PROCEEDINGS OF SCIENCE

Energy and $X_{\max}$ Reconstruction for Cosmic-Ray Events Recorded by a Prototype Station of the IceCube Surface Enhancement

Roxanne Turcotte-Tardif[a,*], **Frank G. Schröder**[a,b] **for the IceCube Collaboration**[c]

[b] *Bartol Research Institute, Department of Physics and Astronomy, University of Delaware, Newark, DE, 19716, USA*

[a] *Karlsruhe Institute of Technology, Institute for Astroparticle Physics, 76021 Karlsruhe, Germany*

[c] *Full author list at* https://icecube.wisc.edu/collaboration/authors/

E-mail: roxanne.turcotte-tardif@kit.edu, fgs@udel.edu

The IceTop array, located at the surface of the IceCube Neutrino Observatory at the South Pole, is currently used as a veto for the in-ice neutrino detection as well as a cosmic-ray detector. Over the years, additional snow has accumulated on the IceTop detectors leading to a reduction in sensitivity and resolution. In order to mitigate this issue as well as further increase the accuracy of cosmic-ray measurements, a detector enhancement is planned in the next few years. The enhanced array will consist of 32 stations, each comprising 8 scintillation detectors and 3 radio antennas, and will span an area of approximately 1 km^2. Specifically, upgrading IceTop with radio antennas will provide precise $X_{\max}$ measurements, a variable widely used to reconstruct the composition of cosmic rays.

In January 2020, a complete prototype station was deployed at the South Pole. Following the measurement of cosmic-ray events with the antennas, we developed the tools necessary to use a template-matching method for energy and $X_{\max}$ reconstruction. This template method uses Monte-Carlo simulations and compares them to recorded data. Thus, a set of simulated air showers is created using air shower parameters reconstructed by IceTop as input to the CORSIKA/CoREAS simulation software for each of the measured events. In this work, the method is applied to measured events, and we will show that it works for a third of the events in the sample. Subsequently, the technique is modified into a log-likelihood minimization, and the results obtained with a simulations-only study of different parameters are discussed. This work concludes with a confirmation that the template method works for air showers recorded with a frequency band of 80 to 300 MHz, and shows that the log-likelihood method has the potential of increasing the accuracy of $X_{\max}$ reconstruction for the complete planned array.

9th International Workshop on Acoustic and Radio EeV Neutrino Detection Activities - ARENA2022
7-10 June 2022
Santiago de Compostela, Spain

*Speaker

PoS(ARENA2022)048

1. Introduction

The IceCube Neutrino Observatory [1] is located at the geographic South Pole and uses the clear ice of Antarctica to detect astrophysical neutrinos. A surface array of ice-Cherenkov tanks, covering an area of 1 km^2 complements the in-ice detector and makes the observatory a powerful cosmic-ray detector. Over the years, additional snow has accumulated over the tanks, reducing their sensitivity and increasing the reconstruction uncertainties. A hybrid array of scintillation panels and radio antennas was designed to mitigate the snow accumulation. The radio antennas will enable an X_{max} reconstruction via the footprint left on the ground from the radio emission in the air showers.

The planned enhanced surface array will be composed of 32 stations spread over 1 km^2 and each station will feature 8 scintillation panels and 3 radio antennas. A complete prototype station was deployed at the South Pole in January 2020, details of the station design can be found in [4]. First coincident measurements of cosmic-ray air showers between the scintillation detectors, the radio antennas, and the IceTop Cherenkov tanks were obtained [5]. From those first coincident measurements, 16 air shower events are used to develop an energy and X_{max} reconstruction based on the state-of-the-art method initially developed by LOFAR [8].

2. Description of the template-fitting reconstruction method

The idea behind the template-fitting method is to compare the two-dimensional lateral distribution function (LDF) of Monte Carlo simulations to the measurements. Therefore, for each event measured with the radio antennas, a set of specific simulations is created for that particular event. The air shower simulations are done with CORSIKA [11] and the radio emission with CoREAS [12]. To transform the simulated electric field into a measurements of the two polarisation channels of each antenna, the instrumental response is folded in the simulations using radcube [3], a specific framework within the IceCube software designed to treat radio data, and filtered between 80-300 MHz. Subsequently, a quantity extracted from the simulated waveform is compared to the same quantity extracted from the measured waveforms. Here, the power in a 50 ns window, defined by

$$P = \frac{\Delta t}{Z\, t_{\mathrm{tot}}} \sum_{i}^{b_{\mathrm{win}}} A_i^2 \tag{1}$$

where Z is the impedance (50 Ω), Δt the sampling time (1 ns), A_i the amplitude of the waveform at bin i, and b_{win} the number of bins over which the summation is made (50 bins $\widehat{=}$ 50 ns), is used. The window of 50 ns is chosen to contain most of the pulses width while limiting the noise-only contribution.

The comparison is implemented in a first step with a χ^2 minimization method inspired from other experiments, e.g. [6, 8]. The polarisations channels of the antennas are treated independently, thus, it follows that for each event, the minimization is done on 6 waveforms. Assuming that the power contained in the noise of the measured waveform (σ_p) can be directly removed from the measured signal (P) and compared to the simulated signal (P_{MC}), it follows that,

$$\chi^2_{\mathrm{power}} = \sum_{i}^{n=6} \left(\frac{(P_i - \sigma_{p_i}) - f^2 \cdot P_{\mathrm{MC}_i}}{\sigma_{p_i}} \right)^2 \tag{2}$$

where i defines the antenna channel where the signal is measured and f is the variable to minimize. The noise contribution σ_{p_i} is the average of the power calculated in multiple time windows of 64 ns of the measured waveform i (where 64 ns is obtained from simulation studies [15]). The variable f is proportional to the radiation energy [8]. However, due to the high altitude of the experiment (~3 km above sea level, or ~690 g/cm^2), some air showers do not fully develop before reaching the detectors. Therefore, as a first estimate, the energy of the air shower is reconstructed by

$$E_{\mathrm{CR}}^{\mathrm{reco}} = f \cdot E_{\mathrm{CR}}^{\mathrm{MC}} \tag{3}$$

where $E_{\mathrm{CR}}^{\mathrm{MC}}$ is the primary energy of the simulations, i.e. the energy from the initial reconstruction with IceTop and $E_{\mathrm{CR}}^{\mathrm{reco}}$ is the estimated primary energy of the cosmic ray reconstructed using the template-fitting method. The reconstruction of X_{max} is done by equaling the χ^2 values of each simulation in a set as function of their respective X_{max}, an example is shown later. Due to the small number of antennas, the core is not let free to fit.

3. Initial reconstruction and event selection

The sample of events in this analysis is composed of 16 air showers recorded by the prototype station between 2020-10-22 and 2021-01-26. These showers are in temporal and spatial coincidence with measurements from the IceTop tanks and the scintillation detectors. There is no particular quality cut applied to the selection, besides that the pulse position must be less than 3.5 ns from its expected position calculated from the IceTop reconstructed air shower direction. The arrival direction of the air showers (zenith and azimuth) and the shower core position used as input for the Monte Carlo simulations are obtained with the IceTop standard reconstruction [2]. Regarding the air shower energy, the IceTop reconstruction is only calibrated for vertical air showers (i.e. for zenith angles θ of $\cos(\theta)$ between 0.8 and 1.0), thus an extrapolation of the calibration is done for the more inclined air showers in the sample, which gives a coarse estimate of the initial energy of the air showers. The energy of the air shower sample is estimated to range between ~50 PeV and ~500 PeV. The reconstructed core position and arrival direction obtained from the IceTop signals for all events in the sample is shown in Figure 1. For each event, a set of 100 CORSIKA simulations is created, which comprises 50 protons and 50 iron nuclei as primary particles to probe a large X_{max} phase-space.

4. Results on measured radio events

The template-fitting method described above is applied to measured radio data from the prototype station. The pulse position, for the simulations as well as for the measurements, is obtained by creating a 200 ns search window around the expected position of the signal (known from the IceTop geometry reconstruction). Within that search window, the position of the maximum of the Hilbert envelope [13] is identified which thereafter defines the pulse position. A symmetrical window totaling 50 ns is created around the pulse position and the power within that window is calculated.

For one measured event, the χ^2 minimization from Equation (2) is applied individually to every simulation in the associated set. Then, a parabola is fit on χ^2 as function of X_{max} , which can be seen in the left plot of Figure 2. The X_{max} value corresponding to the minimum of the parabola

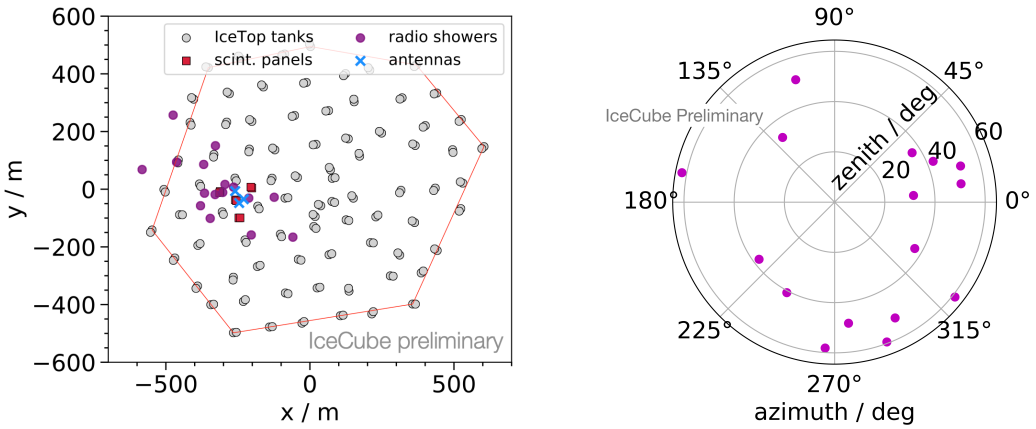

Figure 1: Left: core positions of the air shower event sample relative to the IceTop detector geometry with the prototype station included. Right: arrival directions of the air showers in the sample. Both the shower core positions and the arrival directions are obtained with IceTop reconstruction.

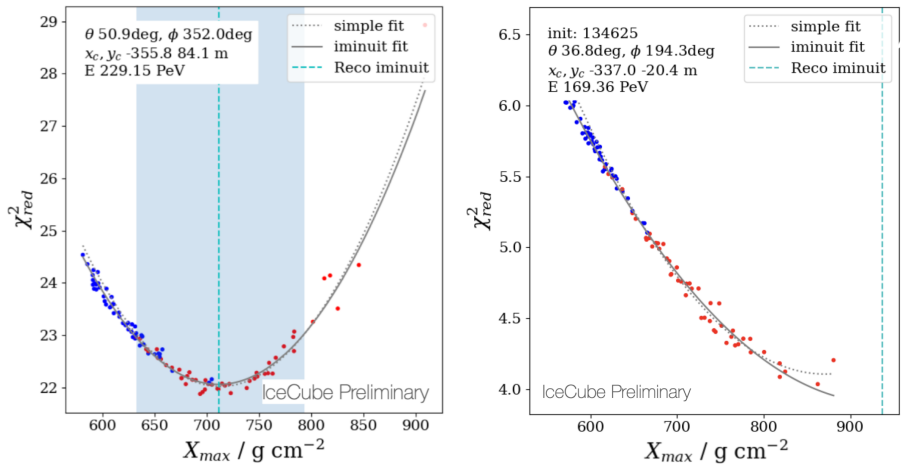

Figure 2: The left and right plots represent the χ^2 plotted against the $X_{\max}$ of each simulations in the set. This shows two different measured events, where the right one exhibits the characteristic parabola and the left one falls outside the simulated phase-space.

represents the reconstructed $X_{\max}$ for the measured event. From the 16 events about a third of the sample has a contained parabola, i.e. both sides of the parabola have data points. The remaining two-thirds fall outside the simulated $X_{\max}$ phase-space. This does not mean that the reconstruction method does not work, only that the initial IceTop reconstruction is too far from the reality and thus the simulations do not overlap with the measured $X_{\max}$ phase-space. An example is shown on the right plot of Figure 2 where the generated set of simulations does not encompass the $X_{\max}$ phase-space of the measured event.

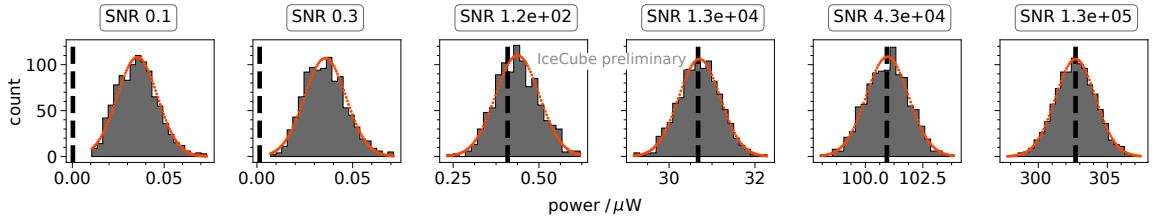

Figure 3: Distributions of the power (P) in a 50 ns window in the waveforms with noise injected. Each distribution correspond to one true pulse and its value is indicated by the vertical line. The distributions are ordered by increasing signal-to-noise values.

5. Migration to a log-likelihood

To evaluate the effects of noise on the signal, two Monte Carlo simulations using an 8-arms star pattern, as described in [3], are used. The idea behind this is to obtain multiple different signal strengths. Treating the two antenna polarisation channels independently results in 320 signal traces for each of the two simulations. To ensure that the geometry of the air shower does not influence the results, the two simulations have different geometries: one has an energy of $1.24 \times 10^{18.0}$ eV, a zenith of 46.7°, an azimuth of 19.5°, and a proton as primary particle. The other has an energy of $3.9 \times 10^{18.0}$ eV, a zenith angle of 67.3°, an azimuth angle of 165.8°, and an iron nucleus as primary.

For each signal trace, 1000 different noise samples are injected. The injected noise consists of the minimal noise expected at the South Pole folded with the instrumental response of the prototype station. More precisely, it is a combination of the non-thermal Galactic radio emission calculated using the Cane model [9], and a 40 K thermal noise from the low-noise amplifier located at the top of the antenna [10]. The spectrum of this modeled noise agrees well in shape and amplitude with the spectrum measured by the antennas of the prototype station except for radio interference that is not present in the model [4]. The power extracted from a thousand noisy waveforms stemming from one true pulse is depicted in each histogram, see Figure 3. In this figure, six histograms from six different pulses are shown, where the true value is indicated by the vertical line. The signal-to-noise ratio (SNR) is defined as the the maximum of the true pulse over the averaged noise of all noise measurements squared, $\mathrm{SNR} = (S_{\mathrm{true}}/\langle N_{\mathrm{RMS}}\rangle)^2$.

Subsequently, a Gaussian fit is applied to each distribution. In Figure 4, the mean (top) and the standard deviation (bottom) are shown as function of the true power of the associated waveform. The functions are parameterized as broken power laws of the form:

$$P'(f, P_{\mathrm{MC}}) = a_0 \left(1 + \left(\frac{f \cdot P_{\mathrm{MC}}}{x_0}\right)^{k_0}\right)^{(b_0/k_0)} \quad \text{and} \quad \sigma'(f, P_{\mathrm{MC}}) = a_1 \left(1 + \left(\frac{f \cdot P_{\mathrm{MC}}}{x_1}\right)^{k_1}\right)^{(b_1/k_1)}. \tag{4}$$

The parameters $a_{0,1}$, $b_{0,1}$, $k_{0,1}$, and $x_{0,1}$ are the parametrisation constants for $P'(f, P_{\mathrm{MC}})$ and $\sigma'(f, P_{\mathrm{MC}})$. Continuing with the assumption that the distributions are Gaussian, the χ^2 can be express as a more general log-likelihood function (LLH) of the form

$$-2 \ln \mathcal{L}_{\mathrm{power}} = \sum_i^{n=6} \left(\ln\left(2\pi\sigma'^2(f^2, P_{\mathrm{MC}})\right) + \left(\frac{P_i - P'_{\mathrm{MC}_i}(f^2, P_{\mathrm{MC}})}{\sigma'(f^2, P_{\mathrm{MC}})}\right)^2 \right). \tag{5}$$

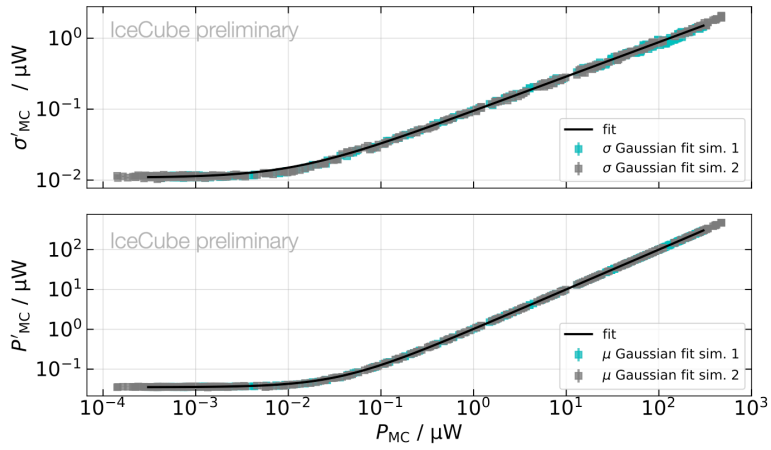

Figure 4: Standard deviation (top) and mean (bottom) of the Gaussian fit on the power distributions, from 1000 waveforms with noise injected as function of the true value of the power of the associated noiseless pulse. The error bars (not visible) represent the error on the fit.

The equation is similar to the χ^2 , where the first term is omitted. As seen, in Figure 4, the variable σ changes with the value to minimize.

6. Simulation study

To compare the χ^2 method to the log-likelihood, to investigate different frequency bands, as well as to estimate the accuracy for the complete array, a simulation study is performed. The simulation study consists of transforming one simulated event out of a simulation set to mock data by injecting modeled noise. This mock event is then compared to the other noiseless simulations in the set. In this fashion, a reconstructed energy (E_{reco}) and a reconstructed X_{max} ($X_{\text{max}_{\text{reco}}}$) are obtained and can be compared to their true values. This step is repeated for each simulation in the set. Finally, an estimation of the resolution and accuracy for a specific event is obtained.

To compare the methods and bands, the mock data technique is applied to every of the 16 simulation sets. The resulting difference between the E_{true} ($X_{\text{max}_{\text{true}}}$) to E_{reco} ($X_{\text{max}_{\text{reco}}}$) is plotted in a cumulative histogram. A Laplacian function is fitted on the distributions and the $\sqrt{2}b$ (which corresponds to one standard deviation σ) and the mean are extracted. The results are shown in Figure 5.

For the prototype station, the χ^2 and the LLH methods result in a comparable resolution of approximately 13% in energy and 60 g/cm^2 in X_{max} . However, for the complete surface enhancement array, the X_{max} reconstruction from the LLH is around 5 g/cm^2 smaller than from the χ^2. The lower frequency band of 50-100 MHz is used to compare to the commonly used 30-80 MHz band of the main radio arrays. The frequencies below 50 MHz for our experiment are strongly attenuated which is the reason why the band of 50-100 MHz is chosen instead. Nevertheless, for the 3 antennas of the prototype station as well as for the 32 planned stations, the higher frequency band performs noticeably better than the lower band. In the high-frequency band, a negligible bias is present for the energy as well as X_{max} .

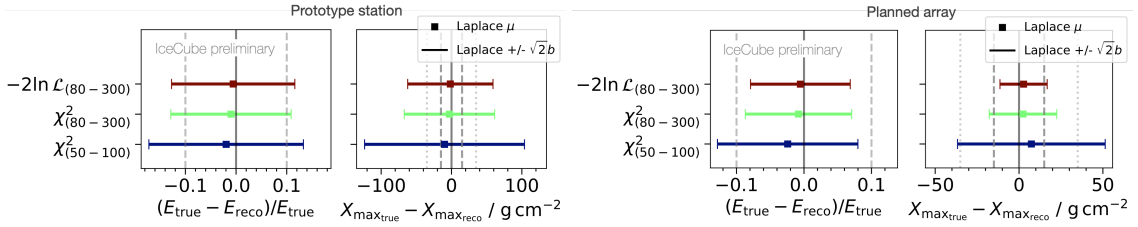

Figure 5: Mean and standard deviation of a Laplacian fit on the cumulative histogram of the reconstruction to the true value of 16 air showers. On the left, for the prototype station, and on the right for the planned complete array of the surface enhancement.

7. Discussion

The template-fitting method is applied for the first time on only three antennas and in a frequency band of 80-300 MHz. The results for the higher frequency band are promising for the reconstruction of $X_{\max}$ with the radio part of the surface array enhancement with a possible resolution of 15 g/cm^2 compared to ~40 g/cm^2 for the lower band. The better resolution is most likely attributable to the sharper gradient close to the Cherenkov-like ring feature of the footprint in higher frequencies [7]. It is important here to remember that these results do not account for the uncertainties in the initial reconstruction, i.e. the mock data was simulated with exactly the parameter of the showers that it is compared to. The application of quality cuts may further improve the resolution. Regarding the three antennas reconstruction of $X_{\max}$, the method is limited by the resolution of the IceTop reconstruction.

Finally, the log-likelihood parametrization seems to describe better the low-to-medium pulse regime, i.e. when the pulse amplitude is close to the noise level. However, the parametrization is limited in the sense that it does not account for fluctuation in the noise level between different waveforms. This could be improved by adding the noise level of the trace as a variable in the parametrization. The last interesting observation is that for low signal-to-noise ratio, the distributions are not described perfectly by a Gaussian distribution, a skewness to the left can be seen. In this regime, a Gumbel function or a Rician function, as proposed in [14], could be used.

8. Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 802729). This work was performed on the HoreKa supercomputer funded by the Ministry of Science, Research and the Arts Baden-Württemberg and by the Federal Ministry of Education and Research.

References

[1] M. G. Aartsen et al. The IceCube Neutrino Observatory: Instrumentation and Online Systems. *JINST*, 12(03):P03012, 2017.

[2] R. Abbasi et al. IceTop: The surface component of IceCube. *Nucl. Instrum. Meth. A*, 700:188–220, 2013.

[3] R. Abbasi et al. Framework and tools for the simulation and analysis of the radio emission from air showers at IceCube. *JINST*, 17(06):P06026, 2022.

[4] Rasha Abbasi et al. Development of a scintillation and radio hybrid detector array at the South Pole. *PoS*, ICRC2021:225, 2021.

[5] Rasha Abbasi et al. First air-shower measurements with the prototype station of the IceCube surface enhancement. *PoS*, ICRC2021:314, 2021.

[6] Pedro Abreu et al. The depth of the shower maximum of air showers measured with AERA. *PoS*, ICRC2021:387, 2021.

[7] Balagopal V. Aswathi, Andreas Haungs, Tim Huege, and Frank G. Schröder. PeVatron Search Using Radio Measurements of Extensive Air Showers at the South Pole. *J. Phys. Conf. Ser.*, 1342(1):012006, 2020.

[8] S. Buitink et al. Method for high precision reconstruction of air shower X_{max} using two-dimensional radio intensity profiles. *Phys. Rev. D*, 90(8):082003, 2014.

[9] Hilary Cane. Spectra of the non-thermal radio radiation from the galactic polar regions. *Monthly Notices of the Royal Astronomical Society*, 189:465–478, 12 1979.

[10] E. de Lera Acedo, N. Drought, B. Wakley, and A. Faulkner. Evolution of skala (skala-2), the log-periodic array antenna for the ska-low instrument. *2015 International Conference on Electromagnetics in Advanced Applications (ICEAA)*, pages 839–843, 09 2015.

[11] D. Heck, J. Knapp, J. N. Capdevielle, G. Schatz, and T. Thouw. CORSIKA: A Monte Carlo code to simulate extensive air showers. 2 1998.

[12] T. Huege, M. Ludwig, and C. W. James. Simulating radio emission from air showers with CoREAS. *AIP Conf. Proc.*, 1535(1):128, 2013.

[13] Alan V Oppenheim. *Discrete-time signal processing*. Pearson Education India, 1999.

[14] H. Schoorlemmer et al. Energy and Flux Measurements of Ultra-High Energy Cosmic Rays Observed During the First ANITA Flight. *Astropart. Phys.*, 77:32–43, 2016.

[15] Roxanne Turcotte-Tardif. *Radio Measurement of Cosmic Rays at the South Pole*. PhD thesis, Karlsruhe Institute for Technology, 2022. Submitted.