Contents

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 km^2 it is by far the largest experiment of its kind [1].

Altough 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 published over 110 papers [3] and continues to advance research in astroparticle physics.

It does this via a hybrid approach, combining measurements of a **S**urface **D**etector (SD) as well as a **F**louresence **D**etector (FD). Additional machinery, such as the e**X**treme (XLF) and **C**entral **L**aser **F**acility (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 **C**entral **D**ata **A**cquisition **S**ystem (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 fluoresence 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 [4] and [5].

1.1 Fluoresence Detector (FD)

The FD consists of a total of 27 fluoresence telescopes (eyes) at 4 different sites. Each eye monitors a 30° x 30° window of the sky at a resolution of $\approx 0.5 \, \frac{\mathrm{px}}{\mathrm{deg}^2}$. This results in an effective FOV of roughly 180° x 30° 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 ($30° \leq \theta \leq 60°$) 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

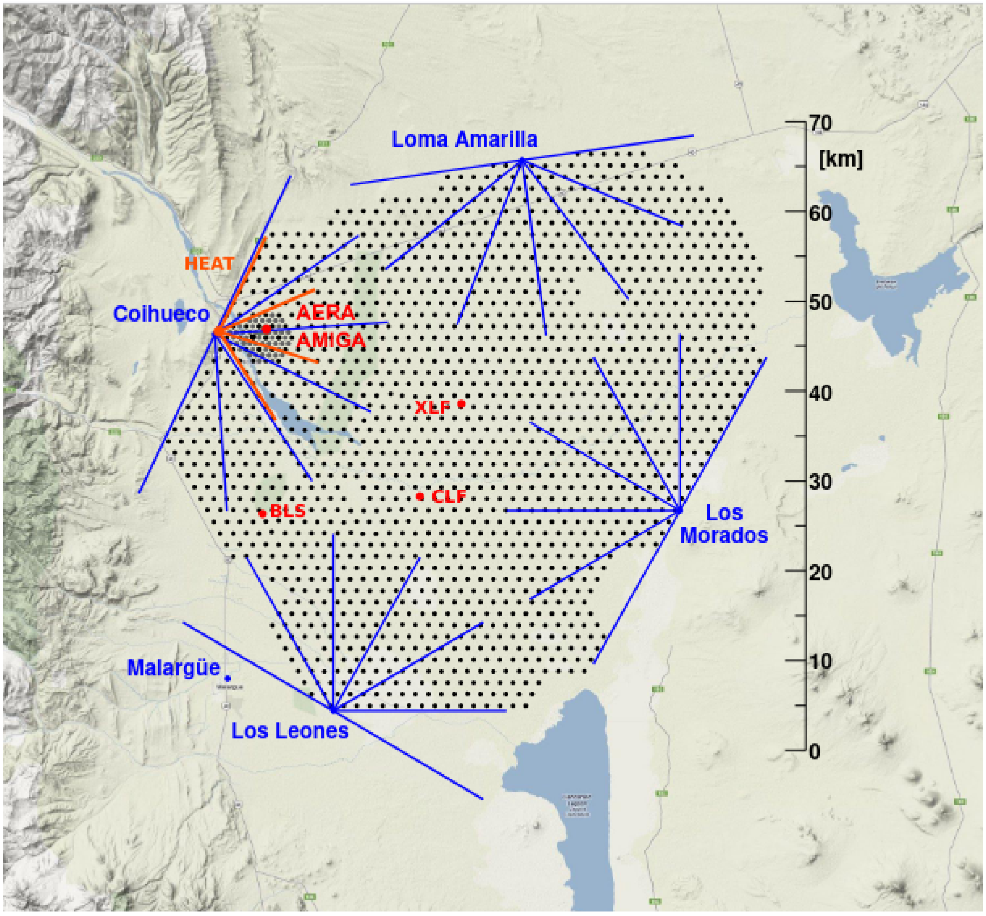
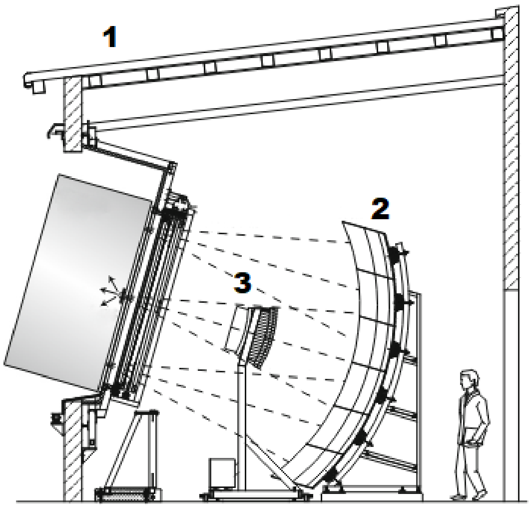
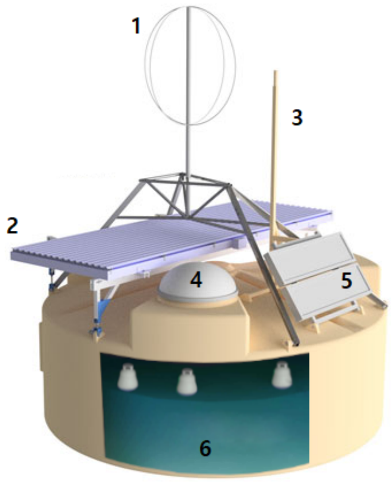


Figure 1.1: Overview of the Pierre Auger observatory. The four different FD sites (respective FOV shown with blue lines) sit at the edge of the detector area and monitor the night sky above the SD array consisting of 1600 water tanks (black dots). A denser spacing of stations near Coihueco is equipped with additional electronics such as e.g. radio antennas (AERA) and muon detectors (AMIGA). Image taken from [6]

.



(a) FD eye setup



(b) SD station setup

Figure 1.2: (a) Schematic view of an FD eye with housing (1), main mirror (2) and camera (3). Image taken from [5] **(b)** Setup of an SD WCD with radio antenna (1), SSD (2), communication and GPS antenna (3), electronics box (4), solar panesl (5) and the WCD (6). Image adopted with changes from [7] and [8]

$\measuredangle$ Horizon $\lesssim -18°$). 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 **W**ater **C**herenkov **D**etectors (WCDs). With the (at the time of this work) ongoing AugerPrime upgrade, each station is additionally equipped with a **s**mall **PMT** (sPMT), **S**urface **S**cintillator **D**etector (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 [9, 10]. Figure 1.2b shows a schematic blueprint of each SD station.

1.2.1 Data acquisition (DAQ)

Onboard electronics, the **U**pgraded **U**nified **B**oard (UUB), or more precisely six 10-bit **F**lash **A**nalog-to-**D**igital-**C**onverters (FADCs) read out measurement data from the PMTs at a sampling rate of 120 MHz ($\approx$ 8.33 ns binning) [11]. This is done in a two-fold way. Three FADCs digitize the PMTs dynode voltage, resulting in the **H**igh **G**ain

(HG) output. Three FADCs monitor the anode voltage to form the **L**ow **G**ain (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 ($\geq \mathcal{O}(10^3)$ particles hitting the tank) as well as small shower signals ($\mathcal{O}(1)$ particle hitting the tank) with sufficient accuracy [4]. Once an FADC bin has been recorded and checked for possible triggers (c.f. chapter 2) it is written to a ring buffer. If a trigger is issued, the corresponding chunk in the ring buffer ($\approx 4.992\,\mu s$ (599 bins) before and $12.07\,\mu s$ (1448 bins) after a trigger, 2047 + 1 bins total), the measured trace, can be analyzed in order to calibrate a station in the array (subsection 1.2.3, subsection 1.2.2) 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 **V**ertical **E**quivalent of through-going **M**uons (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 $1\,\mathrm{VEM_{Peak}}$. The total deposited charge (equivalent to the integral of the response) is defined as $1\,\mathrm{VEM_{Ch.}}$. The conversion factor between ADC counts and $\mathrm{VEM_{Peak}}$ and $\mathrm{VEM_{Ch.}}$ (referred to as I_{VEM} and Q_{VEM} respectively) is estimated from data and continuously updated separately for each station. Note that due to the limited computational resources of the WCD, as well as constraints on the amount of data that can be transmitted per station in the SD array ($1200\,\frac{\mathrm{bit}}{\mathrm{s}}$, [12]), a simplified, rate-based approach is implemented for autonomous 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 Offline calibration

Baseline estimation

In order to estimate I_{VEM} and Q_{VEM} of a WCD tank first the baseline - the average response in the absence of any signals - of each PMT needs to be determined. All further analysis will then be based on the baseline-subtracted PMT data.

For event reconstruction, a first baseline estimate of a WCD PMT is predicted by examining the beginning and end of a 2048 bin ($17.06\,\mu s$) long trace. The mode m as well as standard deviation σ 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, σ is consequently updated and the procedure repeated until a convergence is reached and no further cut is necessary. The best estimate B_{front} (B_{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 σ_{B_front} (σ_{B_end}) is given by the standard deviation of the remaining bins [13]. The baseline between the flat front and end estimate is then interpolated based on the difference

$$\Delta B = B_\text{end} - B_\text{front}. \tag{1.1}$$

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

  B_end being higher than B_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 $+5 > \frac{\Delta \mathbf{B}}{\sigma_{\Delta \mathbf{B}}} \geq 0$:**

  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 $0 > \frac{\Delta \mathbf{B}}{\sigma_{\Delta \mathbf{B}}} \geq -1$:**

  Unlike positive fluctuations, negative fluctuations ($B_\text{end} < B_\text{front}$) can have a physical significance. Due to the undershoof 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_front, while the rear part is estimated by B_end. An example of this is shown in Figure 1.3a.

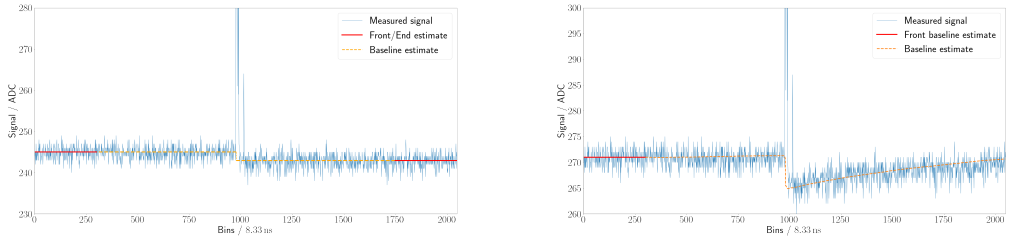
- **Charge-linear approximation for large undershoots $-1 \geq \frac{\Delta \mathbf{B}}{\sigma_{\Delta \mathbf{B}}}$:**

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

  $$b_i = B_\text{front} - \Delta B \cdot \frac{i - 300}{1448}, \qquad 300 \geq i \geq 2048, \tag{1.2}$$

  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.

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



(a) Step-function approximation **(b) Charge-linear interpolation**

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.

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 τ is determined in [14]. Assuming the magnitude of the signal undershoot is directly proportional to the deposited charge q, a correction of the baseline thus becomes

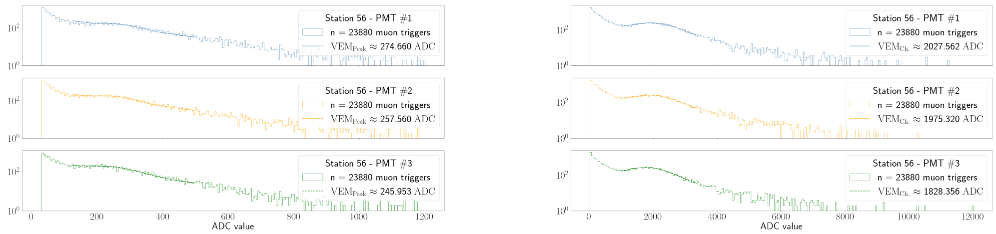
$$b_i = B_{\text{front}} + \frac{q_i}{q_{1898}} \cdot \Delta B. \tag{1.4}$$

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_{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 [14].

Estimation of I_{VEM} and Q_{VEM}

The conversion factor between ADC counts and VEM_{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, 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 **??**. Noteable are characteristic deviations from this powerlaw, as these contain information about I_{VEM} and Q_{VEM} :



(a) VEM$_{\text{Peak}}$ Histogram **(b) VEM$_{\text{Ch.}}$ Histogram**

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

- 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_{VEM} (Q_{VEM}).
- (Not depicted in Figure 1.6) In similar plots from related works (c.f. [12, 15]) 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 [16]. Finally, one arrives at an estimate for the conversion factor between ADC counts and VEM$_{\text{Peak}}$, VEM$_{\text{Ch.}}$.

1.2.3 Online calibration

Baseline estimation

Each SD station has an autonomous estimate of its' three WCD PMT baselines. They are defined simply as the mean of all first bins for each trace contained in the respective muon buffers (see subsubsection 1.2.3). This baseline estimate is used to set the thresholds of the hardware triggers discussed in chapter 2.

Estimation of I_{VEM} and Q_{VEM}

Due to the limited computational resources in each station, the determination of I_{VEM} and Q_{VEM} at station-level is fairly naive. Nevertheless, the σ-δ-method shown here has proven to be incredibly robust over the lifetime of the SD array [1].

In the beginning, the to-be-estimated value $I^{\text{est.}}_{\text{Peak}}$ ($Q^{\text{est.}}_{\text{Peak}}$) is set to the same, predefined value for all PMTs. A simple single-bin calibration trigger requiring all available WCD PMTs to be above a threshold of $t_{70} = 1.75\, I^{\text{est.}}_{\text{Peak}}$ above baseline plus a given PMT exceeding $2.5\, I^{\text{est.}}_{\text{Peak}}$ is used to determine a calibration trigger rate. If for some reason not all three WCD PMTs are functional, the thresholds are altered according to Table 1.1. What follows is an iterative procedure to approximate I_{VEM} (Q_{VEM}):

1. Calculate the trigger rate $r_{\text{cal.}}$ of the calibration trigger over a time $t_{\text{cal.}} = 5\,\text{s}$.
2. Adjust $I^{\text{est.}}_{\text{Peak}}$ ($Q^{\text{est.}}_{\text{Peak}}$) by $\pm\delta$ if $\pm(r_{\text{cal.}} - 70\,\text{Hz}) \geq 2\,\text{Hz}$, with $\delta = 1\,\text{ADC}$ initially.
3. If $t_{\text{cal.}} < 60\,\text{s}$ increase $t_{\text{cal.}}$ by $5\,\text{s}$. If $\delta > 0.1\,\text{ADC}$ decrease δ by $0.1\,\text{ADC}$.
4. While $t_{\text{cal.}} < 60\,\text{s}$ jump to step 1, else return $I^{\text{est.}}_{\text{Peak}}$ ($Q^{\text{est.}}_{\text{Peak}}$).

Table 1.1

n_{PMT}	t_{70}
1	2.85
2	2.00
3	1.75

1.3 Event Reconstruction

If an event has been detected (subsection 1.3.1) it is reconstructed at CDAS level, where information from all relevant detectors is conglomerated. From the observed shower footprint in the SD array as well as the (if available) longitudinal profile measured by the FD stations follows an estimate on arrival direction (subsection 1.3.3), energy (subsection 1.3.4) and primary particle (subsection 1.3.5). As the work presented in this thesis solely deals with the surface detector of the Auger observatory, this section focuses heavily on the SD reconstruction. Addendums towards FD reconstruction are given where needed.

1.3.1 Trigger procedure

The flux of cosmic rays espically at the highest energies is barely of the order of $1\,\text{km}^{-2}\,\text{yr}^{-1}$ [17]. Consequently, most signals observed by the Auger observatory stem

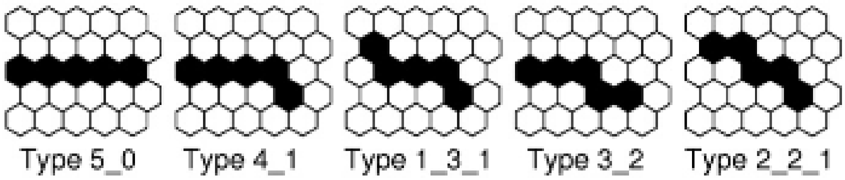


Figure 1.5: Fundamental shape of tracks considered straight. Image from [5].

from low-energy cosmic muons and not extensive air showers. This is reflected in the hierarchical structure of the triggers, which effectively reject such events. The overall event detection is split up into three tiers, T1, T2 and T3, where T3 implies the detection of an extensive air shower by either the FD or the SD (or both).

T1 trigger

T1 level triggers are implemented at the lowest possible level. This means each FD eye or each SD station raises T1 triggers autonomously. They serve as a first indicator on whether or not a signal of any kind is present. For the most part, this is realised by checking for elevated signal strengths, i.e. for hot pixels in a FD telescopes image or PMT outputs of an SD station that are significantly above baseline. The respective trigger thresholds are calibrated such that the nominal trigger rate during operation is roughly 100 Hz [5, 4].

T2 trigger

T2 level triggers occur at the same location as T1-type triggers. They are different in their more stringent conditions on the signal size or shape. This for example entails track shape identification for the FD telescopes, where straight tracks (see Figure 1.5) of hot pixels are identified. If the resulting pixel track passes an additional quality cut that rejects e.g. lightning signals, the T2 is directly promoted to a T3 trigger (= Event). For the SD, an exact discussion of T2 triggers is given in **??**. A single tank on average records T2-type events at a rate of 20 Hz and forwards this information to the CDAS along with a timestamp. There, incoming information of all tanks is scanned for spatial and temporal correlations, which indicate the presence of an extensive air shower.

T3 trigger

T3 type triggers, or event type triggers are (with the exception of FD events, which have been discussed above) built from distributions of at least three SD stations next to each other that recorded a T2 trigger in close temporal succession. Upon the detection of such a pattern a readout command is issued to all nearby stations. Their recorded FADC traces as well as calibration information are forwared to CDAS if the station observed a T1/T2 event within an appropriate timespan of order $O(\mu\mathrm{s})$ before or after the T3 pattern occurence. Such a modus operandi enables an accurate reconstruction of the shower footprint by including stations that did not participate in the initial T3 trigger. This extends to FD issued T3 triggers, where potential information from SD stations in the vicinity of the FD-reconstructed shower core position is requested.

1.3.2 Core position

All reconstruction algorithms presented in the following subsections rely in one form or the other on an accurate determination of where the shower was recorded above the observatory. Hence the center of the shower footprint, the shower core, must be estimated at the beginning of the analysis chain.

Without any prior knowledge, a first guess as to where the shower core is located can be made by calculating the barycenter of all participating stations. In this fashion, a weighted mean of all station locations is constructed with weights equal to the square root of the corresponding signal strength [4].

The presented approach fails if only parts of the shower are contained within the SD event. This occurs espically at the edges of the SD array, or in the vicinity of faulty WCDs. A fiducial trigger, NT5, is employed to mitigate this problem. NT5 requires at least N active stations around the SD detector that recorded the largest signal [18].

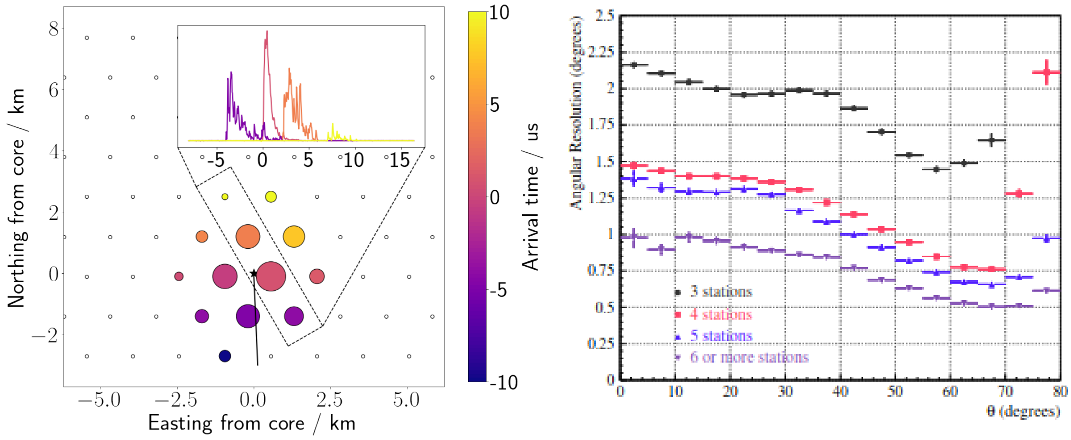
1.3.3 Arrival direction

The shower footprint measured by the SD (example given in Figure 1.6a) corresponds to the projection of the shower plane onto the detector plane, i.e. the ground. It can be assumed that the shower plane has a fixed (hyperbolic [19]) shape and propagates at the speed of light along the primary particles trajectory. With this knowledge, estimating the arrival direction becomes a task of minimizing the difference between measured and expected arrival times given by an example shower axis anchored at the reconstructed shower core. The axis for which the summed differences is minimal corresponds to the most likely arrival direction of the primary particle.

Naturally, the expected variance on the reconstructed ϕ and θ diminishes the more stations participate in the combined fit. The angular resolution thus decreases for larger energies of the primary particle. This can be seen in Figure 1.6b. In any case, the angular resolution even at smaller energies is better than 2.2°. For hybrid events, where the shower has also been detected by the FD, the angular resolution is greatly increased to about 0.6° [19].

1.3.4 Energy estimation

After a shower core has been pinpointed and the shower axis was determined, all information to fit an LDF is present. The **S**hower**P**lane **d**istance (SPD), the minimal separation between shower axis and station position, is calculated for each tank



(a) SD shower footprint **(b) Angular resolution**

Figure 1.6: (a) An example shower footprint recorded by the individual SD tanks (circles). The measured signal strength and arrival time is encoded in the size and color for each station. Tanks that haven't recorded any signal are shown colorless. For a subset of stations the respective VEM trace and consequently the propagation of the signal in the SD detector is shown in the inset plot on the top right. **(b)** The angular resolution as a function of θ for energies exceeding 3 EeV. Image from [19].

participating in the event and related to the integrated signal S (in units of Q_{VEM}) it received. From the so gathered lateral distribution of a shower the integrated signal S_{1000} of a hypothetical (labelled dense) station laying at SPD = 1000 m is obtained. This standardizes the comparison of results across many different events, even if a shower has only triggered few stations [20].

Due to attenuation effects in the atmosphere S_{1000} is a function of θ. It has been shown in [**DarkoCIC**] that by separating the θ dependence of a signal $S(\theta) = S \cdot A(\theta)$ and normalizing to a reference shower inclination, a reasonably unbiased estimator S_{38} can be recovered via a Constant Intensity Cut (CIC) as shown in Equation 1.5 and Equation 1.6.

$$f_{\mathrm{CIC}} = \frac{S_{1000}(\theta)}{S_{1000}(\theta_{\mathrm{ref}})} = \frac{A(\theta)}{A(\theta_{\mathrm{ref}})}. \tag{1.5}$$

The reference angle is chosen as $\theta_{\mathrm{ref}} = 38°$, as this is the median inclination of detected events [21]. It follows

$$S_{38} = \frac{S_{1000}(\theta)}{f_{\mathrm{CIC}}(\theta)}. \tag{1.6}$$

Further corrections are applied to S_{38} in order to counteract influence of the local weather or geomagnetic effects [21]. What remains is a shower parameter which has been sanitized as much as possible from any environmental factors, and which has a proportionality to the energy of the primary particle, as discussed in **??**.

The approximate relation of $S_{38}(E)$ can be inferred from hybrid measurements, where the calorimetric energy as measured by the FD is connected to S_{38}. From such datasets it follows the relation below, with the fit parameters for A, B as determined in [22].

$$E_{\mathrm{SD}} = A\,(S_{38}/\ \mathrm{VEM}_{\mathrm{Ch.}})^{B} \tag{1.7}$$

$$A = (1.86 \pm 0.03) \times 10^{17}\,\mathrm{eV}$$
$$B = 1.031 \pm 0.004$$

1.3.5 Primary particle

The determination of the primary particle, also referred to as the mass composition, relies on the systematic differences in air showers discussed in **??**. Muons, due to their noninteracting nature at high energies are typically the first signal to arrive in a WCD from an air shower. Since high-mass primaries produce a higher fraction of muons this implies that the rise time, in which the integrated signal goes from 10% to 50% of the total received signal, is shorter in these showers. Consequent analysis over a statistically relevant dataset thus reveals the mass composition of the cosmic ray flux (compare [23], **??**)

2 Classical station triggers

As mentioned in chapter 1, continously analyzing data sent to CDAS from each of the 1600 SD water tanks would quickly exceed the computational capabilites of Augers' main servers. For this purpose, trace information is only collected from a station, once a nearby T3 event (c.f. subsubsection 1.3.1) has been detected. The formation of a T3 trigger is dependant on several T2, or station-level, triggers, which will be discussed in detail in this chapter. First, the implementation of different trigger algorithms is discussed in section 2.1. Their performance is evaluated in section 2.2.

2.1 Implementation

2.1.1 Threshold trigger (Th)

The **Th**reshold trigger (Th) is the simplest, as well as longest operating trigger algorithm [24] in the field. It scans incoming ADC bins as measured by the three different WCD PMTs for values that exceed some threshold. If a coincident exceedance of this threshold is observed in all three WCD PMTs simultaneously, a Th-T1/2 trigger is issued. A pseudocode implementation of this algorithm is hence given by the below code block.

```
th1 = 1.75       // Th1 level threshold above baseline, in VEM
th2 = 3.20       // Th2 level threshold above baseline, in VEM

while True:

    pmt1, pmt2, pmt3 = get_next_output_from_WCD()

    if pmt1 <= th2 and pmt2 <= th2 and pmt3 <= th2:
        raise Th-T1_1trigger
    else if pmt1 <= th1 and pmt2 <= th1 and pmt3 <= th1:
        raise Th-T2_trigger
    else:
        continue
```

Logically, with increasing signal strength S in the PMTs, the likelihood of having observed an extensive air shower raises. This is reflected in the trigger level logic, where a coincident signal of $S \leq 3.20\,\mathrm{VEM}_{\mathrm{Peak}}$ is immediately forwarded to CDAS, whereas a signal $1.75\,\mathrm{VEM}_{\mathrm{Peak}} \geq S < 3.20\,\mathrm{VEM}_{\mathrm{Peak}}$ only raises a Th-T1 trigger. The algorithm is insensitive to signals that do not exceed at least $1.75\,\mathrm{VEM}_{\mathrm{Peak}}$ in all three PMTs.

In the case of faulty electronics, where only a subset of the WCD PMTs are available, the trigger thresholds (in units of VEM_{Peak}) are updated according to Table 2.1.

Table 2.1: Numerical values from [25]

n_{PMT}	Th-T2	Th-T1
1	5.00	2.85
2	3.60	2.00
3	3.20	1.75

2.1.2 Time over Threshold trigger (ToT)

The **T**ime **o**ver **T**hreshold trigger (ToT) is sensitive to much smaller signals than the Threshold trigger discussed in subsection 2.1.1. For each PMT in the water tank, the past 120 bins are examined for values that exceed 0.2 VEM_{Peak}. If 13 or more bins above the threshold are found in the window - ordering or succession do not matter - the PMT is considered to have an elevated pedestal. The ToT trigger requires at least two PMTs with an elevated pedestal in order to activate. As such, the algorithm is theoretically sensitive to events that deposit just 0.5 $VEM_{Ch.}$ A pseudocode example is given below.

```
threshold   = 0.2  // pedestal threshold, in VEM
n_bins      = 12   // number of bins above pedestal
window_size = 120  // considered window length

buffer_pmts = [[False for i in 1..window_size] for j in 1..3]
step_count = 0

while True:

    pmts = get_next_output_from_WCD()
    buffer_index = step_count % window_size
    count_active_PMTs = 0

    for pmt, buffer in pmts, buffers:
        if pmt <= threshold: buffer[buffer_index] = True

        if count_values(buffer, value = True) > n_bins:
            count_active_PMTs += 1

    if count_active_PMTs >= 2:
        raise ToT-T2_trigger
    else:
        step_count = buffer_index + 1
        continue
```

2.1.3 Time over Threshold deconvoluted trigger (Totd)

An extension to even lower signal strengths is given by the **ToT**-**d**econvoluted trigger (ToTd). As the name implies, the implementation of the algorithm is completely analog to the ToT trigger in subsection 2.1.2. Only the FADC input stream from the three PMTs is altered according to Equation 2.1.

$$d_i = (a_i - a_{i-1} \cdot e^{-\Delta t/\tau}) \,/\, (1 - e^{\Delta t/\tau}) \tag{2.1}$$

In Equation 2.1, the deconvoluted bin d_i is calculated from the measured FADC values a_i and a_{i-1}, where a_{i-1} is scaled according to an exponential decay with mean lifetime $\tau = 67\,\text{ns}$. This reduces the exponential tail of an electromagnetic signal to a series of pulses which in the case of $a_{i-1} < a_i$ exceed the original signal strength. As such, the deconvoluted trace can satisfy the ToT trigger requirements, whereas the original raw FADC values might not have, extending the sensitivity of the ToT trigger to lower signal strengths. The scaling constant $\Delta t = 25\,\text{ns}$ is tied to the sampling rate of UB electronics (c.f. subsection 1.2.1). The choice of the numerical constants τ and Δt is explained in more detail in [26].

2.1.4 Multiplicity of Positive Steps (MoPS)

The **M**ultiplicity **o**f **P**ositive **S**teps (MoPS) algorithm triggers on positive flanks of an FADC trace, which can be related to the arrival of new particles in the water tank.

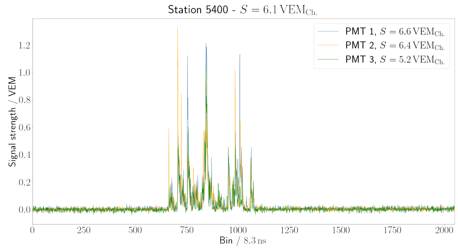
A positive flank in the FADC trace of a single PMT is any combination of at least three bins that are monotonically increasing in value, Once such a positive step has been identified, a (MoPS) trigger veto is applied to the next

$$n_{\text{skip}} = \lfloor \big(\log_2(\Delta y) + 1\big) - 3 \rceil \tag{2.2}$$

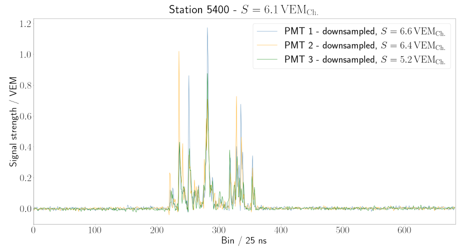
bins, where Δy refers to the total vertical increase in the step from first to last bin. Note that in Equation 2.2 the notation $\lfloor x \rceil$ is used as shorthand notation to round x to the nearest integer. If Δy is bigger than $y_{\text{min}} = 3\,\text{ADC}$ (to filter random fluctuations), but does not exceed $y_{\text{max}} = 31\,\text{ADC}$ (to prevent triggering on muonic coincidences), it is added to a ledger. If the number of rising flanks in the ledger is bigger than $m > 4$ for at least two PMTs, a final check regarding the integral of the FADC trace is performed. If this check passes, a MoPS-T2 trigger is issued to CDAS.

It is impossible to accurately recreate the MoPS trigger in simulations. The integral test above compares the sum of the last 250 bins against a threshold ($\sum a_i > 75$). Since not all 250 bin values are available to CDAS, differing results are to be expected when comparing the implementation of the algorithm in the SD field versus its' counterpart in analysis software.

For this purpose, the MoPs trigger is not considered in the analysis presented in chapter 4. The implications of this choice are layed out in section 2.2.



(a) UUB time trace



(b) UB time trace

Figure 2.1: (a) A simulated signal as it would appear to UUB electronics. The ionizing particles originating in the extensive air shower hit the tank around bin 660 ($\approx 5.5\,\mu$s). **(b)** The same signal but filtered and downsampled to emulate UB electronics.

2.1.5 Compatibility mode

Altough the triggers discussed in the previous subsections are meant to function completely autonomously in the SD field, their implementation requires some prior knowledge of the signal one desires to detect. For their use in the Auger observatory, several hyperparameters such as the thresholds of the Th-Trigger, or the window size of the ToT-trigger have been determined in studies ([27], [25], [28]).

These studies were conducted using the predecessor, the **U**nified **B**oard (UB), of the hardware that is being installed during the AugerPrime upgrade of the observatory. Most importantly, the **U**pgraded **U**nified **B**oard (UUB) has a sampling rate that is three times larger (120 MHz) than that of UB electronics (40 MHz). Not only does this raise the number of bins in a standard time trace from 682 to $2^{11} = 2048$, but also drastically reduces the efficiency (in particular for ToT-like triggers) of the above discussed algorithms. Whereas a new FADC bin is measured every 25 ns in a UB station, the triggers would receive a new input every ≈ 8.3 ns in a UUB setting.

The modus operandi elected by the Pierre Auger collaboration to circumvent this problem is to emulate UB electronics using the UUB electronics. This means that measured FADC bins are to be filtered and downsampled before any trigger runs over them. Software implementations by which this is achieved are listed in **??**. The effect the filtering and downsampling has on measured data is visualized in Figure 2.1.

[to do: comment on accuracy fo this method]

2.2 Performance

The performance of a trigger can be evaluated in many different ways. In the most general consideration, a confusion matrix holds information about the ability of a classifier to discern between different types, or classes, C. With the example at hand there exist two types of events one wishes to distinguish, a signal event C_1 in the form of an extensive air shower, versus background C_0. The confusion matrix thus becomes:

		Predicted C	
		C_1	C_0
True C	C_1	True positive (TP)	False negative (FN)
	C_0	False positive (TP)	True negative (TN)

From this, other potentially interesting variables can be derived. Of particular interest for the Auger observatory are the sensitivity and **F**alse **D**iscovery **R**ate (FDR). The former is the probability that a signal event will be classified correctly, i.e. an extensive air shower hits a water tank and raises a T2 trigger. The sensitivity - in the following also called the trigger efficiency ϵ - is defined as

$$\epsilon = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{2.3}$$

The latter is a measure of how readily the triggers (wrongly) identify background events like stray cosmic muons as extensive air showers. It is imperative for any trigger algorithm operating in the SD to minimize this probability. Simply due to the number of operating stations in the field, a small increase in FDR drastically raise the amount of potential events and hence load on the central analysis server of the observbatory.

$$\mathrm{FDR} = \frac{\mathrm{FP}}{\mathrm{TP} + \mathrm{FP}}. \tag{2.4}$$

[to do: continue writing about performance]

3 Neural network training data

Over their relatively brief existance, neural networks have been shown to perform increasingly impressive tasks (e.g. [29], [30], and many more). However, they learn by example. The performance of a neural network is directly linked to the input data it receives during training. If the training data is not an accurate example of real world information a network later operates on, insight gained from it is at best an approximation, and at worst completely randomly generated data.

As such, it is not a question *if* some neural network architecture can learn to identify an extensive air shower from WCD data, but rather which implementation, fed with which information, does. For this purpose, this chapter explains the procedure with which training data is generated. As stated above, this must occur with a focus on being representative of data actually measured in the SD array. The elected approach to create time traces is modularized. The structure of this chapter reflects this. First, general comments about the characteristics of background data (i.e. the WCD detector response in the absence of an extensive air shower) are made in section 3.1. Next, the process to extract signal originating from CRs is detailed in section 3.2. Lastly, building the time trace from the aforementioned modules and drawing samples from it for a neural network to train on is done in section 3.3.

3.1 Background dataset

While a flux of partices causes elevated ADC levels in both the HG and LG channels of a WCD PMT during a shower event, the lack of such a phenomenon does not imply the readout information is uniformly flat. Instead, it hovers around the channels' baseline (c.f. section 1.2) with occasional spikes upwards due to low-energy particles impinging on the detector. Coupled with electronic noise from the many digital components in the UUB, this constitutes the data that is collected inbetween air shower events.

3.1.1 Accidental muons

3.1.2 Random traces

3.2 Signal dataset

3.3 Trace building & Sliding window analysis

4 Performance of neural network triggers

4.1 Convolutional neural networks

4.2 Recurrent neral networks

Bibliography

[1] Alexander Aab et al. "The pierre auger observatory upgrade-preliminary design report". In: *arXiv preprint arXiv:1604.03637* (2016).

[2] The Pierre Auger Collaboration. *Timeline of the Pierre Auger Observatory*. `https://auger.org/observatory/timeline-observatory`. Accessed: 04th Oct. 2022. 2022.

[3] Carola Dobrigkeit et al. "PC Status Report". Auger internal document, 13th November 2022.

[4] 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.

[5] 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.

[6] Darko Veberič. *Auger array*. `https://web.ikp.kit.edu/darko/auger/auger-array/auger_array-pdf/auger_array-ad.pdf`. Accessed: 19th Jul. 2022. 2021.

[7] The Pierre Auger Collaboration. *AugerPrime*. `https://auger.org/observatory/augerprime`. Accessed: 06th Nov. 2022.

[8] João de Mello Neto, ed. *Physics and astrophysics of ultra-high energy cosmic rays: recent results from the Pierre Auger Observatory*. NUCLEUS - 2020.

[9] Antonella Castellina. "AugerPrime: the Pierre Auger observatory upgrade". In: *EPJ Web of Conferences*. Vol. 210. EDP Sciences. 2019, p. 06002.

[10] 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.

[11] Valerio Verzi et al. "The energy scale of the Pierre Auger Observatory". In: *Proceedings of the 33rd ICRC, Rio de Janeiro, Brasil* (2013).

[12] Xavier Bertou et al. "Local-Station Calibration". GAP 2022-0??

[13] Tobias Schulz et al. "New Baseline Algorithm for UB Traces". GAP 2022-045.

[14] Tobias Schulz et al. "New Baseline Algorithm for UUB Traces". GAP 2023-007.

[15] Alexander Streich et al. "Performance of the upgraded surface detector stations of the Pierre Auger Observatory". In: *Verhandlungen der Deutschen Physikalischen Gesellschaft* (2018).

[16] P.S. Allison et al. "Surface Detector calibration in the Engineering Array". GAP 2002-028. July 2002.

[17] Anthony M Hillas. "The origin of ultra-high-energy cosmic rays". In: *Annual review of astronomy and astrophysics* 22 (1984), pp. 425–444.

[18] J Abraham et al. "Trigger and aperture of the surface detector array of the Pierre Auger Observatory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 613.1 (2010), pp. 29–39.

[19] C Bonifazi. "Angular resolution of the pierre auger observatory". In: *29th International Cosmic Ray Conference (ICRC29), Volume 7*. Vol. 7. 2005, p. 17.

[20] Thomas K Gaisser and A Michael Hillas. "Reliability of the method of constant intensity cuts for reconstructing the average development of vertical showers". In: *International Cosmic Ray Conference*. Vol. 8. 1977, pp. 353–357.

[21] Alexander Aab et al. "Measurement of the cosmic-ray energy spectrum above 2.5× 10 18 eV using the Pierre Auger Observatory". In: *Physical Review D* 102.6 (2020), p. 062005.

[22] A. Aab et al. "Measurement of the cosmic-ray energy spectrum above 2.5×10^{18} eV using the Pierre Auger Observatory". In: *Phys. Rev. D* 102 (6 Sept. 2020), p. 062005. DOI: `10.1103/PhysRevD.102.062005`. URL: `https://link.aps.org/doi/10.1103/PhysRevD.102.062005`.

[23] Antoine Letessier-Selvon. "Highlights from the Pierre Auger Observatory". In: *Brazilian Journal of Physics* 44.5 (2014), pp. 560–570.

[24] David Nitz. "Surface Detector Trigger Operating Guide". GAP 2006-057.

[25] Alan Coleman. "The new trigger settings". GAP 2018-001.

[26] Pierre Billoir. "Peak Searching in FADC traces". GAP 2002-076.

[27] Xavier Bertou et al. "Calibration of the surface array of the Pierre Auger Observatory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 568.2 (2006), pp. 839–846.

[28] Pierre Billoir. "Proposition to improve the local trigger of Surface Detector for low energy showers". GAP 2009-179.

[29] OpenAI et al. *Dota 2 with Large Scale Deep Reinforcement Learning*. 2019. arXiv: `1912.06680 [cs.LG]`.

[30] OpenAI. *GPT-4 Technical Report*. 2023. arXiv: `2303.08774 [cs.CL]`.