Extraction of the Muon Signals Recorded with the Surface Detector of the Pierre Auger Observatory Using Recurrent Neural Networks

Juan Miguel Carceller*

Universidad de Granada

February 2, 2022

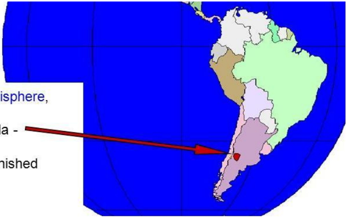
*Now at UCL & CERN working on NOvA and DUNE

Outline

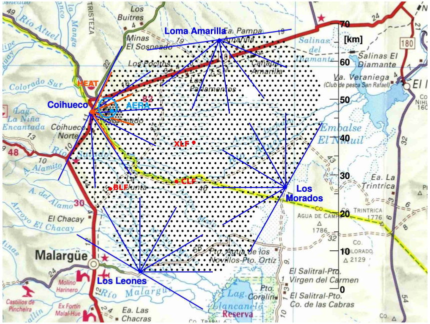
- The Pierre Auger Observatory and the Surface Detector (SD)
- Motivation: why do we want to know the muon signal?
- First work: simple neural network to predict the integral of the muon signal
- Results with LSTM, performance in simulations and comparison to data from other experiments

The Pierre Auger Observatory

- **Hybrid detector**
- Largest detector of cosmic rays built so far
- 1660 surface detectors located in a triangular array covering 3000 km^2
- The array is overlooked by 24 fluorescence telescopes
- Located near Malargüe, in the province of Mendoza in Argentina

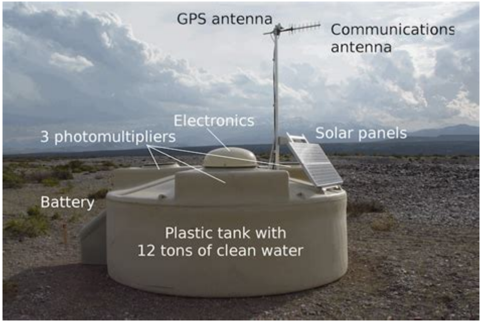


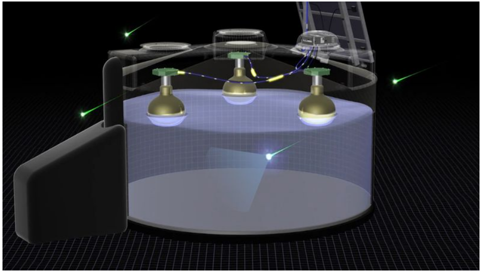
- Southern hemisphere, Malargüe - Pampa Amarilla - Argentina, construction finished mid 2007



The Surface Detector (SD)

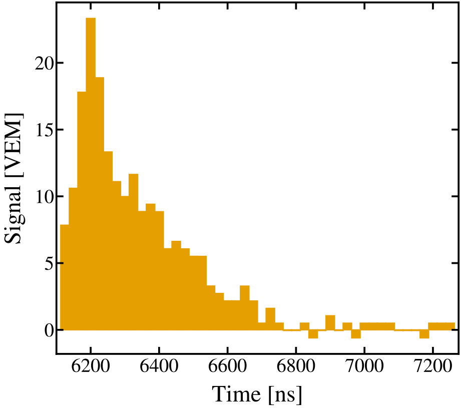
- Measures the arrival time of secondary particles of the shower at the ground
- These particles emit Cherenkov radiation in water that can be detected by the photomultiplier tubes
- Duty cycle $\sim 100\%$

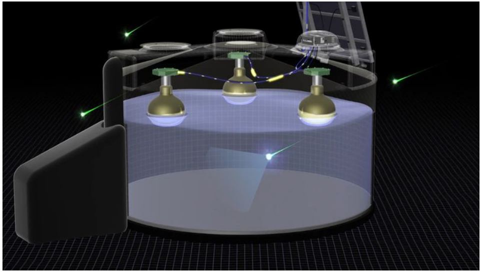




The Surface Detector (SD)

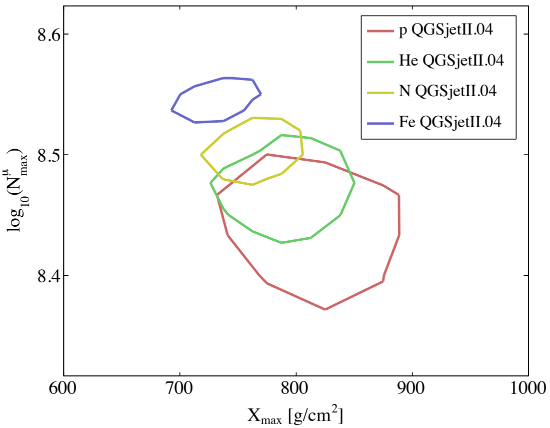
- Measures the arrival time of secondary particles of the shower at the ground
- These particles emit Cherenkov radiation in water that can be detected by the photomultiplier tubes
- Duty cycle $\sim 100\%$





Why is Knowledge about Muons Important?

- Infer information about mass composition
- Study hadronic interactions
- Help to understand differences between data and simulations



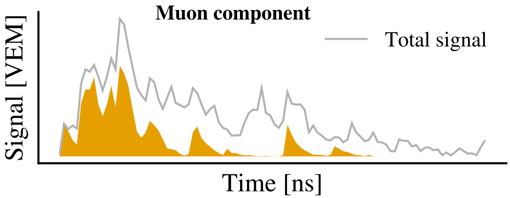
Muon and Electromagnetic Components

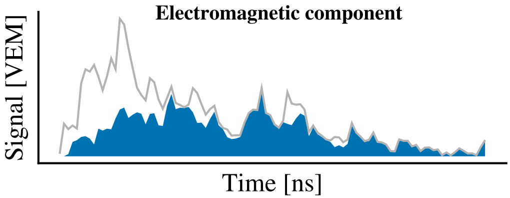
Muon component

- Earlier times
- Usually spiky

Electromagnetic component

- Later times
- Spread and not very spiky

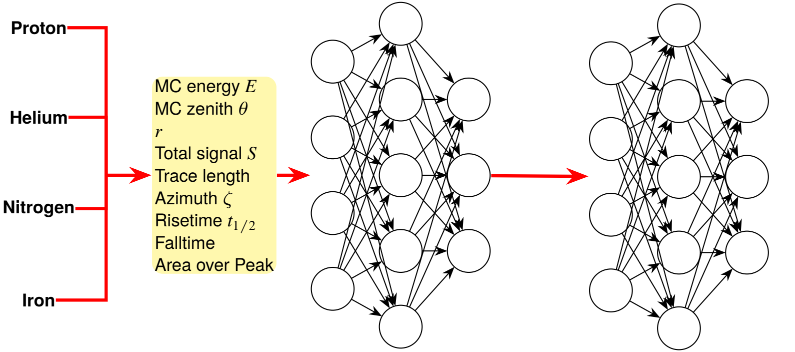




First Work

- In our first work (Astroparticle Physics **111** 12 (2019)) we predict the **integral of the muon signal**

Simulations (Input) → **Neural Network** → **Trained Neural Network** → **Predicted Muon Signal (Output)**



- We use more simulations (that the net has not seen) to test the prediction.

- **One prediction for each station: S^{μ}**

Neural Network

Input: 9 features → Fully connected → Units: 18 Activ.: ReLU → Fully connected → Units: 56 Activ.: ReLU → Fully connected → Units: 11 Activ.: ReLU → Fully connected → Units: 10 Activ.: ReLU → Fully connected → Units: 1 Activ.: Linear Output: $\widehat{S^{\mu}}$

- Inputs are normalized to be between 0 and 1
- The mean squared error is minimized

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{S^{\mu}}_i - S^{\mu}_i\right)^2$$

- $\text{ReLU}(x) = \max(0, x)$

New Method

- Soon after this first work, the method was improved
- The new method is able to **predict the whole temporal sequence** of the muon signal instead of only the integral, JINST 16 P07016 (2021)

Jinst

PUBLISHED BY IOP PUBLISHING FOR SISSA MEDIALAB

RECEIVED: *December 21, 2020*
ACCEPTED: *March 19, 2021*
PUBLISHED: *July 12, 2021*

Extraction of the muon signals recorded with the surface detector of the Pierre Auger Observatory using recurrent neural networks

PIERRE AUGER OBSERVATORY

Results: Example Traces

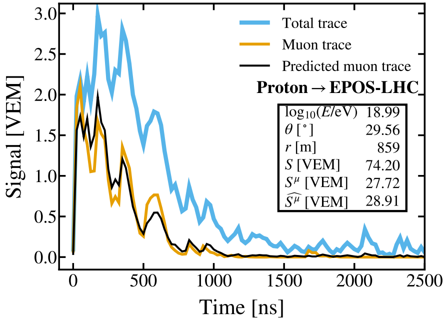
- Objective: Predict the temporal sequence of values in the muon trace
- Predictions follow the shape of the total trace
- Predictions capture the spiky shape of the muon trace

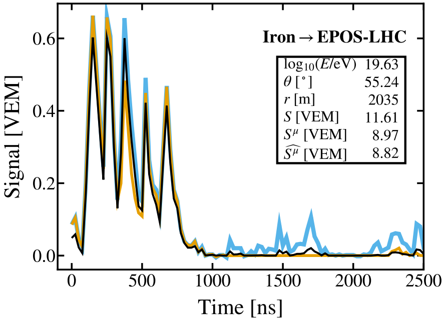
Notation

$\widehat{\ }$ for the predicted quantities

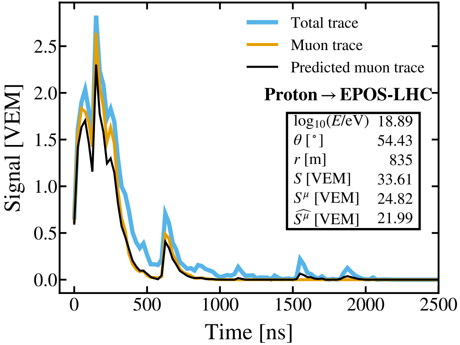
$\widehat{S^{\mu}}$: integral of the predicted muon trace

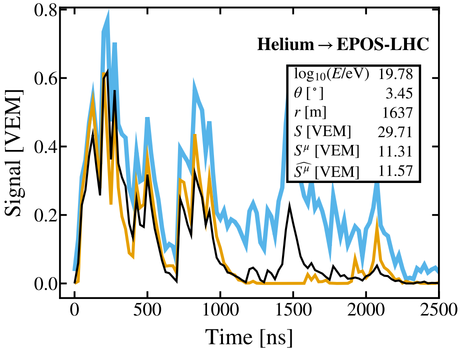
S^{μ}: integral of the true muon trace



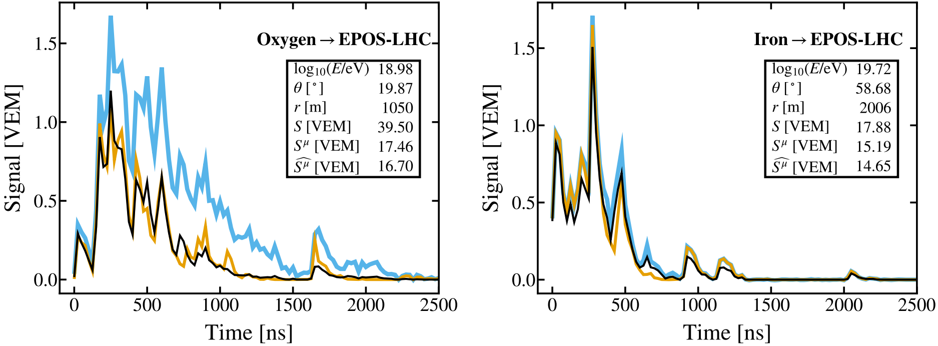


Results: Example Traces II

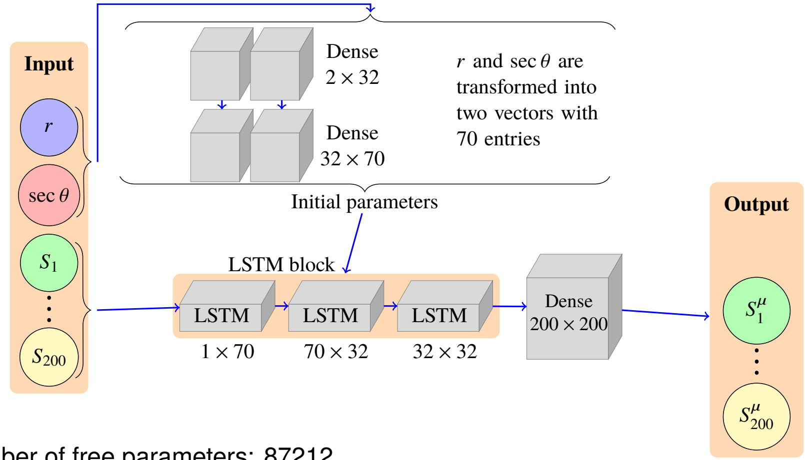




Results: Example Traces III

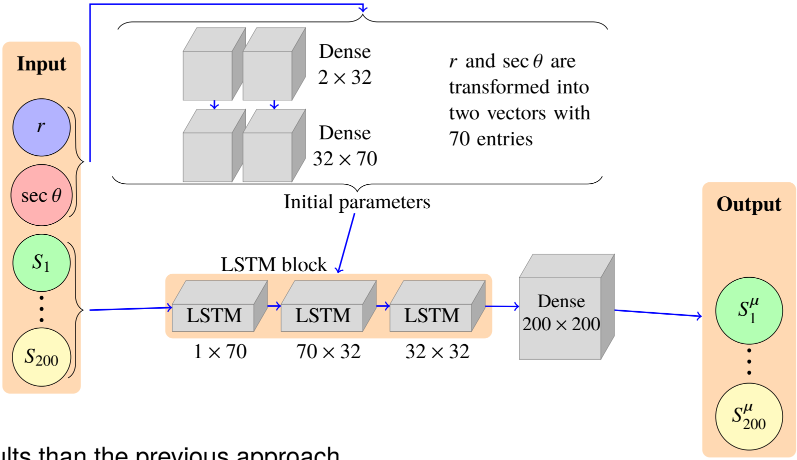


How? Neural Network



- Total number of free parameters: 87212
- r, $\sec\theta$ and $S_1 \ldots S_{200}$ are normalized to be between 0 and 1
- Train with 25

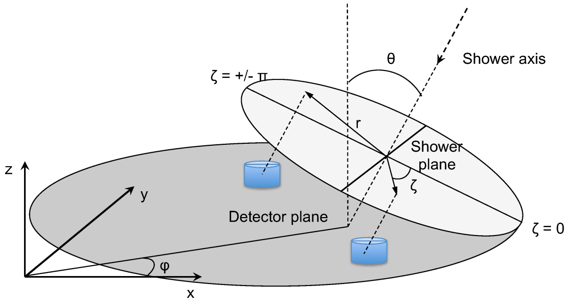
How? Neural Network



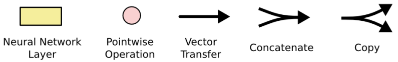
- Better results than the previous approach
- Only two variables used, r and $\sec\theta$ that depend only on the geometric part of the reconstruction
- The energy is not used

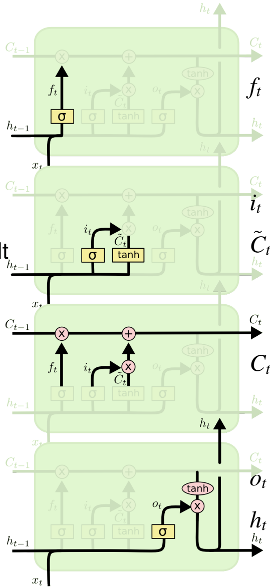
Variables

- The *zenith angle* θ. It is usual to study its secant, $\sec\theta$, because the depth of the atmosphere that a particle travels through is proportional to $\sec\theta$
- The *energy* E of the cosmic ray
- The *core distance* r is the distance from the shower axis to the station in the shower plane
- The *signal* S measured at each station. It is measured in Vertical Equivalent Muons or VEMs
- The *azimuthal angle* ζ



LSTM Layer





- The forget vector f_t is built

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

- The input vector i_t is built
- The candidate cell state $\tilde{C}_t$ is built

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

- The forget gate selects from the previous cell state C_{t-1}
- The input gate selects from the candidate cell $\tilde{C}_t$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

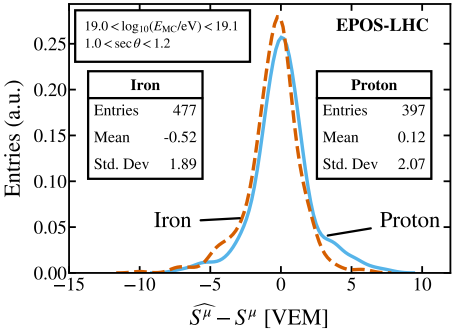
- Output obtained from previous hidden state h_{t-1}
- Hidden state obtained from the current cell state C_t

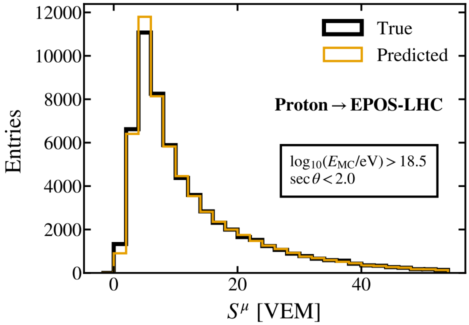
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh(C_t)$$

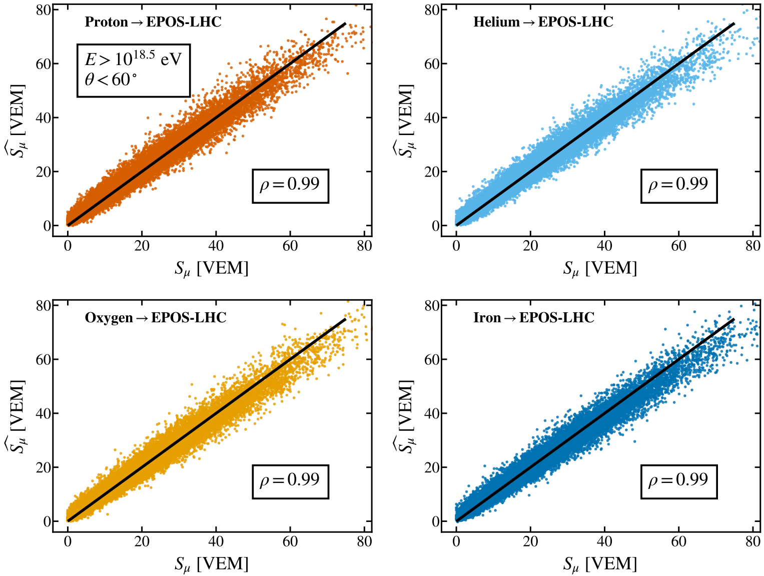
Performance Plots: S^{μ} and $\widehat{S^{\mu}}$

- We compare the integral of the predicted muon trace $\widehat{S^{\mu}}$ to the integral of the true muon trace S^{μ}
- Mean around zero, standard deviation close to 2 VEM (depends heavily on the zenith angle)

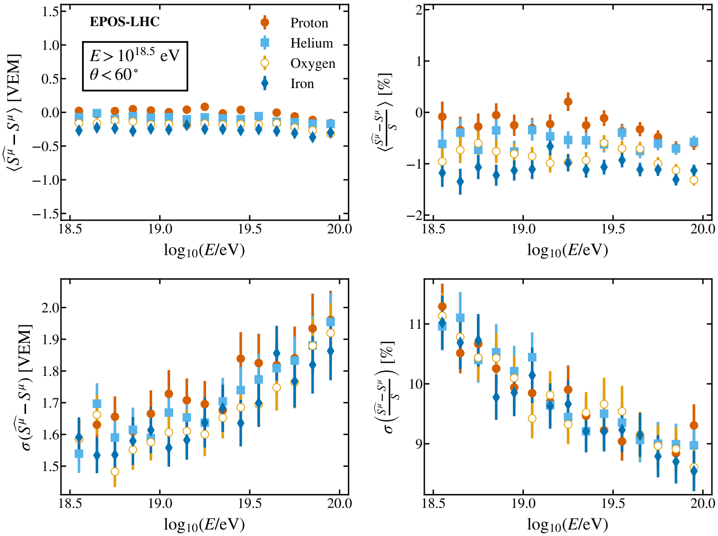




Performance Plots: Correlation

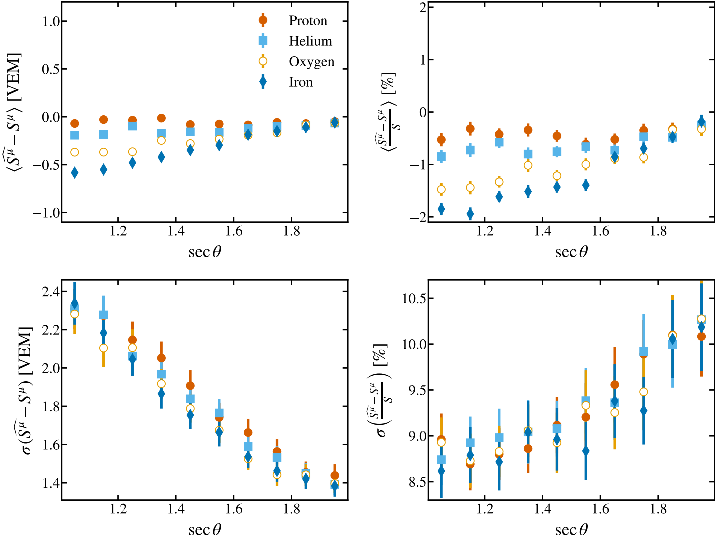


Performance Plots: E



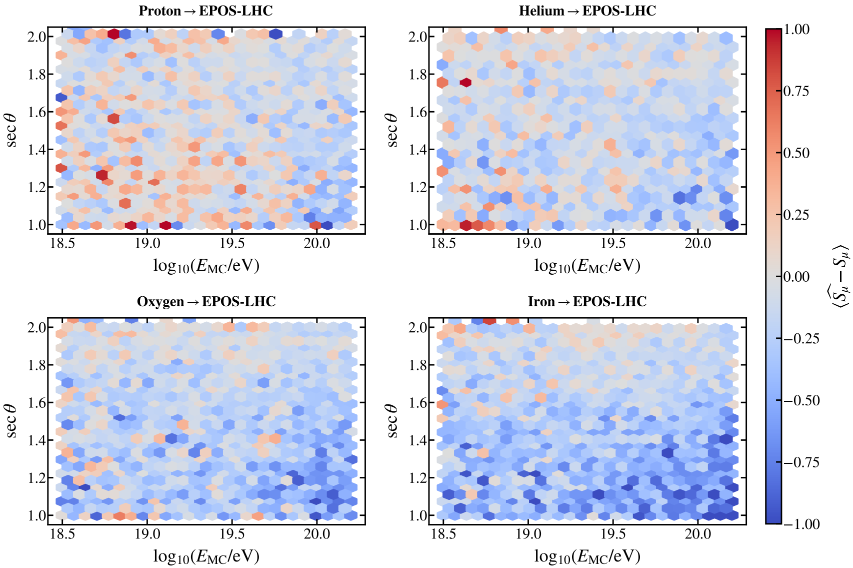
- Resolution better than 11% of the total signal S
- Lower energies have lower signals (rms of absolute difference is lower for lower signals but relative is higher)

Performance Plots: $\sec\theta$

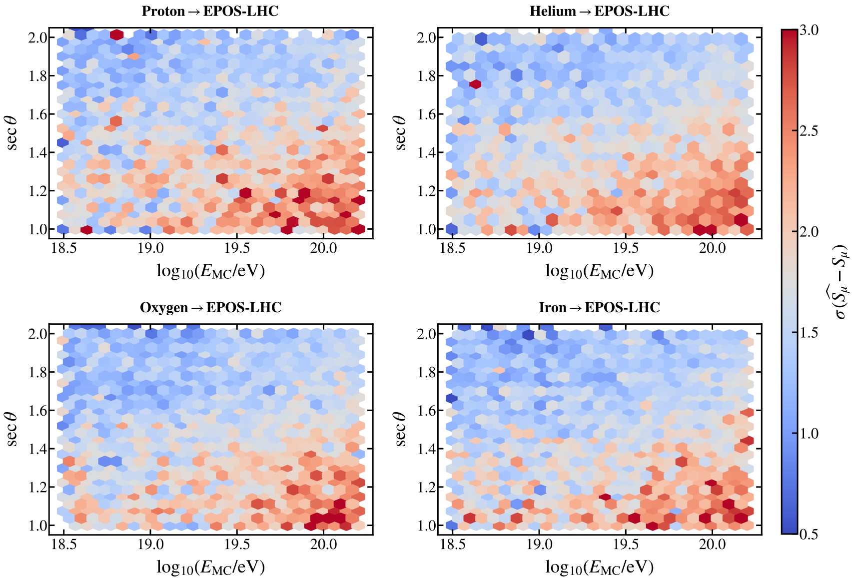


- Resolution better than 11% of the total signal S
- Lower energies have lower signals (rms of absolute difference is lower for lower signals but relative is higher)

Performance Plots: Mean

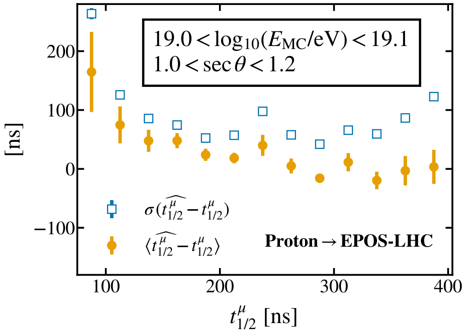


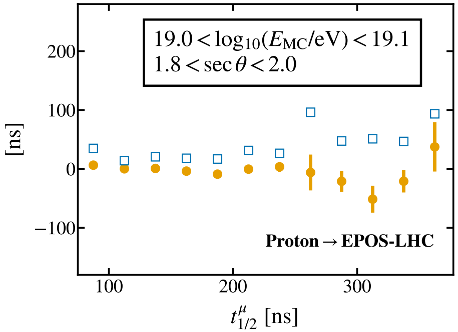
Performance Plots: Std



Performance Plots: Muon Risetime

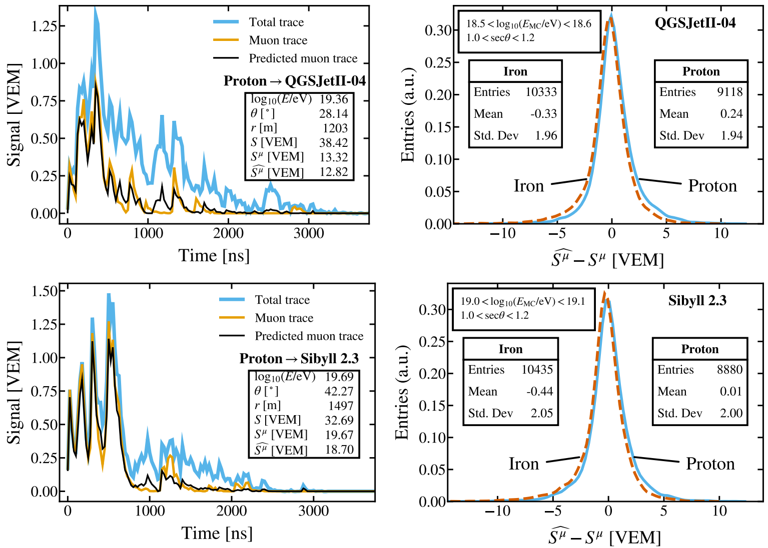
- We compare the risetime of the predicted muon trace $\widehat{t^{\mu}_{1/2}}$ with the risetime of the true muon trace $t^{\mu}_{1/2}$
- Mean close to 0, standard deviation less than 100 ns
- A single muon has a risetime of 15 ns and a decay constant of 60 ns





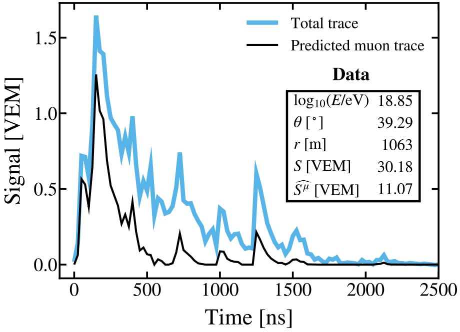
Performance Plots: Other Hadronic Models

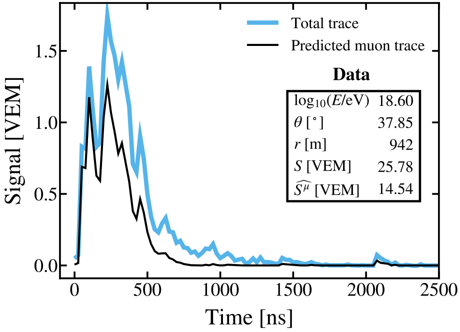
- The predictions are as good when predicting for simulations done with a different hadronic model



Predictions on Data

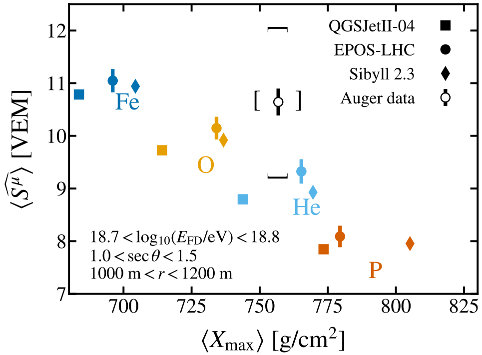
- Two examples of traces for two stations from two events recorded by the SD





Comparing Data and Simulations: Muon Deficit

- We compare predicted muon signals (at ~ 1000 m, by only picking stations with $1000\,\mathrm{m} < r < 1200\,\mathrm{m}$) in simulations and hybrid data
- We obtain a muon deficit in simulations for **vertical events** for the first time
- We compare predicted muon signals in simulations and hybrid data

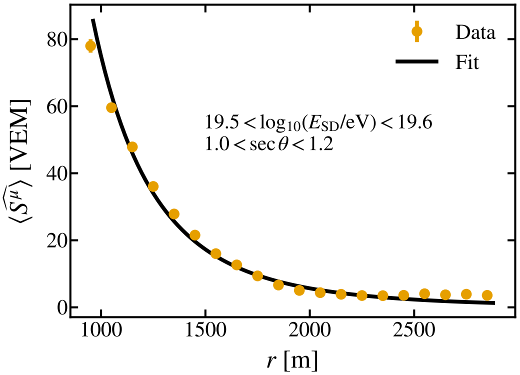


Comparing to Data from Other Experiments

- We fit our data with parameterizations obtained from other experiments, keeping the values of the original parameters

$$\rho_\mu(r) = N_\mu (C_\mu / R_0^2) R^{-\alpha} (1 + R)^{-\beta} [1 + (r/800\text{m})^3]^{-\delta}$$

Akeno: J. Phys. G. Nucl. Part. Phys 21 1101 (1995)

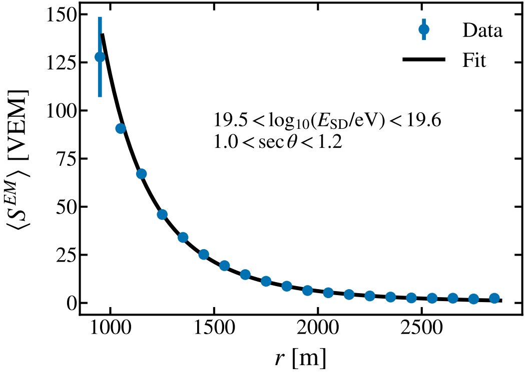


Comparing to Data from Other Experiments

- We fit our data with parameterizations obtained from other experiments, keeping the values of the original parameters
- The electromagnetic signal by subtracting from the total: $\widehat{S^{EM}} = S - \widehat{S^{\mu}}$

Akeno: J. Phys. G. Nucl. Part. Phys 18 423 (1992)

$$\rho_e = N_e C_e R^{-\alpha} (1+R)^{-\eta+\alpha} \left(1 + \frac{r}{2000}\right)^{-0.5}$$



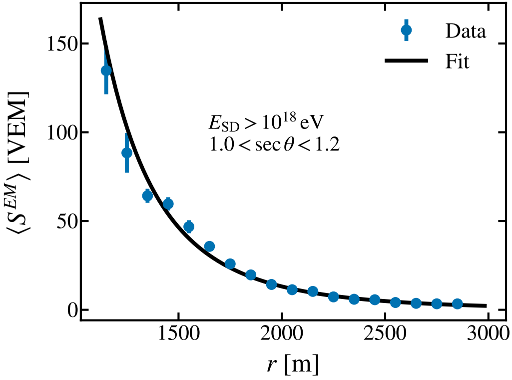
Comparing to Data from Other Experiments

- We fit our data with parameterizations obtained from other experiments, keeping the values of the original parameters
- The electromagnetic signal by subtracting from the total: $\widehat{S^{EM}} = S - \widehat{S^{\mu}}$

$$VR(\alpha,\eta) = \frac{N}{R_0^2} C(\alpha,\eta) \left(\frac{R}{R_0}\right)^{-\alpha} \left(1 + \frac{R}{R_0}\right)^{-\eta+\alpha}$$

*wrong label in the plot, it's $E_{SD} > 10^{19.83}$ eV

Volcano Ranch: Phys. Rev. Lett. 10 146 (1963)

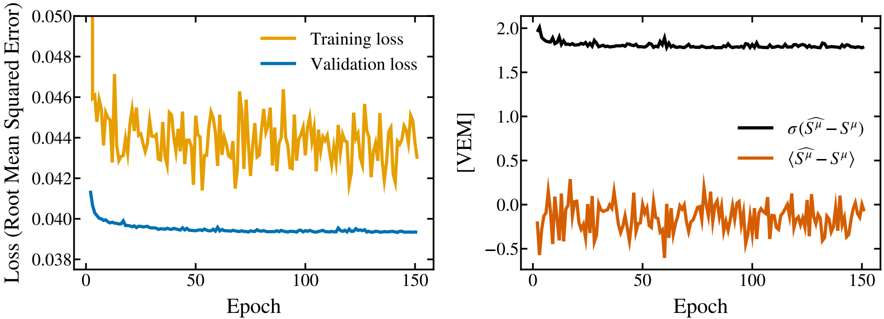


Summary

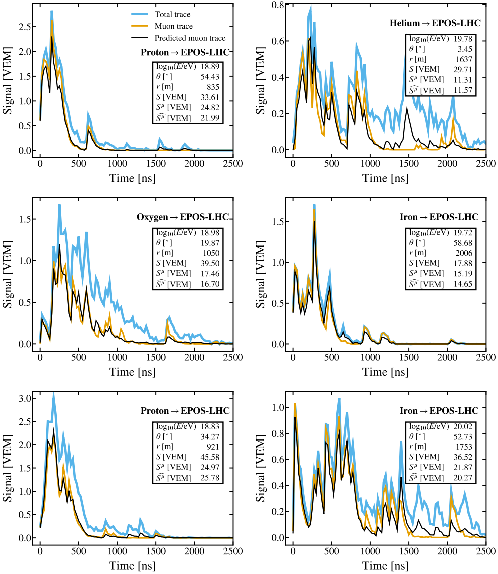
- Using Neural Networks, we can predict the muon signal contained in the simulated traces
- Predictions are precise and do not depend on the hadronic model used
- Predictions show a very small bias with composition
- The lateral distributions of muons and electromagnetic signals extracted from the SD data are well reproduced by published parameterizations
- Next step: validation with data, Auger Prime

Backup

Training: Loss and Other Metrics

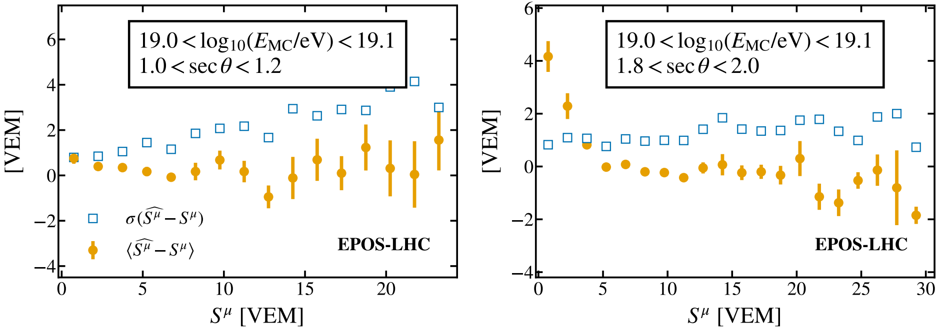


More Examples of Traces



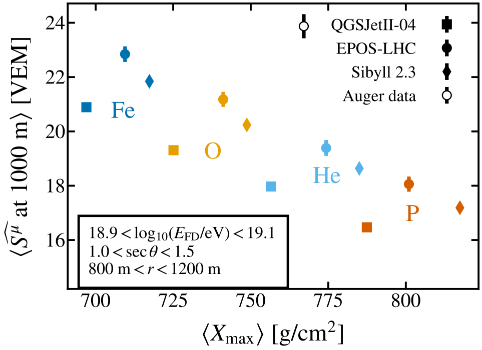
Performance: as a function of S^μ

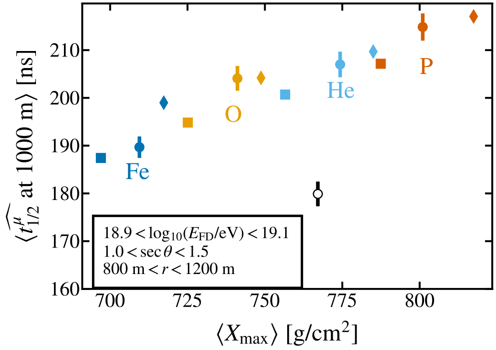
- Mean close to 0
- Performance improved for larger zenith angles



Comparing Data and Simulations: Muon Deficit

- The average muon risetime also points towards a heavier composition than iron





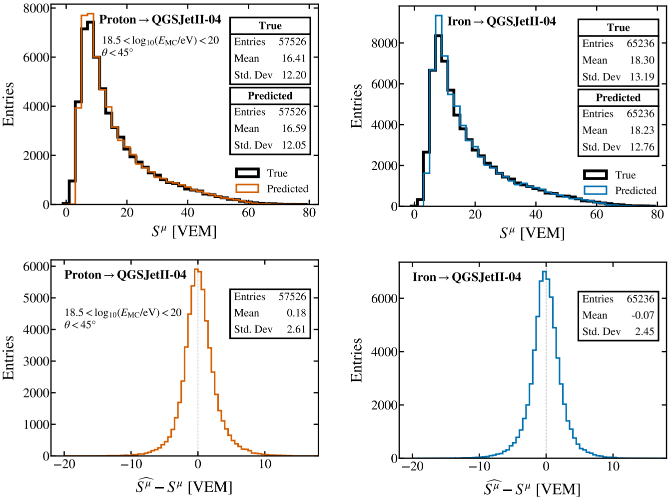
First Work: Data Selection

QGSJetII-04				
		Training	Validation	Test
Primary	# of events	# of detectors		
Proton	19362	16088	4022	57522
Helium	12341	15960	3989	36740
Nitrogen	12201	16071	4017	36069
Iron	19478	16076	4018	65455
EPOS-LHC				
		Training	Validation	Test
Primary	# of events	# of detectors		
Proton	18456	–	–	78063
Iron	18779	–	–	86862

- The trace used is the average of the traces recorded by each of the active PMTs of the station
- Stations with $S > 10$ VEM that don't have saturation in any of the two channels
- $10^{18.5}$ eV $< E < 10^{20}$ eV $0^\circ < \theta < 45^\circ$

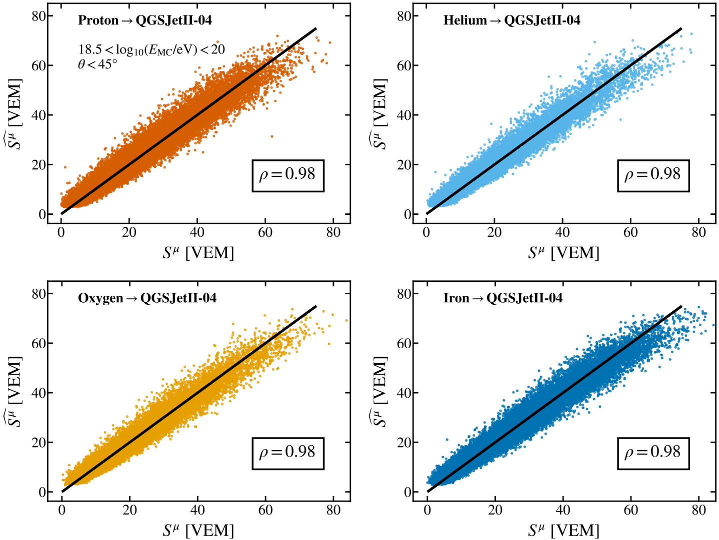
First Work: Performance: $\widehat{S^\mu}$ and S^μ

- Similar predicted and true distribution
- The mean of $\widehat{S^\mu} - S^\mu$ is very close to zero, Std. Dev. ~ 2.5 VEM



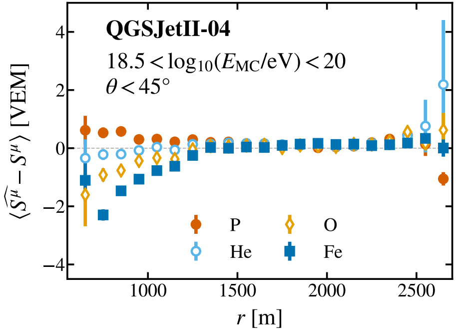
First Work: Performance: Correlation

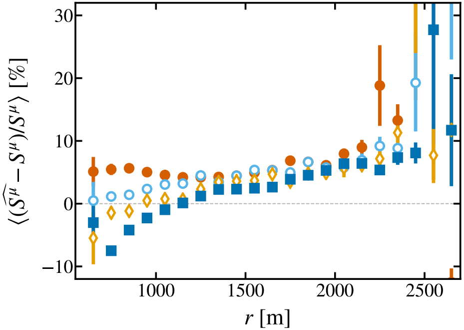
- Pearson correlation coefficient between $\widehat{S^\mu}$ and S^μ is $\rho = 0.98$



First Work: Performance: Dependence or r

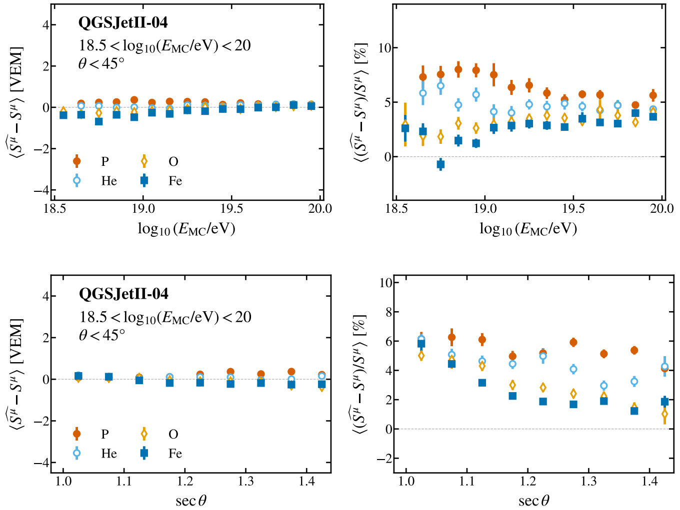
- Relative bias below 10%
- Larger bias for large $r =$ low S





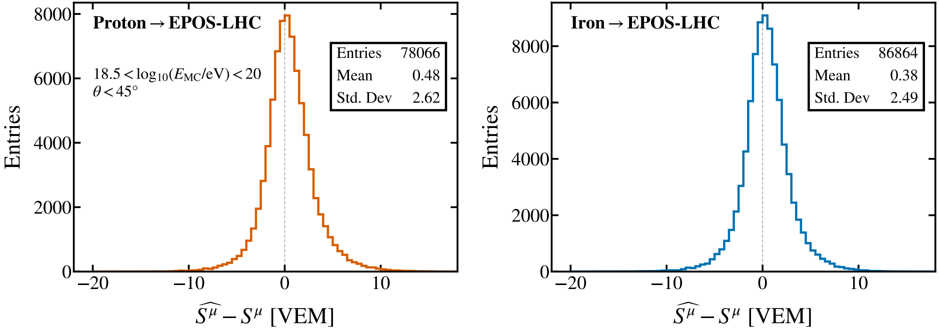
First Work: Performance: Energy and zenith angle dependence

- The bias with E and $\sec\theta$ is ~1 VEM for the absolute bias and below 10% for the relative bias



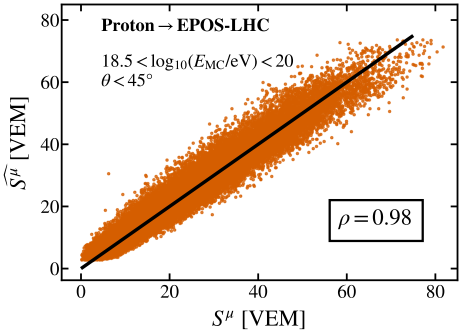
First Work: Performance: Prediction for EPOS-LHC

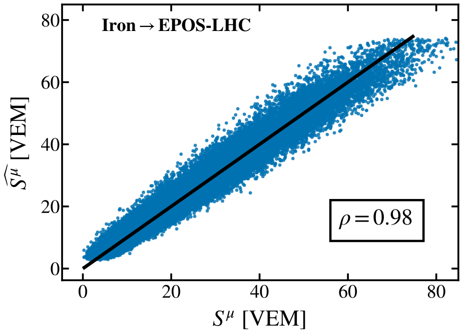
- We use simulations that the neural network has not seen and also done with a different hadronic model than the one used for training
- The estimation performs very similarly to the one done for QGSJetII-04
- The mean of the distribution is slightly biased towards less muons while the RMS of the distribution is still ~ 2.5 VEM



First Work: Performance: Prediction for EPOS-LHC

- The correlation between $\widehat{S^\mu}$ and S^μ is very high, with a Pearson correlation coefficient $\rho = 0.98$





First Work: Training

Training

25% Proton, Helium, Nitrogen and Iron (QGSJetII-04)

Roughly 20000 stations of each primary

Variables used in training

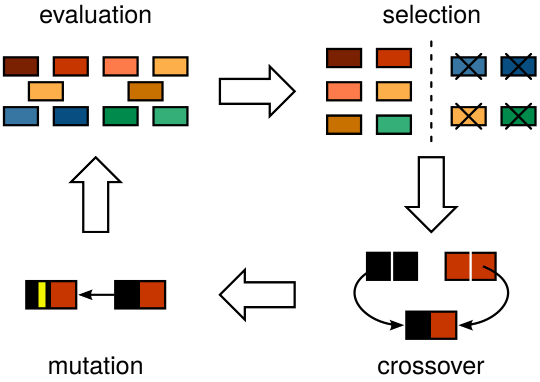
1. MonteCarlo energy E → In $\log_{10}$
2. MonteCarlo zenith θ → In $\sec\theta$
3. Distance to the core r → In m
4. Total signal S → In VEM
5. Trace length
6. Azimuth ζ → In radians
7. Risetime $t_{1/2}$ → In ns
8. Falltime → In ns
9. Area over Peak

Remark: S^{μ} is not strictly given as input to the net, it is used to guide the training process

⚠ Before training, the input is normalized to avoid operating with numbers that are very different in size

First Work: Genetic algorithm

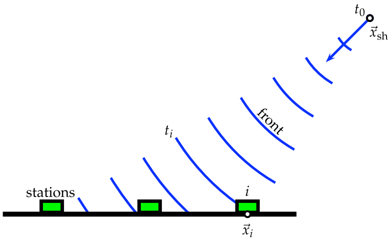
- There are several free parameters that can be fined tuned to achieve better performance. For example, the number of layers of the NN or the number of neurons in each layer
- We have used genetic algorithms to find a good NN



Shower Reconstruction with the SD

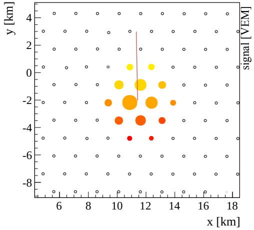
Direction reconstruction

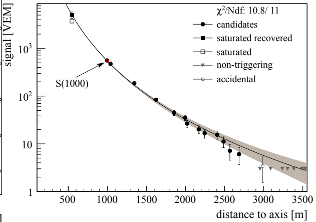
- The direction of the cosmic ray is obtained by fitting a spheric plane to the time of arrival of particles at the stations

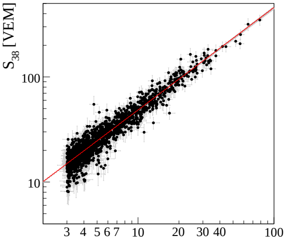


Energy reconstruction

- Energy is obtained from the lateral distribution of particles

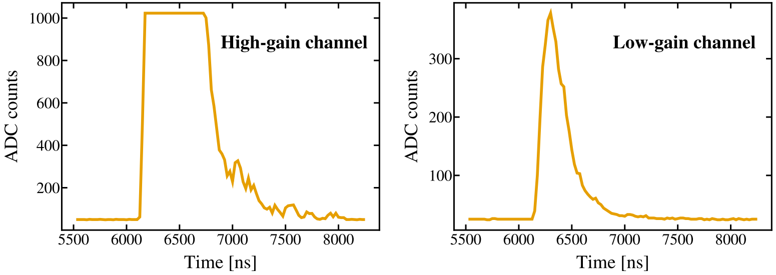






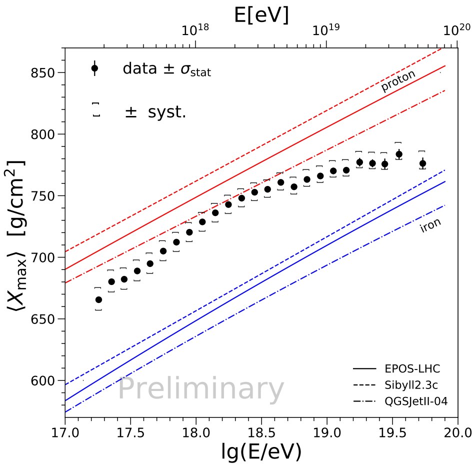
Signal Saturation

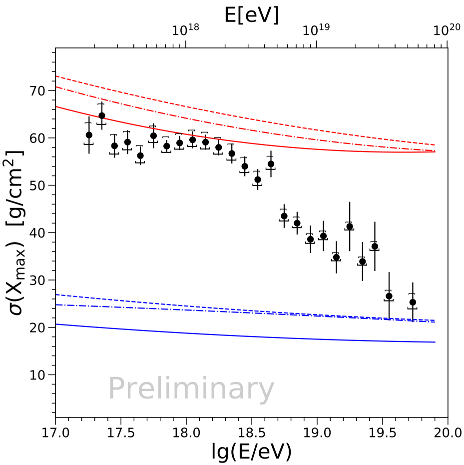
- Signals from PMTs come from the high-gain channel and low-gain channel
- Signals big enough will saturate the high-gain channel first and then the low-gain channel



- We always use signals from the high-gain channel that are not saturated

Recent Results: Mass Composition





Recent Results: Composition Fractions