Cosmic Ray Composition using SPASE-2 and AMANDA-II

K. Andeen[*] **and K. Rawlins**[†] **For the IceCube Collaboration**[‡]

[*]*University of Wisconsin-Madison, 1150 University Ave, Madison, WI*
[†]*University of Alaska Anchorage, 3211 Providence Dr, Anchorage, AK*
[‡]*See special section of these proceedings*

Abstract. **The precise measurement of cosmic ray mass composition in the region of the knee (3 PeV) is critical to understanding the origin of high energy cosmic rays. Therefore, air showers have been observed at the South Pole using the SPASE-2 surface array and the AMANDA-II neutrino telescope which simultaneously measure the electronic air shower component at the surface and the muonic air shower component in deep ice, respectively. These two components, together with a Monte Carlo simulation and a well-understood analysis method will soon yield the relative cosmic ray composition in the knee region. We report on the capabilities of this analysis.**

Keywords: composition, cosmic-ray, neural-network

I. INTRODUCTION

The mass composition of high-energy cosmic rays around and above the knee in the energy spectrum ($\sim$3 PeV) is dependent upon the mechanisms of cosmic ray production, acceleration, and propagation. Therefore, the study of mass composition is critical to understanding the origins of cosmic rays in this energy region. At energies up to 10^{14} eV mass composition can be measured directly using balloon and satellite experiments; however, due to the low flux, composition above 10^{14} eV must be obtained from indirect measurements on the ground. Indirect measurements of composition involve a close examination of the air shower produced as a cosmic ray primary smashes into Earth's atmosphere. By utilizing information from more than one component of the shower, such as the electronic and muonic components, the energy and relative mass can be obtained from primary particles with much higher energies than those currently measurable by direct detection techniques.

II. DATA AND RECONSTRUCTION

One such indirect measurement is possible using the South Pole Air Shower Experiment (SPASE-2) in coincidence with the Antarctic Muon And Neutrino Detector Array (AMANDA-II). The SPASE-2 detector is situated on the surface of the South Pole at an atmospheric depth of $\sim$685 g cm^{-2} and is composed of 30 stations in a 30 m triangular grid. Each station contains four 0.2 m^2 scintillators. The AMANDA-II detector lies deep in the ice such that the center-to-center separation between the deep ice and the surface arrays is $\sim$1730 m with

an angular offset of 12°. AMANDA-II consists of 677 optical modules (OMs) deployed on 19 detector strings spanning depths from 1500-2000 m below the surface of the ice. Each OM contains a photomultiplier tube (PMT) which is optimized for detection of the Cherenkov light emitted by particles—namely muon bundles—passing through the ice. In addition to the composition analysis, this coincident configuration allows for calibration as well as measurement of the angular resolution of the AMANDA-II detector. [1]

For this analysis, coincident data from the years 2003-2005 are used, for a total livetime of around 600 days. For comparison with the data, Monte Carlo simulated proton, helium, oxygen, silicon and iron air showers with energies between 100 TeV and 100 PeV have been produced using the CORSIKA air shower generator with the SIBYLL/GEISHA hadronic interaction models [2]. At the surface the air showers are injected into GEANT4, which simulates the SPASE-II detector response [3]. The showers are then propagated through the ice and the response of AMANDA-II detector is simulated using the standard software package of the AMANDA collaboration. An E^{-1} spectrum is used for generation, but for analysis the events are re-weighted to the cosmic ray energy spectrum of E$^{-2.7}$ at energies below the knee at 3 PeV and E$^{-3.2}$ above. Both the experimental data and the Monte Carlo simulated data are then put through identical reconstruction chains.

The first step in the reconstruction uses information from SPASE-2 only. The goal of this first reconstruction is to find the shower direction, shower size, and core position of the incoming air shower. The direction can be computed from the arrival times of the charged particles in the SPASE-2 scintillators, while the shower core position and shower size are acquired by fitting the lateral distribution of particle density to the Nishimura-Kamata-Greisen (NKG) function. Evaluating the fit at a fixed distance from the center of the shower (in this case 30 m) [4] gives a parameter called S30, which has units of particles/m^2 and will be used throughout this paper as a measure of the electronic part of the air shower.

The next step in the reconstruction provides a measure of the muon component of the air shower from a combined reconstruction which uses both the surface and deep ice detectors. The core position of the shower from the SPASE-only fit is held fixed while θ and ϕ are varied in the ice to find the best fit of the muon track in the AMANDA-II detector. Holding the core fixed at

the surface allows for a lever arm of about 1730 m when calculating directionality, providing a very tight angular resolution for the track. The expected lateral distribution function (LDF) of the photons resulting from the muon bundle in AMANDA-II is computed and corrected for both the ranging out of muons as they progress downward through the detector, as well as the changing scattering length as a function of depth in the ice caused by dust layers. The LDF is then fit to the hit optical modules and evaluated at a perpendicular distance of 50 m from the center of the shower [5]. This parameter, called K50, has units of photoelectrons/OM and will be used throughout the rest of this paper as the measure of the muon component of the air shower.

III. ANALYSIS DETAILS

Once reconstruction has been completed it is important to find and eliminate poorly reconstructed events. Thus events have been discarded which:

- have a reconstructed shower core outside the area of SPASE-2 or a reconstructed muon track passing outside the volume of AMANDA-II,
- have an unreasonable number of hits in the ice given S30 at the surface (these events represent large showers which landed outside of SPASE-2 and were misreconstructed within the array as having a small S30),
- have an unphysical reconstructed attenuation length of light in the ice (an unphysical reconstruction of attenuation length will lead to a misreconstructed value for K50), or
- are reconstructed independently in SPASE-2 and AMANDA-II as coming from significantly different locations in the sky.

After these cuts have been made, it can be seen in Fig. 1 that our two main observables, S30 and K50, form a parameter space in which primary energy and primary mass separate. This is expected, since the showers associated with the heavier primaries develop earlier in the atmosphere and hence have more muons per electron by the time they reach the surface than the showers associated with lighter primaries [6]. This means that K50, which is proportional to the number of muons in the ice, will be higher for heavier primaries than for lighter primaries of the same S30, as is observed.

In the three-year data set used for this analysis, more than 100,000 events survive all quality cuts. It is interesting to notice that in the previous analysis, using the SPASE-2/AMANDA-B10 detector [5], the final number of events for one year was 5,655. Furthermore, the larger detector used here is sensitive to higher energy events. The significant increases in both statistics and sensitivity, along with a new detector simulation and revised reconstruction algorithm for the SPASE-2 array, are the basis for performing a new analysis.
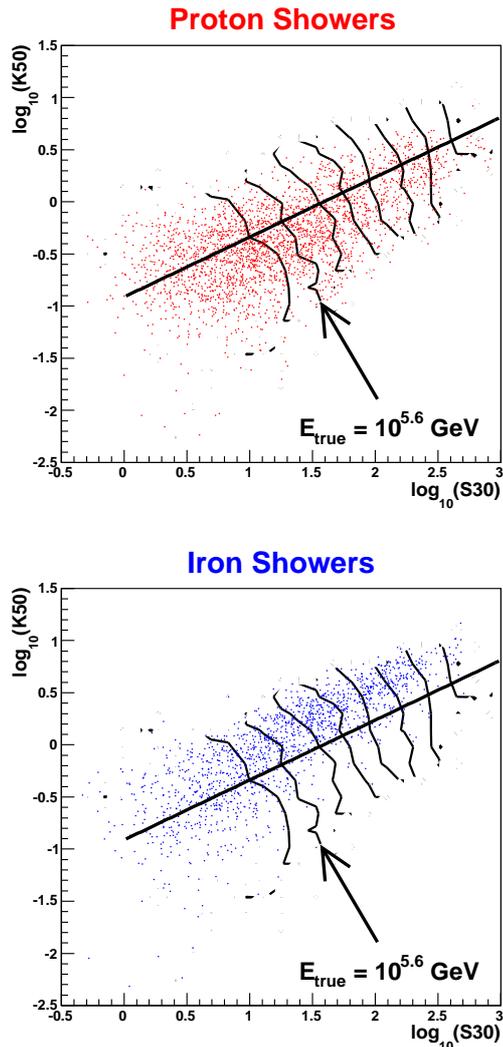

Fig. 1. The two main observables, $\log_{10}(K50)$ vs $\log_{10}(S30)$, in the Monte Carlo simulation with protons above (red) and iron below (blue). The black contours depict lines of constant energy from $5.4 > \log_{10}(E_{true}/GeV) > 6.8$ marked every $\log_{10}(E_{true}/GeV) = 0.2$. The black line along the energy gradient approximates a division between proton and iron showers and is included merely as a reference between the two plots. It is clear that mass and energy are on a roughly linear axis.

A. Calibration

To accurately measure the composition using both electron and muon information, the Monte Carlo simulations must provide an accurate representation of the overall amplitude of light in ice (measured here as K50). However, due to the model-dependencies displayed by air shower simulations, the overall light amplitude is subject to systematic errors. Therefore, it is important to calibrate the composition measurements at low energies where balloon experiments have provided direct measurements of cosmic ray composition. In light of this, a vertical "slice" in S30 is selected for calibration which corresponds with the highest energies measured directly. At these energies, the direct measurements indicate that $<\ln A> \approx 2$, or 50% protons and 50% iron [7]. The K50 values of the data in this S30 "slice" are thus adjusted by
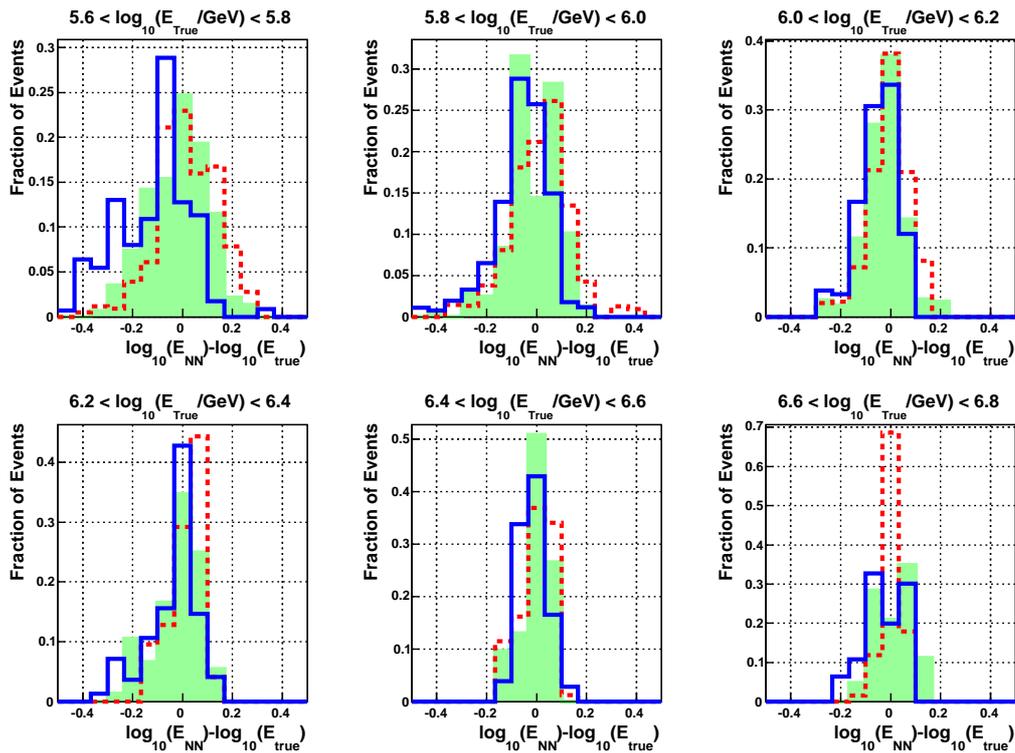
Fig. 2. Energy resolution (the difference between the true primary energy and the energy reconstructed by the neural network) is shown in bins of true energy for iron (solid blue), proton (dashed red) and oxygen (shaded green) primaries. Each energy bin is bounded by two consecutive contours from Fig. 1 (where the indicated energy contour is the lower bound of the first energy bin above). For easier comparison, a Gaussian distribution has been fit to these energy resolution histograms and the mean and sigma of each Gaussian can be found in Table 1.

TABLE I

MEAN AND SIGMA OF A GAUSSIAN DISTRIBUTION FIT TO EACH ENERGY RESOLUTION HISTOGRAM IN FIG. 2.

Shower Type	Gaussian Statistic	True Energy Bins ($\log_{10}$(E) / GeV)					
		5.6 - 5.8	5.8 - 6.0	6.0 - 6.2	6.2 - 6.4	6.4 - 6.6	6.6 - 6.8
Proton	Mean	0.03	0.02	0.00	0.02	0.01	0.00
	σ	0.12	0.11	0.08	0.06	0.05	0.04
Oxygen	Mean	-0.02	-0.01	-0.03	0.00	0.01	0.02
	σ	0.12	0.10	0.07	0.07	0.05	0.09
Iron	Mean	-0.12	-0.05	-0.04	-0.02	-0.01	0.05
	σ	0.15	0.09	0.08	0.08	0.06	0.15

an offset to match the distribution of K50 corresponding to a $50\% - 50\%$ proton and iron sample.

B. Neural Network Reconstruction of Energy and Mass

Similar past investigations utilized the quasi-linear relationship between K50/S30 and mass/energy, as seen in Fig. 1, to employ an analysis wherein, after calibration, the axis is rotated to correspond to the mass/energy coordinate plane [5] [8]. The rotation analysis works quite well up to energies slightly above the knee; however, beyond the knee the relationship between K50/S30 and mass/energy becomes increasingly non-linear as the air showers approach the energy where the shower maximum occurs at the atmospheric depth of the SPASE-2/AMANDA-II detectors. As the data set used here has significantly more statistics at these higher energies than previous studies, it was important

to find a new procedure for extracting the composition. A neural network technique has therefore been developed to resolve the mean logarithmic mass at all energies [9].

The main change to the neural network technique since its development has been to distinguish between the calculation for energy and the calculation for mass by using two separate networks. The first neural network (NN1) is trained to find the primary energy by using $\log_{10}$(K50) and $\log_{10}$(S30) as input parameters, followed by one hidden layer. The second network (NN2) is trained to find the primary mass of the air shower. NN2 also takes as input $\log_{10}$(K50) and $\log_{10}$(S30) and also has a single hidden layer of neurons. In both cases the network is trained through a number of "epochs", or training cycles, on half of the simulated proton and iron showers and tested on the other half of the proton and iron showers. The results of testing determine the
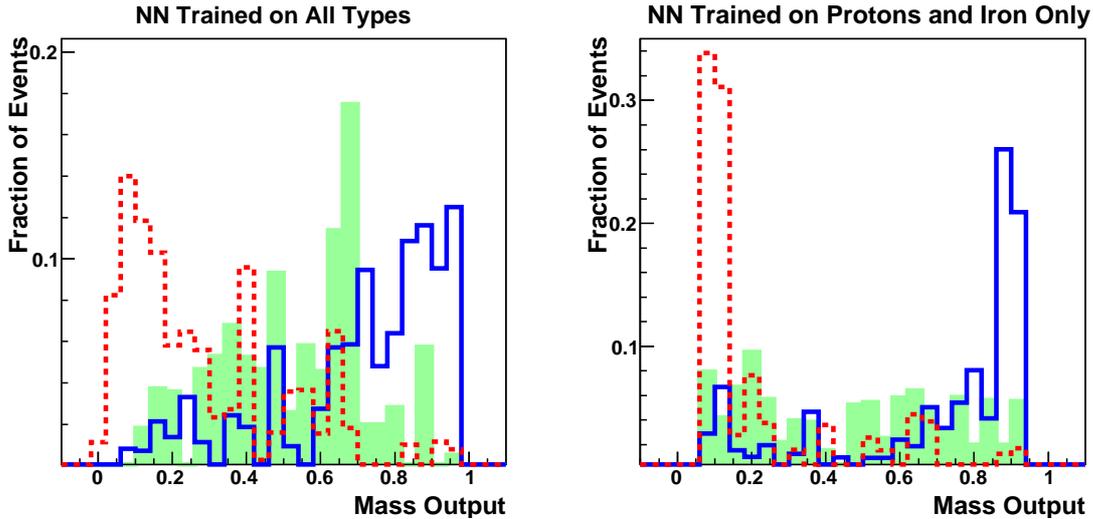
Fig. 3. Mass output of the neural network from the bin $6.2 < \log_{10}(E_{reco}/\text{GeV}) < 6.4$ for three primary types: iron (solid blue), oxygen (shaded green), and protons (dashed red). On the left is the output from a network trained on all particle types. On the right is the output from a network trained only on protons and iron. As expected, when the neural network is trained on more particle types it begins to reconstruct the intermediate primaries in their proper location as opposed to attempting to classify them as proton or iron.

number of "epochs" each network can be trained through without overtraining. (The intermediate primaries are currently used only for checking the mass reconstruction of the network. It is hoped that they will soon be plentiful enough to use as inputs for training as well.)

As NN1 is given a full spectrum of energies on which to train, it very successfully reconstructs the energy of each shower. Plots of energy resolution separated into bins of true energy can be found in Fig. 2. The energy resolution is not only very good but also composition independent. This can be seen more clearly in Table I, which shows the Gaussian mean and sigma of each energy bin from Fig. 2. (Currently only six bins in energy are shown: higher energies are being generated and will be available before the conference.)

The mass network outputs a reconstructed mass for each particle in terms of the primaries on which it has been trained. As the neural network is trained only on proton and iron showers, it reconstructs each shower as some combination of proton and iron. Currently a minimization technique is used to find the mean log mass for each energy bin. This minimization technique has been tested on intermediate primaries and proves to reconstruct them reasonably well for a network trained only on protons and iron. However, as the number of intermediate primaries generated is increasing, it is hoped that the mass network can soon be trained on a wider spectrum of primaries. A test of this has been run and a comparison between the output from a network trained on all particle types and a network trained only on protons and iron is shown in Fig. 3. As expected, it is evident that when the network is trained on intermediate primaries oxygen is reconstructed in its own location between protons and iron and no longer as a fraction of one or the other. This method appears very promising and, as more intermediate primaries are simulated, this is the direction in which the analysis will proceed.

IV. DISCUSSION AND OUTLOOK

The SPASE-2/AMANDA-II cosmic ray composition analysis has lately acquired new Monte Carlo simulation, a new detector simulation of the surface array and a revised surface-array reconstruction algorithm. Aided by these three new features a great deal of progress has been made. It can clearly be seen that the use of a modified version of the neural network technique seen in the ICRC proceedings of 2007 [9] can very accurately reconstruct the energy of cosmic ray primaries in the region of the knee in the cosmic ray spectrum. The inclusion of a larger variety of primary particles for training the neural network is seen to be very promising and, with increased statistics in the Monte Carlo simulation, a composition result will soon follow.

V. ACKNOWLEDGMENTS

REFERENCES

[1] J. Ahrens et al. *Nuclear Instruments and Methods*, 552:347-359, 2004.
[2] D. Heck and T. Pierog, *Extensive Air Shower Simulations with CORSIKA: A User's Guide* http://www-ik.fzk.de/corsika/ usersguide/corsika_tech.html
[3] GEANT4 Collaboration *Nuclear Instruments and Methods, A*, 506:250-303, 2003.
[4] J.E. Dickinson et al. *Nuclear Instruments and Methods A*, 440:95-113, 2000.
[5] K. Rawlins. PhD thesis, University of Wisconsin-Madison, (2001)
[6] T. Gaisser. *Cosmic Rays and Particle Physics*. Cambridge University Press, 1988.
[7] J.R. Hörandel. *International Journal of Modern Physics A*, 20(29):6753-6764, 2005.
[8] J. Ahrens et al. *Astroparticle Physics*, 21: 565-581, 2004.
[9] K. Andeen et al, *Proceedings of the 30th International Cosmic Ray Conference* 2007.