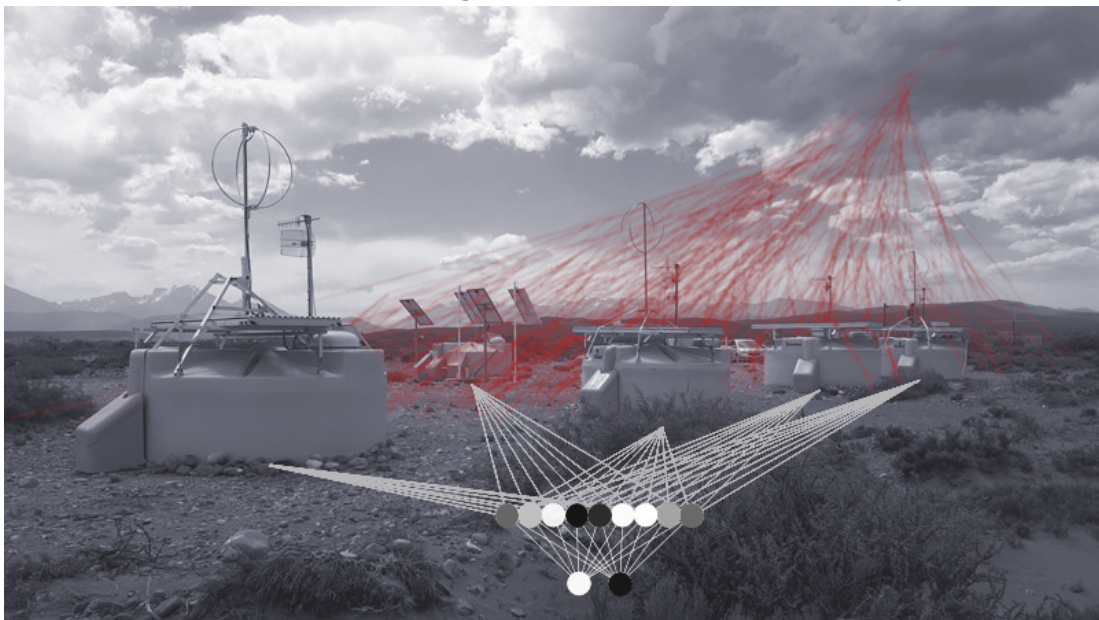


# Harry Potter and the Neural network triggers for the surface detector of the Pierre Auger Observatory



Master's thesis by

**Paul Filip**

at the Institute for Astro Particle Physics

First Reviewer: Prof. Dr. Ralph Engel

Second Reviewer: Prof. Dr. Guido Drexlin

Processing time: 01.04.2022 – 31.03.2022



# Review and Declaration

This thesis has been accepted by the first reviewer of the master thesis.

*Karlsruhe, TBD*

---

Prof. Dr. Ralph Engel

I declare that the work in this thesis was carried out in accordance with the requirements of the university's regulations and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others is indicated as such.

*Karlsruhe, TBD*

---

Paul Filip



# Contents

<b>1</b>	<b>The Pierre Auger Observatory</b>	<b>1</b>
1.1	Fluorescence Detector (FD) . . . . .	1
1.2	Surface Detector (SD) . . . . .	3
1.2.1	Data acquisition (DAQ) . . . . .	3
1.2.2	Online calibration . . . . .	4
1.2.3	Offline calibration . . . . .	4
1.3	Event Reconstruction . . . . .	8



# 1 The Pierre Auger Observatory

Located on the argentinian high-plains of Pampa Amarilla, the Pierre Auger observatory is a hybrid detector designed to detect and study cosmic rays of the highest energies. With an effective area of  $3000 \text{ km}^2$  it is by far the largest experiment of its kind [1].

Although first proposed in 1992, it took 18 years until the idea of a large scale experiment to detect cosmic rays matured and construction of the first prototype started near Mendoza [2]. Some further 20 years later, the Pierre Auger collaboration has co-authored over publications and continues to advance research in astroparticle physics.

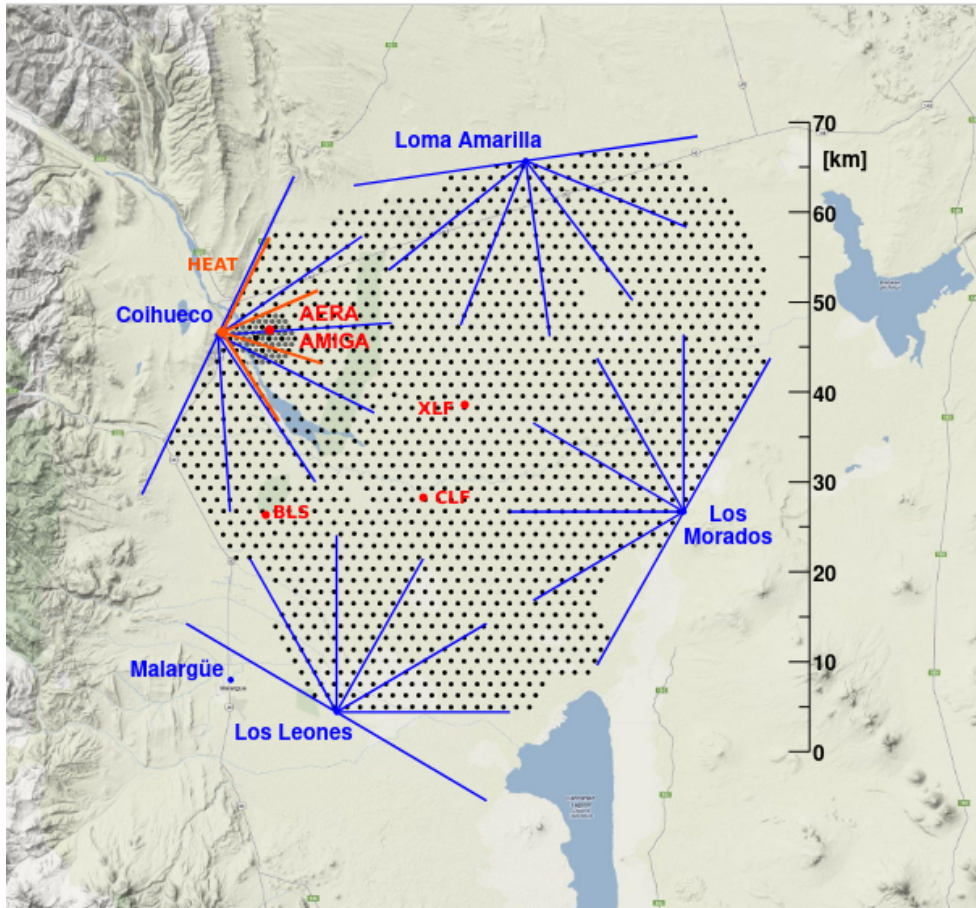
It does this via a hybrid approach, combining measurements of a **Surface Detector** (SD) as well as a **Flouresence Detector** (FD). Additional machinery, such as the **eXtreme** (XLF) and **Central Laser Facility** (CLF), is installed to monitor atmospheric variables. This improves the overall systematic accuracy of predictions made by the experiment. An overview of the site can be seen in Figure 1.1. Data measured by the FD, SD and the atmospheric monitors is sent to a **Central Data Acquisition System** (CDAS) located in the nearby town of Malargüe.

This chapter offers a brief look into the measurement principle and setup of the observatory. Information regarding the fluorescence detector can be found in section 1.1. The SD is described in section 1.2. A more in depth read on detector specifications and design choices is represented by the Pierre Auger observatory design report [1], where a lot of information stated in this chapter is conglomerated from. Notes on the event reconstruction are listed in section 1.3 and summarized from [3] and [4].

## 1.1 Fluorescence Detector (FD)

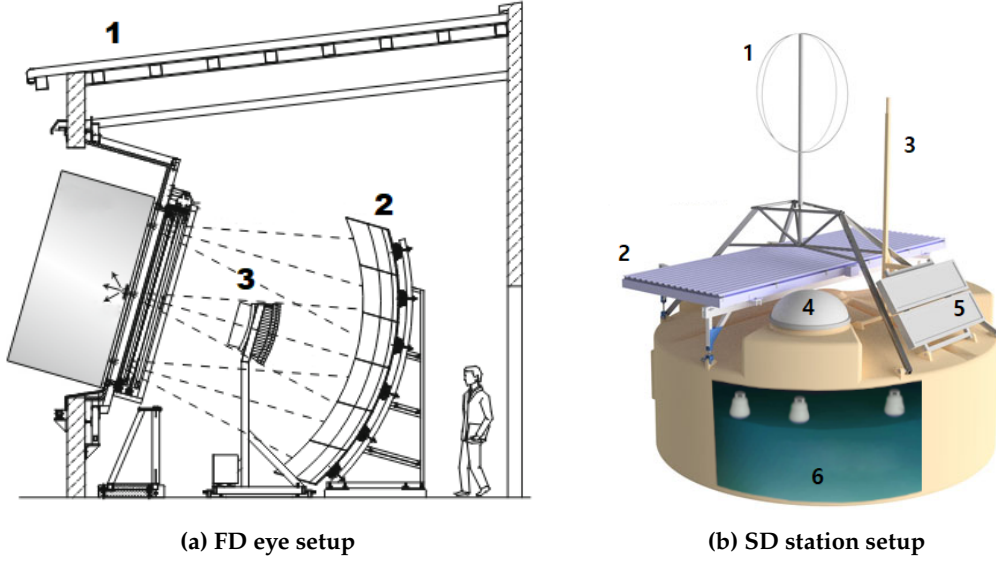
The FD consists of a total of 27 fluorescence telescopes (eyes) at 4 different sites. Each eye monitors a  $30^\circ \times 30^\circ$  window of the sky at a resolution of  $\approx 0.5 \frac{\text{px}}{\text{deg}^2}$ . This results in an effective FOV of roughly  $180^\circ \times 30^\circ$  per FD station, with an exception of Coihueco, where three additional telescopes - HEAT (**H**igh **E**levation **A**uger **T**elescope) - are installed to enable monitoring of higher zenith angles ( $30^\circ \leq \theta \leq 60^\circ$ ) and increase sensitivity for showers of lower energies (compare ??). A schematic of the setup of each eye is given in Figure 1.2a.

The individual telescopes consist of 3.6 m by 3.6 m, convex mirrors. They reflect incoming light onto a set of 440 photomultipliers (PMTs), each corresponding to one pixel in the resulting image seen by an eye. Since the setup needs to be extremely sensitive to UV light in order to detect flouresence caused by extensive air showers, its operation is limited to the relatively noise free moonless astronomical nights (Sun



**Figure 1.1:** Overview of the Pierre Auger observatory. The four different FD sites (respective FOV shown with blue lines) sit at the edge of the detector area and monitor the night sky above the SD array consisting of 1600 water tanks (black dots). A denser spacing of stations near Colihueco is equipped with additional electronics such as e.g. radio antennas (AERA) and muon detectors (AMIGA). Image taken from [5]





**Figure 1.2:** (a) Schematic view of an FD eye with housing (1), main mirror (2) and camera (3). Image taken from [4] (b) Setup of an SD WCD with radio antenna (1), SSD (2), communication and GPS antenna (3), electronics box (4), solar panels (5) and the WCD (6). Image adopted with changes from [6] and [7]

$\angle \text{Horizon} \lesssim -18^\circ$ ). When the FD is operational, this allows the observation of the longitudinal propagation of a shower instead of just its' footprint (as seen by the SD).

## 1.2 Surface Detector (SD)

The SD consists of 1600 individually operating stations, spaced apart on a hexagonal grid with a standard 1.5 km spacing. Each station is made up of a main tank filled with 12 000 L of purified water and reflective inner walls, a solar panel and batteries for power management, as well as an antenna for communication. Within each tank three PMTs detect Cherenkov light originating from shower particles, these are together with the tank referred to as **Water Cherenkov Detectors (WCDs)**. With the (at the time of this work) ongoing AugerPrime upgrade, each station is additionally equipped with a small PMT (sPMT), **Surface Scintillator Detector (SSD)**, and radio antenna atop the tank. This allows for the recording of stronger signals, finer separation of electromagnetic and muonic shower component and detection of highly inclined air showers respectively [8, 9]. Figure 1.2b shows a schematic blueprint of each SD station.

### 1.2.1 Data acquisition (DAQ)

Onboard electronics, the **Upgraded Unified Board (UUB)**, or more precisely six 10-bit **Flash Analog-to-Digital-Converters (FADCs)** read out measurement data from the PMTs at a sampling rate of 120 MHz ( $\approx 8.33$  ns binning) [verzi2013energy]. This is done in a two-fold way. Three FADCs digitize the PMTs dynode voltage, resulting

in the **High Gain (HG)** output. Three FADCs monitor the anode voltage to form the **Low Gain (LG)** output, which can be analyzed if the HG output exceeds a value of  $2^{10}$  ADC counts and becomes saturated. This effectively enables the measurement of both large ( $\geq O(10^3)$  particles hitting the tank) as well as small shower signals ( $O(1)$  particle hitting the tank) with sufficient accuracy. Once an FADC bin has been recorded and checked for possible triggers (c.f. ??) it is written to a ring buffer. If a trigger is issued, the corresponding chunk in the ring buffer ( $\approx 4.992 \mu\text{s}$  (599 bins) before and  $12.07 \mu\text{s}$  (1448 bins) after a trigger,  $2047 + 1$  bins total), the measured trace, can be analyzed in order to calibrate a station in the array (??) or processed by a higher-level CPU for event reconstruction purposes (see section 1.3).

While each station is equipped with the same electronics and runs the same analysis software, variables like the position in the field, station age or slight changes in the manufacturing/installation process cause different stations to age differently. Over the lifetime of the array such differences can sum into potentially drastic discrepancies in gathered data. Put simply, an extensive air shower will look different both to different WCDs at the same time as well as the same WCD at different times. To account for this, measurements are standardized across all stations. ADC counts are related to a **Vertical Equivalent of through-going Muons (VEM)** that would result in the same signal strength. In this fashion, the maximum response that is generated by a PMT from one vertically through-going muon is defined as  $1 \text{ VEM}_{\text{Peak}}$ . The total deposited charge (equivalent to the integral of the response) is defined as  $1 \text{ VEM}_{\text{Ch.}}$ . The conversion factor between ADC counts and  $\text{VEM}_{\text{Peak}}$  and  $\text{VEM}_{\text{Ch.}}$  (referred to as  $I_{\text{Peak}}$  and  $Q_{\text{Peak}}$  respectively) is estimated from data and continuously updated separately for each station. Note that due to the limited computational resources of the WCD, a simplified, rate-based approach is chosen for calibration in the field (Online calibration), this stands in contrast to the more physics-driven histogram method used during event reconstruction (Offline calibration). In any case, both algorithms are listed in the following subsections and discussed in more detail in the referenced literature.

## 1.2.2 Online calibration

### Baseline estimation

#### Estimation of $I_{\text{Peak}}$ and $Q_{\text{Peak}}$

## 1.2.3 Offline calibration

### Baseline estimation

For event reconstruction, a first baseline estimate of a WCD PMT is predicted by examining the beginning and end of a 2048 bin ( $17.06 \mu\text{s}$ ) long trace. The mode  $m$  as well as standard deviation  $\sigma$  of the first (last) 300 bins is calculated. All bins larger or smaller than  $m \pm 2\sigma$  are truncated and removed from the trace window. The value of  $m$ ,  $\sigma$  is consequently updated and the procedure repeated until a convergence is

reached and no further cut is necessary. The best estimate  $B_{\text{front}}$  ( $B_{\text{end}}$ ) for the front (end) of the trace at this point is given by the mean value of all remaining bins. It's statistic uncertainty  $\sigma_{B_{\text{front}}}$  ( $\sigma_{B_{\text{end}}}$ ) is given by the standard deviation of the remaining bins [10]. The baseline between the flat front and end estimate is then interpolated based on the difference

$$\Delta B = B_{\text{end}} - B_{\text{front}}. \quad (1.1)$$

- **Rejection of anomalous upward fluctuations  $\frac{\Delta B}{\sigma_{\Delta B}} \geq +10$ :**

$B_{\text{end}}$  being higher than  $B_{\text{front}}$  often indicates errors in the electronic readout or defect components in the measurement chain. There exists no physical reason why the end baseline should be (significantly) higher than the front. Consequently, traces where this is the case are ignored during event reconstruction.

- **Constant approximation for small upward fluctuations  $+5 > \frac{\Delta B}{\sigma_{\Delta B}} \geq 0$ :**

Small fluctuations of the baseline are expected and the norm. If these fluctuations are positive ( $B_{\text{end}} > B_{\text{front}}$ ) the method of calculating the mode, truncating outliers and repeating both steps is applied to the entire length of the signal, resulting in a constant baseline estimate  $B$  across the trace.

- **Step-function approximation for small downward fluctuations  $0 > \frac{\Delta B}{\sigma_{\Delta B}} \geq -1$ :**

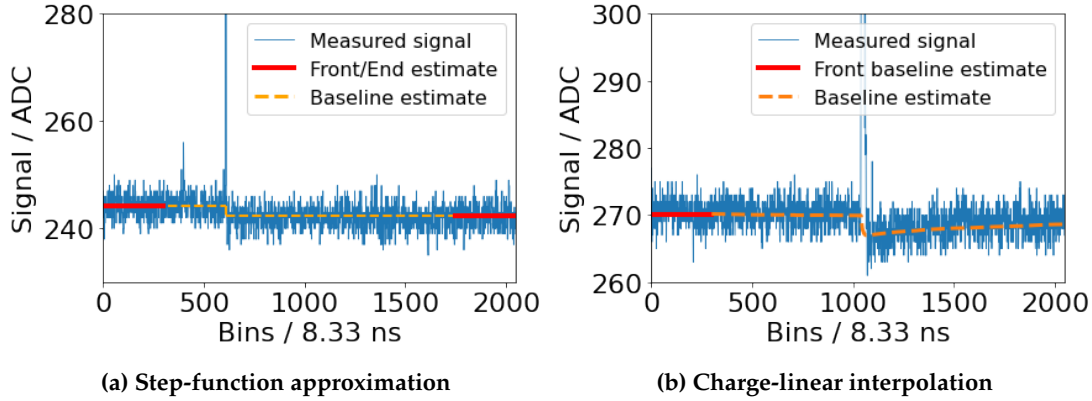
Unlike positive fluctuations, negative fluctuations ( $B_{\text{end}} < B_{\text{front}}$ ) can have a physical significance. Due to the undershoot of PMTs after detecting a signal in the WCD (compare [glietta2008recovery]), the baseline estimate decreasing towards the end of the trace often indicates the presence of shower particles within the tank. For this reason, downward fluctuations are handled differently from upward ones. If the fluctuations are sufficiently small, the baseline across the trace is estimated as a simple step-function; The trace is separated into two parts along its' maximum ADC value. The front part (i.e. before the max. value) has the baseline  $B_{\text{front}}$ , while the rear part is estimated by  $B_{\text{end}}$ . An example of this is shown in Figure 1.3a.

- **Charge-linear approximation for large undershoots  $-1 \geq \frac{\Delta B}{\sigma_{\Delta B}}$ :**

For larger undershoots, the baseline is estimated bin by bin based on the deposited charge in the detector. Starting with a value of  $B_{\text{front}}$  for the bins 1-300, the remaining baseline is first linearly interpolated according to Equation 1.2,

$$b_i = B_{\text{front}} - \Delta B \cdot \frac{i - 300}{1448}, \quad 300 \leq i \leq 2048, \quad (1.2)$$

where the magic numbers 300 and 1448 refer to the last bin of the front baseline estimate and the length of the interpolated baseline respectively. From this, the deposited charge  $q_i$  up to bin  $i$  can be calculated as per Equation 1.3.



**Figure 1.3:** (a) A simple step function is sufficient to accurately model a PMTs' noise level at small downward fluctuations. (b) For larger discrepancies the more involved charge-linear interpolation is used. Note that the signal undershoot is exaggerated for visualization purposes in both examples.

$$q_i = \sum_{k=0}^i (T_k - b_k) \exp\left(-\frac{8.33 \text{ ns}}{\tau} \cdot (i - k)\right) \quad (1.3)$$

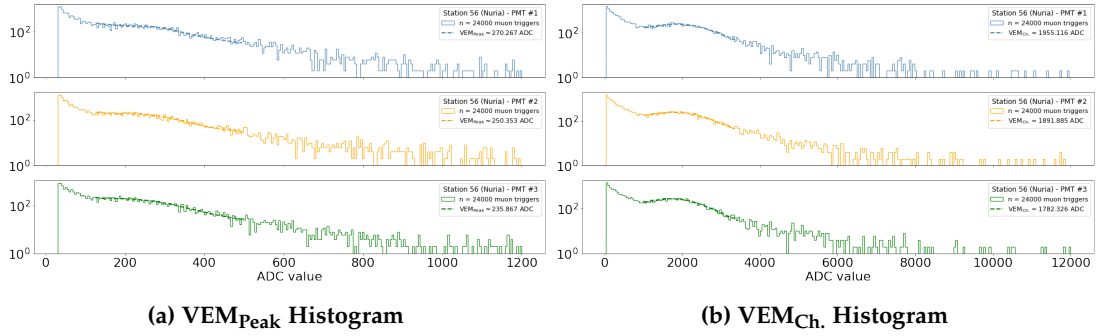
In Equation 1.3,  $T_k$  refers to the numerical value of bin  $k$ . Note that an exponential falloff term has to be added to account for the decay in signal undershoot with a decay time of  $\tau = 45 \mu\text{s}$ . The value of  $\tau$  is determined in [11]. Assuming the magnitude of the signal undershoot is directly proportional to the deposited charge  $q$ , a correction of the baseline thus becomes

$$b_i = B_{\text{front}} + \frac{q_i}{q_{1898}} \cdot \Delta B. \quad (1.4)$$

The parametrization in Equation 1.4 is chosen such that the charge-interpolated baseline at bin 1898 (the center position in the last 300 bins) is exactly equal to the rear baseline estimate  $B_{\text{end}}$ . The prediction can be made more accurate by repeating the above steps, each time recalculating  $q_i$  and readjusting the baseline  $b_i$  in the process. Figure 1.3b shows an example baseline estimate after three such iterations. In general, it converges to a robust estimate within five repetitions [11].

### Estimation of $I_{\text{Peak}}$ and $Q_{\text{Peak}}$

The conversion factor between ADC counts and  $\text{VEM}_{\text{Peak}}$ ,  $\text{VEM}_{\text{Ch}}$  are built from distributions of traces that satisfy the muon trigger, which scans incoming ADC bins for a value exceeding the muon threshold  $t_\mu = b + 30 \text{ ADC}$ , 30 ADC above baseline, for any of the three WCD PMTs. If this requirement is met, 69 bins (19 before, trigger bin, 49 after) are written to the muon buffer, a FIFO (first-in-first-out) type memory storage,



**Figure 1.4:** (a) The maximum value of each muon trace is histogrammed in order to gain information about the current value of  $I_{\text{Peak}}$  of a station. (b) The conversion factor from recorded ADC values to  $Q_{\text{Peak}}$  is given from the histogrammed sum of each muon trace.

that is subsequently filled with low-energy events, which (in general) didn't satisfy any other trigger but still contain useful information [12].

By histogramming the maximum value (sum) of each trace, the plot shown in Figure 1.4a (Figure 1.4b) can be obtained. It becomes apparent that the number of events per bin largely follows a power law with negative spectral index. This is expected considering the discussion in ?? . Notable are characteristic deviations from this powerlaw, as these contain information about  $I_{\text{Peak}}$  ( $Q_{\text{Peak}}$ ):

- Low energy events from e.g.  $e^-$ ,  $e^+$  that deposit their entire energy in the tank give rise to a surplus of events at lower ADC values.
- A characteristic (muon) hump appears in the bins 20–70. This surplus is caused by omni-directional muons impinging onto the detector. Since the energy deposited by such muons is roughly constant, the center of the muon hump serves as an estimate of  $I_{\text{Peak}}$  ( $Q_{\text{Peak}}$ ).
- (Not depicted in Figure 1.4) In similar plots from related works (c.f. [12, 13]) a drastic increase in bin occupations towards the tail end of the histograms can be observed. This is attributed to an increased bin size from 1500 ADC counts onwards, which reduces the amount of data per station sent to CDAS. In the example plots referenced here, a constant binning is chosen instead. This difference is mentioned here to avoid possible confusion.

In this fashion, the average response of the WCD to a through-going muon can be estimated by e.g. fitting a gaussian distribution to the muon hump. However, there exists a systematic difference between the response to a *vertical* or an *omni-directional* muon. Consequently, correctional factors need to be applied to the analysis results. These have been determined in previous experiments [14]. Finally, one arrives at an estimate for the conversion factor between ADC counts and  $\text{VEM}_{\text{Peak}}$ ,  $\text{VEM}_{\text{Ch.}}$ .

## 1.3 Event Reconstruction

If an extensive air shower hits the array,

## Bibliography

- [1] Alexander Aab et al. “The pierre auger observatory upgrade-preliminary design report”. In: *arXiv preprint arXiv:1604.03637* (2016).
- [2] *Timeline of the Pierre Auger Observatory*. <https://auger.org/observatory/timeline-observatory>. Accessed: 04th Oct. 2022. 2022.
- [3] Alexander Aab et al. “Reconstruction of events recorded with the surface detector of the Pierre Auger Observatory”. In: *Journal of Instrumentation* 15.10 (2020), P10021.
- [4] Jorge Abraham et al. “The fluorescence detector of the Pierre Auger Observatory”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 620.2-3 (2010), pp. 227–251.
- [5] Darko Veberič. *Auger array*. [https://web.ikp.kit.edu/darko/auger/auger-array/auger\\_array-pdf/auger\\_array-ad.pdf](https://web.ikp.kit.edu/darko/auger/auger-array/auger_array-pdf/auger_array-ad.pdf). Accessed: 19th Jul. 2022. 2021.
- [6] The Pierre Auger Collaboration. *AugerPrime*. <https://auger.org/observatory/augerprime>. Accessed: 06th Nov. 2022.
- [7] João de Mello Neto, ed. *Physics and astrophysics of ultra-high energy cosmic rays: recent results from the Pierre Auger Observatory*. NUCLEUS - 2020.
- [8] Antonella Castellina. “AugerPrime: the Pierre Auger observatory upgrade”. In: *EPJ Web of Conferences*. Vol. 210. EDP Sciences. 2019, p. 06002.
- [9] Jörg R Hörandel. “Precision measurements of cosmic rays up to the highest energies with a large radio array at the Pierre Auger Observatory”. In: *EPJ Web of Conferences*. Vol. 210. EDP Sciences. 2019, p. 06005.
- [10] Tobias Schulz et al. “New Baseline Algorithm for UB Traces”. GAP 2022-045. Sept. 2022.
- [11] Tobias Schulz et al. “New Baseline Algorithm for UUB Traces”. GAP 2022-0?? Dec. 2022.
- [12] Xavier Bertou et al. “Local-Station Calibration - The Missing Manual”. GAP 2022-0?? Oct. 2022.
- [13] Alexander Streich et al. “Performance of the upgraded surface detector stations of the Pierre Auger Observatory”. In: *Verhandlungen der Deutschen Physikalischen Gesellschaft* (2018).
- [14] P.S. Allison et al. “Surface Detector calibration in the Engineering Array”. GAP 2002-028. July 2002.