground, A is the atomic number, θ is the zenith angle of the primary particle and E_0 is the primary energy. For a muon to reach the InIce detector of IceCube, the minimum energy is supposed to be [47]

$$E_\mu^{min} = \frac{a}{b}\left(e^{bX} - 1\right), \tag{7.2}$$

where $a = 0.24\mathrm{GeV} \cdot kg/m^2$, $b = 3.3 \times 10^{-4} kg/m^2$ and X is the slant depth from muon core at the surface to its end point in the ice. Fig. 7.3 shows proton and iron showers would produce different amounts of muons that can reach the center of the InIce detector. Different muon multiplicities are likely to lead to different muon energy loss and different IceCube signals.

7.2 Cosmic-ray Composition

Between the two energy proxies E_{reco} and E_μ, E_{reco} is a good indicator of the primary energy but has negligible dependence on the primary mass while E_μ is sensitive to both the primary energy and mass. The mass sensitivity of E_μ can be seen in Fig. 7.4, Fig. 7.5 and Appendix A, where the distributions of E_μ for three groups of elements are plotted at a fixed E_{reco}. An easy way to estimate the composition is to compare the experimental data against the simulation and find a match, as exemplified in Fig. 7.6 and Fig. 7.7. The result is listed in Table 7.1.

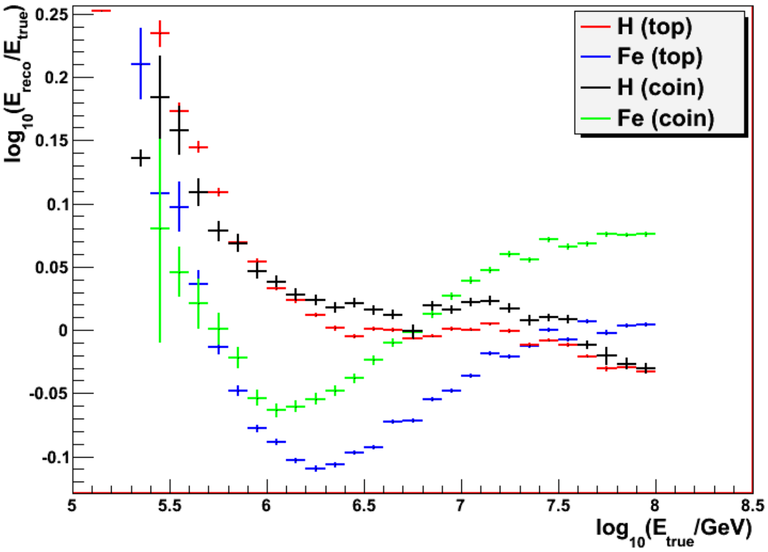


Figure 7.2: The difference between the reconstructed surface energy and true energy in logarithmic scale versus the true energy. Different colors represents proton and iron either with IceTop only or IceTop-InIce coincident events.

Table 7.1: Fit results

$\log_{10} E_{min}$	$\log_{10} E_{max}$	H	O+He	Si+Fe
6.0	6.2	0.362	0.539	0.099
6.2	6.4	0.217	0.781	0.002
6.4	6.6	0.396	0.506	0.098
6.6	6.8	0.303	0.697	0.000
6.8	7.0	0.244	0.755	0.001
7.0	7.2	0.299	0.565	0.136
7.2	7.4	0.000	0.783	0.217
7.4	7.6	0.120	0.880	0.000
7.6	7.8	0.087	0.261	0.652
7.8	8.0	0.000	0.380	0.620
8.0	8.2	0.224	0.485	0.291

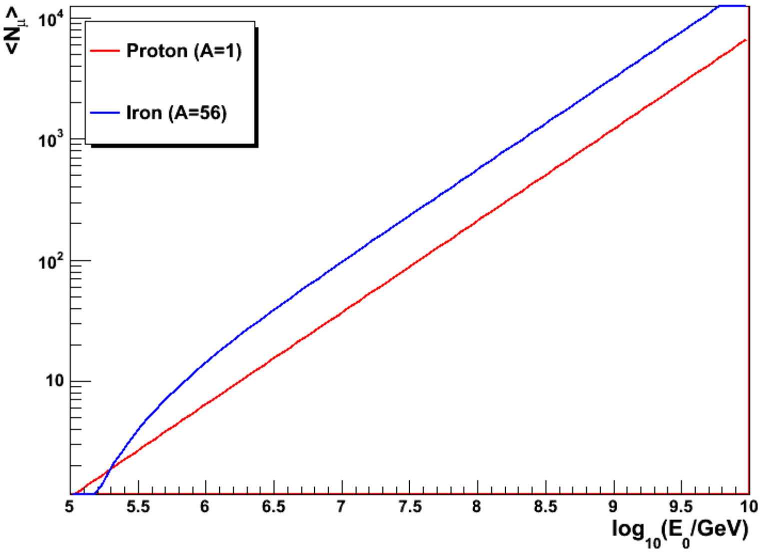


Figure 7.3: The Elbert's formula Eq. 7.1 with $A = 1$ or $A = 56$, $\cos\theta = 1$ and $E_\mu = \frac{0.24}{3.3\times10^{-4}}(e^{3.3\times10^{-4}\times0.92\times2000} - 1)$. A vertical iron shower produces about three times more muons with an energy sufficient to propagate 2000 meters in the ice than a vertical proton shower does for energies above 1PeV.

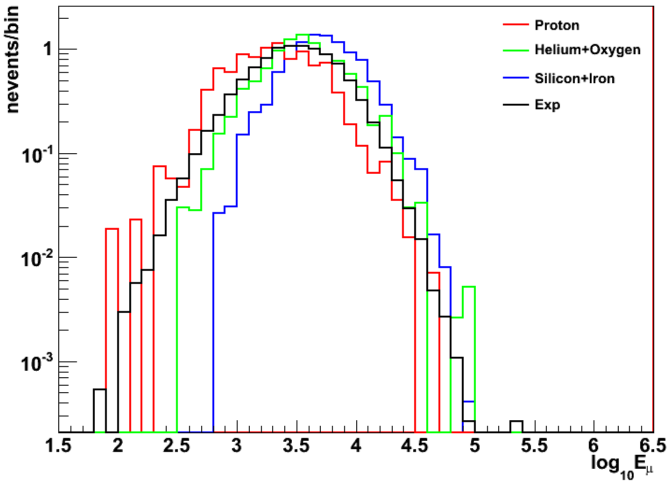


Figure 7.4: Distributions of E_μ for $6.0 \leq \log_{10}(E_{reco}/GeV) < 6.2$.

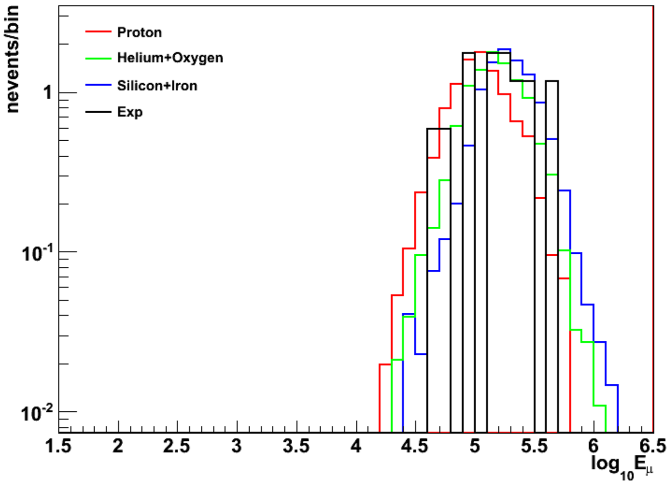


Figure 7.5: Distributions of E_μ for $7.8 \leq \log_{10}(E_{reco}/GeV) < 8.0$.

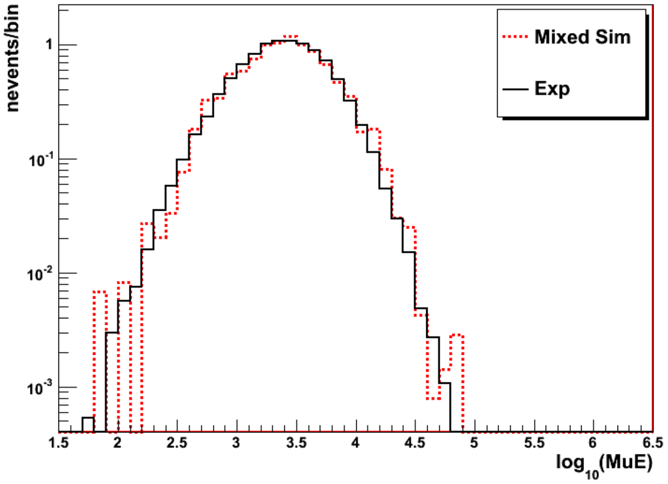


Figure 7.6: Fit result for $6.0 \leq \log_{10}(E_{reco}/GeV) < 6.2$. The fractions of the three groups are 0.36, 0.53 and 0.01.

7.3 Unfolding Strategy

Knowing the two energy proxies of each observed event, we have enough information to infer the energy spectrum and composition. In fact, simply plotting the muon-bundle energy against E_{reco} can give us some sense about the composition

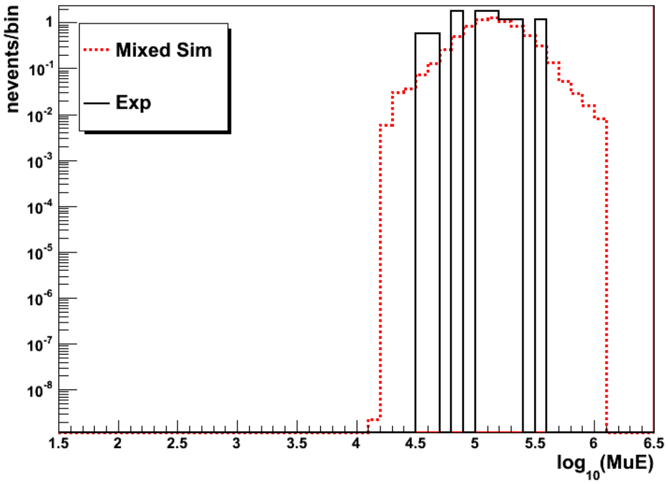


Figure 7.7: Fit result for $7.8 \leq \log_{10}(E_{reco}/GeV) < 8.0$. The fractions of the three groups are 0.22, 0.49 and 0.29. The fit is coarse because of the lack of observed high energy events.

of the measured cosmic rays. In Fig. 7.8, the experimental data lies roughly in the middle of the proton and iron simulation, indicating the measured cosmic rays can roughly be represented by a mix of proton and iron events.

However, it is still hard to solve the spectrum and composition separately, since they depend on each other. Thus, to learn the spectrum, you have to know the composition of the cosmic rays; to learn the composition, you need to apply some energy spectrum. This dilemma can be solved by using neural networks or Bayesian unfolding. The method of neural networks has been discussed by Karen Andeen in her paper about IC40 composition and Ph.D thesis and will not be repeated here. Nonetheless, the results from using NN (neural networks) and Bayesian unfolding techniques have some differences that deserve to be addressed:

- The Bayesian unfolding method is binned while the NN method is not.
- The Bayesian unfolding results lie in the same space as the training data. For example, with a 2-component model, if proton is assigned to 0 and iron is

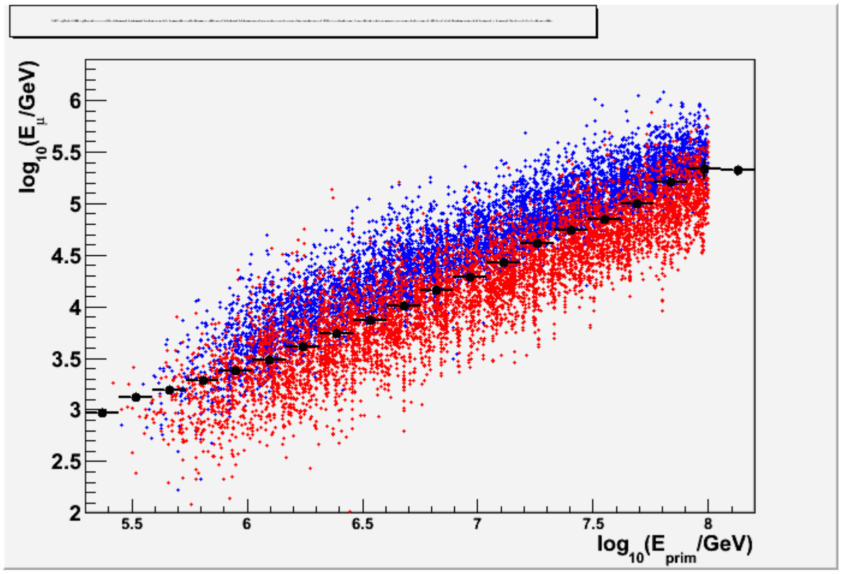


Figure 7.8: $\log_{10}(E_\mu/\mathrm{GeV})$ versus $\log_{10}(E_{\mathrm{prim}}/\mathrm{GeV})$. The red and blue ntuples are simulated proton and iron events. The black profile is the experimental data. The X axis for the experimental data is $\log_{10} E_{reco}$ instead of $\log_{10} E_{prim}$. The simulation and experimental datasets are described in Section 7.5.

assigned to 1, then the unfolded events could only have 0 or 1 as their type indicator. In NN, the results may have anything from 0 to 1.

- In NN, the primary energy and relative heaviness of each measured event [2] can be obtained. The Bayesian unfolding results can give the abundance of events in each energy and particle type bin.
- Detection efficiency is automatically taken care of in Bayesian unfolding because it registers not only those events that pass all the cuts but also the silent events. The direct output of NN is supposed to be the energy spectrum and composition of the measured events that pass all the cuts. To obtain the cosmic ray flux, the result needs to be adjusted for detection efficiency.

---

[2] Events that can pass all the cuts.

The details of the Bayesian unfolding method to this problem will be discussed in Section 7.4.

7.4 Bayesian Unfolding Technique

The principle of this method is described in D'Agostini's paper *A multidimensional unfolding method based on Bayes' theorem* and it is straightforward to use it in solving the composition problem with IC59.

In an experiment, if one observes $n(E_j)$ events with effect E_j $(j = 1, 2, 3 \ldots)$, the expected number of events due to each of the causes C_i is

$$\hat{n}(C_i) = \sum_j n(E_j) P(C_i|E_j),$$

where $P(C_i|E_j)$ is the probability that an effect E_j comes from the cause C_i. Since not every cause produces an effect, the inefficiency needs to be taken into account. The above equation then becomes

$$\hat{n}(C_i) = \frac{1}{\epsilon_i} \sum_j n(E_j) P(C_i|E_j), \tag{7.3}$$

where $0 < \epsilon_i \equiv \sum_j P(E_j|C_i) \leq 1$.

If we define $P(C_i) = \frac{1}{\epsilon_i} \sum_j P(C_i|E_j)$, the Bayes' theorem suggests that

$$P(C_i|E_j) = \frac{P(E_j|C_i) P(C_i)}{\sum_i P(E_j|C_i) P(C_i)}, \tag{7.4}$$

which can be rewritten by substituting an initial probability $P_0(C_i)$ for $P(C_i)$ as

$$P(C_i|E_j) = \frac{P(E_j|C_i) P_0(C_i)}{\sum_i P(E_j|C_i) P_0(C_i)}. \tag{7.5}$$

The initial probability should represent be our best knowledge of the distribution. Since we know that the general cosmic rays follow a $E^{-2.7}$ below the knee and steepen to $E^{-3.1}$, $E^{-2.7}$ will be chosen as initial probability for each energy bin. Because we know much less about the composition, it is reasonable to assume an even distribution among the five elements.

Thus, Eq. 7.3 can be rewritten as

$$\hat{n}(C_i) = \frac{\sum_j n(E_j)P(E_j|C_i)P_0(C_i)}{[\sum_j P(E_j|C_i)][\sum_i P(E_j|C_i)P_0(C_i)]}. \tag{7.6}$$

In this particular case, the cause is binned in logarithmic primary energy and composition and the effective is binned in $\log_{10}(E_{\mathrm{reco}}/\mathrm{GeV})$ and $log_{10}MuE$. So the subscripts in Eq. 7.6 can be changed as $i \longrightarrow iA$ and $j \longrightarrow jk$, where i is the primary energy bin, A is the particle type, j is $\log_{10}(E_{reco}/\mathrm{GeV})$ bin and k is $log_{10}MuE$ bin. Thus, we have

$$\hat{n}(C_{iA}) = \frac{\sum_{jk} n(E_{jk})P(E_{jk}|C_{iA})P_0(C_{iA})}{[\sum_{jk} P(E_{jk}|C_{iA})][\sum_{iA} P(E_{jk}|C_{iA})P_0(C_{iA})]}. \tag{7.7}$$

Several things need to be noted here

- $\sum_{iA} P_0(C_{iA}) = 1.$
- $\sum_{iA} P(C_{iA}|E_{jk}) = 1.$
- $0 < \epsilon_{iA} \equiv \sum_{jk} P(E_{jk}|C_{iA}) \leq 1.$

Once we get $\hat{n}(C_{iA})$, a new probability of a cause can be computed as

$$\hat{P}(C_{iA}) \equiv P(C_{iA}|n(E)) = \frac{\hat{n}(C_{iA})}{\sum_{iA} \hat{n}(C_{iA})}. \tag{7.8}$$

This $\hat{P}(C_{iA})$ is closer to the true distribution of the causes $P_{true}(C_{iA})$. Doing this iteratively, a spectrum very close to the truth can be obtained. The χ^2 at each step can be computed by comparing the $\hat{n}(C)$ of the current and previous step. Section 7.5 will discuss how to prepare the IceCube data for using the Bayesian unfolding method. The uncertainties in the unfolding will be discussed in Section 7.8.

7.5 Data Selection

The Bayesian unfolding method discussed in the previous section requires simulation inputs to calculate the conditional probability $P(E_j|C_i)$ and efficiency

ϵ_i. A series of cuts are built to filter the events and pick up the well-reconstructed events. The philosophy is the more contained and vertical the events, the better the reconstruction works. Those events that can pass all the cuts will be use in computing $P(E_j|C_i)$, but all the events are useful in computing ϵ_i.

7.5.1 Datasets

The analysis uses 21 days experimental data from June 21 to July 11, 2009 and 10^6 simulated cosmic ray events for proton, helium, oxygen, silicon and iron respectively with the IC59 detector [3]. The simulated events follow an E^{-1} energy spectrum [4] from 100 TeV to 100 PeV and are evenly distributed in a space of 0 to $\cos 40°$ zenith angle. Each generated SYBILL CORSIKA shower is thrown to an area with a radius of 1200 meters 100 times. The ice model used in the simulation is SPICE1 implemented with photonics.

7.5.2 Cuts

Not every simulated event can trigger the detector, and not every triggered event can be reconstructed. Various cuts are design to select well reconstructed events by IC59 detector. They are

- IceTopSTA3_InIceCoincidence filter passed
- TopLateralFit converges and gives meaningful results
- $2 \leq \beta < 4.5$
- IceTop Size less than 0.9
- Loudest IceTop station not on the border

---

[3] The corresponding five simulation datasets are 5133, 5185, 5186, 5187 and 5132.

[4] This means the generated events are evenly distributed with $\log_{10}(E/\mathrm{GeV})$.

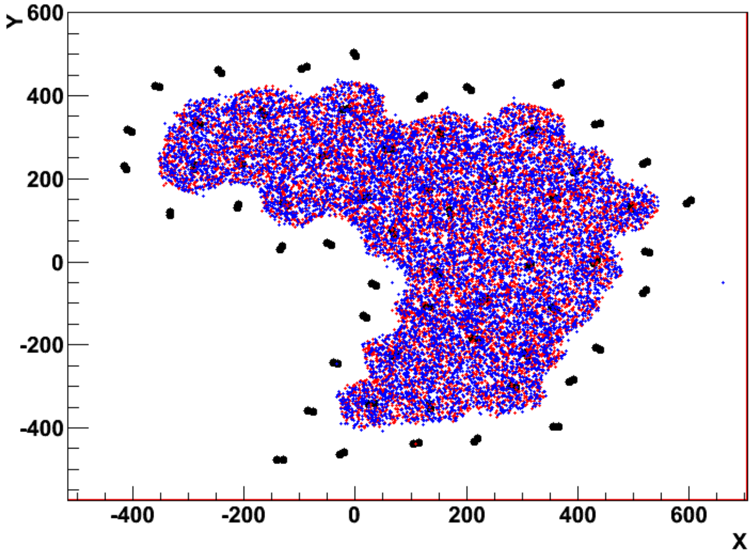


Figure 7.9: True surface core locations (in meters) of simulated events that can pass all the cuts. The black dots represent the IceTop tanks. The red dots are protons while the blue are iron.

- IceTop reconstruction core location error less than 20 meters
- InIce MPEFit converges
- InIce energy reconstruction gives reasonable results
- $0.97 \leq V_{shower}/c < 1.03$ [5]
- Difference between TopLateralFit and MPEFit angles below 4 degree
- Direct InIce track length greater than 100 meters [6]

---

[5] $c = 3 \times 10^8 m/s$ is the speed of light. The shower velocity is computed as the ratio of IceTop and InIce reconstructed core distance to time difference.

[6] The direct track length is compute as the distance between the first and the last hit DOMs in the InIce detector. Only hits with a window of $[-15ns, 75ns)$ from the expectation are counted.

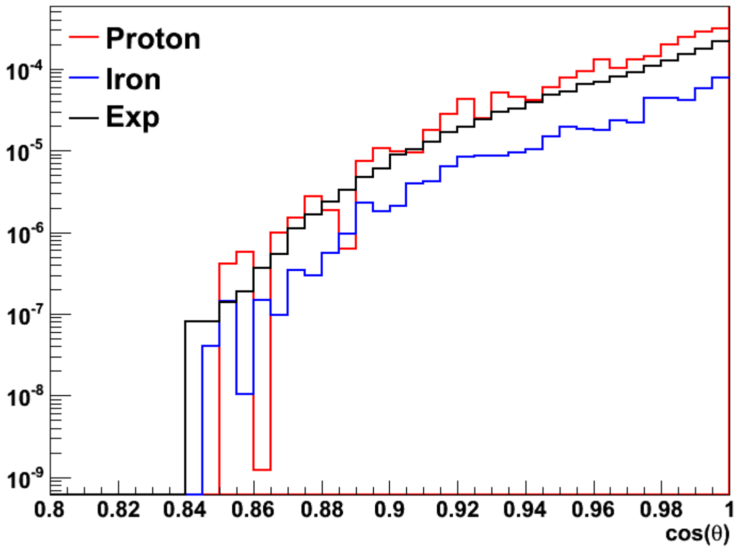


Figure 7.10: True zenith angle of simulated events and IceTop reconstructed zenith of the experimental events in arbitrary units. The red is proton, blue is iron and black is the experiment.

Because simulated showers are thrown to a very large area ($\pi \times 1200^2 m^2$), only a few of them can trigger the detector. Among those that can trigger the detector, only a small fraction can be well reconstructed by both IceTop and InIce detector array. The details of the effects of the cuts can be seen in Table 7.2, Table 7.3 and Table 7.4. For most of the energy bins, about 1% of the total generated events can survive the various trigger, filter conditions and cuts. Because the shower oversampling rate is 100, this roughly makes every event only appear once in the selected events. Those events that do not trigger or pass the cuts are also useful in finding out the detection efficiency.

The events that can pass those cuts are mainly well contained IceTop and InIce vertical cosmic ray showers as shown in Fig. 7.9 and Fig. 7.10. Note there is no explicit cut on the reconstructed zenith angles though both simulation and experiments have almost the same zenith distribution within a narrow range.

Table 7.2: No. of proton events that pass the cuts

$\log_{10}(\frac{E}{\text{GeV}})$	5.0-5.3	5.3-5.6	5.6-5.9	5.9-6.2	6.2-6.5	6.5-6.8	6.8-7.1	7.1-7.4	7.4-7.7	7.7-8.0
Generated [7]	100200	100200	100200	99800	99900	99700	99900	99600	98600	95600
IT Reco	17	638	5027	12893	18870	23989	30365	37249	44738	52638
β	11	549	4696	12371	18007	22203	26801	31151	35426	39513
IC Reco	1	130	1170	3012	4472	5628	7102	8310	9971	11277
IT Size	1	120	895	2132	2498	2524	2803	2905	3361	3310
Loud Sta	1	103	792	1848	2160	2157	2358	2415	2737	2615
IC Len	0	77	622	1481	1725	1689	1787	1799	1978	1810
IT Err	0	77	622	1480	1725	1688	1785	1797	1972	1805
Angle	0	38	374	1000	1236	1249	1332	1376	1504	1449
V_{shower}	0	38	373	995	1235	1245	1329	1371	1492	1437

The numbers of remaining events with the current condition and all the conditions above the current are shown for each energy bin. Items in the first column are

- All events generated
- IceTop reconstruction converges
- $2 \leq \beta < 4.5$
- InIce reconstruction converges
- IceTop size less than 0.9
- Loudest station not on edge
- InIce direct track length greater than 100 meters
- IceTop core location error less than 20 meters
- IceTop and InIce angle difference less than 4 degrees
- $0.97 \leq V_{shower}/c < 1.03$

Table 7.3: No. of iron events that pass the cuts

$\log_{10}(\frac{E_{reco}}{\text{GeV}})$	5.0-5.3	5.3-5.6	5.6-5.9	5.9-6.2	6.2-6.5	6.5-6.8	6.8-7.1	7.1-7.4	7.4-7.7	7.7-8.0
Generated	100200	100200	100200	100000	99900	99900	99800	99800	98800	98300
IT Reco	1	85	1724	8847	16610	23074	30302	38097	47450	58261
β	0	47	1360	8062	15569	20988	26194	30873	35905	41387
IC Reco	0	22	579	2648	4403	6119	7638	9205	11063	12108
IT Size	0	16	486	1984	2531	2939	3150	3353	3620	3504
Loud Sta	0	15	419	1730	2141	2492	2622	2756	2941	2786
IC Len	0	15	339	1365	1618	1845	1884	1952	2005	1874
IT Err	0	15	339	1365	1615	1844	1879	1952	1999	1872
Angle	0	9	223	949	1164	1362	1353	1509	1561	1507
V_{shower}	0	9	223	947	1158	1350	1333	1485	1523	1450

Refer to Table 7.2 for explanations of the numbers.

Table 7.4: No. of experimental events that pass the cuts

$\log_{10}(\frac{E}{\text{GeV}})$	5.0-5.3	5.3-5.6	5.6-5.9	5.9-6.2	6.2-6.5	6.5-6.8	6.8-7.1	7.1-7.4	7.4-7.7	7.7-8.0
ITSTA3_IC	94	14042	143708	216908	114323	45430	17616	6044	1948	590
IT Reco	93	14039	143707	216906	114319	45429	17614	6039	1944	588
β	2	5866	115733	202174	109161	42741	15636	4904	1470	401
IC Reco	2	5863	115611	201962	109030	42681	15617	4900	1465	401
IT Size	2	5510	101928	154750	66089	20319	5358	1446	408	114
IT Loud Sta	2	4701	87842	131708	55173	16924	4452	1179	340	90
IC Len	1	3929	72792	108266	44857	13528	3477	913	248	69
IT Err	1	3923	72694	108150	44795	13509	3467	910	248	69
Angle	1	2135	40204	61083	25725	7900	2017	534	154	47
V_{shower}	1	2130	40033	60792	25578	7847	1994	532	153	39

Refer to Table 7.2 for explanations of the numbers.

7.6 Unfolding Settings

The only input to the unfolding method introduced in Sec. 7.3 is the probability distributions in the observation from any cause $P(E_j|C_i)$ in Eq. 7.6, which can be learned by filtering the simulation that is described in the previous section. The unfolding software RooUnfold-1.0.3 is used to solve the problem. The unfolding settings should be adjusted to accommodate this particular problem.

7.6.1 Data Training

The event simulation is the key for our problem. The two causes we choose to unfold are the logarithmic primary energy and the particle mass. Since the simulation is designed to cover from 10^{14} to 10^{17} eV in primary energy and the number of simulated events decreases as E^{-1} in that range, the number of events with $\log_{10}(\mathrm{E_{prim}}/\mathrm{GeV})$ is almost evenly distributed from 5 to 8. The particle type is represented by five integers 0, 1, 2, 3 and 4 for proton, helium, oxygen, silicon and iron respectively [8]. If the energy dimension is divided into 15 bins, there are about 63000 to 67000 events in each bin as shown in Fig. 7.11.

The observations are also measured in two dimensions: the reconstructed primary energy and muon-bundle energy at in-ice COG in logarithmic scale with 20×20 bins [9]. An example of the probability distribution of a certain cause is shown in Fig. 7.12. The job of unfolding is to bring the measurement Fig. 7.13 back to Fig. 7.11 based on the knowledge about the PDF learned from the simulation. Note that the total number of events in Fig. 7.13 is far less than that in Fig. 7.11 since the majority of simulated events cannot trigger the detector or pass the various cuts, which will be accounted for by detection efficiency ϵ in unfolding.

To test how well the unfolding method works, the simulation is divided randomly into two halves. One for training, the other testing. The result is shown in Fig. 7.14. The first problem is the failure of the first two energy bins. This is

---

[8] This is to make the input data type match the RooUnfold interfaces. The values of those integers have no meaning except they distinguish primary particles with different types and make sure those with the same type fall in the same type bin in the unfolding algorithm.

[9] The binning of the measurement should be equal or finer than that of the truth.

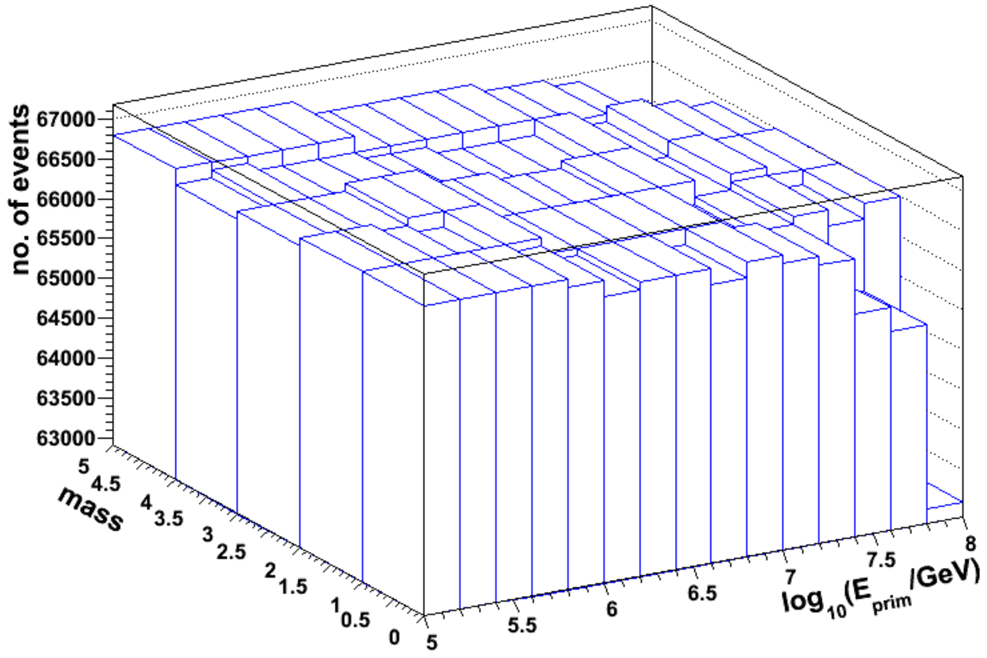


Figure 7.11: The two true causes of simulation in unfolding: $\log_{10}(E_{\mathrm{prim}}/\mathrm{GeV})$ and primary particle type. The number of events in each bin can be considered roughly equal.

actually due to the fact that there are basically no events from the first two energy bins that can trigger the detector and pass the cuts in simulation. Fig. 7.15 is the distribution of those events that caused Fig. 7.13, which clearly shows there is the threshold in the measurement.

The second problem facing us is the deviation in the reconstruction result from the truth. Table 7.5 gives the result and statistics for proton reconstruction. Many bins have a high deficit or excess to the truth compared to the errors that are typically several percent of the truth. Unfortunately, there is no easy way to improve the results. Increasing the number of iterations would make it worse since Bayesian unfolding tends to fluctuate greatly around the truth after certain number of iterations. Four-iteration performs the best in this test. If the same half of the

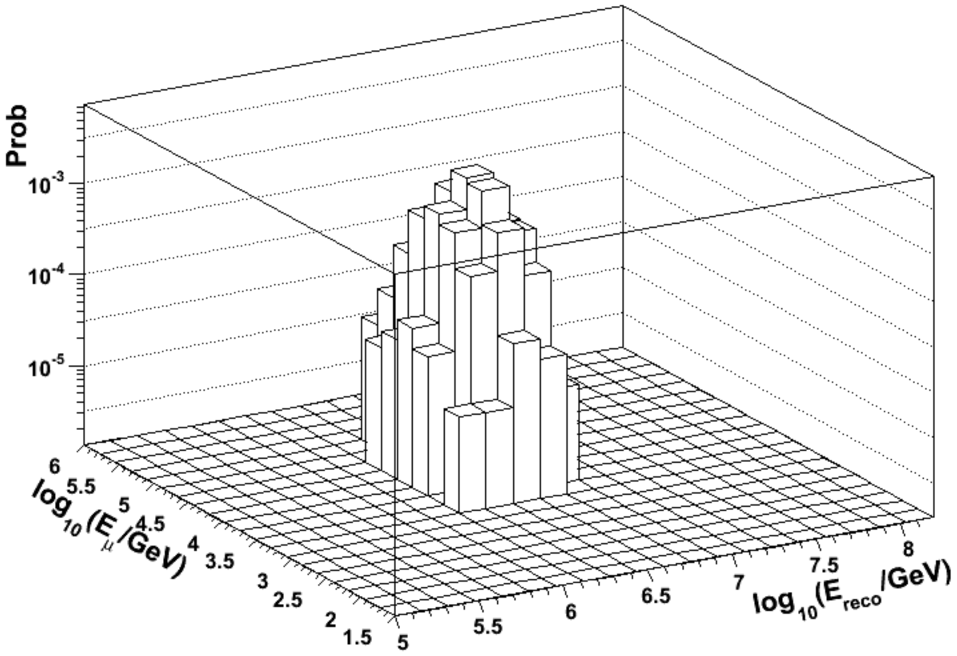


Figure 7.12: The probability distribution of the measurement from a certain cause that are events in the central bin in Fig.7.11 (Oxygen cosmic rays with $6.4 \leq \log_{10}(E_{\mathrm{prim}}/GeV) < 6.6$)

.

simulation is used for training and testing, a very good result can be obtained, which can be seen in Fig. 7.16. The errors in this case just get marginally better than that in Table 7.5 but the deviations from the truth is exactly zero, which indicates the unfolding is already at its best performance. Thus, the difference between the reconstruction and the truth when using two different sets of data for training and unfolding must reflect the difference between the two data sets [10]. The unfolding algorithm does not see the two sets the same, which is a result of the strong fluctuation and poor statistics in the simulated air showers. Tuning on the iteration

---

[10] An interesting thing happened when the training and testing data sets are switched, the deficit and excess bins seemed flipped.

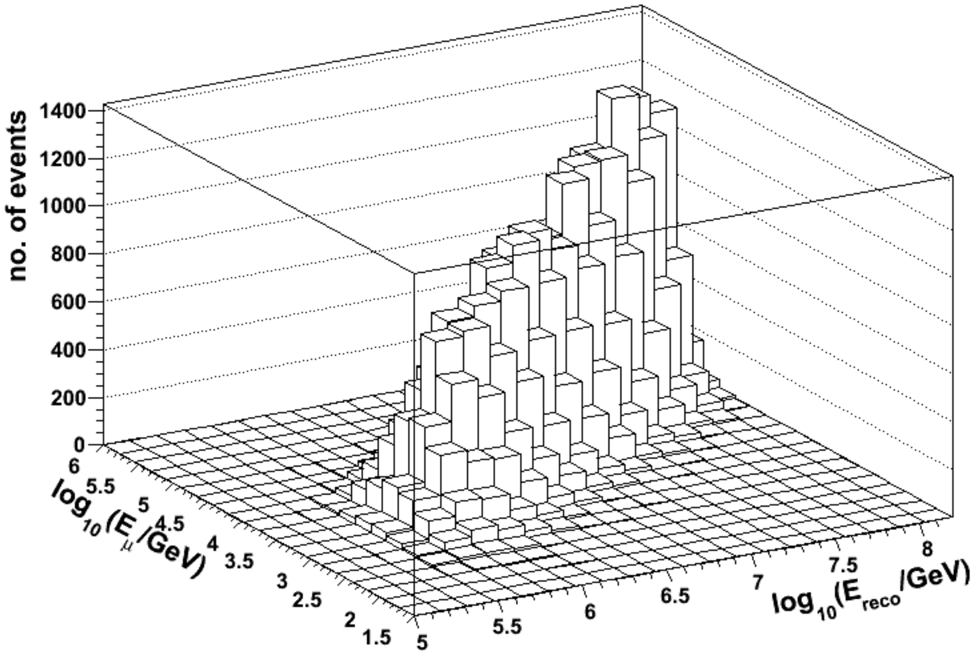


Figure 7.13: The two true observables of simulation in unfolding: $\log_{10}(E_{\mathrm{reco}}/\mathrm{GeV})$ and $\log_{10}(E_\mu/\mathrm{GeV})$.

numbers could magnify or reduce the difference. Discussion on the statistics will continue in the following sections.

7.6.2 Event Weighting

The weighting technique is often used in MC simulation. It assigns a particular weight to each simulated event with respect to its energy in order to generate any energy spectrum that is needed. No weighting is needed if the true cosmic ray spectrum does not differ too much from the simulated one since it does not add any new information in principle. However, this is not the case here. Weighting (other than 1) in unfolding changes two things in RooUnfold:

- the initial probability distribution of the causes $P_0(C_i)$ in Eq. 7.6,

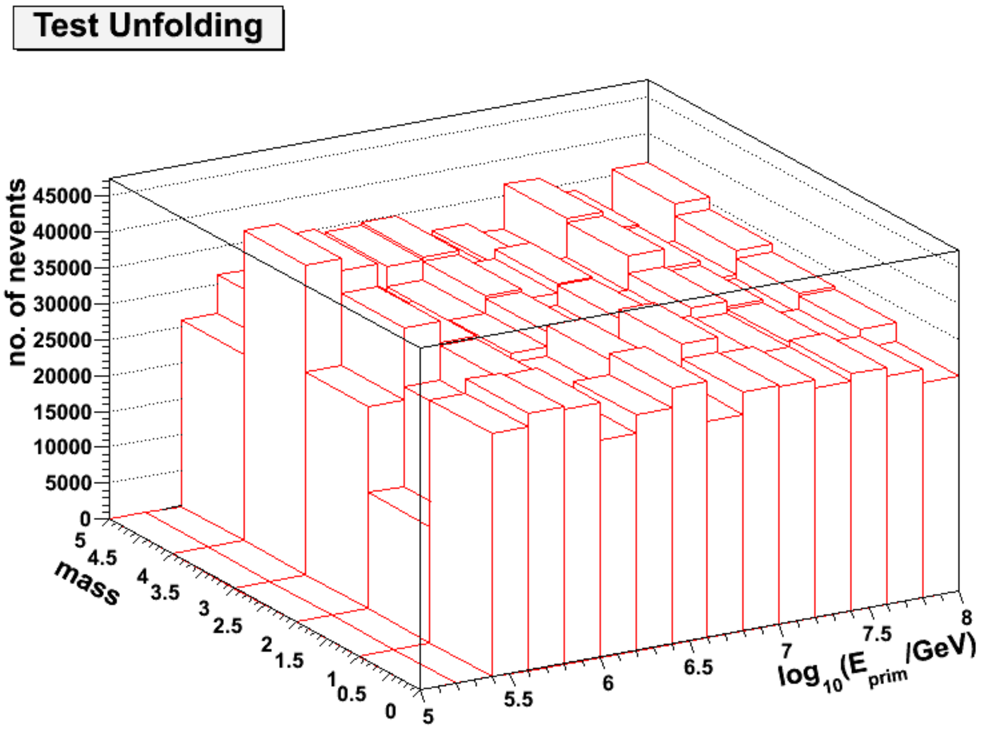


Figure 7.14: Test unfolding result after 4 iterations. The training and testing use different halves of the simulation data.

- and the energy distribution of events within each bin [11].

So weighting the simulation as close to the true spectrum as possible should improve the results in theory. However, complicated weighting schemes make ROOT tree reading very slow, simple weightings between $E^{-1.7}$ and $E^{-2.1}$ can be used. The difference in results by using different weightings will be attributed to the systematics. Weighting the simulation also has side effects. It complicates the unfolding procedure and the computed errors.

---

[11] This effect can also be achieved by using very small bin size without weighting.

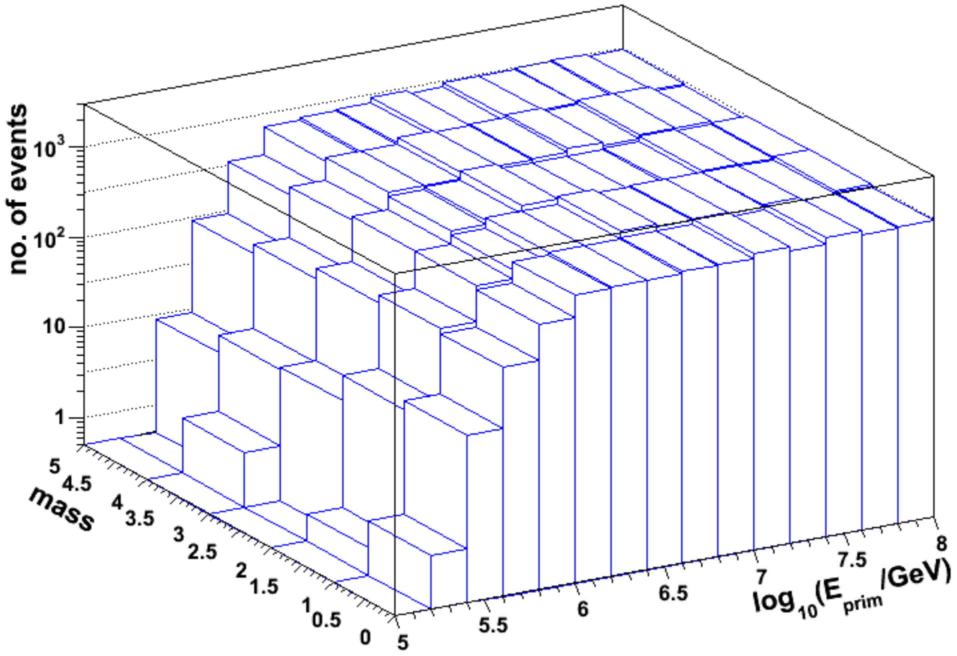


Figure 7.15: The distribution of events that are causes to those in Fig. 7.13. The are almost no events from the first two energy bins, which should be the reason that the reconstruction of the first two bins failed in our test.

7.7 Unfolding Results

Unfolding the experimental measurement is very similar to the test we did on the simulation. Data from June 21 to July 11 [12], 2009 are selected to make the test. Each simulated events is weighted by $E^{-1.7}$ and all events in Fig. 7.11 are used.

7.7.1 Unfolding Output

The unfolding output after 4 iterations is shown in Fig. 7.17 and Table 7.6.

The unfolding results are supposed to be the number of real cosmic rays that arrive in 21 days within a space that is exactly the same as the generation space in

---

[12] The simulation is generated with atmospheric model 12 in CORSIKA, which is supposed to be the atmosphere on July 1st, 1997.

Table 7.5: Test proton unfolding results

Bin	Train Truth	Test Truth	Unfolded Output	Error	Diff
3	33400	33400	33782.7	3247.6	382.7
4	33400	33400	35697.7	1509.2	2297.7
5	33350	33350	35374.0	973.5	2024.0
6	33250	33250	30230.3	698.1	-3019.7
7	33300	33300	32709.2	705.2	-590.8
8	33300	33300	35613.2	734.6	2313.2
9	33200	33200	30112.6	638.6	-3087.4
10	33300	33300	33122.5	630.4	-177.5
11	33250	33250	32954.6	672.8	-295.4
12	33200	33200	31989.0	648.9	-1211.0
13	32850	32850	32959.0	605.4	109.0
14	32700	32700	31723.4	571.9	-976.6
15	31550	31550	30011.0	558.4	-1539.0

Table 7.6: Test proton unfolding results and errors with and without a weighting of $E^{-\gamma}$.

Bin	N ($\gamma = 0$)	σ ($\gamma = 0$)	N ($\gamma = 1.7$)	σ ($\gamma = 1.7$)	N ($\gamma = 2.1$)	σ ($\gamma = 2.1$)
1	0	0	0	0.8	0	25.6
2	2370281.1	148724.2	3117571.1	195613.3	3328386.9	208839.5
3	2679600.8	25642.9	3095835.6	29408.4	3189972.7	30291.9
4	2889266.0	15247.6	2933396.2	15681.6	2944943.7	15805.3
5	2113921.2	8866.8	2059995.3	8674.5	2048139.1	8635.0
6	966753.9	4422.1	1017284.5	4526.5	1029594.4	4559.9
7	471480.2	2995.7	498076.9	3088.4	504326.0	3183.3
8	191463.0	1765.7	199087.4	1799.7	200780.1	1839.4
9	99040.9	1334.5	93551.5	1256.8	93070.4	1577.9
10	33183.1	733.8	34348.0	726.6	34617.2	1191.3
11	11594.5	432.3	11199.2	398.9	11181.0	2189.7
12	4565.6	242.1	4291.0	178.0	4239.8	3932.3
13	1624.8	131.5	1635.4	149.4	1630.5	5860.3
14	659.0	85.3	434.1	295.5	424.0	9702.4
15	245.9	44.8	222.7	448.1	217.2	15957.3

the simulation. The simulation is thrown to an disk area with a radius of 1200m and evenly distributed in a solid angle of $\cos\theta \in [0, \cos 40\,^\circ)$ and $\phi \in [0, 360\,^\circ)$,

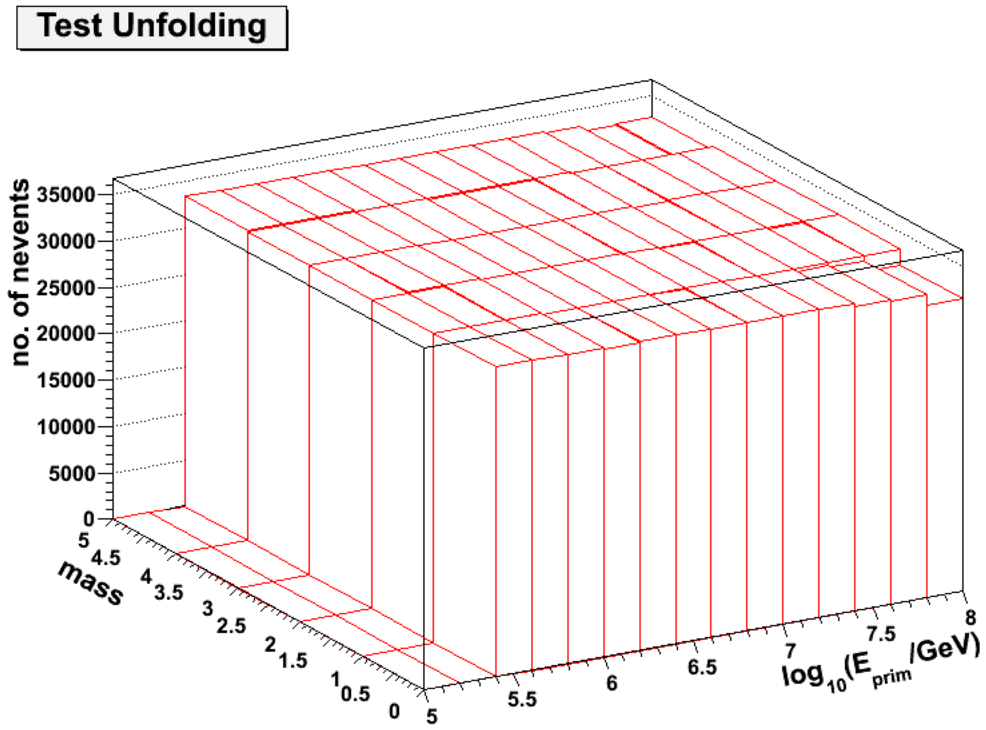


Figure 7.16: Test unfolding result after 4 iterations. The training and testing use the same half of the simulation data.

assuming the incoming cosmic rays are isotropic and only contain the five elements as in simulation.

7.7.2 Bootstrapping

Bootstrapping is a statistical method that estimates the accuracy of the samples by random sampling with replacement from the original dataset. With this method, the iteration number can be increased significantly to achieve a better convergence since multiple alternative datasets can be created to estimate the statistical mean and errors.

The simulated events with each element are grouped into 10 blocks randomly and marked as 0, 1, . . . , 9. In one trial, 10 blocks of data are selected randomly with replacement and unfolding is performed with those blocks. Repeat this procedure

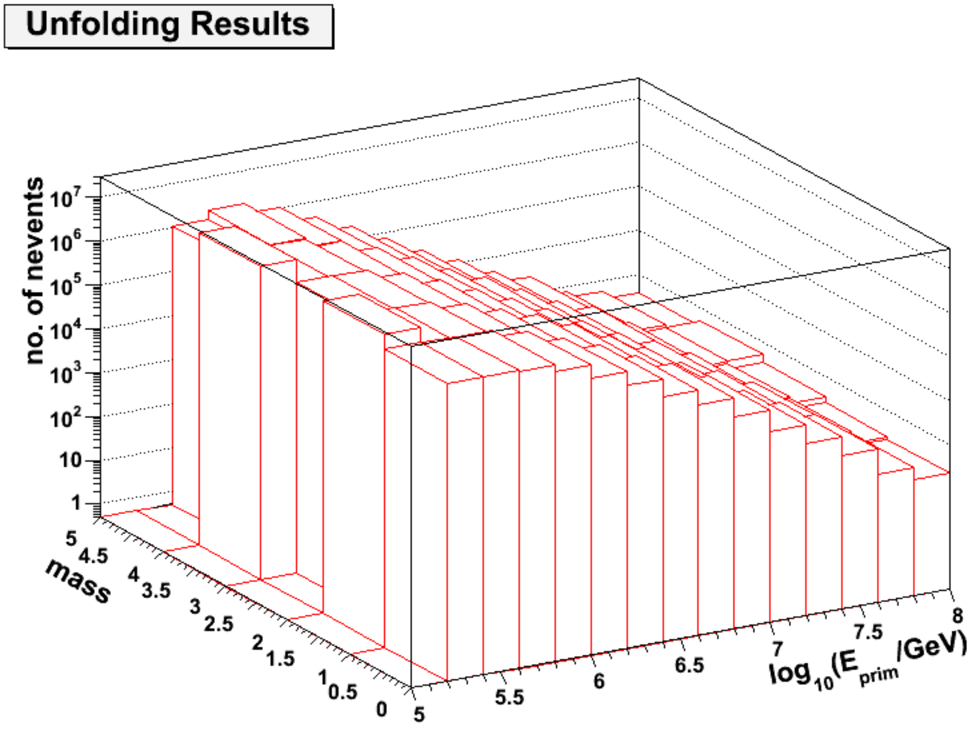


Figure 7.17: The unfolding result with $E^{-1.7}$ weighting in simulation.

for N time and the average of the N results will be the mean [13] $N = 20$ is considered large enough and the blocks used in each trial is listed in Table 7.7.

The unfolding results with bootstrapping are shown in Table 7.8 and Table 7.9. Note that the bins listed are from 3 to 15 because in some trials the first 2 bins cannot be reconstructed, which makes it impossible to calculate the mean and standard deviation.

---

[13] In each resampling, the same block may be used more than once and some blocks not at all, as can be seen by inspecting one of the rows of Table 7.7. Each trial always use 10 blocks.

Table 7.7: Blocks of data selected in each of the 20 trials

Trial	Blocks										Trial	Blocks									
1	4	5	0	6	2	8	4	1	8	6	2	3	1	9	8	7	6	5	5	4	5
3	5	6	7	8	9	1	3	6	8	1	4	4	8	1	5	0	4	9	4	0	5
5	1	8	4	1	8	5	3	1	9	7	6	6	5	4	4	4	4	4	5	6	7
7	9	0	2	5	7	0	3	7	1	5	8	9	3	8	3	9	4	0	7	3	0
9	7	4	2	0	8	6	5	4	3	3	10	3	3	3	4	5	6	7	9	1	3
11	6	9	2	5	9	3	7	2	7	2	12	7	3	9	5	1	8	5	2	0	8
13	6	4	3	2	1	1	1	1	1	2	14	3	4	5	7	9	1	4	7	0	3
15	7	1	5	0	4	9	5	0	6	3	16	9	6	3	0	8	5	4	2	1	0
17	9	8	8	8	8	9	0	1	3	4	18	6	9	1	4	7	0	4	8	2	7
19	2	7	2	8	3	0	6	3	0	7	20	5	2	0	9	7	6	6	5	5	5

Table 7.8: Bootstrapping Unfolding Results (Mean)

Bin	H	He	O	Si	Fe	Total
3	2.60×10^6	2.93×10^6	2.58×10^6	1.66×10^6	2.33×10^6	12.1×10^6
4	4.32×10^6	1.82×10^6	0.85×10^6	1.36×10^6	1.59×10^6	9.94×10^6
5	2.86×10^6	1.49×10^6	0.81×10^9	0.87×10^6	1.31×10^6	7.34×10^6
6	1.14×10^6	0.82×10^6	0.51×10^6	0.59×10^6	0.25×10^6	3.32×10^6
7	503145	557799	280927	183970	63928.4	1.59×10^6
8	177977	250468	96900	85332	28479	639156
9	118716	97561	25476	36698	6179	284630
10	40223	28340	15600	18283	5000	107445
11	9982	14347	7109	4821	506	36765
12	2545	4908	4359	3897	943	16653
13	2166	2342	1200	705	828	6496
14	65	262	649	679	777	2430
15	218	208	69	879	254	1629

7.7.3 Cosmic Ray Flux

The unfolding results obtained are the number of events in area A, solid angle Ω and time period T with respect to the primary energy. The cosmic ray flux can

Table 7.9: Bootstrapping Unfolding Results ($\frac{\sigma}{N} \times 100\%$)

Bin	H	He	O	Si	Fe	Total
3	53.4	43.0	115.1	65.2	147.7	41.6
4	18.1	41.0	58.3	62.7	61.0	17.6
5	24.2	42.5	67.0	54.5	47.8	18.3
6	19.4	51.5	27.3	22.1	27.5	15.6
7	21.1	24.4	48.4	48.4	55.6	15.1
8	31.7	20.7	54.3	38.7	110.3	16.2
9	25.3	22.5	32.5	39.7	29.6	14.4
10	17.7	26.4	37.8	14.4	40.9	11.5
11	37.0	30.4	42.5	38.9	66.2	18.3
12	62.6	42.8	55.9	43.2	71.3	24.1
13	36.1	39.1	68.7	50.5	130.6	23.2
14	118.3	76.3	87.8	85.5	64.2	40.2
15	59.4	108.3	240.5	63.6	129.7	44.1

be calculated by using the unfolding results as

$$\frac{d\phi}{d\ln E} = \frac{N_{\text{events}}}{\Delta(\log_{10} E) \times A \times \Omega \times T} \times \text{Prescale} \times \frac{1}{\ln 10}, \tag{7.9}$$

where the energy bin size $\Delta(\log_{10} E) = 0.2$, area $A = \pi \times 1200^2 \mathrm{m}^2$, $\Omega = 2\pi(1 - \cos 40\,^{\circ}$, time $T = 21 \times 24 \times 3600$ seconds and prescale is 3.

7.7.4 Average Mass

Fig. 7.19 shows the average mass of cosmic rays, which is calculated as

$$< \ln A >= \frac{\phi_H}{\phi}\ln(1) + \frac{\phi_{He}}{\phi}\ln(4) + \frac{\phi_O}{\phi}\ln(16) + \frac{\phi_{Si}}{\phi}\ln(28) + \frac{\phi_{Fe}}{\phi}\ln(56), \tag{7.10}$$

where ϕ is the total flux, ϕ_i is the flux of component i and $i \in \{\mathrm{H, He, O, Si, Fe}\}$.

7.8 Estimation of Uncertainties

The cosmic ray spectrum and composition obtained by using Bayesian unfolding methods are subject to many uncertainties that can be roughly categorized

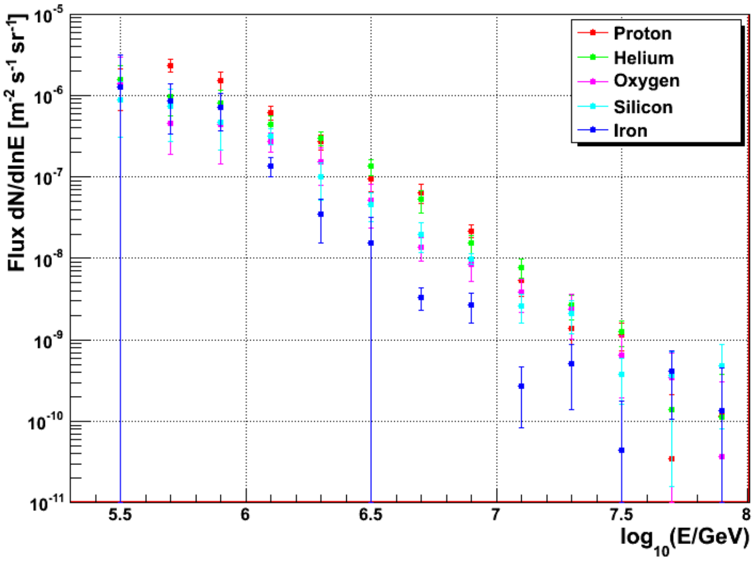


Figure 7.18: Cosmic ray flux calculated from the unfolding results for five different elements. The error bars are the square root of the sum of unfolding and bootstrap variances for each bin.

into to two groups, the statistics and systematics. The statistics are the uncertainties in the unfolding results due to the limited simulation and operation time of the experiment while the systematics arise from the assumptions made about the experiment, which include the atmospheric pressures, temperatures, hadronic models, etc.

7.8.1 Statistics

The statistical uncertainties come from the covariance matrix of $\hat{n}(C_i)$ in Eq. 7.6 and the bootstrapping. The bootstrapping uncertainties can be found in Table 7.9 and will not be repeated here [14]. The two components of the covariance matrix are calculated based on Agostini, 1994 [79]:

---

[14] Note if only a few iterations are used in the unfolding, the uncertainties from the bootstrapping may seem very small. However, a greater residual would occur between the mean of the bootstrapping results and the truth.

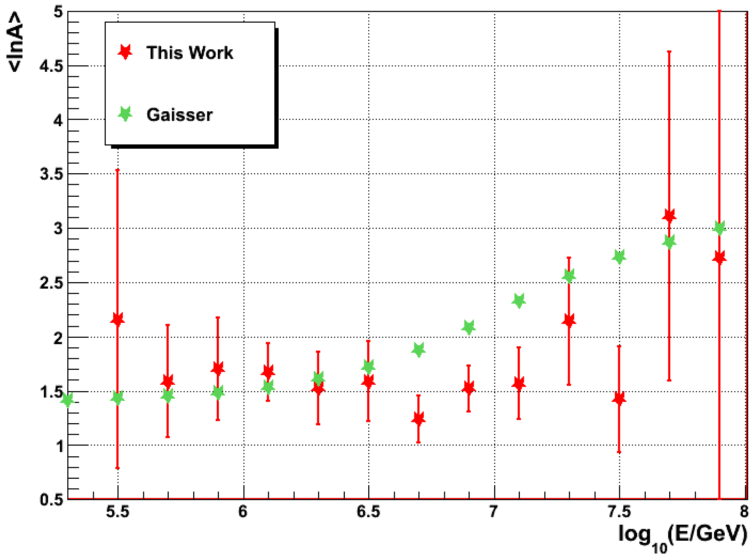


Figure 7.19: The average mass of cosmic rays compared with Tom Gaisser's three-population model (H3a) in [78]. The errors are calculated by propagating the errors in Fig. 7.18 assuming the total flux is constant and no correlations between different elements.

Table 7.10: Errors from unfolding covariance ($\frac{\sigma_{unfolding}}{N} \times 100\%$)

Bin	H	He	O	Si	Fe	Total
3	1.3	1.7	3.0	2.9	3.3	1.1
4	0.5	0.6	0.6	0.7	0.9	0.3
5	0.4	0.4	0.5	0.5	0.5	0.2
6	0.4	0.5	0.4	0.5	0.5	0.2
7	0.6	0.6	0.7	0.7	0.8	0.3
8	0.9	0.8	1.0	1.1	1.5	0.4
9	1.3	1.3	1.4	1.5	1.9	0.7
10	2.1	2.2	2.4	2.4	2.7	1.1
11	3.7	3.5	3.9	3.9	20.0	1.9
12	6.0	5.0	5.7	5.4	14.0	2.7
13	7.7	7.6	15.5	27.8	267.4	6.5
14	510.8	127.5	46.7	42.8	37.7	28.5
15	217.5	222.6	687.9	52.9	186.9	64.6

- contribution from true number of events per bin

$$V_{kl}(n(E)) = \sum_{j=1}^{n_E} M_{kj} M_{lj} n(E_j) \left(1 - \frac{n(E_j)}{\hat{N}_{true}}\right) - \sum_{\substack{i,j=1 \\ i \neq j}}^{n_E} M_{ki} M_{lj} \frac{n(E_i) n(E_j)}{\hat{N}_{true}},$$

  where M_{ij}

$$M_{ij} = \frac{P(E_j|C_i)P(C_i)}{[\sum_{l=1}^{n_E} P(E_l|C_i)][\sum_{l=1}^{n_C} P(E_j|C_l)P(C_l)]}$$

  can be seen as terms of the unfolding matrix.

- and contribution from the unfolding matrix

$$V_{kl}(M) = \sum_{i,j=1}^{n_E} n(E_i) n(E_j) \mathrm{Cov}(M_{ki}, M_{lj}).$$

Table 7.10 gives the total square root of covariances in energy bins from 3 to 15 in the unfolding results with 100 iterations. The sum of bootstrapping and unfolding uncertainties gives us the total statistical errors.

7.8.2 Systematics

- The seasonal variations in the experimental environments should not be a concern since 21 days is a very short period of time in a year. However, if a whole year's data is used with the current simulation, the seasonal changes should be taken into account since the season in simulation is fixed. The seasonal variations include variations in temperature, pressure of the atmosphere, snow depth and any other things that change with the seasons in a year. One way to estimate the seasonal variations is doing a bootstrapping with the experimental data of a year (from May, 2009 to June, 2010). It can also be done by comparing the data at two extreme seasons in a year. Using the same procedure, 21-day data from Dec. 21, 2009 to Jan. 11, 2010 are fed into the unfolding algorithm and the difference is shown in Table 7.11.

Table 7.11: Difference from the season ($\frac{N_{\mathrm{Jul.}} - N_{\mathrm{Jan.}}}{N_{\mathrm{Jul.}} + N_{\mathrm{Jan.}}} \times 100\%$)

Bin	H	He	O	Si	Fe	Total
3	24.0	29.2	21.9	37.1	16.3	24.8
4	28.4	16.4	2.0	7.7	2.1	15.8
5	22.4	12.6	-2.8	-9.6	-8.1	6.7
6	18.9	15.7	-0.8	-10.9	-40.4	1.4
7	7.2	24.4	-8.6	-18.4	-48.7	0.9
8	24.2	14.7	-4.2	-19.5	-42.3	3.4
9	7.4	34.0	-10.8	-10.3	-68.5	4.5
10	20.9	-1.7	-26.5	-0.5	-45.2	-3.1
11	57.9	23.8	-5.9	-32.7	-79.8	4.7
12	-3.3	-18.4	-11.7	13.7	20.4	-6.4
13	88.3	50.9	-30.5	-48.2	-81.9	4.0
14	99.1	73.1	23.8	-38.5	47.2	3.9
15	88.7	57.9	-51.7	39.3	-17.6	22.5

- Even if only 21 days data are used, errors from the snow depth of the IceTop tanks need to be taken care of. This is because the snow depth used in simulation might not be measured on the same of day of the atmospheric model (July 1st). Experimental data from May 20 to May 30, 2009 and experimental data from May 20 to May 30, 2010 are selected to estimate the systematics due to snow height. The results shown in Table 7.12 indicate that the increase in snow height from year 2009 to 2010 tends to lower the reconstruction results by about 10%. But a more significant effect is a heavier composition. This could be because the signals from the abundant low energy e^+e^- and γ are decreased by the snow.

- 1.6×10^5 QGSJet events are generated to estimate the systematics from different hadronic interaction models.

- The prior knowledge about the cosmic ray spectrum affects the weighting on

Table 7.12: Difference from the snow height $(\frac{N_{2009} - N_{2010}}{N_{2009} + N_{2010}} \times 100\%)$

Bin	H	He	O	Si	Fe	Total
3	28.6	-2.0	-22.7	36.9	7.4	6.0
4	10.6	4.1	2.6	15.2	1.7	7.6
5	0.3	7.7	-4.4	4.8	3.3	2.5
6	6.9	5.0	-3.6	-4.9	-6.1	1.5
7	10.5	-4.4	-2.4	-9.8	1.5	0.3
8	19.1	7.7	-15.7	1.8	-17.1	4.5
9	22.3	-4.1	-8.7	-19.9	-67.7	2.0
10	17.8	7.3	6.9	-4.9	-16.8	7.2
11	17.7	3.1	1.4	30.5	44.9	7.8
12	18.1	33.5	-60.3	-84.7	-94.6	1.2
13	-97.3	11.3	10.3	-54.6	7.5	3.3
14	100.0	98.7	-83.4	-92.6	-99.9	17.9
15	87.1	17.1	-50.0	61.2	-7.5	7.3

the simulated events. If weighting $E^{-2.1}$ instead of $E^{-1.7}$ is used, a slightly different result would occur. Table 7.13 shows the difference between the results from using $E^{-2.1}$ and $E^{-1.7}$ in percentage. The difference is so small that it can be neglected.

In summary, the most important uncertainty comes from the limited amount of simulation as revealed by the bootstrapping procedure (Table 7.7). This uncertainty is a reflection of of the uncertainty in simulated cosmic ray events. The big fluctuation in the results indicates the unfolding method is very sensitive to the cosmic ray fluctuation. In order to reduce this uncertainty, more simulation is needed. Suppose it is proportional to the inverse of the square root of the number of simulated events, $\sigma \sim \frac{1}{\sqrt{N}}$, which is not necessarily the case, we need about 100 times more simulation to constrain this uncertainty to less than 10%. Contrary to the statistical uncertainty, the uncertainty from the unfolding method itself assumes the simulation truly represents all the possible components of cosmic rays. Thus, this

Table 7.13: Difference from the weighting ($\frac{N_{1.7} - N_{2.1}}{N_{1.7} + N_{2.1}} \times 100\%$)

Bin	H	He	O	Si	Fe	Total
3	-0.0	0.4	-3.2	-0.5	-0.3	-0.7
4	-0.1	0.7	1.9	-0.1	-2.9	-0.2
5	-0.4	1.2	-1.5	-1.7	-0.2	-0.3
6	-0.7	-0.4	-1.6	-0.4	-0.5	-0.7
7	-0.9	-0.2	0.5	-1.2	-0.6	-0.4
8	0.1	-1.6	0.3	-1.1	-0.8	-0.7
9	-0.0	-1.0	1.4	-0.2	-0.6	-0.3
10	-1.9	-0.3	0.1	-1.1	-1.3	-1.0
11	0.3	-0.4	-1.5	-0.9	-0.3	-0.5
12	-0.6	-0.9	0.1	1.1	-0.6	-0.1
13	-0.8	-0.8	-3.8	-2.1	-3.2	-1.5
14	-7.9	0.3	5.1	1.3	-3.0	0.5
15	2.8	-5.9	-9.4	-1.0	0.9	-1.3

uncertainty is quite small for energy below 50 PeV. However, due to the steepness of the cosmic ray spectrum and few measured high energy events, the uncertainty is several times of the number of events at the high energy end. The increase in snow height during a year can decrease the cosmic ray flux by less than 10% generally. However, the reconstructed composition is sensitive to the snow height.

7.9 Conclusions

The above analysis solved the cosmic ray spectrum by using Bayesian unfolding method. A χ^2 test can give us a sense how well it works (see Fig. 7.20). The unfolding result in Table 7.8 is taken as an artificial cosmic-ray model and forward folding result based on it is compared against the real experimental observables. The average χ^2 for the 400 bins is 14.1 and it seems the fit is much better for the region above the knee. The real χ^2 should be smaller for two reasons. First, some cosmic ray information is lost during the unfolding process. Second, the energy

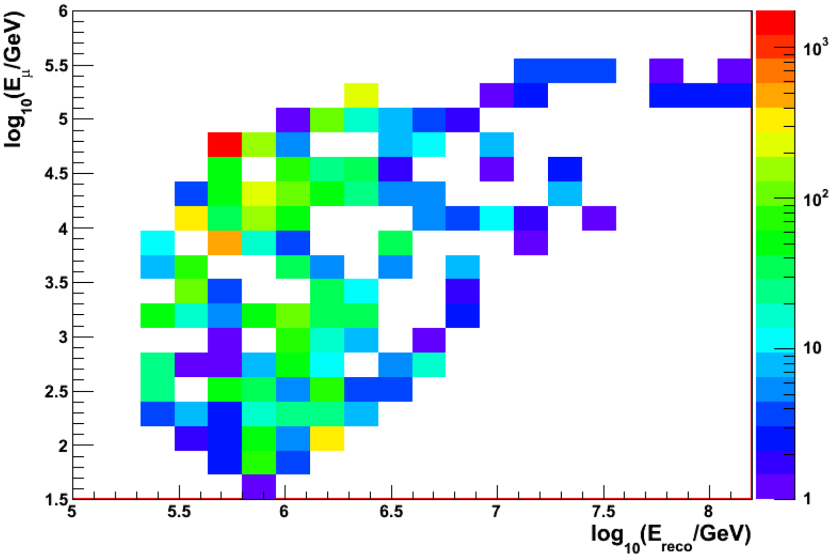


Figure 7.20: χ^2 test on the unfolding results. The χ^2 is computed as $(N_{\text{folded}} - N_{\text{exp}})^2/N_{\text{exp}}$.

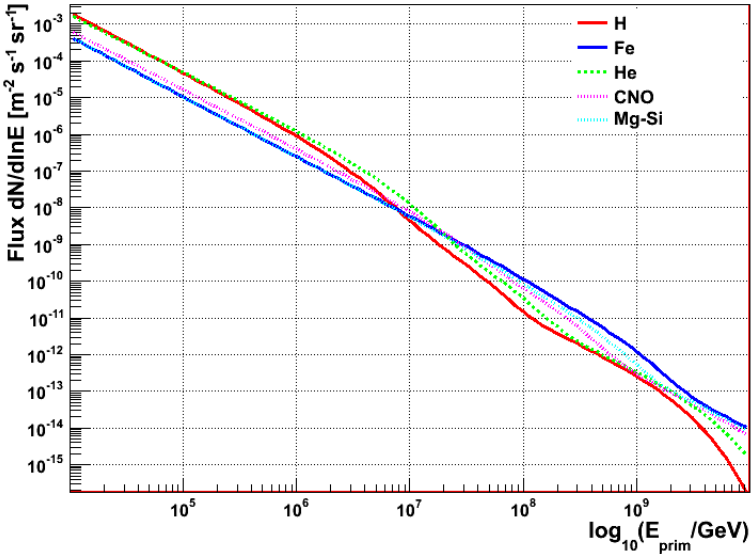


Figure 7.21: Tom Gaisser's H3a cosmic ray model [78].

distribution with a certain bin in simulation is E^{-1} rather than that of the real cosmic rays. The first problem can be reduced by decrease the number of bins in Fig. 7.20 while the second one can only be reduced by increase the number of bins . Compare the results in Fig. 7.18 and Tom Gaisser's H3a model in Fig. 7.21, little resemblance can be found between them with the iron flux not so high as predicted. However, the total flux of the five elements fits the model pretty well for primary energy above the knee. Fig. 7.19 shows that the structure in the heaviness is also different from that in the model. Instead of increasing with energy monotonically, the average mass in the unfolding results decreases first till the knee and then starts increasing again. But due to the low reconstruction quality for energy below the knee, whether this is a true feature of cosmic rays is inconclusive.

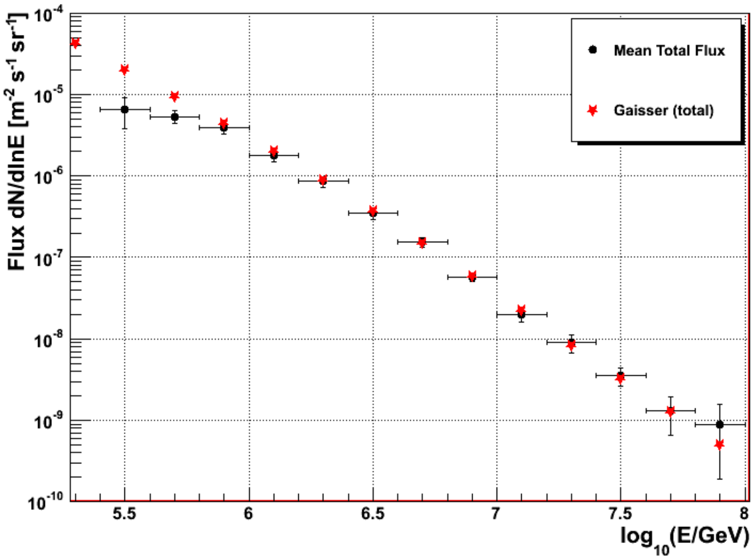


Figure 7.22: The total flux of the five elements compared with Tom Gaisser's H3a model in [78]. The error bars are statistic errors.

7.10 Smoothing the simulation

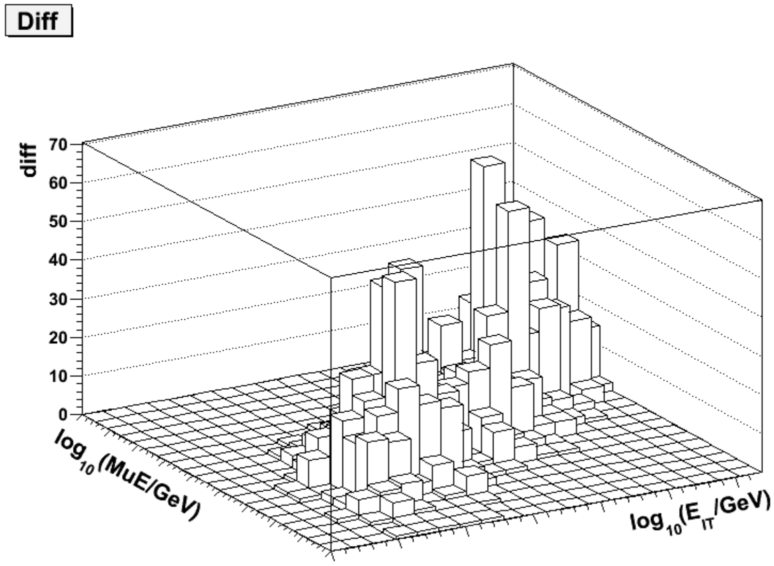


Figure 7.23: Difference in measurement between odd and even events of the simulation data sets $|N_{odd} - N_{even}|$.

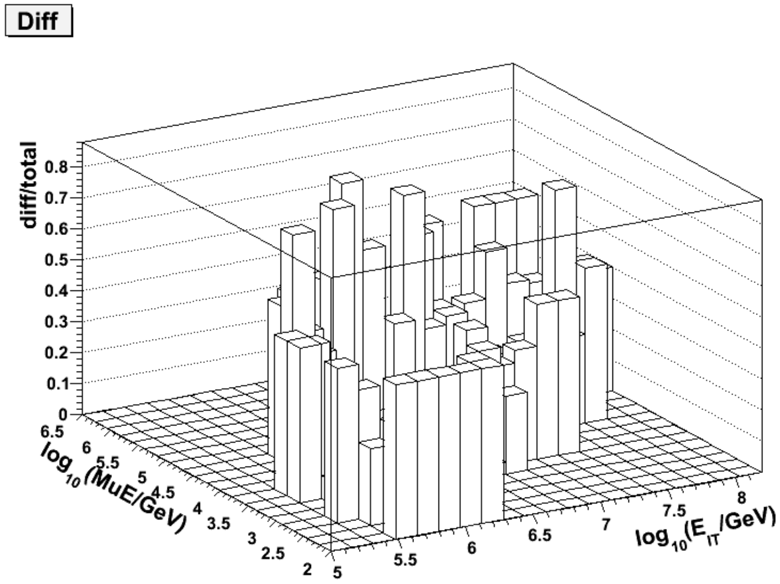


Figure 7.24: Difference in measurement between odd and even events of the simulation data sets as a fraction of the total events $\dfrac{|N_{odd} - N_{even}|}{N_{odd} + N_{even}}$.

The results obtained so far in this chapter suffer from large statistical fluctuations mainly from simulated events. In Fig. 7.12, about 1000 simulated events are distributed into about 400 bins in the E_{reco} and E_{μ} space. The lack of statistics in simulation can also be revealed when odd and even events in the simulation data sets are compared against each other (Fig. 7.23 and Fig. 7.24). One way to reduce the variation in the final results is to smooth the simulated data and use it as the input to the unfolding algorithm instead of the raw simulation. The infomation the unfolding algorithm needs to get from the simulation is $\mathbf{P}(E_{reco}, E_{\mu}|E_{prim}, A)$, which is a rugged distribution exemplified in Fig. 7.12. In smoothing $\mathbf{P}(E_{reco}, E_{\mu}|E_{prim}, A)$, a bivariate normal distribution is used to fit, and the probability distributions are described by five parameters:

$$\begin{aligned}
\bar{x} &: \text{ mean of } \log_{10}(E_{reco}/GeV) \\
\bar{y} &: \text{ mean of } \log_{10}(E_{\mu}/GeV) \\
\sigma_x &: \text{ standard deviation of } x \\
\sigma_y &: \text{ standard deviation of } y \\
\rho &: \text{ correlation of } x \text{ and } y,
\end{aligned}$$

which are dependent on E_{prim} and A. The PDF for a certain E_{prim} and A can be described by

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\bar{x})^2}{\sigma_x^2} + \frac{(y-\bar{y})^2}{\sigma_y^2} - \frac{2\rho(x-\bar{x})(y-\bar{y})}{\sigma_x\sigma_y}\right]\right). \tag{7.11}$$

The same simulation datasets used in previous sections are used to determine the parameters. The simulation is divided into 30 bins. Events with a primary energy between 1PeV and 100PeV are divided into 10 energy bins with respect to $\log_{10}(E_{prim}/GeV)$ and the bin size is 0.2. H, He, O, Si and Fe are grouped into three bins, where the first group is H, the second contains He and O, the last group contains Si and Fe. The simulated events within each bin are weighted by $(E_{prim}/GeV)^{-1.7}$ for $\log_{10}(E_{prim}/GeV) \leq 6.6$ and $(E_{prim}/GeV)^{-2.1}$ for $\log_{10}(E_{prim}/GeV) > 6.6$. An

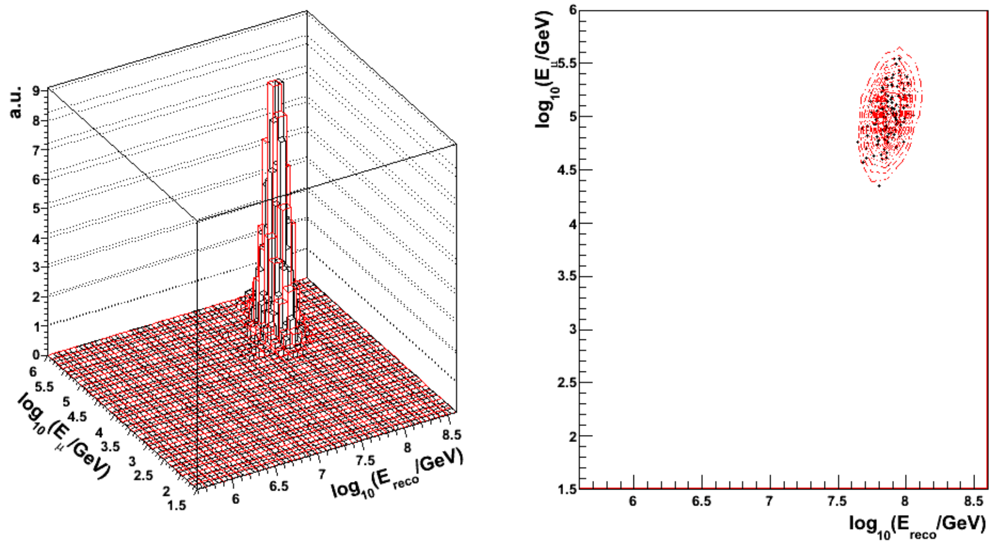


Figure 7.25: Fit result for Si and Fe events within the energy range $7.8 \leq \log_{10}(E_{prim}/GeV) < 8.0$.

example of fit result is shown in Fig. 7.25. The fit result is shown in Table 7.14 and the plots can be found in Appendix A.

Appendix B shows the testing result of the unfolding algorithm with hypothesized events that concentrate on a single or double bins. Here a test on a hypothesized simple model is presented to prove the unfolding is an effective method in solving the composition problem. In the simple model, all three groups of cosmic rays have a spectrum of $E^{-2.7}$ before their cutoff energies and get steeper to $E^{-4.0}$ after the cutoffs. The cutoff energies are 4PeV, 10PeV and 50PeV for the three groups. Fig. 7.26 shows the reconstruction results with 15 iterations. A parameter δ^2 is defined to measure how close the reconstruction is to the truth:

$$\delta^2 = \left(\frac{N_{reco}}{N_{true}} - 1\right)^2, \tag{7.12}$$

which decreases in value with respect to the number of iterations, as shown in Fig. 7.27. Thus we learn that 15 iterations is probably optimal since it balances between the convergence and variation in unfolding result.

Table 7.14: Fit results

Bin	$\bar{x}$	$\bar{y}$	σ_x	σ_y	ρ	Reco Rate
1	6.11822	3.41503	0.107191	0.352849	-0.112226	0.01141641
2	6.31493	3.54915	0.101863	0.321517	-0.013187	0.01199376
3	6.51020	3.78182	0.0941202	0.309182	-0.041147	0.01248741
4	6.69472	3.93835	0.0909438	0.296534	0.0014411	0.01337445
5	6.90238	4.12260	0.0943796	0.295009	0.170434	0.01304604
6	7.12387	4.29558	0.0777602	0.279030	0.199689	0.01455069
7	7.30410	4.44905	0.0923057	0.265015	0.251511	0.01336026
8	7.49610	4.67861	0.0841882	0.272298	0.165232	0.01599453
9	7.67377	4.80143	0.0900683	0.278537	0.361985	0.01472715
10	7.87085	5.01507	0.0843701	0.244534	0.370655	0.01567056
11	6.08959	3.61489	0.108380	0.305075	-0.128556	0.00996432
12	6.29385	3.77763	0.103433	0.297385	-0.088275	0.01186821
13	6.49480	3.94748	0.0973703	0.259038	-0.0448150	0.01379781
14	6.69913	4.14787	0.0901785	0.280339	-0.0166307	0.01437138
15	6.91789	4.31730	0.0904921	0.265551	-0.00821967	0.01350402
16	7.12715	4.49779	0.0814736	0.273319	0.0412998	0.01405467
17	7.32981	4.68761	0.0817637	0.270562	0.146772	0.01406277
18	7.53072	4.87570	0.0784333	0.249023	0.156111	0.01502586
19	7.71935	5.04344	0.0752287	0.253054	0.189423	0.01471362
20	7.91787	5.21719	0.0786094	0.234680	0.225711	0.01556358
21	6.04900	3.72741	0.109901	0.261216	-0.0405904	0.01115232
22	6.24894	3.94824	0.106943	0.269582	-0.0896851	0.01193100
23	6.47253	4.14114	0.102343	0.257298	-0.0640018	0.01300515
24	6.68187	4.31141	0.0989906	0.246594	-0.0742526	0.01398615
25	6.90478	4.49415	0.0917809	0.244484	0.0750787	0.01402143
26	7.13019	4.68350	0.0862470	0.239905	0.0604465	0.01509633
27	7.34303	4.86795	0.0806578	0.245941	0.0532772	0.01458402
28	7.54609	5.02940	0.0769462	0.234985	0.123673	0.01586544
29	7.74705	5.18407	0.0754297	0.230437	0.171810	0.01608435
30	7.94869	5.34762	0.0761249	0.215560	0.153517	0.01456326

Bin 1-10 are proton events, bin 11-20 are combined helium and oxygen events, bin 21-30 are combined silicon and iron event.

The unfolding method is applied to the 21-day data around July 1st, 2009 and the results are shown in Fig. 7.28 and Fig. 7.29.

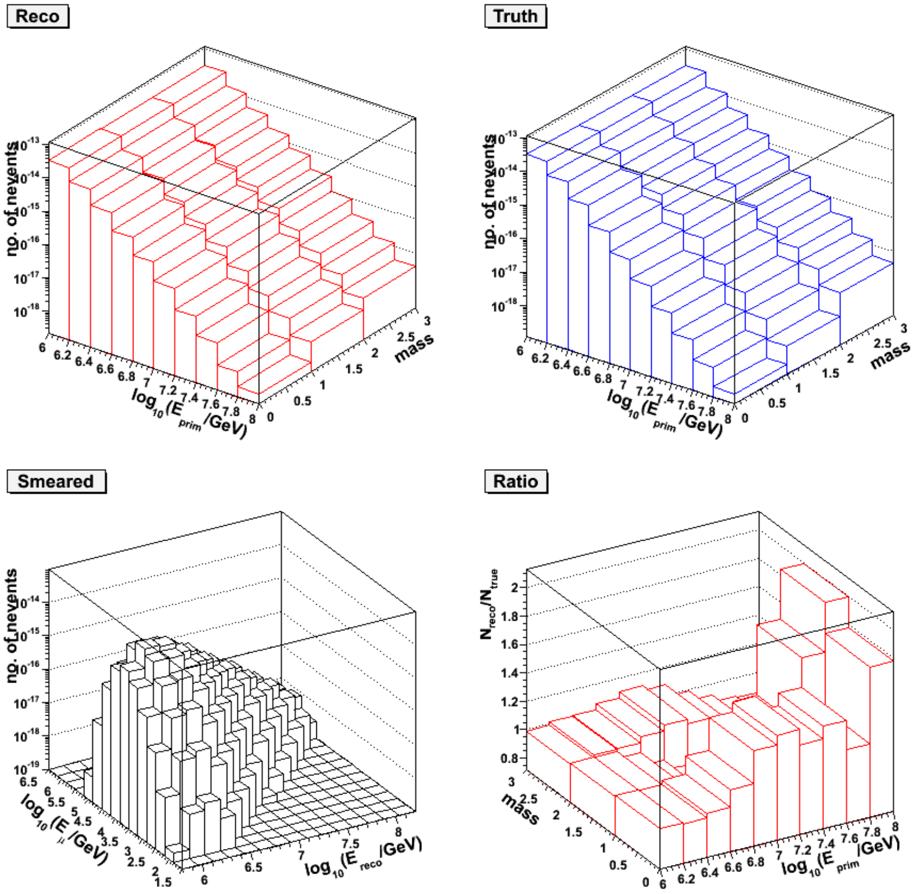


Figure 7.26: Left-up:reconstructed number of events per bin;right-up:true number of events per bin;left-down:observables that would be obtained in this artificial model; right-down:ratio of reconstruction and truth.

An χ^2 is performed to check the unfolded cosmic-ray spectrum (Fig. 7.30), which is computed as

$$\chi^2 = \frac{(N_{reco} - N_{exp})^2}{N_{exp} + 1}, \tag{7.13}$$

where N_{reco} is the forward folded [15] number events per bin in the measurement space and N_{exp} is the experimental measurement in the same bin. The average χ^2 is about

---

[15] A projection of CR spectrum from $(E_{prim}, mass)$ space to (E_{reco}, E_{μ}) space based on simulation.

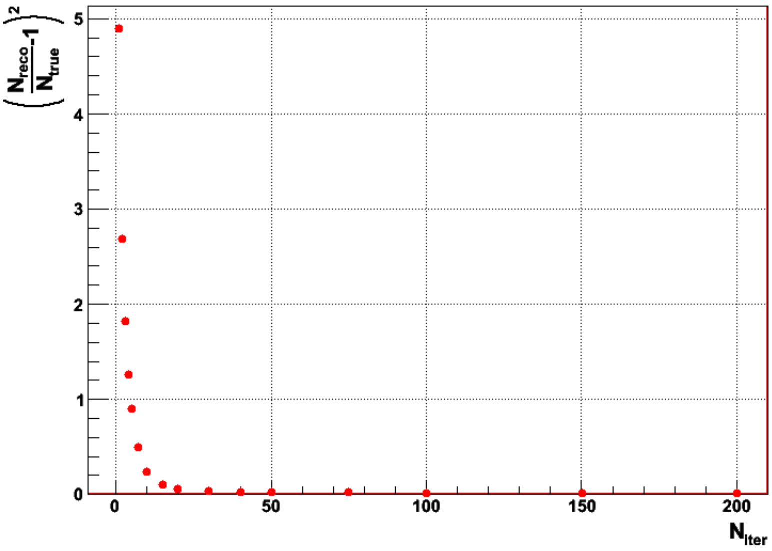


Figure 7.27: χ^2 versus number of iterations.

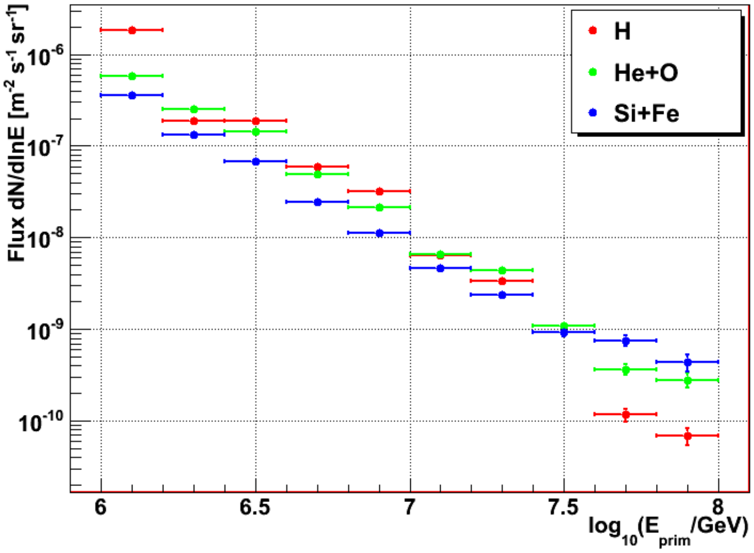


Figure 7.28: Unfolded primary cosmic-ray spectrum.

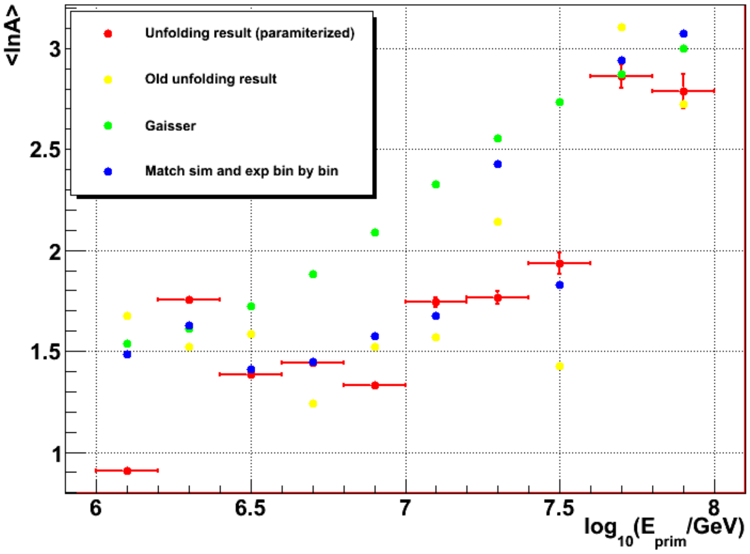


Figure 7.29: Unfolded primary cosmic-ray mass. The blue dots are results in Sec. 7.2.

20, which is considerably large. However, the lowest E_{reco} bin alone contributes to more than 60% of the total χ^2 and χ^2 diminishes as E_{reco} increases.

7.10.1 Uncertainties Revisited

Uncertainties in this analysis can be categorized into two groups: statistical uncertainties and systematic uncertainties. Statistical uncertainties come from the lack of statistics in simulated and experimental data. The former is essentially eliminated if parameterized simulation is used. The latter can be gained by using the bootstrapping method against the experimental data. This uncertainty increases from less than 5% to about 50% in flux, and from 1% to 10% in ln A. The systematic uncertainties are from the assumptions made in the analysis and discrepancies between simulation and experiments. Because of the difficulties in generating simulation, they are hard to estimate. The main systematics that are considered important in this analysis are listed as below

- Atmospheric model. The atmospheric model in simulation is July 1st, 1997.

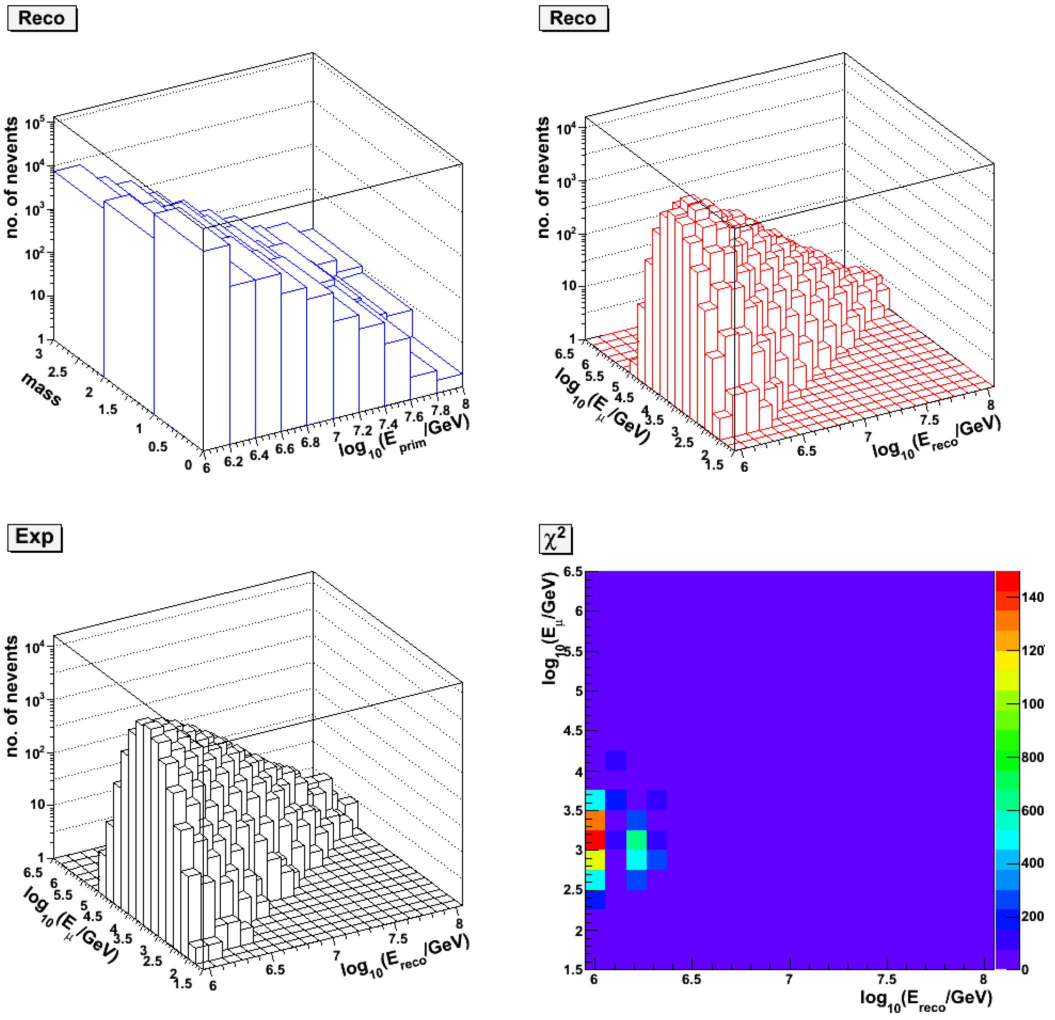


Figure 7.30: Left-up:reconstructed number of events per bin;right-up: forward folded detector measurement of the left spectrum;left-down:experimental measurement; right-down:χ^2.

The period of experimental data is from June 21th to July 11th, 2009. The atmospheric model error is ignored in this analysis because atmospheric conditions in Antarctic mid-winter are assumed to be similar year to year.

- Hadronic interaction model. A small sample of QGSJet-II simulation with energy between 1 to 50 PeV has been generated to estimate this error (The comparison is show in Appendix G. A forward folding approach is used to estimate his error.

- Ice property in simulation. This can be handled by generating simulation with different ice models, but has not been done at the time of the draft of this thesis.

- Different reconstructions. Results from MuE and muon-bundle-reco are compared by using a forward folding approach. Figures are shown in Appendix H.

- Snow height. The snow height above the tank suppresses the IceTop signals. Higher snow depth not only reduces the trigger rate but also makes the event appear heavier in mass. Tanks with different vintage years typically have different snow depth. Thus, detector response to cosmic rays is dependent on where they land at the surface. The difference in the trigger efficiency of different parts of the detector is shown in Fig. 7.31. This effect is simulated by using the snow height measured at the beginning of 2009 as shown Fig. 7.32.

  The difference in detector response among the different parts is shown in

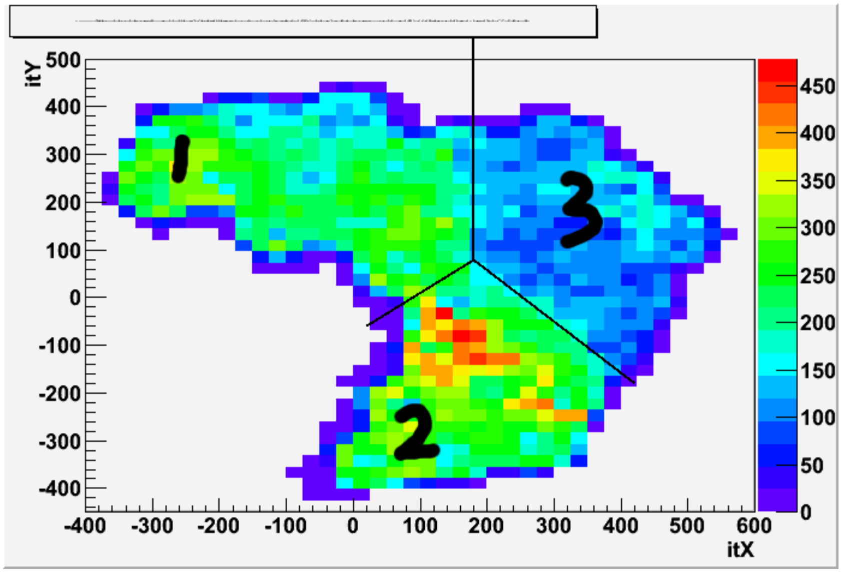


Figure 7.31: Trigger rate at different parts of the array in the experimental data.

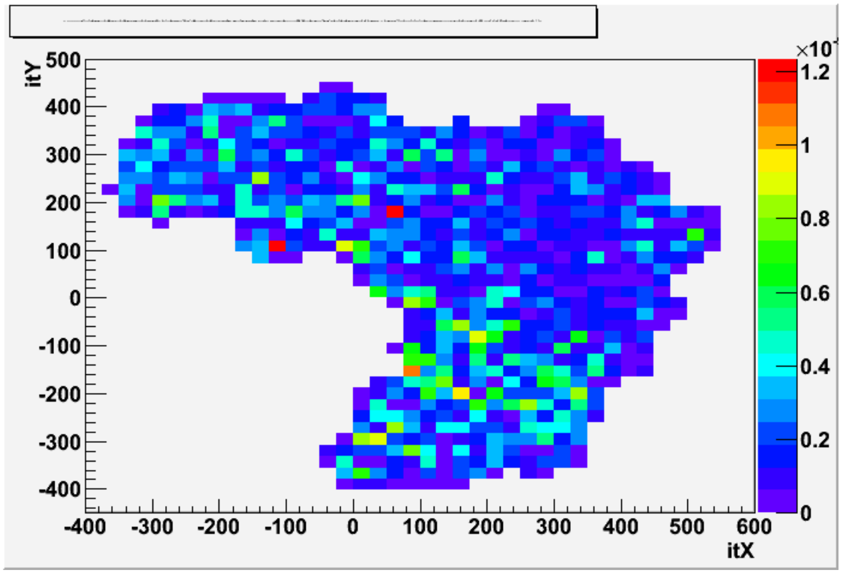


Figure 7.32: Trigger rate at different parts of the array in the simulated proton data.

Fig. 7.33. This effect can also be found in simulation as shown in Fig. 7.34. The effect of snow is estimated by comparing the region 1 and region 2 events in Fig. 7.31 with forward folding method [16].

- Pressure and temperature. The systematics due to the variation of pressure and temperature in the 21-day data is estimated by scaling down the difference between the July and January data.

Table 7.15: Pressure and Temperature

Var.	July (mean/std.)	January (mean/std.)	Scaler
Pressure	662.4/17.7	683.0/14.4	0.425
Temperature	-51.1/17.4	-24.2/4.7	0.295

---

[16] Many fewer events are in Region 3 than in Region 1 and 2.

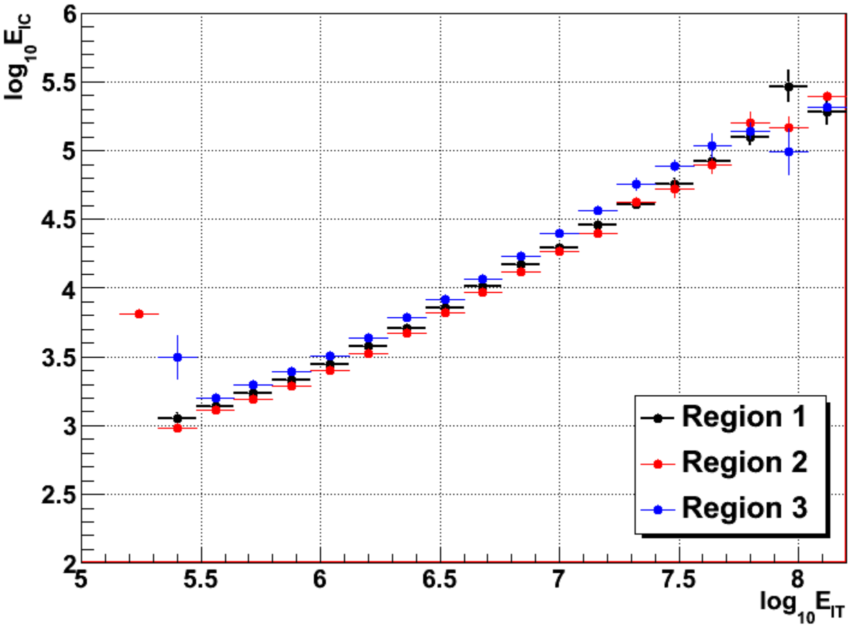


Figure 7.33: Response at different parts of the array in the experimental data.

$$\begin{aligned} \Delta &= S_{\text{Jan}} - S_{\text{July}} \\ \text{Scaler} &= \frac{\sigma^2_{\text{July}}}{\Delta^2 + \sigma^2_{\text{July}}} \end{aligned} \tag{7.14}$$

The scaling factor is computed by using Eq. 7.14. The effect of pressure and temperature can be obtained by multiplying the square root of the scaler to the difference of July and January data normalized to the mean of July and January data.

A summary of the systematics is shown in Table 7.16 where uncertainties are stated as percent of ln A. The uncertainty from the snow (Fig. 7.33 and Fig. 7.34) is computed by doing the experiment with two parts of the IC59 detector and comparing them. The hadronic interaction model effect is estimated by comparing SIBYLL and QGSJetII simulation. The difference of the two simulation can be found in Appendix G. Among the reconstruction algorithms that are describe in Section 6.2.3,

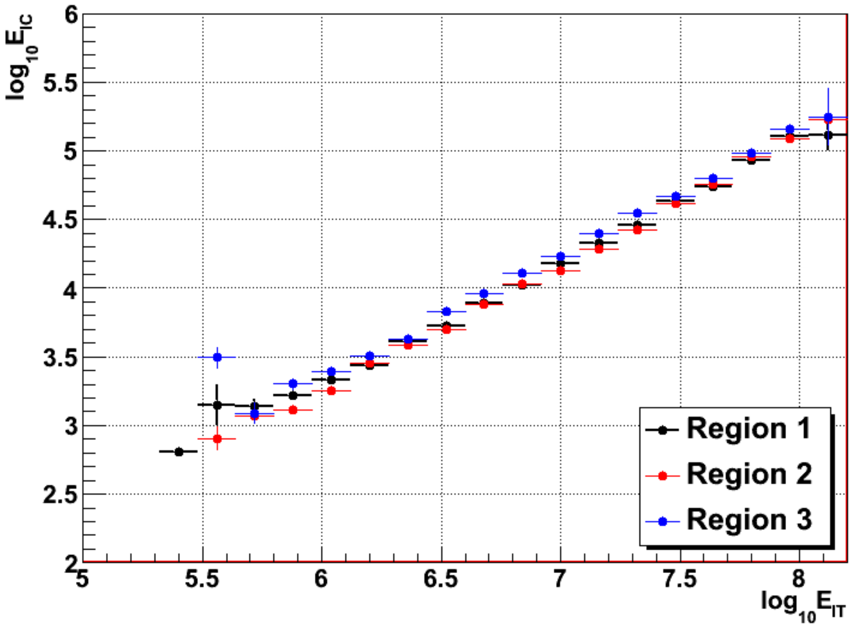


Figure 7.34: Response at different parts of the array in the simulated sim data.

Table 7.16: Systematic results

Snow Depth	10-20%
Hadr. Interaction	5-10%
Reconstruction	5-10%
Pressure/Temperature	10-25%

MuE and K70 are chosen to estimate the reconstruction uncertainty by comparing their the forward folding results. Uncertainty from the pressure fluctuation in 21-day data in July is computed by scaling down the difference between July and January data. The result remains when assuming the seasonal change is due to the change in temperature instead of the pressure [17].

---

[17] The square root of the two scalers $\sqrt{0.425}$ and $\sqrt{0.295}$ are close in Table 7.15.

Chapter 8

SUMMARY

The composition result suggests the mass is rather constant around the knee and jumps quickly after about 30 PeV. However, this result is subject to large uncertainties. The statistical uncertainty due to the low statistics in simulation is addressed by smoothing the data. The error introduced by smoothing is not considered in this thesis. The uncertainty will be reduced by producing more simulation or using more advanced simulation strategies in IceCube.

The estimation of systematic uncertainties only considers the main sources of systematics: snow height, pressure, temperature, hadronic interaction and energy reconstructions. The error from uncertainty in ice properties is not studied here because we have no time to generate simulation with other ice models. The systematics must be further studied in the future simulations in IceCube.

BIBLIOGRAPHY

[1] K.Nakamura, *et al.* (Particle Data Group), J.Phys. G **37**, 075021 (2010).

[2] T.K.Gaisser, *Cosmic Rays and Particle Physics*, (Cambridge, 1990), pp. 148-149.

[3] T.K.Gaisser, *Cosmic Rays and Particle Physics*, (Cambridge, 1990), pp. 157-160.

[4] T.Stanev, *High Energy Cosmic Rays*, (Springer, 2001), pp. 62-71.

[5] T.K.Gaisser, Univ. of Delaware PHYS638 Handout (Unpublished).

[6] T. Stanev, *High Energy Cosmic Ray*, (Springer, 2001), pp. 74-78.

[7] T.K.Gaisser, *Cosmic Rays and Particle Physics*, (Cambridge, 1990), pp. 8-10.

[8] J.A.Simpson, *Ann.Rev.Nucl.Part.Sci.*, **33**, 323-81 (1983).

[9] A.M.Hillas, *Annu.Rev.Astron.Astrophys*, **22**, 425-444 (1984).

[10] D.Allard *et al.*, *29th ICRC*, Pune (2005) **00**, 101-104.

[11] University of Utah, (2009) http://www.telescopearray.org

[12] Pierre Auger Observatory, (1999) http://www.auger.org/observatory/

[13] NASA, (1997), http://imagine.gsfc.nasa.gov/docs/science/how_l2/cerenkov.html

[14] KASCADE-Grande (2011).
http://www-ik.fzk.de/KASCADE/KASCADE_welcome_Grande.html

[15] D. Barnhill *et al*, *29th ICRC*, Pune (2005), **7**, 291-294.

[16] Pierre Auger Observatory, (1999) http://www.auger.org/observatory

[17] D.Chirkin, Muon Propagation Code in Java, (hep-ph/0407075v2, 2008).

[18] T.K.Gaisser, *Neutrino Astronomy: Physics Goals, Detector Parameters*, (Talk given at the OECD Megascience Forum Workshop, Taormina, Sicily, 22/23 May, 1997 arXiv:astro-ph/9707283).

[19] E.Waxman & J.Bahcall, *Phys.Rev.D* **59** 023002 (1999).

[20] The IceCube Collaboration. *IceCube Preliminary Design Document Revision*:1.24, (2001).

[21] M.Merck (2009)
http://wiki.icecube.wisc.edu/index.php/
Bootcamp09_Introduction_to_the_IceCube_Hardware

[22] C.Wendt (2007)
http://wiki.icecube.wisc.edu/index.php/LED_output_time_profile

[23] C.Wendt (2009)
http://wiki.icecube.wisc.edu/index.php/LED_light_output

[24] C.Song (2007)
http://wiki.icecube.wisc.edu/index.php/Interstring_Flasher_Documentation

[25] K.Woschnagg (2008)
http://wiki.icecube.wisc.edu/index.php/Flashers

[26] C. Wendt, *Multi-wavelength Flashers*, (IceCube Teleconference 2010-04-20, Unpublished)

[27] M.Merck (2009)
http://wiki.icecube.wisc.edu/index.php/
Bootcamp09_Introduction_to_the_IceCube_Hardware#The_PMT

[28] R.Abbasi *et al.*, *Nucl.Instrum.Meth.A*, **618**, 139-152 (2010).

[29] A.Lucke, Diploma thesis, DESY-Zeuthen (2008)
http://nuastro-zeuthen.desy.de/e13/e63159/e109/e520/e597/
infoboxContent752/adams-thesis.pdf

[30] R.Abbasi *et al.*, *Nucl.Instrum.Meth.A*, **601**, 294-316 (2009).

[31] NIST GPS Data Archive
http://www.nist.gov/physlab/div847/grp40/gpsarchive.cfm

[32] K.Hoshina, Hits, PEs and Waveforms (IceCube Bootcamp, Unpublished, 2008).

[33] S.Hickford (2009), http://wiki.icecube.wisc.edu/index.php/Local_Coincidence

[34] M.Beimforde, diploma thesis, DESY-Zeuthen (2006) http://nuastro-zeuthen.desy.de/e13/e63159/e109/e520/e597/infoboxContent684/MBeimforde_icecube_200702002_v1.pdf

[35] T.Kuwabara (2011) http://wiki.icecube.wisc.edu/index.php/IceTop_Snow_Height_Estimation_From_VEMCal

[36] J.A.Aguilar & M.Baker, *Data Processing*, (Bootcamp 2010, unpublished).

[37] 2010 DAQ/Trigger Planning, (IceCube internal document, 2009).

[38] E.blaufuss (2009) http://wiki.icecube.wisc.edu/index.php/2009_Online_Filter

[39] J.Diaz Velez, *IceCube Simulation*, (IceCube Bootcamp 2010).

[40] T.D.Straszheim *et al.* (1999) http://software.icecube.wisc.edu/offline-software.trunk/index.html

[41] K.Woschnagg (2010) http://wiki.icecube.wisc.edu/index.php/Surface_coordinates

[42] K.Woschnagg (2010) http://wiki.icecube.wisc.edu/index.php/Coordinate_system

[43] K.Woschnagg (2010) http://wiki.icecube.wisc.edu/index.php/Stage_2_Geometry

[44] J.Kelley & J.Braum, *DOM-Cal XML User's Guide, Version 7.4*, (IceCube internal document, 2009, unpublished).

[45] F.Halzen & S.R.Klein, *Rev.Sci.Instrum*, **81**, 081101 (2010).

[46] S.Eular, diploma thesis, DESY-Zeuthen (2008) http://www.physik.rwth-aachen.de/fileadmin/user_upload/www_physik/Institute/Inst_3B/Forschung/IceCube/publications/thesis_SE.pdf

[47] T.K.Gaisser, *Cosmic Rays and Particle Physics*, (Cambradge, 1990), pp. 75-77.

[48] T.K.Gaisser, Univ. of Delaware PHYS638 Handout (Unpublished)

[49] P.Askebjer *et al.*, *Geophys.Res.Lett.*, **24**, 1355-1358 (1997).

[50] J.T.O.Kirk, *Appl.Opt.*, **38**, 3134-3140 (1999).

[51] M.Ackermann *et al.*, *JGR*, **111**, D13203 (2006).

[52] L.G.Henyey & J.L. Greenstein, *Astrophys.J.*, **93**,70-83 (1941).

[53] P.Askebjer *et al.*, *Appl.Opt.*, **36**, 4168-4180 (1997).

[54] F.Urbach, *Phys.Rev.*, **92**, 1324 (1953).

[55] A.P.Minton, *J.Phys.Chem.*, **75**, 1162-1164 (1971).

[56] J.D.Lindberg & L.S. Laude, *Appl.Opt.*, **13**, 1923-1927 (1974).

[57] J.D.Lindberg, *Appl.Opt.*, **33**, 4314-4319 (1994).

[58] K.Woschanagg, 26th ICRC (Salt Lake City), **2**, 200 (1999).

[59] J. Lundberg *et al.*, *Physics Research A*, **581**, 619-631 (2007).

[60] K.Woschanagg (2007)
http://wiki.icecube.wisc.edu/index.php/Aha

[61] D.Chirkin, Muon Propagation Code (Unpublished)
http://icecube.wisc.edu/ dima/work/WISC/ppc

[62] D.Chirkin, *Study of ice transparency with IceCube flashers*, (Internal report, 2009, icecube/200911002, unpublished).

[63] D.Chrikin (2009)
http://icecube.wisc.edu/ dima/work/WISC/ppc/fit

[64] P.Desiati (2009)
http://wiki.icecube.wisc.edu/index.php/Photonics_vs_PPC

[65] K.Hoshina (2007)
http://wiki.icecube.wisc.edu/index.php/
DOM_Acceptance_Table_for_Photonics_and_Romeo

[66] G.Hill *et al.*, *Tracking the wild groose:further forays into the disagreement between data and simulation*, (Internal document, unpublished).

[67] K.Woschanagg (2007)
http://wiki.icecube.wisc.edu/index.php/Hole_ice

[68] M.D.Agostino, *Freeze-In and Hole Ice Studies with Flashers*, (Internal document, 2006, unpublished).

[69] L.Gerhardt,*Effects of Hole Ice on Hit Time*, (Internal report, 2009, icecube/200904001, unpublished).

[70] Stefan Klepser, Ph.D thesis (2008)
http://nuastro-zeuthen.desy.de/e13/e63159/e109/e517/e588/infoboxContent686/Klepser_icecube_200807002.pdf

[71] Fabian Kislat, diploma thesis (2007)
http://nuastro-zeuthen.desy.de/e13/e63159/e109/e520/e597/infoboxContent685/diplomarbeit.pdf

[72] D.Chirkin, *Feature Extraction of IceCube Waveforms*, (Internal document, 2008, unpublished).

[73] N.Eijndhoven *et al.*, *Astroparticle Physics*, **28**, 4-5, 456-462 (2007)

[74] C.Xu (2009) http://wiki.icecube.wisc.edu/index.php/InIce_Inferred_Surface_Shower_COG

[75] D.Chirkin, *Neutrino Search with IceCube*, (Internal report, 2008, icecube/200807006, unpublished).

[76] S.Grullon, D.J. Boersma & G.Hill, *Photonics-based log-likelihood Reconstruction in IceCube*, (Internal report, icecube/200807001, 2008, unpublished).

[77] T.K.Gaisser, *Cosmic Rays and Particle Physics*, (Cambridge, 1990), pp. 205-207.

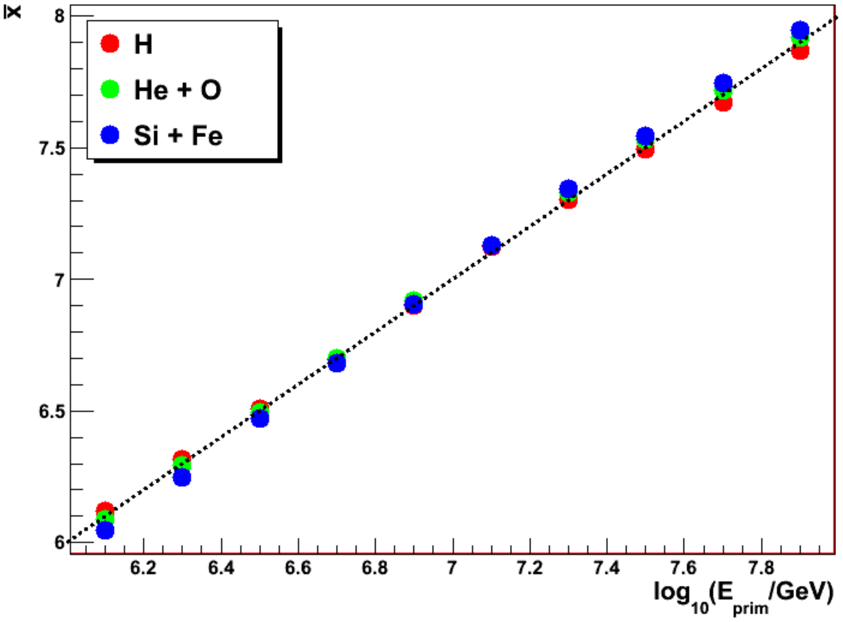
[78] T.K.Gaisser, *Cosmic ray spectrum and composition* $>$ *100TeV*, (Internal report, icecube/201102004, 2011, unpublished).

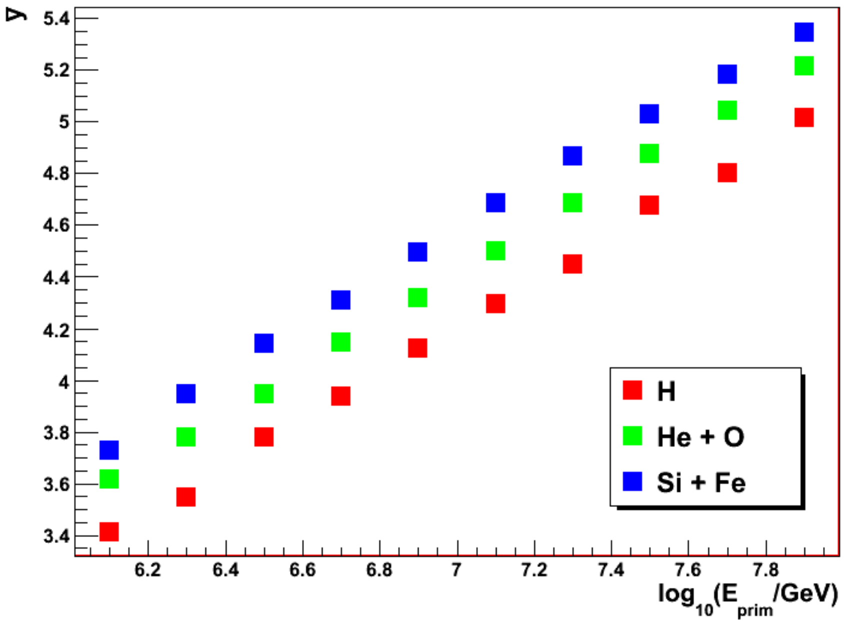
[79] G.D.Agostini, *Nucl.Inst.Meth.Phys.Res.Sec.A*, **362**, 487-498 (1994).

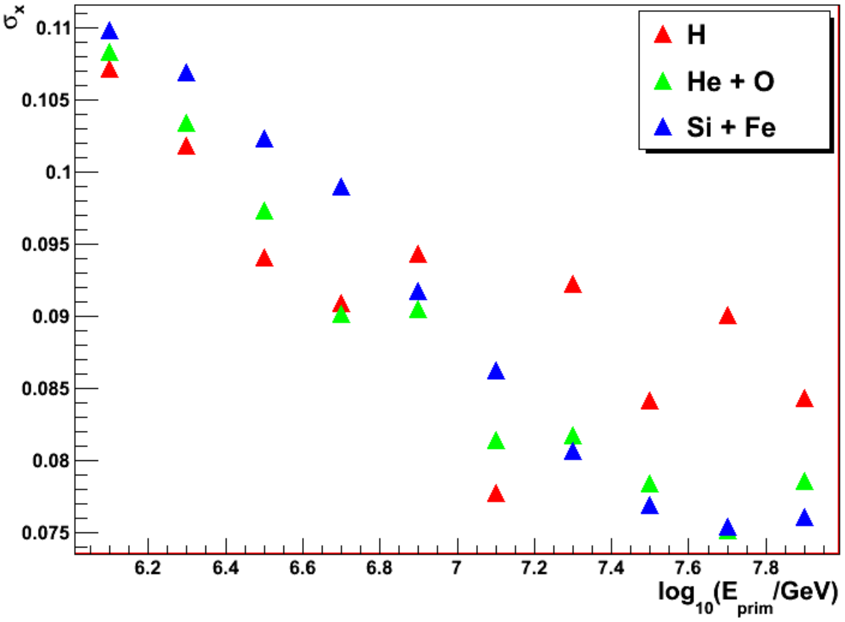
Appendix A

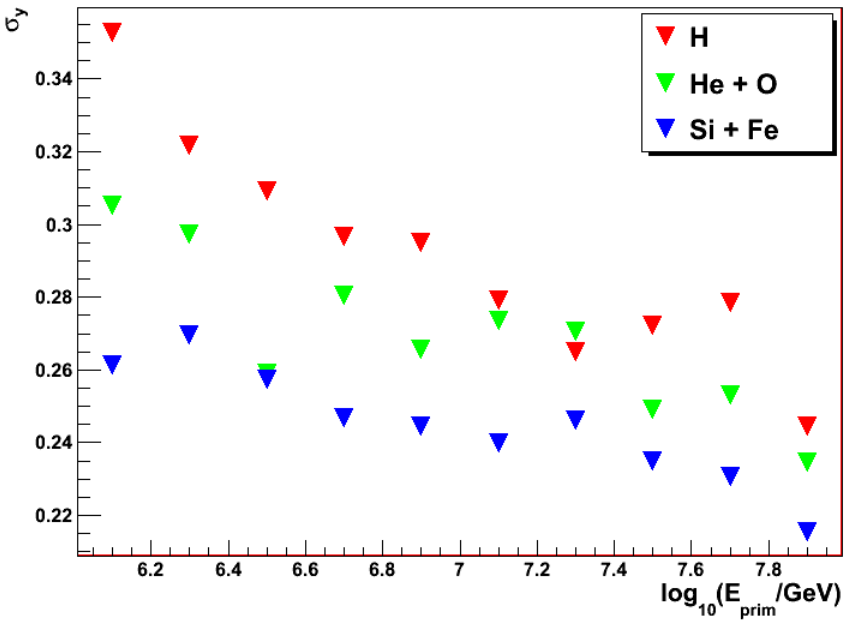
2D FIT RESULTS

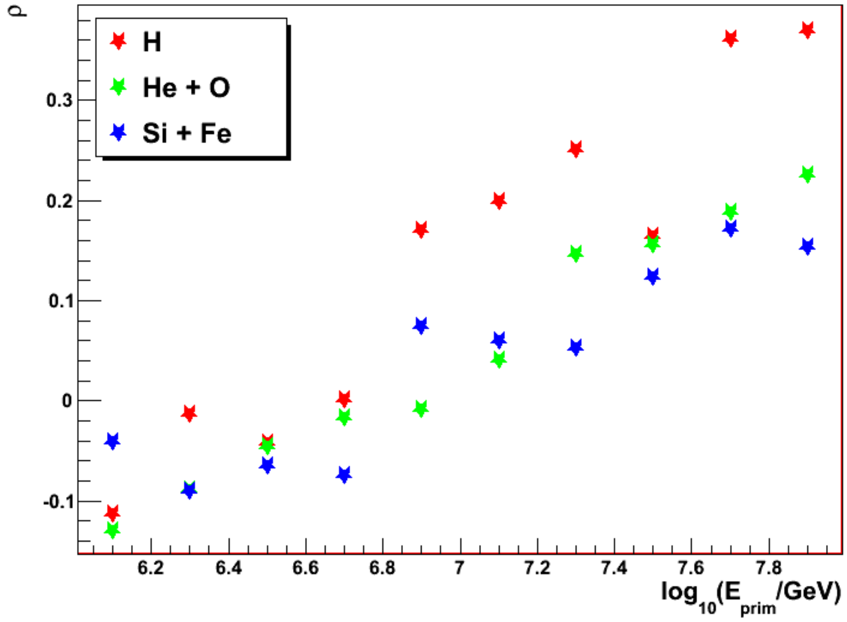
Fit (described in Sec. 7.10) results of IceCube 2D signals ($\log_{10}$ MuE vs $\log_{10}(\mathrm{E_{reco}}/\mathrm{GeV})$): $\bar{x}$, $\bar{y}$, σ_x, σ_y, correlation ρ and detection efficiency.

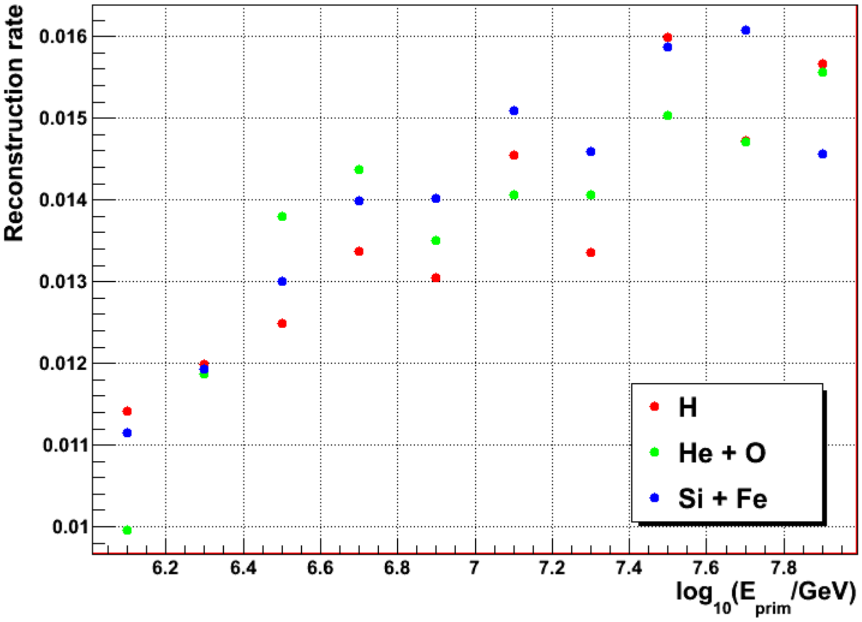



y
5.4
5.2
5
4.8
4.6
4.4
4.2
4
3.8
3.6
3.4
6.2
6.4
6.6
6.8
7
7.2
7.4
7.6
7.8
$\log_{10}(E_{prim}/GeV)$
H
He + O
Si + Fe


σ_x
0.11
0.105
0.1
0.095
0.09
0.085
0.08
0.075
6.2
6.4
6.6
6.8
7
7.2
7.4
7.6
7.8
$\log_{10}(E_{prim}/GeV)$
H
He + O
Si + Fe


σ_y
0.34
0.32
0.3
0.28
0.26
0.24
0.22
6.2
6.4
6.6
6.8
7
7.2
7.4
7.6
7.8
$\log_{10}(E_{prim}/GeV)$
▼ H
▼ He + O
▼ Si + Fe


ρ
0.3
0.2
0.1
0
-0.1
6.2
6.4
6.6
6.8
7
7.2
7.4
7.6
7.8
$\log_{10}(E_{prim}/GeV)$
★ H
★ He + O
★ Si + Fe


Reconstruction rate
0.016
0.015
0.014
0.013
0.012
0.011
0.01
6.2
6.4
6.6
6.8
7
7.2
7.4
7.6
7.8
$\log_{10}(E_{prim}/GeV)$
H
He + O
Si + Fe

Appendix B

UNFOLDING ALGORITHM TESTS

In all the five tests shown below, left (red) is the reconstruction result, middle (blue) is the hypothesized cosmic ray spectrum, and right (black) is the signals that would be generated in IceCube if the hypothesized spectrum realizes.

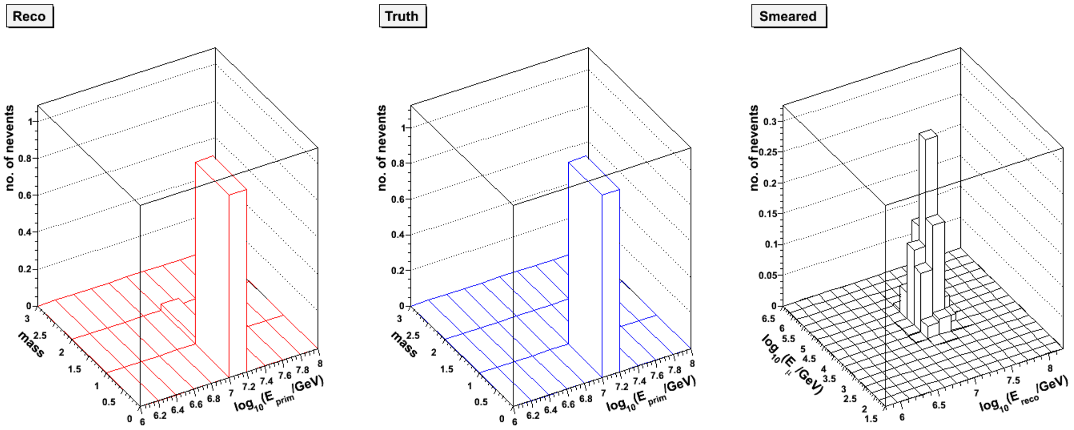


Figure B.1: Test on a single-bin (light composition) spectrum.

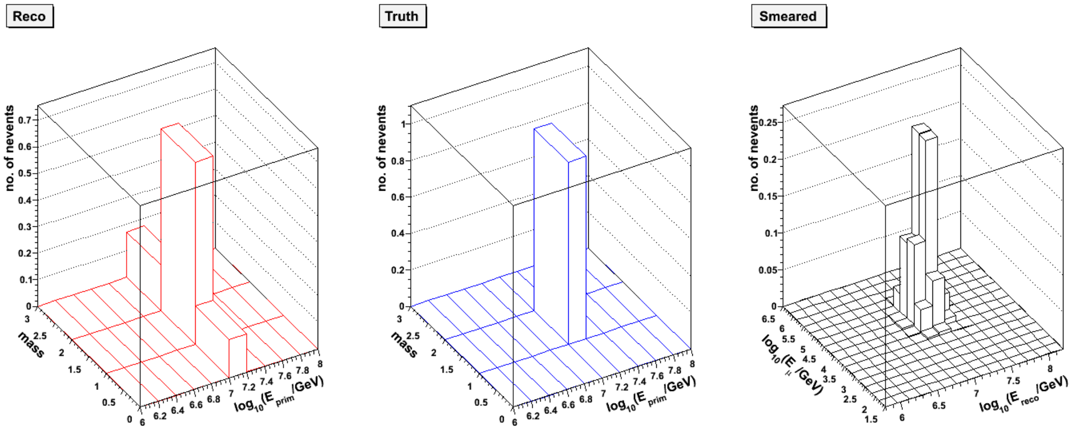


Figure B.2: Test on a single-bin (medium composition) spectrum.

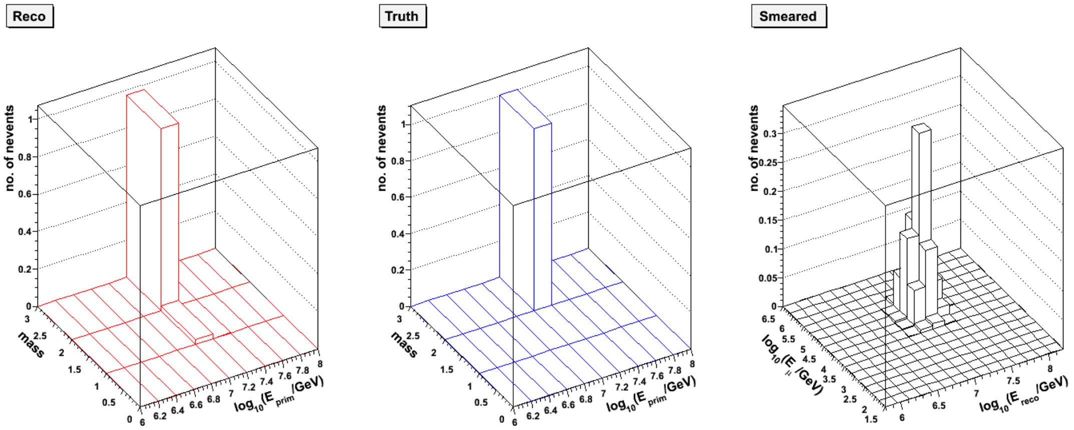


Figure B.3: Test on a single-bin (heavy composition) spectrum.

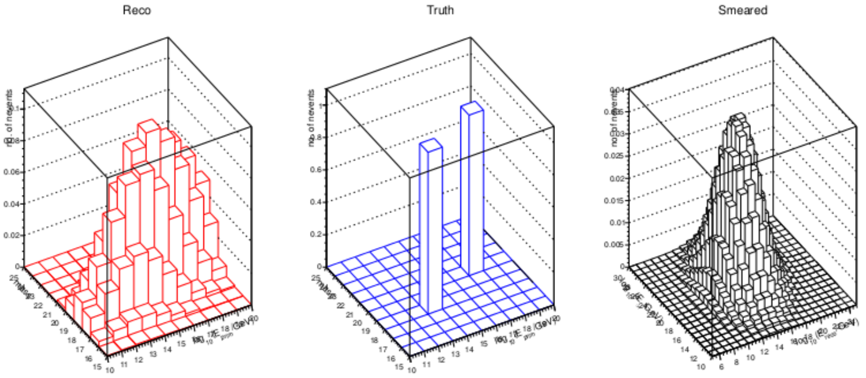


Figure B.4: Test on a double-bin spectrum with wrong initial guess (a flat initial spectrum).

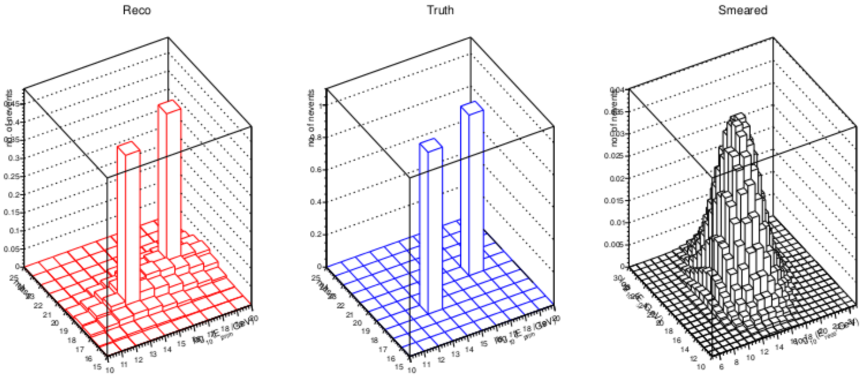


Figure B.5: Test on a double-bin spectrum with right initial guess (a double-spike initial spectrum).

Appendix C

FEATURE EXTRACTION

The feature extractor is the InIce counterpart of IceTop wave processor that determine the arrival times of the arriving photons. It also has independent baseline and droop correction from the DOM calibrator. The methods developed and implemented in feature extractor are described as follows [72]:

1. The waveforms in the three ATWD channels are merged. The higher channels are used for those bins that are saturated in lower channels. When all channels are saturated, the channel 2 is used. The saturation is defined as a bin value is equal to or lower than 20, or equal to or greater than 1000.

2. The baseline is calculated by looking for the most common value in the waveform trace. A parabola is used to fit to the 3 histogram bins closest to the maximum. The mean of the fit is taken as the baseline.

3. In the simple 1-peak extraction, the largest peak is found and assumed to be associated with the most likely photon. The photon arrival time is the leading edge. The total charge is the computed by summing up all the bin values.

4. The multi-peak extraction gives more details in the photon arrival times and break the waveforms to individual SPE pulses. It uses a bayesian unfolding method to decompose the the incoming photon times with an iterative procedure.

5. The feature extraction in FADC traces looks for the first sample above the threshold. It then advances to look for the nearest bin pair between which the rate of ascent is locally at maximum. The leading edge of this pulse is the intersection of the baseline and line fit through samples in these two bins. Summing up the values from the bin found in step 1 to the bin where the waveform falls below the threshold again as the charge of the pulse.

6. The droop correction is required for the fast Bayesian unfolding method for remove the droop introduced mostly by the PMT base circuit by a simple z-transform-based algorithm.

Appendix D

DOM CALIBRATION

The elements of the DOM calibration file are:

- temperature

  The DOM temperature at calibration time, measured in K by the mainboard temperature sensor. This is 10-12K above the ambient temperature, unless the DOM has just been powered on. Note that in general, the calibration is only valid around this temperature.

- frontEndImpedance

  This is the front-end imedance in ohm used in charge calibrations. This value is either 43 or 50 *Omega*.

- speDiscrimCalib

  The front-end SPE discriminator calibration. This calibration translates the SPE discriminator DAC setting to the amplified charge (after the PMT) corresponding to the trigger level. The calibration results are described by a linear fit:

  $$\text{charge trigger level (pC)} = m \cdot \text{DAC} + b.$$

- mpeDiscrimCalib

  The front-end MPE discriminator calibration. This calibration translates the MPE discriminator DAC setting to the amplified charge (after the PMT) corresponding to the trigger level. The calibration results are described by a linear fit:

$$\text{charge trigger level (pC)} = m \cdot \text{DAC} + b.$$

- FADCBaseline

  The FADC baseline records the linear relationship of the FADC pedestal values to the FADC reference DAC setting

  $$\text{FADC baseline} = m \cdot \text{DAC} + b.$$

- FADCGain

  The FADC gain inclues the intrinsic gain of the FADC as well as the channel amplifiers.

- fadcDeltaT

  This is the FADC time offset from ATWD waveform start in ns. The FADC waveform starts earlier than the ATWD waveform and the reported number is negative.

- atwdDeltaT

  This is the element that contains the ATWD time offset from slightly different delays in ATWD digitization start. The ATWD used in the transit time calibration will have an offset of zero, while the other could have up to a few ns offset.

- atwdFreq

  Quadratic fit for each ATWD sampling frequency, one for each chip (0, 1). This can be used to translate the ATWD waveform x-axis into time.

- pmtTransitTime

  The PMT transit time contains a linear fit element relating the total transit time of the DOM (PMT transit time plus the delay board) as a function of high voltage. The relation is as follows:

$$\text{transit time (ns)} = m/\sqrt{V} + b,$$

where V is the high voltage. This calibration can be used to adjust the waveform leading-edge times back to the time of light reaching the DOM.

- hvGainRelation

  This a calibration of the PMT gain versus high voltage, described by the fit:

  $$\log_{10} \text{gain} = m \cdot \log_{10} V + b.$$

- pmtDiscrimCalib

  A refined SPE discriminator calibration generated using the actual PMT pulses.

- TauParam

  Parameters for droop correction.

- ampGains

  Gains for ATWD channel 0, 1 and 2.

- atwdBin0 and atwdBin1

  Results of the linear fit for the ATWD bin calibration. These values convert counts to voltage for each bin of each channel in the ATWD.

- atwdBaselines

  ATWD baseline corrections.

- atwdResponseWidth

  This value stores the response width of the electronics to a pulse (ATWD).

- fadcResponseWidth

  This value stores the response width of the electronics to a pulse (FADC).

- relativeDomEff

  Relative DOM efficiency, normalized to 1.0 for the average DOMs.

- nosiseRate

  DOM noise rate in Hz. This value is from 300 to 1000 Hz.

For IceTop DOMs, there is additional vertical equivalent muon (VEM) calibration data that contains the following items:

- pePerVEM

  This is the average number of PE per vem.

- muPeakWidth

  The value is the width of the average muon peak for this tank in PE.

- hglgCrossOver

  PE threshold for the high gain (HG) pulses over which the corresponding low gain pulses in the tank are used.

- corrFactor

  Optional correction factor to adjust high gain (HG) and low gain DOMs.

Appendix E

DOM STATUS

The DOM status defines the following elements:

- trigMode

  The different kinds of triggers that can be used for the DOM. The standard data-taking operation is SPE.

- lcMode

  This controls how far the local coincidence signal is sent and received.

- DOMGain

  All InIce DOMs and IceTop HG DOMs are running at a gain about 1×10^7 while IceTop LG DOMs are running at a low gain about 5×10^5.

- CableType

  This parameter tells whether the DOm has a terminated or unterminated cable.

- lcWindowPre

  The local coincidence window size before ATWD launch window.

- lcWindowPost

  The local coincidence window size after ATWD launch window.

- lcSpan

  The number of the nearest neighbors to consider in LC search ranging from 1 to 4.

- pmtHV

  This is the real operating PMT voltage, which is around 1200 V.

- speThreshold

  SPE discriminator level.

- mpeThreshold

  MPE discriminator level.

- fePedestal

  The Pedestal voltage of the analog front end.

- dacFADCRef

  The reference setting for FADC baseline.

Appendix F

ATWD TIME CALIBRATION

The ATWD time calibration is the process that translates the leading edge time of the ATWD waveform to the time when the photon hits the PMT. The detailed calibration procedure is listed as follows [44]:

1. Determine the bin position if the feature.
2. Convert the bin offset within the waveform into time T_{offset} by using the sampling speed *atwdfreq* calibration for the appropriate ATWD.
3. Add the feature offset time to the ATWD trigger time T_{launch}.
4. Subtract the transit time $T_{transit}$ for the given operating voltage, determined from the *pmtTransitTime* in calibration data.
5. Subtract the ATWD offset Δ_{ATWD} correction from the *atwd_deltaT* calibration. The ATWD offset is positive if the launch time for this ATWD is later than the ATWD used in the transit time calibration. Thus it can be viewed as a correction to the transit time.

Thus, the PMT hit time can be expressed as

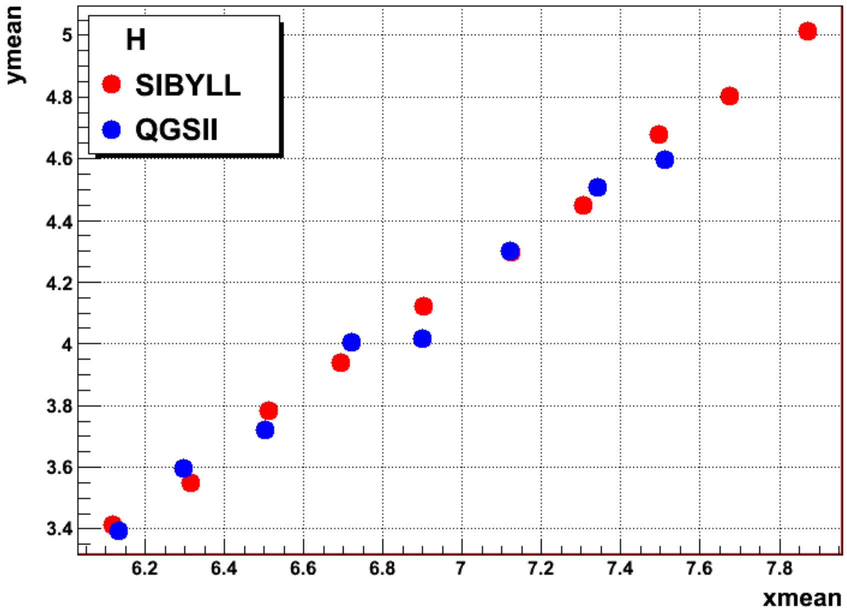
$$\begin{aligned} T_{hit} &= T_{launch} + T_{offset} - (T_{transit} + \Delta_{ATWD}) \\ &= T_{transit} + \frac{bin \cdot 10^3}{F_{ATWD}/MHz} - \left(\frac{m_{TT}}{\sqrt{V}} + b_{TT} + \Delta_{ATWD} \right), \end{aligned}$$

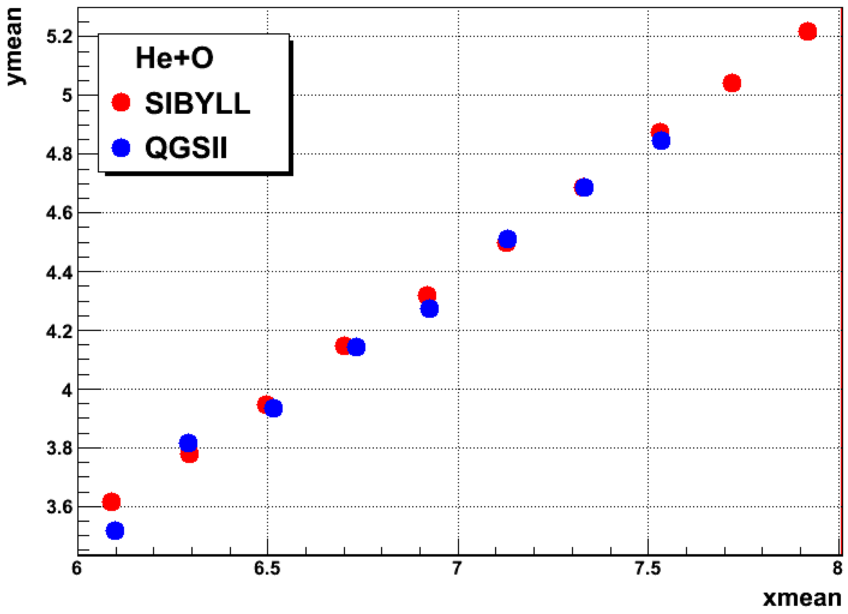
where F_{ATWD} the ATWD sampling frequency is 303 MHz.

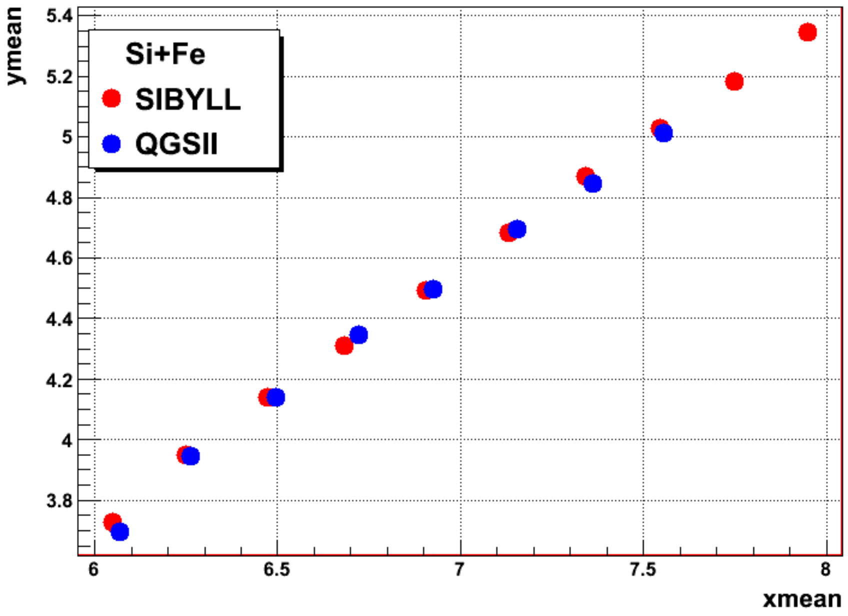
Appendix G

2D FIT RESULTS (SIBYLL VS QGSJET-II)

The following three figures are the discrepancies between SIBYLL and QGSJet-II models for proton, helium plus oxygen, silicon plus iron groups respectively. The X and Y axes of the figures are fitted mean $\log_{10}(E_{reco}/GeV)$ and $\log_{10}$ MuE respectively in a 2D gaussian distribution.




He+O
● SIBYLL
● QGSII
ymean
5.2
5
4.8
4.6
4.4
4.2
4
3.8
3.6
6
6.5
7
7.5
8
xmean


Si+Fe
● SIBYLL
● QGSII
ymean
5.4
5.2
5
4.8
4.6
4.4
4.2
4
3.8
6
6.5
7
7.5
8
xmean

Appendix H

RCONSTRUCTION UNCERTAINTY (K70 AND MUE)

The systematic uncertainty due to reconstruction algorithm is estimated by using K70 and MuE with the forward folding method, in which distributions of K70 and MuE between simulation and experiment are compared at fixed reconstructed primary energy.

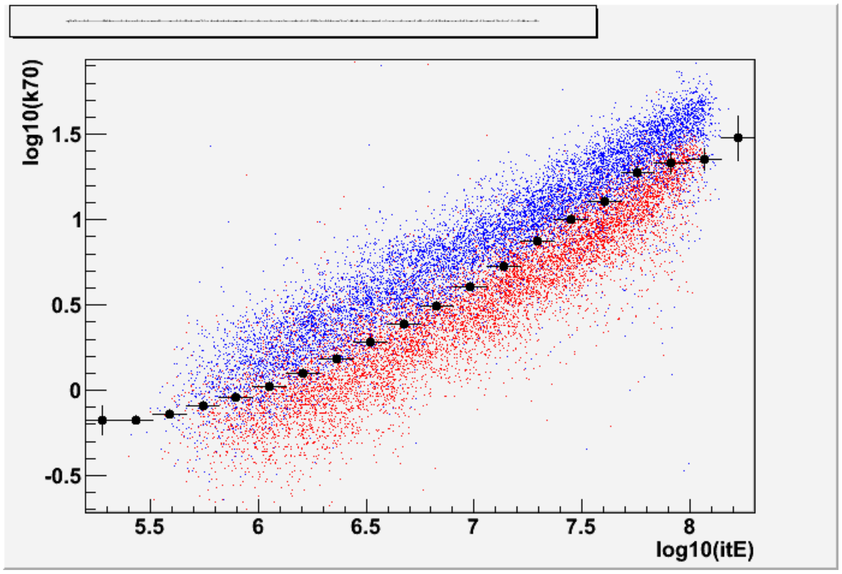


Figure H.1: K70 vs reconstruction primary energy. Red is proton simulation, blue is iron simulation and black is experimental data.

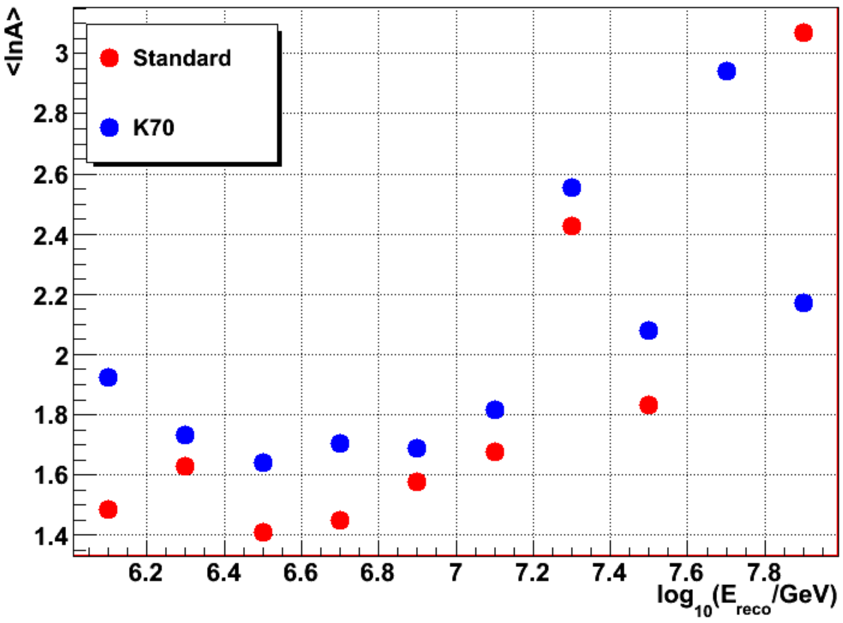


Figure H.2: Composition results by using K70 and MuE. Note the horizontal axis is the reconstructed primary energy.

Appendix I

FORWARD FOLDING RESULTS

Bin 1:

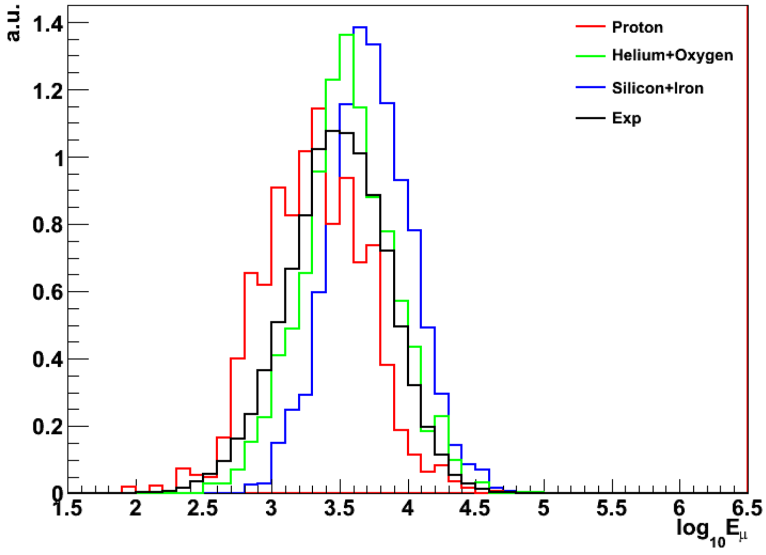


Figure I.1: MuE distribution of bin 1 ($6.0 \leq \log_{10} E_{reco} < 6.2$) in linear scale.

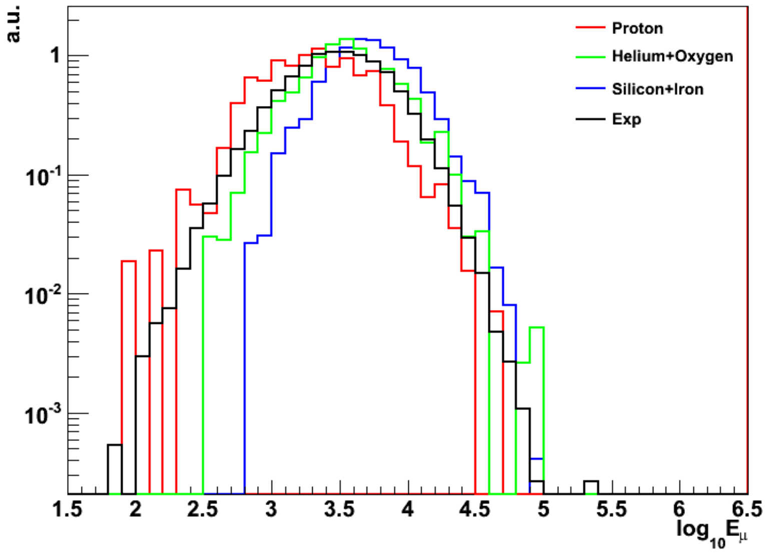


Figure I.2: MuE distribution of bin 1 ($6.0 \leq \log_{10} E_{reco} < 6.2$) in logarithmic scale.

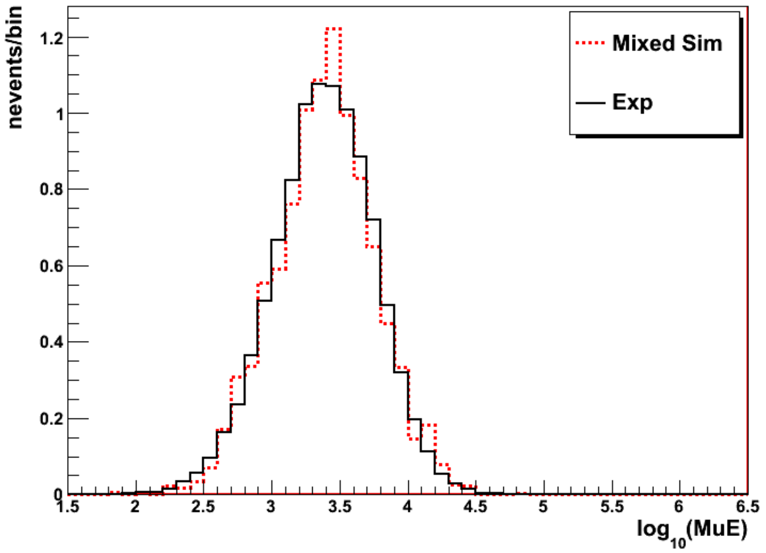


Figure I.3: Foward folding result of MuE of bin 1 ($6.0 \leq \log_{10} E_{reco} < 6.2$) in linear scale.

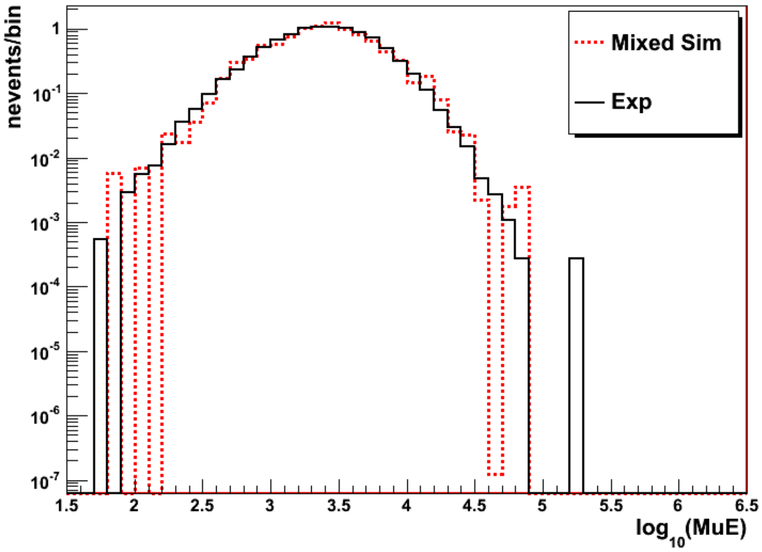


Figure I.4: Foward folding result of MuE of bin 1 ($6.0 \leq \log_{10} E_{reco} < 6.2$) in logarithmic scale.

Bin 2:

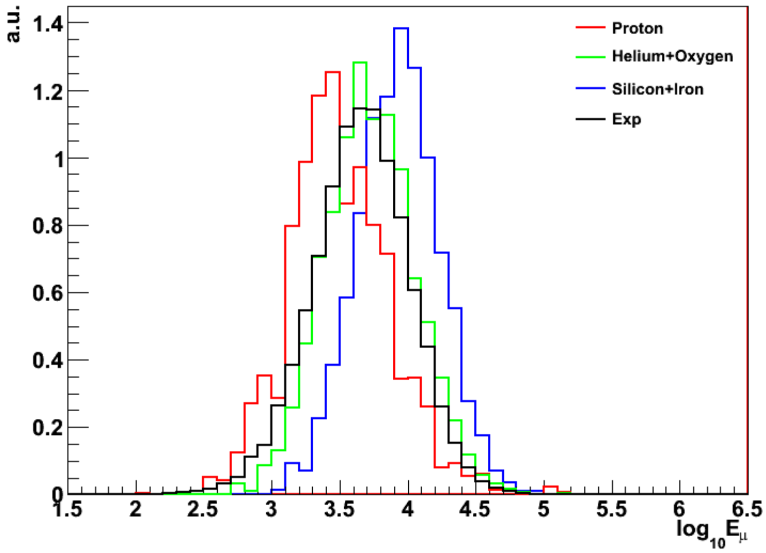


Figure I.5: MuE distribution of bin 2 ($6.2 \leq \log_{10} E_{reco} < 6.4$) in linear scale.

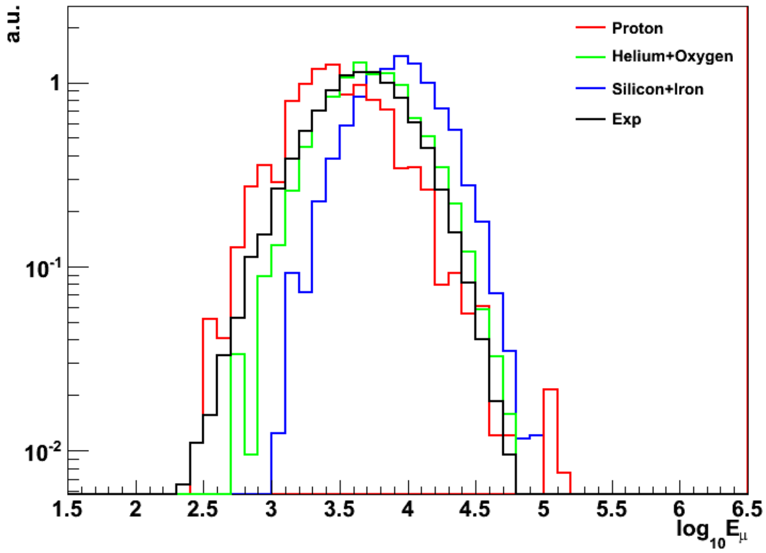


Figure I.6: MuE distribution of bin 2 ($6.2 \leq \log_{10} E_{reco} < 6.4$) in logarithmic scale.

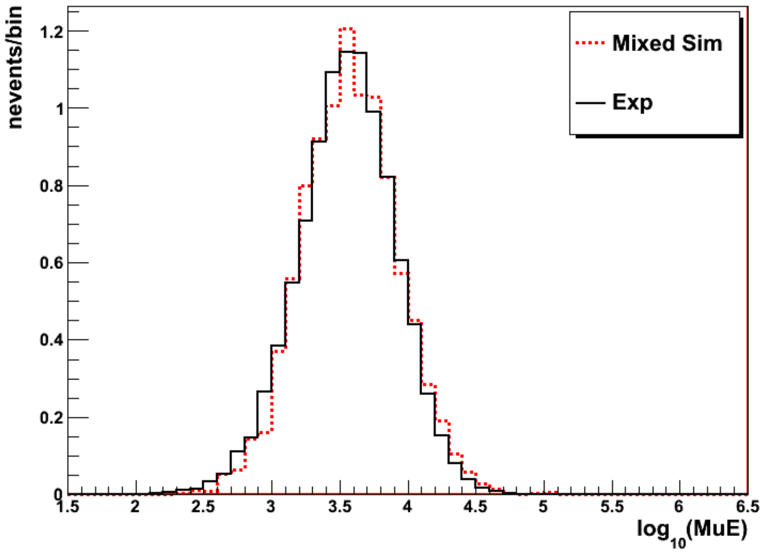


Figure I.7: Foward folding result of MuE of bin 2 ($6.2 \leq \log_{10} E_{reco} < 6.4$) in linear scale.

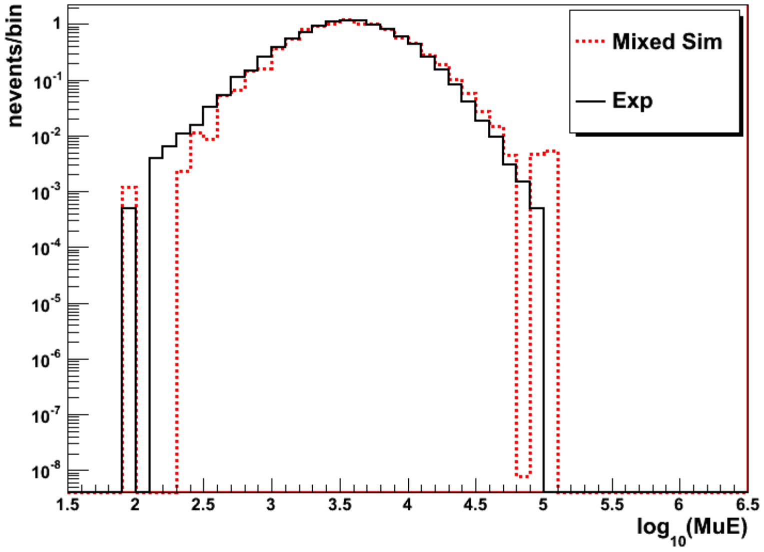


Figure I.8: Foward folding result of MuE of bin 2 ($6.2 \leq \log_{10} E_{reco} < 6.4$) in logarithmic scale.

Bin 3:

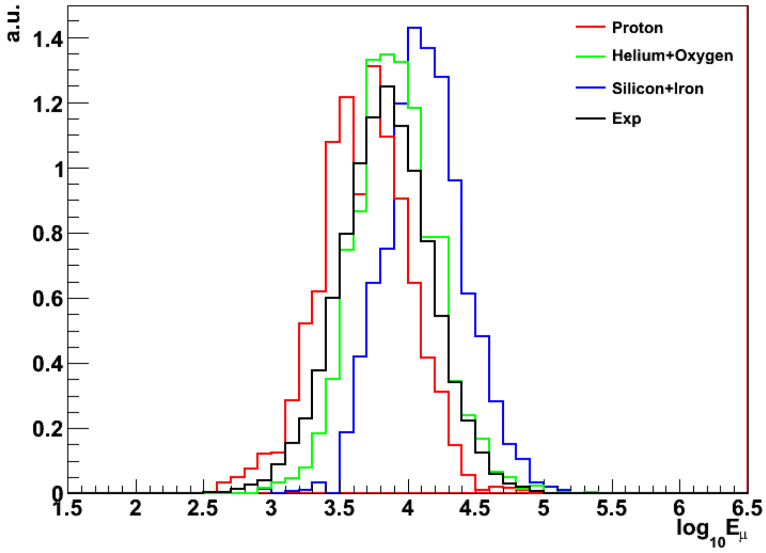


Figure I.9: MuE distribution of bin 3 ($6.4 \leq \log_{10} E_{reco} < 6.6$) in linear scale.

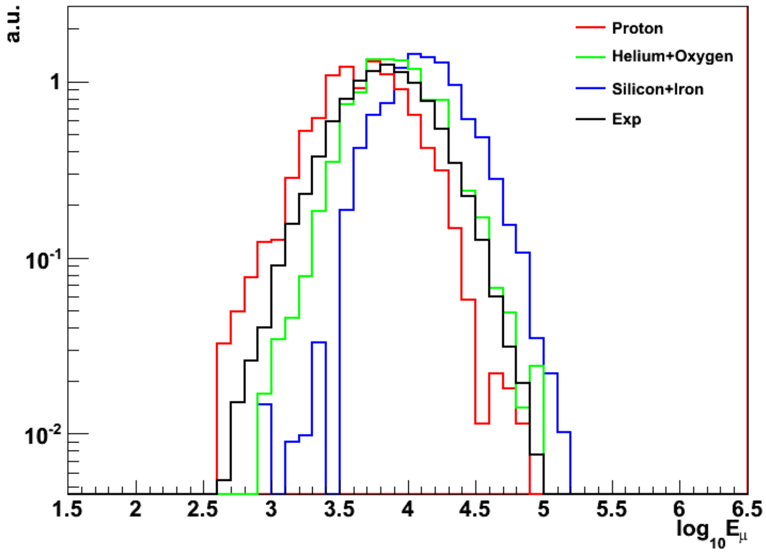


Figure I.10: MuE distribution of bin 3 ($6.4 \leq \log_{10} E_{reco} < 6.6$) in logarithmic scale.

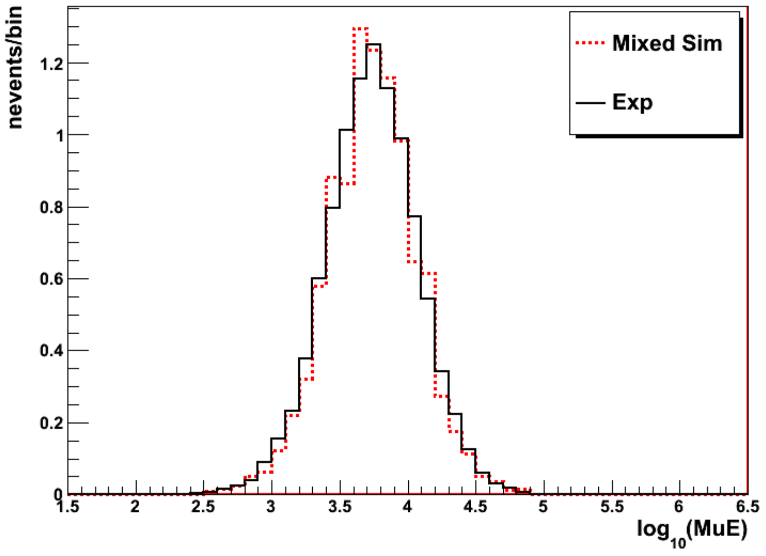


Figure I.11: Foward folding result of MuE of bin 3 ($6.4 \leq \log_{10} E_{reco} < 6.6$) in linear scale.

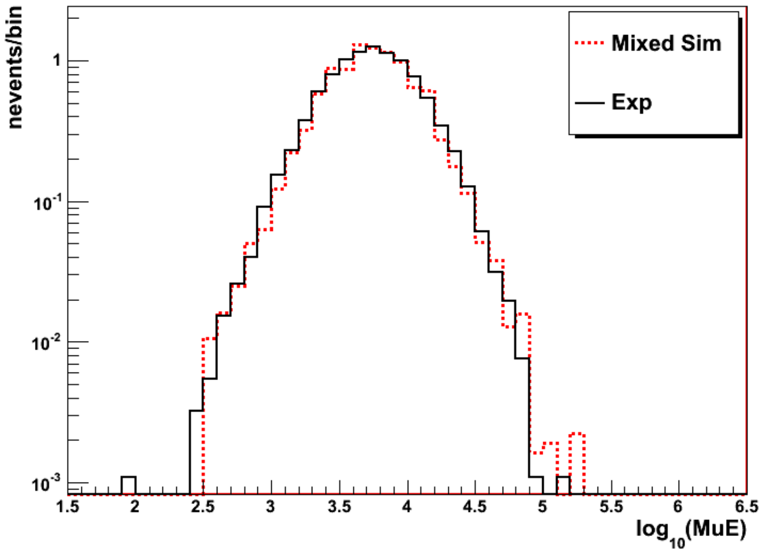


Figure I.12: Foward folding result of MuE of bin 3 ($6.4 \leq \log_{10} E_{reco} < 6.6$) in logarithmic scale.

Bin 4:

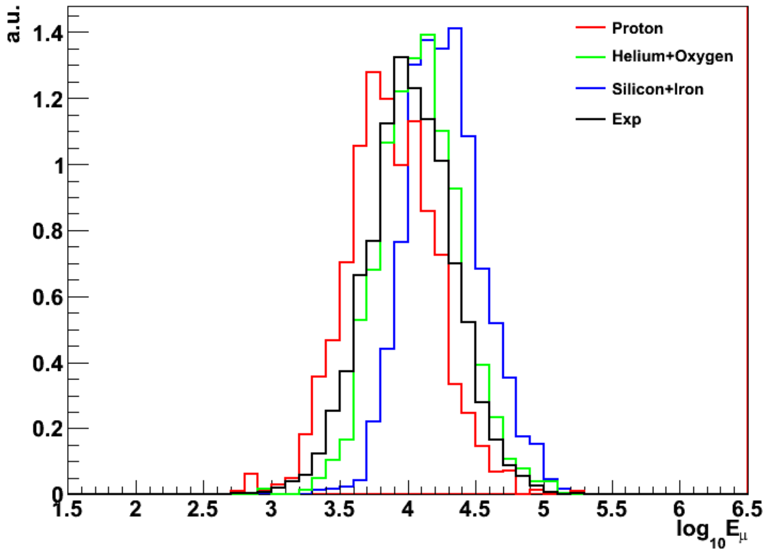


Figure I.13: MuE distribution of bin 4 ($6.6 \leq \log_{10} E_{reco} < 6.8$) in linear scale.

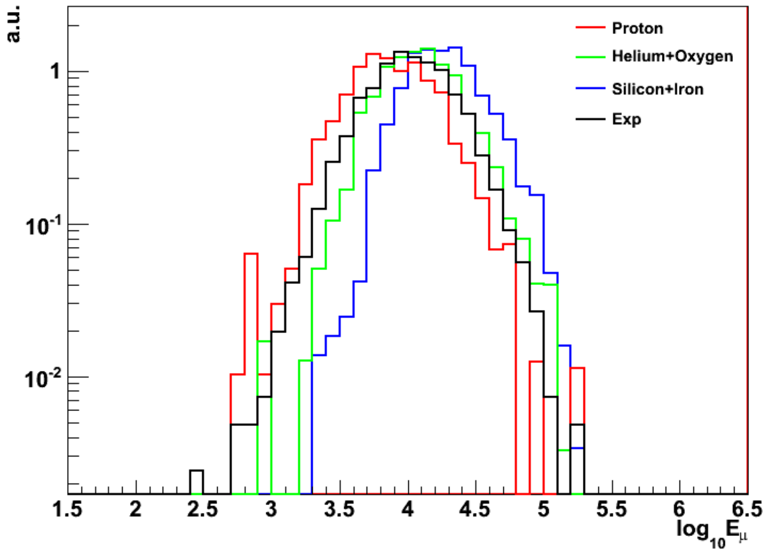


Figure I.14: MuE distribution of bin 4 ($6.6 \leq \log_{10} E_{reco} < 6.8$) in logarithmic scale.

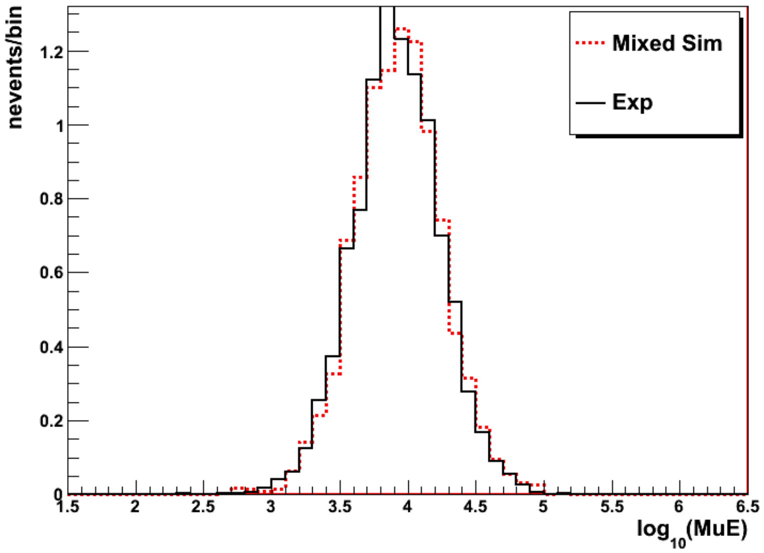


Figure I.15: Foward folding result of MuE of bin 4 ($6.6 \leq \log_{10} E_{reco} < 6.8$) in linear scale.

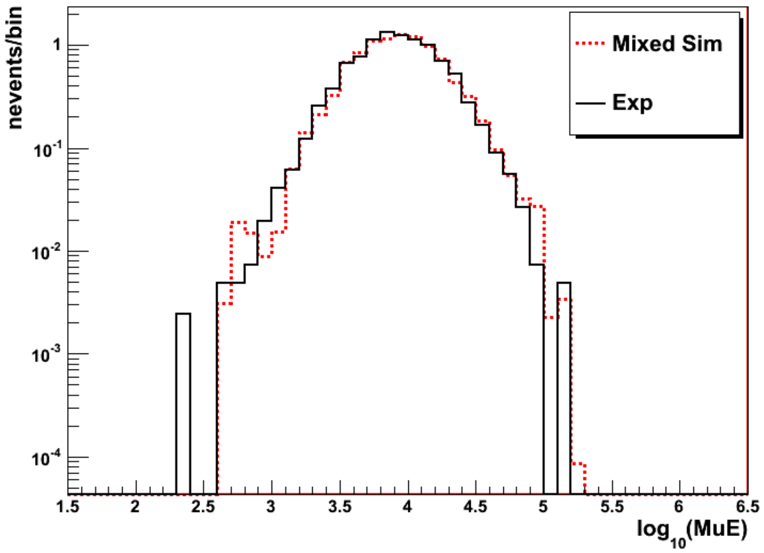


Figure I.16: Foward folding result of MuE of bin 4 ($6.6 \leq \log_{10} E_{reco} < 6.8$) in logarithmic scale.

Bin 5:

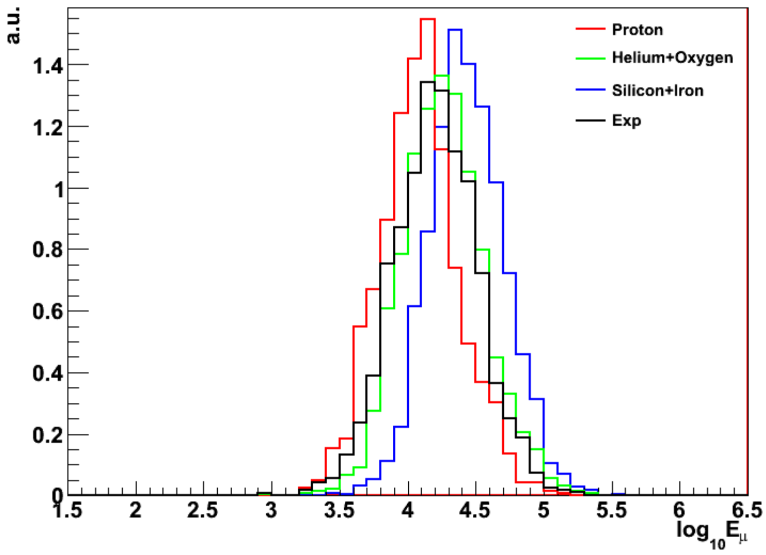


Figure I.17: MuE distribution of bin 5 ($6.8 \leq \log_{10} E_{reco} < 7.0$) in linear scale.

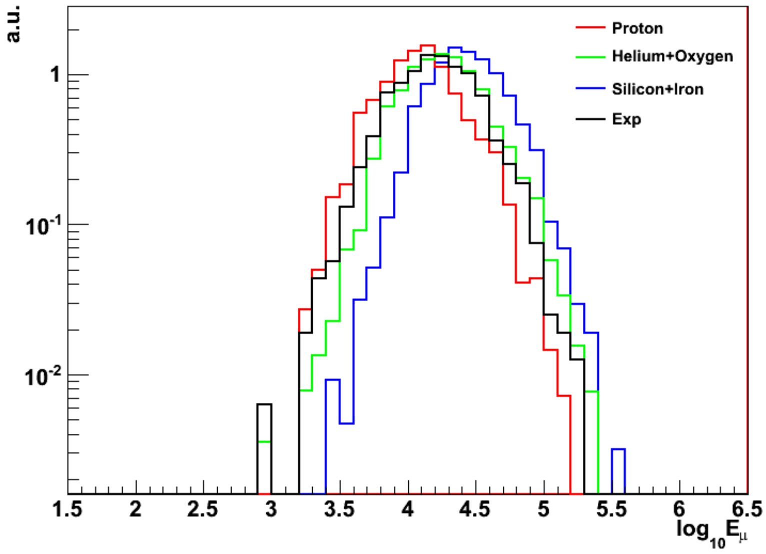


Figure I.18: MuE distribution of bin 5 ($6.8 \leq \log_{10} E_{reco} < 7.0$) in logarithmic scale.

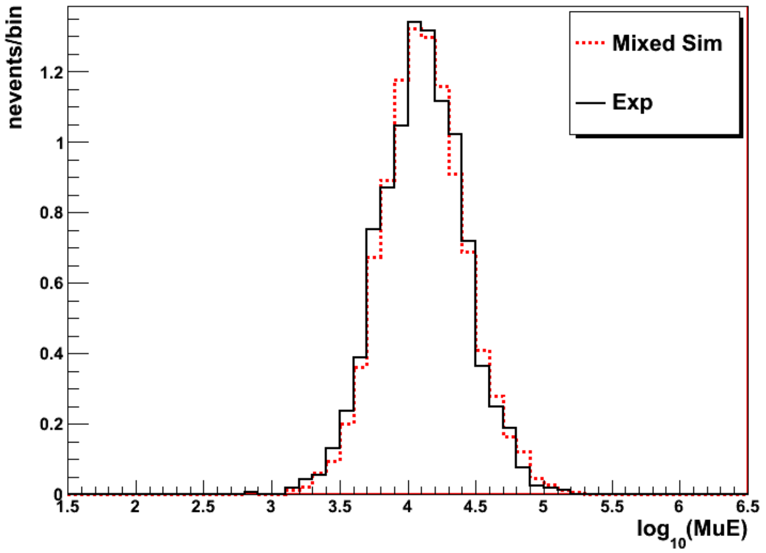


Figure I.19: Foward folding result of MuE of bin 5 ($6.8 \leq \log_{10} E_{reco} < 7.0$) in linear scale.

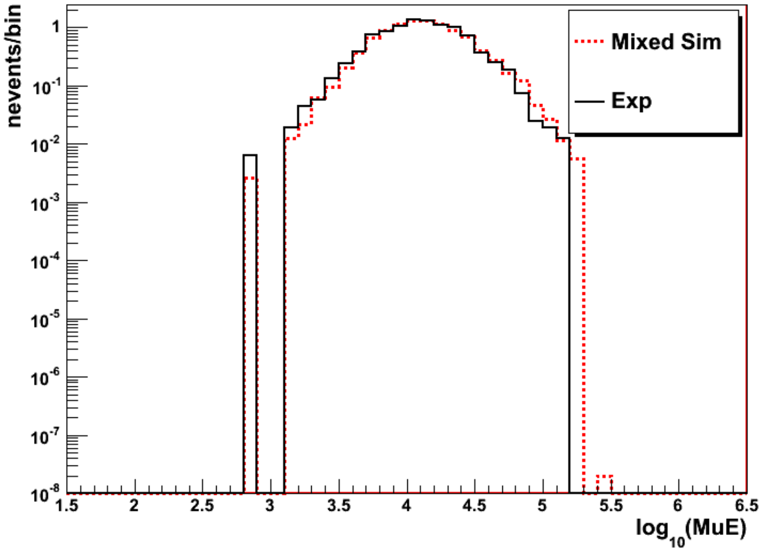


Figure I.20: Foward folding result of MuE of bin 5 ($6.8 \leq \log_{10} E_{reco} < 7.0$) in logarithmic scale.

Bin 6:

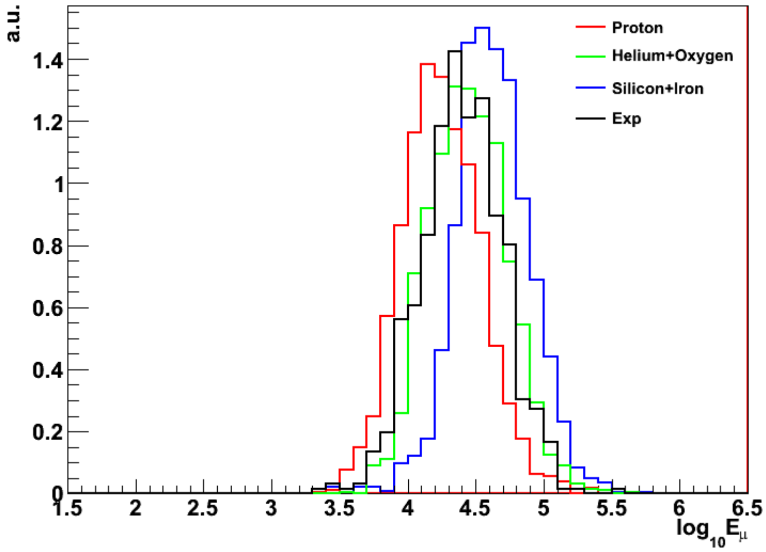


Figure I.21: MuE distribution of bin 6 ($7.0 \leq \log_{10} E_{reco} < 7.2$) in linear scale.

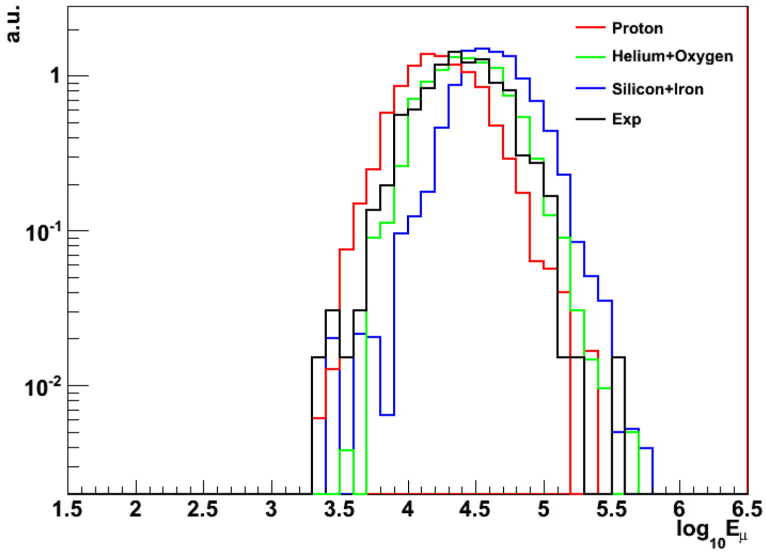


Figure I.22: MuE distribution of bin 6 ($7.0 \leq \log_{10} E_{reco} < 7.2$) in logarithmic scale.

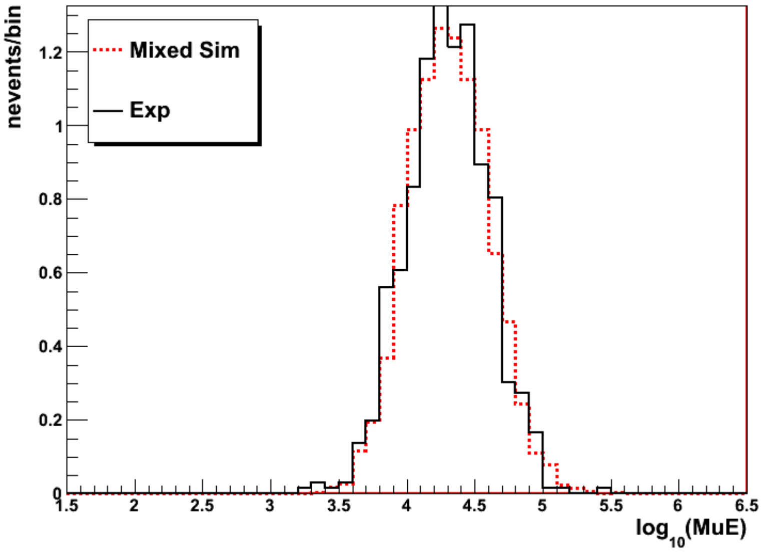


Figure I.23: Foward folding result of MuE of bin 6 ($7.0 \leq \log_{10} E_{reco} < 7.2$) in linear scale.

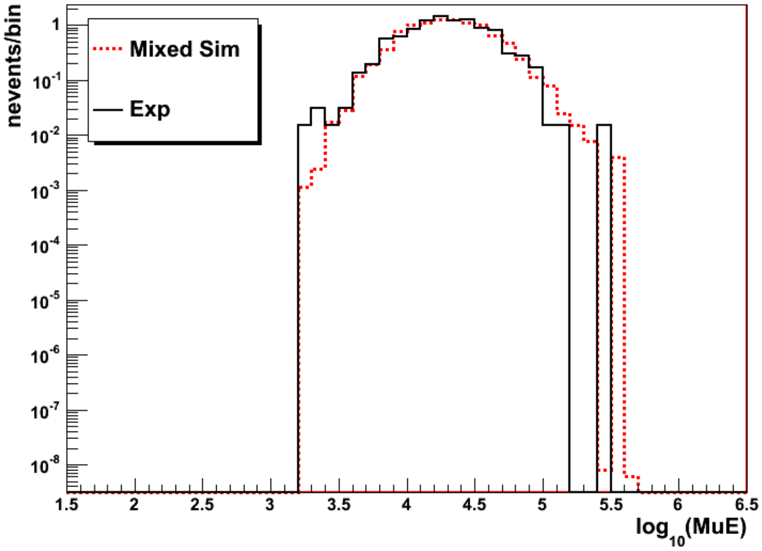


Figure I.24: Foward folding result of MuE of bin 6 ($7.0 \leq \log_{10} E_{reco} < 7.2$) in logarithmic scale.

Bin 7:

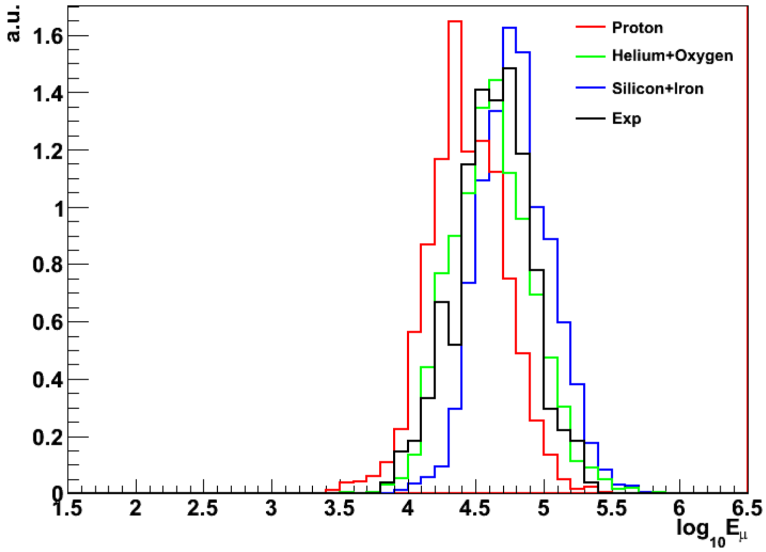


Figure I.25: MuE distribution of bin 7 ($7.2 \leq \log_{10} E_{reco} < 7.4$) in linear scale.

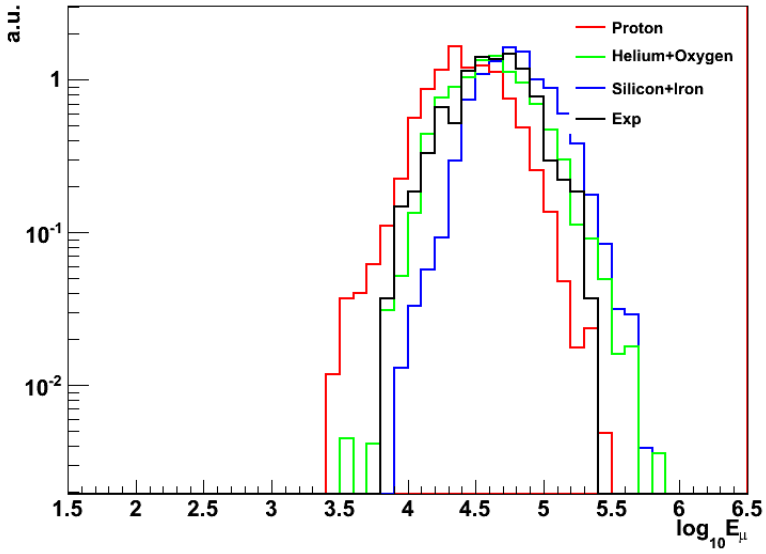


Figure I.26: MuE distribution of bin 7 ($7.2 \leq \log_{10} E_{reco} < 7.4$) in logarithmic scale.

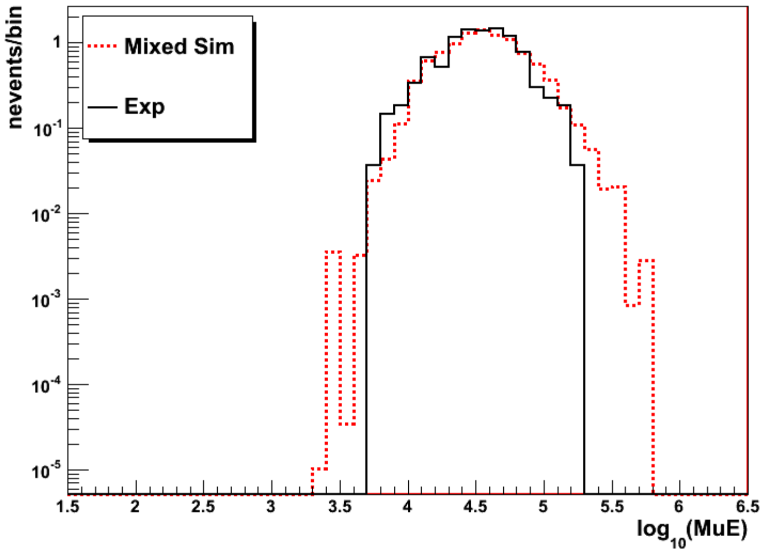


Figure I.27: Foward folding result of MuE of bin 7 ($7.2 \leq \log_{10} E_{reco} < 7.4$) in linear scale.

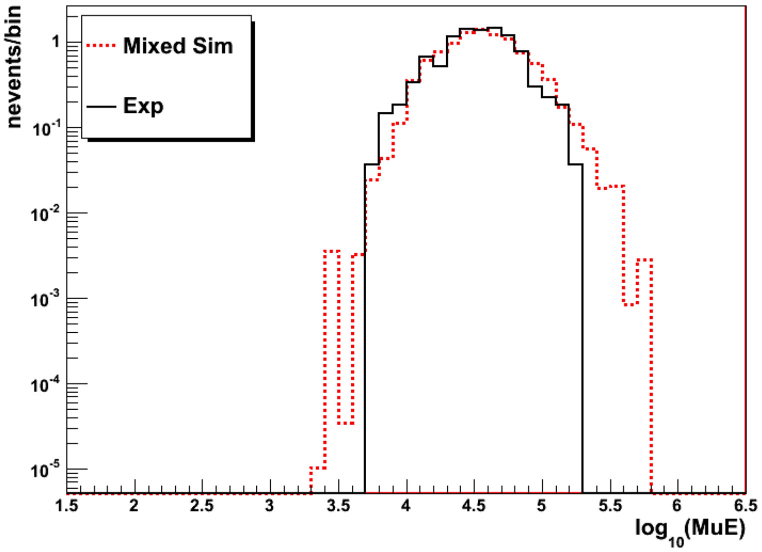


Figure I.28: Foward folding result of MuE of bin 7 ($7.2 \leq \log_{10} E_{reco} < 7.4$) in logarithmic scale.

Bin 8:

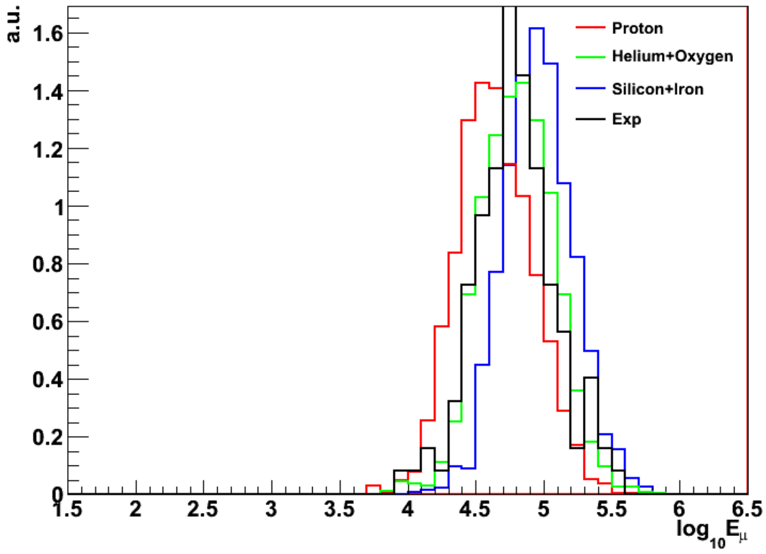


Figure I.29: MuE distribution of bin 8 ($7.4 \leq \log_{10} E_{reco} < 7.6$) in linear scale.

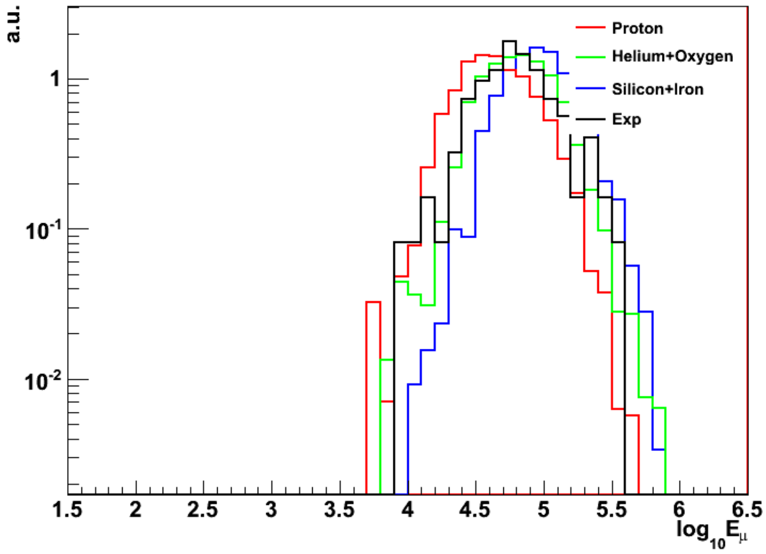


Figure I.30: MuE distribution of bin 8 ($7.4 \leq \log_{10} E_{reco} < 7.6$) in logarithmic scale.

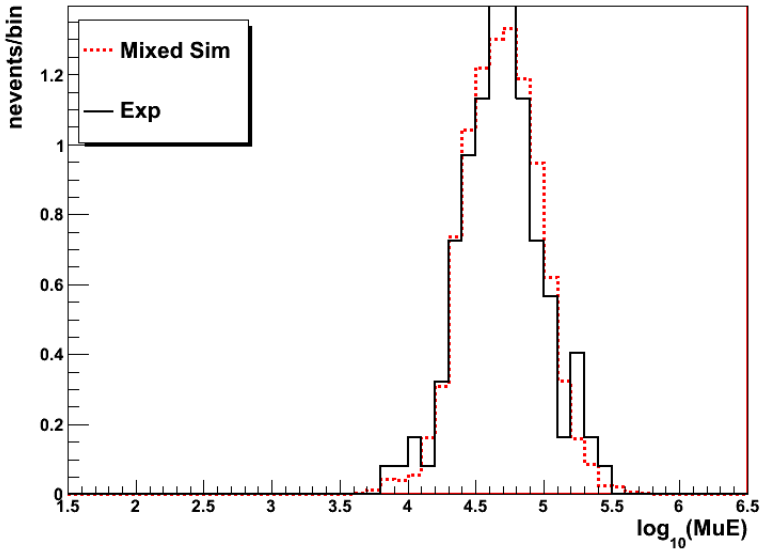


Figure I.31: Foward folding result of MuE of bin 8 ($7.4 \leq \log_{10} E_{reco} < 7.6$) in linear scale.

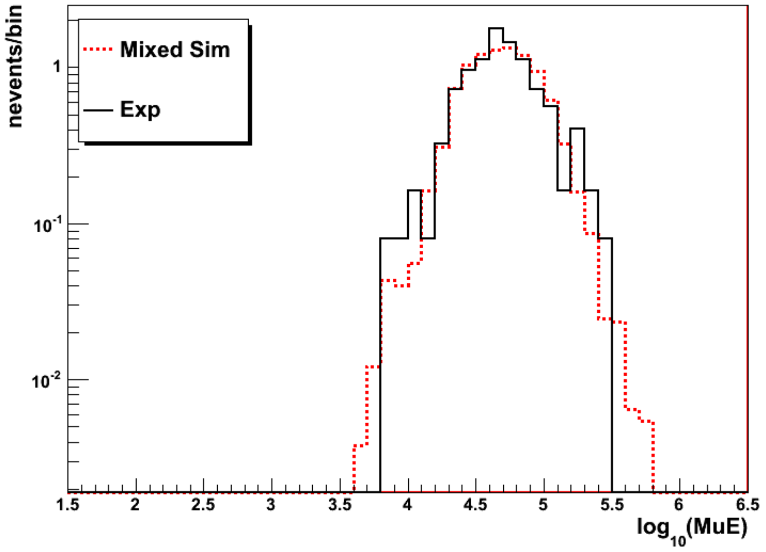


Figure I.32: Foward folding result of MuE of bin 8 ($7.4 \leq \log_{10} E_{reco} < 7.6$) in logarithmic scale.

Bin 9:

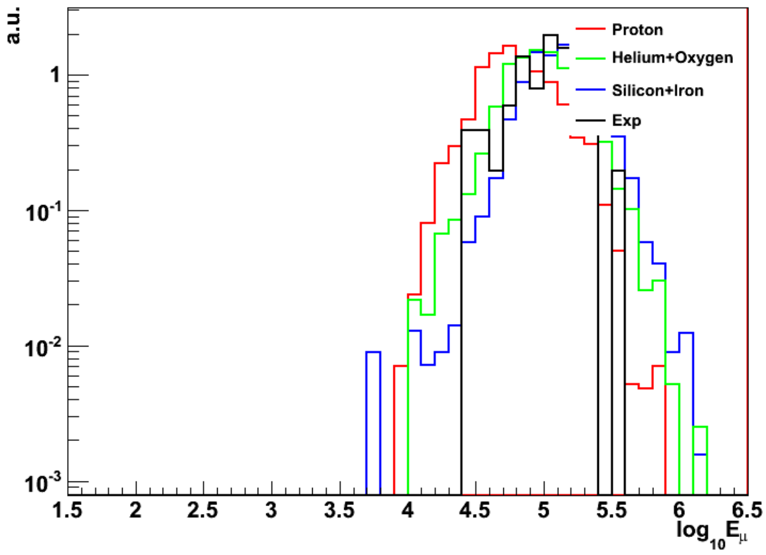


Figure I.33: MuE distribution of bin 9 ($7.6 \leq \log_{10} E_{reco} < 7.8$) in linear scale.

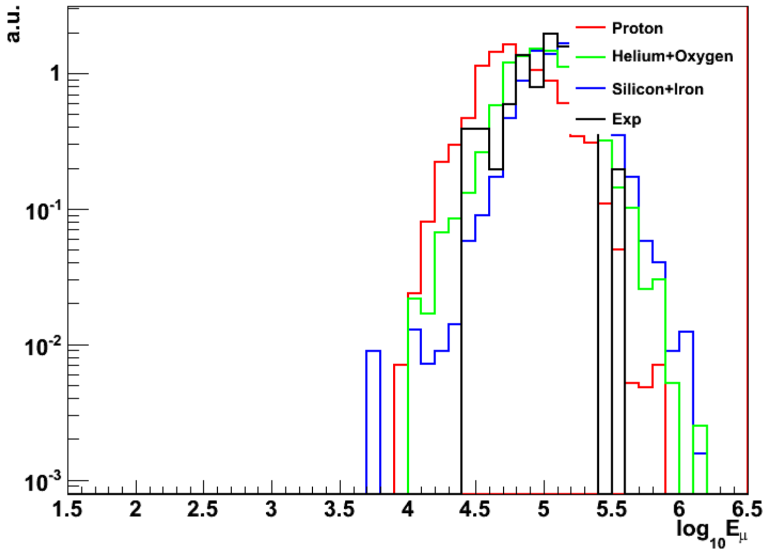


Figure I.34: MuE distribution of bin 9 ($7.6 \leq \log_{10} E_{reco} < 7.8$) in logarithmic scale.

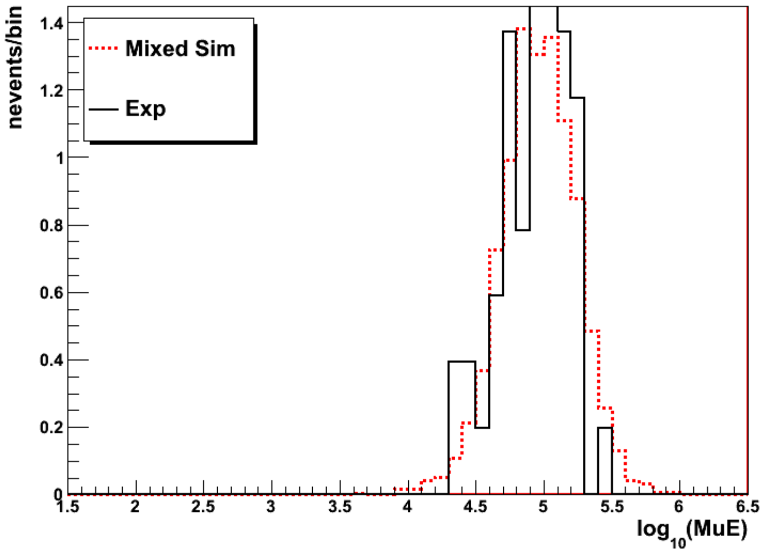


Figure I.35: Foward folding result of MuE of bin 9 ($7.6 \leq \log_{10} E_{reco} < 7.8$) in linear scale.

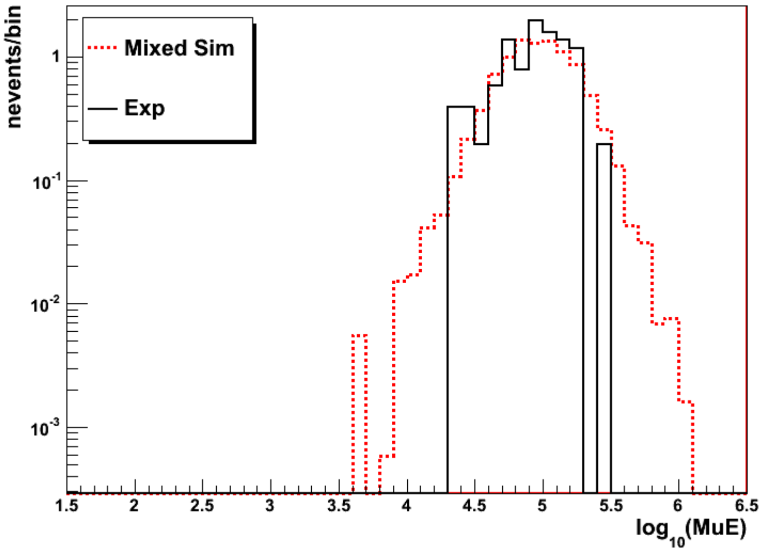


Figure I.36: Foward folding result of MuE of bin 9 ($7.6 \leq \log_{10} E_{reco} < 7.8$) in logarithmic scale.

Bin 10:

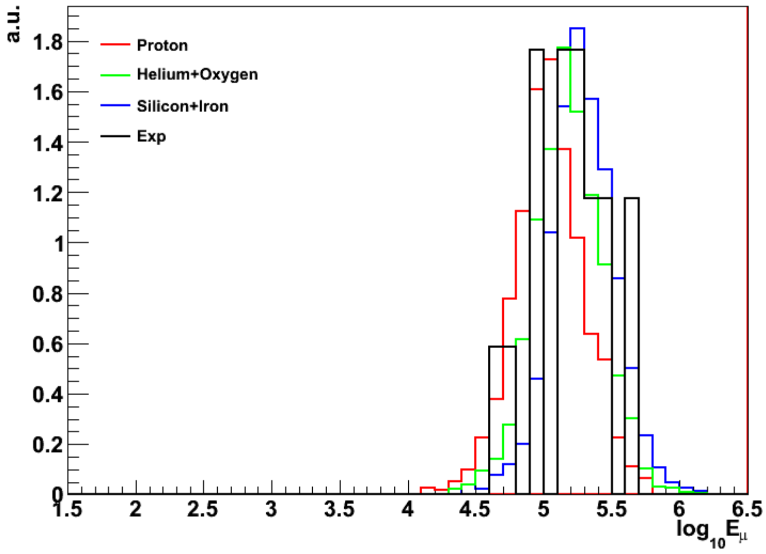


Figure I.37: MuE distribution of bin 10 ($7.8 \leq \log_{10} E_{reco} < 8.0$) in linear scale.

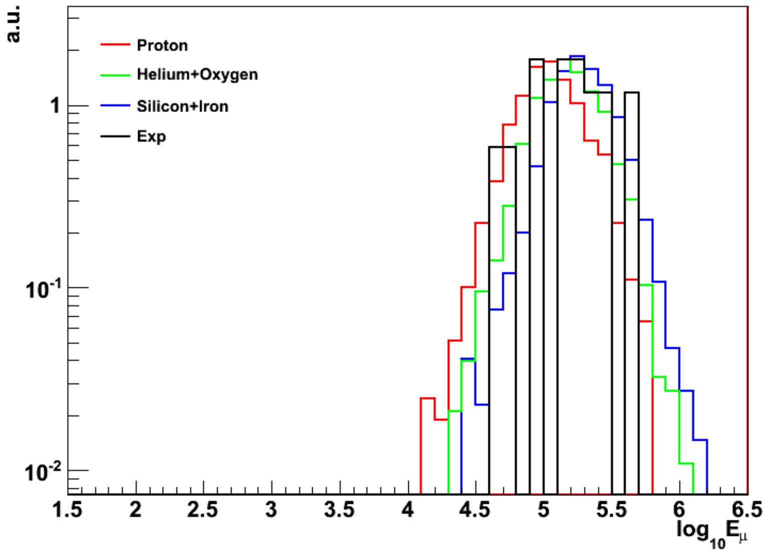


Figure I.38: MuE distribution of bin 10 ($7.8 \leq \log_{10} E_{reco} < 8.0$) in logarithmic scale.

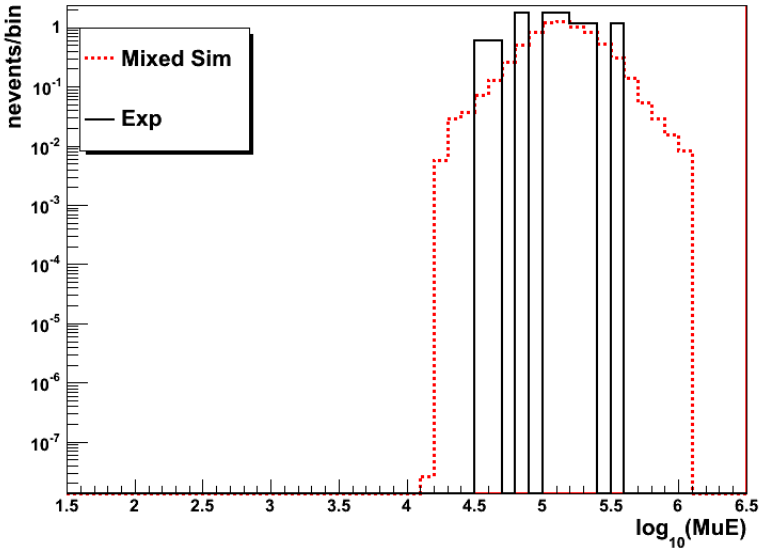


Figure I.39: Foward folding result of MuE of bin 10 ($7.8 \le \log_{10} E_{reco} < 8.0$) in linear scale.

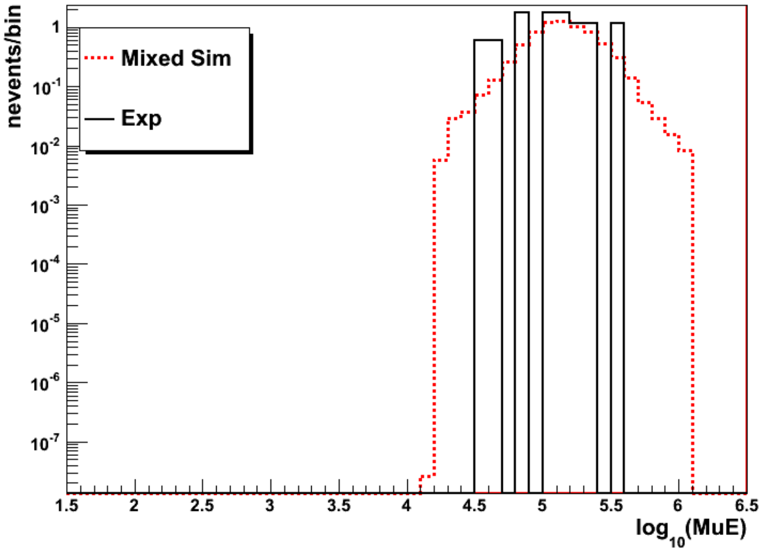


Figure I.40: Foward folding result of MuE of bin 10 ($7.8 \le \log_{10} E_{reco} < 8.0$) in logarithmic scale.

Bin 11:

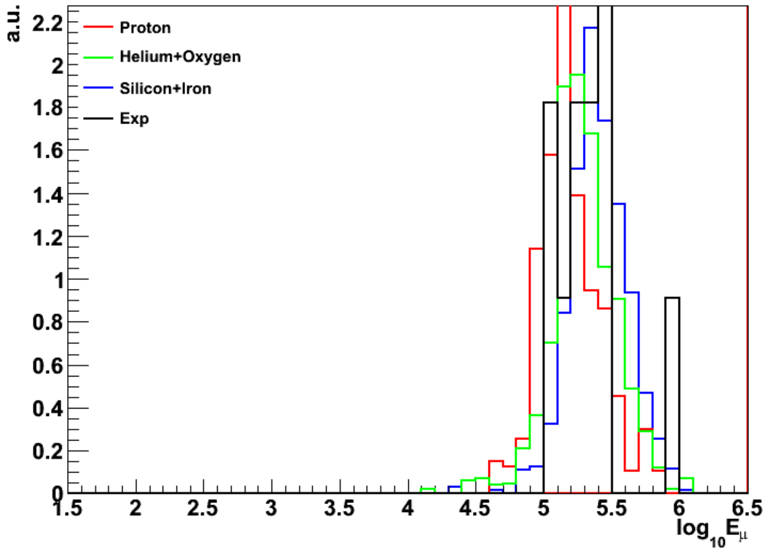


Figure I.41: MuE distribution of bin 11 ($8.0 \leq \log_{10} E_{reco} < 8.2$) in linear scale.

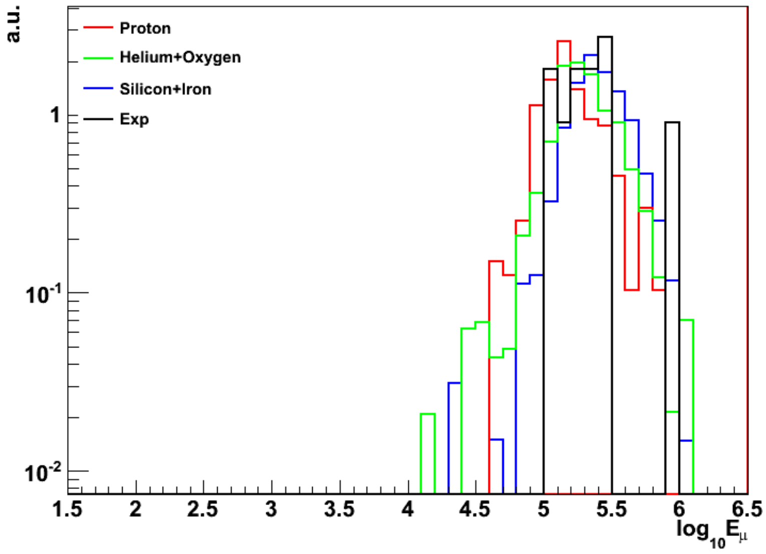


Figure I.42: MuE distribution of bin 11 ($8.0 \leq \log_{10} E_{reco} < 8.2$) in logarithmic scale.

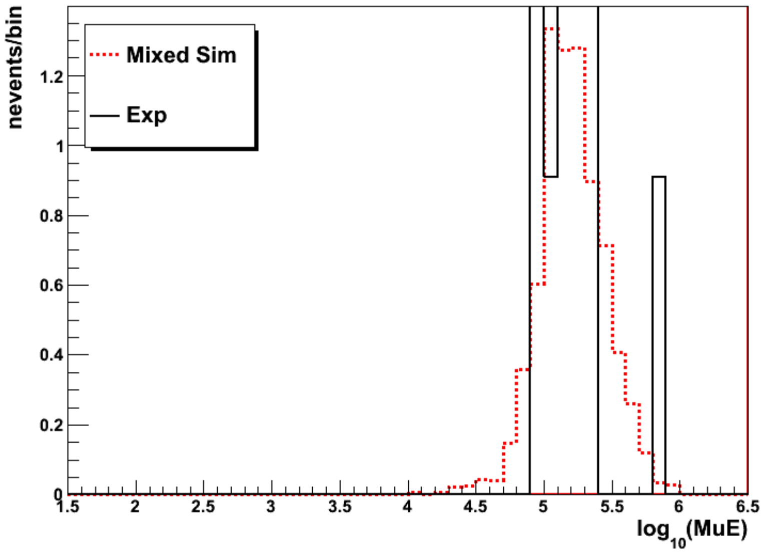


Figure I.43: Foward folding result of MuE of bin 11 ($8.0 \leq \log_{10} E_{reco} < 8.2$) in linear scale.

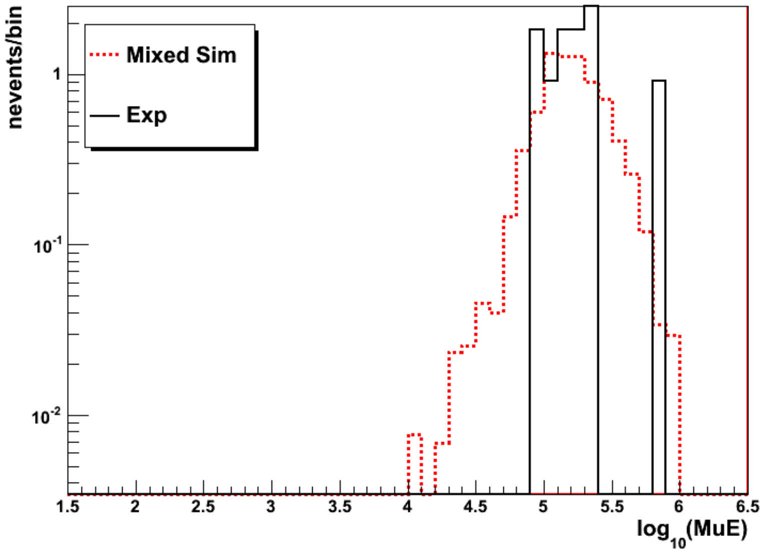


Figure I.44: Foward folding result of MuE of bin 11 ($8.0 \leq \log_{10} E_{reco} < 8.2$) in logarithmic scale.